



Many-Body Physics: From Kondo to Hubbard

Eva Pavarini, Erik Koch and Piers Coleman (Eds.)

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

German Research School for
Simulation Sciences GmbH

**Lecture Notes of the Autumn School on
Correlated Electrons 2015**

Eva Pavarini, Erik Koch, and Piers Coleman (Eds.)

Many-Body Physics: From Kondo to Hubbard

Autumn School organized by
the Forschungszentrum Jülich
and the German Research School
for Simulation Sciences

at Forschungszentrum Jülich
21 – 25 September 2015

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation

Band / Volume 5

ISSN 2192-8525

ISBN 978-3-95806-074-6

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 2015

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 5

ISSN 2192-8525
ISBN 978-3-95806-074-6

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

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

Contents

Preface

1. Heavy Fermions and the Kondo Lattice: A 21st Century Perspective
Piers Coleman
2. Spectroscopic Properties of Mixed-Valence Compounds in the Impurity Model
Kurt Schönhammer
3. Magnetism in Correlated Matter
Eva Pavarini
4. The Kondo Model and Poor Man's Scaling
Andriy Nevidomskyy
5. Numerical Renormalization Group and Multi-Orbital Kondo Physics
Theo Costi
6. NRG with Bosons
Kevin Ingersent
7. Frustrated Spin Systems
Frédéric Mila
8. Introduction to Mean-Field Theory of Spin Glass Models
Václav Janiš
9. The Slave-Boson Approach to Correlated Fermions
Raymond Frésard
10. The Lanczos Method
Erik Koch
11. The Hubbard Model and its Properties
Andreas Mielke
12. The Two-Dimensional Hubbard Model
Robert Eder
13. Quantum Cluster Methods: CPT and CDMFT
David Sénéchal
14. The Dynamical Cluster Approximation and its DCA⁺ Extension
Thomas Maier
15. Electronic Structure of Perovskites: Lessons from Hybrid Functionals
Cesare Franchini

Index

Preface

Many-body physics has the daunting task of describing the collective behavior of vast assemblies of elementary objects. While the fundamental laws are known, exact solutions like the Bethe Ansatz are exceedingly rare. Nonetheless, the past century has witnessed a continuous stream of conceptual breakthroughs, prompted by unforeseen discoveries of new states of matter: superconductivity and superfluidity, antiferromagnetism, the Kondo effect, the Mott transition, symmetry breaking, spin glasses and frustration, heavy Fermions, and high-temperature superconductivity. Each of these cooperative phenomena is an example of emergence at work. Their essence can often be captured by simple model Hamiltonians. Describing the richness of real matter requires, however, to increase the complexity of the models significantly, as emergent phenomena are frequently governed by the interplay of several scales. In this year's school we will highlight the Kondo effect, the physics of the Hubbard model, and frustrated quantum spins, covering the range from fundamental mechanisms to the modeling of real materials.

The aim of the school is to introduce advanced graduate students and up to the essence of emergence and modern approaches for modeling strongly correlated matter.

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 and the German Research School for Simulation Sciences at the Forschungszentrum Jülich provided the major part of the funding and were vital for the organization of the school and the production of this book. The DFG Research Unit FOR 1346 generously supported many of the attending students and the poster session. The Institute for Complex Adaptive Matter (ICAM) offered travel grants for selected international 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. D. Mans of the Graphische 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 proof-reading the manuscripts, often on quite short notice: Michael Baumgärtel, Khaldoon Ghanem, Hoai Le Thi, Julian Mußhoff, Esmaeel Sarvestani, Amin Kiani Sheikhabadi, Guoren Zhang, Qian Zhang, and, in particular, our native speaker Hunter Sims.

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, and Piers Coleman

August 2015

1 Heavy Fermions and the Kondo Lattice: A 21st Century Perspective

Piers Coleman

Center for Materials Theory, Rutgers University

136 Frelinghuysen Road, Piscataway NJ 08854, USA

Contents

1 Heavy electrons	2
1.1 Introduction	2
1.2 Local moments and the Kondo effect	2
1.3 The Kondo lattice	6
2 Kondo insulators: the simplest heavy fermions	10
3 Large-N expansion for the Kondo Lattice	11
3.1 Philosophy and formulation	11
3.2 Mean-field theory	16
3.3 Free energy and saddle point	18
3.4 The composite nature of the f -electron	21
3.5 Cooper pair analogy	22
4 Heavy-fermion superconductivity	23
4.1 Symplectic spins and $SP(N)$	24
4.2 Superconductivity in the Kondo-Heisenberg model	25
5 Topological Kondo insulators	27
6 Coexisting magnetism and Kondo effect	30

1 Heavy electrons

1.1 Introduction

In a world where it is possible to hold a levitated high-temperature superconductor in the palm of one's hand, it is easy to forget the ongoing importance of low-temperature research. Heavy-electron materials are a class of strongly correlated electron material containing localized magnetic moments that, by entangling with the surrounding electrons, profoundly transform the metallic properties. A heavy-fermion metal can develop electron masses 1000 times that of copper; it can also develop unconventional superconductivity, transform into new forms of quantum order, exhibit quantum critical and topological behavior. Although most of these properties develop well below the boiling point of nitrogen, the diversity and highly tunable nature of their ground states make them an invaluable work-horse for exploring and researching the emergent properties of correlated quantum matter.

This lecture will give an introduction to heavy-fermion materials, trying to emphasize a 21st century perspective. More extensive discussion and development of the ideas in these notes can be found in an earlier review article [1] and the later chapters of my book *Introduction to Many-Body Physics* [2].

In the periodic table, the most strongly interacting electrons reside in orbitals that are well localized. In order of increasing localization, partially filled orbitals are ordered as follows:

$$5d < 4d < 3d < 5f < 4f. \quad (1)$$

In addition, when moving along a row of the periodic table, the increasing nuclear charge pulls the orbitals towards the nucleus. These trends are summarized in the *Kmetko-Smith diagram* [3], which is shown in Fig 1. The *d*-orbital metals at the bottom left of this diagram are highly itinerant and exhibit conventional superconductivity. By contrast, in rare earth and actinide metals towards the top right-hand corner, the *f*-shell electrons are localized, forming magnets or antiferromagnets. It is the materials that lie in the cross-over between these two regions that are particularly interesting, for these materials are “on the brink of magnetism.” It is in this cross-over region that many strongly correlated materials reside: It is here for instance that we find cerium and uranium, which are key atoms for a wide range of *4f* and *5f* heavy-electron materials.

1.2 Local moments and the Kondo effect

Heavy-electron materials contain a lattice of localized electrons immersed in a sea of mobile conduction electrons. To understand their physics, we need to first step back and discuss individual localized moments and the mechanism by which they interact with the surrounding conduction sea.

The key feature of a localized moment is that the Coulomb interaction has eliminated the high-frequency charge fluctuations, leaving behind a low-energy manifold of degenerate spin states. In rare earth and actinide ions, the orbital and spin angular momentum combine into a single

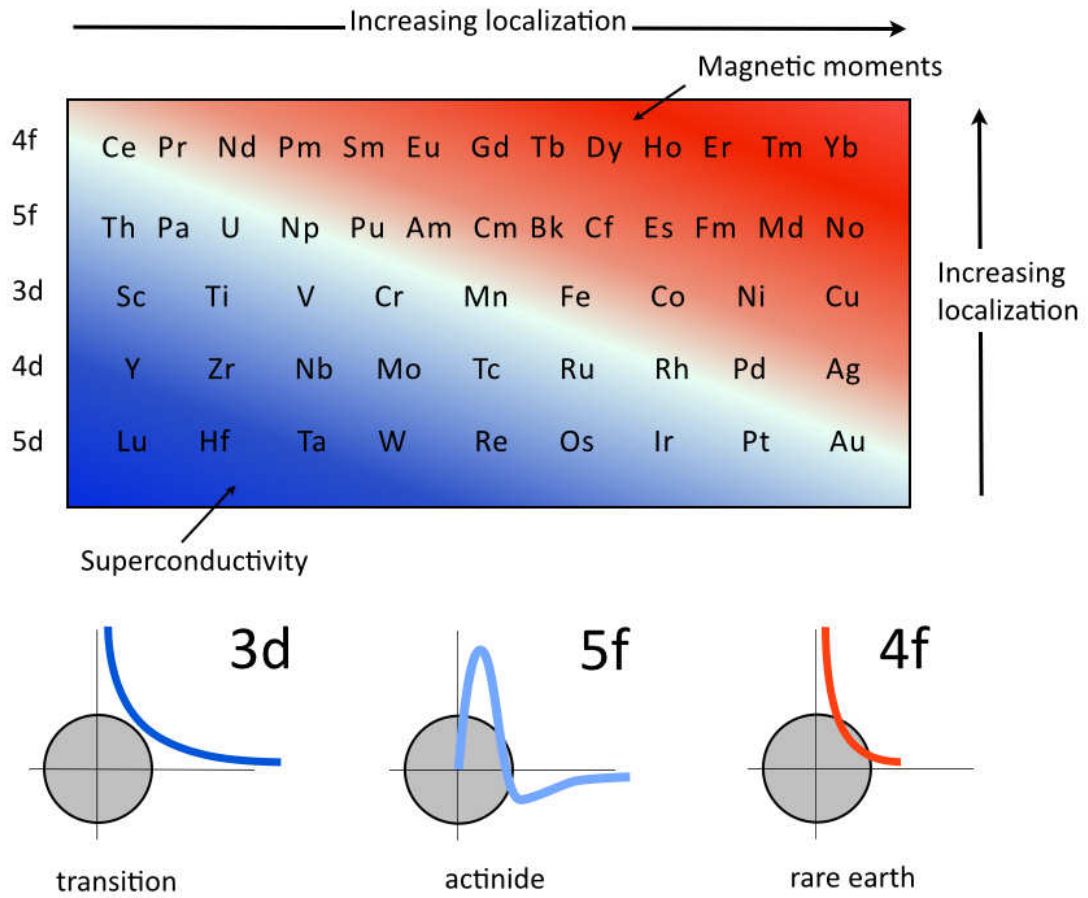


Fig. 1: The Kmetko-Smith diagram [3], showing the broad trends towards increasing electron localization in the d- and f-electron compounds.

entity with angular momentum $\vec{j} = \vec{l} + \vec{s}$. For example, a Ce^{3+} ion contains a single unpaired 4f-electron in the state $4f^1$, with $l = 3$ and $s = 1/2$. Spin-orbit coupling gives rise to a low-lying multiplet with $j = 3 - \frac{1}{2} = \frac{5}{2}$, consisting of $2j + 1 = 6$ degenerate orbitals $|4f^1 : Jm\rangle$, ($m_J \in [-\frac{5}{2}, \frac{5}{2}]$) with an associated magnetic moment $M = 2.64 \mu_B$. In a crystal, the $2j + 1$ -fold degeneracy of such a magnetic ion is split, and provided there are an odd number of electrons in the ion, the Kramers theorem guarantees that the lowest lying state has at least a two-fold degeneracy (Fig. 2 a and b).

One of the classic signatures of localized moments is a high-temperature Curie-Weiss susceptibility, given by

$$\chi \approx \frac{M^2}{3(T + \theta)} n_i \quad M^2 = g^2 \mu_B^2 j(j + 1), \quad (2)$$

where n_i is the concentration of magnetic moments, while M is the magnetic moment with total angular momentum quantum number j and gyromagnetic ratio g (g -factor). θ is the *Curie-Weiss* temperature, a phenomenological scale that takes account of interactions between spins.

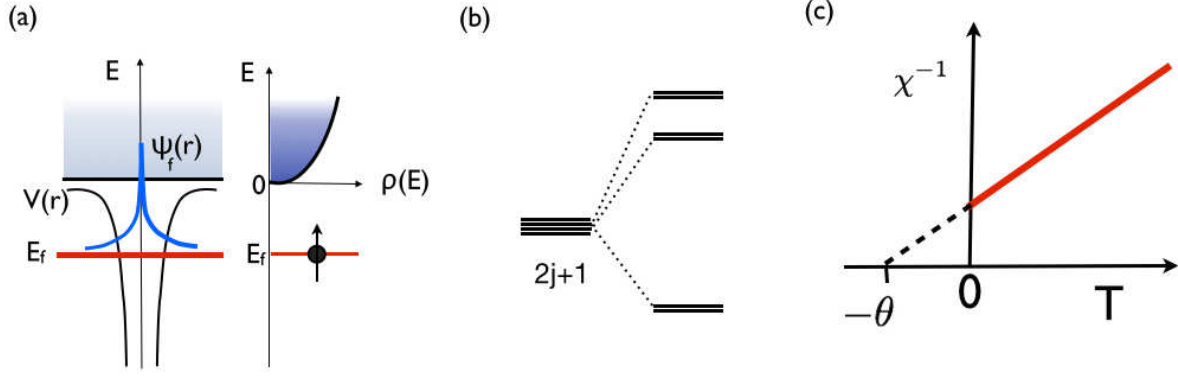


Fig. 2: *a) In isolation, the localized states of an atom form a stable, sharp excitation lying below the continuum. (b) In a crystal, the $2j + 1$ -fold degenerate state splits into multiplets, typically forming a low-lying Kramers doublet. (c) The inverse of the Curie-Weiss susceptibility of local moments χ^{-1} is a linear function of temperature, intersecting zero at $T = -\theta$.*

The presence of such local moments inside a metal profoundly alters its properties. The physics of an isolated magnetic ion is described by the *Kondo model*

$$H = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \overbrace{J \psi^\dagger(0) \vec{\sigma} \psi(0) \cdot \vec{S}_f}^{\Delta H}, \quad (3)$$

where $c_{k\sigma}^\dagger$ creates a conduction electron of energy ε_k and momentum k , and $\psi^\dagger(0) = \mathcal{N}_s^{-1/2} \sum_k c_{k\sigma}^\dagger$ creates a conduction electron at the origin, where $\mathcal{N}_s$ is the number of sites in the lattice. The conduction sea interacts with the local moment via an antiferromagnetic contact interaction of strength J . The antiferromagnetic sign ($J > 0$) of this interaction is an example of *superexchange*, first predicted by Philip W. Anderson [4, 5], which results from high-energy valence fluctuations. Jun Kondo [6] first analyzed the effect of this scattering, showing that, as the temperature is lowered, the effective strength of the interaction grows logarithmically, according to

$$J \rightarrow J(T) = J + 2J^2 \rho \ln \frac{D}{T} \quad (4)$$

where ρ is the density of states of the conduction sea (per spin) and D is the band width. The growth of this interaction enabled Kondo to understand why in many metals at low temperatures the resistance starts to rise as the temperature is lowered, giving rise to a *resistance minimum*. Today, we understand this logarithmic correction as a renormalization of the Kondo coupling constant, resulting from the fact that, as the temperature is lowered, more and more high-frequency quantum spin fluctuations become coherent, and these strengthen the Kondo interaction. The effect is closely analogous to the growth of the strong interaction between quarks, and like quarks, the local moment in the Kondo effect is *asymptotically free* at high energies. However, as you can see from the above equation, once the temperature becomes of the order

$$T_K \sim D \exp \left[-\frac{1}{2J\rho} \right]$$

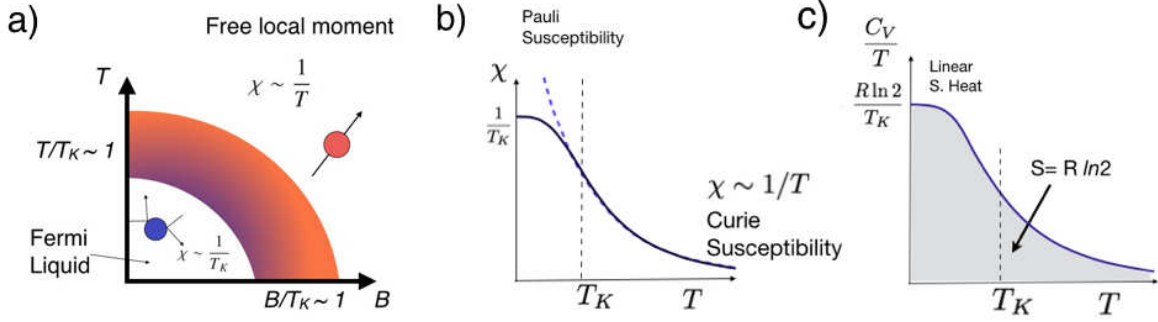


Fig. 3: (a) Schematic temperature-field phase diagram of the Kondo effect. At fields and temperatures large compared with the Kondo temperature T_K , the local moment is unscreened with a Curie susceptibility. At temperatures and fields small compared with T_K , the local moment is screened, forming an elastic scattering center within a Landau-Fermi liquid with a Pauli susceptibility $\chi \sim \frac{1}{T_K}$. (b) Schematic susceptibility curve for the Kondo effect, showing the cross-over from Curie susceptibility at high temperatures to Pauli susceptibility at temperatures below the Kondo temperature T_K . (c) Specific heat curve for the Kondo effect. Since the total area is the full spin entropy $R \ln 2$ and the width is of order T_K , the height must be of order $\gamma \sim R \ln 2 / T_K$. This sets the scale for the zero-temperature specific heat coefficient.

the correction becomes as large as the original perturbation, and at lower temperatures, the Kondo interaction can no longer be treated perturbatively. In fact, non-perturbative methods tell us that this interaction scales to strong coupling at low energies, causing electrons in the conduction sea to magnetically screen the local moment to form an inert *Kondo singlet* denoted by

$$|GS\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right), \quad (5)$$

where the thick arrow refers to the spin state of the local moment and the thin arrow refers to the spin state of a bound electron at the site of the local moment. The key features of the impurity Kondo effect are

- The electron fluid surrounding the Kondo singlet forms a Fermi liquid, with a Pauli susceptibility $\chi \sim 1/T_K$.
- The local moment is a kind of qubit that entangles with the conduction sea to form a singlet. As the temperature T is raised, the entanglement entropy converts to thermal entropy, given by the integral of the specific heat coefficient,

$$S(T) = \int_0^T dT' \frac{C_V(T')}{T'}.$$

Since the total area under the curve, $S(T \rightarrow \infty) = R \ln 2$ per mole, is the high-temperature spin entropy, and since the characteristic width is the Kondo temperature, it follows that the characteristic zero-temperature specific heat coefficient must be of the order of the inverse Kondo temperature: $\gamma = \frac{C_V}{T}(T \rightarrow 0) \sim R \ln 2 / T_K$ (see Fig. 3b).

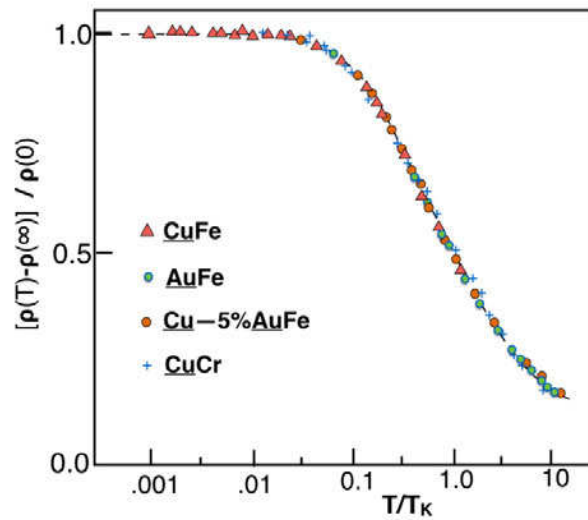


Fig. 4: Temperature dependence of resistivity associated with scattering from an impurity spin from [7, 8]. The resistivity saturates at the unitarity limit at low temperatures, due to the formation of the Kondo resonance. Adapted from [7].

- The only scale in the physics is T_K . For example, the resistivity created by magnetic scattering off the impurity has a universal temperature dependence

$$\frac{R(T)}{R_U} = n_i \Phi\left(\frac{T}{T_K}\right) \quad (6)$$

where n_i is the concentration of magnetic impurities, $\Phi(x)$ is a universal function and ρ_U is the unit of *unitary resistance* (basically resistance with a scattering rate of order the Fermi energy)

$$R_U = \frac{2ne^2}{\pi m\rho}. \quad (7)$$

Experiment confirms that the resistivity in the Kondo effect can indeed be scaled onto a single curve that fits forms derived from the Kondo model (see Fig 4).

- The scattering off the Kondo singlet is resonantly confined to a narrow region of order T_K , called the *Kondo* or *Abrikosov-Suhl* resonance.

1.3 The Kondo lattice

In a heavy-fermion material, containing a lattice of local moments, the Kondo effect develops *coherence*. In a single impurity, a Kondo singlet scatters electrons without conserving momentum, giving rise to a huge build-up of resistivity at low temperatures. However, in a lattice with translational symmetry, this same elastic scattering now conserves momentum, and this leads to coherent scattering off the Kondo singlets. In the simplest heavy-fermion metals, this leads to a dramatic reduction in the resistivity at temperatures below the Kondo temperature.

As a simple example, consider CeCu_6 a classic heavy-fermion metal. Naively, CeCu_6 is just a copper alloy in which 14% of the copper atoms are replaced by cerium, yet this modest replacement radically alters the metal. In this material, it actually proves possible to follow the development of coherence from the dilute single-ion Kondo limit to the dense Kondo lattice by forming the alloy $\text{La}_{1-x}\text{Ce}_x\text{Cu}_6$. Lanthanum is isoelectronic to cerium but has an empty f shell, so the limit $x \rightarrow 0$ corresponds to the dilute Kondo limit, and in this limit the resistivity follows the classic Kondo curve. However, as the concentration of cerium increases, the resistivity curve starts to develop a coherence maximum and in the concentrated limit drops to zero with the characteristic T^2 dependence of a Landau-Fermi liquid (see Fig. 6).

CeCu_6 displays the following classic features of a heavy-fermion metal:

- A Curie-Weiss susceptibility $\chi \sim (T + \theta)^{-1}$ at high temperatures.
- A paramagnetic spin susceptibility $\chi \sim \text{const.}$ at low temperatures.
- A dramatically enhanced linear specific heat $C_V = \gamma T$ at low temperatures, where in CeCu_6 $\gamma \sim 1000 \text{ mJ/mol/K}^2$ is about 1000 times larger than in copper.
- A quadratic temperature dependence of the low-temperature resistivity $\rho = \rho_o + AT^2$.

In a Landau-Fermi liquid, the magnetic susceptibility χ and the linear specific heat coefficient $\gamma = C_V/T|_{T \rightarrow 0}$ are given by

$$\chi = (\mu_B)^2 \frac{N^*(0)}{1 + F_0^a} \quad (8)$$

$$\gamma = \frac{\pi^2 k_B^2}{3} N^*(0) \quad (9)$$

where $N^*(0) = \frac{m^*}{m} N(0)$ is the renormalized density of states, and F_0^a is the spin-dependent part of the s -wave interaction between quasiparticles. One of the consequences of Fermi-liquid theory is that the density of states factors out of the Sommerfeld or Wilson ratio between the susceptibility and linear specific heat coefficient,

$$W = \frac{\chi}{\gamma} = \left(\frac{\mu_B}{2\pi k_B} \right)^2 \frac{1}{1 + F_0^a}. \quad (10)$$

In heavy-fermion metals, this ratio remains approximately fixed across several decades of variation in χ and γ . This allows us to understand heavy-fermion metals as a lattice version of the Kondo effect that gives rise to a renormalized density of states $N^*(0) \sim 1/T_K$.

The discovery of heavy-electron compounds in the 1970s led Mott [9] and Doniach [10] to propose that heavy-electron systems should be modeled as a *Kondo lattice*, where a dense array of local moments interacts with the conduction sea via an antiferromagnetic interaction J . In such a lattice, the local moments polarize the conduction sea, and the resulting Friedel oscillations in the magnetization give rise to an antiferromagnetic RKKY (Ruderman Kittel Kasuya Yosida) magnetic interaction [11–13] that tends to order the local moments. Mott and Doniach realized that this interaction must compete with the Kondo effect.

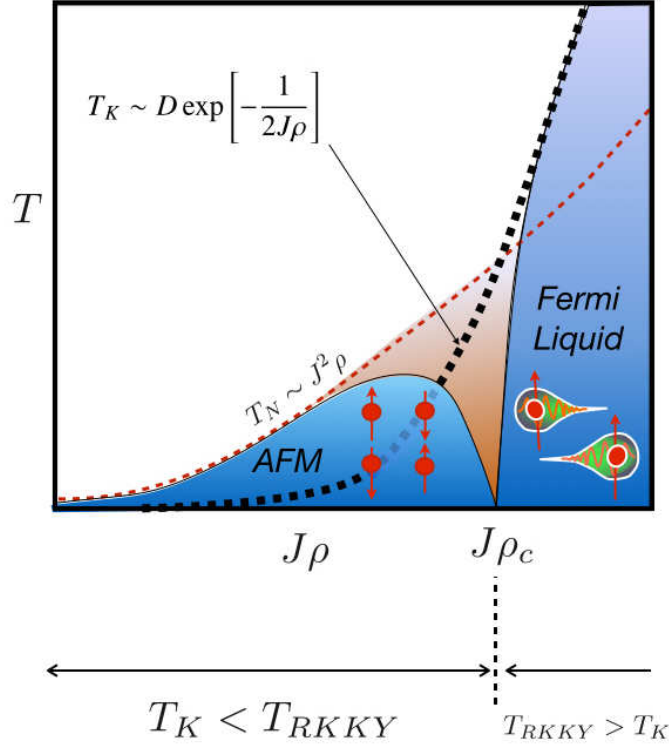


Fig. 5: Doniach phase diagram for the Kondo lattice, illustrating the antiferromagnetic regime and the heavy-fermion regime for $T_K < T_{\text{RKKY}}$ and $T_K > T_{\text{RKKY}}$ respectively. The effective Fermi temperature of the heavy Fermi liquid is indicated as a solid line. Experimental evidence suggests that in many heavy-fermion materials this scale drops to zero at the antiferromagnetic quantum critical point.

The simplest Kondo lattice Hamiltonian [14] is

$$H = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + J \sum_j \vec{S}_j \cdot c_{j\alpha}^\dagger \vec{\sigma}_{\alpha\beta} c_{j\beta}, \quad (11)$$

where

$$c_{j\alpha}^\dagger = \frac{1}{\sqrt{N_s}} \sum_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger e^{-i\mathbf{k} \cdot \mathbf{R}_j} \quad (12)$$

creates an electron at site j . Mott and Doniach [9, 10] pointed out that there are two energy scales in the Kondo lattice: the Kondo temperature $T_K \sim D e^{-1/(2J\rho)}$ and the RKKY scale $E_{\text{RKKY}} = J^2 \rho$.

For small $J\rho$, $E_{\text{RKKY}} \gg T_K$ leading to an antiferromagnetic ground state, but when $J\rho$ is large, $T_K \gg E_{\text{RKKY}}$, stabilizing a ground state in which every site in the lattice resonantly scatters electrons. Based on a simplified one-dimensional *Kondo necklace* model [15], Doniach conjectured [10] that the transition between the antiferromagnet and the dense Kondo ground state is a continuous quantum phase transition. Experiment confirms this conjecture, and today we have several examples of such quantum critical points, including CeCu_6 doped with gold to form $\text{CeCu}_{6-x}\text{Au}_x$ and CeRhIn_5 under pressure [16–18]. In the fully developed Kondo lattice, the ground state Bloch's theorem ensures that the resonant elastic scattering at each site will

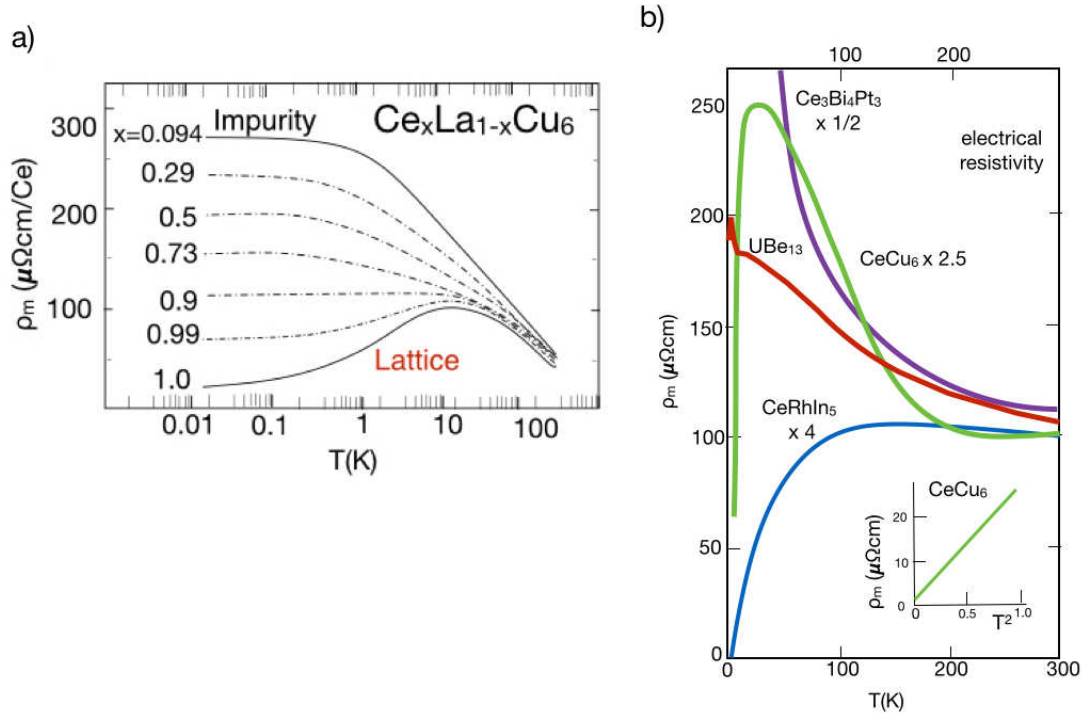


Fig. 6: (a) Resistivity of $Ce_xLa_{1-x}Cu_6$. Dilute Ce atoms in $LaCu_6$ exhibit a classic Kondo resistivity, but as the Ce concentration becomes dense, elastic scattering off each Ce atom leads to the development of a coherent heavy-fermion metal. (b) Resistivities of four heavy-fermion materials showing the development of coherence. A variety of antiferromagnetic, Fermi liquid, superconducting and insulating states are formed (see text).

generate a renormalized f band of width $\sim T_K$. In contrast with the impurity Kondo effect, here elastic scattering at each site acts coherently. For this reason, as the heavy-electron metal develops at low temperatures, its resistivity drops towards zero (see Fig. 6b).

In a Kondo lattice, spin entanglement is occurring on a truly macroscopic scale, but this entanglement need not necessarily lead to a Fermi liquid. Experimentally, many other possibilities are possible. Here are some examples:

- $Ce_3Bi_4Pt_3$, a Kondo insulator in which the formation of Kondo singlets with the Ce moments drives the development of a small insulating gap at low temperatures.
- $CeRhIn_5$, an antiferromagnet on the brink of forming a Kondo lattice, which under pressure becomes a heavy-fermion superconductor with $T_c = 2$ K.
- UBe_{13} a heavy-fermion superconductor that transitions directly from an incoherent metal with resistivity $200 \mu\Omega \text{ cm}$ into a superconducting state.

Each of these materials has qualitatively the same high-temperature Curie-Weiss magnetism and the same Kondo resistivity at high temperatures due to incoherent scattering off the local moments. However at low temperatures, the scattering off the magnetic Ce ions becomes coherent and new properties develop.

2 Kondo insulators: the simplest heavy fermions

In many ways, the Kondo insulator is the simplest ground state of the Kondo lattice. The first Kondo insulator (KI) SmB_6 was discovered almost fifty years ago [19], and today there are several known examples, including $\text{Ce}_3\text{Bi}_4\text{Pt}_3$. At room temperature, these KIs are metals containing a dense array of magnetic moments, yet on cooling they develop a narrow gap due the formation of *Kondo singlets* that screen the local moments [20–23]. We can gain a lot of insight by examining the strong-coupling limit, in which the dispersion of the conduction sea is much smaller than the Kondo coupling J . Consider a simple tight-binding Kondo lattice

$$H = -t \sum_{(i,j)\sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}) + J \sum_{j,\alpha\beta} \vec{\sigma}_j \cdot \vec{S}_j, \quad \vec{\sigma}_j \equiv (c_{j\beta}^\dagger \vec{\sigma}_{\beta\alpha} c_{j\alpha}) \quad (13)$$

in which $t/J \ll 1$ is a small parameter. In this limit, the inter-site hopping is a perturbation to the on-site Kondo interaction,

$$H \xrightarrow{t/J \rightarrow 0} J \sum_{j,\alpha\beta} \vec{\sigma}_j \cdot \vec{S}_j + \mathcal{O}(t), \quad (14)$$

and the corresponding ground state shows the formation of a spin singlet at each site, denoted by the wavefunction

$$|KI\rangle = \prod_j \frac{1}{\sqrt{2}} \left(\uparrow_j \downarrow_j - \downarrow_j \uparrow_j \right) \quad (15)$$

where the double and single arrows denote the localized moment and conduction electron respectively.

Each singlet has a ground-state energy $E = -3J/2$ per site and a singlet-triplet spin gap of magnitude $\Delta E = 2J$. Moreover, if we remove an electron from site i , we break a Kondo singlet and create an unpaired spin with excited energy $3J/2$,

$$|\text{qp}^+, i \uparrow\rangle = \uparrow_i \prod_{j \neq i} \frac{1}{\sqrt{2}} \left(\uparrow_j \downarrow_j - \downarrow_j \uparrow_j \right) = \sqrt{2} c_{i\downarrow} |KI\rangle, \quad (16)$$

as illustrated in Fig 7(a). Similarly, if we add an electron, we create an electron quasiparticle, corresponding to an unpaired local moment and a doubly occupied conduction electron orbital

$$|\text{qp}^-, i \uparrow\rangle = \uparrow_i \left(\uparrow_i \downarrow_i \right) \prod_{j \neq i} \frac{1}{\sqrt{2}} \left(\uparrow_j \downarrow_j - \downarrow_j \uparrow_j \right) = \sqrt{2} c_{j\uparrow}^\dagger |KI\rangle, \quad (17)$$

as illustrated in Fig 7b.

If we now reintroduce the hopping $-t$ between sites, then these quasiparticle excitations become mobile, as illustrated in Fig. 7 a and b. From the explicit form of the states, we find that the nearest-neighbor hopping matrix elements are $\langle \text{qp}^\pm, i\sigma | H | \text{qp}^\pm, j\sigma \rangle = \pm t/2$, giving quasiparticle energies

$$E_{\text{qp}^\pm}(\mathbf{k}) = \pm t(c_x + c_y + c_z) + \frac{3}{2}J. \quad (18)$$

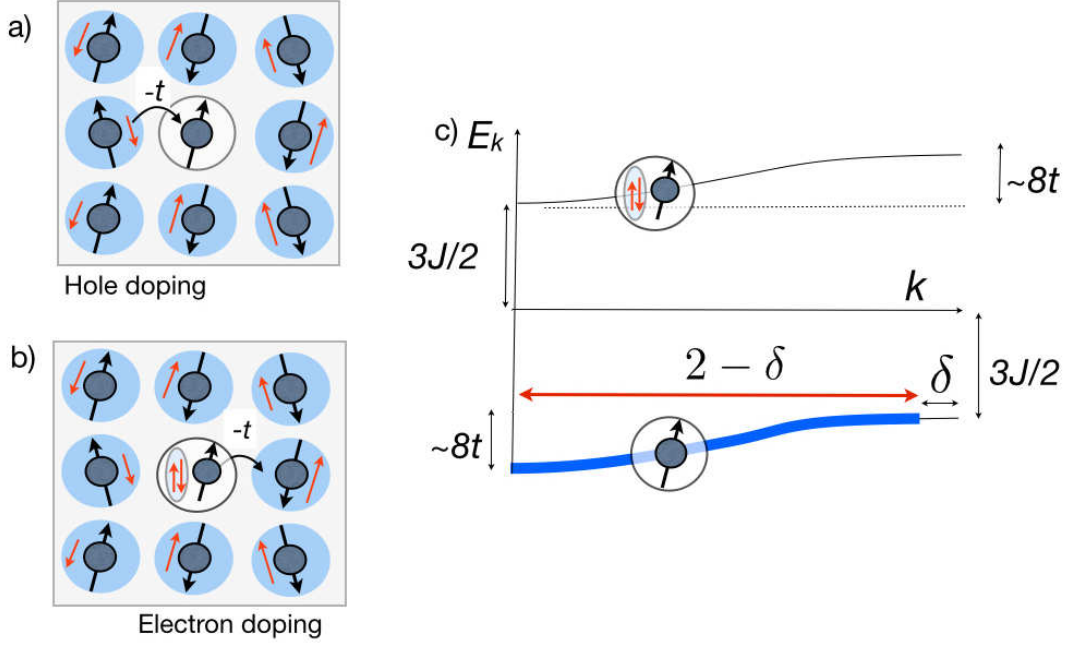


Fig. 7: Showing (a) hole and (b) electron doping of a strong-coupling Kondo insulator. (c) Dispersion of strong-coupling Kondo insulator. A small amount of hole doping δ gives rise to a large Fermi surface containing $2 - \delta$ heavy electrons.

To transform from the quasiparticle to the electron basis, we need to reverse the sign of the hole (qp^+) dispersion to obtain the valence band dispersion, so that the band energies predicted by the strong-coupling limit of the Kondo lattice are

$$E_{\mathbf{k}}^{\pm} = -t(c_x + c_y + c_z) \pm \frac{3}{2}J, \quad (19)$$

separated by an energy $3J$ as shown in Fig. 7c. Note that these are *hard core* fermions that cannot occupy the same lattice site simultaneously.

In this way, the half-filled strong coupling Kondo lattice forms an insulator with a charge gap of size $3J$ and a spin gap of size $2J$. Notice finally that if we dope the insulator with an amount δ of holes, we form a band of heavy fermions. In this way, Kondo insulators can be considered the parent states of heavy-electron materials. However, we would like to examine the physics of a Kondo lattice at weak coupling, and to do this requires a different approach.

3 Large- N expansion for the Kondo Lattice

3.1 Philosophy and formulation

One of the great difficulties with the Kondo lattice is that there is no natural small parameter to carry out an approximate treatment. One way around this difficulty is to use a large- N expansion, in which we extend the number of spin components of the electrons from 2 to N . Historically, Anderson [24] pointed out that the large spin-orbit coupling in heavy-fermion compounds generates (if we ignore crystal fields) a large spin degeneracy $N = 2j + 1$, furnishing a small

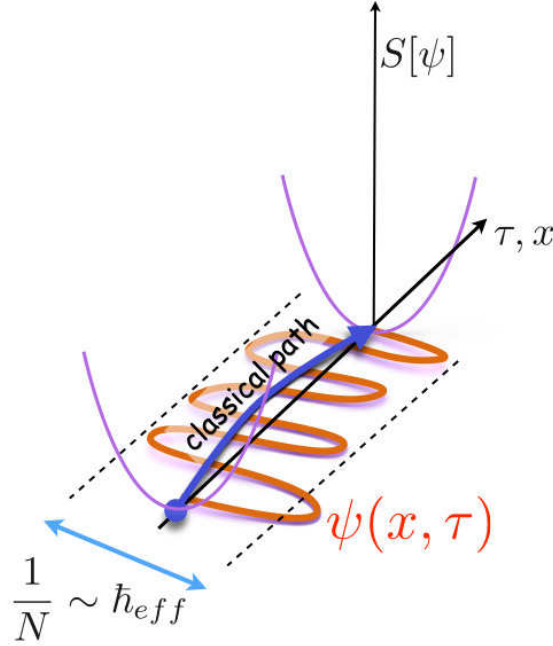


Fig. 8: Illustration of the convergence of a quantum path integral about a semi-classical trajectory in the large- N limit.

parameter $1/N$ for a controlled expansion about the limit $N \rightarrow \infty$. One of the observations arising from Anderson’s idea [25, 26] is that the RKKY interaction becomes negligible (of order $\mathcal{O}(1/N^2)$) in this limit, and the Kondo lattice ground state becomes stable. This observation opened the way to path-integral mean-field treatments of the Kondo lattice [26–31].

The basic idea of the large- N limit is to examine a limit where every term in the Hamiltonian grows extensively with N . In the path integral for the partition function, the corresponding action then grows extensively with N , so that

$$Z = \int \mathcal{D}[\psi] e^{-NS} = \int \mathcal{D}\psi \exp \left[-\frac{S}{1/N} \right] \equiv \int \mathcal{D}[\psi] \exp \left[-\frac{S}{\hbar_{\text{eff}}} \right]. \quad (20)$$

Here $1/N \sim \hbar_{\text{eff}}$ behaves as an effective Planck constant for the theory, focusing the path integral into a non-trivial “semi-classical” or “mean-field” solution as $\hbar_{\text{eff}} \rightarrow 0$. As $N \rightarrow \infty$, the quantum fluctuations of intensive variables $\hat{a}$, such as the electron density per spin, become smaller and smaller, scaling as $\langle \delta a^2 \rangle / \langle a^2 \rangle \sim 1/N$, causing the path integral to focus around a non-trivial mean-field trajectory. In this way, one can obtain new results by expanding around the solvable large- N limit in powers of $1/N$ (Fig. 8).

We will use a simplified Kondo lattice model introduced by Read and Newns [26], in which all electrons have a spin degeneracy $N = 2j + 1$,

$$H = \sum_{\mathbf{k}\alpha} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} + \frac{J}{N} \sum_{j,\alpha\beta} c_{j\beta}^\dagger c_{j\alpha} S_{\alpha\beta}(j). \quad (21)$$

where $c_{j\alpha}^\dagger = \frac{1}{\sqrt{N_s}} \sum_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger e^{-i\mathbf{k} \cdot \mathbf{R}_j}$ creates an electron localized at site j , and the spin of the local moment at position $\mathbf{R}_j$ is represented by pseudo-fermions

$$S_{\alpha\beta}(j) = f_{j\alpha}^\dagger f_{j\beta} - \frac{n_f(j)}{N} \delta_{\alpha\beta}. \quad (22)$$

This representation requires that we set a value for the conserved f occupancy $n_f(j) = Q$ at each site. This interaction can be rewritten in a factorized form

$$H = \sum_{\mathbf{k}\alpha} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} - \frac{J}{N} \sum_{j,\alpha\beta} : \left(c_{j\beta}^\dagger f_{j\beta} \right) \left(f_{j\alpha}^\dagger c_{j\alpha} \right) : \quad (23)$$

Read-Newns model for the Kondo lattice

where the potential scattering terms resulting from the rearrangement of the f -operators have been absorbed into a shift of the chemical potential. Note that

- the model has a global $SU(N)$ symmetry associated with the conserved magnetization.
- the Read-Newns (RN) model is a lattice version of the Coqblin-Schrieffer Hamiltonian [32] introduced to describe the Kondo interaction in strongly spin-orbit coupled rare-earth ions. While the Coqblin-Schrieffer interaction is correct at each site, the assumption that the $SU(N)$ spin is conserved by electron hopping is an oversimplification. (This is a price one pays for a solvable model.)
- in this factorized form, the antiferromagnetic Kondo interaction is attractive.
- the coupling constant has been scaled to vary as J/N to ensure that the interaction grows extensively with N . The interaction involves a product of two terms that scale as $O(N)$, so that $J/N \times O(N^2) \sim O(N)$.
- the RN model also has a **local gauge invariance**: The absence of f charge fluctuations allows us to change the phase of the f -electrons *independently* at each site

$$f_{j\sigma} \rightarrow e^{i\phi_j} f_{j\sigma}. \quad (24)$$

A tricky issue concerns the value we give to the conserved charge $n_f = Q$. In the physical models of interest, $n_f = 1$ at each site, so one might be inclined to explicitly maintain this condition. However, the large- N expansion requires that the action is extensive in N , and this forces us to consider more general classes of solutions where Q scales with N so that the filling factor $q = Q/N$ is finite as $N \rightarrow \infty$. Thus if we are interested in a Kramers-doublet Kondo model, we take the half-filled case $q = 1/2$, $Q = N/2$, but if we want to understand a $j = 7/2$ Yb^{3+} atom without crystal fields, then in the physical system $N = 2j + 1 = 8$, and we should fix $q = Q/N = 1/8$.

The partition function for the Kondo lattice is then

$$Z = \text{Tr} \left[e^{-\beta H} \prod_j \delta(\hat{n}_f(j) - Q) \right] \quad (25)$$

where $\delta(\hat{n}_f(j) - Q)$ projects out the states with $n_f(j) = Q$ at site j . By re-writing the delta function as a Fourier transform, the partition function can be rewritten as a path-integral,

$$Z = \int \mathcal{D}[\psi^\dagger, \psi, \lambda] \exp \left[- \int_0^\beta d\tau \overbrace{(\psi^\dagger \partial_\tau \psi + H[\bar{\psi}, \psi, \lambda])}^{L[\psi^\dagger, \psi, \lambda]} \right] \quad (26)$$

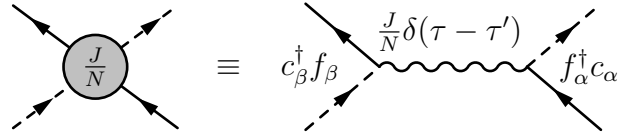
where $\psi^\dagger \equiv (\{c^\dagger\}, \{f^\dagger\})$ schematically represent the conduction and f -electron fields,

$$H[\lambda] = \sum_{\mathbf{k}\alpha} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} - \frac{J}{N} \sum_{j,\alpha\beta} : (c_{j\beta}^\dagger f_{j\beta}) (f_{j\alpha}^\dagger c_{j\alpha}) : + \sum_j \lambda_j (n_{fj} - Q). \quad (27)$$

The field λ_j is a fluctuating Lagrange multiplier that enforces the constraint $n_j = Q$ at each site. Next we carry out a Hubbard-Stratonovich transformation on the interaction,

$$- \frac{J}{N} \sum_{\alpha\beta} (c_{j\beta}^\dagger f_{j\beta}) (f_{j\alpha}^\dagger c_{j\alpha}) \rightarrow \sum_\alpha \left[\bar{V}_j (c_{j\alpha}^\dagger f_{j\alpha}) + (f_{j\alpha}^\dagger c_{j\alpha}) V_j \right] + N \frac{\bar{V}_j V_j}{J}. \quad (28)$$

In the original Kondo model, we started out with an interaction between electrons and spins. Now, by carrying out the Hubbard-Stratonovich transformation, we have formulated the interaction as the exchange of a charged boson



$$\equiv \quad (29)$$

$$- \frac{J}{N} \sum_{\mathbf{k}, \mathbf{k}', \alpha, \beta} (c_{\beta}^\dagger f_{\beta}) (f_{\alpha}^\dagger c_{\alpha}) \quad (30)$$

where the solid lines represent the conduction electron propagators, and the dashed lines represent the f -electron operators. Notice how the bare amplitude associated with the exchange boson is frequency independent, i.e., the interaction is instantaneous. Physically, we may interpret this exchange process as due to an intermediate valence fluctuation.

The path integral now involves an additional integration over the hybridization fields V and $\bar{V}$

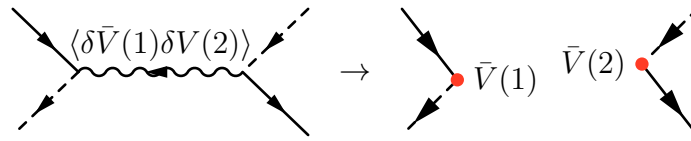
$$Z = \int \mathcal{D}[\bar{V}, V, \lambda] \int \mathcal{D}[\psi^\dagger, \psi] \exp \left[- \int_0^\beta (\psi^\dagger \partial_\tau \psi + H[\bar{V}, V, \lambda]) \right] \quad (31)$$

$$H[\bar{V}, V, \lambda] = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_j \left[\bar{V}_j (c_{j\sigma}^\dagger f_{j\sigma}) + (f_{j\sigma}^\dagger c_{j\sigma}) V_j + \lambda_j (n_{fj} - Q) + N \frac{\bar{V}_j V_j}{J} \right]$$

Read-Newns path integral for the Kondo lattice

where we have suppressed summation signs for repeated spin indices (summation convention).

The RN path integral allows us to develop a mean-field description of the many-body Kondo scattering processes that captures the physics and is asymptotically exact as $N \rightarrow \infty$. In this approach, the condensation of the hybridization field describes the formation of bound states between spins and electrons that cannot be dealt with in perturbation theory. Bound states induce long-range temporal correlations in scattering: Once the hybridization condenses, the interaction lines break-up into independent anomalous scattering events, denoted by



The hybridization V in the RN action carries the local $U(1)$ gauge charge of the f -electrons, giving rise to an important local gauge invariance:

$$f_{j\sigma} \rightarrow e^{i\phi_j} f_{j\sigma}, \quad V_j \rightarrow e^{i\phi_j} V_j, \quad \lambda_j \rightarrow \lambda_j - i\dot{\phi}_j(\tau). \quad (32)$$

Read Newns gauge transformation

This invariance can be used to choose a gauge in which V_j is real by absorbing the phase of the hybridization $V_j = |V_j|e^{i\phi_j}$ into the f -electron. In the radial gauge,

$$Z = \int \mathcal{D}[|V|, \lambda] \int \mathcal{D}[\psi^\dagger, \psi] \exp \left[- \int_0^\beta \overbrace{(\psi^\dagger \partial_\tau \psi + H[|V|, \lambda])}^{S[|V|, \lambda, \psi^\dagger, \psi]} \right]$$

$$H[|V|, \lambda] = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_j \left[|V_j| (c_{j\sigma}^\dagger f_{j\sigma} + f_{j\sigma}^\dagger c_{j\sigma}) + \lambda_j (n_{fj} - Q) + N \frac{|V_j|^2}{J} \right] \quad (33)$$

Read Newns path integral: radial gauge

Subsequently, when we use the radial gauge, we will drop the moduli signs. The interesting feature about this Hamiltonian is that with the real hybridization, the conduction and f electrons now transform under a single global $U(1)$ gauge transformation, i.e the f electrons have become *charged*.

3.2 Mean-field theory

The interior fermion integral in the path integral (33) defines an effective action $S_E[V, \lambda]$ by the relation

$$Z_E = \exp[-NS_E[V, \lambda]] \equiv \int \mathcal{D}[\psi^\dagger, \psi] \exp[-S[V, \lambda, \psi^\dagger, \psi]] , \quad (34)$$

The extensive growth of the effective action with N means that at large N , the integration in (31) is dominated by its stationary points, allowing us to dispense with the integrals over V and λ .

$$Z = \int \mathcal{D}[\lambda, V] \exp[-NS_E[V, \lambda]] \approx \exp[-NS_E[V, \lambda]] \Big|_{\text{Saddle Point}} \quad (35)$$

In practice, we seek uniform, static solutions, $V_j(\tau) = V$, $\lambda_j(\tau) = \lambda$. In this case the saddle-point partition function $Z_E = \text{Tr} e^{-\beta H_{\text{MFT}}}$ is simply the partition function of the static mean-field Hamiltonian

$$\begin{aligned} H_{\text{MFT}} &= \sum_{\mathbf{k}\sigma} \left(c_{\mathbf{k}\sigma}^\dagger, f_{\mathbf{k}\sigma}^\dagger \right) \overbrace{\begin{pmatrix} \varepsilon_{\mathbf{k}} & V \\ \bar{V} & \lambda \end{pmatrix}}^{\underline{h}(\mathbf{k})} \begin{pmatrix} c_{\mathbf{k}\sigma} \\ f_{\mathbf{k}\sigma} \end{pmatrix} + N\mathcal{N}_s \left(\frac{|V|^2}{J} - \lambda q \right) \\ &= \sum_{\mathbf{k}\sigma} \psi_{\mathbf{k}\sigma}^\dagger \underline{h}(\mathbf{k}) \psi_{\mathbf{k}\sigma} + N\mathcal{N}_s \left(\frac{|V|^2}{J} - \lambda q \right). \end{aligned} \quad (36)$$

Here, $f_{\mathbf{k}\sigma}^\dagger = \frac{1}{\sqrt{N_s}} \sum_j f_{j\sigma}^\dagger e^{i\mathbf{k} \cdot \mathbf{R}_j}$ is the Fourier transform of the f -electron field and we have introduced the two-component notation

$$\psi_{\mathbf{k}\sigma} = \begin{pmatrix} c_{\mathbf{k}\sigma} \\ f_{\mathbf{k}\sigma} \end{pmatrix}, \quad \psi_{\mathbf{k}\sigma}^\dagger = \left(c_{\mathbf{k}\sigma}^\dagger, f_{\mathbf{k}\sigma}^\dagger \right), \quad \underline{h}(\mathbf{k}) = \begin{pmatrix} \varepsilon_{\mathbf{k}} & V \\ \bar{V} & \lambda \end{pmatrix}. \quad (37)$$

We should think about H_{MFT} as a renormalized Hamiltonian, describing the low-energy quasi-particles moving through a self-consistently determined array of resonant scattering centers. Later, we will see that the f -electron operators are composite objects, formed as bound states between spins and conduction electrons.

The mean-field Hamiltonian can be diagonalized in the form

$$H_{\text{MFT}} = \sum_{\mathbf{k}\sigma} \left(a_{\mathbf{k}\sigma}^\dagger, b_{\mathbf{k}\sigma}^\dagger \right) \begin{pmatrix} E_{\mathbf{k}^+} & 0 \\ 0 & E_{\mathbf{k}^-} \end{pmatrix} \begin{pmatrix} a_{\mathbf{k}\sigma} \\ b_{\mathbf{k}\sigma} \end{pmatrix} + N\mathcal{N}_s \left(\frac{\bar{V}V}{J} - \lambda q \right). \quad (38)$$

Here $a_{\mathbf{k}\sigma}^\dagger = u_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger + v_{\mathbf{k}} f_{\mathbf{k}\sigma}^\dagger$ and $b_{\mathbf{k}\sigma}^\dagger = -v_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger + u_{\mathbf{k}} f_{\mathbf{k}\sigma}^\dagger$ are linear combinations of $c_{\mathbf{k}\sigma}^\dagger$ and $f_{\mathbf{k}\sigma}^\dagger$ playing the role of quasiparticle operators with corresponding energy eigenvalues

$$\det \left[E_{\mathbf{k}^\pm} \mathbb{1} - \begin{pmatrix} \varepsilon_{\mathbf{k}} & V \\ \bar{V} & \lambda \end{pmatrix} \right] = (E_{\mathbf{k}^\pm} - \varepsilon_{\mathbf{k}})(E_{\mathbf{k}^\pm} - \lambda) - |V|^2 = 0, \quad (39)$$

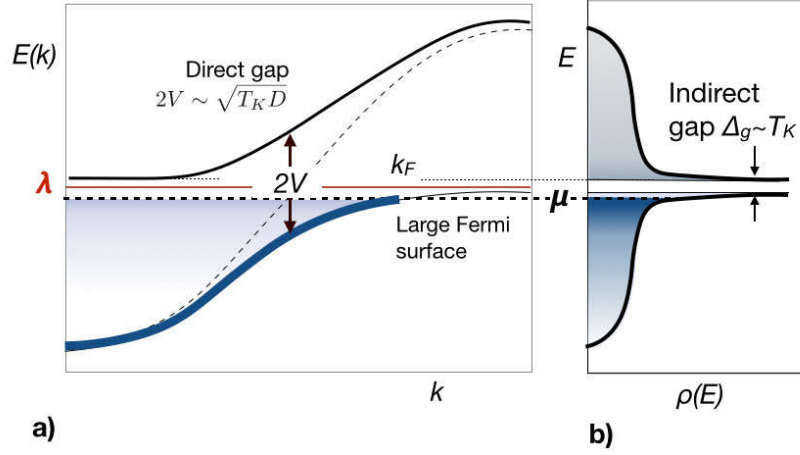


Fig. 9: (a) Dispersion for the Kondo lattice mean-field theory. (b) Renormalized density of states, showing a “hybridization gap” (Δ_g).

or

$$E_{\mathbf{k}\pm} = \frac{\varepsilon_{\mathbf{k}} + \lambda}{2} \pm \left[\left(\frac{\varepsilon_{\mathbf{k}} - \lambda}{2} \right)^2 + |V|^2 \right]^{\frac{1}{2}}, \quad (40)$$

and eigenvectors taking the BCS form

$$\begin{Bmatrix} u_{\mathbf{k}} \\ v_{\mathbf{k}} \end{Bmatrix} = \left[\frac{1}{2} \pm \frac{(\varepsilon_{\mathbf{k}} - \lambda)/2}{2\sqrt{\left(\frac{\varepsilon_{\mathbf{k}} - \lambda}{2}\right)^2 + |V|^2}} \right]^{\frac{1}{2}}. \quad (41)$$

The hybridized dispersion described by these energies is shown in Fig. 9.

Note that

- The Kondo effect injects an f band into the conduction sea, hybridizing with the conduction band to create two bands separated by a *direct* “hybridization gap” of size $2V$ and a much smaller *indirect* gap. If we put $\varepsilon_{\mathbf{k}} = \pm D$, we see that the upper and lower edges of the gap are given by

$$E^{\pm} = \frac{\mp D + \lambda}{2} \pm \sqrt{\left(\frac{\mp D - \lambda}{2} \right)^2 + V^2} \approx \lambda \pm \frac{V^2}{D}, \quad (D \gg \lambda) \quad (42)$$

so the indirect gap has a size $\Delta_g \sim 2V^2/D$, where D is the half bandwidth. We will see shortly that $V^2/D \sim T_K$ is basically the single-ion Kondo temperature, so that $V \sim \sqrt{T_K D}$ is the geometric mean of the bandwidth and Kondo temperature.

- In the case when the chemical potential lies in the gap, a *Kondo insulator* is formed.
- A conduction sea of electrons has been transformed into a heavy Fermi sea of holes.
- The Fermi surface volume *expands* in response to the formation of heavy electrons (see Fig. 10) to accommodate the total number of occupied quasiparticle states

$$N_{tot} = \left\langle \sum_{\mathbf{k}\lambda\sigma} n_{\mathbf{k}\lambda\sigma} \right\rangle = \langle \hat{n}_f + \hat{n}_c \rangle \quad (43)$$

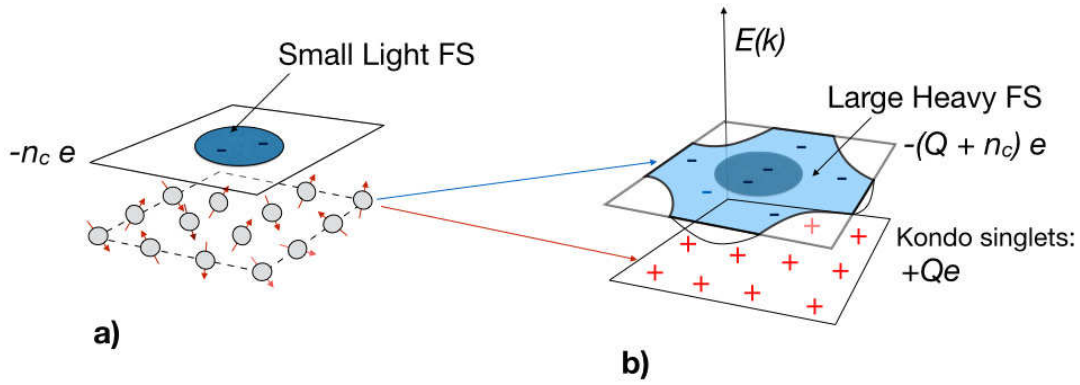


Fig. 10: (a) High-temperature state: small Fermi surface with a background of spins; (b) Low-temperature state: large Fermi surface develops against a background of positive charge. Each spin “ionizes” into Q heavy electrons, leaving behind a background of Kondo singlets, each with charge $+Qe$.

where $n_{k\lambda\sigma} = a_{k\lambda\sigma}^\dagger a_{k\lambda\sigma}$ is the number operator for the quasiparticles and n_c is the total number of conduction electrons. This means

$$N_{tot} = N \frac{V_{FS} a^3}{(2\pi)^3} = Q + n_c, \quad (44)$$

where a^3 is the volume of the unit cell. This is rather remarkable, for the expansion of the Fermi surface implies an increased *negative* charge density in the Fermi sea. Since charge is conserved, we are forced to conclude there is a compensating $+Q|e|$ charge density per unit cell provided by the Kondo singlets formed at each site, as illustrated in Fig. 10.

3.3 Free energy and saddle point

Let us now use the results of the last section to calculate the mean-field free energy F_{MFT} and determine self-consistently the parameters λ and V that set the scales of the Kondo lattice. By diagonalizing the mean-field Hamiltonian, we obtain

$$\frac{F}{N} = -T \sum_{\mathbf{k}, \pm} \ln \left[1 + e^{-\beta E_{\mathbf{k}\pm}} \right] + \mathcal{N}_s \left(\frac{V^2}{J} - \lambda q \right). \quad (45)$$

Let us discuss the ground state, in which only the lower band contributes to the free energy. As $T \rightarrow 0$, we can replace $-T \ln(1 + e^{-\beta E_{\mathbf{k}}}) \rightarrow \theta(-E_{\mathbf{k}}) E_{\mathbf{k}}$, so the ground-state energy $E_0 = F(T=0)$ involves an integral over the occupied states of the lower band:

$$\frac{E_0}{N\mathcal{N}_s} = \int_{-\infty}^0 dE \rho^*(E) E + \left(\frac{V^2}{J} - \lambda q \right), \quad (46)$$

where we have introduced the density of heavy-electron states $\rho^*(E) = \sum_{\mathbf{k}, \pm} \delta(E - E_{\mathbf{k}}^{(\pm)})$. Now by (39) the relationship between the energy E of the heavy electrons and the energy ε of the conduction electrons is

$$E = \varepsilon + \frac{V^2}{E - \lambda}.$$

As we sum over momenta $\mathbf{k}$ within a given energy shell, there is a one-to-one correspondence between each conduction electron state and each quasiparticle state, so we can write $\rho^*(E)dE = \rho(\varepsilon)d\varepsilon$, where the density of heavy electron states

$$\rho^*(E) = \rho \frac{d\varepsilon}{dE} = \rho \left(1 + \frac{V^2}{(E - \lambda)^2} \right). \quad (47)$$

Here we have approximated the underlying conduction electron density of states by a constant $\rho = 1/(2D)$. The originally flat conduction electron density of states is now replaced by a hybridization gap, flanked by two sharp peaks of width approximately $\pi\rho V^2 \sim T_K$ (Fig. 9). Note that the lower bandwidth is lowered by an amount $-V^2/D$. With this information, we can carry out the integral over the energies to obtain

$$\frac{E_o}{N\mathcal{N}_s} = \rho \int_{-D-V^2/D}^0 dE E \left(1 + \frac{V^2}{(E - \lambda)^2} \right) + \left(\frac{V^2}{J} - \lambda q \right), \quad (48)$$

where we have assumed that the upper band is empty and that the lower band is partially filled. Carrying out the integral we obtain

$$\frac{E_o}{N\mathcal{N}_s} = -\frac{\rho}{2} \left(D + \frac{V^2}{D} \right)^2 + \frac{\Delta}{\pi} \int_{-D}^0 dE \left(\frac{1}{E - \lambda} + \frac{\lambda}{(E - \lambda)^2} \right) + \left(\frac{V^2}{J} - \lambda q \right) \quad (49)$$

$$= -\frac{D^2\rho}{2} + \frac{\Delta}{\pi} \ln \left(\frac{\lambda}{D} \right) + \left(\frac{V^2}{J} - \lambda q \right) \quad (50)$$

where we have replaced $\Delta = \pi\rho V^2$, which is the width of an isolated f -resonance, and have dropped terms of order $\mathcal{O}(\Delta^2/D)$. We can rearrange this expression, absorbing the bandwidth D and the Kondo coupling constant into a single Kondo temperature $T_K = De^{-1/J\rho}$ as follows

$$\frac{E_o}{N\mathcal{N}_s} = -\frac{D^2\rho}{2} + \frac{\Delta}{\pi} \ln \left(\frac{\lambda}{D} \right) + \left(\frac{\pi\rho V^2}{\pi\rho J} - \lambda q \right) \quad (51)$$

$$= -\frac{D^2\rho}{2} + \frac{\Delta}{\pi} \ln \left(\frac{\lambda}{D} \right) + \left(\frac{\Delta}{\pi\rho J} - \lambda q \right) \quad (52)$$

$$= -\frac{D^2\rho}{2} + \frac{\Delta}{\pi} \ln \left(\frac{\lambda}{De^{-\frac{1}{J\rho}}} \right) - \lambda q \quad (53)$$

$$= -\frac{D^2\rho}{2} + \frac{\Delta}{\pi} \ln \left(\frac{\lambda}{T_K} \right) - \lambda q. \quad (54)$$

This describes the energy of a family of Kondo lattice models with different $J(D)$ and cutoff D but fixed Kondo temperature. If we impose the constraint $\frac{\partial E_o}{\partial \lambda} = \langle n_f \rangle - Q = 0$ we obtain $\frac{\Delta}{\pi\lambda} - q = 0$, so

$$\frac{E_o(V)}{N\mathcal{N}_s} = \frac{\Delta}{\pi} \ln \left(\frac{\Delta}{\pi q e T_K} \right) - \frac{D^2\rho}{2}, \quad (\Delta = \pi\rho|V|^2) \quad (55)$$

Let us pause for a moment to consider this energy functional qualitatively. There are two points to be made:

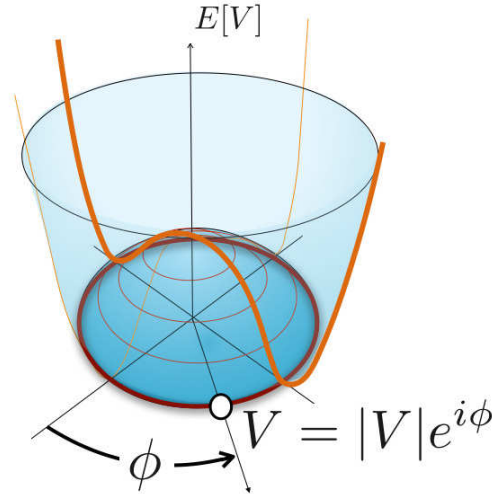


Fig. 11: Mexican hat potential for the Kondo Lattice, evaluated at constant $\langle n_f \rangle = Q$ as a function of a complex hybridization $V = |V|e^{i\phi}$

- The energy surface $E_0(V)$ is actually independent of the phase of $V = |V|e^{i\phi}$ (see Fig. 11), and has the form of Mexican hat at low temperatures. The minimum of this functional will then determine a family of saddle-point values $V = |V_o|e^{i\phi}$, where ϕ can have any value. If we differentiate the ground-state energy with respect to Δ , we obtain

$$0 = \frac{1}{\pi} \ln \left(\frac{\Delta}{\pi q T_K} \right)$$

or

$$\Delta = \pi q T_K$$

confirming that $\Delta \sim T_K$.

- The mean-field value of the constraint field λ is determined relative to the Fermi energy μ . Were we to introduce a slowly varying external potential field to the conduction electron sea, then the chemical potential would become locally shifted so that $\mu \rightarrow \mu + e\phi(t)$. So long as the field $\phi(t)$ is varied at a rate slowly compared with the Kondo temperature, the constraint field will always track with the chemical potential, and since the constraint field is pinned to the chemical potential, $\lambda \rightarrow \lambda + e\phi(t)$. In the process, the constraint term will become

$$\lambda(\hat{n}_f(j) - Q) \rightarrow \lambda(\hat{n}_f(j) - Q) + e\phi(t)(\hat{n}_f(j) - Q). \quad (56)$$

Since the f -electrons now couple to the external potential $e\phi$ we have to ascribe a physical charge $e = -|e|$ to them. By contrast, the $-Q$ term in the constraint must be interpreted as a “background positive charge” $|e|Q \equiv |e|$ per site. These lines of reasoning indicate that we should think of the Kondo effect as a *many-body ionization phenomenon* in which the neutral local moment splits up into a negatively charged heavy electron and a stationary positive background charge we can associate with the formation of a Kondo singlet.

3.4 The composite nature of the f -electron

The matrix Green's function of the Kondo lattice reminds us of the Nambu Green's function in superconductivity. It is given by

$$\mathcal{G}_{\mathbf{k}}(\tau) = -\langle \psi_{\mathbf{k}\sigma}(\tau) \psi_{\mathbf{k}\sigma}^\dagger(0) \rangle \equiv \begin{bmatrix} G_c(\mathbf{k}, \tau) & G_{cf}(\mathbf{k}, \tau) \\ G_{fc}(\mathbf{k}, \tau) & G_f(\mathbf{k}, \tau) \end{bmatrix} \quad (57)$$

where $G_c(\mathbf{k}, \tau) = -\langle c_{\mathbf{k}}(\tau) c_{\mathbf{k}\sigma}^\dagger(0) \rangle$, $G_{cf}(\mathbf{k}, \tau) = -\langle c_{\mathbf{k}}(\tau) f_{\mathbf{k}\sigma}^\dagger(\tau) \rangle$ and so on. The anomalous off-diagonal members of this Green's function remind us of the Gor'kov functions in BCS theory and develop with the coherent hybridization. Using the two component notation (37), this Green's function can be written

$$\mathcal{G}_{\mathbf{k}}(\tau) = -(\partial_\tau + \underline{h}_{\mathbf{k}})^{-1} \xrightarrow{\text{F.T.}} \mathcal{G}_{\mathbf{k}}(i\omega_n) = (i\omega_n - \underline{h}_{\mathbf{k}})^{-1}, \quad (58)$$

where F.T. denotes a Fourier transform in imaginary time ($\partial_\tau \rightarrow -i\omega_n$), or more explicitly,

$$\mathcal{G}_{\mathbf{k}}(z) = (z - \underline{h}_{\mathbf{k}})^{-1} = \begin{pmatrix} z - \varepsilon_{\mathbf{k}} & -V \\ -V & z - \lambda \end{pmatrix}^{-1} = \begin{pmatrix} G_c(\mathbf{k}, z) & G_{cf}(\mathbf{k}, z) \\ G_{fc}(\mathbf{k}, z) & G_f(\mathbf{k}, z) \end{pmatrix} \quad (59)$$

$$= \frac{1}{(z - \varepsilon_{\mathbf{k}})(z - \lambda) - V^2} \begin{pmatrix} z - \lambda & V \\ V & z - \varepsilon_{\mathbf{k}} \end{pmatrix}, \quad (60)$$

where we have taken the liberty of analytically extending $i\omega_r \rightarrow z$ into the complex plane. Now we can read off the Green's functions. In particular, the hybridized conduction electron Green's function is

$$G_c(\mathbf{k}, z) = \Rightarrow \Rightarrow = \frac{z - \lambda}{(z - \varepsilon_{\mathbf{k}})(z - \lambda) - V^2} \quad (61)$$

$$= \frac{1}{z - \varepsilon_{\mathbf{k}} - \frac{V^2}{z - \lambda}} \equiv \frac{1}{z - \varepsilon_{\mathbf{k}} - \Sigma_c(z)} \quad (62)$$

which we can interpret physically as conduction electrons scattering off resonant f -states at each site, giving rise to a momentum-conserving self-energy

$$\Sigma_c(z) = \text{---} \bullet \text{---} \xrightarrow{O(1)} \text{---} \bullet \text{---} = \frac{V^2}{z - \lambda}. \quad (63)$$

We see that the Kondo effect has injected a resonant scattering pole at energy $z = \lambda$ into the conduction electron self-energy. This resonant scattering lies at the heart of the Kondo effect.

3.4.1 An absurd digression: the nuclear Kondo effect

The appearance of this pole in the scattering raises a vexing question in the Kondo effect: *What is the meaning of the f electron?* This might seem like a dumb question, for in electronic materials the Kondo effect certainly involves localized f electrons, and surely we can interpret this pole as the adiabatic renormalization of a hybridized band structure. This is certainly true. Yet as purists, we do have to confess that our starting model was a pure Kondo lattice model with only spin degrees of freedom: They could even have been *nuclear* spins!

This might seem absurd, yet nuclear spins do couple antiferromagnetically with conduction electrons to produce nuclear antiferromagnetism. Leaving aside practical issues of magnitude, we can learn something from the thought experiment in which the nuclear spin coupling to electrons *is* strong enough to overcome the nuclear magnetism. In this case, resonant bound states would form with the nuclear spin lattice giving rise to *charged* heavy electrons, presumably with an expanded Fermi surface.

From this line of argument we see that while it is tempting to associate the heavy fermion with a physical f or d electron localized inside the local moment, from a renormalization group perspective, the heavy electron is an emergent excitation: a fermionic bound state formed between the conduction sea and the neutral localized moments. This alternate point-of-view is useful, because it allows us to contemplate the possibility of new kinds of Kondo effects in states that are not adiabatically accessible from a band insulator or metal.

3.5 Cooper pair analogy

There is a nice analogy with superconductivity that helps one to understand the composite nature of the heavy electron. In a superconductor, electron pairs behave as loose composite bosons described by the relation

$$\overline{\psi_{\uparrow}(x)\psi_{\downarrow}(x')} = -F(x - x'). \quad (64)$$

Here $F(x - x') = -\langle T\psi_{\uparrow}(1)\psi_{\downarrow}(2) \rangle$ is the anomalous Gor'kov Green's function, which determines the Cooper-pair wavefunction, extended over the coherence length $\xi \sim v_F/T_c$. A similar phenomenon takes place in the Kondo effect, but here the bound state develops between spins and electrons, forming a fermion rather than a boson. For the Kondo lattice, it is perhaps more useful to think in terms of a screening time $\tau_K \sim \hbar/T_K$, rather than a length. Both the Cooper pair and heavy electron involve electrons that span decades of energy up to a cutoff, be it the Debye energy ω_D in superconductivity or the (much larger) bandwidth D in the Kondo effect [33,34].

To follow this analogy in greater depth, recall that in the path integral the Kondo interaction factorizes as

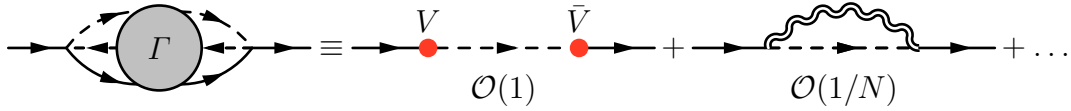
$$\frac{J}{N} c_{\beta}^{\dagger} S_{\alpha\beta} c_{\alpha} \rightarrow \bar{V} (c_{\alpha}^{\dagger} f_{\alpha}) + (f_{\alpha}^{\dagger} c_{\alpha}) V + N \frac{\bar{V}V}{J}, \quad (65)$$

so by comparing the right and left hand side, we see that the composite operators $S_{\beta\alpha} c_{\beta}$ and $c_{\beta}^{\dagger} S_{\alpha\beta}$ behave as a single fermion denoted by the contractions:

$$\frac{1}{N} \sum_{\beta} \overline{S_{\beta\alpha} c_{\beta}} = \left(\frac{\bar{V}}{J} \right) f_{\alpha}, \quad \frac{1}{N} \sum_{\beta} \overline{c_{\beta}^{\dagger} S_{\alpha\beta}} = \left(\frac{V}{J} \right) f_{\alpha}^{\dagger}, \quad (66)$$

Composite Fermion

Physically, this means that the spins bind high-energy electrons, transforming themselves into composites that then hybridize with the conduction electrons. The resulting heavy fermions can be thought of as moments ionized in the magnetically polar electron fluid to form mobile, negatively charged heavy electrons while leaving behind a positively charged “Kondo singlet.” Microscopically, the many-body amplitude to scatter an electron off a local moment develops a bound-state pole, which for large N we can denote by the diagrams



The leading diagram describes a kind of condensation of the hybridization field; the second and higher terms describe the smaller $\mathcal{O}(1/N)$ fluctuations around the mean-field theory.

By analogy with superconductivity, we can define a wavefunction associated with the temporal correlations between spin-flips and conduction electrons, as follows

$$\frac{1}{N} \sum_{\beta} \overline{c_{\beta}(\tau) S_{\beta\alpha}(\tau')} = g(\tau - \tau') \hat{f}_{\alpha}(\tau'), \quad (67)$$

where the spin-flip correlation function $g(\tau - \tau')$ is an analogue of the Gor'kov function, extending over a coherence time $\tau_K \sim \hbar/T_K$. Notice that in contrast to the Cooper pair, this composite object is a fermion and thus requires a distinct operator $\hat{f}_{\alpha}$ for its expression.

4 Heavy-fermion superconductivity

We now take a brief look at heavy-fermion superconductivity. There is a wide variety of heavy-electron superconductors, almost all of which are nodal superconductors, in which the pairing force derives from the interplay of magnetism and electron motion. In the heavy-fermion compounds, as in many other strongly correlated electron systems, superconductivity frequently develops at the border of magnetism, near the *quantum critical point* where the magnetic transition temperature has been suppressed to zero. In some of them, such as UPt_3 ($T_c=0.5\text{K}$) [35] the superconductivity develops out of a well developed heavy Fermi liquid, and in these cases, we can consider the superconductor to be paired by magnetic fluctuations within a well formed heavy Fermi liquid. However, in many other superconductors, such as UBe_{13} ($T_c=1\text{K}$) [36,37], the 115 superconductors CeCoIn_5 ($T_c=2.3\text{K}$) [38], CeRhIn_5 under pressure ($T_c=2\text{K}$) [16], NpAl_2Pd_5 ($T_c=4.5\text{K}$) [39] and PuCoGa_5 ($T_c=18.5\text{K}$) [40,41], the superconducting transition temperature is comparable with the Kondo temperature. In many of these

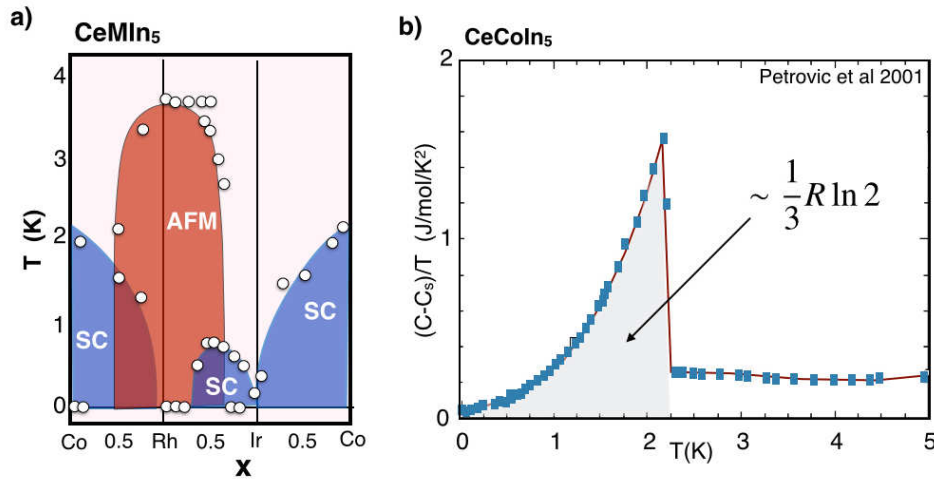


Fig. 12: (a) Phase diagram of 115 compounds CeMIn_5 , adapted from [42], showing magnetic and superconducting phases as a function of alloy concentration. (b) Sketch of specific heat coefficient of CeCoIn_5 , (with nuclear Schottky contribution subtracted), showing the large entropy of condensation associated with the superconducting state. (After Petrovic et al 2001 [38]).

materials, the entropy of condensation

$$S_c = \int_0^{T_c} \frac{C_V}{T} dT \quad (68)$$

can be as large as $(1/3)R \ln 2$ per rare-earth ion, indicating that the spin is in some way entangling with the conduction electrons to build the condensate. In this situation, we need to be able to consider the Kondo effect and superconductivity on an equal footing.

4.1 Symplectic spins and $SP(N)$.

Although the $SU(N)$ large- N expansion provides a very useful description of the normal state of heavy-fermion metals and Kondo insulators, there is strangely no superconducting solution. This shortcoming lies in the very structure of the $SU(N)$ group. $SU(N)$ is perfectly tailored to particle physics, where the physical excitations, the mesons and baryons, appear as color singlets, with the meson a $q\bar{q}$ quark-antiquark singlet while the baryon is an N -quark singlet $q_1 q_2 \dots q_N$, (where of course $N = 3$ in reality). In electronic condensed matter, the meson becomes a particle-hole pair, but there are no two-particle singlets in $SU(N)$ beyond $N = 2$. The origin of this failure can be traced back to the absence of a consistent definition of time-reversal symmetry in $SU(N)$ for $N > 2$. This means that singlet Cooper pairs and superconductivity can not develop at the large- N limit.

A solution to this problem that grew out of an approach developed by Read and Sachdev [43] for frustrated magnetism is to use the symplectic group $SP(N)$, where N must be an even number [44, 45]. This little-known group is a subgroup of $SU(N)$. In fact for $N = 2$, $SU(2) = SP(2)$ are identical, but they diverge for higher N . For example, $SU(4)$ has 15 generators, but its symplectic sub-group $SP(4)$ has only 10. At large N , $SP(N)$ has approximately half the number of generators of $SU(N)$. The symplectic property of the group allows it to consistently

treat time-reversal symmetry of spins, and it also allows the formation of two-particle singlets for any N .

One of the interesting aspects of $SP(N)$ spin operators is their relationship to pair operators. Consider $SP(2) \equiv SU(2)$: The pair operator is $\Psi^\dagger = f_\uparrow^\dagger f_\downarrow^\dagger$, and since this operator is a singlet, it commutes with the spin operators $[\Psi, \vec{S}] = [\Psi^\dagger, \vec{S}] = 0$, which, since Ψ and $\Psi^\dagger$ are the generators of particle-hole transformations, implies that the $SU(2)$ spin operator is particle-hole symmetric. It is this feature that is preserved by the $SP(N)$ group, all the way out to $N \rightarrow \infty$. In fact, we can use this fact to write down $SP(N)$ spins as follows: An $SU(N)$ spin is given by $S_{\alpha\beta}^{SU(N)} = f_\alpha^\dagger f_\beta$. Under a particle-hole transformation $f_\alpha \rightarrow \text{Sgn}(\alpha) f_{-\alpha}^\dagger$. If we take the particle-hole transform of the $SU(N)$ spin and add it to itself we obtain an $SP(N)$ spin,

$$S_{\alpha\beta} = f_\alpha^\dagger f_\beta + \text{Sgn}(\alpha\beta) f_{-\beta}^\dagger f_{-\alpha}, \quad (69)$$

Symplectic Spin operator

where the values of the spin indices are $\alpha, \beta \in \{\pm 1/2, \dots, \pm N/2\}$. This spin operator commutes with the three isospin variables

$$\tau_3 = n_f - N/2, \quad \tau^+ = \sum_{\alpha>0} f_\alpha^\dagger f_{-\alpha}^\dagger, \quad \tau^- = \sum_{\alpha>0} f_{-\alpha} f_\alpha. \quad (70)$$

With these local symmetries, the spin is continuously invariant under $SU(2)$ particle-hole rotations $f_\alpha \rightarrow u f_\alpha + v \text{Sgn}(\alpha) f_{-\alpha}^\dagger$, where $|u|^2 + |v|^2 = 1$, as one can verify. To define an irreducible representation of the spin, we also have to impose a constraint on the Hilbert space, which in its simplest form is $\tau_3 = \tau^\pm = 0$, equivalent to $Q = N/2$ in the $SU(N)$ approach. In other words, the s -wave part of the f pairing must vanish identically.

4.2 Superconductivity in the Kondo-Heisenberg model

Let us take a look at the way this works in a nearest neighbor *Kondo-Heisenberg model* [46],

$$H = H_c + H_K + H_M. \quad (71)$$

Here $H_c = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma}$ describes the conduction sea, whereas H_K and H_M are the Kondo and Heisenberg (RKKY) interactions, respectively. These take the form

$$\begin{aligned} H_K &= \frac{J_K}{N} \sum_j c_{j\alpha}^\dagger c_{j\beta} S_{\beta\alpha}(j) \rightarrow -\frac{J_K}{N} \sum_{i,j} \left((c_{j\alpha}^\dagger f_{j\alpha})(f_{j\beta}^\dagger c_{j\beta}) + \tilde{\alpha}\tilde{\beta}(c_{j\alpha}^\dagger f_{j-\alpha}^\dagger)(f_{j-\beta} c_{j\beta}) \right) \\ H_M &= \frac{J_H}{2N} \sum_{(i,j)} S_{\alpha\beta}(j) S_{\beta\alpha}(j) \rightarrow -\frac{J_H}{N} \sum_j \left[(f_{i\alpha}^\dagger f_{j\alpha})(f_{j\beta}^\dagger f_{i\beta}) + \tilde{\alpha}\tilde{\beta}(f_{i\alpha}^\dagger f_{j-\alpha}^\dagger)(f_{j-\beta} f_{i\beta}) \right] \end{aligned} \quad (72)$$

where we have introduced the notation $\tilde{\alpha} = \text{Sgn}(\alpha)$ and have shown how the interactions are expanded into particle-hole and particle-particle channels. Notice how the interactions are

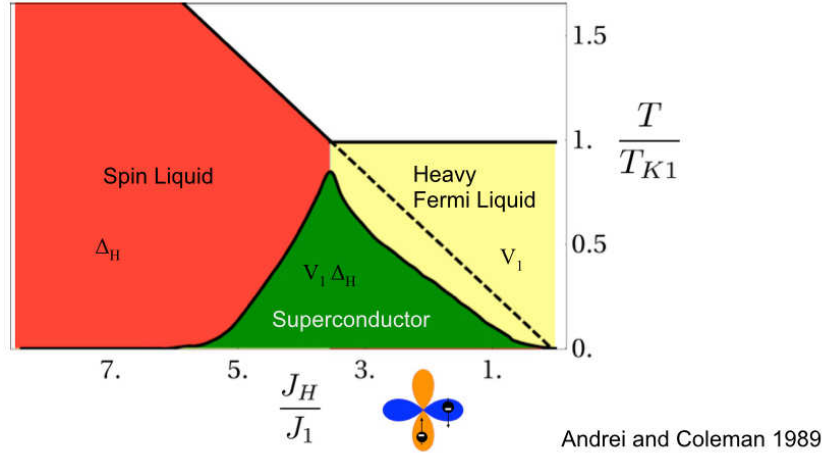


Fig. 13: Phase diagram for the two-dimensional Kondo-Heisenberg model, derived in the $SP(N)$ large- N approach, adapted from [46], courtesy Rebecca Flint.

equally divided between particle-hole and particle-particle channels. When we carry out the Hubbard-Stratonovich decoupling in each of these terms, we obtain

$$H_K \rightarrow \sum_j \left[c_{j\alpha}^\dagger \left(V_j f_{j\alpha} + \tilde{\alpha} \Delta_j^K f_{j-\alpha}^\dagger \right) + \text{H.c.} \right] + N \left(\frac{|V_j|^2 + |\Delta_j^K|^2}{J_K} \right) \quad (73)$$

$$H_H \rightarrow \sum_{(i,j)} \left[t_{ij} f_{i\alpha}^\dagger f_{j\alpha} + \Delta_{ij} \tilde{\alpha} f_{i\alpha}^\dagger f_{j-\alpha}^\dagger + \text{H.c.} \right] + N \left[\frac{|t_{ij}|^2 + |\Delta_{ij}|^2}{J_H} \right] \quad (74)$$

At each site, we can always rotate the f electrons in particle-hole space to remove the Kondo pairing component and set $\Delta_j^K = 0$, but the pairing terms in the Heisenberg component can not be eliminated. This mean-field theory describes a kind of *Kondo-stabilized spin-liquid* [46]. The physical picture is as follows: In practice, a spin-liquid is unstable to magnetism, but its happy coexistence with the Kondo effect brings its energy below that of the antiferromagnet. The hybridization of the f with the conduction sea converts the spinons of the spin liquid into charged fermions. The t_{ij} terms describe various kind of exotic density waves. The Δ_{ij} terms now describe pairing amongst the composite fermions.

To develop a simple theory of the superconducting state, we restrict our attention to uniform, static saddle points, dropping the t_{ij} . Let us look at the resulting mean-field theory. In two dimensions, this becomes

$$H = \sum_{\mathbf{k}, \alpha > 0} (\tilde{c}_{\mathbf{k}\alpha}^\dagger, \tilde{f}_{\mathbf{k}\alpha}) \begin{bmatrix} \varepsilon_{\mathbf{k}} \tau_3 & V \tau_1 \\ V \tau_1 & \vec{w} \cdot \vec{\tau} + \Delta_{H\mathbf{k}} \tau_1 \end{bmatrix} \begin{pmatrix} \tilde{c}_{\mathbf{k}\alpha} \\ \tilde{f}_{\mathbf{k}\alpha} \end{pmatrix} + \mathcal{N}_s N \left(\frac{|V|^2}{J_K} + 2 \frac{|\Delta_H|^2}{J_H} \right) \quad (75)$$

where

$$\tilde{c}_{\mathbf{k}\alpha}^\dagger = (c_{\mathbf{k}\alpha}^\dagger, \tilde{\alpha} c_{-\mathbf{k}, -\alpha}), \quad \tilde{f}_{\mathbf{k}\alpha}^\dagger = (f_{\mathbf{k}\alpha}^\dagger, \tilde{\alpha} f_{-\mathbf{k}, -\alpha}) \quad (76)$$

are Nambu spinors for the conduction and f -electrons. The vector $\vec{W}$ of Lagrange multipliers couples to the isospin of the f electrons: Stationarity of the free energy with respect to this variable imposes the mean-field constraint $\langle \tilde{f}^\dagger \vec{\tau} f \rangle = 0$. The function $\Delta_{H\mathbf{k}} = \Delta_{\mathbf{k}} (\cos k_x -$

$\cos k_y$) is the f -electron pair wavefunction. Here we have chosen a d -wave form-factor. For this choice, the local f pair density automatically vanishes and so we need only choose $\vec{w} = (0, 0, \lambda)$, where λ couples to τ_3 (imposing the constraint $n_f = N/2$). We could have also tried an extended s -wave pair wavefunction, but in this case, the induced s -wave pair density becomes finite, and the effect of the $\vec{w}$ constraint is to suppress the transition temperature. By seeking stationary points in the free energy with respect to variations in Δ_H , V , and λ one can derive the phase diagram for d -wave pairing, shown in Fig. 13. The mean-field theory shows that superconductivity develops at the interface between the Fermi liquid and the spin liquid.

5 Topological Kondo insulators

One of the areas of fascinating development in the last few years is the discovery that Kondo insulators can develop *topological order* to form a *topological Kondo insulator*. Topological order refers to the idea that a quantum mechanical ground state can develop a non-trivial topology. One of the defining features of topological ground states is the development of protected surface states. The best known example of topological order is the integer quantum Hall effect, where an integer-filled Landau level develops topological order that is responsible for the robust quantization of the quantum Hall effect [47–49]. In a remarkable series of discoveries in 2006, [50–57] it became clear that strong spin-orbit coupling can play the role of a synthetic magnetic field, so that band insulators can also develop a non-trivial topology while preserving time-reversal symmetry. Such Z_2 topological band insulators are defined by a single topological $Z_2 = \pm 1$ index that is positive in conventional insulators, but reverses in topological Z_2 insulators. This topological feature manifests itself through the formation of robust conducting surface states.

In 2007, Liang Fu and Charles Kane showed that if an insulator has both time-reversal and inversion symmetry [57], this Z_2 index is uniquely determined by the parities δ_{in} of the Bloch states at the high-symmetry points Γ_i of the valence band

$$Z_2 = \prod_{\Gamma_i} \delta(\Gamma_i) = \begin{cases} +1 & \text{conventional insulator} \\ -1 & \text{topological insulator} \end{cases} \quad (77)$$

Fu Kane formula for the Z_2 index of topological insulators

where $\delta(\Gamma_i) = \prod_n \delta_{in}$ is the product of the parities of the occupied bands at the high-symmetry points in the Brillouin zone. This formula allows one to determine whether an insulator state is topological merely by checking whether the index $Z_2 = -1$, without a detailed knowledge of the ground-state wavefunction.

It used to be thought that Kondo insulators could be regarded as “renormalized silicon.” The discovery of topological insulators forced a reevaluation of this viewpoint. The large spin-orbit coupling and the odd parity of the f states led to the proposal, by Dzero, Sun, Galitski, and the author [58], that Kondo insulators can become topologically ordered. The Fu-Kane formula has

a special significance for Kondo insulators, which contain *odd parity* f electrons hybridizing with *even parity* d electrons. Each time an f electron crosses through the band gap, exchanging with a conduction d state, this changes the Z_2 index, making it highly likely that certain Kondo insulators are topological. The oldest known Kondo insulator SmB_6 , discovered almost 50 years ago, was well known to possess a mysterious low-temperature conductivity plateau [59, 60], and the idea that this system might be a topological Kondo insulator provided an exciting way of explaining this old mystery. The recent observation of robust [61, 62] conducting surface states in the oldest Kondo insulator SmB_6 supports one of the key elements of this prediction, prompting a revival of interest in Kondo insulators as a new route for studying the interplay of strong interactions and topological order.

SmB_6 is really a mixed-valent system, which takes us a little beyond the scope of this lecture. One of the other issues with SmB_6 is that its local crystal field configuration is likely to be a Γ_8 quartet state [63] rather than a Kramers doublet. Nevertheless, key elements of its putative topological Kondo insulating state are nicely illustrated by a *spin-orbit coupled* Kondo-Heisenberg model, describing the interaction of Kramers doublet f states with a d band. The model is essentially identical to (Eq. (71))

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} \psi_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + J_K \sum_j \psi_{j\alpha}^\dagger \psi_{j\beta} S_{\beta\alpha}(j) + J_H \sum_{i,j} S_{\alpha\beta}(i) S_{\beta\alpha}(j) \quad (78)$$

with an important modification that takes into account the large spin-orbit coupling and the odd parity of the f states. This forces the local Wannier states $\Psi_{j\alpha}$ that exchange spin with the local moment to be odd-parity combinations of nearest-neighbor conduction electrons, given by

$$\psi_{j\alpha}^\dagger = \sum_{i,\sigma} c_{i\sigma}^\dagger \Phi_{\sigma\alpha}(\mathbf{R}_i - \mathbf{R}_j). \quad (79)$$

We will consider a simplified model with the form factor

$$\Phi(\mathbf{R}) = \begin{cases} -i\hat{R} \cdot \frac{\vec{\sigma}}{2}, & \mathbf{R} \in \text{n.n} \\ 0 & \text{otherwise.} \end{cases} \quad (80)$$

This form factor describes the spin-orbit mixing between states with orbital angular momentum l differing by one, such as f and d or p and s orbitals. The odd parity of the form-factor $\Phi(\mathbf{R}) = -\Phi(-\mathbf{R})$ derives from the odd-parity f orbitals, while the prefactor $-i$ ensures that the hybridization is invariant under time reversal. The Fourier transform of this form factor, $\Phi(\mathbf{k}) = \sum_{\mathbf{R}} \Phi(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}}$ is then

$$\Phi(\mathbf{k}) = \vec{s}_{\mathbf{k}} \cdot \vec{\sigma} \quad (81)$$

where the vector $\vec{s}_{\mathbf{k}} = (\sin k_1, \sin k_2, \sin k_3)$ is the periodic equivalent of the unit momentum vector $\hat{\mathbf{k}}$. Notice how $\vec{s}(\Gamma_i) = 0$ vanishes at the high symmetry points.

The resulting mean-field Hamiltonian takes the form

$$H_{TKI} = \sum_{\mathbf{k}} \psi_{\mathbf{k}}^\dagger h(\mathbf{k}) \psi_{\mathbf{k}} + \mathcal{N}_s \left[\left(\frac{V^2}{J_K} + \frac{3t^2}{J_H} - \lambda Q \right) \right], \quad (82)$$

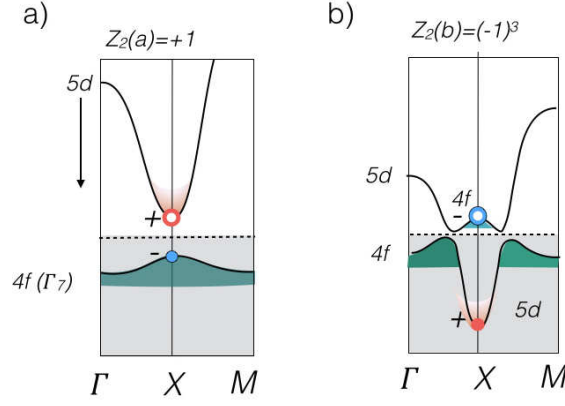


Fig. 14: (a) When the d band is above the filled f band, a trivial insulator is formed. (b) When the d band crosses the f band at the three X -points, the Z_2 parity changes sign, giving rise to a topological insulator.

where $\psi_{\mathbf{k}}^\dagger = (c_{\mathbf{k}\sigma}^\dagger, f_{\mathbf{k}\sigma}^\dagger)$ and

$$h(\mathbf{k}) = \begin{pmatrix} \varepsilon_{\mathbf{k}} & V\vec{\sigma} \cdot \vec{s}_{\mathbf{k}} \\ V\vec{\sigma} \cdot \vec{s}_{\mathbf{k}} & \varepsilon_{f\mathbf{k}} \end{pmatrix} \quad (83)$$

while $\varepsilon_{f\mathbf{k}} = 2t_f(c_x + c_y + c_z) + \lambda$ ($c_l \equiv \cos k_l$) is the dispersion of the f state resulting from a mean-field decoupling of the intersite Heisenberg coupling in the particle-hole channel. For small $\mathbf{k}$, the hybridization in the Hamiltonian $h(\mathbf{k})$ takes the form $V\vec{\sigma} \cdot \mathbf{k}$, closely resembling the topologically non-trivial triplet p -wave gap structure of superfluid He-3B. Like He-3B, the hybridization only develops at low temperatures, making SmB₆ an adaptive insulator.

Let us for the moment treat $h(\mathbf{k})$ as a rigid band structure. Suppose the f band were initially completely filled, with a completely empty d band above it (See Fig. 14a). This situation corresponds to a conventional band insulator with $Z_2 = +1$. Next, let us lower the d conduction band until the two bands cross at a high symmetry point, causing the gap to close and then to re-open. We know, from de Haas-van Alphen studies of the iso-electronic material LaB₆ [64] (whose band-structure is identical to SmB₆ but lacks the magnetic f electrons) and from ARPES studies [65–67] that in SmB₆ the d band crosses through the Fermi surface at the three X points. Once the d band is lowered through the f band around the three X points, the odd-parity f states at the X point move up into the conduction band, to be replaced by even-parity d states. This changes the sign of $Z_2 \rightarrow (-1)^3 = -1$, producing a topological ground state. Moreover, since there are three crossings, we expect there to be three spin-polarized surface Dirac cones.

We end by noting that at the time of writing, our understanding of the physics SmB₆ is in rapid flux on both the experimental and theoretical front. Spin-resolved ARPES [68] measurements have detected the presence of spin textures in the surface Fermi surfaces around the surface $\bar{X}$ point, a strong sign of topologically protected surface states. Two recent theoretical works [69, 70] have shown that the spin textures seen in these experiments are consistent with a spin-quartet ground state in SmB₆. Despite this progress, consensus on the topological nature of SmB₆ has not yet been achieved, and competing groups have offered alternate interpretations of the data, including the possibility of polarity-driven surface metallicity [71] and Rashba-split surface states, both of a non-topological origin [72]. Another area of experimental controversy

concerns the possible de Haas-van Alphen oscillations created by surface topological excitations, with one report of the detection of surface de Haas-van Alphen signals [73] and a recent, very remarkable report of *bulk* de Haas-van Alphen signals associated with unhybridized, quantum critical d electrons [74].

6 Coexisting magnetism and Kondo effect

In this short lecture, I have given a quick introduction to the paramagnetic phases of heavy-fermion systems. One of the major open questions in heavy-fermion and Kondo-lattice physics concerns the physics of magnetism and the right way to describe the development of magnetism within these materials. There is growing evidence that magnetism and the Kondo effect can coexist, sometimes homogeneously and sometimes inhomogeneously. For example, In the 115 superconductor CeRhIn₅ there is evidence for a microscopic and homogeneous coexistence of local-moment magnetism and heavy-fermion superconductivity under pressure [75]. By contrast, in the geometrically frustrated CePdAl [76,77], two thirds of the Cerium sites spontaneously develop magnetism, leaving the other third to undergo a Kondo effect [78]. What is the right way to describe these coexistent states? One possibility that I have worked on with Aline Ramires [79,80] is the use of a *supersymmetric* representation of the spin

$$S_{\alpha\beta} = f_{\alpha}^{\dagger}f_{\beta} + b_{\alpha}^{\dagger}b_{\beta} \quad (84)$$

where the $f_{\alpha}^{\dagger}$ and $b_{\alpha}^{\dagger}$ are fermionic and bosonic creation operators. Such a representation permits in principle, the existence of *two-fluid* ground states, involving a Gutzwiller projection of bosonic and fermionic wavefunctions

$$|\Psi\rangle = P_G|\Psi_F\rangle|\Psi_B\rangle, \quad (85)$$

where $|\Psi_F\rangle$ is the fermionic component of the wavefunction describing the Kondo-quenched local moments, while $|\Psi_B\rangle$ describes the formation of long-range magnetic correlations within a bosonic RVB wavefunction, and

$$P_G = \int \prod_j \frac{d\theta_j}{2\pi} e^{i\theta_j(n_B+n_F-1)} \quad (86)$$

is a Gutzwiller projection operator onto the state with one spin per site. We have been trying to describe such mixed-state wavefunctions in the large- N limit, seeking saddle-point solutions where a bosonic and fermionic fluid coexist [80]. One of the ideas that emerges from this kind of approach is the possibility that the soft modes at a quantum critical point might develop fermionic character, a kind of emergent supersymmetry [81].

Acknowledgment

This article was written with the support of the National Science Foundation under grant no NSF DMR-1309929.

References

- [1] P. Coleman: *Handbook of Magnetism and Advanced Magnetic Materials* (John Wiley and Sons, Amsterdam, 2007), Vol. 1, pp. 95–148
- [2] P. Coleman: *Introduction to Many-Body Physics* (Cambridge University Press, 2015)
- [3] J.L. Smith and E.A. Kmetko, *Journal of the Less Common Metals* **90**, 83 (1983)
- [4] P.W. Anderson, *Phys. Rev.* **115**, 2 (1959)
- [5] P.W. Anderson, *Phys. Rev.* **124**, 41 (1961)
- [6] J. Kondo, *Prog. Theor. Phys.* **32**, 37 (1964)
- [7] R.H. White and T.H. Geballe in: H. Ehrenreich, F. Seitz, and D. Turnbull (eds.): *Solid State Physics* (Academic Press, New York, 1979), Vol. 15, p. 283
- [8] M.D. Daybell: In G. Rado and H. Suhl (Eds.) *Magnetism* (Academic Press, New York, 1973), Vol. 5, p. 121
- [9] N. Mott, *Philosophical Magazine* **30**, 403 (1974)
- [10] S. Doniach, *Physica*, **91 B+C**, 231 (1977)
- [11] M.A. Ruderman and C. Kittel, *Phys. Rev* **96**, 99 (1954)
- [12] T. Kasuya, *Prog. Theo. Phys.* **16**, 45 (1956)
- [13] K. Yosida, *Phys. Rev.* **106**, 896 (1957)
- [14] T. Kasuya, *Prog. Theor. Phys.* **16**, 45 (1956)
- [15] R. Jullien, J. Fields, and S. Doniach, *Phys. Rev. B* **16**, 4889 (1977)
- [16] T. Park, F. Ronning, H.Q. Yuan, M.B. Salmon, R. Movshovich, J.L. Sarrao, and J.D. Thompson, *Nature* **440**, 65 (2006)
- [17] P. Coleman, C. Pépin, Q. Si, and R. Ramazashvili, *J. Phys. Cond. Matt.* **13**, 273 (2001)
- [18] H. v.Löhneysen, A. Rosch, M. Vojta, and P. Wölfle, *Rev. Mod. Phys.* **79**, 1015 (2007)
- [19] A. Menth, E. Buehler, and T.H. Geballe, *Phys. Rev. Lett* **22**, 295 (1969)
- [20] G. Aeppli and Z. Fisk, *Comm. Condens. Matter Phys.* **16**, 155 (1992)
- [21] Z. Fisk, J.L. Sarrao, S.L. Cooper, P. Nyhus, G.S. Boebinger, A. Passner, and P.C. Canfield, *Physica B* **223-224**, 409 (1996)
- [22] H. Tsunetsugu, M. Sigrist, and K. Ueda, *Rev. Mod. Phys.* **69**, 809 (1997)

- [23] P. Riseborough, *Adv. Phys.* **49**, 257 (2000)
- [24] P.W. Anderson: *Valence Fluctuations in Solids* (North Holland, Amsterdam, 1981)
- [25] P. Coleman, *Phys. Rev.* **28**, 5255 (1983)
- [26] N. Read and D. Newns, *J. Phys. C* **16**, 3274 (1983)
- [27] P. Coleman, *Phys. Rev. B* **29**, 3035 (1984)
- [28] N. Read and D.M. Newns, *J. Phys. C* **29**, L1055 (1983)
- [29] A. Auerbach and K. Levin, *Phys. Rev. Lett.* **57**, 877 (1986)
- [30] A.J. Millis, and P.A. Lee, *Phys. Rev. B* **35**, 3394 (1987)
- [31] P. Coleman, *Phys. Rev. B* **35**, 5072 (1987)
- [32] B. Coqblin and J.R. Schrieffer, *Phys. Rev.* **185**, 847 (1969)
- [33] S. Burdin, A. Georges, and D.R. Grempel, *Phys. Rev. Lett* **85**, 1048 (2000)
- [34] T.A. Costi and N. Manini, *Journal of Low Temperature Physics* **126**, 835 (2002)
- [35] G.R. Stewart, Z. Fisk, J.O. Willis, and J.L. Smith, *Phys. Rev. Lett* **52**, 679 (1984)
- [36] H.R. Ott, H. Rudigier, Z. Fisk, and J.L. Smith, *Phys. Rev. Lett* **50**, 1595 (1983)
- [37] H.R. Ott, H. Rudigier, T.M. Rice, K. Ueda, Z. Fisk, and J.L. Smith, *Phys. Rev. Lett* **52**, 1915 (1984)
- [38] C. Petrovic, P.G. Pagliuso, M.F. Hundley, R. Movshovich, J.L. Sarrao, J.D. Thompson, Z. Fisk, and P. Monthoux, *J. Phys. Cond. Matt* **13**, L337 (2001)
- [39] D. Aoki, Y. Haga, T. Matsuda, N. Tateiwa, S. Ikeda, and *et al.*, *Journal of The Physical Society of Japan* (2007)
- [40] J.L. Sarrao, L.A. Morales, J.D. Thompson, B.L. Scott, G.R. Stewart, F. Wastin, J. Rebizant, P. Boulet, E. Colineau, and G.H. Lander, *Nature* (2002)
- [41] N.J. Curro, T. Caldwell, E.D. Bauer, L.A. Morales, M.J. Graf, Y. Bang, A.V. Balatsky, J.D. Thompson, and J.L. Sarrao, *Nature* **434**, 622 (2005)
- [42] J.L. Sarrao and J.D. Thompson, *Journal of the Physical Society of Japan* **76**, 051013 (2007)
- [43] N. Read and S. Sachdev, *Phys. Rev. Lett* **66**, 1773 (1991)
- [44] R. Flint, M. Dzero, P. Coleman, and M. Dzero, *Nature Physics* **4**, 643 (2008)

- [45] R. Flint and P. Coleman, Phys. Rev. Lett. **105**, 246404 (2010)
- [46] P. Coleman and N. Andrei, J. Phys. Cond. Matt. **C 1**, 4057 (1989)
- [47] R.B. Laughlin, Phys. Rev **B23**, 5632 (1981)
- [48] D. Thouless, M. Kohmoto, M. Nightingale, and M. Den Nijs, Phys. Rev. Lett. **49**, 405 (1982)
- [49] F.D.M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988)
- [50] C.L. Kane and E.J. Mele, Phys. Rev. Lett. **95**, 146802 (2005)
- [51] B.A. Bernevig, T.L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006)
- [52] J.E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007)
- [53] R. Roy, Phys. Rev. B **79**, 195321 (2009)
- [54] L. Fu, C.L. Kane, and E.J. Mele, Phys. Rev. Lett. **98**, 106803 (2007)
- [55] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L.W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007)
- [56] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y.S. Hor, R.J. Cava, and M.Z. Hasan, Nature **452**, 970 (2008)
- [57] L. Fu and C. Kane, Physical Review B **76**, 45302 (2007)
- [58] M. Dzero, K. Sun, V. Galitski, and P. Coleman, Phys. Rev. Lett. **104**, 106408 (2010)
- [59] J.W. Allen, B. Batlogg, and P. Wachter, Phys. Rev. B **20**, 4807 (1979)
- [60] J.C. Cooley, M.C. Aronson, A. Lacerda, Z. Fisk, P.C. Canfield, and R.P. Guertin, Phys. Rev. B **52**, 7322 (1995)
- [61] S. Wolgast, C. Kurdak, K. Sun, J.W. Allen, D.-J. Kim, and Z. Fisk, Phys. Rev. B **88**, 180405 (2013)
- [62] D. J. S. Thomas, T. Grant, J. Botimer, Z. Fisk, and J. Xia, Scientific Reports **3**, 3150 (2014)
- [63] V. Alexandrov, M. Dzero, and P. Coleman, Phys. Rev. Lett. **111**, 226403 (2013)
- [64] Z. Fisk, A.J. Arko, J.B. Ketterson, G. Crabtree, F.M. Mueller, L.R. Windmiller, and D. Karim, Phys. Rev. B **13**, 5240 (1976)
- [65] M. Neupane, N. Alidoust, S.Y. Xu, and T. Kondo, Nat. Commun. **4**, 2991 (2013)

- [66] N. Xu, X. Shi, P.K. Biswas, C.E. Matt, R.S. Dhaka, Y. Huang, N.C. Plumb, M. Radović, J.H. Dil, E. Pomjakushina, K. Conder, A. Amato, Z. Salman, D.M. Paul, J. Mesot, H. Ding, and M. Shi, *Phys. Rev. B* **88**, 121102 (2013)
- [67] J.D. Denlinger, J.W. Allen, J.-S. Kang, K. Sun, B.-I. Min, D.-J. Kim, and Z. Fisk, *arXiv:1312.6636v2* (2013)
- [68] N. Xu, P.K. Biswas, J.H. Dil, R.S. Dhaka, G. Landolt, S. Muff, C.E. Matt, X. Shi, N.C. Plumb, M.R. Caute, E. Pomjakushina, K. Conder, A. Amato, S.V. Borisenko, R. Yu, H.M. Weng, Z. Fang, X. Dai, J. Mesot, H. Ding, and M. Shi, *Nat. Commun.* **5**, 1 (2014)
- [69] M. Legner, A. Rüegg, and M. Sigrist, *arXiv:1505.02987v1* (2015)
- [70] P.P. Baruselli and M. Vojta, *arXiv:1505.03507v1* (2015)
- [71] Z.-H. Zhu, A. Nicolaou, G. Levy, N.P. Butch, P. Syers, X.F. Wang, J. Paglione, G.A. Sawatzky, I.S. Elfimov, and A. Damascelli, *Phys. Rev. Lett.* **111**, 216402 (2013)
- [72] P. Hlawenka, K. Siemensmeyer, E. Weschke, A. Varykhalov, J. Sánchez-Barriga, N.Y. Shitsevalova, A.V. Dukhnenko, V.B. Filipov, S. Gabáni, K. Flachbart, O. Rader, and E.D.L. Rienks, *arXiv:1502.01542* (2015)
- [73] G. Li, Z. Xiang, F. Yu, T. Asaba, B. Lawson, P. Cai, C. Tinsman, A. Berkley, S. Wolgast, Y.S. Eo, D.-J. Kim, Ç. Kurdak, J.W. Allen, K. Sun, X.H. Chen, Y.Y. Wang, Z. Fisk, and L. Li, *Science* **346**, 1208 (2014)
- [74] B.S. Tan, Y.T. Hsu, B. Zeng, M.C. Hatnean, N. Harrison, Z. Zhu, M. Hartstein, M. Kiourlappou, A. Srivastava, M.D. Johannes, T.P. Murphy, J.H. Park, L. Balicas, G.G. Lonzarich, G. Balakrishnan, and S.E. Sebastian, *Science* **349**, 287 (2015)
- [75] G. Knebel, D. Aoki, D. Braithwaite, B. Salce, and J. Flouquet, *Phys. Rev. B* **74**, 020501 (2006)
- [76] A. Donni, G. Ehlers, H. Maletta, P. Fischer, H. Kitazawa, and M. Zolliker, *J. Phys.: Condens. Matter* **8**, 11213 (1996)
- [77] C. Dolores Nunez-Regueriro, M. LaCroix and B. Canals, *Physica C* **282-287**, 1885 (1997)
- [78] V. Fritsch, N. Bagrets, G. Goll, W. Kittler, M. J. Wolf, K. Grube, C.-L. Huang, and H. v.Löhneysen, *Phys. Rev. B* **89**, 054416 (2014)
- [79] P. Coleman, C. Pépin, and A.M. Tsvelik, *Phys. Rev. B* **62**, 3852 (2000)
- [80] A. Ramires and P. Coleman, To be published (2015)
- [81] S.-S. Lee, *Phys. Rev. B* **76**, 075103 (2007)

2 Spectroscopic Properties of Mixed-Valence Compounds in the Impurity Model

Kurt Schönhammer

Institut für Theoretische Physik

Universität Göttingen

Contents

1	Introduction	2
2	Basic impurity models	3
2.1	Spin-degenerate single-impurity Anderson model	3
2.2	Impurity models involving core levels	8
3	Anderson impurity model with large orbital degeneracy	8
3.1	Ground-state calculation	9
4	The intermediate states method for spectra	12
4.1	Valence photoemission spectroscopy	12
4.2	Inverse Photoemission	14
4.3	Spectra involving core holes	17
5	Comparison with experiment	18

1 Introduction

The electronic properties of mixed valence lanthanide materials, like Ce compounds, were studied experimentally over a long period of time. In addition to thermodynamic and transport measurements various “high energy” spectroscopies like valence photoemission, inverse photoemission, and core level spectroscopies were used to understand the electronic properties of the f -levels of such systems [1]. It took some time until it was realized that electronic correlations play an essential role for the understanding of the f -spectra. As a first step in the attempt at theoretical understanding, a single rare-earth atom in a simple metal can be studied using the single-impurity Anderson model [2]. In this model, discussed in detail in the following, the energy ε_f of the f -level, the coupling Δ to the conduction electrons, and the Coulomb repulsion U between two electrons in the f -level are the essential parameters that determine quantities like the total f -level occupancy n_f . If spin-orbit coupling and crystal-field splitting are neglected the degeneracy of the f -level is given by $N_f = 14$.

The f -electron spectral function of the Anderson impurity model was a long-standing issue. If the coupling Δ is weak, and the Fermi level falls between ε_f and $\varepsilon_f + U$, the spectrum has a peak near ε_f (seen in photoemission) and a peak near $\varepsilon_f + U$ (seen in inverse photoemission). It was further realized that the spectrum has resonance close to $\varepsilon_F = 0$ usually called the Kondo resonance [3, 4]. Except for some special cases [5] it was, however, for a long time hard to determine even the qualitative properties of the Kondo resonance.

A historically important progress in the treatment of the Anderson impurity model was the realization in the early eighties that $1/N_f$ can be treated as a small parameter [6, 7]. Using this idea O. Gunnarsson and the author developed a method for calculating zero-temperature spectral properties (*intermediate states method*), which becomes exact in the limit $N_f \rightarrow \infty$ [8–11]. In particular, this method makes it possible to study the Kondo peak quantitatively for large values of N_f . Analytical results in the infinite- U limit obtained to leading order in $1/N_f$ are presented in the following sections. Higher-order calculations that require numerical work usually converge quickly for $N_f = 14$. They were successfully used for a comparison to the experimental spectra of Ce compounds [9, 10, 12].

Apart from second quantization, the intermediate states method uses only basic quantum mechanics. The knowledge of more sophisticated many-body techniques, like Feynman diagrams, is not necessary to understand it. This is presumably one of the reasons why it is used frequently by experimental groups for the interpretation of their measured spectra.

At about the same time as the intermediate states method, large N_f approaches (for infinite U) were proposed that allow an extension to finite temperatures [13–16]. Some of the ideas in these papers can be traced back to earlier work [17, 18]. Using different many-body techniques, these approaches lead to the same set of integral equations in the so-called non-crossing approximation (NCA). Usual Feynman diagram techniques can be used in the derivation if a slave boson is introduced [16]. There is a chapter on the slave-boson technique in this book.

At the time of these developments Wilson's numerical renormalization group method (NRG) to calculate ground state properties of the spin-degenerate Anderson impurity model numerically to arbitrary accuracy was known [19], but the extension to calculate the impurity spectral function came more than ten years after Wilson's work [20,21]. There are two chapters on the NRG in this book. Therefore it is not further discussed here.

Later the NCA was generalized to finite values of U [22–24] and further improved [25, 26] to correctly obtain the Kondo scale for $N_f = 2$. Another approach that circumvents many of the earlier deficiencies in the treatment of the spin-degenerate single-impurity Anderson model (SIAM) is the *local moment approach* [27]. It was later extended to include orbital degeneracy [28].

Exact results for ground-state and thermodynamic properties of the spin-degenerate Anderson impurity model were presented using the Bethe ansatz [29, 30]. Later, this approach was extended to models with large orbital degeneracy in the limit $U \rightarrow \infty$ [31]. Unfortunately results for spectral properties by this method do not (yet) exist.

Additional motivation for simple accurate calculational schemes for the impurity spectral function came later from the development of the dynamical mean-field theory (DMFT) [32], in which an extended lattice model of correlated electrons is mapped onto a SIAM with a coupling to a bath whose structure has to be determined self-consistently. Then the coupling of the impurity to the conduction band can have an arbitrary energy dependence.

In section 2 the Anderson impurity model in its basic form as well as the minimal model for an impurity with a core level in a metal are introduced. Important new aspects arising when the orbital degeneracy is taken into account are discussed in section 3. As a first example of the ideas of the $1/N_f$ expansion presented in the following, the ground state of the impurity system is discussed. In section 4 the intermediate states method is introduced and applied to the description of various spectroscopies. The comparison with spectroscopic measurements of mixed-valence compounds is briefly addressed in section 5.

2 Basic impurity models

2.1 Spin-degenerate single-impurity Anderson model

In order to study a single magnetic impurity in simple metals, P.W. Anderson proposed the Hamiltonian [2]

$$H_A = \sum_{\sigma} \left[\varepsilon_d n_{d\sigma} + \sum_k \varepsilon_k n_{k\sigma} + \sum_k V_{dk} \left(\psi_{d\sigma}^{\dagger} \psi_{k\sigma} + H.c. \right) \right] + U n_{d\uparrow} n_{d\downarrow}, \quad (1)$$

where $\psi_{d,\sigma}$ is the annihilation operator of the localized impurity $|d\rangle$ -state with energy ε_d , and the $\psi_{k,\sigma}$ are the annihilation operators of the delocalized band states $|k\rangle$ with energy ε_k . The $n_{d\sigma} = \psi_{d\sigma}^{\dagger} \psi_{d\sigma}$ (and $d \rightarrow k$) are particle number operators. In the body of his 1961 paper Anderson used the “physically unrealistic case” with only spin degeneracy and treated the case

of a doubly degenerate d -orbital in an appendix [2]. As a “physically realistic” case the spin-degenerate model was later used to describe hydrogen chemisorption on metal surfaces ($d \rightarrow a$ for “adsorbate”), where $|a\rangle$ corresponds to the hydrogen $1s$ -level [33]. The last term of the Anderson Hamiltonian describes the local Coulomb repulsion U , which acts when the d -level is doubly occupied. This two-body interaction makes the model highly non-trivial.

Experimentally relevant spectral functions are obtained from the one-particle Green’s functions. The general definition of the retarded functions [34] is

$$\langle\langle A; B \rangle\rangle_z \equiv -i \int_0^\infty \langle [A(t), B]_\pm \rangle e^{izt} dt, \quad (2)$$

where $A(t) = e^{iHt} A e^{-iHt}$ is the operator A in the Heisenberg picture, $\langle \rangle$ denotes the average over the grand canonical ensemble, and z is a complex variable with $\text{Im } z > 0$ in order to ensure the convergence of the time integral. For operators A involving products of an odd (even) number of fermion field operators the anticommutator $[,]_+$ (commutator $[,]_-$) is chosen. The Heisenberg equation of motion (EOM) for $A(t)$ and a partial integration yields the EOM

$$z \langle\langle A; B \rangle\rangle_z - \langle\langle [A, H]; B \rangle\rangle_z = \langle [A, B]_\pm \rangle. \quad (3)$$

This EOM is very useful for discussing the exactly solvable limits of the Anderson impurity model.

The retarded one-particle Green’s functions $G_{ij}(z)$ is obtained by $A \rightarrow \psi_i, B \rightarrow \psi_j^\dagger$

$$G_{ij}(z) \equiv \langle\langle \psi_i; \psi_j^\dagger \rangle\rangle_z. \quad (4)$$

At zero temperature the local Green’s function takes the form

$$\begin{aligned} G_{d\sigma, d\sigma}(z) &\equiv \left\langle E_0(N) \left| \psi_{d\sigma}^\dagger \frac{1}{z + H - E_0(N)} \psi_{d\sigma} + \psi_{d\sigma} \frac{1}{z - H + E_0(N)} \psi_{d\sigma}^\dagger \right| E_0(N) \right\rangle \\ &\equiv G_{d\sigma, d\sigma}^<(z) + G_{d\sigma, d\sigma}^>(z). \end{aligned} \quad (5)$$

The first term is relevant for photoemission and the second one for inverse photoemission. At finite temperature T the spectral functions are obtained as

$$\rho_{dd}^<(\varepsilon) = -\frac{1}{\pi} f(\varepsilon) \text{Im } G_{d\sigma, d\sigma}(\varepsilon + i0) \text{ and } \rho_{dd}^>(\varepsilon) = -\frac{1}{\pi} (1 - f(\varepsilon)) \text{Im } G_{d\sigma, d\sigma}(\varepsilon + i0), \quad (6)$$

where $f(\varepsilon) = 1/(e^{\beta\varepsilon} + 1)$ is the Fermi function with $\beta = 1/k_B T$ and the chemical potential is chosen as the zero of energy. This leads to

$$\rho_{dd}^>(\varepsilon) = e^{\beta\varepsilon} \rho_{dd}^<(\varepsilon). \quad (7)$$

This relation can be read as “photoemission determines inverse photoemission.” It has been used in the present context to get information about a Kondo resonance above the chemical potential by means of photoemission [35]. Unfortunately the relation is of limited practical use.

For $U = 0$ the Anderson impurity model describes noninteracting electrons and is exactly solvable. The generally valid EOMs follow from Eq. (3)

$$(z - \varepsilon_d) G_{d\sigma, d\sigma}(z) - U \langle \langle \psi_{d\sigma} n_{d-\sigma}; \psi_{d\sigma}^\dagger \rangle \rangle_z - \sum_k V_{dk} G_{k\sigma, d\sigma}(z) = 1, \quad (8)$$

$$(z - \varepsilon_k) G_{k\sigma, d\sigma}(z) - V_{dk}^* G_{d\sigma, d\sigma}(z) = 0.$$

For $U = 0$ these equations close and one obtains

$$G_{d\sigma, d\sigma}^{U=0}(z) = \frac{1}{z - \varepsilon_d - \Gamma(z)}, \quad \text{with } \Gamma(z) = \sum_k \frac{|V_{dk}|^2}{z - \varepsilon_k}. \quad (9)$$

For finite systems $\Gamma(z)$ has poles on the real axis. In the thermodynamic limit they go over to a branch cut on the real axis.

Using $\rho_{d\sigma, d\sigma}(\varepsilon) = -\text{Im } G_{d\sigma, d\sigma}(\varepsilon + i0)/\pi$ one obtains the impurity spectral function. The only information about the band states that enters is the coupling function $\Gamma(\varepsilon + i0)$. It determines the width and location of the resonance resulting from the coupling. An often-used approximation for the coupling function is the wide-band limit $\Gamma(\varepsilon \pm i0) = \mp i\Gamma$ with a constant Γ . Then the $U = 0$ impurity spectral function $\rho_{d\sigma, d\sigma}(\varepsilon)$ has a Lorentzian peak of half-width Γ at ε_f . The mathematical structure of the results of the noninteracting limit of the Anderson model first appeared in earlier models by K.O. Friedrichs [36] and T.D. Lee [37].

A long history exists of attempts to solve the Anderson impurity model for finite values of the Coulomb interaction U . It started with Anderson using the Hartree-Fock (HF) approximation [2]

$$U n_{d\uparrow} n_{d\downarrow} \rightarrow U \left(n_{d\uparrow} \langle n_{d\downarrow} \rangle^{HF} + n_{d\downarrow} \langle n_{d\uparrow} \rangle^{HF} - \langle n_{d\uparrow} \rangle^{HF} \langle n_{d\downarrow} \rangle^{HF} \right), \quad (10)$$

which corresponds to a noninteracting model with the bare f -level position given by the replacement $\varepsilon_d \rightarrow \varepsilon_{d\sigma}^{HF} = \varepsilon_d + U \langle n_{d-\sigma} \rangle^{HF}$. This leads to

$$G_{d\sigma, d\sigma}^{HF}(z) = \frac{1}{z - \varepsilon_d - U \langle n_{d-\sigma} \rangle - \Gamma(z)}. \quad (11)$$

In the language of the EOMs the HF-approximation corresponds to the factorization of the retarded function $\langle \langle \psi_{d\sigma} n_{d-\sigma}; \psi_{d\sigma}^\dagger \rangle \rangle_z \rightarrow \langle n_{d-\sigma} \rangle \langle \langle \psi_{d\sigma}; \psi_{d\sigma}^\dagger \rangle \rangle_z$. The discussion of the results of the HF-approximation simplifies in the particle-hole-symmetric case $\varepsilon_d + U/2 = \mu = \varepsilon_F = 0$ for a symmetric band around the chemical potential. Then $\langle n_{d-\sigma} \rangle^{RHF} = 1/2 = \langle n_{d,-\sigma} \rangle^{exact}$ and the R(estricted)HF resonance is at the chemical potential. The shape and position of this RHF spectral function is *independent* of the value of U in this particle-hole symmetric case. At a critical value $U/\Gamma = \pi$ solutions of the HF-equations occur where the occupancies of the impurity level for spin-up and spin-down differ [2]. These “unrestricted Hartree-Fock” (UHF) solutions are an artifact of the approximation as no spontaneous symmetry breaking can occur when the interaction acts in a zero dimensional system. Therefore the spin variable in the Green’s function is suppressed in (most of) the following (e.g. $d\sigma \rightarrow d$).

In order to properly describe the U -dependence of G_{dd} , a better treatment of the self-energy $\Sigma(z)$ defined in the usual way

$$G_{dd}(z) = \frac{1}{z - \varepsilon_d - \Gamma(z) - \Sigma(z)} \quad (12)$$

is necessary. The first order contribution in U to the self-energy Σ is just the HF-term $U \langle n_{d,-\sigma} \rangle$. In the particle-hole symmetric case and the wide-band limit the spectral function takes the form ($\tilde{\Sigma} \equiv \Sigma - U \langle n_{d,-\sigma} \rangle$)

$$\rho_{dd}(\varepsilon) = \frac{1}{\pi} \frac{\Gamma + |\text{Im}\tilde{\Sigma}(\varepsilon + i0)|}{\left(\varepsilon - \text{Re}\tilde{\Sigma}(\varepsilon) \right)^2 + \left(\Gamma + |\text{Im}\tilde{\Sigma}(\varepsilon + i0)| \right)^2} \quad (13)$$

with $\text{Re}\tilde{\Sigma}(\varepsilon)$ an even function of ε . At zero temperature the Fermi liquid property holds: $\text{Im}\tilde{\Sigma}(\varepsilon + i0) \sim \varepsilon^2$ for $\varepsilon \rightarrow 0$. This can be seen easily for the self-energy to second order in U and has been discussed to arbitrary order by Yamada and Yoshida [38]. This implies the exact result for the particle-hole symmetric case at $T = 0$:

$$\rho_{dd}(0) = \rho_{dd}^{RHF}(0) = \frac{1}{\pi\Gamma} . \quad (14)$$

Important additional insights into the energy dependence of $\rho_{dd}(\varepsilon)$ are obtained by considering the exactly solvable *atomic limit*, in which $V_{dk} = 0$ for all values of k . Again the EOMs close, as $(z - \varepsilon_d - U) \langle \langle \psi_{d\sigma} n_{d-\sigma}; \psi_{d\sigma}^\dagger \rangle \rangle_z = \langle n_{d-\sigma} \rangle$ holds for vanishing coupling to the conduction band. Keeping the spin indices one obtains

$$G_{d\sigma,d\sigma}^{V=0}(z) = \frac{1 - \langle n_{d-\sigma} \rangle}{z - \varepsilon_d} + \frac{\langle n_{d-\sigma} \rangle}{z - (\varepsilon_d + U)} . \quad (15)$$

Here we only discuss the most interesting case when ε_d is below the Fermi energy and $\varepsilon_d + U$ is above it. Then the total occupancy of the impurity level is approximately one, which holds exactly in the particle-hole symmetric case. If an electron is removed from the impurity level the empty impurity state created can decay by the tunneling back in of a spin-up *or* spin-down electron, which gives the “atomic peaks” a width *twice* as large as the width of the RHF-Lorentzian:

$$G_{dd}(z) \approx \frac{1/2}{z - U/2 - 2\Gamma(z)} + \frac{1/2}{z + U/2 - 2\Gamma(z)} . \quad (16)$$

A formal way to obtain this result is to calculate a properly defined self-energy matrix to second order in the hybridization V [39]. For $U \gg \Gamma$ most of the spectral weight is in the atomic peaks that are assumed to be well described. At the chemical potential, the spectral weight vanishes in this large- U limit as $\sim \Gamma/U^2$ instead of yielding the exact result $1/(\pi\Gamma)$. This implies that a very narrow peak at the Fermi energy is missing.

Both approximations for the local spectral function in the particle-hole symmetric case presented so far are unable to properly describe this *Kondo resonance* at the Fermi energy. An exact (numerical) calculation was only presented in the eighties with the help of the numerical renormalization group (NRG) [20]. The exact NRG result in the wide band limit for $U/\Gamma = 4\pi$ is shown in Fig. 1. The RHF result agrees with the exact NRG result only for $\varepsilon = 0$ but otherwise fails badly. The naive perturbation theory around the atomic limit fails badly in the low energy region. For larger values of U/Γ than shown in Fig. 1 the high energy features near $\pm U/2$ agree better with the NRG-result.

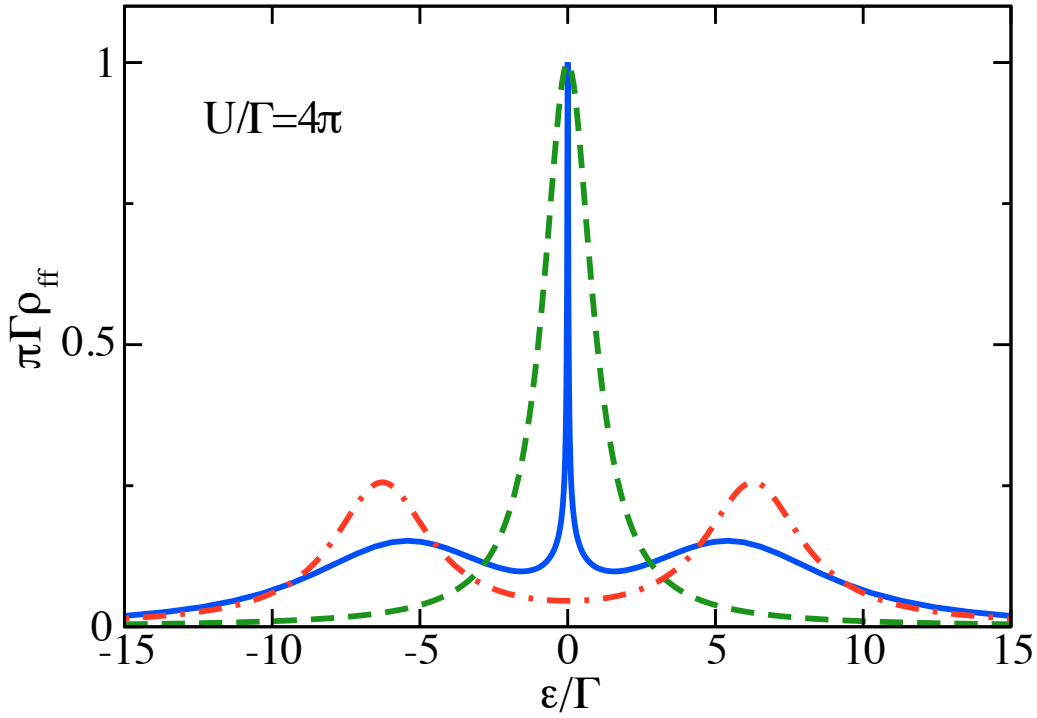


Fig. 1: Result for the impurity spectral function of the spin-symmetric Anderson model in the particle-hole symmetric case in the wide band limit for $U/\Gamma = 4\pi$: exact result from the numerical renormalization group (NRG) with the Kondo peak at the Fermi energy (full line); restricted Hartree-Fock approximation (dashed line); perturbation theory around the atomic limit (dashed-dotted line)

The simple arguments presented in favor of the Kondo resonance in the particle-hole symmetric case give no information about its width and its precise location if $|\varepsilon_d| \neq |\varepsilon_d + U|$. Before the NRG results were available, it was therefore useful to obtain partial answers to these questions in the limit of large additional orbital degeneracy of the impurity level. This is discussed in detail in the following sections.

There is a long history of attempts to obtain a controlled approximation for G_{dd} and the corresponding spectral function that cannot be presented here in detail. We shortly mention decoupling schemes of higher order Green's functions that appear in the EOM of $\langle\langle\psi_{d\sigma}n_{d-\sigma};\psi_{d\sigma}^\dagger\rangle\rangle_z$ or in the EOMs of higher order [40–44]. The quality of the results for G_{dd} is generally hard to judge. The resulting spectral functions can have frequency regions with *negative* spectral weight [43]. Special attention to the large N_f limit has been given by Czyrcholl [44]. To leading order in $1/N_f$, he obtains at zero temperature a sharp peak at the correct Kondo energy.

Additional information on the attempts to understand the physics of the SIAM can be found in the book by A. Hewson [45].

2.2 Impurity models involving core levels

X-ray absorption spectroscopy and X-ray photoemission spectroscopy (XPS) of the core levels of an impurity are useful tools for obtaining information about the properties of the valence electrons. In a minimal model, a single nondegenerate core level of the impurity with energy ε_c is considered, which is filled in the initial state. The creation of the core hole in the photoemission process leads to an additional attractive potential for the valence level of the impurity which lowers it by an amount U_{dc} . The corresponding model Hamiltonian reads

$$H_{tot} = H_A + \varepsilon_c n_c - U_{dc}(1 - n_c) \sum_{\sigma} n_{d\sigma}. \quad (17)$$

As the ground state of the combined system has the form $\psi_c^\dagger |E_0(N)\rangle$, with $|E_0(N)\rangle$ the ground state of the valence system with the core hole present, the time evolution of the remaining pure valence system after removing the core electron is described by the modified Anderson Hamiltonian $\tilde{H}_A$, with the energy ε_d of the impurity level replaced by $\varepsilon_d - U_{dc}$. The creation of the core hole acts as a *quantum quench* for the valence system. The core spectral function is

$$\rho_{cc}(\varepsilon) = \left\langle E_0(N) \left| \delta(\varepsilon - \varepsilon_c - E_0(N) + \tilde{H}_A) \right| E_0(N) \right\rangle. \quad (18)$$

For the case of *noninteracting* valence electrons, i.e., $U = 0$ in Eq. (1), this problem falls into the class of the famous X-ray edge singularity problem [46]. The sharp core-level spectrum without the presence of the valence electrons is replaced by a continuum with a *power law singularity* at the high-energy edge. This is closely related to the Anderson orthogonality catastrophe [47] which states that the overlap of the ground states of H_A and $\tilde{H}_A$ vanishes with a power law in $1/N$ when the number of electrons N tends to infinity. The core-level spectrum can show satellite peaks corresponding to higher-energy eigenstates of $\tilde{H}_A$ due to physical processes that occur on a *finite* time scale [48–50]. This has been addressed in detail, e.g., for core levels of adsorbates at metal surfaces [50]. For small coupling Γ a high energy resonance dominates the core level spectrum if the adsorbate level, initially well above the chemical potential, is pulled well below it when the core hole is created. For finite Coulomb interaction U , the problem cannot be solved exactly, and various approximations were proposed [51, 52]. The treatment within the large-degeneracy limit is discussed in section 4.

3 Anderson impurity model with large orbital degeneracy

Despite the fact that Anderson proposed his model to treat transition-metal impurities in simple metals, the five-fold orbital degeneracy of *d*-orbitals was not treated explicitly. The degeneracy of *f*-orbitals is given by $N_f = 14$ if spin degeneracy is included and spin-orbit coupling and crystal-field splitting is neglected. As mentioned in the introduction, an important progress in the treatment of the Anderson model was the realization that the treatment of $1/N_f$ as a small parameter allows new approximation schemes [6, 7]. In the SIAM Hamiltonian, Eq. (1), the

non-degenerate orbital label d is replaced by the orbital quantum number m and V_{kd} by V_{km} . We assume that in the thermodynamic limit

$$\sum_k V_{km}^* V_{km'} \delta(\varepsilon - \varepsilon_k) = V(\varepsilon)^2 \delta_{mm'} \quad (19)$$

holds [9, 10]. It is useful to introduce new one-particle states

$$|\varepsilon, m\sigma\rangle \equiv V(\varepsilon)^{-1} \sum_k V_{km} \delta(\varepsilon - \varepsilon_k) \quad (20)$$

and to use the combined degeneracy index $\nu \equiv m\sigma$. The orthogonality relation of these states reads $\langle \varepsilon, \nu | \varepsilon', \nu' \rangle = \delta_{\nu\nu'} \delta(\varepsilon - \varepsilon')$.

Despite the fact that it is mathematically more appropriate to write down the many-body Hamiltonian for finite systems and take the thermodynamic limit in the end of the calculation, in the following we formally write it down using creation and annihilation operators of the states defined in Eq. (20). To avoid problems, one can discretize the energies ε and replace the Dirac delta functions $\delta(\varepsilon - \varepsilon')$ by Kronecker deltas $\delta_{\varepsilon\varepsilon'}$. This is done anyway in the higher order numerical treatment of the $1/N_f$ scheme presented in the following sections [9]. Alternatively, one has to subtract the (infinite) energy of the filled Fermi sea.

Keeping these precautions in mind, the N_f -fold-degenerate single-impurity Anderson Hamiltonian used in the following reads

$$H = \sum_{\nu=1}^{N_f} \left[\varepsilon_f \psi_\nu^\dagger \psi_\nu + \int \varepsilon \psi_{\nu\varepsilon}^\dagger \psi_{\nu\varepsilon} d\varepsilon + \int (V(\varepsilon) \psi_\nu^\dagger \psi_{\nu\varepsilon} + H.c.) d\varepsilon \right] + U \sum_{\nu < \mu} n_\nu n_\mu. \quad (21)$$

The Hamiltonian $\tilde{H}_0$, which contains linear combinations of conduction states that do not couple to the f -level, is not included. It can be neglected for the properties studied here.

As in the following $V(\varepsilon)$ enters in the combinations $N_f V(\varepsilon)^2$ and $V(\varepsilon)^2$, it is useful to define

$$\tilde{V}(\varepsilon) \equiv \sqrt{N_f} V(\varepsilon) \quad (22)$$

and require that $\tilde{V}(\varepsilon)$ is *independent* of the degeneracy N_f . This simplifies the discussion of the large degeneracy limit $N_f \rightarrow \infty$.

3.1 Ground-state calculation

The ground-state calculation is performed variationally by classifying the many-electron states shown in Fig. 2 in orders of $1/N_f$.

In the state $|0\rangle$, all conduction states below the Fermi energy are filled, and the f -level is empty. This state couples via H to the states denoted a in Fig. 2. They are of the form

$$|\varepsilon\rangle = \frac{1}{\sqrt{N_f}} \sum_\nu \psi_\nu^\dagger \psi_{\varepsilon\nu} |0\rangle \quad (23)$$

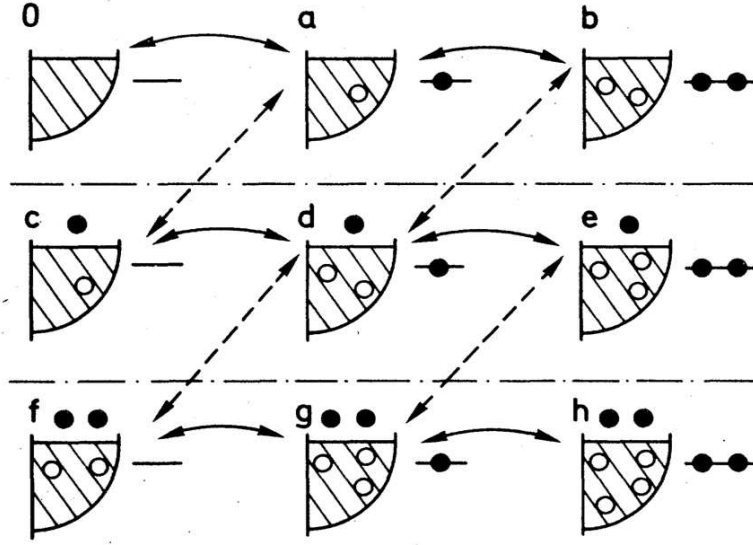


Fig. 2: Schematic representation of the many-electron basis states. Solid circles show electrons and open circles show holes. The hatched part indicates the filled conduction bands and the horizontal lines the f -level. The arrows show which states couple to each other. A solid line indicates the strength $\tilde{V}$ and a dashed line the strength $\tilde{V}/\sqrt{N_f}$.

in which a conduction electron has hopped into the f -level. These states couple to the b states with two electrons in the f -level

$$|\varepsilon, \varepsilon'\rangle = \frac{1}{\sqrt{N_f(N_f - 1)}} \sum_{\nu \neq \nu'} \psi_{\nu}^{\dagger} \psi_{\varepsilon\nu} \psi_{\nu'}^{\dagger} \psi_{\varepsilon'\nu'} |0\rangle, \quad (24)$$

and to the c states with a conduction electron-conduction hole pair

$$|E\varepsilon\rangle = \frac{1}{\sqrt{N_f}} \sum_{\nu} \psi_{E\nu}^{\dagger} \psi_{\varepsilon\nu} |0\rangle, \quad (25)$$

where E refers to a conduction electron state above the Fermi level ($E > \varepsilon_F$). Further states in Fig. 2 can easily be written down [10].

The matrix elements coupling these states are given by

$$\langle \varepsilon | H | 0 \rangle = \tilde{V}(\varepsilon), \quad (26)$$

$$\langle \varepsilon, \varepsilon' | H | \varepsilon'' \rangle = \sqrt{1 - 1/N_f} \left(\tilde{V}(\varepsilon') \delta(\varepsilon - \varepsilon'') + \tilde{V}(\varepsilon) \delta(\varepsilon' - \varepsilon'') \right), \quad (27)$$

$$\langle E\varepsilon | H | \varepsilon' \rangle = \tilde{V}(E)/\sqrt{N_f} \delta(\varepsilon - \varepsilon'). \quad (28)$$

These examples illustrate the general result that within each row in Fig. 2 there are states that couple with strength $\tilde{V}$, while states in different rows at most couple with a strength $\tilde{V}/\sqrt{N_f}$. This allows one to classify the states in orders of $1/N_f$ according to their contribution to the ground state. The states in the first, second and third rows are of the orders $(1/N_f)^0$, $(1/N_f)^1$ and $(1/N_f)^2$, respectively.

As an illustration, we calculate the ground state for $U = \infty$ to lowest order in $1/N_f$. It is written as [53]

$$|E_0\rangle^{(0)} = A \left[|0\rangle + \int_{-B}^0 d\varepsilon a(\varepsilon) |\varepsilon\rangle \right], \quad (29)$$

where the normalization constant A is related to the total occupancy n_f of the f -level by $A^2 = 1 - n_f$. In contrast to the ground-state energy $E_0^{(0)}$, the difference $\Delta E_0 \equiv E_0^{(0)} - \langle 0|H|0\rangle$ is finite also in the thermodynamic limit. Using the coupling matrix elements Eq. (26) and $\langle \varepsilon|(H - \langle 0|H|0\rangle)|\varepsilon'\rangle = (\varepsilon_f - \varepsilon)\delta(\varepsilon - \varepsilon')$, the Schrödinger equation leads to

$$\Delta E_0 = \int_{-B}^0 \tilde{V}(\varepsilon) a(\varepsilon) d\varepsilon \quad \text{and} \quad (\Delta E_0 - \varepsilon_f + \varepsilon) a(\varepsilon) = \tilde{V}(\varepsilon). \quad (30)$$

Therefore ΔE_0 obeys the implicit equation

$$\Delta E_0 = \int_{-B}^0 \frac{\tilde{V}(\varepsilon)^2}{\Delta E_0 - \varepsilon_f + \varepsilon} d\varepsilon \rightarrow \tilde{V}^2 \ln \frac{\varepsilon_f - \Delta E_0}{\varepsilon_f - \Delta E_0 + B}, \quad (31)$$

where the energy integration was performed for the case of an energy-independent $\tilde{V}$. We discuss the solution for the case of constant $\tilde{V}$ and ε_f well below the Fermi energy. Defining the (positive) $\delta \equiv \varepsilon_f - \Delta E_0$, $\tilde{\Delta} \equiv \pi \tilde{V}^2$, and $\tilde{\varepsilon}_f \equiv \varepsilon_f + (\tilde{\Delta}/\pi) \ln(\pi B/\tilde{\Delta})$ the equation for δ simplifies in the Kondo-limit $-\tilde{\varepsilon}_f \gg \tilde{\Delta}$ to

$$\delta = \frac{\tilde{\Delta}}{\pi} e^{\pi(\tilde{\varepsilon}_f - \delta)/\tilde{\Delta}} \rightarrow \delta \approx \frac{\tilde{\Delta}}{\pi} e^{\pi\tilde{\varepsilon}_f/\tilde{\Delta}}. \quad (32)$$

The coefficient

$$a(\varepsilon)^2 = \frac{\tilde{\Delta}}{\pi} \frac{1}{(\varepsilon - \delta)^2} \quad (33)$$

grows on the energy scale δ as the Fermi energy is approached from below. The total f -occupancy is determined by $\int a(\varepsilon)^2 d\varepsilon$. For the case of an energy-independent $\tilde{V}$, one obtains $n_f = \tilde{\Delta}/(\tilde{\Delta} + \pi\delta)$ [9]. The energy scale δ depends *exponentially* on $\pi\tilde{\varepsilon}_f/\tilde{\Delta}$, which suggests that it can be, apart from a factor given by the Boltzmann constant k_B , interpreted as the *Kondo temperature*: $T_K = k_B\delta$. This will be further examined in the following sections.

The infinite- U , lowest-order calculation presented above can be extended to the case when the spin-orbit splitting $\Delta\varepsilon_f$ is taken into account [9]. The single f -level (with $N_f = 14$) is replaced by two levels (with $N_{f_1} = 6$ and $N_{f_2} = 8$ for $j = 5/2$ and $j = 7/2$) at ε_f and $\varepsilon_f + \Delta\varepsilon_f$. For the description of high-resolution experimental spectra of Ce compounds it is important to include the spin-orbit splitting [54, 55].

The $(1/N_f)^0$ calculation of the ground state can also be extended to the *finite* U case. If an infinite three-body interaction is assumed, one just has to take the b states in Fig. 2 into account with an additional term with coefficients $b(\varepsilon, \varepsilon')$ in Eq. (29). The Schrödinger equation then leads to an integral equation for $a(\varepsilon)$ that for $U \gg B$ is of separable form [11]. For a detailed discussion of the explicit bandwidth behavior of the energy difference $\delta(U)$ between the nonmagnetic ground state and the lowest magnetic states that are not totally symmetric in the degeneracy indices, see Ref. [11].

Numerical ground-state calculations of higher order in $1/N_f$ using the states shown in Fig. 2 quickly converge for $N_f = 14$ [9, 10].

4 The intermediate states method for spectra

The theoretical description of photoemission simplifies considerably when the emitted electron in the state $|\kappa\rangle$ is assumed to have no interaction with the remaining $(N - 1)$ -electron system. This *sudden approximation* becomes increasingly accurate as the kinetic energy of the emitted electron is increased. In this approximation the photoelectron current can be calculated using Fermi's golden rule. For a weak energy dependence of the matrix elements $\tau_{\kappa i}$ of the dipole operator, where $|i\rangle$ is a valence state, the current is directly related to the spectral function of one-particle Green' functions $G_{ii}^<$ when interference effects are neglected [56, 10].

In Eq. (5) the zero-temperature local one-particle Green's functions $G^<$ and $G^>$ are expressed as an expectation value of the resolvent of the many body Hamiltonian H . One obtains the well known Lehmann representation by inserting the complete set of $(N \mp 1)$ -electron eigenstates of H . For $G^<$ one can alternatively use the resolution of unity made of an arbitrary complete set $\{|i\rangle\}$ of $(N - 1)$ -electron basis states

$$G_{\nu\nu}^<(z) = \sum_{ij} \langle E_0(N) | \psi_\nu^\dagger | i \rangle \langle i | (z + H - E_0(N))^{-1} | j \rangle \langle j | \psi_\nu | E_0(N) \rangle . \quad (34)$$

The inversion of the matrix $\tilde{H}(z)_{ij} \equiv \langle i | (z + H - E_0(N)) | j \rangle$ would lead to the exact result for $G_{\nu\nu}^<(z)$

$$G_{\nu\nu}^<(z) = \sum_{ij} \langle E_0(N) | \psi_\nu^\dagger | i \rangle (\tilde{H}(z)^{-1})_{ij} \langle j | \psi_\nu | E_0(N) \rangle \quad (35)$$

if the procedure could actually be carried out for a complete set of states. Approximations can be obtained by truncating the set $\{|i\rangle\}$ of *intermediate states*. Useful results can be obtained again using a classification of the states according to their contribution in orders of $1/N_f$. For the calculation of $G^>$ one can proceed the same way but with $(N + 1)$ -electron intermediate states $\{|i\rangle\}$.

4.1 Valence photoemission spectroscopy

Again, we first consider the $U = \infty$ case and work to lowest order in $1/N_f$. Then the ground state is described by Eq. (29). As $\psi_\nu | E_0 \rangle^{(0)} = A \int d\varepsilon a(\varepsilon) \psi_{\varepsilon\nu} | 0 \rangle / \sqrt{N_f}$, we introduce the basis states $|\varepsilon\nu\rangle$ which via H couple to the states $|\varepsilon\varepsilon'\nu\rangle$, where

$$|\varepsilon\nu\rangle \equiv \psi_{\varepsilon\nu} | 0 \rangle, \quad |\varepsilon\varepsilon'\nu\rangle \equiv \frac{1}{\sqrt{N_f}} \sum_{\nu'} \psi_{\nu'}^\dagger \psi_{\varepsilon'\nu'} \psi_{\varepsilon\nu} | 0 \rangle . \quad (36)$$

The matrix $\tilde{H}(z)$ defined before Eq. (35) has the matrix elements

$$\tilde{H}(z)_{\varepsilon,\varepsilon'} = (z - \Delta E_0 - \varepsilon) \delta(\varepsilon - \varepsilon') , \quad (37)$$

$$\tilde{H}(z)_{\varepsilon\varepsilon',\varepsilon''} = \tilde{V}(\varepsilon') \delta(\varepsilon - \varepsilon'') - \tilde{V}(\varepsilon) / \sqrt{N_f} \delta(\varepsilon' - \varepsilon'') , \quad (38)$$

$$\tilde{H}(z)_{\varepsilon\varepsilon',\varepsilon_1\varepsilon_2} = (z - \Delta E_0 + \varepsilon_f - \varepsilon - \varepsilon') \delta(\varepsilon - \varepsilon_1) \delta(\varepsilon - \varepsilon_2) . \quad (39)$$

In leading order we neglect the term $\sim 1/\sqrt{N_f}$ on the right-hand side of the second equation. This leads to the simplification that for each $|\varepsilon\nu\rangle$ one can treat the coupling of this state to a continuum of states with an additional hole at $\varepsilon' < \varepsilon_F = 0$ separately. This greatly simplifies the leading-order calculation of $G_{\nu\nu}^<$.

For the inversion of $\tilde{H}(z)$, it is convenient to use a block matrix form with elements $\tilde{H}_{11}$, $\tilde{H}_{12}$, $\tilde{H}_{21}$ and $\tilde{H}_{22}$, where e.g. $\tilde{H}_{11}$ refers to the $\tilde{H}(z)_{\varepsilon,\varepsilon'}$ and $\tilde{H}_{22}$ to $\tilde{H}(z)_{\varepsilon\varepsilon',\varepsilon_1\varepsilon_2}$. The well known matrix inversion formula

$$\left(\tilde{H}^{-1}\right)_{11} = \left(\tilde{H}_{11} - \tilde{H}_{12}\tilde{H}_{22}^{-1}\tilde{H}_{21}\right)^{-1} \quad (40)$$

simplifies the calculation. Since $\tilde{H}_{22}$ is diagonal its inversion is trivial and one obtains

$$(\tilde{H}(z)^{-1})_{\varepsilon\varepsilon'} = \tilde{g}(z - \Delta E_0 + \varepsilon_f - \varepsilon)\delta(\varepsilon - \varepsilon'), \quad (41)$$

where

$$\tilde{g}(z) = \frac{1}{z - \varepsilon_f - \tilde{\Gamma}(z)} \quad \text{with} \quad \tilde{\Gamma}(z) = \int_{-B}^0 \frac{\tilde{V}(\varepsilon)^2}{z - \varepsilon} d\varepsilon. \quad (42)$$

Note that the energy integration in the definition of $\tilde{\Gamma}(z)$ only extends to $\varepsilon_F = 0$. The function $\tilde{g}(z)$ has the form of the f Green's function of a noninteracting Anderson model with a sharp band cut-off at $\varepsilon = 0$. Finally, performing one of the energy integrations with the help of the delta function in Eq. (41), one obtains for $G_f^<$ to leading order in $1/N_f$

$$G_{\nu\nu}^<(z) = \frac{1}{N_f} A^2 \int_{-B}^0 a(\varepsilon')^2 \tilde{g}(z - \Delta E_0 + \varepsilon_f - \varepsilon') d\varepsilon'. \quad (43)$$

The function $\text{Im } \tilde{g}(\varepsilon \pm i0)$ has a continuum part for $-B \leq \varepsilon \leq 0$ due to the imaginary part of $\tilde{\Gamma}(\varepsilon \pm i0)$. As the transcendental equation (31) for $\Delta E_0^{(0)}$ can be written as $\Delta E_0 = -\tilde{\Gamma}(\delta)$, the function $\tilde{g}(z)$ has a pole at $z = \delta = \varepsilon_f - \Delta E_0$. The strength of the pole $(1 - d\tilde{\Gamma}/dz)^{-1}$ evaluated at $z = \delta$ is given by $1 - n_f$. This pole of $\tilde{g}$ yields for the total f spectral function $\rho_f^<(\varepsilon) = -\sum_{\nu} \text{Im } G_{\nu\nu}^<(\varepsilon + i0)/\pi$ using Eq. (33) and $A^2 = 1 - n_f$ the expression

$$\rho_f^<(\varepsilon) = \frac{(1 - n_f)^2 \tilde{V}(\varepsilon)^2}{(\delta - \varepsilon)^2} \quad \text{for} \quad -\delta \leq \varepsilon \leq 0. \quad (44)$$

There is also a (partial) contribution of this type to $\rho_f^<(\varepsilon)$ for $-B \leq \varepsilon \leq \delta$. As $A^2 \int a(\varepsilon)^2 d\varepsilon = n_f$, the total weight of $\rho_f^<$ resulting from the pole of $\tilde{g}(z)$ at $z = \delta$ is given by $n_f(1 - n_f)$. It becomes very small in the Kondo limit $n_f \approx 1$.

For $\varepsilon < -\delta$ the continuum part of $\text{Im } \tilde{g}(\varepsilon \pm i0)$ also contributes to $\rho_f^<(\varepsilon)$, and there can be a split-off state below the conduction band in $\tilde{g}(z)$.

The low-energy spectral weight described in Eq. (44) rises sharply as ε approaches $\varepsilon_F = 0$ from below. It is the tail of the Kondo resonance present at $\varepsilon \approx \delta$ in the spectral function $\rho_f^>(\varepsilon)$ describing *inverse* photoemission. This is discussed in the next subsection. This low-energy behavior is totally different from the noninteracting case for $N_f = 1$. When ε_f lies below the Fermi energy, the f -spectral density in the $N_f = 1$ case has an ionization peak near ε_f and the spectral density *decreases* when ε approaches $\varepsilon_F = 0$ from below. In the Kondo limit

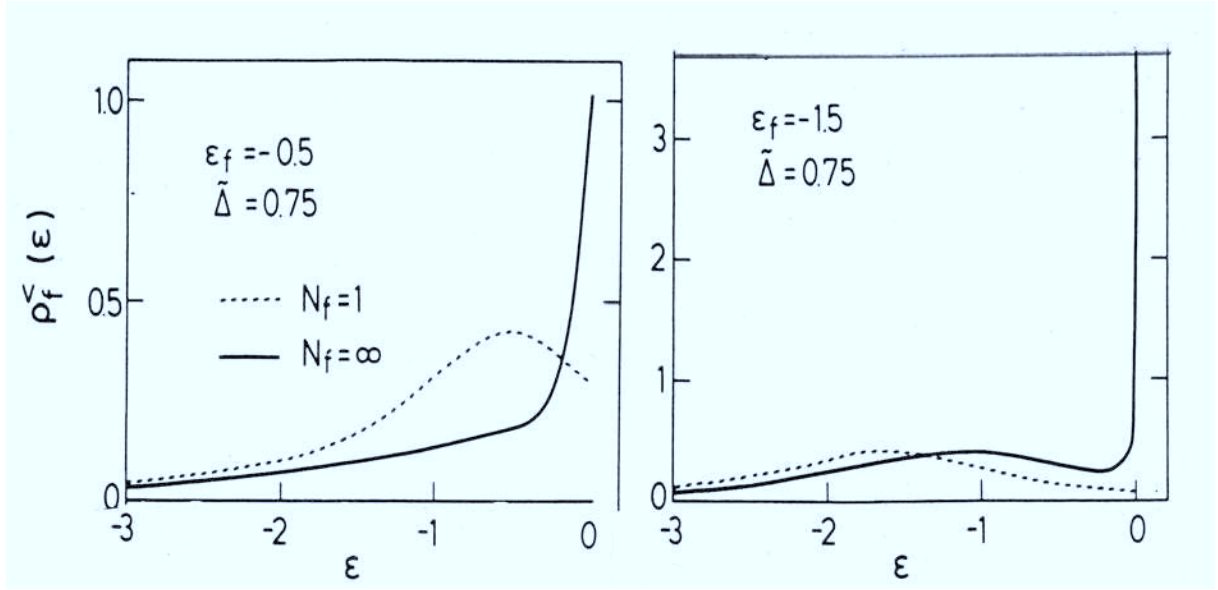


Fig. 3: Comparison of the leading order result for $\rho_f^<(\varepsilon)$ (full lines) with the result of a $N_f = 1$ calculation where Δ is replaced by $\tilde{\Delta}$ (dotted lines) for two different values of ε_f .

$-\varepsilon_f \gg \tilde{\Delta}$, implying $n_f \approx 1$, a similar ionization peak near ε_f dominates $\rho_f^>(\varepsilon)$. In this limit the energy integration with $A^2 a(\varepsilon')^2$ in Eq. (43) for $-\varepsilon \gg \delta$ approximately acts like (one-sided) delta function at the origin and $\rho_f^<(\varepsilon) \approx -\text{Im} \tilde{g}(\varepsilon + i0)/\pi$ holds. The width of this peak is given by $\tilde{\Delta} = N_f \Delta$, where $\Delta = \pi V^2(\varepsilon_f)$ is the half-width of the model for $N_f = 1$. After removing an f -electron from the ground state given by Eq. (29), the probability that a conduction electron in channel ν with energy $\varepsilon \approx \varepsilon_f$ hops into the f -level is given by Δ . Since there are N_f such channels the width is given by $N_f \Delta = \tilde{\Delta}$.

In Fig. 3, we compare results for the leading order result of $\rho_f^>(\varepsilon)$ with the result of a $N_f = 1$ calculation where Δ is replaced by $\tilde{\Delta}$. For both cases shown ε_f is below the Fermi level. In the left part of the figure $|\varepsilon_f| < \tilde{\Delta}$ and the ionization peak of the $N_f = 1$ spectrum only shows as a shoulder in the leading-order result for $\rho_f^>(\varepsilon)$. In the right half of the figure $|\varepsilon_f| = 2\Delta$ and the ionization peak is more asymmetric than the $N_f = 1$ result. This is similar to Fig. 1 where in the exact NRG result the atomic peaks are more asymmetric than the result from simple perturbation theory around the atomic limit.

The leading order calculation of $G_{\nu\nu}^<(z)$ can again be extended to the case of including spin-orbit splitting [9] and the case of finite U [11].

4.2 Inverse Photoemission

In inverse photoemission, earlier called Bremsstrahlung isochromat spectroscopy (BIS), the sample is bombarded by electrons that make radiative transitions into lower-lying $(N + 1)$ -electron states. Here we discuss transitions into the f -level. The theoretical description is in terms of

$$G_{\nu\nu}^>(z) = \left\langle E_0(N) \left| \psi_\nu \frac{1}{z - H + E_0(N)} \psi_\nu^\dagger \right| E_0(N) \right\rangle. \quad (45)$$

As the integrated weight of the total spectral function $\rho_{\nu\nu} = \rho_{\nu\nu}^< + \rho_{\nu\nu}^>$ is unity and $\int \rho_{\nu\nu}^<(\varepsilon) d\varepsilon = n_f/N_f$ holds, with $n_f \leq 1$ in the infinite U case, the integrated weight of $\rho_{\nu\nu}^>$ is given by $1 - n_f/N_f$, i.e., it is larger by a factor of N_f than the integrated weight of $\rho_{\nu\nu}^<$. This is a clear hint that a $1/N_f$ approximation for the full $G_{\nu\nu}$ is problematic. If it does not fulfill $\rho_{\nu\nu}^<(0) = \rho_{\nu\nu}^>(0)$, which is expected for an exact description at any finite N_f , this is an indication of the requirement to treat $G_{\nu\nu}^>$ differently from $G_{\nu\nu}^<$.

If in Eq. (45) the ground state to leading order in $1/N_f$ Eq. (29) is used, one has to calculate the expectation value of the resolvent of the many-body Hamiltonian with

$$\psi_\nu^\dagger |E_0\rangle^{(0)} = A \left[\psi_\nu^\dagger |0\rangle + \frac{1}{\sqrt{N_f}} \sum_{\nu' (\neq \nu)} \int d\varepsilon a(\varepsilon) \psi_\nu^\dagger \psi_{\nu'}^\dagger \psi_{\varepsilon\nu'} |0\rangle \right]. \quad (46)$$

In the first state on the right-hand side the f -level is singly occupied (f^1), while in the states in the second term it is doubly occupied (f^2). Integrating the corresponding expectation values of $\delta(\varepsilon - H + E_0(N))$ shows that the total weight of the f^1 contribution is given by $1 - n_f$ and the f^2 weight by $n_f(1 - 1/N_f)$. For large values of U the two different contribution are energetically well separated.

In a first attempt, one would take the states on the right-hand side of Eq. (46) as the intermediate states to calculate $G_{\nu\nu}^>$. If one focuses on the f^1 -peak in the $U \rightarrow \infty$ limit, only the state $|\nu\rangle = \psi_\nu^\dagger |0\rangle$ plays a role, and one obtains

$$G_{\nu\nu}^>(z) \approx \frac{1 - n_f}{z + \Delta E_0 - \varepsilon_f} = \frac{1 - n_f}{z - \delta}. \quad (47)$$

In this approximation $\rho_{\nu\nu}^>$ has a delta peak at $\varepsilon_f - \Delta E_0 = \delta$. For ε_f well above the Fermi energy $\varepsilon_F = 0$ one has $n_f \ll 1$, and $|\Delta E_0|$ is small compared to ε_f . This leads to a Delta peak of weight ≈ 1 close to ε_f . This is the atomic limit of the trivial empty-level case. Lowering ε_f lowers the peak position but it stays above the Fermi energy. For ε_f well below the Fermi energy the peak is very close to the Fermi energy as $\delta/\tilde{\Delta}$ is exponentially small in $\pi\tilde{\varepsilon}_f/\Delta$ (see Eq. (32)). As $1 - n_f \ll 1$, the weight of the peak is very small. In fact, in this approximation the Kondo peak has zero width.

Obviously when using this leading order description for the calculation of $\rho_{\nu\nu}^<$, the condition $\rho_{\nu\nu}^<(0) = \rho_{\nu\nu}^>(0)$ is not fulfilled. In order to achieve this, one has to go one order higher in $1/N_f$ for the intermediate states inserted in Eq. (45). For ε_f well above the Fermi level, the state $|\nu\rangle$ decays into states $|E\nu\rangle \equiv \psi_{E\nu}^\dagger |0\rangle$ with $E \approx \varepsilon_f$, which leads to a peak with a half-width $\pi V(\varepsilon_f)^2 = \pi \tilde{V}(\varepsilon_f)^2/N_f$. The states $|E\nu\rangle$ couple to the states $|E\varepsilon\nu\rangle \equiv \sum_{\nu'} \psi_{E\nu}^\dagger \psi_{\nu'}^\dagger \psi_{\varepsilon\nu'} |0\rangle / \sqrt{N_f}$ with a matrix element $\tilde{V}(\varepsilon)$, i.e., of order $(1/N_f)^0$. In the infinite- U case these are the additional states to be included. The calculation is similar to the leading-order calculation for $G_{\nu\nu}^<$ presented in the previous subsection but using the inversion formula (40) twice. It leads to [9, 10]

$$G_{\nu\nu}^>(z) = \frac{1 - n_f}{z + \Delta E_0 - \varepsilon_f - \mu(z)} \quad (48)$$

with $\mu(z) = \int_0^B \frac{V(E)^2}{z + \Delta E_0 - E + \tilde{V}(-z - \Delta E_0 + E + \varepsilon_f)} dE.$

The additional term $\tilde{I}$ in the denominator of the integrand of $\mu(z)$ results from including the states $|E\varepsilon\nu\rangle$ with a hole in the conduction band. Neglecting $\tilde{I}$ gives the result for the width of the peak near ε_f well above the Fermi energy, mentioned above. For ε_f well below the Fermi energy it is essential to include $\tilde{I}$. Then the integrand on the right-hand side has a pole at $z = E$ of strength $1 - n_f$ leading to $-\text{Im } \mu(\varepsilon + i0) = (1 - n_f)\pi V(\varepsilon)^2$. This leads to

$$\rho_f^>(\varepsilon) = \frac{(1 - n_f)^2 \tilde{V}(\varepsilon)^2}{(\varepsilon - \delta - \text{Re } \mu(\varepsilon))^2 + ((1 - n_f)\pi V(\varepsilon))^2} \quad \text{for } 0 \leq \varepsilon \leq \delta. \quad (49)$$

In a strict $1/N_f$ expansion of $\rho_f^>$, the contributions from μ in the denominator are of order $1/N_f$ and can be neglected. With this assumption $\rho_f^>(\varepsilon)$ joins smoothly to the low-energy result $\rho_f^<(\varepsilon)$ in Eq. (44). The steep rise found there for $-\delta \leq \varepsilon \leq 0$ continues for $\rho_f^>(\varepsilon)$. In the region $\varepsilon \approx \delta$ the strict $1/N_f$ expansion fails, and the full expression in Eq. (49) has to be used. This gives the Kondo peak a half-width $(1 - n_f)\pi V(\delta)^2 \approx \pi n_f \delta / N_f$. The correct treatment of the energy range $\varepsilon \geq \delta$ within the approximation given by Eq. (48) requires the inclusion of the “continuum part” of μ . Unfortunately the approximation Eq. (48) for $G_{\nu\nu}^>$ leads to an additional weak, unphysical pole slightly below $\varepsilon = 0$. A different type of anomaly appears in the NCA at zero temperature [57]. For large N_f , the NCA properly describes how the weight of the Kondo resonance decreases with increasing T , where the scale is given by the Kondo temperature.

One can summarize the behavior of the f^1 peak as a function of ε_f as follows: Lowering ε_f from well above the Fermi energy to well below it, its position goes from ε_f very close to $\varepsilon_F = 0$ Its weight $(1 - n_f)N_f$ and width $(1 - n_f)\pi \tilde{V}(\delta)^2 / N_f$ is reduced as n_f goes from ≈ 0 to ≈ 1 .

For a comparison with experiment, it is crucial to take into account that U is finite, since this leads to a second f^2 -like peak in the BIS spectrum. Using the leading-order, finite- U ground state and additional intermediate states with a doubly occupied f level and a hole in the conduction band, the resulting matrix has to be inverted numerically [10, 11]. The f^2 peak has a broadening of 2Δ (half-width), as the f^2 state can decay by the hopping of either of the two f electrons into the conduction band. It shows a tailing towards higher energies. The reason is that the intermediate states with two electrons in the f level have a hole in the conduction band. This hole is likely to be close to the Fermi energy but can also be located further down.

In the spin-degenerate case $N_f = 2$ a half-filled symmetric band and $2\varepsilon_f + U = 0$ lead to particle-hole symmetry and the Kondo resonance is at $\varepsilon = 0$ as shown in Fig. 1. For $N_f > 2$ the Kondo resonance is above the Fermi level for $2\varepsilon_f + U = 0$ [24]. In order to obtain the Kondo resonance exactly at $\varepsilon = 0$ for $N_f = 2$ an infinite summation of skeleton diagrams in the generating functional is necessary leading to the symmetrized finite- U NCA [25].

Let us summarize the behavior of the total spectral function $\rho_f = \rho_f^< + \rho_f^>$ in the Kondo regime $-\varepsilon_f \gg \tilde{\Delta}$ for large values of U : The ionization peak near ε_f has weight $n_f \approx 1$; the weight of the f^1 peak (Kondo peak) slightly above ε_F is $(1 - n_f)N_f$, and the f^2 peak near $\varepsilon_f + U$ has a weight $n_f(N_f - 1) \approx N_f - 1$. Therefore, even for $n_f = 0.9$ and $N_f = 14$, the weight of the f^1 peak is higher than that of the ionization peak. Despite the fact that there is a small chance, $1 - n_f$, to find the f level empty, there are N_f different ways to place the electron. The weight of the Kondo peak in the BIS spectrum is a factor N_f larger than the part seen in photoemission.

4.3 Spectra involving core holes

As mentioned in section 2.2, core-level XPS and X-ray absorption spectroscopy provide additional information about the valence electrons. The Hamiltonian used to describe core-level spectra of mixed-valence systems takes the form presented in Eq. (17) with $U_{dc} \rightarrow U_{fc}$ and H_A replaced by the valence Hamiltonian H of Eq. (21). With the assumptions explained in section 2.2 the core spectral function is given by

$$\rho_c(\varepsilon) = \left\langle E_0(N) \left| \delta(\varepsilon - \varepsilon_c - E_0(N) + \tilde{H}) \right| E_0(N) \right\rangle, \quad (50)$$

where $\tilde{H}$ is the Hamiltonian of Eq. (21) with ε_f replaced by $\tilde{\varepsilon}_f \equiv \varepsilon_f - U_{fc}$. In the infinite- U case to order $(1/N_f)^0$ the ground state is given by Eq. (29), and the states $|0\rangle$ and $|\varepsilon\rangle$ defined in Eq. (23) are used as intermediate states in the calculation of $G_{cc} = G_{cc}^<$. The matrix elements $\tilde{H}(z)_{ij} \equiv \langle i | (z + \tilde{H} - E_0(N)) | j \rangle$ are easily written down. The $00, 0\varepsilon, \varepsilon 0$ and $\varepsilon\varepsilon'$ matrix elements of the inverse matrix $\tilde{H}(z)^{-1}$ are all needed. The calculation is analogous to obtaining the Green's functions of a noninteracting Anderson model. With $\tilde{z} \equiv z - \varepsilon_c - \Delta E_0$, this yields e.g., $(\tilde{H}(z)^{-1})_{00} = (\tilde{z} - \tilde{I}(\tilde{z} + \tilde{\varepsilon}_f))^{-1}$ with $\tilde{I}$ defined in Eq. (42). After some algebra [9], the result can be brought into the form

$$\rho_c(\varepsilon + \varepsilon_c) = (1 - n_f) \left(\frac{U_{fc}}{\varepsilon - U_{fc}} \right)^2 \tilde{\rho}_f(\varepsilon - \Delta E_0 + \varepsilon_f - U_{fc}), \quad (51)$$

where

$$\tilde{\rho}_f(\varepsilon) = -\frac{1}{\pi} \text{Im} \frac{1}{\varepsilon + i0 - \varepsilon_f + U_{fc} - \tilde{I}(\varepsilon + i0)}. \quad (52)$$

The same type of expression is obtained in the exact solution of the $N_f = 1$ *filled band model* [58]. In this leading order in $1/N_f$ approximation, the core spectrum is directly related to the valence spectrum $\tilde{\rho}_f$. This clearly shows that core-level spectroscopy gives information about properties of the valence electrons, like n_f, ε_f , and Δ . The multiplying factor $[U_{fc}/(\varepsilon - U_{fc})]^2$ changes the weights in $\tilde{\rho}_f$ but does not normally introduce new structures.

To test the accuracy of the $1/N_f$ method one can study the limit $N_f = 1$, where the exact solution can be obtained by solving the Nozières-de Dominicis integral equation [46] numerically. A comparison of the $1/N_f$ result including the states $0, a, c$, and d in Fig. 2 is shown in Fig. 4. As mentioned in section 2.2 the exact result has an infrared singularity at threshold which is not present in the $1/N_f$ result. To mimic lifetime broadening, the spectra shown were given a Lorentzian broadening of 2Γ (full-width half-max). Despite the fact that the $1/N_f$ calculation includes at most two holes, the asymmetry of the exact solution, which includes an infinite number of electron-hole pairs, is quite well described except very close to threshold.

X-ray absorption spectroscopy of $3d \rightarrow 4f$ transitions has formal similarities with inverse photoemission, as an electron is added to the f -level. The difference is that the final state has the core hole present. The theoretical description therefore, as in core-hole XPS, has to use the Hamiltonian $\tilde{H}$ with $\varepsilon_f \rightarrow \varepsilon_f - U_{fc}$. If one works to lowest order in $1/N_f$, it is possible to obtain an analytical solution even if f^2 configurations are included [9].

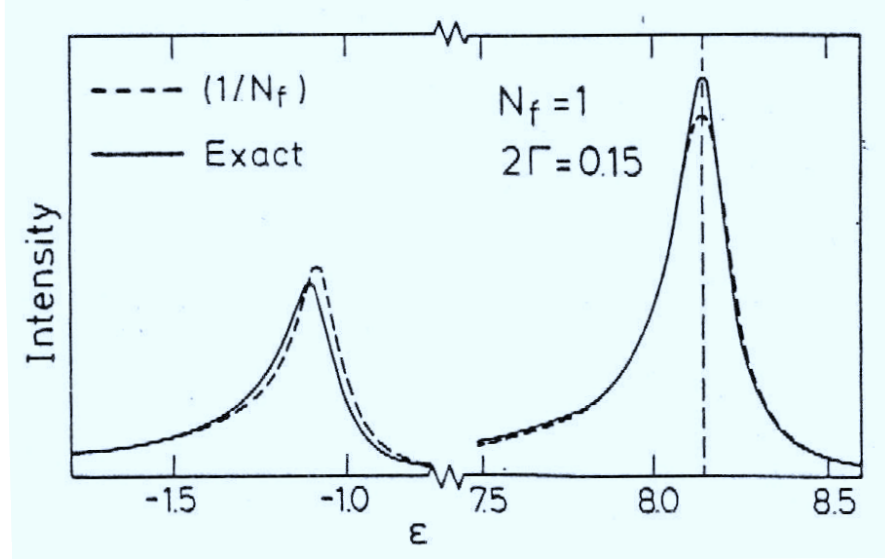


Fig. 4: The core-level spectrum for $N_f = 1$, $\varepsilon_f = 0$, $\Delta = 1.5$, and $U_{fc} = 9$ and a semi-elliptical form of $V(\varepsilon)^2$ with $B = 3$. The spectra are shown with a Lorentzian broadening (FWHM) of 0.15: exact result (full line), using the states 0, a, c, and d of Fig. 2 as intermediate states (dashed line)

5 Comparison with experiment

Model calculations for spectra using the impurity model are frequently used for a comparison with experimental data of lanthanide materials. An example for systems with essentially zero f occupancy in the ground state are La compounds. In Ce systems f^0 and f^1 play the important role. Even in dense systems spectra calculated by the intermediate states method for the Anderson impurity model are often in good agreement with experiment [9, 10, 12, 59]. Switching from the number of electrons in the f level to the number of holes, the formalism presented is easily extended to also describe Yb compounds, as there f^{13} and f^{11} play the same role as the f^1 and f^0 configurations in Ce compounds [54]. To study, e.g., Pr or Nd compounds, the model used here should be generalized.

As the *ab-initio* determination of parameters of model Hamiltonians is a problematic issue, one alternatively adjusts them to experimental data. If, e.g., different spectroscopies are used, a part of the data may suffice to obtain the parameters by fitting to peak positions and their widths. Then additional data can be used as a consistency check. If this turns out to be satisfying for a class of materials, the use of the model in conjunction with a first set of data has predictive power for further measurements. This is what in fact happened with spectra of mixed-valence systems [1]. As an example let us take core-level spectra. The leading peak in Fig. 4 corresponds to final states of mainly f^1 character, while the satellite corresponds to f^0 -like states. These are the important final states for La compounds. For Ce compounds it is important to also take f^2 configurations into account. The core spectrum often has three peaks, corresponding to f^0 , f^1 and f^2 states. The parameters are usually such that the high-energy f^2 peak (shoulder) has a small but observable weight that strongly depends on the value of the coupling parameter Δ [10] and is therefore suitable for its determination.

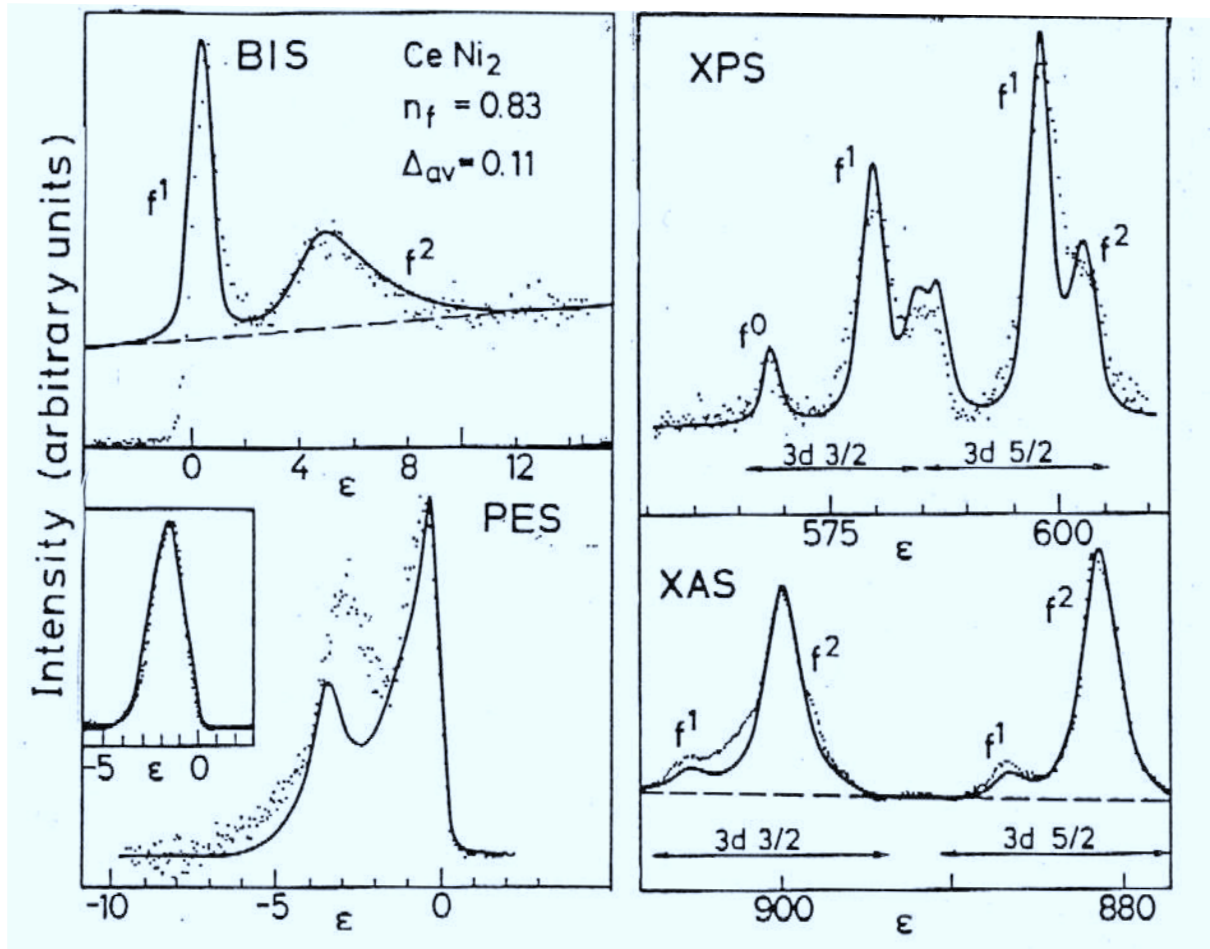


Fig. 5: Comparison of experimental spectra (dots) for CeNi_2 with theoretical results using the impurity model. The results for inverse photoemission (BIS), valence photoemission (PES), 3d X-ray photoemission (XPS), and 3d $\rightarrow$ 4f X-ray absorption (XAS) are discussed in the text.

Fig. 5 shows as an example experimental spectra for CeNi_2 and the attempts to fit them with a single set of parameters using the Anderson impurity model and the methods discussed in section 4. As the energy dependence of $V(\varepsilon)^2$ has to be taken into account [12], the average value Δ_{av} of $\pi V(\varepsilon)^2$, extracted from 3d $\rightarrow$ 4f X-ray absorption spectra, $\Delta_{\text{av}} = 0.11$ eV is presented. The total f occupancy is inferred to be $n_f = 0.83$. The two components of the XAS spectrum in the figure are due to transitions from the spin-orbit-split 3d 3/2 and 3d 5/2 levels. Each component shows an f^1 and f^2 peak with multiplet structure. The relative weights of the 3d 3/2 and 3d 5/2 components were adjusted to the experiment and a weak background was added as shown. A Lorentzian broadening (FWHM=2.0 eV) was used to describe life-time broadening and instrumental resolution. The parameters obtained from the 3d XPS spectrum differ only slightly from the ones from the XAS data. The f^2 and f^1 -Kondo peak of the shown BIS spectrum were obtained using these parameters. Using the same parameters, the valence-band PES spectrum shows the onset of the Kondo peak at energies close to zero. The peak at $\varepsilon \approx -3$ eV is somewhat too low in energy and too narrow ($\varepsilon_f = -1.6$ eV was used). Introducing different features in $V(\varepsilon)^2$ allows one to improve the agreement with experiment [12]. More information

about the Coulomb parameters used can be found in Ref. [12]. It should be mentioned that for CeNi_2 and other Ce compounds the calculation of the static magnetic susceptibility with the parameters from spectroscopy leads to results in good agreement with the measured values.

As mentioned, there were many theoretical developments in the last thirty years that go beyond the methods presented in sections 3 and 4. The intermediate states method nevertheless has remained a valuable tool in the hands of experimentalists.

Acknowledgements

The author would like to thank Olle Gunnarsson for the critical reading the manuscript and useful suggestions. He originally had come up with the idea of the type of $1/N_f$ expansion presented here, which we then worked on together.

References

- [1] For an earlier review see the articles in: *Handbook on the Physics and Chemistry of Rare Earths*, Vol. 10, *High Energy Spectroscopy*, K.A. Gschneider, L. Eyring, and S. Hufner (eds.), (North Holland, 1987)
- [2] P.W. Anderson, Phys. Rev. **124**, 41 (1961)
- [3] A.A. Abrikosov, Physics **2**, 5 (1965)
- [4] H. Suhl, Phys. Rev. **138**, A515 (1965)
- [5] K. Yamada, Prog. Theor. Phys. **53**, 970 (1975)
- [6] T.V. Ramakrishnan, *Perturbative theory of mixed valence systems*, in: *Valence Fluctuations in Solids*, L.M. Falicov, W. Hanke and M.B. Maple (eds.), (North-Holland, 1982);
T.V. Ramakrishnan and K. Sur, Phys. Rev. B **26**, 1798 (1982)
- [7] P.W. Anderson, Summary talk, in *Valence Fluctuations in Solids*, L.M. Falicov, W. Hanke and M.B. Maple (eds.), (North-Holland, 1982)
- [8] O. Gunnarsson and K. Schönhammer, Phys. Rev. Lett. **50**, 604 (1983)
- [9] O. Gunnarsson and K. Schönhammer, Phys. Rev. B **28**, 4315 (1983)
- [10] O. Gunnarsson and K. Schönhammer,
Many-Body Formulation of Spectra of Mixed Valence Systems, in [1]
- [11] O. Gunnarsson and K. Schönhammer, Phys. Rev. B **31** 4815, (1985)
- [12] J. Allen *et al.*, Advances in Physics **35**, 275 (1986)
- [13] Y. Kuramoto, Z. Phys. B **53**, 37 (1983)
- [14] H. Keiter and G. Czychohl, J. Magn. and Magn. Mater. **31**, 477 (1983)
- [15] N. Grewe, Z. Phys. B **52**, 193 (1983)
- [16] P. Coleman, Phys. Rev. B **29**, 3035 (1984)
- [17] H. Keiter and J.C. Kimball, Int. J. Magn. **1**, 233 (1971)
- [18] S.E. Barnes, J. Phys. F **6**, 1375 (1976)
- [19] K. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
- [20] H.O. Frota and L.N. Olivera, Phys. Rev. B **46**, 15207 (1986)
- [21] for a review see: R. Bulla, T.A. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

- [22] J. Holm, Solid State Commun. **69**, 969 (1989)
- [23] T. Pruschke and T. Grewe, Z. Phys. **74**, 439 (1989)
- [24] J. Holm, R. Kree and K. Schönhammer, Phys. Rev. B **48**, 5077 (1993)
- [25] H. Haule, S. Kirchner, J. Kroha, and P. Wölfle, Phys. Rev. B **64**, 155111 (2001)
- [26] S. Kirchner, J. Kroha, and P. Wölfle, Phys. Rev. B **70**, 155301 (2004)
- [27] D.E. Logan, M.P. Eastwood, and M.A. Tusch, J. Phys. Condens. Matter **10**, 2673 (1998)
- [28] M.R. Galpin, A.B. Gilbert, and D.E. Logan, J. Phys. Condens. Matter **21**, 375602 (2009)
- [29] P.W. Wiegmann, Phys. Lett. A **80**, 163 (1980)
- [30] N. Kawakami and A. Okiji, Phys. Lett. A **86**, 483 (1981)
- [31] P. Schlottmann, Phys. Rev. Lett. **50**, 1697 (1983)
- [32] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [33] D. Newns, Phys. Rev. **178**, 1123 (1969)
- [34] D.N. Zubarev, Sov. Phys. Usp. **3**, 320 (1960)
- [35] F. Reinert et al., Phys. Rev. Lett. **87**, 106401 (2001)
- [36] K.O. Friedrichs, Comm. Pure and Appl. Math. **1**, 361 (1948)
- [37] T.D. Lee, Phys. Rev. **95**, 1329 (1954)
- [38] K. Yoshida and K. Yamada, Prog. Theor. Phys. **53**, 1286 (1975)
- [39] W. Brenig and K. Schönhammer, Z. Phys. B **267**, 201 (1974)
- [40] A.C. Hewson, Phys. Rev. **144**, 120 (1966)
- [41] A. Theumann, Phys. Rev. **178**, 120 (1969)
- [42] C. Lacroix, J. Phys. F **11**, 2389 (1981)
- [43] S.-J. Oh and S. Doniach, Phys. Rev. B **26**, 2085 (1982)
- [44] G. Czycholl, Phys. Rev. B **31**, 2867 (1985)
- [45] A. Hewson, *The Kondo Problem to Heavy Fermions*, (Cambridge University Press, 1993)
- [46] P. Nozières and C.T. de Dominicis, Phys. Rev. **178**, 1097 (1969)
- [47] P.W. Anderson, Phys. Rev. Lett. **18**, 1049 (1967)

- [48] M. Combescot and P. Nozières, *J. de Physique* **32**, 913 (1971)
- [49] A. Kotani and Y. Toyozawa, *J. Phys. Soc. Japan* **35**, 563 (1974); **35**, 912 (1974)
- [50] K. Schönhammer and O. Gunnarsson, *Solid State Commun.* **23**, 691 (1977)
- [51] K. Schönhammer and O. Gunnarsson, *Solid State Commun.* **26**, 399 (1978)
- [52] K. Schönhammer and O. Gunnarsson, *Phys. Rev. B* **18**, 6606 (1978)
- [53] C.M. Varma and Y. Yafet, *Phys. Rev. B* **13**, 2950 (1976)
- [54] N.E. Bickers, D.L. Cox, and J.W Wilkins, *Phys. Rev. Lett.* **54**, 230 (1985)
- [55] F. Patthey, B. Delley, W.-D. Schneider, and Y. Baer, *Phys. Rev. Lett.* **55**, 1518 (1985)
- [56] L. Hedin and S. Lundqvist, in *Solid State Physics*, Vol. 23,
H. Ehrenreich, D. Turnbull, and F. Seitz (eds.), (Academic Press, New York, 1969)
- [57] E. Müller-Hartmann, *Z. Phys.* **57**, 281 (1984)
- [58] K. Schönhammer and O. Gunnarsson, *Z. Phys. B* **30**, 297 (1978)
- [59] Y. Baer and W.-D. Schneider,
High-Energy Spectroscopy of Lanthanide Materials – an Overview, in [1]

3 Magnetism in Correlated Matter

Eva Pavarini

Institute for Advanced Simulation

Forschungszentrum Jülich

Contents

1 Overview	2
2 The Hubbard model	8
2.1 Weak-correlation limit	12
2.2 Atomic limit	20
2.3 Strong-correlation limit	24
3 The Anderson model	36
3.1 The Kondo limit	36
3.2 The RKKY interaction	41
4 Conclusion	42
A Formalism	44
A.1 Matsubara Green functions	44
A.2 Linear-response theory	47
A.3 Magnetic susceptibility	49

1 Overview

Around the beginning of last century magnetic phenomena in materials were at the center of a hot scientific debate: What causes ferromagnetic order? At the time, atoms were not fully understood, and there were perhaps more questions than answers. Weiss proposed the molecular mean-field theory of ferromagnetism [1], which dominated the scene. Friedrich Hund formulated his now-famous rules [2] to determine the atomic ground-state multiplets, which turned out to be basically exact. Heisenberg [3] realized that Coulomb exchange leads to ferromagnetic coupling between local magnetic moments. New puzzles emerged. Can other types of long-range order occur in Nature? In 1949 the first observation of antiferromagnetic order was reported, causing a great sensation [4]. Such a state had been predicted by Néel [5] about 20 years before via an extended version of Weiss' mean-field theory. In the mean-time, however, it was clear that the theory had a problem. Indeed, antiferromagnetism was an artifact of the static mean-field approximation. Bethe [6] had found the general solution of the one-dimensional Heisenberg spin chain, which shows that, in the case of antiferromagnetic coupling between the spins, the ground state has a total spin zero, and thus it is not the antiferromagnetic state. Several years later, Anderson understood [7] that the original $SU(2)$ symmetry of the Hamiltonian in spin space, broken in the antiferromagnetic state, is recovered once quantum fluctuations are taken into account; this led to broken-symmetry theory and ultimately to the postulation of the famous Anderson-Higgs boson [8, 9]. While all this was happening, other effects that involved local magnetic moments were discovered. A low-temperature minimum in the resistance of some metals puzzled scientists for long, until in 1964 Kondo understood [10] that it is caused by local spins (magnetic impurities) coupled antiferromagnetically to conduction-electron spins. The theoretical efforts to understand the Kondo effect, described via the Kondo model or the more general Anderson model, fueled the development of new powerful non-perturbative many-body techniques, among which the numerical renormalization group. Experimentally, many f -electron compounds were identified as lattices of Kondo impurities. In 1975, the discovery of a huge electronic specific heat in CeAl_3 below 0.2 K [11] brought a new category of strongly correlated materials to light, the heavy fermions [12]. These are dense Kondo systems with low-temperature Fermi liquid properties but extremely large quasielectron masses. To complicate the scenario, in such materials the Kondo effect competes with long-range magnetic order, and both phenomena are mediated by the same interaction. Furthermore, several members of the heavy-fermion family, such as CeCu_2Si_2 , become unconventional superconductors at low temperature. Later on in 1986, a novel class of unconventional superconductors, high-temperature superconducting cuprates, was found [13]; these systems are believed to be doped Mott insulators, described, at least to first approximation, by the two-dimensional single-band Hubbard model. Around the same time, with the development of the dynamical mean-field theory (DMFT) [14], it emerged that the Anderson model and the Kondo effect are intimately connected with the Mott metal-insulator transition [15, 16] and thus with the Hubbard model. In these lecture notes we will discuss some of the models and methods that made this exciting piece of the history of modern physics [12, 15–22].

Magnetism ultimately arises from the intrinsic magnetic moment of electrons, $\boldsymbol{\mu} = -g\mu_B \mathbf{s}$, where μ_B is the Bohr magneton and $g \simeq 2.0023$ is the electronic g -factor. It is, however, an inherently quantum mechanical effect, the consequence of the interplay between the Pauli exclusion principle, the Coulomb electron-electron interaction, and the hopping of electrons. To understand this let us consider the simplest possible system, an isolated atom or ion. In the non-relativistic limit electrons in a single ion are typically described by the Hamiltonian

$$H_e^{\text{NR}} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \frac{Z}{r_i} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|},$$

where Z is the atomic number and $\{\mathbf{r}_i\}$ are the coordinates of the electrons with respect to the ionic nucleus. Here, as in the rest of this lecture, we use atomic units. If we consider only the external atomic shell with quantum numbers nl , for example the $3d$ shell of transition-metal ions, we can rewrite this Hamiltonian as follows

$$H_e^{\text{NR}} = \varepsilon_{nl} \sum_{m\sigma} c_{m\sigma}^\dagger c_{m\sigma} + \frac{1}{2} \sum_{\sigma\sigma'} \sum_{m\tilde{m}m'\tilde{m}'} U_{mm'\tilde{m}\tilde{m}'}^l c_{m\sigma}^\dagger c_{m'\sigma'}^\dagger c_{\tilde{m}'\sigma'} c_{\tilde{m}\sigma}. \quad (1)$$

The parameter ε_{nl} is the energy of the electrons in the nl atomic shell and m the degenerate one-electron states in that shell. For a hydrogen-like atom

$$\varepsilon_{nl} = -\frac{1}{2} \frac{Z^2}{n^2}.$$

The couplings $U_{mm'\tilde{m}\tilde{m}'}^l$ are the four-index Coulomb integrals. In a basis of atomic functions the bare Coulomb integrals are

$$U_{mm'\tilde{m}\tilde{m}'}^{ijj'} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}(\mathbf{r}_1)} \overline{\psi_{jm'\sigma'}(\mathbf{r}_2)} \psi_{j'\tilde{m}'\sigma'}(\mathbf{r}_2) \psi_{i'\tilde{m}\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

and $U_{mm'\tilde{m}\tilde{m}'}^l = U_{mm'\tilde{m}\tilde{m}'}^{iiii}$, where $m, m', \tilde{m}, \tilde{m}' \in nl$ shell. The eigenstates of Hamiltonian (1) for fixed number of electrons N are the multiplets [23,24]. Since in H_e^{NR} the Coulomb repulsion and the central potential are the only interactions, the multiplets can be labeled with S and L , the quantum numbers of the electronic total spin and total orbital angular momentum operators, $\mathbf{S} = \sum_i \mathbf{s}_i$, and $\mathbf{L} = \sum_i \mathbf{l}_i$. Closed-shell ions have $S = L = 0$ in their ground state. Ions with a partially-filled shell are called *magnetic ions*; the value of S and L for their ground state can be obtained via Hund's rules. The first and second Hund's rules say that the lowest-energy multiplet is the one with

1. the largest value of S
2. the largest value of L compatible with the previous rule

The main relativistic effect is the spin-orbit interaction, which has the form $H_e^{\text{SO}} = \sum_i \lambda_i \mathbf{l}_i \cdot \mathbf{s}_i$. For not-too-heavy atoms it is a weak perturbation. For electrons in a given shell, we can rewrite H_e^{SO} in a simpler manner using the first and second Hund's rule. If the shell filling is $n < 1/2$,

the ground-state multiplet has spin $S = (2l + 1)n = N/2$; thus $\mathbf{s}_i = \mathbf{S}/N = \mathbf{S}/2S$. If, instead, $n > 1/2$, since the sum of $\mathbf{l}_i$ vanishes for the electrons with spin parallel to $\mathbf{S}$, only the electrons with spin antiparallel to $\mathbf{S}$ contribute. Their spin is $\mathbf{s}_i = -\mathbf{S}/N_u = -\mathbf{S}/2S$ where $N_u = 2(2l + 1)(1 - n)$ is the number of unpaired electrons. We therefore obtain the LS Hamiltonian

$$H_e^{\text{SO}} \sim \underbrace{\left[2\Theta(1 - 2n) - 1 \right] \frac{g\mu_B^2}{2S} \left\langle \frac{1}{r} \frac{d}{dr} v_R(r) \right\rangle}_{\lambda} \mathbf{L} \cdot \mathbf{S} = \lambda \mathbf{L} \cdot \mathbf{S}, \quad (2)$$

where Θ is the Heaviside step function and $v_R(r)$ is the effective potential, which includes, e.g., the Hartree electron-electron term [25]. For a hydrogen-like atom, $v_R(r) = -Z/r$. Because of the LS coupling (2) the eigenstates have quantum numbers L, S and J , where $\mathbf{J} = \mathbf{S} + \mathbf{L}$ is the total angular momentum. Since $\mathbf{L} \cdot \mathbf{S} = [\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2]/2$, the value of J in the ground-state multiplet is thus (third Hund's rule)

$$\bullet \text{ total angular momentum } J = \begin{cases} |L - S| & \text{for filling } n < 1/2 \\ S & \text{for filling } n = 1/2 \\ L + S & \text{for filling } n > 1/2 \end{cases}$$

In the presence of spin-orbit interaction a given multiplet is then labeled by $^{2S+1}L_J$, and its states can be indicated as $|JJ_zLS\rangle$. If we consider, e.g., the case of the Cu^{2+} ion, characterized by the $[\text{Ar}] 3d^9$ electronic configuration, Hund's rules tell us that the $3d$ ground-state multiplet has quantum numbers $S = 1/2$, $L = 2$ and $J = 5/2$. A Mn^{3+} ion, which is in the $[\text{Ar}] 3d^4$ electronic configuration, has instead a ground-state multiplet with quantum numbers $S = 2$, $L = 2$ and $J = 0$. The order of the Hund's rules reflects the hierarchy of the interactions. The strongest interactions are the potential $v_R(r)$, which determines ε_{nl} , and the average Coulomb interaction, the strength of which is measured by the average *direct Coulomb integral*,

$$U_{\text{avg}} = \frac{1}{(2l + 1)^2} \sum_{mm'} U_{mm'mm'}^l.$$

For an N -electron state the energy associated with these two interactions is $E(N) = \varepsilon_{nl}N + U_{\text{avg}}N(N - 1)/2$, the same for all multiplets of a given shell. The first Hund's rule is instead due to the average *exchange Coulomb integral*, J_{avg} , defined as

$$U_{\text{avg}} - J_{\text{avg}} = \frac{1}{2l(2l + 1)} \sum_{mm'} (U_{mm'mm'}^l - U_{mm'm'm}^l),$$

which is the second-largest Coulomb term; for transition-metal ions $J_{\text{avg}} \sim 1$ eV. Smaller Coulomb integrals determine the orbital anisotropy of the Coulomb matrix and the second Hund's rule.¹ The third Hund's rule comes, as we have seen, from the spin-orbit interaction which, for not-too-heavy atoms, is significantly weaker than all the rest.

¹For more details on Coulomb integrals and their averages see Ref. [25].

The role of Coulomb electron-electron interaction in determining S and L can be understood through the simple example of a C atom, electronic configuration [He] $2s^2 2p^2$. We consider only the p shell, filled by two electrons. The Coulomb exchange integrals have the form

$$\begin{aligned} J_{m,m'}^p &= U_{mm'm'm}^p = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}(\mathbf{r}_1)} \overline{\psi_{im'\sigma}(\mathbf{r}_2)} \psi_{im\sigma}(\mathbf{r}_2) \psi_{im'\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} \\ &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\phi_{imm'\sigma}(\mathbf{r}_1) \overline{\phi_{imm'\sigma}(\mathbf{r}_2)}}{|\mathbf{r}_1 - \mathbf{r}_2|}. \end{aligned} \quad (3)$$

If we express the Coulomb potential as

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{V} \sum_{\mathbf{k}} \frac{4\pi}{k^2} e^{i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)},$$

we can rewrite the Coulomb exchange integrals in a form that shows immediately that they are always positive

$$J_{m,m'}^p = \frac{1}{V} \sum_{\mathbf{k}} \frac{4\pi}{k^2} |\phi_{imm'\sigma}(\mathbf{k})|^2 > 0.$$

They generate the Coulomb-interaction term

$$-\frac{1}{2} \sum_{\sigma} \sum_{m \neq m'} J_{m,m'}^p c_{m\sigma}^\dagger c_{m\sigma} c_{m'\sigma}^\dagger c_{m'\sigma} = -\frac{1}{2} \sum_{m \neq m'} 2J_{m,m'}^p \left[S_z^m S_z^{m'} + \frac{1}{4} n_m n_{m'} \right].$$

This exchange interaction yields an *energy gain* if the two electrons occupy two different p orbitals with parallel spins, hence it favors the state with the largest spin (first Hund's rule). It turns out that for the p^2 configuration there is only one possible multiplet with $S = 1$, and such a state has $L = 1$. There are instead two excited $S = 0$ multiplets, one with $L = 0$ and the other with $L = 2$; the latter is the one with lower energy (second Hund's rule).

To understand the magnetic properties of an isolated ion we have to analyze how its levels are modified by an external magnetic field $\mathbf{h}$. The effect of a magnetic field is described by

$$H_e^H = \mu_B (g\mathbf{S} + \mathbf{L}) \cdot \mathbf{h} + \frac{h^2}{8} \sum_i (x_i^2 + y_i^2) = H_e^Z + H_e^L. \quad (4)$$

The linear term is the Zeeman Hamiltonian. If the ground-state multiplet is characterized by $J \neq 0$ the Zeeman interaction splits its $2J + 1$ degenerate levels. The second-order term yields Larmor diamagnetism, which is usually only important if the ground-state multiplet has $J = 0$, as happens for ions with closed external shells. The energy $\mu_B h$ is typically very small (for a field as large as 100 T it is as small as 6 meV); it can however be comparable with or larger than the spin-orbit interaction if the latter is tiny (very light atoms). Taking all interactions into account, the total Hamiltonian is

$$H_e \sim H_e^{\text{NR}} + H_e^{\text{SO}} + H_e^H.$$

In a crystal the electronic Hamiltonian is complicated by the interaction with other nuclei and their electrons. The non-relativistic part of the Hamiltonian then takes the form

$$H_e^{\text{NR}} = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq i'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} - \sum_{i\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \frac{1}{2} \sum_{\alpha \neq \alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|},$$

where Z_α is the atomic number of the nucleus located at position $\mathbf{R}_\alpha$. In a basis of localized Wannier functions [25] this Hamiltonian can be written as

$$H_e^{\text{NR}} = - \sum_{ii'\sigma} \sum_{mm'} t_{m,m'}^{i,i'} c_{im\sigma}^\dagger c_{i'm'\sigma} + \frac{1}{2} \sum_{ii'jj'} \sum_{\sigma\sigma'} \sum_{mm'} \sum_{\tilde{m}\tilde{m}'} U_{mm'\tilde{m}\tilde{m}'}^{ijj'j'} c_{im\sigma}^\dagger c_{jm'\sigma'}^\dagger c_{j'\tilde{m}'\sigma'} c_{i'\tilde{m}\sigma}, \quad (5)$$

where

$$t_{m,m'}^{i,i'} = - \int d\mathbf{r} \overline{\psi_{im\sigma}(\mathbf{r})} \left[-\frac{1}{2} \nabla^2 + v_R(\mathbf{r}) \right] \psi_{i'm'\sigma}(\mathbf{r}).$$

The terms $\varepsilon_{m,m'} = -t_{m,m'}^{i,i'}$ yield the crystal-field matrix and $t_{m,m'}^{i,i'}$ with $i \neq i'$ the hopping integrals. The label m indicates here the orbital quantum number of the Wannier function. In general, the Hamiltonian (5) will include states stemming from more than a single atomic shell. For example, in the case of strongly correlated transition-metal oxides, the set $\{im\}$ includes transition-metal $3d$ and oxygen $2p$ states. The exact solution of the many-body problem described by (5) is an impossible challenge. The reason is that the properties of a many-body system are inherently emergent and hence hard to predict *ab-initio* in the absence of any understanding of the mechanism behind them. In this lecture, however, we want to focus on magnetism. Since the nature of cooperative magnetic phenomena in crystals is currently to a large extent understood, we can find realistic approximations to (5) and even map it onto simpler models that still retain the essential ingredients to explain long-range magnetic order.

Let us identify the parameters of the electronic Hamiltonian important for magnetism. The first is the crystal-field matrix $\varepsilon_{m,m'}$. The crystal field at a given site i is a non-spherical potential due to the joint effect of the electric field generated by the surrounding ions and of covalent-bond formation [24]. The crystal field can split the levels within a given shell and therefore has a strong impact on magnetic properties. We can identify three ideal regimes. In the *strong-crystal-field* limit the crystal-field splitting is so large that it is comparable with the average Coulomb exchange responsible for the first Hund's rule. This can happen in $4d$ or $5d$ transition-metal oxides. A consequence of an *intermediate crystal field* (weaker than the average Coulomb exchange but larger than Coulomb anisotropy and spin-orbit interaction) is the quenching of the angular momentum, $\langle \mathbf{L} \rangle = 0$. In this limit the second and third Hund's rule are not respected. This typically happens in $3d$ transition-metal oxides. In $4f$ systems the crystal-field splitting is usually much weaker than the spin-orbit coupling (*weak-crystal-field* limit) and mainly splits states within a given multiplet, leaving a reduced magnetic moment. In all three cases, because of the crystal field, a magnetic ion in a crystal might lose – totally or partially – its spin, its angular momentum, or its total momentum. Or, sometimes, it is the other way around. This happens for Mn^{3+} ions, which should have a $J = 0$ ground state according to the third Hund's rule. In the perovskite LaMnO_3 , however, they behave as $S = 2$ ions because of the quenching of the angular momentum.

Even if the crystal field does not suppress the magnetic moment of the ion, the electrons might delocalize to form broad bands, completely losing their original atomic character. This happens, e.g., if the hopping integrals $t_{m,m'}^{i,i'}$ are much larger than the average on-site Coulomb interaction U_{avg} . Surprisingly, magnetic instabilities arise even in the absence of localized moments. This *itinerant magnetism* is mostly due to band effects, i.e., it is associated with a large one-electron linear static response-function $\chi_0(\mathbf{q}; 0)$. In this limit correlation effects are typically weak. To study them we can exploit the power of the *standard model* of solid-state physics, the density-functional theory (DFT), taking into account Coulomb interaction effects beyond the local-density approximation (LDA) at the perturbative level, e.g., in the random-phase approximation (RPA). With this approach we can understand and describe Stoner instabilities.

In the opposite limit, the *local-moments* regime, the hopping integrals are small with respect to U_{avg} . This is the regime of strong electron-electron correlations, where complex many-body effects, e.g., those leading to the Mott metal-insulator transition, play an important role. At *low enough energy*, however, only spins and spin-spin interactions matter. Ultimately, at integer filling we can integrate out (*downfold*) charge fluctuations and describe the system via effective spin Hamiltonians. The latter typically take the form

$$H_S = \frac{1}{2} \sum_{ii'} \Gamma^{i,i'} \mathbf{S}_i \cdot \mathbf{S}_{i'} + \dots = H_S^H + \dots \quad (6)$$

The term H_S^H given explicitly in (6) is the Heisenberg Hamiltonian, and $\Gamma^{i,i'}$ is the Heisenberg exchange coupling, which can be antiferromagnetic ($\Gamma^{i,i'} > 0$) or ferromagnetic ($\Gamma^{i,i'} < 0$). The Hamiltonian (6) can, for a specific system, be quite complicated, and might include long-range exchange interactions or anisotropic terms. Nevertheless, it represents a huge simplification compared to the unsolvable many-body problem described by (5), since, at least within very good approximation schemes, it can be solved. Spin Hamiltonians of type (6) are the minimal models that still provide a realistic picture of long-range magnetic order in strongly correlated insulators. There are various sources of exchange couplings. Electron-electron repulsion itself yields, via Coulomb exchange, a ferromagnetic Heisenberg interaction, the *Coulomb exchange interaction*. The origin of such interaction can be understood via a simple model with a single orbital, m . The inter-site Coulomb exchange coupling has then the form

$$J^{i,i'} = U_{mmmm}^{ii'ii} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}(\mathbf{r}_1)} \overline{\psi_{i'm\sigma}(\mathbf{r}_2)} \psi_{im\sigma}(\mathbf{r}_2) \psi_{i'm\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

and it is therefore positive, as one can show by following the same steps that we used in Eq. (3) for $J_{m,m'}^P$. Hence, the corresponding Coulomb interaction yields a ferromagnetic Heisenberg-like Hamiltonian with

$$\Gamma^{i,i'} = -2J^{i,i'} < 0.$$

A different source of magnetic interactions are the *kinetic exchange* mechanisms (direct exchange, super-exchange, double exchange, Ruderman-Kittel-Kasuya-Yosida interaction ...), which are mediated by the hopping integrals. Kinetic exchange couplings are typically (with

a few well understood exceptions) antiferromagnetic [26]. A representative example of kinetic exchange will be discussed in the next section.

While the itinerant and local-moment regime are very interesting ideal limiting cases, correlated materials elude rigid classifications. The same system can present features associated with both regimes, although at different temperatures and/or energy scales. This happens in Kondo systems, heavy fermions, metallic strongly correlated materials, and doped Mott insulators.

In this lecture we will discuss in representative cases the itinerant and localized-moment regime and their crossover, as well as the most common mechanisms leading to magnetic cooperative phenomena. Since our target is to understand strongly correlated materials, we adopt the formalism typically used for these systems. A concise introduction to Matsubara Green functions, correlation functions, susceptibilities, and linear-response theory can be found in the Appendix.

2 The Hubbard model

The simplest model that we can consider is the one-band Hubbard model

$$H = \underbrace{\varepsilon_d \sum_i \sum_{\sigma} c_{i\sigma}^{\dagger} c_{i\sigma}}_{H_d} - t \underbrace{\sum_{\langle ii' \rangle} \sum_{\sigma} c_{i\sigma}^{\dagger} c_{i'\sigma}}_{H_T} + U \underbrace{\sum_i n_{i\uparrow} n_{i\downarrow}}_{H_U} = H_d + H_T + H_U, \quad (7)$$

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; $c_{i\sigma}^{\dagger}$ creates an electron in a Wannier state with spin σ centered at site i , and $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$. The Hubbard model is a simplified version of Hamiltonian (5) with $m = m' = \tilde{m} = \tilde{m}' = 1$ and

$$\begin{cases} \varepsilon_d &= -t_{1,1}^{i,i} \\ t &= t_{1,1}^{\langle i,i' \rangle} \\ U &= U_{1111}^{iiii} \end{cases}.$$

In the $U = 0$ limit the Hubbard model describes a system of independent electrons. The Hamiltonian is then diagonal in the Bloch basis

$$H_d + H_T = \sum_{\mathbf{k}\sigma} \left[\varepsilon_d + \varepsilon_{\mathbf{k}} \right] c_{\mathbf{k}\sigma}^{\dagger} c_{\mathbf{k}\sigma}. \quad (8)$$

The energy dispersion $\varepsilon_{\mathbf{k}}$ depends on the geometry and dimensionality d of the lattice. For a hypercubic lattice of dimension d

$$\varepsilon_{\mathbf{k}} = -2t \sum_{\nu=1}^d \cos(k_{r_{\nu}} a),$$

where a is the lattice constant, and $r_1 = x, r_2 = y, r_3 = z$. The energy $\varepsilon_{\mathbf{k}}$ does not depend on the spin. In Fig. 1 we show $\varepsilon_{\mathbf{k}}$ in the one-, two- and three-dimensional cases.

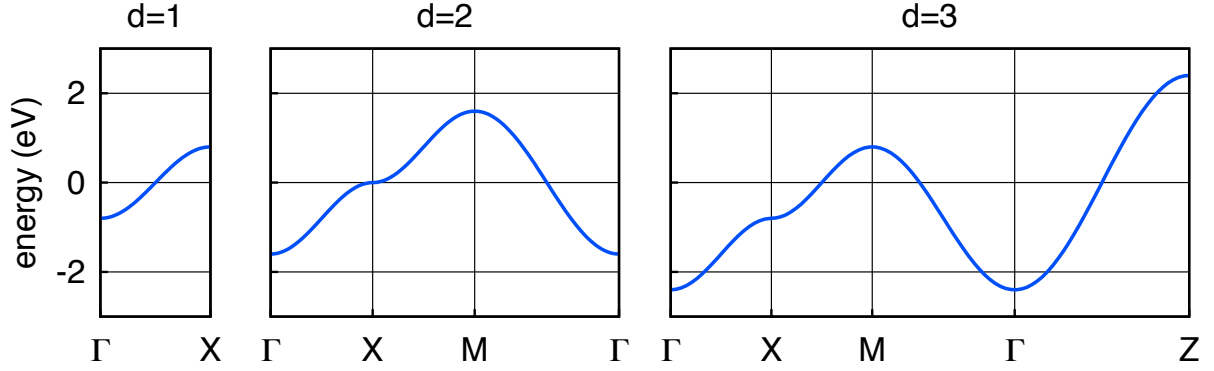


Fig. 1: The band structure of the one-band tight-binding model (hypercubic lattice). The hopping integral is $t = 0.4$ eV. From left to right: one-, two-, and three-dimensional case. At half filling ($n = 1$) the Fermi level is at zero energy.

In the opposite limit ($t = 0$) the Hubbard model describes a collection of isolated atoms. Each atom has four electronic many-body states

$ N, S, S_z\rangle$	N	S	$E(N)$	
$ 0, 0, 0\rangle = 0\rangle$	0	0	0	
$ 1, \frac{1}{2}, \uparrow\rangle = c_{i\uparrow}^\dagger 0\rangle$	1	1/2	ε_d	(9)
$ 1, \frac{1}{2}, \downarrow\rangle = c_{i\downarrow}^\dagger 0\rangle$	1	1/2	ε_d	
$ 2, 0, 0\rangle = c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger 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 $H_d + H_U$ in a form in which the dependence on N_i , S_i , and S_z^i is explicitly given

$$H_d + H_U = \varepsilon_d \sum_i n_i + U \sum_i \left[- (S_z^i)^2 + \frac{n_i^2}{4} \right], \quad (10)$$

where $S_z^i = (n_{i\uparrow} - n_{i\downarrow})/2$ is the z component of the spin operator and $n_i = \sum_\sigma n_{i\sigma} = N_i$.

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

$$H_S = \frac{1}{2} \Gamma \sum_{\langle ii' \rangle} \left[\mathbf{S}_i \cdot \mathbf{S}_{i'} - \frac{1}{4} n_i n_{i'} \right]. \quad (11)$$

The coupling Γ can be calculated by using second-order perturbation theory. For a state in which two neighbors have opposite spin, $|\uparrow, \downarrow\rangle = c_{i\uparrow}^\dagger c_{i'\downarrow}^\dagger |0\rangle$, we obtain the energy gain

$$\Delta E_{\uparrow\downarrow} \sim - \sum_I \langle \uparrow, \downarrow | H_T | I \rangle \langle I | \frac{1}{E(2) + E(0) - 2E(1)} | I \rangle \langle I | H_T | \uparrow, \downarrow \rangle \sim - \frac{2t^2}{U}.$$

Here $|I\rangle$ ranges over the excited states with one of the two neighboring sites doubly occupied and the other empty, $|\uparrow\downarrow, 0\rangle = c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger |0\rangle$, or $|0, \uparrow\downarrow\rangle = c_{i'\uparrow}^\dagger c_{i'\downarrow}^\dagger |0\rangle$; these states can be occupied via virtual hopping processes. For a state in which two neighbors have parallel spins, $|\uparrow, \uparrow\rangle = c_{i\uparrow}^\dagger c_{i'\uparrow}^\dagger |0\rangle$, no virtual hopping is possible because of the Pauli principle, and $\Delta E_{\uparrow\uparrow} = 0$. Thus

$$\frac{1}{2}\Gamma \sim (\Delta E_{\uparrow\uparrow} - \Delta E_{\uparrow\downarrow}) = \frac{1}{2} \frac{4t^2}{U}. \quad (12)$$

The exchange coupling $\Gamma = 4t^2/U$ is positive, i.e., antiferromagnetic.

Canonical transformations [28] provide a scheme for deriving the effective spin model systematically at any perturbation order. Let us consider a unitary transformation of the Hamiltonian

$$H_S = e^{iS} H e^{-iS} = H + [iS, H] + \frac{1}{2} [iS, [iS, H]] + \dots$$

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 H_T into a component H_T^0 that does not change the number of doubly occupied states and two terms that either increase it (H_T^+) or decrease it (H_T^-) by one

$$H_T = -t \sum_{\langle ii' \rangle} \sum_{\sigma} c_{i\sigma}^\dagger c_{i'\sigma} = H_T^0 + H_T^+ + H_T^-,$$

where

$$H_T^0 = -t \sum_{\langle ii' \rangle} \sum_{\sigma} n_{i-\sigma} c_{i\sigma}^\dagger c_{i'\sigma} n_{i'-\sigma} - t \sum_{\langle ii' \rangle} \sum_{\sigma} [1 - n_{i-\sigma}] c_{i\sigma}^\dagger c_{i'\sigma} [1 - n_{i'-\sigma}],$$

$$H_T^+ = -t \sum_{\langle ii' \rangle} \sum_{\sigma} n_{i-\sigma} c_{i\sigma}^\dagger c_{i'\sigma} [1 - n_{i'-\sigma}],$$

$$H_T^- = (H_T^+)^{\dagger}.$$

The term H_T^0 commutes with H_U . The remaining two terms fulfill the commutation rules

$$[H_T^{\pm}, H_U] = \mp U H_T^{\pm}.$$

The operator S can be expressed as a linear combination of powers of the three operators H_T^0 , H_T^+ , and H_T^- . The actual combination, which gives the effective spin model at a given order, can be found via a recursive procedure [28]. At half filling and second order, however, we can simply guess the form of S that leads to the Hamiltonian (11). By defining

$$S = -\frac{i}{U} (H_T^+ - H_T^-)$$

we obtain

$$H_S = H_U + H_T^0 + \frac{1}{U} \left([H_T^+, H_T^-] + [H_T^0, H_T^-] + [H_T^+, H_T^0] \right) + \mathcal{O}(U^{-2}).$$

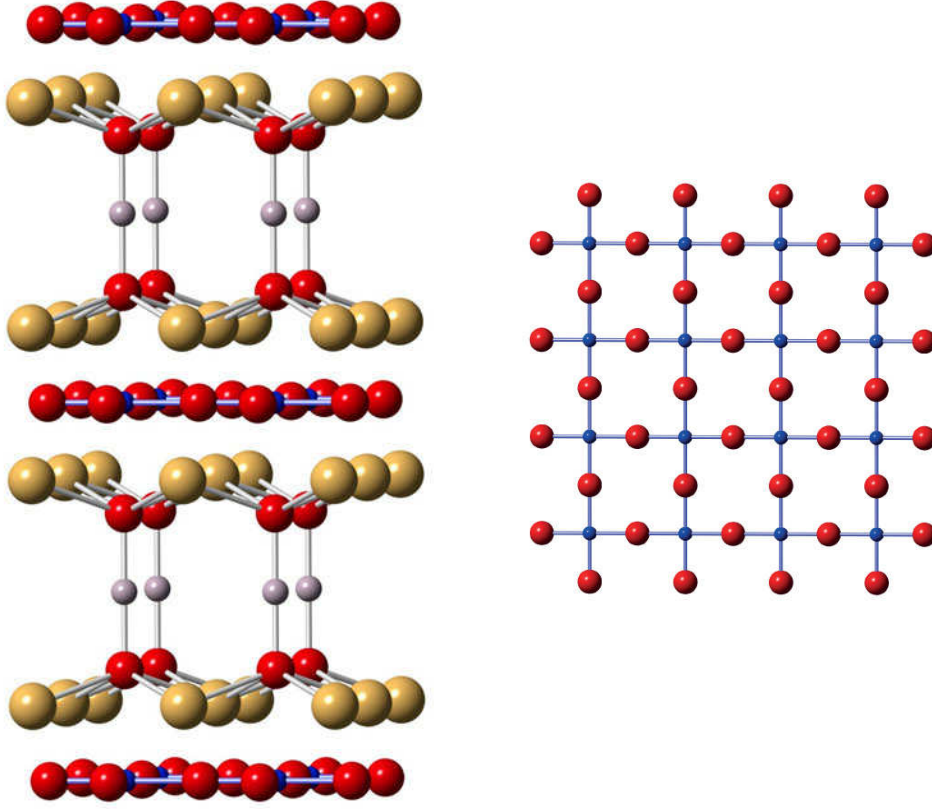


Fig. 2: Left: The crystal structure of $\text{HgBa}_2\text{CuO}_4$ showing the two-dimensional CuO_2 layers. Spheres represent atoms of Cu (blue), O (red), Ba (yellow), and Hg (grey). Right: A CuO_2 layer. The first-nearest-neighbor hopping integral between neighboring Cu sites t is roughly given by $\sim 4t_{pd}^2/\Delta_{dp}$, 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 H_S to the subspace with one electron per site (half filling), no hopping is possible without increasing the number of occupied states; hence, only the term $H_T^- H_T^+$ contributes. After some algebra, we obtain $H_S = H_S^{(2)} + \mathcal{O}(U^{-2})$ with

$$H_S^{(2)} = \frac{1}{2} \frac{4t^2}{U} \sum_{ii'} \left[\mathbf{S}_i \cdot \mathbf{S}_{i'} - \frac{1}{4} n_i n_{i'} \right].$$

The Hubbard model (7) 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 (7). These are strongly correlated organic crystals (one-dimensional case) and high-temperature superconducting cuprates, in short HTSCs (two-dimensional case). An example of HTSC is $\text{HgBa}_2\text{CuO}_4$, whose structure is shown in Fig. 2. 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.

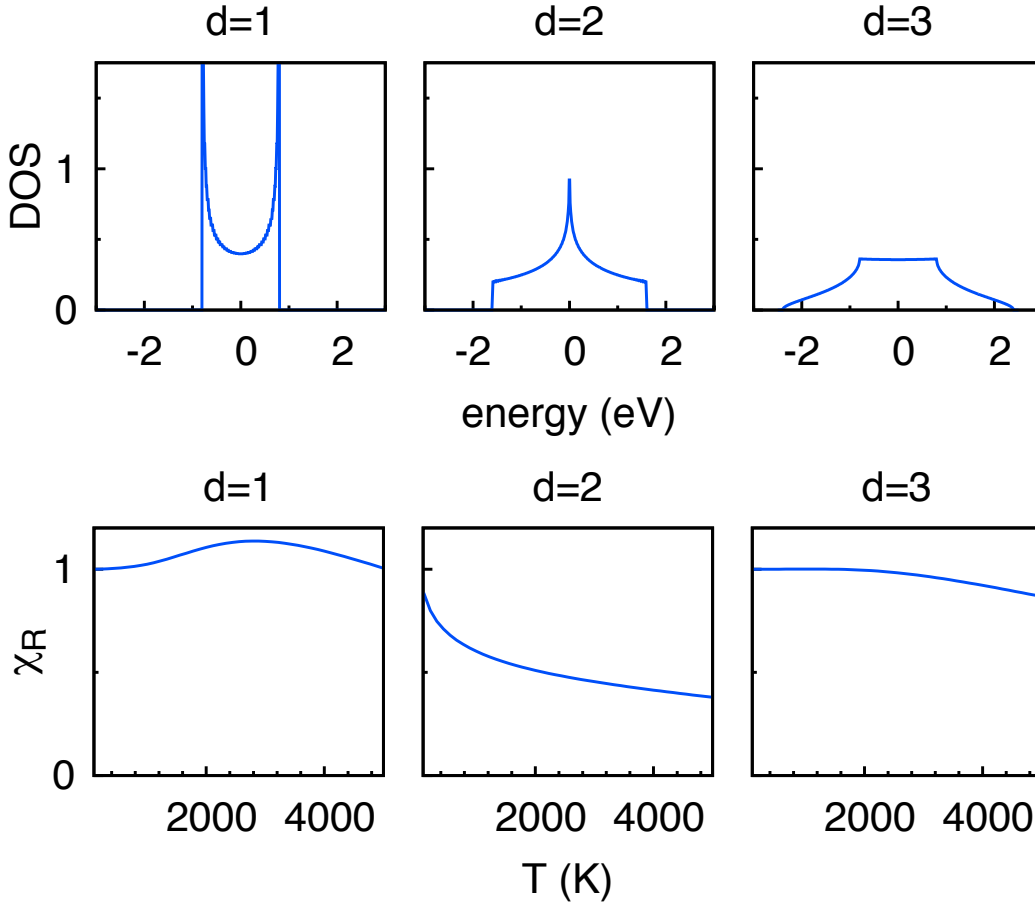


Fig. 3: *Top: Density of states (DOS) per spin, $\rho(\varepsilon)/2$, for a hypercubic lattice in one, two, and three dimensions. The energy dispersion is calculated for $t = 0.4$ eV. The curves exhibit different types of Van-Hove singularities. Bottom: Effect of $\rho(\varepsilon_F)$ on the temperature dependence of $\chi_R = \chi^P(T)/\chi^P(0)$. Up to ~ 1000 K only the logarithmic Van-Hove singularity (two-dimensional case) yields a sizable effect.*

2.1 Weak-correlation limit

2.1.1 The $U = 0$ case: Pauli paramagnetism

Let us consider first the non-interacting limit of the Hubbard model, Hamiltonian (8). In the presence of an external magnetic field $\mathbf{h} = h_z \hat{z}$ the energy $\varepsilon_{\mathbf{k}}$ of a Bloch state is modified by the Zeeman interaction (4) as follows

$$\varepsilon_{\mathbf{k}} \rightarrow \varepsilon_{\mathbf{k}\sigma} = \varepsilon_{\mathbf{k}} + \frac{1}{2} \sigma g \mu_B h_z,$$

where we take the direction of the magnetic field as quantization axis and where on the right-hand side $\sigma = 1$ or -1 depending if the spin is parallel or antiparallel to $\mathbf{h}$. Thus, to linear order in the magnetic field, the $T = 0$ magnetization of the system is

$$M_z = -\frac{1}{2} (g \mu_B) \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} [n_{\mathbf{k}\uparrow} - n_{\mathbf{k}\downarrow}] \sim \frac{1}{4} (g \mu_B)^2 \rho(\varepsilon_F) h_z,$$

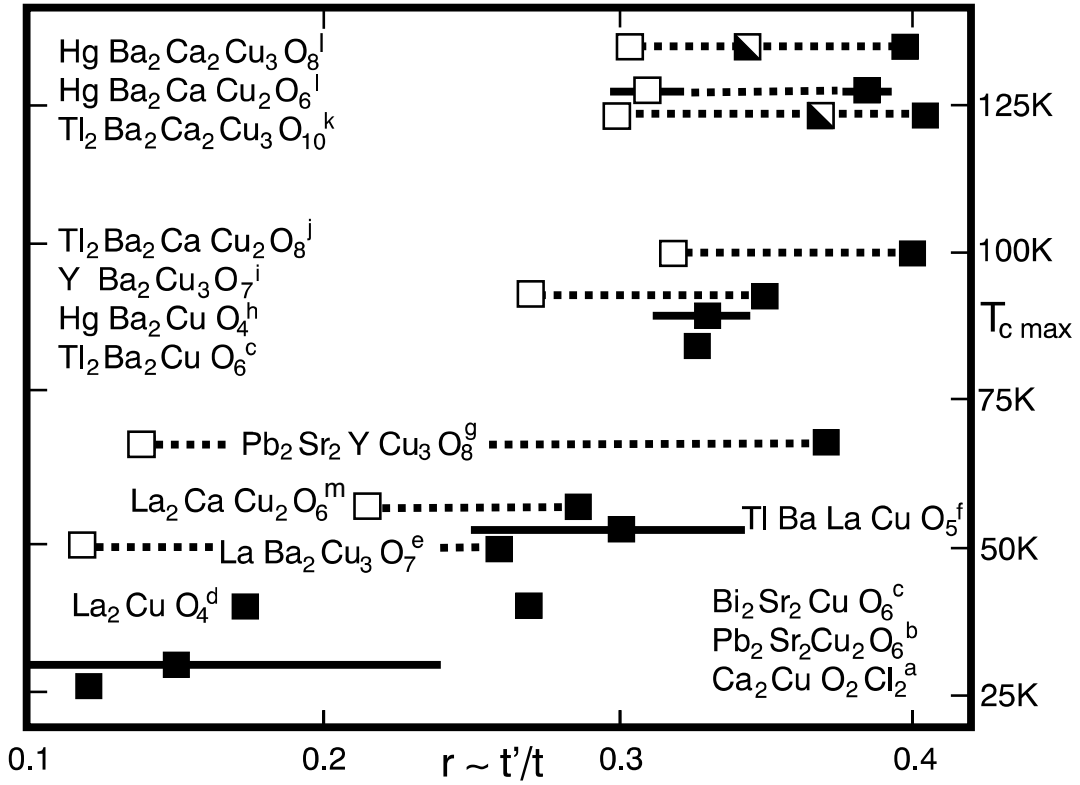


Fig. 4: Band-structure trend in hole-doped cuprates and correlation with $T_{c \text{ max}}$, the maximum value of the critical temperature for superconductivity. From Ref. [29].

where $n_{\mathbf{k}\sigma} = \langle c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} \rangle$ and $N_{\mathbf{k}}$ is the number of $\mathbf{k}$ points; $\rho(\varepsilon_F)$ is the total density of states (DOS) at the Fermi level ε_F . The $T = 0$ susceptibility is then given by the Pauli formula

$$\chi^P(0) = \frac{1}{4} (g\mu_B)^2 \rho(\varepsilon_F).$$

In linear-response theory (see Appendix) the magnetization induced along $\hat{z}$ by an external magnetic field $h_z(\mathbf{q}; \omega) \hat{z}$ oscillating with vector $\mathbf{q}$ is given by

$$M_z(\mathbf{q}; \omega) = \chi_{zz}(\mathbf{q}; \omega) h_z(\mathbf{q}; \omega).$$

The Pauli susceptibility $\chi^P(0)$ is thus the static ($\omega = 0$) and uniform ($\mathbf{q} = 0$) linear response function to an external magnetic field. At finite temperature the Pauli susceptibility takes the form

$$\chi^P(T) = \frac{1}{4} (g\mu_B)^2 \int d\varepsilon \rho(\varepsilon) \left(-\frac{dn(\varepsilon)}{d\varepsilon} \right),$$

where $n(\varepsilon) = 1/(1 + e^{(\varepsilon - \mu)\beta})$ is the Fermi distribution function, $\beta = 1/k_B T$, and μ the chemical potential. $\chi^P(T)$ depends weakly on the temperature; its temperature dependence is more pronounced, however, in the presence of Van-Hove singularities close to the Fermi level (Fig. 3).

2.1.2 The Fermi liquid regime

In *some* limit the independent-particle picture still holds even when the Coulomb interaction is finite. Landau's phenomenological Fermi liquid theory suggests that, at low-enough energy and temperature, the elementary excitations of a weakly interacting system can be described by almost independent fermionic quasiparticles, fermions with effective mass m^* and finite lifetime τ^{QP}

$$\begin{aligned}\varepsilon_{n\mathbf{k}}^{\text{QP}} &= \frac{m}{m^*} \varepsilon_{n\mathbf{k}}, \\ \tau^{\text{QP}} &\propto (aT^2 + b\omega^2)^{-1}.\end{aligned}$$

Remarkably, a very large number of materials do exhibit low-energy Fermi liquid behavior, and the actual violation of the Fermi liquid picture is typically an indication that something surprising is going on. How are quasiparticles related to actual particles, however? Landau postulated that the low-lying states of a weakly-correlated system are well-described by the energy functional

$$E = E_0 + \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma} \delta n_{\mathbf{k}\sigma} + \frac{1}{2} \sum_{\mathbf{k}\sigma} \sum_{\mathbf{k}'\sigma'} f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} \delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'},$$

where E_0 is the ground-state energy, $\delta n_{\mathbf{k}\sigma} = n_{\mathbf{k}\sigma} - n_{\mathbf{k}\sigma}^0$ gives the number of quasiparticles (or quasi-holes), $n_{\mathbf{k}\sigma}$ is the occupation number in the excited state, and $n_{\mathbf{k}\sigma}^0$ is the occupation number of the non-interacting system at $T = 0$. The idea behind this is that $\delta n_{\mathbf{k}\sigma}$ is small with respect to the number of particles and therefore can be used as an expansion parameter. The low-lying elementary excitations are thus fermions with dispersion

$$\varepsilon_{n\mathbf{k}}^{\text{QP}} = \frac{\delta E}{\delta n_{\mathbf{k}\sigma}} = \varepsilon_{\mathbf{k}\sigma} + \sum_{\mathbf{k}'\sigma'} f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} \delta n_{\mathbf{k}'\sigma'},$$

so that

$$E = E_0 + \sum_{\mathbf{k}\sigma} \varepsilon_{n\mathbf{k}}^{\text{QP}} \delta n_{\mathbf{k}\sigma},$$

i.e., the energy of quasiparticles is additive. Remarkably, by definition

$$f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} = \frac{\delta^2 E}{\delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'}}.$$

Hence, $f_{\mathbf{k}\sigma\mathbf{k}'\sigma'}$ is symmetric in all the arguments. If the system has inversion symmetry, $f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} = f_{-\mathbf{k}\sigma-\mathbf{k}'\sigma'}$; furthermore, if the system has time-reversal symmetry, $f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} = f_{-\mathbf{k}-\sigma-\mathbf{k}'-\sigma'}$. It is therefore useful to rewrite it as the sum of a symmetric and an antisymmetric contribution, $f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} = f_{\mathbf{k}\mathbf{k}'}^s + \sigma\sigma' f_{\mathbf{k}\mathbf{k}'}^a$, where

$$\begin{aligned}f_{\mathbf{k}\mathbf{k}'}^s &= \frac{1}{4} \sum_{\sigma\sigma'} f_{\mathbf{k}\sigma\mathbf{k}'\sigma'} \\ f_{\mathbf{k}\mathbf{k}'}^a &= \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' f_{\mathbf{k}\sigma\mathbf{k}'\sigma'}.\end{aligned}$$

Only the symmetric term contributes to the energy of the quasiparticles. Let us consider for simplicity a Fermi gas, the dispersion relation of which has spherical symmetry. Since quasiparticles are only well defined close to the Fermi level, we can assume that $|\mathbf{k}| \sim |\mathbf{k}'| \sim k_F$; therefore, $f_{\mathbf{k}\mathbf{k}'}^s$ and $f_{\mathbf{k}\mathbf{k}'}^a$ depend essentially only on the angle between $\mathbf{k}$ and $\mathbf{k}'$, while the dependence on the $\mathbf{k}$ vector's length is weak. Next, let us expand $f_{\mathbf{k}\mathbf{k}'}^s$ and $f_{\mathbf{k}\mathbf{k}'}^a$ in orthogonal Legendre polynomials $P_l(\cos \theta_{\mathbf{k}\mathbf{k}'})$, where $\theta_{\mathbf{k}\mathbf{k}'}$ is the angle between $\mathbf{k}$ and $\mathbf{k}'$. We have

$$f_{\mathbf{k}\mathbf{k}'}^{s/a} = \rho(\varepsilon_F) \sum_{l=0}^{\infty} F_l^{s/a} P_l(\cos \theta_{\mathbf{k}\mathbf{k}'}),$$

where F_l^s and F_l^a are dimensionless parameters. One can then show that the mass renormalization is given by

$$\frac{m^*}{m} = 1 + \frac{1}{3} F_1^s > 1, \quad F_1^s > 0.$$

Quasiparticles are less compressible than particles, i.e., if κ is the compressibility

$$\frac{\kappa}{\kappa_0} = \frac{m^*}{m} \frac{1}{1 + F_0^s} < 1, \quad F_0^s > 0.$$

They are, however, more spin-polarizable than electrons; correspondingly the system exhibits an enhanced Pauli susceptibility

$$\frac{\chi}{\chi^P} = \frac{m^*}{m} \frac{1}{1 + F_0^a} > 1, \quad F_0^a < 0.$$

It has to be noticed that, because of the finite lifetime of quasiparticles and/or non-Fermi liquid phenomena of various nature, the temperature and energy regime in which the Fermi liquid behavior is observed can be very narrow. This happens, e.g., for heavy-fermion or Kondo systems; we will come back to this point again in the last part of the lecture.

2.1.3 Stoner instabilities

In the presence of the Coulomb interaction $U \neq 0$, finding the solution of the Hubbard model requires many-body techniques. Nevertheless, in the small- U limit, we can already learn a lot about magnetism from Hartree-Fock (HF) static mean-field theory. In the simplest version of the HF approximation we make the following substitution

$$H_U = U \sum_i n_{i\uparrow} n_{i\downarrow} \rightarrow H_U^{\text{HF}} = U \sum_i [n_{i\uparrow} \langle n_{i\downarrow} \rangle + \langle n_{i\uparrow} \rangle n_{i\downarrow} - \langle n_{i\uparrow} \rangle \langle n_{i\downarrow} \rangle].$$

This approximation transforms the Coulomb two-particle interaction into an effective single-particle interaction. Let us search for a ferromagnetic solution and set therefore

$$\langle n_{i\sigma} \rangle = n_\sigma = \frac{n}{2} + \sigma m,$$

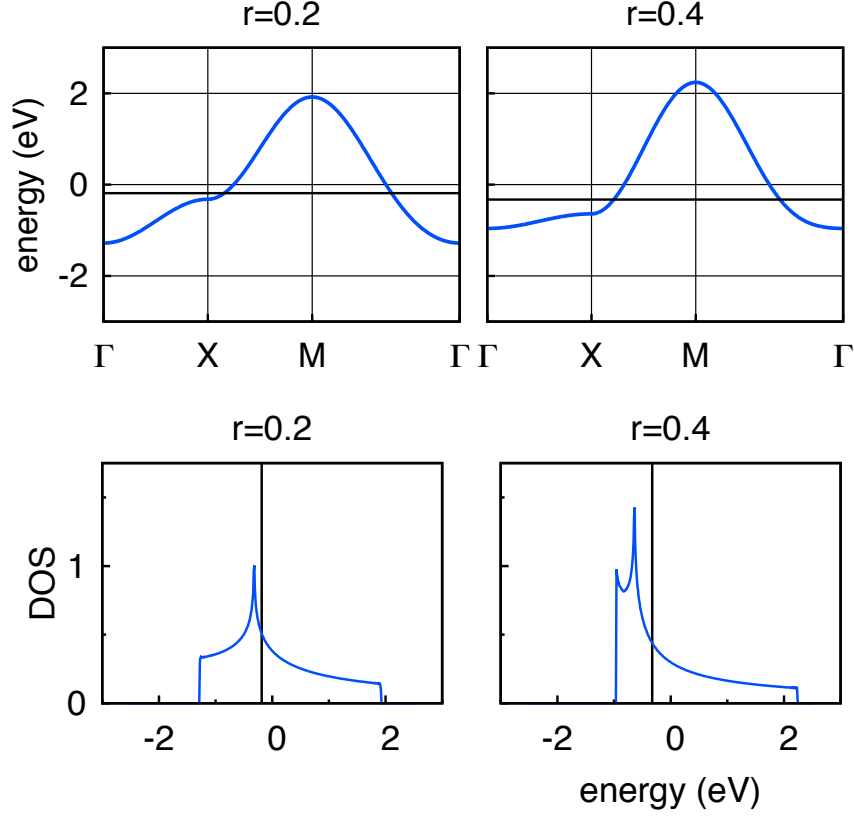


Fig. 5: *Top: Effect of $r = t'/t$ on the band structure of the two-dimensional tight-binding model. Black line: Fermi level at half-filling. Bottom: corresponding density of states per spin.*

where $m = (n_{\uparrow} - n_{\downarrow})/2$ and $n = n_{\uparrow} + n_{\downarrow}$. It is convenient to rewrite the mean-field Coulomb energy as in (10), i.e., as a function of m , n and S_z^i

$$H_U^{\text{HF}} = U \sum_i \left[-2mS_z^i + m^2 + \frac{n^2}{4} \right]. \quad (13)$$

The solution of the problem defined by the Hamiltonian $H_0 + H_U^{\text{HF}}$ amounts to the self-consistent solution of a non-interacting electron system with Bloch energies

$$\varepsilon_{\mathbf{k}\sigma}^U = \varepsilon_{\mathbf{k}} + n_{-\sigma} U = \varepsilon_{\mathbf{k}} + \frac{n}{2} U - \sigma m U.$$

In a magnetic field we additionally have to consider the Zeeman splitting. Thus

$$\varepsilon_{\mathbf{k}\sigma} = \varepsilon_{\mathbf{k}\sigma}^U + \frac{1}{2} g \mu_B h_z \sigma.$$

In the small- U limit and for $T \rightarrow 0$ the magnetization $M_z = -g \mu_B m$ is then given by

$$M_z \sim \chi^P(0) \left[h_z - \frac{2}{g \mu_B} U m \right] = \chi^P(0) [h_z + 2(g \mu_B)^{-2} U M_z].$$

Solving for M_z we find the Stoner expression

$$\chi^S(\mathbf{0}; 0) = \frac{\chi^P(0)}{1 - 2(g \mu_B)^{-2} U \chi^P(0)}.$$

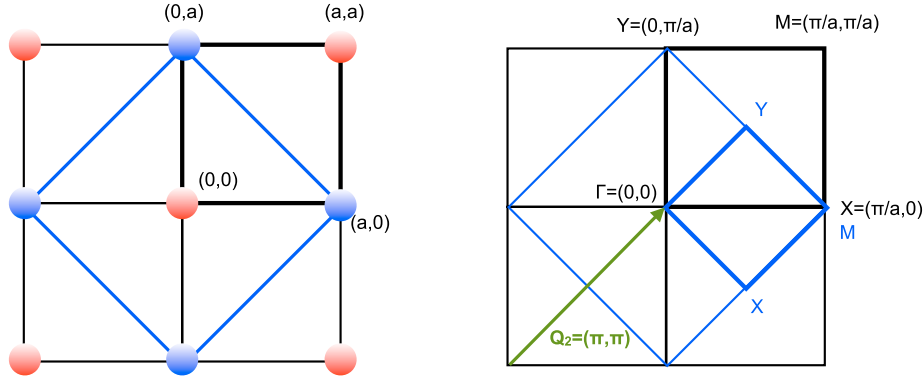


Fig. 6: Doubling of the cell due to antiferromagnetic order and the corresponding folding of the Brillouin zone (BZ) for a two-dimensional hypercubic lattice. The antiferromagnetic $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$ vector is also shown.

Thus with increasing U the $\mathbf{q} = \mathbf{0}$ static susceptibility increases and at the critical value

$$U_c = 2/\rho(\varepsilon_F)$$

diverges, i.e., even an infinitesimal magnetic field can produce a finite magnetization. This means that the ground state becomes unstable against ferromagnetic order.

Let us consider the case of the half-filled d -dimensional hypercubic lattice whose density of states is shown in Fig. 3. In three dimensions the DOS is flat around the Fermi level, e.g., $\rho(\varepsilon_F) \sim 2/W$ where W is the band width. For a flat DOS ferromagnetic instabilities are likely only when $U \sim W$, a rather large value of U , which typically also brings in strong-correlation effects not described by static mean-field theory. In two dimensions we have a rather different situation because a logarithmic Van-Hove singularity is exactly at the Fermi level (Fig. 3); a system with such a density of states is unstable towards ferromagnetism even for very small U . In real materials distortions or long-range interactions typically push the Van-Hove singularities away from the Fermi level. In HTSCs the electronic dispersion is modified as follows by the hopping t' between second-nearest neighbors

$$\varepsilon_{\mathbf{k}} = -2t[\cos(k_x a) + \cos(k_y a)] + 4t' \cos(k_x a) \cos(k_y a).$$

As shown in Fig. 4, the parameter $r \sim t'/t$ ranges typically from ~ 0.15 to 0.4 [29]. Fig. 5 shows that with increasing r the Van-Hove singularity moves downwards in energy.

It is at this point natural to ask ourselves if ferromagnetism is the only possible instability. For a given system, magnetic instabilities with $\mathbf{q} \neq \mathbf{0}$ might be energetically favorable compared to ferromagnetism; an example of a finite- $\mathbf{q}$ instability is antiferromagnetism (see Fig. 6). To investigate finite- $\mathbf{q}$ instabilities we generalize the Stoner criterion. Let us consider a magnetic excitation characterized by the vector $\mathbf{q}$ commensurate with the reciprocal lattice. This magnetic superstructure defines a new lattice; the associated supercell includes $j = 1, \dots, N_j$

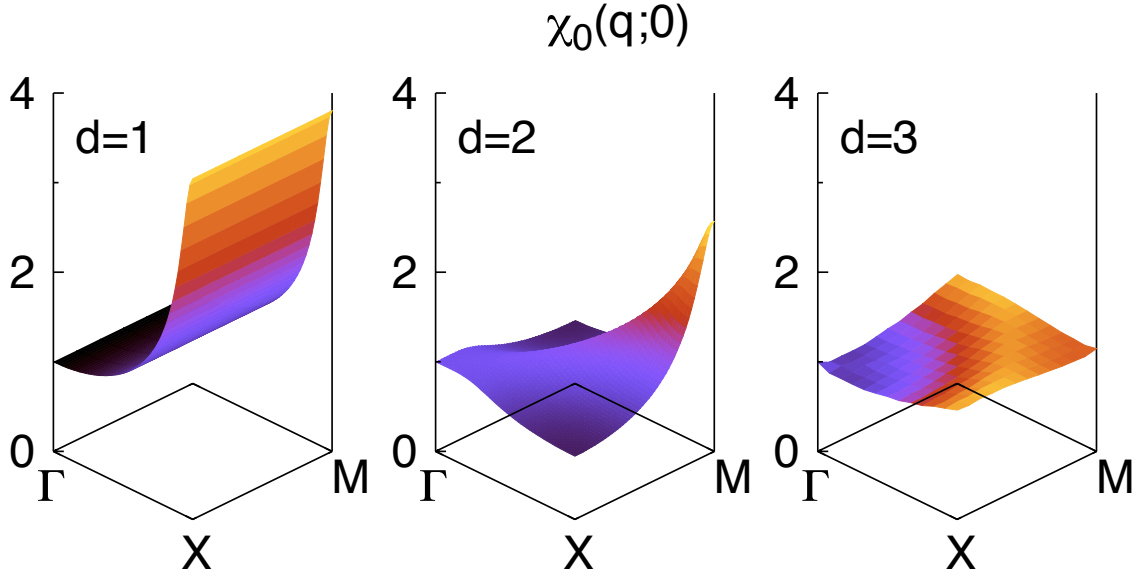


Fig. 7: The ratio $\chi_0(\mathbf{q};0)/\chi_0(\mathbf{0};0)$ in the xy plane for a hypercubic lattice with $t = 0.4$ eV ($T \sim 230$ K) at half filling. From left to right: one, two, and three dimensions.

magnetically inequivalent sites. We therefore define the quantities

$$S_z^i(\mathbf{q}) = \sum_j e^{i\mathbf{q} \cdot \mathbf{R}_j} S_z^{ji},$$

$$\langle S_z^{ji} \rangle = m \cos(\mathbf{q} \cdot \mathbf{R}_j),$$

where j runs over the magnetically inequivalent sites $\{\mathbf{R}_j\}$ and i over the supercells in the lattice. In the presence of a magnetic field oscillating with vector $\mathbf{q}$ and pointing in the z direction, $\mathbf{h}_j = h_z \cos(\mathbf{q} \cdot \mathbf{R}_j) \hat{z}$, the mean-field Coulomb and Zeeman terms can be written as

$$H_U^{\text{HF}} + H_Z = \sum_i \left[\frac{g\mu_B}{2} \left(h_z - \frac{2}{g\mu_B} mU \right) [S_z^i(\mathbf{q}) + S_z^i(-\mathbf{q})] + m^2 + \frac{n^2}{4} \right],$$

where m has to be determined self-consistently. This leads to the generalized Stoner formula

$$\chi^S(\mathbf{q};0) = \frac{1}{2} (g\mu_B)^2 \frac{\chi_0(\mathbf{q};0)}{[1 - U\chi_0(\mathbf{q};0)]}, \quad (14)$$

$$\chi_0(\mathbf{q};0) = -\frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \frac{n_{\mathbf{k}+\mathbf{q}} - n_{\mathbf{k}}}{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}.$$

Expression (14) is also known as the RPA (acronym for random-phase approximation) susceptibility. For $\mathbf{q} = \mathbf{0}$ in the $T \rightarrow 0$ limit we recover the ferromagnetic RPA susceptibility with

$$\chi_0(\mathbf{0};0) = 2 (g\mu_B)^{-2} \chi^P(0) \sim \frac{1}{2} \rho(\varepsilon_F).$$

Figure 7 shows the non-interacting susceptibility in the xy plane for our d -dimensional hy-

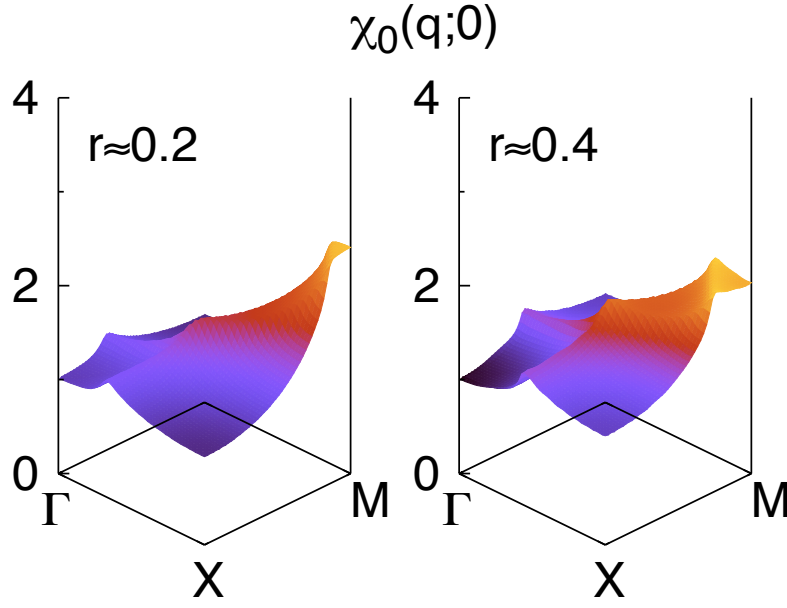


Fig. 8: The ratio $\chi_0(\mathbf{q};0)/\chi_0(0;0)$ in the xy plane for the two-dimensional hypercubic lattice with $t = 0.4$ eV (230 K) at half filling. Left: $t' = 0.2t$. Right: $t' = 0.4t$.

percubic lattice. The figure shows that in the one-dimensional case the susceptibility diverges at the antiferromagnetic vector $\mathbf{Q}_1 = (\pi/a, 0, 0)$; in two dimensions this happens at $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$; in three dimension at $\mathbf{Q}_3 = (\pi/a, \pi/a, \pi/a)$, not shown in the figure. The energy dispersion at these vectors exhibits the property of perfect nesting

$$\varepsilon_{\mathbf{k}+\mathbf{Q}_i} = -\varepsilon_{\mathbf{k}}.$$

Remarkably, the $T = 0$ non-interacting susceptibility $\chi_0(\mathbf{Q}_i;0)$ diverges logarithmically at the nesting vector unless the density of states is zero at the Fermi level ($\varepsilon \rightarrow 0$)

$$\chi_0(\mathbf{Q}_i;0) \propto \frac{1}{4} \int_{-\infty}^{\varepsilon_F=0} d\varepsilon \rho(\varepsilon) \frac{1}{\varepsilon} \rightarrow \infty.$$

Under these conditions an arbitrarily small U can cause a magnetic transition with magnetic vector $\mathbf{Q}_i$. In the two-dimensional case we have reached a similar conclusion for the $T = 0$ ferromagnetic ($\mathbf{q} = 0$) instability. The finite-temperature susceptibility $\chi_0(\mathbf{q};0)$, however, shows that the antiferromagnetic instability is the strongest (Fig. 7). Perfect nesting at $\mathbf{Q}_2$ is suppressed by $t' \neq 0$

$$\varepsilon_{\mathbf{k}+\mathbf{Q}_2} = -\varepsilon_{\mathbf{k}} + 8t' \cos(k_x a) \cos(k_y a).$$

Figure 8 shows how the susceptibility is modified by $t' \neq 0$ (half filling). The $\mathbf{Q}_2$ instability is important even for $t' \sim 0.4t$, but instabilities at incommensurate vectors around it are stronger. As a last remark it is important to notice that the RPA expression (14) depends on the filling only through the density of states, i.e., magnetic instabilities described by the Stoner formula can exist at *any* filling. This is very different from the case of the local-moment regime that we will discuss starting with the next section.

Ion		n	S	L	J	$^{2S+1}L_J$
V^{4+}	Ti^{3+}	$3d^1$	$1/2$	2	$3/2$	$^2D_{3/2}$
	V^{3+}	$3d^2$	1	3	2	$^3F_2$
	Cr^{3+}	$3d^3$	$3/2$	3	$3/2$	$^4F_{3/2}$
	Mn^{3+}	$3d^4$	2	2	0	$^5D_0$
	Fe^{3+}	$3d^5$	$5/2$	0	$5/2$	$^6S_{5/2}$
	Fe^{2+}	$3d^6$	2	2	4	$^5D_4$
	Co^{2+}	$3d^7$	$3/2$	3	$9/2$	$^4F_{9/2}$
	Ni^{2+}	$3d^8$	1	3	4	$^3F_4$
	Cu^{2+}	$3d^9$	$1/2$	2	$5/2$	$^2D_{5/2}$

Table 1: Quantum numbers of the ground-state multiplet for several transition-metal ions with partially filled d shells. In transition-metal oxides the angular momentum is typically quenched because of the crystal field and therefore only the total spin matters.

2.2 Atomic limit

2.2.1 Paramagnetism of isolated ions

As we have seen, the ground-state multiplet of free ions with partially occupied shells can be determined via Hund's rules. In Tab. 1 and Tab. 2 we can find the values of the S , L , and J quantum numbers for the ground-state multiplets of the most common transition-metal and rare-earth ions. If $t = 0$ and $n = 1$, the Hubbard model (7) describes precisely a collection of idealized free ions with an incomplete shell. For such idealized ions the only possible multiplet is the one with quantum numbers $J = S = 1/2, L = 0$. In the presence of a uniform external magnetic field $h_z \hat{z}$ we can then obtain the magnetization per atom as

$$M_z = \langle M_z^i \rangle = -g\mu_B \frac{\text{Tr} [e^{-g\mu_B h_z \beta S_z^i} S_z^i]}{\text{Tr} [e^{-g\mu_B h_z \beta S_z^i}]} = g\mu_B S \tanh(g\mu_B h_z \beta S),$$

and thus

$$\frac{\partial M_z}{\partial h_z} = (g\mu_B S)^2 \frac{1}{k_B T} [1 - \tanh^2(g\mu_B h_z \beta S)].$$

The static uniform susceptibility is then given by the $h \rightarrow 0$ limit

$$\chi_{zz}(0; 0) = (g\mu_B S)^2 \frac{1}{k_B T} = \frac{C_{1/2}}{T}, \quad (15)$$

where $C_{1/2}$ is the $S = 1/2$ Curie constant. If $S = 1/2$, the relation $S^2 = S(S+1)/3$ holds. Thus, for reasons that will become clear shortly, the Curie constant is typically expressed as

$$C_{1/2} = \frac{(g\mu_B)^2 S(S+1)}{3k_B}.$$

If the ions have ground-state total angular momentum J , we can calculate the susceptibility with the same technique, provided that we replace g with the Landé factor g_J

$$g_J = \frac{\langle J J_z L S | (g\mathbf{S} + \mathbf{L}) \cdot \mathbf{J} | J J_z L S \rangle}{\langle J J_z L S | \mathbf{J} \cdot \mathbf{J} | J J_z L S \rangle} \sim \frac{3}{2} + \frac{S(S+1) - L(L+1)}{2J(J+1)}$$

Ion	n	S	L	J	$^{2S+1}L_J$	g_J
Ce ³⁺	$4f^1$	1/2	3	5/2	$^2F_{5/2}$	6/7
Pr ³⁺	$4f^2$	1	5	4	$^3H_4$	4/5
Nd ³⁺	$4f^3$	3/2	6	9/2	$^4I_{9/2}$	8/11
Pm ³⁺	$4f^4$	2	6	4	$^5I_4$	3/5
Sm ³⁺	$4f^5$	5/2	5	5/2	$^6H_{5/2}$	2/7
Eu ³⁺	$4f^6$	3	3	0	$^7F_0$	0
Gd ³⁺	$4f^7$	7/2	0	7/2	$^8S_{7/2}$	2
Tb ³⁺	$4f^8$	3	3	6	$^7F_6$	3/2
Dy ³⁺	$4f^9$	5/2	5	15/2	$^6H_{15/2}$	4/3
Ho ³⁺	$4f^{10}$	2	6	8	$^5I_8$	5/4
Er ³⁺	$4f^{11}$	3/2	6	15/2	$^4I_{15/2}$	6/5
Tm ³⁺	$4f^{12}$	1	5	6	$^3H_6$	7/6
Yb ³⁺	$4f^{13}$	1/2	3	7/2	$^2F_{7/2}$	8/7

Table 2: Quantum numbers of the ground-state multiplet for rare-earth ions with partially filled f shells and corresponding g_J factor. In $4f$ materials the crystal field is typically small; thus the ground-state multiplet is to first approximation close to that of the corresponding free ion.

and calculate the thermal average of the magnetization $\mathbf{M} = -g_J\mu_B\mathbf{J}$, accounting for the $2J + 1$ degeneracy of the multiplet. The result is

$$M_z = \langle M_z^i \rangle = g_J\mu_B J B_J(g_J\mu_B h_z \beta J),$$

where $B_J(x)$ is the Brillouin function

$$B_J(x) = \frac{2J+1}{2J} \coth\left(\frac{2J+1}{2J}x\right) - \frac{1}{2J} \coth\left(\frac{1}{2J}x\right).$$

In the low-temperature ($x \rightarrow \infty$) limit $B_J(x) \sim 1$, and thus the magnetization approaches its saturation value in which all atoms are in the ground state

$$M_z \sim g_J\mu_B J \equiv M_0.$$

In the high-temperature ($x \rightarrow 0$) limit

$$B_J(x) \sim x \frac{J+1}{3J} \left[1 - \frac{2J^2 + 2J + 1}{30J^2} x^2 \right],$$

and thus the susceptibility exhibits the Curie high-temperature behavior

$$\chi_{zz}(\mathbf{0}; 0) \sim \frac{C_J}{T} = \frac{\mu_J^2}{3k_B T},$$

where the generalized Curie constant is

$$C_J = \frac{(g_J\mu_B)^2 J(J+1)}{3k_B}$$

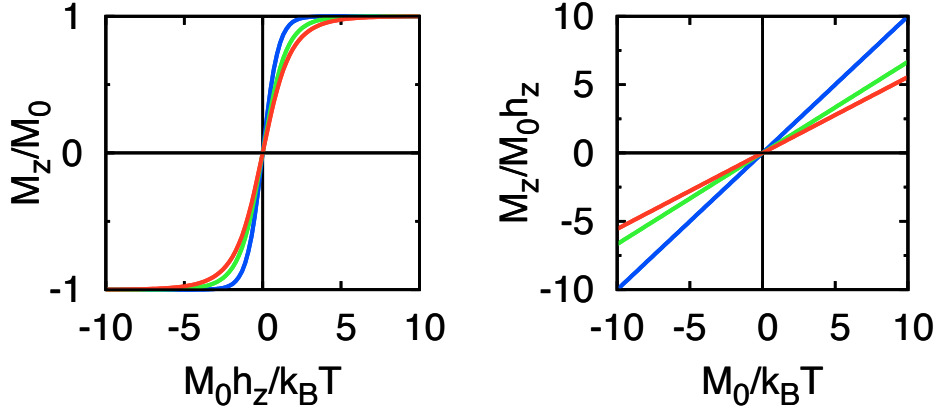


Fig. 9: Left: $M_z/M_0 = B_J(x)$ as a function of $x = h_z M_0 / k_B T$. The different lines correspond to $J = 1/2$ (blue), $J = 1$ (green) and $J = 3/2$ (red). Right: The ratio $M_z/M_0 h_z$ for finite magnetic field in the small x limit; the slope is $(J + 1)/3J$.

and where $\mu_J = g_J \mu_B \sqrt{J(J+1)}$ is the total magnetic moment. Correspondingly, the susceptibility decreases as $1/T$ with increasing T (Fig. 9). We have thus the three limiting cases

$$\chi_{zz}(\mathbf{0}; 0) \sim \begin{cases} 0 & k_B T / |M_0| h_z \rightarrow 0 \\ C_J / T & |M_0| h_z / k_B T \rightarrow 0 \\ C_J / T & h_z \rightarrow 0 \end{cases}.$$

Remarkably, the $T \rightarrow 0$ and $h_z \rightarrow 0$ limit cannot be interchanged. If h_z is finite the susceptibility goes to zero in the $T \rightarrow 0$ limit; if we instead perform the $h_z \rightarrow 0$ limit first it diverges with the Curie form $1/T$. The point $h_z = T = 0$ is a critical point in the phase space.

Let us return to the $S = 1/2$ case, i.e., the one relevant for the Hubbard model. It is interesting to calculate the inter-site spin correlation function $\mathcal{S}_{i,i'}$

$$\mathcal{S}_{i,i'} = \langle (\mathbf{S}_i - \langle \mathbf{S}_i \rangle) \cdot (\mathbf{S}_{i'} - \langle \mathbf{S}_{i'} \rangle) \rangle = \langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle - \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle.$$

We express $\langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle$ in the form $[S(S+1) - S_i(S_i+1) - S_{i'}(S_{i'}+1)]/2$, where $S_i = S_{i'} = 1/2$ and $S = \mathbf{S}_i + \mathbf{S}_{i'}$ is the total spin. Then, since in the absence of a magnetic field $\langle \mathbf{S}_i \rangle = \langle \mathbf{S}_{i'} \rangle = 0$,

$$\mathcal{S}_{i,i'} = [S(S+1) - 3/2]/2 = \begin{cases} 1/4 & S = 1 \\ -3/4 & S = 0 \end{cases}.$$

The ideal paramagnetic state is however characterized by uncorrelated sites. Hence

$$\mathcal{S}_{i,i'} = \langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle \sim \begin{cases} \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle \sim 0 & i \neq i' \\ \langle \mathbf{S}_i \cdot \mathbf{S}_i \rangle = 3/4 & i = i' \end{cases}. \quad (16)$$

The (ideal) paramagnetic phase is thus quite different from a spatially disordered state, i.e., a situation in which each ion has a spin oriented in a given direction but spin orientations are randomly distributed. In the latter case, in general, $\langle \mathbf{S}^i \cdot \mathbf{S}^{i'} \rangle \neq 0$ for $i' \neq i$, even if, e.g., the sum of $\langle S_z^i \cdot S_z^{i'} \rangle$ over all sites i' with $i' \neq i$ is zero

$$\sum_{i' \neq i} \langle S_z^i \cdot S_z^{i'} \rangle \sim 0.$$

The high-temperature static susceptibility can be obtained from the correlation function Eq. (16) using the *fluctuation-dissipation theorem* and the Kramers-Kronig relations (see Appendix). The result is

$$\chi_{zz}(\mathbf{q}; 0) \sim \frac{(g\mu_B)^2}{k_B T} \sum_j \mathcal{S}_{zz}^{i,i+j} e^{i\mathbf{q} \cdot (\mathbf{R}_i - \mathbf{R}_{i+j})} = \chi_{zz}^i(T) = \frac{M_0^2}{k_B T} = \frac{C_{1/2}}{T}. \quad (17)$$

This shows that $\chi_{zz}(\mathbf{q}; 0)$ is $\mathbf{q}$ -independent and coincides with the local susceptibility $\chi_{zz}^i(T)$

$$\chi_{zz}(\mathbf{0}; 0) = \lim_{h_z \rightarrow 0} \frac{\partial M_z}{\partial h_z} = \chi_{zz}^i(T).$$

How can the spin susceptibility (17) be obtained directly from Hamiltonian (10), the atomic limit of the Hubbard model? To calculate it we can use, e.g., the imaginary-time and Matsubara-frequency formalism (see Appendix). Alternatively at high temperatures we can obtain it from the correlation function as we have just seen. The energies of the four atomic states are given by (9) and, at half filling, the chemical potential is $\mu = \varepsilon_d + U/2$. Therefore

$$\chi_{zz}(\mathbf{0}; 0) \sim \frac{(g\mu_B)^2}{k_B T} \left(\frac{\text{Tr} [e^{-\beta(H_i - \mu N_i)} (S_z^i)^2]}{\text{Tr} [e^{-\beta(H_i - \mu N_i)}]} - \left[\frac{\text{Tr} [e^{-\beta(H_i - \mu N_i)} S_z^i]}{\text{Tr} [e^{-\beta(H_i - \mu N_i)}]} \right]^2 \right) = \frac{C_{1/2}}{T} \frac{e^{\beta U/2}}{1 + e^{\beta U/2}}.$$

Thus, the susceptibility depends on the energy scale

$$U = E(N_i + 1) + E(N_i - 1) - 2E(N_i).$$

If we perform the limit $U \rightarrow \infty$, we effectively eliminate doubly occupied and empty states. In this limit, we recover the expression that we found for the spin $S = 1/2$ model, Eq. (17). This is a trivial example of downfolding, in which the low-energy and high-energy sector are decoupled in the Hamiltonian from the start. In the large- U limit the high-energy states are integrated out leaving the system in a magnetic $S = 1/2$ state.

2.2.2 Larmor diamagnetism and Van Vleck paramagnetism

For ions with $J = 0$ the ground-state multiplet, in short $|0\rangle$, is non-degenerate and the linear correction to the ground-state total energy due to the Zeeman term is zero. Remarkably, for open-shell ions the magnetization nevertheless remains finite because of higher-order corrections. At second order there are two contributions for the ground state. The first is the Van-Vleck term

$$M_z^{\text{VV}} = 2h_z \mu_B^2 \sum_I \frac{|\langle 0 | (L_z + gS_z) | I \rangle|^2}{E_I - E_0},$$

where E_I is the energy of the excited state $|I\rangle$ and E_0 the energy of the ground-state multiplet. The Van Vleck term is weakly temperature-dependent and typically small. The second term is the diamagnetic Larmor contribution

$$M_z^{\text{L}} = -\frac{1}{4} h_z \langle 0 | \sum_i (x_i^2 + y_i^2) | 0 \rangle.$$

The Larmor and Van Vleck terms have opposite signs and typically compete with each other.

2.3 Strong-correlation limit

2.3.1 From the Hubbard model to the Heisenberg model

In the large- U limit and at half filling we can map the Hubbard model onto an effective Heisenberg model. In this section we solve the latter using static mean-field theory. In the mean-field approximation we replace the Heisenberg Hamiltonian (11) with

$$H_S^{\text{MF}} = \frac{1}{2} \Gamma \sum_{\langle ii' \rangle} \left[\mathbf{S}_i \cdot \langle \mathbf{S}_{i'} \rangle + \langle \mathbf{S}_i \rangle \cdot \mathbf{S}_{i'} - \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle - \frac{1}{4} n_i n_{i'} \right].$$

In the presence of an external magnetic field $\mathbf{h}$ we add the Zeeman term and have in total

$$H = g\mu_B \sum_i [\mathbf{S}_i \cdot (\mathbf{h} + \mathbf{h}_i^m) + \text{const.}] ,$$

$$\mathbf{h}_i^m = n_{\langle ii' \rangle} \Gamma \langle \mathbf{S}_{i'} \rangle / g\mu_B ,$$

where $n_{\langle ii' \rangle}$ is the number of first nearest neighbors and $\mathbf{h}_i^m$ is the molecular field at site i . We define the quantization axis z as the direction of the external magnetic field, $\mathbf{h} = h_z \hat{z}$, and assume that $\hat{z}$ is also the direction of the molecular field, $\mathbf{h}_i^m = \Delta h_z^i \hat{z}$. Since $\Gamma > 0$ and hypercubic lattices are bipartite, the likely magnetic order is two-sublattice antiferromagnetism. Thus we set $M_z^A = -g\mu_B \langle S_z^i \rangle$, $M_z^B = -g\mu_B \langle S_z^{i'} \rangle$, where A and B are the two sublattices, $i \in A$ and $i' \in B$. In the absence of an external magnetic field, the total magnetization per formula unit, $M_z = (M_z^B + M_z^A)/2$, vanishes in the antiferromagnetic state ($M_z^B = -M_z^A$). We define therefore as the order parameter $\sigma_m = 2m = (M_z^B - M_z^A)/2M_0$, which is zero only above the critical temperature for antiferromagnetic order. We then calculate the magnetization for each sublattice and find the system of coupled equations

$$\begin{cases} M_z^A/M_0 &= B_{1/2} [M_0(h_z + \Delta h_z^A)\beta] \\ M_z^B/M_0 &= B_{1/2} [M_0(h_z + \Delta h_z^B)\beta] \end{cases} , \quad (18)$$

where

$$\begin{cases} \Delta h_z^A &= -(M_z^B/M_0) S^2 \Gamma n_{\langle ii' \rangle} / M_0 \\ \Delta h_z^B &= -(M_z^A/M_0) S^2 \Gamma n_{\langle ii' \rangle} / M_0 \end{cases} .$$

For $h_z = 0$ the system (18) can be reduced to the single equation

$$\sigma_m = B_{1/2} [\sigma_m S^2 \Gamma n_{\langle ii' \rangle} \beta] . \quad (19)$$

This equation always has the trivial solution $\sigma_m = 0$. Figure 10 shows that, for small-enough temperatures it also has a non-trivial solution $\sigma_m \neq 0$. The order parameter σ_m equals ± 1 at zero temperature, and its absolute value decreases with increasing temperature. It becomes zero for $T \geq T_N$ with

$$k_B T_N = \frac{S(S+1)}{3} n_{\langle ii' \rangle} \Gamma .$$

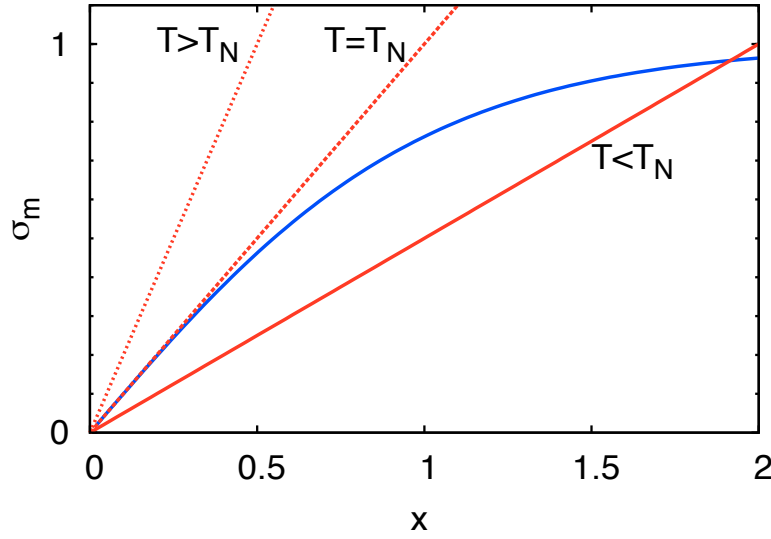


Fig. 10: The self-consistent solution of Eq. (20) for $\sigma_m \geq 0$. The blue line shows the right-hand side of the equation, the Brillouin function $B_{1/2}(x)$, with $x = \sigma_m T_N / T$. The red lines show the left-hand side of the equation, $\sigma_m(x) = \alpha x$, with $\alpha = T / T_N$; the three different curves correspond to representative T / T_N values.

If $T \sim T_N$, we can find the non-trivial solution by first rewriting (19) as

$$\sigma_m = B_{1/2} \left[\frac{T_N}{T} \sigma_m \right]. \quad (20)$$

The inverse of this equation yields T / T_N as a function of σ_m

$$\frac{T}{T_N} = \frac{\sigma_m}{B_{1/2}^{-1}[\sigma_m]}.$$

If $T \sim T_N$, the parameter σ_m is small. We then expand the right-hand-side in powers of σ_m

$$\frac{\sigma_m}{B_{1/2}^{-1}(\sigma_m)} \sim \frac{\sigma_m}{\sigma_m + \sigma_m^3/3 + \dots} \sim 1 - \sigma_m^2/3 + \dots$$

This leads to the following expression

$$\sigma_m = \sqrt{3} \left(1 - \frac{T}{T_N} \right)^{1/2},$$

which shows that the order parameter has a discontinuous temperature derivative at $T = T_N$.

It is interesting to derive the expression of the static uniform susceptibility. For this we go back to the system of equations (18) and calculate from it the total magnetization M_z . In the weak magnetic field limit, $M_z^A \sim -\sigma_m M_0 + \chi_{zz}(\mathbf{0}; 0) h_z$, and $M_z^B \sim \sigma_m M_0 + \chi_{zz}(\mathbf{0}; 0) h_z$. Then, by performing the first derivative of M_z with respect to h_z in the $h_z \rightarrow 0$ limit, we obtain

$$\chi_{zz}(\mathbf{0}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T + (1 - \sigma_m^2)T_N}.$$

The uniform susceptibility vanishes at $T = 0$ and reaches the maximum at $T = T_N$, where it takes the value $C_{1/2}/2T_N$. In the high-temperature regime $\sigma_m = 0$ and

$$\chi_{zz}(\mathbf{0}; 0) \sim \frac{C_{1/2}}{T + T_N},$$

which is *smaller* than the susceptibility of free $S = 1/2$ magnetic ions.

The magnetic linear response is quite different if we apply an external field $\mathbf{h}_\perp$ perpendicular to the spins in the antiferromagnetic lattice. The associated perpendicular magnetization is

$$M_\perp \sim M_0 \frac{\sigma_m(g\mu_B h_\perp)}{\sqrt{(g\mu_B h_\perp)^2 + (4\sigma_m)^2(k_B T_N)^2}},$$

and therefore the perpendicular susceptibility is temperature-independent for $T \leq T_N$

$$\chi_\perp(\mathbf{0}; 0) = \lim_{h_\perp \rightarrow 0} \frac{dM_\perp}{dh_\perp} = \frac{C_{1/2}}{2T_N}.$$

Hence, for $T < T_N$ the susceptibility is anisotropic, $\chi_{zz}(\mathbf{0}; 0) = \chi_\parallel(\mathbf{0}; 0) \neq \chi_\perp(\mathbf{0}; 0)$; at absolute zero $\chi_\parallel(\mathbf{0}; 0)$ vanishes, but the response to $\mathbf{h}_\perp$ remains strong. For $T > T_N$ the order parameter is zero and the susceptibility isotropic, $\chi_\parallel(\mathbf{0}; 0) = \chi_\perp(\mathbf{0}; 0)$.

We have up to now considered antiferromagnetic order only. What about other magnetic instabilities? Let us consider first ferromagnetic order. For a ferromagnetic spin arrangement, by repeating the calculation, we find

$$\chi_{zz}(\mathbf{0}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T - (1 - \sigma_m^2)T_C},$$

where $T_C = -S(S+1)n_{\langle ii' \rangle}\Gamma/3k_B$ is, if the exchange coupling Γ is negative, the critical temperature for ferromagnetic order. Then, in contrast to the antiferromagnetic case, the high-temperature uniform susceptibility is *larger* than that of free $S = 1/2$ magnetic ions.

For a generic magnetic structure characterized by a vector $\mathbf{q}$ and a supercell with $j = 1, \dots, N_j$ magnetically inequivalent sites we make the ansatz

$$\langle M_z^{ji} \rangle = -\sigma_m M_0 \cos(\mathbf{q} \cdot \mathbf{R}_j) = -g\mu_B m \cos(\mathbf{q} \cdot \mathbf{R}_j),$$

where σ_m is again the order parameter, i identifies the supercell, and $\mathbf{R}_j$ the position of the j -th inequivalent site. We consider a magnetic field rotating with the same $\mathbf{q}$ vector. By using the static mean-field approach we then find

$$k_B T_{\mathbf{q}} = \frac{S(S+1)}{3} \Gamma_{\mathbf{q}}, \quad \Gamma_{\mathbf{q}} = - \sum_{ij \neq 0} \Gamma^{00,ij} e^{i\mathbf{q} \cdot (\mathbf{T}_i + \mathbf{R}_j)}, \quad (21)$$

where $\Gamma^{00,ij}$ is the exchange coupling between the spin at the origin and the spin at site $\mathbf{T}_i + \mathbf{R}_j$ (ij in short); $\{\mathbf{R}_j\}$ are vectors inside a supercell and $\{\mathbf{T}_i\}$ are lattice vectors. In our example, $T_0 = T_C$ and $T_{\mathbf{q}_{\text{AF}}} = T_N = -T_C$. Thus we have

$$\chi_{zz}(\mathbf{q}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T - (1 - \sigma_m^2)T_{\mathbf{q}}}, \quad (22)$$

which diverges at $T = T_Q$. The susceptibility $\chi_{zz}(\mathbf{q}; 0)$ reflects the spatial extent of correlations, i.e., the *correlation length* ξ ; the divergence of the susceptibility at T_Q is closely related to the divergence of ξ . To see this we calculate ξ for a hypercubic three-dimensional lattice, assuming that the system has only one instability with vector $\mathbf{Q}$. First we expand Eq. (21) around $\mathbf{Q}$ obtaining $T_q \sim T_Q + \alpha(\mathbf{q} - \mathbf{Q})^2 + \dots$, and then we calculate $\chi_{zz}^{00,ji}$, the Fourier transform of Eq. (22). We find that $\chi_{zz}^{00,ji}$ decays exponentially with $r = |\mathbf{T}_i + \mathbf{R}_j|$, i.e., $\chi_{zz}^{00,ji} \propto e^{-r/\xi}/r$. The range of the correlations is $\xi \propto [T_Q/(T - T_Q)]^{1/2}$, which becomes infinite at $T = T_Q$. It is important to notice that in principle there can be instabilities at any $\mathbf{q}$ vector, i.e., $\mathbf{q}$ need not be commensurate with reciprocal lattice vectors. The value of $\mathbf{q}$ for which T_q is the largest determines (within static mean-field theory) the type of magnetic order that is realized. The antiferromagnetic structure in Fig. 6 corresponds to $\mathbf{q}_{AF} = \mathbf{Q}_2 = (\pi/a, \pi/a, 0)$.

In real systems the spin S is typically replaced by an *effective magnetic moment*, μ_{eff} , and therefore $C_{1/2} \rightarrow C_{\text{eff}} = \mu_{\text{eff}}^2/3k_B$. It follows that μ_{eff} is the value of the product $3k_B T \chi_{zz}(\mathbf{q}; 0)$ in the high-temperature limit (here $T \gg T_Q$). The actual value of μ_{eff} depends, as we have discussed in the introduction, on the Coulomb interaction, the spin-orbit coupling and the crystal field. In addition, the effective moment can be screened by many-body effects, as happens for Kondo impurities; we will discuss the latter case in the last section.

2.3.2 The Hartree-Fock approximation

We have seen that Hartree-Fock (HF) mean-field theory yields Stoner magnetic instabilities in the weak-coupling limit. Can it also describe magnetism in the local-moment regime ($t/U \ll 1$)? Let us focus on the half-filled two-dimensional Hubbard model for a square lattice, and let us analyze two possible magnetically ordered states, the ferro- and the antiferromagnetic state. If we are only interested in the ferromagnetic or the paramagnetic solution, the HF approximation of the Coulomb term in the Hubbard model is given by Eq. (13). Thus the Hamiltonian is $H = H_d + H_T + H_U^{\text{HF}}$ with $H_U^{\text{HF}} = U \sum_i [-2m S_z^i + m^2 + \frac{1}{4}n^2]$. For periodic systems it is convenient to write H in $\mathbf{k}$ space. We then adopt as one-electron basis the Bloch states

$$\Psi_{\mathbf{k}\sigma}(\mathbf{r}) = \frac{1}{\sqrt{N_s}} \sum_i e^{i\mathbf{k} \cdot \mathbf{T}_i} \Psi_{i\sigma}(\mathbf{r}),$$

where $\Psi_{i\sigma}(\mathbf{r})$ is a Wannier function with spin σ , $\mathbf{T}_i$ a lattice vector, and N_s the number of lattice sites. The term H_U^{HF} depends on the spin operator S_z^i , which can be written in $\mathbf{k}$ space as

$$S_z^i = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\mathbf{k}'} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{T}_i} \underbrace{\frac{1}{2} \sum_{\sigma} \sigma c_{\mathbf{k}\sigma}^{\dagger} c_{\mathbf{k}'\sigma}}_{S_z(\mathbf{k}, \mathbf{k}')} = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\mathbf{k}'} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{T}_i} S_z(\mathbf{k}, \mathbf{k}').$$

The term H_U^{HF} has the same periodicity as the lattice and does not couple states with different $\mathbf{k}$ vectors. Thus only $S_z(\mathbf{k}, \mathbf{k})$ contributes, and the Hamiltonian can be written as

$$H = \sum_{\sigma} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + U \underbrace{\sum_{\mathbf{k}} \left[-2m S_z(\mathbf{k}, \mathbf{k}) + m^2 + \frac{n^2}{4} \right]}_{H_U^{\text{HF}} = U \sum_i [-2m S_z^i + m^2 + \frac{1}{4}n^2]},$$

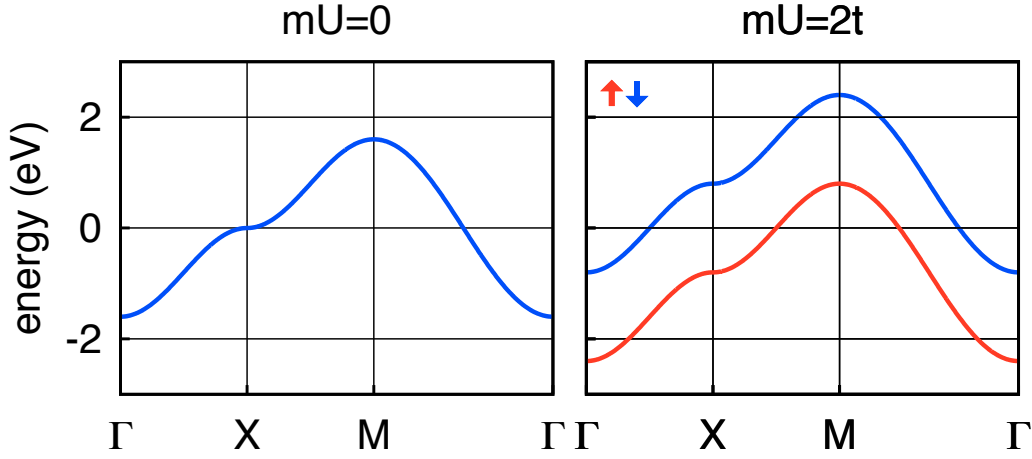


Fig. 11: *Ferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero.*

where $m = (n_{\uparrow} - n_{\downarrow})/2$ and $n = 1$; for simplicity we set $\varepsilon_d = 0$. The HF correction splits the bands with opposite spin, leading to new one-electron eigenvalues, $\varepsilon_{k\sigma} = \varepsilon_{\mathbf{k}} + \frac{1}{2}U - \sigma U m$; the chemical potential is $\mu = U/2$. The separation between $\varepsilon_{k\uparrow} - \mu$ and $\varepsilon_{k\downarrow} - \mu$ is $2mU$, as can be seen in Fig. 11. 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

$$E_F = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} [\varepsilon_{k\sigma} - \mu] = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left[\varepsilon_{\mathbf{k}} - \frac{1}{2}U \right] = -\frac{1}{2}U.$$

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. 6. We rewrite the Bloch states of the original lattice as

$$\Psi_{k\sigma}(\mathbf{r}) = \frac{1}{\sqrt{2}} [\Psi_{k\sigma}^A(\mathbf{r}) + \Psi_{k\sigma}^B(\mathbf{r})], \quad \Psi_{k\sigma}^{\alpha}(\mathbf{r}) = \frac{1}{\sqrt{N_{s_{\alpha}}}} \sum_{i_{\alpha}} e^{i\mathbf{T}_{i_{\alpha}}^{\alpha} \cdot \mathbf{k}} \Psi_{i_{\alpha}\sigma}(\mathbf{r}).$$

Here A and B are the two sublattices with opposite spins and $\mathbf{T}_i^A$ and $\mathbf{T}_i^B$ are their lattice vectors; $\alpha = A, B$. We take as one-electron basis the two Bloch functions $\Psi_{k\sigma}$ and $\Psi_{\mathbf{k}+\mathbf{Q}_2\sigma}$, where $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$ is the vector associated with the antiferromagnetic instability and the corresponding folding of the Brillouin zone, also shown in Fig. 6. Then, in the HF approximation, the Coulomb interaction is given by

$$H_U^{\text{HF}} = \sum_{i \in A} \left[-2m S_z^i + m^2 + \frac{n^2}{4} \right] + \sum_{i \in B} \left[+2m S_z^i + m^2 + \frac{n^2}{4} \right].$$

This interaction couples Bloch states with $\mathbf{k}$ vectors made equivalent by the folding of the Brillouin zone. Thus the HF Hamiltonian takes the form

$$H = \sum_{\mathbf{k}} \sum_{\sigma} \varepsilon_{\mathbf{k}} n_{k\sigma} + \sum_{\mathbf{k}} \sum_{\sigma} \varepsilon_{\mathbf{k}+\mathbf{Q}_2} n_{\mathbf{k}+\mathbf{Q}_2\sigma} + U \underbrace{\sum_{\mathbf{k}} \left[-2m S_z(\mathbf{k}, \mathbf{k} + \mathbf{Q}_2) + 2m^2 + 2\frac{n^2}{4} \right]}_{\text{static mean-field correction } H_U^{\text{HF}}}.$$

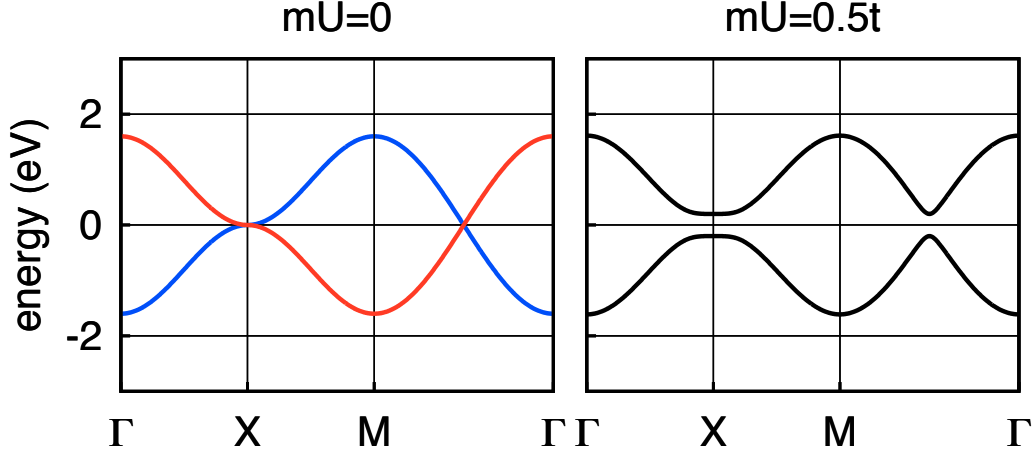


Fig. 12: Antiferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero. Blue: $\varepsilon_{\mathbf{k}}$. Red: $\varepsilon_{\mathbf{k}+\mathbf{Q}_2} = -\varepsilon_{\mathbf{k}}$. The high-symmetry lines are those of the large BZ in Fig. 6.

The sum over $\mathbf{k}$ is restricted to the Brillouin zone of the antiferromagnetic lattice. We find the two-fold degenerate eigenvalues

$$\varepsilon_{\mathbf{k}\pm} - \mu = \frac{1}{2}(\varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}+\mathbf{Q}_2}) \pm \frac{1}{2}\sqrt{(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{Q}_2})^2 + 4(mU)^2}. \quad (23)$$

A gap opens where the bands $\varepsilon_{\mathbf{k}}$ and $\varepsilon_{\mathbf{k}+\mathbf{Q}_2}$ cross, e.g., at the X point of the original Brillouin zone (Fig. 12). At half filling and for $mU = 0$ the Fermi level crosses the bands at the X point too; thus the system is 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 $\varepsilon_{\mathbf{k}}/U$. For the occupied states we find

$$\varepsilon_{\mathbf{k}-} - \mu \sim -\frac{1}{2}U - \frac{\varepsilon_{\mathbf{k}}^2}{U} = -\frac{1}{2}U - \frac{4t^2}{U} \left(\frac{\varepsilon_{\mathbf{k}}}{2t}\right)^2.$$

The ground-state total energy for the antiferromagnetic supercell is then $2E_{\text{AF}}$ with

$$E_{\text{AF}} = -\frac{1}{2}U - \frac{4t^2}{U} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left(\frac{\varepsilon_{\mathbf{k}}}{2t}\right)^2 \sim -\frac{1}{2}U - \frac{4t^2}{U}$$

so that the energy difference per pair of spins between ferro- and antiferro-magnetic state is

$$\Delta E^{\text{HF}} = E_{\uparrow\uparrow}^{\text{HF}} - E_{\uparrow\downarrow}^{\text{HF}} = \frac{2}{n_{\langle ii' \rangle}} [E_{\text{F}} - E_{\text{AF}}] \sim \frac{1}{2} \frac{4t^2}{U} \sim \frac{1}{2} \Gamma, \quad (24)$$

which is similar to the result obtained from the Hubbard model in many-body second order perturbation theory, Eq. (12). Despite the similarity with the actual solution, one has to remember that the spectrum of the Hartree-Fock Hamiltonian has very little to do with the spectrum of the Heisenberg model, the model that describes the actual low-energy behavior of the Hubbard Hamiltonian. If we restrict ourselves to the antiferromagnetic solution, the first excited state 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 case, however, one has to keep in mind that HF suffers from *spin contamination*, i.e., singlet states and $S_z = 0$ triplet states mix [26]. The energy difference per bond $E_{\uparrow\uparrow}^{\text{HF}} - E_{\uparrow\downarrow}^{\text{HF}}$ in Eq. (24) only resembles the exact result, as one can grasp by comparing it with the actual energy difference between triplet and singlet state in the two-site Heisenberg model

$$\Delta E = E_{S=1} - E_{S=0} = \Gamma,$$

which is a factor of two larger. The actual ratio $\Delta E / \Delta E^{\text{HF}}$ might depend on the details of the HF band structures. Thus, overall, Hartree-Fock is not the ideal approach to determine the onset of magnetic phase transitions. Other shortcomings of the Hartree-Fock approximation are in the description of the Mott metal-insulator transition. In Hartree-Fock the metal-insulator transition is intimately related to long-range magnetic order (Slater transition), but in strongly correlated materials the metal-insulator transition can occur in the paramagnetic phase (Mott transition). It is associated with a divergence of the self-energy at low frequencies rather than with the formation of superstructures. This physics, captured by many-body methods such as the dynamical mean-field theory [15], is completely missed by the Hartree-Fock approximation.

2.3.3 The dynamical mean-field theory approach

The modern approach for solving the Hubbard model is the so-called dynamical mean-field theory method [14–16]. In DMFT the lattice Hubbard model is mapped onto a self-consistent quantum impurity model describing an impurity coupled to a non-correlated conduction-electron bath. The quantum impurity model is typically the Anderson model, which will be discussed in detail in the next chapter. Here we do not want to focus on the specific form of the quantum impurity model but rather on the core aspects of the DMFT approach and on the comparison of DMFT with Hartree-Fock. In Hartree-Fock the effective mean field is an energy-independent (static) parameter; in the example discussed in the previous section it is a function of the magnetic order parameter m . In DMFT the role of the effective mean-field is played by the bath Green function $G_0(i\nu_n)$ where ν_n is a fermionic Matsubara frequency; it is frequency dependent (dynamical) and related to the impurity Green function $G(i\nu_n)$ via the Dyson equation

$$[G(i\nu_n)]^{-1} = [G_0(i\nu_n)]^{-1} - \Sigma(i\nu_n), \quad (25)$$

where $\Sigma(i\nu_n)$ is the impurity self-energy. As in any mean-field theory, the effective field is determined by enforcing a self-consistency condition. In DMFT the latter requires that the impurity Green function $G(i\nu_n)$, calculated by solving the quantum impurity model, equals $G_{ii}(i\nu_n)$, the lattice Green function at a site i

$$G_{ii}(i\nu_n) = G(i\nu_n),$$

with

$$G_{ii}(i\nu_n) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} G(\mathbf{k}; i\nu_n) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \frac{1}{i\nu_n - \varepsilon_{\mathbf{k}} - \Sigma(i\nu_n) + \mu}.$$

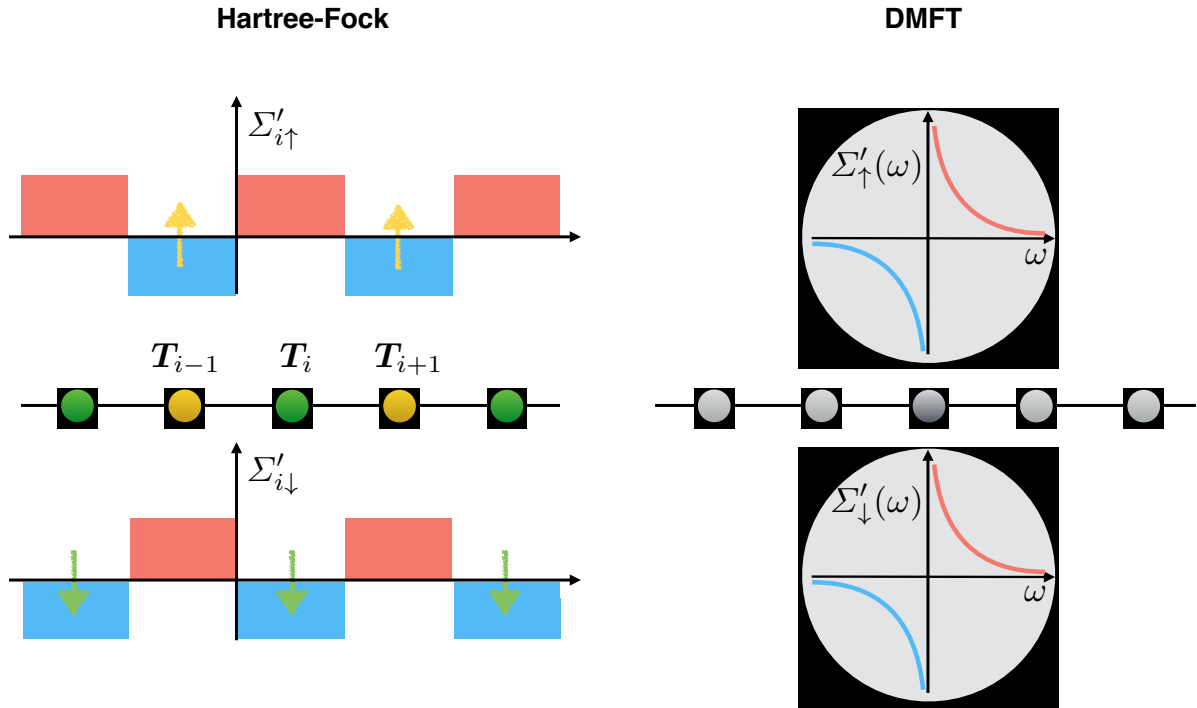


Fig. 13: *Idealized correlated crystal, schematically represented by a half-filled single-band Hubbard chain. Sketch of the real-part of the self-energy in the insulating phase, as described by Hartree-Fock (left-hand side) and DMFT (right-hand side). In HF the self-energy is a spin- and site-dependent 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).*

The Green function on the real axis can be obtained from $G(i\nu_n)$ via analytic continuation; in the non-interacting case, this can be done simply by replacing $i\nu_n$ with $\omega + i0^+$. The self-energy in Eq. (25) is frequency dependent but local (i.e., site- or $\mathbf{k}$ -independent); the locality of the self-energy is, of course, an approximation; it becomes an exact property, however, in the limit of infinite coordination number. DMFT yields the exact result in two opposite limits, $t = 0$ (atomic limit) and $U = 0$ (band limit). The first success of DMFT was the description of the paramagnetic Mott metal-insulator transition in the half-filled one-band Hubbard model. It is interesting to examine in more detail the nature of the Mott transition in DMFT and compare it to the Slater transition described by Hartree-Fock. Let us start with analyzing DMFT results. The poles of the Green function, i.e., the solutions of

$$\omega - \varepsilon_{\mathbf{k}} - \Sigma'(\omega) = 0,$$

where $\Sigma'(\omega)$ is the real part of the self-energy, yield the excitations of our system. In the Fermi liquid regime, the Green function has a pole at zero frequency. Around it, the self-energy has, on the real axis, the following form

$$\Sigma(\omega) \sim \frac{1}{2}U + \left(1 - \frac{1}{Z}\right)\omega - \frac{i}{Z\tau^{\text{QP}}},$$

where the positive dimensionless number Z yields the mass enhancement, $m^*/m \sim 1/Z$, and the positive parameter $\tau^{\text{QP}} \sim 1/(aT^2 + b\omega^2)$ is the quasiparticles lifetime; at higher frequency the self-energy yields two additional poles corresponding to the Hubbard bands. In the Mott insulating regime the central quasiparticle peak disappears, and only the Hubbard bands remain. The self-energy has approximately the form

$$\Sigma(\omega) \sim \frac{rU^2}{4} \left[\frac{1}{\omega} - i\pi\delta(\omega) - if_U(\omega) \right],$$

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 the spin

$$\begin{aligned}\Sigma_\sigma(\omega) &= \Sigma(\omega) \\ G_\sigma(\omega) &= G(\omega) \\ G_\sigma(\mathbf{k}; \omega) &= G(\mathbf{k}; \omega).\end{aligned}$$

Thus DMFT can be seen, to *some* extent, as a complementary approximation to Hartree-Fock. If we write the Hartree-Fock correction to the energies in the form of a self-energy, the latter is a real, static but spin- and site-dependent potential. More specifically, we have at site i

$$\Sigma_{i\sigma}^{\text{HF}}(\omega) = U \left[n_{-\sigma}^i - \frac{1}{2} \right],$$

where n_σ^i is the site occupation for spin σ . Let us consider the antiferromagnetic case. For this, as we have seen, we have to consider two sublattices or a two-site cluster; the magnetization at sites j , nearest neighbors of site i , has opposite sign than at site i . Thus

$$\Sigma_{j\sigma}^{\text{HF}}(\omega) = -U \left[n_{-\sigma}^i - \frac{1}{2} \right].$$

This spatial structure of the self-energy is what opens the gap shown in Fig. 12; this picture of the gap opening is very different from the one emerging from DMFT; as we have just seen, 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. In HF the self-energy resulting from a non-magnetic ($m = 0$) single-site calculation is instead a mere energy shift – the same for all sites and spins – and does not change the band structure at all. Hartree-Fock is not, e.g., the large- U limit of DMFT but merely the large frequency limit of DMFT (or of its cluster extensions). The differences among the two approaches are pictorially shown in Fig. 13 for an idealized one-dimensional crystal.

Let us now focus on the magnetic properties of the Hubbard model, and in particular on the magnetic susceptibility $\chi_{zz}(\mathbf{q}; i\omega_m)$. Since in DMFT we solve the quantum impurity model exactly, we can directly calculate the local linear-response tensor, and therefore the local susceptibility $\chi_{zz}(i\omega_m)$. If we are interested in magnetic order, however, we need also, at least

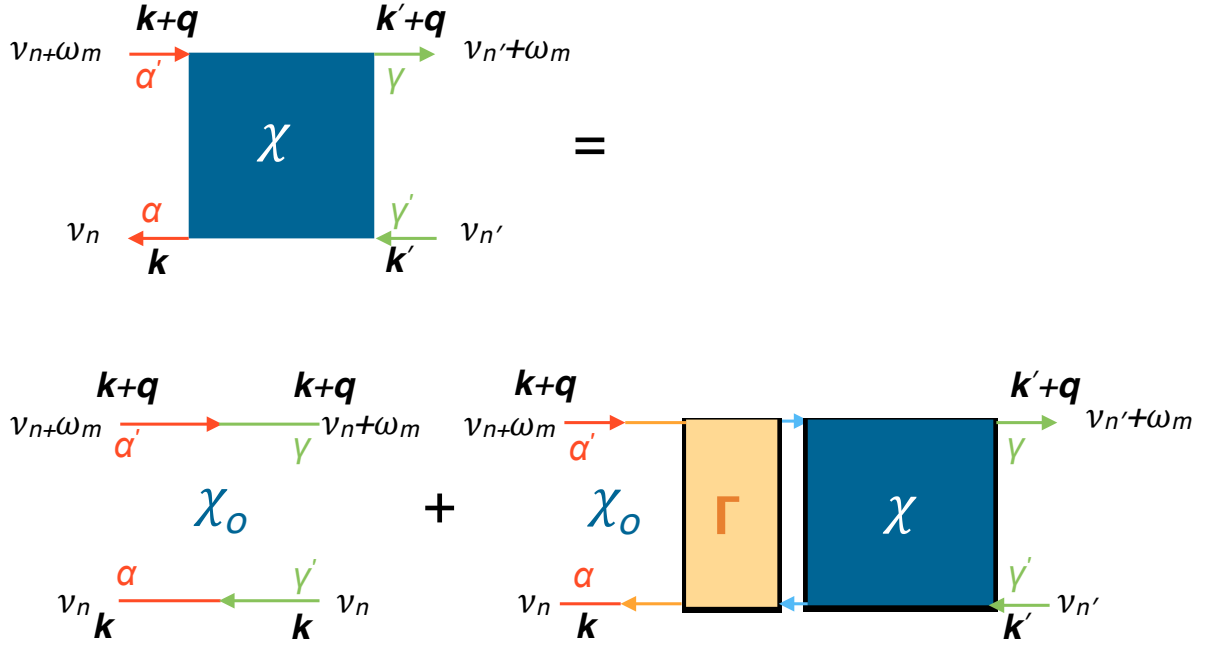


Fig. 14: Diagrammatic representation of the Bethe-Salpeter equation. In the case of the magnetic susceptibility, $\alpha = \alpha' = \sigma$ and $\gamma = \gamma' = \sigma'$.

in an approximate form, the full $\mathbf{q}$ -dependent linear-response function, $\chi_{zz}(\mathbf{q}; i\omega_m)$. How can we obtain it? It is tempting to think that $\chi_{zz}(\mathbf{q}; i\omega_m)$ can be approximated by the bare DMFT susceptibility, $\chi_{zz}^0(\mathbf{q}; i\omega_m)$. In the paramagnetic regime, for the one-band Hubbard model this is given by (see Appendix)

$$\chi_{zz}^0(\mathbf{q}; i\omega_m) = -\frac{(g\mu_B)^2}{4} \frac{1}{\beta N_{\mathbf{k}}} \sum_{\sigma \mathbf{k} n} G_{\sigma}(\mathbf{k}; i\nu_n) G_{\sigma}(\mathbf{k} + \mathbf{q}; i\nu_n + i\omega_m), \quad (26)$$

where ω_m is a bosonic Matsubara frequency and $G_{\sigma}(\mathbf{k}; i\nu_n)$ is the single-particle Green function for spin σ . Since $\chi_{zz}^0(\mathbf{q}; i\omega_m)$ depends only on the single-particle Green function it can be extracted from DMFT calculations with little additional effort. To approximate the actual susceptibility with $\chi_{zz}^0(\mathbf{q}; i\omega_m)$ would be, however, totally incorrect. Indeed, while $\chi_{zz}^0(\mathbf{q}; i\omega_m)$ is exact in the non-interacting limit (i.e., it correctly yields the Pauli susceptibility for $U = 0$), it is incorrect in the atomic limit ($t = 0$) and hence in the whole local-moment regime. Let us see what happens in the atomic limit. By summing Eq. (26) over $\mathbf{q}$, we obtain the local $\chi_{zz}^0(i\omega_m)$, proportional to the sum of products of local Green functions,

$$\chi_{zz}^0(i\omega_m) = -\frac{(g\mu_B)^2}{4} \frac{1}{\beta} \sum_{n\sigma} G_{\sigma}(i\nu_n) G_{\sigma}(i\nu_n + i\omega_m).$$

If we replace the local Green functions with the corresponding atomic Green functions (see Appendix) and then perform the Matsubara sums we find the expression

$$\chi_{zz}^0(0) = \frac{(g\mu_B)^2}{4} \frac{\beta e^{\beta U/2}}{1 + e^{\beta U/2}} \left[\frac{1}{1 + e^{\beta U/2}} + \frac{1}{U\beta} \frac{1 - e^{-\beta U}}{1 + e^{-\beta U/2}} \right] \underset{\beta U \rightarrow \infty}{\sim} \frac{(g\mu_B S)^2}{U},$$

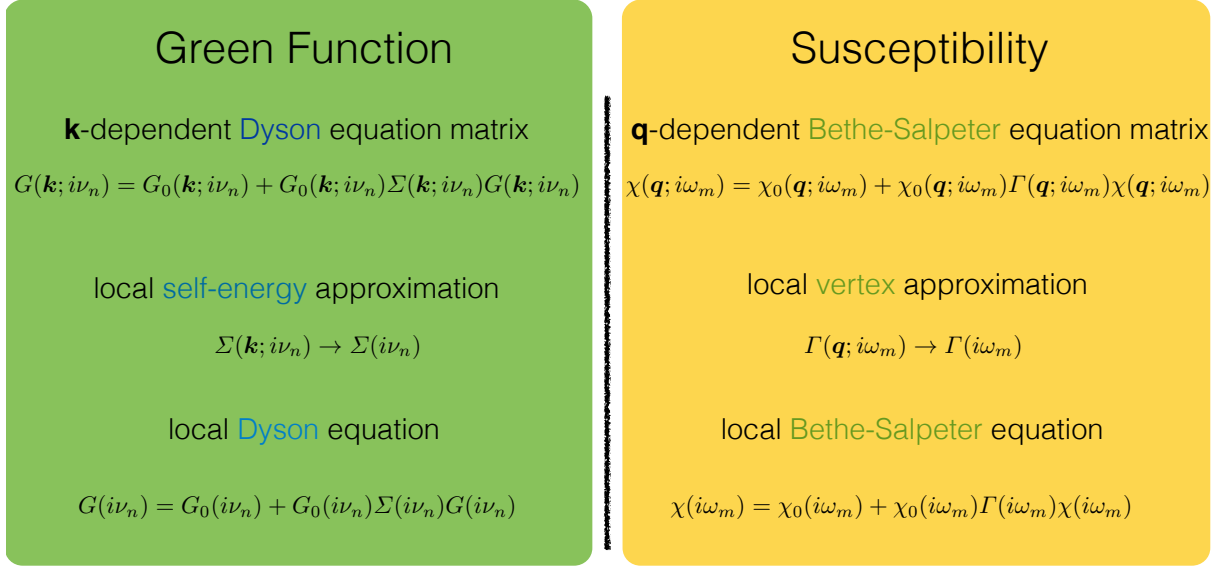


Fig. 15: Analogies between the Green function $G(\mathbf{k}; i\nu_n)$ in the local-self-energy approximation (left) and the response function $\chi(\mathbf{q}; i\omega_m)$ in the local vertex approximation (right). Each term in the Bethe-Salpeter equation is a square matrix of dimension $L_\alpha = N_\alpha N_{\mathbf{k}} N_n$, where $N_{\mathbf{k}}$ is the number of $\mathbf{k}$ points, N_n the number of fermionic Matsubara frequencies, N_α the number of flavors (here: the spin degrees of freedom). The elements of, e.g., the matrix $\chi(\mathbf{q}; i\omega_m)$ can be written as $[\chi(\mathbf{q}; i\omega_m)]_{\sigma\mathbf{k}\nu_n, \sigma'\mathbf{k}'\nu'_n}$.

where the exact result is

$$\chi_{zz}(0) = \frac{1}{4}(g\mu_B)^2 \frac{\beta e^{\beta U/2}}{1 + e^{\beta U/2}} \underset{\beta U \rightarrow \infty}{\sim} \frac{(g\mu_B S)^2}{k_B T}.$$

Since $\chi_{zz}^0(0)$ is incorrect in the large βU limit, it also does not interpolate properly between the two regimes, weak and strong coupling. Let us then take a step back. The local susceptibility $\chi_{zz}(i\omega_m)$ can be obtained, as we have already mentioned, by solving exactly the self-consistent quantum impurity model, e.g., via quantum Monte Carlo. In a similar way, the $\mathbf{q}$ -dependent response function $\chi_{zz}(\mathbf{q}; \omega)$ could in principle be calculated from the solution of the full lattice Hubbard model; the problem is that, unfortunately, the exact solution of the Hubbard model is not available in the general case. If many-body perturbation theory converges, however, we can calculate $\chi_{zz}(\mathbf{q}; \omega)$ by solving the Bethe-Salpeter equation – the analogon of the Dyson equation for the Green function – with $\chi_{zz}^0(\mathbf{q}; i\omega_n)$ as bare susceptibility. The Bethe-Salpeter equation is represented diagrammatically in Fig. 14. By summing up all diagrams in the series we find

$$[\chi(\mathbf{q}; i\omega_n)]^{-1} = [\chi_0(\mathbf{q}; i\omega_n)]^{-1} - \Gamma(\mathbf{q}; i\omega_n). \quad (27)$$

Here each term is a square matrix with dimension $L_\alpha = N_\alpha N_{\mathbf{k}} N_n$, where N_α is the number of flavors (here, the spin degrees of freedom), $N_{\mathbf{k}}$ is the number of $\mathbf{k}$ points and N_n the number of fermionic Matsubara frequencies. The quantity $\Gamma(\mathbf{q}; i\omega_n)$ is the vertex function and hides all diagrams appearing in the many-body perturbation series. Finding the exact $\Gamma(\mathbf{q}; i\omega_n)$ is of

course as difficult as solving the full many-body problem; we therefore have to find a reasonable approximation. In the spirit of DMFT, let us assume that the vertex entering the Bethe-Salpeter equation can be replaced by a local function

$$\Gamma(\mathbf{q}; i\omega_n) \rightarrow \Gamma(i\omega_n).$$

Furthermore, let us assume that the local vertex solves, in turn, a local version of the Bethe-Salpeter equation

$$\Gamma(i\omega_m) = [\chi_0(i\omega_m)]^{-1} - [\chi(i\omega_m)]^{-1}. \quad (28)$$

The local vertex $\Gamma(i\omega_m)$, calculated via Eq. (28), can then be used to compute the susceptibility from the $\mathbf{q}$ -dependent Bethe-Salpeter equation Eq. (27). The analogy between the calculation of the susceptibility in the local vertex approximation and that of the Green function in the local self-energy approximation is shown in Fig. 15.

Let us now qualitatively discuss the magnetic susceptibility of the one-band Hubbard model in the Mott-insulating limit. For simplicity, let us now assume that the vertex is static and thus that we can replace all susceptibility matrices in the Bethe-Salpeter equation with the physical susceptibilities, which we obtain by summing over the fermionic Matsubara frequencies and the momenta. For the magnetic susceptibility this means

$$\begin{aligned} \chi_{zz}(\mathbf{q}; i\omega_m) &= (g\mu_B)^2 \frac{1}{\beta^2} \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' \sum_{nn'} \frac{1}{N_{\mathbf{k}}^2} \sum_{\mathbf{k}\mathbf{k}'} \frac{1}{\beta} [\chi(\mathbf{q}; i\omega_m)]_{\sigma\mathbf{k}\nu_n, \sigma'\mathbf{k}'\nu_{n'}} \\ &= (g\mu_B)^2 \int d\tau e^{i\omega_m\tau} \langle S_z(\tau) S_z(0) \rangle. \end{aligned}$$

In the high-temperature and large- U limit ($\beta U \rightarrow \infty$ and $T \gg T_N$) the static local susceptibility is approximatively given by the static atomic susceptibility

$$\chi_{zz}(0) \sim \frac{\mu_{\text{eff}}^2}{k_B T},$$

where $\mu_{\text{eff}}^2 = (g\mu_B)^2 S(S+1)/3$. Therefore the vertex is roughly given by

$$\Gamma(0) \sim \frac{1}{\chi_{zz}^0(0)} - \frac{k_B T}{\mu_{\text{eff}}^2}.$$

The susceptibility calculated with such a local vertex is

$$\chi_{zz}(\mathbf{q}; 0) \sim \frac{\mu_{\text{eff}}^2}{k_B T - \mu_{\text{eff}}^2 J(\mathbf{q})} \equiv \frac{\mu_{\text{eff}}^2}{k_B (T - T_q)},$$

where the coupling is given by

$$J(\mathbf{q}) = - \left[\frac{1}{\chi_{zz}^0(\mathbf{q}; 0)} - \frac{1}{\chi_{zz}^0(0)} \right].$$

Thus, in the strong-correlation limit, the Bethe-Salpeter equation, solved assuming the vertex is local and, in addition, static, yields a static high-temperature susceptibility of Curie-Weiss form. We had found the same form of the susceptibility by solving the Heisenberg model in the static mean-field approximation. A more detailed presentation of the DMFT approach to calculate linear-response functions can be found in Ref. [27].

3 The Anderson model

The Kondo impurity is a representative case of a system that exhibits both local-moment and Pauli-paramagnetic behavior, although in quite different temperature regimes [12]. The Kondo effect was first observed in diluted metallic alloys, metallic systems in which isolated d or f magnetic impurities are present, and it has been a riddle for decades. A Kondo impurity in a metallic host can be described by the Anderson model

$$H_A = \underbrace{\sum_{\sigma} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{\sigma} \varepsilon_f n_{f\sigma} + U n_{f\uparrow} n_{f\downarrow}}_{H_0} + \underbrace{\sum_{\sigma} \sum_{\mathbf{k}} \left[V_{\mathbf{k}} c_{\mathbf{k}\sigma}^{\dagger} c_{f\sigma} + h.c. \right]}_{H_1}, \quad (29)$$

where ε_f is the impurity level (occupied by $n_f \sim 1$ electrons), $\varepsilon_{\mathbf{k}}$ is the dispersion of the metallic band, and $V_{\mathbf{k}}$ 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

$$\Delta(\varepsilon) = \pi \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} |V_{\mathbf{k}}|^2 \delta(\varepsilon_{\mathbf{k}} - \varepsilon)$$

is such that $\Delta(\mu) \ll |\mu - \varepsilon_f|, |\mu - \varepsilon_f - U|$. The Anderson model is also used as the quantum impurity problem in dynamical mean-field theory. In DMFT the bath parameters $\varepsilon_{\mathbf{k}}$ and $V_{\mathbf{k}}$ have, in principle, to be determined self-consistently. If quantum Monte Carlo is used to solve the Anderson model, it is sufficient to determine the bath Green function self-consistently.

3.1 The Kondo limit

Through the Schrieffer-Wolff canonical transformation [28] one can map the Anderson model onto the Kondo model, in which only the effective spin of the impurity enters

$$H_K = H'_0 + \Gamma \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) = H'_0 + H_{\Gamma}, \quad (30)$$

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, $\mathbf{S}_f$ the spin of the impurity ($S_f = 1/2$), and $\mathbf{s}_c(\mathbf{0})$ 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 $\mathbf{k}$ vector at the Fermi level. The Schrieffer-Wolff canonical transformation works as follows. We introduce the operator S that transforms the Hamiltonian H into H_S

$$H_S = e^S H e^{-S}.$$

We search for an operator S such that the transformed Hamiltonian H_S has no terms of first order in $V_{\mathbf{k}}$. Let us first split the original Hamiltonian H_A into two pieces: H_0 , the sum of all terms except the hybridization term, and H_1 , the hybridization term. Let us choose S linear in $V_{\mathbf{k}}$ and such that

$$[S, H_0] = -H_1. \quad (31)$$

From Eq. (31) one finds that the operator S is given by

$$S = \sum_{\mathbf{k}\sigma} \left[\frac{1 - n_{f-\sigma}}{\varepsilon_{\mathbf{k}} - \varepsilon_f} + \frac{n_{f-\sigma}}{\varepsilon_{\mathbf{k}} - \varepsilon_f - U} \right] V_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{f\sigma} - \text{h.c.}.$$

The transformed Hamiltonian is complicated, as can be seen from explicitly writing the series for a transformation satisfying Eq. (31)

$$H_S = H_0 + \frac{1}{2} [S, H_1] + \frac{1}{3} [S, [S, H_1]] + \dots$$

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 $H_S = H_0 + H_2$, with $H_2 = H_\Gamma + H_{\text{dir}} + \Delta H_0 + H_{\text{ch}}$. The first term is the exchange interaction

$$H_\Gamma = \frac{1}{4} \sum_{\mathbf{k}\mathbf{k}'} \Gamma_{\mathbf{k}\mathbf{k}'} \left[\sum_{\sigma_1\sigma_2} c_{\mathbf{k}'\sigma_1}^\dagger \langle \sigma_1 | \hat{\sigma} | \sigma_2 \rangle c_{\mathbf{k}\sigma_2} \cdot \sum_{\sigma_3\sigma_4} c_{f\sigma_3}^\dagger \langle \sigma_3 | \hat{\sigma} | \sigma_4 \rangle c_{f\sigma_4} \right]$$

where

$$\Gamma_{\mathbf{k}\mathbf{k}'} = V_{\mathbf{k}}^* V_{\mathbf{k}'} \left[\frac{1}{\varepsilon_{\mathbf{k}} - \varepsilon_f} + \frac{1}{\varepsilon_{\mathbf{k}'} - \varepsilon_f} + \frac{1}{U + \varepsilon_f - \varepsilon_{\mathbf{k}}} + \frac{1}{U + \varepsilon_f - \varepsilon_{\mathbf{k}'}} \right].$$

Let us assume that the coupling $\Gamma_{\mathbf{k}\mathbf{k}'}$ is weakly dependent on $\mathbf{k}$ and $\mathbf{k}'$; then by setting $|\mathbf{k}| \sim k_F$, and $|\mathbf{k}'| \sim k_F$ we recover the antiferromagnetic contact coupling in Eq. (30).

The second term is a potential-scattering interaction

$$H_{\text{dir}} = \sum_{\mathbf{k}\mathbf{k}'} \left[A_{\mathbf{k}\mathbf{k}'} - \frac{1}{4} \Gamma_{\mathbf{k}\mathbf{k}'} \hat{n}_f \right] \sum_{\sigma} \hat{c}_{\mathbf{k}'\sigma}^\dagger \hat{c}_{\mathbf{k}\sigma},$$

where

$$A_{\mathbf{k}\mathbf{k}'} = \frac{1}{2} V_{\mathbf{k}}^* V_{\mathbf{k}'} \left[\frac{1}{\varepsilon_{\mathbf{k}} - \varepsilon_f} + \frac{1}{\varepsilon_{\mathbf{k}'} - \varepsilon_f} \right].$$

This term is spin-independent, and thus does not play a relevant role in the Kondo effect. The next term merely modifies the H_0 term

$$\Delta H_0 = - \sum_{\mathbf{k}\sigma} \left[A_{\mathbf{k}\mathbf{k}} - \frac{1}{2} \Gamma_{\mathbf{k}\mathbf{k}} \hat{n}_{f-\sigma} \right] \hat{n}_{f\sigma}.$$

Finally, the last term is a pair-hopping interaction, which changes the charge of the f sites by two electrons and thus can be neglected if $n_f \sim 1$

$$\Delta H_{\text{ch}} = -\frac{1}{4} \sum_{\mathbf{k}\mathbf{k}'\sigma} \Gamma_{\mathbf{k}\mathbf{k}'} c_{\mathbf{k}'-\sigma}^\dagger c_{\mathbf{k}\sigma}^\dagger c_{f\sigma} c_{f-\sigma} + \text{h.c.}.$$

The essential term in H_2 is the exchange term H_Γ , which is the one that yields the antiferromagnetic contact interaction in the Kondo Hamiltonian (30). Remarkably, the Schrieffer-Wolff transformation generates a perturbation series in the hybridization; an analogous perturbation series is also used in the hybridization-expansion continuous-time quantum Monte Carlo approach to solve the quantum impurity problem in dynamical mean-field theory.

3.1.1 The impurity susceptibility

The solution of the problem defined by (29) or (30) is not at all trivial and requires many-body techniques such as the Wilson numerical renormalization group [30] or the Bethe ansatz [31]. Here we only discuss some important exact results concerning the magnetic properties. Let us define the *impurity susceptibility* $\chi_{zz}^f(T)$ as the total susceptibility minus the contribution of the metallic band in the absence of the impurity [30–32]. One can show that at high temperatures $\chi_{zz}^f(T)$ has the following behavior

$$\chi_{zz}^f(T) \sim \frac{(g\mu_B)^2 S_f(S_f + 1)}{3k_B T} \left\{ 1 - \frac{1}{\ln(T/T_K)} \right\}.$$

This expression resembles the Curie susceptibility, apart from the $\ln(T/T_K)$ term. The scale T_K is the Kondo temperature, which, to first approximation, is given by

$$k_B T_K \sim D e^{-2/\rho(\varepsilon_F)\Gamma},$$

where $2D = W$ is the band width of the host conduction band. Because of the $\ln(T/T_K)$ term, the susceptibility apparently diverges at $T \sim T_K$. In reality, however, around T_K there is a crossover to a new regime. For $T \ll T_K$

$$\chi_{zz}^f(T) \sim \frac{C_{1/2}}{\mathcal{W} T_K} \{ 1 - \alpha T^2 + \dots \},$$

where $\mathcal{W}$ is a (universal) Wilson number. Thus the low-temperature system has a Fermi liquid behavior with enhanced density of states, i.e., with heavy masses m^*/m ; furthermore $\chi_{zz}^f(0) = C_{1/2}/\mathcal{W} T_K$ is the Curie susceptibility (Eq. (15)) with the temperature *frozen* at $T = \mathcal{W} T_K$. At $T = 0$ the impurity magnetic moment is screened by the conduction electrons, which form a singlet state with the spin of the impurity. In other words, the effective magnetic moment formed by the impurity magnetic moment and its screening cloud,

$$\mu_{\text{eff}}^2(T) \equiv 3k_B T \chi_{zz}^f(T) \propto \langle S_z^f S_z^f \rangle + \langle S_z^f s_z^c \rangle,$$

vanishes for $T \ll T_K$. The Kondo temperature is typically 10–30 K or even smaller, hence the Fermi liquid behavior is restricted to a very narrow energy and temperature region.

3.1.2 Poor man's scaling

We can understand the existence of a Fermi liquid regime by using a simple approach due to Anderson called *poor man's scaling* [33] 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 ε_q is $D' < \varepsilon_q < D$ or $-D < \varepsilon_q < -D'$, where $D' = D - \delta D$. Next we introduce the operator P_H , which projects onto the high-energy states, and the operator $P_L = 1 - P_H$, 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, $H \equiv H'_0 + H_\Gamma$, as the energy-dependent operator H' , which acts only in the low-energy sector

$$H' = P_L H P_L + \delta H_L = H_L + \delta H_L,$$

$$\delta H_L = P_L H P_H \left(\omega - P_H H P_H \right)^{-1} P_H H P_L.$$

Here H_L is the original Hamiltonian, however in the space in which the high-energy states have been eliminated; the term δH_L is a correction due to the interaction between low and (downfolded) high-energy states. Up to this point, the operator H' has the same spectrum of the original Hamiltonian. To make use of this expression, however, we have to introduce approximations. Thus, let us calculate δH_L using many-body perturbation theory. The first non-zero contribution is of second order in Γ

$$\delta H_L^{(2)} \sim P_L H_\Gamma P_H \left(\omega - P_H H'_0 P_H \right)^{-1} P_H H_\Gamma P_L.$$

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

$$P_H \sim \sum_{q\sigma} c_{q\sigma}^\dagger |FS\rangle \langle FS| c_{q\sigma}, \quad P_L \sim \sum_{k\sigma} c_{k\sigma}^\dagger |FS\rangle \langle FS| c_{k\sigma},$$

where $|\varepsilon_k| < D'$ and $|FS\rangle = \prod_{k\sigma} c_{k\sigma}^\dagger |0\rangle$ is the Fermi sea, i.e., the many-body state corresponding to the metallic conduction band. Thus

$$\delta H_L^{(2)} = -\frac{1}{2} \Gamma^2 \sum_q \frac{1}{\omega - \varepsilon_q} \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) + \dots \sim \frac{1}{4} \rho(\varepsilon_F) \Gamma^2 \frac{\delta D}{D} \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) + \dots$$

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

$$\Gamma \rightarrow \Gamma' = \Gamma + \delta\Gamma,$$

$$\frac{\delta\Gamma}{\delta \ln D} = \frac{1}{2} \rho(\varepsilon_F) \Gamma^2, \quad (32)$$

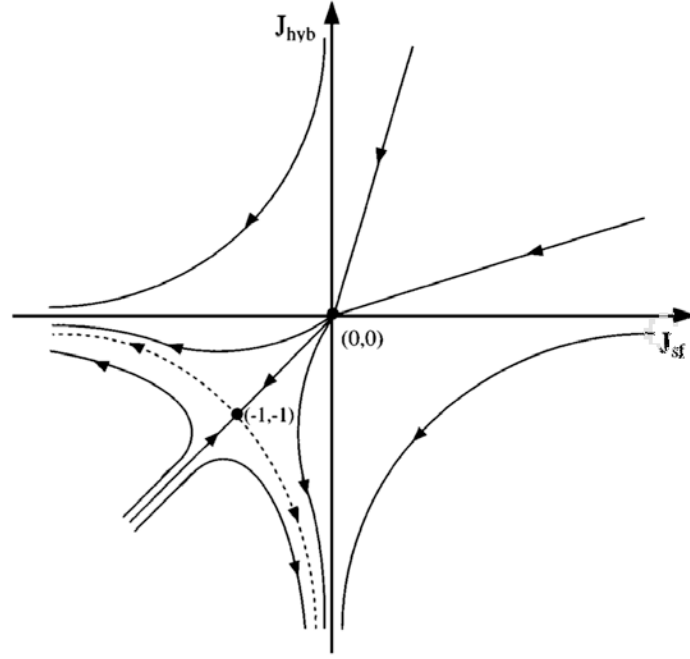


Fig. 16: Sketch of the scaling diagrams for the two-channel Kondo model. $\Gamma = -J_{\text{hyb}}$ and $\Gamma_F = -J_{\text{sf}}$. For $\Gamma > 0$ (antiferromagnetic) and $\Gamma_F < 0$ (ferromagnetic), the antiferromagnetic coupling scales to strong coupling and ferromagnetic one to weak coupling (right bottom quadrant). From Ref. [34].

where $\delta \ln D = \delta D/D$. Equation (32) has two fixed points, $\Gamma = 0$ (weak coupling) and $\Gamma \rightarrow \infty$ (strong coupling). By solving the scaling equations we find

$$\Gamma' = \frac{\Gamma}{1 + \frac{1}{2}\rho(\varepsilon_F)\Gamma \ln \frac{D'}{D}}.$$

If Γ is antiferromagnetic the renormalized coupling constant Γ' diverges for $D' = De^{-2/\Gamma\rho(\varepsilon_F)}$, an energy proportional to the Kondo energy $k_B T_K$. This divergence (scaling to strong coupling) 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 [30]. Nozières [35] 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 [35].

If $\Gamma = \Gamma_F < 0$ (ferromagnetic coupling, as for example the coupling arising from direct Coulomb exchange) the renormalized coupling constant Γ' goes to zero in the $D' \rightarrow 0$ limit (scaling to weak coupling). This means that the local spin becomes *asymptotically free* and yields a Curie-type susceptibility, which diverges for $T \rightarrow 0$. For small but finite coupling we can account for the ferromagnetic interaction perturbatively (expansion in orders of Γ_F). In f -electron materials often both ferro and antiferromagnetic exchange couplings are present, the first, Γ_F , arising from the Coulomb exchange, the second, Γ , from the hybridization. There are therefore two possibilities. If both exchange interactions couple the impurity with the same conduction channel, only the total coupling $\Gamma_F + \Gamma$ matters. Thus a $|\Gamma_F| > \Gamma$ suppresses the Kondo effect. If, however, ferromagnetic and antiferromagnetic exchange interaction couple the impurity to different conduction channels, a $|\Gamma_F| > \Gamma$ does not suppress the Kondo effect (Fig. 16) but merely reduces T_K . In the infinite $|\Gamma_F|$ limit the model describes an undercompensated Kondo effect [34].

3.2 The RKKY interaction

The Kondo Hamiltonian (30) describes a magnetic coupling between a local impurity and a bath of conduction electrons. Thus, in the presence of several Kondo impurities coupled to the same conduction electron bath, an indirect magnetic coupling between the local moments arises. Let us start for simplicity from two Kondo impurities described by the Hamiltonian

$$H_{2K} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{i=1,2} \Gamma \mathbf{S}_f^i \cdot \mathbf{s}_c(\mathbf{0}). \quad (33)$$

Let us calculate the effective magnetic coupling between the impurities by integrating out the degrees of freedom of the conduction electrons; this can be done again via perturbation theory or via a canonical transformation. At second order in Γ , the original Hamiltonian becomes

$$H_{\text{RKKY}} = I_{12}(\mathbf{R}_{12}) \mathbf{S}_f^1 \cdot \mathbf{S}_f^2 \quad (34)$$

where $\mathbf{R}_{12} = \mathbf{R}_1 - \mathbf{R}_2$ and

$$I_{12}(\mathbf{R}_{12}) \sim -\Gamma^2 \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} \Theta(\varepsilon_F - \varepsilon_{\mathbf{k}}) \Theta(\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_F) \frac{\cos \mathbf{q} \cdot \mathbf{R}_{12}}{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}.$$

For a free-electron gas one finds

$$I_{12}(R_{12}) \sim -\frac{\Gamma^2 q_F^4}{\pi^3} \left[\frac{\sin 2q_F R_{12} - 2q_F R_{12} \cos 2q_F R_{12}}{(2q_F R_{12})^4} \right],$$

where $R_{12} = |\mathbf{R}_{12}|$. The coupling $I_{12}(R_{12})$ decays as $\sim 1/R_{12}^3$ with the distance between the two impurities. It oscillates with a behavior similar to Friedel oscillations; the sign of the interaction at a certain distance between the impurities depend on the band filling, and it is negative (ferromagnetic) for $q_F \rightarrow 0$. The exchange Hamiltonian H_{RKKY} is known as the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction and competes with the Kondo effect.

We can understand this competition using the following – well known but naive – argument. For simplicity, we assume $I_{12} > 0$ (antiferromagnetic) as it is often the case. The energy gain obtained forming a singlet (antiferromagnetic state) is $E_S \propto -I_{12} \propto -\Gamma^2$; the Kondo energy gain is instead $E_K \propto -k_B T_K \sim -D e^{-2/\rho(\varepsilon_F)\Gamma}$. If the coupling constant Γ is small $|E_S|$ is larger than $|E_K|$. In this case, the antiferromagnetic order is favored over the Kondo effect, which would lead to the screening of local moments. In the opposite limit, i.e., when Γ is large, the Kondo effect dominates, and the local moments are screened – thus the system does not exhibit any long-range order. Although reality is more complex – the two effects occur together – this simple argument gives a picture of the mechanisms at play in a lattice of Kondo impurities. It is important to understand that the RKKY interaction is an indirect coupling arising from the interaction of a *correlated* impurity with a conduction-electron bath; the coupling I_{12} is proportional to Γ^2 , and the coupling Γ itself was obtained by integrating out high-energy doubly occupied states on the impurity, in a similar way as we have seen for kinetic exchange. In the $U = 0$ limit the coupling Γ diverges and the full construction breaks down. Finally, in a system in which non-perturbative effects – such as the Kondo effect – play a key role, the second order Hamiltonian H_{RKKY} is in general not even sufficient to describe the actual nature of the magnetic phenomena; to obtain H_{RKKY} we have actually integrated out the very interaction leading to the Kondo effect, and this is clearly incorrect in the general case.

4 Conclusion

In this lecture we introduced some of the fundamental aspects of magnetism in correlated systems. We have seen two distinct regimes, the itinerant and the local-moment regime. In the first regime we can, in most cases, treat correlation effects in perturbation theory. In the world of real materials this is the limit in which density-functional theory (DFT), in the local-density approximation or its simple extensions, works best. If the system is weakly correlated we can calculate the linear-response function in the random-phase approximation and understand magnetism within this approach fairly well.

The opposite regime is the strong-correlation limit, in which many-body effects play a key role. In this limit perturbation theory fails and we have in principle to work with many-body methods. If, however, we are interested only in magnetic phenomena, at integer filling a strong simplification comes from mapping the original many-body Hamiltonian into an effective spin model. The exact solution of effective spin models requires in general numerical methods such as the Monte Carlo or quantum Monte Carlo approach, or, when the system is small enough, exact diagonalization or Lanczos. To work with material-specific spin models we need to calculate the magnetic exchange parameters. Typically this is done starting from total-energy DFT calculations for different spin configurations, e.g., in the LDA+ U approximation. The LDA+ U approach is based on the Hartree-Fock approximation, and therefore when we extract the parameters from LDA+ U calculations we have to keep in mind the shortcomings of the method. Furthermore if we want to extract the magnetic couplings from total energy calculations we have to make a guess on the form of the spin model. More flexible approaches, which allow

us to account for actual correlation effects, are based on Green functions and the local-force theorem [36], or on canonical transformations [28, 37].

In strongly correlated materials localized- and itinerant-moment physics can often be observed in the same system, although in different energy or temperature regimes. This is apparent in the case of the Kondo effect. For a Kondo impurity, the susceptibility exhibits a Curie behavior at high temperature and a Fermi liquid behavior at low temperature. In correlated transition-metal oxides Fermi liquid and local-spin magnetism can both play an important role but at different energy scales. Furthermore, in the absence of a large charge gap, downfolding to spin models is not really justified. The modern method to bridge between the localized and itinerant regimes and deal with the actual complications of real systems is the dynamical mean-field theory (DMFT) [14–16]. Within this technique we directly solve generalized Hubbard-like models, albeit in the local self-energy approximation. DMFT is the first flexible approach that allows us to understand the paramagnetic Mott metal-insulator transition and thus also magnetism in correlated materials in a realistic setting. Modern DMFT codes are slowly but steadily becoming as complex and flexible as DFT codes, allowing us to deal with the full complexity of strongly correlated materials. While this is a huge step forward, we have to remember that state-of-the-art many-body techniques have been developed by solving simple models within certain approximations. We have to know these very well if we want to understand real materials and further advance the field. In DMFT we self-consistently solve an effective quantum impurity model, a generalization of the Anderson model. Thus a lot can be learnt from the solution of the Anderson model in the context of the Kondo problem. Much can be understood alone with simple arguments, as Anderson or Nozières have shown us, reaching important conclusions on the Kondo problem with paper and pencil.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Formalism

The formulas in this Appendix are in atomic units: The numerical value of e , m , and $\hbar$ is 1, that of μ_B is either $1/2$ (SI units) or $\alpha/2$ (cgs-Gauss units), where α is the fine-structure constant; the energies are in Hartree.

A.1 Matsubara Green functions

A.1.1 Imaginary-time and frequency Green functions

The imaginary-time Matsubara Green function is defined as

$$G_{\alpha\alpha'}(\boldsymbol{\tau}) = -\langle \mathcal{T} c_\alpha(\tau_1) c_{\alpha'}^\dagger(\tau_2) \rangle = -\frac{1}{Z} \text{Tr} \left[e^{-\beta(H-\mu N)} \mathcal{T} c_\alpha(\tau_1) c_{\alpha'}^\dagger(\tau_2) \right],$$

where $\mathcal{T}$ is the time-ordering operator, $\boldsymbol{\tau} = (\tau_1, \tau_2)$, $Z = \text{Tr} e^{-\beta(H-\mu N)}$ is the partition function, and the imaginary-time operators $o(\tau) = c(\tau), c^\dagger(\tau)$ are defined as

$$o(\tau) = e^{\tau(H-\mu N)} o e^{-\tau(H-\mu N)}.$$

The indices α and α' are the flavors; they can be site and spin indices in the atomic limit and $\mathbf{k}$ and spin indices in the non-interacting-electrons limit. Writing the action of the time-ordering operator explicitly, we obtain

$$G_{\alpha\alpha'}(\boldsymbol{\tau}) = -\Theta(\tau_1 - \tau_2) \langle c_\alpha(\tau_1) c_{\alpha'}^\dagger(\tau_2) \rangle + \Theta(\tau_2 - \tau_1) \langle c_{\alpha'}^\dagger(\tau_2) c_\alpha(\tau_1) \rangle.$$

Using the invariance of the trace of the product of operators under cyclic permutations, one can show that the following properties hold

$$\begin{aligned} G_{\alpha\alpha'}(\boldsymbol{\tau}) &= G_{\alpha\alpha'}(\tau_1 - \tau_2), \\ G_{\alpha\alpha'}(\tau) &= -G_{\alpha\alpha'}(\tau + \beta) \quad \text{for } -\beta < \tau < 0. \end{aligned}$$

The Fourier transform on the Matsubara axis is

$$G_{\alpha\alpha'}(i\nu_n) = \frac{1}{2} \int_{-\beta}^{\beta} d\tau e^{i\nu_n \tau} G_{\alpha\alpha'}(\tau) = \int_0^{\beta} d\tau e^{i\nu_n \tau} G_{\alpha\alpha'}(\tau),$$

with $\nu_n = (2n + 1)\pi/\beta$. The inverse Fourier transform is given by

$$G_{\alpha\alpha'}(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{+\infty} e^{-i\nu_n \tau} G_{\alpha\alpha'}(i\nu_n).$$

The convergence of $G_{\alpha\alpha'}(\tau)$ is only guaranteed in the interval $-\beta < \tau < \beta$. Finally, if n_α is the number of electrons for flavor α , one can show that

$$G_{\alpha\alpha}(\tau \rightarrow 0^+) = -1 + n_\alpha, \quad G_{\alpha\alpha}(\tau \rightarrow \beta^-) = -n_\alpha. \quad (35)$$

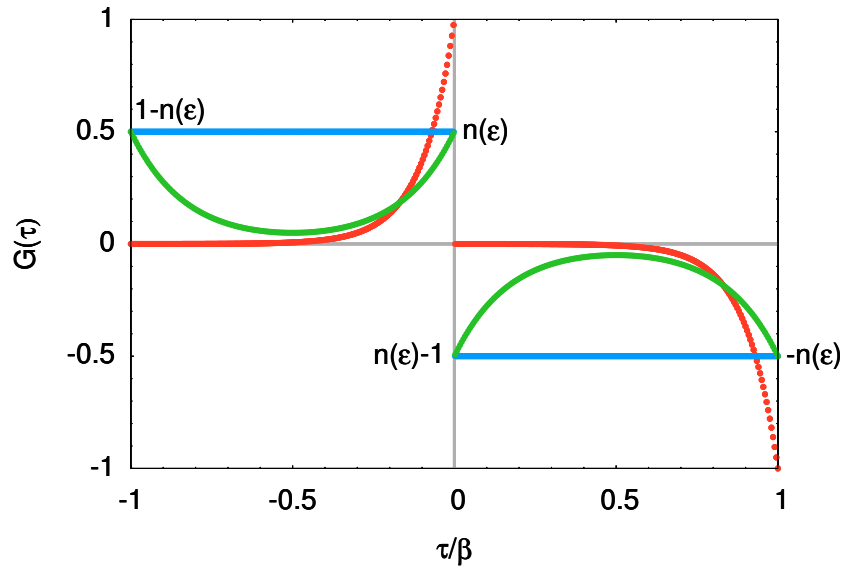


Fig. 17: The function $\mathcal{G}_{k\sigma}(\tau)$ defined in Eq. (37) for a state well below the Fermi level (red) and at the Fermi level (blue) for $\beta = 2 \text{ (eV)}^{-1}$. The green line shows the atomic $G(\tau)$ from Eq. (39) for $U = 6 \text{ eV}$ and magnetic field $h = 0$.

A.1.2 Non-interacting limit

For a non-interacting system described by the Hamiltonian

$$H_0 = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} \quad (36)$$

we can show that the imaginary time Green function $\mathcal{G}_{k\sigma}(\tau)$ is given by

$$\begin{aligned} \mathcal{G}_{k\sigma}(\tau) &= - \left\langle \mathcal{T} \left[c_{\mathbf{k}\sigma}(\tau) c_{\mathbf{k}\sigma}^\dagger(0) \right] \right\rangle \\ &= - \left[\Theta(\tau) (1 - n_\sigma(\varepsilon_{\mathbf{k}})) - \Theta(-\tau) n_\sigma(\varepsilon_{\mathbf{k}}) \right] e^{-(\varepsilon_{\mathbf{k}} - \mu)\tau}, \end{aligned} \quad (37)$$

where $n_\sigma(\varepsilon_{\mathbf{k}})$ is the occupation number

$$n_\sigma(\varepsilon_{\mathbf{k}}) = \frac{1}{1 + e^{\beta(\varepsilon_{\mathbf{k}} - \mu)}}.$$

The Fourier transform of the Green function $\mathcal{G}_{k\sigma}(\tau)$ at the Matsubara frequencies is

$$\mathcal{G}_{k\sigma}(i\nu_n) = \frac{1}{i\nu_n - (\varepsilon_{\mathbf{k}} - \mu)}.$$

To obtain the analytic continuation of this Green function on the real axis we substitute

$$i\nu_n \rightarrow \omega + i0^+.$$

A.1.3 Matsubara sums

The non-interacting Green function $\mathcal{G}_{\mathbf{k}\sigma}(z)$ has a pole at $z_p = \varepsilon_{\mathbf{k}} - \mu$; the Fermi function $n_{\sigma}(z)$ has poles for $z = i\nu_n$ instead. Let us consider the integral

$$\frac{1}{2\pi i} \oint_C \mathcal{F}_{\mathbf{k}\sigma}(z) n_{\sigma}(z) e^{z\tau} dz = 0,$$

where $0 < \tau < \beta$ and where the function $\mathcal{F}_{\mathbf{k}\sigma}(z)$ is analytic everywhere except at some poles $\{z_p\}$. The contour C is a circle in the full complex plane centered at the origin and including the poles of the Fermi function (Matsubara frequencies) and the poles of $\mathcal{F}_{\mathbf{k}\sigma}(z)$. The integral is zero because the integrand vanishes exponentially for $|z| \rightarrow \infty$. Furthermore

$$\text{Res } [n_{\sigma}(i\nu_n)] = -\frac{1}{\beta}.$$

Using Cauchy's integral theorem we then have

$$\frac{1}{\beta} \sum_n e^{i\nu_n \tau} \mathcal{F}_{\mathbf{k}\sigma}(i\nu_n) = \sum_{z_p} \text{Res } [\mathcal{F}_{\mathbf{k}\sigma}(z_p)] n_{\sigma}(z_p) e^{z_p \tau}.$$

We can use this expression and (35) to show that

$$\begin{aligned} \frac{1}{\beta} \sum_n e^{-i\nu_n 0^-} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \mathcal{G}_{\mathbf{k}\sigma}(0^-) = n_{\sigma}(\varepsilon_{\mathbf{k}}), \\ \frac{1}{\beta} \sum_n e^{-i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \mathcal{G}_{\mathbf{k}\sigma}(0^+) = n_{\sigma}(\varepsilon_{\mathbf{k}}) - 1. \end{aligned}$$

In a similar way we can show that

$$\begin{aligned} \frac{1}{\beta} \sum_n e^{i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \frac{dn_{\sigma}(\varepsilon_{\mathbf{k}})}{d\varepsilon_{\mathbf{k}}} = \beta n_{\sigma}(\varepsilon_{\mathbf{k}}) [-1 + n_{\sigma}(\varepsilon_{\mathbf{k}})], \\ \frac{1}{\beta} \sum_n e^{i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(i\nu_n + i\omega_m) &= \frac{n_{\mathbf{k}+\mathbf{q}} - n_{\mathbf{k}}}{-i\omega_m + \varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}, \end{aligned}$$

where in the last relation $\omega_m = 2m\pi/\beta$ is a bosonic Matsubara frequency.

A.1.4 Atomic limit

It is interesting to consider a half-filled idealized atom described by the Hamiltonian

$$H = \varepsilon_d \sum_{\sigma} n_{\sigma} + U \left(\frac{N^2}{4} - S_z^2 \right) + g\mu_B h S_z. \quad (38)$$

For $\tau > 0$ we can calculate explicitly the Green function, obtaining

$$G_{\sigma}(\tau) = -\frac{1}{2} \frac{1}{1 + e^{\beta U/2} \cosh(\beta g\mu_B h/2)} \left[e^{\tau(U - g\mu_B h\sigma)/2} + e^{(\beta - \tau)(U + g\mu_B h\sigma)/2} \right]. \quad (39)$$

The Fourier transform of $G_\sigma(\tau)$ is

$$G_\sigma(i\nu_n) = \left[\frac{w_-}{i\nu_n + (U - g\mu_B h\sigma)/2} + \frac{w_+}{i\nu_n - (U + g\mu_B h\sigma)/2} \right],$$

where

$$w_\pm = \frac{1}{2} \frac{1 + e^{\beta U/2} e^{\pm \beta g\mu_B h\sigma/2}}{1 + e^{\beta U/2} \cosh(\beta g\mu_B h/2)}.$$

Since the Green function is written as the sum of functions with one pole, the analytic continuation is simple, as in the non-interacting case. We replace $i\nu_n$ with $\omega + i0^+$.

A.2 Linear-response theory

A.2.1 Theory

The response of a system described by the Hamiltonian H to a small magnetic field $\mathbf{h}(\mathbf{r}, t)$ is given by the linear correction to the Hamiltonian, i.e.,

$$\sum_\nu \delta H_\nu(\mathbf{r}; t) = - \sum_\nu M_\nu(\mathbf{r}; t) h_\nu(\mathbf{r}; t), \quad (40)$$

where $M(\mathbf{r}; t)$ is the magnetization operator in the Heisenberg representation

$$M_\nu(\mathbf{r}; t) = e^{iHt} M_\nu(\mathbf{r}) e^{-iHt}$$

and $\nu = x, y, z$. To linear order in the perturbation and assuming that the perturbation is turned on adiabatically at $t_0 = -\infty$

$$\underbrace{\langle M_\nu(\mathbf{r}; t) \rangle - \langle M_\nu(\mathbf{r}) \rangle_0}_{\delta \langle M_\nu(\mathbf{r}; t) \rangle} = -i \sum_{\nu'} \int d\mathbf{r}' \int_{-\infty}^t dt' \langle [\Delta M_\nu(\mathbf{r}; t), \delta H_{\nu'}(\mathbf{r}'; t')] \rangle_0.$$

Here $\langle M_\nu(\mathbf{r}) \rangle_0$ is the (equilibrium) thermal average in the absence of the perturbation and $\Delta M_\nu(\mathbf{r}; t) = M_\nu(\mathbf{r}; t) - \langle M_\nu(\mathbf{r}) \rangle_0$. By replacing $\sum_{\nu'} \delta H_{\nu'}(\mathbf{r}'; t')$ with (40) we obtain

$$\delta \langle M_\nu(\mathbf{r}; t) \rangle = i \sum_{\nu'} \int d\mathbf{r}' \int_{-\infty}^t dt' \langle [\Delta M_\nu(\mathbf{r}; t), \Delta M_{\nu'}(\mathbf{r}'; t')] \rangle_0 h_{\nu'}(\mathbf{r}'; t').$$

The function

$$\chi_{\nu\nu'}(\mathbf{r}, \mathbf{r}'; t, t') = i \langle [\Delta M_\nu(\mathbf{r}; t), \Delta M_{\nu'}(\mathbf{r}'; t')] \rangle_0 \Theta(t - t') \quad (41)$$

is the so-called retarded response function. If the Hamiltonian H has time translational invariance symmetry the retarded response function depends only on time differences $t - t'$. For the Fourier transform, we have

$$\delta \langle M_\nu(\mathbf{r}; \omega) \rangle = \sum_{\nu'} \int d\mathbf{r}' \chi_{\nu\nu'}(\mathbf{r}, \mathbf{r}'; \omega) h_{\nu'}(\mathbf{r}'; \omega).$$

For a system with translational invariance, we additionally have

$$\delta \langle M_\nu(\mathbf{q}; \omega) \rangle = \sum_{\nu'} \chi_{\nu\nu'}(\mathbf{q}; \omega) h_{\nu'}(\mathbf{q}; \omega).$$

In the $\omega = 0$ and $\mathbf{q} \rightarrow \mathbf{0}$ limit we have

$$\chi_{\nu\nu'}(\mathbf{0}; 0) = \lim_{h_{\nu'} \rightarrow 0} \frac{\partial M_\nu}{\partial h_{\nu'}},$$

where $h_{\nu'} = h_{\nu'}(\mathbf{0}; 0)$. More details can be found in Ref. [27]. In the rest of the Appendix we replace for simplicity the notation $\langle \dots \rangle_0$ with $\langle \dots \rangle$.

A.2.2 Kramers-Kronig relations and thermodynamic sum rule

Important properties of the spin susceptibility are the Kramers-Kronig relations

$$\begin{aligned} \text{Re}[\chi(\mathbf{q}; \omega)] - \text{Re}[\chi(\mathbf{q}; \infty)] &= \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Im}[\chi(\mathbf{q}; \omega')]}{\omega' - \omega} d\omega', \\ \text{Im}[\chi(\mathbf{q}; \omega)] &= -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Re}[\chi(\mathbf{q}; \omega')] - \text{Re}[\chi(\mathbf{q}; \infty)]}{\omega' - \omega} d\omega', \end{aligned}$$

where $\mathcal{P}$ is the Cauchy principal value, and Re and Im indicate the real and imaginary part. The first Kramers-Kronig relation yields the sum rule

$$\text{Re}[\chi(\mathbf{q}; \omega = 0)] - \text{Re}[\chi(\mathbf{q}; \infty)] = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Im}[\chi(\mathbf{q}; \omega')]}{\omega'} d\omega'. \quad (42)$$

In the $\mathbf{q} = \mathbf{0}$ limit, Eq. (42) is known as *thermodynamic sum rule*.

A.2.3 Fluctuation-dissipation theorem and static susceptibility

We define the spin-spin correlation function

$$\mathcal{S}_{\nu\nu'}(\mathbf{q}; t) = \langle \Delta S_\nu(\mathbf{q}; t) \Delta S_{\nu'}(-\mathbf{q}) \rangle$$

where $\Delta S_\nu(\mathbf{q}; t) = S_\nu(\mathbf{q}; t) - \langle S_\nu(\mathbf{q}; 0) \rangle$ and where the S_ν are spin operators. The Fourier transform of the correlation function in frequency space is $\mathcal{S}_{\nu\nu'}(\mathbf{q}; \omega)$. One can show that

$$\mathcal{S}_{\nu\nu'}(\mathbf{q}; \omega) = e^{\beta\omega} \mathcal{S}_{\nu'\nu}(\mathbf{q}; -\omega).$$

The following formula, known as fluctuation-dissipation theorem, relates the spin-spin correlation function with the magnetic susceptibility

$$\text{Im}[\chi_{\nu\nu'}(\mathbf{q}; \omega)] = \frac{1}{2(1 + n_B)} (g\mu_B)^2 \mathcal{S}_{\nu\nu'}(\mathbf{q}; \omega), \quad n_B(\omega) = \frac{1}{e^{\beta\omega} - 1}.$$

Assuming $k_B T$ large and using Eq. (42) we find

$$\text{Re}[\chi_{\nu\nu'}(\mathbf{q}; \omega = 0)] - \text{Re}[\chi_{\nu\nu'}(\mathbf{q}; \infty)] \sim \frac{(g\mu_B)^2}{k_B T} \mathcal{S}_{\nu\nu'}(\mathbf{q}; t = 0).$$

A.2.4 Imaginary-time and frequency response function

We define the susceptibility in imaginary time as

$$\chi_{\nu\nu'}(\mathbf{q}; \tau, \tau') = \langle \mathcal{T} \Delta M_\nu(\mathbf{q}; \tau) \Delta M_{\nu'}(-\mathbf{q}; \tau') \rangle$$

where $\Delta M_\nu(\mathbf{q}; \tau) = M_\nu(\mathbf{q}; \tau) - \langle M_\nu(\mathbf{q}; 0) \rangle$. As in the case of the Green function, by using the invariance properties of the trace one can show that

$$\chi_{\nu\nu'}(\mathbf{q}; \tau, \tau') = \chi_{\nu\nu'}(\mathbf{q}; \tau - \tau').$$

The response function in imaginary time is related to the response function at the bosonic Matsubara frequency $i\omega_m$ through the Fourier transforms

$$\begin{aligned} \chi_{\nu\nu'}(\mathbf{q}; \tau) &= \frac{1}{\beta} \sum_n e^{-i\omega_m \tau} \chi_{\nu\nu'}(\mathbf{q}; i\omega_m), \\ \chi_{\nu\nu'}(\mathbf{q}; i\omega_m) &= \int d\tau e^{i\omega_m \tau} \chi_{\nu\nu'}(\mathbf{q}; \tau). \end{aligned}$$

A.3 Magnetic susceptibility

A.3.1 Spin and magnetization operators

The spin operators S_ν are defined as

$$S_\nu = \frac{1}{2} \sum_{\sigma\sigma'} c_\sigma^\dagger \sigma_\nu c_{\sigma'},$$

where $\nu = x, y, z$ and σ_ν are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The magnetization operators M_ν are defined as $M_\nu = -g\mu_B S_\nu$.

A.3.2 Matsubara magnetic susceptibility

The magnetic susceptibility for a single-band system can be expressed as

$$\chi_{zz}(\mathbf{q}; \boldsymbol{\tau}) = \frac{(g\mu_B)^2}{4} \sum_{\sigma\sigma'} \sigma\sigma' \chi^{q\sigma\sigma'}(\boldsymbol{\tau}) = \frac{(g\mu_B)^2}{4} \sum_{\sigma\sigma'} \sigma\sigma' \underbrace{\frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\mathbf{k}'} [\chi(\mathbf{q}; \boldsymbol{\tau})]_{\sigma\mathbf{k}, \sigma'\mathbf{k}'}}_{\chi^{q\sigma\sigma'}(\boldsymbol{\tau})} \quad (43)$$

where $\sigma = 1$ or -1 depending on whether the spin is up or down, $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$ and

$$\begin{aligned} [\chi(\mathbf{q}; \boldsymbol{\tau})]_{\sigma\mathbf{k}, \sigma'\mathbf{k}'} &= \langle \mathcal{T} c_{\mathbf{k}\sigma}(\tau_1) c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger(\tau_2) c_{\mathbf{k}'+\mathbf{q}\sigma'}(\tau_3) c_{\mathbf{k}'\sigma'}^\dagger(\tau_4) \rangle \\ &\quad - \langle \mathcal{T} c_{\mathbf{k}\sigma}(\tau_1) c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger(\tau_2) \rangle \langle \mathcal{T} c_{\mathbf{k}'+\mathbf{q}\sigma'}(\tau_3) c_{\mathbf{k}'\sigma'}^\dagger(\tau_4) \rangle. \end{aligned}$$

In Fourier space

$$\begin{aligned}\chi_{zz}(\mathbf{q}; i\omega_m) &= \frac{(g\mu_B)^2}{4} \sum_{\sigma\sigma'} \sigma\sigma' \frac{1}{\beta^2} \sum_{nn'} \chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) \\ &= \frac{(g\mu_B)^2}{4} \sum_{\sigma\sigma'} \sigma\sigma' \frac{1}{\beta^2} \sum_{nn'} \underbrace{\frac{1}{\beta} \frac{1}{N_{\mathbf{k}}^2} \sum_{\mathbf{k}\mathbf{k}'} [\chi(\mathbf{q}; i\omega_m)]_{\sigma\mathbf{k}n, \sigma'\mathbf{k}'n'}}_{\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m)},\end{aligned}$$

where $\omega_m = 2m\pi/\beta$ is a bosonic Matsubara frequency and

$$\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\nu}) = \frac{\beta}{8} \iiint d\boldsymbol{\tau} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\tau}} \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}). \quad (44)$$

The integral for each $\boldsymbol{\tau}$ component is from $-\beta$ to β and $\boldsymbol{\nu} = (\nu_n, -\nu_n - \omega_m, \nu_{n'} + \omega_m, -\nu_{n'})$.

A.3.3 Symmetry properties

Let us now analyze the symmetry properties of (44). The complex conjugate is given by

$$\left[\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) \right]^* = \chi_{-n-1, -n'-1}^{\mathbf{q}\sigma\sigma'}(-i\omega_m),$$

with

$$\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \frac{\beta}{8} \iiint d\boldsymbol{\tau} e^{i(-\omega_m\tau_{23} + \nu_n\tau_{12} + \nu_{n'}\tau_{34})} \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}).$$

By using the fact that, for the cases considered here, the response function is real in τ space and by exchanging the indices 1 and 4, 2 and 3 in the integrand, we find

$$\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \chi_{n',n}^{\mathbf{q}\sigma'\sigma}(i\omega_m),$$

and hence if $\sigma = \sigma'$, $\nu_n = \nu_{n'}$ is a reflection axis. An additional reflection axis can be found by first shifting the frequency $\nu_n = \nu_l - \omega_m$

$$\chi_{l,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \frac{\beta}{8} \iiint d\boldsymbol{\tau} e^{i(-\omega_m\tau_{13} + \nu_l\tau_{12} + \nu_{n'}\tau_{34})} \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau})$$

and then exchanging in the integrand the indices 12 with 34 and vice versa. Hence

$$\chi_{l,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \chi_{n',l}^{\mathbf{q}\sigma'\sigma}(-i\omega_m)$$

so that, if $\sigma = \sigma'$, $\nu_{n+m} = -\nu_{n'}$ is a mirror line

$$\left| \chi_{n+m,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) \right| = \left| \chi_{-n'-1, -n-m-1}^{\mathbf{q}\sigma'\sigma}(i\omega_m) \right|.$$

A.3.4 Non-interacting limit

For a non-interacting system described by Hamiltonian (36) Wick's theorem yields

$$\begin{aligned}\chi^{q\sigma\sigma'}(\boldsymbol{\tau}) &= -\frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \langle \mathcal{T} c_{\mathbf{k}\sigma}(\tau_1) c_{\mathbf{k}+\mathbf{q}\sigma'}^\dagger(\tau_4) \rangle \langle \mathcal{T} c_{\mathbf{k}+\mathbf{q}\sigma'}(\tau_3) c_{\mathbf{k}\sigma}^\dagger(\tau_2) \rangle \\ &= -\frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \mathcal{G}_{\mathbf{k}\sigma}(\tau_{14}) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma'}(-\tau_{23}) \delta_{\sigma,\sigma'}.\end{aligned}$$

Then, in the paramagnetic case, the magnetic susceptibility is given by

$$\chi_{zz}(\mathbf{q}; \boldsymbol{\tau}) = -(g\mu_B)^2 \frac{1}{4} \frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\sigma} \mathcal{G}_{\mathbf{k}\sigma}(\tau_{14}) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(-\tau_{23}).$$

Its Fourier transform is

$$\chi_{zz}(\mathbf{q}; i\omega_m) = (g\mu_B)^2 \frac{1}{4} \frac{1}{\beta^2} \sum_{nn'} \sum_{\sigma} \chi_{n,n'}^{q\sigma\sigma}(i\omega_m),$$

where

$$\sum_{\sigma} \chi_{n,n'}^{q\sigma\sigma}(i\omega_m) = -\beta \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\sigma} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(i\nu_n + i\omega_m) \delta_{n,n'}.$$

Thus, the static susceptibility is

$$\chi_{zz}(\mathbf{q}; 0) = -(g\mu_B)^2 \frac{1}{4} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\sigma} \frac{n_{\sigma}(\varepsilon_{\mathbf{k}+\mathbf{q}}) - n_{\sigma}(\varepsilon_{\mathbf{k}})}{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}.$$

Finally, in the $\mathbf{q} \rightarrow 0$ and $T \rightarrow 0$ limit we find

$$\chi_{zz}(\mathbf{0}; 0) = \frac{1}{4} (g\mu_B)^2 \frac{1}{N_{\mathbf{k}}} \underbrace{\sum_{\mathbf{k}\sigma} \left[-\frac{dn_{\sigma}(\varepsilon_{\mathbf{k}})}{d\varepsilon_{\mathbf{k}}} \right]_{T=0}}_{\rho(\varepsilon_F)} = \frac{1}{4} (g\mu_B)^2 \rho(\varepsilon_F).$$

A.3.5 Atomic limit

In the atomic limit, we sum over $\mathbf{q}$ to obtain the local susceptibility tensor

$$\chi^{\sigma\sigma'}(\boldsymbol{\tau}) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} \chi^{q\sigma\sigma'}(\boldsymbol{\tau}).$$

For Hamiltonian (38), in the sector $\boldsymbol{\tau}^+$ such that $\tau_i > \tau_{i+1}$, the latter has the form

$$\chi^{\sigma\sigma'}(\boldsymbol{\tau}^+) = \frac{1}{\beta} \frac{1}{2(1 + e^{\beta U/2})} (e^{\tau_{12}U/2 + \tau_{34}U/2} + \delta_{\sigma\sigma'} e^{(\beta - \tau_{12})U/2 - \tau_{34}U/2}).$$

The magnetic susceptibility for $\tau_i > \tau_{i+1}$ is then given by

$$\chi_{zz}(\boldsymbol{\tau}^+) = (g\mu_B)^2 \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' \chi^{\sigma\sigma'}(\boldsymbol{\tau}^+) = \frac{(g\mu_B)^2}{4\beta} \frac{1}{(1 + e^{\beta U/2})} e^{(\beta - \tau_{12} - \tau_{34})U/2},$$

which depends only on $\tau = \tau_{12} + \tau_{34}$. In the remaining sectors (labeled with P) the susceptibility has a similar form after appropriate permutation of the imaginary times

$$\chi_{zz}(\tau^P) = (g\mu_B)^2 \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' \chi^{\sigma\sigma'}(\tau^P) = s_P \frac{(g\mu_B)^2}{4\beta} n(-s_P y) e^{-s_P(\tau_{12} + \tau_{34})y},$$

where $y = U/2$ and $s_P = \pm 1$; the derivation can be found in [27]. If we perform the Fourier transform of $\chi_{zz}(\tau)$ we find $\chi_{zz}(i\omega_n) = \chi_{zz}(0)\delta_{\omega_n,0}$. The static susceptibility is

$$\chi_{zz}(0) = (g\mu_B)^2 \frac{1}{4k_B T} \frac{e^{\beta U/2}}{1 + e^{\beta U/2}} = (g\mu_B)^2 \frac{1}{4} \frac{1}{\beta^2} \sum_{nn'} \sum_{\sigma\sigma'} \sigma\sigma' \chi_{n,n'}^{\sigma\sigma'}(0).$$

Here, after setting

$$M_n = \left[\frac{1}{i\nu_n - y} - \frac{1}{i\nu_n + y} \right]$$

we have

$$\begin{aligned} \sum_{\sigma\sigma'} \sigma\sigma' \chi_{n,n'}^{\sigma\sigma'}(0) &= \frac{dM_{n'} M_n}{dy} - \beta n(y) \left[\delta_{n,n'} \frac{dM_n}{dy} + \delta_{n,-n'-1} \frac{dM_{n'}}{dy} \right] + \beta n(-y) M_n M_{n'} \\ &\quad - \frac{1}{y} \left[M_{n'} - \beta n(y) \delta_{n,-n'-1} + \beta n(-y) \delta_{n,n'} \right] M_n. \end{aligned}$$

Most contributions cancel each other when the sums over the Matsubara frequencies are performed. A detailed derivation can be found in Ref. [27]. For $y = 0$ only the terms proportional to $\delta_{n,n'}$ survive, as expected from the Wick theorem.

References

- [1] P. Weiss, J. Phys. Theor. Appl. **6**, 661 (1907)
- [2] F. Hund, Z. Phys. **33**, 855 (1925)
- [3] W. Heisenberg, Z. Phys. **49**, 619 (1928)
- [4] C.G. Shull and J.S. Smart, Phys. Rev. **76**, 1256 (1949)
- [5] L. Néel, Ann. de Phys. **17**, 5 (1932); *ibid.* **5**, 232 (1936); C. R. Acad. Sc. **203**, 304 (1936)
- [6] H. Bethe, Z. Phys. **71**, 205 (1931)
- [7] P.W. Anderson, Phys. Rev. **86**, 694 (1952)
- [8] P.W. Anderson, Phys. Rev. **130**, 439 (1963)
- [9] P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
- [10] J. Kondo, Prog. Theor. Phys. **32**, 37 (1964)
- [11] K. Andres, J.E. Graebner and H.R. Ott, Phys. Rev. Lett. **35**, 1779 (1975)
- [12] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge University Press, 1993)
- [13] J.G. Bednorz, K.A. Müller, Z. Phys. B **64**, 189 (1986)
- [14] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989);
A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [15] 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>
- [16] 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>
- [17] H. Bruus and K. Flensberg: *Many-Body Quantum Theory in Condensed Matter Physics* (Oxford University Press, 2004)
- [18] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)
- [19] D.C. Matthis: *The Theory of Magnetism Made Simple* (World Scientific, Singapore, 2006)

- [20] K. Yosida: *Theory of Magnetism* (Springer, Heidelberg, 1998)
- [21] E. Pavarini, E. Koch, F. Anders, M. Jarrell:
Correlated electrons: from models to materials,
 Reihe Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
<http://www.cond-mat.de/events/correl12>
- [22] E. Pavarini, E. Koch, U. Schollwöck:
Emergent Phenomena in Correlated Matter,
 Reihe Modeling and Simulation, Vol. 3 (Forschungszentrum Jülich, 2013)
<http://www.cond-mat.de/events/correl13>
- [23] See R. Eder, *Multiplets in Transition Metal Ions* in Ref. [21]
- [24] See E. Pavarini, *Crystal-field Theory, Tight-binding Method and Jahn-Teller Effect*,
 in Ref. [21]
- [25] See E. Pavarini, *The LDA+DMFT Approach*, in Ref. [15]
- [26] See E. Koch, *Exchange Mechanisms*, in Ref. [21]
- [27] See E. Pavarini, *Linear-response Theory*, in Ref. [16]
- [28] 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)
- [29] E. Pavarini, I. Dasgupta, T.Saha-Dasgupta, O. Jepsen and O.K. Andersen,
 Phys. Rev. Lett. **87**, 047003 (2001)
- [30] K. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
- [31] N. Andrei, K. Furuya, and J.H. Lowenstein, Rev. Mod. Phys. **55**, 331 (1983);
 A.M. Tsvelik and P.B. Wiegmann, Adv. Phys. **32**, 453 (1983)
- [32] J.E. Gubernatis, J.E. Hirsch, and D.J. Scalapino, Phys. Rev. B **16**, 8478 (1987)
- [33] P.W. Anderson, J. Phys. C: Solid State Phys. **3**, 2436 (1970)
- [34] E. Pavarini and L.C. Andreani, Phys. Rev. Lett. **77**, 2762 (1996)
- [35] P. Nozières, J. Low. Temp. Phys. **17**, 31 (1974)
- [36] 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)

- [37] E. Pavarini, E. Koch, and A.I. Lichtenstein,
Phys. Rev. Lett. **101**, 266405 (2008);
A. Chiesa, S. Carretta, P. Santini, G. Amoretti and E. Pavarini,
Phys. Rev. Lett. **110**, 157204 (2013)

4 The Kondo Model and Poor Man's Scaling

Andriy H. Nevidomskyy

Dept. of Physics and Astronomy, Rice University

6100 Main Street, Houston, TX 77005, USA

Contents

1	The Kondo problem: Introduction	2
2	Concept of renormalization	5
3	Poor man's scaling for the Kondo model	5
3.1	T -matrix description of scattering processes	5
3.2	Renormalization of J_z	7
3.3	Renormalization of $J_{\pm}$	8
3.4	Renormalization group flow	9
3.5	Kondo temperature and breakdown of the perturbative scheme	11
4	Low-temperature properties of the Kondo model	12
4.1	Wilson's numerical renormalization	12
4.2	Ground state of the Kondo model	13
5	Multichannel Kondo problem	14
5.1	Phenomenology and scaling	14
5.2	Two-channel Kondo problem	17
6	Kondo model in the presence of Hund's coupling	19
6.1	Poor man's scaling with Hund's coupling	21
6.2	Experimental ramifications of Hund's coupling	25

The Kondo model has played a very important role in condensed matter physics. Experimentally motivated, it attracted a great deal of theoretical attention in the 1960s and 1970s, resulting in the conclusion that thermodynamic and transport properties depended logarithmically on temperature as $\ln(T/T_K)$, where T_K is called the Kondo temperature. The ideas of summing up the leading logarithmic divergences and establishing how this procedure depended on the high-energy cutoff were instrumental in the development of the scaling theory and the renormalization group, which were initially invented in the 1950s in high-energy physics. Despite this progress, what was very puzzling was that the resulting theoretical predictions for the thermodynamic and transport properties displayed a divergence at $T \approx T_K$, at which point the theory became unusable. Was this logarithmic divergence physical and what was the fate of the model at low temperatures $T \ll T_K$? These questions remained unanswered for almost a decade, until the breakthrough made by Kenneth Wilson in 1974, who invented a numerical algorithm of renormalization, now known as the numerical renormalization group, and showed it to be stable down to very low temperatures [1]. Wilson's work was hugely influential, for which he was awarded the Nobel prize in physics in 1982. At the same time, Nozières had developed a phenomenological low-energy theory of the Kondo model [2], showing it to be a Fermi liquid, in agreement with Wilson's numerical conclusions. This was a triumph of theory, further corroborated when the exact solution of the Kondo model was found in 1980 [3, 4]. From a historical perspective, the Kondo model therefore clearly has an iconic status. However, this is not the only reason why this topic features prominently in several Lectures in this School. It can be said without exaggeration that the ideas of scaling and renormalization group developed *en route* to solving the Kondo problem represent a cornerstone in our current understanding of correlated many-body systems, applicable to both condensed matter and high-energy physics. In this Lecture, I will first briefly introduce the Kondo model, before discussing in detail the elegant renormalization argument invented by P.W. Anderson, the so-called Poor Man's scaling theory [5]. I will then summarize briefly Wilson's numerical renormalization group idea as well as the aforementioned Fermi liquid theory by Nozières. The discussion in these sections is loosely based on the original article by Anderson [5] as well as on the textbooks by Yamada [6] and Hewson [7]. Having thus introduced the concept of scaling and renormalization, I will further illustrate their value by applying these methods to the more complicated incarnations of the Kondo model based on the so-called multichannel Kondo model in Section 5 and Section 6. This lecture is self-contained; however it presumes that the reader is well versed in the language of second quantization and has some familiarity with Feynman diagrams. Other than this, no special prerequisites are necessary.

1 The Kondo problem: Introduction

It was noticed as early as the 1930s that the resistance of noble metals like gold or silver exhibits a minimum as a function of temperature, see Fig. 1. It was later realized that this effect arises from magnetic impurities such as Mn and Fe, which are naturally present in noble metals. In ordinary metals, the electrical resistance originates from the lattice umklapp scattering and

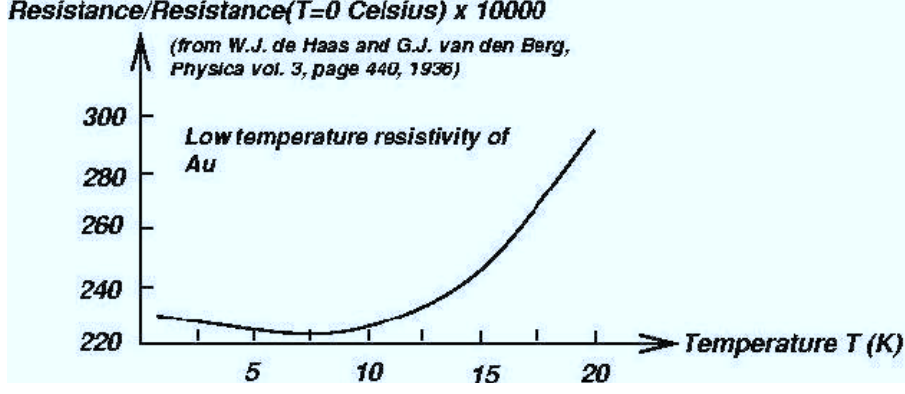


Fig. 1: Normalized resistance of Au with magnetic impurities as a function of temperature. (Reproduced from Ref. [8])

scattering off of impurities as well as lattice vibrations (phonons). When the temperature is lowered from room temperature, the resistance due to phonons decreases proportionally to T^5 . At much lower temperatures, when lattice vibrations are frozen out, the temperature dependence of resistance stems from the electron-electron interaction, which in ordinary metals scales as T^2 , consistent with the prediction of Landau's Fermi liquid theory. In any event, the resistance of a regular metal is a monotonically decreasing function as the temperature is lowered. By contrast, in dilute magnetic alloys the resistance starts increasing again with decreasing temperature.

This behavior of the resistance remained a puzzle until 1964, 30 years after the experimental discovery, when Jun Kondo presented the theory that explains the resistance minimum [9]. Kondo wrote down the model in which the dilute magnetic impurities are described by spin variables $\mathbf{S}(\mathbf{R}_i)$ at positions $\mathbf{R}_i$ that interact with conduction electrons via a spin-spin interaction. Since the impurities are randomly distributed and dilute, it is sufficient to consider one such impurity interacting with conduction electrons:

$$H = \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + 2J\mathbf{s} \cdot \mathbf{S}, \quad (1)$$

where conduction electron spin $\mathbf{s}$ at the impurity site $\mathbf{R} = 0$ is defined as $\mathbf{s}(\mathbf{R}) = \frac{1}{2}c^\dagger(\mathbf{R})\boldsymbol{\sigma}c(\mathbf{R})$ (setting $\hbar = 1$ for convenience). The spin interaction in the last term arises from the exchange interaction between a conduction electron (for instance, in an s -shell of Au) and the localized electron (d -shell in the case of transition metal impurities). The above model is often referred to as the s - d model or, equivalently, as the Kondo model (in what follows, we shall adopt the latter nomenclature). The factor of 2 in front of the interaction is chosen for convenience.

Equivalently, the model can be re-written by Fourier transforming the conduction electron creation/annihilation operators to the reciprocal space as follows:

$$H = \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + J \sum_{\mathbf{k}, \mathbf{k}'} c_{\mathbf{k}'\sigma'}^\dagger \boldsymbol{\sigma}_{\sigma'\sigma} c_{\mathbf{k}\sigma} \cdot \mathbf{S}, \quad (2)$$

with the summation over spin indices σ, σ' implied. One can further generalize this model by allowing anisotropy of the exchange interaction:

$$H = \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + J_z \sum_{\mathbf{k}, \mathbf{k}'} \sum_{\sigma} c_{\mathbf{k}'\sigma}^\dagger \sigma_{\sigma\sigma}^z c_{\mathbf{k}\sigma} \cdot S_z + \frac{1}{2} \sum_{\mathbf{k}, \mathbf{k}'} \left(J_- c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\downarrow} S^- + J_+ c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}\uparrow} S^+ \right), \quad (3)$$

where as usual $S^\pm \equiv S_x \pm iS_y$. In what follows, we shall assume the transverse spin interaction to be isotropic: $J_+ = J_- = J_\pm$ (in which case $J_\pm = J_x = J_y$ also follows).

In order to calculate the resistance of the model in Eq. (2), Kondo computed the scattering probability for conduction electrons using the T -matrix formalism [9, 10]. This formalism will be introduced in detail when discussing the scaling of the Kondo model in Section 3, so in order to avoid an unnecessary repetition, we shall only quote the final result for the resistance obtained by Kondo in the first Born approximation (see Ch. 4 of the book by Yamada [6] for more details):

$$R = R_0 \left[1 - 4J\rho \ln \left(\frac{k_B T}{D} \right) + \dots \right], \quad (4)$$

where R_0 is the residual (temperature-independent) resistance, D is the conduction electron bandwidth and ρ is the density of states at the Fermi level. As temperature decreases, $k_B T \ll D$ and the logarithm is negative, leading to a logarithmic increase of the resistance (and eventual divergence as $T \rightarrow 0$) provided $J > 0$. This is the essence of the Kondo effect, which explains the low-temperature behavior of the resistance in Fig. 1. At high temperatures, on the other hand, the aforementioned T^5 contribution to resistance from phonon scattering dominates, so that the resistance has a non-monotonic behavior with a minimum roughly around $T \sim T_K$.

We note that while historically, the position of the resistance minimum was often taken as a definition of the Kondo temperature, this is unsatisfactory because this definition depends on the details of the phonon scattering and the prefactor R_0 in Eq. (4). Instead, the modern approach is to define the Kondo temperature independently of the resistance. To see how one might go about this, consider the higher scattering processes (beyond the first Born approximation), which are implicitly contained in the “...” in Eq. (4). In fact, Abrikosov showed [11] that these terms yield an even stronger divergence as $T \rightarrow 0$ because they scale as $[J\rho \log(k_B T/D)]^n$. Summing the most divergent terms, Abrikosov obtained the result for resistance [11]

$$R = \frac{R_0}{\left[1 + 2J\rho \ln \left(\frac{k_B T}{D} \right) \right]^2}. \quad (5)$$

The Kondo temperature may be defined as the characteristic temperature at which the resistance diverges, which results in the estimate

$$k_B T_K \sim D \exp \left(-\frac{1}{2J\rho} \right). \quad (6)$$

As mentioned earlier in the introduction, other physical quantities, such as the magnetic susceptibility, were also shown to diverge logarithmically as the temperature $T = T_K$. Clearly, the theory cannot be trusted for low temperatures $T \lesssim T_K$, and this became the stumbling block of the Kondo problem until Wilson’s numerical solution in 1974 [1]. To understand how Wilson’s solution works, we have to first introduce the concept of renormalization and study how it applies to the Kondo model, which will be dealt with in the next two sections.

We note parenthetically that the divergence in Eq. (5) only occurs for the antiferromagnetic sign of the Kondo interaction ($J > 0$); otherwise, the resistance becomes small and converges. We shall explain the physical reason behind this behavior when we study the scaling of the Kondo model in Section 3.

2 Concept of renormalization

Usually, physical phenomena take place on a wide energy scale in condensed matter systems, from the conduction electron bandwidth of the order of several electron-volts, down to the experimentally relevant temperature range of the order of 1 Kelvin ($1 \text{ K} \approx 10^{-4} \text{ eV}$). We are interested in the low-energy (also called infra-red) limit, and the question is how to arrive there starting from the model formulated at high energy scales. The crucial idea is that instead of focusing on the fine details of the high-energy model (such as the exact spatial dependence of the interactions), one can arrive at the low-energy properties by monitoring the behavior of the system as one slowly lowers the cutoff scale Λ , which has the meaning of the energy corresponding to the largest-energy excitations available. If the system has a well-defined low-energy limit, the low-energy excitations will remain immune to this renormalization of the cutoff, and the model will be described by the “fixed point” Hamiltonian. In this case, the entire continuous family of model Hamiltonians $H(\Lambda)$ is said to “flow to the fixed point” and they belong to the same *universality class*. The word “universality” here implies that the low-energy behavior is universal, in other words, independent of the details of the high-energy (ultra-violet) model.

This idea of elucidating the low-energy universal behavior is achieved by the so-called *renormalization group* procedure, which consists of two steps:

1. Rescale the energy cutoff $\Lambda \rightarrow \Lambda' = \Lambda/b$, where $b > 1$, and integrate out the degrees of freedom in the energy range $[\Lambda/b, \Lambda]$. This will result in the change of the Hamiltonian $H(\Lambda) \rightarrow H'$.
2. Rescale the energy scales back so that $\omega = b\omega'$ and the new Hamiltonian $H(\Lambda/b) = bH'$.

These two steps are then repeated successively and in the limit $b \rightarrow 1$, one will obtain a continuous evolution of the model Hamiltonian with Λ . Below, we shall apply this idea to the Kondo model following P.W. Anderson's “Poor Man's scaling” argument [5].

3 Poor man's scaling for the Kondo model

3.1 *T*-matrix description of scattering processes

Following the general renormalization group ideas outlined above, we progressively integrate out the electronic states at the edge of the conduction band in the energy range $[\Lambda - \delta\Lambda, \Lambda]$. The resulting Hamiltonian will depend on the running energy scale Λ :

$$H(\Lambda) = \sum_{|\varepsilon_k| < \Lambda} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + J_z(\Lambda) \sum c_{k'\sigma'}^\dagger \sigma_{\sigma'\sigma}^z c_{k\sigma} S_z + \frac{J_\pm(\Lambda)}{2} \sum \left(c_{k'\uparrow}^\dagger c_{k\downarrow} S_- + c_{k'\downarrow}^\dagger c_{k\uparrow} S_+ \right), \quad (7)$$

where the last two terms correspond to the original Kondo Hamiltonian but with the renormalized coupling constant $J(\Lambda)$. This procedure was first performed by Anderson and Yuval using a

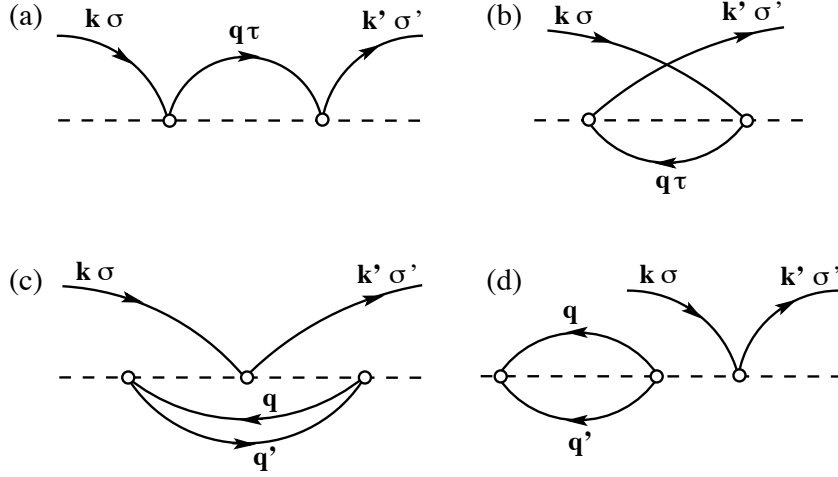


Fig. 2: Feynman diagrams contributing to (a,b) second-order processes in the Kondo interaction vertex (marked with an empty circle); and (c,d) third-order processes in the Kondo interaction. The solid lines denote the conduction electron propagator, whereas the dashed line denotes the impurity spin.

somewhat different method for a one-dimensional model equivalent to the Kondo model [12,13] and later reformulated by Anderson in a simplified form, which he called the “Poor Man’s” scaling approach [5]. The term “poor man” refers to the fact that the bandwidth is not rescaled to its original size after each progressive renormalization. This simplifies the matter as there is no need to rescale the Hamiltonian, eliminating the second step in the renormalization group procedure. Nevertheless, the results obtained via this simplified renormalization procedure are qualitatively accurate and correctly describe the low-energy behavior of the Kondo model.

Following Anderson, we integrate out the high-energy spin fluctuations using the formalism of the T -matrix, which describes the scattering of an electron from initial state $|k\rangle$ into the final state $|k'\rangle$. The matrix elements of such a scattering process constitute the so-called T -matrix, defined as a function of energy ω as follows:

$$T_{k',k}(\omega) = V_{k',k} + V_{k',q} G_0(\omega, \mathbf{q}) T_{q,k}(\omega) = \hat{V} + \hat{V} \frac{1}{\omega - \hat{H}_0} \hat{T}(\omega), \quad (8)$$

where $H_0 = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma}$ is the non-interacting conduction electron Hamiltonian, $\hat{V}$ is the Kondo exchange interaction, and G_0 is the non-interacting Green’s function. In what follows, we shall calculate the T -matrix to second-order in the Kondo interaction $\hat{V} \propto J$, in which case we can replace $\hat{T} \rightarrow \hat{V}$ in the last term in Eq. (8). This corresponds to renormalizing the interaction $\hat{V} \rightarrow \hat{V}'$ with

$$\hat{V}' = \hat{V} + \hat{V} \frac{1}{\omega - \hat{H}_0} \hat{V} = \hat{V} + \Delta \hat{T}. \quad (9)$$

Two kinds of processes contribute to the T -matrix at this order: (a) the electron is scattered directly, as the Feynman diagram in Fig. 2a illustrates; or (b) a virtual electron-hole pair is created in the intermediate state, see Fig. 2b. In both cases, the intermediate state may occur with or without flipping the spin of the conduction electron/hole. Let us first consider the case

when the conduction electron spin is $\uparrow$ both in the initial and in the final state. Consider first the simplest case when the conduction electron spin is not flipped in the intermediate state. The first process in Fig. 2a contributes

$$\Delta T_{\text{no-flip}}^{(a)}(\omega) = \sum_{\mathbf{q}}^{\Lambda > |\varepsilon_{\mathbf{q}}| > \Lambda - \delta\Lambda} (J_z)^2 S_z c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{q}\uparrow} (\omega - \varepsilon_{\mathbf{q}} + \varepsilon_{\mathbf{k}} - \hat{H}_0)^{-1} S_z c_{\mathbf{q}\uparrow}^\dagger c_{\mathbf{k}\uparrow} \quad (10)$$

It is understood that T is a matrix depending on the external momenta and spin polarizations $\{\mathbf{k}' \uparrow, \mathbf{k} \uparrow\}$; however, we drop these indices for brevity. If the energy ω is measured relative to the Fermi level μ , then $\hat{H}_0 = \sum_{k\sigma} (\varepsilon_k - \mu) \hat{n}_k$ can be set to zero in the ground state. Since the summation over $\mathbf{q}$ takes place in the narrow energy window $[\Lambda - \delta\Lambda, \Lambda]$, we can set $\varepsilon_{\mathbf{q}} \sim \Lambda$. Then, $c_{\mathbf{q}\tau} c_{\mathbf{q}\tau}^\dagger = 1 - \hat{n}_{\mathbf{q}}$ can be approximated as 1 in the particle-like intermediate state at low temperatures. Replacing the $\mathbf{q}$ -summation with an integration over the density of states ρ , we thus obtain

$$\Delta T_{\text{no-flip}}^{(a)}(\omega) = \frac{(J_z)^2 |\rho \delta\Lambda| S_z S_z c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow}}{\omega - \Lambda + \varepsilon_{\mathbf{k}}} = \frac{(J_z)^2 |\rho \delta\Lambda|}{4(\omega - \Lambda + \varepsilon_{\mathbf{k}})} c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow}, \quad (11)$$

where we have used $S_z^2 = 1/4$ for a spin 1/2 impurity. This term does not depend on the impurity spin and contributes to the potential scattering only, resulting in an overall energy shift. The same conclusion is reached in the case of the second type of scattering given by Fig. 2b. Such potential scattering is a new term absent from the original Kondo model in Eq. (7); however, it is irrelevant in the renormalization group sense and does not qualitatively alter the behavior of the model.

3.2 Renormalization of J_z

Let us now consider the physically more interesting case where the conduction electron is scattered from a $\uparrow$ to a $\downarrow$ state with a spin-flip in the intermediate state. The first process in Fig. 2a yields the following contribution to the T -matrix:

$$\Delta T_{\uparrow\downarrow}^{(a)}(\omega) = \sum_{\mathbf{q}}^{\Lambda > |\varepsilon_{\mathbf{q}}| > \Lambda - \delta\Lambda} J_+ J_- S^- c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{q}\downarrow} (\omega - \varepsilon_{\mathbf{q}} + \varepsilon_{\mathbf{k}} - \hat{H}_0)^{-1} S^+ c_{\mathbf{q}\downarrow}^\dagger c_{\mathbf{k}\uparrow}. \quad (12)$$

Similar to the earlier case, $\hat{H}_0$ can be set to zero in the ground state, and the intermediate state energy $\varepsilon_{\mathbf{q}} \sim \Lambda$. Given that $c_{\mathbf{q}\tau} c_{\mathbf{q}\tau}^\dagger = 1$ in the particle-like intermediate state at low temperatures, we thus obtain

$$\begin{aligned} \Delta T_{\uparrow\downarrow}^{(a)}(\omega) &= \sum_{\mathbf{q}}^{\Lambda - \delta\Lambda < \varepsilon_{\mathbf{q}} < \Lambda} J_+ J_- S^- S^+ c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow} (\omega - \varepsilon_{\mathbf{q}} + \varepsilon_{\mathbf{k}})^{-1} \\ &\approx J_+ J_- |\rho \delta\Lambda| S^- S^+ c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow} (\omega - \Lambda + \varepsilon_{\mathbf{k}})^{-1}. \end{aligned} \quad (13)$$

Similarly, the second process depicted in Fig. 2b yields

$$\begin{aligned} \Delta T_{\uparrow\uparrow}^{(b)}(\omega) &= \sum_{\mathbf{q}}^{-\Lambda < \varepsilon_{\mathbf{q}} < -\Lambda + \delta\Lambda} J_+ J_- S^+ c_{\mathbf{q}\tau}^\dagger c_{\mathbf{k}'\uparrow} (\omega + \varepsilon_{\mathbf{q}} - \varepsilon_{\mathbf{k}'})^{-1} S^- c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{q}\tau} \\ &\approx J_+ J_- |\rho \delta\Lambda| S^+ S^- c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow} (\omega - \Lambda - \varepsilon_{\mathbf{k}'})^{-1}, \end{aligned} \quad (14)$$

where we used the fact that in the hole-like intermediate state, the summation is near the lower band edge $[-\Lambda, -\Lambda + \delta\Lambda]$ and we can therefore replace $\varepsilon_q = -\Lambda$, with occupation number $c_{q\tau}^\dagger c_{q\tau} = 1$. We can now use the spin commutation relations on the impurity site to deduce that, for spin $1/2$, $S^- S^+ = 1/2 - S_z$, and similarly $S^+ S^- = 1/2 + S_z$ (we have set $\hbar = 1$ for convenience). We conclude that the expressions in Eq. (13) and (14) contribute to the renormalization of the $J_z S_z c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\uparrow}$ term in the Kondo Hamiltonian. Similar expressions, but with the opposite sign, can be obtained starting from the conduction electron in the spin $\downarrow$ state. We conclude that the J_z term in the Kondo interaction is renormalized as follows:

$$V'_z = (J_z + \delta J_z) \sum_{\mathbf{k}, \mathbf{k}'} \left(c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\uparrow} - c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\downarrow} \right) \cdot S_z, \quad (15)$$

with

$$\delta J_z = -J_+ J_- \rho |\delta\Lambda| \left[\frac{1}{\omega - \Lambda + \varepsilon_k} + \frac{1}{\omega - \Lambda - \varepsilon_{k'}} \right]. \quad (16)$$

Note the “−” sign in the above expression. Its importance will become apparent later when we discuss the renormalization flow for the coupling constants.

3.3 Renormalization of $J_\pm$

Finally, let us consider the scattering processes that contribute to the renormalization of the transverse ($J_\pm$) Kondo interaction. These are the processes that involve both the longitudinal and transverse terms, in which the electron is scattered from an initial state $\uparrow$ to a final state $\downarrow$ with a coherent flip of the impurity spin. Repeating the arguments similar to those used to derive Eqs. (13) and (14), one finds that the Feynman diagram in Fig. 2a results in the following contribution to the T matrix:

$$\Delta T_{\downarrow\uparrow}^{(a)}(\omega) = \frac{J_+(-J_z) |\rho \delta\Lambda| S_z S^+ c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}\uparrow}}{\omega - \Lambda + \varepsilon_k} + \frac{J_+ J_z |\rho \delta\Lambda| S^+ S_z c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}\uparrow}}{\omega - \Lambda + \varepsilon_k}. \quad (17)$$

The signs of the two terms are opposite because in the first expression, the spin-flip happens first, so that J_z term scatters two spin- $\downarrow$ states, resulting in the overall minus sign: $-J_z S_z c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{q}\downarrow}$, whereas in the second term the order of spin-flips is the opposite so that $J_z S_z c_{\mathbf{q}\uparrow}^\dagger c_{\mathbf{k}\uparrow}$ contributes with the positive sign. Using the identities $S_z S^+ = S^+/2$ and $S^+ S_z = -S^+/2$, we see that both terms contribute equally to the S^+ term:

$$\Delta T_{\downarrow\uparrow}^{(a)} = -\frac{J_+ J_z |\rho \delta\Lambda| S^+ c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}\uparrow}}{\omega - \Lambda + \varepsilon_k}. \quad (18)$$

Similarly, the diagram in Fig. 2b contributes in two ways

$$\Delta T_{\downarrow\uparrow}^{(b)}(\omega) = \frac{J_+ J_z |\rho \delta\Lambda| S_z S^+ c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\downarrow}}{\omega - \Lambda - \varepsilon_{k'}} + \frac{J_+(-J_z) |\rho \delta\Lambda| S^+ S_z c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\downarrow}}{\omega - \Lambda - \varepsilon_{k'}}. \quad (19)$$

Using the spin identities, we conclude that this results in

$$\Delta T_{\downarrow\uparrow}^{(b)}(\omega) = \frac{J_+ J_z |\rho \delta\Lambda| S^+ c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\downarrow}}{\omega - \Lambda - \varepsilon_{k'}} = -\frac{J_+ J_z |\rho \delta\Lambda| S^+ c_{\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}\uparrow}}{\omega - \Lambda - \varepsilon_{k'}}, \quad (20)$$

where the last equality is obtained by changing the order of the creation/annihilation operators (incurring a minus sign). Collecting together the contributions from Eq. (18) and (20), we find that J_+ is renormalized according to

$$\delta J_+ = -J_+ J_z \rho |\delta \Lambda| \left[\frac{1}{\omega - \Lambda + \varepsilon_k} + \frac{1}{\omega - \Lambda - \varepsilon_{k'}} \right]. \quad (21)$$

A similar result can be obtained for the renormalization of the J_- term, by considering the scattering from spin $\downarrow$ into spin $\uparrow$ state:

$$\delta J_- = -J_- J_z \rho |\delta \Lambda| \left[\frac{1}{\omega - \Lambda + \varepsilon_k} + \frac{1}{\omega - \Lambda - \varepsilon_{k'}} \right]. \quad (22)$$

3.4 Renormalization group flow

Summarizing our results so far, we conclude that elimination of the virtual scattering to the band edges results in a Hamiltonian that retains its Kondo form (neglecting the potential scattering terms such as Eq. 11). However, the coupling constants in Eq. (7) are renormalized as a result of integrating out the high-energy states: $J_\alpha \rightarrow J_\alpha + \delta J_\alpha$. It is said that J_α becomes a *running coupling constant*. By collecting the results obtained in Eqs. (16), (21), and (22) and assuming from now on that $J_+ = J_- = J_\pm$, we conclude that:

$$\delta J_z = -J_\pm^2 \rho |\delta \Lambda| \left[\frac{1}{\omega - \Lambda + \varepsilon_k} + \frac{1}{\omega - \Lambda - \varepsilon_{k'}} \right], \quad (23)$$

$$\delta J_\pm = -J_z J_\pm \rho |\delta \Lambda| \left[\frac{1}{\omega - \Lambda + \varepsilon_k} + \frac{1}{\omega - \Lambda - \varepsilon_{k'}} \right]. \quad (24)$$

The ω dependence underlines the fact that the renormalized interactions are retarded. However, for low-energy excitations relative to the conduction electron bandwidth or the cutoff Λ , the frequency dependence of the interactions can be neglected in the denominator. Similarly, since one is typically interested in the scattering of conduction electrons near the Fermi surface (on energy scales of the order of $k_B T$), the energies $\varepsilon_{k'}$ and ε_k can also be neglected compared to Λ . The resulting renormalization of the coupling constants can then be recast in terms of two coupled differential equations:

$$\frac{dJ_z}{d \ln \Lambda} = -2\rho J_\pm^2 \quad (25)$$

$$\frac{dJ_\pm}{d \ln \Lambda} = -2\rho J_z J_\pm. \quad (26)$$

Note that $\delta \Lambda$ is negative, and therefore $d(\ln \Lambda) = -|d\Lambda|/\Lambda$ in the above equations.

This logarithmic dependence of the coupling strength on the ultra-violet energy cutoff Λ is the essential idea behind the concept of the renormalization group. The above equations can be rewritten more conveniently by introducing the dimensionless coupling constants $g_\alpha \equiv J_\alpha \rho$ ($\alpha = z, \pm$) as follows:

$$\begin{aligned} \frac{dg_z}{d \ln \Lambda} &= -2g_\pm^2 + \mathcal{O}(g^3) \equiv \beta_z(g_\alpha) \\ \frac{dg_\pm}{d \ln \Lambda} &= -2g_z g_\pm + \mathcal{O}(g^3) \equiv \beta_\pm(g_\alpha). \end{aligned} \quad (27)$$

The right-hand side of these relations is called the *beta function*, using the established nomenclature. The isotropic case $J_z = J_\pm$ is particularly instructive, in which case we obtain

$$\frac{dg}{d \ln \Lambda} = -2g^2 + 2g^3 + \mathcal{O}(g^4), \quad (28)$$

where the second term on the right-hand side was obtained by considering the higher-order diagrams depicted in Figs. 2c and d.

Notice that to leading order in the coupling constant, the sign of the β -function in Eq. (28) is negative, meaning that as the energy cutoff Λ decreases, the corresponding coupling strength increases. For ferromagnetic interaction ($g < 0$), the coupling renormalizes to zero, $g \rightarrow 0$; however in the antiferromagnetic case, g remains positive and runs off to infinity as $\Lambda \rightarrow 0$. It is said that the theory tends towards strong coupling. This crucial difference between the ferromagnetic and the antiferromagnetic case is a quantum effect and should be understood as follows: If the impurity couples ferromagnetically to the conduction electrons (the so-called *s-d* model), the effect of such coupling becomes negligible at low temperatures. In other words, the impurity spin decouples from the conduction electron sea and becomes asymptotically free. In the case of antiferromagnetic (Kondo) interaction, on the other hand, the coupling is always relevant at low temperatures, no matter how weak the initial coupling strength. This means that a perturbative treatment of the Kondo model will break down at sufficiently low temperature of the order of the Kondo temperature T_K , and a non-perturbative approach is necessary to determine the low-temperature behavior. It was famously shown by Kenneth Wilson using numerical renormalization group (see Section 4.1) that the ground state of the Kondo model is a spin-singlet [1], forming due to the screening of the impurity spin by the conduction electrons. The antiferromagnetic Kondo model has a very interesting parallel with high-energy physics. In condensed matter physics, we are interested in the low-energy and low-temperature regime, i.e., the infra-red (IR) limit $\Lambda \rightarrow 0$, whereas high-energy particle physics concerns itself with the renormalization in the ultra-violet (UV) regime ($\Lambda \rightarrow \infty$). Bearing this distinction in mind, we note that the negative β -function is equivalent to the statement that the running coupling constants tends to zero as the energy cutoff Λ increases (provided $g > 0$ initially). This is similar to the celebrated phenomenon of the “asymptotic freedom” in quantum chromodynamics (QCD) where the interaction between quarks vanishes in the UV limit [14, 15]. For this reason, the Kondo impurity model is perhaps the simplest model that displays such behavior of the running coupling constant. Of course in condensed matter systems, the UV cutoff is not infinite as in QCD, but rather is fixed to be the conduction electron bandwidth D by the underlying crystalline lattice.

Returning to Eqs. (25-26), note that the following relation between J_z and $J_\pm$ is valid:

$$\frac{dJ_z}{dJ_\pm} = \frac{J_\pm}{J_z}, \quad (29)$$

or, equivalently, $J_z dJ_z = J_\pm dJ_\pm$. Integrating both parts of this equation, we conclude that

$$J_z^2 - J_\pm^2 = \text{const.} \quad (30)$$

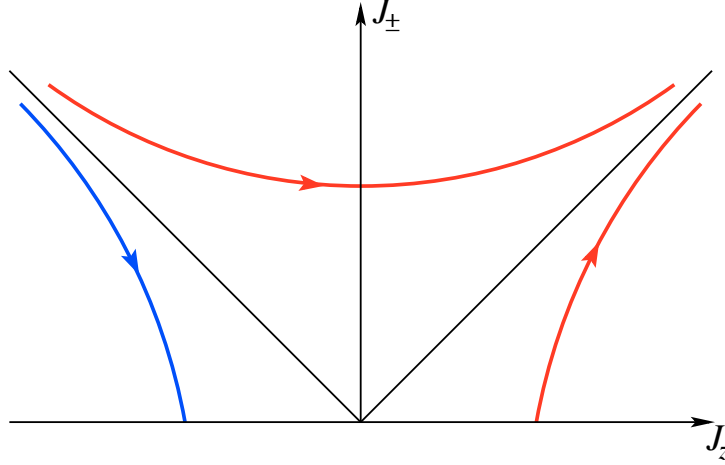


Fig. 3: Renormalization flow diagram of the anisotropic Kondo impurity model. On reducing the cutoff Λ , the coupling constants are scaled along the arrows. In the ferromagnetic region $J_z < 0$ and $|J_\pm| < |J_z|$, the system flows to weak coupling $J_\pm \rightarrow 0$ (blue arrow). In the rest of the parameter regime, the system flows towards the strong coupling regime $J_\pm \rightarrow \infty$ (red arrows).

This is an example of a *scaling law* that holds at any point in the renormalization flow. Consequently, the renormalization group preserves the hyperbolic trajectories expressed by Eq. (30) and depicted in Fig. 3. It follows from Eq. (25) that the β -function for J_z is always negative, meaning that J_z always grows upon renormalization. For antiferromagnetic Kondo interactions, this indicates that the model flows towards the strong-coupling fixed point $(J_z, J_\pm \rightarrow \infty)$, as mentioned above for the isotropic case. The ferromagnetic case $J_z < 0$ requires extra care because the outcome depends on the ratio of $J_\pm/|J_z|$. The case $J_\pm < |J_z|$ corresponds to the constant being positive in the scaling relation (30) and since the β -function for $J_\pm$ is positive, $J_\pm \rightarrow 0$ under the renormalization flow whereas $J_z < 0$ tends to a constant value, as indicated by the blue arrow in Fig. 3. In the other case $J_\pm > |J_z|$, $J_\pm$ initially decreases, however it follows the hyperbolic curve, and at some point J_z changes sign to positive, at which point both coupling constants run off to infinity.

3.5 Kondo temperature and breakdown of the perturbative scheme

Using the above scaling results, we can estimate the temperature scale at which the perturbative approach to the antiferromagnetic Kondo problem breaks down. In what follows, we shall consider the isotropic case $J_z = J_\pm$, in which case the β -function is given by Eq. (28). Integrating both sides of Eq. (28), we obtain:

$$-\int_g^{g^*} \frac{dg}{g^2 - g^3} = 2 \ln \Lambda \Big|_D^{A^*} = -2 \ln \left(\frac{D}{\Lambda^*} \right) \quad (31)$$

The integral in the left-hand side can be evaluated to give

$$-\int \frac{dg}{g^2 - g^3} = \frac{1}{g} + \ln \left| 1 - \frac{1}{g} \right| \quad (32)$$

We expect perturbation theory to fail once the dimensionless running coupling constant becomes large $g^*(\Lambda^*) \gg 1$. Then, the terms of the order $1/g^*$ can be ignored in the left-hand side of Eq. (31) and Eq. (32), resulting in the expression for Λ^*

$$\Lambda^* \sim D \frac{\sqrt{g}}{\sqrt{1-g}} \exp\left(-\frac{1}{2g}\right), \quad (33)$$

which we can identify with the Kondo temperature $k_B T_K \sim \Lambda^*$. Taking into account the fact that the unrenormalized value of $g = \rho J \sim J/D$ is much smaller than 1, we can approximate $\sqrt{1-g} \approx 1$, resulting in the well known expression for the Kondo temperature

$$k_B T_K \sim \sqrt{J\rho} D \exp\left(-\frac{1}{2J\rho}\right). \quad (34)$$

This expression is non-analytic in J , confirming that it cannot be obtained via perturbation theory. Note that had we limited ourselves to the second-order diagrams in J only (Fig. 2a,b), the β -function in Eq. (28) would contain only the $-2g^2$ term, and the corresponding expression for the Kondo temperature would have a slightly different form: $k_B T_K^{(0)} \sim D \exp(-1/2J\rho)$, which only differs by an algebraic prefactor from Eq. (34).

One might worry that higher-order terms in the diagrammatic expansion used to obtain the β -function could generate new terms that are not present in the original Kondo Hamiltonian. However, such terms would behave as a power-law of $(1/\Lambda)^n$, rather than $\ln \Lambda$, and so tend to zero rather than diverge as the cutoff $\Lambda \rightarrow \infty$ (or equivalently, the conduction electron bandwidth $D \rightarrow \infty$). Such higher-order terms are *irrelevant* in the RG sense as they do not affect the low-temperature properties of the Kondo problem.

4 Low-temperature properties of the Kondo model

4.1 Wilson's numerical renormalization

The above scaling argument can be used down to energy scales larger than the Kondo temperature. Beyond that point, the running coupling constant diverges and the theory predictions cannot be trusted. An important breakthrough in this very difficult problem was achieved by Wilson [1], who transformed the model into a form appropriate for computer modeling and used a numerical renormalization algorithm to deduce the properties of the system. Below, we will explain briefly Wilson's line of reasoning. In a spherically symmetric system such as the single-impurity Kondo model, arbitrary real-space interactions $V(\mathbf{r} - \mathbf{R})$ can be expanded in spherical harmonics centered around the impurity site $\mathbf{R}$. Assuming the Kondo interaction to be a δ -function $\delta(\mathbf{r} - \mathbf{R})$, only the s -wave harmonic contributes, allowing one to describe the system as effectively one-dimensional, depending on the radial distance $|\mathbf{r} - \mathbf{R}|$ from the impurity site. Wilson further assumed the conduction electron dispersion to be linear $\varepsilon_k = k$ (here the Fermi velocity was set to 1 in the appropriate units with $-1 \leq k \leq 1$) and replaced it with a spectrum of discrete levels $\varepsilon_n = \Lambda^{-n}$ equally distributed on the logarithmic scale (here we use

Wilson's original notation, Λ should not be confused with the UV cutoff of the previous section). Then, the Hamiltonian of this discrete-level system can be written as a one-dimensional tight-binding chain with the 0-th site corresponding to the impurity position:

$$\mathcal{H}_N = \Lambda^{(N-1)/2} \left\{ \sum_{n=0}^{N-1} \Lambda^{-\frac{n}{2}} (c_n^\dagger c_{n+1} + c_{n+1}^\dagger c_n) - \tilde{J} c_0^\dagger \boldsymbol{\sigma} c_0 \cdot \mathbf{S} \right\}, \quad (35)$$

with the original Hamiltonian obtained after rescaling and taking the limit of an infinitely long chain: $\hat{H} = \lim_{N \rightarrow \infty} \Lambda^{-(N-1)/2} \mathcal{H}_N$.

The prefactor $\Lambda^{(N-1)/2}$ in front of the Hamiltonian is necessary to keep the scale of low-energy excitations constant. Note also that the hopping matrix element is proportional to $\Lambda^{-n/2}$ and decays quickly as a function of the distance from the impurity site.

4.2 Ground state of the Kondo model

The renormalized Kondo interaction $\tilde{J} \cdot \Lambda^{(N-1)/2}$ becomes large as the number of sites N increases, corresponding physically to the formation of a spin-singlet state on the impurity site. Wilson showed by careful numerical simulations that this is indeed the ground state of the Kondo model. Wilson also calculated the ratio of the uniform magnetic susceptibility and the specific heat coefficient (now known as the Wilson ratio) to be

$$W \equiv \lim_{T \rightarrow 0} \frac{T\chi/\chi_0}{C/\gamma_0} = 2. \quad (36)$$

This result is significant because of the conventional Fermi liquid result, where one has

$$\frac{C_{\text{FL}}}{T} \equiv \gamma_0 = \frac{\pi^2}{3} k_B^2 \rho, \quad \chi_{\text{FL}} \equiv \chi_0 = \frac{g^2 \mu_B^2}{4} \rho, \quad (37)$$

resulting in the Wilson ratio $W_{\text{FL}} = 1$ (as before, ρ is the density of states at the Fermi level). In the Kondo impurity case, the Wilson ratio is doubled. The classic work by Nozières [2] explains this as follows: The low-energy excitations of the Kondo model can be understood in the framework of the Fermi liquid theory. However in contrast to the one-body problem (where Wilson's ratio is 1), the interaction between the impurity and an electron with antiparallel spins contributes to the antisymmetric Fermi liquid parameter ϕ^a , which, as Nozières showed, leads to an additional contribution to the Wilson ratio. In the Kondo model, where this interaction becomes infinitely strong in the low-temperature limit, this extra contribution results in the Wilson ratio being 2. This is in line with the more general case of the Anderson Hamiltonian, where the Wilson ratio increases from 1 to 2 with increasing interaction U between antiparallel spins [16].

The Kondo model turns out to be exactly solvable via the Bethe ansatz, as shown independently by N. Andrei [3] and P. Wiegmann [4]. The exact solution fully confirmed Wilson's earlier conclusion that the ground state of the Kondo model is the spin singlet and corroborated Nozières' Fermi liquid theory.

5 Multichannel Kondo problem

Having described the behavior of the spin $1/2$ Kondo impurity model above, it is natural to ask: What happens in the case of impurity spin S larger than $1/2$? This problem was first addressed by Nozières and Blandin in 1980 [17], who formulated what we now refer to as the multichannel Kondo impurity model:

$$H = \sum_{\mathbf{k}, \sigma, \mu} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma\mu}^\dagger c_{\mathbf{k}\sigma\mu} + J \sum_{\mu=1}^K \mathbf{S} \cdot \boldsymbol{\sigma}_\mu, \quad (38)$$

where $\boldsymbol{\sigma}_\mu = \sum_{\mathbf{k}} c_{\mathbf{k}\alpha\mu}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{\mathbf{k}\beta\mu}$ denotes the conduction electron spin in one of K orbital channels labeled by index μ . We shall only focus on the antiferromagnetic coupling J since the ferromagnetic case flows to weak coupling at low energies, as discussed earlier in Section 3.

5.1 Phenomenology and scaling

The multichannel model turns out to harbor rich physics, and depending on the number of conduction electron channels relative to the impurity spin size S , three different scenarios can be realized [17] (for a review, see also Ref. [18]):

- If $K = 2S$, the number of channels is exactly sufficient to fully compensate the impurity spin and the ground state is a spin singlet. This case, referred to as *perfect screening*, gives rise to the usual Fermi liquid behavior similar to the single-channel Kondo model discussed earlier in Section 4.
- If $K < 2S$, the impurity spin is not fully compensated since there are not enough conduction electron degrees of freedom. The dressed impurity remains magnetic with spin $S' = S - K/2$, resulting in the *underscreened* Kondo model.
- If $K > 2S$, the impurity spin is *overscreened*, resulting in the critical non-Fermi-liquid physics characterized by power-law or logarithmic behavior of thermodynamic quantities.

As the perfectly screened case needs little explanation, we will focus here on the latter two scenarios.

(1) Underscreened case $K < 2S$

Consider the spin $\mathbf{S}$ maximally polarized along the z axis, $|S_z = S\rangle$. Because of the antiferromagnetic coupling to K conduction electron channels (each with spin $1/2$), part of the impurity spin will be screened as the energy cutoff Λ becomes lower than the Kondo temperature. The remaining spin $S' = S - K/2$ will be pointing along the z axis as shown in Fig. 4a. This resulting spin can still interact with the conduction electrons because the latter can perform virtual hops onto the impurity site from neighboring sites with characteristic strength $|J'| \sim \Lambda^2/J$ obtained in second-order perturbation theory (here the running cutoff $\Lambda \ll J$ under the RG

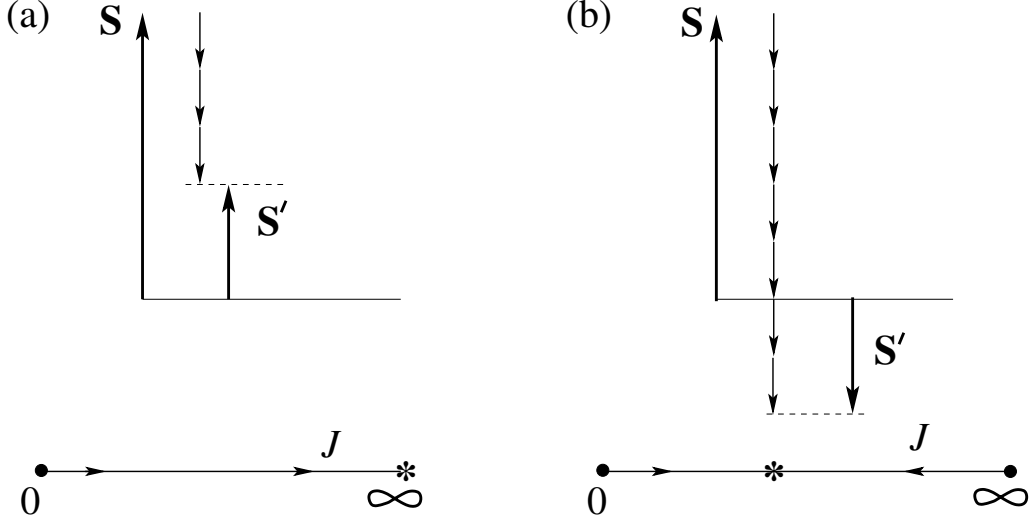


Fig. 4: Schematic depiction of the strong-coupling ground state of the impurity spin S and K conduction electron spins, together with the scaling trajectory for the running coupling constant J when (a) $K < 2S$, (b) $K > 2S$. A fixed point under the RG flow is denoted by asterisk (*).

process). The crucial point is that this coupling J' is *ferromagnetic*. Indeed, a nearby electron with $\downarrow$ spin cannot jump onto the central site because all available orbital channels are already occupied by spin- $\downarrow$ electrons as shown in Fig. 4a. Therefore, down spins do not interact with the unscreened impurity spin. By contrast, spin-up electrons can lower their energy in second-order perturbation theory by interacting with the unscreened spin S' that is also pointing up. Therefore, the residual coupling is indeed ferromagnetic and, as we know from Section 3, scales to weak-coupling under the RG flow, $J' \rightarrow 0$. Therefore, the $J \rightarrow \infty$ fixed point remains stable, as illustrated in the flow diagram in Fig. 4a.

(2) Overscreened case $K > 2S$

As in the previous case, the idea is to consider a two-stage process as the energy cutoff is reduced: First the Kondo singlet with the impurity spin forms, resulting from the strong coupling $J(\Lambda) \rightarrow \infty$. That leaves a residual interaction J' of the partially screened spin with the conduction sea. For a spin S along the z -axis, the conduction electron spins will “pile up” at the impurity site at sufficiently low energy Λ , generating an effective spin $S' = K/2 - S$ that is pointing down, opposite to the initial direction S_z (see Fig. 4b). Similar to the previous case, spin-down electrons do not participate in the virtual hopping onto the impurity site because all the spin-down states are already occupied (Pauli principle). Just as before, spin-up electrons can reduce their energy in second-order perturbation theory, generating an interaction $|J'| \sim \Lambda^2/J$ with the remaining impurity spin. The difference is that now S'_z is pointing *down*, so that the coupling J' to the conduction electrons is *antiferromagnetic*. However, we know that such an antiferromagnetic Kondo interaction scales to strong coupling as the cutoff is reduced, meaning that at sufficiently low Λ , the coupling $J'(\Lambda)$ is going to “blow up.” This is problematic because

our perturbative argument for $J' \sim \Lambda^2/J$ hinges on the fact that J' is small, otherwise perturbation theory does not converge. We conclude that the above two-stage RG process is untenable, meaning that the strong-coupling fixed point $J(\Lambda) \rightarrow \infty$ we assumed is actually unstable. The logical conclusion is that instead $J(\Lambda)$ should get renormalized to some finite value, leading to an *intermediate-coupling fixed point* denoted by the asterisk in Figure 4b.

Formally, one can see the appearance of the intermediate-coupling fixed point as follows. Consider the scaling equation obtained by summing the diagrams in Figures 2a-d. It has the form similar to Eq. (28) we derived in Section 3:

$$\frac{d(J\rho)}{d \ln \Lambda} = -2(J\rho)^2 + 2K(J\rho)^3 + \mathcal{O}(J\rho)^4, \quad (39)$$

except that the prefactor in the last term is now $2K$ instead of 2. This is because the closed loop in the diagrams in Fig. 2c and 2d contributes an additional factor of K due to the summation over the internal channel index $\mu = 1 \dots K$. Notice now that the beta-function on the right-hand side can be made to vanish at a *fixed point*

$$\rho J^* = \frac{1}{K}, \quad (40)$$

provided $J > 0$ (antiferromagnetic). When $K = 1$ as in the single-channel Kondo model, the result is meaningless because the expansion for the β -function in Eq. (39) does not converge. On the other hand for K large, the expansion becomes meaningful because every additional vertex yields a factor $J^* = 1/K$ and every additional loop yields a factor $K(J^*)^2 = 1/K$. Therefore, the expansion at the fixed point

$$\frac{d(\rho J^*)}{d \ln \Lambda} = -2\frac{1}{K^2} + 2K\frac{1}{K^3} + \frac{c}{K^4} + \frac{d}{K^5} + \dots \quad (41)$$

is well defined, thus making plausible the existence of the intermediate-coupling fixed point J^* . This conclusion, originally reached by Nozières and Blandin [17], was later confirmed by the exact solution obtained by the Bethe ansatz [19, 20] and by conformal field theory [21]. Unlike the strong-coupling fixed point in the one-channel Kondo model whose low-energy properties are described by Fermi-liquid theory [2], the intermediate-coupling case is characterized by its non-Fermi-liquid behavior. For instance, the exact solution of the multi-channel problem [19, 20] shows that the magnetic susceptibility and the specific heat both vary as

$$\chi \sim \frac{C}{T} \propto T^{\frac{4}{K+2}-1}, \quad K > 2 \quad (42)$$

In the particular case of $K = 2$, the power-law is replaced by a logarithmic temperature dependence:

$$\frac{C}{T} \sim \chi \propto \ln \left(\frac{T}{T_K} \right). \quad (43)$$

Of particular historical and practical importance is the (overscreened) two-channel Kondo problem for spin $S = 1/2$, which we shall analyze in more detail below.

5.2 Two-channel Kondo problem

The peculiarity of the intermediate-coupling fixed point predicted by Nozières and Blandin [17] is that at low temperatures, the running coupling constant flows to a fixed point with a finite value of J^* , regardless of how strong or weak the initial (bare) coupling is. This is remarkable because unlike the single channel model, nothing cuts off this scaling process and the impurity spin can never be screened. This means that there is no special energy scale analogous to the Kondo temperature in the single-channel model, and on approaching the fixed point, the system looks the same on all length scales. This inherent scale-invariance is the hallmark of a critical point, and like at any critical point, the correlation length ξ (the length at which the impurity spin affects the conduction electrons) diverges. It was shown by Affleck and Ludwig [21, 22] that this critical point is in fact described by a conformal field theory. Because of the criticality, various quantities are expected to scale as power-laws (or logarithmically, which can be viewed as a power-law with exponent zero). Indeed, as already mentioned, the intermediate coupling fixed point in the two-channel Kondo model is characterized by a logarithmic behavior of the magnetic susceptibility and the specific heat. The analysis by Ludwig and Affleck [22] further predicts the resistivity of the non-Fermi-liquid form:

$$\rho \sim \rho_0 + A\sqrt{T}. \quad (44)$$

A critical point generally requires fine-tuning, in other words, it may be destabilized by a relevant perturbation in the space of the model parameters. The two-channel Kondo model provides an illustrative example in that it is very sensitive to external perturbations. Below, we first consider the behavior under the application of the magnetic field that couples to the impurity spin, and then the effect of the channel anisotropy.

(a) Effect of the applied magnetic field

Because the impurity spin is never completely screened at the intermediate-coupling fixed point, the multichannel model displays a residual ground-state entropy. For the case of impurity spin $S = 1/2$, the residual entropy per impurity was calculated by the Bethe ansatz [19, 24]:

$$S(0) = \ln \left[2 \cos \left(\frac{\pi}{K+2} \right) \right]. \quad (45)$$

In the case of the two-channel model, the residual entropy $\frac{1}{2} \ln 2$ per impurity remains. However, application of an external magnetic field that couples to the impurity spin has a dramatic effect: It introduces a new energy scale $T_s \approx T_K (H/T_K)^{1+2/K}$ ($T_s \approx H^2/T_K$ in the two-channel case) that interrupts the scaling and below which the crossover to the screened, Fermi liquid behavior occurs. As a result, the residual entropy is removed in the $T \rightarrow 0$ limit. This is shown in Figure 5a, which displays the entropy as a function of temperature for several applied magnetic field strengths (reproduced from the numerical solution in Ref. [23]). This removal of the residual entropy has a spectacular signature in the temperature dependence of the specific heat. Above the crossover scale $T > T_s$, it behaves as a logarithm according to Eq. (43), but below this scale,

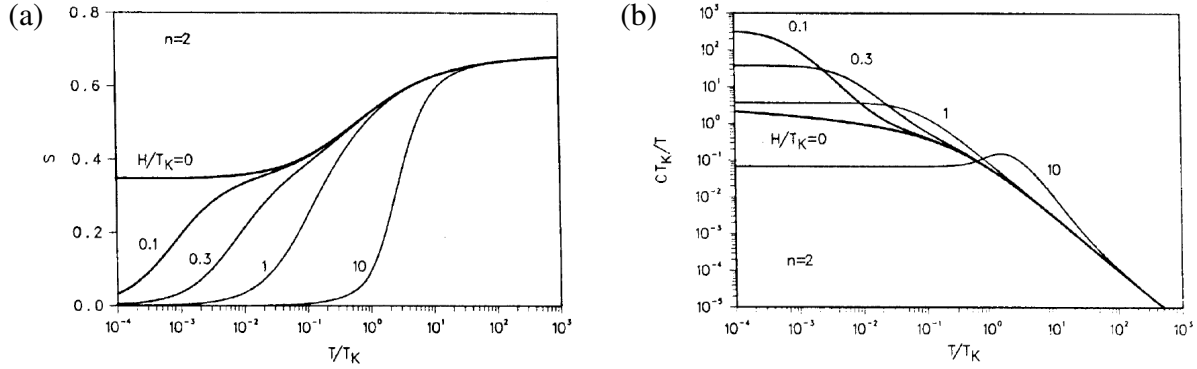


Fig. 5: Results for the two-channel Kondo model in an applied spin (magnetic) field: (a) entropy and (b) specific heat coefficient as function of temperature. The residual ground state entropy is released by the application of the field, while the specific heat develops a Schottky-like anomaly at the crossover scale T_s . (Reproduced from Ref. [23].)

the loss of the residual entropy manifests itself in a Schottky-like anomaly in the specific heat, shown in Fig. 5b. Notice that the value at the peak maximum can greatly exceed the $H = 0$ value of the specific heat, leading to a striking effect as a magnetic field is applied. A similar phenomenon occurs in the heavy fermion compound $Y_{1-x}U_xPd_3$, and it was suggested [25–27] that the multichannel Kondo model may provide an explanation, although of course one must remember that this is a dense Kondo lattice rather than an isolated impurity.

(b) Effect of the channel anisotropy

Unless there is a symmetry argument that requires the two orbital channels to couple to the impurity spin with identical strength, one might consider lifting this degeneracy by assigning two Kondo couplings $J_+ \neq J_-$:

$$H = \sum_{\mathbf{k}, \sigma} \sum_{\mu=\pm} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma\mu}^\dagger c_{\mathbf{k}\sigma\mu} + J_+ \mathbf{S} \cdot \boldsymbol{\sigma}_+ + J_- \mathbf{S} \cdot \boldsymbol{\sigma}_-. \quad (46)$$

Sometimes this is referred to as applying a “channel field” in that it splits the channels by an amount $\Delta J = J_+ - J_-$ similarly to the Zeeman splitting in the case of the “spin field.” It was argued that a real magnetic field can have this effect in the context of a quadrupolar Kondo effect [27]. Like in the case of the spin splitting described above, the channel anisotropy introduces a new crossover scale $T_{ch} \sim (\Delta J)^2/T_K$ that also cuts off the renormalization flow. However in this case, the consequences are much more dramatic: It was shown by Nozières and Blandin [17] that the more strongly coupled channel will tend towards the strong-coupling fixed point (as in the regular Kondo screening), whereas the weakly coupled channel will tend towards the zero-coupling fixed point. The resulting RG flow of the running coupling constants $J_\pm(\Lambda)$ is shown schematically in Fig. 6. The intermediate-coupling fixed point (marked by the black circle) is only stable along the $J_+ = J_-$ line but unstable for any small ΔJ . In particular, the fixed point is unstable along the separatrix shown with the dashed line in Fig. 6. The flow trajectories approach this separatrix on either side of the $J_+ = J_-$ line, depending on the sign of the bare ΔJ .

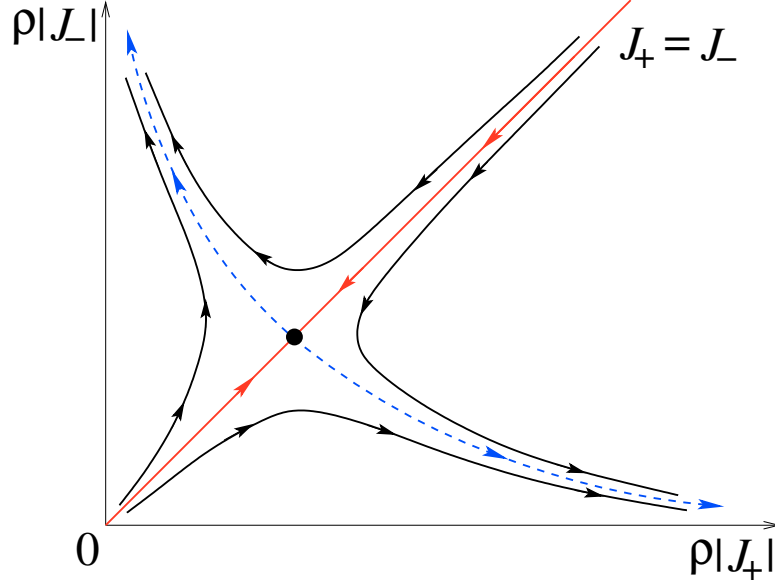


Fig. 6: Renormalization group flow of the anisotropic two-channel Kondo model for impurity spin $S = 1/2$ in the phase space of coupling constants. The flow to the intermediate coupling fixed point (filled circle) is stable only along the red line $J_+ = J_-$. The dashed blue line indicates the separatrix.

In conclusion, to observe the critical behavior of the two-channel spin-1/2 Kondo model and the associated non-Fermi liquid behavior, a perfect channel symmetry is required. In practice, this may be achieved if, for instance, the crystal point-group symmetry protects the system against the channel anisotropy. A conclusive experimental evidence of the non-Fermi-liquid behavior associated with the two-channel Kondo model is still lacking. This is partly due to the fact that the most promising candidates are in the dense Kondo lattice (rather than isolated impurity) limit. Nevertheless, the overscreened multichannel Kondo model displays rich physics and historically has played a very important role in the development of various theoretical tools used to study strongly correlated electron systems.

6 Kondo model in the presence of Hund's coupling

In the previous section, we have introduced the multi-channel Kondo impurity model. In particular, we stated that in the perfectly screened case $K = 2S$, the ground state of the problem is a spin-singlet, similar to the one-channel Kondo impurity model. A practical question arises, which we have not yet addressed: How does the Kondo temperature of such a multi-channel model depend on the size S of the impurity moment? Despite the deceiving simplicity, the answer to this question is not so simple and has not been fully appreciated until fairly recently, although the problem has a very interesting history dating back to the 1960s. One natural way of creating a large moment S on the impurity site is by Hund's coupling between $2S$ singly occupied orbitals, each with spin $s = 1/2$ (we shall use lower case s when referring to such a constituent impurity spin 1/2, not to be confused with the conduction electron spin σ). In this

section, we shall present the Poor Man's scaling theory of this problem and study its dependence on the size S of the impurity moment. The following discussion is based on the results of Ref. [28].

Let us consider K spin $s = 1/2$ impurity spins at a single site, ferromagnetically interacting via Hund's coupling J_H , each coupled to a conduction electron channel of bandwidth D via an antiferromagnetic interaction J :

$$H = \sum_{\mathbf{k}, \sigma, \mu} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma\mu}^\dagger c_{\mathbf{k}\sigma\mu} - J_H \left(\sum_{\mu=1}^K \mathbf{s}_\mu \right)^2 + J \sum_{\mu=1}^K \mathbf{s}_\mu \cdot \boldsymbol{\sigma}_\mu, \quad (47)$$

where $\varepsilon_{\mathbf{k}}$ is the conduction electron energy, $\mu = 1, \dots, K$ is the channel index and $\boldsymbol{\sigma}_\mu = \sum_{\mathbf{k}} c_{\mathbf{k}\alpha\mu}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{\mathbf{k}\beta\mu}$ is the conduction electron spin density in channel μ at the origin. We implicitly assume that Hund's scale KJ_H is smaller than D .

The behavior of this model is well understood in the two extreme limits [17]: for $J_H = \infty$, the K spins lock together, forming a K -channel spin $S = K/2$ Kondo model studied in the previous section. The opposite limit $J_H = 0$ describes K replicas of the spin-1/2 Kondo model. Paradoxically, the leading exponential dependence of the Kondo temperature on the coupling constant $k_B T_K \sim D e^{-1/2 J \rho}$ in these two limits is independent of the size of the spin. Naively, interpolating between these limits, one would conclude that the spin size does not enter into the Kondo temperature. However, it has been long known experimentally that this is not the case: the Kondo temperature of dilute d -metal impurities shows a striking dependence on the impurity spin, as shown in Fig. 7. The left panel is borrowed from the classic review paper by Daybell [29], based on the original experimental finding reported by Daybell and Steyert in 1968 [30]. What they noticed is that the Kondo temperature has a characteristic V-shape when plotted against the occupation of the d -electron level. For clarity, these data have been re-plotted as a function of the impurity spin in Fig. 7b, showing an impressive suppression of the Kondo temperature over five orders of magnitude when S is varied from $S = 1$ in Ti^{2+} and Ni^{2+} to $S = 5/2$ in Mn^{2+} .

Amazingly, this exponential dependence of the Kondo temperature on the impurity spin size had been predicted in 1967 by Schrieffer [31] before the experimental findings (in fact, Daybell and Steyert used Schrieffer's prediction to fit the data in Fig. 7a):

$$T_K^*(S) \approx D \exp \left(-\frac{2S}{2J\rho} \right) = T_K \left(\frac{T_K}{D} \right)^{2S-1}, \quad (48)$$

where $T_K \sim D e^{-1/2 J \rho}$ is the leading exponential term in the Kondo temperature for spin 1/2 (cf. Eq. 34), and we have set $k_B = 1$ for convenience. Schrieffer obtained this result in the limit of infinitely strong Hund's coupling J_H , which we will explain later in this section.

Intriguingly, the experimental results and Schrieffer's early work from the 1960s were largely forgotten and have been re-discovered much more recently by the author and P. Coleman [28], who used the framework of Anderson's Poor Man's scaling to deduce the exponential suppression of the Kondo temperature with the impurity spin and generalized Schrieffer's analysis to finite J_H . This interesting phenomenon, referred to as the *Kondo resonance narrowing* due to Hund's coupling, is the main subject of this section.

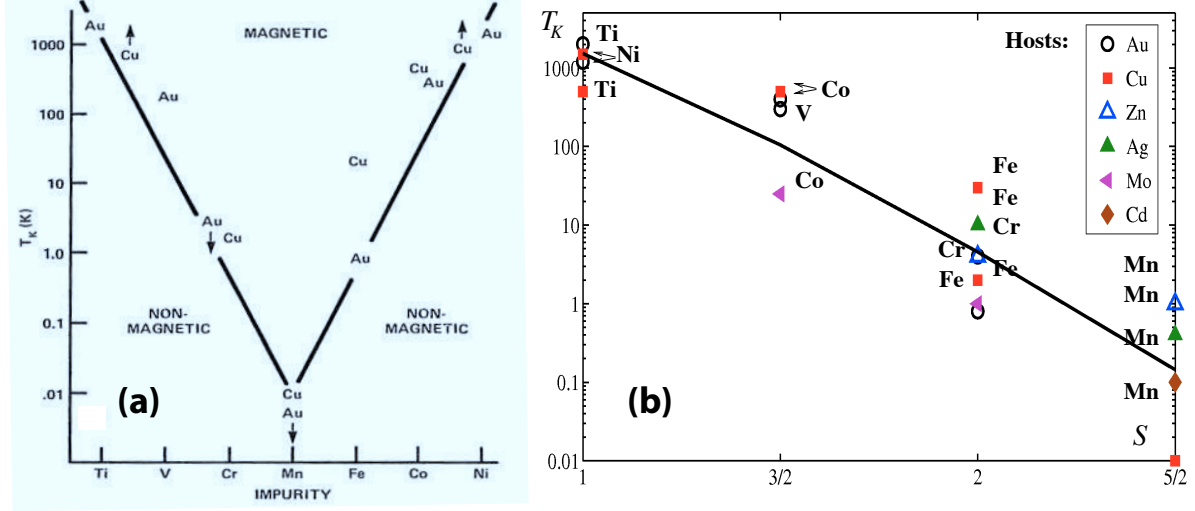


Fig. 7: Measured values of the Kondo temperature T_K^* in host alloys Au, Cu, Zn, Ag, Mo, and Cd containing transition metal impurities: (a) plotted vs. the d -level occupation of the impurity (Reproduced from Ref. [29]); (b) plotted vs. the nominal size S of the spin on the impurity site (Reproduced from Ref. [28]). Solid line is the fit to Eq. (63) with $\Lambda_0 \equiv J_H S$.

6.1 Poor man's scaling with Hund's coupling

We employ the Poor Man's scaling approach [13] described in detail in Section 3, in which the leading renormalization group (RG) flows are followed as the conduction electron degrees of freedom are systematically integrated out from the Hilbert space. The present exposition follows closely that in Ref. [28], where the scaling theory of the Kondo problem in the presence of Hund's coupling was first derived.

In the course of renormalization, one must be careful to consider the cutoff scale Λ relative to the other scales in the problem, in particular Hund's coupling J_H . We thus break up the energy integration into two intervals: (I) $J_H S < \Lambda < D$ and (II) $T_K^* < \Lambda < J_H S$, where T_K^* is the renormalized Kondo temperature (to be determined) below which the problem runs off to strong coupling.

Regime I: $J_H S < \Lambda < D$. In this regime Hund's coupling has no effect to leading order on the renormalization of the Kondo coupling J . In other words, the impurity $s = 1/2$ spins are decoupled from each other at high energies/temperatures, as illustrated schematically in Fig. 9a. We then arrive at the same equation (28) for the β -function obtained earlier in Section 3:

$$\frac{d(J\rho)}{d \ln \Lambda} = -2(J\rho)^2 + 2(J\rho)^3. \quad (49)$$

The Hund's coupling itself also gets renormalized, and the contributions to the β function are captured by the Feynman diagrams in Fig. 8. We shall not go into the details of calculating these diagrams but will quote the final result:

$$\frac{d(J_H \rho)}{d \ln \Lambda} = 4(J\rho)^2 J_H \rho \quad (50)$$

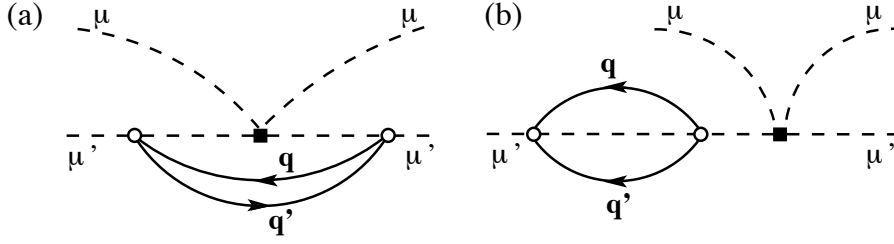


Fig. 8: The lowest-order Feynman diagrams in the RG flow of Hund's coupling. Solid lines denote the conduction electron propagators and dashed lines the impurity spins' (labeled by the channel number μ, μ'). A square vertex denotes Hund's coupling J_H , a circle the Kondo interaction J .

Thus the Hund's strength itself becomes a running coupling constant, whose value is renormalized down as the Kondo coupling $J\rho(\Lambda)$ grows:

$$\ln \left(\frac{\tilde{J}_H(\Lambda)}{J_H} \right) = -4 \int_{\Lambda}^D (\rho J(\Lambda))^2 d(\ln \Lambda). \quad (51)$$

This downward-renormalization of the Hund's coupling is, however, weak and to leading logarithmic order, we can approximate $J_H(\Lambda)$ to be constant.

Integrating both sides of Eq. (49) similarly to the procedure in Section 3.5, we find that as the cutoff Λ is reduced from the electron bandwidth D down to the Hund's scale $J_H S$, to leading logarithmic order we obtain a new renormalized Kondo coupling

$$\frac{1}{2J_I\rho} = \frac{1}{2J\rho} + \ln \Lambda \Big|_{\Lambda=D}^{J_H S}, \quad (52)$$

which grows upon renormalization as expected, $J_I > J$. Expressing the bare Kondo coupling in terms of the Kondo temperature $2J\rho = \ln^{-1}(D/T_K)$, we can rewrite the above expression as follows (we set $k_B = 1$ for convenience):

$$2\rho J_I = \ln^{-1} \left(\frac{J_H S}{T_K} \right). \quad (53)$$

Regime II: $T_K^* < \Lambda < J_H S$. Once Λ is reduced below $J_H S$, the individual local moments become locked into a spin $S = K/2$, as illustrated schematically in Fig. 9. This is the same effect as discussed by Jayaprakash *et al.* in Ref. [32] for the case of two impurities coupled by ferromagnetic RKKY interaction and as realized in the limit of $J_H \rightarrow \infty$ analyzed by Schrieffer in his 1967 paper [31].

The low-energy properties of the system in region II are described by a Kondo model of spin $K/2$ with K conduction electron channels:

$$H_{\text{eff}}^{\text{II}} = J^*(\Lambda) \sum_{\mu=1}^K \mathbf{S} \cdot \boldsymbol{\sigma}_{\mu}, \quad (54)$$

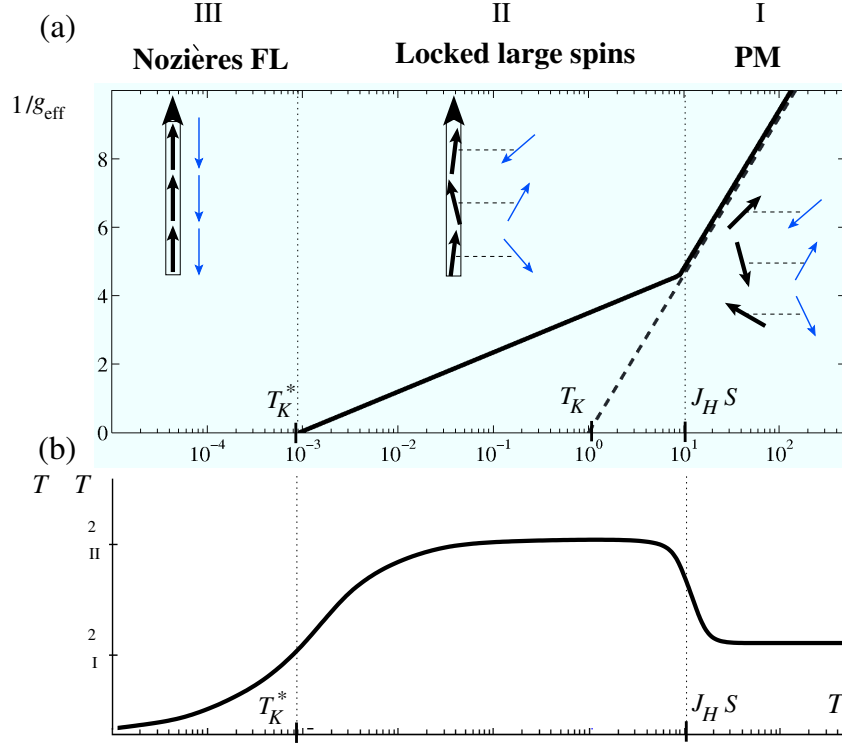


Fig. 9: (a) Schematic showing the behavior of the running coupling constant $g_{\text{eff}}(\Lambda) = J(\Lambda) \rho K_{\text{eff}}$ on a logarithmic scale, with K_{eff} the effective number of conduction electron channels per impurity spin ($K_{\text{eff}} = 1$ in region I and K in regions II and III). (b) Schematic showing effective moment $\mu_{\text{eff}}^2(T) = T \chi(T)$ in terms of the susceptibility $\chi(T)$ showing an enhancement in region II (see Eq. 68) and a loss of localized moments due to Kondo screening in region III.

however with the renormalized value of the Kondo coupling J^* . In order to obtain the value of J^* , we must project the original model onto the subspace of maximum spin S . By the Wigner-Eckart theorem, any vector operator acting in the basis of states $|S_z\rangle$ of spin $S = K/2$ is related by a constant prefactor to $\mathbf{S}$ itself:

$$\langle SS_z | \mathbf{s}_\mu | SS_z \rangle = g_S \langle S_z | \mathbf{S} | S_z \rangle. \quad (55)$$

Summing both sides of the equation over the impurity index $\mu = 1, \dots, K$, one obtains

$$\langle SS_z | \sum_{\mu} \mathbf{s}_\mu | SS_z \rangle = g_S K S_z. \quad (56)$$

However, since $\sum_{\mu} \mathbf{s}_\mu = K \mathbf{s} \equiv \mathbf{S}$, one arrives at the conclusion that $g_S K = 1$, hence determining the value of the constant coefficient $g_S = 1/K$ in Eq. (55). Therefore, the Kondo interaction in the original model (47) can be cast in the form

$$J \sum_{\mu=1}^K \mathbf{s}_\mu \cdot \boldsymbol{\sigma}_\mu = J g_S \sum_{\mu=1}^K \mathbf{S} \cdot \boldsymbol{\sigma}_\mu. \quad (57)$$

Comparing this equation with Eq. (54) and substituting $g_S = 1/K$, we arrive at the following expression for the effective Kondo coupling:

$$J^* = \frac{J}{K} \equiv \frac{J}{2S}. \quad (58)$$

This equation captures the key effect of the crossover from region I to region II in Fig. 9. This result was first derived in the early work on the multi-channel Kondo problem by Schrieffer [31], where the limit of $J_H \rightarrow \infty$ was implicitly assumed, and also appears for the particular case of $K = 2$ in the study of the two-impurity Kondo problem by Jayaprakash *et al.* [32].

Having established the value J^* of the coupling constant, we now proceed with the Poor Man's renormalization of the effective model in Eq. (54). The scaling equation for $J^*(\Lambda)$ in region II involves calculating the same diagrams for the T matrix (see Fig. 2) as we have done earlier for the multichannel Kondo model in Eq. (39), with the result

$$\frac{d(J^*\rho)}{d \ln \Lambda} = -2(J^*\rho)^2 + 2K(J^*\rho)^3, \quad (59)$$

where the prefactor K in the last term appears due to the summation over the channel index $\mu = 1 \dots K$ inside the conduction electron bubble in the diagrams in Fig. 2c and 2d.

To one loop order, the β -function for $J^*(\Lambda)$ in region II is identical to that of region I, see Eq. (49). However, now the value of J^* is K times smaller according to Eq. (58). To avoid the discontinuous jump in the coupling constant at the crossover from region I to region II in Figure 9, it is more convenient to consider a dimensionless coupling constant

$$g_{\text{eff}} \equiv J^*(\Lambda) \rho K_{\text{eff}}, \quad (60)$$

which is designed to be smooth at the crossover from I to II by requiring that the effective number of channels $K_{\text{eff}} = 1$ and K in regions I and II, respectively. It follows from Eq. (59) that this continuous variable satisfies

$$\frac{d g_{\text{eff}}}{d \ln \Lambda} = -\frac{2}{K_{\text{eff}}} g_{\text{eff}}^2 + \frac{2}{K_{\text{eff}}} g_{\text{eff}}^3, \quad (61)$$

so the speed at which it scales to strong coupling becomes K times smaller in region II (see Fig. 9a). Solving this RG equation to leading order, and setting $g_{\text{eff}}(\Lambda = T_K^*) \sim 1$, we obtain $T_K^* \sim (J_H S) (D/J_H S)^K e^{-\frac{K}{2J\rho}}$ for the renormalized Kondo scale. Comparing this with the bare Kondo scale $T_K \sim D e^{-1/2J\rho}$ and denoting $J_H S = \Lambda_0$, we deduce

$$T_K^* \sim \Lambda_0 \left(\frac{T_K}{\Lambda_0} \right)^{2S} \equiv T_K \left(\frac{T_K}{\Lambda_0} \right)^{2S-1}, \quad (62)$$

showing that the effective Kondo temperature is exponentially suppressed by a factor of $(2S - 1)$ compared to its bare value for spin 1/2. When plotted on a logarithmic scale, the Kondo temperature is expected to scale linearly with the size of the impurity spin:

$$\ln T_K^*(S) = \ln \Lambda_0 - (2S) \ln \left(\frac{\Lambda_0}{T_K} \right), \quad (63)$$

which fits well the experimental data by Daybell and Steyert [30] for dilute d -electron impurities in Fig. 7.

Regime III: $A < T_K^*$. When the energy cutoff (or temperature) drops below the Kondo temperature in Eq. (62), the Poor Man's scaling predicts that the Kondo coupling runs off to infinity, as discussed in detail in Section 3. In this regime, the perturbative renormalization scheme employed above breaks down and alternative methods are necessary to establish the fate of the ground state.

In essence, Hund's coupling converts a one channel Kondo model (more precisely, K copies of the one-channel model) to a K -channel Kondo model but with a larger impurity spin $S = K/2$. As discussed in Section 5, the behavior of the perfectly screened multi-channel Kondo model is known from the seminal work of Nozières and Blandin [17] to be a Fermi liquid, with the conduction electrons screening the impurity spin. Of particular interest is the Wilson ratio defined in Eq. (36), which for the K -channel Kondo model was shown [17] to be

$$W = \frac{2(K+2)}{3}, \quad (64)$$

reducing to the value $W = 2$ in the one-channel case [2]. The above result holds in the limit of infinitely strong Hund's coupling. For a finite value of J_H , a crossover occurs between a 1-channel (region I) and a K -channel Kondo model (region II), as illustrated in Fig. 9a. While the Wilson ratio is difficult to determine in this case, we expect its value to depend on the ratio $\eta = U^*/J_H^*$, where U^* is the effective intra-channel interaction in the underlying Anderson model in the limit $T \rightarrow 0$, and J_H^* is the renormalized Hund's coupling (recall that Hund's coupling is a running coupling constant whose value is not constant, see Eq. (50)). We estimate [28] the Wilson ratio to be

$$W(\eta) = 2 \left(1 + \frac{K-1}{1+2(1+\eta)} \right), \quad (65)$$

which reproduces the result of Eq. (64) in the $\eta \rightarrow 0$ limit ($J_H \rightarrow \infty$).

6.2 Experimental ramifications of Hund's coupling

A. Effective moment crossover as a function of temperature

One of the observables sensitive to the Kondo narrowing is the effective spin on the impurity site, which we expect to change from spin 1/2 in regime I at high temperatures ($T > J_H S$) to effective spin $S = K/2$ at temperatures much lower than J_H . The effective impurity moment can be extracted from the Curie magnetic susceptibility, which we estimate to be:

Regime I: A spin-1/2 disordered paramagnet is characterized by a high temperature Curie magnetic susceptibility (g is the electron g-factor):

$$\chi_1(T) = \frac{K}{2} \left(\frac{1}{2} + 1 \right) \frac{(g\mu_B)^2}{3T}, \quad (66)$$

with effective moment $(\mu_{\text{eff}}^I)^2 = 3K/4$ in the conventional units of $(g\mu_B)^2$.

Regime II: The magnetic impurity susceptibility in region II can be calculated perturbatively for $T \gg T_K^*$ by keeping the leading logarithms $\ln(T/T_K)$ (see e.g. Ref. [7]):

$$\chi_{\text{imp}}^* = \frac{(g\mu_B)^2}{3T} S(S+1) \left(1 - \frac{1}{\ln\left(\frac{T}{T_K^*}\right)} + \mathcal{O}\left(\frac{1}{\ln^2\left(\frac{T}{T_K^*}\right)}\right) \right). \quad (67)$$

Substituting $S = K/2$, we see that the magnetic moment is enhanced at the crossover from region I to II as expected, with the ratio given by

$$(\mu_{\text{eff}}^{\text{II}}/\mu_{\text{eff}}^{\text{I}})^2 = (K+2)/3 \quad (68)$$

Note that the above ratio is equal, up to the factor of 2, to the Wilson ratio in Eq. (64), which was obtained in the idealized $J_H \rightarrow \infty$ limit.

Regime III: Below the Kondo scale T_K^* , the Curie contribution to the susceptibility vanishes because the impurity moment becomes completely screened, as already mentioned above.

The temperature behavior of $T\chi(T)$, which is proportional to the effective moment squared, is plotted schematically across these three regimes in Fig. 9b. As described above, it shows an enhancement of the effective moment on the crossover from regime I to II and eventually vanishes due to Kondo screening as $T \rightarrow 0$ in regime III.

B. Suppression of the Kondo temperature

Next, we return to the behavior of the Kondo temperature, Eq. (62) which is the central result of this section. It follows from fitting the experimental data in Fig. 7 to our formula Eq. (63), that the bare value of T_K is of the order of 3000 K ($k_B T_K \sim 0.3$ eV), too large to observe for dilute spin-1/2 impurities. However, the Kondo temperature is drastically suppressed for larger values of spin, for instance $T_K^* \approx 20$ K for Fe impurities ($S = 2$) in Cu [30], from which we extract the ratio $T_K/\Lambda_0 \equiv T_K/(J_H S) \sim 0.2$ consistent with the known value of Hund's coupling $J_H \sim 0.7$ eV in d -electron metals. In the case of Mn impurities ($S = 5/2$), the suppression is so dramatic that T_K^* is unobservably low, probably in the milli-Kelvin range. In this regard, it is fitting to quote a visionary remark from Schrieffer's 1967 paper [31] (written before the experimental discovery!):

“Since $\exp(-1/J^*\rho)$ varies so rapidly with $J^*\rho$, is it not possible that for modest variations of this parameter, T_K could sweep from millidegrees to beyond the melting point of metals?”

C. Kondo resonance narrowing

The exponential suppression of the Kondo temperature as a function of impurity spin in the limit of large Hund's coupling has an important consequence for the electron spectral function $A_{\mathbf{k}}(\omega) \equiv -\frac{1}{\pi} \text{Im}(G_{\mathbf{k}}^R(\omega))$, where G^R is the retarded propagator. Detailed analysis of the spectral function is beyond the subject of this lecture, but for our purposes, it is sufficient for the reader to know that at low temperatures $T \ll T_K$, the formation of the spin singlet in the Kondo

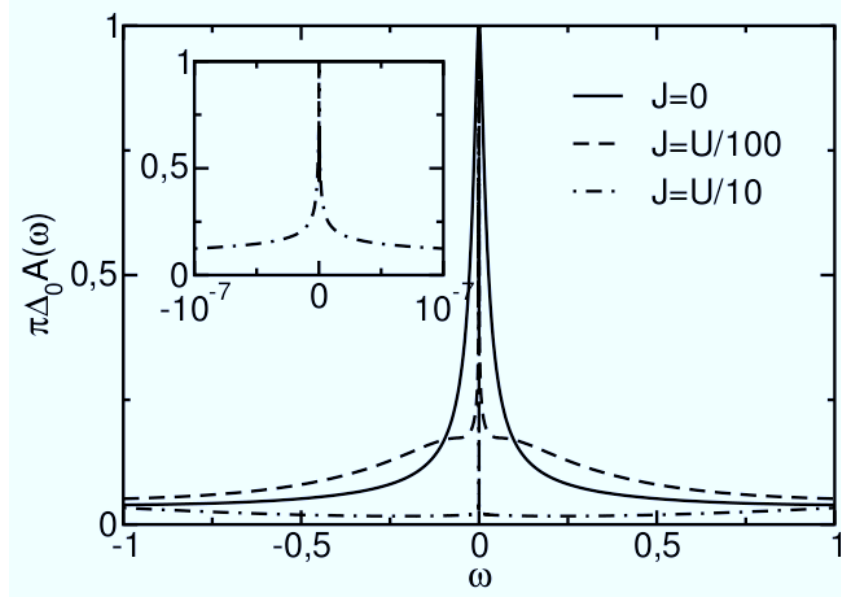


Fig. 10: Local electron density of states calculated with NRG at $T = 0$ for a particle-hole symmetric two-orbital Anderson impurity model with varying strength of Hund's coupling J . The Coulomb parameter U and the hybridization Δ_0 were chosen such that the system is in the Kondo limit. The Kondo resonance width is dramatically reduced upon increasing J (see inset for $J = U/10$). (Reproduced from Ref. [33])

impurity model manifests itself in a sharp peak near zero energy, known as the Abrikosov-Suhl resonance. The width of this peak is known to be of the order of the Kondo temperature. Thus, the suppression of the Kondo temperature because of Hund's coupling on the impurity site should result in the commensurate narrowing of the Kondo resonance peak. Direct experimental measurement of the Kondo resonance is not an easy task, however resonance narrowing was noticed in numerical renormalization group study of the two-orbital Anderson model by Pruschke and Bulla [33], reproduced in Figure 10. A dramatic resonance narrowing, by a factor of 10^{-7} , occurs by introducing a large value of $J_H = U/10$, where U is the interaction strength in the Anderson impurity model. While not understood at the time, this numerical evidence is in perfect agreement with the Kondo resonance narrowing effect described above. More recently, a number of numerical studies have confirmed this effect (for a comprehensive review, see Ref. [34]).

In itinerant systems, such as the iron-based superconductors, Hund's coupling was also found to play an important role, enhancing the effect of electron correlations. In particular, the so-called coherence temperature T^* , as inferred from the broad “hump” in resistivity measurement, was predicted to decrease drastically as the Hund's coupling strength is increased in numerical calculations [35]. Of course in this case, the localized moments are periodically arranged (as described by the Hubbard or periodic Anderson model) rather than centered on isolated impurity sites. Nevertheless it is tempting to associate the coherence temperature T^* with the typical lattice Kondo scale T_K^* . The latter should be suppressed dramatically by the Hund's coupling, explaining qualitatively the behavior in the Hubbard model. For more details, we refer the reader to the review by Georges *et al.* [34].

Conclusions

In this Lecture, we have outlined the basic ideas of the renormalization group (RG) approach and shown how they apply, in the simplest form, to the Kondo impurity model. This simplified renormalization procedure, known as Poor Man's scaling, captures the salient features of the Kondo problem. In particular, we showed how logarithmic singularities arise and how they can be summed up to result in the RG equation for the running coupling constant $J(\Lambda)$ that is cutoff-dependent. The essential feature of the antiferromagnetic Kondo model is that the running coupling constant grows upon renormalization and the resulting theory tends to strong coupling. At that point, the perturbative expansion in J is no longer reliable and an alternative approach is called for. We now know that the ground state of the Kondo model turns out to be a Fermi liquid, with impurity spin completely compensated (screened) by the conduction electrons. This important result was first obtained by Wilson's numerical renormalization and Nozières' phenomenological theory, and later confirmed by the exact solution of the Kondo model.

Having established these results for the simplest Kondo model, we then considered somewhat more complicated models that arise in realistic systems: the multichannel Kondo problem and the effect of Hund's coupling at the impurity site. Naturally, given the time and space constraints, we have only scratched the surface in analyzing these more complex cases. The reader is referred to the original papers and review articles for more detailed information, as well as to an excellent textbook by Yamada [6] and an encyclopedic monograph by Hewson [7]. Nevertheless, we hope to have given the reader a sense of what the Kondo problem is and how one goes about solving it within the scaling approach. The appeal of the Kondo problem also lies in the fact that, in addition to its rich underlying physics, the ideas we have developed in the course of this lecture are quite general and apply to many other branches of physics, both in condensed matter and in high-energy physics. Personally, I find this very satisfying and perhaps this accounts for the reason why, in addition to its historical significance, the Kondo model is such a fascinating topic. Finally, let me add that the lattice generalization of this model, the so-called Kondo lattice model, is still unsolved and remains an active subject of research to date, with important consequences for heavy fermion materials (see lecture by Piers Coleman in this School).

Acknowledgements

The author is grateful for the hospitality of the Institute of Solid State Physics, University of Tokyo, Japan where he held a Visiting Professor appointment in summer 2015 during which this Lecture was conceived.

References

- [1] K.G. Wilson, Rev. Mod. Phys. **47**, 793 (1975)
- [2] P. Nozières, J. Low Temp. Phys. **17**, 31 (1974)
- [3] N. Andrei, Phys. Rev. Lett. **45**, 379 (1980)
- [4] P.B. Wiegmann, Sov. Phys. JETP Lett. **3**, 392 (1980)
- [5] P. W. Anderson, J. Phys. C **3**, 2436 (1970)
- [6] K. Yamada: *Electron Correlations in Metals* (Cambridge University Press, 2004)
- [7] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993)
- [8] W.J. de Haas and G.J. van den Berg, Physica **3**, 440 (1936)
- [9] J. Kondo, Prog. Theor. Phys. **32**, 37 (1964)
- [10] J. Kondo, Solid State Phys. **23**, 183 (1969)
- [11] A.A. Abrikosov, Physics **2**, 5 (1965)
- [12] P.W. Anderson and G. Yuval, Phys. Rev. Lett. **23**, 89 (1969)
- [13] P.W. Anderson and G. Yuval, Phys. Rev. **131**, 1522 (1970)
- [14] D. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973)
- [15] H. Politzer, Phys. Rev. Lett. **30**, 1346 (1973)
- [16] K. Yamada, Prog. Theor. Phys. **53**, 970 (1975)
- [17] P. Nozières and A. Blandin, J. Phys. (Paris) **41**, 193 (1980)
- [18] P. Schlottmann and P.D. Sacramento, Adv. Phys. **42**, 641 (1993)
- [19] N. Andrei and C. Destri, Phys. Rev. Lett. **52**, 364 (1984)
- [20] A.M. Tsvelik and P.B. Wiegmann, Z. Phys. B: Condens. Matter **54**, 201 (1984)
- [21] I. Affleck and A.W.W. Ludwig, Nucl. Phys. B **352**, 849 (1991)
- [22] A.W.W. Ludwig and I. Affleck, Phys. Rev. Lett. **67**, 3160 (1991)
- [23] P.D. Sacramento and P. Schlottmann, Phys. Rev. B **43**, 13294 (1991)
- [24] A.M. Tsvelik, J. Phys. C **18**, 159 (1985)

- [25] C.L. Seaman, M.B. Maple, B.W. Lee, S. Ghamaty, M.S. Torikachvilli, J.-S. Kang, L.Z. Liu, J.W. Allen, and D.L. Cox, *Phys. Rev. Lett.* **67**, 2882 (1991)
- [26] C.L. Seaman, M.B. Maple, B.W. Lee, S. Ghamaty, M.S. Torikachvilli, J.-S. Kang, L.Z. Liu, J.W. Allen, and D.L. Cox, *J. Alloys Compounds* **181**, 327 (1992)
- [27] D. Cox and A. Zawadowski, *Adv. in Phys.* **47**, 599 (1998)
- [28] A.H. Nevidomskyy and P. Coleman, *Phys. Rev. Lett.* **103**, 147205 (2009)
- [29] M. Daybell: In G. Rado and H. Suhl (eds.) *Magnetism* (Academic Press, New York, 1973), Vol. 5, pp. 122–147
- [30] M.D. Daybell and W.A. Steyert, *Rev. Mod. Phys.* **40**, 380 (1968)
- [31] J.R. Schrieffer, *J. Appl. Phys.* **38**, 1143 (1967)
- [32] C. Jayaprakash, H.R. Krishna-murthy, and J.W. Wilkins, *Phys. Rev. Lett.* **47**, 737 (1981)
- [33] T. Pruschke and R. Bulla, *Eur. Phys. J. B* **44**, 217 (2005)
- [34] A. Georges, L. de' Medici, and J. Mravlje, *Ann. Rev. Condens. Matter Phys.* **4**, 137 (2013)
- [35] K. Haule and G. Kotliar, *New J. Phys.* **11**, 025021 (2009)

5 Numerical Renormalization Group and Multi-Orbital Kondo Physics

T.A. Costi

Institute for Advanced Simulation (IAS-3)

Forschungszentrum Jülich

Contents

1	Introduction	2
2	Quantum impurity models	3
3	Wilson's Numerical Renormalization Group method	10
4	Calculation of physical properties	17
5	Complete basis set and full density matrix	24
6	Recent developments: TDNRG and multiorbital Kondo physics	30
7	Summary	34
A	Logarithmic discretization approximation	35
B	Lanczos procedure	36

1 Introduction

This lecture deals with a particular implementation of the renormalization group idea: Wilson's non-perturbative numerical renormalization group (NRG) method for quantum impurity models [1]. The technique was originally developed in the context of the Kondo model for magnetic impurities (such as Fe or Mn) in non-magnetic metals (such as Cu, Au, Ag etc).

The Kondo model is defined by the Hamiltonian

$$H_{\text{KM}} = J \vec{S} \cdot \vec{s}_0 + \sum_{k\sigma} \varepsilon_{k\sigma} c_{k\sigma}^\dagger c_{k\sigma}, \quad (1)$$

and describes a localized impurity spin $\vec{S}$ interacting antiferromagnetically ($J > 0$) with the conduction electrons of the host via their spin density $\vec{s}_0$ at the impurity site. Unlike the case of non-magnetic impurities, or potential scatterers, magnetic impurities have internal dynamical degrees of freedom that result in inelastic scattering of conduction electrons. This makes the Kondo problem, the scattering of electrons from magnetic impurities, a genuine many-body correlation problem. Wilson used the NRG to solve the many-body Hamiltonian (1) and demonstrated conclusively that a $S = 1/2$ magnetic impurity embedded in a non-magnetic metal has its magnetic moment completely screened by the surrounding conduction electrons, provided the temperature is sufficiently low, namely for $T \ll T_K$, where $T_K = \sqrt{JN_F} e^{-1/JN_F}$ is a dynamically generated low-energy scale called the Kondo scale (see Sec. 2). This pioneering work established the formalism and gave a detailed analysis of the fixed points and thermodynamics of the Kondo model and, later, also of the Anderson impurity model. The NRG has since been applied to many more quantum impurity models [2–4]. In addition, it has been extended to the calculation of equilibrium dynamical and transport properties [5–10], e.g., dynamical susceptibilities, resistivities and thermopower or the conductance through quantum dots [11], thereby making the NRG a useful tool for interpreting experiments that probe these quantities.

Despite this progress, the NRG is still under development, and important challenges remain to be addressed. Two such challenges are (i) to extend it to more realistic multi-orbital and multi-channel models (e.g., for use in realistic modeling of materials), and (ii) to extend it to the transient and non-equilibrium steady state response of quantum impurity systems. Recent progress and ideas in these two directions are outlined in Sec. 6.

The outline of this lecture is as follows: Quantum impurity models are introduced in Sec. 2; the linear chain representation of such models is described, and the first step in the NRG procedure is also indicated there (the “zeroth approximation”). Anderson impurity and Kondo models are described, as is the spin-boson model and its fermionic equivalents: the anisotropic Kondo model (AKM) and the interacting resonant level model (IRLM). For a direct treatment of bosonic models within NRG, see the Lecture by K. Ingersent.

Wilson's NRG method is described in Sec. 3, and the calculation of physical properties is outlined in Sec. 4. In Sec. 5, we describe the recently introduced complete basis set [12] and its use in constructing the full density matrix [10]. Applications to thermodynamics and Green functions are given. An outline of some recent developments using the NRG is given in Sec. 6, and Sec. 7 summarizes with possible future directions.

2 Quantum impurity models

Quantum impurity models describe systems where the many-body interaction acts at one or only a few sites, the “impurity,” and the impurity is coupled to a large system, the bath, consisting of a macroscopically large number of non-interacting particles. These particles can be either bosons (e.g., phonons, magnons, photons, particle-hole pairs, etc.) or fermions (e.g., electrons in the conduction band or quasiparticles in superconductors). The “impurity” may be a real impurity, such as an Fe impurity (in Au), or a two-level atom (coupled to the electromagnetic field), or just a confined region behaving like an artificial atom, as in the case of semiconductor quantum dots (coupled to leads). It may also simply represent the lowest two quantum mechanical states of a system with a double-well potential, as in the case of quantum tunneling between macroscopic fluxoid states in a superconducting quantum interference device, which can be used to realize a qubit for quantum computation. Two magnetic impurities in a non-magnetic metal at a distance R apart, interacting via the RKKY indirect exchange J_{RKKY} may also be regarded as a quantum impurity system [13]. Analogues of this in nanostructures, such as double quantum dots attached to leads, also exist. The transfer of electrons between donor and acceptor molecules in photosynthesis and other biological processes may also be crudely described in terms of a two-state system coupled to environmental degrees of freedom (solvent). Concrete models describing the above situations go under the names of (isotropic and anisotropic) single and multi-channel Kondo models, the Anderson impurity model, and the dissipative two-state system [14, 15]. They describe a large number of physical systems of current experimental and theoretical interest. Quantum impurity models are also of relevance in the study of correlated lattice models, such as the Hubbard or Kondo lattice models, since the latter are often well approximated, via the dynamical mean-field theory, by a local impurity model embedded in a medium that has to be determined self-consistently [16].

Historically, interest in quantum impurities arose when magnetic impurities were found to be present, albeit in very low concentrations, even in apparently very pure metals such as Au or Ag. In particular, measurements of the resistivity of Au as early as the 1930’s showed an unexpected minimum at low temperature (Fig. 1). The puzzle of the resistivity minimum was resolved by Kondo in 1964, who showed that a small concentration c_{imp} of *magnetic* impurities modeled by Eq. (1) gives rise to an additional temperature dependent term in the resistivity of the form $\rho_K = -c_{\text{imp}} b \ln(T/D)$, which increases with decreasing temperature. The balance between the decreasing phonon contribution behaving as $\rho_{\text{phonon}} = aT^5$ and the increasing Kondo contribution gives rise to the observed resistivity minimum. The logarithmic contribution to the resistivity, found by Kondo in perturbation theory, cannot hold down to $T = 0$ as the total scattering remains finite in this limit (unitarity limit). Wilson’s non-perturbative NRG provides a way to obtain the correct behavior of the resistivity $\rho(T)$ from high temperatures through a crossover regime at $T \sim T_K$ all the way down to zero temperature [see Fig. 11a showing the analogous quantity for a Kondo correlated quantum dot, the conductance $G(T)$].

The general form of the Hamiltonian for any quantum impurity system is given by

$$H = H_{\text{imp}} + H_{\text{int}} + H_{\text{bath}}, \quad (2)$$

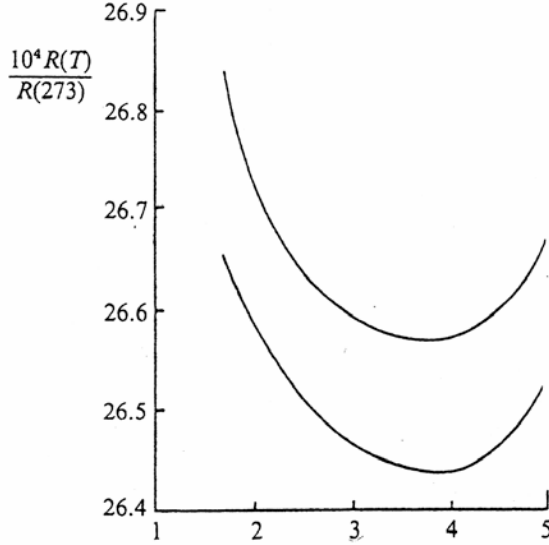


Fig. 1: Resistivity $R(T)$ versus temperature $T[K]$ of two samples of “pure” Au showing the first observation of the resistivity minimum [17]. The expected behavior of $R(T)$ for a pure metal with weak static disorder is a T^5 term due to phonons and a saturation to a constant value, ρ_0 , at $T = 0$ due to static disorder. The former is seen in the experiment, but at low temperature an additional logarithmically increasing contribution is also found.

where H_{imp} describes the impurity, a small quantum mechanical system with only a few degrees of freedom, H_{bath} represents the bath, and H_{int} is the interaction between the two.

We next consider explicit examples and introduce the linear-chain form of quantum impurity models, which is the starting point for an NRG treatment.

Anderson impurity model

The prototype model for strongly correlated systems is the single-band non-degenerate Anderson model [18, 19],

$$H_{AM} = \underbrace{\sum_{\sigma} \varepsilon_d n_{d\sigma} + U n_{d\uparrow} n_{d\downarrow}}_{H_{\text{imp}}} + \underbrace{\sum_{k\sigma} V_{kd} (c_{k\sigma}^{\dagger} d_{\sigma} + d_{\sigma}^{\dagger} c_{k\sigma})}_{H_{\text{int}}} + \underbrace{\sum_{k\sigma} \epsilon_k c_{k\sigma}^{\dagger} c_{k\sigma}}_{H_{\text{bath}}} . \quad (3)$$

The first two terms describe the impurity, represented here by a non-degenerate s -level of energy ε_d (see Sec. 6 for generalizations). Electrons in the local level are subject to a Coulomb repulsion U that acts between spin-up and spin-down electrons. The local level hybridizes with the Bloch states of a non-interacting s -wave conduction band, the last term in H_{AM} , with amplitude V_{kd} . The properties of the model are determined by the hybridization function

$$\Delta(\omega) = \pi \sum_k |V_{kd}|^2 \delta(\omega - \varepsilon_k), \quad (4)$$

which, like the conduction density of states $\rho(\omega) = \sum_k \delta(\omega - \varepsilon_k)$, will in general be a complicated function of energy. In cases where the interest is in the very low-energy physics, it is a good approximation to set $\Delta(\omega) \approx \Delta(\varepsilon_F) \equiv \Delta$. In applications to pseudogap systems [20, 21] or to effective quantum impurities in dynamical mean-field theory, the full frequency dependence has to be retained. In applications to quantum dots, the impurity is attached to two baths, the left and right leads, as shown in Fig. 2.¹

¹Although such dots are attached to two baths (the left and right leads), for a single level on the dot, only the even combination of left and right lead states couples to the dot. When several levels on the dot are active in transport, one will have a two-channel multi-orbital Anderson model with intra- and inter-orbital Coulomb interactions playing a role (e.g. Hund’s exchange).

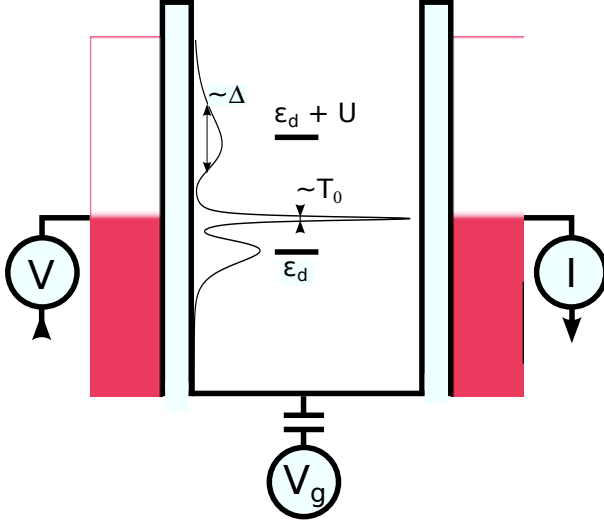


Fig. 2: A quantum dot with charging energy $U \gg \Delta$ and level energy ε_d connected to left/right leads $H_{\alpha=L/R} = \sum_k \epsilon_{k\alpha\sigma} c_{k\alpha\sigma}^\dagger c_{k\alpha\sigma}$ via tunnel barriers. The gate voltage $V_g \sim \varepsilon_d$ allows changing ε_d relative to ϵ_F and thereby the dot occupation n_d from $n_d = 1$ for $\varepsilon_d = -U/2$ (Kondo regime) to $n_d = 0$ through a mixed valence regime with $n_d \approx 0.5$ for $\varepsilon_d \approx 0$. [22, 23]

Kondo impurity model

Closely related to the Anderson model, is the Kondo model, which was briefly mentioned in the introduction. We write its Hamiltonian as

$$H_{KM} = \underbrace{-g\mu_B B S_z}_{H_{\text{imp}}} + \underbrace{J \vec{S} \cdot \vec{s}_0}_{H_{\text{int}}} + \underbrace{\sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma}}_{H_{\text{bath}}}, \quad (5)$$

where we included a magnetic field term $H_{\text{imp}} = -g\mu_B B S_z$ to indicate the impurity spin $\vec{S}$ (taken here to be $S = 1/2$ for simplicity), which interacts via an exchange interaction of strength J with the conduction electron spin-density $\vec{s}_0 = \sum_{\sigma\sigma'} f_{0\sigma}^\dagger \vec{\sigma}_{\sigma\sigma'} f_{0\sigma'}$ at the impurity site, where $f_{0\sigma} = \sum_k c_{k\sigma}$ is the local Wannier state of the conduction electrons at the impurity site. The connection to the Anderson model can be established formally via a Schrieffer-Wolff transformation. In essence, provided $\varepsilon_d < 0$ and $\varepsilon_d + U > 0$ so that a single electron occupies the local level in the Anderson model, the physics of both models will be the same at low temperatures.² In this case, one finds the correspondence $J = 2V^2(1/(U + \varepsilon_d) - 1/\varepsilon_d)$, which reduces to $8V^2/U$ for the symmetric case $\varepsilon_d = -U/2$ (see discussion of zero bandwidth limit below).

Linear chain representation

For a numerical treatment of the Anderson and Kondo models, it is useful to reformulate them in the form of linear chain models [2, 3]. This will allow them to be iteratively diagonalized by a procedure to be described in Sec. 3. We carry this out for the Anderson model: First notice that the impurity state in the Anderson model hybridizes with a local Wannier state $|0\sigma\rangle = f_{0\sigma}^\dagger |\text{vac}\rangle$,

²Strictly speaking, one should also include a potential scattering term in the Kondo model, of the form $\sum_{kk'\sigma} V_{kk'}^{\text{pot}} c_{k\sigma}^\dagger c_{k'\sigma}$ for this to be true.

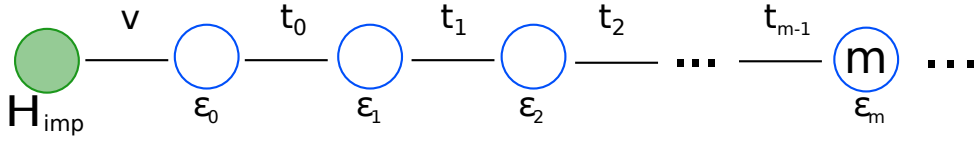


Fig. 3: The linear-chain form of the Anderson model (9). $H_{\text{imp}} = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow}$. The site energies ε_n and hoppings t_n follow from $\Delta(\omega)$.

with $|\text{vac}\rangle$ the vacuum state, and $f_{0\sigma}^\dagger$ given by

$$V f_{0\sigma}^\dagger = \sum_k V_{kd} c_{k\sigma}^\dagger. \quad (6)$$

The value of V follows from the normalization $\{f_{0\sigma}, f_{0\sigma}^\dagger\} = 1$

$$V = \left(\sum_k |V_{kd}|^2 \right)^{1/2}. \quad (7)$$

Using the above local state, one can apply the Lanczos procedure (Appendix B) for tridiagonalizing a Hermitian operator, such as H_{bath} , to obtain

$$H_{\text{bath}} = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} \rightarrow \sum_{\sigma, n=0}^{\infty} [\epsilon_n f_{n\sigma}^\dagger f_{n\sigma} + t_n (f_{n\sigma}^\dagger f_{n+1\sigma} + H.c.)] \quad (8)$$

with site energies ϵ_n and hoppings t_n depending only on the dispersion ε_k and hybridization V_{kd} through the hybridization function $\Delta(\omega)$, resulting in the linear chain form [2]

$$H_{AM} = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + V \sum_{\sigma} \left(f_{0\sigma}^\dagger d_{\sigma} + d_{\sigma}^\dagger f_{0\sigma} \right) + \sum_{\sigma, n=0}^{\infty} \left[\epsilon_n f_{n\sigma}^\dagger f_{n\sigma} + t_n (f_{n\sigma}^\dagger f_{n+1\sigma} + f_{n+1\sigma}^\dagger f_{n\sigma}) \right] \quad (9)$$

depicted in Fig. 3 (with $n_d \equiv \sum_{\sigma} n_{d\sigma}$). Although formally this model looks like the one-dimensional real-space models treated by the DMRG method [24, 25], the interpretation here is not in terms of electrons hopping on a one-dimensional lattice in real-space. Instead, as will become clearer in Sec. 3, each successive site added along the chain corresponds to adding lower-energy degrees of freedom, measured relative to the Fermi level. By considering longer chains one can then access lower energies.

The same procedure can be used to reformulate any quantum impurity model in terms of an impurity site with local interactions attached to a one-dimensional chain of non-interacting sites. For example, the Kondo model (5) can be rewritten as

$$H_{KM} = -g\mu_B S_z + J \vec{S} \cdot \vec{s}_0 + \sum_{\sigma, n=0}^{\infty} \left[\epsilon_n f_{n\sigma}^\dagger f_{n\sigma} + t_n (f_{n\sigma}^\dagger f_{n+1\sigma} + f_{n+1\sigma}^\dagger f_{n\sigma}) \right]. \quad (10)$$

Zeroth-order approximation for Anderson/Kondo models

A zeroth-order (high energy) approximation to the spectrum of the Anderson model can be obtained by considering just the coupling of the $n = 0$ Wannier state to the impurity and neglecting all others (the zero-bandwidth limit),

$$H_{AM} \approx H_0 \equiv \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + V \sum_{\sigma} \left(f_{0\sigma}^{\dagger} d_{\sigma} + d_{\sigma}^{\dagger} f_{0\sigma} \right). \quad (11)$$

There are 16 many-electron states $|n_d, n_0\rangle$, which can be classified by the conserved quantum numbers of total electron number N_{el} , total z -component of spin S_z^{tot} and total spin $\vec{S}$. Using these symmetries we can diagonalize the block matrices H_{N_{el}, S, S_z}^0 to obtain the many-body eigenstates $|N_{el}, S, S_z, r\rangle$ and the corresponding eigenvalues. For example, in the product basis $|n_d\rangle|n_0\rangle$, the Hamiltonian for $N_e = 1, S = 1/2, S_z = \pm 1/2$ is given by

$$H_{N_e=1, S=1/2, S_z=\pm 1/2} = \begin{pmatrix} \varepsilon_d & V \\ V & 0 \end{pmatrix}$$

with eigenvalues

$$E_{\pm} = \left(\varepsilon_d \pm \sqrt{\varepsilon_d^2 + 4V^2} \right) / 2.$$

Proceeding similarly for the other Hilbert spaces (exercise), we find that for the particle-hole symmetric case $\varepsilon_d = -U/2$ in the strong correlation limit $U \gg V^2$, the spectrum separates into two groups of states, one group of low-energy states lying close to the (singlet) ground state with spacings $\mathcal{O}(V^2/U)$ and one group of high-energy states lying at energies $\mathcal{O}(U/2)$ higher and also split by $\mathcal{O}(V^2/U)$. This limit corresponds to a singly occupied impurity level effectively behaving as $S = 1/2$. In fact, the 8 lowest states correspond to those obtained from a zeroth-order approximation to the spectrum of the Kondo model via

$$H_{KM} \approx H_0 \equiv J \vec{S} \cdot \vec{s}_0 = \frac{J}{2} \left[(\vec{S} + \vec{s}_0)^2 - \vec{S}^2 - \vec{s}_0^2 \right]. \quad (12)$$

The Kondo model is therefore the low-energy effective model of the Anderson model in the limit of strong correlations and single occupancy. By comparing the splitting of the two lowest levels in the Kondo model, the singlet and triplet states, with the corresponding splitting of the same levels in the Anderson model one finds the relation between the bare parameters of the models to be $J = 8V^2/U$, in agreement with the value obtained from the Schrieffer-Wolff transformation $J = 2V^2(1/(U + \varepsilon_d) - 1/\varepsilon_d)$ upon setting $\varepsilon_d = -U/2$ [26].

Within the above zeroth-order approximation, $H \approx H_0$, of the Kondo (and Anderson) model, excitations are unrenormalized. The singlet-triplet excitation (splitting) takes the bare value J . The key ingredient of Wilson's NRG, to be discussed in the next section, is a *controlled procedure* for adding the remaining states $n = 1, 2, \dots$ neglected in the above zero bandwidth approximation. As we shall see in the calculation of dynamical quantities below, this leads to a drastic renormalization of the spin and single-particle excitations, such that the relevant excitations of the Kondo model are not on the bare scale J but on the Kondo scale

$T_K = D(\rho J)^{1/2} \exp(-1/\rho J)$, where $\rho = 1/2D$ is the density of conduction states. One can interpret this large renormalization $J \rightarrow T_K$ as a renormalization of a bare tunneling amplitude ($J_\perp = J$) due to the dissipative effects of the bath of conduction electrons by a mapping of the (anisotropic) Kondo model onto the dissipative two-state system (also called the spin-boson model). We introduce this and its fermionic equivalents in the next subsection, partly to make the above connection and partly to show that the linear chain representation, which is the starting point for NRG calculations, applies also to bosonic quantum impurity models. For a detailed discussion of the bosonic models within NRG, we refer the reader to the lecture by K. Ingersent.

Spin-boson model and fermionic equivalents

The Hamiltonian of the spin-boson model (SBM) is given by,

$$H_{\text{SB}} = \underbrace{-\frac{1}{2}\Delta_0\sigma_x + \frac{1}{2}\epsilon\sigma_z}_{H_{\text{imp}}} + \underbrace{\frac{1}{2}\sigma_z \sum_i \lambda_i (a_i + a_i^\dagger)}_{H_{\text{int}}} + \underbrace{\sum_i \omega_i (a_i^\dagger a_i + 1/2)}_{H_{\text{bath}}}. \quad (13)$$

The first term H_{imp} describes a two-level system with bias splitting ϵ and bare tunneling amplitude Δ_0 . The $\sigma_{i=x,y,z}$ are the Pauli spin matrices. The third term, H_{bath} , is the environment and consists of an infinite set of harmonic oscillators ($i = 1, 2, \dots, \infty$) with $a_i(a_i^\dagger)$ the annihilation (creation) operators for a harmonic oscillator of frequency ω_i and $0 \leq \omega_i \leq \omega_c$, where ω_c is an upper cut-off frequency. The non-interacting density of states of the environment is denoted by $g(\omega) = \sum_i \delta(\omega - \omega_i)$ and is finite in the interval $[0, \omega_c]$ and zero otherwise. Finally, $H_{\text{int}} = \frac{1}{2}\sigma_z \sum_i \lambda_i (a_i + a_i^\dagger)$ describes the coupling of the two-state system coordinate σ_z to the oscillators, with λ_i denoting the coupling strength to oscillator i . The function $\Gamma(\omega + i\delta) = \sum_i (\lambda_i/2)^2 / (\omega - \omega_i + i\delta) = \int d\omega' (\lambda(\omega')/2)^2 g(\omega') / (\omega - \omega' + i\delta)$ characterizes the system-environment interaction. For a numerical treatment using the NRG, one proceeds to re-formulate the model (13) in a linear chain form as in (9) and (10) for the Anderson and Kondo models. Thus, one uses the Lanczos procedure and applies H_{bath} repeatedly on the local bosonic orbital $\lambda b_0 = \sum_i \lambda_i a_i$ to tridiagonalize H_{bath} . The resulting linear chain model

$$H_{\text{SB}} = \underbrace{-\frac{1}{2}\Delta_0\sigma_x + \frac{1}{2}\epsilon\sigma_z}_{H_{\text{imp}}} + \underbrace{\frac{1}{2}\sigma_z \lambda (b_0 + b_0^\dagger)}_{H_{\text{int}}} + \underbrace{\sum_{m=0}^{\infty} \epsilon_m b_m^\dagger b_m + t_m (b_m^\dagger b_{m+1} + b_{m+1}^\dagger b_m)}_{H_{\text{bath}}} \quad (14)$$

may then be treated within NRG in a similar way to the treatment of the Anderson and Kondo models [27], see the lecture by K. Ingersent for details. One difference is that the number of bosons in the eigenstates of H_{SB} is arbitrary, requiring an additional approximation even at the first iteration for $H_0 = H_{\text{imp}} + \frac{1}{2}\sigma_z \lambda (b_0 + b_0^\dagger)$ to restrict the maximum number of bosons to a finite number n_b (typically 8 – 10).

Anisotropic Kondo model

It may be shown via bosonization [28] that the ohmic two-state system, specified by Eq. (13) with a spectral function $J(\omega) = -\frac{1}{\pi} \text{Im} \Gamma(\omega + i\delta) \sim \alpha \omega$ for $\omega \rightarrow 0$, where α is the dimensionless dissipation strength, is equivalent to the anisotropic Kondo model

$$H_{\text{AKM}} = \underbrace{\frac{J_{\perp}}{2}(S^+ s_0^- + S^- s_0^+)}_{H_{\text{imp}}} - \underbrace{g\mu_B B S_z}_{H_{\text{int}}} + \underbrace{J_{\parallel} S_z s_0^z + \sum_{k\sigma} \epsilon_k c_{k\sigma}^{\dagger} c_{k\sigma}}_{H_{\text{bath}}}, \quad (15)$$

where $J_{\perp}(J_{\parallel})$ is the transverse (longitudinal) part of the Kondo exchange interaction and B is a local magnetic field. The correspondence is given by $\rho J_{\perp} = -\Delta_0/\omega_c$, $-g\mu_B B = \epsilon$, and $\alpha = (1 + 2\delta/\pi)^2$ where $\delta = \arctan(-\pi\rho J_{\parallel}/4)$ and $\rho = 1/\omega_c$ is the density of states of the conduction electrons in the anisotropic Kondo model [29–31, 14, 15]. The natural low-energy scale of the ohmic two-state system is the renormalized tunneling amplitude $\Delta_r/\omega_c \approx (\Delta_0/\omega_c)^{1/(1-\alpha)}$. A more precise estimate is $\Delta_r/\omega_c = [\sqrt{\Gamma(1-2\alpha)\cos(\pi\alpha)} \Delta_0/\omega_c]^{1/(1-\alpha)}$, yielding the known limits $\Delta_r(\alpha \rightarrow 0) = \Delta_0$ and $\Delta_r(\alpha \rightarrow 1/2) = \frac{\pi}{2}(\Delta_0/\omega_c)^2 \omega_c$ [15]. For $\alpha > 1/2$, further corrections are needed [15, 32]. It is related to the low-energy Kondo scale T_K of the Anisotropic Kondo model. The connection between the (anisotropic) Kondo and ohmic two-state system provides another viewpoint on the local dynamics of a Kondo spin in terms of tunneling and dissipation: The two levels concerned are the two lowest $S_z = 0$ states of H_{imp} , i.e., $|\pm\rangle = 1/\sqrt{2}(|\uparrow\rangle|\downarrow\rangle \pm |\downarrow\rangle|\uparrow\rangle)$ which are connected by $J_{\perp}$ and tunnel-split by $\Delta_0 = J_{\perp}$ when $J_{\parallel} = +\infty$, corresponding to $\alpha = 0$ (decoupled two-level system). A finite $J_{\parallel} < +\infty$, resulting in a finite $\alpha > 0$, couples these states to the environment and leads to a renormalization of the bare tunneling amplitude $\Delta_0 = J_{\perp} \rightarrow \Delta_r = T_K$, which is particularly drastic in the limit of strong dissipation $\alpha \rightarrow 1^-$. Indeed, for $\alpha > 1$, the above correspondence states that $J_{\parallel} < 0$, which corresponds to the ferromagnetic sector of the Kondo model (see the lecture by A. Nevidomskyy). Since in this limit, $J_{\perp}$ is irrelevant [19, 26], it follows that T_K and hence Δ_r vanish for $\alpha > 1$, i.e., the frictional effects of the environment are so large for $\alpha > 1$ that quantum mechanical tunneling is destroyed at $T = 0$ and tunneling between the two states is possible only via thermal activation. Of interest also is the transition from quantum coherent dynamics at weak dissipation $\alpha \ll 1$ to incoherent dynamics at strong dissipation $\alpha \rightarrow 1^-$, see the time-dependent NRG section.

Interacting resonant level model

Finally, the ohmic spin-boson model (SBM) is equivalent to an even simpler fermionic model, the so-called interacting resonant level model (IRLM),

$$H_{\text{IRLM}} = \underbrace{\varepsilon_d n_d + V(f_0^{\dagger} d + d^{\dagger} f_0)}_{H_{\text{imp}}} + \underbrace{U_{dc}(n_d - 1/2)(n_0 - 1/2)}_{H_{\text{int}}} + \underbrace{\sum_k \epsilon_k c_k^{\dagger} c_k}_{H_{\text{bath}}}. \quad (16)$$

This model describes a spinless resonant level with energy ε_d hybridizing with a spinless bath of electrons (where we wrote $f_0 = \sum_k c_k$ and $n_0 = f_0^{\dagger} f_0$) and interacting with the latter via a

non-local Coulomb interaction U_{dc} . Denoting by $\Gamma = \pi\rho V^2$ the width of the resonant level at $U_{dc} = 0$, the correspondence of the SBM to this model is given by $\Delta_0 = 2V \sim \Gamma^{1/2}$, $\varepsilon_d = \epsilon/2$, and $\alpha = (1 + 2\delta/\pi)^2/2$ where $\delta = \arctan(-\pi\rho U_{dc}/2)$. The equivalence between the models can be shown via bosonization and is valid for all $-\infty \leq U_{dc} \leq +\infty$ (describing the sector $2 \geq \alpha \geq 0$) with $U_{dc} = 0$ corresponding to $\alpha = 1/2$ in the SBM. This point marks the crossover from quantum coherent oscillations of the two-level system at $\alpha < 1/2$ to incoherent dynamics at $\alpha > 1/2$, e.g., in the quantity $P(t > 0) \equiv \langle \sigma_z(t) \rangle$ subject to an initial state preparation $\sigma_z = +1$. The physics here is again that of a two-level system (consisting of the single-electron states $|\pm\rangle = (|1\rangle_d|0\rangle_0 \pm |0\rangle_d|1\rangle_0)/\sqrt{2}$ of H_{imp}) tunnel-coupled by $\Delta_0 = \sqrt{\varepsilon_d^2 + 4V^2}$. Note that the non-interacting limit $\alpha = 0$ describing the isolated two level system H_{imp} corresponds to $U_{dc} = +\infty$. Strong dissipation ($\alpha > 1/2$) in this model, corresponds to negative U_{dc} and the quantum critical point $\alpha = 1$ occurs at $U_{dc}^* = -(2/\pi\rho) \tan[\pi(\sqrt{2} - 1)/2] \approx -0.969$. For $U_{dc} < U_{dc}^*$ and $\alpha > 1$, quantum mechanical tunneling is absent ($\Delta_r = 0$), and only tunneling via thermal activation is possible (i.e. at $T > 0$). The IRLM is interesting since it is the simplest model that can capture a part of the Kondo physics contained in the Anderson and Kondo models, in particular, the thermodynamic properties and the spin dynamics of the latter: To capture the spin dynamics of the latter within the IRLM, one notes that under the equivalence $S_z \rightarrow n_d - 1/2$ the dynamic spin susceptibility of the AKM $\chi_{zz}(\omega) = \langle\langle S_z; S_z \rangle\rangle$ corresponds to the dynamic charge susceptibility $\chi_{dd}(\omega) = \langle\langle n_d; n_d \rangle\rangle$ of the IRLM. Since the IRLM is a spinless model, it is simpler to deal with than the Anderson and Kondo models and we shall use it to illustrate much of the NRG in the next section.

3 Wilson's Numerical Renormalization Group method

Wilson's formulation of the RG for the Kondo model is similar in spirit to Anderson's scaling approach (see Hewson's book [26], Ref. [19] or Nevidomskyy's lecture). The main difference lies in the non-perturbative construction of the RG transformation using a numerical representation of the effective Hamiltonians. The scaling approach uses perturbation theory in the initially small dimensionless coupling (J/D in the Kondo model, or V in the Anderson and IRLM models) to construct such a transformation, but since J/D increases with decreasing energy scale, this approach eventually becomes inaccurate. In the Wilson approach, the RG transformation is perturbative only via a small parameter $\Lambda^{-1/2} < 1$, which is related to the momentum rescaling factor $\Lambda > 1$. The accuracy of the transformation is the same at each step and is independent of the size of the running couplings. For this reason, it gave the first correct description of the crossover from the weak-coupling to the strong-coupling regime of the Kondo model. The NRG procedure involves three steps, illustrated schematically in Fig. 4a-c.

Separation of scales and logarithmic discretization approximation

In quantum impurity problems, the behavior of the system typically changes qualitatively over many energy scales as it passes through a crossover between fixed points. In order to describe

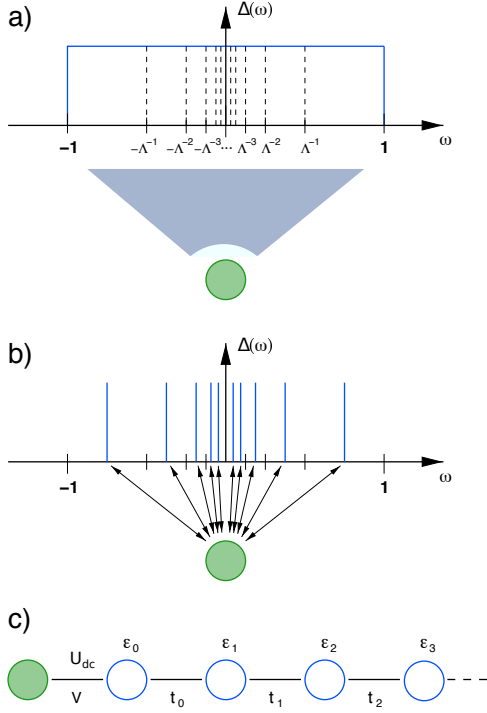


Fig. 4: Steps in the NRG procedure for a quantum impurity model, here shown for the IRLM. The level (filled circle) hybridizes with strength V to a continuum of conduction states and interacts with these via the Coulomb term U_{dc} ($\Delta(\omega) = \pi\rho V^2$ is here constant). (a) Logarithmic discretization of the continuum conduction band about the Fermi level $\epsilon_F = 0$ into discrete intervals $D_n^+ = [\Lambda^{-(n+1)}, \Lambda^{-n}]$ and $D_n^- = [-\Lambda^{-n}, -\Lambda^{-(n+1)}]$, $n = 0, 1, \dots$. (b) Within each discrete interval, choose the conduction state most localized on the impurity. (c) Transform this discretized model to linear-chain form, with hoppings now decreasing along the chain and iteratively diagonalize.

this crossover quantitatively the idea is to separate out the many energy scales in the problem, which arise from the conduction band $[-D, +D]$, and to set up a procedure for treating each scale in turn. We saw in the previous section that it is always possible to rewrite a quantum impurity model in the form of a (semi-infinite) linear chain, see Fig. 3. Truncating this chain to include orbitals $n = 0, \dots, m$, we have for the IRLM

$$H_{\text{IRLM}} \approx H_m = \epsilon_d n_d + V \left(f_0^\dagger d + d^\dagger f_0 \right) + U_{dc} (n_d - 1/2)(n_0 - 1/2) + \sum_{n=0}^m \epsilon_n f_n^\dagger f_n + \sum_{n=0}^{m-1} t_n \left(f_n^\dagger f_{n+1} + f_{n+1}^\dagger f_n \right) \quad (17)$$

with the truncated Hamiltonians H_m satisfying the recursion relation

$$H_{m+1} = H_m + \epsilon_m f_m^\dagger f_m + t_m \left(f_m^\dagger f_{m+1} + f_{m+1}^\dagger f_m \right). \quad (18)$$

Hence, it appears that with this recursion relation, one can iteratively diagonalize the IRLM (and indeed any other QIM) starting from H_0 . For the IRLM, H_0 has four eigenstates classified into three Hilbert spaces by the total electron number $N_e = 0, 1, 2$, a conserved quantity (work them out in the product basis $|n_d\rangle|n_0\rangle$). The two one-electron states form the two-level system and are split by the tunnel splitting $\Delta_0 = \sqrt{\epsilon_d^2 + 4V^2}$. At some point, for sufficiently large m , we will only be able to retain the lowest many-body states of H_m since the Hilbert space grows as $2 \times 2^{m+1}$. The validity of this procedure then depends on whether the perturbation in Eq. (18), the last term involving t_m , is small, once we start neglecting some higher-energy states.³ In practice, for a quasi-continuous band $H_{\text{bath}} = \sum_k \epsilon_k c_k^\dagger c_k = \int_{-D}^{+D} d\epsilon \epsilon c_\epsilon^\dagger c_\epsilon$ the hoppings t_m do

³If we keep extending the system by one orbital at a time without neglecting any states, no error is made. The onsite term in ϵ_m is diagonal and shifts the low-energy levels of H_m .

not decay with increasing m , and the above procedure breaks down after some iterations. For example, it can be easily shown for a semi-elliptic density of states $\rho(\epsilon) = \frac{2}{\pi D^2} \sqrt{D^2 - \epsilon^2}$, that $t_m = D/2$ for all m (see Hewson's book [26]).

In order to have a working procedure involving decreasing hoppings t_m along the chain, and at the same time achieve the energy scale separation described above, Wilson discretized the conduction band into positive and negative energy intervals, $D_n^+ = [\Lambda^{-(n+1)}, \Lambda^{-n}]$ and $D_n^- = [-\Lambda^{-n}, -\Lambda^{-(n+1)}]$, $n = 0, 1, \dots$, about the Fermi level $\epsilon_F = 0$ as shown in Fig. 4a.

The quasi-continuous band is then approximated by a discrete one by keeping only a single conduction state from each interval $D_n^\pm$,

$$H_{\text{bath}} = \int_{-D}^{+D} d\epsilon \epsilon c_\epsilon^\dagger c_\epsilon \approx \sum_{n=0}^{\infty} \left(\epsilon_{-n} c_{-n}^\dagger c_{-n} + \epsilon_{+n} c_{+n}^\dagger c_{+n} \right) \quad (19)$$

with

$$\epsilon_{\pm n} = \pm \frac{1}{2} D (\Lambda^{-n} + \Lambda^{-(n+1)}) = \pm \frac{1}{2} D \Lambda^{-n} (1 + \Lambda^{-1}). \quad (20)$$

The states $c_{\pm n}^\dagger |vac\rangle$ appearing above are the states in each interval $D_n^\pm$ which are most localized near the impurity [2], while the neglected states being orthogonal to these have their wavefunctions localized away from the impurity and are consequently less important (for a more detailed derivation and justification of the logarithmic discretization approximation see Appendix A). By formulating the IRLM as a linear chain using the above logarithmically discretized conduction band, we obtain the same equations (17-18) as above, but, crucially, with hopping parameters t_m (and onsite energies ϵ_m) that now decay exponentially along the chain. For example, for a constant density of states $\rho(\omega) = 1/2D$ and constant hybridization function $\Delta(\omega) = \pi\rho(\omega)V^2 = \Delta_0$ and $\epsilon_F = 0$ one finds for $m = 0, 1, \dots$ [1]

$$\epsilon_m = 0, \quad (21)$$

$$t_m = \frac{1}{2} D (1 + \Lambda^{-1}) \Lambda^{-m/2} \xi_m, \quad (22)$$

$$\xi_m = \frac{1 - \Lambda^{-m-1}}{\sqrt{(1 - \Lambda^{-2m-1})(1 - \Lambda^{-2m-3})}}. \quad (23)$$

The ξ_m converge rapidly to 1 with increasing m and we may write $t_m \approx \frac{1}{2} D (1 + \Lambda^{-1}) \Lambda^{-m/2}$, so that the IRLM becomes

$$\begin{aligned} H_{\text{IRLM}} = & \epsilon_d n_d + V \left(f_0^\dagger d + d^\dagger f_0 \right) + U_{dc} (n_d - 1/2)(n_0 - 1/2) \\ & + \frac{1}{2} D (1 + \Lambda^{-1}) \sum_{n=0}^{\infty} \Lambda^{-n/2} \left(f_n^\dagger f_{n+1} + f_{n+1}^\dagger f_n \right). \end{aligned} \quad (24)$$

This Hamiltonian provides a clear separation of the energy scales $\frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-n/2}$, $n = 1, 2, \dots$ in H and allows the diagonalization of the Hamiltonian in a sequence of controlled steps, each step corresponding to adding an orbital f_n , which is a relative perturbation of strength $\Lambda^{-1/2} < 1$, thereby ensuring convergence of the method. This procedure is described in the following two subsections, where we henceforth restrict ourselves to a constant hybridization with hoppings $t_m \approx \frac{1}{2} D (1 + \Lambda^{-1}) \Lambda^{-m/2}$. The procedure is easily generalized to any hybridization function $\Delta(\omega)$ with hoppings t_m decaying sufficiently fast along the chain.

Renormalization group transformation

A renormalization group transformation relating effective Hamiltonians on successive energy scales $\Lambda^{-n/2}$ and $\Lambda^{-(n+1)/2}$ can be set up as follows. First, H_{IRLM} in (24) is truncated to H_m , whose lowest scale is $D_m = \frac{1}{2}D(1 + \Lambda^{-1})\Lambda^{-(m-1)/2}$. In order to look for fixed points, we define rescaled Hamiltonians $\bar{H}_m \equiv H_m/D_m$ such that the lowest energy scale of $\bar{H}_m$ is always of order $\mathcal{O}(1)$:

$$\bar{H}_m = \Lambda^{(m-1)/2} \left[\sum_{n=0}^{m-1} \Lambda^{-n/2} (f_n^\dagger f_{n+1} + f_{n+1}^\dagger f_n) + \tilde{\varepsilon}_d n_d + \tilde{V} (f_0^\dagger d + d^\dagger f_0) + \tilde{U}_{dc} \left(n_d - \frac{1}{2} \right) \left(n_0 - \frac{1}{2} \right) \right] \quad (25)$$

$$\tilde{\varepsilon}_d = \frac{\varepsilon_d}{\frac{1}{2}D(1 + \Lambda^{-1})}, \quad \tilde{V} = \frac{V}{\frac{1}{2}D(1 + \Lambda^{-1})}, \quad \tilde{U}_{dc} = \frac{U_{dc}}{\frac{1}{2}D(1 + \Lambda^{-1})}, \quad (26)$$

from which we can recover H as

$$H = \lim_{m \rightarrow \infty} \frac{1}{2}D(1 + \Lambda^{-1}) \Lambda^{-(m-1)/2} \bar{H}_m. \quad (27)$$

The sequence of rescaled Hamiltonians $\bar{H}_m$ satisfies the recursion relation

$$\bar{H}_{m+1} = \Lambda^{1/2} \bar{H}_m + \left(f_m^\dagger f_{m+1} + f_{m+1}^\dagger f_m \right) \quad (28)$$

and allows a RG transformation $\mathcal{T}$ to be defined:

$$\bar{H}_{m+1} = \mathcal{T}[\bar{H}_m] \equiv \Lambda^{1/2} \bar{H}_m + \left(f_m^\dagger f_{m+1} + f_{m+1}^\dagger f_m \right) - \bar{E}_{G,m+1} \quad (29)$$

with $\bar{E}_{G,m+1}$ the ground-state energy of $\bar{H}_{m+1}$. In fact $\mathcal{T}$ defined in (29) may not have fixed points since it relates a Hamiltonian with an even number of orbitals to a Hamiltonian with an odd number of orbitals. If this happens, then $\mathcal{R} = \mathcal{T}^2$ can be defined as the RG transformation, and this will have fixed points, a set of even m fixed points and a set of odd m fixed points:

$$\bar{H}_{m+2} = \mathcal{R}[\bar{H}_m] \equiv \mathcal{T}^2[\bar{H}_m]. \quad (30)$$

Iterative diagonalization scheme

The transformation R relates effective Hamiltonians $H_m = D_m \bar{H}_m$ and $H_{m+1} = D_{m+1} \bar{H}_{m+1}$ on decreasing scales $D_m > D_{m+1}$. It can be used to iteratively diagonalize the Anderson Hamiltonian by the following sequence of steps:

1. the local part

$$\bar{H}_0 = \Lambda^{-1/2} \left[\tilde{\varepsilon}_d n_d + \tilde{V} (f_0^\dagger d + d^\dagger f_0) + \tilde{U}_{dc} (n_d - 1/2)(n_0 - 1/2) \right], \quad (31)$$

which contains the many-body interactions, is diagonalized to yield the eigenstates $|q\rangle$ (the “zeroth” order step described above and in Sec. 2),

2. assuming that $\bar{H}_m$ has been diagonalized for some $m \geq 0$,

$$\bar{H}_m = \sum_q \bar{E}_q^m |q\rangle\langle q|$$

we add a “site” and use Eq. (29) to set up the matrix for $\bar{H}_{m+1}$ within a product basis $|q, \alpha_{m+1}\rangle = |q\rangle_m |\alpha_{m+1}\rangle$ consisting of the eigenstates $|q\rangle_m$ of $\bar{H}_m$ and the states of the next orbital along the chain $|\alpha_{m+1}\rangle$ ($|0\rangle$ for $\alpha_{m+1} = 1$ and $|1\rangle$ for $\alpha_{m+1} = 2$),

$$\begin{aligned} \langle q, \alpha_{m+1} | \bar{H}_{m+1} | q', \alpha'_{m+1} \rangle &= A^{1/2} \delta_{\alpha_{m+1}, \alpha'_{m+1}} \delta_{q, q'} \bar{E}_q^m \\ &+ (-1)^{N_{e, q'}} {}_m \langle q | f_m^\dagger | q' \rangle_m \langle \alpha_{m+1} | f_{m+1} | \alpha'_{m+1} \rangle \\ &+ (-1)^{N_{e, q}} \langle \alpha_{m+1} | f_{m+1}^\dagger | \alpha'_{m+1} \rangle_m \langle q | f_m | q' \rangle_m, \end{aligned} \quad (32)$$

with $N_{e, q}, N_{e, q'}$ the number of electrons in $|q\rangle$ and $|q'\rangle$, respectively. This is diagonalized and the procedure is repeated for the next energy shell as depicted in Fig. 4c. Since $\bar{H}_m$ is already diagonalized, the off-diagonal matrix elements, involving ${}_m \langle q | f_{m\sigma}^\dagger | q' \rangle_m = {}_m \langle q' | f_{m\sigma} | q \rangle_m^\dagger$, can be expressed in terms of the known eigenstates of $\bar{H}_m$ by using the unitary transformation relating product states $|q\rangle_{m-1} |\alpha_m\rangle$ to eigenstates $|q\rangle_m$ of $\bar{H}_m$

$$|q\rangle_m = \sum_{r, \alpha_m} U_m(r, \alpha_m, q) |q\rangle_{m-1} |\alpha_m\rangle, \quad (33)$$

where U_m is the matrix of eigenvectors of $\bar{H}_m$.

Equation (33) also shows that the NRG eigenstates have the form of so called matrix product states (MPS) [33], a feature of NRG shared also by the density matrix renormalization group method (DMRG) for one-dimensional quantum systems [24, 25]. In order to see this, we introduce the notation $A_{q_m q_{m-1}}^{\alpha_m} \equiv U_m(q_{m-1}, \alpha_m, q_m)$ with $|q_m\rangle \equiv |q\rangle_m$ and repeatedly apply Eq. (33) to obtain

$$\begin{aligned} |q_m\rangle &= \sum_{q_{m-1}, \alpha_m} A_{q_{m-1} q_m}^{\alpha_m} |q_{m-1}\rangle |\alpha_m\rangle \\ &= \sum_{q_{m-1}, \alpha_m} A_{q_{m-1} q_m}^{\alpha_m} \left[\sum_{q_{m-2}, \alpha_{m-1}} A_{q_{m-2} q_{m-1}}^{\alpha_{m-1}} |\alpha_{m-1}\rangle \right] |\alpha_m\rangle \\ &= \sum_{q_{m_0}, \alpha_{m_0+1}, \dots, \alpha_m} (A^{\alpha_{m_0+1}} \dots A^{\alpha_m})_{q_{m_0} q_m} |\alpha_{m_0+1}\rangle \dots |\alpha_m\rangle, \end{aligned} \quad (34)$$

where $m_0 \geq 0$.

3. In order to reduce the size of the matrices that need to be diagonalized, one uses available symmetries, such as conservation of total electron number, or in models with spin degrees of freedom with rotational symmetry, conservation of total spin. For multi-channel models, such as the three-channel Kondo model, additional symmetries, such as $SU(3)$, may be used to significantly reduce the numerical effort [34, 35]. The use of symmetries, beyond the advantage of reducing the computational cost, also improves the accuracy of the calculations once one starts to neglect high-energy states (see next subsection), since it avoids the possibility of splitting up degenerate states within a multiplet carrying the same conserved quantum numbers.

Truncation

In practice, since the number of many-body states in $\bar{H}_m$ grows as $2 \times 2^{m+1}$ for the IRLM and as $4 \times 4^{m+1}$ for the Anderson or Kondo models, it is not possible to retain all states after a given iteration $m = m_0$. Keeping 1024 states results in $m_0 = 8$ for the IRLM and $m_0 = 3$ for the Anderson model. We denote the retained states of $\bar{H}_m$ by $|k\rangle_m$, while the higher-energy states neglected are denoted by $|l\rangle_m$, see Fig. 12 in Sec. 5. While only kept states are used to set up and diagonalize the sequence of Hamiltonians $\bar{H}_m, m = m_0, m_0 + 1, \dots$ up to a maximum chain size of length $m = N$, we shall see later that the discarded states $|l\rangle_m$ from each iteration $m \geq m_0$ prove to be very useful for calculating physical properties.

The truncation of the spectrum of $\bar{H}_m$ restricts the range of eigenvalues in $H_m = D_m \bar{H}_m$ to be such that $0 \leq E_q^m \leq K D_m$ where $K = K(\Lambda)$ depends on Λ and the number of states retained. For 1000 states and $\Lambda = 3$, $K(\Lambda) \approx 10$. However, eigenvalues below D_m are only approximate eigenvalues of the infinite system H , since states with energies below D_m are calculated more accurately in subsequent iterations $m+1, m+2, \dots$. Therefore the part of the spectrum of H_m that is close to the spectrum of H is restricted to $D_m \leq E_q^m \leq K(\Lambda) D_m$. This allows the whole spectrum of H to be recovered by considering the spectra of the sequence of Hamiltonians $H_m, m = 0, 1, \dots$. In this way, the many-body eigenvalues and eigenstates are obtained on all energy scales. Due to the smallness of the perturbation (of $\mathcal{O}(\Lambda^{-1/2}) < 1$) in adding an energy shell to go from H_m to H_{m+1} , the truncation of the high-energy states turns out, in practice, to be a very good approximation.

Fixed points

The analysis of fixed points is important to gain a conceptual understanding of the model and for accurate analytic calculations in the vicinity of a fixed point [2].

From (30), a fixed point H^* of $\mathcal{R} = \mathcal{T}^2$ is defined by

$$H^* = \mathcal{R}[H^*]. \quad (35)$$

Proximity to a fixed point is identified by ranges of $m, m_1 \leq m \leq m_2$, where the energy levels $\bar{E}_p^m$ of $\bar{H}_m$ are approximately independent of m : $\bar{E}_p^m \approx \bar{E}_p$ for $m_1 \leq m \leq m_2$. A typical energy level flow diagram showing regions of m where the energy levels are approximately constant is shown in Fig. 5a for the anisotropic Kondo model (AKM) [30]

$$H_{AKM} = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \frac{J_\perp}{2} \left(S^+ f_{0\downarrow}^\dagger f_{0\uparrow} + S^- f_{0\uparrow}^\dagger f_{0\downarrow} \right) + \frac{J_\parallel}{2} S_z \left(f_{0\uparrow}^\dagger f_{0\uparrow} - f_{0\downarrow}^\dagger f_{0\downarrow} \right). \quad (36)$$

There is an unstable high-energy fixed point (small m) and a stable low-energy fixed point (large m). The low-energy spectrum is identical to that of the isotropic Kondo model at the strong-coupling fixed point $J = \infty$ in [1] (e.g. the lowest single-particle excitations in Fig. 5a, $\eta_1 = 0.6555$, $\eta_2 = 1.976$ agree with the $\Lambda = 2$ results of the isotropic model in [1]). The crossover from the high-energy to the low-energy fixed point is associated with the Kondo scale

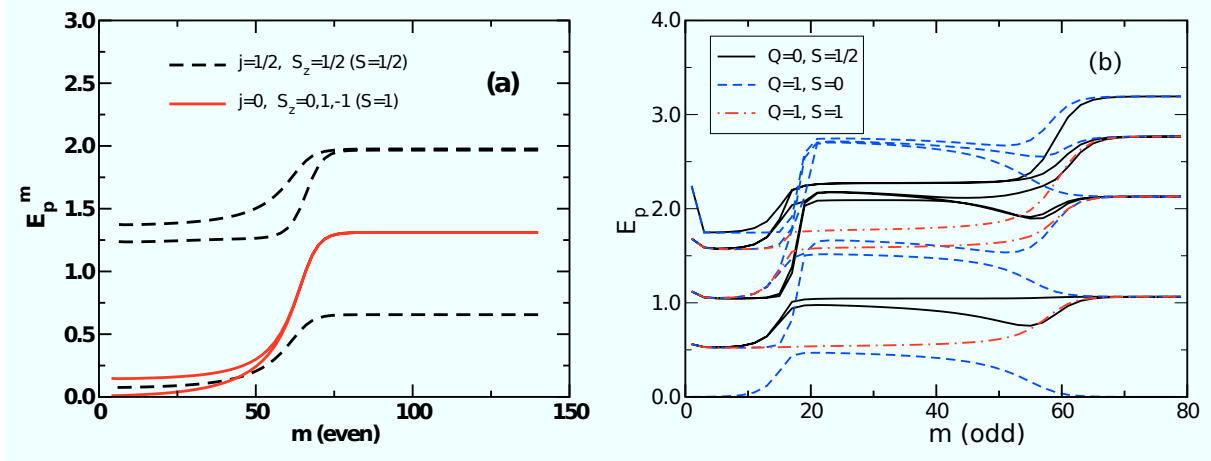


Fig. 5: (a) The lowest rescaled energy levels of the AKM for $J_{\parallel} = 0.443$ and $J_{\perp} = 0.01$. The states are labeled by conserved pseudospin j and total S_z (adapted from [30]). (b) The flow of the lowest many-body energy levels of the Anderson model for $\varepsilon_d = -U/2$, $U/\pi\Delta_0 = 12.66$, and $\Lambda = 2.5$. States are labeled by quantum numbers for total charge Q and total spin S (adapted from [4, 2]).

T_K . Spin-rotational invariance, broken at high energies, is restored below this scale (e.g., the $j = 0$ states with $S_z = 0$ and $S_z = \pm 1$ become degenerate below T_K and can be classified by the same total spin S as indicated in Fig. 5a). In Fig. 5b typical energy level flows for the symmetric Anderson impurity model $\varepsilon_d = -U/2$ in the strongly correlated Kondo regime are shown. Here, one sees three fixed points: an unstable free orbital fixed point for $m < 10$, a marginal fixed point for $10 < m < 50$ corresponding to formation of a local moment interacting weakly via the antiferromagnetic Kondo exchange with the conduction electrons. In this region, the effective Hamiltonian is essentially the Kondo model. Finally, for $m > 50$ there is a flow to the stable strong-coupling fixed point, characterized by a fixed-point spectrum obtained by setting $J = \infty$, i.e., the local spin and local conduction orbital are frozen out. The fixed-point spectrum is then that of a free electron chain with one site removed, i.e., there is a crossover to an even m fixed-point spectrum. The freezing out of the local spin implies that inelastic scattering processes are blocked as $m \rightarrow \infty$ ($T \rightarrow 0$), and one obtains the picture of a renormalized Fermi liquid at low temperatures.

Analytic calculations can be carried out in the vicinity of these various fixed points by setting up effective Hamiltonians $H_{\text{eff}} = H^* + \sum_{\lambda} \omega_{\lambda} O_{\lambda}$, where the leading deviations O_{λ} about H^* can be obtained from general symmetry arguments. This allows, for example, thermodynamic properties to be calculated in a restricted range of temperatures, corresponding to the restricted range of m where $\bar{H}_m$ can be described by a simple effective Hamiltonian H_{eff} . In this way Wilson could show that the ratio of the impurity susceptibility, χ_{imp} , and the impurity contribution to the linear coefficient of specific heat, γ_{imp} , at $T = 0$, is twice the value of a non-interacting Fermi liquid: $R = 4\pi^2 \chi_{\text{imp}} / 3\gamma_{\text{imp}} = 2$. We refer the reader to the detailed description of such calculations in [1, 2], and we turn now to the numerical procedure for calculating physical properties, which can give results at all temperatures, including the crossover regions.

4 Calculation of physical properties

The ability of the method to yield thermodynamic, dynamic, and transport properties makes it very useful for interpreting experimental results.⁴ We shall first describe the calculation of thermodynamics and dynamics using conventional approaches (without use of the complete basis set, but including reduced density matrices for dynamics) [2, 7, 8]. In Sec. 5 we shall then discuss more recent approaches using the complete basis set and full density matrix [12, 10] (this division, however, is somewhat arbitrary).

Thermodynamics: conventional approach

Suppose we have diagonalized exactly the Hamiltonian for a quantum impurity model such as the Kondo model and have all the many-body eigenvalues E_q and eigenstates $|q\rangle$

$$H = \sum_q E_q |q\rangle\langle q| \equiv \sum_q E_q X_{qq}. \quad (37)$$

We can then calculate the partition function

$$Z(T) \equiv \text{Tr} e^{-H/k_B T} = \sum_q e^{-E_q/k_B T} \quad (38)$$

and hence the thermodynamics via the impurity contribution to the free energy $F_{\text{imp}}(T) = -k_B T \ln Z/Z_c$, where $Z_c = \text{Tr} e^{-H_c/k_B T}$ is the partition function for the non-interacting conduction electrons. In the NRG procedure we can only calculate the “shell partition functions” Z_m for the sequence of truncated Hamiltonians H_m

$$Z_m(T) \equiv \text{Tr} e^{-H_m/k_B T} = \sum_q e^{-E_q^m/k_B T} = \sum_q e^{-D_m \bar{E}_q^m/k_B T}. \quad (39)$$

We will have $Z_m(T) \approx Z(T)$ provided

1. we choose $k_B T = k_B T_m \ll E_{\text{max}}^m = D_m K(\Lambda)$ so that the contribution to the partition function from excited states $E_q^m > D_m K(\Lambda)$, not contained in Z_m , is negligible, and
2. the truncation error made in replacing H by H_m in equating (38) and (39) is small. This error has been estimated in [2] to be approximately $\Lambda^{-1} D_m/k_B T_m$.

Combining these two conditions requires that

$$\frac{1}{\Lambda} \ll \frac{k_B T_m}{D_m} \ll K(\Lambda). \quad (40)$$

The choice $k_B T = k_B T_m \approx D_m$ is reasonable and allows the thermodynamics to be calculated at a sequence of decreasing temperatures $k_B T_m \sim D_m$, $N = 0, 1, \dots$ from the truncated partition functions Z_m . The procedure yields essentially exact results. For small $\Lambda \lesssim 3$, the window

⁴Spatial correlations may also be investigated, see Ref. [36]

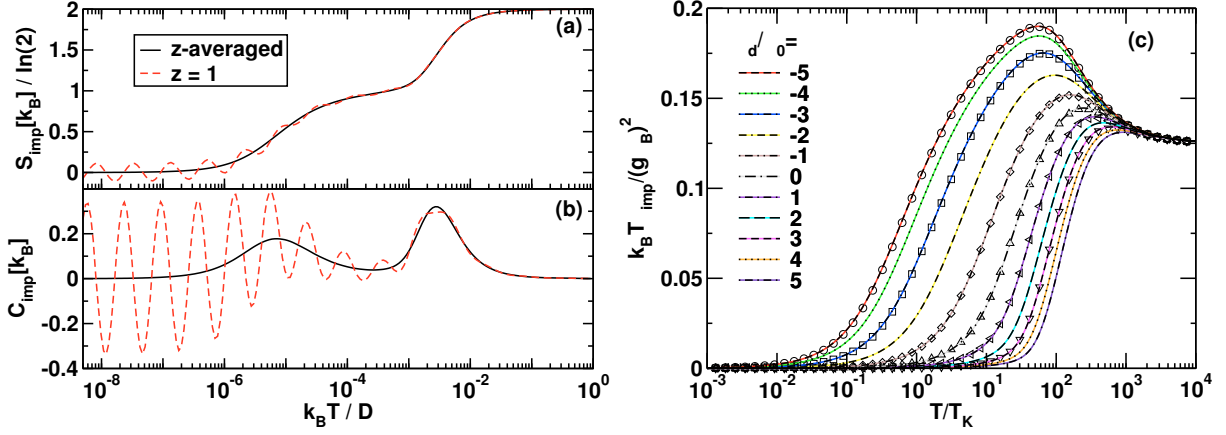


Fig. 6: Temperature dependence of, (a), the impurity entropy, $S_{\text{imp}}(T)$, and, (b), the impurity specific heat, $C_{\text{imp}}(T)$, for the symmetric Anderson model with $U/\Delta_0 = 12$ and $\Delta_0 = 0.001D$. The calculations are for $\Lambda = 4$, without z -averaging [$n_z = 1$, $z = 1$ (dashed lines)], and with z -averaging [$n_z = 2$, $z = 1/4, 3/4$ (solid lines)]. For $\Lambda = 4$ two z values suffice to eliminate the discretization oscillations [37]. (c) Impurity susceptibility, $\chi_{\text{imp}}(T)$, vs T/T_K for the asymmetric Anderson model with the same parameters as above and several values of ε_d/Δ_0 with T_K for the symmetric model. Broken lines: FDM approach. Solid lines: conventional approach. Symbols: Bethe ansatz (for selected values of $\varepsilon_d/\Delta_0 = -5, -3, -1, 0, +1, +3$). NRG parameters: $\Lambda = 10$ with z -averaging [$n_z = 4$, $z = 1/8, 1/2, 3/8, 3/4$] [38].

for choosing the temperature T_m to satisfy Eq. (40) is small, and typically only one such temperature is used for each shell. For larger $\Lambda \gg 1$ one can use many temperatures T_m^i , $i = 1, \dots, n_T$ that satisfy the above condition; however, for large $\Lambda = 4 - 10$, discretization oscillations become important [39, 40]. This problem is overcome by averaging the results over several discretizations of the band, i.e., one carries out several calculations with discretizations of the band $\pm D, \pm D\Lambda^{-(1-z_k)}, \pm D\Lambda^{-(2-z_k)}, \dots$ and averages the results for several z_k , $k = 1, \dots, N_z$. Figures 6a and b illustrate this for the entropy and specific heat of the Anderson model. In this way, the conventional approach can recover essentially exact results for thermodynamics. Fig. 6c shows a comparison for the impurity static spin susceptibility of the Anderson impurity model

$$\chi_{\text{imp}}(T) = \frac{(g\mu_B)^2}{k_B T} \left[\frac{1}{Z} \text{Tr} (S_z^{\text{tot}})^2 e^{-H/k_B T} - \frac{1}{Z_c} \text{Tr} (S_{z,c}^{\text{tot}})^2 e^{-H_c/k_B T} \right]$$

to both Bethe-ansatz results and results obtained within the more recent full density matrix approach to be described below.

Figure 7a illustrates the evolution of the impurity specific heat of the IRLM with increasing dissipation strength α : for $\alpha \ll 1$ the specific heat curve fits that of an isolated two-level system $C(T) \sim (T_0/T)^2 / \cosh^2(T_0/T)$ except at $T \ll T_0$ where the behavior is linear in T (as expected for any finite α), while for $\alpha \rightarrow 1$ the specific heat curves approach the universal Kondo specific heat curve for the isotropic Kondo model. Fig. 7b demonstrates also the very good agreement of the IRLM results with corresponding Bethe Ansatz calculations for the AKM, further illustrating the equivalence of these models for thermodynamic properties.

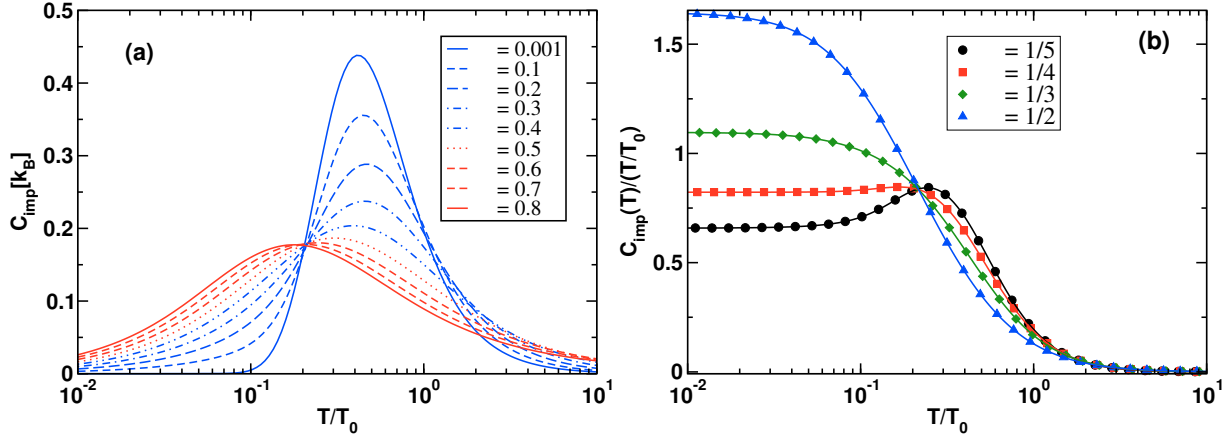


Fig. 7: (a) Evolution of the impurity specific heat, $C_{\text{imp}}(T)$, of the IRLM with increasing dissipation strength α , and, (b), $C_{\text{imp}}(T)/(T/T_0)$, for the IRLM at several α (symbols) compared to the corresponding Bethe Ansatz results for the equivalent AKM [31] (lines). The conventional approach was used for the NRG calculations with $\Lambda = 4$ and z -averaging with $n_z = 8$. The low-energy scale T_0 was extracted from the local charge susceptibility of the IRLM and corresponds to the renormalized tunneling amplitude Δ_r of this model, up to factors of order unity.

Dynamics: conventional approach without reduced density matrices

We consider now the application of the NRG method to the calculation of dynamic properties of quantum impurity models [41, 6–8]. For definiteness we consider the Anderson impurity model and illustrate the procedure for the impurity spectral density $A_{d\sigma}(\omega, T) = -\frac{1}{\pi} \text{Im} G_{d\sigma}(\omega, T)$ with

$$G_{d\sigma}(\omega, T) = \int_{-\infty}^{+\infty} d(t-t') e^{i\omega(t-t')} G_{d\sigma}(t-t') \quad (41)$$

$$G_{d\sigma}(t-t') = -i\theta(t-t') \langle [d_{\sigma}(t), d_{\sigma}^{\dagger}(t')]_+ \rangle_{\varrho} \quad (42)$$

with the density matrix ϱ of the system.

Suppose we have all the many-body eigenstates $|q\rangle$ and eigenvalues E_q of the Anderson impurity Hamiltonian H . Then the density matrix, $\varrho(T)$, of the full system at temperature $k_B T = 1/\beta$ can be written

$$\varrho(T) = \frac{1}{Z(T)} \sum_q e^{-\beta E_q} |q\rangle \langle q|, \quad (43)$$

the impurity Green function can be written in the Lehmann representation as

$$G_{d\sigma}(\omega, T) = \frac{1}{Z(T)} \sum_{q, q'} |\langle q | d_{\sigma} | q' \rangle|^2 \frac{e^{-E_q/k_B T} + e^{-E_{q'}/k_B T}}{\omega - (E_{q'} - E_q)} \quad (44)$$

and the corresponding impurity spectral density $A_{d\sigma}$ as

$$A_{d\sigma}(\omega, T) = \frac{1}{Z(T)} \sum_{q, q'} |M_{q, q'}|^2 (e^{-E_q/k_B T} + e^{-E_{q'}/k_B T}) \delta(\omega - (E_{q'} - E_q)) \quad (45)$$

with $M_{q, q'} = \langle q | d_{\mu} | q' \rangle$.

Consider first the case $T = 0$ ($T > 0$ is described in the next section), then

$$A_{d\sigma}(\omega, T = 0) = \frac{1}{Z(0)} \sum_q |M_{q,0}|^2 \delta(\omega + (E_q - E_0)) + \frac{1}{Z(0)} \sum_{q'} |M_{0,q'}|^2 \delta(\omega - (E_{q'} - E_0)), \quad (46)$$

with $E_0 = 0$ the ground-state energy. In order to evaluate this from the information which we actually obtain from an iterative diagonalization of H , we consider the impurity spectral densities corresponding to the sequence of Hamiltonians H_m , $m = 0, 1, \dots, N$,

$$A_{d\sigma}^m(\omega, T = 0) = \frac{1}{Z_m(0)} \sum_q |M_{q,0}^m|^2 \delta(\omega + E_q^m) + \frac{1}{Z_m(0)} \sum_{q'} |M_{0,q'}^m|^2 \delta(\omega - E_{q'}^m). \quad (47)$$

From the discussion on the spectrum of H_m in the previous section, it follows that the ground-state excitations of H_m that are representative of the infinite system H are those in the range $D_m \leq \omega \leq K(\Lambda)D_m$. Lower energy excitations and eigenstates are calculated more accurately at subsequent iterations, and higher energy excitations are not contained in H_m due to the elimination of the higher energy states at each m . Hence, for fixed m , we can approximately evaluate the spectral density at a characteristic frequency $\omega \approx \omega_m \equiv k_B T_m$ via

$$A_{d\sigma}(\omega, T = 0) \approx A_{d\sigma}^m(\omega, T = 0), \quad m = 0, 1, \dots, N. \quad (48)$$

In making this approximation, we are assuming that the matrix elements $M_{0,q'}^m$ of the finite-system Hamiltonian are the same as those of the infinite system $M_{0,q'}$. This assumption fails when an applied field strongly affects the groundstate and low lying excited states, thereby making also the matrix elements for the finite-size system $M_{0,q'}^m$ appreciably different from those of the infinite system. We shall come back to this point below, when we introduce the reduced density matrix approach to Green functions [8]. Returning to the calculation of spectral densities, a typical choice for the characteristic frequency to evaluate $A_{d\sigma}(\omega, 0)$ from $A_{d\sigma}^m(\omega, 0)$ is $\omega = 2\omega_m$ for $\Lambda = 2$. In this way $A_{d\sigma}(\omega, T = 0)$ can be calculated at a sequence of decreasing frequencies $\omega = 2\omega_m$, $m = 0, 1, \dots, N$ from the quantities $A_{d\sigma}^m$. In practice we are not interested in the discrete spectra $A_{d\sigma}^m(\omega) = \sum_q w_q^m \delta(\omega - E_q^m)$ of the Hamiltonians H_m but in continuous spectra that can be compared with experiment. Smooth spectra can be obtained from the discrete spectra by replacing the delta functions $\delta(\omega - E_q^m)$ by smooth distributions $P_m(\omega - E_q^m)$. A natural choice for the width η_m of P_m is D_m , the characteristic scale for the energy level structure of H_m . Two commonly used choices for P are the Gaussian and the Logarithmic Gaussian distributions [7, 41, 42]. More refined schemes also exist [43, 44], as well as different band discretizations to reduce artifacts close to band edges [45]. A peak of intrinsic width Γ at frequency Ω_0 will be well resolved by the above procedure provided that $\Omega_0 \ll \Gamma$, which is the case for the Kondo resonance and other low-energy resonances. In the opposite case, the low (logarithmic) resolution at higher frequencies may be insufficient to resolve the intrinsic widths and heights of such peaks. Usually such higher frequency peaks are due to single-particle processes and can be adequately described by other methods (exceptions include interaction dominated features in the ohmic two-state system, see below, and in strongly correlated lattice models in high dimensions [16, 46–48]). In both cases, $\Omega_0 \ll \Gamma$ and $\Omega_0 \gg \Gamma$, the

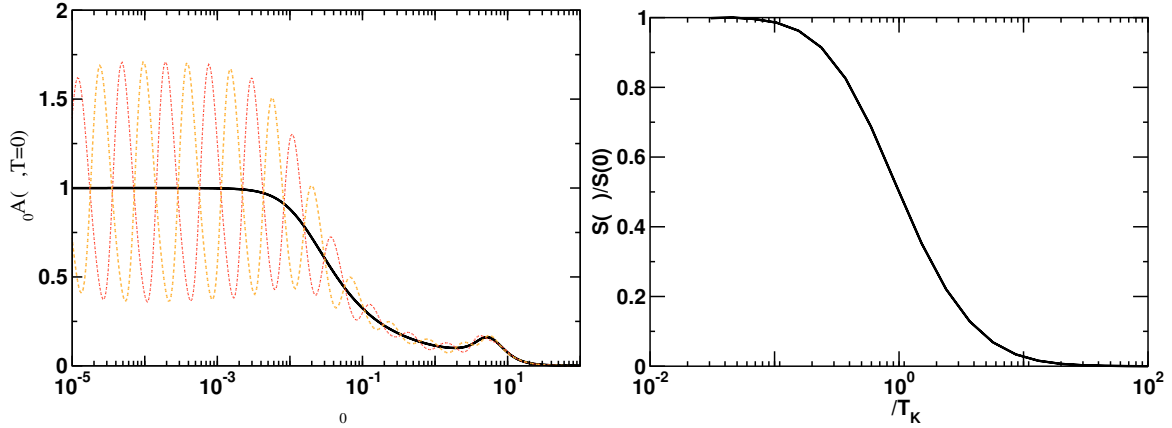


Fig. 8: (a) The impurity spectral density for the symmetric Anderson model for $U/\Delta_0 = 12$ at large $\Lambda = 10$ showing discretization oscillations for two values of z . Averaging over 8 z -values yields the smooth curve. (b). Longitudinal spin relaxation function $S(\omega)/S(0)$ versus ω/T_K for the isotropic Kondo model, showing that the spin relaxes incoherently.

positions and intensities of such peaks are given correctly. An alternative procedure for obtaining smooth spectra, which in principle resolves finite frequency peaks with the same resolution as the low-energy peaks, has been proposed in [49] and uses the averaging over several band discretizations, described above for the thermodynamics. This procedure allows carrying out calculations for spectral functions at larger Λ . An example is shown in Fig. 8a for the symmetric Anderson model. As in the thermodynamics, calculations of the dynamics at large $\Lambda \gg 1$ exhibit discretization oscillations, see Fig. 8a, which may be eliminated by averaging over several band discretizations.

How accurate is the NRG for dynamic properties? A good measure of the accuracy of the procedure is given by the Friedel sum rule, a Fermi liquid relation which states that [26]

$$A_{d\sigma}(0) = \frac{1}{\pi\Delta_0} \sin^2(\pi n_d/2), \quad n_d = \int_{-\infty}^0 d\omega A_{d\sigma}(\omega). \quad (49)$$

From Fig. 8a we find that $\pi\Delta_0 A_{d\sigma}(0,0) = 1 \pm 10^{-3}$, i.e., the Friedel sum rule is satisfied to within 0.1% relative error. More important, however, is that this error remains small independent of the interaction strength $0 \leq U \leq \infty$. Two-particle Green functions and response functions, such as the longitudinal dynamical spin susceptibility $\chi_{zz} = \langle\langle S_z; S_z \rangle\rangle$, and the corresponding relaxation function, $S(\omega) = -\frac{1}{\pi} \chi''_{zz}(\omega)/\omega$, of the Anderson impurity model and of the (anisotropic) Kondo model can also be calculated with comparable accuracy to single-particle spectral functions [30]. The spin relaxation function for the Kondo model is shown in Fig. 8b and illustrates the statement made in Sec. 2 that the spin excitations of the Kondo model are drastically renormalized from the bare value of J down to the Kondo scale T_K due to the frictional effects of the environment.

The procedure for calculating finite temperature dynamical quantities, like $A_{d\sigma}(\omega, T)$, required as input for calculating transport properties is similar to that for the $T = 0$ dynamics described above [7]. The spectral density $A_{d\sigma}(\omega, T)$ at fixed temperature T is evaluated as above at

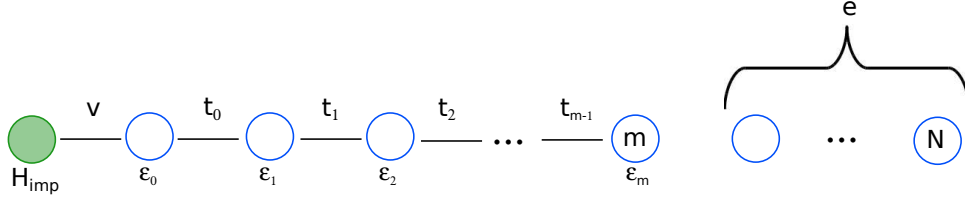


Fig. 9: The Hilbert space of H_m is supplemented with $N - m$ environment degrees of freedom $e = (\alpha_{m+1}, \dots, \alpha_N)$ [8].

frequencies $\omega \approx 2\omega_m, m = 0, 1, \dots, M \leq N$ until $2\omega_M$ becomes of order $k_B T$, i.e., $2\omega_M = \alpha k_B T$ with $\alpha \approx 1$. To calculate the spectral density at frequencies $\omega < k_B T$ a smaller “cluster” is used. This is done because when $k_B T$ is larger than the frequency at which the spectral density is being evaluated, it is the excited states of order $k_B T$ contained in previous clusters that are important and not the excitations very much below $k_B T$. This approach suffers from the same criticism as the $T = 0$ approach above, namely one is using a finite cluster H_m to approximate $M_{q',q} \approx M_{q',q}^m$ (and also $Z(T) \approx Z_m(T)$). In particular, for $\omega < k_B T$, the use of a small cluster of size $M < N$ does not capture the full information available, a deficiency that is corrected by the full density matrix approach. Nevertheless, this early approach gives remarkably good results for finite temperature spectra and transport properties [7].

Dynamics: conventional approach with reduced density matrices

A way of reducing finite-size errors, inherent in the above approach to Green functions, has been proposed by Hofstetter [8] and further developed within the full density matrix approach. As mentioned above, there are situations when a small field can strongly polarize the low-energy states of H_m , thereby strongly affecting the matrix elements $M_{q',q}^m$ and hence the spectra. For example, a magnetic field $B \approx T_K$ in the Anderson model is sufficiently strong to polarize the groundstate such that $n_{d\uparrow} \approx 1$ and $n_{d\downarrow} \approx 0$ at $T \ll T_K$. In this case, the use of the canonical density matrix $\varrho(T) \approx \varrho_m(T) = \frac{1}{Z_m(T)} \sum_q e^{-\beta E_q^m} |q\rangle \langle q|$ in evaluating the spectra on scales $\omega_m \gg T_K$ can result in large errors. A solution to this is to use $\varrho_N(T) = \frac{1}{Z_N(T)} \sum_q |q\rangle e^{-\beta E_q^N} \langle q|$ for the longest chain diagonalized and to evaluate the Green functions on scales $\omega_m > \omega_N$ by tracing out intermediate degrees of freedom $e = (\alpha_{m+1}, \dots, \alpha_N)$ in ϱ_N . Since the longest chain H_N is close to the infinite system limit, this should provide a better description of the spectra, particularly at higher frequencies. In order to carry out this procedure, the Hilbert space of each H_m is extended to that of H_N by adding the $N - m$ environment degrees of freedom e , see Fig. 9. Evaluating the reduced density matrix $\varrho_m^{\text{red}} = \text{Tr}_e [\varrho_N]$ appearing in Eq. (42) leads to a Lehmann representation for the spectral density at $T = 0$

$$A_{d\sigma}(\omega, T = 0) = \sum_{kk'} C_{kk'}^N M_{kk'}^N \delta(\omega - (E_\kappa^N - E_{k'}^N)) \quad (50)$$

$$C_{kk'}^N = \sum_p \varrho_{pk'}^{\text{red}} M_{p\kappa}^N + \sum_p \varrho_{kp}^{\text{red}} M_{k'p}^N \quad (51)$$

in place of (46). In Fig. 10 we show a comparison of this approach with results from the

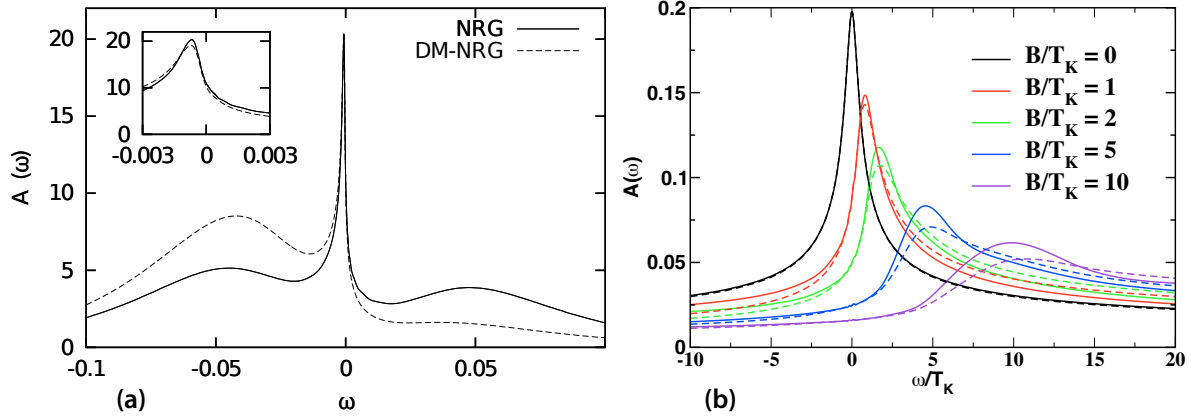


Fig. 10: Spin-up spectral density $A(\omega, T = 0)$ for, (a), the symmetric Anderson model, with $U/\Delta_0 = 10$, $\Delta_0 = 0.01$ and $B = \Delta_0/10$, with and without reduced density matrices (DM-NRG/NRG) [8], and, (b), for the Kondo model in several magnetic fields B , with and without reduced density matrices (dashed/solid lines, respectively) [11, 50], showing that the low-energy Kondo resonance is sufficiently well captured in the conventional approach.

previous approach for the Anderson and Kondo models in a magnetic field. A field-induced rearrangement of spectral weight at $\omega \approx \varepsilon_d, \varepsilon_d + U$ is well captured by the reduced density matrix approach (Fig. 10a). The low-energy Kondo resonance is less subject to finite size corrections, since this part of the spectrum is already calculated from sufficiently long chains, such that the corrections in using reduced density matrices are small (Fig. 10b).

Transport properties

The transport properties of quantum impurity models require knowledge of both the frequency and temperature dependence of the impurity spectral density, a topic that was addressed above. The linear-response conductance $G(T)$ and thermopower $S(T)$ through a quantum dot described by the Anderson model are given by the following expressions

$$G(T) = \frac{e^2}{h} \int d\omega \left(-\frac{\partial f}{\partial \omega} \right) \sum_{\sigma} \mathcal{T}_{\sigma}(\omega, T, B), \quad (52)$$

$$S(T) = -\frac{1}{|e|T} \frac{\int d\omega \omega (-\partial f / \partial \omega) \sum_{\sigma} \mathcal{T}_{\sigma}(\omega)}{\int d\omega (-\partial f / \partial \omega) \sum_{\sigma} \mathcal{T}_{\sigma}(\omega)}, \quad (53)$$

where the transmission function $\mathcal{T}_{\sigma}(\omega, T)$ through a quantum dot symmetrically coupled to left and right leads is related to $A_{d\sigma}(\omega, T)$ via

$$\mathcal{T}_{\sigma}(\omega, T) = 2\pi \Delta_0 A_{d\sigma}(\omega T).$$

Note that the discrete form of the spectral function may be directly substituted into the expressions for $G(T)$ and $S(T)$ above without the necessity of broadening [51]. For the conductance, this leads to

$$G(T) = \frac{\gamma\beta}{Z} \sum_{\sigma} \sum_{m,n} |M_{mn}^{\sigma}|^2 \frac{1}{e^{\beta E_m} + e^{\beta E_n}}, \quad (54)$$

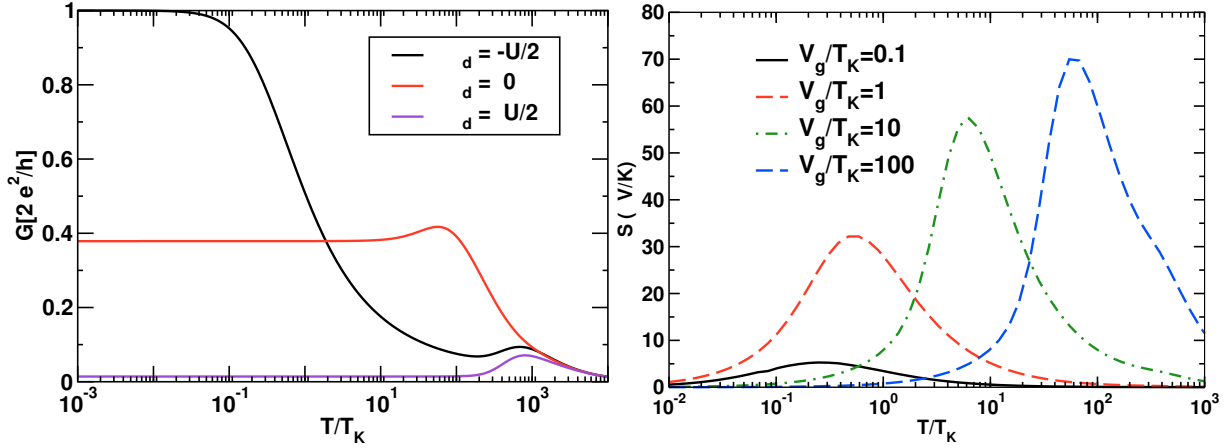


Fig. 11: (a) Linear conductance $G(T)$ versus T/T_K for $U/\Delta_0 = 16$ and several values of $\epsilon_d = -U/2, 0, +U/2$ using the approach of Yoshida et al. [51]. The resistivity of a Kondo impurity $\rho(T)$ is similar to $G(T)$ for the $\epsilon_d = -U/2$ curve. NRG parameters were for $\Lambda = 4$, $n_z = 2$ [52]. (b). Thermopower of a negative $U = -16\Delta_0$ quantum dot, exhibiting a large enhancement for gate voltages $V_g \geq T_K$ [53] (calculated within the full density matrix approach to spectral functions).

with $\gamma = 2\pi\Delta_0 \frac{e^2}{h}$. Results for the temperature dependence of the conductance of the Anderson model using this procedure are shown in Fig. 11a. Thermoelectric properties have also been investigated for quantum dots with repulsive onsite Coulomb interactions [54] and for attractive onsite interactions [53]. The latter provide a mechanism for enhancing thermopower, as shown in Fig. 11b. The method gives uniformly accurate results at high and low temperatures, as well as correctly describing the crossover region $T \approx T_K$ (detailed comparisons of the resistivity of dilute magnetic impurities with known results at high and low temperature can be found in [7]). These calculations, as well as similar resistivity calculations for dilute impurities, provide a quantitative interpretation of experiments for $S = 1/2$ realizations of the Kondo effect. They have also been extended using the full density matrix approach to describe the resistivity and dephasing rates of real Fe impurities in Au and Ag by using a 3-channel Kondo model [35,55].

5 Complete basis set and full density matrix

We noted in Sec. 3 that at each m , the states generated, denoted $|qm\rangle$, are partitioned into the lowest-energy retained states, denoted $|km\rangle$, and the high-energy eliminated (or discarded) states, $|lm\rangle$. In order to avoid an exponential increase in the dimension of the Hilbert space, only the former are used to set up and diagonalize the Hamiltonian for the next iteration $m + 1$. The eliminated states, while not used in the iterative NRG procedure, may be used to set up a complete orthonormal basis set [12]. This complete basis set is very powerful and allows evaluating correlation functions $\langle A(t)B(0) \rangle$, transient quantities, and even thermodynamic expressions in an unambiguous way, avoiding any possible double counting of excitations. Eliminated states from different iterations have no overlap, see Fig. 12, in contrast to the retained states. Hence,

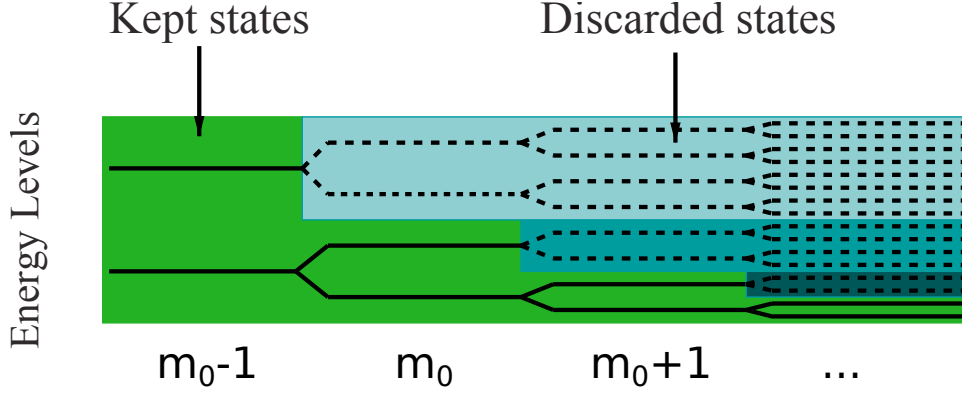


Fig. 12: For iterations $m < m_0$ all states are kept, while for $m \geq m_0$ only the lowest 1000 or so states generated are used to set up the Hamiltonian for the next iteration $m + 1$ (adapted from [57]).

using the latter to carry out calculations of physical quantities restricts one to using excitations from a single shell only. However, for finite-temperature Green functions and non-equilibrium quantities, multiple-shell contributions become important [56], and the complete basis set offers a way to evaluate these quantities [12].

The complete basis set is defined by the product states $|lem\rangle = |lm\rangle|e\rangle$, $m = m_0, \dots, N$, where m_0 is the first iteration at which truncation occurs, and $|e\rangle = |\alpha_{m+1}\rangle|\alpha_{m+2}\rangle \dots |\alpha_N\rangle$ are environment states at iteration m such that the product states $|lem\rangle$, for each $m = m_0, m_0 + 1, \dots, N$, reside in the same Fock space (that of the largest system diagonalized, $m = N$). By “e” we shall henceforth denote the collection $e = \{\alpha_{m+1} \dots \alpha_N\}$. The eliminated states satisfy the completeness relation [12, 58]

$$\sum_{m=m_0}^N \sum_{le} |lem\rangle \langle lem| = 1, \quad (55)$$

where for $m = N$ all states are counted as discarded (i.e., there are no kept states at iteration $m = N$). We shall also use the following representations of this relation [12, 58]

$$1 = 1_m^- + 1_m^+, \quad (56)$$

$$1_m^- = \sum_{m'=m_0}^m \sum_{l'e'} |l'e'm'\rangle \langle l'e'm'| \quad (57)$$

$$1_m^+ = \sum_{m'=m+1}^N \sum_{l'e'} |l'e'm'\rangle \langle l'e'm'| = \sum_{ke} |kem\rangle \langle kem|. \quad (58)$$

By using the complete basis set, we can construct the full density matrix FDM [10, 59]

$$\rho = \frac{1}{Z(T)} \sum_{m=m_0}^N \sum_{le} e^{-\beta E_l^m} |lem\rangle \langle lem|, \quad \text{Tr } \rho = 1 \Rightarrow \quad (59)$$

$$Z(T) = \sum_{m=m_0}^N 4^{N-m} \sum_l e^{-\beta E_l^m} \equiv \sum_{m=m_0}^N 4^{N-m} Z_m(T) \quad (60)$$

where $Z(T)$ is the partition function made up of the complete spectrum, i.e., it contains the eliminated states from all H_m , $m = m_0, m_0 + 1, \dots, N$. Consequently, it can be used to evaluate the impurity thermodynamics at arbitrary temperatures.

Consider the following density matrix for the m -th shell (defined, however, in the Hilbert space of H_N),

$$\tilde{\rho}_m = \sum_{le} |lem\rangle \frac{e^{-\beta E_l^m}}{\tilde{Z}_m} \langle lem|. \quad (61)$$

Normalization $\text{Tr} [\tilde{\rho}_m] = 1$ implies that

$$1 = \sum_l \frac{e^{-\beta E_l^m}}{\tilde{Z}_m} 4^{N-m} = 4^{N-m} \frac{Z_m}{\tilde{Z}_m}, \quad (62)$$

where $Z_m = \sum_l e^{-\beta E_l^m}$. Then the FDM can be written as a sum of weighted density matrices for shells $m = m_0, \dots, N$

$$\rho = \sum_{m=m_0}^N w_m \tilde{\rho}_m \quad (63)$$

$$w_m = 4^{N-m} \frac{Z_m}{Z}; \quad \sum_{m=m_0}^N w_m = 1 \quad (64)$$

Application to thermodynamics

Substituting $\rho = \sum_m w_m \tilde{\rho}_m$ into the expression for the thermodynamic average $\langle \hat{O} \rangle$ of a local observable of the impurity (e.g., n_d or $n_{d\uparrow} n_{d\downarrow}$) and making use of the decomposition of unity Eq. (55), we have

$$\begin{aligned} \langle \hat{O} \rangle_\rho &= \text{Tr} [\rho \hat{O}] = \sum_{l'e'm'} \left\langle l'e'm' \left| \hat{O} \sum_{lem} w_m \right| lem \right\rangle \frac{e^{-\beta E_l^m}}{\tilde{Z}_m} \langle lem | l'e'm' \rangle \\ &= \sum_{lem} O_{ll}^m w_m \frac{e^{-\beta E_l^m}}{\tilde{Z}_m} = \sum_{lm} 4^{N-m} w_m O_{ll}^m \frac{e^{-\beta E_l^m}}{4^{N-m} Z_m} = \sum_{m=m_0, l}^N w_m O_{ll}^m \frac{e^{-\beta E_l^m}}{Z_m}, \end{aligned} \quad (65)$$

where orthonormality $\langle lem | l'e'm' \rangle = \delta_{ll'} \delta_{ee'} \delta_{mm'}$, and the trace over the $N - m$ environment degrees of freedom $\sum_{lem} \dots = \sum_{lm} 4^{N-m} \dots$ has been used, and $O_{ll}^m = \langle lm | \hat{O} | lm \rangle$. For other observables, such as the specific heat or the susceptibility, one requires a similar calculation for the conduction band contribution $\langle \hat{O} \rangle_{\rho_0}$, with ρ_0 the FDM of the non-interacting band. The impurity contribution is then obtained as $O_{\text{imp}} = \langle \hat{O} \rangle_\rho - \langle \hat{O} \rangle_{\rho_0}$.

For each temperature T and shell m , we require $w_m(T)$ and the factor $B_l^m(T) = e^{-\beta E_l^m} / Z_m$ where $Z_m = \sum_l e^{-\beta E_l^m}$. Numerical problems due to large exponentials are avoided by calculating $B_l^m(T) = e^{-\beta(E_l^m - E_0^m)} / Z'_m$ where $Z'_m = e^{\beta E_0^m} Z_m$ and E_0^m is the lowest discarded energy for shell m . Figure 13 shows results for the double occupancy of the Anderson model obtained within the FDM approach and comparisons with the conventional approach of Sec. 4.

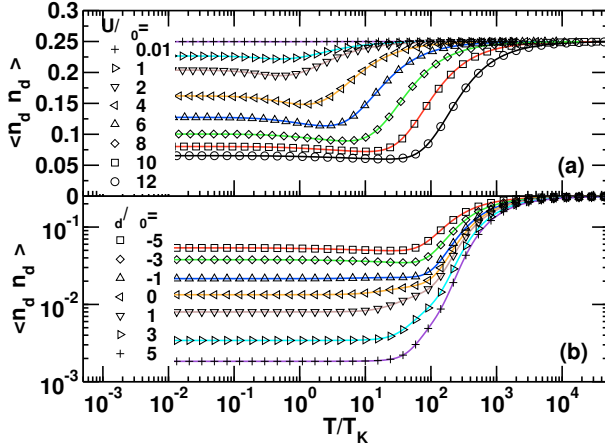


Fig. 13: Double occupancy $D_{\text{occ}} = \langle n_{d\uparrow} n_{d\downarrow} \rangle$ as a function of temperature T/T_K for, (a), the symmetric Anderson model at different U/Δ_0 , and, (b), for the asymmetric Anderson model at $U/\Delta_0 = 12$ and increasing values of $\varepsilon_d/\Delta_0 = -5, -3, -1, 0, +1, +3, +5$. FDM (solid lines), conventional approach (symbols). $\Lambda = 10$ with z -averaging [$n_z = 4$, $z = 1/8, 1/2, 3/8, 3/4$] [38].

Application to dynamics

We consider a general fermionic/bosonic ($s = \pm 1$) retarded Green function

$$\begin{aligned} G_{AB}(t) &= -i\theta(t)\langle [A(t), B]_s \rangle \equiv -i\theta(t) \text{Tr} [\rho(A(t)B + sBA(t))] \\ &= -i\theta(t) [C_{A(t)B} + sC_{BA(t)}], \end{aligned} \quad (66)$$

where A and B are fermionic/bosonic operators, e.g., for the d -level Green function of our quantum dot $A = d_\sigma$ and $B = d_\sigma^\dagger$. The trace is evaluated using the complete basis set. We outline the derivation of $C_{A(t)B}$, with the expression for $C_{BA(t)}$ obtained in a similar manner. We have

$$\begin{aligned} C_{A(t)B} &= \text{Tr} [\rho A(t)B] = \sum_{lem} \langle lem | e^{iHt} A e^{-iHt} B \rho | lem \rangle \\ &= \sum_{lem} \sum_{l'e'm'} \langle lem | e^{iHt} A e^{-iHt} | l'e'm' \rangle \langle l'e'm' | B \rho | lem \rangle, \end{aligned} \quad (67)$$

which consists of three contributions with $m' = m$, $m' > m$ and $m' < m$. Consider the first contribution ($m' = m$), denoted by $C_{A(t)B}^{(i)}$. Using the NRG approximation $e^{-iHt} |l'e'm\rangle \approx e^{-iH_m t} |l'e'm\rangle = e^{-iE_{l'}^m t} |l'e'm\rangle$ and $\langle lem | A | l'e'm \rangle = \delta_{ee'} \langle lm | A | l'm \rangle = \delta_{ee'} A_{ll'}^m$, we have

$$C_{A(t)B}^{(i)} = \sum_{lm} \sum_{l'} e^{-i(E_{l'}^m - E_l^m)t} A_{ll'}^m \sum_e \underbrace{\langle l'em | B \rho | lem \rangle}_{(B\rho)_{l'e,le}^m}.$$

Inserting the FDM expression $\rho = \sum_m w_m \tilde{\rho}_m$ into $(B\rho)_{l'e,le}^m$ yields

$$\sum_e (B\rho)_{l'e,le}^m = B_{l'l}^m w_m e^{-\beta E_l^m} / Z_m,$$

hence we have

$$C_{A(t)B}^{(i)} = \sum_m \frac{w_m}{Z_m} \sum_l \sum_{l'} e^{-i(E_{l'}^m - E_l^m)t} A_{ll'}^m B_{l'l}^m e^{-\beta E_l^m}. \quad (68)$$

The off-diagonal contributions with $m' > m$ and $m' < m$ in Eq. (67), which we label by $C_{A(t)B}^{(ii)}$ and $C_{A(t)B}^{(iii)}$, may be put into diagonal form by using $1_m^+ = \sum_{m'=m+1}^N \sum_{l'e'} |l'e'm'\rangle \langle l'e'm'| =$

$\sum_{ke} |kem\rangle\langle kem|$ [Eq. (58)], thereby introducing kept states at iteration m (or m') in place of discarded states at iterations $m' > m$ (or $m > m'$),

$$\begin{aligned}
C_{A(t)B}^{(ii)} &= \sum_{l'e'm' > m} \sum_{lem} \langle lem | e^{iHt} A e^{-iHt} | l'e'm' \rangle \langle l'e'm' | B \rho | lem \rangle \\
&= \sum_{lem} \sum_{ke'} \langle lem | e^{iHt} A e^{-iHt} | ke'm \rangle \langle ke'm | B \rho | lem \rangle \\
&\approx \sum_{lm} \sum_k e^{-i(E_k^m - E_l^m)t} A_{lk}^m \sum_e (B\rho)_{ke,le}^m \\
&= \sum_{lm} \sum_k e^{-i(E_k^m - E_l^m)t} A_{lk}^m B_{kl}^m e^{-\beta E_l^m} \frac{w_m}{Z_m}
\end{aligned} \tag{69}$$

$$\begin{aligned}
C_{A(t)B}^{(iii)} &= \sum_{lem > m'} \sum_{l'e'm'} \langle lem | e^{iHt} A e^{-iHt} | l'e'm' \rangle \langle l'e'm' | B \rho | lem \rangle \\
&= \sum_{l'e'm'} \sum_{ke'} \langle l'e'm' | B \rho | ke'm \rangle \langle ke'm | e^{iHt} A e^{-iHt} | l'e'm' \rangle \\
&\approx \sum_{l'm'} (B\rho)_{l'e',ke'} e^{-i(E_{l'}^{m'} - E_k^{m'})t} A_{kl'}^{m'}
\end{aligned} \tag{70}$$

where the NRG approximation has been used together with $\sum_e (B\rho)_{ke,le}^m = B_{kl}^m e^{-\beta E_l^m} \frac{w_m}{Z_m}$. It is also easy to show that [54],

$$\text{Tr}_e [(B\rho)_{le,ke}^m] \equiv \sum_e (B\rho)_{le,ke}^m = \sum_{k'} B_{lk'}^m \underbrace{\sum_e \langle k'em | \rho | kem \rangle}_{R_{\text{red}}^m(k',k)} \tag{71}$$

where $R_{\text{red}}^m(k', k)$ is the reduced density matrix obtained from the FDM ρ by tracing out the degrees of freedom $e = (\alpha_{m+1}, \dots, \alpha_N)$ [10, 54], hence the contribution $C_{A(t)B}^{(iii)}$ may be written as

$$C_{A(t)B}^{(iii)} = \sum_{lm} (B\rho)_{le,ke} e^{-i(E_l^m - E_k^m)t} A_{kl}^m = \sum_{lkk'm} e^{-i(E_l^m - E_k^m)t} A_{kl}^m B_{lk'}^m R_{\text{red}}^m(k', k), \tag{72}$$

and

$$\begin{aligned}
C_{A(t)B} &= C_{A(t)B}^{(i)} + C_{A(t)B}^{(ii)} + C_{A(t)B}^{(iii)} = \sum_m \frac{w_m}{Z_m} \sum_{ll'} e^{-i(E_{l'}^m - E_l^m)t} A_{ll'}^m B_{l'l}^m e^{-\beta E_l^m} \\
&\quad + \sum_m \frac{w_m}{Z_m} \sum_{lk} e^{-i(E_k^m - E_l^m)t} A_{lk}^m B_{kl}^m e^{-\beta E_l^m} \\
&\quad + \sum_m \sum_{lkk'} e^{-i(E_l^m - E_k^m)t} A_{kl}^m B_{lk'}^m R_{\text{red}}^m(k', k).
\end{aligned} \tag{73}$$

Similar arguments lead to an expression for $C_{BA(t)}$ (exercise),

$$\begin{aligned}
C_{BA(t)} &= C_{BA(t)}^{(i)} + C_{BA(t)}^{(ii)} + C_{BA(t)}^{(iii)} = \sum_m \frac{w_m}{Z_m} \sum_{ll'} e^{-i(E_{l'}^m - E_l^m)t} A_{ll'}^m B_{l'l}^m e^{-\beta E_{l'}^m} \\
&\quad + \sum_m \frac{w_m}{Z_m} \sum_{lk} e^{-i(E_l^m - E_k^m)t} A_{kl}^m B_{lk}^m e^{-\beta E_l^m} \\
&\quad + \sum_m \sum_{lkk'} e^{-i(E_k^m - E_l^m)t} A_{lk}^m B_{k'l}^m R_{\text{red}}^m(k, k').
\end{aligned} \tag{74}$$

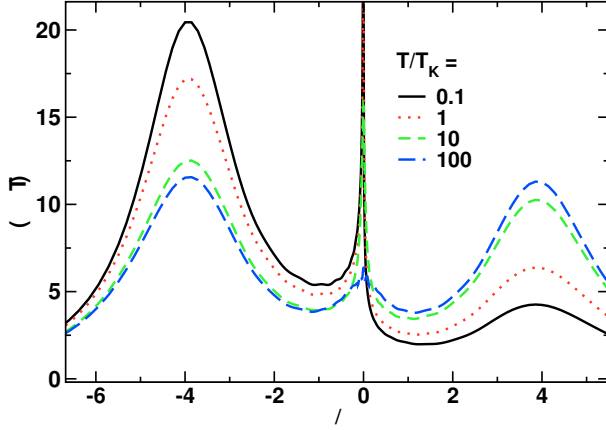


Fig. 14: Spectral function of a negative- U quantum dot modeled by the Anderson model with $U/\Gamma = -8$ at finite gate voltage $V_g = T_K$ and several temperatures, calculated within the FDM approach (adapted from [53]). The spectral function, polarized at low temperatures, becomes unpolarized at $T \gg T_K$.

Fourier transforming $-i\theta(t)(C_{A(t)B} + sC_{BA(t)})$ using

$$\int_{-\infty}^{+\infty} -i\theta(t)e^{-i(E_{q'}-E_q)t}e^{i(\omega+i\delta)t} = 1/(\omega + i\delta - (E_{q'} - E_q)) \quad (75)$$

finally yields the Green function

$$\begin{aligned} G_{AB}(\omega + i\delta) = & \sum_{m=m_0}^N \frac{w_m}{Z_m} \sum_{ll'} A_{ll'}^m B_{l'l}^m \frac{e^{-\beta E_l^m} + s e^{-\beta E_{l'}^m}}{\omega + i\delta - (E_{l'}^m - E_l^m)} \\ & + \sum_{m=m_0}^{N-1} \frac{w_m}{Z_m} \sum_{lk} A_{lk}^m B_{kl}^m \frac{e^{-\beta E_l^m}}{\omega + i\delta - (E_k^m - E_l^m)} \\ & + s \sum_{m=m_0}^{N-1} \frac{w_m}{Z_m} \sum_{lk} A_{kl}^m B_{lk}^m \frac{e^{-\beta E_l^m}}{\omega + i\delta - (E_l^m - E_k^m)} \\ & + \sum_{m=m_0}^{N-1} \sum_{lkk'} A_{kl}^m B_{lk'}^m \frac{R_{\text{red}}^m(k', k)}{\omega + i\delta - (E_l^m - E_k^m)} \\ & + s \sum_{m=m_0}^{N-1} \sum_{lkk'} A_{lk}^m B_{k'l}^m \frac{R_{\text{red}}^m(k, k')}{\omega + i\delta - (E_k^m - E_l^m)}. \end{aligned} \quad (76)$$

The reduced density matrices appearing in Eq. (76) can be evaluated efficiently at all temperatures in a recursive manner [10]. The use of the complete basis set to calculate finite-temperature Green functions ensures that the spectral sum rule $\int d\omega A_\sigma(\omega, T) = 1$ holds as an identity [10]. Furthermore, calculations at $\omega < T$ may be carried out without the need to restrict to a smaller cluster $M < N$, as was the case with the approach described in Sec. 4. Fig. 14 shows the spectral function of the negative- U Anderson model calculated from Eq. (76) at several temperatures.

For an application of this approach to thermoelectric properties of quantum dots see Ref. [54], and for a recent application to the magnetoresistivity and dephasing rate of multi-channel Kondo models see Ref. [35].

6 Recent developments: TDNRG and multiorbital Kondo physics

The NRG has proved to be a reliable method for dealing with equilibrium properties of strongly correlated quantum impurity systems. Nevertheless, the method is still under development. In this section, we describe two areas where significant progress has been made but where further work is needed. The first is in the transient response of a quantum impurity following either a quantum quench, a pulse of finite duration, or a periodic train of pulses [12, 56, 58, 60, 61]. This is relevant, for example, in many pump-probe experiments [62, 63]. The second area is in developing ways to deal with real quantum impurities in metals or on surfaces, such as impurities with partially filled d - or f -levels, in which multiple channels (or bands) of the host may couple to the impurity.

Time-dependent NRG (TDNRG)

We are interested in the dynamics of a local observable $\hat{O}$ following a quantum quench in which one or more system parameters of H change suddenly at $t = 0$. Thus, the time-dependence of H is described by $H(t) = \theta(-t)H^i + \theta(t)H^f$, with H^i and H^f being time-independent initial ($t < 0$) and final state ($t > 0$) Hamiltonians, respectively [56]. The time evolution of $\hat{O}$ at $t > 0$ is then given by $O(t) = \text{Tr} [\rho(t)\hat{O}]$ where $\rho(t) = e^{-iH^f t} \rho e^{iH^f t}$ is the time-evolved density matrix and $\rho = e^{-\beta H^i} / \text{Tr} [\rho]$ is the density matrix of the initial state at inverse temperature β . In terms of the complete basis set, we have

$$\begin{aligned} O(t) &= \text{Tr} \left[e^{-iH^f t} \rho e^{iH^f t} \hat{O} \right] = \sum_{m=m_0}^N \sum_{le} \langle lem | e^{-iH^f t} \rho e^{iH^f t} \hat{O} | lem \rangle_f \\ &= \sum_{mm'=m_0}^N \sum_{lel'e'} \langle lem | e^{-iH^f t} \rho e^{iH^f t} | l'e'm' \rangle_f \langle l'e'm' | \hat{O} | lem \rangle_f. \end{aligned} \quad (77)$$

Making use of $1_m^+ = \sum_{m'=m+1}^N \sum_{l'e'} |l'e'm'\rangle \langle l'e'm'| = \sum_{ke} |kem\rangle \langle kem|$ [Eq. (58)], allows us to write [12]

$$\begin{aligned} O(t) &= \sum_{m=m_0}^N \sum_{rs \notin KK'} \sum_e \langle sem | e^{-iH^f t} \rho e^{iH^f t} | rem \rangle_f \langle rem | \hat{O} | sem \rangle_f \\ &\approx \sum_{m=m_0}^N \sum_{rs \notin KK'} \sum_e \langle sem | e^{-iH_m^f t} \rho e^{iH_m^f t} | rem \rangle_f \langle rem | \hat{O} | sem \rangle_f \\ &= \sum_{m=m_0}^N \sum_{rs \notin KK'} \left(\sum_e \langle sem | \rho | rem \rangle_f \right) e^{-i(E_s^m - E_r^m)t} O_{rs}^m \\ &= \sum_{m=m_0}^N \sum_{rs \notin KK'} \rho_{sr}^{i \rightarrow f}(m) e^{-i(E_s^m - E_r^m)t} O_{rs}^m, \end{aligned} \quad (78)$$

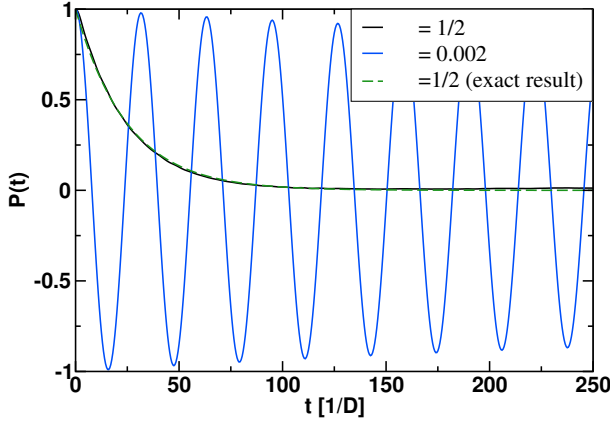


Fig. 15: $P(t) = \langle \sigma_z(t) \rangle$ of the IRLM (with $\sigma_z = 2n_d - 1$) using the TDNRG as formulated in Ref. [60]. In the initial state $\varepsilon_d = -\infty$ so that $P(t \leq 0) = +1$, while in the final state at $t > 0$ the level is shifted to $\varepsilon_d = 0$ so that the two-level system can relax to its new groundstate. For weak dissipation $\alpha \ll 1$, one observes weakly damped coherent oscillations. These vanish for $\alpha \geq 1/2$. The NRG parameters used are $\Lambda = 4$, $n_z = 32$ with resonant level width $\Gamma = 0.02 D$, where $D = 1$ is the half-width of the IRLM (and a semi-elliptic density of states was used).

in which r and s may not both be kept states, $O_{rs}^m = {}_f \langle lem | \hat{O} | rem \rangle_f$ are the final state matrix elements of $\hat{O}$, which are independent of e , the NRG approximation

$$H^f |rem\rangle \approx H_m^f |rem\rangle = E_r^m |rem\rangle, \quad (79)$$

is adopted [in the second line of Eq. (78)], and $\rho_{sr}^{i \rightarrow f}(m) = \sum_e {}_f \langle sem | \rho | rem \rangle_f$ represents the reduced density matrix of the initial state projected onto the basis of final states (henceforth called the *projected density matrix*). The latter has been evaluated for the special choice of a density matrix defined on the longest Wilson chain

$$\rho = \sum_l |lN\rangle_i \frac{e^{-\beta E_l^N}}{Z_N} {}_i \langle lN|, \quad (80)$$

with $Z_N = \sum_l e^{-\beta E_l^N}$, in which only the discarded states of the last NRG iteration enter [12,58]. More recently, the projected density matrix has been evaluated for a general initial density matrix, given by the full density matrix of the initial state [60]. This allows calculations to be carried out at arbitrary finite temperature. While the short-time limit $O(t \rightarrow 0^+)$ in the TDNRG recovers the exact thermodynamic value $O^i = \text{Tr}[\rho O]$ in the initial state, the long-time limit suffers from an error of a few percent. In addition, significant noise is observed at intermediate $t\Gamma \gtrsim 1$ to long times $t\Gamma \gg 1$. Attempts to further improve the method may be found in Ref. [60] and references therein. A generalization of the single-quench TDNRG formalism to multiple quenches, allowing applications to systems subject to general pulses or periodic driving fields, may be found in Ref. [61]. Figure 15 shows results for the quantity $P(t) = \langle \sigma_z(t) \rangle$ of the ohmic spin boson model (calculated via the IRLM using the equivalence between models discussed earlier). The aforementioned error of a few percent in the long time limit is evident in the case $\alpha = 1/2$, whose exact result is $P(t) = e^{-2\Gamma t}$. This is likely due to the finite heat capacity of the logarithmically discretized bath, which implies that the energy change following a quench cannot be fully dissipated into such a bath [64].

Multi-orbital and multi-channel Kondo models

The Anderson impurity model is a starting point for describing many different systems, from the classic examples of transition-metal magnetic impurities such as Fe or Mn in non-magnetic metals such as Au, to rare-earth magnetic impurities in non-magnetic metals, such as Ce in LaAl_2 [26] or magnetic ions such as Co, Fe and Ti adsorbed on surfaces of non-magnetic metals such as Cu or $\text{Cu}_2\text{N}/\text{Cu}$ (where the Cu_2N monolayer reduces the hybridization V to the substrate [66]). Of course, the relevant correlated orbitals in these systems are not the non-degenerate “ s -levels” as in (3), but would be the 5-fold or 7-fold degenerate partially filled d - or f -orbitals in the case of transition metal or rare earth metal impurities, respectively. Furthermore, electrons in these partially filled shells would be subject to Coulomb, Hund’s exchange, spin-orbit and crystal-field interactions, often leading to non-degenerate low-energy multiplets. In addition, these low-energy multiplets would hybridize with conduction channels of appropriate symmetry, and in general with many channels, not just one as in (3). Such a non-degenerate multi-channel Anderson model capable of describing a real transition-metal impurity would then look more complicated than Eq. (3), e.g., the following model (but still neglecting spin-orbit and crystal field interactions),

$$\begin{aligned}
 H = & \sum_{m\sigma} \varepsilon_{dm} n_{m\sigma} + \frac{U}{2} \sum_{m\sigma} n_{m\sigma} n_{m-\sigma} + \frac{U'}{2} \sum_{m \neq m'\sigma} n_{m\sigma} n_{m'-\sigma} + \frac{U' - J}{2} \sum_{m \neq m'\sigma} n_{m\sigma} n_{m'\sigma} \\
 & - \frac{J}{2} \sum_{m \neq m'\sigma} d_{m\sigma}^\dagger d_{m-\sigma} d_{m'-\sigma}^\dagger d_{m'\sigma} - \frac{J'}{2} \sum_{m \neq m'\sigma} d_{m\sigma}^\dagger d_{m-\sigma}^\dagger d_{m'-\sigma} d_{m'\sigma} \\
 & + \sum_{km\sigma} \epsilon_{km\sigma} c_{km\sigma}^\dagger c_{km\sigma} + \sum_{km\sigma} V_{km\sigma} (c_{km\sigma}^\dagger d_{m\sigma} + h.c.)
 \end{aligned}$$

would be closer to describing a real transition-metal impurity such as Mn in Cu. Despite its apparent complexity, this model, just like its simpler counterpart in Eq. (3), has the same general structure as Eq. (2) describing a general quantum impurity model, namely all many-body interactions (U, U', J, J') are contained in a local part H_{imp} , while the multi-channel bath H_{bath} represents non-interacting electrons coupling via a one-body hybridization to H_{imp} . While the NRG can be applied to such multi-channel models, for N_c -channels the Hilbert space grows as 4^{N_c} instead of 4 as for a single channel. The fraction of states that can be retained at each iteration is correspondingly smaller ($1/4^{N_c}$) than for a single channel ($1/4$), making accurate calculations difficult, particularly for dynamical quantities. While implementing all available symmetries ($U(1), SU(2), SU(3)$, parity etc.), in order to increase the fraction of states that can be retained at each iteration, will help, such symmetries are not always present. At present, reliable NRG calculations for dynamics can be carried out for three-channel models [35]. It should be emphasized that the difficulty is with the number of channels that couple to the impurity, not the complexity or number of orbitals on the impurity. Thus, while resistivity calculations for three-channel fully screened Kondo models, such as those shown in Fig. 16a, are demanding and require full use of all available symmetries, it is relatively straightforward to deal with underscreened Kondo models [67] with high spin values, as shown in Fig. 16b

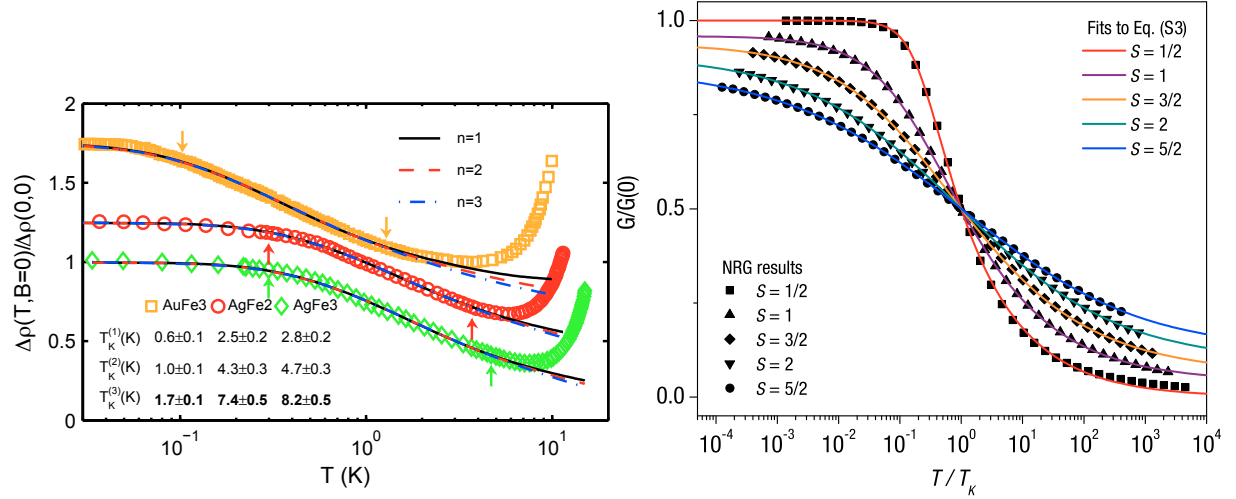


Fig. 16: Temperature dependence of, (a), the Kondo resistivity of one, two and three-channel fully screened Kondo models compared to experimental data for the Kondo contribution $\Delta\rho$ to the resistivity of Fe impurities in Au and Ag [35, 55]. Consistency in the extracted Kondo scales T_K^n for resistivity and dephasing measurements allows one to conclude that only the $S = 3/2$ three-channel Kondo model could fit all experimental data. (b), conductance curves $G(T)$ for $S > 1/2$ single-channel underscreened Kondo models (contrasted with the fully screened $S = 1/2$ case) from Ref. [65].

for the resistivity curves for the single-channel underscreened Kondo model [65, 68]. Increasing the number of channels to five would be a significant development, allowing many interesting realistic systems to be investigated with NRG in combination with *ab-initio* methods to extract the relevant model parameters [55, 69, 70]. We mention here one recent proposal for achieving this, which, however, has so far only been benchmarked on a three-channel model [71]. As in the single-band case, we rewrite the above model in linear-chain form with $H_{\text{bath}} = \sum_{m=1}^{N_c} \sum_{k\sigma} \epsilon_{km\sigma} c_{km\sigma}^\dagger c_{km\sigma} \rightarrow \sum_{m=1}^{N_c} \sum_{n=0}^{\infty} \sum_{\sigma} t_{mn} (f_{mn\sigma}^\dagger f_{mn+1\sigma} + H.c.)$, where $t_{mn} \sim D_m \Lambda^{-n/2}$, $m = 1, \dots, N_c$ for N_c channels with half-bandwidths D_m . If all channels have the same half-bandwidth $D_m = D$, $m = 1, \dots, N_c$, the hoppings within a shell t_{mn} , $m = 1, \dots, N_c$ are constant, one has to add all orbitals $f_{mn+1\sigma}$, $m = 1, \dots, N_c$ of the next shell $n + 1$ in going from H_n to H_{n+1} in the NRG procedure of Sec. 3 before truncating the spectrum of H_{n+1} , hence leading to the above growth of the Hilbert space at each iteration. Choosing band-widths D_m with $D_1 > D_2 > \dots > D_{N_c}$, as suggested in Ref. [71], leads to an energy scale separation of the orbitals within each shell, i.e., $t_{mn} \sim D_m \Lambda^{-n/2}$ for fixed n decrease with $m = 1, \dots, N_c$. This allows adding the orbitals $f_{mn+1\sigma}$, $m = 1, \dots, N_c$ of a given shell sequentially while simultaneously truncating the spectrum after each orbital is added. The calculation then resembles a single-channel calculation. The above energy scale separation is guaranteed provided $D_m/D_{m+1} = 1/g = \Lambda^{-1/2N_c}$, implying $t_{m+1n}/t_{mn} = g < 1$. Since the hoppings in this approach decrease by a factor $\Lambda^{-1/2N_c}$, a larger Λ will be required to obtain the same accuracy as a single-channel calculation. In this way, the authors obtained accurate results for three-channel and three-impurity models.

7 Summary

Wilson's non-perturbative NRG transformation for the Kondo model has become a powerful tool for the study of quantum impurity models in general. It gives information on the many-body eigenvalues and eigenstates of such models on all energy scales and thereby allows the direct calculation of their thermodynamic, dynamic, and transport properties. Recently, it has been further developed to yield the transient response of these systems to a sudden perturbation (a quantum quench) [12], the time-dependent NRG (TDNRG). Extensions of the TDNRG to general pulses using multiple quenches have also been made [60, 61]. The NRG also has potential to give information on the non-equilibrium steady-state transport through correlated impurity systems such as quantum dots. Recent work tries to construct a non-equilibrium density matrix for such systems by using the TDNRG to time-evolve from a known initial density matrix [72].

The method has been extended in new directions, such as to models with bosonic baths to study spin-boson models [27] and the interplay of correlations and phonon effects in Anderson-Holstein models [73]. It has also been used successfully to make progress on understanding the Mott transition, heavy fermion behavior, and other phenomena in correlated lattice systems [42, 74–76]. There is room for further improvement and extensions of the method both technically and in the investigation of more complex systems such as multi-impurity and multi-channel models [35, 71, 77].

Acknowledgement

Useful discussions with R. Bulla, H. Nghiem, L. Merker, D. Kennes and V. Meden are acknowledged (as is use of figures from joint work). Many relevant papers have not been cited in these lecture notes and we apologize to the authors for this. For a more comprehensive survey of the literature on NRG work, see Ref. [4].

Appendices

A Logarithmic discretization approximation

The approximation

$$H_c = \int_{-1}^{+1} d\varepsilon \varepsilon c_{\varepsilon\sigma}^\dagger c_{\varepsilon\sigma} \approx \sum_{n=0}^{\infty} \left(\varepsilon_{-n} c_{-n\sigma}^\dagger c_{-n\sigma} + \varepsilon_{+n} c_{+n\sigma}^\dagger c_{+n\sigma} \right) \quad (81)$$

used to replace the continuum band by the discrete one can be analyzed by introducing a complete orthonormal basis set of states for the conduction electrons in each interval $\pm[\Lambda^{-(n+1)}, \Lambda^{-n}]$ using the following wavefunctions

$$\psi_{np}^\pm(\varepsilon) = \begin{cases} \frac{\Lambda^{n/2}}{(1-\Lambda^{-1})^{1/2}} e^{\pm i\omega_n p \varepsilon} & \text{for } \Lambda^{-(n+1)} < \pm\varepsilon < \Lambda^{-n} \\ 0 & \text{otherwise} \end{cases} \quad (82)$$

Here p is a Fourier harmonic index and $\omega_n = 2\pi\Lambda^n/(1-\Lambda^{-1})$. The operators $c_{\varepsilon\sigma}$ can be expanded in terms of a complete set of new operators $a_{np\sigma}, b_{np\sigma}$ labeled by the interval n and the harmonic index p

$$c_{\varepsilon\sigma} = \sum_{np} [a_{np\sigma} \psi_{np}^+(\varepsilon) + b_{np\sigma} \psi_{np}^-(\varepsilon)]. \quad (83)$$

In terms of these operators, the Kondo Hamiltonian becomes

$$\begin{aligned} H_{KM} = & \frac{1}{2} (1 + \Lambda^{-1}) \sum_{np} \Lambda^{-n} (a_{np\sigma}^\dagger a_{np\sigma} - b_{np\sigma}^\dagger b_{np\sigma}) \\ & + \frac{(1 - \Lambda^{-1})}{2\pi i} \sum_n \sum_{p \neq p'} \Lambda^{-n} (a_{np\sigma}^\dagger a_{np'\sigma} - b_{np\sigma}^\dagger b_{np'\sigma}) e^{\frac{2\pi i(p-p')}{1-\Lambda^{-1}}} \\ & + J \sum_{\sigma\sigma'} f_{0\sigma}^\dagger \vec{\sigma}_{\sigma\sigma'} f_{0\sigma'} \cdot \vec{S}, \end{aligned} \quad (84)$$

where in terms of the new operators, $f_{0\sigma} = \frac{1}{\sqrt{2}} \int_{-1}^{+1} d\varepsilon c_{\varepsilon\sigma}$ contains only $p = 0$ states:

$$f_{0\sigma} = \frac{1}{\sqrt{2}} \int_{-1}^{+1} d\varepsilon c_{\varepsilon\sigma} = \left[\frac{1}{2} (1 - \Lambda^{-1}) \right]^{1/2} \sum_{n=0}^{\infty} \Lambda^{-n/2} (a_{n0\sigma} + b_{n0\sigma}). \quad (85)$$

We notice that only the $p = 0$ harmonic appears in the local Wannier state. This is a consequence of the assumption that the Kondo exchange is independent of k . Hence the conduction electron orbitals a_{np}, b_{np} for $p \neq 0$ only couple to the impurity spin indirectly via their coupling to the a_{n0}, b_{n0} in the second term of Eq. (84). This coupling is weak, being proportional to $(1 - \Lambda^{-1})$, and vanishes in the continuum limit $\Lambda \rightarrow 1$, so these states may be expected to contribute little to the impurity properties compared to the $p = 0$ states. This is indeed the case as shown by explicit calculations in [1, 2]. The logarithmic discretization approximation consists of neglecting conduction electron states with $p \neq 0$, resulting in H_c given by Eq. (81) with $c_{+n\sigma} \equiv a_{n,0\sigma}$ and $c_{-n\sigma} \equiv b_{n,0\sigma}$ and a discrete Kondo Hamiltonian given by Eq. (10).

B Lanczos procedure

Neglecting spin indices, the conduction electron operator is

$$H_c = \sum_k \varepsilon_k c_k^\dagger c_k .$$

The Lanczos algorithm for tridiagonalizing this operator by repeated action on the normalized conduction electron Wannier state $|0\rangle = \frac{1}{\sqrt{N}} \sum_k c_k^\dagger |vac\rangle$, with $|vac\rangle$ the vacuum state and N the number of sites in the crystal, is

$$|1\rangle = \frac{1}{t_0} (H_c |0\rangle - |0\rangle \langle 0 | H_c | 0 \rangle) \quad (86)$$

$$|n+1\rangle = \frac{1}{t_n} (H_c |n\rangle - |n\rangle \langle n | H_c | n \rangle - |n-1\rangle \langle n-1 | H_c | n \rangle) \quad (87)$$

yielding

$$H_c = \sum_{n=0}^{\infty} \epsilon_n f_n^\dagger f_n + t_n (f_n^\dagger f_{n+1} + H.c.) , \quad (88)$$

where the site energies are given by $\epsilon_n = \langle n | H_c | n \rangle$ and the hoppings t_n are obtained as normalizations from Eqs. (86)-(87).

References

- [1] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
- [2] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1003 (1980)
- [3] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1044 (1980)
- [4] R. Bulla, T.A. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)
- [5] H.O. Frota and L.N. Oliveira, Phys. Rev. B **33**, 7871 (1986)
- [6] T.A. Costi and A.C. Hewson, Philosophical Magazine Part B **65**, 1165 (1992)
- [7] T.A. Costi, A.C. Hewson, and V. Zlatić, J. Phys.: Condens. Matter **6**, 2519 (1994)
- [8] W. Hofstetter, Phys. Rev. Lett. **85**, 1508 (2000)
- [9] R. Peters, T. Pruschke, and F.B. Anders, Phys. Rev. B **74**, 245114 (2006)
- [10] A. Weichselbaum and J. von Delft, Phys. Rev. Lett. **99**, 076402 (2007)
- [11] T.A. Costi, Phys. Rev. Lett. **85**, 1504 (2000)
- [12] F.B. Anders and A. Schiller, Phys. Rev. Lett. **95**, 196801 (2005)
- [13] B.A. Jones, C.M. Varma, and J.W. Wilkins, Phys. Rev. Lett. **61**, 125 (1988)
- [14] A.J. Leggett, S. Chakravarty, A.T. Dorsey, M.P.A. Fisher, A. Garg, and W. Zwerger, Rev. Mod. Phys. **59**, 1 (1987)
- [15] U. Weiss: *Quantum dissipative systems*, Vol. 13 (World Scientific Pub Co Inc, 2008)
- [16] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [17] W. de Haas, J. de Boer, and G. van den Berg, Physica **1**, 1115 (1934)
- [18] P.W. Anderson, Phys. Rev. **124**, 41 (1961)
- [19] P.W. Anderson, Journal of Physics C: Solid State Physics **3**, 2436 (1970)
- [20] R. Bulla, T. Pruschke, and A.C. Hewson, Journal of Physics: Condensed Matter **9**, 10463 (1997)
- [21] C. Gonzalez-Buxton and K. Ingersent, Phys. Rev. B **57**, 14254 (1998)
- [22] D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, and M.A. Kastner, Nature **391**, 156 (1998)
- [23] D. Goldhaber-Gordon, J. Göres, M.A. Kastner, H. Shtrikman, D. Mahalu, and U. Meirav, Phys. Rev. Lett. **81**, 5225 (1998)

- [24] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
- [25] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)
- [26] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1997)
- [27] R. Bulla, H.-J. Lee, N.-H. Tong, and M. Vojta, Phys. Rev. B **71**, 045122 (2005)
- [28] J. von Delft and H. Schoeller, Annalen der Physik **7**, 225 (1998)
- [29] F. Guinea, V. Hakim, and A. Muramatsu, Phys. Rev. B **32**, 4410 (1985)
- [30] T.A. Costi and C. Kieffer, Phys. Rev. Lett. **76**, 1683 (1996)
- [31] T.A. Costi and G. Zarand, Physical Review B **59**, 12398 (1999)
- [32] F. Lesage and H. Saleur, Phys. Rev. Lett. **80**, 4370 (1998)
- [33] F. Verstraete, V. Murg, and J.I. Cirac, Advances in Physics **57**, 143 (2008)
- [34] A. Weichselbaum, Annals of Physics **327**, 2972 (2012)
- [35] M. Hanl, A. Weichselbaum, T.A. Costi, F. Mallet, L. Saminadayar, C. Bäuerle, and J. von Delft, Phys. Rev. B **88**, 075146 (2013)
- [36] L. Borda, Phys. Rev. B **75**, 041307 (2007)
- [37] L. Merker and T.A. Costi, Phys. Rev. B **86**, 075150 (2012)
- [38] L. Merker, A. Weichselbaum, and T.A. Costi, Phys. Rev. B **86**, 075153 (2012)
- [39] W.C. Oliveira and L.N. Oliveira, Phys. Rev. B **49**, 11986 (1994)
- [40] V.L. Campo and L.N. Oliveira, Phys. Rev. B **72**, 104432 (2005)
- [41] O. Sakai, Y. Shimizu, and T. Kasuya, Journal of the Physical Society of Japan **58**, 3666 (1989)
- [42] R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64**, 045103 (2001)
- [43] A. Freyn and S. Florens, Phys. Rev. B **79**, 121102 (2009)
- [44] R. Zitko, Computer Physics Communications **180**, 1271 (2009)
- [45] R. Zitko and T. Pruschke, Phys. Rev. B **79**, 085106 (2009)
- [46] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [47] G. Kotliar and D. Vollhardt, Physics Today **57**, 53 (2004)
- [48] D. Vollhardt, Annalen der Physik **524**, 1 (2012)

- [49] M. Yoshida, M.A. Whitaker, and L.N. Oliveira, Phys. Rev. B **41**, 9403 (1990)
- [50] T.A. Costi, eprint arXiv:cond-mat/0212651 (2002)
- [51] M. Yoshida, A.C. Seridonio, and L.N. Oliveira, Phys. Rev. B **80**, 235317 (2009)
- [52] L. Merker, S. Kirchner, E. Muñoz, and T.A. Costi, Phys. Rev. B **87**, 165132 (2013)
- [53] S. Andergassen, T.A. Costi, and V. Zlatić, Phys. Rev. B **84**, 241107 (2011)
- [54] T.A. Costi and V. Zlatić, Phys. Rev. B **81**, 235127 (2010)
- [55] T.A. Costi, L. Bergqvist, A. Weichselbaum, J. von Delft, T. Micklitz, A. Rosch, P. Mavropoulos, P.H. Dederichs, F. Mallet, L. Saminadayar, and C. Bäuerle, Phys. Rev. Lett. **102**, 056802 (2009)
- [56] T.A. Costi, Phys. Rev. B **55**, 3003 (1997)
- [57] A.I. Tóth, C.P. Moca, O. Legeza, and G. Zaránd, Phys. Rev. B **78**, 245109 (2008)
- [58] F.B. Anders and A. Schiller, Phys. Rev. B **74**, 245113 (2006)
- [59] A. Weichselbaum, Phys. Rev. B **86**, 245124 (2012)
- [60] H.T.M. Nghiem and T.A. Costi, Phys. Rev. B **89**, 075118 (2014)
- [61] H.T.M. Nghiem and T.A. Costi, Phys. Rev. B **90**, 035129 (2014)
- [62] L. Perfetti, P.A. Loukakos, M. Lisowski, U. Bovensiepen, H. Berger, S. Biermann, P.S. Cornaglia, A. Georges, and M. Wolf, Phys. Rev. Lett. **97**, 067402 (2006)
- [63] S. Loth, M. Etzkorn, C.P. Lutz, D.M. Eigler, and A.J. Heinrich, Science **329**, 1628 (2010)
- [64] A. Rosch, The European Physical Journal B **85**, 1 (2012)
- [65] J.J. Parks, A.R. Champagne, T.A. Costi, W.W. Shum, A.N. Pasupathy, E. Neuscamman, S. Flores-Torres, P.S. Cornaglia, A.A. Aligia, C.A. Balseiro, G.K.-L. Chan, H.D. Abruña, and D.C. Ralph, Science **328**, 1370 (2010)
- [66] A. Otte, M. Ternes, K. Von Bergmann, S. Loth, H. Brune, C. Lutz, C. Hirjibehedin, and A. Heinrich, Nature Physics **4**, 847 (2008)
- [67] A. Posazhennikova and P. Coleman, Phys. Rev. Lett. **94**, 036802 (2005)
- [68] N. Roch, S. Florens, T.A. Costi, W. Wernsdorfer, and F. Balestro, Phys. Rev. Lett. **103**, 197202 (2009)
- [69] P. Lucignano, R. Mazzarello, A. Smogunov, M. Fabrizio, and E. Tosatti, Nature Materials **8**, 563 (2009)

- [70] R. Requist, S. Modesti, P.P. Baruselli, A. Smogunov, M. Fabrizio, and E. Tosatti, Proceedings of the National Academy of Sciences pp. 69–74 (2013)
- [71] A.K. Mitchell, M.R. Galpin, S. Wilson-Fletcher, D.E. Logan, and R. Bulla, Phys. Rev. B **89**, 121105 (2014)
- [72] F.B. Anders, Phys. Rev. Lett. **101**, 066804 (2008)
- [73] D. Meyer, A.C. Hewson, and R. Bulla, Phys. Rev. Lett. **89**, 196401 (2002)
- [74] R. Bulla, Phys. Rev. Lett. **83**, 136 (1999)
- [75] T.A. Costi and N. Manini, Journal of Low Temperature Physics **126**, 835 (2002).
- [76] W. Koller, D. Meyer, and A.C. Hewson, Phys. Rev. B **70**, 155103 (2004)
- [77] T. Pruschke and R. Bulla, The European Physical Journal B **44**, 217 (2005)

6 NRG with Bosons

Kevin Ingersent

Department of Physics, University of Florida

P.O. Box 188440, Gainesville, FL 32611-8440, USA

Contents

1	Introduction	2
2	NRG with local bosons	3
2.1	The Anderson-Holstein model	3
2.2	NRG solution method	4
2.3	Results	6
3	Bosonic NRG	9
3.1	The spin-boson model	9
3.2	NRG solution method	10
3.3	Results	12
4	Bose-Fermi NRG	16
4.1	The Bose-Fermi Kondo model	16
4.2	NRG solution method	18
4.3	Results	19
5	Closing	23

1 Introduction

The lecture notes in this Autumn School describe many quantum impurity problems of current interest in connection with the physics of strongly correlated electrons, as well as some of the techniques that have been devised to solve these problems. One such technique that has historically been very influential in the understanding of quantum impurity systems is the numerical renormalization group (NRG) method [1–3]. The NRG remains very important for the study of a variety of topical issues (see, e.g., the lecture notes by T. Costi [4]).

The NRG method was developed to provide a robust account of the low-energy properties of Hamiltonians describing the coupling of a local dynamical degree of freedom (a spin or a localized electronic level) to a gapless band of delocalized electrons. These lecture notes focus on extensions of the NRG to treat quantum impurity problems that involve bosonic degrees of freedom. We will consider three classes of problem of increasing complexity:

1. *Local-bosonic* models in which a localized degree of freedom couples not only to band fermions but also to one or more discrete bosonic modes, each representing perhaps an optical phonon mode or a monochromatic light source. Such models arise, for example, in the description of single-molecule devices in which the molecular charge couples to a localized vibration.
2. *Pure-bosonic* models in which the impurity couples to an environment of dispersive bosonic excitations that acts as a source of decoherence on the impurity degrees of freedom. The canonical example of such a problem is the spin-boson model [5, 6].
3. *Bose-Fermi* models that couple an impurity both to band fermions and to dispersive bosons, the latter representing, e.g., acoustic phonons or some effective magnetic fluctuations. Such models, which describe not only the key physics of certain nanodevices but also form approximate descriptions of heavy-fermion systems, manifest the phenomenon of critical Kondo destruction.

In each of these cases, it proves possible to preserve the essential strategy of the NRG approach of introducing an artificial separation of energy scales that allows the Hamiltonian to be solved iteratively to provide controlled approximations to physical quantities at a sequence of energy scales extending arbitrarily close to zero. However, the presence of bosons imposes greater computational challenges since the Fock space of the problem is of infinite dimension even in the atomic limit where one neglects the energy dispersion of the environmental excitations. As a result, state truncation plays a greater role than in conventional NRG calculations and it proves to be very important to make an appropriate choice of bosonic basis states.

The sections that follow treat in turn the three classes of problem identified above. For each class, a representative model is introduced and physically motivated. The extension of the conventional (pure-fermionic) NRG method to solve this problem is described and illustrative results are presented.

2 NRG with local bosons

2.1 The Anderson-Holstein model

The Anderson-Holstein model has been studied since the 1970s in connection with mixed valence [7–10], negative- U centers in superconductors [11–13], and most recently, single-molecule devices [14–16]. Its Hamiltonian is $H_{\text{AH}} = H_{\text{A}} + H_{\text{LB}}$, where [17]

$$H_{\text{A}} = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \frac{1}{\sqrt{N_k}} \sum_{\mathbf{k}, \sigma} V_{\mathbf{k}} (d_\sigma^\dagger c_{\mathbf{k}\sigma} + \text{H.c.}) \quad (1)$$

describes the standard Anderson impurity model [18] in which d_σ annihilates an electron of spin z component $\sigma = \pm \frac{1}{2}$ (or $\sigma = \uparrow, \downarrow$) and energy ε_d in the impurity level, $n_d = n_{d\uparrow} + n_{d\downarrow}$ (with $n_{d\sigma} = d_\sigma^\dagger d_\sigma$) is the total impurity occupancy, and $U > 0$ is the Coulomb repulsion between two electrons in the impurity level. $V_{\mathbf{k}}$ is the hybridization matrix element between the impurity and a conduction-band state of energy $\varepsilon_{\mathbf{k}}$ annihilated by fermionic operator $c_{\mathbf{k}\sigma}$, while N_k is the number of unit cells in the host metal and, hence, the number of inequivalent $\mathbf{k}$ values. The local boson Hamiltonian term

$$H_{\text{LB}} = \omega_0 b^\dagger b + \lambda(n_d - 1)(b + b^\dagger). \quad (2)$$

describes the linear coupling of the impurity occupancy to the displacement of a local bosonic mode of frequency ω_0 . Without loss of generality, we can take the electron-boson coupling λ to be real and non-negative.

The conduction-band dispersion $\varepsilon_{\mathbf{k}}$ and the hybridization $V_{\mathbf{k}}$ affect the impurity degrees of freedom only through the hybridization function $\Delta(\varepsilon) \equiv (\pi/N_k) \sum_{\mathbf{k}} V_{\mathbf{k}}^2 \delta(\varepsilon - \varepsilon_{\mathbf{k}})$. To focus on universal physics of the model, we assume a featureless form

$$\Delta(\varepsilon) = \Delta \Theta(D - |\varepsilon|), \quad (3)$$

where D is the half-bandwidth and $\Theta(x)$ is the Heaviside step function.

In the case $\Delta = 0$ of an isolated impurity, the impurity occupancy n_d is fixed, and it is possible to eliminate the electron-boson coupling (the second term in H_{LB}) from H_{AH} via the substitution $b \rightarrow \tilde{b} = b + (\lambda/\omega_0)(n_d - 1)$, which maps the Anderson-Holstein model to an Anderson model plus a free boson mode: $H_{\text{AH}} = \tilde{H}_{\text{A}} + \omega_0 \tilde{b}^\dagger \tilde{b}$, where $\tilde{H}_{\text{A}}$ is identical to H_{A} apart from the replacement of the level energy ε_d and the Coulomb repulsion U by

$$\tilde{\varepsilon}_d = \varepsilon_d + \omega_p, \quad \tilde{U} = U - 2\omega_p, \quad \text{where} \quad \omega_p = \lambda^2/\omega_0 \quad (4)$$

is called the *polaron energy* in the context of electron-phonon coupling. The physical content of Eq. (4) is that H_{LB} describes a quantum harmonic oscillator displaced by a constant force proportional to $\lambda(n_d - 1)$. For $n_d \neq 1$, the ground state of the displaced oscillator is a coherent state of energy $-\omega_p$ (relative to the undisplaced ground state for $n_d = 1$) in which the boson occupation $n_b = b^\dagger b$ follows a Poisson distribution $P(n_b) = \exp(-\bar{n}_b) (\bar{n}_b)^{n_b}/n_b!$ with mean $\bar{n}_b = (\lambda/\omega_0)^2$. This lowering of the ground-state energy can be captured by the effective

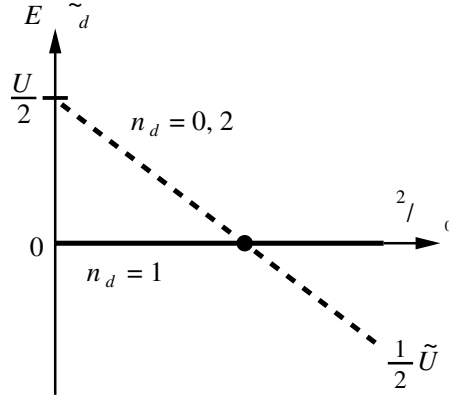


Fig. 1: Evolution with polaron energy $\omega_p = \lambda^2/\omega_0$ of $E - \tilde{\epsilon}_d$, where E is the lowest energy in each n_d sector of the Anderson-Holstein model for $\varepsilon_d = -\frac{1}{2}U$ (particle-hole symmetry) and $\Delta = 0$ (no hybridization), while $\tilde{\epsilon}_d$ is the energy of the lowest $n_d = 1$ spin doublet. The gap $\frac{1}{2}\tilde{U}$ to the lowest energy in the sectors $n_d = 0, 2$ vanishes at $\omega_p = \frac{1}{2}U$. For $\omega_p > \frac{1}{2}U$, the system has a charge-doublet ground state. Adapted from [19].

renormalization of Anderson model parameters according to Eq. (4). As shown in Fig. 1, for $\omega_p > U/2$ the effective Coulomb interaction on the impurity site becomes negative, and for the special case $\varepsilon_d = -\frac{1}{2}U$ of exact particle-hole symmetry the ground state passes from a spin doublet to a charge doublet.

When the impurity-band hybridization is switched on, the effect of the electron-boson coupling remains fully captured by Eq. (4) only if ω_0 is so large that the boson distribution adjusts essentially instantaneously each time that n_d changes. More generally, though, each hybridization event causes the emission and absorption of a cloud of bosons that relaxes with a characteristic time scale ω_0^{-1} toward the distribution favored by the new impurity occupancy [9]. If ω_0^{-1} is comparable with or longer than the characteristic time scale for impurity-band tunneling, the relaxation is incomplete by the time the next hybridization event unleashes another boson cloud. This creates inertia in the system that manifests itself as a reduction in the effective hybridization width Δ . The resulting interplay of impurity charge fluctuations, strong electron-electron correlations, and electron-boson coupling can be treated analytically only in certain limiting cases [9, 14]. In order to obtain a nonperturbative account of the physics over the full parameter space, an NRG treatment of the Anderson-Holstein model is very desirable.

2.2 NRG solution method

As described in greater detail in the other lecture notes [4], there are three essential steps in the NRG treatment of a pure-fermionic problem such as that described by H_A :

1. Division of the band energies $-D \leq \varepsilon < D$ into logarithmic bins spanning $D\Lambda^{-(m+1)} \leq \pm\varepsilon < D\Lambda^{-m}$ for $\Lambda > 1$ and $m = 0, 1, 2, \dots$. Within each bin, the continuum of band states is replaced by a discrete state, namely, the linear combination [weighted according to the hybridization function $\Delta(\varepsilon)$] that interacts with the impurity. The states from adjacent bins have average energies that differ by a factor of Λ .

2. The Lanczos method [20] is applied to perform a canonical transformation on the discrete bin states, mapping the conduction band onto a semi-infinite tight-binding chain that couples to the impurity only at the end site $n = 0$:

$$H_{\text{band}} = \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} \simeq D \sum_{n=0}^{\infty} \sum_{\sigma} t_{n+1} (f_{n\sigma}^\dagger f_{n+1,\sigma} + \text{H.c.}), \quad (5)$$

where the chain-site creation and annihilation operators obey $\{f_{n\sigma}, f_{n'\sigma'}^\dagger\} = \delta_{n,n'} \delta_{\sigma,\sigma'}$ and the dimensionless hopping coefficients drop off as $t_n \simeq \Lambda^{-n/2}$ due to the separation of energy scales in the discretized band.

3. Iterative diagonalization of scaled Hamiltonians H_N on chains truncated at length $N + 1$, starting with (for the Anderson impurity model)

$$H_0 = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + \sqrt{\frac{2\Delta D}{\pi}} \sum_{\sigma} (d_{\sigma}^\dagger f_{0\sigma} + \text{H.c.}). \quad (6)$$

The basis of H_N has dimension $d_N = 4^{N+2}$, requiring storage $\propto d_N^2$ and a solution time $\propto d_N^3$. It therefore becomes necessary after only a few iterations to truncate the basis. In most cases, one retains at the end of iteration N just the N_s states of lowest energy (with N_s typically lying in the range 100 to 1 000) so that the basis of H_{N+1} has a reduced dimension $d_{N+1} = 4N_s$. Under this procedure, the low-lying many-body eigenstates of H_N (a) describe the essential physics on energy and temperature scales of order $D\Lambda^{-N/2}$, and (b) provide a good starting point for finding the low-lying eigenstates of $H_{N+1} = \Lambda^{1/2}(H_N - E_N^{(0)}) + D t_{N+1} \Lambda^{(N+1)/2} \sum_{\sigma} (f_{N\sigma}^\dagger f_{N+1,\sigma} + \text{H.c.})$, where $E_N^{(0)}$ is the ground-state energy of iteration N . The rescaling of H_{N+1} by a multiplicative factor of $\sqrt{\Lambda}$ relative to H_N facilitates the identification of renormalization-group fixed-points characterized by self-similar many-body spectra [1, 2].

Going from H_A to H_{AH} does not require any modification of steps 1 and 2 above. At step 3, however, H_{LB} in Eq. (2) must be added into H_0 in Eq. (6). For the Anderson model, the Fock space of iteration 0 has dimension 4 (for the impurity) $\times$ 4 (for chain site 0) = 16, which makes numerical diagonalization of H_0 a trivial matter. The inclusion of bosonic degrees of freedom that are not limited by the Pauli exclusion principle immediately has the effect of raising the Fock-space dimension to infinity. Since diagonalization of infinite matrices is computationally infeasible, one is forced to introduce the additional approximation (relative to the pure-fermionic NRG) of truncating the basis of H_0 .

The low-lying many-body states of H_0 should be superpositions of configurations in which n_d takes each of its possible values. We therefore expect to have to be able to capture both (a) configurations with low values of n_b that are energetically favorable for $n_d = 1$ and (b) configurations with boson distributions close to the coherent states favored for $n_d = 0$ and 2. Given the rapid fall-off of the coherent-state boson occupation distribution $P(n_b) = \exp(-\bar{n}_b) (\bar{n}_b)^{n_b} / n_b!$ for $n_b \gg \bar{n}_b = (\lambda/\omega_0)^2$, one can hope to work with a bosonic basis consisting of all occupation number eigenstates with $0 \leq n_b < N_b$. Hewson and Meyer [9] established a criterion

$N_b \geq 4\bar{n}_b$. However, in view of the standard deviation $\sigma_b = \sqrt{\bar{n}_b}$ of the Poisson distribution it seems probable that for large $\bar{n}_b$ it would be more efficient to select $N_b = \bar{n}_b + c\sigma_b$ with c being a fixed number of order 5. Better still might be a basis that directly includes the ground state and low-lying excitations of the displaced oscillators, i.e., eigenstates of well-defined and small $\tilde{b}^\dagger \tilde{b}$. However, exploration of such an option has been rendered unnecessary by the success of the simpler basis $0 \leq n_b < N_b$. This basis increases the CPU time for iteration 0, which is proportional to $(16N_b)^3$, but leaves unaffected the generally much greater CPU time $\propto (4N_s)^3$ for high iteration numbers.

2.3 Results

We begin by considering results for the symmetric Anderson-Holstein model ($\varepsilon_d = -\frac{1}{2}U$) where the impurity-boson subsystem has a level-crossing from a magnetic ground state (for $\lambda < \lambda_c$) to a non-magnetic charge-doublet ground state (for $\lambda > \lambda_c$). Figure 2 plots three temperature scales extracted from thermodynamic properties calculated using the NRG:

- $T_s = 0.103/\chi_s(T = 0)$, where $\chi_s(T)$ is the impurity contribution to the system's static spin susceptibility, i.e., the difference between the spin susceptibility $(\langle S_z^2 \rangle - \langle S_z \rangle^2)/T$ (S_z being the z component of the system's total spin) with and without the impurity. The impurity spin degree of freedom is quenched for $T \lesssim T_s$, and in the Kondo regime $0 < \Delta \ll -\varepsilon_d, U + \varepsilon_d$ of the Anderson model [2], T_s coincides with the Kondo temperature.
- $T_c = 0.412/\chi_c(T = 0)$, where $\chi_c(T)$ is the impurity contribution to the static charge susceptibility $(\langle Q^2 \rangle - \langle Q \rangle^2)/T$ with Q being the total electron number measured from half-filling. The impurity charge is quenched for $T \lesssim T_c$, and in the regime $0 < \Delta \ll \varepsilon_d, -(U + \varepsilon_d)$ of the negative- U Anderson model, T_c is the Kondo temperature characterizing a many-body screening of the impurity charge directly analogous to the standard (spin) Kondo effect [21, 22]. The different coefficients in the definitions of T_s and T_c reflect the values $\chi_s = 1/(4T)$ for a free spin doublet and $\chi_c = 1/T$ for a free charge doublet.
- T_K defined via the impurity contribution to the entropy via the condition $S_{\text{imp}}(T_K) = 0.383$. T_K can be regarded as the crossover temperature for the suppression of all impurity degrees of freedom and coincides with the relevant Kondo temperature in the spin-Kondo and charge-Kondo regimes of the Anderson model.

Figure 2 provides evidence for a smooth crossover with increasing electron-boson coupling from a spin-Kondo effect to a charge-Kondo effect. As ω_p increases from zero, T_s rises rapidly as the impurity loses its local-moment character and the system crosses from the strongly correlated spin-Kondo regime to the weakly correlated mixed-valence regime as $\tilde{U}$ falls toward zero from its initial value of U . At the same time, T_c decreases from very large values at $\omega_p = 0$ and becomes exponentially small in the charge-Kondo regime $\omega_p \gg \frac{1}{2}U = 0.1D$ where $\tilde{U}$ is large and negative. Meanwhile, T_K evolves from following T_s deep in the spin-Kondo regime

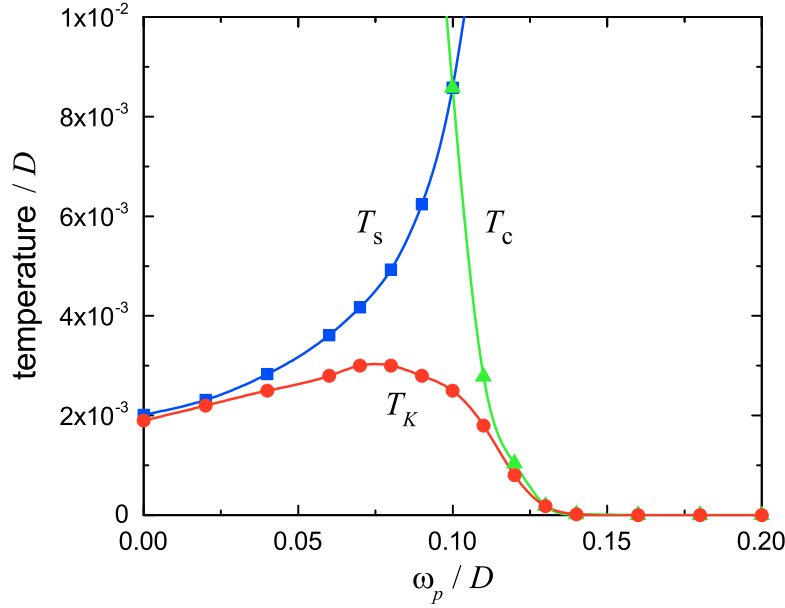


Fig. 2: Variation with the polaron energy $\omega_p = \lambda^2/\omega_0$ of three energy scales for the particle-hole-symmetric Anderson-Holstein model with $U = -2\varepsilon_d = 0.2D$, $\Delta = 0.032D$, and $\omega_0 = 0.05D$: the spin-screening temperature T_s , the charge-screening temperature T_c , and the Kondo temperature T_K extracted from the impurity entropy via the condition $S_{\text{imp}}(T_K) = 0.383$. NRG results obtained for $\Lambda = 2.5$ with bosonic cutoff $N_b = 16$, retaining up to 2000 many-body states (spin multiplets) after each iteration.

to tracking T_c deep in the charge-Kondo regime. In the intermediate region near $\omega_p = \frac{1}{2}U$, T_K is much smaller than either T_s or T_c , pointing to a many-body Kondo effect of mixed spin and charge character.

The plot of T_K vs. ω_p in Fig. 2 is clearly asymmetric about its peak near $\omega_p = \frac{1}{2}U$. This asymmetry can be seen more clearly in Fig. 3, which plots an effective Kondo energy scale (in this case, extracted from the impurity spectral function) vs λ/ω_0 . For $\omega_p \ll \frac{1}{2}U$ (which for the parameters used in Fig. 3 means $\lambda/\omega_0 \ll 1$), $\Gamma(\lambda)$ is captured by a perturbative mapping [14] onto the Kondo model that incorporates effects beyond the replacement of ε_d and U by $\tilde{\varepsilon}_d$ and $\tilde{U}$ given in Eq. (4). Upon increase of ω_p beyond $\frac{1}{2}U$, Γ drops off extremely fast (note the logarithmic vertical axis in Fig. 3) in a manner that is described quite well by a perturbative mapping to an *anisotropic* charge-Kondo model in which the rate $J_\perp$ of charge-flip impurity-band scattering ($n_d = 0 \rightarrow 2$ and its time reverse, both via an $n_d = 1$ virtual state) is smaller than the rate $J_\parallel$ of charge-conserving scattering ($n_d = 0 \rightarrow 0$ and $n_d = 2 \rightarrow 2$, also both via $n_d = 1$) by a factor $J_\perp/J_\parallel \simeq \exp(-2\lambda^2/\omega_0^2)$. This confirms the extremely strong suppression of real (non-virtual) charge fluctuations caused by the small overlap between the bosonic ground states in each sector of different n_d . In the vicinity of $\omega_p = \frac{1}{2}U$, neither perturbative approach is satisfactory, and one must rely on the full machinery of the NRG to provide reliable results.

Finally, in this section we consider the effect of moving away from particle-hole symmetry of the impurity level. For $\varepsilon_d \neq -\frac{1}{2}U$, the $n_d = 2$ curve in Fig. 1 is raised above the solid line (which now represents just $n_d = 0$) by an amount $U + 2\varepsilon_d$ independent of the electron-boson

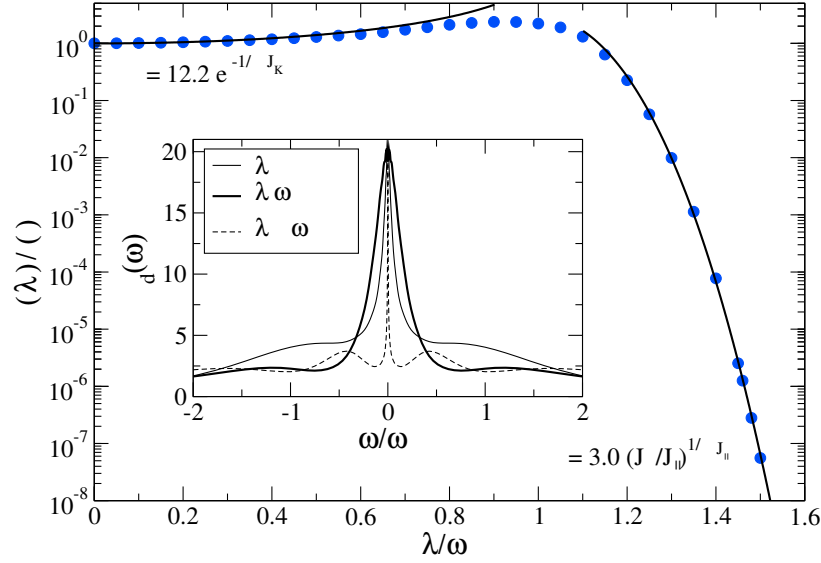


Fig. 3: Variation with λ/ω_0 of an effective Kondo energy scale Γ for the particle-hole-symmetric Anderson-Holstein model with $U = -2\varepsilon_d = 0.1D$, $\Delta = 0.012D$, and $\omega_0 = 0.05D$. Γ is the width of the Abrikosov-Suhl resonance in the impurity spectral function $\rho_d(\omega)$ (see inset). Circles represent NRG data while solid lines are the results of analytical approximations. Reprinted figure with permission from P.S. Cornaglia, H. Ness, and D.R. Grempel, *Phys. Rev. Lett.* **93**, 147201 (2004). Copyright 2004 by the American Physical Society.

coupling. This shift splits the impurity charge doublet in a manner analogous to the action of a magnetic field on the $n_d = 1$ spin doublet, and in the charge-Kondo regime an impurity asymmetry $|U + 2\varepsilon_d| \gg T_c$ suppresses the Kondo effect.

Fig. 4 shows the linear conductance G through a nanodevice in which a single molecule bridges the gap between two electrical leads. The transport is assumed to be dominated by a single, strongly correlated molecular level of energy ε_d (which may be tuned via a voltage applied to an electrical gate) and whose charge couples to a local vibrational mode in a manner described by the Anderson-Holstein model. In such a device [14],

$$G = \frac{2e^2}{h} \pi \Delta \int_{-\infty}^{\infty} d\omega \left(-\frac{\partial f}{\partial \omega} \right) \rho_d(\omega), \quad (7)$$

where $\rho_d(\omega) = -\pi^{-1} \text{Im } G_{dd}(\omega)$, with G_{dd} being the retarded Green's function for the active impurity level. For $\lambda = 0$, Fig. 4a, the conductance at a comparatively high temperature $T = \Delta$ exhibits the phenomenon of Coulomb blockade, where the strong interactions in the molecular level suppress conductance via sequential tunneling of electrons from the source electrode into the molecule and then off into the drain electrode, except near the point $\varepsilon_d = 0$ (or $\varepsilon_d = -U$) of degeneracy between states of occupancy $n_d = 0$ and 1 (or $n_d = 1$ and 2). At $T = 0$, however, G is nonzero due to the formation of the collective Kondo ground state that allows electrons to pass from one lead to another without incurring an energy penalty U . For $\lambda = 0.4\omega_0$, Fig. 4b, the physics is similar, except the spacing between the Coulomb blockade peaks has diminished from U to roughly $\tilde{U}$. For still larger values of λ such that $\tilde{U} < 0$, Fig. 4d, the high-temperature

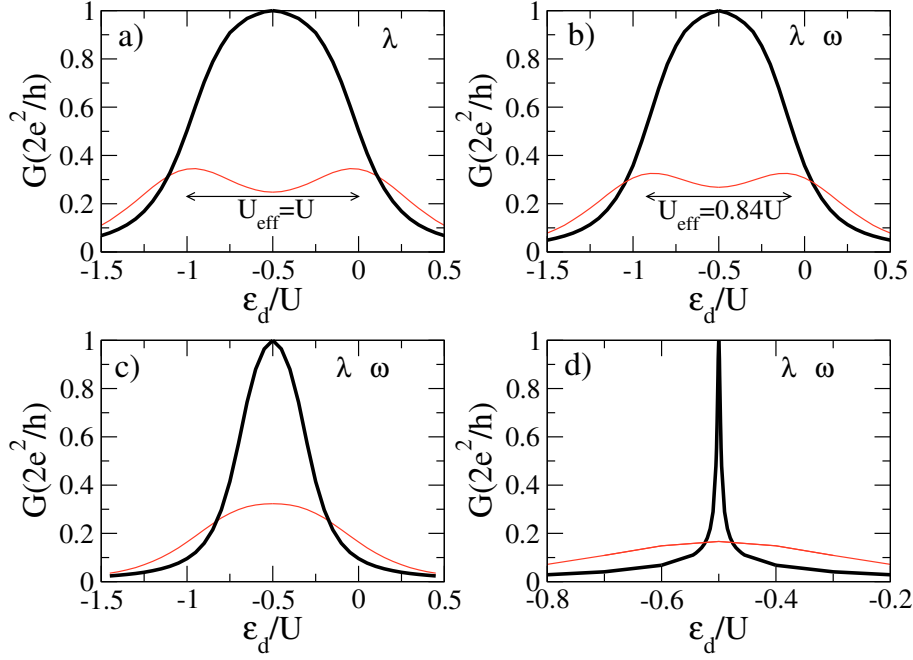


Fig. 4: Linear conductance G vs. level energy ε_d for a single-molecule device described by the Anderson-Holstein model with $U = 0.1$, $\Delta = 0.016$, $\omega = 0.05$, and different values of λ . Thin (thick) lines show NRG results for temperature $T = \Delta$ ($T = 0$). Reprinted figure with permission from P.S. Cornaglia, H. Ness, and D.R. Grempel, *Phys. Rev. Lett.* **93**, 147201 (2004). Copyright 2004 by the American Physical Society.

conductance is suppressed for all values of ε_d because there is no point of degeneracy between ground states differing by 1 in their charge. A charge-Kondo peak remains centered at particle-hole symmetry, but it is very narrow since the many-body Kondo state is essentially destroyed once the charge-doublet splitting $|U + 2\varepsilon_d|$ exceeds the Kondo scale.

3 Bosonic NRG

3.1 The spin-boson model

The spin-boson model for a dissipative two-state system has been heavily studied in many contexts [5, 6], including chemical reactions, motion of defects in solids, biological molecules, and quantum information. Its Hamiltonian can be written [17] $H_{\text{SB}} = -\Delta S_x - h S_z + H_{\text{B}}$, where

$$H_{\text{B}} = \sum_q \omega_q a_q^\dagger a_q + \frac{S_z}{\sqrt{N_q}} \sum_q \lambda_q (a_q + a_q^\dagger). \quad (8)$$

Here, S_x and S_z are the x and z components of the spin (or pseudospin) of a two-state impurity system and a_q annihilates a boson of energy ω_q [17]. Δ is the matrix element for tunneling between states $|\uparrow\rangle$ ($S_z = \frac{1}{2}$) and $|\downarrow\rangle$ ($S_z = -\frac{1}{2}$), h is a (pseudo)magnetic field that couples to the z component of the local spin, N_q is the number of boson modes, and λ_q is a linear coupling between the displacement of mode q and the local spin z . The values of ω_q and λ_q

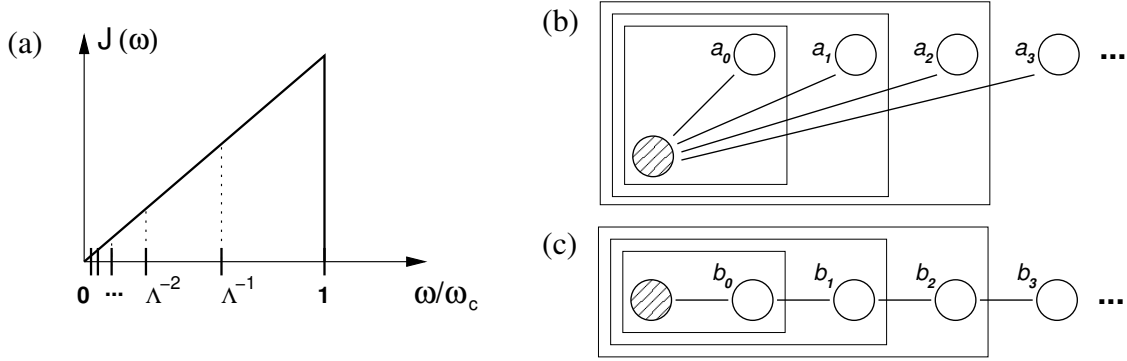


Fig. 5: NRG treatment of a bosonic bath: (a) The bath spectral function is divided into logarithmic bins. (b) The impurity (shaded circle) interacts with one representative state (open circle) from each logarithmic bin $m = 0, 1, 2, \dots$ in a “star” Hamiltonian form. (c) The Lanczos procedure maps the Hamiltonian to a “chain” form where the impurity interacts with just the end site $n = 0$. In (b) and (c), boxes from innermost to outermost enclose the degrees of freedom treated at NRG iterations $N = 0, 1$, and 2 . Reprinted figures with permission from R. Bulla, H.-J. Lee, N.H. Tong, and M. Vojta, *Phys. Rev. B* **71**, 045122 (2005). Copyright 2005 by the American Physical Society.

enter the problem only in a single combination, the bosonic bath spectral function $J(\omega) = (\pi/N_q) \sum_q \lambda_q^2 \delta(\omega - \omega_q)$, which in the thermodynamic limit $N_q \rightarrow \infty$ generally becomes a smooth function. The most important feature of $J(\omega)$ is its asymptotic low-frequency behavior, so it is conventional to study power-law spectra

$$J(\omega) = 2\pi\alpha\omega_c(\omega/\omega_c)^s\Theta(\omega)\Theta(\omega_c - \omega), \quad (9)$$

where α is a dimensionless dissipation strength, ω_c is a high-frequency cut-off and the bath exponent must satisfy $s > -1$ to allow normalization. The most subtle physics occurs for bath exponents $0 < s \leq 1$, which admit two distinct phases distinguished by an order parameter $\phi = \lim_{h \rightarrow 0^+} \langle S_z \rangle$. In the *delocalized* phase ($\alpha < \alpha_c$), the effective value of α renormalizes to zero, leading to a singlet ground state and $\phi = 0$. In the *localized* phase ($\alpha > \alpha_c$), Δ renormalizes to zero, asymptotically confining the impurity to one or other of its two states and yielding (at least for bias field $h = 0$) a doublet ground state with $\phi > 0$. In the heavily studied case $s = 1$ of an ohmic bath, the quantum phase transition occurs at $\alpha_c = 1 + O(\Delta/\omega_c)$ and is known to be Kosterlitz-Thouless-like [5], involving a jump in the order parameter but a correlation length that diverges on approach from the delocalized side. For sub-ohmic bath exponents $0 < s < 1$, the transition takes place at $\alpha_c \propto (\Delta/\omega_c)^{1-s}$ and is believed to be continuous [23].

3.2 NRG solution method

For $s = 1$, the spin-boson model may be mapped to the anisotropic Kondo model [24] and thus can be treated using the conventional NRG method [25]. However, no such mapping exists for general values of s . The direct NRG treatment of the spin-boson model was pioneered by Bulla

et al. [26, 27]. One can follow the same three essential steps found in the conventional NRG (see Sec. 2.2). After logarithmic binning of the bath, Fig. 5a, the impurity interacts with one representative state from each bin, destroyed by an operator a_m , see Fig. 5b, allowing the bath part of H_{SB} to be written

$$H_{\text{bath}} = \sum_q \omega_q a_q^\dagger a_q \simeq \omega_c \sum_{m=0}^{\infty} \xi_m a_m^\dagger a_m, \quad (10)$$

where the operators a_m obey canonical bosonic commutation relations $[a_m, a_{m'}^\dagger] = \delta_{m,m'}$ and have dimensionless oscillator energies

$$\xi_m = \int_{\omega_c \Lambda^{-(m+1)}}^{\omega_c \Lambda^{-m}} \omega J(\omega) d\omega \Big/ \int_{\omega_c \Lambda^{-(m+1)}}^{\omega_c \Lambda^{-m}} J(\omega) d\omega = \frac{1+s}{2+s} \frac{1 - \Lambda^{-(2+s)}}{1 - \Lambda^{-(1+s)}} \Lambda^{-m}. \quad (11)$$

Application of the Lanczos procedure converts this “star form” of the bath Hamiltonian to a tight-binding “chain form”, see Fig. 5c

$$H_{\text{bath}} \simeq \omega_c \sum_{n=0}^{\infty} \left[\varepsilon_n b_n^\dagger b_n + \tau_{n+1} (b_n^\dagger b_{n+1} + \text{H.c.}) \right], \quad (12)$$

where $[b_n, b_{n'}^\dagger] = \delta_{n,n'}$. The fact that $J(\omega) = 0$ for $\omega < 0$ causes the dimensionless on-site energies ε_n and hopping coefficients τ_n to take values of order Λ^{-n} , dropping off with increasing n at a rate twice as fast as the parameters $t_n \approx \Lambda^{-n/2}$ in the fermionic NRG.

Bulla *et al.* [26, 27] constructed two different iterative NRG procedures for the spin-boson model, one based on the star form of H_{bath} and the other based on the chain form:

1. The star-based NRG procedure is illustrated schematically in Fig. 5(b). It starts from an initial Hamiltonian

$$H_0 = -\Delta S_x - h S_z + \omega_c \left[\xi_0 a_0^\dagger a_0 + \sqrt{\frac{2\alpha}{1+s}} S_z \gamma_0 (a_0 + a_0^\dagger) \right] \quad (13)$$

that includes only the bosonic operator a_0 representing the highest-energy logarithmic bin and proceeds to incorporate one more bin at each subsequent iteration according to

$$H_{N+1} = \Lambda (H_N - E_N^{(0)}) + \Lambda^{N+1} \omega_c \left[\xi_{N+1} a_{N+1}^\dagger a_{N+1} + \sqrt{\frac{2\alpha}{1+s}} S_z \gamma_{N+1} (a_{N+1} + a_{N+1}^\dagger) \right]. \quad (14)$$

In Eqs. (13) and (14), γ_m is a positive normalization constant satisfying

$$\gamma_m^2 = \frac{1+s}{2\pi\alpha} \int_{\omega_c \Lambda^{-(m+1)}}^{\omega_c \Lambda^{-m}} d\omega J(\omega) = [1 - \Lambda^{-(1+s)}] \Lambda^{-(1+s)m}. \quad (15)$$

Each operator a_m couples only to the impurity, allowing the bosonic basis to be optimized (at least for $\Delta = 0$, where the impurity becomes static) by transforming to displaced oscillators with annihilation operators $\tilde{a}_m = a_m \pm \theta_m$, where $\theta_m = \sqrt{\alpha/2(1+s)} \gamma_m / \xi_m \sim$

$\Lambda^{(1-s)m/2}$. Since the ground state of the displaced oscillator corresponds to $\langle b_m^\dagger b_m \rangle = \theta_m^2 \sim \Lambda^{(1-s)m}$, a basis of eigenstates of $b_m^\dagger b_m$ restricted to n_{b_m} less than some finite N_b will prove inadequate for capturing the low-energy behavior for any sub-ohmic case $s < 1$. It will be shown in Sec. 3.3 that the same conclusion holds throughout the localized phase of the full sub-ohmic spin-boson model with $\Delta > 0$, but that success can be achieved using a suitably chosen basis of N_b displaced oscillator states optimized for the value of θ_m .

2. The chain-based NRG procedure, which is illustrated schematically in Fig. 5c, starts from an initial Hamiltonian

$$H_0 = -\Delta S_x - h S_z + \omega_c \left[\varepsilon_0 b_0^\dagger b_0 + \sqrt{\frac{2\alpha}{1+s}} S_z (b_0 + b_0^\dagger) \right] \quad (16)$$

where $b_0 = \sum_{m=0}^{\infty} \gamma_m a_m$ and $\varepsilon_0 = (1+s)/(2+s)$. The iteration relation is

$$H_{N+1} = \Lambda (H_N - E_N^{(0)}) + \Lambda^{N+1} \left[\varepsilon_{N+1} b_{N+1}^\dagger b_{N+1} + \tau_{N+1} (b_N^\dagger b_{N+1} + \text{H.c.}) \right], \quad (17)$$

where the Λ^{-N} decay of the tight-binding coefficients ε_N and τ_N dictates a rescaling of H_{N+1} by a factor of Λ (instead of $\sqrt{\Lambda}$ as in the fermionic NRG).

For $\Delta = h = 0$, H_0 describes a displaced harmonic oscillator having a ground state in which the occupation number $n_{b_0} \equiv \langle b_0^\dagger b_0 \rangle$ has a mean value $\bar{n}_{b_0} = \alpha/[2(1+s)\varepsilon_0^2] = \alpha(2+s)^2/[2(1+s)^3]$. The α values of greatest interest are those near α_c , of order 1 or smaller. Therefore, just as in the NRG treatment of the Anderson-Holstein model [Sec. 2.2], it should be satisfactory to use a basis of bosonic number eigenstates with $n_{b_0} < N_b$, where $N_b \geq 4\bar{n}_{b_0}$. However, what is unclear *a priori* is whether a bosonic truncation $n_{b,N+1} < N_b$ will prove satisfactory during subsequent iterations of Eq. (17).

The star and the chain NRG formulations will be seen in Sec. 3.3 to have different strengths and weaknesses. In both cases, a key challenge is to find a bosonic basis size N_b for each site sufficiently large that the physical results are a good approximation to those for $N_b \rightarrow \infty$ while keeping the computational time ($\propto B_b^3$) within acceptable bounds.

3.3 Results

3.3.1 Truncation effects in the chain and star formulations

Bulla *et al.* systematically investigated the effects of basis truncation in the star and chain versions of the bosonic NRG [27]. Figure 6a shows results for the chain NRG, which was the first of the two to be implemented [26]. Throughout the delocalized phase and on the phase boundary, $\langle b_{N+1}^\dagger b_{N+1} \rangle$ for a large fixed N (the figure illustrates $N = 20$ for $s = 0.6$ and $\Delta = 0.01\omega_c$) converges rapidly with increasing N_b to a value much smaller than N_b . For $\alpha > \alpha_c$, however, $\langle b_{N+1}^\dagger b_{N+1} \rangle$ continues to grow with N_b . Although the average boson occupancy saturates for sufficiently large values of N_b , the size of the basis required to eliminate truncation effects

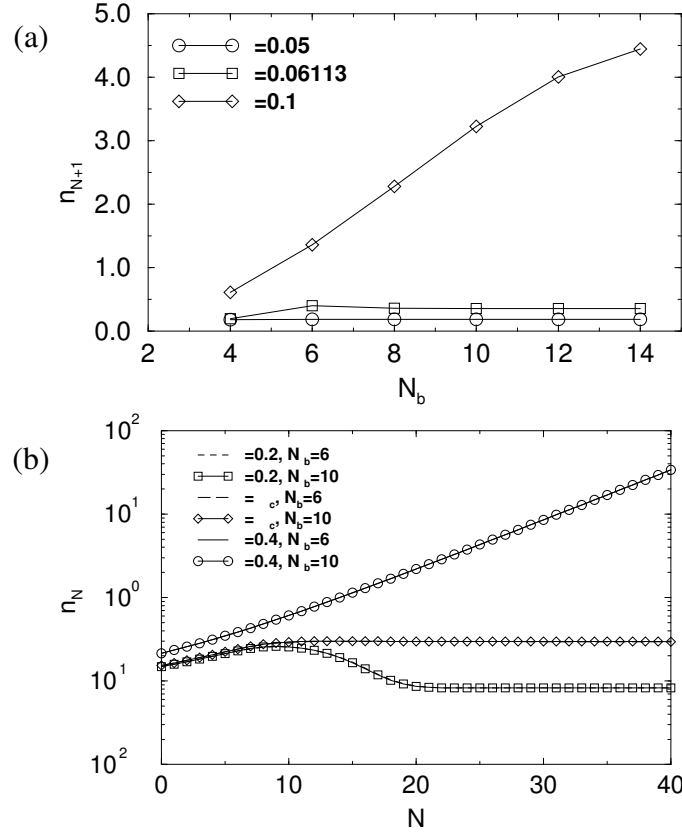


Fig. 6: Effects of bosonic basis truncation in the NRG treatment of the sub-ohmic spin-boson model: (a) Average occupancy of bosonic chain site 21 vs. bosonic truncation parameter N_b within the chain NRG. Data for $s = 0.6$, $\Delta = 0.01\omega_c$, $\Lambda = 2$, and for α below, at, and above $\alpha_c = 0.06113$. (b) Average occupancy $n_N = \langle b_N^\dagger b_N \rangle$ vs. bosonic bin index N within the star NRG. Data for $s = 0.8$, $\Delta = 0.01\omega_c$, and $\Lambda = 2$, obtained for α below, at, and above α_c using an optimized displaced oscillator basis of $N_b = 6$ (lines) or $N_b = 10$ (symbols) states per bin. Reprinted figures with permission from R. Bulla, H.-J. Lee, N.H. Tong, and M. Vojta, *Phys. Rev. B* **71**, 045122 (2005). Copyright 2005 by the American Physical Society.

grows with both N and α . The implication is that the chain NRG cannot access the asymptotic low-energy physics in the localized phase of the sub-ohmic spin-boson model. No such problem affects the localized phase of the ohmic case $s = 1$, or super-ohmic ($s > 1$) baths where the ground state is delocalized for any $\Delta \neq 0$.

Figure 6b illustrates results obtained using the star NRG. Since the oscillator shift θ_N is known analytically only for $\Delta = 0$, these calculations used a basis of N_b orthogonalized oscillator states chosen by minimizing the ground-state energy over multiple trial values of θ_N [27]. The figure demonstrates (for $s = 0.8$) that in both phases and on the phase boundary, $\langle b_N^\dagger b_N \rangle$ shows negligible difference for $N_b = 6$ and $N_b = 10$ and that the optimized basis seems to provide robust values for the boson occupancies, even in the localized phase where $\langle b_N^\dagger b_N \rangle$ is diverging. Despite this promising behavior, the star NRG proves to be unreliable in the delocalized phase and on the phase boundary, where its results are inconsistent with those obtained by other methods including chain NRG [27]. The reasons for this failure are not fully understood.

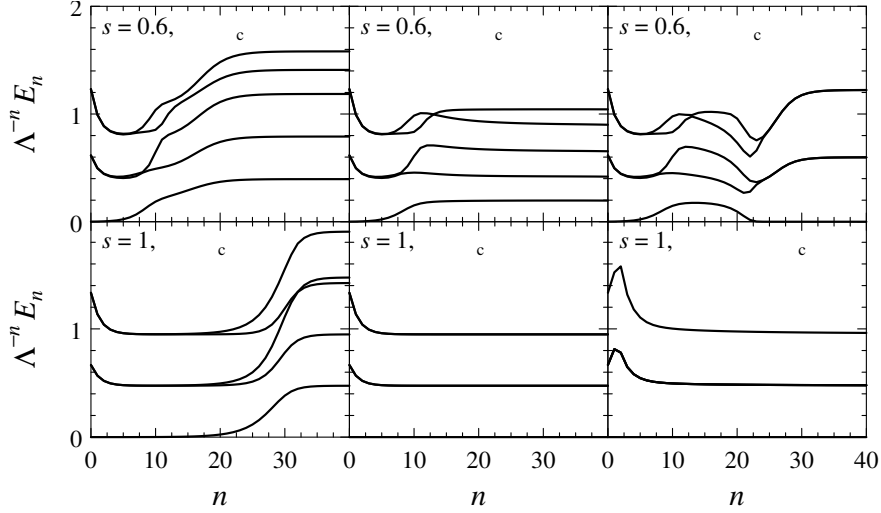


Fig. 7: NRG energies E_N vs. iteration number N for low-lying many-body eigenstates in the spin-boson model with $s = 0.6$ (top panels) and $s = 1$ (bottom panels) at dissipation strengths $\alpha < \alpha_c$ (left), $\alpha = \alpha_c$ (middle), and $\alpha > \alpha_c$ (right). Reprinted figure with permission from R. Bulla, N.-H. Tong, and M. Vojta, *Phys. Rev. Lett.* **91**, 170601 (2003). Copyright 2003 by the American Physical Society.

The conclusion from [27] is that neither the star NRG nor the chain NRG can provide a fully reliable treatment of all cases. Although the star NRG has been preferred in a study of a related model with ohmic dissipation [28], studies of the sub-ohmic spin-boson model have focused overwhelmingly on the chain approach. The next subsection will highlight a few successes and failures of the method.

3.3.2 Chain NRG results for the spin-boson model

Figure 7 illustrates for $s = 0.6$ and for $s = 1$ one of the primary outputs of the bosonic NRG method: the evolution of the low-lying many-body spectrum with iteration number N . When combined with the matrix elements of appropriate operators between the many-body eigenstates, the spectrum can yield information on static and dynamic quantities of interest, such as the static magnetization and the dynamical susceptibility of the impurity spin. Dynamical properties are not discussed in these lecture notes for reasons of space. The left panels in Fig. 7 exemplify the delocalized phase, where a rapid initial change in the energies E_N over the first few iterations is followed by one or two intermediate plateaus before final approach to a delocalized fixed-point spectrum. This spectrum, which is identical for all values of Δ and $\alpha < \alpha_c(\Delta)$, is just that of free bosons described by H_{bath} given in Eq. (12), reflecting the renormalization of the dissipative coupling α to zero throughout the delocalized phase.

Upon fine tuning of α extremely close to its critical value, as shown in the center panels of Fig. 7, an intermediate plateau (quite well developed in the left panel for $s = 1$, but barely visible in the $s = 0.6$ example) stretches beyond $N = 40$. For $s < 1$, this plateau spectrum is entirely distinct from that at the delocalized fixed point but is the same for all combinations $\Delta \neq 0$, $\alpha = \alpha_c(\Delta)$; it characterizes a *critical fixed point* located at nonzero critical couplings

$\Delta = \Delta^*(s)$, $\alpha = \alpha^*(s)$. In the localized phase (upper right panel of Fig. 7), the spectrum should in principle converge to a free-boson spectrum for a set of displaced oscillators, leading to a set of energies identical to those at the delocalized fixed point. However, due to bosonic truncation effects, the chain NRG instead yields a different, artificial fixed-point spectrum.

For $s = 1$, by contrast, the quantum phase transition at $\alpha = \alpha_c(\Delta)$ is governed by a *critical end point* at $\Delta^* = 0$, $\alpha^* = 1$, which terminates a line of localized fixed points $\Delta^* = 0$, $\alpha^* \geq 1$ [5]. For points (Δ, α) not too deep into the localized phase, α^* is sufficiently small that the chain NRG can faithfully reproduce the appropriate displaced oscillator ground state and its excitations, so the many-body spectrum for $N \rightarrow \infty$ is the same for all values of α (see the lower panels of Fig. 7).

Most NRG studies of the sub-ohmic spin-boson model have focused on the critical properties, which are very hard to access using algebraic methods. A particular focus has been the evaluation of critical exponents such as β and δ entering the variations

$$\phi(\alpha > \alpha_c, T = h = 0) \propto (\alpha - \alpha_c)^\beta \quad \text{and} \quad \phi(\alpha = \alpha_c, T = 0) \propto |h|^{1/\delta} \quad (18)$$

of the order parameter $\phi = \lim_{h \rightarrow 0^+} \langle S_z \rangle$, the exponents γ and x entering the variations

$$\chi(\alpha < \alpha_c, T = h = 0) \propto (\alpha_c - \alpha)^{-\gamma} \quad \text{and} \quad \chi(\alpha = \alpha_c, h = 0) \propto T^{-x} \quad (19)$$

of the order-parameter susceptibility $\chi = \partial\phi/\partial h|_{h=0}$, and the correlation-length exponent ν characterizing the vanishing according to

$$T^* \propto |\alpha - \alpha_c|^\nu \quad (20)$$

of the energy scale (extracted from data such as those shown in Fig. 7) at which the many-body spectrum flows away from the critical spectrum to that of either the delocalized or the localized fixed point. The chain NRG allows all these exponents to be determined to an unprecedented degree of numerical precision [26, 29]. Although they vary with s , the exponents are found to obey to within estimated numerical errors the *scaling relations*

$$\delta = (1 + x)/(1 - x), \quad \beta = \gamma(1 - x)/(2x), \quad \text{and} \quad \nu = \gamma/x \quad (21)$$

expected [30] to hold at an interacting quantum critical point in a magnetic impurity model below its upper critical dimension.

For $\frac{1}{2} < s < 1$, the observed scaling of exponents is consistent with expectations based on a mapping (carried out within a path-integral formulation [23]) of the spin-boson model onto a classical model for a chain of Ising spins with a long-range ferromagnetic interaction that decays with separation d like $d^{-(1+s)}$. The Ising model has a phase transition that is interacting for cases corresponding to $\frac{1}{2} < s < 1$ [31, 32]. Within this range, the NRG is fully consistent with the mapped classical problem, and the values of the exponents it produces agree with analytical limiting results where they are available. There is every indication that the chain-form bosonic NRG is yielding correct results over this range of weakly sub-ohmic bath exponents.

In contrast, for more slowly decaying interactions, corresponding in the spin-boson model to $0 < s < \frac{1}{2}$, the quantum critical point of the classical Ising chain is noninteracting and is characterized by mean-field exponents corresponding to $\beta = x = \frac{1}{2}$, $\delta = 3$, $\gamma = 1$, and $\nu = 1/s$. Of these values, only ν is in agreement with the NRG results. This discrepancy has led to considerable debate about the validity of the quantum-to-classical mapping. However, a preponderance of the evidence now points to deficiencies of the bosonic chain NRG in the treatment of mean-field (noninteracting) critical points:

- It has been pointed out [33] that above the upper critical dimension, the order-parameter exponent β and the magnetic exponent δ are properties not just of the vicinity of the critical point (where the chain NRG seems to be valid) but of the full flow to the delocalized fixed point (where truncation errors are known to be inevitable [27]). Indeed, a solution of the sub-ohmic spin-boson Hamiltonian in its NRG chain form using a *variational matrix product state* method that selects an optimized bosonic basis for chain site has shown for $s = 0.2, 0.3$, and 0.4 that the exponents take their mean-field (classical) values [34]. This result strongly highlights the importance of basis selection in NRG treatments of problems involving bosonic baths.
- A second (seemingly independent) effect has been proposed [33, 35, 36] to account for the difference between the thermal critical exponent $x = s$ found within the NRG and the classical value $x = \frac{1}{2}$. The basic idea [35] is that since the NRG at iteration N , corresponding to temperature $T \simeq \omega_c \Lambda^{-N}$, neglects all oscillator weight at frequencies $\omega \lesssim T$, the distance $\alpha - \alpha_c$ from criticality acquires temperature-dependent corrections. As a result, χ^{-1} calculated at $\alpha = \alpha_c(T = 0)$ acquires a spurious term $\propto T^s$ that dominates the underlying mean-field $T^{1/2}$ term. An *ad hoc* procedure for correcting this problem has been proposed [35], but it leads to some apparent inconsistencies [37]. Whether or not there is a rigorous fix for the mass-flow problem remains an important open question.

4 Bose-Fermi NRG

4.1 The Bose-Fermi Kondo model

The Bose-Fermi Kondo impurity model with Ising-symmetric bosonic coupling is described by the Hamiltonian $H_{\text{BFK}} = H_K + H_B$, where H_B is as given in Eq. (8) and

$$H_K = \sum_{\mathbf{k}, \sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \frac{J}{2N_k} \mathbf{S} \cdot \sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} c_{\mathbf{k}\sigma}^\dagger \boldsymbol{\sigma}_{\sigma\sigma'} c_{\mathbf{k}'\sigma'} \quad (22)$$

is the standard Kondo Hamiltonian for the antiferromagnetic exchange coupling (with strength J) between an impurity spin- $\frac{1}{2}$ degree of freedom and the on-site spin of a conduction band. For the hybridization function in Eq. (3), H_K is the effective Hamiltonian to which the Anderson impurity model [Eq. (1)] reduces in the limit $0 < \Delta \ll -\varepsilon_d$, $U + \varepsilon_d$ in which real fluctuations

of the impurity occupancy are frozen out, and only the impurity spin degree of freedom remains active.

In this section, the conduction band dispersion $\varepsilon_{\mathbf{k}}$ is assumed to give rise to a density of states (per unit cell per spin orientation)

$$\rho(\varepsilon) \equiv N_k^{-1} \sum_{\mathbf{k}} \delta(\varepsilon - \varepsilon_{\mathbf{k}}) = \rho_0 |\varepsilon/D|^r \Theta(D - |\varepsilon|). \quad (23)$$

The case $r = 0$ represents a standard metal, while $r > 0$ describes a pseudogapped or semimetallic host. The bosonic bath is again taken to have a spectral function of the form given in Eq. (9) with the conventional replacement $2\pi\alpha \rightarrow (K_0 g)^2$, where K_0 is a density of states and g an energy.

The metallic ($r = 0$) Bose-Fermi Kondo model was originally introduced in the context of an extended dynamical mean-field theory for the two-band extended Hubbard model [38]. It has received most attention in connection with heavy-fermion quantum criticality, arising as an effective impurity problem in an extended dynamical mean-field treatment of the Kondo lattice model [39, 40]; here, the bosonic bath in the impurity problem embodies the effects, at a given Kondo lattice site, of the fluctuating magnetic field generated (via the Ruderman-Kittel-Kasuya-Yosida interaction) by local moments at other lattice sites. The $r = 0$ Bose-Fermi Kondo model with $s = 1$ also describes certain mesoscopic qubit devices, where the bosonic bath represents gate-voltage fluctuations [41], and the model has been invoked with $s = \frac{1}{2}$ and $s = \frac{1}{3}$ within a dynamical large- N treatment of a single-electron transistor coupled to both the conduction electrons and spin waves of ferromagnetic leads [42].

Perturbative renormalization-group studies of the $r = 0$ Bose-Fermi Kondo model [43, 44] indicate that for $0 < s \leq 1$, the competition between the conduction band and the bosonic bath for control of the impurity spin gives rise to a continuous quantum phase transition at $g = g_c(J)$ between a Kondo phase (for $g < g_c$) and a localized phase (for $g > g_c$). Just as in the spin-boson model, the phases can be distinguished by an order parameter $\phi = \lim_{h \rightarrow 0^+} \langle S_z \rangle$, where h is a local field that enters the Hamiltonian through a term $-hS_z$ added to H_{BFK} . In the localized phase, the order parameter increases as $(g - g_c)^\beta$ [cf. Eq. (18)]. For $g < g_c$, $\phi = 0$ but the effective Kondo temperature [the crossover scale to the low-temperature Fermi-liquid regime] vanishes continuously as $T_K \propto (g_c - g)^\nu$ [cf. Eq. (20)] describing a critical destruction of the Kondo many-body state. It is worth pointing out that although H_K exhibits SU(2) spin symmetry, the impurity-boson coupling lowers the overall symmetry of H_{BFK} to a U(1) invariance associated with conservation of the z component of local spin. This means that within any renormalization-group treatment, the Kondo exchange coupling evolves from $J\mathbf{S} \cdot \mathbf{s}_c$ (where $\mathbf{s}_c$ is the conduction-band spin at the impurity site) to an effective form $J_z S_z s_{c,z} + \frac{1}{2} J_\perp (S^+ s_c^- + S^- s_c^+)$. It is the spin-flip coupling $J_\perp$ that necessarily scales to infinity in the Kondo phase and to zero in the localized phase.

The fermionic pseudogap Kondo model described by Eqs. (22) and (23) with $r > 0$ has served as a paradigm for impurity quantum phase transitions [30, 45–47]. The suppression of the density of conduction states near the Fermi energy gives rise for $0 < r < \frac{1}{2}$ to a transition between

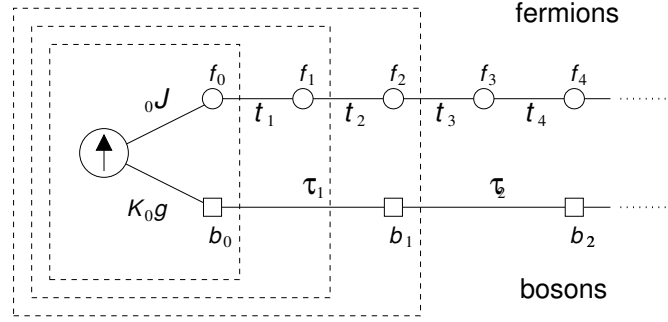


Fig. 8: Schematic representation of the Bose-Fermi NRG procedure for situations where the conduction half bandwidth D and the bosonic cutoff ω_c are of similar magnitudes. Since the bosonic tight-binding coefficients τ_n (and ξ_n , not shown) vary as Λ^{-n} , decaying twice as fast as the fermionic coefficients $t_n \propto \Lambda^{-n/2}$, the bosonic chain is extended by one site only at every other iteration. Dashed boxes from innermost to outermost enclose the degrees of freedom treated at NRG iterations $N = 0, 1$, and 2 . Adapted from [50].

an unscreened or local-moment phase for $J < J_c$ (where $\rho_0 J_c \simeq r$ for $r \ll \frac{1}{2}$) and a phase exhibiting partial Kondo screening of the impurity spin for $J > J_c$. The critical coupling J_c diverges as r approaches $\frac{1}{2}$ from below, and for $r \geq \frac{1}{2}$ the system is always in the local-moment phase where the impurity has a spin-doublet ground state. The pseudogap variant of the Bose-Fermi Kondo model has been proposed as a setting to explore the interplay between fermion- and boson-induced critical destruction of the Kondo effect [48].

4.2 NRG solution method

The Bose-Fermi Kondo model can be treated by suitably combining [49, 50] elements of the NRG treatment of pure-fermionic models (as summarized at the start of Sec. 2.2) and the NRG formulation of pure-bosonic models (as described in Sec. 3.2). Since the conduction band part of the Hamiltonian is mapped onto the tight-binding form in Eq. (5), it is convenient also to treat the bosonic bath part in the chain representation [Eq. (12)] rather than the alternative star formulation. The NRG iteration scheme must take into account that the fermionic hopping coefficients are proportional to $\Lambda^{-n/2}$ whereas the bosonic tight-binding coefficient decay as Λ^{-n} . It is in the spirit of the NRG for each iteration to treat fermions and bosons of the same energy scale. This can be achieved by adding one site to the fermionic chain at each iteration but extending the bosonic chain only at every other iteration. In situations where D and ω_c are not too different in magnitude, one can adopt the scheme shown schematically in Fig. 8 where bosonic site n is introduced at iteration $N = 2n$. (If $\omega_c \ll D$, it is more appropriate to delay the incorporation of bosonic site 1 until some iteration $N = M > 2$, and then to introduce bosonic sites $n > 1$ at iteration $N = M + 2n - 2$.)

To date, all NRG calculations for Bose-Fermi models such as H_{BFK} have been performed using a bosonic basis of eigenstates of $b_n^\dagger b_n$ with eigenvalues n_b satisfying $0 < n_b < N_b$. If one retains N_s many-body eigenstates after each iteration, then the CPU time is proportional to $(4N_b N_s)^3$ at any iteration where the bosonic chain is extended, and is otherwise proportional to $(4N_s)^3$.

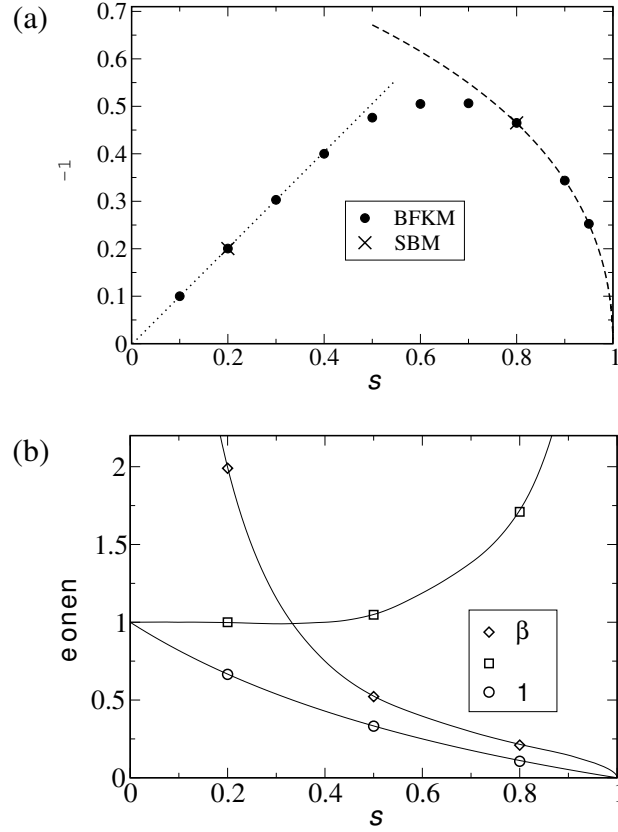


Fig. 9: Critical exponents of the Bose-Fermi Kondo model vs. bath exponent s : (a) Reciprocal correlation length exponent ν^{-1} , comparing NRG results for the Bose-Fermi Kondo model (circles) with those for the spin-boson model (crosses). The dotted line plots the mean-field dependence $\nu^{-1} = s$, while the dashed line shows the form $\nu^{-1} = \sqrt{2(1-s)} + C$ with $C = O(1)$ that arises in a perturbative expansion about $s = 1$. (b) Exponents β , γ , and $1/\delta$. Symbols show values directly computed within the NRG, while the lines come from substituting NRG values for $\nu(s)$ and $x(s)$ into the scaling relations in Eqs. (21). Reprinted from [50].

The additional factor of 4^3 at iterations of the former type compared to the iteration time in pure-bosonic problems such as the spin-boson model makes the NRG treatment of H_{BFK} quite computationally intensive. In comparison with a standard Kondo or Anderson problem, the computational time grows by a factor even greater than N_b^3 because it is generally necessary to increase N_s for the Bose-Fermi problem to achieve similar levels of accuracy. For $N_b = 8$, the overall time increase may be of order 10^4 – 10^5 .

4.3 Results

4.3.1 Metallic band ($r = 0$)

The NRG scheme described in Sec. 4.2 has been used to carry out a detailed study of the Bose-Fermi Kondo model with Ising-symmetric bosonic coupling [49, 50]. The principal finding of this work is that for sub-ohmic exponents $0 < s < 1$, the quantum critical point of the model is described by s -dependent critical exponents that are identical within numerical error to those

for the spin-boson model with the same bath exponent. This conclusion is illustrated in Fig. 9a, which compares Bose-Fermi Kondo and spin-boson values for the correlation-length exponent ν [defined in Eq. (20)]. Figure 9b compares directly computed values of β , γ , and δ [see Eqs. (18) and (19)] with ones derived from NRG values for $\nu(s)$ and $x(s)$ using the scaling relations in Eqs. (21). The NRG gives $x = s$ across the entire range $0 < s < 1$.

Given the lessons learned from the spin-boson model (see Sec. 3.3), it seems quite possible that the finding of an interacting critical point in the Bose-Fermi Kondo model for $0 < s < \frac{1}{2}$ is an artifact of the NRG treatment of the bosonic bath. This is not absolutely certain because there is a symmetry difference between the two problems. The Bose-Fermi Kondo model has $U(1)$ spin-rotational invariance about the z axis plus an additional Z_2 symmetry for longitudinal field $h = 0$. There are no such symmetries in the spin-boson model due to the nonzero value of Δ that must be present to induce a quantum phase transition. A recent study of a two-bath generalization of the spin-boson model in which the baths couple to different components of the impurity spin has found that a quantum critical point is classical in the presence of a transverse field but non-classical in the absence of the field where an additional Z_2 symmetry exists [34, 51]. Whether such a difference exists between the Bose-Fermi Kondo and one-bath spin-boson models is an interesting question.

4.3.2 Pseudogapped band ($r > 0$)

This subsection is devoted to the pseudogap variant of the Bose-Fermi Kondo model described by a density of states of the form of Eq. (23) with $0 < r < \frac{1}{2}$ and a bosonic bath exponent $\frac{1}{2} < s < 1$. For $g = 0$, the bosons decouple from the rest of the system and for $r > 0$ the model exhibits fermion-driven critical destruction of the Kondo effect at some $J = J_c(r, g = 0)$ [30, 45–47]. For $r = 0$, as described in the preceding paragraphs, the model instead features boson-driven Kondo destruction. This raises the question of the nature of the quantum critical point or points in situations where the fermionic and bosonic Kondo-destruction mechanisms are both present. The issue has been elucidated in [48], which presents NRG solutions of H_{BFK} for numerous combinations of exponents (r, s) , corroborated by continuous-time quantum Monte Carlo treatments of the corresponding Bose-Fermi Anderson model ($H_{\text{BFA}} = H_A + H_B$) for $(r, s) = (0.4, 0.6)$ and $(0.4, 0.8)$. In all cases $g > 0$, a continuous quantum phase transition occurs at some $J = J_c(r, s, g)$ between Kondo ($J > J_c$) and localized ($J < J_c$) phases. The physics in the former phase is essentially that of the Kondo-screened phase of the pseudogap Kondo model, modified by an irrelevant coupling to the bosons, while the localized phase behaves like that of the spin-boson model with irrelevant corrections from the Kondo coupling.

While continuous quantum phase transitions between phases in which the impurity degree of freedom is respectively quenched and asymptotically free are found in a number of models, including several discussed above, an interesting new feature of the pseudogap Bose-Fermi Kondo model is the existence of three qualitatively different types of quantum criticality, each accessed within a different region of the (r, s) space, as shown in Fig. 10:

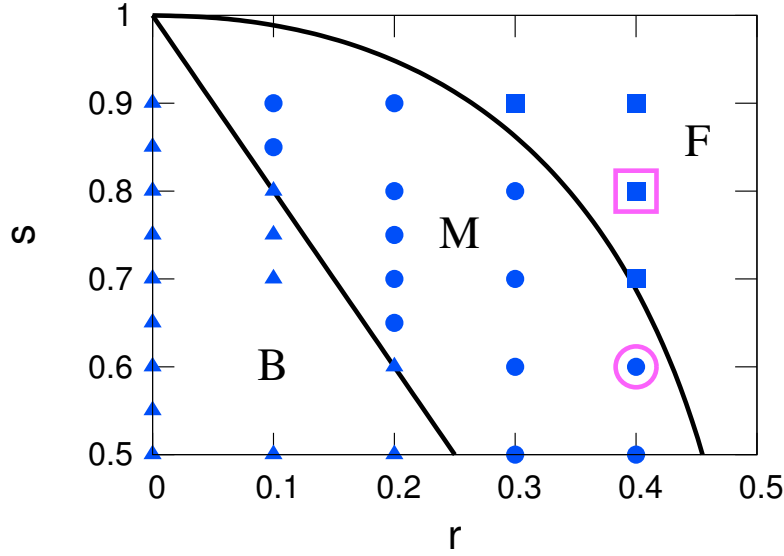


Fig. 10: Summary of the quantum criticality found in the pseudogap Bose-Fermi Kondo model for different band exponents r and bath exponents s . Squares, triangles, and circles respectively denote quantum criticality of the F, B, and M types described in the text. Filled symbols represent NRG results for the Bose-Fermi Kondo model while open symbols represent continuous-time quantum Monte Carlo results for the Bose-Fermi Anderson model. Solid lines show the conjectured boundaries $s = 1 - 2r$ and $x_B(s) = s = x_F(r)$ between the different types of criticality. Reprinted from [48].

- *Fermionic- or F-type criticality* arises in cases where the critical spin fluctuations induced by the band pseudogap are more divergent for temperatures $T \rightarrow 0$ than those resulting from the bosonic coupling. Specifically, F-type criticality occurs for all (r, s) such that the thermal critical exponent $x_F(r)$ of the pseudogap Kondo model—which is given by $x_F \simeq 1 - (\rho_0 J_c)^2$ for $0 < r \ll \frac{1}{2}$ [30] but in general must be determined numerically—is smaller than that of the spin-boson model $x_B(s) = s$ [see Eq. (19)]. The asymptotic low-energy spectrum calculated within the NRG recovers the $SU(2)$ spin symmetry broken by the bosonic coupling g , and this spectrum decomposes into a direct product of the spectrum of free bosons with bath exponent s [that of H_b in Eq. (12)] and the critical spectrum of the pseudogap Kondo model with band exponent r , i.e., (BF critical) = (B free) $\otimes$ (F critical). All calculated critical exponents are identical to those of the pure-fermionic pseudogap Kondo model with the same r .
- *Bosonic- or B-type criticality* is fully governed by the bosonic bath, a condition that (for reasons that have yet to be fully understood) occurs for $s < 1 - 2r$. The fixed-point spectrum exhibits $SU(2)$ spin symmetry and decomposes into a direct product of the critical spectrum of the spin-boson model with bath exponent s and the Kondo-phase spectrum of the pseudogap Kondo model with band exponent r , i.e., (BF critical) = (B critical) $\otimes$ (F Kondo). All calculated critical exponents are identical to those of the metallic ($r = 0$) Bose-Fermi Kondo model with the same s .

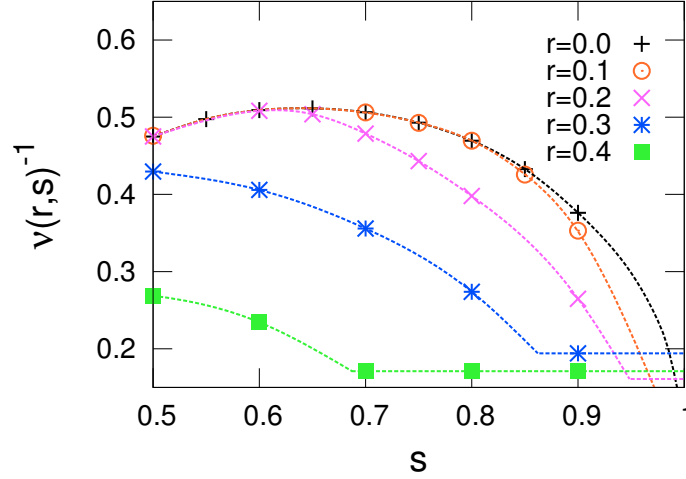


Fig. 11: Reciprocal of the Bose-Fermi Kondo model correlation length exponent $1/\nu$ vs. bath exponent s for the metallic case $r = 0$ and for pseudogaps described by $r = 0.1$ – 0.4 . The $r = 0$ exponents coincide with those of the corresponding spin-boson model and represent the values $1/\nu_B(s)$ describing pure-bosonic criticality. Each horizontal line segment shows a pure-fermionic value $1/\nu_F(r)$. The Bose-Fermi Kondo exponent coincides with $1/\nu_B(s)$ for $s < 1 - 2r$ and with $1/\nu_F(r)$ for $s \geq x_F(r)$. For $1 - 2r < s < x_F(r)$, $1/\nu(r, s)$ lies in between the bosonic and fermionic values. Reprinted from [48].

- *Mixed- or M-type criticality* is found across the range of intermediate s values such that $1 - 2r < s < x_F(r)$. The fixed-point spectrum exhibits broken $SU(2)$ spin symmetry and does not decompose into a direct product of bosonic and fermionic parts. The thermal exponent x takes its spin-boson model value $x_B(s) = s$ but the correlation-length exponent lies between the values for the spin-boson model and the pseudogap Kondo model, i.e., $\nu_F^{-1}(r) < \nu^{-1}(r, s) < \nu_B^{-1}(s)$, as illustrated in Fig. 11. Just as for the F- and B-types, all other calculated exponents obey the scaling relations in Eqs. (21), indicating that the quantum critical point is interacting.

It is important to consider whether the NRG accurately captures the quantum phase transitions of the pseudogap Bose-Fermi Kondo model. Given that over the entire region $0 < r < \frac{1}{2}$, $\frac{1}{2} < s < 1$ the exponent x determined using the NRG is *larger* than its mean-field counterpart $x = \frac{1}{2}$, it does not seem possible that the true critical behavior is being masked by the mass-flow problem identified in the spin-boson model for $0 < s < \frac{1}{2}$. Since the exponent ν can be determined entirely in the Kondo phase, it should be immune to the truncation errors that plague the localized phase (an immunity that has been found to be present in the spin-boson model [29]). These observations provide strong grounds for believing that the Bose-Fermi NRG method correctly accounts for the quantum critical behavior over the range of bath exponents considered.

Mixed quantum criticality arising from a nontrivial interplay between fermionic and bosonic critical fluctuations constitutes a new universality class of impurity quantum phase transitions distinct from those of the pseudogap Kondo model and the spin-boson model. This intriguing finding provides a motivation to search for yet other universality classes in models that may be more readily realized than the pseudogap Bose-Fermi Kondo model.

5 Closing

The goal of these lectures notes has been to introduce the main technical issues surrounding the extension of the NRG method to treat quantum impurity models that include bosonic degrees of freedom, to lay out the steps that have been developed to addresses these issues, and to provide an idea of the capabilities and limitations of the method that results. Space limitations have precluded description of the calculation of Green's functions and correlation functions, details of which can be found in some of the references, and have forced omission of many interesting applications as well as discussion of topical issues such as the treatment of systems out of equilibrium.

Acknowledgments

I acknowledge support from the National Science Foundation Materials World Network program under Grant No. DMR-1107814.

References

- [1] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
- [2] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1003 (1980); *ibid.* **21**, 1044 (1980)
- [3] R. Bulla, T.A. Costi, and Th. Pruscke, Rev. Mod. Phys. **80**, 395 (2008)
- [4] T. Costi: *NRG and Multiorbital Kondo Physics* in: E. Pavarini, E. Koch, and P. Coleman (eds.): *Many-Body Physics: From Kondo to Hubbard*, Modeling and Simulation, Vol. 5 (Forschungszentrum Jülich, 2015)
<http://www.cond-mat.de/events/correl15>
- [5] A.J. Leggett, S. Chakravarty, A.T. Dorsey, M.P.A. Fisher, A. Garg, and W. Zwerger, Rev. Mod. Phys. **59**, 1 (1987)
- [6] U. Weiss: *Quantum Dissipative Systems* (4th ed., World Scientific, Singapore, 2012)
- [7] F.D.M. Haldane, Phys. Rev. B **15**, 2477 (1977)
- [8] K. Schönhammer and O. Gunnarsson, Phys. Rev. B **30**, 3141 (1984)
- [9] A.C. Hewson and D. Meyer, J. Phys.: Condens. Matter **14**, 427 (2002)
- [10] H.C. Lee and H.-Y. Choi, Phys. Rev. B **69**, 075109 (2004)
- [11] E. Šimánek, Solid State Commun. **32**, 731 (1979)
- [12] C.S. Ting, D.N. Talwar, and K.L. Ngai, Phys. Rev. Lett. **45**, 1213 (1980)
- [13] H.-B. Schüttler, and A.J. Fedro, Phys. Rev. B **38**, 9063 (1988)
- [14] P.S. Cornaglia, H. Ness, and D.R. Grempel, Phys. Rev. Lett. **93**, 147201 (2004)
- [15] P.S. Cornaglia, D.R. Grempel, and H. Ness, Phys. Rev. B **71**, 075320 (2005)
- [16] J. Paaske and K. Flensberg, Phys. Rev. Lett. **95**, 176801 (2005)
- [17] For compactness of notation, we drop all factors of the reduced Planck constant $\hbar$, Boltzmann's constant k_B , the impurity magnetic moment $g\mu_B$, and the electronic charge e , as well as the zero-point energy of every harmonic oscillator.
- [18] P.W. Anderson, Phys. Rev. **124**, 41 (1961)
- [19] M. Cheng, M.T. Glossop, and K. Ingersent, Phys. Rev. B **80**, 165113 (2009)
- [20] C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950)
- [21] G. Iche and A. Zawadowski, Solid State Commun. **10**, 1001 (1972)

- [22] A. Taraphder and P. Coleman Phys. Rev. Lett. **66**, 2814 (1991)
- [23] H. Spohn and R. Dümcke, J. Stat. Phys. **41**, 389 (1985)
- [24] F. Guinea, V. Hakim, and A. Muramatsu, Phys. Rev. B **32**, 4410 (1985)
- [25] T.A. Costi and C. Kieffer, Phys. Rev. Lett. **76**, 1683 (1996)
- [26] R. Bulla, N.-H. Tong, and M. Vojta, Phys. Rev. Lett. **91**, 170601 (2003)
- [27] R. Bulla, H.-J. Lee, N.H. Tong, and M. Vojta, Phys. Rev. B **71**, 045122 (2005)
- [28] J. Sabio, L. Borda, F. Guinea, and F. Sols, Phys. Rev. B **78**, 085439 (2008)
- [29] M. Vojta, N.-H. Tong, and R. Bulla, Phys. Rev. Lett. **94**, 070604 (2005)
- [30] K. Ingersent and Q. Si, Phys. Rev. Lett. **89**, 076403 (2002)
- [31] M.E. Fisher, S. Ma, and B.G. Nickel, Phys. Rev. Lett. **29**, 917 (1972)
- [32] E. Luijten and H.W.J. Blöte, Phys. Rev. B **56**, 8945 (1997)
- [33] M. Vojta, N.-H. Tong, and R. Bulla, Phys. Rev. Lett. **102**, 249904(E) (2009)
- [34] C. Guo, A. Weichselbaum, J. von Delft, and M. Vojta, Phys. Rev. Lett. **108**, 160401 (2012)
- [35] M. Vojta, R. Bulla, F. Güttge, and F. Anders, Phys. Rev. B **81**, 075122 (2010)
- [36] M. Vojta, Phys. Rev. B **85**, 115113 (2012)
- [37] S. Kirchner, K. Ingersent, and Q. Si, Phys. Rev. B **85**, 075113 (2012)
- [38] J.L. Smith and Q. Si, Europhys. Lett. **45**, 228 (1999)
- [39] Q. Si, S. Rabello, K. Ingersent, and J.L. Smith, Nature (London) **413**, 804 (2001)
- [40] Q. Si, S. Rabello, K. Ingersent, and J.L. Smith, Phys. Rev. B **68**, 115103 (2003)
- [41] K. Le Hur, Phys. Rev. Lett. **92** 196804 (2004)
- [42] S. Kirchner, L. Zhu, Q. Si, and D. Natelson, Proc. Natl. Acad. USA **102**, 18824 (2005)
- [43] L. Zhu and Q. Si, Phys. Rev. B **66**, 024426 (2002)
- [44] G. Zaránd and E. Demler, Phys. Rev. B **66**, 024427 (2002)
- [45] D. Withoff and E. Fradkin, Phys. Rev. Lett. **64**, 1835 (1990)
- [46] C. Gonzalez-Buxton and K. Ingersent, Phys. Rev. B **57**, 14254 (1998)
- [47] L. Fritz and M. Vojta, Phys. Rev. B **70** 214427 (2004)

- [48] J. Pixley, S. Kirchner, K. Ingersent, and Q. Si, Phys. Rev. B **88**, 245111 (2013)
- [49] M.T. Glossop and K. Ingersent, Phys. Rev. Lett. **95**, 067202 (2005)
- [50] M.T. Glossop and K. Ingersent, Phys. Rev. B **75**, 104410 (2007)
- [51] B. Bruognolo, A. Weichselbaum, C. Guo, J. von Delft, I. Schneider, and M. Vojta, Phys. Rev. B **90**, 245130 (2014)

7 Frustrated Spin Systems

Frédéric Mila

Institute of Theoretical Physics

Ecole Polytechnique Fédérale de Lausanne

1015 Lausanne, Switzerland

Contents

1	Introduction	2
2	Competing interactions and degeneracy	3
2.1	Ising	3
2.2	Heisenberg	6
3	Classical ground-state correlations	8
3.1	Algebraic correlations	8
3.2	Dipolar correlations	10
4	Order by disorder	12
4.1	Quantum fluctuations	12
4.2	Thermal fluctuations	14
5	Alternatives to magnetic long-range order in Heisenberg models	14
5.1	Spin gap	15
5.2	Resonating Valence Bond spin liquids	17
5.3	Algebraic spin liquids	21
5.4	Chiral spin liquids	23
5.5	Nematic order	24
6	Conclusion	27

1 Introduction

In the field of magnetism, the word frustration was first introduced in the context of spin glasses to describe the impossibility of simultaneously satisfying all exchange processes. In this lecture, we are primarily interested in disorder-free systems that can be described by a periodic Hamiltonian. In that case, frustration is more precisely described as *geometrical frustration*, a concept that has received the following general definition: One speaks of *geometrical frustration* when a local condition is unable to lead to a simple pattern for an extended system [1]. Typical examples are paving problems, where some figures such as triangles in two dimensions lead to regular, packed pavings while others such as pentagons are unable to lead to a compact, periodic structure.

In this lecture, we will be dealing mostly with two models of magnetism: the *Ising* model

$$H = \sum_{(i,j)} J_{ij} S_i S_j, \quad S_i, S_j = \pm 1 \text{ or } \uparrow, \downarrow \quad (1)$$

and the *Heisenberg* model

$$H = \sum_{(i,j)} J_{ij} \vec{S}_i \cdot \vec{S}_j \quad (2)$$

where the spins $\vec{S}_i$ are unit vectors in the *classical* case, and components of a quantum spin in the *quantum* case: $[S_i^\alpha, S_i^\beta] = i\epsilon^{\alpha\beta\gamma} S_i^\gamma$, and $\vec{S}_i^2 = S(S+1)$. In both cases, i and j are sites of a periodic lattice, and J_{ij} is assumed to depend only on their relative position.

Frustration can only occur if at least some exchange paths are antiferromagnetic, i.e., if some of the exchange integrals J_{ij} are positive, since, if all exchange paths are ferromagnetic with negative exchange integrals, the configuration with all spins parallel is clearly the ground state. However, even when all bonds are antiferromagnetic, geometrical frustration is not necessarily realized. Indeed, for bipartite lattices such as the square lattice that can be divided into two sublattices such that each spin of one sublattice is only coupled to spins on the other sublattice, the energy of the Ising model or of the classical Heisenberg model is simply minimized by the Néel configuration, in which the spins of one sublattice are parallel to each other and antiparallel to all spins of the other sublattice.

A necessary condition to satisfy the general condition of geometrical frustration with only antiferromagnetic exchange interactions is to have loops of odd length. This is however *not* sufficient. Indeed, as we shall see, it is often possible to minimize the energy of the classical antiferromagnetic Heisenberg model with a simple helical arrangement of spins, and if this defines a unique ground state, as in the case of the triangular lattice with nearest-neighbor interactions, the system is strictly speaking not geometrically frustrated: Geometrical frustration occurs when there is no unique way to minimize the energy but when there are other ways with less simple (often non periodic) structures to reach the ground-state energy.

The objective of this lecture is to review the physical consequences of this degeneracy from a theoretical perspective. For Ising spins, degeneracy can lead to all types of zero-temperature behaviors: long-range order, algebraic order, dipolar correlations, or complete disorder. For

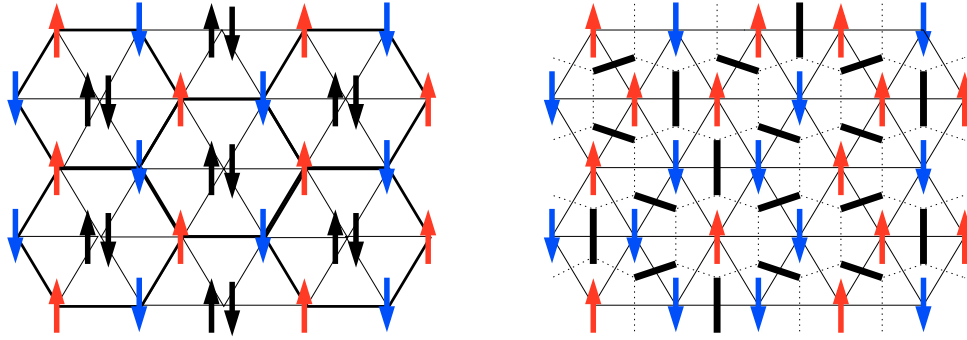


Fig. 1: *Left panel: Example of degenerate ground states of the antiferromagnetic Ising model on the triangular lattice. Once spins have been arranged antiferromagnetically on a honeycomb sublattice (thick solid line), the spins in the center of the hexagons can point up or down. Right panel: Mapping between a ground state of the Ising model on the triangular lattice and a dimer covering on the dual honeycomb lattice. The bonds of the triangular lattice with parallel spins are crossed by a dimer.*

Heisenberg models, fluctuations (thermal or quantum) play a major role. They can order the system by picking one ordered state out of the ground-state manifold, but they can also destroy any kind of magnetic long-range order. This opens the way to new types of ground states such as spin nematics (where the order parameter is not the local spin but a more complicated object), valence-bond crystals (completely non-magnetic states with a broken translational symmetry), or quantum spin liquids where both the rotational $SU(2)$ symmetry in spin space and the translation symmetry in real space are preserved.

2 Competing interactions and degeneracy

2.1 Ising

For the Ising model, competing interactions generally lead to an infinite degeneracy. As a first hint, let us consider the antiferromagnetic Ising model on the triangular lattice. On a triangle, the best one can do is to satisfy two bonds out of three, and any configuration with two up spins and one down spin or two down spins and one up spin on each triangle minimizes the energy. A simple way to satisfy this condition is to look at the triangular lattice as a centered honeycomb lattice. Then, if the honeycomb lattice is in a Néel state, the condition will be automatically satisfied regardless of the orientation of the spins inside the hexagons (see left panel of Fig. 1). So the ground state degeneracy is at least equal to $2^{N/3}$, where N is the number of sites, and there is a residual entropy per site bounded from below by $(1/3) \ln 2 \simeq 0.2310$. In fact, the residual entropy is much larger, as first shown by Wannier [2], who derived the exact result $S/N = 0.3230\dots$

A useful way to derive this result for the purpose of this lecture is to map the problem onto that of the dimer coverings on the dual honeycomb lattice by putting a dimer across each unsatisfied bond of a ground state of the Ising model [3] (see right panel of Fig. 1). Up to a factor 2 associ-

ated with flipping all spins, there is a one-to-one correspondence. The problem of counting the dimer coverings on a planar graph was addressed in the early sixties [4, 5], and the following theorem has been proven:

Theorem: If one can attach to each adjacent pair of sites an arrow such that around each loop with an even number of sites N the number of arrows in each direction is odd, the total number of dimer coverings can be expressed as the determinant of a periodic, skew-symmetric matrix.

Proof: The number of dimer coverings is clearly given by

$$Z = \frac{1}{\left(\frac{N}{2}\right)! 2^{N/2}} \sum_P b(p_1, p_2) b(p_3, p_4) \cdots b(p_{N-1}, p_N)$$

where the sum runs over the permutations $P = \{p_1, \dots, p_N\}$ of $1, \dots, N$, and

$$b(i, j) = \begin{cases} 1 & \text{if } i, j \text{ adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Now, define the skew-symmetric Kasteleyn matrix $a(i, j)$ by:

$$a(i, j) = \begin{cases} 1 & \text{if } i, j \text{ adjacent and } i \rightarrow j \\ -1 & \text{if } i, j \text{ adjacent and } i \leftarrow j \\ 0 & \text{otherwise.} \end{cases}$$

where $i \rightarrow j$ means that the arrow goes from i to j . Then,

$$Z = \left| \frac{1}{\left(\frac{N}{2}\right)! 2^{N/2}} \sum_P \varepsilon(P) a(p_1, p_2) \cdots a(p_{N-1}, p_N) \right|$$

where $\varepsilon(P)$ is the signature of the permutation. Indeed, this will be clearly true if, for all permutations, i.e., for all dimer coverings, the sign of $\varepsilon(P) a(p_1, p_2) \cdots a(p_{N-1}, p_N)$ is the same. Now, consider 2 dimer coverings C and C' . To go from one to the other, one just has to shift sites around loops of even length. For the sites of a given loop, the product of the matrix elements of C and C' will be negative because, thanks to the hypothesis of the theorem, the number of arrows in each direction is odd. Besides, the signature of the permutation is negative because the number of sites is even. So, each loop contributes a factor with the same sign for C and C' , and the terms in the sum corresponding to C and C' have the same sign, which implies that the terms in the sum have the same sign for all dimer coverings.

Now, the sum over P is the Pfaffian of the skew-symmetric matrix $a(i, j)$, and its square is equal to the determinant of a . So finally,

$$Z = \sqrt{\det a}$$

The determinant is the product of the eigenvalues of a , which, on a periodic lattice, can be easily calculated using the Bloch theorem. In the case of the honeycomb lattice, this leads to

$$\frac{1}{N_{hc}} \ln Z = \frac{1}{4} \int_0^1 dx \int_0^1 dy \ln |3 + 2 \cos(2\pi y) - 2 \cos(2\pi(x + y)) - 2 \cos(2\pi x)| \simeq 0.1615$$

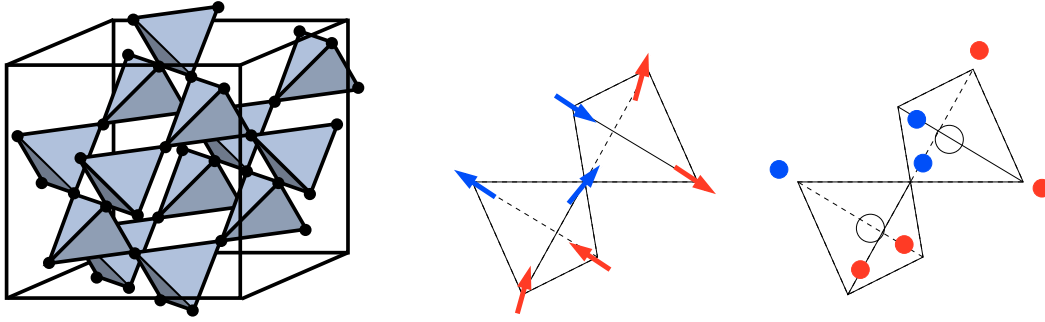


Fig. 2: Left panel: sketch of the pyrochlore lattice. Middle panel: example of a 2-in 2-out structure stabilized in spin ice. Right panel: same configuration as that of the middle panel, but for water ice, the blue and red dots standing for H atoms. Each oxygen atom (empty circle) forms an H_2O molecule with the two hydrogen atoms close to it.

Since the number of sites of the honeycomb lattice $N_{hc} = N/2$, the entropy per site for the Ising model is given by $S/N = 0.3230 \dots$, in agreement with Wannier's result [2].

When the mapping on an exactly soluble dimer covering problem is not possible, one can always resort to a numerical enumeration of all configurations to estimate the residual entropy. However, it is often possible to get a good estimate with the help of a simple argument developed by Pauling in the context of ice [6]. Let us consider the case of $Dy_2Ti_2O_7$ and $Ho_2Ti_2O_7$ [7]. In these systems, known as spin-ice systems, classical spins sit on a pyrochlore lattice (see left panel of Fig. 2). They have to point along the direction joining the site on which they sit to the centers of the neighboring tetrahedra, and the interaction is ferromagnetic. Then, on a tetrahedron, the energy is minimized for configurations with two spins in and two spins out (see middle panel of Fig. 2). This model can be mapped onto the problem of water ice, in which oxygen atoms sit at the center of the tetrahedra of a pyrochlore lattice, and the hydrogen atoms sit in principle on the pyrochlore lattice, but on each tetrahedron two of them get closer to the oxygen to form a water molecule H_2O (see right panel of Fig. 2). Remarkably, it can also be mapped onto the antiferromagnetic Ising model on the pyrochlore lattice, with two spins up and two spins down on each tetrahedron.

To estimate the entropy, Pauling noted that, on a given tetrahedron, there are in total $2^4 = 16$ configurations, but only 6 of them satisfy the constraint. So, to estimate the number of configurations for a system of N sites, he suggested to multiply the total number of configurations 2^N by 6/16 per tetrahedron. Since the number of tetrahedra is $N/2$, this gives:

$$Z \approx 2^N \left(\frac{6}{16} \right)^{\frac{N}{2}} = \left(\frac{3}{2} \right)^{N/2}$$

leading to

$$\frac{S}{N} = \frac{1}{2} \ln \left(\frac{3}{2} \right) = 0.2027$$

This estimate compares remarkably well with the 'exact' numerical result $S/N = 0.20501\dots$ [8]. Even more remarkably, this value has been measured experimentally in $Dy_2Ti_2O_7$ [9].

Indeed, the entropy per site at high temperature is equal to $\ln 2$, and if the ground state is degenerate, the residual entropy can be determined as:

$$S/N = \ln 2 - \int_{0+}^{+\infty} \frac{C(T)dT}{T} \quad (3)$$

where $C(T)$ is the specific heat.

2.2 Heisenberg

For the Heisenberg model, frustration is often used, by analogy with the Ising case, as a synonym of competing interactions, but the effects to be discussed below occur when the competition is so severe that for classical spins the ground state is infinitely degenerate. That this is not always the case is clear from the following theorem [10]:

Theorem: On a Bravais lattice, i.e., a lattice with one site per unit cell, the classical energy is minimized by a helical structure $\vec{S}_{\vec{R}_i} = (\cos(\vec{k} \cdot \vec{R}_i), \sin(\vec{k} \cdot \vec{R}_i), 0)$, where $\vec{k}$ is a minimum of the Fourier transform of the coupling constant $J(\vec{k}) = \sum_{\vec{R}_j} J_{\vec{R}_i \vec{R}_j} \exp[i\vec{k} \cdot (\vec{R}_j - \vec{R}_i)]$.

Proof: This theorem is easily proven by first replacing the local constraint $||\vec{S}_i||^2 = 1$ by a global one $\sum_i ||\vec{S}_i||^2 = N$, and by looking for a specific solution of the global constraint that satisfies the local constraint.

In fact, even on non-Bravais lattices, it is often possible to minimize the energy with some kind of helical state. However, the fact that the energy can be minimized by a regular structure does not imply that this is the only one. Let us demonstrate this in a few representative cases.

2.2.1 J_1 - J_2 model on the square lattice

For the Heisenberg model on the square lattice with nearest-neighbor coupling J_1 and next-nearest-neighbor coupling J_2 (see left panel of Fig. 3), the Fourier transform of the coupling constant is given by:

$$J(\vec{k}) = 2J_1(\cos k_x + \cos k_y) + 4J_2 \cos k_x \cos k_y \quad (4)$$

As long as $J_1 > 2J_2 > 0$, the minimum is reached for $\vec{k} = (\pi, \pi)$, and the ground state has Néel order. However, when $J_2 > J_1/2 > 0$, $J(\vec{k})$ is minimized by two wave-vectors: $\vec{k} = (0, \pi)$ and $\vec{k} = (\pi, 0)$. So the minimum energy is reached for two helical states. Quite remarkably, the energy per site is given by $E/N = -2J_2$ and does not depend on J_1 . This is a consequence of the fact that in these states the two sublattices are Néel ordered, so that J_1 couples any spin to two pairs of spins pointing in opposite directions and drops from the energy. But this remains true regardless of the relative angle θ between the spins of each sublattice, so that the ground-state manifold consists of all states with Néel-ordered sublattices (see left panel of Fig. 3). The ground state is thus infinitely degenerate, and the degeneracy is controlled by a continuous parameter, the angle θ between the sublattices.

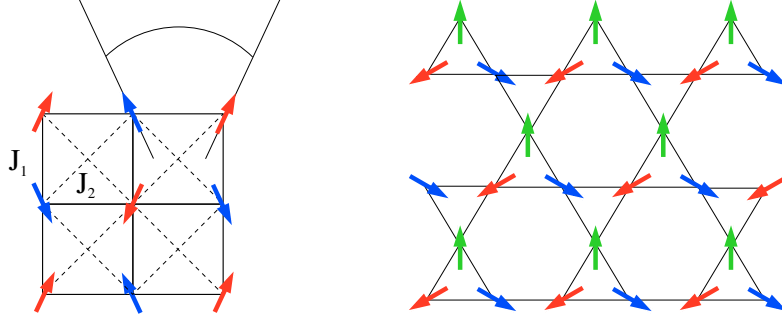


Fig. 3: Left panel: sketch of the ground state of the J_1 - J_2 model on the square lattice for $J_2 > J_1/2$. Each sublattice has Néel order. The ground-state energy is independent of the angle θ between the spins on the two sublattices. Right panel: $\vec{q} = \vec{0}$ ground state of the classical Heisenberg model on the kagome lattice. New, non-coplanar ground states can be generated by rotating the spins along any line (e.g. the red and blue spins of a horizontal line) around the common direction of the spins to which they are coupled (e.g. the vertical direction of the green spins)

2.2.2 Kagome lattice

For three spins, the classical Heisenberg model is proportional to

$$\vec{S}_1 \cdot \vec{S}_2 + \vec{S}_2 \cdot \vec{S}_3 + \vec{S}_3 \cdot \vec{S}_1 = \frac{1}{2}(\vec{S}_1 + \vec{S}_2 + \vec{S}_3)^2 - \frac{3}{2}$$

This expression is minimized as soon as $\vec{S}_1 + \vec{S}_2 + \vec{S}_3 = \vec{0}$. So, for antiferromagnets built out of triangles, a sufficient condition to minimize the energy is to find a configuration for which this constraint is satisfied on all triangles. For the triangular lattice, this condition leads, up to a global rotation of the spins, to a unique three-sublattice helical state. However, for the kagome lattice, which consists of corner-sharing triangles, this can be satisfied in an infinite number of ways. For coplanar configurations, once the direction of one spin has been chosen, the problem is equivalent to the antiferromagnetic 3-state Potts model, which is known to have an infinitely degenerate ground state with an extensive residual entropy [11]. This leads to an infinite, discrete degeneracy. But the situation is far richer. Consider for instance the ground state in which all up triangles have the same pattern (see right panel of Fig. 3). Non-coplanar ground states can be constructed by rotating the spins along any line around the common direction of the spins adjacent to it. This can be done independently on all lines parallel to a given direction. More generally, starting from any ground state, the spins inside a cluster surrounded by identical spins can be freely rotated. So the degeneracy is controlled by an infinite number of continuous variables.

2.2.3 Pyrochlore

The situation is similar for the three-dimensional pyrochlore lattice. All the states that satisfy the constraint $\vec{S}_1 + \vec{S}_2 + \vec{S}_3 + \vec{S}_4 = \vec{0}$ on each tetrahedron are ground states. This constraint can be satisfied in many ways, and new ground states can often be generated by rotations of subsets of spins.

3 Classical ground-state correlations

When the classical ground state of a model is infinitely degenerate, zero-temperature correlations are naturally defined as the average over all ground states. Naively, one could imagine that averaging over an infinite number of different states will generally lead to effective disorder, but quite remarkably this is not the case, and most systems that have an infinite but discrete ground-state degeneracy turn out to exhibit algebraic correlations, sometimes of dipolar type, although some are completely disordered (e.g. the antiferromagnetic model on the kagome lattice [12]), and some long-range ordered (e.g. the 3-state Potts model on the dice lattice [13] or the bilinear-biquadratic Heisenberg model on the triangular lattice [14]).

3.1 Algebraic correlations

A well known example of algebraic correlations is provided by the AF Ising model on the triangular lattice. The original proof by Stephenson [15] that the correlations decay as $1/r^{1/2}$ is too involved to be reproduced here, but this result can be made plausible thanks to the mapping onto the dimer problem on the honeycomb lattice. Indeed, the Pfaffian that enters the calculation of the number of dimer coverings can be reformulated in terms of an integral over Grassmann variables, and the Green function between Grassmann variables at sites i and j is the matrix element $(a^{-1})_{ij}$, where a is the Kasteleyn matrix. Now, in Fourier space, the spectrum of the Kasteleyn matrix has a Dirac point at zero eigenvalue for the honeycomb lattice. This implies that the Green function decays algebraically at long distance. Using the Wick theorem, this in turn implies that more complicated correlations such as dimer-dimer correlations in the dimer covering problem or spin-spin correlations in the Ising model should also decay algebraically. Intuitively, the presence of algebraic correlations suggests that the system is almost ordered and that configurations that are not too far from a specific one dominate the sum. A plausible condition for this specific configuration is to be maximally flippable, i.e., connected by individual spin flips to a maximal number of allowed configurations. To implement this idea, it has been suggested to map these models onto height models [16, 17], the maximally flippable state corresponding to a flat surface. Such models represent the fluctuations of the surface of a solid by assigning to each point a height. They describe the roughening transition of a solid between a flat surface, where the height difference is bounded from above, and a rough surface, where height differences diverge [18].

For the triangular Ising antiferromagnet, the mapping onto height variables proceeds in two steps. First of all, microscopic height variables $z(\vec{r})$ are defined on the vertices of the original triangle lattice (see left panel of Fig. 4). To each ground state of the Ising model, one first associates a dimer covering of the dual honeycomb lattice. Then, after choosing the height of one site, one associates to each dimer covering a height configuration following the prescription that, when going clockwise around an up triangle, the height difference between neighboring sites is equal to 2 if one crosses a dimer and -1 if one does not cross a dimer. With this prescription, the height differences around all triangles (up and down) is zero, and the assignment is consistent.

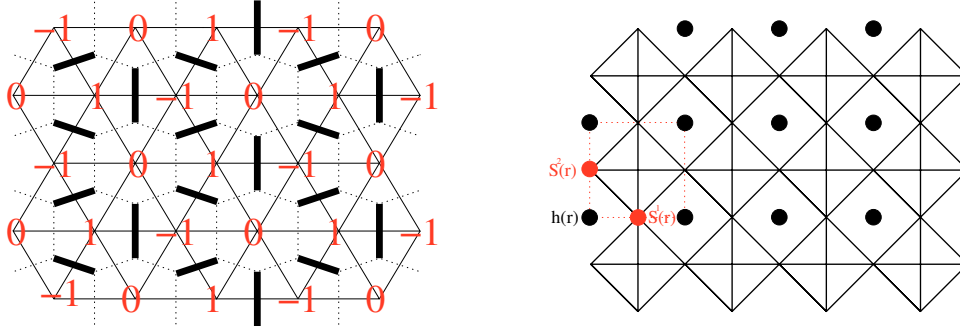


Fig. 4: Left panel: height mapping of the Ising model on the triangular lattice. The configuration represented here is the flat one. It corresponds to the maximally flippable configuration in terms of Ising spins (not shown). Right panel: sketch of the mapping of the Ising model onto the checkerboard lattice on a height model. The dotted red line is the unit cell.

Next, one defines a smooth height field on the dual lattice $h(\vec{x})$ by coarse-graining the field $z(\vec{r})$, i.e., by averaging it on each triangular plaquette:

$$h(\vec{x}) = [z(\vec{r}_1) + z(\vec{r}_2) + z(\vec{r}_3)]/3$$

where $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ are the sites of a triangle, and $\vec{x} = (\vec{r}_1 + \vec{r}_2 + \vec{r}_3)/3$ is a site of the dual lattice.

The idea is now to describe the ground-state manifold as fluctuations around the flat surface, where the height variable is almost the same everywhere. With the prescription above, this corresponds to the configuration shown in the left panel of Fig. 4, in which $h(\vec{x}) = 0$ on all triangles. So it is natural to assume that this height configuration has a maximal weight, and that other height configurations will be penalized. The simplest assumption is a Gaussian weight. Going to a continuous height variable, one assumes a free energy of the form:

$$F(\{h(\vec{x})\}) = \int d\vec{x} \frac{K}{2} |\vec{\nabla} h(\vec{x})|^2 = \sum_{\vec{q}} \frac{K}{2} q^2 |h(\vec{q})|^2$$

For a mode $h(\vec{q})$, the mean value of $h(\vec{q})^2$ is given by:

$$\langle |h(\vec{q})|^2 \rangle = \frac{\int dh h^2 \exp(-\frac{K}{2} q^2 h^2)}{\int dh \exp(-\frac{K}{2} q^2 h^2)} = \frac{1}{K q^2},$$

which implies that

$$C(\vec{r}) \equiv \frac{1}{2} \langle |h(\vec{r}) - h(\vec{0})|^2 \rangle = \int dq (1 - e^{i\vec{q}\cdot\vec{r}}) \frac{1}{K q^2} \propto \frac{1}{2\pi K} \ln \frac{\pi r}{\alpha}, \quad r \rightarrow +\infty$$

To extract the correlations of the Ising model, one must relate the spin variables $\sigma_{\vec{r}_i}$ to the height variables. This relation is not simple, but it is enough to realize that the spin variables have to be periodic functions of the height variables of period six. Indeed, turning the spins around a triangle twice leads to the same configuration, while the height has changed by 6. So any local operator of the spin $O(\vec{r})$ can be expanded as

$$O(\vec{r}) = \sum_{G \neq 0} O_G e^{iGh(\vec{r})}$$

with $G = 2\pi n/6$, n integer. Using the identity

$$\langle e^{-iG[h(\vec{r})-h(\vec{0})]} \rangle = e^{-\frac{1}{2}G^2 \langle |h(\vec{r})-h(\vec{0})|^2 \rangle}$$

it is clear that the long-range correlations are dominated by the smallest value of G . So one can assume $\sigma(\vec{r}) \propto e^{i\frac{2\pi}{6}h(\vec{r})}$ to calculate the long-range correlations, which leads to:

$$\langle \sigma(\vec{r})\sigma(\vec{0}) \rangle \propto \left(\frac{\pi r}{a}\right)^{-\left(\frac{2\pi}{6}\right)^2 \frac{1}{2\pi K}} = \left(\frac{\pi r}{a}\right)^{-\frac{2\pi}{36K}}$$

Comparing this result with the exact result that the correlations decay as $1/r^{1/2}$, one gets $K = \pi/9$. This shows that the system is in the rough phase where $\langle |h(\vec{r}) - h(\vec{0})|^2 \rangle$ diverges logarithmically since this phase is stable as long as $K < \pi/2$.

3.2 Dipolar correlations

The checkerboard lattice (see right panel of Fig. 4) is another example of a frustrated lattice on which the Ising model is infinitely degenerate, the rule in the ground state being that $\sum_{\square} S_i = 0$ for all plaquettes with diagonal bonds. This is not a planar graph (bonds are crossing), and a mapping onto a dimer model is not possible. So the form of the long-range correlations cannot be guessed from the spectrum of a Kasteleyn matrix. The height mapping is in that case very useful, and it brings an unexpected new feature of the correlations [17].

The mapping works as follows (see Fig. 4, right panel): The height variables are numbers defined in the squares without diagonal bonds and related to the spins by

$$\begin{cases} S^1(\vec{r}) = (-1)^{x+y} \Delta_x h \\ S^2(\vec{r}) = (-1)^{x+y} \Delta_y h \end{cases}$$

where $S^1(\vec{r})$ and $S^2(\vec{r})$ refer to the two spins in the unit cell.

In that case, the spin is related to the height field by:

$$S^i(\vec{r}) = (-1)^{x+y} \partial_{x_i} h + e^{i\pi h} + \dots$$

The dominant contribution to the correlations comes from the first term:

$$\langle S^i(\vec{r}) S^j(\vec{0}) \rangle = (-1)^{x+y} \partial_{x_i} \partial_{x_j} C(\vec{r}) \propto \frac{(-1)^{x+y}}{2\pi K} \partial_{x_i} \partial_{x_j} \ln \left(\frac{\pi r}{\alpha} \right),$$

which finally leads to

$$\langle S^i(\vec{r}) S^j(\vec{0}) \rangle = \frac{(-1)^{x+y}}{2\pi K} \frac{r^2 \delta_{ij} - 2x_i x_j}{r^4} \quad (5)$$

These correlations have a dipolar form.¹

¹The mapping onto a height model leads in general to algebraic and often dipolar correlations. What makes the height mapping possible? The mapping is possible if a height configuration leads to a single spin configuration and a spin configuration to a single height configuration up to a global shift. Since the height is single valued, one has to go back to the same height around basic loops. For this to occur, elementary loops for the height should be constrained by a local condition in the ground state manifold. This is the case for the triangular lattice ($\sum_i \sigma_i = \pm 1$, never 3 or -3), and for the checkerboard lattice ($\sum_{\square} S_i = 0$).

The situation is very similar for the pyrochlore lattice with antiferromagnetic couplings [19], but the effective model is a lattice gauge theory. In the ground state, the rule for each tetrahedron is that

$$\sum_{\boxtimes} S_i = 0$$

Consider the dual lattice of the pyrochlore lattice. This is the diamond lattice, a bipartite lattice. Divide it into two sublattices A and B and on each bond κ define a unit vector $\hat{e}_\kappa$ from A to B . Next, define a field on the site of the original lattice, hence on the bond of the diamond lattice, by

$$\vec{B}_\kappa = S_\kappa \hat{e}_\kappa$$

The lattice version of the integral of the divergence around a site $\vec{x}$ of the diamond lattice is given by

$$\sum_{\kappa(\vec{x})} \vec{B}_\kappa \cdot \hat{e}_\kappa = \sum_{\kappa(\vec{x})} S_\kappa = 0$$

because of the local rule in the ground-state manifold of the AF Ising model on the pyrochlore lattice.

So, one can look in the continuum for a field with zero divergence. Upon coarse graining, the configurations with small coarse-grained field $\vec{B}$ are favored. Indeed, to go from one ground state to the other, one must flip the spins along a loop of alternating spins to fulfill the constraint $\sum_{\boxtimes} S_i = 0$. Along such loops, the sum of the field $\vec{B}_\kappa = \vec{0}$. This implies that configurations with a very small $\vec{B}$ will have small loops of $\vec{B}_\kappa$ and will be very flippable.

So, one can postulate a weight

$$S(\vec{B}(\vec{x})) = \exp \left[-\frac{K}{2} \int d^3 \vec{r} \vec{B}(\vec{r})^2 \right] \quad (6)$$

Since $\text{div } \vec{B} = 0$, one can choose a vector potential $\vec{A}$ such that $\vec{B} = \vec{\nabla} \times \vec{A}$. With the gauge $\text{div } \vec{A} = 0$, its correlations are given by:

$$\langle |A(\vec{q})|^2 \rangle = \frac{1}{Kq^2}$$

or in real space by

$$\langle A_i(\vec{r}) A_j(\vec{0}) \rangle = \frac{\delta_{ij}}{4\pi K r}$$

This leads to the correlations of the $\vec{B}$ field

$$\langle B_i(\vec{r}) B_j(\vec{0}) \rangle = \frac{1}{4\pi K} \frac{3x_i x_j - \delta_{ij} r^2}{r^5}$$

and finally to the spin-spin correlations

$$\langle S_\alpha(\vec{r}) S_\beta(\vec{0}) \rangle = \frac{1}{4\pi K} \frac{3(\hat{e}_\alpha \cdot \vec{r})(\hat{e}_\beta \cdot \vec{r}) - (\hat{e}_\alpha \cdot \hat{e}_\beta) r^2}{r^5} \quad (7)$$

where α and β keep track of the position of the spin in the unit cell.

The dipolar form of the correlations leads to a very specific signature, namely the presence of pinch points in diffuse scattering [20, 21].

4 Order by disorder

If the ground state of a classical Heisenberg model has no long-range magnetic order because of geometrical frustration, it does not necessarily imply that this remains true at $T > 0$ or for the quantum version of the model [22–24]. Indeed, the spectrum of fluctuations usually depends on the ground state, and this can lead to a selection mechanism known as *order by disorder* that can lead to long-range magnetic order.

4.1 Quantum fluctuations

Let us start by discussing the effect of quantum fluctuations. When spins are quantum operators, the fluctuations around a given ground state can be described at the harmonic level with the help of the Holstein-Primakoff transformation:

$$\begin{cases} S_i^{z_i} = S - a_i^\dagger a_i \\ S_i^+ = \sqrt{2S - a_i^\dagger a_i} a_i \\ S_i^- = a_i^\dagger \sqrt{2S - a_i^\dagger a_i} \end{cases} \quad (8)$$

where the quantization axis z_i is in the direction of the spin at site i in the ground state under consideration. Keeping only terms of order S^2 and S leads to a quadratic Hamiltonian. If the environment of all sites is the same up to a rotation (a condition that is often met even if the ground state is not periodic, as for instance in helical states on a Bravais lattice), the Hamiltonian takes the general form in Fourier space:

$$H = E_{\text{classical}} + \sum_{\vec{k}} \left[B_{\vec{k}} a_{\vec{k}}^\dagger a_{\vec{k}} + \frac{1}{2} A_{\vec{k}} \left(a_{\vec{k}}^\dagger a_{-\vec{k}}^\dagger + a_{\vec{k}} a_{-\vec{k}} \right) \right]$$

where $E_{\text{classical}}$ is proportional to S^2 , and $A_{\vec{k}}$ and $B_{\vec{k}}$ are coefficients proportional to S that depend on the exchange integrals and on the ground state. This Hamiltonian can be put in diagonal form

$$\mathcal{H} = E_0 + \sum_{\vec{k}} \omega_{\vec{k}} \left(\alpha_{\vec{k}}^\dagger \alpha_{\vec{k}} + \frac{1}{2} \right)$$

where the operators $\alpha_{\vec{k}}^\dagger$ and $\alpha_{\vec{k}}$ are related to the Holstein-Primakoff operators by a Bogoliubov transformation:

$$\alpha_{\vec{k}} = u_{\vec{k}} a_{\vec{k}} + v_{\vec{k}} a_{-\vec{k}}^\dagger$$

with

$$\omega_{\vec{k}} = \sqrt{B_{\vec{k}}^2 - A_{\vec{k}}^2}, \quad u_{\vec{k}} = \sqrt{\frac{B_{\vec{k}} + \omega_{\vec{k}}}{2\omega_{\vec{k}}}}, \quad v_{\vec{k}} = \text{sign}(A_{\vec{k}}) \sqrt{\frac{B_{\vec{k}} - \omega_{\vec{k}}}{2\omega_{\vec{k}}}}$$

In Eq. (4.1), the energy E_0 is the sum of the classical energy, which is independent of the ground state, and of a quantum correction also independent of the ground state. For antiferromagnetic bonds, its expression is the same as that of the classical energy up to the replacement $S^2 \rightarrow S(S+1)$.

Then, the energy of the quantum state obtained by 'dressing' a classical state with quantum fluctuations includes a zero-point contribution and is given by

$$E(\theta) = E_0 + \frac{1}{2} \sum_{\vec{k}} \omega_{\vec{k}}^2(\theta) \quad (9)$$

In that expression, θ stands for a parameter (or a set of parameters) that runs over the degenerate ground-state manifold, and we have made it explicit that the right-hand side depends on θ through $\omega_{\vec{k}}$, which is the case because the coefficients $A_{\vec{k}}$ and $B_{\vec{k}}$ depend on the ground state. If the minimization of this energy with respect to θ selects a value of θ for which the classical ground state has long-range order, then, at this level of approximation, the quantum system possesses this type of order.

As an example, let us consider again the J_1 - J_2 model on the square lattice for $J_2/J_1 > 1/2$. The ground state is infinitely degenerate, the energy being independent of the angle θ between the two Néel sublattices (see left panel of Fig. 3). Although there are four sites per unit cell, thus in principle 4 branches of excitations, the Hamiltonian describing harmonic fluctuations has the periodicity of the square lattice, and the spectrum is simply given by $\omega_{\vec{k}} = \sqrt{B_{\vec{k}}^2 - A_{\vec{k}}^2}$ with

$$A_{\vec{k}}/S = -2J_1 \left(\cos^2 \frac{\theta}{2} \cos k_x + \sin^2 \frac{\theta}{2} \cos k_y \right) - 4J_2 \cos k_x \cos k_y$$

and

$$B_{\vec{k}}/S = 2J_1 \left(\sin^2 \frac{\theta}{2} \cos k_x + \cos^2 \frac{\theta}{2} \cos k_y \right) + 4J_2$$

The zero-point energy is minimal for $\theta = 0$ and $\theta = \pi$, which corresponds to the two helical ground states of wave vector $(0, \pi)$ and $(\pi, 0)$. Remarkably, these are the collinear states of the ground-state manifold. This is a general trend. Whenever possible, coplanar states are favored over non-coplanar ones and collinear states over non-collinear ones because their spectrum is softer, and their zero-point energy lower [24].

The presence of two ground states and not simply one has an interesting consequence: these states can be distinguished by an Ising order parameter defined on the dual lattice,

$$\sigma_x = (\vec{S}_i - \vec{S}_k) \cdot (\vec{S}_j - \vec{S}_l) / \left| (\vec{S}_i - \vec{S}_k) \cdot (\vec{S}_j - \vec{S}_l) \right| \quad (10)$$

where (i, j, k, l) are the corners with diagonal (i, k) and (j, l) of the plaquette centered at the site x of the dual lattice. As a consequence, there is an Ising transition at finite temperature, a remarkable effect for a system with continuous symmetry [25, 26].

This harmonic selection process does not always lead to long-range order, however. For instance, for the kagome lattice, the coplanar ground states are favored, in agreement with the general trend, but the zero-point energy is exactly the same for all coplanar ground states [27]. One would have to push the expansion to higher order in $1/S$ to select among these states.

4.2 Thermal fluctuations

For a purely classical model, thermal fluctuations have a similar effect. Deviations from a classical ground state can be parametrized by two local coordinates according to

$$\vec{S}_i = \left(x_i, y_i, \sqrt{1 - x_i^2 - y_i^2} \right) \quad (11)$$

in the local reference frame of the ground state ($x_i = y_i = 0$ in the ground state). Expanding the energy to second order in x_i and y_i leads to a quadratic form for the energy. This allows the calculation of the partition function as a Gaussian integral. Although the coefficients of the quadratic form are not equal to the coefficients $A_{\vec{k}}$ and $B_{\vec{k}}$ of the quantum case, they are related to them, and the low temperature free energy can be expressed in terms of the frequencies of the quantum case as ($k_B = 1$):

$$F = F_0 - \frac{1}{2} N_h T \ln T + T \sum_{\vec{k}} \ln \omega_{\vec{k}} \quad (12)$$

where F_0 is independent of the ground state, and N_h is the number of harmonic modes, i.e., of non-zero frequencies. In general, the state that minimizes $\sum_{\vec{k}} \ln \omega_{\vec{k}}$ is selected [24]. This is often the same state as the one selected by quantum fluctuations, but this does not need to be the case since quantum fluctuations minimize another function of the frequencies, namely their sum $\sum_{\vec{k}} \omega_{\vec{k}}$. Besides, zero frequencies play a different role. For the quantum case, they just give a vanishing contribution to the zero point energy, but for the classical case, they lead to an integral that is not Gaussian. If the next term in the expansion of the non-quadratic modes is quartic, which is generally the case, the low-temperature free energy takes the form

$$F = F_0 - \frac{1}{2} N_h T \ln T - \frac{1}{4} N_q T \ln T + \dots \quad (13)$$

where N_q is the number of quartic modes. The factor $\frac{1}{2}$ is replaced by $\frac{1}{4}$ if the mode is quartic. As a consequence, the state with the maximal number of zero modes will always be favored by thermal fluctuations [27], whereas this is not necessarily the case for quantum fluctuations.

5 Alternatives to magnetic long-range order in Heisenberg models

The presence of classical degeneracy in the ground state of a frustrated Heisenberg model typically leads to the presence of additional zero-frequency spin-wave modes on top of those that can be expected on general grounds (at $\vec{q} = \vec{0}$ and at $\vec{q} = \pm \vec{Q}$ in a helical state with pitch vector $\vec{Q}$). This degeneracy often leads to lines of zero modes (e.g. J_1 - J_2 model at $J_2 = J_1/2$) or even to a plane of zero modes (e.g. kagome) in the Brillouin zone, which, to harmonic order, implies a divergence of the correction to the local magnetization

$$\delta_m \equiv S - \langle S_i^z \rangle = \frac{1}{N} \sum_{\vec{k}} \langle a_{\vec{k}}^\dagger a_{\vec{k}} \rangle \quad (14)$$

since, in the ground state (the vacuum of the Bogoliubov α particles), $\langle a_{\vec{k}}^\dagger a_{\vec{k}} \rangle = v_{\vec{k}}^2 \propto 1/\omega_{\vec{k}}$. The spin-1/2 J_1 - J_2 model is believed to be disordered around $J_2 = J_1/2$ for that reason [28]. If the system is to maintain long-range order, higher-order corrections will have to open a gap at these accidental zero modes to restore a finite value of the corrections. Self-consistent calculations have been done that confirm this scenario, but for the calculation to be fully consistent, the corrections must still be smaller than the spin S , which is often not the case, at least for small enough spin. This implies that one has to look for alternatives to magnetic long-range order. One possibility is to break the $SU(2)$ symmetry with an order parameter that is not the local spin, in which case one speaks of a spin nematic (see below). When the $SU(2)$ symmetry is not broken, the resulting ground state is generally called a *spin liquid* [29]. The rest of this chapter reviews some of these possibilities.

5.1 Spin gap

A simple alternative to magnetic long-range order consists in opening a spin gap by adopting a ground-state configuration in which spins are paired up to make local singlets rather than developing long-range order. In that respect, there is an important difference between systems with half-integer or integer spin per unit cell:

- If the total spin per unit cell is integral, then singlet pairing can be achieved without any spatial symmetry breaking, and the ground state is non-degenerate. This is in particular the case of integer spin chains [30] and of spin ladders [31]. In spin ladders, there are two sites per unit cell, and the ground state is adiabatically connected to that of the strong rung limit, which simply consists of a product of singlets on the rungs. In integer spin chains, the simple picture relies on representing the local spin as a set of $2S$ spins 1/2, the ground state being adiabatically connected to the product of bond singlets constructed out of S spins 1/2 at each end of the bond [32]. In 2D, long-range antiferromagnetic order is generally realized in the absence of frustration (square lattice, honeycomb lattice) unless the system consists of weakly coupled dimers. With frustration however, the dimers do not need to be weakly coupled for this to be realized, as in the case of the Shastry-Sutherland model [33] (see below).
- If the total spin per unit cell is half-integral, singlet pairing can only be achieved through a spontaneous symmetry breaking that enlarges the unit cell to accommodate an integer total spin. The first and most famous example is the spin-1/2 J_1 - J_2 Heisenberg chain [34], which has in fact inspired the construction of the Shastry-Sutherland model.

The Majumdar-Ghosh model

The Hamiltonian of the spin-1/2 J_1 - J_2 Heisenberg chain is given by:

$$\mathcal{H}_{J_1-J_2} = \sum_i (J_1 \vec{S}_i \cdot \vec{S}_{i+1} + J_2 \vec{S}_i \cdot \vec{S}_{i+2}) \quad (15)$$

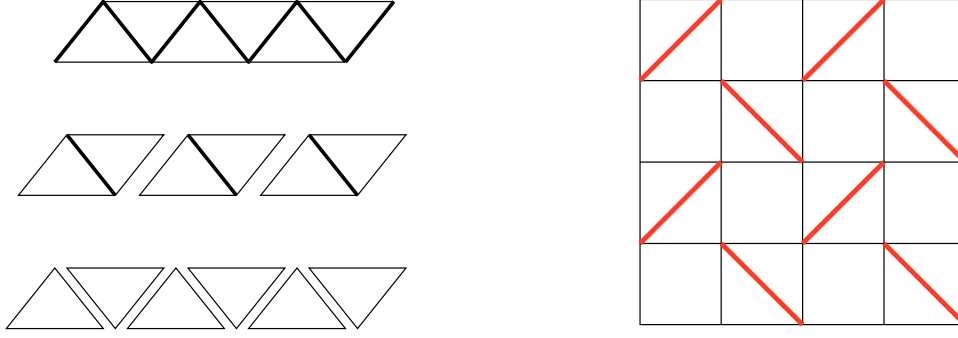


Fig. 5: Left panel: from top to bottom, sketch of the J_1 - J_2 chain, of its decomposition in terms of diamonds, and of its decomposition in terms of triangles. Thick solid lines stand for bonds of strength $J_1 = 2J_2$, thin solid lines for bonds of strength J_2 . Right panel: Shastry-Sutherland lattice.

At the Majumdar-Ghosh point $J_2 = J_1/2$, the Hamiltonian can be rewritten (see middle left panel of Fig. 5)

$$\mathcal{H}_{J_1-J_2} = \sum_{i \text{ odd}} \left(J_1 \vec{S}_i \cdot \vec{S}_{i+1} + \frac{J_1}{2} (\vec{S}_i + \vec{S}_{i+1}) \cdot \vec{S}_{i-1} + \frac{J_1}{2} (\vec{S}_i + \vec{S}_{i+1}) \cdot \vec{S}_{i+2} \right)$$

With this rewriting, it is clear that the wave function

$$|\psi_{\text{odd}}\rangle = \prod_{i \text{ odd}} |S(i, i+1)\rangle, \quad |S(i, i+1)\rangle = \text{singlet}$$

is an eigenstate of $\mathcal{H}_{J_1-J_2}$ of energy $E_{\text{odd}} = -(N/2)(3/4) J_1$, where N is the number of sites, since

$$(S_i^\alpha + S_{i+1}^\alpha) |S(i, i+1)\rangle = 0, \quad \alpha = x, y, z.$$

Now, the Hamiltonian can also be rewritten (see bottom left panel of Fig. 5)

$$\mathcal{H}_{J_1-J_2} = \sum_i \frac{J_1}{2} (\vec{S}_i \cdot \vec{S}_{i+1} + \vec{S}_i \cdot \vec{S}_{i+2} + \vec{S}_{i+1} \cdot \vec{S}_{i+2}) = \sum_i \mathcal{H}_\Delta(i).$$

Since the ground-state energy of a triangle of coupling $J_1/2$ is given by

$$E_0(\mathcal{H}_\Delta(i)) = -\frac{3}{4} \frac{J_1}{2}$$

the ground state energy of $\mathcal{H}_{J_1-J_2}$ satisfies the inequality

$$E_0(\mathcal{H}_{J_1-J_2}) \geq \sum_i E_0(\mathcal{H}_\Delta(i)) = -\frac{3}{4} \frac{J_1}{2} N = E_{\text{odd}}$$

which, by the variational principle, proves that $|\psi_{\text{odd}}\rangle$ is a ground state. The same reasoning can be done for

$$|\psi_{\text{even}}\rangle = \prod_{i \text{ even}} |S(i, i+1)\rangle$$

so that the ground state is two-fold degenerate.

More recently, this calculation has been generalized to arbitrary spin [35]. The trick is to consider the following generalization of the Hamiltonian

$$\mathcal{H}_{J_1-J_3} = J_1 \sum_i \vec{S}_i \cdot \vec{S}_{i+1} + J_3 \sum_i \left[(\vec{S}_{i-1} \cdot \vec{S}_i)(\vec{S}_i \cdot \vec{S}_{i+1}) + \text{h.c.} \right] \quad (16)$$

This Hamiltonian reduces to the J_1 - J_2 model with $J_2 = J_3/2$ for spin 1/2 but not for larger spin. It has been shown that the fully dimerized states are degenerate ground states for $J_3/J_1 = 1/(4S(S+1) - 2)$.

The Shastry-Sutherland model

In an attempt to find higher-dimensional analogs of the Majumdar-Ghosh chain, Shastry and Sutherland came across a very interesting 2D model known as the Shastry-Sutherland model [33] and depicted in the right panel of Fig. 5. This model can be seen as an orthogonal dimer model or as a square lattice with some diagonal couplings. Its Hamiltonian can be written as

$$\mathcal{H}_{\text{Shastry-Sutherland}} = \sum_{\langle\langle i,j \rangle\rangle} J \vec{S}_i \cdot \vec{S}_j + \sum_{\langle i,j \rangle} J' \vec{S}_i \cdot \vec{S}_j \quad (17)$$

where $\langle\langle i,j \rangle\rangle$ stands for the diagonal bonds, and $\langle i,j \rangle$ for the bonds of the square lattice. The same type of argument as the first step of the proof of the Majumdar-Ghosh model shows that the product of singlets on J bonds is always an eigenstate, and variational arguments show that this will be the case as long as J'/J is not too large, less than 0.675 for spin-1/2 according to the most recent numerical estimate [36]. Remarkably enough, this model provides a very accurate description of the layered Cu oxide $\text{SrCu}_2(\text{BO}_3)_2$, a system in which the frustration not only opens a gap, but is also at the origin of a sequence of magnetization plateaus at 1/8, 2/15, 1/6, 1/4, 1/3, and 1/2 [37, 38].

5.2 Resonating Valence Bond spin liquids

For a spin-1/2 antiferromagnet with an odd number of spins per unit cell, as for instance the square and triangular lattices (1 spin per unit cell) and the kagome lattice (3 spins per unit cell), a product of singlet dimers has to break the spatial symmetry. In 1973, Anderson suggested that the spatial symmetry might be restored if the ground state is a superposition of all possible singlet dimer coverings that he called a Resonating Valence Bond (RVB) state (see Fig. 6), and he suggested that this might take place for the triangular lattice [39, 40]. It is now widely believed that the spin-1/2 Heisenberg model on the triangular lattice has 3-sublattice long-range order [41], but the RVB state remains a strong candidate for the spin-1/2 kagome antiferromagnet.

Proving directly that a RVB state is realized for a given spin-1/2 model is very challenging because it requires performing numerical simulations on very large clusters, and quantum Monte Carlo simulations cannot be performed for frustrated magnets because of the minus sign problem. However, the possibility of realizing an RVB state has been demonstrated in effective

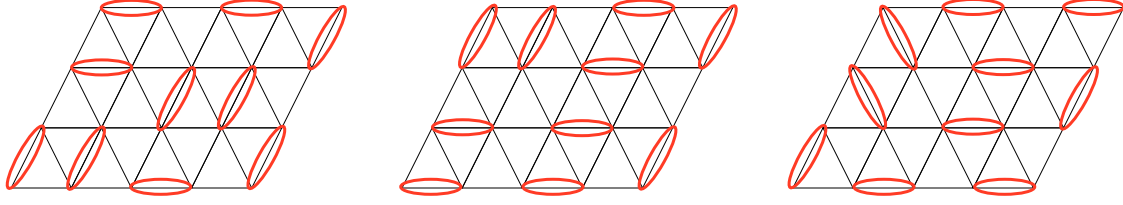


Fig. 6: Sketch of the Resonating Valence Bond state on the triangular lattice. Red ellipses stand for singlet dimers for the Heisenberg model and simply for dimers for the ground state of the Quantum Dimer Model on the triangular lattice at the Rokhsar-Kivelson point.

models of the singlet sector known as Quantum Dimer Models (QDM). Let us thus start with a brief review of their properties.

The Hilbert space of a QDM is defined by all nearest-neighbor dimer coverings of a lattice, which are assumed to be orthogonal (note that this would not be the case for spin-1/2 singlets). The Hamiltonian typically contains kinetic terms that flip dimers around plaquettes and potential terms proportional to the number of flippable plaquettes. The simplest model on a square lattice can be written as:

$$H_{\text{QDM}} = \sum_{\text{Plaquettes}} [-J (|\begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}\rangle \langle \begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}| + h.c.) + v (|\begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}\rangle \langle \begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}| + |\begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}\rangle \langle \begin{smallmatrix} \vdots \\ \vdots \end{smallmatrix}|)] \quad (18)$$

Rokhsar and Kivelson [42] asked the question whether, for some values of v/J , such a model can sustain a resonating valence bond (RVB) phase. As a first step toward an answer to this question, they proved the following theorem:

Theorem: For $v = J$ (RK-point), the sum of all configurations with equal weight is a ground state.

Proof: Consider $|\psi\rangle_{\text{RK}} = \sum_c |c\rangle$, where $|c\rangle$ is a dimer configuration. Let us write $H_{\text{QDM}} = \sum_i H_{\text{QDM}}^i$, where the sum over i runs over plaquettes, and consider a plaquette i . Then one can distinguish two types of configurations:

- c does not contain parallel dimers on plaquette i (plaquette i is not flippable in c). Then $H_{\text{QDM}}^i |c\rangle = 0$.
- c does contain parallel dimers. Then there is a companion configuration c' obtained from c by changing the orientation of dimers on plaquette i such that

$$\begin{aligned} H_{\text{QDM}}^i |c\rangle &= -J |c'\rangle + v |c\rangle \\ H_{\text{QDM}}^i |c'\rangle &= -J |c\rangle + v |c'\rangle \\ \Rightarrow H_{\text{QDM}}^i (|c\rangle + |c'\rangle) &= (v - J) (|c'\rangle + |c\rangle) \end{aligned}$$

Then, if $v = J$, $H_{\text{QDM}}^i (|c\rangle + |c'\rangle) = 0$. This implies that

$$H_{\text{QDM}}^i \sum_c |c\rangle = 0, \quad \forall i$$

and finally that

$$H_{\text{QDM}} |\psi\rangle_{\text{RK}} = 0.$$

So $|\psi\rangle_{\text{RK}}$ is an eigenstate with eigenvalue 0.

Let us now prove that this is the ground state energy. For that purpose, let us enlarge the Hilbert space to allow for empty sites (sites without dimers). Let us define the operator $a_{\uparrow\downarrow}$ as the operator that destroys $\uparrow\downarrow$ on a given plaquette provided there are such dimers on a given configuration, and that gives 0 otherwise, and let us denote the adjoint operator by $a_{\uparrow\downarrow}^\dagger$. Then the Hamiltonian

$$H = -J \sum_{\text{Plaquettes}} \left(a_{\uparrow\downarrow}^\dagger a_{\uparrow\downarrow} + a_{\uparrow\downarrow}^\dagger a_{\uparrow\downarrow} \right) + v \sum_{\text{Plaquettes}} \left(a_{\uparrow\downarrow}^\dagger a_{\uparrow\downarrow} + a_{\uparrow\downarrow}^\dagger a_{\uparrow\downarrow} \right)$$

is equal to H_{QDM} when restricted to the subspace where all sites belong to a dimer. So, the ground state energy of H_{QDM} must be larger or equal to the ground state of H . Now, for $v = J$, one can write

$$H = v \sum_{\text{Plaquettes}} \left(a_{\uparrow\downarrow}^\dagger - a_{\uparrow\downarrow} \right) \left(a_{\uparrow\downarrow} - a_{\uparrow\downarrow}^\dagger \right) = \sum_i A_i^\dagger A_i$$

with $A_i = a_{\uparrow\downarrow} - a_{\uparrow\downarrow}^\dagger$, and the sum over i runs over all the plaquettes. Then,

$$\langle \psi | H | \psi \rangle = v \sum_i \langle \psi | A_i^\dagger A_i | \psi \rangle = v \sum_i \|A_i \psi\|^2 \geq 0$$

This implies that the ground state energy of H is non-negative, hence that the ground state energy of H_{QDM} is non-negative. Since the sum of all dimer configurations is a zero energy eigenstate, it is thus a ground state of H_{QDM} .

Does this imply that the QDM on the square lattice at the RK point is an RVB liquid? Not quite, for several reasons. First of all, there are other ground states. Indeed all configurations with no flippable plaquette are zero energy eigenstates, and there are many of them. This is not a final blow, however, because their energy depends on the value of v/J in different ways. For $v > J$, these non-flippable configurations remain zero energy eigenstates and can be expected to be the only ground states, whereas for $v < J$, the state emanating from $|\psi\rangle_{\text{RK}}$ can be expected to be lower in energy than the non-flippable states.

More importantly, this RVB state can only be expected to extend into an RVB phase for $v < J$ if the spectrum is gapped at the RK-point, i.e., if correlations decay exponentially. Now, the dimer-dimer correlations at the RK ground state are exactly the same as the classical dimer-dimer correlations averaged over all dimer coverings considered in Section 3. The decay is itself related to the property of the Kasteleyn matrix: If all eigenvalues stay at a finite distance from 0, dimer-dimer correlations decay exponentially. Otherwise, they decay algebraically [43, 44]. For the square lattice, it is easy to show that, as for the honeycomb lattice, the spectrum of the Kasteleyn matrix has no gap. As a consequence, the correlations decay algebraically, the

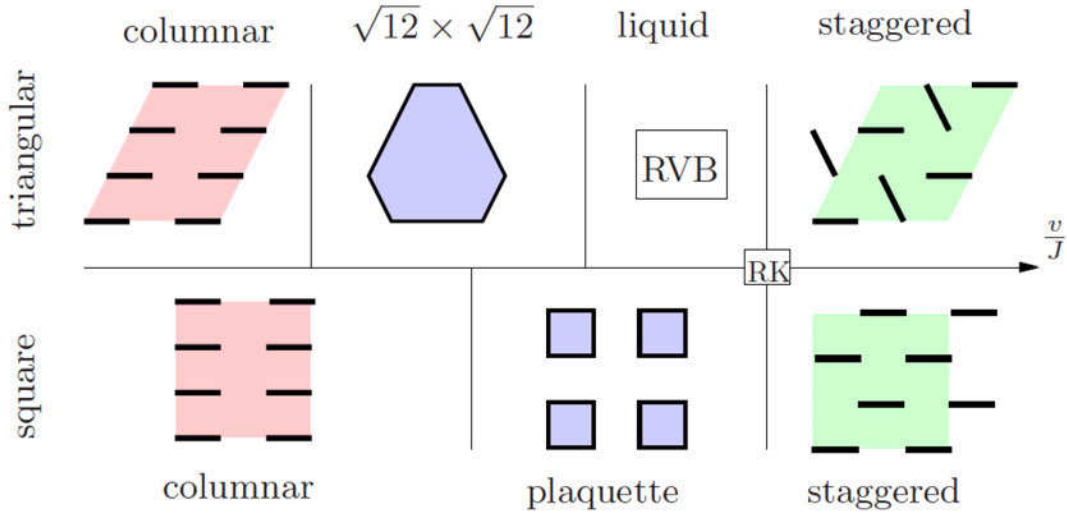


Fig. 7: Sketch of the phase diagram of the Rokhsar-Kivelson model on the triangular lattice (top) and on the square lattice (bottom). The staggered phases are degenerate, the configurations shown being the most symmetric ones.

spectrum of H_{QDM} has no gap, and there is no RVB phase, a result confirmed by numerical simulations. By contrast, for the triangular lattice, the Kasteleyn matrix is gapped, and the correlations decay exponentially. The spectrum is thus gapped, and one can expect to stabilize an RVB phase for $v < J$, in agreement with numerical results [45, 46] (see Fig. 7).

This state, although it is gapped, differs from a simple gapped state described, e.g., by a product of singlet dimers by non-trivial topological properties:

- The ground state has a degeneracy that depends on the topology of the lattice (non degenerate on a finite lattice with open boundary conditions, two-fold degenerate on a cylinder, four-fold degenerate on a torus,...).
- The elementary excitations are fractionalized. In a finite system, they can only be created by pairs and consist in multiplying the configurations in the sum by $(-1)^{n_c}$, where n_c is the number of dimers cut along a line going from i to j , where i and j refer to the location of the elementary excitations called 'visons' [47].

Apart from spin models specifically constructed to have this property, the main candidate for an RVB spin liquid in quantum antiferromagnets is the spin-1/2 Heisenberg model on the kagome lattice, with strong numerical evidence from density-matrix renormalization group calculations of a gap in both the singlet and triplet sectors [48]. The possibility of describing the singlet sector of this RVB liquid with the help of a Quantum Dimer Model is still debated. According to the first attempt at deriving an effective QDM, the system should be in a crystalline phase that breaks translational symmetry [49]. However, the derivation of an effective QDM has been recently revisited, with the opposite conclusion that this effective model does indeed have an RVB ground state [50].

5.3 Algebraic spin liquids

For half-integer spin, spin chains are not gapped, and spin-spin correlations decay to zero algebraically, in contrast to integer spin chains, in which they decay exponentially. It is thus natural to consider the possibility of algebraic decay as well in 2D frustrated magnets. To implement this goal, the most convenient way is to use a fermionic representation of spin-1/2 operators sometimes known as Abrikosov fermions:

$$\begin{cases} S_i^+ = c_{i\uparrow}^\dagger c_{i\downarrow} \\ S_i^- = c_{i\downarrow}^\dagger c_{i\uparrow} \\ S_i^z = \frac{1}{2} (n_{i\uparrow} - n_{i\downarrow}) \end{cases} \quad (19)$$

with the constraint $n_{i\uparrow} + n_{i\downarrow} = 1$. In this representation, the Heisenberg model takes the form:

$$H = \frac{1}{2} \sum_{i,j} J_{ij} \left[\frac{1}{2} (c_{i\uparrow}^\dagger c_{i\downarrow} c_{j\downarrow}^\dagger c_{j\uparrow} + \text{h.c.}) + \frac{1}{4} (c_{i\uparrow}^\dagger c_{i\uparrow} - c_{i\downarrow}^\dagger c_{i\downarrow}) (c_{j\uparrow}^\dagger c_{j\uparrow} - c_{j\downarrow}^\dagger c_{j\downarrow}) \right] \quad (20)$$

This Hamiltonian is not quadratic, and the only simple (but of course approximate) solution relies on performing a mean-field decoupling of the four-fermion operators, and on treating the constraint on average with a Lagrange parameter. To describe algebraic liquids, it is convenient to introduce the bond operator

$$\chi_{ij} = c_{i\uparrow}^\dagger c_{j\uparrow} + c_{i\downarrow}^\dagger c_{j\downarrow} \quad (21)$$

which satisfies the identity

$$\vec{S}_i \cdot \vec{S}_j = \frac{1}{4} - \frac{1}{2} \chi_{ij}^\dagger \chi_{ij} \quad (22)$$

and to decouple the Hamiltonian by introducing the order parameter $\chi_{ij}^0 = \langle \chi_{ij} \rangle$.

Following Affleck and Marston in their seminal work [51], let us consider the square lattice. As is often the case with mean-field theory, the self-consistent equations possess several solutions. Let us concentrate on two of them:

- $\chi_{ij}^0 = \chi_0$ on a set of bonds that constitute a dimer covering of the lattice, and $\chi_{ij}^0 = 0$ on the other bonds. This solution describes a spontaneously dimerized state.
- $\chi_{ij}^0 = \chi_0 e^{i\theta_{ij}}$, where the phases θ_{ij} are chosen in such a way that they lead to a π -flux per plaquette. A possible choice of gauge is given by $\theta_{ij} = \pi/4$ if the arrow goes from i to j in Fig. 8 (left panel). This solution is called the π flux state. The dispersion of the fermionic spectrum is given by:

$$E = \pm J \chi_0 \sqrt{\cos^2 k_x + \cos^2 k_y} \quad (23)$$

This spectrum is a Dirac spectrum, with 4 Dirac points. Since the system must be half-filled to satisfy the constraint $n_{i\uparrow} + n_{i\downarrow} = 1$, the Fermi energy is given by $E_F = 0$, and the Fermi surface consists of 4 points: $k_x, k_y = \pm\pi/2$.

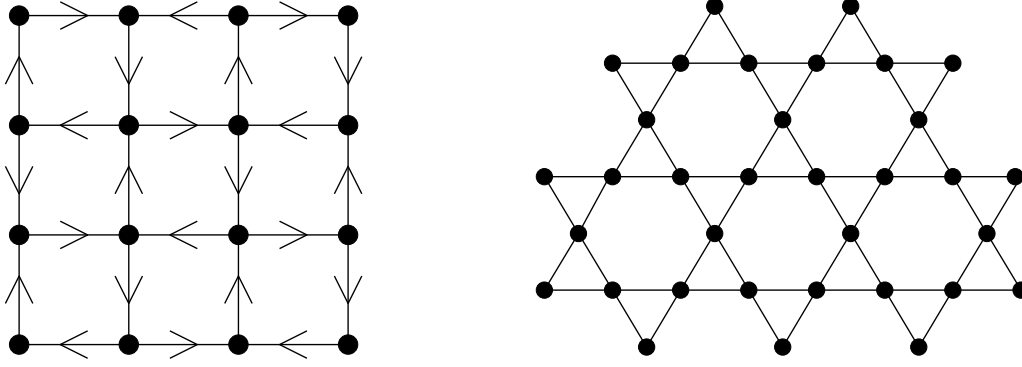


Fig. 8: *Left panel: Affleck-Marston π -flux phase on the square lattice. The phase of the hopping integral is equal to $\pi/4$ on each bond in the direction of the arrow. Right panel: flux pattern of the algebraic spin liquid proposed as a ground state for the spin-1/2 Heisenberg model on the kagome lattice.*

In this mean-field ground state, the spin-spin correlations decay algebraically, which gives rise to power laws at low temperature. It has been argued that, because of the Dirac nature of the spectrum, this mean-field solution is locally stable [52], and that the state described by this solution could be considered as a consistent theory of an algebraic spin liquid. Note that the order parameter χ_{ij} is invariant under the gauge transformation $c_{i\uparrow}^\dagger \rightarrow e^{i\theta} c_{i\uparrow}^\dagger$, where θ can take any value. There is thus a $U(1)$ gauge symmetry in the problem, and this type of spin liquid is sometimes referred to as a $U(1)$ spin liquid. The order parameter breaks the original $SU(2)$ symmetry, however, and different order parameters can correspond to the same physical solutions. A proper classification of the solutions relies on the projective symmetry group introduced by Wen [53].

To compare the various solutions, one can of course compare the mean-field energies. However, these energies are not variational because the local constraint is only satisfied on average, and the wave functions obtained within mean-field theory contain a lot of configurations with doubly occupied sites. A numerically tractable way to go beyond the mean-field approximation consists of projecting into the sector with no doubly occupied sites using the Gutzwiller projector

$$P_G = \prod_i (1 - n_{i\uparrow} n_{i\downarrow}) \quad (24)$$

The energy of the state $P_G|\psi_{\text{MF}}\rangle$, where $|\psi_{\text{MF}}\rangle$ is a mean-field solution, can be evaluated with a Monte Carlo simulation [54]. In fact, one can generalize the method and use as a variational subspace a set of functions of the form $P_G|\psi_{\text{fermion}}\rangle$, where $|\psi_{\text{fermion}}\rangle$ is the ground state of quadratic fermionic Hamiltonians not necessarily related to a mean-field solution of the original problem. This method goes by the name of variational Monte Carlo (VMC) [55].

For the square lattice, the best VMC state is a flux state with a uniform flux less than π per plaquette. This is clearly not the true ground state, which has long-range Néel order. However, the VMC approach, as a way to look for alternatives to long-range magnetic order, is very versatile since it allows the comparison of several types of quantum liquids (including the RVB

states discussed in the previous section and the chiral states discussed in the next one). It has been implemented for the spin-1/2 kagome antiferromagnet, with the conclusion that the lowest-energy state is the one depicted in the right panel of Fig. 8, with a flux π per hexagon and no flux through the triangles [56]. The resulting state is an algebraic spin liquid, and it is one of the main candidates for the ground state of the spin-1/2 kagome antiferromagnet.

Note that one can also introduce a pairing operator $\eta_{ij} = c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger - c_{i\downarrow}^\dagger c_{j\uparrow}^\dagger$ to decouple the Hamiltonian. This leads to a mean-field theory of the RVB state discussed in the previous section. This order parameter is only invariant upon the transformation $c_{i\uparrow}^\dagger \rightarrow e^{i\theta} c_{i\uparrow}^\dagger$ with $\theta = \pi$, i.e., $c_{i\uparrow}^\dagger \rightarrow -c_{i\uparrow}^\dagger$. There is thus only a Z_2 gauge symmetry, and these states are sometimes referred to as Z_2 spin liquids.

5.4 Chiral spin liquids

All the states discussed so far as alternatives to magnetic long-range order leave the time-reversal symmetry unbroken. It has been suggested however that another family of quantum spin liquids related to Fractional Quantum Hall (FQH) states might be stabilized by frustration [57]. These states break time-reversal and parity symmetry, but not their product. A possible order parameter is the mixed product $\vec{S}_1 \cdot (\vec{S}_2 \times \vec{S}_3)$, also called the scalar chirality [58].

Although the symmetry-based definition is the most fundamental one, the discussion of the properties of chiral spin liquids is most conveniently done using the language of Gutzwiller projected wave-functions. In that language, a chiral state is obtained by applying a Gutzwiller projection to a fermionic state that is the ground state of a Hamiltonian with a fractional flux through some plaquettes, i.e., a flux that is neither equal to 0 nor to π , so that time reversal symmetry is broken [59]. In the mean-field language, the order parameter is of the χ_{ij} type, but it has to break translational symmetry to accommodate a fractional flux per plaquette. The resulting band structure is gapped, and the effective low-energy theory is a pure gauge theory. As a consequence, the ground state is expected to have a topological degeneracy equal to twice that of the corresponding FQH state, the extra factor of 2 coming from the broken time-reversal symmetry. The elementary excitations are anyons, with fractional statistics.

The best evidences so far in favor of such ground states in frustrated magnets have been obtained for extensions of the Heisenberg model on the kagome lattice, possibly a consequence of the fact that, when further-neighbor couplings are introduced, the classical kagome antiferromagnet has non-coplanar ground states [60]. In particular, when second- and third-neighbor interactions are present and for specific values of the couplings the best VMC state is a flux state with a complicated flux pattern. This chiral state has been confirmed by numerical results, which have reported a set of four low-lying states, in agreement with the expected 4-fold degenerate ground state in the thermodynamic limit [61].

Another model consists of a generalization of the Heisenberg model that explicitly breaks the chiral symmetry by introducing a term of the form $\vec{S}_i \cdot (\vec{S}_j \times \vec{S}_k)$ on each plaquette. Since the chiral symmetry is explicitly broken, one expects the ground state to be only two-fold degenerate, in agreement with DMRG results [62].

Note that the situation is quite different in 3D, where the gauge theory is equivalent to electromagnetism, and where one can expect to have a linearly dispersive mode, sometimes referred to as a photon [63].

5.5 Nematic order

Another completely different way for a system to avoid magnetic ordering is to break the $SU(2)$ symmetry without developing long-range magnetic order. This requires that the operator that develops long-range correlations is not simply the local spin operator but a more complicated local operator, very similar to liquid crystals, which breaks spatial symmetries but does not have standard crystalline order. By order of increasing complexity, the next possibility is an operator constructed out of two spin operators [64]. Let us thus consider two spins $\vec{S}_i$ and $\vec{S}_j$. By combining them, one can construct *a priori* nine operators:

$$S_i^\alpha S_j^\beta, \quad \alpha, \beta = x, y, z \quad (25)$$

Out of these nine operators, one can construct a scalar, a vector, and a rank 2 tensor:

- $\vec{S}_i \cdot \vec{S}_j$: This is a scalar, and it cannot break the $SU(2)$ symmetry.
- $\vec{S}_i \times \vec{S}_j$: This operator has three components and is a vector. If such an operator develops long-range correlations, the system is said to be *p-nematic*.
- The last five independent operators constitute a rank-2 tensor. They can be conveniently arranged in a five-component vector:

$$\vec{Q}_{ij} = \begin{pmatrix} S_i^x S_j^x - S_i^y S_j^y \\ \frac{1}{\sqrt{3}} (3S_i^z S_j^z - \vec{S}_i \cdot \vec{S}_j) \\ S_i^x S_j^y + S_i^y S_j^x \\ S_i^y S_j^z + S_i^z S_j^y \\ S_i^z S_j^x + S_i^x S_j^z \end{pmatrix} \quad (26)$$

If this operator develops long-range order, the system is said to be *n-nematic*

Now, let us briefly see which possibilities can be realized depending on the value of the spin, and on whether the sites i and j are equal or different.

- *Classical spin systems*: A purely local order parameter of the p-nematic type cannot be realized because $\vec{S}_i \times \vec{S}_i = \vec{0}$. However, a purely local n-nematic order parameter $\vec{Q}_i \equiv \vec{Q}_{ii}$, whose components involve products of components of the spin at site i , can be realized. If we relax the requirement of a purely local order parameter, then both the vector chirality $\vec{S}_i \times \vec{S}_j$ and the quadrupolar operator $\vec{Q}_{ij}$ can be realized.
- *S=1/2 systems*: For spins 1/2, there are no new purely local order parameters: $\vec{S}_i \times \vec{S}_i = i\vec{S}_i$ is proportional to the local spin, and $\vec{Q}_i$ vanishes identically because $(S_i^\alpha)^2 = 1/4$ and $S_i^\alpha S_i^\beta = -S_i^\beta S_i^\alpha$ for spin-1/2 operators. So the only possible nematic operators are bond operators: the vector chirality $\vec{S}_i \times \vec{S}_j$ and the bond quadrupolar operator $\vec{Q}_{ij}$.

- $S=1$ systems: The identity $\vec{S}_i \times \vec{S}_i = i\vec{S}_i$ still holds, so there is no p-nematic local order parameter, but, by contrast to the spin-1/2 case, the local quadrupolar operator $\vec{Q}_i$ does not vanish identically. For $i \neq j$, as for spin 1/2, both the bond vector chirality and the bond quadrupolar operators can be realized.
- $S>1$ systems: There is no difference from the spin-1 case regarding operators that involve two spins. There is a difference however when one considers more than 2 spins. For spin $S > 1$, purely local operators involving $2S > 2$ spins can be constructed, corresponding to octupolar order, etc.

Two cases have been discussed at length in the recent literature: purely local quadrupolar order in spin-1 antiferromagnets and bond n-nematic order in spin-1/2 antiferromagnets.

Local quadrupolar order parameter in spin-1 bilinear-biquadratic models

The possibility for a spin-1 system to develop purely quadrupolar order can be understood in very simple terms. Indeed, let us consider the S^z basis $\{|-1\rangle, |0\rangle, |1\rangle\}$ of a spin 1. The states $|-1\rangle$ and $|1\rangle$ are clearly magnetic and correspond to a spin pointing opposite to z or along z . By contrast, the state $|0\rangle$ is not magnetic. Indeed, one can easily check that it satisfies $\langle 0|S^\alpha|0\rangle = 0$ for $\alpha = x, y, z$. However, it is not invariant by rotation. Indeed, $\langle 0|(S^x)^2|0\rangle = \langle 0|(S^y)^2|0\rangle = 1$, while $\langle 0|(S^z)^2|0\rangle = 0$. This state is a quantum state that describes fluctuations perpendicular to z . Note that there is no sign attached to the direction perpendicular to which fluctuations take place, and this direction is called the *director* of the quadrupolar state. This director can point in any direction, and one can in fact define a purely quadrupolar, time-reversal invariant basis with directors along x, y and z :

$$|x\rangle = i \frac{|1\rangle - |\bar{1}\rangle}{\sqrt{2}}, \quad |y\rangle = \frac{|1\rangle + |\bar{1}\rangle}{\sqrt{2}}, \quad |z\rangle = -i|0\rangle \quad (27)$$

In this basis, a general state can be decomposed as

$$|\vec{d}\rangle = d_x |x\rangle + d_y |y\rangle + d_z |z\rangle$$

where $\vec{d}$ is a complex vector of norm 1. To keep track of the nature of the state (magnetic, quadrupolar, or mixed), it is convenient to parametrize $\vec{d}$ according to [65]:

$$\vec{d} = \vec{u} + i\vec{v}, \quad \vec{u} \text{ and } \vec{v} \text{ real}, \quad \|\vec{u}\|^2 + \|\vec{v}\|^2 = 1, \quad \vec{u} \cdot \vec{v} = 0$$

Then, $\langle \vec{d} | \vec{S} | \vec{d} \rangle = 2\vec{u} \times \vec{v}$. A state is purely magnetic if $\|\vec{u}\|^2 = \|\vec{v}\|^2 = \frac{1}{2}$, and it is purely quadrupolar if $\vec{u} = \vec{0}$ or $\vec{v} = \vec{0}$, with director along the non-zero vector.

Let us now consider the bilinear-biquadratic spin-1 model on the triangular lattice defined by the Hamiltonian

$$H = \sum_{\langle i,j \rangle} \left[J_{\text{bil}} \vec{S}_i \cdot \vec{S}_j + J_{\text{biq}} \left(\vec{S}_i \cdot \vec{S}_j \right)^2 \right] \quad (28)$$

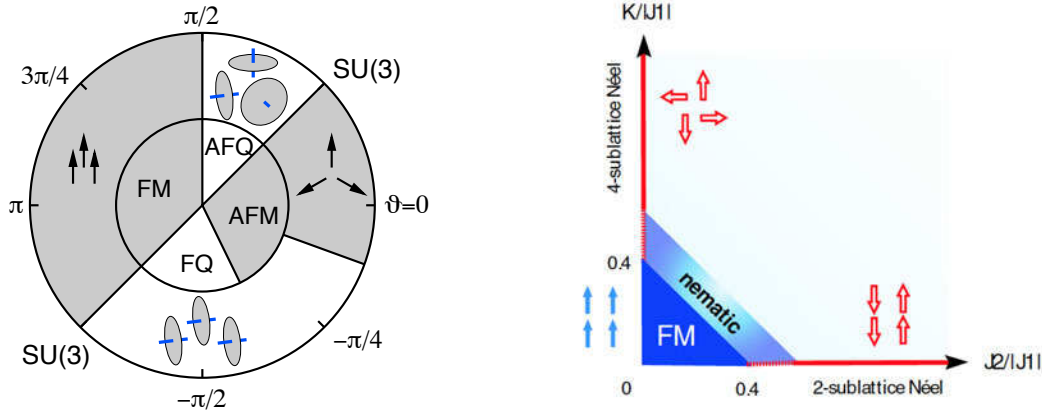


Fig. 9: Left panel: phase diagram of the bilinear-biquadratic spin-1 Heisenberg model on the triangular lattice. The angle θ keeps track of the relative magnitude of the bilinear coupling ($J_{\text{bil}} \propto \cos \theta$) to the biquadratic one ($J_{\text{biq}} \propto \sin \theta$). The inner circle is the mean-field phase diagram, the outer one the numerical one obtained with flavor-wave theory and exact diagonalization. There are two quadrupolar phases: a ferroquadrupolar phase for negative biquadratic coupling and a 3-sublattice antiferroquadrupolar one for positive biquadratic coupling. (After Läuchli et al. [67]) Right panel: sketch of the phase diagram of the spin-1/2 Heisenberg model on the square lattice with ferromagnetic nearest-neighbor coupling $J_1 < 0$, antiferromagnetic next-nearest neighbor coupling J_2 , and four-spin plaquette interaction K . There is a nematic phase close to the ferromagnetic one. (After Shannon et al. [69])

The presence of quadrupolar order with a local order parameter is plausible if biquadratic interactions are present because the Hamiltonian can be rewritten with the help of quadrupolar operators using the identity

$$J_{\text{bil}} \vec{S}_1 \cdot \vec{S}_2 + J_{\text{biq}} (\vec{S}_1 \cdot \vec{S}_2)^2 = \left(J_{\text{bil}} - \frac{J_{\text{biq}}}{2} \right) \vec{S}_1 \cdot \vec{S}_2 + \frac{J_{\text{biq}}}{2} \vec{Q}_1 \cdot \vec{Q}_2 + \frac{4}{3} J_{\text{biq}} \quad (29)$$

Then, the first step is to determine a mean-field phase diagram by looking for the product state that minimizes the energy as function of θ , with $J_{\text{bil}} = J \cos \theta$ and $J_{\text{biq}} = J \sin \theta$. This minimization is easily done using the following expectation values:

$$\begin{aligned} \langle \vec{d}_1, \vec{d}_2 | \vec{S}_1 \cdot \vec{S}_2 | \vec{d}_1, \vec{d}_2 \rangle &= |\vec{d}_1 \cdot \vec{d}_2^*|^2 - |\vec{d}_1 \cdot \vec{d}_2|^2 = 4(\vec{u}_1 \times \vec{v}_1) \cdot (\vec{u}_2 \times \vec{v}_2) \\ \langle \vec{d}_1, \vec{d}_2 | (\vec{S}_1 \cdot \vec{S}_2)^2 | \vec{d}_1, \vec{d}_2 \rangle &= 1 + |\vec{d}_1 \cdot \vec{d}_2|^2 = 1 + (\vec{u}_1 \cdot \vec{u}_2 - \vec{v}_1 \cdot \vec{v}_2)^2 + (\vec{u}_1 \cdot \vec{v}_2 + \vec{v}_1 \cdot \vec{u}_2)^2 \end{aligned}$$

The conclusion is that if J_{biq} is large and positive, directors tend to be perpendicular on neighboring sites, while if J_{biq} is large and negative, directors tend to be parallel. The resulting phase diagram on the triangular lattice is depicted in the left panel of Fig. 9. In addition to the standard ferromagnetic and 3-sublattice antiferromagnetic phase, it contains a ferroquadrupolar phase and a 3-sublattice antiferroquadrupolar phase. These simple results have been confirmed by more sophisticated calculations [66, 67].

Bond nematic phases

For spin-1/2 antiferromagnets, there is no purely local order parameter. However, bond operators can give rise to nematic order. An instability towards a p-nematic state was suggested early on for the square and kagome lattices with further-neighbor interactions in the context of a semiclassical analysis [68]. More recently, extensive numerical results for spin-1/2 have suggested that there is an instability towards an n-nematic phase for the J_1 - J_2 model on the square lattice for negative (ferromagnetic) J_1 and positive J_2 , close to the ferromagnetic phase [69]. The absence of a purely local order parameter prevents one from performing a simple analysis in terms of a product wave-function. In that case, it is the proximity of a ferromagnetic state that leads to a simple picture: The instability that is responsible for the disappearance of ferromagnetism is a condensation of two-magnon bound states, and the order parameter that describes the resulting order is a two-spin n-nematic operator (see right panel of Fig. 9). A similar effect has been predicted close to saturation.

6 Conclusion

It should be clear to the reader at this stage that frustrated magnetism has become a vast subject over the years, and although I have tried to cover (sometimes very briefly) several aspects of current interest, I am aware of the fact that entire subfields have been left aside. For instance, I have chosen to concentrate on the Ising and Heisenberg models, but frustration is also a source of fascinating phenomena for the XY model [70], with a number of specificities that I would not have been able to cover properly. For similar reasons, I have also decided to say very little about experimental results, but of course frustrated magnetism is to a large extent an experimental subject [71]. If, in spite of all its imperfections and shortcomings, this chapter manages to make some of the advanced theoretical concepts of frustrated magnetism accessible to non-specialists, it will have reached its main objective.

It is a pleasure to thank all the colleagues with whom I have had the privilege to interact on frustrated magnetism over the years, in particular M. Albrecht, V.I. Anisimov, F. Becca, C. Berthier, P. Carretta, S. Capponi, J. Cisarova, T. Coletta, P. Corboz, S. Dommange, J. Dorier, M. Elhajal, P. Fazekas, M. Ferrero, J.-B. Fouet, J. Gavilano, A. Honecker, M. Horvatic, D. Ivanov, S. Korshunov, V. Kotov, B. Kumar, C. Lacroix, N. Laflorencie, A. Läuchli, C. Lhuillier, M. Mambrini, S. Manmana, V. Mazurenko, P. Mendels, L. Messio, F. Michaud, P. Millet, G. Misguich, S. Miyahara, R.M. Noack, B. Normand, H.R. Ott, K. Penc, J.-D. Picon, D. Poilblanc, A. Ralko, I. Rousochatzakis, K. Schmidt, J. Strecka, M. Takigawa, O. Tchernyshyov, T. Toth, M. Troyer, F. Vernay, C. Weber, S. Wenzel, H.Y. Yang, F.C. Zhang, and M. Zhitomirsky. This work has been supported by the Swiss National Science Foundation.

References

- [1] J.-F. Sadoc and R. Mosseri, Geometrical Frustration, (Cambridge University Press, 1999)
- [2] G.H. Wannier, Phys. Rev. **79**, 357 (1950)
- [3] R. Moessner, S.L. Sondhi, and P. Chandra, Phys. Rev. Lett. **84**, 4457 (2000)
- [4] M.E. Fisher, Phys. Rev. **124**, 1664 (1961)
- [5] P.W. Kasteleyn, J. Math. Phys. **4**, 287 (1963)
- [6] L. Pauling, J. Am. Chem. Soc. **57**, 2680 (1935)
- [7] J.S. Gardner, M.J.P. Gingras and J.E. Greedan, Rev. Mod. Phys. **82**, 53 (2010)
- [8] J.F. Nagle, Journal of Mathematical Physics **7**, 1484 (1966)
- [9] A.P. Ramirez, A. Hayashi, R.J. Cava, R. Siddharthan, and B. S. Shastry, Nature **399**, 333 (1999)
- [10] J.M. Luttinger and L. Tisza, Phys. Rev. **70**, 954 (1946)
- [11] R.J. Baxter, J. Math. Phys. **11**, 784 (1970)
- [12] A. Sütö, Z. Phys. B **44**, 121 (1981)
- [13] R. Kotecký, J. Salas, and A.D. Sokal, Phys. Rev. Lett. **101**, 030601 (2008)
- [14] S. Korshunov, F. Mila, and K. Penc, Phys. Rev. B **85**, 174420 (2002)
- [15] J. Stephenson, J. Math. Phys. **5**, 1009 (1964)
- [16] C.L. Henley, J. Stat. Phys. **89**, 483 (1997)
- [17] R. Moessner and S.L. Sondhi, Phys. Rev. B **63**, 224401 (2001)
- [18] H.W.J. Blöte and H.J. Hilhorst, J. Phys. A: Math. Gen. **15**, L631 (1982)
- [19] S.V. Isakov, K. Gregor, R. Moessner, and S.L. Sondhi, Phys. Rev. Lett. **93**, 167204 (2004)
- [20] D.A. Garanin and B. Canals, Phys. Rev. B **59**, 443 (1999)
- [21] T. Fennell, P.P. Deen, A.R. Wildes, K. Schmalzl, D. Prabhakaran, A.T. Boothroyd, R.J. Aldus, D.F. McMorrow, and S.T. Bramwell, Science **326**, 415 (2009)
- [22] E.F. Shender, Zh. Eksp. Teor. Fiz. **83**, 326 (1982) [Sov. Phys. JETP **56**, 178 (1982)]
- [23] J. Villain, R. Bidaux, J.P. Carton and R. Conte, J. Phys. (Paris) **41**, 1263 (1980)
- [24] C.L. Henley, Phys. Rev. Lett. **62**, 2056 (1989)

- [25] P. Chandra, P. Coleman, and A.I. Larkin, Phys. Rev. Lett. **64**, 88 (1990)
- [26] C. Weber, L. Capriotti, G. Misguich, F. Becca, M. Elhajal, and F. Mila, Phys. Rev. Lett. **91**, 177202 (2003)
- [27] J.T. Chalker, P.C.W. Holdsworth, and E.F. Shender, Phys. Rev. Lett. **68**, 855 (1992)
- [28] P. Chandra and B. Douçot, Phys. Rev. B **38**, 9335(R) (1988)
- [29] L. Balents, Nature **464**, 199 (2010)
- [30] F.D.M. Haldane., Phys. Lett. A **93**, 464 (1983)
- [31] E. Dagotto and T.M. Rice, Science **271**, 618 (1996)
- [32] I. Affleck, T. Kennedy, E.H. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987)
- [33] B. Sriram Shastry and B. Sutherland, Physica (Amsterdam) **108B+C**, 1069 (1981)
- [34] C.K. Majumdar and D.K. Ghosh, J. Math. Phys. (N.Y.) **10**, 1388 (1969)
- [35] F. Michaud, F. Vernay, S.R. Manmana, and F. Mila, Phys. Rev. Lett. **108**, 127202 (2012)
- [36] P. Corboz and F. Mila, Phys. Rev. B **87**, 115144 (2013)
- [37] H. Kageyama, K. Yoshimura, R. Stern, N.V. Mushnikov, K. Onizuka, M. Kato, K. Kosuge, C.P. Slichter, T. Goto, and Y. Ueda, Phys. Rev. Lett. **82**, 3168 (1999)
- [38] For a recent review, see M. Takigawa and F. Mila, p. 241 in [71]
- [39] P.W. Anderson, Mater. Res. Bull. **8**, 153 (1973)
- [40] P. Fazekas and P.W. Anderson, Philos. Mag. **30**, 23 (1974)
- [41] B. Bernu, C. Lhuillier, and L. Pierre, Phys. Rev. Lett. **69**, 2590 (1992)
- [42] D.S. Rokhsar and S.A. Kivelson, Phys. Rev. Lett. **61**, 2376 (1988)
- [43] P. Fendley, R. Moessner, and S.L. Sondhi, Phys. Rev. B **66**, 214513 (2002)
- [44] A. Ioselevich, D.A. Ivanov, and M.V. Feigel'man, Phys. Rev. B **66**, 174405 (2002)
- [45] R. Moessner and S.L. Sondhi, Phys. Rev. Lett. **86**, 1881 (2001)
- [46] A. Ralko, M. Ferrero, F. Becca, D. Ivanov, and F. Mila, Phys. Rev. B **71**, 224109 (2005)
- [47] T. Senthil and M.P.A. Fisher, Phys. Rev. B **62**, 7850 (2000)
- [48] S. Yan, D.A. Huse, and S.R. White, Science **332**, 1173 (2011)
- [49] D. Schwandt, M. Mambrini, and D. Poilblanc, Phys. Rev. B **81**, 214413 (2010)

- [50] I. Rousochatzakis, Y. Wan, O. Tchernyshyov, and F. Mila, Phys. Rev. B **90**, 100406(R) (2014)
- [51] J.B. Marston and I. Affleck, Phys. Rev. B **39**, 11538 (1989)
- [52] M. Hermele, T. Senthil, and M.P.A. Fisher, Phys. Rev. B **72**, 104404 (2005)
- [53] X.-G. Wen, Phys. Rev. B **44**, 2664 (1991)
- [54] C. Gros, Ann. Phys. (N.Y.) **189**, 53 (1989)
- [55] For a recent review, see F. Becca, L. Capriotti, A. Parola, and S. Sorella, p. 379 in [71]
- [56] Y. Ran, M. Hermele, P.A. Lee, and X.-G. Wen, Phys. Rev. Lett. **98**, 117205 (2007)
- [57] V. Kalmeyer and R.B. Laughlin, Phys. Rev. Lett. **59**, 2095 (1987)
- [58] X.G. Wen, F. Wilczek, and A. Zee, Phys. Rev. B **39**, 11413 (1989)
- [59] X.G. Wen, Phys. Rev. B **40**, 7387 (1989)
- [60] L. Messio, C. Lhuillier, and G. Misguich, Phys. Rev. B **83**, 184401 (2011)
- [61] S.-S. Gong, W. Zhu, and D.N. Sheng, Sci. Rep. **4**, 6317 (2014)
- [62] B. Bauer, L. Cincio, B.P. Keller, M. Dolfi, G. Vidal, S. Trebst, and A.W.W. Ludwig, Nature Communications **5**, 5137 (2014)
- [63] M. Hermele, M.P.A. Fisher, and L. Balents, Phys. Rev. B **69**, 064404 (2004)
- [64] A.F. Andreev and I.A. Grishchuk, Zh. Eksp. Teor. Fiz. **87**, 467 (1984)
[Sov. Phys. JETP **60**, 267 (1984)]
- [65] B.A. Ivanov and A.K. Kolezhuk, Phys. Rev. B **68**, 052401 (2003)
- [66] H. Tsunetsugu and M. Arikawa, J. Phys. Soc. Jpn. **75**, 083701 (2006)
- [67] A. Läuchli, F. Mila, and K. Penc, Phys. Rev. Lett. **97**, 087205 (2006)
- [68] P. Chandra and P. Coleman, Phys. Rev. Lett. **66**, 100 (1991)
- [69] N. Shannon, T. Momoi, and P. Sindzingre, Phys. Rev. Lett. **96**, 027213 (2006)
- [70] For a review, see S. Korshunov, Physics - Uspekhi **49** (3), 225 (2006)
- [71] For a recent review of all aspects of the fields (materials, experiments and theory), see C. Lacroix, P. Mendels, and F. Mila (eds.): *Introduction to Frustrated Magnetism* (Springer, New York, 2011)

8 Introduction to Mean-Field Theory of Spin Glass Models

Václav Janiš

Institute of Physics

Academy of Sciences of the Czech Republic

Prague

Contents

1	Introduction	2
1.1	Electronic background of spin-glass interactions	2
1.2	Models of spin glasses	2
1.3	Replica trick and the Sherrington-Kirkpatrick mean-field solution	4
2	Fundamental concepts of the full mean-field theory of spin glasses	6
2.1	Ergodicity, thermodynamic limit and thermodynamic homogeneity	6
2.2	Real replicas and phase-space scalings	8
2.3	Hierarchical construction of mean-field free energies	10
2.4	Continuous limit: Parisi solution	15
3	Asymptotic solutions of mean-field models: K-level replica symmetry breaking	17
3.1	Ising glass	17
3.2	Potts glass	20
3.3	p -spin glass	22
4	Conclusions	24

1 Introduction

1.1 Electronic background of spin-glass interactions

Electrons are responsible for most of the low-temperature electromagnetic phenomena in solids. That is why there is a vast effort to reliably understand the behavior of systems with many electrons in crystals. Electrons possess charge and spin and, since their mass is small, also non-negligible kinetic degrees of freedom. Moreover, at low temperatures the Fermi statistics and the Pauli principle add to the complexity of many-electron systems. The combination of Coulomb repulsion, kinetic energy, and Fermi statistics leads to a large scale of collective quantum phenomena in which electrons are either major agents or act indirectly as mediators. The latter is the case for metallic spin glasses.

Spin glasses are magnetic systems in which the interaction between the well formed and immobile magnetic moments is frustrated and the forces from different sources are in conflict with each other due to a frozen structural disorder. Hence, no conventional long-range order establishes at low temperatures. Nevertheless, these systems exhibit a freezing or ordering temperature signaling emergence of a new “glassy” phase. The classical examples of spin glasses are noble metals (Au, Ag, Cu, Pt) weakly diluted with transition metal ions, mostly Fe or Mn. Although only the interaction between the local magnetic moments of the transition metal ions is relevant, the metal matrix in which they are diluted supplies the mediating particles, the conduction electrons. The scattering of the conduction electrons on magnetic impurities leads to an indirect Ruderman, Kittel, Kasuya and Yosida (RKKY) [1–3] exchange interaction between the spins that strongly oscillates with distance $\mathbf{R}$,

$$J(\mathbf{R}) = J_0 \frac{\cos(2k_F R + \phi_0)}{(k_F R)^3}, \quad R \rightarrow \infty.$$

Here J_0 and ϕ_0 are material constants and k_F is the Fermi wave number of the host metal. Since the magnetic ions are strongly diluted in the host, the distances between the spins are effectively random, leading to a random distribution of the spin exchange with no preference for a homogeneous magnetic order. Positive and negative signs of the exchange are then equally likely and we can model the RKKY interaction by a randomly fluctuating exchange $J(R) = \pm J_0/k_F R^3$. Since the alloys are very dilute, the average distance between the active spins is large, and a mean-field approximation is well justified. The first mean-field model of spin glasses was proposed by Edwards and Anderson [4], just a few years after the first experimental evidence of spin-glass behavior was discovered [5].

1.2 Models of spin glasses

The spin of the conduction electrons of the noble metals serve only to mediate the interaction between the local spin moments of the transition metal alloys. In modeling the spin-glass behavior, we can resort to purely immobile spins distributed regularly on a lattice with a random spin exchange with no preference for either ferro or antiferromagnetic long-range order. For a spin

Hamiltonian $H[J, S]$, we have to evaluate the free energy of the system in the thermodynamic limit. We assume that the free energy is self-averaging and the system is ergodic, hence

$$-\beta F(T) = \lim_{N \rightarrow \infty} \ln \text{Tr}_S \exp \{-\beta H[J, S]\} = \left\langle \ln \text{Tr}_S \exp \{-\beta H[J, S]\} \right\rangle_J. \quad (1)$$

We will later analyze whether any of the fundamental assumptions of the standard statistical mechanics is broken in the mean-field theory of spin glasses. Although Heisenberg spins are physically most interesting, the glassy behavior in random frustrated spin systems becomes so complex that full understanding of the new phenomena originating from the glassy phase demands either further simplifications or alternative spin models to study. We will deal with three spin-glass models that show different scenarios of the paramagnetic to spin-glass transition.

1.2.1 Heisenberg and Ising models

The simplest lattice spin system consists of spins with the lowest value $\hbar/2$. We set $\hbar/2 = 1$ and $k_B = 1$ to simplify the resulting formulas. Since the transition temperature is relatively high, the spins can be treated classically. The Hamiltonian of the Heisenberg spin model in a homogeneous external magnetic field $\mathbf{h}$ on a regular lattice is

$$H[J, S] = - \sum_{i < j}^N J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j - \sum_i \mathbf{h} \cdot \mathbf{S}_i \quad (2)$$

with a normalization condition $\mathbf{S}_i \cdot \mathbf{S}_i = 1$. Although the spin rotational degrees of freedom influence the way the system can go from the paramagnetic to the spin-glass phase [6, 7], it is actually the Ising model that has been mostly studied in theories of spin glasses. Only the projection S^z of the spin vector onto the easy axis, determined by the external magnetic field, is significant in the Ising model. We use the Ising model as the generic case for the demonstration of properties of the mean-field solutions of spin glasses. The mean-field limit of the Ising spin glass is called the Sherrington-Kirkpatrick (SK) model.

1.2.2 Potts model

The p -state Potts model is a generalization of the Ising model to $p > 2$ spin components. The original formulation of Potts [8] with the Hamiltonian $H_p = - \sum_{i < j} J_{ij} \delta_{n_i, n_j}$ where $n_i = 0, \dots, p-1$ is an admissible value of the spin projection of the p -state model on lattice site $\mathbf{R}_i$, is unsuitable for practical calculations. The Potts Hamiltonian can, however, be represented via interacting spins [9]

$$H_P[J, S] = -\frac{1}{2} \sum_{i, j}^N J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j - \sum_i \mathbf{h} \cdot \mathbf{S}_i, \quad (3)$$

where $\mathbf{S}_i = \{s_i^1, \dots, s_i^{p-1}\}$ are Potts vector variables taking values from a set of state vectors $\{\mathbf{e}_A\}_{A=1}^p$. Functions on vectors $\mathbf{e}_A$ are in equilibrium fully defined through their scalar product

$$\sum_{A=1}^p e_A^\alpha = 0, \quad \sum_{A=1}^p e_A^\alpha e_A^\beta = p \delta^{\alpha\beta}, \quad e_A^\alpha e_B^\alpha = p \delta_{AB} - 1$$

for $\alpha \in \{1, \dots, p-1\}$. We use the Einstein summation convention for repeating Greek indices of the vector components indicating a scalar product of the Potts vectors. Using these properties we can construct an explicit representation of the Potts spin vectors

$$e_A^\alpha = \begin{cases} 0 & A < \alpha \\ \sqrt{\frac{p(p-\alpha)}{p+1-\alpha}} & A = \alpha \\ \frac{1}{\alpha-p} \sqrt{\frac{p(p-\alpha)}{p+1-\alpha}} & A > \alpha \end{cases}.$$

The Potts model with a random spin exchange shows a transition to a glassy phase, but the scenario depends on the number of spin components p .

1.2.3 p -spin model

The Potts model is not the only interesting extension of the Ising model. Another generalization is the so-called p -spin model. It describes a system of Ising spins where the spin exchange connects a cluster of p spins. The Hamiltonian of such a model reads [10]

$$H_p[J, S] = \sum_{1 \leq i_1 < i_2 < \dots < i_p}^N J_{i_1 i_2 \dots i_p} S_{i_1} S_{i_2} \dots S_{i_p}. \quad (4)$$

The p -spin model is interesting in that we can analytically study the limit $p \rightarrow \infty$ for which we know the exact solution being the random-energy model [11, 12]. It is equivalent to the one-level replica-symmetry breaking solution from the replica trick [13]. There were hopes that one could understand better the genesis of the full mean-field solution of the Ising glass by using the inverse number of the coupled spins, $1/p$, as a small parameter starting from the random-energy model. We show later on that the $1/p$ expansion does not work, since it does not cure the negative entropy at zero temperature for $p < \infty$.

1.3 Replica trick and the Sherrington-Kirkpatrick mean-field solution

Given the spin Hamiltonian one has to resolve the free energy by evaluating the right-hand side of Eq. (1). We use the regular lattice with a frustrated random nearest-neighbor interaction to simulate the magnetic properties of the diluted magnetic ions in noble metals. Since the average distance between the ions is large, one assumes a long-range spin exchange in the spin-glass models, which leads us naturally to a mean-field theory. The modern understanding of the mean-field approximation is the mathematical limit $d \rightarrow \infty$ of the model on a d -dimensional hypercubic lattice. Lattice mean-field theories with long-range interactions are also called models on fully connected graphs, where each node of the graph is connected to any other node.

The limit of infinite dimensions or the long-range interaction introduces a new large scale. To make the thermodynamic limit meaningful the dependence of the energy on this new large scale must be compensated by rescaling the non-local spin exchange so that the energy remains linearly proportional to the volume or the number of lattice sites (spins). Since the linear contribution from the spin exchange J_{ij} is missing in the spin-glass models due to frustration that

prefers neither ferromagnetic nor antiferromagnetic order, we have to rescale it as $J = J_{ij}/\sqrt{N}$. To simplify the numerics, one usually chooses a Gaussian random distribution

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi J^2}} \exp \left\{ -\frac{N J_{ij}^2}{2J^2} \right\}. \quad (5)$$

Even with the mean-field simplification one faces the hard problem of averaging the logarithm of the partition sum, that is, to evaluate the following multiple integral

$$-\beta \langle F_N \rangle_{av} = \int \prod_{i<j}^N d[J_{ij}] P[J_{ij}] \ln \int \prod_{i=1}^N d[\mathbf{S}_i^a] \rho[\mathbf{S}_i^a] \exp \{ -\beta H[J, \mathbf{S}] \}. \quad (6)$$

The next simplification is introduced by the application of the replica trick introduced in Ref. [4]. We create ν copies of the original spin variables and average the replicated system over the random spin exchange. Simply counting the diagrammatic contributions to the free energy where the spin exchange is represented by a bond, one easily finds that each closed loop contributes ν times. Hence, the free energy can be represented as

$$\beta F = -\lim_{\nu \rightarrow 0} \left[\frac{1}{\nu} \lim_{N \rightarrow \infty} \left(\langle Z_N^\nu \rangle_{av} - 1 \right) \right]. \quad (7)$$

Representation (7) simplifies the averaging over randomness tremendously, in particular when the Gaussian distribution of the random spin exchange from Eq. (5) is used. The integration over the spin exchange can be explicitly performed for the replicated partition sum and then, assuming *ergodicity* and the existence of the thermodynamic limit together with the validity of linear response, one can get an explicit representation for the density of the free energy of the Ising spin glass with a single order parameter $q = N^{-1} \sum_i m_i^2$ in the glassy phase

$$f(q) = -\frac{\beta}{4}(1-q)^2 - \frac{1}{\beta} \int_{-\infty}^{\infty} \frac{d\eta}{\sqrt{2\pi}} e^{-\eta^2/2} \ln 2 \cosh [\beta (h + \eta\sqrt{q})] \quad (8)$$

derived by Sherrington and Kirkpatrick [14].

At first sight, there is nothing wrong with the derivation of the Sherrington-Kirkpatrick free energy, hence one would not expect any unphysical behavior. Nevertheless, the authors themselves had already found that the entropy, calculated from the free energy as $S(T) = -\partial F(T)/\partial T$ leads to a negative value at zero temperature, $S(0) = -\sqrt{2/\pi} k_B \approx -0.798 k_B$. And this is too bad. This unexpected result triggered an avalanche of attempts to resolve the enigma. It was found soon by Monte Carlo simulations that the entropy of the Sherrington-Kirkpatrick model is non-negative and approaches zero when lowering the temperature [15, 16]. Analyzing the stability of the SK solution, de Almeida and Thouless showed that it is actually unstable in the whole spin-glass phase. The stability condition to be satisfied by the SK solution

$$A = 1 - \beta^2 \left\langle \left(1 - \tanh^2 [\beta (h + \eta\sqrt{q})] \right)^2 \right\rangle_{\eta} \geq 0 \quad (9)$$

is broken everywhere below the transition temperature along the de Almeida-Thouless (AT) line in arbitrary magnetic field [17]. Initially it was suggested that the replica trick and averaging over the spin exchange prior to averaging the thermal fluctuations is responsible for the

instability [18]. The theory of Thouless, Anderson and Palmer, where thermal averaging was performed for a fixed configuration of the spin exchanges, did not, however, resolve the problem [19]. When averaging over the random exchange is applied in their theory by assuming ergodicity, one ends up with the SK solution [20]. After a number of unsuccessful or only partially successful attempts it was Giorgio Parisi who proposed a scheme of replica-symmetry breaking that would lead to a (marginally) stable and thermodynamically consistent equilibrium state [21–26]. Indeed, two decades later it was rigorously proven that the Parisi construction of the replica symmetry breaking leads to the exact free energy of the Sherrington-Kirkpatrick model [27, 28].

There is a vast literature documenting the way to the full mean-field solution of the spin glass models, and we refer the reader to review articles [29, 30] or books [31–35] for details. Here we use an offbeat road to the full solution of the mean-field spin-glass models by trying to identify the thermodynamic origin of the failure of the SK mean-field solution [36–39].

2 Fundamental concepts of the full mean-field theory of spin glasses

We now know that the replica trick and the way we use it is not the cause of the instability of the SK mean-field solution. One has to find a physical argument or a hole in the derivation of the SK free energy. Although the SK solution is unstable within the whole glassy phase, the major problem of this simple mean-field theory is negative entropy at very low temperatures. This constitutes a severe, intolerable problem. One has to find the reason for this behavior. In deriving the mean-field solution, we assumed that the fundamental principles of statistical mechanics are valid, the thermodynamic limit exists, and different statistical ensembles are equivalent. Only if this is true we can equate the entropy calculated from the free energy of the canonical ensemble with the entropy from the microcanonical one. The entropy in the latter ensemble is positive by definition and hence the canonical and microcanonical ensembles in the Sherrington-Kirkpatrick solution are not equivalent and do not lead to the same physical results. The assumptions for the existence of the thermodynamic limit must then be revisited.

2.1 Ergodicity, thermodynamic limit and thermodynamic homogeneity

The very fundamental basis on which statistical mechanics is built is the ergodic hypothesis. It is the means by which the long-time evolution of a microscopic state is related to statistical averaging over the allowed states in phase space. The Birkhoff ergodic theorem asserts that for ergodic systems [40]

$$\langle f \rangle_T \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_0+T} f(X(t)) dt = \frac{1}{\Sigma_E} \int_{S_E} f(X) dS_E \equiv \langle f \rangle_S$$

holds. Here T is a macroscopic time scale, $X(t)$ the classical trajectory of a microscopic state, S_E the constant energy subspace of the phase space and Σ_E its volume. It means that the classi-

cal trajectory of the microscopic state covers almost all of the allowed phase space constrained by external macroscopic conditions. It is, however, highly nontrivial to determine which are the allowed states indeed with only homogeneous macroscopic parameters. This is actually the most difficult task in constructing the proper phase space in statistical models. That is, to find out which points of the phase space are infinitesimally close to the trajectory of the many-body system extended to infinite times. Since we never solve the equation of motion of the statistical system, we have only static means to check the validity of ergodicity. We do it by testing the validity of consequences of the ergodic hypothesis and its impact on the behavior of the equilibrium state. The most important consequence of ergodicity of statistical systems is the existence of the thermodynamic limit.

The trajectory of the many-body system covers almost the whole allowed phase space. It means that the space covered by such trajectory does not depend on the initial state in non-chaotic systems. In ergodic systems then the thermodynamic limit does not depend on the specific form of the volume in which the macroscopic state is confined nor on its surrounding environment. The ergodic macroscopic systems can either be isolated or embedded in a thermal bath. The thermodynamic equilibrium, the equilibrium state in the thermodynamic limit, is the same with vanishing relative statistical fluctuations. The thermodynamic equilibrium can then be reached by limiting any partial volume of the whole to infinity. The ergodic equilibrium state is homogeneous in the thermodynamic limit.

Thermodynamic homogeneity is usually expressed via Euler's lemma [41]

$$\alpha F(T, V, N, \dots, X_i, \dots) = F(T, \alpha V, \alpha N, \dots, \alpha X_i, \dots),$$

telling us that the thermodynamic potential, the free energy F in this case, is an extensive variable and is a first-order homogeneous function of all its extensive variables; volume V , number of particles N , and the other model-dependent extensive variables X_i . As a consequence of Euler's lemma, we obtain that thermodynamic equilibrium is attained as a one-parameter scaling limit where we have only one independent, large-scale, extensive variable, be it either the volume or the number of particles, and the other extensive variables enter the thermodynamic potentials as volume or particle densities insensitive to fluctuations of the scaling variable.

Although the foundations of statistical mechanics are based on ergodicity, lack of ergodicity is widespread in physical phenomena [42]. Typical examples of ergodicity breaking are phase transitions with a symmetry breaking in the underlying Hamiltonian. Broken ergodicity is sometimes used as a generalization of spontaneous symmetry breaking [43]. Although broken global symmetry is always accompanied by broken ergodicity, the converse does not hold. Ergodicity is broken in the mean-field spin-glass models without any symmetry of the Hamiltonian being simultaneously broken.

Broken ergodicity represents an obstruction in the application of fundamental thermodynamic laws. It hence must be recovered. When ergodicity is broken in a phase transition breaking a symmetry of the Hamiltonian, one introduces a symmetry-breaking field into the Hamiltonian, being the Legendre conjugate to the extensive variable that is not conserved in the broken symmetry transformation in the low-temperature phase. The symmetry-breaking field allows one

to circumvent the critical point of the symmetry-breaking phase transition and simultaneously restores ergodicity. Systems in which no symmetry of the Hamiltonian is broken at the phase transition do not offer natural external symmetry-breaking fields. Other techniques must be employed to find the proper portion of the phase space covered by the trajectory of the microscopic state in the low-temperature phase and to restore ergodicity.

2.2 Real replicas and phase-space scalings

Thermodynamic homogeneity allows us to use the scaling of the original phase space. Thermodynamic quantities remain unchanged if we arbitrarily rescale the phase space and then divide the resulting thermodynamic potential by the chosen scaling (geometric) factor. We can do that by scaling the energy E of the equilibrium state. If we use a scaling factor ν , which can be an arbitrary positive number, then the following identities hold for entropy $S(E)$ of the microcanonical and free energy $F(T)$ of the canonical ensemble with energy E and temperature T , respectively:

$$S(E) = k_B \ln \Gamma(E) = \frac{k_B}{\nu} \ln \Gamma(E)^\nu = \frac{k_B}{\nu} \ln \Gamma(\nu E) , \quad (10a)$$

$$F(T) = - \frac{k_B T}{\nu} \ln [\text{Tr} e^{-\beta H}]^\nu = - \frac{k_B T}{\nu} \ln [\text{Tr} e^{-\beta \nu H}] , \quad (10b)$$

where we denoted by $\Gamma(E)$ the phase-space volume of the isolated system with energy E .

The scaling of the phase space with an integer scaling factor ν can be simulated by replicating the extensive variables ν times. That is, instead of a single phase space, we use ν replicas of the original space. The reason for introducing replicas of the original variables is to extend the space of the available states in the search for the allowed space in equilibrium. The replicas are independent when introduced. We use the replicated variables to study the stability of the original system with respect to fluctuations in the surrounding thermal bath. For this purpose, we break the independence of the replica variables by switching on a (homogeneous) infinitesimal interaction between the replicas that we denote μ^{ab} . We then add a small interacting part $\Delta H(\mu) = \sum_i \sum_{a < b}^\nu \mu^{ab} X_i^a X_i^b$ to the replicated Hamiltonian with dynamical extensive variables X_i . The original system is then stable with respect to fluctuations in the bath, represented by the interaction between the replicated variables, if the linear response to the perturbation μ is not broken. If the linear response holds, then, after switching the perturbation μ off, the perturbed free energy per replica relaxes to the original one in the thermodynamic limit

$$-\beta F_\nu(\mu) = \frac{1}{\nu} \ln \text{Tr}_\nu \exp \left\{ -\beta \sum_{a=1}^\nu H^a - \beta \Delta H(\mu) \right\} \xrightarrow{\mu \rightarrow 0} \ln \text{Tr} \exp \{ -\beta H \} , \quad (11)$$

where Tr_ν refers to the trace in the ν -times-replicated phase space. If the linear response to the inter-replica interaction is broken, the thermodynamic limit of the original system is not uniquely defined and depends on properties of the thermal bath represented by the replicated variables. If there are no apparent physical fields breaking the symmetry of the Hamiltonian, the phase-space scaling represented by replicas of the dynamical variables introduces shadow or auxiliary symmetry-breaking fields, inter-replica interactions $\mu^{ab} > 0$. They induce new

order parameters in the response of the system to these fields that need not vanish in the low-temperature phase, when the linear response breaks down. The real replicas offer a way to disclose a degeneracy when the thermodynamic limit is not uniquely defined by a single extensive scale and densities of the other extensive variables. The inter-replica interactions are not measurable and hence to restore the physical situation we have to switch off these fields at the end. If the system is thermodynamically homogeneous, we must fulfill the following identity

$$\lim_{\mu \rightarrow 0} \frac{dF_\nu(\mu)}{d\nu} \equiv 0 \quad (12)$$

for arbitrary ν . This quantification of the global thermodynamic homogeneity, and thermodynamic independence of the scaling parameter ν , will lead us to the construction of a stable solution of mean-field spin glass models. To use equation (12) in the replica approach we will need to analytically continue the replica-dependent free energy to arbitrary positive scaling factors $\nu \in \mathbb{R}^+$. Specific assumptions on the symmetry of the matrix μ^{ab} will have to be introduced. It is evident from Eq. (11) that the linear response to inter-replica interactions can be broken only if the replicas are mixed in the ν -times-replicated free energy F_ν .

2.2.1 Replicated Sherrington-Kirkpatrick model

We apply real replicas to test the thermodynamic homogeneity of the Sherrington-Kirkpatrick model. We replicate the Ising Hamiltonian ν times

$$[H]_\nu = \sum_{a=1}^{\nu} H^a = \sum_{\alpha=1}^{\nu} \sum_{\langle ij \rangle} J_{ij} S_i^a S_j^a$$

and add a small replica-mixing perturbation $\Delta H(\mu) = \frac{1}{2} \sum_{a \neq b} \sum_i \mu^{ab} S_i^a S_i^b$. We assume that ergodicity and linear response hold in the replicated phase space. The averaging over the long-range spin exchange leads to mixing of the replicated spins, and after performing averaging over the random spin exchange we obtain in the limit $\mu \rightarrow 0$ an extended Sherrington-Kirkpatrick free-energy density with the Sherrington-Kirkpatrick order parameter q and new off-diagonal parameters χ^{ab} [36]

$$f_\nu = \frac{\beta J^2}{4} \left[\frac{1}{\nu} \sum_{a \neq b}^{\nu} \left\{ (\chi^{ab})^2 + 2q\chi^{ab} \right\} - (1-q)^2 \right] - \frac{1}{\beta\nu} \int_{-\infty}^{\infty} \frac{d\eta}{\sqrt{2\pi}} e^{-\eta^2/2} \ln \text{Tr}_\nu \exp \left\{ \beta^2 J^2 \sum_{a < b}^{\nu} \chi^{ab} S^a S^b + \beta \bar{h} \sum_{a=1}^{\nu} S^a \right\}. \quad (13)$$

The new parameters in the extended phase space are the response functions conjugate to the inter-replica interaction μ^{ab} and are the inter-replica susceptibilities $\chi^{ab} = \langle \langle S^a S^b \rangle_T \rangle_{av} - q$, in complete analogy to the real magnetic field to which the magnetic susceptibility is a linear response. Since we deal with a mean-field model, the susceptibilities are local. Here $\langle f(S) \rangle_T$ denotes averaging over thermal fluctuations and $\langle f(S) \rangle_{av}$ over the spin exchange. The SK order parameter is $q = \langle \langle S^a \rangle_T^2 \rangle_{av}$ and $\bar{h} = h + \eta\sqrt{q}$.

$$\begin{pmatrix} 0 & \chi_1 & \chi_2 & \chi_2 & \chi_3 & \chi_3 & \chi_3 & \chi_3 \\ \chi_1 & 0 & \chi_2 & \chi_2 & \chi_3 & \chi_3 & \chi_3 & \chi_3 \\ \chi_2 & \chi_2 & 0 & \chi_1 & \chi_3 & \chi_3 & \chi_3 & \chi_3 \\ \chi_2 & \chi_2 & \chi_1 & 0 & \chi_3 & \chi_3 & \chi_3 & \chi_3 \\ \chi_3 & \chi_3 & \chi_3 & \chi_3 & 0 & \chi_1 & \chi_2 & \chi_2 \\ \chi_3 & \chi_3 & \chi_3 & \chi_3 & \chi_1 & 0 & \chi_2 & \chi_2 \\ \chi_3 & \chi_3 & \chi_3 & \chi_3 & \chi_2 & \chi_2 & 0 & \chi_1 \\ \chi_3 & \chi_3 & \chi_3 & \chi_3 & \chi_2 & \chi_2 & \chi_1 & 0 \end{pmatrix}$$

Fig. 1: Matrix of overlap susceptibilities χ^{ab} for $\nu = 8$ and with three levels (hierarchies) of symmetry breaking $K = 3$ exemplifying the structure allowing for analytic continuation to arbitrary positive ν .

Replicated spin variables introduced new order parameters that should be determined from stationarity of the free-energy density from Eq. (13). There are, however, two problems with this free energy. First, it is not a closed expression that would give its analytic dependence on replication index ν . Second, representation (13) holds only for integer numbers ν . The replicated spins do not have a direct physical meaning and were introduced only to simulate scalings of the phase space to test thermodynamic homogeneity of the resulting free energy. Hence, one has to analytically continue the free energy (13) to a positive real replication index ν . A specific symmetry of the matrix of the overlap susceptibilities must be assumed to reach this objective. Such a symmetry was found by G. Parisi and is independent of the replica trick and the limit of the number of replicas to zero.

2.3 Hierarchical construction of mean-field free energies

First what one notices is that the replicated free energy, Eq. (13), contains too many order parameters, $\nu(\nu - 1)/2$. It means that the solution is degenerate. Actually, the number of independent parameters in the ν -times-replicated phase space should not be bigger than the number of replicas ν . Parisi assumed the following structure of the overlap susceptibilities

$$\chi^{aa} = 0, \quad \chi^{ab} = \chi^{ba}, \quad \sum_c (\chi^{ac} - \chi^{bc}) = 0. \quad (14)$$

It means that each row/column contains the same elements, only in a different order. This is a consequence of the third condition in Eq. (14). Hence the matrix of the overlap susceptibilities has just $\nu - 1$ independent numbers. The values of the overlap susceptibilities can appear multiple times in each row. Let us assume that we have just K different values of the overlap susceptibilities that we denote $\chi_1, \chi_2, \dots, \chi_K$. Let the corresponding multiplicities be $m_1, m_2, \dots, m_K$. A sum rule $\nu - 1 = \sum_{l=1}^K m_l$ then holds. An example of such a matrix for $\nu = 8$, $K = 3$, and $m_l = 2^{l-1}$ is illustrated in Fig. 1.

It is now straightforward to perform the sum over the spin configurations and find a closed expression for the free energy with matrices χ^{ab} fulfilling criteria (14). We label the overlap

susceptibilities so that they form a decreasing succession $\chi_l > \chi_{l+1}$. The averaging over the thermal fluctuations of the replicated spins in the free energy from Eq. (13) can now be performed explicitly and we obtain [36]

$$f_K(q, \{\chi\}; \{m\}) = -\frac{\beta}{4}(1-q)^2 + \frac{\beta}{4} \sum_{l=1}^K (m_l - m_{l-1}) \chi_l (2q + \chi_l) + \frac{\beta}{2} \chi_1 \\ - \frac{1}{\beta m_K} \int_{-\infty}^{\infty} \frac{d\eta}{\sqrt{2\pi}} e^{-\eta^2/2} \ln \left[\int_{-\infty}^{\infty} \frac{d\lambda_K}{\sqrt{2\pi}} e^{-\lambda_K^2/2} \left\{ \dots \int_{-\infty}^{\infty} \frac{d\lambda_1}{\sqrt{2\pi}} e^{-\lambda_1^2/2} \right. \right. \\ \left. \left. \left\{ 2 \cosh \left[\beta \left(h + \eta \sqrt{q} + \sum_{l=1}^K \lambda_l \sqrt{\chi_l - \chi_{l+1}} \right) \right] \right\}^{m_1} \dots \right\}^{m_K/m_{K-1}} \right] \quad (15a)$$

with $\chi_{K+1} = 0$ and $m_0 = 1$. It may appear convenient to rewrite the free-energy density in another equivalent form

$$f_K(q; \Delta\chi_1, \dots, \Delta\chi_K, m_1, \dots, m_K) = -\frac{\beta}{4} \left(1 - q - \sum_{l=1}^K \Delta\chi_l \right)^2 - \frac{1}{\beta} \ln 2 \\ + \frac{\beta}{4} \sum_{l=1}^K m_l \Delta\chi_l \left[2 \left(q + \sum_{i=l}^K \Delta\chi_i \right) - \Delta\chi_l \right] - \frac{1}{\beta} \int_{-\infty}^{\infty} \mathcal{D}\eta \ln Z_K, \quad (15b)$$

where we ordered the parameters so that $\Delta\chi_l = \chi_l - \chi_{l+1} \geq \Delta\chi_{l+1} \geq 0$. We further used a short-hand notation for iterative partition functions

$$Z_l = \left[\int_{-\infty}^{\infty} \mathcal{D}\lambda_l Z_{l-1}^{m_l} \right]^{1/m_l}$$

with an abbreviation for a Gaussian differential $\mathcal{D}\lambda \equiv d\lambda e^{-\lambda^2/2}/\sqrt{2\pi}$. The initial partition function for the Ising spin glass is $Z_0 = \cosh \left[\beta \left(h + \eta \sqrt{q} + \sum_{l=1}^K \lambda_l \sqrt{\Delta\chi_l} \right) \right]$. The free energy f_K is an analytic function of the multiplicities (geometric parameters) m_l , and hence they can now be arbitrary positive numbers. The equilibrium state in the replicated phase space is determined from the extremal point with respect to variations of the overlap susceptibilities, the order parameters in the glassy phase. With the symmetry from Eq. (14), the order parameters are the independent values of the overlap susceptibilities χ_l in representation (15a) or their differences $\Delta\chi_l$ from (15b) and their multiplicities m_l . The equilibrium values are determined from the extremal point of the respective free energy functional. The type of the extremum from which the stable equilibrium state is determined depends, however, on the values of the multiplicities m_l . If $m_l > 1$ then the equilibrium free energy is minimal with respect to variations of this parameter. If $m_l < 1$ then the equilibrium free energy is *maximal*. The value $m_l = 1$ is a degeneracy point at which the free energy is independent of $\Delta\chi_l$. It appears that the stable solution is generated by multiplicities being all between zero and one. The free energy of the spin glass is hence maximal in the replicated phase space.

Each independent value of the overlap susceptibility χ_l determines a replica hierarchy. K is then the number of replica hierarchies. The number of replica hierarchies is also related to the level of

the replica-symmetry breaking (RSB). The concept of replica-symmetry breaking comes from the replica trick used to average the free energy via the replicated partition function, Eq. (7) where the order parameters in the replicated space are $q + \chi_l$. The number of order parameters is $2K + 1$. If all $\chi_l = 0$ and $K = 0$, we have a single order parameter q and the solution is called replica symmetric. It is the SK solution. Then non-zero values of K are called K -level replica-symmetry breaking (K RSB).

It is clear that the complexity of the solution increases rapidly with increasing number of different values of the overlap susceptibilities or their differences $\Delta\chi_i$, that is, with the number K . We give here an example of the lowest replica-symmetry breaking free energy ($K = 1$) as the next step beyond the SK solution

$$f_1(q; \chi_1, m_1) = -\frac{\beta}{4}(1 - q - \chi_1)^2 + \frac{\beta}{4}m_1\chi_1(2q + \chi_1) - \frac{1}{\beta m_1} \int_{-\infty}^{\infty} \mathcal{D}\eta \ln \int_{-\infty}^{\infty} \mathcal{D}\lambda_1 \{2 \cosh [\beta (h + \eta\sqrt{q} + \lambda_1\sqrt{\chi_1})]\}^{m_1} . \quad (16)$$

It has three parameters, q , χ_1 , and m_1 to be determined from the stationarity of the free-energy functional Eq. (16). It represents a free energy with first level of ergodicity breaking or replica-symmetry breaking (1RSB).

Generally, the free energy f_K stands for ergodicity breaking on K levels, K generations of replicas. When the replica symmetry is broken, it also means that ergodicity is broken and the thermodynamic limit of the original system depends on the behavior of the spins of the surrounding bath, the replicated spins. The physical interpretation of the breaking of replica symmetry is ergodicity breaking. The hierarchical replication of the system variables is then an iterative way to restore ergodicity or thermodynamic homogeneity in a larger phase space. The order parameters from the replicated space m_l and $\Delta\chi_l$ then play the role of Legendre conjugate parameters controlling the energy exchange between the original system and its simulated thermal bath.

Free energy $f_K(q; \Delta\chi_1, \dots, \Delta\chi_K, m_1, \dots, m_K)$ contains $2K + 1$ variational parameters, q , $\Delta\chi_i$, m_i for $i = 1 \dots K$ that are determined from the stationarity of the free energy with respect to small fluctuations of these parameters. The replica construction introduced a new parameter K that is not *a priori* determined. It can assume any integer value in the true equilibrium. The number of replica hierarchies is in this construction determined from stability conditions that restrict admissible solutions, stationarity points. A solution with K levels is locally stable if it does not decay into a solution with $K + 1$ hierarchies. A new order parameter in the next replica generation $\Delta\chi$ may emerge so that $\Delta\chi_l > \Delta\chi > \Delta\chi_{l+1}$ for arbitrary l . That is, the new order parameter may peel off from $\Delta\chi_l$ and shifts the enumeration of the order parameters for $i > l$ in the existing K -level solution. To guarantee that this does not happen and that the averaged free energy depends on no more geometric parameters than $m_1, \dots, m_K$ we have to fulfill a set of $K + 1$ generalized stability criteria that for our hierarchical solution read for $l = 0 \dots K$

$$\Lambda_l^K = 1 - \beta^2 \left\langle \left\langle \left\langle 1 - t^2 + \sum_{i=0}^l m_i (\langle t \rangle_{i-1}^2 - \langle t \rangle_i^2) \right\rangle_l^2 \right\rangle_K \right\rangle_{\eta} \geq 0 \quad (17)$$

with $m_0 = 0$ and formally $\langle t \rangle_{-1} = 0$. We introduced the following short-hand notation $t \equiv \tanh \left[\beta \left(h + \eta \sqrt{q} + \sum_{l=1}^K \lambda_l \sqrt{\Delta \chi_l} \right) \right]$ and $\langle t \rangle_l(\eta; \lambda_K, \dots, \lambda_{l+1}) = \langle \rho_l \dots \langle \rho_1 t \rangle_{\lambda_1} \dots \rangle_{\lambda_l}$ with $\langle X(\lambda_l) \rangle_{\lambda_l} = \int_{-\infty}^{\infty} \mathcal{D}\lambda_l X(\lambda_l)$ and $\rho_l = Z_{l-1}^{m_l} / \langle Z_{l-1}^{m_l} \rangle_{\lambda_l}$. The lowest K for which all stability conditions Eq. (17) are fulfilled is an allowed equilibrium state. It need not, however, be the true equilibrium state, since the stability conditions test only local stability and cannot decide which of several extremal points is the true ground state. The stability conditions Eq. (17) are necessary for the system to be thermodynamically homogeneous. They are, however, not sufficient to guarantee global thermodynamic homogeneity. Note that the stability conditions from Eq. (17) guarantee only local homogeneity, since they hold for the optimal geometric parameters m_l determined by the stationarity equations. The global thermodynamic homogeneity Eq. (12) would demand $\Lambda_K^K \geq 0$ for arbitrary positive m_K . This is generally valid if $\Delta \chi_K = 0$.

The free energy with K hierarchies of replicated spin variables Eq. (15) was derived by the standard procedure utilizing ergodicity in the extended, replicated phase space. Real replicas in the thermodynamic approach, that is without the replica trick, were introduced to include control over thermodynamic homogeneity of the equilibrium state. The necessity to analytically continue the free energy to arbitrary positive replication index forced us to introduce a specific structure of the matrix χ^{ab} as exemplified in Fig. 1. Since $\chi^{ab} \geq 0$, we can consider it as a distance between replicas a and b . The reduction rules Eq. (14) lead to an *ultrametric distance* in the space of replicas [13]. An ultrametric space is characterized by the existence of only equilateral or isosceles triangles. That is, for three replica indices a, b, c either $\chi^{ab} = \chi^{ac} = \chi^{bc}$ or at least one of these equalities holds. Ultrametricity is generated by the hierarchical structure of successive replications used to derive the free energy $f_K(q, \{\chi\}; \{m\})$.

The construction of the order parameters in the replicated phase space suggests that the overlap susceptibilities χ^{ab} measure the interaction strength with which different copies (replicas) of the spins thermodynamically influence each other. That is, thermal averaging of one spin copy (a) depends on the values of spins of another copy (b) if $\chi^{ab} > 0$. We cannot separate these replicas, although only one spin replica represents the physical system under consideration. The non-replicated original phase variables together with temperature and chemical potential are hence insufficient to completely describe the equilibrium thermodynamic states. To get rid of the dependence of the thermodynamic states on the boundary or initial conditions we have to average over all initial/boundary values and external variables that influence the thermodynamics of the investigated system. In the long-range, completely connected models, the degeneracy of the solutions of the mean-field equations is reflected in the dependence on the initial spin configurations. We simulated this dependence by self-consistent interactions between the original and replicated spin variables, where both spin species are subject to the same thermal equilibration.

To understand the role of the geometric parameters (replication indices) m_l we look at the l th level of replication. It appears that when ergodicity and the linear response with respect to inter-replica interaction is broken, the solution in the replicated space has the replication indices ordered in a decreasing succession $1 > m_1 > \dots m_K \geq 0$. It allows us to give a straightforward interpretation of the successive replications. Let us take the partition sum Z_{l-1} and perform the

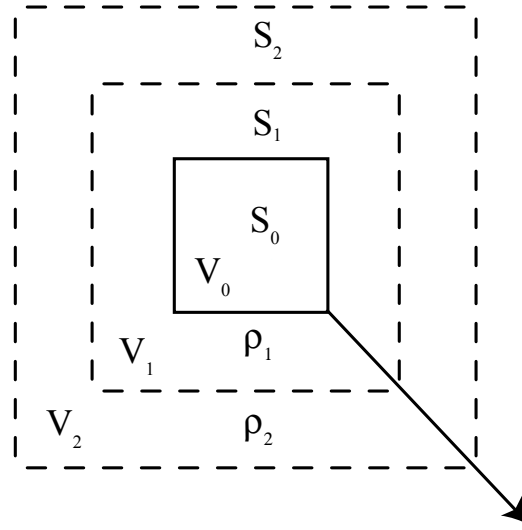


Fig. 2: The thermodynamic limit of a system of spins S_0 confined in a volume $V_0 \rightarrow \infty$ can either be performed in a free space or embedded in a bath of external spins S_1 in a larger volume $V_1 \rightarrow \infty$. The bath spins have a density matrix ρ_1 . If the two thermodynamic limits are different, we must embed the system into a next bath of spins with a density matrix ρ_2 . We continue with the embeddings so long until the smaller and the larger system lead to the same thermodynamic limit.

next replication of the spin variables from Z_{l-1} . The averaged interaction strength of the original and the replicated spins is $\Delta\chi_l$ and the averaged number of the original spins affected by the interaction with the replicated spins is $m_l N$. The spins from the subspace with $l-1$ replica hierarchies are affected by thermal fluctuations of the spins from the next, l th hierarchy. The Gaussian spins from the l th hierarchy are λ_l and their thermal fluctuations are represented by a Gaussian integral. The free energy density of the systems with l replica hierarchies then is

$$f_l(\bar{h}_l) = \frac{1}{m_l} \ln \int \mathcal{D}\lambda_l Z_{l-1}^{m_l} \left(\beta, \bar{h}_l + \lambda_l \sqrt{\Delta\chi_l} \right),$$

where $\bar{h}_l$ is an internal magnetic field.

These hierarchical replications of the mean-field models can be understood as hierarchical embeddings of a finite volume V into larger volumes with the surrounding spins, see Fig. 2. The density matrix of the spins from the l th shell is $\rho_l = Z_{l-1}^{m_l} / \langle Z_{l-1}^{m_l} \rangle_{\lambda_l}$. We can now perform the thermodynamic limit of the volume $V_{l-1} \rightarrow \infty$ with no surrounding spins or together with the next shell $V_l \rightarrow \infty$. If the system in the volume V_{l-1} is ergodic, the thermodynamic limit should not depend on the behavior of the surrounding spins. Here is the core of the problems and instabilities of simple mean-field solutions of spin-glass models. They break ergodicity in the whole low-temperature phase and the replica-symmetry breaking is a mathematical representation of ergodicity breaking. The hierarchical construction is a way to incorporate the influence of the thermal bath on the original spin systems in the thermodynamic limit.

2.4 Continuous limit: Parisi solution

The free energy $f_K(q; \Delta\chi_1, \dots, \Delta\chi_K, m_1, \dots, m_K)$ is an analytic function of its order parameters derived from K hierarchical embeddings. Due to the embedded structure of the partition sums Z_l , it is practically impossible to find explicit solutions for $K > 2$ in the whole temperature range of the glassy phase. Parisi found that in the case of the Ising spin glass the solutions for $K = 1$ and 2 are unstable and assumed that the same holds also for all finite K . He then performed the limit $K \rightarrow \infty$ and derived a continuous version of the infinitely replicated system by assuming $\Delta\chi_l = \Delta\chi/K \rightarrow dx$. Second and higher powers of $\Delta\chi_l$ with fixed index l are neglected [24, 44, 45]. When performing the limit $K \rightarrow \infty$ in representations of Eqs. (15) the free-energy functional can be represented as [37, 38]

$$f(q, X; m(x)) = -\frac{\beta}{4}(1 - q - X)^2 - \frac{1}{\beta} \ln 2 + \frac{\beta}{2} \int_0^X dx m(x) [q + X - x] - \frac{1}{\beta} \left\langle g(X, h + \eta\sqrt{q}) \right\rangle_\eta, \quad (18)$$

where $\langle X(\eta) \rangle_\eta = \int_{-\infty}^{\infty} \mathcal{D}\eta X(\eta)$. This free energy is only implicit since its interacting part $g(X, h)$ can be expressed only via an integral representation containing the solution itself

$$g(X, h) = \mathbb{E}_0(h; X, 0) \circ g_0(h) \equiv \mathbb{T}_x \exp \left\{ \frac{1}{2} \int_0^X dx [\partial_h^2 + m(x)g'(x; h + \bar{h})\partial_{\bar{h}}] \right\} g(h + \bar{h}) \Big|_{\bar{h}=0}, \quad (19a)$$

with $g(h) = \ln [\cosh \beta h]$. The “time-ordering” operator $\mathbb{T}_x$ orders products of x -dependent non-commuting operators from left to right in an x -decreasing succession. The exponent of the ordered exponential contains the function $g'(x; h) = \partial g(x; h)/\partial h$ for $x \in [0, X]$ and is not known when $g(x; h)$ is not known on the whole definition interval. This derivative can also be expressed via an ordered exponential

$$g'(X, h) = \mathbb{E}(h; X, 0) \circ g'_0(h) \equiv \mathbb{T}_x \exp \left\{ \int_0^X dx \left[\frac{1}{2} \partial_h^2 + m(x)g'(x; h + \bar{h})\partial_{\bar{h}} \right] \right\} g'_0(h + \bar{h}) \Big|_{\bar{h}=0}. \quad (19b)$$

It is an implicit but closed functional equation for the derivative $g'(x; h)$ on interval $[0, X]$ for a given function $m(x)$. We have to know the full dependence of this function on parameter x to evaluate the free energy with continuous replica-symmetry breaking. It is important to note that the free energy $f(q, X; m(x))$ defines a thermodynamic theory independently of the replica method within which it was derived. It means that we can look for equilibrium states of spin-glass models without the necessity to go through instabilities of the discrete hierarchical replica-symmetry breaking solutions.

Analogously, we can perform the limit $K \rightarrow \infty$ with $m_l - m_{l-1} = -\Delta m/K \rightarrow -dm$ in representation Eq. (15a). The minus sign is used, since we expect ordering $m_l < m_{l-1}$. Further

on, we use $\Delta\chi_l \rightarrow -x(m)dm$. The limiting continuous free energy can then be represented as [39]

$$f(q, \chi_1, m_1, m_0; x(m)) = -\frac{\beta}{4}(1 - q - \chi_1)^2 + \frac{\beta}{4} [m_1 (q + \chi_1)^2 - m_0 q^2] - \frac{\beta}{4} \int_{m_0}^{m_1} dm [q + \chi_1 - X(m)]^2 - \frac{1}{\beta} \langle g_1(m_0, h + \eta\sqrt{q}) \rangle_\eta, \quad (20)$$

where we denoted $X(m) = \int_m^{m_1} dm' x(m')$ and

$$g_1(m_0, h) = \mathbb{E}_0(m_0, m_1, h) \circ g_1(h) \equiv \mathbb{T}_m \exp \left\{ -\frac{1}{2} \int_{m_0}^{m_1} dm x(m) [\partial_h^2 + m g'_1(m; h + \bar{h}) \partial_{\bar{h}}] \right\} g_1(h + \bar{h}) \Big|_{\bar{h}=0}, \quad (21a)$$

and

$$g_1(h) \equiv g_1(m_1, h) = \frac{1}{m_1} \ln \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} e^{-\phi^2/2} [2 \cosh(\beta(h + \phi\sqrt{\chi_1}))]^{m_1}. \quad (21b)$$

This free energy better suits the case when the solution with continuous RSB peels off from a solution with one-level RSB or when the two solutions coexist. The space of order parameters is restricted to an interval $1 \geq m_1 \geq m_0$ and $1 \geq \chi_1 \geq X(m) \geq 0$.

If the free energy f_K is not locally stable and at least one of the stability conditions Eq. (17) is broken for all K , it is still a question whether the continuous limit $K \rightarrow \infty$ is locally stable. It can be shown that the continuous free energy $f(q, X; m(x))$ is marginally stable and fulfills the continuous version of stability conditions with equality [37]

$$1 = \langle \mathbb{E}(h_\eta; X, x) \circ [g''_\mu(x, h_\eta)^2] \rangle_\eta. \quad (22)$$

This equation is a consequence of the stationarity equation for the order-parameter function $m(x)$. We recall that the prime stands for the derivative with respect to the magnetic field and that the second derivative of $g_\mu(x, h)$ obeys an integral equation

$$\frac{\partial^2 g_\mu(x, h)}{\partial h^2} = \mathbb{E}(h; x, 0) \circ g''_\mu(h) + \int_0^x dy \mu(y) \mathbb{E}(h; x, y) \circ [g''_\mu(y, h)^2]. \quad (23)$$

The continuous free energy hence does not break ergodicity and is always marginally ergodic in the whole spin-glass phase. The ferromagnetic model is marginally ergodic only at the critical point, since the order parameter makes the ordered phase ergodic.

Both continuous free energies Eq. (18) and (20) were derived as the limit of the number of replica hierarchies $K \rightarrow \infty$ where the distance between the neighboring hierarchies is infinitesimal, that is $\Delta\chi_l \propto K^{-1}$ and $\Delta\chi_l/\Delta m_l < \infty$ for each $l \leq K$. Representation (20) is, however, more general, since it does not assume $\chi_K \rightarrow 0$ in the limit $K \rightarrow \infty$ as the free energy Eq. (18) does. On the other hand, the condition $\chi_K \rightarrow 0$ guarantees thermodynamic homogeneity of the resulting free energy.

The continuous free energies were derived for the Ising spin glass but they can be straightforwardly generalized to other spin-glass models. The symmetry of the order parameters has to be adapted and the input single-site free energy g or g_1 must be appropriately modified [38, 46].

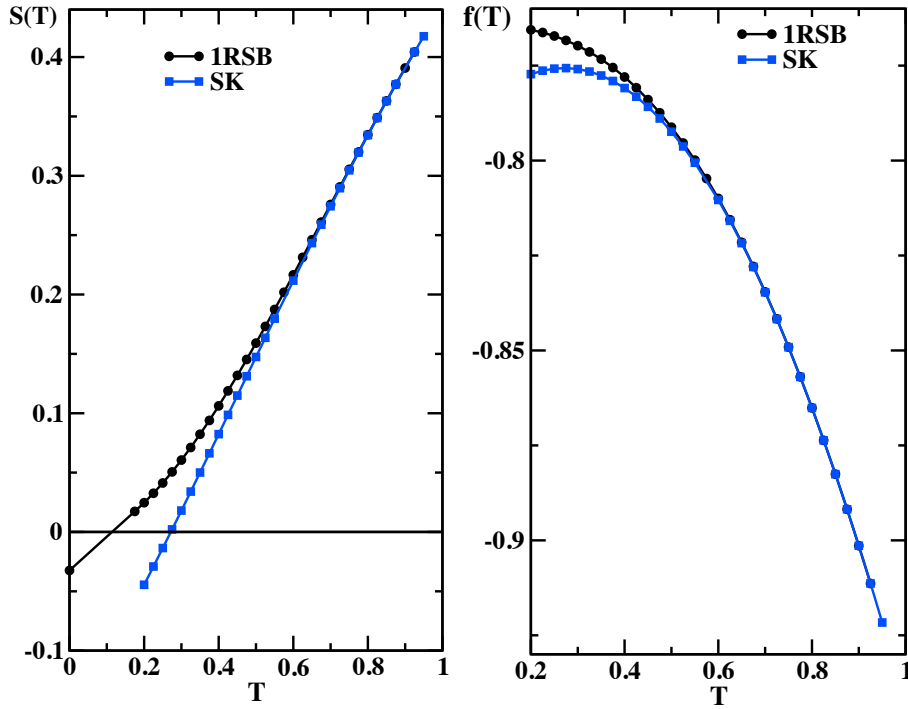


Fig. 3: Entropy (left panel) and free energy (right panel) in the glassy phase of the Sherrington-Kirkpatrick model at zero magnetic field for the replica-symmetric and 1RSB solutions. Improvements at very low temperatures are evident, in particular, the unphysical negative value of entropy is significantly reduced.

3 Asymptotic solutions of mean-field models: K -level replica symmetry breaking

3.1 Ising glass

The Sherrington-Kirkpatrick model is the paradigm for the mean-field theory of spin glasses. It is this model for which Parisi derived a free energy with a continuous replica-symmetry breaking [23–25]. It was also later proved that the hierarchical scheme of replica-symmetry breaking covers the exact equilibrium state [27, 28]. The rigorous proof does not, however, tell us whether the equilibrium state is described only by a finite number of replica hierarchies or a continuous limit is needed. Only a few years ago we resolved the hierarchical free energy $f_K(q; \Delta\chi_1, \dots, \Delta\chi_K, m_1, \dots, m_K)$ for arbitrary K via the asymptotic expansion below the transition temperature in a small parameter $\theta = 1 - T/T_c$ [47]. Only this asymptotic solution was able to resolve the question of the structure of the equilibrium state, at least close to the transition temperature.

3.1.1 Zero magnetic field

The order parameters in the spin-glass phase of the general K -level free energy from Eqs. (15) cannot be solved explicitly in the whole temperature range unless we resort to solutions with only a few hierarchical levels ($K = 1, 2$). The replica-symmetric and 1RSB solutions in the

whole low-temperature phase are plotted in Fig. 3, where one can see the improvements of the replica-symmetry breaking scheme in free energy and entropy. One has to go to higher replica hierarchies. The only chance to analyze the behavior of the entire hierarchical construction with an arbitrary number of hierarchies is to expand the solution near the critical point where the order parameters are small.

The strategy to solve the stationarity equations for the hierarchical free energy asymptotically near the critical temperature is to expand the partition function into powers of the small order parameters and to restrict the solution only to a functional subspace generated by a fixed polynomial expansion. We first use such an expansion to derive the leading asymptotic limit of the equations for the physical parameters $q, \Delta\chi_l$. At this stage we do not need to assume smallness of the geometric parameters. Smallness of $m_l, l = 1, \dots, K$ at zero magnetic field is utilized later on when deriving the asymptotic form of mean-field equations for them.

After a rather lengthy and tedious calculation, one finds the leading order of the order parameters [47]

$$\Delta\chi_l^K \doteq \frac{2}{2K+1} \theta, \quad (24a)$$

$$m_l^K \doteq \frac{4(K-l+1)}{2K+1} \theta, \quad (24b)$$

$$q^K \doteq \frac{1}{2K+1} \theta. \quad (24c)$$

The result proves that the limit $K \rightarrow \infty$ leads indeed to the Parisi continuous replica-symmetry breaking, since all the order parameters are of order K^{-1} . Each solution with a finite number of replica generations is unstable

$$A_l^K = -\frac{4}{3} \frac{\theta^2}{(2K+1)^2} < 0, \quad (24d)$$

from which it follows that the Parisi solution is the true equilibrium state in the glassy phase of the Sherrington-Kirkpatrick model near the transition point.

Other physical quantities in the asymptotic limit are the Edwards-Anderson parameter defined as $Q^K = q + \sum_{l=1}^K \Delta\chi_l^K$ with its asymptotics

$$Q^K \doteq \theta + \frac{12K(K+1)+1}{3(2K+1)^2} \theta^2, \quad (25a)$$

the local spin susceptibility

$$\chi_T = \beta \left(1 - Q^K + \sum_{l=1}^K m_l \Delta\chi_l \right) \doteq 1 - \frac{\theta^2}{3(2K+1)^2} \quad (25b)$$

and the free energy difference to the paramagnetic state

$$\Delta f \doteq \left(\frac{1}{6} \theta^3 + \frac{7}{24} \theta^4 + \frac{29}{120} \theta^5 \right) - \frac{1}{360} \theta^5 \left(\frac{1}{K} \right)^4. \quad (25c)$$

Differences between different levels of RSB manifest themselves in the free energy first in fifth order to which one had to expand the free energy in the order parameters.

3.1.2 Non-zero magnetic field

The glassy phase in the Sherrington-Kirkpatrick model also exists in an arbitrarily strong applied magnetic field. The transition boundary separating the paramagnetic from the glassy phase is the de Almeida-Thouless line. It is defined by the vanishing of the stability function $\Lambda = 0$ of Eq. (9). The hierarchical solution behaves in an external magnetic field differently from the rotationally invariant case. The physical order parameter q and the geometric order parameters m_l remain finite at the transition to the paramagnetic phase, and only the differences of the overlap susceptibilities $\Delta\chi_l$ are the genuine small parameters controlling the expansion around the critical point. A small expansion parameter below the AT line can be chosen as

$$\alpha = \beta^2 \langle (1 - t_0^2)^2 \rangle_\eta - 1 > 0, \quad (26)$$

where we denoted $t_0 = \tanh \beta (h + \eta \sqrt{q_0})$, and $q_0 = \langle t_0^2 \rangle_\eta$.

One must first expand the SK parameter q to the two lowest nontrivial orders in α . The solution is then used to determine the lowest asymptotic order of χ_l and m_l . We obtain $m_l = m_1 + O(\alpha)$ and $\chi_1 = \sum_{l=1}^K \Delta\chi_l^K$ with [48]

$$m_1 = \frac{2 \langle t_0^2 (1 - t_0^2)^2 \rangle_\eta}{\langle (1 - t_0^2)^3 \rangle_\eta} \quad (27)$$

and

$$\chi_1 = \frac{\alpha}{2\beta^2 m_1} \frac{1}{1 - 3\beta^2 \langle t_0^2 (1 - t_0^2)^2 \rangle_\eta} + O(\alpha^2). \quad (28)$$

These two parameters do not depend on the number of hierarchical levels used. Parameter m_1 is of order unity even at the boundary of the spin-glass phase (AT line), where the small parameter α vanishes.

To disclose the leading asymptotic behavior of each separate parameter $\Delta\chi_l$ and m_l for $l = 1, \dots, K$, we must go beyond the leading orders in the parameters m and χ_1 . It is first the fourth order in α from which we find that $\Delta\chi_l \doteq \chi_1/K$ and

$$m_l^K \doteq m_1 + \frac{K+1-2l}{K} \Delta m \quad (29)$$

where we added a superscript to specify the number of hierarchical levels used to determine the order parameters χ_l, m_l . Further on, we introduced a parameter independent of the number of hierarchies $\Delta m = m_1^2 - m_2^2$. This parameter has an explicit asymptotic representation

$$\Delta m \doteq \frac{\beta^2 \chi_1 \left\langle (1 - t_0^2)^2 \left(2(1 - 3t_0^2)^2 + 3(t_0^2 - 1)m(8t_0^2 + (t_0^2 - 1)m) \right) \right\rangle_\eta}{\langle (1 - t_0^2)^3 \rangle_\eta} \quad (30)$$

Both parameters χ_1 and Δm are linearly proportional to α . The former, however, exists already in 1RSB, while the latter first emerges in 2RSB. Since they do not depend on the number of hierarchies used and the latter determines a uniform distribution of parameters m_l for $l = 3, \dots, K$, we demonstrated that all characteristic features of the asymptotic solution near the

AT instability line are contained already in 2RSB. What was, however, highly nontrivial was to unveil equidistant distributions of both parameters χ_l and m_l .

The stability conditions

$$\Lambda_K = -\frac{2\beta^2}{3K^2} \frac{\chi_1 \Delta m}{m+2} \quad (31)$$

indicate that in an applied magnetic field the equilibrium state is also described by the continuous replica-symmetry breaking $K \rightarrow \infty$.

3.2 Potts glass

The Potts model with p states reduces to the Ising model for $p = 2$, but differs from it for $p > 2$ in that it breaks the spin-reflection symmetry. This property was used to argue that the Parisi scheme fails to describe the equilibrium state [49]. It had been long believed that it is the one-level replica-symmetry breaking that determines the equilibrium state below the transition temperature [50]. The Potts glass displays a discontinuous transition to the replica-symmetry-broken state for $p > 4$ [51]. Discontinuous transitions do not allow us to use an asymptotic expansion in a small parameter below the transition temperature. It is, nevertheless, possible to test the ordered phase of the Potts glass for $2 < p < 4$. We did it in Refs. [46, 52] and found an unexpected behavior.

Studying the discrete replica-symmetry breaking we found two 1RSB solutions with the same geometric parameter

$$m \doteq \frac{p-2}{2} + \frac{36-12p+p^2}{8(4-p)}\theta. \quad (32)$$

One non-trivial 1RSB solution then leads to order parameters

$$q^{(1)} \doteq 0, \quad (33a)$$

$$\Delta\chi^{(1)} \doteq \frac{2}{4-p}\theta + \frac{228-96p+p^2}{6(4-p)^3}\theta^2 \quad (33b)$$

while the second one has both parameters nonzero

$$q^{(2)} \doteq \frac{-12+24p-7p^2}{3(4-p)^2(p-2)}\theta^2, \quad (34a)$$

$$\Delta\chi^{(2)} \doteq \frac{2}{4-p}\theta - \frac{360-204p-6p^2+13p^3}{6(4-p)^3(p-2)}\theta^2. \quad (34b)$$

Both the solutions have the same asymptotic free energy to the fifth asymptotic order

$$\begin{aligned} \frac{\beta}{p-1} f_{1RSB} \doteq & \frac{\theta^3}{3(4-p)} + \frac{(p(11p-102)+204)\theta^4}{12(4-p)^3} \\ & - \frac{(p(p((18744-1103p)p-120648)+325728)-317232)\theta^5}{720(4-p)^5}. \end{aligned} \quad (35)$$

We can see that the asymptotic expansion with small parameters q and $\Delta\chi$ already breaks down at $p = 4$ above which we expect a discontinuous transition from the paramagnetic to a 1RSB

state at $T_0 > T_c = 1$. Note that a transition to the replica-symmetric solution $q > 0$, $\Delta\chi = 0$ is continuous up to $p = 6$.

The 1RSB solution has a higher free energy than the replica-symmetric one. The difference is of order θ^3 ,

$$f_{1RSB} - f_{RS} \doteq \frac{(p-2)^2(p-1)\theta^3}{3(4-p)(6-p)^2}. \quad (36)$$

The two stationary states of the 1RSB free energy behave differently as a function of parameter p . The former solution is physical for all values of p unlike the latter that becomes unphysical for $p > p^* \approx 2.82$ where $q^{(2)}$ from Eq. (34a) turns negative. It is also the region of the parameter p where the first solution is locally stable as can be seen from the stability function

$$A_0^1 \doteq \frac{\theta^2(p-1)}{6(4-p)^2} (7p^2 - 24p + 12) > 0. \quad (37)$$

That is why the solution with $q = 0$ was assumed to be the true equilibrium and a solution with a continuous replica-symmetry breaking had not been expected to exist. We, however, found that there is a Parisi-like solution even in the region of stability of the solution from Eq. (33). The second 1RSB solution is unstable and decays to solutions with higher numbers of replica hierarchies as

$$q^K \doteq -\frac{1}{3K^2} \frac{12 - 24p + 7p^2}{(4-p)^2(p-2)} \theta^2, \quad (38a)$$

$$\Delta\chi_l^K \doteq \frac{1}{K} \frac{2}{(4-p)} \theta, \quad (38b)$$

$$m_l^K \doteq \frac{p-2}{2} + \frac{2}{4-p} \left[3 + \frac{3}{2}p - p^2 + \left(3 - 6p + \frac{7}{4}p^2 \right) \frac{2l-1}{2K} \right] \theta. \quad (38c)$$

We can see that the K RSB solution behaves unphysically in the same way that the second 1RSB solution does. The averaged square of the local magnetization is negative for $p > p^*$ where the first 1RSB solution is locally stable. Negativity of q means that local magnetizations are imaginary and the solution is unphysical. This deficiency, however, decreases with increasing number of spin hierarchies and disappears in the limit $K \rightarrow \infty$. It means that the resulting solution with a continuous replica-symmetry breaking shows no unphysical behavior. It is analogous to the negativity of entropy in the low-temperature solutions of K RSB approximations of the Sherrington-Kirkpatrick model.

The Potts glass hence shows a degeneracy for $p^* < p < 4$ with a marginally stable solution continuously breaking the replica symmetry and a locally stable one-level replica-symmetry breaking. To decide which one is the true equilibrium state one has to compare free energies. The difference of the continuous free energy f_c and that of the K RSB solution is

$$\beta(f_c - f_{KRSB}) \doteq \frac{(p-1)(p(7p-24)+12)^2\theta^5}{720K^4(4-p)^5} \quad (39a)$$

and that of the replica-symmetric one reads

$$\beta(f_c - f_{RS}) \doteq \frac{(p-1)(p-2)^2\theta^3}{3(4-p)(6-p)^2}. \quad (39b)$$

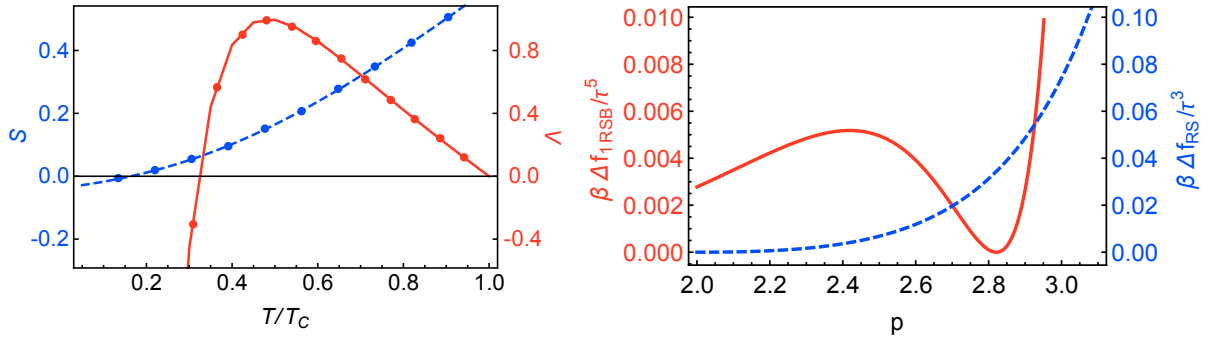


Fig. 4: Left panel: Entropy S (left scale, dashed line) and local stability Λ (right scale, solid line) of the 1RSB solution from Eq. (33) of the 3-state Potts glass. The solution becomes locally unstable at $\approx 0.33T_c$ and entropy negative at $T \approx 0.16T_c$. Right panel: The free-energy differences between the solution with continuous replica-symmetry breaking and 1RSB (left scale, solid line) and RS (right scale, dashed line) solutions.

We see that the solution with the continuous RSB has the highest free energy as the true equilibrium state should have for geometric factors $m < 1$. In this situation entropy reaches a minimum and the free energy a maximum in the phase space of the order parameters. The locally stable 1RSB solution becomes unstable at lower temperatures and entropy turns negative at very low temperatures as demonstrated on the 3-state model in Fig. 4. This leads us to the conclusion that the Parisi solution with a continuous replica-symmetry breaking represents the equilibrium state for the Potts glass with $p < 4$.

3.3 p -spin glass

The spin model generalized to random interactions connecting p spins, the p -spin glass, was used to simulate the dynamical transition in real glasses [53,54]. This model, analogously to the Potts glass, generalizes the Ising spin glass to $p > 2$ and allows one to study the behavior of the equilibrium state as a function of parameter p . In particular, the limit $p \rightarrow \infty$ is accessible [10] and is exactly solvable. It coincides with the random energy model of Derrida [11, 12]. For this reason the p -spin glass was also intended to be used to study and understand the genesis of the Parisi free energy when studying the asymptotic limit $p \rightarrow \infty$.

To cover both the boundary solutions $p = 2$ and $p = \infty$ we have to mix up the one-level RSB scheme and the Parisi continuous RSB. Such a free energy density of the mean-field p -spin glass reads [39]

$$\begin{aligned}
 f_T^{(p)}(q, \chi_1, \mu_1, \mu_0; x(\mu)) = & -\frac{\beta}{4} \left[1 - p(q + \chi_1) + (p-1)(q + \chi_1)^{p/(p-1)} \right] \\
 & + \frac{p-1}{4} \left[\mu_1(q + \chi_1)^{p/(p-1)} - \mu_0 q^{p/(p-1)} \right] \\
 & - \frac{p-1}{4} \int_{\mu_0}^{\mu_1} d\mu [q + \chi_1 - X(\mu)]^{p/(p-1)} - g_1(\mu_0, h), \quad (40)
 \end{aligned}$$

with

$$g_1(\mu_0, h) = \mathbb{E}_1(\mu_1, \mu_0, h) \circ [g_1(h)] \\ \equiv \mathbb{T}_\mu \exp \left\{ -\frac{p}{4} \int_{\mu_0}^{\mu_1} d\mu x(\mu) \left[\partial_h^2 + \mu g_1'(\mu, h + \bar{h}) \partial_{\bar{h}} \right] \right\} g_1(h + \bar{h}) \Big|_{\bar{h}=0}, \quad (41a)$$

where $X(\mu) = \int_{\mu}^{\mu_1} d\mu' x(\mu')$. The generating free energy is

$$g_1(h) = \frac{1}{\mu_1} \ln \int_{-\infty}^{\infty} \frac{d\phi}{\sqrt{2\pi}} e^{-\phi^2/2} \left[2 \cosh \left(\beta(h + \phi \sqrt{p\chi_1/2}) \right) \right]^{\mu_1/\beta}. \quad (41b)$$

We rescaled the function $m \rightarrow \mu = \beta m$. If $\mu = 0$ or $\mu_1 = \mu_0$, the free energy $f^{(p)}$ reduces to the 1RSB approximation. On the other hand, if $\mu_0 = 0$ or $\mu_0 = \beta$, the free energy $f^{(p)}$ coincides with that of the Parisi solution with continuous replica-symmetry breaking.

The p -spin glass can be used to investigate analytically not only the $p \rightarrow \infty$ limit but also the $T \rightarrow 0$ limit. In this limit simple solutions of mean-field models lead to negative entropy. It is easy to calculate the zero-temperature entropy in the 1RSB solution. We obtain

$$S_0(h) \propto -\frac{p(p-1)}{8} \left[\frac{\exp\{-\mu_1^2 p \chi_1/4\}}{\sqrt{\pi p \chi_1}} \frac{\exp\{-h^2/p \chi_1\}}{2CH_\mu(h)} \right]^2, \quad (42)$$

where we used the following notation

$$2CH_\mu(h) = e^{\mu_1 h} E_\mu^{(p)}(-h) + e^{-\mu_1 h} E_\mu^{(p)}(h), \quad (43a)$$

$$E_\mu^{(p)}(h) = \int_{h/\sqrt{p\chi_1/2}}^{\infty} \frac{d\phi}{\sqrt{2\pi}} e^{-(\phi - \mu_1 \sqrt{p\chi_1/2})^2/2}. \quad (43b)$$

Negativity of the low-temperature entropy indicates that 1RSB cannot produce a stable ground state for arbitrary $p < \infty$. The negativity of the entropy decreases with increasing p , see Fig. 5, but only if a condition $\mu_1^2 p \chi_1 = \infty$ is fulfilled; the 1RSB solution ($\mu_1 > 0$) leads to zero entropy at zero temperature. Nonnegative entropy is a necessary condition for the physical consistency of the low-temperature solution. It then means that the low-temperature equilibrium state for $p < \infty$ must contain the Parisi continuous order-parameter function $x(\mu)$ with $\beta > \mu_1 > \mu_0$. It can also be seen from the asymptotic free energy for $p \rightarrow \infty$ that reads

$$f_T^{(p \rightarrow \infty)}(q, \chi_1, \mu_1) = -\frac{1}{4T} [1 - (q + \chi_1)(1 - \ln(q + \chi_1))] - \frac{1}{\mu_1} \ln [2 \cosh(\mu_1 h)] \\ - \frac{\mu_1}{4} [\chi_1 - (q + \chi_1) \ln(q + \chi_1)] - \frac{\mu_1 q}{4} [\ln q + p(1 - \tanh^2(\mu_1 h))], \quad (44)$$

giving the leading-order solution for the variational parameters χ_1, μ_1 , and q . The first two parameters are of order one while the latter is exponentially small for large p ,

$$\chi_1 = 1 - q, \quad (45a)$$

$$q = \exp\{-p(1 - \tanh^2(\mu_1 h))\}, \quad (45b)$$

$$\mu_1 = 2\sqrt{\ln [2 \cosh(\mu_1 h)] - h \tanh(\mu_1 h)}. \quad (45c)$$

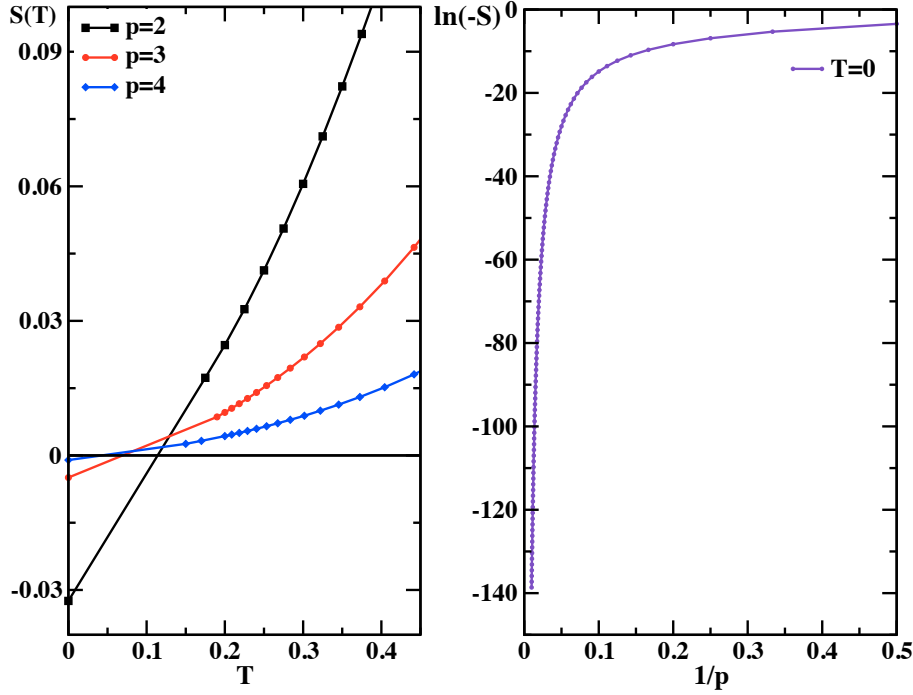


Fig. 5: Entropy of the 1RSB solution of the p -spin glass. Temperature dependence for different values of p (left panel). Logarithm of the negative part of the entropy at zero temperature as a function of $1/p$ (right panel), from Ref. [39].

The above nontrivial solution holds only if $\beta > 2\sqrt{\ln[2 \cosh(\beta h)] - h \tanh(\beta h)}$ (low-temperature phase), otherwise $\mu_1 = \beta$ and $\chi_1 + q = 0$ (high-temperature phase). To derive an equation for the order-parameter function $x(\mu)$ one needs to include the next-to-leading order contributions. To go beyond the leading asymptotic order one can use the Landau-type theory for the order-parameter function developed in Ref. [38]. Note that the asymptotic solution of Eqs. (45) with $\mu(x) = 0$ suffers from a negative entropy as can be seen from Eq. (42) and is plotted in Fig. 5. Note that the transition to the ordered phase in the p -spin glass is discontinuous and hence, an asymptotic expansion below the transition temperature is not applicable. Only an asymptotic expansion $p \rightarrow \infty$ makes sense.

4 Conclusions

The mean-field models of spin glasses have equilibrium states demanding nonstandard tools to describe them quantitatively. Moreover, the replica trick standardly used is suspicious, and it is difficult to understand and give a physical meaning to the order parameters and the functions with the replica variables in this construction. The standard way to reach the equilibrium states is to use the replica trick, derive representations with the replica variables and then try to give the results a physical meaning. The reason for using the replicas remains nevertheless veiled in this way and the physical meaning of the replica-symmetry breaking unclear [13, 45]. One has to analyze the physical reasons for instabilities and negative entropy in non-replicated mean-

field solutions to find the genuine reason why the true equilibrium of the mean-field spin-glass models cannot be reached without replicas.

The most prominent feature of the mean-field spin-glass models is that the combination of randomness and frustration of the spin exchange leads to *ergodicity breaking* that is not accompanied by any symmetry breaking in the generic Hamiltonian. To restore ergodicity in the glassy phase one can use replications of the phase space of the dynamical variables. A small inter-replica interaction is introduced as a perturbation and the linear response of the system is calculated. If the linear response (replica independence or symmetry) is broken, replica variables influence the behavior of the original system, and inter-replica interactions become non-zero. The replication is then used hierarchically until one restores at least local thermodynamic homogeneity. The principal step in this procedure is to select an adequate symmetry-breaking of the replicated variables so as to make thermodynamic potentials analytic functions of the originally integer replication index. Only then it is possible to test and restore thermodynamic homogeneity and ergodicity. If infinitely many replications are needed, the Parisi solution with the continuous replica-symmetry breaking is obtained. Real replicas allow us to restore ergodicity in hierarchical steps by breaking successively the independence of the replicated spaces.

The solution of the full mean-field models of spin glasses is unreachable. One has to resort to approximations or perturbative expansions. We applied the real replicas to the mean-field Ising, p -state Potts, and p -spin glass models and calculated their asymptotic solutions below the transition temperature to the glassy phase. We thereby demonstrated that the Ising spin glass below the transition point is described by a solution with infinitely many replica hierarchies with the continuous order-parameter function of Parisi. While the Ising spin glass is known to have continuously broken replica symmetry in the equilibrium state, the Potts and p -spin glasses allow for locally stable solutions with a one-level discrete replica-symmetry breaking. Since the solution with the continuous replica-symmetry breaking exists independently of the stability of the solutions with finitely many replica hierarchies, the continuous RSB and 1RSB coexist in the p -state Potts and the p -spin glass models. In both cases for $p < \infty$ the 1RSB state leads to negative entropy at very low temperatures and the ultimate equilibrium state for the mean-field spin-glass models breaks the replica symmetry in a continuous form as suggested by Parisi in the Ising model. Our analysis indicates that a continuous RSB is indispensable to keep entropy non-negative down to zero temperature. Spin reflection symmetry is hence not substantial for the existence of a solution with continuous replica-symmetry breaking as was previously assumed.

Acknowledgement

I would like to thank my collaborators A. Klíč and A. Kauch for their numerical calculations and the figures I used in these lecture notes. I also thank the Fulbright Commission for financing my stay at Louisiana State University in Baton Rouge where I compiled this contribution.

References

- [1] M.A. Ruderman and C. Kittel, *Physical Review* **96**, 99 (1954)
- [2] T. Kasuya, *Progress of Theoretical Physics* **16**, 45 (1956)
- [3] K. Yosida, *Physical Review* **106**, 893 (1957)
- [4] S.F. Edwards and P.W. Anderson, *Journal of Physics F: Metal Physics* **5**, 965 (1975)
- [5] V. Cannella and J.A. Mydosh, *Physical Review B* **6**, 4220 (1972)
- [6] M. Gabay and G. Toulouse, *Physical Review Letters* **47**, 201 (1981)
- [7] D.M. Cragg, D. Sherrington, and M. Gabay, *Physical Review Letters* **49**, 158 (1982)
- [8] R.B. Potts,
 Mathematical Proceedings of the Cambridge Philosophical Society **48**, 106 (1952)
- [9] F.Y. Wu, *Reviews of Modern Physics* **54**, 235 (1982)
- [10] D. Gross and M. Mézard, *Nuclear Physics B* **240**, 431 (1984)
- [11] B. Derrida, *Physical Review Letters* **45**, 79 (1980)
- [12] B. Derrida, *Physical Review B* **24**, 2613 (1981)
- [13] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro,
 Journal de Physique, France **45**, 843 (1984)
- [14] D. Sherrington and S. Kirkpatrick, *Physical Review Letters* **35**, 1792 (1975)
- [15] S. Kirkpatrick, *Physical Review B* **16**, 4630 (1977)
- [16] S. Kirkpatrick and D. Sherrington, *Physical Review B* **17**, 4384 (1978)
- [17] J.R.L. de Almeida and D.J. Thouless,
 Journal of Physics A: Mathematical and General **11**, 983 (1978)
- [18] J.L. van Hemmen and R.G. Palmer,
 Journal of Physics A: Mathematical and General **12**, 563 (1979)
- [19] D.J. Thouless, P.W. Anderson, and R.G. Palmer, *Philosophical Magazine* **35**, 593 (1977)
- [20] V. Janiš, *Physical Review B* **74**, 054207 (2006)
- [21] G. Parisi, *Physical Review Letters* **43**, 1754 (1979)
- [22] G. Parisi, *Physics Letters A* **73**, 203 (1979)

- [23] G. Parisi, Journal of Physics A: Mathematical and General **13**, L115 (1980)
- [24] G. Parisi, Journal of Physics A: Mathematical and General **13**, 1101 (1980)
- [25] G. Parisi, Journal of Physics A: Mathematical and General **13**, 1887 (1980)
- [26] G. Parisi, Physics Reports **67**, 25 (1980)
- [27] F. Guerra, Communications in Mathematical Physics **233**, 1 (2003)
- [28] M. Talagrand, Annals of Mathematics **163**, 221 (2006)
- [29] K. Binder and A.P. Young, Reviews of Modern Physics **58**, 801 (1986)
- [30] F. Zamponi, preprint arXiv:1008.4844 (2010)
- [31] M. Mézard, G. Parisi, and M. Virasoro: *Spin glasses and beyond* (World Scientific, Singapore, 1987)
- [32] K.H. Fischer and J.A. Hertz: *Spin Glasses* (Cambridge University Press, 1991)
- [33] V. Dotsenko: *Introduction to the replica theory of disordered statistical systems* (Cambridge University Press, 2001)
- [34] H. Nishimori: *Statistical physics of spin glasses and information processing: an introduction* (Oxford University Press, 2001)
- [35] C.D. Dominicis and I. Giardinà: *Random fields and spin glasses: a field theory approach* (Cambridge University Press, 2006)
- [36] V. Janiš, Physical Review B **71**, 214403 (2005)
- [37] V. Janiš, Physical Review B **77**, 104417 (2008)
- [38] V. Janiš, A. Kauch, and A. Klíč, Physical Review B **87**, 054201 (2013)
- [39] V. Janiš, A. Kauch, and A. Klíč, Phase Transitions **88**, 245 (2015)
- [40] G.D. Birkhoff, Proceedings of the National Academy of Sciences of the United States of America **17**, 656 (1931)
- [41] L.E. Reichl: *A Modern Course in Statistical Physics* (University of Texas Press, Austin, 2nd ed., 1980)
- [42] R. Palmer, Advances in Physics **31**, 669 (1982)
- [43] F.T. Bantilan and R.G. Palmer, Journal of Physics F: Metal Physics **11**, 261 (1981)
- [44] B. Duplantier, Journal of Physics A: Mathematical and General **14**, 283 (1981)

- [45] G. Parisi, *Physical Review Letters* **50**, 1946 (1983)
- [46] V. Janiš and A. Klíč, *Journal of Physics-Condensed Matter* **23**, 022204 (2011)
- [47] V. Janiš and A. Klíč, *Physical Review B* **74**, 054410 (2006)
- [48] V. Janiš, A. Klíč, and M. Ringel,
Journal of Physics A: Mathematical and General **41**, 324004 (2008)
- [49] P. Goldbart and D. Elderfield, *Journal of Physics C: Solid State Physics* **18**, L229 (1985)
- [50] G. Cwilich and T.R. Kirkpatrick,
Journal of Physics A: Mathematical and General **22**, 4971 (1989)
- [51] D.J. Gross, I. Kanter, and H. Sompolinsky, *Physical Review Letters* **55**, 304 (1985)
- [52] V. Janiš and A. Klíč, *Physical Review B* **84**, 064446 (2011)
- [53] T.R. Kirkpatrick and D. Thirumalai, *Physical Review B* **36**, 5388 (1987)
- [54] T.R. Kirkpatrick and P.G. Wolynes, *Physical Review B* **36**, 8552 (1987)

9 The Slave-Boson Approach to Correlated Fermions

Raymond Frésard

Laboratoire CRISMAT, UMR CNRS-ENSICAEN 6508
6, Bld. Maréchal Juin, 14050 Caen CEDEX 4 , France

Contents

1	Introduction	2
2	Slave-boson representations	4
2.1	Barnes's representation	4
2.2	Kotliar and Ruckenstein representation	6
2.3	Spin-rotation-invariant representation	7
2.4	Multi-band systems	8
3	Gauge symmetry and radial slave-boson fields	8
4	Saddle-point approximations	10
4.1	Saddle-point approximation to the Barnes representation	10
4.2	Saddle-point approximation to the Kotliar and Ruckenstein representation . . .	12
4.3	Saddle-point approximation to the multi-band Hubbard model	15
4.4	A concrete example	19
4.5	Magnetic order in the Anderson lattice model	22
5	Fluctuation corrections to the saddle-point approximation: SRI representation of the Hubbard model	23
5.1	Magnetic and striped phases	25
6	Extended Hubbard model	26
6.1	Saddle-point approximation to the extended Hubbard model	26
6.2	Landau parameters	27
7	Summary	30

1 Introduction

The immense and steadily increasing field of strongly correlated electrons has emerged as a central theme of many-body physics over the past three decades (for a review see [1]). Among them, the so-called heavy-fermion metallic compounds [2] and cuprate superconductors [3] have received particular attention. It is acknowledged that fully accounting for their properties is a challenging task, but it is believed that their key properties are embodied in model Hamiltonians, such as the Anderson or Kondo lattice models in the former case, and in the Hubbard model or possibly the t - J -model in the latter one [4]. The difficulty in solving these models is rooted in the fact that conventional many-body perturbation theory (including infinite resummations), does not work in these cases. This failure is obvious in lattice models with on-site repulsion U exceeding the band width D .

Take the Hubbard model with large on-site repulsion U , where each lattice site can either be empty (state $|0\rangle$), singly occupied ($|\uparrow\rangle, |\downarrow\rangle$), or doubly occupied ($|2\rangle$). The dynamics of an electron will be very different according to whether it resides on a singly or doubly occupied site. For large U , the doubly occupied states will be pushed far up in energy and will only marginally contribute to the low-energy physics. This leads effectively to a projection of the Hilbert space onto a subspace devoid of doubly occupied states. It turns out to be difficult to effect the projection within conventional many-body theory, as was realized early on in the context of the magnetic impurity problem. Indeed, this difficulty is at the heart of the single-impurity Kondo problem, for which a sound physical picture and quantitative analytical and numerical methods of solution have been developed over a period of 40 years [5]. We will discuss impurity models briefly in a later section. More details can be found in the review [6].

A powerful technique for describing the projection in Hilbert space is the method of auxiliary particles (slave bosons, pseudofermions [7–12]): One assigns an auxiliary field or particle to each of the four states $|0\rangle, |\downarrow\rangle, |\uparrow\rangle, |2\rangle$ at a given lattice site (considering one strongly correlated orbital per site). The Fermi character of the electrons requires that two of the auxiliary particles are fermions, e.g., the ones representing $|\downarrow\rangle, |\uparrow\rangle$ and the remaining two are bosons. Introducing new particles for the states $|0\rangle, |2\rangle$ allows one to express the projection to the Hilbert space of states without double occupancy as $n_0 + n_\uparrow + n_\downarrow = 1$, where n_α are the occupation numbers of the states $|\alpha\rangle$; i.e., each lattice site is either empty or singly occupied. There are various ways of defining auxiliary particles for a given problem. It is wise to choose the one that is best adapted to the physical properties of the system.

Compared to alternative ways of performing the projection, the auxiliary-particle method has the advantage of allowing one to use the machinery of quantum field theory, i.e. Wick's theorem, diagrammatic perturbation theory and infinite resummations of diagrams, provided the constraint can be incorporated in a satisfactory way.

Historically, auxiliary particle representations were first introduced in the context of spin models. Spin operators may be represented by Bose operators (Holstein-Primakoff [7], Schwinger [8]) or in the case of spin 1/2 (and with additional complications for higher spins as well) by Fermi operators (Abrikosov [9], Coqblin-Schrieffer [10]). Electron operators necessarily in-

volve a combination of auxiliary fermion and boson operators. The simplest such representation was proposed for the Anderson impurity problem by Barnes [11], and for lattice problems by Coleman [12]. A more complex representation of electron operators, incorporating the result of the Gutzwiller approximation [13] on the slave-boson saddle-point approximation level was developed by Kotliar and Ruckenstein [14]. Generalizations of the latter to manifestly spin-rotation invariant form [15, 16] and to particle-hole and spin rotation invariant form [16] have also been proposed. Generalizations to multi-band Hubbard models have been introduced as well [17–20].

Quite generally, auxiliary particle theories have to deal with two problems: the treatment of the constraint and the approximate description of the dynamics. An accurate control of the constraint alone does not yet make a good theory! In fact, the latest attempts suggest that the price to pay to exactly implement constraints is to have to diagonalize the many-body Hamiltonian matrix [21, 22].

Besides, the effect of strong Coulomb interaction in systems with orbital degeneracy plays a prominent role. Such a situation is realized in almost all transition metals and transition-metal oxides. These systems contain d -electrons in cubic or trigonal environments, the crystal field can only partially lift the degeneracy of the d -bands, for instance down to two as is the case of V_2O_3 [23] or down to three for perovskites such as $LaTiO_3$. On top of high- T_c superconductivity, a whole series of application-oriented, fundamental properties of correlated electronic systems arose in recent years, in particular for transition-metal oxides. They include colossal magnetoresistance (see, e.g., [24]), transparent conducting oxides (see, e.g., [25]), high-capacitance heterostructures [26], and large thermopower (see, e.g., [27]), to quote a few. In addition, they also entail fascinating phenomena such as superconductivity at the interface of two insulators [28], peculiar magnetism in low-dimensional systems [29], high-temperature ferromagnetism in vanadate superlattices [30], all of them providing strong motivation to investigate these systems from the theory side.

Given the variety of systems and properties of interest, it is desirable to have an approximate scheme amenable to the computation of the desired quantities as functions of interaction strength and density in the thermodynamic limit. Slave-boson approaches showed a great potential toward that aim. While the solution of the Ising chain is the only example so far that could be solved exactly through slave-boson calculations [31], approximate approaches such as the self-consistent T-matrix approximation to the single-impurity Anderson model and the slave-boson saddle-point approximation to the Hubbard model have been widely used. Part of the success of the latter follows from the fact that it is variationally controlled in the large-dimensionality limit, and it is exact in the large-degeneracy limit. Further, it can be improved systematically by performing a loop-expansion around the saddle point.

In Section 2 we review the various slave-boson representations of the most prominent models. Section 3 is devoted to the gauge symmetry of the Barnes slave-boson representation of the single-impurity Anderson model and to the concept of a radial slave-boson field, which is shown in general to possess an exact, non-vanishing expectation value. The saddle-point approximation is applied to several models in Section 4. Fluctuation corrections, calculation of

the spin and charge autocorrelation functions as well as magnetic phases are addressed in Section 5. Recent applications to a Hubbard model extended by long-ranged Coulomb interactions are presented in Section 6, and the results are summarized in Section 7.

2 Slave-boson representations

The goal of a slave-boson (SB) representation is to describe an interacting fermionic system by means of an action that is bilinear in fermionic fields. In these frameworks one avoids Hubbard-Stratonovich decoupling the interaction terms, which typically shares the difficulties and limitations of perturbation theory. It necessitates introducing auxiliary fermionic fields, which will be denoted below by the doublet $\{f_\sigma\}$, and slave-boson fields, say $\{e, p, d\}$, in terms of which one needs to rewrite the physical electron operators $\{a_\sigma\}$. All of them satisfy canonical commutations. By doing so, one increases the number of degrees of freedom (DoF), implying that the auxiliary fields need to satisfy constraints to ensure a faithful representation of the original model. These constraints can be handled in the functional integral formalism [32]. A particularity of SB approaches is the apparition of radial slave boson fields: They are bosonic fields with their amplitude as sole DoF. Being phase-free, their exact expectation value may be finite in accordance with Elitzur's theorem, as discussed below.

A natural basis of the Hilbert space of electrons in a local orbital may be chosen as consisting of four states: two with single occupancy (representing a local spin 1/2) and the empty and doubly occupied states. Forming a doublet, the singly occupied states manifestly have fermionic character, while the remaining two states have bosonic character. Below, we shall create these states out of a vacuum state $|\text{vac}\rangle$, which is annihilated by any of the above auxiliary fields. These four states may then be created by fermionic or bosonic auxiliary operators. This may be achieved in a multitude of ways. We will concentrate here on the representations introduced by Barnes for the single-impurity Anderson model [11], by Kotliar and Ruckenstein [14], and by Wölfle *et al.* for the Hubbard model [15, 16], as well as an extension to multi-band systems [17].

2.1 Barnes's representation

The basic idea consists in locally decomposing the electronic excitations into spin and charge components. There are many different ways to achieve this goal. For instance, it could be reached by means of a suitable Hubbard-Stratonovich decoupling of the interaction term but would likely be limited to weak interaction. Instead, in his pioneering work, Barnes suggested representing the spin and charge degrees of freedom by fermionic and bosonic operators, respectively [11]. Being more numerous than the original (physical) operators, the auxiliary operators span a Fock space that is larger than the physical one. They therefore need to fulfill constraints for such a representation to be faithful. In fact, it can be shown that one constraint suffices. Specifically, Barnes considered the single-impurity Anderson model (SIAM):

$$H = \sum_{\vec{k}\sigma} \varepsilon_{\vec{k}} c_{\vec{k}\sigma}^\dagger c_{\vec{k}\sigma} + \varepsilon_f \sum_{\sigma} a_{\sigma}^\dagger a_{\sigma} + V \sum_{\vec{k}\sigma} \left(c_{\vec{k}\sigma}^\dagger a_{\sigma} + a_{\sigma}^\dagger c_{\vec{k}\sigma} \right) + U a_{\uparrow}^\dagger a_{\uparrow} a_{\downarrow}^\dagger a_{\downarrow}. \quad (1)$$

For the f electrons that are described by this model, the interaction strength is large and often treated in the $U \rightarrow \infty$ limit. To that aim, Barnes introduced auxiliary fermionic $\{f_\sigma\}$ and bosonic $\{e, d\}$ operators that satisfy canonical commutation relations. In terms of these, the physical electron operators a_σ read

$$a_\sigma = e^\dagger f_\sigma + \sigma f_{-\sigma}^\dagger d . \quad (2)$$

The a_σ operators obey the ordinary fermion anticommutation relations. Yet this property is not automatically preserved when using the representation Eq. (2), even when the fermionic and bosonic auxiliary operators obey canonical commutation relations. In addition, the constraint

$$Q \equiv e^\dagger e + \sum_\sigma f_\sigma^\dagger f_\sigma + d^\dagger d = 1 \quad (3)$$

must be satisfied. Eqs. (2-3) make for a faithful representation of the physical electron operator in the sense that both a_σ and its expression in terms of auxiliary particles Eq. (2) possess the same matrix elements in the physical Hilbert subspace with $Q = 1$. The above representation has been widely used, in particular in the $U \rightarrow \infty$ limit where the operator d (linked to double occupancy) drops out. One can implement the constraint by means of a functional integral representation. For example, for $U \rightarrow \infty$ the partition function, projected onto the $Q = 1$ subspace, reads

$$Z = \int_{-\pi/\beta}^{\pi/\beta} \frac{\beta d\lambda}{2\pi} e^{i\beta\lambda} \int \prod_\sigma D[f_\sigma, f_\sigma^\dagger] \int \prod_{\vec{k}\sigma} D[c_{\vec{k}\sigma}, c_{\vec{k}\sigma}^\dagger] \int D[e, e^\dagger] e^{-\int_0^\beta d\tau (\mathcal{L}_f(\tau) + \mathcal{L}_b(\tau))} \quad (4)$$

with the fermionic and bosonic Lagrangians

$$\begin{aligned} \mathcal{L}_f(\tau) &= \sum_{\vec{k}\sigma} c_{\vec{k}\sigma}^\dagger(\tau) (\partial_\tau + \varepsilon_{\vec{k}} - \mu) c_{\vec{k}\sigma}(\tau) + \sum_\sigma f_\sigma^\dagger(\tau) (\partial_\tau + \varepsilon_f - \mu + i\lambda) f_\sigma(\tau) \\ &\quad + V \sum_{\vec{k}\sigma} \left(c_{\vec{k}\sigma}^\dagger(\tau) f_\sigma(\tau) e^\dagger(\tau) + h. c. \right) \\ \mathcal{L}_b(\tau) &= e^\dagger(\tau) (\partial_\tau + i\lambda) e(\tau) . \end{aligned} \quad (5)$$

Here the role of the λ integration is to enforce the constraint. Since the latter commutes with the Hamiltonian, one single integration is sufficient, and introducing a time-dependent λ to integrate over would be superfluous. Furthermore, the fermions may be integrated out since the Lagrangian is bilinear in the fermionic fields. Remarkably, this has been achieved without decoupling the interaction term. As a matter of principle one should verify the correctness of the representation. This can be done in, e.g., the $V \rightarrow 0$ limit by carrying out all integrals. By virtue of the substitution $z = e^{-i\beta\lambda}$, $\beta d\lambda = idz/z$, the λ integral in Eq. (4) is transformed into a contour integral along the complex unit circle. Observing that this substitution implies a 2nd-order pole at $z = 0$ (i.e., at $i\lambda \rightarrow +\infty$, real), one obtains the expected result:

$$Z = 1 + 2 e^{-\beta(\varepsilon_f - \mu)} . \quad (6)$$

Alternatively, Eq. (4) may be viewed as the projection of the product of two non-interacting partition functions for each spin projection onto the $U = \infty$ subspace. Indeed Eq. (4) may be rewritten as:

$$Z = P \prod_{\sigma} \det [S_{\sigma}[e(\tau), \lambda)]. \quad (7)$$

Here, $\det [S_{\sigma}[e(\tau), \lambda)]$ is the fermionic determinant for one spin species involving an effective time-dependent hybridization ($V e^{\dagger}(\tau)$), while the projection operator is given by

$$P = \int_{-\pi/\beta}^{\pi/\beta} \frac{\beta d\lambda}{2\pi} e^{i\beta\lambda} \int D[e, e^{\dagger}] e^{-\int_0^{\beta} d\tau \mathcal{L}_b(\tau)}. \quad (8)$$

Having checked that the representation is faithful is certainly satisfactory, but there is an asymmetry in the representation of charge and spin degrees of freedom. While the former can be expressed in terms of bosons, this is not the case in the latter, and may cause unnecessary errors in any approximate treatment (for details see Ref. [16]).

With this motivation Kotliar and Ruckenstein introduced a representation where spin and charge degrees of freedom may be expressed by bosons.

2.2 Kotliar and Ruckenstein representation

Kotliar and Ruckenstein (KR) extended Barnes's representation through the introduction of two additional Bose operators linked to the spin degrees of freedom, $p_{\uparrow}$ and $p_{\downarrow}$ [14]. In this approach, the physical electron operators are represented as:

$$a_{\sigma} = \tilde{z}_{\sigma} f_{\sigma} \quad \text{with} \quad \tilde{z}_{\sigma} = e^{\dagger} p_{\sigma} + p_{-\sigma}^{\dagger} d. \quad (9)$$

The first term corresponds to the transition from the singly occupied state to the empty one, and the second term to the transition from the doubly occupied state to the singly occupied one. The representation may again be seen to be faithful, under the condition that the auxiliary operators obey canonical commutation relations and satisfy constraints. They read

$$\begin{aligned} 1 &= e^{\dagger} e + \sum_{\sigma} p_{\sigma}^{\dagger} p_{\sigma} + d^{\dagger} d \\ f_{\sigma}^{\dagger} f_{\sigma} &= p_{\sigma}^{\dagger} p_{\sigma} + d^{\dagger} d \quad \sigma = \uparrow, \downarrow, \end{aligned} \quad (10)$$

and need to be satisfied on each site. They may be enforced in a functional integral representation with Lagrange multipliers in a fashion analogous to the one we encountered with the Barnes representation. Moreover, the correctness of the representation may be explicitly verified in the limit $V \rightarrow 0$ through the exact evaluation of the partition function or Green's functions. This representation allows one to express the density operator ($\sum_{\sigma} p_{\sigma}^{\dagger} p_{\sigma} + 2d^{\dagger} d$) and the z -component of the spin operator ($\frac{1}{2} \sum_{\sigma=\pm} \sigma p_{\sigma}^{\dagger} p_{\sigma}$) in terms of bosons. These DoFs may therefore be treated on equal footing. We show in Section 2.4 how this procedure is extended to multiband models.

2.3 Spin-rotation-invariant representation

Though faithful, the Kotliar and Ruckenstein representation is not manifestly spin-rotation-invariant (SRI). Indeed, the transverse components of the spin operator may not be simply represented in terms of auxiliary operators since $S^{x(y)}$ is neither related to $\frac{1}{2} \sum_{\sigma\sigma'} f_{\sigma}^{\dagger} \tau_{\sigma\sigma'}^{x(y)} f_{\sigma'}$ nor to $\frac{1}{2} \sum_{\sigma\sigma'} p_{\sigma}^{\dagger} \tau_{\sigma\sigma'}^{x(y)} p_{\sigma'}$. Therefore fluctuations associated with the transverse modes are not treated on the same footing as the ones associated with the longitudinal mode. To overcome this shortcoming a manifestly SRI formulation has been introduced [15, 16]. In this setup, instead of using the doublet $\{p_{\sigma}\}$ [14] one introduces a scalar (S=0) field p_0 and a vector (S=1) field $\vec{p} = (p_x, p_y, p_z)$. The state $|\sigma\rangle = a_{\sigma}^{\dagger}|0\rangle$ may be represented in terms of them as

$$|\sigma\rangle = \sum_{\sigma'} p_{\sigma\sigma'}^{\dagger} f_{\sigma'}^{\dagger} |\text{vac}\rangle, \text{ with } p_{\sigma\sigma'}^{\dagger} = \frac{1}{2} \sum_{\mu=0,x,y,z} p_{\mu}^{\dagger} \tau_{\sigma\sigma'}^{\mu}, \quad (11)$$

and τ^{μ} the Pauli matrices. The bosons p_{μ} obey canonical commutation relations. Again, all auxiliary operators annihilate the vacuum ($f_{\sigma}|\text{vac}\rangle = e|\text{vac}\rangle = \dots|\text{vac}\rangle = 0$). With this at hand the electron operators may be written as:

$$a_{\sigma} = \sum_{\sigma'} f_{\sigma'} \tilde{z}_{\sigma'\sigma}, \text{ with } \tilde{z}_{\sigma'\sigma} = e^{\dagger} p_{\sigma'\sigma} + \sigma' \sigma p_{-\sigma, -\sigma'}^{\dagger} d. \quad (12)$$

The constraints that the auxiliary operators need to satisfy read

$$1 = e^{\dagger} e + \sum_{\mu=0,x,y,z} p_{\mu}^{\dagger} p_{\mu} + d^{\dagger} d \quad (13)$$

$$\sum_{\sigma} f_{\sigma}^{\dagger} f_{\sigma} = \sum_{\mu=0,x,y,z} p_{\mu}^{\dagger} p_{\mu} + 2d^{\dagger} d \quad (14)$$

$$\sum_{\sigma, \sigma'} f_{\sigma'}^{\dagger} \vec{\tau}_{\sigma\sigma'} f_{\sigma} = p_0^{\dagger} \vec{p} + \vec{p}^{\dagger} p_0 - i \vec{p}^{\dagger} \times \vec{p}. \quad (15)$$

The density operator n , the density of doubly occupied sites operator D , and the spin operator $\vec{S}$ may all be expressed in terms of bosons. They read

$$n = \sum_{\mu} p_{\mu}^{\dagger} p_{\mu} + 2d^{\dagger} d, \quad D = d^{\dagger} d, \quad \vec{S} = \sum_{\sigma\sigma'\sigma_1} \vec{\tau}_{\sigma\sigma'} p_{\sigma\sigma_1}^{\dagger} p_{\sigma_1\sigma'}. \quad (16)$$

The latter expression is especially useful in the context of the t - J model, in particular because the spin degrees of freedom need not be expressed in terms of the original fermions. Using the above, one can tackle models of correlated electrons such as the single-impurity Anderson model, the Anderson lattice model, the t - J or the Hubbard model. However, while the spin and charge degrees of freedom have been mapped onto bosons, anomalous propagators necessarily vanish on a saddle-point level as the Lagrangian is bilinear in the fermionic fields, independent of the model. Here they are not treated on equal footing with the spin and charge degrees of freedom. This gave sufficient motivation to introduce a manifestly spin- and charge-rotation-invariant formulation [16].

2.4 Multi-band systems

A generic Hamiltonian describing the low-energy properties of systems with orbital degeneracy can be written as

$$H = \sum_{i,j,\sigma,\rho} t_{i,j} a_{i,\rho,\sigma}^\dagger a_{j,\rho,\sigma} + U \sum_{i,\rho} n_{i,\rho,\uparrow} n_{i,\rho,\downarrow} + U_2 \sum_{i,\rho' \neq \rho} n_{i,\rho,\uparrow} n_{i,\rho',\downarrow} + U_3 \sum_{i,\sigma,\rho' < \rho} n_{i,\rho,\sigma} n_{i,\rho',\sigma}, \quad (17)$$

where σ is a spin index for the up and down states, ρ is labeling the M bands, and $U_n \equiv U - nJ_H$. Taking J_H finite accounts for the Hund's rule coupling, which favors the formation of magnetic moments.

For this model with on-site interaction, a SB representation can be introduced. Generalizing the Kotliar and Ruckenstein representation one may rewrite any atomic state with the help of a set of pseudo-fermions $\{f_\alpha\}$ and slave bosons $\{\psi_{\alpha_1, \dots, \alpha_m}^{(m)}\}$ ($0 \leq m \leq 2M$). $\psi_{\alpha_1, \dots, \alpha_m}^{(m)}$ is the SB associated with the atomic state consisting of m electrons in states $|\alpha_1, \dots, \alpha_m\rangle$, where α is a composite spin and band index. By construction, it is symmetric under any permutation of two indices and 0 if any two indices are equal. The annihilation operator of a physical electron may be expressed in terms of the auxiliary particles as

$$a_\alpha = \tilde{z}_\alpha f_\alpha, \quad (18)$$

where $\tilde{z}_\alpha$ describes the change in the boson occupation numbers when an electron in state α is annihilated as:

$$\tilde{z}_\alpha = \sum_{m=1}^{2M} \sum_{\alpha_1 < \dots < \alpha_{m-1}} \psi_{\alpha_1, \dots, \alpha_{m-1}}^{\dagger(m-1)} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{(m)} \quad \alpha_i \neq \alpha. \quad (19)$$

The operators $\tilde{z}_\alpha$ in Eq. (19) describe the change in the slave-boson occupation as a many-channel process. Now, the redundant degrees of freedom are projected out with the constraints

$$1 = \sum_{m=0}^{2M} \sum_{\alpha_1 < \dots < \alpha_m} \psi_{\alpha_1, \dots, \alpha_m}^{\dagger(m)} \psi_{\alpha_1, \dots, \alpha_m}^{(m)} \quad (20)$$

$$f_\alpha^\dagger f_\alpha = \sum_{m=1}^{2M} \sum_{\alpha_1 < \dots < \alpha_{m-1}} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{\dagger(m)} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{(m)}. \quad (21)$$

3 Gauge symmetry and radial slave-boson fields

When representing the electron operators a_σ as $\tilde{z}_\sigma f_\sigma$, one may infer that a group of transformations will leave this expression invariant, assuming that it acts on the fields in such a way that

$$f_\sigma(\tau) \longrightarrow f_\sigma(\tau) e^{i\phi(\tau)}, \quad \text{and} \quad \tilde{z}_\sigma(\tau) \longrightarrow \tilde{z}_\sigma(\tau) e^{-i\phi(\tau)}. \quad (22)$$

This is indeed the case when considering the $U \rightarrow \infty$ Barnes representation for the SIAM since $\tilde{z}_\sigma$ is given by $e^\dagger$. This local $U(1)$ gauge symmetry was first realized by Read and Newns [33].

One may make use of it to gauge away the phase of the slave boson, which remains as a purely radial field, while the Lagrange constraint parameter is promoted to a time-dependent field. Since standard textbooks do not mention representations of such radial fields that are set up on a discretized time mesh from the beginning, the key steps are presented below, following Ref. [31]. In this scheme the partition function takes a form analogous to Eqs. (4-5). However the projection operator does not mix the N time steps and may be written as

$$P = \lim_{N \rightarrow \infty} \lim_{W \rightarrow \infty} \prod_{n=1}^N P_n, \quad \text{with}$$

$$P_n = \int_{-\infty}^{\infty} \frac{\beta}{N} \frac{d\lambda_n}{2\pi} \int_{-\infty}^{\infty} dx_n e^{-\frac{\beta}{N}(i\lambda_n(x_n-1) + Wx_n(x_n-1))}. \quad (23)$$

Here the constraint parameter λ_n is defined for each time step n , i.e., it is a time-dependent field, and x_n represents the radial slave-boson field at time step n . In the discrete time-step form, the fermionic part of the action reads

$$S_f = \sum_{n=1}^N \left\{ \sum_{\vec{k}\sigma} c_{\vec{k},n,\sigma}^\dagger \left(c_{\vec{k},n,\sigma} - e^{-\frac{\beta}{N}(\varepsilon_{\vec{k}} - \mu)} c_{\vec{k},n-1,\sigma} \right) + \sum_{\sigma} f_{n,\sigma}^\dagger \left(f_{n,\sigma} - e^{-\frac{\beta}{N}(\varepsilon_f + i\lambda_n - \mu)} f_{n-1,\sigma} \right) \right\}$$

$$+ \frac{\beta}{N} \sum_{n=1}^N \sum_{\vec{k}\sigma} V x_n \left(c_{\vec{k},n,\sigma}^\dagger f_{n-1,\sigma} + f_{n,\sigma}^\dagger c_{\vec{k},n-1,\sigma} \right). \quad (24)$$

The integration over the fermionic fields can be manifestly carried out. This allows one to obtain the partition function by projecting the resulting fermionic determinant:

$$Z = P \prod_{\sigma} \det [S_{\sigma} [\{x_n\}, \{\lambda_n\}]] \quad (25)$$

with the above projection operator Eq. (23). The expectation value of the hole density operator takes the simple form:

$$\langle n_h(\tau_m) \rangle = \langle x_m \rangle = \frac{1}{Z} P \left\{ x_m \prod_{\sigma} \det [S_{\sigma} [\{x_n\}, \{\lambda_n\}]] \right\}. \quad (26)$$

It is easily seen to be time-independent. In contrast to a Bose condensate, $\langle x_m \rangle$ is generally finite and may only vanish for zero hole concentration [21]. It is not related to a broken symmetry. The radial slave-boson field exhibits another specific feature: For any power $a > 0$, one finds $\langle x_m^a \rangle = \langle x_m \rangle^a$, as the corresponding projections of the fermionic determinant all yield the same value.

As concerns the hole autocorrelation function, it is conveniently expressed as a projection of the fermionic determinant. It reads

$$\langle n_h(\tau_n) n_h(\tau_m) \rangle = \langle x_n x_m \rangle = \frac{1}{Z} P \left\{ x_n x_m \prod_{\sigma} \det [S_{\sigma} [\{x_n\}, \{\lambda_n\}]] \right\}. \quad (27)$$

Hence it can be obtained without first determining a self-energy.

Regarding the Kotliar and Ruckenstein representation, it took a long discussion to determine the gauge symmetry group [34–37, 16]. It was finally agreed that it reads $U(1) \times U(1) \times U(1)$. In fact, it could be shown that one may gauge away the phase of three bosonic fields by promoting all three constraint parameters to fields. The fourth one, for example d , remains complex. Therefore, in the $U \rightarrow \infty$ limit ($d \rightarrow 0$), all three remaining bosonic fields are radial slave-boson fields. In functional integral language they may be handled in the same fashion as the above x -field. For an example, see [22].

4 Saddle-point approximations

The exact evaluation of a quantity represented by a functional integral is an ambitious task. So far, the results are limited to a very small cluster [21, 22] or to the Ising chain [31]. Hence we rather focus on an economical way to determine observable quantities in the SB framework. It is provided by a saddle-point approximation (SPA) to the functional integral and often yields physically reasonable results. This is equivalent to allowing for a finite expectation value of a Bose field amplitude. Strictly speaking, a finite expectation value of a Bose field operator violates gauge invariance and should not exist. In contrast, a finite saddle-point amplitude of the radial slave-boson fields is compatible with Elitzur's theorem. Besides, the saddle-point approximation is exact in the large-degeneracy limit, and the Gaussian fluctuations provide the $1/N$ corrections [16]. Moreover it obeys a variational principle in the limit of large spatial dimensions where the Gutzwiller approximation becomes exact for the Gutzwiller wave function [38]. Furthermore, it could be shown in this limit that longer-ranged interactions are not dynamical and reduce to their Hartree approximation [39]. Therefore, this approach also obeys a variational principle in this limit when applied to the extended Hubbard model Eq. (74).

4.1 Saddle-point approximation to the Barnes representation

In its simplest form, the SPA consists of replacing the boson field operators e_i at each lattice site, or e at the impurity site, by the modulus of its expectation value, in accordance with the above. This yields a non-interacting model, which is easily solved. Below we briefly discuss the solutions for the Anderson impurity model and the Anderson lattice model.

4.1.1 Kondo effect in the Anderson impurity model

In SPA the Anderson impurity Hamiltonian Eq. (1) takes for $U \rightarrow \infty$ the form

$$H = \sum_{\vec{k}\sigma} \varepsilon_{\vec{k}} c_{\vec{k}\sigma}^\dagger c_{\vec{k}\sigma} + \varepsilon_f \sum_{\sigma} f_{\sigma}^\dagger f_{\sigma} + V \sum_{\vec{k}\sigma} e_0 \left(c_{\vec{k}\sigma}^\dagger f_{\sigma} + f_{\sigma}^\dagger c_{\vec{k}\sigma} \right) + \lambda(Q - 1). \quad (28)$$

The conserved charge is $Q = \sum_{\sigma} f_{\sigma}^\dagger f_{\sigma} + e_0^2 = 1$, and λ is the corresponding Lagrange multiplier. One recognizes that Eq. (28) describes a resonant-level model with renormalized parameters. They are $\tilde{\varepsilon}_f = \varepsilon_f + \lambda$ and $\tilde{V} = Ve_0$. Introducing $\tilde{\Delta} = e_0^2 \Delta = \pi N_F^{(0)} \tilde{V}^2$, where $\Delta = \pi N_F^{(0)} V^2$

($N_F^{(0)} = 1/2D$ is the conduction electron DOS at the Fermi level) allows one to write the saddle-point equations in the form of two conditions for the level position $\tilde{\varepsilon}_f$ and the level width $\tilde{\Delta}$. They are

$$\tilde{\varepsilon}_f = \varepsilon_f - \frac{2\Delta}{\pi} \ln \frac{\sqrt{\tilde{\varepsilon}_f^2 + \tilde{\Delta}^2}}{D} \quad \text{and} \quad \tilde{\Delta} = \Delta - \frac{2\Delta}{\pi} \tan^{-1} \frac{\tilde{\Delta}}{\tilde{\varepsilon}_f}. \quad (29)$$

In the limit of $\Delta \ll |\varepsilon_f|$, the occupation of the local level $n_f = \frac{2/\pi^{-1}}{\tan} \frac{\tilde{\Delta}}{\tilde{\varepsilon}_f}$ approaches unity. This means that a local moment forms at higher temperature. Below a characteristic temperature, the Kondo temperature T_K , the local moment gets screened by the conduction electron spins, which form a resonance state with the local moment. It is located close to the Fermi energy, at $\tilde{\varepsilon}_f$, and is of width $\tilde{\Delta} \approx T_K = D \exp \frac{-|\varepsilon_f|}{2N_F^{(0)}V^2} = D \exp \frac{-1}{2N_F^{(0)}J}$, where $J = \frac{V^2}{|\varepsilon_f|}$ is the antiferromagnetic spin exchange coupling constant of the local spin and the local conduction electron spin density. The low-temperature behavior of Kondo systems is reasonably well described by SB mean-field theory. Yet at higher temperatures, a spurious first-order transition to the local-moment regime is found in this approximation rather than a continuous crossover.

As an alternative scheme, Kroha *et al.* [40] developed an approximation that guarantees local gauge invariance in a conserving approximation and allows for Fermi-liquid as well as non-Fermi liquid behavior for the investigated multi-channel Anderson impurity problem.

4.1.2 Heavy fermions in the Anderson lattice model

The Anderson lattice model in the limit $U \rightarrow \infty$ has been investigated in the SB mean-field approximation [33], in which the Hamiltonian reads again as a single-particle Hamiltonian, but for two hybridized bands

$$H = \sum_{\vec{k}\sigma} \varepsilon_{\vec{k}} c_{\vec{k}\sigma}^\dagger c_{\vec{k}\sigma} + \varepsilon_f \sum_{i,\sigma} f_{i\sigma}^\dagger f_{i\sigma} + V \sum_{i,\sigma} e_0 \left(c_{i\sigma}^\dagger f_{i\sigma} + f_{i\sigma}^\dagger c_{i\sigma} \right) + \sum_i \lambda_i (Q_i - 1). \quad (30)$$

The saddle-point condition with respect to the field λ_i leads to the condition $\langle Q_i \rangle = 1$. For a translationally invariant state it is independent of the lattice position $\vec{R}_i$. As in the impurity problem, the f -level position is shifted by correlation effects to $\tilde{\varepsilon}_f$, and the square of the boson amplitude is related to the f -level occupation n_f through:

$$\tilde{\varepsilon}_f = \varepsilon_f - 2N_F^{(0)}V^2 \ln \frac{\tilde{\varepsilon}_f - \varepsilon_F}{D} \quad (31)$$

$$e_0^2 = 1 - n_f = 1 - \frac{2N_F^{(0)}V^2 e_0^2}{\tilde{\varepsilon}_f}, \quad (32)$$

under the assumption $|\tilde{\varepsilon}_f| \ll D$. In the case where ε_f is sufficiently below the Fermi level ε_F we have $|\tilde{\varepsilon}_f| \ll |\varepsilon_f|$ and, from Eq. (31), we observe that $\tilde{\varepsilon}_f - \varepsilon_F = D \exp \frac{-|\varepsilon_f|}{2N_F^{(0)}V^2} = T_K$, equal to the single-impurity Kondo temperature. In this limit $e_0^2 \approx |\tilde{\varepsilon}_f|/2N_F^{(0)}V^2 \ll 1$. Thus, the hybridization amplitude is substantially reduced, leading to heavy quasiparticle bands of energy

$$E_{\vec{k}}^\pm = \frac{1}{2} \left[\varepsilon_{\vec{k}} + \tilde{\varepsilon}_f \pm \sqrt{(\varepsilon_{\vec{k}} + \tilde{\varepsilon}_f)^2 + V^2 e_0^2} \right]. \quad (33)$$

4.2 Saddle-point approximation to the Kotliar and Ruckenstein representation

Extending this approach to the Hubbard Model on the lattice has also been achieved. Yet at this stage of the formulation, the representation suffers from the drawback that neither the low-charge-carrier-density limit nor the non-interacting limit are properly recovered on the SPA level [14], in contrast to more conventional approaches. Kotliar and Ruckenstein overcame this difficulty by noticing that there is no unique SB representation but rather infinitely many different ones. If faithful, they are all equivalent when the functional integral is exactly evaluated but differ on the saddle-point level. Kotliar and Ruckenstein provided us with a representation of the kinetic energy that solves both aspects of the above drawback. The KR representation consists of replacing the operators $\tilde{z}_\sigma$ in Eq. (9) by

$$z_\sigma = e^\dagger L_\sigma R_\sigma p_\sigma + p_{-\sigma}^\dagger L_\sigma R_\sigma d, \text{ with} \quad (34)$$

$$L_\sigma = \frac{1}{\sqrt{1 - p_\sigma^\dagger p_\sigma - d^\dagger d}} \quad \text{and} \quad R_\sigma = \frac{1}{\sqrt{1 - p_{-\sigma}^\dagger p_{-\sigma} - e^\dagger e}} \quad (35)$$

and of consistently using $a_\sigma = z_\sigma f_\sigma$ in the representation of the kinetic energy operator. In this form, the SPA to the KR representation is equivalent to the Gutzwiller approximation (GA) to the Gutzwiller wave function [14]. As the GA yields the exact energy of the Gutzwiller wave function in the large-dimensionality limit, the SPA to the KR representation acquires a variational principle in this limit. In addition it turns exact in several large N limits [16], or for particular toy models [41]. These properties are shared by the SRI formulation [16]. Indeed, introducing $\tilde{p}_{\sigma\sigma'} \equiv \sigma\sigma' p_{-\sigma', -\sigma}$, the z operator reads

$$\underline{z} = e^\dagger \underline{L} \underline{M} \underline{R} \underline{p} + \tilde{p}^\dagger \underline{L} \underline{M} \underline{R} d \quad (36)$$

with

$$\underline{M} = \left[1 + e^\dagger e + \sum_\mu p_\mu^\dagger p_\mu + d^\dagger d \right]^{\frac{1}{2}}, \quad (37)$$

$$\underline{L} = \left[(1 - d^\dagger d) \underline{1} - 2 \underline{p}^\dagger \underline{p} \right]^{-\frac{1}{2}} \quad \text{and} \quad \underline{R} = \left[(1 - e^\dagger e) \underline{1} - 2 \tilde{p}^\dagger \tilde{p} \right]^{-\frac{1}{2}}. \quad (38)$$

Eq. (36) and Eq. (38) correct Eq. (22) in [16] and Eq. (38) corrects Eq. (3) in [42].

4.2.1 Mott-Hubbard metal-to-insulator transition

The KR and SRI representations have been used to characterize a broad range of phases of the Hubbard Model [43–57], as they are able to capture interaction effects beyond the physics of a Slater determinant. These representation encompass the Brinkman-Rice mechanism [58, 59], described below, allowing for the description of the Mott metal-to-insulator transition. This transition is a genuine interaction-driven transition that is not linked to a period doubling resulting from, e.g., an antiferromagnetic instability. On the contrary, it arises when considering the paramagnetic saddle point. In the SRI representation, it corresponds to setting the bosonic

fields $\vec{p}_i(\tau)$ and the constraint fields enforcing Eq. (15) to zero and to replacing the remaining bosonic and constraint fields by their mean value. The free energy then reads

$$F = -T \sum_{\vec{k}, \sigma} \ln \left(1 + e^{-\frac{E_{\vec{k}\sigma}}{T}} \right) + U d^2 + \alpha (e^2 + p_0^2 + d^2 - 1) - \beta_0 (p_0^2 + 2d^2). \quad (39)$$

Here the Lagrange multiplier α (β_0) enforces the constraints of Eq. (13) and (14). With

$$z_0 = \frac{1}{\sqrt{2}} \frac{p_0(e + d)}{\sqrt{1 - d^2 - \frac{1}{2}p_0^2} \sqrt{1 - e^2 - \frac{1}{2}p_0^2}} \quad (40)$$

the quasiparticle dispersion relation is given by

$$E_{\vec{k}\sigma} = z_0^2 t_{\vec{k}} + \beta_0 - \mu. \quad (41)$$

z_0^2 plays the role of both a mass renormalization factor and of a quasiparticle residue. In the parameter range in which it vanishes, a Mott insulating state is realized. Solving the saddle-point equations at half filling $\rho = 1$ yields

$$z_0^2 = 1 - \left(\frac{U}{U_c} \right)^2, \text{ with } U_c = -4 \sum_{\vec{k}, \sigma} t_{\vec{k}} f_F(z_0^2 t_{\vec{k}}), \quad (42)$$

where f_F is the Fermi function. Therefore, the quasiparticle residue continuously varies from 1 down to 0 for $U \rightarrow U_c$. At this point, the quasiparticle mass diverges and its residue vanishes, signaling a metal-to-insulator transition. As an additional signature of a transition, a Mott gap opens. Indeed, solving the equation for the chemical potential of the quasiparticles for $U > U_c$ and $\rho \rightarrow 1$ yields [35]

$$\mu(\rho) = \frac{U}{2} \left[1 - \frac{1 - \rho}{|1 - \rho|} \sqrt{1 - \frac{U_c}{U}} \right]. \quad (43)$$

The discontinuity in μ across $\rho = 1$ indicates a pair of first-order phase transitions from the metallic phase at $\rho < 1$ (with finite z_0) to the insulating phase at $\rho = 1$ (with chemical potential $\mu = U/2$) and back to the metallic phase at $\rho > 1$ (with finite z_0). This discontinuity vanishes for $U \rightarrow U_c^+$, which is therefore a critical point. In the insulating phase the quasiparticle contribution to doubly occupied sites vanishes. This does not imply that the latter is predicted to be zero but that it purely results from fluctuations, which we address in Sec. 5.

The saddle-point equations following from the free energy Eq. (39) read

$$\begin{aligned} p_0^2 + e^2 + d^2 - 1 &= 0, \\ p_0^2 + 2d^2 &= \rho, \\ \frac{1}{2e} \frac{\partial z_0^2}{\partial e} \bar{\varepsilon} &= -\alpha, \\ \frac{1}{2p_0} \frac{\partial z_0^2}{\partial p_0} \bar{\varepsilon} &= \beta_0 - \alpha, \\ \frac{1}{2d} \frac{\partial z_0^2}{\partial d} \bar{\varepsilon} &= 2(\beta_0 - \alpha) + \alpha - U. \end{aligned} \quad (44)$$

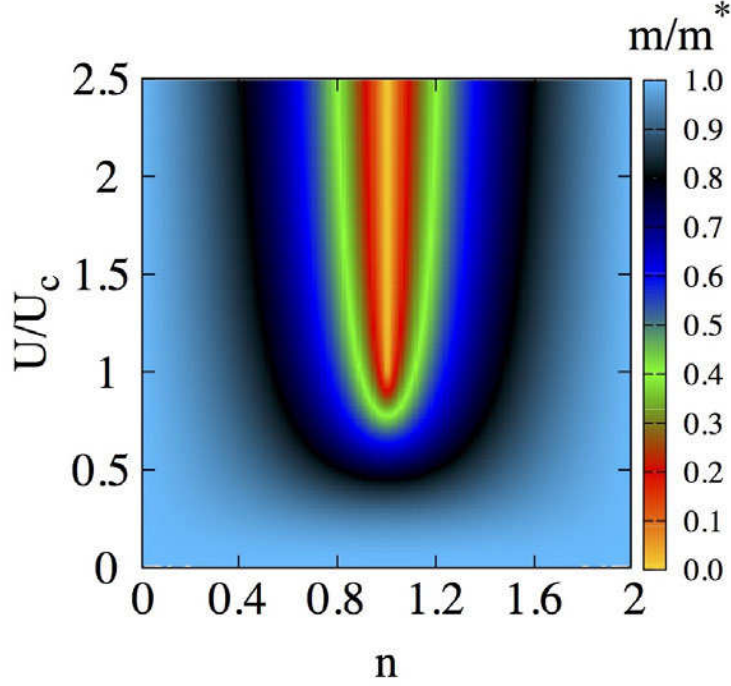


Fig. 1: Inverse effective mass z_0^2 for the Hubbard model on the cubic lattice.

Here we have introduced the averaged kinetic energy,

$$\bar{\varepsilon} = \int d\omega \rho(\omega) \omega f_F(z_0^2 \omega + \beta_0 - \mu), \quad (45)$$

the determination of which involves the density of states $\rho(\omega)$. Introducing the doping away from half filling $\delta = 1 - \rho$, the Coulomb parameter $u = U/(-8\bar{\varepsilon})$, and $y \equiv (e + d)^2$, the saddle-point equations can be cast into a single one that finally reads

$$y^3 + (u - 1)y^2 = u\delta^2. \quad (46)$$

For more details see Ref. [16, 59]. In the case of a 3D cubic lattice the quasiparticle mass diverges at half filling for $U_c \simeq 16.04 t$. This behavior is general, and the transition occurs for other lattices in a qualitatively equivalent way. For instance, in the case of a 2D square lattice, the metal-to-insulator transition occurs at $U_c = 2(8/\pi)^2 t$. Note that the ratio of the critical interaction for the 3D cubic lattice to the one of the 2D square lattice (1.24) is somewhat smaller than the naive estimate that would be obtained from the corresponding ratio of the number of nearest neighbors (3/2). In the case of a rectangular DOS one has $U_c = W$.

Regarding the doping dependence of the quasiparticle residue, Fig. 1 shows that a mass renormalization larger than 2 is only realized in the regime of large $U > \frac{3}{4}U_c$ and doping $|\delta| < 0.25$. In the limits $U \rightarrow 0$ and $|\delta| \rightarrow 1$, the saddle-point approximation correctly yields the exact result $z_0^2 = 1$. Further, calculations performed for the 2D square lattice yield a figure very similar to Fig. 1.

4.3 Saddle-point approximation to the multi-band Hubbard model

We now turn to the multi-band Hubbard model that is often used for the description of transition-metal oxides. As in the case of the one-band Hubbard model, the non-interacting limit is not properly recovered when representing the electron operator a_α according to Eq. (18) as $\tilde{z}_\alpha f_\alpha$ in the kinetic energy term. This difficulty may be overcome in a fashion analogous to the one followed in the framework of the Kotliar and Ruckenstein representation to the one-band Hubbard model: One represents the electron operator a_α as $z_\alpha f_\alpha$ where

$$z_\alpha = \sum_{m=1}^{2M} \sum_{\alpha_1 < \dots < \alpha_{m-1}} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{(m)} L_\alpha R_\alpha \psi_{\alpha_1, \dots, \alpha_{m-1}}^{\dagger(m-1)} \quad (47)$$

involves the normalization factors L_α and R_α . They are now given by

$$\begin{aligned} R_\alpha &= \left[1 - \sum_{m=0}^{2M-1} \sum_{\alpha_1 < \dots < \alpha_m} \psi_{\alpha_1, \dots, \alpha_m}^{\dagger(m)} \psi_{\alpha_1, \dots, \alpha_m}^{(m)} \right]^{-\frac{1}{2}} \quad \alpha_i \neq \alpha \\ L_\alpha &= \left[1 - \sum_{m=1}^{2M} \sum_{\alpha_1 < \dots < \alpha_{m-1}} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{\dagger(m)} \psi_{\alpha, \alpha_1, \dots, \alpha_{m-1}}^{(m)} \right]^{-\frac{1}{2}}. \end{aligned} \quad (48)$$

Namely L_α normalizes to 1 the probability that no electron in state $|\alpha\rangle$ is present on a site before one such electron hops to that particular site, and R_α makes sure that it happened. Clearly the eigenvalues of the operators L_α and R_α are 1 in the physical subspace.

We now proceed to the saddle-point approximation, and we investigate the Mott transition at commensurate integer filling n for an M -band model. In order to highlight general features of the model, we first consider the paramagnetic, paraorbital phase at $J_H = 0$. The latter is obtained after having integrated out the fermions, setting all bosonic fields to their averaged value, and, for given m , demanding that the various $\psi_{\alpha_1 < \dots < \alpha_m}^{(m)}$ are equal to one another. The Mott transition that occurs at commensurate density n is most conveniently discussed by projecting out occupancies that are larger than $n + 1$ and smaller than $n - 1$ (if any), as they would at most play a subleading role. The constraint allows for eliminating the variables $\psi^{(n-1)}$ and $\psi^{(n)}$ obtaining the free energy at filling n as

$$F(D) = (1 - 2D^2) D^2 \left(\sqrt{b_{n,M}} + \sqrt{c_n} \right)^2 \bar{\varepsilon} + U \left(D^2 + \binom{n}{2} \right), \quad (49)$$

with $\bar{\varepsilon} \equiv \int d\epsilon \epsilon \rho(\epsilon) f_F(z^2 \epsilon + \lambda_0 - \mu)$, $D^2 \equiv \binom{2M}{n+1} \psi^{(n+1)2}$, $b_{n,M} \equiv (2M - n + 1)/(2M - n)$, and $c_n \equiv (n + 1)/n$. Here, $\rho(\epsilon)$ is the total DOS. Minimizing Eq. (49) with respect to D yields a critical interaction strength at which D vanishes. It depends on n and M and reads

$$U_c^{(n,M)} = -\bar{\varepsilon} \left(\sqrt{b_{n,M}} + \sqrt{c_n} \right)^2, \quad (50)$$

which reproduces the results of the Gutzwiller approximation [13, 76]. This locates the Mott transition. In the often considered case of a rectangular DOS, the critical interaction strength may be related to the band width W through

$$U_c^{(n,M)} = \frac{nW}{4M} (2M - n) \left(\sqrt{b_{n,M}} + \sqrt{c_n} \right)^2. \quad (51)$$

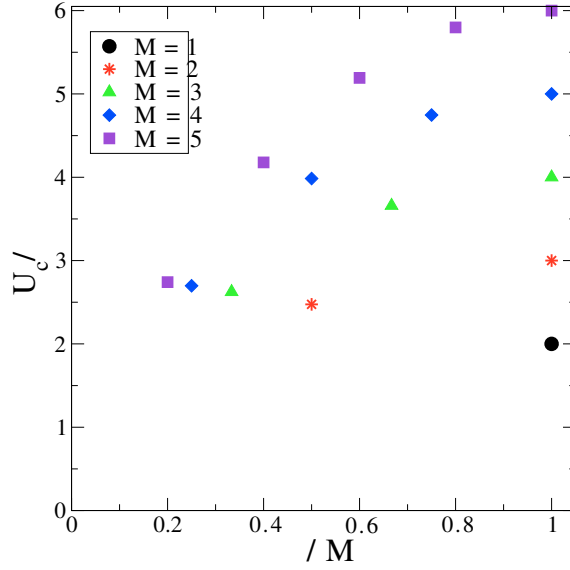


Fig. 2: *Dependence of the location of the Mott transition on the filling n and the band degeneracy M for the particle-hole symmetric rectangular density of states.*

As shown in Fig. 2, it weakly depends on the band degeneracy for fixed filling n , but quite significantly on n for a given band degeneracy.

The effective mass of the quasiparticles diverges when reaching the Mott point. We obtain the analytical behavior as

$$\frac{m}{m^*} = z_0^2 = \frac{(\sqrt{b_{n,M}} + \sqrt{c_n})^2}{8} \frac{U_c^{(n,M)2} - U^2}{U_c^{(n,M)2}}. \quad (52)$$

The dependence on the band degeneracy is weak as a consequence of the particular form of the coefficients $b_{n,M}$ and c_n . As the critical interaction strength increases with M the quasiparticle residue $Z = z_0^2$ increases slightly with M . However, for small values of U and without projecting out higher occupancies, Z actually decreases with increasing M . There is therefore a crossover value of the interaction strength beyond which the system becomes more metallic with increasing M [17].

As a function of the hole doping δ , the quasi particle residue vanishes for δ going to 0 above $U_c^{(n,M)}$ as

$$z_0^2 = \frac{\delta}{2} (b_{n,M} - c_n) + \frac{|\delta|}{2} \left((b_{n,M} + c_n) \sqrt{1 + 4\varphi_{n,M}} + 4\sqrt{b_{n,M}c_n\varphi_{n,M}} \right) \quad (53)$$

where we introduced:

$$\varphi_{n,M} \equiv \frac{U_c^{(n,M)2} \frac{b_{n,M}c_n}{(\sqrt{b_{n,M}} + \sqrt{c_n})^4}}{(U - U_c^{(n,M)}) \left(U - U_c^{(n,M)} \left(\frac{\sqrt{b_{n,M}} - \sqrt{c_n}}{\sqrt{b_{n,M}} + \sqrt{c_n}} \right)^2 \right)}. \quad (54)$$

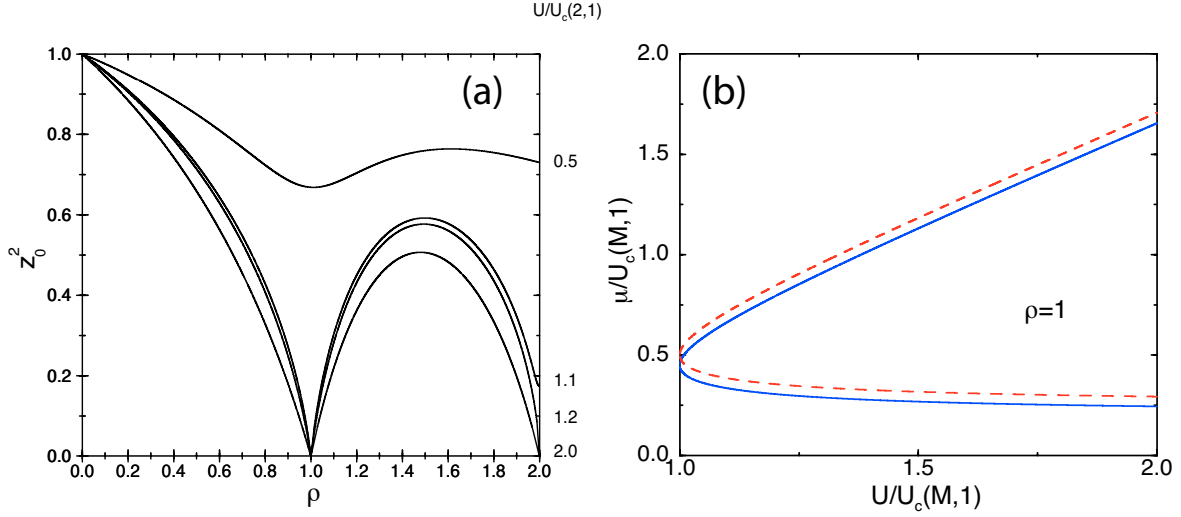


Fig. 3: (a) Inverse effective mass in the two-band model as a function of density for several values of U . (b) Chemical potential for $n = 1$ for the one-band (dashed line) and two-band (full line) models.

This expression of the quasiparticle residue consists of two contributions that are either symmetric or antisymmetric with respect to particle or hole doping. The particle-hole symmetry requires the antisymmetric contribution to vanish for $n = M$. We observe that the asymmetry of z_0^2 on particle or hole doping increases under an increase of $|n - M|$. It vanishes more slowly for hole doping (for $n \leq M$) than for particle doping, for increasing degeneracy at fixed n , for increasing degeneracy at $n = M$, and under an increase of U . As an example, we calculate the effective mass for the two-band model, which has been calculated without projecting out higher occupancies, and show it in Fig. 3a.

Analogously to the one-band case we obtain a Mott gap. Indeed, the number of quasiparticles is a continuous function of their chemical potential $\mu - \lambda_0/2$. The constraint parameters Λ and λ as well as μ jump when going through the Mott gap. The Mott gap $\Delta \equiv \lim_{\delta \rightarrow 0^-} \mu(\delta) - \lim_{\delta \rightarrow 0^+} \mu(\delta)$ results as

$$\Delta = \sqrt{(U - U_c^{(n,M)}) \left(U - U_c^{(n,M)} \left(\frac{\sqrt{b_{n,M}} - \sqrt{c_n}}{\sqrt{b_{n,M}} + \sqrt{c_n}} \right)^2 \right)}. \quad (55)$$

In the limiting case of $U \gg U_c^{(n,M)}$, the Mott gap is given by U , while it closes at $U_c^{(n,M)}$ as $\Delta \sim U_c^{(n,M)} \sqrt{U/U_c^{(n,M)} - 1}$, the square root behavior being typical of slave-boson mean-field theories. It is displayed in Fig. 3b, where it is compared to the one-band case as obtained by Lavagna [35]. Clearly, no big difference in the Mott gap is found when going from one band to two bands. In fact $\Delta/U_c^{(n,M)}$ is independent of M at $n = M$, while for fixed n the dependence on M is very weak. For a comparison to the experimental situation in the titanates see Ref. [17].

4.3.1 Influence of Hund's rule coupling

The Hund's rule coupling turns out to have a deep influence on the nature of the Mott transition. As an example we treat here the two-band model around the $n = 1$ Mott insulating lobe. At density $\rho = 1$ we obtain the saddle-point free energy as

$$F = \frac{4}{3}\bar{\varepsilon} (1 - 2r^2) \left(r + (d_0 + d_x + \Delta_0)/\sqrt{2} \right)^2 + (U + 3J) \Delta_0^2 + (U + J) d_x^2 + U d_0^2. \quad (56)$$

with $\bar{\varepsilon} \equiv \int d\epsilon \epsilon \rho(\epsilon) f_F(z^2\epsilon + \lambda_0 - \mu)$, $d_0 \equiv (\psi_{\uparrow,\uparrow}^{(2)} + \psi_{\downarrow,\downarrow}^{(2)})/\sqrt{2}$, $d_x \equiv (\psi_{\uparrow,\downarrow}^{(2)} + \psi_{\downarrow,\uparrow}^{(2)})/\sqrt{2}$, $\Delta_0 \equiv (\psi_{\uparrow,\downarrow,0}^{(2)} + \psi_{0,\downarrow,\uparrow}^{(2)})/\sqrt{2}$, $r^2 \equiv d_0^2 + d_x^2 + \Delta_0^2$, and $\lambda \equiv \sum_{\alpha} \lambda_{\alpha}/2$, and we have used the constraints to remove the variables $\psi^{(0)}$ and $\psi^{(1)}$. Let us notice that this expression differs from an ordinary Ginzburg-Landau free energy in that it cannot be written as a fourth-order polynomial in the variables d_0 , d_x , and Δ_0 . Therefore, a critical point for one field would be critical for the other ones as well. Lacking an analytical expression for the location of the Mott transition for arbitrary J_H/U , we first focus on the small J_H/U regime. We find

$$U_c^{(1,2)}(J_H) = U_c^{(1,2)}(0) \left(1 - \frac{4}{3} \frac{J_H}{U} + \mathcal{O}(J_H/U)^2 \right). \quad (57)$$

Hence $U_c^{(1,2)}$ first decreases linearly with J_H . Another regime of interest is the large J_H regime. There we obtain the location of the Mott transition as

$$U_c^{(1,2)} = -\frac{2}{3}\bar{\varepsilon} (3 + 2\sqrt{2}) \left(1 - \frac{8}{9} \frac{\bar{\varepsilon}}{J_H} \right) + \mathcal{O} \left(\frac{\bar{\varepsilon}}{J_H} \right)^2 \quad (58)$$

and thus decreasing J from ∞ leads to an increase of the critical interaction. Another intriguing feature of transition-metal oxides such as V_2O_3 is the metal-to-insulator transition that occurs in the vicinity of the tri-critical point under an increase of temperature. It has been interpreted [60] as the transition from a Fermi liquid with finite quasiparticle residue Z to an insulator with $Z = 0$. In other words, there is a finite coherence temperature T_{coh} at which the coherence of the Fermi liquid (and Z) vanishes. This result was obtained in the dynamical mean-field approximation to the one-band model, which becomes exact in the limit of large dimensions and is recovered in the Gutzwiller approximation [61]. At finite T there is a first-order metal-to-insulator transition at a critical $U_c^{(1,M)}(T)$

$$U_c^{(1,M)}(T) = U_c^{(1,M)}(0) - \sqrt{8U_c^{(1,M)}(0) T \ln 2M}. \quad (59)$$

Thus an increase in temperature may produce a metal-to-insulator transition, which is consistent with the experimental situation in V_2O_3 . In the dynamical mean-field approximation at finite temperatures there is an interaction strength $U_{c2}(T)$ at which the metallic solution ceases to exist. This quantity can also be evaluated in this SB scheme and is given by

$$U_{c2}^{(1,M)}(T) = U_c^{(1,M)}(0) \left(1 - \alpha_M (T/W)^{\frac{2}{3}} \right), \quad (60)$$

with $\alpha_1 \sim 2.53$ for the one-band model, and $\alpha_2 \sim 3.32$ for the two-band model.

4.4 A concrete example

We now proceed with an example that builds on a model for the anisotropic superconductor Sr_2RuO_4 . This two-band model includes a finite J_H and an effective kinetic energy term that derives from a tight-binding Hamiltonian. As suggested by Noce and Cuoco [62], the bands crossing the Fermi energy build on the Ru $4d$ d_{xz} - and d_{yz} -orbitals as well as on the O $2p_z$ -orbitals. Following Ref. [63], we integrate out the latter. This yields the effective model

$$H_0 = \sum_{\vec{k}, \sigma} \left(d_{xz, \vec{k}, \sigma}^\dagger, d_{yz, \vec{k}, \sigma}^\dagger \right) \begin{pmatrix} e_{\vec{k}} & a_{\vec{k}} \\ a_{\vec{k}} & f_{\vec{k}} \end{pmatrix} \begin{pmatrix} d_{xz, \vec{k}, \sigma} \\ d_{yz, \vec{k}, \sigma} \end{pmatrix}, \quad (61)$$

with $a_{\vec{k}} = -4t' \sin k_x \sin k_y$, $e_{\vec{k}} = t \cos k_x$, and $f_{\vec{k}} = t \cos k_y$. The two pairs of bands $E_{\vec{k}, \nu, \sigma}$ with

$$E_{\vec{k}, \nu, \sigma} = \frac{1}{2} (e_{\vec{k}} + f_{\vec{k}}) + \frac{1}{2} \nu \sqrt{(e_{\vec{k}} - f_{\vec{k}})^2 + 4a_{\vec{k}}^2}, \quad \nu = \pm 1 \quad (62)$$

acquire two-dimensional character because of the finite t' .

On the slave-boson SPA level the free energy reads

$$\begin{aligned} F = & -\frac{1}{\beta} \sum_{\vec{k}, \nu, \sigma} \ln \left(1 + e^{-\beta E_{\vec{k}, \nu, \sigma}} \right) + U \sum_i \left(\sum_{\alpha < \alpha'} d_{i, \alpha \alpha'}^2 + 3 \sum_{\alpha < \alpha' < \alpha''} t_{i, \alpha \alpha' \alpha''}^2 + 6q_i^2 \right) \\ & + J_H \sum_i \left(\sum_{\sigma} d_{i, xz\sigma, yz-\sigma}^2 + 3 \sum_{\rho} d_{i, \rho \uparrow, \rho \downarrow}^2 + 4 \sum_{\alpha < \alpha' < \alpha''} t_{i, \alpha \alpha' \alpha''}^2 + 8q_i^2 \right) \\ & + \sum_i \Lambda_i \left(e_i^2 + \sum_{\alpha} p_{i\alpha}^2 + \sum_{\alpha < \alpha'} d_{i, \alpha \alpha'}^2 + \sum_{\alpha < \alpha' < \alpha''} t_{i, \alpha \alpha' \alpha''}^2 + q_i^2 - 1 \right) \\ & - \sum_{i, \alpha} \beta_{i, \alpha} \left(p_{i, \alpha}^2 + \sum_{\alpha'} d_{i, \alpha \alpha'}^2 + \sum_{\alpha' \alpha''} t_{i, \alpha \alpha' \alpha''}^2 + q_i^2 \right) \end{aligned} \quad (63)$$

Here the bosons e , p_{α} , $d_{\alpha\alpha'}$, $t_{\alpha\alpha'\alpha''}$, and q refer to occupancies zero, one, two, three, and four, respectively. The Lagrange multipliers Λ and β_{α} enforce the constraints Eq. (20) and Eq. (21), respectively. In a paramagnetic or a ferromagnetic phase, the dispersions of the quasiparticles are given by

$$E_{k, \nu, \sigma} = \frac{1}{2} \left[\beta_{xz, \sigma} + \beta_{yz, \sigma} - 2\mu + \left(z_{xz, \sigma}^2 \tilde{e}_k + z_{yz, \sigma}^2 \tilde{f}_k \right) + \nu \sqrt{\left(z_{xz, \sigma}^2 \tilde{e}_k - z_{yz, \sigma}^2 \tilde{f}_k \right)^2 + 4z_{xz, \sigma}^2 z_{yz, \sigma}^2 \tilde{a}_k^2} \right] \quad (64)$$

where the dependence on σ is effective in the ferromagnetic phase only. The mass renormalization factors are constructed according to Eq. (47).

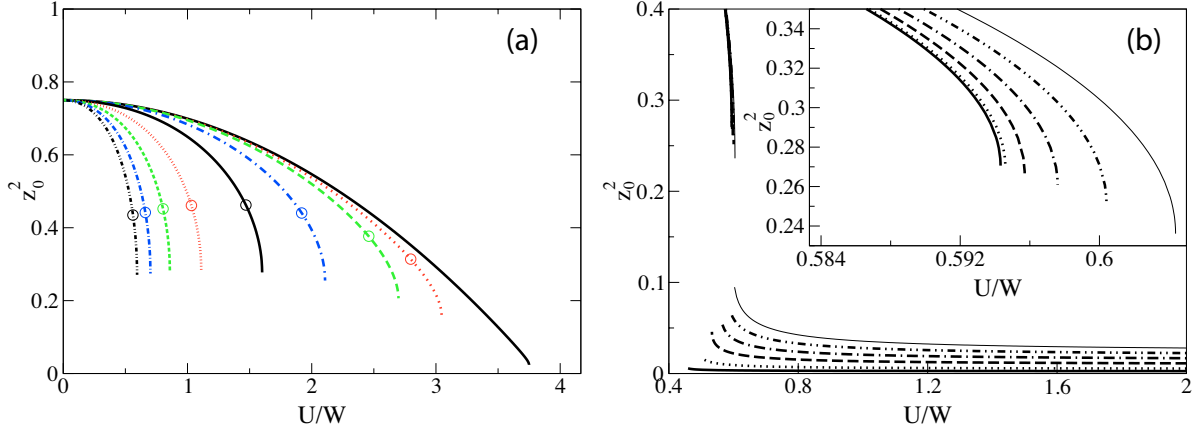


Fig. 4: (a) Effective mass renormalization at $\rho = 2$ for $J_H/U = 0$ (thick full line), 0.01 (thick dotted line), 0.02 (thick dashed line), 0.05 (thick dashed-dotted line), 0.1 (thin full line), 0.2 (thin dotted line), 0.3 (thin dashed line), 0.4 (thin dashed-dotted line) and 0.5 (thin dashed-dotted-dotted line). The circles indicate the location of the first order transition. (b) Effective mass renormalization off half filling for $J_H/U = 0.5$ and $\rho = 2.005$ (full line), 2.01 (dotted line), 2.02 (dashed line), 2.03 (dashed-dotted line), 2.04 (dashed-dotted-dotted line), and 2.05 (thin full line). Inset: Blow-up of the metallic solutions with the same parameters.

The saddle-point equations have been solved on a 800×800 lattice, at a temperature $T = t/1000$. We neglected four-fold occupancies and empty configurations since the electronic density in the ruthenates under study is $\rho \sim 2$. This approximation is justified in the vicinity of the Mott transition but breaks down for densities above three (below one) and for weak coupling, where our results should be taken with care. It is now well established that the Hund's rule coupling has a strong influence on the Mott transition. While the latter is second-order for $J_H = 0$ and $\rho = 2$ or for any J_H for $\rho = 1$ or 3, it becomes first-order for finite J_H at half filling as shown in Fig. 4a. In fact, no diverging effective mass is found. Instead, the metallic solution of the saddle-point equations ceases to exist at a critical value U_{c2} . Moreover the effective mass is at most renormalized by a factor of five for $J_H/U \geq 0.01$, in contrast to the one-band case. The saddle-point equations also possess an insulating paramagnetic solution: It is characterized by a vanishing value of all bosons except $d_{xz,\sigma}$ and $d_{yz,\sigma}$ and therefore a diverging effective mass (for finite J_H). It extends down to $U_{c1} = 0$. We remark that U_{c2} is only slightly larger than U_c , where the energy of the metallic and insulating solutions coincide. Consequently, the effective mass renormalization is even more modest in the metallic phase. Finally, Fig. 4a also shows that U_c strongly depends on J_H .

Once the system is doped the situation changes little by little. For small electron doping, the first-order transition remains but gradually vanishes with increasing electron concentration as shown in Fig. 4b. The metallic solution is only modestly affected, except that it allows for decreasing values of z in the vicinity of the Mott transition. The insulating solution becomes metallic under electron doping, and the truly insulating state is only found for integer fillings. However the effective mass renormalization remains very large, and accordingly the quasi particle residue is small. Under these circumstances, magnetic or even striped phases are likely to set in, and in addition the system may well be strongly influenced by other interaction terms,

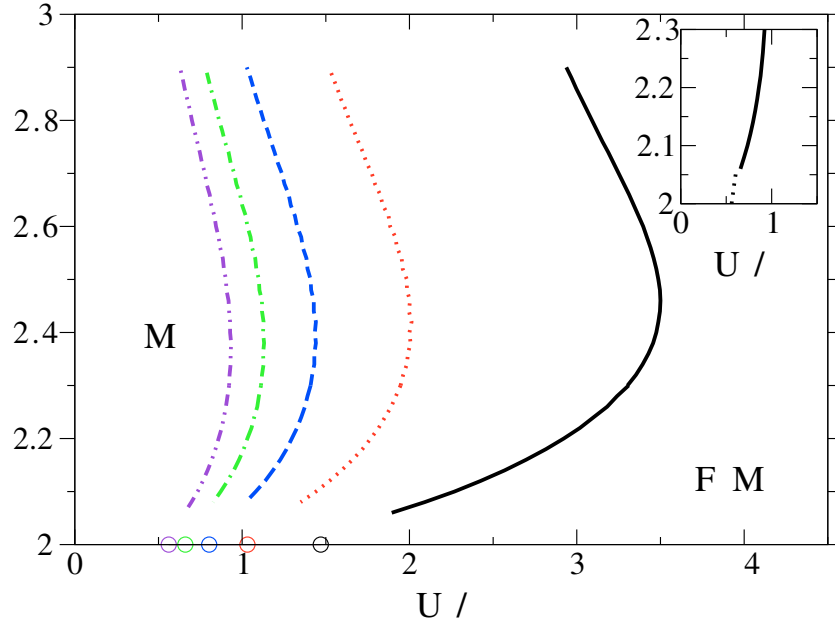


Fig. 5: Instability line towards ferromagnetism in the ρ - U plane for $J_H/U = 0.1$ (full line), 0.2 (dotted line), 0.3 (dashed line), 0.4 (dashed-dotted line), and 0.5 (dashed-dotted-dotted line). The circles are indicating the corresponding U_c at half filling. Inset: Instability line (full line), and first-order transition line (dotted line), for $J_H/U = 0.5$.

as reviewed by Vollhardt *et al.* [64], or disorder effects. This gives a qualitative explanation as to why many transition-metal oxides remain insulating even upon substantial doping, such as $\text{La}_{1-x}\text{Ca}_x\text{VO}_3$ [65] (for a review, see [66]).

The instabilities of the paramagnetic phases towards ferromagnetism are collected in Fig. 5, for several values of J_H/U . For large values of the latter, the range of stability of the paramagnetic phase is seen to depend weakly on density. In contrast, it may extend to large interaction strengths for $J_H/U = 0.1$. On top, there is a strong asymmetry around $\rho = 2.5$, which mostly follows from the difference between $U_c^{(2,2)}$ and $U_c^{(3,2)}$ and not from the neglecting of four-fold occupancies. As displayed in the inset of Fig. 5, the instability lines connect to the first-order transition line separating two paramagnetic solutions, where the latter ends, within numerical accuracy. No ferromagnetic solution, even with very small magnetization, has been found for very small doping and $U > U_c$.

When comparing this phase diagram to La-doped Ca_2RuO_4 , we see that a small amount of electron doping turns a Mott insulator into a ferromagnet, in agreement with experiment [67].

It should also be remarked that ferromagnetic instabilities only arise in the doped Mott insulating regime or, in other words, that ferromagnetism is a property of electrons undergoing strong local interactions.

An experimental attempt to reach such a ferromagnetic instability by enhancing the electronic correlations due to the reduction of the bandwidth in two-dimensional superlattices resulted in a ferromagnetic state with a high Curie temperature [30]. Yet the exact underlying effects seem to be more complex than the sole reduction of the dimensionality [68].

4.5 Magnetic order in the Anderson lattice model

The Anderson lattice model is believed to describe the physics of many transition-metal oxides as well as rare-earth and actinide compounds, including the so-called heavy fermion compounds. It is one of the archetypical models of correlated electrons on a lattice: It consists of a light conduction band hybridized with a strongly correlated narrow f -electron band. The physics is influenced by the strength of the onsite Coulomb repulsion in the f orbital, the hybridization strength, and the band filling. Depending on the values of these parameters, the model describes either localized moments interacting via spin exchange interaction (e.g. the RKKY interaction), which usually order at low temperature, or Kondo screened moments and heavy quasiparticles. The competition between these two ground states gives rise to a quantum phase transition [69, 70]. A qualitatively correct description (excluding the critical behavior at the quantum critical point, which requires a different approach) may be obtained within the SRI slave-boson SPA. The Hamiltonian of the Anderson lattice model reads

$$H = \sum_{\vec{k}\sigma} \varepsilon_{\vec{k}} c_{\vec{k}\sigma}^\dagger c_{\vec{k}\sigma} + \varepsilon_a \sum_{i,\sigma} a_{i\sigma}^\dagger a_{i\sigma} + V \sum_{i,\sigma} \left(c_{i\sigma}^\dagger a_{i\sigma} + a_{i\sigma}^\dagger c_{i\sigma} \right) + U \sum_i a_{i\uparrow}^\dagger a_{i\uparrow} a_{i\downarrow}^\dagger a_{i\downarrow}, \quad (65)$$

where $c_{i\sigma} = \sum_{\vec{k}} e^{i\vec{k} \cdot \vec{R}_i} c_{\vec{k}\sigma}$ and $\vec{R}_i$ is the lattice vector at site i . H may be represented in terms of SRI slave-boson operators as

$$\begin{aligned} H = & \sum_{\vec{k}\sigma} \varepsilon_{\vec{k}} c_{\vec{k}\sigma}^\dagger c_{\vec{k}\sigma} + \varepsilon_a \sum_i \left(\sum_{\mu} p_{i\mu}^\dagger p_{i\mu} + 2d_i^\dagger d_i \right) + V \sum_{i,\sigma,\sigma'} \left(c_{i\sigma}^\dagger z_{i\sigma'\sigma} f_{i\sigma'} + h.c. \right) \\ & + U \sum_i d_i^\dagger d_i + \sum_i \alpha_i \left(e_i^\dagger e_i + \sum_{\mu} p_{i\mu}^\dagger p_{i\mu} + d_i^\dagger d_i - 1 \right) \\ & + \sum_i \left\{ \beta_{i0} \left(\sum_{\mu} p_{i\mu}^\dagger p_{i\mu} + 2d_i^\dagger d_i \right) + \vec{\beta}_i \cdot (p_{i0}^\dagger \vec{p}_i + \vec{p}_i^\dagger p_{i0}) \right\}. \end{aligned} \quad (66)$$

An application of the SPA to this Hamiltonian describing spiral magnetic states has been considered in [71]. There, the nonmagnetic boson saddle-point amplitudes e, d, p_0 and Lagrange parameters α, β_0 have been assumed spatially uniform, while the magnetic parameters $\vec{p}_i$ and $\vec{\beta}_i$ were taken to have the spatial dependence of a spiral vector field, $\vec{p}_i = p(\cos \phi_i, \sin \phi_i, 0)$ and $\vec{\beta}_i = \beta(\cos \phi_i, \sin \phi_i, 0)$ oriented perpendicular to the z -axis in spin space, and $\phi_i = \vec{q} \cdot \vec{R}_i$. The spatial periodicity characterized by the wave vector $\vec{q}$ leads to a coupling of Bloch states at wave vectors $\vec{k}$ and $\vec{k} + \vec{q}$. The energy matrix of the hybridized bands then takes the form

$$\epsilon_{\vec{k}} = \begin{pmatrix} \epsilon_{\vec{k}} - \mu & Vz_+ & 0 & Vz_- \\ Vz_+ & \epsilon_a + \beta_0 - \mu & Vz_- & \beta \\ 0 & Vz_- & \epsilon_{\vec{k}+\vec{q}} - \mu & Vz_+ \\ Vz_- & \beta & Vz_+ & \epsilon_a + \beta_0 - \mu \end{pmatrix} \quad (67)$$

where the weight factors $z_{\pm}$ are defined by

$$z_{\pm} = \frac{ep_+ + dp_-}{\sqrt{1-d^2-p_+^2} \sqrt{1-e^2-p_-^2}} \pm \frac{ep_- + dp_+}{\sqrt{1-d^2-p_-^2} \sqrt{1-e^2-p_+^2}} \quad (68)$$

with $p_{\pm} = (p_0 \pm p)/\sqrt{2}$. Requiring that the free energy

$$F = -\frac{1}{\beta} \sum_{\vec{k}\sigma\alpha} \ln \left[1 + e^{\frac{-E_{\vec{k},\sigma,\alpha}}{T}} \right] + U d^2 + \alpha(e^2 + p_0^2 + p^2 + d^2) - \beta_0(p_0^2 + p^2 + 2d^2) + 2\beta p_0 p \quad (69)$$

be stationary yields the saddle-point values. Here $E_{\vec{k},\sigma,\alpha}$ are the eigenvalues of the energy matrix $\epsilon_{\vec{k}}$ given in Eq. (67).

The zero-temperature phase diagram in the (t/U) - δ plane (with the t nearest-neighbor hopping amplitude and δ the filling factor of the conduction band) has been calculated in [71]. Spiral magnetic states have been found in a wide region, with wave vector $\vec{q}$ approaching the edge of the Brillouin zone at $\delta = 1$ (antiferromagnetic order). Approaching the limit $\delta = 0$, one finds a ferromagnetic region, followed by another antiferromagnetic state very close to $\delta = 0$. These findings have been confirmed by quantum Monte Carlo simulations [72]. One should keep in mind that the spatial dimension enters only through the energy dispersion of the conduction electrons. These results are therefore best applicable in three or higher dimensions, where fluctuation effects are expected to be small.

5 Fluctuation corrections to the saddle-point approximation: SRI representation of the Hubbard model

The spin and charge response functions of the Hubbard model have been considered as well. In particular, in the SRI representation they may be directly evaluated, as all degrees of freedom have been mapped onto bosons. Indeed, the spin and density fluctuations may be expressed as

$$\sum_{\sigma} \sigma \delta n_{\sigma} = \delta(p_0^{\dagger} p_3 + p_3^{\dagger} p_0) \equiv \delta S_z \quad \text{and} \quad \sum_{\sigma} \delta n_{\sigma} = \delta(d^{\dagger} d - e^{\dagger} e) \equiv \delta N. \quad (70)$$

This allows one to write the spin and charge autocorrelation functions in terms of the slave-boson correlation functions as

$$\begin{aligned} \chi_s(k) &= \sum_{\sigma, \sigma'} \sigma \sigma' \langle \delta n_{\sigma}(-k) \delta n_{\sigma'}(k) \rangle = \langle \delta S_z(-k) \delta S_z(k) \rangle \\ \chi_c(k) &= \sum_{\sigma \sigma'} \langle \delta n_{\sigma}(-k) \delta n_{\sigma'}(k) \rangle = \langle \delta N(-k) \delta N(k) \rangle. \end{aligned} \quad (71)$$

Performing the calculation to one-loop order, one can make use of the propagators given in the appendix of Ref. [42] to obtain

$$\begin{aligned} \chi_s(k) &= 2p_0^2 S_{77}^{-1}(k) \\ \chi_c(k) &= 2e^2 S_{11} S^{-1}(k) - 4ed S_{12}^{-1}(k) + 2d^2 S_{22}^{-1}(k). \end{aligned} \quad (72)$$

It should be emphasized here that Fermi liquid behavior is obtained when considering the above $\chi_s(k)$ and $\chi_c(k)$ in the long-wavelength and low-frequency limit [73, 74]. The obtained Landau parameters involve effective interactions that differ in the spin channel and in the charge

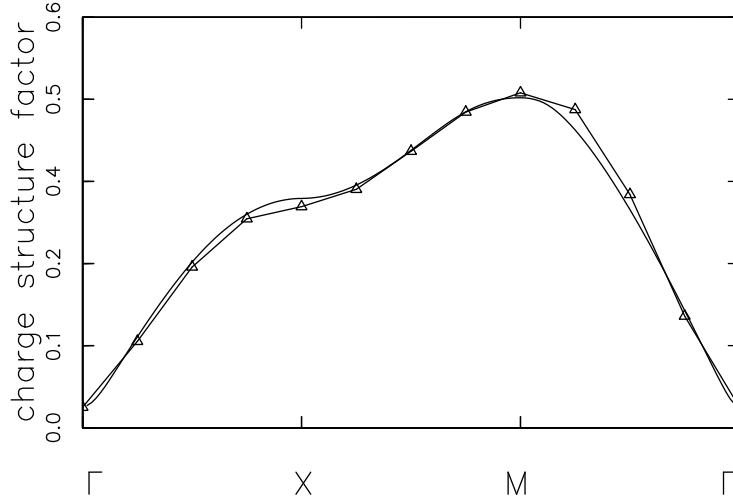


Fig. 6: Comparison of the Quantum Monte Carlo (triangles) and Slave-Boson (full line) charge structure factors for $U = 4t$, $\delta = 0.275$ and $\beta = 6/t$.

channel. Performing the algebra at half filling yields for a rectangular DOS

$$F_0^a = -1 + \frac{1}{(1 + U/U_c)^2} \quad \text{and} \quad F_0^s = \frac{U(2U_c - U)}{(U_c - U)^2}. \quad (73)$$

As can be seen in Eq. (73) F_0^a remains larger than -1 when reaching the Mott transition, while F_0^s diverges (for a recent manifestation of a related behavior see [75]).

Ferromagnetic instabilities have been investigated, too [77], as well as ferromagnetic phases. In particular, in the limit $U \rightarrow \infty$, it could be shown analytically that the fully polarized ferromagnetic ground state and the paramagnetic ground state are degenerate at density $\rho = 2/3$ for any bipartite lattice [43]. For lower densities the ground state is paramagnetic.

Yet, in such an analysis, the focus is put on a ferromagnetic instability only, while other commensurate or even incommensurate instabilities should be considered as well. This analysis has been carried out for the Hubbard model on the square lattice [77]. Off half filling it turned out that the leading instabilities are systematically towards incommensurate states characterized by a wave vector (Q_x, π) for $U < 57t$ with Q_x smoothly varying from π for $U = 0^+$ down to 0 for $U = 57t$. While the Hubbard model was initially introduced, *inter alia*, to describe metallic magnetism [78, 79], this result shows that ferromagnetism is confined to the very large U regime. Further, for the largest U , the wave vector characterizing the instability is rather of the form $(0, Q_y)$, with $Q_y \simeq \pi$.

The computation of the charge structure factor has been performed, too, in particular with the aim of putting forward charge instabilities [42]. The result turned negative, even for the t - t' - U repulsive Hubbard model [80]. Instead, the charge structure factor quite systematically consists of one broad peak centered at (π, π) . As an example, we compare in Fig. 6 the slave-boson result with quantum Monte Carlo simulations by Dzierzawa [81], for $U = 4t$ and $\delta = 0.275$ at temperature $T = t/6$. The agreement between both approaches is excellent, as the difference does not exceed a few percent.

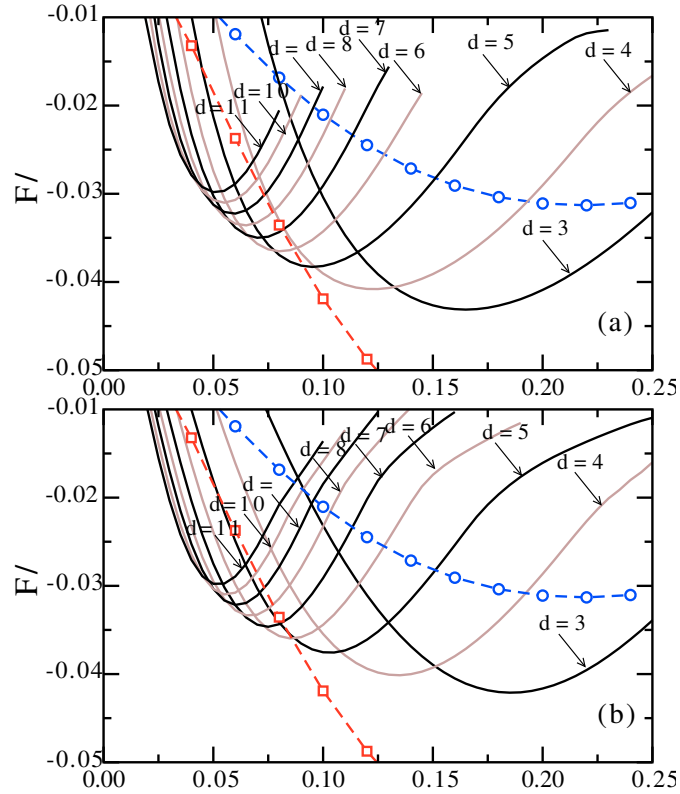


Fig. 7: Free energy gain δF per site with respect to the AF phase as a function of doping x , obtained for the t - t' - U Hubbard Model with $U = 12t$ and $t' = -0.3t$ for: (a) Vertical site-centered striped phases; (b) vertical bond-centered striped phases. Domain walls are separated by $d = 3, \dots, 11$ lattice constants. Circles and squares show the corresponding data for vertical and diagonal spiral order, respectively.

5.1 Magnetic and striped phases

Since the leading instabilities of the paramagnetic phase are generally towards incommensurate phases, spiral and striped phases have been thoroughly investigated [43–46, 48–50, 57]. Comparison of ground-state energies in spiral phases with numerical simulations showed very good agreement. For instance, for $U = 4t$ it could be shown that the SB ground state energy is larger than its counterpart by less than 3% [44]. For larger values of U , it has been obtained that the SB ground state energy exceeds the exact diagonalization data by less than 4% (7%) for $U = 8t$ ($20t$) and doping larger than 15%. The discrepancy increases when the doping is lowered [46]. Regarding the pure Hubbard model, calculations on $L \times L$ clusters with $L > 100$ showed that magnetic striped phases are generally slightly more stable than spiral phases. However, the situation is more intricate for the t - t' - U repulsive Hubbard model. As shown in Fig. 7 for an intermediate value of t' , a large number of phases compete. While the vertical site-centered striped phases are generally lower in energy than the vertical bond-centered striped phases at low doping δ , the opposite result is found at larger δ . For instance, for $U = 12t$, the transition occurs at $\delta \simeq 0.16$ for $t' = -0.15t$ and at $\delta \simeq 0.18$ for $t' = -0.3t$. Yet, in the latter case, the diagonal spiral phase is lower in energy for $\delta \geq 0.09$, in contrast to the former case [50].

6 Extended Hubbard model

The Hubbard model assumes a perfect screening of the long-range part of the Coulomb interaction. This may be questionable and the relevance of this approximation may be assessed by considering the extended Hubbard model that reads

$$H = \sum_{i,j,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i \left(n_{i\uparrow} - \frac{1}{2} \right) \left(n_{i\downarrow} - \frac{1}{2} \right) + \frac{1}{2} \sum_{i,j} V_{ij} (1 - n_i)(1 - n_j) + \frac{1}{2} \sum_{i,j} J_{ij} \vec{S}_i \cdot \vec{S}_j. \quad (74)$$

It includes intersite Coulomb V_{ij} and exchange J_{ij} interactions. These elements decay fast with increasing distance $|\vec{R}_i - \vec{R}_j|$ but extend in general beyond nearest neighbors. The particle-hole symmetric form for both density-density interaction terms is consistently used.

Although one expects that $V_{ij} > 0$, in certain cases effective intersite Coulomb interactions may be attractive [82]. Therefore, $\{V_{ij}\}$ may be treated as effective parameters. Similarly, for the exchange elements $\{J_{ij}\}$ both antiferromagnetic ($J_{ij} > 0$) and ferromagnetic ($J_{ij} < 0$) exchange elements may be considered. For more details see [83].

In the SRI representation [15, 16] the Hamiltonian Eq. (74) may be represented as

$$\begin{aligned} H = & \sum_{i,j,\sigma} t_{ij} \sum_{\sigma\sigma'\sigma_1} z_{i\sigma_1\sigma}^\dagger f_{i\sigma}^\dagger f_{j\sigma'} z_{j\sigma'\sigma_1} + U \sum_i \left(d_i^\dagger d_i - \frac{1}{2} \sum_{\sigma} f_{i\sigma}^\dagger f_{j\sigma'} + \frac{1}{4} \right) \\ & + \frac{1}{4} \sum_{i,j} V_{ij} \left[\left(1 - \sum_{\sigma} f_{i\sigma}^\dagger f_{i\sigma} \right) Y_j + Y_i \left(1 - \sum_{\sigma} f_{j\sigma}^\dagger f_{j\sigma} \right) \right] \\ & + \frac{1}{2} \sum_{i,j} J_{ij} \sum_{\sigma\sigma'\sigma_1} \vec{\tau}_{\sigma\sigma'} p_{i\sigma\sigma_1}^\dagger p_{i\sigma_1\sigma'} \cdot \sum_{\rho\rho'\rho_1} \vec{\tau}_{\rho\rho'} p_{j\rho\rho_1}^\dagger p_{j\rho_1\rho'}, \end{aligned} \quad (75)$$

where we used the representation of the physical quantities in terms of slave bosons Eq. (16) and expressed the hole doping operator as $Y_i \equiv e_i^\dagger e_i - d_i^\dagger d_i$.

6.1 Saddle-point approximation to the extended Hubbard model

In the paramagnetic phase the saddle-point approximation to the extended Hubbard model (74) runs in a fashion analogous to section 4.2 [83], though with the quasiparticle dispersion (Eq. (41)) modified into

$$E_{\vec{k}\sigma} = z_0^2 t_{\vec{k}} + \beta_0 - \frac{1}{2} U - \frac{1}{2} V_0 Y - \mu, \quad (76)$$

in which the Fourier transform of the intersite Coulomb repulsion, $V_{\vec{k}} = \frac{1}{L} \sum_{i,j} V_{ij} e^{-i\vec{k} \cdot (\vec{R}_j - \vec{R}_i)}$, was introduced. The steps leading to the saddle-point equations Eqs. (44) can be repeated, and the final saddle-point equation is again given by Eq. (46). We therefore obtain the remarkable result that the slave-boson mean values are independent of $\{J_{ij}\}$ and $\{V_{ij}\}$. Hence, in a paramagnetic phase, the intersite interactions only influence the fluctuations and do not change electron localization due to strong onsite interaction U . In particular, the nearest-neighbor Coulomb interaction V has no influence on the Mott gap, and the results obtained by Lavagna [35] also apply to the extended Hubbard model.

6.2 Landau parameters

In this section we present the homogenous spin and charge instabilities that are generated or magnified by the intersite Coulomb and exchange interactions. We follow the derivation of Lhottellier *et al.* [83] and make use of the inverse propagator matrix they derived. Moreover, spin and charge fluctuations separate at the one-loop order, and the intersite Coulomb (exchange) elements have no effect on the value of F_0^a (F_0^s). We recall that a ferromagnetic (charge) instability is identified by $F_0^a = -1$ ($F_0^s = -1$).

6.2.1 Ferromagnetic instability — F_0^a parameter

Fortunately enough, an analytical expression of the Landau parameter F_0^a at half filling can be obtained. It reads

$$F_0^a = 2N_F^{(0)} \bar{\varepsilon} \left\{ \frac{u(2+u)}{(1+u)^2} - \frac{J_0/U_c}{1-u^2} \right\}, \quad (77)$$

where we have introduced $u = U/U_c$ and the bare density of states at the Fermi energy $N_F^{(0)}$. Eq. (77) consists of a regular and a singular part. The regular part generalizes the result Eq. (73) to an arbitrary DOS. It follows from the Hubbard model and has been discussed in much detail [59, 74]. In particular, in the metallic phase at half filling, it systematically yields values of F_0^a larger than -1 for generic lattices such as the cubic lattice for which $2N_F^{(0)} \bar{\varepsilon} = -1.14$. The singular part depends solely on the $\vec{k} = 0$ component of the exchange coupling, $J_0 \equiv J_{\vec{k}=0}$ and not on the details of $\{J_{ij}\}$ [84]. It triggers a ferromagnetic (FM) instability in the metallic phase for $J_0 < 0$ regardless of its value, while it stiffens the spin response for $J_0 > 0$, as shown in Fig. 8a. We emphasize that the ferromagnetic instability deduced from Eq. (77) is general and occurs in all cases *below* the metal-to-insulator transition when $J_0 < 0$. This result follows from the band narrowing when $U \rightarrow U_c$, which amplifies the effects of the intersite exchange interaction.

The physical origin of Eq. (77) lies in the fact that, in the limit of vanishing hopping, the Hubbard model at half filling favors the formation of localized magnetic moments that order according to the exchange couplings, for instance ferromagnetically for $J_0 < 0$. Further, our result suggests that a minimum of coherence of the quasi particles z_F^2 is necessary to destabilize the ferromagnetic order. It only depends on $j_0 \equiv J_0/U$ and, for a rectangular DOS for which $2N_F^{(0)} \bar{\varepsilon} = -1$, reads

$$z_F^2 = \frac{4j_0 + j_0^2 + (1-j_0)\sqrt{1-6j_0+j_0^2} - 1}{4j_0^2}. \quad (78)$$

It behaves as $z_F^2 \simeq -2j_0$ for small FM exchange. Hence, for $J_0 \rightarrow 0^-$ the FM instability takes place at $U = U_c$, while it is absent for $J_0 = 0$. This is the only case for which the spin susceptibility is finite at the Brinkman-Rice point U_c . Fig. 8a shows that the location of the FM instability depends rather sensitively on the FM coupling, from U_c^- for $J = 0^+$ down to $0.33 U_c$ for $J/U = -0.2$.

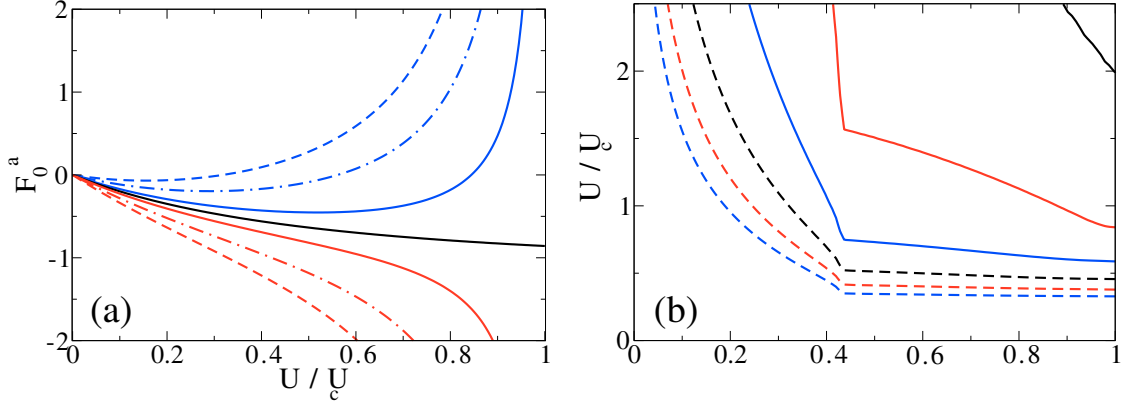


Fig. 8: (a) Landau parameter F_0^a for the extended Hubbard model at half filling on the cubic lattice with different lines from top to bottom for decreasing $J/U = 0.2, 0.1, 0.04, 0, -0.04, -0.1, -0.2$. (b) Instability lines of the unpolarized state towards FM order as given by the divergence of the magnetic susceptibility (Landau parameter $F_0^a = -1$) for the extended Hubbard model with FM exchange $J < 0$ on the cubic lattice with different lines from top to bottom for $J/U = 0, -0.01, -0.05, -0.1, -0.15, -0.2$.

Also, away from half filling, finite FM exchange coupling $J_0 < 0$ triggers the FM instability at significantly lower values of U . For instance, Fig. 8b shows that, in the case of the cubic lattice, $J/U = -0.01$ already brings this instability down to the values of $U \sim 20t$ for the doping $\delta < 0.57$ where the DOS is almost independent of energy. When $J/U = -0.05$, the FM instability occurs at $U < 10t$ in the same doping regime and comes down also for lower electron fillings. For lower J_0 , the FM instability occurs at even lower values of U . This is in contrast to the calculations for the two-band model presented in section 4.4, where the FM instability was only found in the doped Mott insulator regime. In that case, no intersite FM coupling is needed and the FM instability follows from Hund's exchange.

On the contrary, an antiferromagnetic coupling suppresses the FM instability, and the value of $J/U = 0.1$ totally removes ferromagnetism.

6.2.2 Charge instability — F_0^s parameter

The symmetric Landau parameter F_0^s , which stands for the charge response, has to be evaluated numerically even at half filling, except for $V = 0$. As expected, F_0^s vanishes for $U = 0$, as F_0^a does. Otherwise, the symmetric parameter F_0^s increases with U in the entire regime of filling $0 < \rho \leq 1$. This increase is stronger near half filling, where $F_0^s > 10$ for $U/U_c > 0.7$ in a range of small doping away from half filling, see Fig. 9a. At half filling the value of the positive F_0^s is given by Eq. (73). It rapidly increases and finally diverges at the metal-to-insulator transition (we recall that for the simple cubic lattice $U_c \simeq 16.04t$). Away from $\rho = 1$, the increase of F_0^s is moderate, and it follows the same pattern as $1/z^2$ in Fig. 1, being another manifestation of strong electron correlations near half filling. The increase of F_0^s with increasing U/U_c is enhanced by a positive intersite Coulomb repulsion in the extended Hubbard model. When $V > 0$, one finds even a stronger increase of F_0^s near half filling, and finally it becomes even larger than $F_0^s = 10$ in a broad range of filling $\rho > 0.6$ for the cubic lattice at $V = 0.2U$. The

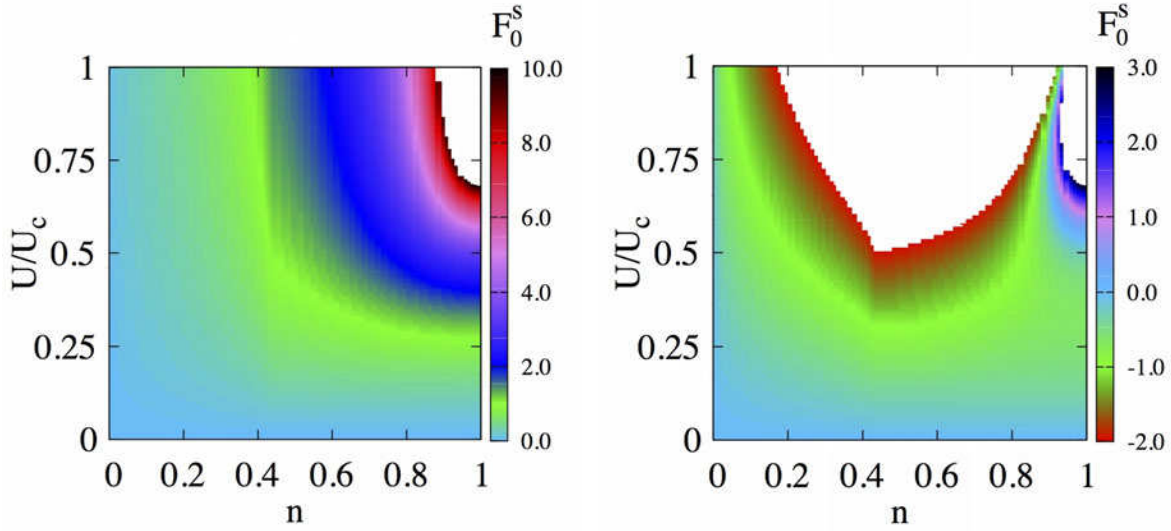


Fig. 9: (a) Landau parameter F_0^s for the Hubbard model on the cubic lattice. Here the white region stands for values $F_0^s > 10$. No instability is found. (b) Landau parameter F_0^s for the extended Hubbard model on the cubic lattice with attractive intersite interaction $V/U = -0.2$. Large values of $F_0^s > 3$ are found only near $\rho = 1$, while the charge instability $F_0^s = -1$ occurs for a broad range of $0.045 < n < 0.93$. Note that the instability line $F_0^s = -1$ extends to $n = 1^-$, and stops at $U \simeq 1.246 U_c$.

uniform charge distribution is therefore more robust in the regime of $\rho \simeq 1$, if $V/U > 0$.

The uniform charge distribution is destabilized by attractive charge interactions $V < 0$, particularly in the regime near quarter filling. At $V = -0.2U$ the value of F_0^s decreases with increasing U for any filling ρ and this decrease is fastest near quarter filling. For $U < U_c$ one finds the charge instability at $F_0^s = -1$ in a broad range of $\rho \in (0.045, 0.93)$. This instability is related to the shape of the DOS and is easiest to realize at $\rho \simeq 0.42$, where the DOS has a van Hove singularity. Remarkably, U and V cooperate to cause this striking tendency towards phase separation that is absent for $V = 0$.

The data of Fig. 9b suggest that in the case of charge response the regime near the metal-to-insulator transition at half filling is robust and the Landau parameter F_0^s is here always enhanced, even for $V < 0$.

We now inspect the case $\rho = 1$ in more detail. It can be noticed in Fig. 10 that F_0^s is reduced for attractive V while it is enhanced for repulsive V . The reduction of F_0^s occurs only for sufficiently large $-V$ and is visible in Fig. 10 for $V/U = -0.15$, and beyond. As a result, a minimum in F_0^s develops at $U \simeq 0.4 U_c$, the minimal value of F_0^s decreases with increasing $-V$, and a charge instability may be found at the critical value $V/U < -0.234$, see the inset in Fig. 10. The instability moves towards lower values of U with decreasing V when the minimum of F_0^s becomes deeper with further decreasing V . Particularly interesting is the non-monotonic behavior of F_0^s with increasing U for $V < 0$. We therefore suggest that a sufficiently strong intersite Coulomb attraction $-V/U > 0.234$ is necessary to induce phase separation. The instability is absent for repulsive V , where the uniform charge distribution is locally stable. The nature of the instabilities at finite wave vector is an open question.

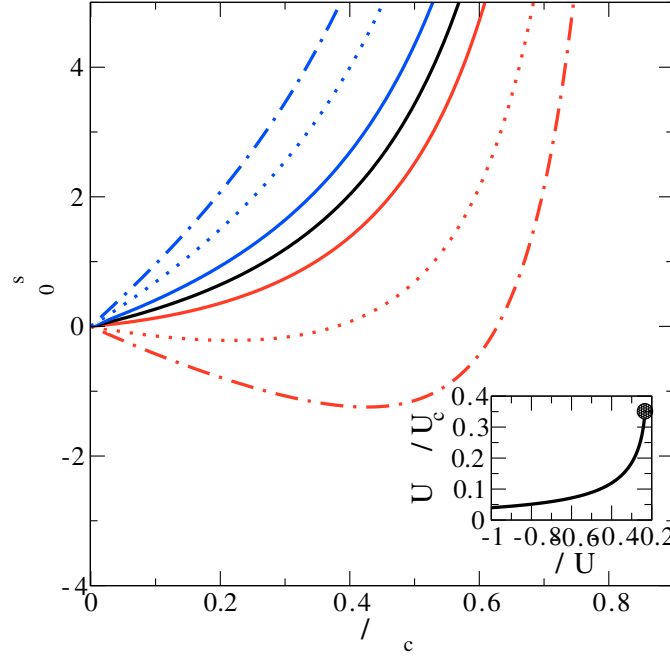


Fig. 10: Landau parameter F_0^s for the extended Hubbard model on the cubic lattice at half filling ($\rho = 1$) for selected decreasing values of intersite Coulomb interaction V from top to bottom: $V/U=0$ (black line), $V/U > 0$ (blue): $V/U=0.05$ (solid line), $V/U=0.15$ (dotted) and $V/U=0.25$ (dashed-dotted line), and $V/U < 0$ (red): $V/U=-0.05$ (solid line), $V/U=-0.15$ (dotted) and $V/U=-0.25$ (dashed-dotted line). The inset shows the instability value U_{inst}/U_c for $V/U \in [-1.0, -0.2]$. Its end point is marked by a solid circle.

7 Summary

We have reviewed the most prominent auxiliary particle techniques and their applications to strongly correlated electron systems, using a variety of approximation schemes, ranging from saddle-point approximations, possibly with Gaussian fluctuations, to exact evaluation of quantities represented in the functional integral formalism.

It has been shown how to handle the radial SB fields that appear when making use of the gauge symmetry associated to a particular SB representation to gauge away the phase degree of freedom of the SB. It was further made evident that the exact expectation value of a radial SB field is generally finite and unrelated to a Bose condensation.

It was seen that the Kotliar-Ruckenstein representation, especially in its spin-rotation invariant formulation, is particularly useful for identifying complex spin- and/or charge-ordered ground states in saddle-point approximations, since it treats all spin and charge states on a lattice site on the same footing. Regarding the Hubbard model on the square lattice, unrestricted Hartree-Fock calculations point towards a huge number of solutions. An indication that this is also realized using slave bosons on the saddle-point level is provided by Fig. 7, but identifying the numerous competing phases remains a challenge. Yet ferromagnetic ground states could only be found in the very large $U > 8W$ regime, as a reminiscence of Nagaoka ferromagnetism.

The Kotliar-Ruckenstein slave-boson technique has also been applied to the orbitally degener-

ate case. The main results are the following a) low-energy single-particle quantities such as the critical value of the interaction strength at which the transition occurs, the quasiparticle residue and the single-particle Mott-Hubbard gap depend very weakly on degeneracy, justifying the agreement between theory and experiment when it was applied to orbitally degenerate systems, b) the degeneracy temperature decreases with increasing band degeneracy, c) the Mott-Hubbard transition depends strongly on J_H , d) there is a coexistence region of metallic-like and insulating-like solutions of the saddle-point equations, e) ferromagnetism appears as a property of doped Mott insulators.

Results have been presented for a Hubbard model extended with long-ranged Coulomb and exchange interactions. It was shown that they have no effect on the Mott-Hubbard gap in the paramagnetic phase. Calculations of the Landau parameter F_0^s show that attractive interactions lead to charge instabilities in a broad density range away from half filling, signaling a tendency towards phase separation. The presented calculations of F_0^a predict a ferromagnetic instability in a strongly correlated metallic system with globally ferromagnetic exchange. The analytic result for F_0^a Eq. (77) uncovers that, for any lattice, the Hubbard model at half filling is on the verge of a ferromagnetic instability, which is triggered by an infinitesimal ferromagnetic inter-site exchange. This result provides a new context for the original idea of Kanamori [79], who introduced the Hubbard model as the simplest model of itinerant ferromagnetism.

Acknowledgment

I am deeply indebted to Peter Wölfle and Thilo Kopp for all the enlightening discussions we had along many years that form the foundation for this review, and Klaus Doll, Michael Dzierzawa, Hans Kroha, Gabi Kotliar, Grégoire Lhoutellier, Ulrike Lüders, Burkhard Möller, Andrzej Oleś, Henni Ouerdane, Marcin Raczkowski, and Walter Zimmermann for valuable collaboration. I gratefully acknowledge financial support by the Région Basse-Normandie and the Ministère de la Recherche.

References

- [1] M. Imada, A. Fujimori, and Y. Tokura, *Rev. Mod. Phys.* **70**, 1039 (1998)
- [2] P.A. Lee, T.M. Rice, J. Serene, L.J. Sham and J.W. Wilkins, *Comments on Condensed Matter Physics*, **12**, 99 (1986)
- [3] J.G. Bednorz and K.A. Müller, *Z. Physik B* **64**, 189 (1986);
B. Raveau, C. Michel, M. Hervieu, and D. Groult:
Crystal Chemistry of High- T_c Superconducting Copper Oxides,
Springer Series in Materials Science **15** (Springer, Heidelberg, 1991)
- [4] P.W. Anderson, *Science* **235**, 1196 (1987)
- [5] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993)
- [6] R. Frésard, J. Kroha, and P. Wölfle in: *Theoretical Methods for Strongly Correlated Systems*,
A. Avella and F. Mancini (eds.), Springer Series in Solid-State Sciences **171**
(Springer, Heidelberg, 2012) pp. 65–101
- [7] T. Holstein and H. Primakoff, *Phys. Rev.* **58**, 1094 (1940)
- [8] J. Schwinger: *Quantum Theory of Angular Momentum*,
H. Biedenharn and H. Van Dam (eds.) (Academic, New York, 1965)
- [9] A.A. Abrikosov, *Physics* (Long Island City, N.Y.) **2**, 5 (1965)
- [10] B. Coqblin and J.R. Schrieffer, *Phys. Rev.* **185**, 847 (1969)
- [11] S.E. Barnes, *J. Phys. F* **6**, 1375 (1976); **7**, 2637 (1977)
- [12] P. Coleman, *Phys. Rev. B* **29**, 3035 (1984)
- [13] M. Gutzwiller, *Phys. Rev. Lett.* **10**, 159 (1963)
- [14] G. Kotliar and A.E. Ruckenstein, *Phys. Rev. Lett.* **57**, 1362 (1986)
- [15] T.C. Li, P. Wölfle, and P.J. Hirschfeld, *Phys. Rev. B* **40**, 6817 (1989)
- [16] R. Frésard and P. Wölfle, *Int. J. of Mod. Phys. B* **6**, 685 (1992),
Int. J. of Mod. Phys. B **6**, 3087 (1992)
- [17] R. Frésard and G. Kotliar, *Phys. Rev. B* **56**, 12 909 (1997)
- [18] H. Hasegawa, *J. Phys. Soc. Jpn.* **66**, 1391 (1997)
- [19] F. Lechermann, A. Georges, G. Kotliar, and O. Parcollet, *Phys. Rev. B* **76**, 155102 (2007)
- [20] J. Bünamann, *Phys. Stat. Sol. (b)* **248**, 203 (2011)

- [21] R. Frésard, H. Ouerdane, and T. Kopp, Nucl. Phys. B **785**, 286 (2007)
- [22] R. Frésard and T. Kopp, Ann. Phys. (Berlin) **524**, 175 (2012)
- [23] C. Castellani, C.R. Nanoli and J. Ranninger, Phys. Rev. B **18**, 4945 (1978)
- [24] R. von Helmolt, J. Wecker, B. Holzapfel, L. Schultz, and K. Samwer, Phys. Rev. Lett. **71**, 2331 (1993);
Y. Tomioka, A. Asamitsu, Y. Moritomo, H. Kuwahara, and Y. Tokura, Phys. Rev. Lett. **74**, 5108 (1995)
B. Raveau, A. Maignan, and V. Caignaert, J. Solid State Chem. **117**, 424 (1995);
A. Maignan, C. Simon, V. Caignaert, and B. Raveau, Solid State Commun. **96**, 623 (1995)
- [25] H. Kawazoe, H. Yasakuwa, H. Hyodo, M. Kurota, H. Yanagi, and H. Hosono, Nature **389**, 939 (1997)
- [26] L. Li, C. Richter, S. Paetel, T. Kopp, J. Mannhart, and R.C. Ashoori, Science **332**, 825 (2011)
- [27] H. Ohta et al, Nature Mat. **6**, 129 (2007);
R. Frésard, S. Hébert, A. Maignan, L. Pi, and J. Hejtmanek, Phys. Lett. A **303**, 223 (2002);
A. Maignan, C. Martin, R. Frésard, V. Eyert, E. Guilmeau, S. Hébert, M. Poienar, and D. Pelloquin, Solid State Commun. **149**, 962 (2009)
- [28] N. Reyren, S. Thiel, A.D. Caviglia, L. Fitting Kourkoutis, G. Hammerl, C. Richter, C.W. Schneider, T. Kopp, A.-S. Rüetschi, D. Jaccard, M. Gabay, D.A. Muller, J.-M. Triscone, and J. Mannhart, Science **317**, 1196 (2007)
- [29] V. Eyert, U. Schwingenschlögl, C. Hackenberger, T. Kopp, R. Frésard, and U. Eckern, Prog. Solid State Chem. **36**, 156 (2008)
- [30] U. Lüders, W.C. Sheets, A. David, W. Prellier, and R. Frésard, Phys. Rev. B **80**, 241102(R) (2009)
- [31] R. Frésard and T. Kopp, Nucl. Phys. B **594**, 769 (2001)
- [32] The correctness of a particular representation may be verified through the exact evaluation of, e.g., the partition function and Green's functions in the atomic limit.
- [33] N. Read, D.M. Newns, J. Phys. C **16**, L1055 (1983); *ibid.* 3273 (1983);
D.M. Newns and N. Read, Adv. in Physics **36**, 799 (1987)
- [34] J.W. Rasul, T. Li, J. Phys. C **21**, 5119 (1988); T.C. Li and J.W. Rasul, Phys. Rev. B **39**, 4630 (1989);
J.W. Rasul, T.C. Li and H. Beck, Phys. Rev. B **39**, 4191 (1989)

- [35] M. Lavagna, Phys. Rev. B **41**, 142 (1990); Helvetica Physica Acta **63**, 310 (1990); Int. J. Mod. Phys. B **5**, 885 (1991)
- [36] Th. Jolicœur and J.C. Le Guillou, Phys. Rev. B **44**, 2403 (1991)
- [37] Y. Bang, C. Castellani, M. Grilli, G. Kotliar, R. Raimondi and Z. Wang, Int. J. of Mod. Phys. B **6**, 531 (1992)
- [38] W. Metzner, and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989); Phys. Rev. B **37**, 7382 (1988); W. Metzner, Z. Phys. B **77**, 253 (1989)
- [39] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989)
- [40] J. Kroha, P. Hirschfeld, K.A. Muttalib, and P. Wölfle Solid State Comm. **83**, 1003 (1992); J. Kroha, P. Wölfle, and T.A. Costi, Phys. Rev. Lett. **79**, 261 (1997); J. Kroha and P. Wölfle, Acta Phys. Pol. B **29**, 3781 (1998); S. Kirchner and J. Kroha, J. Low Temp. Phys. **126**, 1233 (2002)
- [41] R. Frésard and T. Kopp, Phys. Rev. B. **78**, 073108 (2008)
- [42] W. Zimmermann, R. Frésard, and P. Wölfle, Phys. Rev. B **56**, 10 097 (1997)
- [43] B. Möller, K. Doll, and R. Frésard, J. Phys. Condensed Matter **5** 4847 (1993)
- [44] R. Frésard, M. Dzierzawa and P. Wölfle, Europhys. Lett. **15**, 325 (1991)
- [45] E. Arrigoni and G.C. Strinati, Phys. Rev. B **44**, 7455 (1991)
- [46] R. Frésard and P. Wölfle, J. Phys.: Condensed Matter **4** 3625 (1992)
- [47] P.A. Igoshev, M.A. Timirgazin, A.K. Arzhnikov, and V.Y. Irkhin, JETP Lett. **98**, 150 (2013)
- [48] G. Seibold, E. Sigmund, and V. Hizhnyakov, Phys. Rev. B **57**, 6937 (1998)
- [49] M. Raczkowski, R. Frésard, and A.M. Oleś, Phys. Rev. B **73**, 174525 (2006)
- [50] M. Raczkowski, R. Frésard, and A.M. Oleś, Europhys. Lett. **76**, 128 (2006)
- [51] A. Isidori and M. Capone, Phys. Rev. B **80**, 115120 (2009)
- [52] A. Ruegg, S. Pilgram, and M. Sigrist, Phys. Rev. B **75**, 195117 (2007)
- [53] P. Korbel, W. Wojcik, A. Klejnberg, J. Spałek, M. Acquarone, and M. Lavagna, Eur. Phys. J. B, **32**, 315 (2003)
- [54] G. Kotliar, E. Lange, and M.J. Rozenberg, Phys. Rev. Lett. **84**, 5180 (2000)
- [55] G. Seibold, Phys. Rev. B **58**, 15520 (1998)

- [56] N. Pavlenko and T. Kopp, Phys. Rev. Lett. **97**, 187001 (2006)
- [57] J. Lorenzana and G. Seibold, Low Temp. Phys. **32**, 320 (2006)
- [58] W.F. Brinkman and T.M. Rice, Phys. Rev. B. **2**, 4302 (1970)
- [59] D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984);
D. Vollhardt, P. Wölfle, and P.W. Anderson, Phys. Rev. B. **35**, 6703 (1987)
- [60] M.J. Rozenberg *et al*, Phys. Rev. Lett. **75**, 105 (1995)
- [61] R. Frésard and K. Doll, Proceedings of the NATO ARW *The Hubbard Model: Its Physics and Mathematical Physics*, D. Baeriswyl, D.K. Campbell, J.M.P. Carmelo, F. Guinea, and E. Louis (eds.), San Sebastian, 1993 (Plenum Press, 1995), p. 385
- [62] C. Noce and M. Cuoco, Phys. Rev. B **59**, 2659 (1999)
- [63] R. Frésard and M. Lamboley, J. Low Temp. Phys. **126**, 1091 (2002)
- [64] D. Vollhardt, N. Blümer, K. Held, M. Kollar, J. Schlimpf, and M. Ulmke, Z. Phys. B **103**, 283 (1997)
- [65] H.C. Nguyen and J.B. Goodenough, Phys. Rev. B **52**, 8776 (1995)
- [66] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)
- [67] G. Cao, S. McCall, V. Dobrosavljevic, C.S. Alexander, J.E. Crow, and R.P. Guertin, Phys. Rev. B **61**, R5053 (2000)
- [68] C. Schuster, U. Lüders, R. Frésard, and U. Schwingenschlögl, EPL **103**, 37003 (2012)
- [69] S. Doniach, Physica B&C **91**, 231 (1977)
- [70] H. v. Löhneysen, A. Rosch, M. Vojta and P. Wölfle, Rev. Mod. Phys. **79**, 1015 (2007)
- [71] B. Möller and P. Wölfle, Phys. Rev. B **48**, 10320 (1993)
- [72] M. Guerrero and R.M. Noack, Phys. Rev. B **53**, 3707 (1996)
- [73] T.C. Li, Y.S. Sun and P. Wölfle, Z. Phys. B – Condensed Matter **82**, 369 (1991)
- [74] T.C. Li and P. Bénard, Phys. Rev. B. **50**, 17 387 (1994)
- [75] P. Limelette, H. Muguerra, and S. Hébert, Phys. Rev. B **82**, 035123 (2010)
- [76] J.P. Lu, Phys. Rev. B **49**, 5687 (1994)
- [77] K. Doll, M. Dzierzawa, R. Frésard, and P. Wölfle, Z. Phys. B – Condensed Matter **90**, 297 (1993)

- [78] J. Hubbard, Proc. R. Soc. London A **276**, 238 (1963)
- [79] J. Kanamori, Prog. Theor. Phys. **30**, 275 (1963)
- [80] R. Frésard and W. Zimmermann, Phys. Rev. B **58**, 15288 (1998)
- [81] M. Dzierzawa, unpublished
- [82] R. Micnas, J. Ranninger, and S. Robaszkiewicz, Rev. Mod. Phys. B **62**, 113 (1990)
- [83] G. Lhoutellier, R. Frésard, and A.M. Oleś, Phys. Rev. B **91**, 224410 (2015)
- [84] In the specific case of J_{ij} restricted to nearest neighbor on the cubic lattice one has $J_0 = 6J$.

10 The Lanczos Method

Erik Koch

Computational Materials Science

German Research School for Simulation Sciences

Jülich

Contents

1	Lanczos Method	3
1.1	Krylov space	5
1.2	Spectral functions	8
2	Application to the Hubbard model	11
2.1	Representation of basis and Hamiltonian	13
2.2	Green functions	15
2.3	Parallelization strategies	15
3	Application to DMFT	18
3.1	Cluster methods	20
3.2	Anderson impurity model	22
3.3	Hybridization sum rules	22
3.4	Symmetries	26
4	Conclusions	29

The Lanczos iteration [1] was conceived as a method for tridiagonalizing Hermitian matrices. Like the related Arnoldi method [2] for non-Hermitian matrices, it initially received widespread attention. Iterative approaches were, however, soon eclipsed by direct methods (Householder transformations and QR factorization) that are better suited for solving the eigenvalue problem for general matrices. Actually, the Lanczos method is particularly efficient for the determination of extreme eigenvalues and -vectors. Therefore, it was rediscovered in the 1970s [3], when computers had become sufficiently powerful to treat matrices large enough for the Lanczos algorithm to outperform general methods, nicely illustrating the Fundamental Law of Computer Science: the faster the computer, the greater the importance of the speed of algorithms [4]. By now iterative methods are an integral part of the numerical linear algebra curriculum [4–6].

For finding eigenvalues of a matrix H of dimension N , the Lanczos method requires the evaluation of matrix-vector products $H \cdot v$ as the only problem-specific step. This matrix-vector product can be calculated particularly efficiently when the matrix H is sparse, i.e., when the number of non-zero matrix elements per row does not scale with the matrix dimension. Storing such a matrix takes only $\mathcal{O}(N)$ memory and $H \cdot v$ can be evaluated in $\mathcal{O}(N)$ time. Calculating the extremal eigenvalues requires $\mathcal{O}(1)$ iterations, i.e., overall $\mathcal{O}(N)$ time. For comparison, direct diagonalization takes $\mathcal{O}(N^2)$ for storing the matrix and $\mathcal{O}(N^3)$ time to diagonalize. Besides their favorable scaling for sparse matrix problems, iterative methods have the advantage that they systematically approach the desired result. Typically the iteration converges geometrically and can be stopped as soon as the desired accuracy is reached. In contrast, direct methods appear to make no progress towards the solution until all $\mathcal{O}(N^3)$ operations are completed and the full result is obtained.

Since the Lanczos method is particularly suited for dealing with large sparse Hamiltonians, it is the method of choice for systems with short-range interactions. For band-structure calculations in a linear combination of atomic orbitals (LCAO) or tight-binding (TB) basis, it is known as the recursion method [7]. The basic idea here is to switch from the Bloch picture of a perfectly periodic solid to a local picture, replacing the solution of the Schrödinger equation in terms of Bloch waves by the calculation of the local density of states. The crucial technical point is to calculate the density of states not via a spectral representation (in terms of Bloch waves), but by repeated application of the Hamiltonian H to a localized single-electron state. With each application of H the electron explores more and more sites. Thus, if the hopping matrix elements beyond a certain distance are zero, such calculations can be performed without having to restrict the system to finite size.

For many-body models like quantum-spin- or Hubbard-models [8] this is unfortunately not possible. They have to be defined on a finite cluster, giving rise to a finite-dimensional Hamiltonian matrix. Since the size of the Hilbert space grows exponentially with system-size, actual calculations are restricted by the available computer memory. In a typical simulation, first the ground-state is calculated by a Lanczos iteration. Building on this, spectral functions are calculated in a similar way as in the recursion method. The great advantage of this approach is that it gives the dynamical properties of the ground state ($T=0$) directly on the real axis. The price is the restriction to (small) finite-size systems.

1 Lanczos Method

We can find the ground-state $|\Psi_0\rangle$ and its energy E_0 for a Hamiltonian H from the variational principle. The wave-function functional

$$E[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (1)$$

is minimized for $\Psi = \Psi_0$, with $E[\Psi_0] = E_0$. The functional gradient

$$\frac{\delta E[\Psi]}{\delta \langle \Psi |} = \frac{H|\Psi\rangle - E[\Psi]|\Psi\rangle}{\langle \Psi | \Psi \rangle} = |\Psi_a\rangle \quad (2)$$

gives the direction of steepest ascent of the functional from the point $|\Psi\rangle$. Moving in the opposite direction will thus result in a wave function with lower energy expectation value: $E[\Psi - \alpha\Psi_a] < E[\Psi]$ for small, positive α .

To find the optimum value of α , we minimize $E[\Psi - \alpha\Psi_a]$. For this, it is convenient to introduce an orthogonal basis in the space spanned by the two vectors $|\Psi\rangle$ and $|\Psi_a\rangle$. From (2) we see that $\text{span}(|\Psi\rangle, |\Psi_a\rangle) = \text{span}(|\Psi\rangle, H|\Psi\rangle)$. As first basis vector, we normalize $|\Psi\rangle$

$$|v_0\rangle = |\Psi\rangle / \sqrt{\langle \Psi | \Psi \rangle},$$

for the second vector we orthogonalize $H|v_0\rangle$ to $|v_0\rangle$

$$|\tilde{v}_1\rangle = H|v_0\rangle - |v_0\rangle\langle v_0|H|v_0\rangle \quad (3)$$

and normalize to obtain $|v_1\rangle$. With $a_n = \langle v_n | H | v_n \rangle$ and $b_1^2 = \langle \tilde{v}_1 | \tilde{v}_1 \rangle$ we thus have

$$H|v_0\rangle = b_1|v_1\rangle + a_0|v_0\rangle \quad (4)$$

from which we see that $\langle v_1 | H | v_0 \rangle = b_1$.

We can then write any normalized wave function in $\text{span}(|\Psi\rangle, H|\Psi\rangle) = \text{span}(|v_0\rangle, |v_1\rangle)$ as

$$|v\rangle = \cos(\theta)|v_0\rangle + \sin(\theta)|v_1\rangle. \quad (5)$$

Minimizing the expectation value

$$\langle v | H | v \rangle = a_0 \cos^2(\theta) + 2b_1 \sin(\theta) \cos(\theta) + a_1 \sin^2(\theta), \quad (6)$$

with respect to θ , we obtain, dividing by $\cos^2(\theta)$, the quadratic equation

$$b_1 \tan^2(\theta) + (a_0 - a_1) \tan(\theta) - b_1 = 0. \quad (7)$$

Solving for θ , we find the lowest-energy state on the subspace spanned by $|v_0\rangle$ and $H|v_0\rangle$. Alternatively, we can diagonalize the Hamiltonian matrix on the two-dimensional subspace, which in the basis $|v_0\rangle, |v_1\rangle$ is given by

$$H_{\text{span}(|\Psi\rangle, H|\Psi\rangle)} = \begin{pmatrix} a_0 & b_1 \\ b_1 & a_1 \end{pmatrix}. \quad (8)$$

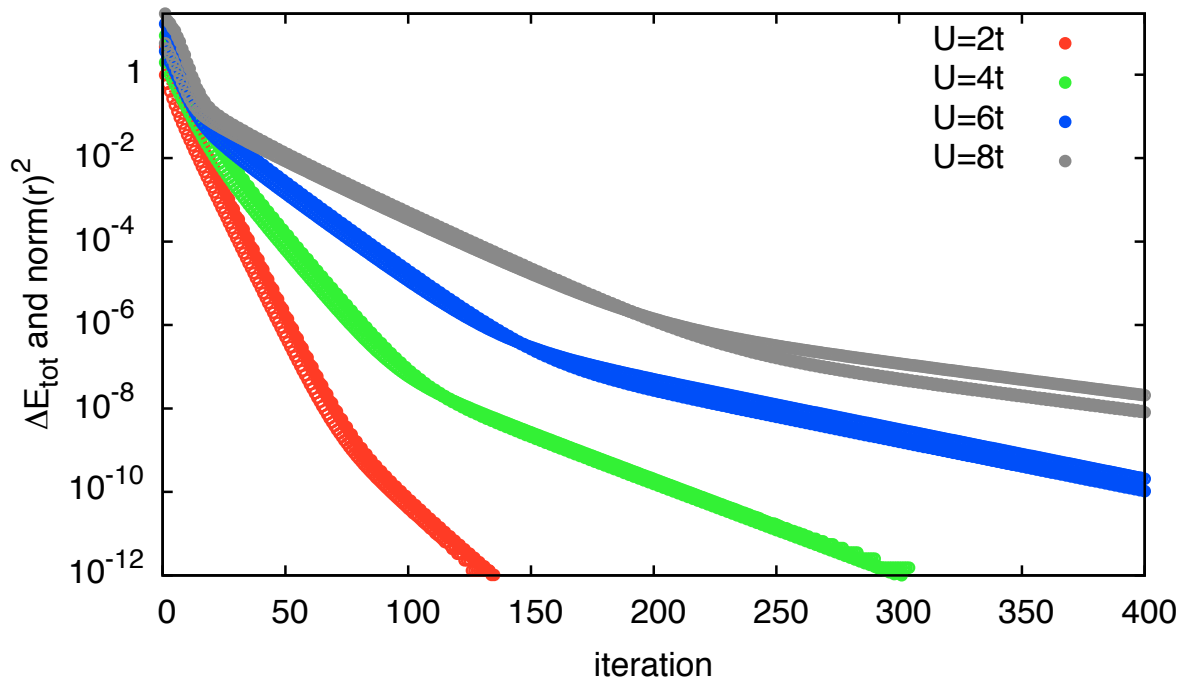


Fig. 1: Convergence of the residual (filled circles) and the lowest eigenvalue (open circles) for a steepest-descent minimization of a Hubbard-chain of 10 sites at half filling, starting from a random initial vector.

Naturally, we can use the variational state of lowest energy

$$|\Psi^{(2)}\rangle = \cos(\theta_{\min})|v_0\rangle + \sin(\theta_{\min})|v_1\rangle \quad (9)$$

as the starting point for another steepest-descent minimization. Doing this repeatedly, we obtain a series of vectors with decreasing energy expectation value, which rapidly converges to a minimum. For a generic functional, this would usually be a local, not the global minimum, which makes the optimization of high-dimensional functions a hard problem. The energy functional (1), however, only has minima for the ground states; all other stationary points are saddle points. We can thus expect rapid convergence to the ground state, examples of which are given in Figure 1, except in the unlikely case where the gradient (2) vanishes, i.e., if $|\Psi\rangle$ happens to be an eigenfunction of H .

For checking convergence of this steepest-descent method, introduced by Kantorovich [9] and, independently, by Hestenes and Karush [10], we can monitor the change in the energy expectation value or determine when the residual

$$r[\Psi] = \| (H - E[\Psi])|\Psi\rangle \|^2 = \langle \Psi | H^2 | \Psi \rangle - E[\Psi]^2, \quad (10)$$

which measures the quality of the eigenstate, becomes sufficiently small.

1.1 Krylov space

If we apply the method of steepest descent L times, starting from a vector $|v_0\rangle$, the resulting vector will lie in $\mathcal{K}^L(|v_0\rangle) = \text{span}(|v_0\rangle, H|v_0\rangle, H^2|v_0\rangle, \dots, H^L|v_0\rangle)$, the $L + 1$ -dimensional *Krylov space* [11] of H over $|v_0\rangle$. Instead of repeatedly minimizing the energy in two-dimensional subspaces, we could directly find the state of lowest energy in $\mathcal{K}^L(|v_0\rangle)$. Having more degrees of freedom for the minimization will lead to even faster convergence.

To implement this idea, we construct an orthonormal basis $|v_n\rangle$ of the Krylov space. We start with the normalized vector $|v_0\rangle$. The second basis vector $|v_1\rangle$ is constructed as in the steepest-descent method (3):

$$b_1|v_1\rangle = |\tilde{v}_1\rangle = H|v_0\rangle - a_0|v_0\rangle. \quad (11)$$

The next basis vector is likewise constructed as $H|v_n\rangle$ orthogonalized to all previous vectors, and normalized

$$b_2|v_2\rangle = |\tilde{v}_2\rangle = H|v_1\rangle - \sum_{i=0}^1 |v_i\rangle\langle v_i|H|v_1\rangle = H|v_1\rangle - a_1|v_1\rangle - b_1|v_0\rangle. \quad (12)$$

where $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. The fourth basis vector is

$$b_3|v_3\rangle = |\tilde{v}_3\rangle = H|v_2\rangle - \sum_{i=0}^2 |v_i\rangle\langle v_i|H|v_2\rangle = H|v_2\rangle - a_2|v_2\rangle - b_2|v_1\rangle. \quad (13)$$

Here the last term in the orthogonalization vanishes when H is Hermitian: (11) together with the orthogonality of the basis vectors for $n = 0 \dots 2$ implies $\langle v_2|H|v_0\rangle = 0$. When H is Hermitian it follows that $\langle v_0|H|v_2\rangle = 0$.

The construction of the further basis vectors follows the same scheme

$$b_{n+1}|v_{n+1}\rangle = |\tilde{v}_{n+1}\rangle = H|v_n\rangle - \sum_{i=0}^n |v_i\rangle\langle v_i|H|v_n\rangle = H|v_n\rangle - a_n|v_n\rangle - b_n|v_{n-1}\rangle$$

with $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. Rearranging shows that H is tridiagonalized

$$H|v_n\rangle = b_n|v_{n-1}\rangle + a_n|v_n\rangle + b_{n+1}|v_{n+1}\rangle$$

which in turn implies that $H|v_i\rangle$ is orthogonal to all basis states except $|v_i\rangle$ and $|v_{i\pm 1}\rangle$. This tridiagonalization of H is the essence of the *Lanczos method* [1].

After L steps the Hamiltonian on the $L + 1$ -dimensional Krylov space is given by

$$H_{\mathcal{K}^L(|v_0\rangle)} = \begin{pmatrix} a_0 & b_1 & 0 & 0 & \dots & 0 & 0 \\ b_1 & a_1 & b_2 & 0 & \dots & 0 & 0 \\ 0 & b_2 & a_2 & b_3 & & 0 & 0 \\ 0 & 0 & b_3 & a_3 & & 0 & 0 \\ & \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & & a_{L-1} & b_L \\ 0 & 0 & 0 & 0 & \dots & b_L & a_L \end{pmatrix} \quad (14)$$

<code>v=init</code>	
<code>b0=norm2(v)</code>	not part of tridiagonal matrix
<code>scal(1/b0,v)</code>	$v = v_0\rangle$
<code>w=0</code>	
<code>w=w+H*v</code>	$w = H v_0\rangle$
<code>a[0]=dot(v,w)</code>	
<code>axpy(-a[0],v,w)</code>	$w = \tilde{v}_1\rangle = H v_0\rangle - a_0 v_0\rangle$
<code>b[1]=norm2(w)</code>	
<code>for n=1,2,...</code>	
<code>if abs(b[n])<eps then exit</code>	invariant subspace
<code>scal(1/b[n],w)</code>	$w = v_n\rangle$
<code>scal(-b[n],v)</code>	$v = -b_n v_{n-1}\rangle$
<code>swap(v,w)</code>	
<code>w=w+H*v</code>	$w = H v_n\rangle - b_n v_{n-1}\rangle$
<code>a[n]=dot(v,w)</code>	$a[n] = \langle v_n H v_n\rangle - b_n\langle v_n v_{n-1}\rangle$
<code>axpy(-a[n],v,w)</code>	$w = \tilde{v}_{n+1}\rangle$
<code>b[n+1]=norm2(w)</code>	
<code>diag(a[0]..a[n], b[1]..b[n])</code>	getting a_{n+1} needs another $H v\rangle$
<code>if converged then exit</code>	
<code>end</code>	

Table 1: *The implementation of the Lanczos iteration requires only two N -dimensional vectors for tridiagonalizing H and thus for calculating the ground-state energy. Constructing the Lanczos-approximation of the ground-state vector requires a second iteration and one additional N -dimensional vector. The most expensive operation by far is the matrix-vector product.*

If we do not normalize the basis vectors, we obtain an iteration of the form

$$|\Phi_{n+1}\rangle = H|\Phi_n\rangle - \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle}|\Phi_n\rangle - \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle}|\Phi_{n-1}\rangle \quad (15)$$

where $|\Phi_n\rangle = \prod_{i=1}^n b_i |v_n\rangle$ in terms of which we have

$$a_n = \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle}, \quad b_n^2 = \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle}. \quad (16)$$

In this unnormalized basis the Hamiltonian appears non-Hermitian

$$H|\Phi_n\rangle = b_n^2|\Phi_{n-1}\rangle + a_n|\Phi_n\rangle + |\Phi_{n+1}\rangle, \quad (17)$$

but it actually is:

$$\langle\Phi_{n+1}|H|\Phi_n\rangle = \langle\Phi_{n+1}|\Phi_{n+1}\rangle = b_{n+1}^2\langle\Phi_n|\Phi_n\rangle = \langle\Phi_n|H|\Phi_{n+1}\rangle. \quad (18)$$

The numerical implementation only requires keeping two N -dimensional vectors in memory. It is shown in Table 1.

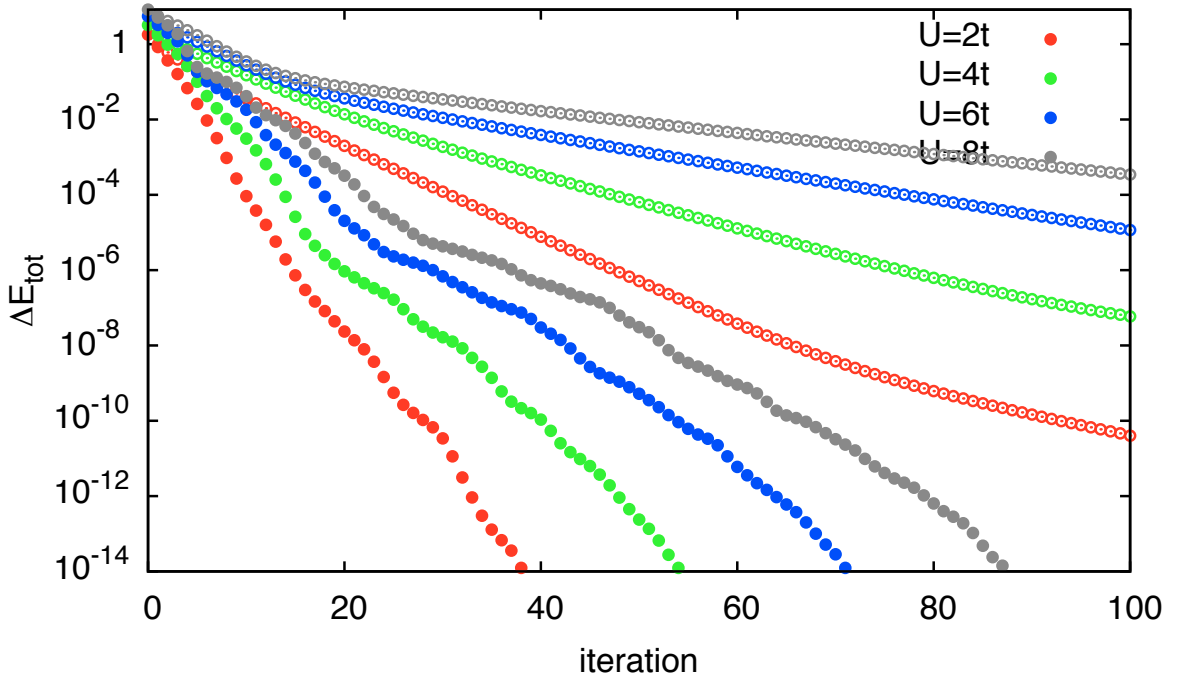


Fig. 2: Convergence of the lowest eigenvalue for a Lanczos iteration (full circles) compared to steepest-descent minimization (open circles) of a 10-site Hubbard chain at half-filling, starting from a random initial vector. Due to the additional variational degrees of freedom, Lanczos converges significantly faster. Overall, convergence for the half-filled system gets harder for larger U , as the distance to the lowest excited states is reduced ($\sim t^2/U$) and the spectrum widens ($\sim U$). In all cases, convergence is reached after less than $L \approx 100$ Lanczos iterations, to be compared to the dimension $N=63\,504$ of the Hilbert space.

Diagonalizing (14), after fewer than one hundred iterations, the lowest eigenvalue of the tridiagonal representation of H on the Krylov space gives an excellent approximation to the ground-state energy of H in the full Hilbert space (Fig. 2). A formal estimate of the convergence was given by Kaniel and Paige [5]. For an $N+1$ -dimensional, symmetric matrix H with eigenvalues E_n , the lowest eigenvalue $\check{E}_0$ of the tridiagonal representation of H on the $(L+1)$ -dimensional Krylov space over $|v_0\rangle$ fulfills

$$\frac{\check{E}_0 - E_0}{E_N - E_0} \leq \left(\frac{\tan(\arccos(\langle \check{\Psi}_0 | \Psi_0 \rangle))}{T_L \left(1 + 2 \frac{E_1 - E_0}{E_N - E_1} \right)} \right)^2 \quad (19)$$

where $T_L(x)$ is the Chebyshev polynomial of order L and $\langle \check{\Psi}_0 | \Psi_0 \rangle$ the overlap of the Lanczos approximation to the ground-state $\check{\Psi}_0$ with the ground-state of H . Thus, if the initial state $|v_0\rangle$ is not orthogonal to the non-degenerate ground-state, convergence is exponential with a rate roughly increasing with the square root of the gap to the first excited state measured in units of the width of the spectrum.

The approximate ground-state vector is given by the linear combination

$$|\check{\Psi}_0\rangle = \sum_{n=0}^L u_{0n} |v_n\rangle, \quad (20)$$

where u_0 is the ground-state vector of the $L + 1$ -dimensional tridiagonal matrix (14). Instead of storing all $L + 1$ basis vectors $|v_n\rangle$, we can restart the Lanczos iteration from the same $|v_0\rangle$, accumulating the sum (20) iteration by iteration. This only requires keeping one additional N -dimensional vector in memory.

So far we have tacitly assumed that the Krylov vectors $H^n|v_0\rangle$ are linearly independent. If not, there will be a vector $H|\tilde{v}_m\rangle$ that vanishes when orthogonalized to the previous states, i.e., $b_n = 0$. This means that the Krylov space $\text{span}(|v_0\rangle, |v_1\rangle, \dots, |v_m\rangle)$ is invariant under H , i.e., we have found an exact eigenspace of H . For a large matrix H it is quite unlikely to be that lucky. Still, as the Lanczos iteration approaches the ground-state, we encounter a similar situation: Close to an eigenstate, the functional (1) becomes almost stationary, i.e., the coefficients b_n almost vanish. Normalization of the very short vector $|\tilde{v}_n\rangle$ then amplifies numerical noise in the small vector. This makes the numerical $|v_n\rangle$, which in theory should automatically be orthogonal to all $|v_m\rangle$ with $m < n - 2$, actually have finite overlaps with these vectors. This loss of orthogonality manifests itself in the appearance of multiple copies of eigenvectors (ghost states) that are unrelated to the actual multiplicities of the eigenvalues. This is the problem that makes the Lanczos method impractical for tridiagonalizing dense matrices. For the ground state, the variational principle prevents severe problems from the loss of orthogonality. An example of the appearance of ghost states is shown in Figure 3.

If we want to reliably obtain excited states, we need to explicitly orthogonalize to the previous basis states. This leads to the Lanczos method with (complete) reorthogonalization [5]. A similar orthogonalization is performed in the Arnoldi method [2], which, however, is devised for non-symmetric matrices.

1.2 Spectral functions

Given the orthogonality problems of the Lanczos method, it appears hopeless to use it to obtain the matrix elements of the resolvent, as they contain information about the full spectrum $H|\Psi_n\rangle = E_n|\Psi_n\rangle$. Still, we are tempted to approximate the Lehmann representation

$$G_c(z) = \left\langle \Psi_c \left| \frac{1}{z - H} \right| \Psi_c \right\rangle = \sum_{n=0}^N \frac{\langle \Psi_c | \Psi_n \rangle \langle \Psi_n | \Psi_c \rangle}{z - E_n} \quad (21)$$

in terms of the eigenstates on the Krylov space $\mathcal{K}^L(|\Psi_c\rangle)$

$$\check{G}_c(z) = \left\langle \Psi_c \left| \frac{1}{z - \check{H}_c} \right| \Psi_c \right\rangle = \sum_{n=0}^L \frac{\langle \Psi_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \Psi_c \rangle}{z - \check{E}_n}. \quad (22)$$

This is straightforward to calculate: We run L Lanczos iterations, starting from the (normalized) vector $|\Psi_c\rangle$, to create the tridiagonal $\check{H}_c$. The matrix element of the resolvent is the top left

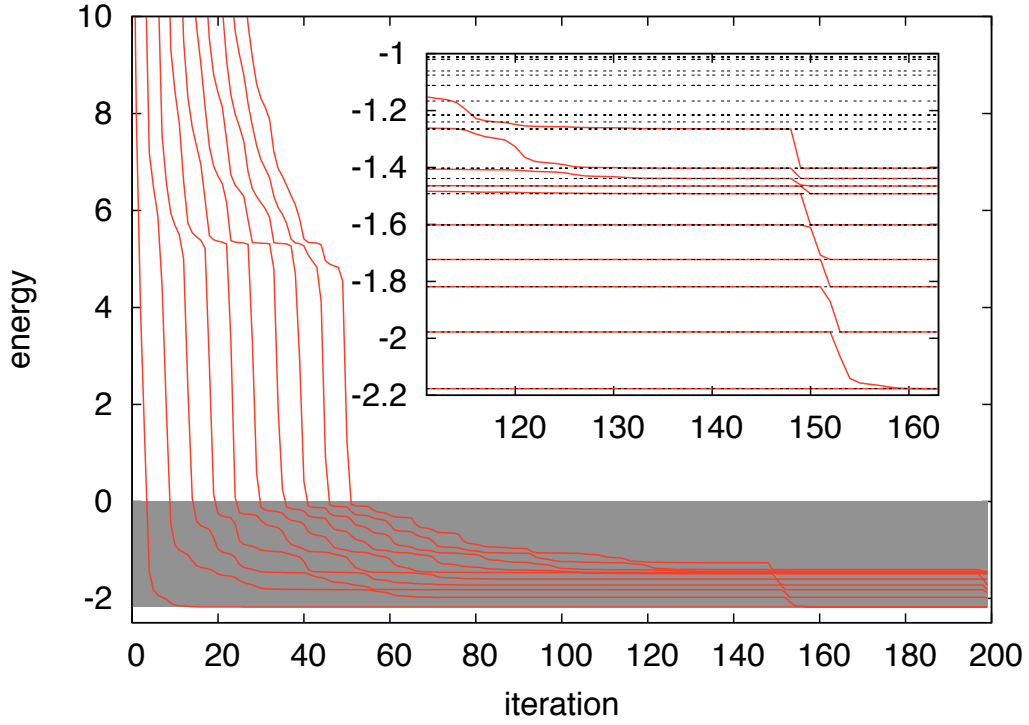


Fig. 3: Appearance of ghost states in an overconverged Lanczos iteration. The ground-state energy for a half-filled 8-site Hubbard chain with $U = 10t$ is converged to numerical accuracy (10^{-16}) after about 85 iterations. Forcing the Lanczos iteration to continue, we see that at first the higher excited states also converge to the exact eigenvalues (dashed lines). But, as shown in the inset, they eventually start collapsing to the ground state. The appearance of these ghost states is due to the orthogonality problem introduced by small normalization parameters b_n , when the iteration is very close to a stationary point.

matrix element of the inverse of

$$z - \check{H}_c = \left(\begin{array}{c|cccccc} z - a_0 & -b_1 & 0 & 0 & \cdots & 0 & 0 \\ \hline -b_1 & z - a_1 & -b_2 & 0 & \cdots & 0 & 0 \\ 0 & -b_2 & z - a_2 & -b_3 & \cdots & 0 & 0 \\ 0 & 0 & -b_3 & z - a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & z - a_{L-1} & -b_L \\ 0 & 0 & 0 & 0 & \cdots & -b_L & z - a_L \end{array} \right). \quad (23)$$

This is easily determined by partitioning the matrix as indicated

$$z - \check{H}_c = \begin{pmatrix} z - a_0 & B^{(1)T} \\ B^{(1)} & z - \check{H}_c^{(1)} \end{pmatrix} \quad (24)$$

and inverting the block matrix, giving

$$[(z - \check{H}_c)^{-1}]_{00} = \left(z - a_0 - B^{(1)T} (z - \check{H}_c^{(1)})^{-1} B^{(1)} \right)^{-1} = (z - a_0 - b_1^2 [(z - \check{H}_c^{(1)})^{-1}]_{00})^{-1}.$$

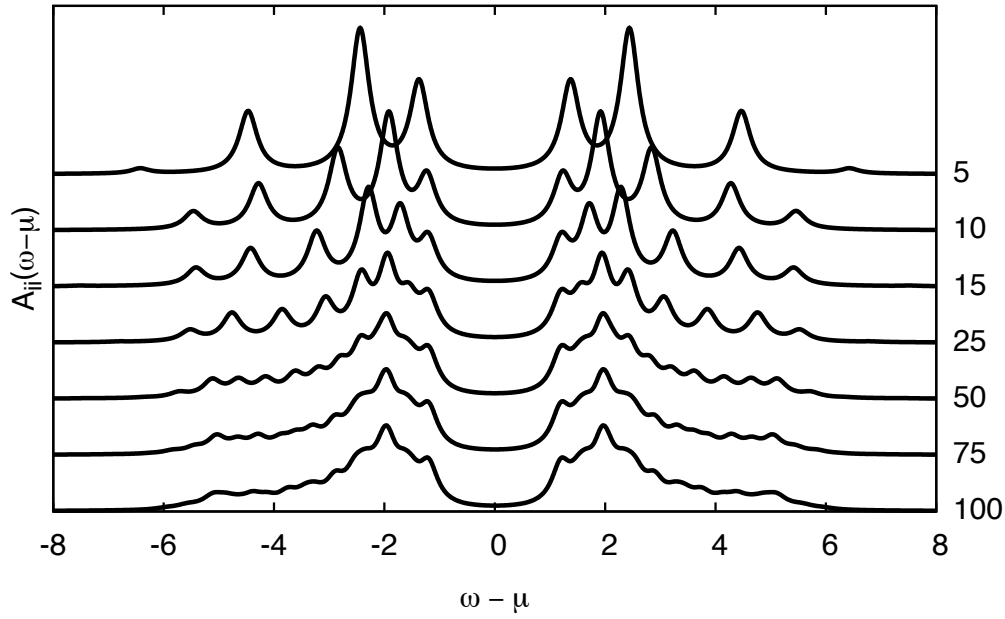


Fig. 4: Convergence of the spectral function with increasing number of Lanczos steps, $L=5, 10, 15, 25, 50, 75$, and 100 , for a 14-site Hubbard chain with $U = 5t$ at half filling. With increasing L , more and more moments of the photoemission and inverse photoemission part of the spectrum are reproduced correctly.

Repeating inversion by partitioning for the submatrices $\check{H}^{(n)}$ we obtain the continued fraction

$$\check{G}_c(z) = [(z - \check{H}_c)^{-1}]_{00} = \frac{1}{z - a_0 - \frac{b_1^2}{z - a_1 - \frac{b_2^2}{z - a_2 - \dots}}}, \quad (25)$$

which terminates with $-b_L^2/(z - a_L)$. The spectral representation (22) is obtained by diagonalizing the Lanczos matrix $\check{H}_c$ giving us the $L + 1$ eigenvalues $\check{E}_n$ and eigenvectors u_n . Since

$$|\check{\Psi}_n\rangle = \sum_{l=0}^L u_{nl}|v_l\rangle \quad (26)$$

the matrix elements are given by $\langle \check{\Psi}_n | \Psi_c \rangle = u_{n0}$. Thus

$$\check{G}_c(z) = \sum_{n=0}^L \frac{|u_{n0}|^2}{z - \check{E}_n} \quad (27)$$

The spectral function

$$\check{A}(\omega \pm i\eta) = \mp \frac{1}{\pi} \text{Im } \check{G}(\omega \pm i\eta) \quad (28)$$

obtained this way converges very quickly. An example is shown in Figure 4.

To understand how the $L + 1$ eigenstates of $\check{H}$ can represent the full spectrum so well, we consider the moments of the spectral function

$$\int_{-\infty}^{\infty} d\omega \omega^m \check{A}(\omega) = \sum_{n=0}^L |u_{n0}|^2 \check{E}_n^m = \sum_{n=0}^L \langle \Psi_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \Psi_c \rangle \check{E}_n^m = \langle \Psi_c | \check{H}^m | \Psi_c \rangle \quad (29)$$

Since $\check{H}$ is the projection of H onto the Krylov space $\mathcal{K}^L(|\Psi_c\rangle)$, we have $\check{H}^m |\Psi_c\rangle = H^m |\Psi_c\rangle$ for $m \leq L$. Thus the Lanczos representation $\check{A}(z)$ correctly reproduces the first $2L + 1$ moments of the spectral function $A(z)$. A further Lanczos step adds one new level to the continued fraction (25), leaving all previous terms unchanged. $b_m^2 = 0$ then implies that the continued fraction terminates, and all moments are given correctly. A near vanishing $b_m^2 \approx 0$, which gives rise to the loss of orthogonality of the Lanczos vectors, for the spectral function merely means that further terms in the continued fraction hardly contribute any more.

So far we have only considered diagonal elements of the resolvent. Off-diagonal matrix elements

$$G_{c_1, c_2}(z) = \left\langle \Psi_{c_2} \left| \frac{1}{z - H} \right| \Psi_{c_1} \right\rangle \quad (30)$$

are easily obtained by considering the diagonal elements for the linear combinations

$$\left\langle \Psi_{c_1} \pm \Psi_{c_2} \left| \frac{1}{z - H} \right| \Psi_{c_1} \pm \Psi_{c_2} \right\rangle = G_{c_1, c_1}(z) \pm G_{c_1, c_2}(z) \pm G_{c_2, c_1}(z) + G_{c_2, c_2}(z). \quad (31)$$

2 Application to the Hubbard model

The Hubbard model

$$H = -t \sum_{\langle i, j \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad (32)$$

describes the fundamental dichotomy between itinerancy and locality for correlated electrons on a lattice: The hopping tends to delocalize electrons and is diagonal in k -space. This makes it possible to solve the band-structure problem for the infinite solid. In k -space the single electron Hamiltonian is block-diagonal. For the one-band Hubbard model each block is just the band energy ε_k . In general, each block defines the band structure problem for one k -point. Including electron-electron repulsion destroys this symmetry. The two-body Coulomb term is diagonal in real space, while in k -space it is dense

$$H = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \frac{U}{M} \sum_{k, k', q} c_{k\uparrow}^\dagger c_{k-q, \uparrow} c_{k'\downarrow}^\dagger c_{k'+q, \downarrow}. \quad (33)$$

This has two important consequences:

1. Since we know no general approach to transform the full Hamiltonian into finite-dimensional blocks, we have to restrict ourselves to finite-dimensional systems. For a cluster

M	$N_{\uparrow}$	$N_{\downarrow}$	dim of Hilbert space	memory
2	1	1	4	
4	2	2	36	
6	3	3	400	
8	4	4	4 900	
10	5	5	63 504	
12	6	6	853 776	6 MB
14	7	7	11 778 624	89 MB
16	8	8	165 636 900	1263 MB
18	9	9	2 363 904 400	17 GB
20	10	10	34 134 779 536	254 GB
22	11	11	497 634 306 624	3708 GB
24	12	12	7 312 459 672 336	53 TB
20	1	1	400	
20	2	2	36 100	
20	3	3	1 299 600	9 MB
20	4	4	23 474 025	179 MB
20	5	5	240 374 016	1833 MB
20	6	6	1 502 337 600	11 GB
20	7	7	6 009 350 400	44 GB
20	8	8	15 868 440 900	118 GB
20	9	9	28 210 561 600	210 GB
20	10	10	34 134 779 536	254 GB

Table 2: Dimension of Hilbert space $\dim(\mathcal{H})$ and computer memory required for storing a single many-body wave function for Hubbard models with M orbitals and $N_{\uparrow} + N_{\downarrow}$ electrons. The first group of numbers gives the dimensions for half filling, where the Hilbert space is largest. The second group shows how the dimension grows with the filling (dimensions are symmetric about half filling). Note that the fourth column resembles a semi-logarithmic plot of $\dim(\mathcal{H})$ as a function of system size or filling.

of M sites with $N_{\uparrow}$ electrons with spin up and $N_{\downarrow}$ with spin down, the dimension of the Hilbert space is

$$\dim(\mathcal{H}) = \dim(\mathcal{H}_{\uparrow}) \times \dim(\mathcal{H}_{\downarrow}) = \binom{M}{N_{\uparrow}} \times \binom{M}{N_{\downarrow}}. \quad (34)$$

Examples for the single-band Hubbard model, illustrating the enormous growth of the Hilbert space are given in table 2. Actual calculations are therefore limited to quite small systems.

2. For a tight-binding system where hopping matrix-elements are restricted to close neighbors, the many-body Hamiltonian is a sparse matrix when expressed in a real-space basis of localized orbitals (32). The basis states are then configurations $|\{n_{i\sigma}\}\rangle = \prod (c_{i\sigma}^{\dagger})^{n_{i\sigma}} |0\rangle$, characterized by their occupation numbers $\{n_{i\sigma}\}$.

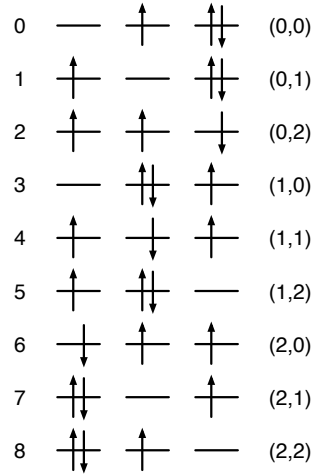


Fig. 5: Basis configurations for a three site system with two up and one down spin electron. The left label denotes the index of the configuration. Equivalently, a state is also unambiguously identified by the tuple of its up- and down-configuration index.

2.1 Representation of basis and Hamiltonian

Since the many-body basis states

$$|\{n_{i\sigma}\}\rangle = \prod_{i\sigma} \left(c_{i\sigma}^\dagger\right)^{n_{i\sigma}} |0\rangle \quad (35)$$

can be represented by fermionic occupation numbers, it is natural to encode them in a string of bits. For a Hamiltonian like (32) that conserves spin, we can write

$$|\{n_{i\sigma}\}\rangle = \prod_{i=0}^{L-1} \left(c_{i\downarrow}^\dagger\right)^{n_{i\downarrow}} \left(c_{i\uparrow}^\dagger\right)^{n_{i\uparrow}} |0\rangle \quad (36)$$

with $\sum n_{i\sigma} = N_\sigma$, and encode each spin-component as the integer $m_\sigma = \sum n_{i\sigma} 2^i$. Enumerating all basis states with N_σ electrons on L sites is then as simple as looping over all integers from 0 to $2^L - 1$ and storing each integer m_σ with N_σ bits set. For $N_\uparrow = 2$ and $N_\downarrow = 1$ electrons on $L = 3$ sites we obtain

$m_\uparrow$	bits	state	$i_\uparrow$	$m_\downarrow$	bits	state	$i_\downarrow$
0	000			0	000		
1	001			1	001	$c_{0\downarrow}^\dagger 0\rangle$	0
2	010			2	010	$c_{1\downarrow}^\dagger 0\rangle$	1
3	011	$c_{0\uparrow}^\dagger c_{1\uparrow}^\dagger 0\rangle$	0	3	011		
4	100			4	100	$c_{2\downarrow}^\dagger 0\rangle$	2
5	101	$c_{0\uparrow}^\dagger c_{2\uparrow}^\dagger 0\rangle$	1	5	101		
6	110	$c_{1\uparrow}^\dagger c_{2\uparrow}^\dagger 0\rangle$	2	6	110		
7	111			7	111		

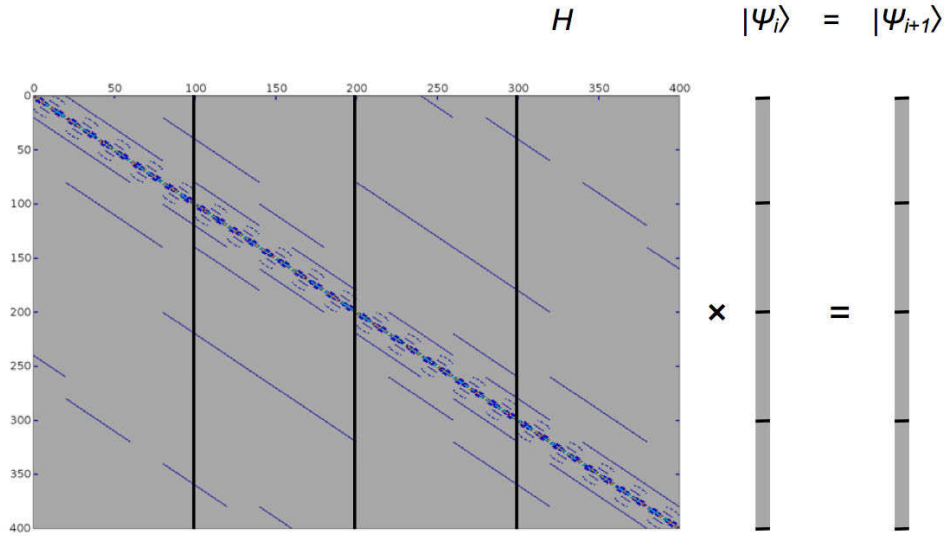


Fig. 6: Matrix-vector product for the Lanczos iteration showing (in blue) the non-zero elements of the Hamiltonian matrix for a 6-site Hubbard model at half-filling. Vector Ψ_i is only read, Ψ_{i+1} can be written sequentially: $\Psi_{i+1,n} = \sum_m H_{n,m} \Psi_{i,m}$. Access to elements of Ψ_i is highly non-local.

We number the basis states in the order they are found $i_\sigma = 0 \dots \dim(\mathcal{H}_\uparrow)$. A full basis state (36) is then indexed by the $i = i_\downarrow + \dim(\mathcal{H}_\downarrow) \cdot i_\uparrow$. This corresponds to writing the basis as the tensor product of the up and down states. Alternatively we could use $i = i_\uparrow + \dim(\mathcal{H}_\uparrow) \cdot i_\downarrow$. The corresponding configuration is given by the integers $m_\uparrow[i_\uparrow]$ and $m_\downarrow[i_\downarrow]$

For calculating the matrix elements, it is convenient to store the two lookup tables for converting between the integer encoding the basis state m_σ and its index i_σ in the basis. Since $\dim(\mathcal{H}_\sigma)$ is normally (i.e, close to half-filling) much smaller than $\dim(\mathcal{H})$, this does not use much memory.

The hopping term connects basis states that differ only in two occupation numbers of the same spin, e.g., $n_{i\sigma}$ and $n_{j\sigma}$. The matrix element is $\mp t_{ij}$, where the sign depends on $N_{i,j}$, the number of electrons of spin σ between site i and site j : $-(-1)^{N_{i,j}} t_{i,j}$. As an example we give the matrix of the hopping between the basis states for $N_\uparrow = 2$ electrons on a linear cluster with $L = 3$ sites, nearest neighbor hopping t and with periodic boundary conditions:

$$T_\uparrow = \begin{pmatrix} 0 & -t & +t \\ -t & 0 & -t \\ +t & -t & 0 \end{pmatrix}. \quad (37)$$

The full hopping matrix is then given by the tensor product of $T_\uparrow$ and $T_\downarrow$. The matrix above looks fairly dense; for larger system the T quickly becomes very sparse, as shown in Figure 6.

2.2 Green functions

In a basis of spin-orbitals α and β , the elements of the Green matrix are given by

$$\begin{aligned} G_{\alpha\beta}(\omega) &= \left\langle \Psi_0 \left| c_\alpha^\dagger \frac{1}{\omega + (H - E_0 - i\eta)} c_\beta \right| \Psi_0 \right\rangle + \left\langle \Psi_0 \left| c_\alpha \frac{1}{\omega - (H - E_0 - i\eta)} c_\beta^\dagger \right| \Psi_0 \right\rangle \\ &= \sum_n \frac{\langle \Psi_0 | c_\alpha^\dagger | \Psi_n^{(N-1)} \rangle \langle \Psi_n^{(N-1)} | c_\beta | \Psi_0 \rangle}{\omega + \left(E_n^{(N-1)} - E_0^{(N)} \right) - i\eta} + \sum_n \frac{\langle \Psi_0 | c_\alpha | \Psi_n^{(N+1)} \rangle \langle \Psi_n^{(N+1)} | c_\beta^\dagger | \Psi_0 \rangle}{\omega - \left(E_n^{(N+1)} - E_0^{(N)} \right) + i\eta}, \end{aligned} \quad (38)$$

where the sums are over the eigenstates of the Hilbert space with one electron less (first term) and one additional electron (second term). Diagonal elements are calculated in Lanczos as described in Section 1.2: To find $G_{\alpha\alpha}(\omega)$, we need the ground state vector $|\Psi_0\rangle$ and two additional Lanczos runs, giving the two terms in (38). For the first term, we start the Lanczos iteration from the normalized vector $|\Psi_c^<\rangle = c_\alpha |\Psi_0\rangle / \sqrt{n_\alpha}$, where $n_\alpha = \langle \Psi_0 | c_\alpha^\dagger c_\alpha | \Psi_0 \rangle$. Likewise, for the second term, we start from $|\Psi_c^>\rangle = c_\alpha^\dagger |\Psi_0\rangle / \sqrt{1 - n_\alpha}$. The Green function is then given, in terms of the Lanczos coefficients, by

$$\check{G}_{\alpha\alpha}(\omega) = \frac{n_\alpha}{\omega - E_0 - i\eta + a_0^< - \frac{b_1^{<2}}{\omega - E_0 - i\eta + a_1^< - \dots}} + \frac{1 - n_\alpha}{\omega + E_0 + i\eta - a_0^> - \frac{b_1^{>2}}{\omega + E_0 + i\eta - a_1^> - \dots}}. \quad (39)$$

If the ground state is degenerate, e.g., for $N_\uparrow \neq N_\downarrow$, where $E_0(N_\uparrow, N_\downarrow) = E_0(N_\downarrow, N_\uparrow)$, we average the Green functions calculated from the different ground states. This is the $T \rightarrow 0$ limit of the finite-temperature Green function

$$G_{\alpha\alpha}(\omega) = \frac{1}{Z} \sum_m e^{-\beta E_m^{(N)}} G_{\alpha\alpha}^{(m)}(\omega), \quad (40)$$

where $Z = \sum_n e^{-\beta E_n^{(N)}}$ is the partition function and $G_{\alpha\alpha}^{(m)}(\omega)$ has the same form as (38), only with Ψ_0 replaced by Ψ_m . For finite, but sufficiently low temperatures, the Boltzmann factor is negligibly small, except for the lowest few states. If we calculate those, taking care of the orthogonality problem (ghost states), we can easily obtain the finite-temperature Green function. A more elaborate method is given in [12].

Off-diagonal elements of the Green matrix are calculated from diagonal elements of linear combinations of spin-orbitals, e.g., $(c_\alpha^\dagger \pm c_\beta^\dagger) |\Psi_0\rangle$, as described in Section 1.2.

2.3 Parallelization strategies

Because of the enormous size of the many-body Hilbert space, see Table 2, Lanczos calculations are limited by the available memory. On shared-memory systems the most time consuming operation of the Lanczos iteration, the multiplication of the Hamiltonian matrix with a many-body vector, can be parallelized very easily when it is written such that the elements of the resulting vector are calculated independently: As illustrated in Figure 6, different threads can

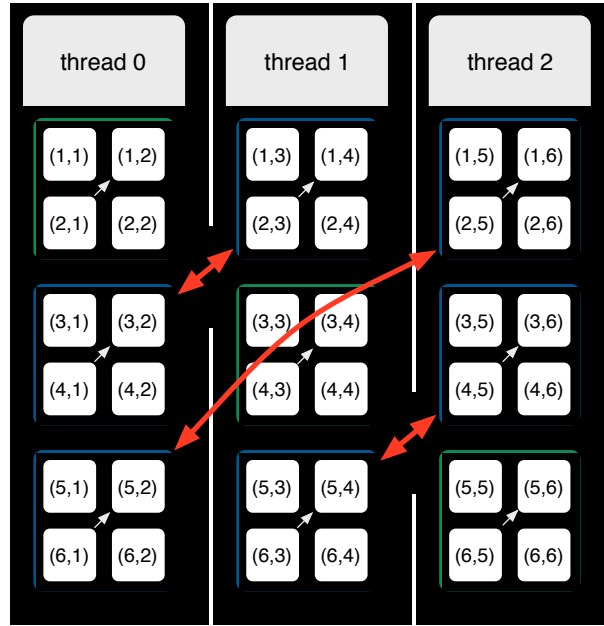


Fig. 7: *Transpose operation that makes memory access thread-local when calculating the operation of the Hamiltonian on the state-vector. The communication (red arrows) is realized by a call to `MPI_Alltoall`, which is very efficiently implemented on Blue Gene. The small grey arrows indicate the local operations needed to complete the matrix-transpose.*

work on different chunks of $|\Psi_{i+1}\rangle$. The off-diagonal elements of the kinetic energy part of (32) lead to non-local memory access, but the elements of $|\Psi_i\rangle$ as well as the matrix elements are only read, so that there is no need for locking. An OpenMP parallelization thus needs only a single pragma. Parallelizing the scalar products in a similar way, we obtain almost ideal speedup. Such an implementation is, however, limited to one node. To use significantly more memory than available on a single node we need to find an implementation that can efficiently use distributed memory.

A naive approach on distributed memory systems uses MPI2 one-sided communication to emulate the shared-memory approach by direct remote memory access. This leads, however, to a severe speed-down, i.e., the more processors we use, the longer we have to wait for the result.

An efficient distributed-memory implementation [14] is instead based on the fact that hopping does not change spin. Hopping of the up-electron mixes only different up-hopping configurations, while the down-electron configuration remains unchanged. If we group all up configurations for a fixed down configuration together in a single thread, this hopping can be carried out locally. Figure 5 illustrates this: for a fixed index $i_\downarrow$, all $i_\uparrow$ configurations are stored in adjacent memory locations and can be stored in a thread. We see, that this basis can be naturally indexed by a tuple $(i_\downarrow, i_\uparrow)$ (right labels in Figure 5) instead of the global index (left labels). We can therefore equivalently regard the vectors as matrices $v(i_\downarrow, i_\uparrow)$ with indices $i_\downarrow$ and $i_\uparrow$. Now it is easy to see that a matrix transpose reshuffles the data elements such that the down configurations are sequential in memory and local to the thread.

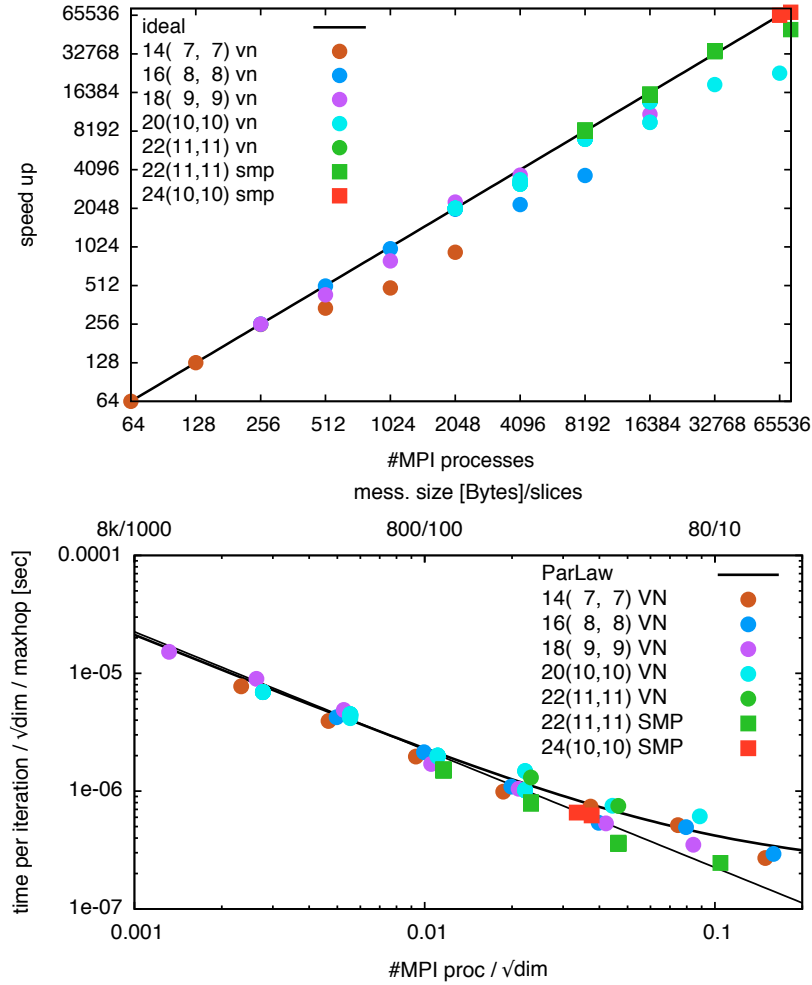


Fig. 8: Timings of the parallel implementation of the Lanczos algorithm for the Hubbard model on the Jülich IBM BlueGene. Sizes of the state vectors of the half-filled systems are given in Table 2. For the 24-site system with 10+10 electrons, $\dim(\mathcal{H}) = 3\,846\,525\,097\,536$, the state vector takes about 28 TBytes. The simulation of such a system requires the entire machine, using only one processor per node (SMP mode), to make most efficient use of the available memory. For smaller systems we can use all four processors per node (VN mode). Despite massive communication in each iteration, the code shows excellent speed up. Only when the message size per process becomes too small, performance degrades because of network latency. This is shown in the lower plot. Properly scaling the execution times we obtain a universal scaling (ParLaw) for system sizes ranging over more than five and process counts ranging over three orders of magnitude.

We implement an efficient matrix transpose using `MPI_Alltoall`. This routine expects that the data packages which will be sent to a given process to be stored contiguously in memory. This does not apply to our case, since we would like to store the spin-down electron configurations sequentially in memory. Thus, the matrix is stored column wise. For `MPI_Alltoall` to work properly, we would have to bring the data elements into row-major order. This could be done by performing a local matrix transpose. The involved matrices are, however, in general rectangular, leading to expensive local-copy and reordering operations. We can avoid this by calling `MPI_Alltoall` for each column separately (red arrows in Figure 7). After this, only

a local strided transposition has to be performed (small white arrows) to obtain the fully transposed matrix or Lanczos vector. The implementation described so far uses `MPI_Alltoall` which assumes that the matrix to be transposed is a square matrix and that the dimension $\dim_{\uparrow} = \dim_{\downarrow}$ is divisible by the number of MPI processes. To overcome these restrictions we have generalized the algorithm to `MPI_Alltoallv`. This is the implementation that is used in practice. The speed-up shown at the top of Figure 8 shows that our parallelization based on collective communication is indeed very efficient. Even for a system of 24 sites with 10 electrons of either spin, where a single many-body vector takes about 28 TB of memory, our implementation works extremely well despite the fact that in each Lanczos iteration 28 TB of data have to be moved across the entire machine twice.

The lower plot in Figure 8 shows that the execution times for runs of systems varying by more than five orders of magnitude in size (of the Hilbert space) and for processor counts varying over three orders of magnitude fall on a universal curve, which is only determined by the bandwidth and the latency of the network. This suggests that the implementation should scale to even larger systems than the present Jülich BlueGene with almost 300 000 CPUs and an aggregate memory of 144 terabytes.

3 Application to DMFT

Using the Lanczos method as a solver for DMFT gives results at zero temperature and directly on the real axis. An important limitation is, however, the need to approximate the bath Green function

$$\mathcal{G}^{-1}(\omega) = \omega + \mu - \int_{-\infty}^{\infty} d\omega' \frac{\Delta(\omega')}{\omega - \omega'} \quad (41)$$

by a discretized version, e.g., of the form

$$\mathcal{G}_{\text{And}}^{-1}(\omega) = \omega + \mu - \sum_{l=1}^{N_b} \frac{V_l^2}{\omega - \varepsilon_l}, \quad (42)$$

corresponding to an Anderson impurity model with a finite number of sites

$$H_{\text{And}} = \varepsilon_0 \sum_{\sigma} n_{\sigma} + U n_{\uparrow} n_{\downarrow} + \sum_{\sigma} \sum_{l=1}^{N_b} \left(\varepsilon_l n_{l\sigma} + V_l \left(a_{l\sigma}^{\dagger} c_{\sigma} + c_{\sigma}^{\dagger} a_{l\sigma} \right) \right), \quad (43)$$

where $c_{\sigma}^{\dagger}$ and $a_{l\sigma}^{\dagger}$ create an electron of spin σ on the impurity or bath site l , respectively, $n_{\sigma} = c_{\sigma}^{\dagger} c_{\sigma}$ and $n_{l\sigma} = a_{l\sigma}^{\dagger} a_{l\sigma}$. Writing the non-interacting part of H_{And} as a matrix

$$H_{\text{And}}^0 = \begin{pmatrix} 0 & V_1 & V_2 & V_3 & \cdots \\ V_1 & \varepsilon_1 & 0 & 0 & \\ V_2 & 0 & \varepsilon_2 & 0 & \\ V_3 & 0 & 0 & \varepsilon_3 & \\ \vdots & & & & \ddots \end{pmatrix} \quad (44)$$

we see that (42) is easily recovered from inversion by partitioning.

Since practical calculations are limited by the rapidly increasing size of the Hilbert space to small numbers of bath sites N_b , it is crucial for the reliability of the calculations to find a good representation $\mathcal{G}_{\text{And}}^{-1}$ for the bath Green function. The most common approach is to use a least squares fit [15]: Because of the spectral poles on the real axis, such a fit in practice is done on the imaginary axis, where the Green functions are smooth and the optimization of the distance function is not easily trapped in local minima. One then minimizes a function of the form

$$\chi^2(\{V_l, \varepsilon_l\}) = \sum_{n=0}^{n_{\max}} |\mathcal{G}^{-1}(i\omega_n) - \mathcal{G}_{\text{And}}^{-1}(i\omega_n)|^2 \quad (45)$$

on a set of Matsubara frequencies corresponding to some fictitious temperature. The choice of this temperature and of $n_{\max}$ essentially determines the relative weighing of high- versus low-frequency features in the fit. If low $i\omega_n$ are weighted too little, the fit easily becomes underdetermined, since for large imaginary frequencies the hybridization function contains only little information about the system (which is the reason why the analytic continuation back to the real axis is so difficult). To emphasize different frequency ranges, it is possible to introduce frequency-dependent weight functions in (45).

Instead of fitting, we could use a moment expansion of the Weiss function $W(\omega) = \int d\omega' \frac{\Delta(\omega')}{\omega - \omega'}$ similar to that discussed in Section 1.2. Such an approach [16] has been used for the Bethe lattice with infinite coordination, where the self-consistency condition simplifies to $W(\omega) = t^2 G_{\text{imp}}(\omega)$: As Lanczos gives a continued-fraction representation for the photoemission and inverse-photoemission part separately, the hybridization function is written as

$$W^<(\omega) + W^>(\omega) = t^2 G^<(\omega) + t^2 G^>(\omega) = \frac{t^2 b_0^{<2}}{\omega + a_0^< - \frac{b_1^{<2}}{\omega + a_1^< - \dots}} + \frac{t^2 b_0^{>2}}{\omega - a_0^> - \frac{b_1^{>2}}{\omega - a_1^> - \dots}} \quad (46)$$

Truncating the continued fractions at $N_b^<$ and $N_b^>$, this corresponds to the impurity model with

$$H_{\text{And}}^0 = \begin{pmatrix} 0 & t^2 b_0^< & & \dots & t^2 b_0^> \\ t^2 b_0^< & -a_0^< & b_1^< & & \\ & b_1^< & -a_1^< & b_2^< & \\ & & b_2^< & -a_2^< & \ddots \\ \vdots & & & \ddots & \ddots \\ t^2 b_0^> & & & & a_0^> & b_1^> \\ & & & & b_1^> & a_1^> & b_2^> \\ & & & & & b_2^> & a_2^> & \ddots \\ & & & & & & & \ddots \end{pmatrix}, \quad (47)$$

where the bath forms two chains, coupled to the impurity. Diagonalizing the bath, it is easily brought to the form (43). This bath parametrization works very well for systems with a large gap. An example is shown in Figure 9. Since the approach uses moment expansions for the

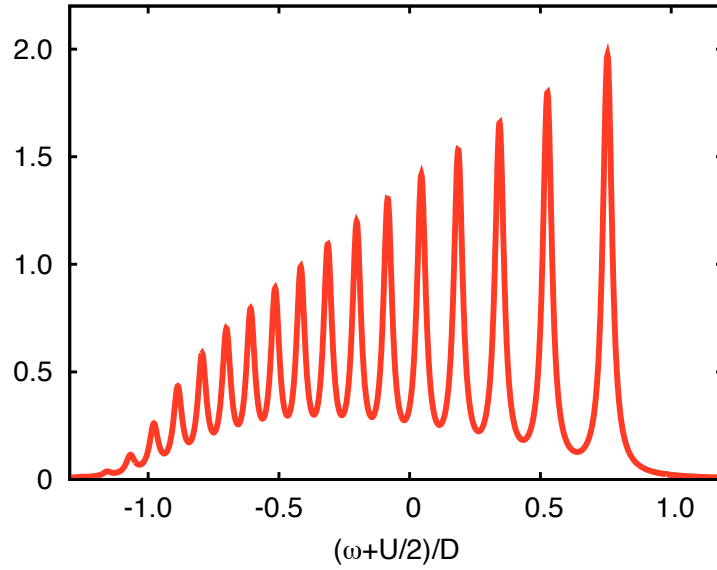


Fig. 9: Spectral function of the lower Hubbard band of a Hubbard model on the infinite Bethe lattice with half bandwidth D and $U = 8D$ in antiferromagnetic DMFT. The bath was obtained from the continued-fraction expansion of the impurity Green's functions, $N_b = 24$ [18].

two parts $W^<$ and $W^>$ of the Weiss function separately, it does not converge as quickly as a moment expansion for the full hybridization function would. This makes itself particularly felt when the gap is small or the system is even metallic. To improve the description of the hybridization function we can combine the separate continued fractions for photoemission and inverse photoemission into a single one that can, with the same number of bath parameters, reproduce twice as many moments of the bath spectral function [17].

3.1 Cluster methods

For cluster versions of DMFT we can use exact sum rules and symmetries to find the structure of the bath. Our discussion will closely follow [19]. To fix the notation we briefly sketch the self-consistency loop for cellular DMFT (CDMFT) and the dynamical cluster approximation (DCA) using Lanczos as impurity solver. Let N_c be the number of cluster sites, N_b the number of bath sites. For simplicity we suppress spin indices.

Given an $N_c \times N_c$ bath Green matrix $\mathcal{G}^{-1}$,

1. fit the parameters of an Anderson model with N_b bath sites

$$\mathcal{G}_{\text{And}}^{-1}(\omega) = \omega + \mu - \mathbf{H}_c - \mathbf{\Gamma} [\omega - \mathbf{E}]^{-1} \mathbf{\Gamma}^\dagger \quad (48)$$

to $\mathcal{G}^{-1}$, where $\mathbf{\Gamma}$ is the $N_c \times N_b$ -dimensional hybridization matrix, and $\mathbf{E}$ the $N_b \times N_b$ -dimensional bath-matrix. $\mathbf{H}_c$ is specified below,

2. solve the $(N_c + N_b)$ -site Anderson model H_{And} (specified below) to obtain the $N_c \times N_c$ cluster Green matrix $\mathbf{G}_c$,

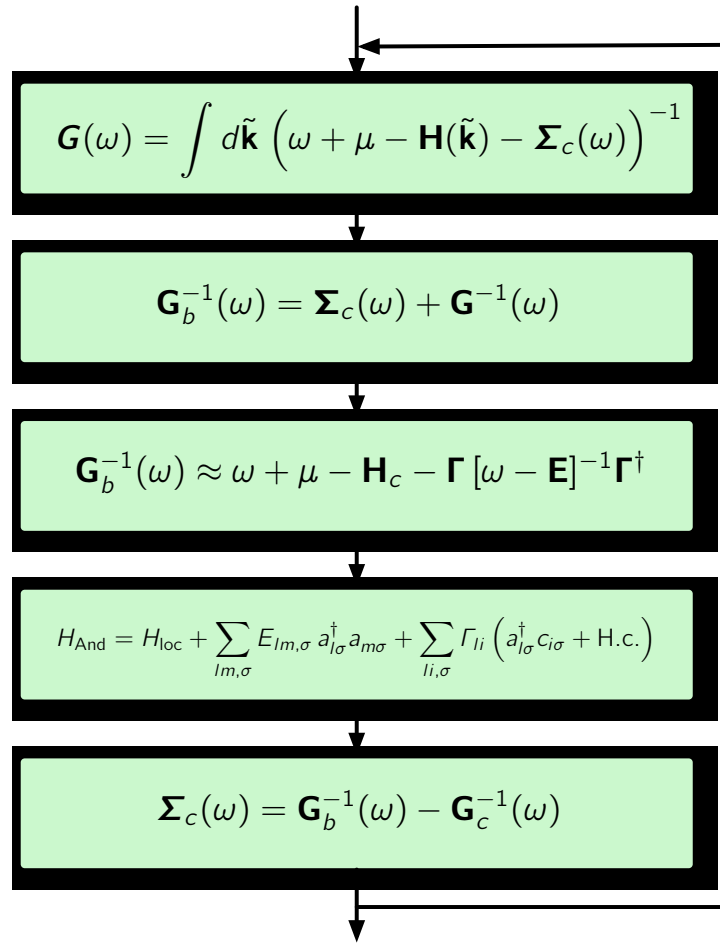


Fig. 10: Self-consistency loop for (cluster) DMFT.

3. get the cluster self-energy matrix

$$\Sigma_c(\omega) = \mathcal{G}^{-1}(\omega) - \mathbf{G}_c^{-1}(\omega), \quad (49)$$

4. calculate the local Green matrix for the cluster by integrating over the reduced Brillouin zone of the cluster

$$\mathbf{G}(\omega) = \int d\tilde{\mathbf{k}} \left(\omega + \mu - \mathbf{H}(\tilde{\mathbf{k}}) - \Sigma_c(\omega) \right)^{-1}, \quad (50)$$

where $\mathbf{H}(\tilde{\mathbf{k}})$ is the single-electron part of the Hubbard Hamiltonian (32) in the reduced Brillouin zone of the cluster,

5. determine the new bath Green matrix (self-consistency condition)

$$\mathcal{G}^{-1}(\omega) = \Sigma_c(\omega) + \mathbf{G}^{-1}(\omega). \quad (51)$$

These steps are iterated to self-consistency.

3.2 Anderson impurity model

The Anderson model to be solved in step 2 is given by

$$H_{\text{And}} = H_{\text{clu}} + \sum_{lm,\sigma} E_{lm,\sigma} a_{l\sigma}^\dagger a_{m\sigma} + \sum_{li,\sigma} \Gamma_{il} \left(a_{l\sigma}^\dagger c_{i\sigma} + \text{H.c.} \right) \quad (52)$$

where the operator $a_{l\sigma}^\dagger$ creates an electron of spin σ on bath site l . The cluster Hamiltonian H_{clu} is obtained from the lattice Hamiltonian by transforming to the reciprocal space of the super-lattice of the clusters and projecting to the cluster. Writing the single-electron part of $H(\tilde{\mathbf{k}})$ as the matrix $\mathbf{H}(\tilde{\mathbf{k}})$, the single-electron part of H_{clu} is given by

$$\mathbf{H}_c = \int d\tilde{\mathbf{k}} \mathbf{H}(\tilde{\mathbf{k}}). \quad (53)$$

The (local) interaction terms are simply those of the lattice model, restricted to the cluster. The Hamiltonian $H(\tilde{\mathbf{k}})$ in the reciprocal space of the super-lattice $\{\tilde{\mathbf{r}}\}$ of clusters can be obtained by changing to the basis of operators

$$\tilde{c}_{\mathbf{R}_i\sigma}^{\text{CDMFT}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i\tilde{\mathbf{k}}\tilde{\mathbf{r}}} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (54)$$

The resulting quantum cluster approximation is CDMFT. Alternatively, we can start from the operators in the reciprocal space of the *lattice* to obtain

$$\tilde{c}_{\mathbf{R}_i\sigma}^{\text{DCA}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i\tilde{\mathbf{k}}(\tilde{\mathbf{r}}+\mathbf{R}_i)} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (55)$$

Now we obtain the DCA. The choice of the operators in the two approaches differs just by local phase factors. In CDMFT this Kohn gauge [20] is chosen such that phases appear only in matrix elements involving *different* clusters. Thus all matrix elements on the cluster are the same as in the original Hamiltonian. The price for retaining the original matrix elements on the cluster is a breaking of the translation symmetry of the original lattice. DCA opts instead to retain this symmetry by distributing the phase change uniformly over the cluster sites. The price for retaining translation invariance is that the matrix elements in the cluster Hamiltonian differ from those in the original Hamiltonian (coarse graining). In both cases, CDMFT and DCA, the eigenvalues of $\mathbf{H}(\tilde{\mathbf{k}})$ are identical to the eigenvalues of the non-interacting part of H . Obviously, we could construct other cluster extensions to DMFT by different choices of the Kohn gauge $\varphi(\tilde{\mathbf{k}}; \mathbf{R}_i)$ on the cluster

$$\tilde{c}_{\mathbf{R}_i\sigma}^\varphi(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i(\tilde{\mathbf{k}}\tilde{\mathbf{r}}+\varphi(\tilde{\mathbf{k}};\mathbf{R}_i))} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (56)$$

3.3 Hybridization sum rules

While the most general parametrization for the bath is given by expression (48), we can always diagonalize the hopping matrix $\mathbf{E}$ among the bath sites to obtain

$$\mathcal{G}_{\text{And}}^{-1}(\{\varepsilon_l, \mathbf{V}_l\}; \omega) = \omega + \mu - \mathbf{H}_c - \sum_l \frac{\mathbf{V}_l \mathbf{V}_l^\dagger}{\omega - \varepsilon_l}. \quad (57)$$

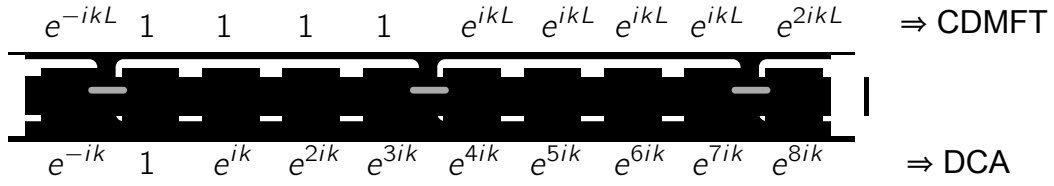


Fig. 11: Phase choice on the cluster that leads to CDMFT or DCA.

The hybridization matrix is then given by the tensor product of the vectors $\mathbf{V}_l$, where

$$V_{l,i} = \sum_m \Gamma_{i,m} \phi_{l,m} \quad (58)$$

and ϕ_l are the eigenvectors of $\mathbf{E}$ with eigenvalues ε_l .

To obtain sum rules for the hybridizations, we write the inverse of the bath Green matrix as

$$\mathcal{G}^{-1}(\omega) = \Sigma_c(\omega) + \left(\int d\tilde{\mathbf{k}} \left(\omega + \mu - \mathbf{H}(\tilde{\mathbf{k}}) - \Sigma_c(\omega) \right)^{-1} \right)^{-1}.$$

Considering the limit $\omega \rightarrow \infty$, expanding to order $1/\omega^2$, using (53), and comparing to (57) we find

$$\sum_l \mathbf{V}_l \mathbf{V}_l^\dagger = \int d\tilde{\mathbf{k}} \mathbf{H}^2(\tilde{\mathbf{k}}) - \left(\int d\tilde{\mathbf{k}} \mathbf{H}(\tilde{\mathbf{k}}) \right)^2. \quad (59)$$

To illustrate this hybridization sum rule we consider a representative set of examples.

Single site

We consider a d -dimensional lattice with hoppings t_n to the z_n n^{th} -nearest neighbors. For $N_c = 1$ we have $\mathbf{H}(\mathbf{k}) = \varepsilon_{\mathbf{k}}$. Thus we find for the hybridizations

$$\sum_l V_l^2 = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} d^d \mathbf{k} \varepsilon_{\mathbf{k}}^2 = \sum_n z_n t_n^2, \quad (60)$$

where the integral is just the second moment of the density of states, so that the last equation follows as in the recursion method [7]. For a Bethe lattice of connectivity z with hopping matrix element $t/\sqrt{z}$ the sum rule reduces to $\sum_l V_l^2 = t^2$.

CDMFT

We start by considering a linear chain with nearest-neighbor hopping t and a 3-site cluster $N_c = 3$. In the CDMFT gauge we have

$$\mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & 1 & e^{-3i\tilde{k}} \\ 1 & 0 & 1 \\ e^{3i\tilde{k}} & 1 & 0 \end{pmatrix} \quad (61)$$

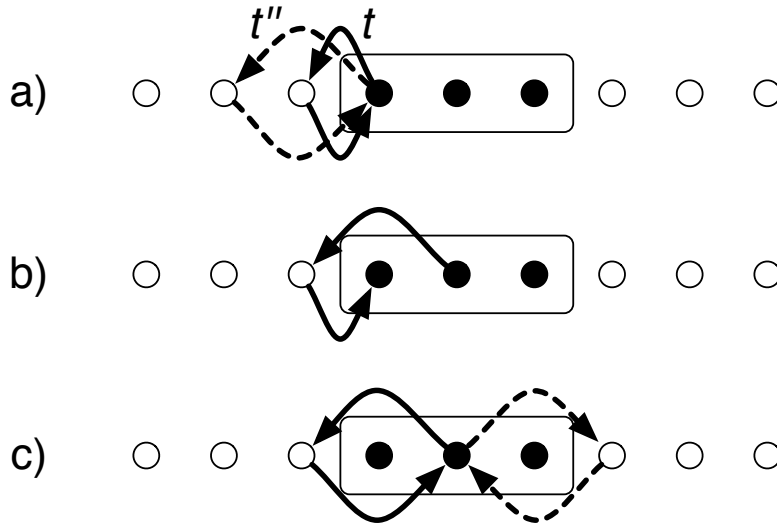


Fig. 12: CDMFT sum rules for a one-dimensional 3-site cluster with nearest and next-nearest-neighbor hoppings t and t'' , respectively: a) $\sum_l |V_{l,1}|^2 = t^2 + t''^2$, b) $\sum_l \bar{V}_{l,1} V_{l,2} = t t''$, and c) $\sum_l |V_{l,2}|^2 = 2t''^2$. The hybridizations are given by the two-step hopping processes that are lost when the cluster is cut out of the original lattice.

so that $\mathbf{H}_c$ is the original single-electron Hamiltonian restricted to the cluster:

$$\mathbf{H}_c = \frac{3}{2\pi} \int_{-\pi/3}^{\pi/3} d\tilde{k} \mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (62)$$

The sum rule (59) then is

$$\left(\sum_l V_{l,i} \bar{V}_{l,j} \right) = \begin{pmatrix} t^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & t^2 \end{pmatrix}, \quad (63)$$

i.e., only the sites on the surface of the cluster couple to the bath. If we also allow second-nearest-neighbor hopping with matrix element t'' , we find

$$\left(\sum_l \bar{V}_{l,\mu} V_{l,\nu} \right) = \begin{pmatrix} t^2 + t''^2 & t t'' & 0 \\ t t'' & 2t''^2 & t t'' \\ 0 & t t'' & t^2 + t''^2 \end{pmatrix}. \quad (64)$$

The general CDMFT hybridization sum rule (59) can be easily visualized: The integral over the Brillouin zone of the cluster projects the single-electron part of the full Hamiltonian to the cluster (see Eqn. (53)). The matrix elements of $\mathbf{H}_c^2$ are thus the two-step hoppings that are possible on the cluster. Likewise the integral over the Hamiltonian squared gives the second moments, but note that here the intermediate site is not restricted to the cluster. Thus the sum-rule matrix is given by the second-order paths between cluster sites that proceed via a site outside the cluster. This is illustrated in Figure 12. As a special case, for a single site we recover the second equality in (60).

The vanishing of a matrix element in the sum rule merely implies that the corresponding matrix element of the bath Green matrix decays faster than $1/\omega$ for large ω . For a diagonal element, however, all terms in $\sum_l V_{l,i} \bar{V}_{l,i}$ are positive. Thus a vanishing sum means that all terms must be zero. Hence the sum rule implies that cluster sites that are so far in the interior that they cannot be reached by hopping from outside the cluster do not couple to the bath and that all matrix elements of the bath Green function involving such a site i are given by $\mathcal{G}_{ij}^{-1}(\omega) = \omega + \mu - (\mathbf{H}_c)_{ij}$ for all ω . In that sense the bath hybridizes only to the surface of the cluster and we see that the hybridization strength to these sites does not decrease for increasing cluster size N_c .

DCA

We start again by considering the 3-site cluster. In the DCA gauge we write

$$\mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & e^{i\tilde{k}} & e^{-i\tilde{k}} \\ e^{-i\tilde{k}} & 0 & e^{i\tilde{k}} \\ e^{i\tilde{k}} & e^{-i\tilde{k}} & 0 \end{pmatrix}. \quad (65)$$

Now $\mathbf{H}_c$ has translational symmetry, but the hopping is rescaled by $\sin(\pi/N_c)/(\pi/N_c)$

$$\mathbf{H}_c = \frac{3}{2\pi} \int_{-\pi/3}^{\pi/3} d\tilde{k} \mathbf{H}(\tilde{k}) = -\frac{3\sqrt{3}}{2\pi} t \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad (66)$$

Since all matrices in (59) are periodic, it is convenient to transform to k -space. With $V_{l,K} = \sum_i V_{l,i} e^{iKr_i}/\sqrt{N_c}$ and the coarse-graining factor $\tau = 3\sqrt{3}/2\pi$ we find

$$\begin{aligned} \sum_l |V_{l,K=0}|^2 &= (2 + \tau - 4\tau^2) t^2 \\ \sum_l |V_{l,K=\pm 2\pi/3}|^2 &= (2 - \tau/2 - \tau^2) t^2. \end{aligned}$$

The hybridization sum rule (59) is then, likewise, diagonal in the cluster momenta $\mathbf{K}$

$$\sum_l |V_{l,\mathbf{K}}|^2 = \int d\tilde{\mathbf{k}} \varepsilon_{\mathbf{K}+\tilde{\mathbf{k}}}^2 - \left(\int d\tilde{\mathbf{k}} \varepsilon_{\mathbf{K}+\tilde{\mathbf{k}}} \right)^2, \quad (67)$$

while all terms $V_{l,\mathbf{K}} \bar{V}_{l,\mathbf{K}'}$ mixing different cluster momenta vanish. As a special case, for a single site the above sum rule is just the first equality in (60). Expanding $\varepsilon_{\mathbf{K}+\mathbf{k}}$ around $\mathbf{K}$, we find that for a d -dimensional system $\sum_l |V_{l,\mathbf{K}}|^2$ decreases with cluster size as $1/N_c^{2/d}$, while all cluster sites couple with the same strength to the bath.

Discussion

Besides providing exact relations for the bath parametrization, in particular which sites need not be coupled to a bath, the sum rules contain important information about the scaling of cluster methods with cluster size: In CDMFT individual hybridizations are independent of cluster size,

while for DCA they decrease with cluster size as $N_c^{-2/d}$. Interestingly this means that for a d -dimensional system in CDMFT the overall coupling to the bath scales as $N_c^{(d-1)/d}$, while in DCA it scales as $N_c^{(d-2)/d}$. For non-local properties, a DCA calculation is therefore expected to converge faster with cluster size. For a calculation where we represent the bath by discrete degrees of freedom this decrease in hybridization strength does not, however, help very much as we still need bath sites to fit the hybridizations, even if they are small. With increasing DCA cluster size we thus have to parametrize N_c baths, one for each $\mathbf{K}$. In CDMFT the situation is more fortunate, as the sum rules imply that many hybridizations vanish and we only need to parametrize the coupling of surface sites to the bath.

The lack of translational invariance in CDMFT has, however, two important practical implications. First, the full Green matrix has to be calculated, instead of just its diagonal. Second, when calculating local quantities, like the density per site, in CDMFT we have a choice of considering each inequivalent site or the average over all sites. In a gapped system the best choice is the innermost site. In such a situation it might, however, be better to do a straight Lanczos calculation with $N_c + N_b$ cluster sites, instead of using N_b bath sites.

3.4 Symmetries

In the absence of spontaneous symmetry breaking, the symmetries of the cluster (point symmetries in CDMFT and additionally translational symmetry in DCA) are reflected in the Green matrix. In a symmetry-broken state with long-range order, like an antiferromagnet or a charge-density wave, the symmetry of the Green matrix is accordingly lowered. To exploit the symmetry we introduce vectors on the cluster that transform according to its irreducible representations. We write these vectors as $\mathbf{w}_{I,\nu}$ where I is the irreducible representation and $\nu = 1 \dots N_I$ counts linearly independent vectors transforming according to I . On an N_c -site cluster we can choose N_c such vectors that are orthonormal. Defining the matrix $\mathbf{W} = (\mathbf{w}_{I,\nu})$ of these vectors, we can block-diagonalize the bath Green matrix: $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ has blocks of dimension N_I corresponding to the different irreducible representations I . Since $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ is block diagonal for all ω , it follows from equation (57), that $\mathbf{W}$ must also block-diagonalize the individual hybridization matrices $\mathbf{V}_l \mathbf{V}_l^\dagger$. Therefore the hybridization vectors must transform according to an irreducible representation: They can be written as $\mathbf{V}_l = \sum_\nu V_{l;I,\nu} \mathbf{w}_{I,\nu}$ for some irreducible representation I . If the $\mathbf{V}_l$ also had components $\mathbf{w}_{J,\nu}$ of a different irreducible representation $J \neq I$ this would produce a hybridization matrix that could not be block-diagonalized. This can only happen for bath sites with identical energy ε_l (accidental degeneracy): Assume $\mathbf{V}_l$ and $\mathbf{V}_{l'}$ are the hybridizations for two bath sites with $\varepsilon_l = \varepsilon_{l'}$. Then we can form arbitrary linear combinations of the hybridization matrices and hence of the hybridization vectors. For all these linear combinations the sum of the hybridization matrices must be block diagonal, and hence we can choose the hybridization vectors such that they transform according to irreducible representations.

We thus find that the bath sites can be arranged into sets corresponding to the different irreducible representations. For fitting a block of the symmetrized bath Green matrix we need

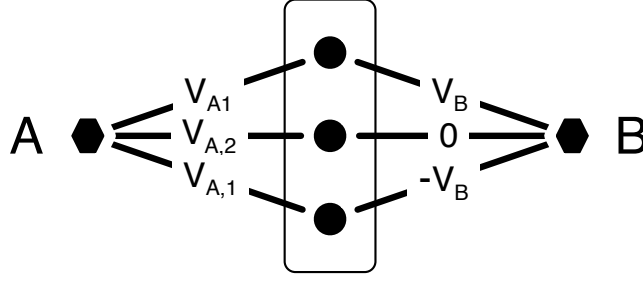


Fig. 13: Hybridization of bath sites of symmetry A and B to a 3-site cluster. As defined in table 3, A is the unit representation, so a bath site of type A has the same hybridization V to all cluster sites that are equivalent by symmetry. B is the antisymmetric representation, so the hybridization of a bath site of type B to cluster sites that are related by mirror symmetry has the opposite sign. Consequently the hybridization to the central site of a linear cluster with an odd number of sites vanishes in the B representation.

then only consider bath sites of the respective irreducible representation. If the block is one-dimensional we can choose the corresponding hybridizations to be real.

Sectors corresponding to different irreducible representations are only coupled through the Hubbard interaction U when solving the Anderson impurity model. Note that the coupling to bath sites corresponding to an irreducible representation other than the unit representation lowers the symmetry of the impurity Hamiltonian with respect to that of the Green matrix.

CDMFT

As an example we consider a linear cluster of 3 sites as shown in Figure 13. The symmetry is C_2 (see Table 3). Transforming to the basis vectors $\mathbf{w}_{A,1} = (|1\rangle + |3\rangle)/\sqrt{2}$ and $\mathbf{w}_{A,2} = |2\rangle$ of symmetry A (see Table 3) and $\mathbf{w}_B = (|1\rangle - |3\rangle)/\sqrt{2}$, we find the transformed bath Green matrix

$$\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W} = \begin{pmatrix} \mathcal{G}_{11}^{-1} + \mathcal{G}_{13}^{-1} & \sqrt{2} \mathcal{G}_{12}^{-1} & 0 \\ \sqrt{2} \mathcal{G}_{21}^{-1} & \mathcal{G}_{22}^{-1} & 0 \\ 0 & 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1} \end{pmatrix}.$$

A bath site of irreducible representation A contributes to the first block and has the same hybridization $V_{A,1}$ to the outer cluster sites plus an independent hybridization parameter $V_{A,2}$ to the central site. A bath site of irreducible representation B contributes to the second block. For such a bath site the hybridization to cluster sites that are related by mirror symmetry have opposite signs. Consequently, the hybridization to the central site vanishes.

The situation is slightly more complicated when the symmetry group has irreducible representations of dimension higher than one. The simplest example is the 2×2 cluster with C_{4v} symmetry. With $\mathbf{w}_{A_1} = (|1\rangle + |2\rangle + |3\rangle + |4\rangle)/2$, $\mathbf{w}_{B_2} = (|1\rangle - |2\rangle + |3\rangle - |4\rangle)/2$, and the pair $\mathbf{w}_{E,1} = (|1\rangle - |2\rangle - |3\rangle + |4\rangle)/2$, $\mathbf{w}_{E,2} = (|1\rangle + |2\rangle - |3\rangle - |4\rangle)/2$ we find that $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ is

C_2	E	σ_v		C_{2v}	E	C_2	σ_v	σ'_v
A	1	1		A_1	1	1	1	1
B	1	-1		A_2	1	1	-1	-1
				B_1	1	-1	1	-1
				B_2	1	-1	-1	1

C_{3v}	E	$2C_3$	$3\sigma_v$		C_{4v}	E	$2C_4$	C_4^2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1		A_1	1	1	1	1	1
A_2	1	1	-1		A_2	1	1	1	-1	-1
E	2	-1	0		B_1	1	-1	1	1	-1
					B_2	1	-1	1	-1	1
					E	2	0	-2	0	0

Table 3: Character tables of the point groups C_{1v} , C_{2v} , C_{3v} , and C_{4v} .

diagonal with diagonal elements

$$\begin{aligned}
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{11} &= \mathcal{G}_{11}^{-1} + 2\mathcal{G}_{12}^{-1} + \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{22} &= \mathcal{G}_{11}^{-1} - 2\mathcal{G}_{12}^{-1} + \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{33} &= \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{44} &= \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1}
\end{aligned}$$

A bath site of symmetry A_1 has the same hybridization to all cluster sites while for a bath site of symmetry B_2 the hybridizations have alternating signs: $\mathbf{V}_l^\dagger = \bar{V}_l (1, -1, 1, -1)$. To realize the two-dimensional representation E we need two bath sites l_1 and l_2 with degenerate energies $\varepsilon_{l_1} = \varepsilon_{l_2} = \varepsilon_l$ and hybridizations: $\mathbf{V}_{l_1}^\dagger = \bar{V}_l (1, -1, -1, 1)$ and $\mathbf{V}_{l_2}^\dagger = \bar{V}_l (1, 1, -1, -1)$. This is illustrated in Figure 14.

DCA

As an example for DCA, we consider a 3-site cluster with periodic boundary conditions. The symmetry group is C_{3v} (translations and inversion). Hence we introduce the basis vector $\mathbf{w}_{A_1} = (|1\rangle + |2\rangle + |3\rangle)/\sqrt{3}$, corresponding to $k = 0$, while the vectors formed by $\sin(2\pi/3)$ and $\cos(2\pi/3)$ give the E representation: $\mathbf{w}_{E,1} = (|1\rangle - |2\rangle)/\sqrt{2}$ and $\mathbf{w}_{E,2} = (|1\rangle + |2\rangle - 2|3\rangle)/\sqrt{6}$.

$$\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W} = \begin{pmatrix} \mathcal{G}_{11}^{-1} + 2\mathcal{G}_{12}^{-1} & 0 & 0 \\ 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{12}^{-1} & 0 \\ 0 & 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{12}^{-1} \end{pmatrix}.$$

In general bath sites corresponding to the Γ point have the same hybridization to all cluster sites, while those corresponding to $k = \pi$ have alternating hybridizations. For all other k -points we need two degenerate bath sites, with hybridizations $V_{l_1,\mu} = V_l \sin(k\mu)$ and $V_{l_2,\mu} = V_l \cos(k\mu)$ to cluster site μ .

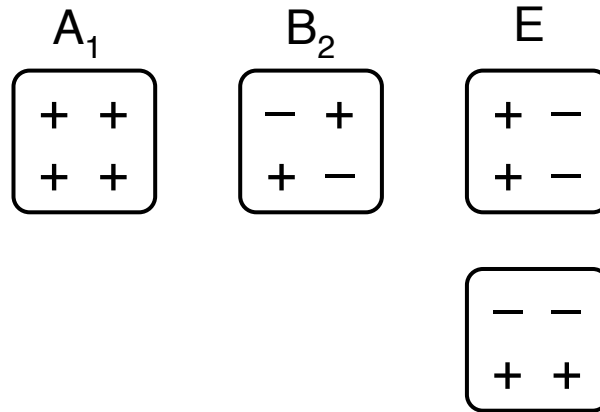


Fig. 14: Hybridization of bath sites of symmetry A_1 , B_2 , and E to a 2×2 cluster. For a given irreducible representation the absolute value of the hybridization to all cluster sites is the same, while the signs are indicated in the Figure. Non-trivial hybridizations corresponding to irreducible representations A_2 or B_1 only appear for larger clusters.

4 Conclusions

We have seen that the Lanczos method is unbelievably efficient for calculating ground-state and dynamical response functions of many-body Hamiltonians. The determination of the ground-state takes only about $\mathcal{O}(\dim(\mathcal{H}))$ in time and memory. The iteration already converges after about a hundred steps, even for Hilbert spaces with dimensions in the trillions. This astounding convergence is based on the idea of steepest descent to the ground state, which the Lanczos method even improves upon. In addition, we can very efficiently calculate Green functions. Here the rapid convergence is due to the fact that the Lanczos iteration reproduces more and more moments of the spectral function.

The great advantages of the Lanczos approach to strongly correlated systems is that it provides us with expressions for the Green function on the entire complex plane, i.e., in particular for real frequencies. The ground state (i.e. $T = 0$) is calculated directly, but the method can easily be extended to finite but low temperatures.

The greatest disadvantage is the need to store full many-body vectors. Calculations are therefore restricted by the available memory to relatively small systems. To minimize the effects of finite system size it is therefore crucial to (i) efficiently use the vast distributed memories of current massively parallel machines and to (ii) find bath parametrizations that minimize the effect of truncating it to finite size.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] C. Lanczos: *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators*, J. Res. Nat. Bur. Stand. **49**, 255 (1950)
- [2] W.E. Arnoldi, Quarterly of Applied Mathematics **9**, 17 (1951)
- [3] C.C. Paige: *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices* (PhD thesis, London University, 1971)
- [4] L.N. Trefethen and D. Bau III: *Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 1997)
- [5] G.H. Golub and C.F. van Loan: *Matrix Computations* (Johns Hopkins University Press, 1996)
- [6] J.K. Cullum and R.A. Willboughby: *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. 1: Theory, Vol. 2: Programs* (Birkhäuser, Boston, 1985)
- [7] R. Haydock: *The Recursive Solution of the Schrödinger Equation*, in H. Ehrenreich, F. Seitz, and D. Turnbull (eds.): *Solid State Physics* **35**, 216 (Academic Press, 1980)
- [8] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994)
- [9] L.V. Kantorovich, Uspekhi Mat. Nauk, **3**, 89 (1948)
- [10] M.R. Hestenes and W. Karush, J. Res. Nat. Bureau Standards **47**, 45 (1951)
- [11] A.N. Krylov, Izv. Akad. Nauk SSSR, Otd. Mat. Estest. **7**, 491 (1931)
- [12] J. Jaklič and P. Prelovšek, Adv. Phys. **49**, 1 (2000)
M. Aichhorn *et al.*, Phys. Rev. B **67**, 161103 (2003)
- [13] A. Dolfen: *Massively parallel exact diagonalization of strongly correlated systems* (Diploma Thesis, RWTH Aachen University, 2006)
- [14] A. Dolfen, T.L. Luo, and E. Koch, Advances in Parallel Computing **15**, 601 (2008)
- [15] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
- [16] Q. Si, M.J. Rozenberg, G. Kotliar, and A.E. Ruckenstein, Phys. Rev. Lett. **72**, 2761 (1994)
- [17] M. Bugeanu: *Simulations of strongly correlated materials: Clusters and DMFT using a Lanczos solver*, MSc Thesis, German Research School for Simulation Sciences, 2012
- [18] G. Sangiovanni, A. Toschi, E. Koch, *et al.*, Phys. Rev. B **73**, 205121 (2006)
- [19] E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008)
- [20] W. Kohn, Phys. Rev. **133**, A171 (1964)

11 The Hubbard Model and its Properties

Andreas Mielke

Institut für Theoretische Physik, Universität Heidelberg

Philosophenweg 19, 69120 Heidelberg, Germany

Contents

1	Introduction	2
2	The Hubbard model	2
2.1	Definition	2
2.2	Symmetries of the Hubbard model	3
3	Some rigorous results	5
3.1	Lieb's Theorem	5
3.2	The Mermin-Wagner theorem	8
3.3	Nagaoka's theorem	9
3.4	Flat-band systems	10
3.5	Uniform density theorem	14
3.6	Further rigorous results	14
4	(Functional) Renormalization	15
4.1	General idea	15
4.2	Field-theoretic representation of the Hubbard model	15
4.3	Renormalization group equations for G_{eff}	19
4.4	Numerical solutions	20
4.5	Some results	21
5	Summary, conclusions, and outlook	22

1 Introduction

The Hubbard model, although highly oversimplified, contains the main ingredients to describe interacting quantum mechanical particles, originally fermions, moving in a solid. Its basis is a tight binding description. The Hamiltonian defining the model contains two parts: a single-particle part and a two-particle interaction. The idea is that only one or few energy bands close to the Fermi energy contribute. Therefore, the single-particle part, often called kinetic energy, describes particles hopping on a lattice which may have a single or a few bands. The Coulomb interaction is assumed to be screened. In the Hubbard model, the interaction taken into account is just an on-site interaction, the range of the interaction is thus the shortest possible. With this setup, the Hubbard model is certainly only a caricature of a realistic description of electrons in a solid. Nevertheless, the Hubbard model exhibits almost all interesting phenomena one observes in nature: magnetic ordering of any kind, a metal-insulator transition, superconductivity (it is even used in the context of high temperature superconductivity), a Tomonaga-Luttinger liquid in one space dimension, and even more sophisticated transitions like a Pomeranchuk instability. The Hubbard model can thus be viewed as the simplest possible model of correlated fermions. The Hubbard model contains few independent parameters. If we assume that hopping is only allowed between nearest-neighbor lattice sites, and if we assume translational invariance, the hopping is described by a single parameter. Under the same assumptions, the interaction is described by a single parameter as well. Since the energy scale can be chosen freely, only the ratio between these two parameters is important. The second parameter is the electron density on the lattice. And the third parameter is the lattice itself. In fact, we will see that the lattice is essential for the properties of the model.

Despite its simplicity, only few properties of the Hubbard model have been proven rigorously. Nevertheless, for almost all the different phenomena, rigorous results exist and serve as landmarks for any kind of approximation that is applied to the Hubbard model. Therefore, in this lecture, I concentrate on the rigorous results for the Hubbard model.

2 The Hubbard model

2.1 Definition

The Hamiltonian of the Hubbard model is given by

$$H = H_{\text{kin}} + H_{\text{int}} = \sum_{x,y \in V, \sigma} t_{xy} c_{x,\sigma}^\dagger c_{y,\sigma} + \sum_x U_x c_{x\uparrow}^\dagger c_{x\downarrow}^\dagger c_{x\downarrow} c_{x\uparrow} \quad (1)$$

The model was proposed independently by J. Hubbard [1] for the description of transition metals, by J. Kanamori [2] for the description of itinerant ferromagnetism, and by M.C. Gutzwiller [3] for the description of the metal-insulator transition. In Chemistry, the model is popular as well, and was introduced ten years earlier [4–6]. Under the name Pariser-Parr-Pople model it has been used to describe extended π -electron systems.

Typically, one assumes that the vertex set V forms a translationally invariant lattice and that U_x is independent of x , *i.e.*, $U_x = U$. But more general settings are possible. Especially in the quantum chemical context, V is just a general graph, t_{xy} and U_x depend on the lattice sites.

On a regular lattice, one often assumes nearest-neighbor hopping, *i.e.*, $t_{xy} = t$ for nearest-neighbored sites such that $|x - y| = 1$, and $t_{xy} = 0$ otherwise. Sometimes, a next-nearest-neighbor hopping $t_{xy} = t'$ for $|x - y| = 2$ is introduced. In Sect. 4.5 we will see that such a next nearest-neighbor hopping may change the physical behavior of the system drastically.

For small U and in two or more dimensions, one expects that the Hubbard model describes a Fermi liquid. We will come back to that point later, in Sect. 4, where we sketch how renormalization theory is used to obtain instabilities of the Fermi liquid. Typically, one is interested in the case where the model describes strongly interacting electrons, *i.e.* correlated electrons. In that situations, the interaction U is as large as or larger than typical values of t_{xy} .

For a general overview on the Hubbard model and on correlated fermions in general I refer to the book of Fulde [7]. An overview on rigorous results for the Hubbard model can be found in the article of Lieb [8], an overview on ferromagnetism in the Hubbard model in [9].

2.2 Symmetries of the Hubbard model

The Hubbard model has several symmetries:

Gauge symmetry:

$$c_{x\sigma}^\dagger \rightarrow \exp(i\alpha) c_{x\sigma}^\dagger, \quad c_{x\sigma} \rightarrow \exp(-i\alpha) c_{x\sigma} \quad (2)$$

The Hamiltonian remains invariant if this transformation is applied. As a consequence, the particle number $N_e = \sum_{x\sigma} c_{x\sigma}^\dagger c_{x\sigma}$ is conserved. This is a generic property of almost all models in condensed matter theory that describe fermions.

Spin symmetry: With the help of the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3)$$

we define local

$$S_{\alpha,x} = \frac{1}{2} \sum_{\sigma,\sigma'} c_{x\sigma}^\dagger (\sigma_\alpha)_{\sigma,\sigma'} c_{x\sigma'}, \quad \alpha = x, y, z, \quad \mathbf{S}_x = (S_{x,x}, S_{x,y}, S_{x,z}) \quad (4)$$

and global spin operators.

$$S_\alpha = \sum_x S_{\alpha,x}, \quad \mathbf{S} = (S_x, S_y, S_z) \quad (5)$$

Often one uses

$$S_\pm = S_x \pm iS_y, \quad S_+ = \frac{1}{2} \sum_x c_{x\uparrow}^\dagger c_{x\downarrow}, \quad S_- = S_+^\dagger \quad (6)$$

These operators form an $SU(2)$ algebra. The Hamiltonian commutes with these operators: it has a $SU(2)$ -symmetry. We have

$$[S_x, S_y] = iS_z \quad (7)$$

H , $\mathbf{S}^2$ and S_z can be diagonalized simultaneously. We denote the eigenvalues of $\mathbf{S}^2$ as $S(S+1)$, where S is the spin of the eigenstate. $S \propto N_e$, *i.e.*, an extensive value for S , means that the state is ferro- or ferri-magnetic.

Particle-hole transformations: Using the transformation

$$c_{x\sigma}^\dagger \rightarrow c_{x\sigma}, \quad c_{x\sigma} \rightarrow c_{x\sigma}^\dagger \quad (8)$$

the Hamiltonian becomes

$$\begin{aligned} H \rightarrow H' &= \sum_{x,y,\sigma} t_{xy} c_{x\sigma} c_{y\sigma}^\dagger + U \sum_x c_{x\uparrow} c_{x\downarrow} c_{x\downarrow}^\dagger c_{x\uparrow}^\dagger \\ &= - \sum_{x,y,\sigma} t_{xy} c_{y\sigma}^\dagger c_{x\sigma} + U \sum_x (1 - c_{x\uparrow}^\dagger c_{x\uparrow})(1 - c_{x\downarrow}^\dagger c_{x\downarrow}) \\ &= - \sum_{x,y,\sigma} t_{xy} c_{x\sigma}^\dagger c_{y\sigma} + U \sum_x c_{x\uparrow}^\dagger c_{x\downarrow}^\dagger c_{x\downarrow} c_{x\uparrow} + U(|V| - N_e) \end{aligned} \quad (9)$$

$|V|$ is the number of vertices.

Thus, the particle-hole transformation is not a symmetry, but it can be used to obtain eigenstates from other eigenstates.

For a bipartite lattice, *i.e.*, a lattice that splits into two sub-lattices A and B so that $t_{xy} = 0$ if both x and y belong to the same sub-lattice, it is possible to introduce the following transformation:

$$c_{x\sigma}^\dagger \rightarrow c_{x\sigma}^\dagger \text{ if } x \in A, \quad c_{x\sigma}^\dagger \rightarrow -c_{x\sigma}^\dagger \text{ if } x \in B \quad (10)$$

This transformation changes the sign of the kinetic energy. Applying this transformation together with the particle-hole transformation at half filling (*i.e.* $N_e = |V|$) maps the Hamiltonian onto itself. Thus, we have another symmetry for this class of lattices, a particle-hole symmetry. The transformation (10) alone is of some importance since it can be used to change the sign of the hopping matrix elements. Typically, it is assumed that $t_{xy} < 0$ is the natural choice of the sign, at least for nearest neighbors. For bipartite lattices the sign can be changed. In general, the assumption $t_{xy} < 0$, although popular, has no compelling reason [8].

On bipartite lattices at half filling, one can use the particle-hole symmetry to obtain a second $SU(2)$ symmetry. The generators are

$$\hat{S}_z = \frac{1}{2}(N_e - |V|), \quad \hat{S}_+ = \sum_{x \in A} c_{x\uparrow}^\dagger c_{x\downarrow}^\dagger - \sum_{x \in B} c_{x\uparrow}^\dagger c_{x\downarrow}^\dagger, \quad \hat{S}_- = \hat{S}_+^\dagger \quad (11)$$

These generators can be obtained from the original $SU(2)$ generators by performing a particle-hole transformation together with a transformation of type (10) for spin down only. The model

has thus a $SU(2) \times SU(2) = SO(4)$ symmetry at half filling. In discussions concerning high temperature superconductivity, even an approximate $SO(5)$ -symmetry has been proposed. The additional symmetry of the Hubbard model on a bipartite lattice at half filling ($N_e = |V|$) is essential for several of the rigorous results that are valid in this case. The two most important are Lieb's theorem [10], see Sect. 3.1, and the uniform density theorem, Sect. 3.5.

Lattice symmetries: On translationally invariant lattices, the model has the symmetries of the lattice.

The one-dimensional case: The one dimensional Hubbard model has an infinite set of invariants. A special form of the Bethe ansatz yields exact eigenstates of the Hamiltonian. This was first shown by E. Lieb and F. Wu [11]. The ground state is part of these Bethe ansatz eigenstates. Not all the eigenstates of the one-dimensional Hubbard model are Bethe ansatz states. But it was shown by Essler, Korepin, and Schoutens [12], that for even $|V|$, where the lattice is bipartite, all other eigenstates can be obtained by applying the operators $\hat{S}_{\pm}$ to the Bethe ansatz states.

For an exactly solvable model one should expect an infinite set of invariants. A first attempt to find those is a paper by Heilmann and Lieb [13]. Later Shastry [14] and Grosse [15] presented a large set of such invariants.

The one-dimensional Hubbard model has been investigated by many people, the literature is vast, and a complete overview would be a course in its own. I will not discuss the one-dimensional Hubbard model in this course.

3 Some rigorous results

Most of the rigorous results on the Hubbard model concern the magnetic behavior in the ground state, *i.e.*, at $T = 0$. I discuss the most important rigorous results in the following subsections. For each of the theorems mentioned below I try to explain the main idea of the proof. For the mathematical details I refer to the original papers.

3.1 Lieb's Theorem

In 1989, E. Lieb [10] proved an important theorem and an even more important corollary on the Hubbard model. The theorem is about the attractive Hubbard model. It holds for arbitrary hoppings t_{xy} , with the only assumption that the graph of the hopping matrix is connected. The interaction U_x may depend on x .

Theorem (Lieb 1989) Let H be the Hamiltonian in (1) with real t_{xy} , the graph of $T = (t_{xy})$ should be connected, and negative $U_x < 0$. Let the particle number N_e be even. Then, the ground state is unique and has a total spin $S = 0$.

For the proof, I refer to the original paper by Lieb. He uses a technique called spin reflection positivity. For some details see the remarks below.

On a bipartite lattice, using a particle-hole transformation for spin-down only, together with a transformation (10), the kinetic energy remains the same but the signs of U_x are switched. In that way, one can obtain a result for the attractive Hubbard model. Since S_z transforms with the above transformation to $\hat{S}_z$, one obtains a result for $\hat{S}_z = 0$, *i.e.*, $N_e = |V|$, *i.e.*, half filling. Therefore, the following corollary holds

Corollary Let H be the Hamiltonian in (1) with real t_{xy} . The graph of t_{xy} should be connected and bipartite, and positive $U_x = U > 0$. Let the particle number $N_e = |V|$. Then, the ground state is unique in the subspace $S_z = 0$. The total spin is $S = \frac{1}{2}||A| - |B||$.

The last statement $S = \frac{1}{2}||A| - |B||$ does not follow directly from the theorem because the theorem makes no statement about $\hat{S}$. It can be understood in two ways.

The first is to look at weak interactions. For a bipartite lattice, the spectrum of $T = (t_{xy})_{xy \in V}$ is symmetric with respect to 0. For any eigenvalue ε there exists an eigenvalue $-\varepsilon$. Half filling now means that for arbitrary weak interaction all single particle eigenstates with energies $\varepsilon < 0$ are completely filled with two electrons and that the eigenstates with $\varepsilon = 0$ are filled with one electron. For the latter, Hund's rule [16] applies, which means that all electrons have the same spin. The degeneracy of the eigenvalue 0 is $||A| - |B||$, therefore we obtain the a total spin $S = \frac{1}{2}||A| - |B||$.

The second idea is to look at strong interactions and to use a unitary transformation $\exp(R)$ to transform the Hamiltonian to a form where the number of doubly occupied sites is conserved. The ansatz is

$$R = \sum_{x,y,\sigma} r_{x,y,\sigma} c_{x,\sigma}^\dagger c_{y,\sigma} \quad (12)$$

We assume that U is large and expand $\exp(R)H\exp(-R)$ to obtain

$$H \rightarrow H_{\text{int}} + H_{\text{kin}} + [R, H_{\text{int}}] + [R, H_{\text{kin}}] + \frac{1}{2}[R, [R, H_{\text{int}}]] + \dots \quad (13)$$

The kinetic energy H_{kin} can be written as

$$H_{\text{kin}} = H_{\text{kin},0} + H_{\text{kin},1} \quad (14)$$

$H_{\text{kin},0}$ does not change the number of doubly occupied sites, $H_{\text{kin},1}$ changes it by ± 1 . We have

$$H_{\text{kin},1} = \sum_{x,y,\sigma} t_{xy} (n_{x,-\sigma} - n_{y,-\sigma})^2 c_{x,\sigma}^\dagger c_{y,\sigma} \quad (15)$$

We choose R so that

$$H_{\text{kin},1} + [R, H_{\text{int}}] = 0 \quad (16)$$

This yields

$$H \rightarrow H_{\text{eff}} = H_{\text{int}} + H_{\text{kin},0} - \frac{1}{2}[R, [R, H_{\text{int}}]] + \dots \quad (17)$$

We have

$$[R, H_{\text{int}}] = -U \sum_{x,y,\sigma} r_{x,y,\sigma} (n_{x,-\sigma} - n_{y,-\sigma}) c_{x,\sigma}^\dagger c_{y,\sigma} \quad (18)$$

and therefore

$$r_{x,y,\sigma} = \frac{t_{xy}}{U} (n_{x,-\sigma} - n_{y,-\sigma}) \quad (19)$$

Let P_0 be the projector onto states for which each site is occupied by one electron, which is the ground state at half filling and U arbitrarily large. If we restrict the Hilbert space to these states, we get

$$\begin{aligned} H_{\text{eff}} &= P_0 R H R P_0 \\ &= U P_0 R^2 P_0 \\ &= U P_0 \sum_{x,y,\sigma} \frac{t_{xy}}{U} (n_{x,-\sigma} - n_{y,-\sigma}) c_{x,\sigma}^\dagger c_{y,\sigma} \sum_{x',y',\sigma'} \frac{t_{x',y'}}{U} (n_{x',-\sigma'} - n_{y',-\sigma'}) c_{x',\sigma'}^\dagger c_{y',\sigma'} P_0 \\ &= -\frac{1}{U} P_0 \sum_{x,y,\sigma,\sigma'} t_{xy}^2 c_{x,\sigma}^\dagger c_{y,\sigma} c_{y,\sigma'}^\dagger c_{x,\sigma'} P_0 \\ &= \frac{1}{U} P_0 \sum_{x,y,\sigma,\sigma'} t_{xy}^2 c_{x,\sigma}^\dagger c_{y,\sigma} c_{y,\sigma'}^\dagger c_{x,\sigma} P_0 - \frac{1}{U} \sum_{x,y} t_{xy}^2 \\ &= \sum_{x,y} \frac{2t_{xy}^2}{U} \mathbf{S}_x \cdot \mathbf{S}_y P_0 + \frac{1}{U} P_0 \sum_{x,y,\sigma,\sigma'} t_{xy}^2 c_{x,\sigma}^\dagger c_{x,\sigma} c_{y,\sigma'}^\dagger c_{y,\sigma'} P_0 - \frac{1}{U} \sum_{x,y} t_{xy}^2 \\ &= \sum_{x,y} \frac{2t_{xy}^2}{U} \mathbf{S}_x \cdot \mathbf{S}_y P_0 \end{aligned} \quad (20)$$

This transformation is of importance on its own. It shows that the Hubbard model at half filling and for large U can be mapped to the anti-ferromagnetic Heisenberg model. For the corollary above it has the consequence that the total spin of the Hubbard model at half filling and for large U is the same as for the Heisenberg model, therefore $S = \frac{1}{2}||A| - |B||$.

Since the ground state is unique for all U , it is sufficient to know the total spin S for small or large U : due to uniqueness it cannot change.

Lieb's theorem suggests anti-ferromagnetism or ferrimagnetism (depending on whether the two sub-lattices A and B have the same size or not) for the Hubbard model at half filling. But, whereas for the anti-ferromagnetic Heisenberg model long-range order was proven in two dimensions in the ground state and in three dimensions for sufficiently low temperatures, there is no proof for long-range order for the Hubbard model up to now. The methods to prove long-range order for the Heisenberg model cannot be applied to the Hubbard model. The simple reason is that the Hubbard model is much more complicated and allows for a wider variety of phenomena. Nevertheless, many results including those from renormalization (see Sect. 4) indicate the existence of long range order for large U .

Lieb's proof uses the fact that for an even number of fermions, there is always a ground state with $S_z = 0$ due to the $SU(2)$ spin symmetry. This means that the ground state can be written in the form $\psi = \sum_{\alpha,\beta} W_{\alpha\beta} \psi_{\alpha,\uparrow} \psi_{\beta,\downarrow}$ where $\psi_{\alpha,\sigma}$ form an orthonormal basis of multi-particle states

with $N_e/2$ particles with spin σ . Since t_{xy} and U_x are real, one can assume that the matrix W is self adjoint. The expectation value of the Hamiltonian in the state ψ can be written in a quadratic form $E(W)$ in W and it can be shown that for non-positive interactions $E(W) \geq E(|W|)$. $|W|$ is the positive semi-definite matrix satisfying $W^2 = |W|^2$. It is then easy to see that the ground state corresponding to $|W|$ has $S = 0$. Uniqueness is shown by assuming that a second ground state with some W exists. Then, $R = |W| - W$ is a ground state as well. A lengthy but easy to understand argument that uses the fact that the graph of T is connected then shows that $W = \pm|W|$ and therefore that the ground state is unique. Compared to many other proofs, Lieb's proof is very elegant and compact, only somewhat more than one page in a letter. I recommend that everyone read it.

3.2 The Mermin-Wagner theorem

The term Mermin-Wagner Theorem is usually used for a huge class of theorems that state that for lattice models in one or two spacial dimensions with a continuous symmetry, like an $SU(2)$ symmetry, there is no long range order at finite temperature. Originally, Mermin and Wagner [17] showed in 1966 that in the one- or two-dimensional Heisenberg model there is no long-range order, neither anti-ferromagnetic nor ferromagnetic. This result was extended to the Hubbard model by Walker and Ruijgrok in 1968 [18] and by Ghosh in 1971 [19]. Further, Hohenberg [20] showed in 1967 that there cannot be superconductivity or long range crystalline order in one or two dimensions. The proof for the Hubbard model was considerably simplified and somewhat extended by Koma and Tasaki [21].

Theorem (Koma, Tasaki 1992) For a Hubbard model in one and two dimensions with finite-ranged hopping (*i.e.* $t_{xy} = 0$ if the distance $|x - y|$ lies above some finite value) in the thermodynamic limit, the following bounds hold for the correlation functions

$$|\langle c_{x\uparrow}^\dagger c_{x\downarrow}^\dagger c_{y\downarrow} c_{y\uparrow} \rangle| \leq \begin{cases} |x - y|^{-\alpha f(\beta)} & \text{for } d = 2 \\ \exp(-\gamma f(\beta)|x - y|) & \text{for } d = 1 \end{cases} \quad (21)$$

$$|\langle \mathbf{S}_x \cdot \mathbf{S}_y \rangle| \leq \begin{cases} |x - y|^{-\alpha f(\beta)} & \text{for } d = 2 \\ \exp(-\gamma f(\beta)|x - y|) & \text{for } d = 1 \end{cases} \quad (22)$$

for some $\alpha > 0$, $\gamma > 0$, $f(\beta) > 0$ where $\langle \dots \rangle$ denotes the expectation value at inverse temperature β and $f(\beta)$ is a decreasing function of β that behaves like $f(\beta) \approx 1/\beta$ for $\beta \gg \beta_0$ and $f(\beta) \approx (2/\beta_0)|\ln(\beta)|$ for $\beta \ll \beta_0$, where β_0 is some constant.

This result rules out long-range spin-order or superconductivity at finite temperatures in one or two dimensions. The power laws for $d = 2$ are certainly not optimal for high temperatures, where one expects an exponential decay of correlation functions. But they are sufficient to exclude long-range order. This means that a Kosterlitz-Thouless transition may occur [22].

The interesting point of the proof is that it only needs a $U(1)$ symmetry. Thus, any lattice model with a $U(1)$ symmetry in one or two dimensions cannot have superconducting or magnetic long-range order at finite temperature in one or two dimensions.

The older proof of Ghosh [19] uses the $SU(2)$ spin symmetry and the Bogoliubov inequality and is easy to understand.

The result by Koma and Tasaki is more general; their proof uses a method developed by McBryan and Spencer [22] for classical spin systems and its extension to quantum spin systems developed by Ito [23]. The proof uses the fact that for an arbitrary observable A one has $\text{Tr}(A \exp(-\beta H)) = \text{Tr}(G(\theta) A G(\theta)^{-1} \exp(-\beta G(\theta) H G(\theta)^{-1}))$. $G(\theta)$ is a local transformation. The right-hand side can be bounded using some Schwartz inequality stating that for hermitian matrices O and P one has $\text{Tr}(OP) \leq (\text{Tr}(O^*O)\text{Tr}(P^*P))^{1/2}$, and the Golden-Symanzik-Thompson inequality $\text{Tr} \exp(O + P) \leq \text{Tr}(\exp(O) \exp(P))$. Suitable choices for A and $G(\theta)$ then yield the bounds.

In one dimension, with nearest-neighbor hopping only, and for finite U_x , the Lieb-Mattis theorem [24] says that the minimal energy in the subspace with fixed spin S is strictly lower than the minimal energy in the subspace with $S + 1$. This clearly rules out ferromagnetism in one dimension in the ground state.

3.3 Nagaoka's theorem

The so called Nagaoka Theorem was actually first proven by Thouless [25] 1965 for some special bipartite lattices. The proof of Nagaoka [26], only one year later, is more general and applies to non-bipartite lattices as well. Therefore, the result is called the Nagaoka theorem today. The most general proof is due to Tasaki [27]. It states the following:

Theorem (Tasaki 1989) The Hubbard model (1) with non-negative t_{xy} , $N_e = |V| - 1$, and a hard-core repulsion $U_x = \infty$ for all $x \in V$ has a ground state with a total spin $S = \frac{1}{2}N_e$. The ground state is unique except for the usual $(2S + 1)$ -fold spin degeneracy provided a certain connectivity condition for t_{xy} holds.

This theorem is remarkable, because it states that there is a unique ferromagnetic ground state in the vicinity of half filling, where an anti-ferromagnetic spin order is assumed to be present. The proof of the theorem uses the Schwarz inequality to show that a ferromagnetic ground state exists. To show uniqueness, it uses the Perron-Frobenius theorem, which states that for a matrix with only non-negative entries and for which the graph is connected (the matrix is irreducible), the eigenstate with the largest eigenvalue is unique and has non-negative entries. The theorem can be applied here by finding a suitable basis for the multi-particle Hilbert space of the Hubbard model. The connectivity condition in the theorem ensures that the graph of the Hamiltonian in that basis obeys the irreducibility needed in the Perron-Frobenius theorem. Essentially it states that through the hopping of particles, arbitrary permutations of the particles can be realized. This holds for almost any lattice except the one-dimensional chain.

The Nagaoka theorem made people believe that for many lattices, *e.g.*, also for hyper-cubic lattices, a large region in the parameter space (U large and a density close to but not at half filling) exists where the Hubbard model has ferromagnetic ground states. But any attempt to prove that has failed so far. Instead, many variational calculations by various groups mainly

in the early '90s showed that the Nagaoka state is not very stable. Changing the conditions a bit, either putting more than one hole in the system or lowering U causes the Nagaoka state to become unstable against single spin flips, *i.e.*, $E(S = N_e/2 - 1) < E(S = N_e/2)$ (for details see *e.g.* [28]). Exact diagonalization of small systems yields the same result. On the other hand, for some special non-bipartite lattices, these calculations indicate that the Nagaoka state may be more stable and that a larger region in the parameter space exists where the ground state is ferromagnetic.

3.4 Flat-band systems

A first example of a lattice having a flat band is a bipartite lattice with $|A| = n|B|$. A simple example which Lieb [10] used as an illustration for his theorem is the quadratic lattice with additional lattice sites on each edge. If there is only nearest-neighbor hopping, the original lattice sites of the quadratic lattice form one of the sub-lattices, say B and the new lattice sites form the second sub-lattice A . There are twice as many lattice sites on A as on B . Each elementary cell contains one lattice site from B and two from A , in total three. We have thus a three-band model. Since the lattice is bipartite, the single-particle spectrum is symmetric with respect to 0. There is one energy band in the center, which is completely flat. The flat band causes the extensive magnetization $S = \frac{1}{2}||A| - |B|| = \frac{1}{2}|B|$, as we pointed out already in Sect. 3.1. Since this magnetization is related to the existence of two sub-lattices, the system is ferrimagnetic.

Two years after Lieb, first examples of lattices with a flat band at the bottom of the spectrum were published [29–33]. One class of such lattices are line graphs, the other are decorated lattices. Since the construction of a line graph is elementary and since we need it later, we give a more detailed description here.

Let $G = (V, E)$ be a graph with a vertex set V and an edge set E . Any lattice can be regarded as a graph. The lattice sites are the vertices and there are edges between two vertices if there is a non-vanishing hopping matrix element connecting the two. If we allow only for nearest neighbor hopping, the hopping matrix is (up to a factor t) the adjacency matrix $A(G) = (a_{xy})_{xy \in V}$ of the graph. $a_{xy} = 1$ if $\{x, y\} = e \in E$ is an edge of the graph, 0 otherwise.

The line graph $L(G)$ of a graph G is constructed as follows: The vertex set $V(L(G))$ of the line graph is the edge set $E(G)$ of the original graph and two vertices of the line graph are connected, if the corresponding edges in G have a vertex in common.

Figure (1) shows an illustration of the construction of a line graph. Let G be a part of the hexagonal lattice, as shown in black. Now we put a new vertex in the middle of each edge and connect two new vertices if the edges of the original hexagonal lattice have a vertex in common. This procedure yields a new lattice built of hexagons surrounded by triangles, shown in red. The new lattice constructed that way is the line graph of the hexagonal lattice, it is called the kagome lattice. You may take any lattice or even any graph G and construct the line graph in that way.

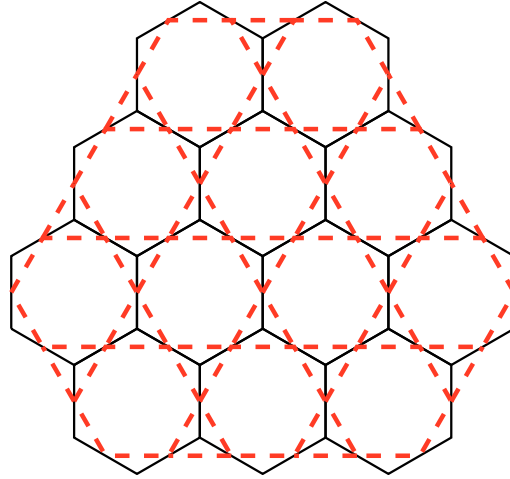


Fig. 1: The kagome lattice (red, dashed) as the line graph of the hexagonal lattice (black).

Let us now investigate the spectral properties of the adjacency matrix $A(L(G))$ of a line graph. To do this, we first introduce a new matrix $B(G) = (b_{xe})_{x \in V, e \in E}$, the so called edge vertex incidence matrix. The matrix elements $b_{xe} = 1$ if the edge e connects to the vertex x , $b_{xe} = 0$ otherwise. Note that B has $|V|$ columns and $|E|$ rows. Except for a graph without loops or with only one loop, $|E| > |V|$ and the kernel of B has a dimension $\geq |E| - |V|$. The adjacency matrix of the line graph and the incidence matrix of the original graph are related via $A(L(G)) = B(G)^T B(G) - 2$. As a consequence, -2 is a lower bound of the spectrum of $A(L(G))$ and becomes the lowest eigenvalue with degeneracy at least $|E| - |V|$ if $|E| > |V|$. In fact, one can show that the degeneracy is $N_d = |E| - |V| + 1$ if G is bipartite and connected, $N_d = |E| - |V|$ if G is not bipartite and connected.

This fact can now be applied to a lattice. If G is a translationally invariant lattice with one or more energy bands, $L(G)$ is a lattice as well and the lowest energy band lies at energy $-2t$ and is completely flat. A lattice that is a line graph, *e.g.* the kagome lattice, has a lowest flat band. This makes it easy to construct ground states of the Hubbard model, at least for $N_e \leq N_d$. In that case, any state with all electrons having the same spin is a ground state, since this state minimizes both the kinetic energy and the interaction. This construction is indeed trivial. The interesting question is whether there are other ground states and whether they can be characterized completely. This is indeed possible for $N_e = N_d$ as the following theorem shows [29, 30]:

Theorem (Mielke 1991) Let H be the Hubbard model on a line graph $L(G)$ of a two-connected bipartite graph G or a three-connected graph G and let $N_e = N_d$, and $U_x > 0$ for all x . Then the ground state has a spin $S = \frac{N_d}{2}$ and is unique up to the $(2S+1)$ -fold degeneracy due to the $SU(2)$ spin symmetry.

The kagome lattice is obviously an example for this theorem.

The original proof of the theorem uses some graph theoretical notions. We will not present it here since later a more general and simpler result has been shown which does not use the notion of a line graph.

On the other hand, let us discuss the single-particle ground states with energy $-2t$ a bit further since they serve for many easy illustrations we may need later. Let p be a self-avoiding closed path $(x_1, x_2, \dots, x_n)$ of even length n on G . It obviously translates to an even path on $L(G)$. Let us now construct the single-particle state $\psi_p(e)$ as follows. $\psi_p(e) = 0$ if e lies not on p . On p , $\psi_p(e) = \pm 1$ with alternating sign for subsequent edges of G . It is easy to see that $B\psi_p = 0$. ψ_p is therefore a ground state of $A(L(G))$ with eigenvalue -2 . It can be shown that these states form an over-complete basis of the eigenspace of the eigenvalue -2 .

If G is a bipartite plane graph, like the hexagonal lattice, each face f is surrounded by a self-avoiding path, let us call it f as well. Let F be the set of faces. Due to Euler's theorem, $|F| = |E| - |V| + 2$. One of the faces is the outer face of the graph, there are exactly $|E| - |V| + 1$ inner faces. It is easy to see that the states ψ_f corresponding to the inner faces f are linearly independent. They thus form a basis (not orthonormal) of the eigenspace of the ground state energy. Using this construction, it is possible to construct all ground states for $N_e \leq N_d$.

One year later, 1992, Tasaki [32] published a class of decorated lattices with lowest flat bands, for which he proved a similar result. In 1993 [33], we investigated these lattices further and showed how one can construct all ground states with $N_e \leq N_d$ for these decorated lattices. We further showed that for these lattices the characterization of the ground states can be mapped to a percolation problem. This allows us to show that the system remains ferromagnetic with an extensive but not saturated total spin S up to some critical density. Below that density the system is paramagnetic.

This construction is most easily understood for line graphs of planar bipartite graphs, see the kagome lattice in Fig. 1 as an example. For these graphs, the faces yield the single-particle ground state. The inner faces form a basis (not orthogonal). Neighboring faces of G have an edge in common. Therefore, putting electrons with different spin on neighboring faces may produce a double occupancy on that edge. This yields a higher energy. To obtain a ground state, electrons on neighboring faces should have the same spin. But if $N_e < N_d$ not all faces are occupied and one may form non-touching clusters with total spins pointing in different directions. Constructing non-touching clusters of faces is a percolation problem on the dual graph of G . This percolation problem has a percolation threshold, above which one large extended cluster is formed. This cluster has an extensive spin, whereas all other finite clusters have a finite spin. Therefore, above the percolation threshold the system is ferromagnetic. The percolation problem is not the classical percolation problem since each cluster with a spin S has a $2S + 1$ -fold degeneracy.

Since 1993, more classes of lattices with flat bands have been found and investigated. A general result, which covers all these cases, is available as well. It needs a condition on the projector $\rho = (\rho_{xy})_{x,y \in V}$ onto the space of single-particle ground-states [16, 34, 35].

Theorem (Mielke 1993, 1999) The Hubbard model with an N_d -fold degenerate single-particle ground state, $U_x > 0$, and $N_e \leq N_d$ electrons has a unique $(2S + 1)$ -fold degenerate ferromagnetic ground state with $S = N_d/2$ if and only if $N_e = N_d$ and ρ_{xy} is irreducible.

The original proof of this theorem was complicated and used a special construction for a non-orthonormal basis of single-particle ground-states. The later proof is simpler. First, the following result is shown:

Theorem (Mielke 1999) The Hubbard model with an N_d -fold degenerate single-particle ground-state, $U_x > 0$, and $N_e \leq N_d$ electrons has a multi-particle ground state with $S < N_e/2 - 1$ if it has a single spin-flip ground state with $S = N_e/2 - 1$.

In other words: To prove stability of ferromagnetism, it is sufficient to show that there is no single spin-flip ground state. This is indeed easy if $N_e = N_d$ and ρ_{xy} is irreducible. Therefore, the first theorem is a consequence of the second. Note that the second theorem is not trivial at all. For other lattices, you may easily construct cases where a ferromagnetic state is stable with respect to single spin flips but where it nevertheless is not the ground state of the system.

The last two results are very general; they hold for arbitrary lattices and arbitrary, even complex, hopping matrix elements t_{xy} . This is important because the flat band physics started to attract much attention in the past few years for mainly two reasons. First, using optical lattices it is now possible to investigate these systems experimentally. For instance the kagome lattice was build using that technique by Jo *et al.* in 2012 [36]. Second, people became interested recently in so-called topological flat bands. Here, the flat (often quasi flat) band arises from special choices for the phases of complex t_{xy} .

In 2003, Tanaka and Ueda [37] showed that for the special case of the kagome lattice, the ferromagnetic ground states remains stable if one introduces a special perturbation that yields a small dispersion to the lowest flat band, provided U is not too small. Similarly, Tasaki [38] showed in 1996 that for some decorated lattices the ferromagnetic ground state remains stable. These results are important because they indicate that flat band ferromagnetism is not something exotic like the Nagaoka ferromagnet.

Another interesting question is what happens if ρ_{xy} is not irreducible. Batista and Shastry [39] were the first to investigate an example for such a lattice; today many examples are known. One can show the following general results [40]:

Let ρ have the following properties:

1. ρ is reducible. It can be decomposed into N_r irreducible blocks ρ_k , $k = 1, \dots, N_r$. N_r should be an extensive quantity, *i.e.*, $N_r \propto N_d \propto |V|$, so that in the thermodynamic limit the density of degenerate single-particle ground states and the density of irreducible blocks are both finite.
2. Let V_k be the support of ρ_k , *i.e.*, the set of vertices for which at least one element of ρ_k does not vanish. $\rho_{k,xy} = 0$ if $x \notin V_k$ or $y \notin V_k$. One has $V_k \cap V_{k'} = \emptyset$ if $k \neq k'$ because of the fact that ρ_k are irreducible blocks of the reducible matrix ρ and $\bigcup_k V_k \subseteq V$.
3. We choose the basis B such that the support of each basis state $\psi_i(x) \in B$ is a subset of exactly one V_k . We denote the number of states belonging to the cluster V_k as ν_k . One has $\sum_k \nu_k = N_d$.
4. $\nu_{\max} = \max_k \{\nu_k\}$ is $O(1)$, *i.e.*, not an extensive quantity.

With these properties one can show

Theorem (Mielke 2012) For Hubbard models with a lowest single-particle eigenenergy 0 which is N_d -fold degenerate and for which the projector onto the eigenspace of 0 fulfils the properties listed above, the following results hold for $N_e \leq N_d$:

1. The ground state energy is 0.
2. Let A_x be an arbitrary local operator, *i.e.*, an arbitrary combination of the four creation and annihilation operators $c_{x\sigma}^\dagger$ and $c_{x\sigma}$. The correlation function $\rho_{A,xy} = \langle A_x A_y \rangle - \langle A_x \rangle \langle A_y \rangle$ has a finite support for any fixed x and vanishes if x and y are out of different clusters V_k . The system has no long-range order.
3. The system is paramagnetic.
4. The entropy at zero temperature $S(c)$ is an extensive quantity, $S(c) = \mathcal{O}(N_e)$. It increases as a function of $c = N_e/N_d$ from 0 for $c = 0$ to some maximal value $S_{\max} \geq \sum_k [(\nu_k - 1) \ln 2 + \ln(\nu_k + 2)]$ and then decays to $S(1) = \sum_k \ln(\nu_k + 1)$.

These models have therefore no long-range order. The most interesting aspect is the finite entropy at zero temperature.

3.5 Uniform density theorem

The uniform density theorem [41–43] is valid on a bipartite lattice and at half filling. The proof makes use of the particle-hole symmetry which is valid in that case. It states

Theorem (MacLachlan 1959, 1961; Lieb, Loss, McCann 1993) For the Hubbard model on a bipartite lattice and at half filling, either in a canonical ensemble with $N_e = |V|$ at $T > 0$ or in the ground state at $T = 0$ or in a grand canonical ensemble with $\mu = 0$, the density matrix $\rho_{\sigma,xy} = \langle c_{x\sigma}^\dagger c_{y\sigma} \rangle$ has the property

$$\rho_{\sigma,xy} = \frac{1}{2} \delta_{xy} \quad \text{if } x, y \in A \text{ or } x, y \in B \quad (23)$$

The theorem may appear to be trivial if one has a translationally invariant lattice in mind. The point is, it holds for arbitrary t_{xy} and arbitrary U_x on an arbitrary bipartite graph, translational invariance is not used and not necessary. The theorem is therefore of large importance in quantum chemistry, *i.e.*, for the Pariser-Parr-Pople variant of the Hubbard model.

3.6 Further rigorous results

There are further rigorous results on the Hubbard model. Many of them deal with the absence of ferromagnetism or at least with the absence of a fully polarized ground state under certain conditions. For details I refer to [8, 9, 44].

4 (Functional) Renormalization

4.1 General idea

First ideas on renormalization were developed in the '50s and '60s, mainly in the context of field theory, especially QED. The main idea is to separate different scales, often energy scales. If one is interested in the physics at low temperature, for example, one wants to separate degrees of freedom at higher energies from those at lower energies. Typically, this is done in many discrete steps or in a continuous form, either by integrating out the higher degrees of freedom or by separating them from the lower degrees of freedom. In that way, new effective interactions between the lower degrees of freedom are generated. If one is able to control them in some way, one can derive an effective theory at low energies.

Today, two ways have been used to apply this general idea to the Hubbard model. One idea is to use continuous unitary transformations to (block-) diagonalize the Hamiltonian of the system. This approach was developed independently by Wegner [45] and by Glazek and Wilson [46] and was applied to the Hubbard model by Wegner and coworkers [47–49]. The continuous unitary transformation brings the Hamiltonian to a block-diagonal form where in each block the number of quasi-particles is conserved. The transformation creates new interactions between the quasi-particles. These interactions, if they become strong, can cause instabilities of the Fermi liquid. A stability analysis shows which kinds of long-range order may occur.

The second, older, and more popular method is to use a field theoretic formulation and to integrate out higher degrees of freedom. Here as well, effective interactions occur and a stability analysis shows which instabilities occur.

In this section, we explain briefly the field theoretic renormalization. For a detailed and mathematical introduction I refer to the book of Salmhofer [50]. For a recent review I refer to [51].

4.2 Field-theoretic representation of the Hubbard model

In the following we want to deal with a system of interacting particles. Let us assume that we have a usual two-particle interaction. The single-particle contribution typically contains the kinetic energy and some single-particle potential. We assume that this contribution can be diagonalized, and we choose the basis for the representation of the Hamiltonian such that it is diagonal. Furthermore, we assume that we have a finite system where the single-particle energies ε_i are discrete. Eventually, we may take the thermodynamic limit. Then, the Hamiltonian is of the form

$$\hat{H} = \sum_i \varepsilon_i c_i^\dagger c_i + \sum_{i,j,k,l} V_{i,j,k,l} c_i^\dagger c_j^\dagger c_l c_k \quad (24)$$

Having such a system, one typically wants to calculate expectation values of some operators $A = A(\{c_i^\dagger, c_i\})$. At finite temperatures, they are

$$\langle A(\{c_i^\dagger, c_i\}) \rangle = Z^{-1} \text{Tr} \left[A(\{c_i^\dagger, c_i\}) \exp(-\beta(\hat{H} - \mu\hat{N})) \right] \quad (25)$$

where $\beta = 1/T$ is the inverse temperature, μ is the chemical potential and

$$Z = \text{Tr} \exp(-\beta(\hat{H} - \mu\hat{N})) \quad (26)$$

is the grand canonical partition function of the system. The traces are calculated over the entire Fock space. We now use the standard way to obtain a field-theoretic representation for the grand canonical partition function. For a general introduction see the book of Negele and Orland [52]. It uses two main ingredients:

1. Coherent states for fermions, constructed with the help of Grassmann variables ξ_i . A coherent state for Fermions is defined as an eigenstate of the annihilation operators and has the form

$$|\xi\rangle = \exp\left(-\sum_i \xi_i c_i^\dagger\right) |\text{vac.}\rangle = \prod_i (1 - \xi_i c_i^\dagger) |\text{vac.}\rangle \quad (27)$$

These states form an over-complete basis with the completeness relation

$$\int D[\xi] \exp\left(-\sum_i \xi_i^* \xi_i\right) |\xi\rangle \langle \xi| = 1 \quad (28)$$

2. The representation

$$\exp\left(-\beta(\hat{H} - \mu\hat{N})\right) = \left[\exp\left(-\frac{\beta}{M}(\hat{H} - \mu\hat{N})\right)\right]^M \quad (29)$$

together with the fact that for large M , *i.e.*, small $\varepsilon = \beta/M$ one has

$$\exp\left(-\varepsilon(\hat{H} - \mu\hat{N})\right) =: \exp\left(-\varepsilon(\hat{H} - \mu\hat{N})\right) : + \mathcal{O}(\varepsilon^2) \quad (30)$$

The symbols $: \cdot :$ denote normal ordering, which means that all creation operators stand left of all annihilation operators. We need the normal ordering to calculate matrix elements of operators between two coherent states. If the operator is normal-ordered, we obtain its matrix element simply by replacing the creation and annihilation operators by the corresponding Grassmann variables.

These two ingredients are used to write the partition function in the form

$$Z = \lim_{M \rightarrow \infty} \int D[\xi] \exp\left(-S_M[\xi]\right) \quad (31)$$

$$S_M[\xi] = \varepsilon \sum_{k=1}^M \left[\sum_i \xi_{i,k}^* \left(\frac{\xi_{i,k} - \xi_{i,k-1}}{\varepsilon} - \mu \xi_{i,k} \right) + H(\{\xi_{i,k}^*, \xi_{i,k-1}\}) \right] \quad (32)$$

where $\varepsilon = \beta/M$. Typically one introduces in the limit $M \rightarrow \infty$ the function $\xi_i(\tau)$, where $\xi_{i,k} = \xi_i(\varepsilon k)$. Then we may write

$$\frac{\xi_{i,k} - \xi_{i,k-1}}{\varepsilon} = \frac{\xi_i(\varepsilon k) - \xi_i(\varepsilon k - \varepsilon)}{\varepsilon} \rightarrow \frac{\partial \xi_i(\tau)}{\partial \tau} \quad (33)$$

$$\varepsilon \sum_{k=1}^M \rightarrow \int_0^\beta d\tau \quad (34)$$

and

$$S[\xi] = \int_0^\beta d\tau \left(\sum_i \xi_i^*(\tau) \left(\frac{\partial}{\partial \tau} - \mu \right) \xi_i(\tau) + H(\{\xi_i^*(\tau), \xi_i(\tau)\}) \right) \quad (35)$$

$$Z = \int_{\xi_i(\beta) = -\xi_i(0)} D[\xi] \exp(-S[\xi]) \quad (36)$$

Here and in the following I give only a brief overview, for details I refer to standard text books like [52].

To use this method for a renormalization of the Hubbard model, we formulate the Hubbard model in momentum space

$$H = \sum_{\vec{k}, \sigma} \varepsilon_{\vec{k}} c_{\vec{k}, \sigma}^\dagger c_{\vec{k}, \sigma} + \frac{1}{2} \sum_{k_1 \dots k_4, \sigma_1 \dots \sigma_4} V_{\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4} c_{\vec{k}_1, \sigma_1}^\dagger c_{\vec{k}_2, \sigma_2}^\dagger c_{\vec{k}_4, \sigma_4} c_{\vec{k}_3, \sigma_3} \quad (37)$$

Due to translational invariance, the single-particle Hamiltonian H_{kin} is diagonal. Further, we use the fact that the functions $\xi_i(\tau)$ are anti-periodic and use a Fourier representation ϕ with Fourier frequencies ω_n . The partition function now becomes

$$Z = \int D[\phi] \exp(S[\phi^*, \phi]) \quad (38)$$

$$S[\phi^*, \phi] = \sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \phi_K^* \phi_K - V[\phi^*, \phi] \quad (39)$$

where $K = (\omega_n, \vec{k}, \sigma)$ is a multi index, which contains the wave vector, the frequencies ω_n , and the spin. The interaction is still a generic interaction.

The physics of the system can be obtained by calculating expectation values, in particular propagators. This can be done using the generating functions. There are several ways to do that. A generating function that yields all connected propagators is

$$W[J^*, J] = \ln \left\langle \exp \left(-V[\phi^*, \phi] + \sum_K (J_K^* \phi_K + \phi_K^* J_K) \right) \right\rangle_0 \quad (40)$$

Here $\langle \dots \rangle_0$ denotes the expectation value in the non-interacting system, *i.e.*,

$$\langle A[\phi^*, \phi] \rangle_0 = \frac{\int D[\phi] A[\phi^*, \phi] \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \phi_K^* \phi_K \right)}{\int D[\phi] \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \phi_K^* \phi_K \right)} \quad (41)$$

The main idea of renormalization is simple. We introduce a cut-off Λ and perform all integrals in the expression for W over fields ϕ_K and ϕ_K^* for which $|i\omega_n - \varepsilon_{\vec{k}} + \mu| > \Lambda$. The remaining integrals can again be written as

$$W[J^*, J] = \ln \left\langle \exp \left(-V_\Lambda[\phi^*, \phi, J^*, J] + \sum_{K: |i\omega_n - \varepsilon_{\vec{k}} + \mu| < \Lambda} (J_K^* \phi_K + \phi_K^* J_K) \right) \right\rangle_{\Lambda, 0} \quad (42)$$

The non-interacting part now depends on Λ since the integration over some of the fields introduces new quadratic contributions. Since the original system is translationally invariant, the new quadratic contributions are translationally invariant and hence diagonal. Therefore, the average depends on Λ . The new interaction $V_\Lambda[\phi^*, \phi, J^*, J]$ also depends explicitly on Λ and on the fields J_K and J_K^* with $|i\omega_n - \varepsilon_{\vec{k}} + \mu| > \Lambda$. In the new integral, divergences are shifted, since the single-particle energies $\varepsilon_{\vec{k}}$ and eventually μ as well are shifted. Since we are interested in propagators near the Fermi surface, the fields J_K and J_K^* in $V_\Lambda[\phi^*, \phi, J^*, J]$ are uninteresting. We will never calculate derivatives with respect to these fields. Therefore we may set them to 0 explicitly.

In a next step, we introduce a new cut-off Λ_1 and perform the integral over all fields ϕ_K and ϕ_K^* for which $|i\omega_n - \varepsilon_{\vec{k}} + \mu| > \Lambda_1$. This yields again a new V_{Λ_1} and shifted single-particle energies. We iterate this procedure and obtain finally an effective theory that depends only on fields with small values of $|i\omega_n - \varepsilon_{\vec{k}} + \mu|$. This is now the effective theory we are looking for. *A priori* it is not clear that the procedure converges. For special cases and for not too large interaction, one can eventually prove convergence.

Let us mention that there are various technically different variants of the procedure described above. Instead of a discrete series of steps one can vary the cut-off continuously. Instead of a hard cut-off as described above, one can introduce a soft cut-off (these notions will be explained later). Instead of W one can use a different generating function. For Fermi systems it is often easier not to use W but

$$G_{\text{eff}}[\psi^*, \psi] = \ln \langle \exp(-V[\phi^* + \psi^*, \phi + \psi]) \rangle_0 \quad (43)$$

W and G_{eff} contain the same information and can be mapped to each other. To see that, we calculate

$$\begin{aligned} & Z_0 \langle \exp(-V[\phi^* + \psi^*, \phi + \psi]) \rangle_0 \\ &= \int D[\phi] \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \phi_K^* \phi_K - V[\phi^* + \psi^*, \phi + \psi] \right) \\ &= \int D[\phi] \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) (\phi_K^* - \psi_K^*) (\phi_K - \psi_K) - V[\phi^*, \phi] \right) \\ &= \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \psi_K^* \psi_K \right) \\ &\quad \times \int D[\phi] \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \phi_K^* \phi_K \right. \\ &\quad \left. - V[\phi^*, \phi] - \sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) (\phi_K^* \psi_K + \psi_K^* \phi_K) \right) \\ &= Z_0 \exp \left(\sum_K (i\omega_n - \varepsilon_{\vec{k}} + \mu) \psi_K^* \psi_K + W[C(K)^{-1} \psi_K^*, C(K)^{-1} \psi_K] \right) \end{aligned} \quad (44)$$

and therefore

$$G_{\text{eff}}[\psi^*, \psi] = \sum_K \psi_K^* C(K)^{-1} \psi_K + W[C(K)^{-1} \psi_K^*, C(K)^{-1} \psi_K] \quad (45)$$

where

$$C(K) = \frac{1}{i\omega_n - \varepsilon_{\vec{k}} + \mu} \quad (46)$$

Without going into the details, let us mention that since W is the generating function for connected propagators, one can show that G_{eff} is the generating function for connected, amputated propagators. Due to the factors $C(K)^{-1}$ in the argument of W , one multiplies the final result by a factor $C(K)^{-1} = (i\omega_n - \varepsilon_{\vec{k}} + \mu)$ for each derivative one takes, so that the factors $C(K)$ for the external lines are cancelled.

4.3 Renormalization group equations for G_{eff}

In this subsection we derive the renormalization group equation for the effective action. In contrast to the above description for W we will use a continuous renormalization. For computational and mathematical details we refer to [50].

Let us introduce the modified propagator

$$C^\Lambda(K) = \frac{\Theta_\Lambda(K)}{i\omega_n - (\varepsilon_{\vec{k}} - \mu)} \quad (47)$$

Here $\Theta_\Lambda(K)$ is a cut-off function. The most simple case would be a hard cut-off, *e.g.*,

$$\Theta_\Lambda(K) = \theta(|\varepsilon_{\vec{k}} - \mu - i\omega_n| - \Lambda) \quad (48)$$

For analytic calculations, a cut-off function that is differentiable might be more suitable. This is called a weak cut-off. For $|\varepsilon_{\vec{k}} - \mu - i\omega_n| \gg \Lambda$ one has $\Theta_\Lambda(K) = 1$ and $C^\Lambda(K) = C(K)$. For $|\varepsilon_{\vec{k}} - \mu - i\omega_n| \ll \Lambda$ one has $\Theta_\Lambda(K) = 0$ and therefore $C^\Lambda(K) = 0$. I define

$$G_{\text{eff}}^\Lambda[\psi^*, \psi] = \ln \frac{\int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K - V[\phi^* + \psi^*, \phi + \psi] \right)}{\int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K \right)} \quad (49)$$

For values of K with $\Theta_\Lambda(K) = 0$ only $\phi_K = 0$ contributes. G_{eff}^Λ is then given by $-V$. For values of K with $\Theta_\Lambda(K) = 1$, G_{eff}^Λ is given by G_{eff} . G_{eff}^Λ interpolates between $-V$ and G_{eff} . The goal is to derive a differential equation for G_{eff}^Λ which has the initial condition $-V$ and can be solved. We start with

$$F[\psi^*, \psi] = F \left[\frac{\partial}{\partial \eta}, \frac{\partial}{\partial \eta^*} \right] \exp \left(\sum_K (\eta_K^* \psi_K + \eta_K \psi_K^*) \right) \Big|_{\eta=\eta^*=0} \quad (50)$$

$$\begin{aligned} & \int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K + \sum_K (\eta_K^* \phi_K + \eta_K \phi_K^*) \right) \\ &= \exp \left(\sum_K \eta_K^* C^\Lambda(K) \eta_K \right) \int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K \right) \end{aligned} \quad (51)$$

The second equation can be written as

$$\begin{aligned} \left\langle \exp \left(\sum_K (\eta_K^* \phi_K + \eta_K \phi_K^*) \right) \right\rangle_{\Lambda,0} &= \frac{\int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K + \sum_K (\eta_K^* \phi_K + \eta_K \phi_K^*) \right)}{\int D[\phi] \exp \left(\sum_K \phi_K^* (C^\Lambda(K))^{-1} \phi_K \right)} \\ &= \exp \left(\sum_K \eta_K^* C^\Lambda(K) \eta_K \right) \end{aligned} \quad (52)$$

This yields

$$\begin{aligned} \exp(G_{\text{eff}}^\Lambda[\psi^*, \psi]) &= \langle \exp(-V[\phi^* + \psi^*, \phi + \psi]) \rangle_{\Lambda,0} \\ &= \exp \left(-V \left[\frac{\partial}{\partial \eta}, \frac{\partial}{\partial \eta^*} \right] \right) \left\langle \exp \left(\sum_K (\eta_K^* (\phi_K + \psi_K) + \eta_K (\phi_K^* + \psi_K^*)) \right) \right\rangle \Big|_{\eta=\eta^*=0} \\ &= \exp \left(-V \left[\frac{\partial}{\partial \eta}, \frac{\partial}{\partial \eta^*} \right] \right) \exp \left(\sum_K \eta_K^* C^\Lambda(K) \eta_K + \sum_K (\eta_K^* \psi_K + \eta_K \psi_K^*) \right) \Big|_{\eta=\eta^*=0} \\ &= \exp \left(-V \left[\frac{\partial}{\partial \eta}, \frac{\partial}{\partial \eta^*} \right] \right) \exp \left(\sum_K \frac{\partial}{\partial \psi_K} C^\Lambda(K) \frac{\partial}{\partial \psi_K^*} \right) \exp \left(\sum_K (\eta_K^* \psi_K + \eta_K \psi_K^*) \right) \Big|_{\eta=\eta^*=0} \\ &= \exp \left(\sum_K \frac{\partial}{\partial \psi_K} C^\Lambda(K) \frac{\partial}{\partial \psi_K^*} \right) \exp \left(-V \left[\frac{\partial}{\partial \eta}, \frac{\partial}{\partial \eta^*} \right] \right) \exp \left(\sum_K (\eta_K^* \psi_K + \eta_K \psi_K^*) \right) \Big|_{\eta=\eta^*=0} \\ &= \exp \left(\sum_K \frac{\partial}{\partial \psi_K} C^\Lambda(K) \frac{\partial}{\partial \psi_K^*} \right) \exp(-V[\psi_K^*, \psi_K]) \end{aligned} \quad (53)$$

and

$$\frac{\partial}{\partial \Lambda} \exp(G_{\text{eff}}^\Lambda[\psi^*, \psi]) = \sum_K \frac{\partial}{\partial \psi_K} \frac{\partial C^\Lambda(K)}{\partial \Lambda} \frac{\partial}{\partial \psi_K^*} \exp(G_{\text{eff}}^\Lambda[\psi^*, \psi]) \quad (54)$$

finally

$$\frac{\partial}{\partial \Lambda} G_{\text{eff}}^\Lambda[\psi^*, \psi] = \sum_K \frac{\partial}{\partial \psi_K} \frac{\partial C^\Lambda(K)}{\partial \Lambda} \frac{\partial G_{\text{eff}}^\Lambda[\psi^*, \psi]}{\partial \psi_K^*} + \sum_K \frac{\partial G_{\text{eff}}^\Lambda[\psi^*, \psi]}{\partial \psi_K} \frac{\partial C^\Lambda(K)}{\partial \Lambda} \frac{\partial G_{\text{eff}}^\Lambda[\psi^*, \psi]}{\partial \psi_K^*} \quad (55)$$

This is the differential equation for G_{eff}^Λ we were looking for. It is an exact renormalization group equation.

4.4 Numerical solutions

In general, it is not possible to solve the renormalization equation exactly. Furthermore, it is not at all clear that the results do not diverge. It is well known that for sufficiently low temperatures, a divergence occurs that leads to a superconducting instability. The Fermi liquid becomes a superconductor. This effect is called the Kohn-Luttinger effect. It was found by

Kohn and Luttinger [53] in a second order perturbative calculation in 1965 and can be treated in a mathematically rigorous way [50].

The exact flow equation (55) is the basis of various approximations. Approximations are mostly truncations of the effective action G_{eff} . In general, it contains arbitrary many-body terms. Most approximations truncate it to a two-particle term. Often, further truncations to special two-particle terms are used. It turns out that rather simple truncations are sufficient to describe the many-body phenomena one is interested in. Compared to other techniques, the functional renormalization group has many advantages. The most important is that it allows for a very precise description of low energy scales. For a detailed review of truncation schemes I refer to [51].

But even after a suitable truncation, the resulting equations for the expansion coefficients can often not be treated analytically. A numerical solution is possible, but needs a discretization of k -space. Often, one uses a discretization along the Fermi surface (in 2D, a line) and splits the k -space in sectors, each containing exactly one of the segments of the Fermi surface. First results applying such a truncation and discretization scheme were obtained by Halboth and Metzner [54,55]. Their approach neglected the frequency dependence. More recent approaches are more sophisticated, using a more elaborate truncation scheme or taking the frequency dependence into account.

4.5 Some results

Without going into details, we list some of the important instabilities found using renormalization theory, mainly in two dimensions. For a detailed review we refer to the article of Metzner *et al.* [51].

1. On the square lattice with nearest-neighbor hopping t , next-nearest-neighbor hopping t' , and (if not stated otherwise) a repulsive interaction $U > 0$ one finds several instabilities of the Fermi liquid:

- (a) Anti-ferromagnetism at or close to half filling ($\mu = 0$) for $t' = 0$.

This is consistent with the expectations for large U and with Lieb's theorem, see Sect. 3.1.

- (b) d -wave Cooper pairing at small negative values of t' and away from half filling.

The $d_{x^2-y^2}$ pairing is found in a large parameter region using various different truncation schemes. It is therefore now widely believed that the Hubbard model has a d -wave superconducting ground state at weak repulsive interactions $U > 0$.

The interplay between anti-ferromagnetism and d -wave Cooper pairing is interesting since it occurs as well in high-temperature cuprate superconductors.

- (c) A Pomeranchuk instability (an anisotropic deformation of the Fermi surface) leading to orientational symmetry breaking, at sufficiently large $|t'|$.

According to the uniform density theorem, there is no Pomeranchuk instability on a bipartite lattice at half filling, see Sect. 3.5

- (d) A ferromagnetic instability, which may occur if one varies t' and μ simultaneously so that the system stays at the van Hove singularity. At the van Hove singularity, the density of states is infinite. A large or infinite density of states at the Fermi level is believed to favor ferromagnetism, according to the Stoner criterion, which is a simple translationally invariant mean-field criterion.

Although there is no extensive degeneracy of states in the single-particle spectrum at the Fermi surface, this results is similar to flat-band ferromagnetism, see Sect. 3.4.

Let us mention that some of the functional renormalization schemes indicate a direct competition between ferromagnetism and superconductivity.

- (e) s -wave Cooper pairing at negative U .

2. On other two-dimensional lattices

- (a) Unconventional superconductivity or non-magnetic insulating states on the triangular lattice.

On the triangular lattice, the anti-ferromagnetic order is strongly suppressed because the lattice is not bipartite and the anti-ferromagnetic spin interaction is strongly frustrated.

At weak interactions, only a superconducting instability is expected.

- (b) On the hexagonal (honeycomb) lattice at half filling, the Fermi surface consists of a set of Dirac points. The density of states vanishes. There are no instabilities at small interactions. At stronger interactions, various instabilities have been found, including a spin liquid and f -wave Cooper pairing.

A more complete summary of these results can be found in [51].

5 Summary, conclusions, and outlook

The aim of this lecture was to present an introduction to the Hubbard model and its properties so that the reader has an overview and knows the very recent advances in the field as well. I put the focus on rigorous results, since they already provide much insight into the physics of the Hubbard model and serve as landmarks, and shortly described the method of functional renormalization and the results obtained.

The reader who is interested in rigorous results can take the three short reviews of Lieb [8] and Tasaki [9, 44] as a starting point or can dive into the original proofs using the references. There are still many open questions in this field.

The reader who wants to better understand functional renormalization should consult the review of Metzner *et al.* [51], or, if he is interested in the mathematical aspects, the book of Salmhofer [50].

Many other methods have been used to study the Hubbard model. Some of them are introduced and discussed in the other lectures in this School.

The fermionic Hubbard model has been investigated during the last 50 years, but there are still lots of open questions. In the past five years, the interest in the bosonic Hubbard model has increased, mainly because it is possible now to experimentally investigate such systems using optical lattices. Much less is known for the bosonic case. Most of the proof techniques that worked well for fermions cannot be applied here, so that there are only very few rigorous results. Other techniques that have been applied to the fermionic case are difficult as well. But due to the current and increasing interest, I expect many new results in this fascinating field in the near future.

References

- [1] J. Hubbard, Proc. Roy. Soc. A **276**, 238 (1963)
- [2] J. Kanamori, Prog. Theo. Phys. **30**, 275 (1963)
- [3] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)
- [4] R. Pariser and R.G. Parr, J. Chem. Phys. **21**, 466 (1953)
- [5] R. Pariser and R.G. Parr, J. Chem. Phys. **21**, 767 (1953)
- [6] J. Pople, Trans. Faraday Soc. **49**, 1375 (1953)
- [7] P. Fulde: *Electron correlations in molecules and solids* (Springer-Verlag, Berlin, Heidelberg, New-York 1991)
- [8] E. Lieb, in *In Advances in Dynamical Systems and Quantum Physics*, pp. 392–412 (Internat. Press 1995), also cond-mat/9311033v2
- [9] H. Tasaki, cond-mat/9712219 (1997)
- [10] E. Lieb, Phys. Rev. Lett. **62**, 1201 (1989)
- [11] E. Lieb and F. Wu, Phys. Rev. Lett. **20**, 1445 (1968)
- [12] F. Essler, V. Korepin, and K. Schoutens, Phys. Rev. Lett. **67**, 3848 (1991)
- [13] O.J. Heilmann and E.H. Lieb, Trans. New York Acad. Sci. **33**, 116 (1971)
- [14] B. Sriram Shastry, Phys. Rev. Lett. **56**, 2453 (1986)
- [15] H. Grosse, Lett. Math. Phys. **18**, 151 (1989)
- [16] A. Mielke, Phys. Lett. A **174**, 443 (1993)
- [17] N. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966)
- [18] M. Walker and T. Ruijgrok, Phys. Rev. **171**, 513 (1968)
- [19] D. Gosh, Phys. Rev. Lett. **27**, 1584 (1971)
- [20] P. Hohenberg, Phys. Rev. **158**, 383 (1967)
- [21] T. Koma and H. Tasaki, Phys. Rev. Lett. **68**, 3248 (1992)
- [22] O. McBryan and T. Spencer, Comm. Math. Phys. **53**, 299 (1977)
- [23] K. Ito, J. Stat. Phys. **29**, 747 (1982)
- [24] E. Lieb and D. Mattis, Phys. Rev. **125**, 164 (1962)

- [25] D. Thouless, Proc. Phys. Soc. **86**, 893 (1965)
- [26] Y. Nagaoka, Phys. Rev. **147**, 392 (1966)
- [27] H. Tasaki, Phys. Rev. B **40**, 9192 (1989)
- [28] A. Sütö, Comm. Math. Phys. **140**, 43 (1991)
- [29] A. Mielke, J. Phys. A: Math. Gen. **24**, 3311 (1991)
- [30] A. Mielke, J. Phys. A: Math. Gen. **24**, L73 (1991)
- [31] A. Mielke, J. Phys. A: Math. Gen. **25**, 4335 (1992)
- [32] H. Tasaki, Phys. Rev. Lett. **69**, 1608 (1992)
- [33] A. Mielke and H. Tasaki, Commun. Math. Phys. **158**, 341 (1993)
- [34] A. Mielke, J. Phys. A, Math. Gen. **32**, 8411 (1999)
- [35] A. Mielke, Phys. Rev. Lett. **82**, 4312 (1999)
- [36] G.-B. Jo, J. Guzman, C. Thomas, P. Hosur, A. Vishwanath, and D. Stamper-Kurn, Phys. Rev. Lett. **108**, 045305 (2012)
- [37] A. Tanaka and H. Ueda, Phys. Rev. Lett. **90**, 067204 (2003)
- [38] H. Tasaki, J. Stat. Phys. **84**, 535 (1996)
- [39] C.D. Batista and B.S. Shastry, Phys. Rev. Lett. **91**, 116401 (2003)
- [40] A. Mielke, Eur. Phys. J. B **85**, 1 (2012)
- [41] A. MacLachlan, Mol. Phys. **2**, 271 (1959)
- [42] A. MacLachlan, Mol. Phys. **4**, 49 (1961)
- [43] E.H. Lieb, M. Loss, and R.J. McCann, J. Math. Phys. **34**, 891 (1993)
- [44] H. Tasaki, cond-mat/9512169 (1995)
- [45] F. Wegner, Ann. Physik (Leipzig) **3**, 77 (1994)
- [46] S.D. Glazek and K.G. Wilson, Phys. Rev. D **48**, 5863 (1993)
- [47] I. Grote, E. Körding, and F. Wegner, J. Low Temp. Phys. **126**, 1385 (2002)
- [48] V. Hankevych and F. Wegner, Eur. Phys. J. B **31**, 333 (2003)
- [49] F. Wegner and V. Hankevych, Acta Phys. Pol. B **34**, 497 (2003)

- [50] M. Salmhofer: *Renormalization: An Introduction*
(Springer Verlag, Berlin, Heidelberg, New-York 1998)
- [51] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer,
Rev. Mod. Phys. **84**, 299 (2012)
- [52] J.W. Negele and H. Orland: *Quantum many-particle systems*
(Addison-Wesley, Redwood City, Menlo Park, Reading 1988)
- [53] W. Kohn and J. Luttinger, Phys. Rev. Lett. **15**, 524 (1965)
- [54] C. Halboth and W. Metzner, Phys. Rev. Lett. **85**, 5162 (2000)
- [55] C. Halboth and W. Metzner, Phys. Rev. B **61**, 7364 (2000)

12 The Two-Dimensional Hubbard Model

Robert Eder

Karlsruhe Institute of Technology

Institute for Solid State Physics

Contents

1	The Hubbard model	2
2	The Hubbard-I approximation	3
3	The Gutzwiller wave function	9
4	Strong coupling theory	15
5	Spin waves	18
6	Single hole problem	21
7	Summary and discussion	29
A	The bosonic Bogoliubov transformation	31

1 The Hubbard model

The Hubbard model was proposed in the 1960's 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, if 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 will be 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 the lattice site i by $c_{i,\sigma}^\dagger$ the model thus can be written as

$$H = \sum_{i,j} \sum_{\sigma} t_{i,j} c_{i,\sigma}^\dagger c_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} = H_t + H_U. \quad (1)$$

Here $n_{i,\sigma} = c_{i,\sigma}^\dagger c_{i,\sigma}$ 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 basic approximations and to discuss some of the problems which so far precluded a full solution.

We consider (1) for a two-dimensional square lattice with N sites and periodic boundary conditions. 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 concern for cuprate superconductors is $n_e \geq 0.8$, i.e., the lightly doped Mott-insulator, and $U/t \approx 10$.

2 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’s functions.

We consider the case 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 : $d_{i,\sigma}^{\dagger} h_{j,-\sigma}^{\dagger} |0\rangle$. 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 $c_{i,\sigma}^{\dagger}$ 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:

$$H_{\text{eff},1} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} \left(t_{i,j} d_{i,\sigma}^{\dagger} h_{j,-\sigma}^{\dagger} + H.c. \right) + U \sum_{i,\sigma} d_{i,\sigma}^{\dagger} d_{i,\sigma}. \quad (2)$$

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 $\downarrow$. Again, we assume that the probability for this is $1/2$. We therefore can write down the second term for the effective Hamiltonian

$$H_{\text{eff},2} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} t_{i,j} \left(d_{i,\sigma}^{\dagger} d_{j,\sigma} - h_{i,-\sigma}^{\dagger} h_{j,-\sigma} \right). \quad (3)$$

The negative sign of the hopping term for holes is due to the fact that the original hopping term has to be rewritten as $-t_{i,j} c_{j,\sigma} c_{i,\sigma}^{\dagger}$ to describe the propagation of a hole. Addition of (2) and (3) and Fourier transformation gives

$$H_{\text{eff}} = \sum_{\mathbf{k},\sigma} \left(\left(\frac{\varepsilon_{\mathbf{k}}}{2} + U \right) d_{\mathbf{k},\sigma}^{\dagger} d_{\mathbf{k},\sigma} - \frac{\varepsilon_{\mathbf{k}}}{2} h_{\mathbf{k},\sigma}^{\dagger} h_{\mathbf{k},\sigma} \right) + \sum_{\mathbf{k},\sigma} \frac{\varepsilon_{\mathbf{k}}}{2} \left(d_{\mathbf{k},\sigma}^{\dagger} h_{-\mathbf{k},-\sigma}^{\dagger} + H.c. \right), \quad (4)$$

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 Bogoliubov transformation

$$\begin{aligned}\gamma_{-,\mathbf{k},\sigma} &= u_{\mathbf{k}}d_{\mathbf{k},\sigma} + v_{\mathbf{k}}h_{-\mathbf{k},-\sigma}^\dagger \\ \gamma_{+,\mathbf{k},\sigma} &= -v_{\mathbf{k}}d_{\mathbf{k},\sigma} + u_{\mathbf{k}}h_{-\mathbf{k},-\sigma}^\dagger\end{aligned}\quad (5)$$

this can be solved resulting in the dispersion relations for the lower and upper Hubbard band

$$E_{\mathbf{k},\pm} = \frac{1}{2} \left(\varepsilon_{\mathbf{k}} + U \pm \sqrt{\varepsilon_{\mathbf{k}}^2 + U^2} \right). \quad (6)$$

In the limit $U/t \gg 1$ this simplifies to $E_{\mathbf{k},-} = \varepsilon_{\mathbf{k}}/2$, $E_{\mathbf{k},+} = \varepsilon_{\mathbf{k}}/2 + U$ so that the original band with dispersion $\varepsilon_{\mathbf{k}}$ is split into two bands, separated by a gap of U , each having half of the original width. For the case of particle-hole symmetry, the chemical potential is $U/2$ [5], so the lower band is completely filled and 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 that are created in pairs and propagate, whereby the double-occupancies have a large ‘energy of formation’ of U .

To compute the spectral weight of the bands we translate the electron annihilation operator as follows:

$$c_{\mathbf{k},\sigma} = \frac{1}{\sqrt{2}} \left(d_{\mathbf{k},\sigma} + h_{-\mathbf{k},-\sigma}^\dagger \right) = \frac{1}{\sqrt{2}} \left((u_{\mathbf{k}} + v_{\mathbf{k}})\gamma_{-,\mathbf{k},\sigma} + (u_{\mathbf{k}} - v_{\mathbf{k}})\gamma_{+,\mathbf{k},\sigma} \right).$$

Namely, annihilation of an electron on a singly occupied site creates a hole, whereas annihilation on a doubly occupied site annihilates a double occupancy. The factor of $1/\sqrt{2}$ takes into account that both processes are possible with a probability of $1/2$ in the disordered state. We obtain

$$Z_{\pm}(\mathbf{k}) = \frac{1}{2} (u_{\mathbf{k}} \mp v_{\mathbf{k}})^2 = \frac{1}{2} \left(1 \pm \frac{\varepsilon_{\mathbf{k}}}{\sqrt{\varepsilon_{\mathbf{k}}^2 + U^2}} \right). \quad (7)$$

Taking again the limit $U/t \gg 1$, the spectral weight of each of the bands is only $\approx 1/2$ per k-point.

Next, we derive these results in a more rigorous fashion following Hubbard’s original paper [1]. We split the electron operator into the two eigenoperators of the interaction part H_U in (1):

$$c_{i,\sigma} = c_{i,\sigma} n_{i,-\sigma} + c_{i,\sigma} (1 - n_{i,-\sigma}) = \hat{d}_{i,\sigma} + \hat{c}_{i,\sigma}, \quad (8)$$

which obey $[\hat{d}_{i,\sigma}, H_U] = U \hat{d}_{i,\sigma}$ and $[\hat{c}_{i,\sigma}, H_U] = 0$. Then we define the four time-ordered zero-temperature Green’s functions [6]

$$G_{\alpha,\beta}(\mathbf{k}, t) = -i \langle T \alpha_{\mathbf{k},\sigma}(t) \beta_{\mathbf{k},\sigma}^\dagger \rangle, \quad (9)$$

where $\alpha, \beta \in \{\hat{c}, \hat{d}\}$. These Green's functions obey the equations of motion (with $\hbar = 1$)

$$i\partial_t G_{\alpha,\beta}(\vec{k}, t) = \delta(t) \langle \{\beta_{\mathbf{k},\sigma}^\dagger, \alpha_{\mathbf{k},\sigma}\} \rangle - i \langle T[\alpha_{\mathbf{k},\sigma}, H](t) \beta_{\mathbf{k},\sigma}^\dagger \rangle.$$

The commutators $[\alpha_{\mathbf{k},\sigma}, H_U]$ are trivial but the commutators with the kinetic term H_t are involved. After some algebra, using the identity $n_{i,\sigma} = n_i/2 + \sigma S_i^z$, we find:

$$\begin{aligned} [\hat{c}_{i,\uparrow}, H_t] &= \sum_j t_{ij} \left[\left(1 - \frac{n_e}{2}\right) c_{j,\uparrow} + (c_{j,\uparrow} S_i^z + c_{j,\downarrow} S_i^-) - \frac{1}{2} c_{j,\uparrow} (n_i - n_e) + c_{j,\downarrow}^\dagger c_{i,\downarrow} c_{i,\uparrow} \right], \\ [\hat{d}_{i,\uparrow}, H_t] &= \sum_j t_{ij} \left[\frac{n_e}{2} c_{j,\uparrow} - (c_{j,\uparrow} S_i^z + c_{j,\downarrow} S_i^-) + \frac{1}{2} c_{j,\uparrow} (n_i - n_e) - c_{j,\downarrow}^\dagger c_{i,\downarrow} c_{i,\uparrow} \right]. \end{aligned} \quad (10)$$

The first term on the right-hand side describes the ‘simple’ propagation of the hole. The second term is the Clebsch-Gordan contraction of the spin-1 operator 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 i . Similarly, the third term describes hopping combined with creation of a density excitation at site j whereas the last term describes the coupling to 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 sides – obviously a rather crude approximation. After Fourier transformation we obtain

$$\begin{aligned} [\hat{c}_{\mathbf{k},\uparrow}, H] &\approx \left(1 - \frac{n_e}{2}\right) \varepsilon_{\mathbf{k}} (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) \\ [\hat{d}_{\mathbf{k},\uparrow}, H] &\approx \frac{n_e}{2} \varepsilon_{\mathbf{k}} (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) + U \hat{d}_{\mathbf{k},\uparrow} \end{aligned}$$

If we set $n_e = 1/2$ and identify

$$\begin{aligned} \sqrt{2} \hat{d}_{\mathbf{k},\sigma} &\rightarrow d_{\mathbf{k},\sigma} \\ \sqrt{2} \hat{c}_{\mathbf{k},\sigma} &\rightarrow h_{\mathbf{k},-\sigma}^\dagger \end{aligned} \quad (11)$$

exactly the same equations of motion are obtained from the heuristic Hamiltonian H_{eff} (4) (the significance of the factor $\sqrt{2}$ will become clear in a moment).

Using the anticommutator relations $\{\hat{d}_{i,\sigma}^\dagger, \hat{d}_{i,\sigma}\} = n_{i,-\sigma}$, $\{\hat{c}_{i,\sigma}^\dagger, \hat{c}_{i,\sigma}\} = (1 - n_{i,-\sigma})$, $\{\hat{d}_{i,\sigma}^\dagger, \hat{c}_{i,\sigma}\} = \{\hat{c}_{i,\sigma}^\dagger, \hat{d}_{i,\sigma}\} = 0$ and putting $\langle n_{i,\sigma} \rangle = n_e/2$ we obtain the Fourier transformed equations of motion:

$$\begin{pmatrix} \omega - (1 - n_e/2) \varepsilon_{\mathbf{k}} & -(1 - n_e/2) \varepsilon_{\mathbf{k}} \\ -n_e/2 \varepsilon_{\mathbf{k}} & \omega - n_e/2 \varepsilon_{\mathbf{k}} - U \end{pmatrix} \begin{pmatrix} G_{\hat{c},\hat{c}} & G_{\hat{c},\hat{d}} \\ G_{\hat{d},\hat{c}} & G_{\hat{d},\hat{d}} \end{pmatrix} = \begin{pmatrix} 1 - n_e/2 & 0 \\ 0 & n_e/2 \end{pmatrix}. \quad (12)$$

If we consider again $n_e = 1/2$ and multiply both sides of (12) by 2 we have the unit matrix on the right while all Green's functions are multiplied by a factor of 2 – which would originate from the additional factor of $\sqrt{2}$ in (11). The Hamiltonian (4) would therefore produce exactly the same equations of motion as the Hubbard-I approximation, which demonstrates the equivalence. We continue with arbitrary n_e and use the identity (which holds for any 2×2 matrix)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

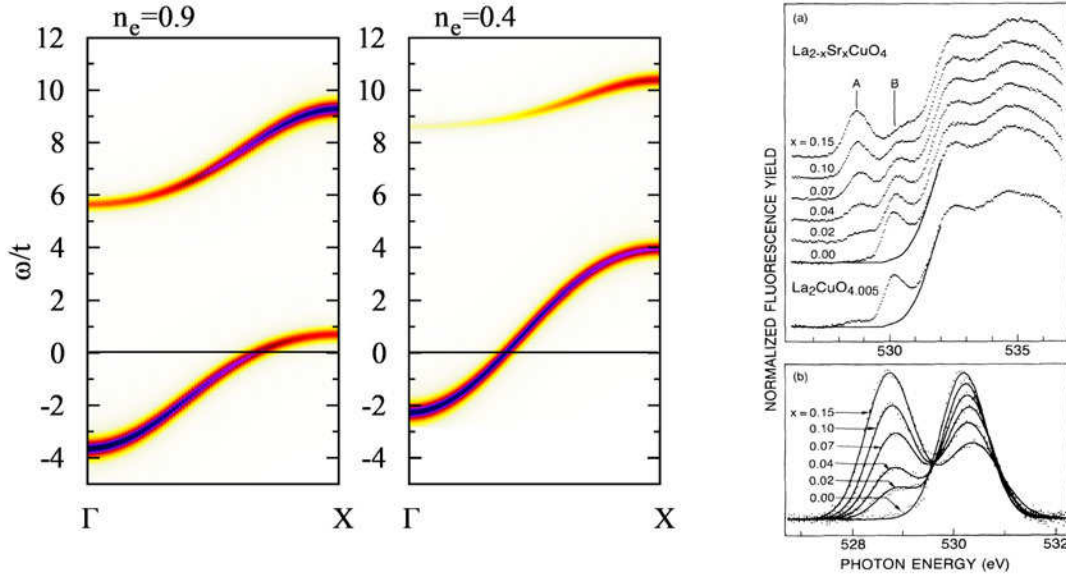


Fig. 1: Left: Single particle spectral function obtained from the Green's function (13) for two different electron densities. The Fermi energy is at zero. Right: X-ray luminescence spectra for $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ show the unoccupied part of the lower Hubbard-band (A) and the upper Hubbard-band (B). With decreasing $n_e \approx 1 - x$ the upper Hubbard-band rapidly loses intensity. Reprinted with permission from [7], Copyright 1991 by the American Physical Society.

to solve for the Green's function matrix $\mathbf{G}(\mathbf{k}, \omega)$. Recalling that $c_{\mathbf{k},\sigma} = \hat{c}_{\mathbf{k},\sigma} + \hat{d}_{\mathbf{k},\sigma}$, the standard electron Green's function $G(\mathbf{k}, \omega)$ is given by $G = G_{\hat{c},\hat{c}} + G_{\hat{c},\hat{d}} + G_{\hat{d},\hat{c}} + G_{\hat{d},\hat{d}}$ and after some algebra this can be brought to the form

$$G(\mathbf{k}, \omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma(\omega)}$$

$$\Sigma(\omega) = \frac{n_e}{2} U + \frac{n_e}{2} \left(1 - \frac{n_e}{2}\right) \frac{U^2}{\omega - \left(1 - \frac{n_e}{2}\right) U} \quad (13)$$

from which also the self-energy $\Sigma(\omega)$ corresponding to the Hubbard-I approximation can be read off. In order to fix the Fermi energy E_F we write the operator of electron number as

$$\hat{N}_e = 2 \sum_i n_{i,\uparrow} n_{i,\downarrow} + \sum_i \left(n_{i,\uparrow} (1 - n_{i,\downarrow}) + n_{i,\downarrow} (1 - n_{i,\uparrow}) \right) = \sum_{i,\sigma} \left(\hat{d}_{i,\sigma}^\dagger \hat{d}_{i,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{c}_{i,\sigma} \right).$$

The expectation value of $\hat{N}_e$ then can be expressed in terms of the Green's functions (9)

$$\begin{aligned} \langle N_e \rangle &= -2i \sum_{\mathbf{k}} \left(G_{\hat{d},\hat{d}}(\mathbf{k}, t = 0^-) + G_{\hat{c},\hat{c}}(\mathbf{k}, t = 0^-) \right) \\ &= -2i \sum_{\mathbf{k}} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega 0^+} \left(G_{\hat{d},\hat{d}}(\mathbf{k}, \omega) + G_{\hat{c},\hat{c}}(\mathbf{k}, \omega) \right) \\ &= 2 \sum_{\mathbf{k}} \int_{-\infty}^{\mu} d\omega \left(A_{\hat{d},\hat{d}}(\mathbf{k}, \omega) + A_{\hat{c},\hat{c}}(\mathbf{k}, \omega) \right), \end{aligned} \quad (14)$$

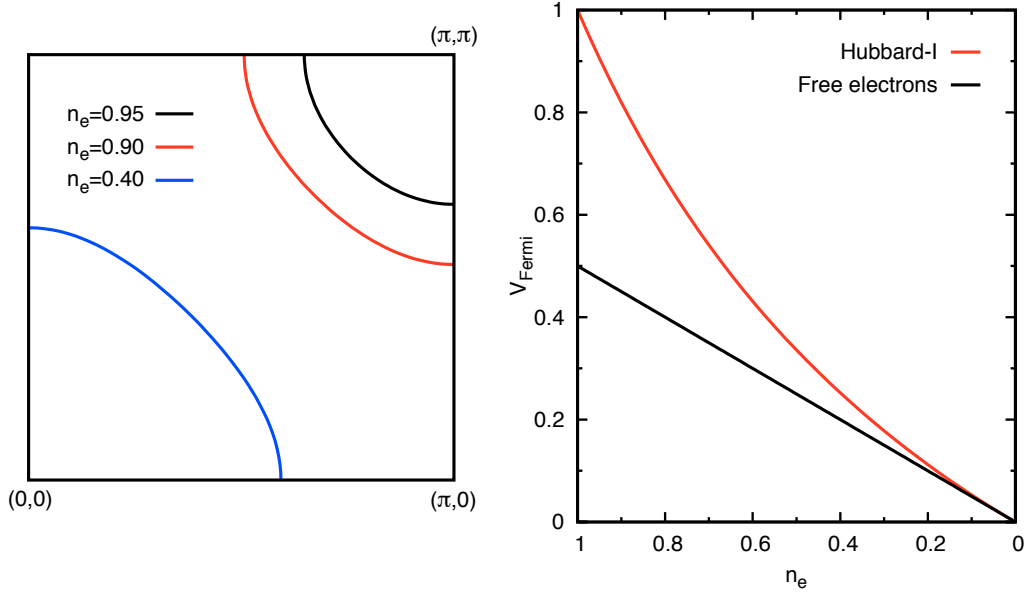


Fig. 2: Left: Fermi surface for different electron densities. Right: Fermi surface volume as a function of electron density n_e .

where $A_{\hat{d},\hat{d}}$ and $A_{\hat{c},\hat{c}}$ are the spectral densities of the respective Green's function. The resulting expression for N_e is somewhat lengthy so we do not write it explicitly. Close to half-filling it can be written as

$$N_e = \Theta(E_F - E_{\mathbf{k},-}) + \mathcal{O}(1 - n_e).$$

For $n_e = 1$ this means that the completely filled lower Hubbard band with N occupied momenta per spin-direction corresponds to N electrons (as it has to be), whereas for $n_e < 1$ it means that the lower Hubbard band is doped with holes. Figure 1 shows the spectral density obtained from the Green's function (13) for $U/t = 8$ and two different band fillings, whereas Figure 2 shows the resulting Fermi surfaces and the dependence of the Fermi surface volume, obtained from (14), on the electron density. In Figure 1, 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) so that the Fermi surface takes the form of a small pocket around $X = (\pi, \pi)$, see Figure 2, whereby the area of the pocket is roughly proportional to the hole density $n_h = 1 - n_e$.

An interesting feature seen in Figure 1 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 loses 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

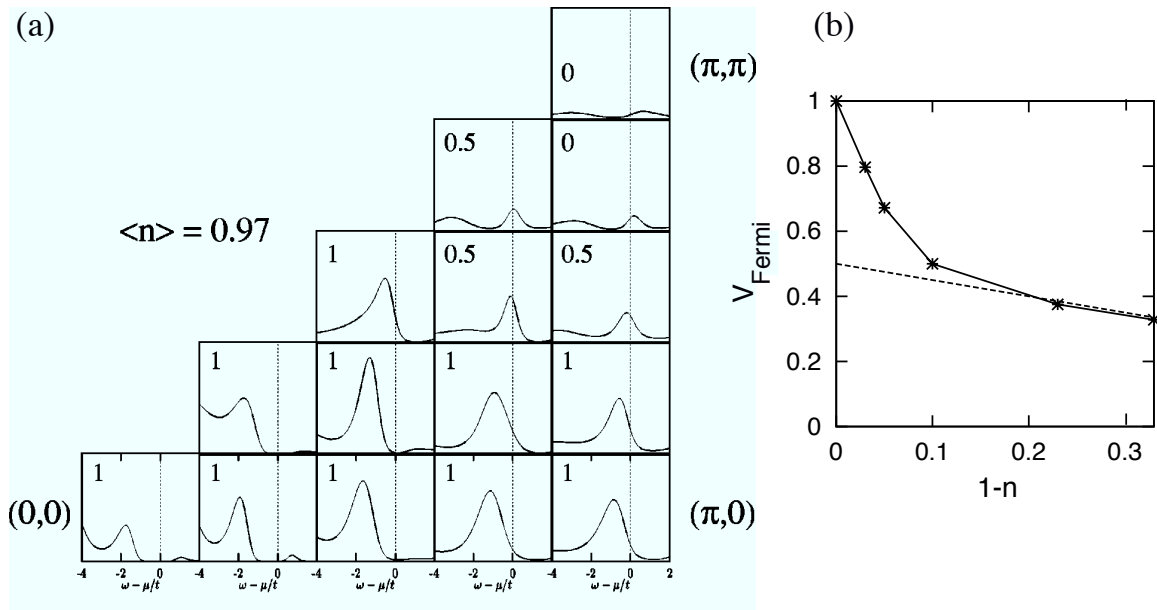


Fig. 3: Left: Single particle spectral function $A(\mathbf{k}, \omega)$ obtained by Quantum Monte-Carlo simulations on an 8×8 cluster at $k_B T = t$. Right: Fermi surface volume (15) deduced from $A(\mathbf{k}, \omega)$ versus electron density. The dashed line corresponds to free electrons. Reprinted with permission from [8], Copyright 2000 by the American Physical Society.

semiconductor or insulator is one of the fingerprints of strong correlations and can be observed experimentally in cuprate superconductors. An example is shown in Figure 1 [7]. 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 2 also shows the dependence of the Fermi surface volume V_{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 the Fermi surface volume for free electrons, where $V_{\text{Fermi}} = n_e/2$. The Hubbard-I approximation gives $V_{\text{Fermi}} \rightarrow 1$, a completely filled band, as $n_e \rightarrow 1$, and approaches the free electron behavior for small n_e . This leads to a peculiar nonlinear dependence in $V_{\text{Fermi}}(n_e)$, 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, which will be seen to be $J = 4t^2/U$. Figure 3 shows the result of a quantum Monte-Carlo calculation of the spectral density for an 8×8 cluster at the rather high temperature $k_B T = t$. The 8×8 cluster has the allowed momenta $(n\pi/4, m\pi/4)$ with integer m and n and Figure 3 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 an 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 Figure 1

for $n_e = 0.9$. To study V_{Fermi} , an ‘occupation number’ $n_{\mathbf{k}}$ of 1, 0.5 or 0 was assigned to each momentum $\mathbf{k}$, depending on 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

$$V_{\text{Fermi}} = \frac{1}{64} \sum_{\mathbf{k}} n_{\mathbf{k}}, \quad (15)$$

where 64 is the number of momenta in the 8×8 cluster. The obtained estimate for V_{Fermi} is also shown in Figure 3 as a function of electron density and indeed has a rough similarity to the result for the Hubbard-I approximation.

3 The Gutzwiller wave function

The Gutzwiller wave function 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 projection operator. More precisely, the Gutzwiller wave function reads [3]

$$|\Phi_G\rangle = \prod_i (1 - \lambda n_{i,\uparrow} n_{i,\downarrow}) |FS\rangle,$$

where λ is a variational parameter to be determined by minimizing the energy E_G . First we rewrite the Fermi sea as a superposition of real space configurations. Suppressing the spin index we have

$$\begin{aligned} \prod_{j=1}^M c_{\mathbf{k}_j}^\dagger |0\rangle &= \frac{1}{\sqrt{N^M}} \sum_{i_1, i_2, i_3, \dots, i_M} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_j}\right) \prod_{j=1}^M c_{i_j}^\dagger |0\rangle \\ &= \frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} \sum_{\sigma} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}}\right) \prod_{j=1}^M c_{i_{\sigma(j)}}^\dagger |0\rangle \end{aligned}$$

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 $\prod_{j=1}^M c_{i_{\sigma(j)}}^\dagger$ we permute the $c_i^\dagger$ operators back to the ordered sequence $c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger$. The permutation that brings $\sigma(i) \rightarrow i$ obviously is σ^{-1} and since the Fermi sign of σ^{-1} is equal to that of σ we obtain

$$\begin{aligned} &\frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} \sum_{\sigma} (-1)^\sigma \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}}\right) c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger |0\rangle \\ &= \frac{1}{\sqrt{N^M}} \sum_{i_1 > i_2 > i_3 \dots > i_M} D(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M | i_1, i_2, \dots, i_M) c_{i_1}^\dagger c_{i_2}^\dagger \dots c_{i_M}^\dagger |0\rangle, \end{aligned}$$

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

$$c_{i_{1,\uparrow}}^\dagger c_{i_{2,\uparrow}}^\dagger c_{i_{3,\uparrow}}^\dagger \dots c_{i_{N_\uparrow,\uparrow}}^\dagger c_{j_{1,\downarrow}}^\dagger c_{j_{2,\downarrow}}^\dagger c_{j_{3,\downarrow}}^\dagger \dots c_{j_{N_\downarrow,\downarrow}}^\dagger |0\rangle$$

that 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. 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 of $W(N_d) = \langle\Phi(N_d)|\Phi(N_d)\rangle$ and we consider which N_d gives the largest contribution in this sum. To compute norms, we need to evaluate expressions such as

$$\begin{aligned} D^*(\mathbf{k}_j|i_j) D(\mathbf{k}_j|i_j) &= \sum_{\sigma,\sigma'} (-1)^\sigma (-1)^{\sigma'} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}})\right) \\ &= M! + \sum_{\sigma \neq \sigma'} (-1)^\sigma (-1)^{\sigma'} \exp\left(i \sum_{j=1}^M \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}})\right). \end{aligned} \quad (16)$$

where in the first term we have collected the $M!$ terms with $\sigma = \sigma'$. At this point, we make an important approximation: (16) still has to be summed over $i_1, i_2, i_3 \dots i_M$. 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 all 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 combinatorial 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

$$\begin{aligned} \frac{d}{dN_d} \log(W(N_d)) &= \log\left((1 - \lambda)^2 \frac{(n_\uparrow - n_d)(n_\downarrow - n_d)}{n_d(1 - n_\uparrow - n_\downarrow + n_d)}\right), \\ \frac{d^2}{dN_d^2} \log(W(N_d)) &= -\frac{1}{N} \left(\frac{1}{n_d} + \frac{1}{n_\uparrow - n_d} + \frac{1}{n_\downarrow - n_d} + \frac{1}{1 - n_\uparrow - n_\downarrow + n_d} \right) = -\frac{c}{N} \end{aligned}$$

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 $n_\uparrow = n_\downarrow = 1/2$ where

$$n_{d,0} = \frac{1 - \lambda}{2(2 - \lambda)}. \quad (17)$$

For the noninteracting case $\lambda \rightarrow 0$ this gives $n_{d,0} = 1/4$ as it has to be. From the second equation we find

$$\begin{aligned} \log(W(N_d)) &= \log(W(N_{d,0})) - \frac{c}{2N} (N - N_{d,0})^2 + \dots \\ W(N_d) &= W(N_{d,0}) \exp\left(-\frac{c}{2N} (N_d - N_{d,0})^2\right) = W(N_{d,0}) \exp\left(-\frac{Nc}{2} (n_d - n_{d,0})^2\right), \end{aligned}$$

which shows that as a function of n_d the weight $W(N_d)$ is a Gaussian with a 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 $\langle H_U \rangle = NU n_{d,0}$. The expectation value of the kinetic energy 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

$$c_{i,\sigma}^\dagger c_{j,\sigma} = \hat{d}_{i,\sigma}^\dagger \hat{d}_{j,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{d}_{j,\sigma} + \hat{d}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} + \hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma}.$$

If the number of double occupancies is decreased, the expectation value of the first term on the right-hand side 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 right-hand side 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 $D(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M | i_1, i_2, \dots, i_M)$ 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 [9]. 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

$$B_i = \frac{|i, 0\rangle + \alpha_\uparrow|i, \uparrow\rangle + \alpha_\downarrow|i, \downarrow\rangle + \beta|i, \uparrow\downarrow\rangle}{\sqrt{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2}}$$

$$|\Psi\rangle = \prod_i B_i$$

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

$$\langle N_\sigma \rangle = N \frac{\alpha_\sigma^2 + \beta^2}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2},$$

$$\langle N_d \rangle = N \frac{\beta^2}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2}. \quad (18)$$

These equations can be reverted to give

$$\alpha_\sigma = \sqrt{\frac{n_\sigma - n_d}{1 - n_\uparrow - n_\downarrow + n_d}},$$

$$\beta = \sqrt{\frac{n_d}{1 - n_\uparrow - n_\downarrow + n_d}}. \quad (19)$$

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 $|\Psi\rangle$ that has precisely $\langle N_\uparrow \rangle$ $\uparrow$ -electrons, etc. Next, the representation of the electron annihilation operator $c_{i,\sigma}$ is

$$\tilde{c}_{i,\sigma} = |i, 0\rangle \langle i, \sigma| + |i, -\sigma\rangle \langle i, \uparrow\downarrow|.$$

Here a subtle detail should be noted: in the expression on the right-hand side it is assumed that a double occupancy is always 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

$$r(\sigma, n_\uparrow, n_\downarrow, n_d) = \frac{\langle \Psi' | \tilde{c}_{i,\sigma}^\dagger \tilde{c}_{j,\sigma} | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle}. \quad (20)$$

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 $|\Psi\rangle$ 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 that deviate by at most $N^{-1/2}$ from the average values (18) have an appreciable weight. This means, however, that $\mathcal{P}$ simply can be dropped so that we replace $|\Psi'\rangle \rightarrow |\Psi\rangle$ in (20). Then, since $|\Psi\rangle$ is normalized, the denominator can be dropped. 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 $\tilde{c}_{i,\sigma}^\dagger$ and $\tilde{c}_{j,\sigma}$ are identical. So

$$\begin{aligned} r(\sigma, n_\uparrow, n_\downarrow, n_d) &= \langle \Psi | \tilde{c}_{i,\sigma}^\dagger | \Psi \rangle^2 = \left(\frac{\alpha_\sigma + \alpha_{-\sigma}\beta}{1 + \alpha_\uparrow^2 + \alpha_\downarrow^2 + \beta^2} \right)^2 \\ &= \left(\sqrt{n_\sigma - n_d} \sqrt{1 - n_\uparrow - n_\downarrow + n_d} + \sqrt{n_d} \sqrt{n_{-\sigma} - n_d} \right)^2, \end{aligned}$$

where the second line has been obtained by inserting (19). 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 (17). Lastly, we divide r by the value for $U \rightarrow 0$ where $n_d = n_\uparrow \cdot n_\downarrow$ to obtain the proper limiting value for $U = 0$ and finally obtain

$$\eta(\sigma, n_\uparrow, n_\downarrow, n_d) = \left(\frac{\sqrt{n_\sigma - n_d} \sqrt{1 - n_\uparrow - n_\downarrow + n_d} + \sqrt{n_d} \sqrt{n_{-\sigma} - n_d}}{\sqrt{n_\sigma(1 - n_\sigma)}} \right)^2. \quad (21)$$

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

$$e_G = \eta(n_\sigma, n_d) t_0 + n_d U. \quad (22)$$

where $e_G = E_G/N$ and t_0 is the (kinetic) energy of the Fermi sea per site. Using (21) 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 seems 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 the new state $|\Phi_G(\mathbf{k})\rangle$ amounts to replacing $e_G \rightarrow e_G - \tilde{\varepsilon}_{\mathbf{k}}/N$, $t_0 \rightarrow t_0 - \varepsilon_{\mathbf{k}}/N$, $n_\uparrow \rightarrow n_\uparrow - 1/N$ and $n_d \rightarrow n_d + \delta n_d/N$ where δn_d is the as-yet-unknown shift of n_d . Inserting into (22) and expanding we find

$$e_G - \frac{1}{N} \tilde{\varepsilon}_{\mathbf{k}} = \left(\eta - \frac{1}{N} \frac{\partial \eta}{\partial n_\uparrow} + \frac{1}{N} \frac{\partial \eta}{\partial n_d} \delta n_d \right) \left(t_0 - \frac{1}{N} \varepsilon_{\mathbf{k}} \right) + n_d U + \frac{1}{N} \delta n_d U.$$

The terms of zeroth order in $1/N$ cancel due to (22), and collecting the first order terms gives

$$\tilde{\varepsilon}_{\mathbf{k}} = \eta \varepsilon_{\mathbf{k}} + t_0 \frac{\partial \eta}{\partial n_{\uparrow}} - \left(\frac{\partial \eta}{\partial n_d} t_0 + U \right) \delta n_d.$$

The last term on the right-hand side vanishes because the expression in the bracket is $\frac{de_G}{dn_d}$. The second term on the right-hand side is a $\mathbf{k}$ -independent shift that can be absorbed into a shift of E_F . The quasiparticle dispersion $\tilde{\varepsilon}_{\mathbf{k}}$ therefore follows the original dispersion but is 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 $n_{\mathbf{k}} = 2\langle c_{\mathbf{k},\uparrow}^\dagger c_{\mathbf{k},\uparrow} \rangle$. This may be obtained as the functional derivative of the ground state energy with respect to $\varepsilon_{\mathbf{k}}$, which means under a change $t_{ij} \rightarrow t_{ij} + \delta t_{ij}$ so that $\varepsilon_{\mathbf{k}} \rightarrow \varepsilon_{\mathbf{k}} + \delta \varepsilon_{\mathbf{k}}$, the change of the ground state energy is

$$e_G \rightarrow e_G + 2 \sum_{\mathbf{k}} n_{\mathbf{k}} \delta \varepsilon_{\mathbf{k}}.$$

From (22), we obtain the variation of e_G as

$$\delta e_G = 2\eta \sum_{\mathbf{k}} n_{\mathbf{k}}^{(0)} \delta \varepsilon_{\mathbf{k}} + \delta n_d \left(\frac{\partial \eta}{\partial n_d} t_0 + U \right),$$

where $n_{\mathbf{k}}^{(0)} = \Theta(E_F - \varepsilon_{\mathbf{k}})$ is the momentum distribution of the Fermi sea. Again, the second term on the right-hand side vanishes due to the extremum condition for n_d so that $n_{\mathbf{k}} = \eta n_{\mathbf{k}}^{(0)}$. 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 $n_{\mathbf{k}}^{(0)}$ 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 $(1 - \eta) n_e/2$. In fact, for any $\varepsilon_{\mathbf{k}}$ that can be represented by hopping integrals $t_{i,j}$, one has $\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} = 0$, so 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 step of magnitude η at the position of the original Fermi surface. Let us now consider in more detail the case $n_\sigma = 1/2$ in which the Mott-insulator should be realized for large U/t . We find from Eq. (21)

$$\eta(n_d) = 16 n_d \left(\frac{1}{2} - n_d \right).$$

Minimizing (22) 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 $1/4$, 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.

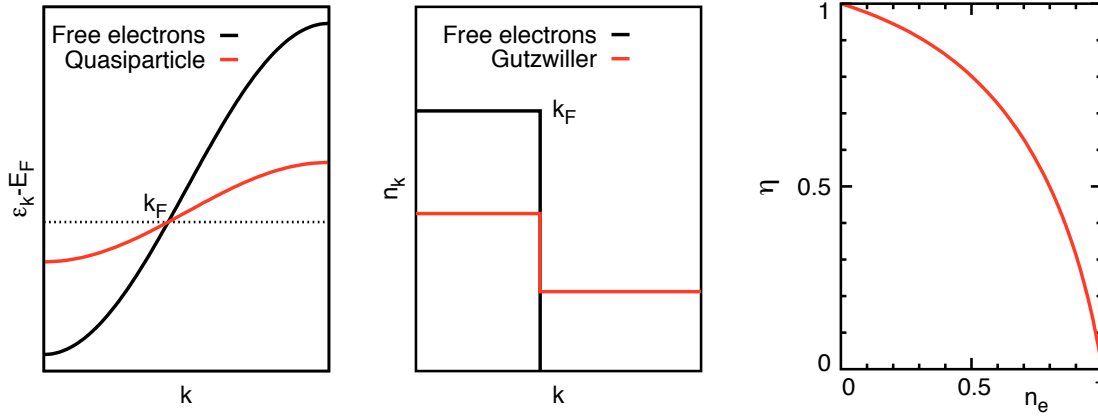


Fig. 4: Dispersion relation (left) and momentum distribution function (middle) obtained from the Gutzwiller wave function compared to the free electron case. The right part shows the dependence $\eta(n_e)$ for the two-dimensional Hubbard model with $U/t = 16$.

This is commonly interpreted as a metal-to-insulator transition as a function of increasing U , the so-called Brinkman-Rice transition [10]. Brinkman and Rice also could show that the magnetic susceptibility diverges at the transition as one would expect 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 4 then shows the dependence of η on n_e for $U/t = 16$, i.e., for $U > U_c$. As $n_e \rightarrow 1$ the renormalization factor $\eta \rightarrow 0$ so that both the bandwidth and the step in n_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 a 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 $n_e \rightarrow 1$, 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 $n_e \rightarrow 1$.

4 Strong coupling theory

The approximations we have considered so far – Hubbard-I and Gutzwiller wave functions – neglect the coupling of the electrons to the collective excitations of the strongly correlated electron system. This interaction with collective excitations on the one hand has a massive impact on the dispersion and lifetimes of the electrons, but on the other hand is very hard to treat. In the following, we illustrate this effect by studying the problem of a single hole in a quantum antiferromagnet, a problem for which a reasonably accurate solution is possible. As a first step, we derive an effective Hamiltonian that describes only the lower Hubbard band.

As usual, we consider the limit $U/t \gg 1$. In this limit most of the sites are occupied by at most one electron of either spin direction and double occupancies exist only as short-lived interme-

diate states. In this section we derive an effective Hamiltonian that operates in the sector of the Hilbert space with at most one electron per site but takes into account the effect of the ‘virtual’ double occupancies by suitable correction terms. This can be achieved by a technique called canonical perturbation theory. The basic assumption of this approximation is that the Hilbert space can be decomposed into ‘sectors’ that are energetically well separated. For the Hubbard model an obvious decomposition is according to the number N_d of double occupancies. Accordingly, we decompose the Hamiltonian into a part H_0 having matrix elements only between states in the same sector and a part H_1 that connects different sectors [11]:

$$\begin{aligned} H_0 &= \sum_{i,j} \sum_{\sigma} t_{i,j} \left(\hat{c}_{i,\sigma}^{\dagger} \hat{c}_{j,\sigma} + \hat{d}_{i,\sigma}^{\dagger} \hat{d}_{j,\sigma} \right) + U \sum_i n_{i,\uparrow} n_{i,\downarrow}, \\ H_1 &= \sum_{i,j} \sum_{\sigma} t_{i,j} \left(\hat{d}_{i,\sigma}^{\dagger} \hat{c}_{j,\sigma} + \hat{c}_{i,\sigma}^{\dagger} \hat{d}_{j,\sigma} \right). \end{aligned} \quad (23)$$

Obviously, H_0 does not change the number of double occupancies whereas H_1 decreases or increases this number by one. We now consider unitary transformations within the Hilbert space that act on the states $|\Psi\rangle$ and operators $\hat{O}$ as follows

$$\begin{aligned} |\Psi'\rangle &= e^S |\Psi\rangle \\ \hat{O}' &= e^S \hat{O} e^{-S} = \hat{O} + [S, \hat{O}] + \frac{1}{2!} [S, [S, \hat{O}]] + \frac{1}{3!} [S, [S, [S, \hat{O}]]] + \dots, \end{aligned}$$

where the second line is the Baker-Campbell-Hausdorff theorem. Unitarity of e^S requires that the so-called generator S is anti-Hermitian, $S^{\dagger} = -S$ (for this reason one can often see this written as $S \rightarrow iS'$ with a Hermitian S' in the literature). We now seek a generator S such that the ‘inter sector part’ H_1 in (23) is eliminated from the transformed Hamiltonian H' . The approach obviously makes sense only if S is small so that the expansion of the transformed operators can be terminated after some low order, usually second order in S . Since

$$H' = H_0 + H_1 + [S, H_0] + [S, H_1] + \frac{1}{2!} [S, [S, H_0]] + \frac{1}{2!} [S, [S, H_1]] + \dots$$

the generator S obviously has to fulfill $H_1 + [S, H_0] = 0$ in order to eliminate H_1 . To second order in S the transformed Hamiltonian then becomes

$$\begin{aligned} H' &= H_0 + [S, H_1] + \frac{1}{2!} [S, [S, H_0]] + \frac{1}{2!} [S, [S, H_1]] + \frac{1}{3!} [S, [S, [S, H_0]]] + \dots \\ &= H_0 - \frac{1}{2} [S, H_1] + \frac{1}{2!} [S, [S, H_1]] + \dots \end{aligned}$$

For the Hubbard model an additional complication occurs: namely H_0 in (23) contains two terms of different orders of magnitude, i.e., a part of the hopping term $\propto t_{ij}$ and the Coulomb term $H_U \propto U \gg t_{ij}$. Accordingly, we demand $H_1 + [S, H_U] = 0$. Using $[H_U, \hat{d}_{i,\sigma}^{\dagger}] = U \hat{d}_{i,\sigma}^{\dagger}$ it is easy to see that S is given by

$$S = \sum_{i,j} \sum_{\sigma} \frac{t_{i,j}}{U} \left(\hat{d}_{i,\sigma}^{\dagger} \hat{c}_{j,\sigma} - \hat{c}_{i,\sigma}^{\dagger} \hat{d}_{j,\sigma} \right),$$

so that indeed $S \sim \frac{t}{U} \ll 1$ and the truncation of the expansion in powers of S is meaningful. The correction term to the Hamiltonian then becomes

$$H'_c = \frac{1}{2} [S, H_1] = \frac{1}{2} \sum_{i,j,l,m} \sum_{\sigma,\sigma'} \frac{t_{i,j} t_{l,m}}{U} \left[\hat{d}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} - \hat{c}_{i,\sigma}^\dagger \hat{d}_{j,\sigma}, \hat{d}_{l,\sigma'}^\dagger \hat{c}_{m,\sigma'} + \hat{c}_{l,\sigma'}^\dagger \hat{d}_{m,\sigma'} \right] \quad (24)$$

Expanding the commutator on the right-hand side will produce a considerable number of terms. However, keeping in mind the goal of the present calculation, namely the derivation of an effective Hamiltonian that describes the lower Hubbard band, i.e., the sector of the Hilbert space with $N_d = 0$, most of these terms can be discarded. Namely each of the products of four operators resulting from writing out (24) contains two $\hat{d}$ -operators, and these must appear exactly in the sequence $\hat{d}_{i,\sigma'}^\dagger \hat{d}_{i,\sigma}$. This is because neither the initial nor the final state must contain a double occupancy and the above combination is the only one that obeys this constraint (for the same reason, we neglect the commutator of the kinetic part $\propto t_{ij}$ of H_0 with S ; this would produce terms with one $\hat{d}$ -operator). Dropping the undesired terms we obtain

$$\begin{aligned} H'_c &= -\frac{1}{2} \sum_{i,j,l} \sum_{\sigma,\sigma'} \frac{t_{i,j} t_{l,i}}{U} \hat{c}_{l,\sigma'}^\dagger \hat{d}_{i,\sigma'} \hat{d}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} - \frac{1}{2} \sum_{i,j,m} \sum_{\sigma,\sigma'} \frac{t_{i,j} t_{j,m}}{U} \hat{c}_{i,\sigma}^\dagger \hat{d}_{j,\sigma} \hat{d}_{j,\sigma'}^\dagger \hat{c}_{m,\sigma'} \\ &= -\sum_{i,j,l} \sum_{\sigma,\sigma'} \frac{t_{i,l} t_{l,j}}{U} \hat{c}_{i,\sigma}^\dagger \hat{d}_{l,\sigma'} \hat{d}_{l,\sigma}^\dagger \hat{c}_{j,\sigma} \end{aligned}$$

where the second line was obtained by exchanging $(i, l) \rightarrow (l, i)$ in the first term and $(j, m) \rightarrow (l, j)$ in the second. Next we use $\hat{d}_{l,\uparrow}^\dagger \hat{d}_{l,\uparrow} = n_{l,\downarrow}$ and $\hat{d}_{l,\downarrow}^\dagger \hat{d}_{l,\downarrow} = -S_l^+$ and find

$$H' = -\sum_{i,j,l} \frac{t_{i,l} t_{l,j}}{U} \left((\hat{c}_{i,\uparrow}^\dagger n_{l,\downarrow} \hat{c}_{j,\uparrow} - \hat{c}_{i,\downarrow}^\dagger S_l^+ \hat{c}_{j,\uparrow}) + (\hat{c}_{i,\downarrow}^\dagger n_{l,\uparrow} \hat{c}_{j,\downarrow} - \hat{c}_{i,\uparrow}^\dagger S_l^- \hat{c}_{j,\downarrow}) \right)$$

In the special case where $i = j$ there is an additional factor of 2 because the ‘intermediate’ double occupancy may be formed either at i or at j . The respective terms become

$$H' = -2 \sum_{i,j} \frac{t_{i,j}^2}{U} (n_{i,\uparrow} n_{j,\downarrow} + n_{i,\downarrow} n_{j,\uparrow} - S_i^+ S_j^- - S_i^- S_j^+) = 4 \sum_{i,j} \frac{t_{i,j}^2}{U} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{n_i n_j}{4} \right)$$

where we used $n_\uparrow = n/2 + S^z$ and $n_\downarrow = n/2 - S^z$. With the abbreviation $J_{ij} = 4t_{i,j}^2/U$, the complete effective Hamiltonian for the lower Hubbard band thus becomes

$$\begin{aligned} H_{\text{sc}} &= \sum_{i,j} \sum_{\sigma} t_{i,j} \hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} + \sum_{i,j} J_{ij} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{n_i n_j}{4} \right) \\ &\quad - \sum_{i,j,l} \frac{t_{i,l} t_{l,j}}{U} \left((\hat{c}_{i,\uparrow}^\dagger n_{l,\downarrow} \hat{c}_{j,\uparrow} - \hat{c}_{i,\downarrow}^\dagger S_l^+ \hat{c}_{j,\uparrow}) + (\uparrow \leftrightarrow \downarrow) \right), \end{aligned} \quad (25)$$

The first two terms together are called the t - J model. It has been proposed by Zhang and Rice as an effective model for the CuO_2 planes of cuprate superconductors [4]. The second line is frequently referred to as the ‘conditional hopping terms’ or ‘three-site hopping terms.’

5 Spin waves

We first recall the goal of the present discussion, which is the study of the interaction between electrons and collective modes in a doped Mott insulator. In the following, we discuss the most important collective excitations of the undoped Mott insulator, namely spin waves. Spin waves are frequently discussed using the Holstein-Primakoff transformation [12] but for the case of spin 1/2 a simpler and more transparent derivation is possible, which is outlined below.

We consider the strong-coupling Hamiltonian (25) for the case of exactly one electron per site. Since no empty site is present, the terms in (25) that transport an electron from one site to another can be dropped and we are left with

$$H_{sc} = \sum_{i,j} J_{i,j} \left(\mathbf{S}_i \cdot \mathbf{S}_j - \frac{n_i n_j}{4} \right)$$

We assume that there are nonvanishing $t_{i,j}$, and hence also $J_{i,j}$, only between nearest neighbors. The terms $\propto n_i n_j$ give only an unimportant constant shift of $-J/4$ per bond, so we omit them. Finally we arrive at the Heisenberg antiferromagnet (note that by definition $J_{i,j} > 0$)

$$H_{HAF} = J \sum_{\langle i,j \rangle} \left(S_i^z S_j^z + \frac{1}{2} (S_i^+ S_j^- + S_i^- S_j^+) \right), \quad (26)$$

where $\langle i,j \rangle$ denotes a sum over pairs of nearest neighbors. If only the term $\propto S_i^z S_j^z$ were present, the ground state of (26) would be the Néel state, shown in Figure 5 (a). In this state, the square lattice is divided into two sublattices and all sites of the A -sublattice are occupied by an $\uparrow$ -electron, whereas all sites of the B -sublattice are occupied by a $\downarrow$ -electron. However, the Néel state is not an eigenstate of the total Hamiltonian: acting, e.g., with the term $S_i^- S_j^+$ contained in the second term of (26), the spins at the sites i and j are inverted, see Figure 5 (b), and the resulting state is orthogonal to the Néel state. To deal with these so-called quantum fluctuations we proceed as follows: we interpret the Néel state as the vacuum $|0\rangle$ and we model an inverted spin on the site i of the A sublattice as the presence of a Boson, created by $a_i^\dagger$. Similarly, an inverted spin on the site j of the B sublattice is modelled by the presence of a Boson created by $b_j^\dagger$. The state Figure 5 (b) thus would be $a_i^\dagger b_j^\dagger |0\rangle$. We use Bosons to represent the inverted spins because the spin-flip operators acting on different sites commute. Since any given spin can be inverted only once, a state like $(a_i^\dagger)^2 |0\rangle$ is meaningless. Accordingly, we have to impose the additional constraint that at most one Boson can occupy a given site. We call this the hard-core constraint. An inverted spin on either sublattice is parallel to its $z = 4$ nearest neighbors and the energy changes from $-J/4$ to $+J/4$ for each of these z bonds. Accordingly, we ascribe an ‘energy of formation’ of $zJ/2$ to each Boson. The transverse part creates pairs of inverted spins on nearest neighbors, with the matrix element being $J/2$ and we can thus write down the following Hamiltonian to describe the quantum fluctuations:

$$H_{SW} = \frac{zJ}{2} \left(\sum_{i \in A} a_i^\dagger a_i + \sum_{i \in B} b_i^\dagger b_i \right) + \frac{J}{2} \sum_{i \in A} \sum_{\mathbf{n}} \left(a_i^\dagger b_{i+\mathbf{n}}^\dagger + b_{i+\mathbf{n}} a_i \right). \quad (27)$$

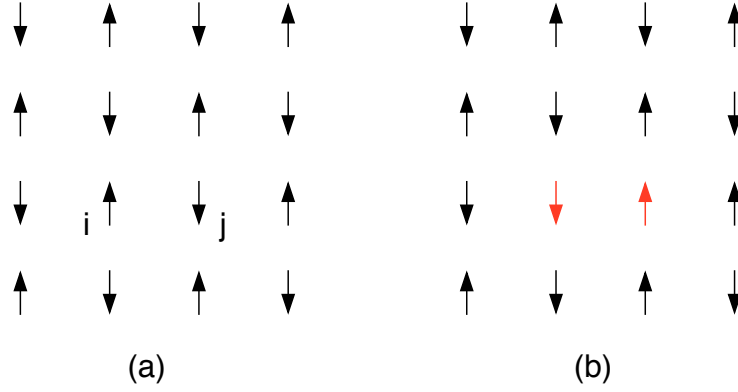


Fig. 5: The Néel state (a) is not the ground state of the Heisenberg antiferromagnet. By acting, e.g., with the term $J/2 S_i^- S_j^+$ in (26) the state (b) is generated, which is orthogonal to (a).

Here $\mathbf{n}$ are the z vectors which connect a given site with its z nearest neighbors. The Hamiltonian (27) is in quadratic form, but we recall that the Bosons are not free particles but have to obey the hard-core constraint. However, we now simply ignore this and treat the Bosons as free particles – we will return to this issue later on. Fourier transformation of (27) gives

$$H_{SW} = \frac{zJ}{2} \sum_{\mathbf{k}} \left(a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + b_{\mathbf{k}}^\dagger b_{\mathbf{k}} + \gamma_{\mathbf{k}} (a_{\mathbf{k}}^\dagger b_{-\mathbf{k}}^\dagger + b_{-\mathbf{k}} a_{\mathbf{k}}) \right),$$

$$\gamma_{\mathbf{k}} = \frac{1}{z} \sum_{\mathbf{n}} e^{i\mathbf{k} \cdot \mathbf{n}} = \frac{1}{4} \left(2 \cos(k_x) + 2 \cos(k_y) \right) \quad (28)$$

where $\mathbf{k}$ is a wave vector in the antiferromagnetic Brillouin zone. We can solve (28) by a Bosonic Bogoliubov transformation, i.e., we make the ansatz

$$\begin{aligned} \gamma_{a,\mathbf{k}}^\dagger &= u_{\mathbf{k}} a_{\mathbf{k}}^\dagger + v_{\mathbf{k}} b_{-\mathbf{k}} \\ \gamma_{b,-\mathbf{k}}^\dagger &= u_{\mathbf{k}} b_{-\mathbf{k}}^\dagger + v_{\mathbf{k}} a_{\mathbf{k}}. \end{aligned} \quad (29)$$

Demanding that $[\gamma_{a,\mathbf{k}}, \gamma_{a,\mathbf{k}}^\dagger] = [\gamma_{b,\mathbf{k}}, \gamma_{b,\mathbf{k}}^\dagger] = 1$ gives the condition $u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2 = 1$. An equation for $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ is obtained by demanding $[H, \gamma_{a,\mathbf{k}}^\dagger] = \omega_{\mathbf{k}} \gamma_{a,\mathbf{k}}^\dagger$. This is explained in detail in the Appendix. Using the formulae from the Appendix, we find the spin wave dispersion and the coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$:

$$\omega_{\mathbf{k}} = \frac{zJ}{2} \sqrt{1 - \gamma_{\mathbf{k}}^2}, \quad u_{\mathbf{k}} = \sqrt{\frac{1 + \nu_{\mathbf{k}}}{2\nu_{\mathbf{k}}}}, \quad v_{\mathbf{k}} = \sqrt{\frac{1 - \nu_{\mathbf{k}}}{2\nu_{\mathbf{k}}}},$$

where $\nu_{\mathbf{k}} = \sqrt{1 - \gamma_{\mathbf{k}}^2}$. For $\mathbf{k} \rightarrow 0$ we have $\gamma_{\mathbf{k}} \rightarrow 1 - (k_x^2 + k_y^2)/4$ so that $\omega_{\mathbf{k}} \rightarrow \sqrt{2}J|\mathbf{k}|$, i.e., the spin waves have a cone-shaped dispersion and reach zero frequency at $\mathbf{k} = (0, 0)$ but also at $\mathbf{k} = (\pi, \pi)$. To compute observables we revert the transformation

$$\begin{aligned} a_{\mathbf{k}}^\dagger &= u_{\mathbf{k}} \gamma_{a,\mathbf{k}}^\dagger - v_{\mathbf{k}} \gamma_{b,-\mathbf{k}}, \\ b_{-\mathbf{k}} &= -v_{\mathbf{k}} \gamma_{a,\mathbf{k}}^\dagger + u_{\mathbf{k}} \gamma_{b,-\mathbf{k}}. \end{aligned} \quad (30)$$

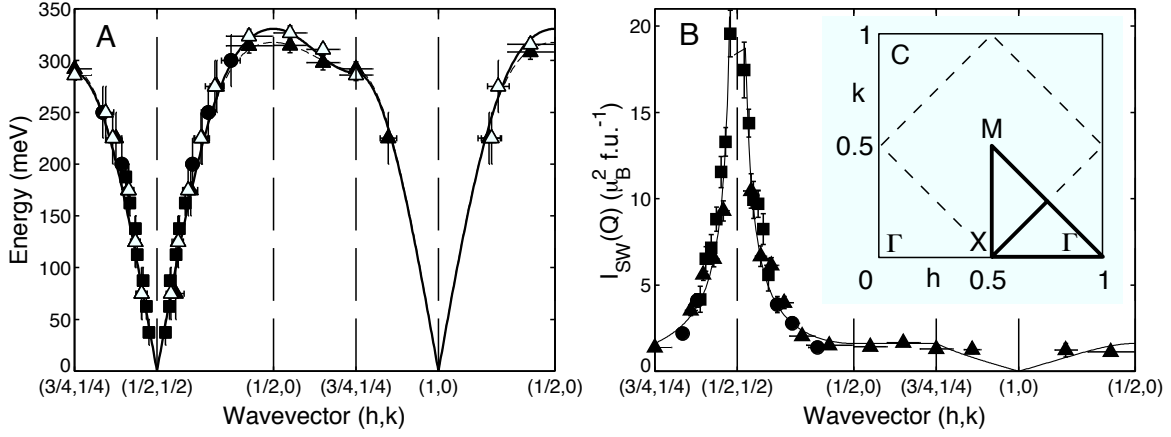


Fig. 6: Comparison of spin-wave theory to experiments on La_2CuO_4 . Panel A shows the dispersion of the frequency $\omega_{\mathbf{k}}$ of the spin waves, panel B the $\mathbf{k}$ -dependence of the peak intensity, which is easily expressed in terms of the coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$. The dots are experimental data, the lines the prediction of spin-wave theory. Reprinted with permission from [14], Copyright 2001 by the American Physical Society.

As an application, let us consider the ground state energy. The energy of the Néel state, which was the zero of energy for the spin wave Hamiltonian (28), was $-J/4$ per bond, and there are $zN/2$ bonds in the system. To this we add the expectation value of (28) calculated in the ground state, i.e., the vacuum for the γ 's, thereby using (30). We obtain

$$\langle H_{sw} \rangle = \frac{zJ}{2} \sum_{b,fk} (2v_{\mathbf{k}}^2 - 2\gamma_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}}) = \frac{zJN}{4} \left[\frac{2}{N} \sum_{\mathbf{k}} \left(\sqrt{1 - \gamma_{\mathbf{k}}^2} - 1 \right) \right].$$

The expression in the square bracket can be converted to an integral over the antiferromagnetic zone and evaluated numerically. The result is that for the two-dimensional square lattice the ground state energy per bond is lowered from $-0.25 J$ to $-0.328974 J$ due to the quantum fluctuations. Monte-Carlo simulations for the two-dimensional Heisenberg antiferromagnet give a ground state energy of $-0.33 J$ per bond [13]. The spin waves can also be observed experimentally by inelastic neutron scattering. An example is shown in Figure 6. It is quite obvious that the agreement with experiment is excellent and in fact spin-wave theory is an extraordinarily successful description of many properties of magnetic Mott insulators.

To conclude this section, we return to the issue of the hard-core constraint which the $a^\dagger$ and $b^\dagger$ Bosons had to obey and which we simply ignored. To address this question, we calculate the density of these Bosons, i.e.,

$$n_a = \frac{2}{N} \sum_{\mathbf{k}} \langle a_{\mathbf{k}}^\dagger a_{\mathbf{k}} \rangle = \frac{2}{N} \sum_{\mathbf{k}} v_{\mathbf{k}}^2 = \frac{2}{N} \sum_{\mathbf{k}} \frac{1 - \nu_{\mathbf{k}}}{2\nu_{\mathbf{k}}}.$$

Numerical evaluation for a 2D square lattice gives $n_a = 0.19$. The probability that two of the Bosons occupy the same site and violate the constraint therefore is $\approx n_a^2 = 0.04 \ll 1$ and our assumption of relaxing the constraint is justified *a posteriori*.

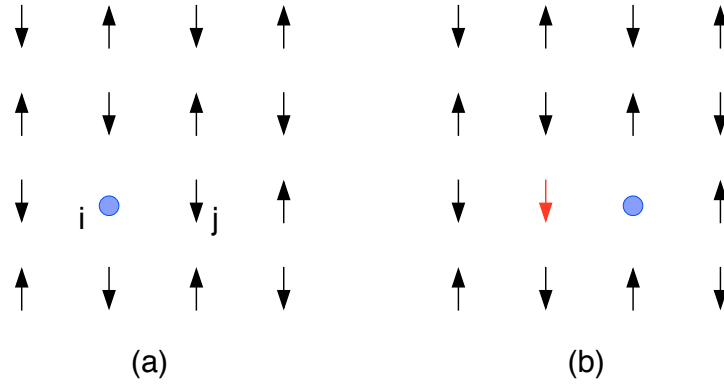


Fig. 7: A hole hopping in the Néel state displaces spins, which thus become misaligned.

6 Single hole problem

In the preceding section, we considered the strong coupling Hamiltonian (25) for $N_e = N$, where it reduces to the Heisenberg antiferromagnet, and we studied the collective excitations of the Mott insulator, i.e., the spin waves. Now we consider the case $N_e = N - 1$, i.e., a single mobile hole in an antiferromagnet, and we study the interaction between this hole and the spin waves. Again we assume that there is a nonvanishing hopping element, denoted by $-t$, only between nearest neighbors, and for simplicity we drop the three-site hopping terms in (25); that means we study the t - J model. Since we saw that the ground state still has antiferromagnetic order – although reduced – we again start from the Néel state and assume that the electron on site i , belonging to the $\uparrow$ -sublattice, is removed – see Figure 7 (a). Then, the hopping terms in (25) become active and an electron from a neighboring site j can hop to i , resulting in the state in Figure 7 (b). Since this electron has ‘switched sublattices’, however, its spin now is opposite to that of the electron originally at site i . This inverted spin at site i then may be viewed as a spin wave as discussed in the preceding section. In other words, the propagating hole ‘radiates off’ spin waves and thus is coupled to the spin excitations. This process would be described precisely by the term $c_{j,\downarrow} S_i^-$ in the commutator relations (10) – which were neglected in the Hubbard-I approximation. In the following we will see, however, that the coupling of the single hole to the spin excitations modifies the ‘band structure’ drastically.

We continue to use the Bosons $a_i^\dagger$ and $b_j^\dagger$ defined in the preceding section and introduce an additional ‘particle’ namely a hole created by the fermionic creation operators $h_{a,i}^\dagger$ for $i \in A$ and $h_{b,j}^\dagger$ for $j \in B$. We have introduced two species of hole creation operators, $h_{a,i}^\dagger$ and $h_{b,j}^\dagger$, because we continue to use the antiferromagnetic Brillouin zone, which in turn necessitates the two-sublattice structure. However, both particles simply stand for a ‘hole.’ Since we are considering only a single hole the statistics of $h_i^\dagger$ moreover is irrelevant and we might as well describe it by a Bosonic operator. The two states shown in Figure 7 would then be expressed as $h_{a,i}^\dagger |0\rangle$ and $h_{b,j}^\dagger a_i^\dagger |0\rangle$, and generalizing this we can immediately write down the Hamiltonian

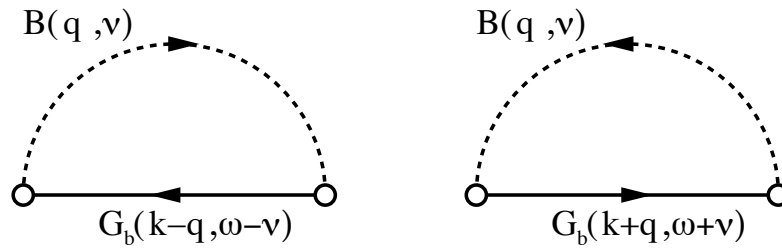


Fig. 8: Simplest Feynman diagrams for the self-energy of a hole.

for the interaction between the hole and the spin defects:

$$H_{\text{int}} = t \sum_{i \in A} \sum_{\mathbf{n}} \left(h_{b,i+\mathbf{n}}^\dagger h_{a,i} a_i^\dagger + H.c. \right) + t \sum_{j \in B} \sum_{\mathbf{n}} \left(h_{a,j+\mathbf{n}}^\dagger h_{b,j} b_j^\dagger + H.c. \right).$$

Note that the hopping integral for a hole is $+t$. Upon Fourier transformation this becomes

$$H_{\text{int}} = \sqrt{\frac{2}{N}} \sum_{\mathbf{k}, \mathbf{q}} \left(\left(\varepsilon_{\mathbf{k}-\mathbf{q}} h_{b,\mathbf{k}-\mathbf{q}}^\dagger h_{a,\mathbf{k}} a_{\mathbf{q}}^\dagger + H.c. \right) + (a \leftrightarrow b) \right).$$

where $\varepsilon_{\mathbf{k}} = 2t(\cos(k_x) + \cos(k_y))$. Next, we use the inverse Bogoliubov transformation (19) to replace $a^\dagger$ and $b^\dagger$ by $\gamma_i^\dagger$ and add the spin-wave Hamiltonian to describe the dynamics of the $\gamma_i^\dagger$. The total Hamiltonian for the coupled hole-magnon system then reads

$$H_{\text{tot}} = \sqrt{\frac{2}{N}} \sum_{\mathbf{k}, \mathbf{q}} \left(\left(M(\mathbf{k}, \mathbf{q}) h_{b,\mathbf{k}-\mathbf{q}}^\dagger h_{a,\mathbf{k}} \gamma_{a,\mathbf{q}}^\dagger + H.c. \right) + (a \leftrightarrow b) \right) + \sum_{\mathbf{q}} \omega_{\mathbf{q}} \left(\gamma_{a,\mathbf{q}}^\dagger \gamma_{a,\mathbf{q}} + \gamma_{b,\mathbf{q}}^\dagger \gamma_{b,\mathbf{q}} \right), \quad (31)$$

where $M(\mathbf{k}, \mathbf{q}) = \varepsilon_{\mathbf{k}-\mathbf{q}} u_{\mathbf{q}} - \varepsilon_{\mathbf{k}} v_{\mathbf{q}}$. Obviously the first term describes how a hole with momentum $\mathbf{k}$ ‘radiates off’ a spin wave with momentum $\mathbf{q}$ thereby changing its own momentum to $\mathbf{k} - \mathbf{q}$. The Hamiltonian (31) no longer is a quadratic form and requires more sophisticated techniques for its solution. More precisely, we will use the so-called self-consistent Born approximation (SCB) to derive an equation for the self-energy of the hole. We again define the time-ordered zero-temperature Green’s functions [6]

$$\begin{aligned} G_{\alpha}(\mathbf{k}, t) &= -i \langle T h_{\alpha,\mathbf{k}}(t) h_{\mathbf{k}}(0)^\dagger \rangle \\ B_{\alpha}(\mathbf{q}, t) &= -i \langle T \gamma_{\alpha,\mathbf{q}}(t) \gamma_{\mathbf{q}}(0)^\dagger \rangle, \end{aligned}$$

where $\langle \dots \rangle$ denotes the expectation value in the empty state containing neither spin waves nor a hole. The simplest Feynman diagrams for the self-energy of the hole are shown in Fig. 8, and the corresponding expression is [6]

$$\Sigma_a(\mathbf{k}, \omega) = \frac{i}{2\pi} \frac{2}{N} \sum_{\mathbf{q}} \int d\nu M^2(\mathbf{k}, \mathbf{q}) [B_a(\mathbf{q}, \nu) G_b(\mathbf{k}-\mathbf{q}, \omega-\nu) + B_a(\mathbf{q}, \nu) G_b(\mathbf{k}+\mathbf{q}, \omega+\nu)]. \quad (32)$$

Since a single hole in a macroscopic system is not expected to have any effect on the spin wave spectrum we may use the noninteracting Green's function for the spin waves, which is given by

$$B_\alpha^{(0)}(\mathbf{q}, t) = -ie^{-i\omega_{\mathbf{q}}t} \left(\Theta(t) \langle \gamma_{\alpha, \mathbf{q}} \gamma_{\alpha, \mathbf{q}}^\dagger \rangle + \Theta(-t) \langle \gamma_{\alpha, \mathbf{q}}^\dagger \gamma_{\alpha, \mathbf{q}} \rangle \right) = -i\Theta(t) e^{-i\omega_{\mathbf{q}}t},$$

with Fourier transform $B_\alpha^{(0)}(\mathbf{q}, \omega) = (\omega - \omega_{\mathbf{q}} + i0^+)^{-1}$. Next, we note that (with $\alpha \in a, b$)

$$0 = \langle h_{\alpha, \mathbf{k}}^\dagger h_{\alpha, \mathbf{k}} \rangle = -iG_\alpha(\mathbf{k}, t = 0^-) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega 0^+} G_\alpha(\mathbf{k}, \omega).$$

Due to the presence of the factor $e^{i\omega 0^+}$ we can close the integration path around the upper ω half-plane. Since the integral must vanish, we can conclude that the hole Green's function is analytical in the upper half-plane (this also would have followed from the fact that for the special case of an empty ground state the time-ordered Green's function is equal to the retarded one, and the retarded Green's function is analytical in the upper half-plane). We now insert $B_\alpha^{(0)}(\mathbf{q}, \omega)$ into the expression for the self-energy, (32), and obtain

$$\Sigma_a(\mathbf{k}, \omega) = \frac{i}{2\pi} \frac{2}{N} \sum_{\mathbf{q}} \int d\nu M^2(\mathbf{k}, \mathbf{q}) \left[\frac{G_b(\mathbf{k} - \mathbf{q}, \omega - \nu)}{\nu - \omega_{\mathbf{q}} + i0^+} + \frac{G_b(\mathbf{k} + \mathbf{q}, \omega + \nu)}{\nu - \omega_{\mathbf{q}} + i0^+} \right].$$

Since $G_b \propto |\omega|^{-1}$ for large $|\omega|$ the integrand behaves like $|\nu|^{-2}$ and we can close the integration contour around either the upper or the lower ν -half-plane. For the first term we choose the lower half-plane, where the integrand has a pole at $\nu = \omega_{\mathbf{q}} - i0^+$. If ν has a negative imaginary part and ω is real, the frequency argument $\omega - \nu$ of G_b in this term has a positive imaginary part, and since G_b is regular in the upper half-plane there are no singularities from the factor of G_b . In the second term, we close the integration contour around the upper half-plane. For ν in the upper half-plane, both $B(\nu)$ and $G(\omega + \nu)$ are regular. The second term thus vanishes and the result is

$$\Sigma_a(\mathbf{k}, \omega) = \frac{2}{N} \sum_{\mathbf{q}} M^2(\mathbf{k}, \mathbf{q}) G_b(\mathbf{k} - \mathbf{q}, \omega - \omega_{\mathbf{q}})$$

Since we expect that $G_a = G_b = G$ and hence $\Sigma_a = \Sigma_b = \Sigma$ and $G^{-1}(\mathbf{k}, \omega) = \omega - \Sigma(\mathbf{k}, \omega)$ we finally obtain the self-consistency equation for $\Sigma(\mathbf{k}, \omega)$ [15]

$$\Sigma(\mathbf{k}, \omega) = \frac{2}{N} \sum_{\mathbf{q}} \frac{M^2(\mathbf{k}, \mathbf{q})}{\omega - \omega_{\mathbf{q}} - \Sigma(\mathbf{k} - \mathbf{q}, \omega - \omega_{\mathbf{q}})}.$$

This can be solved numerically on a finite $\mathbf{k}$ -mesh and ω -grid [15], whence the hole spectral density $A(\mathbf{k}, \omega) = -\frac{1}{\pi} \text{Im} G(\mathbf{k}, \omega + i0^+)$ of the hole Green's function can be calculated. This can in principle be compared to the ARPES spectra of a Mott-insulator. Figure 9 shows the result of such a calculation, taken from Ref. [15]. The spectrum is spread out over an energy range of several t , and a considerable amount of spectral weight is distributed over this range. The spread-out weight is frequently referred to as 'incoherent continua.' At the bottom of the spectrum, around $\omega \approx -2t$, there is a relatively intense isolated peak, often referred to as the 'quasiparticle peak.' The corresponding eigenstate is the ground state of the hole for the respective momentum. Figure 9 also shows $A(\mathbf{k}, \omega)$ obtained by Lanczos diagonalization of a

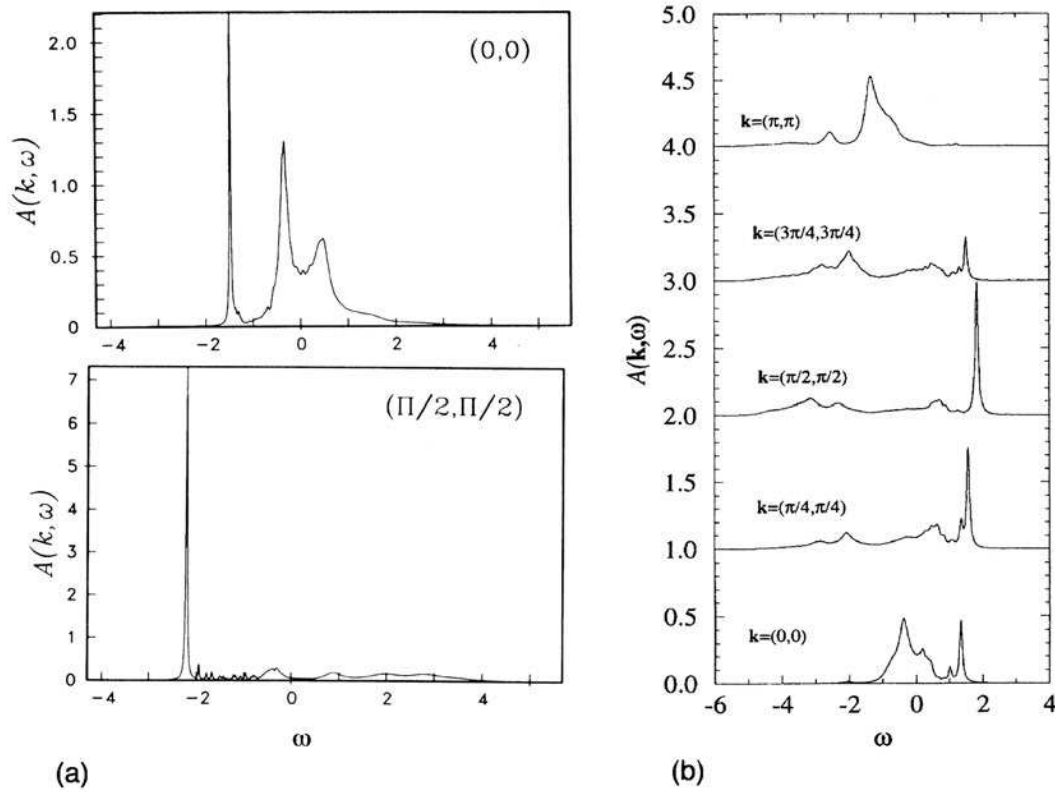


Fig. 9: (a) Single-particle spectral function $A(\mathbf{k}, \omega)$ obtained from the SCB for the t - J model [15] with $J/t = 0.3$ for $\mathbf{k} = (0, 0)$ and $\mathbf{k} = (\pi/2, \pi/2)$. (b) Single particle spectral function of the t - J model obtained by exact diagonalization of a 32-site cluster [16]. Note that this Figure has an inverted ω -axis and actually shows $A(\mathbf{k}, -\omega)$. Repinted with permission from [15], Copyright 1991 by the American Physical Society and from [16] Copyright 1995 by the American Physical Society

32-site cluster [16]. Note that the Lanczos spectra are ‘upside down’ as compared to the SCB spectra, which means the quasiparticle peak is at the top, rather than the bottom of the spectra. The Lanczos spectra also show the quasiparticle peak and the incoherent continua, and in fact even the $\mathbf{k}$ -dependence of the incoherent continua shows some similarity with the results of the SCB-approximation. Figure 10 compares the dispersion of the quasiparticle peak as obtained by Lanczos (dots) and SCB (line). Obviously the agreement is very good. The whole spectrum – quasiparticle peak plus incoherent continua – would now replace the lower Hubbard band in the Mott-insulator, whereby the quasiparticle peak would form the top of the photoemission spectrum, as in the Lanczos spectra. Comparing with the prediction of the Hubbard-I approximation, the width of the quasiparticle band is reduced drastically. Moreover one can see from Figure 10 that the band maximum is shifted from (π, π) to $(\pi/2, \pi/2)$. The hole pocket predicted by the Hubbard-I approximation therefore should form around this momentum.

The self-consistent Born approximation in fact not only agrees very well with numerical spectra but also with experiments on insulating cuprates. After the publication of ARPES spectra for the antiferromagnetic Mott-insulator $\text{Sr}_2\text{CuO}_2\text{Cl}_2$ by Wells *et al.* [17], considerable theoretical effort was put into reproducing these spectra, and after adding additional hopping integrals

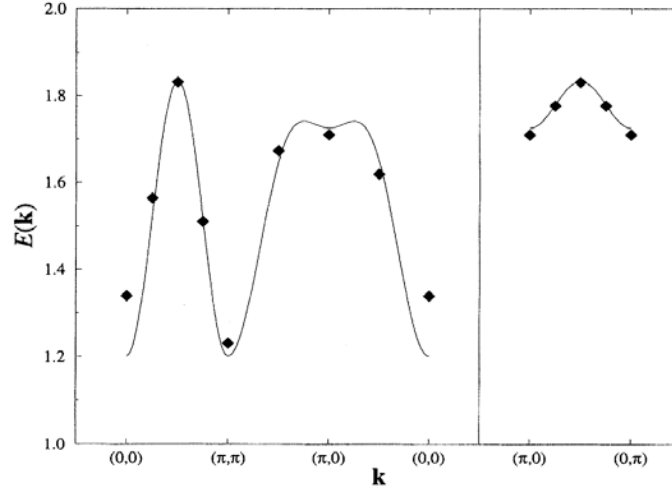


Fig. 10: Dispersion of the quasiparticle peak from the SCB (line) and by Lanczos diagonalization of a 32-site cluster. Reprinted with permission from [16], Copyright 1995 by the American Physical Society.

between 2nd and 3rd nearest neighbors, good agreement with experiment could be achieved using the self-consistent Born approximation; see for example Ref. [18].

The self-consistent Born approximation thus is quite successful but somewhat technical. For this reason we now give a simplified discussion of the states that form the quasiparticle band. We decompose the t - J Hamiltonian as $H = H_0 + H_1$ whereby

$$\begin{aligned} H_0 &= -t \sum_{\langle i,j \rangle} \sum_{\sigma} \left(\hat{c}_{i,\sigma}^{\dagger} \hat{c}_{j,\sigma} + H.c. \right) + J \sum_{\langle i,j \rangle} \left(S_i^z S_j^z - \frac{n_i n_j}{4} \right), \\ H_1 &= \frac{J}{2} \sum_{\langle i,j \rangle} (S_i^+ S_j^- + H.c.). \end{aligned} \quad (33)$$

In the absence of any hole, the ground state of H_0 is again the Néel state with energy NJ , and we choose this as the zero of energy. We assume that a hole is created at some site i , which raises the energy by $zJ/2$. By the action of the hopping term, the hole then starts to propagate. As was discussed above, however, in every step the hole shifts one electron to the opposite sublattice where its spin is opposite to the Néel order; see Figure 11. The hole thus

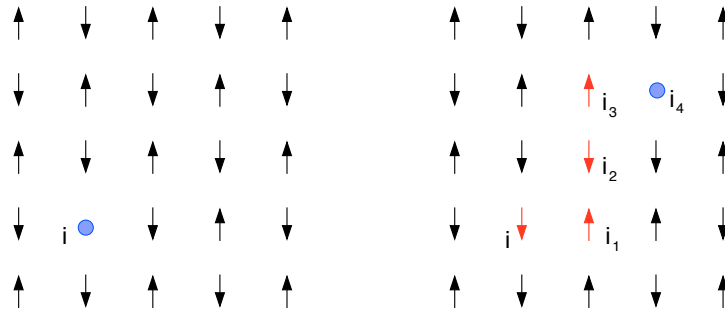


Fig. 11: A hole hopping in the Néel state creates a ‘string’ of misaligned spins.

leaves behind a trace of misaligned spins so that the magnetic energy increases roughly linearly with the distance travelled by the hole. We call such a state that is created by a hole hopping in the Néel state a ‘string’ and denote it by $|i_0, i_1, \dots, i_\nu\rangle$, where $i_1, i_2, i_{\nu-1}$ are the sites which the hole has visited, and the hole itself is at the site i_ν . In the first step the hole generates z different string states, whereas any subsequent hop from any string of length ν generates $z - 1$ strings of length $\nu + 1$. The number of strings of length ν thus is $n_\nu = z(z - 1)^{\nu-1}$ for $\nu \geq 1$, while $n_0 = 1$. Since each displaced spin is parallel to $z - 2$ neighbors – compare Figure 11 – the magnetic energy increases by $J(z - 2)/2$. The only exception is the first hop away from i where the energy increases by $J(z - 1)/2$. Accordingly, the exchange energy for a string of length $\nu > 0$ is

$$I_\nu = \frac{(z + 1)J}{2} + \nu \frac{(z - 2)J}{2}, \quad (34)$$

where $I_0 = zJ/2$. It may happen that the path that the hole has taken is folded or self-intersecting, and in this case (34) is not correct. However, it will be correct for *most* possible paths of the hole: in particular, it is correct for $\nu \leq 2$, so we will use this expression. Neglecting the possibility of self-intersection or folding of the string is an approximation known as the Bethe lattice. Since the magnetic energy increases linearly with the number of hops the hole has taken, we conclude that the hole is self-trapped. To describe the resulting localized state we make the *ansatz*

$$|\Psi_i\rangle = \sum_{\nu=0}^{\infty} \alpha_\nu \sum_{i_1, i_2, \dots, i_\nu} |i, i_1, i_2, \dots, i_\nu\rangle \quad (35)$$

where it is understood that the second sum runs only over those ν -tuples of sites that correspond to a true string starting at i . Since we assume that the magnetic energy is the same for all strings of length ν , the coefficient α_ν also depends only on the length of the string. The α_ν in (35) are to be determined by minimizing the energy. The norm and magnetic energy are

$$\langle \Psi_i | \Psi_i \rangle = \sum_{\nu=0}^{\infty} n_\nu |\alpha_\nu|^2, \quad (36)$$

$$\langle \Psi_i | H_I | \Psi_i \rangle = \sum_{\nu=0}^{\infty} n_\nu I_\nu |\alpha_\nu|^2. \quad (37)$$

To obtain the expectation value of the kinetic energy, we consider a string of length $\nu \geq 1$ with coefficient α_ν . By acting with the hopping term, we obtain $z - 1$ strings of length $\nu + 1$ with coefficient $\alpha_{\nu+1}$ and one string of length $\nu - 1$ with coefficient $\alpha_{\nu-1}$. For $\nu = 0$, we obtain z strings of length 1. In this way, we find

$$\langle \Psi_i | H_t | \Psi_i \rangle = t \left(z \alpha_0 \alpha_1 + \sum_{\nu=1}^{\infty} n_\nu \alpha_\nu (\alpha_{\nu-1} + (z - 1) \alpha_{\nu+1}) \right) = 2t \sum_{\nu=0}^{\infty} n_{\nu+1} \alpha_\nu \alpha_{\nu+1}. \quad (38)$$

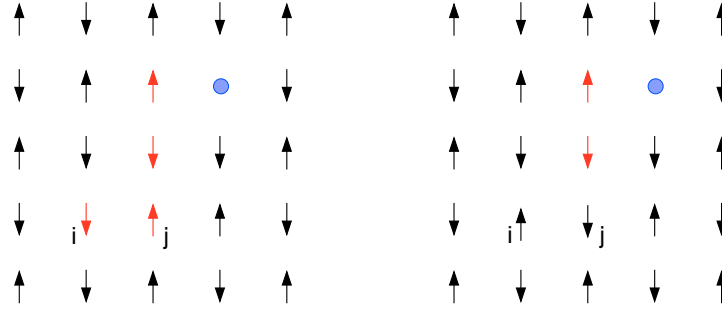


Fig. 12: By acting with the term $J/2 S_i^+ S_j^-$, the first two defects created by the hole can be ‘healed’ and the starting point of the string be shifted to a neighbor.

Then we demand

$$0 = \frac{\partial E_{\text{loc}}}{\partial \alpha_\nu} = \frac{\partial}{\partial \alpha_\nu} \frac{\langle \Psi_i | H_0 | \Psi_i \rangle}{\langle \Psi_i | \Psi_i \rangle} = \frac{1}{\langle \Psi_i | \Psi_i \rangle^2} \left[\frac{\partial \langle \Psi_i | H_0 | \Psi_i \rangle}{\partial \alpha_\nu} \langle \Psi_i | \Psi_i \rangle - \langle \Psi_i | H_0 | \Psi_i \rangle \frac{\partial \langle \Psi_i | \Psi_i \rangle}{\partial \alpha_\nu} \right]$$

$$= \frac{1}{\langle \Psi_i | \Psi_i \rangle} \left[\frac{\partial \langle \Psi_i | H_0 | \Psi_i \rangle}{\partial \alpha_\nu} - E_{\text{loc}} \frac{\partial \langle \Psi_i | \Psi_i \rangle}{\partial \alpha_\nu} \right]$$

whence

$$\frac{\partial \langle \Psi_i | H_0 | \Psi_i \rangle}{\partial \alpha_\nu} - E_{\text{loc}} \frac{\partial \langle \Psi_i | \Psi_i \rangle}{\partial \alpha_\nu} = 0.$$

Inserting equations (37), (38), and (36) we obtain [19]

$$I_\nu \alpha_\nu + t (\alpha_{\nu-1} + (z-1) \alpha_{\nu+1}) = E_{\text{loc}} \alpha_\nu \quad \text{for } \nu \geq 1$$

$$I_0 \alpha_0 + z t \alpha_1 = E_{\text{loc}} \alpha_0.$$

This is a non-Hermitian eigenvalue problem, but by introducing $\beta_\nu = \sqrt{n_\nu} \alpha_\nu$ it can be made Hermitian. So far it seems that the hole in the Néel state is localized. However, it is easy to see that the part H_1 in (33) that was neglected so far can assist the trapped hole in escaping from the string potential, see Figure 12. Namely, by acting on the first two sites of a string, the spins that were inverted by the hole are inverted a second time and thus fit with the Néel order again:

$$H_1 |i, i_1, i_2, i_3, \dots, i_\nu\rangle = \frac{J}{2} |i_2, i_3, \dots, i_\nu\rangle$$

The initial site of the string thus is shifted to a $(2, 0)$ - or $(1, 1)$ -like neighbor while simultaneously the length ν is increased or decreased by two. This suggests that there is a nonvanishing matrix element of H_1 between states $|\Psi_i\rangle$ and $|\Psi_j\rangle$, and it is straightforward to see that

$$\langle \Phi_i | H_1 | \Phi_{i+2\hat{x}} \rangle = J \sum_{\nu=0}^{\infty} (z-1)^\nu \alpha_\nu \alpha_{\nu+2} = J \cdot m,$$

$$\langle \Phi_i | H_1 | \Phi_{i+\hat{x}+\hat{y}} \rangle = 2J \cdot m.$$

The factor of 2 in the second matrix element is due to the fact that a string to a $(1, 1)$ -like neighbor can pass either through $(1, 0)$ or $(0, 1)$ and the contributions from these two different paths are additive.

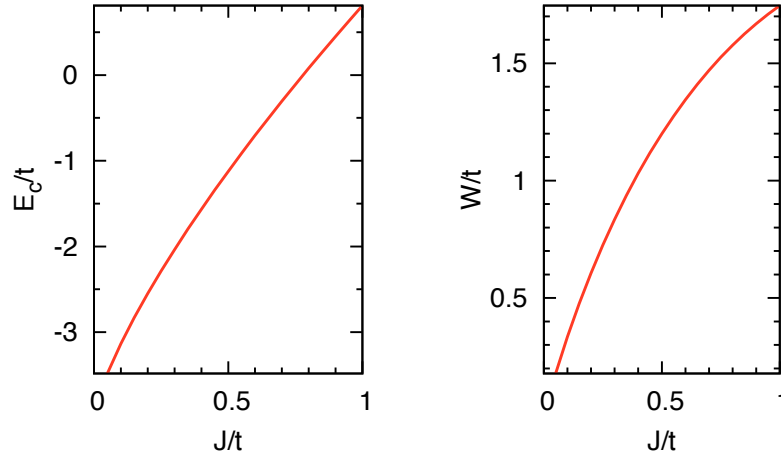


Fig. 13: Center of gravity E_c and bandwidth W of the quasiparticle band as obtained from the variational calculation in the string basis.

If the full Hamiltonian $H_0 + H_1$ is taken into account, the hole therefore can propagate through the entire lattice, and we describe this by the Bloch state

$$|\Psi_{\mathbf{k}}\rangle = \sqrt{\frac{2}{N}} \sum_{i \in A} e^{-i\mathbf{k} \cdot \mathbf{R}_j} |\Phi_i\rangle. \quad (39)$$

Since the matrix element for a $(1, 1)$ -like neighbor is twice that for a $(2, 0)$ -like neighbor we obtain the dispersion

$$\begin{aligned} E(\mathbf{k}) &= E_{\text{loc}} + 2Jm \cdot 4 \cos(k_x) \cos(k_y) + Jm \cdot 2(\cos(2k_x) + \cos(2k_y)) \\ &= E_{\text{loc}} + 4Jm (\cos(k_x) + \cos(k_y))^2 - 4Jm. \end{aligned}$$

This has a degenerate minimum along the line $k_y = \pi - k_x$ (i.e. $(\pi, 0) \rightarrow (0, \pi)$) and the symmetry equivalent lines, the maxima are at $(0, 0)$ and (π, π) . This is very similar to the dispersion in Figure 10 (note that the dispersion from SCB is upside down in this Figure!). The center of the band is $E_c = E_{\text{loc}} + 4Jm$, where E_{loc} is relatively large, of order $2-3t$. The width of the band is $W = 16mJ$. Both quantities are shown in Figure 13. At $J/t = 0.3$ we have $E_c = -2t$ and $W = 0.83t$ – in fact Figure 9 (a) shows that the center of the band is around $-2t$ and Figure 10 shows that the bandwidth is $\approx 0.7t$. The trial wave function (39) thus describes the single hole ground state that gives rise to the quasiparticle peak quite well.

The above derivation shows that due to the interaction with the spin waves the motion of the hole in an antiferromagnet is very different from free propagation: the hole executes a rapid ‘zig-zag’ motion around a given site i with hopping integral t under the influence of the string potential thereby forming the quasi-localized states $|\Psi_i\rangle$. The spin-flip part of the Hamiltonian then enables tunneling between such quasi-localized states with the much smaller energy scale $\propto J$, which accordingly determines the bandwidth of the quasiparticle.

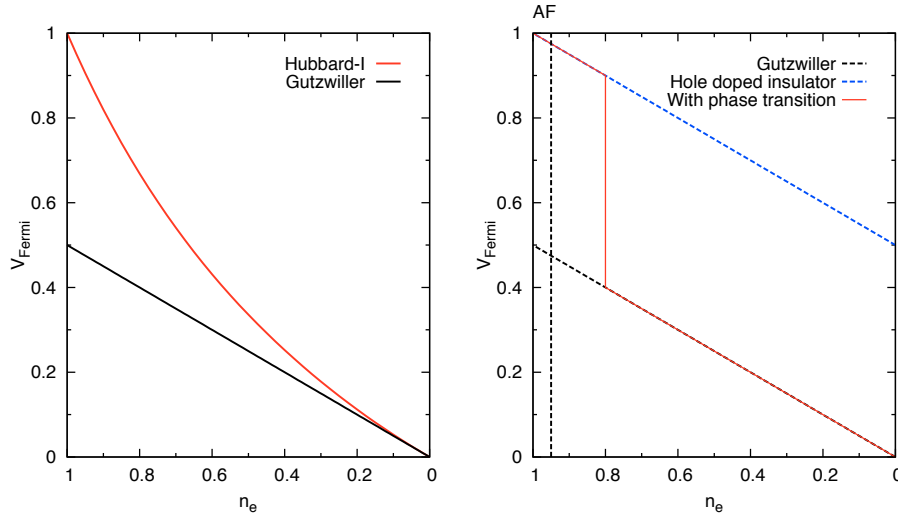


Fig. 14: Left: Fermi surface volume as obtained from the Hubbard-I approximation and the Gutzwiller wave function. Right: a possible compromise with a phase transition between two phases with different $V_{\text{Fermi}}(n_e)$.

7 Summary and discussion

Since a generally accepted theory of the lightly doped 2-dimensional Hubbard model does not 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 solved.

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 $n_e \rightarrow 1$ versus a free-electron-like Fermi surface with volume $n_e/2$ formed by a band whose mass diverges as $n_e \rightarrow 1$. It should be noted that for n_e very close to 1 antiferromagnetic order sets in, but superconductivity is observed at values of n_e where there is no more antiferromagnetism. The more relevant question therefore is, which Fermi surface is realized up to the onset of antiferromagnetism. A possible compromise between the two approximations could be as shown in Figure 14: near $n_e = 1$ but outside the antiferromagnetic doping range 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 one scenario which the Hubbard-I approximation ‘tries to reproduce.’ Viewed this way, the ‘pseudogap phase’ of cuprate superconductors could 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 transition first of all must reproduce the two Hubbard bands – otherwise the hole-doped lower Hubbard band cannot be reproduced. Next, the two different phases and the transition between them would have to be described, 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.

The second major problem is that – as we have seen above – the particles in the Hubbard bands are heavily renormalized due to their interaction with, mainly, spin excitations. The quasi-particle band produced by the self-consistent Born approximation or the string wave function describes a very complicated state in which the hole is heavily dressed by spin excitations. It is this strong dressing with spin waves which leads for example to the shift of the band maximum from (π, π) , where it is predicted by the Hubbard-I approximation, to $(\pi/2, \pi/2)$, where the self-consistent Born approximation puts it. It is likely that similar effects will also occur in the lightly doped Mott insulator, and clearly it is rather hopeless to try and reproduce these heavily renormalized quasiparticles by any ‘simple’ theory, such as a mean-field theory. From the calculation with the string states it appears that long-range antiferromagnetic order is not absolutely necessary – rather a similar formation of strings may already occur in a state that has antiferromagnetic correlations with a range of a few lattice spacings. However, for the situation without antiferromagnetic order but only short ranged antiferromagnetic correlations, there is no state analogous to the Néel state that would allow for a similarly simple development as in the single-hole theory.

A The bosonic Bogoliubov transformation

We consider the following Hamiltonian, whereby $a^\dagger$ and $b^\dagger$ are bosonic operators that satisfy $[a, a^\dagger] = [b, b^\dagger] = 1$:

$$H = \varepsilon (a^\dagger a + b^\dagger b) + V a^\dagger b^\dagger + V^* b a, \quad (40)$$

To solve this we make the ansatz

$$\gamma_1^\dagger = u a^\dagger + v b \quad (41)$$

$$\gamma_2^\dagger = u b^\dagger + v a \quad (42)$$

In order for the γ' s to fulfil the bosonic commutator relations $[\gamma_i, \gamma_i^\dagger] = 1$, we must have $|u|^2 - |v|^2 = 1$. The relation $[\gamma_1^\dagger, \gamma_2^\dagger] = 0$ on the other hand is fulfilled automatically. Moreover, we require that when expressed in terms of the γ' s the Hamiltonian should take the simple form

$$H = \sum_{i=1}^2 \omega \gamma_i^\dagger \gamma_i + \text{const}$$

which means that the γ' s obey

$$[H, \gamma_i^\dagger] = \omega \gamma_i^\dagger \quad (43)$$

We now insert (40) and $\gamma_1^\dagger$ from (42) into (43), use the commutator relations for $a^\dagger$ and $b^\dagger$, and equate the coefficients of $a^\dagger$ and b on both sides of (43). This gives the non-Hermitian eigenvalue problem

$$\begin{pmatrix} \varepsilon & -V \\ V^* & -\varepsilon \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \omega \begin{pmatrix} u \\ v \end{pmatrix}$$

which is easily solved to give

$$\begin{aligned} \omega &= \sqrt{\varepsilon^2 - |V|^2} \\ u &= \frac{V}{\sqrt{2\omega(\varepsilon - \omega)}} \\ v &= \sqrt{\frac{\varepsilon - \omega}{2\omega}} \end{aligned}$$

The eigenvalue problem has the property that if (u, v) is an eigenvector with eigenvalue ω then (v^*, u^*) is an eigenvector with eigenvalue $-\omega$. This guarantees that automatically $[H, \gamma_1] = -\omega \gamma_1$. Since H is symmetric under $a \leftrightarrow b$ the requirement $[H, \gamma_2^\dagger] = \omega \gamma_2^\dagger$ gives the same system of equations. To apply this to the spin wave Hamiltonian (28), we obviously need to set $\varepsilon = 1$ and $V = \gamma_{\mathbf{k}}$.

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] See the lecture by A. Mielke in this School.
- [6] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, San Francisco, 1971)
- [7] 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)
- [8] C. Gröber, R. Eder, and W. Hanke, Phys. Rev. B **62** 4336 (2000)
- [9] T. Ogawa, K. Kanda, and T. Matsubara, Progr. Theor. Phys. **53**, 614 (1975)
- [10] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970)
- [11] K.A. Chao, J. Spalek, and A.M. Oles, J. Phys. C **10**, 1271 (1972)
- [12] P.W. Anderson, Phys. Rev. **86**, 694 (1952)
- [13] J.D. Reger and A.P. Young, Phys. Rev. B **37**, 5978(R) (1988)
- [14] R. Coldea, S.M. Hayden, G. Aeppli, T.G. Perring, C.D. Frost, T.E. Mason, S.-W. Cheong, and Z. Fisk, Phys. Rev. Lett. **86**, 5377 (2001)
- [15] G. Martinez and P. Horsch, Phys. Rev. B **44**, 317 (1991)
- [16] P.W. Leung and R.J. Gooding Phys. Rev. B **52**, R15711(R) (1995)
- [17] B.O. Wells, Z.-X. Shen, A. Matsuura, D.M. King, M.A. Kastner, M. Greven, and R.J. Birgeneau, Phys. Rev. Lett. **74**, 964 (1995)
- [18] O.P. Sushkov, G.A. Sawatzky, R. Eder and H. Eskes, Phys. Rev. B. **56**, 11769 (1997)
- [19] R. Eder und K.W. Becker, Z. Phys. B **78**, 219 (1990)

13 Quantum Cluster Methods: CPT and CDMFT

David Sénéchal

Département de physique

Université de Sherbrooke, Québec, Canada

Contents

1	Introduction	2
1.1	The Hubbard model and Green functions	2
1.2	Clusters	6
1.3	Cluster Perturbation Theory	7
2	Periodic systems	8
2.1	Cluster kinematics	8
2.2	The CPT Green function and periodization	11
3	The exact diagonalization method	15
4	Cellular Dynamical Mean-Field Theory	16
4.1	The dynamical mean field	16
4.2	The equivalent Anderson impurity model	18
4.3	The self-consistency condition	20
4.4	Inhomogeneous systems	22
5	Applications	23
5.1	The Mott transition	23
5.2	Superconductivity	25
6	Extended interactions	28

1 Introduction

1.1 The Hubbard model and Green functions

Quantum cluster methods are schemes used to obtain approximate solutions of models of interacting electrons on a lattice. These models are used to describe classes of materials, such as high-temperature superconductors, in which electron-electron interactions are strong and for which the – otherwise successful – ideas behind Fermi liquid theory do not seem useful. The prototype of such models is the one-band Hubbard model. Other lectures in this volume offer extensive background material on this model; let us nevertheless write its Hamiltonian once more, in order to establish notation:

$$H = H_0 + H_1 \quad H_0 = \sum_{\mathbf{r}, \mathbf{r}', \sigma} t_{\mathbf{r}\mathbf{r}'} c_{\mathbf{r}\sigma}^\dagger c_{\mathbf{r}'\sigma} \quad H_1 = U \sum_{\mathbf{r}} n_{\mathbf{r}\uparrow} n_{\mathbf{r}\downarrow} \quad (1)$$

H_0 is the one-body term, which defines the band structure of the model, whereas H_1 is the electron-electron interaction. Sites of the Bravais lattice are indexed by the associated lattice vectors $\mathbf{r}$. The operator $c_{\mathbf{r}\sigma}$ destroys an electron in a Wannier orbital of spin projection σ centered at site $\mathbf{r}$. The number of electrons in that orbital is $n_{\mathbf{r}\sigma} = c_{\mathbf{r}\sigma}^\dagger c_{\mathbf{r}\sigma}$ and the total number of electrons at that site is $n_{\mathbf{r}} = n_{\mathbf{r}\uparrow} + n_{\mathbf{r}\downarrow}$. The Hermitian hopping matrix $t_{\mathbf{r}\mathbf{r}'}$ defines the band structure; its Fourier transform is the dispersion relation $\varepsilon(\mathbf{k})$:

$$t_{\mathbf{r}\mathbf{r}'} = \frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \varepsilon(\mathbf{k}) . \quad (2)$$

Periodic boundary conditions are used, and the number N of sites in the system is assumed to be very large. For convenience, we will include the chemical potential in the one-body part H_0 , i.e., $t_{\mathbf{r}\mathbf{r}} = -\mu$.

A more general model might include more than one band; a band index n is then needed to label one-body states: $(\mathbf{r}, \sigma) \rightarrow (n, \mathbf{r}, \sigma)$. The hopping matrix then becomes a more general hybridization matrix $t_{n\mathbf{r}, n'\mathbf{r}'}$, and there may be inter-band Coulomb interactions, Hund's couplings, and so on. The methods presented in this chapter may also be applied to such cases, but we will base our arguments on model (1), possibly augmented by longer-range Coulomb interactions (Sect. 6). In order to keep the discussion as general as possible, we will introduce a general index $\alpha = (n, \mathbf{r}, \sigma)$ that is a composite of position, spin and band indices. The first letters of the Greek alphabet ($\alpha, \beta, \dots$) will be used for that purpose.

The one-particle Green function A complete solution to model (1) at zero temperature would be provided by the many-body ground state $|\Omega\rangle$. Such an object, even if it were known, would be too unwieldy and would contain much more information than what is necessary to make useful predictions. We will instead seek approximate solutions for the one-particle Green function, defined at zero-temperature and as a function of complex frequency z as

$$G_{\alpha\beta}(z) = \left\langle \Omega \left| c_\alpha \frac{1}{z - H + E_0} c_\beta^\dagger \right| \Omega \right\rangle + \left\langle \Omega \left| c_\beta^\dagger \frac{1}{z + H - E_0} c_\alpha \right| \Omega \right\rangle, \quad (3)$$

where E_0 is the ground state energy associated with the Hamiltonian H , which, let us not forget, includes the chemical potential. $G_{\alpha\beta}(z)$ contains dynamical information about one-particle excitations, such as the spectral weight measured in ARPES experiments. We will generally use a boldface matrix notation ($\mathbf{G}$) for quantities carrying two one-body indices ($G_{\alpha\beta}$).

A finite-temperature expression for the Green function (3) is obtained by simply replacing the ground state expectation value by a thermal average. Practical computations at finite temperature are mostly done using Monte Carlo methods, which rely on the path integral formalism and are performed as a function of imaginary time, not directly as a function of real frequencies. In the limited scope of this chapter, we will confine ourselves to the zero-temperature formalism.

Green function in the time domain The expression (3) may be unfamiliar to those used to a definition of the Green function in the time domain. Let us just mention the connection. We define the spectral function in the time domain and its Fourier transform as

$$A_{\alpha\beta}(t) = \langle \{c_\alpha(t), c_\beta^\dagger(0)\} \rangle \quad A_{\alpha\beta}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} A_{\alpha\beta}(t) \quad (4)$$

where $\{\cdot, \cdot\}$ is the anticommutator and z is a complex frequency. The time dependence is defined in the Heisenberg picture, i.e., $c_\alpha(t) = e^{iHt} c_\alpha(0) e^{-iHt}$, where H includes the chemical potential. Then it can be shown that the Green function is related to $A_{\alpha\beta}(z)$ by

$$G_{\alpha\beta}(z) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{A_{\alpha\beta}(\omega)}{z - \omega} . \quad (5)$$

The retarded Green function $G_{\alpha\beta}^R(t)$ is defined in the time domain as

$$G_{\alpha\beta}^R(t) = -i\Theta(t) \langle \{c_\alpha(t), c_\beta^\dagger(0)\} \rangle = -i\Theta(t) A_{\alpha\beta}(t) \quad (6)$$

where $\Theta(t)$ is the Heaviside step function. Since the Fourier transform of the latter is

$$\mathcal{F}(\Theta)(\omega) = \int_0^{\infty} dt e^{i\omega t} = \frac{1}{\omega + i0^+} , \quad (7)$$

a simple convolution shows that

$$G_{\alpha\beta}^R(\omega) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{A_{\alpha\beta}(\omega')}{\omega - \omega' + i0^+} = G_{\alpha\beta}(\omega + i0^+) . \quad (8)$$

In fact, this connection can be established easily from the spectral representation, introduced next.

Spectral representation Let $\{|r\rangle\}$ be a complete set of eigenstates of H with one particle *more* than the ground state, where r is a positive integer label. Likewise, let us use negative integer labels to denote eigenstates of H with one particle *less* than the ground state. Then, by inserting completeness relations,

$$G_{\alpha\beta}(z) = \sum_{r>0} \langle \Omega | c_\alpha | r \rangle \frac{1}{z - E_r + E_0} \langle r | c_\beta^\dagger | \Omega \rangle + \sum_{r<0} \langle \Omega | c_\beta^\dagger | r \rangle \frac{1}{z + E_r - E_0} \langle r | c_\alpha | \Omega \rangle . \quad (9)$$

By setting

$$Q_{\alpha r} = \begin{cases} \langle \Omega | c_\alpha | r \rangle & (r > 0) \\ \langle r | c_\alpha | \Omega \rangle & (r < 0) \end{cases} \quad \text{and} \quad \omega_r = \begin{cases} E_r - E_0 & (r > 0) \\ E_0 - E_r & (r < 0) \end{cases} \quad (10)$$

we write

$$G_{\alpha\beta}(z) = \sum_r \frac{Q_{\alpha r} Q_{\beta r}^*}{z - \omega_r} . \quad (11)$$

This shows how the Green function is a sum over poles located at $\omega_r \in \mathbb{R}$, with residues that are products of overlaps of the ground state with energy eigenstates with one more ($\omega_r > 0$) or one less ($\omega_r < 0$) particle. The sum of residues is normalized to the unit matrix, as can be seen from the anticommutation relations:

$$\begin{aligned} \sum_r Q_{\alpha r} Q_{\beta r}^* &= \sum_{r>0} \langle \Omega | c_\alpha | r \rangle \langle r | c_\beta^\dagger | \Omega \rangle + \sum_{r<0} \langle \Omega | c_\beta^\dagger | r \rangle \langle r | c_\alpha | \Omega \rangle \\ &= \langle \Omega | (c_\alpha c_\beta^\dagger + c_\beta^\dagger c_\alpha) | \Omega \rangle = \delta_{\alpha\beta} . \end{aligned} \quad (12)$$

Thus, in the high-frequency limit, $\mathbf{G}(z \rightarrow \infty) = \mathbf{1}/z$ ($\mathbf{1}$ stands for the unit matrix).

The same procedure applied to the spectral function (4) leads easily to

$$A_{\alpha\beta}(\omega) = 2\pi \sum_r Q_{\alpha r} Q_{\beta r}^* \delta(\omega - \omega_r) , \quad (13)$$

and this demonstrates the connection (5) between $A_{\alpha\beta}(\omega)$ and $G_{\alpha\beta}(z)$. The property (12) amounts to saying that $A_{\alpha\alpha}(\omega)$ is a probability density:

$$A_{\alpha\alpha}(\omega) = 2\pi \sum_r |Q_{\alpha r}|^2 \delta(\omega - \omega_r) \quad \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A_{\alpha\alpha}(\omega) = 1 \quad (14)$$

The identity

$$-\frac{1}{\pi} \text{Im} \frac{1}{\omega + i0^+} = \delta(\omega) \quad (15)$$

implies that

$$A_{\alpha\alpha}(\omega) = -2 \text{Im} G_{\alpha\alpha}(\omega + i0^+) . \quad (16)$$

From the definition of $Q_{\alpha r}$, one sees that $A_{\alpha\alpha}(\omega)$ is the probability density for an electron added or removed from the ground state in the one-particle state α to have an energy ω . The density of states $\rho(\omega)$ is simply the trace

$$\rho(\omega) = \frac{1}{N} \sum_\alpha A_{\alpha\alpha}(\omega) = -\frac{2}{N} \text{Im} \text{tr} \mathbf{G}(\omega + i0^+) . \quad (17)$$

Self-energy In the absence of interactions ($H_1 = 0$) the Hamiltonian reduces to $H_0 = \sum_{\alpha,\beta} t_{\alpha\beta} c_\alpha^\dagger c_\beta$. Since the matrix $\mathbf{t}$ is Hermitian, there exists a basis $\{|\ell\rangle\}$ of one-body states that diagonalizes it: $H_0 = \sum_\ell \varepsilon_\ell c_\ell^\dagger c_\ell$. The ground state is then the filled Fermi sea:

$$|\Omega\rangle = \prod_{\varepsilon_\ell < 0} c_\ell^\dagger |0\rangle \quad (18)$$

and one-particle excited states are $c_\ell^\dagger|\Omega\rangle$ ($\varepsilon_\ell > 0$) with $E_\ell - E_0 = \varepsilon_\ell$ and $c_\ell|\Omega\rangle$ ($\varepsilon_\ell < 0$) with $E_\ell - E_0 = -\varepsilon_\ell$. The spectral representation is in that case extremely simple and the matrix $\mathbf{G} = \mathbf{G}_0$ is diagonal:

$$G_{0,\ell\ell'}(z) = \frac{\delta_{\ell\ell'}}{z - \varepsilon_\ell}. \quad (19)$$

In any other basis of one-body states in which $\mathbf{t}$ is not diagonal, the expression is simply

$$\mathbf{G}_0(z) = \frac{1}{z - \mathbf{t}}. \quad (20)$$

In the presence of interactions, the Green function takes the following general form:

$$\mathbf{G}(z) = \frac{1}{z - \mathbf{t} - \mathbf{\Sigma}(z)}, \quad (21)$$

where all the information related to H_1 is buried within the self-energy $\mathbf{\Sigma}(z)$. The relation (21), called *Dyson's equation*, may be regarded as a definition of the self-energy. It can be shown that the self-energy has a spectral representation similar to that of the Green function:

$$\Sigma_{\alpha\beta}(z) = \Sigma_{\alpha\beta}^\infty + \sum_r \frac{S_{\alpha r} S_{\beta r}^*}{z - \sigma_r}, \quad (22)$$

where the σ_r are poles located on the real axis (they are zeros of the Green function). By contrast with the Green function, the self-energy may have a frequency-independent piece $\Sigma_{\alpha\beta}^\infty$, which has the same effect as a hopping term; in fact, within the Hartree-Fock approximation, this is the only piece of the self-energy that survives.

Averages of one-body operators Many physical observables are one-body operators, of the form

$$\mathcal{O} = \sum_{\alpha,\beta} s_{\alpha\beta} c_\alpha^\dagger c_\beta. \quad (23)$$

The ground-state expectation value of such operators can be computed from the Green function $G_{\alpha\beta}(z)$. Let us explain how.

From the spectral representation (11) of the Green function, we see that $\langle c_\alpha^\dagger c_\beta \rangle$ is given by the integral of the Green function along a contour $C_<$ surrounding the negative real axis counter-clockwise:

$$\langle c_\alpha^\dagger c_\beta \rangle = \int_{C_<} \frac{dz}{2\pi i} G_{\beta\alpha}(z). \quad (24)$$

Therefore the expectation value we are looking for is

$$\bar{\mathcal{O}} = \frac{1}{N} \sum_{\alpha,\beta} s_{\alpha\beta} \langle c_\alpha^\dagger c_\beta \rangle = \frac{1}{N} \int_{C_<} \frac{dz}{2\pi i} \text{tr}[\mathbf{s} \mathbf{G}(z)] \quad (25)$$

(we divide by N to find an intensive quantity). The trace includes a sum over lattice sites, spin and band indices.

The contour $C_<$ can be taken as the imaginary axis (from $-iR$ to iR), plus the left semi-circle of radius R . Since $\mathbf{G}(z) \rightarrow \mathbf{1}/z$ as $z \rightarrow \infty$, the semi-circular part will contribute, but this

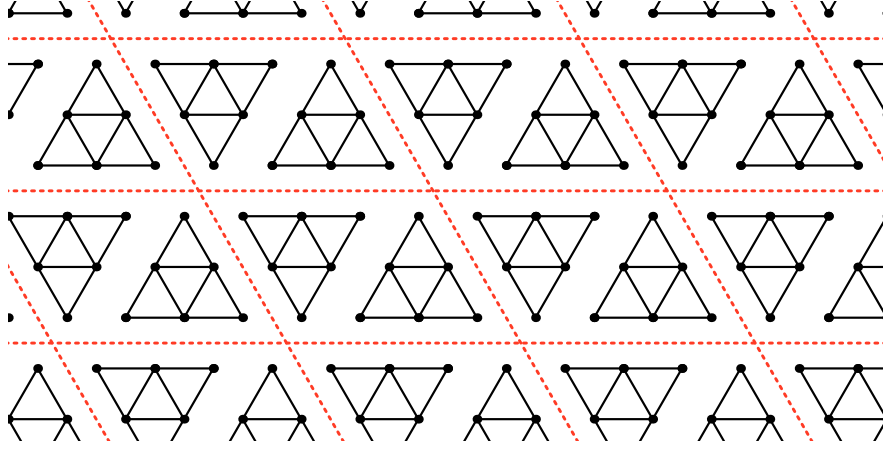


Fig. 1: Tiling of the triangular lattice by 6-site triangular clusters. The super-cells are delimited by dashed lines.

contribution may be canceled by subtracting from $\mathbf{G}(z)$ a term like $1/(z - p)$, with $p > 0$: the added term does not contribute to the integral, since its only pole lies outside the contour, yet it cancels the dominant z^{-1} behavior as $z \rightarrow \infty$, leaving a contribution that vanishes on the semi-circle as $R \rightarrow \infty$. We are left with

$$\bar{\mathcal{O}} = \frac{1}{N} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left\{ \text{tr} [\mathbf{s} \mathbf{G}(i\omega)] - \frac{\text{tr} \mathbf{s}}{i\omega - p} \right\}. \quad (26)$$

If the operator $\mathcal{O}$ is Hermitian, so is the matrix $\mathbf{s}$. By virtue of the property $\mathbf{G}(z)^\dagger = \mathbf{G}(z^*)$, easily seen from (11), we have $\text{tr} [\mathbf{s} \mathbf{G}(-i\omega)] = \text{tr} [\mathbf{s} \mathbf{G}(i\omega)]^*$; this implies that $\bar{\mathcal{O}}$ is real.

1.2 Clusters

How can we compute $\mathbf{G}(z)$ if the many-body ground state $|\Omega\rangle$ is not known? Quantum cluster methods provide an approximate solution by dividing the original system into smaller parts. The original lattice γ is tiled into small, manageable and disconnected clusters. For instance, the triangular lattice may be tiled by 6-site clusters, as illustrated in Fig. 1; in that case, two distinct clusters are included in each repeated unit, or *super-cell*. On a cluster labeled j , a Hamiltonian $H^{(j)}$ is defined, whose interaction part $H_1^{(j)}$ coincides with that of the original problem:

$$H_1 = \sum_j H_1^{(j)}, \quad (27)$$

but whose one-body part $H_0^{(j)}$ will depend on the particular quantum cluster method used. In order to stay general, we have left open the possibility that the cluster Hamiltonians $H^{(j)}$ are all different from one another, for instance because of a position-dependent potential added to the basic Hubbard model. However, in most cases, they will all be identical to one or a few clusters forming a repeated super-cell.

Each cluster's one-particle Green function $\mathbf{G}^{(j)}(z)$ needs to be computed. Various numerical methods may be harnessed for this task. At zero temperature, exact diagonalization techniques

are generally used; this entails computing the many-body ground state $|\Omega^{(j)}\rangle$ of $H^{(j)}$ and applying the definition (3) to find $\mathbf{G}^{(j)}$, as explained summarily in Sect. 3.

Then the self-energy $\Sigma^{(j)}(z)$ associated with $\mathbf{G}^{(j)}(z)$ is extracted from Dyson's equation, and the following approximation for the lattice self-energy is assembled:

$$\Sigma = \begin{pmatrix} \Sigma^{(1)} & 0 & \dots & 0 \\ 0 & \Sigma^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma^{(n)} \end{pmatrix}. \quad (28)$$

This equation defines the basic assumption behind quantum cluster methods: the self-energy can be approximated by the direct sum of the self-energies of all clusters. The lattice Green function is then constructed from the Dyson equation (21). Note that in the simple case of a super-cell made of a single cluster, all self-energies $\Sigma^{(j)}$ are identical.

Many relevant physical properties of the model can be extracted from the one-particle Green function, as shown in Eq. (26), but not all. Other functions of interest relevant to experiments are the dynamical susceptibilities, where the creation and annihilation operators of Eq. (3) are replaced by one-body operators, such as the spin or electron densities. However, the methods described in this chapter will not provide us with approximate ways to compute these properties, beyond computing them within each cluster.

1.3 Cluster Perturbation Theory

The simplest of all quantum cluster methods is *Cluster Perturbation Theory* (CPT) [1, 2]. In CPT, each cluster's one-body Hamiltonian $H_0^{(j)}$ is simply the restriction to the cluster of the full one-body Hamiltonian H_0 . If hopping terms connecting sites located on cluster i to those of cluster j are collected into a matrix $\mathbf{t}^{(i,j)}$, then the full one-body matrix may be expressed as

$$\mathbf{t} = \begin{pmatrix} \mathbf{t}^{(1,1)} & \mathbf{t}^{(1,2)} & \dots & \mathbf{t}^{(1,n)} \\ \mathbf{t}^{(2,1)} & \mathbf{t}^{(2,2)} & \dots & \mathbf{t}^{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{t}^{(n,1)} & \mathbf{t}^{(n,2)} & \dots & \mathbf{t}^{(n,n)} \end{pmatrix} \quad (29)$$

The one-body matrix defining $H_0^{(j)}$ in CPT is then simply the diagonal block $\mathbf{t}^{(j,j)}$ and each cluster's interacting Green function obeys the relation

$$\mathbf{G}^{(j)-1}(z) = z - \mathbf{t}^{(j,j)} - \Sigma^{(j)}(z) \quad (30)$$

The basic approximation (28), when combined with Eqs (21) and (29), leads to the following formula for the approximate, or CPT, Green function:

$$\mathbf{G}_{\text{cpt}}^{-1}(z) = \bigoplus_j \mathbf{G}^{(j)-1}(z) - \mathbf{t}_{\text{ic}}, \quad (31)$$

where $\oplus$ stands for the direct sum, and the matrix t_{ic} is obtained from t in Eq. (29) by removing all diagonal blocks; it is the inter-cluster hopping matrix.

One may collectively denote by H' the sum of cluster Hamiltonians and by G' the direct sum of cluster Green functions:

$$H' = \sum_j H^{(j)} \quad G'(z) = \bigoplus_j G^{(j)}(z) \quad (32)$$

We may then write simpler-looking formulas:

$$H = H' + \sum_{\alpha,\beta} (t_{ic})_{\alpha\beta} c_{\alpha}^{\dagger} c_{\beta} \quad (33)$$

$$G_{\text{cpt}}^{-1}(z) = G'^{-1}(z) - t_{ic} \quad (34)$$

Cluster Perturbation Theory is called this way because it can be derived by treating the second term of Eq. (33) as a perturbation on H' . It can be shown that, at lowest order in t_{ic} , the Green function is indeed given by Eq. (34) [2, 3].

Cluster Perturbation Theory has the following characteristics:

1. Although it is derived using strong-coupling perturbation theory, it is exact in the $U \rightarrow 0$ limit, since the self-energy disappears in that case.
2. It is also exact in the strong-coupling limit $t_{rr'}/U \rightarrow 0$.
3. It provides an approximate lattice Green function for arbitrary wave-vectors, as explained in Sect. 2.2 below, hence its usefulness in comparing with ARPES data.
4. Although formulated as a lowest-order result of strong-coupling perturbation theory, it is not controlled by including higher-order terms in that perturbation expansion – this would be extremely difficult – but rather by increasing the cluster size.
5. It cannot describe broken-symmetry states. This is accomplished by more sophisticated approaches like the Variational Cluster Approximation (VCA) and Cluster Dynamical Mean Field Theory (CDMFT), which can both be viewed as extensions or refinements of CPT. But even in these approaches, formula (34) still applies. The difference lies in the use of different cluster Hamiltonians $H_0^{(j)}$ and therefore different cluster Green functions $G^{(j)}$.

2 Periodic systems

2.1 Cluster kinematics

Typically, the clusters that ‘tile’ the lattice are repeated: one or a few of them form a repeated *super-cell*, like the two 6-site clusters of Fig. 1. Mathematically, this corresponds to introducing a super-lattice Γ , whose sites form a subset of the original lattice γ that will be labeled by the positions $\tilde{r}$. This super-lattice is generated by basis vectors $\{e_1, e_2, e_3\}$: every site $\tilde{r}$ of the super-lattice may be expressed as an integer combination of these basis vectors. Associated with each site of Γ is a super-cell containing L sites. The super-cell is made of one cluster, or

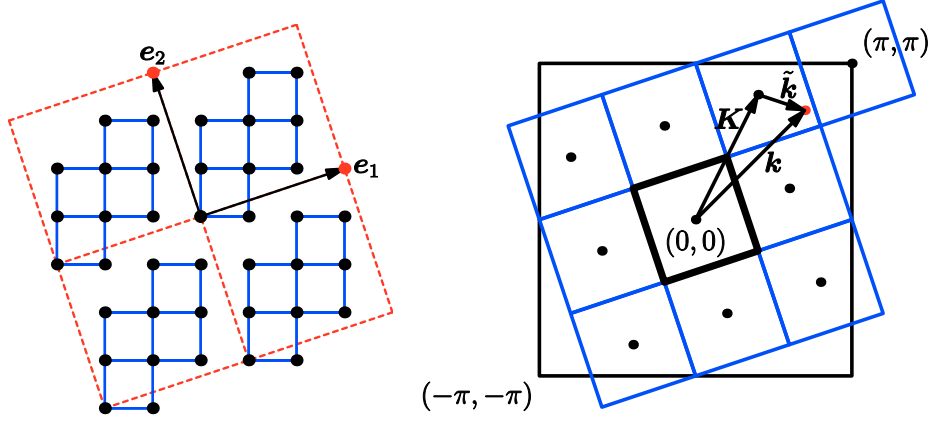


Fig. 2: Left panel: Tiling of the square lattice with identical ten-site clusters ($L = 10$). The vectors $e_{1,2}$ define a super-lattice of clusters. Right panel: the corresponding Brillouin zones. The reduced Brillouin zone (tilted black square) is associated with the super-lattice and L copies of it can be fitted within the original Brillouin zone (large square).

sometimes of a few clusters, as in Fig. 1. Note that the shape of these clusters is not uniquely determined by the super-lattice structure. The sites within each super-cell will be labeled by their vector position (in capitals): $\mathbf{R}$, $\mathbf{R}'$, etc. Each site $\mathbf{r}$ of the original lattice γ can therefore be expressed uniquely as $\mathbf{r} = \tilde{\mathbf{r}} + \mathbf{R}$.

The Brillouin zone of the original lattice, denoted BZ_γ , contains L points belonging to the reciprocal super-lattice Γ^* . The Brillouin zone of the super-lattice BZ_Γ has a volume L times smaller than that of BZ_γ . Any wave-vector $\mathbf{k}$ of the original Brillouin zone can be uniquely expressed as $\mathbf{k} = \mathbf{K} + \tilde{\mathbf{k}}$, where $\mathbf{K}$ belongs both to the reciprocal super-lattice and to BZ_γ , and $\tilde{\mathbf{k}}$ belongs to BZ_Γ (see Fig. 2).

The passage between momentum space and real space, via discrete Fourier transforms can be done either directly ($\mathbf{r} \leftrightarrow \mathbf{k}$) or independently for cluster and super-lattice sites ($\tilde{\mathbf{r}} \leftrightarrow \tilde{\mathbf{k}}$ and $\mathbf{R} \leftrightarrow \mathbf{K}$). This can be encoded into unitary matrices U^γ , U^Γ and U^c defined as follows:

$$U_{\mathbf{k},\mathbf{r}}^\gamma = \frac{1}{\sqrt{N}} e^{-i\mathbf{k}\cdot\mathbf{r}}, \quad U_{\tilde{\mathbf{k}},\tilde{\mathbf{r}}}^\Gamma = \sqrt{\frac{L}{N}} e^{-i\tilde{\mathbf{k}}\cdot\tilde{\mathbf{r}}}, \quad U_{\mathbf{K},\mathbf{R}}^c = \frac{1}{\sqrt{L}} e^{-i\mathbf{K}\cdot\mathbf{R}}, \quad (35)$$

where again N is the (large) number of sites in the original lattice γ , which can be treated with periodic boundary conditions at the edges. The matrix U^γ is of order N , U^Γ is of order N/L and U^c of order L .

The discrete Fourier transforms on a generic one-index quantity f are then

$$f(\mathbf{k}) = \sum_{\mathbf{r}} U_{\mathbf{k},\mathbf{r}}^\gamma f_{\mathbf{r}} \quad , \quad f(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} U_{\tilde{\mathbf{k}},\tilde{\mathbf{r}}}^\Gamma f_{\tilde{\mathbf{r}}} \quad , \quad f_{\mathbf{K}} = \sum_{\mathbf{R}} U_{\mathbf{K},\mathbf{R}}^c f_{\mathbf{R}} \quad (36)$$

or, in reverse,

$$f_{\mathbf{r}} = \sum_{\mathbf{k}} U_{\mathbf{k},\mathbf{r}}^{\gamma*} f(\mathbf{k}) \quad , \quad f_{\tilde{\mathbf{r}}} = \sum_{\tilde{\mathbf{k}}} U_{\tilde{\mathbf{k}},\tilde{\mathbf{r}}}^{\Gamma*} f(\tilde{\mathbf{k}}) \quad , \quad f_{\mathbf{R}} = \sum_{\mathbf{K}} U_{\mathbf{K},\mathbf{R}}^{c*} f_{\mathbf{K}} \quad (37)$$

These discrete Fourier transforms close by virtue of the following identities

$$\frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} = \delta_{\mathbf{r}} \quad \frac{L}{N} \sum_{\tilde{\mathbf{k}}} e^{i\tilde{\mathbf{k}} \cdot \tilde{\mathbf{r}}} = \delta_{\tilde{\mathbf{r}}} \quad \frac{1}{L} \sum_{\mathbf{K}} e^{i\mathbf{K} \cdot \mathbf{R}} = \delta_{\mathbf{R}} \quad (38)$$

$$\frac{1}{N} \sum_{\mathbf{r}} e^{-i\mathbf{k} \cdot \mathbf{r}} = \delta_{\mathbf{k}} \quad \frac{L}{N} \sum_{\tilde{\mathbf{r}}} e^{-i\tilde{\mathbf{k}} \cdot \tilde{\mathbf{r}}} = \delta_{\tilde{\mathbf{k}}} \quad \frac{1}{L} \sum_{\mathbf{R}} e^{-i\mathbf{K} \cdot \mathbf{R}} = \delta_{\mathbf{K}} \quad (39)$$

where $\delta_{\mathbf{r}}$ is the usual Kronecker delta, used for all labels (since they are all discrete):

$$\delta_{\alpha} = \begin{cases} 1 & \text{if } \alpha = 0 \\ 0 & \text{otherwise} \end{cases} \quad \delta_{\alpha\beta} \equiv \delta_{\alpha-\beta} , \quad (40)$$

It is implicit that the Kronecker deltas are periodic, i.e., that $\delta_{\tilde{\mathbf{k}}} = \delta_{\tilde{\mathbf{k}}+\mathbf{K}}$, for instance.¹

A one-index quantity like the destruction operator $c_{\mathbf{r}} = c_{\tilde{\mathbf{r}}+\mathbf{R}}$ can be represented in a variety of ways through partial Fourier transforms:

$$\begin{aligned} c_{\mathbf{R}}(\tilde{\mathbf{k}}) &= \sum_{\tilde{\mathbf{r}}} U_{\tilde{\mathbf{k}}\tilde{\mathbf{r}}}^{\Gamma} c_{\tilde{\mathbf{r}}+\mathbf{R}} & c_{\mathbf{K}}(\tilde{\mathbf{k}}) &= \sum_{\tilde{\mathbf{r}}, \mathbf{R}} U_{\tilde{\mathbf{k}}\tilde{\mathbf{r}}}^{\Gamma} U_{\mathbf{K}\mathbf{R}}^c c_{\tilde{\mathbf{r}}+\mathbf{R}} \\ c_{\tilde{\mathbf{r}}, \mathbf{K}} &= \sum_{\mathbf{R}} U_{\mathbf{K}\mathbf{R}}^c c_{\tilde{\mathbf{r}}+\mathbf{R}} & c(\mathbf{k}) &= \sum_{\mathbf{r}} U_{\mathbf{k}\mathbf{r}}^{\gamma} c_{\mathbf{r}} \end{aligned} \quad (41)$$

Note that

$$\mathbf{k} \cdot \mathbf{r} = (\tilde{\mathbf{k}} + \mathbf{K}) \cdot (\tilde{\mathbf{r}} + \mathbf{R}) = \tilde{\mathbf{k}} \cdot \tilde{\mathbf{r}} + \mathbf{K} \cdot \mathbf{R} + \tilde{\mathbf{k}} \cdot \mathbf{R} + \mathbf{K} \cdot \tilde{\mathbf{r}} . \quad (42)$$

By definition, the last term is a multiple of 2π ($\mathbf{K}$ is an element of the reciprocal lattice γ^*). Therefore the two representations $c(\mathbf{k})$ and $c_{\mathbf{K}}(\tilde{\mathbf{k}})$ are not identical, since the phases involved in (41), $\mathbf{k} \cdot \mathbf{r}$ and $\tilde{\mathbf{k}} \cdot \tilde{\mathbf{r}} + \mathbf{K} \cdot \mathbf{R}$, differ by $\tilde{\mathbf{k}} \cdot \mathbf{R}$. These two representations are obtained respectively by applying the unitary matrices $\mathbf{S} \equiv \mathbf{U}^{\Gamma} \otimes \mathbf{U}^c$ and $\mathbf{U}^{\gamma}$ on the $\mathbf{r}$ basis, and these two operations are different since the $N \times N$ matrix $\mathbf{A} \equiv \mathbf{U}^{\gamma} \mathbf{S}^{-1}$ is not trivial:

$$\Lambda_{\mathbf{k}\mathbf{k}'} = \delta_{\tilde{\mathbf{k}}\tilde{\mathbf{k}}} \Lambda_{\mathbf{K}\mathbf{K}'}^c(\tilde{\mathbf{k}}) \quad \text{where} \quad \Lambda_{\mathbf{K}\mathbf{K}'}^c(\tilde{\mathbf{k}}) = \frac{1}{L} \sum_{\mathbf{R}} e^{-i\mathbf{R} \cdot (\tilde{\mathbf{k}} + \mathbf{K} - \mathbf{K}')} . \quad (43)$$

The matrix $\Lambda^c(\tilde{\mathbf{k}})$ is $L \times L$ and connects the $(\mathbf{K}, \tilde{\mathbf{k}})$ basis to the $\mathbf{k} = \tilde{\mathbf{k}} + \mathbf{K}$ basis:

$$c(\tilde{\mathbf{k}} + \mathbf{K}) = \sum_{\mathbf{K}'} \Lambda_{\mathbf{K}\mathbf{K}'}^c(\tilde{\mathbf{k}}) c_{\mathbf{K}'}(\tilde{\mathbf{k}}) . \quad (44)$$

A two-index quantity like the hopping matrix $t_{\mathbf{r}\mathbf{r}'}$ or the Green function $G_{\mathbf{r}\mathbf{r}'}$ has a number of different representations. The first index transforms like $c_{\mathbf{r}}$ and the second like $c_{\mathbf{r}'}^{\dagger}$. For instance,

$$t(\mathbf{k}, \mathbf{k}') = \sum_{\mathbf{r}, \mathbf{r}'} U_{\mathbf{k}\mathbf{r}}^{\gamma} U_{\mathbf{k}'\mathbf{r}'}^{\gamma*} t_{\mathbf{r}\mathbf{r}'} . \quad (45)$$

¹Such periodic Kronecker deltas are sometimes called Laue functions.

Due to translation invariance on the lattice, this matrix is diagonal when expressed in momentum space: $t(\mathbf{k}, \mathbf{k}') = \varepsilon(\mathbf{k}) \delta_{\mathbf{k}, \mathbf{k}'}$, $\varepsilon(\mathbf{k})$ being the dispersion relation (2). However, in practice we most often use the mixed representation

$$t_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{i\tilde{\mathbf{k}} \cdot \tilde{\mathbf{r}}} t_{\mathbf{r}\mathbf{r}'} \quad \text{where} \quad \mathbf{r} = \mathbf{R} \quad \text{and} \quad \mathbf{r}' = \tilde{\mathbf{r}} + \mathbf{R}' . \quad (46)$$

For instance, if we tile the one-dimensional lattice with clusters of length $L = 2$, the nearest-neighbor hopping matrix, corresponding to the dispersion relation $\varepsilon(k) = -2t \cos(k) - \mu$, has the following mixed representation:

$$\mathbf{t}(\tilde{k}) = - \begin{pmatrix} \mu & t(1 + e^{-2i\tilde{k}}) \\ t(1 + e^{2i\tilde{k}}) & \mu \end{pmatrix} . \quad (47)$$

Finally, let us point out that the space E of one-electron states is larger than the space of lattice sites γ , as it also includes spin and maybe band degrees of freedom, which form a set B . We could therefore write $E = \gamma \otimes B$. The transformation matrices defined above ($\mathbf{U}^\gamma$, $\mathbf{U}^\Gamma$ and $\mathbf{U}^c$) should, as necessary, be understood as tensor products ($\mathbf{U}^\gamma \otimes \mathbf{1}$, $\mathbf{U}^\Gamma \otimes \mathbf{1}$ and $\mathbf{U}^c \otimes \mathbf{1}$) acting trivially in B . This should be clear from the context. The total number of degrees of freedom in the super-cell is therefore an integer multiple of L , which we shall denote by M_{sc} .

2.2 The CPT Green function and periodization

The most convenient representation for periodic systems is the $(\mathbf{R}, \tilde{\mathbf{k}})$ scheme, which uses real-space indices in the super-cell and reduced wave-vectors. Because of translational invariance, the inter-cluster hopping matrix of Eq. (34) is diagonal in $\tilde{\mathbf{k}}$ and t_{ic} , becoming effectively a M_{sc} -dimensional, $\tilde{\mathbf{k}}$ -dependent matrix $\mathbf{t}_{\text{ic}}(\tilde{\mathbf{k}})$. If the super-cell contains more than one cluster, like the example of Fig. 1, the matrix $\mathbf{t}_{\text{ic}}$ also contains $\tilde{\mathbf{k}}$ -independent terms from the hopping terms between those. Likewise, the restriction of $\mathbf{G}'$ (32) to the super-cell is M_{sc} -dimensional, the same for all super-cells, and therefore $\tilde{\mathbf{k}}$ -independent. We will also denote it by $\mathbf{G}'$, even though this is a slight abuse of notation. Therefore, Eq. (34) becomes

$$\mathbf{G}_{\text{cpt}}^{-1}(z, \tilde{\mathbf{k}}) = \mathbf{G}'^{-1}(z) - \mathbf{t}_{\text{ic}}(\tilde{\mathbf{k}}) \quad (48)$$

In that relation, all matrices are of size M_{sc} . The super-cell Green function $\mathbf{G}'$ is either a single cluster Green function $\mathbf{G}^{(j)}$, directly computed from the impurity solver, or a direct sum of the cluster Green functions making up the super-cell.

Relation (48) is the most convenient way to compute the CPT Green function. The cluster self-energies do not need to be extracted explicitly.

A supplemental ingredient of CPT is the *periodization* formula, which provides a fully $\mathbf{k}$ -dependent Green function out of the mixed representation $\mathbf{G}_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z)$. It was proposed in Ref. [2] to define the following *periodized* Green function:²

$$G_{\text{per.}}(\mathbf{k}, z) = \frac{1}{L} \sum_{\mathbf{R}, \mathbf{R}'} e^{-i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')} G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z) . \quad (49)$$

²In the following the spin and band indices are muted: the left hand side is still a matrix in those indices, but we will focus here on the spatial and wave-vector indices only, in order to lighten the notation.

Let us explain. Treating intra-cluster and inter-cluster hopping terms differently breaks the original translational symmetry of the model: The Green function (48) is not translationally invariant on the original lattice γ . This means that it is not diagonal when expressed in the $\mathbf{k}$ -scheme, i.e., $G(\mathbf{k}, \mathbf{k}') \neq 0$ if $\mathbf{k} \neq \mathbf{k}'$. However, because of the residual super-lattice translational invariance, $\mathbf{k}'$ and $\mathbf{k}$ must correspond to the same $\tilde{\mathbf{k}}$ and differ by an element of the reciprocal super-lattice: $\mathbf{k}' = \mathbf{k} + \mathbf{K}$. Thus, in the $(\mathbf{K}, \tilde{\mathbf{k}})$ basis, the matrix $\mathbf{G}$ has the following form:

$$G_{\mathbf{K}\mathbf{K}'}(\tilde{\mathbf{k}}, z) = \frac{1}{L} \sum_{\mathbf{R}, \mathbf{R}'} e^{-i(\mathbf{K} \cdot \mathbf{R} - \mathbf{K}' \cdot \mathbf{R}')} G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z). \quad (50)$$

Since an element of the reduced Brillouin zone is defined up to a vector belonging to the reciprocal super-lattice Γ^* , one may replace $\tilde{\mathbf{k}}$ by $\mathbf{k}$ in $G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z)$, i.e., $G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z) = G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}} + \mathbf{K}, z)$. This form can be further converted to the full wave-vector basis ($\mathbf{k} = \mathbf{K} + \tilde{\mathbf{k}}$) by use of the unitary matrix Λ^c of Eq (43):

$$\begin{aligned} G(\tilde{\mathbf{k}} + \mathbf{K}, \tilde{\mathbf{k}} + \mathbf{K}') &= \left(\Lambda^c(\tilde{\mathbf{k}}) \mathbf{G} \Lambda^{c\dagger}(\tilde{\mathbf{k}}) \right)_{\mathbf{K}\mathbf{K}'} \\ &= \frac{1}{L^2} \sum_{\mathbf{R}, \mathbf{R}', \mathbf{K}_1, \mathbf{K}'_1} e^{-i(\tilde{\mathbf{k}} + \mathbf{K} - \mathbf{K}_1) \cdot \mathbf{R}} e^{i(\tilde{\mathbf{k}} + \mathbf{K}' - \mathbf{K}'_1) \cdot \mathbf{R}'} G_{\mathbf{K}_1 \mathbf{K}'_1} \\ &= \frac{1}{L} \sum_{\mathbf{R}, \mathbf{R}'} e^{-i(\tilde{\mathbf{k}} + \mathbf{K}) \cdot \mathbf{R}} e^{i(\tilde{\mathbf{k}} + \mathbf{K}') \cdot \mathbf{R}'} G_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}, z). \end{aligned} \quad (51)$$

The periodization formula (49) amounts to picking the diagonal piece of the Green function ($\mathbf{K} = \mathbf{K}'$, or $\mathbf{k} = \mathbf{k}'$) and discarding the rest. This makes sense since the density of states $\rho(\omega)$ is the trace of the imaginary part of the Green function:

$$\rho(\omega) = -\frac{2}{N} \text{Im} \sum_{\mathbf{r}} G_{\mathbf{r}\mathbf{r}}(\omega + i0^+) = -\frac{2}{N} \text{Im} \sum_{\mathbf{k}} G(\mathbf{k}, \omega + i0^+), \quad (52)$$

and the spectral function $A(\mathbf{k}, \omega)$, as a partial trace, involves only the diagonal part. Moreover, because of the sum rule (12), which is basis independent, the frequency integral of the imaginary part of the off-diagonal components of the Green function vanishes.

Another possible formula for periodization is to apply relation (49) to the self-energy Σ instead. This is appealing since Σ is an irreducible quantity, as opposed to $\mathbf{G}$, and amounts to throwing out the off-diagonal components of Σ before applying Dyson's equation to get $\mathbf{G}$, as opposed to discarding the off-diagonal part at the last step, once the matrix inversion towards $\mathbf{G}$ has taken place. Unfortunately, this turns out not to work, which is not surprising given the nonlinear relation between Σ and the spectral function. As Fig. 3 shows, periodizing the Green function (Eq. (49)) reproduces the expected features of the spectral function of the one-dimensional Hubbard model: In particular, the Mott gap that opens at arbitrarily small U (as known from the exact solution). On the other hand, periodizing the self-energy leaves spectral weight within the Mott gap for an arbitrarily large value of U . This illustrates the correctness of Green function periodization.

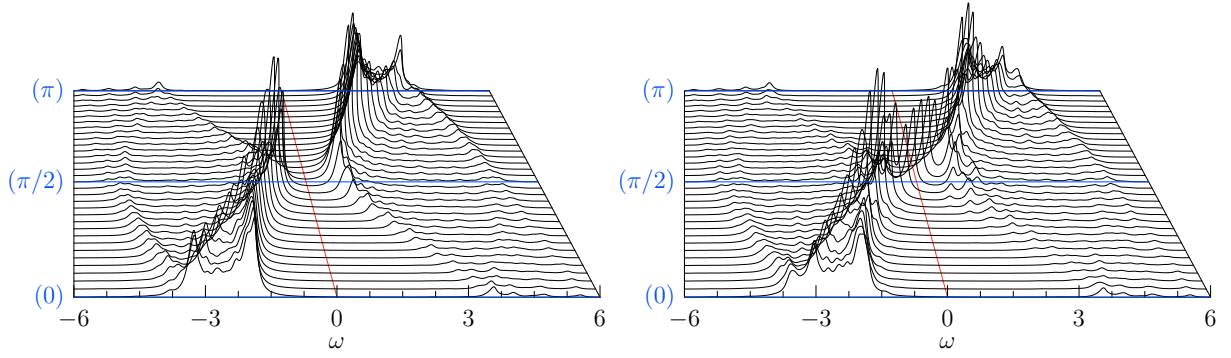


Fig. 3: Left: CPT spectral function of the one-dimensional, half-filled Hubbard model with $U = 4$, $t = 1$, with Green function periodization ($L = 16$). Right: the same, with self-energy periodization instead; notice the important spectral weight in the middle of the Mott gap.

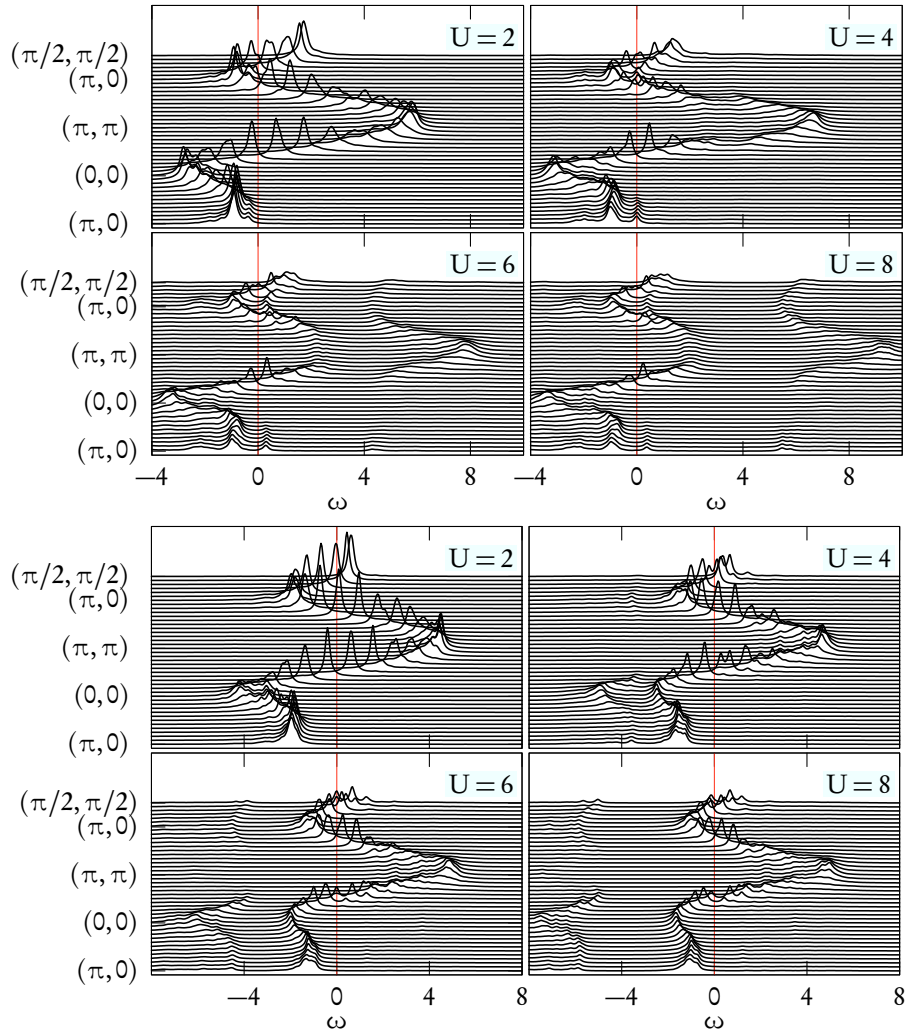


Fig. 4: Spectral function of the two-dimensional Hubbard model with band parameters $t = 1$, $t' = -0.3$ and $t'' = 0.2$. Top panel: different values of U for hole doping $1/6$. Bottom panel: the same for electron doping $1/6$. The pseudo-gap phenomenon manifests itself as the disappearance of the quasi-particle peak at the Fermi level along the side $(\pi, 0) - (\pi, \pi)$ in the hole-doped case, and along the diagonal $(0, 0) - (\pi, \pi)$ in the electron-doped case. Adapted from [4].

As an illustration of what can be revealed by the periodized CPT Green function, Fig. 4 shows the spectral function of the two-dimensional Hubbard model, with band parameters appropriate for cuprate superconductors: $t'/t = -0.3$ and $t''/t = 0.2$, where t' and t'' are second (diagonal) and third-neighbor hopping amplitudes, respectively. Several values of U were used, and two values of electron density: $5/6$ (top) and $7/6$ (bottom). We can see the emergence of the Hubbard bands and how the chemical potential ($\omega = 0$) is pinned to the lower and upper Hubbard bands, respectively. In the top panel, the quasi-particle weight at the Fermi surface disappears along the direction $(\pi, 0) - (\pi, \pi)$, whereas a well-defined quasi-particle peak remains along the diagonal direction $(0, 0) - (\pi, \pi)$. This is accentuated as U increases. In the electron-doped case, the roles of the two directions are reversed.

Averages of one-body operators In the $(\mathbf{R}, \tilde{\mathbf{k}})$ basis, Formula (26) for the ground state average of a one-body operator (23) becomes

$$\bar{\mathcal{O}} = \frac{1}{N} \sum_{\tilde{\mathbf{k}}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left\{ \text{tr} \left[\mathbf{s}(\tilde{\mathbf{k}}) \mathbf{G}(\tilde{\mathbf{k}}, i\omega) \right] - \frac{\text{tr} \mathbf{s}(\tilde{\mathbf{k}})}{i\omega - p} \right\} \quad (53)$$

where all matrices are now of size M_{sc} and we assume that the matrix $\mathbf{s}$ is diagonal in $\tilde{\mathbf{k}}$ (translational invariance over the super-lattice).

The result (25) is quite general and could formally be expressed as $\bar{\mathcal{O}} = \text{Tr}(\mathbf{s} \mathbf{G})$ where the symbol Tr (with a capital ‘T’) stands for a functional trace, i.e., includes an integral over frequencies as well as a trace over site and band indices, including even the convergence correction (the last term of Eq. (53)). The above expression is basis-independent; in the full wave-vector basis of one-particle states, the frequency summand would take the following form:

$$\frac{1}{N} \sum_{\tilde{\mathbf{k}}, \tilde{\mathbf{k}}', \mathbf{K}, \mathbf{K}'} s(\tilde{\mathbf{k}} + \mathbf{K}, \tilde{\mathbf{k}}' + \mathbf{K}') G(\tilde{\mathbf{k}}' + \mathbf{K}', \tilde{\mathbf{k}} + \mathbf{K}, i\omega) \quad (54)$$

If the operator $\mathcal{O}$ is translationally invariant, as it usually is, then

$$s(\tilde{\mathbf{k}} + \mathbf{K}, \tilde{\mathbf{k}}' + \mathbf{K}') = \delta_{\mathbf{K}\mathbf{K}'} \delta_{\tilde{\mathbf{k}}\tilde{\mathbf{k}}'} s(\mathbf{k}) = \delta_{\mathbf{k}\mathbf{k}'} s(\mathbf{k}), \quad (55)$$

and the above reduces to

$$\sum_{\mathbf{k}} s(\mathbf{k}) G_{\text{per.}}(\mathbf{k}, i\omega), \quad (56)$$

where $G_{\text{per.}}(\mathbf{k}, i\omega)$ is the periodized Green function (49). This means that expectation values of translationally invariant, one-body operators, computed in the periodization scheme (49), coincide with those computed without periodization, i.e., with Eq. (53). This does not hold for other periodization schemes (e.g. periodizing the self-energy), as it crucially depends on our discarding the off-diagonal elements of $\mathbf{G}$ in the full wave-vector basis, which is possible because we take the trace of $\mathbf{G}$ against a matrix $\mathbf{s}$ that is itself diagonal in that basis.

3 The exact diagonalization method

This lecture is not about numerics. Nevertheless, it is important to understand some basic facts about the exact diagonalization technique for correlated systems at zero temperature in order to understand some of the constraints imposed on quantum cluster methods by computing resources.

Let us therefore consider a single cluster with M degrees of freedom labeled α, β , etc. and Hamiltonian H_c . The essential steps involved in computing the cluster Green function $G_{\alpha\beta}(z)$ in the exact diagonalization method are the following:

1. **Coding the basis states.** States in the Hilbert space are represented as $|\psi\rangle = \sum_i \psi_i |b_i\rangle$, where the $|b_i\rangle$ form a basis in which it must be convenient to compute the matrix elements of H_c , i.e., the matrix elements should not be too numerous. Because of the local two-body interaction term, the most convenient basis is local, i.e., is defined by occupation numbers of the Wannier orbitals:

$$|b(n_\alpha)\rangle = (c_1^\dagger)^{n_1} (c_2^\dagger)^{n_2} \cdots (c_M^\dagger)^{n_M} |0\rangle \quad n_\alpha = 0 \quad \text{or} \quad 1 \quad (57)$$

For M degrees of freedom (spin included), there are 2^M such states. However, various symmetries will make the Hamiltonian block-diagonal in this basis. The most obvious ones are particle number and spin conservation, if applicable. Then only a subset of dimension D of the 2^M basis states is needed. The actual states are then specified by a D -dimensional array ψ_i . If point group symmetries are taken into account, then things are slightly more complicated, but easily manageable if we only take care of Abelian symmetries. The important point is that the dimension of the Hilbert space grows exponentially with M and therefore only small clusters can be used. Even though a ground state computation can be performed on the Hubbard model with slightly over 20 sites if all symmetries are used in the normal state, such a size is not realistic for quantum cluster methods, in which the Green function must also be computed and where many sequential solutions are needed within self-consistent or variational procedures. At this time the sweet spot still lies around 12 sites (i.e. $M = 24$ with spin).

2. **Building the Hamiltonian.** The matrix elements of H_c must be computed. In principle they need not be stored in memory, i.e., they could be computed ‘on the fly’, as needed. However, if memory is less a problem than computing time, it is advantageous to store them and to build a sparse matrix for H_c , especially in a way that allows the matrix to be quickly updated when the parameters of H_c (the coefficients of the various terms) change, for instance between successive iterations of a self-consistent procedure.
3. **Computing the ground state.** Once a representation of the Hamiltonian H_c is at hand, the ground state $|\Omega\rangle$ is typically computed using the Lanczos method. The latter is an iterative procedure that starts from a random vector in the Hilbert space and, through successive applications of the matrix H_c , finds the lowest eigenvector of H_c with numerical

accuracy. For instance, if the dimension of the Hilbert space is $D \sim 10^6$, an accurate ground state is found with only a few hundred iterations. The procedure actually builds an orthonormal basis in the K -dimensional Krylov space based on a random initial vector $|\phi_0\rangle$:

$$\mathcal{K} = \text{span} \{|\phi_0\rangle, H_c|\phi_0\rangle, H_c^2|\phi_0\rangle, \dots, H_c^{K-1}|\phi_0\rangle\} \quad (58)$$

In that basis, the Hamiltonian H_c is truncated into a simple tridiagonal form that is easily diagonalized. It can be shown that extreme eigenvectors are very well represented in $\mathcal{K}$.

4. **Computing the Green function.** Once the ground state $|\Omega\rangle$ is known, we need to apply definition (3) to compute the Green function. Once again, Krylov spaces are constructed. Two variants of the Lanczos method may be used, leading to two different data representations of the Green function. The first one proceeds by applying the usual Lanczos procedure, but on the states $c_\alpha|\Omega\rangle$ and $(c_\alpha + c_\beta)|\Omega\rangle$ (and their creation operator equivalent) instead of a random state. The Krylov spaces thus constructed provide a good representation of the action of the operator $(z \pm H)^{-1}$ of Eq. (3). The second variant, called the *band Lanczos method*, constructs a generalized Krylov space generated from the set $\{c_\alpha|\Omega\rangle\}$ (or $\{c_\alpha^\dagger|\Omega\rangle\}$) by successive application of H_c , and then uses the projection of H_c on that space to compute (3). It requires more memory than the first method but is faster. It also provides a spectral representation (11) of the Green function, albeit with a few hundred poles ω_r instead of a number of the order of the dimension of the Hilbert space, as in the exact result. Both representations of the cluster Green function allow it to be computed at any complex frequency z , provided we avoid the poles ω_r located on the real axis.

4 Cellular Dynamical Mean-Field Theory

Let us go back to the cluster decomposition (32) of the Hamiltonian. The main problem with CPT, i.e., with the prescription that $H_0^{(j)}$ is simply the restriction of H_0 to the cluster, is that the corresponding self-energy does not feel the effect of the lattice at all. It is the self-energy of a small system and thus cannot account for complex phenomena, such as phase transitions and spontaneously broken symmetry.

Therefore the main thrust of quantum cluster methods is to define $H_0^{(j)}$ in such a way as to represent as well as possible the effect of the lattice on the cluster. Perhaps the most elegant way to describe this is through the path integral formalism. The material of Sect. 4.1 is meant for readers familiar with that formalism. Others may skip straight to Sect. 4.2 if they wish.

4.1 The dynamical mean field

In the path integral formulation, the basic object is the partition function, whose expression is

$$Z = \int \prod_{\alpha} [dc_{\alpha} d\bar{c}_{\alpha}] \exp(iS[c, \bar{c}]) \quad (59)$$

where the action S is

$$S[c, \bar{c}] = \int dt \left\{ \sum_{\alpha, \beta} \bar{c}_\alpha(t) (i\delta_{\alpha\beta} \partial_t - t_{\alpha\beta}) c_\beta(t) - H_1(c, \bar{c}) \right\}. \quad (60)$$

The integral is carried out over a continuum of conjugate Grassmann variables $c_\alpha(t)$ and $\bar{c}_\alpha(t)$. An imaginary-time formulation is also possible and is necessary when dealing with finite temperatures; it is used to apply Monte Carlo sampling techniques.

In terms of the noninteracting Green function G_0 in the time domain, this may be written as

$$S[c, \bar{c}] = \int dt dt' \left\{ \sum_{\alpha, \beta} \bar{c}_\alpha(t) G_{0, \alpha\beta}^{-1}(t - t') c_\beta(t') - H_1(c, \bar{c}) \delta(t - t') \right\}. \quad (61)$$

Following a tiling of the lattice with clusters, that action may be expressed as

$$S = \sum_j S^{(j)} + \sum_{i,j} S^{(i,j)}, \quad (62)$$

where $S^{(j)}$ is the restriction of S to the cluster labeled j , and $S^{(i,j)}$ involve sites belonging to clusters i and j , typically inter-cluster hopping terms contained in the matrix $\mathbf{t}^{(i,j)}$ of Eq. (29). CDMFT assumes that the effect of the environment of each cluster can be well approximated by an effective action

$$\sum_{i \in \Gamma} S^{(i,j)} \rightarrow S_{\text{env.}}^{(j)}. \quad (63)$$

Replacing the sum on the left by a single term for each cluster effectively decouples them. This contribution from the environment is assumed to be uncorrelated, i.e., to be quadratic in c . Thus, the total effective action for a given cluster takes the general form

$$S_{\text{eff}}[c, \bar{c}] = \int dt dt' \sum_{\alpha, \beta} \bar{c}_\alpha(t) \mathcal{G}_{0, \alpha\beta}^{-1}(t - t') c_\beta(t') + \int dt H_1(c, \bar{c}) \quad (64)$$

where $\mathcal{G}_0$ is the *dynamical mean field*. The indices α, β are now restricted to the same cluster, and likewise for the interaction Hamiltonian H_1 .

In the frequency domain the dynamical mean field can be written, in matrix form, as

$$\mathcal{G}_0^{-1}(\omega) = \omega - \mathbf{t}_c - \mathbf{\Gamma}(\omega), \quad (65)$$

where $\mathbf{t}_c$ is the restriction of the hopping matrix to the cluster and $\mathbf{\Gamma}(\omega)$, the *hybridization function*, represents the dynamical hybridization of the cluster orbitals with their effective environment. This is better expressed in terms of an Anderson impurity model [5]; let us explain.

In order for this effective action to make sense, the dynamical mean field $\mathcal{G}_0$ must be causal. This implies that it must have the analytic properties of a Green function: The poles and zeros of its eigenvalues must lie on the real axis and the associated residues must be positive. In addition, $\mathcal{G}_0(\omega)$ must behave like $1/\omega$ at large frequencies. In other words, $\mathcal{G}_0(\omega)$ must have

a spectral representation like Eq. (11). Consequently, the hybridization function must have a spectral representation (22) like that of a self-energy:³

$$\Gamma_{\alpha\beta}(z) = \sum_r^{N_b} \frac{\theta_{\alpha r} \theta_{\beta r}^*}{z - \varepsilon_r} \quad (66)$$

as a function of a complex frequency z , where ε_r and $\theta_{\alpha r}$ form a collection of parameters that can be adjusted to fit any causal hybridization function as closely as needed. N_b is the number of poles deemed necessary to adequately represent the hybridization function.

4.2 The equivalent Anderson impurity model

Now, let us connect back with the Hamiltonian formalism. The effect of the cluster's environment, which is modeled by the hybridization function (66), may equivalently be represented by coupling the cluster with a set of N_b ancillary orbitals labeled by r , with annihilation operators a_r . The orbitals are uncorrelated, and form a *bath* in which the cluster is immersed, so to speak. The cluster dynamics is then described by the following Anderson impurity model:

$$H_{\text{AIM}} = \sum_{\alpha,\beta} t_{c,\alpha\beta} c_{\alpha}^{\dagger} c_{\beta} + \sum_{\alpha,r} (\theta_{\alpha r} c_{\alpha}^{\dagger} a_r + \text{H.c.}) + \sum_r^{N_b} \varepsilon_r a_r^{\dagger} a_r + H_1. \quad (67)$$

Electrons can hop between the cluster sites labeled α, β and the bath orbitals.

Let us show how the hybridization function (66) emerges from this model. The Green function associated with the noninteracting Anderson model (if we drop H_1 from (67)) is simply

$$\mathbf{G}_0^{\text{full}}(z) = \frac{1}{z - \mathbf{T}} \quad (68)$$

where the full hopping matrix $\mathbf{T}$ for the combined cluster and bath system is

$$\mathbf{T} = \begin{pmatrix} \mathbf{t}_c & \boldsymbol{\theta} \\ \boldsymbol{\theta}^{\dagger} & \boldsymbol{\varepsilon} \end{pmatrix}. \quad (69)$$

$\mathbf{t}_c$ is the $M \times M$ hopping matrix within cluster degrees of freedom only, $\boldsymbol{\theta}$ is the $M \times N_b$ hopping matrix between bath and cluster orbitals, and $\boldsymbol{\varepsilon}$ the diagonal $N_b \times N_b$ matrix of bath energies ε_{α} . The Green function obtained by tracing out the bath degrees of freedom is simply the restriction of $\mathbf{G}_{\text{full}}$ to the cluster degrees of freedom only. The mathematical problem at hand is simply to invert a 2×2 block matrix

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} \quad (70)$$

where $A_{11} = z - \mathbf{t}_c$, $A_{12} = A_{21}^{\dagger} = \boldsymbol{\theta}$, $A_{22} = z - \boldsymbol{\varepsilon}$, and B_{11} is the Green function we are looking for. By working out the inverse matrix condition, we find in particular that

$$A_{11} B_{11} + A_{12} B_{21} = \mathbf{1} \quad B_{21} = -A_{22}^{-1} A_{21} B_{11} \quad (71)$$

³ $\mathbf{T}$ could also have a frequency-independent piece $\mathbf{T}^{\infty}$, but that piece turns out to be always zero in CDMFT, although this is not so in Potthoff's dynamical impurity approximation (DIA).

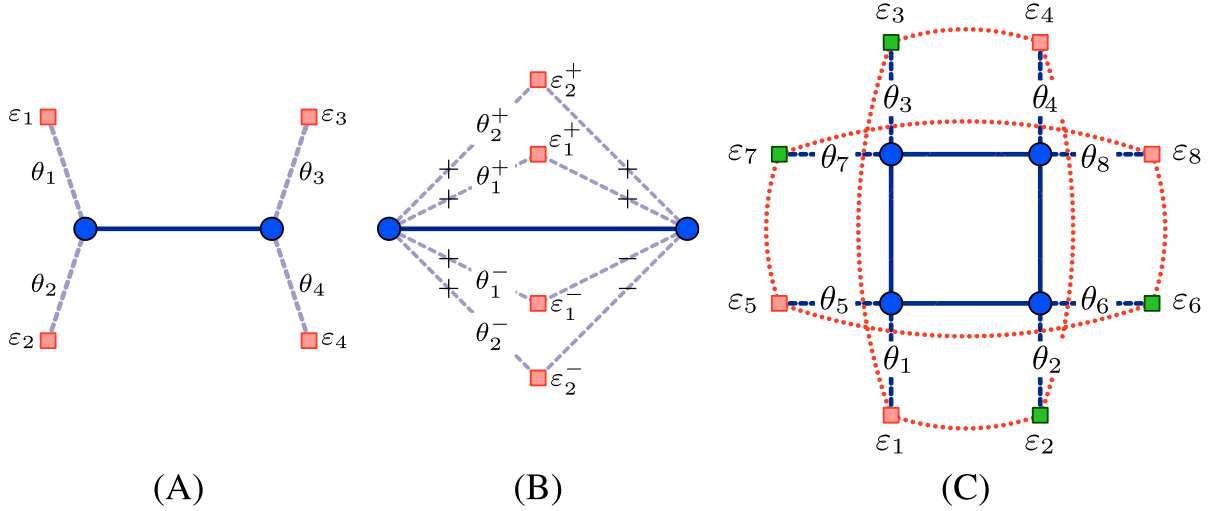


Fig. 5: Example cluster-bath systems used in the ED implementation of CDMFT. (A) and (B): 2-site, 4-bath systems used to study the one-dimensional Hubbard model. (C): a 4-site, 8-bath system used for the two-dimensional Hubbard model. See text for explanations.

and therefore

$$(A_{11} - A_{12}A_{22}^{-1}A_{21}) B_{11} = \mathbf{1} . \quad (72)$$

The noninteracting Green function of the cluster, G_{0c} , is thus given by

$$G_{0c}^{-1} = z - t_c - \Gamma(z) \quad \Gamma(z) = \boldsymbol{\theta} \frac{1}{z - \boldsymbol{\varepsilon}} \boldsymbol{\theta}^\dagger \quad (73)$$

where we recognize the hybridization function (66) in the last term. Note that the bath energy matrix $\boldsymbol{\varepsilon}$ can always be chosen to be diagonal, but does not have to be; Eq. (73) is valid even if $\boldsymbol{\varepsilon}$ is not diagonal.

In the interacting case, the only difference lies in the existence of the self-energy, which has no component in the bath since the latter is uncorrelated. We then have the relations

$$G_c^{-1}(z) = z - t_c - \Gamma(z) - \Sigma(z) = \mathcal{G}_0^{-1}(z) - \Sigma(z) \quad (74)$$

Solving the AIM on the cluster can be done in a variety of ways, for instance by exact diagonalization, but also by more traditional quantum Monte Carlo (QMC) approaches like the Hirsch-Fye method. Generally, the method used to compute $G_c^{-1}(z)$ is called the *impurity solver*, because of the AIM context, even though the problem does not involve physical impurities.

The continuous-time quantum Monte Carlo solver [6, 7] (CT-QMC) has been developed especially for the purpose of DMFT and CDMFT. It is an exact method, free of systematic errors, that works at finite temperature. It is however quite computationally intensive, and is also limited in practice to small clusters (but the bath is essentially infinite). All QMC methods are also affected by the fermion sign problem.

When using an exact diagonalization solver, the number of bath orbitals is very limited. Examples of cluster-bath systems are illustrated on Fig. 5. In system (A), each of the two cluster sites is hybridized with 2 bath orbitals exclusively. Left-right symmetry imposes the constraints

$\varepsilon_3 = \varepsilon_1$, $\varepsilon_4 = \varepsilon_2$, $\theta_3 = \theta_1$ and $\theta_4 = \theta_2$. System (B) is more general: the bath orbitals are not hybridized with specific cluster sites, but with combinations of a given symmetry (here, even and odd orbitals $c_{\pm} = c_1 \pm c_2$). This approach, proposed in Ref. [8], is potentially more accurate, since the number of variational parameters is generally larger. System (C) is a 4-site cluster coupled to 8 bath orbitals. Again, symmetry considerations in the normal state will impose constraints on the values of the 16 parameters ε_i and θ_i . In principle, one may also introduce hopping terms between the bath orbitals (indicated by curved dashed lines); this would introduce additional bath parameters and therefore a richer representation of the hybridization function $\Gamma(z)$. In practice, anomalous hopping will be introduced along the curved dashed lines when studying superconductivity (see Sect. 5.2).

4.3 The self-consistency condition

How does one determine the hybridization function $\Gamma(z)$ so as to best represent the effect of the lattice environment on the cluster? Just like ordinary mean-field theory can be formulated equivalently either as a variational principle (Hartree-Fock theory) or via a self-consistency condition, so can dynamical mean-field theory. The variational approach is based on Potthoff's self-energy functional theory [9, 10] and is called the *dynamical impurity approximation* (DIA, or CDIA for its cluster extension). It has many advantages, both from the formal point of view and in practice, for some systems with small values of N_b solved by exact diagonalization, but we will not describe it here for lack of space. Interested readers can turn to Refs [11–13] for details. Historically and in most applications, the hybridization function is determined by applying a self-consistency principle that we will now describe.⁴

The approximate Green function of the lattice model that follows from the effective Hamiltonian (67) is constructed from the principle (28):

$$\mathbf{G}^{-1}(\tilde{\mathbf{k}}, z) = z - \mathbf{t}(\tilde{\mathbf{k}}) - \mathbf{\Sigma}(z) , \quad (75)$$

where $\mathbf{t}(\tilde{\mathbf{k}})$ is the exact dispersion of the lattice model, expressed as a partial Fourier transform, i.e., as matrix in cluster indices with a dependence on the reduced wave-vector $\tilde{\mathbf{k}}$ defined in the Brillouin zone of the super-lattice.

Let us assume, for the time being, that the super-cell is made of a single cluster, so that $\mathbf{\Sigma}$ above is computed from the cluster Green function $\mathbf{G}_c$ (we will treat the case of multiple clusters later). Let us then Fourier transform $\mathbf{G}(\tilde{\mathbf{k}}, z)$ back to real-space, in order to project it onto the super-cell located at $\tilde{\mathbf{r}} = 0$:

$$\bar{\mathbf{G}}(z) = \frac{L}{N} \sum_{\tilde{\mathbf{k}}} \left[z - \mathbf{t}(\tilde{\mathbf{k}}) - \mathbf{\Sigma}(z) \right]^{-1} \quad (76)$$

Ideally, this projected Green function $\bar{\mathbf{G}}$ should coincide with the cluster Green function $\mathbf{G}_c$ calculated from the dynamical mean field $\mathcal{G}_0$. The condition $\bar{\mathbf{G}} = \mathbf{G}_c$ closes a self-consistency

⁴Contrary to mean-field theory, where the variational and self-consistent approaches yield the same solution, the self-consistent method followed in CDMFT does not yield the same solution as CDIA when the number of bath sites is finite.

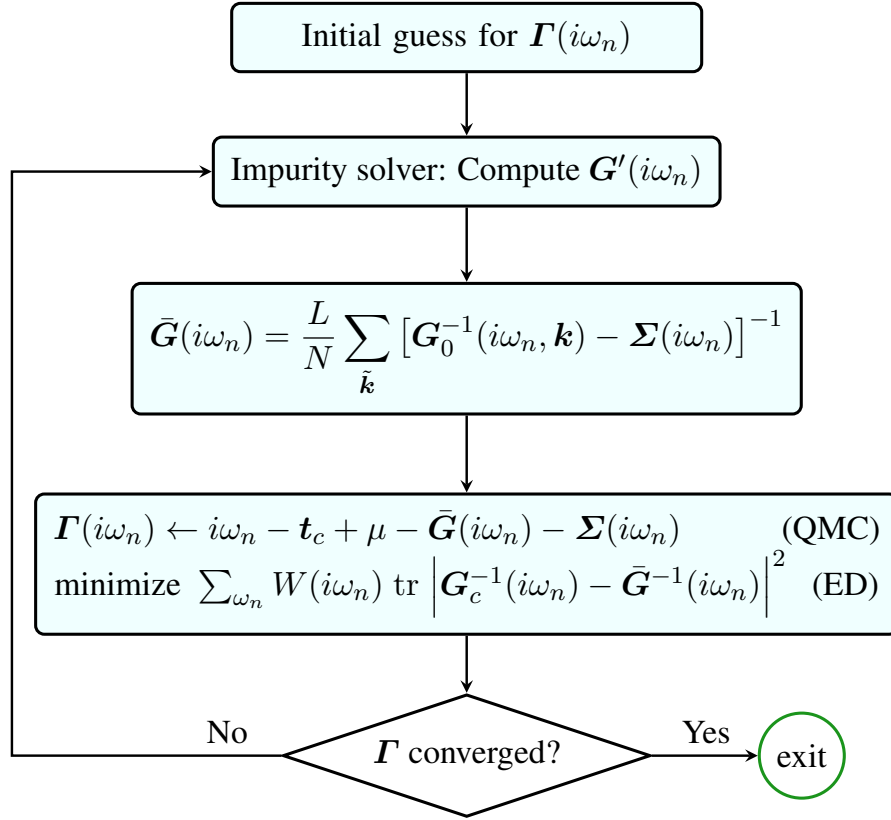


Fig. 6: The CDMFT self-consistency loop.

loop, illustrated in Fig. 6, that hopefully converges towards an optimal dynamical mean field $\mathcal{G}_0$. Let us summarize the elements of this loop:

1. An initial trial value of the dynamical mean field $\mathcal{G}_0$ is selected. With a QMC solver, this means choosing a Matsubara-frequency-dependent matrix $\Gamma(i\omega_n)$. With an ED solver applied to the Anderson impurity model (67), this means choosing an initial set of bath parameters (θ, ε) .
2. The impurity solver is applied, and G_c is computed, as well as the associated self-energy Σ , from Eq. (74).
3. The CPT Green function (75) is computed, as well as its projection $\bar{G}$ onto the cluster.
4. The next step depends on whether one uses a QMC or an ED solver to compute G_c . In the QMC solver, one deals with the path integral formulation directly and the hybridization function $\Gamma(i\omega_n)$ is defined at Matsubara frequencies without reference to a bath of ancillary orbitals. Then the dynamical mean field is updated by substituting $G_c \rightarrow \bar{G}$ into Eq. (74):

$$\mathcal{G}_0^{-1}(i\omega_n) \leftarrow \bar{G}^{-1}(i\omega_n) + \Sigma(i\omega_n) \quad (77)$$

In the ED solver, the hybridization function is only known through Eq. (73) and the bath parameters θ and ε . Moreover, the number of adjustable bath parameters is finite, and therefore the self-consistency condition $\bar{G} = G_c$ cannot be satisfied exactly for all frequencies; it can only be optimized. This is done by minimizing the following “distance

function”

$$d(\boldsymbol{\theta}, \varepsilon) = \sum_{i\omega_n} W(i\omega_n) \operatorname{tr} \left| \mathbf{G}_c^{-1}(i\omega_n) - \bar{\mathbf{G}}^{-1}(i\omega_n) \right|^2 \quad (78)$$

where the sum is carried over a finite set of Matsubara frequencies associated with a fictitious temperature T , with weights $W(i\omega_n)$ used to emphasize low frequencies. This minimization is done numerically by any classic optimization method, simply by applying Eq. (73), without recomputing $\boldsymbol{\Sigma}$; it usually does not contribute appreciably to the total computing time. The outcome is a new set of bath parameters $(\boldsymbol{\theta}, \varepsilon)$, and therefore a new hybridization function.

5. One goes back to step 2, until $\mathcal{G}_0$ (or $\boldsymbol{\Gamma}$) converges.

4.4 Inhomogeneous systems

Let us now consider a super-cell made of $n > 1$ clusters. This may be needed if a single cluster does not tile the lattice, but two are necessary, like in Fig. 1. It is also needed for inhomogeneous systems where translational invariance is lost, at least on a length scale that involves many clusters. The notation of Eqs. (28) and (29) may still be used, except that n is no longer the total number of clusters on the lattice, but the number of clusters in the super-cell, so that each of the blocks $\mathbf{t}^{(i,j)}$ will depend on the reduced wave-vector $\tilde{\mathbf{k}}$. The inverse CPT Green function therefore has the form

$$\begin{aligned} \mathbf{G}^{-1}(\tilde{\mathbf{k}}, z) &= z - \mathbf{t}(\tilde{\mathbf{k}}) - \boldsymbol{\Sigma}(z) \\ &= \begin{pmatrix} z - \mathbf{t}^{(11)}(\tilde{\mathbf{k}}) - \boldsymbol{\Sigma}_1(z) & -\mathbf{t}^{(12)}(\tilde{\mathbf{k}}) & \dots & -\mathbf{t}^{(1n)}(\tilde{\mathbf{k}}) \\ -\mathbf{t}^{(21)}(\tilde{\mathbf{k}}) & z - \mathbf{t}^{(22)}(\tilde{\mathbf{k}}) - \boldsymbol{\Sigma}_2(z) & \dots & -\mathbf{t}^{(2n)}(\tilde{\mathbf{k}}) \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{t}^{(n1)}(\tilde{\mathbf{k}}) & -\mathbf{t}^{(n2)}(\tilde{\mathbf{k}}) & \dots & z - \mathbf{t}^{(nn)}(\tilde{\mathbf{k}}) - \boldsymbol{\Sigma}_n(z) \end{pmatrix} \quad (79) \end{aligned}$$

Some of the clusters may be connected only to other clusters of the same super-cell, not to clusters of the neighboring super-cells; therefore for those clusters, the matrix $\mathbf{t}^{(ij)}$ is constant, i.e., does not depend on $\tilde{\mathbf{k}}$.

As before, we define the projected local Green function (76), but this time $\bar{\mathbf{G}}$ has a block-matrix structure like that of Eq. (79). Let $\bar{\mathbf{G}}^{(j)}$ denote the j^{th} diagonal block of $\bar{\mathbf{G}}$. The self-consistency condition is modified so as to match $\bar{\mathbf{G}}^{(j)}$ to the cluster Green function $\mathbf{G}^{(j)}$ computed from the impurity solver. In particular, in the ED framework, the distance function (78) has the form

$$d(\boldsymbol{\theta}, \varepsilon) = \sum_{i\omega_n, j} W(i\omega_n) \operatorname{tr} \left| \mathbf{G}^{(j)-1}(i\omega_n) - \bar{\mathbf{G}}^{(j)-1}(i\omega_n) \right|^2. \quad (80)$$

The bath parameters of all clusters in the super-cell contribute to any $\bar{\mathbf{G}}^{(j)}$, and thus all clusters are coupled. The distance function is a sum over n separate terms (one for each cluster), each with its own set of bath parameters and hybridization function $\boldsymbol{\Gamma}^{(j)}(z)$. Thus minimizing each of these terms separately will minimize their sum and vice versa.

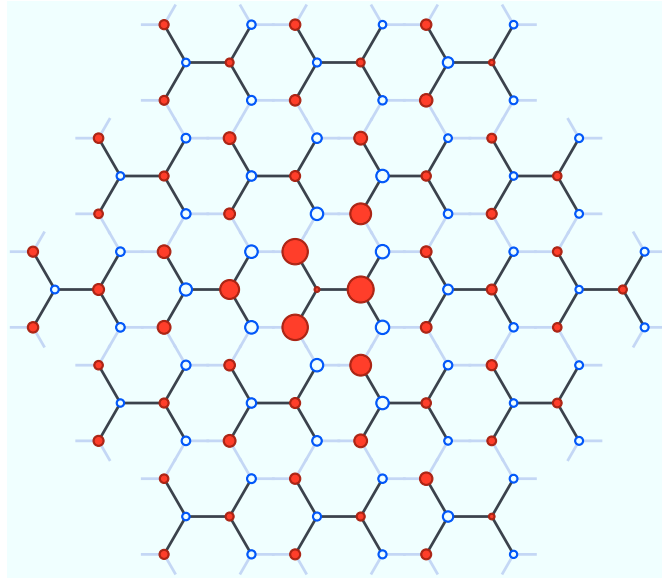


Fig. 7: Super-cell of 19 six-site clusters used in Ref. [14] in order to study the effect of a repeated non-magnetic impurity (located at the central site) on the Hubbard model defined on the graphene lattice. The local magnetization is indicated by filled red (up) and open blue (down) circles respectively (the area of each circle is proportional to $\langle S_r^z \rangle$).

This approach was called I-CDMFT in Ref. [14]. In that work, it was applied to the problem of a repeated, non-magnetic impurity in graphene. From previous mean-field calculations, it was known that antiferromagnetic correlations arising from the impurity go well beyond nearest neighbors [15, 16]. In order to isolate the magnetism resulting from a single impurity, and at the same time avoid edge effects, the impurity was repeated periodically, i.e., a super-cell of 19 six-site clusters was defined (see Fig. 7). The impurity, characterized by a local energy $H_{\text{imp.}} = h n_0$, where 0 labels the impurity site, is located on the middle cluster. Fig. 7 shows the local magnetization induced by such an impurity on the Hubbard model defined on the graphene lattice, with parameters $U = 2$ and $h = 11$ (the nearest-neighbor hopping amplitude t is set to unity).

5 Applications

5.1 The Mott transition

A key success of Dynamical Mean-Field Theory is the picture it provides of the Mott metal-insulator transition. Consider Fig. 8, which shows the qualitative phase diagram in the U - T plane of the half-filled, particle-hole symmetric Hubbard model. The left panel shows the prediction of single-site DMFT [17]. The Mott transition in the U - T plane is of first order, indicated as a red line on the figure. This first-order line ends at a finite-temperature critical point and at $U_{c1(0)}$, with a region (colored area) where the metallic and insulating phases may coexist. This single-site DMFT picture of the Mott transition has been criticized, mainly on

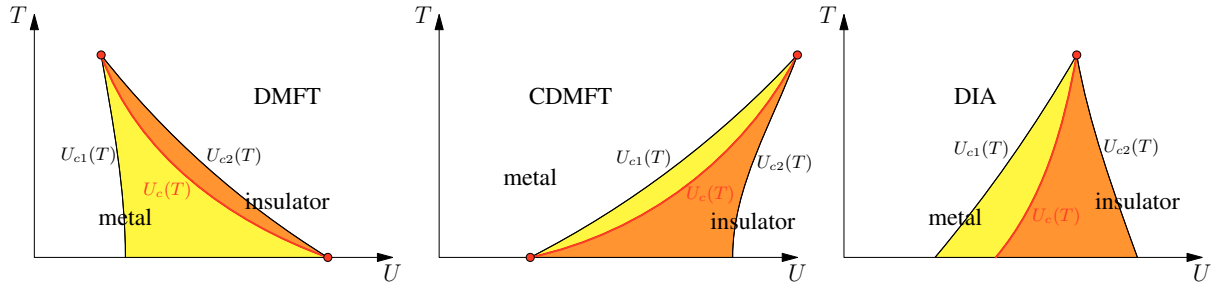


Fig. 8: Schematic phase diagram of the half-filled, particle-hole symmetric Hubbard model using single-site DMFT (left), CDMFT with a CT-QMC solver on a 2×2 plaquette (center) and the Dynamical Impurity Approximation (right). The red dots are the end of the first-order lines and therefore critical points. Adapted from [19].

the ground that the absence of feedback of magnetic correlations on single particle excitations yields a nonzero ground-state entropy ($S = N \ln 2$, N being the number of sites) in the Mott phase, and that this exaggerates the stability of the insulating phase at nonzero temperature. Cluster Dynamical Mean Field Theory modifies this picture by the addition of short-range correlation effects or, said otherwise, by adding a momentum-dependence to the self-energy. This provides a feedback of short-range antiferromagnetic fluctuations into single-particle properties. However, the main features of the DMFT picture are not affected by these refinements. The middle diagram of Fig. 8 emerges from a CDMFT study using the continuous-time QMC solver [18]. The essential difference with the single-site result is that the first-order line ends at $U_{c2}(0)$ instead of $U_{c1}(0)$ at zero temperature. The zero-temperature points are the result of an extrapolation, since QMC solvers were used in both cases. The right panel of Fig 8 is a modified scenario inspired by the Dynamical Impurity Approximation (DIA) [19], in which only the zero-temperature axis was actually calculated, with an exact diagonalization solver. It was shown in Ref. [19] that the transition is of first order even at zero temperature; the transition (red line) occurs when the ground state energies of the metallic and insulating solutions cross. As a function of U , a hysteresis loop was observed for the various bath parameters.

Regarding Fig. 8, we note that the slope of the first-order line is negative in the single-site DMFT solution, but positive in the CDMFT solution. Through the Clausius-Clapeyron equation, this implies that the entropy is larger in the Mott phase than in the metallic phase according to the single-site solution, whereas the opposite is true according to the CDMFT solution. This is another sign that the degeneracy of a single site exaggerates the entropy of the insulating state in DMFT.

Figure 9 shows the evolution of the local density of states $A(\omega)$ (DoS) across the Mott transition. The DoS on the left half of each panel ($U = 5.0$ and $U = 5.6$) are computed in the metallic solution, and those on the right ($U = 5.4$ and $U = 5.8$) in the insulating solution. The middle points ($U = 5.4$ and $U = 5.6$) are in the coexistence region. The bath used in the CT-QMC solution is effectively infinite, whereas the bath used with the ED solver has only 8 orbitals and is based on the system illustrated in Fig. 5C. Accordingly, the spectral function contains

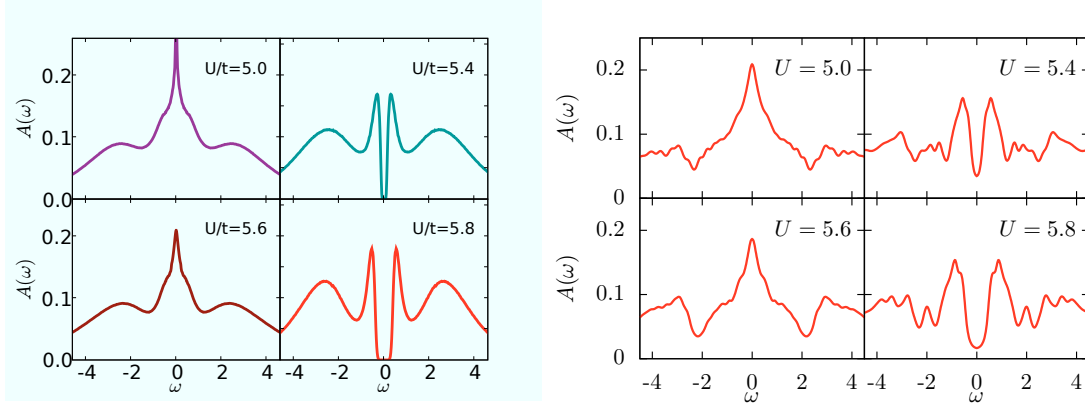


Fig. 9: Evolution of the local density of states $A(\omega)$ across the Mott transition, for the square-lattice Hubbard model. Left panel: CT-QMC solver, from Ref. [18]. Right panel, ED solver, but with the CDIA, based on solutions found in Ref. [19].

traces of the discreteness of the bath, even though a Lorentzian broadening of the poles was used. Nevertheless, the same physics occurs at the same value of U/t in the two methods. The self-consistency condition $\bar{G} = G'$, in the bath parametrization used in [19], does not show a first-order transition (and a coexistence region) in that case; the latter can only be seen by solving the more exact conditions defined in the CDIA.

5.2 Superconductivity

Many of the theoretical approaches described in this volume were motivated by the ambition to explain the origin of high-temperature superconductivity, more specifically to answer the following question: does the Hubbard model contain the key elements to explain superconductivity in the cuprates? Single-site DMFT alone cannot answer this question, as it lacks the short-range correlation effects needed to even describe d -wave superconductivity. But the question has been addressed by quantum cluster methods: DCA, VCA, and CDMFT.

First of all, let us explain how superconductivity may be incorporated in the Green function formalism. The standard theoretical description of superconductivity involves a spontaneous violation of charge conservation, which translates into a nonzero expectation value of the operator $\hat{\Psi}$ that creates a uniform distribution of Cooper pairs. In a one-band model, this is

$$\hat{\Psi} = \frac{1}{N} \sum_{\mathbf{r}, \mathbf{r}'} g_{\mathbf{r}\mathbf{r}'} [c_{\mathbf{r}\uparrow} c_{\mathbf{r}'\downarrow} - c_{\mathbf{r}\downarrow} c_{\mathbf{r}'\uparrow}]. \quad (81)$$

In momentum space, this becomes

$$\hat{\Psi} = \frac{1}{N} \sum_{\mathbf{k}} g(\mathbf{k}) [c_{\uparrow}(\mathbf{k}) c_{\downarrow}(-\mathbf{k}) - c_{\downarrow}(\mathbf{k}) c_{\uparrow}(-\mathbf{k})], \quad (82)$$

with the correspondence

$$g(\mathbf{k}) = \sum_{\mathbf{r}} g_{\mathbf{r},0} e^{-i\mathbf{k}\cdot\mathbf{r}}, \quad (83)$$

where the pairing amplitude $g_{\mathbf{r}\mathbf{r}'}$ is assumed to depend only on $\mathbf{r} - \mathbf{r}'$. This amplitude is, roughly speaking, the Cooper pair's wave-function as a function of the relative position of the two electrons.

A nonzero value of $\langle \hat{\Psi} \rangle$ can only be established if the anomalous Green function (the *Gorkov function*) is nonzero:

$$F_{\mathbf{r}\mathbf{r}'}(z) = \langle \Omega | c_{\mathbf{r}\uparrow} \frac{1}{z - H + E_0} c_{\mathbf{r}'\downarrow} | \Omega \rangle + \langle \Omega | c_{\mathbf{r}'\downarrow} \frac{1}{z + H - E_0} c_{\mathbf{r}\uparrow} | \Omega \rangle \quad (84)$$

The order parameter ψ is then the frequency integral of the Gorkov function, just like ordinary operators in Eq. (53):

$$\psi = \langle \hat{\Psi} \rangle = \frac{1}{N} \sum_{\tilde{\mathbf{k}}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left[g_{\mathbf{R}\mathbf{R}'}(\tilde{\mathbf{k}}) F_{\mathbf{R}'\mathbf{R}}(\tilde{\mathbf{k}}, i\omega) \right] \quad (85)$$

(we have used the $(\mathbf{R}, \tilde{\mathbf{k}})$ basis in the above formula). In order to smoothly integrate the Gorkov function into the formalism we have developed so far, the Nambu formalism is used: we perform a particle-hole transformation on the spin down sector and work with the operators $d_{\mathbf{r}\sigma} = (c_{\mathbf{r}\uparrow}, c_{\mathbf{r}\downarrow}^\dagger)$. These still obey the anticommutation relations $\{d_{\mathbf{r}\sigma}, d_{\mathbf{r}'\sigma'}^\dagger\} = \delta_{\mathbf{r}\mathbf{r}'}\delta_{\sigma\sigma'}$. In terms of these operators, the pairing operator $\hat{\Psi}$ takes the form of a hopping term because only destruction operators of opposite spins are multiplied. Thus, if we apply to the $d_{\mathbf{r}\sigma}$ and $d_{\mathbf{r}\sigma}^\dagger$ the definition (3) of the Green function and write it in block form to reveal the spin components, we find

$$\mathbf{G}(z) = \begin{pmatrix} \mathbf{G}_\uparrow(z) & \mathbf{F}(z) \\ \mathbf{F}^\dagger(z) & -\mathbf{G}_\downarrow(z) \end{pmatrix} \quad (86)$$

where $\mathbf{G}_\uparrow$ and $\mathbf{G}_\downarrow$ only involve up and down spins, respectively. As long as there are no spin-flip terms in the Hamiltonian, the Gorkov function can be included in this way in the Green function (spin-flip terms would spoil the process, as they would look like anomalous terms after the particle-hole transformation).

In order to probe for superconductivity in CDMFT, the hybridization function (73) must contain anomalous terms. Within the ED solver, this was at first accomplished as illustrated on Fig. 5C, i.e., by adding to the bath Hamiltonian d -wave pairing terms between bath orbitals themselves (the red dotted lines of Fig. 5C). In this parametrization, the bath is seemingly made of two “ghost clusters” whose pairing terms mimic the broken symmetry state that could take place on the cluster itself. Note that CDMFT does not tamper with the cluster Hamiltonian when probing broken symmetries. The agents of symmetry breaking are rather concentrated in the bath.

Any study of d -wave superconductivity within the square-lattice Hubbard model must also take into account the possibility of antiferromagnetic order in competition, or in coexistence with, superconductivity. This requires a more general bath parametrization, with different bath energies and hybridization for the two magnetic sub-lattices and spin projections. The reduced symmetry would translate into the following constraints on the parameters defined in Fig. 5C:

$$\varepsilon_{1\sigma} = \varepsilon_{4\sigma} = \varepsilon_{2-\sigma} = \varepsilon_{3-\sigma} \quad \text{and} \quad \varepsilon_{5\sigma} = \varepsilon_{8\sigma} = \varepsilon_{6-\sigma} = \varepsilon_{7-\sigma} \quad (87)$$

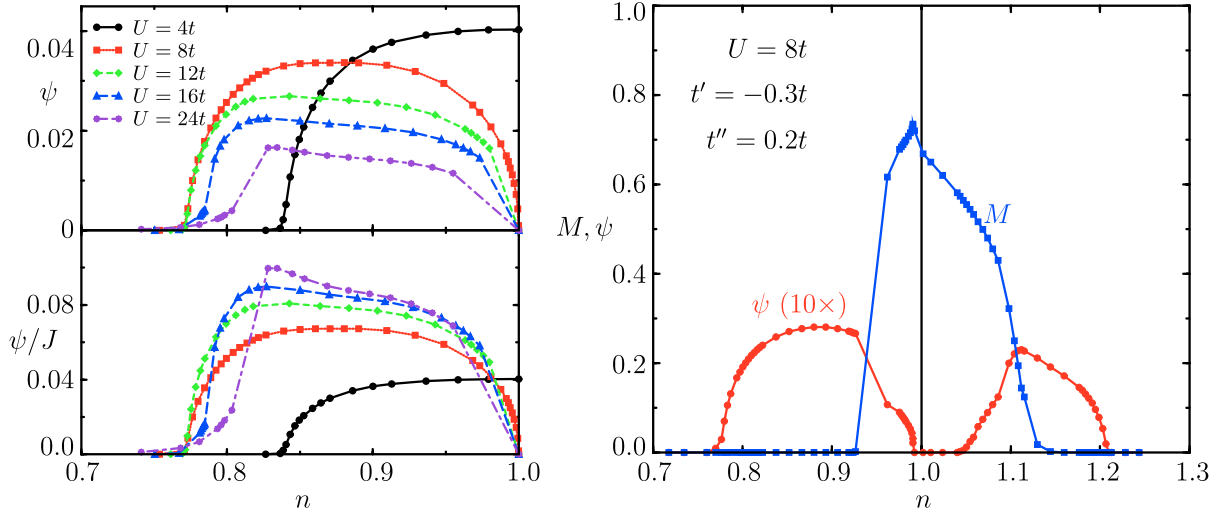


Fig. 10: Left panel top: d -wave order parameter ψ as a function of electron density n , for various values of U and $t' = -0.3t$, $t'' = 0.2t$. Bottom: the same, scaled by $J = 4t^2/U$. Right panel: d -wave (ψ) and antiferromagnetic (M) order parameters vs. n , from a common solution where they are allowed to compete. Taken from Ref. [20].

and likewise for θ_i . The independent bath parameters would then be $\varepsilon_{1,\uparrow}$, $\varepsilon_{1,\downarrow}$, $\varepsilon_{2,\uparrow}$, $\varepsilon_{2,\downarrow}$, the corresponding values of θ_i , and two in-bath pairing operators.

Figure 10, taken from Ref. [20], shows the result of a CDMFT computation on the one-band Hubbard model with a tight binding dispersion appropriate for a cuprate superconductor like YBCO, with diagonal hopping $t' = -0.3t$ and third-neighbor hopping $t'' = 0.2t$. Both the d -wave order parameter ψ and the antiferromagnetic order parameter M are shown, where

$$M = \langle \hat{M} \rangle \quad \hat{M} = \frac{1}{N} \sum_{\mathbf{r}} e^{i\mathbf{Q} \cdot \mathbf{r}} (n_{\mathbf{r}\uparrow} - n_{\mathbf{r}\downarrow}) \quad \mathbf{Q} = (\pi, \pi) . \quad (88)$$

There is a region of doping where the two phases compete and coexist at a microscopic scale (right panel of Fig. 10). One sees (left panel of Fig. 10) that the superconducting order parameter scales like $J = 4t^2/U$, at least in the under-doped region.

The CT-QMC solver was applied to the superconductivity problem in Ref. [21], in the special case $t' = 0$. The CT-QMC allows for an estimate of T_c , but such a computation is very resource-intensive because of critical slowing down. Of course, this T_c has a mean-field character: The Mermin-Wagner theorem forbids the spontaneous breakdown of continuous symmetries in a purely two-dimensional system at nonzero temperature. In a finite cluster, the long wavelength pair fluctuations that would destabilize a superconducting phase at finite temperature are not at work and cannot be accounted for by a fermionic bath.

A fair criticism of the above results on broken symmetries in the Hubbard model is the lack of finite-size analysis. The existence of broken symmetry phases, such as antiferromagnetism and d -wave superconductivity, can only be established firmly in the limit of infinite cluster size; an infinite bath-size is not sufficient. In principle infinite-size extrapolations should be performed in order to assess the robustness of CDMFT predictions in the thermodynamic limit. In

practice, this requires vast amounts of computing resources and a solver that can accommodate larger clusters, such as the Hirsch-Fye QMC or the auxiliary-field CT-QMC [22]. This was accomplished within the dynamical cluster approximation (DCA) in Ref. [23], with cluster sizes ranging up to $L = 32$ and special attention paid to the cluster shape in relation to the periodic boundary conditions used in DCA. The existence of a Kosterlitz-Thouless transition was confirmed in the square-lattice Hubbard model at $T_c \approx 0.023t$ for $U = 4t$ and $\delta \approx 10\%$. The cost of these computations precluded a wider exploration of parameter space. The assumption underlying current work on superconductivity using cluster approaches is that the thermodynamic limit will bring important renormalizations but will not qualitatively affect the dependence of the superconducting order upon band parameters, interaction strength, or doping. Thus, cluster approaches are important tools in exploring the space of models that can potentially lead to superconductivity or other broken symmetry phases.

6 Extended interactions

The cluster methods discussed above only apply to systems with on-site interactions, such as the simple Hubbard model (1), as the interaction $H_1^{(j)}$ on each cluster was supposed to be the restriction to the cluster of the full interaction H_1 . This is no longer true in the *extended* Hubbard model, in which the Coulomb interaction between electrons residing on different Wannier orbitals is included:

$$H_V = \frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} V_{\mathbf{r}'\mathbf{r}} n_{\mathbf{r}'} n_{\mathbf{r}} \quad (89)$$

(the factor $1/2$ is there to avoid double counting the pairs $(\mathbf{r}, \mathbf{r}')$). Treating such a model with quantum cluster approaches requires an additional approximation: One applies the Hartree approximation on the extended interactions that straddle different clusters but treats exactly all interactions within each cluster. This is called the dynamical Hartree approximation (DHA). It has been used in Ref. [24, 25] in order to assess the effect of such interactions on strongly-correlated superconductivity.

In the DHA, the extended interaction (89) is separated into two terms:

$$\frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} V_{\mathbf{r}, \mathbf{r}'}^c n_{\mathbf{r}} n_{\mathbf{r}'} + \frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}} n_{\mathbf{r}} n_{\mathbf{r}'} \quad (90)$$

where $V_{\mathbf{r}, \mathbf{r}'}^c$ denotes the extended interaction between sites belonging to the same cluster and $V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}}$ those interactions between sites belonging to different clusters. Each number operator appearing in the second term is then written as $n_{\mathbf{r}} = \bar{n}_{\mathbf{r}} + \delta n_{\mathbf{r}}$, where $\bar{n}_{\mathbf{r}} = \langle n_{\mathbf{r}} \rangle$ is the average value of $n_{\mathbf{r}}$ and $\delta n_{\mathbf{r}}$, by definition, its fluctuation. The classic mean-field treatment is then applied:

$$n_{\mathbf{r}} n_{\mathbf{r}'} = \bar{n}_{\mathbf{r}} \bar{n}_{\mathbf{r}'} + \delta n_{\mathbf{r}} \bar{n}_{\mathbf{r}'} + \bar{n}_{\mathbf{r}} \delta n_{\mathbf{r}'} + \delta n_{\mathbf{r}} \delta n_{\mathbf{r}'} \quad (91)$$

and the last term is dropped, as fluctuations are deemed small. Substituting $\delta n_{\mathbf{r}} = n_{\mathbf{r}} - \bar{n}_{\mathbf{r}}$ in

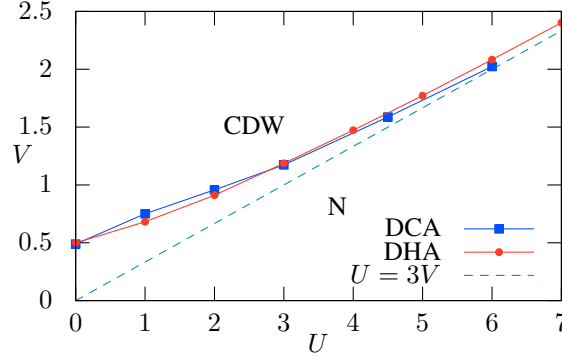


Fig. 11: Phase diagram of the half-filled extended Hubbard model defined on the graphene lattice. U and V are the on-site and nearest-neighbor repulsion, respectively. See text for details. Adapted from Refs [25] and [26].

the rest, we end up with the replacement

$$\frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}} n_{\mathbf{r}} n_{\mathbf{r}'} \rightarrow \hat{V}^{\text{ic}} = \frac{1}{2} \sum_{\mathbf{r}, \mathbf{r}'} V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}} (\bar{n}_{\mathbf{r}} n_{\mathbf{r}'} + n_{\mathbf{r}} \bar{n}_{\mathbf{r}'} - \bar{n}_{\mathbf{r}} \bar{n}_{\mathbf{r}'}). \quad (92)$$

These are local, one-body terms only and therefore are included in the non-interacting piece H_0 . Because of translation symmetry on the super-lattice, the averages $\bar{n}_{\mathbf{r}}$ must be the same in each super-cell, and therefore the above term can be determined within a super-cell only, and the matrix $V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}}$ can be viewed as acting on the sites of the super-cell only. Of course, interactions straddling different super-cells are “folded back” into a single super-cell.

The averages $\bar{n}_{\mathbf{r}}$ must be determined self-consistently. In a CPT computation, this calls for a self-consistent procedure whereby $\bar{n}_{\mathbf{r}}$ is computed at each step with the help of Eq. (53) and its value injected in a new H_0 , etc. In CDMFT, computing the averages can be done within the existing self-consistency loop of Fig. 6, e.g., at the same step where $\bar{\mathbf{G}}$ is computed. Thus, the mean fields $\bar{n}_{\mathbf{r}}$ are converged at the same time as the hybridization function $\mathbf{\Gamma}$ and the DHA adds very little overhead to the existing CDMFT procedure, although it is conceivable that more iterations are needed for convergence.

Instead of treating each operator $n_{\mathbf{r}}$ separately in the DHA, it is often advantageous to diagonalize the symmetric matrix $V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}}$ and to express the mean field problem in terms of eigenoperators m_k , which are linear combinations of the different $n_{\mathbf{r}}$ ’s of the cluster:

$$\hat{V}^{\text{ic}} = \sum_k V_k^{\text{ic}} \left[\bar{m}_k m_k - \frac{1}{2} \bar{m}_k^2 \right] \quad (93)$$

where V_k^{ic} is an eigenvalue of the matrix $V_{\mathbf{r}, \mathbf{r}'}^{\text{ic}}$ defined in the super-cell. The use of eigenoperators allows one to select or suppress different channels for charge-density-wave instabilities.

For instance, Fig. 11 shows the phase diagram of the extended Hubbard model on the graphene lattice. Only on-site (U) and nearest-neighbor (V) interactions were considered. If V is large enough, a charge-density-wave sets in with different densities on the A and B sub-lattices. The

DHA was used to find the phase boundary between the homogeneous solution (N) and the charge density wave (CDW), by including selectively the eigenoperator m_k associated with the CDW and by comparing the energy of the two solutions. The transition between the two phases becomes discontinuous (first-order) at some value of U . The results of the DHA are compared with those of DCA computations performed on larger clusters in Ref. [26]. The strong coupling limit of the phase boundary is known to be the straight line $U = 3V$ (dashed line on the figure).

Concluding remarks

Quantum cluster methods provide a unique window into the physics of strongly correlated materials. They capture short-distance correlations exactly in models of strongly correlated electrons and thus can describe phenomena that are not accessible to mean-field-like approaches. For instance, the effective, dynamic attraction between electrons located on nearest-neighbor sites in the Hubbard model, which leads to pairing and superconductivity in that model, is captured by CDMFT on a 4-site cluster. These methods are so far limited to rather small clusters or by other issues affecting various impurity solvers, like the sign problem in QMC, etc. But they constitute a framework that motivates continuous improvement in impurity solvers and that opens up lines of inquiry that were previously limited.

Acknowledgments

Discussions with A.-M.S. Tremblay, A. Foley and S. Verret are gratefully acknowledged. Computing resources were provided by Compute Canada and Calcul Québec. This work was supported by NSERC grant no RGPIN-2015-05598 (Canada).

References

- [1] C. Gros and R. Valenti, Phys. Rev. B **48**, 418 (1993)
- [2] D. Sénéchal, D. Perez, and M. Pioro-Ladrière, Phys. Rev. Lett. **84**, 522 (2000)
- [3] D. Sénéchal, D. Perez, and D. Plouffe, Phys. Rev. B **66**, 075129 (2002)
- [4] D. Sénéchal and A.M.S. Tremblay, Phys. Rev. Lett. **92**, 126401 (2004)
- [5] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
- [6] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)
- [7] P. Sémon, C.-H. Yee, K. Haule, and A.M.S. Tremblay, Phys. Rev. B **90**, 075149 (2014)
- [8] A. Liebsch and H. Ishida, Journal of Physics: Condensed Matter **24**, 053201 (2012)
- [9] M. Potthoff, Eur. Phys. J. B **32**, 429 (2003)
- [10] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003)
- [11] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003)
- [12] D. Sénéchal, Phys. Rev. B **81**, 235125 (2010)
- [13] M. Potthoff in *DMFT at 25: Infinite dimensions*, E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.), Lecture Notes of the Autumn School on Correlated Electrons 2014 (Forschungszentrum Jülich, 2014)
- [14] M. Charlebois, D. Sénéchal, A.-M. Gagnon, and A.M.S. Tremblay, Phys. Rev. B **91**, 035132 (2015)
- [15] H. Kumazaki and D.S. Hirashima, Journal of the Physical Society of Japan **76** (2007)
- [16] O.V. Yazyev and L. Helm, Phys. Rev. B **75**, 125408 (2007)
- [17] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [18] H. Park, K. Haule, and G. Kotliar, Phys. Rev. Lett. **101**, 186403 (2008)
- [19] M. Balzer, B. Kyung, D. Sénéchal, A.M.S. Tremblay, and M. Potthoff, Europhys. Lett. **85**, 17002 (2009)
- [20] S.S. Kancharla, B. Kyung, D. Sénéchal, M. Civelli, M. Capone, G. Kotliar, and A.M.S. Tremblay, Phys. Rev. B **77**, 184516 (2008)
- [21] K. Haule and G. Kotliar, Phys. Rev. B **76**, 104509 (2007)

- [22] E. Gull, P. Werner, O. Parcollet, and M. Troyer, *Europhys. Lett.* **82**, 57003 (2008)
- [23] T.A. Maier, M. Jarrell, T.C. Schulthess, P.R.C. Kent, and J.B. White, *Phys. Rev. Lett.* **95**, 237001 (2005)
- [24] D. Sénéchal, A.G.R. Day, V. Bouliane, and A.M.S. Tremblay, *Phys. Rev. B* **87**, 075123 (2013)
- [25] J.P.L. Faye, P. Sahebsara, and D. Sénéchal, arXiv e-prints (2015)
- [26] W. Wu and A.M.S. Tremblay, *Phys. Rev. B* **89**, 205128 (2014)

14 The Dynamic Cluster Approximation and its DCA^+ Extension

Thomas A. Maier

Computer Science and Mathematics Division

Center for Nanophase Materials Sciences

Oak Ridge National Laboratory

Contents

1	Introduction	2
2	The dynamic cluster approximation	3
2.1	General formalism	4
2.2	Quantum Monte Carlo cluster solver	6
2.3	Comparison with finite-size calculations	8
2.4	Calculation of response functions	10
3	The DCA^+ method	13
3.1	General formalism	13
3.2	Determination of lattice quantities from cluster quantities	16
3.3	Reduction of the QMC sign problem	20
4	Applications to the 2D Hubbard model	21
4.1	Antiferromagnetism	21
4.2	Pseudogap	22
4.3	Superconductivity	23
5	General discussion and concluding remarks	24
A	DCA and DCA^+ as self-energy functional approximations	28

1 Introduction

Theoretical studies of materials governed by strong electronic interactions remain one of the most challenging problems in condensed matter physics. While the Schrödinger equation describing these systems is known, the complexity of its solution increases exponentially with the number of electrons so that approximations have to be made. Mainstream theoretical modeling based on density-functional theory attempts to solve the full equation by mapping the many-body problem onto a single-particle problem. But its implementation in terms of the local density approximation fails to account for the strong correlations between the electrons and thus is unsuccessful in describing many of the many-body phenomena in these systems, such as magnetism or superconductivity. To study such behavior, a different approach is usually employed in order to make the problem tractable: A simplified model Hamiltonian is devised to provide an accurate description of the important low-energy degrees of freedom and higher energy states are left out. Because of this reduction in complexity, the model can then be solved with accurate many-body methods that treat the interactions accurately and thus are able to describe phenomena due to strong electronic correlations. Much of the work in this area is based on using methods such as exact diagonalization or quantum Monte Carlo (QMC) to determine the exact state of a finite-size lattice and regarding this state as an approximation to the bulk thermodynamic limit. Quantum cluster methods [1] such as the dynamic cluster approximation (DCA) [2–4], the subject of this lecture, use a different philosophy, in which the bulk, infinite-size, lattice problem is replaced by a finite-size cluster embedded in a mean-field bath designed to represent the remaining degrees of freedom. In contrast to finite-size calculations, quantum cluster methods give approximate results for the thermodynamic limit. They have been used extensively since their initial development in 1988 [2] and have provided insight into many important questions of condensed matter science. In this lecture, we present a pedagogical discussion of the DCA framework and its recent DCA⁺ extension [5, 6], together with selected applications that showcase the ability of these methods to provide insight into the physics of strongly correlated systems.

Preliminaries

In general, the Hamiltonian describing the physics of an interacting system is divided into a non-interacting part H_0 and an interacting part H_1

$$H = H_0 + H_{\text{int}} . \quad (1)$$

As a simple example, we will focus one of the most studied models in this field, the single-band Hubbard model [7]

$$H = \sum_{ij,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} . \quad (2)$$

Here $c_{i\sigma}^{(\dagger)}$ destroys (creates) an electron on site i with spin σ , and $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ is the corresponding number operator. The first (non-interacting) term describes the hopping of electrons

between sites i and j with amplitude t_{ij} , and the second (interaction) term raises the energy by the Coulomb repulsion U when two electrons with opposite spin reside on the same site.

The single-particle dynamics of the Hamiltonian at finite temperatures is described by the thermodynamic Green's function

$$G_{ij,\sigma} = -\langle T_\tau c_{i\sigma}(\tau) c_{j\sigma}^\dagger \rangle, \quad (3)$$

$$G_{ij,\sigma}(i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} G_{ij,\sigma}(\tau), \quad \omega_n = (2n+1)\pi/\beta, \quad (4)$$

$$G_\sigma(\mathbf{k}, i\omega_n) = \frac{1}{N} \sum_{ij} e^{i\mathbf{k}(\mathbf{r}_i - \mathbf{r}_j)} G_{ij,\sigma}(i\omega_n). \quad (5)$$

Here τ is the imaginary time, T_τ the time-ordering operator, $\beta = 1/T$ the inverse temperature, and ω_n are the fermionic Matsubara frequencies. For problems with translational symmetry in space and time, the Green's function becomes diagonal in momentum $\mathbf{k}$ and frequency $i\omega_n$ as stated in Eqs. (4) and (5). The Green's function G_0 of the non-interacting system, i.e. $H = H_0$, is given by

$$G_0(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n + \mu - \varepsilon_{\mathbf{k}}}, \quad (6)$$

where μ is the chemical potential and $\varepsilon_{\mathbf{k}}$ the dispersion, obtained from a Fourier transform of the hopping t_{ij} . For example, for a two-dimensional (2D) model with nearest neighbor hopping t and next-nearest-neighbor hopping t' , one has

$$\varepsilon_{\mathbf{k}} = -2t(\cos k_x + \cos k_y) - 4t' \cos k_x \cos k_y \quad (7)$$

with $\mathbf{k} = (k_x, k_y)$. Finally, the Dyson equation

$$G(\mathbf{k}, i\omega_n) = \frac{1}{G_0^{-1}(\mathbf{k}, i\omega_n) - \Sigma(\mathbf{k}, i\omega_n)}. \quad (8)$$

defines the self-energy $\Sigma(\mathbf{k}, i\omega_n)$ as the difference between the (inverse) non-interacting Green's function G_0 and the fully renormalized Green's function G and thus describes the effects of the interaction term H_{int} on the single-particle dynamics.

2 The dynamic cluster approximation

Calculating the Green's function G and the self-energy Σ exactly in the thermodynamic limit is prohibitively expensive as the problem size grows exponentially in the number of degrees of freedom. Finite-size methods, such as determinantal QMC [8] or Lanczos diagonalization [9], make the problem tractable by restricting the sums over sites in Eq. (2) to those of a finite size $L \times L$ cluster, small enough to be able to calculate the Green's function of the cluster, $G_c(\mathbf{K}, i\omega_n)$, and the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$, where $\mathbf{K}$ are the momenta of the finite-size cluster. Then, one could imagine carrying out calculations on a set of $L \times L$ lattices and then scaling to the thermodynamic limit $L \rightarrow \infty$. The DCA [1–3] takes a different approach: Similar to a finite-size calculation, it represents the system by a reduced number of cluster degrees of

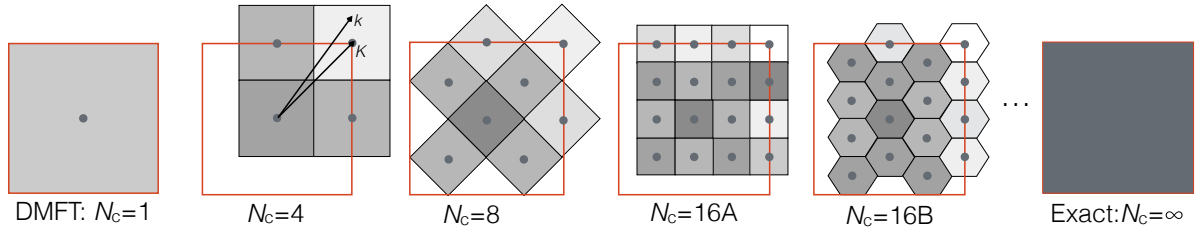


Fig. 1: Coarse-graining of momentum space: At the heart of the DCA method is a partitioning of the first Brillouin zone into N_c patches over which the Green's function is coarse-grained (averaged) to represent the system by a reduced number of N_c “cluster” degrees of freedom. The bulk degrees of freedom not included on the cluster are taken into account as a mean-field. For $N_c = 1$, the dynamical mean-field approximation is recovered, while for $N_c \rightarrow \infty$, one obtains the exact result. For a given cluster size N_c , one can have different locations and shapes of the coarse-graining patches, as illustrated for $N_c = 16A$ and $16B$.

freedom, but instead uses coarse-graining to retain information about the degrees of freedom not contained in the cluster. In the Appendix, we provide a rigorous derivation of both the DCA and DCA⁺ algorithms based on approximations of the grand potential. In the following, we give a more physically motivated discussion of these algorithms.

2.1 General formalism

To coarse-grain the degrees of freedom, the Brillouin zone is split into N_c patches of equal size. As illustrated in Fig. 1, each patch is represented by the cluster momentum $\mathbf{K}$ at the center of the cell, and a patch function

$$\phi_{\mathbf{K}}(\mathbf{k}) = \begin{cases} 1, & \text{if } \mathbf{k} \text{ in patch } \mathbf{K}. \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

is used to restrict momentum sums over momenta $\mathbf{k}$ inside the $\mathbf{K}^{th}$ patch. There can be different numbers N_c of patches, with different size and shape. The basic assumption of the DCA then is that the self-energy is only weakly momentum dependent, so it can be approximated on a coarse grid of $\mathbf{K}$ -points of a finite-size cluster

$$\Sigma(\mathbf{k}, i\omega_n) \simeq \Sigma_c(\mathbf{K}, i\omega_n). \quad (10)$$

Since the self-energy describes energy shift and life-time effects due to the interaction of an electron with other electrons, the dynamics of which is represented by the Green's function $G(\mathbf{k}, i\omega_n)$, it is generally a functional of $G(\mathbf{k}, i\omega_n)$, i.e., $\Sigma(\mathbf{k}, i\omega_n) = \Sigma[G(\mathbf{k}, i\omega_n)]$. In finite-size methods, the degrees of freedom are reduced to those of a cluster by calculating the self-energy from the cluster Green's function, i.e., $\Sigma_c(\mathbf{K}, i\omega_n) = \Sigma_c[G_c(\mathbf{K}, i\omega_n)]$. In contrast, in the DCA, all the degrees of freedom of the bulk lattice are retained by calculating the self-energy

$\Sigma_c(\mathbf{K}, i\omega_n) = \Sigma_c[\bar{G}(\mathbf{K}, i\omega_n)]$ from a coarse-grained Green's function

$$\bar{G}(\mathbf{K}, i\omega_n) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) G(\mathbf{k}, i\omega_n) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \frac{1}{i\omega_n - \varepsilon_{\mathbf{k}} + \mu - \Sigma^{DCA}(\mathbf{k}, i\omega_n)}. \quad (11)$$

Here the patch function $\phi_{\mathbf{K}}(\mathbf{k})$ restricts the sum to momenta $\mathbf{k}$ inside the $\mathbf{K}^{th}$ patch and the DCA self-energy

$$\Sigma^{DCA}(\mathbf{k}, i\omega_n) = \sum_{\mathbf{K}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma_c(\mathbf{K}, i\omega_n) \quad (12)$$

is approximated by a constant self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$ within the $\mathbf{K}^{th}$ patch, but varies between patches. $\bar{G}(\mathbf{K}, i\omega_n)$ represents a Green's function in which the degrees of freedom not contained on the cluster are coarse-grained or averaged out. The corresponding non-interacting Green's function

$$\mathcal{G}_0(\mathbf{K}, i\omega_n) = [\bar{G}^{-1}(\mathbf{K}, i\omega_n) + \Sigma_c(\mathbf{K}, i\omega_n)]^{-1} \quad (13)$$

is obtained by removing the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$. To calculate $\Sigma_c(\mathbf{K}, i\omega_n)$, an effective cluster model is set up using $\mathcal{G}_0$ together with the interaction term H_{int} of the Hamiltonian Eq. (2),

$$S[\phi^*, \phi] = - \int_0^\beta d\tau \int_0^\beta d\tau' \sum_{ij,\sigma} \phi_{i\sigma}^*(\tau) \mathcal{G}_{0,ij,\sigma}(\tau - \tau') \phi_{j\sigma}(\tau) + \int_0^\beta d\tau \sum_i U \phi_{i\uparrow}^*(\tau) \phi_{i\uparrow} \phi_{i\downarrow}^*(\tau) \phi_{i\downarrow}(\tau), \quad (14)$$

where ϕ and ϕ^* are the Grassmann variables corresponding to the operators c and $c^\dagger$, respectively. From this the cluster Green's function

$$G_{c,ij,\sigma}(\tau - \tau') = \frac{1}{Z} \int \mathcal{D}[\phi^* \phi] \phi_{i\sigma}(\tau) \phi_{j\sigma}^*(\tau') e^{-S[\phi^*, \phi]}, \quad (15)$$

where

$$Z = \int \mathcal{D}[\phi^* \phi] e^{-S[\phi^*, \phi]} \quad (16)$$

is the partition function, is calculated using, for example, the QMC algorithm discussed in Sec. 2.2 and used to determine the cluster self-energy

$$\Sigma_c(\mathbf{K}, i\omega_n) = \mathcal{G}_0^{-1}(\mathbf{K}, i\omega_n) - G_c^{-1}(\mathbf{K}, i\omega_n). \quad (17)$$

Then, using this new result for $\Sigma_c(\mathbf{K}, i\omega_n)$ in Eq. (12), steps (11) to (17) are iterated to self-consistency, i.e., until $\Sigma_c(\mathbf{K}, i\omega_n)$ does not change anymore between iterations. A sketch of the self-consistency loop to obtain the DCA self-energy $\Sigma^{DCA}(\mathbf{k})$ is given in Fig. 2, and Fig. 3 shows a typical result taken from Ref. [5] for a Hubbard model with $t' = -0.15t$ and $U = 7t$ at a filling $\langle n \rangle = 0.95$ and temperature $T = 0.33t$, which has been obtained with the quantum Monte Carlo cluster solver described in Sec. 2.2. Here, the imaginary part of $\Sigma^{DCA}(\mathbf{k}, \pi T)$ is plotted for the first Matsubara frequency versus $\mathbf{k}$ along a high-symmetry path in the Brillouin zone. For this case, the self-energy has a pronounced momentum dependence and one sees the basic approximation of the DCA: The self-energy is constant within a $\mathbf{K}$ -patch and has a

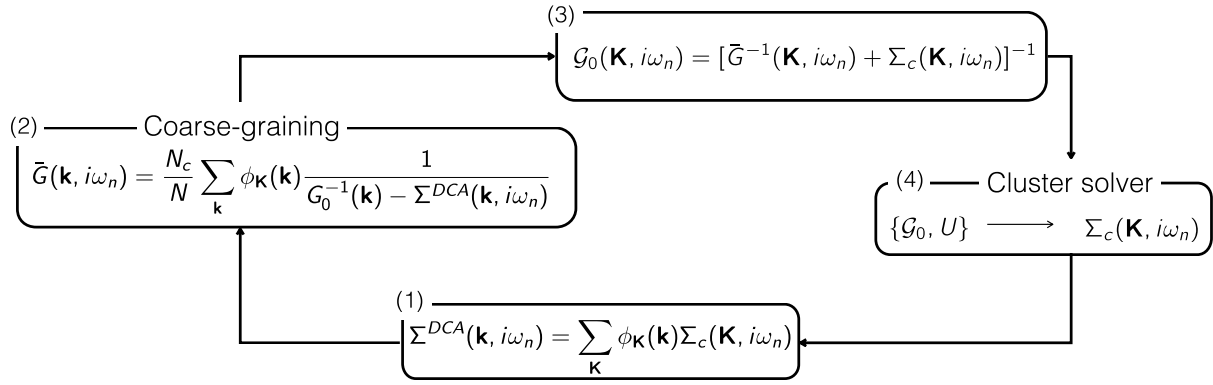


Fig. 2: The DCA self-consistency loop: (1) To initialize the algorithm, one starts with a cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$, which is usually taken from a previous calculation, for example at a higher temperature, or set to zero. From $\Sigma_c(\mathbf{K}, i\omega_n)$, the lattice self-energy $\Sigma^{DCA}(\mathbf{k}, i\omega_n)$ is constructed using a piecewise-constant continuation of $\Sigma_c(\mathbf{K}, i\omega_n)$. (2) $\Sigma^{DCA}(\mathbf{k}, i\omega_n)$ is then used in the coarse-graining of the Green's function to give $\tilde{G}(\mathbf{K}, i\omega_n)$. (3) From $\tilde{G}(\mathbf{K}, i\omega_n)$, one then calculates the corresponding bare Green's function of the cluster, $\mathcal{G}_0(\mathbf{K}, i\omega_n)$, by removing the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$. (4) The bare cluster Green's function $\mathcal{G}_0(\mathbf{K}, i\omega_n)$ is used together with the interaction U to set up the effective cluster model, which is solved with a cluster solver such as the QMC algorithm discussed in Sec. 2.2. This provides a new result for the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$, which is used in the next iteration in step (1).

step discontinuity at the boundary between the patches. In addition, it is apparent that different clusters can give large differences in $\Sigma^{DCA}(\mathbf{k})$. This is especially clear in the case of the 16A and 16B clusters (see Fig. 1), which have the same size but different locations and shapes of the coarse-graining patches. In principle, these finite-size effects can be mitigated by carrying out calculations for larger clusters. In some cases, calculations for very large clusters are possible and one can perform finite size scaling to obtain the exact infinite cluster size result. Usually, however, the increase in numerical complexity associated with going to larger clusters limits calculations to relatively small clusters for which this problem persists.

2.2 Quantum Monte Carlo cluster solver

The DCA algorithm requires the calculation of the Green's function and the self-energy of the effective cluster problem as defined by the action in Eq. (14). While the determinantal QMC (DQMC) technique [8] is the method of choice for unbiased calculations of 2D finite-size lattices, here one deals with a finite-size cluster embedded in a dynamic mean field, and other methods that can treat this coupling to a fermionic bath are necessary. The Hirsch-Fye QMC algorithm was originally developed to study the properties of magnetic impurities hybridized with conduction electrons in metals [10] and later extended to solve the effective cluster problem in the DCA [11]. Just like the DQMC algorithm, this method uses a Suzuki-Trotter decomposition of the partition function and therefore has time discretization errors. In recent years, however,

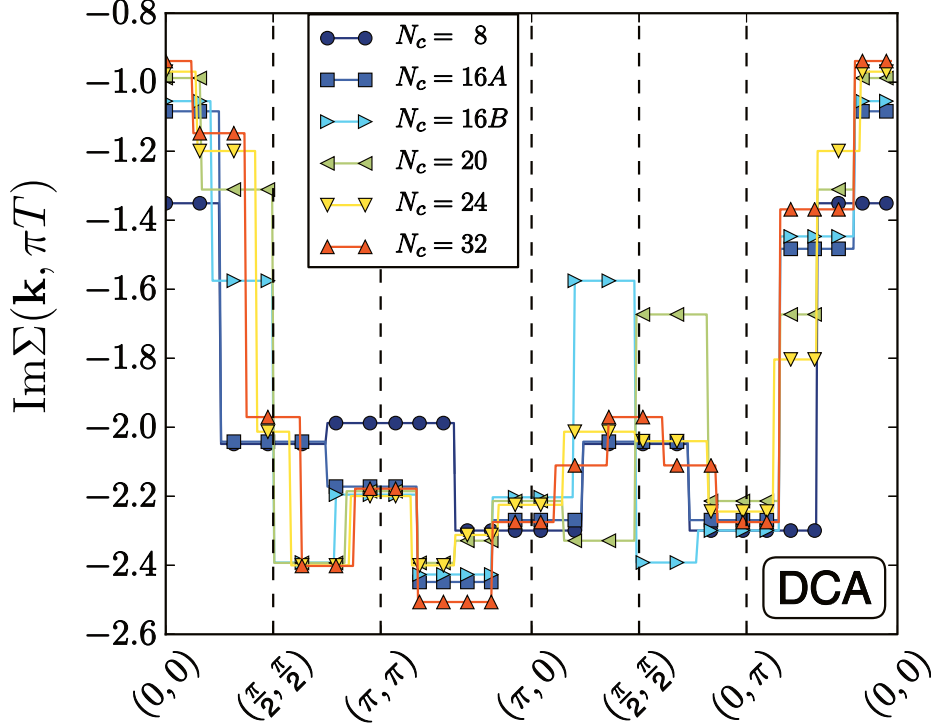


Fig. 3: DCA self-energy: DCA results for the imaginary part of the lattice self-energy $\Sigma(\mathbf{k}, \pi T)$ for a Hubbard model with $U = 7t$, $t' = -0.15t$, $\langle n \rangle = 0.95$ at a temperature $T = 0.33t$. The DCA approximation gives a self-energy with jump discontinuities between the coarse-graining patches and which depends strongly on the location and shape of the patches ($N_c = 16A$ vs. $16B$).

a number of continuous-time QMC (CT-QMC) methods have been developed [12], which are free from time discretization errors and which are more efficient than the Hirsch-Fye algorithm. Of these, we will discuss the continuous-time auxiliary-field (CT-AUX) QMC algorithm [13], which has been developed specifically for the type of large cluster DCA calculations we are interested in.

The CT-AUX algorithm is formally similar to the Hirsch-Fye QMC algorithm in that it employs an auxiliary-field decoupling of the interaction term, H_{int} in Eq. (2). But instead of the time discretization, it performs a weak-coupling expansion of the interaction term. Monte Carlo sampling is then performed in the combined space of perturbation expansion order and time-ordered configurations for a given order. The method yields numerically exact results that are continuous in time since the positions along the imaginary time axis are variable and not fixed as in Hirsch-Fye. In the following, we sketch the basic idea of this approach and refer the reader to Refs. [12, 13] for a detailed discussion.

One starts by expanding the partition function, Eq. (16), in powers of H_{int} to which an arbitrary constant term $-K/\beta$ with non-zero K has been added. Then, one applies the auxiliary-field decomposition [14]

$$1 - \frac{\beta U}{K} \sum_i \left[n_{i\uparrow} n_{i\downarrow} - \frac{1}{2} (n_{i\uparrow} + n_{i\downarrow}) \right] = \frac{1}{2N_c} \sum_{i, s_i = \pm 1} e^{\gamma s_i (n_{i\uparrow} - n_{i\downarrow})}, \quad (18)$$

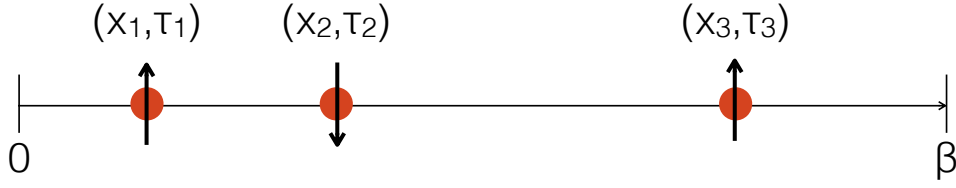


Fig. 4: Configuration in the CT-AUX QMC algorithm: The CT-AUX QMC algorithm samples different expansion orders in an interaction expansion and time-ordered configurations of vertices with cluster site x , imaginary time τ , and auxiliary Ising spin s for a given expansion order k . Here we show a representation of an order $k = 3$ configuration $((x_1, \tau_1, s_1), (x_2, \tau_2, s_2), (x_3, \tau_3, s_3))$.

with $\cosh(\gamma) = 1 + U\beta N_c/2K$. With this, the partition function

$$Z = \sum_{k=0}^{\infty} \sum_{s_1 \dots s_k = \pm 1} \int_0^{\beta} d\tau_1 \dots \int_{\tau_{k-1}}^{\beta} d\tau_k \left(\frac{K}{2\beta N_c} \right)^k Z_k(\{x, \tau, s\}_k) \quad (19)$$

then becomes a sum over expansion orders k and auxiliary-spin configurations $s_1 \dots s_k$ of terms

$$Z_k(\{x, \tau, s\}_k) = Z_0 \prod_{\sigma} \det N_{\sigma}^{-1}(\{x, \tau, s\}_k). \quad (20)$$

Here, $\{x, \tau, s\}_k$ is a configuration of k vertices with cluster site x , imaginary time τ and auxiliary spin s (see Fig. 4) and N_{σ} is a $k \times k$ matrix

$$[N_{\sigma}^{-1}]_{ij} = [e^{\gamma(-1)^{\sigma} s_i}] \delta_{ij} - \mathcal{G}_{0,\sigma}(x_i, \tau_i; x_j, \tau_j). \quad (21)$$

This equation expresses the weight of a configuration $\{x, \tau, s\}_k$ for expansion order k in terms of the bare cluster Green's function $\mathcal{G}_{0,\sigma}(x_i, \tau_i; x_j, \tau_j)$, the interaction U (through γ), and the product of determinants of two matrices N_{σ} .

The Monte Carlo algorithm then samples the partition function by randomly creating and removing auxiliary spins s at different times τ and locations x and updating the N_{σ} matrices using fast update formulas [13]. In the case of a single-spin update, this takes the form of a rank-1 update (vector outer product). In order to improve the efficiency of the algorithm, k_s subsequent rank-1 updates can be rewritten as a rank- k_s update [15]. Measurements of the single-particle and two-particle Green's functions are made in a similar manner as in the Hirsch-Fye algorithm. That is, after the auxiliary field has been introduced, one has a Wick's theorem for decomposing products of time-ordered operators. To avoid time-discretization problems and for improved efficiency, it is beneficial to Fourier-transform to frequency and momentum space before carrying out the measurements [12]. Due to the random times of the vertices, non-equidistant fast Fourier transform algorithms are employed to further improve efficiency [16].

2.3 Comparison with finite-size calculations

Both types of calculations, finite-size and DCA, make the assumption that correlations are short-ranged and contained within a finite-size cluster. In contrast to studies of finite-size systems,

however, in which the exact state of a cluster with N_c sites is determined and regarded as an approximation to the bulk thermodynamic limit, the DCA, for a given cluster size N_c , gives approximate results for the thermodynamic limit. Both methods give the exact result as $N_c \rightarrow \infty$, but only the DCA has a non-trivial limit as $N_c \rightarrow 1$. For $N_c = 1$, the single DCA patch extends over the full Brillouin zone (see Fig. 1), the coarse-grained Green's function in Eq. (11) becomes the local Green's function, and the equations of the Dynamical Mean-Field Approximation [17] are recovered.

The difference between a finite-size and a DCA calculation becomes clear when writing the bare cluster Green's function $\mathcal{G}_0(\mathbf{K}, i\omega_n)$ that enters the bilinear part of the effective action given in Eq. (14) as [1]

$$\mathcal{G}_0(\mathbf{K}, i\omega_n) = \frac{1}{i\omega_n + \mu - \bar{\varepsilon}_{\mathbf{K}} - \Gamma(\mathbf{K}, i\omega_n)}. \quad (22)$$

Here, $\bar{\varepsilon}_{\mathbf{K}} = N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \varepsilon_{\mathbf{k}}$ is the coarse-grained average of the dispersion, and the hybridization function

$$\Gamma(\mathbf{K}, i\omega_n) = \frac{\frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \delta t_{\mathbf{K}}^2(\mathbf{k}) G(\mathbf{k}, i\omega_n)}{1 + \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \delta t_{\mathbf{K}}(\mathbf{k}) G(\mathbf{k}, i\omega_n)} \quad (23)$$

with $\delta t_{\mathbf{K}}(\mathbf{k}) = \varepsilon_{\mathbf{k}} - \bar{\varepsilon}_{\mathbf{K}}$ describes the coupling of the cluster degrees of freedom to the remaining sites of the bulk lattice in an averaged, mean-field manner. This is in contrast to the finite-size case, in which one uses the bare Green's function of an isolated cluster, i.e. $G_0(\mathbf{K}, i\omega_n) = [i\omega_n + \mu - \varepsilon_{\mathbf{K}}]^{-1}$, in the action. For the DCA, $\mathcal{G}_0$ in Eq. (22) has the form of a Green's function of a non-interacting cluster with momenta $\mathbf{K}$, where each $\mathbf{K}$ is coupled to a dynamic mean field given by $\Gamma(\mathbf{K}, i\omega_n)$. As one sees from Eq. (23), $\Gamma(\mathbf{K}, i\omega_n)$ describes the effects of the $\mathbf{k}$ momenta surrounding $\mathbf{K}$ in an averaged fashion. As a consequence, for a given finite cluster, the DCA gives results that are usually closer to the thermodynamic limit than a finite-size calculation [11, 18].

Sign problem

The most significant challenge of QMC calculations of fermionic systems is the so-called fermion sign problem [19]. In the general case of a Hubbard model at finite doping, it arises from a negative product of determinants in the weight Z_k (Eq. (20)) of a configuration that is used to calculate the probability of accepting an update in the Monte Carlo procedure. For a simple Hubbard model, it leads to a statistical error that grows exponentially in the number of cluster sites N_c , the inverse temperature β , and the size of the Coulomb repulsion U . Just like finite-size QMC calculations, this sign problem is also encountered in DCA and DCA⁺ calculations when a QMC algorithm such as the CT-AUX algorithm described in Sec. 2.2 is used to solve the effective cluster problem.

DCA QMC calculations were shown to have a much less severe sign problem than finite-size QMC calculations (see Fig. 5). Lacking a rigorous mathematical justification, this was attributed to the action of the mean-field host on the cluster [11]. In any case, this significant reduction of the severity of the sign problem in the DCA has enabled access to much lower temperatures than those that can be reached in finite-size systems.

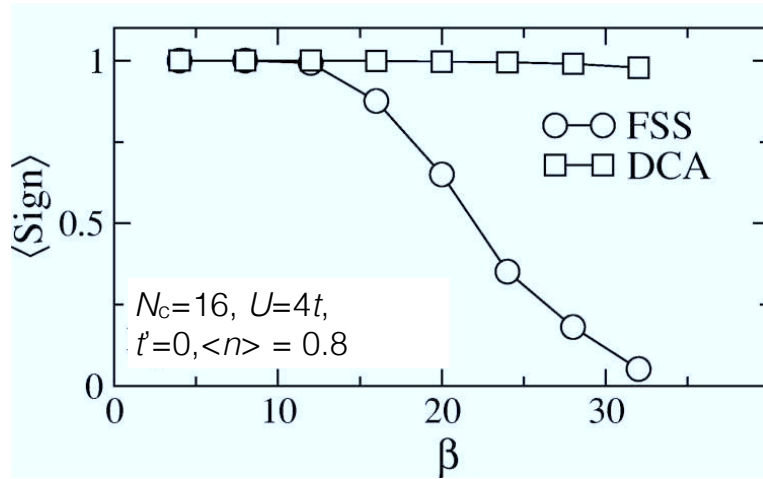


Fig. 5: DCA reduction of the sign problem: The average QMC sign compared between finite-size calculation (FSS) and DCA for a 16-site cluster with $U = 4t$ and $\langle n \rangle = 0.8$. The DCA algorithm reduces the sign problem of finite-size QMC significantly. Figure taken from Ref. [11].

2.4 Calculation of response functions

Response functions, such as the magnetic susceptibility, provide information on the response of a system to an external field, as well as on the nature of the dominant fluctuations and possible instabilities of the system towards spontaneous symmetry breaking. The calculations of these functions requires extensions to the single-particle formalism described in the previous sections. They either require an explicit calculation on the single-particle level in the presence of a symmetry-breaking field or a calculation of two-particle Green's functions, from which the susceptibilities can be calculated [1]. Here we discuss the second approach and refer the reader to Ref. [1] for the first approach.

General formalism

As an illustrative example, let us consider the “pair-field” susceptibility

$$P_{\varphi}(T) = \int_0^{\beta} d\tau \langle \Delta_{\varphi}(\tau) \Delta_{\varphi}^{\dagger}(0) \rangle \quad (24)$$

which gives the superconducting response of a system to an external “pair-field” that couples to the pairing operator

$$\Delta_{\varphi}^{\dagger} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} g_{\varphi}(\mathbf{k}) c_{\mathbf{k}\uparrow}^{\dagger} c_{-\mathbf{k}\downarrow}^{\dagger}. \quad (25)$$

Here, $g_{\varphi}(\mathbf{k})$ is a form factor that describes the momentum-dependence and symmetry of the pair wavefunction. For a pair with internal $d_{x^2-y^2}$ -wave symmetry, for example, a state that is relevant to the copper-oxide high-temperature superconductors as well as to the 2D Hubbard model, one has

$$g_{d_{x^2-y^2}}(\mathbf{k}) = \cos k_x - \cos k_y \quad (26)$$

where $\mathbf{k} = (k_x, k_y)$. This arranges the electrons of the corresponding real-space pair on nearest-neighbor sites with a $d_{x^2-y^2}$ phase (+1 along $\pm x$ and -1 along $\pm y$). For a conventional s -wave pair, on the other hand, one simply has $g_s(\mathbf{k}) = 1$.

The pair-field susceptibility $P_\varphi(T)$ may be calculated from the two-particle Green's function

$$G_{2,\sigma_1\ldots\sigma_4}(x_1, x_2; x_3, x_4) = -\langle T_\tau c_{\sigma_1}(x_1) c_{\sigma_2}(x_2) c_{\sigma_3}^\dagger(x_3) c_{\sigma_4}^\dagger(x_4) \rangle. \quad (27)$$

Fourier-transforming on both the space and time variables gives $G_{2\sigma_1\ldots\sigma_4}(k_4, k_3; k_2, k_1)$ with $k = (\mathbf{k}, i\omega_n)$. With this, one has

$$P_\varphi(T) = \frac{T^2}{N^2} \sum_{k, k'} g_\varphi(\mathbf{k}) G_{2,\uparrow\downarrow\uparrow}(k, -k, -k', k') g_\varphi(\mathbf{k}') \quad (28)$$

The way G_2 is calculated in the DCA algorithm is similar to the way G is calculated at the single-particle level. Just as the Dyson equation (8) relates the Green's function to the self-energy, the Bethe-Salpeter equation (BSE) relates G_2 to the irreducible vertex function $\Gamma_\alpha(k_1, k_2; k_3, k_4)$ for channel α .¹ For example, for the particle-particle channel, which is relevant here, one has

$$\begin{aligned} G_2(k, q - k, q - k', k') &= G_\uparrow(k) G_\downarrow(q - k) \delta_{k, k'} \\ &- \frac{T}{N} \sum_{k''} G_\uparrow(k) G_\downarrow(q - k) \Gamma_{pp}(k, q - k, q - k'', k'') G_2(k'', q - k'', q - k', k'). \end{aligned} \quad (29)$$

Here we have used momentum, energy and spin conservation to reduce the dependence on four variables $k_1 \ldots k_4$ to three variables k, k' , and q , where $q = (\mathbf{q}, i\omega_m)$ is the combined transferred momentum and bosonic frequency ω_m . A similar expression is obtained in the particle-hole channels. Furthermore, because of the spin-rotational invariance of the Hubbard model, it is convenient to separate the particle-particle channel into singlet and triplet parts and the particle-hole channel into a magnetic part, which carries spin $S = 1$, and a charge density part, which has $S = 0$.

Approximation of the irreducible vertex

Just like the DCA self-energy $\Sigma(\mathbf{k})$ is approximated by the cluster self-energy $\Sigma_c(\mathbf{K})$, the DCA irreducible vertex function is approximated by a piecewise constant continuation of the corresponding cluster irreducible vertex function [11]

$$\Gamma_\alpha(k, k') = \sum_{\mathbf{K}, \mathbf{K}'} \phi_{\mathbf{K}}(\mathbf{k}) \Gamma_{c,\alpha}(K, K') \phi_{\mathbf{K}'}(\mathbf{k}'). \quad (30)$$

Here, for simplicity, we have used an abbreviated notation and included the transferred momentum q in the “channel” α , so that, for example, $\Gamma_{pp,q}(k, k') \equiv \Gamma_{pp}(k, q - k, q - k', k')$ in the

¹Depending on how particles and holes are involved in the scattering, one can define particle-particle and particle-hole channels

particle-particle channel. Like the self-energy, the cluster-irreducible vertex $\Gamma_{c,\alpha}(k, k')$ is determined from the solution of the cluster problem, that is by calculating the cluster two-particle correlation function $G_{2,c,\alpha}(k, k')$ and extracting $\Gamma_{c,\alpha}(k, k')$ from the BSE on the cluster [11]. For example, in the particle-particle channel for $Q = 0$

$$G_{2,c}(K, K') = G_{c,\uparrow}(K) G_{c,\downarrow}(-K) \delta_{K,K'} - \frac{T}{N} \sum_{K''} G_{c,\uparrow}(K) G_{c,\downarrow}(-K) \Gamma_{c,pp}(K, K'') G_{2,c}(K'', K') \quad (31)$$

Writing this in matrix notation in K, K' , one then has

$$\mathbf{\Gamma}_{c,pp} = -\frac{N}{T} \left[[\mathbf{G}_{2,c}^0]^{-1} - [\mathbf{G}_{2,c}]^{-1} \right], \quad (32)$$

where $G_{2,c}^0(K, K') = G_{c,\uparrow}(K) G_{c,\downarrow}(-K) \delta_{K,K'}$. The approximation in Eq. (30) is then used in the lattice BSE in Eq. (29) to calculate the two-particle Green's function on the lattice and from it the response function. Since $\Gamma_{c,\alpha}(K, K')$ depends on the cluster momenta $\mathbf{K}$ only, the problem may be further simplified by coarse-graining over momenta $\mathbf{k}$ within a patch around $\mathbf{K}$ for cases where the form factor $g_\varphi = 1$. For cases with a momentum-dependent form-factor $g_\varphi(\mathbf{k})$, a slightly modified version of this procedure is necessary [11].

Bethe-Salpeter eigenvalues and eigenfunctions

An alternative approach to determining the nature of the low-energy fluctuations is based on calculating the eigenvalues and eigenvectors of the BSE kernel [20–22]. For example, in the particle-particle channel with $q = 0$, one solves the eigenvalue equation

$$-\frac{T}{N} \sum_{k'} \Gamma_{pp}(k, k') G_\uparrow(k') G_\downarrow(-k') g_\alpha(k') = \lambda_\alpha g_\alpha(k), \quad (33)$$

with similar equations for the particle-hole channels. Note that it is possible to reconstruct the two-particle Green's function from these eigenvalues and eigenvectors

$$G_{2,pp}(k, k') = \sum_{\alpha} G_\uparrow(k) G_\downarrow(-k) \frac{g_\alpha(k) g_\alpha^*(k')}{1 - \lambda_\alpha}. \quad (34)$$

From this it becomes clear that an instability occurs when the leading eigenvalue λ_α becomes 1, and the momentum and frequency structure of the interaction is then reflected in the structure of the corresponding eigenvector $g_\alpha(k)$. This approach is in many ways more powerful than calculating the response function directly, because here one does not have to assume a given form factor $g_\alpha(\mathbf{k})$ and therefore cannot “miss” the structure of the dominant correlations.

Using the DCA (30) for the lattice vertex $\Gamma_{pp}(k, k')$ and assuming that the eigenvectors (as irreducible quantities) only depend on the cluster momenta $\mathbf{K}$, one can then sum (coarse-grain) over the Green's function legs to obtain an equation that only depends on coarse-grained and cluster quantities [21, 22]

$$-\frac{T}{N_c} \sum_{K'} \Gamma_{c,pp}(K, K') \chi_{0,pp}(K') g_\alpha(K') = \lambda_\alpha g_\alpha(K), \quad (35)$$

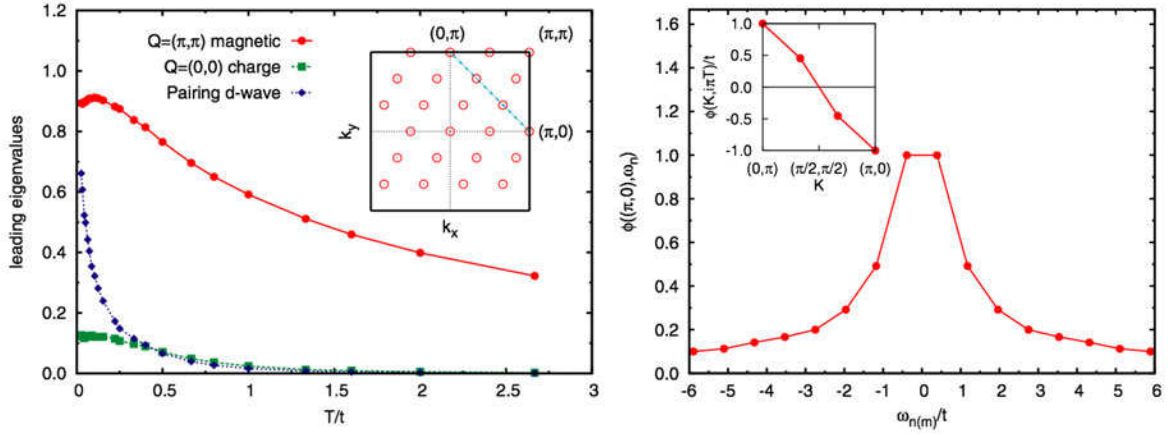


Fig. 6: Example of the solution of the Bethe-Salpeter equation in the various two-particle channels for a Hubbard model with $U = 4t$ and $\langle n \rangle = 0.85$: *Left panel: The leading eigenvalue in the $\mathbf{Q} = (\pi, \pi)$, $\omega_m = 0$ particle-hole magnetic channel dominates but saturates at low temperatures. The leading eigenvalue in the singlet $\mathbf{Q} = 0$, $\omega_m = 0$ particle-particle channel is found to have $d_{x^2-y^2}$ symmetry and increases towards 1 at low temperatures. The largest eigenvalue in the charge density particle-hole channel remains small. Right panel: The momentum dependence of the leading eigenvector $\phi_d(\mathbf{K}, \pi T)$ in the singlet particle-particle channel shows its $d_{x^2-y^2}$ dependence. Its frequency dependence reflects the frequency dependence of the pairing interaction Γ_{pp} . Figures taken from Ref. [21].*

with $\chi_{0,pp}(K) = N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) G_{\uparrow}(k) G_{\downarrow}(-k)$. While this reduces the complexity significantly, it also lowers the momentum resolution to the discrete set of cluster momenta $\mathbf{K}$. An example of the resulting leading eigenvalues and eigenvectors in the singlet particle-particle and also particle-hole magnetic and charge channels is shown in Fig. 6. These results illustrate how the calculation and analysis of the eigenvalues and eigenvectors of the BSE kernel provides a useful, unbiased method for determining the nature of the leading correlations of interacting many-body systems.

3 The DCA⁺ method

The DCA⁺ algorithm was developed as an extension of the DCA in order to introduce a lattice self-energy with continuous momentum $\mathbf{k}$ dependence and thus reduce its cluster-shape and size dependence [5]. Formally, this is achieved by expanding the lattice self-energy $\Sigma(\mathbf{k})$ into an arbitrarily large set of basis functions, instead of the patch functions that are used in the expansion of the DCA self-energy in Eq. (12). This means that in contrast to the DCA, where the lattice self-energy $\Sigma(\mathbf{k})$ is given by the cluster self-energy $\Sigma_c(\mathbf{K})$, in the DCA⁺ they are generally different.

3.1 General formalism

Using the identity

$$\frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \phi_{\mathbf{K}'}(\mathbf{k}) = \delta_{\mathbf{K}\mathbf{K}'}, \quad (36)$$

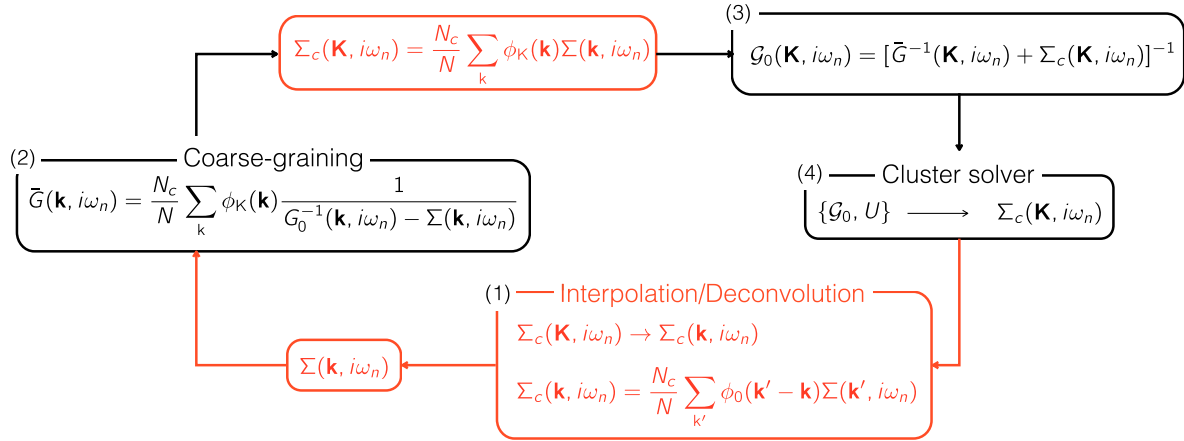


Fig. 7: DCA⁺ self-consistency loop: Parts colored in black are shared with the DCA algorithm, parts colored in red are additional steps. **(1)** To initialize the algorithm, one starts with a cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$. This is usually taken from a previous calculation, for example at a higher temperature, or set to zero. From $\Sigma_c(\mathbf{K}, i\omega_n)$, the lattice self-energy $\Sigma(\mathbf{k}, i\omega_n)$ is constructed using an interpolation of $\Sigma_c(\mathbf{K})$ followed by a deconvolution. Here, the interpolated cluster self-energy is also symmetrized according to the point-group symmetries of the lattice. **(2)** $\Sigma(\mathbf{k}, i\omega_n)$ is then used in the coarse-graining of the Green's function to give $\tilde{G}(\mathbf{K}, i\omega_n)$. **(3)** From $\tilde{G}(\mathbf{K}, i\omega_n)$, one then calculates the corresponding bare Green's function of the cluster, $\mathcal{G}_0(\mathbf{K}, i\omega_n)$, by removing the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$. This cluster/coarse-grained self-energy is obtained by coarse-graining the lattice self-energy $\Sigma(\mathbf{k}, i\omega_n)$. It is equal to the cluster self-energy that was used in the interpolation/deconvolution step unless it did not converge. **(4)** The bare cluster Green's function $\mathcal{G}_0(\mathbf{K}, i\omega_n)$ is used together with the interaction U to set up the effective cluster model, which is solved with a cluster solver such as the QMC algorithm discussed in Sec. 2.2. This provides a new result for the cluster self-energy $\Sigma_c(\mathbf{K}, i\omega_n)$, which is used in the next iteration in step (1).

for the patch functions defined in Eq. (9), one reverses the relation in Eq. (12) to a coarse-graining relation for the self-energy

$$\Sigma_c(\mathbf{K}) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma(\mathbf{k}). \quad (37)$$

This relation forms the basic equation of the DCA⁺ algorithm. It is fundamentally different from the DCA equation (12) in that it allows the lattice self-energy $\Sigma(\mathbf{k})$ to have continuous momentum dependence. Eq. (37) is trivially satisfied by the DCA approximation to $\Sigma(\mathbf{k})$ in Eq. (12). In general, however, Eq. (37) allows for a more physical approximation in which the lattice self-energy $\Sigma(\mathbf{k})$ can retain a smooth and continuous $\mathbf{k}$ dependence without the step discontinuities of the standard DCA. The inversion of Eq. (37) and extraction of a lattice self-energy $\Sigma(\mathbf{k})$ is a challenging task, which in Ref. [5] is solved by an initial interpolation of the cluster self-energy $\Sigma_c(\mathbf{K})$ and a subsequent deconvolution of a generalization of Eq. (37). This approach is described in Sec. 3.2. Assuming that this equation can be inverted to find the lattice

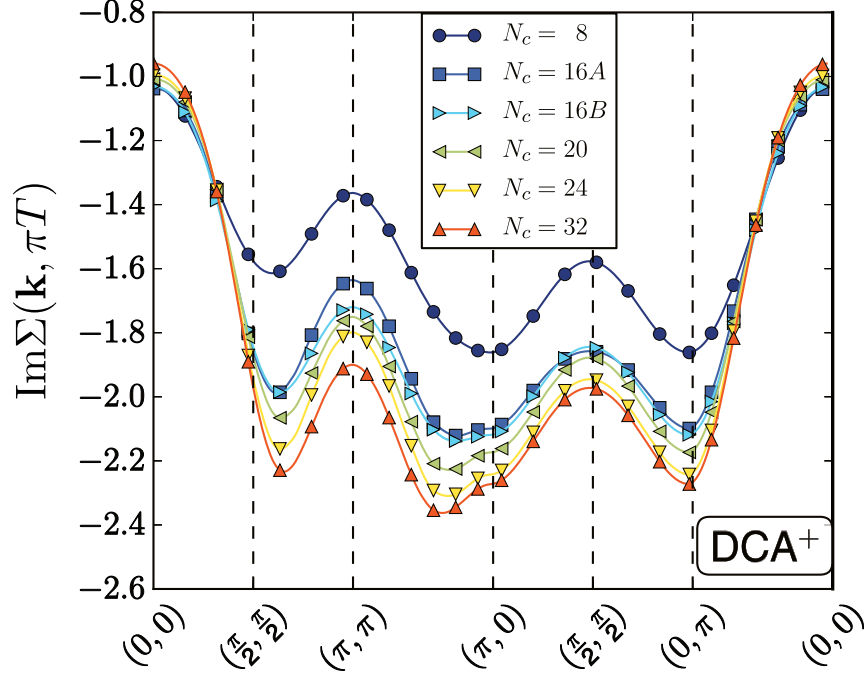


Fig. 8: DCA⁺ self-energy: DCA⁺ results for the imaginary part of the lattice self-energy $\Sigma(\mathbf{k}, \pi T)$ for a Hubbard model with the same parameters as those that were used in the plot of the DCA self-energy in Fig. 3, i.e., $U = 7t$, $t' = -0.15t$, $\langle n \rangle = 0.95$ and $T = 0.33t$. Compared with the DCA self-energy in Fig. 3, the DCA⁺ approximation gives a smooth momentum dependence with much weaker cluster shape and size dependence. Figure taken from Ref. [5].

self-energy $\Sigma(\mathbf{k})$, one then proceeds by calculating the coarse-grained Green's function

$$\bar{G}(\mathbf{K}, i\omega_n) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \frac{1}{G_0^{-1}(\mathbf{k}) - \Sigma(\mathbf{k}, i\omega_n)}. \quad (38)$$

just as in the standard DCA algorithm. Note, however, that in contrast to the DCA, here the full lattice self-energy with continuous momentum $\mathbf{k}$ dependence enters in the coarse-graining of the Green's function. The remainder of the algorithm remains identical to the DCA algorithm, including in particular the structure of the effective cluster problem. A detailed description of the self-consistent loop of the DCA⁺ is given in Fig. 7.

Results for the lattice self-energy $\Sigma(\mathbf{k})$ obtained in the DCA⁺ algorithm are shown in Fig. 8 for a Hubbard model with $t' = -0.15t$ and $U = 7t$ at a filling $\langle n \rangle = 0.95$. Here, the imaginary part of the lattice self-energy $\Sigma(\mathbf{k}, \pi T)$ at the first Matsubara frequency is shown for various cluster sizes. Compared to the corresponding DCA results in Fig. 3, the DCA⁺ algorithm by construction gives a smooth and therefore much more physical momentum dependence. One clearly sees that it gives a much more systematic convergence by reducing the cluster shape and size dependence of the DCA. This is especially clear for the 16A and 16B clusters. While the DCA self-energies in Fig. 3 differ significantly between these two clusters, the DCA⁺ results are almost identical with the exception of a small region around (π, π) . One should also stress that in the DCA⁺, the full lattice point-group symmetries are restored in the interpolation/deconvolution step for clusters that are not fully symmetric, while this is not the case in the DCA. This can be seen, for example, in the case of the 16B cluster.

3.2 Determination of lattice quantities from cluster quantities

The non-trivial task in the DCA⁺ algorithm is the generation of a lattice self-energy $\Sigma(\mathbf{k})$ with continuous momentum $\mathbf{k}$ dependence that satisfies the coarse-graining relation in Eq. (37) [5]. This inversion problem is obviously underdetermined since there are more $\mathbf{k}$ points than cluster $\mathbf{K}$ points. One can imagine interpolating the cluster self-energy $\Sigma_c(\mathbf{K})$ onto the fine $\mathbf{k}$ grid to give $\Sigma_c(\mathbf{k})$. This is discussed in more detail below. Then, it is convenient to generalize the coarse-graining relation in Eq. (37) to a convolution

$$\Sigma_c(\mathbf{k}) = \frac{N_c}{N} \sum_{\mathbf{k}'} \phi_0(\mathbf{k}' - \mathbf{k}) \Sigma(\mathbf{k}'). \quad (39)$$

Here, we have used the fact that one can write $\phi_{\mathbf{K}}(\mathbf{k}') = \phi_{\mathbf{K}=0}(\mathbf{k}' - \mathbf{K})$ and then generalized the cluster momentum $\mathbf{K}$ to $\mathbf{k}$ to give $\phi_0(\mathbf{k}' - \mathbf{k})$. Eq. (39) now has the form of a convolution, which needs to be inverted. One could, for example, imagine expanding the lattice self-energy in terms of cubic Hermite splines for fixed frequency $i\omega_n$ [5]

$$\Sigma(\mathbf{k}) = \sum_i \mathcal{B}_{i\omega_n}(\mathbf{k} - \mathbf{k}_i) \sigma(\mathbf{k}_i), \quad (40)$$

where $\sigma(\mathbf{k}_i)$ are the expansion coefficients. Using this expansion, Eq. (39) then becomes

$$\Sigma_c(\mathbf{k}_i) = \sum_j P_{ij} \sigma(\mathbf{k}_j) \quad (41)$$

with the projection matrix

$$P_{ij} = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_0(\mathbf{k} - \mathbf{k}_i) \mathcal{B}_{i\omega_n}(\mathbf{k} - \mathbf{k}_j) \quad (42)$$

and the problem of deconvoluting Eq. (39) has now become a problem of inverting the projection matrix P . The properties of this matrix were discussed in detail in Ref. [5]. There it was shown that P is a near-singular matrix with eigenvalues that become smaller the more the corresponding eigenvector is delocalized. An inversion of P is therefore only possible if eigenvalues smaller than some cut-off ϵ and their corresponding eigenvectors are neglected. If the cluster self-energy is long-ranged and extends beyond the boundary of the real-space cluster, it cannot be represented by the retained smaller-range eigenvectors and eigenvalues of P and the DCA⁺ calculation does not converge. In this case the cluster size has to be increased. The analysis of the projection matrix and the convergence of a DCA⁺ calculation thus provides a useful measure of whether the cluster is large enough to capture the correlations.

3.2.1 Determination of lattice self-energy

In practice, the inversion of the projection operator P is numerically unstable, and it is therefore favorable to extract the lattice self-energy $\Sigma(\mathbf{k})$ by deconvoluting the interpolated cluster self-energy $\Sigma_c(\mathbf{k})$. To this end, a Bayesian based Richardson-Lucy algorithm is used after the cluster self-energy has been interpolated [5].

Interpolation of cluster self-energy

Instead of interpolating $\Sigma_c(\mathbf{K})$ directly, it was shown in Ref. [5] that a better result is obtained when $\Sigma_c(\mathbf{K})$ is first transformed to $\mathcal{T}(\Sigma_c(\mathbf{K}))$ with $\mathcal{T}(z) = (z - i)^{-1}$. This form is similar to the cumulant used in Ref. [23] for the interpolation of cellular dynamical mean-field results. One then applies a Wannier-interpolation [24] to obtain the interpolated cluster self-energy

$$\Sigma_c(\mathbf{k}) = \mathcal{T}^{-1} \left[\sum_{\mathbf{R}} e^{-i\mathbf{k}\mathbf{R}} \underbrace{\left(\sum_{\mathbf{K}} e^{-i\mathbf{k}\mathbf{R}} \mathcal{T}(\Sigma_c(\mathbf{K})) \right)}_{(\mathcal{T}\Sigma)_R} \right]. \quad (43)$$

One should always verify that $(\mathcal{T}\Sigma)_R$ does not extend beyond the cluster boundary. When this is the case, this procedure does not introduce any artificial features in the interpolated self-energy, which is often the case with splines, where overshoots are not uncommon. Once the cluster self-energy is interpolated, lattice point-group symmetries are imposed on the interpolated self-energy $\Sigma(\mathbf{k})$. This is especially important for clusters that do not have the full lattice symmetry, such as the 16B cluster in Fig. 1. The interpolated cluster self-energy $\Sigma_c(\mathbf{k})$ from Eq. (43) is then used in the convolution in Eq. (39) to determine the lattice self-energy $\Sigma(\mathbf{k})$.

Deconvolution of lattice self-energy

Since the patch function $\phi_0(\mathbf{k})$ in Eq. (39) acts as a box-car filter on $\Sigma(\mathbf{k})$, one can apply the Richardson-Lucy algorithm [25, 26] to solve this deconvolution problem. This algorithm is based on Bayesian inference, which finds the most likely solution $\Sigma(\mathbf{k})$ by using an iteration procedure [5]

$$\Sigma_{i+1}(\mathbf{k}) \leftarrow \Sigma_i(\mathbf{k}) \int d\mathbf{k}' \frac{\phi_0(\mathbf{k} - \mathbf{k}') \Sigma_c(\mathbf{k}')}{\int d\mathbf{k}'' \phi_0(\mathbf{k}' - \mathbf{k}'') \Sigma_i(\mathbf{k}'')}. \quad (44)$$

This procedure is carried out separately for the real and imaginary parts of the self-energy. As discussed in Ref. [5], this algorithm converges well when correlations are short-ranged and contained within the cluster. When the range of the correlations exceeds the cluster, one generally finds that this algorithm converges slowly and gives an estimate for the lattice self-energy $\Sigma(\mathbf{k})$ that, after coarse-graining, deviates considerably from the cluster self-energy $\Sigma_c(\mathbf{K})$ and thus does not satisfy the DCA⁺ constraint in Eq. (37). Fig. 9 shows the result of this procedure for a calculation of a 32-site cluster, for which the cluster self-energy does not exceed the cluster size and thus the Richardson-Lucy procedure converges quickly and well. In this case, one sees that the coarse-grained lattice self-energy $\bar{\Sigma}(\mathbf{K}) = N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma(\mathbf{k})$ agrees very well with the cluster self-energy $\Sigma_c(\mathbf{K})$.

3.2.2 Deconvolution of the irreducible vertex function

In order to extend the DCA⁺ algorithm to the two-particle level, it is necessary to determine an irreducible vertex function $\Gamma_{\alpha}(k, k')$ with continuous momentum dependence. Just as for

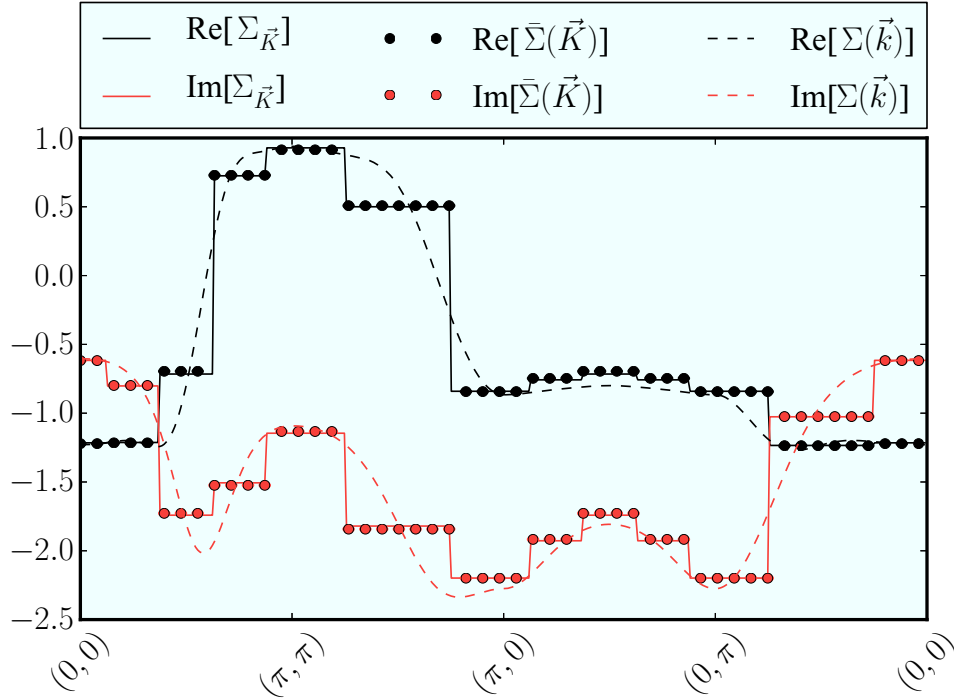


Fig. 9: Richardson-Lucy deconvolution of the lattice self-energy: DCA^+ calculation for a 32-site cluster with $t' = -0.15t$ and $U = 7t$ at a filling of $\langle n \rangle = 0.95$ and $T = 0.33t$. Shown are real (black) and imaginary (red) parts of the continuous lattice self-energy $\Sigma(\mathbf{k})$, its coarse-grained result $\bar{\Sigma}(\mathbf{K}) = N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma(\mathbf{k})$, and the cluster self-energy $\Sigma_c(\mathbf{K})$ for the lowest Matsubara frequency $\omega_0 = \pi T$ for a high-symmetry path in the Brillouin zone. The coarse-grained lattice self-energy $\bar{\Sigma}(\mathbf{K})$ agrees very well with the cluster self-energy $\Sigma_c(\mathbf{K})$.

the self-energy, one again uses the identity in Eq. (36), in this case twice, to reverse the DCA relation (30) to get

$$\Gamma_{c,\alpha}(K, K') = \frac{N_c^2}{N^2} \sum_{\mathbf{k}, \mathbf{k}'} \phi_{\mathbf{K}}(\mathbf{k}) \Gamma_{\alpha}(k, k') \phi_{\mathbf{K}'}(\mathbf{k}'). \quad (45)$$

In analogy to Eq. (37) for the self-energy, the DCA^+ lattice irreducible vertex function is thus related to the cluster irreducible vertex $\Gamma_{c,\alpha}$ through a coarse-graining relation. In the standard DCA algorithm, where $\Gamma_{\alpha}(k, k')$ is a piecewise-constant continuation (30) of the cluster $\Gamma_{c,\alpha}(K, K')$, this requirement is trivially satisfied. In the DCA^+ algorithm, however, one wants to find a $\Gamma_{\alpha}(k, k')$ with continuous momentum dependence and without step discontinuities that satisfies Eq. (45). Assuming that this equation can be inverted to find the lattice $\Gamma_{\alpha}(k, k')$, one can then solve the lattice BSE Eq. (29) to calculate the two-particle Green's function G_2 , or equivalently, solve the eigenvalue problem (33). Since the full lattice $\mathbf{k}$ dependence is retained in the DCA^+ formalism, the resulting eigenvectors $g_{\alpha}(\mathbf{k})$ will have smooth momentum dependence.

Evidently, the difficult task is the solution or inversion of the integral equation (45) to obtain the lattice $\Gamma_{\alpha}(k, k')$. Here, one carries out a similar procedure as for the self-energy by starting

with an interpolation. For this, one uses a singular value decomposition

$$\Gamma_{c,\alpha}(K, K') = \sum_i \sigma_i U_i(K) V_i(K') \quad (46)$$

to write the cluster vertex in a separable representation. Here, σ_i are the singular values, the columns of U contain the left singular vectors and V contains the right singular vectors. Then, one proceeds by interpolating $U_i(\mathbf{K}, i\omega_n)$ and $V_i(\mathbf{K}, i\omega_n)$ onto a fine $\mathbf{k}$ grid for each $i\omega_n$ to give an interpolated cluster vertex

$$\Gamma_{c,\alpha}(k, k') = \sum_i \sigma_i U_i(k) V_i(k'), \quad (47)$$

where $U_i(k)$ and $V_i(k')$ with $k = (\mathbf{k}, i\omega_n)$ are cubic spline interpolations in momentum space of $U_i(K)$ and $V_i(K')$, respectively. In the following we drop the frequency arguments for simplicity. Just as for the self-energy, the coarse-graining in Eq. (45) is then generalized to a convolution

$$\Gamma_{c,\alpha}(k_1, k_2) = \frac{N_c^2}{N^2} \sum_{\mathbf{k}, \mathbf{k}'} \phi_0(\mathbf{k} - \mathbf{k}_1) \Gamma_\alpha(k, k') \phi_0(\mathbf{k}' - \mathbf{k}_2), \quad (48)$$

where again the generalized patch function $\phi_0(\mathbf{k} - \mathbf{K}) = \phi_{\mathbf{K}}(\mathbf{k})$. Then, the lattice vertex is expanded into a set of basis functions $\{\mathcal{B}\}$.² If cubic Hermite splines [27] are used as basis-functions, the continuous lattice vertex is expanded as

$$\Gamma_\alpha(k, k') = \sum_{i,j} \mathcal{B}_{i\omega_n}(\mathbf{k} - \mathbf{k}_i) \gamma_\alpha(k_i, k_j) \mathcal{B}_{i\omega_n'}(\mathbf{k}' - \mathbf{k}_j). \quad (49)$$

Here, the vectors $\mathbf{k}_i$ span a fine $\mathbf{k}$ grid that covers the first Brillouin zone. Using the expansion in Eq. (49), one can rewrite Eq. (48) as a matrix-equation,

$$\begin{aligned} \Gamma_{c,\alpha}(k, k') &= \sum_{i,j} \Phi_{i\omega_n}(\mathbf{k}_1, \mathbf{k}_i) \gamma^\alpha(k_i, k_j) \Phi_{i\omega_n'}(\mathbf{k}_2, \mathbf{k}_j) \\ \Phi_{i\omega_n}(\mathbf{k}_i, \mathbf{k}_j) &= \frac{N_c}{N} \sum_{\mathbf{k}} \phi_0(\mathbf{k} - \mathbf{k}_i) \mathcal{B}_{i\omega_n}(\mathbf{k} - \mathbf{k}_j) \end{aligned} \quad (50)$$

Then, using a singular value decomposition of the matrix Φ ,

$$\Phi_{i\omega_n}(\mathbf{k}_i, \mathbf{k}_j) = \sum_i \sigma_i^\Phi u_i^\Phi(k_i) v_i^\Phi(k_j), \quad (51)$$

(note that all quantities on the right hand side carry an implicit $i\omega_n$ -dependence) one can formally invert Eq. (50) and obtain an explicit form for the lattice-vertex $\Gamma(k, k')$

$$\begin{aligned} \Gamma_\alpha(k, k') &= \sum_i \sigma_i \tilde{u}_i(k) \tilde{v}_i(k') \\ \tilde{u}_i(k) &= \sum_j v_j^\Phi(k) \frac{\langle u_j^\Phi(k), U_i(k) \rangle}{\sigma_j^\Phi}, \\ \tilde{v}_i(k) &= \sum_j \frac{\langle V_i(k), v_j^\Phi(k) \rangle}{\sigma_j^\Phi} u_j^\Phi(k). \end{aligned} \quad (52)$$

²The set of basis-functions $\{\mathcal{B}\}$ can be freely chosen, since the DCA⁺ is not dependent on the choice of the basis-functions. Usually, one uses cubic Hermite splines [27].

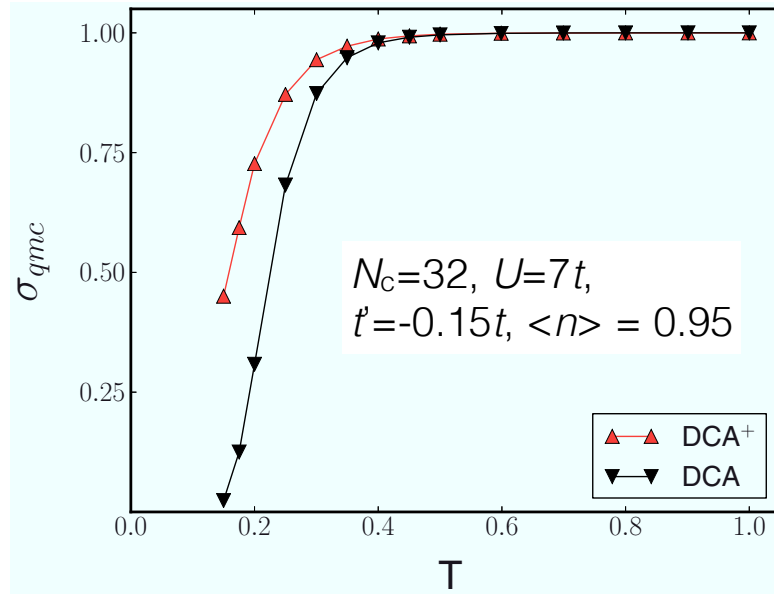


Fig. 10: DCA⁺ reduction of the sign problem: *The average QMC sign compared between DCA and DCA⁺ for a 32-site cluster with $U = 7t$, $t' = -0.15t$ and $\langle n \rangle = 0.95$. The DCA⁺ algorithm reduces the sign problem of the DCA further. Figure reproduced from Ref. [5].*

Here, $\langle a, b \rangle$ represents the usual dot-product between vectors a and b . One generally finds that the singular values of the Φ -matrix decay rapidly. Note that this set of equations has a stable solution only if the expansion coefficients $\langle u_j^\Phi(k), U_i(k) \rangle$ and $\langle V_i(k), v_j^\Phi(k) \rangle$ decay faster than the singular values. For numerical reasons, one generally imposes an upper bound to the inverse of the singular values σ_i^Φ . Due to the statistical noise of the Monte Carlo solution, the expansion coefficients may become small, but usually not zero. To stabilize the inversion, it is helpful to convert $1/\sigma_i^\Phi$ to the value $\min\{1/\epsilon, 1/\sigma_i^\Phi\}$, where ϵ is a small but finite number. In this way, all components are taken into account [5].

3.3 Reduction of the QMC sign problem

As discussed in Sec. 2.3, DCA QMC calculations have a much less severe sign problem than finite-size QMC calculations. Fig. 10 compares the QMC sign between DCA⁺ and DCA calculations. One sees that the DCA⁺ further reduces the sign problem (at a given low temperature it has a larger average sign). Intuitively, this can be understood in the following way: Empirically, it is known that the sign problem gets worse when the strength of the correlations increases. The step discontinuities in the k -dependence of the DCA self-energy lead to artificial long-range correlations in real space. The smooth momentum dependence of the DCA⁺ lattice self-energy removes these artificial long-range correlations, and this is believed to reduce the sign problem. In any case, with the larger average sign in the DCA⁺ it is possible to study the physics of fermionic systems in larger clusters, at lower temperatures or larger interaction strength.

4 Applications to the 2D Hubbard model

DCA simulations of the single-band Hubbard model [1, 28, 29] have found behavior that is reminiscent of what is observed in the cuprate high-temperature superconductors. In the paramagnetic state, the low-energy spin excitations become suppressed below the crossover temperature T^* , and a pseudogap opens in the density of states at the chemical potential. At lower temperatures, one finds a finite-temperature transition to antiferromagnetic long-range order at low doping, while at larger doping, an instability to a $d_{x^2-y^2}$ -wave superconducting state is found.

4.1 Antiferromagnetism

The finite-temperature antiferromagnetic instability found in DCA calculations of the Hubbard model is an apparent contradiction to the Mermin-Wagner theorem [30], which demands that such a transition is suppressed to zero temperature. In the DCA, it results from the fact that long-range (small q) fluctuations beyond the cluster size, which suppress ordering, are neglected and replaced by a mean-field. With increasing cluster size, however, the DCA progressively includes longer-ranged fluctuations while retaining some mean-field character. Larger clusters are thus expected to systematically drive the Néel temperature to zero and hence recover the Mermin-Wagner theorem in the infinite-cluster-size limit. T_N is the temperature at which the spin susceptibility

$$\chi_s(\mathbf{q}) = \int_0^\beta d\tau \langle T_\tau S^z(\mathbf{q}, \tau) S^z(-\mathbf{q}, 0) \rangle, \quad (53)$$

diverges for $\mathbf{q} = (\pi, \pi)$. Here $S^z(\mathbf{q}) = 1/N \sum_{\mathbf{k}} (c_{\mathbf{k}+\mathbf{q}\uparrow}^\dagger c_{\mathbf{k}\uparrow} - c_{\mathbf{k}+\mathbf{q}\downarrow}^\dagger c_{\mathbf{k}\downarrow})$ is the z -component of the spin operator. $\chi_s(\mathbf{q})$ is calculated from the two-particle Green's function in the particle-hole magnetic channel, in a similar fashion as the pair-field susceptibility described in Sec. 2.4. The basic approximation is that the irreducible vertex function in the Bethe-Salpeter equation for this two-particle Green's function is replaced by the irreducible vertex function calculated in the effective cluster problem.

Fig. 11 shows DCA results for the Néel temperature T_N of a half-filled Hubbard model with $U = 8t$ reproduced from Ref. [31]. T_N decreases slowly with increasing cluster size N_c . This decrease can be understood from a scaling argument in which one assumes that an antiferromagnetic transition occurs when the correlation length ξ reaches the linear cluster size, i.e., $\xi(T_N) = \sqrt{N_c}$ in 2D. Since correlations develop exponentially with decreasing temperature in 2D, i.e. $\xi(T) \sim e^{C/T}$, one obtains a logarithmic decrease of $T_N(N_c)$ with N_c as seen in the data for $N_c > 4$. This decrease is consistent with $T_N \rightarrow 0$ in the infinite-size cluster limit as required by the Mermin-Wagner theorem. The clusters with $N_c = 2$ and 4 are special because their coordination number is reduced from four. In these clusters, local singlet states form below a temperature of $J = 4t^2/U$ and suppress antiferromagnetism.

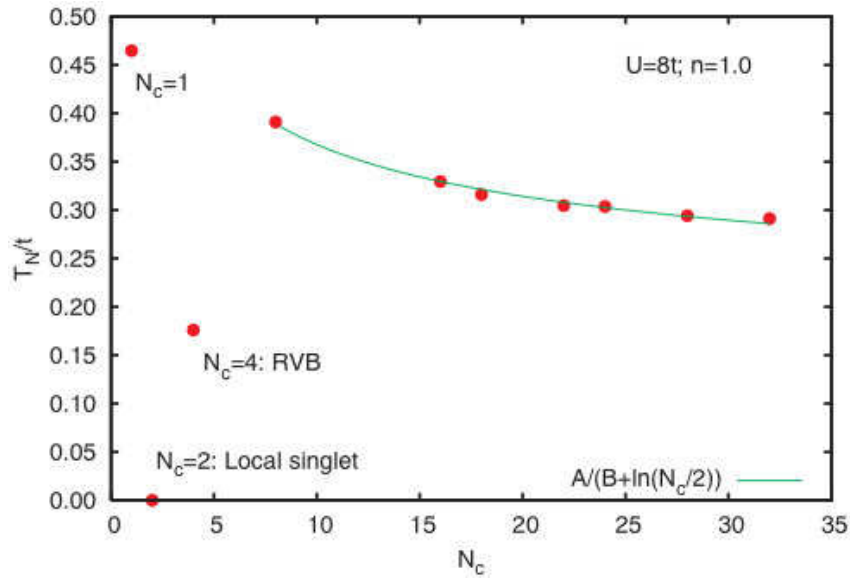


Fig. 11: DCA calculation of the Néel temperature T_N versus cluster size in a half-filled Hubbard model with $U = 8t$: $T_N \rightarrow 0$ for $N_c \rightarrow \infty$ as required by the Mermin-Wagner theorem. The solid line represents a fit to the function $A/[B + \ln(N_c)/2]$ obtained from the scaling ansatz $\xi(T_N) = \sqrt{N_c}$. For $N_c = 2$ and 4 local singlets form in the cluster that suppress the AF state. Figure reproduced from Ref. [31].

4.2 Pseudogap

Measurements of magnetic susceptibility and Knight shifts have found the opening of a pseudogap, a partial suppression of low energy spin fluctuations, in the underdoped cuprates [32]. ARPES experiments have found that this gap is anisotropic, opening in the anti-nodal $(\pi, 0)$ and $(0, \pi)$ regions of the Fermi surface. This state appears below a temperature T^* , which rises with decreasing hole doping as the Mott-insulating half-filled state is approached. DCA calculations of the 2D Hubbard model have found a similar pseudogap in the bulk spin susceptibility $\chi_s(\mathbf{q} = 0)$ [11, 33] and in calculations of the single-particle spectral weight [28, 34].

To illustrate these results, and to further compare the convergence properties between DCA and DCA^+ , we show in Fig. 12 DCA (left panel) and DCA^+ results³ (right panel) for the temperature dependence of the $\mathbf{q} = 0$ spin susceptibility χ_s . The DCA and DCA^+ results display similar behavior for $\chi_s(\mathbf{q} = 0, T)$ with a peak at $T^*(N_c)$ and a decrease below T^* , reflecting the opening of the pseudogap. For the DCA, however, one observes a strong cluster size dependence and poor convergence even for the largest clusters that can still be simulated before the QMC sign problem makes calculations impossible. In the DCA^+ results, one sees that convergence is reached much sooner. And in addition, because of the reduced sign problem, larger clusters can be reached, for which one sees that T^* is converged.

³For the DCA, $\chi_s(\mathbf{q} = 0, T)$ was determined in the same way as the $\mathbf{q} = (\pi, \pi)$ susceptibility in the previous section, using the same algorithm as that described for the pair-field susceptibility in Sec. 2.4, just for the magnetic spin $S = 1$ particle-hole channel instead of the particle-particle channel. For the DCA^+ results, the full DCA^+ lattice self-energy $\Sigma(\mathbf{k})$ was used in the Green's functions entering this formalism, but the deconvolution of the irreducible vertex function described in Sec. 3.2 was not performed. Instead this quantity was set equal to the cluster irreducible vertex function, just as in the DCA. This approximation is justified because of the weak internal $\mathbf{k}$ dependence of the correlations in the antiferromagnetic particle-hole channel.

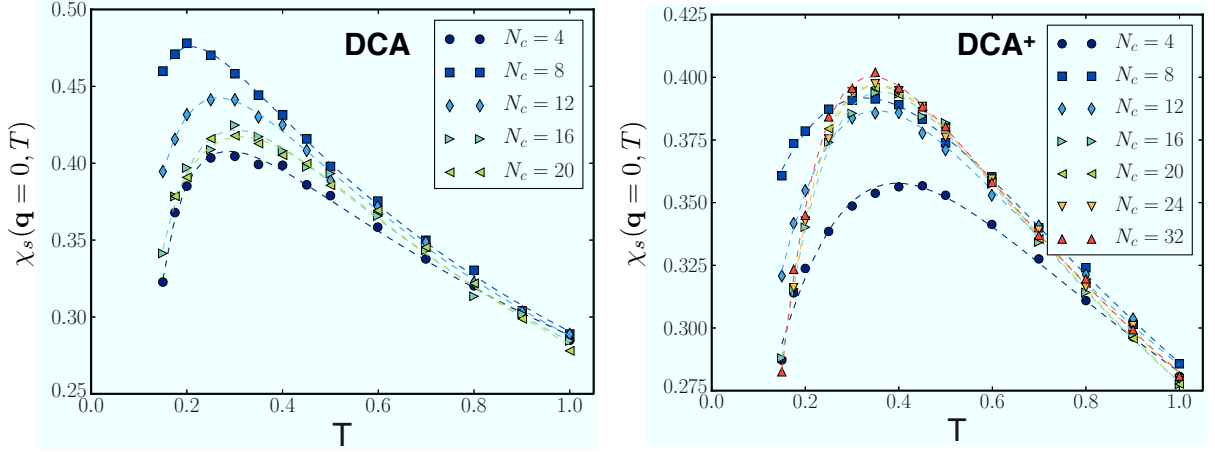


Fig. 12: Pseudogap in the uniform spin susceptibility: *DCA* (left panel) and *DCA*⁺ (right panel) results for the uniform spin susceptibility $\chi_s(\mathbf{q} = 0, T)$ versus temperature for a Hubbard model with $U = 7t$, $t' = -0.15t$ and $\langle n \rangle = 0.95$. $\chi_s(\mathbf{q} = 0, T)$ peaks at a temperature $T^*(N_c)$ and decreases at lower temperatures as the pseudogap opens. The *DCA*⁺ converges $T^*(N_c)$ more rapidly and its reduced sign problem enables calculations at lower temperatures. Figures reproduced from Ref. [5].

4.3 Superconductivity

DCA methods have also been used extensively to investigate the possibility and nature of superconductivity in the 2D Hubbard model [1, 21, 29, 31], because of its relevance to the cuprate high-temperature superconductors [35, 36]. Early DCA calculations used a 2×2 cluster to map out the temperature versus doping phase diagram for $U = 8t$ and found $d_{x^2-y^2}$ -wave superconductivity over a large finite-doping region with a maximum $T_c \sim 0.08t$ at a filling $\langle n \rangle = 0.95$ [33]. A DCA study of larger clusters up to 26 sites was then carried out for $U = 4t$ and found a superconducting transition at $T_c \approx 0.023t$ for $\langle n \rangle = 0.9$ [31]. But convergence was poor due to the cluster shape and size dependence of the standard DCA. Recent *DCA*⁺ calculations for the same parameters could reach clusters up to 52 sites, for which the results were asymptotically converged, and a finite-size scaling analysis similar to that discussed in Sec. 4.1 gave a $T_c = 0.02t$ [6].

As an illustration of recent progress on this issue, Fig. 13 shows the results of standard DCA and *DCA*⁺ calculations for a more realistic $U = 7t$ and $\langle n \rangle = 0.9$ taken from Ref. [6]. The *DCA*⁺ results in this figure were obtained by using a different coarse-graining [37], in which the patch function $\phi_{\mathbf{K}}(\mathbf{k})$ for a given $\mathbf{K}$ is finite over several intervals instead of just a single one. This leads to a further reduction of the sign problem, which is also reflected in the difference in the maximum cluster size between the DCA and *DCA*⁺ results in Fig. 13. The standard DCA calculations with the usual coarse-graining are limited to only 12 sites or fewer, for this value of U , and T_c is clearly not converged. The *DCA*⁺ calculations with modified coarse-graining, however, are able to reach clusters of up to 28 sites. In particular, one sees that for clusters between 12 and 26 sites, T_c changes by only 10% between different N_c , and one can estimate $T_c \approx 0.052t$. For $N_c = 24$ and $T = 0.05t$, the right panel of Fig. 13 displays the momentum dependence of the leading eigenvector $\phi_d(\mathbf{k}, \pi T)$ of the BSE along the diagonal from $(\pi, 0)$ to

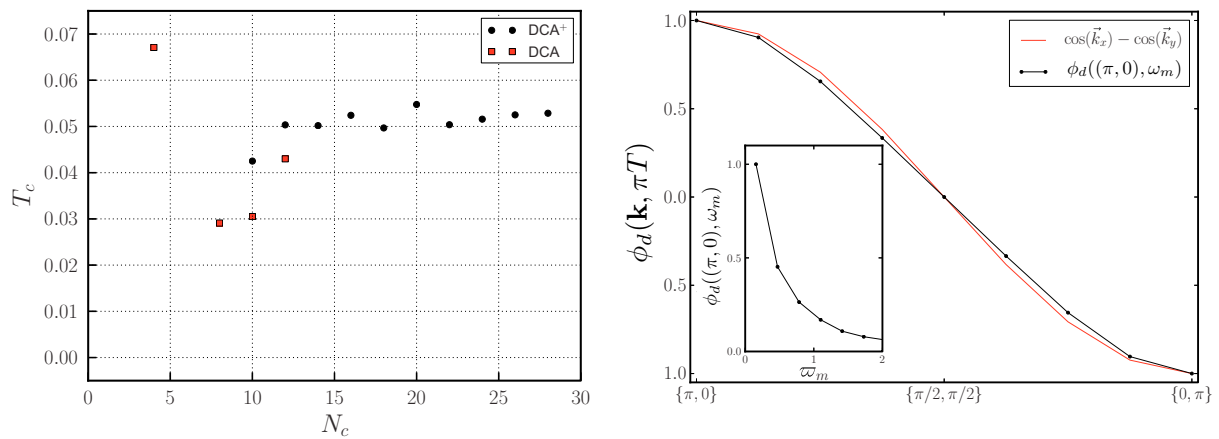


Fig. 13: Superconducting transition temperature T_c and BSE leading eigenvector of a Hubbard model with $U = 7t$ and $\langle n \rangle = 0.9$: (Left panel) The QMC sign problem limits the largest accessible cluster size. DCA results (red squares) for T_c are not converged. The DCA⁺ with modified coarse-graining (see text) allows calculations for larger clusters, for which T_c hardly changes (black circles). (Right panel) The $\mathbf{k}$ -dependence of the leading eigenvector $\phi_d(\mathbf{k}, \pi T)$ of the particle-particle BSE for $T = 0.05t$ and $N_c = 24$ is close to $\cos k_x - \cos k_y$; its frequency dependence is plotted in the inset. Figures reproduced from Ref. [6].

$(0, \pi)$. Compared to the coarse $\mathbf{K}$ -dependence of the DCA result in Fig. 6, the DCA⁺ provides information with much higher resolution in $\mathbf{k}$, and thus allows one to estimate deviations from the simple d -wave $\cos k_x - \cos k_y$ form factor with higher precision.

5 General discussion and concluding remarks

We have seen that the DCA approximation and its DCA⁺ extension are quantum cluster methods that map the bulk lattice problem onto a finite-size cluster embedded in a mean-field host that is designed to represent the remaining degrees of freedom. This is achieved through a coarse-graining of the momentum space, which effectively reduces the degrees of freedom to those of a cluster, while retaining the effects of the remaining bulk degrees of freedom as a mean-field, which the cluster is coupled to. Correlations on the cluster are treated accurately using, for example, quantum Monte Carlo methods, while longer-range correlations are described on the mean-field level. Because of translational invariance of the bulk lattice, the properties of the mean-field host are calculated from the effective cluster problem in a self-consistent manner. In the DCA, the mean-field host reflects the correlations described by the cluster self-energy directly. In the DCA⁺, in contrast, the mean-field is calculated using a lattice self-energy with continuous bulk momentum dependence that is generated from the cluster self-energy. This leads to an improved cluster shape and size dependence, and to a reduction of the sign problem in the underlying QMC cluster solver. In what follows, we give a brief discussion of several other fundamental features that are common to both the DCA and DCA⁺ algorithms.

Nature of approximation and limitations

The basic assumption of the coarse-graining approximation in both the DCA and DCA⁺ methods is that correlations are short-ranged so that the $\mathbf{k}$ -dependence of the self-energy $\Sigma(\mathbf{k}, i\omega_n)$ and the irreducible vertex functions $\Gamma_\alpha(\mathbf{k}, i\omega_n, \mathbf{k}', \omega_{n'})$ is well approximated by a coarse grid of cluster momenta $\mathbf{K}$ at intervals $\Delta K = 2\pi/L_c$, where L_c is the linear cluster size. Obviously, this approximation assumes a self-energy and irreducible vertex functions that are only weakly $\mathbf{k}$ -dependent, or equivalently, correlations that are short-ranged and do not extend beyond a length of $L_c/2$. This type of approximation is therefore expected to be appropriate for cases with significant screening, where correlations are short-ranged. It is clear that such an approximation breaks down in the proximity of classical or quantum phase transitions, where the critical behavior is governed by the long wave-length fluctuations. Since the fluctuations beyond $L_c/2$ are replaced by a mean-field, the DCA and DCA⁺ approaches give mean-field critical behavior and possibly predict instabilities at finite temperature even when prohibited by the Mermin-Wagner theorem, as is the case, for example, for the observed finite-temperature antiferromagnetic transition in the half-filled 2D Hubbard model discussed in Sec. 4.1. The fact that instabilities occur when the correlation length extends beyond the cluster size, however, allows one to use finite-size scaling in combination with calculations on increasingly larger clusters to extrapolate to the exact infinite cluster size limit.

Causality

A particular challenge in the development of quantum cluster methods such as the DCA and DCA⁺ has been the requirement of causality [1], i.e., that the algorithm give a self-energy with negative imaginary part, i.e. $\text{Im } \Sigma(\mathbf{k}, \omega + i0^+) < 0$. For the standard DCA algorithm, it is possible to prove causality [1,3]. Hettler *et al.* also remarked [3] that the use of a simple interpolation of the cluster self-energy to generate a lattice self-energy in the coarse-graining step Eq. (11) will lead to causality violations if $\text{Im } \Sigma_c(\mathbf{K}, i\omega_n)$ has a minimum somewhere in the Brillouin zone, which is generally the case. In Ref. [38], such causality violations were more fundamentally related to the “ringing” phenomenon in Fourier-analysis. To avoid these problems in the DCA, one interpolates the cluster self-energy only after the calculation is converged. The DCA⁺ algorithm, however, is fundamentally different from a simple interpolation of the cluster self-energy. The constraint of the DCA⁺ algorithm in Eq. (37) demands that the coarse-grained lattice self-energy $N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma(\mathbf{k})$ is equal to the cluster self-energy $\Sigma_c(\mathbf{K})$. Except for very special cases, this results in a lattice self-energy $\Sigma(\mathbf{k})$ that is different from a simple interpolation of the cluster self-energy $\Sigma_c(\mathbf{K})$. The arguments for causality violations that arise when a simple interpolation is used in Refs. [3, 38] thus do not apply, and the DCA⁺ algorithm does not necessarily violate causality. Although a rigorous proof for the causality of the DCA⁺ is not available, violations of causality have not yet been observed in any of the applications of this method [5].

Thermodynamic consistency

Thermodynamic consistency in the Baym and Kadanoff sense [39] ensures that observables calculated from the single-particle Green's function agree with those calculated from the two-particle Green's function. For example, the pair-field susceptibility may be determined from the two-particle formalism described in Sec. 2.4. Alternatively, one may extend the single-particle formalism to allow for calculations in the superconducting phase by introducing an anomalous Green's function $F_{ij}(\tau) = -\langle T_\tau c_{i\uparrow}(\tau) c_{j\downarrow}(0) \rangle$. The s -wave pair-field susceptibility, for example, is then obtained from $P_s = \partial \Delta_s / \partial \Psi|_{\Psi=0}$, where Ψ is the applied pair-field and $\Delta_s = T/N \sum_{\mathbf{k}, \omega_n} F(\mathbf{k}, i\omega_n)$. Thermodynamically consistent algorithms will give the same result, irrespective of whether P_s is calculated on the single-particle level or whether it is calculated on the two-particle level. An algorithm is thermodynamically consistent if it is self-consistent and if the two-particle irreducible vertex function Γ is related to the single-particle self-energy through $\Gamma = \delta \Sigma[G] / \delta G$. This is the case in the DCA where both Σ and Γ are obtained as derivatives of the same Baym-Kadanoff Φ -functional (see Appendix and [1]) and approximated by their respective cluster quantities Σ_c and Γ_c . In the DCA⁺, it is important that the relation between Γ and Γ_c is consistent with the relation between Σ and Σ_c . In both cases, the cluster quantities are related to their lattice counterparts by a coarse-graining relation, as seen from Eqs. (37) and (45), and it was shown in Ref. [6] that these relations indeed satisfy $\Gamma = \delta \Sigma[G] / \delta G$. Accordingly, the DCA⁺ algorithm is thermodynamically consistent if a similar procedure is used to extract the lattice quantities.

Conclusions

In this lecture we have tried to communicate that the DCA and DCA⁺ approximations are powerful theoretical approaches that enable reliable and, in certain cases, controlled studies of the rich phenomenology in systems dominated by strong electronic correlations. They are non-perturbative in nature and, in conjunction with numerically exact QMC cluster solvers, provide an unbiased tool to understand the physics in these systems. They enable the calculation of static and dynamic single-particle and two-particle correlation functions and thus allow for the determination of various experimentally accessible observables, such as those related to photoemission, transport, magnetism, and superconductivity. Despite the reduction in complexity, the solution of the effective cluster problem remains a challenging task, and its complexity increases rapidly as the number of cluster degrees of freedom increases. Hence these methods have been applied almost exclusively to low-energy effective descriptions of the full many-body problem, such as the single-band Hubbard model discussed in this lecture. Applications to more complex multi-orbital models are required when the specifics of certain materials need to be included. This can be achieved along the lines of the local density approximation + DMFT approach, where the models studied by DMFT or DCA are parametrized by electronic structure calculations. While this remains an important and challenging task, the progress in [40–42] demonstrates that calculations along these lines are within reach.

Acknowledgments

I would like to acknowledge discussions with E. Gull, M. Jarrell, J. Keller, A. Macridin, S. Okamoto, T. Pruschke, D.J. Scalapino, T.C. Schulthess and P. Staar that have led to a deeper understanding and appreciation of the formalisms presented in this lecture. Support from the Center for Nanophase Materials Sciences and resources of the Oak Ridge Leadership Computing Facility, which are DOE Office of Science User Facilities, are also greatly acknowledged.

Appendix

A DCA and DCA⁺ as self-energy functional approximations

General methods may be derived from the Green's function based formalism of Luttinger and Ward [43] and Baym and Kadanoff [39], where the grand potential

$$\Omega[\mathbf{G}] = \text{Tr} \ln[-\mathbf{G}] - \text{Tr} [(\mathbf{G}_0^{-1} - \mathbf{G}^{-1}) \mathbf{G}] + \Phi[\mathbf{G}, \mathbf{U}] \quad (54)$$

is expressed in terms of the single-particle Green's function $\mathbf{G}$.⁴ Here, $\Phi[\mathbf{G}]$ is the so-called Baym-Kadanoff functional described in Ref. [39]. The self-energy Σ is obtained from the functional derivative of $\Phi[\mathbf{G}]$ with respect to $\mathbf{G}$

$$\Sigma = \frac{\delta \Phi[\mathbf{G}]}{\delta \mathbf{G}} \quad (55)$$

and is related to the Green's function via the Dyson equation (8)

$$\mathbf{G}^{-1} = \mathbf{G}_0^{-1} - \Sigma. \quad (56)$$

These two relations imply that the free energy Ω is stationary with respect to $\mathbf{G}$, i.e.

$$\frac{\delta \Omega[\mathbf{G}]}{\delta \mathbf{G}} = 0. \quad (57)$$

In principle, the exact Green's function $\mathbf{G}$ and self-energy Σ can be determined from the self-consistent solution of Eqs. (55) and (56). However, since the functional $\Phi[\mathbf{G}]$ is usually unknown, an approximation is required that replaces the exact $\Phi[\mathbf{G}]$ by a known or computable functional. Conserving approximations replace the exact $\Phi[\mathbf{G}]$ by an approximate functional, which considers certain sub-classes of diagrams that can be summed up analytically and that are thought to capture the dominant physics. This generally results in a weak coupling approximation such as second-order perturbation theory or the fluctuation exchange approximation [44]. These approaches usually fail, however, when the interaction U gets larger than several times the hopping t .

DCA

The DCA therefore takes a different approach. From Eq. (12), one sees that fundamentally the DCA is an approximation of the self-energy where

$$\Sigma^{\text{DCA}}(\mathbf{k}) \simeq \sum_{\mathbf{K}} \Phi_{\mathbf{K}}(\mathbf{k}) \Sigma_c(\mathbf{K}). \quad (58)$$

To proceed, it is therefore convenient to follow the work of Potthoff [45] and Okamoto [38] and express the grand potential in terms of the self-energy instead of the Green's function

$$\Omega[\Sigma] = \text{Tr} \ln [-(\mathbf{G}_0^{-1} - \Sigma)] - (\mathcal{L}\Phi)[\Sigma]. \quad (59)$$

⁴Here we have used a matrix equation for the Green's functions $\mathbf{G}$ and $\mathbf{G}_0$ and the trace Tr sums over momenta $\mathbf{k}$ and frequencies ω_n .

Here, the functional $(\mathcal{L}\Phi)[\Sigma]$ is obtained from $\Phi[\mathbf{G}]$ through a Legendre transform

$$(\mathcal{L}\Phi)[\Sigma] = \Phi - \text{Tr}[\Sigma \mathbf{G}]. \quad (60)$$

At stationarity, i.e. $\delta\Omega[\Sigma]/\delta\Sigma = 0$, one again obtains the usual Dyson equation $[\mathbf{G}_0^{-1} - \Sigma]^{-1} = \mathbf{G}$, where $\mathbf{G} = -\delta(\mathcal{L}\Phi)[\Sigma]/\delta\Sigma$. However, just as $\Phi[\mathbf{G}]$ is unknown, the functional $(\mathcal{L}\Phi)[\Sigma]$ is generally unknown. Instead of replacing the functional by a low order weak coupling approximation, the DCA keeps the same functional but reduces the complexity by replacing the exact self-energy $\Sigma(\mathbf{k})$ by the approximation in Eq. (58), i.e. $\Sigma(\mathbf{k}) \simeq \Sigma^{DCA}(\mathbf{k}) = \sum_{\mathbf{K}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma_c(\mathbf{K})$. This replacement reduces the degrees of freedom over which the functional is evaluated to those of a finite-size cluster,

$$(\mathcal{L}\Phi)[\Sigma_c] = \Phi - \frac{N}{N_c} \sum_{\mathbf{K}} \text{Tr}[\Sigma_c(\mathbf{K}) G_c(\mathbf{K})] \quad (61)$$

with

$$G_c(\mathbf{K}) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) G(\mathbf{k}) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \frac{1}{G_0^{-1}(\mathbf{k}) - \Sigma^{DCA}(\mathbf{k})}. \quad (62)$$

Relation (62) implies that the parameters of the effective cluster problem are fixed by the requirement that the cluster Green's function G_c is equal to the coarse-grained lattice Green's function on the right-hand side. Using the relation $\delta(\mathcal{L}\Phi)[\Sigma_c]/\delta\Sigma_c = -N/N_c G_c$ together with

$\delta\Sigma^{DCA}(\mathbf{k})/\delta\Sigma_c(\mathbf{K}) = \phi_{\mathbf{K}}(\mathbf{k})$, it is straightforward to show that the relation (62) implies that the DCA approximation for the grand potential

$$\Omega^{DCA}[\Sigma_c] = \text{Tr} \ln [-(\mathbf{G}_0^{-1} - \Sigma^{DCA})] + \Phi - \sum_{\mathbf{K}} \text{Tr}[\Sigma_c(\mathbf{K}) G_c(\mathbf{K})]. \quad (63)$$

is stationary, i.e. $\delta\Omega[\Sigma_c]/\delta\Sigma_c(\mathbf{K}) = 0$.

DCA⁺

Just like in the DCA, the DCA⁺ approach replaces the self-energy in the functional $(\mathcal{L}\Phi)[\Sigma]$ by the piecewise-constant cluster approximation $\sum_{\mathbf{K}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma_c(\mathbf{K})$ to limit the degrees of freedom over which this term is calculated to those of a finite-size cluster. But in contrast to the DCA, this replacement is not made in the first $\text{Tr} \ln [-(\mathbf{G}_0^{-1} - \Sigma)]$ term of the grand-potential $\Omega[\Sigma]$. Instead, here one retains the full lattice self-energy $\Sigma(\mathbf{k})$ to give the DCA⁺ approximation

$$\Omega^{DCA^+}[\Sigma] = \text{Tr} \ln [-(\mathbf{G}_0^{-1} - \Sigma)] + \Phi - \frac{N}{N_c} \sum_{\mathbf{K}} \text{Tr}[\Sigma_c(\mathbf{K}) G_c(\mathbf{K})]. \quad (64)$$

Note that in contrast to the DCA where the grand potential reduces to a functional of the cluster self-energy, i.e. $\Omega[\Sigma] \rightarrow \Omega^{DCA}[\Sigma_c]$, here, the grand potential remains a functional of the lattice self-energy $\Sigma(\mathbf{k})$, which enters in the first term. Then again, one can ask for stationarity of the grand potential. When the cluster self-energy Σ_c is related to the lattice self-energy Σ through the DCA⁺ constraint

$$\Sigma_c(\mathbf{K}) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \Sigma(\mathbf{k}), \quad (65)$$

one obtains from $\delta\Omega^{DCA^+}[\Sigma(\mathbf{k})]/\delta\Sigma(\mathbf{k}) = 0$

$$[G_0^{-1}(\mathbf{k}) - \Sigma(\mathbf{k})]^{-1} = \sum_{\mathbf{K}} \phi_{\mathbf{K}}(\mathbf{k}) G_c(\mathbf{K}). \quad (66)$$

To derive the right-hand side we have used $\delta\Sigma_c(\mathbf{K})/\delta\Sigma(\mathbf{k}) = N_c/N \phi_{\mathbf{K}}(\mathbf{k})$ according to Eq. (65). Using the identity Eq. (36) and multiplying both sides with $N_c/N \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k})$ results in the DCA⁺ coarse-graining equation Eq. (38), i.e.

$$G_c(\mathbf{K}) = \frac{N_c}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) \frac{1}{G_0^{-1}(\mathbf{k}) - \Sigma(\mathbf{k})}. \quad (67)$$

This proves that the DCA⁺ algorithm, just like the DCA algorithm gives results for the self-energy and the Green's function that correspond at self-consistency to a stationary solution of the DCA⁺ grand potential.

References

- [1] T. Maier, M. Jarrell, T. Pruschke, and M. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005)
- [2] M.H. Hettler, A.N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H.R. Krishnamurthy, *Phys. Rev. B* **58**, R7475 (1998)
- [3] M. Hettler, M. Mukherjee, M. Jarrell, and H. Krishnamurthy, *Phys. Rev. B* **61**, 12739 (2000)
- [4] T. Maier, M. Jarrell, T. Pruschke, and J. Keller, *Eur. Phys. J. B* **13**, 613 (2000)
- [5] P. Staar, T. Maier, and T.C. Schulthess, *Phys. Rev. B* **88**, 115101 (2013)
- [6] P. Staar, T. Maier, and T.C. Schulthess, *Phys. Rev. B* **89**, 195133 (2014)
- [7] J. Hubbard, *Proc. Royal. Soc. London* **276**, 238 (1963)
- [8] R.R. dos Santos, *Braz. J. Phys.* **33**, 36 (2003)
- [9] E. Dagotto, *Rev. Mod. Phys.* **66**, 763 (1994)
- [10] J. Hirsch and R. Fye, *Phys. Rev. Lett.* **56**, 2521 (1986)
- [11] M. Jarrell, T. Maier, C. Huscroft, and S. Moukouri, *Phys. Rev. B* **64**, 195130 (2001)
- [12] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, *Rev. Mod. Phys.* **83**, 349 (2011)
- [13] E. Gull, P. Werner, O. Parcollet, and M. Troyer, *Europhys. Lett.* **82**, 57003 (2008)
- [14] S. Rombouts, K. Heyde, and N. Jachowicz, *Phys. Rev. Lett.* **82**, 4155 (1999)
- [15] E. Gull, P. Staar, S. Fuchs, P. Nukala, M. Summers, T. Pruschke, T.C. Schulthess, and T.A. Maier, *Phys. Rev. B* **83**, 075122 (2011)
- [16] P. Staar, T.A. Maier, and T.C. Schulthess, *J. Phys.: Conf. Series* **402**, 012015 (2012)
- [17] A. Georges, W. Krauth, and M. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [18] S. Moukouri and M. Jarrell, *Phys. Rev. Lett.* **87**, 167010 (2001)
- [19] M. Troyer and U. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005)
- [20] N. Bulut, D.J. Scalapino, and S.R. White, *Phys. Rev. B* **47**, 14599 (1993)
- [21] T.A. Maier, M.S. Jarrell, and D.J. Scalapino, *Phys. Rev. Lett.* **96**, 047005 (2006)
- [22] T.A. Maier, M. Jarrell, and D. Scalapino, *Phys. Rev. B* **74**, 094513 (2006)

- [23] T.D. Stanescu and G. Kotliar, Phys. Rev. B **74**, 125110 (2006)
- [24] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012)
- [25] W.H. Richardson, J. Opt. Soc. Am. **62**, 55 (1972)
- [26] L.B. Lucy, Astron. J. **79**, 745 (1974)
- [27] R. Keys, IEEE Trans. Acoust. Speech **29**, 1153 (1981)
- [28] E. Gull, M. Ferrero, O. Parcollet, A. Georges, and A. Millis, Phys. Rev. B **82**, 155101 (2010)
- [29] E. Gull, O. Parcollet, and A.J. Millis, Phys. Rev. Lett. **110**, 216405 (2013)
- [30] N. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966)
- [31] T.A. Maier, M. Jarrell, T.C. Schulthess, P.R.C. Kent, and J.B. White, Phys. Rev. Lett. **95**, 237001 (2005)
- [32] T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999)
- [33] M. Jarrell, T. Maier, M.H. Hettler, and A. Tahvildarzadeh, Europhys. Lett. **56**, 563 (2001)
- [34] C. Huscroft, M. Jarrell, T. Maier, S. Moukouri, and A. Tahvildarzadeh, Phys. Rev. Lett. **86**, 139 (2001)
- [35] P.W. Anderson, Science **235**, 1196 (1987)
- [36] F. Zhang and T. Rice, Phys. Rev. B **41**, 7243 (1990)
- [37] P. Staar, M. Jiang, U. Hähner, T. Maier, and T. Schulthess, in preparation (2015)
- [38] S. Okamoto, A.J. Millis, H. Monien, and A. Fuhrmann, Phys. Rev. B **68**, 195121 (2003)
- [39] G. Baym and L. Kadanoff, Phys. Rev. **124**, 287 (1961)
- [40] A.I. Poteryaev, A.I. Lichtenstein, and G. Kotliar, Phys. Rev. Lett. **93**, 086401 (2004)
- [41] P.R.C. Kent, T. Saha-Dasgupta, O. Jepsen, O.K. Andersen, A. Macridin, T. Maier, M. Jarrell, and T.C. Schulthess, Phys. Rev. B **78**, 035132 (2008)
- [42] Y. Nomura, S. Sakai, and R. Arita, Phys. Rev. B **89**, 195146 (2014)
- [43] J. Luttinger and J. Ward, Phys. Rev. **118**, 1417 (1960)
- [44] N. Bickers, D. Scalapino, and S. White, Phys. Rev. Lett. **62**, 961 (1989)
- [45] M. Potthoff, Eur. Phys. J. B **32**, 429 (2003)

15 Electronic Structure of Perovskites: Lessons from Hybrid Functionals

Cesare Franchini

University of Vienna

Sensengasse 3/8, 1090 Vienna, Austria

Contents

1	Introduction	2
2	Hybrid functionals: overview and basic concepts	4
2.1	Density-functional theory	4
2.2	Hartree-Fock theory	6
2.3	The adiabatic connection formula: The birth of hybrid functionals	7
2.4	Hybrid functionals: Historical overview	9
2.5	Screened exchange : Hybrid functionals meet GW	11
3	Applications: Hybrid functionals for perovskites	14
3.1	The 3d perovskite data set	15
3.2	4d and 5d perovskites	20
3.3	Metal to insulator transition: LaMnO ₃ and BaBiO ₃	23
3.4	Multiferroics	28

1 Introduction

Named in honor of the Russian mineralogist Lev Perovski after the discovery of the calcium titanium oxide mineral CaTiO_3 in the Ural Mountains by Gustav Rose in 1839, the perovskite structure identifies the wide class of compounds with general chemical formula ABX_3 . In this ideally cubic structure, A is a large cation, usually an alkaline-earth or rare-earth element, located on the corners of the lattice; B, in the center of the lattice, is a small $3d$, $4d$, or $5d$ transition metal (TM), whereas the X-site anions are normally oxygen atoms that form an octahedral environment around the B ion. The stability and structural distortions of perovskite crystal structures are often discussed in terms of Goldschmidt's tolerance factor $t = (R_A + R_X)/\sqrt{2}(R_B + R_X)$, where R_A , R_B , and R_X are the ionic radii of the A, B, and X ions, respectively [1]. $t = 1$ represents the ideal cubic condition, which appears in a few cases. Different distorted structural variants are formed for $0.75 < t < 1$, the range of stability of the perovskite structure.

This high degree of chemical and structural flexibility combined with a inhomogeneous distribution of the partially filled d states (tendency to electron localization) lead to the coexistence of several physical interactions (spin, charge, lattice, and orbital), which are all simultaneously active and give rise to a wide array of physical properties and functionalities: colossal magnetoresistance (manganites) [2, 3], (multi)ferroelectricity (BaTiO_3 , BiFeO_3) [4], superconductivity (cuprates) [5], metal-insulator transition (LaMnO_3) [6], ferromagnetism (SrRuO_3) [7], band gaps spanning the visible and ultraviolet [8], surface chemical reactivity from active to inert [9], etc. This rich array of behaviors uniquely suits perovskites for novel solutions in different fields of application, including optoelectronics, spintronics, (photo)catalysis, and piezoelectric devices. Moreover, the last few years have witnessed the rapid emergence of a new class of solar cells based on mixed organic-inorganic halide perovskites ($\text{CH}_3\text{NH}_3\text{PbX}_3$, $\text{X}=\text{Cl}, \text{Br}, \text{I}$ and related compounds) that are revolutionizing the field of solar technologies.

The study of the strong competition and interplay among the various orbital, structural, and spin orderings is a central issue for the understanding of the physics and chemistry of perovskites and poses great challenges at both the technological and fundamental level. The theoretical description of perovskites requires methods at the frontier of materials modeling, capable of describing the complex entanglement of the distinct interactions and to provide an accurate account of basic properties such as structural distortions, band gap, and magnetic moment.

From an methodological perspective, the theoretical study of so-called strongly correlated materials and specifically perovskites has been mainly conducted within two historically distinct electronic structure communities: model Hamiltonians and *first principles*. Spin-orbital effective Hamiltonians are based on a simplified lattice-fermion models such as the Hubbard model [10], in which the many-body problem is reduced to a small number of relevant bands and short-ranged electron interactions. On the other hand, in first principles schemes the intractable Schrödinger equation involving many interacting electrons is mapped onto a simplified problem that includes an approximate treatment of the exchange and correlation (XC) electron interaction. Prominent examples of this class of schemes are the density-functional theory [11, 12] (DFT) and the Hartree-Fock theory [13, 14] (HF). In the density-functional theory [11, 12] the

full many-body problem is mapped onto a system of non-interacting electrons moving in an effective Kohn-Sham potential. In DFT, XC effects are accounted for by approximated XC functionals such as the local-density approximation (LDA) and the generalized gradient approximation (GGA) [15]. The HF method replaces the many-body Schrödinger equation by a set of one-particle equations, with each electron moving in an effective (Fock) mean field. The electronic many-body wavefunction is written in terms of one particle spin-orbitals arranged in a Slater determinant [16]. This Slater picture guarantees an exact account of the exchange term but completely neglects correlation effects.

DFT and HF are considered to be first principles methods (or likewise *ab-initio*) because they rely on basic and fundamental quantum-mechanical laws without invoking additional assumptions or adjustable parameters. Both types of approaches, first principles and many-body model Hamiltonians, have specific assets but also limitations. On the one hand, DFT takes into account the realistic electronic structure, but standard XC functionals are generally unable to correctly describe the localized picture typical of strongly correlated compounds. In this sense, the most pressing aspect of DFT is the construction of improved XC functionals able to overcome this drawback. Many different approaches have been proposed for reducing these inaccuracies: the weighted-density approximation [17], the DFT+U approximation [18], the self-interaction correction method [19], the screened-exchange approximation [20], the optimized effective potential [21], various meta-GGA potentials [22], and hybrid functionals [23]. Alternative and more sophisticated routes to go beyond DFT/HF are the GW approximation, in which the self energy of a many-body system of electrons is computed explicitly by making use of the single particle Green function G and the screened Coulomb interaction W [24], and post Hartree-Fock methods (Møller-Plesset perturbation theory [25], configuration interaction [26], and coupled cluster [27]). On the other hand, model Hamiltonian approaches solve the many-body problem more accurately but do not deal with the realistic band structure, some aspect of the electronic correlation are not included (like two-particle correlation and non-local correlation) and the results are dependent on a set of adjustable parameters. In this respect, dynamical mean field theory (DMFT) [28] and its extensions have marked a methodological breakthrough in this field. With the aim of constructing many-body model Hamiltonians entirely from first principles and taking advantage of the specific assets of both types of approaches, in the past decade much effort has been made to combine DFT with many-body model Hamiltonians, an example of which is DFT+DMFT [29, 30].

In this abundant cluster of methods there no best choice. The decision to adopt one method or the other should be guided by the specific intent of the research and by the computational resources at disposal. If the aim is to achieve a realistic and predictive account of the coupling between structural distortions and electronic/spin degrees of freedom in an extended, strongly correlated system like perovskites, beyond-LDA/GGA approaches appears to be a practical and convenient solution. The reason being that these methods allow for computationally efficient calculations of the forces acting on the ions and simultaneously furnish a generally adequate (self-consistent) account of electron localization, spin/orbital ordering, and total energies. The unfavorable aspect is that these approaches give only an approximate description of electronic

correlation, inferior to post-Hartree-Fock methods and last-generation many-body Hamiltonians and GW.

The purpose of these lecture notes is to present the essential ideas and a fairly detailed physical picture of hybrid functionals and to provide an overview of recent applications of hybrid functionals to perovskites covering several different physical scenarios, including 3d, 4d, and 5d perovskites, metal-insulator transitions (MIT), magnetic transitions and ordering temperatures, Jahn-Teller distortions, polarons, multiferroism, and surface energies. These notes are based on a series of articles dealing with the applications of hybrid functionals to perovskites [31–42] and on several methodological papers (cited along the text).

2 Hybrid functionals: overview and basic concepts

Hybrid functionals are a class of XC functionals in DFT that are constructed by a suitable mixing of LDA/GGA XC functionals with a certain portion of the exact HF exchange [23]. Before discussing the origin and framework of hybrid functionals in detail we start from a very brief recap of DFT and HF. Both methods are designed to provide a solution to the Schrödinger equation for a many-body system, defined by the Hamiltonian (in atomic units)

$$\hat{H} = \underbrace{-\frac{1}{2} \sum_{i=1}^N \nabla_i^2}_{T_e} - \underbrace{\sum_{n=1}^M \frac{1}{2M_n} \nabla_n^2}_{T_n} + \underbrace{\frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{V_{ee}} - \underbrace{\sum_{n,i}^{M;n} \frac{Z_n}{|\mathbf{r}_i - \mathbf{R}_n|}}_{V_{en}} + \underbrace{\frac{1}{2} \sum_{\substack{n,m=1 \\ n \neq m}}^M \frac{Z_n Z_m}{|\mathbf{R}_n - \mathbf{R}_m|}}_{V_{nn}} \quad (1)$$

where T_e and T_n represent the kinetic energy operators for the electrons and nuclei in the system, respectively, and V_{ee} , V_{en} , and V_{nn} are the electron-electron, electron-nuclei, and nuclei-nuclei potential energies, respectively. Within the Born-Oppenheimer approximation, the global problem can be separated into an electronic problem with fixed nuclei and a nuclear problem under an effective potential generated by the electrons. In compact form, the electronic Hamiltonian can be written as:

$$\hat{H}_{el} = T_e + V_{ee} + V_{en}, \quad (2)$$

and the corresponding Schrödinger equation has the form

$$\hat{H}_{el} \Psi = E \Psi, \quad (3)$$

where E is the energy and Ψ the many-particle wave function.

2.1 Density-functional theory

DFT is a way to solve the Schrödinger equation for a many-body system of interacting electrons. In DFT the real system is described by an effective Kohn-Sham (KS) one-electron system in which the electron density $n(\mathbf{r})$ (which only depends on the three spatial coordinates), rather than the complex many-electron wave function Ψ plays the crucial role, and the electron-electron interactions are taken into account via approximate XC functionals. [11, 12, 43] The

effective KS potential is constructed such that the ground-state density obtained from the fictitious, noninteracting electron system is equal to the ground-state density of the given real, interacting electron system.

In the KS formulation of DFT, the density-dependent KS potential V_{KS} is the sum of the classical Hartree potential V_H , the XC potential V_{xc} , and the external potential of the nuclei V_{en} (usually labeled V_{ext}), and the KS single-particle equations read

$$\left(-\frac{\hbar^2}{2m_e} \nabla^2 + V_{ext}(\mathbf{r}) + V_H + V_{xc}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \epsilon_i^{KS} \phi_i(\mathbf{r}), \quad (4)$$

where ϕ_i and ϵ_i^{KS} are the KS single-particle electronic orbitals and energies, respectively, with the density calculated as a sum over filled orbitals,

$$n(\mathbf{r}) = \sum_{i, \text{occupied}} |\phi_i(\mathbf{r})|^2. \quad (5)$$

According to the DFT prescriptions, the self-consistent solution of this set of one-electron Schrödinger equations should yield the correct ground-state density of the original system. In DFT the total energy can be expressed as a functional of the density

$$E = T_{KS}[n] + \int d^3r n(\mathbf{r}) V_{ext}(\mathbf{r}) + \frac{1}{2} \iint d^3r d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n]. \quad (6)$$

The quality of the solutions depends on the form of the XC potential V_{xc} that is defined to include everything else omitted from the first three terms in Eq. (4) and is determined from the functional derivative of the XC energy functional E_{xc} ,

$$V_{xc}([n], \mathbf{r}) = \frac{\delta E_{XC}}{\delta n(\mathbf{r})}. \quad (7)$$

Since the exact form of E_{xc} is unknown, in practice it must be approximated. The two most popular approximations are the LDA [44],

$$E_{xc}^{LDA}[n] = \int d^3r e_{xc}(n(\mathbf{r})), \quad (8)$$

where $e_{xc}(n(\mathbf{r}))$ is the known XC energy per particle for an electron gas of uniform spin densities, and the GGA [45],

$$E_{xc}^{GGA}[n] = \int d^3r f(n(\mathbf{r}), \nabla n(\mathbf{r})), \quad (9)$$

where the input function f is not unique [46].

Accurate approximations to E_{xc} are essential to obtain good predictions of materials properties. Despite the great success of DFT, its predictive power has been limited by the strong underestimation of band gaps (up to 50%, see Fig. 1). This band gap problem is to a large extent due to two major deficiencies, the presence of self interaction and the fact that the KS gaps are missing a term coming from the discontinuity of derivatives of the exchange-correlation functional [47]. The self-interaction problem [48] arises from the non complete cancellation of the spurious

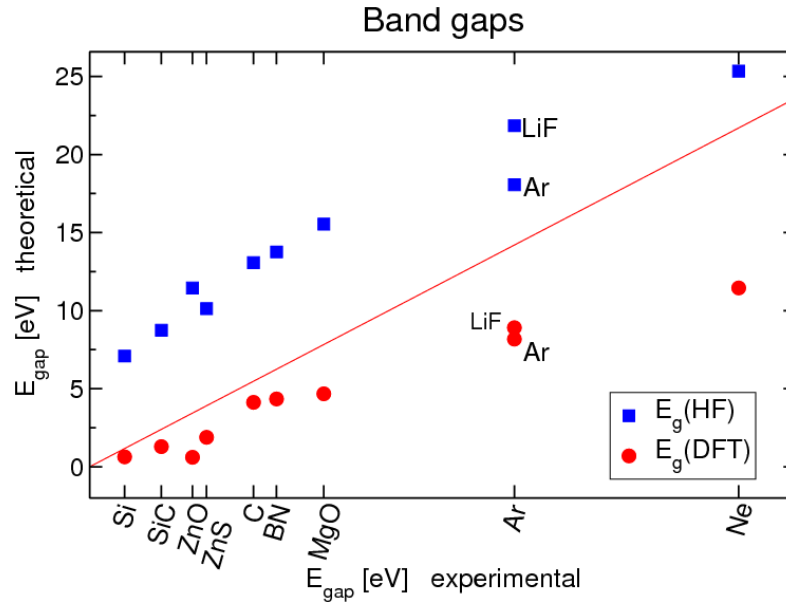


Fig. 1: DFT and HF band gaps compared to the measured values (red line).

repulsion of each electron from itself, included in the Hartree term V_H , that is not completely accounted for in XC functionals. Since this spurious Coulomb term $\Delta\epsilon$ for a given electron has the general form

$$\Delta\epsilon_1 = \iint d\mathbf{r} d\mathbf{r}' \frac{n_1(\mathbf{r})n_1(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (10)$$

the self-interaction error is larger for localized states, which are typical of strongly-correlated systems. The second drawback, the discontinuity problem, is caused by the absence of any derivative discontinuity in the LDA and GGA functionals at integer particle number, which should be present as demonstrated by Perdew and coworkers [49].

2.2 Hartree-Fock theory

The Hartree-Fock method is a variational, wavefunction-based approach. Unlike DFT, the many-body Schrödinger equation is solved with respect to the wavefunction, which takes the form of a Slater determinant

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \dots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \dots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \dots & \phi_N(\mathbf{r}_N) \end{vmatrix}, \quad (11)$$

where the ϕ are single-particle spin orbitals. Thus, similarly to DFT, the HF method follows a single-particle approach, i.e., the electrons are considered as occupying single-particle orbitals making up the wavefunction. With this type of wave function a mean-field approximation is implied: Each electron feels the presence of the other electrons indirectly through an effective

potential. The Slater determinant is fully antisymmetric: Exchanging any two particle's coordinates is equivalent to interchanging the corresponding rows of the Slater determinant. As a result, exchanging any two particles' coordinates changes the sign of the wavefunction. This assures that the exchange term is taken into account exactly. The Slater determinant vanishes when two or more orbitals are the same, i.e., the spurious self-interaction in the Hartree term is canceled. However, due to the mean-field approximation, electron correlation effects are completely neglected.

The spin-orbitals are determined by minimizing the ground-state energy using the variational principle with this Slater determinant

$$E_{HF} = \frac{\langle \Psi | \hat{H}_e | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (12)$$

and finally one arrives to the HF equations

$$\left(-\frac{\hbar^2}{2m_e} \nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) \right) \phi_i(\mathbf{r}) + \int d^3r' V_x(\mathbf{r}, \mathbf{r}') \phi_i(\mathbf{r}') = \epsilon_i^{HF} \phi_i(\mathbf{r}). \quad (13)$$

This set of equations has strong similarities with the DFT KS equations Eq. (4). Apart from the specific form of the single-particle orbitals ϕ , the only fundamental difference is the way XC effects (the last term on the left side) are treated: In DFT both exchange and correlation contributions are taken into account in an approximate way through the XC potential V_{xc} , whereas in HF theory correlation effects are completely neglected (mean-field approximation) but the exchange part is exact (through the Slater determinant), without invoking any type of approximation. Specifically, the exchange is given by

$$V_x(\mathbf{r}, \mathbf{r}') = -\frac{\sum_j f_j \phi_j(\mathbf{r}) \phi_j^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (14)$$

where f_j are the occupation weights of the orbitals ϕ_j . This is not a simple term since it is nonlocal, orbital dependent, and requires knowledge of all (occupied) orbitals.

Going back to Fig. 1, where DFT approximated functionals lead to small band gaps, the lack of correlation in HF yields to a serious overestimation of band gaps. An uneducated and rough interpretation of Fig. 1 might suggest that the inclusion of a scrap of HF in DFT might eventually result in a better estimation of band gaps. Nonsense? Not really, but the route to such envisioned *hybrid functionals* should start from elsewhere.

2.3 The adiabatic connection formula: The birth of hybrid functionals

The starting point for constructing the bridge between DFT and HF is the adiabatic connection. In DFT, the adiabatic connection continuously transforms the non-interacting system to the physical interacting one:

$$E_{XC} = \int_0^1 E_{xc,\lambda} d\lambda, \quad E_{xc,\lambda} = \langle \Psi_\lambda | V_{ee} | \Psi_\lambda \rangle - E_H, \quad (15)$$

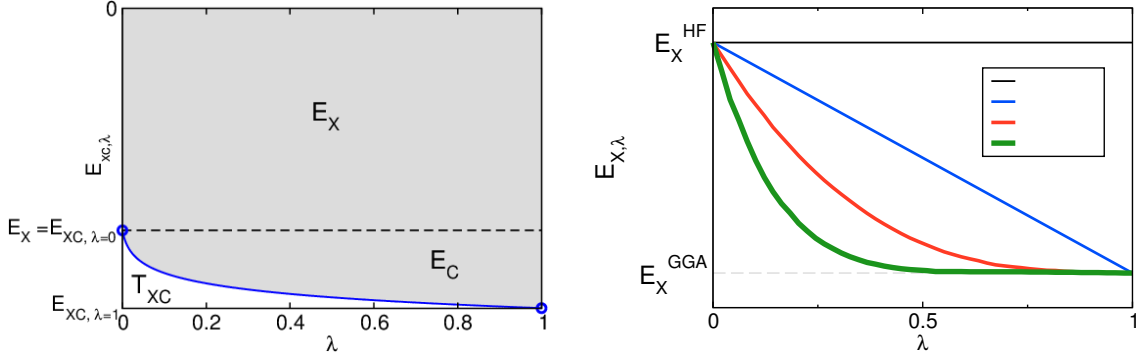


Fig. 2: (left) A cartoon of the integrand in the adiabatic connection formula. The Kohn-Sham exchange-correlation energy $E_{XC,\lambda}$ (shadow area) as a function of the coupling constant λ . The extreme values at $\lambda=0$ and $\lambda=1$ (indicated by the circles) correspond to the non-interacting (exchange only, E_X) and truly interacting system (exchange plus full correlation, $E_{XC,\lambda=1}$). (right) Sketch of the dependence of $E_{XC,\lambda}$ on the mixing parameter α in the PBE0 hybrid functional; the comparison with Møller-Plesset perturbation theory suggests $\alpha=0.25$ as “optimum” fraction [50]. Adaped from a picture in Martin Schlipf, “Beyond LDA and GGA: Hybrid Functionals.”

where λ is an intraelectronic coupling-strength parameter that tunes the electron-electron Coulomb potential V_{ee} ,

$$E_{XC} = \int_0^1 E_{xc,\lambda} d\lambda, \quad V_{ee,\lambda} = \sum_{ij} \frac{\lambda}{|\mathbf{r} - \mathbf{r}'|} \quad (16)$$

and Ψ_λ is the λ -dependent wavefunction.

This formula connects the noninteracting Kohn-Sham system ($\lambda = 0$, effective KS potential, $\Psi_{\lambda=0}$ = KS wavefunction) continuously with the fully interacting real system ($\lambda = 1$, true external potential, $\Psi_{\lambda=1}$ = exact many-body wavefunction) through intermediate partially interacting ($0 < \lambda < 1$) systems (see Fig. 2). Only $\lambda = 1$ is real or physical. The $\lambda = 0$ limit and all the intermediate values of λ , are convenient mathematical fictions.

The exchange-correlation energy E_{XC} corresponds to the shaded area in Fig. 2 and is equivalent to the full ($\lambda = 1$) many-body exchange-correlation energy minus the kinetic contribution to the XC term (T_{XC} , white area in Fig. 2). However, as already mentioned, the exact XC functional is unknown. The calculation of the complete adiabatic connection curves is prohibitive and can only be done for small reference systems using Quantum Monte Carlo or beyond-HF quantum-chemistry approaches. This is the core of DFT: XC functionals are constructed by combining known exact limiting values with accurate reference energies.

But what we want here is to take advantage of the fact that the exact exchange E_X ($\lambda = 0$) can be calculated from the KS orbitals, Eq. (14), and that XC functionals provide an approximation for the $\lambda = 1$ limit. A first evident way to approximate the integral of Eq. (16) is a linear interpolation, which leads to the following approximate XC energy

$$E_{XC} \approx \frac{1}{2} E_{XC,\lambda=0} + \frac{1}{2} E_{XC,\lambda=1} \quad (17)$$

that forms the basis for the half-half hybrid functional proposed by Becke [51]

$$E_{XC}^{\text{Hybrid}} = \frac{1}{2}E_X^{\text{HF}} + \frac{1}{2}E_{XC}^{\text{LDA}}. \quad (18)$$

Approaches like hybrid functionals (but also DFT+U) involving nonlocal exchange, go beyond standard KS DFT and are usually referred to as generalized KS approaches [52, 53]. It should be mentioned already at this stage that the nonlocality of the exchange potential increases the computational load substantially due to a double summation over the Brillouin zone. In plane-wave codes the computing time increases by a factor of 100 to 1000 for multiple $\mathbf{k}$ points.

2.4 Hybrid functionals: Historical overview

After proposing the half-half hybrid, Becke introduced a parametric hybrid functional including exact exchange and local (LDA) and gradient-corrected (GGA) exchange and correlation that has become very popular in the quantum chemistry community with the abbreviation B3LYP [54, 55]. The B3LYP depends on three parameters. It incorporates only 20% of the exact HF exchange and has the form

$$E_{XC}^{\text{B3LYP}} = E_{XC}^{\text{LDA}} + \alpha_1(E_X^{\text{HF}} - E_X^{\text{LDA}}) + \alpha_2(E_X^{\text{GGA}} - E_X^{\text{LDA}}) + \alpha_3(E_C^{\text{GGA}} - E_C^{\text{LDA}}), \quad (19)$$

where the three mixing parameters $\alpha_1 = 0.2$, $\alpha_2 = 0.72$, and $\alpha_3 = 0.81$ are determined by fitting experimental atomization energies, electron and proton affinities, and ionization potentials of the molecules in Pople's G1 data set. The B3LYP has been intensively and successfully adopted for atomic and molecular calculation, but its application to periodic systems is not equally satisfactory, because the B3LYP functional does not reproduce the correct exchange-correlation energy for the free-electron gas. This is particularly problematic for metals and heavier elements beyond the 3d transition-metal series [56].

A more appropriate hybrid functional for solid-state applications is the PBEh proposed by Perdew, Burke, and Ernzerhof [57] (also referred to as PBE0) [58], which reproduces the homogeneous electron gas limit and significantly outperforms B3LYP in solids, especially in the case of systems with itinerant character (metals and small-gap semiconductors) [56].

The starting point for the construction of the PBE0 functional is the application of the mean value theorem to the monotonically decreasing integrand $E_{XC,\lambda}$ in the adiabatic connection formula, Eq. (15). One obtains an approximate E_{XC} that depends on one parameter only, α :

$$E_{XC} \approx \alpha E_{XC,\lambda=0} + (1 - \alpha) E_{XC,\lambda=1}, \quad \alpha \in [0, 1]. \quad (20)$$

The final form of the PBE0 functional is obtained by replacing $E_{XC,\lambda=1}$ with the GGA-type functional of Perdew, Burke, and Ernzerhof (PBE):

$$E_{XC}^{\text{PBE0}} = E_{XC}^{\text{PBE}} + \alpha(E_X^{\text{HF}} - E_X^{\text{PBE}}). \quad (21)$$

By analyzing the dependence of $E_{X,\lambda}$ on α and by a direct comparison with Møller-Plesset perturbation-theory reference energies, Perdew, Ernzerhof and Burke found that the choice

$\alpha=0.25$ yields the best atomization energies of typical molecules [50]. Although this 0.25 choice has become a standard in PBE0 calculations, the same authors have warned that an ideal hybrid should have an optimum α for each system and property. They arrived to this conclusion by treating α as an index of correlation strength, and showing that different adiabatic connection curves would be obtained for different values of α , as exemplified in Fig. 2. The important message is that $\alpha=0.25$ is the best compromise, but a system-specific and property-specific α would better mimic the real system. Various attempts to propose a system-dependent α will be discussed later.

The PBE0 method has been successfully applied to solids [59,60], but its widespread application was hindered by the numerical complications in calculating the slow-decaying long-range (lr) part of the exchange integrals and exchange potential. This is particularly difficult for metals, where a dense $\mathbf{k}$ -point sampling is needed, leading to very slow convergence [56].

To solve this issue, Heyd, Scuseria, and Ernzerhof proposed to replace the lr-exchange by the corresponding density-functional counterpart [61, 62], i.e., by replacing the exact exchange V_x by a screened version V_{sx}

$$V_{sx}(\mathbf{r}, \mathbf{r}') = - \sum_j f_j \phi_j(\mathbf{r}) \phi_j^*(\mathbf{r}') \frac{\text{erfc}(\mu|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|}, \quad (22)$$

where μ is the critical inverse screening length. In the resulting screened hybrid functional, usually referred to as HSE06, only the sort-range (sr) exchange is treated at HF level

$$E_{XC}^{\text{HSE}} = \alpha E_X^{\text{HF},sr}(\mu) + (1-\alpha) E_X^{\text{PBE},sr}(\mu) + E_X^{\text{PBE},lr}(\mu) + E_C^{\text{PBE}}, \quad \alpha \in [0, 1], \quad \mu \in [0, \infty]. \quad (23)$$

The allowed range of variation of μ is $0 \leq \mu \leq \infty$, but in practical calculations its value is usually chosen in a relatively narrow (and physically relevant) range [63]. Based on molecular tests the value of μ was set to $0.2 \text{ \AA}^{-1}$ (corresponding to a screening length $r_s = 2/\mu = 10 \text{ \AA}$), which is routinely considered as the standard choice for HSE calculations [62].

As the intent of HSE was to achieve accuracy equivalent to PBE0 at a reduced computational effort, the Fock-exchange fraction was initially routinely set to its PBE0 value of $\alpha=0.25$, but it has been shown that the value α that lead to the best agreement with experimental band gaps is material-specific and often deviates significantly from 0.25, for both PBE0 and HSE [32, 63–65]. A few comprehensive reviews of applications of the HSE functional are available in literature [31, 66, 67].

The decomposition of the Coulomb kernel into a lr and sr part in the HSE functional is obtained by splitting the Coulomb operator by way of the error function (erf)

$$\frac{1}{r} = \underbrace{\frac{\text{erfc}(\mu r)}{r}}_{\text{sr}} + \underbrace{\frac{\text{erf}(\mu r)}{r}}_{\text{lr}}, \quad (24)$$

where $r = |\mathbf{r} - \mathbf{r}'|$, erfc is the complementary error function, i.e., $\text{erfc}(\mu r) = 1 - \text{erf}(\mu r)$, and μ is the screening parameter that controls the range separation. The inverse of the screening parameter, μ^{-1} , represents the critical distance at which the short-range Coulomb interactions

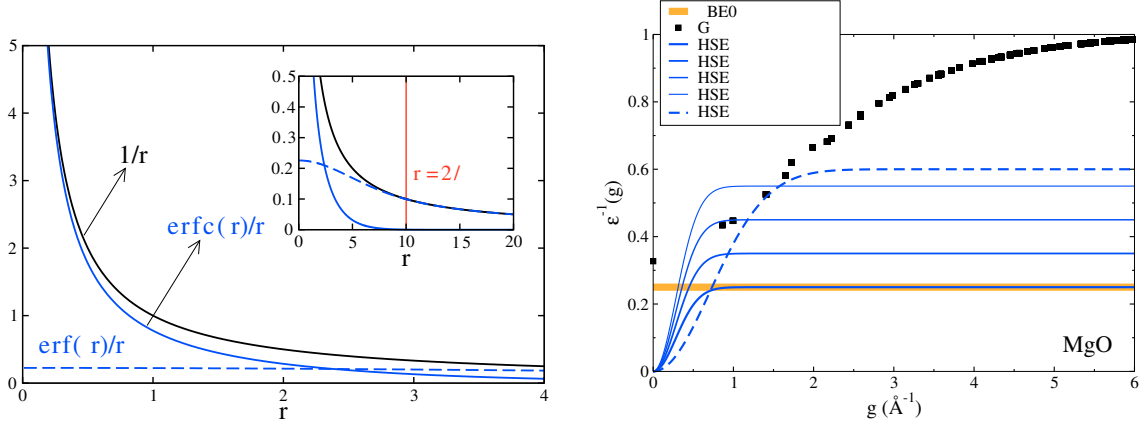


Fig. 3: (left) Plots of $1/r$, $\text{erf}(\mu r)/r$, and $\text{erfc}(\mu r)/r$ for the standard value of $\mu = 0.2$. The inset shows the three curves around the critical screening distance $r_s = 2/\mu = 10 \text{ \AA}$, at which the short-range Coulomb interactions become negligible. (right) Diagonal part of the electronic contribution to the inverse of the macroscopic dielectric function vs. wave vector g computed at the GW level for MgO compared with the HSE screening given for different values of the mixing parameter α at the fixed screening length $\mu = 0.2 \text{ \AA}^{-1}$, as well as the parametrization proposed by Kresse, $\mu = 0.5 \text{ \AA}^{-1}$, $\alpha = 0.6$ [68]. The thick straight line is the PBE0 curve.

can be assumed to be negligible. This is shown graphically in Fig. 3 (left panel). The HSE functional reduces to PBE0 for $\mu = 0$, and to PBE for $\alpha = 0$ or $\mu \rightarrow \infty$. Moussa *et al.* have shown that the PBE limit is indeed achieved for a much smaller value of μ of about 1 [63]. In the next section we will discuss a few important aspects of this screened exchange in more detail.

2.5 Screened exchange : Hybrid functionals meet GW

Besides the computational convenience there is a formal justification for the incorporation of a certain amount of screening in the hybrid functional formalism, based on the consideration that in multi-electron systems the unscreened HF exchange is effectively reduced by the presence of the other electrons in the system. The concept of screening immediately links hybrid functionals to the GW approximation, in which the screening is a natural and fundamental ingredient [24, 31, 63, 67–70].

In the GW quasiparticle equations, written in a similar fashion to the DFT (Eq. (4)) and HF (Eq. (13)) equations,

$$\left(-\frac{\hbar^2}{2m_e} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) \right) \phi(\mathbf{r}) + \int d\mathbf{r}' \Sigma(\mathbf{r}, \mathbf{r}', \omega) \phi(\mathbf{r}') = \epsilon_i^{\text{GW}} \phi_i(\mathbf{r}), \quad (25)$$

the self-energy Σ is calculated explicitly in terms of the single-particle Green function G and the screened Coulomb interaction W

$$\Sigma(\omega) = iG(\omega) W(\omega), \quad (26)$$

and the screened Coulomb interaction is given by

$$W(\omega) = \epsilon(\omega)^{-1} v, \quad (27)$$

where v is the bare Coulomb interaction $1/|\mathbf{r} - \mathbf{r}'|$ and ϵ is the frequency-dependent dielectric function. By replacing the screening in W by an effective static ($\omega = 0$) dielectric constant $\epsilon_\infty = 1/\alpha$, we obtain the effective W of HSE

$$W^{HSE} = \alpha \frac{\text{erfc}(\mu|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|}. \quad (28)$$

Inspired by the discussion in Ref. [68] we illustrate this connection by comparing in Fig. 3 the diagonal part of the electronic contributions to the static dielectric function $\epsilon^{-1}(g, \omega = 0)$ as a function of the reciprocal lattice vector g ($|\mathbf{g}|$) at the GW and HSE levels for the wide-bandgap material MgO. The GW calculation are done using the Vienna Ab Initio Simulation Package (VASP) [71, 72]. In reciprocal space the relation between $\epsilon^{-1}(g, \omega = 0)$ and the screened potential at zero frequency W is

$$W^{static}(\mathbf{g}, \mathbf{g}) = \frac{4\pi}{|\mathbf{g}|^2} \epsilon^{-1}(|\mathbf{g}|, \omega = 0), \quad (29)$$

where $4\pi/|\mathbf{g}|^2$ is the Coulomb kernel and $\epsilon^{-1}(g)$ is the actual screening. At large g , $\epsilon^{-1}(g)$ approaches 1, meaning that nonlocal exchange is not screened and a purely HF picture is recovered; for smaller g the dielectric function is progressively reduced indicating that a continuously larger portion of the nonlocal exchange is screened by the other electrons. The value of $\epsilon^{-1}(g)$ at $g = 0$ (usually referred to as the ion-clamped dielectric constant ϵ_∞) should be compared with the inverse of the measured dielectric constant that is about 0.3 for MgO.

In HSE the model screening in reciprocal space (also plotted in Fig. 3) is given by the following expression

$$\epsilon^{-1}(g) = \frac{1}{\alpha} \left(1 - e^{g^2/4\mu^2} \right). \quad (30)$$

From Fig. 3 it is evident that HSE does not appear as an appropriate approximation. The first problem is that at very short wave vector HSE does not include any exact exchange, thus mimicking an infinite dielectric constant ϵ_∞ . This is good for metals but fundamentally not correct for insulators and in the vacuum. However, it has been shown that the removal of the long-range part of the exact exchange does not severely affect the prediction of band gaps for semiconductors and insulators [68]. Rather, the main problem of HSE is the average amount of nonlocal exchange. It is clear that HSE with only 1/4 of the exact exchange is not capable of describing wide-bandgap materials such as MgO that have a generally small dielectric constant. In fact, Fig. 3 shows that at any reciprocal lattice vector g (i.e. for any wavelength), the standard 0.25 HSE screening is always smaller than the MgO $\epsilon^{-1}(g)$, indicating that the amount of nonlocal exchange always exceeds 0.25. This problem could be solved by increasing α to a value at least equal to the inverse dielectric constant ϵ_∞ , which represents the minimum of $\epsilon^{-1}(g)$ (found at $g = 0$). Kresse has proposed an alternative parametrization ($\mu=0.5$, $\alpha=0.6$) that improves the band gaps from small to wide-band-gaps materials as compared to standard HSE [68]. The shape of $\epsilon^{-1}(g)$ corresponding to this parametrization is also included in Fig. 3: It incorporates a larger amount of nonlocal exchange but the slope at small g is sizably lower. This parametrization is efficient for bandgap predictions but properties such as formation energies are worse than

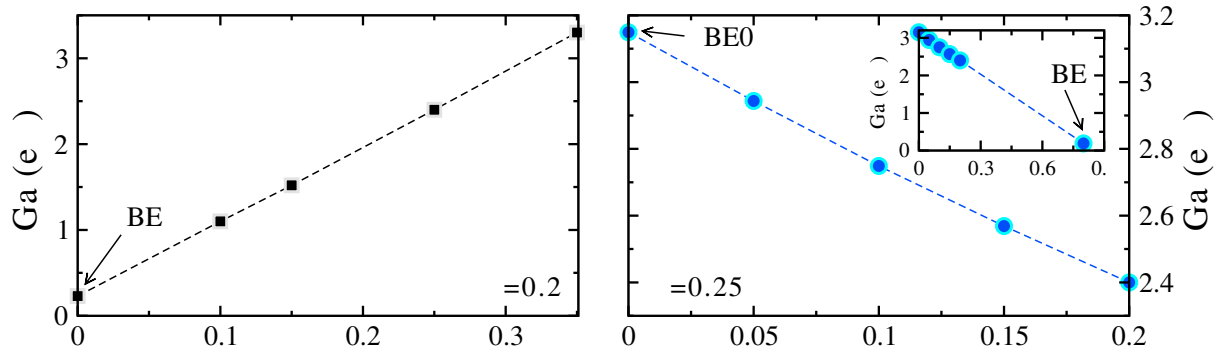


Fig. 4: Effect of the HSE parameters α and μ on the band gap of LaMnO_3 . (left) α dependence of the band gap with μ fixed to its standard value, 0.208. The PBE result for $\alpha=0$ is also indicated. (right) μ dependence of the band gap with α fixed to its standard value, 0.25. The PBE0 ($\mu=0$) and PBE ($\mu \rightarrow \infty$, inset) results are indicated by arrows.

those obtained by conventional HSE, supporting the conclusions of Perdew, Ernzerhof in their seminal paper [50].

The bottom line of the above discussion is that the link between GW and HSE provides a rough instruction on how to set an optimal α parameter through the relation

$$\alpha_{opt} \approx \frac{1}{\epsilon_{\infty}}. \quad (31)$$

The ϵ_{∞} dependent α has proven to yield good band gaps [65, 73] but becomes impractical when ϵ_{∞} is unknown due to the complexity of the procedure to compute the dielectric constant. In addition, this choice is more justified for unscreened PBE0-like hybrid functionals than for screened hybrid functionals (like HSE). In screened hybrids, screening is already present to some extent in the range separation. In Ref. [32] the effect of the HSE screening on α is quantified in a downward shift of about 0.07.

Several protocols have been proposed to obtain system-dependent parameters, either based on the similarity between hybrid functionals and quasiparticle theory [32, 65, 73–77] (these schemes rely on the $\alpha \approx 1/\epsilon_{\infty}$ assumption, Eq. (31)) or by means of a systematic analysis of the HSE parameter space [63] (for the SC/40 semiconductor set the best accuracy is achieved for $\alpha=0.313$ and $\mu = 0.185 \text{ \AA}^{-1}$ [63]).

As a practical illustration of the role of α and μ , Fig. 4 illustrates the dependence of the band gap as a function of α and μ for one of the most studied strongly correlated perovskites, LaMnO_3 . By increasing α for a fixed value of μ , we include more nonlocal exchange and reduce the screening; as expected, the band gap increases linearly, approaching a HF-like description. On the contrary, the increase of μ at fixed α reduces the already small nonlocal exchange at short wave vector g (large distances) thus inducing a progressive reduction of the band gap.

As a final note before concluding this section, it should be emphasized that, although hybrid functionals are a step forward with respect to standard local and semilocal functionals as they incorporate a portion of nonlocal exchange, the treatment of correlation effects is not improved

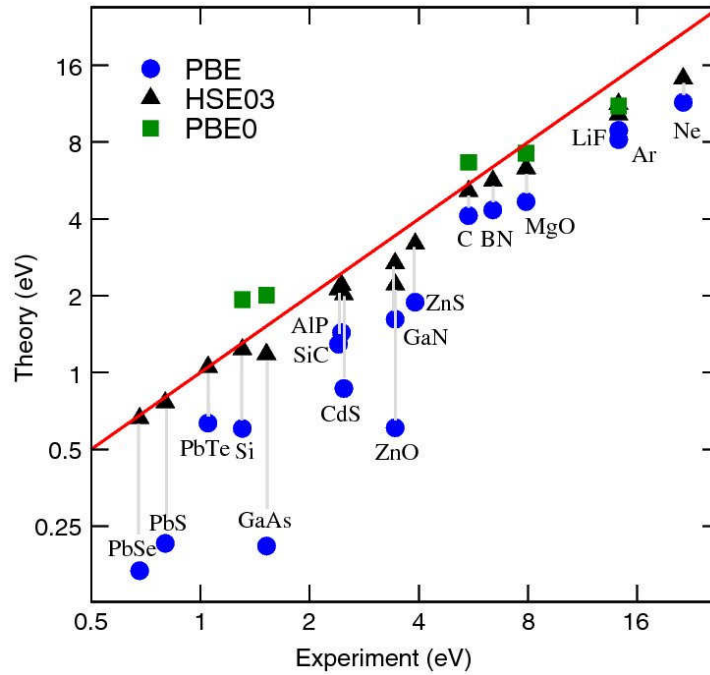


Fig. 5: Comparison between computed (PBE, PBE0 and HSE) and experimental band gaps for a representative data set of insulators and semiconductors. Taken from Ref. [78].

with respect to conventional functionals. The fractional nonlocal exchange, however, is sufficient to cure the LDA/GGA underestimation of the derivative discontinuities at integer numbers, which is ultimately one of the core arguments in favor of hybrid functionals. Fig. 5 manifests the improved estimation of band gaps obtained by HSE and PBE0 as compared to standard PBE.

3 Applications: Hybrid functionals for perovskites

Hybrid functionals have been increasingly and successfully employed for a wide range of solid-state problems. Initially, the applications were limited to monatomic systems or binary compounds but the upsurge of interest in transition-metal perovskite-based compounds motivated by their technological relevance and functional ductility has motivated the use of hybrid functionals in this class of materials. Most of the applications of hybrid functionals to perovskites have focused on the “classical” $3d$ subclass, in particular titanates, manganites, nickelates, ferrites, aluminates, and to a lesser extent cobaltates and cuprates (See Ref. [31] and references therein). The range of physical phenomena investigated includes band-gap prediction, orbital/charge ordering, metal-to-insulator transition (MIT), structural distortions, phonons, ordering temperatures, ferroelectricity and multiferroism, exchange integrals, dielectric properties, spin-orbit coupling effects, etc.

In this section, we will present an ample number of examples, aiming to cover a wide variety of physical effects. Most of the example will use the screened HSE method.

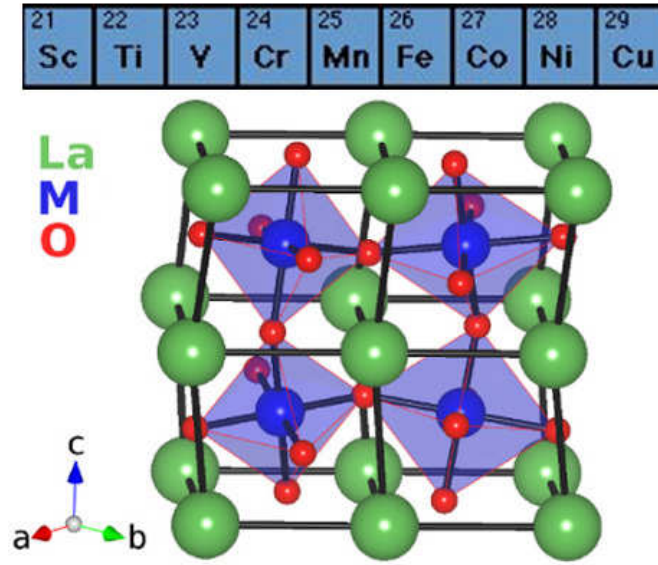


Fig. 6: The perovskite structure.

3.1 The 3d perovskite data set

We start by presenting the structural, electronic, and magnetic properties of the series of 3d transition-metal-oxide-perovskites LaMO_3 , with M ranging from Sc to Cu [32]. This data set is representative of a large variety of behavior arising from the different filling of the 3d manifold (from d^0 to d^8) and different values of the tolerance factor (from 0.91 to 1.01), as summarized in Table 1. It comprises Mott-Hubbard (MH) and charge-transfer (CT) insulators; (correlated) metals; A-type, C-type, and G-type antiferromagnets as well non-magnetic and paramagnetic systems; and crystal symmetry spanning orthorhombic, monoclinic, rhombohedral, and tetragonal symmetries with different degrees of Jahn-Teller distortions. The typical structure, general for all perovskites, is given in Fig. 6.

From a statistical analysis of the structural properties (lattice constant and internal distortions), band gaps, and magnetic properties (ordering and spin moment) computed by HSE for different values of the mixing factor α we have constructed a mean absolute relative error (MARE, with respect to experimentally available results) map (shown in Fig. 7) from which a few conclusions can be drawn:

Table 1: Summary of the fundamental ground-state properties of LaMO_3 : (i) Crystal structure: O=orthorhombic, M=monoclinic, R=rhombohedral, and T=tetragonal; (ii) 3d filling; (iii) electronic character: I=insulator and M=metal; (iv) Magnetic ordering: NM=non-magnetic, different type of AFM arrangements, and PM=paramagnetic.

LaScO_3	LaTiO_3	LaVO_3	LaCrO_3	LaMnO_3	LaFeO_3	LaCoO_3	LaNiO_3	LaCuO_3
O- P_{nma}	O- P_{nma}	M- $P_{21/b}$	O- P_{nma}	O- P_{nma}	O- P_{nma}	R- R_{3c}	R- R_{3c}	T- $P_{4/m}$
d^0	$t_{2g}^\uparrow$	$t_{2g}^{\uparrow\uparrow}$	$t_{2g}^{\uparrow\uparrow\uparrow}$	$t_{2g}^{\uparrow\uparrow\uparrow}e_g^\uparrow$	$t_{2g}^{\uparrow\uparrow\uparrow}e_g^{\uparrow\uparrow}$	$t_{2g}^{\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow}$	$t_{2g}^{\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow}e_g^\uparrow$	$t_{2g}^{\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow}e_g^{\uparrow\downarrow}$
I	I	I	I	I	I	I	M	M
NM	G-AFM	C-AFM	G-AFM	A-AFM	G-AFM	PM	PM	PM

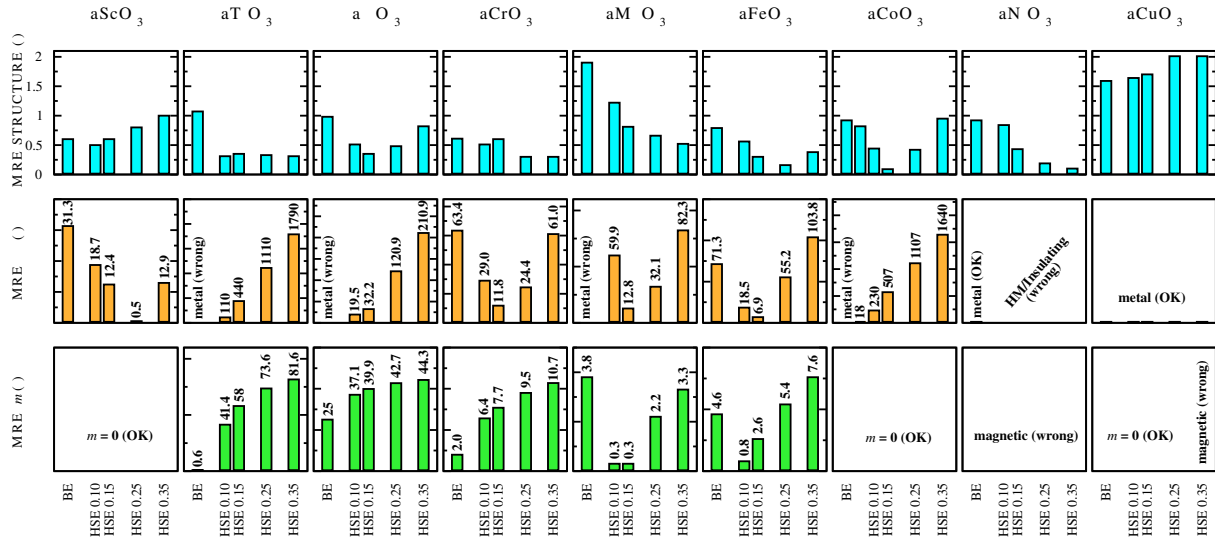


Fig. 7: Summary of the MARE for the structural properties (top), band gap Δ (middle), and magnetic moment m (bottom) at the PBE and HSE level. For the band gap Δ and the magnetic moment m , the MARE are indicated by the numbers associated with each bar. A few comments for the labels 'OK' and 'wrong': (i) LaScO_3 , m : all methods correctly predict a non-magnetic ground state; (ii) LaCoO_3 , m : all methods correctly predict zero magnetic moment; (iii) LaNiO_3 , Δ : PBE is the only approach that correctly finds a metallic solution; (iv) LaNiO_3 , m : all methods wrongly predict a magnetic ground state; (v) LaCuO_3 , Δ : all methods correctly predict a metallic solution; (vi) LaCuO_3 , m : PBE and HSE (0.05, 0.15, and 0.25) correctly predict zero magnetic moment, whereas HSE 0.35 wrongly stabilizes a magnetic ground state.

1. The results clearly depend on the value of α , but as a general trend HSE is capable of capturing the correct electronic and magnetic ground state for all insulating compounds ($M=\text{Sc to Co}$), thereby improving the deficient DFT-based predictions (but DFT delivers a better value of the magnetic moments for $M=\text{Ti, V, and Cr}$).
2. For the structural properties, on the other hand, PBE performs rather well, delivering optimized geometry within 1%. So in this sense PBE would not need any adjustment. The only important exception are the JT parameters in LaMnO_3 , which PBE finds 60% smaller than experiment (not shown, see [31]).
3. The complex nature of the PM correlated metals LaNiO_3 and LaCuO_3 is only marginally accounted for by PBE and rather poorly treated at the HSE level. This is mostly due to underlying dynamical correlation effects which cannot be easily treated at the DFT/HF level. For these compounds, PBE might be considered to be a good starting point for more elaborated many-body approaches such as GW.
4. From this plot it is possible to derive a set of phenomenological optimum values of α , α_{HSE} . Fig. 8 shows a nice correlation between α_{HSE} and the optimum values derived from the inverse dielectric constant relation ($1/\epsilon_\infty$) Eq. (31) using the experimental ϵ_∞ . The 0.07 shift between the two curves should be attributed to the fact that Eq. (31) holds for standard *unscreened* hybrid functionals such as PBE0, as discussed in the previous section. HSE is a range-separated screened hybrid functional which already contains a certain degree of screening (controlled by the screening factor μ).

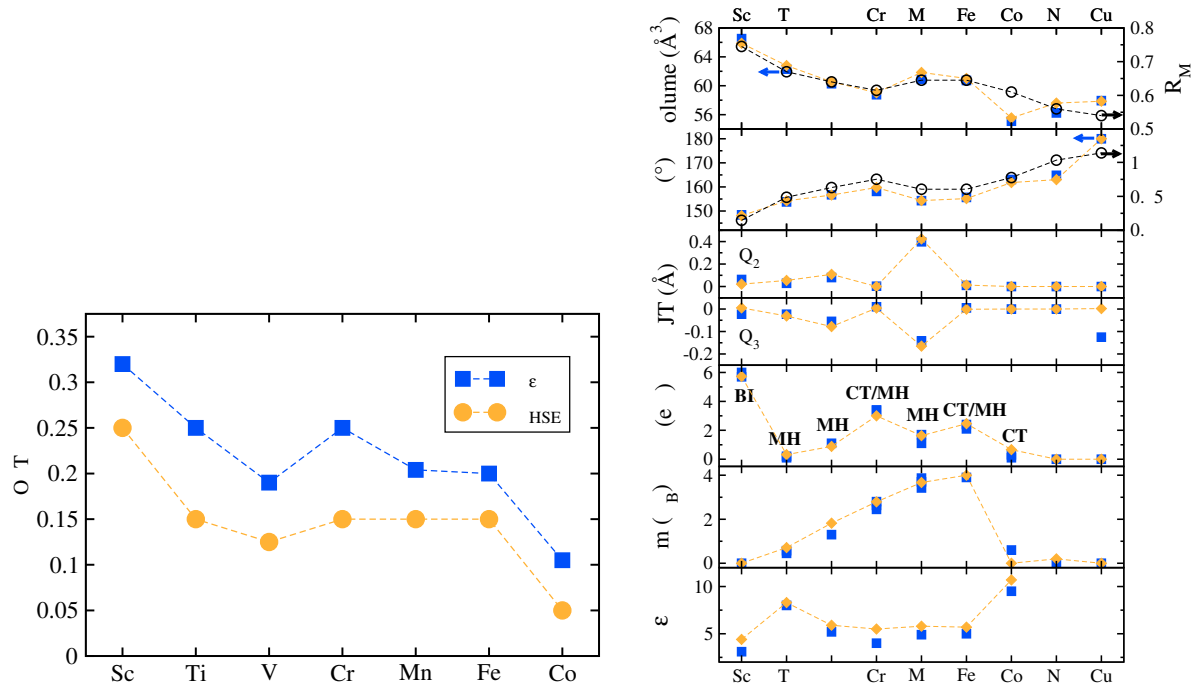


Fig. 8: (left) optimum values of α obtained from HSE, α_{HSE} , compared to those derived from the inverse dielectric constant relation ($1/\epsilon_\infty$), Eq. (31). (right) Trend of selected structural (Volume V , tilting angle θ , and JT distortions Q_2 and Q_3), electronic (band gap Δ), magnetic (magnetic moment m), and dielectric (ϵ_∞) quantities along the LaMO_3 series from $M=\text{Sc}$ to $M=\text{Cu}$. We also show the trend of the tolerance factor t as well as R_M . For LaTiO_3 we used $\alpha=0.1$. The character of the insulating gap is also indicated (BI = band insulator, CT = charge transfer, MH = Mott-Hubbard, CT/MH = mixed CT and MH character).

Within this set the overall comparison with available experimental data for a large set of quantities is excellent. Figure 8 shows the remarkably good agreement between the calculated and measured values of the volume (V), tilting angle (θ), JT distortion, band gap (Δ), magnetic moment (m), and dielectric constant (ϵ_∞). From a physical point of view, the progressive reduction of the volume from Sc to Cu is clearly associated with the almost monotonic decrease of the M ionic radius R_M , whose size is determined by the competition between the size of the $4s$ shell (where extra protons are pulled in) and the additional screening due to the increasing number of $3d$ electrons. Adding protons should lead to a decreased atom size, but this effect is hindered by repulsion of the $3d$ and, to a lesser extent, $4s$ electrons. The V/R_M curves show a plateau at about half filling (Cr-Mn-Fe) indicating that for this trio of elements these two effects are essentially balanced and atom size does not change much. The volume contraction is associated with a rectification of the average $(M - \widehat{\text{O}_1} - M + M - \widehat{\text{O}_2} - M)/2$ tilting angle θ , which follows the evolution of the tolerance factor t very well. This indicates that the tolerance factor is indeed a good measure of the overall stability and degree of distortion of perovskite compounds. The variation of the magnetic moment as a function of M can be easily understood in terms of the progressive t_{2g} and e_g band filling in the high-spin compounds LaTiO_3 ($t_{2g}^\uparrow$, $m=0.51 \mu_B$), LaVO_3 ($t_{2g}^{\uparrow\uparrow}$, $m=1.3 \mu_B$), LaCrO_3 ($t_{2g}^{\uparrow\uparrow\uparrow}$, $m=2.63 \mu_B$), LaMnO_3 ($t_{2g}^{\uparrow\uparrow\uparrow}e_g^\uparrow$, $m=3.66 \mu_B$), and LaFeO_3 ($t_{2g}^{\uparrow\uparrow\uparrow}e_g^{\uparrow\uparrow}$, $m=3.9\text{--}4.6 \mu_B$).

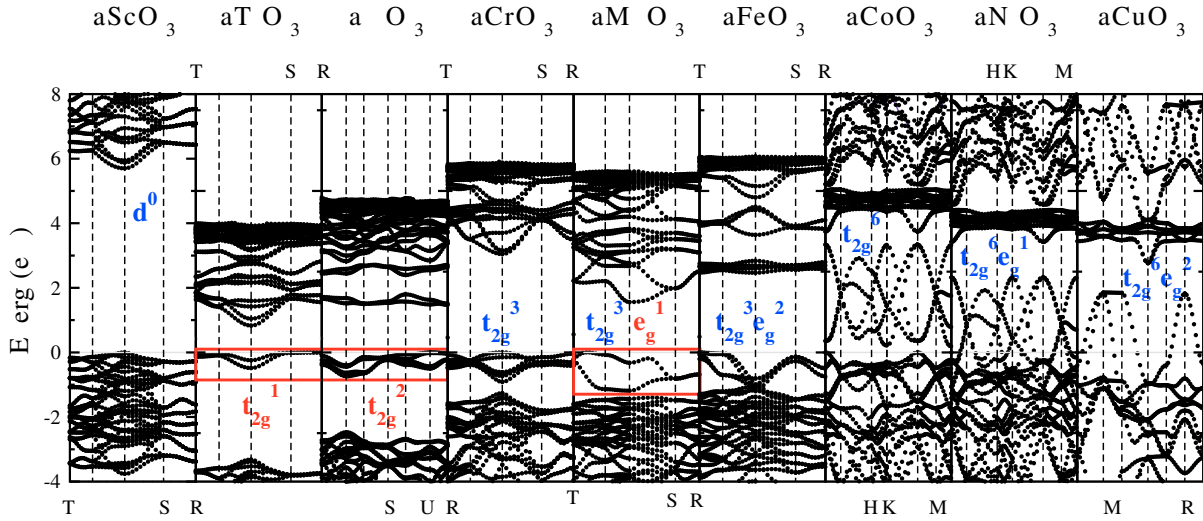


Fig. 9: Summary of the HSE electronic dispersion relations showing the complete trend from the band insulator LaScO_3 to metallic LaCuO_3 . The thick (red) lines enclose the d bands responsible for the observed orbital-ordering in LaTiO_3 (t_{2g}), LaVO_3 (t_{2g}) and LaMnO_3 (e_g).

We conclude this overview of the $3d$ perovskite data set by showing the band structures (Fig. 9) and associated orbital ordering for LaTiO_3 (t_{2g}), LaVO_3 (t_{2g}) and LaMnO_3 (e_g) (Fig. 10).

Starting from the d^0 band insulator LaScO_3 , the addition of one d electron creates a highly localized t_{2g} state right below E_F in LaTiO_3 . The gradual filling of this t_{2g} manifold leads to a continuous increase of the bandwidth from t_{2g}^1 (LaTiO_3) to t_{2g}^3 (LaCrO_3), connected with a gradual increase of the crystal-field splitting. In LaMnO_3 , the fully occupied t_{2g} band is pushed down in energy and the valence band maxima are dominated by the half-filled e_g^1 subbands. The e_g orbital gets completely filled in LaFeO_3 , which is the last member of the series having a predominantly MH gap. The inclusion of one additional electron yields a sudden change of the band structure characterized by a high increase of p - d hybridization and bandwidth around E_F , which finally leads to the onset of a metallic state in LaNiO_3 and LaCuO_3 .

Three members of the LaMO_3 family (LaTiO_3 , LaVO_3 , and LaMnO_3) are known to display orbital-ordering (OO) associated with the partially filled t_{2g} and e_g orbitals located at the top of the valence band (enclosed by thick lines in Fig. 9). In the following we describe briefly the most important characteristics of the observed OO states.

(i) In LaTiO_3 , where the OO originates from the single t_{2g} electron, the lobes have a quasi cigar-like shape with asymmetric contributions along the two main directions, indicating an almost identical occupation of the three xy , xz , and yz t_{2g} -shells. Coplanar lobes are arranged in a checkerboard-like way with a sign alternation along z , in good agreement with previously reported theoretical [79] and experimental works [80]. There is a clear connection between this checkerboard-like Ti d^1 ordering and the JT structural instability, which is manifested by the tendency of the occupied t_{2g} state to lie along the longer Ti-O bond. This also explains why the checkerboard-like OO in LaTiO_3 is not as evident as in LaMnO_3 : In LaTiO_3 the difference between the distinct Ti-O bond lengths Ti-O_s , Ti-O_m , and Ti-O_l , quantified by the JT parameters Q_2 and Q_3 , is about one order of magnitude smaller than in LaMnO_3 [31].

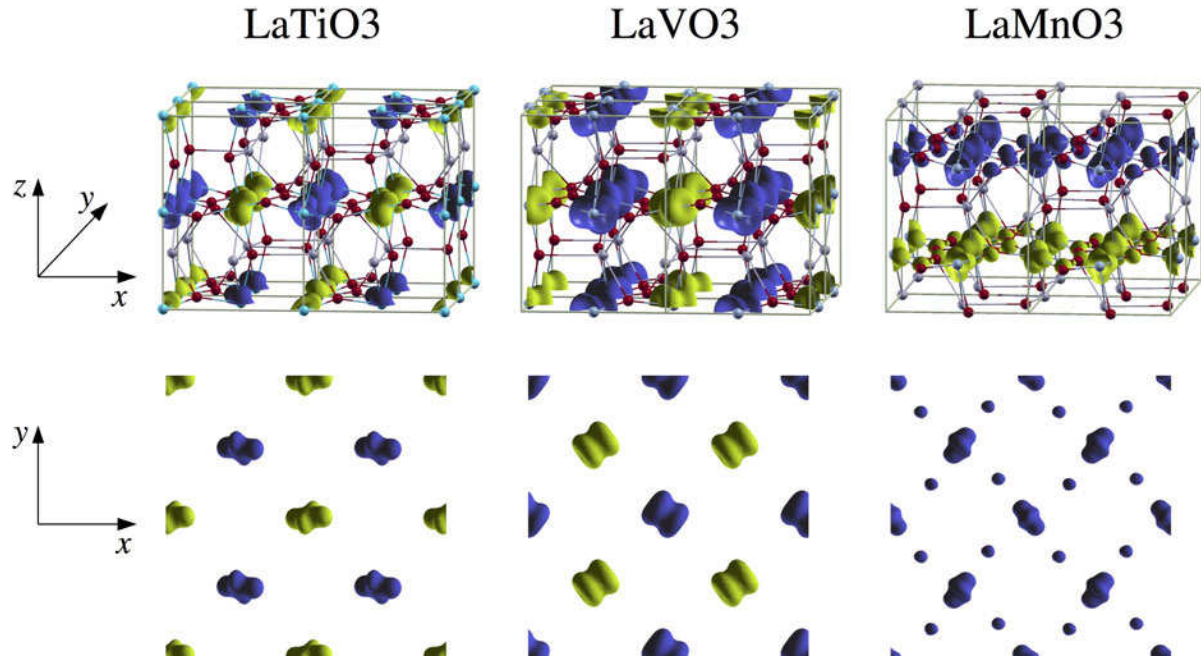


Fig. 10: *Isosurfaces of the magnetic orbitally ordered charge density for LaTiO_3 , LaVO_3 , and LaMnO_3 associated with the topmost occupied bands highlighted in the insets of Fig. 9. Light (yellow) and dark (blue) areas represent spin down and spin up, respectively, indicating the different types of spin orderings in LaTiO_3 (G-type), LaVO_3 (C-type), and LaMnO_3 (A-type). Top panel: three dimensional view; Bottom: projection onto the xy plane.*

(ii) The V^{3+} ions in LaVO_3 can accommodate two electrons in the three possible orbital states d_{xy} , d_{xz} , and d_{yz} . The spins are arranged according to C-type ordering, whereas the OO state is found to be G-type, in accordance with the Goodenough-Kanamori rules [81] and in agreement with X-ray diffraction [82] and previous GGA [83] and HF [84] calculations. The distribution of the t_{2g} orbitals in the G-type OO state follows the cooperative JT-induced V-O bond-alternation in the xy plane and along the z axis, i.e., the t_{2g} charge density in one specific V site is rotated by 90° with respect to that in the 6 neighboring V sites (four in-plane and two in the adjacent vertical planes). As already observed for LaTiO_3 , the t_{2g} orbitals are preferentially occupied along the long-bond direction.

(iii) The C-type OO in LaMnO_3 , originating from the singly occupied e_g state of the Mn^{+3} $3d$ electrons in the high-spin configuration $t_{2g}^3 e_g^1$ has been extensively studied both experimentally [85], and theoretically [86]. We have also recently addressed this issue through a maximally localized Wannier function representation of the e_g states [37]. This C-type OO state can be written in the form $|\theta\rangle = \cos\frac{\theta}{2}|3z^2 - r^2\rangle + \sin\frac{\theta}{2}|x^2 - y^2\rangle$ with the sign of $\theta \sim 108^\circ$ alternating along x and y and repeating along z , as correctly represented by our HSE charge density plots.

3.2 $4d$ and $5d$ perovskites

In recent years there has been an upsurge of interest in $4d$ and $5d$ transition metal oxides in which exotic states may emerge from the subtle interplay between Hubbard's U , Hund's J , the bandwidth, the spin-orbit coupling (SOC), and the splitting of the crystal field. It is commonly expected that $4d$ and $5d$ oxides are more metallic and less magnetic than their $3d$ counterparts because of the more extended nature of the $4d$ and $5d$ orbitals. In contrast with these expectations, many ruthenates, technetates, and iridates are found to be magnetic insulators and to display a large array of phenomena rarely or never seen in other materials, including relativistic Mott-insulators, Slater insulators, Hund's correlated metals, molecular insulators, etc. Considering the complexity of the issues at hand, hybrid functionals appear again to be particularly adequate, but so far their application to $4d$ and $5d$ perovskites has been very limited (see Ref. [31] and references therein).

3.2.1 $RTcO_3$ ($R=Ca, Sr$ and Ba)

A few years ago, antiferromagnetism with a huge ordering temperature of up to 1000 K was found for $CaTcO_3$ and $SrTcO_3$. This is surprising not only because magnetism in $4d$ materials with more extended orbitals has been rarely found, but also because the reported Néel temperature (T_N) is by far the highest among materials not incorporating $3d$ transition metals. Since Tc is a radioactive element, those perovskites have been rarely investigated, and for many of them only the structural properties are known. This represents a great challenge for *ab-initio* schemes, and indeed several groups have in the last few years intensively studied this class of materials [35, 87–89]. These studies have shown that the proper treatment of the attenuated (but still important) electronic correlation and its coupling with the magnetic exchange interactions is capable of explaining the onset of the remarkable magnetic ordering temperatures.

Here we summarized the HSE results discussed in more detail in Ref. [35]. The main objective is to be able to predict the Néel temperature. From an *ab-initio* perspective, this can be achieved in a simple way by mapping total energies for different magnetic configurations onto a general Heisenberg Hamiltonian, extracting the nearest-neighbor (NN, J_1) and next-nearest-neighbor (NNN, J_2) magnetic exchange parameters, and then using Monte Carlo simulations to compute T_N . To obtain good results, very accurate total energies and spin moments are needed. In this case LDA/PBE are not of great help as the magnetic moments for some magnetic orderings are seriously underestimated [90]. HSE provides very stable magnetic solutions for different types of spin orderings with almost identical values of the magnetic moment of Tc ($\approx 2 \mu_B$).

The first question is Why is T_N so large? DMFT attributes this large T_N to the fact that $SrTcO_3$ is on the border between the itinerant-metallic and localized-insulating regimes [88]. From an *ab-initio* point of view it is instructive to start from Anderson's theory of super-exchange (SE) interactions [91], which links the strength of the SE coupling constant to the actual hybridization between the metal and the mediating atom, and from Van Vleck's theory of antiferromagnetism [92], which connects the strength of the SE interaction to the magnetic ordering temperature. In this case, the HSE antiferromagnetic NN J_1 coupling constant is the dominant

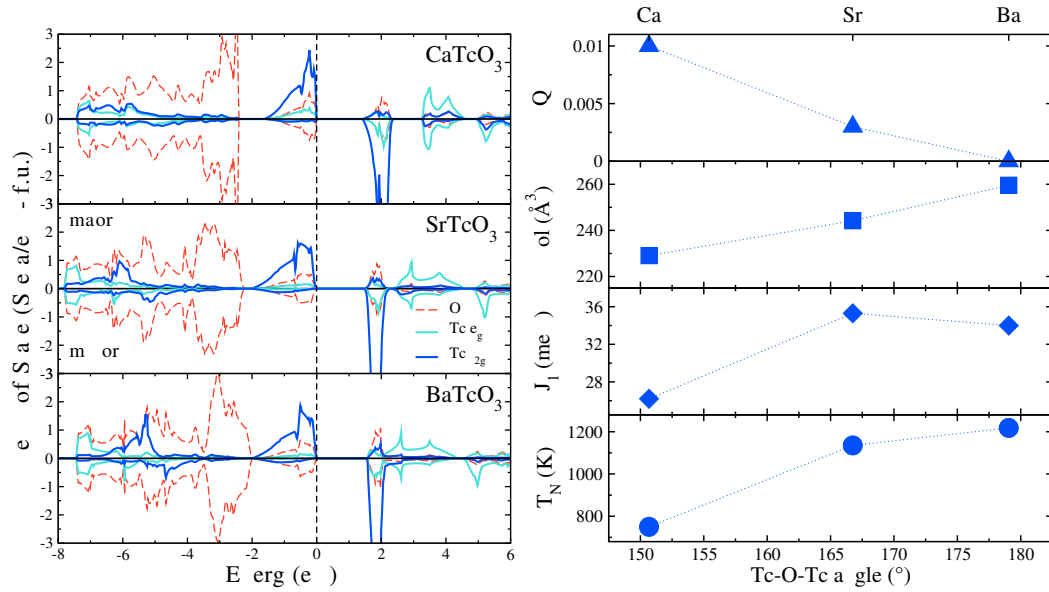


Fig. 11: (left) Density of states. (right) Dependence of the relevant magnetic (T_N and J_1) and structural [Volume and $Q=1/2(Q_1+Q_2)$] quantities on the $\widehat{\text{Tc}-\text{O}-\text{Tc}}$ angle (average between $\widehat{\text{Tc}-\text{O1}-\text{Tc}}$ and $\widehat{\text{Tc}-\text{O2}-\text{Tc}}$) in the $RT\text{cO}_3$ series.

parameter (~ -30 meV), almost two orders of magnitude larger than J_2 (~ -0.5 meV). The calculated density of states (DOS) displayed in Fig. 11 shows that this huge J_1 value arises from the strong covalency between the Tc t_{2g} and O p orbitals evolving along the wide $4d$ t_{2g} manifold, in particular for the topmost valence states spreading from $-1.5/2$ eV to the Fermi level (E_F). The increasing bandwidth (w) of this group of hybridized bands observed when going from CaTcO_3 ($w = 1.5$ eV) to SrTcO_3 and BaTcO_3 ($w = 2.0$ eV) associated with the enhanced t_{2g} - p hybridization in the 3-eV-wide t_{2g} band around -5.5 eV explains the larger J_1 and the correspondingly larger T_N for SrTcO_3 and BaTcO_3 .

The next fundamental question is Why does the Tc-O hybridization increase along the $RT\text{cO}_3$ series when r_R get larger? The answer comes from the interpretation of the right panel of Fig. 11. Going from Ca to Ba we observe (i) volume enhancement; (ii) quenching of the JT distortions Q_2 and Q_3 ($Q_2 = 2(l-s)/\sqrt{2}$ and $Q_3 = 2(2m-l-s)/\sqrt{6}$ with l , s , and m being the long, short, and medium Tc-O distances, respectively); (iii) the decrease of the cooperative rotation of the TcO_6 octahedra represented by the $\widehat{\text{Tc}-\text{O}-\text{Tc}}$ bond angles. The monotonic increase of the $\widehat{\text{Tc}-\text{O}-\text{Tc}}$ bond angles leads to the progressive rectification of the NN super-exchange paths. This generates, in a tight-binding framework, an enhanced Tc- t_{2g} /O- p hybridization, as confirmed by the DOS. T_N steeply increases from CaTcO_3 (750 K) to SrTcO_3 (1135 K) as a consequence of the observed larger change in $\widehat{\text{Tc}-\text{O}-\text{Tc}}$, which goes from 151° to 167° . When moving from SrTcO_3 to BaTcO_3 (1218 K) the rise of T_N is weaker due to a smaller change in $\widehat{\text{Tc}-\text{O}-\text{Tc}}$ (from 167° to 179°) and to a further reduction and sign change in J_2 . In analogy with the RMnO_3 perovskites, the increase of T_N for larger Tc-O-Tc angles correlates with a progressive reduction of the JT distortions (i.e., a decrease of the associated structural ordering temperature).

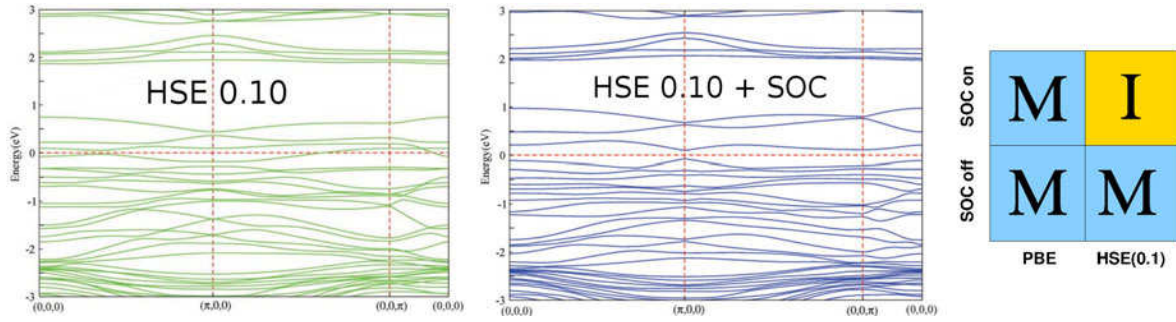


Fig. 12: Band structure and minimal phase diagram of BaIrO_3 . The relativistic Mott-Hubbard insulating state emerges from the inclusion of a small fraction of exact-exchange at HSE level and spin-orbit coupling.

3.2.2 BaIrO_3

As another example we discuss BaIrO_3 . Iridium oxides are a primary class of materials which has recently attracted a great attention. Iridates lie at the intersection of strong SOC and electron correlation, in which the electrons entangle the orbital and spin degrees of freedom. One of the most stunning examples is the Ruddlesden-Popper compound $\text{Sr}_{n+1}\text{Ir}_n\text{O}_{3n+1}$ ($n = 1, 2, \dots \infty$). The origin of the unusual insulating state in the parent $n = 1$ compound was long debated, but a general consensus exists in attributing the opening of the gap to a relativistic Mott-Hubbard type mechanism [93]. This compound has not yet been studied at the HSE level.

BaIrO_3 does not crystallize in the usual perovskite structure, as its tolerance factor is substantially larger than 1; it assumes a quasi one-dimensional phase consisting of two characteristic nonequivalent Ir_3O_{12} clusters, each of which is made of three face-sharing IrO_6 octahedra [94]. In Fig. 12, we show an unpublished HSE+SOC band structure and a minimal PBE/HSE-SOC phase diagram showing the onset of a SOC-Mott state in BaIrO_3 . Similar conclusions have been deduced from a recent LDA+U study [95]. At the PBE level (with and without SOC, not shown) the system is metallic, in disagreement with the expected insulating ground state (with a very small band gap). An equally wrong picture is obtained by HSE if SOC is not included. Only the inclusion of SOC within a beyond-DFT level calculation (HSE or LDA+U) leads to a correct description of this compound, which is categorized as a relativistic Mott insulator, i.e., the gap is opened by the concerted action of both electronic correlation and spin-orbit coupling.

It has been proposed that this is another example of a $J_{\text{eff}} = 1/2$ state, similar to the one observed in Sr_2IrO_4 [93]. The diagram of this $J_{\text{eff}} = 1/2$ state is schematized in Fig. 13: The crystal field splitting separates the e_g and t_{eg} manifold; the bonding between oxygen and inequivalent Ir sites yields the formation of two continuous subbands (i.e. without band gap); the strong SOC splits the t_{eg} band into effective $J_{\text{eff}} = 3/2$ and $J_{\text{eff}} = 1/2$ bands with the $J_{\text{eff}} = 1/2$ energetically higher than the $J_{\text{eff}} = 3/2$; finally the Hubbard U opens up a gap among $J_{\text{eff}} = 1/2$ states.

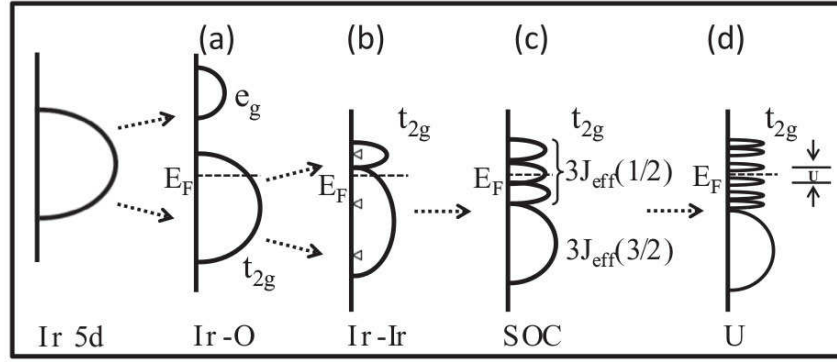


Fig. 13: Schematic diagram of the $J_{\text{eff}} = 1/2$ state in BaIrO_3 .
 (a) Crystal field splitting; (b) Chemical bonding; (c) SOC; (d) Hubbard U .

3.3 Metal to insulator transition: LaMnO_3 and BaBiO_3

Metal-insulator transitions are among the most studied processes in solid state physics and in particular in oxides. The driving force behind this type of transition can have different origins: electronic correlation, magnetic ordering, spin-orbit coupling, temperature, pressure, chemical doping, electron-phonon coupling, etc. [6]. For a method aiming to describe the MIT it is necessary to provide a sound description of both the insulating and the metallic regime. In this sense, DFT approximations are not apt, as the insulating state is often treated incorrectly. Hybrid functionals, specifically HSE and PBE0, represent a convenient solution as they can attain the homogeneous electron gas limit and are in principle applicable to the metallic state. Clearly, due to the dense $\mathbf{k}$ -point integration required to describe the metallic state and the associated huge computational cost, screened hybrids like HSE are highly preferable over PBE0 for practical applications, especially for perovskites.

Here we discuss two paradigmatic metal-to-insulator transitions: the well-known pressure-induced MIT in the prototypical Mott-Hubbard/JT perovskite LaMnO_3 [37] and the more exotic electron-phonon mediated MIT in BaBiO_3 [33, 34].

3.3.1 LaMnO_3 under pressure

At zero pressure (volume V_0) and low temperature, LaMnO_3 is a type-A AFM insulator characterized by staggered JT and GdFeO_3 -type (GFO) distortions, manifested by long and short Mn-O in-plane distances and medium Mn-O vertical ones and by the tilting of the Mn^{3+}O_6 octahedra (see Fig. 14(d)).

The application of hydrostatic pressure progressively quenches the cooperative JT distortions and leads to a MIT at $P_c = 32$ GPa [96]. The persistence of the structural distortions up to P_c indicates that the MIT is not of Mott-Hubbard type. This conclusion was initially proposed by LDA+U and dynamical mean-field theory studies [97, 98] and only very recently confirmed by high-pressure Raman measurements [99]. Baldini and coworkers [99] have also reported the coexistence of domains of distorted and regular octahedra in the pressure range 3-34 GPa, and connected the onset of metallicity with the increase of undistorted MnO_6 octahedra beyond a

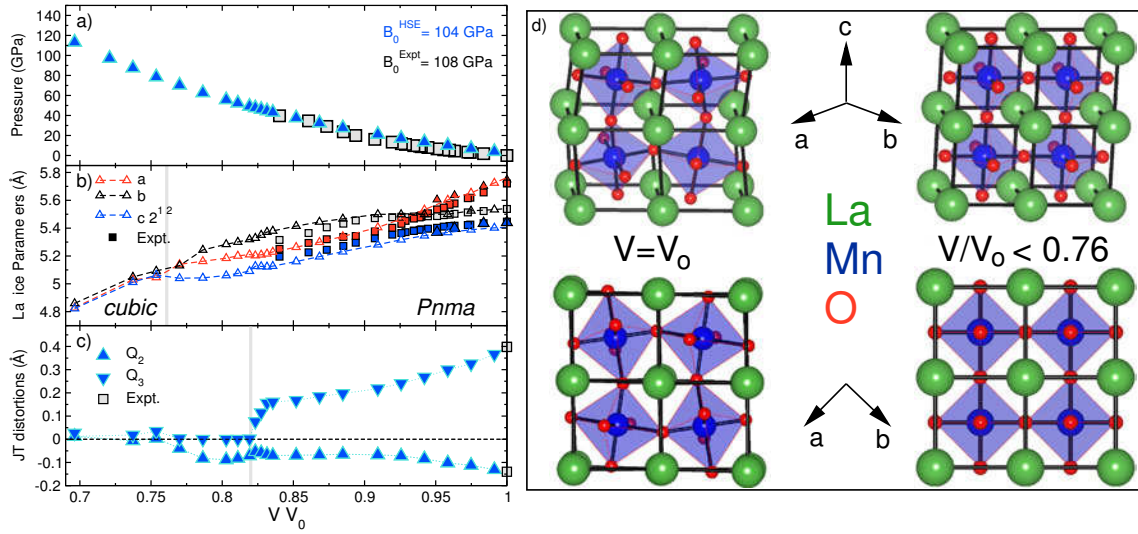


Fig. 14: Evolution of the structural properties of LaMnO_3 as a function of $v = V/V_0$ as predicted by HSE (triangles) compared to the experimental data (squares) (see Ref. [37] and references therein). (a) Equation of state (the inset gives the bulk modulus); (b) Structural parameters: at $v_3 = 0.76$ LaMnO_3 undergoes an orthorhombic ($a \neq b \neq \sqrt{c}/2$) to cubic ($a = b = \sqrt{c}/2$) transition marked by the vertical line. (c) Progressive quenching of the cooperative Jahn-Teller local modes $Q_2 = 2(l - s)/\sqrt{2}$ and $Q_3 = 2(2m - l - s)/\sqrt{6}$ with increasing pressure, where l , s , and m indicate the long, short, and medium Mn-O bond lengths; the JT modes are almost completely quenched at the onset of metallicity, marked by the vertical line at $v_2 = 0.82$; (d) Side (top) and top (bottom) view of the Pnma (left, $V/V_0 = 1$) and cubic (right, $V/V_0 < 0.76$) phases of LaMnO_3 , underlining the suppression of the JT and GFO structural distortion in the perfect cubic phase.

critical threshold. The concomitant presence of two distinct phases in this pressure range was confirmed by the X-ray absorption spectroscopy experiments of Ramos *et al.* [100].

The progression of the structural properties of compressed LaMnO_3 computed by HSE as a function of $v = V/V_0$ is summarized in Fig. 14, whereas the corresponding trends in the electronic and magnetic properties are shown in Figs. 15. The MIT can be described as follows.

In the pressure range 0–35 GPa, HSE results are in very good agreement with measurements in terms of (i) the pressure-volume equations of state and bulk modulus B_0 ($B_0^{\text{Expt}} = 108$ GPa, $B_0^{\text{HSE}} = 104$ GPa, see Fig. 14(a)), (ii) the pressure-induced changes in the structural parameters (Fig. 14(b)), and (iii) the concurrent suppression of the JT modes Q_2 and Q_3 and the band gap at the same compression ($v_2 = 0.82$, slightly smaller than the experimental one, $V/V_0 = 0.86$, see Fig. 14(b) and Fig. 15(a)); the $P = 0$ HSE gap opened between occupied and empty e_g states, $E_g = 1.45$ eV (Fig. 15(c)), is well within the measured range, 1.1–1.7 eV. Similarly the HSE ground state values of Q_2 and Q_3 match exactly the experimental values. The incremental compression of LaMnO_3 leads to a continuous structural transformation from the $P = 0$ distorted Pnma phase to a perfect cubic structure via a gradual quenching of the JT modes, the rectification of the GFO tilting angles, and the alignment of the a , b , and c lattice parameters towards the same value, ≈ 5.1 Å at $v_3 = 0.76$ as outlined in Fig. 14(b-d).

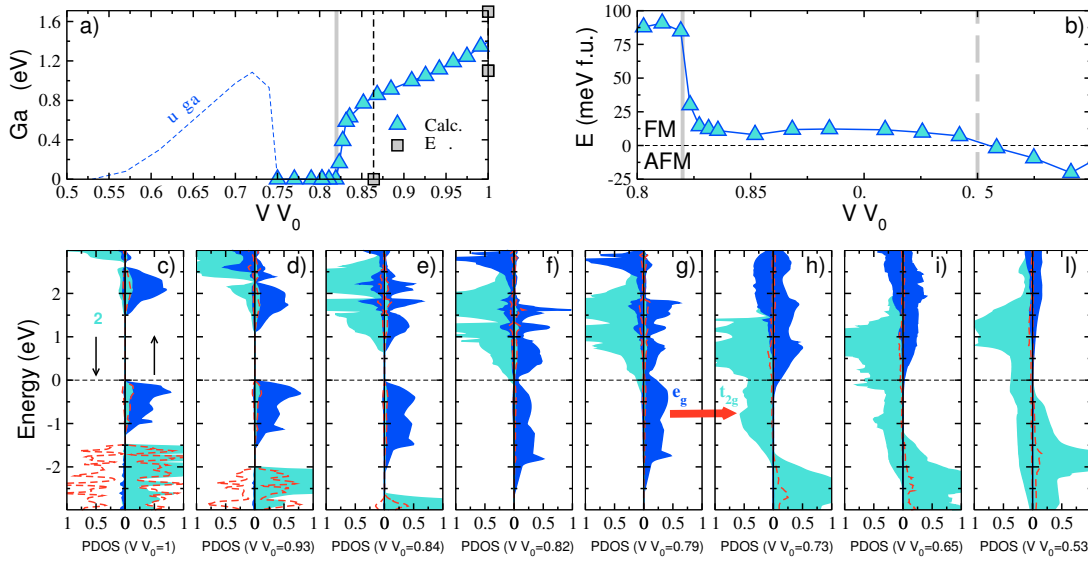


Fig. 15: (Color online) Evolution of the calculated electronic properties and magnetic ordering of LaMnO_3 under pressure. (a) Band gap: MIT at $v_2 = 0.82$ (marked by the vertical full line; the dashed line refers to the corresponding experimental onset). (b) Energy difference ΔE between the AFM and FM spin arrangements: AFM/FM crossover at $v_1 = 0.95$ (indicated by a vertical dashed line) and stabilization of the FM state at $v_2 = 0.82$. (c-l) Changes in the e_g and t_{2g} density of states around the Fermi level with pressure. The dashed (red) lines refer to the Oxygen p states, whereas the thick (red) arrow indicates the transfer of one electron from the e_g to the t_{2g} sub-bands.

As for the electronic properties, e_g bands around the Fermi energy (E_F) come progressively closer until the gap is closed (Fig. 15(c-f)). Concomitantly, the unoccupied t_{2g} bands are pushed down in energy and ultimately cross E_F at $v_2 = 0.82$, the onset of metallicity (see Fig. 15(f)). At this critical volume HSE predicts a jump in the relative stability between the AFM and FM ordering, with the latter becoming the more favorable by about 90 meV/f.u., as illustrated in Fig. 15(b). At low/intermediate compressions ($V/V_0 > v_2 = 0.82$) the data displayed in Fig. 15(b) shows a strong competition between the AFM and FM phases. HSE predicts a crossover between the AFM and FM phases at $v_1 = 0.95$ (corresponding to a pressure of 11 GPa), below which the AFM and FM ordering become almost degenerate ($\Delta E < 12$ meV/f.u.). Considering that in the FM phase the JT/GFO distortions are almost completely inhibited, this result strongly supports the latest Raman [99] and X-ray absorption spectroscopy [100] studies reporting the formation of a mixed state of domains of distorted and regular MnO_6 octahedra in the range 13–34 GPa, which compare well with the corresponding theoretical pressure range, 11–50 GPa ($v_2 < V/V_0 < v_1$).

The FM transition at $V/V_0 = 0.82$ comes right before a high spin (HS, $S = 2$) to low spin (LS, $S = 1$) moment collapse, which is correlated with the e_g and t_{2g} orbital occupations: Under compression the Mn^{3+} ion retains its $P = 0$ ($t_{2g}^{\uparrow\uparrow\uparrow}(e_g)^{\uparrow}$) orbital configuration down to $V/V_0 = 0.80$, with a magnetic moment of $3.7 \mu_B$; further compression yields a rapid reduction of the magnetic moment down to $1.7 \mu_B$ due the redistribution of electrons within the $3d$ shell

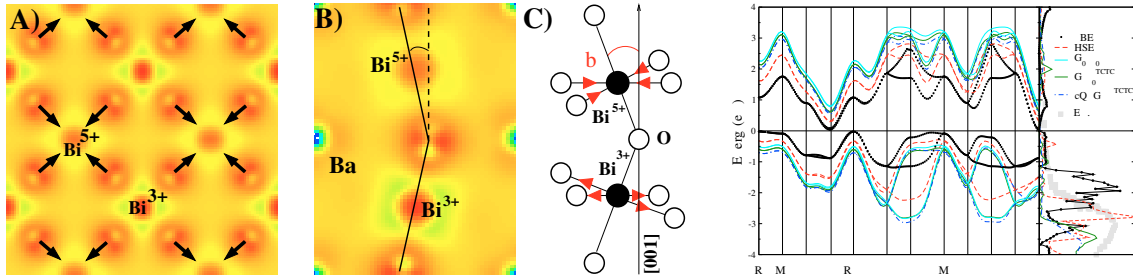


Fig. 16: (left) Charge density plots showing (a) the breathing distortion in the (100) plane and (b) the tilting instability projected in the (001) plane. Bi⁵⁺ (Bi1) and Bi³⁺ (Bi2) indicate the two inequivalent bismuth ions, whereas the arrows indicate the breathing displacement of the oxygen atoms. (c) Schematic view of the BaBiO₃ unit cell showing the BiO₆ tilted octahedra. Color coding: darker (red) areas indicates a high value of the charge density. (right) Calculated band structure of monoclinic BaBiO₃ within PBE, HSE G₀W₀, GW₀^{TCTC} (TCTC=test-charge test-charge) and scQPGW^{TCTC} (scQP=selfconsistent quasiparticle) and corresponding total density of states.

that ultimately leads to the low-spin configuration $(t_{2g})^{\uparrow\uparrow\downarrow}(e_g)^0$. This HS-orthorhombic to LS-cubic transition is reflected in the DOS (Fig. 15 (g-i)), whose evolution from $V/V_0 = 0.79$ to $V/V_0 = 0.73$ clearly indicates the transfer of one electron from the e_g to the t_{2g} sub-bands and the subsequent realization of a nearly FM half-metallic state with a metallic minority t_{2g} band and a quasi-insulating majority channel with a residual density of e_g electrons at the bottom of the conduction band.

Finally, by computing spin-dependent transport properties of the FM phase, we found a spin polarizations of 80%-90%, values very similar to those reported for the doped colossal magnetoresistance (CMR) manganite La_{0.7}Sr_{0.3}MnO₃. Thus, it can be concluded that the high pressure FM cubic phase of LaMnO₃ is a *transport* half-metal (tHM). The FM-tHM regime being the crucial common ingredient of all CMR manganites, its realization in the undoped (stoichiometric) phase of the CMR parent compound LaMnO₃ in a wide interval of compressions could give new fundamental insights into the elusive phenomenon of CMR.

3.3.2 Hole-doped BaBiO₃

We now focus on the charge-ordered mixed-valence insulating perovskite BaBiO₃. Despite its apparently simple *sp* nature, BaBiO₃ exhibits a plethora of fascinating and unique behaviors, including charge-density-wave (CDW) formation, superconductivity, polaron formation, two-dimensional electron gas, and topological effects.

Let us start with the ground state. BaBiO₃ is a prime example of a multivalent Peierls compound, whose semiconducting nature can be tuned into a metallic/superconducting one by chemical doping [101]. The primitive cell of BaBiO₃ can be described as Ba₂²⁺Bi³⁺Bi⁵⁺O₆²⁻, where Bi⁵⁺ and Bi³⁺ cations occur in equal parts. The two Bi species are alternately ordered in a distorted cubic (monoclinic) structure, in which Bi⁵⁺ is surrounded by Bi³⁺ neighbors (and vice versa), with alternate breathing-in and breathing-out distortions of oxygen octahedra

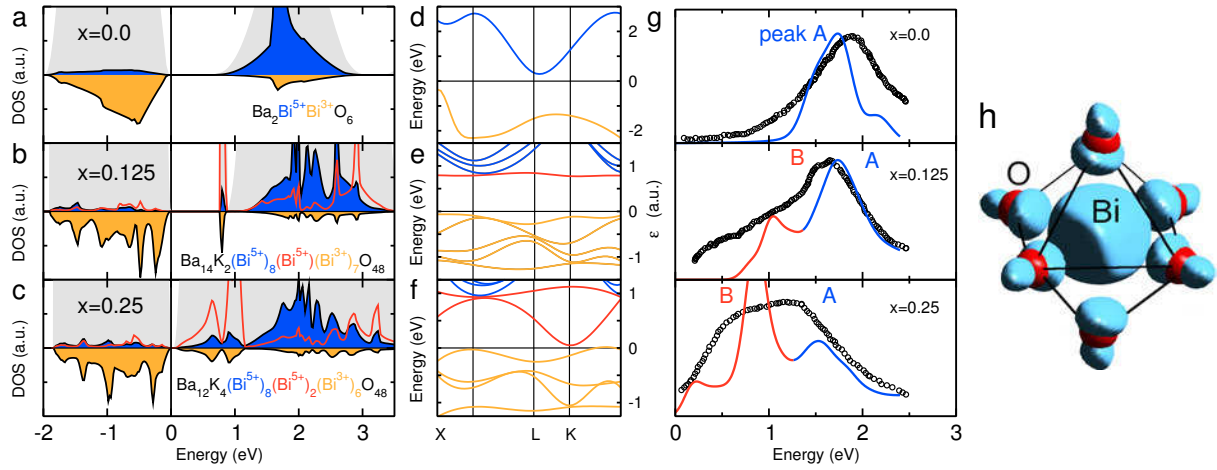


Fig. 17: Polaron-driven insulator-to-metal transition in K -doped $\text{Ba}_{x-1}\text{K}_x\text{BiO}_3$ (a-c) Evolution of the DOS and (d-f) corresponding bandstructures. The gray shadows indicate the total density of states, and red curves indicate the bipolaronic states. (g) Comparison between the theoretical and measured imaginary part of the dielectric function for $x = 0, 0.125$, and 0.25 . Peak (A) corresponds to excitations from Bi^{3+} into Bi^{5+} states, peak (B) corresponds to excitations from Bi^{3+} into bipolaronic states [red curves in panel (b),(c),(e), and (f)]. (h) Charge density corresponding to the bipolaronic band (red line) of panel (e) localized around the BiO_6 octahedron at the converted $\text{Bi}^{3+} \rightarrow \text{Bi}^{5+}$ atom [33].

around inequivalent Bi sites, as sketched in Fig. 16. As a consequence of this charge ordering, the formally expected metallic state for the cubic perovskite $\text{BaBi}^{4+}\text{O}_3$ is replaced by an insulating regime characterized by a large direct optical response $E_d \approx 2.0$ eV and an indirect optical transition E_i , which is well described by HSE, as shown in the band structure of Fig. 16. Upon hole doping through $\text{Ba} \rightarrow \text{K}$ substitution $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ undergoes an insulating to metal transition for $x \approx 0.33$, eventually turning into a superconductor for higher doping. HSE explains this MIT as a progressive reduction of the Bi-O distortions modulated by the formation of hole-polarons, i.e., the coupling between the excess holes induced by the K-doping trapped in Bi^{3+} sites and the surrounding phonon field [102].

This is shown in Figure 17: (i) At $x = 0$ a band gap is opened between the occupied Bi^{3+} s states and the unoccupied Bi^{5+} s band. The optical spectrum is characterized by a main peak in agreement with experiment. (ii) At $x = 0.125$ a very localized bi-polaronic mid gap states emerges, which is also recognizable in the optical spectrum. (iii) Upon further hole doping additional bi-polaronic states are formed, which start to interact among each other as reflected by the width of the mid-gap states, consistent with the experimental signal. As a consequence the band gap is progressively reduced and ultimately closes for $x > 0.25$, in good agreement with experiment.

The hole trapping is accompanied by a fairly large local relaxation involving the breaking of the perfectly checkerboard CDW ($x = 0$) and the rearrangement of the oxygen octahedra around the Bi ions, since each single Bi^{3+} cation captures two holes and therefore tends to attract the negatively charged oxygens more.

The electron-phonon mechanism behind the formation of the superconducting state has been discussed by Yin *et al.* by combining HSE (along with other theoretical schemes) with a model approach to evaluate the electron-phonon coupling and the critical temperature (T_C) based on the McMillan equation [103]. It was shown that HSE corrects the LDA/GGA overscreening (which causes the underestimation of the electron-phonon coupling) and leads to a nice agreement with experiment in terms of T_C for a Coulomb pseudopotential of $\mu^* = 0.1$.

3.4 Multiferroics

Multiferroics constitute an important class of compounds in which different ferroic orders such as ferromagnetism, ferroelectricity and/or ferroelasticity may coexist in a single compound. Many multiferroics are transition-metal oxides with perovskite crystal structure, including, among others, rare-earth manganites and the famous BiFeO_3 , and BiMnO_3 .

The application of hybrid functionals to the classical model ferroelectric oxides SrTiO_3 and BaTiO_3 has been discussed in detail in the early works of Bilc *et al.* [104] and Wahl *et al.* [105], as well as in the more recent study of the Scuseria group [106]. A further valid illustration of the applicability of HSE to multiferroic materials is supplied by the work of Stroppa and Picozzi, in particular the detailed study of the structural, electronic, magnetic, and ferroelectric properties of the two prototypical *proper* and *improper* multiferroic systems BiFeO_3 and orthorhombic HoMnO_3 , respectively [107].

In recent times, it has been found that pressure can be used as a means to alter the electronic bonding state, the lattice, and thus the physical properties of perovskite compounds, and thereby to induce the onset of multiferroic behaviors. As an example, Inaguma *et al.* have synthesized two novel high-pressure polymorphs of the transition metal perovskites PbNiO_3 and CdPbO_3 characterized by a hexagonal lithium niobate (LiNbO_3)-type (LNO) structure with space group $R3c$ [108, 109]. The application of HSE has revealed that both these compounds are electrically polarized with a pretty large electric polarization of $\sim 100 \mu\text{C}/\text{cm}^2$ [38, 40] and $\sim 52.3 \mu\text{C}/\text{cm}^2$ [39], respectively.

As a final example we discuss the multiferroic character of PbNiO_3 as predicted by HSE. The low-pressure phase of PbNiO_3 has a rhombohedrally non-centrosymmetric structure with space group $R3c$, which is isostructural to the most common multiferroic material BiFeO_3 (see Fig. 18). More interestingly, the rhombohedral phase undergoes an antiferromagnetic transition at 205 K and exhibits semiconducting transport properties. This therefore suggests a possible multiferroic behavior, i.e., co-existence of ferroelectric and magnetic properties.

A compact representation of the structural, electronic and ferroelectric properties of PbNiO_3 is given in Figure 18. In this compound the polarization is driven by the large Pb-O polar displacement along the [111] direction, which is typical of the acentric LNO structure (like BiFeO_3). The origin of the electric polarization in PbNiO_3 is revealed by the comparison between the paraelectric and ferroelectric density of states (see Figure 18(b)) showing the 2 eV downshift and broadening of Pb 6s-O 2p spectral weight occurring along with the centrosymmetric-to-ferroelectric transformation. For PbNiO_3 GGA not only gives inaccurate ferroelectric distor-

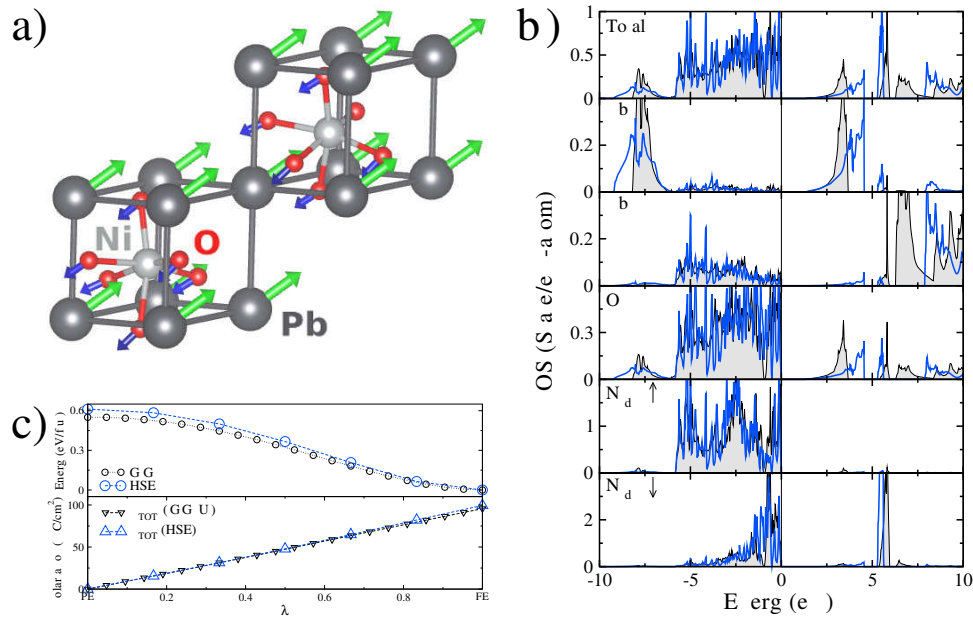


Fig. 18: Ferroelectric properties of PbNiO₃. (a) Structural model showing the ferroelectric displacements. (b) HSE calculated density of states in the acentric ferroelectric (full lines) and centric paraelectric (shadow) phases. (c) Total energy E and total polarization $\mathbf{P}_{\text{tot}}$ as a function of the polar distortion λ .

tions (with errors exceeding 2 %) but also yields an incorrect metallic ground state, which prevents any further possibility of exploring ferroelectric features [38]. In contrast, both HSE and PBE+U lead to the correct picture, ultimately delivering an almost identical electric polarization of about 100 C/cm² (see Figure 18(c)). The spontaneous polarization P of rhombohedral PbNiO₃ was calculated by HSE considering as centrosymmetric structural reference the $R\bar{3}c$ symmetry, and we linearly interpolate the atomic positions between the centric and the polar phase, i.e., the so-called adiabatic path.

Acknowledgment

I gratefully acknowledge valuable discussions on hybrid functionals and great collaborations over the years with Jiangang He, Xianfeng Hao, Sun Yan, G. Kresse, R. Podloucky, A. Filipetti, S. Sanvito, Xing-Qiu Chen and A. Stroppa. I would like to thank Tommaso and Massimo for the kind hospitality during the writing of these notes.

References

- [1] V. Goldschmidt, *Naturwissenschaften* **14**, 477 (1926).
- [2] R. von Helmolt, J. Wecker, B. Holzapfel, L. Schultz, and K. Samwer, *Phys. Rev. Lett.* **71**, 2331 (1993)
- [3] M.B. Salamon and M. Jaime, *Rev. Mod. Phys.* **73**, 583 (2001)
- [4] S. Picozzi and C. Ederer, *Journal of Physics: Condensed Matter* **21**, 303201 (2009)
- [5] J. Bednorz and K. Müller, *Zeitschrift für Physik B Condensed Matter* **64**, 189 (1986)
- [6] M. Imada, A. Fujimori, and Y. Tokura, *Rev. Mod. Phys.* **70**, 1039 (1998)
- [7] F. Fukunaga and N. Tsuda, *Journal of the Physical Society of Japan* **63**, 3798 (1994)
- [8] T. Arima, Y. Tokura, and J.B. Torrance, *Phys. Rev. B* **48**, 17006 (1993)
- [9] H. Tanaka and M. Misono, *Current Opinion in Solid State and Materials Science* **5**, 381 (2001)
- [10] J. Hubbard, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **276**, 238 (1963)
- [11] W. Kohn and L.J. Sham, *Phys. Rev.* **140**, A1133 (1965)
- [12] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964)
- [13] V. Fock, *Zeitschrift für Physik* **61**, 126 (1930)
- [14] C.C.J. Roothaan, *Rev. Mod. Phys.* **23**, 69 (1951)
- [15] J.P. Perdew, A. Ruzsinszky, J. Tao, V.N. Staroverov, G.E. Scuseria, and G.I. Csonka, *The Journal of Chemical Physics* **123**, 062201 (2005)
- [16] J.C. Slater, *Phys. Rev.* **34**, 1293 (1929)
- [17] J.A. Alonso and L.A. Girifalco, *Phys. Rev. B* **17**, 3735 (1978)
- [18] V.I. Anisimov, J. Zaanen, and O.K. Andersen, *Phys. Rev. B* **44**, 943 (1991)
- [19] J.P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981)
- [20] D.M. Bylander and L. Kleinman, *Phys. Rev. B* **41**, 7868 (1990)
- [21] M. Grüning, A. Marini, and A. Rubio, *The Journal of Chemical Physics* **124**, 154108 (2006)

- [22] J. Tao, J.P. Perdew, V.N. Staroverov, and G.E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003)
- [23] A.D. Becke, The Journal of Chemical Physics **98** (1993)
- [24] L. Hedin, Phys. Rev. **139**, A796 (1965)
- [25] C. Møller and M.S. Plesset, Phys. Rev. **46**, 618 (1934)
- [26] J. Paldus, The Journal of Chemical Physics **61** (1974)
- [27] J. Cizek, The Journal of Chemical Physics **45** (1966)
- [28] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989);
A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [29] 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>
- [30] 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>
- [31] C. Franchini, Journal of Physics: Condensed Matter **26**, 253202 (2014)
- [32] J. He and C. Franchini, Phys. Rev. B **86**, 235117 (2012)
- [33] C. Franchini, G. Kresse, and R. Podloucky, Phys. Rev. Lett. **102**, 256402 (2009)
- [34] C. Franchini, A. Sanna, M. Marsman, and G. Kresse, Phys. Rev. B **81**, 085213 (2010)
- [35] C. Franchini, T. Archer, J. He, X.-Q. Chen, A. Filippetti, and S. Sanvito,
Phys. Rev. B **83**, 220402 (2011)
- [36] J. He, M.-X. Chen, X.-Q. Chen, and C. Franchini, Phys. Rev. B **85**, 195135 (2012)
- [37] C. Franchini, R. Kovacik, M. Marsman, S.S. Murthy, J. He, C. Ederer, and G. Kresse,
Journal of Physics: Condensed Matter **24**, 235602 (2012)
- [38] X.F. Hao, A. Stroppa, S. Picozzi, A. Filippetti, and C. Franchini,
Phys. Rev. B **86**, 014116 (2012)
- [39] Y. Xu, X. Hao, C. Franchini, and F. Gao, Inorganic Chemistry **52**, 1032 (2013)
- [40] X.F. Hao, A. Stroppa, P. Barone, A. Filippetti, C. Franchini, and S. Picozzi,
New Journal of Physics **16**, 015030 (2014)

- [41] J. He and C. Franchini, Phys. Rev. B **89**, 045104 (2014)
- [42] X. Hao, Y. Xu, C. Franchini, and F. Gao, physica status solidi (b) **252**, 626 (2015)
- [43] R. Parr and W. Yang: *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1994)
- [44] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)
- [45] D.C. Langreth and M.J. Mehl, Phys. Rev. B **28**, 1809 (1983)
- [46] J.P. Perdew and S. Kurth: *A Primer in Density Functional Theory* (Springer, Heidelberg, 2003)
- [47] S. Kümmel and L. Kronik, Rev. Mod. Phys. **80**, 3 (2008)
- [48] E. Fermi and E. Amaldi, R. Accad. d'Italia. Memorie **6**, 119 (1934)
- [49] J.P. Perdew, R.G. Parr, M. Levy, and J.L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982)
- [50] J.P. Perdew, M. Ernzerhof, and K. Burke, The Journal of Chemical Physics **105** (1996)
- [51] A.D. Becke, The Journal of Chemical Physics **98** (1993)
- [52] T.L. Gilbert, Phys. Rev. B **12**, 2111 (1975)
- [53] A. Seidl, A. Görling, P. Vogl, J.A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996)
- [54] A.D. Becke, The Journal of Chemical Physics **98** (1993)
- [55] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, J. Phys. Chem. **98**, 11623 (1994)
- [56] J. Paier, M. Marsman, and G. Kresse, The Journal of Chemical Physics **127**, 024103 (2007)
- [57] J.P. Perdew, M. Ernzerhof, and K. Burke, The Journal of Chemical Physics **105** (1996)
- [58] C. Adamo and V. Barone, The Journal of Chemical Physics **110** (1999)
- [59] J. Paier, M. Marsman, K. Hummer, G. Kresse, I.C. Gerber, and J.G. Angyan, The Journal of Chemical Physics **124**, 154709 (2006)
- [60] C. Franchini, R. Podloucky, J. Paier, M. Marsman, and G. Kresse, Phys. Rev. B **75**, 195128 (2007)
- [61] J. Heyd, G.E. Scuseria, and M. Ernzerhof, The Journal of Chemical Physics **118**, 8207 (2003)

- [62] J. Heyd, G.E. Scuseria, and M. Ernzerhof,
The Journal of Chemical Physics **124**, 219906 (2006)
- [63] J.E. Moussa, P.A. Schultz, and J.R. Chelikowsky,
The Journal of Chemical Physics **136**, 204117 (2012)
- [64] T. Onishi, International Journal of Quantum Chemistry **108**, 2856 (2008)
- [65] A. Alkauskas, P. Broqvist, and A. Pasquarello, *physica status solidi (b)* **248**, 775 (2011)
- [66] B.G. Janesko, T.M. Henderson, and G.E. Scuseria,
Phys. Chem. Chem. Phys. **11**, 443 (2009)
- [67] T.M. Henderson, J. Paier, and G.E. Scuseria, *physica status solidi (b)* **248**, 767 (2011)
- [68] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and
C.G. Van de Walle, Rev. Mod. Phys. **86**, 253 (2014)
- [69] R. Del Sole, L. Reining, and R.W. Godby, Phys. Rev. B **49**, 8024 (1994)
- [70] F. Gygi and A. Baldereschi, Phys. Rev. Lett. **62**, 2160 (1989)
- [71] G. Kresse and J. Hafner, Phys. Rev. B **48**, 13115 (1993)
- [72] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996)
- [73] M.A.L. Marques, J. Vidal, M.J.T. Oliveira, L. Reining, and S. Botti,
Phys. Rev. B **83**, 035119 (2011)
- [74] T. Shimazaki and Y. Asai, Chemical Physics Letters **466**, 91 (2008)
- [75] D. Koller, P. Blaha, and F. Tran, Journal of Physics: Condensed Matter **25**, 435503 (2013)
- [76] V. Atalla, M. Yoon, F. Caruso, P. Rinke, and M. Scheffler,
Phys. Rev. B **88**, 165122 (2013)
- [77] J.H. Skone, M. Govoni, and G. Galli, Phys. Rev. B **89**, 195112 (2014)
- [78] M. Marsman, J. Paier, A. Stroppa, and G. Kresse,
Journal of Physics: Condensed Matter **20**, 064201 (2008)
- [79] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and
O.K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004)
- [80] M. Cwik, T. Lorenz, J. Baier, R. Müller, G. André, F. Bourée, F. Lichtenberg,
A. Freimuth, R. Schmitz, E. Müller-Hartmann, and M. Braden,
Phys. Rev. B **68**, 060401 (2003)
- [81] J.B. Goodenough, Progress in Solid State Chemistry **5**, 145 (1971)

- [82] Y. Ren, A.A. Nugroho, A.A. Menovsky, J. Strempfer, U. Rütt, F. Iga, T. Takabatake, and C.W. Kimball, *Phys. Rev. B* **67**, 014107 (2003)
- [83] H. Sawada, N. Hamada, K. Terakura, and T. Asada, *Phys. Rev. B* **53**, 12742 (1996)
- [84] I.V. Solovyev, *Phys. Rev. B* **73**, 155117 (2006)
- [85] Y. Murakami, J.P. Hill, D. Gibbs, M. Blume, I. Koyama, M. Tanaka, H. Kawata, T. Arima, Y. Tokura, K. Hirota, and Y. Endoh, *Phys. Rev. Lett.* **81**, 582 (1998)
- [86] E. Pavarini and E. Koch, *Phys. Rev. Lett.* **104**, 086402 (2010)
- [87] E.E. Rodriguez, F. Poineau, A. Llobet, B.J. Kennedy, M. Avdeev, G.J. Thorogood, M.L. Carter, R. Seshadri, D.J. Singh, and A.K. Cheetham, *Phys. Rev. Lett.* **106**, 067201 (2011)
- [88] L. de' Medici, J. Mravlje, and A. Georges, *Phys. Rev. Lett.* **107**, 256401 (2011)
- [89] J. Mravlje, M. Aichhorn, and A. Georges, *Phys. Rev. Lett.* **108**, 197202 (2012)
- [90] M. Avdeev, G.J. Thorogood, M.L. Carter, B.J. Kennedy, J. Ting, D.J. Singh, and K.S. Wallwork, *Journal of the American Chemical Society* **133**, 1654 (2011)
- [91] P.W. Anderson, *Phys. Rev.* **115**, 2 (1959)
- [92] J.H. Van Vleck, *The Journal of Chemical Physics* **9** (1941)
- [93] B.J. Kim, H. Jin, S.J. Moon, J.-Y. Kim, B.-G. Park, C.S. Leem, J. Yu, T.W. Noh, C. Kim, S.-J. Oh, J.-H. Park, V. Durairaj, G. Cao, and E. Rotenberg, *Phys. Rev. Lett.* **101**, 076402 (2008)
- [94] T. Siegrist and B. Chamberland, *Journal of the Less Common Metals* **170**, 93 (1991)
- [95] W. Ju, G.-Q. Liu, and Z. Yang, *Phys. Rev. B* **87**, 075112 (2013)
- [96] I. Loa, P. Adler, A. Grzechnik, K. Syassen, U. Schwarz, M. Hanfland, G.K. Rozenberg, P. Gorodetsky, and M.P. Pasternak, *Phys. Rev. Lett.* **87**, 125501 (2001)
- [97] G. Trimarchi and N. Binggeli, *Phys. Rev. B* **71**, 035101 (2005)
- [98] A. Yamasaki, M. Feldbacher, Y.-F. Yang, O.K. Andersen, and K. Held, *Phys. Rev. Lett.* **96**, 166401 (2006)
- [99] M. Baldini, V.V. Struzhkin, A.F. Goncharov, P. Postorino, and W.L. Mao, *Phys. Rev. Lett.* **106**, 066402 (2011)
- [100] A.Y. Ramos, N.M. Souza-Neto, H.C.N. Tolentino, O. Bunau, Y. Joly, S. Grenier, J.-P. Itie, A.-M. Flank, P. Lagarde, and A. Caneiro, *EPL (Europhysics Letters)* **96**, 36002 (2011)

- [101] R.J. Cava, B. Batlogg, J.J. Krajewski, R. Farrow, L.W. Rupp, A.E. White, K. Short, W.F. Peck, and T. Kometani, **332**, 814.
- [102] A.S. Alexandrov and J.T. Devrese: *Advances in Polaron Physics* (Springer, Heidelberg, 2010)
- [103] Z.P. Yin, A. Kutepov, and G. Kotliar, Phys. Rev. X **3**, 021011 (2013)
- [104] D.I. Bilc, R. Orlando, R. Shaltaf, G.-M. Rignanese, J. Íñiguez, and P. Ghosez, Phys. Rev. B **77**, 165107 (2008)
- [105] R. Wahl, D. Vogtenhuber, and G. Kresse, Phys. Rev. B **78**, 104116 (2008)
- [106] F. El-Mellouhi, E.N. Brothers, M.J. Lucero, and G.E. Scuseria, Phys. Rev. B **84**, 115122 (2011)
- [107] A. Stroppa and S. Picozzi, Phys. Chem. Chem. Phys. **12**, 5405 (2010)
- [108] Y. Inaguma, K. Tanaka, T. Tsuchiya, D. Mori, T. Katsumata, T. Ohba, K.-I. Hiraki, T. Takahashi, and H. Saitoh, Journal of the American Chemical Society **133**, 16920 (2011)
- [109] Y. Inaguma, M. Yoshida, T. Tsuchiya, A. Aimi, K. Tanaka, T. Katsumata, and D. Mori, Journal of Physics: Conference Series **215**, 012131 (2010)

Index

A

algebraic correlations, 7.8
algebraic spin liquid, 7.22
Anderson impurity model, 2.2, 5.4, 6.3,
6.16, 9.10, 13.18
Anderson lattice model, 9.11, 9.22
Anderson-Holstein model, 6.3
anisotropic Kondo model, 5.2, 5.9, 5.15
antiferromagnetic Heisenberg model, 11.7
antiferromagnetism, 3.17, 3.24, 9.23, 11.7,
11.21, 14.21, 15.15
atomic limit, 2.6

B

Barnes's representation, 9.4, 9.10
bath symmetries, 10.26
Baym-Kadanoff functional, 14.28
beta function, 4.10
Bethe ansatz, 2.3
Bethe lattice, 12.26
Bethe-Salpeter equation, 3.34, 14.11
bipartite lattice, 11.4
Bogoliubov transformation, 12.31
Bose-Fermi Kondo model, 6.16, 6.19
pseudogap version, 6.18, 6.20
bosonic coherent states, 6.3, 6.5
Brinkman-Rice transition, 12.15

C

canonical perturbation theory, 12.16
causality, 14.25
CDMFT, 10.22, 13.16
Ce compounds, 2.18
charge density wave, 13.30
charge fluctuations, 9.23
charge structure factor, 9.24
charge-transfer insulator, 15.15
checkerboard lattice, 7.10
chiral spin liquid, 7.23
cluster perturbation theory, 13.7
coherence, 1.6, 1.7, 1.9
colossal magnetoresistance, 15.26
competing interactions, 7.3
complete basis set, 5.24
conductance, 5.23

continued fraction, 10.10
Coqblin-Schrieffer interaction, 1.13
core hole, 2.8
Coulomb exchange, 3.7
Coulomb repulsion, 2.2
CT-QMC, 13.19
Curie-Weiss susceptibility, 1.3, 1.4

D

d-wave Cooper pairing, 11.21
DCA, 10.22, 14.3
DCA⁺, 14.13
de Almeida-Thouless line, 8.5
deconvolution, 14.17
density-functional theory, 15.4
dimer covering, 7.3, 7.18
dipolar correlations, 7.10
disorder, frozen, 8.2
dissipative two-state system, 5.3
distributed memory, 10.16
Doniach phase diagram, 1.8
dynamic cluster approximation, 14.3
dynamic properties, 5.19
dynamical Hartree approximation, 13.28
dynamical impurity approximation, 13.20
dynamical mean-field theory, 2.3, 3.30
Dyson's equation, 13.7

E

Edwards-Anderson parameter, 8.18
effective magnetic moment, 3.27
entropy, negative, 8.5, 8.23
equation of motion, 2.4
ergodic theorem, 8.6
ergodicity, 8.5–8.7
breaking, 8.7, 8.14
exact diagonalizations, 13.15
exchange interaction
random frustrated, 8.2, 8.5
extended Hubbard model, 9.26
extended interactions, 13.28

F

Fermi liquid, 4.13, 4.14, 5.16
theory, 3.14

ferrimagnetism, 11.7
 ferromagnetism, 3.15, 9.21, 9.23, 9.24,
 9.27, 11.9, 11.11
 finite-size calculations, 14.8
 first principles, 15.2
 fixed point, 4.5, 4.16, 5.15
 flat band ferromagnetism, 11.12
 fluctuation corrections, 9.23
 fluctuation-dissipation theorem, 3.48
 free energy, maximal, 8.11
 frustration, 7.2
 full density matrix, 5.25

G

gauge symmetry, 9.8, 9.10, 11.3
 geometrical frustration, 7.2
 ghost states, 10.8
 Gorkov function, 13.26
 grand potential functional, 14.28
 graphene, non-magnetic impurity in, 13.23
 Green function, 13.2
 computing averages from, 13.14
 periodization, 13.11
 Gutzwiller projection, 7.22, 7.23
 Gutzwiller wave function, 12.9

H

harmonic oscillator
 displaced, 6.3, 6.11
 Hartree-Fock, 2.5, 3.27, 15.6
 heavy-fermion materials, 1.2
 heavy-fermion superconductor, 1.23
 height model, 7.8
 Heisenberg antiferromagnet, 12.18
 Heisenberg model, 3.7, 7.2, 7.6
 Heyd-Scuseria-Ernzerhof functional (HSE),
 15.10
 Hubbard bands, 12.4
 Hubbard model, 3.8, 9.12, 9.23, 11.2,
 12.2, 13.2, 14.2
 Hubbard-I approximation, 12.3
 Hund's rules coupling, 4.19
 hybrid functional, 15.4
 hybridization function, 13.17
 hybridization sum rule, 10.22

I

I-CDMFT, 13.23

imaginary time Green function, 3.44
 interacting resonant level model, 5.9
 intermediate states method, 2.2
 intermediate-coupling fixed point, 4.16
 inverse photoemission, 2.2
 irreducible vertex, 14.11
 Ising model, 7.3, 7.8, 7.9
 classical long-range, 6.15

J

J_1 - J_2 model, 7.6, 7.13
 Jahn-Teller, 15.23

K

kagome lattice, 7.7, 7.20, 11.10, 11.13
 Kasteleyn matrix, 7.4, 7.8, 7.20
 Kohn gauge, 10.22
 Kohn-Luttinger effect, 11.20
 Kondo effect, 1.4–1.6, 3.36, 4.4, 5.24,
 6.6, 9.10
 charge analog, 6.6
 critical destruction of, 6.2, 6.17, 6.18,
 6.20
 Kondo insulator, 1.9–1.11
 Kondo lattice, 1.6–1.8
 Kondo model, 1.4, 4.3, 5.2, 6.16
 multichannel, 4.14
 pseudogap version, 6.17
 two-channel, 4.17
 Kondo resonance, 2.2
 narrowing, 4.20, 4.26
 Kondo scale, 5.8
 Kondo singlet, 1.5, 1.6
 Kondo temperature, 1.5, 4.4, 4.11
 Kotliar-Ruckenstein representation, 9.6,
 9.12
 Kramers-Kronig relations, 3.48
 Krylov space, 10.5, 13.16

L

Lanczos method, 10.5
 Lanczos procedure, 5.6
 Landau parameters, 9.24
 Landau-Fermi liquid, 1.7
 large degeneracy limit, 2.9
 large- N expansion, 1.11
 Lehmann representation, 5.22
 Lieb's theorem, 11.6

line graph, 11.10
linear response
 function, 3.49
 theory, 3.47
local moments, 1.2, 1.4
local vertex approximation, 3.35

M
magnetic impurities, 5.2
magnetic ion, 3.3
magnetic phases, 9.22, 9.25
Majumdar-Ghosh point, 7.16
Matsubara Green function, 3.44
Matsubara sums, 3.46
mean-field model
 Ising glass, 8.3, 8.17
 p-spin glass, 8.4, 8.22
 Potts glass, 8.4, 8.20
Mermin-Wagner theorem, 11.8
metal insulator transition, 15.23
mixed valence, 2.2
moment expansion, 10.11
Mott transition, 13.23
Mott-Hubbard insulator, 15.15
Mott-Hubbard transition, 9.12, 9.15
multi-band Hubbard model, 9.8, 9.15
multi-channel Anderson model, 5.32
multiferroism, 15.28

N
Nagaoka theorem, 11.9
Nambu formalism, 13.26
nematic, 7.24
non-crossing approximation, 2.2
NRG, 5.2
numerical renormalization group, 2.3, 5.2, 6.2
 bosonic chain formulation, 6.12
 bosonic star formulation, 6.11
 bosonic truncation effects, 6.12, 6.16
 choice of bosonic basis, 6.5, 6.12
 for Bose-Fermi models, 6.18
 for bosonic models, 6.10
 for models with local bosons, 6.5
 mass flow effects, 6.16
 time dependent, 5.30

O

ohmic two-state system, 5.9
one-particle Green's function, 2.4
operator
 time ordering, 8.15
orbital degeneracy, 2.8
order by disorder, 7.12
ordering temperature, 15.20
orthogonality catastrophe, 2.8

P
pair-field susceptibility, 14.10
paramagnetism of isolated ions, 3.20
Parisi, 8.6, 8.10
 solution, 8.15
particle-hole transformation, 11.4
Pauli paramagnetism, 3.12
percolation, 11.12
periodization, 13.11
perovskite, 15.2
phase separation, 9.29
polaron, 15.27
poor man's scaling, 3.39, 4.6
projection, 9.6, 9.9
pseudogap, 14.22
pyrochlore lattice, 7.5, 7.7, 7.11

Q

quadrupolar, 7.25
quantum cluster methods, 13.2, 14.2
quantum dimer model, 7.18
quantum fluctuations, 7.12
quantum impurity models, 5.3
quantum Monte Carlo, 14.6
 continuous time, 13.19
quantum quench, 2.8
quantum-to-classical mapping, 6.15

R

radial slave-boson fields, 9.8
random-phase approximation (RPA), 3.19
recursion method, 10.2
reduced density matrix, 5.22
renormalization, 1.4, 11.15
renormalization group, 4.5, 5.2
 equation, 11.19
 numerical, 4.12
replica
 overlap susceptibilities, 8.9

generations, 8.12
 hierarchy, 8.11, 8.14
 interactions, 8.8
 phase space, 8.8
 real, 8.8
 symmetry breaking, 8.10, 8.12
 continuous, 8.15
 trick, 8.5, 8.12
 residual entropy, 7.5
 resistivity minimum, 5.3
 resonating valence bond, 7.17
 response functions, 14.10
 RG transformation, 5.13
 Richardson-Lucy algorithm, 14.17
 RKKY interaction, 1.7, 1.8, 1.12, 1.25, 3.41
 running coupling constant, 4.9, 4.10
 ruthenates, 9.19, 9.21

S
 saddle-point approximation, 9.10, 9.12, 9.15, 9.19
 scaling law, 4.11
 screened exchange, 15.11
 screening, 4.10
 self-consistent Born approximation, 12.22
 self-energy, 2.5, 13.5
 self-energy functional theory, 13.20
 self-interaction, 15.5
 shared-memory, 10.15
 Shastry-Sutherland model, 7.17
 Sherrington-Kirkpatrick
 model, 8.3
 replicated, 8.9
 order parameter, 8.9
 solution, 8.5
 sign problem, 14.9, 14.20
 singular value decomposition, 14.19
 slave boson, 2.2
 constraint, 9.5–9.8
 representations, 9.4
 $SO(4)$ symmetry, 11.5
 solution
 asymptotic, 8.17
 degenerate, 8.21
 $SP(N)$, 1.24
 spectral density, 5.19

spectral function, 2.2, 10.8
 spectral representation, 13.3, 13.18
 spin fluctuations, 9.23
 spin gap, 7.15
 spin ice, 7.5
 spin models, 3.24
 spin waves, 12.18
 spin-boson model, 5.8, 6.9
 spin-rotation-invariant representation, 9.7, 9.23
 spiral phases, 9.22, 9.23, 9.25
 stability
 criteria, 8.12
 marginal, 8.16
 steepest-descent, 10.3
 Stoner instabilities, 3.15
 striped phases, 9.25
 strong-coupling fixed point, 4.10, 4.11
 $SU(2)$ spin symmetry, 11.3
 sudden approximation, 2.12
 superconductivity, 13.25, 14.23
 susceptibility, 3.49, 14.10
 symmetry-breaking fields, 8.8

T
 t - J model, 12.17
 thermal fluctuations, 7.14
 thermodynamic
 homogeneity, 8.7
 global, 8.9
 limit, 2.9, 8.7
 thermodynamic consistency, 14.26
 thermopower, 5.23
 topological Kondo insulator, 1.27
 transition
 discontinuous, 8.20
 transport properties, 5.23
 triangular lattice, 7.3, 7.8

U
 $U(1)$ spin liquid, 7.22
 ultrametric space, 8.13
 uniform density theorem, 11.14
 universality, 4.5

V
 valence photoemission, 2.2
 vanadates, 9.21

W

Wilson ratio, 4.13, 4.25

X

X-ray

absorption spectroscopy, 2.8

edge singularity, 2.8

Z

Z_2 spin liquid, 7.23

Zeeman interaction, 3.5

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, and 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, and 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