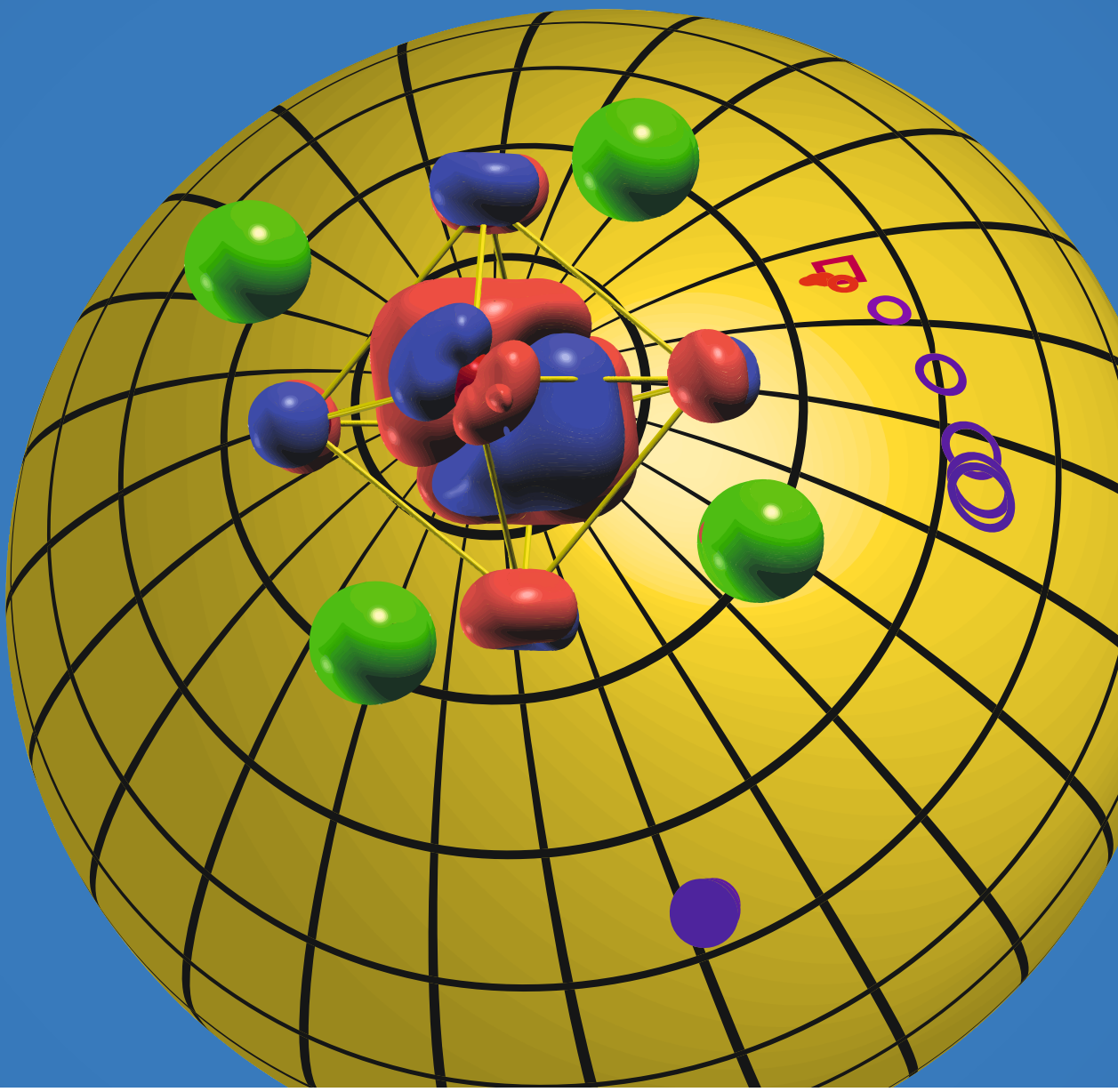


Dynamical Mean-Field Theory of Correlated Electrons

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

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 12

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

Lecture Notes of the Autumn School on
Correlated Electrons 2022

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

Dynamical Mean-Field Theory of Correlated Electrons

Autumn School organized by
the Institute for Advanced Simulation
at Forschungszentrum Jülich
4 – 7 October 2022

Schriften des Forschungszentrums Jülich
Modeling and Simulation Band / Volume 12

ISSN 2192-8525 ISBN 978-3-95806-619-9

Bibliographic information published by the Deutsche Nationalbibliothek.
Die 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:	Schloemer & Partner GmbH, Düren
Copyright:	Forschungszentrum Jülich 2022
Distributor:	Forschungszentrum Jülich Zentralbibliothek, Verlag 52425 Jülich Phone +49 (0)2461 61-5368 · Fax +49 (0)2461 61-6103 e-mail: zb-publikation@fz-juelich.de www.fz-juelich.de/zb

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation, Band / Volume 12

ISSN 2192-8525
ISBN 978-3-95806-619-9

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER) unter www.fz-juelich.de/zb/openaccess

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

Contents

Preface

1. Why Calculate in Infinite Dimensions?
   Dieter Vollhardt

2. Fermi Liquids
   Giovanni Vignale

3. The Physics of Quantum Impurity Models
   Jan von Delft

4. Machine Learning as a Solver for DMFT
   Cedric Weber

5. Quantum Monte Carlo Impurity Solvers
   Philipp Werner

6. Analytic Continuation of Quantum Monte Carlo Data
   Erik Koch

7. LDA+DMFT for Strongly Correlated Materials
   Alexander Lichtenstein

8. DMFT for Linear Response Functions
   Eva Pavarini

9. DFT+DMFT for Oxide Heterostructures
   Frank Lechermann

10. Dynamical Mean-Field Theory for Correlated Topological Phases
    Michael Potthoff

11. Beyond DMFT: Spin Fluctuations, Pseudogaps and Superconductivity
    Karsten Held

12. The GW+EDMFT Method
    Ferdi Aryasetiawan

13. DMFT and GW+DMFT for Systems out of Equilibrium
    Martin Eckstein

Index

Preface

Correlated systems are characterized by strong electron-electron interactions, resulting in many-body effects eluding a static mean-field description. They thus proved largely intractable until the development of dynamical mean-field theory (DMFT). In this approach, mapping the lattice onto an auxiliary impurity problem gives a local dynamical self-energy, which captures the essence of electronic correlations. DMFT is exact in the limit of infinite dimensions, and often provides an excellent approximation for real crystal structures. This crucial insight represented a paradigm shift in the study of correlations. Solving ever more complex impurity problems nowadays allows the simulation of real materials, making contact with experiment via the calculation of spectra, dynamical response functions, and non-equilibrium properties. Even subtle non-local effects can be captured using various approaches, including clusters of impurities or diagrammatic expansions. Thus DMFT finds application not only in correlated bulk systems but also in heterostructures, and can even be employed to understand the properties of topological phases of strongly correlated electron systems.

This year's school will introduce the concept of dynamical mean fields and explore how it can be used to understand the physics of real materials. Lectures will range from Fermi liquids and the limit of infinite dimensions to the physics of quantum impurities and their relation to the properties of correlated lattice systems.

A school of this size and scope requires backing from many sources. This is even more true during the Corona pandemics, which provided scores of new challenges. We are very grateful for all the practical and financial support we have received. The Institute for Advanced Simulation at the Forschungszentrum Jülich and the Jülich Supercomputer Centre provided the major part of the funding and were vital for the organization of the school as well as for the production of this book. The Institute for Complex Adaptive Matter (ICAM) supplied additional funds and ideas for successful online formats.

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 Grafische Betriebe for providing their expert support in producing the present volume on a tight schedule. We heartily thank our students and postdocs who helped with proofreading the manuscripts, often on quite short notice: Elaheh Adibi, Neda Samani, and Xue-Jing Zhang.

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

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

August 2022

1 Why Calculate in Infinite Dimensions?

Dieter Vollhardt
Center for Electronic Correlations and Magnetism
University of Augsburg

Contents

1 Dimensions in physics: From zero to infinity **2**
1.1 Integer and continuous spatial dimensions 2
1.2 Simplifications arising in infinite dimensions 3
1.3 Example: Derivation of Bohr's atomic model from the Schrödinger equation . . 4

2 Construction of classical mean-field theories in infinite dimensions **6**
2.1 Ising model 6
2.2 Ising model with random coupling: The spin glass problem 8
2.3 Hard-sphere fluid 10

3 Correlated electrons in solids **11**
3.1 From the Ising model to the Hubbard model 11
3.2 Characteristic features of the Hubbard model 13

4 Static mean-field theories of the Hubbard model **15**
4.1 Hartree approximation 15
4.2 Gutzwiller approximation 16

5 Lattice fermions in infinite dimensions **18**
5.1 Simplifications of diagrammatic quantum many-body theory 19
5.2 The Hubbard model in $d = \infty$ 20

6 Dynamical mean-field theory (DMFT) of correlated electrons and its applications **24**
6.1 The self-consistent DMFT equations 24
6.2 Mott transition, ferromagnetism, and topological properties 27
6.3 Correlated electrons in bulk materials, surfaces, and nanostructures 30

7 Beyond mean-field theory **31**
7.1 $1/d$ corrections 32
7.2 Beyond DMFT 33

8 Summary **34**

1 Dimensions in physics: From zero to infinity

In physics the term *dimension*[1] has a double meaning. On the one hand, it denotes the number of coordinates required to specify the location of a point within an object (or, more generally, a mathematical space). In particular, the dimension d of a point is $d = 0$, of a line $d = 1$, of a surface $d = 2$, and of a solid body $d = 3$. On the other hand, Maxwell introduced the dimension of a physical quantity, which is expressed in powers of fundamental units of mass, length, time, etc. [2]. Thereby it became possible to analyze the relationship between different physical quantities ("dimensional analysis"), which was later put on a firm theoretical basis by renormalization group theory.
In this lecture the term dimension refers to *spatial* dimensions.

1.1 Integer and continuous spatial dimensions

1.1.1 Integer dimensions

In theoretical physics the dimensionality of a microscopic model with finite-range interactions (or, for quantum-mechanical models in tight-binding approximation with finite-range hopping amplitude) in all directions, corresponds to the dimension of the space considered. This is particularly evident in the case of models on regular lattices, such as the hypercubic lattice, which is a straightforward generalization of the simple cubic lattice ($d = 3$) to lower ($d = 0, 1, 2$) or higher ($d = 4, 5, 6, ..., \infty$) dimensions. The usefulness of investigations of models in dimensions higher than the "real" dimension $d = 3$ is not self-evident. In their 1964 paper on the Ising model on general d-dimensional hypercubic lattices Fisher and Gaunt [3] found it necessary to write:

> *Of course the behavior of model physical systems in four or more space-like dimensions is not directly relevant to comparison with experiment! We can hope, however, to gain theoretical insight into the general mechanism and nature of phase transitions.*

The dimensionality of a lattice may be deduced from the total number of sites which can be visited in a walk of N steps away from a given site. The higher the dimension d, the more sites there are. For regular d-dimensional lattices and large N the number of sites in such a "sphere" of radius N is proportional to N^d. This can be formalized as follows [4]: if c_N denotes the number of sites which are N steps from a given site, one finds for hypercubic lattices

$$\lim_{N\to\infty} \frac{\ln c_N}{\ln N} = d, \tag{1}$$

where d is the dimensionality of the lattice. This relation, which also holds for regular two- and three-dimensional lattices, may therefore be used as a definition of the dimensionality d of a lattice-like structure. More specifically, for regular lattices both a dimension d and a *coordination number* Z (the number of nearest neighbors of a site) can be defined; Z is then

[1] **Dimension**: from Latin *dimensio* "a measuring", noun from past participle stem of *dimetiri* "to measure" [1].

determined by the dimension d and the lattice structure. In particular, for a hypercubic lattice one has $Z = 2d$. But there exist lattice-like structures without a relation between d and Z. A famous example is the Bethe lattice (which is not a lattice but an infinite, loop-free graph ("Cayley tree") without lattice periodicity), where each site has the same number of nearest neighbors Z. In this case the number of sites which are N steps away from a given site is $c_N = Z\big((Z-1)^N - 1\big)/(Z-2)$ for $Z \geq 2$, i.e., it increases exponentially with N. This is a faster increase than N^d for any finite d. According to the definition (1) the Bethe lattice is therefore effectively infinite-dimensional.

1.1.2 Continuous dimensions

Historically, the concept of spatial dimensions implied that the dimension is a fixed, integer number, e.g., $d = 3$. But, both from a mathematical and physical point of view this is an unnecessary restriction.[2] In particular, with the formulation of the renormalization group theory [8] it became apparent that it is useful to regard the spatial dimension d as a continuous parameter and to analytically continue d-dimensional integrals from integer to continuous values of d. This led to the "ε-expansion", with $\varepsilon = 4 - d$, which for continuous values of d close to the critical dimension $d = 4$ of the Ginzburg-Landau theory can be arbitrarily small ($\varepsilon \ll 1$). Expansions in ε then make it possible to perform explicit perturbative calculations around $d = 4$, as expressed by the striking title "Critical Exponents in 3.99 Dimensions" of a paper by Wilson and Fisher [9].

1.2 Simplifications arising in infinite dimensions

The mathematical equations describing classical or quantum-mechanical many-body systems can almost never be solved exactly in dimension $d = 3$. In many problems there does not even exist a small physical parameter ("control parameter"), in which an expansion could be performed (if such an expansion is tractable at all). One way out is to study problems in d=1, where mathematical solutions are more readily available. Such investigations have indeed led to a wealth of insight. However, one-dimensional systems are quite special and do not describe behavior which is characteristic for systems in $d > 2$, e.g., thermal phase transitions. An alternative is to go in the opposite direction. Indeed, the coordination number of a three-dimensional lattice can already be quite large, e.g., $Z = 6$ for a simple cubic lattice, $Z = 8$ for a bcc lattice and $Z = 12$ for an fcc-lattice. It is then interesting to check whether the limit $Z \to \infty$ leads to some simplifications. Such investigations do not go far back in time. In fact, Z was originally regarded as a measure of the *range* of the interaction between spins in the Ising model, and thus of the number of spins in the range of the interaction [10]. In this case the limit $Z \to \infty$

[2] For example, there exists a class of fractal objects [5] ("fractals"), whose defining property is their scale invariance or self-similarity. A quantitative measure of a fractal is its dimension ("fractal dimension") which is, in general, not an integer. For example, a rocky coast line has a fractal dimension between $d = 1$ and $d = 2$. It is even possible to design and characterize electrons in fractal geometries [6]. A discussion of the notion of dimension in a more general context can be found in ref. [7].

describes an infinitely long-ranged interaction.[3] Since a particle or spin then interacts with infinitely many other particles or spins (which are all "neighbors", i.e., the system has 'infinite connectivity"'), this limit was originally referred to as "limit of high density" [10] and later as "limit of infinite dimensions" [3]. Starting with Fisher and Gaunt [3] the Ising model and other classical models were investigated on general d-dimensional hypercubic lattices. Today Z denotes the coordination number, i.e., the number of nearest neighbors.

1.2.1 Infinite dimensions and mean-field behavior

In the statistical theory of classical and quantum-mechanical systems mean-field theories play an important role since they provide an approximate understanding of the properties of a model. While in the full many-body model a particle or spin experiences a complicated, fluctuating field generated by the other particles or spins, a mean-field theory introduces an average field ("mean field") instead. Usually the full interaction problem then reduces to an effective, single-particle problem, which is described by a self-consistent field theory.
A mean-field theory can often be derived systematically by increasing the range of an interaction (e.g., the coupling between spins) or the size of a parameter (e.g., the spin or orbital degeneracy N, the spatial dimension d, or the coordination number Z) to infinity. In particular, the mean-field theory of the Ising model (the so called "molecular" or "Weiss" mean-field theory, which can be derived by the Bragg-Williams approximation [12]), becomes exact in the limit $d \to \infty$; see section 2.1.[4] This also answers the question serving as the title of this lecture *Why calculate in infinite dimensions?* Namely, equations which depend on the dimension d and which are too complicated to be solved in $d = 3$, often simplify in the limit $d \to \infty$ to such a degree that they can be solved exactly.[5] At best this leads to an approximate solution which retains characteristic features of the problem in $d < \infty$ and provides insights into the (unknown) solution in $d = 3$. Several examples will be discussed in this lecture. To this end I will address mainly many-body systems (classical and quantum) of condensed matter physics and the simplifications arising in $d = \infty$. However, to demonstrate the usefulness of this approach I first discuss the solution of a famous quantum-mechanical *two*-body problem, the hydrogen atom, in the limit $d \to \infty$.

1.3 Example: Derivation of Bohr's atomic model from the Schrödinger equation

Bohr's model of the hydrogen atom is semiclassical: it postulates that in the ground state the electron moves in a circular orbit around the proton, like a planet around the sun in classical mechanics, and assumes in addition that the angular momentum and the energy of the electron are

[3] This limit is even useful for one-dimensional models. For example, for a solvable one-dimensional particle model it was shown [11] that in this limit the equation of state reduces to the van der Waals equation.

[4] Due to the simplicity of the Ising model the limit of infinitely long-ranged spin coupling J and of infinite dimensions d both yield the same mean-field theory. However, for more complicated models, in particular quantum models with itinerant degrees of freedom, this will in general not be the case.

[5] To avoid trivial results or divergencies when taking the limit $d \to \infty$, an appropriate scaling of parameters or coupling constants is necessary, as will be discussed later.

quantized. Although this model was extremely useful in the pre-history of quantum mechanics[6] its connection to the "proper" quantum mechanics was unclear for a long time. Today we understand that the main features of Bohr's model may be derived by solving the hydrogen atom with the Schrödinger equation in infinite dimensions. This approach was introduced by Witten [15] to demonstrate the usefulness of expansions in the inverse of some large parameter for deriving approximate solutions of hard problems such as quantum chromodynamics. The results for the hydrogen problem (which is, of course, exactly solvable in $d = 3$) showed that even atomic physics greatly simplifies as one goes from $d = 3$ to $d = \infty$, and that expansions in $1/d$ provide valuable approximations for problems which are no longer exactly solvable [15, 16].
Following ref. [16] the radial Schrödinger equation in d-dimensional spherical coordinates for the electron of a hydrogen atom with infinite proton mass and orbital angular momentum l is given by

$$\left(-\frac{1}{2}\left(\frac{\mathrm{d}^2}{\mathrm{d}\varrho^2}+\frac{d-1}{\varrho}\frac{\mathrm{d}}{\mathrm{d}\varrho}\right)+\frac{l(l+d-2)}{2\varrho^2}-\frac{1}{\varrho}\right)\psi=\varepsilon\psi, \tag{2}$$

where we used atomic units, i.e., $\varrho = r/a_0$ is the radial distance in units of the Bohr radius $a_0 = \hbar^2/(e^2m)$ and $\varepsilon = E/(e^2/a_0)$ is the energy of the electron in units of Hartree.[7] To eliminate the first-order derivative in (2) the wave function is rescaled as $\varphi = \varrho^{(d-1)/2}\psi$, whereby (2) reduces to

$$\left(-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\varrho^2}+\frac{\Lambda(\Lambda+1)}{2\varrho^2}-\frac{1}{\varrho}\right)\varphi=\varepsilon\varphi. \tag{3}$$

The d-dependence now enters only in the centrifugal potential, where $\Lambda = l+(d-3)/2$ replaces the usual orbital angular momentum. As a consequence, the main quantum number n in $d = 3$ becomes $n + (d-3)/2$ in $d \geq 3$. To make sure that the centrifugal term in (3) remains finite in the limit $d \to \infty$ another dimensional rescaling of the radial coordinate and the energy is performed as $\mathcal{R} = (\frac{d-1}{2})^{-2}\varrho$ and $\mathcal{E} = (\frac{d-1}{2})^2\varepsilon$, respectively. This brings (3) into the form (we only discuss the ground state, $l = 0$)

$$\left(-\frac{2}{(d-1)^2}\frac{\mathrm{d}^2}{\mathrm{d}\mathcal{R}^2}+\frac{1}{2}\,\frac{d-3}{d-1}\,\frac{1}{\mathcal{R}^2}-\frac{1}{\mathcal{R}}\right)\varphi=\mathcal{E}\varphi. \tag{4}$$

Since the form of (4), which determines the radial probability amplitude φ in $d \geq 3$, is the same as that in $d = 3$, all properties of the d-dimensional hydrogen atom can be related to those of the well-known solution in $d = 3$. In particular, one finds that, as in $d = 3$, the radial probability $|\varphi|^2$ is maximal at the distance $\mathcal{R} = 1$, corresponding to $r = (\frac{d-1}{2})^2 a_0$. As d increases the width of the electron distribution *decreases*, i.e., the electron is confined to a thinner and thinner spherical shell of radius $\mathcal{R}$ around the nucleus. In addition, the factor $(\sin\vartheta)^{d-2}$ in the d-dimensional volume element restricts the polar angle to $\vartheta \to \pi/2$ for $d \to \infty$. Hence for large d the electronic probability distribution approaches the shape of a planar, circular orbit as described by the Bohr model!

[6] Even today the Bohr model plays a useful role, e.g., in the description of highly excited Rydberg atoms [13] and cavity quantum electrodynamics [14].

[7] Note that the Coulomb potential of the hydrogen atom is that in $d = 3$, not in general d (in which case it would have a $1/\varrho^{d-2}$ dependence for $d > 2$).

In $d = \infty$ itself the derivative in (4), which originates from the kinetic energy of the electron, is suppressed and the equation reduces to the classical energy expression

$$\mathcal{E}(\mathcal{R}) = \frac{1}{2\mathcal{R}^2} - \frac{1}{\mathcal{R}}. \tag{5}$$

This is the energy of the hydrogen atom in Bohr's model. Its minimum determines the radial coordinate and energy of the electron in $d = \infty$ as $\mathcal{R} = 1$, $\mathcal{E} = -1/2$. Hence in infinite dimensions the electron no longer orbits around the nucleus, but is located at the minimum of the effective potential. Quantum fluctuations are then completely quenched.[8] For that reason the electron does not emit radiation (photons) as had to be assumed by Bohr in 1913. This assumption is found to be justified in $d = \infty$.

The large-d approach to the Schrödinger equation outlined above can also be applied to more complicated problems of atomic physics, such as the helium atom or the hydrogen molecule. It turns out that perturbative calculations in the small parameter $1/d$ yield surprisingly accurate results for the two-electron bond [16].

As explained above, the solution of classical and quantum-mechanical many-body problem obtained in the limit $d = \infty$ ("mean-field theory") may provide important physical insights, which are not available otherwise. Since the mean-field theory of the classical Ising model is particularly simple and its validity in $d = \infty$ follows directly from the law of large numbers, we start with this famous example, then discuss the simplifications arising in $d = \infty$ in other classical model systems, and finally turn to interacting fermions and derive the dynamical mean-field theory (DMFT) of correlated electrons.

2 Construction of classical mean-field theories in infinite dimensions

2.1 Ising model

In 1906 Weiss [17] introduced a model of magnetic domains, where he assumed that the alignment of the "elementary magnets" in each domain is caused by some sort of "molecular field", today referred to as "Weiss mean field". But how does this molecular field arise in the first place? To answer this question and to explain ferromagnetism in three-dimensional solids from a truly microscopic point of view, Ising investigated a minimal microscopic model of interacting classical spins with a *non-magnetic* interaction between neighboring elementary magnets [18], which had been proposed to him by his thesis advisor Lenz in 1922. The Hamiltonian function for the Ising model with coupling J_{ij} between two spins at lattice sites $\boldsymbol{R}_i$, $\boldsymbol{R}_j$ is given by[9]

$$H = -\frac{1}{2} \sum_{\boldsymbol{R}_i, \boldsymbol{R}_j} J_{ij}\, S_i S_j. \tag{6}$$

[8] This does not violate the Heisenberg uncertainty relation since in the product of length and momentum the dimensional scaling factors cancel.

[9] This notation of the Ising model, as it is used today, was actually introduced by Pauli at a Solvay conference in 1930; a detailed discussion of the history of the Ising model and its many applications can be found in ref. [19].

In particular, if the coupling is restricted to nearest-neighbor spins it takes the form

$$H = -\frac{1}{2}J \sum_{\langle \boldsymbol{R}_i, \boldsymbol{R}_j \rangle} S_i S_j, \tag{7}$$

where we assume ferromagnetic coupling ($J > 0$) and $\langle \boldsymbol{R}_i, \boldsymbol{R}_j \rangle$ indicates summation over all nearest-neighbor sites (the factor $1/2$ prevents double counting of sites). This can also be written as

$$H = -\sum_{\boldsymbol{R}_i} h_i S_i, \tag{8}$$

where now every spin S_i interacts with a site-dependent, i.e., locally fluctuating, field

$$h_i = J \sum_{\boldsymbol{R}_j}^{(i)} S_j \tag{9}$$

generated by the coupling of the spin S_i to its neighboring sites; here the superscript (i) on the summation symbol indicates summation over the Z nearest-neighbor sites of $\boldsymbol{R}_i$.

2.1.1 Weiss mean field

In the Weiss mean-field theory the interaction of a spin with its local field in (8) is decoupled (factorized), i.e., h_i is replaced by a mean field $h_{\rm MF}$, which leads to the mean-field Hamiltonian

$$H^{\rm MF} = -h_{\rm MF} \sum_{\boldsymbol{R}_i} S_i + E_{\rm shift}. \tag{10}$$

Now a spin S_i interacts only with a global field $h_{\rm MF} = JZS$ (the "molecular" or "Weiss" field), where $S \equiv \langle S_i \rangle = (1/L) \sum_{i=0}^{L} S_i$ is the average value of S_i, $E_{\rm shift} = LJZS^2/2$ is a constant energy shift, and L is the number of lattice sites of the system.
Next we show that in infinite dimensions or for infinite coordination number Z this decoupling arises naturally. First we have to rescale the coupling constant J as

$$J = \frac{J^*}{Z} \ , \ J^* = \text{const}, \tag{11}$$

so that $h_{\rm MF}$ and the energy (or the energy density in the thermodynamic limit) remain finite in the limit $Z \to \infty$. Writing $S_i = S + \delta S_i$, where δS_i is the deviation of S_i from its average S, (9) becomes $h_i = J^*(S + \Delta S_i)$, where

$$\Delta S_i = \frac{1}{Z} \sum_{\boldsymbol{R}_j}^{(i)} \delta S_j \tag{12}$$

is the sum over the fluctuations δS_j of the Z nearest-neighbor spins of S_i per nearest neighbor. These fluctuations are assumed to be uncorrelated, i.e., random. The law of large numbers then implies that the sum increases only as $\sqrt{Z}$ for $Z \to \infty$, such that ΔS_i altogether *decreases* as

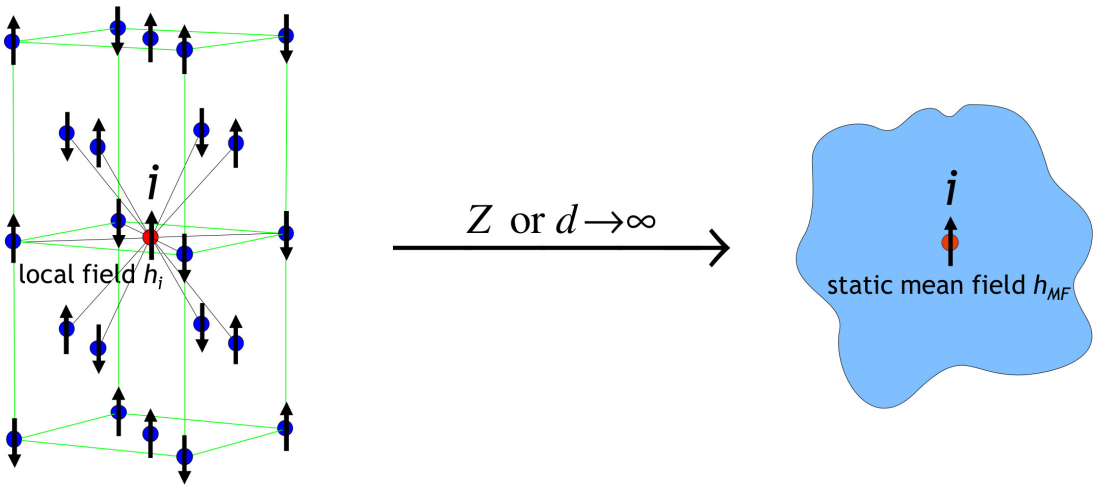


Fig. 1: *Already in three dimensions ($d = 3$) the coordination number Z of a lattice can be quite large, as in the face-centered cubic lattice shown on the left, where $Z = 12$. In the limit $Z \to \infty$, or equivalently $d \to \infty$, the Ising model effectively reduces to a single-site problem where the local field h_i is replaced by a global ("molecular") mean field $h_{\rm MF}$.*

$1/\sqrt{Z}$ in this limit. As a consequence the local field h_i can indeed be replaced by its mean $h_{\rm MF}$ (central limit theorem). Hence the Hamiltonian function (6) becomes purely local

$$H^{\rm MF} = \sum_{\boldsymbol{R}_i} H_i + E_{\rm shift}, \tag{13}$$

where $H_i = -h_{\rm MF} S_i$. Thereby the problem reduces to an effective single-site problem (see Fig. 1). We note that S corresponds to the magnetization m of the system ($S \equiv m$). In the paramagnetic phase, where $m = 0$, the mean field $h_{\rm MF}$ vanishes; hence (10) and (13) are only non-trivial in the presence of ferromagnetic order. The magnetization m is obtained from the partition function of the mean-field Hamiltonian (10) as $m = \tanh(\beta h_{\rm MF})$ where $\beta = 1/(k_B T)$ [12]. The self-consistency condition $h_{\rm MF} = JZS \equiv J^* m$ then yields the well-known self-consistent equation for the magnetization m as

$$m = \tanh\bigl(\beta J^* m\bigr). \tag{14}$$

The Weiss mean-field theory is seen to become exact in the limit of infinite coordination number Z or dimension d. In this case $1/Z$ or $1/d$ serve as a small parameter which can be used, in principle, to improve the mean-field theory systematically (see section 7.1). This mean-field theory contains no unphysical singularities, is applicable for all values of the input parameters (temperature and/or external magnetic field) and is often viewed as a prototypical mean-field theory in statistical mechanics.

2.2 Ising model with random coupling: The spin glass problem

The term *spin glass* was introduced in the early 1970s to characterize the behavior of certain disordered magnetic systems, i.e., alloys of a non-magnetic material with a few percent of randomly distributed magnetic impurities, such as manganese in zinc oxide. At low temperatures

a phase transition occurs, where the magnetic moments "freeze" into a state in which the spin orientations are not aligned, but randomly oriented. The disorder is due to the random local distribution of the spins and the fact that their coupling can be ferro- or antiferromagnetic. Theoretical investigations of this experimental finding started in 1975 with Edwards and Anderson [20] who proposed a classical Heisenberg model with random finite-range couplings J_{ij}, which can be positive or negative. In particular, they constructed an order parameter for the spin glass phase and introduced the famous "replica trick". Namely, to compute the averaged free energy they considered n copies ("replicas") of the system, calculated the corresponding partition function $\mathcal{Z}^n$ to obtain $\ln \mathcal{Z} = \lim_{n\to 0} \left((\mathcal{Z}^n - 1)/n\right)$ by taking the limit $n \to 0$, and then averaged $\ln \mathcal{Z}$ over the randomness. Later in the year Sherrington and Kirkpatrick [21] introduced a mean-field version of the Edwards-Anderson model by assuming Ising spins instead of Heisenberg spins, but with random *infinite-range* spin couplings J_{ij} in (6), whereby all spins are equally coupled. For random J_{ij} with zero mean ($\overline{J}_{ij} = 0$) but $\overline{J_{ij}^2} \neq 0$ the spin couplings need to be scaled as $J_{ij} \to J_{ij}^*/\sqrt{L}$, where L is the number of lattice sites, to keep the energy density of the system finite in the thermodynamic limit $L \to \infty$. As mentioned in section 1.2 the mean-field theory of the Ising model derived for infinite range-coupling is equivalent to that for finite-range coupling in infinite dimensions d or for infinite coordination number Z. In the latter case, and for spins with random nearest-neighbor couplings J_{ij}, the scaling

$$J_{ij} = J_{ij}^*/\sqrt{Z} \tag{15}$$

applies; in both cases the model has infinite connectivity. Thereby a mean-field investigation of the spin glass problem became possible which, however, resulted in a negative entropy at low temperatures [21]. This non-physical result was found to originate from the assumption of "replica symmetry" [22]. The question of how to break this symmetry ("replica symmetry breaking") was answered by Parisi in 1979 [23], who realized that there must be an infinite number of order parameters in the spin glass phase. His analytic procedure led to a consistent and stable mean-field solution of the spin glass problem (for details see ref. [24]).[10]

The mean-field theory of the spin-glass problem illustrates particularly well that, in spite of the simplifications introduced by a mean-field theory, the solution of this mean-field theory can still be extraordinarily complicated. This will become evident again in the context of the dynamical mean-field theory (DMFT) of correlated electrons (see section 6.1).

The limit of infinite dimensions and the methods of solution developed in the context of the spin-glass mean-field theory have also been very useful in the study of classical liquids and amorphous systems, such as glasses and granular matter, where calculations are notoriously difficult. This made it possible to explore not only the thermodynamic properties but even the general dynamics of liquids and structural glasses, including the famous glass transition and the rheology of glasses, on a mean-field level [25, 26].

[10] Parisi was awarded one half of the Nobel Prize in Physics 2021 "for groundbreaking contributions to our understanding of complex systems". The Nobel Committee highlighted the great influence which Parisi's concept of broken replica symmetry and his method of solution of the mean-field theory of the spin-glass problem had on the statistical physics of systems exhibiting multiple equilibria.

2.3 Hard-sphere fluid

As mentioned above, the limit of infinite spatial dimensions leads to theoretical simplifications not only in the study of lattice systems (in which case the coordination number Z tends to infinity) but also in the case of systems in the continuum. This may be demonstrated by calculating the equation of state of a classical dilute hard-sphere fluid[11]; here we follow ref. [27]. While at low densities the virial series, i.e., the expansion of the equation of state in terms of the density, is a very useful technique to examine thermodynamic properties of a fluid in $d = 3$, this approach breaks down at high densities. By contrast, these calculations are possible in infinite spatial dimensions[12] since in this limit the virial expansion $\beta PV = N \sum_{l=0}^{\infty} B_{l+1} n^l$ terminates after the second term, such that only the second virial coefficient B_2 needs to be calculated (the first virial coefficient takes the value $B_1 = 1$ and characterizes the classical ideal gas). Let us consider a classical fluid of hard spheres in d dimensions as the simplest case. An interaction parameter does not explicitly enter in this model, since the interaction is either zero if the hard spheres are not in contact, or infinite if they touch. The only coupling parameter is the radius a of the hard spheres. The grand partition function $\mathcal{Z} = \exp(\beta PV)$ is expanded in powers of z/λ^d, where $z = \exp(\beta\mu)$ is the fugacity and $\lambda = \hbar\sqrt{2\pi\beta/m}$ is the thermal wave length. Each term of this "Mayer cluster expansion" can be expressed in terms of graphs [12]. Due to the linked cluster theorem $\mathcal{Z}$ reduces to a sum of connected graphs, $\mathcal{Z}_c$, with $\mathcal{Z} = \exp(\mathcal{Z}_c - 1)$, which can be written as

$$\mathcal{Z}_c = \sum_{l=0}^{\infty} b_l \left(\frac{z}{\lambda^d}\right)^l . \tag{16}$$

Here the coefficients b_l are determined by all possible connected graphs multiplied by a weight which is determined by configurational integrals over the volume. The typical scale of volume is the volume of a hard sphere with radius a in d dimensions, $v_d(a) = v_d(1)a^d$, where $v_d(1) = \pi^{d/2}/\Gamma(1+d/2)$ is the volume of a d-dimensional unit sphere.

2.3.1 Equation of state

By differentiating the grand potential $\Omega(T, V, \mu) = -\beta^{-1} \ln \mathcal{Z} = -PV$ with respect to the volume V or the chemical potential μ, one obtains the pressure P or particle number N, respectively, from which the equation of state is obtained by eliminating z/λ^d. In the limit $d \to \infty$ the evaluation of the coefficients b_l simplifies significantly because for large d every loop in a connected graph is suppressed exponentially by a factor $(\sqrt{3}/2)^d/\sqrt{d}$. This is due to the fact that $v_d(1)$ decreases exponentially for $d \to \infty$, which implies that the cross section of the spheres vanishes. This leaves only loop-less ("tree") diagrams. As a consequence, only the second virial coefficient, given by $B_2 = (1/2)v_d(a)$ remains, while all higher virial coefficients vanish

[11] A fluid is a substance which continuously deforms under tangential stress. Any liquid or gas is therefore a fluid (but not all fluids are liquids).

[12] In the following P is the pressure, V the volume, N the particle number, $n = N/V$ the number density, μ the chemical potential, and m the mass of the hard sphere particles.

in the limit $d \to \infty$.[13] However, B_2 has also a d dependence and vanishes for $d \to \infty$ since the volume of a d-dimensional unit sphere $v_d(1)$ goes to zero in this limit. Therefore a scaling of B_2 with the dimension d is required to reach a proper mean-field limit $d \to \infty$.[14] For this purpose we scale the hard-sphere radius a, the only coupling parameter in the problem, as $a = d^\nu a^*$, with a^* as the scaled radius. The exponent ν has to be determined such that B_2 remains finite for $d \to \infty$, i.e., $v_d(a) = v_d(d^\nu a^*) \equiv v(a^*) = \text{const}$. With $\Gamma(1+d/2) \sim (d/2)^{d/2}$ for $d \to \infty$ one finds $\nu = 1/2$, which implies the scaling

$$a \to \sqrt{d} a^*, a^* = \text{const} \tag{17}$$

for $d \to \infty$. The shrinking of the volume (and thus of the cross section) of a d-dimensional unit sphere for increasing d is thus compensated by a corresponding increase of the radius of the hard sphere. Thereby B_2 becomes $B_2 = (1/2)\, v(a^*)$. The mean-field equation of state of a classical hard sphere fluid then takes the form

$$\frac{PV}{k_B T} = N \left(1 + \frac{1}{2} v(a^*)\, n \right). \tag{18}$$

The interactions in the hard-sphere fluid are here described entirely by the second virial coefficient, which characterizes the interaction between two particles.

3 Correlated electrons in solids

3.1 From the Ising model to the Hubbard model

One of the most striking solid-state phenomena is *ferromagnetism* as observed in magnetite (Fe_3O_4) and elemental iron (Fe). How can ferromagnetism by explained? A crucial first step in the development of a microscopic theory of ferromagnetism was the formulation of the Ising model [18] discussed in Sect. 2.1. Ising solved the model in $d = 1$, found that a transition to a ferromagnetic phase does not occur, and concluded (incorrectly) that this holds also in $d = 3$. The Ising model is a classical spin model. But it had already been shown by Bohr (1911) and van Leeuwen (1919) that magnetism is a quantum effect. Therefore another important step in the development of a theory of ferromagnetism was Heisenberg's formulation of a quantum spin model in 1928 [29]. With this model is was possible to explain the origin of the Weiss molecular field as the result of a quantum-mechanical exchange process. Clearly, a model of localized spins cannot explain ferromagnetism observed in $3d$ transition *metals* such as iron, cobalt and nickel. As pointed out by Bloch [30] in 1929 an appropriate model had to include the mobile nature of the electrons, i.e., their wave character, which, in a solid, leads to electronic bands. The conditions for ferromagnetism which he obtained for free electrons where quite unrealistic.

[13] This result for B_2 is plausible since B_2 has the dimension of a volume and $v_d(a)$ is the only characteristic volume in the problem.

[14] To this end a scaled (dimensionless) "density per dimension" ϱ was defined through $n v_d(a) = \varrho^d$ in ref. [28]. However, thereby thermodynamic variables are scaled rather than the coupling parameter a of the system, in contrast to the general construction principle of mean-field theories.

Obviously one had to go beyond free electrons and also take the mutual interaction between their electric charges into account. This immediately leads to an enormously difficult many-body problem, which is made particularly complicated by the fermionic nature of the electrons, their long-range Coulomb interaction, their high density in the metallic state, and the presence of a periodic lattice potential. Attempts by Slater [31] in 1936 to explain ferromagnetism in Ni by including the Coulomb interaction within Hartree-Fock theory were also not successful.

3.1.1 Electronic correlations

It became clear that one had to include *correlations*, i.e., effects of the interaction between electrons which cannot be explained within a single-particle picture and therefore go beyond the physics described by factorization approximations such as Hartree or Hartree-Fock mean-field theory. Wigner [32] was apparently the first who tried to calculate the contribution of the mutual electronic interaction to the ground state energy beyond the Hartree-Fock result, which he referred to as "correlation energy".[15] Such an approach is faced by two intimately connected problems: the need for a sufficiently simple model of correlated electrons which unifies the competing approaches by Heisenberg and Bloch (namely the picture of localized and itinerant electrons, respectively), and its solution within a more or less controlled approximation scheme. Progress in this direction was slow.[16] A microscopic model of ferromagnetism in metals ("band ferromagnetism") did not emerge until 1963, when a model of correlated lattice electrons was proposed independently by Gutzwiller [44], Hubbard [45], and Kanamori [46] to explain ferromagnetism in $3d$ transition metals. Today this model is referred to as "Hubbard model" [47].

[15] Electronic correlations play a fundamental role in modern condensed matter physics. They are known to be strong in materials with partially filled d and f electron shells and narrow energy bands, as in the $3d$ transition metals or the rare–earths and their compounds, but they also occur in materials without rare-earth or actinide elements, such as "Moiré" heterostructures of van der Waals materials, e.g., twisted bilayers of graphene [33] and two-layer stacks of TaS_2 [34]. Electronic correlations in solids lead to the emergence of complex behavior, resulting in rich phase diagrams. In particular, the cooperation between the various degrees of freedom of the correlated electrons (spin, charge, orbital momentum) on a lattice with specific dimensionality and topology result in a wealth of correlation and ordering phenomena such as heavy fermion behavior [35], high temperature superconductivity [36], colossal magnetoresistance [37], Mott metal-insulator transitions [38], and Fermi liquid instabilities [39]. The surprising discovery of a multitude of correlation phenomena in Moiré heterostructures [33] reinforced this interest in correlation physics. The unusual properties of correlated electron systems are not only of interest for fundamental research but may also be relevant for technological applications. Indeed, the exceptional sensitivity of correlated electron materials upon changes of external parameters such as temperature, pressure, electromagnetic fields, and doping can be employed to develop materials with useful functionalities [40]. In particular, Moiré heterostructures may enable "twistronics", a new approach to device engineering [41]. Consequently there is a great need for the development of appropriate models and theoretical investigation techniques which allow for a comprehensive, and at the same time reliable, exploration of correlated electron materials [42].

[16] One reason for the slow development was that in the nineteen-thirties and forties nuclear physics attracted more attention than solid-state physics, with a very specific focus of research during the 2nd World War. But apart from that, the sheer complexity of the many-body problem itself did not allow for quick successes. High hurdles had to be overcome, both regarding the development of appropriate mathematical techniques (field-theoretic and diagrammatic methods, Green functions, etc.) and physical concepts (multiple scattering, screening of the long-range Coulomb interaction, quasiparticles and Fermi liquid theory, electron-phonon coupling, superconductivity, metal-insulator transitions, disorder, superexchange, localized magnetic states in metals, etc.). A discussion of the many-body problem and of some of the important developments up to 1961 can be found in the lecture notes and reprint volume by Pines [43].

The single-band Hubbard model is the minimal microscopic lattice model of interacting fermions with local interaction.[17] The Hamiltonian consists of the kinetic energy $\hat{H}_0$ and the interaction energy $\hat{H}_{\rm int}$ (in the following operators are denoted by a hat):

$$\hat{H} = \hat{H}_0 + \hat{H}_{\rm int}, \tag{19a}$$

$$\hat{H}_0 = \sum_{\boldsymbol{R}_i,\boldsymbol{R}_j} \sum_{\sigma} t_{ij} \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} = \sum_{\boldsymbol{k},\sigma} \varepsilon_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k}\sigma}, \tag{19b}$$

$$\hat{H}_{\rm int} = U \sum_{\boldsymbol{R}_i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \equiv U \hat{D}. \tag{19c}$$

Here $\hat{c}^\dagger_{i\sigma}(\hat{c}_{i\sigma})$ are creation (annihilation) operators of fermions with spin σ at site $\boldsymbol{R}_i$ (for simplicity denoted by i), $\hat{n}_{i\sigma} = \hat{c}^\dagger_{i\sigma}\hat{c}_{i\sigma}$, and $\hat{D} = \sum_{\boldsymbol{R}_i} \hat{D}_i$ is the operator of total double occupation of the system with $\hat{D}_i = \hat{n}_{i\uparrow}\hat{n}_{i\downarrow}$ as the operator of double occupation of a lattice site i. The Fourier transform of the kinetic energy in (19b), where t_{ij} is the amplitude for hopping between sites i and j, defines the dispersion $\varepsilon_{\boldsymbol{k}}$ and the momentum distribution operator $\hat{n}_{\boldsymbol{k}\sigma}$. In the following the hopping is restricted to nearest-neighbor sites i and j, such that $-t \equiv t_{ij}$. A schematic picture of the Hubbard model is shown in Fig. 2.

3.2 Characteristic features of the Hubbard model

In the Hubbard model the Coulomb interaction between two electrons is assumed to be so strongly screened that it can be described by a purely local interaction which occurs only *on* a lattice site.[18] Due to the Pauli principle a local interaction is only possible if the two electrons have opposite spin.[19] This distinguishes a local interaction from other model interactions since it has no classical counterpart. The interaction is therefore completely independent of the lattice structure and spatial dimension of the system. The kinetic energy $\hat{H}_0$ is diagonal in momentum space and reflects the wave nature of the electrons, while the interaction energy $\hat{H}_{\rm int}$ is diagonal in position space and characterizes their particle nature. In view of the uncertainty principle the two terms are therefore extremely "quantum incompatible".

[17] It is interesting to note that Anderson had introduced the main ingredient of the Hubbard model, namely a local interaction between spin-up and spin-down d electrons with strength U, already in his 1959 paper on the theory of superexchange interactions [48] and, even more explicitly in his 1961 paper on localized magnetic states in metals, where he formulated a model of s and d electrons referred to today as "single impurity Anderson Model" (SIAM) or "Anderson impurity model" (AIM) [49]. The latter paper inspired Wolff [50] to study the occurrence of localized magnetic moments in dilute alloys with a single-band model of noninteracting d electrons which interact on a single site. In this sense the Hubbard model could be called "periodic Wolff model" in analogy to the standard terminology "periodic Anderson model", which generalizes the AIM by extending the interaction to all sites of the lattice. Apparently Gutzwiller, Hubbard and Kanamori did not know (or were not influenced) by the earlier work of Anderson and Wolff; at least they did not reference these papers.

[18] Thereby the Hubbard model applies particularly well to cold fermionic atoms in optical lattices where the bare interaction is indeed extremely short-ranged [51].

[19] It appears as if the interaction between the electrons was spin-dependent. But the Coulomb interaction is, of course, a spin-independent two-body interaction. The fact that the operators in (19c) contain spin indices is merely a consequence of the occupation-number formalism.

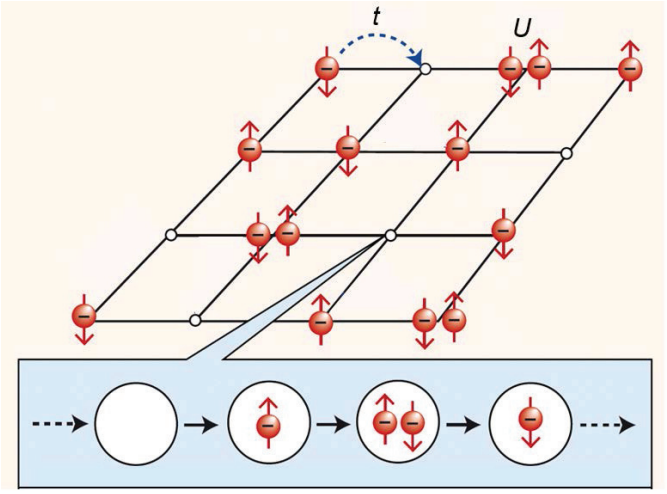


Fig. 2: *Schematic illustration of interacting electrons in a solid described by the Hubbard model. The ions enter only as a rigid lattice, here represented by a square lattice ($d = 2$, $Z = 4$). The electrons, which have mass, negative charge, and spin ($\uparrow$ or $\downarrow$), are quantum particles which "hop" from one lattice site to the next with a hopping amplitude t. Together with the lattice structure this determines the band structure of the non-interacting electrons. The quantum dynamics leads to fluctuations in the occupation of lattice sites as indicated by the sequence: a lattice site can either be unoccupied, singly occupied ($\uparrow$ or $\downarrow$), or doubly occupied. When two electrons meet on a lattice site, which is only possible if they have opposite spins because of the Pauli exclusion principle, they encounter a local interaction U.*

The physics described by the Hubbard model is very different from that of electrons with a long-range Coulomb interaction in the continuum. Therefore the Hubbard model is far from obvious. Its formulation required fundamentally new insights into the nature of the many-body problem of interacting fermions (see footnote 16). In particular, screening is a basic ingredient of the many-body problem of metals.

A direct interaction between electrons with equal spin direction, e.g., on neighboring sites, is not described by the model, but can be easily included. Similarly the model can be generalized to more than one band. Indeed, in Wannier basis the Hubbard model can be derived systematically from a general Hamiltonian of interacting electrons including the kinetic energy, ionic potential $U_{\rm ion}(\boldsymbol{r})$, and two-body Coulomb interaction $V_{\rm ee}(\boldsymbol{r}-\boldsymbol{r}')$ [45].[20]

Since the single-band Hubbard model is obtained from a general interaction Hamiltonian it is the fundamental quantum lattice model of interacting fermions[21]. As a consequence, many

[20] In Wannier basis the general Hamiltonian has infinitely many bands and model parameters. In the one-band model (19) all other bands are projected onto a single effective s band. The matrix elements of the kinetic energy and the Coulomb interaction are expected to fall off quickly with distance. Therefore one usually retains only the first few contributions. Thus hopping is restricted to nearest-neighbor sites i and j, such that $-t \equiv t_{ij}$. It is also natural to assume that the local part of the interaction ("Hubbard U") is the largest matrix element of the Coulomb interaction. Keeping only these parameters one obtains the Hubbard model, (19b), (19c). However, nearest-neighbor interactions (density-density, bond-charge, exchange interactions and hopping of doubly occupied sites) may be of appreciable size [45]. For a detailed discussion see ref. [52].

[21] More generally, the Hubbard model is the fundamental lattice model of quantum particles, since it may also be used for interacting bosons [53, 51].

well-known models can be derived from the Hubbard model in certain limits of the model parameters. For example, at half filling and in the limit $U \gg t$ the Hubbard model corresponds to the Heisenberg model with antiferromagnetic exchange coupling $J = 4t^2/U$.[22]

3.2.1 Does the Hubbard model explain band ferromagnetism?

As discussed above, the Hubbard model was originally introduced to provide a microscopic explanation of ferromagnetism in $3d$ transition metals [44–46]. Indeed, the Hubbard interaction favors a ferromagnetic ground state, since this corresponds to the state with the lowest energy (zero energy) due to the absence of doubly occupied sites. However, this argument ignores the kinetic energy. While the lattice structure and spatial dimension do not influence the Hubbard interaction, both play an important role in the kinetic energy, where they determine, for example, the density of states of the electronic band at the Fermi energy (see section 6.2.2).
In spite of the extreme simplifications of the Hubbard model compared with interacting electrons in a real solid, the model still cannot be solved analytically, except in dimension $d = 1$ for nearest-neighbor hopping [54]. For dimensions $d = 2, 3$, approximations are required.[23] Here mean-field theories play an important role.

4 Static mean-field theories of the Hubbard model

4.1 Hartree approximation

Lattice fermion models such as the Hubbard model are much more complicated than models with localized spins. Therefore the construction of a mean-field theory with the comprehensive properties of the mean-field theory of the Ising model will be more complicated, too. The simplest static mean-field theory of the Hubbard model is the Hartree approximation [56]. To clarify the characteristic features of this mean-field theory we proceed as in the derivation of the mean-field theory of the Ising model and factorize the interaction term. To this end we rewrite the Hubbard interaction in the form of (8), i.e., we let an electron with spin σ at site $\boldsymbol{R}_i$ interact with a local field $\hat{h}_{i\sigma}$ (an operator, which has a dynamics) produced by an electron with opposite spin on that site:

$$\hat{H}_{\text{int}} = \sum_{\boldsymbol{R}_i} \sum_{\sigma} \hat{h}_{i\sigma} \hat{n}_{i\sigma}, \tag{20}$$

where $\hat{h}_{i\sigma} = \frac{1}{2} U \hat{n}_{i-\sigma}$ (the factor $1/2$ is due to the summation over both spin directions). Next we replace the operator $\hat{h}_{i\sigma}$ by its expectation value $\langle \hat{h}_{i\sigma} \rangle$, now a real number, and obtain the

[22] This can be shown using second order degenerate perturbation theory or by a unitary (Schrieffer-Wolff) transformation.

[23] In view of the complexity of the many-body problem, progress in this field often relies on making good approximations. As Peierls wrote: "*... the art of choosing a suitable approximation, of checking its consistency and finding at least intuitive reasons for expecting the approximation to be satisfactory, is much more subtle than that of solving an equation exactly.*" [55].

mean-field Hamiltonian

$$\hat{H}^{\rm MF} = \hat{H}_{\rm kin} + \sum_{\boldsymbol{R}_i,\sigma} \langle \hat{h}_{i\sigma} \rangle \hat{n}_{i\sigma} + E_{\rm shift}, \tag{21}$$

where $E_{\rm shift}$ is a constant energy shift. Now a σ-electron at site $\boldsymbol{R}_i$ interacts only with a local *static* field $\langle \hat{h}_{i\sigma} \rangle = \frac{1}{2} U \langle \hat{n}_{i-\sigma} \rangle$. The above decoupling of the operators corresponds to the Hartree approximation[24] whereby correlated fluctuations *on* the site $\boldsymbol{R}_i$ are neglected.

It should be noted that although (21) is now an effective one-particle problem it cannot be solved exactly since, in general, the mean field $\langle \hat{h}_{i\sigma} \rangle$ may vary from site to site, leading to solutions without long-range order. This is a new feature originating from the quantum-mechanical kinetic energy in the Hamiltonian.

The Hartree approximation is valid in the weak-coupling ($U \to 0$) and/or low-density limit ($n \to 0$), but clearly does not become exact in the limit $d \to \infty$, since the Hubbard interaction between two electrons is purely local and hence does not dependent on the spatial dimension. Therefore the physics behind the factorizations (13) and (21) is very different. Namely, (13) describes the decoupling of a spin from a bath of infinitely many neighboring spins whose fluctuations become unimportant in the limit $d \to \infty$, while (21) corresponds to the decoupling of an electron from *one* other electron (with opposite spin) on the same site.

While the Hartree approximation is useful for investigations at weak coupling, it will lead to fundamentally wrong results at strong coupling, when the double occupation of a lattice site becomes energetically highly unfavorable and is therefore suppressed. Indeed, a factorization of the local correlation function $\langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle \to \langle \hat{n}_{i\uparrow} \rangle \langle \hat{n}_{i\downarrow} \rangle$ eliminates correlation effects generated by the local quantum dynamics. Hence the nature of the Hartree mean-field theory of spin-$\frac{1}{2}$ electrons with a local interaction is very different from the mean-field theory of spins with nearest-neighbor coupling.

4.2 Gutzwiller approximation

Another useful approximation scheme for quantum many-body systems makes use of variational wave functions [57]. Starting from a physically motivated many-body trial wave function the energy expectation value is calculated and then minimized with respect to the variational parameters. Although variational wave functions usually yield only approximate results, they have several advantages: they are physically intuitive, can be custom tailored to a particular problem, can be used even when standard perturbation methods fail or are inapplicable, and provide a rigorous upper bound for the exact ground state energy by the Ritz variational principle.

To investigate the properties of the electronic correlation model which he had introduced [44] (but which was later named after Hubbard), Gutzwiller also proposed a simple variational wave function [44]. This "Gutzwiller wave function" introduces correlations into the wave function for non-interacting particles by a purely local correlation factor in real space, which is con-

[24] Since the Hubbard interaction acts only between electrons with opposite spin on the same lattice site an exchange (Fock) term does not arise.

structed from the double occupation operator $\hat{D}$, (19c), as

$$|\Psi_G\rangle = g^{\hat{D}} |\mathrm{FS}\rangle \tag{22a}$$

$$= \prod_{\boldsymbol{R}_i} \left(1 - (1-g)\hat{D}_i\right) |\mathrm{FS}\rangle. \tag{22b}$$

Here $|\mathrm{FS}\rangle$ is the wave function of the non-interacting Fermi sea and g is a variational parameter with $0 \leq g \leq 1$. The projector $g^{\hat{D}}$ globally reduces the amplitude of those spin configurations in $|\mathrm{FS}\rangle$ with too many doubly occupied sites for given repulsion U. The limit $g = 1$ describes the non-interacting case, while $g \to 0$ corresponds to $U \to \infty$. The Gutzwiller wave function can be used to calculate the expectation value of an operator, e.g., the ground state energy of the Hubbard model, (19), as

$$E_{\mathrm{G}}(g, U) \equiv \frac{\langle \Psi_G | \hat{H} | \Psi_G \rangle}{\langle \Psi_G | \Psi_G \rangle}. \tag{23}$$

By computing the minimum of $E_{\mathrm{G}}(g, U)$ with respect to the variational parameter g, the latter is determined as a function of the interaction parameter U.

In general the evaluation of expectation values in terms of $|\Psi_G\rangle$ cannot be performed exactly. Therefore Gutzwiller introduced a non-perturbative approximation scheme whereby he obtained an explicit expression for the ground state energy of the Hubbard model [44, 58, 59]. The results of Gutzwiller's rather complicated approach were later re-derived by counting classical spin configurations [60]; for details see ref. [61] and section 3 of ref. [62]. The Gutzwiller approximation is therefore a semiclassical approximation. In the next section we will see that it can also be derived by calculating the expectation values of operators in terms of the Gutzwiller wave function in the limit $d \to \infty$.

4.2.1 Brinkman-Rice metal-insulator transition

The results of the Gutzwiller approximation [44,58] describe a correlated, normal-state fermionic system at zero temperature whose momentum distribution has a discontinuity q at the Fermi level, with $q = 1$ in the non-interacting case, which is reduced to $q < 1$ by the interaction as in a Landau Fermi liquid. In 1970 Brinkman and Rice [63] noticed that in the case of a half-filled band ($n_\uparrow = n_\downarrow = 1/2$) the Gutzwiller approximation describes a transition at a finite critical interaction strength U_c from an itinerant to a localized state, where lattice sites are singly occupied and the discontinuity q vanishes. This "Brinkman-Rice transition" therefore corresponds to a correlation induced ("Mott") metal-insulator transition. They argued [63] that the inverse of q can be identified with the effective mass of Landau quasiparticles, $q^{-1} = m^*/m \geq 1$, which diverges at U_c.

The Gutzwiller approximation yields results which are physically very reasonable. In the 1970s and 80s it was the only approximation scheme which was able to describe a Mott metal-insulator transition at a finite value of the interaction *and* was in accord with basic properties of Landau

Fermi liquid theory.[25] This was confirmed by a detailed investigation of the assumptions and implications of the Gutzwiller approximation, where I showed that the Gutzwiller-Brinkman-Rice theory was not only in qualitative [66], but even in good quantitative agreement with experimentally measured properties of normal-liquid ^{3}He [61]; for a discussion see section 3 of ref. [62].

4.2.2 Can the Gutzwiller approximation be derived by quantum many-body methods?

The semiclassical Gutzwiller approximation clearly has features of a mean-field theory, since the kinetic energy of the correlated system is merely a renormalization of the kinetic energy of the non-interacting system. This also explains why the results obtained for the Hubbard lattice model can even be used to describe liquid ^{3}He [61, 67]. However, the validity of the Gutzwiller approximation was unclear for a long time. In particular, it was not known how to improve the approximation systematically. This was clarified a few years later, when the Gutzwiller approximation was re-derived in two different ways: as a slave-boson mean-field theory [68], and by an exact diagrammatic calculation of expectation values of operators in terms of the Gutzwiller wave function in the limit $d \to \infty$ [69–71][26]. The latter derivation will be discussed next.

5 Lattice fermions in infinite dimensions

The expectation values of the kinetic and the interaction energy of the Hubbard model (19) can, in principle, be calculated in terms of the Gutzwiller wave function for arbitrary dimensions d, using diagrammatic quantum many-body theory. Introducing a new analytic approach in which expectation values are expressed by sums over different lattice sites such that Wick's theorem leads to contractions which involve only anticommuting numbers, Walter Metzner and I showed that in $d = 1$ the diagrams can be calculated analytically to all orders [69];[27] for a discussion see section 4 of ref. [62]. We also observed that the values of these diagrams could be calculated

[25] Other well-known approximation schemes, in particular those proposed by Hubbard, do not have these important properties: in the Hubbard-I approximation [45], which interpolates between the atomic limit and the non-interacting band, a band gap opens for any $U > 0$, while the Hubbard-III approximation [64], which corresponds to the coherent potential approximation [65] for disordered systems, the Fermi surface volume is not conserved.

[26] Kotliar and Ruckenstein [68] formulated a functional integral representation of the Hubbard and Anderson models in terms of auxiliary bosons, whose simplest saddle-point approximation ("slave-boson mean-field theory") reproduces the results of the Gutzwiller approximation. Thus they showed that the results of the Gutzwiller approximation can also be obtained without the use of the Gutzwiller variational wave function. By applying a gauge transformation to the Gutzwiller wave function Gebhard [72] later found that the calculation of expectation values with this wave function can be performed in $d = \infty$ even without diagrams. This provided a direct link between the slave-boson mean-field theory and the results obtained with the Gutzwiller wave function in $d = \infty$. The latter approach was generalized by Gebhard and collaborators to multi-band Hubbard models, which can be used to describe the effect of correlations in real materials ("Gutzwiller density functional theory") [73].

[27] The diagrams have the same form as the usual Feynman diagrams in quantum many-body theory, but a line corresponds to the time-independent one-particle density matrix $g^0_{ij,\sigma} = \langle \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} \rangle_0$ of the non-interacting system rather than to the one-particle propagator $G^0_{ij,\sigma}(t)$ since the variational approach involves only static quantities.

in $d = \infty$ when the momentum conservation at a vertex was neglected. When summing over all diagrams this approximation gave exactly the results of the Gutzwiller approximation [69]. Thus the Gutzwiller approximation had been derived systematically by diagrammatic quantum many-body methods in $d = \infty$. This showed that the limit $d \to \infty$ was not only useful for the investigation of spin models, but also in the case of lattice fermion models.[28]

5.1 Simplifications of diagrammatic quantum many-body theory

The simplifications of diagrammatic many-body perturbation theory arising in the limit $d \to \infty$ are due to a collapse of irreducible diagrams in position space, which implies that only site-diagonal ("local") diagrams, i.e., diagrams which only depend on a single site, remain [70, 71]. To understand the reason for this diagrammatic collapse let us, for simplicity, consider diagrams where lines correspond to the one-particle density matrix $g^0_{ij,\sigma}$ as they enter in the calculation of expectation values with the Gutzwiller wave function (note that, since $g^0_{ij,\sigma} = \lim_{t\to 0^-} G^0_{ij,\sigma}(t)$, the following arguments are equally valid for the one-particle Green function $G^0_{ij,\sigma}(t)$ or its Fourier transform).

The one-particle density matrix $g^0_{ij,\sigma}$ may be interpreted as the quantum amplitude of the hopping of an electron with spin σ between sites $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$. The square of its magnitude is therefore proportional to the *probability* of an electron to hop from $\boldsymbol{R}_j$ to a site $\boldsymbol{R}_i$. For nearest-neighbor sites $\boldsymbol{R}_i$, $\boldsymbol{R}_j$ on a lattice with coordination number Z this implies $|g^0_{ij,\sigma}|^2 \sim \mathcal{O}(1/Z)$, such that on a hypercubic lattice, where $Z = 2d$, and large d one finds [70, 71]

$$g^0_{ij,\sigma} \sim \mathcal{O}\Big(\frac{1}{\sqrt{d}}\Big), \tag{24}$$

and for general i, j one obtains [71, 75]

$$g^0_{ij,\sigma} \sim \mathcal{O}\Big(1/d^{\|\boldsymbol{R}_i - \boldsymbol{R}_j\|/2}\Big). \tag{25}$$

Here $\|\boldsymbol{R}\| = \sum_{n=1}^{d} |R_n|$ is the length of $\boldsymbol{R}$ in the "Manhattan metric", where electrons only hop along horizontal or vertical lines, but never along a diagonal; for further discussions of diagrammatic simplifications see ref. [76].

For non-interacting electrons at $T = 0$ the expectation value of the kinetic energy is given by

$$E^0_{\mathrm{kin}} = -t \sum_{\langle \boldsymbol{R}_i, \boldsymbol{R}_j \rangle} \sum_{\sigma} g^0_{ij,\sigma}. \tag{26}$$

[28] In the limit $d \to \infty$ simplifications occur even for fermions in the continuum. In the ground state of a d-dimensional Fermi gas in the continuum, spatial isotropy implies that $\boldsymbol{k}$-states occupy a Fermi sphere. As is well-known from the discussion of the classical ideal gas in the microcanonical ensemble in the limit $d \to \infty$ the volume of a high-dimensional sphere is located essentially *at* the surface. Therefore in the high dimensional Fermi sphere essentially all states lie at the Fermi energy E_F. This is easily confirmed by calculating the energy per particle in d dimensions, ${E_0}^{(d)}/N = \big(d/(d{+}2)\big)E_F$ [74], which indeed gives ${E_0}^{(\infty)}/N = E_F$ in $d = \infty$. It implies that the pressure of the Fermi gas at $T = 0$ ("Fermi pressure"), $P = 2{E_0}^{(d)}/Vd$, which is larger than zero in finite dimensions, goes to zero in this limit.

The sum over nearest neighbors leads to a factor $\mathcal{O}(Z)$ (which is $\mathcal{O}(d)$ for a hypercubic lattice). In view of the $1/\sqrt{d}$ dependence of $g^0_{ij,\sigma}$ it is therefore necessary to scale the nearest-neighbor hopping amplitude t as [70, 71]

$$t = \frac{t^*}{\sqrt{d}},\ t^* = \text{const.}, \tag{27}$$

so that the kinetic energy remains finite for $d \to \infty$. The same result may be derived in a momentum-space formulation.[29] It is important to bear in mind that, although $g^0_{ij,\sigma} \sim 1/\sqrt{d}$ vanishes for $d \to \infty$, the electrons are still mobile. Indeed, even in the limit $d \to \infty$ the off-diagonal elements of $g^0_{ij,\sigma}$ contribute, since electrons may hop to $Z \sim \mathcal{O}(d)$ many nearest neighbors with amplitude $t^*/\sqrt{d}$. [30]

5.2 The Hubbard model in $d = \infty$

A rescaling of the microscopic parameters of the Hubbard model with d is only required in the kinetic energy, since the interaction term is independent of the spatial dimension.[31] Altogether this implies that only the Hubbard Hamiltonian with a rescaled kinetic energy

$$\hat{H} = -\frac{t^*}{\sqrt{d}} \sum_{\langle \boldsymbol{R}_i, \boldsymbol{R}_j \rangle} \sum_{\sigma} \hat{c}^\dagger_{i\sigma} \hat{c}_{j\sigma} + U \sum_{\boldsymbol{R}_i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \tag{28}$$

has a non-trivial $d \to \infty$ limit where both the kinetic energy and the interaction contribute. Namely, it is the *competition* between the two terms which leads to interesting many-body

[29] The need for the scaling (27) also follows from the density of states of non-interacting electrons. For nearest-neighbor hopping on a d-dimensional hypercubic lattice, $\varepsilon_{\boldsymbol{k}}$ has the form $\varepsilon_{\boldsymbol{k}} = -2t \sum_{i=1}^{d} \cos k_i$ (here and in the following we set Planck's constant $\hbar$, Boltzmann's constant k_B, and the lattice spacing equal to unity). The density of states corresponding to $\varepsilon_{\boldsymbol{k}}$ is given by $N_d(\omega) = \sum_{\boldsymbol{k}} \delta(\omega - \varepsilon_{\boldsymbol{k}})$, which is the probability density for finding the value $\omega = \varepsilon_{\boldsymbol{k}}$ for a random choice of $\boldsymbol{k} = (k_1, \ldots, k_d)$. If the momenta k_i are chosen randomly, $\varepsilon_{\boldsymbol{k}}$ is the sum of d many independent (random) numbers $-2t \cos k_i$. The central limit theorem then implies that in the limit $d \to \infty$ the density of states is given by a Gaussian, i.e., $N_d(\omega) \stackrel{d\to\infty}{\longrightarrow} \frac{1}{2t\sqrt{\pi d}} \exp\left(-\left(\frac{\omega}{2t\sqrt{d}}\right)^2\right)$. Only if t is scaled with d as in (27) does one obtain a non-trivial density of states $N_\infty(\omega)$ in $d = \infty$ [77, 70] and thus a finite kinetic energy. The density of states on other types of lattices in $d = \infty$ can be calculated similarly [78, 79].

[30] Due to the diagrammatic collapse in $d = \infty$ it is possible to sum the diagrams representing the expectation value of operators in terms of the Gutzwiller wave function exactly [70, 71]. Lines in a diagram correspond to the one-particle density matrix $g^0_{ij,\sigma}$ rather than the one-particle Green's function $G^0_{ij,\sigma}(t)$. The results obtained in infinite dimensions are found to coincide with those of the Gutzwiller approximation, which was originally derived by classical counting of spin configurations (section 4.2). Obviously the Gutzwiller approximation is a mean-field theory – but what is the corresponding mean field? To answer this question we note that the one-particle irreducible self-energy (i.e., the quantity which has the same topological structure as the self-energy in the Green's function formalism) and which is denoted by S^* in [70, 71], becomes site-diagonal in infinite dimensions. In the translationally invariant, paramagnetic phase the self-energy is then site-independent, i.e., constant. Since the self-energy encodes the effect of the interaction between the particles (see footnote 32) it may be viewed as the *global mean field* of the Gutzwiller approximation; its value is given by eq. (10) in [70]. Expectation values can be evaluated exactly even when the Fermi sea $|\mathrm{FS}\rangle$ in (22) is replaced by an arbitrary, not necessarily translationally invariant one-particle starting wave function $|\Phi_0\rangle$ [70, 71]. In general, the mean field S^*_{ii} then depends on the lattice site i (but is still site-diagonal) and is determined by the self-consistent eq. (11) in [70], which reduces to eq. (10) in the paramagnetic case. Details of the calculation of S^*_{ii} can be found in the paper by Metzner [71].

[31] In the limit $d \to \infty$, interactions beyond the Hubbard interaction, e.g., nearest-neighbor interactions such as $\hat{H}_{nn} = \sum_{\langle \boldsymbol{R}_i, \boldsymbol{R}_j \rangle} \sum_{\sigma\sigma'} V_{\sigma\sigma'} \hat{n}_{i\sigma} \hat{n}_{j\sigma'}$ have to be scaled, too. In this case a scaling as in the Ising model, $V_{\sigma\sigma'} \to V^*_{\sigma\sigma'}/Z$, is required [80]. In $d = \infty$ non-local contributions therefore reduce to their (static) Hartree substitute and only the Hubbard interaction remains dynamical.

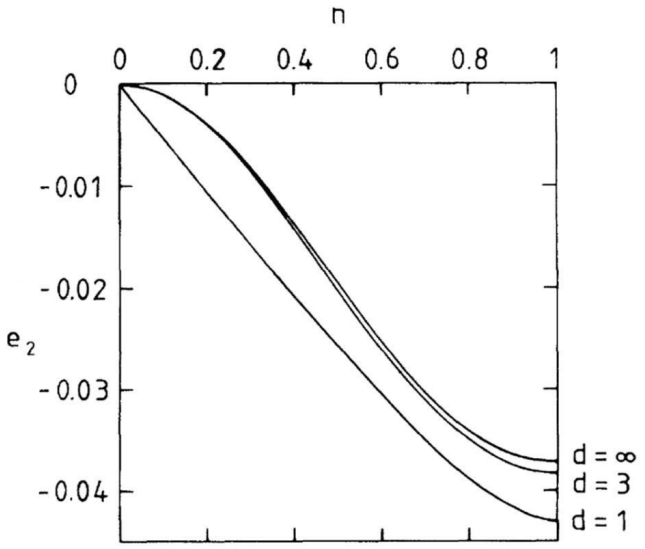


Fig. 3: *Correlation energy of the Hubbard model in second-order Goldstone perturbation theory in U in units of $2U^2/|\varepsilon_0|$ vs. density n in dimensions $d = 1, 3, \infty$. Here ε_0 is the kinetic energy for $U = 0$ and $n = 1$; adapted from ref. [70].*

physics. Most importantly, the Hubbard model still describes nontrivial correlations among the fermions even in $d = \infty$. This is already apparent from the evaluation of the second-order diagram in Goldstone perturbation theory for the correlation energy at weak coupling [70]. The integral over the three internal momenta (which lead to a nine-dimensional integral in $d = 3$) reduces to a single integral in $d = \infty$. Obviously the calculation is much simpler in $d = \infty$ than in finite dimensions. More importantly, the results for the energy obtained in $d = \infty$ turn out to be very close to those in $d = 3$ (Fig. 3) and therefore provide a computationally simple, but quantitatively reliable approximation.

These results showed that microscopic calculations for correlated lattice fermions in $d = \infty$ dimensions were useful and very promising. Further insights followed quickly: (i) Müller-Hartmann [80] proved that in infinite dimensions only the Hubbard interaction remains dynamical and that the self-energy becomes $\boldsymbol{k}$-independent, i.e., local in position space, as in the Gutzwiller approximation [70, 71] (see footnote 30), but retains its dynamics[32]

$$\Sigma_\sigma(\boldsymbol{k}, \omega) \overset{d\to\infty}{\equiv} \Sigma_\sigma(\omega), \tag{29}$$

whereby typical Fermi liquid features are preserved [85] (for a discussion see section 5.2.2),

[32]This result may be understood as follows [81, 82]: The interaction between particles influences their motion. This effect is described by a complex, spatially dependent and dynamical field, the self-energy $\Sigma_\sigma(\boldsymbol{k}, \omega)$. On a lattice with a very large number of nearest neighbors the *spatial* dependence of this field becomes increasingly unimportant and vanished completely in $d = \infty$, as in the mean-field theory of the Ising model. So the field becomes a mean field in position space but retains its full dynamics. In this respect there is a direct analogy to non-interacting electrons in the present of static ("quenched") disorder, where the self-energy also becomes $\boldsymbol{k}$-independent in the limit $d \to \infty$ ("coherent potential"). The coherent potential approximation [65] is a single-site theory where an electron moves through an effective medium described by the self-energy $\Sigma_\sigma(\omega)$, and becomes exact in $d = \infty$ [81, 83, 84]. In the case of the Hubbard model in the limit $d \to \infty$ the coherent potential is more complicated due to the interaction between the particles (see footnote 38).

(ii) Schweitzer and Czycholl [86] demonstrated that calculations for the periodic Anderson model also become much simpler in high dimensions, and (iii) Brandt and Mielsch [87] derived the exact solution of the Falicov-Kimball model in infinite dimensions by mapping the lattice problem onto a solvable atomic problem in a generalized, time-dependent external field; they also indicated that in principle such a mapping was even possible for the Hubbard model.[33]
Due to the $\boldsymbol{k}$-independence of the irreducible self-energy the most important obstacle for diagrammatic calculations in finite dimensions $d \geq 1$, namely the integration over intermediate momenta, is removed. At the same time the limit $d \to \infty$ does not affect the dynamics of the system. Hence, in spite of the simplifications in position or momentum space, the many-electron problem retains its full dynamics in $d = \infty$.

5.2.1 One-particle propagator and spectral function

In $d = \infty$ the one-particle propagator of an interacting lattice fermion system (the "lattice Green function") at $T = 0$ is then given by

$$G_{\boldsymbol{k},\sigma}(\omega) = \frac{1}{\omega - \varepsilon_{\boldsymbol{k}} + \mu - \Sigma_\sigma(\omega)}. \tag{30}$$

The $\boldsymbol{k}$-dependence of $G_{\boldsymbol{k}}(\omega)$ now comes entirely from the energy dispersion $\varepsilon_{\boldsymbol{k}}$ of the *non*-interacting particles. This means that in a homogeneous system described by the propagator

$$G_{ij,\sigma}(\omega) = L^{-1} \sum_{\boldsymbol{k}} G_{\boldsymbol{k},\sigma}(\omega) e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_i - \boldsymbol{R}_j)} \tag{31}$$

its local part, $G_{ii,\sigma} \equiv G_\sigma$, is given by

$$G_\sigma(\omega) = L^{-1} \sum_{\boldsymbol{k}} G_{\boldsymbol{k},\sigma}(\omega) = \int\limits_{-\infty}^{\infty} dE \frac{N_0(E)}{\omega - E + \mu - \Sigma_\sigma(\omega)}, \tag{32}$$

where $N_0(E)$ is the density of states of the non-interacting system. The spectral function of the interacting system (also often called density of states) is given by

$$A_\sigma(\omega) = -\frac{1}{\pi} \mathrm{Im} G_\sigma(\omega + i0^+). \tag{33}$$

5.2.2 $\boldsymbol{k}$-independence of the self-energy and Fermi liquid behavior

The $\boldsymbol{k}$-independence of the self-energy allows one to make contact with Fermi liquid theory [85]. In general, i.e., even when Σ has a $\boldsymbol{k}$-dependence, the Fermi surface is defined by the $\omega = 0$ limit of the denominator of (30) (in the paramagnetic phase we can suppress the spin index)[34]

$$\varepsilon_{\boldsymbol{k}} + \Sigma_{\boldsymbol{k}}(0) = E_F. \tag{34a}$$

[33] Alternatively, it can be shown that in the limit $Z \to \infty$ the dynamics of the Falicov-Kimball model reduces to that of a non-interacting, tight-binding model on a Bethe lattice with coordination number $Z = 3$ which can be solved analytically [88].

[34] We note that the notion of a $(d-1)$-dimensional Fermi surface of electrons on d-dimensional lattices is no longer meaningful in $d = \infty$ (see footnote 28). Therefore the following discussion addresses the consequences of a momentum independent self-energy for finite-dimensional rather than strictly infinite-dimensional systems.

According to Luttinger and Ward [89] the volume within the Fermi surface is not changed by interactions, provided the latter can be treated in perturbation theory.[35] This is expressed by

$$n = \sum_{\boldsymbol{k}\sigma} \Theta\big(E_F - \varepsilon_{\boldsymbol{k}} - \Sigma_{\boldsymbol{k}}(0)\big), \tag{34b}$$

where n is the electron density and $\Theta(x)$ is the step function. The $\boldsymbol{k}$-dependence of $\Sigma_{\boldsymbol{k}}(0)$ in (34a) implies that, in spite of (34b), the shape of the Fermi surface of the interacting system will be quite different from that of the non-interacting system, except for the rotationally invariant case $\varepsilon_{\boldsymbol{k}} = f(|\boldsymbol{k}|)$. By contrast, for lattice fermion models in $d = \infty$, where $\Sigma_{\boldsymbol{k}}(\omega) \equiv \Sigma(\omega)$, the Fermi surface itself, and hence the enclosed volume, is not changed by the interaction. The Fermi energy is simply shifted uniformly from its non-interacting value E_F^0 to $E_F = E_F^0 + \Sigma(0)$, to keep n in (34b) constant. Thus $G(0)$, the $\omega = 0$ value of the local lattice Green function, and the spectral function $A(0) = -\frac{1}{\pi}\mathrm{Im}G(i0^+)$ are not changed by the interaction at all. This "pinning behavior" is well-known from the single-impurity Anderson model [91]. A renormalization of $N(0)$ can only be due to a $\boldsymbol{k}$-dependence of Σ.
For $\omega \to 0$ the self-energy has the property [85]

$$\mathrm{Im}\ \Sigma(\omega) \propto \omega^2, \tag{34c}$$

which implies Fermi liquid behavior. The effective mass of the quasiparticles

$$\frac{m^*}{m} = 1 - \frac{d\mathrm{Re}\Sigma}{d\omega}\bigg|_{\omega=0} = 1 + \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega\ \frac{\mathrm{Im}\Sigma(\omega + i0^-)}{\omega^2} \geq 1 \tag{34d}$$

is seen to be enhanced. In particular, the momentum distribution

$$n_{\boldsymbol{k}} = \frac{1}{\pi}\int_{-\infty}^{0} d\omega\ \mathrm{Im}G_{\boldsymbol{k}}(\omega) \tag{35}$$

has a discontinuity at the Fermi surface given by $n_{k_F^-} - n_{k_F^+} = (m^*/m)^{-1}$, where $k_F^\pm = k_F \pm 0^+$.

5.2.3 Is there a unique $d \to \infty$ limit of the Hubbard model?

The motivation for the scaling of the hopping amplitude in the limit $d \to \infty$, (27), deserves a more detailed discussion. To obtain a physically meaningful mean-field theory of a model it is necessary that its internal or free energy remains finite in the limit d or $Z \to \infty$. While for the Ising model the scaling $J = J^*/Z$, $J^* =$ const., is rather obvious, this is not so for more complicated models. Namely, fermionic or bosonic many-particle systems are typically described by a Hamiltonian with non-commuting terms, e.g., a kinetic energy and an interaction, each of which is associated with a coupling parameter, usually a hopping amplitude and an interaction, respectively. In such a case the question of how to scale these parameters has no unique answer since this depends on the physical effects one wishes to explore. The scaling

[35] Recently necessary and sufficient conditions for the validity of Luttinger's theorem [89] based on the Atiyah-Singer index theorem were derived, by which the topological robustness of a generalized Fermi surface may be quantified [90].

should be performed such that the model remains non-trivial and "physically interesting" and that its energy stays finite in the $d, Z \to \infty$ limit. Here "non-trivial" means that not only $\langle \hat{H}_0 \rangle$ and $\langle \hat{H}_{\rm int} \rangle$ but also their competition, expressed by the commutator $\left\langle [\hat{H}_0, \hat{H}_{\rm int}] \right\rangle$, should remain finite. In the case of the Hubbard model it would be possible to scale the hopping amplitude as in the mean-field theory of the Ising model, i.e., $t \to t^*/Z,\ t^* = {\rm const.}$, but then the kinetic energy would be reduced to zero in the limit $d, Z \to \infty$, making the resulting model uninteresting (but not unphysical) for most purposes. For the bosonic Hubbard model the situation is even more subtle, since the kinetic energy has to be scaled differently depending on whether it describes the normal or the Bose-Einstein condensed fraction; for a discussion see ref. [92]. Hence, in the case of a many-body system described by a Hamiltonian with several terms, a mean-field solution in the limit $d \to \infty$ depends crucially on how the scaling of the model parameters is chosen.

6 Dynamical mean-field theory (DMFT) of correlated electrons and its applications

The diagrammatic simplifications of quantum many-body theory in infinite spatial dimensions provide the basis for the construction of a comprehensive mean-field theory of lattice fermions which is diagrammatically controlled and whose free energy has no unphysical singularities. The construction is based on the scaled Hamiltonian (28). The self-energy is then momentum independent, but retains its frequency dependence and thereby describes the full many-body dynamics of the interacting system.[36] The resulting theory is both mean-field-like and dynamical and therefore represents a *dynamical* mean-field theory (DMFT) for lattice fermions which is able to describe genuine correlation effects as will be discussed next.

6.1 The self-consistent DMFT equations

In $d = \infty$, lattice fermion models with a local interaction effectively reduce to a single site embedded in a dynamical mean field provided by the other interacting fermions as illustrated in Fig. 4. The self-consistent DMFT equations can be derived in different ways depending on the physical interpretation of the correlation problem emerging in the limit $d, Z \to \infty$ [81, 93, 94]; for a discussion see ref. [76]. The mapping of the lattice electron problem onto a single-impurity Anderson model[37] with a self-consistency condition in $d = \infty$ introduced by Georges and Kotliar [93], which was also employed by Jarrell [94], turned out to be the most useful approach, since it made a connection with the well-studied theory of quantum impurities [91], for whose solution efficient numerical codes such as the quantum Monte-Carlo (QMC) method

[36] This is in contrast to Hartree(-Fock) theory where the self-energy acts only as a static potential.

[37] The mapping itself can be performed without approximation, but leads to a complicated coupling between the impurity and the bath which makes the problem intractable. This can be solved in the limit $d \to \infty$ when the momentum dependence of the self-energy drops out.

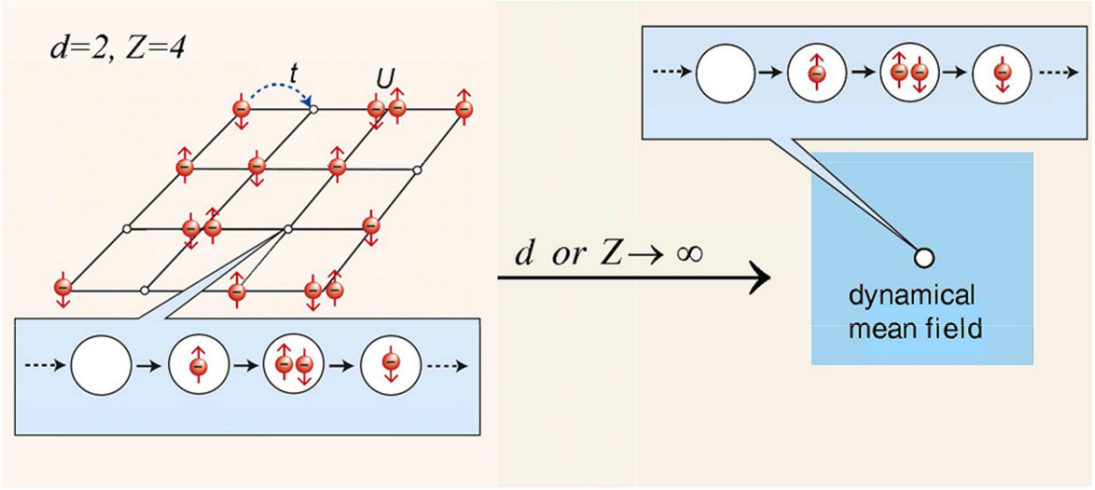


Fig. 4: *In the limit d or $Z \to \infty$ the Hubbard model effectively reduces to a dynamical single-site problem which may be viewed as a lattice site embedded in a $\boldsymbol{k}$-independent, dynamical fermionic mean field. Electrons can hop from the mean field onto this site and back, and interact on the site as in the original Hubbard model (see Fig. 2). The local dynamics of the electrons is independent of the dimension or coordination number and therefore remains unchanged.*

[95] had already been developed and were readily available.[38] For a detailed discussion of the foundations of DMFT see the review by Georges, Kotliar, Krauth, and Rozenberg [96] and the lecture by Kollar at the Jülich Autumn School 2018 [97]; an introductory presentation can be found in ref. [98].

For $T > 0$ the self-consistent DMFT equations are given by:

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

$$G_\sigma(i\omega_n) = -\frac{1}{\mathcal{Z}} \int \prod_\sigma Dc_\sigma^* Dc_\sigma [c_\sigma(i\omega_n)\, c_\sigma^*(i\omega_n)]\, \exp\bigl(S_{\rm loc}\bigr) \tag{36}$$

with the partition function

$$\mathcal{Z} = \int \prod_\sigma Dc_\sigma^* Dc_\sigma \exp\bigl(-S_{\rm loc}\bigr) \tag{37}$$

and the local action

$$S_{\rm loc} = -\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \sum_\sigma c_\sigma^*(\tau_1)\, \mathcal{G}_\sigma^{-1}(\tau_1 - \tau_2)\, c_\sigma(\tau_2) + U \int_0^\beta d\tau\, c_\uparrow^*(\tau) c_\uparrow(\tau) c_\downarrow^*(\tau) c_\downarrow(\tau). \tag{38}$$

Here $\mathcal{G}_\sigma$ is the effective local propagator (also called "bath Green function", or "Weiss mean

[38] Alternatively Janiš derived the self-consistent DMFT equations by generalizing the coherent potential approximation (CPA) [81]. In the CPA quenched disorder acting on non-interacting electrons is averaged and produces a mean field, the "coherent potential". In the case of the Hubbard model in $d = \infty$ the fluctuations generated by the Hubbard interaction may be treated as "annealed" disorder acting on non-interacting electrons [82] which, after averaging, produce a mean field, the self-energy. Numerical solutions of the DMFT equations starting from the CPA point of view have not yet been developed.

field"[39]), which is defined by a Dyson equation

$$\mathcal{G}_\sigma(i\omega_n) = \Big(\big(G_\sigma(i\omega_n)\big)^{-1} + \Sigma_\sigma(i\omega_n)\Big)^{-1}. \tag{39}$$

Furthermore, by identifying the local propagator (36) with the Hilbert transform of the lattice Green function

$$G_{\boldsymbol{k}\,\sigma}(i\omega_n) = \frac{1}{i\omega_n - \varepsilon_{\boldsymbol{k}} + \mu - \Sigma_\sigma(i\omega_n)}, \tag{40}$$

(this identification is exact in $d = \infty$ [96]), one obtains
(II) the *self-consistency condition*

$$G_\sigma(i\omega_n) = \frac{1}{L}\sum_{\boldsymbol{k}} G_{\boldsymbol{k}\,\sigma}(i\omega_n) \;=\; \int\limits_{-\infty}^{\infty} d\varepsilon \frac{N(\omega)}{i\omega_n - \varepsilon + \mu - \Sigma_\sigma(i\omega_n)} \tag{41}$$

$$= \; G^0_\sigma\big(i\omega_n - \Sigma_\sigma(i\omega_n)\big). \tag{42}$$

In (41) the ionic lattice enters only through the density of states of the non-interacting electrons.[40]

The self-consistent DMFT equations can be solved iteratively: starting with an initial guess for the self-energy $\Sigma_\sigma(i\omega_n)$ one obtains the local propagator $G_\sigma(i\omega_n)$ from (41) and thereby the bath Green function $\mathcal{G}_\sigma(i\omega_n)$ from (39). This determines the local action (38) which is needed to compute a new value for the local propagator $G_\sigma(i\omega_n)$ from (36). By employing the old self-energy a new bath Green function $\mathcal{G}_\sigma$ is calculated and so on, until convergence is reached.

It should be stressed that although the DMFT corresponds to an effectively local problem, the propagator $G_{\boldsymbol{k}}(\omega)$ depends on the crystal momentum $\boldsymbol{k}$ through the dispersion relation $\varepsilon_{\boldsymbol{k}}$ of the non-interacting electrons. But there is no *additional* momentum-dependence through the self-energy, since this quantity is local within the DMFT.

Solutions of the self-consistent DMFT equations require the extensive application of numerical methods, in particular quantum Monte-Carlo (QMC) simulations [94, 96] with continuous-time QMC [99] still as the method of choice, the numerical renormalization group [100], the density matrix renormalization group [101], exact diagonalization [96], Lanczos procedures [102], and solvers based on matrix product states [103] or tensor networks [104]. Here the recent development of impurity solvers making use of machine learning opens new perspectives [105].

[39] This expresses the fact that $\mathcal{G}$ describes the coupling of a single site to the rest of the system, similar to the Weiss mean-field $h_{\rm MF}$ in the mean-field theory of the Ising model (see section 2.1). However, in the case of the DMFT the mean field depends on the frequency, i.e., is dynamical. It should be noted that, in principle, both local functions $\mathcal{G}_\sigma(i\omega_n)$ and $\Sigma_\sigma(i\omega_n)$ can be viewed as a dynamical mean field since both enter in the bilinear term of the local action (38).

[40] Eq. (42) illustrates the mean-field nature of the DMFT equations very clearly: the local Green function of the interacting system is given by the non-interacting Green function G^0_σ at the renormalized energy $i\omega_n - \Sigma_\sigma(i\omega_n)$, which corresponds to the energy measured relative to the energy $\Sigma_\sigma(i\omega_n)$ of the surrounding fermionic bath, the dynamical mean field.

6.1.1 Characteristic features of DMFT

In DMFT the mean field is dynamical, whereby local quantum fluctuations are fully taken into account, but is local (i.e., spatially independent) because of the infinitely many neighbors of every lattice site ("single-site DMFT"). The only approximation of DMFT when applied in $d < \infty$ is the neglect of the $\boldsymbol{k}$-dependence of the self-energy. DMFT provides a comprehensive, non-perturbative, thermodynamically consistent and diagrammatically controlled approximation scheme for the investigation of correlated lattice models at all interaction strengths, densities, and temperatures [96, 98], which can resolve even low energy scales. It describes fluctuating moments, the renormalization of quasiparticles and their damping, and is especially valuable for the study of correlation problems at intermediate couplings, where no other investigation methods are available. Unless a symmetry is broken the $\boldsymbol{k}$-independence of the self-energy implies typical Fermi-liquid properties of the DMFT solution.
Most importantly, with DMFT it is possible to compute electronic correlation effects quantitatively in such a way that they can be tested experimentally, for example, by electron spectroscopies. Namely, DMFT describes the correlation induced transfer of spectral weight and the finite lifetime of quasiparticles through the real and imaginary part of the self-energy, respectively. This greatly helps to understand and characterize the Mott metal-insulator transition (MIT) to be discussed next.

6.2 Mott transition, ferromagnetism, and topological properties

Intensive theoretical investigations of the Hubbard model and related correlation models using DMFT over the last three decades have provided a wealth of new insights into the physics described by this fundamental fermionic interaction model. In the following subsection only a few exemplary results will be discussed; more detailed presentations can be found in refs. [96, 76], and the lecture notes of the Jülich Autumn Schools in 2011, 2014, and 2018 [42].

6.2.1 Metal-insulator transitions

Mott-Hubbard transition The interaction-driven transition between a paramagnetic metal and a paramagnetic insulator, first discussed by Mott [106] and referred to as "Mott metal-insulator transition" (MIT), or "Mott-Hubbard MIT" when studied within the Hubbard model, is one of the most intriguing phenomena in condensed matter physics [107,108]. This transition is a consequence of the quantum-mechanical competition between the kinetic energy of the electrons and their interaction U: the kinetic energy prefers the electrons to be mobile (a wave effect) which invariably leads to their interaction (a particle effect). For large values of U doubly occupied sites become energetically too costly. The system can reduce its total energy by localizing the electrons, which leads to a MIT. Here the DMFT provided detailed insights into the nature of the Mott-Hubbard-MIT for all values of the interaction U and temperature T [96, 109, 98, 76]. A microscopic investigation of the Mott MIT obtained within DMFT from a Fermi-liquid point of view was performed only recently [110].

While at small U the interacting system can be described by coherent quasiparticles whose spectral function ("density of states" (DOS)) still resembles that of the free electrons, the DOS in the Mott insulating state consists of two separate incoherent "Hubbard bands" whose centers are separated approximately by the energy U (here we discuss only the half-filled case without magnetic order). At intermediate values of U the spectrum then has a characteristic three-peak structure which is qualitatively similar to that of the single-impurity Anderson model [111] and which is a consequence of the three possible occupations of a lattice site: empty, singly occupied (up or down), and doubly occupied.
At $T = 0$ the width of the quasiparticle peak vanishes at a critical value of U which is of the order of the band width. So the Mott-Hubbard MIT occurs at intermediate coupling and therefore belongs to the hard problems in many-body theory, where most analytic approaches fail and investigations have to rely on numerical methods. Therefore several features of the Mott-Hubbard MIT near the transition point are still not sufficiently understood even within DMFT. Actually, it was recently argued that the Mott-Hubbard transition in the infinite-dimensional one-band Hubbard model may be understood as a topological phase transition, where the insulating state is the topological phase, and the transition from the metallic (Fermi liquid) to the insulating state involves domain wall dissociation [112].
At $T > 0$ the Mott-Hubbard MIT is found to be first order and is associated with a hysteresis region in the interaction range $U_{c1} < U < U_{c2}$ where U_{c1} and U_{c2} are the values at which the insulating and metallic solution, respectively, vanish [96,109]; for more detailed discussions see refs. [96, 98, 76]. The hysteresis region terminates at a critical point, above which the transition becomes a smooth crossover from a "bad metal" to a "bad insulator"; for a schematic plot of the phase diagram see fig. 3 of ref. [98]. Transport in the incoherent region above the critical point shows remarkably rich properties, including scaling behavior [113].
Mott-Hubbard MITs are found, for example, in transition metal oxides with partially filled bands. For such systems band theory typically predicts metallic behavior. One of the most famous examples is V_2O_3 doped with Ti or Cr [114]. However, it is now known that certain organic materials are better realizations of the single-band Hubbard model without magnetic order and allow for much more controlled investigations of the Mott state and the Mott MIT [115].

Metal-insulator transitions in the presence of disorder DMFT also provides a theoretical framework for the investigation of correlated electrons in the presence of disorder. When the effect of local disorder is taken into account through the arithmetic mean of the local DOS (LDOS) one obtains, in the absence of interactions, the coherent potential approximation (CPA) [83,84]; for a discussion see ref. [76]. However, CPA cannot describe Anderson localization. To overcome this deficiency a variant of the DMFT was formulated where the *geometrically* averaged LDOS is computed from the solutions of the self-consistent stochastic DMFT equations and is then fed into the self-consistency cycle [116]. This corresponds to a "typical medium theory" which can describe the Anderson transition of non-interacting electrons. By implementing this scheme into DMFT to study the properties of disordered electrons in the presence of interactions it is possible to compute the phase diagram of the Anderson-Hubbard model [117].

6.2.2 Band ferromagnetism

The Hubbard model was introduced in 1963 [44–46] in an attempt to explain metallic ("band") ferromagnetism in $3d$ metals such as Fe, Co, and Ni starting from a microscopic point of view. However, at that time investigations of the model employed strong and uncontrolled approximations. Therefore it was uncertain for a long time whether the Hubbard model can explain band ferromagnetism at realistic temperatures, electron densities, and interaction strengths in $d > 1$ at all. Three decades later it was shown with DMFT that on generalized fcc-type lattices in $d = \infty$ (i.e., on "frustrated" lattices with large spectral weight at the lower band edge) the Hubbard model indeed describes metallic ferromagnetic phases in large regions of the phase diagram [79,118]. In the paramagnetic phase the susceptibility χ_F obeys a Curie-Weiss law [119], where the Curie temperature T_C is now much lower than that obtained by Stoner theory, due to many-body effects. In the ferromagnetic phase the magnetization M is consistent with a Brillouin function as originally derived for localized spins, even for a *non-integer* magneton number as in $3d$ transition metals. Therefore, DMFT accounts for the behavior of both the magnetization and the susceptibility of band ferromagnets [79, 118]; see section 6.1 of my lecture notes at the 2018 Jülich Autumn School [97]. As the Mott MIT, band ferromagnetism is a hard intermediate-coupling problem.

6.2.3 Topological properties of correlated electron systems

The interest in non-trivial topological properties of electronic systems sparked by the theory of the quantum Hall effect greatly increased when it was realized that the spin-orbit interaction can generate topologically insulating behavior [120]. Initially, investigations focused on topological features of non-interacting systems. But during the last decade the influence of electronic interactions on these topological properties received more and more attention. Here DMFT turned out to be a useful tool. For example, DMFT was employed to study interaction effects in two-dimensional topological insulators [121], and to analyze the robustness of the Chern number in the Haldane-Hubbard model [122] as well as of the topological quantization of the Hall conductivity of correlated electrons at $T > 0$ [123]. Furthermore, to better understand the topological phase transition from a Weyl-semimetal to a Mott insulator the topological properties of quasiparticle bands were computed [124]. DMFT also made it possible to explore topological phase transitions in the Kitaev model in a magnetic field and to calculate the corresponding phase diagrams [125]. Correlation-induced topological effects can even arise from non-Hermitian properties of the single-particle spectrum in equilibrium systems [126]. Recently it was demonstrated explicitly that a topologically nontrivial multiorbital Hubbard model remains well-defined and nontrivial in the limit $d \to \infty$ for arbitrary, but even, d [127].

6.2.4 Nonequilibrium DMFT

The study of correlated electrons out of equilibrium by employing a generalization of the DMFT has become another fascinating new research area [128]. Nonequilibrium DMFT is able to explain, for example, the results of time-resolved electron spectroscopy experiments, where femtosecond pulses are now available in a wide frequency range. In such experiments a probe

is excited and the subsequent relaxation is studied. One recent example is the ultrafast dynamics of doubly occupied sites in the photo-excited quasi-$2d$ transition-metal dichalcogenide $1T$-TaS_2 [129]. Such excitations may even result in long-lived, metastable ("hidden") states [130].

6.3 Correlated electrons in bulk materials, surfaces, and nanostructures

The development and application of theoretical techniques to explore the basic features of the physics described by the one-band Hubbard model took several decades. During that time first-principles investigations of the even more complicated many-body problem of correlated *materials* were clearly out of reach. The electronic properties of solids were mainly studied within density-functional theory (DFT) [131], e.g., in the local density approximation (LDA) [132], the generalized gradient approximation (GGA) [133], and the LDA+U method [134]. Those approaches can describe the ground state properties of many simple elements and semiconductors, and even of some insulators, quite accurately, and often predict the magnetic and orbital properties [132] as well as the crystal structures of many solids correctly [135]. However, these methods fail to describe the electronic and structural properties of correlated paramagnetic materials since they miss characteristic features of correlated electron systems such as heavy quasiparticle behavior and Mott physics. This situation changed dramatically with the advent of DMFT.

6.3.1 DFT+DMFT and *GW*+DMFT

The computational scheme introduced by Anisimov *et al.* [136] and Lichtenstein and Katsnelson [137], which merges material-specific DFT-based approximations with the many-body DMFT and which is now denoted by DFT+DMFT (or rather more specifically LDA+DMFT, GGA+DMFT, etc.), provides a powerful new method for the microscopic computation of the electronic, magnetic, and structural properties of correlated materials from first principles even at finite temperatures [138–142]. In particular, this approach naturally accounts for the existence of local moments in the paramagnetic phase. By construction, DFT+DMFT includes the correct quasiparticle physics and the corresponding energetics, and reproduces the DFT results in the limit of weak Coulomb interaction U. Most importantly, DFT+DMFT describes the correlation-induced many-body dynamics of strongly correlated electron materials at all values of the Coulomb interaction and doping.

As in the case of the single-band Hubbard model the many-body model of correlated materials constructed within the DFT+DMFT scheme consists of two parts: an effective kinetic energy obtained by DFT which describes the material-specific band structure of the uncorrelated electrons, and the local interactions between the electrons in the same orbital as well as in different orbitals. Here the static contribution of the electronic interactions already included in the DFT-approximations must be subtracted to avoid double counting [136–141]. Such a correction is not necessary in the fully diagrammatic, but computationally very demanding GW+DMFT approach, where the LDA/GGA input is replaced by the GW approximation [143]. The resulting many-particle problem with its numerous energy bands and local interactions is then solved within DMFT, typically by CT-QMC.

Bulk materials The application of DFT+DMFT made investigations of correlated materials much more realistic and led to the discovery of novel physical mechanisms and correlation phenomena. Take, for example, the Mott MIT. Within the single-band Hubbard model the Mott MIT was originally explained as a transition where the effective mass of quasiparticles diverges ("Brinkman-Rice scenario") [63]. After DMFT had opened the way to study multi-band models, an orbitally-selective Mott MIT was identified [144]. Then, with the advent of DFT+DMFT, a site-selective Mott MIT was discovered in Fe_2O_3 with its numerous energy bands and local interactions [145]. With DFT+DMFT it was also shown that the non-trivial topological properties of α-Ce are driven by the f-d band inversion, which originates from the formation of a coherent $4f$ band around the Fermi energy [146].
DFT+DMFT has been remarkably successful in the investigation of correlated materials, including transition metals and their oxides, manganites, fullerenes, Bechgaard salts, f-electron materials, magnetic superconductors, and Heusler alloys [138–141]. In particular, the study of Fe-based pnictides and chalcogenides led to the new insight that in metallic multi-orbital materials the intra-atomic exchange J can also induce strong correlations [147]. Clearly DMFT-based approaches will be very useful for the future design of correlated materials [148], such as materials with a high thermopower for thermoelectric devices which can convert waste heat into electric energy [149].

Surfaces, layers, and nanostructures DMFT studies of inhomogeneous systems greatly improved our understanding of correlation effects at surfaces and interfaces, in thin films and multi-layered nanostructures [150] and, most recently, in infinite layer nickelates [151]. Thereby they provided a new understanding of potential functionalities of such structures and their application in electronic devices. DMFT has been extended to study correlations also in finite systems such as nanoscopic conductors and molecules [152]. In this way many-body effects were shown to be important even in biological matter, e.g., in the kernel of hemoglobin and molecules with important biological functions [153].

7 Beyond mean-field theory

Mean-field approximations provide useful information on the general physical properties of many-body problems. In particular, DMFT with its dynamical but local self-energy has been a breakthrough for the investigation and explanation of electronic correlation effects in models and materials. Although it is an approximation when used in $d < \infty$, experiments with cold atoms in optical lattices demonstrated that DMFT can be remarkably accurate in $d = 3$ [154]. A dynamical, local self-energy was also shown to be well justified in iron pnictides and chalcogenides [155] as well as in Sr_2RuO_4 [156]. Nevertheless mean-field results clearly cannot explain correlation phenomena occurring on finite length scales, the critical behavior at thermal or quantum phase transitions, or unconventional superconductivity observed in finite-dimensional systems. In such cases corrections to mean-field theory must be included.

7.1 $1/d$ corrections

For mean-field theories derived in the limit $d \to \infty$ corrections can be obtained, in principle, by performing an expansion in the parameter $1/d$ around the mean-field results:

Mean-field theory of the Ising model $1/d$-expansions [3] around this paradigmatic mean-field theory[41] have been performed in great detail, resulting in high-order expansions of the free energy and of susceptibilities [158]. Expansions in $1/d$ to determine the critical temperature T_C in $d < \infty$ appear to be asymptotic [159].

Gutzwiller approximation Systematic $1/d$-expansions around the Gutzwiller approximation lead to excellent agreement with results obtained by variational Monte-Carlo techniques in $d = 2, 3$ [75, 71, 72]. However, finite orders in $1/d$ do not remove the Brinkman-Rice transition (section 4.2.1) found in $d = \infty$.

Coherent potential approximation (CPA) The CPA corresponds to the exact solution of the Anderson disorder model in $d = \infty$ [83, 84], but does not describe Anderson localization. Corrections to the CPA self-energy in $1/d$ can be calculated, but bare perturbation theory leads to unphysical divergencies at the band edges [83]. Indeed, such expansions do not fulfill the condition $\mathrm{Im}\big(z - \Sigma(z)\big) \lesseqgtr 0$ for $\mathrm{Im}z \lesseqgtr 0$ (Herglotz analyticity) which would guarantee the non-negativity of the density of states. This also holds for $1/d$-corrections to the conductivity of disordered electrons. The conductivity has the correct analytic properties only if the self-energy is properly connected to the irreducible vertex function by a Ward identity [160, 161]. To this end vertex corrections need to be included which, however, mix corrections with different powers in $1/d$. Self-consistent $1/d$-expansions around CPA which are able to describe the Anderson localization transition are presently not in sight. Here numerical cluster approaches to be discussed in section 7.2 are more successful.

DMFT of correlated lattice electrons Systematic calculations of $1/d$-corrections to the DMFT start with a Luttinger-Ward functional of the non-local Green function $G_{ij,\sigma}(i\omega_n)$, from which the non-local self-energy $\Sigma_{ij,\sigma}(i\omega_n)$ is obtained by functional derivative. To calculate first-order corrections in $1/d$ one needs to consider only one pair of nearest-neighbor sites [162]. This generalizes the single-impurity problem of the DMFT to a two-impurity problem. It is then possible, in principle, to formulate a self-consistent, thermodynamically consistent approximation which is correct to order $1/d$ [162, 160]. This scheme requires an exact cancellation of certain diagrams in the approximation. Unfortunately, numerical computations within this approach are unstable and can easily lead to acausal solutions. Although the scheme was modified such that the diagrammatic cancellation is assured at each iteration step [163] and thereby provided a causal solutions in test calculations for the Hubbard model, acausal behavior cannot be ruled out

[41] For the ferromagnetic Ising model high-temperature expansions can be converted to $1/d$-expansions around mean-field theory [157].

in general. Therefore it is still not clear whether, and how, controlled and thermodynamically consistent $1/d$ expansions around DMFT can be constructed. Apparently analytic calculations in this directions were not further pursued in view of the successes of numerical cluster approaches which, although not systematic in $1/d$, are explicitly causal [164] (see below).

7.2 Beyond DMFT

There are different, mostly numerical, techniques to include non-local correlations into the DMFT; for a review see ref. [165]; in the following we mention four approaches:

Extended DMFT This is an early strategy to include intersite quantum fluctuations into DMFT [166], where the interaction strength of a nearest-neighbor density-density interaction is scaled such that its fluctuation part contributes even in the large d limit.

Cluster extensions The dynamical cluster approximation (DCA) [167] and the cellular DMFT (CDMFT) [168] map a lattice model onto a cluster of sites, which is then self-consistently embedded in a dynamical mean field. Thereby it has become possible to compute, for example, typical features of unconventional superconductivity in the Hubbard model in $d = 2$, such as the interplay of antiferromagnetism and d-wave pairing as well as pseudogap behavior [164], but also signatures of Anderson localization in disordered systems [169].

Diagrammatic generalizations By extending the DMFT on a diagrammatic level through the inclusion of non-local contributions, corrections to the local self-energy of the DMFT can be calculated explicitly. Here the dynamical vertex approximation (DΓA) [170] and the dual fermion theory [171] are powerful approaches. For example, they provided new insights into the mechanism of superconductivity arising from purely repulsive interactions, e.g., in the two-dimensional Kondo lattice model [172] and the Hubbard model [173]. In particular, in the repulsive Hubbard model a specific set of local particle-particle diagrams was identified which describe a strong screening of the bare interaction at low frequencies. Thereby antiferromagnetic spin fluctuations are suppressed, which in turn reduce the pairing interaction. Thus dynamical vertex corrections were found to reduce T_c strongly [173]. With these approaches one can also determine critical behavior, not only in the vicinity of thermal phase transitions ($T > 0$) [174] but also near quantum phase transitions ($T = 0$) [175].

Functional renormalization group (fRG) In this approach the fRG flow [176] does not start from the bare action of the system, but rather from the DMFT solution [177]. Local correlations are thus included from the beginning, and nonlocal correlations are generated by the fRG flow, as demonstrated for the two-dimensional Hubbard model [177].

8 Summary

In the absence of exact solutions of models in statistical and condensed matter physics in dimensions $d = 2$ and 3, mean-field solutions give indispensable, albeit approximate, information about the properties of these models. A well-established strategy for deriving a mean-field theory is the solution of a model in infinite dimensions or on a lattice with infinite coordination. In fact, solutions obtained in this way are sometimes referred to as "mean-field solution", even if the physical meaning of the "mean field" is not directly apparent.
In particular, the exact solution of the Hubbard model in $d = \infty$, which corresponds to a dynamical mean-field theory (DMFT) of correlated lattice fermions, has now become the generic mean-field theory of correlated electrons. It provides a comprehensive, non-perturbative and thermodynamically consistent approximation scheme for the investigation of correlated fermions, especially electrons in solids and cold fermionic atoms in optical lattices, in finite dimensions. Non-local extensions of the DMFT make it possible to explore and explain correlation effects which occur on the scale of several lattice constants and at thermal and quantum phase transitions. Furthermore, the combination of DMFT with methods for the computation of electronic band structures resulted in a conceptually new theoretical framework for the realistic study of correlated materials.
The further development of this approach and its applications is a subject of active research.

Acknowledgments

I thank Krzysztof Byczuk, Václav Janiš, Marcus Kollar, and Walter Metzner for discussions and useful comments.

References

[1] *The Oxford Dictionary of English Etymology* (Oxford University Press, 1966)

[2] J.J. Roche: *The Mathematics of Measurement: A Critical History* (Athlone, 1998); S.T. Bramwell, Nat. Phys. **13**, 820 (2017)

[3] M.E. Fisher and D.S. Gaunt, Phys. Rev. **133**, A224 (1964)

[4] R.J. Baxter: *Exactly Solved Models in Statistical Mechanics* (Academic Press, 1982)

[5] B.B. Mandelbrot: *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982)

[6] S.N. Kempkes, M.R. Slot, S.E. Freeney, S.J.M. Zevenhuizen, D. Vanmaekelbergh, I. Swart, and C. Morais Smith, Nat. Phys. **15**, 127 (2019)

[7] S.M. Bhattacharjee, in *Topology and Condensed Matter Physics*, eds. S.M. Bhattacharjee, M. Mj, A. Bandyopadhyay, Texts and Readings in Physical Sciences, vol. 19 (Springer, Singapore, 2017)

[8] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975); M.E. Fisher, Rev. Mod. Phys. **70**, 653 (1998); P.C. Hohenberg and A.P. Krekhov, Physics Reports, **572**, 1 (2015)

[9] K.G. Wilson and M.E. Fisher, Phys. Rev. Lett. **28**, 240 (1972)

[10] R. Brout, Phys. Rev. **118**, 1009 (1960)

[11] M. Kac, G.E. Uhlenbeck, and P.L. Hemmer, J. Math. Phys. **4**, 216 (1963); *ibid.* **4**, 229 (1963); *ibid.* **5**, 60 (1964)

[12] K. Huang, *Statistical Mechanics* (Wiley, New York, 1987)

[13] T.F. Gallagher: *Rydberg Atoms* (Cambridge University Press, 1994)

[14] S. Haroche, Rev. Mod. Phys. **85**, 1083 (2013)

[15] E. Witten, Phys. Today **33**, No. 7, 38 (1980)

[16] A. Svidzinsky, M. Scully, and D. Herschbach, Phys. Today **67**, No. 1, 33 (2014)

[17] P. Weiss, Comptes Rendus **143**, 1136 (1906)

[18] E. Ising, Z. Phys. **31**, 253 (1925)

[19] T. Ising, R. Folk, R. Kenna, B. Berche, and Y. Holovatch, Journ. Phys. Stud. **21** No. 3, 3002 (2017) [arXiv:1706.01764]

[20] S.F. Edwards and P.W. Anderson, J. Phys. F: Metal Phys. **5**, 965 (1975)

[21] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **32**, 1792 (1975)

[22] J.R.L. de Almeida, D.J. Thouless, J. Phys. **A** 11, 983 (1978)

[23] G. Parisi, Phys. Lett. A **73**, 203 (1979); G. Parisi, Phys. Rev. Lett. **43**, 1754 (1979)

[24] A. Billoire, Lect. Notes Phys. **736**, 11 (2008);
C. Middleton, Phys. Today **74**, No. 12, 17 (2021)

[25] G. Parisi and F. Zamponi, Rev. Mod. Phys. **82**, 789 (2010);
A. Manacorda, G. Schehr, and F. Zamponi, J. Chem. Phys. **152**, 164506 (2020)

[26] G. Parisi, P. Urbani, and F. Zamponi, *Theory of Simple Glasses: Exact Solutions in Infinite Dimensions* (Cambridge University Press, 2020)

[27] H.L. Frisch, N. Rivier, and D. Wyler, Phys. Rev. Lett. **54**, 2061 (1985); erratum *ibid.* **55**, 550 (1985); D. Wyler, N. Rivier, and H.L. Frisch, Phys. Rev. A **36**, 2422 (1987)

[28] H.L. Frisch and J.K. Percus, Phys. Rev. E **60**, 2942 (1999)

[29] W. Heisenberg, Z. Phys. **49**, 619 (1928)

[30] F. Bloch, Z. Phys. **57**, 545 (1929)

[31] J.C. Slater, Phys. Rev. **49**, 537 (1936)

[32] E. Wigner, Phys. Rev. **46**, 1002 (1934)

[33] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Nature **556**, 43 (2018); L. Balents, C.R. Dean, D.K. Efetov, and A.F. Young, Nat. Phys. **16**, 725 (2020); D.M. Kennes, M. Claassen, L. Xian, A. Georges, A.J. Millis, J. Hone, C.R. Dean, D.N. Basov, A.N. Pasupathy, and A. Rubio, Nat. Phys. **17**, 155 (2021); A. Ghiotto, E.-M. Shih, G.S.S.G. Pereira, D.A. Rhodes, B. Kim, J. Zang, A.J. Millis, K. Watanabe, T. Taniguchi, J.C. Hone, L. Wang, C.R. Dean, A.N. Pasupathy, Nature **597**, 345 (2021)

[34] V. Vaňo, M. Amini, S.C. Ganguli, G. Chen, J.L. Lado, S. Kezilebieke, and P. Liljeroth, Nature **599**, 582 (2021)

[35] N. Grewe and F. Steglich, in K.A. Gschneidner Jr. and L. Eyring (eds.): *Handbook on the Physics and Chemistry of Rare Earths*, vol. 14 (North Holland, 1991), p. 343

[36] J.R. Schrieffer (ed.): *Handbook of High-Temperature Superconductivity* (Springer, 2007)

[37] E. Dagotto: *Nanoscale Phase Separation and Collossal Magnetoresistance* (Springer, 2002)

[38] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

[39] H. v. Löhneysen, A. Rosch, M. Vojta, and P. Wölfle, Rev. Mod. Phys. **79**, 1015 (2007)

[40] Y. Tokura, Phys. Today **56**, No. 7, 50 (2003); E. Dagotto, Science **309**, 257 (2005)

[41] Z. Hennighausen and S. Kar, Electron. Struct. **3**, 014004 (2021)

[42] The Lecture Notes of the Jülich Autumn Schools on Correlated Electrons published by E. Pavarini, E. Koch and co-editors annually since 2011 provide up-to-date information on correlated electron systems and theoretical techniques for their investigation.
`https://www.cond-mat.de/events/correl.html`

[43] D. Pines: *The Many-Body Problem* (W.A. Benjamin, Inc., Reading, 1962)

[44] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)

[45] J. Hubbard, Proc. Roy. Soc. London A **276**, 238 (1963)

[46] J. Kanamori, Prog. Theor. Phys. **30**, 275 (1963)

[47] A. Montorsi (ed.): *The Hubbard Model – A Reprint Volume* (World Scientific, 1992)

[48] P.W. Anderson, Phys. Rev. **115**, 2 (1959)

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

[50] P.A. Wolff, Phys. Rev. **124**, 1030 (1961)

[51] I. Bloch, J. Dalibard, and W. Zwerger, Rev. Mod. Phys. **80**, 885 (2008)

[52] M. Kollar, R. Strack, and D. Vollhardt, Phys. Rev. B **53**, 9225 (1996)

[53] M.P.A. Fisher, P.B. Weichman, G. Grinstein, and D.S. Fisher, Phys. Rev. B **40**, 546 (1989); D. Jaksch, C. Bruder, J.I. Cirac, C.W. Gardiner, and P. Zoller, Phys. Rev. Lett. **81**, 3108 (1998)

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

[55] R. Peierls, Contemp. Phys. **21**, 3 (1980)

[56] W. Langer, M. Plischke, and Mattis, Phys. Rev. Lett. **23**, 1448 (1969);
K. Dichtel, R.J. Jelitto, and H. Koppe, Z. Phys. **246**, 248 (1971)

[57] E. Feenberg: *Theory of Quantum Fluids* (Academic, New York, 1969);
V.R. Pandharipande and R.B. Wiringa, Rev. Mod. Phys. **51**, 821 (1979);
P. Fulde: *Electron Correlations in Molecules and Solids* (Springer, 2002)

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

[59] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism*
(World Scientific, Singapore, 1999)

[60] T. Ogawa, K. Kanda, and T. Matsubara, Prog. Theor. Phys. **53**, 614 (1975)

[61] D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984)

[62] D. Vollhardt, in E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.): *DMFT at 25: Infinite Dimensions*, Modeling and Simulation, vol. 4 (FZ Jülich, 2014) https://www.cond-mat.de/events/correl14/manuscripts/

[63] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970)

[64] J. Hubbard, Proc. Roy. Soc. London A **281**, 401 (1964)

[65] R.J. Elliott, J.A. Krumhansl, and P.L. Leath, Rev. Mod. Phys. **46**, 465 (1974)

[66] P.W. Anderson and W.F. Brinkman, in K.H. Bennemann and J.B. Ketterson (eds.): *The Physics of Liquid and Solid Helium*, Part II (Wiley, New York, 1978), p. 177

[67] D. Vollhardt, P. Wölfle, and P.W. Anderson, Phys. Rev. B **35**, 6703 (1987)

[68] G. Kotliar and A.E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986)

[69] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **59**, 121 (1987); Phys. Rev. B **37**, 7382 (1988); erratum *ibid.* **39**, 12339 (1989)

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

[71] W. Metzner, Z. Phys. B **77**, 253 (1989)

[72] F. Gebhard, Phys. Rev. B **41**, 9452 (1990)

[73] J. Bünemann, F. Gebhard, and W. Weber, Found. Phys. **30**, 2011 (2000); T. Schickling, J. Bünemann, F. Gebhard, and W. Weber, New J. Phys. **16**, 93034 (2014)

[74] M. Acharyya, Eur. J. Phys. **31**, L89 (2010)

[75] P.G.J. van Dongen, F. Gebhard, and D. Vollhardt, Z. Phys. **76**, 199 (1989)

[76] D. Vollhardt, in A. Avella and F. Mancini (eds.): *Lectures on the Physics of Strongly Correlated Systems XIV*, *AIP Conference Proceedings*, vol. 1297 (American Institute of Physics, Melville, 2010), p. 339; http://arxiv.org/abs/1004.5069v3

[77] U. Wolff, Nucl. Phys. B **225**, 391 (1983)

[78] E. Müller-Hartmann, in E. Talik and J. Szade (eds.): *Proc. of the V. Symp. Phys. of Metals* (Poland, 1991), p. 22; G.S. Uhrig, Phys. Rev. Lett. **77**, 3629 (1996)

[79] M. Ulmke, Euro. Phys. J. B **1**, 301 (1998)

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

[81] V. Janiš, Z. Phys. B **83**, 227 (1991)

[82] V. Janiš and D. Vollhardt, Int. J. Mod. Phys. **B6**, 731 (1992)

[83] R. Vlaming and D. Vollhardt, Phys. Rev. B **45**, 4637 (1992)

[84] V. Janiš and D. Vollhardt Phys. Rev. B **46**, 15712 (1992)

[85] E. Müller-Hartmann, Z. Phys. B **76**, 211 (1989)

[86] H. Schweitzer and G. Czycholl, Solid State Commun. **69**, 171 (1989)

[87] U. Brandt and C. Mielsch, Z. Phys. B **75**, 365 (1989)

[88] P.G.J. van Dongen and D. Vollhardt, Phys. Rev. Lett. **65**, 1663 (1990)

[89] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960);
J.M. Luttinger, Phys. Rev. **119**, 1153 (1960)

[90] J.T. Heath and K.S. Bedell, New J. Phys. **22** 063011 (2020)

[91] A.C. Hewson: *The Kondo Problem to Heavy Fermions*
(Cambridge University Press, 1997)

[92] K. Byczuk and D. Vollhardt, Phys. Rev. **B 77**, 235106 (2008)

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

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

[95] J.E. Hirsch and R.M. Fye, Phys. Rev. Lett. **56**, 2521 (1986)

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

[97] E. Pavarini, E. Koch, A. Lichtenstein, and D. Vollhardt (eds.): *DMFT: From Infinite Dimensions to Real Materials*, Modeling and Simulation, vol. 8 (FZ Jülich, 2018)
`https://www.cond-mat.de/events/correl18/manuscripts/`

[98] G. Kotliar and D. Vollhardt, Phys. Today **57**, 53 (2004)

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

[100] R. Bulla, T.A. Costi, and Th. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

[101] K. Hallberg, D.J. García, P.S. Cornaglia, J.I. Facio, and Y. Núñez-Fernández,
EPL **112**, 17001 (2015)

[102] E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008);
Y. Lu and M.W. Haverkort, Eur. Phys. J. Special Topics **226**, 2549 (2017)

[103] A. Holzner, A. Weichselbaum, and J. von Delft, Phys. Rev. B **81**, 125126 (2010); F.A. Wolf, A. Go, I.P. McCulloch, A.J. Millis, and U. Schollwöck, Phys. Rev. X **5**, 041032 (2015); N.-O. Linden, M. Zingl, C. Hubig, O. Parcollet, and U. Schollwöck, Phys. Rev. B **101**, 041101(R) (2020)

[104] D. Bauernfeind, M. Zingl, R. Triebl, M. Aichhorn, and H.G. Evertz, Phys. Rev. X **7**, 031013 (2017)

[105] L.-F. Arsenault, A. Lopez-Bezanilla, O.A. von Lilienfeld, and A.J. Millis, Phys. Rev. B **90**, 155136 (2014); E. Sheridan, C. Rhodes, F. Jamet, I. Rungger, and C. Weber, Phys. Rev. B **104**, 205120 (2021)

[106] N.F. Mott, Proc. Phys. Soc. London, Ser. A **62**, 416 (1949); N.F. Mott, Rev. Mod. Phys. **40**, 677 (1968)

[107] N.F. Mott: *Metal–Insulator Transitions* (Taylor and Francis, London, 1990), 2nd ed.

[108] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

[109] R. Bulla, Phys. Rev. Lett. **83**, 136 (1999); M.J. Rozenberg, R. Chitra, and G. Kotliar, Phys. Rev. Lett. **83**, 3498 (1999); J. Joo and V. Oudovenko, Phys. Rev. B **64**, 193102 (2001); R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64**, 045103 (2001); N. Blümer, *Metal-Insulator Transition and Optical Conductivity in High Dimensions* (Shaker Verlag, Aachen, 2003)

[110] F. Krien, E.G.C.P. van Loon, M.I. Katsnelson, A.I. Lichtenstein, and M. Capone, Phys. Rev. B **99**, 245128 (2019)

[111] O. Gunnarsson and K. Schönhammer, Phys. Rev. B **31**, 4815 (1985); R. Bulla, A.C. Hewson, and Th. Pruschke, J. Phys.: Condens. Matter **10**, 8365 (1998)

[112] S. Sen, P.J. Wong, and A.K. Mitchell, Phys. Rev. B **102**, 081110 (2020)

[113] H. Terletska, J. Vučičević, D. Tanasković, and V. Dobrosavljević, Phys. Rev. Lett. **107**, 026401 (2011); M.M. Radonjić, D. Tanasković, V. Dobrosavljević, K. Haule, and G. Kotliar, Phys, Rev. B **85**, 085133 (2012); X. Deng, J. Mravlje, R. Žitko, M. Ferrero, G. Kotliar, and A. Georges, Phys. Rev. Lett. **110**, 086401 (2013)

[114] D.B. McWhan, A. Menth, J.P. Remeika, W.F. Brinkman, and T.M. Rice, Phys. Rev. B **7**, 1920 (1973); P. Limelette, A. Georges, D. Jerome, P. Wzietek, P. Metcalf, and J.M. Honig, Science **302**, 89 (2003); P. Hansmann, A. Toschi, G. Sangiovanni, T. Saha-Dasgupta, S. Lupi, M. Marsi, and K. Held, Phys. Status Solidi B **250**, 1251 (2013); J.C. Leiner, H.O. Jeschke, R. Valentí, S. Zhang, A.T. Savici, J.Y.Y. Lin, M.B. Stone, M.D. Lumsden, Jiawang Hong, O. Delaire, Wei Bao, and C.L. Broholm, Phys. Rev. X **9**, 011035 (2019)

[115] P. Limelette, P. Wzietek, S. Florens, A. Georges, T.A. Costi, C. Pasquier, D. Jérome, C. Mézière, and P. Batail, Phys. Rev. Lett. **91**, 016401 (2003); F. Kagawa, T. Itou, K. Miyagawa, and K. Kanoda, Phys. Rev. B **69**, 064511 (2004); A. Pustogow, M. Bories, A. Löhle, R. Rösslhuber, E. Zhukova, B. Gorshunov, S. Tomić, J.A. Schlueter, R. Hübner, T. Hiramatsu, Y. Yoshida, G. Saito, R. Kato, T.-H. Lee, V. Dobrosavljević, S. Fratini, and M. Dressel, Nat. Mater. **17**, 773 (2018); A. Pustogow, Y. Saito, A. Löhle, M.S. Alonso, A. Kawamoto, V. Dobrosavljević, M. Dressel, S. Fratini, Nat. Commun. **12**, 1571 (2021)

[116] V. Dobrosavljević and G. Kotliar, Phys. Rev. Lett. **78**, 3943 (1997); V. Dobrosavljević, A.A. Pastor, and B.K. Nikolić, Europhys. Lett. **62**, 76 (2003); S. Mahmoudian, S. Tang, and V. Dobrosavljević, Phys. Rev. B **92**, 144202 (2015)

[117] K. Byczuk, W. Hofstetter, and D. Vollhardt, Phys. Rev. Lett. **94**, 056404 (2005); ibid. **102**, 146403 (2009); K. Byczuk, W. Hofstetter, and D. Vollhardt, in E. Abrahams (ed.): *Fifty Years of Anderson Localization* (World Scientific, Singapore, 2010), p. 473 [reprinted in Int. J. Mod. Phys. B **24**, 1727 (2010)]

[118] D. Vollhardt, N. Blümer, K. Held, and M. Kollar, Lecture Notes in Physics **580**, 191 (Springer, Heidelberg, 2001)

[119] K. Byczuk and D. Vollhardt, Phys. Rev. B **65**, 134433 (2002)

[120] M.Z. Hasan and C.L. Kane, Rev. Mod. Phys. **82**, 3045 (2010)

[121] Y. Tada, R. Peters, M. Oshikawa, A. Koga, N. Kawakami, and S. Fujimoto, Phys. Rev. B **85**, 165138 (2012)

[122] T. Mertz, K. Zantout, and R. Valentí, Phys. Rev. B **100**, 125111 (2019)

[123] A.A. Markov, G. Rohringer, and A.N. Rubtsov, Phys. Rev. B **100**, 115102 (2019)

[124] B. Irsigler, T. Grass, J.-H. Zheng, M. Barbier, and W. Hofstetter, Phys. Rev. B **103**, 125132 (2021)

[125] S. Liang, M.-H. Jiang, W. Chen, J.-X. Li, and Q.-H. Wang, Phys. Rev. B **98**, 054433 (2018)

[126] T. Yoshida, R. Peters, and N. Kawakami, Phys. Rev. B **98**, 035141 (2018); Y. Nagai, Y. Qi, H. Isobe, V. Kozii, L. Fu, Phys. Rev. Lett. **125**, 227204 (2020)

[127] D. Krüger and M. Potthoff Phys. Rev. Lett. **126**, 196401 (2021)

[128] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, and P. Werner, Rev. Mod. Phys. **86**, 779 (2014)

[129] M. Ligges, I. Avigo, D. Golež, H.U.R. Strand, Y. Beyazit, K. Hanff, F. Diekmann, L. Stojchevska, M. Kalläne, P. Zhou, K. Rossnagel, M. Eckstein, P. Werner, and U. Bovensiepen, Phys. Rev. Lett. **120**, 166401 (2018); erratum *ibid.* **122**, 159901 (2019)

[130] J. Li, H.U.R. Strand, P. Werner, and M. Eckstein, Nat. Commun. **9**, 4581 (2018)

[131] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964);
W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)

[132] R.O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989)

[133] J.P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)

[134] V.I. Anisimov, J. Zaanen, and O.K. Andersen, Phys. Rev. B **44**, 943 (1991)

[135] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001)

[136] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar,
J. Phys.: Condens. Matter **9**, 7359 (1997)

[137] A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **57**, 6884 (1998)

[138] K. Held, I.A. Nekrasov, G. Keller, V. Eyert, N. Blümer, A.K. McMahan, R.T. Scalettar, Th. Pruschke, V.I. Anisimov, and D. Vollhardt, Psi-k Newsletter **56**, 65 (2003) [reprinted in Phys. Status Solidi B **243**, 2599 (2006)]; K. Held, Adv. Phys. **56**, 829 (2007)

[139] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[140] M.I. Katsnelson, V. Yu. Irkhin, L. Chioncel, A.I. Lichtenstein, and R.A. de Groot, Rev. Mod. Phys. **80**, 315 (2008)

[141] D. Vollhardt and A.I. Lichtenstein (eds.), *Dynamical Mean-Field Approach with Predictive Power for Strongly Correlated Materials*, Eur. Phys. J. Spec. Top. **226** (2017)

[142] D. Vollhardt, JPS Conf. Proc. **30**, 011001 (2020)

[143] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003); S. Biermann, J. Phys.: Condens. Matter **26**, 173202 (2014); J.M. Tomczak, P. Liu, A. Toschi, G. Kresse, and K. Held, Eur. Phys. J. Special Topics **226**, 2565 (2017)

[144] L. de' Medici, A. Georges, and S. Biermann, Phys. Rev. B **72**, 205124 (2005)

[145] E. Greenberg, I. Leonov, S. Layek, Z. Konopkova, M.P. Pasternak, L. Dubrovinsky, R. Jeanloz, I.A. Abrikosov, and G. Kh. Rozenberg, Phys. Rev. X **8**, 031059 (2018)

[146] J. Kim, D. Ryu, C.-J. Kang, K. Kim, H. Choi, T.-S. Nam, and B.I. Min,
Phys. Rev. B **100**, 195138 (2019)

[147] K. Haule and G. Kotliar, New. J. Phys. **11**, 025021 (2009); A. Georges, L de' Medici, and J. Mravlje, Ann. Rev. Cond. Matter Phys. **4**, 137 (2013); L. de'Medici, in *Iron-based Superconductivity*, Springer Series in Materials Science **211**, 409 (2015); D. Guterding, S. Backes, M. Tomić, H.O. Jeschke, R. Valentí, Phys. Status Solidi B **254**, 1600164 (2017)

[148] R. Adler, C.-J. Kang, C.-H. Yee, and G. Kotliar, Rep. Prog. Phys. **82**, 012504(2019)

[149] G.G. Pálsson and G. Kotliar, Phys. Rev. Lett. **80**, 4775 (1998); P. Wissgott, A. Toschi, H. Usui, K. Kuroki, and K. Held, Phys. Rev. B **82**, 201106(R) (2010); W. Xu, K. Haule, and G. Kotliar, Phys. Rev. Lett. **111**, 036401 (2013); J.M. Tomczak, J. Phys.: Condens. Matter **30**, 183001 (2018); M. Zingl, G.J. Kraberger, and M. Aichhorn, Phys. Rev. Materials **3**, 075404 (2019)

[150] M.J. Han, X. Wang, C.A. Marianetti, and A.J. Millis, Phys. Rev. Lett. **107**, 206804 (2011); J.K. Freericks, *Transport in multilayered nanostructures — The dynamical mean-field approach*, 2nd edition (Imperial College Press, London, 2016); Z. Zhong, M. Wallerberger, J.M. Tomczak, C. Taranto, N. Parragh, A. Toschi, G. Sangiovanni, and K. Held, Phys. Rev. Lett. **114**, 246401 (2015); R. Peters, Y. Tada, and N. Kawakami, Phys. Rev. B **94**, 205142 (2016); I. Di Marco, A. Held, S. Keshavarz, Y.O. Kvashnin, and L. Chioncel, Phys. Rev. B **97**, 035105 (2018); F. Lechermann, Phys. Rev. Materials **2**, 085004 (2018)

[151] H. Chen, A. Hampel, J. Karp, F. Lechermann, A.J. Millis, Front. Phys. **10**, 835942 (2022)

[152] D. Jacob, K. Haule, and G. Kotliar, Phys. Rev. B **82**, 195115 (2010); N. Lin, C.A. Marianetti, A.J. Millis, and D.R. Reichman, Phys. Rev. Lett. **106**, 096402 (2011); L. Chioncel, C. Morari, A. Östlin, W.H. Appelt, A. Droghetti, M.M. Radonjić, I. Rungger, L. Vitos, U. Eckern, and A.V. Postnikov, Phys. Rev. **B** 92, 054431 (2015); M. Schüler, S. Barthel, T. Wehling, M. Karolak, A. Valli, and G. Sangiovanni, Eur. Phys. J. Special Topics **226**, 2615 (2017); P. Pudleiner, A. Kauch, K. Held, G. Li, Phys. Rev. B 100, 075108 (2019)

[153] C. Weber, D.D. O'Regan, N.D.M. Hine, P.B. Littlewood, G. Kotliar, and M.C. Payne, Phys. Rev. Lett. **110**, 106402 (2013); M.A. al-Badri, E. Linscott, A. Georges, D.J. Cole, and C. Weber, Commun. Phys. **3**, 4 (2020)

[154] U. Schneider, L. Hackermüller, S. Will, Th. Best, I. Bloch, T.A. Costi, R.W. Helmes, D. Rasch, and A. Rosch, Science **322**, 1520 (2008)

[155] P. Sémon, K. Haule, and G. Kotliar, Phys. Rev. B **95**, 195115 (2017)

[156] A. Tamai, M. Zingl, E. Rozbicki, E. Cappelli, S. Riccò, A. de la Torre, S. McKeown Walker, F.Y. Bruno, P.D.C. King, W. Meevasana, M. Shi, M. Radović, N.C. Plumb, A.S. Gibbs, A.P. Mackenzie, C. Berthod, H. Strand, M. Kim, A. Georges, and F. Baumberger, Phys. Rev. X **9**, 021048 (2019)

[157] A. Georges and J.S. Yedidia, J. Phys. A: Math. Gen. **24**, 2173 (1991)

[158] P. Butera and M. Pernici, Phys. Rev. B **86**, 011139 (2012)

[159] M.E. Fisher and R.R.P. Singh, in G. Gimmet and D.J.A. Welsh (eds.): *Disorder in Physical Systems* (Oxford University Press, 1990), p. 87

[160] V. Janiš, Phys. Rev. Lett. **83**, 2781 (1999)

[161] V. Janiš and D. Vollhardt, Phys. Rev. B **63**, 125112 (2001)

[162] A. Schiller and K. Ingersent, Phys. Rev. Lett. **75**, 113 (1995)

[163] G. Zaránd, D.L. Cox, and A. Schiller, Phys. Rev. B **62**, R16227 (2000)

[164] Th. Maier, M. Jarrell, Th. Pruschke, and M.H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005)

[165] G. Rohringer, H. Hafermann, A. Toschi, A.A. Katanin, A.E. Antipov, M.I. Katsnelson, A.I. Lichtenstein, A.N. Rubtsov, and K. Held, Rev. Mod. Phys. **90**, 025003 (2018)

[166] Q. Si and J.L. Smith (1996), Phys. Rev. Lett. **77**, 3391 (1996); J.L. Smith and Q. Si, Phys. Rev. B **61**, 5184 (2000); R. Chitra and G. Kotliar, Phys. Rev. Lett. **84**, 3678 (2000); P. Sun and G. Kotliar, Phys. Rev. B **66**, 085120 (2002)

[167] M.H. Hettler, A.N. Tahvildar-Zadeh, M. Jarrell, Th. Pruschke, and H.R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998)

[168] A.I. Lichtenstein, and M.I. Katsnelson, Phys. Rev. B **62**, R9283 (2000); G. Kotliar, S.Y. Savrasov, G. Pálsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001)

[169] C.E. Ekuma, H. Terletska, K.-M. Tam, Z.-Y. Meng, J. Moreno, and M. Jarrell, Phys. Rev. B **89**, 081107(R) (2014); H. Terletska, Y. Zhang, K.M. Tam, T. Berlijn, L. Chioncel, N.S. Vidhyadhiraja, and M. Jarrell, Appl. Sci. **8**, 2401 (2018)

[170] A. Toschi, A.A. Katanin, and K. Held, Phys. Rev. B **75**, 045118 (2007)

[171] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **77**, 033101 (2008)

[172] J. Otsuki, Phys. Rev. Lett. **115**, 036404 (2015)

[173] M. Kitatani, T. Schäfer, H. Aoki, and K. Held, Phys. Rev. B **99**, 041115(R) (2019)

[174] T. Schäfer, A.A. Katanin, K. Held, and A. Toschi, Phys. Rev. Lett. **119**, 046402 (2017)

[175] D. Hirschmeier, H. Hafermann, and A.I. Lichtenstein, Phys. Rev. B **97**, 115150 (2018); T. Schäfer, A.A. Katanin, M. Kitatani, A. Toschi, and K. Held, Phys. Rev. Lett. **122**, 227201 (2019)

[176] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer, Rev. Mod. Phys. **84**, 299 (2012)

[177] C. Taranto, S. Andergassen, J. Bauer, K. Held, A. Katanin, W. Metzner, G. Rohringer, and A. Toschi, Phys. Rev. Lett. **112**, 196402 (2014); D. Vilardi, C. Taranto, and W. Metzner, Phys. Rev. B **99**, 104501 (2019)

2 Fermi Liquids

Giovanni Vignale
University of Missouri
Columbia, MO 65211

Contents

1 A tale of many fermions **2**

2 Phenomenological theory **6**
2.1 The Landau energy functional 6
2.2 The heat capacity 8
2.3 The Landau parameters 8
2.4 Compressibility and spin susceptibility 9
2.5 Galilean invariance and effective mass 11
2.6 Measuring m^*, K and χ_S 13

3 The lifetime of quasiparticles **15**
3.1 General formulas 15
3.2 Three-dimensional electron gas 17
3.3 Two-dimensional electron gas 19
3.4 Measuring the quasiparticle lifetime 21
3.5 The kinetic equation 22

4 Microscopic basis of the Landau theory of Fermi liquids **26**
4.1 Existence of quasiparticles and self-energy 26
4.2 Landau interaction function and scattering amplitude 28

5 Fermi liquid of massless Dirac fermions **30**

6 Non-Fermi-liquid behavior **33**
6.1 Disordered electron liquid 33
6.2 Luttinger liquid 34
6.3 Fractional quantum Hall liquid 35

1 A tale of many fermions

One of the triumphs of quantum mechanics in the 20th century was the explanation of the atomic structure (periodic table of the elements) in terms of elementary fermions (electrons) endowed with intrinsic spin $\hbar/2$ and obeying Pauli's exclusion principle, such that no two fermions can occupy the same quantum state. When many fermions (N) are present in a finite region of space the Pauli exclusion principle creates a kind of effective repulsion between the particles: this repulsion exists even if we pretend that there are no physical interactions between the particles, e.g., in the limiting case of the ideal Fermi gas. Under such conditions the minimum energy state is attained by singly occupying the N lowest lying single-particle states: this maximally compact state is known (for large N) as the "Fermi sea", and the energy of the highest occupied state is known as the Fermi energy, denoted by ϵ_F. When many interacting fermions condense in a state close to the minimum energy allowed by the Pauli exclusion principle (i.e., such that the average occupation numbers of the N lowest single-particle states are close to 1) one obtains a *Fermi liquid*. Physical realizations of the Fermi liquids concept range from interacting electrons in metals and semiconductors (our main interest here), to liquid ^{3}He, to gases of cold fermionic atoms, to nuclear matter, electrons in white dwarves, and neutron stars.

The behavior of Fermi liquids confronts us at the outset with a puzzle. In spite of strong mutual interactions the particles appear in many measurements to behave as if they were essentially non-interacting. Many properties of electrons in metals, for example heat capacity and electric conductivity, can be qualitatively understood in terms of the Sommerfeld picture, which is based on the degenerate ideal Fermi gas model. But this model is not easily justified. The average interaction energy per electron can be roughly estimated to be on the order of $e^2 n^{\frac{1}{3}}$, where $n = N/V$ is the average electronic density and V is the volume. For electrons in Na this works out to be $\simeq 4.29$ eV, which is already larger than the value $\epsilon_F \simeq 3.24$ eV of the Fermi energy. There is no sense in which the Coulomb interaction can be considered a small perturbation, and yet the electrons behave by and large as noninteracting particles, while the presence of the interaction manifests itself in rather subtle ways.

It was not until the late 1950s that this puzzling state of affairs was clarified from a theoretical point of view by L.D. Landau [1]. Although Landau did not provide a rigorous solution of the problem, he did provide a firm basis for the understanding of the "normal" low-energy behavior of interacting Fermi systems. The solidity of this basis was subsequently confirmed by theoretical and experimental work.

Perhaps the most striking feature of Landau's approach is that he completely sidestepped what most physicists would have considered a prerequisite for further progress, namely a complete description of the interacting ground state. Instead, he focused on the excited states. His basic idea was that, under very broad conditions, the low-lying excitations of a system of interacting Fermions with repulsive interactions are connected to the low-lying states of a non-interacting ideal Fermi liquid by a suitably slow switching-on of the interaction between the particles. There are several subtleties in the specification of the switching-on process, beginning with a precise definition of the words "suitably slow". For the time being we will not delve into these

subtleties, but simply notice that the switching-on procedure establishes a one-to-one correspondence between the eigenstates of the ideal system and a set of (approximate) eigenstates of the interacting system. Since the eigenstates of the noninteracting system are specified by a set of occupation numbers $\{\mathcal{N}_{\vec{k}\sigma}\}$ of single-particle momentum eigenstates, it follows that the corresponding low-lying excitations of the interacting system can also be described by the same set of occupation numbers. It is important to understand that the quantum numbers $\mathcal{N}_{\vec{k}\sigma}$ that specify an excited state of the Fermi liquid are not the true momentum occupation numbers $n_{k\sigma} = \langle \hat{a}^\dagger_{\vec{k}\sigma} \hat{a}_{\vec{k}\sigma} \rangle$ for that state. Rather, they are momentum occupation numbers of the ideal system from which the excited state has evolved.

Because in an interacting system the $n_{\vec{k}\sigma}$'s are not constants of the motion, one could have naively expected that any memory of the initial noninteracting momentum distribution $\{\mathcal{N}_{\vec{k}\sigma}\}$ would be completely lost at the end of the switching-on process. Landau's insight was the recognition that, for states that are weakly excited (i.e., close to the noninteracting Fermi distribution) the occupation numbers change very slowly even when particle-particle interactions are strong. The main consequence of this fact is that the $\mathcal{N}_{\vec{k}\sigma}$ retain their validity as approximate quantum numbers, which specify an excited state. Thus, low-energy elementary excitations of an interacting Fermi liquid can be described in terms of addition or removal of individual quasiparticles from a filled Fermi sphere of radius k_F, where k_F is the Fermi momentum of a non interacting electron gas of the same density. In other words, the interacting system has a Fermi surface that coincides with that of the non-interacting system – a statement that is known as *Luttinger's theorem*. For example, a state of the ideal system containing a particle of momentum $\hbar\vec{k}$ with $k \geq k_F$ outside the non-interacting Fermi sphere evolves into an excited state of the interacting system containing one quasiparticle of momentum $\hbar\vec{k}$ outside the interacting Fermi sphere. Likewise, a state of the noninteracting system containing one empty state (a hole) of momentum $\hbar\vec{k}$ within the non-interacting Fermi sphere evolves into an excited state of the interacting system containing a quasihole of the same momentum. More complex excitations consisting of multiple quasiparticles and quasiholes can be constructed in a similar fashion. By definition, the ground-state has $\mathcal{N}_{\vec{k}\sigma} = \Theta(k_F - k)$. In contrast to this, the momentum occupation numbers $n_{\vec{k}\sigma}$ in the ground-state decrease discontinuously by an amount Z (with $0 < Z \leq 1$) as $\vec{k}$ crosses the Fermi sphere from inside to outside. The discontinuity, Z, is of course 1 in the noninteracting system and less than 1 in the interacting system. The existence of a discontinuity in the momentum occupation number at $k = k_F$ is one of the distinctive signatures of the normal Fermi liquid.

The physical basis of the Landau theory rests on the surprising ineffectiveness of electron-electron scattering to change the momentum distribution of quasiparticles near the Fermi level. What happens is that most of the states into which two quasiparticles near the Fermi surface might end up after a collision are already occupied by other electrons, and therefore, according to the Pauli exclusion principle, unavailable. Because of this "Pauli blocking" effect, which operates irrespective of the strength of the interaction, the rate at which a quasiparticle is scattered out of a state of momentum $k \simeq k_F$ vanishes for $k \to k_F$. This result can be obtained from a simple phase space argument.

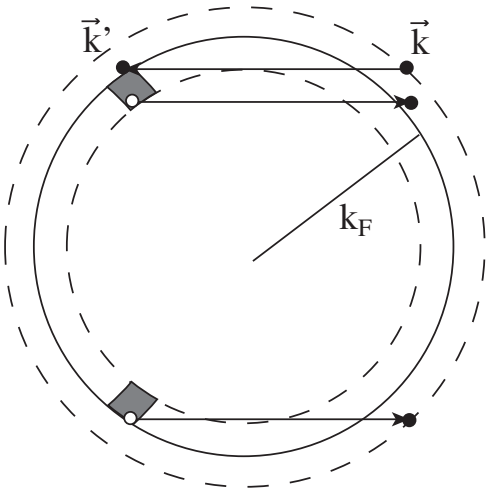


Fig. 1: *Schematic illustration of two possible decay processes for a quasiparticle near the Fermi surface. The quasiparticle makes a transition from $\vec{k}$ to $\vec{k}'$, producing an electron-hole pair in the process. Momentum and energy conservation restrict the momentum of the hole to the shaded regions.*

Consider a quasiparticle with initial wave vector $\vec{k}$ with $k > k_F$. At zero temperature the empty states into which the quasiparticle can decay lie within a shell of thickness $|k-k_F|$ just above the Fermi surface. The number of states in this region is clearly proportional to $|k-k_F|$ – a result valid in one, two and three dimensions. Now, through the Coulomb interaction, the momentum and energy change of the quasiparticle will be offset by the momentum and energy of an electron-hole pair. In two and three dimensions, the state of the hole must lie within a shell of thickness $|k - k_F|$ below the Fermi surface (see Fig. 1). This contributes another factor of $|k-k_F|$ to the probability of decay, which, as anticipated, is thus found to be proportional to $(k-k_F)^2$ in three dimensions.[1]

Accordingly in two and three dimensions in the limit $k \to k_F$, the inverse of the scattering rate, i.e., the scattering time $\tau_{\vec{k}}$, becomes long enough to include many cycles of oscillation of an external field that is able to create the quasiparticle excitation out of the ground-state (the frequency of this field being proportional to the excitation energy which is of order $|k-k_F|$). Thus, on a time scale that is short compared to τ_k (but still long compared to the inverse excitation frequency) the occupation number $\mathcal{N}_{\vec{k}\sigma}$ can be regarded as a good quantum number for the excited state.[2]

Strictly speaking however the $\mathcal{N}_{\vec{k}\sigma}$ are not exact quantum numbers (for that to be true the scattering rate would have to actually vanish), and if one waits long enough, i.e., up to times $t \gg \tau_{\vec{k}}$,

[1] In two dimensions, a more accurate calculation shows that the scattering rate vanishes at a somewhat slower rate $-(k-k_F)^2 \ln|k-k_F|$ (see Section 3.3). In the presence of disorder the quasiparticle inelastic lifetime is shorter. This is discussed in Section 6.1.

[2] The alert reader will notice that the phase-space argument is circular: one assumes the existence of quasiparticles to deduce that their lifetime is long. This does not prove the existence of quasiparticles, but shows that one can assume their existence without falling in a contradiction. This discussion does also clarify that a "suitably slow" switching-on must be carried out in a time intermediate between the fast time scale of order $\frac{1}{v_F|k-k_F|}$ related to the resolution of a quasiparticle state and the quasiparticles lifetime which we have shown to be of order $\frac{\epsilon_F}{\hbar(v_F|k-k_F|)^2}$.

one will see them change. Thus a quasiparticle state is *not* an exact eigenstate of the interacting Fermi liquid. Rather, it must be understood as a superposition of closely spaced exact eigenstates with energies spread over a range of width $\hbar/\tau_{\vec{k}} \ll \varepsilon_{\vec{k}}$ about the median quasi-particle energy $\varepsilon_{\vec{k}}$. Such a state decays with a characteristic lifetime $\tau_{\vec{k}}$, and can be regarded as stationary only for times much shorter than $\tau_{\vec{k}}$. It turns out that these "quasi-eigenstates" completely determine the response of the system to macroscopic perturbations.

So strong has been the impact of the Landau theory in condensed matter physics that the Landau Fermi liquid is often referred to as the Fermi liquid *tout-court*. It is important therefore to keep in mind that there are situations in which the Landau description of the Fermi liquid fails completely. A well known case is that of one-dimensional electronic systems. In this case the spectral density of electron-hole pairs exhibits a peak at small wavevector q and frequency $\omega = v_F q$. All the electron-hole pairs with given momentum $\hbar q$ have essentially the same energy comprised between $\hbar|\omega_-(q)| = \left|\hbar v_F q - \frac{\hbar^2 q^2}{2m}\right|$ and $\hbar\omega_+(q) = \hbar v_F q + \frac{\hbar^2 q^2}{2m}$. Due to the massive quasi-degeneracy of the noninteracting spectrum, an arbitrarily weak interaction causes a complete reconstruction of the many-body eigenstates, and, in particular, destroys the "Fermi surface" (actually, two points in 1D). Thus, there is no Landau Fermi liquid in one dimension. The correct paradigm in this case is the *Luttinger liquid*, where the low-energy excitations are collective charge and spin density waves and single-particle excitations are no longer sharply defined. Whether similar departures from Fermi liquid theory can also occur in two dimensions remains a very active and contentious area of research.

In spite of its occasional failures, the concept of a quasiparticle is an encompassing one. It applies not only to electrons in metals and doped semiconductors (where the renormalization factors $\frac{m^*}{m}$ and Z remain close to 1) but also to ^{3}He atoms in the liquid phase where $\frac{m^*}{m} \simeq 3$ and to highly correlated heavy fermion systems where $\frac{m^*}{m}$ can run in the hundreds. Considering the diversity in coupling constants and physical character of these systems it is quite amazing that they can be subsumed under the same generic theoretical paradigm.

This Chapter is organized as follows:

In Section 2 we summarize the basic results of the phenomenological theory of the Fermi liquids due to and mostly developed by Landau.

Section 3 presents a simple theory of the quasiparticle inelastic lifetime based on the Fermi golden rule of elementary quantum mechanics.

The microscopic underpinning of the Landau theory is briefly reviewed in Section 4 and some numerical results for quasiparticle properties are presented.

Section 5 presents the Fermi liquid theory for massless Dirac Fermions in graphene.

Finally, Section 6 illustrates some ways in which the standard theory of Fermi liquids may fail due to reduced dimensionality, disorder effects, and high magnetic fields.

The material presented in this chapter is largely adapted from the discussion of the normal Fermi liquid theory in "Quantum theory of the electron liquid" (G.F. Giuliani and G. Vignale, Cambridge University Press, 2005) [2] and in "The Theory of Quantum Liquids, Vol. I" (D. Pines and P. Noziéres, W.A. Benjamin 1966) [3], to which you are referred for more detailed derivations and discussions.

2 Phenomenological theory

2.1 The Landau energy functional

As discussed in the previous section, a set of noninteracting occupation numbers $\mathcal{N}_{k\sigma} = 0$ or 1 defines, by continuation, a quasi-eigenstate of the interacting Fermi liquid. Similarly, a distribution of fractional occupation numbers $0 \leq \mathcal{N}_{k\sigma} \leq 1$ defines an ensemble of quasiparticle states in which the state of momentum $\hbar\vec{k}$ and spin σ has a probability $\mathcal{N}_{k\sigma}$ to be occupied and $1-\mathcal{N}_{k\sigma}$ to be empty. At the heart of Landau's macroscopic theory of the Fermi liquids lies an *Ansatz* for the functional dependence of the energy of the liquid on the quasiparticle distribution function $\mathcal{N}_{\vec{k}\sigma}$. This functional is in fact an expansion for the energy to second order in the deviation of the quasiparticle distribution function from its ground-state value $\mathcal{N}^{(0)}_{\vec{k}\sigma} = \Theta(k_F - k)$:

$$E\big[\{\mathcal{N}_{\vec{k}\sigma}\}\big] = E_0 + \sum_{\vec{k}\sigma} \mathcal{E}_{\vec{k}\sigma}\, \delta\mathcal{N}_{\vec{k}\sigma} + \frac{1}{2}\sum_{\vec{k}\sigma,\vec{k}'\sigma'} f_{\vec{k}\sigma,\vec{k}'\sigma'}\, \delta\mathcal{N}_{\vec{k}\sigma}\delta\mathcal{N}_{\vec{k}'\sigma'}\ , \tag{1}$$

where E_0 is the ground-state energy (which needs not be specified!), $\mathcal{E}_{\vec{k}\sigma}$ is the energy of a single quasiparticle, $f_{\vec{k}\sigma,\vec{k}'\sigma'}$ is the Landau interaction function and $\delta\mathcal{N}_{\vec{k}\sigma} = \mathcal{N}_{\vec{k}\sigma} - \mathcal{N}^{(0)}_{\vec{k}\sigma}$ is the deviation of the quasiparticle distribution function from the ideal Fermi distribution at $T = 0$.
Because the quasiparticles are well defined only in the immediate vicinity of the Fermi surface, it is evident that this expansion makes sense only when $\delta\mathcal{N}_{\vec{k}\sigma}$ is restricted to a thin shell of momentum space surrounding the Fermi surface. In addition, since every wave vector sum introduces a factor L^d (the d-dimensional volume), the interaction function $f_{\vec{k}\sigma,\vec{k}'\sigma'}$ must scale as the inverse of the volume $\frac{1}{L^d}$ in order to keep the energy proportional to the volume in the thermodynamic limit.
Both the quasiparticle energy and the interaction function (as well as the Landau parameters introduced below) are phenomenological quantities that the Landau theory of Fermi liquid assumes to be given. In practice, they must be either determined from measurements of physical properties, or calculated by a microscopic many-body theory.
The energy $\mathcal{E}_{\vec{k}\sigma}$ of a single quasiparticle can be formally viewed as the functional derivative of the energy with respect to the quasiparticle distribution function evaluated at the ground-state:

$$\mathcal{E}_{\vec{k}\sigma} = \left(\frac{\delta E}{\delta\mathcal{N}_{\vec{k}\sigma}}\right)_{\mathcal{N}_{\vec{k}\sigma}=\mathcal{N}^{(0)}_{\vec{k}\sigma}}\ . \tag{2}$$

Since the ground-state of the $N+1$-particle system is obtained by adding a quasiparticle of wavevector k_F to the ground state of the N-particle system, it is evident that

$$\mathcal{E}_{k_F\sigma} = \mu\ , \tag{3}$$

where μ is the chemical potential.
In an isotropic liquid, for $|\vec{k}|$ close to k_F, the quasiparticle energy can be expanded as

$$\mathcal{E}_{\vec{k}\sigma} \simeq \mu + \hbar v^*_F\,(k - k_F), \tag{4}$$

where

$$v_F^\star \;=\; \frac{1}{\hbar}\left|\frac{\partial \mathcal{E}_{\vec{k}\sigma}}{\partial \vec{k}}\right|_{k=k_F} \tag{5}$$

defines the effective Fermi velocity of a quasiparticle. v_F^* can be conveniently written as

$$v_F^* \;=\; \frac{\hbar k_F}{m^\star}\;, \tag{6}$$

which defines the quasiparticle effective mass $m^\star$. The effective mass in turn determines $N^\star(0)$, the quasiparticle density of states (per unit volume) at the Fermi level μ. This is given by $N^\star(0) \;=\; \frac{m^\star}{m}N(0)$, where $N(0)$ is the density of states at the Fermi surface of a non interacting electron gas, $N(0) = \frac{nd}{2\epsilon_F}$ in d-dimensions.

A fundamental role in the Landau theory is played by the quantity

$$\tilde{\mathcal{E}}_{\vec{k}\sigma} = \frac{\delta E}{\delta \mathcal{N}_{\vec{k}\sigma}} = \mathcal{E}_{\vec{k}\sigma} + \sum_{\vec{k}'\sigma'} f_{\vec{k}\sigma,\vec{k}'\sigma'}\,\delta\mathcal{N}_{\vec{k}'\sigma'}\;, \tag{7}$$

sometimes referred to as the *local quasiparticle energy*. This can be interpreted as the energy of a quasiparticle modified by its interaction with other quasiparticles. From the form of this equation it is clear that, within the Landau theory, this effect is treated in a mean field approximation.

We next turn our attention to the quasiparticle interaction function. An inspection of Eq. (1) reveals that this quantity can be expressed in terms of functional derivatives of the Landau energy functional with respect to the quasiparticle distribution function as

$$f_{\vec{k}\sigma,\vec{k}'\sigma'} = \frac{\delta^2 E}{\delta\mathcal{N}_{\vec{k}\sigma}\delta\mathcal{N}_{\vec{k}'\sigma'}} = \frac{\delta\tilde{\mathcal{E}}_{\vec{k}\sigma}}{\delta\mathcal{N}_{\vec{k}'\sigma'}}\;, \tag{8}$$

where the functional derivatives are evaluated at the ground-state distribution. Notice that in order to correctly perform the second derivative appearing in Eq. (8), one needs to know the energy functional $E[\{\mathcal{N}_{\vec{k}\sigma}\}]$ up to second order in $\delta\mathcal{N}_{\vec{k}\sigma}$. This implies, for instance, that to derive the expression for the interaction function in a paramagnetic system one needs the knowledge of the energy functional appropriate to an infinitesimally polarized electron liquid. This complication does not arise in the case of the quasiparticle energy, since its calculation only requires a knowledge of the Landau energy functional up to first order in $\delta\mathcal{N}_{\vec{k}\sigma}$. Finally, in order to calculate thermal properties at finite temperature T one also needs an expression for the entropy of the liquid. This is given by

$$S\big[\{\mathcal{N}_{\vec{k}\sigma}\}\big] = -k_B\sum_{\vec{k}\sigma}\left(\mathcal{N}_{\vec{k}\sigma}\ln\mathcal{N}_{\vec{k}\sigma} + (1-\mathcal{N}_{\vec{k}\sigma})\ln\left(1-\mathcal{N}_{\vec{k}\sigma}\right)\right)\;, \tag{9}$$

which coincides with the entropy of the noninteracting ensemble of origin, and vanishes in the ground-state. This is a direct consequence of the assumed one-to-one correspondence between states of the interacting and non-interacting systems.

Eqs. (1) and (9) are widely used to calculate, from a macroscopic point of view, the thermal equilibrium properties, the response functions, and the transport properties of an interacting Fermi liquid, and to establish relationships between different such properties. What follows is a summary of the main results.

2.2 The heat capacity

As it turns out, the low-temperature specific heat of a Fermi liquid coincides with that of a noninteracting Fermi gas comprised of particles of mass m^*: it is therefore given by

$$c_v(T) = \frac{\pi^2}{3} N^*(0) L^d k_B^2 T \,, \tag{10}$$

where, as we have seen, $N^*(0)$, the density of quasiparticle energy states per unit volume at the Fermi level, differs from the corresponding quantity in the non interacting case by the substitution of the bare electronic mass m with m^*.[3] This is a direct consequence of Eq. (9) for the entropy, which in turn implies that the quasiparticle distribution function at thermal equilibrium is given by

$$\mathcal{N}^{\rm eq}_{\vec{k}\sigma}(\mu, T) = \frac{1}{e^{\beta(\mathcal{E}_{\vec{k}\sigma} - \mu)} + 1} \,, \tag{11}$$

where $\beta = 1/k_B T$. Notice that the Landau interaction function does not appear in this expression. This is because the thermal excitation of the system does not contribute to the quasiparticle energy. Thus the effective mass can in principle be directly measured from the heat capacity, or by any other measurement that is sensitive only to the quasiparticle density of states. The situation is quite different when the excitation is caused by an external field such as pressure or magnetic field, as we show next.

2.3 The Landau parameters

It is useful to introduce at this point the *Landau Fermi liquid parameters*. One starts from the observation that within the dynamically relevant shell in which $\delta\mathcal{N}_{\vec{k}'\sigma'}$ is finite, the Landau interaction function depends only on the cosine of the angle θ between $\vec{k}$ and $\vec{k}'$. Accordingly we can set $f_{\vec{k}\sigma,\vec{k}'\sigma'} \simeq f_{\sigma\sigma'}(\cos\theta)$ and introduce the dimensionless quantities $F_\ell^{s,a}$ which are defined in terms of spin symmetric (s) and spin antisymmetric (a) angular averages of $f_{\sigma\sigma'}(\cos\theta)$ as follows

$$F_\ell^{s,a} \;=\; \frac{L^d N^*(0)}{2} \int \frac{d\Omega_d}{\Omega_d} \big(f_{\uparrow\uparrow}(\cos\theta) \pm f_{\uparrow\downarrow}(\cos\theta) \big) \begin{cases} P_\ell(\cos\theta) & , \; 3D \\ \cos \ell\theta & , \; 2D \end{cases} \,, \tag{12}$$

where the $+$ and $-$ signs are associated with s and a respectively, $\Omega_d = 2^{d-1}\pi$ is the solid angle in $d = 3$ or 2 dimensions, and $P_\ell(\cos\theta)$ is the ℓ-th Legendre polynomial.

It must be noted that this definition (introduced in Ref. [2]) differs from the one commonly used in previous texts and in large part of the literature. Nervous readers can revert to the standard notation by simply making the substitution $F_\ell^{s,a} \to \frac{F_\ell^{s,a}}{2\ell+1}$ in three dimensions and $F_\ell^{s,a} \to \frac{F_\ell^{s,a}}{2}(1+\delta_{\ell 0})$ in two dimensions.

[3] We are considering here only the mass renormalization that arises from interactions between the particles. In a crystalline environment the crystal potential and the electron-phonon coupling produce additional mass renormalizations.

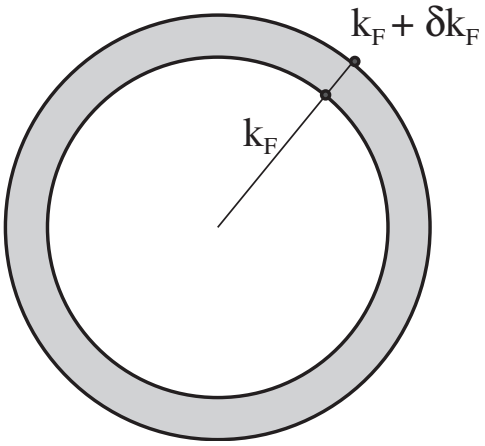


Fig. 2: *Calculation of the compressibility in the Landau theory of Fermi liquids. The chemical potential, i.e., the energy of a quasiparticle at the Fermi surface, changes due to (i) the variation of the Fermi momentum, (ii) the addition of quasiparticles to the shaded region.*

The inverse of Eq. (12) is

$$f_{\uparrow\uparrow}(\cos\theta) \pm f_{\uparrow\downarrow}(\cos\theta) \;=\; \frac{2}{L^d N^*(0)} \sum_{\ell=0}^{\infty} F_\ell^{s,a} \begin{cases} (2\ell{+}1) P_\ell(\cos\theta) & , \; 3D \\ (2{-}\delta_{\ell 0}) \cos\ell\theta & , \; 2D \end{cases} , \tag{13}$$

where $+$ is associated with s and $-$ with a.

2.4 Compressibility and spin susceptibility

An important property of a Fermi liquid is the proper compressibility K,[4] given by the relation

$$\frac{1}{K} \;=\; n^2 \frac{\partial\mu}{\partial n} \;=\; \frac{n k_F}{d} \frac{\partial\mu}{\partial k_F} \, . \tag{14}$$

The compressibility determines, among other things, the magnitude of the screening wave vector and the hydrodynamic sound velocity $s = \frac{1}{\sqrt{nmK}}$. For a non interacting system one simply has the result

$$K_0 \;=\; \frac{N(0)}{n^2} \, . \tag{15}$$

To evaluate the derivative $\frac{\partial\mu}{\partial k_F}$ within the Landau theory of Fermi liquids one must recall that, according to Eq. (3), the change in μ as the Fermi surface expands to accommodate the additional density δn is the sum of two terms: one is the change in the bare quasiparticle energy when the wave vector varies from k_F to $k_F{+}\delta k_F$; the other is the interaction energy with the additional quasiparticles created by the expansion of the Fermi sphere (see Fig. 2). The first term is responsible for changing the density of states in Eq. (15) from $N(0)$ to $N^*(0)$, but the

[4]For a charged Fermi liquid the *proper* compressibility is calculated under the assumption that the system remains charge-neutral during the compression and there is no energy cost associated with the compression of the neutralizing background of charge.

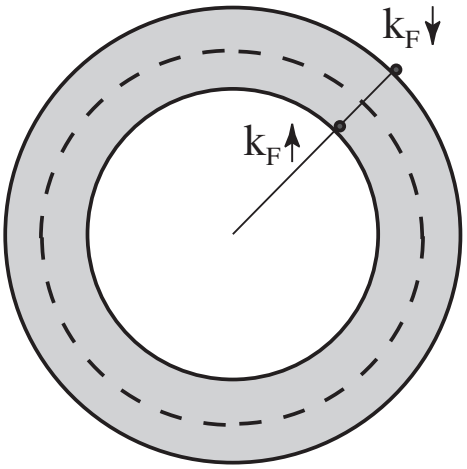


Fig. 3: *Calculation of the spin susceptibility in the Landau theory of Fermi liquids. In the presence of a magnetic field the up- and down-spin Fermi surfaces split in such a way that the energies of two quasiparticles at the two Fermi surfaces are equal. The additional quasiparticles are down-spin electrons above the unperturbed Fermi surface (dashed line) and up-spin holes below it.*

second term involves the Landau parameter F_0^s, since the additional quasiparticle distribution is spherically symmetric ($\ell = 0$) and spin-independent (superscript s). The result is

$$\frac{K}{K_0} = \frac{m^\star/m}{1+F_0^s} \,. \tag{16}$$

Thus, the interaction enters the proper compressibility not only through the effective mass, but also, explicitly, through the spin symmetric spherical average of the Landau interaction function.

The spin susceptibility can be calculated in a completely analogous way (see Fig. 3). In the presence of an external magnetic field B the Hamiltonian is modified by the addition of the Zeeman energy term

$$\hat{H}_Z = \frac{g\mu_B}{2} B \hat{S}_z \,, \tag{17}$$

where g is the *noninteracting* g-factor for the electrons and $\hat{S}_z$ is the z-component of the spin in units of $\frac{\hbar}{2}$. Here $\frac{g\mu_B}{2}$ is the magnetic moment of the electron.[5] The spin susceptibility is defined as the derivative of the magnetization with respect to the magnetic field at zero magnetic field. In an ideal Fermi gas it works out to be

$$\chi_{S0} = \left(\frac{g\mu_B}{2}\right)^2 N(0) \,. \tag{18}$$

In a Fermi liquid the energy of a quasiparticle of wave vector $\vec{k}$ and spin σ in the presence of the magnetic field becomes

$$\mathcal{E}_{\vec{k}\sigma}(B) = \mathcal{E}_{\vec{k}\sigma} + \frac{1}{2} g\mu_B B\sigma \,, \tag{19}$$

where, it must be noted, the g-factor of the quasiparticle coincides with that of the bare electron (as long as spin-orbit interactions are neglected), because the many-body state described by the

[5] For free electrons $g \simeq 2$, but this value can be considerably different for electrons in a solid state environment, due to the spin-orbit interaction: for example in GaAs one has $g = -0.44$.

quasiparticle at $\vec{k}\sigma$ is an eigenstate of $\hat{S}_z$ with eigenvalue $\hbar\sigma/2$. Because of the Zeeman energy, the Fermi surfaces of up-spin and down-spin electrons shift by equal amounts in opposite directions, i.e., the radius of the down-spin Fermi surface increases by an amount $\delta k_{F\downarrow}$, while the radius of the up-spin Fermi surface decreases by the same amount.

The equilibrium value of $\delta k_{F\downarrow}$ is determined by the condition that the energy of an up-spin quasiparticle at the up-spin Fermi surface be equal to that of a down-spin quasiparticle at the down-spin Fermi surface: if this were not the case one could gain energy by transferring quasiparticles from one Fermi surface to the other. The common value of the energy is, of course, the chemical potential (see Eq. (3)). The mathematical form of the equilibrium condition is thus

$$\tilde{\mathcal{E}}_{k_{F\uparrow}\uparrow} + \frac{1}{2}g\mu_B B = \tilde{\mathcal{E}}_{k_{F\downarrow}\downarrow} - \frac{1}{2}g\mu_B B \; . \tag{20}$$

Now, by making use of Eq. (7) and following the same procedure we used for the case of the compressibility, we obtain the elegant result

$$\frac{\chi_S}{\chi_{S0}} \; = \; \frac{m^\star/m}{1+F_0^a} \; , \tag{21}$$

where the interaction enters both through the effective mass and through the Landau parameter F_0^a. This has the same structure as Eq. (16). The $\ell = 0$ component is selected by the spherical symmetry of the quasiparticle distribution, and the a subscript reflects the spin-antisymmetry of that distribution.

Looking at Eqs. (16) and (21) we see that measurements of K and χ_S, combined with a knowledge of the effective mass from the heat capacity allow us to determine the values of the Landau parameters F_0^s and F_0^a. Negative values of these parameters, arising from the exchange interaction enhance both the proper compressibility and the spin susceptibility. Because these two quantities must be finite and positive in a stable ground state, we conclude that the uniform and paramagnetic state will be unstable if F_0^a or F_0^s become less than -1.[6]

2.5 Galilean invariance and effective mass

The effective mass of quasiparticles that we have discussed in the previous section arises entirely from the interaction between the Fermions in a translationally invariant Fermi liquid. Translational invariance is a good assumption for liquid ^{3}He or for nuclear matter, but hardly so for electrons in a solid state environment. The interaction of the electron with the periodic crystal potential and with lattice vibrations is an important source of effective mass renormalization. But if the system is translationally invariant, in the sense that momentum is strictly conserved, and if in addition the kinetic energy is of the Galilean-invariant form $p^2/2m$ with m the bare particle mass, then an exact relation exists between the quasiparticle effective mass and the interaction function:

$$\frac{m^*}{m} = 1 + F_1^s \; , \tag{22}$$

where F_1^s is the $\ell = 1$ (dipolar) component of the spin-symmetric Landau parameter.

[6]For the electron gas only the spin instability is real, as the density instability is preempted by the "improper" contribution arising from the charged background.

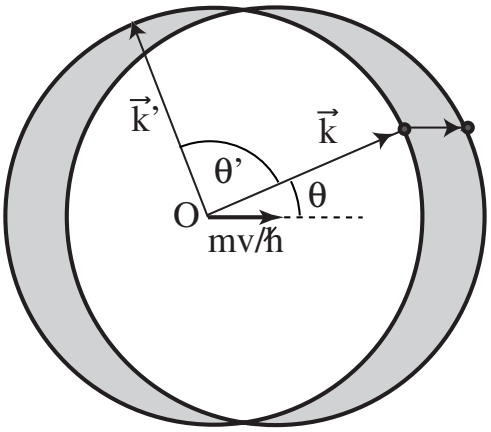


Fig. 4: *Relation between the effective mass and the Landau parameters. When the original Fermi surface centered at O is viewed from a reference frame moving at speed v it appears to be shifted by an amount $mv/\hbar$. The shift can be described in terms of quasiparticles and quasiholes added to the original Fermi distribution in the shaded crescent-shaped regions.*

The origin of this relation is illustrated in Fig. 4. We consider a quasiparticle of momentum $\vec{p} = \hbar\vec{k}$ in the reference frame in which the center of mass of the liquid is at rest. If we now change the reference frame to one in which the center of mass of the liquid moves to the right with velocity $\vec{v}$ the same quasiparticle will appear to have momentum $\hbar\vec{k} + m\vec{v}$ where m is the bare mass of the particle, not the quasiparticle mass! This is because under this transformation the total momentum of the fluid changes by $m\vec{v}$. But the momentum of the quasiparticle is not the only thing that changes. The entire momentum distribution shifts by $m\vec{v}$ and the net result of this shift can be described as the creation of quasiparticles and quasiholes in the shaded crescent-shaped regions of Fig. 4, with quasiparticles residing in the right crescent, and an equal number of quasiholes residing in the left crescent.
The corresponding change in energy of the quasiparticles (calculated to first order in $\vec{v}$) has two contributions:

(i) the change of the single quasiparticle energy $\mathcal{E}_{\vec{k}\sigma}$ due to the fact that the quasiparticle momentum is shifted from $\hbar\vec{k}$ to $\hbar\vec{k} + m\vec{v}$

(ii) the energy of interaction between the quasiparticle of momentum $\hbar\vec{k}$ and the additional quasiparticles and quasiholes that appear in the crescent-shaped regions of Fig. 4

The first contribution involves the effective mass of the quasiparticle, and the second contribution involves the dipolar component of the Landau interaction function (reflecting the dipolar structure of the additional quasiparticle distribution). Combining the two contribution must yield the exact change in the energy of the system under the Galilean transformation, namely $\Delta E = \hbar\vec{v}\cdot\vec{k}$, which does not depend the particle mass. Clearly, this is possible only if a precise relation exists between the quasiparticle mass and the interaction function, and this is what Eq. (22) gives us.
The importance of momentum conservation in the above discussion cannot be overemphasized. Consider, for example, the following question: what is the *spin-current* $j_\uparrow - j_\downarrow$, carried by a

r_s	F_0^s	F_0^a	F_1^s	F_1^a	F_2^s	F_2^a
1	-0.21	-0.17	-0.04	-0.0645	-0.0215	-0.0181
2	-0.37	-0.25	-0.03	-0.0825	-0.0168	-0.0126
3	-0.55	-0.32	-0.02	-0.0915	-0.0107	-0.0073
4	-0.74	-0.37	0.0	-0.0956	-0.0047	-0.0022
5	-0.95	-0.40	0.03	-0.0965	+0.0009	+0.0023

Table 1: *Calculated values of Landau Fermi liquid parameters of the three-dimensional electron liquid. The values of F_1^a, F_2^s, and F_2^a were calculated by Yasuhara and Ousaka [5].*

quasiparticle of wave vector $\vec{k}$ and spin $\uparrow$? Recall that the "spin current" is the difference between the current carried by spin-up particles and that carried by spin-down particles, where the spin of a particle (as well as the spin of a quasiparticle) is a good quantum number. One might be tempted to answer "$\hbar\vec{k}/m$" on the (wrong) assumption that a spin-up quasiparticle carries no down-spin current, but this is incorrect because the difference between the total up- and down-spin momenta of the particles $\hat{\vec{P}}_\uparrow - \hat{\vec{P}}_\downarrow$ is not a constant of the motion. In fact, the magnitude of the *spin-current* is smaller than $\frac{\hbar\vec{k}}{m}$ [4]. What happens is that in the process of switching-on the interaction some momentum is transferred from the up- to the down-spin component of the electron liquid. This reduces the spin current without altering the total momentum and spin. The reduction can be expressed in terms of an effective *spin mass* $m_S > m$ such that $j_\uparrow - j_\downarrow = \frac{\hbar k}{m_S}$. The relation between m_S and m^* has the same form as the relation (22) between the "charge mass", m, and m^*, i.e.,

$$\frac{m^*}{m_S} \;=\; 1 \;+\; F_1^a \;. \tag{23}$$

Numerical values of several Landau parameters of the uniform electron liquid (in the jellium model) obtained from approximate microscopic calculations [5] are listed in Table 2.5 for several different densities.

2.6 Measuring m^*, K and χ_S

The effective mass, the proper compressibility, and the spin susceptibility of liquid ^{3}He have been the object of many experimental studies by a variety of techniques (see, for example, Ref. [6]). Values of the effective mass, the spin susceptibility and the compressibility are reported in Ref. [7]. The effective mass ratio m^*/m ranges between 3 and 6 as the pressure is increased from 0 to 33 bar. In the same interval of pressures the spin susceptibility enhancement factor $(1+F_0^a)^{-1}$ is nearly constant at a value between 3.5 and 4, while the compressibility ratio K/K_0 decreases from 6 to 1 [7].

For electron liquids the situation is generally complicated by the presence of solid-state effects which are hard to identify and eliminate from the analysis. This is especially the case in what would appear to be the best realization of the three-dimensional jellium model, namely elemental metals. In spite of this difficulty valiant attempts have been made to determine a variety of Fermi liquid properties in these systems.

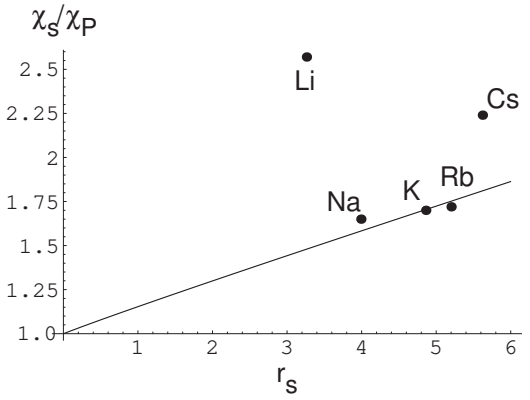


Fig. 5: *Experimentally determined spin susceptibilities of elemental metals vs theoretical values obtained from an analytical fit to quantum Monte Carlo results (see Ref. [2] for details).*

Fig. 5 shows a comparison between the spin susceptibilities of some elemental metals determined via the measurement of the Knight shift in the nuclear magnetic resonance, versus the spin susceptibility calculated directly from the second derivative of the energy with respect to spin polarization, which is known from quantum Monte Carlo calculations. The agreement is quite good for the three middle elements (with intermediate values of r_s) but it is much less satisfactory for the low-density metal Cs, and absolutely disappointing for Li.

The experimental situation improves considerably in two-dimensional electronic systems, which can be made essentially free of lattice and disorder effects. Here the electronic density can be changed by electrical means (gates), thus allowing a systematic study of interaction effects. Thus, for example, measurements of the capacitance of double-layer GaAs quantum wells (Eisenstein, 1994) [8] have allowed a precise determination of the proper compressibility of the two-dimensional electron liquid, confirming, in particular, that this quantity becomes negative below a certain density, without implying an instability of the system (stability being ensured by the presence of remote neutralizing charges on the gates or on the donors).

Coming to the effective mass and the spin susceptibility, the most accurate measurements so far have been done on systems such as n-type Silicon inversion layers, n-doped GaAs/AlGaAs, and p-doped GaAs (see Ref. [2] for details). While the early experiments were done at densities of the order of 10^{12} cm^{-2}, corresponding to $r_s \sim 3$ in Si, the most recent ones have reached considerably lower densities in both GaAs (1.7×10^9 cm^{-2}, i.e. $r_s = 13.4$) and in Si inversion layers (10^{11} cm^{-2}, i.e. $r_s = 8.4$). At such low densities, the many-body renormalizations (i.e., the Landau parameters) are strong: for example, the spin susceptibility enhancement $\frac{m^* g^*}{m_b g_b}$ (where m_b and g_b are the "bare" mass and g-factor determined by the band structure of the host semiconductors: $g_b \sim 2$ in Si and $g_b \sim 0.44$ in GaAs) can be as large as 5. [9] This is consistent with the theoretical expectation that the 2DEG should become ferromagnetic at sufficiently low density: however, no ferromagnetic instability has been observed so far.

Recently, tilted field experiments have allowed the determination of the spin susceptibility of the two-dimensional electron liquid in very narrow AlAs quantum wells in a broad range of densities [10]. Fig. (6) shows the comparison, obviously very satisfactory, between the measured data and the values obtained from quantum Monte Carlo calculations.

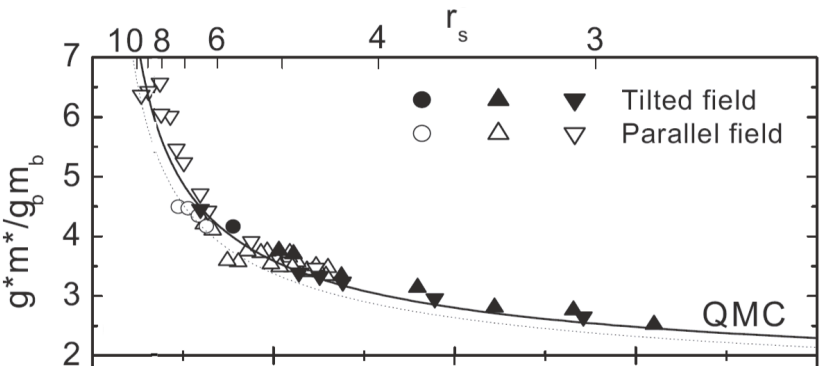


Fig. 6: *Spin susceptibility enhancement vs r_s for the 2DEL in narrow AlAs quantum wells. Circles and triangles denote experimental data taken in different samples and for different orientations of the magnetic field. The quantum Monte Carlo predictions at zero and full spin polarization are shown by the dotted and solid lines respectively. From Vakili et al. [10].*

3 The lifetime of quasiparticles

3.1 General formulas

Up to this point we have treated quasiparticles as if they were exact eigenstates of the many-fermion system. In reality, as we pointed out in the introduction, quasiparticles are not exact eigenstates, and do decay in time by transferring their momentum and energy to other quasiparticles. The key point of the Landau theory of Fermi liquids is that the decay rate tends to zero much more rapidly than the quasiparticle energy (relative to the chemical potential) when the latter tends to zero. It is this circumstance that allows us in a first approximation to neglect the decay rate and to treat the interaction between quasiparticles at the mean-field level, i.e., simply as a modification of the quasiparticle energy which is describable in terms of the Landau interaction function. However, the residual interactions, beyond the mean-field approximation, cause quasiparticles to decay. This decay process plays an essential role in the study of transport phenomena, where an external field drives the quasiparticles out of their thermal equilibrium distribution. In this situation the residual interactions between quasiparticles pull the quasiparticle distribution back towards equilibrium, thus providing a kind of restoring force which determines the magnitude of the response to the external field.

The most relevant process contributing to the decay of a quasiparticle state in a Fermi liquid is particle-hole pair production. At zero temperature, given a quasiparticle close to the Fermi surface, there is a certain probability that, because of the interaction, part of the energy and momentum of the quasiparticle will be lost by exciting a single electron-hole pair out of the Fermi sea. Upon losing part of its initial momentum and energy the quasiparticle makes a transition to an available lower energy state. At finite temperatures the scenario is slightly complicated by the fact that the available final states are neither definitely occupied nor definitely empty, but the basic physical picture remains the same.

In principle a quasiparticle can also lose momentum and energy by exciting *multiple* particle-hole pairs and/or collective modes (zero sound in a neutral Fermi liquid or plasmons in the electron liquid). As it turns out both processes are irrelevant for quasiparticles near the Fermi

surface at sufficiently low temperature. Decay by emission of collective modes is forbidden by energy and momentum conservation, since collective modes owe their very existence to having frequencies that are significantly higher than particle-hole excitation energies at the same momentum. As for multiple particle-hole pair excitations, we note that the spectral density of these excitations vanishes, at low energy, much more rapidly than the spectral density of single particle-hole pairs.

The simplest way to estimate the particle-hole contribution to the quasiparticle decay rate is to make use of the *Fermi golden rule* to compute the transition probability between the initial and the final state of the system. Within this approach the rate at which a quasiparticle of spin σ and momentum $\hbar\vec{k}$ is scattered by the Coulomb interaction into an empty state of momentum $\hbar(\vec{k}-\vec{q})$, while a second quasiparticle of spin σ' and momentum $\hbar k'$ is scattered into an empty state of momentum $\hbar(\vec{k}'+\vec{q})$ is given by

$$\frac{2\pi}{\hbar}\left|\frac{W(\vec{q})}{L^d}\right|^2 \delta\big(\varepsilon_{\vec{k}-\vec{q}\sigma}+\varepsilon_{\vec{k}'+\vec{q}\sigma'}-\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}'\sigma'}\big), \tag{24}$$

where $\frac{W(\vec{q})}{L^d}$ is the matrix element of an effective interaction between quasiparticles, taken between the initial and final plane-wave states. The δ-function ensures that the energy is conserved through the collision.

Eq. (24) is approximate in more than one way. First of all, the correct two-particle scattering amplitude is a function of $\vec{k}, \vec{k}'$, and $\vec{q}$, as well as the relative spin orientation, not just of the momentum transfer $\vec{q}$. Secondly, Eq. (24) violates the indistinguishability of the particles, since the scattering amplitude is not antisymmetric upon interchange of the two final plane-wave states of parallel spin. Finally, Eq. (24) is not entirely self-consistent, since the quasiparticle energy is approximated by the bare particle energy. In spite of these defects, Eq. (24) is still an excellent starting point to begin to understand the microscopic physics of the Fermi liquid. In what follows I focus on the case of the electron liquid, where the long range of the Coulomb interaction requires some special attention: however the final result has the same form for all Fermi liquids.

The inverse lifetime $1/\tau^{(e)}_{\vec{k}}$ of a plane wave state initially occupied by an electron of momentum $\hbar\vec{k}$ and spin σ, is given by the sum of the probabilities of all the allowed decay processes:

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} = \frac{2\pi}{\hbar}\sum_{\vec{q}\vec{k}'\sigma'}\left|\frac{W(\vec{q})}{L^d}\right|^2 n_{\vec{k}'\sigma'}\big(1-n_{\vec{k}'+\vec{q}\sigma'}\big)\big(1-n_{\vec{k}-\vec{q}\sigma}\big)\,\delta\big(\varepsilon_{\vec{k}-\vec{q}\sigma}+\varepsilon_{\vec{k}'+\vec{q}\sigma'}-\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}'\sigma'}\big), \tag{25}$$

where the Fermi occupation factors guarantee that the plane-wave state $\vec{k}'\sigma'$ is indeed occupied by an electron, while the final states $\vec{k}'+\vec{q}\,\sigma'$, and $\vec{k}-\vec{q}\,\sigma'$ are empty and therefore available for occupation after the scattering event.

A nice feature of Eq. (25) is that part of the calculation can be carried out without specifying the form of the scattering amplitude $W(\vec{q})$. We work, for simplicity, in the paramagnetic state, and approximate $n_{\vec{k}\sigma}$ by the *noninteracting* occupation numbers $n^{(0)}_{\vec{k}\sigma}$. Then we make use of the fluctuation-dissipation theorem for the non-interacting electron gas (see Ref. [2] for details) to

evaluate the sum over $\vec{k}'$ and σ':

$$\frac{\pi}{\hbar L^d}\sum_{\vec{k}'\sigma'} n^{(0)}_{\vec{k}'\sigma'}\left(1-n^{(0)}_{\vec{k}'+\vec{q}\,\sigma'}\right)\delta\left(\frac{\varepsilon_{\vec{k}'+\vec{q}\,\sigma'}-\varepsilon_{\vec{k}'\sigma'}}{\hbar}-\omega\right) \;=\; -\frac{\mathrm{Im}\,\chi_0(q,\omega)}{1-e^{-\beta\hbar\omega}}\,. \tag{26}$$

Here $\chi_0(q,\omega)$ is the density-density response function of the noninteracting Fermi liquid, also known as the Lindhard function [2]. Naturally, the appearance of the spectral density of particle-hole pairs, $-\,\mathrm{Im}\,\chi_0(q,\omega)$, shows that these excitations play a central role in the process. Eq. (25) can now be rewritten as

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} = -\frac{2}{(2\pi)^d}\int_{-\infty}^{\infty}d\omega\frac{1-n_F(\varepsilon_{\vec{k}\sigma}-\hbar\omega-\mu)}{1-e^{-\beta\hbar\omega}}\int_0^{\infty}dq q^{d-1}\left|W(\vec{q})\right|^2\mathrm{Im}\,\chi_0(q,\omega)\int d\Omega_d\delta\left(\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}-\vec{q}\sigma}-\hbar\omega\right) \tag{27}$$

where $n_F(x)=1/(e^{\beta x}+1)$ is the Fermi-Dirac distribution function at zero chemical potential, so that $n_F(\varepsilon_{\vec{k}\sigma}-\mu)=1/(e^{\beta(\varepsilon_{\vec{k}\sigma}-\mu)}+1)=n^{(0)}_{\vec{k}\sigma}$ is the noninteracting occupation number. In obtaining this expression we have introduced the variable $\hbar\omega=\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}-\vec{q}\sigma}$ through the introduction of an auxiliary delta function and its corresponding integration. Notice that the angular integration only involves the delta function.

The corresponding formula for the lifetime of a quasihole ($\tau^{(h)}_{\vec{k}\sigma}$) is obtained from Eq. (27) by performing the replacements $1-n_F(\varepsilon_{\vec{k}\sigma}-\hbar\omega-\mu)\to n_F(\varepsilon_{\vec{k}\sigma}-\hbar\omega-\mu)$ and $1-e^{-\beta\hbar\omega}\to 1-e^{\beta\hbar\omega}$, and changing the overall sign. This can be used to demonstrate that the principle of detailed balance (see Eq. (50) below), is satisfied at the appropriate level of accuracy, i.e., we have

$$\frac{n_{\vec{k}\sigma}}{\tau^{(e)}_{\vec{k}\sigma}} \;=\; \frac{1-n_{\vec{k}\sigma}}{\tau^{(h)}_{\vec{k}\sigma}}\,. \tag{28}$$

This important result guarantees (see Section 3.5) that the equilibrium Fermi-Dirac distribution of quasiparticles is stable against quasiparticle decay processes.

We shall henceforth concentrate only on the calculation of the quasiparticle lifetime. The calculation will be carried out separately for three and two-dimensional systems below. In one dimension, a calculation of $1/\tau^{(e)}_{\vec{k}\sigma}$ based on Eq. (27) would result in a divergent integral at finite temperatures: we can conclude that the Landau Fermi liquid picture cannot be consistently applied to 1D systems, as mentioned in the introduction.

3.2 Three-dimensional electron gas

In this case the angular integration is rather simple once the z-axis is taken along the direction of $\vec{q}$. We have:

$$\int_0^{2\pi}d\phi\int_0^{\pi}d\theta\sin\theta\,\delta\left(\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}-\vec{q}\sigma}-\hbar\omega\right)=\frac{2\pi m}{\hbar^2 kq}\,\Theta\left(1-\left|\frac{q^2+\frac{2m\omega}{\hbar}}{2kq}\right|\right), \tag{29}$$

where $\Theta(x)$ is the familiar Heaviside step function which mandates a precise behavior for the limits of the remaining quadratures. As it turns out however, these limiting conditions are

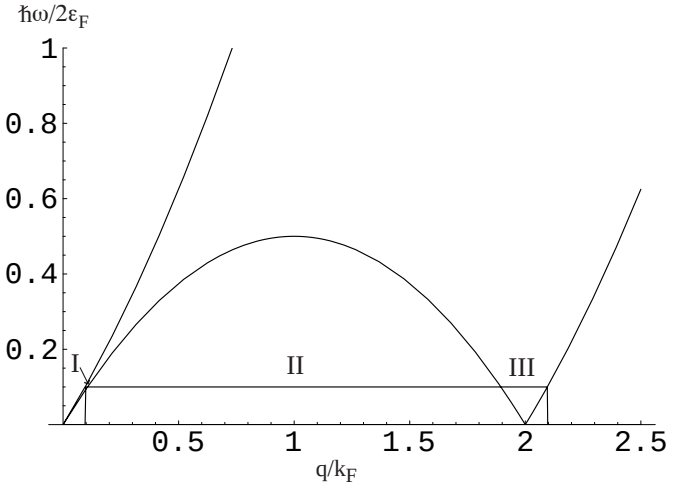


Fig. 7: *Regions of wave vector integration for the calculation of the low-energy quasiparticle lifetime. For $k \to k_F$ and $\hbar\omega \ll \epsilon_F$ the leading contribution to the wave vector integral comes from region II where $\frac{|\omega|}{v_F} < q < 2k_F - \frac{|\omega|}{v_F}$. Two-dimensional plasmon excitations do not contribute to the lifetime and hence are not shown.*

irrelevant, in view of the behavior of the integrand. This can be seen as follows: for positive frequencies the Fermi thermal occupation factor $1 - n_F(\varepsilon_{\vec{k}\sigma} - \hbar\omega - \mu)$ cuts off the integral for ω of the order of $|\varepsilon_{\vec{k}\sigma} - \mu|$, an energy which, by definition, is much smaller than the Fermi energy ϵ_F. For negative frequencies, on the other hand, it is the thermal occupation factor $1/(1 - e^{-\beta\hbar\omega})$ that cuts off the frequency integral for ω of the order of $k_B T$, an energy scale that we assume to be much smaller than ϵ_F. At very low frequency the most stringent limits on the q-integral come from the factor $\mathrm{Im}\,\chi_0(q,\omega)$ which contains the structure of the particle-hole continuum and differs from zero only along the segment shown in Fig. 7. This sets the lower limit of the q integral at $q \sim \frac{|\omega|}{v_F} \sim 0$ and the upper limit at $q \sim 2k_F + \frac{|\omega|}{v_F} \sim 2k_F$. Notice that the dominant contribution to the integral (for $\omega \to 0$) comes from the region labeled as II in Figure 7. In this region it is legitimate to approximate $\mathrm{Im}\,\chi_0(q,\omega)$ by its zero temperature and low-frequency form

$$\mathrm{Im}\,\chi_0(q,\omega) \simeq -\frac{\pi\omega}{2v_F q} N(0)\,. \tag{30}$$

Accordingly the formula for $1/\tau_k^{(e)}$ in three dimensions is seen to be proportional to

$$\frac{1}{\tau_{\vec{k}\sigma}^{(e)}} \;\propto\; \int_{-\infty}^{\infty} \frac{d\omega\;\omega}{\left(1 + e^{\beta(\hbar\omega - \varepsilon_{\vec{k}\sigma} + \mu)}\right)\left(1 - e^{-\beta\hbar\omega}\right)} \int_0^{2k_F} dq\,\left|W(\vec{q})\right|^2 . \tag{31}$$

The frequency integral can be calculated analytically and is given by

$$\frac{1}{2\hbar^2} \frac{(\varepsilon_{\vec{k}\sigma} - \mu)^2 + (\pi k_B T)^2}{1 + e^{-\beta(\varepsilon_{\vec{k}\sigma} - \mu)}}\,. \tag{32}$$

The integral over the wave vector q deserves special attention. It is quite obvious at this point that one cannot make use of the bare Coulomb interaction. The integral would emphatically diverge. This is of course a consequence of the long range of the Coulomb interaction. It is then

quite natural to employ for $W(\vec{q})$ some sort of screened interaction. In this case the q integral presents no problems. A reasonable approximation is provided by the choice $W(\vec{q}) \simeq \frac{v_q}{\varepsilon(q,0)}$ so that

$$\int_0^{2k_F} dq \left|W(\vec{q})\right|^2 \simeq \frac{1}{N^2(0)} \int_0^{2k_F} dq \left(\frac{N(0)v_q}{\varepsilon(q,0)}\right)^2 . \tag{33}$$

Further simplification can be achieved by making use of the Thomas-Fermi approximate dielectric function, that is the static long wavelength limit of the RPA dielectric function, given by $\epsilon(q,0) = 1 + \frac{4\pi e^2 N(0)}{q^2}$. This gives the result

$$\int_0^{2k_F} dq \left|W(\vec{q})\right|^2 \simeq \frac{2k_F}{N^2(0)} \xi_3(r_s) , \tag{34}$$

where the function $\xi_3(r_s)$ is given by

$$\xi_3(r_s) = \sqrt{\frac{\alpha_3 r_s}{4\pi}} \tan^{-1} \sqrt{\frac{\pi}{\alpha_3 r_s}} + \frac{1}{2\left(1+\frac{\pi}{\alpha_3 r_s}\right)} . \tag{35}$$

Recall that $r_s (= \left(\frac{3}{4\pi n a^3}\right)^{1/3}$ in 3D) is the average distance between electrons in units of the Bohr radius $a = \frac{\hbar^2}{me^2}$. For most densities in the metallic range $\xi_3(r_s) \simeq 1$. Notice that $\xi_3(r_s) \sim \sqrt{r_s}$ as $r_s \to 0$: thus, due to the non-perturbative nature of the screening, the quasiparticle decay rate turns out being proportional to the electron charge e, rather than e^4, as one could have naively expected.

Collecting the various factors we finally obtain for the inelastic quasiparticle lifetime in three dimensions the following result

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} \simeq \frac{\pi}{8\hbar\epsilon_F} \frac{(\varepsilon_{\vec{k}\sigma}-\epsilon_F)^2 + (\pi k_B T)^2}{1 + e^{-\beta(\varepsilon_{\vec{k}\sigma}-\epsilon_F)}} \xi_3(r_s) \qquad \text{in } 3D, \tag{36}$$

where we have approximated μ with ϵ_F and k with k_F.

The inverse lifetime of a quasiparticle at the Fermi surface ($k = k_F$) vanishes as T^2 at small temperatures. On the other hand, at $T = 0$, the inverse lifetime vanishes as $(\varepsilon_{\vec{k}\sigma}-\epsilon_F)^2$. One power of $\varepsilon_{\vec{k}\sigma}-\epsilon_F$ (or T) arises from the phase space restrictions on the scattering process. The second one stems from the linearly vanishing density of particle-hole pair excitations. The numerical prefactor is simply a Fermi surface average of the statically screened Coulomb interaction. This is the expected behavior, an indication that the Landau theory of the electron liquid is consistent with the microscopic perturbative approach.

3.3 Two-dimensional electron gas

The two-dimensional case presents a few new twists. The most important difference to the three-dimensional case is the q dependence of the integrand of Eq. (27), which must be handled with special care in the regions $q \simeq 0$ and $q \simeq 2k_F$. This necessitates a more precise treatment of the limits of integration.

We begin by considering the angular integration which in this case gives the interesting result

$$\int_0^{2\pi} d\phi\, \delta\big(\varepsilon_{\vec{k}\sigma}-\varepsilon_{\vec{k}-\vec{q}\sigma}-\hbar\omega\big) \;=\; \frac{2\Theta\left(1-\left|\frac{q^2+\frac{2m\omega}{\hbar}}{2kq}\right|\right)}{\sqrt{\left(\frac{\hbar^2 kq}{m}\right)^2-\left(\hbar\omega+\frac{\hbar^2q^2}{2m}\right)^2}}\,, \tag{37}$$

an expression that features an extra frequency dependence with respect to the three-dimensional case. The other necessary ingredient is the expression for $\mathrm{Im}\,\chi_0(q,\omega)$ in two dimensions, which, at low frequency and in region II of Fig. 7, is approximately given by

$$\mathrm{Im}\,\chi_0(q,\omega) \simeq -\frac{\omega}{qv_F}\frac{N(0)}{\sqrt{1-\left(\frac{q}{2k_F}\right)^2}} \qquad (2D), \tag{38}$$

where $N(0)=\frac{m}{\pi\hbar^2}$. Within the necessary accuracy, we can set $k=k_F$ in the argument of the square root appearing in Eq. (37), which can then be rewritten as

$$\frac{\hbar^2}{2m}\sqrt{\left(q^2-\left(\frac{|\omega|}{v_F}\right)^2\right)\left(4k_F^2-q^2\right)}\,. \tag{39}$$

Accordingly we see that the contribution of region II to the q integral of Eq. (27) is

$$8\pi k_F\int_{\frac{|\omega|}{v_F}}^{2k_F-\frac{|\omega|}{v_F}}\frac{\big|N(0)W(\vec{q})\big|^2 dq}{\sqrt{q^2-\left(\frac{|\omega|}{v_F}\right)^2}\,(4k_F^2-q^2)} \qquad (2D). \tag{40}$$

This integral can be evaluated rather easily. Notice that in the limit $\omega\to 0$ it presents a logarithmic divergence originating from the regions $q\simeq 0$ and $q\simeq 2k_F$. To extract the exact coefficient of the logarithmic singularity we set $q=0$ and $q=2k_F$ in the regular parts of the integrand, when evaluating the contributions of $q\simeq 0$ and $q\simeq 2k_F$ respectively. Up to corrections that remain finite as $\omega\to 0$ the integral is then found to be equal to

$$\frac{\pi\big(|N(0)W(0)|^2+|N(0)W(2k_F)|^2/2\big)}{k_F}\,\ln\frac{4\epsilon_F}{|\omega|}\,. \tag{41}$$

It is convenient, at this point, to define the "coupling constant"

$$\xi_2(r_s)\equiv|N(0)W(0)|^2+\frac{1}{2}|N(0)W(2k_F)|^2\,, \tag{42}$$

which, in the Thomas-Fermi approximation (see [2]), depends on r_s in the following manner:

$$\xi_2(r_s)=1+\frac{1}{2}\left(\frac{r_s}{r_s+\sqrt{2}}\right)^2\,. \tag{43}$$

Notice that, unlike its three-dimensional counterpart, $\xi_2(r_s)$ tends to a constant, 1, in the high-density limit.[7] Combining Eqs. (27), (37), (38), and (40), we find that the quasiparticle lifetime

[7] The surprising fact that the inverse lifetime fails to vanish in the noninteracting limit $r_s\to 0$ is an artifact due to our asymptotic expansion of the integral (40), which requires the limit $k\to k_F$ to be taken *before* the limit $r_s\to 0$. The expansion fails for $r_s<|k/k_F-1|$.

is given by the integral

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} \simeq \frac{\hbar\xi_2(r_s)}{2\pi\epsilon_F}\int_{-\infty}^{\infty}\frac{\omega\ln\frac{4\epsilon_F}{|\omega|}}{\left(1+e^{\beta(\hbar\omega-\varepsilon_{\vec{k}\sigma}+\mu)}\right)\left(1-e^{-\beta\hbar\omega}\right)}\,d\omega \qquad \text{in } 2D. \tag{44}$$

Consider first the "zero-temperature" situation $|\varepsilon_{\vec{k}\sigma}-\epsilon_F| \gg k_BT$. In this case it is clear that the main contribution to the integral comes from the region $\omega \simeq \varepsilon_{\vec{k}\sigma}-\epsilon_F$. Now, since in this region the logarithm is slowly varying, we can take it out of the integration to give the factor $\ln\frac{4\epsilon_F}{|\varepsilon_{\vec{k}\sigma}-\epsilon_F|}$. This leaves us with a frequency integral which coincides with that of Eq. (31), which we calculated exactly. The only difference is that in this case we need to take the limit $\frac{k_BT}{|\varepsilon_{\vec{k}\sigma}-\epsilon_F|} \to 0$. By making use of Eq. (32) we therefore obtain the result

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} \simeq \xi_2(r_s)\frac{(\varepsilon_{\vec{k}\sigma}-\epsilon_F)^2}{4\pi\hbar\epsilon_F}\ln\frac{4\epsilon_F}{|\varepsilon_{\vec{k}\sigma}-\epsilon_F|}\,, \quad k_BT \ll |\varepsilon_{\vec{k}\sigma}-\epsilon_F|\,. \tag{45}$$

The other relevant case is that of $k_BT \gg |\varepsilon_{\vec{k}\sigma}-\epsilon_F|$, which corresponds to the case of a quasiparticle lying on the Fermi surface. In this case a direct inspection of Eq. (44) shows that the most relevant contributions to the integral come from a region of the order of $k_BT/\hbar$ centered about the origin. In this situation the logarithm can again be taken out of the integral[8] as to give a factor of $\ln\frac{4\epsilon_F}{k_BT}$. The remaining integral can then again be evaluated by means of Eq. (32) in the limit of $\frac{|\varepsilon_{\vec{k}\sigma}-\epsilon_F|}{k_BT} \to 0$. This immediately gives[9]

$$\frac{1}{\tau^{(e)}_{\vec{k}\sigma}} \simeq \xi_2(r_s)\frac{(\pi k_BT)^2}{8\pi\hbar\epsilon_F}\ln\frac{4\epsilon_F}{k_BT}\,, \quad |\varepsilon_{\vec{k}\sigma}-\epsilon_F| \ll k_BT\,. \tag{46}$$

The only significant difference with the three-dimensional case is the appearance here of the logarithmic factors. This fact was first discovered by Giuliani and Quinn in Ref. [11].
The basic comments made about the three-dimensional result continue to apply. We emphasize that the above calculation focused on the leading term in the low-energy/low-temperature expansion of the inverse lifetime.[10] The complete calculation of the "subleading" contributions of order $(\varepsilon_{\vec{k}\sigma}-\epsilon_F)^2$ and $(k_BT)^2$ is more tricky: in particular, it can be shown that the regions I and III in the q integral (see Fig. 7) do contribute to the result at this order.

3.4 Measuring the quasiparticle lifetime

Thanks to the great improvements in the manufacture of high-quality quantum well systems it has become possible to directly measure, by means of precise and elegant tunneling experiments between parallel identical quantum wells, the quasiparticle lifetime of a two-dimensional electron liquid (Murphy *et al.*, Ref. [12]).

[8] This is due to the fact that if $f(x)$ is a well behaved function in the interval $[-a,a]$, then in the limit $a \to 0$, with logarithmic accuracy, $\int_{-a}^{a}\ln|x|f(x)\,dx \simeq \ln|a|\int_{-a}^{a}f(x)\,dx$ as one can readily verify.

[9] The coefficient of Eq. (46) can also be quickly inferred from Eq. (45) by making use of the general result (32).

[10] It is somewhat sobering to remark that a surprisingly vast variety of contradicting results for the coefficients of Eqs. (45) and (46) have appeared in the literature.

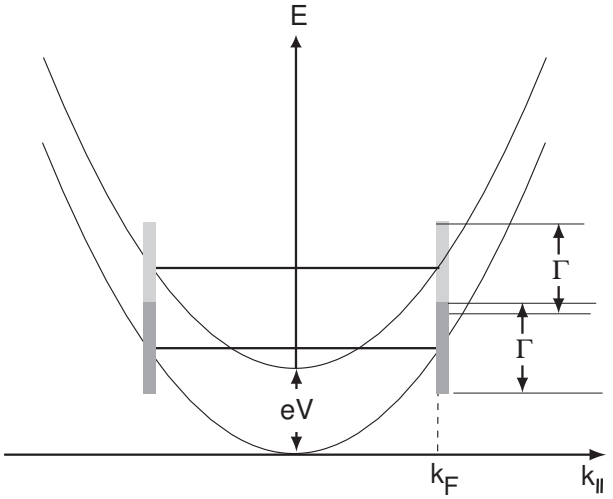


Fig. 8: *Momentum-conserving tunneling between two identical free-electron bands separated by a potential difference eV. Because of energy conservation, the tunneling probability decreases rapidly when eV exceeds the spectral width Γ of the single-particle states in each band.*

The basic idea of the experiment is shown in Fig. 8. The two identical parabolas separated by an energy eV represent the energy vs wave vector relation of the quasiparticle states in the two quantum wells, and the shaded regions show the energy spread of these states due to finite lifetime of a quasiparticle at the Fermi surface (we are, of course, at finite temperature). Here $k_\parallel$ is the two-dimensional in-plane-wave vector, V is the electric potential difference between the two quantum wells, and Γ is the width of the quasiparticle peak in the spectral function at $k_\parallel = k_F$, which is directly related to the quasiparticle lifetime. Under the assumption that electron-impurity and electron-phonon scattering are negligible the two-dimensional wave vector of the tunneling electrons is conserved and overall energy conservation causes the tunneling probability to decrease sharply when eV exceeds Γ. More precisely, a plot of the tunneling conductance vs voltage is approximately a Lorentzian centered at zero voltage with width at half maximum equal to 2Γ, as shown in the inset of Fig. 9. From this width the quasiparticle lifetime can be inferred.
In practice, the interpretation of the experimental data is complicated by the presence of disorder, which leads to imperfect momentum conservation and a finite linewidth even in the limit of zero temperature. This extrinsic contribution, however, is expected to be nearly independent of temperature, and when one subtracts it from the data one obtains values that are in reasonably good agreement with the theory presented in this section (see Fig. 9).

3.5 The kinetic equation

Perhaps the technical centerpiece of the Landau theory of Fermi liquids is the *kinetic equation*, which governs the time evolution of the quasiparticle distribution function in out-of-equilibrium situations, such as in the presence of slowly varying external fields. The main idea is to treat the system as an assembly of quasiparticle wave packets characterized by an average position $\vec{r}$ and an average momentum $\hbar\vec{k}$ ($k \simeq k_F$). The quantum mechanical uncertainties in position and energy are assumed to be negligible on the scale of spatial and temporal variation of the external

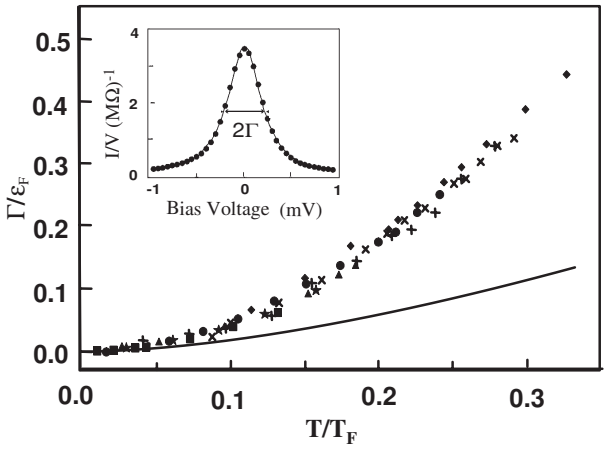


Fig. 9: *Inset: Lorentzian lineshape of the current-voltage (I-V) relation in tunneling between two-dimensional GaAs quantum wells. Main figure: a plot of the half-width Γ, identified with the inverse of the quasiparticle lifetime, vs temperature for systems of different density (a residual T=0 contribution, attributed to disorder, has been subtracted). The solid line is the theoretical prediction from Eq. (46) which is only applicable asymptotically. Adapted from Ref. [12].*

fields. This description makes sense only if the wavevector and frequency of the external field are much smaller than the Fermi wave vector and the Fermi energy respectively. In addition, the thermal energy k_BT must be much smaller than the Fermi energy in order that the notions of Fermi surface and quasiparticles be well defined. Under these assumptions the quasiparticle wave packets can be treated as classical particles, with canonical coordinates and momenta $\vec{r}$ and $\hbar\vec{k}$, described by a "classical" Hamiltonian

$$H_{cl}(\vec{r},\hbar\vec{k},\sigma) = \mathcal{E}_{\vec{k}\sigma} - e\phi_\sigma(\vec{r},t) + \sum_{\vec{k}'\sigma'} f_{\vec{k}\sigma,\vec{k}'\sigma'}\,\delta\mathcal{N}_{\vec{k}'\sigma'}(\vec{r},t) \tag{47}$$

where $\phi_\sigma(\vec{r},t)$ is a (generally spin-dependent) scalar potential. The last term on the right hand side of Eq. (47) describes the effect of the *short-range* interaction between the quasiparticles. It has the form of a mean effective potential whose strength is controlled by the Landau interaction function. The long-range electrostatic potential (Hartree potential) is self-consistently included in the external field.

The equation of motion for the quasiparticle distribution function follows immediately from Liouville's theorem for a classical flow in phase space

$$\frac{\partial\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial t} + \frac{1}{\hbar}\frac{\partial H_{cl}}{\partial\vec{k}}\cdot\frac{\partial\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial\vec{r}} - \frac{1}{\hbar}\frac{\partial H_{cl}}{\partial\vec{r}}\cdot\frac{\partial\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial\vec{k}} = \left(\frac{\partial\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial t}\right)_{\rm coll}. \tag{48}$$

The collisional time derivative on the right hand side of Eq. (48) takes into account the fact that the evolution of the quasiparticle distribution function is affected by collision processes that are not included in the classical mean field Hamiltonian. As discussed in previous sections, quasiparticle collisions result in a finite lifetime of quasiparticles ($\tau^{(e)}_{\vec{k}\sigma}$) and quasiholes ($\tau^{(h)}_{\vec{k}\sigma}$) near the Fermi surface. We can therefore write

$$\left(\frac{\partial\mathcal{N}_{\vec{k}\sigma}}{\partial t}\right)_{\rm coll} = -\frac{\mathcal{N}_{\vec{k}\sigma}}{\tau^{(e)}_{\vec{k}\sigma}} + \frac{1-\mathcal{N}_{\vec{k}\sigma}}{\tau^{(h)}_{\vec{k}\sigma}}\,, \tag{49}$$

where the first term represents the rate at which quasiparticles leave the state $\vec{k}\sigma$ and the second is the rate at which they are scattered into it. Besides the interaction contributions derived in Section 3, the total decay rates will in general include contributions from electron-impurity and electron-phonon scattering.

It is evident that the collisional derivative must vanish when $\mathcal{N}_{\vec{k}\sigma}$ is the thermal equilibrium distribution $\mathcal{N}^{\rm eq}_{\vec{k}\sigma}$, i.e., the Fermi-Dirac distribution with energy $\mathcal{E}_{\vec{k}\sigma}$ (see Eq. (11)). This *principle of detailed balance* leads to an exact relation between quasiparticle and quasihole lifetimes:

$$\frac{\mathcal{N}^{\rm eq}_{\vec{k}\sigma}}{\tau^{(e)}_{\vec{k}\sigma}} \;=\; \frac{1-\mathcal{N}^{\rm eq}_{\vec{k}\sigma}}{\tau^{(h)}_{\vec{k}\sigma}}\,, \tag{50}$$

which is satisfied (at the appropriate level of accuracy) by the formulas presented in Section 3 (see Eq. (28)). Expanding Eq. (48) to first order in the strength of the external fields we obtain the *linearized kinetic equation* for the deviation of the distribution function from equilibrium. This equation has the form

$$\frac{\partial\delta\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial t} + \vec{v}_{\vec{k}\sigma}\cdot\frac{\partial\delta\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial\vec{r}} + \vec{v}_{\vec{k}\sigma}\cdot\vec{\mathcal{F}}_{\vec{k}\sigma}(\vec{r},t)\,\delta(\mathcal{E}_{\vec{k}\sigma}-\mu) = \left(\frac{\partial\delta\mathcal{N}_{\vec{k}\sigma}(\vec{r},t)}{\partial t}\right)_{\rm coll} \tag{51}$$

where $v_{\vec{k}\sigma} = \hbar\vec{k}/m^*$ is the quasiparticle velocity, and

$$\vec{\mathcal{F}}_{\vec{k}\sigma}(\vec{r},t) = -\vec{\nabla}_{\vec{r}}\left(-e\phi_\sigma(\vec{r},t) + \sum_{\vec{k}'\sigma'} f_{\vec{k}\sigma,\vec{k}'\sigma'}\,\delta\mathcal{N}_{\vec{k}'\sigma'}(\vec{r},t)\right) \tag{52}$$

is the classical force acting on the quasiparticle. This equation is the starting point for most applications of the Landau theory of Fermi liquids.

One outstanding application of the kinetic equation is the study of the macroscopic dynamics of the quasiparticle distribution function in the absence of external fields, leading to the prediction of self-sustained collective modes (i.e., normal modes of oscillation of the Fermi surface) of different symmetries. In this manner one can obtain the (long-wavelength) dispersion of the zero-sound mode in the neutral Fermi liquid and plasmons in the electron liquid. Focusing, for example, on plasmons, we neglect the collision term (justified, since the collision rate is much smaller than the plasmon frequency) and notice that $\phi_\sigma(\vec{r},t)$ is the self-consistent electrostatic potential (Hartree potential) associated with a density fluctuation $\delta n(\vec{r},t) = \sum_{\vec{k}'\sigma'}\delta\mathcal{N}_{\vec{k}'\sigma'}(\vec{r},t)$. We take advantage of linearity by performing a Fourier transformation with respect to the variables $\vec{r}$ and t in Eqs. (51) and (52). This gives us

$$\big(\omega-\vec{q}\cdot\vec{v}_{\vec{k}\sigma}\big)\delta\mathcal{N}_{\vec{k}\sigma}(\vec{q},\omega) + \vec{q}\cdot\vec{v}_{\vec{k}\sigma}\,\delta\big(\mathcal{E}_{\vec{k}\sigma}-\mu\big)\sum_{\vec{k}'\sigma'}\big(v_q+f_{\vec{k}\sigma,\vec{k}'\sigma'}\big)\delta\mathcal{N}_{\vec{k}'\sigma'}(\vec{q},\omega) = 0\,, \tag{53}$$

where $v_q = 4\pi e^2/q^2$ in 3D and $v_q = 2\pi e^2/q$ in 2D. A nontrivial solution $\delta\mathcal{N}_{\vec{k}'\sigma'}(\vec{q},\omega) \neq 0$ of this equation exists only if ω equals the plasmon frequency $\omega_p(q)$. In the long wavelength limit, $q\to 0$, we get $\omega_p(q) = (4\pi ne^2/m)^{1/2}$ in 3D and $\omega_p(q) = (2\pi ne^2 q/m)^{1/2}$ in 2D. Notice that these results involve the bare electron mass m, not the quasiparticle mass, which appears

in $\vec{v}_{\vec{k}\sigma}$. How did the quasiparticle mass get converted back to the bare mass? The answer is that the Landau interaction function, acting on the self consistent solution of Eq. (53) reinstates the bare mass according to the Galilean invariance relation discussed in section 2.5. Inclusion of the quasiparticle collision term does not change these results (if translational and Galilean invariance are in force), but affects the dispersion and introduces damping of the collective modes at finite wave vector.

The other classic application of the kinetic equation for quasiparticles is the calculation of the transport coefficients of a Fermi liquid. These are the spin diffusion constant, D_s, the shear and bulk viscosities, η and ζ respectively, and the thermal conductivity κ. D_s is the constant of proportionality between the spin current and the gradient of spin density that drives it. Similarly, η and ζ can be viewed as the traceless and traceful components of a diffusion tensor for the momentum density – with the momentum current being driven by a gradient in the macroscopic velocity field. Lastly, κ is the constant of proportionality between the heat (entropy) current and the gradient of temperature that drives it. Quasiparticle collisions, which are responsible for the finite quasiparticle lifetime, are absolutely essential to calculate these transport coefficient. In fact, these coefficients would all be infinite if those collisions were neglected, which of course becomes a better and better approximation as the temperature is reduced. This counterintuitive result (divergence of the transport coefficients for $T \to 0$) follows from the "asymptotic freedom" of the Landau quasiparticles in this limit. Indeed, one can show that the transport coefficient are qualitatively described by the following formulas

$$D_s \sim v_F^2 \tau_s \,, \quad \eta \sim S\tau_\eta \,, \quad \zeta \sim B\tau_\zeta \,, \quad \kappa = n c_v v_F^2 \tau_q \,, \tag{54}$$

where S and B, are, respectively, the high-frequency shear modulus and the bulk modulus (both on the order of $n\epsilon_F$) and c_v is the heat capacity (per particle) of the Fermi liquid. Here τ_s, τ_η etc... are transport relaxation times which are related to the quasiparticle lifetime (since they all depend on the same scattering probabilities) but are not identical with it or with each other. All these scattering times diverge in the limit of zero temperature as $1/T^2$ in three dimensions. In two dimensions the situation is more delicate as the scattering times associated with spin diffusion and thermal conductivity diverge as $1/(T^2 \ln T)$, while the scattering time associated with the viscosity continues to diverge as $1/T^2$ [14]. This is due to the fact that the scattering processes that are responsible for the logarithmic divergence have zero momentum transfer and therefore do not contribute to the transfer of momentum within the liquid. The divergence is eventually cut off when the quasiparticle mean free path becomes comparable to the macroscopic size of the system, at which point the coefficients lose their hydrodynamic significance. You might wonder why the density diffusion constant D_n does not appear in our list of transport coefficients. The answer is that in the absence of impurities or external potentials quasiparticle collisions cannot change the total particle current: this leads to an infinite conductivity and then, via the Einstein relation – which connects the conductivity to the diffusion constant – to an infinite diffusion constant! The physical interpretation of this surprising result is that a density imbalance in a Fermi liquid does not relax via a diffusion process, but through the emission of sound waves.

The transport coefficients (or, equivalently, the transport relaxation times defined by Eq. (54)) can be calculated with the help of the kinetic equation. For the neutral Fermi liquid ^{3}He this was first done by Abrikosov and Khalatnikov [13]. Explicit formulas for the transport coefficients in terms of angular averages of quasiparticle scattering probability are known for the three-dimensional case: see Eqs. (1.171) and (1.151) of Ref. [3]. The scattering probability is obtained from the Fermi golden rule, just as in the calculation of the quasiparticle lifetime. We refer the interested reader to Ref. [3] for details.

4 Microscopic basis of the Landau theory of Fermi liquids

Landau guessed the theory of interacting Fermi liquids largely on a basis of physical intuition. Shortly afterwards, it was shown that indeed the theory could be "derived" (or, more accurately, shown to be self-consistent) from the microscopic Hamiltonian under certain assumptions of continuity and regularity. Nowadays the Landau theory is recognized as an early example of *renormalization*, whereby the exact many-body Hamiltonian is transformed, through recursive elimination of fast degrees of freedom, into an effective Hamiltonian of weakly interacting quasiparticles. A "poor man" version of this theory, based on seminal work by Hamann and Overhauser [15], can be found in Section 8.6 of Ref. [2].
Without going into technical details, which can be easily found in the literature (including Ref. [2]), we summarize the main correspondences between the Landau theory and the microscopic theory.

4.1 Existence of quasiparticles and self-energy

The existence of long-lived quasiparticles, with an inverse lifetime that scales as $|k-k_F|^2$ for $k \to k_F$ corresponds to the fact that the microscopic retarded Green function has the form

$$G(\vec{k},\omega) \;=\; G^{(\mathrm{reg})}(\vec{k},\omega) + \frac{Z_{\vec{k}}}{\omega - \frac{\mathcal{E}_{\vec{k}}}{\hbar} + \frac{i}{2\tau_{\vec{k}}}}\,, \tag{55}$$

where $G^{(\mathrm{reg})}(\vec{k},\omega)$ is a regular function of $\vec{k}$ and ω, and we have omitted the spin dependence for simplicity. Thus, the retarded Green function is dominated by a single pole of strength $Z_{\vec{k}}$ ($0 < Z_{\vec{k}} < 1$) at the complex frequency $z = \frac{\mathcal{E}_{\vec{k}}}{\hbar} - \frac{i}{2\tau_{\vec{k}}}$ in the lower half of the complex plane (as required by causality). The imaginary part of the frequency at the pole implies an exponential decay, with a characteristic time $\tau_{\vec{k}}$, of the squared amplitude of a plane-wave state. The quasiparticle energy, the strength of the pole (also known as the *renormalization constant*), and the plane-wave lifetime are determined by the (retarded) self-energy function[11] as follows

$$\mathcal{E}_{\vec{k}} = \varepsilon_{\vec{k}} + \mathrm{Re}\,\Sigma^{\mathrm{ret}}_{\sigma}(\vec{k},\mathcal{E}_{\vec{k}})\,, \tag{57}$$

[11] We remind the reader that the self-energy is defined as the difference between the inverse noninteracting Green function and the inverse Green function:

$$\Sigma = G_0^{-1} - G^{-1} \tag{56}$$

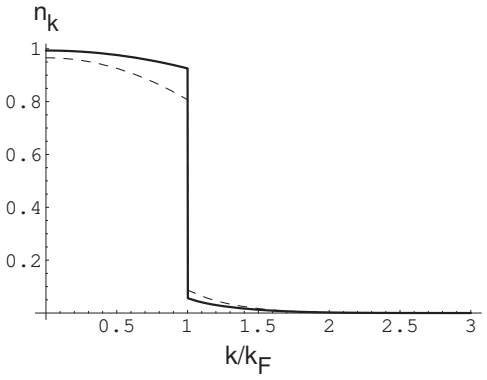


Fig. 10: *Behavior of the plane-wave states average occupation number in three dimensions at $r_s = 2$ (solid line) and $r_s = 5$ (dashed line). The jump of n_k at $k = k_F$ is the renormalization constant Z_{k_F}.*

where $\varepsilon_{\vec{k}}$ is the bare particle energy,

$$Z_{\vec{k}} \equiv \left(1 - \frac{1}{\hbar}\frac{\partial}{\partial\omega}\,\mathrm{Re}\,\Sigma^{\mathrm{ret}}(k,\omega)\Big|_{\hbar\omega=\mathcal{E}_{\vec{k}}}\right)^{-1}, \tag{58}$$

and

$$\frac{\hbar}{2\tau_{\vec{k}}} \;\equiv\; Z_{\vec{k}}\,\big|\,\mathrm{Im}\,\Sigma^{\mathrm{ret}}(k,\mathcal{E}_{\vec{k}\sigma})\big|\,. \tag{59}$$

Crucially, the validity of the Fermi liquid scenario requires that

$$\mathrm{Im}\,\Sigma^{\mathrm{ret}}(\vec{k},\mathcal{E}_{\vec{k}\sigma}) \sim -a\big(k-k_F\big)^2, \tag{60}$$

where a is a positive constant, and the approximate equality $\sim$ disregards the possibility of logarithmic terms $\ln|k-k_F|$. Microscopic calculations of the self-energy confirm that the asymptotic form (60) is, at the very least, self-consistent, i.e., the presence of a pole of the form (55) in the Green function guarantees the vanishing of the imaginary part of the self-energy, which in turn implies a divergence of the quasiparticle life time.[12]

The existence of an infinitely sharp quasiparticle peak of strength $Z_{\vec{k}\sigma}$ in the spectral function (defined as the negative of the imaginary part of the retarded Green function) implies that the momentum state occupation number $n_{\vec{k}}$ has a discontinuity as a function of k at $k = k_F$, the magnitude of the discontinuity being given by Z_{k_F}. This is shown in Fig. 10. The fact that the discontinuity occurs precisely at $k = k_F$ where k_F is related to density by the ideal gas relation, is known in the literature as *Luttinger's theorem*. Notice that $Z_{\vec{k}}$ and the ground state occupations $n_{\vec{k}\sigma}$ are "invisible" in the Landau theory of Fermi liquids, which abstracts from the detailed structure of the ground state. On the other hand, all the parameters of the Landau theory can be calculated from the microscopic theory following well-defined procedures. For

[12] It must be noted that the plane-wave lifetime of Eqs. (55) and (59) is not exactly the same as the lifetime of a quasiparticle and a quasihole. The inverse of the former differs from the inverse of latter by factors $1 - n_{\vec{k}}$ and $n_{\vec{k}}$ respectively. See Ref. [2] for details.

example, the effective mass of the quasiparticle works out to be

$$\frac{m^*}{m} = \frac{1}{Z_{k_F}\left(1 + \frac{m}{\hbar^2 k_F}\frac{\partial}{\partial k}\,\mathrm{Re}\,\Sigma^{\mathrm{ret}}(k,\mu)\Big|_{k=k_F}\right)}\,. \tag{61}$$

The results of several microscopic calculations of the effective mass of quasiparticles in the 2D and 3D electron gas are presented and critically discussed in Chapter 8 of Ref. [2].

4.2 Landau interaction function and scattering amplitude

What about the Landau interaction function? What is its representation in the microscopic theory? Considering Eq. (7) and the microscopic expression, Eq. (57), for the quasiparticle energy, combined with Eq. (58) for the renormalization constant we arrive at

$$f_{\vec{k},\vec{k}'} = Z_{\vec{k}} Z_{\vec{k}'}\,\frac{\delta\Sigma(\vec{k},\mathcal{E}_{\vec{k}})}{\delta\mathcal{N}_{\vec{k}'}}\,, \tag{62}$$

where we continue to ignore the spin for simplicity. The problem with this expression is that the quasiparticle occupation number $\mathcal{N}_{\vec{k}'}$ is not a well-defined microscopic quantity. However, one can calculate the change in the Green function that follows from a change in the corresponding occupation number $n_{\vec{k}'}$ of the non-interacting ground state from which the interacting state is supposed to arise under adiabatic switching-on of the interaction. According to Landau's hypothesis of continuity, such a change will result in an identical change of the quasiparticle occupation number, while at the same time propagating through the expression for the self-energy to produce the desired $\delta\Sigma$. The analysis is quite complex (see Section 8.5.5 of Ref. [2] for details) but the final result is simple, at least formally:

$$f_{\vec{k},\vec{k}'} = Z_{\vec{k}} Z_{\vec{k}'} \lim_{\omega\to 0}\lim_{q\to 0} \Gamma_{\vec{k}\mathcal{E}_{\vec{k}};\vec{k}'\mathcal{E}_{\vec{k}'}}(q,\omega)\,, \tag{63}$$

where $\vec{k}$ and $\vec{k}'$ lie on the Fermi surface and $\Gamma_{\vec{k}\mathcal{E}_{\vec{k}};\vec{k}'\mathcal{E}_{\vec{k}'}}(q,\omega)$ is the probability *amplitude* for the two particles with momenta $\vec{k}$ and $\vec{k}'$ and energies $\mathcal{E}_{\vec{k}}$ and $\mathcal{E}_{\vec{k}'}$ respectively to scatter against each other with momentum and energy transfers q and ω, where both q and ω tend to zero in the order specified by Eq. (63) (i.e., q must tend to zero before ω does). Notice that this is the scattering amplitude between two particles embedded in the many-body system. As such it has a very complex diagrammatic representation but it can, in principle be calculated by the methods of many-body theory. In the simplest approximation, i.e., first order perturbation theory, the calculation becomes trivial and we find

$$f_{\vec{k},\vec{k}'} = -v_{\vec{k}-\vec{k}'}\,, \tag{64}$$

where $v_{\vec{k}-\vec{k}'}$ is the Fourier transform of the Coulomb interaction.[13]

[13] This formula is valid for same-spin electrons. If the electrons have opposite spin the first order interaction is zero.

The Landau interaction function $f_{\vec{k},\vec{k}'}$, from which the Landau parameters are extracted, should not be confused with the effective interaction between quasiparticles which is used to calculate the quasiparticle lifetime and the collision term in the kinetic equation of Section 3.5. The essential difference, already evident from Eq. (63), is that the Landau interaction function is the limit of the microscopic scattering amplitude for zero momentum and energy transfer, with the additional specification that the momentum transfer q tends to zero before the frequency transfer ω. In contrast to this, the effective interaction determines the scattering amplitude for all momentum transfers such that $0 < q \simeq 2k_F$, which connect two wave vectors on the Fermi surface and, again, for small energy transfers $\omega \ll \epsilon_F$, such that the quasiparticles remain close to the Fermi surface. Even in the limit of small q this is different from the Landau interaction function because the energy transfer goes to zero first![14]

This leaves us with the problem of extending the Landau interaction function to finite q, thus generating what is known as an effective interaction between quasiparticles [16–19]. This is an extremely difficult task, even if one limits oneself to seeking a "local" interaction, which depends only on q and ω and is averaged over the quasiparticle momenta $\vec{k}$ and $\vec{k}'$ over the Fermi surface. A very useful concept in this context has been that of the dimensionless *many-body local field factors*, denoted by $G^{s(a)}(q)$. The local field factors were originally introduced to generate approximate expressions for the density and spin response functions of a Fermi liquid beyond the random phase approximation: this concept is described in detail in Section 5.4 of Ref. [2]. It is now understood [17] that the local field factors are in a very precise sense the finite-q extension of the $\ell = 0$ Landau parameters:

$$-\lim_{q\to 0} v(q)\, G^{s(a)}(q) = f_0^{s(a)}\,, \tag{66}$$

where $v(q)$ is the Fourier transform of the bare interaction (Coulomb interaction for an electron liquid).[15] Armed with this understanding and with a decent microscopic calculation of the local field factors (see Ref. [2], Appendix 11) it is relatively easy to construct a local effective interaction appropriate to the problem at hand.[16] The most famous effective interaction between quasiparticles is perhaps the Kukkonen-Overhauser (KO) interaction [16], which is especially suitable for calculations of the superconducting transition temperature. The explicit form of this interaction is

$$V^s_{\text{eff}}(q,\omega) = v(q) + \big(v(q)(1-G^s(q))\big)^2 \frac{\chi_0(q,\omega)}{1 - v(q)(1-G^s(q))\chi_0(q,\omega)}\,, \tag{67}$$

[14] It can be shown that the relation between the amplitudes calculated for $q \to 0$ with these two different orders of limits is

$$A_\ell^{s(a)} = f_\ell^{s(a)}/(1+F_\ell^{s(a)}) \tag{65}$$

where $F_\ell^{s(a)}$ are the Landau parameters and $A_\ell^{s(a)}$ is the ℓ-th component in the Legendre-polynomial expansion of the $q \to 0$ limit of the scattering amplitude in the symmetric (s) or antisymmetric (a) spin channel.

[15] The quantity $-v(q)\, G^{s(a)}(q)$, often generalized to finite frequency, is known in the density functional theory literature as the "exchange-correlation kernel".

[16] Different effective interactions arise depending on whether one consider the interaction between two particles extraneous to the Fermi liquid, e.g., two charged impurities in the electron gas, an extraneous particle and a quasiparticle, or two quasiparticles. Even in the latter case differences arise depending on if one wants to calculate a superconducting transition temperature or a self-energy, since different classes of diagrams are involved in the two cases.

for the spin-symmetric channel (s), and

$$V^a_{\rm eff}(q,\omega) = \big(v(q)G^a(q)\big)^2 \frac{\chi_0(q,\omega)}{1+v(q)G^a(q)\chi_0(q,\omega)}\,, \tag{68}$$

for the spin antisymmetric channel (a), where $\chi_0(q,\omega)$ is the well-known Lindhard function, i.e., the density/spin response function of the non-interacting electron gas, and $V^{s(a)}_{\rm eff} \equiv V^{\uparrow\uparrow}_{\rm eff} + (-)V^{\uparrow\downarrow}_{\rm eff}$. An application of the KO interaction to a study of superconductivity in an electron-hole liquid can be found in Ref. [18]. For a very recent application of this interaction to the study of superconductivity in elemental metals see Ref. [19].

5 Fermi liquid of massless Dirac fermions

We now discuss some peculiarities of the Fermi liquid formed by "massless Dirac fermions" (MDF). MDFs occur near the crossing of two bands, such as occurs in a single layer of C atoms (graphene) and in numerous Dirac and Weyl semimetals: the band dispersion is linear in k near the crossing point, which we set at $k = 0$ (see Fig. 11): $\epsilon_k = \hbar v k$ where v is the fermion velocity, independent of momentum. A Fermi liquid is realized when the Fermi level crosses one of the bands (say the upper one) in the vicinity of the crossing. The existence of a non-vanishing Fermi wave vector k_F, related to the density of electrons in the upper band by the usual Fermi liquid relation, $k_F^2 = 2\pi n$, and a non-vanishing Fermi energy $\epsilon_F = \hbar v k_F$, are sufficient to establish the existence of a normal Fermi liquid at sufficiently low temperature $k_B T \ll \epsilon_F$. This is because the structure of the low-energy excitations in the MDF system is indistinguishable from that of the low-energy excitations in an ordinary parabolic band:

$$\epsilon_k = \epsilon_F + \hbar v\big(k-k_F\big) = \epsilon_F + \hbar\frac{k_F}{m_c}\big(k-k_F\big)\,, \tag{69}$$

where $m_c \equiv \hbar k_F/v$, also known as the "cyclotron mass", plays the same role as the bare electron mass for parabolic bands. Unlike the bare electron mass, however, m_c is density-dependent, scaling as $n^{1/2}$ in 2D, and thus vanishing as the Fermi level approaches the crossing point for $k_F \to 0$. In view of this correspondence, we expect the Fermi liquid concept to be robust in this new situation, in particular we expect to find long-lived quasiparticles with an effective mass m^* that is somewhat different from m_c due to the presence of electron-electron interaction. Nevertheless there are some important differences to be kept in mind. We list them below.

1. First of all, the range of validity of the Fermi liquid theory shrinks to zero as the Fermi energy approaches the crossing point of the bands, also known as the "Dirac point". When $\epsilon_F = 0$ there is no Fermi surface and no Fermi liquid. The length and time scales provided by k_F^{-1} and $\hbar\epsilon_F^{-1}$ diverge, and the system becomes scale-invariant (or quantum-critical). The only energy scale left is $k_B T$ itself and the inverse lifetime of electrons and holes must be proportional to $k_B T$, which is of the same order of magnitude as the energy of the excitations. This defines the so-called "Planckian regime", and we see that the whole

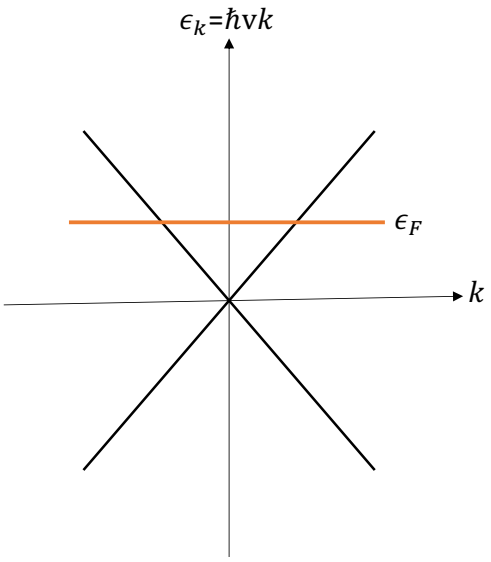


Fig. 11: *Band dispersion for massless Dirac Fermions in the vicinity of the crossing point (Dirac point). Also shown is the Fermi level which defines our Fermi liquid.*

concept of quasiparticle falls apart in this regime. The nature of the quantum critical state at $\epsilon_F = 0$ (also known as the charge neutrality point) is not completely understood at present. Strong electron-electron interactions have been dealt with by a hydrodynamic description, where individual particles are superseded by collective variables such as density and current. Alternatively, it has been suggested that electrons and holes in the upper and lower band bind together to produce a gapped state known as "excitonic insulator".

2. Although, for finite k_F, the structure of the low-energy excitations remains the same as in the standard parabolic case, there are some important differences in the structure of higher energy excitations. In particular, the presence of electrons in the fully occupied lower band cannot be ignored, as it gives a significant contribution to the Fermi liquid properties. These contributions fall into two categories: (i) contributions to the Landau interaction function arising from interactions between the electrons near the Fermi surface and those in the fully occupied bands, (ii) contributions to the quasiparticle lifetime arising from interband transitions at energies lower than or comparable to the Fermi energy.

   Concerning (i) it must be noted that the relative strength of the electron-electron interaction, as measured by the ratio of the potential energy to the kinetic energy, is no longer density-dependent: rather it becomes a fixed constant $\alpha = \frac{e^2}{\hbar v}$ of order 1. However, there is now another measure of the importance of interaction effects, and that is Λ/k_F where Λ is an ultraviolet cutoff wave vector, which determines the largest momentum of the occupied states for which the linear (massless) band model is still valid. This cutoff is poorly defined, but is expected to be of order $1/a$, where a is the lattice constant. So even though α is constant, interaction effects become stronger as k_F tends to zero, which is similar to the familiar situation, but leads to very different phenomenology in this case. For example it can be shown that the inverse compressibility, proportional to $\partial\mu/\partial n$,

is increased by interactions rather than decreased [20]. This happens because when the electronic density is increased the Fermi level in the upper band moves farther away from the lower band: the negative exchange energy that is lost due to this effect outweighs the negative exchange energy that is gained by having more electrons in the upper band. The same phenomenon is observed for the spin susceptibility, which is now suppressed, rather than enhanced, by interactions.

Concerning point (ii), a detailed analysis presented in Ref. [21] shows that the quasiparticle lifetime (in the upper band, $+$) is given by

$$\frac{1}{\tau^{(\mathrm{e})}_{\boldsymbol{k},+}} \simeq \frac{\varepsilon_{\mathrm{F}}}{\hbar}\frac{1}{\pi N(0)}\left(\frac{\xi_{\boldsymbol{k},+}}{\varepsilon_{\mathrm{F}}}\right)^2 \ln\left(\frac{\Lambda}{\xi_{\boldsymbol{k},+}}\right), \tag{70}$$

at zero temperature and

$$\frac{1}{\tau^{(\mathrm{e})}_{\boldsymbol{k},+}} \simeq \frac{\varepsilon_{\mathrm{F}}}{\hbar}\frac{\pi}{2N(0)}\left(\frac{k_{\mathrm{B}}T}{\varepsilon_{\mathrm{F}}}\right)^2 \ln\left(\frac{\Lambda}{k_{\mathrm{B}}T}\right), \tag{71}$$

at finite temperature. This is essentially the Giuliani-Quinn result [11] discussed earlier in Section 3.3. Three main differences with respect to the classic calculation for an ordinary two-dimensional electron gas have been identified in Ref. [21]: i) a simple Fermi golden rule approach with statically screened Coulomb interactions is not viable in graphene as it yields logarithmically-divergent intra-band scattering rates due to the collinear scattering singularity; ii) the leading-order contribution to the quasiparticle decay rate in the low-energy and low-temperature limits is completely controlled by scattering events with small momentum transfer: the $2k_{\mathrm{F}}$ contributions are suppressed by the chiral nature of massless Dirac carriers in graphene; iii) because of point ii), the leading order contribution to the quasiparticle decay rate is completely independent of the strength of the background dielectric constant ϵ: the result is therefore *universal* in that it does not depend on the substrate on which graphene is placed.

3. As the Fermi level approaches the crossing point with decreasing density, the Fermi liquid concept remains in force as long as $k_F > 0$, but the effective mass of quasiparticles is found to be logarithmically suppressed relative to the noninteracting cyclotron mass [22]. Another way of saying this is that the Fermi velocity v is renormalized to $v^*(k_F) > v$ where $v^*(k_F)$ diverges logarithmically in first order perturbation theory as $k_F \to 0$:

$$\frac{v^*(k_F)}{v} = 1 + \frac{\alpha}{4}\ln\left(\frac{\Lambda}{k_F}\right). \tag{72}$$

While an increase of the velocity, leading to a reshaping of the bands near the Dirac point, has been experimentally observed, the divergence of $v^*(k_F)$ poses a problem of legitimacy for the microscopic perturbation theory on which this prediction is based. Clearly a non-perturbative approach is needed to analyze the $k_F \to 0$ limit. This approach is provided by the renormalization group, which is reviewed in Ref. [22], and generally confirms the predictions of the weak coupling theory.

6 Non-Fermi-liquid behavior

6.1 Disordered electron liquid

Thus far we have only considered translationally invariant electron liquids. But real electron liquids are inevitably subjected to potentials that break translational invariance, such as the periodic potential in a crystal lattice, or the potential from randomly distributed impurities.

A general approach to inhomogeneous interacting Fermi liquids begins with a consideration of the exact eigenstates $|\phi_\alpha\rangle$ of a single particle – an electron in our case – in the external potential. These are Bloch waves in a perfectly periodic crystal lattice, but have no definite symmetry in the presence of a random impurity potential. In either case, the ground-state of the noninteracting system is obtained by singly occupying the eigenstates with the N lowest energies ε_α. The highest occupied eigenvalue defines a Fermi energy $\epsilon_F = \varepsilon_N$, but not a Fermi surface. Excited states are obtained by promoting some electrons from below the Fermi level to above the Fermi level. All these states are described by a set of occupation numbers $\mathcal{N}_\alpha$.

We now start from one of these states, and slowly turn on the electron-electron interaction, expecting to generate long-lived interacting states characterized by a quasiparticle distribution $\mathcal{N}_\alpha$. But here comes an important difference. While in the homogeneous case the interaction between quasiparticles is hindered by the Pauli exclusion principle *and* by the conservation of momentum and energy, in an inhomogeneous system the constraint of momentum conservation does not exist. It is therefore expected that the quasiparticles will interact more strongly and decay more rapidly than in the homogeneous case. How much more rapidly? This is an essential question, since the very existence of the Fermi liquid requires that the decay rate of a quasiparticle of energy ε tend to zero more rapidly than $\varepsilon - \epsilon_F$ when $\varepsilon \to \epsilon_F$.

The answer depends on whether the system is periodic or disordered. In periodic systems crystal momentum (the Bloch wave vector) is conserved up to reciprocal lattice vectors. While the occurrence of "umklapp" processes can alter the numerical value of the lifetime it does not lead to qualitative departures from the homogeneous picture.

In disordered systems the lack of momentum conservation has more serious consequences. First of all, if disorder is sufficiently strong, it can lead to *localization* of the quasiparticle states and hence change the electrical properties of the system from metal to insulator. We will not pursue this scenario here. Even in the weak disorder regime, i.e., for $k_F \ell \gg 1$, where ℓ is the electron mean free path, the combined effects of disorder and interactions can be significant. Because density fluctuations relax at a slower pace than in a perfect crystal the electrons within them stay together for a longer time and hence interact more strongly: this "electron loitering" leads to an enhanced quasiparticle decay rate. In three dimensions the decay rate goes as $\frac{1}{\tau^{(e)}} \sim (\varepsilon - \epsilon_F)^{3/2}$ and $\frac{1}{\tau^{(e)}} \sim (k_B T)^{3/2}$ in the limits of $k_B T \ll \varepsilon - \epsilon_F$ and $k_B T \gg \varepsilon - \epsilon_F$ respectively [23]. This is considerably larger than $(\varepsilon - \epsilon_F)^2$ and $(k_B T)^2$ yet still small compared to $\varepsilon - \epsilon_F$. In two dimension, on the other hand, one finds $\frac{1}{\tau^{(e)}} \sim (\varepsilon - \epsilon_F)$ at $T = 0$, and $\frac{1}{\tau^{(e)}} \sim (k_B T) \ln(k_B T)$ at finite temperature [24, 23], which implies that the conventional Fermi liquid picture is, at best, marginally valid.

In concomitance with the enhancement of the quasiparticle decay rate, the single-particle density of states, $N(\varepsilon)$, is *reduced* in the vicinity of the Fermi level, going as $(\varepsilon-\epsilon_F)^{1/2}$ in three dimensions, and $|\ln(\varepsilon-\epsilon_F)|$ in two dimensions [25]. These results are based on perturbation theory, and therefore cannot be trusted when the correction to the unperturbed density of states becomes too large (e.g., in two dimensions, for $\varepsilon \to \epsilon_F$). Notice, however, that a suppression of the density of states at the Fermi level is also predicted in the limit of strong disorder, even though the physics appears to be quite different in that regime. Other thermodynamic and transport properties, such as the heat capacity and the conductivity, show non-analytic behavior in T at low temperature. Remarkably, there is no non-analytic correction to the compressibility.[17]
Going beyond perturbation theory, in the true spirit of the Landau theory of Fermi liquid, has turned out to be an extremely difficult task, and the problem remains unsolved to date (for a relatively recent review, see Ref. [27]). Suffice it to say that the scattering amplitudes in the antisymmetric spin channel have been predicted to diverge in the low energy sector, suggesting a diverging spin susceptibility associated with the formation of ferromagnetic domains at the length scale of the divergence. At the same time the inverse conductance is seen to initially increase, as if the system were going towards an insulating state, but it then reaches a maximum and begins to decrease as the divergence of γ^a takes over. These findings show very clearly that the noninteracting theory of localization needs serious revision in the presence of electron-electron interactions.

6.2 Luttinger liquid

An interesting phenomenon, known by the colorful name of "orthogonality catastrophe", lies at the heart of the distinction between Fermi liquids and non Fermi liquids. Imagine injecting an extra electron into the ground-state of a strongly correlated N-electron system. Since the new electron lacks the appropriate correlations with the pre-existing electrons, the state of the $N+1$-electron system after the injection is essentially orthogonal to the ground-state. In mathematical terms, the orthogonality catastrophe implies that the renormalization constant Z, defined in Section 4, vanishes in the thermodynamic limit.[18] We take this to be the defining feature of a non-Fermi liquid state.
A classic example of non-Fermi liquid behavior is the so-called "Luttinger liquid" [28], which is realized in quasi-one dimensional electronic systems such as Bechgaard salts, TTF-TCNQ, and carbon nanotubes, as well as in confined systems of fermionic cold atoms. The reduced effective dimensionality of these systems hinders single particle motion to the point that the particles must be regarded as strongly correlated even when their interactions are weak.
An immediate consequence of this situation is the disappearance of the quasiparticle δ-function peak in the spectral function $A(k_F, \omega)$ at the chemical potential: there are no single-electron

[17] It should mentioned, for completeness, that additional non-analytic corrections to the density of states and the specific heat arise when the transverse electromagnetic interaction between the electrons is taken into account. See Ref. [26] for details.

[18] It can be shown that the renormalization constant is the square of the overlap between the excited state of the system immediately after the injection of an electron and the ground-state of the system.

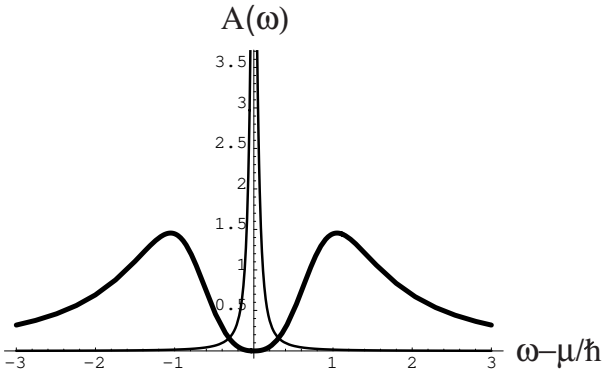


Fig. 12: *Schematic behavior of the local spectral function at $A(k_F, \omega)$ for a Luttinger liquid in the weak coupling regime (thin line) and in the strong coupling regime (thick line). In both cases $A(k_F, \omega) \sim \frac{1}{|\omega - \mu/\hbar|^\nu}$ for $\omega \to \mu/\hbar$, with the exponent ν tending to 1 for weak interactions and becoming negative for strong interactions. Notice the absence of the quasiparticle δ-function peak at $\omega = \mu/\hbar$.*

quasiparticles. For weak interactions the δ-function peak is replaced by a power-law divergence for $\omega \to \mu/\hbar$. With increasing coupling strength a sort of energy gap develops, whereby $A(k_F, \omega)$ vanishes with a power law for $\omega \to \mu/\hbar$ as shown in Fig. 12. The position of the lateral maxima in the spectral function is a rough measure of the energy of the disturbance created by the injection of the new electron in the liquid, while the "width" of these maxima is inversely proportional to the time needed for the many-electron system to adjust to the presence of the new electron. Another consequence of the vanishing of Z is that the plane-wave occupation number n_k is no longer discontinuous at $k = k_F$, even though a singularity persists in its derivatives with respect to k.

Luttinger liquids exhibit anomalous transport properties. For example, the electrical conductivity is expected to vanish at zero temperature. One might find this not so surprising since it is known that in a one dimensional system any amount of random disorder causes localization of the one-electron states, and hence a vanishing conductivity at $T = 0$. But, in the Luttinger liquid any perturbation that breaks translational invariance, e.g., *even a single impurity*, leads to an insulating state at $T = 0$. The physical reason for this effect is that the perfectly clean system is on the verge of spontaneously forming a charge density wave (CDW) of wave vector $2k_F$. Under these conditions even a single impurity can pin down an insulating CDW state.

6.3 Fractional quantum Hall liquid

It is still an open question whether non-Fermi liquid behavior occurs in more than one dimension. Although many theories suggest that this should be the case, for example in the normal phase of the superconducting cuprates, the experimental evidence remains somewhat ambiguous. A notable exception is the fractional quantum Hall liquid which occurs in the two-dimensional electron liquid at high magnetic field. This system presents the most radical departures from Landau Fermi liquid theory so far encountered in any condensed matter system, and the experimental evidence is extremely strong.

The basic reason for this exotic behavior is the *quenching of the kinetic energy* at high magnetic field. The kinetic energy of an electron in a magnetic field B is quantized in units of $\hbar\omega_c$ (where $\omega_c = \frac{eB}{mc}$ is the cyclotron frequency) so that its admissible values are

$$\epsilon_n = \left(n + \frac{1}{2}\right)\hbar\omega_c \tag{73}$$

where n is a non-negative integer. These energy levels are known as *Landau levels*. The number of degenerate states in each Landau level is proportional to the magnetic field and to the area ($\mathcal{A}$) of the system, and is given by

$$N_L = \frac{eB\mathcal{A}}{hc}\ . \tag{74}$$

Thus, in the limit of large magnetic field all the electrons fall in the lowest Landau level and the kinetic energy becomes a constant $N\frac{\hbar\omega_c}{2}$, where N is the number of electrons. The key parameter that controls the properties of the system is the *filling factor*

$$\nu = \frac{N}{N_L}\ , \tag{75}$$

which is less than 1 for a fractionally filled lowest Landau level. In this situation, the non-interacting ground-state is highly degenerate because the electrons can be distributed in many different ways among orbitals of the lowest Landau level. The degeneracy is removed, however, by the interaction. Under these conditions, the concept of an adiabatic connection between non-interacting and interacting states, which is central to the Landau theory of Fermi liquids, has no meaning, since the non-interacting limit of an interacting state is a completely ill-defined notion. Indeed, the incompressible quantum Hall liquid state, introduced by Laughlin [29] to explain the quantum Hall effect at filling factors of the form $\nu = \frac{1}{2k+1}$, where k is a positive integer, is not connected in any obvious way to a non-interacting state. The many-body wave function that describes it has multiple zeroes (of order $2k+1$) on the hypersurfaces of configuration space on which two electrons come in contact: in this sense, the zeroes of the wave function are "bound" to the particles. These multiple zeroes are far more powerful than the simple zeroes required by the Pauli exclusion principle in a non-interacting, and are ultimately responsible for giving the Laughlin state a particularly low interaction energy.

An exotic connection to a non-interacting state continues to exist, however, as has been pointed out by Jain, based on the beautiful idea of *composite fermions* [30]. Composite fermions are constructed by attaching to each electron an infinitely thin flux tube carrying an even number $2k$ of magnetic flux quanta $\Phi_0 \equiv \frac{hc}{e}$. These composite particles, like the original electrons, are subject to the external magnetic field B and interact with each other via the Coulomb interaction. The vector potential produced by the magnetic flux tubes exerts no force on the particles, and therefore does not contribute to the particle-particle interaction. When two composite particles are adiabatically interchanged along a path that does not enclose other particles the wave function is multiplied by an Aharonov-Bohm phase factor $e^{i\pi(2k+1)} = -1$: this indicates that the composite particles are fermions and the problem of interacting composite particles is mathematically identical to the original problem of interacting electrons. While the transformed

problem appears at first sight more difficult than the original, the advantage of this description appears when one introduces a mean field approximation. To this end, the fictitious magnetic fluxes attached to the particles are spread out, creating a uniform and constant magnetic flux proportional to the particle density. This average magnetic flux combines with the external physical flux, yielding an effective flux

$$B^*\mathcal{A} = B\mathcal{A} - 2k\Phi_0 N \tag{76}$$

where the flux tubes have been assumed to be antiparallel to the external field for $k > 0$ and parallel for $k < 0$. The composite fermions feel the effective field B^* and their effective filling factor ν^* is therefore given by (see Eqs.(74), (75) and (76))

$$\frac{1}{\nu^*} = \frac{1}{\nu} - 2k \; . \tag{77}$$

Thus, by cleverly choosing the values of $2k$ and ν, it is possible to convert a strongly-correlated problem of electrons at fractional filling ν into a weakly correlated problem of composite fermions at *integral* filling factor ν^*. For example if $\nu = \frac{1}{3}$ and $2k = 2$, then we have $\nu^* = 1$, which corresponds to a full Landau level of composite fermions. But the non-interacting ground-state at integer filling factor is perfectly well defined (one simply has to completely fill an integer number of Landau level) and provides a suitable starting point for an analytic continuation when the interactions are turned on. Although this conversion is inspired by a questionable mean field approximation, it provides a conceptual framework for constructing correlated wave functions of excellent quality *for electrons* starting from uncorrelated wave functions for composite fermions. The composite fermion picture also establishes a beautiful connection between the seemingly different phenomena of the integral and fractional quantum Hall effects. According to this picture, the fractional quantum Hall effect of electrons is nothing but the integral quantum Hall effect of composite fermions.

The special case of a half-filled Landau level ($\nu = \frac{1}{2}$) provides an extreme example of the power of the composite fermion idea. In this case, Eq. (77), with $2k = 2$, yields $\nu^* = \infty$, implying that the composite fermions experience no magnetic field on the average (i.e., the number of occupied Landau levels of composite fermions is infinite). If the mean field approximation makes sense one would then expect to see here some of the characteristic signatures of a two-dimensional Fermi liquid, such as a two-dimensional Fermi surface [31]. In fact, surface acoustic wave propagation experiments have provided considerable evidence in support of this prediction. However, one must keep in mind that the "Fermi liquid" behavior is limited to density and current response properties: it certainly does not apply to single-particle properties. For example, tunneling experiments show a pseudogap in the spectral density of one-electron excitations, characteristic of an "orthogonality catastrophe", and no sign of the quasiparticle peak characteristic of an ordinary Fermi liquid.

The *edge* of a 2D electron liquid at high magnetic field provides yet another example of non Fermi liquid behavior. It can be shown that the collective oscillations of the density in such an edge are dynamically equivalent to the collective oscillations of a *chiral Luttinger Liquid* [32],

i.e., a Luttinger liquid in which, after diagonalizing the electron-electron interaction, only one half of the bosons – those propagating to the right *or* those propagating to the left – are retained. Unlike previous putative realizations of the Luttinger liquid paradigm, the edges of the 2D electron liquid are essentially free of disorder effects and, more importantly, the Luttinger liquid coupling constant coincides with the bulk filling factor ν. This has allowed a rather detailed experimental verification of the non-universal exponents in the power-law decay of the correlation functions of the Luttinger liquid. Studies of the tunneling current between edges in the fractional quantum Hall regime have also provided the first convincing evidence of fractionally charged excitations in condensed matter systems.

References

[1] L.D. Landau, Sov. Phys. JETP **3**, 920 (1957); *ibid.* **5**, 101 (1957); *ibid.* **8**, 70 (1959)

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

[3] D. Pines and P. Noziéres: *The Theory of Quantum Liquids, Vol. I* (W.A. Benjamin, 1966)

[4] Z. Qian, G. Vignale, and D.C. Marinescu, Phys. Rev. Lett. **93**, 106601 (2004)

[5] H. Yasuhara and Y. Ousaka, Int. J. Mod. Phys. B **6**, 3089 (1992)

[6] D.S. Greywall, Phys. Rev. A **27**, 2747 (1983)

[7] D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984)

[8] J.P. Eisenstein, L.N. Pfeifer, and K.W. West, Phys. Rev. Lett. **68**, 674 (1994)

[9] J. Zhu *et al.*, Phys. Rev. Lett. **90**, 226401 (2004)

[10] K. Vakili *et al.*, Phys. Rev. Lett. **92**, 056805 (2003)

[11] G.F. Giuliani and J.J Quinn, Phys. Rev. B **26**, 4421 (1982)

[12] S. Murphy: *A Brief Guide to Electronic Lifetimes in 2D*, in G.F. Giuliani and G. Vignale (eds.): *The electron liquid paradigm in condensed matter physics*, Proceedings of the International School of Physics "Enrico Fermi", Course CLVII (IOS Press, Amsterdam, 2004)

[13] A.A Abrikosov and I.M. Khalatnikov, Rep. Prog. Phys. **22**, 329-367 (1959)

[14] H.-H. Fu and C. Ebner, Phys. Rev. A **10**, 338 (1974)

[15] D.R. Hamann and A.W. Overhauser, Phys. Rev. **143**, 183 (1966)

[16] C.A. Kukkonen and A.W. Overhauser, Phys. Rev. B **20**, 550 (1979)

[17] G. Vignale and K.S. Singwi, Phys. Rev. B **32**, 2156 (1985)

[18] G. Vignale and K.S. Singwi, Phys. Rev. B **31**, 2729 (1984)

[19] C.A. Kukkonen and Kun Chen, Phys. Rev. B **104**,195142 (2021)

[20] M. Polini *et al.*, Solid State Commun. **143**, 58–62 (2007)

[21] M. Polini and G. Vignale: *The quasiparticle lifetime in a doped graphene sheet*, in M. Polini, G. Vignale, V. Pellegrini, and J.K. Jain (eds.): *No Nonsense Physicist – an overview of Gabriele Giuliani's work and life*, Edizioni della Normale (2015)

[22] V.N. Kotov *et al.*, Rev. Mod. Phys. **84**, 1067 (2012)

[23] A. Schmid, Z. Phys. **271**, 251 (1974)

[24] E. Abrahams, P.W. Anderson, P.A Lee, and T.V. Ramakrishnan, Phys. Rev. B **24**, 6783 (1981)

[25] B.L. Altshuler and A.G. Aronov, in A.L. Efros and M. Pollak (eds.): *Electron-electron interactions in disordered systems* (North Holland, Amsterdam, 1985), p. 1–153

[26] M. Yu. Reizer, Phys. Rev. B **40**, 11571 (1989)

[27] C. Di Castro and R. Raimondi: *Disordered electron systems*, in G.F. Giuliani and G. Vignale (eds.): *The electron liquid paradigm in condensed matter physics*, Proceedings of the International School of Physics "Enrico Fermi", Course CLVII (IOS Press, Amsterdam, 2004)

[28] F.D.M. Haldane, J. Phys. C: Solid State Phys. **14**, 2585 (1981)

[29] R.B. Laughlin, Phys. Rev. B **23**, 5632 (1983)

[30] J.K. Jain, Phys. Rev. Lett. **63**, 199 (1989)

[31] B.I. Halperin, P.A. Lee, and N. Read, Phys. Rev. B **47**, 7312 (1993)

[32] X.G. Wen, Phys. Rev. B **41**,12838 (1990)

3 The Physics of Quantum Impurity Models

Jan von Delft
Arnold Sommerfeld Center for Theoretical Physics
Ludwig-Maximilians-Universität München
Theresienstr. 37, 80333, Munich

Contents

1 Introduction **2**

2 Single-impurity Anderson model: local moment formation **3**

3 Kondo model: spin-exchange interaction **5**

4 Numerical renormalization group **8**

5 NRG results for the single-impurity Anderson model **12**

6 Two-channel Kondo model **17**

A Schrieffer-Wolff transformation **19**

1 Introduction

A *quantum impurity model* describes a localized, discrete quantum system, the *impurity*, coupled to noninteracting excitations with a continuous (non-discrete) excitation spectrum, the *bath*. Such models were first invoked in the 1960s to explain certain anomalies in the resistivity of magnetic alloys, where the impurities are localized magnetic moments (or localized spins, for short). In subsequent decades, quantum impurity models turned out to be relevant in several other contexts, including transport through quantum dots and nanotubes, whose discrete energy levels act as impurities; dissipation and decoherence of qubits, described as impurities coupled to bosonic baths; and dynamical mean field theory, where a single lattice site is viewed as an impurity coupled to a self-consistently determined bath. This lecture aims to introduce the physics of quantum impurity models from a renormalization group perspective. By way of introduction, we begin with some historical remarks.

In 1934, de Haas, de Boer and van den Berg [1] discovered an unexpected phenomenon when measuring the resistance of gold as a function of temperature, $\rho(T)$: it showed a *minimum* at $T = 3.70$ K, whereas it had been expected that the resistance would decrease monotonically with decreasing temperature as phonons freeze out, reaching its smallest value at $T = 0$. The fact that $\rho(T)$ instead turns upward once T drops below 3.70 K indicates that in this regime the electron scattering rate increases with decreasing temperature—a finding that was very surprising at the time. In subsequent years, similar resistance minima were found in numerous so-called dilute magnetic alloys: metals such as Cu, Ag, Au containing a small concentration of magnetic impurities such as Cr, Mn, Fe.

The origin of the resistance minima observed for magnetic alloys remained puzzling until 1964, when Kondo offered an explanation [2]. He considered a model, since known as the *Kondo model*, describing a dilute concentration of localized spin-1/2 impurities immersed in a metallic conduction band, with a spin-exchange coupling between the localized spins and the conduction band. This coupling causes spin-flip scattering events: when a conduction electron scatters off a localized spin, the spin of both can flip. Kondo computed the corresponding scattering rate to lowest nontrivial order in the exchange coupling, and found that it *increases* with decreasing temperature, causing the resistance $\rho(T)$ to likewise increase. This explained the resistance minimum.

Moreover, Kondo found that the scattering rate, $\gamma(T)$, increased *logarithmically* with decreasing T, like $\ln(D/T)$, where D is a large energy scale such as the bandwidth of the conduction band. This meant that while he had resolved an experimental puzzle, he had simultaneously discovered a theoretical one, which came to be known as the *Kondo problem*: what happens to the scattering rate in the limit $T \to 0$? The perturbative result, implying a logarithmic divergence, clearly can not be trusted in this limit.

To answer this question, nonperturbative methods were called for. Over the years, a wide variety of approaches were applied to the Kondo problem, including the numerical renormalization group (NRG) [3, 4], the Bethe Ansatz [5, 6], conformal field theory [7], bosonization [8, 9], and Monte Carlo methods [10]. As a result, the answer to the above question is now very well

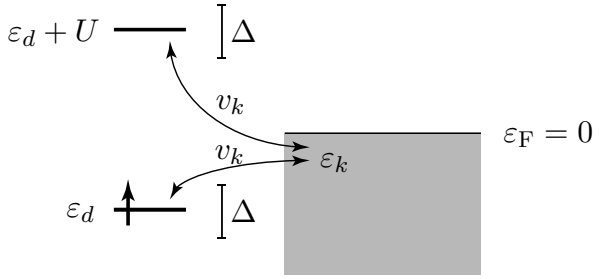
established: the scattering rate $\gamma(T)$ approaches a constant value for $T \to 0$, because the local spin is *screened* by conduction electrons to form a *spin singlet*. The screened singlet acts just like a static (i.e. nonmagnetic) impurity, off which other electrons scatter very strongly, albeit without flipping their spin. The screening of a localized spin by conduction electrons has come to be known as the *Kondo effect*.

The crossover, with decreasing T, from an unscreened spin to a spin singlet is a paradigmatic example of a renormalization group (RG) flow between two RG fixed points. The goal of these notes is to elucidate the physics of quantum impurity models from such an RG perspective.

In the following sections, we will introduce two paradigmatic models describing spin screening: the single-impurity Anderson model and the Kondo model; explain why a perturbative treatment fails at low temperatures; describe Wilson's numerical renormalization group approach for reaching low temperature; and discuss various physical quantities exhibiting signatures of spin screening. In the final section, we briefly discuss a model that does not show full spin screening—the two-channel Kondo model. We set $\hbar = 1$ and $k_{\mathrm{B}} = 1$ throughout.

2 Single-impurity Anderson model: local moment formation

In 1961, P.W. Anderson introduced a "highly simplified" model, since known as the single-impurity Anderson model (SIAM), to explain the formation of local moments in magnetic alloys. In such systems, the magnetic impurities have d or f orbitals that are strongly localized, yet nevertheless hybridize with the host metal, in such a manner that they behave as localized spins (or magnetic moments) acting as dynamical degrees of freedom. A fully realistic description is very challenging, since d or f shells typically host several electrons, with strong local Coulomb interactions. Anderson argued that the essential physics could be captured by considering just a single, local level rather than an entire shell. He thus studied the Hamiltonian $H_{\mathrm{SIAM}} = H_{\mathrm{bath}} + H_{\mathrm{loc}} + H_{\mathrm{hyb}}$, with



$$H_{\mathrm{bath}} = \sum_{ks} \varepsilon_k \hat{n}_{ks} \,, \qquad \hat{n}_{ks} = c_{ks}^\dagger c_{ks} \,, \tag{1a}$$

$$H_{\mathrm{loc}} = \sum_{s} \left(\varepsilon_d - \tfrac{1}{2} hs\right) \hat{n}_{ds} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow} \,, \qquad \hat{n}_{ds} = d_s^\dagger d_s \,, \tag{1b}$$

$$H_{\mathrm{hyb}} = \sum_{ks} v_k \left(c_{ks}^\dagger d_s + d_s^\dagger c_{ks}\right) . \tag{1c}$$

Here, H_{bath} describes a band of free (noninteracting) electrons with spin $s \in \{\uparrow, \downarrow\}$ and energy ε_k (k is a continuous momentum index), H_{loc} describes a discrete, localized level with energy ε_d, spin s, Zeeman splitting h if a magnetic field is applied (we set $g\mu_{\text{B}} = 1$), and Coulomb energy cost U if the level is doubly occupied. H_{hyb} describes *hybridization* between the d level and the conduction band, allowing electrons to hop back and forth between the two with amplitude v_k (we choose it real and constant, $v_k = v$). As a result, the d level acquires a finite width, Δ.
Since H_{bath} and H_{hyb} are quadratic, the conduction band can be integrated out, in principle. One then finds that the dependence of the d level dynamics on bath parameters enters only via the *hybridization function*

$$\Delta(\omega) = \sum_k \frac{v_k^2}{\omega - \varepsilon_k + \mathrm{i}0^+} \,. \tag{2}$$

When this function is expressed through its spectral representation, $\Delta(\omega) = \int \mathrm{d}\epsilon \frac{\Gamma(\epsilon)}{\omega - \epsilon + \mathrm{i}0^+}$, the corresponding *hybridization spectrum* is given by

$$\Gamma(\epsilon) = \sum_k |v_k|^2 \, \delta(\epsilon - \varepsilon_k) \,. \tag{3}$$

It characterizes how strongly a conduction band level with energy ϵ couples to the impurity. Anderson assumed the hybridization spectrum to be constant near the Fermi energy; for concreteness, we will use the box-shaped form,

$$\Gamma(\epsilon) = (\Delta/\pi)\, \Theta\big(D - |\epsilon|\big) \,, \qquad \Delta/\pi = v^2 \nu \,. \tag{4}$$

Here, D is the half-bandwidth of a flat band centered on the Fermi energy at $\varepsilon_F = 0$, related to the density of states per spin, ν, by $\nu = 1/(2D)$, and Δ is the width acquired by the d level through hybridizing with the bath. (This follows from Eq. (23) below.) Unless stated otherwise, we measure energy in units of $D = 1$.
The Hilbert space of the local level is spanned by four states, $|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, $|\uparrow\downarrow\rangle$, describing the local level being empty, occupied by one electron with spin $\uparrow$ or $\downarrow$, or doubly occupied, respectively. They are eigenstates of H_{loc}, with eigenenergies $E_0 = 0$, $E_\uparrow = \epsilon_d - \frac{1}{2}h$, $E_\downarrow = \epsilon_d + \frac{1}{2}h$, and $E_{\uparrow\downarrow} = 2\varepsilon_d + U$, respectively. Unless stated otherwise, we will consider the case of zero magnetic field, $h = 0$. We are interested in the *local moment regime*, where the parameters are chosen such that the empty and doubly-occupied levels lie well above the (broadened) singly-occupied levels; this implies $E_0 - E_s > \Delta$ and $E_{\uparrow\downarrow} - E_s > \Delta$, i.e.

$$\varepsilon_d + \Delta < 0, \qquad \varepsilon_d + U > \Delta, \tag{5}$$

Then, the average occupancy of the local level, $n_d = \sum_s \langle \hat{n}_{ds} \rangle$, is $\simeq 1$. Thus, the d level hosts a local moment, containing a spin-up or down electron with equal probability, $n_{d\uparrow} = n_{d\downarrow} \simeq \frac{1}{2}$.
Transitions between these two spin states can occur via second-order hopping processes involving $|0\rangle$ or $|\uparrow\downarrow\rangle$ as *virtual*, high-energy intermediate states. Their net result is that the spin on the

d level has flipped from $|\uparrow\rangle$ to $|\downarrow\rangle$, or vice versa, and a particle-hole excitation has been created in the band. The effective spin-flip rate due to such processes is found to be

$$J = \frac{v^2}{\varepsilon_d + U} - \frac{v^2}{\varepsilon_d} = -\frac{Uv^2}{\varepsilon_d\,(\varepsilon_d + U)} = -\frac{2U\Delta}{\pi\varepsilon_d\,(\varepsilon_d + U)}\,. \tag{6}$$

Note that $J > 0$, since in the local moment regime of Eq. (5), $\varepsilon_d < 0$. If one is interested only in the physics of these spin-flip processes, it is convenient to consider the Kondo model, which we discuss next.

3 Kondo model: spin-exchange interaction

The Kondo model (KM) is obtained from the SIAM by projecting the latter onto the subspace in which the local occupancy is strictly $n_d = 1$, i.e., only the impurity states $|\uparrow\rangle$ and $|\downarrow\rangle$ are considered, and focusing on the scattering of low-energy band excitations only. Formally, this can be achieved using a so-called Schrieffer-Wolff [11] transformation (outlined in App. A). Then, one obtains an effective Hamiltonian of the form $H_{\rm KM} = H_{\rm bath} + H_{\rm spin} + H_{\rm exchange}$, with

$$H_{\rm bath} = \sum_{ks} \varepsilon_k \hat{n}_{ks}\,, \qquad H_{\rm spin} = h\hat{\rm S}^z_d\,, \tag{7a}$$

$$H_{\rm exchange} = J\hat{\mathbf{S}}_d \cdot \hat{\mathbf{s}}_c, \qquad \hat{\mathbf{s}}_c = \sum_{ks,k's'} c^\dagger_{ks}\tfrac{1}{2}\boldsymbol{\sigma}_{ss'}c_{k's'}\,. \tag{7b}$$

This defines the KM. Here, $H_{\rm bath}$ again describes a free conduction band; $\hat{\mathbf{S}}_d$ are spin-$\frac{1}{2}$ operators acting in the space spanned by $|\uparrow\rangle$ and $|\downarrow\rangle$; and $\hat{\mathbf{s}}_c$ are conduction band spin operators, describing its spin density at the impurity site. The local spin couples to the conduction electron spin density via an antiferromagnetic exchange interaction, $J > 0$, favoring an anti-alignment of the spins of the impurity and the conduction band. The KM describes the low-energy behavior of the SIAM, at energy scales for which charge fluctuations involving $|0\rangle$ and $|\uparrow\downarrow\rangle$ are "frozen out", i.e., can occur only virtually.

The KM is an interesting model in its own right. It has been studied using a wide variety of theoretical methods, both analytical and numerical, and can be regarded as thoroughly understood. Below, we briefly summarize some of its salient properties. These will be elaborated in subsequent sections.

Perturbative treatment: When the KM is used to perturbatively compute the scattering rate of conduction electrons off the local spin in an expansion in powers of J, this rate is found to increase logarithmically with decreasing temperature, $\gamma(T) \sim J + \nu J^2 \log(D/T)$.

Effective coupling: The perturbation expansion can be expressed through an effective, T-dependent dimensionless coupling, $g(T)$, having the form

$$g(T) = \frac{1}{1/g_0 - \ln(D/T)}, \qquad g_0 = \nu J\,. \tag{8}$$

At large temperatures, $T \simeq D$, the effective coupling reduces to its bare value, g_0. However, it increases with decreasing T due to the logarithm in the denominator.

Kondo temperature: When the temperature becomes sufficiently small, the perturbative correction is no longer small and the perturbative treatment breaks down (becomes invalid). The characteristic crossover scale at which this happens is called the Kondo temperature, $T_{\rm K}$. One way of defining it is as the temperature at which the effective coupling diverges, $g(T_{\rm K}) = \infty$, i.e., where the denominator in Eq. (8) vanishes, $1/g_0 = \ln(D/T_{\rm K})$. This yields

$$T_{\rm K} \sim D{\rm e}^{-1/(J\nu)} \,. \tag{9}$$

The Kondo temperature is exponentially small in the exchange coupling, reflecting the fact that $g(T)$ depends logarithmically on T.

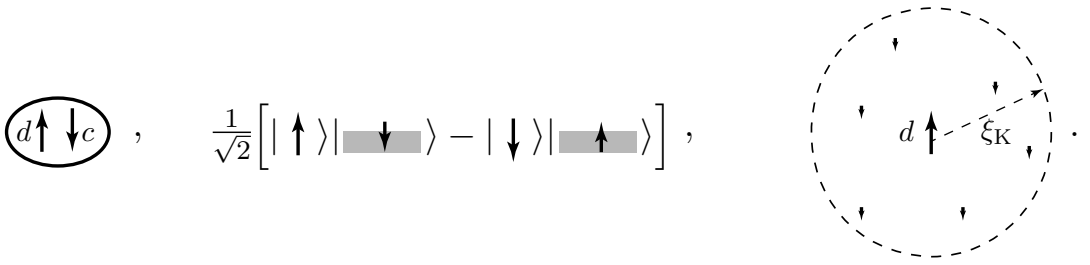
Universality: For $T > T_{\rm K}$, the effective coupling can be expressed in the form

$$g(T) = \frac{1}{\ln(D/T_{\rm K}) - \ln(D/T)} = \frac{1}{\ln(T/T_{\rm K})} \,. \tag{10}$$

($g(T)$ is not defined for $T < T_{\rm K}$, since its perturbative derivation becomes invalid.) This beautiful expression shows that the physics depends on the bare parameters of the model, J, ν and D, only in the combination $T_{\rm K}$. When physical quantities, such as the resistivity or spin susceptibility are measured for different materials, or computed using different bare parameters, they will collapse onto a universal scaling curve when plotted as functions of $T/T_{\rm K}$.

Spin singlet ground state: The regime $T < T_{\rm K}$ is beyond the reach of perturbative methods. However, the nature of the ground state can be anticipated from the fact that the effective coupling becomes large with decreasing temperature: the ground state will be such that the exchange interaction in the Hamiltonian, $\hat{\mathbf{S}}_d \cdot \hat{\mathbf{s}}_c$, is minimized. Consequently, the ground state of the KM is a spin singlet, i.e., has total spin $S = 0$.

Kondo cloud: The singlet ground state can be visualized as follows: A cloud of conduction electrons with a net spin of $1/2$, the so-called Kondo cloud, binds the local spin $1/2$ into a singlet. Commonly used schematic depictions of this singlet include the following:



The singlet may be viewed as a bound state with binding energy $T_{\rm K}$, albeit not a very tightly bound one: since $T_{\rm K}$ is an exponentially small energy scale, the spatial extent of the Kondo cloud scales is exponentially large, scaling as $\xi_{\rm K} \sim v_{\rm F}/T_{\rm K}$, where $v_{\rm F}$ is the Fermi velocity.

Spin screening: The effect of the Kondo cloud on the local spin is called spin screening: by binding the local spin into a singlet, the cloud hides the presence of a local dynamical degree of freedom from the rest of the system. This has striking consequences for the temperature dependence of various physical quantities: as the temperature is decreased from above to below

$T_{\rm K}$, their behavior changes from reflecting the presence of a free spin to reflecting its absence (since it is screened into a singlet). Correspondingly, $T \gg T_{\rm K}$ is called the *local moment (LM) regime*, and $T \ll T_{\rm K}$ the *strong-coupling (SC) regime*.

We next discuss the leading T dependence of three physical quantities in these regimes: the impurity entropy $S(T)$, the impurity spin susceptibility $\chi(T)$, and the spin scattering rate, $\gamma(T)$. Corresponding numerical results will be shown in Sec. 5 when discussing the SIAM, whose low-energy behavior matches that of the KM.

Impurity entropy: The high- and low-temperature limits of the impurity entropy are given by

$$S(T) \simeq \begin{cases} \ln(2)\,, & T \gg T_{\rm K}\,, \\ \ln(1) = 0\,, & T \ll T_{\rm K}\,. \end{cases} \tag{11}$$

These reflect the fact that the free, unscreened impurity has two degenerate states, $|\uparrow\rangle$ and $|\downarrow\rangle$, whereas the ground state is a non-degenerate singlet.

Spin susceptibility: The static spin susceptibility, defined as the linear response of the local spin to an applied magnetic field, has the following limiting behaviors:

$$\chi(T) = \left.\frac{d\langle \mathrm{S}^z_d(h)\rangle_T}{\mathrm{d}h}\right|_{h=0} \simeq \begin{cases} \dfrac{1}{4T}\Big(1 - \mathcal{O}\big(\ln(T/T_{\rm K})\big)\Big)\,, & T \gg T_{\rm K}\,, \\ \chi(0)\left(1 - \mathcal{O}(T^2/T_{\rm K}^2)\right), & T \ll T_{\rm K}\,. \end{cases} \tag{12}$$

For high temperatures, it shows the Curie behavior, $\chi(T) \sim 1/T$, characteristic of a free spin, with logarithmic corrections reflecting the onset of spin screening. As spin screening becomes stronger with decreasing temperature, the Curie behavior is cut off, and for $T \to 0$ the susceptibility approaches a constant, comparable in magnitude to $1/T_{\rm K}$. In fact, this constant can be used to *define* the Kondo temperature, setting $T_{{\rm K}\chi} = 1/\big(4\chi(0)\big)$ [4, 12]. (This definition has the advantage that it does not depend on bare parameters of the model. It differs from that of Eq. (9) by a constant of order unity, but that is no cause for concern, since the Kondo temperature in any case is a *crossover* scale, whose prefactor is a matter of convention.) A rough characterization of the crossover from high to low temperatures is provided by the Curie-Weiss expression $\chi(T) \simeq 1/\big(4(T+T_{{\rm K}\chi})\big)$, but it should be recognized that this does not properly account for the leading high- and low-T corrections indicated in Eq. (12).

Electron scattering: The rate at which conduction electrons scatter off the impurity has the limiting forms

$$\gamma(T) \sim \frac{1}{\ln(T/T_{\rm K})}\,, \qquad T \gg T_{\rm K}\,, \tag{13a}$$

$$\gamma(T) = \gamma(0)\Big(1 - \mathcal{O}(T/T_{\rm K})^2\Big), \qquad T/T_{\rm K} \ll 1\,. \tag{13b}$$

At high temperatures, it is simply proportional to the effective coupling, $g(T)$. Its logarithmic growth is cut off as screening sets in, and it approaches a constant as $T \to 0$.

Fermi liquid behavior: Once the local spin is fully screened into a singlet, the "remaining" conduction electrons scattering off it can no longer flip their spins; instead, they merely acquire

a phase shift. In this sense, the screened singlet acts as a static impurity, without any dynamics. However, it is a strong scatterer: for an electron incident at the Fermi energy, the phase shift acquired during scattering turns out to be $\delta(\epsilon{=}0) = \pi/2$, the maximum value possible for scattering off static impurities.

Nozières has developed a Fermi-liquid description [13] of the strong-coupling regime. It involves no dynamical impurity at all (since the latter is fully screened), only low-energy quasiparticle excitations, i.e., dressed versions of the original, bare conduction electrons of the KM. The quasiparticles experience mutual interactions that are weak and purely local, i.e., act only at the site of the screened impurity. The parameters of the Hamiltonian are fully determined by the energy dependence of the scattering phase shift, $\delta(\varepsilon)$, and can be extracted from zero-temperature properties of the KM (which in turn are known exactly from a Bethe Ansatz solution [5, 6]). A perturbative treatment of these interactions readily yields the leading quadratic temperature dependence, $(T/T_{\rm K})^2$, indicated in Eqs. (12) for $\chi(T)$ and (13a) for $\gamma(T)$. Fermi-liquid theory, combined with results from the exact Bethe Ansatz solution of the Kondo model [5,6], can also be used to analytically derive the prefactors of the quadratic terms [14, 12], yielding

$$\frac{\chi(T)}{\chi(0)} - 1 = -0.821\frac{T^2}{T_{\rm K\chi}^2}, \qquad \frac{\gamma(T)}{\gamma(0)} - 1 = -\frac{\pi^4}{16}\frac{T^2}{T_{\rm K\chi}^2}, \qquad T/T_{\rm K\chi} \ll 1\,. \tag{14}$$

These prefactors are universal constants, characteristic of the strong-coupling regime of the KM. They serve as useful and stringent consistency checks for numerical treatments of the KM.

4 Numerical renormalization group

A description of the crossover from high to low temperatures, and of the strong-coupling regime, requires nonperturbative approaches [3–10]. Here, we focus on the numerical renormalization group (NRG), developed by Ken Wilson in 1975 as a specific implementation of his general RG ideas [3]. Although almost 50 years old, NRG remains the gold standard for solving quantum impurity models, due to its great flexibility, ability to resolve low-energy properties with high accuracy, and the insights it yields about the RG evolution of a system from high to low energies. Here, we discuss NRG for the example of the SIAM; since its low-energy behavior is governed by the Kondo model, the SIAM, too, exhibits a Kondo effect, with an exponentially small Kondo temperature. Hence, NRG results for the SIAM also serve to illustrate the physics for the KM discussed in the previous section. The SIAM was first treated with NRG by Krishna-murthy, Wilkens and Wilson in 1980 [15, 16]. For NRG reviews, see Refs. [4, 17].

Since the Kondo effect involves an exponentially small energy scale, Wilson's goal was to devise a numerical scheme capable of resolving arbitrarily low energy scales. He achieved this goal by (i) discretizing the hybridization spectrum on a logarithmic grid that becomes arbitrarily dense around the Fermi energy; (ii) mapping the resulting discrete model onto a semi-infinite chain, called a *Wilson chain*, containing the impurity at the beginning and exponentially decaying hopping matrix elements along the chain; and (iii) diagonalizing this chain iteratively, while

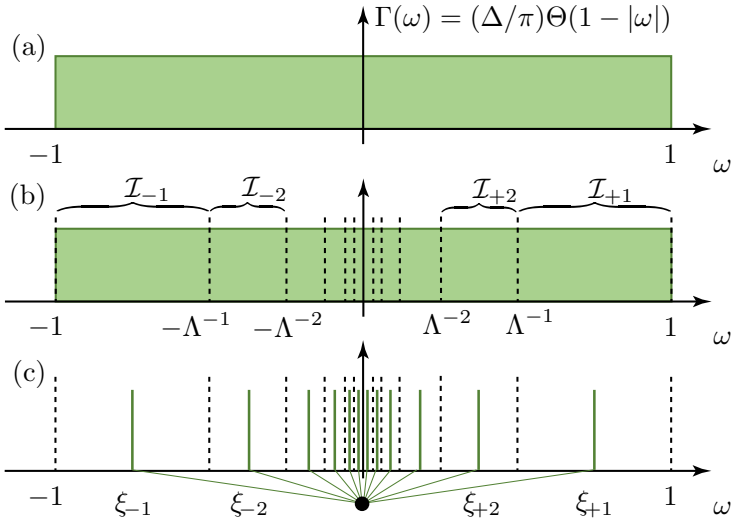


Fig. 1: *(a) Box-shaped hybridization spectrum $\Gamma(\omega)$. (b) Partitioning of the support of $\Gamma(\omega)$ into logarithmically spaced intervals, $\mathcal{I}_{\pm n}$. (c) Hybridization function after logarithmic discretization.*

retaining only low-energy states. This procedure yields a set of approximate eigenstates of the Hamiltonian spanning all energy scales of the model. We now discuss these steps one by one.
(i) *Logarithmic discretization:* The bath spectrum, defined on the interval $\epsilon \in [-1, 1]$, is discretized on a logarithmic grid. Its grid points are chosen at $\pm\Lambda^{-n}$, with $n = 1, 2, 3, \ldots, N$, where $\Lambda > 1$ is a discretization parameter, and N is large enough that Λ^{-N} is smaller than the smallest physical energy scale of interest (e.g. T). (In practice, a typical choice is $\Lambda = 2$, and N is chosen such that Λ^{-N} is below machine precision, $\mathcal{O}(10^{-16})$.) The grid defines a set of intervals $\mathcal{I}_{\pm n}$, defined as $\mathcal{I}_{+n} = [\Lambda^{-n}, \Lambda^{-n+1}]$ and $\mathcal{I}_{-n} = [-\Lambda^{-n+1}, \Lambda^{-n}]$ (see Fig. 1(b)). Next, one represents each interval $\mathcal{I}_{\pm n}$ by a single state, with energy $\xi_{\pm n}$ and hopping amplitude (to the impurity) $\gamma_{\pm n}$ (see Fig. 1(c)), and replaces the bath Hamiltonian and hybridization terms by

$$H_{\text{bath}} \to H_{\text{bath}}^{\text{star}} = \sum_{\pm n,s} \xi_{\pm n} a_{\pm ns}^\dagger a_{\pm ns} \,, \tag{15a}$$

$$H_{\text{hyb}} \to H_{\text{hyb}}^{\text{star}} = \sum_{\pm n,s} \gamma_{\pm n} (a_{\pm ns}^\dagger d_s + \text{h.c.}) \,, \tag{15b}$$

where $a_{\pm ns}$ annihilates the representative bath mode $\pm n$ with spin s. This is known as the *star geometry*, because the depiction of the impurity-bath couplings in Fig. 1 is reminiscent of rays emanating from a star. If the discrete bath energies and couplings are chosen as [18],

$$\gamma_{\pm n}^2 = \int_{\mathcal{I}_{\pm n}} \mathrm{d}\omega \, \Gamma(\omega) \,, \qquad \xi_{\pm n} = \frac{\int_{\mathcal{I}_{\pm n}} \mathrm{d}\omega \, \Gamma(\omega)}{\int_{\mathcal{I}_{\pm n}} \mathrm{d}\omega \, \Gamma(\omega)/\omega} \,, \tag{16}$$

then the resulting hybridization spectrum, $\Gamma^{\text{star}}(\epsilon) = \sum_{\pm n} (\gamma_\pm)^2 \, \delta(\epsilon - \xi_{\pm n})$, serves as a representation of the original $\Gamma(\epsilon)$ that becomes increasingly good the lower the energy [4]. (More sophisticated choices are available to improve the approximation, see Refs. [19–21].)

(ii) *Wilson chain:* Next, the star geometry is mapped to a chain. To this end, note that in $H_{\rm hyb}^{\rm star}$, the impurity operator d_s couples to discretized bath modes only in the combination $t_d f_{0s}^\dagger = \sum_{\pm n} \gamma_{\pm n} a_{\pm ns}^\dagger$. ($t_d$ is chosen such that f_{0s} obeys $\{f_{0s}, f_{0s}^\dagger\} = 1$.) It is therefore advisable to perform a unitary transformation from the $2N$ discrete modes $\{a_{\pm n,s}\}$ to a new set of $\mathscr{L}{+}1{=}2N$ orthonormal modes containing f_{0s}, say $\{f_{\ell s}\}$, with $\ell = 0, 1, 2, \ldots, \mathscr{L}$. This can be achieved by tridiagonalizing the single-particle Hamiltonian matrix for $H_{\rm bath}^{\rm star} + H_{\rm hyb}^{\rm star}$, say $h^{\rm star}$, which has dimensions $(2N{+}1){\times}(2N{+}1)$, using Lanczos tridiagonalization. For example, for $N = 2$ we obtain

$$h^{\rm star} = \begin{pmatrix} \xi_{+2} & 0 & \gamma_{+2} & 0 & 0 \\ 0 & \xi_{+1} & \gamma_{+1} & 0 & 0 \\ \gamma_{+2} & \gamma_{+1} & 0 & \gamma_{-1} & \gamma_{-2} \\ 0 & 0 & \gamma_{-1} & \xi_{-1} & 0 \\ 0 & 0 & \gamma_{-2} & 0 & \xi_{-2} \end{pmatrix} \rightarrow h^{\rm chain} = \begin{pmatrix} 0 & t_d & 0 & 0 & 0 \\ t_d & \epsilon_0 & t_0 & 0 & 0 \\ 0 & t_0 & \epsilon_1 & t_1 & 0 \\ 0 & 0 & t_1 & \epsilon_2 & t_2 \\ 0 & 0 & 0 & t_2 & \epsilon_3 \end{pmatrix}, \tag{17}$$

where the central row and column of $h^{\rm star}$ represent the mode d_s, and the Lanczos scheme is initialized with the vector $(0, 0, 1, 0, 0)^T$. Such a tridiagonalization converts $H_{\rm bath}^{\rm star}$ into a form representing a tight-binding chain of $\mathscr{L}{+}1$ sites,

$$H_{\rm bath}^{\rm chain} = \sum_{\ell=0}^{\mathscr{L}-1} \sum_s t_\ell \big(f_{\ell s}^\dagger f_{\ell+1,s} + \text{h.c.} \big) + \sum_{\ell=0}^{\mathscr{L}} \sum_s \epsilon_\ell \, f_{\ell s}^\dagger f_{\ell s} \,, \tag{18a}$$

with the impurity d level (at site $\ell = -1$) coupled to site $\ell = 0$ with amplitude t_d:

$$H_{\rm hyb}^{\rm chain} = \sum_s t_d \big(d_s^\dagger f_{0s} + \text{h.c.} \big). \tag{18b}$$

The corresponding *Wilson chain* representation of the SIAM for a chain with largest site index $\mathscr{L}$ is given by $H^{\mathscr{L}} = H_{\rm loc} + H_{\rm bath}^{\rm chain} + H_{\rm hyb}^{\rm chain}$, with $H_{\rm loc}$ given by Eq. (1b).

An important consequence of logarithmic discretization is that the hopping matrix elements t_ℓ and on-site energies ϵ_ℓ decay exponentially along the chain: the integration intervals $\mathcal{I}_{\pm n}$ in Eqs. (16) have width $\Lambda^{-n}(\Lambda{-}1)$, thus $\gamma_{\pm n} \sim \Lambda^{-n/2}$ and $\xi_{\pm n} \sim \Lambda^{-n}$, which leads to $t_\ell \sim \Lambda^{-\ell/2}$ and $\epsilon_\ell \sim \Lambda^{-\ell}$. The characteristic energy scale of site ℓ is given by the hopping amplitude onto that site, $t_{\ell-1} \sim \Lambda^{-(\ell-1)/2}$. Thus, Wilson chains exhibit *energy scale separation*: different parts of the chain represent different energy scales, with the characteristic energy of site ℓ decreasing exponentially with increasing ℓ.

(iii) *Iterative diagonalization:* The energy scale separation along a Wilson chain can be exploited to numerically diagonalize $H^{\mathscr{L}}$ in iterative fashion, adding one site at a time, starting from site $\ell = 0$. In doing so, the resolution of the low-energy part of the eigenspectrum will increase exponentially. However, the number of eigenstates will increase exponentially, too. Therefore, a systematic truncation scheme is needed to limit the number of eigenstates that are computed explicitly.

Wilson proposed an energy-based truncation scheme: Suppose that H^ℓ has been diagonalized, yielding a set of eigenstates satisfying $H^\ell |\alpha\rangle_\ell = E_\alpha^\ell |\alpha\rangle_\ell$. This set of eigenstates is called *shell* ℓ.

These are divided into two groups: the N_{keep} lowest-lying states are *kept*, the remaining ones *discarded*. (Typical choices for N_{keep} range between 500 and 5000.) Then a new site, $\ell{+}1$, is added, and $H^{\ell+1}$ is diagonalized in the direct product space of the new site and the kept states of shell ℓ, whereas the contribution from its discarded states is neglected. Their neglect is justified due to energy scale separation: the energies of the discarded, high-lying states from shell ℓ are significantly larger than the characteristic energy of that shell, $E_\ell^{\text{high}} \gg \Lambda^{-(\ell-1)/2}$, which in turn is larger than the coupling to the new site, $t_\ell = \Lambda^{-\ell/2}$. The contribution of discarded states on the eigenstates of the new shell $\mathscr{L}{+}1$, estimated by a second-order perturbation argument, is of order $t_\ell^2/E_\ell^{\text{high}}$. This is $\ll (\Lambda^{-\ell/2})^2/\Lambda^{-(\ell-1)/2} = \Lambda^{-(\ell+1)/2}$, and hence much smaller than the characteristic energy of the new shell, $\Lambda^{-\ell/2}$. Therefore, the effect of discarded states can be safely neglected.

For a chain of total length ℓ, Wilson's iterative diagonalization scheme yields a set of N_{keep} eigenstates for each shell $\ell \leq \mathscr{L}$. In combination, they resolve the spectrum of the Hamiltonian on all energy scales from 1 to $\Lambda^{-(\mathscr{L}-1)/2}$. Conceptually, the eigenenergies of shell ℓ represent the low-energy part of the finite-size spectrum of the impurity plus bath put in a spherical box of radius $R_\ell \sim \Lambda^{(\ell-1)/2}$, centered on the impurity. As ℓ increases, the box radius increases exponentially, and the finite-size level spacing, $1/R_\ell \sim \Lambda^{-(\ell-1)/2}$, decreases exponentially. From an RG perspective, increasing R_ℓ amounts to decreasing the infrared cutoff, here given by the finite-size level spacing, and thus probing the system at lower energy scales. A detailed discussion of this perspective may be found in Refs. [3, 22].

To analyze how the structure of the spectrum changes with ℓ, Wilson defined a set of rescaled excitation energies for each shell,

$$\mathcal{E}_\alpha^\ell = \Lambda^{(\ell-1)/2}\big(E_\alpha^\ell - E_{\text{GS}}^\ell\big). \tag{19}$$

These are measured with respect to the ground state (GS) energy of that shell, and scaled such that the spacing of the lowest excitations is $\mathcal{O}(1)$. A plot of the rescaled energies $\mathcal{E}_\alpha^\ell$ versus shell number ℓ on a linear scale, or equivalently, vs. the energy scale $\Lambda^{-\ell/2}$ on a logarithmic scale, with ℓ increasing in steps of 2, is called an *NRG energy level flow diagram*. It reveals how the rescaled eigenspectra evolve ("flow") with increasing system size. (ℓ is increased in steps of 2 because of an even-odd effect inherited from $H_{\text{bath}}^{\text{chain}}$: its single-particle eigenspectra for $\mathscr{L}$ even or odd are structurally different, causing the same to be true for the many-body eigenspectra of $H^{\mathscr{L}}$.) An example of an NRG energy level flow diagram is shown in Fig. 2 below.

Once a Wilson chain has been iteratively diagonalized, the approximate eigenstates so obtained can be used to compute physical quantities. These include thermodynamic quantities, such as the impurity entropy $S(T)$ and spin susceptibility $\chi(T)$, but also dynamical, frequency-dependent quantities, such as the dynamic spin susceptibility $\chi(\omega)$, or the impurity spectral function, $\mathcal{A}(\omega)$ and the retarded self-energy, $\Sigma^R(\omega)$. For details, see Refs. [23, 24, 17].

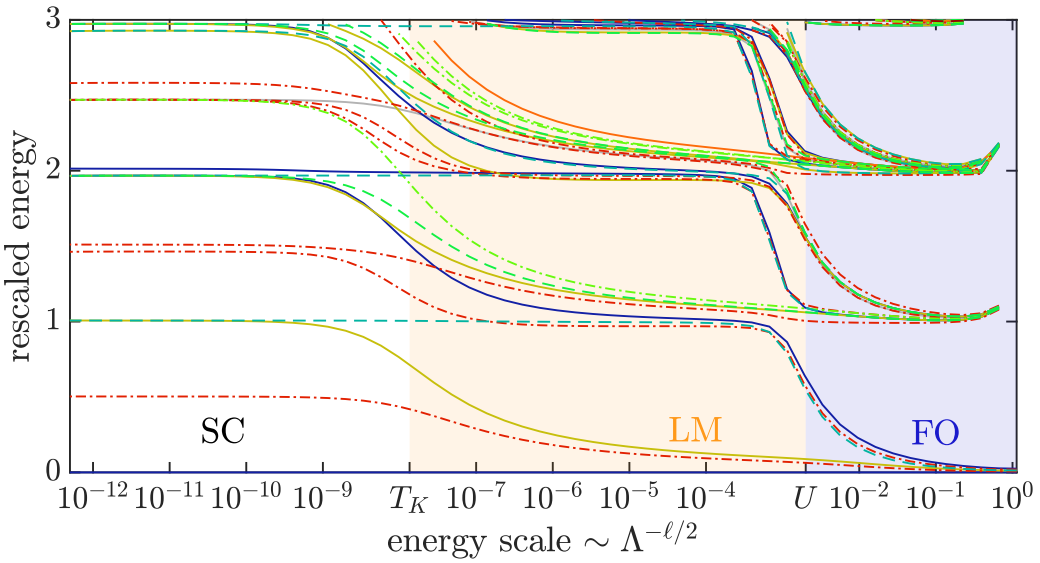


Fig. 2: *SIAM: NRG energy level flow diagrams (for ℓ = even), showing three fixed-point regimes, associated with the free orbital (FO), local moment (LM) and strong-coupling (SC) fixed points.*

5 NRG results for the single-impurity Anderson model

In the following, we illustrate the RG flow of the SIAM with NRG results. Before we start, let us specify our definition of the Kondo temperature for the SIAM. Following Haldane [25, 26] and Refs. [14, 12], we use

$$T_{\rm K} = \frac{1}{4\chi(0)} \simeq \sqrt{U\Delta/2}\, {\rm e}^{-\pi U/8\Delta + \pi\Delta/2U}\, {\rm e}^{x^2}\,, \qquad x = \big(\varepsilon_d + U/2\big)\frac{\pi}{\sqrt{2U\Delta}}\,. \tag{20}$$

The first equality for $T_{\rm K}$ corresponds to $T_{{\rm K}\chi}$ from our discussion of the spin susceptibility of the Kondo model; the second follows from analytical results for $\chi(0)$ from the exact Bethe-Ansatz solution of the Anderson model [5]. The parameter x measures the distance to the particle-hole symmetric point, where $\varepsilon_d = -U/2$.

Or numerical results were computed using the following parameters (all energies are given in units of D): $U = 2 \cdot 10^{-3}$, $\varepsilon_d = -U/2 = -10^{-3}$, $\Delta = 0.04\,U$, and $h = 0$ (unless stated otherwise). We are thus in the wide-band limit, $\Delta, U \ll D = 1$. The corresponding Kondo temperature from Eq. (20) is $T_{\rm K} = 1.3533 \cdot 10^{-8}$. The numerical NRG parameters were chosen as $\Lambda = 1.7$, $N_{\rm keep} = 5000$.

Figure 2 shows an energy level flow diagram for the SIAM. We observe that there are three energy ranges during which the flow of the finite-size spectrum is almost stationary (ℓ independent). Within each of these, $\Lambda^{-\ell/2}$ lies within an energy regime governed by one of the RG *fixed points* of the model. Here, the nomenclature "fixed point" really is appropriate: the spectrum literally remains fixed while changing ℓ (conceptually, increasing the system size R_ℓ). The energy ranges in between, where the levels shift and/or cross each other, are associated with crossovers between fixed points.

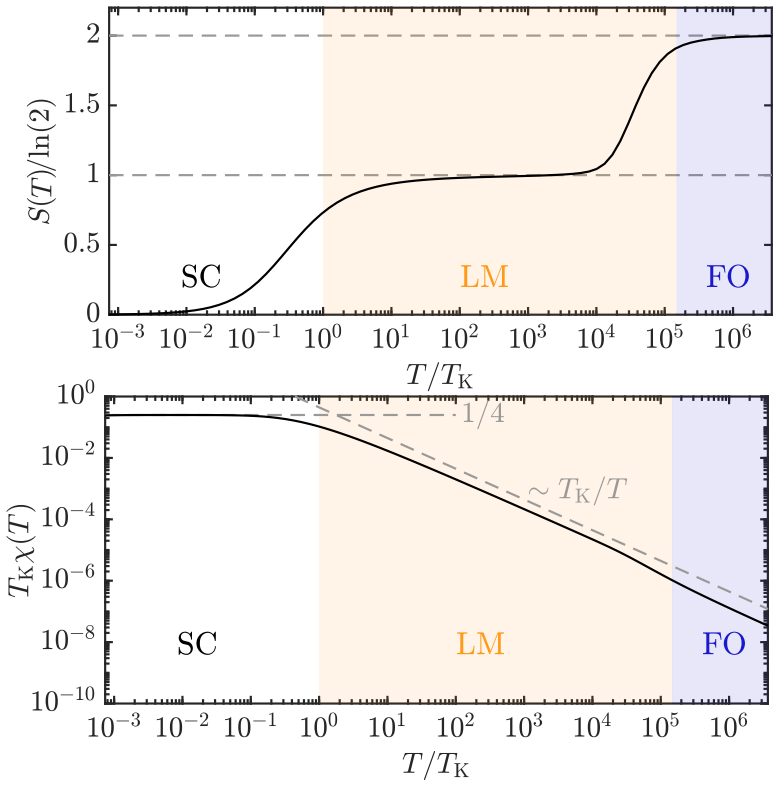


Fig. 3: *SIAM: Top: Temperature dependence of the impurity entropy, $S(T)$. The plateaus at $\ln(4) = 2\ln(2)$, $\ln(2)$ and $\ln(1) = 0$ in the FO, LM and SC regimes, respectively, reflect the number of states that are energetically available to the impurity in these regimes: four ($|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, $|\uparrow\downarrow\rangle$) in the FO regime, two ($|\downarrow\rangle$, $|\uparrow\downarrow\rangle$) in the LM regime, and one (screened singlet) in the SC regime. Bottom: Temperature dependence of the static spin susceptibility, $\chi(T)$. In the LM regime it shows Curie behavior, $\chi(T) \sim 1/T$. In the SC regime it saturates at a value, $\chi(0)$, which we use as definition for the Kondo temperature, setting $T_K = 1/(4\chi(0))$.*

For the SIAM, the RG flow has three fixed point regimes (first analyzed with NRG in Refs. [15, 16]), clearly visible in Fig. 2. (I) The *free orbital* (FO) regime involves excitation energies $\gtrsim U$. It is governed by "free" (not hindered by U) charge fluctuations, arising through real transitions between all four states of the d level, $|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, and $|\uparrow\downarrow\rangle$. (II) The *local moment regime* (LM) involves excitation energies in the range between the Kondo temperature, T_K and U. Here, charge fluctuations can occur only virtually, giving rise to spin-flip transitions between $|\uparrow\rangle$ and $|\downarrow\rangle$. Real transitions to $|0\rangle$ or $|\uparrow\downarrow\rangle$ are frozen out, since the available excitation energies are too small to overcome the associated cost in Coulomb energy, U. (III) The *strong-coupling* (SC) regime involves excitation energies well below T_K. It features a screened spin-singlet, off which Fermi-liquid quasiparticles scatter without spin-flip scattering (cf. the discussion on p. 7). Spin-flip scattering is frozen out, since that would require breaking up the spin singlet, which has binding energy $\sim T_K$.

This interpretation of the fixed point spectra is beautifully confirmed by the temperature dependence of two *thermodynamic quantities*, the impurity entropy, $S(T)$, and the static local spin susceptibility, $S(T)$, shown in Fig. 3. As explained in the caption of these plots, both quantities show distinctly different behaviors in the three fixed point regimes, characteristic of a free orbital, a local moment or a screened singlet, respectively.

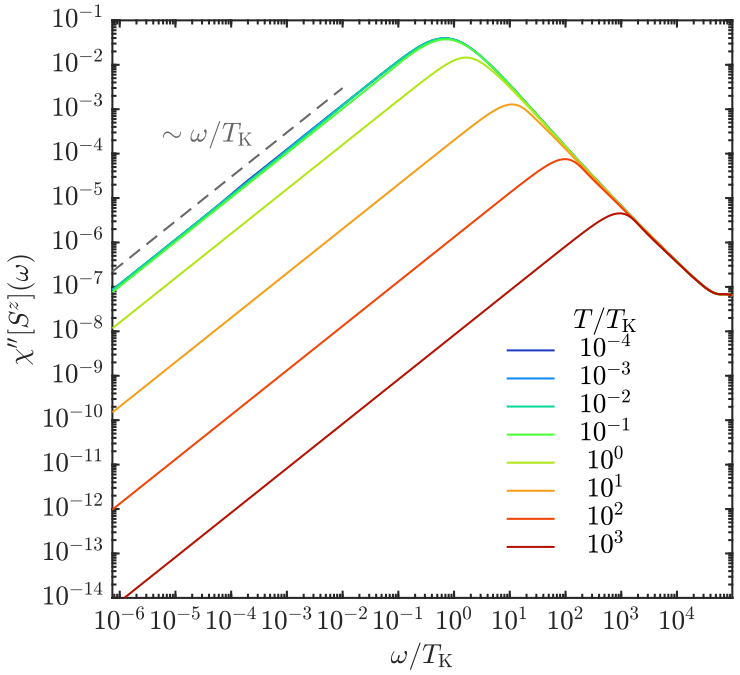


Fig. 4: *SIAM: Frequency dependence of the imaginary part of the dynamical spin susceptibility, $\chi''(\omega)$, plotted on a log-log scale, for a range of different temperatures.*

Next, we turn to *dynamical quantities*. First, we consider the dynamical spin susceptibility, defined as

$$\chi(\omega) = \chi'(\omega) + \mathrm{i}\chi''(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}(-\mathrm{i})\,\Theta(t)\,\big\langle\big[\mathrm{S}^z(t), \mathrm{S}^z(0)\big]\big\rangle_T\,. \tag{21}$$

The static spin susceptibility discussed earlier, $\chi(T)$, is equal to the zero-frequency limit of the real part here, $\chi'(\omega = 0)$. The frequency dependence of the imaginary part, $\chi''(\omega)$, is shown in Fig. 4. For $T = 0$ and $\omega \gg T_\mathrm{K}$, it increases with decreasing ω, characteristic of an almost free spin in the LM regime. It reaches a maximum at $\omega \simeq T_\mathrm{K}$, where spin screening becomes strong, and for $\omega \to 0$ decreases linearly with ω. This linear dependence is characteristic of the Fermi liquid behavior of the system in the SC regime (energies $\ll T_\mathrm{K}$), and indicative of a well-screened impurity spin. For temperatures $T \gg T_\mathrm{K}$, the maximum in $\chi''(\omega)$ occurs at T, and the peak value of χ'' decreases with increasing T. This illustrates how a finite temperature cuts off the RG flow towards strong coupling.

Another dynamical quantity of great interest is the retarded correlator for the d level,

$$G_s(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t}(-\mathrm{i})\,\Theta(t)\,\big\langle\{d_s(t), d_s(0)\}\big\rangle_T = \int \mathrm{d}\omega'\, \frac{\mathcal{A}_s(\omega')}{\omega - \omega' + \mathrm{i}0^+}\,. \tag{22}$$

Specifically, we are interested in its spectral function, $\mathcal{A}_s(\omega) = -\frac{1}{\pi}\mathrm{Im}G_s(\omega)$, which characterizes the local density of states associated with the d level, and in the corresponding retarded self-energy, $\Sigma_s(\omega)$, defined via

$$G_s(\omega) = \frac{1}{\omega - \varepsilon_d - \Delta(\omega) - \Sigma_s(\omega)}\,, \tag{23}$$

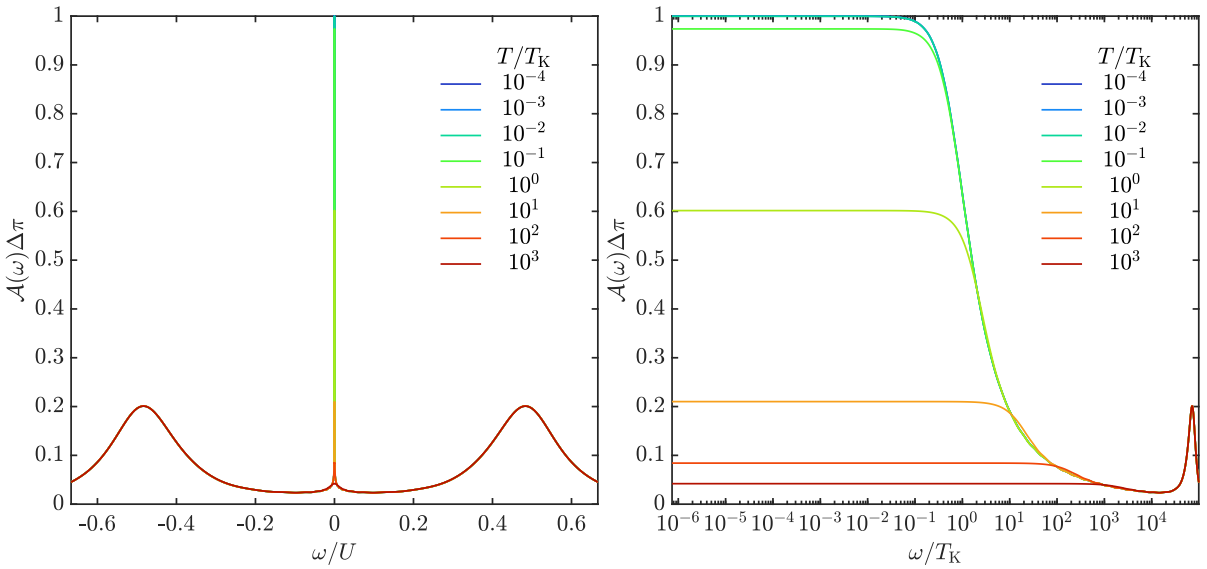


Fig. 5: *SIAM: Local spectral function, $A(\omega)$, for eight different temperatures, shown on a linear scale (left) and a logarithmic scale (right).*

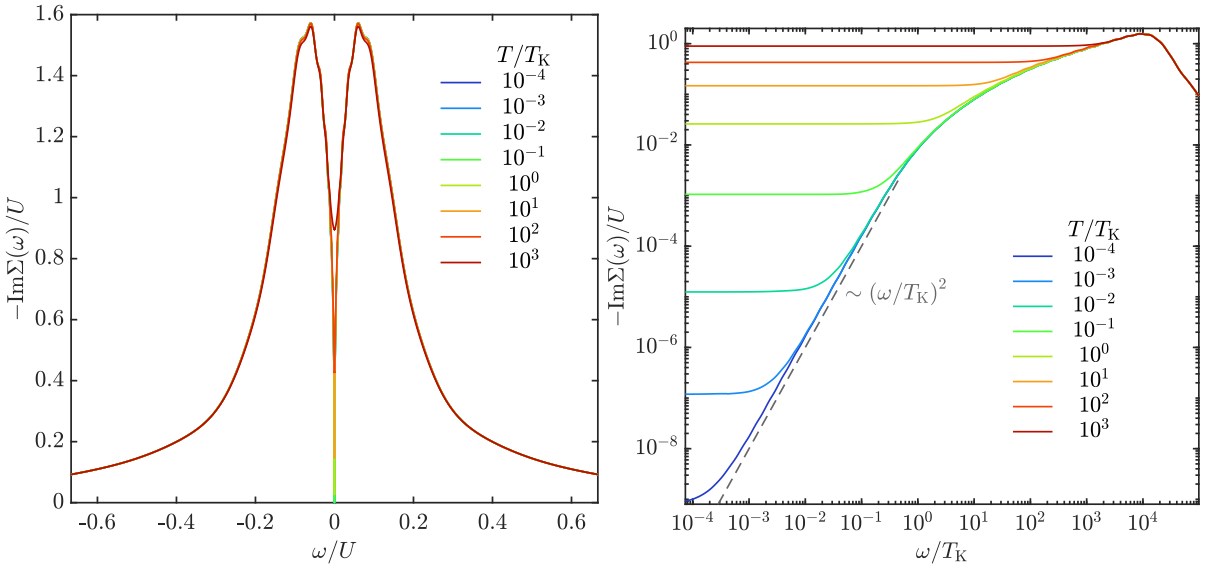


Fig. 6: *Local self-energy, $\Sigma(\omega)$, for eight different temperatures, shown on a linear scale (left) and a logarithmic scale (right).*

where $\Delta(\omega)$ is the *hybridization function* from Eq. (2). Note that $\Delta(0) = -\mathrm{i}\Delta$, which is why Δ sets the hybridization-induced width of the d level. (For the wide-band limit (Δ, $U \ll D$) considered here, $\Delta(\omega) \simeq -\mathrm{i}\Delta$ actually holds to very good approximation throughout the regime $|\omega| \leq U/2$.) Since we will mostly consider the case $h = 0$ (no magnetic field), we will henceforth drop the spin index s.

Figure 5 shows the local spectral function, $\mathcal{A}(\omega)$, for a number of different temperatures, ranging from $T \ll T_\mathrm{K}$ to $T \gg T_\mathrm{K}$. It features two broad peaks at $\omega = \pm U/2$, the so-called Hubbard side-bands. These arising from real transitions between the d level and the band and are essentially independent of T. Additionally, once T drops below T_K, $\mathcal{A}(\omega)$ develops a sharp peak at $\omega = 0$, known as the *Kondo resonance*. At zero temperature, its width is essentially given

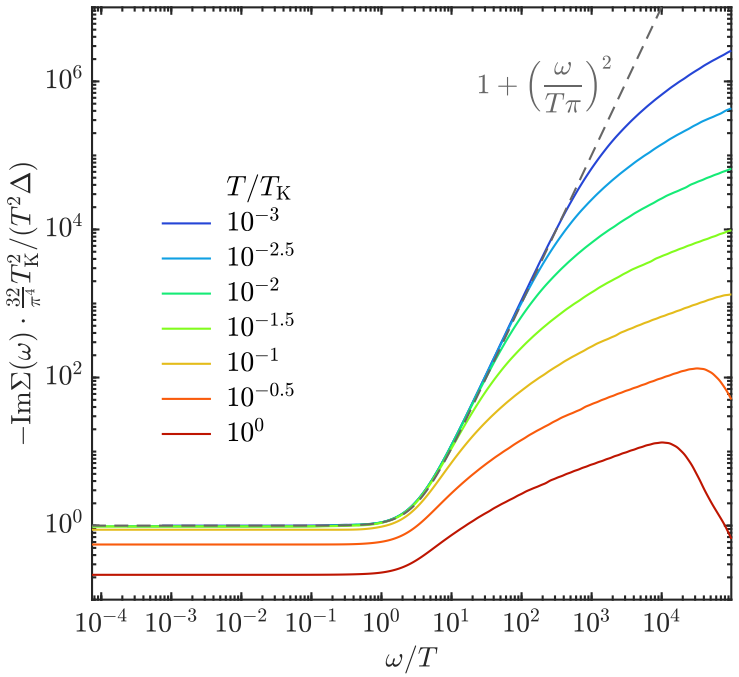


Fig. 7: *SIAM: Fermi-liquid scaling behavior of the local self-energy, $\Sigma(\omega,T)$. For $\omega,T \ll T_{\rm K}$, the rescaled data collapses onto a universal scaling curve in accordance with Eq.* (24).

by $T_{\rm K}$ and its height is $\mathcal{A}(0) = 1/(\pi\Delta)$. The latter relation follows from Eq. (23) by using $\mathrm{Im}\Sigma(0)\big|_{T=0} = 0$, a well-established analytical result that follows from Nozières' Fermi-liquid treatment of the SC regime [13, 14, 12]. Figure 6, which shows the imaginary part of the self-energy, $-\mathrm{Im}\Sigma(\omega)$, confirms this fact: At $T=0$, the self-energy indeed drops to zero in the energy range $|\omega| \ll T_{\rm K}$.

The occurrence of the Kondo resonance is perhaps the most striking manifestation of spin screening. Intuitively, it arises because at very low energies, $\omega, T \ll T_{\rm K}$, Fermi-liquid quasi-particles scatter purely elastically off the screened singlet: they acquire a phase shift, but their energy and spin does not change.

When ω and/or T increase from 0, inelastic scattering sets in and $-\mathrm{Im}\Sigma$ becomes nonzero. As long as $\omega, T \ll T_{\rm K}$, this can still be described using Fermi-liquid theory. By combining known relations from Refs. [27, 12] one finds

$$-\mathrm{Im}\Sigma(\omega,T) = \frac{\pi^2\Delta\big(\omega^2+\pi^2T^2\big)}{32T_{\rm K}^2}\,, \qquad \omega,T \ll T_{\rm K}\,. \tag{24}$$

This implies that a plot of $-\mathrm{Im}\Sigma(\omega,T)/T^2$ vs. ω/T should yield a scaling collapse. Indeed, when the self-energy is computed using state-of-the-art NRG methodology [28,29], this scaling behavior is beautifully recovered, as shown in Fig. 7.

The application of a local magnetic field h favors the state $|\uparrow\rangle$ over $|\downarrow\rangle$. It thus tends to polarize the local spin, thereby disrupting spin screening. As a result, the RG flow to strong coupling is cut off at the energy scale h. This has dramatic consequences for the local spectral function: The Kondo resonances for $\mathcal{A}_\uparrow(\omega)$ and $\mathcal{A}_\downarrow(\omega)$ shift apart and get suppressed with increasing field. This behavior is shown in Fig. 8.

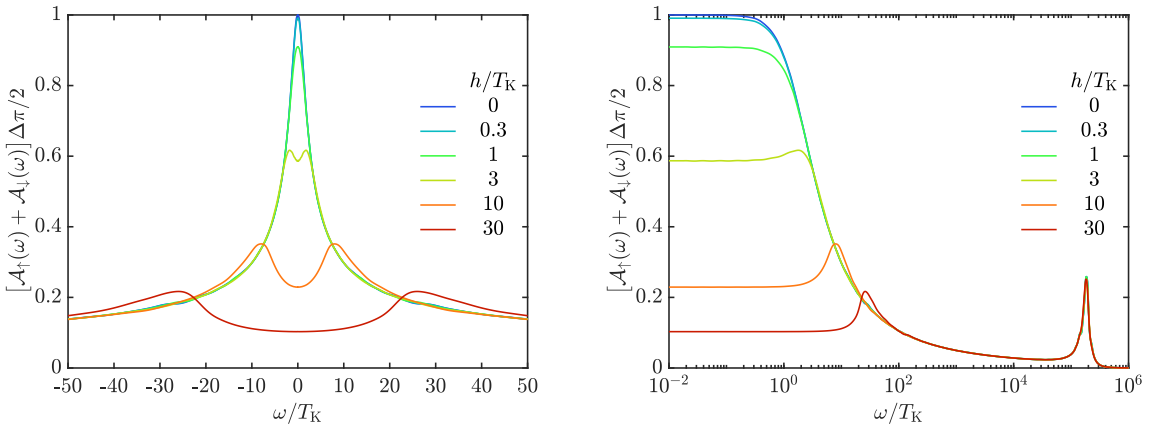


Fig. 8: *SIAM: Dependence of the local spectral function on magnetic field, computed at* $T = 0$.

We conclude our discussion of the SIAM by putting it into the context of dynamical mean-field theory (DMFT) [30]. There, one considers a correlated lattice model but assumes the self-energy to be purely local. This allows the model to be mapped onto a quantum impurity model with a hybridization function that has to be determined self-consistently. For the one-band Hubbard model, the corresponding impurity model is the SIAM. DMFT self-consistency then has drastic consequences: at sufficiently high temperature or interaction strength, the Kondo resonance is completely suppressed [31,32], and the system undergoes a Mott transition from a metallic state to an insulating state.

6 Two-channel Kondo model

We end these notes by making a few remarks about a quantum impurity model that does *not* show spin screening: the two-channel Kondo (2CK) model [33]. Its Hamiltonian is similar to that of the usual (one-channel) Kondo model, but now there are two conduction bands, labeled by a channel index $j = 1, 2$, that couple symmetrically to the local spin-$\frac{1}{2}$:

$$H_{\mathrm{2CK}} = \sum_{ks} \sum_{j=1,2} \varepsilon_k \hat{n}_{kjs} + J \hat{\mathbf{S}}_d \cdot \hat{\mathbf{s}}_c, \qquad \hat{\mathbf{s}}_c = \sum_{ks,k's'} \sum_{j=1,2} c^\dagger_{kjs} \tfrac{1}{2} \boldsymbol{\sigma}_{ss'} c_{k'js'} \,. \tag{25}$$

This model also involves some screening of the local spin by conduction electrons, but full spin-screening is not possible: if the local spin-$\frac{1}{2}$ would form a spin singlet with a Kondo cloud from one channel, that would break channel symmetry. If two Kondo clouds from both channels try to screen the spin-$\frac{1}{2}$ without breaking channel symmetry, they together constitute a spin-1 cloud and *overscreen* the spin-$\frac{1}{2}$, yielding another spin-$\frac{1}{2}$ object. Repeating this heuristic argument, one concludes that a spin singlet cannot be formed as long as channel symmetry is respected. Therefore, the quasiparticles do not behave as a Fermi liquid—instead, the model shows *non-Fermi-liquid* behavior.

This has striking consequences for physical quantities. Here, we only mention two zero-temperature properties that are decidedly non-Fermi-liquid in character. Figure 9 shows corresponding NRG results.

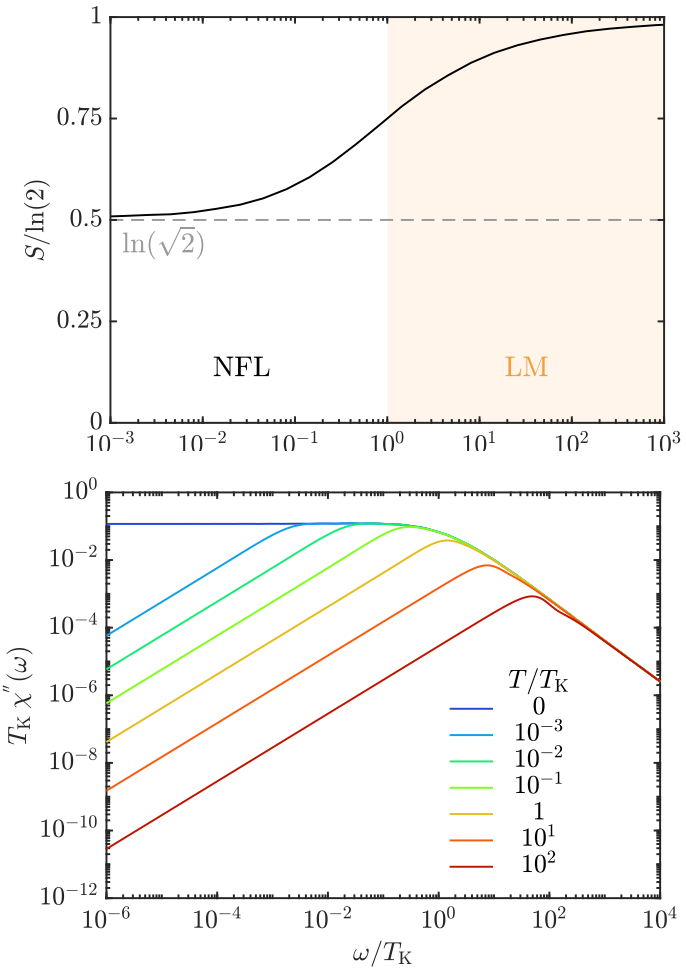


Fig. 9: *Two-channel Kondo model. Top: temperature dependence of the impurity entropy,* $S(T)$*. We defined* T_K *through the condition* $S(T_K) = \big(S(\infty)+S(0)\big)/2$ *such that the crossover is centered on* T_K*. Bottom: Frequency dependence of the imaginary part of the dynamical spin susceptibility,* $\chi''(\omega)$*, plotted on a log-log scale, for a range of different temperatures.*

Impurity entropy: At $T=0$, the impurity entropy $S(0)$ does not equal $\ln(1) = 0$, but $\frac{1}{2}\ln(2) = \ln(\sqrt{2})$. The unusual fractional prefactor reflects the fact that the impurity retains some dynamical character, since it is not fully screened.
Dynamical spin susceptibility: At $T=0$, the imaginary part of the susceptibility approaches a constant as $\omega \to 0$, $\chi''(\omega) \sim \text{const}$ (in contrast to the Fermi liquid behavior, $\chi''(\omega) \sim \omega$, of the one-channel Kondo model). That, in turn, implies that the real part diverges logarithmically, $\chi'(\omega) \sim \ln(T_{\rm K}/\omega)$. This, too, indicates that the impurity is not fully screened but retains some dynamical character, even at $T = 0$.

Acknowledgement

I warmly thank Andreas Gleis and Jeongmin Shim for doing the NRG computations reported in these notes and creating the corresponding figures, and Andreas Gleis for establishing Eq. (24).

A Schrieffer-Wolff transformation

The KM can be obtained from the SIAM using a Schrieffer-Wolff transformation [11] and a projection onto the subspace were $n_d = 1$. Here, we briefly outline the key steps of this derivation. The starting point is the SIAM Hamiltonian, written in the form $H_{\rm SIAM} = H_0 + H_1$, with $H_0 = H_{\rm bath} + H_{\rm loc}$ being $\mathcal{O}(v^0)$ and $H_1 = H_{\rm hyb}$ being $\mathcal{O}(v^1)$, where $v = v_k$ is the hybridization parameter. The goal is to find a unitary transformation such that the new Hamiltonian, $\widetilde{H} = e^A\, H_{\rm SIAM}\, e^{-A}$, contains no terms of order $\mathcal{O}(v^1)$. Here, A is anti-unitary, $A^\dagger = -A$, with an hybridization expansion of the form $A = \mathcal{O}(v^1) + \mathcal{O}(v^2) + \ldots$. Expanding $\widetilde{H}$ in powers of A using the Baker-Campbell-Hausdorff formula, one obtains

$$\widetilde{H} = (H_0{+}H_1) + [A,\, H_0{+}H_1] + \tfrac{1}{2}\big[A,\, [A,\, H_0{+}\cancel{H_1}]\big] + \mathcal{O}(v^3) \tag{26}$$

(The term slashed out is $\mathcal{O}(v^3)$.) To ensure that the right side contains no terms $\mathcal{O}(v^1)$, we must choose A such that $H_1 = -[A, H_0]$, implying

$$\widetilde{H} = H_0 + \tfrac{1}{2}[A, H_1] + \mathcal{O}(v^3)\,. \tag{27}$$

One may readily verify that the requirement $H_1 = -[A, H_0]$ is satisfied by

$$A = \sum_{ks} v\left(\frac{1}{\varepsilon_k{-}\varepsilon_d}\, c^\dagger_{ks} d_s + \frac{U}{(\varepsilon_d - \varepsilon_k)(\varepsilon_d + U - \varepsilon_k)}\, d^\dagger_{-s} d_{-s} c^\dagger_{ks} d_s\right) - \text{h.c.} \tag{28}$$

Inserting this into Eq. (27) and dropping terms that are zero in the subspace where $n_d = 1$, one obtains

$$\widetilde{H}|_{n_d=1} = \sum_{ks} \varepsilon_k \hat{n}_{ks} + \sum_{kk'} \widetilde{v}_{kk'}\, \hat{\mathbf{S}}_d \cdot c^\dagger_{ks} \tfrac{1}{2} \boldsymbol{\sigma}_{ss'} c_{k's'} + \ldots \tag{29}$$

where the coupling matrix elements are given by

$$\widetilde{v}_{kk'} = \frac{-\frac{1}{2} v^2 U}{(\varepsilon_d - \varepsilon_k)(\varepsilon_d + U - \varepsilon_k)} + (k \leftrightarrow k')\,. \tag{30}$$

In Eq. (29), the ellipsis $\ldots$ stands for terms of the form $\sum_{kk'} c^\dagger_{ks} c_{k's}$, describing potential scattering, i.e., scattering without any spin dependence. These are usually ignored when discussing the Kondo effect, because in contrast to the spin-exchange interaction of the second term in Eq. (29), they do *not* give rise to a logarithmic temperature dependence when treated perturbatively. In renormalization group (RG) terminology, spin-exchange is a relevant perturbation, whereas potential scattering is not.

Now consider low-energy excitations close to the Fermi energy, then $\widetilde{v}_{kk'}$ simplifies to

$$\widetilde{v}_{kk'} \simeq -\frac{v^2 U}{\varepsilon_d\,(\varepsilon_d + U)} = J, \qquad \forall |\varepsilon_k|,\, |\varepsilon_{k'}| \ll |\varepsilon_d|,\, |\varepsilon_d{+}U|\,. \tag{31}$$

In this limit, $\widetilde{H}$ of Eq. (29) thus reduces to the Kondo Hamiltonian $H_{\rm KM}$ of Eq. (7), with an exchange coupling constant J given by Eq. (31). Note that this equals the expression for J given in Eq. (6) above, found by considering 2nd order spin-flip processes.

To summarize the above discussion: the low-energy behavior of the SIAM, at energy scales for which charge fluctuations involving $|0\rangle$ and $|\!\uparrow\downarrow\rangle$ are "frozen out" (can occur only virtually), is described by the KM.

References

[1] W.J. de Haas, J. de Boer, and G.J. van den Berg, Physica **1**, 1115 (1934)

[2] J. Kondo, Prog. Theor. Phys. **32**, 37 (1964)

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

[4] R. Bulla, T.A. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

[5] A. Tsvelick and P. Wiegmann, Adv. Phys. **32**, 453 (1983)

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

[7] I. Affleck and A. Ludwig, Nucl. Phys. B **352**, 849 (1991)

[8] V.J. Emery and S. Kivelson, Phys. Rev. B **46**, 10812 (1992)

[9] G. Zaránd and J. von Delft, Phys. Rev. B **61**, 6918 (2000)

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

[11] J.R. Schrieffer and P.A. Wolff, Phys. Rev. **149**, 491 (1966)

[12] M. Filippone, P. Moca, A. Weichselbaum, J. von Delft, and C. Mora, Phys. Rev. B **98**, 075404 (2018)

[13] P. Nozières, J. Low Temp. Phys. **17**, 31 (1974)

[14] C. Mora, C.P. Moca, J. von Delft, and G. Zaránd, Phys. Rev. B **92**, 075120 (2015)

[15] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1003 (1980)

[16] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**, 1044 (1980)

[17] A. Weichselbaum, Phys. Rev. B **86**, 245124 (2012)

[18] J. Vivaldo L. Campo and L.N. Oliveira, Phys. Rev. B **72**, 104432 (2005)

[19] R. Žitko, Comp. Phys. Commun. **180**, 1271 (2009)

[20] R. Žitko and T. Pruschke, Phys. Rev. B **79**, 085106 (2009)

[21] B. Bruognolo, N.O. Linden, F. Schwarz, S.S.B. Lee, K. Stadler, A. Weichselbaum, M. Vojta, F.B. Anders, and J. von Delft, Phys. Rev. B **95**, 121115 (2017)

[22] J. von Delft, G. Zaránd, and M. Fabrizio, Phys. Rev. Lett. **81**, 196 (1998)

[23] R. Peters, T. Pruschke, and F.B. Anders, Phys. Rev. B **74**, 245114 (2006)

[24] A. Weichselbaum and J. von Delft, Phys. Rev. Lett. **99**, 076402 (2007)

[25] F.D.M. Haldane, Phys. Rev. Lett. **40**, 416 (1978)

[26] F.D.M. Haldane, J. Phys. C: Solid State Phys. **11**, 5015 (1978)

[27] A.C. Hewson, J. Phys.: Condens. Matter **13**, 10011 (2001)

[28] F.B. Kugler, Phys. Rev. B **105**, 245132 (2022)

[29] A. Gleis (unpublished) (2022)

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

[31] R. Bulla, Phys. Rev. Lett. **83**, 136 (1999)

[32] R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64**, 045103 (2001)

[33] P. Nozières and P. Blandin, J. Phys. **41**, 193 (1980)

4 Machine Learning as a Solver for DMFT

Cedric Weber
King's College London
The Strand, WC2R 2LS London

Contents

1 Introduction **2**
1.1 Supervised learning and linear regression 4
1.2 A single layer neural network: the perceptron model 5
1.3 Neural networks 10
1.4 Back-propagation 11

2 Generating a quantum database for the Anderson impurity model **13**
2.1 Polynomial basis method 13
2.2 Generating the validation dataset with exact diagonalization 14
2.3 Data representation 16
2.4 Defining the unknowns: from learning solutions to learning errors 18
2.5 Constructing the database 19

3 Training a model solver to solve the Anderson impurity model **20**
3.1 Data processing: symmetry, augmentation and transformation 22
3.2 Activation function with many-body quantities 23
3.3 An error correction approach for solving DMFT 24
3.4 Database of solutions for the Anderson Impurity Model 24
3.5 A Neural Network Impurity Solver 26
3.6 A data-driven approach to the Mott transition 28

4 Conclusion and code availability **28**

1 Introduction

Widespread interest has been devoted in the last three decades to strongly correlated materials, which are being used in emerging technologies, such as spintronics, quantum computing and high temperature superconductivity for example. They are characterized by strong electronic interactions between their d and f-band valence electrons. The interplay between charge localization (Mott physics) and itinerant behavior (quasi-particles) provides a challenge for standard electronic theories, such as density function theories. In recent years significant progress in understanding the underlying physics of strong electron correlation effects has been made by the dynamical mean-field theory (DMFT) method [1], in particular with the harnessing of DMFT to widespread materials modelling methods, such as density functional theory (DFT), with the DFT+DMFT combined approach [2].

Despite this formidable achievement, practical challenges remain with the DFT+DMFT approach. At the heart of the theory is the choice of the quantum engine that solves the many-body Anderson Impurity Model (AIM), an underlying model that provides the local Green function of the compound of interest via a self-consistent mapping. Currently, methods of reference at finite temperature are based on statistical sampling, with for instance the continuous-time quantum Monte Carlo solver [3]. The latter provide exact results (within sampling error bars) on the imaginary axis for the fully rotationally-invariant AIM Hamiltonian.

Other solvers can provide robust information on the real axis, such as numerical renormalization group approach [4], but their application remains currently challenging for the case multi-orbital systems for the DFT+DMFT approach.

Another well-known method that provides information on the real-axis is the exact diagonalization (ED) approach. Here, the DMFT hybridization is approximated by a finite number (N_b) of effective bath orbitals. In practice, N_b is restricted because of the exponential growth of the Fock space with the total number of sites (bath and impurity). The growth of the Hilbert space is mitigated by Lanczos-based algorithms, which utilize the sparse nature of the Hamiltonian, and to a large degree enable routine calculations for systems up to $N_s \approx 12$ [5]. Overall, the limitation N_s limits the scope of applicability of this method, and typically discretization effects remain large for multi-orbital systems.

In summary, the quantum engine at the heart of DMFT remains an active topic of research, and there is no single silver bullet that provides a general solution of the AIM problem for all contexts and all parameters ranges. Generally, with a few exceptions, most quantum solvers are also computationally demanding, adding a significant overhead to the DFT calculation. Until the latter is addressed, high-throughput material design will remain beyond the reach of most researchers, without access to dedicated large HPC infrastructures: the quantum solver used to provide the solution of the embedding theory, as for every calculation applied to each material of interest, has to be solved hundred's of time with a large inherent computational cost. Furthermore, if approaches such as DMFT are made widely available to engineers and scientists, it would be repeated all across the world, at the cost of thousands of computing hours for every single calculation, generating large hallmarks in terms of energy consumption, data storage and

computational slow-downs. Approximate methods provide cost effective alternatives, such as perturbation based theories, but they are qualitatively wrong in known limits of the theory.
The fundamental problem lies in the fact that the effects of the Coulomb interaction and of the specific material cannot be separated. Otherwise, one could calculate the interaction contributions once and forever, store them and add them every time a new material is calculated.
The most prominent example of the strategy is indeed the local density approximation (LDA) to density functional theory (DFT), where the Kohn-Sham exchange-correlation potential is taken from a homogeneous electron gas that is calculated at the density of the real system in the same point. In this way, DFT profits from the existence of tabulated and interpolated Quantum Monte Carlo (QMC) results, shared with the entire scientific community.
In the latter context, the Anderson impurity model (AIM) plays a central role in the dynamical mean-field theory (DMFT), very similarly to the role of the local density approximation in DFT, where the exchange functional involves solving the quantum problem of a gas of interacting electrons. Contrary to DMFT however, where the AIM is solved every single time without data storing or sharing, the success of DFT has been established by a simple but efficient strategy: the difficult problem of the interacting electronic gas has been solved at a large computational cost with Quantum Monte Carlo, once and for all, and for everyone, shared with the wider scientific community for the benefit of all [6].
It is very surprising that only moderate work has been done on storing and sharing solutions of the AIM obtained at large computational costs (typically in the several thousands of core-hours for every single-point calculation on a given compound), as the obtained many-body solution of the AIM is not specific to a given material. Furthermore, repeatedly solving the AIM correlated problem at large computational cost is also required when only small variations are introduced by finite displacements (for lattice dynamics properties). This limits the scope of predictions for structure optimization or other interesting applications under applied constraints, such as external pressure, due to the overwhelming number of required calculations, when relaxing different chemical compositions with numerous single-point calculations of the same chemical composition are required. This latter fundamental limitation of DMFT, and more generally quantum embedding approaches, can be addressed with advances in the field of data science.
Application of machine learning (ML) to quantum physics as neural networks and gaussian processes to accelerate material discovery have been well studied, for instance for accelerating DFT [7–9] optimizing QMC in terms of speed [10–12] and accuracy [13], and similarly for analytic continuation [14].
In the context of DMFT, the pioneering work of Arseneault *et al.* [15] has opened new avenues in applying ML to the Anderson impurity model for a broad range of parameters. Recently, the application of neural networks (NN) combined with different exact solvers has further established the validity of ML approaches in the context of DMFT [16–18]. These approaches have pioneered in applying ML to quantum many-body systems.
We note that ML and its applicability is limited in terms of offering an overall general solution of quantum many-body systems, for reasonably sized data-bases, due to the rich and complex nature of the many-body problem.

Here, we review the literature on machine learning and neural networks within the context of condensed matter. We also address a simpler task than providing the whole solution of the AIM from a data-science perspective. Instead, we use here physically inspired approximate and cost efficient solvers to DMFT, and outline an approach to provide many-body corrections to account for the obtained errors of the fast solver [19]. The credit to this work goes to the lead author of this work, Dr Evan Sheridan (Phasecraft), and to Dr Francois Jamet (UK National Physics Laboratory) and Zelong Zhao (King's College London).

1.1 Supervised learning and linear regression

We illustrate here the concept of supervised learning within the framework of linear regression. Let's consider for instance a data set that we'd like to fit with a simple regression model. For sake of illustration we'll consider data obtained from electric vehicles, where we relate the mileage in miles per gallon equivalent (MPGe) to the vehicle weight and battery capacity [20].

Vehicle List		
Vehicle weight (Kg)	Battery Capacity (kWh)	Mileage (MPGe)
1000	54	108
1500	81	103
2000	108	98
2500	135	93
3000	162	88
3500	189	83
4000	217	78

The data set is composed of two-dimensional vectors, where x_1^i is the vehicle weight, and x_2^i the battery capacity of a given vehicle. We note that the given features need to be chosen with care for a given problem. To perform any sort of learning, we need to represent a model function for the mileage

$$h_\theta(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \,. \tag{1}$$

The θ parameters, or weights, are the model parameters that will be learned throughout the supervised learning process. Introducing the extra term $x_0 = 1$, the notation simplifies to

$$h_\theta(\mathbf{x}) = \sum_{i=0}^{d} \theta_i x_i \,. \tag{2}$$

In the spirit of regression, we need to identify the optimal parameters θ_i that provide the best model for the known dataset (training set), and ultimately for inferring the mileage on future vehicles (inference process). How can we learn from the available data at hand, and obtain the model parameters? To address this question, we need to define a figure of merit of our model, or a cost function

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2, \tag{3}$$

where y^i are the known mileage obtained from the table above. Minimizing the cost function will hence provide the theoretical model. This can be achieved with a gradient descent algorithm for instance, starting from a initial guess for θ, and repeatedly updating the parameter values by following the steepest gradient,

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta). \tag{4}$$

The gradient descent update is performed over all parameters θ_j simultaneously and the procedure iterated until convergence is achieved. The parameter α plays the role of a *learning rate*, indeed for small α changes in the model parameters θ are small and progress is slow, with the caveat that a large number of iterations is required to reach convergence, whereas for large α changes in the model parameters are rapid, but convergence might be hampered from sudden jumps in the iteration process. To implement the algorithm, one can analytically calculate the derivative and one can easily check that the following is obtained

$$\frac{\partial}{\partial \theta_j} J(\theta) = \big(h_\theta(x) - y\big) x_j, \tag{5}$$

and combined with equation (4) we obtain the *training rule*, also known as Widrow-Hoff rule

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}. \tag{6}$$

As mentioned above, the learning rate is proportional to the proportional coefficient α, but perhaps counter-intuitively the amplitude of the learning is proportional to the error rate $y^{(i)} - h_\theta(x^{(i)})$. Thus, this algorithm learns most from large deviation from the sample, i.e., when our prediction has a large error and $h_\theta(x^{(i)})$ deviates most from $y^{(i)}$. Some practical considerations are absent from the discussion above. In particular, the parameter θ can be updated for every measurement or known data point of the training set, albeit the latter is in reality finite. This highlights the importance of the learning rate. Furthermore, our model is limiting in terms of dependencies and extensions: what if we would like to add a parameter in the model related to the weight of the battery, which is a function of both the total vehicle weight and battery capacity? Such a parameter should feed from the previous variable and provide a higher level model parameter. In the next section we will extend this simple model and learning process to a more general framework that allows for this flexibility. The formulae obtained above are of course nothing else than the well known linear regression method, but it provides a means to set the terminology and general extension to neural networks, discussed in the section hereafter.

1.2 A single layer neural network: the perceptron model

Neural networks are a flexible ensemble of data-driven models, largely inspired by the human's brain network of synapses and neurons, that provide non-linear neural connections. In contrast to a biological neuron, inside the artificial neural network the neuron, or perceptron, is in the form of a simple function whose operation is to take an input vector $\mathbf{x} = \{x_1, \ldots, x_n\}$ (the *feature vector*) and activate a logical threshold if the signal is large enough,

$$f(\mathbf{x}) = \sigma\big(\mathbf{x}.\mathbf{w} + b\big), \tag{7}$$

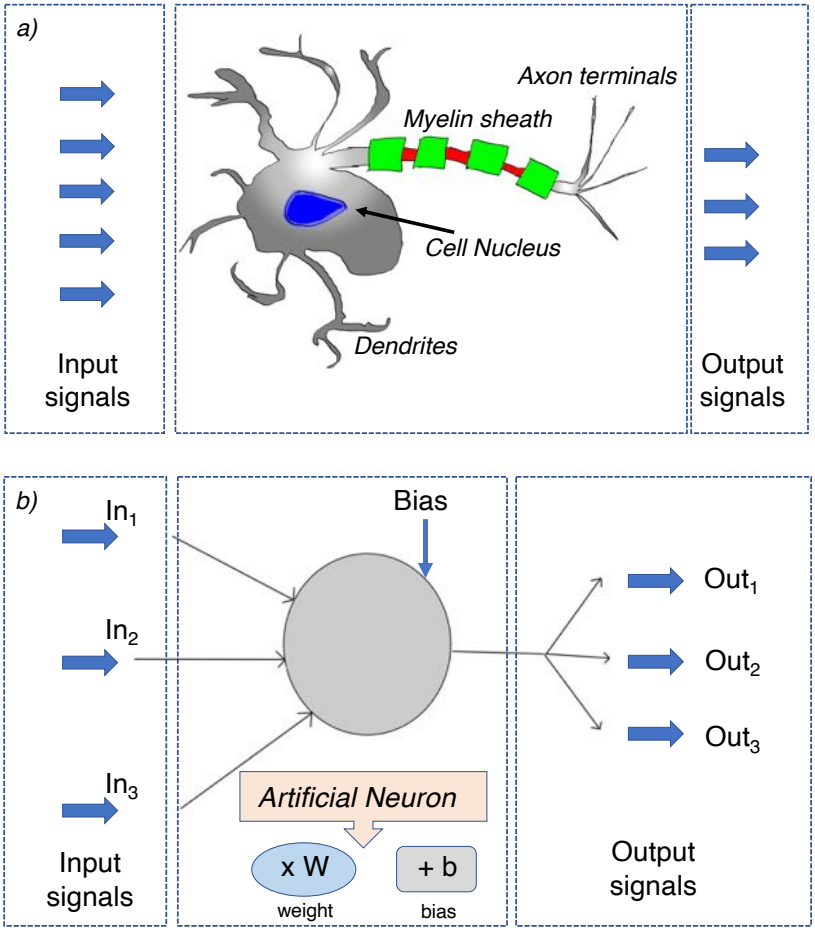


Fig. 1: *a) Schematic of a biological neuron, which consist of a cell nucleus, synapses connected via dendrites to the cell, myelin sheath that embeds the axon that ultimately provides the output signal via the axon terminal to another neuron. b) Simplified artificial neuron model: input signals In_i are multiplied by a proportional weighting factor W and a constant bias b is added to the signal, providing the output Out_i that connect with the other neurons.*

where σ is a continuous activation function, $\mathbf{w}$ is a set of parameters that are specific to the neuron, and b is the bias parameter. Connecting a network of neurons together and adjusting the weights to match the value of a target output provide a mean to use the network to build non-linear predictive responses for different given inputs, which is foundation of learning.

The perceptron model was originally introduced by Frank Rosenblatt, who simulated and built purpose hardware for this model in the early 1960s that provided a direct and parallel implementation of perceptron learning [21]. This model is the first neural network learning model introduced, which is simple and limited, but provides the basic concepts and is a good learning tool. The original motivation for deriving this theory was related to the physiology of the brain learning process, and in particular pattern recognition. The theory is based on a simplified model for the brain neuron: the latter consist of a complex interplay between input signals carried by synapses, interconnected with the neuron cell which provides a time-dependent output signal transported via the axon terminals (see Fig. 1.a). Many complex physiologic phenomena occur via the brain neuron cells in the learning process. The simplest model of an artificial neuron con-

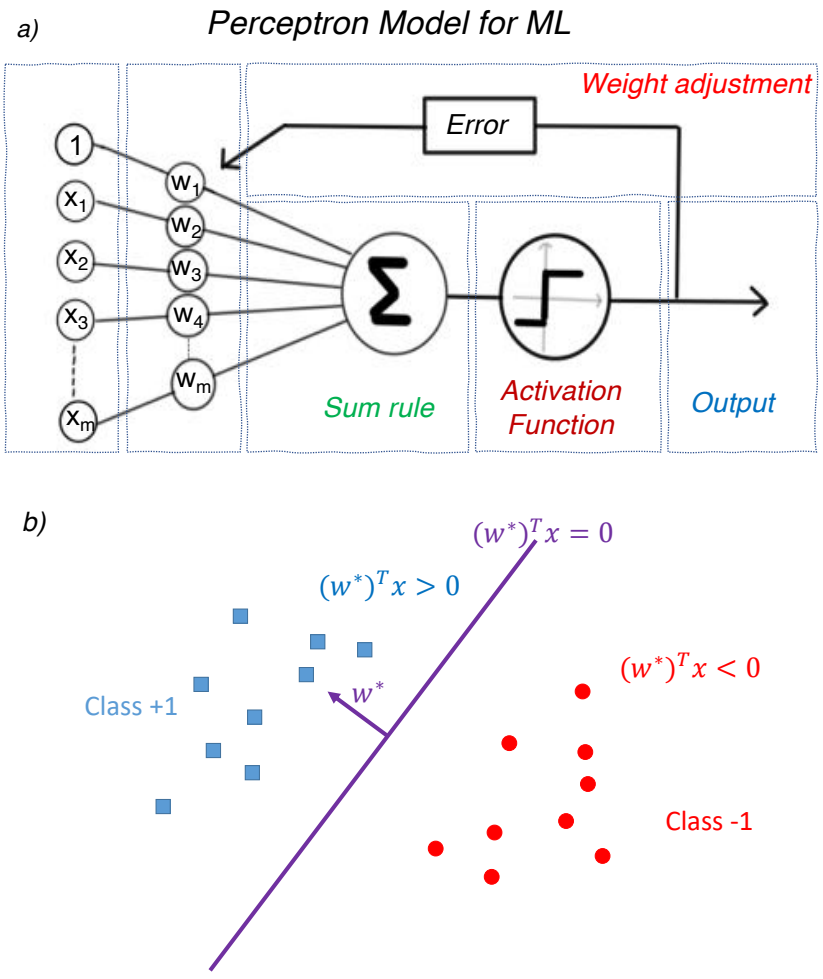


Fig. 2: *a) Schematic of the perceptron model. Inputs x_i are weighted with w_i, and collected by a sum-rule Σ, before entering a threshold logic unit and triggering the activation function. The output is then obtained and compared to the training set. The perceptron learns by the error correction method, where the weights are updated based on the obtained error. b) A typical classification task where the training set consists in a set of data points labeled as circles or squares. After the learning process, the perceptron weights w_i correspond to the equation of the separation line.*

siders a simple proportional relation between input and output signals via a weight coefficient w_i and a constant applied bias b_i (see Fig. 1.b). This model omits the time-dependence of the output signal and many other factors, but provides a basic building block for inter-neuron connections. Typically, the perceptron model consist of a layer of artificial neuron cells, connected to a set of input signals x_i (see Fig. 2a). To mimic the learning process, a summation is applied to the neuron layer, which collects the weighted sum of all input signals. A threshold logic unit is then applied which determines the outcome of the final output binary signal, typically the output signal being $z = 1$ if the learning outcome is positive, and $z = 0$ in the alternative. This provides typically a means to classify data in two categories (classifier). A typical example is a set of data points in Euclidean space which are delimited in two classes, as to whether they

lie above a delimiting line, or below it (see Fig. 2b). The line coefficients w_i are unknown, but instead we know for a group of points whether they belong to the class $+1$ or -1.
In this example, we are provided with a given training data set $\big\{(x_i,\, y_i)\,\big|\, i{=}1,\ldots,n\big\}$. We define the activation function $f_w(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$

- $y = +1, \qquad$ if $\mathbf{w}^T\mathbf{x} > 0$
- $y = -1, \qquad$ if $\mathbf{w}^T\mathbf{x} < 0$

The prediction of the perceptron model is the sign of the activation function $\text{sign}\big(f_w(\mathbf{x})\big)$. The aim of this approach is to learn from the data set and minimize the classification error.
We will use a two neuron model, and following the general recipe, we will set the weights to random initial values: $w_1 = 0.4$ and $w_2 = -0.2$ (note that weights can be either positive or negative). Our training set is set as follow:

Training set		
x_1	x_2	outcome
0.2	0.3	1
0.4	0.1	0

The learning process occurs by testing the algorithm on the training set and to adjust in turns the network weights in the learning process. Weights are typically adjusted by comparing the prediction of the network on a given data point, and correcting for errors obtained in the evaluation. We provide here a simple example and recipe to optimize weights in the single layer neural network, with a simple learning algorithm and objective function. To be more specific, we use the objective function $\mathcal{C}$

$$z \;= 1, \qquad \text{if} \qquad \sum_{i=1}^{n} x_i\, w_i \leq \theta$$

$$z \;= 0, \qquad \text{if} \qquad \sum_{i=1}^{n} x_i\, w_i > \theta$$

where θ is the threshold value. This part defines the logical activation function that converts the signal, modulated by the network weights, into a prediction. The task of the learning process is to train the network weights, for a given objective function, such that the training set is reproduced accurately. For the sake of illustration, we use here a threshold value $\theta = 0.1$.

Training set					
x_1	x_2	ω_1	ω_2	Prediction $\mathcal{P}$	Dataset $\mathcal{D}$
0.2	0.3	0.4	–0.2	1	1
0.4	0.1	0.4	–0.2	1	0

Applying our randomized neural network, we observe that the first training data point is actually well classified by the network with original choice of weights. However, for the second data point, our network produces a wrong prediction. After every error of the network, we perform a weight update with the following learning rule

$$\Delta w_i = \alpha\,(t - z)\,x_i, \tag{8}$$

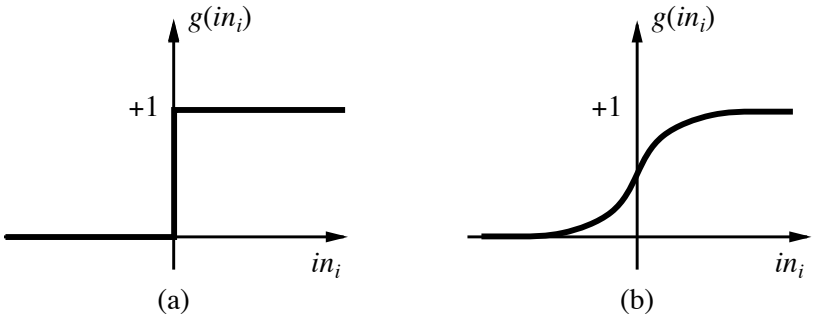


Fig. 3: *a) Activation step-function. b) Activation sigmoid function.*

where t is the target value (training set), z the current output, α the learning rate, w_i the weight associated with input i, x_i the corresponding input value. We note the direct correspondence with the parameter update obtained in equation (6) in the steepest descent approach. The following pseudo-code provides the general approach for a single or multiple layer neural network

Algorithm 1 Neural network pseudo-code

inputs for sample data point: θ,w_i,x_i
Require: $\mathcal{C} = 0$
call Evaluation function: x_i, α, θ, w_i, $\mathcal{D}_i$ output w_i
Prediction $\mathcal{P}_i \leftarrow \sum_{i=1}^{n} x_i w_i \leq \theta$
Cost function $\leftarrow \|\mathcal{D}-\mathcal{P}\|$
Learning $w_i \leftarrow \alpha\,(\mathcal{D}-\mathcal{P})\,x_i$
Iterate over training set

The weights obtained from the neural network will eventually provide a means to predict the class of an unknown data point y_i. The weights have a very simple geometrical interpretation, they represent the line parameters (see Fig. 2.b). In the simple perceptron model, the relation between outputs and inputs remains linear, due to the limited complexity of the model.
We note that in our approach, the choice of the threshold value θ and logical rule to determine whether the weighted signal falls within class A or B is entirely arbitrary. The mathematical formulation of the logical threshold unit is denoted as *activation function*. In general, there is a breadth of possible choices available and studied in the literature, providing different learning efficiency and resilience towards noise, typical examples are the step function and the sigmoid function $(1/\,(1+e^{-z}))$, ReLU $(\max(z,0))$, or tanh (see Fig. 3).
To expand the scope of this approach, and allow for identifying non-linear boundaries between more complex sample sets, the perceptron model can simply be extended by allowing for several neuron layers between inputs and the logical threshold function. The current is simply modulated several times, by the weights of the respective layers. Furthermore, we can also allow for connections between a neuron of one layer with multiple neurons of the next layer, allowing for a large number of weight parameters. This is the realization of a so-called *neural network*.

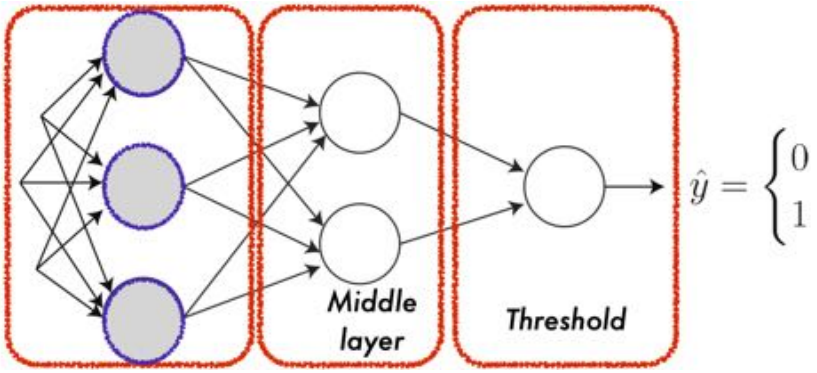


Fig. 4: *Simple extension of the perceptron model, where we have now an intermediate neuron layer between the input fully connected layer and the binary output.*

1.3 Neural networks

A neural network is a direct extension of the perceptron model. It has two main components: (i) the network architecture in terms of number of layers, neurons per layer, and how neurons are inter-connected, and (ii) the parameters defining connections (weights), with the task to learn the parameters for achieving a given task. In our application to DMFT, this task will consist in learning the errors of a given approximate solver to the Anderson impurity model.

In our context, the input will consist of a Green function $G(\tau)$ represented in imaginary time, which is essentially a vector of d dimension $\left(x_1^{(i)}, x_2^{(i)}, ..., x_d^{(i)}\right)$. We'll come back to our main aim in the next sections, but for now we extend the example of the perceptron model where we want to classify an input in a binary class $\hat{y} = 0$ or $\hat{y} = 1$ (see Fig. 4). On the left of this figure, the input is connected to the first layer, the *fully connected layer*. The second layer is denoted as a *hidden layer*, as its presence merely provide additional degrees of freedoms to propagate the information forward to the logical threshold unit. It is worth noting that we have so far only considered forward propagation of the information throughout the network. After doing a first single forward pass through the network, for a given initial choice of weights w_i and a given input vector x_i, we need to update the parameters for the learning process.

Here, we need to generalize the learning formula introduced in the context of the perceptron model. This extension leads to the concepts of training loss, and validation loss. The former is a metric used to assess how the model fits the training data, i.e., it assesses the error of the model on the training set. It is worth noting that the training set is a portion of the entire dataset used to initially train the model. Computationally, the training loss is calculated by taking the sum of errors for each example in the training set. It is also important to note that the training loss is measured after each batch, that is usually visualized by plotting a curve of the training loss after each update of the weights. The latter (validation loss) is a metric used to assess the performance of the learning model on the validation set. The validation set is another portion of the dataset set aside to validate the performance of the model, usually a smaller portion of the dataset as the largest chunk is used to train the model instead (typical splits of the entire dataset in validation/training are 20% validation and 80% training).

The loss function used in neural networks is based on the binary cross-entropy formula

$$\mathcal{L} = -\big(y_i \log_2(\hat{y}_i) + (1-y_i)\log_2(1-\hat{y}_i)\big), \tag{9}$$

where y represents the expected outcome and $\hat{y}$ the outcome produced by the model. Let's have a look at a simple example: for a neural network that tries to determine whether a picture contains a cat, the outcome is either of 1 (cat) or 0 (no cat). With a sample that has two pictures, the first of which contain cats, whilst the second doesn't. Let's imagine that the neural network is 80% confident that the first image contains a cat: $y = 1$ and $\hat{y} = 0.8$. The loss function in equation (9) gives $\mathcal{L} = 0.32$. For the second image, the network gives a 90% probability that there aren't any cats in the picture, $y = 0$ and $\hat{y} = 0.1$, with $\mathcal{L} = 0.15$. The loss function is designed such that either the first term $y_i \log_2(\hat{y}_i)$ or the second term $(1-y_i)\log_2(1-\hat{y}_i)$ are naught or small when the network has a large confidence in asserting the classification, whilst the loss function is large in uncertain situations. The loss function can be averaged over the training sample (or the validation sample), leading to the overall cost function $\mathcal{C}$

$$\mathcal{C} = -\frac{1}{N}\sum_{i=1}^{N}\big(y_i \log_2(\hat{y}_i) + (1-y_i)\log_2(1-\hat{y}_i)\big). \tag{10}$$

For DMFT we are however not focusing on binary classification, and instead our predicted and model values are in general real. The fairly straightforward extension to a real number can be achieved simply by generalizing the cost function with a regression model, for instance

$$\mathcal{C} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{11}$$

1.4 Back-propagation

Once the error is established, the weight update is less obvious than in the case of the perceptron model, where the error obtained on each components x_i could directly be linked and associated with a well defined weight w_i.

Although the gradients of the loss function will provide eventually the weight update, as a generalization of the linear update that we have seen in the previous section, the connection between error and weights is convoluted due to the multiple intermediate layers. Let us first introduce more specific notation for the neural network. Fig. 5 provides a schematic of a maximally connected feed-forward network, where the web of neuron connections is illustrated, with the associated activation values h and the prediction made in the final layer h_1^4. In what follows, we consider training the network on input data $\mathbf{X} = \{x_1, \ldots, x_N\}$ and their associated outputs $\mathbf{Y} = \{y_1, \ldots, y_N\}$. The first-pass through the network consists of assigning $h_i^0 = x_i$ and evaluating the activation functions at each layer of the network as

$$h_j^l = \sigma\Big(\sum_k w_{jk}^l h_k^{l-1} + b_j^l\Big) = \sigma\big(z_j^l\big)\,, \tag{12}$$

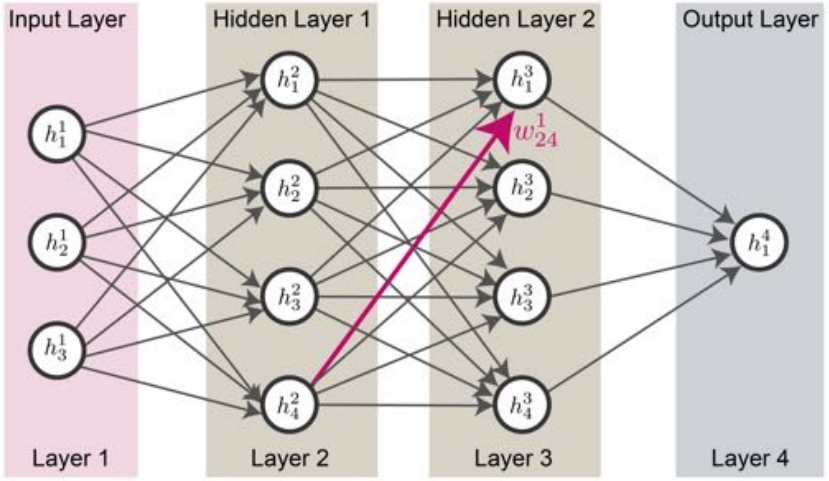


Fig. 5: *Schematic of a feed-forward neural network. An input vector is propagated through the system, via a number of hidden layers, triggering the activation functions h_i^j of each neuron i in layer j.*

where $z_j^l = \sum_k w_{jk}^l h_k^{l-1} + b_j^l$ which can be more compactly expressed as,

$$\mathbf{h}^l = \sigma\big(\mathbf{w}^l \mathbf{h}^{l-1} + \mathbf{b}^l\big). \tag{13}$$

At the end of the first-pass through the network the aforementioned cost function is evaluated,

$$C = \frac{1}{2N} \sum_i^N \big|y_i - h_i^L\big|^2, \tag{14}$$

where h_i^L is a nested function of x_i with many intermediate evaluations of σ,

$$h_i^L = \sigma^L\bigg(\sum_k w_{ik}^L \sigma^{L-1}\Big(\sum_k w_{ik}^L \sigma^{L-2}\big(\dots x_i \dots\big) + b_i^{L-1}\Big) + b_i^L\bigg). \tag{15}$$

Given the nested structure of this function and the sheer number of parameters that $\mathbf{W}_{ij}$ can have, it is prohibitive to find analytical solutions for the combination of weights that minimize equation (14). Instead, established gradient descent methods are applied [22] and define what is now known as the *backpropagation* approach.

The central question of the backpropagation method is to calculate the variation of the cost function with respect to all of the network parameters, $\partial C/\partial w_{jk}^l$ and $\partial C/\partial b_j^l$, and to use these gradients to update the weights. The first step to obtaining these expressions is to express the error in the j-th neuron of layer l as

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} = \frac{\partial C}{\partial h_j^l}\frac{\partial h_j^l}{\partial z_j^l} = \frac{\partial C}{\partial h_j^l}\,\sigma'\big(z_j^l\big), \tag{16}$$

where we have applied the chain rule. In this form, δ_j^l doesn't exploit the connectivity of the overall network. Indeed, calculating

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{h_j^l}\frac{\partial h_j^l}{\partial w_{jk}^l} = \frac{\partial C}{\partial h_j^l}\,\sigma'\big(z_j^l\big)\,h_k^{l-1} = \delta_j^l\,h_k^{l-1}, \tag{17}$$

shows that variation of the cost in layer l with respect to its weight is dependent on the activated output in the preceding layer. Hence, relating the errors from layer-to-layer can allow for a systematic way to calculate the variation of the cost in each layer of the network

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} = \sum_k \frac{\partial C}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \delta_k^{l+1} \frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \delta_k^{l+1}\, w_{kj}^{l+1}\, \sigma'\big(z_j^l\big). \tag{18}$$

Similarly, $\partial C/\partial b_j^l$ can be found as $\frac{\partial C}{\partial b_j^l} = \delta_j^l$. All weights in the network can then be updated by gradient descent in the following manner,

$$w_{jk}^l \to w_{jk}^l - \eta\, \frac{\partial C}{\partial w_{jk}^l}, \quad \text{and} \qquad b_j^l \to b_j^l - \eta\, \frac{\partial C}{\partial b_j^l}, \tag{19}$$

where η is the so-called *learning rate*. The origin of the name *backpropagation* refers to the equations above as for the update δ_j^l one must first know all errors in the subsequent layer δ_k^{l+1}, and hence the error propagates backward through the network.
In summary, the learning process in the neural network consists of repeated forward- and backward-passes throughout the network, after each pass the cost function is reduced. The forward-pass propagates the input forward for the evaluation of the cost function, while the backward-pass updates the network weights starting in the output layer and back-propagating the information to the input layer, that in turn implements the weight updates that reduce the cost function on the next forward pass.

2 Generating a quantum database for the Anderson impurity model

We have now introduced all the concepts in the field of data science required for designing a data-driven approach for solving quantum many-body hamiltonians. To apply the learning process outlined above, we need first to decide on a compact representation of the many-body quantities that will be used in the neural network.

2.1 Polynomial basis method

One way to represent the Green function in a compact formulation is via a polynomial support basis. We expand $G(\tau)$ in an arbitrary orthogonal polynomial basis $P_i^{(k)}\big(x(\tau)\big)$ (e.g., Legendre, Chebyshev, or else) where i is the polynomial order, k is the species and $x(\tau) = 2\tau/\beta - 1$ is a transformation to absorb the temperature dependence from $[-1, +1]$ to $[0, \beta]$. The expansion is

$$G^{(k)}(\tau) = \sum_{i \geq 0} P_i^{(k)}\big(x(\tau)\big)\, G_i^{(k)}. \tag{20}$$

Applying the orthogonality constraints obeyed by the polynomials

$$\int_0^\beta P_i^{(k)}\big(x(\tau)\big) P_{i'}^{(k)}\big(x(\tau)\big) W\big(x(\tau)\big)\, d\tau = \tilde{W}(i)\, \delta_{i,i'}, \tag{21}$$

provides a means to obtain the basis coefficients that can be calculated as

$$\int_0^\beta G^{(k)}(\tau)P_i^{(k)}\big(x(\tau)\big)W(x(\tau))d\tau = \int_0^\beta \sum_{i'\geq 0} G_{i'}^{(k)}P_{i'}^{(k)}\big(x(\tau)\big)P_i^{(k)}\big(x(\tau)\big)\,W\big(x(\tau)\big)d\tau = \tilde{W}(i)G_i^{(k)}. \tag{22}$$

For the purpose of this discussion one restrict the analysis only to the Legendre polynomials, where $W(x) = 1$ and $\tilde{W}(i) = 4/\beta/(2i{+}1)$.
With the G_i, we can express $G(\tau)$ on an arbitrarily fine imaginary time τ grid, absent from discretization constraints. For a given $G(\tau)$, the procedure outlined above provides us with the g_l coefficients, or as an alternative a simple fitting procedure can be achieved, which amounts to the minimization of the function

$$\min_{\{g_l\}}\big(G(\tau) - G_{\text{FIT}}^{\{g_l\}}(\tau)\big), \tag{23}$$

where the fitted (model) function $G_{\text{FIT}}^{\{g_l\}}(\tau)$ is parametrized by the basis coefficients g_l, i.e.

$$G_{\text{FIT}}^{\{g_l\}}(\tau) = \sum_{l\geq 0}^{N_l} \frac{\sqrt{2l{+}1}}{\beta} P_l\big(x(\tau)\big)g_l. \tag{24}$$

It is straightforward to find $G_{\text{FIT}}^{\{G_l\}}(\tau)$ using the conjugate gradient method for example, as the number of Legendre coefficients N_l is generally of the order of a few tenths, typically $N_l \approx 20$. This procedure is indeed exceptionally computationally efficient. Furthermore, it can also be used to filter out noise if a Monte Carlo solver is used to generate the database, shifting the paradigm of dealing with a statistical problem to that of an optimization one. It allows the advantage of including *a-priori* information, such as tail exponents etc.

2.2 Generating the validation dataset with exact diagonalization

Now that we have introduced the representations of our dataset, we need to provide a training set of reference data against which the neural network can learn and validate. As a reference solver for the AIM, it is essential that the solver can efficiently compute solutions in fast and stable ways over a wide range of parameters (Coulomb U, bandwidth W and inverse temperature $\beta = 1/T$). A variety of solvers are available for use in DMFT calculations, and they can be selected to match the computational resources available to complete a calculation in a time-efficient way. As we will see hereafter, we will consider two families of solvers

i. A selection of fast approximate solvers that have low computational cost an provide a reasonable solution of the AIM in some limits of the parameter space, failing in other regions of parameters (e.g. strong or weak interaction limits).

ii. A choice of benchmark solver against which the solvers in (i) above can be compared.

We discuss first the choice of an exact solver for validation and training in (ii). We use here the exact diagonalization method, a typical approach known for solving the AIM in DMFT [23],

where the infinite lattice surrounding the impurity site is approximated by a discretized bath of finite size N_b. The first step is to parametrize the Weiss field $\mathcal{G}_0$ for each orbital and spin in terms of a finite number of bath orbitals by approximating the Weiss field in terms of a non-interacting Anderson impurity model

$$\mathcal{G}^{-1}_{\text{And},m}(i\omega_n) = i\omega_n + \mu - \epsilon^{\text{imp}}_m - \sum_{k=1}^{k=N_b} \frac{V_{m,k}V^*_{m,k}}{i\omega_n - \epsilon_{m,k}} \tag{25}$$

where k is the index for the bath level, m is the index for each orbital/spin. This entails minimizing a distance between the Weiss field and the parametrized impurity Green function obtained in equation (25), using a cost function

$$\chi^2(\epsilon_k, V_k) = \sum_{n=0}^{n=n_c} \mathcal{A}_n \left| \mathcal{G}_{\text{And}}(i\omega_n;\, \{\epsilon_k, V_k\}) - \mathcal{G}_0(i\omega_n) \right|^2. \tag{26}$$

For using ED as a solver for DMFT, it is common practice to weight the cost function towards smaller imaginary frequencies by using a prefactor $\mathcal{A}_n \approx 1/i\omega_n^2$. This avoids fitting cost on the asymptotic regions of the Green function, which are known analytically, but instead provides a good solution in the low energy regime.

Once a set of ϵ_k, V_k has been identified, the calculation proceeds as a standard brute force matrix diagonalization by scanning quantum sectors of the AIM (either total spin S^z or numbers of up and down particles). The focus in DMFT-AI is however reversed: one can limit the database to the large ensemble of parameters ϵ_k, V_k which are tractable with typically 6, 7 or 8 bath sites, such that the solution of the AIM remains exact without the need for iterative solvers (Lanczos, Arnoldi, ...) and the computational cost reasonable. We hence limit ourselves to a large but finite set of corresponding hybridization functions that will be used to train the neural network.

We now turn to the discussion of the *fast and approximate solvers*. Perturbation theory is a well-known and successful diagrammatic method for solving quantum many-body problems in the weak-coupling limit. It is quite often the first port of call in a scientist's arsenal when tackling the quantum many-body problem. The goal of weakly perturbative methods for the AIM is to approximate $\Sigma(i\omega_n)$ analytically with diagrammatic expansions in the Coulomb repulsion U/t for (t is the hopping term in the Hubbard model)

$$G_0^{-1}(i\omega_n) = \Sigma(i\omega_n) + G^{-1}(i\omega_n). \tag{27}$$

Weak coupling expansions in U/t up to second order, $\mathcal{O}(2)$, were successfully used for the AIM [1] to capture the main features of the Mott transition at strong coupling in the nonperturbative regime. This only applies at half-filling, and can be attributed to the simple form of the atomic Green function in the $t/U \to 0$ limit being proportional to U^2 [1]. Nonetheless, iterative perturbation theory (IPT) is extremely useful for generating solutions for the AIM at low computational cost, in spite of the parameter regimes where the solution can be qualitatively wrong. Specifically, the second order perturbation of $\Sigma(i\omega_n)$ at inverse temperature β and half-filling is given by,

$$\Sigma^{\text{IPT}}(i\omega_n) = \Sigma^1(i\omega_n) + \Sigma^2(i\omega_n) = \frac{U}{2} + U^2 \int_0^\beta d\tau\, e^{i\omega_n\tau} G_0^2(\tau)\, G_0(-\tau), \tag{28}$$

where $\Sigma^{1,2}(i\omega_n)$ consist of the non-skeleton $\mathcal{O}(2)$ representations of the self-energy. Iterating over (27) with the above form of the self-energy is the foundation of the IPT theory, that successfully captures the Mott transition. Higher order diagrams can also be readily incorporated into this approach, to provide further corrections, but the complexity rapidly increases with the diagrammatic order, limiting the scope of manually correcting this approach. Practically, this amounts to replacing (28) with

$$\Sigma(i\omega_n) = \Sigma^1(i\omega_n) + \Sigma^2(i\omega_n) + \Sigma^3(i\omega_n), \tag{29}$$

where $\Sigma^3(i\omega_n)$ encapsulates all of irreducible third-order processes allowed, coming at an additional computation cost due to calculating the integrals associated with the higher-order diagrams and their additional interaction vertices. IPT is a good example to illustrate the data-driven approach outlined in these notes: the neural network that we will discuss below learns the error obtained in IPT and provides a highly non-linear solution in a multi-dimensional space to account for $\Sigma^n(i\omega_n)$ with $n > 2$, absent in standard IPT.

There are of course other approaches than neural networks that deal with corrections beyond second-order perturbation theories, and out of half-filling. For instance the Non-Crossing Approximation (NCA) is the lowest order strong-coupling perturbative method that sums up all diagrams without crossing hybridization lines. In this scheme the propagator of the impurity is mapped to an integro-differential Volterra equation that is solved for the strongly-coupled form of the self-energy in equation (27).

Both NCA, IPT, and the exact-diagonalization solver with a very small number of bath sites (typically zero, also known as Hubbard-1, or with $N_b = 1, 2, 3$ bath sites) represent a valid ensemble of approximate solvers which all introduce a negligible overhead in terms of calculations, and also all need corrections for providing quantitatively accurate solutions of the AIM, covering both the weak- and strong-coupling limits in the phase-space of the AIM. Finally, we note that those solvers are tractable and can also provide solutions in both real and Matsubara frequencies, but we'll limit the discussion in a first instance to the imaginary time formalism hereafter.

2.3 Data representation

The construction of a high-quality database of training samples is of key importance for any data-driven approach. Specifically, there must be sufficient *representative* samples, such that after the training process the inference process will produce the most likely outputs. Strategies for generating databases in machine learning are key for the success of any data-driven approach, and require great care in identifying robust and well thought strategies.

Before presenting the database construction at great lengths, we first need to view the AIM from a data-science perspective. Bearing this in mind, we look at the AIM from a black box perspective, and regard it only in terms of its inputs and outputs. In doing so, the AIM merely provides a relation between input $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_{N_S}}\}$ and output samples $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y_{N_S}}\}$, where N_S is the number of database samples or images.

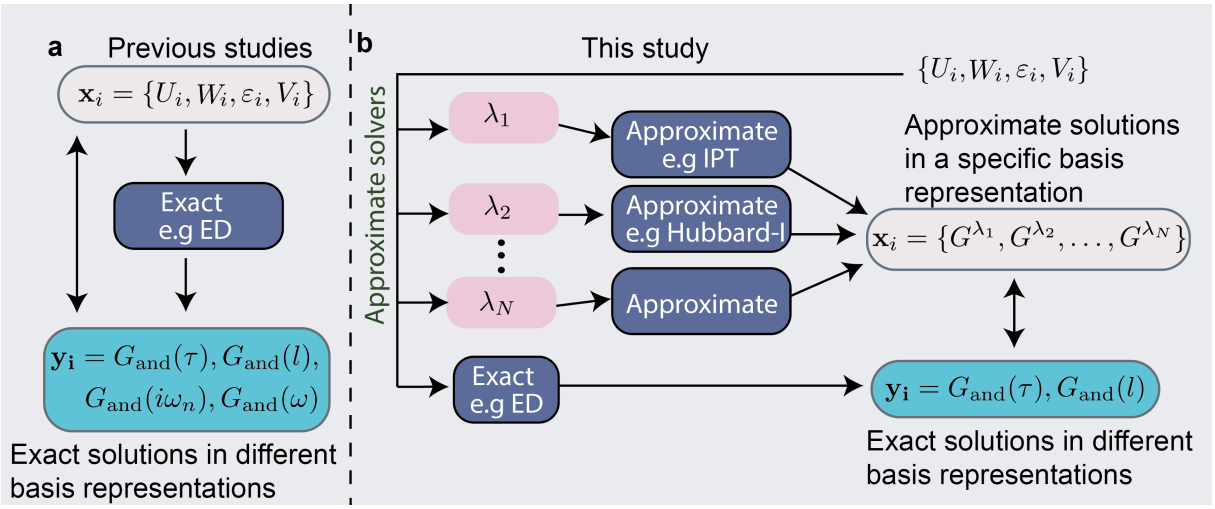


Fig. 6: *Depiction of two approaches for machine learning solutions of the Anderson Impurity Model.* ***a*** *The approach taken by [15] which uses as input features* $\mathbf{x}_i$ *the physical parameters of the AIM* $\{U_i, W_i, \varepsilon_i, V_i\}$ ***b*** *The approach taken here uses approximate solutions, expressed in different bases,* (λ_k)*, of the AIM as the feature vectors.*

We limit our discussion to the single-orbital AIM, which is completely described by the set of parameters $\{U, W, \varepsilon, \beta, V\}$, where U is the Hubbard parameter, W is the half-bandwidth of the bath-states, ε is the impurity on-site energy, β is the inverse temperature and V characterizes the impurity-bath coupling. Furthermore, we can split these contributions into the different components of the AIM, where $\{U, \varepsilon\}$ represent the physics of the impurity while $\{W, V\}$ represent respectively the bath. β is viewed from the perspective of the grand canonical as a fixed parameter of the entire system and the database hence generated at constant temperature. This doesn't preclude generating data for various temperatures, as indeed we'll discuss below the differences that occur in the network as we move from high to low temperature.
In practice, the AIM can absorb inputs in different representations, for instance a set of hamiltonian parameters for ED, or a function in imaginary time for IPT and NCA (the Weiss field). Typically, one might have

- inputs: $\mathbf{x}_i = \{U_i, W_i, \varepsilon_i, V_i\}$
- outputs: $\mathbf{y}_i = \{G(\tau), G(l), G(i\omega_n), G(\omega)\}$,

depending on the choice of basis. The basis choices above are the imaginary-time, Legendre, imaginary-frequency and real-frequency bases, respectively.
Outputs can however be seen as vectors of given length, and in this case, each input vector has a dimension of 4, whereas the output has the number of imaginary-time points (usually $N_\tau > 200$) for the imaginary time Green function, while it is known that the number of Legendre coefficients $N_l < 50$ is more compact [15]. The mapping is illustrated in Fig. 6a. In Ref [15], the authors employ a similar approach for the input features, but attempt to learn the spectral function $A(\omega)$ instead.

2.4 Defining the unknowns: from learning solutions to learning errors

The approach discussed in this lecture is different with regard to the input vectors $\mathbf{x}_i$, where instead we use a combination of approximate solutions to the AIM. We illustrate the idea in Fig. 6b for the i-th AIM instance, which is still characterized by the set $\{U_i, W_i, \varepsilon_i, V_i\}$, but the feature space increases by $N_\lambda N_b - 4$, where N_λ is the number of approximate solvers used, λ refers to the different impurity solver types, and N_b is the number of basis coefficients used.

For example, if we consider using two different approximate solvers $\{\lambda_1, \lambda_2\}$, then the input features for a model would be (let's say for IPT and NCA),

$$\mathbf{x}_i = \left\{G^{\lambda_1}_{\text{and}}(\tau), G^{\lambda_2}_{\text{and}}(\tau)\right\} \quad \text{or} \qquad \mathbf{x}_i = \left\{G^{\lambda_1}_{\text{and}}(l), G^{\lambda_2}_{\text{and}}(l))\right\}, \tag{30}$$

in the imaginary-time or Legendre polynomial bases respectively, and the output vectors are given by,

$$\mathbf{y}_i = G_{\text{and}}(\tau) \quad \text{or} \qquad \mathbf{y}_i = G_{\text{and}}(l). \tag{31}$$

In this case, the length of the feature vector is $2N_b$ and the length of the output vector is N_b. This is a straightforward generalization, we will discuss how this impacts the cost function hereafter.

Moreover, instead of learning the relationship between the input parameters $\{U_i, W_i, \varepsilon_i, V_i\}$ and the exact solution of the AIM, we attempt to learn the *error* between a set of approximate solutions of the AIM and its exact solution, generated using a reference solver such as ED, but equivalently, quantum Monte Carlo can be used as a valid exact benchmark at higher computational costs (we note however that the compute cost for generating a database is not a large concern, as these calculations only need to be performed once and for all).

We then learn the error of a set of approximations, rather than using just one, noting that different approximate AIM solvers have their merits in different parts of the AIM parameter space.

As mentioned in the previous section, IPT is a well-known and successful diagrammatic method for solving quantum many-body problems in the weak-coupling limit, and can also capture some features of the strongly interacting limit. Similarly, the Hubbard-I approach is exact in the weakly hybridized atomic limit, but can qualitatively fail outside of this parameter regimes.

We note that while the weak-coupling expansion has its merits, for scenarios out at strong-coupling it can be qualitatively wrong. To address this, for example, the Non-Crossing Approximation (NCA) is the lowest order strong-coupling perturbative method that sums up all diagrams without crossing hybridization lines.

It is also possible to use basis-truncated approximate ED solutions that use fewer bath sites to represent the Weiss field, which thus provide a more consistent coverage of the parameter AIM space, albeit with larger errors when small numbers of bath fitting parameters are used.

Thus, a natural extension for data driven methods lies with the combination of different approximations, which span a wider range of the AIM parameter space, rather than just using one quantum solver alone.

2.5 Constructing the database

Having established the form of the inputs and their associated outputs, we now discuss the construction of the database. We first deal with inverse temperature, β: unlike the rest of the parameters, the Green function is explicitly dependent on it, so each database is constructed at a specific β. The next parameters to decide are $\{U, W, \varepsilon\}$, all of which are randomly distributed between their extremal values. Therefore, for each instance of an AIM, random samples of each are drawn from the uniform distributions $U \in [U_{\text{min}}, U_{\text{max}}]$, $W \in [W_{\text{min}}, W_{\text{max}}]$ and $\varepsilon \in [\varepsilon_{\text{min}}, \varepsilon_{\text{max}}]$.

Determining the hybridization parameters is slightly more difficult, as regards to how it manifests itself in the impurity solver. The ED solver necessitates a discrete representation of the bath parameters, while this is not true for other impurity solvers, and there is a connection between the lattice bandwidth W and the range of hybridizations that need to be considered in the database.

Furthermore, we want to be able to deal with both discrete and continuous representations of the bath to allow for various approximate solvers. For illustration, we take the Hilbert transform of specific form of the density of states, i.e.,

$$G(z) = \int_{-\infty}^{\infty} \frac{A(\epsilon)}{z - \epsilon}\, d\epsilon \tag{32}$$

with two typical density of states samples being the semi-circular DOS given by

$$A(\epsilon) = \frac{2\sqrt{-\epsilon^2}}{(\pi W)^2} \Theta\big(W - |\epsilon|\big) \tag{33}$$

or the constant DOS, given by,

$$A(\epsilon) = \frac{\Theta\big(W^2 - \epsilon^2\big)}{2W}, \tag{34}$$

where Θ is the Heaviside step-function.

Using either, truncates the limits of integration in Eq. (32) from $-W$ to $+W$, where W is the half-bandwidth, and provides hence the energy scale associated with the hybridization parameters V in the Weiss field ($z = i\omega_n$), the latter being given by,

$$\Delta(i\omega_n) = \sum_{i=1}^{N} \frac{V_i^2}{i\omega_n - \epsilon_i}, \tag{35}$$

where V_i and ϵ_i are the bath parameters, and which become an additional parameter in the database construction.

Specifically, along with the number of samples in the database and the inverse temperature β, the number of bath parameters determines the overall time it takes for the construction of the database. After the number of bath sites is chosen, random samples of each are drawn from the uniform distributions $V \in [V_{\text{min}}, V_{\text{max}}]$ and $\epsilon \in [\epsilon_{\text{min}}, \epsilon_{\text{max}}]$.

To ensure that the discrete representation of the hybridization function retains consistent physical characteristics, its bath parameters are all scaled to the chosen values of W, such that both ϵ_i

and V_i are normalized by it and ϵ_i is centered on its weighted arithmetic mean with respect to V_i^2. Alternatively, it is possible to create discrete representations of the bath by treating ϵ_i and V_i as fit parameters in Eq. (35) to a continuous representation generated from the half-bandwidth W. The next step is to generate the database that will be used for the training of the data-driven model. To do this, each instance of $\{U_i, W_i, \varepsilon_i, \Delta_i(i\omega_n)\}$ is passed to the set of approximate solvers $\{\lambda_1, \ldots, \lambda_N\}$ as well as one exact solver. In this case, the exact solution is obtained by the ED algorithm using a large number of bath sites, generally between 4–6 is enough to ensure a converged solution for the single-site AIM.

3 Training a model solver to solve the Anderson impurity model

We now outline the details of the multivariate maximally connected neural network regression model that is used for the training against the database we have just constructed. As established above, the set of inputs for the model are $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_{N_s}}\}$, where $\mathbf{x_i}$ is a set of different approximate solutions of the AIM, while the outputs are $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y_{N_s}}\}$, where $\mathbf{y_i}$ is an exact solution of the AIM given by ED. For all models trained in this section, they proceed by minimizing the cost function,

$$C(\mathbf{X}, \mathbf{Y}, \boldsymbol{\alpha}) = \frac{1}{N_s} \sum_{j}^{N_s} \big(\mathbf{y}_j - g_{\boldsymbol{\alpha}}(\mathbf{x}_j)\big)^2, \tag{36}$$

with respect to the parameters α to produce a model $g_{\boldsymbol{\alpha}}(x_i) := G^{\mathcal{M}}(x_i)$, where $G^{\mathcal{M}}(x_i)$ is the model Green function of the problem. $G^{\mathcal{M}}(x_i)$ is constructed such that the error between it and the true solution $\mathbf{y}_i$ is minimized, and therefore $G^{\mathcal{M}}(x_i)$ corrects for the error between the approximate solution x_i and the exact one y_i, for all N_s entries in the database. The neural network we use is shown as a schematic in Fig. 7. In the input layer, each neuron evaluates

$$f(\mathbf{x}\cdot\mathbf{w} + \mathbf{b}) = f\Big(\sum_{i,j} w_{ij} G^{\lambda_j}(k_i) + b_1\Big) \tag{37}$$

with $f(\ldots)$ being the activation function of the input layer neurons (colored pink), index i is associated to the feature (i.e. mesh point) and index j indicates the approximate solver used, w_{ij} are the set of neural weights, and $\mathbf{x}_i$ is in general of the format λ_N entries, despite the depiction in Figure 7 that suggests that the number of approximate solvers is 2. This procedure then repeats itself as the values propagate forward through the network such that $f(\mathbf{x}\cdot\mathbf{w} + \mathbf{b})$ of each neuron are used as the inputs for the next layer in the network, until eventually the output layer is reached. As the neural network is being used to solve a regression problem, the output layer applies a linear activation function to its neurons, which does not modify its input data. Therefore, when the output layer is reached the cost function (36) is evaluated for a "mini-batch" of samples, after which the weights throughout the entire network are updated in accordance with the backpropagation method. This procedure is then repeated until $G^{\mathcal{M}}$ is found with weights α that minimize $C(\mathbf{X}, \mathbf{Y}, \boldsymbol{\alpha})$.

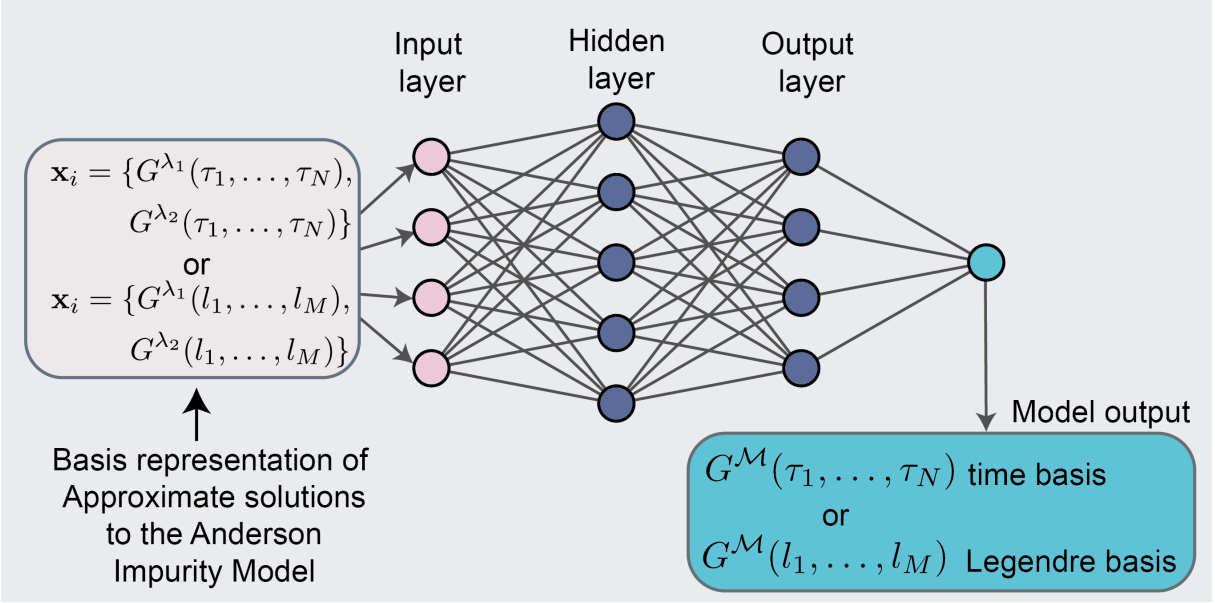


Fig. 7: *Depiction of the maximally connected feed-forward artificial neural network used to generate solutions of the Anderson Impurity Model to produce a model output $G^{\mathcal{M}}(\tau)$ or $G^{\mathcal{M}}(l)$. In the schematic, there is 1 hidden layer with 5 neurons and the output layer has 5 neurons.*

It is important to keep in mind the number of parameters in the model $g_\alpha(x_i)$ so as to avoid potential over-fitting scenarios. For the neural network presented in Fig. 7 the total number of parameters N_α can be determined by the following equation,

$$N_\alpha = \sum_l \left(N_N^l N_N^{l-1} + 1\right), \tag{38}$$

where N_N^l is the number of neurons in layer l. For example, if there are 20 neurons in the input and hidden layer, 100 neurons in the output layer and 200 is the length of the input vector, then the total number of parameters of the network broken down per layer is given by:

$$N_\alpha = \underbrace{20(200{+}1)}_{\text{input layer}} + \underbrace{20(20{+}1)}_{\text{hidden layer}} + \underbrace{100(20{+}1)}_{\text{output layer}} = \underbrace{6540}_{\text{total}}. \tag{39}$$

The value of N_α is pertinent when considering sources of data over-fitting, as it should not exceed to the total number of feature observations in the database. In addition to what determines the number of weights in the neural network, the following series of adjustable parameters have an effect on its performance They are usually referred to as hyperparameters:

- **Learning rate:** step-size update for the weights of the network
- **Mini-batch size:** number of samples after which the neural network weights are updated
- **Epochs:** number of sweeps of the neural network
- **Activation functions:** family of non-linear neuron activation functions, including $\tanh$, elu and relu

- **Cross-validation split:** $\%$-split of the database between training/validation samples
- **Basis functions:** equidistant τ-basis, adaptive τ-basis or Legendre G_l-basis

Typically a hyperparameter grid search is employed over these parameters and by doing so, the model is trained iteratively as the learning steps occur across different values of its parameters. Ultimately it will produce the final optimized value of the cost function for both the training and validation datasets, where the minimum cost function provides a measure for the optimum choice of parameters to be used for the inference process in future DMFT calculations.

3.1 Data processing: symmetry, augmentation and transformation

Once the database of approximate and exact solutions is obtained, and before the data is passed on to the machine learning algorithm for training, there are a number transformation operations that allow the database to be augmented through symmetry operations. Without having to re-run the impurity solver, there are a number of ways to both extend and transform the database in ways that are optimal for learning a model.
For simplicity, we assume a database under consideration is expressed in either the imaginary-time or in the Legendre polynomial basis.
The first symmetry operation makes use of the fact that the Green function can be decomposed into its symmetric and anti-symmetric contributions by decomposing it into the Legendre basis,

$$G^{\mathrm{S}}(\tau) = \sum_{\substack{l\geq 0 \\ \text{even}}} \frac{\sqrt{2l+1}}{\beta} P_l\big(x(\tau)\big)\, G_l \quad \text{and} \qquad G^{\mathrm{AS}}(\tau) = \sum_{\substack{l\geq 0 \\ \text{odd}}} \frac{\sqrt{2l+1}}{\beta} P_l\big(x(\tau)\big)\, G_l, \tag{40}$$

where $G^{\mathrm{S}}(\tau)$ and $G^{\mathrm{AS}}(\tau)$ are respectively the symmetric and anti-symmetric parts of total Green function which give the total Green function when summed, i.e.

$$G(\tau) = G^{\mathrm{S}}(\tau) + G^{\mathrm{AS}}(\tau). \tag{41}$$

In practice, if performed in the τ basis, the latter requires an intermediate step of generating the Legendre coefficients, or reading them in from a database which has them stored already. For the size of the databases dealt with in this lecture (typically less than 40k samples), the latter can be added practically at no additional computational cost. Physically, the symmetric part of the Green function represents the physics at half-filling while the anti-symmetric component encodes the information away from half-filling. This operation need not only be used for the augmentation of the database, it can similarly be used for partitioning it. Specifically, instead of training a model on both the symmetric and anti-symmetric components simultaneously, two separate models can be trained on the symmetric and anti-symmetric components separately, after which they are recombined to produce the total answer in Eq. (41).
The same procedure can be followed in the Legendre basis, where the symmetric part of G_l is encoded in the even coefficients and the anti-symmetric part in its odd ones. For both bases, this operation allows the database to be augmented by a factor of two.

The strategy to learn different features on a given dataset is very much akin to the concepts developed in deep learning: an image would be decomposed in different features with different characteristics, and the neuron model optimized for such.
The second symmetry operation on $G(\tau)$ that we can consider is particle-hole equivalence, i.e., $G^e(\tau) = G^h(\beta-\tau)$ where G^e is the electron Green function and G^h is the hole Green function. Therefore, for every entry in the database that is away from half-filling, the corresponding electron (if hole-type) or hole (if electron-type) Green function can be generated simply by flipping $G(\tau)$. If, however, the database is expressed in the Legendre basis instead, this transformation requires that the odd coefficients be multiplied by -1. We note that for both basis representations, this can double the size of the database.

3.2 Activation function with many-body quantities

In addition to augmenting the database by exploiting symmetry operations, the data must also be transformed into a representation appropriate for how the training data will be manipulated, an in particular for designing a suitable activation function.
Scaling the input and output variables so that they are normalized is a standard technique when preparing data for training algorithms such as neural networks. One example for instance: standard activation functions, as seen in the early chapters of this lecture, are dealing usually with positive signals, so care will be needed to manipulate and transform the Green functions in a suitable format.
Another need for the mapping lies with protecting the weights that are learned in the model from becoming too large or biased towards large input values. Specifically, this is essential for when input variables are the Legendre coefficients, as the Legendre basis has no inherent scale for the coefficients. On the other hand, while an inherent scale exists for $G(\tau)$, i.e., $-1 \leq G(\tau) \leq 0$ when $\tau > 0$, it is also possible to create a family of scaling transformations and test their efficacy throughout the training process. The following scaling transformations work regardless whether the aforementioned symmetrization or augmentation procedures have been followed.
$\mathcal{T}_0$ is the unscaled Green function and each transformation is a function of $\mathcal{T}_0$. For $G(\tau)$ the situation is quite simple, there are only a few transformations that can be done to normalize in between the range $[0, 1]$ or $[-1, 1]$. We note that if $G^{\mathrm{AS}}(\tau)$ is used, i.e., the anti-symmetric part of the Green function, it is important to ensure that the scaling operations do not shift the data out of the scaling range, and so an extra constant shift should be applied in these cases to counteract this behavior. For the Legendre basis, it is clear that the unscaled data is not normalized. Fortunately, by applying a *tanh* function this can readily be achieved. In the example shown, we see the first anti-symmetric component of the Green function, G_1, is scaled to be much closer to G_0 and G_2, the either-side symmetric components. As stated above for the τ basis, the dependence of training the model is also assessed as a function of these transformations.
Moreover, we emphasize that to recover the physical Green function it is necessary to apply the relevant inverse transformation $\mathcal{T}^{-1}$, which are given in Table 1.

a)

	$\mathcal{T}_r$	$\mathcal{T}_r^{-1}$
$\mathcal{T}_0$	$G(\tau)$	$G(\tau)$
$\mathcal{T}_1$	$-G(\tau)$	$-G(\tau)$
$\mathcal{T}_2$	$G(\tau)+1/2$	$G(\tau)-1/2$
$\mathcal{T}_3$	$-2\big(G(\tau)+1/2\big)$	$-G(\tau)/2-1/2$
$\mathcal{T}_4$	$2\big(G(\tau)+1/2\big)$	$G(\tau)/2-1/2$

b)

	$\mathcal{T}_r$	$\mathcal{T}_r^{-1}$
$\mathcal{T}_0$	G_l	G_l
$\mathcal{T}_1$	$\tanh(G_l)$	$\tanh^{-1}(G_l)$
$\mathcal{T}_2$	$-\tanh(G_l)$	$-\tanh^{-1}(G_l)$

Table 1: *Scaling transformations for a): $G(\tau)$ and b): G_l*

3.3 An error correction approach for solving DMFT

Here we present results that pertain to the machine learning framework outlined above. We begin with a discussion of the different aspects of the generated databases, and this is followed by details on the training of an artificial neural network with those generated databases. We conclude by illustrating how the generated data-driven solver is able to capture the Mott transition in the half-filled single-band Hubbard model the using DMFT scheme presented above.

3.4 Database of solutions for the Anderson Impurity Model

Using the procedure outlined in the previous section we generate a database of size $N_s = 10^3$ at inverse temperatures of $\beta = \{1, 2, 5, 10, 20, 50\}$ eV^{-1} over the parameter ranges indicated in Table 2 for discrete sets of bath parameters. The range of temperatures chosen represent the high-temperature and intermediate temperature limits, whereby the features of the Green function are smoother, and hence our choice of the range. Each database entry constitutes a random combination of all parameters in Table 2. The parameters chosen cover a range of physical features, for example the Hubbard U is uniformly randomly sampled over the range $\{1, \ldots, 10\}$, in addition to $W \in \{1, \ldots, 10\}$ and $\varepsilon \in \{-1, \ldots, 1\}$, then the various combinations of U, W, ε result in metals or insulators. Take for instance if $\{U, W, \varepsilon\} = \{8, 2, 0\}$ then the result is insulating, and if $\{U, W, \varepsilon\} = \{2, 8, 0\}$ the result is metallic. In Table 2 we clarify the notation for the approximate solvers ED-[1,2,3]. ED-1 means solving the AIM with one bath site only, and hence results in a truncated approximation to the exact ED solution of the AIM (which in this case uses 4 bath sites). We note that these latter ED solvers are significantly faster than the ED solution obtained at large cost for $N_b > 6$, due to the exponential increase of the Hilbert space. Of course the latter are themselves approximate solutions, similar to IPT or Hubbard-I, and the large error induced by the finite size effects of the bath discretization. We note that interestingly the machine learning framework does act in this respect as a *Hilbert space extrapolation* tool, inferring information on small Hilbert spaces that remains pertinent for larger dimensions.
Furthermore, in Fig. 8 we show the distribution of all chosen parameters for the $10,000$ samples in the database corresponding to $\beta = 20$ eV^{-1}. As expected, $\{U, W, \varepsilon\}$ are distributed evenly, $N_b = 4$ is constant as the number of bath-sites is not changed, and $\{\epsilon_i, V_i\}$ are chosen by normalizing to W. While the illustrated database is not the only one that could be considered, it is not a special choice. For all other databases we analyzed, the distribution of parameters behaves similarly to the $\beta = 20$ eV^{-1} case presented.

U (eV)	$\{1, \ldots, 10\}$
N_{bath}, ϵ_i, V_i	4
W (eV)	$\{1, \ldots, 10\}$
ε	$\{-1, \ldots, 1\}$
β (eV^{-1})	$\{1, 2, 5, 10, 20, 50\}$
N_{samples}	10,000
$\mathcal{S}$	Hubbard-I, IPT, NCA, ED-[1,2,3]

Table 2: *Parameter selection for the database of AIM solutions shown in Fig. 8 where* $\{p_i, \ldots, p_f\}$ *denotes that a parameter is randomly selected from this interval* $[p_i, p_f]$*. U is the Hubbard interaction, N_{bath} stands for the number of bath sites, W is the Half-bandwidth, ε is the electron on-site energy, β is the inverse temperature, $N_{samples}$ is the number of database entries, and $\mathcal{S}$ denotes the total ensemble of approximate quantum solvers used in the ML approach. ED-[1,2,3] denotes the exact diagonalization solver with respectively* $1, 2, 3$ *bath sites.*

We review the strength and weaknesses of the Hubbard-I, IPT and NCA solvers for representative samples of the database against the corresponding exact diagonalization results. The latter provide valid approximations of the AIM in different limits (Hubbard-I and IPT are good in the weakly hybridized limit, NCA is a good approximation for stronger interactions). In general, the Hubbard-I, IPT and NCA solvers are however in quantitative and qualitative disagreement.

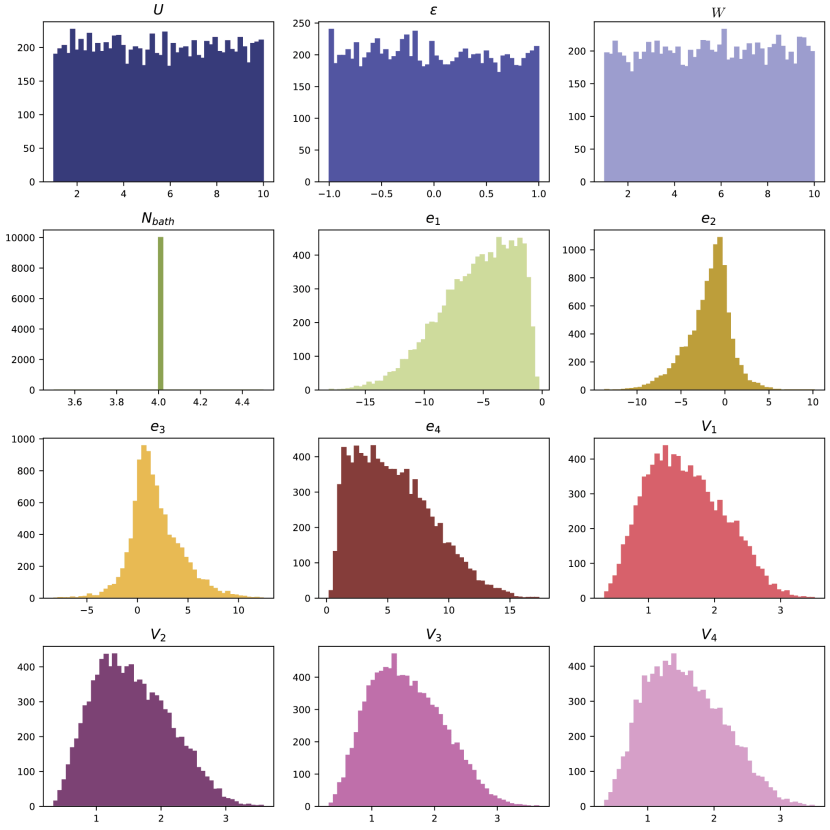


Fig. 8: *Typical statistical distribution of the Anderson Impurity Model parameter space for a database for* $\beta = 20\ eV^{-1}$*,* $N_{bath} = 4$ *with 10,000 entries.*

Hyperparameter	Range	Optima
η	$[0.0001, 0.0002, 0.001, 0.01]$	0.0002
Mini-batch size	$\{4, 8, 16, 32, 64\}$	8
Optimizer	{Adamax, Nadam}	Nadam
Activation functions	{elu, relu, *tanh* }	*tanh*
Hidden Layers	$\{1, 2, 3, 4\}$	1

Table 3: *Hyperparameter grid-search over the neural network parameters with a fixed number epochs = 20 and* $\beta = 1$ *eV*$^{-1}$

For example, the Hubbard-I solver indicates for highly hybridized AIM an insulating solution when in reality the system is metallic. Nevertheless, we emphasize that this behavior is expected and welcome, since the end-goal is to systematically correct for the error between the approximate and exact solutions.

3.5 A Neural Network Impurity Solver

The very first step to training a machine learning model is the hyperparameter grid-search over its independent parameters using tensorflow [24]. Specifically, for our neural network we coarse-grain the number of epochs to 20, set the inverse temperature to $\beta = 1$ eV^{-1} and scan across all combinations of parameters in Table 3. Ultimately, 401 different neural networks are trained and the combination of parameters with the minimum cost function $\sim 10^{-6}$ is given by that combination of parameters shown in the "Optima" column of Table 3. Additional fixed parameters in the grid search are: evenly spaced time grid, Hubbard-I, IPT and NCA solvers as inputs $\mathbf{x}_i$ to the neural network as they all require minimal computational resources in comparison to the ED methods, no data augmentation, and the $\mathcal{T}_4$ imaginary-time transformation from Table 1. It is noteworthy that either increasing the complexity of the network, i.e., increasing its depth beyond 1, or increasing the learning rate to an order beyond 10^{-2} has detrimental effects on the minimization of the cost function. Practically, it would be computationally prohibitive to perform this grid search for every β and their additional free-parameters. In what follows, all networks use the optimal values as specified in Table 3 and use an $80/20$ cross-validation split, i.e., 80% training data and 20% validation data.

We propose a collection of data scaling transformations of the input and output data which improve the training of the neural network in the imaginary-time and Legendre bases. Fig. 9 presents the validation loss for these scenarios, for $\beta = 1$ eV^{-1}. For the Legendre basis the effect of data transformations is quite significant, as shown in Fig. 9. Here we see that by applying a *tanh*-type Legendre transformations, the final value of the loss can be improved by at least 2 orders of magnitude, reduced from 10^{-4} to 10^{-6}. We expect the effect of this transformation to be enhanced for larger values of β (lower temperatures), where the range of G_l can greatly exceed the value of unity, and therefore necessitates the application of an appropriate data transformation. At higher temperatures (*i.e* lower β), the Legendre coefficients are often bounded close to unity, and so the neural network is less sensitive to the untransformed input as compared to lower temperatures.

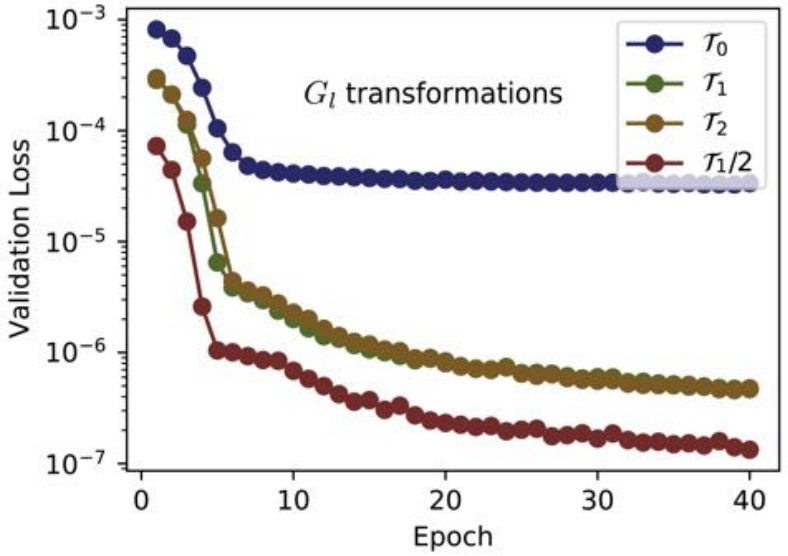


Fig. 9: *Validation loss for the Legendre basis transformations at $\beta = 1$ eV^{-1}. The transformations applied on the database lead to different figures of merit for the network predictions.*

In Figure 10 we show the value of the cost function when trained in the Legendre basis using the $\mathcal{T}_1/2$ data transformation. We observe for the training in the Legendre basis that higher temperatures are more amenable to the training procedure and that including more approximate solvers increases accuracy of the final validation loss. Therefore, we see that by executing suitable basis transformations which are supplemented by a multitude of different approximate solvers, the accuracy of the overall predictive quality of the neural network can be improved. We note that production of high quality data on larger values of β requires a larger number of imaginary time slices or Legendre coefficients.

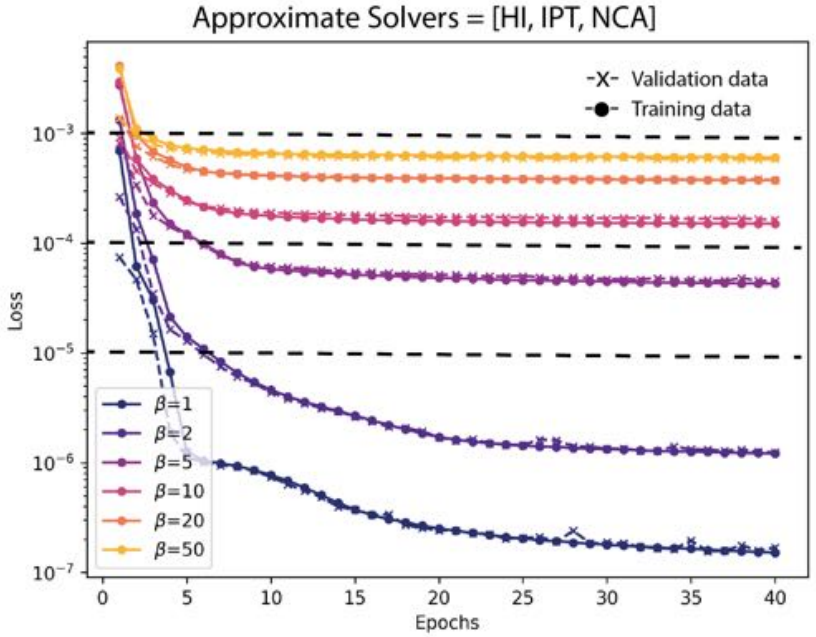


Fig. 10: *Cost function for the Legendre mesh for different inverse temperatures β using as input the approximate solvers Hubbard-I, IPT and NCA.*

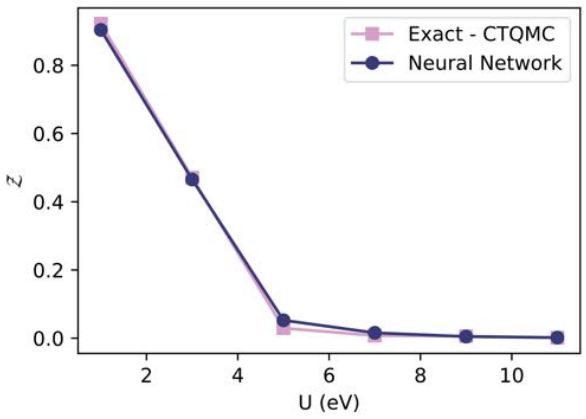


Fig. 11: *CTQMC and Neural Network solvers used for a DMFT prediction of the quasiparticle weight $\mathcal{Z}$ as a function of U at $\beta = 20$ eV^{-1} and $W = 1.0$ eV for the single-band half-filled Hubbard model on a Bethe lattice.*

3.6 A data-driven approach to the Mott transition

The motivation behind developing the data-driven impurity solver is to alleviate DMFT calculations from the intensive computational burden imposed by Exact Diagonalization and Monte Carlo methods. In Fig. 11 we illustrate how the neural network solver, used in a DMFT calculation, can predict the Mott transition at $\beta = 20$ eV^{-1}, $W = 1.0$ eV (at half-filling). This is compared with the equivalent exact CTQMC results.
For each value of U, both solvers are run for 30 iterations to a self-consistent solution. As U is increased the Mott transition occurs at $U/D \approx 4$–6, consistent with other calculations in the literature [1], up to a factor of 2 due to the choice of $D = 2eV$. We emphasize that the network uses approximate solutions as its input during its cycle, for which it predicts the error-free corrected output instantly. By contrast the CTQMC has to be run long enough to mitigate its statistical error bars. This proof-of-concept calculation highlights the power of the data-driven method for a prototypical strongly correlated system, where the solver runs nearly instantaneously, without any significant overheads.

4 Conclusion and code availability

We reviewed neural networks as a data-driven framework that can readily be trained for providing solutions of the Anderson impurity model. This provides an impurity solver capable of capturing the Mott transition using DMFT for the Hubbard model. So far this approach remains robust at higher temperatures, using approximate solutions results in consistently reliable results. We anticipate that improved results at lower temperatures could be attained by extending this method to larger databases or more compact representations of the Green function.
The code discussed in these notes, coined *Data driven Dynamical Mean Field Theory* (D^{3}MFT) is available on GitHub at *http://github.com/zelong-zhao/d3mft*. For the installation of this package, please make sure that you have the Anaconda manager installed on your system, then simply run *./install.sh d3mft*. Once installed, there are different examples which can be run.

References

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

[2] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[3] A.N. Rubtsov, V.V. Savkin, and A.I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005)

[4] A.K. Mitchell and L. Fritz, Phys. Rev. B **88**, 075104 (2013)

[5] C.A. Perroni, H. Ishida, and A. Liebsch, Phys. Rev. B **75**, 045125 (2007)

[6] M Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)

[7] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007)

[8] Z. Li, J.R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015)

[9] A.P. Bartok, M.C. Payne and G. Csanyi, Phys. Rev. Lett. **104**, 136403 (2010)

[10] J. Liu, Y. Qi and L. Fu, Phys. Rev. B **95**, 041101 (2017)

[11] L. Huang and L. Wang, Phys. Rev. B **95**, 035105 (2017)

[12] H. Shen, J. Liu and L. Fu, Phys. Rev. B **97**, 205140 (2018)

[13] P. Broecker, J. Carrasquilla and S. Trebst, Sci. Rep. **7**, 8823 (2017)

[14] L.F. Arsenault, R. Neuberg and A.J. Millis, Inverse Problems **33**, 115007 (2017)

[15] L.F. Arsenault, A. Lopez-Bezanilla, and A.J. Millis, Phys. Rev. B **90**, 155136 (2014)

[16] E.J. Sturm, M.R. Carbone, and R.M. Konik, Phys. Rev. B **103**, 245118 (2021)

[17] N. Walker, S. Kellar, and K.M. Tam, arXiv:2008.12331

[18] H. Hafermann, C. Jung, and A.I. Lichtenstein, Euro. Phys. Lett. **85**, 27007 (2009)

[19] E. Sheridan, C. Rhodes, F. Jamet, I. Rungger, and C. Weber, Phys. Rev. B **104**, 205120 (2021)

[20] H. Jung, R. Silva and M. Han, World Electric Vehicle Journal **9**, 46 (2018)

[21] F Rosenblatt, Psychological Review, **65**, 386 (1958)

[22] C.M. Bishop: *Pattern Recognition and Machine Learning, Information Science and Statistics* (Springer, 2006)

[23] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)

[24] M. Abadi et al.: *TensorFlow: Large-scale machine learning on heterogeneous systems* `http://tensorflow.org` (2015)

5 Quantum Monte Carlo Impurity Solvers

Philipp Werner
University of Fribourg
Chemin du Musée 3, 1700 Fribourg, Switzerland

Contents

1 Quantum impurity models **2**
1.1 Action formulation 3
1.2 Dynamical mean-field theory 4

2 Continuous-time QMC solvers – General formalism **6**

3 Weak-coupling approach **7**
3.1 Sampling 9
3.2 Determinant ratios and fast matrix updates 10
3.3 Measurement of the Green function 11
3.4 Multi-orbital and cluster impurity problems 12

4 Hybridization-expansion approach **12**
4.1 Sampling 14
4.2 Measurement of the Green function 15
4.3 Generalizations – Matrix and Krylov formalisms 17

5 Scaling of the algorithms **20**

6 Electron-boson systems **23**
6.1 Local phonons 23
6.2 Frequency-dependent interactions 26
6.3 Boson distribution function 27

1 Quantum impurity models

A quantum impurity model describes an atom or molecule embedded in some host with which it exchanges electrons or spin. This exchange allows the impurity to make transitions between different quantum states, and these transitions lead to non-trivial dynamical properties. Quantum impurity models play a prominent role, for example, in the theoretical description of dilute metal alloys and in theoretical studies of quantum dots and molecular conductors. These models also appear as an auxiliary problem whose solution yields the dynamical mean-field description of correlated lattice models.
The Hamiltonian of a general impurity model has the form

$$H = H_{\rm loc} + H_{\rm bath} + H_{\rm mix}, \tag{1}$$

where $H_{\rm loc}$ describes the impurity, characterized by a small number of degrees of freedom (typically spin and orbital degrees of freedom denoted by a, b, ...), and $H_{\rm bath}$ describes an infinite reservoir of free electrons, labeled by a continuum of quantum numbers p and a discrete set of quantum numbers ν (typically spin). $H_{\rm mix}$ describes the exchange of electrons between the impurity and the bath in terms of hybridization amplitudes $V^a_{p\nu}$. Denoting the impurity creation operators by $d^\dagger$ and the bath creation operators by $c^\dagger$, the three terms are

$$\begin{aligned} H_{\rm loc} &= \sum_{ab} \epsilon^{ab} d^\dagger_a d_b + \frac{1}{2}\sum_{abcd} U^{abcd} d^\dagger_a d^\dagger_b d_c d_d, & (2)\\ H_{\rm bath} &= \sum_{p\nu} \varepsilon_p c^\dagger_{p\nu} c_{p\nu}, & (3)\\ H_{\rm mix} &= \sum_{pa\nu} \left(V^a_{p\nu}\, d^\dagger_a c_{p\nu} + (V^a_{p\nu})^* c^\dagger_{p\nu} d_a \right). & (4) \end{aligned}$$

In most of the following discussions, we focus on the single-orbital Anderson impurity model, where the local Hamiltonian

$$\begin{aligned} H_{\rm loc} &= H_\mu + H_U, & (5)\\ H_\mu &= -\mu(n_\uparrow + n_\downarrow), & (6)\\ H_U &= U\, n_\uparrow n_\downarrow, & (7) \end{aligned}$$

has a Hilbert space of dimension four. The discrete quantum number labeling the impurity states is the spin σ, $n_\sigma = d^\dagger_\sigma d_\sigma$ is the density operator for impurity electrons with spin σ, and the chemical potential is $\mu = -\epsilon$. The bath and mixing terms are

$$\begin{aligned} H_{\rm bath} &= \sum_{p\sigma} \varepsilon_p c^\dagger_{p\sigma} c_{p\sigma}, & (8)\\ H_{\rm mix} &= \sum_{p\sigma} \left(V_{p\sigma}\, d^\dagger_\sigma c_{p\sigma} + V^*_{p\sigma} c^\dagger_{p\sigma} d_\sigma \right). & (9) \end{aligned}$$

An illustration of the Anderson impurity model is shown in Fig. 1.

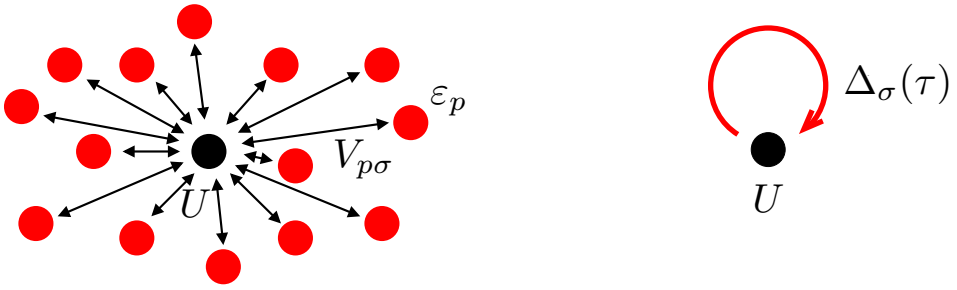


Fig. 1: *Schematic representation of the Anderson impurity model. The left panel illustrates the Hamiltonian representation. Spin up and down electrons on the impurity (black dot) interact with an on-site energy U and hop to a continuum of non-interacting bath levels with energy ε_p. The amplitudes for these transitions are given by the hybridization parameters $V_{p\sigma}$. Right panel: Action representation of the Anderson impurity model, where the bath is replaced by the hybridization function $\Delta_\sigma(\tau)$.*

1.1 Action formulation

For analytical and numerical studies of equilibrium impurity problems, it can be useful to express the partition function and the imaginary-time Green function in terms of the imaginary-time action. By integrating out the bath degrees of freedom in the path integral formalism one obtains the partition function of the Anderson impurity model as

$$Z = \mathrm{Tr}_d\big(\mathcal{T} e^{-S}\big),$$

with the impurity action $S = S_{\mathrm{mix}} + S_{\mathrm{loc}}$ given by

$$S_{\mathrm{mix}} = \sum_\sigma \int_0^\beta d\tau d\tau'\, d^\dagger_\sigma(\tau')\Delta^\sigma(\tau'-\tau)d_\sigma(\tau), \tag{10}$$

$$S_{\mathrm{loc}} = \int_0^\beta d\tau \Big(-\mu\big(n_\uparrow(\tau)+n_\downarrow(\tau)\big) + U\, n_\uparrow(\tau)n_\downarrow(\tau)\Big). \tag{11}$$

$\mathcal{T}$ is the time-ordering operator. The impurity Green function becomes

$$G(\tau) = -\big\langle \mathcal{T} d(\tau)d^\dagger(0)\big\rangle_S = -\frac{1}{Z}\mathrm{Tr}_d\big(\mathcal{T} e^{-S} d(\tau)d^\dagger(0)\big).$$

The imaginary-time and Matsubara-frequency representations are related by

$$G(i\omega_n) = \int_0^\beta d\tau\, e^{i\omega_n\tau} G(\tau), \qquad G(\tau) = \frac{1}{\beta}\sum_n e^{-i\omega_n\tau} G(i\omega_n),$$

where the fermionic Matsubara frequencies are $\omega_n = (2n+1)\pi/\beta$ and $\beta = 1/T$ is the inverse temperature.

The hybridization function $\Delta^\sigma(\tau'-\tau)$ in Eq. (10) represents the amplitude for hopping from the impurity into the bath at time τ and back onto the impurity at time τ'. It is a function of the

bath energies and hybridization amplitudes and is most conveniently expressed in Matsubara frequency space:

$$\Delta^\sigma(i\omega_n) = \sum_p \frac{|V_{p\sigma}|^2}{i\omega_n - \varepsilon_p}. \tag{12}$$

It is also useful to introduce the Green function of the non-interacting impurity, $\mathcal{G}_0$, which is related to the hybridization function by

$$[\mathcal{G}_0^\sigma]^{-1}(i\omega_n) = i\omega_n + \mu - \Delta^\sigma(i\omega_n). \tag{13}$$

1.2 Dynamical mean-field theory

Quantum impurity models are a key ingredient of the dynamical mean-field theory (DMFT), which provides an approximate description of correlated lattice models [1]. The success of DMFT created a demand for accurate and versatile impurity solvers and triggered the development of the continuous-time impurity solvers. These solvers have been discussed in detail in various lecture notes [2], reviews [3] and books [4]. Our presentation here follows closely Chapter 8 in Ref. [4].

In this section, we briefly introduce the DMFT approximation, which maps an interacting lattice model, such as the Hubbard model, onto an effective single-site problem (impurity model) subject to a self-consistency condition for the bath.

The Hubbard model

$$H_{\text{Hubbard}} = -t \sum_{\langle ij\rangle\sigma} \left(d^\dagger_{i\sigma} d_{j\sigma} + d^\dagger_{j\sigma} d_{i\sigma}\right) + U \sum_i n_{i\uparrow} n_{i\downarrow} - \mu \sum_{i\sigma} n_{i\sigma}$$

describes electrons hopping between nearest neighbor sites of some lattice with amplitude t. Two electrons on the same site interact with energy U. The chemical potential term has been added because we will work in the grand canonical ensemble. The noninteracting dispersion ϵ_k is obtained as the Fourier transform of the hopping matrix. For example, in the case of a one-dimensional lattice with lattice spacing a, $\epsilon_k = -2t\cos(ka)$.

Inspired by the Weiss molecular-field theory [1], we focus on one particular site of the lattice and replace the remaining degrees of freedom of the model by a bath of non-interacting levels and a hybridization term that connects the interacting site to the bath. The effective single-site problem thus becomes an Anderson impurity model,[1]

$$H_{\text{imp}} = \sum_{p\sigma} \varepsilon_p c^\dagger_{p\sigma} c_{p\sigma} + \sum_{p\sigma} \left(V_{p\sigma} d^\dagger_\sigma c_{p\sigma} + V^*_{p\sigma} c^\dagger_{p\sigma} d_\sigma\right) + U n_\uparrow n_\downarrow - \mu(n_\uparrow + n_\downarrow). \tag{14}$$

Here, the $d^\dagger$ create electrons on the impurity, $n_\sigma = d^\dagger_\sigma d_\sigma$, and the $c^\dagger_p$ create electrons in bath states labeled by a quantum number p. In this effective single-site model, hoppings from the impurity into the bath and back represent processes in the original Hubbard model where an

[1] In the DMFT context, the bath energy levels ε_p of the impurity model are not directly related to the dispersion of the lattice model, ϵ_k.

electron hops from a given site into the lattice and returns to it after some excursion through the lattice. The hybridization parameters V_p give the amplitudes for such processes.

The DMFT procedure optimizes the parameters ε_p and V_p such that the bath of the Anderson impurity model mimics the lattice environment as closely as possible. If we work with the impurity action, the bath properties are encoded in $\Delta(\tau)$ or $\mathcal{G}_0(\tau)$ and these functions thus play the role of the mean field. It is a *dynamical* mean field, because the hybridization function Δ or Weiss Green function $\mathcal{G}_0$ depends on (imaginary) time or frequency.

The self-consistent solution is constructed in such a way that the impurity Green function $G_{\text{imp}}(i\omega_n)$ reproduces the *local* lattice Green function $G_{\text{loc}}(i\omega_n) \equiv G_{i,i}(i\omega_n)$. In other words, if $G(k, i\omega_n)$ is the momentum-dependent lattice Green function of the Hubbard model, we seek bath parameters and hybridizations such that the *DMFT self-consistency condition*

$$\int (dk)\, G(k, i\omega_n) \equiv G_{\text{imp}}(i\omega_n) \tag{15}$$

is satisfied.[2] The solution of Eq. (15) is obtained by iteration. To define a practical procedure, we have to relate the left-hand-side of Eq. (15) to impurity model quantities. This step involves, as the essential approximation of the DMFT method, a significant simplification of the momentum-dependence of the lattice self-energy [5].

The self-energy describes the effect of interactions on the propagation of electrons. In the non-interacting model, the lattice Green function is $G_0(k, i\omega_n) = \left(i\omega_n + \mu - \epsilon_k\right)^{-1}$, with ϵ_k being the Fourier transform of the hopping matrix. The Green function of the interacting model is $G(k, i\omega_n) = \left(i\omega_n + \mu - \epsilon_k - \Sigma(k, i\omega_n)\right)^{-1}$ with $\Sigma(k, i\omega_n)$ the lattice self-energy. Therefore

$$\Sigma(k, i\omega_n) = G_0^{-1}(k, i\omega_n) - G^{-1}(k, i\omega_n).$$

Similarly, we obtain the impurity self-energy

$$\Sigma_{\text{imp}}(i\omega_n) = \mathcal{G}_0^{-1}(i\omega_n) - G_{\text{imp}}^{-1}(i\omega_n),$$

with $\mathcal{G}_0^{-1}$ defined in Eq. (13). The DMFT approximation is the identification of the lattice self-energy with the momentum-independent impurity self-energy,

$$\Sigma(k, i\omega_n) \approx \Sigma_{\text{imp}}(i\omega_n).$$

This approximation allows us to rewrite the self-consistency equation (15) as

$$\int (dk)\, \left(i\omega_n + \mu - \epsilon_k - \Sigma_{\text{imp}}(i\omega_n)\right)^{-1} \equiv G_{\text{imp}}(i\omega_n). \tag{16}$$

Since both $G_{\text{imp}}(i\omega_n)$ and $\Sigma_{\text{imp}}(i\omega_n)$ are determined by the impurity model parameters ε_p and V_p (or the function $\Delta(\tau)$ or $\mathcal{G}_0(\tau)$), Eq. (16) defines a self-consistency condition for these parameters (or functions).

We now formulate the self-consistency loop for the Weiss Green function $\mathcal{G}_0(i\omega_n)$. Starting from an arbitrary initial $\mathcal{G}_0(i\omega_n)$, for example, the local Green function of the noninteracting lattice model, we iterate the following steps until convergence:

[2] $\int (dk)$ denotes a normalized integral over the Brillouin zone.

1. Solve the impurity problem, that is, compute the impurity Green function $G_{\rm imp}(i\omega_n)$ for the given $\mathcal{G}_0(i\omega_n)$,

2. Extract the self-energy of the impurity model: $\Sigma_{\rm imp}(i\omega_n) = \mathcal{G}_0^{-1}(i\omega_n) - G_{\rm imp}^{-1}(i\omega_n)$,

3. Identify the lattice self-energy with the impurity self-energy, $\Sigma(k, i\omega_n) = \Sigma_{\rm imp}(i\omega_n)$ (DMFT approximation), and compute the local lattice Green function:

$$G_{\rm loc}(i\omega_n) = \int (dk) \big(i\omega_n + \mu - \epsilon_k - \Sigma_{\rm imp}(i\omega_n) \big)^{-1},$$

4. Apply the DMFT self-consistency condition, $G_{\rm loc}(i\omega_n) = G_{\rm imp}(i\omega_n)$, and use it to define a new Weiss Green function $\mathcal{G}_0^{-1}(i\omega_n) = G_{\rm loc}^{-1}(i\omega_n) + \Sigma_{\rm imp}(i\omega_n)$.

The computationally expensive step is the solution of the impurity problem (Step 1). When the loop converges, the bath contains information about the lattice (through the density of states), and about the phase (metal, Mott insulator, antiferromagnetic insulator, ...). The impurity, which exchanges electrons with the bath, thus feels, at least to some extent, as if it were a site of the lattice.

2 Continuous-time QMC solvers – General formalism

Quantum impurity models are (0+1)-dimensional quantum field theories and as such are computationally much more tractable than interacting lattice models. The main objective is computing the impurity Green function

$$G(\tau) = -\langle \mathcal{T} d(\tau) d^\dagger(0) \rangle = -\frac{1}{Z} \mathrm{Tr} \Big(e^{-(\beta-\tau)H} d\, e^{-\tau H} d^\dagger \Big), \tag{17}$$

where $Z = \mathrm{Tr}\, e^{-\beta H}$ is the impurity model partition function, β the inverse temperature, $\mathcal{T}$ is the (imaginary) time-ordering operator, and $\mathrm{Tr} = \mathrm{Tr}_d \mathrm{Tr}_c$ the trace over the impurity and bath states. In the last expression we assumed that $0 \le \tau < \beta$.

Continuous-time Monte Carlo algorithms expand the partition function into a series of "diagrams" and stochastically sample these diagrams [3]. We represent the partition function as a sum (or more precisely as an integral) over configurations C with weight w_C,

$$Z = \sum_C w_C, \tag{18}$$

and implement a random walk $C_1 \to C_2 \to C_3 \to \cdots$ in configuration space in such a way that ergodicity and detailed balance are satisfied. Using sign-weighted averages, the impurity Green function can be estimated from a finite number M of measurements as

$$G = \sum_C \frac{w_C\, G_C}{Z} = \frac{\sum_C |w_C| \,\mathrm{sign}_C\, G_C}{\sum_C |w_C| \,\mathrm{sign}_C} \approx \frac{\sum_{i=1}^M \mathrm{sign}_{C_i} G_{C_i}}{\sum_{i=1}^M \mathrm{sign}_{C_i}} \equiv \frac{\langle \mathrm{sign} \cdot G \rangle_{\rm MC}}{\langle \mathrm{sign} \rangle_{\rm MC}}. \tag{19}$$

To derive the general framework for the continuous-time solvers it is useful to express the partition function as an imaginary-time ordered exponential in an *interaction representation*. To do this, we split the Hamiltonian into two parts, $H = H_1 + H_2$, and define the imaginary-time dependent operators in the interaction representation as $O(\tau) = e^{\tau H_1} O e^{-\tau H_1}$. In this representation, the partition function becomes $Z = \mathrm{Tr}\big(e^{-\beta H_1}\mathcal{T} e^{-\int_0^\beta d\tau H_2(\tau)}\big)$.[3]
Next, we expand the time-ordered exponential into a power series,

$$Z = \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \cdots \int_{\tau_{n-1}}^\beta d\tau_n \,\mathrm{Tr}\,\Big(e^{-(\beta-\tau_n)H_1}(-H_2)\cdots e^{-(\tau_2-\tau_1)H_1}(-H_2)e^{-\tau_1 H_1}\Big). \qquad (20)$$

This yields a representation of the partition function of the form (18), namely, as an infinite sum over the weights of certain configurations. The configurations are collections of time-points on the imaginary-time interval: $C = \{\tau_1, \ldots, \tau_n\}$, $n = 0, 1, \ldots$, where we assume the imaginary-time ordering $\tau_i < \tau_{i+1}$ and the restriction $\tau_i \in [0, \beta)$. The expression for the Monte Carlo weights is

$$w_C = \mathrm{Tr}\Big(e^{-(\beta-\tau_n)H_1}(-H_2)\cdots e^{-(\tau_2-\tau_1)H_1}(-H_2)e^{-\tau_1 H_1}\Big)(d\tau)^n. \qquad (21)$$

There are two complementary continuous-time Monte Carlo techniques: (i) the *weak-coupling approach*, which scales favorably with system size (that is, the number of correlated sites or orbitals in the impurity model) and allows the efficient simulation of relatively large impurity clusters with simple interactions, and (ii) the *hybridization-expansion approach*, which can handle impurity models with strong interactions among multiple orbitals. For simplicity, we continue to focus on the single-orbital Anderson impurity model defined in Eqs. (6)–(9). In this case, the weak-coupling continuous-time Monte Carlo approach expands Z in powers of the interaction U in an interaction representation where the imaginary-time evolution is determined by the *quadratic* part $H_\mu + H_{\mathrm{bath}} + H_{\mathrm{mix}}$ of the Hamiltonian. The complementary hybridization-expansion approach expands Z in powers of the impurity-bath hybridization term H_{mix} in an interaction representation where the imaginary-time evolution is determined by the *local* part $H_\mu + H_U + H_{\mathrm{bath}}$ of the Hamiltonian. The details of how the weights (21) are sampled and how the observables are measured depend on the specific continuous-time method.

3 Weak-coupling approach

The weak-coupling continuous-time impurity solver [6] expands the partition function in powers of $H_2{=}H_U$.[4] Equation (21) then gives the weight of a configuration of n *interaction vertices*. Since $H_1 = H - H_2 = H_\mu + H_{\mathrm{bath}} + H_{\mathrm{mix}}$ is quadratic, we can use Wick's theorem to evaluate the trace. The result is a product of two determinants of $n{\times}n$ matrices (one for each spin). The

[3] We can understand this formula by defining the operator $A(\beta) = e^{\beta H_1} e^{-\beta H}$ and writing the partition function as $Z = \mathrm{Tr}(e^{-\beta H_1} A(\beta))$. The operator $A(\beta)$ satisfies $dA/d\beta = e^{\beta H_1}(H_1 - H)e^{-\beta H} = -H_2(\beta)A(\beta)$, the solution of which is $A(\beta) = \mathcal{T}\exp\big(-\int_0^\beta d\tau H_2(\tau)\big)$.

[4] A related algorithm, based on an expansion in powers of $H_U - K/\beta$ (with K some non-zero constant), is the *continuous-time auxiliary field method* [7].

elements of these matrices are the Weiss Green functions $\mathcal{G}_0^\sigma$ for the time intervals defined by the vertex positions

$$\frac{w_C}{Z_0} = (-Ud\tau)^n \frac{1}{Z_0}\mathrm{Tr}\left(e^{-(\beta-\tau_n)H_1} n_\uparrow n_\downarrow \cdots e^{-(\tau_2-\tau_1)H_1} n_\uparrow n_\downarrow e^{-\tau_1 H_1}\right) = (-Ud\tau)^n \prod_\sigma \det M_\sigma^{-1},$$

where

$$\left(M_\sigma^{-1}\right)_{ij} = \mathcal{G}_0^\sigma(\tau_i - \tau_j) \quad \text{with} \quad \mathcal{G}_0^\sigma(\tau) = -\frac{1}{Z_0}\mathrm{Tr}\left(e^{-\beta H_1}\mathcal{T} d(\tau) d^\dagger(0)\right),$$

and $Z_0 = \mathrm{Tr}\, e^{-\beta H_1}$ is the partition function of the non-interacting model.[5] For the diagonal elements, we adopt the convention $(M_\sigma^{-1})_{ii} = \mathcal{G}_0^\sigma(0^-)$.

At this point, one notices a potential sign problem. In the paramagnetic phase, where $\mathcal{G}_0^\uparrow = \mathcal{G}_0^\downarrow$, the product of determinants is positive, which means that for a repulsive interaction ($U > 0$) odd perturbation orders yield negative weights. Except in the particle-hole symmetric case, where odd perturbation orders vanish, these odd order configurations cause a sign problem. Fortunately, we can solve this sign problem by shifting the chemical potentials for up and down spins in an appropriate way [6]. To do so, we rewrite the interaction term as [8]

$$H_U = \frac{U}{2}\sum_s \prod_\sigma \left(n_\sigma - \alpha_\sigma(s)\right) + \frac{U}{2}(n_\uparrow + n_\downarrow) + U\left(\left(\frac{1}{2}+\delta\right)^2 - \frac{1}{4}\right), \tag{22}$$

with

$$\alpha_\sigma(s) = \frac{1}{2} + \sigma s\left(\frac{1}{2}+\delta\right). \tag{23}$$

Here, δ is some constant and $s = \pm 1$ is an auxiliary Ising variable. This construction is not a Hubbard-Stratonovich transformation, but simply a shift in the zero of energy. The constant $U\left((\frac{1}{2}+\delta)^2 - \frac{1}{4}\right)$ in Eq. (22) is irrelevant and will be ignored in the following. We absorb the contribution $\frac{1}{2}U(n_\uparrow + n_\downarrow)$ into the non-interacting Green function by shifting the chemical potential as $\mu \to \mu - \frac{1}{2}U$. Explicitly, the Weiss Green function is redefined as[6]

$$\left(\mathcal{G}_0^\sigma\right)^{-1} = i\omega_n + \mu - \Delta^\sigma \;\to\; \left(\tilde{\mathcal{G}}_0^\sigma\right)^{-1} = i\omega_n + \mu - \tfrac{1}{2}U - \Delta^\sigma.$$

The introduction of an Ising variable s_i at each vertex position τ_i enlarges the configuration space exponentially. A configuration C now corresponds to a collection of auxiliary spin variables defined on the imaginary-time interval: $C = \{(\tau_1, s_1), (\tau_2, s_2), \ldots, (\tau_n, s_n)\}$. The weight of these configurations is

$$w_C = \tilde{Z}_0(-Ud\tau/2)^n \prod_\sigma \det \tilde{M}_\sigma^{-1}, \tag{24}$$

where

$$\left(\tilde{M}_\sigma^{-1}\right)_{ij} = \tilde{\mathcal{G}}_0^\sigma(\tau_i - \tau_j) - \alpha_\sigma(s_i)\delta_{ij}. \tag{25}$$

[5] We note that in the DMFT framework discussed in Section 1.2, the function $\mathcal{G}_0^\sigma$ is determined directly by the self-consistency loop, without reference to a Hamiltonian. For the purpose of the present discussion, we may however assume that we know H_{bath} and H_{mix} terms whose parameters yield $\mathcal{G}_0^\sigma$ through Eqs. (12) and (13).

[6] In a DMFT calculation, this means that the shifted chemical potential is used within the self-consistency loop.

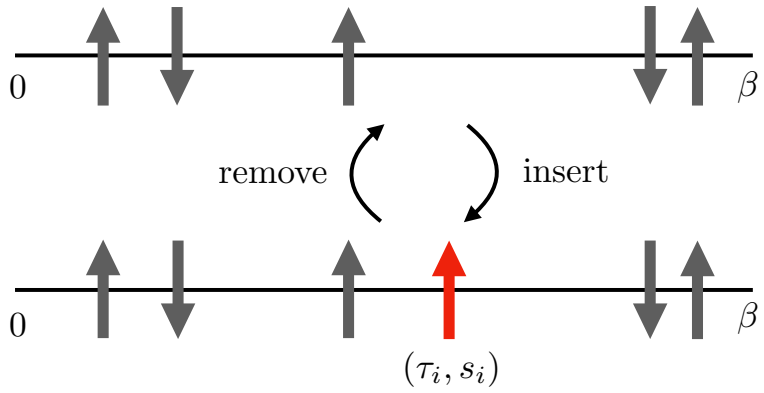


Fig. 2: *Local update in the weak-coupling method. The horizontal line represents the imaginary-time interval* $[0,\beta)$*. We increase the perturbation order by adding an auxiliary spin with random orientation at a random time and decrease it by removing a randomly chosen auxiliary spin.*

3.1 Sampling

For ergodicity it is sufficient that the sampling inserts the auxiliary spins with random orientation at random times and removes randomly chosen spins. Detailed balance requires that the probability $p(C\to C')$ to move from configuration C to C' satisfies $w(C)p(C\to C') = w(C')p(C'\to C)$. Splitting $p(C\to C') = p^{\text{prop}}(C\to C')p^{\text{acc}}(C\to C')$ into a proposal and acceptance probability, and using the Metropolis-Hastings algorithm [4], we have

$$p^{\text{acc}}(C\to C') = \min\big(1, \mathcal{R}(C\to C')\big),$$

where

$$\mathcal{R}(C\to C') = \frac{w(C')\,p^{\text{prop}}(C'\to C)}{w(C)\,p^{\text{prop}}(C\to C')}$$

and Eq. (24) is used to compute the ratio of the weights. To complete the description of the sampling we need to specify proposal probabilities for the insertion and removal of an auxiliary spin. A simple and reasonable procedure is illustrated in Fig. 2. For the insertion, we pick a random time in $[0,\beta)$ and a random orientation for the new spin, while for the removal, we simply pick a random spin. The corresponding proposal probabilities are

$$p^{\text{prop}}(n\to n{+}1) = \tfrac{1}{2}(d\tau/\beta), \quad p^{\text{prop}}(n{+}1\to n) = 1/(n{+}1). \tag{26}$$

The first step is choosing with equal probability whether we insert or remove. If we insert, then we are going from a configuration with n spins to a configuration with $n{+}1$ spins, and from Eq. (24) and the above choices for p^{prop}, the acceptance probability becomes $p^{\text{acc}}(n\to n{+}1) = \min\big(1, \mathcal{R}_{\text{insert}}(n\to n{+}1)\big)$ with

$$\mathcal{R}_{\text{insert}}(n\to n{+}1) = \frac{-\beta U}{n{+}1}\prod_\sigma \frac{\det\big(\tilde{M}_\sigma^{(n+1)}\big)^{-1}}{\det\big(\tilde{M}_\sigma^{(n)}\big)^{-1}}. \tag{27}$$

The acceptance probability for the removal follows from

$$\mathcal{R}_{\text{remove}}(n{+}1\to n) = 1/\mathcal{R}_{\text{insert}}(n\to n{+}1). \tag{28}$$

3.2 Determinant ratios and fast matrix updates

From Eq. (27), we see that each update requires the calculation of a ratio of two determinants. At first sight, one might think that for a matrix of size $n\times n$ this is an $\mathcal{O}(n^3)$ operation. However, each insertion or removal of a vertex (or spin) merely changes one row and one column of the matrix M_σ^{-1} (or $\tilde{M}_\sigma^{-1}$).[7] It is thus possible to evaluate this ratio in a time $\mathcal{O}(n^2)$ for insertion and $\mathcal{O}(1)$ for removal [3].

We first note that the objects which are stored and manipulated, besides the lists of the times $\{\tau_i\}$ (or times and spins $\{(\tau_i, s_i)\}$), are the matrices $M_\sigma = (\mathcal{G}_0^\sigma)^{-1}$, not $M_\sigma^{-1} = \mathcal{G}_0^\sigma$. Inserting a vertex (or auxiliary spin) adds a new row and column to M_σ^{-1}. We imagine inserting this row and column on the border of the given matrix and write the resulting matrix in a block matrix form (omitting the σ index for simplicity):

$$\left(M^{(n+1)}\right)^{-1} = \begin{pmatrix} \left(M^{(n)}\right)^{-1} & Q \\ R & S \end{pmatrix}.$$

The analogous blocks of the M matrix are defined as

$$M^{(n+1)} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix}. \tag{29}$$

Here Q, R, and S are $n\times 1$, $1\times n$, and 1×1 matrices which contain the functions $\mathcal{G}_0$ evaluated at time intervals determined by the position of the new vertex (spin). They can be easily computed. We want to find $\tilde{P}$, $\tilde{Q}$, $\tilde{R}$, and $\tilde{S}$, and the ratio of determinants. Using the expression for the block inversion of a matrix and for the determinant of a block matrix, the determinant ratio needed for the acceptance probability becomes

$$\frac{\det\left(M^{(n+1)}\right)^{-1}}{\det\left(M^{(n)}\right)^{-1}} = \det\left(S - RM^{(n)}Q\right) = S - RM^{(n)}Q\,. \tag{30}$$

Because we store $M^{(n)}$, computing the acceptance probability of an insertion move is just an $\mathcal{O}(n^2)$ operation. If the move is accepted, the new matrix $M^{(n+1)}$ can be computed from $M^{(n)}$, Q, R, and S, also in a time $\mathcal{O}(n^2)$:

$$\tilde{S} = \left(S - (R)\,(M^{(n)}\,Q)\right)^{-1}, \tag{31}$$

$$\tilde{Q} = -(M^{(n)}Q)\,\tilde{S}, \tag{32}$$

$$\tilde{R} = -\tilde{S}\,(R\,M^{(n)}), \tag{33}$$

$$\tilde{P} = M^{(n)} + (M^{(n)}Q)\,\tilde{S}\,(RM^{(n)}). \tag{34}$$

In the case of removing a spin we imagine removing a bordering row and column. It follows from Eqs. (30) and (31) that

$$\frac{\det\left(M^{(n)}\right)^{-1}}{\det\left(M^{(n+1)}\right)^{-1}} = \det\tilde{S} = \tilde{S}. \tag{35}$$

[7] In the following, we write the formulas without the tildes, that is, for the sampling of interaction vertices. For the algorithm with auxiliary spins, it suffices to replace $M\to\tilde{M}$ and $\mathcal{G}_0\to\tilde{\mathcal{G}}_0$.

$\tilde{S}$ is just a 1×1 matrix so its determinant is trivial to compute. The above formulas also imply that the elements of the reduced matrix are

$$M^{(n)} \;=\; \tilde{P} - (\tilde{Q})(\tilde{R})/\tilde{S}. \tag{36}$$

The calculation of the removal probability is thus $\mathcal{O}(1)$, while the calculation of the new $M^{(n)}$ matrix is $\mathcal{O}(n^2)$.

3.3 Measurement of the Green function

To compute the contribution of a configuration C to the Green function, $G^\sigma_C(\tau)$, we insert in the right-hand side of Eq. (21) a creation operator $d^\dagger$ at time 0 and an annihilation operator d at time τ and divide by w_C. Wick's theorem and Eq. (30) then lead to the expression [6]

$$G^\sigma_C(\tau) \;=\; \mathcal{G}^\sigma_0(\tau) - \sum_k \mathcal{G}^\sigma_0(\tau-\tau_k) \sum_l \big(M_\sigma\big)_{kl}\, \mathcal{G}^\sigma_0(\tau_l). \tag{37}$$

The estimate for the impurity Green function for a given imaginary-time then follows from Eq. (19). To avoid unnecessary and time-consuming summations during the Monte Carlo simulation (evaluation of Eq. (37) for many τ-values), we accumulate the quantity [7]

$$S_\sigma(\tilde{\tau}) \equiv \sum_k \delta(\tilde{\tau}-\tau_k) \sum_l \big(M_\sigma\big)_{kl}\, \mathcal{G}^\sigma_0(\tau_l),$$

by binning the time points $\tilde{\tau}$ on a fine grid. After the simulation is finished, we compute the Green function as[8]

$$G^\sigma(\tau) = \mathcal{G}^\sigma_0(\tau) - \int_0^\beta d\tilde{\tau}\, \mathcal{G}^\sigma_0(\tau-\tilde{\tau}) \big\langle S_\sigma(\tilde{\tau}) \big\rangle_{\text{MC}}. \tag{38}$$

It is also possible to measure the Matsubara components of the Green function directly. Using the imaginary-time translational invariance of the Green functions, one finds

$$G^\sigma_C(i\omega_n) \;=\; \mathcal{G}^\sigma_0(i\omega_n) - \mathcal{G}^\sigma_0(i\omega_n) \sum_{kl} \frac{1}{\beta} e^{i\omega_n(\tau_k-\tau_l)} \big(M_\sigma\big)_{kl}\, \mathcal{G}^\sigma_0(i\omega_n),$$

so that

$$G^\sigma(i\omega_n) = \mathcal{G}^\sigma_0(i\omega_n) - \frac{1}{\beta}\big(\mathcal{G}^\sigma_0(i\omega_n)\big)^2 \left\langle \sum_{kl} e^{i\omega_n(\tau_k-\tau_l)} \big(M_\sigma\big)_{kl} \right\rangle_{\text{MC}}. \tag{39}$$

We note that because the Weiss Green function has the high-frequency behavior $\mathcal{G}_0(i\omega_n) \sim 1/i\omega_n$, the measured impurity Green function automatically inherits the correct high-frequency tail.

[8] Comparison of this equation with the Dyson equation $G = \mathcal{G}_0 + \mathcal{G}_0 \star \Sigma \star G$ (where the $\star$ symbol denotes a convolution in imaginary time) shows that his procedure amounts to measuring $-\Sigma \star G$.

3.4 Multi-orbital and cluster impurity problems

The generalization of the weak-coupling method to impurity clusters is straightforward. All we have to do is to add a site index to the interaction vertices (or auxiliary Ising spin variables) and sample the vertices (auxiliary spins) on a family of n_{sites} imaginary-time intervals.
General four-Fermion terms as in Eq. (2) are, at least in principle, also easily dealt with. We simply expand the partition function in powers of the interactions U^{abcd}. The trace over the impurity and bath degrees of freedom again yields a determinant of a matrix whose order equals the total perturbation order. In general there is a sign problem. To reduce the sign problem, it is advantageous to introduce auxiliary fields α and replace

$$\frac{1}{2}\sum_{abcd} U^{abcd}\, d_a^\dagger d_b^\dagger d_c d_d \rightarrow -\frac{1}{2}\sum_{abcd} U^{abcd}\big(d_a^\dagger d_c - \alpha_{ac}\big)\big(d_b^\dagger d_d - \alpha_{bd}\big),$$

with an appropriate shift in the quadratic part of the Hamiltonian. However, in general, it is not possible to completely eliminate the sign problem by a suitable choice of α parameters. Furthermore, since the number of interaction terms grows like $\mathcal{O}(n_{\text{orbitals}}^4)$, the computational cost rapidly escalates. In practice, the approach discussed in the following section is a more suitable approach for single-site multi-orbital impurity problems with general interactions.

4 Hybridization-expansion approach

While the Monte Carlo weights in the weak-coupling method are expressed in terms of the Weiss Green function $\mathcal{G}_0$, the hybridization-expansion method, which is in many ways complementary to the weak-coupling approach, naturally involves the hybridization function Δ. It follows from Eq. (13) that the Weiss Green function $\mathcal{G}_0$ and hybridization function Δ contain the same information, and the DMFT procedure sketched in Sec. 1.2 could be just as well written as a self-consistency loop fixing the hybridization function Δ.
The hybridization-expansion approach [9] is based on an expansion of the partition function in powers of the impurity-bath hybridization term. Here, we decompose the Hamiltonian as $H_2 = H_{\text{mix}}$ and $H_1 = H - H_2 = H_\mu + H_U + H_{\text{bath}}$. Since $H_2 \equiv H_2^{d^\dagger} + H_2^d = \sum_{p\sigma} V_{p\sigma}\, d_\sigma^\dagger c_{p\sigma} + \sum_{p\sigma} V_{p\sigma}^*\, c_{p\sigma}^\dagger d_\sigma$ has two terms, corresponding to electrons hopping from the bath to the impurity and from the impurity back to the bath, only even perturbation orders contribute to Eq. (20). Furthermore, at perturbation order $2n$, only the $(2n)!/(n!)^2$ terms corresponding to n creation operators $d^\dagger$ and n annihilation operators d contribute. We therefore write the partition function as a sum over configurations $\{\tau_1, \ldots, \tau_n; \tau_1', \ldots, \tau_n'\}$ that are collections of imaginary-time points corresponding to these n annihilation and n creation operators:

$$Z = \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \cdots \int_{\tau_{n-1}}^\beta d\tau_n \int_0^\beta d\tau_1' \cdots \int_{\tau_{n-1}'}^\beta d\tau_n' \,\text{Tr}\Big(e^{-\beta H_1}\mathcal{T} H_2^d(\tau_n) H_2^{d^\dagger}(\tau_n') \cdots H_2^d(\tau_1) H_2^{d^\dagger}(\tau_1')\Big). \tag{40}$$

Since the imaginary-time evolution operator $e^{-\tau H_1}$ does not rotate the spin in the case of the Anderson impurity model, the configurations must contain an equal number of creation and

annihilation operators for each spin. Taking this additional constraint into account and using the explicit expressions for H_2^d and $H_2^{d^\dagger}$, we find

$$\begin{aligned} Z &= Z_{\rm bath} \sum_{\{n_\sigma\}} \prod_\sigma \int_0^\beta d\tau_1^\sigma \cdots \int_{\tau_{n_\sigma-1}^\sigma}^\beta d\tau_{n_\sigma}^\sigma \int_0^\beta d\tau_1'^\sigma \ldots \int_{\tau_{n_\sigma-1}'^\sigma}^\beta d\tau_{n_\sigma}'^\sigma \\ &\quad \times {\rm Tr}_d \Big(e^{-\beta H_{\rm loc}} \mathcal{T} \prod_\sigma d_\sigma(\tau_{n_\sigma}^\sigma) d_\sigma^\dagger(\tau_{n_\sigma}'^\sigma) \ldots d_\sigma(\tau_1^\sigma) d_\sigma^\dagger(\tau_1'^\sigma) \Big) \\ &\quad \times \frac{1}{Z_{\rm bath}} {\rm Tr}_c \Big(e^{-\beta H_{\rm bath}} \mathcal{T} \prod_\sigma \sum_{p_1 \ldots p_{n_\sigma}} \sum_{p_1' \ldots p_{n_\sigma}'} V_{p_1\sigma}^* V_{p_1'\sigma} \cdots V_{p_{n_\sigma}\sigma}^* V_{p_{n_\sigma}'\sigma} \\ &\qquad c_{p_{n_\sigma}\sigma}^\dagger(\tau_{n_\sigma}^\sigma) c_{p_{n_\sigma}'\sigma}(\tau_{n_\sigma}'^\sigma) \ldots c_{p_1\sigma}^\dagger(\tau_1^\sigma) c_{p_1'\sigma}(\tau_1'^\sigma) \Big), \end{aligned}$$

where to separate the d and c operators we used the fact that H_1 does not mix the impurity and the bath. The local Hamiltonian $H_{\rm loc}$ is defined in Eq. (5) and $Z_{\rm bath} = {\rm Tr}_c\, e^{-\beta H_{\rm bath}}$.
Introducing the β-antiperiodic hybridization function (12), which in the time-domain reads

$$\Delta^\sigma(\tau) = \sum_p \frac{|V_{p\sigma}|^2}{e^{\varepsilon_p \beta} + 1} \begin{cases} -e^{-\varepsilon_p(\tau-\beta)} & 0 < \tau < \beta \\ e^{-\varepsilon_p \tau} & -\beta < \tau < 0 \end{cases},$$

the trace over the bath states can be expressed as

$$\begin{aligned} &\frac{1}{Z_{\rm bath}} {\rm Tr}_c \Big(e^{-\beta H_{\rm bath}} \mathcal{T} \prod_\sigma \sum_{p_1 \ldots p_{n_\sigma}} \sum_{p_1' \ldots p_{n_\sigma}'} V_{p_1\sigma}^* V_{p_1'\sigma} \cdots V_{p_{n_\sigma}\sigma}^* V_{p_{n_\sigma}'\sigma} \\ &\qquad c_{p_{n_\sigma}\sigma}^\dagger(\tau_{n_\sigma}^\sigma) c_{p_{n_\sigma}'\sigma}(\tau_{n_\sigma}'^\sigma) \cdots c_{p_1\sigma}^\dagger(\tau_1^\sigma) c_{p_1'\sigma}(\tau_1'^\sigma) \Big) = \prod_\sigma \det M_\sigma^{-1}, \end{aligned}$$

where M_σ^{-1} is the $(n_\sigma \times n_\sigma)$ matrix with elements

$$\big(M_\sigma^{-1}\big)_{ij} = \Delta^\sigma(\tau_i'^\sigma - \tau_j^\sigma).$$

In the hybridization expansion approach, the configuration space consists of all sequences $C = \{\tau_1^\uparrow, \ldots, \tau_{n_\uparrow}^\uparrow; \tau_1'^\uparrow, \ldots, \tau_{n_\uparrow}'^\uparrow | \tau_1^\downarrow, \ldots, \tau_{n_\downarrow}^\downarrow; \tau_1'^\downarrow, \ldots, \tau_{n_\downarrow}'^\downarrow\}$ of $n_\uparrow$ creation and annihilation operators for spin up ($n_\uparrow = 0, 1, \ldots$) and $n_\downarrow$ creation and annihilation operators for spin down ($n_\downarrow = 0, 1, \ldots$). The weight of such a configuration is

$$w_C = Z_{\rm bath} {\rm Tr}_d \Big(e^{-\beta H_{\rm loc}} \mathcal{T} \prod_\sigma d_\sigma(\tau_{n_\sigma}^\sigma) d_\sigma^\dagger(\tau_{n_\sigma}'^\sigma) \cdots d_\sigma(\tau_1^\sigma) d_\sigma^\dagger(\tau_1'^\sigma) \Big) \prod_\sigma \det M_\sigma^{-1} (d\tau)^{2n_\sigma}. \quad (41)$$

The trace factor represents the contribution of the impurity, which fluctuates between different quantum states as electrons hop in and out. The determinants sum up all bath evolutions which are compatible with the given sequence of transitions.
To evaluate the trace factor, we may for example use the eigenbasis of $H_{\rm loc}$. In this basis, the imaginary-time evolution operator $e^{-\tau H_{\rm loc}}$ is diagonal while the operators d_σ and $d_\sigma^\dagger$ produce transitions between eigenstates with amplitude ± 1. Because the time evolution does not flip the electron spin, the creation and annihilation operators for a given spin alternate. This observation

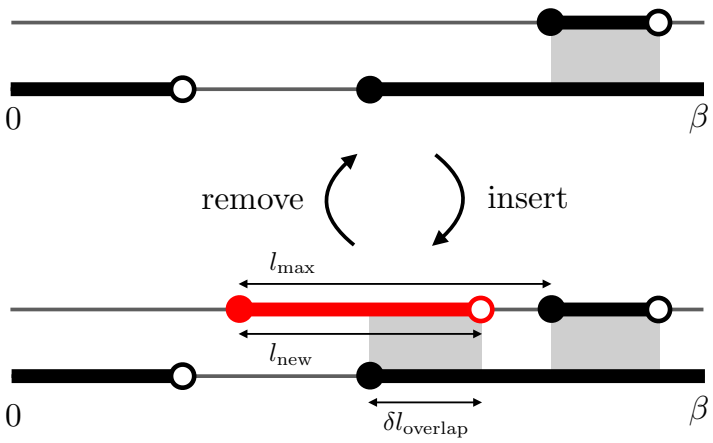


Fig. 3: *Local update in the segment picture. The two segment configurations correspond to spin up and down electrons. Each segment depicts a time interval in which an electron of the corresponding spin resides on the impurity. The segment end points are the locations of the operators $d^\dagger$ (full circles) and d (empty circles). We increase the perturbation order by adding a segment or anti-segment of random length for random spin and decrease it by removing a randomly chosen segment or anti-segment.*

allows us to separate the operators for spin up from those for spin down and to depict the time evolution by a *collection of segments* with each segment representing an imaginary-time interval in which an electron of spin up or down resides on the impurity (Fig. 3). We call an unoccupied time-interval between two segments an "anti-segment".
At each time, the eigenstate of the impurity follows immediately from the segment representation, and the trace factor becomes

$$\mathrm{Tr}_d\Big(e^{-\beta H_{\rm loc}}\mathcal{T}\prod_\sigma d_\sigma(\tau^\sigma_{n_\sigma})d^\dagger_\sigma(\tau'^\sigma_{n_\sigma})\cdots d_\sigma(\tau^\sigma_1)d^\dagger_\sigma(\tau'^\sigma_1)\Big) = \mathcal{S}\exp\Big(\mu(l_\uparrow+l_\downarrow) - U l_{\rm overlap}\Big), \quad (42)$$

with $\mathcal{S}$ being a permutation sign, l_σ the total "length" of the segments for spin σ, and $l_{\rm overlap}$ the total length of the overlap between spin-up and spin-down segments. The lower panel of Fig. 3 shows a configuration with two segments for spin up and one segment for spin down (note the periodic boundary conditions). The time intervals where segments overlap, indicated by gray rectangles, correspond to a doubly occupied impurity and cost a repulsion energy U.

4.1 Sampling

For ergodicity, it is sufficient to insert and remove pairs of creation and annihilation operators (segments or anti-segments) for spin up and down. One possible strategy for inserting a segment is the following: We select a random time in $[0,\beta)$ for the creation operator. If it falls on an existing segment, the impurity is already occupied and the move is rejected. If it falls on an empty space, we compute $l_{\rm max}$, the length from this selected time to the next segment (in the direction of increasing τ, taking into account the periodic boundary conditions).[9] Then we

[9] If there are no segments for the given spin, $l_{\rm max} = \beta$.

choose the position of the new annihilation operator randomly in this interval of length $l_{\max}$ (Fig. 3). If in the inverse procedure we propose to remove a randomly chosen segment for this spin, then the proposal probabilities for the insertion and removal are

$$p^{\text{prop}}(n_\sigma \to n_\sigma{+}1) = \frac{d\tau}{\beta}\frac{d\tau}{l_{\max}}, \quad p^{\text{prop}}(n_\sigma{+}1 \to n_\sigma) = \frac{1}{n_\sigma + 1}.$$

The acceptance probability for the insertion of a segment becomes $p^{\text{acc}}(n_\sigma \to n_\sigma{+}1) = \min\bigl(1, \mathcal{R}_{\text{insert}}(n_\sigma \to n_\sigma{+}1)\bigr)$, with

$$\mathcal{R}_{\text{insert}}(n_\sigma \to n_\sigma{+}1) = \frac{\beta l_{\max}}{n_\sigma + 1}\, e^{\mu l_{\text{new}} - U\delta l_{\text{overlap}}} \frac{\det\left(M_\sigma^{(n_\sigma+1)}\right)^{-1}}{\det\left(M_\sigma^{(n_\sigma)}\right)^{-1}}, \tag{43}$$

while the acceptance probability for a removal is obtained from

$$\mathcal{R}_{\text{remove}}(n_\sigma{+}1 \to n_\sigma) = 1/\mathcal{R}_{\text{insert}}(n_\sigma \to n_\sigma{+}1). \tag{44}$$

Here, l_{new} is the length of the new segment, and $\delta l_{\text{overlap}}$ is the change in the overlap (see Fig. 3). We compute the ratio of determinants using the fast update formulas discussed in Section 3.2.

4.2 Measurement of the Green function

The strategy is to create configurations which contribute to the Green function measurement by decoupling the bath from a given pair of creation and annihilation operators in C. We start by expressing the expectation value for the Green function as

$$G(\tau) = -\frac{1}{Z}\sum_C w_C^{d(\tau)d^\dagger(0)} = -\frac{1}{Z}\sum_C w_C^{(\tau,0)}\, \frac{w_C^{d(\tau)d^\dagger(0)}}{w_C^{(\tau,0)}},$$

where $w_C^{d(\tau)d^\dagger(0)}$ denotes the weight of the configuration C with an additional operator $d^\dagger(0)$ and $d(\tau)$ in the trace factor and $w_C^{(\tau,0)}$ denotes the complete weight corresponding to the enlarged operator sequence (including enlarged hybridization determinants). Since the trace factors of both weights are identical, up to a permutation sign $(-1)^{i+j}$,

$$\frac{w_C^{d(\tau)d^\dagger(0)}}{w_C^{(\tau,0)}} = \frac{(-1)^{i+j}\det\left(M_C\right)^{-1}}{\det\left(M_C^{(\tau,0)}\right)^{-1}} = \left(M_C^{(\tau,0)}\right)_{ji},$$

with i and j denoting the row and column corresponding to the additional operators $d^\dagger$ and d in the enlarged $\big(M_C^{(\tau,0)}\big)^{-1}$. Hence, the measurement formula for the Green function becomes[10]

$$G(\tau) = -\frac{1}{Z}\sum_C w_C^{(\tau,0)}\big(M_C^{(\tau,0)}\big)_{ji} = -\frac{1}{Z}\sum_{\tilde{C}} w_{\tilde{C}}\tilde{n}^2\delta(\tau_{\tilde{n}}-\tau)\delta(\tau'_{\tilde{n}}-0)\big(M_{\tilde{C}}\big)_{\tilde{n}\tilde{n}}$$
$$= -\frac{1}{Z}\sum_{\tilde{C}} w_{\tilde{C}}\tilde{n}^2\frac{1}{\beta}\delta(\tau,\tau_{\tilde{n}}-\tau'_{\tilde{n}})\big(M_{\tilde{C}}\big)_{\tilde{n}\tilde{n}},$$

with $\delta(\tau,\tau') = \delta(\tau-\tau')$ for $\tau' > 0$, and $\delta(\tau,\tau') = -\delta(\tau-\tau'-\beta)$ for $\tau' < 0$. In the first step, we went from a sum over configurations C with n creation and annihilation operators in addition to $d(\tau)$ and $d^\dagger(0)$ to a sum over configurations $\tilde{C}$ with $\tilde{n} = n+1$ operator pairs, while in the last step, we used the translational invariance and the β-anti-periodicity of the Green function. We finally replace the factor $\tilde{n}^2$ (which comes from the $1/(n!)^2$ factor in the Monte Carlo weights without time ordering) by a sum over all pairs i,j of creation and annihilation operators, to obtain the measurement formula $G(\tau) = -\frac{1}{Z}\sum_{\tilde{C}} w_{\tilde{C}}\sum_{ij}\frac{1}{\beta}\delta(\tau,\tau_j-\tau'_i)\big(M_{\tilde{C}}\big)_{ji}$, or

$$G(\tau) = \left\langle -\sum_{ij}\frac{1}{\beta}\delta(\tau,\tau_i-\tau'_j)M_{ij}\right\rangle_{\text{MC}}. \tag{45}$$

Fourier transformation of Eq. (45) yields the measurement formula

$$G(i\omega_n) = \left\langle -\sum_{ij}\frac{1}{\beta}e^{i\omega_n(\tau_i-\tau'_j)}M_{ij}\right\rangle_{\text{MC}} \tag{46}$$

for the Fourier coefficients of the Green function. Note that in contrast to the weak-coupling approach, where we measure the Green function as a $\mathcal{O}(1/(i\omega_n)^2)$ correction to the Weiss Green function, Eq. (46) does not automatically yield the correct high frequency tail.

An elegant way to suppress the noise in $G(i\omega_n)$ at large ω_n and to obtain a compact representation of the Green function is to measure the expansion coefficients in a basis of orthogonal

[10] For the purpose of this derivation, it is convenient to use configurations C and $\tilde{C}$ without time ordering, that is, we write the Green function as

$$\begin{aligned}
G(\tau) =& -\frac{Z_{\text{bath}}}{Z}\sum_n\frac{1}{n!^2}\int_0^\beta d\tau_1\cdots d\tau_n\int_0^\beta d\tau'_1\cdots d\tau'_n\\
&\times \mathrm{Tr}_d\Big(e^{-\beta H_{\text{loc}}}\mathcal{T}d(\tau)d^\dagger(0)d(\tau_n)d^\dagger(\tau'_n)\cdots d(\tau_1)d^\dagger(\tau'_1)\Big)\det\big(M^{(\tau,0)}\big)^{-1}\big(M^{(\tau,0)}\big)_{n+1,n+1}\\
=& -\frac{Z_{\text{bath}}}{Z}\sum_n\frac{(n+1)^2}{\big((n+1)!\big)^2}\int_0^\beta d\tau_1\cdots d\tau_{n+1}\int_0^\beta d\tau'_1\cdots d\tau'_{n+1}\delta(\tau_{n+1}-\tau)\delta(\tau'_{n+1}-0)\\
&\times \mathrm{Tr}_d\Big(e^{-\beta H_{\text{loc}}}\mathcal{T}d(\tau_{n+1})d^\dagger(\tau'_{n+1})d(\tau_n)d^\dagger(\tau'_n)d(\tau_1)d^\dagger(\tau'_1)\Big)\det\big(M^{(\tau,0)}\big)^{-1}\big(M^{(\tau,0)}\big)_{n+1,n+1}\\
=& -\frac{Z_{\text{bath}}}{Z}\sum_{\tilde{n}}\frac{\tilde{n}^2}{(\tilde{n}!)^2}\int_0^\beta d\tau_1\cdots d\tau_{\tilde{n}}\int_0^\beta d\tau'_1\cdots d\tau'_{\tilde{n}}\delta(\tau_{\tilde{n}}-\tau)\delta(\tau'_{\tilde{n}}-0)\\
&\times \mathrm{Tr}_d\Big(e^{-\beta H_{\text{loc}}}\mathcal{T}d(\tau_{\tilde{n}})d^\dagger(\tau'_{\tilde{n}})\cdots d(\tau_1)d^\dagger(\tau'_1)\Big)\det\big(M^{(\tilde{n})}\big)^{-1}\big(M^{(\tilde{n})}\big)_{\tilde{n}\tilde{n}}.
\end{aligned}$$

polynomials [10]. A suitable choice are the *Legendre polynomials* $P_l(x)$ defined on $x \in [-1,1]$ through the recursion relation

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ (l{+}1)P_{l+1}(x) &= (2l{+}1)xP_l(x) - lP_{l-1}(x). \end{aligned}$$

The P_l furthermore satisfy $\int_{-1}^{1} dx\, P_k(x)P_l(x) = \frac{2}{2l+1}\delta_{kl}$. Defining $x(\tau) = 2\tau/\beta{-}1$, one can thus express the Green function on the interval $\tau \in [0,\beta]$ as

$$G(\tau) = \sum_{l\geq 0} \frac{\sqrt{2l+1}}{\beta} P_l\big(x(\tau)\big) G_l, \tag{47}$$

$$G_l = \sqrt{2l+1} \int_0^\beta d\tau P_l\big(x(\tau)\big) G(\tau). \tag{48}$$

The advantage of the Legendre representation over the Matsubara representation is a much faster decay of the expansion coefficients with increasing order. The Matsubara Fourier transform requires anti-periodization of the Green function with discontinuities at $\tau = m\beta$, which leads to slowly decaying Matsubara coefficients ($G(i\omega_n) \sim 1/i\omega_n$ for large ω_n). On the other hand, the Legendre basis represents the smooth function $G(\tau)$ on the interval $[0,\beta]$. In practice, 30 to 50 Legendre coefficients are enough to reproduce the Green function with high precision and neglecting the higher orders acts as a convenient noise filter.
From Eqs. (45) and (48) it follows that

$$G_l = \left\langle -\sum_{ij} \frac{\sqrt{2l+1}}{\beta} \tilde{P}_l(\tau_i{-}\tau'_j) M_{ij} \right\rangle_{\text{MC}}, \tag{49}$$

with $\tilde{P}_l(\tau) = P_l\big(x(\tau)\big)$ for $\tau > 0$ and $\tilde{P}_l(\tau) = -P_l\big(x(\tau{+}\beta)\big)$ for $\tau < 0$.
The Matsubara coefficients are obtained from the Legendre coefficients as $G(i\omega_n) = \sum_{l\geq 0} T_{nl} G_l$, with the unitary transformation T_{nl} given by $T_{nl} = (-1)^n i^{l+1}\sqrt{2l+1}\, j_l(\frac{1}{2}\beta\omega_n)$, which involves the spherical Bessel functions $j_l(z)$. In the limit $n \to \infty$, T_{nl} decays $\sim 1/(i\omega_n)$ for n even and $\sim 1/(i\omega_n)^2$ for n odd.
An even more compact representation of Green functions can be obtained with the so-called intermediate representation introduced in Ref. [11].

4.3 Generalizations – Matrix and Krylov formalisms

4.3.1 Matrix formalism

It is obvious from the above derivation that the hybridization-expansion formalism is applicable to general classes of impurity models [12]. Because we compute the trace factor in the weight (41) exactly, H_{loc} can contain arbitrary local interactions (for example, spin-exchange terms in multi-orbital models), degrees of freedom (for example, spins in Kondo-lattice models) or constraints (for example, 'no double occupancy' in the t-J model).

For multi-orbital impurity models with $H_{\rm loc}$ diagonal in the occupation number basis, such as models with density-density interactions, the segment formalism illustrated in Fig. 3 is still applicable, but there is now a collection of segments for each *flavor* α (orbital, spin, etc.). The trace factor can again be computed from the length of the segments (the chemical potential contribution) and the overlaps between segments of different flavor (the interaction contribution). This allows a very efficient simulation of models with 5, 7, and in principle even more orbitals, despite the fact that the corresponding Hilbert spaces ($4^5 = 1024$ for 5 orbitals, $4^7 = 16384$ for 7 orbitals) are quite large.

If $H_{\rm loc}$ is not diagonal in the occupation number basis defined by the $d^\dagger_\alpha$, the calculation of

$$\mathrm{Tr}_d\Big(e^{-\beta H_{\rm loc}}\mathcal{T}\prod_\alpha d_\alpha(\tau^\alpha_{n_\alpha})\,d^\dagger_\alpha(\tau'^\alpha_{n_\alpha})\cdots d_\sigma(\tau^\alpha_1)\,d^\dagger_\alpha(\tau'^\alpha_1)\Big) \tag{50}$$

becomes rather involved and for a model with a large Hilbert space also computationally expensive. An obvious idea is to evaluate the trace in the eigenbasis where the imaginary-time evolution operators $e^{-H_{\rm loc}\tau}$ become diagonal. On the other hand, the operators d_α and $d^\dagger_\alpha$, which are simple and sparse in the occupation number basis, become complicated matrices in the eigenbasis. The evaluation of the trace factor in the eigenbasis thus involves the multiplication of matrices whose size scales as the dimension of the Hilbert space of the local problem. Since the dimension of this Hilbert space grows *exponentially* with the number of flavors, the calculation of the trace factor becomes the computational bottleneck of the simulation, and the matrix formalism is therefore restricted to a relatively small number of flavors.

It is important to identify and use *conserved quantum numbers* [13]. Typically, these are particle number for spin up and spin down and momentum. If we group the eigenstates of $H_{\rm loc}$ according to these quantum numbers, the operator matrices acquire a sparse block structure. For example, the operator $d^\dagger_{\uparrow,q}$ connects states corresponding to the quantum numbers $m = \{n_\uparrow, n_\downarrow, k, \ldots\}$ to those with $m' = \{n_\uparrow{+}1, n_\downarrow, k{+}q, \ldots\}$ (if they exist). Checking the compatibility of the operator sequence with the different starting blocks allows us to identify the blocks which contribute to the trace without performing any expensive matrix-matrix multiplications.

Let us take as a simple example a two-orbital model with conserved quantum numbers $n_\uparrow$ and $n_\downarrow$. The operator sequence $d^\dagger_\uparrow(\tau_4)\,d^\dagger_\uparrow(\tau_3)\,d_\uparrow(\tau_2)\,d_\uparrow(\tau_1)$ (with $\tau_1 < \tau_2 < \tau_3 < \tau_4$) is compatible with the starting blocks $\{n_\uparrow = 2; n_\downarrow = 0, 1, 2\}$, since the quantum numbers evolve as

$$\{n_\uparrow = 2; n_\downarrow\} \underset{d_\uparrow}{\to} \{n_\uparrow = 1; n_\downarrow\} \underset{d_\uparrow}{\to} \{n_\uparrow = 0; n_\downarrow\} \underset{d^\dagger_\uparrow}{\to} \{n_\uparrow = 1; n_\downarrow\} \underset{d^\dagger_\uparrow}{\to} \{n_\uparrow = 2; n_\downarrow\},$$

whereas the blocks $\{n_\uparrow = 0, 1; n_\downarrow = 0, 1, 2\}$ do not contribute to the weight, since, for example,

$$\{n_\uparrow = 1; n_\downarrow\} \underset{d_\uparrow}{\to} \{n_\uparrow = 0; n_\downarrow\} \underset{d_\uparrow}{\to} \emptyset\,.$$

Having identified the contributing blocks, the trace calculation reduces to a block matrix multiplication of the form

$$\sum_{\substack{\text{contributing}\\ m}} \mathrm{Tr}_m\Big(\cdots\big(O\big)_{m''m'}\big(e^{-(\tau'-\tau)H_{\rm loc}}\big)_{m'}\big(O\big)_{m'm}\big(e^{-\tau H_{\rm loc}}\big)_m\Big), \tag{51}$$

where O is either a creation or annihilation operator, m denotes the index of the matrix block, and the sum runs over those starting sectors which are compatible with the operator sequence. Using the block structure imposed by the conserved quantum numbers, it is possible to efficiently simulate 3-orbital models or 4-site clusters. However, since the matrix blocks are dense and the largest blocks grow exponentially with system size, the simulation of 5-orbital models already becomes quite expensive and the simulation of 7-orbital models with 5, 6 or 7 electrons is doable only if we truncate the size of the blocks.
In fact, one should distinguish two types of truncations:

1. Restriction of the trace $\sum_{\text{contributing } m} \mathrm{Tr}_m(\ldots)$ to those quantum number sectors or states which give the dominant contribution,

2. Reduction of the size of the operator blocks $\big(O\big)_{m'm''}$ by eliminating high-energy states.

Truncations of type (1) have little effect at low enough temperature, because they restrict the possible states only at a single point on the imaginary-time interval. Truncations of the type (2) are more problematic and possibly lead to systematic errors which are difficult to estimate and control when the system size is large.
Accumulating a histogram of the states or quantum number sectors visited during the sampling can be very instructive. For example, in the study of correlated materials with multiple partially filled orbitals, interesting questions are the typical valence or the dominant spin state, and the importance of fluctuations to other charge and spin states. Dynamical mean-field theory allows us to address these issues by adopting a real-space representation of the solid as a collection of atoms and treating the local fluctuations on a given site through the effective impurity model construction. The strong-coupling solver, which treats the local part of the impurity problem exactly, is ideally suited for such an analysis.

4.3.2 Krylov formalism

An alternative strategy [14] to evaluate the trace factor (50) is to

1. Adopt the occupation number basis in which we can easily apply the d_α and $d_\alpha^\dagger$ operator matrices to any state and in which we can exploit the sparse nature of H_{loc} during the imaginary-time evolutions,

2. Approximate the trace by a sum over the lowest energy states, that is, by a truncation of type (1) described in the previous subsection.

Instead of evaluating the matrix corresponding to the product of operators, we propagate each retained state in the trace through the sequence of time-evolution, creation and annihilation operators. This computation only involves matrix-vector multiplications of the type $d_\alpha|v\rangle$, $d_\alpha^\dagger|v\rangle$, and $H_{\text{loc}}|v\rangle$ with *sparse* operators d_α, $d_\alpha^\dagger$ and H_{loc} and is thus possible for systems for which the multiplication of dense matrix blocks becomes prohibitively expensive. Furthermore, the approach does not require any approximation of type (2), so all excited states remain accessible at intermediate τ. While the sparsity of H_{loc} depends on the number of interaction terms,

this number grows at most proportionally to the number of orbitals squared. In contrast, the dimension of the matrix grows exponentially with the number of orbitals.
The expensive step is the calculation of the time evolution from one operator to the next. We evaluate the matrix exponentials applied to a vector, $\exp(-\tau H_{\text{loc}})|v\rangle$, by iteratively constructing the *Krylov space*

$$\mathcal{K}_p(|v\rangle) = \text{span}\{|v\rangle, H_{\text{loc}}|v\rangle, H_{\text{loc}}^2|v\rangle, \ldots, H_{\text{loc}}^p|v\rangle\}$$

and by approximating the full matrix exponential by the matrix exponential of the Hamiltonian projected onto $\mathcal{K}_p(|v\rangle)$. The iteration number p is determined by tracking the convergence of $\exp(-\tau H_{\text{loc}})|v\rangle$ and stopping the calculation if the difference between iteration p and $p+1$ drops below some cutoff value. The number of iterations depends on the time interval τ, but typically, convergence occurs for very small iteration numbers $p \ll N_{\text{dim}}$, with N_{dim} the dimension of the Hilbert space.
In the limit where the dimension of the local Hilbert space N_{dim} is large, the Krylov approach is more efficient than an implementation based on a matrix representation of the operators d_α, $d_\alpha^\dagger$ and an evaluation of the trace of the matrix product. If the Monte Carlo configuration has n creation and n annihilation operators and we perform the trace over $N_{\text{tr}} \leq N_{\text{dim}}$ states, the Krylov calculation of the trace scales as

$$\mathcal{O}\big(N_{\text{tr}} N_{\text{dim}} 2n(1+\langle p\rangle)\big),$$

where the first term comes from the application of the creation and annihilation operators and the second term, proportional to the average dimension $\langle p\rangle$ of the Krylov space, from the application of the time-evolution operators. If we retain all the states in the trace calculation, $N_{\text{tr}} = N_{\text{dim}}$, and the trace calculation scales as N_{dim}^2. If we restrict the trace to a small number of low-energy states, then N_{tr} is $\mathcal{O}(1)$ and the trace computation becomes *linear* in N_{dim}. This scaling should be compared with a computational effort of $\mathcal{O}(2nN_{\text{dim}}^3)$ for the evaluation of the trace based on matrix multiplications (without truncation of the matrix blocks).[11]
While in theory the Krylov space approach is the method of choice due to its superior N_{dim} scaling, in practice the precise numbers of N_{tr}, $\langle p\rangle$, and N_{dim} determine which one of the two approaches performs better for a given problem. Experience shows that for five orbital problems the Krylov approach becomes superior to the matrix method.

5 Scaling of the algorithms

In the weak-coupling and hybridization-expansion algorithms, the average expansion orders have a simple physical interpretation: In a DMFT calculation, they yield highly accurate measurements for the potential and kinetic energy.
Let us first consider the weak-coupling algorithm, where after the introduction of auxiliary fields (Eqs. (22) and (23)) and the shifting of the chemical potential one obtains $H = H_1 + H_2$, with

[11] In the truncated trace approach, it is important to measure the various local observables at $\tau = \frac{1}{2}\beta$ where they are least affected by the truncation at $\tau = 0$ and $\tau = \beta$. Also, it is important not to destroy the multiplet structure when truncating the trace.

$H_1 = H_\mu + \frac{1}{2}U(n_\uparrow{+}n_\downarrow) + H_{\text{bath}}$ and $H_2 = Un_\uparrow n_\downarrow - \frac{1}{2}U(n_\uparrow{+}n_\downarrow)$.[12] It follows from Eq. (20) that

$$\begin{aligned}\langle -H_2\rangle &= \frac{1}{\beta}\int_0^\beta d\tau \left\langle -H_2(\tau)\right\rangle \\ &= \frac{1}{\beta}\frac{1}{Z}\sum_{n=0}^{\infty}\frac{n{+}1}{(n{+}1)!}\int_0^\beta d\tau\int_0^\beta d\tau_1\cdots\int_0^\beta d\tau_n \mathrm{Tr}\Big(e^{-\beta H_1}\mathcal{T}(-H_2(\tau))(-H_2(\tau_n))\cdots(-H_2(\tau_1))\Big) \\ &= \frac{1}{\beta}\frac{1}{Z}\sum_C n_C w_C = \frac{1}{\beta}\langle n\rangle, \end{aligned} \tag{52}$$

and therefore the average perturbation order $\langle n\rangle$ is related to the potential energy by

$$\langle n\rangle_{\text{weak-coupling}} = -\beta U\langle n_\uparrow n_\downarrow\rangle + \tfrac{1}{2}\beta U\langle n_\uparrow{+}n_\downarrow\rangle = -\beta E_{\text{pot}} + \tfrac{1}{2}\beta U\langle n_\uparrow{+}n_\downarrow\rangle. \tag{53}$$

We learn from this formula that the average perturbation order is roughly proportional to the inverse temperature β and the interaction strength U.

In the hybridization-expansion case, the average perturbation order is proportional to the kinetic energy. In single-site DMFT, we can express the kinetic energy

$$E_{\text{kin}} = \sum_{k\sigma}\epsilon_k G_{k\sigma}(0^-)$$

in terms of the Green function and hybridization function:[13]

$$E_{\text{kin}} = \sum_\sigma\int_0^\beta d\tau\, G_\sigma(\tau)\Delta^\sigma(-\tau).$$

[12] For simplicity, we have chosen $\delta = 0$.

[13] The first step in the derivation of this formula is to switch to the Fourier representation:

$$E_{\text{kin}} = \sum_{k\sigma}\epsilon_k G_{k\sigma}(0^-) = \sum_{k\sigma}\epsilon_k\frac{1}{\beta}\sum_n e^{-i\omega_n 0^-}G_{k\sigma}(i\omega_n) = \sum_{k\sigma}\epsilon_k\frac{1}{\beta}\sum_n e^{i\omega_n 0^+}\frac{1}{i\omega_n{+}\mu{-}\epsilon_k - \Sigma_\sigma(i\omega_n)}.$$

Introducing the density of states $\mathcal{D}(\epsilon)$, we can then write

$$\begin{aligned}E_{\text{kin}} &= \sum_\sigma\frac{1}{\beta}\sum_n e^{i\omega_n 0^+}\int d\epsilon\,\frac{\epsilon}{i\omega_n{+}\mu{-}\epsilon - \Sigma_\sigma(i\omega_n)}\,\mathcal{D}(\epsilon) \\ &= \sum_\sigma\frac{1}{\beta}\sum_n e^{i\omega_n 0^+}\int d\epsilon\,\frac{-\big(i\omega_n{+}\mu{-}\epsilon - \Sigma_\sigma(i\omega_n)\big) + \big(i\omega_n{+}\mu{-}\Sigma_\sigma(i\omega_n)\big)}{i\omega_n{+}\mu{-}\epsilon - \Sigma_\sigma(i\omega_n)}\,\mathcal{D}(\epsilon) \\ &= \sum_\sigma\frac{1}{\beta}\sum_n e^{i\omega_n 0^+}\Big(-1 + \big(i\omega_n{+}\mu - \Sigma_\sigma(i\omega_n)\big)G^\sigma_{\text{loc}}(i\omega_n)\Big),\end{aligned}$$

with G_{loc} the local lattice Green function, which after convergence of the DMFT calculation is identical to the impurity Green function G. The latter is related to the hybridization function by $G = (i\omega_n{+}\mu - \Sigma - \Delta)^{-1}$. Hence, we obtain

$$E_{\text{kin}} = \sum_\sigma\frac{1}{\beta}\sum_n e^{i\omega_n 0^+}G_\sigma(i\omega_n)\Delta^\sigma(i\omega_n) = \sum_\sigma\int d\tau\, G_\sigma(\tau)\Delta^\sigma(-\tau).$$

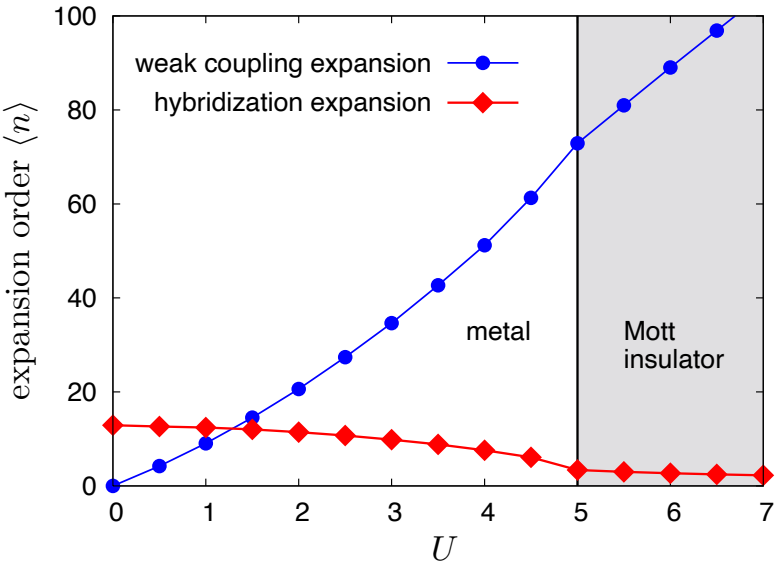


Fig. 4: *Average perturbation order for the weak-coupling and hybridization-expansion algorithms. These results correspond to the DMFT solution of the one-band Hubbard model with semi-circular density of states of bandwidth* 4 *and temperature* $T = 1/30$ *[15]. The bath is therefore different for each data point.*

Substituting the strong-coupling measurement formula (45) for G into this expression, one finds

$$E_{\rm kin} = \sum_\sigma \int_0^\beta d\tau \Big\langle -\sum_{ij} \frac{1}{\beta}\, \delta(\tau, \tau_i - \tau'_j)\, \big(M_\sigma\big)_{ij} \Big\rangle_{\rm MC} \Delta^\sigma(-\tau)$$
$$= -\sum_\sigma \Big\langle \frac{1}{\beta} \sum_{ij} \big(M_\sigma\big)_{ij}\, \Delta^\sigma(\tau'_j - \tau_i) \Big\rangle_{\rm MC} .$$

Now we use that $\big(M_\sigma\big)_{ij} = (-1)^{i+j} \det M_\sigma^{-1}[j,i] / \det M_\sigma^{-1}$, where $M_\sigma^{-1}[j,i]$ denotes the hybridization matrix with row j and column i removed. Hence, the sum

$$\sum_j (-1)^{i+j} \det M_\sigma^{-1}[j,i] \Delta^\sigma(\tau'_j - \tau_i) = \det M_\sigma^{-1}$$

appearing in the numerator is nothing but the expansion of the determinant of the hybridization matrix along column i. The expression for the kinetic energy thus simplifies to

$$E_{\rm kin} = -\sum_\sigma \Big\langle \frac{1}{\beta} \sum_i \frac{\det M_\sigma^{-1}}{\det M_\sigma^{-1}} \Big\rangle_{\rm MC} = -\frac{1}{\beta} \sum_\sigma \langle n_\sigma \rangle \,, \tag{54}$$

and the average total perturbation order $\langle n \rangle$ of the Monte Carlo configuration is related to the kinetic energy by

$$\langle n \rangle_{\text{hybridization-expansion}} = -\beta E_{\rm kin}.$$

While the average expansion order in both the weak-coupling and hybridization-expansion methods scales as β, the scaling of the expansion order with the interaction strength is very different. In the weak-coupling approach it grows roughly proportional to U, while in the

Solver	Scaling		Use
Weak-coupling	β^3	L^3	Impurity clusters with density-density interaction
Hybridization expansion (segment formalism)	β^3	L	Single-site multi-orbital models with density-density interaction
Hybridization expansion (matrix/Krylov formalism)	β	e^L	Single-site multi-orbital models with general U_{ijkl}

Table 1: *Scaling of the different impurity solvers with inverse temperature β and system size L. In the case of the segment algorithm, we assume that the calculation of the determinant ratios dominates the overlap calculations. In the matrix or Krylov case, we assume that the trace calculation dominates the calculation of the determinant ratios.*

hybridization-expansion approach, it decreases with increasing U (Fig. 4). In the case of the Anderson impurity model, this behavior leads to a significant computational speed-up for the hybridization-expansion approach in the intermediate- and large-U regime. Since local updates are $\mathcal{O}(n^2)$, a full sweep (update of all vertices in a configuration) is order $\mathcal{O}(n^3)$.
For impurity clusters, or models with complicated interaction terms, which require the matrix or Krylov formalisms discussed in Section 4.3, the hybridization-expansion method scales exponentially with system size, and we can only apply it to relatively small systems. Here, the weak-coupling approach, if applicable, can be the method of choice. Table 1 gives a summary of the different scalings (assuming a diagonal hybridization) and indicates which solver is appropriate for which type of problem. The weak-coupling solvers are mainly used in cluster DMFT calculations of the Hubbard model, where the polynomial scaling allows to treat clusters of up to 100 sites [16], at least in parameter regimes where there is no serious sign problem. The strong-coupling approach, on the other hand, is useful in particular for the study of (single-site) multi-orbital problems with complicated local interactions. Such problems typically have to be solved in single-site DMFT studies of strongly correlated materials, or in realistic simulations of transition metal impurities [17].

6 Electron-boson systems

6.1 Local phonons

In this section, we consider a quantum impurity model in which dispersionless phonons of frequency ω_0 couple to the electron density on the impurity site. The local term of the Anderson-Holstein impurity Hamiltonian $H = H_{\text{loc}} + H_{\text{mix}} + H_{\text{bath}}$ is

$$H_{\text{loc}} = -\mu(n_\uparrow + n_\downarrow) + U n_\uparrow n_\downarrow + g\big(n_\uparrow + n_\downarrow - 1\big)\big(b^\dagger + b\big) + \omega_0\, b^\dagger b. \tag{55}$$

Here, b and $b^\dagger$ denote the phonon annihilation and creation operators. An impurity model of this type has to be solved in single-site DMFT simulations of the Holstein-Hubbard model.

The bosonic sector of the Hilbert space of $H_{\rm loc}$ contains an infinite number of states. Hamiltonian-based impurity solvers truncate the Hilbert space to a finite number of phonon states, but treating even a truncated space may be computationally expensive. An attractive feature of the action-based continuous-time Monte Carlo formalism is that the phonons are integrated out, which both in the weak-coupling and the hybridization-expansion algorithms allows to treat the bosonic contribution in an elegant and efficient way.

We only discuss here the hybridization-expansion approach [18] which is based on a canonical transformation [19] called the *Lang-Firsov transformation*. This transformation decouples the electrons and phonons in the local Hamiltonian and applies to the physically relevant situation where the phonons couple to the total charge on the impurity atom. In this particular case, the electron-phonon coupling can be treated at essentially no additional computation cost.

At expansion order n_σ for spin σ, the $n_\sigma!$ diagrams corresponding to a given time sequence of fermionic creation and annihilation operators can be summed up into a determinant of a matrix M_σ^{-1}, as discussed in Sec. 4, so that the weight of the Monte Carlo configuration can be expressed as

$$w(\{O_i(\tau_i)\}) = \mathrm{Tr}_c\Big\langle \mathcal{T} e^{-\int_0^\beta H_{\rm loc}(\tau)} O_{2n}(\tau_{2n})\cdots O_1(\tau_1)\Big\rangle_b d\tau_1\cdots d\tau_{2n}\prod_\sigma(\det M_\sigma^{-1})s_\sigma, \qquad (56)$$

where the $O_i(\tau_i)$ denote the (time-ordered) creation or annihilation operators and s_σ is 1 (-1) if the spin-σ operator with the lowest time argument is a creation (annihilation) operator. To decouple the electrons and phonons by a Lang-Firsov transformation, we rewrite the local Hamiltonian (55) as

$$H_{\rm loc} = -\mu(n_\uparrow{+}n_\downarrow) + Un_\uparrow n_\downarrow + \sqrt{2}g(n_\uparrow{+}n_\downarrow{-}1)X + \frac{\omega_0}{2}\big(X^2+P^2\big). \qquad (57)$$

Here the phonon coordinate X and momentum P, satisfying $[P,X] = i$, are related to the phonon creation and annihilation operators by $X = (b^\dagger + b)/\sqrt{2}$ and $P = i(b^\dagger - b)/\sqrt{2}$. We decouple the boson and fermion operators in $H_{\rm loc}$ by shifting X by

$$X_0 = (\sqrt{2}g/\omega_0)(n_\uparrow{+}n_\downarrow{-}1) \qquad (58)$$

using the unitary transformation e^{iPX_0}. The transformed Hamiltonian $\tilde{H}_{\rm loc} = e^{iPX_0}H_{\rm loc}e^{-iPX_0}$ becomes

$$\tilde{H}_{\rm loc} = -\tilde{\mu}(\tilde{n}_\uparrow{+}\tilde{n}_\downarrow) + \tilde{U}\tilde{n}_\uparrow\tilde{n}_\downarrow + \frac{\omega_0}{2}\big(X^2+P^2\big).$$

The first two terms of $\tilde{H}_{\rm loc}$ correspond to the local terms of the Anderson impurity model with modified chemical potential $\tilde{\mu}$ and interaction strength $\tilde{U}$, where

$$\tilde{\mu} = \mu - g^2/\omega_0, \qquad (59)$$
$$\tilde{U} = U - 2g^2/\omega_0. \qquad (60)$$

The impurity electron creation and annihilation operators are transformed to polaron operators,

$$\tilde{d}_\sigma^\dagger = e^{iPX_0}d_\sigma^\dagger e^{-iPX_0} = e^{\frac{g}{\omega_0}(b^\dagger - b)}d_\sigma^\dagger,$$
$$\tilde{d}_\sigma = e^{iPX_0}d_\sigma e^{-iPX_0} = e^{-\frac{g}{\omega_0}(b^\dagger - b)}d_\sigma.$$

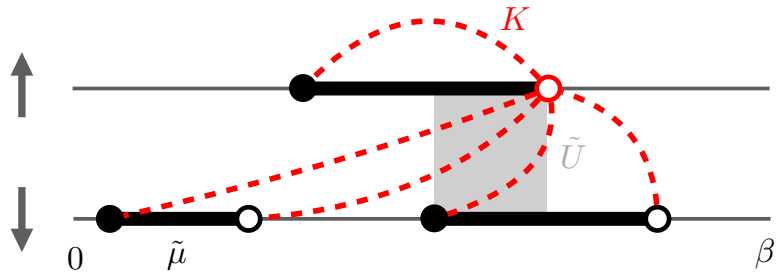


Fig. 5: *Illustration of an order $n = 3$ hybridization-expansion diagram for the Anderson-Holstein impurity model. Empty and full circles represent hybridization events. Dashed lines indicate interactions $K(\tau)$ connecting all pairs of hybridization events. We only show the lines attached to the red operator.*

After the transformation, the phonon expectation value $\langle \cdots \rangle_b$ becomes the product of a term involving electron operators, which is analogous to that computed for the Anderson impurity model without phonons, and a phonon term which is the expectation value of a product of exponentials of boson operators. The total weight of a configuration thus has the form

$$w\big(\{O_i(\tau_i)\}\big) = w_b\big(\{O_i(\tau_i)\}\big)\tilde{w}_{\text{AIM}}\big(\{O_i(\tau_i)\}\big).$$

Here, $\tilde{w}_{\text{AIM}}$ is the weight of a corresponding configuration in the Anderson impurity model with parameters modified according to Eqs. (59) and (60), while the phonon contribution is

$$w_b\big(\{O_i(\tau_i)\}\big) = \left\langle e^{s_{2n}A(\tau_{2n})}\, e^{s_{2n-1}A(\tau_{2n-1})} \cdots e^{s_1 A(\tau_1)} \right\rangle_b$$

with $0 \le \tau_1 < \tau_2 < \ldots < \tau_{2n} < \beta$, and $s_i = 1$ or (-1) if the i^{th} operator is a creation or annihilation operator. The operator in the exponent is $A(\tau) = \frac{g}{\omega_0}\big(e^{\omega_0\tau}b^\dagger - e^{-\omega_0\tau}b\big)$. The expectation value is to be taken in the thermal state of free bosons, and with the disentangling of operators $e^{X+Y} = e^X e^Y e^{-\frac{1}{2}[X,Y]}$ one finds $e^{sA(\tau)} = e^{-\frac{g^2}{2\omega_0^2}} e^{s\frac{g}{\omega_0}e^{\omega_0\tau}b^\dagger} e^{-s\frac{g}{\omega_0}e^{-\omega_0\tau}b}$, which leads to the expression

$$w_b\big(\{O_i(\tau_i)\}\big) = e^{-n\frac{g^2}{\omega_0^2}}\, e^{-\sum_{2n\ge i>j\ge 1}\frac{s_i s_j g^2}{\omega_0^2}e^{-\omega_0(\tau_i-\tau_j)}} \left\langle e^{\sum_j s_j\frac{g}{\omega_0}e^{\omega_0\tau_j}b^\dagger} e^{-\sum_j s_j\frac{g}{\omega_0}e^{-\omega_0\tau_j}b} \right\rangle_b .$$

Using $\left\langle e^{ub^\dagger} e^{vb} \right\rangle_b = e^{uv/(e^{\beta\omega_0}-1)}$ to evaluate the thermal expectation value, we finally obtain

$$w_b\big(\{O_i(\tau_i)\}\big) = \exp\left(-\frac{g^2/\omega_0^2}{e^{\beta\omega_0}-1}\left(n\big(e^{\beta\omega_0}+1\big) + \sum_{2n\ge i>j\ge 1} s_i s_j \big(e^{\omega_0(\beta-(\tau_i-\tau_j))} + e^{\omega_0(\tau_i-\tau_j)}\big)\right)\right). \quad (61)$$

This phonon contribution can be interpreted as originating from an interaction $K(\tau-\tau')$ between all pairs of operators (see Fig. 5 and Ref. [20]) of the form ($0 \le \tau \le \beta$)

$$K(\tau) = -\frac{g^2}{\omega_0^2}\frac{\cosh(\omega_0(\tau-\beta/2)) - \cosh(\omega_0\beta/2)}{\sinh(\omega_0\beta/2)}, \quad (62)$$

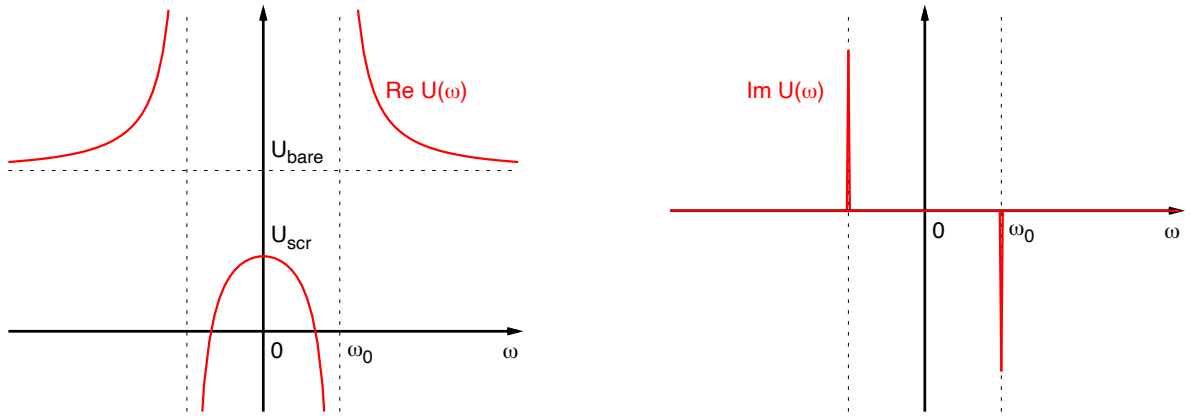


Fig. 6: *Frequency-dependent interaction $U(\omega)$ corresponding to the Anderson-Holstein impurity model with interaction $U = U_{bare}$, bosonic frequency ω_0 and electron-boson coupling g. The difference between the bare interaction U_{bare} and the screened interaction U_{scr} is $2g^2/\omega_0$ [21].*

keeping the sign factors s_i associated with creation/annihilation operators. The inclusion of phonons is thus possible without any truncation and with a negligible extra computational cost, since the computational bottleneck is the update of the determinants of hybridization functions, and not the evaluation of the nonlocal interaction between operator pairs. The phonon coupling has little effect on the average perturbation order, except very close to a bipolaronic phase.

6.2 Frequency-dependent interactions

The Anderson-Holstein impurity model corresponds to the frequency-dependent interaction $U(\omega)$ sketched in Fig. 6. In the high-frequency limit, the real part of this interaction reaches $U_{\rm bare} = U$, while the static value corresponds to the screened interaction $U_{\rm scr} = \tilde{U}$ defined in Eq. (60). The imaginary part of this frequency-dependent interaction consists of δ-functions at $\omega = \pm\omega_0$, with weight $\mp g^2\pi$ [21]. An arbitrary $U(\omega)$ can thus be thought of as arising from a Holstein-type coupling to a continuum of bosonic modes with energies ω and coupling strengths g_ω given by $g_\omega^2 = -{\rm Im}U(\omega)/\pi$. According to Eq. (62), each boson contributes an effective "interaction" $s_i s_j K(\tau_i - \tau_j) = -\frac{g_\omega^2}{\omega^2}\frac{\cosh(\omega(\beta/2-(\tau_i-\tau_j))-\cosh(\beta\omega/2)}{\sinh(\beta\omega/2)}$ between impurity creation or annihilation operators at imaginary times τ_i and τ_j. Hence, the hybridization-expansion Monte Carlo simulation for a model with general $U(\omega)$ proceeds exactly as in the case of the Anderson-Holstein impurity model, but with the K-function (62) replaced by [20]

$$K(\tau) = \int_0^\infty d\omega\, \frac{{\rm Im}U(\omega)}{\pi\omega^2}\frac{\cosh\big(\omega(\beta/2-\tau)\big) - \cosh\big(\beta\omega/2\big)}{\sinh\big(\beta\omega/2\big)} \tag{63}$$

and the shifted interaction and chemical potential (Eqs. (59) and (60)) given by

$$\tilde{\mu} = \mu + \int_0^\infty d\omega\, \frac{{\rm Im}U(\omega)}{\pi\omega}, \tag{64}$$

$$\tilde{U} = U + 2\int_0^\infty d\omega\, \frac{{\rm Im}U(\omega)}{\pi\omega} = U_{\rm scr}. \tag{65}$$

The last identity follows from the Kramers-Kronig relation and the anti-symmetry of ${\rm Im}U(\omega)$.

6.3 Boson distribution function

To measure the boson distribution function $p(x) = \big\langle \delta(x-X) \big\rangle_{\text{MC}}$, we calculate the expectation values $\big\langle \cos(\alpha X) \big\rangle_{\text{MC}}$ for different α. In order to derive the measurement formula, let us first discuss the measurement of $\langle e^{i\alpha X} \rangle_{\text{MC}}$. This measurement formula is obtained by inserting the operator $e^{i\alpha X}$ at $\tau = 0$ into the expression (56), which defines

$$w^X\big(\{O_i(\tau_i)\}\big) = \text{Tr}_c \Big\langle T_\tau e^{-\int_0^\beta H_{\text{loc}}(\tau)} O_{2n}(\tau_{2n}) \dots O_1(\tau_1) e^{i\alpha X} \Big\rangle_b \, d\tau_1 \dots d\tau_{2n} \prod_\sigma (\det M_\sigma^{-1}) s_\sigma .$$

During the Monte Carlo sampling, we then measure the ratio $w^X\big(\{O_i(\tau_i)\}\big)/w\big(\{O_i(\tau_i)\}\big)$. Since the additional $e^{i\alpha X}$ operator only modifies the bosonic factor, this amounts to measuring the ratio $w_b^X\big(\{O_i(\tau_i)\}\big)/w_b\big(\{O_i(\tau_i)\}\big)$, where $w_b^X\big(\{O_i(\tau_i)\}\big)$ is the bosonic weight factor obtained with the additional operator $e^{i\alpha X}$ at $\tau = 0$. This ratio can be expressed as

$$\begin{aligned} \frac{w_b^X\big(\{O_i(\tau_i)\}\big)}{w_b\big(\{O_i(\tau_i)\}\big)} &= \exp\left(-\frac{\alpha^2}{4}\frac{e^{\beta\omega_0}+1}{e^{\beta\omega_0}-1}\right) \\ &\quad \times \exp\left(-i\alpha X_0(\tau{=}0) - \frac{i}{e^{\beta\omega_0}-1}\sum_j s_j \frac{g}{\omega_0}\frac{\alpha}{\sqrt{2}}\big(e^{\omega_0(\beta-\tau_i)} - e^{\omega_0\tau_i}\big)\right). \end{aligned} \tag{66}$$

Note that because of the Lang-Firsov shift, this expression depends on $X_0(\tau{=}0)$, with X_0 defined in Eq. (58), and hence on the occupation of the impurity at $\tau = 0$ in the measured configuration. Since the first factor is independent of the Monte Carlo configuration, the measurement formula for $\big\langle \cos(\alpha X) \big\rangle_{\text{MC}}$ becomes

$$\begin{aligned} \big\langle \cos(\alpha X) \big\rangle_{\text{MC}} &= \exp\left(-\frac{\alpha^2}{4}\frac{e^{\beta\omega_0}+1}{e^{\beta\omega_0}-1}\right) \\ &\quad \times \left\langle \cos\Big(\alpha\sqrt{2}\frac{g}{\omega_0}\big(n_\uparrow(\tau{=}0)+n_\downarrow(\tau{=}0)-1\big) + \frac{1}{e^{\beta\omega_0}-1}\sum_j s_j \frac{g}{\omega_0}\frac{\alpha}{\sqrt{2}}\big(e^{\omega_0(\beta-\tau_i)} - e^{\omega_0\tau_i}\big)\Big)\right\rangle_{\text{MC}} . \end{aligned} \tag{67}$$

In the Monte Carlo simulation $\tilde{p}(\alpha) = \big\langle \cos(\alpha X) \big\rangle_{\text{MC}}$ is measured on a fine α-grid, which then allows to compute the boson distribution function as

$$p(x) = \frac{1}{2\pi}\int d\alpha \, \tilde{p}(\alpha) \cos(\alpha x). \tag{68}$$

References

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

[2] P. Werner: *Lecture notes for the International Summer School on Numerical Methods for Correlated Systems in Condensed Matter*, Sherbrooke, Canada (2008-2016)

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

[4] J.E. Gubernatis, N. Kawashima, and P. Werner: *Quantum Monte Carlo Methods* (Cambridge University Press, 2016)

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

[6] A.N. Rubtsov, V.V. Savkin, and A.I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005)

[7] E. Gull, P. Werner, O. Parcollet, and M. Troyer, Europhys. Lett. **82**, 57003 (2008)

[8] F.F. Assaad, T.C. Lang, Phys. Rev. B **76**, 035116 (2007)

[9] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A.J. Millis, Phys. Rev. Lett. **97**, 076405 (2006)

[10] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, Phys. Rev. B **84**, 075145 (2011)

[11] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, Phys. Rev. B **96**, 035147 (2017)

[12] P. Werner and A.J. Millis, Phys. Rev. B **74**, 155107 (2006)

[13] K. Haule, Phys. Rev. B **75**, 155113 (2007)

[14] A. Läuchli and P. Werner, Phys. Rev. B **80**, 235117 (2009)

[15] E. Gull, P. Werner, A.J. Millis, and M. Troyer, Phys. Rev. B **76**, 235123 (2007)

[16] S. Fuchs, E. Gull, L. Pollet, E. Burovski, E. Kozik, T. Pruschke, and M. Troyer, Phys. Rev. Lett. **106**, 030401 (2011)

[17] B. Surer, M. Troyer, P. Werner, A.M. Läuchli, T.O. Wehling, A. Wilhelm, and A.I. Lichtenstein, Phys. Rev. B **85**, 085114 (2012)

[18] P. Werner and A.J. Millis, Phys. Rev. Lett. **99**, 146404 (2007)

[19] I.G. Lang and Y.A. Firsov, Sov. Phys. JETP **16**, 1301 (1962)

[20] P. Werner and A.J. Millis, Phys. Rev. Lett. **104**, 146401 (2010)

[21] P. Werner and M. Casula, J. Phys.: Condens. Matter **28**, 383001 (2016)

6 Analytic Continuation of Quantum Monte Carlo Data

Erik Koch
Jülich Supercomputer Centre and
Institute for Advanced Simulation
Forschungszentrum Jülich

Contents

1 Setting the stage **3**
1.1 Analytic continuation 3
1.2 Analytic properties of the integral equations 6
1.3 Preparing the data 8

2 Optimization methods **9**
2.1 Least squares and singular values 9
2.2 Non-negative least-squares 12
2.3 Linear regularization 13
2.4 Maximum entropy 17

3 Average spectrum method **20**

4 Conclusions **24**

A Technical appendices **25**
A.1 Blocking method for correlated data 25
A.2 Non-negative least-squares algorithm (NNLS) 27
A.3 Shannon entropy 29
A.4 Sampling from a truncated normal distribution 32

The analytic continuation of Monte Carlo data may appear as an exercise in achieving the unachievable. To understand why, let us consider the example of a fermionic finite-temperature Matsubara Green function $G(\tau)$. For imaginary times $\tau \in [0,\beta]$ it is related to the spectral function $\rho(\omega)$ by the integral equation

$$G(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\omega\tau}}{1+\mathrm{e}^{-\beta\omega}}\,\rho(\omega)\,d\omega\,.$$

While calculating $G(\tau)$ from $\rho(\omega)$ is a straightforward integral, the inverse problem is hard. This is not because we have to solve a Fredholm equation of the first kind, the difficulty rather arises from the remarkable insensitivity of the imaginary-time data on changes in the spectral function. To illustrate this, we write the spectral function as a sum of delta-peaks $w_i\,\delta(\omega-\varepsilon_i)$, for which the imaginary-time Green function becomes a linear combination of exponentials

$$G(\tau) = -\frac{1}{2\pi}\sum_i w_i\big(1-n_{\mathrm{FD}}(\varepsilon_i)\big)\,\mathrm{e}^{-\varepsilon_i\tau} = -\frac{1}{2\pi}\sum_i w_i\,n_{\mathrm{FD}}(\varepsilon_i)\,\mathrm{e}^{+\varepsilon_i(\beta-\tau)},$$

where we have introduced the Fermi-Dirac distribution $n_{\mathrm{FD}}(\varepsilon) = 1/\big(\mathrm{e}^{+\beta\varepsilon}+1\big)$. While a peak at zero energy simply contributes a constant to $G(\tau)$, the contribution of peaks at large frequencies, $|\varepsilon| \gg 0$, is only noticeable close to $\tau = 0$ or β, while inside the interval $(0,\beta)$ it becomes exponentially small. To reconstruct the spectral function reliably over the entire ω-range, we thus need to know $G(\tau)$ very accurately very close to the boundaries of the interval $(0,\beta)$.
Numerical simulations can give, however, only a finite number of data points, $G(\tau_j)$. Obviously, this does not provide enough information to reconstruct a continuous spectral function: we expect that there are many different spectral functions $\rho(\omega)$ that reproduce a given set of data points $\{G(\tau_j)\}$. Such a problem without a well-defined solution is called ill posed [1]. If we insist on obtaining a unique result, we need to add constraints, e.g., by including additional information about what kind of solution we consider reasonable. In addition, Monte Carlo data are noisy. When reconstructing the spectral function, we thus need to take the accuracy of the data into account and quantify how reliable the result is, given the noise in the input. Both types of information, the estimate of the reliability of the data and our expectations about a reasonable solution of the inverse problem, can be handled using Bayesian reasoning [2].
In the following we will introduce the analytic properties that allow the continuation of Green and correlation functions. We then describe how to quantify the statistical errors in the numerical data and to set up the inverse problem. In the main part we use this to give an overview of methods to solve the inverse problem. The most straightforward approach simply performs a least-squares fit to the data points. We explain why this approach is ill posed and how it fails spectacularly. We then discuss the idea of regularization by introducing assumptions about a reasonable solution. This makes the problem well posed, but dependent on prior information. The effect of the prior information included in the regularizer can be quantified using Bayesian techniques. We discuss how they are used to argue for the different flavors of the Maximum Entropy method. Finally we introduce the average spectrum method which tries to avoid introducing prior information by calculating $\rho(\omega)$ as a functional integral over the space of all possible spectral functions.

1 Setting the stage

1.1 Analytic continuation

A system at finite temperature with time-independent Hamiltonian H is described as an ensemble of eigenstates, $H|n\rangle = E_n|n\rangle$, weighted by their Boltzmann factor. The expectation value of an operator A is thus given by

$$\langle A\rangle = \frac{\sum_n e^{-\beta E_n}\langle n|A|n\rangle}{\sum_n e^{-\beta E_n}} = \frac{1}{Z}\operatorname{Tr}\left(e^{-\beta H}A\right). \tag{1}$$

For a canonical ensemble the trace is over the N-electron Hilbert space. For a grand-canonical ensemble we get the same expression when measuring energies relative to the chemical potential, i.e., choosing $\mu = 0$, and taking the trace over the entire Fock space.
Time correlation functions can be calculated using the Heisenberg picture

$$\big\langle A(t_0{+}t)B(t_0)\big\rangle = \frac{1}{Z}\operatorname{Tr} e^{-\beta H}\, e^{\mathrm{i}H(t_0+t)}Ae^{-\mathrm{i}H(t_0+t)}\, e^{\mathrm{i}Ht_0}Be^{-\mathrm{i}Ht_0} = \big\langle A(t)B\big\rangle = \big\langle AB(-t)\big\rangle\,, \tag{2}$$

where the t_0-independence follows from the cyclic property of the trace $\operatorname{Tr} ABC = \operatorname{Tr} CAB$. Monte Carlo techniques are ideal to evaluate the high-dimensional sums needed to calculate such traces [3]. But since the time-evolution leads to complex coefficients, Monte Carlo sampling will have to fight with a serious phase-problem. This can be avoided using a Wick rotation, i.e., working in imaginary time. For this we need to analytically continue (2). This is straightforward: simply replace t in the analytic expression by the complex variable $\zeta = t{-}i\tau$ and determine for what values of ζ the result is well defined. This is most easily done using the spectral representation, i.e., evaluating the trace in the basis of eigenfunctions

$$\big\langle A(t{-}\mathrm{i}\tau)B\big\rangle = \frac{1}{Z}\operatorname{Tr}\mathrm{e}^{(\mathrm{i}t+\tau-\beta)H}A\,\mathrm{e}^{-(\mathrm{i}t+\tau)H}B = \frac{1}{Z}\sum_{n,m}\mathrm{e}^{(\mathrm{i}t+\tau-\beta)E_n}\,\mathrm{e}^{-(\mathrm{i}t+\tau)E_m}\langle n|A|m\rangle\langle m|B|n\rangle. \tag{3}$$

For systems with a finite number of states the sum is always analytic, while for systems whose spectrum is not bounded from above, we need $\beta \geq \tau \geq 0$ to maintain absolute convergence. Thus (2) can be analytically continued to a stripe below the real axis $\big\{\zeta \in \mathbb{C}\,\big|{-}\beta \leq \operatorname{Im}\zeta \leq 0\big\}$. We can then use quantum Monte Carlo to sample the function $C_{AB}(\tau) := \big\langle A(-\mathrm{i}\tau)B\big\rangle$ for $\tau \in [0,\beta]$. The analytic continuation back to the real axis is a bit less obvious, since QMC only gives us the function values, i.e., the left hand side of (3) for $t = 0$, but not the explicit functional form on the right hand side, for which we would have to know all eigenenergies and matrix elements. We can, however, define a spectral function that neatly contains all the required information by taking the Fourier transform

$$\int_{-\infty}^{\infty} dt\,\mathrm{e}^{\mathrm{i}\omega t}\big\langle A(t)B\big\rangle = \frac{2\pi}{Z}\sum_{n,m}\mathrm{e}^{-\beta E_n}\langle n|A|m\rangle\langle m|B|n\rangle\,\delta\big(\omega-(E_m{-}E_n)\big) =: \rho_{AB}(\omega) \tag{4}$$

in terms of which we can write (3) as

$$\big\langle A(t{-}\mathrm{i}\tau)B\big\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\,\mathrm{e}^{-(\mathrm{i}t+\tau)\omega}\rho_{AB}(\omega)\,. \tag{5}$$

For the special case $t = 0$ this gives us an integral equation directly relating $\rho_{AB}(\omega)$ to $C_{AB}(\tau)$

$$C_{AB}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \mathrm{e}^{-\omega\tau}\rho_{AB}(\omega)\,, \tag{6}$$

which is, however, not suited for practical calculations since the integral kernel, $\exp(-\omega\tau)$, diverges for $\omega \to -\infty$. We can get around this problem by modifying the kernel, dividing it by a function that makes it finite, and correspondingly multiplying the spectral function to leave the integral unchanged

$$C(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \frac{\mathrm{e}^{-\omega\tau}}{\mu(\omega)}\, \underbrace{\mu(\omega)\rho_{AB}(\omega)}_{=:\tilde\rho(\omega)}\,. \tag{7}$$

A suitable kernel modification would be $\mu(\omega) = 1 \pm \mathrm{e}^{-\beta\omega}$, which makes the kernel finite for $\omega \to -\infty$ as long as $\tau \leq \beta$, while keeping it finite for $\omega \to +\infty$. To analytically continue $C_{AB}(\tau) = \langle A(-\mathrm{i}\tau)B\rangle$ to the real axis we then solve the integral equation (with finite kernel)

$$C_{AB}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \frac{\mathrm{e}^{-\omega\tau}}{1 \pm \mathrm{e}^{-\beta\omega}}\, \tilde\rho^{\pm}_{AB}(\omega) \tag{8}$$

for $\tilde\rho^{\pm}_{AB}(\omega)$ and use $\rho_{AB}(\omega) = \tilde\rho^{\pm}_{AB}(\omega)/(1\pm\mathrm{e}^{-\beta\omega})$ in (5) to calculate the analytical continuation on the real axis. For the plus sign $1/(1\pm\mathrm{e}^{-\beta\omega})$ is related to the Fermi-Dirac function $n_{\mathrm{FD}}(-\omega) = 1-n_{\mathrm{FD}}(\omega)$, while for minus to the Bose-Einstein function $-n_{\mathrm{BE}}(-\omega) = n_{\mathrm{BE}}(\omega)-1$.
It is reasonable to expect that $\tilde\rho^{\pm}_{AB}(\omega)$ is a spectral function in its own right. Reordering the spectral representation (4), we can write it as

$$\begin{aligned}
\tilde\rho^{\pm}_{AB}(\omega) &= \rho_{AB}(\omega) \pm \rho_{AB}(\omega)\,\mathrm{e}^{-\beta\omega} \\
&= \rho_{AB}(\omega) \pm \frac{2\pi}{Z}\sum_{n,m} \mathrm{e}^{-\beta E_n}\langle n|A|m\rangle\langle m|B|n\rangle\, \delta\big(\omega - (E_m{-}E_n)\big)\, \mathrm{e}^{-\beta(E_m-E_n)} \\
&= \rho_{AB}(\omega) \pm \rho_{BA}(-\omega)\,.
\end{aligned} \tag{9}$$

Comparing with (3) and (4) we see that $\tilde\rho^{\pm}_{AB}(\omega)$ is the spectral function of

$$\mathrm{i}G^{\pm}_{AB}(t) := \big\langle A(t)B\big\rangle \pm \big\langle B(-t)A\big\rangle = \big\langle A(t)B\big\rangle \pm \big\langle BA(t)\big\rangle = \big\langle [A(t),\, B]_{\pm}\big\rangle, \tag{10}$$

which, for $t > 0$, is the retarded correlation function $G^{R\pm}_{AB}(t) = \Theta(t)\, G^{\pm}_{AB}(t)$, with Θ the step function, $\Theta(t{>}0) = 1$ and $\Theta(t{<}0) = 0$. As discussed above, the first term can be analytically continued to $\big\{\zeta \in \mathbb{C}\,\big|{-\beta} \leq \operatorname{Im}\zeta \leq 0\big\}$, while the second term can be continued to the stripe of width β above the real axis. It is thus natural to define the Matsubara function

$$-\,G^{M\pm}_{AB}(\tau) := \big\langle\, \mathcal{T}^{\pm}_{\tau}\, A(-\mathrm{i}\tau)B(0)\big\rangle \tag{11}$$

with the imaginary-time ordering $\mathcal{T}^{\pm}_{\tau}\, A(-\mathrm{i}\tau)B(0) = \Theta(\tau)A(-\mathrm{i}\tau)B(0) \mp \Theta(-\tau)B(0)A(-\mathrm{i}\tau)$ taking care of selecting the appropriate analytic term for the given τ. This introduces a discontinuity at $\tau = 0$

$$G^{M\pm}_{AB}(0^+) - G^{M\pm}_{AB}(0^-) = -\big\langle [A,\, B]_{\pm}\big\rangle. \tag{12}$$

From the cyclic property of the trace in (3), it follows that the Matsubara functions for positive and negative τ are related (anti)symmetrically, i.e. for $\tau \in (0, \beta)$

$$G^{M\pm}_{AB}(\beta - \tau) = -\big\langle A(-\mathrm{i}(\beta-\tau))B\big\rangle = -\big\langle B(-\mathrm{i}\tau)A\big\rangle = -\big\langle BA(\,\mathrm{i}\tau)\big\rangle = \mp G^{M\pm}_{AB}(-\tau). \quad (13)$$

For $\tau \in (0, \beta]$ we obviously have (remember the sign introduced in (11)) $G^{M\pm}_{AB}(\tau) = -C_{AB}(\tau)$, so that from (8) we obtain

$$G^{M\pm}_{AB}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \frac{\mathrm{e}^{-\omega\tau}}{1 \pm \mathrm{e}^{-\beta\omega}}\, \tilde{\rho}^{\pm}_{AB}(\omega) \quad \text{for } \tau \in [0, \beta]. \quad (14)$$

It is convenient to choose the sign in the kernel modification to obtain a simple relation for the sum rule, which directly follows from the spectral representation, using $|n\rangle\langle n| = \mathbb{1}$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \tilde{\rho}^{\pm}_{AB}(\omega) = \big\langle [A,\, B]_{\pm}\big\rangle\,. \quad (15)$$

For observables and bosonic operators we thus choose the commutator, while for fermionic Green functions it is more convenient to choose the anticommutator.
For the special case $B = A^\dagger$ we find

$$\tilde{\rho}^{\pm}_{AA^\dagger}(\omega) = \frac{2\pi}{Z}\sum_{n,m}\big(\mathrm{e}^{-\beta E_n} \pm \mathrm{e}^{-\beta E_m}\big)\,\big|\langle n|A|m\rangle\big|^2\,\delta\big(\omega - (E_m - E_n)\big), \quad (16)$$

which is obviously non-negative for the fermionic case, for the bosonic sign choice it is non-negative for $\omega = (E_m - E_n) > 0$, non-positive for $\omega < 0$, and vanishes at least linearly at $\omega = 0$. We can thus define a non-negative function $\tilde{\rho}^{-}_{AB}(\omega)/\omega$ which is regular at $\omega = 0$

$$\lim_{\omega\to 0}\frac{\tilde{\rho}^{-}_{AB}(\omega)}{\omega} = \frac{2\pi\beta}{Z}\sum_{n,m}\mathrm{e}^{-\beta E_n}\,\big|\langle n|A|m\rangle\big|^2\,\delta(E_n - E_m) \quad (17)$$

so that we can rewrite (14) with non-negative functions as

$$G^{M+}_{AA^\dagger}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \frac{\mathrm{e}^{-\omega\tau}}{1 + \mathrm{e}^{-\beta\omega}}\, \tilde{\rho}^{+}_{AA^\dagger}(\omega) \quad (18)$$

$$G^{M-}_{AA^\dagger}(\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \frac{\omega\,\mathrm{e}^{-\omega\tau}}{1 - \mathrm{e}^{-\beta\omega}}\, \frac{\tilde{\rho}^{-}_{AA^\dagger}(\omega)}{\omega}\,, \quad (19)$$

which, when A is an annihilator, applies to the diagonal elements of Green functions.
When A is an observable, we see from (9) that $\tilde{\rho}^{-}_{AA}(\omega) = -\tilde{\rho}^{-}_{AA}(-\omega)$, so that we can restrict the integral to $\omega > 0$

$$G^{M-}_{AA}(\tau) = -\frac{1}{2\pi}\int_{0}^{\infty} d\omega\, \frac{\omega\big(\mathrm{e}^{-\omega\tau} + \mathrm{e}^{-\omega(\beta-\tau)}\big)}{1 - \mathrm{e}^{-\beta\omega}}\, \frac{\tilde{\rho}^{-}_{AA}(\omega)}{\omega} \quad \text{when } A \text{ hermitian.} \quad (20)$$

We could actually cancel the factor ω in the integrand since $\tilde{\rho}^{-}_{AA}(\omega \geq 0)$ is non-negative by itself, but when calculating susceptibilities it is common to keep it, since it shows the behavior for $\omega \to 0$, (17), more clearly.

1.2 Analytic properties of the integral equations

We can gain some insight into the integral equations (18) and (19) by realizing that they are intimately related to the Euler and Bernoulli polynomials [4]. Introducing the reduced variables $x = \beta\omega$ and $y = \tau/\beta \in [0,1]$ and the functions $f(x) = \tilde{\rho}^{\pm}(x/\beta)/\beta$ (scaled to conserve the sum rule) and $g(y) = G^{M\pm}(\beta y)$ we obtain, for the fermionic case

$$g(y) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} dx\, \frac{\mathrm{e}^{-xy}}{1+\mathrm{e}^{-x}}\, f(x)\,, \tag{21}$$

from which we see that, for fixed kernel, the spectral function is spread out over an ever wider range as we go to lower temperatures. The scaled kernel of this equation is essentially the generating function of the Euler polynomials $E_n(s)$ on $s \in [0,1]$, which are defined by

$$\frac{2\mathrm{e}^{st}}{\mathrm{e}^{t}+1} = \sum_{n=0}^{\infty} E_n(s)\,\frac{t^n}{n!}\,. \tag{22}$$

With $s = \tau/\beta$ and $t = -\beta\omega$ we find from (18)

$$G^{M+}(\tau) = -\frac{1}{4\pi}\sum_{n=0}^{\infty} E_n(\tau/\beta)\;\frac{(-\beta)^n}{n!}\int_{-\infty}^{\infty} d\omega\,\omega^n\,\tilde{\rho}^{+}(\omega) \tag{23}$$

that the fermionic Matsubara function is a linear combination of Euler polynomials, where the expansion coefficients of $E_n(\tau/\beta)$ is proportional to the n-th moment of the spectral function. Since the Euler polynomials are not orthogonal, to determine the moments of $\tilde{\rho}$ from $G^{M+}(\tau)$, we first have to find the dual functions $E^n(s)$ with $\int_0^1 ds\, E^n(s)\, E_m(s) = \delta_{n,m}$. Integrating them with the generating function (22) we obtain

$$\int_0^1 ds\, E^n(s)\,\mathrm{e}^{st} = \frac{t^n}{n!}\,\frac{e^t+1}{2}\,, \tag{24}$$

which is solved by

$$E^n(s) = \frac{(-1)^n}{2\,n!}\Big(\delta^{(n)}(s-1) + \delta^{(n)}(s)\Big)\,, \tag{25}$$

where $\delta^{(n)}(s-a)$ is the n-th derivative of the delta function at $s = a$ (to make the evaluation for $a = 0$ and 1 unique, we consider the limit from inside the interval of integration). Integration by parts then produces $(-1)^n$ times the n-th derivative of the rest of the integrand at a. Using this in (23) and rewriting the Matsubara function at β as that at 0^-, eq. (13), we find that the discontinuity in the n-th derivative of the Matsubara function is proportional to the n-th moment of the spectral function

$$\frac{d^n G^{M+}(\beta)}{d\,\tau^n} + \frac{d^n G^{M+}(0)}{d\,\tau^n} = \frac{d^n G^{M+}(0^+)}{d\,\tau^n} - \frac{d^n G^{M+}(0^-)}{d\,\tau^n} = -\frac{(-1)^n}{2\pi}\int_{-\infty}^{\infty} d\omega\;\omega^n\,\tilde{\rho}^{+}(\omega)\,. \tag{26}$$

The higher moments contain the information about the spectral function at large frequencies. Extracting the derivatives from Monte Carlo data for $G(\tau)$ is difficult. Instead, they can be sampled directly: For $\tau > 0$ we have, (11),

$$-\,G^{M+}(\tau) = \big\langle A(-\mathrm{i}\tau)B\big\rangle = \frac{1}{Z}\,\mathrm{Tr}\,\mathrm{e}^{-\beta H}\mathrm{e}^{\tau H}A\mathrm{e}^{-\tau H}B\,. \tag{27}$$

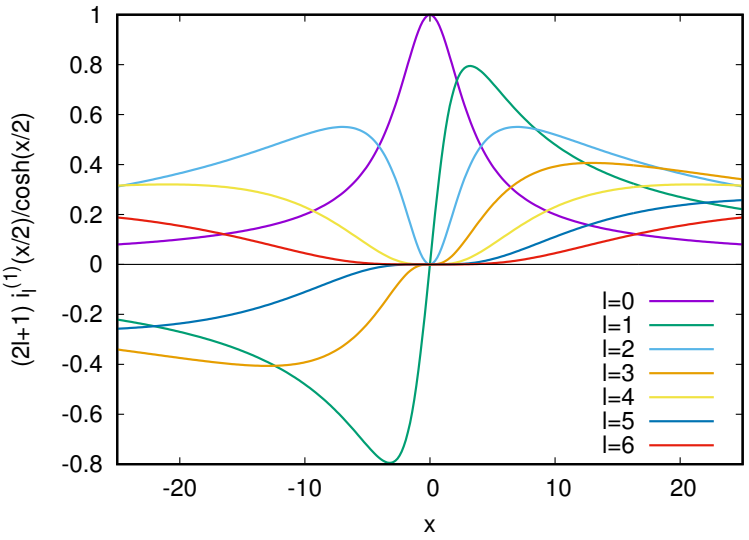


Fig. 1: *Dependence of the scaled Legendre kernel* $(2l{+}1)\, i_l^{(1)}(x/2)/\cosh(x/2)$ *on the order* l*. For* $l = 0$*,* G_l *contains information about the spectral function close to the Fermi level, while for increasing* l *it probes ever larger frequencies. As the Legendre polynomials themselves, the kernel is even/odd for even/odd* l*.*

Taking the derivative with respect to τ brings down the Hamiltonian to the left and the right of A, producing $\langle [H,\, A(-\mathrm{i}\tau)]\, B\rangle$. Repeated derivatives produce repeated commutators defined by $[H;\, A]_n := \big[H,\, [H;\, A]_{n-1}\big]$ and $[H;\, A]_0 := A$ as in the Baker-Campbell-Hausdorff formula. The moments can then be determined directly by sampling the expectation values

$$\Big\langle \big[[H;\, A]_n,\, B\big] \Big\rangle = -\frac{(-1)^n}{2\pi} \int_{-\infty}^{\infty} d\omega\; \omega^n\, \tilde{\rho}^+(\omega)\,. \tag{28}$$

Working with the Euler polynomials can become cumbersome due to their lack of orthogonality. This inconvenience can be overcome by expressing them in terms of orthogonal polynomials, e.g., shifted Legendre polynomials $P_l(2y-1)$. When the Matsubara function is expanded as [5]

$$G^{M+}(\tau) = \sum_{l=0}^{\infty} \frac{\sqrt{2l{+}1}}{\beta}\, G_l\, P_l(2\tau/\beta - 1) \quad \text{with} \quad G_l = \sqrt{2l{+}1} \int_0^\beta d\tau\, P_l(2\tau/\beta - 1)\, G^{M+}(\tau)$$

the expansion coefficients are related to the spectral function via (18) by

$$G_l = (-1)^{l+1} \sqrt{2l{+}1}\, \frac{\beta}{4\pi} \int_{-\infty}^{\infty} d\omega\, \frac{\mathrm{i}_l^{(1)}(\beta\omega/2)}{\cosh(\beta\omega/2)}\, \tilde{\rho}(\omega)\,, \tag{29}$$

where $\mathrm{i}_l^{(1)}(x)$ are the modified spherical Bessel functions of first kind. As shown in Fig. 1, for increasing l the integral kernel probes spectral features at higher and higher frequencies. From the derivatives of the recursion relation $(2l{+}1)P_l(x) = P'_{l+1}(x) - P'_{l-1}(x)$ and (26) we find that the n-th moment of the spectral function is given by a sum over all even or odd Legendre coefficients, starting at $l = n$

$$(-1)^{n+1}\, \frac{2}{n!} \sum_{k=0}^{\infty} \frac{\sqrt{4k+2n+1}}{\beta^{2k+n+1}}\, \frac{\big(2(k{+}n)\big)!}{\big(2(k{-}n)\big)!}\, G_{2k+n} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\; \omega^n\, \tilde{\rho}^+(\omega)\,. \tag{30}$$

For bosonic Matsubara functions we can obtain similar results using the Bernoulli polynomials $B_n(s)$ whose generating function is directly related to the bosonic kernel.

1.3 Preparing the data

Certainly the most important aspect of preparing Monte Carlo data for analytic continuation is the decision what data to sample. As we have seen, the information about spectral features further away from the chemical potential is concentrated in the Matsubara function extremely close to $\tau = 0$ and β. Reconstructing the spectral function from data given on a uniform τ-grid, we can therefore only expect to get reasonable results close to the chemical potential. Using, on the other hand, the derivatives of the Matsubara function at $\tau = 0$ and β gives us the moments of the spectral function, which, as we know, e.g., form the Lanczos method [6], accurately characterize the spectral function over the entire frequency range using just a few tens of the lowest moments.

The second concern is to properly characterize the statistical errors in the Monte Carlo data. Considering the integral equation

$$g(y) = \int K(y,x)\, f(x)\, dx\,, \tag{31}$$

the actual numerical data is not given as the function $g(y)$ but as vectors of M discrete data points $\boldsymbol{g} = (g_1, \dots, g_M)^\dagger$ representing $g(y)$. The mean over K independent samples is then

$$\bar{\boldsymbol{g}} = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{g}_k \tag{32}$$

with its statistical uncertainty being characterized by the $M \times M$ covariance matrix

$$\boldsymbol{C} = \frac{1}{K(K-1)} \sum_{k=1}^{K} \big(\boldsymbol{g}_k - \bar{\boldsymbol{g}}\big)\big(\boldsymbol{g}_k - \bar{\boldsymbol{g}}\big)^\dagger. \tag{33}$$

By the central limit theorem the probability density of measuring $\bar{\boldsymbol{g}}$ given the covariance matrix $\boldsymbol{C}$ instead of the exact result $\boldsymbol{g}_{\text{exact}}$ is then

$$p(\bar{\boldsymbol{g}}|\, \boldsymbol{g}_{\text{exact}}, \boldsymbol{C}) = \frac{1}{(2\pi)^{M/2} \det \boldsymbol{C}} \mathrm{e}^{-(\bar{\boldsymbol{g}}-\boldsymbol{g}_{\text{exact}})^\dagger \boldsymbol{C}^{-1} (\bar{\boldsymbol{g}}-\boldsymbol{g}_{\text{exact}})/2}. \tag{34}$$

This probability will play a central role in the reconstruction of the spectral function representing $\boldsymbol{g}_{\text{exact}}$. It is, therefore, crucial to have an accurate estimate of $\boldsymbol{C}$. Rewriting it as

$$\boldsymbol{C} = \frac{1}{K(K-1)} \sum_k (\boldsymbol{g}_k - \bar{\boldsymbol{g}})(\boldsymbol{g}_k - \bar{\boldsymbol{g}})^\dagger = \frac{1}{K(K-1)} \sum_k \boldsymbol{g}_k\, \boldsymbol{g}_k^\dagger - \frac{1}{K-1}\, \bar{\boldsymbol{g}}\, \bar{\boldsymbol{g}}^\dagger$$

and realizing that $\boldsymbol{g}\, \boldsymbol{g}^\dagger$ is the (scaled) projector onto $\boldsymbol{g}$, we see that the covariance matrix is a linear combination of K projectors to one-dimensional subspaces. We therefore need $K > M$ independent samples $\boldsymbol{g}_k$ in (33) to have a chance of obtaining a non-singular covariance matrix. Thus, reducing the discretization error requires taking more samples. The easiest way for obtaining independent samples are independent Monte Carlo runs, e.g., on a parallel computer. If we do not have enough CPUs available, we need to construct independent samples from a sequential run. This can be done, e.g., using the blocking technique described in appendix, A.1.

For the numerical solution of the integral equation (31) we also have to discretize $f(x)$ into a vector $\boldsymbol{f} = (f_1, \ldots, f_N)^\dagger$, e.g., by representing it as a piecewise constant function of value f_n on interval n. The integral equation then becomes a simple linear equation $\boldsymbol{g} = \boldsymbol{K}\boldsymbol{f}$, where the kernel matrix is obtained, e.g., from the Riemann sum [7]

$$g(y_m) = \sum_n K(y_m, x_n)\, w_n\, f(x_n)\,, \tag{35}$$

with w_n the width of interval n or, when the functions are expanded in an orthonormal set of functions $|\psi_m\rangle$ like in the Legendre expansion of the Green function, it is given by

$$g_m = \sum_n \int dy \int dx\, \overline{\psi_m(y)}\, K(y,x)\, \varphi_n(x)\; f_n = \sum_n \langle \psi_m | K | \varphi_n \rangle\, f_n\,. \tag{36}$$

Assuming $\boldsymbol{f}$ is the exact model, i.e., it gives the exact data, $\boldsymbol{K}\boldsymbol{f} = \boldsymbol{g}_{\text{exact}}$, it follows from (34)

$$p(\bar{\boldsymbol{g}}|\, \boldsymbol{f}, \boldsymbol{C}) \propto e^{-(\bar{\boldsymbol{g}}-\boldsymbol{K}\boldsymbol{f})^\dagger \boldsymbol{C}^{-1}(\bar{\boldsymbol{g}}-\boldsymbol{K}\boldsymbol{f})/2} \tag{37}$$

Factorizing the inverse covariance matrix, $\boldsymbol{C}^{-1} = \boldsymbol{T}^\dagger\boldsymbol{T}$, e.g., by Cholesky decomposition, we can absorb the explicit dependence on $\boldsymbol{C}$ by introducing $\tilde{\boldsymbol{g}} := \boldsymbol{T}\bar{\boldsymbol{g}}$ and $\tilde{\boldsymbol{K}} := \boldsymbol{T}\boldsymbol{K}$

$$\big(\bar{\boldsymbol{g}} - \boldsymbol{K}\boldsymbol{f}\big)^\dagger \boldsymbol{C}^{-1} \big(\bar{\boldsymbol{g}} - \boldsymbol{K}\boldsymbol{f}\big) = \big(\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{T}}\boldsymbol{f}\big)^\dagger \big(\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{T}}\boldsymbol{f}\big) = \big\|\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{T}}\boldsymbol{f}\big\|^2. \tag{38}$$

The covariance of the transformed data $\tilde{\boldsymbol{g}}$ is then the unit matrix, i.e. the transformation produces uncorrelated data point $\tilde{g}_n$ that all have the same (unit) errorbar.

2 Optimization methods

After discretization of model $\boldsymbol{f}$ and data $\boldsymbol{g}$ and transformation to $\tilde{\boldsymbol{g}}$, analytic continuation is reduced to solving the linear system

$$\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{K}}\boldsymbol{f}\,. \tag{39}$$

Nothing could be easier than that! When the number of data points M we are given equals the number of points N at which we want to know the model, the solution is unique, $\boldsymbol{f} = \tilde{\boldsymbol{K}}^{-1}\tilde{\boldsymbol{g}}$, as long as the kernel is not singular. When $M > N$ the model is overdetermined so that in general there will be no solution. Normally, however, we want to know the model at many more positions than we are given data points, $M < N$ so that the solution is underdetermined. A natural choice is then the $\boldsymbol{f}$ that gives the best fit to the data.

2.1 Least squares and singular values

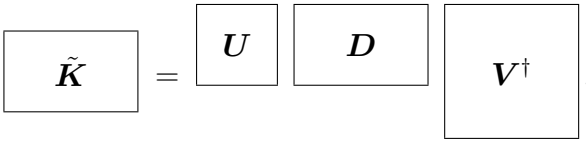
When we ask for a best-fit, we first have to define what we mean by that. Least-squares methods define "best" in terms of the Euclidian norm: minimize $\chi^2(\boldsymbol{f}) := \|\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{T}}\boldsymbol{f}\|^2$. We can justify this choice using Bayesian reasoning: As we have noted in (37), the probability of measuring $\tilde{\boldsymbol{g}}$ when the true model is $\boldsymbol{f}$ is given by $p(\tilde{\boldsymbol{g}}|\, \boldsymbol{f}) = (2\pi)^{-M/2} \exp(-\chi^2(\boldsymbol{f})/2)$. We can invert

this relation using Bayes' theorem [2], $p(B|\,A)\,p(A) = p(A,B) = p(A|\,B)\,p(B)$, stating that the probability of outcome A and B can be written as the probability of B given A times the probability of A, or, equivalently, as the probability of A given B times that of B. For the relation between model and data this implies

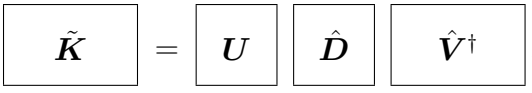
$$p(\boldsymbol{f}|\,\tilde{\boldsymbol{g}}) = \frac{p(\tilde{\boldsymbol{g}}|\,\boldsymbol{f})\,p(\boldsymbol{f})}{p(\tilde{\boldsymbol{g}})}\,. \tag{40}$$

The most probable model $\boldsymbol{f}$ given $\tilde{\boldsymbol{g}}$ thus maximizes $p(\tilde{\boldsymbol{g}}|\,\boldsymbol{f})\,p(\boldsymbol{f})$. In the absence of any further information about possible models it is reasonable to assume that $p(\boldsymbol{f})$ is the same for all $\boldsymbol{f}$, i.e., to use an "uninformative prior". A model that maximizes $p(\boldsymbol{f}|\,\tilde{\boldsymbol{g}})$ is then one that maximizes $\exp\big(-\chi^2(\boldsymbol{f})/2\big)$. It is called a "maximum likelihood estimator" and gives a best fit in the least-squares sense. Since the rank of the kernel matrix, $\operatorname{rank}\boldsymbol{K} \le \min(N,M)$, for $M < N$ the least-squares solution will not be unique: We can add any vector that is mapped by $\boldsymbol{K}$ into zero, without changing the fit. The least-squares problem is thus ill-posed. The usual way of making the solution unique is to ask in addition that $\boldsymbol{f}_{\mathrm{LS}}$ has vanishing overlap with any vector that is mapped to zero, i.e., $\boldsymbol{f}_{\mathrm{LS}}$ is orthogonal to the null space of $\tilde{\boldsymbol{K}}$.

A convenient tool for the theoretical analyzing least-squares problems is the singular value decomposition (SVD) of the matrix $\tilde{\boldsymbol{K}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\dagger$, where $\boldsymbol{U}$ is a unitary $M\times M$ matrix whose column vectors $|\boldsymbol{u}_m\rangle$ define an orthonormal basis in data space and $\boldsymbol{V}$ likewise is a unitary $N\times N$ matrix with columns $|\boldsymbol{v}_n\rangle$ spanning the space of models, while $\boldsymbol{D}$ is a diagonal $M\times N$ matrix with diagonal elements $d_1 \ge d_2 \ge \ldots \ge d_{\min(N,M)} \ge 0$. For the underdetermined case, $M < N$, the singular value decomposition can be pictured as



For the least-squares solution it is convenient to define the reduced singular value decomposition, where the null space of $\tilde{\boldsymbol{K}}$ is dropped in $\boldsymbol{V}$, pictorially,



The singular value decomposition provides a spectral representation of the kernel

$$\tilde{\boldsymbol{K}} = \sum_{i=1}^{\min(M,N)} |\boldsymbol{u}_i\rangle\, d_i\, \langle \boldsymbol{v}_i| \tag{41}$$

which allows us to write the residue vector for $M < N$ as

$$|\tilde{\boldsymbol{g}}\rangle - \tilde{\boldsymbol{K}}|\boldsymbol{f}\rangle = |\tilde{\boldsymbol{g}}\rangle - \sum_i |\boldsymbol{u}_i\rangle\, d_i\, \langle \boldsymbol{v}_i|\boldsymbol{f}\rangle = \sum_i |\boldsymbol{u}_i\rangle \Big(\langle \boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle - d_i \langle \boldsymbol{v}_i|\boldsymbol{f}\rangle \Big) \tag{42}$$

so that the least-squares solution (for which the residue vanishes when $d_M > 0$) is

$$|\boldsymbol{f}_{\mathrm{LS}}\rangle = \sum_i \frac{\langle u_i|\tilde{\boldsymbol{g}}\rangle}{d_i}\, |\boldsymbol{v}_i\rangle \quad \text{or, equivalently,} \quad \boldsymbol{f}_{\mathrm{LS}} = \hat{\boldsymbol{V}}\hat{\boldsymbol{D}}^{-1}\boldsymbol{U}^\dagger\tilde{\boldsymbol{g}}\,. \tag{43}$$

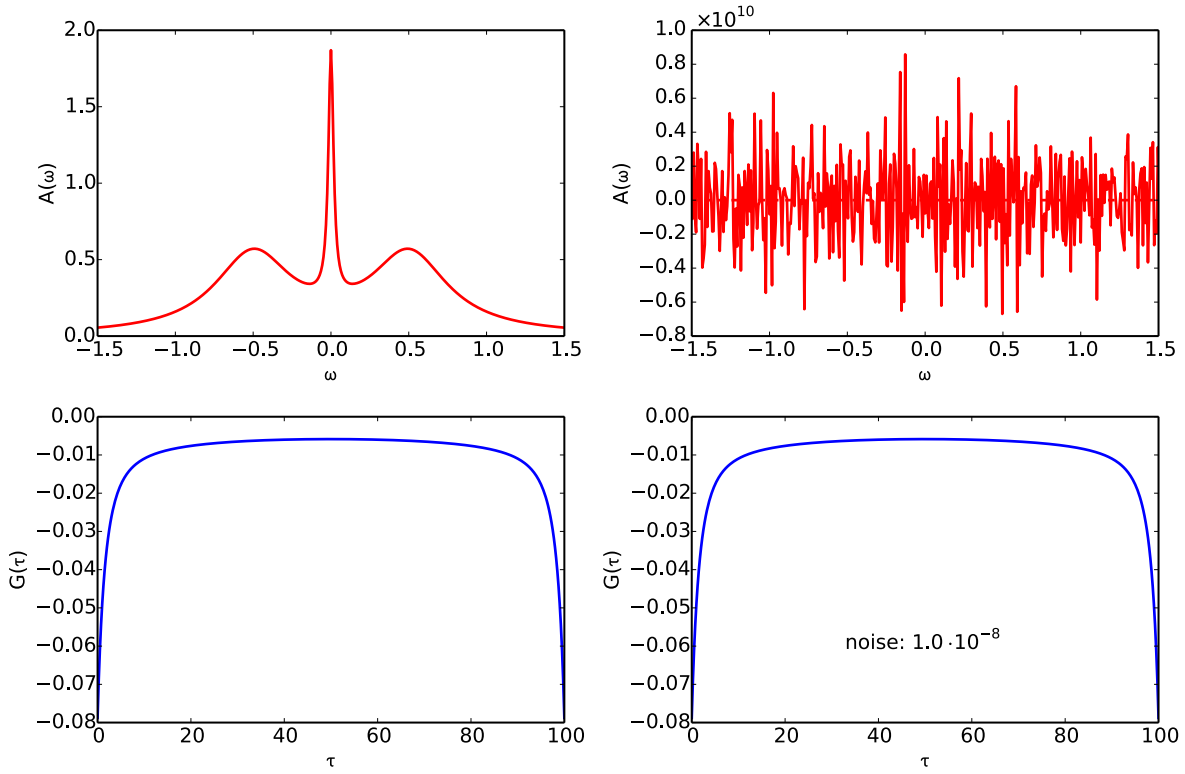


Fig. 2: *Least-squares solution for the analytical continuation of a fermionic imaginary-time Green function. The exact data (bottom left) is constructed from a simple model spectral function consisting of three Lorenz peaks (top left). We add noise of amplitude* 10^{-8} *to the data (bottom right). The least-squares solution given the noisy data is shown in the top right panel. It varies over ten orders of magnitude showing no resemblance at all to the original model.*

As simple and elegantly the least-squares solution can be constructed, as useless it is for the analytic continuation problem. This is illustrated in Fig. 2, showing that $\boldsymbol{f}_{\mathrm{LS}}$, despite giving a perfect fit to the data and, in particular, fulfilling the sum rule for $\sum_n f_n$, is completely dominated by numerical noise. What is the reason for this catastrophic failure? Making the noise in the data explicit, $\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{g}}_{\mathrm{exact}} + \Delta\tilde{\boldsymbol{g}}$ we see that

$$|\boldsymbol{f}_{\mathrm{LS}}\rangle = |\boldsymbol{f}_{\mathrm{exact}}\rangle + \sum \frac{\langle \boldsymbol{u}_i | \Delta\tilde{\boldsymbol{g}}\rangle}{d_i} \, |\boldsymbol{v}_i\rangle \, . \tag{44}$$

When the kernel has close to vanishing singular values, the noise component is divided by a number close to numerical accuracy. This is, in fact, what we are seeing in Fig. 2: dividing noise of order 10^{-8} by the numerical epsilon of double precision numbers (of order 10^{-16}), we would expect the least-squares solution to vary over about eight orders of magnitude. We can verify this picture more quantitatively by looking at the singular values of the kernel matrix, shown in Fig. 3. The exponential decay of the singular values seen in this example actually is the hallmark of an ill-conditioned problem. It is a consequence of the orthogonality of the modes $|\boldsymbol{v}_i\rangle$: With increasing i they develop more and more nodes. Integrating over these oscillating modes with the positive fermionic Green function kernel means that the integral will decrease with the number of nodes. Once the singular value reaches machine precision, the singular modes become numerically degenerate. These modes contribute negligibly to the fit of the data, but cause the catastrophic numerical instability of the least-squares result.

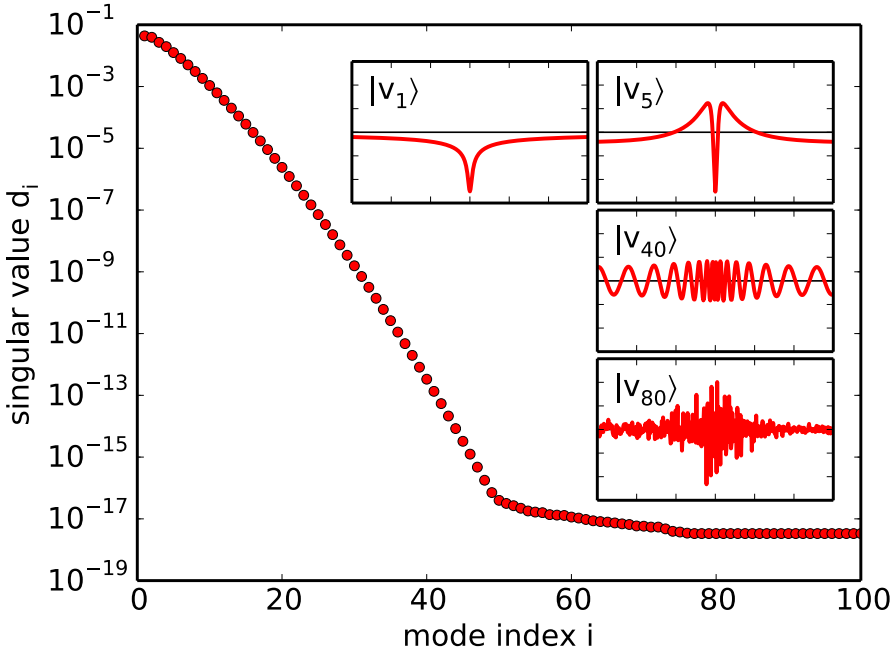


Fig. 3: *Singular values of the kernel matrix used in Fig. 2 on a logarithmic scale. The singular values decay exponentially until leveling off at a value determined by the numerical accuracy of the calculation. The insets show some of the singular modes $|\boldsymbol{v}_i\rangle$. With increasing mode index i, i.e., decreasing singular value, they have an increasing number of nodes. Once the singular value reaches the numerical accuracy, the singular modes become numerically degenerate so that the SVD routine returns arbitrary linear combinations as exemplified here for $|\boldsymbol{v}_{80}\rangle$.*

2.2 Non-negative least-squares

When motivating the least-squares approach using Bayesian reasoning, (40), one assumption was that we have no knowledge whatsoever about the possible models. When we are interested, e.g., in a diagonal spectral function, this is not quite true: We actually do know that $\boldsymbol{f}$ cannot be negative, cf. (18). To incorporate this information, the prior probability $p(\boldsymbol{f})$ should, in fact, vanish when $\boldsymbol{f}$ has a component $f_n < 0$. In other words, we really should maximize the likelihood over non-negative models only: $\max_{\boldsymbol{f}\geq 0} \exp(-\chi^2(\boldsymbol{f}))$. This approach is called non-negative least squares fitting (NNLS). A practical algorithm is discussed in A.2. It will, in general, not give a perfect fit, $\chi^2(\boldsymbol{f}_{\text{NNLS}}) > 0$, but what is not fitted is the part of the data that is incompatible with a non-negative model, i.e., pure noise.

As shown in Fig. 4, using non-negative least squares gives a dramatic improvement over the least-squares solution. Just incorporating the information about the non-negativity of the model reduces the oscillations in the result by nine orders of magnitude, bringing it into a reasonable range. This is because the amplitude of oscillating modes is now strongly limited by non-negativity. In fact, the constraints give the modes with small singular value or in the null space an important role: All modes except the first have nodes, so they can often not be included in the solution with their optimal value (43) without violating the constraint. Since the contribution of the modes with tiny singular value to the fit is tiny, they are free to arrange such that the modes with larger d_i can move closer to their optimum. Thus in NNLS the behavior of all modes is coupled, making the fit much more robust. Moreover, the non-negativity constraint makes the

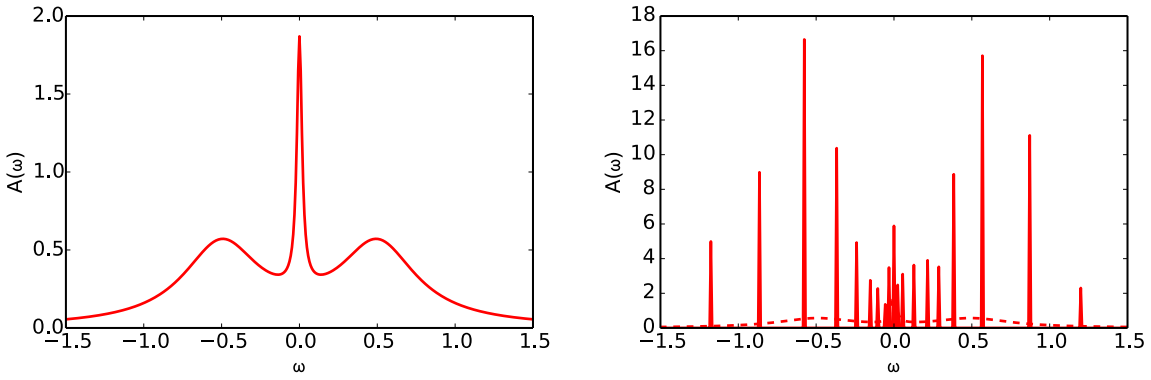


Fig. 4: *Non-negative least-squares solution of the same problem as in Fig. 2. Using our knowledge about the non-negativity of the spectral function gives a dramatic improvement, bringing the solution form a scale of the order of* $\pm 10^{10}$ *to a positive function with peaks of the order of* 10^2 *so that the original model shown on the left can actually also be seen in the plot on the right (dashed line). While the NNLS solution does show some resemblance to the original function it is far to spiky, even in the present case of exceedingly small noise (*$\sim 10^{-8}$*) in the data.*

problem well posed, i.e., giving a unique solution. Still, the spiky NNLS solutions indicate that we are still overfitting the noise in the data and this problem becomes stronger when considering data with noise levels larger than the $\sim 10^{-8}$ used for the example.

While the least-squares approaches do take information about the covariance of the data into account, via the modification of the kernel from $\boldsymbol{K}$ to $\tilde{\boldsymbol{K}}$, so that the data points that are given with higher accuracy have more weight, the results are completely independent of the absolute scale of $\boldsymbol{C}$: Multiplying it by a scalar σ^2 simply rescales all singular values of $\tilde{\boldsymbol{K}}$ by $1/\sigma$, which is compensated but the same rescaling of $\tilde{\boldsymbol{g}}$, leaving the solution unchanged. Thus the least-squares type solutions completely neglect the information about the overall noise in the data. This problem can be addressed when we include our intuition that the "true" solution should show some degree of smoothness. We then have to introduce a measure of smoothness, which puts an absolute scale in the fitting problem. This is the idea behind regularization approaches.

2.3 Linear regularization

To understand the failure of the least-squares methods better, we expand the noisy data and the fit in their respective singular modes $|\boldsymbol{u}_m\rangle$ and $|\boldsymbol{v}_n\rangle$. For the example of Fig. 2 this is shown in Fig. 5. It shows that, initially, the expansion parameters of $\boldsymbol{g}$ decrease somewhat faster with the mode index i than the singular values. Consequently the expansion of the least-squares solution also decrease with i. But once the $\langle \boldsymbol{u}_i|\boldsymbol{g}\rangle$ reach the level of the noise in the data, here $\sigma = 10^{-8}$, at $i \approx 30$, the expansion coefficients of the data remain constant while the singular values decrease further, leading to exponentially increasing contributions of the highly oscillating modes with large i that render the least-squares solution useless. The situation is quite similar for the non-negative least squares solution. The main difference being that the contributions of the modes with small or vanishing singular value are bounded $|\langle \boldsymbol{v}_n|\boldsymbol{f}_{\mathrm{NNLS}}\rangle| \lesssim 1$ by the non-negativity combined with the sum-rule for the model.

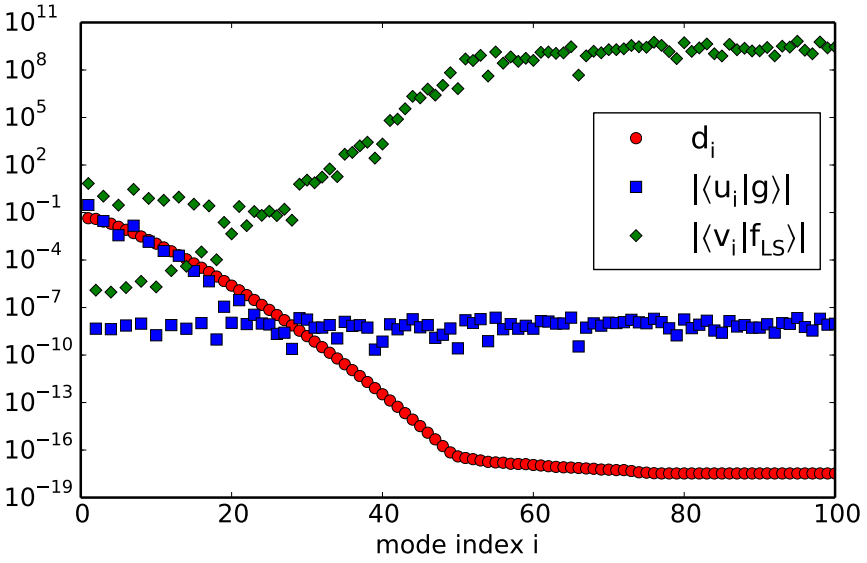


Fig. 5: *Picard plot for the example of Fig. 2. Since the model is symmetric, the expansion coefficients for the odd modes should vanish. For noisy data, instead of vanishing, the coefficients of odd modes are at the noise level. The even coefficients initially decay somewhat faster than the singular values so that the corresponding coefficients of the least-squares solution decrease with i. Once the $\langle \boldsymbol{u}_i|\boldsymbol{g}\rangle$ have decreased to the noise level, here 10^{-8}, they remain at that level while the singular values decrease further. This leads to exponentially increasing contributions of the corresponding modes to the least-squares solution.*

By maximizing the likelihood $\mathrm{e}^{-\chi^2(\boldsymbol{f})/2}$, with or without non-negativity constraint, we apparently overfit the noise that becomes most visible in the modes for which the singular value is below their contribution to the (noisy) data. The assumption behind this is that the exact solution cannot be dominated by the highly oscillating modes with vanishing singular value, i.e., that $\langle \boldsymbol{u}_i|\boldsymbol{g}_{\text{exact}}\rangle/d_i$, for large i decreases with the mode index. This is called the Picard condition. When it is not fulfilled, the reconstruction of the exact model is hopeless, since the relevant information is contained in vanishingly small coefficients $\langle \boldsymbol{u}_i|\boldsymbol{g}_{\text{exact}}\rangle$ that will be completely masked by the noise, cf. (44). When the exact model is not highly oscillating, the Picard condition holds and we have a chance of reconstructing the model from noisy data.

When the Picard condition is fulfilled we can get rid of a large part of the noise by suppressing the contribution of modes with singular value below the noise level in the data. This amounts to a least-squares fit with a truncated singular value decomposition, where the singular values beyond a limiting index are set to zero, $d_{i>i_{\text{trunc}}} := 0$.

A somewhat more refined method is to continuously switch off the small singular modes. This is called Tikhonov regularization. Introducing a regularization parameter α, the Tikhonov solution is given by

$$\boldsymbol{f}_{\text{T}}(\alpha) = \sum_{i=1}^{M} \frac{d_i}{d_i^2+\alpha^2}\langle \boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle, \tag{45}$$

which in the limit $\alpha \to 0$ becomes the least-squares solution (43), while for $\alpha \to \infty$ the solution vanishes. For finite regularization parameter, modes with large singular value $d_i \gg \alpha$ are hardly affected, while the contribution of small singular values to $\boldsymbol{f}_{\text{T}}(\alpha)$ vanishes. To employ

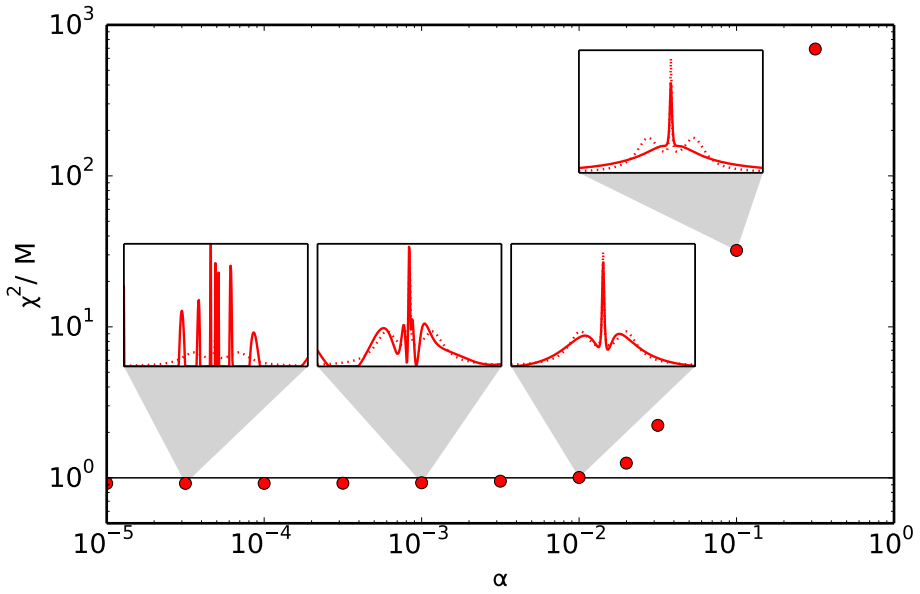


Fig. 6: *Non-negative Tikhonov regularization for the example of Fig. 2, but with noise level increased from* 10^{-8} *to* 10^{-4}*. The insets show the solutions* $\boldsymbol{f}_{\mathrm{T}}(\alpha)$ *at selected values of* α*. The dotted line shows the exact model for comparison. For small* α *the method overfits the noise, leading to strongly oscillating solutions, while the quality of the fit changes little. For large* α *the method underfits the data, leading to a rapid increase in* $\chi^2(\boldsymbol{f}_{\mathrm{T}}(\alpha)$ *and a loss of structure in the reconstructed models. The solid line indicates the expected noise in the data,* $\chi^2 = M$*, relevant for the discrepancy principle.*

Tikhonov regularization for non-negative models we need to formulate it as an optimization problem. Expanding in singular modes and completing the square, it can be written as

$$\begin{aligned}\|\tilde{\boldsymbol{K}}\boldsymbol{f}-\tilde{\boldsymbol{g}}\|^2+\alpha^2\|\boldsymbol{f}\|^2 &= \sum_{i=1}^{M}\left(\langle\boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle - d_i\langle\boldsymbol{v}_i|\boldsymbol{f}\rangle\right)^2 + \alpha^2\sum_{n=1}^{N}\langle\boldsymbol{v}_i|\boldsymbol{f}\rangle^2 \\ &= \sum_{i=1}^{M}\left(\frac{\alpha^2\langle\boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle}{d_i^2+\alpha^2} + \left(\frac{d_i\langle\boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle}{\sqrt{d_i^2+\alpha^2}} + \sqrt{d_i^2+\alpha^2}\langle\boldsymbol{v}_i|\boldsymbol{f}\rangle\right)^2\right) + \alpha^2\sum_{i=M+1}^{N}\langle\boldsymbol{v}_i|\boldsymbol{f}\rangle^2\end{aligned} \tag{46}$$

which attains its minimum $\sum_i \alpha^2\langle\boldsymbol{u}_i|\tilde{\boldsymbol{g}}\rangle/(d_i^2+\alpha^2)$ for the unique solution (45). In Bayesian terms, (40), Tikhonov regularization chooses $p(\boldsymbol{f}) \propto \mathrm{e}^{-\alpha^2\|\boldsymbol{f}\|^2/2}$ as prior probability.
Alternatively, we can express Tikhonov regularization as a least-squares problem with an expanded kernel and data as

$$\min_{\boldsymbol{f}}\left(\|\tilde{\boldsymbol{K}}\boldsymbol{f}-\tilde{\boldsymbol{g}}\|^2+\alpha^2\|\boldsymbol{f}\|^2\right) = \min_{\boldsymbol{f}}\left\|\begin{pmatrix}\tilde{\boldsymbol{K}}\\ \alpha\mathbb{1}_N\end{pmatrix}\boldsymbol{f}-\begin{pmatrix}\tilde{\boldsymbol{g}}\\ 0_N\end{pmatrix}\right\|^2. \tag{47}$$

Performing the minimization over all models gives Tikhonov regularization, restricting the optimization to $\boldsymbol{f} \geq 0$ defines the non-negative Tikhonov regularization method.
The crucial question is how to choose the regularization parameter α. Fig. 6 shows the results of non-negative Tikhonov regularization for the example of Fig. 2 increasing, however, the noise level from 10^{-8} to 10^{-4} to make the problem not too easy. For small α the solutions show

strong oscillations, while the mean-square misfit, χ^2, increases only little with α. For large α the solution becomes featureless except for the peak at the Fermi level, which is already present in the leading singular mode, cf. Fig. 3, while $\chi^2(\boldsymbol{f}_\mathrm{T}(\alpha))$ rapidly gets worse. A compromise between overfitting of the noise in the data and smoothness of the model should be reached when α is chosen such that the deviation from the optimum fit $\chi^2(\boldsymbol{f}_\mathrm{T}(\alpha)) = \|\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}_\mathrm{T}(\alpha)\|^2$ equals the noise expected in M data points $\tilde{g}_m$ with unit covariance: $\chi^2 = M$. This criterion for choosing the regularization parameter is called the discrepancy principle [8]. We can formulate it as a constrained optimization problem with α^{-2} playing the role of the Lagrange parameter:

$$\min_{\boldsymbol{f}} \|\boldsymbol{f}\|^2 + \frac{1}{\alpha^2}\left(\left\|\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f}\right\|^2 - M\right) \tag{48}$$

has the same variational equation as (46).

The regularization parameter α is the crucial ingredient of any regularization approach. Its role is to strike a balance between fitting the noisy data and keeping the solution smooth in some sense. While it is clear that with increasingly accurate data the chosen α should get smaller, there is no unique procedure for actually determining its value. The discrepancy principle is just one very reasonable way of choosing α but there is a plethora of other approaches, see [8] for a first overview. Likewise, the choice of the regularizer is not unique. Instead of $\|\boldsymbol{f}\|^2 = \langle\boldsymbol{f}|\mathbb{1}|\boldsymbol{f}\rangle$ we could choose any positive semidefinite $N\times N$ matrix $\boldsymbol{M}$ and use $\langle\boldsymbol{f}|\boldsymbol{M}|\boldsymbol{f}\rangle \geq 0$ instead. An obvious choice follows when we remember that $\boldsymbol{f}$ is the discretized version of the model function $f(x)$. As in (35), assuming a uniform x-grid, we can then write

$$\frac{1}{N}\|\boldsymbol{f}\|^2 = \frac{1}{N}\sum_{n=1}^{N}|f_n|^2 \approx \int dx\,|f(x)|^2. \tag{49}$$

Changing the integration variable from x to z, the integral and its Riemann sum in the new coordinates becomes

$$\int dx\,|f(x)|^2 = \int dz\,\frac{dx}{dz}\left|f\big(x(z)\big)\right|^2 \approx \frac{1}{N}\sum_{n=1}^{N}\frac{dx(z_n)}{dz}\left|f\big(x(z_n)\big)\right|^2 \tag{50}$$

so that Tikhonov regularization, $\boldsymbol{M} = \mathbb{1}$, on the old grid becomes regularization on the z-grid with a diagonal matrix $\boldsymbol{M}$ that contains the Jacobian factors $M_{nn} = \frac{dx(z_n)}{dz}$ on the diagonal.
Alternative choices of $\boldsymbol{M}$ impose smoothness by implementing finite-difference versions of the first or higher derivatives, choosing, e.g.,

$$\sum_{n=1}^{N-1}|f_n - f_{n+1}|^2 = \langle\boldsymbol{f}|\begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & 0 & & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}|\boldsymbol{f}\rangle. \tag{51}$$

A regularizer that penalizes the k-th derivative does not have full rank. For, e.g., the first derivative matrix in (51), all constant models give zero. In practice there is, however, no problem since the information about the low moments of the model are usually well contained in the modes with the largest singular values.

Going back to the Tikhonov regularizer, we might wonder why it actually has the effect of smoothing the solution. After all, $\|\boldsymbol{f}\|^2 = \sum_n |f_n|^2$ is local, i.e., does not depend on the change in neighboring values. So if we permuted the coordinate values $\{1, \ldots, N\}$ in an arbitrary way, the value of $\|\boldsymbol{f}\|^2$ would remain unchanged. The main reason why the identity regularizer $\boldsymbol{M} = \mathbb{1}$ leads to smooth models is that it reduces the effect of modes with small singular value. As we have seen in Fig. 3 these modes are highly oscillatory, while the modes that are least affected are the ones with few nodes that are relatively smooth. Still, even the leading mode is not entirely featureless. While a simple first derivative regularizer like (51) would reduce the contribution of such a mode, that is usually strongly supported by the data, Tikhonov will leave it largely unaffected. In that sense, Tikhonov regularization respects the variations in the important modes. To emulate this with a derivative regularizer would require to laboriously taylor an $\boldsymbol{M}$ suitable for every specific kernel $\tilde{\boldsymbol{K}}$.

There is also a second aspect. As we noted above, $\boldsymbol{f}_{\mathrm{T}}(\alpha \to \infty) = 0$. When we impose a sum rule, however, we force the solution to be finite and find from

$$\min_{\boldsymbol{f}} \left(\sum_n f_n^2 + \lambda_0 \Big(1 - \sum_n f_n \Big) \right) \tag{52}$$

that the Tikhonov regularizer prefers a flat solution, $f_n = 1/N$, or, in the case of a general diagonal matrix, $f_n \propto 1/M_{nn}$. These are the models resulting in the absence of data, i.e., for diverging variance resulting in vanishing $\tilde{\boldsymbol{K}}$ and $\tilde{\boldsymbol{g}}$, except for the 0-th moment sum-rule. They are called the default model of the regularizer.

We are, of course, not limited to bilinear regularizers of the type $\langle \boldsymbol{f}|M|\boldsymbol{f}\rangle$. An important non-linear regularizer is the entropy of the model. It is the basis of the maximum entropy approach.

2.4 Maximum entropy

Maximum entropy methods differ from Tikhonov-type regularization in the assumptions they make about the solutions. While Tikhonov is based on the Picard condition giving preference to the modes with large singular value, maximum entropy favors models that contain as little information as possible. This is measured by the information entropy, see A.3 for details. Using the generalized entropy (89)

$$H(\boldsymbol{f}; \boldsymbol{\rho}) = -\sum_n \left(f_n \ln \frac{f_n}{\rho_n} - f_n + \rho_n \right) \tag{53}$$

as regularizer that should be maximized, we have to solve the non-linear optimization problem

$$\min_{\boldsymbol{f}} \Big(\chi^2(\boldsymbol{f})/2 - \alpha H(\boldsymbol{f}; \boldsymbol{\rho}) \Big), \tag{54}$$

where we use α instead of α^2 as in (46) to conform with the conventions used, e.g., in [9]. A convenient property of the non-linear entropy regularizer is that it automatically ensures the positivity of the solution since the gradient

$$-\frac{\partial H(\boldsymbol{f};\boldsymbol{\rho})}{\partial f_n} = \ln\frac{f_n}{\rho_n} \tag{55}$$

diverges for $f_n \to 0$ while the gradient of the fit function

$$\frac{\partial \frac{1}{2}\chi^2(\boldsymbol{f})}{\partial f_n} = \frac{\partial}{\partial f_n}\frac{1}{2}\sum_m \Big(\tilde{g}_m - \sum_{n'}\tilde{K}_{mn'}f_{n'}\Big)^2 = -\sum_m \tilde{K}^\dagger_{nm}\Big(\tilde{g}_m - \sum_{n'}\tilde{K}_{mn'}f_{n'}\Big) \tag{56}$$

is always finite, so that any solution of the variational equations f_n has the same sign as ρ_n. Since it is easy to also calculate the Hessians

$$\frac{\partial^2 \frac{1}{2}\chi^2(\boldsymbol{f})}{\partial f_{n'}\partial f_n} = \sum_m \tilde{K}^\dagger_{nm}\tilde{K}_{mn'} \quad \text{and} \quad -\frac{\partial^2 H(\boldsymbol{f};\boldsymbol{\rho})}{\partial f_{n'}\partial f_n} = \frac{1}{f_n}\,\delta_{nn'} \tag{57}$$

it is straightforward to solve the non-linear minimization problem using, e.g., the Levenberg-Marquardt method. The only slight complication being that finite steps in the iteration might change the sign of some component f_n.

In the absence of data, minimizing (54), i.e, setting (55) to zero, we find $\boldsymbol{f} = \boldsymbol{\rho}$. Thus $\boldsymbol{\rho}$ is the default model and, as in (88) or (50) it can be related to the choice of the grid. Even when we have decided on a default model, we still have to determine the value of the regularization parameter. Choosing it according to the discrepancy principle is called historic MaxEnt [9].

Other flavors of the maximum entropy method determine the regularization parameter using Bayesian methods. For this we write the entropy regularizer as a prior probability

$$p(\boldsymbol{f}\,|\boldsymbol{\rho},\alpha) \propto \mathrm{e}^{+\alpha H(\boldsymbol{f};\boldsymbol{\rho})} \tag{58}$$

so that the minimization (54) becomes equal to maximizing the posterior probability, cf. (40),

$$p(\boldsymbol{f}\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho},\alpha) = \frac{p(\tilde{\boldsymbol{g}}\,|\boldsymbol{f},\boldsymbol{\rho},\alpha)\,p(\boldsymbol{f}\,|\boldsymbol{\rho},\alpha)}{p(\tilde{\boldsymbol{g}})} \propto \mathrm{e}^{-\chi^2(\boldsymbol{f})/2+\alpha H(\boldsymbol{f};\boldsymbol{\rho})}\,, \tag{59}$$

where we have used that the QMC data $\tilde{\boldsymbol{g}}$ is actually independent of our choice of regularization parameter and default model. The Bayesian approach to determining the regularization parameter uses the posterior probability of α

$$p(\alpha\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho}) = \int\prod_n \frac{df_n}{\sqrt{f_n}}\; p(\boldsymbol{f},\alpha\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho}) = \int\prod_n 2d\sqrt{f_n}\;\frac{p(\boldsymbol{f}\,|\tilde{\boldsymbol{g}},\boldsymbol{\rho},\alpha)\,p(\tilde{\boldsymbol{g}},\boldsymbol{\rho},\alpha)}{p(\tilde{\boldsymbol{g}},\boldsymbol{\rho})} \tag{60}$$

obtained from marginalizing out $\boldsymbol{f}$, i.e., integrating over the space of models $\boldsymbol{f}$. The peculiar choice of the integration measure, $2d\sqrt{f_n}$, is discussed in [9]. It naturally appears in the expression for the entropy when using Stirling's approximation to one order higher than in the derivation given in A.3, which rather suggests that the factor $1/\sqrt{f_n}$ should be considered part of the entropic prior and not the integration measure.

For historic MaxEnt it was not necessary to know the normalized probabilities (58) and (59). When, however, we want to compare the probabilities of different renormalization parameters, we need to determine the normalization of the distributions that depend on α by, again, integrating over $\boldsymbol{f}$

$$Z_{\chi^2}(\tilde{\boldsymbol{g}}) = \int_{-\infty}^{\infty} \prod_n df_n \, \mathrm{e}^{-\chi^2(\boldsymbol{f})} = (2\pi)^{M/2} \tag{61}$$

$$Z_H(\boldsymbol{\rho}, \alpha) := \int \prod_n df_n \, \frac{\mathrm{e}^{+\alpha H(\boldsymbol{f},\boldsymbol{\rho})}}{\prod_{n'} \sqrt{f_{n'}}} \tag{62}$$

Since the likelihood is a Gaussian, the integral is straightforward, cf. (34). The normalization of the entropic prior is more difficult. In MaxEnt such functional integrals are approximated by Gaussian integrals obtained from expanding the exponent to second order about its maximum. As already discussed above, the entropy term is maximized when the model equals the default model. The second-order expansion (57) is thus given by the diagonal matrix $-\delta_{nn'}/\rho_n$, i.e., the entropy becomes just a Tikhonov regularizer with general diagonal matrix. Also expanding $1/\sqrt{f_n} \approx \big(1-(f_n-\rho_n)/2\rho_n\big)/\sqrt{\rho_n}$, the normalization of the entropy prior thus is approximated by that of the simple Tikhonov prior without non-negativity constraint [9]

$$Z_H(\boldsymbol{\rho}, \alpha) \approx \prod_n \int_{-\infty}^{\infty} df_n \, \frac{\mathrm{e}^{-\frac{\alpha(f_n-\rho_n)^2}{2\rho_n}}}{\sqrt{\rho_n}} \left(1 - \frac{f_n - \rho_n}{2\rho_n}\right) = \left(\frac{2\pi}{\alpha}\right)^{N/2}. \tag{63}$$

To calculate $p(\alpha \,|\tilde{\boldsymbol{g}}, \boldsymbol{\rho})$, (60), we still have to choose a prior probability $p(\tilde{\boldsymbol{g}}, \boldsymbol{\rho}, \alpha)$. From the discrepancy principle it seems reasonable that it should be independent of the data normalized to have a unit covariance matrix. If we also assume that α is independent of the default model, we only have to choose $p(\alpha)$. Assuming that the prior is scale invariant

$$p(\alpha)\, d\alpha \stackrel{!}{=} p(s\alpha)\, d(s\alpha) \tag{64}$$

one obtains the Jeffreys prior $p(\alpha) \propto 1/\alpha$ [2], which might not be the most appropriate choice, since the scale of the regularization parameter is fixed by the noise in the data, which we know. Using all this for calculating the posterior probability of the renormalization parameter, there are two different flavors of how α enters the analytical continuation: Historic MaxEnt chooses the α that maximizes $p(\alpha \,|\tilde{\boldsymbol{g}}, \boldsymbol{\rho})$ to determine $\boldsymbol{f}_{\text{historic}} = \boldsymbol{f}(\alpha_{\max})$. Bryan's method no longer insists on picking a specific value of α. The approach rather determines the model as the average over all regularization parameters, weighted with their posterior probability

$$\boldsymbol{f}_{\text{Bryan}} = \int_0^{\infty} d\alpha \, \boldsymbol{f}(\alpha)\, p(\alpha \,|\tilde{\boldsymbol{g}}, \rho)\,. \tag{65}$$

It might seem that the MaxEnt approaches could be improved by actually performing the integrals over model space exactly rather than using simple Gaussian approximations that even violate the non-negativity of the models, which is one of the precious priors that we are sure

of. But doing these integrals is fiendishly hard. A second drawback of MaxEnt, or rather of all regularization approaches, is the need to deal with a regularization parameter, introducing the need for making assumptions about its behavior, its prior probability and the like, for which there is no apparent solution. If only we could efficiently integrate over model space there might be a way of eliminating all these complication arising from the need to regularize. Instead of looking for a solution that maximizes some posterior probability, we could ask for the average over all possible models, weighted with their likelihood. This approach which is free of explicit regularization parameters is the average spectrum method.

3 Average spectrum method

The average spectrum methods is an appealing alternative to the optimization approaches. It was probably first proposed by White [10] and reinvented several times after. The basic idea is of striking elegance: The spectral function is obtained as the average over all physically admissible functions, weighted by how well they fit the data

$$f_{\mathrm{ASM}}(x) := (2\pi)^{-M/2} \int_{f(x)\geq 0} \mathcal{D}f \; f(x)\, \mathrm{e}^{-\chi^2[f]/2}. \tag{66}$$

Due to the ill-conditioning of the inverse problem there are very many functions that differ drastically but essentially fit the data equally well. Taking the average, we can thus expect that the spectral features not supported by the data will be smoothed out, providing a regularization without the need for explicit parameters. So far the practical application of this conceptually appealing approach has, however, been hampered by the immense computational cost of numerically implementing the functional integration.

It is worth emphasisizing that the non-negativity constraint is essential. An unconstrained integration over the Gaussian in (66) actually produces a least-squares solution. Since the width of the likelihood increases with the inverse of the singular value this would, of course, be numerically very inefficient, and since the width diverges for the modes in the null space of the kernel, their contribution will never converge to a definite value, reflecting that the problem is underdetermined.

When we discretize the model function $f(x)$ as discussed in Sec. 1.3, the functional integral becomes

$$\boldsymbol{f}_{\mathrm{ASM}} \propto \prod_{n=1}^{N} \int_0^\infty df_n \, \boldsymbol{f} \, \mathrm{e}^{-\chi^2(\boldsymbol{f})/2}, \tag{67}$$

where the N-dimensional integral can be evaluated by Monte Carlo techniques. The most straightforward approach is to perform a random walk in the space of non-negative vectors $\boldsymbol{f}$, updating a single component, $f_n \to f'_n$, at a time. Detailed balance is fulfilled when we sample the new component f'_n for the conditional distribution $\propto \mathrm{e}^{-\chi^2(\boldsymbol{f};f'_n)/2}$ with

$$\chi^2(\boldsymbol{f}; f'_n) := \big\| \underbrace{\tilde{\boldsymbol{g}} - \tilde{\boldsymbol{K}}\boldsymbol{f} + \tilde{\boldsymbol{K}}_n f_n}_{=:\tilde{\boldsymbol{g}}_n} - \tilde{\boldsymbol{K}}_n f'_n \big\|^2 = \tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{K}}_n \left(f'_n - \tilde{\boldsymbol{K}}_n^\dagger \, \tilde{\boldsymbol{g}}_n / \tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{K}}_n \right)^2, \tag{68}$$

where $\tilde{\boldsymbol{K}}_n$ is the n-th column of $\tilde{\boldsymbol{K}}$. We thus have to sample f'_n from a univariate Gaussian of width $\sigma = 1/\|\tilde{\boldsymbol{K}}_n\|$ centered at $\mu = \tilde{\boldsymbol{K}}_n^\dagger \tilde{\boldsymbol{g}}_n / \|\tilde{\boldsymbol{K}}_n\|^2$ and truncated to the non-negative values $f'_n \in [0, \infty)$. This can be done very efficiently, e.g., as described in A.4.

Still, sampling components can be very slow because the width of the Gaussian (68) is, in general set by the inverse of the largest singular value, i.e., the random walk performs only exceedingly small steps. This is even more evident when sampling spectral functions, where we cannot change just a single f_n without violating the sum-rule (15). A way around is to introduce global moves along the principal axes of χ^2, i.e., along the singular modes.

Transforming to the new bases $\boldsymbol{h} := \boldsymbol{U}^\dagger \tilde{\boldsymbol{g}}$ in data and $\boldsymbol{e} := \boldsymbol{V}^\dagger \boldsymbol{f}$ in model space diagonalizes

$$\chi^2(\boldsymbol{f}) = \left\| \boldsymbol{U}^\dagger \tilde{\boldsymbol{g}} - \boldsymbol{D}\boldsymbol{V}^\dagger \boldsymbol{f} \right\|^2 = \sum_{i=1}^{M} (h_i - d_i e_i)^2 \tag{69}$$

so that in the new basis the integral (67) factorizes into Gaussian integrals

$$e_i^{\mathrm{ASM}} \propto \int_{\boldsymbol{f} \geq 0} de_i \, e_i \, \mathrm{e}^{-(h_i - d_i e_i)^2/2}. \tag{70}$$

Without non-negativity constraint the integrals would be independent and result in a least-squares solution. With the constraint they are coupled via their range of integration. Updating modes $e_i \to e'_i$ is restricted by the condition $\boldsymbol{f}' = \boldsymbol{f} + (e'_i - e_i)\boldsymbol{V}_i \geq 0$, where $\boldsymbol{V}_i$ is the i-th column vector of $\boldsymbol{V}$. This is equivalent to $e'_i \geq e_i - f_n / V_{ni}$ for $V_{ni} > 0$ and correspondingly for $V_{ni} < 0$ so that e'_i is constrained to

$$\max \left\{ \frac{f_n}{V_{ni}} \middle| V_{ni} < 0 \right\} \leq e_i - e'_i \leq \min \left\{ \frac{f_n}{V_{ni}} \middle| V_{ni} > 0 \right\}. \tag{71}$$

Sampling modes is usually much more efficient than sampling components: For modes with large singular value the Gaussian is narrow so that the random walk quickly jumps close to the optimal value h_i/d_i and then takes small steps around there. For modes with small or zero singular value the distribution is very broad so that the random walk can take large steps, allowing for an efficient sampling. Still, sampling may become inefficient when non-negativity restricts a mode to a very narrow interval. This will happen when $\boldsymbol{f}$ has regions where it becomes very small, e.g., in the tail of the spectral function. Then the scale for the step size is not given by the singular value but rather by the width of the interval (71). Also this problem can be overcome by using a real space renormalization group technique, introducing blocks of varying size in which modes are sampled. This way the method can interpolate efficiently between sampling components, i.e., blocks of size 1, and sampling modes, i.e., blocks of size N. Details of the method and its performance are given in [11, 12].

Using this approach, we find that the results of the average spectrum method actually depend on the choice of the discretization (67). This is not a problem of the particular method, but a general feature of the functional integral and would also affect, e.g., MaxEnt were it to do the normalization and marginalization integrals exactly, see e.g. Sec. 6.2 of [2]. We find that the choice of the coordinates for the discretization grid plays the role of a default model, while the number of grid points N acts as a regularization parameter.

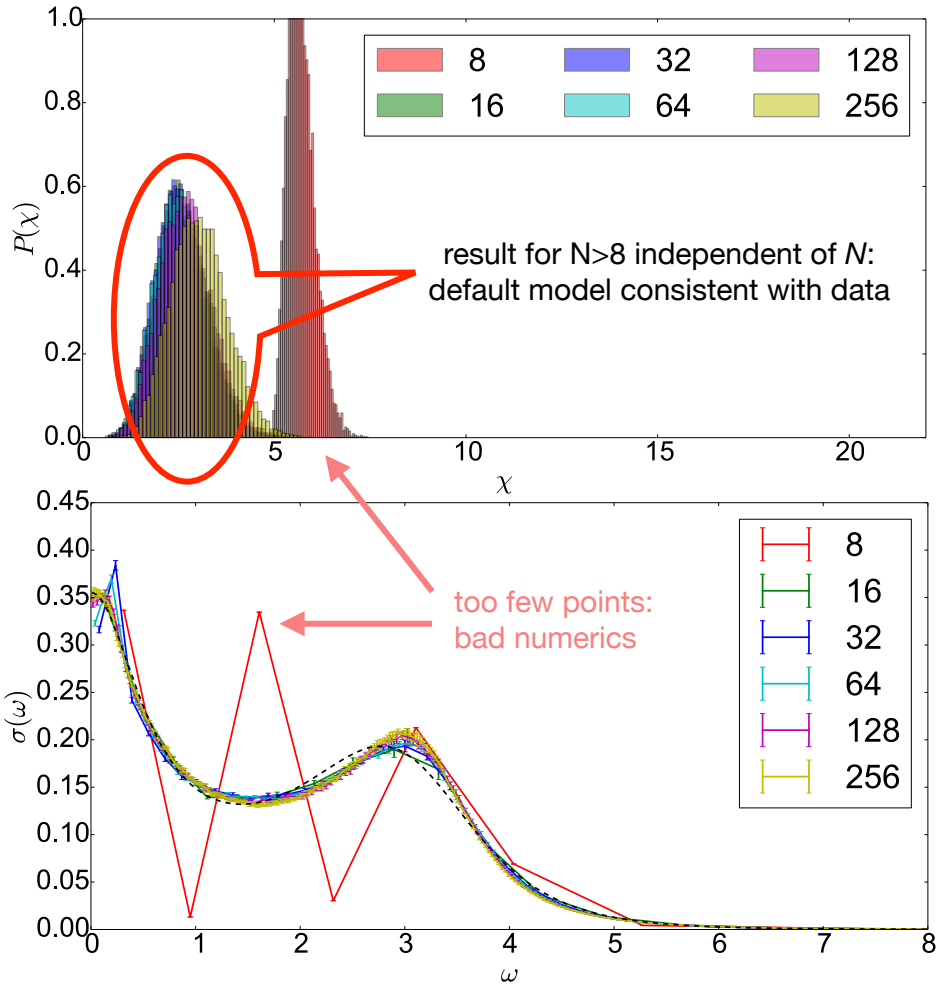


Fig. 7: *Average spectrum method on a Gaussian grid determined from NNLS for an optical conductivity reconstructed from Matsubara data [11]. The resulting spectral shown below are rather insensitive to the choice of the regularization parameter, the number of grid points N, except for $N = 8$ where the coarse grid leads to discretization errors when evaluating the Fredholm integral. The exact model is shown as the dashed line for comparison. The top panel shows the histograms of χ taken during the Monte Carlo sampling of the functional integral. The small variations in the histograms indicate a robust choice of the grid.*

The reason for this is that the notion of sampling uniformly, i.e., with a flat prior, is tied to the choice of a specific grid. This is most easily understood when we consider what happens when we double the number of grid points. On the original grid we sample $f \in [0, \infty)$. On the denser grid we represent f over the large interval by two values $\hat{f}_1$ and $\hat{f}_2$ over intervals of half the width, so that $f = \hat{f}_1 + \hat{f}_2$. If we sample the $\hat{f}_i \in [0, \infty)$ with a flat prior, $p(\hat{f}_i) = \text{const.}$, this implies a probability distribution for f

$$p(f) = \int_0^\infty d\hat{f}_1\, p(\hat{f}_1) \int_0^\infty d\hat{f}_2\, p(\hat{f}_2)\; \delta\big(\hat{f}_1 + \hat{f}_2 - f\big) \propto \int_0^f d\hat{f}_1 = f \tag{72}$$

which is *not* flat. Properly defining the flat prior as $p(\hat{f}_i) = \lim_{\lambda\to\infty} \mathrm{e}^{-\hat{f}_i/\lambda}/\lambda$, $p(f)$ becomes a gamma distribution. More generally, we find that sampling with a flat distribution on a particular grid defines a measure for the functional integral (66) represented by a gamma process [11].

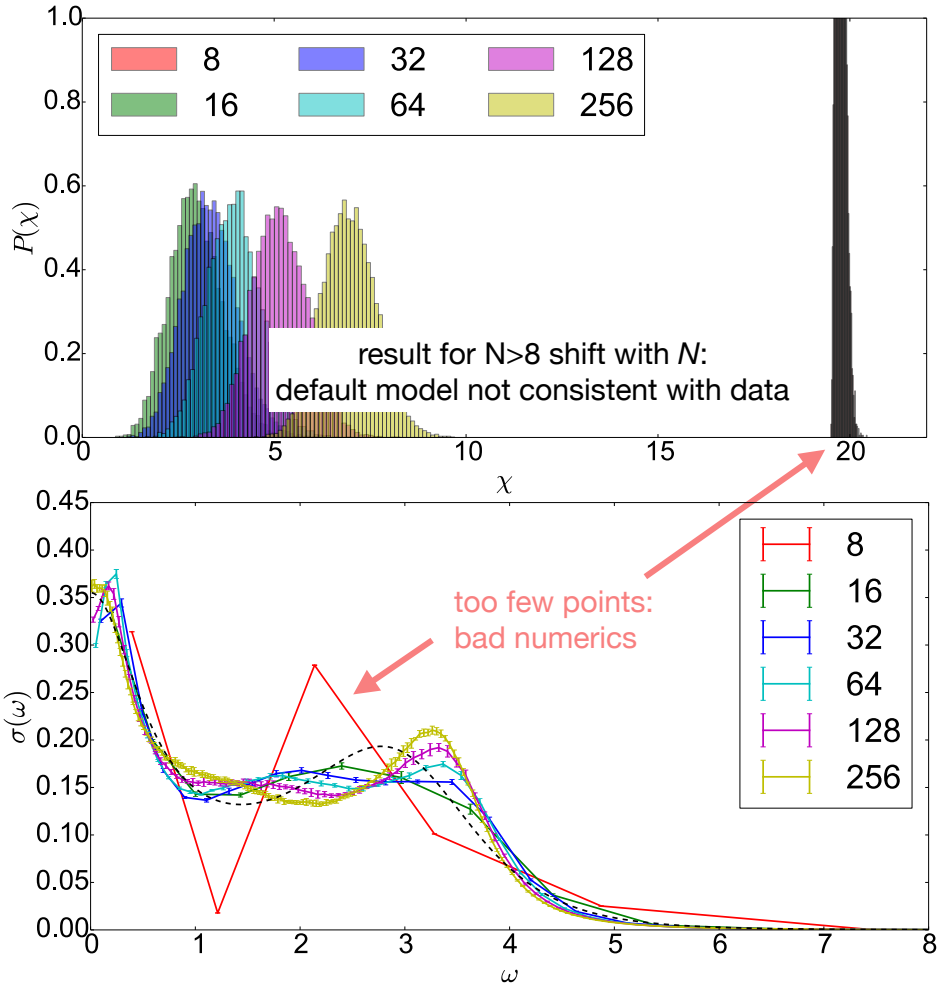


Fig. 8: *Average spectrum method on a Lorentzian grid of the same width as the Gaussian in Fig. 7. The resulting spectral functions shown below are now quite sensitive to the choice of the regularization parameter, the number of grid points N. In the top panel we see from the histograms of χ that the fit sizeably deteriorates with increasing N, indicating problems with the choice of the grid.*

Still, we can give a practical recipe for determining the regularization parameters and checking the quality of the results. To find the grid type (default model), we use non-negative least-squares to determine the width of the spectrum. This implies a Gaussian grid of particular width. We then vary the number of grid points to check how the results change with increasing N. When N is too small, the result will be inaccurate because of discretization errors in evaluating the integral (31) entering χ^2. When N becomes to large there is a rapidly increasing number of vectors $\boldsymbol{f}$ that, despite having a small weight $\mathrm{e}^{-\chi^2(\boldsymbol{f})/2}$, contribute to the average due to their sheer number. In between there will be a region, where the results are fairly independent of the actual choice of N. This is shown in Fig. 7. When the grid is not chosen well, as in Fig. 8, where the grid uses a Lorentzian density of the same width as the Gaussian in the previous figure, results vary strongly with N.

This approach gives already reliable and robust results. When we have to deal with particularly difficult cases, we can use Bayesian techniques to make the method even more robust by sampling over different grids, albeit at an increased computational cost [11].

4 Conclusions

As for so many problems, there is no magic solution to the problem of analytic continuation. Any method can only reconstruct what is in the data and must substitute missing information ideally by exact prior knowledge or, otherwise, by mere assumptions about the solution. The most important aspect of analytic continuation is thus encountered already before the solution of the inverse problem is even started. Depending on what features of the model we are interested in, we have to decide where to measure the data. If we want, e.g., to reconstruct the spectral function far from the Fermi level, it does not help to just have highly accurate values for the Green function when they are not close enough to $\tau = 0$ or β to give information about the discontinuity in its derivatives.

Moreover, it is deceiving to just look at the single result returned by a regularization approach. There is not "the" solution, rather every method produces an expected solution with its uncertainty quantified by a non-intuitive $N \times N$ covariance matrix. This is, however, rarely analyzed because it is hard to calculate and difficult to interpret. Still, there are approaches to estimate the error in observables derived by integrating over the spectrum. They are nicely discussed in [9] and should be used wherever possible.

We have presented the approaches to the analytic continuation problem in the order of increasing sophistication and accuracy — and numerical cost. The analysis of QMC data should ideally follow this progression until the desired information about the spectral function has been reliably obtained. A Picard plot will give a first impression of how much information is actually contained in the data and what can be expected from a straightforward linear regularization. Despite the uncontrolled approximations in the practical flavors of MaxEnt, the approach has developed into the standard approach for analytical continuation. The average spectrum method, that is now numerically competitive, provides an appealing alternative since it makes all assumptions via the choice of the discretization explicit, while being numerically exact.

The most important lesson is that results of analytic continuation must not be overinterpreted. When the results depend on the details of the method, they rather reflect the choices made by the approach than the data. Thus before interpreting details of the spectral function, we have to make sure that they are robust under (reasonable) variations in the regularization parameters. The discrepancy principle and the fit histogram are practical methods for doing this.

A Technical appendices

A.1 Blocking method for correlated data

Let us assume we have an ergodic Markov chain Monte Carlo method, e.g., using Metropolis sampling, that generates a set of K data points $m_1, \ldots, m_K$ drawn from a probability distribution $p(m)\,dm$ and we are interested in the mean value $\mu = \int dm\, p(m)\, m$. The obvious estimate for μ is the average $\bar{m} = \sum_{k=1}^{K} m_k/K$. It will, of course, be different for different Monte Carlo runs, but, by the central limit theorem, for large K the averages $\bar{m}$ of different runs will tend to be distributed as a Gaussian centered at μ with variance

$$\sigma^2(\bar{m}) = \langle \bar{m}^2 \rangle - \langle \bar{m} \rangle^2 = \frac{1}{K^2} \sum_{k,l=0}^{K} \Big(\langle m_k m_l \rangle - \langle m_k \rangle \langle m_l \rangle \Big), \tag{73}$$

where $\langle \cdot \rangle$ is the average over all possible Monte Carlo runs producing K data points. How can we estimate $\sigma^2(\bar{m})$ from the simulation data of a single run? Splitting the double sum

$$\sigma^2(\bar{m}) = \frac{1}{K} \underbrace{\frac{1}{K} \sum_{k=1}^{K} \Big(\langle m_k^2 \rangle - \mu^2 \Big)}_{= \langle m^2 \rangle - \mu^2 =: s_0} + \frac{1}{K^2} \sum_{k \neq l} \Big(\langle m_k m_l \rangle - \mu^2 \Big) \tag{74}$$

we see that for uncorrelated data, $\langle m_k m_l \rangle = \langle m_k \rangle \langle m_l \rangle$ for $k \neq l$, the variance is given by s_0/K. But in general, samples obtained from Markov chain Monte Carlo will be positively correlated, so that $\sigma^2(\bar{m}) \geq s_0/K$. We can eliminate this correlation using an elegant renormalization group technique [13]. For this we consider the transformation of the original data set $m_1, \ldots, m_K$ of K samples (assuming K is even) into half as many data points, obtained by averaging

$$\hat{m}_{\hat{k}} := \frac{m_{2\hat{k}-1} + m_{2\hat{k}}}{2}\,. \tag{75}$$

Obviously, the average of the new data points $\sum_{\hat{k}=1}^{K/2} \hat{m}_{\hat{k}}/(K/2)$ is still $\bar{m}$ and thus must have the same distribution as the averages of the original data. Consequently, $\sigma^2(\bar{m})$ must remain invariant under the blocking transformation (75). Looking at the uncorrelated part of the variance for the blocked data $\hat{m}_{\hat{k}}$ and remembering that the ensemble average $\langle m_k^2 \rangle$ is independent of k, we see that

$$\hat{s}_0 = \frac{1}{K/2} \sum_{\hat{k}=1}^{K/2} \left(\frac{\langle m_{2\hat{k}-1}^2 + 2 m_{2\hat{k}-1} m_{2\hat{k}} + m_{2\hat{k}}^2 \rangle}{4} - \mu^2 \right) = \frac{s_0}{2} + \frac{1}{2\,K/2} \sum_{\hat{k}=1}^{K/2} \big(\langle m_{2\hat{k}-1} m_{2\hat{k}} \rangle - \mu^2 \big) \tag{76}$$

contains part of the correlations not contained in s_0. Therefore $\hat{s}_0/(K/2) \geq s_0/K$. Under repeated blocking transformations the uncorrelated part of the variance will thus increase. When it reaches a plateau, i.e., a fixed-point under the blocking transformation, it becomes equal to $\sigma^2(\bar{m})$ and the blocked data has become uncorrelated.

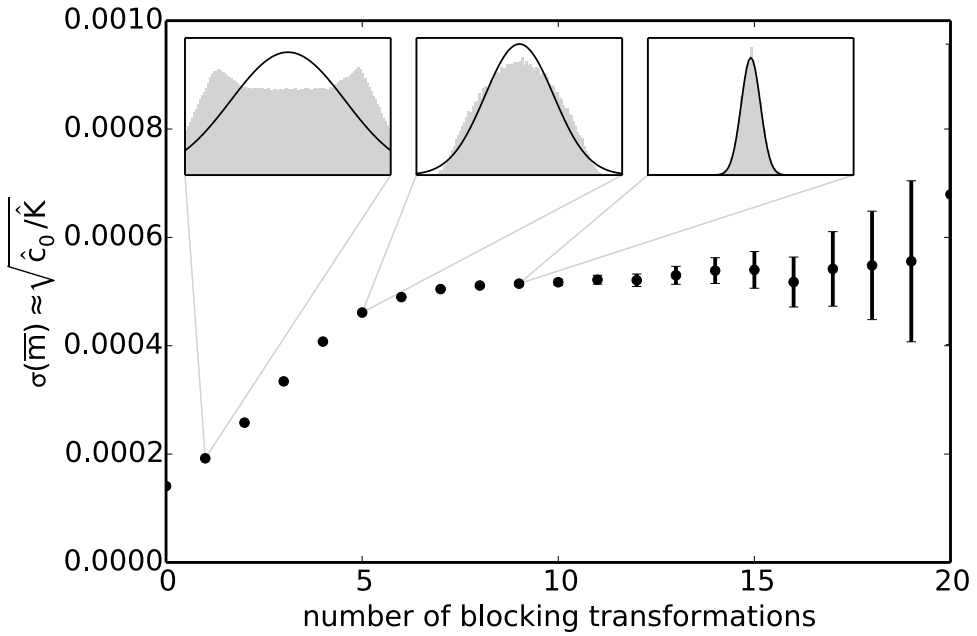


Fig. 9: *Estimate of the standard deviation of the average of correlated data obtained with the blocking method. Initially the variance is severely underestimated but the estimate increases with each blocking step until a plateau is reached, at which point the blocked data has become uncorrelated. By that time the distribution of the $\hat{m}_{\hat{k}}$ has become Gaussian of width $\hat{c}_0$, as shown in the insets. Eventually the number of blocked samples, $\hat{K} = K/2^n$, become so small that the estimates become unreliable.*

We can try to estimate the ensemble average s_0 from the data from one specific simulation run as $\sum_{k=1}^{K}(m_k^2 - \bar{m}^2)/K$. Taking the ensemble average and comparing to s_0

$$\frac{1}{K}\sum_{k=1}^{K}\left(\langle m_k^2\rangle - \langle\bar{m}^2\rangle\right) = \frac{1}{K}\sum_{k=1}^{K}\Big(\big(\langle m_k^2\rangle - \langle\bar{m}\rangle^2\big) - \big(\langle\bar{m}^2\rangle - \langle\bar{m}\rangle^2\big)\Big) = s_0\left(1-\frac{1}{K}\right)$$

we find that the unbiased estimator actually is

$$s_0 \approx c_0 := \frac{1}{K-1}\sum_{k=1}^{K}\big(m_k^2 - \bar{m}^2\big) \quad\Rightarrow\quad \sigma^2(\bar{m}) \approx \frac{1}{K(K-1)}\sum_{k=1}^{K}\big(m_k^2 - \bar{m}^2\big). \tag{77}$$

In an actual implementation of the blocking method, we repeatedly block the data and calculate the corresponding estimator of the uncorrelated variance $\hat{s}_0/\hat{K}$. An example is shown in Fig. 9. As expected, $\hat{c}_0/\hat{K}$ increases with each blocking step until it reaches a plateau. There the blocked data $\hat{m}_{\hat{k}}$ are uncorrelated and, by the central limit theorem, approach Gaussian variables of variance $\sigma^2(\hat{m}) = \hat{K}\sigma^2(\bar{m})$. For such variables the variance of the variance $\sigma^2(\bar{m})$ is given by $\langle(\hat{c}_0/\hat{K})^2\rangle - \langle\hat{c}_0/\hat{K}\rangle^2 = 2\sigma^4(\bar{m})/(\hat{K}-1)$, which provides us with the errorbars. Since the number of blocked data points is halved in each step, eventually the blocked sample becomes very small and $\hat{c}_0/\hat{K}$ starts to fluctuate, also indicated by rapidly increasing errorbars. We can then identify the plateau by checking when $\hat{s}_0/\hat{K}$ does not change between blocking steps within its error bar.

A.2 Non-negative least-squares algorithm (NNLS)

The model $\boldsymbol{f}$ fitting a given data vector $\boldsymbol{g}$ best in the least-squares sense minimizes the norm of the residual vector $\chi^2(\boldsymbol{f}) = \|\boldsymbol{K}\boldsymbol{f} - \boldsymbol{g}\|^2$. At the minimum $\boldsymbol{f}_{\mathrm{LS}}$ the gradient $\boldsymbol{w}$ vanishes

$$w_n(\boldsymbol{f}_{\mathrm{LS}}) := \frac{1}{2}\,\frac{\partial\,\chi^2(\boldsymbol{f})}{\partial f_n}\bigg|_{\mathrm{LS}} = \mathrm{Re}\left(\boldsymbol{K}^\dagger(\boldsymbol{K}\boldsymbol{f}_{\mathrm{LS}}-\boldsymbol{g})\right)_n = 0 \quad \forall n\,. \tag{78}$$

Since χ^2 is a non-negative quadratic form in $\boldsymbol{f}$ stationary points must be minima

$$\chi^2(\boldsymbol{f}_{\mathrm{LS}}+\boldsymbol{\delta}) = \chi^2(\boldsymbol{f}_{\mathrm{LS}}) + 2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{w}(\boldsymbol{f}_{\mathrm{LS}}) + \|\boldsymbol{K}\boldsymbol{\delta}\|^2 \geq \chi^2(\boldsymbol{f}_{\mathrm{LS}})\,. \tag{79}$$

The least-squares fit can be found from the singular value decomposition (SVD) $\boldsymbol{K}^\dagger = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^\dagger$

$$\boldsymbol{f}_{\mathrm{LS}} = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\dagger\boldsymbol{g}\,, \tag{80}$$

where the diagonal matrix $\boldsymbol{D}$ has dimension $K = \mathrm{rank}(\boldsymbol{K})$ while the matrix $\boldsymbol{U}$ is $M\times K$ and $\boldsymbol{V}$ is $N\times K$-dimensional. In terms of the SVD the gradient (78) is given by

$$\boldsymbol{w}(\boldsymbol{f}) = \mathrm{Re}\,\boldsymbol{V}\boldsymbol{D}(\boldsymbol{D}\boldsymbol{V}^{\mathsf{T}}\boldsymbol{f} - \boldsymbol{U}^\dagger\boldsymbol{g})\,. \tag{81}$$

This way of calculating the gradient is numerically more stable than calculating it directly in terms of $\boldsymbol{K}$. It also immediately shows that the gradient vanishes for $\boldsymbol{f}_{\mathrm{LS}}$.
Finding the best fit, $\min\|\boldsymbol{K}\boldsymbol{f} - \boldsymbol{g}\|^2$, under the constraint $\boldsymbol{f} \geq 0$ (non-negative least-squares, NNLS) is more complicated. When all components of the unconstrained solution are non-negative, $(\boldsymbol{f}_{\mathrm{LS}})_n \geq 0$, it is obviously also the solution of the constrained problem. When there are components $(\boldsymbol{f}_{\mathrm{LS}})_n < 0$ we might expect that the constrained fit assumes its minimum on the boundary, $(\boldsymbol{f}_{\mathrm{NNLS}})_n = 0$, where the gradient is positive $w_n > 0$. These are the Karush-Kuhn-Tucker conditions [14]:

$$f_n > 0 \quad \text{and} \quad w_n = 0 \qquad \text{or} \qquad f_n = 0 \quad \text{and} \quad w_n \geq 0\,. \tag{82}$$

We distinguish the two cases by defining the two sets $\mathcal{P} = \{n|\, f_n > 0\}$ and $\mathcal{Z} = \{n|\, f_n = 0\}$ which partition the set of indices, $\mathcal{P} \cup \mathcal{Z} = \{1, \ldots, N\}$.
When $\boldsymbol{f}_{\mathrm{KT}}$ fulfills the Karush-Kuhn-Tucker conditions, it minimizes $\chi^2(\boldsymbol{f})$ on $\boldsymbol{f} \geq 0$. To see this we consider a vector $\boldsymbol{f}_{\mathrm{KT}} + \boldsymbol{\delta}$ with $\delta_z \geq 0$ so that $\boldsymbol{f}_{\mathrm{KT}} + \boldsymbol{\delta} \geq 0$. Then

$$\chi^2(\boldsymbol{f}_{\mathrm{KT}}+\boldsymbol{\delta}) = \chi^2(\boldsymbol{f}_{\mathrm{KT}}) + 2\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{w}(\boldsymbol{f}_{\mathrm{KT}}) + \|\boldsymbol{K}\boldsymbol{\delta}\|^2 \geq \chi^2(\boldsymbol{f}_{\mathrm{KT}}) \tag{83}$$

since $\boldsymbol{\delta}^{\mathsf{T}}\boldsymbol{w} = \sum_n \delta_n w_n = \sum_{n\in\mathcal{P}} \delta_n w_n + \sum_{n\in\mathcal{Z}} \delta_n w_n \geq 0$, where the first sum vanishes because of the gradient, while in the second sum both factors in each term are non-negative. Conversely, when $\boldsymbol{f}_{\mathrm{NNLS}}$ solves the non-negative least-squares problem it must fulfill the Karush-Kuhn-Tucker condition, otherwise an infinitesimal change (respecting non-negativity) of a component violating it could lower χ^2. Thus, to solve the NNLS problem we just have to find a vector that fulfills the Karush-Kuhn-Tucker conditions. For this we can simply go through all possible partitionings of the indices $\{1, \ldots, N\} = \mathcal{P} \cup \mathcal{Z}$. For a given partitioning we determine the

least-squares solution on the indices in $\mathcal{P}$, i.e., we minimize $\|\boldsymbol{K}\boldsymbol{P}_\mathcal{P}\boldsymbol{f}-\boldsymbol{g}\|^2$, where $\boldsymbol{P}_\mathcal{P}$ is the projector to the space spanned by the components in $\mathcal{P}$. This makes sure that the gradients for these components vanish, $w_p=0$, while $f_z=0$. If also $f_p\geq 0$ for all $p\in\mathcal{P}$ and $w_z\geq 0$ for all $z\in\mathcal{Z}$, we have found the NNLS solution, otherwise we try the next partitioning. The only problem is that there are 2^N partitionings of the N indices (each index can be either in $\mathcal{P}$ or $\mathcal{Z}$).

A practical algorithm [14] considers possible partitionings in a much more efficient way. We start from some partition for which $\boldsymbol{f}_\mathcal{P}>0$ and $\boldsymbol{w}_\mathcal{P}=0$, e.g., an empty positive set, $\mathcal{P}=\{\}$ and $\boldsymbol{f}_{\{\}}=0$. Given a set $\mathcal{P}$ and the corresponding $\boldsymbol{f}_\mathcal{P}$ for which the Karush-Kuhn-Tucker condition is not yet fulfilled, we add the component i with the most negative gradient. Least-squares fitting on the expanded set $\mathcal{P}'{=}\mathcal{P}\cup\{i\}$ will produce an improved fit $\chi^2(\boldsymbol{f}_{\mathcal{P}'}){<}\chi^2(\boldsymbol{f}_\mathcal{P})$: because of the negative gradient, the new component will not stay at zero but rather take a positive value. In case $\boldsymbol{f}_{\mathcal{P}'}\geq 0$, we calculate the new gradient. If it is non-negative, we have found the Karush-Kuhn-Tucker solution, otherwise we repeat the procedure. Each iteration will produce a non-negative solution with improved fit, so that we will converge to the minimum of $\chi^2(\boldsymbol{f})$ under the constraint $\boldsymbol{f}\geq 0$.

In general, however, the least squares solution $\boldsymbol{f}_{\mathcal{P}'}$ will have negative components. In this case, we can find a mixing with the previous fit $\boldsymbol{f}_\alpha=(1{-}\alpha)\boldsymbol{f}_\mathcal{P}{+}\alpha\boldsymbol{f}_{\mathcal{P}'}$ with $\alpha\in(0,1)$, that brings the most negative component of $\boldsymbol{f}_{\mathcal{P}'}$ to zero. Since $\chi^2(\boldsymbol{f}_\alpha)\leq(1{-}\alpha)\chi^2(\boldsymbol{f}_{\mathcal{P}'})+\alpha\chi^2(\boldsymbol{f}_\mathcal{P})<\chi^2(\boldsymbol{f}_\mathcal{P})$ the fit will still be improved. We remove the components where $\boldsymbol{f}_\alpha$ vanishes from $\mathcal{P}'$ and perform a least-squares fit on the new set, repeating the procedure until we get a non-negative least squares solution. This must happen after a finite number of iterations, since in each step at least one element is removed from the positive set while the resulting $\boldsymbol{f}_\alpha\geq 0$ keeps improving the fit $\chi^2(\boldsymbol{f}_\alpha)<\chi^2(\boldsymbol{f}_\mathcal{P})$. Thus we can continue the outer loop with $\boldsymbol{f}_\alpha$, calculating the new gradient and adding the component where it is most negative to the positive set.

Since each step produces a $\boldsymbol{f}\geq 0$ with improved fit, the algorithm does not visit any partitioning twice and will thus always converge. At worst it may take 2^N steps, but in practice the down-hill search produces a solution after trying less than hundred partitionings. Numerically, the most delicate part is the calculation of the gradient, which should be stabilized using a factorization of the kernel. Obviously, checking the Karush-Kuhn-Tucker condition for the gradient must take the numerical accuracy into account. Moreover, the implementation may not converge, when, after adding the component with the most negative gradient, the least-squares fit gives that component a negative value. This can only happen as a consequence of numerical errors. In this case we rather include the component with the second most negative gradient in $\mathcal{P}$.

Note that the non-negative least squares solution $\boldsymbol{f}_\mathrm{NNLS}$ is unique, unlike the least-squares solution (80), to which we can add any multiple of a vector with zero singular value without changing the fit. While the least-squares problem is thus ill-posed when there are vectors that do not contribute to the fit, these vectors play a crucial role in non-negative least squares fitting: They take values such that the modes that are important for the fit can approach their optimal value as closely as possible without violating the constraint. Thus NNLS is well posed.

function NNLS($\boldsymbol{K}$, $\boldsymbol{g}$)
  $\boldsymbol{f} \leftarrow 0$
  $\mathcal{Z} \leftarrow \{1, \ldots, N\}$ ▷ below we use the abbreviation $\mathcal{P} = \{1, \ldots, N\} \setminus \mathcal{Z}$
  loop
    $\boldsymbol{w} \leftarrow \boldsymbol{K}^{\dagger}(\boldsymbol{K}\boldsymbol{f} - \boldsymbol{g})$ ▷ for robust calculation use, e.g., SVD
    if $\boldsymbol{w}[\mathcal{Z}] \geq 0$ **then return** $\boldsymbol{f}$
    end if
    $i \leftarrow \operatorname{argmin}(\boldsymbol{w}[\mathcal{Z}])$ ▷ find component with most negative gradient
    $\mathcal{Z} \leftarrow \mathcal{Z} \setminus \{i\}$
    loop
      $\boldsymbol{f}' \leftarrow LS(\boldsymbol{K}\boldsymbol{P}\boldsymbol{f}, \boldsymbol{g})$ ▷ LS solution on components $\mathcal{P}$
        ▷ $\boldsymbol{f}_i > 0$, if not: numerical error in gradient! Do not pick i again this round
      if $\boldsymbol{f}'[\mathcal{P}] > 0$ **then**
        $\boldsymbol{f} \leftarrow \boldsymbol{f}'$
        break
      end if
      $\alpha \leftarrow \min\left\{ \frac{\boldsymbol{f}_i}{\boldsymbol{f}_i - \boldsymbol{f}'_i} \middle| i \in \mathcal{P} \wedge \boldsymbol{f}'_i \geq 0 \right\}$
      $\boldsymbol{f} \leftarrow (1-\alpha)\boldsymbol{f} + \alpha\boldsymbol{f}'$ ▷ now $\boldsymbol{f} \geq 0$ and $\boldsymbol{f}_i = 0$ for $i = \operatorname{argmin}$
      $\mathcal{P} \leftarrow \mathcal{P} \setminus \{i | \boldsymbol{f}_i = 0\}$
    end loop
  end loop
end function

Fig. 10: *Function that returns the non-negative least-squares solution $\boldsymbol{f} \geq 0$ of $\boldsymbol{g} = \boldsymbol{K}\boldsymbol{f}$.*

A.3 Shannon entropy

When developing *The Mathematical Theory of Communication*, Claude Shannon introduced the bit as the amount of information needed to decide between two equally probable events [15]. Receiving an unlikely (surprising) message should convey more information than receiving a likely one, and the information contained in two independent messages should be the sum of the information carried by each individually. These axioms lead to $-\log_2 p_i$ as the information contained in receiving a message of probability p_i. Summing over a set of M possible messages of probabilities p_i and weighting the information contained in them by their probability defines the average information or entropy of an information source

$$H(\{p_i\}) = -\sum p_i \log_2 p_i \,. \tag{84}$$

It gives a lower limit to the number of bits needed for encoding the N messages. The maximum number of bits, $\log_2 N$, is needed when we know least about which message to expect, i.e., when all probabilities are the same. In the opposite limit, when one of the messages is certain, we need not encode it at all. Thus the entropy of an information source measures our ignorance before receiving one of the possible messages.

Changing the base of the logarithm, $\log_b(x) = \log_2(x)/\log_2(b)$, for $b > 1$ simply multiplies the entropy by a positive constant, i.e., changes the units in which we measure information. For convenience, we use the natural logarithm, $\ln$, working in natural units, 1 nat $\approx$ 1.44 bits.

An alternative derivation [2] of (84) starts by considering microstates representing the probabilities as $p_i = n_i/M$ by placing N (distinguishable) objects into M bins. Since we can place any of the N objects into any of the M bins, there are M^N such states. The number of different ways of placing the objects into bins and obtaining the same set of $\{n_i\}$, i.e., the same macrostate, is also easily determined: We can pick any n_1 of the N objects and put them into the first bin. Then we can pick any n_2 of the remaining $N - n_1$ objects and put them into the second bin. The probability of realizing a macrostate $\{n_1, \dots, n_M\}$ is thus

$$\frac{1}{M^N}\binom{N}{n_1}\binom{N-n_1}{n_2}\binom{N-n_1-n_2}{n_3}\cdots\binom{n_M}{n_M} = \frac{1}{M^N}\,\frac{N!}{n_1!\,n_2!\cdots n_M!}\,.$$

Taking the logarithm and using the Stirling approximation $\ln n! \approx n\ln n - n$ we find

$$-N\ln M + N\ln N - N - \sum_{i=1}^{M}\big(n_i\ln n_i - n_i\big) = N\left(\ln\frac{1}{M} - \sum_i p_i \ln p_i\right),$$

which is proportional to the $H(\{p_i\})$ minus the entropy of a flat distribution $\{1/M\}$.
Subtracting the entropy of the flat distribution becomes crucial when taking the limit of a continuous probability distribution: encoding an infinite number of messages will, in general, take an infinite number of bits. Subtracting $-\log 1/M$ keeps the limit $M \to \infty$ finite. To see this, we discretize a continuous distribution $p(x)$ on an equidistant grid of M points, $p_i = p(x_i)\Delta x$ with $\Delta x = (x_{\max} - x_{\min})/M = (\int dx)/M$, and take the limit of the Riemann sum

$$-\sum_{i=1}^{M} p_i \ln\left(\frac{p_i}{1/M}\right) = -\sum_{i=1}^{M}\Delta x\, p(x_i)\ln\left(\frac{p(x_i)\Delta x}{1/M}\right) \to -\int dx\, p(x)\ln\left(\frac{p(x)}{1/\int dx}\right).$$

This defines the entropy of a distribution $p(x)$

$$H[p] = -\int dx\, p(x)\ln\left(\frac{p(x)}{1/\int dx}\right). \tag{85}$$

We can find the $p(x)$ that maximizes this functional from the variational principle. Remembering that the functional derivatives are defined by the expansion

$$H[p+\delta p] = H[p] + \int dx\,\frac{\delta H[p]}{\delta p(x)}\,\delta p(x) + \frac{1}{2}\int dx\,dx'\,\frac{\delta^2 H[p]}{\delta p(x')\delta p(x)}\,\delta p(x')\,\delta p(x) + \mathcal{O}^3(\delta p)$$

we read off the first variation from

$$H[p+\delta p] = -\int dx\,(p+\delta p)\big(\underbrace{\ln(p+\delta p)}_{=\ln p+\ln(1+\frac{\delta p}{p})=\ln p+\frac{\delta p}{p}+\mathcal{O}^2} + \ln\textstyle\int dx\big) = H[p] - \int dx\,\underbrace{\left(1+\ln\frac{p(x)}{1/\int dx}\right)}_{=-\frac{\delta H[p]}{\delta p(x)}}\delta p + \mathcal{O}^2,$$

where we have used $\ln(1+x) = x - x^2/2 + \cdots$. In second order we find $\delta^2 H[p]/\delta p(x')\,\delta p(x) = -2\delta(x-x')/p(x) \le 0$ so that the stationary points are maxima. Imposing normalization of the 0-th moment via a Lagrange parameter, the variational equation becomes

$$0 = \frac{\delta}{\delta p(x)}\,H[p] + \lambda_0\big(1 - \textstyle\int dx\, p(x)\big) = -1 - \ln p(x) - \ln\int dx - \lambda_0$$

which is solved by the constant distribution, where $\lambda_0 = -1$ is fixed by normalization

$$p(x) = \frac{1}{\int dx}\, \mathrm{e}^{-(1+\lambda_0)} = \frac{1}{\int dx}\,. \tag{86}$$

Inserting into (85) we find $H[1/\int dx] = 0$, as it must by construction.
Likewise, we can ask which distribution maximizes the entropy, when we know in addition its first moment $\mu = \int dx\, x\, p(x)$. The variational equation then contains two Lagrange parameters

$$0 = \frac{\delta}{\delta p(x)}\, H[p] + \lambda_0\big(1 - \int dx\, p(x)\big) + \lambda_1\big(\mu - \int dx\, x\, p(x)\big) = -1 - \ln \frac{p(x)}{1\int dx} - \lambda_0 - \lambda_1\, x$$

and we obtain a Boltzmann distribution

$$p_\mu(x) = \frac{1}{\int dx}\, \mathrm{e}^{-(1+\lambda_0+\lambda_1 x)}, \tag{87}$$

where λ_0 and λ_1 are fixed by solving the system $\int dx\, p_\mu(x) = 1$ and $\int dx\, x\, p_\mu(x) = \mu$. Likewise, when we also know the variance of $p(x)$, maximizing the entropy results in a Gaussian.
Given the entropy functional, it is natural to ask what happens under a change of variable. Remembering that density functions transform as $p(x)\, dx = p(z)\, dz$, we find

$$H[p] = -\int \frac{dx}{dz}\, dz\;\; p(z)\frac{dz}{dx}\;\; \ln\left(\frac{p(z)\frac{dz}{dx}}{1/\int dx}\right) = -\int dz\, p(z) \ln\left(\frac{p(z)}{\rho(z)}\right),$$

where we introduced $\rho(z)\, dz = dx/\int dx$. It reflects how the intervals on x change under the under transformation to z. When we define $\rho(x) = 1/\int dx$, we see that the form of the entropy functional is invariant under coordinate transformations

$$H[p\,|\,\rho] = -\int dx\, p(x) \ln \frac{p(x)}{\rho(x)}\,. \tag{88}$$

This is the relative entropy or Kullback-Leibler divergence. From $\ln x \leq x-1$ it follows that $H[p] \leq 0$. By construction, the maximum is attained for $p(x) = \rho(x)$. The relative entropy describes the average information contained in the distribution $p(x)$ when what we expected was the distribution $\rho(x)$. The prior $\rho(x)$ plays the role of a density of states: from the functional derivative of the relative entropy $\delta H[p]/\delta p(x) = -1 - \ln(p(x)/\rho(x))$ we see that the solutions of the variational equations for $p(x)$ derived above become proportional to $\rho(x)$.
For convenience we might want to allow non-normalized densities of states $\tilde\rho(x)$ and correspondingly drop the normalization constraint for $\tilde p(x)$. If we write [16]

$$\tilde H[\tilde p\,|\,\tilde\rho] = \int dx \left(\tilde p(x) - \tilde\rho(x) - \tilde p(x) \ln \frac{\tilde p(x)}{\tilde\rho(x)}\right) \tag{89}$$

we obtain from the variational equation $\delta\tilde H/\delta\tilde p = 0$ (without normalization constraint) the solution $\tilde p(x) = \tilde\rho(x)$ with $\tilde H[\tilde\rho\,|\,\tilde\rho] = 0$ as for (88).
We note that a flat prior $\rho(x) = \mathrm{const.}$ is bound to the choice of variable: Given any $\rho(z)$ we can always transform to $x(z) \propto \int^z \rho(z')\, dz'$ to obtain $\rho(x) = \rho(z)\, dz/dx = \mathrm{const.}$, where x must be restricted to a finite interval to be normalizable.

A.4 Sampling from a truncated normal distribution

It is straightforward to generate a random variable x with a normal probability distribution

$$p_n(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2} \tag{90}$$

using, e.g., the Box-Muller method [17]. When taking a constraint into account, we need, however, variables with a normal distribution restricted to some interval, $x \in [a, b]$.

$\boldsymbol{x \geq a}$: When x is restricted to be larger than some value a, a straightforward approach is to sample normally distributed values x until we find an $x > a$. The probability for finding such an x is, on average, just the integral over the Gaussian

$$\bar{P}_n(a) = \int_a^\infty dx\, p_u(x) = I(a)\,. \tag{91}$$

This is easily written in terms of the complementary error function $\mathrm{erfc}(z) = 2/\sqrt{\pi}\int_z^\infty dt\,\mathrm{e}^{-t^2}$. For $a > 0$

$$I(a \geq 0) = \frac{1}{2}\,\mathrm{erfc}(a/\sqrt{2})\,, \tag{92}$$

while for $a < 0$

$$I(a \leq 0) = 1 - I(-a)\,. \tag{93}$$

The average acceptance probability (91) is shown in figure 11. For a to the left of the peak of the normal distribution it is very likely that a proposed random variable x is larger than a and thus is accepted. For $a > 0$ this probability is, however, rapidly decreasing to zero, meaning that we would have to propose very many normally distributed variables x until we find one that is larger than a. This is very inefficient, so for $a > 0$ we need a better approach. Following [18], we generate random variables $x \geq a$ with an exponential probability distribution

$$p_{\mathrm{exp}}(x) = \alpha\,\mathrm{e}^{-\alpha(x-a)}\,. \tag{94}$$

These are easily obtained as $x = a - \ln(u/\alpha)/\alpha$ from uniformly distributed random numbers $u \in [0, 1)$. To transform these exponentially distributed random numbers $x \geq a$ into the desired normally distributed random numbers we use the rejection method [17], accepting x with probability proportional to the ratio $p_n(x)/p_{\mathrm{exp}}(x)$ of the desired and the proposed probability distribution functions. To obtain a probability, we introduce a prefactor to make sure that for no $x \geq a$ the ratio exceed one. Completing the square in the exponential we find

$$p_{\mathrm{acc}}(x; A) = \frac{1}{A}\,\frac{\mathrm{e}^{-x^2/2}}{\mathrm{e}^{-\alpha(x-a)}} = \frac{1}{A}\,\underbrace{\mathrm{e}^{-(x-\alpha)^2/2}}_{\leq 1}\,\mathrm{e}^{\alpha^2/2-\alpha a} \overset{!}{\leq} 1\,. \tag{95}$$

For $x \geq a$ the choice $A = \mathrm{e}^{\alpha^2/2-\alpha a}$ maximizes the acceptance probability, which becomes

$$p_{\mathrm{acc}}(x) = \mathrm{e}^{-(x-\alpha)^2/2}\,. \tag{96}$$

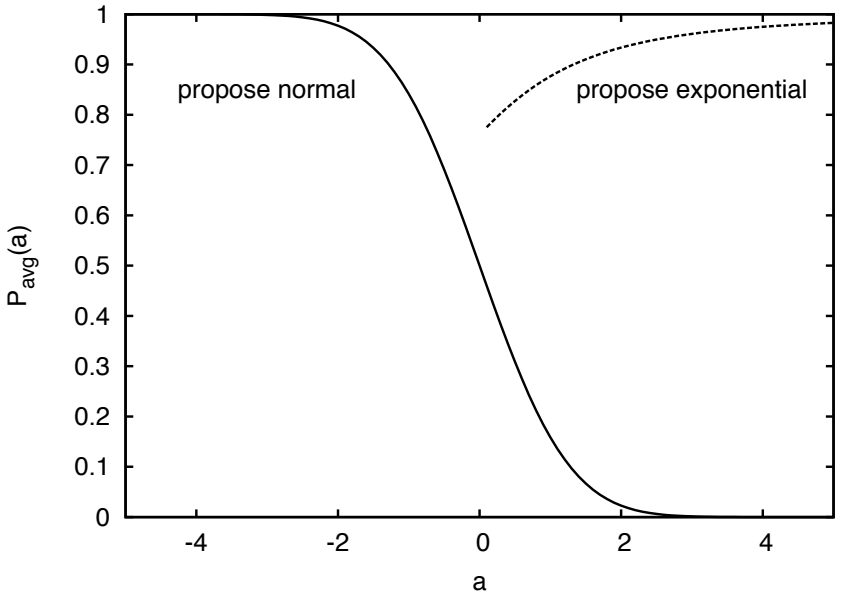


Fig. 11: *Efficiency of the methods for sampling from a normally distributed variable x restricted to the interval $[a, \infty)$: For $a < 0$ the average acceptance probability for an unrestricted normally distributed random variable ($\bar{P}_n(a)$, full line) is larger than $1/2$, while for $a > 0$ it rapidly approaches zero. For positive a we therefore propose exponentially distributed random numbers ($\bar{P}_{\exp}(a \geq 0)$, dotted line), for which the average acceptance probability is at least $\sqrt{\pi/2\mathrm{e}} \approx 0.76$.*

The corresponding average acceptance probability is then the integral of the product of the probability for proposing a value x times the probability for accepting it

$$\bar{P}_{\exp}(a \geq 0) = \int_a^\infty dx\, p_{\exp}(x)\, p_{\mathrm{acc}}(x) = \sqrt{2\pi}\, \alpha\, \mathrm{e}^{-\alpha^2/2+\alpha a}\, I(a) \tag{97}$$

In this expression α is still a free parameter, which we choose to maximize the average acceptance. Solving the variational equation we obtain

$$\alpha = \frac{a + \sqrt{a^2+4}}{2}\,. \tag{98}$$

We note that for $a \geq 0$, $\bar{P}_{\exp}(a)$ has the same form as (91), differing, however, by a prefactor $\gamma_{\exp} = \sqrt{2\pi}\, \alpha\, \mathrm{e}^{-\alpha^2/2+\alpha a}$ which grows faster than the complementary error function decays. Therefore, as can be seen from Fig. 11, for $a > 0$ this method is dramatically more efficient than sampling from an unbounded uniform distribution. Thus for $a < 0$ we choose the method with $\gamma_n = 1$, while for $a \geq 0$ we choose $\gamma_{\exp}$, obtaining an average acceptance probability $\bar{P}_\gamma(a) = \gamma(a)\, I(a)$.

$\boldsymbol{x \geq b}$: When we need to sample a normal variable constrained from above we can use the same methods as above, sampling $-x \geq -b$, with average acceptance probability $\bar{P}_\gamma(-b)$.

$\boldsymbol{a \leq x \leq b}$: When the random variable is constrained to a finite interval, an obvious approach is to first sample an $x \in [a, \infty)$ and to accept it if $x \leq b$. The average acceptance probability is then $\bar{P}_\gamma(a)$ for proposing an $x \in [a, \infty)$, times $\int_a^b dx\, p_n(x) / \int_a^\infty dx\, p_n(x)$ for accepting it, i.e.,

$$\bar{P}_\gamma(a, b) = \gamma\big(I(a) - I(b)\big)\,. \tag{99}$$

For large intervals this will be an efficient approach, while for narrow intervals the acceptance will go to zero. In this case it becomes more efficient to propose x uniformly distributed on $[a,b]$ and accept them with probability

$$p_{\rm acc} = \mathrm{e}^{-(x^2-m^2)/2}\,, \tag{100}$$

where m is the coordinate at which the normal distribution assumes its maximum value on $[a,b]$, ensuring that $p_{\rm acc} \le 1$. When $0 \in [a,b]$ then $m=0$, otherwise $m = \min(|a|,|b|)$. The average acceptance probability for this approach is

$$\bar{P}_u(a,b) = \int_a^b dx\, \frac{1}{b-a}\,\mathrm{e}^{-(x^2-m^2)/2} = \frac{\mathrm{e}^{m^2/2}}{b-a}\,\sqrt{2\pi}\big(I(a)-I(b)\big). \tag{101}$$

Again, (101) differs from (99) only by its prefactor γ_u, which increases as the width of the interval $b-a$ becomes smaller.
For a given interval $[a,b]$ we then choose the most efficient method:

- $a<0<b$: Since $a<0$ we have to choose between normal sampling with $\gamma_n = 1$ and uniform sampling with $\gamma_u = \sqrt{2\pi}/(b-a)$. For $\gamma_n = \gamma_u$ both methods have the same average acceptance probability. Solving this gives us the critical width $w_0 = \sqrt{2\pi}$. For intervals $b-a<w_0$ we thus use uniform sampling with γ_u, otherwise γ_n. The worst case is $\bar{P}_{\gamma=1}(0,w_0) = I(0)-I(w_0) = \mathrm{erf}(\sqrt{\pi})/2 \approx 0.49$.

- $0<a<b$: Since $a>0$ we choose between exponential sampling with $\gamma_{\rm exp} = \sqrt{2\pi}\,\alpha\,\mathrm{e}^{-\alpha^2/2+\alpha a}$ and $\gamma_u = \sqrt{2\pi}\,\mathrm{e}^{a^2/2}/(b-a)$. Solving $\gamma_{\rm exp} = \gamma_u$ gives the critical width $w_>(a) = e^{(\alpha-a)^2/2}/\alpha$. For intervals $b-a<w_>(a)$ we use uniform sampling with γ_u, otherwise $\gamma_{\rm exp}$. The worst case is $\bar{P}_\gamma(0,w_>(0)) = \gamma_{\rm exp}\big(I(0)-I(\sqrt{e})\big) = \sqrt{\pi/2e}\,\mathrm{erf}(\sqrt{\mathrm{e}/2}) \approx 0.68$.

- $a<b<0$: We sample $-x$ in the interval from $-b$ to $-a$ as described above.

Overall, we can thus sample from a truncated normal distribution with an average acceptance larger than $\mathrm{erf}(\sqrt{\pi})/2 \approx 0.49$.
The generalization to sampling from a Gaussian distribution $\exp(-(x-\mu)^2/2\sigma^2)/\sqrt{2\pi}\sigma$ of variance σ centered at μ restricted to $x\in[a,b]$ is then straightforward: use $x{=}\sigma\tilde{x}{+}\mu$, where $\tilde{x}$ is sampled, as described above, from a normal distribution (90) on the interval $[(a{-}\mu)/\sigma, (b{-}\mu)/\sigma]$.

References

[1] J. Hadamard, Princeton Univ. Bull. **13**, 49–52 (1902): ...ces problèmes se présentait comme parfaitement bien posé, je veux dire comme *possible* et *déterminé*.

[2] D.S. Sivia and J. Skilling: *Data Analysis: A Bayesian Tutorial* (Oxford Univ. Press, 2006)

[3] A. Flesch, E. Gorelov, E. Koch, and E. Pavarini, Phys. Rev. B **87**, 195141 (2013)

[4] A. Erdélyi (ed.): *Higher Transcendental Functions* (McGraw-Hill, New York, 1953)

[5] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, Phys. Rev. B **84**, 075145 (2011)

[6] E. Koch: *The Lanczos Method* in E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (eds.): *The LDA+DMFT approach to strongly correlated materials* Modeling and Simulation Vol. 1 (Forschungszentrum Jülich, 2011)

[7] J. Waldvogel inn: W. Gautschi, G. Mastroianni, T. Rassias (eds): *Approximation and Computation*, Springer Optimization and Its Applications, Vol. 42 (Springer, New York, 2010) p. 267

[8] P.C. Hansen: *Discrete Inverse Problems* (SIAM, Philadelphia, 2010)

[9] M. Jarrell: *The Maximum Entropy Method: Analytic Continuation of QMC Data* in E. Pavarini, E. Koch, F. Anders, and M. Jarrell (eds.): *Correlated Electrons: From Models to Materials* Modeling and Simulation Vol. 2 (Forschungszentrum Jülich, 2012)

[10] S.R. White in D.P. Landau, K.K. Mon, and B.B. Schüttler: *Computer Simulation Studies in Condensed Matter Physics III* (Springer, 1991) p. 145–153

[11] K. Ghanem: *Stochastic Analytic Continuation: A Bayesian Approach* PhD Thesis (RWTH Aachen, 2017)

[12] K. Ghanem and E. Koch, Phys. Rev. B **101**, 085111 (2020); *ibid.* **102**, 035114 (2020)

[13] H. Flyvbjerg and H.G. Petersen, J. Chem. Phys. **91**, 461(1989)

[14] C.L. Lawson and R.J. Hanson: *Solving Least Squares Problems* (SIAM, 1995)

[15] C. Shannon, Bell Syst. Tech. J. **27**, 379 and 623 (1948)

[16] J. Skilling, in J. Skilling (ed.): *Maximum Entropy and Bayesian Methods* (Kluwer, Dordrecht, 1989) p. 45–52

[17] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery: *Numerical Recipes* (Cambridge University Press, 2007)

[18] C.P. Robert Statistics and Computing **5**, 121 (1995)

7 LDA+DMFT for Strongly Correlated Materials

Alexander Lichtenstein
I. Institut für Theoretische Physik
Universität Hamburg, 20355 Hamburg

Contents

1 DFT and DMFT: Role of reference systems **2**

2 Functional approach **4**

3 Density Functional Theory **6**

4 Dual Fermion approach with a general reference system **8**

5 Perturbation in Dual Space **12**

6 LDA+DMFT scheme for real materials **15**

7 Conclusions **19**

A Path integrals for fermions **20**

1 DFT and DMFT: Role of reference systems

In this lecture we give an introduction to the theoretical description of interacting electron system based on non-perturbative, strong-coupling expansions around optimal reference systems. The Density Functional Theory (DFT) and its Local Density Approximation (LDA) based on the simplest reference system related to the homogeneous electron gas with constant external potential with the same Coulomb electron-electron interactions (see Fig. 1). Such a reference system can be solved via the numerically exact diffusion Quantum Monte Carlo scheme for the ground state energy as function of electron density [1]. On other hand, the Dynamical Mean-Field Theory (DMFT) for strongly interacting fermionic systems is based on the strong coupling expansion around an effective impurity reference system (Fig. 1). This scheme become exact in the limit of infinite lattice dimension ($z \to \infty$) [2].

For finite lattice dimension we can start from the DMFT reference system and use analytical perturbation theory for non-local correlation effects. The frequency dependent effective impurity DMFT problem can nowadays be efficiently solved within the continuous-time Quantum Monte Carlo (ct-QMC) scheme [3]. Therefore the perturbation theory needs to be formulated in the action path integral formalism. We give a brief introduction to the path integral over fermionic Grassmann fields and formulate a general scheme for the expansion around the DMFT solution using a special dual space transformation. We discuss here a general way to include nonlocal correlations beyond DMFT [4] which based on the dual-fermion path-integral formalism [5].

Consider the noninteracting, "kinetic" part H_t of the Hubbard model first [6]. This is fixed by specifying the hopping-matrix elements t_{ij} between sites i and j. In the absence of the local Hubbard-interaction term, H_t is easily diagonalized. For a Hubbard model on a lattice, diagonalization is achieved by Fourier transformation of hopping parameters to $\mathbf{k}$-space, and one has the normal "band structure". For a single orbital model it is particularly simple: $\varepsilon_{\mathbf{k}} = t_{\mathbf{k}}$ with band width W. If, on the other hand, only the local part of the Hamiltonian is kept, i.e., the Hubbard interaction H_U with interaction strength U and the local term of H_t fixed by the on-site energy ε_0, the diagonalization of the Hamiltonian is trivial again and reduces to the diagonalization of a single "Hubbard atom".

The great success of the DMFT approach is related with the interpolation between these two limits [7]. For the half-filled Hubbard model on the infinite-dimension Bethe lattice at half-filling the DMFT gives the exact description of the Mott-transition [8] between the weak-coupling ($U/W \ll 1$) metallic state and the strong-coupling ($U/W \gg 1$) insulating paramagnetic state [9]. In a nutshell, DMFT maps the correlated Hubbard lattice problem onto the self-consistent solution of an effective Anderson impurity problem with a single interacting Hubbard atom (interaction strength U) in a non-interacting fermionic bath (which mimics the rest of the crystal).

Now we can think how to incorporate nonlocal correlations beyond the DMFT: since the Hubbard and the Anderson-impurity model share the same interaction part, one can think of the Hubbard model as the impurity model plus a residual term $\propto (t_{\mathbf{k}} - \Delta_\nu)$ and treat this *perturbatively*. Since this term is frequency dependent we need a novel perturbation theory based on the

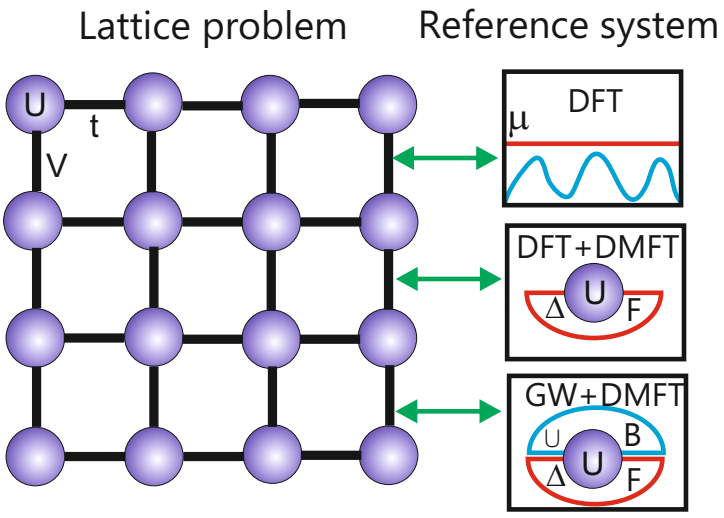


Fig. 1: *Schematic representation of reference systems in many-body approaches to lattice-fermion models: (i) Density-functional theory (DFT) with the interacting homogeneous electron gas as a reference system, defined by a constant external potential μ. (ii) Dynamical mean-field theory (DMFT) with an effective impurity problem as a reference system, defined by a fermionic bath, specified by a hybridization function Δ. (iii) GW+DMFT with correlated atom in a fermionic (Δ) and a bosonic bath (Λ) due to effects of the frequency-dependent screening of long-range Coulomb (V) interactions.*

action formalism. One may view this idea as a generalization of the Kohn-Sham idea in density functional theory (DFT) [10, 11] of an optimal reference system, but with a crucial difference. Here, not an interacting homogeneous electron gas, but an effective impurity model, tailored to the problem of strong correlations, serves as the reference system, see Fig. 1. Since at the zeroth order of this perturbative expansion, i.e., on the level of the DMFT problem, we already have an interacting problem and since the perturbation is momentum and frequency dependent, one is forced to replace the Hamiltonians by actions within the path-integral formalism. Note that the fermion path integral can also be used to formulate the DMFT itself [7,13]. Now, the separation of the local and nonlocal terms is achieved by a Hubbard-Stratonovich transformation applied to the single-particle $(t_{\mathbf{k}} - \Delta_\nu)$-term [5]. This provides us with a new action. Moreover, it is formally possible to integrate out the original local degrees of freedom and in this way generate an effective action in the transformed, so-called dual-fermion representation [5]. Note that integrating out the local degrees of freedom is not only a formal step but can be achieved in practice, by solving the impurity problem within the numerically exact ct-QMC method.

The dual action consists of a bare dual propagator (non-local part of the DMFT Green function) $\tilde{G}^0_{\mathbf{k},\nu}$ and a local but frequency-dependent effective potential related to scattering processes of two, three, and more dual particles on the impurity site. The simplest two-particle dual potential coincides with the fully connected part of the screened impurity interaction vertex $\gamma^{\omega}_{\nu\nu'}$, which can be calculated with the impurity ct-QMC solver as a function of bosonic (ω) and fermionic (ν, ν') Matsubara frequencies. Normally, correlations between three particles on the DMFT impurity site are much weaker than two-particle correlations and can be ignored. The same ap-

plies to higher-order terms. One can think of the dual-fermion formalism as an expansion in the order of local multi-particle correlation functions. This means that "bare" interactions between dual fermions are related with the connected part of the screened impurity vertex. Standard diagrammatic techniques can be applied for calculations of the bold dual propagator $\tilde{G}_{\mathbf{k},\nu}$, which allows to obtain the nonlocal self-energy for the original fermions [5] and to describe nonlocal correlations beyond the DMFT.

The dual-fermion approach is not necessarily bound to a specific starting point. However, the DMFT starting point is very efficient. Namely, it corresponds to the elimination of all local diagrams for any n-particle correlation of dual fermions when using the DMFT self-consistency equation, Eq. (46). In the dual space, this simply reduces to $\sum_{\mathbf{k}} \tilde{G}^0_{\mathbf{k},\nu} = 0$ and means that, on average over the whole Brillouin zone, Δ_ν optimally approximates the electron spectrum $\varepsilon_{\mathbf{k}}$, including its local correlation effects. Therefore, the noninteracting dual fermions correspond to strongly correlated DMFT quasiparticles, and the remaining nonlocal effects can be quite small and reasonably described by perturbative summations of dual diagrams. This also explains the notion "dual fermions".

2 Functional approach

We introduce a general functional approach which will cover the DFT, Dynamical Mean Field Theory (DMFT) and Baym-Kadanoff (BK) theories [12, 13]. Let us start from the full many–body Hamiltonian describing electrons moving in the periodic external potential of ions $V(\mathbf{r})$ with the chemical potential μ and interacting via Coulomb law: $U(\mathbf{r}-\mathbf{r}') = 1/|\mathbf{r}-\mathbf{r}'|$. We use the atomic units $\hbar = m = e = 1$. In the field-operator representation, the Hamiltonian has the form

$$H = \sum_\sigma \int d\mathbf{r}\, \hat{\psi}^\dagger_\sigma(\mathbf{r}) \left(-\frac{\nabla^2}{2} + V(\mathbf{r}) - \mu \right) \hat{\psi}_\sigma(\mathbf{r}) + \frac{1}{2} \sum_{\sigma\sigma'} \int d\mathbf{r} \int d\mathbf{r}' \hat{\psi}^\dagger_\sigma(\mathbf{r}) \hat{\psi}^\dagger_{\sigma'}(\mathbf{r}')\, U(\mathbf{r}-\mathbf{r}')\, \hat{\psi}_{\sigma'}(\mathbf{r}') \hat{\psi}_\sigma(\mathbf{r}) \tag{1}$$

We can always use the single-particle orthonormal basis set in solids $\varphi_n(\mathbf{r})$ for example Wannier orbitals with full set of quantum numbers, e.g., site, orbital and spin index: $k = (i, m, \sigma)$ and expand the fields in creation and annihilation operators

$$\hat{\psi}(\mathbf{r}) = \sum_n \varphi_n(\mathbf{r})\, \hat{c}_n \qquad \hat{\psi}^\dagger(\mathbf{r}) = \sum_n \varphi^*_n(\mathbf{r})\, \hat{c}^\dagger_n. \tag{2}$$

Going from fermionic operators to the Grassmann variables $\{c^*_k, c_k\}$ we can write the functional integral representation for partition function of the many-body Hamiltonian in the imaginary time domain using the Euclidean action S

$$Z = \int \mathcal{D}[c^*, c]\, e^{-S}\,, \qquad S = \sum_{12} c^*_1 (\partial_\tau + t_{12}) c_2 + \frac{1}{2} \sum_{1234} c^*_1 c^*_2\, U_{1234}\, c_4 c_3, \tag{3}$$

where the one- and two-electron matrix elements are defined as

$$t_{12} = \int d\mathbf{r}\, \varphi_1^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \mu \right) \varphi_2(\mathbf{r}) \tag{4}$$

$$U_{1234} = \int d\mathbf{r} \int d\mathbf{r}'\, \varphi_1^*(\mathbf{r}) \varphi_2^*(\mathbf{r}')\, U(\mathbf{r}-\mathbf{r}')\, \varphi_3(\mathbf{r}) \varphi_4(\mathbf{r}').$$

and we use the following short definition of the sum:

$$\sum_1 \cdots \equiv \sum_{im} \int d\tau\, \cdots \tag{5}$$

The one-electron Green function defined via a simplest non-zero correlation function for fermions [14] can be written in the path-integral formalism [15–17] as

$$G_{12} = -\left\langle c_1 c_2^* \right\rangle_S = -\frac{1}{Z} \int \mathcal{D}[c^*, c]\, c_1 c_2^*\, \exp(-S). \tag{6}$$

The Baym-Kadanoff functional [12] gives the one–particle Green function and the total free energy at its stationary point. In order to construct the exact functional of the Green function (Baym-Kadanoff) we modify the action by introducing the source term J

$$S[J] = S + \sum_{12} c_1^* J_{12}\, c_2. \tag{7}$$

The partition function Z, or equivalently the free energy of the system F, becomes a functional of the auxiliary source field

$$Z[J] = e^{-F[J]} = \int \mathcal{D}[c^*, c]\, e^{-S'[J]}. \tag{8}$$

Variation of this source function gives all correlation functions, for example the Green function

$$G_{12} = \left. \frac{\delta F[J]}{\delta J_{21}} \right|_{J=0}. \tag{9}$$

The Baym-Kadanoff functional Γ is obtained by the Legendre transform from variable J to G

$$\Gamma[G] = F[J] - \mathrm{Tr}\, JG, \tag{10}$$

This is a non-trivial step and uses the possibility of inverting the Eq. (9) for arbitrary J and find $J = J(G)$ which may be multivalued for strongly correlated systems [18].
We can use the standard decomposition of the free energy functional Γ into the single particle part and the correlated part

$$\Gamma[G] = \mathrm{Tr} \ln G - \mathrm{Tr}\, \Sigma G + \Phi[G], \tag{11}$$

were Σ is single particle self-energy and $\Phi[G]$ is a correlated part of the Baym-Kadanoff functional and is equal to the sum of all two-particle irreducible diagrams. At the stationary point this functional gives the free energy of the system. In practice, $\Phi[G]$ is not known for interacting electron systems, which is similar to the problem in the density functional theory. There is a formal expression for the "exchange-correlation" part of this functional via all connecting two-particle diagrams [13, 19, 20].

3 Density Functional Theory

The general functional approach reduces to density-functional theory, if one only uses the diagonal part in space-time of the Green function, which corresponds to the one-electron density

$$n(1) = G_{12}\,\delta_{12} = \langle c_1^* c_1 \rangle_S, \tag{12}$$

Consider a general interacting Hamiltonian with "λ-scaled" interaction

$$\hat{H} = \hat{T} + \lambda \hat{U}, \tag{13}$$

which depends on the coupling constant λ as a parameter. The same is true for the effective action functional [20]

$$\Gamma = \Gamma\left[n, \lambda\right]. \tag{14}$$

Clearly n, λ are to be considered as two independent variables. Note, however, that this does not prevent the exact expectation value n_g from depending on λ: this dependence is fixed by the variational principle

$$\left(\frac{\delta \Gamma\left[n, \lambda\right]}{\delta n}\right)_{n_g} = 0. \tag{15}$$

The functional $\Gamma\left[n, \lambda\right]$ is defined as

$$\Gamma\left[n, \lambda\right] = F\left[J, \lambda\right] - J\left(1\right) n\left(1\right), \tag{16}$$

where J is a functional of n and λ. This functional dependence is provided by the equation

$$\frac{\delta F\left[J, \lambda\right]}{\delta J\left(1\right)} = n\left(1\right).$$

We can find a formally exact expression for the $\Gamma[n]$ functional within the inversion method [19]. Let us expand all the quantities in Eqn (16) in terms of λ;

$$\begin{aligned} J\left[n, \lambda\right] &= J_0\left[n\right] + \lambda J_1\left[n\right] + \lambda^2 J_2\left[n\right] + \ldots, \\ F\left[J, \lambda\right] &= F_0\left[J\right] + \lambda F_1\left[J\right] + \lambda^2 F_2\left[J\right] + \ldots, \\ \Gamma\left[n, \lambda\right] &= \Gamma_0\left[n\right] + \lambda \Gamma_1\left[n\right] + \lambda^2 \Gamma_2\left[n\right] + \ldots. \end{aligned} \tag{17}$$

Comparison of the two sides in Eqn (16) for different orders of λ,

$$\sum \lambda^i \Gamma_i\left[n\right] \;=\; \sum \lambda^i F_i \left(\sum \lambda^k J_k[n]\right) - \sum \lambda^i J_i\left(1\right) n\left(1\right), \tag{18}$$

leads to the formally exact expression for $\Gamma_i\left[n\right]$,

$$\begin{aligned} \Gamma_i\left[n\right] = F_i\left[J_0\right] &+ \sum_{k=1}^{i} \frac{\delta F_{i-k}\left[J_0\right]}{\delta J_0\left(1\right)} J_k\left(1\right) - J_i\left(1\right) n\left(1\right) \\ &+ \sum_{m=2}^{i} \frac{1}{m!} \sum_{k_1,\ldots,k_m \geq 1}^{k_1+\ldots+k_m \leq i} \frac{\delta^m F_{i-(k_1+\ldots+k_m)}\left[J_0\right]}{\delta J_0\left(1\right) \ldots \delta J_0\left(m\right)} J_{k_1}\left(1\right) \cdots J_{k_m}\left(m\right). \end{aligned}$$

The functionals $\left\{F_i\left[J_0\right]\right\}$ and their derivatives are assumed to be known and obtained via a standard many-body Quantum Monte Carlo scheme. Since n and λ are considered to be independent, the functionals $J_i\left[n\right]$ can be obtained using,

$$\frac{\delta\Gamma_i\left[n\right]}{\delta n\left(1\right)} = -J_i\left(1\right). \tag{19}$$

Of special importance for the DFT scheme is the zero-th order term

$$\Gamma_0\left[n\right] = F_0\left[J_0\right] - J_0\left(1\right)n\left(1\right). \tag{20}$$

Using Eqn (19)

$$-J_0\left(1\right) = \frac{\delta F_0\left[J_0\right]}{\delta J_0\left(1'\right)}\frac{\delta J_0\left(1'\right)}{\delta n\left(1\right)} - J_0\left(1\right) - n\left(1'\right)\frac{\delta J_0\left(1'\right)}{\delta n\left(1\right)} \;\Rightarrow \left(\frac{\delta F_0\left[J_0\right]}{\delta J_0\left(1'\right)} - n\left(1'\right)\right)\frac{\delta J_0\left(1'\right)}{\delta n\left(1\right)} = 0.$$

Strict convexity of $\Gamma_0\left[n\right]$ prohibits $\left(\delta J_0\left(1'\right)/\delta n\left(1\right)\right)$ from having zero eigenvalues. Thus we obtain that J_0 obeys the equation:

$$n\left(1\right) = \frac{\delta F_0\left[J_0\right]}{\delta J_0\left(1\right)}. \tag{21}$$

Hence J_0 is determined as a potential which generates the expectation value n in the *noninteracting* $(\lambda = 0)$ system. Notice that the same exact notion appears in the Kohn-Sham formalism [11]. We refer to this noninteracting system as Kohn-Sham (KS) system and J_0 as Kohn-Sham potential.

The Kohn–Sham potential for interacting system $V_{\text{KS}} = V_{\text{ext}} + V_H + V_{\text{xc}}$ plays the role of the "constrained fields" J_i. Here V_{ext} is am external potential and V_H the Hartree potential. In principle, the exchange-correlation potential V_{xc} is known only for the homogeneous electron gas, therefore in all practical applications one uses a so-called local density approximation to DFT. In this case the DFT functional is defined in the following way

$$F_{\text{DFT}}[n] = T_0[n] + V_{\text{ext}}[n] + V_H[n] + V_{\text{xc}}[n] \tag{22}$$

where T_0 is the kinetic energy of the non-interacting systems. Finally, if we define the total electron density and exchange density as

$$n(\mathbf{r}) = \sum_k \varphi_k^*(\mathbf{r})\varphi_k(\mathbf{r}) \quad \text{and} \quad n(\mathbf{r},\mathbf{r}') = \sum_k \varphi_k^*(\mathbf{r})\varphi_k(\mathbf{r}') \tag{23}$$

the DFT approach can be formulated as

$$T_0[n] + V_{\text{ext}}[n] = \sum_k \int d\mathbf{r}\,\varphi_k^*(\mathbf{r})\Big(-\frac{1}{2}\bigtriangledown^2 + V_{\text{ext}}(\mathbf{r}) - \mu\Big)\varphi_k(\mathbf{r}) \tag{24}$$

$$V_H[n] = \frac{1}{2}\int d\mathbf{r}\, n(\mathbf{r})\, U(\mathbf{r}-\mathbf{r}')\, n(\mathbf{r}') \tag{25}$$

$$V_{\text{xc}}[n] = -\frac{1}{2}\int d\mathbf{r}\, n(\mathbf{r},\mathbf{r}')\, U(\mathbf{r}-\mathbf{r}')\, n(\mathbf{r}',\mathbf{r}) + \sum_{i=2}^{\infty}\Gamma_i\left[n\right] \tag{26}$$

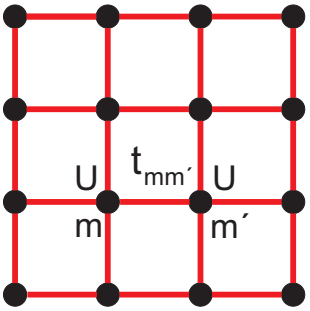


Fig. 2: *Generic Hubbard lattice for correlated lattice fermions with the local Coulomb interaction U and hopping parameters t: m can label different orbitals or lattice sites.*

and the local density approximation (LDA) to the DFT reads

$$V_{\rm xc}[n] = \int d\mathbf{r}\, n(\mathbf{r})\, \varepsilon_{\rm xc}\big(n(\mathbf{r})\big) \tag{27}$$

where $\varepsilon_{\rm xc}(n)$ is exchange correlation density for homogeneous electron Coulomb gas which can be calculated with the QMC scheme [1].
In the DFT scheme we lose information about the non-equal time Green function, which gives the single particle excitation spectrum as well as the k-dependence of the spectral function, and restrict ourself to only the ground-state energy of the many-electron system. Moreover, we lose information about all collective excitations in solids, such as plasmons or magnons, which can be obtained from the generalized susceptibility.
One of the most successful approaches to correlated materials is based on the combination of the DFT scheme with a Dynamical Mean Field Theory (DMFT) for strongly interacting d- or f-electrosn in the crystal. In the DMFT scheme one can obtain the numerically exact solution for the correlated part of the local functional. In the following section we discuss the general strong-coupling perturbation theory based on so-called dual fermion (DF) transformations [5] in the path integral formalism, which allows us to introduce the DMFT scheme as zero-order DF-expansion and shows a perturbative way to go beyond the DMFT approximations.

4 Dual Fermion approach with a general reference system

We start with a general lattice fermion model with the local Hubbard-like interaction vertex U. Generalization to the multi-orbital case is straightforward [21]. All equations will be written in matrix form, giving an idea of how to rewrite the dual fermion (DF) formula to the multi-orbital or multi-site case. The general strategy is related with the formally exact separation of the local and non-local correlation effects. We introduce auxiliary dual fermionic fields which will couple local correlated impurities or clusters back to the original lattice [5].
Using the path-integral formalism (Appendix A) the partition function of a general fermionic lattice system (Fig. 2) can be written in the following form as a functional integral over Grassmann variables $[c^*, c]$

$$Z = \int \mathcal{D}[c^*, c]\, \exp\big(-S_L[c^*, c]\big). \tag{28}$$

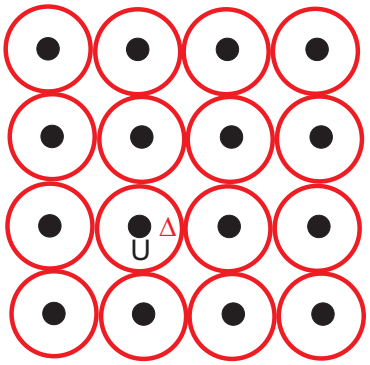


Fig. 3: *Schematic view on the real-space DMFT reference system.*

The original lattice action of interacting lattice fermions, similar to Eq. (3), can be written in Matsubara space as a sum of the lattice one-electron contributions with the Fourier transformed hopping $t_{\mathbf{k}}$ (or energy spectrum in the single-orbital case) and the local interaction part U

$$S_L[c^*,c] = -\sum_{\mathbf{k}\nu\sigma} c^*_{\mathbf{k}\nu\sigma}\big(i\nu{+}\mu{-}t_{\mathbf{k}}\big)\, c_{\mathbf{k}\nu\sigma} + \sum_i \int_0^\beta d\tau\, U\, n^*_{i\tau\uparrow} n_{i\tau\downarrow}\,. \tag{29}$$

In the following, $\nu = (2n{+}1)\pi/\beta, (\omega = 2n\pi/\beta), n = 0, \pm1, \ldots$ are the fermionic (bosonic) Matsubara frequencies, β is the inverse temperature, $\tau \in [0,\beta)$ the imaginary time, and μ the chemical potential. The index i labels the lattice sites, m refers to different orbitals, σ is the spin projection and the $\mathbf{k}$-vectors are quasimomenta. In order to keep the notation simple, it is useful to introduce the combined index $|1\rangle \equiv |i,m,\sigma,\tau\rangle$ and assume summation over repeated indices. Translational invariance is assumed for simplicity in the following, although a real space formulation is straightforward. The local part of the action, S_U, may contain any type of local multi orbital interaction.

In order to formulate an expansion around the best possible reference action, Fig. 3, a quantum impurity (cluster) problem is introduced by a general frequency-dependent hybridization function Δ_ν and the same local interaction

$$S_\Delta[c_i^*,c_i] = -\sum_{\nu,\sigma} c^*_{i\nu\sigma}\big(i\nu{+}\mu{-}\Delta_\nu\big) c_{i\nu\sigma} + \sum_\nu U n^*_{i\nu\uparrow} n_{i\nu\downarrow}\,, \tag{30}$$

where Δ_ν is the effective hybridization matrix describing the coupling of the impurity to an auxiliary fermionic bath. The main motivation for rewriting the lattice action in terms of a quantum impurity model is that such a reference system can be solved numerically exactly for an arbitrary hybridization function using ct-QMC methods [3]. Using the locality of the hybridization function Δ_ν, the lattice action Eq. (29) can be rewritten exactly in terms of individual impurity models and the effective one-electron coupling $\Delta_\nu{-}t_{\mathbf{k}}$ between different impurities, Fig. 4,

$$S_L[c^*,c] = \sum_i S_\Delta[c_i^*,c_i] - \sum_{\mathbf{k}\nu\sigma} c^*_{\mathbf{k}\nu\sigma}\big(\Delta_\nu{-}t_{\mathbf{k}}\big) c_{\mathbf{k}\nu\sigma}\,. \tag{31}$$

We will find the condition for the optimal choice of the hybridization function later. Although we can solve the individual impurity model exactly, the effect of spatial correlations due to

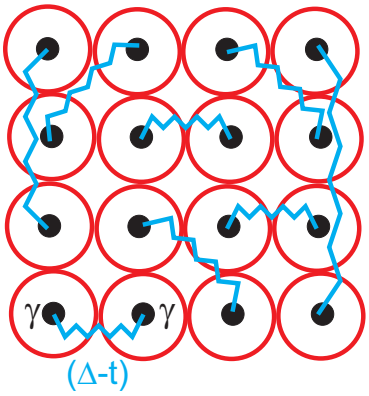


Fig. 4: *Schematic view on the non-local DF perturbation beyond a DMFT solution.*

the second term in Eq. (31) is very hard to treat, even perturbatively, since the impurity action is non-Gaussian and one cannot use of the Wick theorem. The main idea of the dual fermion transformation is the change of variables from strongly correlated fermions (c^*, c) to weakly correlated "dual" Grassmann fields (d^*, d) in the path-integral representation for the partition function of Eq. (3), followed by a simple perturbation treatment. The new variables were introduced through the following Hubbard-Stratonovich(HS)-transformation with the matrix $\widetilde{\Delta}_{\mathbf{k}\nu} = \Delta_\nu - t_\mathbf{k}$ in real space

$$e^{c_1^*\,\widetilde{\Delta}_{12}\,c_2} = \det\widetilde{\Delta}\int\mathcal{D}\,[d^*,d]\;e^{-d_1^*\widetilde{\Delta}_{12}^{-1}d_2 - d_1^*c_1 - c_1^*d_1}. \tag{32}$$

We can immediately seen that using this HS-transformation we "localize" the $[c_i^*, c_j]$ fermions: while on the left-hand side they are still "hopping" through the lattice, on the right-hand side they are localized on one site $[c_i^*, c_i]$.
With this reference system the lattice partition function becomes

$$\frac{Z}{Z_d} = \int\mathcal{D}[c^*,c,d^*,d]\exp\bigl(-S[c^*,c,d^*,d]\bigr) \tag{33}$$

with $Z_d = \det\widetilde{\Delta}$. The lattice action transforms to

$$S[c^*,c,d^*,d] = \sum_i S_\Delta^i + \sum_{\mathbf{k},\nu,\sigma} d^*_{\mathbf{k}\nu\sigma}\bigl(\Delta_\nu - t_k\bigr)^{-1} d_{\mathbf{k}\nu\sigma}\,. \tag{34}$$

Hence the coupling between sites is transferred to a local coupling to the auxiliary fermions

$$S_\Delta^i[c_i^*,c_i,d_i^*,d_i] = S_\Delta[c_i^*,c_i] + \sum_{\nu,\sigma}\bigl(d^*_{i\nu\sigma}\;c_{i\nu\sigma} + c^*_{i\nu\sigma}\;d_{i\nu\sigma}\bigr) \tag{35}$$

For the last term we use the invariance of the trace so that the sum over all states labeled by $\mathbf{k}$ could be replaced by the equivalent summation over all sites by a change of basis in the second term. The crucial point is that the coupling to the auxiliary fermions is purely local and S_Δ^i decomposes into a sum of local terms. The lattice fermions can therefore be integrated out from S_Δ^i for each site i separately. This completes the change of variables

$$\frac{1}{Z_\Delta}\int\mathcal{D}[c^*,c]\exp\bigl(-S_\Delta^i[,c_i^*,c_i,d_i^*d_i]\bigr) = \exp\Bigl(-\sum_{\nu\,\sigma} d^*_{i\nu\sigma}\;g_\nu d_{i\nu\sigma} - V_i[d_i^*d_i]\Bigr), \tag{36}$$

where Z_Δ is the partition function of the impurity action Eq. (30) and g_ν is the exact impurity Green function

$$g_{12} = -\langle c_1 c_2^* \rangle_\Delta = \frac{1}{Z_\Delta} \int \mathcal{D}[c^*, c]\, c_1 c_2^*\, e^{-S_\Delta[c^*,c]}. \tag{37}$$

The above equation may be viewed as the defining equation for the dual potential $V[d^*, d]$. The choice of matrices in Eq. (32) ensures a particularly simple form of this potential. An explicit expression is found by expanding both sides of Eq. (36) and equating the resulting expressions order by order. Formally this can be done up to all orders and in this sense the transformation to the dual fermions is exact. For most applications, the dual potential is approximated by the first non-trivial interaction vertex

$$V[d^*, d] = \frac{1}{4} \sum_{1234} \gamma_{1234}\, d_1^* d_2^* d_4 d_3\,, \tag{38}$$

where for the local vertex the combined index $1 \equiv \{m\nu\sigma\}$ comprises orbital degrees of freedom (or cluster sites), frequency, and spin. γ is the exact, fully antisymmetric, reducible two-particle vertex of the local quantum impurity problem. With the present choice of normalization in the HS-transformation we did not "amputate" the impurity "legs" or g_{12} Greens function which will be a very useful choice for ct-QMC calculations of the local vertex for multi-orbital case. It is given then by the connected part of the local two-particle correlations function

$$\gamma_{1234} = \chi_{1234} - \chi^0_{1234} \tag{39}$$

with the two-particle Green function of the local impurity (reference system) being defined as

$$\chi_{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_\Delta = \frac{1}{Z_\Delta} \int \mathcal{D}[c^*, c]\, c_1 c_2 c_3^* c_4^*\, e^{-S_\Delta[c^*,c]}\,. \tag{40}$$

The disconnected part of a generalized susceptibility reads

$$\chi^0_{1234} = g_{14} g_{23} - g_{13} g_{24}\,. \tag{41}$$

The single- and two-particle Green functions can be calculated using the ct-QMC Monte Carlo algorithms [3]. After integrating-out the lattice fermions, the dual action depends on the new variables only and for the one-orbital paramagnetic case reads

$$\tilde{S}[d^*, d] = -\sum_{\mathbf{k}\,\nu\sigma} d^*_{\mathbf{k}\nu\sigma}\, \tilde{G}^{-1}_{0\mathbf{k}\nu}\, d_{\mathbf{k}\nu\sigma} + \sum_i V_i[d_i^*, d_i] \tag{42}$$

while the bare dual Green function is has the form

$$\tilde{G}^0_{\mathbf{k}\nu} = \Big(\big(t_{\mathbf{k}} - \Delta_\nu\big)^{-1} - g_\nu \Big)^{-1}. \tag{43}$$

This formula involves only the local Green function g_ν of the impurity model. It is important to note, that the HS-transformation to dual fermion variables, allows us to "perform the analytical amputation" of impurity "legs" which causes enormous problems in the multi-orbital ct-QMC

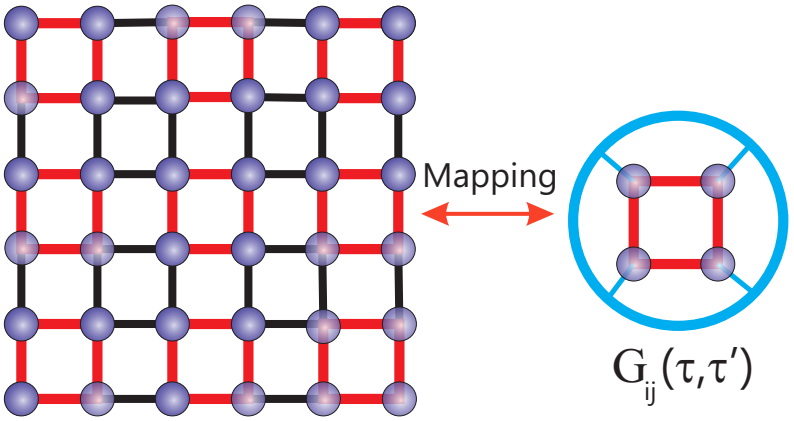


Fig. 5: *Schematic representation of a plaquette cluster-reference system for the square lattice.*

formalism. Transformation to the original DF-normalization where both dual G_d and real Green function have the same dimension unit reads

$$G_d = g\,\widetilde{G}\,g = G_{\rm DMFT} - g \qquad G_{\rm DMFT} = \big(g_\nu + \Delta_\nu - t_{\mathbf{k}}\big)^{-1}. \tag{44}$$

The Dual Fermion transformation allows us to use arbitrary reference systems and transform the strongly correlated lattice fermion problem to an effective action of weakly coupled dual quasiparticles. This is related with the fact that the bare dual Green function Eq. (44) related with the small non-local part of the DMFT lattice Green function and the main two-particle part of the bare interaction among dual fermions Eq. (38) is exactly equal to the fully screened impurity vertex of the reference system. The rest of the dual fermion problem is related with an optimal perturbation scheme for such an effective action Eq. (42).

5 Perturbation in Dual Space

For general multi-orbital multi-site dual fermion perturbation technique from arbitrary reference system (Fig. 5) we use the particle-hole notation for the local vertex and write the exact spin and imaginary time structure of the generalized connected susceptibility [5, 22]

$$\gamma^{\sigma\sigma'}_{1234}(\tau_1,\tau_2,\tau_3,\tau_4) = -\big\langle c_{1\sigma} c^*_{2\sigma} c_{3\sigma'} c^*_{4\sigma'} \big\rangle_\Delta + g^{\sigma}_{12} g^{\sigma'}_{34} - g^{\sigma}_{14} g^{\sigma}_{32}\,\delta_{\sigma\sigma'}.$$

Then the bare vertex of dual-fermion perturbation is related with the full impurity vertex, which in Matsubara space depends on two fermionic (ν,ν') and one bosonic (ω) frequencies. We also symmetrized vertex for charge density d- and spin s-channels

$$\gamma^{d/m}_{1234}(\nu,\nu',\omega) = \gamma^{\uparrow\uparrow}_{1234}(\nu,\nu',\omega) \pm \gamma^{\uparrow\downarrow}_{1234}(\nu,\nu',\omega).$$

Now we can write the first-order, local in site (i) DF-correction to the dual self-energy (Fig.6):

$$\widetilde{\Sigma}^{(1)i}_{12}(\nu) = \sum_{\nu',3,4} \gamma^{d}_{1234}(\nu,\nu',0)\widetilde{G}^{ii}_{43}(\nu'). \tag{45}$$

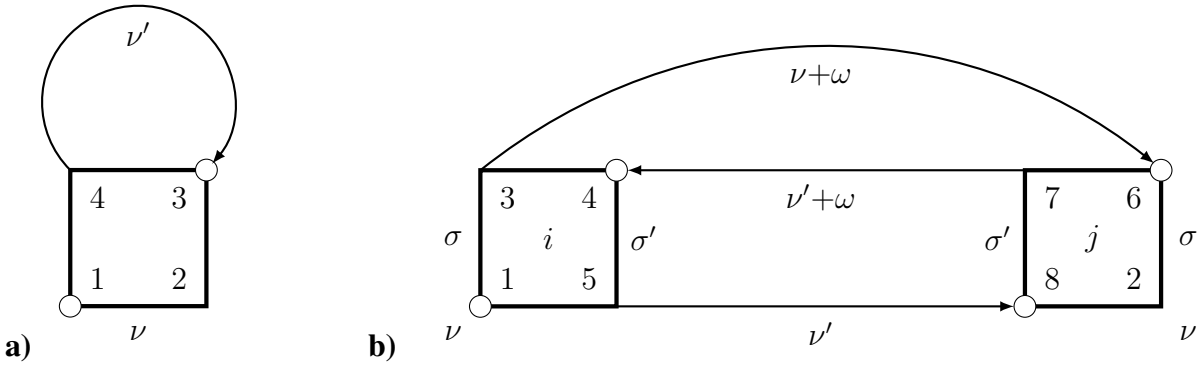


Fig. 6: *Feynman diagram for the 1-st order (a) and the 2-nd order (b) of dual fermion perturbation for the self-energy $\widetilde{\Sigma}$: a line represents the non-local $\widetilde{G}_{43}$ and a box is the local γ_{1234}.*

We now can use the freedom to chose the hybridization function Δ_ν in order to eliminate the main first-oder dual fermion correction Eq. (45) and all other local DF-diagram. Since the vertex function γ_{1234} is purely local, it is enough to ensure that the local part of dual Green function vanishes $\widetilde{G}_{loc} = 0$. This condition is exactly equivalent to the DMFT self-consistent condition for the hybridization function Δ_ν (Fig.7)

$$\sum_{\mathbf{k}} \left(g_\nu^{-1} + \Delta_\nu - t_{\mathbf{k}} \right)^{-1} = g_\nu \,. \tag{46}$$

The effective impurity model Eq. (30), which is fully determined by the local hybridization function Δ_ν on fermionic Matsubara frequencies $i\nu_n$ is solved using the numerically exact ct-QMC scheme [3] and the exact local Green function g_ν is obtained. The self-consistent DMFT condition for the hybridization function equates the local part of the lattice Green function with that of the impurity, which shows that DMFT minimizes, in a local sense, the distance $|t_{\mathbf{k}} - \Delta_\nu|$. It worthwhile to point here that the "free" or non-interacting dual fermions are equivalent to the full solution of the DMFT problem, This is why dual fermions are only "weakly interacting" so that this perturbation scheme can be very efficient, provided a good reference system.

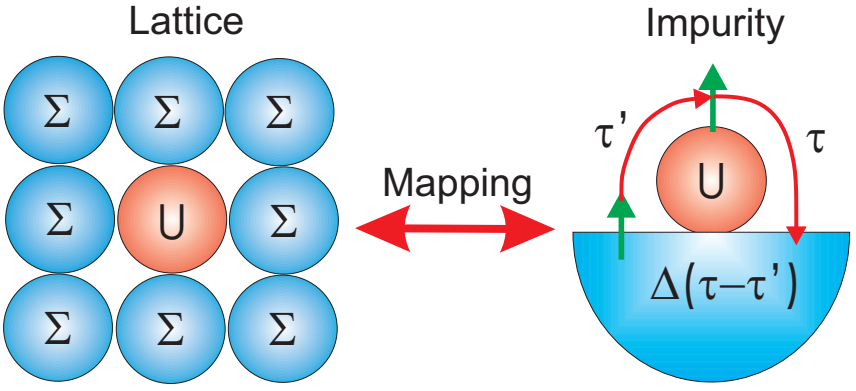


Fig. 7: *Schematic representation of the DMFT reference system for correlated lattice models.*

The second-order Feynman diagram for DF-perturbation (Fig.6) in real space ($\mathbf{R_{ij}}$) has density- and spin-channel contributions with corresponding constants ($c_d = -1/4$ and $c_m = -3/4$)

$$\widetilde{\Sigma}^{(2)ij}_{12}(\nu) = \sum_{\nu'\omega} \sum_{3-8} \sum_{\alpha=d,m} c_\alpha \gamma^{\alpha,i}_{1345}(\nu,\nu',\omega)\, \widetilde{G}^{ij}_{36}(\nu{+}\omega)\, \widetilde{G}^{ji}_{74}(\nu'{+}\omega)\, \widetilde{G}^{ij}_{58}(\nu')\, \gamma^{\alpha,j}_{8762}(\nu',\nu,\omega)\,.$$

Using the exact relation between dual Green function and real Green function [5,22], we can express the total lattice self-energy as the sum of the reference contribution Σ^0 (e.g. the impurity) and corrections Σ' which related with the dual self-energy $\widetilde{\Sigma}$

$$\begin{aligned} \Sigma_{\mathbf{k}\nu} &= \Sigma^0_\nu + \Sigma'_{\mathbf{k}\nu} \\ \Sigma'_{\mathbf{k}\nu} &= g^{-1}_\nu - \left(g_\nu{+}\widetilde{\Sigma}_{\mathbf{k}\nu}\right)^{-1}. \end{aligned}$$

We pointe out that this expression is related with the exact transformation between dual- and real-spaces [4] which has general consequences for similar combinations including strong-coupling techniques, like DMFT or DF.

What is much more important for numerical calculations is that, starting from the full DF-Green function: $\tilde{G}^{-1}_{\mathbf{k}\nu} = \tilde{G}^{-1}_{0\mathbf{k}\nu} - \widetilde{\Sigma}_{\mathbf{k}\nu}$ and the exact relation between dual and real Green function [5,22], we can find an expression for the lattice Green function including only the reference impurity Green function and the dual self-energy [4]

$$G_{\mathbf{k}\nu} = \left(\left(g_\nu{+}\widetilde{\Sigma}_{\mathbf{k}\nu}\right)^{-1} - \widetilde{\Delta}_{\mathbf{k}\nu}\right)^{-1} \tag{47}$$

This formula is perfectly suitable for ct-QMC calculations for realistic multi-orbital correlated matter, where from noisy Monte-Carlo data one only needs local one- and two-particle Green functions, without knowledge of a local self-energy and an "amputated" vertex function.

In principle, one can go beyond second order perturbation theory and include dual ladder diagrams [22], dual parquet diagrams [23] or even try to sum up all dual diagrams with two-particle vertex γ_{1234} stochastically, using a diagrammatic Monte Carlo in dual space [24, 25]. We also can make the diagrammatic series self-consistent, using the "bold" line and update the bare dual Green function with the dual self-energy. Finally, one can also "upgrade" the reference system, which is not anymore "best" for dual fermion theory, but only for the DMFT approach with non-interacting dual fermions.

As numerical example of the DF-perturbation, we discuss a simple, but non-trivial case of a half-field Hubbard model in the strong-coupling case $U = W = 8t$ ($t = 1$, U is equal to the bandwidth $W = 8t$) and the temperature $\beta = 5$. We use a free periodic plaquette as a reference system and perform second-order DF-perturbation. We use exact-diagonalization solver and the Padé analytical continuation to the real axes [7]. Results of the second-order DF superperturbation are shown on Fig. 8 together with numerical exact lattice Monte Carlo results and exact diagonalization of a $4{\times}4$ periodic cluster. The DOS for the second order dual fermion plaquette perturbations are in a good agreements with the two numerically exact schemes, ED and DQMC. The DF theory reproduces the so-called four-peak structure of the half-filled Hubbard model, which is a standard feature of lattice QMC calculations [26].

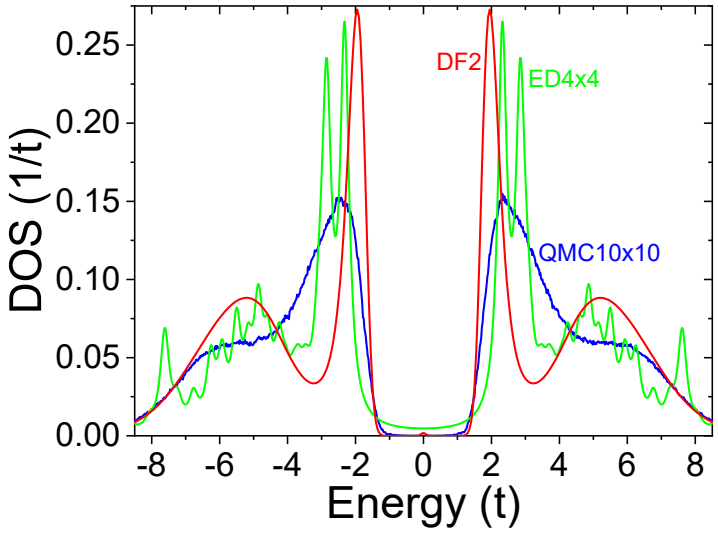


Fig. 8: *Density of states for the half-filled Hubbard model with $U = W = 8t$ in the second-order DF approximation (DF2) in comparison to ED results for a 4×4 cluster and lattice QMC simulation of a 10×10 system.*

6 LDA+DMFT scheme for real materials

In order to investigate real correlated systems with the local DMFT scheme we need to have an efficient scheme of partitioning the space and orbital degrees of freedom. For example in the high-temperature superconducting oxide $YBa_2Cu_3O_7$ the strongly correlated electrons are the Cu $3d$'s, and moreover there is only one per non-equivalent copper $d_{x^2-y^2}$ band which crosses the Fermi level with strong many-body fluctuations. Just a few percent of the total number of electronic states need to be included in the DMFT calculations. Therefore the simplest realistic correlated scheme would be a DFT+DMFT approach [27, 28] with partitioning of the orbital space to the normal band electrons $|K\rangle$ described by the DFT Bloch basis and the correlated local orbitals $|L\rangle$ described by some optimal Wannier basis (see Fig. 9 for an illustration).

The treatment of correlated electron systems requires the calculation of Green functions and hybridization functions in terms of local orbitals. This is readily achieved when using a basis set, which is localized in real space, such as linear (or N-th order) muffin-tin orbitals (NMTO) [29], plane-waves [30], Gaussian basis sets [31], or a projector scheme [32]. However, many implementations of density functional theory use a delocalized plane-wave basis set. This has the advantage, that the basis set is simple, universal, and its convergence is controlled in principle by a single parameter, the energy cutoff. The projector augmented wave method (PAW) [33], being a representative of a plane-wave based methods, can be used as a simple example of the general projection scheme from the Bloch to the local basis: $\langle K|L\rangle$ (Fig. 9).

Following the general projection scheme of Ref. [32,31], the desired quantity for an implementation of the DFT+DMFT method is a projection $\mathcal{P}^{\mathcal{C}} = \sum_L |L\rangle\langle L|$ of the full DFT Kohn-Sham Green function $G_{\rm KS}(\omega)$ on a set of localized orbitals $\{\,|L\rangle\,\}$

$$G^{\mathcal{C}}(\omega) = \mathcal{P}^{\mathcal{C}}\, G^{\rm KS}(\omega)\, P^{\mathcal{C}}. \tag{48}$$

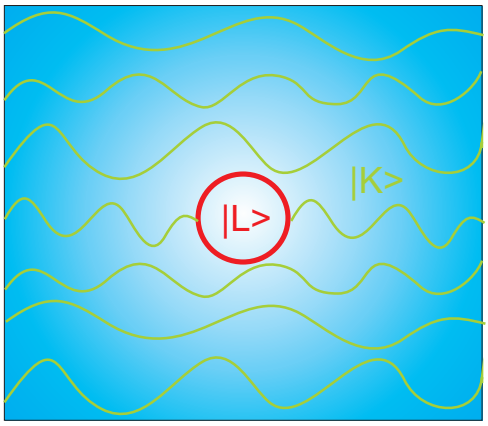


Fig. 9: *Schematic representation of the projection scheme from the Bloch basis to a local Wannier correlated subset.*

The subspace $\mathcal{C} = \mathrm{span}(\{ |L\rangle \})$ is usually called the correlated subspace. It is the subspace of orbitals in which many-body fluctuations play a major role and where the DMFT corrections to the DFT will be considered. In plane-wave based calculations, $G^{\mathrm{KS}}(\omega)$ in Matsubara space is available in terms of an almost complete set of Bloch states $|K\rangle$ that are eigenstates of the Kohn-Sham Hamiltonian $H_{\mathrm{KS}} |K\rangle = \varepsilon_K |K\rangle$

$$G_{\mathrm{KS}}(\omega) = \sum_K \frac{|K\rangle \langle K|}{i\omega + \mu - \varepsilon_K}. \tag{49}$$

Inserting equation (49) into (48) shows that one needs to evaluate projections of type $\langle L|K\rangle$ in order to access the matrix elements $G^{\mathcal{C}}_{LL'}(\omega)$ of the local Green function. In most cases the correlated orbitals are d or f orbitals, which, to a good approximation, are localized inside the PAW augmentation spheres. For $|L\rangle$ within these spheres and given the PAW decomposition of a Bloch state $|K\rangle$ [33] one obtains

$$\langle L|K\rangle = \sum_i \langle L|\varphi_i\rangle \langle \tilde{p}_i|\tilde{K}\rangle. \tag{50}$$

The index i of the augmentation functions $|\varphi_i\rangle$ includes site s, angular momentum l and m, as well as an index ν labeling the radial function: $i = (s, l, m, \nu)$, while the $|\tilde{p}_i\rangle$ are projectors of the PAW scheme.

In the described projections scheme the $|L\rangle\langle L|$ matrices are not properly normalized for two reasons: (1) the Bloch basis is incomplete since only a limited number of Bloch bands is included and (2) the PAW augmentation functions are in general not orthonormal. The simplest way is to orthonormalized the projection matrices by the following Wannier type construction: By definition, the localized states $|L\rangle$ are labeled by site and angular momentum indices: $L = (s, l, m)$. We split the site index $s = \mathbf{R}+\mathbf{T}$ such that $\mathbf{R}$ labels the position within the unit cell and $\mathbf{T}$ is the Bravais lattice vector of the unit cell in which s is located. This allows us to construct the Bloch transform of the localized states,

$$|L_{\mathbf{k}}\rangle = \sum_T e^{i\mathbf{k}\mathbf{T}} |L_{\mathbf{T}}\rangle \,, \tag{51}$$

where $\mathbf{k}$ is from the first Brillouin zone and $|L_{\mathbf{T}}\rangle \equiv |L\rangle = |s,l,m\rangle$. The sum in equation (51) runs over the Bravais lattice. Labeling the Bloch states $|K\rangle = |\mathbf{k},n\rangle$ by their crystal momentum, $\mathbf{k}$, and band index, n, we normalize our projection matrices $\mathcal{P}^{\mathcal{C}}_{Ln}(\mathbf{k}) = \langle L_{\mathbf{k}}|\mathbf{k},n\rangle$ using the overlap operator

$$O_{LL'}(\mathbf{k}) = \sum_n \mathcal{P}^{\mathcal{C}}_{Ln}(\mathbf{k})\, \mathcal{P}^{*\mathcal{C}}_{L'n}(\mathbf{k}) \tag{52}$$

in

$$\bar{\mathcal{P}}^{\mathcal{C}}_{Ln}(\mathbf{k}) = \sum_{L'} O^{-1/2}_{LL'}(\mathbf{k}) P^{\mathcal{C}}_{L'n}(\mathbf{k}). \tag{53}$$

These orthonormalized projection matrices are calculated once at the beginning of any calculation and can then be used to obtain the local Green function of the correlated orbitals from the full Bloch Green function $G^B_{nn'}$

$$G^{\mathcal{C}}_{LL'}(\omega) = \sum_{\mathbf{k},nn'} \bar{\mathcal{P}}^{\mathcal{C}}_{Ln}(\mathbf{k})\, G^B_{nn'}(\mathbf{k},\omega)\, \bar{\mathcal{P}}^{*\mathcal{C}}_{L'n'}(\mathbf{k}).$$

Similarly the hybridization function, $\Delta(\omega)$, is available. It is related to the local Green function by

$$G^{-1}(\omega) = i\omega - \epsilon_d - \Delta(\omega), \tag{54}$$

where ϵ_d is the static crystal field. Equation (54) is a matrix equation with G, Δ, and ϵ_d being $(\dim \mathcal{C}) \times (\dim \mathcal{C})$ matrices, in general. To separate the hybridization from the static DFT crystal field, we numerically evaluate the limit $\omega \to \infty$, where $\omega - G^{-1}(\omega) \to \epsilon_d$.

In a DFT+DMFT calculation the projection matrices $\bar{\mathcal{P}}^{\mathcal{C}}_{Ln}(k)$ are used for up- and downfolding quantities like the Green function and the self energy in the course of the iterative DMFT procedure in exactly the same way as shown for the local Green function above. For example, the self energy obtained by an impurity solver for the effective impurity model $\Sigma^{\mathcal{C}}_{LL'}(\omega)$ can be upfolded to the Bloch basis as

$$\Sigma^B_{nn'}(\mathbf{k},\omega) = \sum_{LL'} \overline{P}^{*\mathcal{C}}_{Ln}(\mathbf{k})\; \Sigma^{\mathcal{C}}_{LL'}(\omega)\; \overline{P}^{\mathcal{C}}_{L'n'}(\mathbf{k}).$$

Since the self energy in DMFT is a purely local quantity, the index $\mathbf{k}$ on $\Sigma^B_{nn'}(\mathbf{k},\omega)$ reflects the momentum dependence brought about by the projection matrices. The presented projection scheme allows for the inclusion of both correlated and uncorrelated states in the procedure. Therefore, information about the interplay of correlated orbitals with their uncorrelated ligands can be obtained.

Figure 10 shows that the DFT+DMFT calculation commences with the solution of the Kohn-Sham equations of DFT. In a second step the projection onto the correlated subset is computed. The Kohn-Sham Green function is then computed and used as an initial guess for the mean-field $\mathcal{G}$ of the DMFT cycle, which consists of the usual steps detailed before. In usual applications, the DFT+DMFT loop will stop after DMFT self-consistency is obtained. It has, however, recently become possible to continue the cycle supplying the DFT code with an altered charge density that includes correlation effects. In such a unified approach, changes in the charge

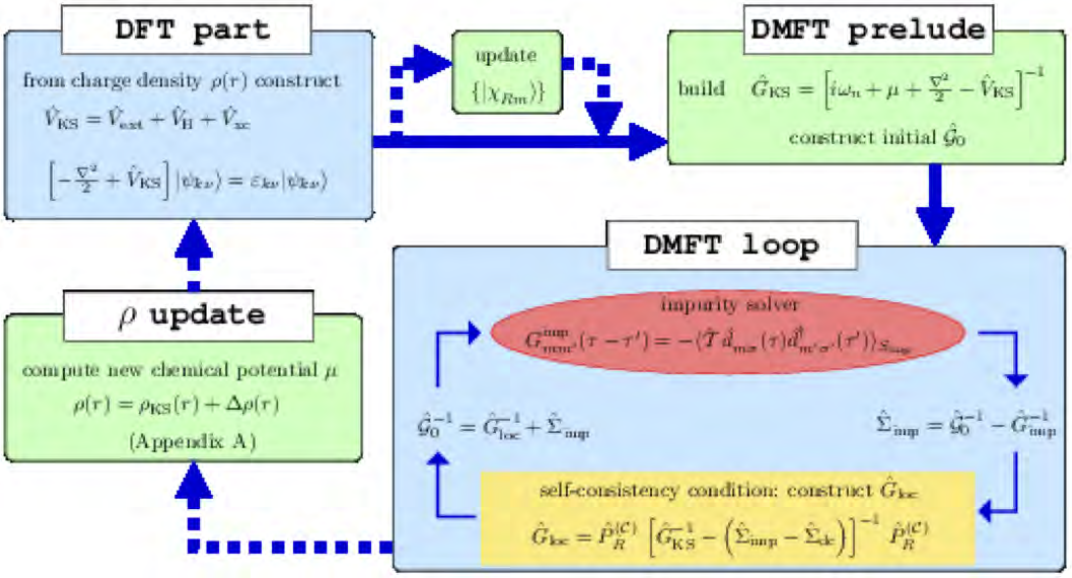


Fig. 10: *Illustration of the DFT+DMFT procedure. As a first step, the Kohn-Sham (KS) equations, determining the Kohn-Sham potential and thus the Hamiltonian, are solved. Secondly, the KS Green function and from it the starting value for the bath Green function $\mathcal{G}$ is constructed and passed on to the DMFT loop, which consists of the usual steps described before. A potential self-consistency over the charge density is also indicated [31].*

density induced by correlations can be studied [31, 34]. It allows furthermore for the accurate calculation of total energies, that allow the determination of crystal structures and other coupled electronic and structural effects. We will briefly sketch here the most important elements of the charge self-consistent implementation of DFT+DMFT following Ref. [31].

As example, we show a realistic DFT+DMFT calculation of ferromagnetic spectral function for Fe and Ni at $T = 380$ K in Fig. 11 [35], in comparison with experimental ARPES spectra for Fe [36] and Ni [37]. For the DFT part the VASP code with theWannier projector scheme was used [32] and for DMFT impurity solver numerical exact ct-QMC was used [3]. One can see a good agreement for renormalization of the band dispersion near the Fermi level for Ni, but some discrepancy for Fe (specially in P-H-Γ direction). This indicates that one needs to include non-local self-energy effects in Fe beyond the DMFT, such as in the multi-orbital DF-scheme.

We do not discuss the problem of "double counting" in the LDA+DMFT formalism: since any DFT approximation includes average of Coulomb interactions, one should subtract these from the DMFT part. In practice it is very hard to fulfill, since we do not know this part analytically and normally we chose the double counting correction from the LDA+U scheme as a static version of LDA+DMFT [27, 28]. The main problem is related with the fact that both DFT and DMFT are non-perturbative, strong-coupling theories which need to be "aligned" on some reference level, like the chemical potential. One can try to search for this "reference level" as a free parameter [38] and compare the resulting spectral function of the LDA+DMFT with experimental photoemission spectrum. Another possibility is to start from an analytical

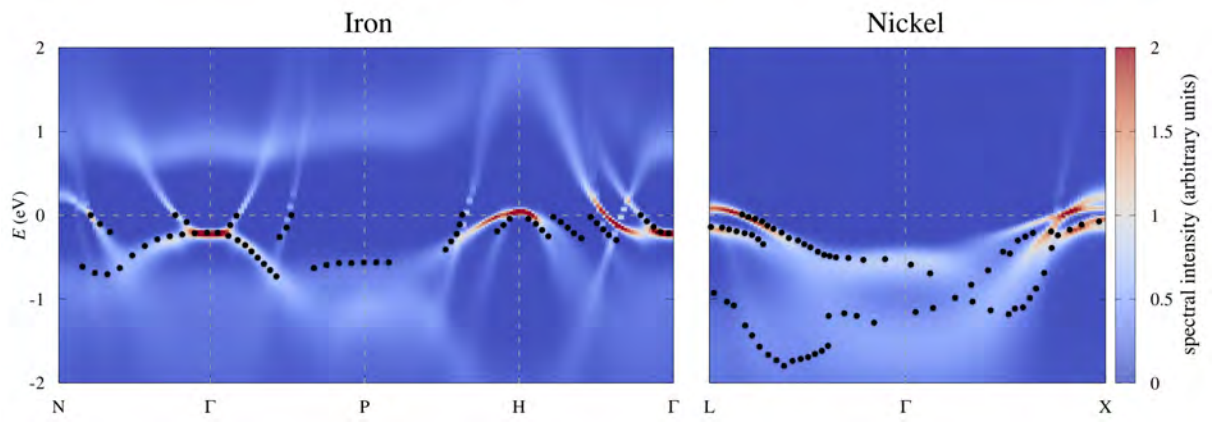


Fig. 11: *Momentum resolved spectral function for ferromagnetic Fe and Ni obtained by LDA+DMFT for T=380 K in comparison with photoemission data (dots) [35].*

many-body scheme like GW and design a GW+DMFT approach where one knows the local correlated part analytically [39]. If we would like to incorporate non-local self-energy effects beyond the DMFT, then such a simple "double counting" schemes may not work any more and more elaborat dual fermion/boson schemes should be considered. There is also the complicated problem of estimating the effective local Coulomb matrix elements in the Wannier basis used in LDA+DMFT. One of the most accurate schemes is related with a constrained RPA scheme [40]. We can mention a recent discussion of the cRPA scheme for strongly correlated insulators [41]. One of the recent accurate charge self-consistent LDA+DMFT software packages is based on combination of Quantum ESPRESSO (DFT), Wannier90 (Projectors), and TRIQS (ct-QMC) [42].

7 Conclusions

We have discussed the path integral expansion for general reference systems which help to understand the DFT scheme and the correlated DMFT approach for interacting lattice fermions and their combination for realistic systems. The transformation to dual variables allowed to go beyond the local DMFT approximation for lattice self-energy. The accurate multi-orbital DFT+DMFT scheme for strongly correlated transition metals and rare earth systems can described the spin, orbital and charge fluctuations in the d- or f-shell which play the crucial role in photoemission spectra, as well as magnetic and optical excitations. The numerically exact ct-QMC solution of the quantum impurity problem gives an effective local exchange-correlation functional for given correlated materials in a specific external hybridization. The accurate combination of the DMFT scheme with first-principle DFT approaches gives a very useful tool for the investigation of correlated electronic materials.

Acknowledgment

The support by the European Research Council via Synergy Grant 854843 – FASTCORR are greatly acknowledged.

A Path integrals for fermions

We first introduce a formalism of the path integral over fermionic fields [15]. Let us consider a simple case of a single quantum state $|i\rangle$ occupied by fermionic particles [16]. Due to the Pauli principle the many-body Hilbert space is spanned by only two orthonormal states $|0\rangle$ and $|1\rangle$. In the second quantization scheme for fermions with annihilation $\hat{c}_i$ and creation $\hat{c}_i^\dagger$ operators with anticommutation relations $\{\hat{c}_i, \hat{c}_j^\dagger\} = \delta_{ij}$ we have the following simple rules

$$\hat{c}_i\,|1\rangle = |0\rangle \quad \hat{c}_i\,|0\rangle = 0 \quad \text{and} \quad \hat{c}_i^\dagger\,|0\rangle = |1\rangle \quad \hat{c}_i^\dagger\,|1\rangle = 0\,. \tag{55}$$

Moreover, the density operator and the Pauli principle have the form

$$\hat{c}_i^\dagger \hat{c}_i\,|n\rangle = n_i\,|n\rangle \quad \text{and} \quad \hat{c}_i^2 = (\hat{c}_i^\dagger)^2 = 0\,.$$

The central object here are the so-called fermionic coherent states $|c\rangle$, which are eigenstates of annihilation operator $\hat{c}_i$ with eigenvalue c_i

$$\hat{c}_i\,|c\rangle = c_i\,|c\rangle\,. \tag{56}$$

It is worthwhile to note that such a left-eigenbasis has only annihilation operators, due to the fact that they are bounded from the below and one can rewrite one of equations from Eq. (55) in the following "eigenvalue" form

$$\hat{c}_i\,|0\rangle = 0\,|0\rangle\,.$$

Due to the anti-commutation relations for the fermionic operators the eigenvalues of coherent states c_i are so-called Grassmann numbers with the multiplication rules [17]

$$c_i c_j = -c_j c_i \quad \text{and} \quad c_i^2 = 0\,. \tag{57}$$

It is convenient to assume that the Grassmann numbers also anti-commute with the fermionic operators

$$\{c, \hat{c}\} = \{c, \hat{c}^\dagger\} = 0\,.$$

An arbitrary function of one Grassmann variable can be represented by only the first two Taylor coefficients

$$f(c) = f_0 + f_1 c\,. \tag{58}$$

One can prove the following general many-body representation of coherent states

$$|c\rangle = e^{-\sum_i c_i \hat{c}_i^\dagger}\,|0\rangle\,. \tag{59}$$

Let us show this for the simple case of one fermionic state

$$\hat{c}\,|c\rangle = \hat{c}\big(1 - c\hat{c}^\dagger\big)\,|0\rangle = \hat{c}\big(\,|0\rangle - c\,|1\rangle\,\big) = -\hat{c}c\,|1\rangle = c\,|0\rangle = c\,|c\rangle\,. \tag{60}$$

One can also define a "left" coherent state $\langle c|$ as the left-eigenstates of creation operators $\hat{c}_i^\dagger$

$$\langle c|\,\hat{c}_i^\dagger = \langle c|\,c_i^*.$$

Note that the new eigenvalue c_i^* is just another Grassmann number, not the complex conjugate of c_i. The left coherent state can be obtained similar to Eq. (59)

$$\langle c| = \langle 0|\, e^{-\sum_i \hat{c}_i c_i^*}.$$

A general function of two Grassmann variables can, analogously to Eq. (58), be represented by only four Taylor coefficients

$$f(c^*, c) = f_{00} + f_{10}c^* + f_{01}c + f_{11}c^*c\,. \tag{61}$$

Using this expansion we can define a derivative of Grassmann variables in the natural way

$$\frac{\partial c_i}{\partial c_j} = \delta_{ij}\,.$$

One needs to be careful with the "right order" of such a derivative and remember the anti-commutation rules, i.e.,

$$\frac{\partial}{\partial c_2} c_1 c_2 = -c_1\,.$$

For the case of the general two-variable function in Eq. (61) we have

$$\frac{\partial}{\partial c^*}\frac{\partial}{\partial c} f(c^*, c) = \frac{\partial}{\partial c^*}\big(f_{01} - f_{11}c^*\big) = -f_{11} = -\frac{\partial}{\partial c}\frac{\partial}{\partial c^*} f(c^*, c).$$

One also needs a formal definition of the integration over Grassmann variables, and the natural way consists of the following rules [17]

$$\int 1\, dc = 0 \quad \text{and} \quad \int c\, dc = 1\,,$$

which just shows that the integration over a Grassmann variable is equivalent to differentiation

$$\int \cdots\, dc \to \frac{\partial}{\partial c} \cdots$$

The coherent states are not orthonormal and the overlap of any two such states is equal to

$$\langle c|c\rangle = e^{\sum_i c_i^* c_i}$$

which is easy to see for the case of one particle

$$\langle c|c\rangle = \big(\langle 0| - \langle 1|\, c^*\big)\big(|0\rangle - c\,|1\rangle\big) = 1 + c^*c = e^{c^*c}.$$

An important property of coherent states is the resolution of unity

$$\int dc^* \int dc\, e^{-\sum_i c_i^* c_i}\, |c\rangle\langle c| = \hat{1} = \iint dc^* dc\, \frac{|c\rangle\,\langle c|}{\langle c|c\rangle}.$$

For simplicity we demonstrate this relation only for one fermion state

$$\begin{aligned}\iint dc^* dc\, e^{-c^*c}\, |c\rangle\langle c| &= \iint dc^* dc\, \big(1 - c^*c\big)\big(|0\rangle - c\,|1\rangle\big)\big(\langle 0| - \langle 1|\, c^*\big) \\ &= -\iint dc^* dc\; c^*c\, \big(|0\rangle\langle 0| + |1\rangle\langle 1|\big) = \sum_n |n\rangle\langle n| = \hat{1}\,.\end{aligned}$$

Matrix elements of normally ordered operators are very easy to calculate in the coherent basis by operating with $\hat{c}^\dagger$ on the states to the right and $\hat{c}$ to the left:

$$\langle c^*| \hat{H}(\hat{c}^\dagger, \hat{c}) |c\rangle = H(c^*, c) \langle c^*|c\rangle . = H(c^*, c) e^{\sum_i c_i^* c_i} \tag{62}$$

Within the manifold of coherent states we can map the fermionic operators to the Grassmann variables $(\hat{c}_i^\dagger, \hat{c}_i) \to (c_i^*, c_i)$.
Finally, we prove the so-called "trace-formula" for arbitrary fermionic operators in normal order (in one-fermion notation)

$$\begin{aligned} \mathrm{Tr}\ \hat{O} &= \sum_{n=0,1} \langle n| \hat{O} |n\rangle = \sum_{n=0,1} \iint dc^* dc\, e^{-c^* c} \langle n| c\rangle\langle c| \hat{O} |n\rangle = \\ &= \iint dc^* dc\, e^{-c^* c} \sum_{n=0,1} \langle -c| \hat{O} |n\rangle\langle n| c\rangle = \iint dc^* dc\, e^{-c^* c} \langle -c| \hat{O} |c\rangle . \end{aligned}$$

The fermionic "minus" sign in the left coherent states come from the commutation of the (c^*) and (c) coherent state in such a transformation: $\langle n|c\rangle \langle c|n\rangle = \langle -c|n\rangle \langle n|c\rangle$. One has to use the standard Grassmann rules: $c_i^* c_j = -c_j c_i^*$ and $|-c\rangle = |0\rangle + c\,|1\rangle$.
We are ready now to write the partition function for the grand-canonical quantum ensemble with $H = \hat{H} - \mu\hat{N}$ and inverse temperature β. One has to use the N-slices Trotter decomposition for the partition function in $[0, \beta)$ with imaginary time $\tau_n = n\Delta\tau = n\beta/N$ $(n = 1, ..., N)$, and insert N times the resolution of unity as follows

$$\begin{aligned} Z = \mathrm{Tr}\, e^{-\beta H} &= \iint dc^* dc\, e^{-c^* c} \big\langle -c \big| e^{-\beta H} \big| c \big\rangle \\ &= \int \Pi_{n=1}^N dc_n^* dc_n\, e^{-\sum_n c_n^* c_n} \langle c_N|\, e^{-\Delta\tau H}\, |c_{N-1}\rangle \langle c_{N-1}|\, e^{-\Delta\tau H}\, |c_{N-2}\rangle \ldots \langle c_1|\, e^{-\Delta\tau H}\, |c_0\rangle \\ &= \int \Pi_{n=1}^N dc_n^* dc_n\, e^{-\Delta\tau \sum_{n=1}^N \left(c_n^* (c_n - c_{n-1})/\Delta\tau + H(c_n^*, c_{n-1}) \right)} \end{aligned}$$

In the continuum limit $(N \to \infty)$

$$\Delta\tau \sum_{n=1}^{N} \cdots \to \int_0^\beta d\tau \cdots , \qquad \frac{c_n - c_{n-1}}{\Delta\tau} \to \partial_\tau \quad \text{and} \quad \Pi_{n=0}^{N-1} dc_n^* dc_n \to \mathcal{D}\,[c^*, c]$$

with antiperiodic boundary conditions for fermionic Grassmann variables in imaginary time $c(\tau)$ and $c^*(\tau)$

$$c(\beta) = -c(0), \qquad c^*(\beta) = -c^*(0)$$

we end up in the standard path-integral formulation of the partition function

$$Z = \int \mathcal{D}\,[c^*, c]\, e^{-\int_0^\beta d\tau \left(c^*(\tau)\partial_\tau c(\tau) + H(c^*(\tau), c(\tau)) \right)} . \tag{63}$$

It is useful to mention the general form of the Gaussian path-integral for a non-interacting "quadratic" fermionic action, which is equivalent to the Hubbard-Stratonovich transformation

used in the dual-fermion derivation Eq. (32). For an arbitrary matrix M_{ij} and Grassmann vectors J_i^* and J_i one can calculate analytically the following integral

$$Z_0\left[J^*,J\right]=\int\mathcal{D}\left[c^*c\right]e^{-\sum_{i,j=1}^{N}c_i^*M_{ij}c_j-\sum_{i=1}^{N}\left(c_i^*J_i+J_i^*c_i\right)}=\det M\;e^{\sum_{i,j=1}^{N}J_i^*(M^{-1})_{ij}J_j}\,. \qquad (64)$$

To prove this relation one needs to first change variables in order to eliminate J_i^* and J_i and expand the exponential function (only the N-th oder is non-zero)

$$e^{-\sum_{i,j=1}^{N}c_i^*M_{ij}c_j}=\frac{1}{N!}\Bigg(-\sum_{i,j=1}^{N}c_i^*M_{ij}c_j\Bigg)^N.$$

Finally, different permutations of c_i^* and c_j and integration over Grassmann variables will give $\det M$. As a small exercise we will check such integrals for the first two many-particle dimensions. For N=1 it is trivial

$$\int\mathcal{D}\left[c^*c\right]e^{-c_1^*M_{11}c_1}=\int\mathcal{D}\left[c^*c\right]\left(-c_1^*M_{11}c_1\right)=M_{11}=\det M$$

and for N=2 we have

$$\int\mathcal{D}\left[c^*c\right]e^{-c_1^*M_{11}c_1-c_1^*M_{12}c_1-c_2^*M_{21}c_1-c_2^*M_{22}c_2}=$$

$$=\frac{1}{2!}\int\mathcal{D}\left[c^*c\right]\left(-c_1^*M_{11}c_1-c_1^*M_{12}c_1-c_2^*M_{21}c_1-c_2^*M_{22}c_2\right)^2=M_{11}M_{22}-M_{12}M_{21}=\det M.$$

For a change of variables in the path integral one uses the following transformation with unit Jacobian: $c\rightarrow c+M^{-1}J$ and

$$c^*Mc+c^*J+J^*c=\left(c^*+J^*M^{-1}\right)M\left(c+M^{-1}J\right)-J^*M^{-1}J\,.$$

Using the Gaussian path-integral it is very easy to calculate any correlation function for a non-interaction action (Wick-theorem)

$$\left\langle c_ic_j^*\right\rangle_0=-\frac{1}{Z_0}\frac{\delta^2Z_0\left[J^*,J\right]}{\delta J_i^*\,\delta J_j}\bigg|_{J=0}=M_{ij}^{-1}$$

$$\left\langle c_ic_jc_k^*c_l^*\right\rangle_0=\frac{1}{Z_0}\frac{\delta^4Z_0\left[J^*,J\right]}{\delta J_i^*\delta J_j^*\delta J_l\delta J_k}\bigg|_{J=0}=M_{il}^{-1}M_{jk}^{-1}-M_{ik}^{-1}M_{jl}^{-1}.$$

Corresponding bosonic path-integrals can be formulated in a similar way with complex variables and periodic boundary conditions on imaginary time. The Gaussian path-integral over bosonic fields is equal to inverse of the M-matrix determinant [15].

References

[1] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)

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

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

[4] G. Rohringer, H. Hafermann, A. Toschi, A.A. Katanin, A.E. Antipov, M.I. Katsnelson, A.I. Lichtenstein, A.N. Rubtsov, and K. Held, Rev. Mod. Phys. **90**, 025003 (2018)

[5] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **77**, 033101 (2008)

[6] F. Lechermann, A.I. Lichtenstein, and M. Potthoff, Eur. Phys. J. **226**, 2591 (2017)

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

[8] N.F. Mott, *Metal-Insulator Transitions*, 2nd edn. (Taylor and Francis, London, 1990)

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

[10] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964)

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

[12] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)

[13] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[14] A.A. Abrikosow, L.P. Gorkov, and I.E. Dzyaloshinski: *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964)

[15] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Addison-Wesley, 1988)

[16] A. Kamenev: *Field Theory of Non-Equilibrium Systems* (Cambridge Univ. Press, 2011)

[17] F.A. Berezin *Method of Second Quantization* (Academic Press , New York, 1966)

[18] K. Van Houcke, E. Kozik, R. Rossi, Y. Deng, and F. Werner, arXiv:2102.04508

[19] R. Fukuda, T. Kotani, Y. Suzuki, and S. Yokojima, Prog. Theor. Phys. **92**, 833 (1994)

[20] M. Valiev and G.W. Fernando, Phys. Rev. B **54**, 7765 (1996)

[21] H. Hafermann, F. Lechermann, A.N. Rubtsov, M.I. Katsnelson, A. Georges, and A.I. Lichtenstein, Lecture Notes in Physics **843**, 145 (2012)

[22] H. Hafermann: *Numerical Approaches to Spatial Correlations in Strongly Interacting Fermion Systems* (Cuvillier Verlag, Göttingen, 2010)

[23] F. Krien and A. Valli, Phys. Rev. B **100**, 245147 (2019)

[24] S. Iskakov, A.E. Antipov, and E. Gull, Phys. Rev. B **94**, 035102 (2016)

[25] J. Gukelberger, E. Kozik, and H. Hafermann, Phys. Rev. B **96**, 035152 (2017)

[26] D. Rost, E.V. Gorelik, F. Assaad, and N. Blümer, Phys. Rev. B **86**, 155109 (2012)

[27] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, J. Phys.: Condens. Matter **9**, 7359 (1997)

[28] A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **57**, 6884 (1998)

[29] O.K. Andersen and T. Saha-Dasgupta, Phys. Rev. B **62**, R16219 (2000)

[30] V.I. Anisimov, D.E. Kondakov, A.V. Kozhevnikov, I.A. Nekrasov, Z.V. Pchelkina, J.W. Allen, S.-K. Mo, H.-D. Kim, P. Metcalf, S. Suga, A. Sekiyama, G. Keller, I. Leonov, X. Ren and D. Vollhardt, Phys. Rev. B **71**, 125119 (2005)

[31] F. Lechermann, A. Georges, A. Poteryaev, S. Biermann, M. Posternak, A. Yamasaki, and O.K. Andersen, Phys. Rev. B **74**, 125120 (2005)

[32] M. Karolak, T.O. Wehling, F. Lechermann, and A.I. Lichtenstein, J. Phys.: Condens. Matter **23**, 085601 (2011)

[33] P.E. Blöchl, Phys. Rev. B **50**, 17953 (1994)

[34] D. Grieger, C. Piefke, O.E. Peil, F. Lechermann, Phys. Rev. B **86**, 155121 (2012)

[35] A. Hausoel, M. Karolak, E. Şaşıoğlu, A. Lichtenstein, K. Held, A. Katanin, A. Toschi, and G. Sangiovanni, Nat. Commun. **8**, 16062 (2017)

[36] J Schäfer, M. Hoinkis, E. Rotenberg, P. Blaha, and R. Claessen, Phys. Rev. B **72**, 155115 (2005)

[37] F.J. Himpsel, J.A. Knapp, D.E. Eastman, Phys. Rev. B **19**, 2919 (1979)

[38] M. Karolak, G. Ulm, T. Wehling, V. Mazurenko, A. Poteryaev, and A.I. Lichtenstein, J. Electr. Spectr. **181**, 11 (2010)

[39] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003)

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

[41] E.G.C.P. van Loon, M. Rösner, M.I. Katsnelson, and T.O. Wehling,
Phys. Rev. B **104**, 045134 (2021)

[42] S. Beck, A. Hampel, O. Parcollet, C. Ederer, and A. Georges,
J. Phys.: Condens. Matter **34**, 235601 (2022)

8 DMFT for Linear Response Functions

Eva Pavarini
Institute for Advanced Simulation
Forschungszentrum Jülich

Contents

1 Introduction: the many-body problem **2**

2 From DMFT to LDA+DMFT **4**
2.1 DMFT for a toy model: The Hubbard dimer 4
2.2 Non-local Coulomb interaction 12
2.3 Quantum-impurity solvers: Continuous-time quantum Monte Carlo 14
2.4 DMFT for the single- and multi-orbital Hubbard model 19
2.5 LDA+DMFT: Model building 24

3 Linear response functions **28**
3.1 Definitions 28
3.2 DMFT and the Bethe-Salpeter equation 30
3.3 The local susceptibility: Legendre representation 33
3.4 Magnetic susceptibility for the single-band Hubbard model 34

4 Conclusion **45**

A Eigenstates of Hubbard dimer and Anderson molecule **46**

B Lehmann representation of the local Green function **50**

C Atomic magnetic susceptibility **51**

1 Introduction: the many-body problem

The electronic many-body problem, in the non-relativistic limit and in the Born-Oppenheimer approximation, is described by the Hamiltonian

$$\hat{H}_e = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_i \sum_\alpha \frac{Z_\alpha}{|\boldsymbol{r}_i - \boldsymbol{R}_\alpha|} + \sum_{i>j} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{\alpha>\alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_{\alpha'}|}, \tag{1}$$

where $\{\boldsymbol{r}_i\}$ are electron coordinates, $\{\boldsymbol{R}_\alpha\}$ nuclear coordinates and Z_α the nuclear charges. Using a complete one-electron basis, for example the basis $\{\varphi_a(\boldsymbol{r})\}$, where $\{a\}$ are the quantum numbers, we can write this Hamiltonian in second quantization as

$$\hat{H}_e = \underbrace{-\sum_{ab} t_{ab} c_a^\dagger c_b}_{\hat{H}_0} + \underbrace{\frac{1}{2}\sum_{aa'bb'} U_{aa'bb'}\; c_a^\dagger c_{a'}^\dagger c_{b'} c_b}_{\hat{H}_U}.$$

Here the hopping integrals are given by

$$t_{ab} = -\int d\mathbf{r}\; \overline{\varphi_a}(\boldsymbol{r}) \left(-\frac{1}{2}\nabla^2 \underbrace{- \sum_\alpha \frac{Z_\alpha}{|\boldsymbol{r} - \boldsymbol{R}_\alpha|}}_{v_{\rm en}(\boldsymbol{r})} \right) \varphi_b(\boldsymbol{r}),$$

while the elements of the Coulomb tensor are

$$U_{aa'bb'} = \int d\mathbf{r}_2 \int d\mathbf{r}_2\; \overline{\varphi_a}(\boldsymbol{r}_1)\, \overline{\varphi_{a'}}(\boldsymbol{r}_2)\; \frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\; \varphi_{b'}(\boldsymbol{r}_2)\, \varphi_b(\boldsymbol{r}_1).$$

In principle, all complete one-electron bases are equivalent. In practice, since, in the general case, we cannot solve the N-electron problem exactly, some bases are better than others. One possible choice for the basis are the Kohn-Sham orbitals, $\left\{\varphi_a^{\rm KS}(\boldsymbol{r})\right\}$, obtained, e.g., in the local density approximation (LDA).[1] In this case, it is useful to replace the electron-nuclei interaction $v_{\rm en}(\boldsymbol{r})$ with the DFT potential $v_R(\boldsymbol{r})$, which includes in addition the Hartree term $v_{\rm H}(\boldsymbol{r})$ and the (approximate) exchange-correlation potential $v_{\rm xc}(\boldsymbol{r})$

$$v_{\rm R}(\boldsymbol{r}) = v_{\rm en}(\boldsymbol{r}) + \underbrace{\int d\mathbf{r}'\; \frac{n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}}_{v_{\rm H}(\boldsymbol{r})} + v_{\rm xc}(\boldsymbol{r}),$$

so that

$$\tilde{t}_{ab} = -\int d\mathbf{r}\; \overline{\varphi_a^{\rm KS}}(\boldsymbol{r}) \left(-\frac{1}{2}\nabla^2 + v_{\rm R}(\boldsymbol{r}) \right) \varphi_b^{\rm KS}(\boldsymbol{r}). \tag{2}$$

To avoid double counting (DC), we have however to subtract from $\hat{H}_U$ the term $\hat{H}_{\rm DC}$, which describes the Coulomb terms already included in the hopping integrals

$$\hat{H}_e = \underbrace{-\sum_{ab} \tilde{t}_{ab}\, c_a^\dagger c_b}_{\hat{H}_0 = \hat{H}_e^{\rm LDA}} + \underbrace{\frac{1}{2}\sum_{aba'b'} \tilde{U}_{aa'bb'}\; c_a^\dagger c_{a'}^\dagger c_{b'} c_b - \hat{H}_{\rm DC}}_{\Delta\hat{H}_U}.$$

[1] For the purpose of many-body calculations the differences between LDA, GGA, or their plain extensions are in practice negligible; for simplicity, in the rest of the lecture, we thus adopt LDA as representative functional.

For weakly-correlated systems, in the Kohn-Sham basis, the effects included in $\Delta\hat{H}_U$ can, in first approximation, either be neglected or treated as a perturbation. This implies that $\hat{H}_e^{\rm LDA} \sim \hat{H}_{\rm eff}$, where $\hat{H}_{\rm eff}$ is the effective model which provides a good description of the system (at least) at low energy, and which describes emergent effective "elementary particles" and their interactions. Hypothetically, one could imagine that $\hat{H}_{\rm eff}$ is obtained via a canonical transformation, so that $\hat{H}_{\rm eff} \sim \hat{S}^{-1}\hat{H}_e\,\hat{S}$, although the exact form of the operator $\hat{S}$ is unknown.

A defining feature of strong-correlation effects is that they cannot be described via a single-electron Hamiltonian. An effective model $\hat{H}_{\rm eff}$ of form $\hat{H}_e^{\rm LDA}$ does not capture the Mott metal-insulator transition, no matter what the specific values of the parameters $\tilde{t}_{ab}$ are.[2] Thus for strongly-correlated materials the low-energy effective model must be different. The canonical Hamiltonian used to describe the Mott transition is the Hubbard model

$$\hat{H} = -\sum_{\sigma}\sum_{ii'} t^{i,i'} c^{\dagger}_{i\sigma} c_{i'\sigma} + U\sum_{i} \hat{n}_{i\uparrow}\hat{n}_{i\downarrow}, \tag{3}$$

which includes, in addition to a single-electron term, the on-site Coulomb repulsion. This model captures the essence of the Mott transition. At half filling, for $U{=}0$ it describes a paramagnetic metal, and for $t^{i,i'}(1{-}\delta_{i,i'}){=}0$ an insulating set of paramagnetic atoms. Unfortunately, differently from Hamiltonians of type $\hat{H}_e^{\rm LDA}$, Hubbard-like models cannot be solved exactly in the general case. Remarkably, till 30 years ago, no method for describing the complete phase diagram of (3) in a single coherent framework, including the paramagnetic insulating phase, was actually known. This changed between 1989 and 1992, when the dynamical mean-field theory (DMFT) was developed [1–4]. The key idea of DMFT consists in mapping the Hubbard model onto a self-consistent *auxiliary quantum-impurity problem*, which can be solved exactly. The mapping is based on the *local dynamical self-energy approximation*, very good for realistic three-dimensional lattices—and becoming exact in the infinite coordination limit [1, 2].

DMFT was initially applied only to simple cases, due to limitations in model building, computational power, and numerical methods for solving the auxiliary impurity problem (the quantum-impurity solvers). In the last twenty years remarkable progress lifted many of these limitations. First, reliable schemes to build realistic low-energy materials-specific Hubbard-like models have been devised, in particular using Kohn-Sham localized Wannier functions. This is astonishing, given that we do not know the exact operator $\hat{S}$ which gives the effective low-energy Hamiltonian, and thus a truly systematic derivation is not possible. Second, key advances in quantum-impurity solvers and increasingly more powerful supercomputers made it possible to study always more complex many-body Hamiltonians. The approach which emerged, consisting in solving, within DMFT, materials-specific many-body Hamiltonians constructed via LDA, is known as the LDA+DMFT method [5–7]. This technique (and its extensions) is now the state-of-the-art for describing strongly-correlated materials. In this lecture I will outline the basic ideas on which the method is based, its successes and its limitations. This manuscript extends the one of last year's school—in which more details on the model building aspects can be found—to the calculation of linear response functions.

[2] One can obtain an insulator by reducing the symmetry, e.g, by increasing the size of the primitive cell. This Slater-type insulator has however different properties than a Mott-type insulator.

2 From DMFT to LDA+DMFT

In this section we introduce the basics of dynamical mean-field theory. We start from a case for which we can perform analytic calculations, the two-site Hubbard Hamiltonian. This is a toy model, useful to illustrate how the method works, but for which, as we will see, DMFT is not a good approximation. Indeed, the Hubbard dimer is the worst case for DMFT, since the coordination number is the lowest possible. Next we extend the formalism to the one-band and then to the multi-orbital Hubbard Hamiltonian. For three-dimensional lattices the coordination number is typically large and thus DMFT is an excellent approximation. Finally, we summarize the modern schemes to construct materials-specific many-body models. They are based on Kohn-Sham Wannier orbitals, calculated, e.g, using the LDA functional. The solution of such models via DMFT defines the LDA+DMFT method.

2.1 DMFT for a toy model: The Hubbard dimer

The two-site Hubbard model is given by

$$\hat{H} = \varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_{\sigma} \left(c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma} \right) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{4}$$

with $i = 1, 2$. The ground state for $N = 2$ electrons (half filling) is the singlet[3]

$$|G\rangle_H = \frac{a_2(t,U)}{\sqrt{2}} \left(c^\dagger_{1\uparrow} c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow} c^\dagger_{2\uparrow} \right) |0\rangle + \frac{a_1(t,U)}{\sqrt{2}} \left(c^\dagger_{1\uparrow} c^\dagger_{1\downarrow} + c^\dagger_{2\uparrow} c^\dagger_{2\downarrow} \right) |0\rangle \tag{5}$$

with

$$a_1^2(t,U) = \frac{1}{\Delta(t,U)} \frac{\Delta(t,U) - U}{2}, \qquad a_2^2(t,U) = \frac{4t^2}{\Delta(t,U)} \frac{2}{\Delta(t,U) - U}, \tag{6}$$

and

$$\Delta(t,U) = \sqrt{U^2 + 16t^2}. \tag{7}$$

The energy of this state is

$$E_0(2) = 2\varepsilon_d + \frac{1}{2}\big(U - \Delta(t,U)\big). \tag{8}$$

In the $T \to 0$ limit, using the Lehmann representation (see Appendix B), one can show that the local Matsubara Green function for spin σ takes then the form

$$\begin{aligned} G^\sigma_{i,i}(i\nu_n) = & \frac{w_+(t,U)}{i\nu_n - \big(\underbrace{E_0(2) - \varepsilon_d + t - \mu}_{E_0(2) - E_-(1) - \mu}\big)} + \frac{w_-(t,U)}{i\nu_n - \big(\underbrace{-E_0(2) + U + 3\varepsilon_d + t - \mu}_{E_+(3) - E_0(2) - \mu}\big)} \\ & + \frac{w_-(t,U)}{i\nu_n - \big(\underbrace{E_0(2) - \varepsilon_d - t - \mu}_{E_0(2) - E_+(1) - \mu}\big)} + \frac{w_+(t,U)}{i\nu_n - \big(\underbrace{-E_0(2) + U + 3\varepsilon_d - t - \mu}_{E_-(3) - E_0(2) - \mu}\big)}, \end{aligned} \tag{9}$$

[3] Eigenstates and eigenvalues of the Hubbard dimer for arbitrary filling can be found in Appendix A.

where $\nu_n = \pi(2n+1)/\beta$ are fermionic Matsubara frequencies and $\mu = \varepsilon_d + U/2$ is the chemical potential. The weights are

$$w_\pm(t,U) = \frac{1}{4}\big(1 \pm w(t,U)\big), \qquad \text{with} \qquad w(t,U) = 2a_1(t,U)a_2(t,U) = \frac{4t}{\Delta(t,U)}. \tag{10}$$

The local Green function can be rewritten as the average of the Green function for the bonding ($k=0$) and the anti-bonding state ($k=\pi$), i.e.,

$$G^\sigma_{i,i}(i\nu_n) = \frac{1}{2}\left(\underbrace{\frac{1}{i\nu_n + \mu - \varepsilon_d + t - \Sigma^\sigma(0;i\nu_n)}}_{G^\sigma(0;i\nu_n)} + \underbrace{\frac{1}{i\nu_n + \mu - \varepsilon_d - t - \Sigma^\sigma(\pi;i\nu_n)}}_{G^\sigma(\pi;i\nu_n)} \right). \tag{11}$$

The self-energy is given by

$$\Sigma^\sigma(k;i\nu_n) = \frac{U}{2} + \frac{U^2}{4}\frac{1}{i\nu_n + \mu - \varepsilon_d - \frac{U}{2} - e^{ik}\,3t}. \tag{12}$$

The self-energies $\Sigma^\sigma(0;i\nu_n)$ and $\Sigma^\sigma(\pi;i\nu_n)$ differ due to the phase $e^{ik} = \pm 1$ in their denominators. The local self-energy is, by definition, the average of the two

$$\begin{aligned}\Sigma^\sigma_l(i\nu_n) = \frac{1}{2}\Big(\Sigma^\sigma(\pi;i\nu_n) + \Sigma^\sigma(0;i\nu_n)\Big) &= \frac{U}{2} + \frac{U^2}{4}\,\frac{i\nu_n + \mu - \varepsilon_d - \frac{U}{2}}{\big(i\nu_n + \mu - \varepsilon_d - \frac{U}{2}\big)^2 - (3t)^2} \\ &= \frac{U}{2} + \frac{U^2}{4}\,\frac{i\nu_n}{(i\nu_n)^2 - (3t)^2}.\end{aligned} \tag{13}$$

The difference

$$\begin{aligned}\Delta\Sigma^\sigma_l(i\nu_n) = \frac{1}{2}\Big(\Sigma^\sigma(\pi;i\nu_n) - \Sigma^\sigma(0;i\nu_n)\Big) &= -\frac{U^2}{4}\,\frac{3t}{\big(i\nu_n + \mu - \varepsilon_d - \frac{U}{2}\big)^2 - (3t)^2} \\ &= -\frac{U^2}{4}\,\frac{3t}{(i\nu_n)^2 - (3t)^2},\end{aligned} \tag{14}$$

thus measures the importance of non-local effects; it would be zero if the self-energy was independent of k. Next we define the hybridization function

$$F^\sigma(i\nu_n) = \frac{\big(t + \Delta\Sigma^\sigma_l(i\nu_n)\big)^2}{i\nu_n + \mu - \varepsilon_d - \Sigma^\sigma_l(i\nu_n)} \tag{15}$$

which for $U{=}0$ becomes

$$F^\sigma_0(i\nu_n) = \frac{t^2}{i\nu_n}. \tag{16}$$

By using these definitions, we can rewrite the local Green function as

$$G^\sigma_{i,i}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - F^\sigma(i\nu_n) - \Sigma^\sigma_l(i\nu_n)}. \tag{17}$$

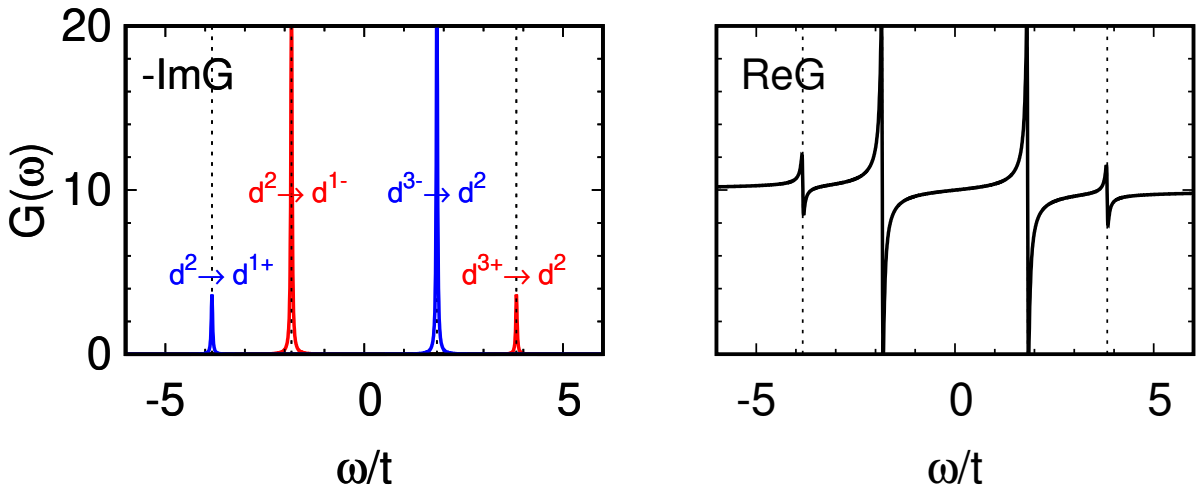


Fig. 1: *Hubbard dimer: Imaginary (left) and real (right) part of the retarded Green function, obtained setting $i\nu_n \to \omega + i\delta$ (analytic continuation) in Eq. (9). Red lines: $k = 0$ contribution. Blue lines: $k = \pi$ contribution. Dashed lines: Poles of the retarded Green function. Parameters: $t = 1$, $U = 4$. The weight of the poles yielding the smaller peaks, $w_-(t, U)$, defined in Eq. (10), goes to zero for $U \to 0$. In the atomic limit, instead, all four poles have the same weight; the energies of the two positive (negative) poles become identical, however.*

The associated retarded Green function, obtained via analytic continuation ($i\nu_n \to \omega + i\delta$), is shown in Fig. 1. It is important to point out that, as one may see from the formulas just discussed, the local Green function and the local self-energy satisfy the *local Dyson equation*

$$\Sigma_l^\sigma(i\nu_n) = \frac{1}{\mathfrak{G}_{i,i}^\sigma(i\nu_n)} - \frac{1}{G_{i,i}^\sigma(i\nu_n)}, \tag{18}$$

where $\mathfrak{G}_{i,i}^\sigma(i\nu_n)$ is given by

$$\mathfrak{G}_{i,i}^\sigma(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - F^\sigma(i\nu_n)}. \tag{19}$$

Thus, one could think of mapping the Hubbard dimer into an auxiliary quantum-impurity model, chosen such that, within certain approximations, the impurity Green function is as close as possible to the local Green function of the original problem. How can we do this? Let us adopt as auxiliary model the Anderson molecule

$$\hat{H}^{\rm A} = \varepsilon_s \sum_\sigma \hat{n}_{s\sigma} - t \sum_\sigma \left(c_{d\sigma}^\dagger c_{s\sigma} + c_{s\sigma}^\dagger c_{d\sigma} \right) + \varepsilon_d \sum_\sigma \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow} \,, \tag{20}$$

where s labels the uncorrelated bath site and d the correlated quantum-impurity site. The first constraint would be that Hamiltonian (20) has a ground state with the same occupations of the 2-site Hubbard model, i.e., at half filling, $n_d = n_s = 1$. Such a *self-consistency condition* is satisfied if $\varepsilon_s = \mu = \varepsilon_d + U/2$. This can be understood by comparing the Hamiltonian matrices of the two models in the Hilbert space with $N = 2$ electrons. To this end, we first order the

two-electron states of the Hubbard dimer as

$$\begin{aligned} |1\rangle &= c^\dagger_{1\uparrow}c^\dagger_{2\uparrow}|0\rangle, & |4\rangle &= \tfrac{1}{\sqrt{2}}(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow}c^\dagger_{2\uparrow})|0\rangle, \\ |2\rangle &= c^\dagger_{1\downarrow}c^\dagger_{2\downarrow}|0\rangle, & |5\rangle &= c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}|0\rangle, \\ |3\rangle &= \tfrac{1}{\sqrt{2}}(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{1\downarrow}c^\dagger_{2\uparrow})|0\rangle, & |6\rangle &= c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}|0\rangle. \end{aligned} \tag{21}$$

In this basis the Hamiltonian of the Hubbard dimer has the matrix form

$$\hat{H}_2(\varepsilon_d, U, t) = \begin{pmatrix} 2\varepsilon_d & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d{+}U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_d{+}U \end{pmatrix}. \tag{22}$$

The ground state, the singlet given in Eq. (5), can be obtained by diagonalizing the lower $3{\times}3$ block. For the Anderson molecule, ordering the basis in the same way ($1 \to d, 2 \to s$), this Hamiltonian becomes

$$\hat{H}^A_2(\varepsilon_d, U, t; \varepsilon_s) = \begin{pmatrix} \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_d{+}\varepsilon_s & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d{+}U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_s \end{pmatrix}. \tag{23}$$

Comparing the lower $3{\times}3$ block of $\hat{H}^A_2(\varepsilon_d, U, t; \varepsilon_s)$ with the corresponding block of $\hat{H}_2(\varepsilon_d, U, t)$ we can see that, unless $\varepsilon_s = \mu = \varepsilon_d + U/2$, the two doubly occupied states $|5\rangle$ and $|6\rangle$ have different energies and thus the two sites $i = 1, 2$ are differently occupied in the ground state. By setting $\varepsilon_s = \mu$ we find that

$$\hat{H}^A_2(\varepsilon_d, U, t; \mu) = \hat{H}_2(\varepsilon_d{+}\tfrac{U}{4}, \tfrac{U}{2}, t). \tag{24}$$

The $N = 2$ ground state of $\hat{H}^A_2(\varepsilon_d, U, t; \mu)$ has thus the form of the ground-state for the Hubbard dimer

$$|G\rangle_A = \frac{a_2(t, U/2)}{\sqrt{2}}\left(c^\dagger_{d\uparrow}c^\dagger_{s\downarrow} - c^\dagger_{d\downarrow}c^\dagger_{s\uparrow}\right)|0\rangle + \frac{a_1(t, U/2)}{\sqrt{2}}\left(c^\dagger_{d\uparrow}c^\dagger_{d\downarrow} + c^\dagger_{s\uparrow}c^\dagger_{s\downarrow}\right)|0\rangle, \tag{25}$$

and the condition $n_s = n_d = 1$ is satisfied. Since $\varepsilon_s \neq \varepsilon_d$, however, the eigenstates of $\hat{H}^{\mathrm{A}}$ for one electron ($N = 1$) or one hole ($N = 3$) are not the bonding and antibonding states of the

Hubbard dimer.[4] The impurity Green function is then given by

$$G^{\sigma}_{d,d}(i\nu_n) = \frac{1}{4}\Bigg(\frac{1+w'(t,U)}{i\nu_n - (E_0(2) - E_-(1) - \mu)} + \frac{1-w'(t,U)}{i\nu_n - (E_+(3) - E_0(2) - \mu)} \\ \frac{1-w'(t,U)}{i\nu_n - (E_0(2) - E_+(1) - \mu)} + \frac{1+w'(t,U)}{i\nu_n - (E_-(3) - E_0(2) - \mu)} \Bigg), \tag{26}$$

where

$$E_0(2) - E_\pm(1) - \mu = -\Big(E_\pm(3) - E_0(2) - \mu \Big) = -\frac{1}{4}\Big(2\Delta(t,U/2) \pm \Delta(t,U) \Big) \tag{27}$$

and

$$w'(t,U) = \frac{1}{2}\frac{32t^2 - U^2}{\Delta(t,U)\Delta(t,U/2)}\,. \tag{28}$$

After some rearrangement we obtain a much simpler expression

$$G^{\sigma}_{d,d}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}^{\sigma}_0(i\nu_n) - \Sigma^{\sigma}_A(i\nu_n)}\,. \tag{29}$$

The impurity self-energy equals the local self-energy of the Hubbard dimer

$$\Sigma^{\sigma}_A(i\nu_n) = \frac{U}{2} + \frac{U^2}{4}\frac{i\nu_n}{(i\nu_n)^2 - (3t)^2}\,, \tag{30}$$

as one may see comparing it to equation (13). The hybridization function is given by

$$\mathcal{F}^{\sigma}_0(i\nu_n) = \frac{t^2}{i\nu_n}\,, \tag{31}$$

as for the non-interacting Hubbard dimer, Eq. (16). For $U = 0$, $G^{\sigma}_{d,d}(i\nu_n)$ equals the non-interacting impurity Green function

$$G^{0\sigma}_{d,d}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}^{\sigma}_0(i\nu_n)}\,. \tag{32}$$

The impurity Green function thus satisfies the *impurity Dyson equation*

$$\Sigma^{\sigma}_A(i\nu_n) = \frac{1}{G^{0\sigma}_{d,d}(i\nu_n)} - \frac{1}{G^{\sigma}_{d,d}(i\nu_n)}\,. \tag{33}$$

In Fig. 2 we show the retarded impurity Green function of the Anderson molecule (orange, right panels) and the retarded local Green function of the 2-site Hubbard model, both in the local self-energy approximation (blue, right panels) and exact (blue, left panels). Comparing left and right panels we can see that setting $\Delta\Sigma^{\sigma}_l(\omega) = 0$ yields large errors. The right panels demonstrate, however, that the spectral function of the Hubbard dimer with $\Delta\Sigma^{\sigma}_l(\omega) = 0$ is quite similar to the one of the Anderson molecule. The small remaining deviations come from the fact that,

[4] The complete list of eigenvalues and eigenvectors of the Anderson molecule for $\varepsilon_s = \varepsilon_d + U/2$ and arbitrary electron number N can be found in Appendix A.

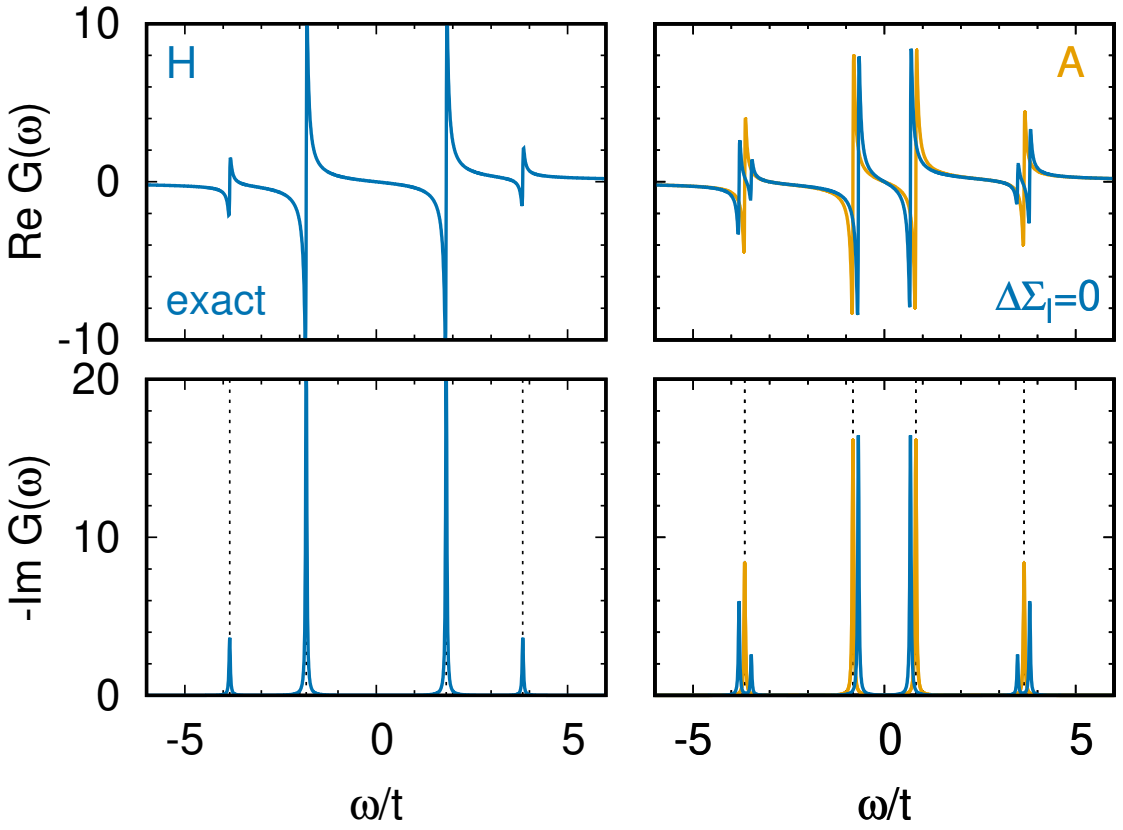


Fig. 2: *Retarded Green function of the Hubbard dimer ($t = 1$, $U = 4$) and of the Anderson molecule ($\varepsilon_s = \varepsilon_d + U/2$) in the zero temperature limit. Left panels: Hubbard dimer, exact Green function. Right panels, blue: Hubbard dimer in the local self-energy approximation, i.e., with $\Delta\Sigma_l^\sigma(\omega) = 0$. Right panels, orange: Anderson molecule. Dashed lines: Poles for the Hubbard dimer (left) or the Anderson molecule (right).*

for the Hubbard dimer, in the impurity Dyson equation, the non-interacting impurity Green function is replaced by $\mathfrak{G}^\sigma_{i,i}(i\nu_n)$ in the local self-energy approximation, i.e., by the *bath* Green function

$$\mathcal{G}^\sigma_{i,i}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}^\sigma_l(i\nu_n)}\,, \tag{34}$$

where

$$\mathcal{F}^\sigma_l(i\nu_n) = \frac{t^2}{i\nu_n + \mu - \varepsilon_d - \Sigma^\sigma_A(i\nu_n)}\,. \tag{35}$$

We are now in the position of explaining how DMFT works for the Hamiltonian of the Hubbard dimer, choosing the Anderson molecule Hamiltonian (20) as the auxiliary quantum-impurity model. The procedure can be split in the following steps

1. Build the initial quantum impurity model with $G^{0\sigma}_{d,d}(i\nu_n) = G^{0\sigma}_{i,i}(i\nu_n)$. The initial bath is thus defined by energy $\varepsilon_s = \varepsilon_d$ and hopping t.

2. Calculate the local Green function $G^\sigma_{d,d}(i\nu_n)$ for the auxiliary model.

3. Use the local Dyson equation to calculate the impurity self-energy

$$\Sigma_A^\sigma(i\nu_n) = \frac{1}{G_{d,d}^{0\sigma}(i\nu_n)} - \frac{1}{G_{d,d}^{\sigma}(i\nu_n)}.$$

4. Calculate the local Green function of the Hubbard dimer setting the self-energy equal to the one of the quantum-impurity model

$$G_{i,i}^\sigma(i\nu_n) \sim \frac{1}{2}\left(\frac{1}{i\nu_n + \mu - \varepsilon_d + t - \Sigma_A^\sigma(i\nu_n)} + \frac{1}{i\nu_n + \mu - \varepsilon_d - t - \Sigma_A^\sigma(i\nu_n)}\right).$$

5. Calculate a new bath Green function $\mathcal{G}_{i,i}^\sigma(i\nu_n)$ from the local Dyson equation

$$\mathcal{G}_{i,i}^\sigma(i\nu_n) = \frac{1}{\Sigma_A^\sigma(i\nu_n) + 1/G_{i,i}^\sigma(i\nu_n)}.$$

6. Build a new $G_{d,d}^{0\sigma}(i\nu_n)$ from $\mathcal{G}_{i,i}^\sigma(i\nu_n)$.

7. Restart from the second step.

8. Iterate till self-consistency, i.e., here till $n_d^\sigma = n_i^\sigma$ and $\Sigma_A^\sigma(i\nu_n)$ does not change any more.

The Anderson molecule satisfies the self-consistency requirements for $\varepsilon_s = \mu$. The remaining difference between $G_{d,d}^\sigma(i\nu_n)$, the impurity Green function, and $G_{i,i}^\sigma(i\nu_n)$, the local Green function of the Hubbard dimer in the local self-energy approximation, arises from the difference in the associated hybridization functions

$$\Delta\mathcal{F}_l(i\nu_n) = \mathcal{F}_l^\sigma(i\nu_n) - \mathcal{F}_0^\sigma(i\nu_n) = t^2p^2\left(-\frac{2}{i\nu_n} + \frac{1}{i\nu_n - \varepsilon_a} + \frac{1}{i\nu_n + \varepsilon_a}\right), \tag{36}$$

where $p^2 = U^2/8\varepsilon_a^2$ and $\varepsilon_a = \sqrt{9t^2 + U^2/4}$. The error made is small, however, as shown in the right panels of Fig. 2. To further improve we would have to modify the auxiliary model adding more bath sites. Staying with the Anderson molecule, in Fig. 3 we compare in more detail its spectral function with the exact spectral function of the Hubbard dimer. The figure emphasizes several important conclusions. The top right panel reminds us that DMFT is not a good approximation for molecular complexes with two (or few) correlated sites. This is because in such systems the coordination number is the lowest possible, the worst case for dynamical mean-field theory. In three-dimensional crystals, instead, the coordination number is typically large enough to make dynamical mean-field theory an excellent approximation. The bottom left panel of Fig. 3 shows that, in the local-self-energy approximation, the agreement between Anderson and Hubbard Green functions remains very good for any U value. This indicates that when the local-self-energy approximation works well, as in the case of three-dimensional crystals, it can be successfully used to study the behavior of a given system as a function of U. Leaving for a moment DMFT aside, the two bottom panels of Fig. 3 show that the evolution with U is different for the impurity Green function of the Anderson molecule and the exact local Green function of the Hubbard dimer. Anticipating the discussion of later sections, if we compare to

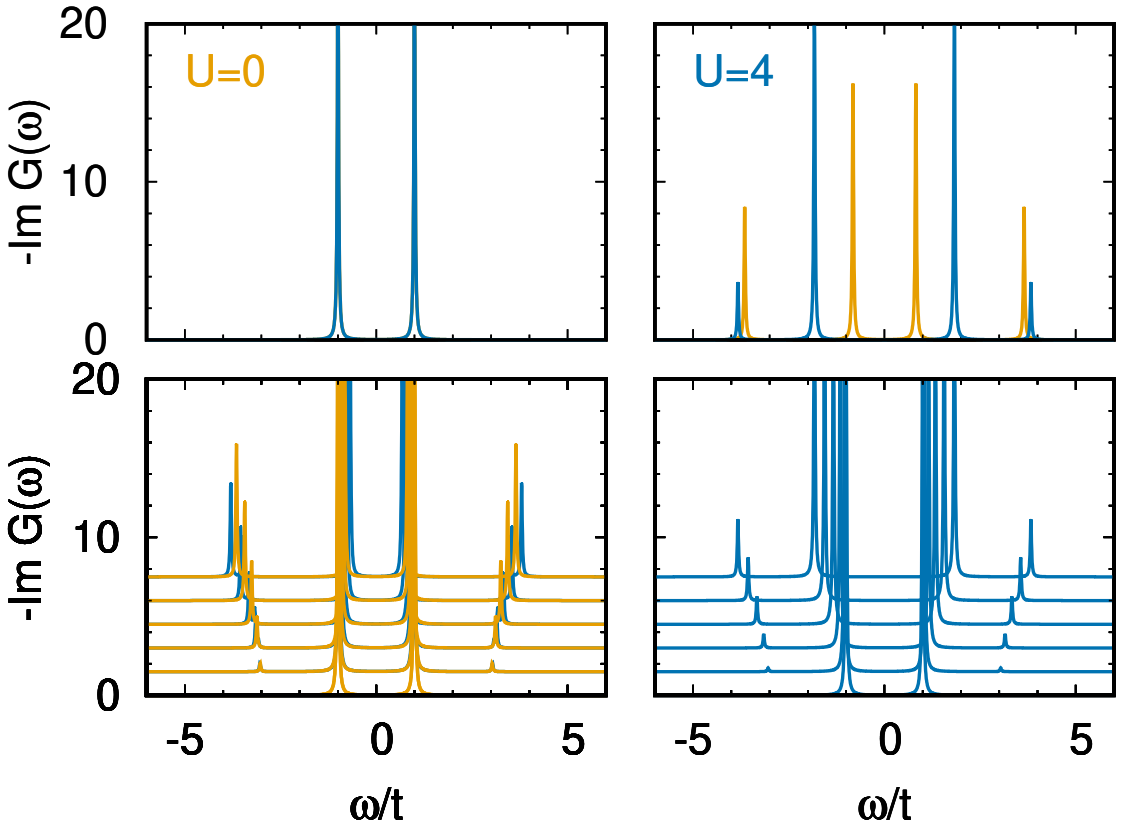


Fig. 3: *Imaginary part of the retarded Green function of the Anderson molecule (orange) and Hubbard dimer (blue) in the zero temperature limit. In the bottom left panel the local self-energy approximation is adopted for the Hubbard dimer; in all other cases the exact Green function of the Hubbard dimer is shown. Parameters:* $t = 1$, $\varepsilon_s = \mu$. *Top:* $U = 0$ *(left) and* $U = 4t$ *(right). Bottom: Evolution with increasing* U *from* 0 *to* $4t$ *in equal steps.*

the spectral function of the actual *lattice* Hubbard model, we could say that the Hubbard dimer captures well the evolution of the Hubbard bands and the gap in the large-U limit. On the other hand, the Anderson molecule partially captures the behavior of the central "quasi-particle" or "Kondo" peak, although the Kondo effect itself is unrealistically described; as a matter of fact, the Kondo energy gain (the "Kondo temperature") is perturbative ($\propto t^2/U$) in the case of the Anderson molecule, while it is exponentially small for a Kondo impurity in a metallic bath. Going back to DMFT, this also points to the possible shortcomings of calculations in which the quantum-impurity model for the lattice Hubbard model is solved via exact diagonalization, however using a single bath site or very few; this might perhaps be sufficient in the large-gap limit,[5] but is bound to eventually fail approaching the metallic regime. Indeed, this failure is one of the reasons why the solution of the Kondo problem required the development of—at the time new—non-perturbative techniques such as the numerical renormalization group.

[5]For a discussion of bath parametrization in exact diagonalization and the actual convergence with the number of bath sites for the lattice Hubbard model see Ref. [8].

2.2 Non-local Coulomb interaction

In Sec. 2.1 we have seen that the local Coulomb interaction gives rise, *alone*, to non-local self-energy terms, which can be very important. What is, instead, the effect of the non-local part of the Coulomb interaction? For a Hubbard dimer, extending the Coulomb interaction to first neighbors leads to the Hamiltonian

$$\begin{aligned}\hat{H} =&\varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_{\sigma} \left(c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma} \right) + U \sum_{i=1,2} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \\ &+ \sum_{\sigma\sigma'} \left(V - 2J_V - J_V\, \delta_{\sigma\sigma'} \right) \hat{n}_{1\sigma} \hat{n}_{2\sigma'} - J_V \sum_{i\neq i'} \left(c^\dagger_{i\uparrow} c_{i\downarrow} c^\dagger_{i'\downarrow} c_{i'\uparrow} + c^\dagger_{i'\uparrow} c^\dagger_{i'\downarrow} c_{i\uparrow} c_{i\downarrow} \right), \end{aligned} \tag{37}$$

where the parameters in the last two terms are the intersite direct (V) and exchange (J_V) Coulomb interaction. For two electrons the Hamiltonian, in a matrix form, becomes

$$\hat{H}_2^{\rm NL} = \begin{pmatrix} 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d + V - J_V & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d + U & -J_V \\ 0 & 0 & 0 & -\sqrt{2}t & -J_V & 2\varepsilon_d + U \end{pmatrix},$$

where the basis is defined in Eq. (21). In the atomic ($t = 0$) limit, the triplet states, $|1\rangle$, $|2\rangle$ and $|3\rangle$, have lower energy than the singlet states, $|4\rangle$, $|5\rangle$ and $|6\rangle$, as one can see by comparing the diagonal elements of the upper and lower 3×3 block of the matrix $\hat{H}_2^{\rm NL}$ here above. This is due to the fact that J_V is positive (ferromagnetic) and $V < U$. The triplet can remain the ground multiplet even for finite t. If, however, J_V is sufficiently small, the ground state is a singlet, as in the case $V{=}J_V{=}0$. Setting for simplicity $J_V = 0$, we notice that $\hat{H}_2^{\rm NL} = \hat{H}_2(\varepsilon'_d, U', t)$, where the right-hand-side term is the $N{=}2$-electron Hamiltonian of the $J_V{=}V{=}0$ Hubbard dimer, Eq. (22), with parameters $\varepsilon'_d = \varepsilon_d + V/2$ and $U'{=}U{-}V$. The $N{=}2$ ground state is thus still given by Eq. (5), provided that we replace U with U' in the coefficients. Eventually, in the limiting case $U{=}V$, $\hat{H}_2^{\rm NL}$ equals the corresponding Hamiltonian of an effective non-correlated dimer, $\hat{H}_2(\varepsilon'_d, 0, t)$. What happens away from half filling? For $N{=}1$ electrons, eigenvectors and eigenvalues are the same as in the $V{=}0$ case; for $N{=}3$ electrons all energies are shifted by $2V$. Summarizing, we can obtain the Green function for $V \neq 0$ from Eq. (9) setting

$$\begin{aligned} E_\pm(N{=}1, U; V) &= E_\pm(N{=}1, U; 0) &&= \varepsilon_d \pm t \\ E_\pm(N{=}3, U; V) &= E_\pm(N{=}3, U; 0) + 2V &&= 3\varepsilon_d \pm t + U + 2V \\ E_0(N{=}2, U; V) &= E_0(N{=}2, U{-}V; 0) + V &&= E_0(2, U{-}V) + V \\ \mu(U; V) &= \mu(U) + V &&= \mu + V \\ w_\pm(t, U; V) &= w_\pm(t, U{-}V; 0). \end{aligned}$$

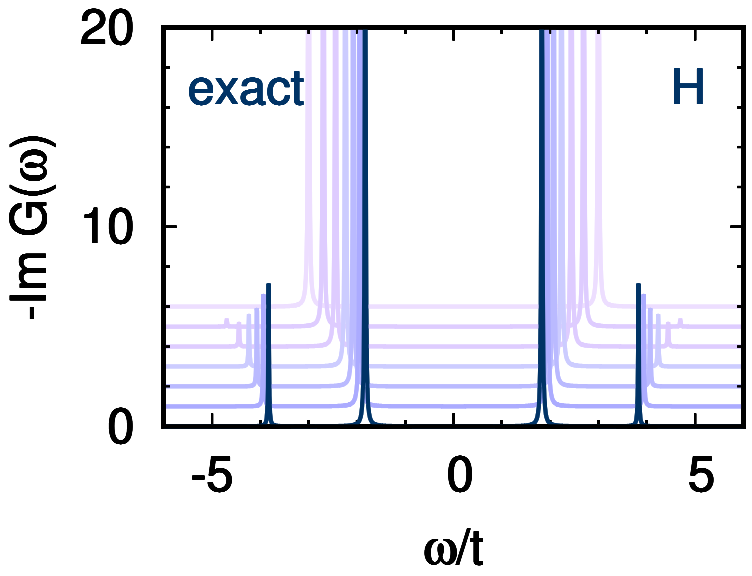


Fig. 4: *Imaginary part of the retarded Green function of the Hubbard dimer in the zero temperature limit (U=4, t=1), increasing the intersite Coulomb repulsion V from 0 to V=U=4 in equal steps; we have set $J_V = 0$. The dark blue line corresponds to $V = 0$.*

Thus we have, recalling that $U' = U-V$,

$$G^{\sigma}_{i,i}(i\nu_n) = \frac{w_+(t,U')}{i\nu_n-\left(E'_0(2)-\frac{1}{2}V-\varepsilon_d+t-\mu'\right)}+\frac{w_-(t,U')}{i\nu_n-\left(-E'_0(2)+U'+\frac{1}{2}V+3\varepsilon_d+t-\mu'\right)}$$
$$+\frac{w_-(t,U')}{i\nu_n-\left(E'_0(2)-\frac{1}{2}V-\varepsilon_d-t-\mu'\right)}+\frac{w_+(t,U')}{i\nu_n-\left(-E'_0(2)+U'+\frac{1}{2}V+3\varepsilon_d-t-\mu'\right)}, \qquad (38)$$

where we set $\mu' = \mu-V/2 = \varepsilon_d+U'/2$ and $E'_0(2) = E_0(2,U')$. The associated spectral function is shown in Fig. 4. The figure illustrates that increasing V from 0 to U makes the spectra progressively closer to the one of a non-correlated system. Eventually, for U=V, only two poles contribute, since $w_-(t,U') = 0$. In this limit, the spectral function is identical to the one of the non-interacting Hubbard dimer, however with an enhanced effective hopping, $t \longrightarrow t + V/2$. We can thus say that, in first approximation, the (positive) intersite coupling V effectively reduces the strength of correlations in the Hubbard dimer. In conclusion, the case of the Hubbard dimer explains why *strong-correlation* effects typically appear when the *local term of the electron-electron repulsion dominates*, i.e., when it is much larger than long-range terms. A hypothetical system in which the Coulomb interaction strength is independent on the distance between sites (for the dimer, U=V) is likely to be already well described via an *effective weakly correlated model*. Of course, in real materials, the effects of long-range Coulomb repulsion can be much more complicated than in the two-site model just discussed, but the general considerations made here remain true even in realistic cases.

2.3 Quantum-impurity solvers: Continuous-time quantum Monte Carlo

For the case of the Anderson molecule exact diagonalization is the simplest quantum-impurity solver and the one that provides most insights. Methods based on Quantum Monte Carlo (QMC) sampling are often, however, the only option for realistic multi-orbital and/or multi-site models. Thus, here we explain how to obtain the impurity Green function of the Anderson molecule via hybridization-expansion continuous-time QMC [9], a very successful QMC-based quantum-impurity solver. In this approach, the first step consists in splitting the Hamiltonian into bath ($\hat{H}_{\rm bath}$), hybridization ($\hat{H}_{\rm hyb}$), and local ($\hat{H}_{\rm loc}$) terms

$$\hat{H}^{\rm A} = \underbrace{\varepsilon_s \sum_\sigma \hat{n}_{s\sigma}}_{\hat{H}_{\rm bath}} \underbrace{-t \sum_\sigma \left(c^\dagger_{d\sigma} c_{s\sigma} + c^\dagger_{s\sigma} c_{d\sigma} \right)}_{\hat{H}_{\rm hyb}} + \underbrace{\varepsilon_d \sum_\sigma \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow}}_{\hat{H}_{\rm loc}} . \tag{39}$$

Next, we write the partition function Z as a perturbation series in the hybridization. To this end, we define $\hat{H}_0 = \hat{H}_{\rm bath} + \hat{H}_{\rm loc}$ and rewrite the partition function as

$$Z = {\rm Tr} \left(e^{-\beta(\hat{H}_0 - \mu \hat{N})} \hat{V}(\beta) \right) \tag{40}$$

where the operator $\hat{V}(\beta)$ is given by

$$\hat{V}(\beta) = e^{\beta(\hat{H}_0 - \mu\hat{N})}\, e^{-\beta(\hat{H}_0 + \hat{H}_{\rm hyb} - \mu\hat{N})} = \sum_m \underbrace{\int_0^\beta d\tau_1 \cdots \int_{\tau_{m-1}}^\beta d\tau_m}_{\int d\boldsymbol{\tau}^m} \underbrace{(-1)^m \prod_{l=m}^1 \hat{H}_{\rm hyb}(\tau_l)}_{\hat{O}^m(\boldsymbol{\tau})} , \tag{41}$$

and

$$\hat{H}_{\rm hyb}(\tau_l) = e^{\tau_l(\hat{H}_0 - \mu\hat{N})}\, \hat{H}_{\rm hyb}\, e^{-\tau_l(\hat{H}_0 - \mu\hat{N})} = -t \sum_\sigma \left(c^\dagger_{d\sigma_l}(\tau_l) c_{s\sigma_l}(\tau_l) + c^\dagger_{s\sigma_l}(\tau_l) c_{d\sigma_l}(\tau_l) \right). \tag{42}$$

In this expansion, the only terms that contribute to the trace are even order ones ($m = 2k$) and they are products of impurity (d) and bath (s) creator-annihilator pairs. We can thus rewrite

$$\int d\boldsymbol{\tau}^{2k} \longrightarrow \int d\boldsymbol{\tau}^k \int d\bar{\boldsymbol{\tau}}^k \quad \text{and} \qquad \hat{O}^{2k}(\boldsymbol{\tau}) \longrightarrow \sum_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}} \hat{O}^{2k}_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau}, \bar{\boldsymbol{\tau}}) \tag{43}$$

where

$$\hat{O}^{2k}_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau}, \bar{\boldsymbol{\tau}}) = (t)^{2k} \prod_{i=1}^k \left(c^\dagger_{d\bar{\sigma}_i}(\bar{\tau}_i) c_{s\bar{\sigma}_i}(\bar{\tau}_i) c^\dagger_{s\sigma_i}(\tau_i) c_{d\sigma_i}(\tau_i) \right). \tag{44}$$

The vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_k)$ gives the spins $\{\sigma_i\}$ associated with the k impurity annihilators at imaginary times $\{\tau_i\}$, while $\bar{\boldsymbol{\sigma}} = (\bar{\sigma}_1, \bar{\sigma}_2, ..., \bar{\sigma}_k)$ gives the spins $\{\bar{\sigma}_i\}$ associated with the k impurity creators at imaginary times $\{\bar{\tau}_i\}$. It follows that the local and bath traces can be

decoupled and the partition function can be rewritten as

$$\frac{Z}{Z_{\rm bath}} = \sum_k \int d\boldsymbol{\tau}^k \int d\bar{\boldsymbol{\tau}}^k \sum_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}} d^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}})\, t^k_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) \tag{45}$$

$$d^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) = \frac{t^{2k}}{Z_{\rm bath}}\, {\rm Tr}_{\rm bath}\left(e^{-\beta(\hat{H}_{\rm bath}-\mu\hat{N}_s)}\mathcal{T}\prod_{i=k}^{1} c^\dagger_{s\sigma_i}(\tau_i)c_{s\bar{\sigma}_i}(\bar{\tau}_i)\right) \tag{46}$$

$$t^k_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) = {\rm Tr}_{\rm loc}\left(e^{-\beta(\hat{H}_{\rm loc}-\mu\hat{N}_d)}\mathcal{T}\prod_{i=k}^{1} c_{d\sigma_i}(\tau_i)c^\dagger_{d\bar{\sigma}_i}(\bar{\tau}_i)\right), \tag{47}$$

where $Z_{\rm bath} = 1 + 2e^{-\beta(\varepsilon_s-\mu)} + e^{-2\beta(\varepsilon_s-\mu)}$ and

$$c_{d\sigma}(\tau) = e^{\tau(\hat{H}_{\rm loc}-\mu\hat{N}_d)}c_{d\sigma}e^{-\tau(\hat{H}_{\rm loc}-\mu\hat{N}_d)}, \qquad c_{s\sigma}(\tau) = e^{\tau(\hat{H}_{\rm bath}-\mu\hat{N}_s)}c_{s\sigma}e^{-\tau(\hat{H}_{\rm bath}-\mu\hat{N}_s)}.$$

The trace involving only bath operators is simple to calculate, since $\hat{H}_{\rm bath}$ describes an independent-electron problem for which Wick's theorem holds. It is given by the determinant

$$d^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) = \det\left(\mathcal{F}^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}})\right) \tag{48}$$

of the $k\times k$ non-interacting hybridization-function matrix, with elements

$$\left(\mathcal{F}^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}})\right)_{i',i} = \mathcal{F}^0_{\bar{\sigma}_{i'},\sigma_i}(\bar{\tau}_{i'}-\tau_i), \tag{49}$$

where

$$\mathcal{F}^0_{\bar{\sigma},\sigma}(\tau) = \delta_{\bar{\sigma},\sigma}\frac{t^2}{1+e^{-\beta(\varepsilon_s-\mu)}} \times \begin{cases} -e^{-\tau(\varepsilon_s-\mu)} & \tau > 0, \\ +e^{-(\beta+\tau)(\varepsilon_s-\mu)} & \tau < 0. \end{cases} \tag{50}$$

This is the imaginary time Fourier transform of the hybridization function introduced previously

$$\mathcal{F}^0_{\bar{\sigma},\sigma}(i\nu_n) = \frac{t^2}{i\nu_n - (\varepsilon_s-\mu)}\,\delta_{\bar{\sigma},\sigma}. \tag{51}$$

The calculation of the local trace is in general more complicated. In the case discussed here, the Hamiltonian does not flip spins. Thus only terms with an equal number of creation and annihilation operators *per spin* contribute to the local trace, and we can express the partition function in expansion orders per spin, k_σ. This yields [10]

$$\frac{Z}{Z_{\rm bath}} = \left(\prod_\sigma \sum_{k_\sigma=0}^{\infty} \int d\boldsymbol{\tau}_\sigma^{k_\sigma} \int d\bar{\boldsymbol{\tau}}_\sigma^{k_\sigma}\right) d^k_{\bar{\boldsymbol{\sigma}},\boldsymbol{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) t^k_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}), \tag{52}$$

where the vectors $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_\uparrow, \boldsymbol{\sigma}_\downarrow)$ and $\bar{\boldsymbol{\sigma}} = (\bar{\boldsymbol{\sigma}}_\uparrow, \bar{\boldsymbol{\sigma}}_\downarrow)$ have $(k_\uparrow, k_\downarrow)$ components, and for each k_σ component $\sigma_i = \bar{\sigma}_i = \sigma$. Thus

$$t^k_{\boldsymbol{\sigma},\bar{\boldsymbol{\sigma}}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) = {\rm Tr}_{\rm loc}\left(e^{-\beta(\hat{H}_{\rm loc}-\mu\hat{N}_d)}\,\mathcal{T}\prod\nolimits_\sigma\prod\nolimits_{i=k_\sigma}^{1} c_{d\sigma}(\tau_{\sigma_i})c^\dagger_{d\sigma}(\bar{\tau}_{\bar{\sigma}_i})\right). \tag{53}$$

The latter can be calculated analytically. To do this, first we parametrize all configurations for a given spin via a timeline $[0,\beta)$ plus a number of creator/annihilator pairs which define segments

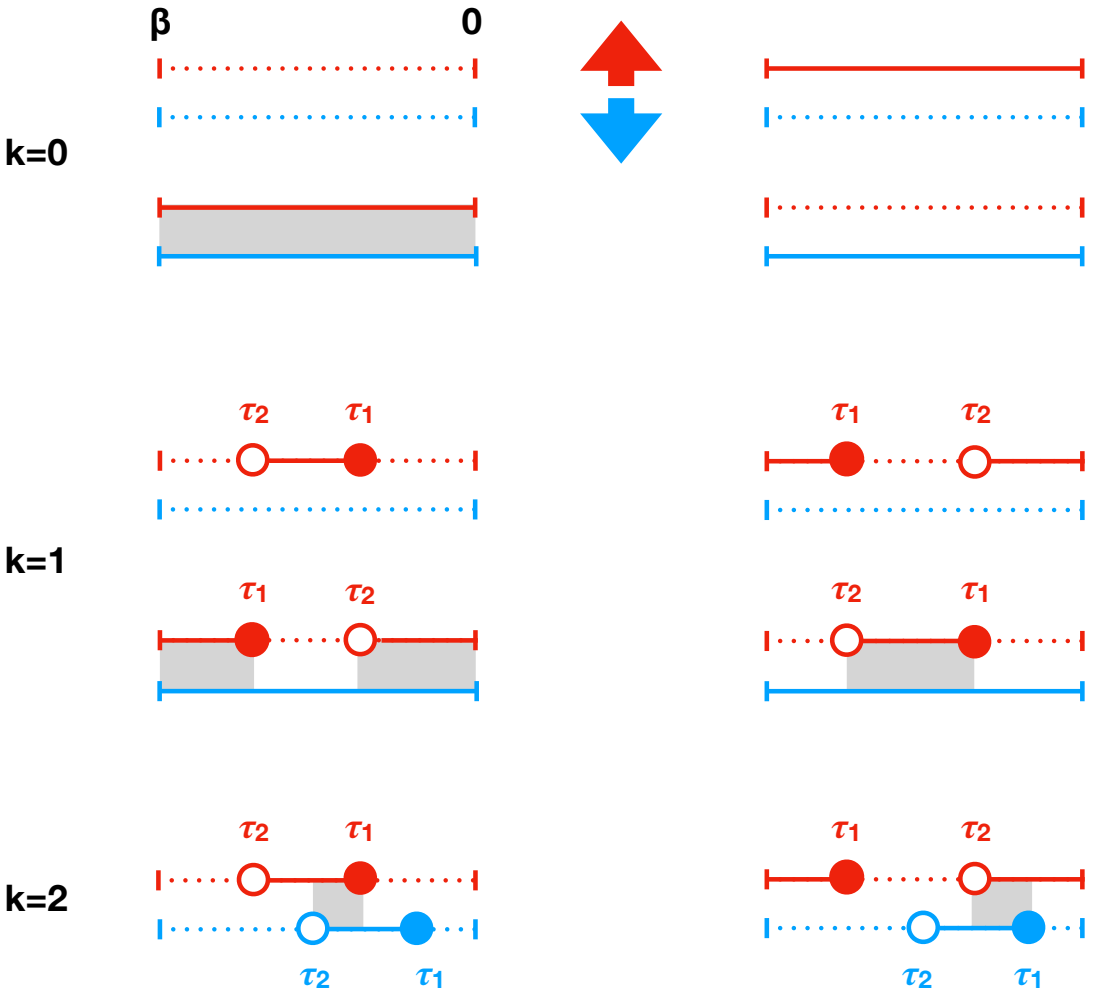


Fig. 5: *Representative configurations contributing to the local trace at zeroth, first and second order. The timelines for spin up are red and those for spin down are blue. The filled circles correspond to the insertion of a creator (time τ_1), and the empty circles to the insertion of an annihilator (time τ_2). Dotted lines represent the vacuum state for a given spin, full lines the occupied state. The grey boxes indicate the regions in which $l_{\uparrow,\downarrow} \neq 0$.*

on the timeline. At zeroth order two possible configurations exist per spin, an empty timeline, which corresponds to the vacuum state $|0\rangle$, and a full timeline, which corresponds to the state $c^\dagger_{d\sigma}|0\rangle$. A given configuration yields, at order $k = k_\uparrow + k_\downarrow$

$$t^k_{\sigma,\bar{\sigma}}(\boldsymbol{\tau},\bar{\boldsymbol{\tau}}) = \left(\prod_\sigma s_\sigma^{k_\sigma}\right) e^{-\sum_{\sigma\sigma'}((\varepsilon_d-\mu)\delta_{\sigma\sigma'}+\frac{U}{2}(1-\delta_{\sigma,\sigma'}))l_{\sigma,\sigma'}}, \tag{54}$$

where $l_{\sigma,\sigma'}$ is the length of the overlap of the τ segments for spins σ and σ', respectively, while $s_\sigma = \mathrm{sign}(\tau_{\sigma_1} - \bar{\tau}_{\sigma_1})$ is the fermionic sign. Possible configurations at order $k = 0, 1, 2$ are shown in Fig. 5. At order $k = 0$, summing up the contribution of the four configurations shown in Fig. 5 yields the local partition function $Z_{\rm loc} = 1 + 2e^{-\beta(\varepsilon_d-\mu)} + e^{-\beta(2(\varepsilon_d-\mu)+U)}$. Order $k = 1$ is already more complicated. Setting $\varepsilon_s = \mu$ as in the self-consistent solution, the contribution

to the bath trace in this case is

$$d^1_{\bar\sigma\sigma}(\tau_1,\tau_2) = \mathcal{F}^0_{\bar\sigma\sigma}(\tau_1,\tau_2) = -\frac{t^2}{2}\delta_{\sigma,\bar\sigma}\ \mathrm{sign}(\tau_1-\tau_2). \tag{55}$$

The local trace at the same order is instead given by

$$t^1_{\sigma\bar\sigma}(\tau_2,\tau_1) = \mathrm{Tr}_{\rm loc}\Big(e^{-\beta(\hat H_{\rm loc}-\mu\hat N_d)}\mathcal{T}c_{d\sigma}(\tau_2)c^\dagger_{d\sigma}(\tau_1)\Big). \tag{56}$$

We can now calculate the contribution at half filling of the four $k=1$ configurations shown in Fig. 5. In the case $k_\uparrow = 1$ and $k_\downarrow = 0$ we have, going from left to right in each row

$$t^1_{\uparrow\uparrow}(\tau_2,\tau_1) = \begin{cases} +e^{-(\tau_2-\tau_1)(\varepsilon_d-\mu)} & = +e^{+\tau_{21}U/2} \\ -e^{-(\beta-(\tau_1-\tau_2))(\varepsilon_d-\mu)} & = -e^{(\beta+\tau_{21})U/2} \\ -e^{-\beta(2(\varepsilon_d-\mu)+U)+(\tau_1-\tau_2)(\varepsilon_d-\mu+U)} & = -e^{-\tau_{21}U/2} \\ +e^{-(\tau_2-\tau_1)(\varepsilon_d-\mu+U)-\beta(\varepsilon_d-\mu)} & = +e^{(\beta-\tau_{21})U/2} \end{cases} \tag{57}$$

where $\tau_{21} = \tau_2 - \tau_1$ and $\mu = \varepsilon_d + U/2$. Similar results can be obtained for $k_\uparrow = 0$ and $k_\downarrow = 1$. Summing up all terms up to order one we find

$$\begin{aligned} \frac{Z}{Z_{\rm bath}} &\sim Z_{\rm loc} + \sum_\sigma \int_0^\beta d\tau_2 \int_0^\beta d\tau_1\, d^1_{\sigma\sigma}(\tau_1,\tau_2)\, t^1_{\sigma\sigma}(\tau_2,\tau_1) \\ &\sim Z_{\rm loc}\left(1 - \beta\frac{1-e^{\frac{\beta U}{2}}}{1+e^{\frac{\beta U}{2}}}\frac{2t^2}{U}\right). \end{aligned} \tag{58}$$

The exact formula of the partition function can be obtained from the eigenvalues and eigenvectors in Appendix A

$$\frac{Z}{Z_{\rm bath}} = Z_{\rm loc}\frac{3(1+e^{\frac{\beta U}{2}}) + e^{\frac{\beta U}{4}}\left(4e^{-\frac{\beta\Delta(t,U)}{4}} + 4e^{+\frac{\beta\Delta(t,U)}{4}} + e^{+\frac{\beta\Delta(t,U/2)}{2}} + e^{-\frac{\beta\Delta(t,U/2)}{2}}\right)}{8\big(1+e^{\frac{\beta U}{2}}\big)}. \tag{59}$$

Its Taylor expansion in powers of t/U yields, at second order, the expression above. Going back to Eq. (56), one can observe that, for $k=1$, the local trace is proportional to the local Green function, $G^\sigma_{d,d}(\tau)$. Indeed, $G^\sigma_{d,d}(\tau)$ can be calculated using the configurations just described—provided that we start from $k=1$ and we divide by the hybridization function. More specifically, for $k=1$ and $\tau>0$ we have

$$G^\sigma_{d,d}(\tau) \sim -\frac{1}{\beta}\int_0^\beta\int_0^\beta \underbrace{d\tau_2 d\tau_1\ d^1_{\sigma\sigma}(\tau_1,\tau_2)t^1_{\sigma\sigma}(\tau_2,\tau_1)}_{w^1}\ \delta\big(\tau-(\tau_2-\tau_1)\big)\ \frac{1}{\mathcal{F}^0_{\sigma\sigma}(\tau_1-\tau_2)}. \tag{60}$$

We are now ready to generalize to arbitrary order. Taking all k values into account, the partition function can be expressed as the sum over all configurations $\{c\}$, i.e., in short

$$Z = \sum_c w_c = \sum_c |w_c|\ \mathrm{sign}\ w_c. \tag{61}$$

In a compact form, we can write $w_c = d\boldsymbol{\tau}_c \; d_c \; t_c$ where $d\boldsymbol{\tau}_c = \prod_\sigma \prod_i^{k_\sigma} d\tau_{\sigma_i} d\bar{\tau}_{\bar{\sigma}_i}$, and d_c and t_c are the bath and local traces for the configuration c. This expression of the partition function shows that we can interpret $|w_c|$ as the sampling weight of configuration c. A generic observable $\hat{O}$ can then be obtained as the Monte Carlo average on a finite number of configurations N_c

$$\langle \hat{O} \rangle = \frac{\sum_c \langle \hat{O} \rangle_c |w_c| \,\mathrm{sign}\; w_c}{\sum_c |w_c| \,\mathrm{sign}\; w_c} = \frac{\sum_c \mathrm{sign}\; w_c \langle \hat{O} \rangle_c \; |w_c| / \sum_c |w_c|}{\sum_c \mathrm{sign}\; w_c \; |w_c| / \sum_c |w_c|} \approx \frac{\frac{1}{N_c} \sum_c^{N_c} \langle \hat{O} \,\rangle_c \,\mathrm{sign}\; w_c}{\frac{1}{N_c} \sum_c \mathrm{sign}\; w_c}. \quad (62)$$

The term $\frac{1}{N_c} \sum_c \mathrm{sign}\; w_c$ in the denominator is the average fermionic sign. When this is small, much longer runs are required to obtain data of the same quality; eventually the computational time can become so long that the calculation is unfeasible—in these cases we have a sign problem. In practice, the QMC simulation starts from a random configuration c. Next we propose an update $c \to c'$. Within the Metropolis-Hastings algorithm, the acceptance ratio is

$$R_{c\to c'} = \min\left(1, \frac{p_{c'\to c}}{p_{c\to c'}} \frac{|w_{c'}|}{|w_c|}\right), \quad (63)$$

where $p_{c\to c'}$ is the proposal probability for the update $c \to c'$. In the approach described here, known as *segment solver*, the basic updates are addition and removal of segments, antisegments (segments winding over the borders of the timeline, see Fig. 5), or complete lines. As example, let us consider the insertion of a segment for spin σ. A segment is made by a creator and an annihilator. The creator is inserted at time $\tau_{\rm in}$; the move is rejected if $\tau_{\rm in}$ is in a region where a segment exists. If created, the segment can have at most length $l_{\rm max}$, given by the distance between $\tau_{\rm in}$ and the time at which the next creator is, hence

$$p_{c\to c'} = \frac{d\bar{\tau}}{\beta} \frac{d\tau}{l_{\rm max}} \,. \quad (64)$$

The proposal probability of the reverse move (removing a segment) is instead given by the inverse of the number of existing segments

$$p_{c'\to c} = \frac{1}{k_\sigma + 1} \,. \quad (65)$$

The acceptance ratio for the insertion of a segment becomes then

$$R_{c\to c'} = \min\left(1, \frac{\beta l_{max}}{k_\sigma + 1} \left| \frac{d_{c'}}{d_c} \frac{t_{c'}}{t_c} \right| \right). \quad (66)$$

For the impurity Green function, here the most important observable, the direct average yields

$$\langle \hat{O} \rangle_c = \langle G^\sigma_{d,d} \rangle_c = \sum_{\sigma'} \sum_{i=1}^{k_\sigma} \sum_{j=1}^{k_\sigma} \Delta(\tau, \tau_{\sigma' j} - \bar{\tau}_{\sigma' j}) \left(M^{k'_\sigma}\right)_{\sigma' j, \sigma' i} \delta_{\sigma, \sigma_{\sigma' j}} \delta_{\sigma, \bar{\sigma}_{\sigma' i}} \quad (67)$$

where $M^k = \left(\mathcal{F}^k\right)^{-1}$ is the inverse of the hybridization matrix and

$$\Delta(\tau, \tau') = -\frac{1}{\beta} \begin{cases} \delta(\tau - \tau') & \tau' > 0 \\ -\delta(\tau - (\tau' + \beta)) & \tau' < 0 \end{cases} . \quad (68)$$

One can verify that at order $k = 1$ this indeed returns Eq. (60).

2.4 DMFT for the single- and multi-orbital Hubbard model

The Hubbard Hamiltonian (3) is in principle the simplest model for the description of the Mott metal-insulator transition. In the tight-binding approximation it becomes

$$\hat{H} = \varepsilon_d \sum_{\sigma i} \hat{n}_{i\sigma} - t \sum_{\sigma \langle ii' \rangle} c^\dagger_{i\sigma} c_{i'\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{69}$$

where $\langle ii' \rangle$ is a sum over first neighbors. As discussed in the introduction, for $U = 0$, at half-filling, this Hamiltonian describes a metallic band. For $t = 0$ it describes an insulating collection of disconnected atoms. Somewhere in-between, at a critical value of t/U, a metal to insulator transition must occur. In this section we will discuss the DMFT solution of (69) and the picture of the metal-insulator transition emerging from it. The first step consists in mapping the original many-body Hamiltonian into an effective quantum-impurity model, such as the Anderson Hamiltonian

$$\hat{H}^A = \underbrace{\sum_{\boldsymbol{k}\sigma} \varepsilon^s_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k}\sigma}}_{\hat{H}_{\rm bath}} + \underbrace{\sum_{\boldsymbol{k}\sigma} \left(V^s_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k}\sigma} c_{d\sigma} + \text{h.c.} \right)}_{\hat{H}_{\rm hyb}} + \underbrace{\varepsilon_d \sum_\sigma \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow}}_{\hat{H}_{\rm imp}}. \tag{70}$$

In this model the on-site Coulomb repulsion U appears only in the impurity Hamiltonian, $\hat{H}_{\rm imp}$, while the terms $\hat{H}_{\rm bath}$ and $\hat{H}_{\rm hyb}$, describe, respectively, the bath and the bath-impurity hybridization. In the next step, the quantum-impurity model is solved. Differently from the case of the Anderson molecule, this cannot be done analytically. It requires non-perturbative numerical methods, such as exact diagonalization, the numerical renormalization group, density-matrix renormalization group or QMC. Here we describe the DMFT self-consistency loop for a QMC quantum-impurity solver. Solving the quantum-impurity model yields the impurity Green function $G^\sigma_{d,d}(i\nu_n)$. From the impurity Dyson equation we can calculate the impurity self-energy

$$\Sigma^\sigma_A(i\nu_n) = \left(G^{0\sigma}_{d,d}(i\nu_n) \right)^{-1} - \left(G^\sigma_{d,d}(i\nu_n) \right)^{-1}. \tag{71}$$

Next, we adopt the local self-energy approximation, i.e., we assume that the self-energy of the Hubbard model equals the impurity self-energy. Then, the local Green function is given by

$$G^\sigma_{i_c,i_c}(i\nu_n) = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \frac{1}{i\nu_n + \mu - \varepsilon_{\boldsymbol{k}} - \Sigma^\sigma_A(i\nu_n)}, \tag{72}$$

where $N_{\boldsymbol{k}}$ is the number of $\boldsymbol{k}$ points. The local Dyson equation is used once more, this time to calculate the bath Green function $\mathcal{G}^\sigma(i\nu_n)$, which in turn defines a new quantum-impurity model. This procedure is repeated until self-consistency is reached, i.e., the number of electrons is correct and the self-energy does not change anymore (within a given numerical accuracy). In this situation we have

$$G^\sigma_{i_c,i_c}(i\nu_n) \sim G^\sigma_{d,d}(i\nu_n). \tag{73}$$

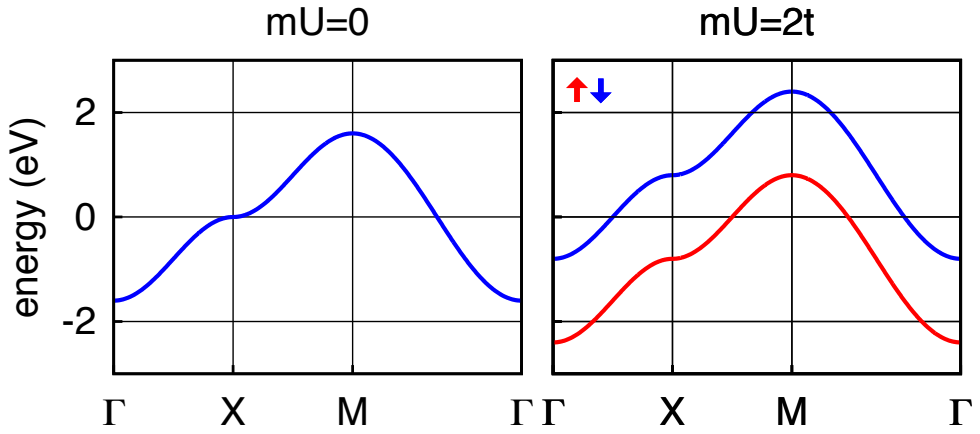


Fig. 6: *The metal-insulator transition in ferromagnetic Hartree-Fock. The calculation is for a square lattice tight-binding model with dispersion* $\varepsilon_{\boldsymbol{k}} = -2t(\cos k_x + \cos k_y)$.

It is important to underline that self-consistency is key to the success of DMFT in describing the metal-to-insulator transition. This can, perhaps, be best understood looking once more at the effects of self-consistency in a simpler approach, the static mean-field Hartree-Fock method.[6] If we chose the same primitive cell as in DMFT, the Hartree-Fock self-energy matrix is

$$\Sigma^{\sigma}_{i,i'}(i\nu_n) = U\left(\frac{n}{2} - s_\sigma m\right)\delta_{i,i'}, \tag{74}$$

where $s_\sigma = +1$ for spin up and $s_\sigma = -1$ for spin down and $m = m_+ = (n_\uparrow - n_\downarrow)/2$, with $n_\sigma = n_{i\sigma}$. The approximation is then identical to replacing the Hubbard Hamiltonian with

$$\hat{H}_{\rm HF} = \sum_{\boldsymbol{k}\sigma}\left(\varepsilon_{\boldsymbol{k}} + U\left(\frac{1}{2} - s_\sigma m\right)\right)\hat{n}_{\boldsymbol{k}\sigma}. \tag{75}$$

This shows that $h_{\rm eff} = 2Um$ plays the role of an effective magnetic field (Weiss field). The self-consistency criterion is

$$\bar{n}_\sigma = \bar{n}_{i\sigma} = \langle \hat{n}_{i\sigma}\rangle_{\rm HF}\,, \tag{76}$$

where the expectation value $\langle \hat{n}_{i\sigma}\rangle_{\rm HF}$ is calculated using the Hamiltonian $\hat{H}_{\rm HF}$, which in turn depends on $\bar{n}_\sigma$ via m. This gives the self-consistency equation

$$m = \frac{1}{2}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}\sigma}\frac{\sigma e^{-\beta(\varepsilon_{\boldsymbol{k}} + U(\frac{1}{2} - s_\sigma m) - \mu)}}{1 + e^{-\beta(\varepsilon_{\boldsymbol{k}} + U(\frac{1}{2} - s_\sigma m) - \mu)}}\,. \tag{77}$$

If we set $m = 0$ the equation is satisfied; for such a trivial solution the static mean-field correction in Eq. (75) merely redefines the chemical potential and has therefore no effect. For sufficiently large U, however, a non-trivial solution ($m \neq 0$) can be found. If $m \neq 0$ the spin up and spin down bands split, and eventually a gap can open. This is shown in Fig. 6. The static mean-field correction in Eq. (75) equals the contribution of the Hartree diagram to the self-energy, $\Sigma^{\sigma}_{\rm H}(i\nu_n) = U\bar{n}_{-\sigma}$. In many-body perturbation theory, however, $\bar{n}_\sigma = 1/2$, i.e., $m = 0$.

[6] Keeping in mind that many self-consistent solutions obtained with the Hartree-Fock method are spurious.

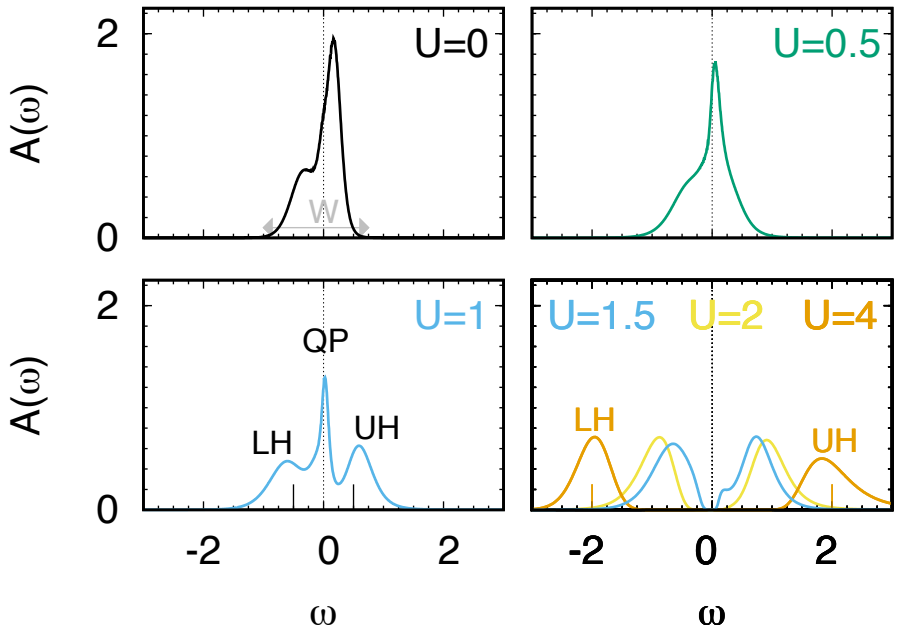


Fig. 7: *$VOMoO_4$: LDA+DMFT spectral function at finite temperature for $0 \leq U \leq 4$. Energies are in eV and spectral functions in states/spin/eV. The calculations have been done using a continuous-time hybridization-expansion QMC solver [10]. A detailed LDA+DMFT study of the electronic and magnetic properties $VOMoO_4$ can be found in Ref. [11].*

In the self-consistent static mean-field approximation, instead, m can differ from zero, and a phenomenon not described by the mere Hartree diagram can be captured, ferromagnetism in a correlated metal. If the band splitting, given by $h_{\rm eff}{=}2Um$, is larger than the bandwidth, the system can even become an insulator.

In DMFT the role of the Weiss mean field is played by the bath Green function $\mathcal{G}^{\sigma}_{i,i}(i\nu_n)$. The emerging picture of the Mott transition is described in Fig. 7 for a representative single-band material. In the $U = 0$ limit, the spectral function $A_0(\omega)$ is metallic at half filling (top left panel). For finite U, if we set $\Sigma^{\sigma}_A(\omega) = 0$ as initial guess, the DMFT self-consistency loop starts with $A(\omega) = A_0(\omega)$. For small U/t, the *converged* spectral function $A(\omega)$ is still similar to $A_0(\omega)$. This can be seen comparing the $U = 0.5$ and $U = 0$ panels in Fig. 7. Further increasing U/t, sizable spectral weight is transferred from the zero-energy quasi-particle peak to the lower (LH) and upper (UH) Hubbard bands, centered at $\omega \sim \pm U/2$. This can be observed in the $U = 1$ panel of Fig. 7. The system is still metallic, but with strongly renormalized masses and short lifetimes, reflected in the narrow quasi-particle (QP) peak. Finally, for U larger than a critical value ($U \geq 1.5$ in the figure) a gap opens and the system becomes a Mott insulator. When this happens the self-energy diverges at low frequency, where

$$\Sigma^{\sigma}_A(\omega + i0^+) \sim \frac{U}{2} + \frac{U^2}{4}\frac{a(t,U)}{\omega + i0^+}. \tag{78}$$

In the large U/t limit the gap increases linearly with the Coulomb repulsion, i.e., $E_g(1) \sim U - W$, where W is the bandwidth.

The single-band Hubbard model describes the essence of the metal-insulator transition. In order to understand this phenomenon in materials, however, we have to study multi-orbital Hubbard-like Hamiltonians. These have the form

$$\hat{H} = \hat{H}_0 + \hat{H}_U$$
$$\hat{H}_0 = -\sum_{ii'}\sum_{\sigma}\sum_{mm'} t^{i,i'}_{m\sigma,m'\sigma'}\, c^{\dagger}_{im\sigma} c_{i'm'\sigma'}$$
$$\hat{H}_U = \frac{1}{2}\sum_{i}\sum_{\sigma\sigma'}\sum_{mm'}\sum_{pp'} U_{mpm'p'}\, c^{\dagger}_{im\sigma} c^{\dagger}_{ip\sigma'} c_{ip'\sigma'} c_{im'\sigma},$$

where m, m' and p, p' are different orbitals and the Coulomb tensor is local. The DMFT approach can be extended to solve models of this type, mapping them to multi-orbital quantum-impurity models. The main changes with respect to the formalism introduced in the previous section are then the following

$$\varepsilon_{\boldsymbol{k}} \to (H_{\boldsymbol{k}})_{m\sigma,m'\sigma'} \qquad (i\nu_n+\mu) \to (i\nu_n+\mu)\,\hat{1}_{m\sigma,m'\sigma'}$$
$$t^{i,i'} \to t^{i,i'}_{m\sigma,m'\sigma'} \qquad \varepsilon_d \to \varepsilon^{i,i'}_{m\sigma,m'\sigma'} = -t^{i,i}_{m\sigma,m'\sigma'}$$

where $\hat{1}$ is the identity matrix. As a consequence, the local Green function, the bath Green function, the hybridization function and the self-energy also become matrices

$$\mathcal{G}^{\sigma}(i\nu_n) \to \mathcal{G}^{\sigma,\sigma'}_{m,m'}(i\nu_n) \qquad G^{\sigma}(i\nu_n) \to G^{\sigma,\sigma'}_{m,m'}(i\nu_n) \qquad \Sigma^{\sigma}(i\nu_n) \to \Sigma^{\sigma,\sigma'}_{m,m'}(i\nu_n).$$

The corresponding generalization of the self-consistency loop is shown schematically in Fig. 8. Although the extension of DMFT to Hubbard models with many orbitals might appear straightforward, in practice it is not. The bottleneck is the solution of the generalized multi-orbital quantum-impurity problem. The most flexible solvers available so far are all based on QMC. Despite being flexible, QMC-based approaches have limitations. These can be classified in two types. First, with increasing the number of degrees of freedom, calculations become very quickly computationally too expensive—how quickly depends on the specific QMC algorithm used and the actual implementation. Thus, going beyond a rather small number of orbitals and reaching the zero-temperature limit is unfeasible in practice. The second type of limitation is more severe. Increasing the number of degrees of freedom leads, eventually, to the infamous sign problem; when this happens, QMC calculations cannot be performed at all. In order to deal with limitations of the first type, it is crucial to restrict QMC calculations to the essential degrees of freedom; furthermore, we should exploit symmetries, develop fast algorithms and use the power of massively parallel supercomputers to reduce the actual computational time. For the second type of problems not a lot can be done; nevertheless, it has been shown that a severe sign problem might appear earlier with some basis choices than with others [10]. Although eventually we cannot escape it, this suggests that the model set up can be used as a tool to expand the moderate sign-problem zone. For what concerns symmetries, in the paramagnetic

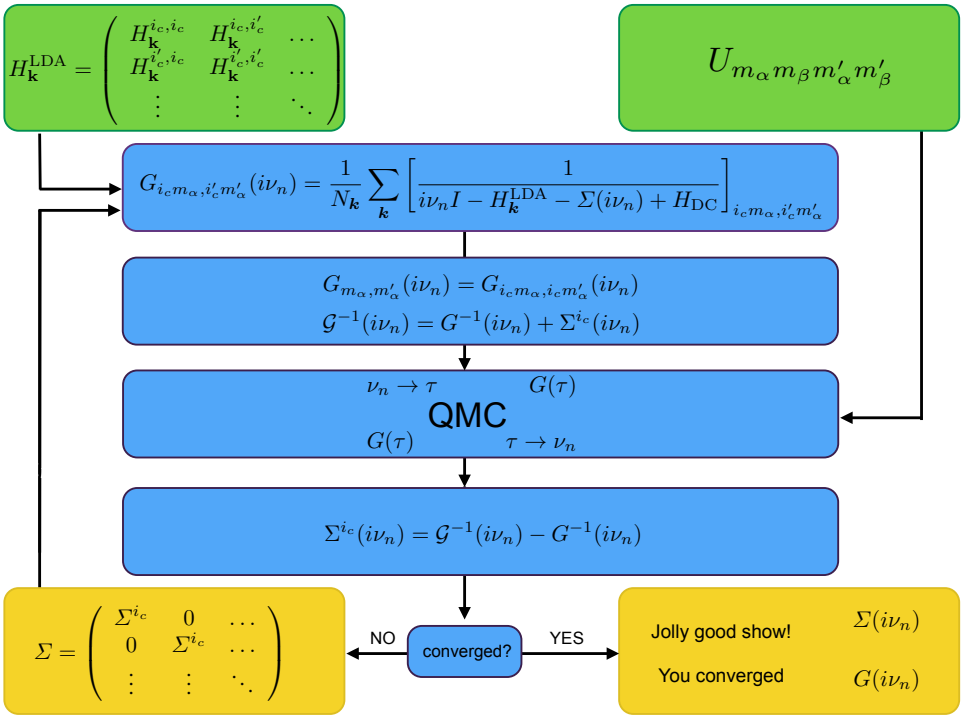


Fig. 8: *LDA+DMFT self-consistency loop. The one-electron Hamiltonian is built in the basis of Bloch states obtained from localized Wannier functions, for example in the local-density approximation (LDA); this gives* $H_{\boldsymbol{k}}^{\mathrm{LDA}}$*. The set* $\{i_c\}$ *labels the equivalent correlated sites inside the unit cell. The local Green-function matrix is at first calculated using an initial guess for the self-energy matrix. The bath Green-function matrix is then obtained via the Dyson equation and used to construct an effective quantum-impurity model. The latter is solved via a quantum-impurity solver, here quantum Monte Carlo (QMC). This yields the impurity Green-function matrix. Through the Dyson equation the self-energy is then obtained, and the procedure is repeated until self-consistency is reached.*

case and in absence of spin-orbit interaction or external fields, an obvious symmetry to exploit is the rotational invariance of spins, from which follows

$$X^{\sigma,\sigma'}_{m,m'}(i\nu_n) = \delta_{\sigma,\sigma'}\, X_{m,m'}(i\nu_n),$$

where $X = \mathcal{G}, G, \Sigma$. In addition, if we use a basis of real functions, the local Green-function matrices are real and symmetric in imaginary time τ, hence

$$X^{\sigma,\sigma'}_{m,m'}(i\nu_n) = \delta_{\sigma,\sigma'}\, X_{m,m'}(i\nu_n) = \delta_{\sigma,\sigma'}\, X_{m',m}(i\nu_n).$$

Finally, often the unit cell contains several equivalent correlated sites, indicated as $\{i_c\}$ in Fig. 8. In order to avoid expensive cluster calculations, we can use space-group symmetries to construct the matrices $\mathcal{G}, G, \Sigma$ at a given site i'_c from the corresponding matrices at an equivalent site, e.g., $i_c = 1$. Space-group symmetries also tell us if some matrix elements are zero. For example, for a model with only t_{2g} (or only e_g) states, in cubic symmetry, in the paramagnetic case and in absence of spin-orbit interaction or external fields, we have

$$X^{\sigma,\sigma'}_{m,m'}(i\nu_n) = \delta_{\sigma,\sigma'}\, X_{m,m}(i\nu_n)\, \delta_{m,m'}.$$

2.5 LDA+DMFT: Model building

The state-of-the art approach for building realistic Hubbard-like models relies on constructing, for a given system, *materials-specific* Kohn-Sham Wannier functions $\varphi^{\rm KS}_{im\sigma}(\boldsymbol{r})$. These can be obtained via electronic structure calculations based on density-functional theory [5–7], e.g., in the LDA approximation.[7] After we have built the complete one-electron basis, the first steps in model-building are those already described in the introduction. We recall here the essential points and then discuss the next stage. The many-body Hamiltonian can be expressed as $\hat{H} = \hat{H}_0 + \hat{H}_U - \hat{H}_{\rm DC}$, with

$$\hat{H}_0 = \hat{H}^{\rm LDA} = -\sum_{\sigma}\sum_{ii'}\sum_{mm'} t^{i,i'}_{m,m'} c^{\dagger}_{im\sigma} c_{i'm'\sigma},$$

$$\hat{H}_U = \frac{1}{2}\sum_{ii'jj'}\sum_{\sigma\sigma'}\sum_{mm'pp'} U^{iji'j'}_{mp\,m'p'} c^{\dagger}_{im\sigma} c^{\dagger}_{jp\sigma'} c_{j'p'\sigma'} c_{i'm'\sigma}.$$

The double-counting correction $\hat{H}_{\rm DC}$ arises from the fact that the hopping integrals are calculated replacing the electron-nuclei interaction $v_{\rm en}(\boldsymbol{r})$ with the self-consistent DFT reference potential

$$v_{\rm R}(\boldsymbol{r}) = v_{\rm en}(\boldsymbol{r}) + \underbrace{\int d\boldsymbol{r}'\, \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}}_{v_{\rm H}(\boldsymbol{r})} + v_{\rm xc}(\boldsymbol{r}),$$

which includes the long-range Hartree term $v_{\rm H}(\boldsymbol{r})$ and the exchange-correlation contribution $v_{\rm xc}(\boldsymbol{r})$. We thus have to subtract from $\hat{H}_U$ the effects already included in $\hat{H}_0$

$$\hat{H}_U \to \Delta\hat{H}_U = \hat{H}_U - \hat{H}_{\rm DC}.$$

Unfortunately we do not know which important correlation effects are indeed included in $\hat{H}_0$ via $v_{\rm R}(\boldsymbol{r})$, and therefore the exact expression of $\Delta\hat{H}_U$ is also unknown. The remarkable successes of the LDA suggest, however, that in many materials the LDA is overall a good approximation, and therefore, in those systems at least, the term $\Delta\hat{H}_U$ can be completely neglected. What about strongly-correlated materials? Even in correlated systems, most likely the LDA works rather well for the delocalized electrons or in describing the average or the long-range Coulomb effects. Thus one can think of separating the electrons into *uncorrelated* and *correlated*; only for the latter we do take the correction $\Delta\hat{H}_U$ into account explicitly, assuming furthermore that $\Delta\hat{H}_U$ is local or almost local [5], since we know that it is the local term which is responsible for most non-trivial many-body effects. Typically, correlated electrons are those that partially retain their atomic character, e.g., those that originate from localized d and f shells; for convenience, here we assume that in a given system they stem from a single atomic shell l (e.g., d for

[7] Using GGA or similar functionals in place of LDA yields minor differences in the many-body Hamiltonian; instead, using LDA+U or similar approximations yields Hartree-Fock-like effects that would have to be subtracted via the double-counting correction.

transition-metal oxides or f for heavy-fermion systems) and label their states with the atomic quantum numbers l and $m = -l, \ldots, l$ of that shell. Thus

$$U^{iji'j'}_{mp\,m'p'} \sim \begin{cases} U^l_{mp\,m'p'} & iji'j' = iiii \quad \wedge \quad mp, m'p' \in l \\ 0 & iji'j' \neq iiii \quad \vee \quad mp, m'p' \notin l. \end{cases}$$

Within this approximation $\Delta\hat{H}_U$ is replaced by $\Delta\hat{H}^l_U = \hat{H}^l_U - \hat{H}^l_{\rm DC}$, where $\hat{H}^l_{\rm DC}$ is, e.g., given by the static mean-field contribution of $\hat{H}^l_U$. There is a drawback in this procedure, however. By splitting electrons into correlated and uncorrelated we implicitly assume that the main effect of the latter is the renormalization or *screening* of parameters for the former, in particular of the Coulomb interaction. The computation of screening effects remains, unfortunately, a challenge. The calculation of exact screening would require the solution of the original many-body problem, taking all degrees of freedom into account, an impossible task. Commonly-used approximate schemes are the constrained LDA approximation (cLDA) and the constrained random-phase approximation (RPA) [5–7]. Both methods give reasonable estimates of screened Coulomb parameters for DMFT calculations. Typically cRPA calculations include more screening channels and are performed for less localized bases than cLDA calculations; thus cRPA parameters turn out to be often smaller than cLDA ones. To some extent, the difference can be taken as an estimate of the error bar.

After we have selected the electrons for which we think it is necessary to include explicitly the Hubbard correction, we have to build the final Hamiltonian for DMFT calculations. To this end, it is often convenient to integrate out or *downfold*, in part or completely, the weakly correlated states. There are different degrees of downfolding. The two opposite extreme limits are (i) *no downfolding*, i.e., keep explicitly in the Hamiltonian all weakly-correlated states (ii) *massive downfolding*, i.e., downfold all weakly correlated states. If we perform massive downfolding, e.g., downfold to the d (or e_g or t_{2g}) bands at the Fermi level, the Hamiltonian relevant for DMFT takes a simpler form. The LDA part is limited to the selected orbitals or bands, which, in the ideal case, are decoupled from the rest

$$\hat{H}^{\rm LDA} = -\sum_{\sigma} \sum_{ii'} \sum_{m_\alpha m'_\alpha} t^{i,i'}_{m_\alpha, m'_\alpha}\, c^\dagger_{im_\alpha\sigma}\, c_{i'm'_\alpha\sigma}.$$

The local *screened* Coulomb interaction for this set of orbitals is the on-site tensor

$$\hat{H}^l_U = \frac{1}{2} \sum_i \sum_{\sigma\sigma'} \sum_{m_\alpha m'_\alpha} \sum_{m_\beta m'_\beta} U_{m_\alpha m_\beta m'_\alpha m'_\beta}\, c^\dagger_{im_\alpha\sigma} c^\dagger_{im_\beta\sigma'} c_{im'_\beta\sigma'} c_{im'_\alpha\sigma}.$$

It is important to point out that the level of downfolding does not modify the hardness of the quantum-impurity problem. If, for example, in studying a transition-metal oxide, we plan to treat only $3d$ bands as correlated, it does not matter if we perform calculations with a Hamiltonian containing also, e.g., O p states, or we rather downfold all states but the $3d$ and work with a set of Wannier basis spanning the $3d$-like bands only. The number of correlated orbitals in the quantum-impurity problem is the same.[8]

[8] The choice might influence how severe the QMC sign problem is, however.

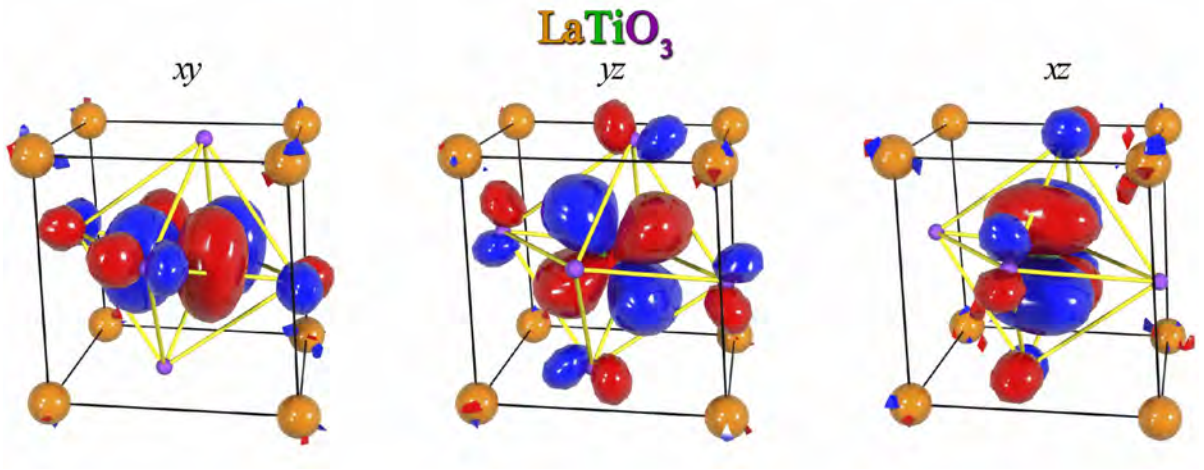


Fig. 9: *NMTO Wannier-like orbitals for t_{2g} states in $LaTiO_3$ obtained via massive downfolding to the t_{2g} bands. The t_{2g}-like orbitals have O p tails at the neighboring O sites reflecting the distortions of the lattice. The figure has been taken from Ref. [12].*

One advantage of massive downfolding is that the double-counting correction typically becomes a shift of the chemical potential, and it is therefore not necessary to calculate it explicitly. A second important advantage is that the interpretation of the final results is simpler. Instead, a disadvantage is that the basis functions are less localized, and therefore the approximation of the Coulomb interaction to a local operator might be less justified, and in some cases it might be necessary to include non-local Coulomb terms. The effect of downfolding on the localization of Wannier functions is illustrated for example in Fig. 9. Finally, another disadvantage of massive downfolding is that the energy window in which the model is valid is more narrow.

All advantages and disadvantages considered, what is then the best way of performing DMFT calculations? There is no universal answer to this question; it depends on the problem we are trying to solve and the system we are studying. Independently of the degree of downfolding we choose, it is important to point out that a clear advantage of Wannier functions in general is that they carry information about the lattice, bonding, chemistry, and distortions. This can be seen once more in Fig. 9, where orbitals are tilted and deformed by the actual structure and chemistry of the compound. Indeed, one might naively think of using a "universal" basis, for example atomic functions, the same for all systems, and thus calculating the hopping integrals using simply the electron-nuclear interaction $v_{en}(\boldsymbol{r})$. Besides the complications arising from the lack of orthogonality, such a basis has no built-in materials-specific information, except lattice positions. It is therefore a worse starting point for describing the electronic structure, even in the absence of correlations: larger basis sets are required to reach the same accuracy. From the point of view of LDA+DMFT, an advantage of an universal basis would be that it is free from double-counting corrections; on the other hand, however, exactly because we do not use the LDA potential and LDA orbitals to calculate the hopping integrals, we also cannot count on the successes of LDA in the description of average and long-range Coulomb effects. The hopping integrals would not even include the long-range Hartree term. For these

reasons *ab-initio* Wannier functions remain so far the basis of choice. They can be built via the Nth-Order Muffin-Tin Orbital (NMTO) method [12], the maximal-localization scheme [13], or projectors. Fig. 9 shows examples of NMTO-based Wannier functions. No matter what construction procedure is used, a common characteristic of *ab-initio* Wannier functions is that they are site-centered and localized.[9] A question naturally arises: How crucial is it to use localized functions as one-electron basis? This is an important point, since we have seen that strong-correlation effects arise in systems in which the on-site Coulomb interaction is much larger than longer-range terms. Let us consider therefore two opposite extreme limits. The first is the case in which the basis functions are independent of the lattice position (i.e., they are totally delocalized). For such a basis choice the Coulomb interaction parameters would be the same for every pair of lattice sites, no matter how distant. Thus a Hubbard-like model would be hard to justify. In the second extreme case, we adopt a hypothetical basis so localized that $\psi_{im\sigma}(\boldsymbol{r})\overline{\psi_{i'm'\sigma'}}(\boldsymbol{r}) \sim \delta_{i,i'}\,\delta(\boldsymbol{r}-\boldsymbol{T}_i)$. Even for such a basis choice, the unscreened Coulomb interaction is *not* local, but given by

$$U^{iji'j'}_{mp\,m'p'} \propto \frac{\delta_{i,i'}\delta_{j,j'}}{|\boldsymbol{T}_i-\boldsymbol{T}_j|},$$

hence it decays slowly with distance, although the (divergent) on-site term dominates. More generally, we can conclude that by increasing the localization of the basis we enhance the importance of the on-site Coulomb repulsion with respect to long-range terms; this better justifies Hubbard-like models—although we have to remember that most of the long-range part of the Coulomb interaction is in any case subtracted via the double-counting correction $\hat{H}_{\rm DC}$. The extreme case of the $\delta(\boldsymbol{r}-\boldsymbol{T}_i)$ functions also illustrates, however, how far we can go. A major problem with the extremely localized basis discussed above is that it would make it impossible to properly describe bonding, since the hopping integrals would be zero. Although such a basis is, of course, never used to build many-body models, there is a tempting approximation that has similar flaws. If one uses DFT-based electronic-structure techniques that tile the space in interstitial and non-overlapping atomic spheres (e.g., the LAPW method), it is tempting to use as basis for correlated electrons the atomic functions *defined inside the atomic spheres*. These functions are, by construction, much more localized than Wannier orbitals (even when no downfolding is performed in the Wannier construction). However, they *do not form a complete basis set* in the space of square-integrable functions. This is obvious because such a basis does not even span the LDA bands; to reproduce the bands we need, in addition, functions defined in the interstitial region. This is illustrated in Fig. 10 for a simple example of two quantum well potentials.[10] We therefore cannot use it to write the many-body Hamiltonian in the usual form, $\hat{H}_0 + \hat{H}_U$. In conclusion, a basis which, as *ab-initio* Wannier functions, is complete and indeed spans the bands, is better justified, although we somewhat lose in localization.

[9] Differences in localizations between the various construction procedures are actually small for the purpose of many-body calculations, provided that the same bands are spanned in the same way.

[10] Another, but less severe, problem of atomic sphere truncations is that the results will depend on the sphere size, in particular when atomic spheres are small.

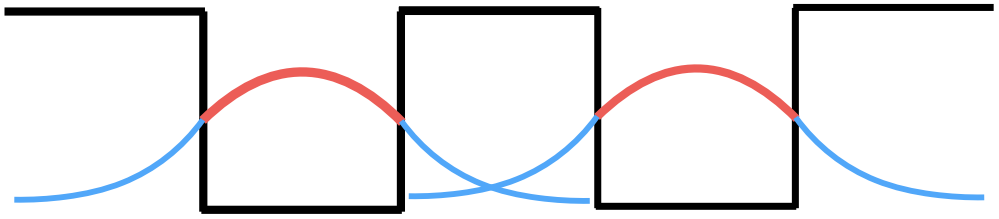


Fig. 10: *The problem of two quantum wells. The figure shows (schematically) for each well the wavefunction of a bound state. If we consider only the part of the wavefunction inside its own well (red in the figure), the differential overlap (and hence the hopping integral) between functions centered on different wells would be zero.*

3 Linear response functions

Linear response functions are key to understanding many experimental results. In this section we explain how to calculate them within the LDA+DMFT approach. First we introduce the generalized susceptibility, which yields the linear response to a given external perturbation. Next we present the method used to calculate it and discuss the approximations adopted. Last we analyze in detail the case of the magnetic susceptibility for the one-band Hubbard model.

3.1 Definitions

Let us start by introducing the site susceptibility in imaginary time. This is given by

$$\chi_{\hat{P}^i_\nu\,\hat{O}^{i'}_{\nu'}}(\boldsymbol{\tau}) = \left\langle \mathcal{T} \Delta\hat{P}^i_\nu(\tau_1,\tau_2)\Delta\hat{O}^{i'}_{\nu'}(\tau_3,\tau_4)\right\rangle_0, \tag{79}$$

where $\boldsymbol{\tau} = (\tau_1,\tau_2,\tau_3,\tau_4)$ and $\mathcal{T}$ is the time-ordering operator. The site-dependent operators are defined via the equations

$$\hat{P}^i_\nu(\tau_1,\tau_2) = \sum_{\boldsymbol{\alpha}} p^\nu_{\boldsymbol{\alpha}}\; c^\dagger_{i\alpha'}(\tau_2)c_{i\alpha}(\tau_1), \qquad \Delta\hat{P}^i_\nu(\tau_1,\tau_2) = \hat{P}^i_\nu(\tau_1,\tau_2) - \left\langle \hat{P}^i_\nu(\tau_1,\tau_2)\right\rangle$$
$$\hat{O}^{i'}_{\nu'}(\tau_3,\tau_4) = \sum_{\boldsymbol{\gamma}} o^{\nu'}_{\boldsymbol{\gamma}}\; c^\dagger_{i'\gamma'}(\tau_4)c_{i'\gamma}(\tau_3), \qquad \Delta\hat{O}^{i'}_{\nu'}(\tau_3,\tau_4) = \hat{O}^{i'}_{\nu'}(\tau_3,\tau_4) - \left\langle \hat{O}^{i'}_{\nu'}(\tau_3,\tau_4)\right\rangle.$$

The labels $\boldsymbol{\alpha} = (\alpha,\alpha')$, $\boldsymbol{\gamma} = (\gamma,\gamma')$ are collective flavors. For the multi-band Hubbard model they may include spin (σ) and orbital (m) quantum number, plus a fractional lattice vector identifying a correlated basis atom in the unit cell (i_c). The weight factors $o^\nu_{\boldsymbol{\alpha}}$ and $p^{\nu'}_{\boldsymbol{\gamma}}$, in general complex numbers, identify the type of response. We can then rewrite Eq. (79) as

$$\chi_{\hat{P}^i_\nu\,\hat{O}^{i'}_{\nu'}}(\boldsymbol{\tau}) = \sum_{\boldsymbol{\alpha\gamma}} p^\nu_{\boldsymbol{\alpha}} o^{\nu'}_{\boldsymbol{\gamma}}\; \chi_{i\boldsymbol{\alpha},i'\boldsymbol{\gamma}}(\boldsymbol{\tau}),$$

with

$$\chi_{i\boldsymbol{\alpha},i'\boldsymbol{\gamma}}(\boldsymbol{\tau}) = \left\langle \mathcal{T} c_{i\alpha}(\tau_1)c^\dagger_{i\alpha'}(\tau_2)c_{i'\gamma}(\tau_3)c^\dagger_{i'\gamma'}(\tau_4)\right\rangle - G_{i\alpha,i\alpha'}(\tau_1,\tau_2)G_{i'\gamma,i'\gamma'}(\tau_3,\tau_4). \tag{80}$$

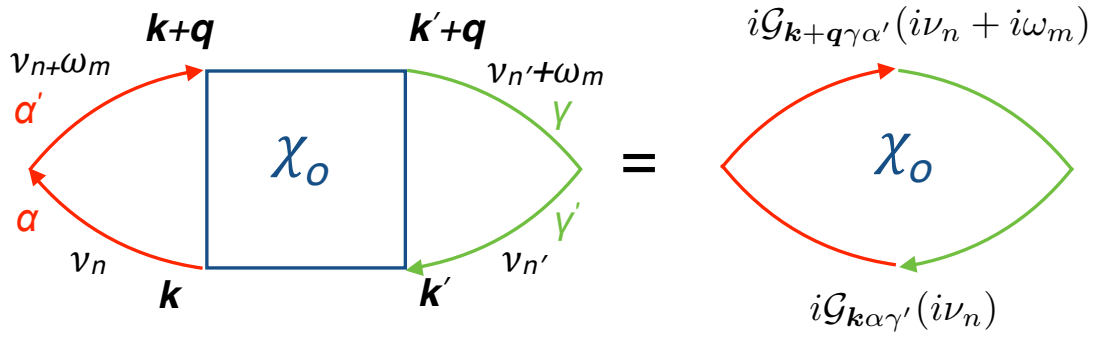


Fig. 11: *Diagram contributing to the linear susceptibility for a non-interacting system. The red lines indicates that the creator/annihilator is originally from the operator $\hat{P}_{\nu'}$ and the green lines indicate that the creator/annihilator is from the operator $\hat{O}_{\nu}$. The corresponding frequencies and momenta are explicitly assigned.*

Performing the Fourier transform from imaginary time to Matsubara frequencies we obtain

$$\chi_{i\boldsymbol{\alpha},i'\boldsymbol{\gamma}}(\boldsymbol{\nu}) = \frac{1}{16}\iiint d\boldsymbol{\tau}\; e^{i\boldsymbol{\nu}\cdot\boldsymbol{\tau}}\chi_{i\boldsymbol{\alpha},i'\boldsymbol{\gamma}}(\boldsymbol{\tau}), \tag{81}$$

where $\boldsymbol{\nu} = (\nu_1, -\nu_2, \nu_3, -\nu_4)$. Due to the conservation of energy, only three of the four ν_i frequencies are independent. Hence, for convenience we set $\nu_1 = \nu_n$, $\nu_2 = \nu_n + \omega_m$, $\nu_3 = \nu_{n'} + \omega_m$, and $\nu_4 = \nu_{n'}$. Next we perform the Fourier transform from site to momentum space. Due to the conservation of lattice momentum, only three of the four $\boldsymbol{k}_i$-vectors are independent. After redefining $\boldsymbol{k}_1 = \boldsymbol{k}$, $\boldsymbol{k}_2 = \boldsymbol{k}+\boldsymbol{q}$, $\boldsymbol{k}_3 = \boldsymbol{k}'+\boldsymbol{q}$ and $\boldsymbol{k}_4 = \boldsymbol{k}'$, we find the expression

$$\chi_{\hat{P}_\nu\,\hat{O}_{\nu'}}(\boldsymbol{q};\boldsymbol{\nu}) = \sum_{\boldsymbol{\alpha}\boldsymbol{\gamma}} p^{\nu}_{\boldsymbol{\alpha}} o^{\nu'}_{\boldsymbol{\gamma}} \sum_{ii'} e^{i(\boldsymbol{T}_i - \boldsymbol{T}_{i'})\cdot\boldsymbol{q}}\chi_{i\boldsymbol{\alpha},i'\boldsymbol{\gamma}}(\boldsymbol{\nu}) = \sum_{\boldsymbol{\alpha}\boldsymbol{\gamma}} p^{\nu}_{\boldsymbol{\alpha}} o^{\nu'}_{\boldsymbol{\gamma}} \underbrace{\frac{1}{N_{\boldsymbol{k}}^2}\sum_{\boldsymbol{k}\boldsymbol{k}'}[\chi(\boldsymbol{q};i\omega_m)]_{\boldsymbol{k}\nu_n\boldsymbol{\alpha},\boldsymbol{k}'\nu_{n'}\boldsymbol{\gamma}}}_{[\chi(\boldsymbol{q};\omega_m)]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}}}.$$

In this expression, by summing over $\boldsymbol{k}$ and $\boldsymbol{k}'$ we obtained $[\chi(\boldsymbol{q};\omega_m)]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}}$. The physical linear response function is given by the sum over the fermionic Matsubara frequencies

$$\chi_{\hat{P}_\nu\,\hat{O}_{\nu'}}(\boldsymbol{q};i\omega_m) = \sum_{\boldsymbol{\alpha}\boldsymbol{\gamma}} p^{\nu}_{\boldsymbol{\alpha}} o^{\nu'}_{\boldsymbol{\gamma}} \underbrace{\frac{1}{\beta^2}\sum_{nn'}\Big[\chi(\boldsymbol{q};\omega_m)\Big]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}}}_{[\chi(\boldsymbol{q};\omega_m)]_{\boldsymbol{\alpha},\boldsymbol{\gamma}}}. \tag{82}$$

In this lecture we will consider the example of the magnetic susceptibility. In this case the operators $\hat{P}^i_\nu$ and $\hat{O}^{i'}_{\nu'}$ are the three components of the magnetization operator. In the single-orbital limit ($\alpha = \alpha' = \sigma$ and $\gamma = \gamma' = \sigma'$), we thus have, e.g.,

$$o^z_{\boldsymbol{\alpha}} = -g\mu_B\langle\sigma|\hat{\sigma}_z|\sigma\rangle, \qquad p^z_{\boldsymbol{\alpha}} = -g\mu_B\langle\sigma'|\hat{\sigma}_z|\sigma'\rangle.$$

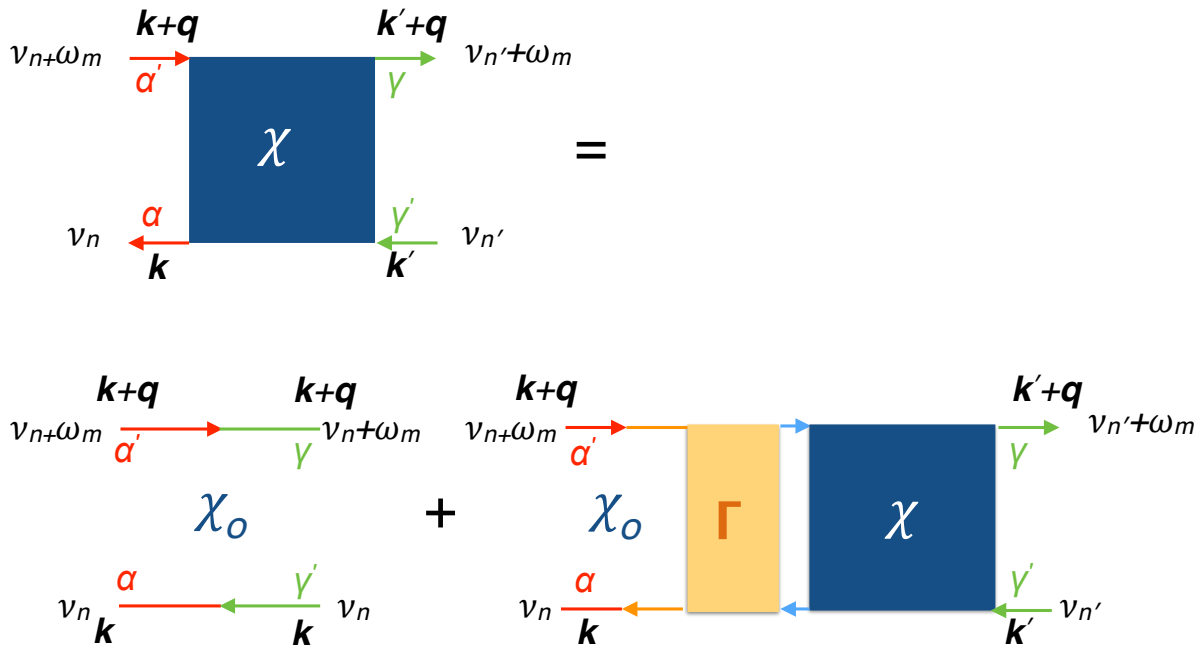


Fig. 12: *Diagrammatic representation of the Bethe-Salpeter equation for the linear susceptibility. The red lines indicate a creator/annihilator stemming from the operator $\hat{P}_\nu$ and the green lines from the operator $\hat{O}_{\nu'}$. The box labeled with Γ is the vertex function, the one labeled with χ the full susceptibility, and χ_0 is the pair-bubble term.*

3.2 DMFT and the Bethe-Salpeter equation

The linear response functions for the multi-band Hubbard model $\hat{H}_0 + \hat{H}_U$ can be in principle obtained via standard many-body perturbation theory, i.e., using the Coulomb interaction U as the expansion parameter. In this approach the expansion point is the linear response function for the non-interacting Hamiltonian $\hat{H}_0$. This term, due to Wick's theorem, can be written as follows

$$\Big[\chi_0(\boldsymbol{q};i\omega_m)\Big]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}} = \frac{1}{N_{\boldsymbol{k}}^2}\sum_{\boldsymbol{k}\boldsymbol{k}'}\underbrace{\Big[-\beta N_{\boldsymbol{k}}\mathcal{G}_{\boldsymbol{k}\alpha\gamma'}(i\nu_n)\,\mathcal{G}_{\boldsymbol{k}'+\boldsymbol{q}\alpha'\gamma}(i\nu_{n'}+i\omega_m)\,\delta_{n,n'}\delta_{\boldsymbol{k},\boldsymbol{k}'}\Big]}_{\Big[\chi_0(\boldsymbol{q};i\omega_m)\Big]_{\boldsymbol{k}\nu_n\boldsymbol{\alpha},\boldsymbol{k}'\nu_{n'}\boldsymbol{\gamma}}}. \quad (83)$$

The associated Feynman diagram is shown in Fig. 11, and we will refer to it as the bubble term. Once we switch on the Coulomb interaction, many-body perturbation theory leads to the Bethe-Salpeter equation, pictorially shown in Fig. 12. The susceptibility can then be expressed via the relation

$$\Big[\chi(\boldsymbol{q};i\omega_m)\Big]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}} = \frac{1}{N_{\boldsymbol{k}}^2}\sum_{\boldsymbol{k}\boldsymbol{k}'}\Big[\chi_0(\boldsymbol{q};i\omega_m) + \frac{1}{N_{\boldsymbol{k}}^2}\chi_0(\boldsymbol{q};i\omega_m)\Gamma(\boldsymbol{q};i\omega_m)\chi(\boldsymbol{q};i\omega_m)\Big]_{\boldsymbol{k}\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{k}'\boldsymbol{\gamma}}.$$

Replacing recursively $\chi(\boldsymbol{q};i\omega_m)$ one obtains an infinite series in the vertex $\Gamma(\boldsymbol{q};i\omega_m)$, which plays the role of the self-energy in the Dyson equation.

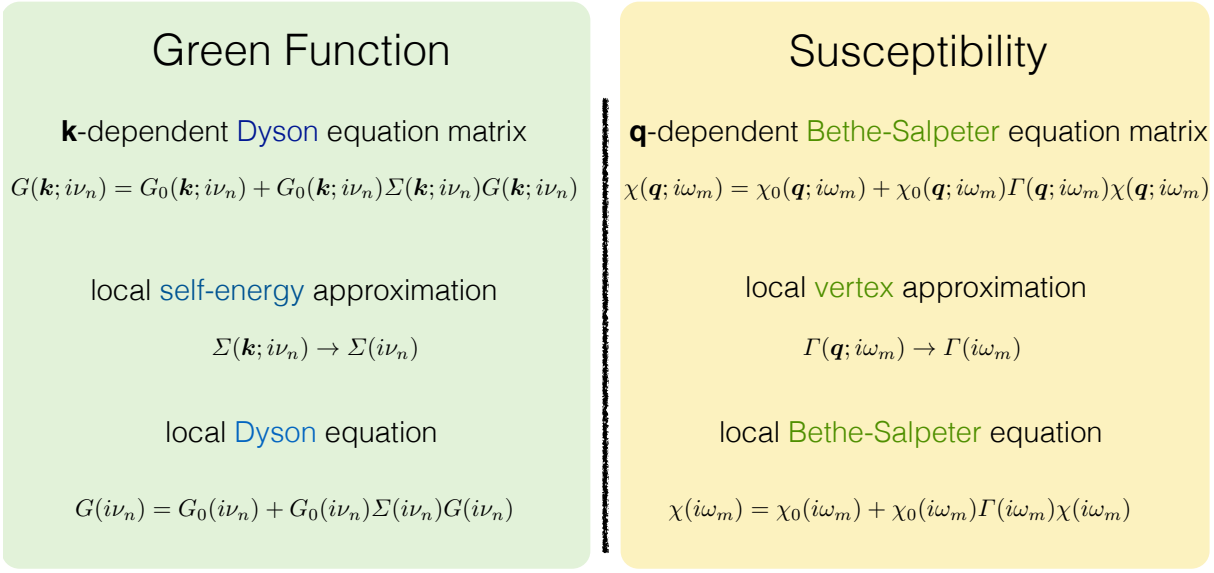


Fig. 13: *Analogies between the calculation of the Green function $G(\boldsymbol{k};i\nu_n)$ in the local-self-energy approximation (left) and the calculation of the response function $\chi(\boldsymbol{q};i\omega_m)$ in the local vertex approximation (right). Each term in the general Bethe-Salpeter equation can be viewed as a square matrix of dimension $N_{\boldsymbol{k}}N_nN_\alpha$, where $N_{\boldsymbol{k}}$ is the number of $\boldsymbol{k}$ points, N_n the number of fermionic Matsubara frequencies, N_α the number of flavors.*

Instead of expanding around the non-interacting limit, for correlated systems it is more convenient to construct a diagrammatic series starting from the bubble term calculated using the DMFT Green functions, i.e., replacing $\mathcal{G} \longrightarrow G$ in Eq. (83). The result is

$$[\chi_0(\boldsymbol{q};i\omega_m)]_{\nu_n\boldsymbol{\alpha},\nu_{n'}\boldsymbol{\gamma}} = -\beta\delta_{nn'}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}} G_{\alpha\gamma'}(\boldsymbol{k};i\nu_n)\, G_{\alpha'\gamma}(\boldsymbol{k}+\boldsymbol{q};i\nu_n+i\omega_m). \tag{84}$$

If we follow this approach, the unknown vertex $\Gamma(\boldsymbol{q};i\omega_m)$ in the Bethe-Salpeter equation differs from the one obtained in standard many-body perturbation theory. In order to calculate it we adopt two approximations. The first is that the vertex is approximately local. In the infinite dimension limit it has been shown that the vertex can be replaced by a local [4, 14] quantity, $\Gamma(i\omega_m)$. Assuming that, in the spirit of the dynamical mean-field approximation, for a real 3-dimensional system we can do the same, we thus set

$$\Gamma(\boldsymbol{q};i\omega_m) \longrightarrow \Gamma(i\omega_m). \tag{85}$$

Thus, dropping for simplicity the flavor indices, after performing the $\boldsymbol{k}$ sums, the Bethe-Salpeter equation becomes

$$\chi(\boldsymbol{q};i\omega_m) = \chi_0(\boldsymbol{q};i\omega_m) + \chi_0(\boldsymbol{q};i\omega_m)\Gamma(i\omega_m)\chi(\boldsymbol{q};i\omega_m). \tag{86}$$

By solving it we find, formally

$$\chi^{-1}(\boldsymbol{q};i\omega_m) = \chi_0^{-1}(\boldsymbol{q};i\omega_m) - \Gamma(i\omega_m). \tag{87}$$

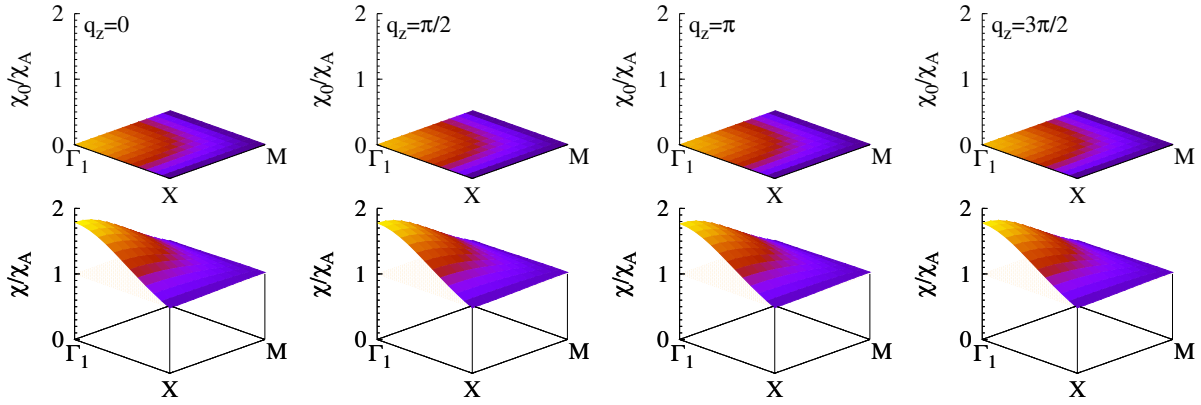


Fig. 14: *$VOMoO_4$: Static magnetic susceptibility $\chi(\boldsymbol{q};0)/\chi_A(0)$ in the q_x, q_y plane for representative q_z values, $T \sim 380$ K and $U = 5$ eV. The normalization $\chi_A(0) \sim \mu^2_{\rm eff}/k_BT$ is the atomic susceptibility in the large βU limit. Top panels: $\Gamma = 0$. Bottom panels: $\Gamma \neq 0$. Special points: $\Gamma_1 = (2\pi, 0)$, X$= (\pi, 0)$ and M$= (\pi, \pi)$. Rearranged from Ref. [11].*

Next we assume that $\Gamma(i\omega_m)$ equals the vertex for the quantum impurity. We define $\chi(i\omega_m)$ the impurity susceptibility obtained via the quantum-impurity solver in the final iteration of the DMFT self-consistency loop, and $\chi_0(i\omega_m)$ the average of Eq. (84) over momenta

$$\chi_0(i\omega_m) = \frac{1}{N_{\boldsymbol{q}}} \sum_{\boldsymbol{q}} \chi_0(\boldsymbol{q}; i\omega_m). \tag{88}$$

The local vertex $\Gamma(i\omega_m)$ is then obtained by solving a local Bethe-Salpeter equation

$$\Gamma(i\omega_m) = \chi_0^{-1}(i\omega_m) - \chi^{-1}(i\omega_m). \tag{89}$$

Replacing $\Gamma(i\omega_m)$ into Eq. (87) finally yields the $\boldsymbol{q}$-dependent susceptibility. It has to be noticed that, although the two equations (87) and (89) look innocent, solving them numerically is a delicate task because the local susceptibility is in general not diagonal in n, n' and does not decay very fast with the frequencies. There are, however, various ways to reduce the computational costs, e.g., via extrapolations [11] or using compact representations based on auxiliary polynomials [15, 16]. The method just illustrated for the calculation of linear response functions in the local vertex approximation bears resemblance with the approach adopted for the calculation of the Green functions in the local self-energy approximation. These analogies are schematically pointed out in Fig. 13. Instead, in Fig. 14 we show as an example the case of the static magnetic susceptibility for a one-band system, the S=1/2 frustrated Mott insulator $VOMoO_4$. The figure shows both the bubble term $\chi_0(\boldsymbol{q}; i\omega_m)$ (top panels) and the full susceptibility $\chi(\boldsymbol{q}; i\omega_m)$ (bottom panels). The two differ sizably in absolute value. In addition, as we will discuss later, the $\chi_0(\boldsymbol{q}; i\omega_m)$ term alone is very weakly dependent on the temperature. The expected Curie-Weiss-like behavior is only recovered when $\Gamma(i\omega_m)$ is taken into account.

3.3 The local susceptibility: Legendre representation

The core of the approach described in the previous section is the calculation of the local susceptibility tensor, $\chi_{\boldsymbol{\alpha\gamma}}(\boldsymbol{\tau})$. In DMFT all local observables $\langle\hat{O}\rangle$ are obtained via the quantum-impurity solver, for example the continuous-time hybridization expansion QMC technique presented in Section 2.3. Susceptibilities, however, require sizably longer computational time than Green-function matrices. Thus, instead of calculating directly $\chi_{\boldsymbol{\alpha\gamma}}(\boldsymbol{\tau})$, it is convenient to express the tensor elements in a basis of orthogonal functions $f_l^m(\tau)$, chosen such that the representation is as compact as possible. A successful choice [15, 16] is

$$f_l^m(\tau) = e^{-i\varphi_m(\tau)} \begin{cases} \sqrt{2l+1}\; p_l(x(\tau)), & \tau > 0 \\ -(-1)^m\sqrt{2l+1}\; p_l(x(\tau+\beta)), & \tau < 0 \end{cases}$$

where $p_l(x(\tau))$ is a Legendre polynomial of degree l, with $x(\tau) = 2\tau/\beta - 1$; here the factor $(-1)^m$ in the second row ensures anti-periodicity for all values of m, which is the index for the bosonic Matsubara frequency ω_m. Via the orthogonality properties of the polynomials we obtain

$$\chi_{\boldsymbol{\alpha\gamma}}(i\omega_m) = \frac{1}{\beta^2}\sum_{ll'} f_l^{-m}(0^+)\; \chi^{l,l'}_{\boldsymbol{\alpha\gamma}}(i\omega_m)\; f_{l'}^{-m}(0^+). \tag{90}$$

The expansion coefficients in Eq. (90) take the form

$$\chi^{l,l'}_{\boldsymbol{\alpha\gamma}}(i\omega_m) = \int_0^\beta d\tau_{23}\int_0^\beta d\tau_{12}\int_0^\beta d\tau_{34}\; e^{-i\omega_m\tau_{23}} f_l^m(\tau_{12})\chi^{\alpha\alpha'}_{\gamma\gamma'}(\tau_{14},\tau_{24},\tau_{34},0) f_{l'}^m(\tau_{34}), \tag{91}$$

where $\tau_{ij} = \tau_i - \tau_j$, with $\tau_{14} = \tau_{12}+\tau_{23}+\tau_{34}$, and $\tau_{24} = \tau_{23}+\tau_{34}$. The phase defining the gauge is $\varphi_m(\tau) = \omega_m\tau/2$ and does not depend on l. As we have seen, in quantum Monte Carlo the observables are obtained as the average over the visited configurations c. Splitting (91) into two terms [16] we have

$$\left\langle \chi^{l,l'}_{\boldsymbol{\alpha\gamma}}(i\omega_m)\right\rangle_c = \left\langle \mathcal{C}^{l,l'}_{\boldsymbol{\alpha\gamma}}(i\omega_m)\right\rangle_c - \beta\delta_{m,0}\left\langle G^l_{\boldsymbol{\alpha}}\right\rangle_c\left\langle G^{l'}_{\boldsymbol{\gamma}}\right\rangle_c.$$

The first term can be expressed as

$$\left\langle C^{l,l'}_{\boldsymbol{\alpha\gamma}}(i\omega_m)\right\rangle_c = \frac{1}{\beta}\sum_{bb'dd'}^{N_B}\sum_{i,j}^{k_b,k_d}\sum_{i',j'}^{k_{b'},k_{d'}} f_l^m(\tau_{dj}-\bar\tau_{bi}) f_{l'}^m(\tau_{d'j'}-\bar\tau_{b'i'}) c^{db,d'b'}_{ji,j'i'}(i\omega_m)\delta_{\boldsymbol{\alpha},(\alpha_{dj},\bar\alpha_{bi})}\delta_{\boldsymbol{\gamma},(\alpha_{d'j'},\bar\alpha_{b'i'})}$$

where

$$c^{db,d'b'}_{ji,j'i'}(i\omega_m) = \left(w^{db}_{ji}w^{d'b'}_{j'i'} - w^{d'b}_{j'i}w^{db'}_{ji'}\right) e^{-i\omega_m(\bar\tau_{bi}-\tau_{d'j'})}.$$

Here the imaginary times τ_{bi} and $\bar\tau_{bi}$ all vary in the interval $[0,\beta)$. The letters b and d label the N_B flavors decoupled by symmetry, e.g., $\{\uparrow,\downarrow\}$. Finally, $w^{db}_{ji} = \delta_{b,d}\mathcal{M}^{k_b}_{bj,bi}$, where the matrix $\mathcal{M}^{k_b} = [\mathcal{F}_0^{k_b}]^{-1}$ is the inverse of the hybridization function matrix $\mathcal{F}_0^{k_b}$ for expansion order k_b. The Green functions in the second term are instead given by

$$\left\langle G^l_{\boldsymbol{\alpha}}\right\rangle_c = -\frac{1}{\beta}\sum_{b}^{N_B}\sum_{ij}^{k_b} f_l^0(\tau_{bj}-\bar\tau_{bi})\; w^{bb}_{ji}\;\delta_{\boldsymbol{\alpha},(\alpha_{bj},\bar\alpha_{bi})}.$$

3.4 Magnetic susceptibility for the single-band Hubbard model

The magnetic susceptibility is the linear response to an external magnetic field applied along a direction, here defined as $\hat{z}$. The associated site susceptibility is

$$\chi^{i,i'}_{zz}(\boldsymbol{\tau}) = \left\langle \mathcal{T}\hat{M}^i_z(\tau)\hat{M}^{i'}_z(0)\right\rangle_0 - \left\langle \hat{M}^i_z\right\rangle_0\left\langle \hat{M}^{i'}_z\right\rangle_0,$$

where $\hat{M}^i_z = -g\mu_B\hat{S}^i_z$ is the magnetization for lattice site i. Its Fourier transform is

$$\begin{aligned}\chi_{zz}(\boldsymbol{q}; i\omega_m) &= \sum_{ii'} e^{i\boldsymbol{q}\cdot(\boldsymbol{T}_i-\boldsymbol{T}_{i'})}\int d\tau\; e^{i\omega_m\tau}\chi^{i,i'}_{zz}(\tau)\\ &= \left\langle \hat{M}_z(\boldsymbol{q};\omega_m)\hat{M}_z(-\boldsymbol{q};0)\right\rangle_0 - \left\langle \hat{M}_z(\boldsymbol{q})\right\rangle_0\left\langle \hat{M}_z(-\boldsymbol{q})\right\rangle_0, \end{aligned} \tag{92}$$

where ω_m is a bosonic Matsubara frequency. In this section we will discuss the example of the single-band Hubbard model on a hypercubic lattice and at half filling. Furthermore, unless differently specified, we will adopt the tight-binding dispersion

$$\varepsilon_{\boldsymbol{k}} = -2t\sum_{n=1}^{d}\cos k_d a, \tag{93}$$

where $d = 1, 2, 3$ is the dimension and a the length of the unit vectors defining the lattice. For a single-band model the magnetization operator can be expressed in the Bloch basis as

$$\hat{M}_z(\boldsymbol{q}) = -\frac{g\mu_B}{2}\sum_{\boldsymbol{k}}\sum_{\sigma} s_\sigma c^\dagger_{\boldsymbol{k}+\boldsymbol{q}\sigma}c_{\boldsymbol{k}\sigma}, \tag{94}$$

where $s_\sigma = 1$ for $\sigma =\uparrow$ and $s_\sigma = -1$ for $\sigma =\downarrow$. To obtain the magnetic response function we thus have to calculate the imaginary-time tensor with elements

$$\begin{aligned}\left[\chi(\boldsymbol{q};\boldsymbol{\tau})\right]_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} &= \left\langle \mathcal{T}c_{\boldsymbol{k}\sigma}(\tau_1)c^\dagger_{\boldsymbol{k}+\boldsymbol{q}\sigma}(\tau_2)c_{\boldsymbol{k}'+\boldsymbol{q}\sigma'}(\tau_3)c^\dagger_{\boldsymbol{k}'\sigma'}(\tau_4)\right\rangle_0\\ &\quad - \left\langle \mathcal{T}c_{\boldsymbol{k}\sigma}(\tau_1)c^\dagger_{\boldsymbol{k}+\boldsymbol{q}\sigma}(\tau_2)\right\rangle_0\left\langle \mathcal{T}c_{\boldsymbol{k}'+\boldsymbol{q}\sigma'}(\tau_3)c^\dagger_{\boldsymbol{k}'\sigma'}(\tau_4)\right\rangle_0. \end{aligned} \tag{95}$$

The associated imaginary-time magnetic susceptibility is then given by

$$\chi_{zz}(\boldsymbol{q};\boldsymbol{\tau}) = (g\mu_B)^2\frac{1}{4}\sum_{\sigma\sigma'} s_\sigma s_{\sigma'}\underbrace{\frac{1}{\beta}\frac{1}{N^2_{\boldsymbol{k}}}\sum_{\boldsymbol{k}\boldsymbol{k}'}\left[\chi(\boldsymbol{q};\boldsymbol{\tau})\right]_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}}_{\chi_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q};\boldsymbol{\tau})}. \tag{96}$$

After we Fourier transform with respect to imaginary time and sum over the fermionic Matsubara frequencies, we have

$$\chi_{zz}(\boldsymbol{q}; i\omega_m) = (g\mu_B)^2\frac{1}{4}\sum_{\sigma\sigma'} s_\sigma s_{\sigma'}\frac{1}{\beta^2}\sum_{nn'}\chi^{n,n'}_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q}; i\omega_m), \tag{97}$$

where

$$\chi^{n,n'}_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q}; i\omega_m) = \frac{1}{16}\iiint d\boldsymbol{\tau}\; e^{i\boldsymbol{\nu}\cdot\boldsymbol{\tau}}\chi_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q};\boldsymbol{\tau}). \tag{98}$$

For $\omega_n = 0$ we obtain the static response function.

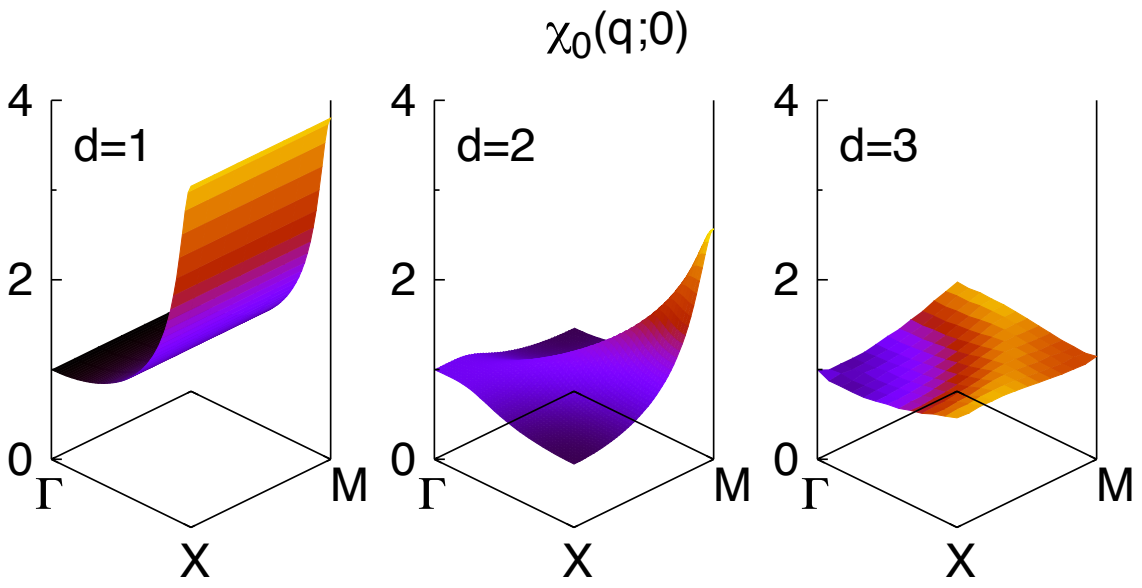


Fig. 15: *The ratio $\chi_0(\boldsymbol{q};0)/\chi_0(\boldsymbol{0};0)$ in the x-y plane for a hypercubic lattice with $t = 0.4$ eV ($T \sim 230$ K) at half filling. From left to right: one, two, and three dimensions.*

3.4.1 Non-interacting limit

In the non-interacting limit we can use Wick's theorem to simplify Eq. (95). It follows that the elements of the two-particle Green function tensor vanish if $\boldsymbol{k} \neq \boldsymbol{k}'$. Thus Eq. (96) becomes

$$\chi_{zz}(\boldsymbol{q};\boldsymbol{\tau}) = -(g\mu_B)^2 \frac{1}{4}\frac{1}{\beta}\frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{\sigma} \mathcal{G}_{\boldsymbol{k}\sigma}(\tau_{14}) \mathcal{G}_{\boldsymbol{k}+\boldsymbol{q}\sigma}(-\tau_{23}).$$

For the frequency-dependent magnetic susceptibility Eq. (97) we have instead

$$\chi_{zz}(\boldsymbol{q};i\omega_m) = (g\mu_B)^2 \frac{1}{4}\frac{1}{\beta^2} \sum_{nn'} \sum_{\sigma} \chi^{n,n'}_{\sigma\sigma\sigma\sigma}(\boldsymbol{q};i\omega_m),$$

where

$$\sum_{\sigma} \chi^{n,n'}_{\sigma\sigma\sigma\sigma}(\boldsymbol{q};i\omega_m) = -\frac{\beta}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{\sigma} \mathcal{G}_{\boldsymbol{k}\sigma}(i\nu_n)\, \mathcal{G}_{\boldsymbol{k}+\boldsymbol{q}\sigma}(i\nu_n + i\omega_m)\, \delta_{n,n'}. \tag{99}$$

The actual dynamical susceptibility is then given by

$$\chi_{zz}(\boldsymbol{q};i\omega_m) = -\,(g\mu_B)^2\, \frac{1}{4}\frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{\sigma} \frac{n_\sigma(\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}) - n_\sigma(\varepsilon_{\boldsymbol{k}})}{\varepsilon_{\boldsymbol{k}+\boldsymbol{q}} - \varepsilon_{\boldsymbol{k}} + i\omega_m},$$

where $n_\sigma(x) = n_F(x)$ is the Fermi distribution function. Figure 15 shows the static spin susceptibility for a d-dimensional hypercubic lattice. For $T \to 0$, it diverges at the antiferromagnetic vector $\boldsymbol{q}_C$, which in two dimension is the M point. Indeed, since $\varepsilon_{\boldsymbol{k}+\boldsymbol{q}_C} = -\varepsilon_{\boldsymbol{k}}$ (perfect nesting) we have

$$\chi_0(\boldsymbol{q}_C;0) \propto \frac{1}{4}\int_{-\infty}^{\varepsilon_F} d\varepsilon\, \frac{\rho(\varepsilon)}{\varepsilon}.$$

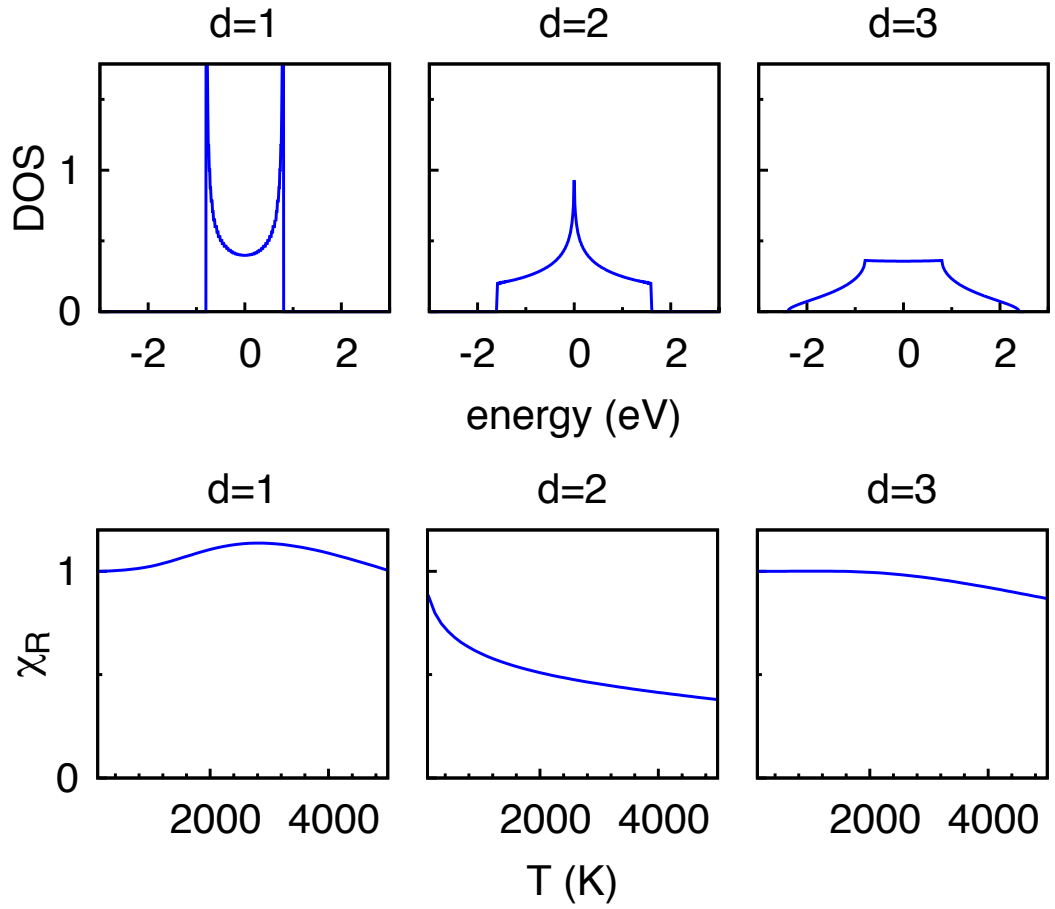


Fig. 16: *Effect of $\rho(\varepsilon_F)$ on the temperature dependence of $\chi_R = \chi_P(T)/\chi_P(0)$ for a hypercubic lattice with $t = 0.4$ eV and at half filling. Up to ~ 1000 K only the logarithmic Van-Hove singularity (two-dimensional case) yields a sizable effect.*

In the $\boldsymbol{q} \to 0$ and $T \to 0$ limit, setting $\omega_m = 0$ we recover the Pauli susceptibility

$$\chi_{zz}(\mathbf{0};0) = \frac{1}{4}\left(g\mu_B\right)^2 \rho(\varepsilon_F),$$

$$\rho(\varepsilon_F) = -\sum_\sigma \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \left.\frac{dn_\sigma(\varepsilon_{\boldsymbol{k}})}{d\varepsilon_{\boldsymbol{k}}}\right|_{T=0}.$$

Fig. 16 shows that, assuming the dispersion given in Eq. (93), the Pauli susceptibility is weakly temperature dependent in three dimensions, but not when a van-Hove singularity is close to the Fermi level, as it happens for the $d = 2$ case.

3.4.2 Small U/t limit: Hartree Fock approximation and Stoner model

In the small U/t limit one can expand around the non-interacting susceptibility and treat the effect of Coulomb repulsion in the static mean-field or Hartree-Fock approximation. We have previously seen (Section 2.4 and Fig. 6) that in the ferromagnetic case this means that the band for spin up and spin down electrons acquire different energy. The energy splitting equals

the one obtained in the presence of an effective magnetic field $h_{\rm eff} = 2mU$, where m is the magnetization. We can generalize this result to a magnetic structure characterized by a vector $\boldsymbol{q}$; in this case we have $h_{\rm eff}(\boldsymbol{q}) = 2m(\boldsymbol{q})U$, where $m(\boldsymbol{q})$ is the associated order parameter. The associated static magnetic response function is thus

$$\chi(\boldsymbol{q};0) = \frac{\chi_0(\boldsymbol{q};0)}{1 - \Gamma\chi_0(\boldsymbol{q};0)}, \tag{100}$$

where $\Gamma{=}2U$. This is a simplified version of the Bethe-Salpeter equation obtained in standard many-body perturbation theory, with, however, a first-order frequency- and momentum-independent vertex. In the case of a hypercubic lattice with dispersion (93) the susceptibility $\chi_0(\boldsymbol{q};0)$ is larger at the nesting vector; this favors instabilities towards antiferromagnetism.

3.4.3 Atomic limit

Let us now consider the opposite extreme, the atomic limit. First we adopt a simple approach. Since all atoms are decoupled, only on-site terms $i = i'$ contribute. We then can calculate the right-hand side of Eq. (92) by summing up the contributions of the four atomic states, $|0\rangle$, $c^\dagger_\uparrow|0\rangle$, $c^\dagger_\downarrow|0\rangle$, $c^\dagger_\uparrow c^\dagger_\downarrow|0\rangle$, obtaining at half filling

$$\chi_{zz}(\boldsymbol{q};i\omega_m) = (g\mu_B)^2 \, \frac{1}{4k_BT} \frac{e^{\beta U/2}}{1+e^{\beta U/2}} \, \delta_{\omega_m,0}. \tag{101}$$

The same expression can be derived following the general procedure outlined in the previous pages, i.e., starting from the two-particle Green function tensor $\chi_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q};\boldsymbol{\tau})$, defined in Eq. (96). In the atomic limit, it is convenient to work in real space, since

$$\chi_{\sigma\sigma\sigma'\sigma'}(\boldsymbol{q};\boldsymbol{\tau}) = \frac{1}{\beta} \sum_i \chi_{i\sigma\sigma,i\sigma'\sigma'}(\boldsymbol{\tau}).$$

Thanks to the symmetries of the tensor in imaginary time, it is sufficient to calculate $\chi_{i\sigma\sigma,i\sigma'\sigma'}(\boldsymbol{\tau})$ for positive times $0 < \tau_{j4} < \beta$, where $\tau_{j4} = \tau_j - \tau_4$ with $j = 1,2,3$. Due to the time ordering operator we have, however, to consider separately six different imaginary-time sectors. In the Appendix one can find a list of all these sectors and their contributions. For simplicity, we discuss here explicitly only the case $\tau_{14} > \tau_{24} > \tau_{34} > 0$ and label the corresponding $\boldsymbol{\tau}$-vector as $\boldsymbol{\tau}^+$. Calculating the trace we obtain

$$\chi_{i\sigma\sigma,i\sigma'\sigma'}(\boldsymbol{\tau}^+) = \frac{e^{\tau_{12}U/2+\tau_{34}U/2} + \delta_{\sigma\sigma'} e^{(\beta-\tau_{12})U/2-\tau_{34}U/2}}{2(1+e^{\beta U/2})} - G^\sigma_{i,i}(\tau_{12})\, G^{\sigma'}_{i,i}(\tau_{34}).$$

The mean-field terms $G^\sigma_{i,i}(\tau_{12})\, G^{\sigma'}_{i,i}(\tau_{34})$ cancel out in the actual magnetic linear response function, so here we do not give their form explicitly and we will neglect them in the rest of the calculations. For a single atom, the contribution of the $\boldsymbol{\tau}^+$ sector to the imaginary-time magnetic susceptibility is

$$\chi_{zz}(\boldsymbol{\tau}^+) = (g\mu_B)^2 \, \frac{1}{4}\frac{1}{\beta} \sum_{\sigma\sigma'} s_\sigma s_{\sigma'} \chi_{i\sigma\sigma,i\sigma'\sigma'}(\boldsymbol{\tau}^+) = \frac{(g\mu_B)^2}{4\beta} \frac{1}{(1+e^{\beta U/2})} e^{(\beta-\tau_{12}-\tau_{34})U/2}.$$

Summing up the contributions of all imaginary-time sectors and performing the Fourier transform we obtain $\chi^{n,n'}_{\sigma\sigma\sigma'\sigma'}(i\omega_n)$, defined in Eq. (98). For $U \neq 0$ this tensor is non-diagonal in the fermionic Matsubara frequencies. For $\omega_n = 0$ we have [11]

$$\sum_{\sigma\sigma'} \sigma\sigma' \, \chi^{n,n'}_{i\sigma\sigma,i\sigma'\sigma'}(0) = M_{n'}\frac{dM_n}{dy} + M_n\frac{dM_{n'}}{dy} - \beta n(y)\left[\delta_{n,n'} + \delta_{n,-n'}\right]\frac{dM_n}{dy} + \beta n(-y) M_n M_{n'} - \frac{1}{y}\left\{ M_{n'} - \beta\left[n(y)\delta_{n,-n'} - n(-y)\delta_{n,n'}\right]\right\} M_n \tag{102}$$

where

$$M_n = \frac{1}{i\nu_n - y} - \frac{1}{i\nu_n + y}. \tag{103}$$

We can now calculate the magnetic susceptibility via Eq. (97), recovering the expected result, Eq. (101). The resulting atomic magnetic susceptibility is thus proportional to $1/k_B T$, i.e., has a Curie-like behavior; furthermore it is zero at finite frequency. The temperature dependence can be remarkably different from the $U = 0$ limit. Indeed, if the density of states is flat around the Fermi level, as it is often the case in three-dimensional lattices, the non-interacting Pauli susceptibility $\chi_{zz}(\mathbf{0}; 0)$ is weakly temperature dependent. As we have seen, a strong temperature dependence can be found, however, if, e.g., a logarithmic van-Hove singularity is at the Fermi level, as in the case of the square lattice at half filling shown in Fig. 16.

3.4.4 DMFT: $\chi_0(\mathbf{q}; \omega)$ and the Bethe-Salpeter equation

In order to calculate the magnetic susceptibility with DMFT we need, first of all, $\chi_0(\mathbf{q}; \omega)$, the bubble term calculated from the DMFT Green functions. We consider here the small t/U or Mott insulating regime. In this case we can derive an approximate local self-energy starting from the atomic limit. The t=0 local Green function is

$$G^{\sigma}_{i,i}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \Sigma^{\sigma}_l(i\nu_n)},$$

where the local self-energy is given by

$$\Sigma^{\sigma}_l(i\nu_n) = \frac{U}{2} + \frac{U^2}{4}\frac{1}{i\nu_n + \mu - \varepsilon_d - \frac{U}{2}}, \tag{104}$$

and $\mu = \varepsilon_d + \frac{U}{2}$. In the Mott insulating regime we can assume that the local self-energy has the same form (104), with $U^2/4$ replaced by a quantity which plays the role of a dimensionless *order parameter* [17] for the insulating phase

$$\frac{1}{r_U}\frac{4}{U^2} = \int_{-\infty}^{+\infty} d\varepsilon \, \frac{\rho(\varepsilon)}{\varepsilon^2}. \tag{105}$$

Here $\rho(\varepsilon)$ is the density of states per spin. The integral in Eq. (105) diverges in the metallic phase. The Green function can then be rewritten as

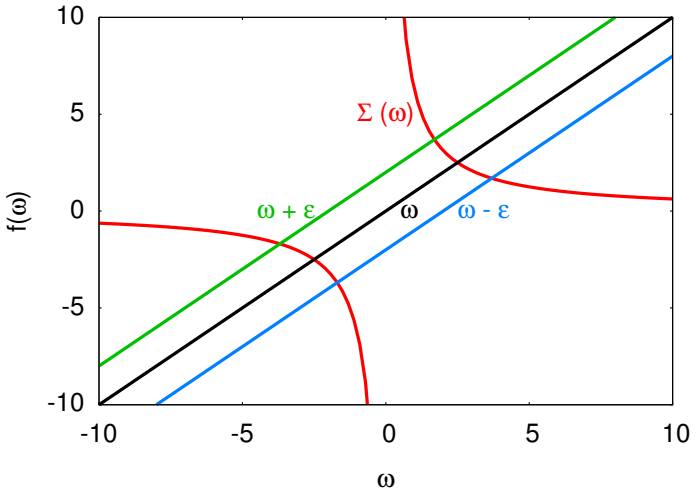


Fig. 17: *Graphical solution of the equation $\omega - \varepsilon_{\boldsymbol{k}} = \Sigma_l^\sigma(\omega)$ yielding the poles $E_{\boldsymbol{k}}^+$ and $E_{\boldsymbol{k}}^-$ of the Green function defined in Eq. (106).*

$$G^\sigma(\boldsymbol{k}; i\nu_n) = \frac{1}{i\nu_n - \Sigma_l^\sigma(i\nu_n) - \varepsilon_{\boldsymbol{k}}} = \frac{1}{E_{\boldsymbol{k}}^+ - E_{\boldsymbol{k}}^-}\left[\frac{E_{\boldsymbol{k}}^+}{i\nu_n - E_{\boldsymbol{k}}^+} - \frac{E_{\boldsymbol{k}}^-}{i\nu_n - E_{\boldsymbol{k}}^-}\right] \tag{106}$$

where $E_{\boldsymbol{k}}^+$ and $E_{\boldsymbol{k}}^-$ are the two roots of the equation $\omega - \Sigma_l^\sigma(\omega) - \varepsilon_{\boldsymbol{k}} = 0$,

$$E_{\boldsymbol{k}}^\pm = \frac{1}{2}\varepsilon_{\boldsymbol{k}} \pm \frac{1}{2}\sqrt{\varepsilon_{\boldsymbol{k}}^2 + r_U\, U^2}.$$

By performing the Matsubara sums, one finds

$$\begin{aligned}\chi_{zz}^0(\boldsymbol{q}; 0) &= (g\mu_B)^2 \frac{1}{4}\sum_\sigma \frac{1}{\beta^2}\sum_n \chi_{\sigma\sigma\sigma\sigma}^{n,n}(0) \\ &= (g\mu_B)^2 \frac{1}{2}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}}\Big[\underbrace{-I_{\boldsymbol{k},\boldsymbol{q}}^{++} - I_{\boldsymbol{k},\boldsymbol{q}}^{--}}_{A_{\boldsymbol{k},\boldsymbol{q}}} + \underbrace{I_{\boldsymbol{k},\boldsymbol{q}}^{+-} + I_{\boldsymbol{k},\boldsymbol{q}}^{-+}}_{B_{\boldsymbol{k},\boldsymbol{q}}},\Big]\end{aligned}$$

where, setting $\alpha = \pm$ and $\gamma = \pm$,

$$I_{\boldsymbol{k},\boldsymbol{q}}^{\alpha\gamma} = \frac{E_{\boldsymbol{k}}^\alpha\, E_{\boldsymbol{k}+\boldsymbol{q}}^\gamma}{\big(E_{\boldsymbol{k}}^+ - E_{\boldsymbol{k}}^-\big)\big(E_{\boldsymbol{k}+\boldsymbol{q}}^+ - E_{\boldsymbol{k}+\boldsymbol{q}}^-\big)}\frac{n(E_{\boldsymbol{k}}^\alpha) - n(E_{\boldsymbol{k}+\boldsymbol{q}}^\gamma)}{E_{\boldsymbol{k}}^\alpha - E_{\boldsymbol{k}+\boldsymbol{q}}^\gamma}.$$

In the $\boldsymbol{q} \to \boldsymbol{0}$ limit

$$\begin{aligned}A_{\boldsymbol{k},\boldsymbol{0}} &= \beta\,\left[\frac{(E_{\boldsymbol{k}}^+)^2}{\varepsilon_{\boldsymbol{k}}^2 + r_U\, U^2}\, n(E_{\boldsymbol{k}}^+)\big(1 - n(E_{\boldsymbol{k}}^+)\big) + \frac{(E_{\boldsymbol{k}}^-)^2}{\varepsilon_{\boldsymbol{k}}^2 + r_U\, U^2}\, n(E_{\boldsymbol{k}}^-)\big(1 - n(E_{\boldsymbol{k}}^-)\big)\right] \\ B_{\boldsymbol{k},\boldsymbol{0}} &= \frac{r_U\, U^2}{2(\varepsilon_{\boldsymbol{k}}^2 + r_U\, U^2)^{3/2}}\,\big(n(E_{\boldsymbol{k}}^-) - n(E_{\boldsymbol{k}}^+)\big)\,.\end{aligned}$$

In the large βU limit, the $A_{\boldsymbol{k},\boldsymbol{0}}$ term, proportional to the density of states at the Fermi level, vanishes exponentially; the $B_{\boldsymbol{k},\boldsymbol{0}}$ term yields the dominant contribution. Hence

$$\chi_{zz}^0(\boldsymbol{0}; 0) \sim (g\mu_B)^2 \frac{1}{4}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}} \frac{r_U\, U^2}{(\varepsilon_{\boldsymbol{k}}^2 + r_U\, U^2)^{3/2}} \sim (g\mu_B)^2 \frac{1}{4\sqrt{r_U}\, U}\left[1 - \frac{3}{2}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}} \frac{\varepsilon_{\boldsymbol{k}}^2}{r_U\, U^2} + \dots\right].$$

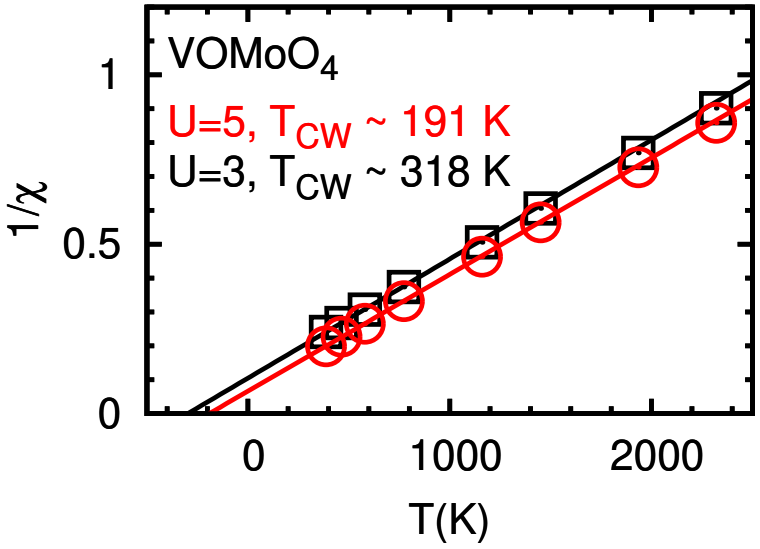


Fig. 18: *$VOMoO_4$: The Curie-Weiss behavior of the uniform magnetic susceptibility at half filling, obtained with the LDA+DMFT approach. Rearranged from Ref. [11].*

The right-hand side is equal to the atomic term $\chi^0_{zz}(0)$ minus a correction of order t^2/U^3. As we can see, $\chi^0_{zz}(\mathbf{0};0)$ is small and weakly dependent on the temperature. Here for simplicity we will discuss only the case of the two-dimensional square lattice at half filling. In the Mott-insulating regime, due to the superexchange interaction, this model exhibits an antiferromagnetic instability at $\boldsymbol{q}_C = (\pi/a, \pi/a, 0)$. Let us then calculate $\chi^0_{zz}(\boldsymbol{q}_C;0)$ and compare it with $\chi^0_{zz}(\mathbf{0};0)$. Since, as we have seen, $\varepsilon_{\boldsymbol{k}+\boldsymbol{q}_C} = -\varepsilon_{\boldsymbol{k}}$, we find

$$A_{\boldsymbol{k},\boldsymbol{q}_C} = \frac{1}{2}\frac{r_U\,U^2}{\varepsilon^2_{\boldsymbol{k}} + r_U\,U^2}\frac{n(E^+_{\boldsymbol{k}}-\varepsilon_{\boldsymbol{k}}) - n(E^+_{\boldsymbol{k}})}{\varepsilon_{\boldsymbol{k}}}$$
$$B_{\boldsymbol{k},\boldsymbol{q}_C} = \frac{1}{2}\frac{\varepsilon^2_{\boldsymbol{k}}}{\varepsilon^2_{\boldsymbol{k}} + r_U\,U^2}\frac{n(E^+_{\boldsymbol{k}}-\varepsilon_{\boldsymbol{k}}) - n(E^+_{\boldsymbol{k}})}{\varepsilon_{\boldsymbol{k}}} - \frac{1}{2}\frac{1}{\sqrt{\varepsilon^2_{\boldsymbol{k}} + r_U\,U^2}}\left(n(E^+_{\boldsymbol{k}}) - n(E^-_{\boldsymbol{k}})\right),$$

and therefore

$$\chi_0(\boldsymbol{q}_C;0) \sim (g\mu_B)^2\frac{1}{4\sqrt{r_U}U}\left(1 - \frac{1}{2}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}}\frac{\varepsilon^2_{\boldsymbol{k}}}{r_U U^2}\right).$$

Thus $\chi_0(\boldsymbol{q};0)$ is indeed larger at $\boldsymbol{q} = \boldsymbol{q}_C$ than at $\boldsymbol{q} = \mathbf{0}$; it is however weakly temperature dependent and does not exhibit Curie-Weiss instabilities. The calculation presented above can be generalized to any $\boldsymbol{q}$ vector [11], obtaining the expression

$$\chi_0(\boldsymbol{q};0) \sim (g\mu_B)^2\frac{1}{4\sqrt{r_U}U}\left(1 - \frac{1}{2}\frac{J_{\mathbf{0}}}{\sqrt{r_U}U} - \frac{1}{4}\frac{J_{\boldsymbol{q}}}{\sqrt{r_U}U}\right), \tag{107}$$

where $J_{\boldsymbol{q}} = J\big(\cos q_x + \cos q_y\big)$, and the super-exchange coupling is $J = 4t^2/U$. For the next step we need to calculate the local vertex. This requires, as we have seen, the solution of the self-consistent quantum-impurity model via the quantum-impurity solver. Here, for the purpose of illustrating how the approach works, we approximate the local susceptibility with the atomic

susceptibility in the large βU limit. Furthermore we work with the susceptibilities obtained *after* the Matsubara sums have been performed. Thus

$$\chi^0_{zz}(0) \sim (g\mu_B)^2 \frac{1}{4\sqrt{r_U}U}, \qquad \chi_{zz}(0) \sim \frac{1}{4k_BT}.$$

The local vertex is then approximately given by

$$\Gamma \sim \frac{1}{\chi^0_{zz}(0)} - \frac{1}{\chi_{zz}(0)} \sim \frac{1}{(g\mu_B)^2}\left[4\sqrt{r_U}U\left(1+\frac{1}{2}\frac{J_{\mathbf{0}}}{\sqrt{r_U}U}\right) - 4k_BT\right].$$

The last step consists in solving the Bethe-Salpeter equation

$$\chi_{zz}(\boldsymbol{q};0) = \frac{1}{(\chi^0_{zz}(\boldsymbol{q};0))^{-1} - \Gamma} \sim \frac{(g\mu_B)^2}{4}\frac{1}{k_BT + J_{\boldsymbol{q}}/4} = \frac{(g\mu_B)^2}{4k_B}\,\frac{1}{T-T_{\boldsymbol{q}}}.$$

This shows that including the local vertex correction we recover the Curie-Weiss behavior, as expected for a system described by local spins coupled by a Heisenberg-like exchange; we also correctly find the antiferromagnetic instability, since $\boldsymbol{q}_C$ is the vector for which the critical temperature $T_{\boldsymbol{q}}$ is the largest. In conclusion, we have seen that $\Gamma(i\omega_m)$ is essential to properly describe the magnetic response function of strongly-correlated systems. This can be seen in Fig. 14 for the Mott insulator $VOMoO_4$. In the figure we can compare the very weak linear magnetic response $\chi_0(\boldsymbol{q};0)$ (upper panels) with the LDA+DMFT result $\chi(\boldsymbol{q};0)$ (lower panels). The latter is not only strongly enhanced with respect to $\chi_0(\boldsymbol{q};0)$, but also exhibits the expected Curie-Weiss like behavior, as can be seen in Fig. 18 for $\boldsymbol{q} = \mathbf{0}$.

3.4.5 DMFT: Static and dynamical susceptibility below the critical temperature

In this section we will discuss the magnetic response in the anti-ferromagnetic phase, i.e., for $T \ll T_N$, where T_N is the transition temperature in DMFT. We will consider as representative case the single-band Hubbard model on a square lattice with dispersion

$$\varepsilon_{\boldsymbol{k}} = \underbrace{-2t(\cos k_x + \cos k_y)}_{\alpha_{\boldsymbol{k}}} + \underbrace{4t' \cos k_x \cos k_y}_{\gamma_{\boldsymbol{k}}} + \dots . \tag{108}$$

a model typically adopted for describing the Cu $3d$ $x^2{-}y^2$ states at the Fermi level in high-temperature superconducting cuprates. We consider again the half-filled case in the small t/U and t'/U limit, yielding an insulating ground state. The DMFT results presented for this model are from Ref. [19], where more details can be found. In the paramagnetic phase, the static susceptibility exhibits a Curie-Weiss behavior, for the same reasons we discussed in the previous section. This can be seen in Fig. 19.

In the antiferromagnetic phase, the square lattice can be divided into two sublattices, A and B, describing sites with opposite magnetic moment, m (sub-lattice A) and $-m$ (sub-lattice B); the unit cell contains $n_s{=}2$ Cu atoms, labeled with $i_c = 1, 2$. In this case the local self-energy can be approximated by the site-dependent static Hartree-Fock term

$$\Sigma^{\sigma}_{i_c,i'_c}(i\nu_n) \approx \left(-\mu + (-1)^{i_c-1} s_\sigma m U\right)\delta_{i_c,i'_c}. \tag{109}$$

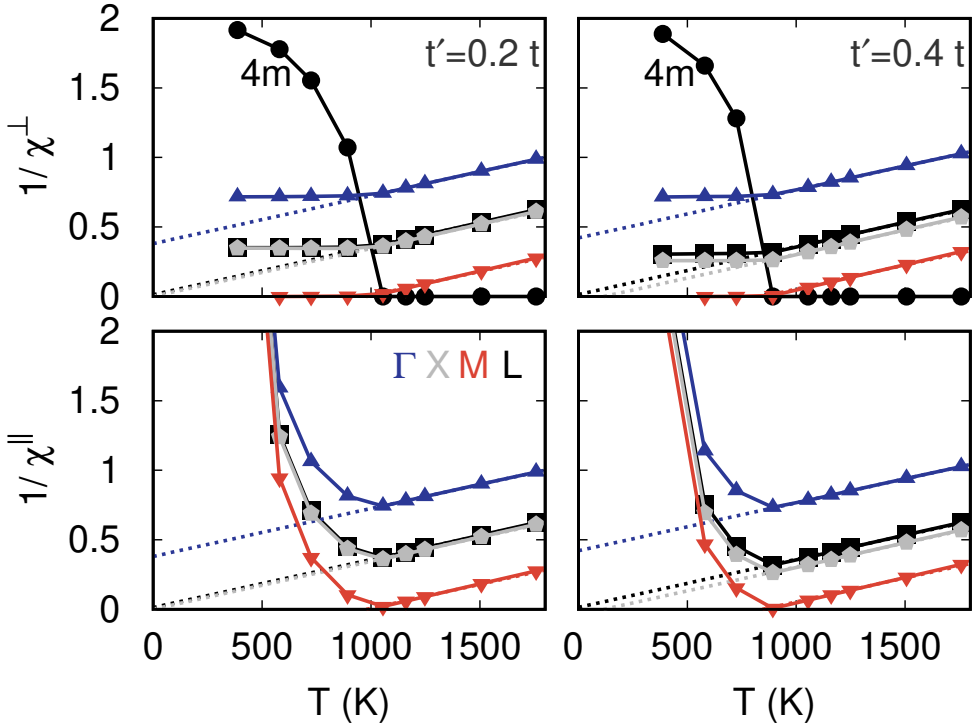


Fig. 19: *Static inverse transverse and longitudinal susceptibility $\chi(\boldsymbol{q};0)$ as a function of temperature. Triangles: Γ and M point. Gray pentagons: X point. Black circles: local (L) susceptibility. Rearranged from Ref. [19].*

For a given spin we can thus write the associated $n_s \times n_s$ Green function matrix as

$$G^{\sigma}(\boldsymbol{k};i\nu_n) \approx \frac{1}{D_{\boldsymbol{k}}(i\nu_n)} \begin{pmatrix} i\nu_n - \gamma_{\boldsymbol{k}} - s_\sigma mU & \alpha_{\boldsymbol{k}} e^{-ik_x a} \\ \alpha_{\boldsymbol{k}} e^{ik_x a} & i\nu_n - \gamma_{\boldsymbol{k}} + s_\sigma mU \end{pmatrix}, \tag{110}$$

where $D_{\boldsymbol{k}}(i\nu_n) = (i\nu_n - \gamma_{\boldsymbol{k}})^2 - \alpha_{\boldsymbol{k}}^2 - (mU)^2$. The elements of $G^{\sigma}(\boldsymbol{k};i\nu_n)$ can be re-expressed as

$$G^{\sigma}_{i_c,i'_c}(\boldsymbol{k};i\nu_n) = \sum_{p=\pm} \frac{w^{i_c i'_c}_{\sigma \boldsymbol{k} p}}{i\nu_n - E^p_{\boldsymbol{k}}}, \tag{111}$$

where the poles corresponds to the Hartree-Fock energies $E^{\pm}_{\mathbf{k}} = \gamma_{\boldsymbol{k}} \pm \sqrt{\alpha^2_{\mathbf{k}} + (mU)^2} = \gamma_{\boldsymbol{k}} \pm \Delta\alpha_{\boldsymbol{k}}$, shown in Fig. 20 for $\gamma_{\boldsymbol{k}} = 0$. The weights are

$$w^{11}_{\sigma \boldsymbol{k} p} = \frac{1}{2}\left(1 - p \frac{s_\sigma mU}{\sqrt{\alpha^2_{\boldsymbol{k}} + (mU)^2}}\right) = w^{22}_{-\sigma \mathbf{k} p}, \tag{112}$$

$$w^{12}_{\sigma \boldsymbol{k} p} = \frac{p}{2} \frac{\alpha_{\boldsymbol{k}}}{\sqrt{\alpha^2_{\mathbf{k}} + (mU)^2}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{T}_1 - \boldsymbol{T}_2)} = \left[w^{21}_{\sigma \boldsymbol{k} p}\right]^*. \tag{113}$$

Performing the Matsubara sum

$$\chi^{0;i_c i'_c}_{\sigma\sigma'\sigma'\sigma}(\mathbf{q};i\omega_m) = -\frac{1}{\beta N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}n} G^{\sigma}_{i_c,i'_c}(\boldsymbol{k};i\nu_n)\, G^{\sigma'}_{i'_c,i_c}(\boldsymbol{k}+\boldsymbol{q};i\nu_n + i\omega_m) \tag{114}$$

$$\approx \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{pp'} w^{i_c i'_c}_{\sigma \boldsymbol{k} p}\, w^{i'_c i_c}_{\sigma' \boldsymbol{k}+\boldsymbol{q} p'}\, \mathcal{I}^{pp'}_{\mathbf{k},\boldsymbol{q}}(i\omega_m), \tag{115}$$

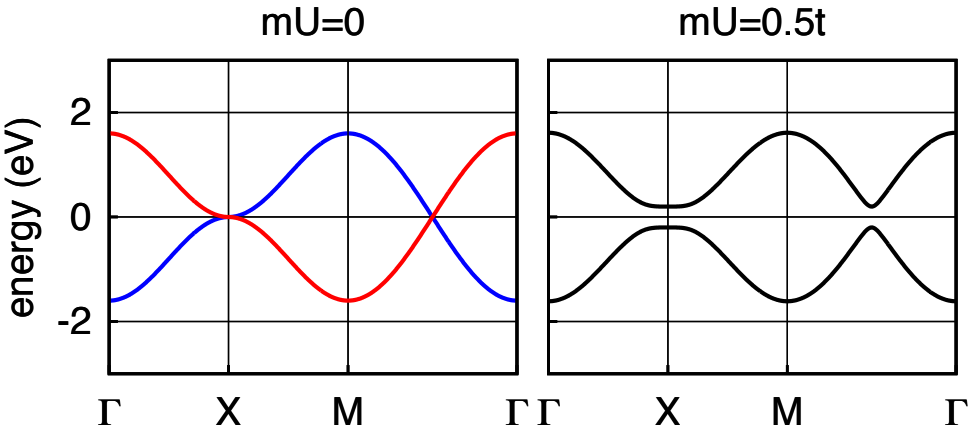


Fig. 20: *The metal-insulator transition in anti-ferromagnetic Hartree-Fock. The calculation is for a square lattice tight-binding model with dispersion* $\varepsilon_{\boldsymbol{k}} = -2t(\cos k_x + \cos k_y)$.

with

$$\mathcal{I}^{pp'}_{\boldsymbol{k},\boldsymbol{q}}(i\omega_m) = -\frac{n_F(E^p_{\boldsymbol{k}}) - n_F(E^{p'}_{\boldsymbol{k}+\boldsymbol{q}})}{i\omega_m + E^p_{\mathbf{k}} - E^{p'}_{\boldsymbol{k}+\boldsymbol{q}}}. \tag{116}$$

Below the critical temperature we have to distinguish two types (T) of response. Assuming that the ordered magnetic moments are along $\hat{z}$, a magnetic field parallel to $\hat{z}$ yields the longitudinal ($T = \parallel$) linear response, and one perpendicular to $\hat{z}$ the transverse ($T = \perp$) linear susceptibility. They can be expressed in a compact way as

$$\chi^T_0(\mathbf{q}; i\omega_m) = \frac{(g\mu_B)^2}{4} \sum_{\sigma\sigma'} \frac{1}{2} \sum_{i_c i'_c} f^T_{\sigma\sigma'}\, \chi^{0;i_c i'_c}_{\sigma\sigma'\sigma'\sigma}(\mathbf{q}; i\omega_m) e^{i\boldsymbol{q}\cdot(\boldsymbol{T}_{i_c} - \boldsymbol{T}_{i'_c})} \tag{117}$$

$$\approx \frac{(g\mu_B)^2}{4} \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{pp'} v^{T,pp'}_{\boldsymbol{k},\mathbf{q}} \mathcal{I}^{pp'}_{\boldsymbol{k},\boldsymbol{q}}(i\omega_m) \tag{118}$$

where $f^{\parallel}_{\sigma\sigma'} = \delta_{\sigma,\sigma'}$ and $f^{\perp}_{\sigma\sigma'} = \delta_{\sigma,-\sigma'}$, while

$$v^{T,pp'}_{\boldsymbol{k},\mathbf{q}} = \frac{1}{2} \sum_{\sigma\sigma'} \sum_{i_c i'_c} w^{i_c i'_c}_{\sigma\boldsymbol{k}p} w^{i'_c i_c}_{\sigma'\boldsymbol{k}+\mathbf{q}p'} f^T_{\sigma\sigma'} e^{i\varphi^{i_c i'_c}_{\boldsymbol{q}}} \tag{119}$$

$$= \frac{1}{2}\left(1 + pp' \frac{\alpha_{\boldsymbol{k}}\alpha_{\boldsymbol{k}+\boldsymbol{q}} + f^T_{\sigma\sigma'} s_\sigma s_{\sigma'} (mU)^2}{\Delta\alpha_{\mathbf{k}} \Delta\alpha_{\boldsymbol{k}+\boldsymbol{q}}}\right). \tag{120}$$

We note that $v^{\parallel,pp}_{\boldsymbol{k},\mathbf{0}}=1$ and $v^{\parallel,p-p}_{\boldsymbol{k},\mathbf{0}}=0$, while $v^{\perp,p-p}_{\boldsymbol{k},M} = 1$ and $v^{\parallel,pp}_{\boldsymbol{k},M} = 0$. For simplicity in the discussion that follows we set $t' = 0$. In the low-temperature limit only the terms with $p = -p'$ contribute. For $t \ll U$ we have

$$\chi^T_0(\boldsymbol{q}; i\omega_m) \approx \frac{(g\mu_B)^2}{4} \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} v^{T,+-}_{\boldsymbol{k},\boldsymbol{q}} \left(\frac{1}{i\omega_m + U} - \frac{1}{i\omega_m - U}\right). \tag{121}$$

Here we can see that the excitation energies are of order U, and not of the order of the superexchange couplings as we would expect for the Hubbard model.

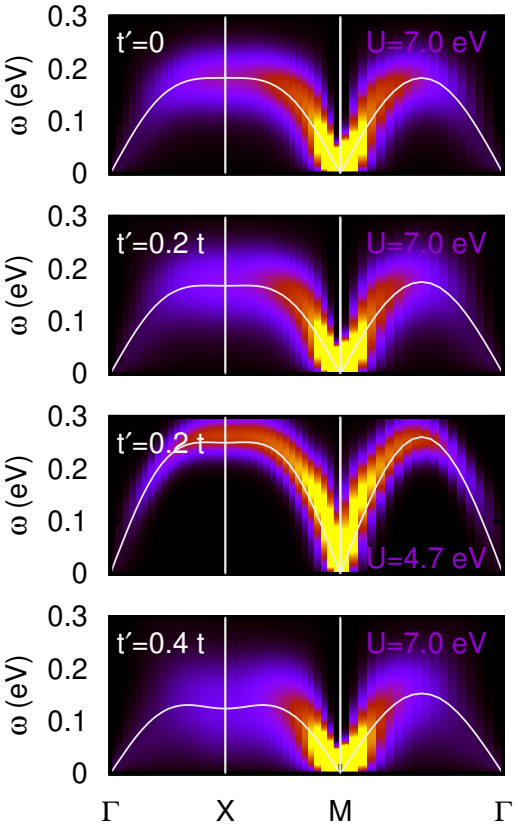


Fig. 21: *DMFT spin-wave spectra for fixed* t *and representative values of* t'*. The white lines are linear spin-wave curves calculated using super-exchange couplings* $J_1 = \frac{4t^2}{U}$ *and* $J_2 = \frac{4(t')^2}{U}$*. Rearranged from Ref. [19].*

In order to recover the correct behavior we have to solve the Bethe-Salpeter equation. To this end we have first to return to the site-dependent tensor, since we have to invert the associated $n_s \times n_s$ matrices. Using the transverse susceptibility as representative case, the relevant matrix elements in the ordered phase are

$$\chi^{0;i_c i'_c}_{\sigma-\sigma-\sigma\sigma}(\boldsymbol{q}; i\omega_m) \approx \left(-\frac{a^{i_c i'_c}_{\sigma}(\boldsymbol{q})}{i\omega_n - U} + \frac{a^{i_c i'_c}_{-\sigma}(\boldsymbol{q})}{i\omega_n + U} \right) e^{-i\boldsymbol{q}\cdot(\boldsymbol{T}_{i_c} - \boldsymbol{T}_{i'_c})}, \tag{122}$$

where

$$a^{i_c i'_c}_{\sigma}(\boldsymbol{q}) = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \sum_{pp'} w^{i_c i'_c}_{\sigma \boldsymbol{k} p} w^{i'_c i_c}_{-\sigma \boldsymbol{k}+\boldsymbol{q} p'} \delta_{p,+} \delta_{p,-}. \tag{123}$$

In the $t \ll U$ limit, at linear order in J_1/U we have

$$a^{11}_{\sigma}(\boldsymbol{q}) = a^{22}_{-\sigma}(\boldsymbol{q}) \approx \frac{1}{4}(1-s_\sigma)^2 + s_\sigma(1-s_\sigma)\frac{J_1}{U}, \tag{124}$$

$$a^{12}_{\sigma}(\boldsymbol{q}) = a^{21}_{-\sigma}(\boldsymbol{q}) \approx -\frac{J_1}{U} f_{\boldsymbol{q}}, \tag{125}$$

where $f_{\boldsymbol{q}} = (\cos q_x + \cos q_y)/2$. The $i_c{=}i'_c$ elements are therefore independent on $\boldsymbol{q}$ at order J_1/U. By inverting the susceptibility matrix with the elements defined above we thus obtain, at

sufficiently low frequency

$$\left[\frac{1}{\chi^0(\boldsymbol{q};i\omega_m)}-\frac{1}{\chi^0(i\omega_m)}\right]^{i_c i'_c}_{\sigma-\sigma-\sigma\sigma}\approx 2J_1 f_{\boldsymbol{q}}(1-\delta_{i_c i'_c})e^{-i\boldsymbol{q}\cdot(\boldsymbol{T}_{i_c}-\boldsymbol{T}_{i'_c})} \tag{126}$$

By adding to this the inverse of the local susceptibility matrix

$$\chi^{i_c i'_c}_{\sigma-\sigma-\sigma\sigma}(i\omega_m)\approx\left(-\frac{a^{i_c i'_c}_{\sigma}}{i\omega_n-2J_1}+\frac{a^{i_c i'_c}_{-\sigma}}{i\omega_n+2J_1}\right)\delta_{i_c,i'_c}$$

with $a^{i_c i'_c}_{\sigma}=\frac{1}{N_{\boldsymbol{q}}}\sum_{\boldsymbol{q}}a^{i_c i'_c}_{\sigma}(\boldsymbol{q})$. Inverting again, we finally obtain

$$\chi^{\perp}(\boldsymbol{q};i\omega_m)=\frac{(g\mu_B)^2}{4}\sum_{\sigma}\frac{1}{2}\sum_{i_c i'_c}\chi^{i_c i'_c}_{\sigma-\sigma-\sigma\sigma}(\mathbf{q};i\omega_m)e^{i\boldsymbol{q}\cdot(\boldsymbol{T}_{i_c}-\boldsymbol{T}_{i'_c})} \tag{127}$$

$$\approx(g\mu_B)^2\frac{J_1(1-f_{\boldsymbol{q}})}{\omega_m^2+4J_1^2(1-f_{\boldsymbol{q}}^2)}, \tag{128}$$

which is the expected behavior for a Heisenberg antiferromagnet. The result of actual DMFT calculations is shown in Fig. 19 and Fig. 21. More details can be found in Ref. [19].

4 Conclusion

The LDA+DMFT approach and its extension has proved very successful for describing correlated materials. It has shown us that materials details do matter, contrarily to what often was assumed in the past; for example a crystal field much smaller than the bandwidth can favor the Mott metal-insulator transition [18]. The method is becoming progressively more and more versatile. It is now possible, e.g., to study multi-orbital Hubbard-like models including the full Coulomb vertex and the spin-orbit interaction. Successful extension schemes, e.g., cluster methods, account, at least in part, for the $\boldsymbol{q}$-dependence of the self-energy. In this lecture, we have seen how to use the LDA+DMFT approach to calculate not only Green and spectral functions but also linear-response functions. In the scheme presented, the local susceptibility is obtained via the quantum-impurity solver at the end of the self-consistency loop; the $\boldsymbol{q}$-dependent susceptibility is, instead, calculated solving in addition the Bethe-Salpeter equation in the local-vertex approximation. As representative case we have studied the magnetic susceptibility of the one-band Hubbard model at half filling. The extension of the LDA+DMFT approach to the calculation of generalized susceptibilities makes it possible to put the method and the approximations adopted to more stringent tests. This is key for further advancing the theoretical tools for the description of strong correlation effects in real materials.

Appendix

A Eigenstates of Hubbard dimer and Anderson molecule

The Hamiltonian of the **Hubbard dimer** is given by

$$\hat{H} = \varepsilon_d \sum_\sigma \sum_{i=1,2} \hat{n}_{i\sigma} - t \sum_\sigma \left(c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma} \right) + U \sum_{i=1,2} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}.$$

It commutes with the number of electron operator $\hat{N}$, with the total spin $\hat{S}$ and with $\hat{S}_z$. Thus we can express the many-body states in the atomic limit as

$\|N,S,S_z\rangle$			N	S	$E(N,S)$
$\|0,0,0\rangle$	$=$	$\|0\rangle$	0	0	0
$\|1,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}\|0\rangle$	1	$1/2$	ε_d
$\|1,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}\|0\rangle$	1	$1/2$	ε_d
$\|2,1,1\rangle$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{1\uparrow}\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,1,-1\rangle$	$=$	$c^\dagger_{2\downarrow}c^\dagger_{1\downarrow}\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,1,0\rangle$	$=$	$\frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right)\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,0,0\rangle_0$	$=$	$\frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right)\|0\rangle$	2	0	$2\varepsilon_d$
$\|2,0,0\rangle_1$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$
$\|2,0,0\rangle_2$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$
$\|3,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	3	$1/2$	$3\varepsilon_d + U$
$\|3,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	3	$1/2$	$3\varepsilon_d + U$
$\|4,0,0\rangle$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	4	0	$4\varepsilon_d + 2U$

Let us order the $N = 1$ states as in the table above, first the spin up and then spin down block. For finite t the Hamiltonian matrix for $N = 1$ electrons takes then the form

$$\hat{H}_1 = \begin{pmatrix} \varepsilon_d & -t & 0 & 0 \\ -t & \varepsilon_d & 0 & 0 \\ 0 & 0 & \varepsilon_d & -t \\ 0 & 0 & -t & \varepsilon_d \end{pmatrix}.$$

This matrix can be easily diagonalized and yields the *bonding* $(-)$ and *antibonding* $(+)$ states

$\lvert 1,S,S_z\rangle_\alpha$		$E_\alpha(1,S)$	$d_\alpha(1,S)$
$\lvert 1,1/2,\sigma\rangle_+$	$= \frac{1}{\sqrt{2}}\big(\lvert 1,1/2,\sigma\rangle_1 - \lvert 1,1/2,\sigma\rangle_2\big)$	$\varepsilon_d{+}t$	2
$\lvert 1,1/2,\sigma\rangle_-$	$= \frac{1}{\sqrt{2}}\big(\lvert 1,1/2,\sigma\rangle_1 + \lvert 1,1/2,\sigma\rangle_2\big)$	$\varepsilon_d{-}t$	2

where $d_\alpha(N)$ is the spin degeneracy of the α manifold.
For $N = 2$ electrons (half filling), the hopping integrals only couple the three $S = 0$ states, and therefore the Hamiltonian matrix is given by

$$\hat{H}_2 = \begin{pmatrix} 2\varepsilon_d & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d + U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_d + U \end{pmatrix}.$$

The eigenvalues and the corresponding eigenvectors are

$\lvert 2,S,S_z\rangle_\alpha$		$E_\alpha(2,S)$	$d_\alpha(2,S)$
$\lvert 2,0,0\rangle_+$	$= a_1\lvert 2,0,0\rangle_0 - \frac{a_2}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + \frac{1}{2}\big(U{+}\Delta(t,U)\big)$	1
$\lvert 2,0,0\rangle_o$	$= \frac{1}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 - \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + U$	1
$\lvert 2,1,m\rangle_o$	$= \lvert 2,1,m\rangle$	$2\varepsilon_d$	3
$\lvert 2,0,0\rangle_-$	$= a_2\lvert 2,0,0\rangle_0 + \frac{a_1}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + \frac{1}{2}\big(U{-}\Delta(t,U)\big)$	1

where

$$\Delta(t,U) = \sqrt{U^2 + 16t^2},$$

and

$$a_1^2 = a_1^2(t,U) = \frac{1}{\Delta(t,U)}\frac{\Delta(t,U) - U}{2} \qquad a_2^2 = a_2^2(t,U) = \frac{4t^2}{\Delta(t,U)}\frac{2}{(\Delta(t,U) - U)},$$

so that $a_1 a_2 = 2t/\Delta(t,U)$. For $U = 0$ we have $a_1 = a_2 = 1/\sqrt{2}$, and the two states $|2,0,0\rangle_-$ and $|2,0,0\rangle_+$ become, respectively, the state with two electrons in the bonding orbital and the state with two electrons in the antibonding orbital; they have energy $E_\pm(2,0) = 2\varepsilon_d \pm 2t$; the remaining states have energy $2\varepsilon_d$ and are non-bonding. For $t > 0$, the ground state is unique and it is always the singlet $|2,0,0\rangle_-$; in the large U limit its energy is

$$E_-(2,0) \sim 2\varepsilon_d - 4t^2/U.$$

In this limit the energy difference between the first excited state, a triplet state, and the singlet ground state is thus equal to the Heisenberg antiferromagnetic coupling

$$E_o(2,1) - E_-(2,0) \sim 4t^2/U = \Gamma.$$

Finally, for $N = 3$ electrons, eigenstates and eigenvectors are

$\|3,S,S_z\rangle_\alpha$		$E_\alpha(3)$	$d_\alpha(3,S)$
$\|3,1/2,\sigma\rangle_+$	$= \frac{1}{\sqrt{2}}\big(\|1,1/2,\sigma\rangle_1 + \|1,1/2,\sigma\rangle_2\big)$	$3\varepsilon_d + U + t$	2
$\|3,1/2,\sigma\rangle_-$	$= \frac{1}{\sqrt{2}}\big(\|1,1/2,\sigma\rangle_1 - \|1,1/2,\sigma\rangle_2\big)$	$3\varepsilon_d + U - t$	2

If we exchange holes and electrons, the $N = 3$ case is identical to the $N = 1$ electron case. This is due to the particle-hole symmetry of the model.

The Hamiltonian of the **Anderson molecule** is given by

$$\hat{H} = \varepsilon_s \sum_\sigma \hat{n}_{2\sigma} - t \sum_\sigma \left(c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma}\right) + \varepsilon_d \sum_\sigma \hat{n}_{1\sigma} + U \hat{n}_{1\uparrow}\hat{n}_{1\downarrow}.$$

In the atomic limit, its eigenstates states can be classified as

$\|N,S,S_z\rangle$			N	S	$E(N,S)$
$\|0,0,0\rangle$	$=$	$\|0\rangle$	0	0	0
$\|1,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}\|0\rangle$	1	1/2	ε_d
$\|1,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}\|0\rangle$	1	1/2	ε_s
$\|2,1,1\rangle$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{1\uparrow}\|0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$\|2,1,-1\rangle$	$=$	$c^\dagger_{2\downarrow}c^\dagger_{1\downarrow}\|0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$\|2,1,0\rangle$	$=$	$\frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right)\|0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$\|2,0,0\rangle_0$	$=$	$\frac{1}{\sqrt{2}}\left(c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right)\|0\rangle$	2	0	$\varepsilon_d + \varepsilon_s$
$\|2,0,0\rangle_1$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$
$\|2,0,0\rangle_2$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	2	0	$2\varepsilon_s$
$\|3,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	3	1/2	$\varepsilon_d + 2\varepsilon_s$
$\|3,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	3	1/2	$2\varepsilon_d + \varepsilon_s + U$
$\|4,0,0\rangle$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	4	0	$2\varepsilon_d + 2\varepsilon_s + U$

For $N = 1$ electrons the Hamiltonian can be written in the matrix form

$$\hat{H}_1 = \begin{pmatrix} \varepsilon_d & -t & 0 & 0 \\ -t & \varepsilon_s & 0 & 0 \\ 0 & 0 & \varepsilon_d & -t \\ 0 & 0 & -t & \varepsilon_s \end{pmatrix}.$$

The eigenstates are thus

$\lvert 1, S, S_z\rangle_\alpha$	$E_\alpha(1,S)$	$d_\alpha(1,S)$
$\lvert 1,1/2,\sigma\rangle_+ = \alpha_1\lvert 1,1/2,\sigma\rangle_1 - \alpha_2\lvert 1,1/2,\sigma\rangle_2$	$\frac{1}{2}\big(\varepsilon_d+\varepsilon_s+\sqrt{(\varepsilon_d-\varepsilon_s)^2+4t^2}\big)$	2
$\lvert 1,1/2,\sigma\rangle_- = \alpha_2\lvert 1,1/2,\sigma\rangle_1 + \alpha_1\lvert 1,1/2,\sigma\rangle_2$	$\frac{1}{2}\big(\varepsilon_d+\varepsilon_s-\sqrt{(\varepsilon_d-\varepsilon_s)^2+4t^2}\big)$	2

where $d_\alpha(N)$ is the spin degeneracy of the α manifold. For $\varepsilon_s = \varepsilon_d + U/2$ the eigenvalues are

$$E_\pm(1,S) = \varepsilon_d + \frac{1}{4}\big(U \pm \Delta(t,U)\big),$$

while the coefficients are $\alpha_1 = a_1(t,U)$ and $\alpha_2 = a_2(t,U)$.
For N=2 electrons, the hopping integrals only couple the S=0 states. The Hamiltonian is

$$\hat{H}_2 = \begin{pmatrix} \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_d{+}\varepsilon_s & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_d{+}\varepsilon_s & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d{+}U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_s \end{pmatrix}$$

For $\varepsilon_s = \varepsilon_d + U/2$ the eigenvalues and the corresponding eigenvectors are

$\lvert 2, S, S_z\rangle_\alpha$	$E_\alpha(2,S)$	$d_\alpha(2,S)$
$\lvert 2,0,0\rangle_+ = b_1\lvert 2,0,0\rangle_0 - \frac{b_2}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + \frac{U}{2} + \frac{1}{4}\big(U{+}2\Delta(t,\frac{U}{2})\big)$	1
$\lvert 2,0,0\rangle_o = \frac{1}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 - \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + U$	1
$\lvert 2,1,m\rangle_o = \lvert 2,1,m\rangle$	$2\varepsilon_d + \frac{U}{2}$	3
$\lvert 2,0,0\rangle_- = b_2\lvert 2,0,0\rangle_0 + \frac{b_1}{\sqrt{2}}\big(\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\big)$	$2\varepsilon_d + \frac{U}{2} + \frac{1}{4}\big(U{-}2\Delta(t,\frac{U}{2})\big)$	1

where $b_1 = a_1(t,U/2)$ and $b_2 = a_2(t,U/2)$. These states have the same form as in the case of the Hubbard dimer; the ground state energy and the weight of doubly occupied states in $\lvert 2,0,0\rangle_-$ differ, however. Finally, for $N = 3$ electrons, the eigenstates are

$\lvert 3, S, S_z\rangle_\alpha$	$E_\alpha(3,S)$	$d_\alpha(3,S)$
$\lvert 3,1/2,\sigma\rangle_+ = \alpha_2\lvert 1,1/2,\sigma\rangle_1 + \alpha_1\lvert 1,1/2,\sigma\rangle_2$	$3\varepsilon_d + U + \frac{1}{4}\big(U{+}\Delta(t,U)\big)$	2
$\lvert 3,1/2,\sigma\rangle_- = \alpha_1\lvert 1,1/2,\sigma\rangle_1 - \alpha_2\lvert 1,1/2,\sigma\rangle_2$	$3\varepsilon_d + U + \frac{1}{4}\big(U{-}\Delta(t,U)\big)$	2

B Lehmann representation of the local Green function

For a single-orbital model, the local Matsubara Green function for a given site i is defined as

$$G^{\sigma}_{i,i}(i\nu_n) = -\int_0^{\beta} d\tau e^{i\nu_n\tau}\big\langle \mathcal{T} c_{i\sigma}(\tau)c^{\dagger}_{i\sigma}(0)\big\rangle,$$

where $\mathcal{T}$ is the time-ordering operator, $\beta = 1/k_BT$, and ν_n a fermionic Matsubara frequency. Let us assume we know all eigenstates $|N_l\rangle$ and their energy $E_l(N)$, for arbitrary number of electrons N. Thus, formally

$$G^{\sigma}_{i,i}(i\nu_n) = -\frac{1}{Z}\sum_{Nl}\int_0^{\beta} d\tau e^{i\nu_n\tau}e^{-\Delta E_l(N)\beta}\big\langle N_l\big|c_{i\sigma}(\tau)c^{\dagger}_{i\sigma}(0)\big|N_l\big\rangle,$$

where $Z = \sum_{Nl} e^{-\Delta E_l(N)\beta}$ is the partition function, $\Delta E_l(N) = E_l(N) - \mu N$ with μ the chemical potential, and $c^{\dagger}_{i\sigma}(0) = c^{\dagger}_{i\sigma}$. We now insert a complete set of states, obtaining

$$\begin{aligned} G^{\sigma}_{i,i}(i\nu_n) &= -\frac{1}{Z}\sum_{ll'NN'}\int_0^{\beta} d\tau e^{i\nu_n\tau}e^{-\Delta E_l(N)\beta}\big\langle N_l\big|c_{i\sigma}(\tau)|N'_{l'}\big\rangle\big\langle N'_{l'}\big|c^{\dagger}_{i\sigma}\big|N_l\big\rangle \\ &= -\frac{1}{Z}\sum_{ll'NN'}\int_0^{\beta} d\tau e^{-\Delta E_l(N)\beta}e^{(i\nu_n+\Delta E_l(N)-\Delta E_{l'}(N'))\tau}\big|\langle N'_{l'}|c^{\dagger}_{i\sigma}|N_l\rangle\big|^2 \\ &= \frac{1}{Z}\sum_{ll'NN'}\frac{e^{-\Delta E_{l'}(N')\beta}+e^{-\Delta E_l(N)\beta}}{i\nu_n+\Delta E_l(N)-\Delta E_{l'}(N')}\big|\langle N'_{l'}|c^{\dagger}_{i\sigma}|N_l\rangle\big|^2. \end{aligned}$$

Due to the weight $\big|\langle N'_{l'}|c^{\dagger}_{i\sigma}(0)|N_l\rangle\big|^2$ only the terms for which $N' = N+1$ contribute. Thus, after exchanging the labels $l'N' \leftrightarrow lN$ in the first addend, we obtain the Lehmann representation

$$G^{\sigma}_{i,i}(i\nu_n)=\sum_{ll'N}\frac{e^{-\beta\Delta E_l(N)}}{Z}\left(\frac{\big|\langle(N-1)_{l'}|c_{i\sigma}|N_l\rangle\big|^2}{i\nu_n-\Delta E_l(N)+\Delta E_{l'}(N-1)}+\frac{\big|\langle(N+1)_{l'}|c^{\dagger}_{i\sigma}|N_l\rangle\big|^2}{i\nu_n-\Delta E_{l'}(N+1)+\Delta E_l(N)}\right).$$

Let us consider as example the atomic limit of the Hubbard model at half filling. In this case all sites are decoupled; there are four eigenstates per site, the vacuum $|0\rangle$, with $\Delta E(0) = 0$, the doublet $|1_\sigma\rangle = c^{\dagger}_{i\sigma}|0\rangle$, with $\Delta E_\sigma(1) = -U/2$, and the doubly-occupied singlet $|2\rangle = c^{\dagger}_{i\uparrow}c^{\dagger}_{i\downarrow}|0\rangle$, with $\Delta E(2) = 0$. Furthermore, $Z = 2(1+e^{\beta U/2})$ and

$$\big|\langle(N-1)_{l'}|c_{i\sigma}|N_l\rangle\big|^2=\begin{cases}1 & \text{if } |N_l\rangle=|2\rangle\vee|1_\sigma\rangle\\ 0 & \text{otherwise}\end{cases}\qquad \big|\langle(N+1)_{l'}|c^{\dagger}_{i\sigma}|N_l\rangle\big|^2=\begin{cases}1 & \text{if } |N_l\rangle=|0\rangle\vee|1_{-\sigma}\rangle\\ 0 & \text{otherwise}\end{cases}$$

Thus, after summing up the four non-zero contributions, we find

$$G^{\sigma}_{i,i}(\nu_n) = \frac{1}{2}\left(\frac{1}{i\nu_n+U/2}+\frac{1}{i\nu_n-U/2}\right).$$

C Atomic magnetic susceptibility

Let us consider an idealized single-level atom described by the Hamiltonian

$$\hat{H} = \varepsilon_d\big(\hat{n}_\uparrow + \hat{n}_\downarrow\big) + U n_\uparrow n_\downarrow.$$

The eigenstates of this system, $|\Psi_i^N\rangle$, as well as their expectation values at half filling are

$\lvert\Psi_i^N\rangle$	N	$\Delta E_i = \langle\Psi_i^N\lvert\hat{H}-\mu\hat{N}\rvert\Psi_i^N\rangle$
$\lvert 0\rangle$	0	0
$c_\sigma^\dagger\lvert 0\rangle$	1	$-\frac{U}{2}$
$c_\uparrow^\dagger c_\downarrow^\dagger\lvert 0\rangle$	2	0

The magnetic susceptibility in Matsubara space is given by

$$\Big[\chi_{zz}(i\omega_m)\Big]_{nn'} = \beta\frac{1}{4}(g\mu_B)^2\sum_P \mathrm{sign}(P) f_P$$

$$f_P(i\omega_{P_1}, i\omega_{P_2}, i\omega_{P_3}) = \int_0^\beta d\tau_{14}\int_0^{\tau_{14}} d\tau_{24}\int_0^{\tau_{24}} d\tau_{34}\; e^{i\omega_{P_1}\tau_{14}+i\omega_{P_2}\tau_{24}+i\omega_{P_3}\tau_{34}} f_P(\tau_{14},\tau_{24},\tau_{34})$$

where $P = A, B, \ldots$ are the six possible permutations of the indices (123) and

$$\begin{aligned} f_P(\tau_{14},\tau_{24},\tau_{34}) &= \frac{1}{Z}\sum_{\sigma\sigma'}\sigma\sigma'\mathrm{Tr}\; e^{-\beta(\hat{H}-\mu\hat{N})}\left[\hat{o}_{P_1}(\tau_{14})\hat{o}_{P_2}(\tau_{24})\hat{o}_{P_3}(\tau_{34})c_{\sigma'}^\dagger\right] \\ &= \frac{1}{Z}\sum_{\sigma\sigma'}\sigma\sigma'\sum_{ijkl} e^{-\beta\Delta E_i}\langle i|\hat{o}_{P_1}|j\rangle\langle j|\hat{o}_{P_2}|k\rangle\langle k|\hat{o}_{P_3}|l\rangle\langle l|c_{\sigma'}^\dagger|i\rangle \\ &\quad\times\left[e^{\Delta E_{ij}\tau_{14}+\Delta E_{jk}\tau_{24}+\Delta E_{kl}\tau_{34}}\right], \end{aligned}$$

where $\Delta E_{ij} = \Delta E_i - \Delta E_j$. For the identity permutation the operators are $\hat{o}_{P_1} = c_\sigma$, $\hat{o}_{P_2} = c_\sigma^\dagger$, and $\hat{o}_{P_3} = c_{\sigma'}$ and the frequencies are $\omega_1 = \nu_n$, $\omega_2 = -\omega_m - \nu_n$, $\omega_3 = \omega_m + \nu_{n'}$. This expression can be used to calculate the magnetic susceptibility of any one-band system whose eigenvalues and eigenvectors are known, e.g., via exact diagonalization. In the case of our idealized atom

$$f_E(\tau_{14},\tau_{24},\tau_{34}) = \frac{1}{(1+e^{\beta U/2})}\; e^{\beta U/2}\; e^{-(\tau_{12}+\tau_{34})U/2} = \frac{1}{(1+e^{\beta U/2})}\; g_E(\tau_{14},\tau_{24},\tau_{34}).$$

The frequencies and functions $f_P(\tau_{14},\tau_{24},\tau_{34})$ for all permutations are given in the table below

	ω_{P_1}	ω_{P_2}	ω_{P_3}	$g_P(\tau_{14},\tau_{24},\tau_{34})$	$\mathrm{sign}(P)$
$E(123)$	ν_n	$-\omega_m-\nu_n$	$\omega_m+\nu_{n'}$	$e^{\beta U/2}\; e^{-(\tau_{12}+\tau_{34})U/2}$	$+$
$A(231)$	$-\omega_m-\nu_n$	$\omega_m+\nu_{n'}$	ν_n	$e^{\beta U/2}\; e^{-(\tau_{12}+\tau_{34})U/2}$	$+$
$B(312)$	$\omega_m+\nu_{n'}$	ν_n	$-\omega_m-\nu_n$	$-e^{+(\tau_{12}+\tau_{34})U/2}$	$+$
$C(213)$	$-\omega_m-\nu_n$	ν_n	$\omega_m+\nu_{n'}$	$-e^{\beta U/2}e^{-(\tau_{12}+\tau_{34})U/2}$	$-$
$D(132)$	ν_n	$\omega_m+\nu_{n'}$	$-\omega_m-\nu_n$	$e^{+(\tau_{12}+\tau_{34})U/2}$	$-$
$F(321)$	$\omega_m+\nu_{n'}$	$-\omega_m-\nu_n$	ν_n	$e^{+(\tau_{12}+\tau_{34})U/2}$	$-$

The missing ingredient is the integral

$$
\begin{aligned}
I_P(x,-x,x;i\omega_{P_1},i\omega_{P_2},i\omega_{P_3}) &= \int_0^\beta d\tau_{14}\int_0^{\tau_{14}} d\tau_{24}\int_0^{\tau_{24}} d\tau_{34}\, e^{i\omega_{P_1}\tau_{14}+i\omega_{P_2}\tau_{24}+i\omega_{P_3}\tau_{34}} e^{x(\tau_{14}-\tau_{24}+\tau_{34})} \\
&= +\int_0^\beta d\tau_{14}\int_0^{\tau_{14}} d\tau\int_0^{\tau_{14}-\tau} d\tau'\, e^{(i\omega_{P_1}+i\omega_{P_2}+i\omega_{P_3}+x)\tau_{14}-i(\omega_{P_2}+\omega_{P_3})\tau} e^{-(i\omega_{P_3}+x)\tau'} \\
&= +\frac{1}{i\omega_{P_3}+x}\frac{1}{-i\omega_{P_2}+x}\left[\frac{1}{i\omega_{P_1}+x}\,\frac{1}{n(x)}+\beta\delta_{\omega_{P_1}+\omega_{P_2}}\right] \\
&\quad +\frac{1}{i\omega_{P_3}+x}\frac{1-\delta_{\omega_{P_2}+\omega_{P_3}}}{i(\omega_{P_2}+\omega_{P_3})}\left[\frac{1}{i\omega_{P_1}+x}-\frac{1}{i(\omega_{P_1}+\omega_{P_2}+\omega_{P_3})+x}\right]\frac{1}{n(x)} \\
&\quad +\delta_{\omega_{P_2}+\omega_{P_3}}\frac{1}{i\omega_{P_3}+x}\left\{\left[\frac{1}{(i\omega_{P_1}+x)}\right]^2\frac{1}{n(x)}-\beta\left[\frac{1}{(i\omega_{P_1}+x)}\right]\frac{1-n(x)}{n(x)}\right\}.
\end{aligned}
$$

where $x=\pm U/2$, depending on the permutation. Summing up all terms we obtain the final expression for $\omega_m=0$. Setting $y=U/2$ we have in total [11]

$$
\begin{aligned}
\sum_{\sigma\sigma'}\sigma\sigma'\,\chi^{n,n'}_{i\sigma\sigma,i\sigma'\sigma'}(0) =& M_{n'}\frac{dM_n}{dy}+M_n\frac{dM_{n'}}{dy}-\beta n(y)\left[\delta_{n,n'}+\delta_{n,-n'}\right]\frac{dM_n}{dy}+\beta n(-y)M_nM_{n'} \\
&-\frac{1}{y}\left\{M_{n'}-\beta\left[n(y)\delta_{n,-n'}-n(-y)\delta_{n,n'}\right]\right\}M_n
\end{aligned}
\tag{129}
$$

where

$$
M_n=\frac{1}{i\nu_n-y}-\frac{1}{i\nu_n+y}. \tag{130}
$$

The finite frequency term (not given here) vanishes once we sum over n,n'.

References

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

[2] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989); Z. Phys. B **76**, 211 (1989); Int. J. Mod. Phys. B **3**, 2169 (1989)

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

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

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

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

[7] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (Eds.): *DMFT: From Infinite Dimensions to Real Materials*, Reihe Modeling and Simulation, Vol. 8 (Forschungszentrum Jülich, 2018) http://www.cond-mat.de/events/correl18

[8] E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008)

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

[10] A. Flesch, E. Gorelov, E. Koch, E. Pavarini, Phys. Rev. B **87**, 195141 (2013)

[11] A. Kiani and E. Pavarini, Phys. Rev. B **94**, 075112 (2016)

[12] E. Pavarini, A. Yamasaki, J. Nuss, and O.K. Andersen, New. J. Phys **7**, 188 (2005)

[13] N. Marzari *et al.*, Rev. Mod. Phys. **84**, 1419 (2012)

[14] V. Slatić and B. Horvatić, Sol. Stat. Comm. **75**, 263 (1990)

[15] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann and O. Parcollet, Phys. Rev. B **84**, 075145 (2011)

[16] J. Musshoff, G. Zhang, E. Koch, and E. Pavarini, Phys. Rev. B **100**, 045116 (2019)

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

[18] E. Pavarini *et al.*, Phys. Rev. Lett. **92**, 176403 (2004)

[19] J. Musshoff, A. Kiani and E. Pavarini, Phys. Rev. B **103**, 075136 (2021)

9 DFT+DMFT for Oxide Heterostructures

Frank Lechermann
Institut für Theoretische Physik III
Ruhr-Universität Bochum
44780 Bochum

Contents

1 Introduction **2**

2 DFT+DMFT in a nutshell **3**

3 Band-insulator/band-insulator heterostructure **6**
3.1 $LaAlO_3/SrTiO_3$: defect-free interface 6
3.2 $LaAlO_3/SrTiO_3$: oxygen-deficient interface 7
3.3 Oxygen-deficient $SrTiO_3$ surface 9

4 Mott-insulator/band-insulator heterostructure **10**
4.1 $LaTiO_3/SrTiO_3$ 10
4.2 δ-doping of titanate Mott insulators 11

5 Natural oxide heterostructures: delafossites **12**
5.1 Basic DFT characterization 13
5.2 Impact of correlation effects 15
5.3 Theoretical Mott design 19

6 Further systems **23**

7 Concluding remarks **23**

1 Introduction

The physics of transition-metal (TM) oxides is an ongoing highlight topic in condensed matter research, with its origins dating back to the earliest days of the quantum description of the solid state. The phenomenology spans from the various manifestations of basic electronic structure theory, i.e., conventional metallicity and band-insulating regimes, to the peculiar realizations of strong correlation mechanisms, as exemplified by, e.g., intriguing magnetic orderings and the discovery of unconventional superconducting phases. Beyond nature's original conception, the last 20-30 years witnessed an enormous boost in the fabrication of TM oxides, resulting in the nearly free design of nano-architectural combinations of individual oxide compounds (e.g. [1–4]).

This engineering path becomes especially interesting based on the fact that deviations from the conventional band-theory picture of weakly-interacting lattice electrons are most often encountered in bulk oxides with partially-filled TM(d) shells. Therefore by heterostructuring these materials, a unique playground for investigating, tailoring and eventually designing correlation effects may be opened. The challenges appear twofold in this context. First, *known* bulk-oxide correlation phenomena are transferred into the heterostructure environment and become tunable by various means. This, e.g., applies to magnetic, superconducting or further transport properties. Second, by interfacing different electronic bulk phases, *new* interface phases may emerge, possibly even without a distinct bulk analogon. A rather strict, obvious separation between these both research directions is albeit delicate.

In the first challenge, the proximity to an interface can enable, e.g., different structural/geometrical relaxations/constraints, symmetry breakings due to layering (cf. Fig. 1) or polarization effects because of a heterostructure-adapted electric field. Already a lot of work has been performed in this context and Refs. [5–7] may serve as overviews. For a concrete example, the physics of doped Mott insulators is a key research field in strongly correlated condensed matter. Bulk doping however poses many difficulties in view of a well-defined theoretical description. Most importantly, the intertwining with disorder mechanisms in the electronic and structural sector often hinders a straightforward modelling. Those problems may be overcome in Mott-oxide heterostructures, since electron-, hole-, or structural (i.e. effects only due to the different ion size of the valence-identical dopant) doping can nowadays be realized by respective doping layers. In this respect, the correlated doping physics becomes accessible in a well-defined manner by model-Hamiltonian and/or first-principles techniques. In general, accompanied by this progress, the examination of the influence of defects in strongly correlated materials has gained renewed interest, being explored by considering the detailed defect chemistry together with a state-of-the-art treatment of electronic correlations.

The second challenge is even more demanding and highlight examples in this regard might be provided by the physics of conducting quasi-two-dimensional electron interface systems (2DES) emerging in oxide heterostructures consisting of bulk band insulators. But the clarification of the uniqueness of such engineered electron phases beyond respective bulk counterparts is still a matter of frontline research. Furthermore, topological nontrivial electronic states are

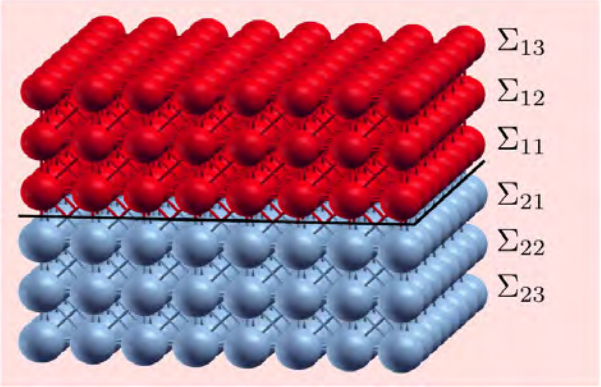


Fig. 1: *Principle issue of symmetry breaking induced by an interface between to oxide materials joint in an heterostructure architecture, leading to layer-dependent electronic self-energies* Σ_{ij}.

surely also to be considered in the oxide-heterostructure context, but the explicit interplay between topology and electronic correlations is still largely uncharted territory. In the following, these latter material possibilities will not be covered, but the interested reader finds ideas on this in a recent review [8].

The present brief overview on the first-principles many-body description of oxide heterostructures focusses on the applicative physics aspect, and is not intended to provide a thorough introduction to the DFT+DMFT formalism. The interested reader in view of this main theoretical/methodological aspect is referred to other chapters within the extended series of the Jülich School. We here provide only a very short summary on the key features of the DFT+DMFT framework in section 2. The sections 3, 4 and 5 deal with different applications and the specific physics of selected oxide heterostructure problems. Since the existing general theoretical work in this field is rather extensive, the given short review cannot provide a comprehensive and detailed survey. Instead, focus is here mainly on the corresponding work, directly associated with the author.

2 DFT+DMFT in a nutshell

The electronic density functional theory marks one cornerstone in condensed matter research, and will remain a key step in the atomistic investigation of matter. In essence, the problem of interacting electrons is mapped within DFT onto the problem of noninteracting particles within a complicated effective potential. In the present context it is vital to note that although DFT represents *in principle* a complete many-body account of interacting electrons, the most common Kohn-Sham representation based on the local-density or generalized-gradient approximations (i.e. LDA or GGA) brand this method as an effective single-particle approach. For condensed matter problems, it therefore describes electrons via band-theory arguments, and the original many-body effects due to the mutual Coulomb repulsion between electrons are cast into the aforementioned effective single-particle potential.

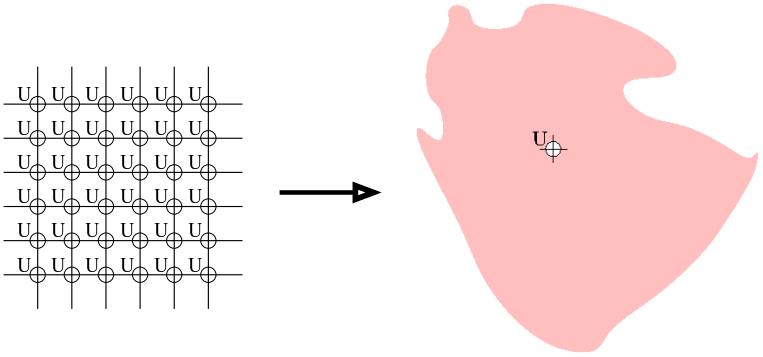


Fig. 2: *Sketch of the DMFT mapping of an interacting lattice problem (left) onto a problem of an impurity within an energy-dependent bath (right), assuming a sole onsite Coulomb interaction U, and hopping t between neighboring sites.*

Dynamical mean-field theory [9, 10] can be seen as the quantum many-body scheme for correlated electrons with the best compromise between generality, accuracy and performance. Notably, also DMFT describes a mapping, i.e., from the problem of interacting lattice electrons onto the problem of a quantum impurity embedded within a self-consistent energy-dependent bath, as sketched for the Hubbard model in Fig. 2.

For chemical potential μ, and Hamiltonian $H(\mathbf{k})$ at wave vector $\mathbf{k}$, the one-particle Green function reads

$$G(\mathbf{k}, i\omega_\mathrm{n}) = \big(i\omega_\mathrm{n} + \mu - H(\mathbf{k}) - \Sigma(\mathbf{k}, i\omega_\mathrm{n})\big)^{-1}, \tag{1}$$

where $\omega_\mathrm{n} := (2n+1)\pi T$ are fermionic Matsubara frequencies employed to emphasize the treatment at *finite* temperature. The analytical continuation to real frequencies ω in actual calculations may, e.g., be performed via the maximum entropy method (see e.g. [11, 12] for more details).

The local Green function is approximated in DMFT with the help of a k-independent impurity self-energy $\Sigma_\mathrm{imp}(i\omega_\mathrm{n})$, i.e.,

$$G^\mathrm{DMFT}_\mathrm{loc}(i\omega_\mathrm{n}) = \sum_\mathbf{k} \big(i\omega_\mathrm{n} + \mu - H(\mathbf{k}) - \Sigma_\mathrm{imp}(i\omega_\mathrm{n})\big)^{-1}, \tag{2}$$

and the corresponding impurity problem reads

$$\Sigma_\mathrm{imp}(i\omega_\mathrm{n}) = \mathcal{G}_0(i\omega_\mathrm{n})^{-1} - G_\mathrm{imp}(i\omega_\mathrm{n})^{-1}. \tag{3}$$

The so-called Weiss field $\mathcal{G}_0(i\omega_\mathrm{n})$ is a unique function of the local Hamiltonian (expressed within a localized basis) and the DMFT self-consistency condition implies $G_\mathrm{imp} = G^\mathrm{DMFT}_\mathrm{loc}$. The calculational loop is depicted in the 'DMFT loop' box of Fig. 3. Quantum-impurity solvers based, e.g., on quantum Monte Carlo, Exact Diagonalization, etc. yield the solution. For more details we refer to [12] for an early review. Note that local-interaction diagrams are included to all orders in this non-perturbative theory. The vital energy dependence of the Weiss field ensures the qualitatively correct description of low-energy quasiparticle (QP) features as well as high-energy incoherent (Hubbard) excitations.

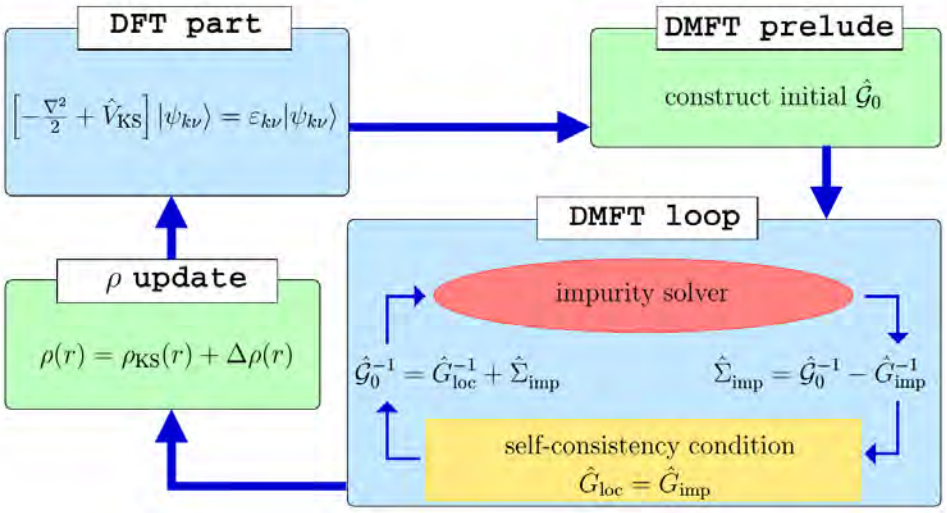


Fig. 3: *State-of-the-art charge self-consistent DFT+DMFT loop (after [21]). The calculation usually starts from a self-consistent Kohn-Sham solution. The correlated subspace is defined and the initial Weiss field $\mathcal{G}_0$ constructed. Afterwards, a single (or more) DMFT step is performed. The obtained self-energies are upfolded and an updated charge density $n(\mathbf{r})$ is computed. A new charge density implies a new Kohn-Sham potential, and a single new Kohn-Sham step is performed, therefrom a new Weiss field is generated, etc..*

In the hybrid scheme of DFT+DMFT [13, 14], the many-body part incorporates a Hubbard-model(-like) picturing of a suitably chosen *correlated subspace* [15, 16]. This correlated subspace is understood as a quantum-numbered real-space region where correlated electrons tend to reside. The key interfacing blocks of the complete DFT+DMFT self-consistency cycle [17–20] (cf. Fig. 3) are marked by the downfolding of the full-problem Bloch space to the correlated subspace, and the upfolding of the DMFT self-energy back to the original space. The main corresponding formulas for sites $\mathbf{R}$, local orbitals mm' and band indices $\nu\nu'$, read

$$G^{\mathbf{R},\mathrm{imp}}_{mm'}(i\omega_\mathrm{n}) = \sum_{\mathbf{k},(\nu\nu')\in\mathcal{W}} \bar{P}^{\mathbf{R}}_{m\nu}(\mathbf{k})\, G^{\mathrm{bloch}}_{\nu\nu'}(\mathbf{k},i\omega_\mathrm{n})\, \bar{P}^{\mathbf{R}*}_{\nu'm'}(\mathbf{k})\,, \tag{4}$$

$$\Delta\Sigma^{\mathrm{bloch}}_{\nu\nu'}(\mathbf{k},i\omega_\mathrm{n}) = \sum_{\mathbf{R},mm'} \bar{P}^{\mathbf{R}*}_{\nu m}(\mathbf{k})\, \Delta\Sigma^{\mathbf{R},\mathrm{imp}}_{mm'}(i\omega_\mathrm{n})\, \bar{P}^{\mathbf{R}}_{m'\nu'}(\mathbf{k})\,, \tag{5}$$

with $\bar{P}$ denoting the normalized projection between Bloch space and correlated subspace [15]. The object $\Delta\Sigma^{\mathrm{bloch}}_{\nu\nu'}$ describes the k-dependent self-energy in Bloch space after double-counting correction (which takes care of the fact that some correlations are already handled on the DFT level). Note that there is a choice for the range $\mathcal{W}$ of included Kohn-Sham bands in the downfolding. In the upfolding operation, the charge density will also be affected by correlation effects, i.e.,

$$\rho(\mathbf{r}) = \sum_{\mathbf{k},\nu\nu'} \langle\mathbf{r}|\Psi_{\mathbf{k}\nu}\rangle \Big(f(\tilde{\epsilon}_{\mathbf{k}\nu})\,\delta_{\nu\nu'} + \Delta N_{\nu\nu'}(\mathbf{k}) \Big) \langle\Psi_{\mathbf{k}\nu'}|\mathbf{r}\rangle \quad , \tag{6}$$

where Ψ denotes Kohn-Sham states, f the associated Fermi function and ΔN is the DMFT self-energy correction term [21, 15]. Note finally that this first-principles many-body scheme works, at heart, at finite temperature T. Electron states are therefore subject to the full thermal impact, beyond sole occupational Fermi-function modification.

In oxide heterostructures, as in various other multi-atom unit-cell problems, the correlated subspace invokes not only a single lattice site. In the case of symmetry-equivalent sites, the self-energy is determined for a representative site and transferred to the remaining sites via proper symmetry relations. A different impurity problem is defined for each symmetry-inequivalent site j through [22]

$$\mathcal{G}_0^{(j)}(i\omega_\mathrm{n})^{-1} = G^{(j)}(i\omega_\mathrm{n})^{-1} + \Sigma_\mathrm{imp}^{(j)}(i\omega_\mathrm{n}) \quad , \tag{7}$$

and the coupling is realized via the DFT+DMFT self-consistency condition invoking the computation of the complete lattice Green function.
If not otherwise stated, the materials investigations discussed in sections 3, 4 and 5 involving the author were performed using charge-selfconsistent DFT+DMFT based on a mixed-basis pseudopotential code [23] and hybridization-expansion continuous-time quantum Monte Carlo [24] in the TRIQS package [25] as an impurity solver. For more technical details on the implementation the reader is referred to Ref. [20].

3 Band-insulator/band-insulator heterostructure

An emerging metallic two-dimensional electron system (2DES) at the n-type interface between band-insulating $LaAlO_3$ (LAO) and $SrTiO_3$ (STO) discovered by Ohtomo and Hwang [2] marks one of the most appreciated discoveries in the oxide-heterostructure context. A polar-catastrophe mechanism [26] is believed to be dominantly at the root of the 2DES establishment. In terms of electronic correlations, this and related interface systems between band insulators [27] appear quite subtle since the basic constituents do not already host strongly correlated electrons in partially-filled d- or f-states. However, the follow-up experimental studies revealed, e.g., the possibility for superconductivity [28] as well as magnetic order [29] in the LAO/STO interface, which may be fingerprints for the relevance of electronic correlations. Furthermore, defect-related physics is believed to be important in these interfaces, which can directly induce local-based self-energy effects beyond the effective single-particle realm.

3.1 $LaAlO_3$/$SrTiO_3$: defect-free interface

Interface metallicity is easily achieved within DFT for the stoichiometric defect-free (DF) LAO/STO (see Fig. 4b,c). Two electrons occupy the dominant Ti$(3d)$ low-energy manifold, matching the number for the polar-catastrophe avoidance, which predicts $Ti^{3.5+}O_2^{4-}$ at the interface [26]. Figure 4b shows that the electrons which form the 2DES are confined to the STO part, with dominant localization near $Ti^{(12)}$, i.e., directly at the boundary towards the LAO part. There, the major orbital Ti$(3d)$ character is of d_{xy} kind.
Note that although there is an apparent effect of larger onsite Hubbard interactions on the quasiparticle formation/shape as visualized in Fig. 4d, sizable spectral-weight transfer to higher energies far from the Fermi level remains absent in the DF case. This is understandable from the highly fractional filling of the interface Ti-d_{xy} orbitals. Note also that in either case, DFT or DFT+DMFT level, the DF system remains far from a ferromagnetic (FM) instability.

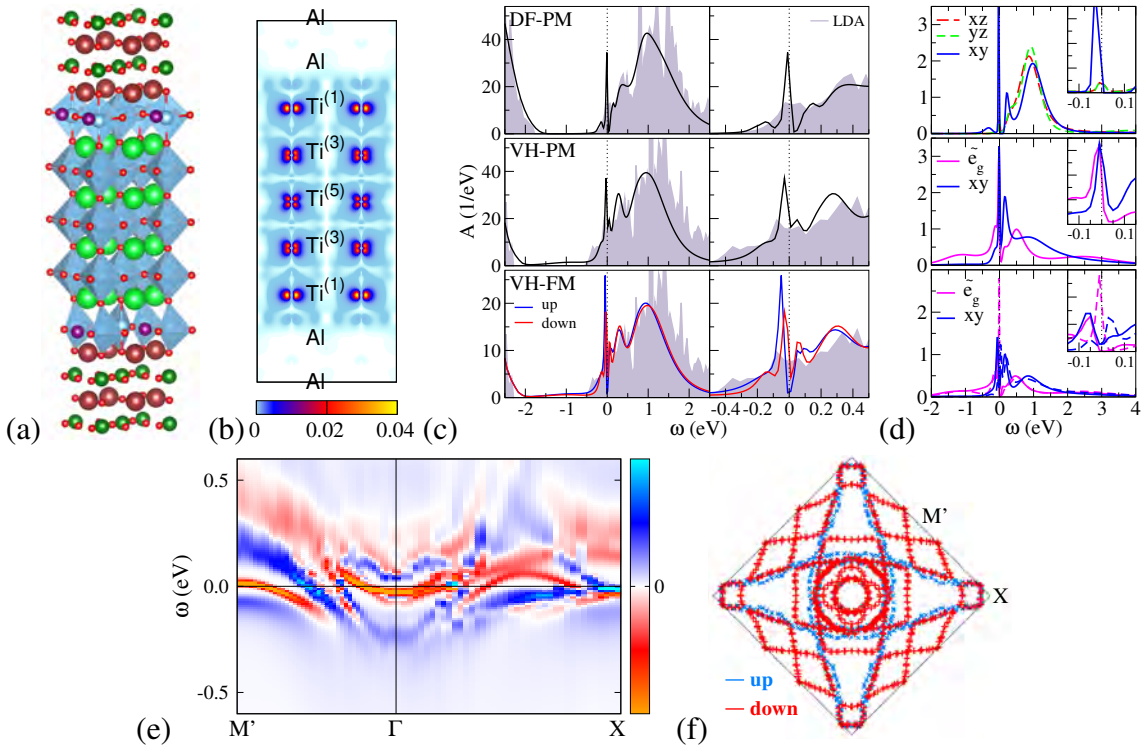


Fig. 4: *Influence of oxygen vacancies in the LAO/STO interface based on DFT+DMFT calculations for a dense-defect scenario (after [30]). (a) 80-atom superlattice, La (brown), Al (dark green), Sr (green), Ti (blue), O (red), OV (violet). (b) DFT valence charge density of the occupied low-energy manifold in the DF structure. (c) Total spectral function, left: larger window, right: smaller window. Top: defect-free paramagnetic (DF-PM), middle: vacancy-hosting paramagnetic (VH-PM) and bottom: vacancy-hosting ferromagnetic (VH-PM). (d) Local spectral function for selected effective Ti(3d) states, vertical ordering as in (c). (e) spin contrast of the k-dependent spectral function $A(\mathbf{k}, \omega)$ in the ferromagnetic phase. (f) FM Fermi surface.*

3.2 $LaAlO_3/SrTiO_3$: oxygen-deficient interface

While the clean interface hosts the polar-catastrophe scenario but lacks strong (local) electron correlation effects, things change when allowing for defects at or near the interface. For instance, relevant local correlations are evident from the co-appearance of oxygen vacancies and ferromagnetism in LAO/STO [31]. Point defects are in fact an important ingredient of the interface physics between band-insulating systems. In this regard, oxygen vacancies (OVs) so far appear to have the major impact [32]. In a DFT+DMFT study of the LAO/STO interface [30] it is shown that stable ferromagnetic order builds up on both, OVs *and* electronic correlations. While Ti-$3d(t_{2g})$ orbitals dominate the states directly above the STO gap, an oxygen vacancy leads to an in-gap state of Ti-$3d(e_g)$ kind, here termed $\tilde{e}_g$. In a minimal model, the correlated electronic structure at the interface may be described by the interplay between $\tilde{e}_g$ and an in-plane xy orbital from the t_{2g} threefold (see Fig. 4). A reduced Hubbard $U = 2.5\,\mathrm{eV}$ and Hund's exchange $J_\mathrm{H} = 0.5\,\mathrm{eV}$ is appropriate to elucidate this physics in the smaller correlated subspace. Note that a dense-defect scenario is assumed in that superlattice assessment, i.e., there is an OV at every other O site in the interface TiO_2 plane. Still, an in-gap $\tilde{e}_g$-like state at $\varepsilon_\mathrm{IG} \sim -1.2\,\mathrm{eV}$ is well reproduced in agreement with experiment [33]. This finding was also verified by even

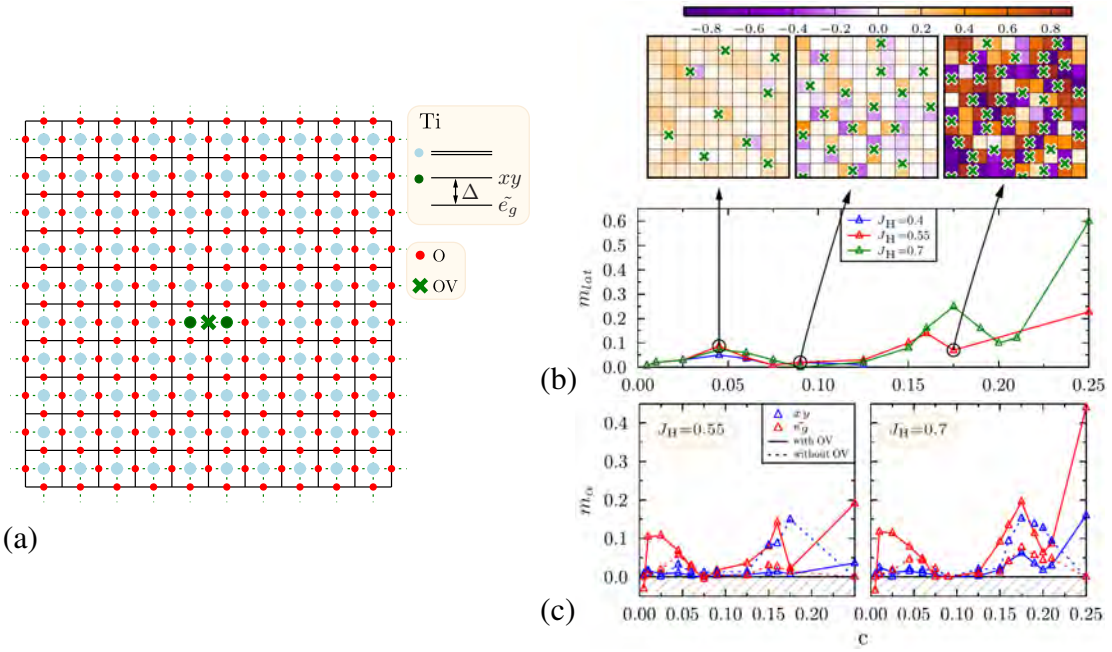


Fig. 5: *Influence of oxygen vacancies in the LAO/STO interface (after [35]). (a) Model setting:* 10×10 *square lattice, two orbitals per site, oxygen degrees of freedom are integrated out, nearest-neighbor hopping* $t = 0.2\,eV$*. A crystal-field splitting* $\Delta = 0.3\,eV$ *between* $\tilde{e}_g$ *and* xy *is applied, if an OV is located nearby. (b) Top: color-coded ordered local magnetic moments on the lattice for selected OV concentration. Bottom: lattice net moment per site for different Hund's exchange* $J_{\rm H}$*. (c) Orbital-resolved magnetic moment, averaged per site.*

larger supercell calculations [34]. Close to an oxygen vacancy, the formal oxdiation state of titanium is close to Ti^{3+}, i.e., locally a $3d^1$ occupation is realized by the appearance of defects. But the emergent spin polarization in the given dense-defect limit in not of purely local kind, it develops substantial dispersive behavior in reciprocal space (cf. Fig. 4d,e). In general, these defect-based studies show that strong electron correlations, describable within DFT+DMFT, may be introduced also in band-band insulator architectures of oxide heterostructures.

Approaching significantly lower OV concentrations in a first-principles manner asks for the handling of much larger supercells. Due to the numerical heaviness of DFT+DMFT, this is yet not easily achievable. Therefore instead, a model-Hamiltonian approach equipped with the relevant ingredients from the dense-defect limit, appears more adequate. Figure 5 displays the setting and some main results of such an ansatz [35]. A two-orbital $\tilde{e}_g$-xy Hubbard model is solved on a 10×10 square lattice resembling the TiO_2 interface plane. The efficient rotational-invariant slave boson (RISB) scheme [36, 37], employing a self-energy which has a linear-in-frequency term and a static contribution, is put into practice for a simplified treatment of the correlation problem. Focussing on the magnetic order, it can be shown that there are three regimes with growing number of OVs. At very small concentration, a Ruderman-Kittel-Kasuya-Yoshida (RKKY) coupling leads to FM order, whereas at larger concentration a double-exchange mechanism dominates a different FM phase. In-between local antiferromagnetic (AFM) pairs (or, in an advanced self-energy modeling, possibly cluster singlets) result in a nearly absent net magnetic moment. This intricate and highly nontrivial OV-dependent magnetic exchange is in line with experimental findings of a strongly probe-dependent magnetic response.

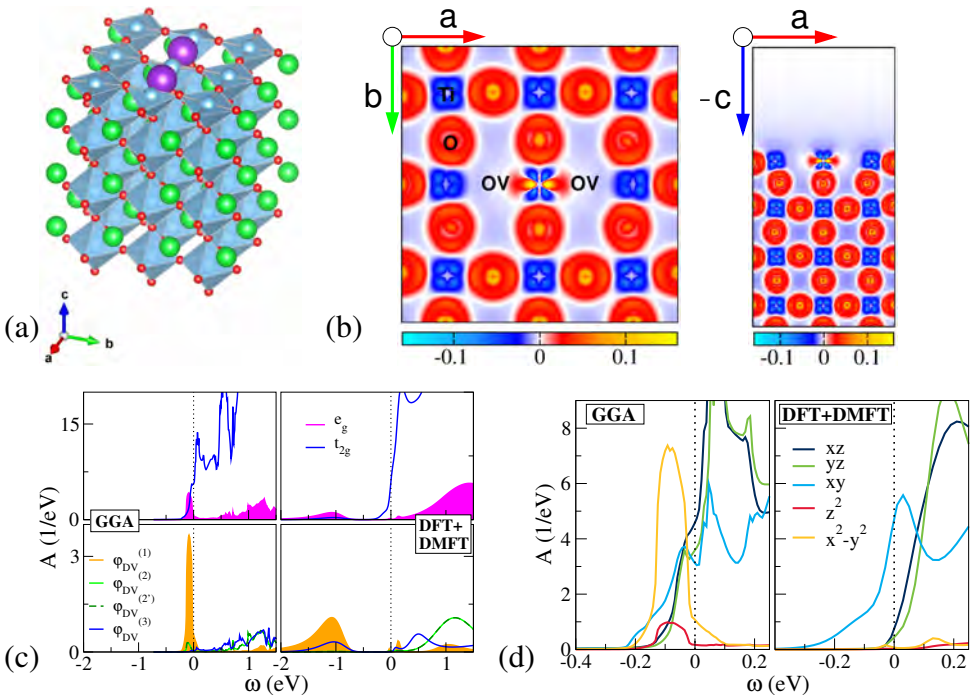


Fig. 6: *Influence of a double vacancy in the $SrTiO_3$ (001) surface (after [38]). (a) 180-atom supercell with a TiO_2-terminated surface layer of STO: Sr (green), Ti (blue), O (red), OV (violet). (b) Bond charge density $n_{\mathrm{total}}(\mathbf{r}) - n_{\mathrm{atomic}}(\mathbf{r})$: top view (left), side view (right). (c) Spectral-weight comparison of the summed Ti(t_{2g},e_g) (top) and of the dominant Ti-based effective orbital between DFT(GGA) (left) and DFT+DMFT (right). The $\varphi^{(1)}_{\mathrm{DV}}$ Wannier-like orbital is an effective inplane e_g orbital located on the Ti site between the OVs. (d) Low-energy spectral-weight comparison of the Ti($3d$) states.*

3.3 Oxygen-deficient $SrTiO_3$ surface

In parallel to the STO-based heterostructure investigations, studying the $SrTiO_3$ surface attracted significant attention. Interface and surface properties of a chosen oxide are often related and the comparison between both planar-extended defects to the bulk enables a better understanding of emergent phenomena. A 2DES was initially also found on the oxygen-deficient (001) surface [39, 40] and soon after also confirmed for other cleavage planes, e.g., in (111) direction [41]. As a difference to the LAO/STO interface, the defect-free STO surface is believed to be insulating. Due to a missing interface-driven polar-catastrophe mechanism, again defects such as OVs are essential to metallize the surface. Similar to the interface spectrum, an in-gap state at a very similar position, i.e. $\sim\!-1.3$ eV, has been detected on the STO surface early on [42]. Recent DFT+DMFT considerations of the STO(001) surface with OVs indeed verified this in-gap state [38], which is again dominantly formed by Ti-$3d(e_g)$ weight. (cf. Fig 6). Furthermore, the low-energy structure dominated by Ti-$3d(t_{2g})$ states is also in accordance with experiment. A double-vacancy defect provided the best matching with experiment, however only two distinct vacancy configuration were examined because of the large numerical effort within DFT+DMFT. Nonetheless, the spectral separation of e_g at high energy and t_{2g} is a clear generic feature of the study. Note that conventional DFT is not sufficient to provide this orbital separation (cf. Fig. 6c,d).

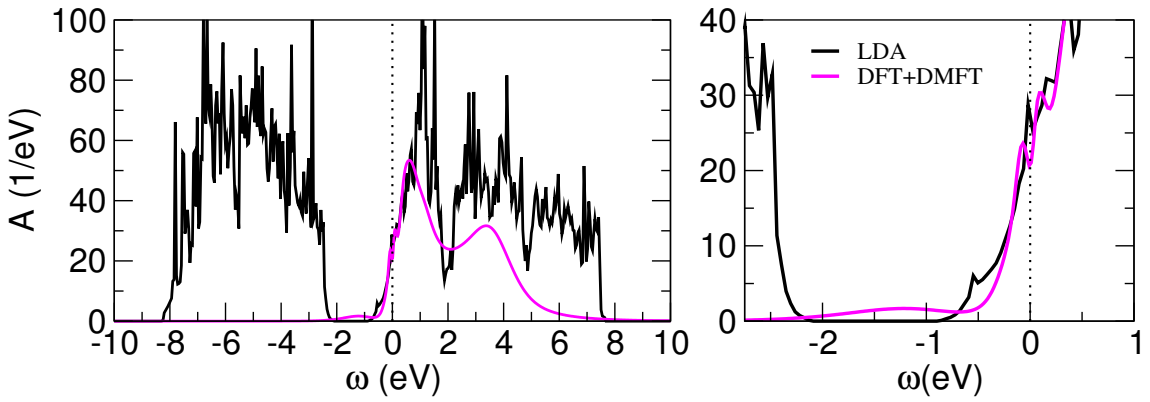


Fig. 7: *Spectral information for a $LaTiO_3$/$SrTiO_3$ superlattice with 4 SrO and 4 LaO layers (after [43]). Left: total spectrum, right: closer to Fermi level.*

4 Mott-insulator/band-insulator heterostructure

4.1 $LaTiO_3$/$SrTiO_3$

Oxide heterostructures composed of a Mott insulator and a band insulator have also been studied early on, initially via experimental work on the $LaTiO_3$/$SrTiO_3$ (LTO/STO) interface [44]. Contrary to the previous combination of pure band insulators, the Mott-insulating part of course injects the correlated electron problem into the heterostructure architecture right from the start. While STO is an ideal cubic perovskite at ambient temperature, LTO marks a distorted perovskite with orthorhombic crystal symmetry. Initial model-Hamiltonian Hartree-Fock studies of a Hubbard-model application to such an interface yielded an intricate phase diagram with the number of La-layers and Hubbard U [45]. Simplified realistic DMFT for LTO/STO [46] emphasized the structural orthorhombic-versus-tetragonal aspect of the LTO part. Transport within a Mott-band insulator heterostructure has also been studied again within a model-Hamiltonian approach [47].

A superlattice DFT+DMFT investigation of LTO/STO [43] revealed the realistic competition of both insulating systems stacked along the c-axis, giving rise to a metallic interface state. The correlated subspace can be chosen to be spanned by Ti-$3d(t_{2g})$ states, i.e., consists locally of three correlated orbitals. A Hubbard $U = 5\,\mathrm{eV}$ and Hund's exchange $J_\mathrm{H} = 0.7\,\mathrm{eV}$ is applied in that subspace. Orbital-dependent charge transfers lead to a strong Ti-$3d(xy)$, i.e., inplane electronic polarization in the interface TiO_2 layer. Surely, on the STO side the Ti^{4+} oxidation state with nominal $3d^0$ occupation is quickly reached, while on the LTO side the Ti^{3+}-$3d^1$ establishes and a sizably lower Hubbard band at $\sim\!-1.1\,\mathrm{eV}$ is identified in the spectral function (see also Fig. 7). In those calculations, the lattice constant was fixed to the cubic STO value, but local structural relaxations on the DFT level were allowed. It is to be noted that the lattice degrees of freedom are an important aspect in oxide heterostructures. Simplified DFT+DMFT bulk-like studies revealed, e.g., the impact of strain on the Mott-insulating state of $LaTiO_3$ [48] and $LaVO_3$ [49].

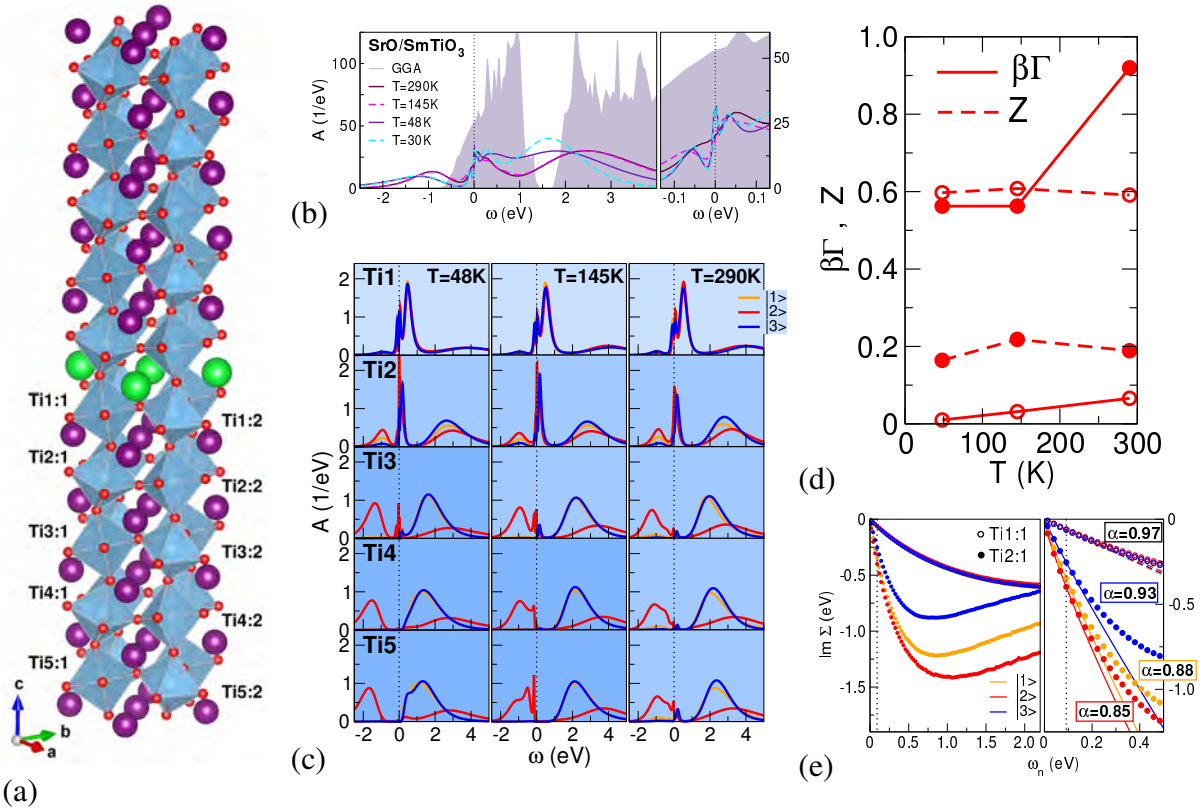


Fig. 8: *Paramagnetic DFT+DMFT data of δ-doping $SmTiO_3$ with a SrO monolayer (after [50]). (a) 100-atom unit cell in a superlattice architecture: Sr (green), Sm (purple), Ti (blue), O (red). Two inequivalent Ti sites are handled in each TiO_2 layer. Structural relaxations are performed on the DFT(GGA) level. (b) Total spectral function at different temperatures. (c) Layer-, orbital- and T-resolved spectral function. The correlated subspace consists of three effective t_{2g} orbitals at each Ti site. (d) QP weight Z and dimensionless electron-electron scattering rate $\beta\Gamma = -Z\,\mathrm{Im}\,\Sigma(i0^+)$ for the dominant state $|2\rangle$ in the first (open circles) and second (filled circles) TiO_2 layer. (e) Orbital-resolved imaginary part of the self-energy on the Matsubara axis. Left: larger frequency range, right: low-frequency region with fitting functions $\mathrm{Im}\,\Sigma(\omega_\mathrm{n}) = C_0 + A\,\omega_\mathrm{n}^\alpha$ (dashed/full lines). Exponential-fitting cutoff n_c is denoted by the dotted line. Values $\alpha = 1$ and $C_0 = 0$ mark the Fermi-liquid regime.*

4.2 δ-doping of titanate Mott insulators

Even more intriguing physics may be found when starting directly from a doped-Mott state within the heterostructure setting. Motivated by experimental work [51–53], a first-principles many-body investigation of δ-doping the rare-earth titanates $LaTiO_3$, $GdTiO_3$ and $SmTiO_3$ with a single SrO layer was undertaken [54, 50]. Especially the δ-doped $SmTiO_3$ case displays puzzling physics in experiment, namely NFL transport that switches to FL-like characteristics upon adding further SrO layers. In the rare-earth titanate $3d(t_{2g}^1)$ series of distorted-perovskite Mott insulators, the magnetic low-temperature state changes from AFM to FM with the size of the rare-earth ion. The samarium titanate is still AFM, but just on the border towards ferromagnetism. Structurally well-defined hole doping introduced by the SrO monolayer renders $SmTiO_3$ metallic. The DFT+DMFT results (see Fig. 8) reveal significant spectral-weight trans-

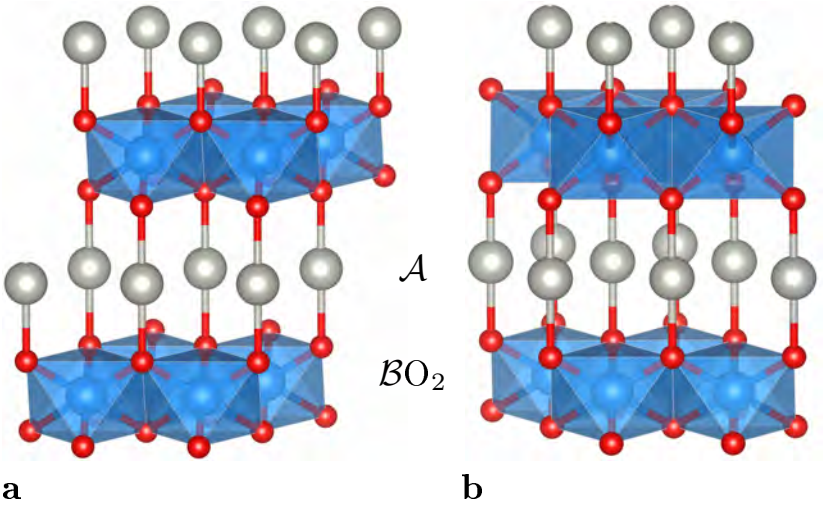


Fig. 9: *Delafossite $\mathcal{AB}O_2$ crystal structure. (***a***) $R\bar{3}m$ and (***b***) $P6_3/mmc$ symmetry. $\mathcal{A}$ (grey), $\mathcal{B}$ (blue) and O (red). From Ref. [57].*

fer to higher energies compared to DFT, and in addition a complex layer-resolved electronic structure. While far from the doping layer the system resides in an orbital-polarized Mott-insulating state, it is conducting in an orbital-balanced manner just at the interface TiO_2 layer (cf. Fig. 8c). Both regimes are joined by an orbital-polarized doped-Mott layer with a largely renormalized QP peak at the Fermi level. A detailed analysis of the layer- and orbital-dependent self-energies shown in Fig. 8d,e indeed reveals signatures of a non-Fermi-liquid (NFL) exponent for the dominant effective t_{2g} orbital in the second, i.e., orbital-polarized doped-Mott, TiO_2 layer. Further investigations hint towards competing AFM-FM fluctuations in the Mott-critical zone as a possible cause for NFL behavior [50]. A pseudogap(-like) structure in the theoretical spectral function subject to such fluctuations has indeed been identified in experimental studies [55]. In an extension of this study, it was shown that the addition of further SrO layers establishes an extra band-insulating regime in the formed $SrTiO_3$-like region, with a stronger inplane xy-polarized metallic layer at the boundary [56].

5 Natural oxide heterostructures: delafossites

Delafossites, named after the french crystallographer Gabriel Delafosse (1796-1878), are oxides of type $\mathcal{AB}O_2$, where $\mathcal{A}$ and $\mathcal{B}$ denote different metallic elements. Their unique crystal structure (see Fig. 9) consists of an alternate stacking of triangular $\mathcal{A}$ lattices and planes of edge-sharing $\mathcal{B}O_6$ octahedra along the c-axis. Importantly these two different layer types are connected via oxygen in a so-called dumbbell position. There are two possible stacking types, namely the more common rhombohedral one with $R\bar{3}m$ space-group symmetry and the hexagonal one giving rise to $P6_3/mmc$ symmetry. The metallic ions are in the formal oxidation state $\mathcal{A}^+$ and $\mathcal{B}^{3+}$, respectively.

Delafossites split into a larger insulating and a smaller metallic class of compounds. In a series of papers [58–60], Shannon *et al.* in 1971 described the novel synthesis and single-crystal growth of several delafossites with $\mathcal{A}$=Pd, Pt and Ag. Among those, there are oxides with

exceptionally high electrical conductivity at room temperature, e.g., $PdCoO_2$ and $PtCoO_2$, in combination with an outstanding single-crystal purity. This combined feature of such delafossites has started to become an intense field of research in recent years (see e.g. Refs. [61, 62] for reviews). That specific group of delafossites in terms of metallic properties, includes also the $PdCrO_2$ compound, which, among further challenging physics, hosts Mott-insulating CrO_2 layers [63–65].

This brings us to a very relevant aspect in view of our general heterostructure context. The delafossite architecture gives rise to a natural heterostructure, in which individual layers may attain a distinct character of their own. In most known layered materials, e.g., cuprates, cobaltates, etc., there is usually one 'active' layer and the remaining part mainly provides the glue. However in delafossites, e.g., the $\mathcal{A}$ layer can manage the metallic transport, while the $\mathcal{B}O_2$ layers account for the magnetic ordering. This not only entails exciting physical processes in the pure compound, but furthermore possibly allows for a kind of 'meta oxide-heterostructure' physics upon additional (nano-)engineering.

5.1 Basic DFT characterization

Let us first focus on a DFT assessment of selected Co- and Cr-based delafossites. Figure 10 displays the band theoretical results for the spectrum of $PdCoO_2$, $PdCrO_2$ and $AgCrO_2$, namely density of states (DOS) and band structures. For $PdCoO_2$ we also provide plots of the Wannier-like Pd$(4d)$ orbitals. In that compound, the Pd$(4d)$ states are largely occupied with a bandwidth (W) hierarchy of $W_{a_{1g}} > W_{e'_g} > W_{e_g}$. As shown in Fig. 10b, the Co$(3d)$ weight is mostly located in the bands close to and above the Fermi level $\varepsilon_{\rm F}$, with a single band crossing $\varepsilon_{\rm F}$. The latter dispersion, denoted in the following 'cPd', is dominantly of mixed Pd$(4d)$ and partial Co$(3d)$ kind. Note that the e'_g orbitals are the ones with the strongest in-plane character (see Fig. 10b). Though the band-filling Co$(3d)$ character resembles an original Co$(3d^6)$ picture, from the hybridizations at the Fermi level a completely inert Co-t_{2g} subshell is not truly justified. The DFT fermiology and dispersions at low energy are in good agreement with data from angle-resolved photoemission spectroscopy (ARPES) measurements [66] and de Haas-van Alphen studies [67]. Thus plain DFT seemingly provides already an adequate description of key $PdCoO_2$ features.

Contrary to $PdCoO_2$, the $\mathcal{B}$-site states of Cr$(3d)$ character are much less filled, the three t_{2g}-dominated bands are right at the Fermi level, and the Pd$(4d)$ character at $\varepsilon_{\rm F}$ is minor (see Figs. 10c, d). This however strongly disagrees with available experimental data from ARPES [68, 63] and quantum oscillations [69, 70]. In experiment, there is also only a single band crossing the Fermi level, quite similarly as in $PdCoO_2$. This discrepancy is due to the neglect of strong electronic correlations in conventional DFT, which misses the Mott-localized character of the CrO_2 layers. Partial agreement with experiment concerning the dispersions can be achieved within spin-polarized DFT [71, 68, 63, 72], accounting also for the magnetic ordering at low temperatures. But this Slater-type handling of the Cr$(3d)$ states is not describing the underlying physics comprehensively.

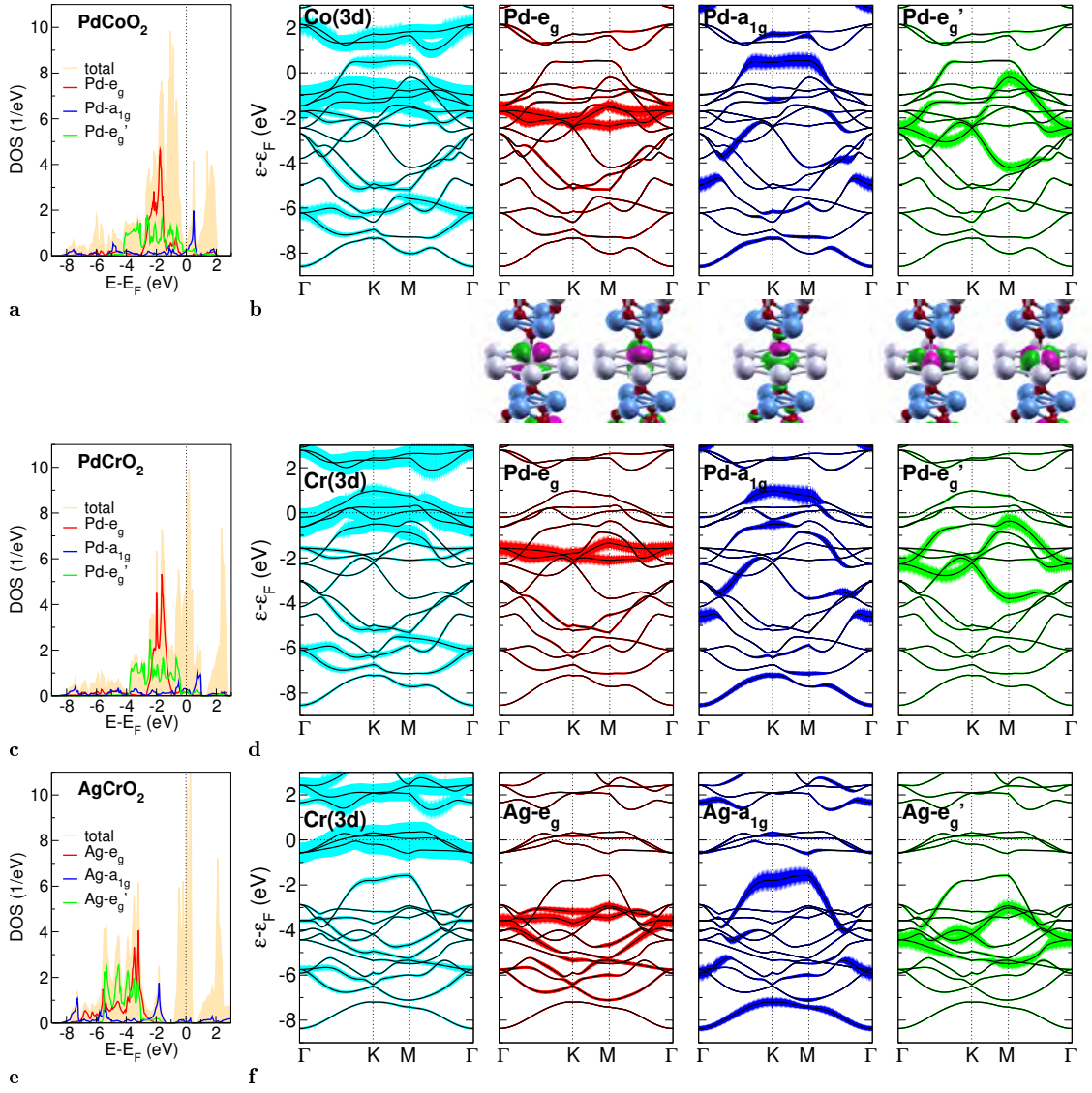


Fig. 10: *DFT electronic structure of $PdCoO_2$, $PdCrO_2$ and $AgCrO_2$. (a) Total and orbital-resolved $\mathcal{A}$-site$(4d)$ density of states. (b) Band structure along high symmetry lines in the $k_z = 0$ plane, with fatbands marking the $\mathcal{B}$-site$(3d)$ and the orbital-resolved $\mathcal{A}$-site$(4d)$ character. Additionally, the corresponding real-space Pd$(4d)$ projected local orbitals are provided for the case of $PdCoO_2$: Pd (grey), Cr (light blue) and O (red). (c, d) Same as (a, b) but for $PdCrO_2$. (e, f) Same as (a, b) but for $AgCrO_2$. From Ref. [57].*

Finally, the $AgCrO_2$ compound would be insulating in DFT if the Cr-t_{2g} states were not located again at the Fermi level (see Figs. 10e, f). The Ag$(4d)$ states are filled and would give rise to a band insulator. The missing correlation effects on Cr become most evident here. The prominent Ag-a_{1g} dominated band just below the Cr-t_{2g} bands in energy and with a nearly flat dispersion along K-M bears striking resemblance to the former low-energy cPd band in $PdCoO_2$. In fact as we will see in the following, this present band will just form the highest valence band in true $AgCrO_2$ once correlations are properly included. Furthermore for the same reason, the akin band in $PdCrO_2$ (yet there with stronger Pd-e_g' character) will be shifted to $\varepsilon_{\rm F}$.

5.2 Impact of correlation effects

For the charge self-consistent DFT+DMFT calculations the correlated subspace is build up from the five effective $\mathcal{B}$-site Wannier-like $3d$ functions as obtained from the projected-local-orbital formalism [15, 16], using as projection functions the linear combinations of atomic $3d$ orbitals which diagonalize the $\mathcal{B}$-site $3d$ orbital-density matrix. A five-orbital Slater-Kanamori Hubbard Hamiltonian, i.e., including density-density, spin-flip and pair-hopping terms, is utilized in the correlated subspace, parametrized by a Hubbard U and a Hund exchange $J_{\rm H}$. The $\mathcal{B}$ site will here be either of Co or Cr type and a value of $J_{\rm H} = 0.7\,$eV is proper for TM oxides of that kind. Concerning the Hubbard interaction, the value $U = 3\,$eV is assigned to the $3d$ states of Co and Cr in $PdCoO_2$ and $PdCrO_2$, respectively [73, 74]. For Cr$(3d)$ in $AgCrO_2$, the somewhat larger value of $U = 4\,$eV is used in order to comply with the weaker screening because of the (nearly) filled Ag$(4d)$ shell. Note that no further Hubbard interactions are assigned to the $\mathcal{A}$ site. The d orbitals on those sites are here of $4d$ kind, of formal d^9 filling and only weakly hybridizing with oxygen. Thus by any means, Coulomb interactions are expected much smaller than on the $\mathcal{B}$ site. Spin-orbit coupling is neglected in the crystal calculations. Though the $\mathcal{A}CrO_2$ delafossites order antiferromagnetically at low temperatures within the CrO_2 layers, the investigations remain at still higher T assuming paramagnetism for all studied cases. If not otherwise stated, the system temperature is set to $T = 290\,$K. Further details on the calculational settings can be found in Ref. [57] and see also Refs. [64, 75].

Figure 11 exhibits the spectral DFT+DMFT summary for the three compounds. In the case of the Co compound, the changes compared to DFT appear minor, as already expected from the simplest picture of a closed Co-t_{2g} subshell. The dispersions, which now describe true quasiparticle features, are hardly modified at lower energy. Quite on the contrary, the QP dispersion for $PdCrO_2$ has changed dramatically (see Fig. 11b); the DFT-original Cr bands at $\varepsilon_{\rm F}$ have disappeared and instead, a single cPd dispersion as in $PdCoO_2$ crosses the Fermi level. This result brings theory eventually in line with experimental findings [68, 63, 69, 70]. Also for $AgCrO_2$, the DFT+DMFT approach settles the comparison with experiment, namely by identifying the insulating nature with a compatible gap of $\sim 1.8\,$eV. While the latter delafossite shows of course no Fermi surface, the fermiology of $PdCoO_2$ and $PdCrO_2$ in Figs. 11d,e becomes rather similar with interactions. A single-sheet interacting Fermi surface, comprising a single electron, is centered around Γ and has a hexagonal shape with some warping. Note that this warping is somewhat stronger in the case of the Co compound.

Two functions are provided to discuss the $\mathbf{k}$-integrated spectra (see Fig. 11h-j). First, the site- and orbital-projected spectral function $A_{\rm proj}(\omega)$, defined by projecting the Bloch-resolved spectral function $A_\nu(\mathbf{k},\omega)$ with Bloch index ν onto a chosen site-orbital and summing over $\nu, \mathbf{k}$. Note that this function is comparable but strictly not identical to the local spectral function $A_{\rm loc}$, which is obtained from analytical continuation of the local Green function. Second, it proves instructive to also plot directly $A_\nu(\omega)$, i.e., the $\mathbf{k}$-integrated Bloch-resolved spectrum. This allows us to trace the behavior of the former DFT bands upon interaction and displays the QP formation originating in Bloch space.

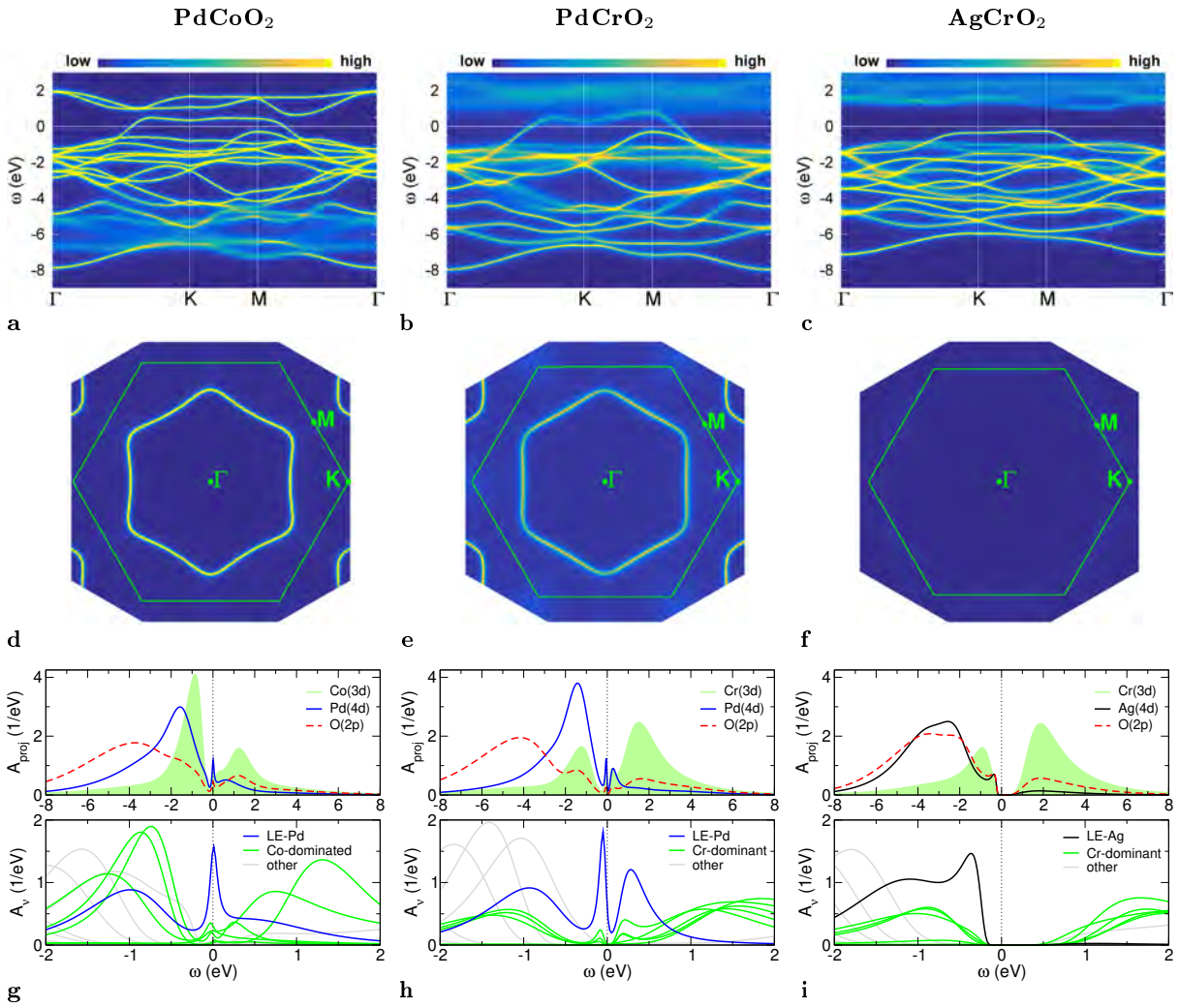


Fig. 11: *Spectral properties of $PdCoO_2$, $PdCrO_2$ and $AgCrO_2$ from paramagnetic DFT+DMFT. (a-c)* **k***-resolved spectral function $A(\mathbf{k},\omega)$ along high symmetry lines in the $k_z = 0$ plane. (d-f) Interacting Fermi surface in the $k_z = 0$ plane. (g-i) Site- and orbital-projected spectral function $A_{\rm proj}(\omega)$ (top) and Bloch-resolved spectral function $A_\nu(\omega)$, see text (bottom), respectively. From Ref. [57].*

The projected spectrum of $PdCoO_2$ exhibits the near subshell filling of Co$(3d)$ and the Pd dominance of the low-energy QP peak at the Fermi level. In $PdCrO_2$, the Cr-t_{2g} states are in a Mott state and therefore their spectral weight is shifted to upper and lower Hubbard bands. The Cr-e_g states are mostly empty, but show also strong incoherence effects in Fig. 11b. Mott criticality in the CrO_2 layers has been originally suggested by several experiments from strong hints for localized Cr^{3+} $S = 3/2$ spins [76, 63, 70]. The QP peak at low energy is of dominant Pd$(4d)$ character, therefore confirming the previously announced mechanism of a correlation-induced shift of a DFT-original deeper lying Pd-dominated band towards $\varepsilon_{\rm F}$. The projected $AgCrO_2$ spectrum shows again the Mott-insulating Cr$(3d)$ part along with the band-insulating Ag$(4d)$ part. The plots of $A_\nu(\omega)$ render obvious that for $PdCoO_2$ the low-energy QP is for the most

part constituted from a single Pd-dominated Bloch dispersion, which we call LE-Pd. The same holds for the $PdCrO_2$ case. Yet importantly, both QP peaks display the hybridizing contribution of Co/Cr-dominated functions, and the LE-Pd function moreover exhibits significant energy dependence. Both features point to the relevance of the subtle impact of electronic correlation onto the low-energy regime. Or in other words, the 'single-band' dispersion crossing $\varepsilon_{\rm F}$, though not dominated by the strongly-interacting $\mathcal{B}$-site $3d$ orbitals, still carries subtle effects of correlations which most certainly rule (parts of) the challenging delafossite physics. But be aware of our difference in nomenclature; 'cPd' denotes the complete single low-energy dispersion, while 'LE-Pd' marks the most dominant A_ν contribution to it.

The impact of electronic correlations is especially crucial and intriguing for $PdCrO_2$. On a more formal level, interactions lead to a metal-to-metal transition between a system with Cr$(3d)$-dominated threefold dispersion at weak coupling and a system with Pd$(4d)$-dominated single dispersion at strong coupling. The question arises how this apparent quantum phase transition takes place with increasing interaction strength U. Figure 12a displays the spectral function and fatbands for the DFT limit ($U = 0$) and for $U = 1.5$ eV, i.e., half the assumed correct interaction strength in $PdCrO_2$. The comparison renders the mechanism for the transition clear: The three bands at the Fermi level are filled with two electrons, and hence four electrons populate the altogether four bands when counting down in energy from $\varepsilon_{\rm F}$. These four bands are of mixed Cr$(3d)$, Pd$(4d)$ character, with dominance from the $3d$ sector. Due to the given band entanglement, strong correlations transform three of them into Hubbard bands, and leave a resulting one with half filling at the Fermi level. Interestingly, for the intermediate coupling (right panel of Fig. 12a), the system adopts a 'strange' situation. The Cr-t_{2g} dispersions are very incoherent and not yet Mott localized, and the cPd dispersion is not yet fully established coherently. Note that especially the latter dispersion, which appears weakly-interacting at strong and weak coupling, is intriguingly affected by correlations close to the given metal-to-metal transition. This underlines the intricate inter-layer physics that is at work in $PdCrO_2$ with its 'hidden-Mott' state.

From a model perspective of correlated electrons, Kondo-lattice type of Hamiltonians describing strongly-interacting sites within a Fermi sea [77, 78] have been discussed as a starting perspective for $PdCrO_2$ [64]. Such a framework has then indeed been put into practice in order to account for the coupling of the Cr spins to the Pd layer in the magnetically-ordered state [65]. However, as already mentioned in Ref. [64], a standard Kondo-lattice model of spins coupled to free electrons appears too simplistic to cover the full complexity of the above described hidden-Mott physics [79]. Modelling the electronic correlations that originate from the CrO_2 layer and spanning over to the Pd layer in a comprehensive way has most definitely to account for the outlined metal-to-metal transition.

Because of the $4d^{10}$ state of silver in $AgCrO_2$, an intricate band entanglement as in $PdCrO_2$ is missing. In the DFT limit, the three Cr-t_{2g} bands at the Fermi level are already half filled with three electrons. Thus the internal Mott transition in the CrO_2 layers does not lead to a metal-to-metal transition, but to a more ordinary metal-to-insulator transition with increasing U. But there is a twist; the valence-band maximum of insulating $AgCrO_2$ is dominated by silver (and

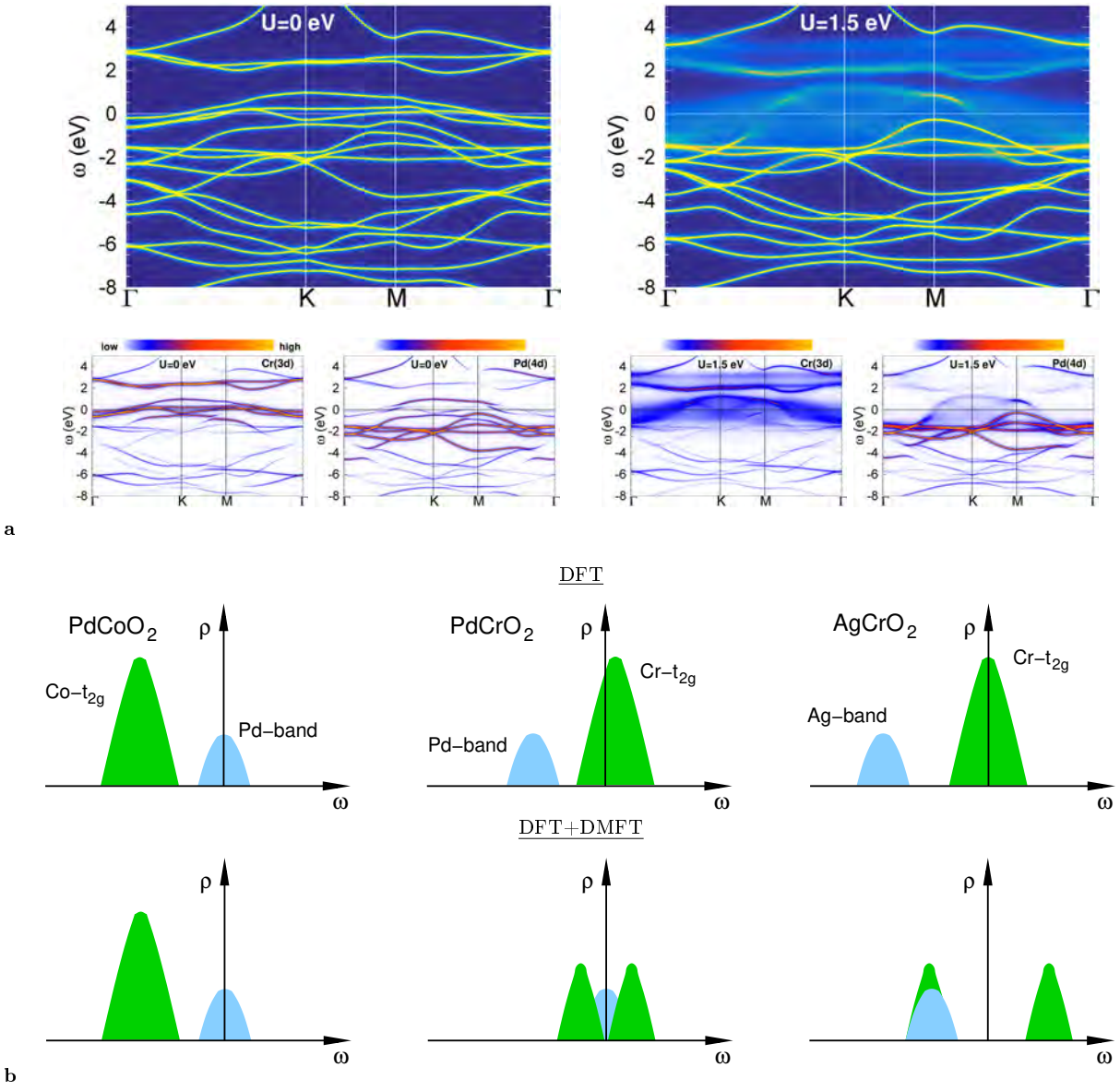


Fig. 12: *Correlation features in delafossites. (a) Spectral properties of $PdCrO_2$ for different interaction strengths. Left panel: $U = 0\,eV$, i.e., DFT bands. Right panel: $U = 1.5\,eV$. Top: spectral function, bottom: fatbands for Cr$(3d)$ (left) and Pd$(4d)$ (right) (from [75]). (b) Sketch of the basic differences in the* **k***-integrated spectral function $\rho(\omega)$ for the most-relevant four bands close to the Fermi level among the studied delafossites. Top: DFT, bottom: DFT+DMFT. Left: $PdCoO_2$: Co-t_{2g} bands occupied, uppermost Pd$(4d)$-band half filled; middle: hidden-Mott $PdCrO_2$: Cr-t_{2g} bands 1/3 filled, uppermost Pd$(4d)$-band occupied; right: band-Mott insulating $AgCrO_2$: Cr-t_{2g} bands half filled, uppermost Ag$(4d)$-band occupied. From Ref. [57].*

oxygen) character, highlighting the band-insulating aspect of the system. The compound is therefore best coined as band-Mott insulator.

To emphasize the key differences of the given delafossites from a minimal perspective, Fig. 12b summarizes the main features from the noted four-band perspective of $\mathcal{B}$-site derived t_{2g} bands and $\mathcal{A}$-site derived uppermost $4d$ band part.

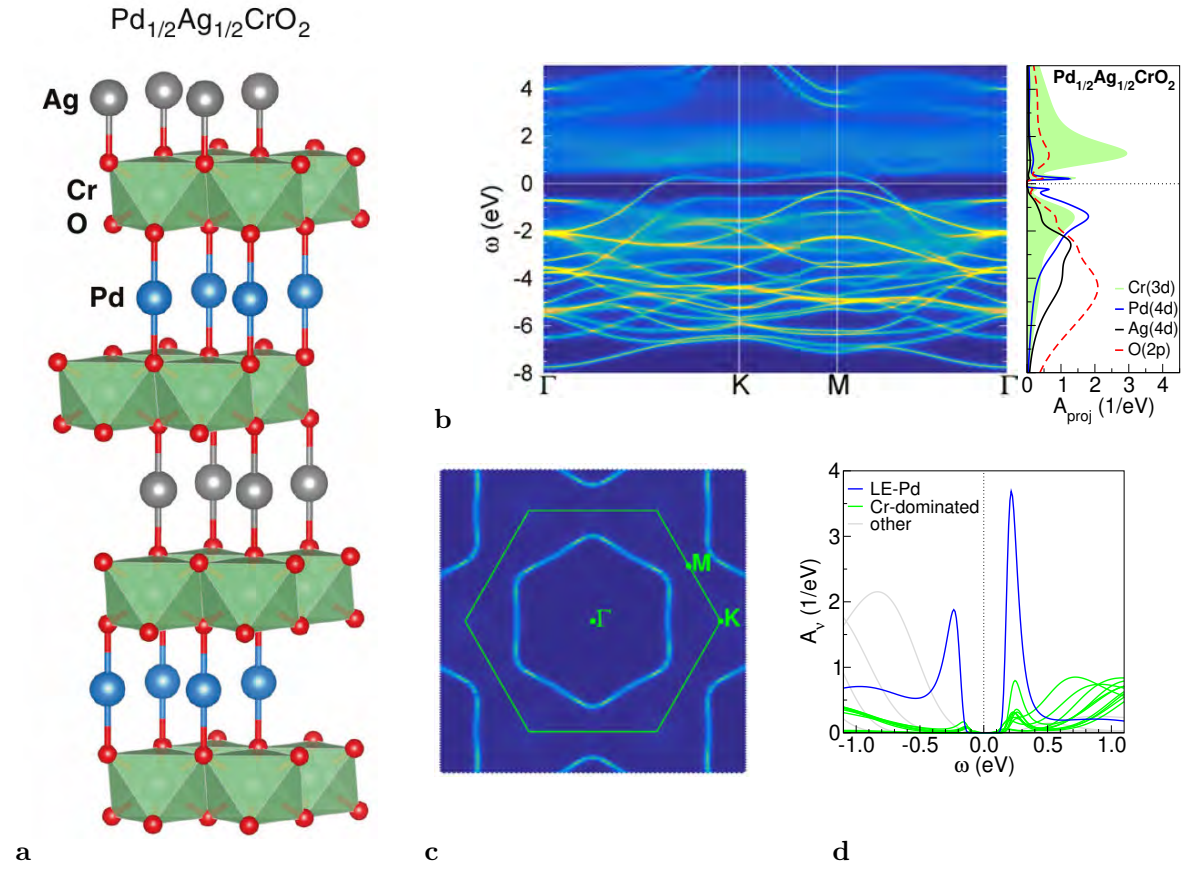


Fig. 13: *Paramagnetic DFT+DMFT spectral data for the out-of-plane heterostructure. $Pd_{1/2}Ag_{1/2}CrO_2$. (a) Crystal structure, with Pd: blue, Ag: grey, Cr: green and O: red. (b) Spectral function $A(\mathbf{k},\omega)$ along high-symmetry lines in the $k_z = 0$ plane of reciprocal space (left) and* **k***-integrated site- and orbital-projected spectral function (right). (c) Fermi surface for $k_z = 0$ within the first Brillouin zone (green hexagon). (d)* **k***-integrated Bloch contribution $A_\nu(\omega)$ with characterization of dominance. After Ref. [75].*

5.3 Theoretical Mott design

As noted earlier, delafossites may be viewed as natural heterostructures with different electronic characteristics in the $\mathcal{A}$- and $\mathcal{B}O_2$-layers. It may be therefore obvious that a merging of delafossite physics and the explicit field of oxide heterostructures could turn out as a fruitful combination. In fact, heterostructures from combining $PdCrO_2$ and $AgCrO_2$ may be of particular interest. Both compounds have similar lattice parameters, resulting in a minor mismatch, and differ only by one electron in the $\mathcal{A}$-site valence. However their electronic phenomenology, i.e., hidden-Mott metal vs. band-Mott insulator, is quite different. Heterostructuring both delafossites provides therefore a specific doping scenario: by keeping the local environment rather undisturbed, filling modifications in the Cr-t_{2g} manifold may be triggered.

Alternate stackings of Pd, CrO_2 and Ag layers, i.e., straightforward out-of-plane heterostructures, are a natural realization of such a scenario. Let us discuss a specific $PdCrO_2$/$AgCrO_2$ stacking along the c-axis, namely $Pd_{1/2}Ag_{1/2}CrO_2$ (see Fig. 13a), with the Hubbard U (identical

on every Cr site) being also correspondingly interpolated from the limiting delafossite cases (see Ref. [75] for further details). From the $PdCrO_2$ perspective, the additional blocking layers of Ag kind as well as the stronger Mott-insulating character induced therefrom into the CrO_2 layers should increase correlations within the Pd layers, too. But this will again happen in a more subtle way than in standard correlated systems, where an associated Coulomb repulsion for such a layer is increased when looking for stronger correlation effects. Remember that there is no Hubbard U on Pd and all correlation increase has to take place in a "nonlocal" way from the surrounding layers. Thus, the present heterostructures pose a quite original correlation problem at low energy; a half-filled Pd layer without intra-layer interaction, subject to rising 'Coulomb pressure and confinement' imposed from the neighboring layers. How does the single electron of dominant Pd$(4d)$ character cope with that situation?

Figure 13 displays the DFT+DMFT spectral properties at room temperature for the given meta-heterostructure. A detailed discussion, also concerning stability issues and with extension to lower-temperature properties, can be found in Ref. [75]. The main result, may be extracted from the low-energy comparison of the $\mathbf{k}$-resolved and the $\mathbf{k}$-integrated data: while there are still QP-like dispersions visible in $A(\mathbf{k},\omega)$ at the Fermi level, the integrated spectra shows vanishing spectral weight at ε_{F}. We coin this puzzling electronic state as correlation-induced semimetal (CIS), which is obviously a result of the intriguing correlation scenario described above. Upon rising obstruction of transport, the key Pd$(4d)$ electron can neither localize in real space (as in a Mott insulator) nor rest in a filled band (as in a band insulator). Hence it reduces the low-energy spectral weight as much as possible for an intact half-filled band, resulting in the CIS state. Note that this finding is not an artifact of the analytical continuation from Matsubara space to real frequencies, as the result is confirmed from both, maximum-entropy as well as Padé methods [75] and therefore a robust feature. In some sense it amounts to a very strong reduction of the usual QP coherence scale of strongly correlated electrons, yet by still keeping the 'coherence' of the original dispersion. To our knowledge, such a rather exotic electronic state has not yet been reported in correlated matter and it awaits experimental verification.

Finally, let us push the limits of conceivable delafossite engineering even somewhat further, by interpreting the metallic implication of Pd in $PdCrO_2$ and the band-insulating implication of Ag in $AgCrO_2$ theoretically footloose. Instead of engineering $PdCrO_2$ 'out of plane' from replacing Pd layers by Ag layers, one may imagine an 'in-plane' alternation from replacing Pd sites by Ag sites in the periodically-repeated $\mathcal{A}$ layer. As a result, a novel natural-heterostructure delafossite emerges, but now with a decorated $\mathcal{A}$ layer. The viewpoint behind arises from picturing the $\mathcal{A}$ layer in hidden-Mott delafossite as a canonical single-band triangular lattice at half filling, embedded in a Mott-insulating background. By manipulating the features of this triangular lattice, a platform for studying correlation effects in such a Mott background may be generated. The simplest manipulations in this regard are given by the straightforward transformations of the original triangular lattice via the K- and M-point ordering instabilities, associated with the honeycomb (K) and the kagomé (M) lattice (see Fig. 14a). Realizing those lattices within a Mott background is exciting because they host Dirac-semimetallic and, in the case of the kagomé lattice, additionally flat-band dispersions. The study of these dispersion features under

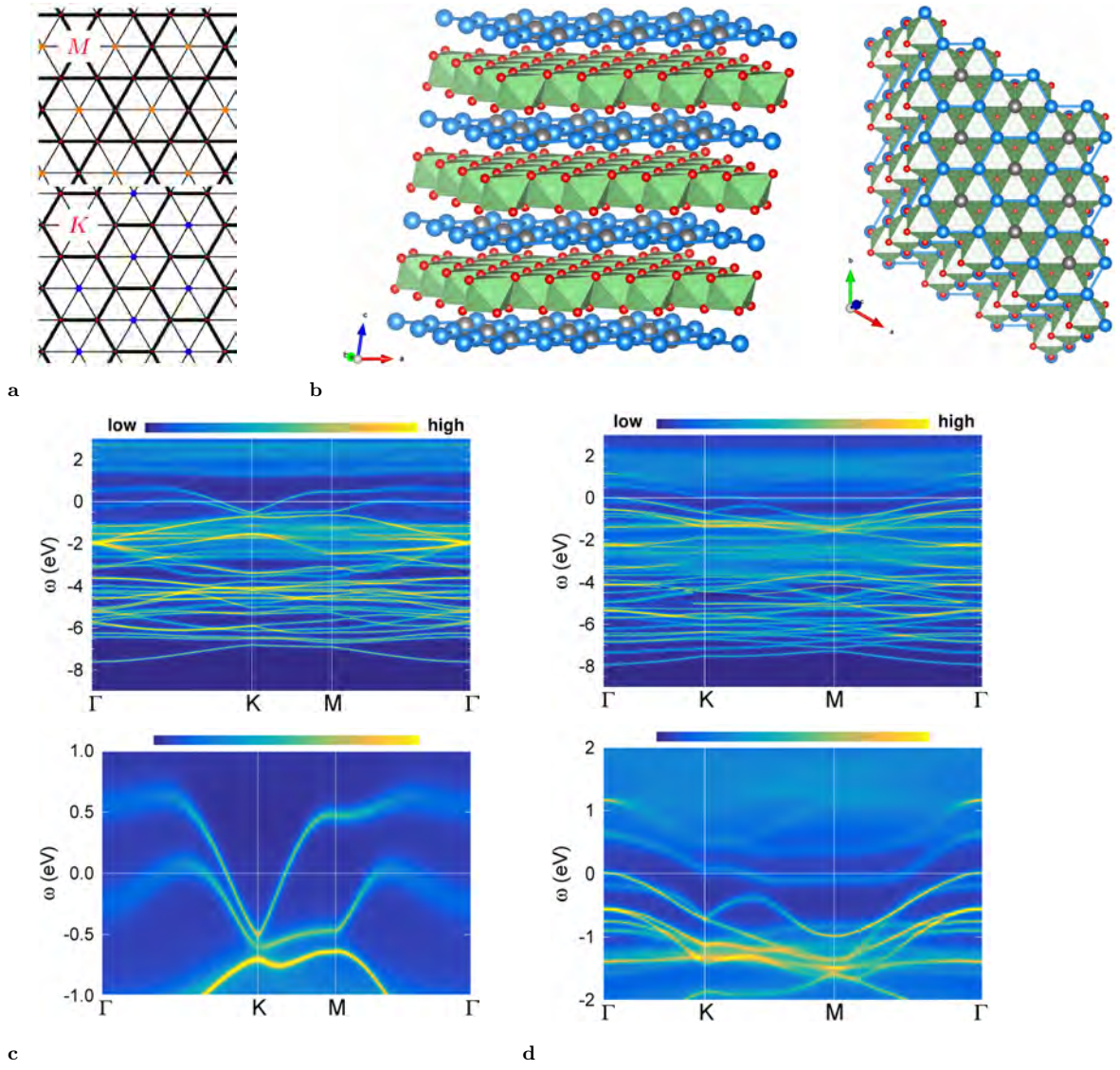


Fig. 14: *Effective-honeycomb and -kagomé (Pd,Ag)CrO_2 delafossite. (a) In-plane design of the $\mathcal{A}$ layer via the natural triangular-lattice ordering instabilities of K-point (honeycomb) and M-point (kagomé) kind (from [80]). (b) Designed crystal structure from understanding Pd(Ag) positions as active(blocking) sites; Pd: blue, Ag: grey, O: red; Cr: green. Left: 3D view and right: view onto the effective honeycomb lattice of Pd sites. (c, d) Interacting spectral information from DFT+DMFT at $T = 193\,K$. (c) Spectral function in (top) larger and (bottom) smaller energy window for effective-honeycomb structure. (d) same as (c) but for effective-kagomé structure.*

the possible influence of strong correlations is a recent emerging research field in condensed matter, see e.g. Refs. [81–83].

To facilitate this reductions of the triangular lattice, the introduction of periodic blocking sites which ideally disconnect hopping processes, can be a promising route. From the nominal Ag$(4d)$ filled-shell in the CrO_2-based delafossites, we assume in the present context that Ag sites within the Pd layer may serve as such blocking sites. For sure, finite covalency will only realize a partial blocking, however, this may still be sufficient to mimic basic honeycomb- or

kagomé-lattice features. To realize a honeycomb(kagomé) lattice in that spirit within the $\mathcal{A}$ layer, one out of three(four) in-plane Pd sites has to be replaced by Ag. Note that though preparing such orderings in the lab will be surely demanding, a layer-by-layer growth might still be feasible from tailoring the layer stoichiometry. If the given Pd-Ag in-plane orderings are thermodynamically stable (for given temperature, pressure, strain, etc.), nature will take its course in realizing the periodic effective honeycomb/kagomé pattern. For the honeycomb case, Fig. 14b depicts the designed delafossite structure. The DFT+DMFT calculations for both effective-lattice systems are again performed with corresponding linear interpolation of the chosen Hubbard U values for the bulk compounds. The honeycomb(kagomé) structure asks for a supercell of three(four) original formula units. Note that we utilized a somewhat lower system temperature of $T = 193$ K for both structural cases, since from general inspection of the spectral properties, the coherence scale for stable quasiparticles appears smaller for the given in-plane alternations compared to the out-of-plane heterostructures.

In Fig. 14c,d we show the resulting spectral functions. From graphene studies it is well known that the nearest-neighbor (NN) tight-binding electronic structure of the half-filled honeycomb lattice is semimetallic, with prominent Dirac dispersions (i.e. massless Dirac fermions) at the K point in reciprocal space [84]. The low-energy spectrum of the present effective-honeycomb lattice in the delafossite setting shows indeed some resemblance of this feature. First, the CrO_2 planes remain Mott-insulating upon the in-plane (Pd, Ag) structuring and there are two Pd-dominated dispersions close to $\varepsilon_{\rm F}$ (see Fig. 14a). A Dirac-like dispersion around K is indicated, yet shifted, with different filling and different overall dispersion compared to graphene. Still, some blocking behavior of Ag is realized, transforming the original $PdCrO_2$ low-energy dispersion in direction towards the canonical honeycomb dispersion. The NN tight-binding electronic structure of the kagomé lattice is known for its flat-band feature at one side of the band edge, as well as for the Dirac dispersion at 4/3 (2/3) filling, depending on the sign of the NN hopping (e.g. [81]). In the present case, the flat-band feature should appear at the upper band edge and thus the Dirac point at 2/3 filling. Figure 14d depicts the resulting spectral function of effective-kagomé (Pd,Ag)CrO_2, and from a brief look the canonical kagomé features are hard to decipher. The intriguing effect of correlations and only-partial Ag blocking render things hard to read. Yet after a closer look, and after also comparing with the non-interacting DFT states, the remains of the flat-band feature can be located around 1 eV above the Fermi level. Interestingly, the interactions in the delafossite structure seemingly transfer spectral weight from there towards $\varepsilon_{\rm F}$. Close to Γ, a waterfall-like spectral signature may be observed. Hence a flat, but at $T = 193$ K rather incoherent, low-energy feature ranging from K to M and halfway to Γ emerges. The Dirac-dispersion feature at K from the canonical kagomé lattice can also be identified at about -0.6 eV.

In conclusion, expectedly neither the effective-honeycomb nor the effective-kagomé lattice realization in the modified $PdCrO_2$ structure enables canonical textbook dispersions at the Fermi level. But as a proof of principle, the in-plane engineering of delafossites may be a route to create nontrivial low-energy dispersions which are subject to the puzzling layer-entangled correlation delafossite physics.

6 Further systems

So far, the present review mainly focussed on early transition-metal titanate heterostructures, but theoretical work also dealt with other oxide designs. For instance, the late-transition-metal rare-earth (R) nickelate series $R\mathrm{NiO_3}$ provides important material building blocks, too [85]. Originally, there was the idea to realize cuprate-like physics within nickelates by proper heterostructuring [86, 87], however experimental success remained absent for a long time. It was only 2019 that Li *et al.* detected superconductivity in thin films of reduced $\mathrm{NdNiO_2}$ on an STO substrate [88]. If this superconducting state is cuprate-like is, however, still a matter of ongoing debate (see e.g. [89]).
Bulk perovskite-like vanadates split into correlated metals, e.g., $\mathrm{SrVO_3}$ and $\mathrm{CaVO_3}$, and Mott insulators, e.g., $\mathrm{LaVO_3}$ and $\mathrm{YVO_3}$. Heterostructures based on $\mathrm{SrVO_3}$ were studied in view of a possible loss of metallicity in a small-layer limit [90, 91]. Moreover, transition-metal oxides from the $4d$ and $5d$ series serve as further building blocks. For instance, the $4d^4$ physics of strontium and/or calcium ruthenates poses a longstanding problem, which can be tuned by heterostructuring [92].

7 Concluding remarks

We here provided a short overview on selected features of DFT+DMFT applications to oxide heterostructures, mainly guided by the author's own work. Note that especially the experimental side of this frontier field of condensed matter physics extends to various further phenomena, e.g., detailed magnetism engineering in interfaces, which can be found in other, more general reviews on the oxide-heterostructure topic.
From the DMFT perspective, one may state that present state-of-the-art DFT+DMFT is truly capable of addressing many challenges of oxide heterostructures. In other words, such 'large-scale' quantum materials science problems are not out of reach for first-principles many-body techniques. Of course, therefore one has to leave the 'safer ground' of model Hamiltonians or more straightforward bulk scenarios. Hence especially in such a regime, DFT+DMFT is not a 'black-box' machinery, but physical intuition and careful weighing of the calculational setting (i.e. choice of correlated subspace, etc.) goes hand in hand with demanding numerics. But this may not be seen as a contradicting endeavor, but quite on the contrary: heavy numerics guided by proper physical insight seems indispensable to tackle the multiorbital and multisite correlated electron problem of oxide heterostructures and related modern materials complexities.

References

[1] J.N. Eckstein and I. Bozovic, Annu. Rev. Mater. Sci. **25**, 679 (1995)

[2] A. Ohtomo and H.Y. Hwang, Nature **427**, 423 (2004)

[3] S. Stemmer and A.J. Millis, MRS Bulletin **38**, 1032 (2013)

[4] D. Lu, D.J. Baek, S.S. Hong, L.F. Kourkoutis, Y. Hikita, and H.Y. Hwang, Nat. Mater. **15**, 1255 (2016)

[5] H.Y. Hwang, Y. Iwasa, M. Kawasaki, B. Keimer, N. Nagaosa, and Y. Tokura, Nat. Mater. **11**, 103 (2012)

[6] P. Zubko, S. Gariglio, M. Gabay, P. Ghosez, and J.-M. Triscone, Annu. Rev. Condens. Matter Phys. **2**, 141 (2011)

[7] J. Chakhalian, J.W. Freeland, A.J. Millis, C. Panagopoulos, and J.M. Rondinelli, Rev. Mod. Phys. **86**, 1189 (2014)

[8] B. Keimer and J.E. Moore, Nat. Phys. **13**, 1045 (2017)

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

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

[11] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

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

[13] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, J. Phys.: Condens. Matter **9**, 7359 (1997)

[14] A.I. Lichtenstein and M. Katsnelson, Phys. Rev. B **57**, 6884 (1998)

[15] B. Amadon, F. Lechermann, A. Georges, F. Jollet, T.O. Wehling, and A.I. Lichtenstein, Phys. Rev. B **77**, 205112 (2008)

[16] V.I. Anisimov, D.E. Kondakov, A.V. Kozhevnikov, I.A. Nekrasov, Z.V. Pchelkina, J.W. Allen, S.-K. Mo, H.-D. Kim, P. Metcalf, S. Suga, A. Sekiyama, G. Keller, I. Leonov, X. Ren, and D. Vollhardt, Phys. Rev. B **71**, 125119 (2005)

[17] S.Y. Savrasov, G. Kotliar, and E. Abrahams, Nature **410**, 793 (2001)

[18] J. Minár, L. Chioncel, A. Perlov, H. Ebert, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **72**, 045125 (2005)

[19] L.V. Pourovskii, B. Amadon, S. Biermann, and A. Georges, Phys. Rev. B **76**, 235101 (2007)

[20] D. Grieger, C. Piefke, O.E. Peil, and F. Lechermann, Phys. Rev. B **86**, 155121 (2012)

[21] F. Lechermann, A. Georges, A. Poteryaev, S. Biermann, M. Posternak, A. Yamasaki, and O.K. Andersen, Phys. Rev. B **74**, 125120 (2006)

[22] M. Potthoff and W. Nolting, Phys. Rev. B **59**, 2549 (1999)

[23] B. Meyer, C. Elsässer, F. Lechermann, and M. Fähnle: *FORTRAN 90 Program for Mixed-Basis-Pseudopotential Calculations for Crystals* (2012)

[24] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A.J. Millis, Phys. Rev. Lett. **97**, 076405 (2006)

[25] P. Seth, I. Krivenko, M. Ferrero, and O. Parcollet, Comput. Phys. Commun. **200**, 274 (2016)

[26] N. Nakagawa, H.Y. Hwang, and D.A. Muller, Nat. Mater. **5**, 204 (2006)

[27] Y.Z. Chen, N. Bovet, F. Trier, D.V. Christensen, F.M. Qu, N.H. Andersen, T. Kasama, W. Zhang, R. Giraud, J. Dufouleur, T.S. Jespersen, J.R. Sun, A. Smith, J. Nygård, L. Lu, B. Büchner, B.G. Shen, S. Linderoth, and N. Pryds, Nat. Commun. **4**, 1372 (2013)

[28] N. Reyren, S. Thiel, A.D. Caviglia, L.F. 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] A. Brinkman, M. Huijben, M. van Zalk, J. Huijben, U. Zeitler, J.C. Maan, W.G. van der Wiel, G. Rijnders, D.H.A. Blank, and H. Hilgenkamp, Nat. Mater. **6**, 493 (2007)

[30] F. Lechermann, L. Boehnke, D. Grieger, and C. Piefke, Phys. Rev. B **90**, 085125 (2014)

[31] M. Salluzzo, S. Gariglio, D. Stornaiuolo, V. Sessi, S. Rusponi, C. Piamonteze, G.M. DeLuca, M. Minola, D. Marré, A. Gadaleta, H. Brune, F. Nolting, N.B. Brookes, and G. Ghiringhelli, Phys. Rev. Lett. **111**, 087204 (2013)

[32] N. Pavlenko, T. Kopp, E.Y. Tsymbal, J. Mannhart, and G.A. Sawatzky, Phys. Rev. B **86**, 064431 (2012)

[33] G. Berner, M. Sing, H. Fujiwara, A. Yasui, Y. Saitoh, A. Yamasaki, Y. Nishitani, A. Sekiyama, N. Pavlenko, T. Kopp, C. Richter, J. Mannhart, S. Suga, and R. Claessen, Phys. Rev. Lett. **110**, 247601 (2013)

[34] A. Chikina, F. Lechermann, M.-A. Husanu, M. Caputo, C. Cancellieri, X. Wang, T. Schmitt, M. Radovic, and V.N. Strocov, ACS Nano **12**, 7927 (2018)

[35] M. Behrmann and F. Lechermann, Phys. Rev. B **92**, 125148 (2015)

[36] F. Lechermann, A. Georges, G. Kotliar, and O. Parcollet, Phys. Rev. B **76**, 155102 (2007)

[37] T. Li, P. Wölfle, and P.J. Hirschfeld, Phys. Rev. B **40**, 6817 (1989)

[38] F. Lechermann, H.O. Jeschke, A.J. Kim, S. Backes, and R. Valentí, Phys. Rev. B **93**, 121103(R) (2016)

[39] A.F. Santander-Syro, O. Copie, T. Kondo, F. Fortuna, S. Pailhè, R. Weht, X.G. Qiu, F. Bertran, A. Nicolaou, A. Taleb-Ibrahimi, P.L. Fèvre, G. Herranz, M. Bibes, N. Reyren, Y. Apertet, P. Lecoeur, A. Barthélémy, and M.J. Rozenberg, Nature **469**, 189 (2011)

[40] W. Meevasana, P.D.C. King, R.H. He, S.-K. Mo, M. Hashimoto, A. Tamai, P. Songsiriritthigul, F. Baumberger, and Z.-X. Shen, Nat. Mat. **10**, 114 (2011)

[41] S. McKeown Walker, A. de la Torre, F.Y. Bruno, A. Tamai, T.K. Kim, M. Hoesch, M. Shi, M.S. Bahramy, P.C. King, and F. Baumberger, Phys. Rev. Lett. **113**, 177601 (2014)

[42] H. Tanaka, T. Matsumoto, T. Kawai, and S. Kawai, Jpn. J. Appl. Phys. **32**, 1405 (1993)

[43] F. Lechermann, L. Boehnke, and D. Grieger, Phys. Rev. B **87**, 241101(R) (2013)

[44] A. Ohtomo, D.A. Muller, J.L. Grazul, and H.Y. Hwang, Nature **419**, 378 (2002)

[45] S. Okamoto and A.J. Millis, Nature **428**, 630 (2004)

[46] H. Ishida and A. Liebsch, Phys. Rev. B **77**, 115350 (2008)

[47] A. Rüegg, S. Pilgram, and M. Sigrist, Phys. Rev. B **75**, 195117 (2007)

[48] K. Dymkowski and C. Ederer, Phys. Rev. B **89**, 161109(R) (2014)

[49] G. Sclauzero and C. Ederer, Phys. Rev. B **92**, 235112 (2015)

[50] F. Lechermann, Sci. Rep **7**, 1565 (2017)

[51] P. Moetakef, C.A. Jackson, J. Hwang, L. Balents, S.J. Allen, and S. Stemmer, Phys. Rev. B **86**, 201102(R) (2012)

[52] C.A. Jackson, J.Y. Zhang, C.R. Freeze, and S. Stemmer, Nat. Commun. **5**, 4258 (2014)

[53] E. Mikheev, C.R. Freeze, B.J. Isaac, T.A. Cain, and S. Stemmer, Phys. Rev. B **91**, 165125 (2015)

[54] F. Lechermann and M. Obermeyer, New J. Phys. **17**, 043026 (2015)

[55] P.B. Marshall, E. Mikheev, S. Raghavan, and S. Stemmer, Phys. Rev. Lett. **117**, 046402 (2016)

[56] F. Lechermann, John von Neumann Institute for Computing (NIC) NIC Series **49** (2018)

[57] F. Lechermann, npj Comput. Mater. **7**, 120 (2021)

[58] R.D. Shannon, D.B. Rogers, and C.T. Prewitt, Inorg. Chem. **10**, 713 (1971)

[59] C.T. Prewitt, R.D. Shannon, and D.B. Rogers, Inorg. Chem. **10**, 719 (1971)

[60] D.B. Rogers, R.D. Shannon, and C.T. Prewitt, Inorg. Chem. **10**, 723 (1971)

[61] R. Daou, R. Frésard, V. Eyert, S. Hébert, and A. Maignan, Sci. Technol. Adv. Mater. **18**, 919 (2017)

[62] A.P. Mackenzie, Rep. Prog. Phys. **80**, 032501 (2017)

[63] H.-J. Noh, J. Jeong, B. Chang, D. Jeong, H.S. Moon, E.-J. Cho, J.M. Ok, J.S. Kim, K. Kim, B.I. Min, H.-K. Lee, J.-Y. Kim, B.-G. Park, H.-D. Kim, and S. Lee, Sci. Rep. **4**, 3680 (2014)

[64] F. Lechermann, Phys. Rev. Mater. **2**, 085004 (2018)

[65] V. Sunko, F. Mazzola, S. Kitamura, S. Khim, P. Kushwaha, O.J. Clark, M. Watson, I. Markovic, D. Biswas, L. Pourovskii, T.K. Kim, T.L. Lee, P.K. Thakur, H. Rosner, A.G.R. Moessner, T. Oka, A.P. Mackenzie, and P.D.C. King, Sci. Adv. **6**, eaaz0611 (2020)

[66] H.-J. Noh, J. Jeong, J. Jeong, E.-J. Cho, S.B. Kim, K. Kim, B.I. Min, and H.-D. Kim, Phys. Rev. Lett. **102**, 256404 (2009)

[67] C.W. Hicks, A.S. Gibbs, A.P. Mackenzie, H. Takatsu, Y. Maeno, and E.A. Yelland, Phys. Rev. Lett. **109**, 116401 (2012)

[68] J.A. Sobota, K. Kim, H. Takatsu, M. Hashimoto, S.-K. Mo, Z. Hussain, T. Oguchi, T. Shishidou, Y. Maeno, B.I. Min, and Z.-X. Shen, Phys. Rev. B **88**, 125109 (2013)

[69] J.M. Ok, Y.J. Jo, K. Kim, T. Shishidou, E.S. Choi, H.-J. Noh, T. Oguchi, B.I. Min, and J.S. Kim, Phys. Rev. Lett. **111**, 176405 (2013)

[70] C.W. Hicks, A.S. Gibbs, L. Zhao, P. Kushwaha, H. Borrmann, A.P. Mackenzie, H. Takatsu, S. Yonezawa, Y. Maeno, and E.A. Yelland, Phys. Rev. B **92**, 014425 (2015)

[71] K.P. Ong and D.J. Singh, Phys. Rev. B **85**, 134403 (2012)

[72] D. Billington, D. Ernsting, T.E. Millichamp, C. Lester, S.B. Dugdale, D. Kersh, J.A. Duffy, S.R. Giblin, J.W. Taylor, P. Manuel, D.D. Khalyavin, and H. Takatsu, Sci. Rep. **5**, 12428 (2015)

[73] M.A. Korotin, V.I. Anisimov, D.I. Khomskii, and G.A. Sawatzky, Phys. Rev. Lett. **80**, 4305 (1998)

[74] F. Lechermann, Phys. Rev. Lett. **102**, 046403 (2009)

[75] F. Lechermann and R. Richter, Phys. Rev. Res. **2**, 013352 (2020)

[76] H. Takatsu, H. Yoshizawa, S. Yonezawa, and Y. Maeno, Phys. Rev. B **79**, 104424 (2009)

[77] N. Mott, Phil. Mag. **30**, 403 (1974)

[78] S. Doniach, Physica B **91**, 231 (1977)

[79] E.V. Komleva, V.Y. Irkhin, I.V. Solovyev, M.I. Katsnelson, and S.V. Streltsov, Phys. Rev. B **102**, 174438 (2020)

[80] L. Boehnke and F. Lechermann, Phys. Rev. B **85**, 115128 (2012)

[81] I.I. Mazin, H.O. Jeschke, F. Lechermann, H. Lee, M. Fink, R. Thomale, and R. Valentí, Nat. Commun. **5**, 4261 (2014)

[82] L. Ye, M. Kang, J. Liu, F. von Cube, C.R. Wicker, T. Suzuki, C. Jozwiak, A. Bostwick, E. Rotenberg, D.C. Bell, L. Fu, R. Comin, and J.G. Checkelsky, Nature **555**, 638 (2018)

[83] N.J. Ghimire and I.I. Mazin, Nat. Mater. **19**, 137 (2020)

[84] P.R. Wallace, Phys. Rev. **71**, 622 (1947)

[85] J. Hwang, J. Son, J.Y. Zhang, A. Janotti, C.G. Van de Walle, and S. Stemmer, Phys. Rev. B **87**, 060101(R) (2013)

[86] P. Hansmann, X. Yang, A. Toschi, G. Khaliullin, O.K. Andersen, and K. Held, Phys. Rev. Lett. **103**, 016401 (2009)

[87] M.J. Han, X. Wang, C.A. Marianetti, and A.J. Millis, Phys. Rev. Lett. **107**, 206804 (2011)

[88] D. Li, K. Lee, B.Y. Wang, M. Osada, S. Crossley, H.R. Lee, Y. Cui, Y. Hikita, and H.Y. Hwang, Nature **572**, 624 (2019)

[89] M. Norman, Physics **13**, 85 (2020)

[90] S. Okamoto, Phys. Rev. B **84**, 201305(R) (2011)

[91] Z. Zhong, M. Wallerberger, J.M. Tomczak, C. Taranto, N. Parragh, A. Toschi, G. Sangiovanni, and K. Held, Phys. Rev. Lett. **114**, 246401 (2015)

[92] L. Si, O. Janson, G. Li, Z. Zhong, Z. Liao, G. Koster, and K. Held, Phys. Rev. Lett. **119**, 026402 (2017)

10 Dynamical Mean-Field Theory for Correlated Topological Phases

Michael Potthoff

Department of Physics, University of Hamburg,
Notkestraße 9-11, 22607 Hamburg

The Hamburg Centre for Ultrafast Imaging,
Luruper Chaussee 149, 22761 Hamburg

Contents

1 Motivation **2**

2 Chern number **4**

3 Chern insulator **13**

4 Electron correlations **18**

5 Exact interplay of correlations and topology in $D = \infty$ **20**

6 Concluding discussion **30**

1 Motivation

Consider two objects A and B. These could be physical objects, such as elementary particles, molecules, or macroscopically large bodies. They could be mathematical structures, but also poems or novels, the mood of two people, etc. It is obvious that the question "Is A equal to B?" rarely deserves a positive answer. In the case of "Is A essentially equal to B?" on the other hand, it seems to be a question worthy of discussion in almost all situations. All the complexity here is cleverly hidden in the word "essentially". This leaves room for us to define criteria by which we can decide about essential equality or inequality. An instructive example is a cat and a dog. Are these "objects" the same? No, but it may be reasonable to regard them as essentially the same, depending on the criterion chosen, e.g., that of being an animal.

Why do we bother? The relation "is essentially the same" is an equivalence relation because the following statements appear to be immediately true: "A is essentially the same as A", "if A is essentially the same as B, then B is essentially the same as A", and "if A is essentially the same as B and B is essentially the same as C, then A is essentially the same as C". Any equivalence relation can be used to classify objects, and classification is an important goal of almost any science. Classification of life forms, for example, is one of the most important goals of biology.

In physics, there are various concepts for classifying systems or the static or dynamic states of systems. These find expression, for example, in the periodic table of elements or in equilibrium phase diagrams of macroscopically large physical bodies. Despite the same chemical composition, we distinguish, for example, the condensed and the liquid state of water. Decision-making criteria in many cases, and also in the distinction between solid and liquid, can be based on symmetry arguments, i.e., on invariance under a group of transformations.

Let us discuss a well-known example from mathematics, which is frequently used in quantum physics: A separable Hilbert space is a linear space over the field of real or complex numbers, i.e., addition and scalar multiplication are defined and satisfy a couple of postulates. In addition the space is equipped with a bilinear or sesquilinear inner product, which defines a distance function for which it is a complete metric space, and has a countable basis. Two separable Hilbert spaces $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$ are essentially the same, $H_\mathrm{A} \cong H_\mathrm{B}$, if there exists an isomorphism $\mathcal{H}_\mathrm{A} \to \mathcal{H}_\mathrm{B}$, i.e., an invertible map preserving the mathematical structure. Due to the rich structure, the equivalence relation $\cong$ is very strong. This "explains" the simplicity of the classification, namely *all* (infinite-dimensional) separable Hilbert spaces are essentially the same and, e.g., isomorphic to L^2.

A more fruitful, finer classification is obtained with a "weaker" equivalence relation. Groups, for example, are mathematical structures defined with a minimum of postulates. Structure-conserving maps between two groups G_1 and G_2, group isomorphisms, therefore define a much weaker equivalence relation $G_1 \cong G_2$. In fact, a complete classification is an entirely open mathematical problem. Narrowing a bit and asking for essential equality between finite, simple groups, brings us to one of the major milestones of group theory. Here, a complete classification has been achieved in fact with a proof that is documented several hundred papers published over the last 70 years.

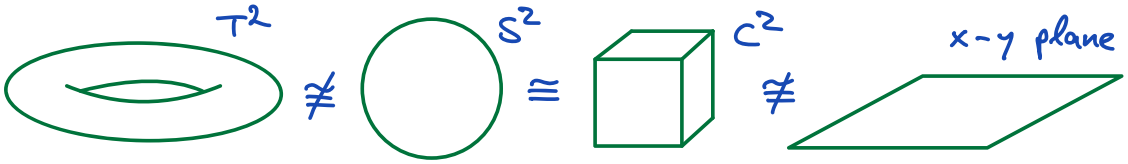


Fig. 1: *Various topological subspaces of* $\mathbb{R}^3$.

An equivalence relation that has proven very fruitful in the context of condensed-matter physics is "A can continuously and reversibly be deformed into B". Continuity is a rather weak property of functions and the main concept of topology. To give a mathematical example again, we may consider $\mathbb{R}^3$ as a topological space and forget about its capability to carry a linear structure. Among the topological subspaces of $\mathbb{R}^3$ we have, e.g., the 2-sphere $S^2 = \{x \in \mathbb{R}^3 \mid ||x|| = 1\}$, the Cartesian product of 1-spheres $S^1 \times S^1$, the surface of a cube C^2, the 2-torus ("donut") T^2, the open or closed 3-ball with the 2-sphere as its boundary, and the like, see Fig. 1. We find that $S^1 \times S^1 \cong T^2 \ncong S^2 \cong C^2 \ncong x$-$y$ plane etc.

An important tool of topological classification are *topological invariants*. These are all quantities that are preserved under continuous transformations that have a continuous inverse transformation. The number of holes is such an invariant. Two subspaces with different number of holes in fact cannot be continuously deformed into each other. This explains, why S^2 (no hole) and the torus T^2 (one hole) are topologically different. On the other hand, two subspaces with the same number of holes are not necessarily topologically equivalent. An example is the 2-sphere and the x-y plane. These may differ in another invariant.

In electron band-structure theory we are interested in the topological classification of more complex structures. Consider a system of independent electrons on an infinite periodic lattice. Its Bloch Hamiltonian $H(\boldsymbol{k})$ parametrically depends on the wave vectors $\boldsymbol{k}$ in a reciprocal unit cell. In two dimensions, for example, this parameter space forms a torus T^2 because wave vectors at the boundaries of the reciprocal unit cell must be identified. However, it is not (only) the nontrivial topology of the parameter-space geometry that is the focus. We are rather interested in the topological classification of the *bundle* of ground states of $H(\boldsymbol{k})$ over T^2. Questions behind this are: Can these ground states be deformed into each other continuously? Can we topologically classify band structures in this way? In case of band insulators, this seemingly simple question has been answered affirmatively and comprehensively. The answer is surprisingly deep and requires substantial mathematical background. It has led to a classification scheme called "the 10-fold way", in which symmetry arguments and different topological invariants play the essential role.

The goal of this contribution is twofold: First, with the Chern number, we exemplarily pick a central topological invariant and, starting from basic concepts, discuss its precise definition, its range of applicability for the construction of topological phase diagrams, its connection with time-reversal symmetry, its experimental significance, etc. This elucidates the concept of "class A" Chern insulators, i.e., band insulators with a certain topological twist, which fall into one of the mentioned 10 classes.

Second, despite the great success of topological band theory, it cannot be straightforwardly extended to *interacting* electron systems. The development of novel concepts for the topological characterization and classification of the states of strongly correlated electron systems is one the major activities in theoretical physics, and in fact it is one of the most fascinating endeavors. One promising route is paved by combining dynamical mean-field theory with topological band theory, and the discussion of this idea and of an illuminating example will take a major part of the present work.
The presentation heavily makes use of various ideas from textbooks on topology and its applications in physics [1–4], on the results of the M.Sc. thesis of David Krüger, essentially published in Ref. [5], and of the B.Sc. thesis of Lara Heyl [6].

2 Chern number

The Chern number is a topological invariant. To define this invariant and to prove its quantization, we have to introduce and discuss a couple of basic theoretical concepts, including the adiabatic theorem, the Berry connection, gauge transformations and the Berry phase, and finally the Berry curvature.

Adiabatic theorem

Consider a system Hamiltonian H which smoothly depends on r parameters $\boldsymbol{R} = (R_1, ..., R_r)$

$$H = H(\boldsymbol{R}) = H(R_1, ..., R_r)\,. \tag{1}$$

The space of parameters $\mathcal{M}$ is a smooth r-dimensional manifold. We will further assume that the ground state $\big|\Phi_0(\boldsymbol{R})\big\rangle$ of $H(\boldsymbol{R})$ is unique (nondegenerate) for all $\boldsymbol{R}$. This implies that there is a finite gap $\Delta(\boldsymbol{R})$ to the first excited state $\big|\Phi_1(\boldsymbol{R})\big\rangle$ on the entire manifold $\mathcal{M}$. For each parameter configuration $\boldsymbol{R} \in \mathcal{M}$ the Hilbert space $\mathcal{H}$ is "locally" spanned by an orthonormal eigenbasis $\{\big|\Phi_j(\boldsymbol{R})\big\rangle\}$ of $H(\boldsymbol{R})$, i.e., we have

$$H(\boldsymbol{R})\big|\Phi_j(\boldsymbol{R})\big\rangle = E_j(\boldsymbol{R})\big|\Phi_j(\boldsymbol{R})\big\rangle\,. \tag{2}$$

Suppose that the parameters evolve in time as described by a trajectory $\boldsymbol{R} = \boldsymbol{R}(t)$ in $\mathcal{M}$. This means that the Hamiltonian acquires an explicit time dependence $H(t) = H\big(\boldsymbol{R}(t)\big)$. The formal solution of Schrödinger's equation provides us with the state of the system at time t: $\big|\Psi(t)\big\rangle = \mathcal{T}\exp\big(-i\int_0^t d\tau\, H\big(\boldsymbol{R}(\tau)\big)\big)\big|\Psi(0)\big\rangle$, where $\mathcal{T}$ is the time ordering and where have set $\hbar \equiv 1$. We assume that the state at time $t = 0$ is the ground state of $H\big(\boldsymbol{R}(t=0)\big)$, i.e., at $t = 0$ we have $\big|\Psi(0)\big\rangle = \big|\Phi_0\big(\boldsymbol{R}(0)\big)\big\rangle$.
The adiabatic theorem [7,8] states that, *if the time evolution of the parameters $\boldsymbol{R}(t)$ is sufficiently slow*, the system will remain in its ground state. More precisely, the state of the quantum system at time t is the instantaneous ground state of the Hamiltonian $H\big(\boldsymbol{R}(t)\big)$, i.e., we have $\big|\Psi(t)\big\rangle \propto \big|\Phi_0\big(\boldsymbol{R}(t)\big)\big\rangle$. This is a strong statement and it is nontrivial to precisely formulate the conditions under which this holds true. Roughly, its validity requires that the typical time scale

τ_{typical} of the parameter dynamics is much larger than the inverse of the gap: $\tau_{\text{typical}} \gg 1/\Delta$. Here, we take the adiabatic theorem for granted and will focus on the phase of the state. The adiabatic theorem, as we have formulated it, tells us that the state at time t is a ground state of $H\big(\boldsymbol{R}(t)\big)$,

$$\big|\Psi(t)\big\rangle = e^{-i\gamma(t)}\,\big|\Phi_0\big(\boldsymbol{R}(t)\big)\big\rangle\;, \tag{3}$$

but it leaves the phase factor $e^{-i\gamma(t)}$ yet undetermined.

Berry connection

Schrödinger's equation can be derived from the action principle $\delta S = \delta\int dt\, L = 0$ with the Lagrangian

$$L = L\Big(|\Psi\rangle, \partial_t|\Psi\rangle, \langle\Psi|, \partial_t\langle\Psi|,\, t\Big) = \big\langle\Psi\big|i\partial_t - H\big(\boldsymbol{R}(t)\big)\big|\Psi\big\rangle\;. \tag{4}$$

In fact, $0 = \partial L/\partial\langle\Psi| - \partial_t\big(\partial L/\partial(\langle\partial_t\Psi|)\big) = \big(i\partial_t - H\big(\boldsymbol{R}(t)\big)\big)|\Psi\rangle$, while the second Euler-Lagrange equation for $\langle\Psi|$ is just the adjoint Schrödinger equation. The Lagrange formalism is a useful tool, as one can incorporate the result of the adiabatic theorem as a holonomic constraint with the goal to get a simple equation of motion for the yet unknown phase $\gamma(t)$ only. Using Eq. (3) to express the "generalized coordinates" $\big|\Psi(t)\big\rangle$ and $\big\langle\Psi(t)\big|$ in terms of $\big|\Phi_0\big(\boldsymbol{R}(t)\big)\big\rangle$ and $\big\langle\Phi_0\big(\boldsymbol{R}(t)\big)\big|$, we get an effective Lagrangian,

$$L_{\text{eff}} = e^{i\overline{\gamma}(t)}\,\Big\langle\Phi_0\big(\boldsymbol{R}(t)\big)\Big|\big(i\partial_t - H\big(\boldsymbol{R}(t)\big)\big)\,e^{-i\gamma(t)}\Big|\Phi_0\big(\boldsymbol{R}(t)\big)\Big\rangle\;, \tag{5}$$

which indeed is a functional of the unknown phase of the system state only, while the local ground states $\big|\Phi_0\big(\boldsymbol{R}(t)\big)\big\rangle$ have fixed selected phases. Since $|\Psi\rangle$ and $\langle\Psi|$ must be treated as independent in Eq. (4), γ and $\overline{\gamma}$ have to be treated as independent as well. Hence, the Lagrangian L_{eff} has the functional form $L_{\text{eff}}(\gamma, \overline{\gamma}, \partial_t\gamma, \partial_t\overline{\gamma},\, t)$ with an explicit time dependence due to $\boldsymbol{R}(t)$. Carrying out the time derivative $i\partial_t$ of the terms on the right and using Eq. (2) for $j = 0$, we arrive at

$$L_{\text{eff}} = e^{i\overline{\gamma}(t)}e^{-i\gamma(t)}\Big(\partial_t\gamma(t) - E_0\big(\boldsymbol{R}(t)\big) + \boldsymbol{A}_0\big(\boldsymbol{R}(t)\big)\cdot\dot{\boldsymbol{R}}(t)\Big)\;. \tag{6}$$

Here,

$$\boldsymbol{A}_0(\boldsymbol{R}) = i\Big\langle\Phi_0(\boldsymbol{R})\Big|\frac{\partial}{\partial\boldsymbol{R}}\Big|\Phi_0(\boldsymbol{R})\Big\rangle \tag{7}$$

is the Berry connection, which mediates between the ground states at $\boldsymbol{R}$ and $\boldsymbol{R}+d\boldsymbol{R}$ on the manifold $\mathcal{M}$, and $\boldsymbol{A}_0(\boldsymbol{R})\cdot\dot{\boldsymbol{R}} = \sum_{\rho=1}^{r} A_{0,\rho}(\boldsymbol{R})\cdot\dot{R}_\rho$ with $\dot{R}_\rho = \partial_t R_\rho$. Note that the normalization condition for the state implies that $\boldsymbol{A}_0(\boldsymbol{R})$ is real. The resulting Euler-Lagrange equations for γ and $\overline{\gamma}$ are equivalent, such that eventually $\gamma = \overline{\gamma}$. The equation for γ,

$$\partial_t\gamma(t) = E_0\big(\boldsymbol{R}(t)\big) - \boldsymbol{A}_0\big(\boldsymbol{R}(t)\big)\cdot\dot{\boldsymbol{R}}(t)\;, \tag{8}$$

is easily solved by

$$\gamma(t) = \gamma(0) + \int_0^t E_0\big(\boldsymbol{R}(\tau)\big)\,d\tau - \int_0^t \boldsymbol{A}_0\big(\boldsymbol{R}(\tau)\big)\cdot\dot{\boldsymbol{R}}(\tau)\,d\tau\;, \tag{9}$$

Finally, with Eq. (3) and choosing w.l.o.g. $\gamma(0)=0$,

$$\big|\Psi(t)\big\rangle = \exp\left(-i\int_0^t E_0\big(\boldsymbol{R}(\tau)\big)\,d\tau\right)\exp\left(i\int_{\mathcal{C},\boldsymbol{R}(0)}^{\boldsymbol{R}(t)}\boldsymbol{A}_0(\boldsymbol{R})\cdot d\boldsymbol{R}\right)\big|\Phi_0\big(\boldsymbol{R}(t)\big)\big\rangle\,. \tag{10}$$

The first phase factor is called "dynamical", the second "geometrical". For the second, the substitution rule has been used in the last step. We see that it only depends on the path $\mathcal{C}$ from $\boldsymbol{R}(0)$ to $\boldsymbol{R}(t)$ in $\mathcal{M}$, a geometrical object, but not on the time dependence of the trajectory $\boldsymbol{R}(t)$, on "velocity", "acceleration" etc.

Gauge transformations and Berry phase

After having demonstrated the relevance of the Berry connection for the adiabatic time evolution of a quantum system, we will now turn to our main question, namely its relevance for the (time-independent) topological properties of the system and reconsider the bundle of ground states $\big|\Phi_0(\boldsymbol{R})\big\rangle$ with $\boldsymbol{R}\in\mathcal{M}$. The choice of an orthonormal energy eigenbasis for each $\boldsymbol{R}$, see Eq. (2), also implies the choice of a particular phase $\varphi(\boldsymbol{R})$ for each of the ground states in $\mathcal{M}$. This choice is arbitrary, so it is important to ask what happens if we make a different choice. Our naive expectation is that all observable ground-state properties of the system are invariant under a transformation of the phase, i.e., under a gauge transformation of the form

$$\big|\Phi_0(\boldsymbol{R})\big\rangle \mapsto e^{i\varphi(\boldsymbol{R})}\,\big|\Phi_0(\boldsymbol{R})\big\rangle\,. \tag{11}$$

Such an $\boldsymbol{R}$-dependent gauge transformation is called "local", as opposed to a "global" transformation $\big|\Phi_0(\boldsymbol{R})\big\rangle \mapsto e^{i\varphi}\big|\Phi_0(\boldsymbol{R})\big\rangle$. The system's topological properties are governed by the map $\boldsymbol{R}\mapsto e^{i\varphi(\boldsymbol{R})}$ from the parameter manifold $\mathcal{M}$ into $U(1)=\big\{z\in\mathbb{C}\,\big|\,|z|=1\big\}$. The fact that this map or, equivalently, the $\boldsymbol{R}$-dependent choice of the phase factors, has observable consequences as has been pointed out by M. Berry [9].
Under a local gauge transformation, Eq. (11), the Berry connection transforms as

$$\boldsymbol{A}_0(\boldsymbol{R}) \mapsto \boldsymbol{A}_0'(\boldsymbol{R}) = \boldsymbol{A}_0(\boldsymbol{R}) - \frac{\partial\varphi(\boldsymbol{R})}{\partial\boldsymbol{R}}\,. \tag{12}$$

This implies that $\boldsymbol{A}_0(\boldsymbol{R})$ is gauge dependent. It has no direct physical meaning but depends on the *arbitrary* choice of the phases $\varphi(\boldsymbol{R})$ of the local ground states $|\Phi_0(\boldsymbol{R})\rangle$. The situation is particularly interesting, if $\mathcal{C}$ is a *closed* path in $\mathcal{M}$, given by some parametrization $\boldsymbol{R}=\boldsymbol{R}(\lambda)$, where the real parameter λ *could* stand for the time. In this case $\gamma_{\mathcal{C}}=\oint_{\mathcal{C}}\boldsymbol{A}_0(\boldsymbol{R})\cdot d\boldsymbol{R}$ is called the Berry phase. Under a gauge transformation, the Berry phase transforms as $\gamma_{\mathcal{C}}\mapsto\gamma_{\mathcal{C}}-\oint_{\mathcal{C}}\big(\partial\varphi(\boldsymbol{R})/\partial\boldsymbol{R}\big)\cdot d\boldsymbol{R}$, see Eq. (12). If certain rather general conditions are satisfied, as specified by the Poincaré lemma and regarding, in particular, the continuity of $\partial\varphi(\boldsymbol{R})/\partial\boldsymbol{R}$, we can argue that the Berry phase is invariant $\gamma_{\mathcal{C}}\mapsto\gamma_{\mathcal{C}}-\oint_{\mathcal{C}}d\varphi(\boldsymbol{R})=\gamma_{\mathcal{C}}$.
However, the integral along $\mathcal{C}$ can be nonzero in fact. Consider, for example, $\mathcal{M}=\mathbb{R}^3$ and the circle $\mathcal{C}=\big\{\boldsymbol{R}=(X,Y,Z)^T\,\big|\,X^2{+}Y^2=R^2\,,\,Z{=}0\big\}$ of radius R around $\boldsymbol{R}=0$ in the X-Y plane. Suppose $\boldsymbol{A}_0'(\boldsymbol{R})=\boldsymbol{A}_0(\boldsymbol{R})-(-Y,X,0)^T/(X^2{+}Y^2)$. *Locally*, we have $\boldsymbol{A}_0'(\boldsymbol{R})=$

$\boldsymbol{A}_0(\boldsymbol{R}) - (\partial/\partial\boldsymbol{R})\arctan(Y/X)$. However, the map $\varphi(\boldsymbol{R}) = \arctan(Y/X)$ is discontinuous on the $X = 0$ plane, and $\partial\varphi(\boldsymbol{R})/\partial\boldsymbol{R}$ is singular on the Z axis. In fact, it is straightforward to see that $\oint_{\mathcal{C}}\big(\partial\varphi(\boldsymbol{R})/\partial\boldsymbol{R}\big)\cdot d\boldsymbol{R} = 2\pi \neq 0$.
Generally, under an arbitrary gauge transformation, we have

$$\gamma_{\mathcal{C}} = \oint_{\mathcal{C}} \boldsymbol{A}_0(\boldsymbol{R})\cdot d\boldsymbol{R} \mapsto \gamma_{\mathcal{C}} + 2\pi k \tag{13}$$

with $k \in \mathbb{Z}$. This is equivalent with the statement that the Berry phase factor $e^{i\gamma_{\mathcal{C}}}$ is always gauge invariant. To verify this, we rewrite the integral as the limit of a Riemann sum

$$\begin{aligned}\exp\left(i\gamma_{\mathcal{C}}\right) &= \exp\left(i\oint_{\mathcal{C}} \boldsymbol{A}_0(\boldsymbol{R})\cdot d\boldsymbol{R}\right) = \lim_{n\to\infty}\exp\Big(i\boldsymbol{A}_0(\boldsymbol{R}_n)\cdot\Delta\boldsymbol{R}_n + \cdots + i\boldsymbol{A}_0(\boldsymbol{R}_1)\cdot\Delta\boldsymbol{R}_1\Big)\\ &= \lim_{n\to\infty}\big(1+i\boldsymbol{A}_0(\boldsymbol{R}_n)\cdot\Delta\boldsymbol{R}_n\big)\cdots\big(1+i\boldsymbol{A}_0(\boldsymbol{R}_1)\cdot\Delta\boldsymbol{R}_1\big)\,.\end{aligned} \tag{14}$$

This is a Volterra product integral. The k-th factor can be written as $1 + i\boldsymbol{A}_0(\boldsymbol{R}_k)\cdot\Delta\boldsymbol{R}_k = 1 - \big\langle\Phi_0(\boldsymbol{R}_k)\big|\Big(\big|\Phi_0(\boldsymbol{R}_k)\big\rangle - \big|\Phi_0(\boldsymbol{R}_k-\Delta\boldsymbol{R}_k)\big\rangle\Big) = \big\langle\Phi_0(\boldsymbol{R}_k)\big|\Phi_0(\boldsymbol{R}_k-\Delta\boldsymbol{R}_k)\big\rangle$, with the help of Eq. (7). We have $\boldsymbol{R}_k - \Delta\boldsymbol{R}_k = \boldsymbol{R}_{k-1}$ and, since $\mathcal{C}$ is a closed path, we can identify $\boldsymbol{R}_0$ with $\boldsymbol{R}_n$. This yields

$$\begin{aligned}\exp\left(i\gamma_{\mathcal{C}}\right) &= \lim_{n\to\infty}\big\langle\Phi_0(\boldsymbol{R}_n)\big|\Phi_0(\boldsymbol{R}_{n-1})\big\rangle\cdots\big\langle\Phi_0(\boldsymbol{R}_1)\big|\Phi_0(\boldsymbol{R}_0)\big\rangle\\ &= \lim_{n\to\infty}\mathrm{tr}\Big(\big|\Phi_0(\boldsymbol{R}_n)\big\rangle\big\langle\Phi_0(\boldsymbol{R}_n)\big|\cdots\big|\Phi_0(\boldsymbol{R}_1)\big\rangle\big\langle\Phi_0(\boldsymbol{R}_1)\big|\Phi_0(\boldsymbol{R}_0)\big\rangle\big\langle\Phi_0(\boldsymbol{R}_0)\big|\Big)\,.\end{aligned} \tag{15}$$

where the latter representation is manifestly gauge invariant, since the phases cancel out in each of the dyadic products.

Example: Toy model with two-dimensional Hilbert space

As an example, we will consider the following toy Hamiltonian on a three-dimensional parameter manifold

$$H = -\frac{1}{2}\boldsymbol{R}\cdot\boldsymbol{\tau}\,, \tag{16}$$

Here $\boldsymbol{R} \in \mathcal{M} = \mathbb{R}^3 \setminus \{0\}$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)^T$ is the vector of Pauli matrices. Actually, this form of H is generic for a quantum system with a two-dimensional Hilbert space. The 2×2 matrix representing the Hamiltonian can be expanded in the Hermitian basis $\{\mathbf{1}, \tau_1, \tau_2, \tau_3\}$, which yields Eq. (16) after dropping the uninteresting term $\propto \mathbf{1}$.
When interpreting the three parameters $\boldsymbol{R}$ as the components of a field strength $\boldsymbol{B}$, the model Eq. (16) just describes a local spin $1/2$ in an external magnetic field. We thus expect a Zeeman splitting of the two eigenstates of H. In fact we have $E_0(\boldsymbol{R}) = -R/2$ and $E_1(\boldsymbol{R}) = +R/2$ where $R = |\boldsymbol{R}|$. Note that we have intentionally excluded the point $\boldsymbol{R} = 0$ from the manifold $\mathcal{M}$, as here the ground state would be degenerate and our entire construction would break down. The $\boldsymbol{R}$-dependent normalized ground state of the model is readily calculated and is conveniently expressed in terms of spherical coordinates $\boldsymbol{R} = R\big(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta\big)^T$ as

$$\big|\Phi_0(\boldsymbol{R})\big\rangle = \big|\Phi_0(\theta,\phi)\big\rangle = \begin{pmatrix}\cos(\theta/2)\\ e^{i\phi}\sin(\theta/2)\end{pmatrix}\,. \tag{17}$$

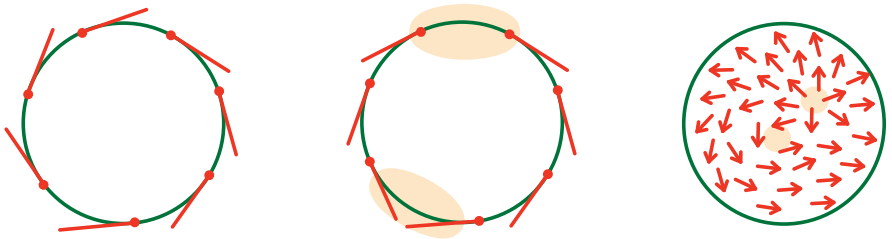


Fig. 2: *Hairy-ball theorem: There is no everywhere nonvanishing and continuous tangent vector field on an n-sphere, if n is even.* Left*: Sketch of a continuous tangent vector field for $n = 1$.* Middle*: Tangent vector field with two twirls for $n = 1$.* Right*: Tangent vector field on S^2 with (already) two twirls. At each $\boldsymbol{R} \in S^2$, the direction of the arrow with unit length specifies the phase $\varphi(\boldsymbol{R})$.*

In fact, the ground state is independent of R, and it is *almost everywhere* a smooth and invertible function of θ and ϕ. The angle variables provide us with a chart of the 2-sphere $S^2 = \{\boldsymbol{R} \,|\, R{=}1\}$. However, the state that is obtained in the limit $\theta \to \pi$, i.e., when approaching the south pole of S^2, will depend on ϕ. This means that the map $(\theta, \phi) \mapsto \boldsymbol{R} \mapsto |\Phi_0(\boldsymbol{R})\rangle$ is discontinuous at the south pole. Hence, the parameter manifold S^2 cannot be described by a single chart. A second one is necessary for a complete atlas.

Choosing a different gauge is another way to deal with the singularity since this will not affect any *physical*, gauge-invariant properties. In fact, the gauge transformation $|\Phi_0(\boldsymbol{R})\rangle \mapsto e^{-i\phi}|\Phi_0(\boldsymbol{R})\rangle$ does the job. It removes the singularity at the south pole — but it also introduces a similar one at the north pole. Clearly, the problem cannot be solved with a *continuous* gauge transformation, and the proposed one is itself *discontinuous*. As we have seen above, $\boldsymbol{R} = (X, Y, Z)^T \mapsto \varphi(\boldsymbol{R}) = \phi = \arctan(Y/X)$ is discontinuous on the entire $X = 0$ plane. Generally, there is no gauge such that the phase of $|\Phi_0(\boldsymbol{R})\rangle$ is continuous on the entire manifold $\mathcal{M}$, which essentially is the 2-sphere. The 2-sphere, with a phase factor $e^{i\varphi(\boldsymbol{R})} \in U(1) \cong S^1$ attached to each point, is like a hairy sphere, so that we can apply the "hairy-ball theorem" [4], see Fig. 2. This states that any attempt to comb a hairy billiard ball (the 2-sphere) smoothly, without twirls, must fail and that there must be at least one singular point.

On the equator $\theta = \pi/2$, oriented from west to east, and for arbitrary $\phi \in [0, 2\pi[$, the ϕ component of the Berry connection obtained from Eq. (17) is

$$\boldsymbol{A}_0(\theta, \phi)\, R\boldsymbol{e}_\phi = i\langle \Phi_0(\theta, \phi)|\partial_\phi|\Phi_0(\theta, \phi)\rangle = -\frac{1}{2}\,, \tag{18}$$

where $\boldsymbol{e}_\phi = (1/R)\partial\boldsymbol{R}/\partial\phi$. Here, we have used the definition Eq. (12) and the representation $\partial/\partial\boldsymbol{R} = \boldsymbol{e}_\phi(R\sin\theta)^{-1}(\partial/\partial\phi) + \boldsymbol{e}_\theta(\cdots) + \boldsymbol{e}_R(\cdots)$ with $\theta = \pi/2$ and $R = 1$. Computing the line integral along the equator, where $d\boldsymbol{R} = R\boldsymbol{e}_\phi\, d\phi$, yields the Berry phase

$$\gamma_{\text{equator}} = \oint_{\text{equator}} \boldsymbol{A}_0(\boldsymbol{R}){\cdot}d\boldsymbol{R} = \int_0^{2\pi} \boldsymbol{A}_0(\theta, \phi){\cdot}R\boldsymbol{e}_\phi\, d\phi = -\pi\,. \tag{19}$$

In the context of adiabatic dynamics, this has the interesting consequence that the system state $|\Psi\rangle$ picks up a sign $e^{i\gamma_C} = -1$ when $\boldsymbol{R}$ is slowly steered once around the equator C.

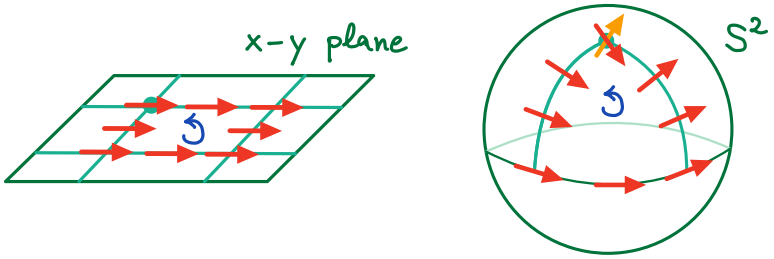


Fig. 3: *Parallel transport along a closed path on a flat and on a curved space.*

In the above-discussed alternative gauge after the transformation $\big|\Phi_0(\boldsymbol{R})\big\rangle \mapsto e^{-i\phi}\big|\Phi_0(\boldsymbol{R})\big\rangle = \big(e^{-i\phi}\cos(\theta/2), \sin(\theta/2)\big)^T$, we get a different result for the Berry connection, $\boldsymbol{A}_0'(\theta,\phi)\cdot R\boldsymbol{e}_\phi = +\frac{1}{2}$, showing again that $\boldsymbol{A}_0'(\boldsymbol{R})$ is not gauge invariant. Computing the corresponding Berry phase yields $\gamma_{\text{equator}} = +\pi$. This is consistent with our general reasoning: The Berry phase $\gamma_{\mathcal{C}}$ is gauge invariant up to an integer multiple of 2π only, but the Berry phase factor $e^{i\gamma_{\mathcal{C}}}$ is fully gauge invariant.

Berry curvature

More elegantly, we can compute the Berry phase $\gamma_{\mathcal{C}}$ with the help of the Stokes theorem

$$\gamma_{\mathcal{C}} = \oint_{\mathcal{C}} \boldsymbol{A}_0(\boldsymbol{R})\cdot d\boldsymbol{R} = \int_{\mathcal{S}} \boldsymbol{\Omega}_0(\boldsymbol{R})\cdot d\boldsymbol{S}\ . \tag{20}$$

Here $\boldsymbol{S}$ is the unit normal vector of the surface $\mathcal{S}$ on the sphere of fixed radius R that is bounded by the closed path $\mathcal{C} = \partial\mathcal{S}$, and $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}}\times\boldsymbol{A}_0(\boldsymbol{R})$ is the curl of the Berry connection. The orientation of the path $\mathcal{C}$ is inherited from the orientation of $\mathcal{S}$ in the usual way by the right-hand rule. $\boldsymbol{\Omega}_0(\boldsymbol{R})$ is called the (ground-state) Berry curvature. In this form, the Stokes theorem is formulated for applications to a three-dimensional parameter space only, and we will stick to the above toy model. Qualitatively, however, the entire discussion holds for arbitrary r-dimensional manifolds $\mathcal{M}$, and a brief discussion of the general case will be given later.

The Berry phase γ_{C} is a quantity, which measures the degree to which parallel transport fails to preserve the geometrical data that is being transported. Parallel transport along closed loops is in a way trivial on flat spaces but leaves a detectable footprint for a curved space, see Fig. 3. In fact, Eq. (20) demonstrates that a finite Berry phase requires a finite Berry curvature $\boldsymbol{\Omega}_0(\boldsymbol{R})$. Eq. (19) shows that parallel transport of states in Hilbert space does not follow descriptive Euclidean geometry. Rather, it takes two circumnavigations of the equator instead of just one to get back to the same state.

The Berry curvature is a highly convenient quantity since it is invariant under arbitrary gauge transformations. This follows from its definition, from the transformation behavior (12) of the Berry connection, and from the fact that the curl of any gradient field vanishes. Another important property is $\partial_{\boldsymbol{R}}\boldsymbol{\Omega}_0(\boldsymbol{R}) = 0$, i.e., the Berry curvature is divergence-free. Again, this holds by definition since $\partial_{\boldsymbol{R}}\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}}\cdot\partial_{\boldsymbol{R}}\times\boldsymbol{A}_0(\boldsymbol{R}) = 0$.

Vice versa, there is the mathematical question, under which conditions a divergence-free vector field $\boldsymbol{\Omega}_0(\boldsymbol{R})$ can be expressed as the curl of another vector field on the entire manifold $\mathcal{M}$. This is answered by the following form of the Poincaré lemma: If the manifold $\mathcal{M}$ is an open ball or, more generally, a star-shaped open set in $\mathbb{R}^3$, one can find a single field $\boldsymbol{A}_0(\boldsymbol{R})$ such that $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}} \times \boldsymbol{A}_0(\boldsymbol{R})$ *globally*, i.e., for *all* $\boldsymbol{R} \in \mathcal{M}$. Note that *locally* a representation as the curl of a vector field is always possible.

In case of topologically more complicated manifolds, a global representation may not be possible. The manifold $\mathcal{M} = \mathbb{R}^3 \backslash \{0\}$ relevant for the toy model or, since the ground state does not depend on R, the submanifold S^2, are in fact examples for manifolds that cannot be covered by a single chart and are thus topologically different from $\mathbb{R}^3$ or from $\mathbb{R}^2$, respectively. Consequently, the Stokes theorem Eq. (20) cannot be applied if $\mathcal{S} = S^2$. We would mistakenly get $\gamma_{\mathcal{C}} = 0$ since the path $\mathcal{C}$ vanishes in this case.

Let us compute the Berry curvature for our toy model concretely. Its α-th component

$$\Omega_{0,\alpha}(\boldsymbol{R}) = \frac{1}{2} \sum_{\beta\gamma} \varepsilon_{\alpha\beta\gamma}\, \Omega_{0,\beta\gamma}(\boldsymbol{R}) \tag{21}$$

can be expressed in terms of an antisymmetric real matrix

$$\begin{aligned} \Omega_{0,\beta\gamma}(\boldsymbol{R}) &= \partial_\beta A_{0,\gamma}(\boldsymbol{R}) - \partial_\gamma A_{0,\beta}(\boldsymbol{R}) = i\Big(\big\langle \partial_\beta \Phi_0(\boldsymbol{R}) \big| \partial_\gamma \Phi_0(\boldsymbol{R}) \big\rangle - \big\langle \partial_\gamma \Phi_0(\boldsymbol{R}) \big| \partial_\beta \Phi_0(\boldsymbol{R}) \big\rangle\Big) \\ &= -2\,\mathrm{Im} \big\langle \partial_\beta \Phi_0(\boldsymbol{R}) \big| \partial_\gamma \Phi_0(\boldsymbol{R}) \big\rangle = -2\,\mathrm{Im} \sum_{j\neq 0} \big\langle \partial_\beta \Phi_0(\boldsymbol{R}) \big| \Phi_j(\boldsymbol{R}) \big\rangle \big\langle \Phi_j(\boldsymbol{R}) \big| \partial_\gamma \Phi_0(\boldsymbol{R}) \big\rangle\,. \end{aligned} \tag{22}$$

For the second equality, we have used Eq. (7) and the fact that partial derivatives commute in case of a smooth $\boldsymbol{R}$ dependence. For the last, we have inserted a resolution of the identity (at $\boldsymbol{R}$). Note that the term $j = 0$ does not contribute as for $j = 0$ the matrix elements are real. Using the identity $\big\langle \Phi_j(\boldsymbol{R}) \big| \partial_\gamma \Phi_0(\boldsymbol{R}) \big\rangle = \big\langle \Phi_j(\boldsymbol{R}) \big| \partial_\gamma H(\boldsymbol{R}) \big| \Phi_0(\boldsymbol{R}) \big\rangle / \big(E_0(\boldsymbol{R}) - E_j(\boldsymbol{R})\big)$, valid for $j \neq 0$, we finally have

$$\Omega_{0,\beta\gamma}(\boldsymbol{R}) = -2\,\mathrm{Im} \sum_{j\neq 0} \frac{\big\langle \Phi_0(\boldsymbol{R}) \big| \partial_\beta H(\boldsymbol{R}) \big| \Phi_j(\boldsymbol{R}) \big\rangle \big\langle \Phi_j(\boldsymbol{R}) \big| \partial_\gamma H(\boldsymbol{R}) \big| \Phi_0(\boldsymbol{R}) \big\rangle}{\big(E_0(\boldsymbol{R}) - E_j(\boldsymbol{R})\big)^2}\,. \tag{23}$$

In case of the toy model, the j-sum merely consists of a single term, referring to the excited state $\big|\Phi_1(\boldsymbol{R})\big\rangle$. Furthermore, $\partial_\beta H(\boldsymbol{R}) = -\tau_\beta/2$, see Eq. (16), and $\big(E_0(\boldsymbol{R}) - E_1(\boldsymbol{R})\big)^2 = R^2$. Hence

$$\Omega_{0,\beta\gamma}(\boldsymbol{R}) = -\frac{1}{2}\,\mathrm{Im} \frac{\big\langle \Phi_0(\boldsymbol{R}) \big| \tau_\beta \big| \Phi_1(\boldsymbol{R}) \big\rangle \big\langle \Phi_1(\boldsymbol{R}) \big| \tau_\gamma \big| \Phi_0(\boldsymbol{R}) \big\rangle}{R^2} \tag{24}$$

and with Eq. (21)

$$\boldsymbol{\Omega}_0(\boldsymbol{R}) = \frac{1}{2} \sum_{\alpha\beta\gamma} \varepsilon_{\alpha\beta\gamma}\, \Omega_{0,\beta\gamma}(\boldsymbol{R})\, \boldsymbol{e}_\alpha = -\frac{1}{4}\,\mathrm{Im} \frac{\big\langle \Phi_0(\boldsymbol{R}) \big| \boldsymbol{\tau} \big| \Phi_1(\boldsymbol{R}) \big\rangle \times \big\langle \Phi_1(\boldsymbol{R}) \big| \boldsymbol{\tau} \big| \Phi_0(\boldsymbol{R}) \big\rangle}{R^2}\,. \tag{25}$$

We can choose the z axis to point in the direction of $\boldsymbol{R}$, such that $\big|\Phi_0(\boldsymbol{R})\big\rangle = (1,0)^T$ with Eq. (17) and $\big|\Phi_1(\boldsymbol{R})\big\rangle = e^{i\phi}(0,1)^T$. Evaluating the matrix element

$$\big\langle\Phi_0(\boldsymbol{R})\big|\boldsymbol{\tau}\big|\Phi_1(\boldsymbol{R})\big\rangle = e^{i\phi}(1,0)\,\boldsymbol{\tau}\begin{pmatrix}0\\1\end{pmatrix} = e^{i\phi}\begin{pmatrix}1\\-i\\0\end{pmatrix}, \quad \big\langle\Phi_1(\boldsymbol{R})\big|\boldsymbol{\tau}\big|\Phi_0(\boldsymbol{R})\big\rangle = e^{-i\phi}\begin{pmatrix}1\\i\\0\end{pmatrix}, \tag{26}$$

and remembering that we had assumed $\boldsymbol{R} = R\boldsymbol{e}_z$, we find

$$\boldsymbol{\Omega}_0(\boldsymbol{R}) = -\frac{1}{2}\frac{\boldsymbol{R}}{R^3}\,. \tag{27}$$

Since $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}}\times\boldsymbol{A}_0(\boldsymbol{R})$, one may ask the question, if there is a similarly compact result for the Berry connection. In fact, one may verify by a direct calculation that, for an arbitrarily chosen unit vector $\boldsymbol{e}$, the curl of

$$\boldsymbol{A}_0(\boldsymbol{R}) = -\frac{1}{2}\frac{1}{R^2}\,\frac{\boldsymbol{e}\times\boldsymbol{R}}{1+\boldsymbol{e}\boldsymbol{R}/R} \tag{28}$$

yields the Berry curvature (27), as has already been noted by P. Dirac in 1931 [10].

Equations (27) and (28) are interesting as they allow us to interpret $\boldsymbol{A}_0(\boldsymbol{R})$ as the vector potential of the magnetic field $\boldsymbol{\Omega}_0(\boldsymbol{R})$ of a "magnetic monopole" with "magnetic point charge" $q=-1/2$ located at the origin $\boldsymbol{R}=0$. Note that $\boldsymbol{A}_0(\boldsymbol{R})$ is singular for all $\boldsymbol{R}$ with $\boldsymbol{R}=-R\boldsymbol{e}$, i.e., on the ray (Dirac string) starting at $\boldsymbol{R}=0$ to infinity in $-\boldsymbol{e}$ direction. Different choices for $\boldsymbol{e}$ just correspond to different gauges. The singular point $\boldsymbol{R}=-\boldsymbol{e}$, where the ray intersects S^2, can be moved around on S^2 by choosing different gauges but it cannot be removed for any particular gauge. Hence, $\boldsymbol{\Omega}_0(\boldsymbol{R})=\partial_{\boldsymbol{R}}\times\boldsymbol{A}_0(\boldsymbol{R})$ holds locally but not *globally* on S^2 or on $\mathbb{R}^3\backslash\{0\}$. This is consistent with our earlier discussion that the Poincaré lemma does not apply to S^2.

There is also a descriptive meaning of the Berry phase $\gamma_{\mathcal{C}}$. Equation (20) tells us that $\gamma_{\mathcal{C}}$ is the flux of the "magnetic field" through a surface that is bounded by the closed path $\mathcal{C}$. The right-hand rule tells us that, for $\mathcal{C}$ being the equator, oriented from west to east, $\mathcal{S}$ is the upper hemisphere. Using Eq. (27) with $d\boldsymbol{S}=R^2(\boldsymbol{R}/R)\,dS$, we see that $\gamma_{\mathcal{C}} = -\frac{1}{2}\int_S(\boldsymbol{R}/R^3)d\boldsymbol{S} = -\pi$ for any great circle on the sphere of arbitrary radius. This generalizes our earlier result, Eq. (19). We can also integrate over the lower hemisphere $\pi/2<\theta<\pi$ to compute the Berry phase via the Stokes theorem. According to Eq. (27), the collected flux is the same: $\int_S\boldsymbol{\Omega}_0(\boldsymbol{R})d\boldsymbol{S} = -\pi$. However, in applying the Stokes theorem with the same path $\mathcal{C}$ (the equator) we have to take care of an additional sign, $\oint_{\mathcal{C}}\boldsymbol{A}_0(\boldsymbol{R})d\boldsymbol{R} = -\int_{\mathcal{S}}\boldsymbol{\Omega}_0(\boldsymbol{R})\,d\boldsymbol{S}$, since the orientations of $\mathcal{C}$ and $\mathcal{S}$ are not consistent in this case (the right-hand rule is violated). There is a second point to be considered: The Stokes theorem holds if $\boldsymbol{A}_0(\boldsymbol{R})$ is smooth on (an open set containing) the lower hemisphere, which is not the case for the Berry connection derived from Eq. (17). In the above-discussed alternative gauge after the transformation $|\Phi_0(\boldsymbol{R})\rangle\mapsto e^{-i\phi}|\Phi_0(\boldsymbol{R})\rangle$, the Berry connection $\boldsymbol{A}_0'(\boldsymbol{R})$ is in fact smooth on the lower hemisphere. However, $\oint_{\mathcal{C}}\boldsymbol{A}_0'(\boldsymbol{R})d\boldsymbol{R} = \pi$, opposed to Eq. (19), but consistent with the Stokes theorem: $\int_{\mathcal{S}}\boldsymbol{\Omega}_0(\boldsymbol{R})\,d\boldsymbol{S} = -\oint_{\mathcal{C}}\boldsymbol{A}_0'(\boldsymbol{R})\,d\boldsymbol{R} = -\pi$.

Chern number

We are now in a position to define the Chern number and to understand why this is a topological invariant. Given the Hamiltonian $H(\boldsymbol{R})$ of a quantum system with a nondegenerate ground state for all $\boldsymbol{R}$ on its parameter manifold $\mathcal{M}$, our goal is to show that the first Chern number, defined as

$$C = \frac{1}{2\pi} \oint_{\mathcal{S}} \boldsymbol{\Omega}_0(\boldsymbol{R}) \cdot d\boldsymbol{S} \,, \tag{29}$$

is quantized: $C \in \mathbb{Z}$.

Here, we consider a *closed* two-dimensional surface $\mathcal{S}$ embedded in a two- or higher-dimensional manifold $\mathcal{M}$. The Chern number will generally be dependent on our choice for $\mathcal{S}$. For simplicity, we still stick to a three-dimensional parameter space, e.g., to $\mathcal{M} = \mathbb{R}^3 \backslash \{0\}$, which applies to our toy model. The Berry curvature $\boldsymbol{\Omega}_0(\boldsymbol{R})$ is divergence-free. Hence, locally $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}} \times \boldsymbol{A}_0(\boldsymbol{R})$. If $\mathcal{M}$ could be covered by a single chart, i.e., if it is essentially given by $\mathbb{R}^3$, then the Poincaré lemma would tell us that $\boldsymbol{\Omega}_0(\boldsymbol{R})$ is *globally* given as the curl of the Berry connection $\boldsymbol{A}_0(\boldsymbol{R})$ defined on the *entire* manifold $\mathcal{M}$. In this case $C = 0$, Namely, the Stokes theorem gives $\oint_{\mathcal{S}} \partial_{\boldsymbol{R}} \times \boldsymbol{A}_0(\boldsymbol{R}) d\boldsymbol{S} = \oint_{\partial \mathcal{S}} \boldsymbol{A}_0(\boldsymbol{R}) d\boldsymbol{R} = 0$ since the boundary of $\mathcal{S}$ vanishes.

Consider now the case of a nontrivial $\mathcal{M}$ and assume that the closed surface $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are surfaces, which both share the same closed path $\mathcal{C}$ as their boundary, except for the different orientation: $\partial \mathcal{S}_1 = -\partial \mathcal{S}_2 = \mathcal{C}$. Let $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}} \times \boldsymbol{A}_{0,1}(\boldsymbol{R})$ on $\mathcal{S}_1$ and $\boldsymbol{\Omega}_0(\boldsymbol{R}) = \partial_{\boldsymbol{R}} \times \boldsymbol{A}_{0,2}(\boldsymbol{R})$ on $\mathcal{S}_2$. On an environment of $\mathcal{C}$, the Berry connections $\boldsymbol{A}_{0,1}(\boldsymbol{R})$ and $\boldsymbol{A}_{0,2}(\boldsymbol{R})$ satisfy $\partial_{\boldsymbol{R}} \times \big(\boldsymbol{A}_{0,1}(\boldsymbol{R}) - \boldsymbol{A}_{0,2}(\boldsymbol{R})\big) = 0$, and thus differ by a gradient field, i.e., transform into each other via a gauge transformation. Thus, the corresponding Berry phase factors are equal, $e^{i\gamma_{1,\mathcal{C}}} = e^{i\gamma_{2,\mathcal{C}}}$, while the Berry phases are equal modulo $2\pi k$ with $k \in \mathbb{Z}$:

$$\oint_{\mathcal{C}} \boldsymbol{A}_{0,1}(\boldsymbol{R}) \cdot d\boldsymbol{R} - \oint_{\mathcal{C}} \boldsymbol{A}_{0,2}(\boldsymbol{R}) \cdot d\boldsymbol{R} = 2\pi k \,. \tag{30}$$

We choose the orientation of $\mathcal{C}$ consistent with the orientation of $\mathcal{S}_1$ according to the right-hand rule. The Stokes theorem then gives

$$\int_{\mathcal{S}_1} \boldsymbol{\Omega}(\boldsymbol{R}) \cdot d\boldsymbol{S}_1 = \oint_{\mathcal{C}} \boldsymbol{A}_{0,1}(\boldsymbol{R}) \cdot d\boldsymbol{R} \quad \text{and} \quad \int_{\mathcal{S}_2} \boldsymbol{\Omega}(\boldsymbol{R}) \cdot d\boldsymbol{S}_2 = -\oint_{\mathcal{C}} \boldsymbol{A}_{0,2}(\boldsymbol{R}) \cdot d\boldsymbol{R} \,, \tag{31}$$

where the sign in the second equation comes from the fact that the orientations of $\mathcal{C}$ and $\mathcal{S}_2$ are opposite. With Eq. (30) and Eq. (31) we conclude that

$$\oint_{\mathcal{S}} \boldsymbol{\Omega}(\boldsymbol{R}) \cdot d\boldsymbol{S} = \int_{\mathcal{S}_1} \boldsymbol{\Omega}(\boldsymbol{R}) \cdot d\boldsymbol{S}_1 + \int_{\mathcal{S}_2} \boldsymbol{\Omega}(\boldsymbol{R}) \cdot d\boldsymbol{S}_2 = \oint_{\mathcal{C}} \boldsymbol{A}_{0,1}(\boldsymbol{R}) \cdot d\boldsymbol{R} - \oint_{\mathcal{C}} \boldsymbol{A}_{0,2}(\boldsymbol{R}) \cdot d\boldsymbol{R} \in 2\pi\mathbb{Z}. \tag{32}$$

This is exactly the proposition made and thus concludes the proof.

Inserting the result we have obtained for the toy model Eq. (27) in the definition of the Chern number Eq. (29), we find a nontrivial (nonzero) invariant

$$C = \frac{1}{2\pi} \oint_{\mathcal{S}} \left(-\frac{1}{2} \frac{\boldsymbol{R}}{R^3} \right) \cdot d\boldsymbol{S} = -\frac{1}{2} \frac{1}{2\pi} 4\pi = -1 \in \mathbb{Z} \,, \tag{33}$$

if we choose the closed surface $\mathcal{S}$ such that it once encloses the "magnetic monopole" at $\boldsymbol{R} = 0$. Taking a sphere centered around $\boldsymbol{R} = 0$ is the most simple choice, and the result Eq. (33) is

obtained straightforwardly. Let us emphasize, however, that it also holds for any continuous deformation of the sphere as long as $\boldsymbol{R} = 0$ is included, i.e., if there is no gap closure during the deformation.

We conclude the discussion with two general points: First, it should be stressed that the Chern number is assigned to a *bundle* $\{|\Phi_0(\boldsymbol{R})\rangle\}_{\mathcal{S}}$ of ground states of Hamiltonians $H(\boldsymbol{R})$ with parameters $\boldsymbol{R}$ belonging to a closed two-dimensional manifold $\mathcal{S}$. This can be indicated by writing $C(\mathcal{S})$ or, more generally, $C(\mathcal{S}, \boldsymbol{\lambda})$, where we made explicit the dependence of the Hamiltonian $H = H(\boldsymbol{R}, \boldsymbol{\lambda})$ on additional parameters $\boldsymbol{\lambda}$, which are unrelated to $\boldsymbol{R}$ and which are called *control parameters*. Examples for $\boldsymbol{\lambda}$ will be seen in the next section. A nonzero Chern number can be interpreted as a nontrivial twist of this bundle. We can now state that the bundle $\{|\Phi_0(\boldsymbol{R}, \boldsymbol{\lambda}_1)\rangle\}_{\mathcal{S}}$ cannot be deformed continuously into the bundle $\{|\Phi_0(\boldsymbol{R}, \boldsymbol{\lambda}_2)\rangle\}_{\mathcal{S}}$, if the corresponding Chern numbers differ from each other, i.e., if $C(\mathcal{S}, \boldsymbol{\lambda}_1) \neq C(\mathcal{S}, \boldsymbol{\lambda}_2)$. In passing from $\boldsymbol{\lambda}_1$ to $\boldsymbol{\lambda}_2$ in control-parameter space, one would necessarily have to go through a critical point $\boldsymbol{\lambda}_c$, where there is a gap closure at one (or several) points $\boldsymbol{R} \in \mathcal{S}$, since the Chern number cannot change as long the ground state is nondegenerate.

Second, the entire construction can be generalized from three-dimensional parameter manifolds $\mathcal{M}$ to manifolds of arbitrary dimension r using the calculus of differential forms and the generalized Stokes theorem. In this language, the Berry connection is a 1-form, and the Berry curvature Ω a 2-form, which can be integrated over a closed two-dimensional submanifold of $\mathcal{M}$, and are thus relevant for $r = 3$. By construction Ω is *closed*, i.e., $d\Omega = 0$. In case of a topologically nontrivial manifold $\mathcal{M}$, however, it is not necessarily *exact*, i.e., of the form $\Omega = dA$. For a bundle of Hamiltonians $H(\boldsymbol{R})$ with parameters $\boldsymbol{R}$ living in an $r = 2n+1$-dimensional manifold $\mathcal{M}$, and for a given closed $2n$-dimensional submanifold $\mathcal{S}$ of $\mathcal{M}$, one can associate the n-th Chern number $C_{2n}(\mathcal{S})$. It is defined as a $2n$-dimensional integral over the $2n$-form $\frac{i^n}{(2\pi)^n}\frac{1}{n!}\int_{\mathcal{S}} \mathrm{tr}(\Omega^n)$ over $\mathcal{S}$, where $\Omega = dA + A^2$ is the Berry-curvature 2-form given in terms of the nonabelian (matrix-valued) Berry-connection 1-form A. Let us also mention that the case of even r must be treated differently. The details of the mathematical formulation are too complicated to be discussed here. One may have a look at Refs. [3, 11–13].

3 Chern insulator

Our goal is to apply the concept of the Chern number as a topological invariant to a system of noninteracting electrons on a lattice. An additional Hubbard interaction is discussed later.

Generic model of a noninteracting insulator

A lattice model for noninteracting electrons is uniquely characterized by a hopping matrix $\boldsymbol{t}$ and has the Hamiltonian

$$H = \sum_{i,i'=1}^{L} \sum_{\alpha,\alpha'=1}^{M} t_{i\alpha,i'\alpha'}\, c^\dagger_{i\alpha} c_{i'\alpha'} \,. \tag{34}$$

Here $i = 1, ..., L$ labels the sites of a D-dimensional Bravais lattice of position vectors $\boldsymbol{r}_i$ with periodic boundary conditions, and $\alpha = 1, ..., M$ is an orbital index. For simplicity we consider spinless fermions. Furthermore, $M = 2$ different orbitals per site are necessary at least to allow for a band-insulating state with a gapped band structure. We choose $M = 2$ for simplicity.

We recap a few simple facts from band theory: The lattice is spanned by basis vectors $\boldsymbol{a}_r$ with $r = 1, ..., D$. Hence, $\boldsymbol{r}_i = \sum_{r=1}^{D} n_{ir}\boldsymbol{a}_r$ with $n_{ir} \in \mathbb{Z}$. The discrete Fourier transformation $c^\dagger_{\boldsymbol{k}\alpha} = L^{-1/2}\sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\boldsymbol{r}_i} c^\dagger_{i\alpha}$ block diagonalizes the Hamiltonian: $H = \sum_{\boldsymbol{k}\alpha\alpha'} t_{\alpha\alpha'}(\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}\alpha} c_{\boldsymbol{k}\alpha'}$. Here, the wave vectors $\boldsymbol{k}$ lie in the unit cell of the reciprocal lattice spanned by the basis vectors $\boldsymbol{b}_s$ with $s = 1, ..., D$, defined as the unique solution of the inhomogeneous linear system of equations $\boldsymbol{a}_r\boldsymbol{b}_s = 2\pi\delta_{rs}$. Periodic boundary conditions for the real-space lattice imply that the wave-vector components are discrete: $k_s = (m_s/L_s)\boldsymbol{b}_s$ with $m_s = 0, ..., L_s{-}1$. There are $L = L_1 L_2 \cdots L_D$ wave vectors in any reciprocal unit cell. Typically, one chooses the first Brillouin zone (1BZ). Furthermore, we have $t_{\alpha\alpha'}(\boldsymbol{k}{+}\boldsymbol{g}) = t_{\alpha\alpha'}(\boldsymbol{k})$ for any reciprocal lattice vector $\boldsymbol{g} = \sum_{s=1}^{D} m_s\boldsymbol{b}_s$.

The last point leads to the important observation that *in the thermodynamic limit* $L \to \infty$ the 1BZ is actually a smooth D-dimensional manifold without a boundary, namely a D-dimensional torus $T^D \cong S^1 \times \cdots \times S^1$. This is nontrivial in the sense that the Poincaré lemma does not apply when integrating over the 1BZ. Furthermore, for each $\boldsymbol{k} \in$ 1BZ $\cong T^D$ the "Bloch Hamiltonian" $\boldsymbol{t}(\boldsymbol{k})$ is a $2{\times}2$ matrix (in general $M{\times}M$). We conclude that for $D = 2$ we have exactly the situation as discussed above in the toy-model example: A bundle of Hamiltonians with two-dimensional Hilbert spaces attached to the points of a smooth nontrivial parameter manifold 1BZ $\cong T^2$ (replacing the 2-sphere).

However, the present situation is much more pleasant: The different Bloch Hamiltonians in the bundle actually all derive from one and the same Hamiltonian of the system. Furthermore, the closed two-dimensional surface $\mathcal{S} =$ 1BZ, over which the integration is performed, is an intrinsic property of the system, namely of its geometry. There is no need to indicate the dependence of the Chern number $C(\mathcal{S}, \boldsymbol{\lambda})$ on the choice for $\mathcal{S}$. There is no choice, since $\mathcal{S}$ is given *a priori*. Hence, $C(\boldsymbol{\lambda})$ is a material property and depends on material (control) parameters only. It can be used to topologically classify the ground states in $\boldsymbol{\lambda}$ space.

For each $\boldsymbol{k} \in$ 1BZ we can expand the Bloch Hamiltonian in the basis of Hermitian $2{\times}2$ matrices as $\boldsymbol{t}(\boldsymbol{k}) = d_0(\boldsymbol{k})\mathbf{1} + \boldsymbol{d}(\boldsymbol{k}){\cdot}\boldsymbol{\tau}$. Actually, this form is quite convenient to *define* interesting models, rather than specifying the real-space hopping matrix $t_{i\alpha,i'\alpha'}$. As an example, we will consider the Qi-Wu-Zhang (QWZ) model [14, 2] on the square lattice. This is defined by $d_0(\boldsymbol{k}) = 0$ and

$$\boldsymbol{d}(\boldsymbol{k}) = \begin{pmatrix} t\sin k_x \\ t\sin k_y \\ m + t\cos k_x + t\cos k_y \end{pmatrix} . \tag{35}$$

Here, $t > 0$ is the hopping parameter, which can be fixed to specify the energy scale. Further, m is the so-called "mass" parameter. The eigenvalues of $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k}){\cdot}\boldsymbol{\tau}$ are given by $\epsilon_\pm(\boldsymbol{k}) = \pm|\boldsymbol{d}(\boldsymbol{k})| = \pm t\sqrt{\sin^2 k_x + \sin^2 k_y + \big(m/t{+}\cos k_x{+}\cos k_y\big)^2}$. We see that there is a (fully occupied) valence band and an (unoccupied) conduction band separated by a finite

gap $\Delta(\boldsymbol{k}) = 2|\boldsymbol{d}(\boldsymbol{k})| > 0$ in the entire 1BZ ($-\pi < k_x, k_y \leq \pi$), except for special values of the mass m, for which $|\boldsymbol{d}(\boldsymbol{k})| = 0$. We immediately see that a gap closure requires $\boldsymbol{k} = \boldsymbol{k}_c$ with $\boldsymbol{k}_c \in \big\{(0,0),\, (0,\pi),\, (\pi,0),\, (\pi,\pi)\big\}$, i.e., it can take place at critical wave vectors $\boldsymbol{k}_c$ given by the high-symmetry points (HSPs) in the 1BZ. Furthermore, from the condition $m/t + \cos k_x + \cos k_y = 0$, evaluated at the HSPs, we find the critical mass parameters $m_c = -2t,\, 0,\, 2t$, at which there is a band closure (at $\boldsymbol{k}_c = (0,0)$, $\boldsymbol{k}_c = (\pi,0),\, (0,\pi)$, and $\boldsymbol{k}_c = (\pi,\pi)$, respectively).

Band and Chern insulators

Our toy model Eq. (16) is recovered with the substitution $\boldsymbol{d} \leftrightarrow -\frac{1}{2}\boldsymbol{R}$. Hence, we immediately get the Berry curvature (note that ground and excited states switch their roles and that the prefactor $\frac{1}{2}$ is the same, see Eq. (23))

$$\boldsymbol{\Omega}_0(\boldsymbol{d}) = \frac{1}{2}\frac{\boldsymbol{d}}{d^3}\,. \tag{36}$$

At least this is the Berry curvature when regarding $\boldsymbol{d} = (d_x, d_y, d_z)^T$ as the parameters. One would then have to choose, more or less arbitrarily, a closed two-dimensional surface embedded in three-dimensional $\boldsymbol{d}$ space to compute the Chern number. The *physical* parameter manifold $\mathcal{S}$, however, is the 1BZ. This yields the Chern number as a material property. If the resulting Chern number is nonzero, we refer to such a material as a "Chern insulator", as opposed to a "trivial" band insulator with $C = 0$.

For noncritical mass parameters, Eq. (35) defines a smooth map $\boldsymbol{d} : \mathcal{S} \to \mathbb{R}^3\backslash\{0\}$. Since $\mathcal{S} = \text{1BZ} \cong T^2$ is a closed two-dimensional surface, its image under $\boldsymbol{d}$,

$$\mathcal{D} \equiv \big\{\boldsymbol{d} = \boldsymbol{d}(\boldsymbol{k}) \in \mathbb{R}^3\backslash\{0\} \;\big|\; \boldsymbol{k} = (k_x, k_y) \in \text{1BZ}\big\}\,, \tag{37}$$

is a closed two-dimensional surface as well, which is embedded in $\mathbb{R}^3\backslash\{0\}$. We can define its infinitesimal vectorial surface element as $d\boldsymbol{S} = \frac{\partial \boldsymbol{d}(\boldsymbol{k})}{\partial k_x} \times \frac{\partial \boldsymbol{d}(\boldsymbol{k})}{\partial k_y}\, dk_x dk_y$, which is perpendicular to $\mathcal{D}$. This can be used, together with the substitution rule for functions of several variables, to compute the Chern number as

$$C = \frac{1}{2\pi} \oint_{\mathcal{D}} \left(\frac{1}{2}\frac{\boldsymbol{d}}{d^3}\right) d\boldsymbol{S} = \frac{1}{4\pi} \oint_{\mathcal{S}} \frac{\boldsymbol{d}(\boldsymbol{k})}{d(\boldsymbol{k})^3} \cdot \frac{\partial \boldsymbol{d}(\boldsymbol{k})}{\partial k_x} \times \frac{\partial \boldsymbol{d}(\boldsymbol{k})}{\partial k_y}\, dk_x dk_y\,. \tag{38}$$

The two-dimensional integral in Eq. (38) with the QWZ $\boldsymbol{d}(\boldsymbol{k})$ from Eq. (35) is easily evaluated numerically, and one can verify in this way that $C \in \mathbb{Z}$.

In fact, $C = 0$ for $-\infty < m < -2t$, $C = +1$ for $-2t < m < 0$, $C = -1$ for $0 < m < 2t$, and $C = 0$ for $2t < m < \infty$. This provides us with a topological phase diagram. Systems in the phase diagram with different Chern numbers cannot be deformed continuously into each other. Upon varying m, one necessarily has to pass a critical value m_c, where there is a band closure at a HSP in the 1BZ and thus a degenerate ground state.

Wrapping number

Now that we have the topological phase diagram, we may consider continuous deformations of the Hamiltonian that do not lead to a gap closure and thus ensure that the Chern number does not change. For example, we may treat the modulus of $\boldsymbol{d}(\boldsymbol{k})$ as a $\boldsymbol{k}$-dependent control parameter. A continuous deformation $d(\boldsymbol{k}) = |\boldsymbol{d}(\boldsymbol{k})| \mapsto 1$ does not close the gap $\Delta(\boldsymbol{k}) = 2d(\boldsymbol{k})$. Hence, the Hamiltonian with normalized $d(\boldsymbol{k}) = 1$ has the same Chern number as the original one.
This deformation is advantageous, since this implies $\partial|d(\boldsymbol{k})|^2/\partial k_x = 0$ (and the same for k_y), and thus $\boldsymbol{d}(\boldsymbol{k})$ becomes perpendicular to $\partial\boldsymbol{d}(\boldsymbol{k})/\partial k_x$, and to $\partial\boldsymbol{d}(\boldsymbol{k})/\partial k_y$ as well. Therewith, $\boldsymbol{d}(\boldsymbol{k})$ becomes parallel to the surface element $d\boldsymbol{S}$, and $\boldsymbol{d}(\boldsymbol{k})d\boldsymbol{S} = \pm dS$ with $dS > 0$. Eq. (38) then reads

$$C = \pm\frac{1}{4\pi}\oint_{\mathcal{D}} dS\,, \tag{39}$$

where $\mathcal{D}$ is the image of the 1BZ $\cong T^2$ under the continuous map $T^2 \to \mathbb{R}^3\backslash\{0\}$ with $\boldsymbol{k} \mapsto \boldsymbol{d}(\boldsymbol{k})$, see Eq. (37).
Since the 1BZ $\cong T^2$ is a closed two-dimensional surface and since $d(\boldsymbol{k}) = 1$, the image of T^2 is a closed two-dimensional surface embedded in the 2-sphere S^2. This implies that the image $\mathcal{D}$ must cover the entire S^2 — once or several times. The Chern number is just the area of this surface, divided by 4π. Since the surface area of S^2 is just 4π, we see that the Chern number is just the count, how often S^2 is covered. Its sign reflects the orientation of the image with its surface normal on S^2 pointing inwards or outwards. The Chern number gives us the information how often T^2 wraps around S^2. It can be interpreted as a *wrapping number*.

Time-reversal symmetry

It is important to understand that a nonzero Chern number requires that time-reversal symmetry is broken. The identity and time reversal form a group of transformations isomorphic to $\mathbb{Z}_2$. According to Wigner's theorem [8], time reversal is represented in Hilbert space by a unitary or anti-unitary operator $\mathcal{K}$. In single-particle quantum mechanics, it is defined via its action of the position, momentum and spin operator as $\mathcal{K}\boldsymbol{r}\mathcal{K}^\dagger = \boldsymbol{r}$, $\mathcal{K}\boldsymbol{p}\mathcal{K}^\dagger = -\boldsymbol{p}$, and $\mathcal{K}\boldsymbol{s}\mathcal{K}^\dagger = -\boldsymbol{s}$. This implies $\mathcal{K}[x, p_x]\mathcal{K}^\dagger = -[x, p_x]$, and since $[x, p_x] = i\hbar$ we must have $\mathcal{K}i\hbar\mathcal{K}^\dagger = -i\hbar$. This means $\mathcal{K}$ is anti-unitary, i.e., anti-linear and unitary with $\mathcal{K}^\dagger\mathcal{K} = 1$. For a many-body system with an even (odd) number of spin-1/2 fermions we have $\mathcal{K}^2 = 1$ ($\mathcal{K}^2 = -1$). Here, we consider spinless fermions, where $\mathcal{K}^2 = 1$. On the Fock space of a spinless-fermion model, the anti-unitary operator $\mathcal{K}$ can be defined to act trivially on the annihilation and creation operators with respect to the site-orbital basis, $\mathcal{K}c_{i\alpha}\mathcal{K}^\dagger = c_{i\alpha}$ (and the same for $c^\dagger_{i\alpha}$). Since $c^\dagger_{\boldsymbol{k}\alpha} = L^{-1/2}\sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\boldsymbol{r}_i}c^\dagger_{i\alpha}$, anti-linearity of $\mathcal{K}$ implies $\mathcal{K}c_{\boldsymbol{k}\alpha}\mathcal{K}^\dagger = c_{-\boldsymbol{k}\alpha}$.
Let us assume that the parameter-dependent Hamiltonian $H(\boldsymbol{R})$ of a spinless-fermion system is time-reversal symmetric for all $\boldsymbol{R}$, i.e., $H(\boldsymbol{R}) = \mathcal{K}H(\boldsymbol{R})\mathcal{K}^\dagger$. Hence, we may assume that the eigenstates of $H(\boldsymbol{R})$ can be chosen among the eigenstates of $\mathcal{K}$. In fact, we may choose the phases of $\big|\Phi_j(\boldsymbol{R})\big\rangle$ such that $\mathcal{K}\big|\Phi_j(\boldsymbol{R})\big\rangle = +\big|\Phi_j(\boldsymbol{R})\big\rangle$. These relations can be used to analyze

the matrix element in the representation Eq. (23) of the Berry curvature. We have

$$\begin{aligned}\big\langle \Phi_0(\boldsymbol{R})\big|\partial_\beta H(\boldsymbol{R})\big|\Phi_j(\boldsymbol{R})\big\rangle &= \big\langle \mathcal{K}\Phi_0(\boldsymbol{R})\big|\partial_\beta(\mathcal{K}H(\boldsymbol{R})\mathcal{K}^\dagger)\big|\mathcal{K}\Phi_j(\boldsymbol{R})\big\rangle \\ &= \big\langle \Phi_0(\boldsymbol{R})\big|\mathcal{K}^\dagger\mathcal{K}\partial_\beta H(\boldsymbol{R})\mathcal{K}^\dagger\mathcal{K}\big|\Phi_j(\boldsymbol{R})\big\rangle^* \\ &= \big\langle \Phi_0(\boldsymbol{R})\big|\partial_\beta H(\boldsymbol{R})\big|\Phi_j(\boldsymbol{R})\big\rangle^* , \end{aligned} \tag{40}$$

where we have used that $\langle \mathcal{A}u|v\rangle = \langle u|\mathcal{A}^\dagger v\rangle^*$ for any anti-linear operator $\mathcal{A}$ and arbitrary states $|u\rangle, |v\rangle$. Hence, the matrix element is real. With Eq. (23) we conclude that the Berry curvature vanishes for any time-reversal symmetric Hamiltonian. The Chern number is zero.
Consider a multi-orbital spinless-fermion model, where $H = \sum_{\boldsymbol{k}\alpha\alpha'} t_{\alpha\alpha'}(\boldsymbol{k}) c^\dagger_{\boldsymbol{k}\alpha} c_{\boldsymbol{k}\alpha'}$ is specified by the Bloch Hamiltonian $\boldsymbol{t}(\boldsymbol{k})$. We have

$$\mathcal{K}H\mathcal{K}^\dagger = \mathcal{K}\sum_{\boldsymbol{k}\alpha\alpha'} t_{\alpha\alpha'}(\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}\alpha} c_{\boldsymbol{k}\alpha'}\mathcal{K}^\dagger = \sum_{\boldsymbol{k}\alpha\alpha'} t^*_{\alpha\alpha'}(\boldsymbol{k})\, c^\dagger_{-\boldsymbol{k}\alpha} c_{-\boldsymbol{k}\alpha'} = \sum_{\boldsymbol{k}\alpha\alpha'} t^*_{\alpha\alpha'}(-\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}\alpha} c_{\boldsymbol{k}\alpha'} , \tag{41}$$

and thus H is invariant under time reversal, $\mathcal{K}H\mathcal{K}^\dagger = H$, if $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{t}^*(-\boldsymbol{k})$. Specifically, for the QWZ model $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k})\boldsymbol{\tau}$ with $\boldsymbol{d}(\boldsymbol{k})$ given by Eq. (35). We have

$$\boldsymbol{t}^*(-\boldsymbol{k}) = \boldsymbol{d}(-\boldsymbol{k})\boldsymbol{\tau}^* = \big(-d_x(\boldsymbol{k}), -d_y(\boldsymbol{k}), +d_z(\boldsymbol{k})\big)\cdot\big(\tau_x, -\tau_y, \tau_z\big) , \tag{42}$$

showing that time-reversal symmetry is broken. Thus, a topologically nontrivial phase of the model signalled by a nonzero Chern number is not excluded.

The quantum Hall effect

Typically, a topological invariant is a rather abstract, mathematical concept, and its only significance is to discriminate topologically different phases that cannot be deformed into each other continuously. Sometimes, on top of this important classification property, a topological invariant can be an easily accessible physical observable. The Chern number of a two-dimensional Chern insulator is exactly such a case, and it plays the central role in the quantum Hall effect.
The quantum Hall effect (QHE) is observed in a two-dimensional electron gas, which can be realized, e.g., in gallium-arsenide heterostructures. Applying a magnetic field perpendicular to the heterostructure, the transverse (Hall) resistivity R_{xy} can be measured as function of the magnetic field strength. From the standard theory of electromagnetism and from Drude theory one expects a linear dependence. At sufficiently low temperatures, however, plateaus in the field dependence are found [15]. In fact, the zero-temperature Hall conductivity is quantized: $1/R_{xy} = \nu e^2/h$ with integers ν characteristic for the Hall plateaus. This striking quantization effect has been employed for a new practical standard for electrical resistance as it comes with an extremely high accuracy. A theoretical derivation [16] shows that the integer ν is given by the first Chern number. This means that the quantum Hall effect is eventually traced back to topological properties of the two-dimensional electron gas.
The QWZ model must be seen as a simplified model related to the quantum anomalous Hall effect (QAHE), i.e., the quantum version of the anomalous Hall effect. The latter requires a combination of magnetic polarization and spin-orbit coupling to generate a Hall effect rather than an external magnetic field. In fact, broken time-reversal symmetry rather than an orbital magnetic field and Landau levels is more fundamental for the concept of a Chern insulator [17].

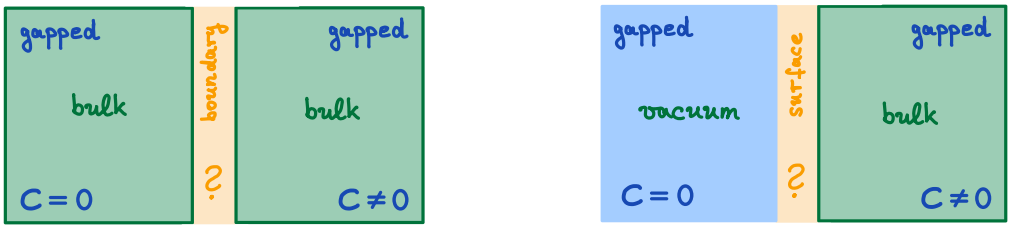


Fig. 4: *Gapless modes must exist at the boundary between two bulk systems with different Chern numbers or at the surface of a Chern insulator.*

Bulk-boundary correspondence

Another important and experimentally observable consequence of the existence of the Chern number as a topological invariant in systems with broken time-reversal symmetry is the necessary presence of gapless modes at the *surfaces* of a Chern insulator. The finite and quantized conductivity in the quantum-Hall systems is in fact carried by these boundary modes. Their number is exactly given by the Chern number, and each carries a flux quantum e^2/h.
The existence of gapless boundary modes can be explained with the so-called bulk-boundary correspondence. This is a rather deep concept in topology but can be sketched in simple terms as follows:
Consider a D-dimensional system that is spatially cut into two subsystems along a $D-1$-dimensional boundary, see Fig. 4. For each of the two D-dimensional subsystems, the bulk band structure shall be gapped and topologically characterized by a Chern number. We assume that the two Chern numbers are different and imagine an extended boundary which smoothly interpolates between the bulk Hamiltonians of the two subsystems. Then, somewhere along a path starting in the bulk of the first subsystem through the boundary and ending in the bulk of the second, the energy gap must vanish. Otherwise it would be impossible for the topological invariant to change. Hence, we conclude that at the boundary between two insulators with different Chern numbers, or, since the vacuum has vanishing Chern number, at the surface of a Chern insulator, there must be low-energy electronic states crossing the gap. These boundary modes have the fascinating and experimentally detectable property of being topologically protected. Their existence is a fundamental consequence of a gapped and topologically nontrivial band structure and thus not affected by arbitrary local perturbations at the boundary.

4 Electron correlations

Up to this point we have concentrated on the topological properties of noninteracting lattice-fermion models. We will now shift the focus to interacting models, such as the QWZ model with an additional Hubbard interaction as an instructive prototype and example for concrete calculations. To this end, we first define the *spinful* QWZ model with Hamiltonian $H^{(0)} = \sum_{\boldsymbol{k}\alpha\alpha'\sigma} t_{\alpha\alpha'}(\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}\alpha\sigma} c_{\boldsymbol{k}\alpha'\sigma}$, where $\sigma = \uparrow, \downarrow$ denotes the spin projection and where $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k})\cdot\boldsymbol{\tau}$ with $\boldsymbol{d}(\boldsymbol{k})$ given by Eq. (35). The QWZ+U model is then given by $H = H^{(0)} + H^{(1)}$ where the (intra-orbital) Hubbard term reads as $H^{(1)} = U \sum_{i\alpha} n_{i\alpha\uparrow} n_{i\alpha\downarrow}$ with $n_{i\alpha\sigma} = c^\dagger_{i\alpha\sigma} c_{i\alpha\sigma}$.

Generally, a lattice-fermion model is defined via its hopping matrix $t_{i\alpha,i'\alpha'}$ (see Eq. (34)) or by the corresponding Bloch Hamiltonian $\boldsymbol{t}(\boldsymbol{k})$ that is obtained after Fourier transformation to reciprocal space and can trivially be made spinful as above (if this is not already the case) and equipped with a Hubbard or with the full Coulomb interaction. We will restrict ourselves to local interactions, such as the Hubbard interaction, since in those cases one can straightforwardly apply dynamical mean-field theory (DMFT).

Dynamical mean-field theory

For the following discussion, we recall some important facts related to DMFT. An overview is given by Ref. [18].
(i) DMFT is an approximate mean-field theory for D-dimensional interacting lattice-fermion models with local interactions. As opposed to other mean-field approaches, it becomes exact in the limit of infinite spatial dimensions $D \to \infty$. This requires a proper scaling of the hopping-matrix elements with D. In case of the hypercubic lattice, for example, the nearest-neighbor hopping must be scaled as $t = t^*/\sqrt{D}$ to maintain the balance between the noninteracting and the interacting parts, $H^{(0)}$ and $H^{(1)}$, of the Hamiltonian in the $D \to \infty$ limit.
(ii) The central quantity of DMFT is the single-particle Green function $\boldsymbol{G}(\boldsymbol{k},\omega)$. In the noninteracting case, its elements are given by

$$G^{(0)}_{\alpha\alpha'}(\boldsymbol{k},\omega) = \left(\frac{1}{\omega+\mu-\boldsymbol{t}(\boldsymbol{k})}\right)_{\alpha\alpha'}, \tag{43}$$

assuming a spin-diagonal and spin-independent hopping. Here, μ is the chemical potential. The interacting Green function is obtained from $\boldsymbol{G}^{(0)}(\boldsymbol{k},\omega)$ in terms of the local, i.e., $\boldsymbol{k}$-independent self-energy $\boldsymbol{\Sigma}(\omega)$ as

$$\boldsymbol{G}(\boldsymbol{k},\omega) = \frac{1}{\left(\boldsymbol{G}^{(0)}(\boldsymbol{k},\omega)\right)^{-1} - \boldsymbol{\Sigma}(\omega)}\,. \tag{44}$$

Locality of the self-energy is the decisive assumption of DMFT and is an exact property in the limit $D \to \infty$.
(iii) Operationally, DMFT self-consistently maps the lattice model onto an Anderson impurity model with the same local interaction term. The self-consistency cycle may be started with a guess for $\boldsymbol{\Sigma}(\omega)$. Via Eq. (44) this fixes $\boldsymbol{G}(\boldsymbol{k},\omega)$ and the local Green function $\boldsymbol{G}_{\text{loc}}(\omega) = \frac{1}{L}\sum_{\boldsymbol{k}} \boldsymbol{G}(\boldsymbol{k},\omega)$ in particular. The latter is used to define the noninteracting Green function matrix $\mathcal{G}(\omega)$ of a (multi-impurity) Anderson impurity model as

$$\mathcal{G}_{\alpha\alpha'}(\omega) = \left(\frac{1}{\left(\boldsymbol{G}_{\text{loc}}(\omega)\right)^{-1} + \boldsymbol{\Sigma}(\omega)}\right)_{\alpha\alpha'}, \tag{45}$$

and thereby fixes its noninteracting Hamiltonian $H^{(0)}_{\text{imp}}$ or action. The interacting impurity problem $H_{\text{imp}} = H^{(0)}_{\text{imp}} + H^{(1)}$ must be solved for the same interaction, chemical potential and temperature T (here assumed as $T = 0$) with the goal to get an updated self-energy that can be fed back to the start of the cycle. The DMFT solution of the lattice problem is obtained by iterating the cycle until self-consistency is reached. Clearly, an "impurity solver" is needed to run the algorithm (see Ref. [18], for examples).

Topological Hamiltonian

A noninteracting electron system on a lattice is specified by the Bloch Hamiltonian $\boldsymbol{t}(\boldsymbol{k})$. We assume that (i) $\boldsymbol{t}(\boldsymbol{k})$ is gapped, i.e., that there is an energy μ_0 such that, *for all* $\boldsymbol{k} \in$ 1BZ, its eigenvalues $\varepsilon_m(\boldsymbol{k}) < \mu_0$ if $m = 1, ..., m_0$ ("occupied bands") and $\varepsilon_m(\boldsymbol{k}) > \mu_0$ if $m = m_0{+}1, ..., M$ ("unoccupied bands"). We further assume that (ii) the total electron number is $N = 2m_0L$. Then, the state $|\chi_0(\boldsymbol{k})\rangle = \prod_{m\le m_0}\prod_\sigma c^\dagger_{\boldsymbol{k}m\sigma}|\text{vac.}\rangle$ can be seen as the ground state of $H(\boldsymbol{k}) = \sum_{m\sigma}\varepsilon_m(\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}m\sigma}c_{\boldsymbol{k}m\sigma} = \sum_{\alpha\alpha'\sigma} t_{\alpha\alpha'}(\boldsymbol{k})c^\dagger_{\boldsymbol{k}\alpha\sigma}c_{\boldsymbol{k}\alpha'\sigma}$ at a given $\boldsymbol{k} \in$ 1BZ in the invariant subspace with $2m_0$ electrons. The ground state of $H^{(0)}$ is $|\Phi_0\rangle = \prod_{\boldsymbol{k}} |\chi_0(\boldsymbol{k})\rangle$.
Apart from the trivial spin degeneracy which can be disregarded here, we have a Hamiltonian $H^{(0)}(\boldsymbol{k})$ with a smooth dependence on the parameters of a two-dimensional closed manifold (the 1BZ) with a unique, i.e., nondegenerate ground state for all $\boldsymbol{k}$ and thus a gap to the lowest excited state. As discussed in the preceding section, we can thus define the corresponding Berry curvature and infer that its integral over the parameter manifold is given by 2π times an integer, the Chern number.
Within DMFT, this argument holds true if we replace $\boldsymbol{t}(\boldsymbol{k})$ by $\boldsymbol{t}_{\text{top}}(\boldsymbol{k}) = \boldsymbol{t}(\boldsymbol{k}) + \boldsymbol{\Sigma}(\omega{=}0)$, *provided that* $\boldsymbol{t}_{\text{top}}(\boldsymbol{k})$ is still gapped and that $\boldsymbol{\Sigma}(\omega)$ does not diverge as $\omega \to 0$. The bundle of matrices $\boldsymbol{t}_{\text{top}}(\boldsymbol{k})$ is called the "topological Hamiltonian" [19]. Note that the self-energy at zero frequency $\boldsymbol{\Sigma}(\omega{=}0)$ is a Hermitian matrix in the orbital indices as can be deduced from the fundamental properties of the single-particle Green function and Dyson's equation (44). We also note that the Slater determinants $|\chi_0(\boldsymbol{k})\rangle$ do not have a direct physical meaning in case of an interacting electron system and must be seen as auxiliary quantities. Nevertheless, the topological Hamiltonian provides us with a means for computing an "interacting Chern number" and thereby with a topological classification of gapped interacting electron systems on a lattice.
We will proceed with the concept of topological Hamiltonian in the context of DMFT. However, in the extended context of interacting topological insulators there are some partly open questions. However, we will postpone their discussion to the final section.

5 Exact interplay of correlations and topology in $D = \infty$

Among the various mean-field theories, DMFT has an exceptional standing, since it is internally consistent and nonperturbative, and since there is a nontrivial limit, in which this mean-field approach becomes exact [20]. This is the limit of infinite spatial dimensions. For the Hubbard model $H = \sum_{ii'\sigma} t_{ii'}\, c^\dagger_{i\sigma}c_{i'\sigma} + U\sum_i n_{i\uparrow}n_{i\downarrow}$ on a D-dimensional hypercubic lattice, one has to properly scale the nearest-neighbor hopping amplitudes to keep a nontrivial balance between the tight-binding and the Hubbard-interaction term in the limit $D \to \infty$, namely $t = t^*/\sqrt{D}$ with fixed t^* (e.g. $t^*{=}1$). In this limit, DMFT provides the exact solution and, while one can profit from certain simplifications such as the locality of the self-energy [21], the remaining physics is highly nontrivial. This is demonstrated, for example, with the famous phase diagram for the Mott metal-insulator transition [18]. Here, we consider the $D \to \infty$ limit of a multi-orbital

Hubbard model, which already for $U = 0$ has a nontrivial topological phase diagram including band and Chern insulator phases. Our goal is to construct the model such that its correlations *and* its topological properties remain nontrivial, such that we are able to study, *on a numerically exact level*, the topological properties of a paradigmatic strongly interacting electron system.

Spinful QWZ model on a D-dimensional hypercubic lattice

To this end, we start from the noninteracting QWZ model, for which an extension to arbitrary even dimensions $D = 2, 4, 6, ...$ is known [12], namely $H^{(0)} = \sum_{k\alpha\alpha'\sigma} t_{\alpha\alpha'}(\boldsymbol{k})\, c^\dagger_{\boldsymbol{k}\alpha\sigma} c_{\boldsymbol{k}\alpha'\sigma}$, where $\alpha = 1, ..., M$ is the orbital index and $t_{\alpha\alpha'}(\boldsymbol{k})$ are the elements of the spin-independent Bloch Hamiltonian

$$\boldsymbol{t}(\boldsymbol{k}) = \left(m + t\sum_{r=1}^{D} \cos k_r\right) \boldsymbol{\gamma}_D^{(0)} + t\sum_{r=1}^{D} \sin k_r\, \boldsymbol{\gamma}_D^{(r)} \,. \tag{46}$$

Here, $t = t^*/\sqrt{D}$ is the hopping and m the mass parameter. We set t^*=1 to fix the energy scale. This Bloch Hamiltonian is an $M\times M$ matrix for each wave vector $\boldsymbol{k} = (k_1, ..., k_D) \in$ 1BZ with $-\pi < k_r \leq \pi$ and is constructed with the help of the $M\times M$ matrices $\gamma_D^{(\mu)}$, where $\mu = 0$ or $\mu = r = 1, ..., D$. These satisfy the anticommutation relations of a Clifford algebra

$$\left\{\boldsymbol{\gamma}_D^{(\mu)},\, \boldsymbol{\gamma}_D^{(\nu)}\right\} = 2\delta^{(\mu\nu)} \,. \tag{47}$$

This algebra is a very convenient one, e.g., for the computation of the eigenvalues of $\boldsymbol{t}(\boldsymbol{k})$. We have $\boldsymbol{t}(\boldsymbol{k}) = \sum_{\mu=0}^{M} d_\mu(\boldsymbol{k})\boldsymbol{\gamma}^{(\mu)}$ with $d_0(\boldsymbol{k}) \equiv m + t\sum_{r=1}^{D} \cos k_r$ and $d_r(\boldsymbol{k}) \equiv t \sin k_r$ for $r = 1, ..., D$. Using Eq. (47) we immediately get

$$\boldsymbol{t}(\boldsymbol{k})^2 = \sum_\mu d_\mu(\boldsymbol{k})\boldsymbol{\gamma}^{(\mu)} \sum_\nu d_\nu(\boldsymbol{k})\boldsymbol{\gamma}^{(\nu)} = \frac{1}{2}\sum_{\mu\nu} d_\mu(\boldsymbol{k}) d_\nu(\boldsymbol{k}) \left(\boldsymbol{\gamma}^{(\mu)}\boldsymbol{\gamma}^{(\nu)} + \boldsymbol{\gamma}^{(\nu)}\boldsymbol{\gamma}^{(\mu)}\right) = \sum_\mu d_\mu(\boldsymbol{k})^2 \,. \tag{48}$$

This means that there are two $M/2$-fold degenerate bands with dispersions given by $\epsilon_\pm(\boldsymbol{k}) = \pm\left(d_0(\boldsymbol{k})^2 + \sum_r d_r(\boldsymbol{k})^2\right)^{1/2}$, when disregarding the spin degree of freedom. Note that, since the γ matrices are traceless, $\boldsymbol{t}(\boldsymbol{k})$ is traceless as well and, therefore, there must be as many negative as positive eigenvalues.

The orbital-dependent density of states (DOS) is given in terms of the free retarded Green function as

$$\rho_\alpha(\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,\frac{1}{L}\sum_{\boldsymbol{k}} G^{(0)}_{\alpha\alpha}(\boldsymbol{k}, \omega + i0^+) \,. \tag{49}$$

Again, thanks to the Clifford algebra, we can easily compute

$$G^{(0)}_{\alpha\alpha}(\boldsymbol{k}, \omega) = \left(\frac{1}{\omega + \mu^{(0)} - \boldsymbol{\epsilon}(\boldsymbol{k})}\right)_{\alpha\alpha} = \left(\frac{1}{\omega - \sum_\mu d_\mu(\boldsymbol{k})\boldsymbol{\gamma}_D^\mu}\right)_{\alpha\alpha} = \left(\frac{\omega + \sum_\mu d_\mu(\boldsymbol{k})\boldsymbol{\gamma}_D^\mu}{\omega^2 - \sum_\mu d_\mu(\boldsymbol{k})^2}\right)_{\alpha\alpha} . \tag{50}$$

We consider a half-filled system and have set the chemical potential of the free system to zero. For $D = 2$, one obtains the QWZ model discussed Sec. 3 when choosing $\boldsymbol{\gamma}_2^{(1)} = \boldsymbol{\tau}_x$ and $\boldsymbol{\gamma}_2^{(2)} = \boldsymbol{\tau}_y$, and the so-called chiral element $\boldsymbol{\gamma}_2^{(0)} = -i\boldsymbol{\tau}_x\boldsymbol{\tau}_y = \boldsymbol{\tau}_z$ for the γ matrices. Note that they satisfy the anticommutation relations Eq. (47), i.e., the Pauli matrices form a two-dimensional representation of the algebra, $M = 2$. We have $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k})\cdot\tau$ with $\boldsymbol{d}(\boldsymbol{k})$ given by Eq. (35).

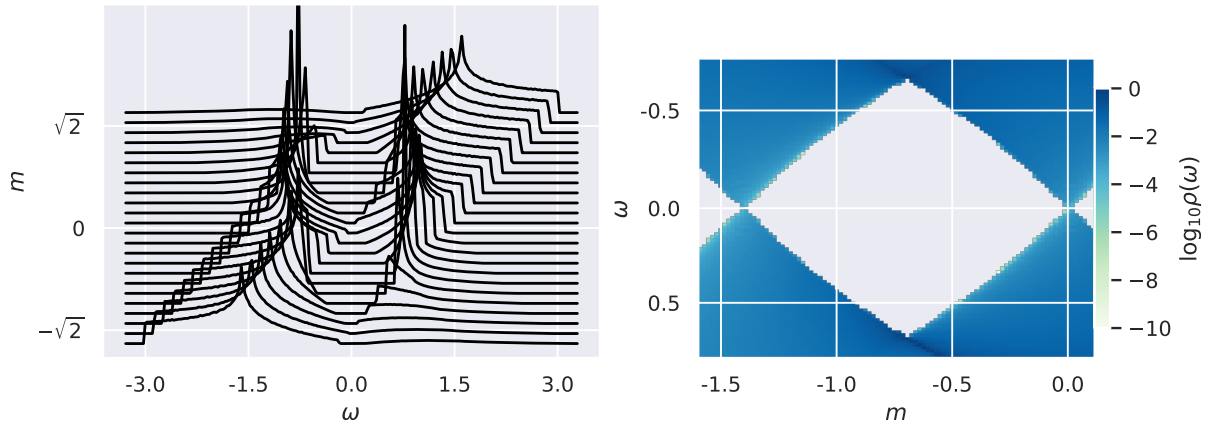


Fig. 5: *Density of states on the A-orbitals for the $D = 2$ QWZ model for different m (left). Right: Density of states (color code) at low excitation energies. $t = t^*/\sqrt{2}$ with $t^* = 1$ sets the energy scale. Figure taken from Ref. [5].*

The matrices $\boldsymbol{\gamma}^{(r)}$ have vanishing diagonal elements, such that there is a contribution from $\boldsymbol{\gamma}^{(0)}$ only. We have $\gamma^{(0)}_{\alpha\alpha} \equiv z_\alpha{=}{+}1$ for "A orbitals" $\alpha = 1, ..., M/2$ and $\gamma^{(0)}_{\alpha\alpha} \equiv z_\alpha{=}{-}1$ for "B orbitals" $\alpha = (M/2){+}1, ..., M$. This holds for general D and M (for the $D = 2$ QWZ model: $M = 2$) and implies that the orbital-resolved free Green functions $G^{(0)}_{\alpha\alpha}(\boldsymbol{k}, \omega)$ for $\alpha = 1, ..., M$ can be divided into two classes with representatives $G^{(0)}_A(\boldsymbol{k}, \omega)$ and $G^{(0)}_B(\boldsymbol{k}, \omega)$.
The local DOS on the A orbitals is displayed in Fig. 5 (left) as a function of the mass parameter m. We note the symmetry $\rho_A(\omega) \to \rho_A(-\omega)$ for $m \to -m$. Furthermore, we have $\rho_A(-\omega) = \rho_B(\omega)$. In the right plot for small frequencies, the evolution of the gap with m can be read off. Gap closures are found at $m = 0$ and $m = \pm\sqrt{2}$ (in units of $t^* = 1$, i.e., $t = 1/\sqrt{2}$), as discussed earlier.

Bulk-boundary correspondence

For $D = 2$ dimensions and for mass parameter $m = 1$ the system is a Chern insulator with $C = -1$. Hence, according to the bulk-boundary correspondence, there must be a gap closure at the surfaces of the bulk system. This is uncovered, e.g., by a calculation for the system in a ribbon geometry, i.e., we assume that the system is infinitely extended (with periodic boundary conditions) is x direction, while in y direction it consists of a finite number d of one-dimensional chains labelled by a chain index $\lambda = 1, ..., d$. As the vacuum (for $\lambda < 1$ and $\lambda > d$) has Chern number $C = 0$, we expect gapless edge modes localized at the two edges.
For the actual calculation we can only profit from translational symmetry in x direction. We perform a two-dimensional Fourier transformation of the bulk Bloch Hamiltonian, Eq. (46), or specifically, for $D = 2$, of $\boldsymbol{t}(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k})\cdot\boldsymbol{\tau}$ with $\boldsymbol{d}(\boldsymbol{k})$ given by Eq. (35) to real space, cut the hopping parameters in Eq. (34) at the two edges, and perform a one-dimensional Fourier transform from x space back to k_x space. The result is a $d{\times}d$ matrix for each wave "vector" k_x, which must be diagonalized numerically. The d eigenvalues as functions of k_x represent the band structure of the QWZ model in a ribbon geometry.

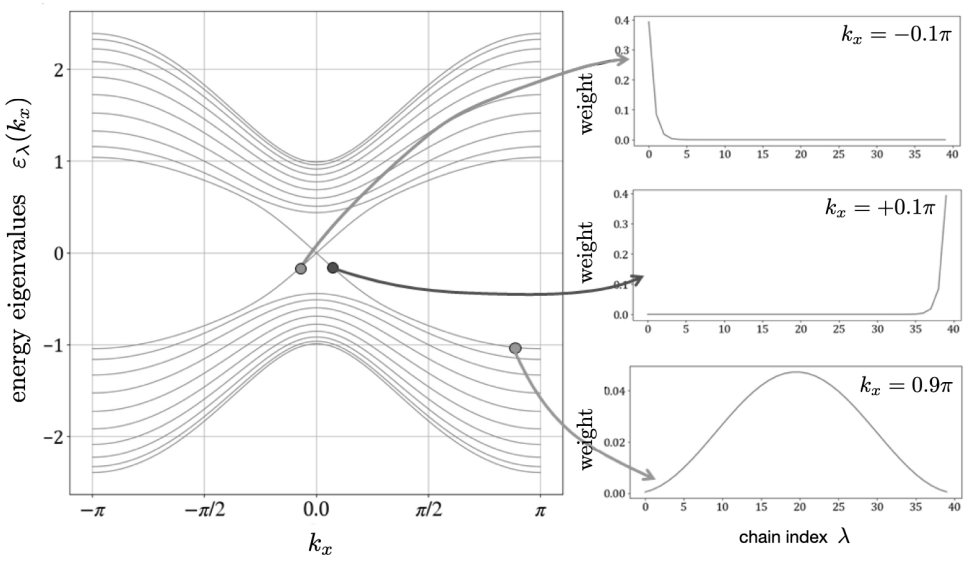


Fig. 6: *Dispersions of the $D = 2$ QWZ model in a ribbon geometry at $m = 1$ (left). Calculations for $d = 10$ chains of infinite length in x direction with edges at $\lambda = 1$ and $\lambda = 10$ in y direction. Right: Weight of various eigenstates in the different chains $\lambda = 1, ..., d$ of a ribbon with thickness $d = 40$, as indicated. Calculations for $t = t^*/\sqrt{2}$ with t^*=1. Figure taken from Ref. [6].*

They are displayed in Fig. 6 (left) for a ribbon thickness of $d = 10$. In fact, we find two low-energy eigenmodes visible around $k_x = 0$, which split off the "bulk continuum", which is here represented by 8 bands only, at wave vectors $k_x \approx \pm 0.36\pi$. Each of these modes bridges the bulk band gap. Furthermore, the analysis of the corresponding eigenstates in the right part of Fig. 6 shows that one edge mode is localized at the top edge $\lambda = 1$ and the other one at the opposite edge $\lambda = 40$. Note that here the calculations have been done for a thicker ribbon with $d = 40$. The corresponding weight factors $|\langle i, \alpha | k_x, \lambda \rangle|^2$ decay exponentially with increasing distance to the edges, opposed to bulk states which extend over the entire ribbon. Let us emphasize once more, that these two gapless edge modes must necessarily exist and that they are topologically protected against local perturbations.

Topological phase diagram of the $D = 2$ QWZ+U model

We now add a local Hubbard-type interaction $H^{(1)} = (U/2) \sum_{i\alpha\sigma} n_{i\alpha\sigma} n_{i\alpha-\sigma}$ to the Hamiltonian and consider the resulting two-dimensional QWZ+U model at half filling. Within the DMFT, the self-energy is site-diagonal, i.e., k-independent. Furthermore, as the $H^{(1)}$ term is an intra-orbital interaction only, it is diagonal in orbital space, $\Sigma_{\alpha\beta}(k, \omega) = \Sigma_{\alpha\beta}(\omega) = \delta_{\alpha\beta}\Sigma_\alpha(\omega)$. Analogous to the discussion of the density of states above, the orbital-dependent diagonal elements $\Sigma_\alpha(\omega)$ can be divided into two classes A and B. With the definition $\Sigma_\pm(\omega) = \frac{1}{2}\big(\Sigma_A(\omega) \pm \Sigma_B(\omega)\big)$, we have the following decomposition

$$\boldsymbol{\Sigma}(\omega) = \Sigma_+(\omega)\mathbf{1} + \Sigma_-(\omega)\,\boldsymbol{\gamma}^{(0)}_{D=2}\,, \tag{51}$$

where $\boldsymbol{\gamma}^{(0)}_2 = \boldsymbol{\tau}_z$ for $D = 2$. We see that the topological Hamiltonian

$$\boldsymbol{t}_{\text{top}}(\boldsymbol{k}) = \boldsymbol{t}(\boldsymbol{k}) + \boldsymbol{\Sigma}(\boldsymbol{k}, \omega{=}0) = \boldsymbol{t}(\boldsymbol{k}) + \Sigma_+(0)\mathbf{1} + \Sigma_-(0)\boldsymbol{\tau}_z \tag{52}$$

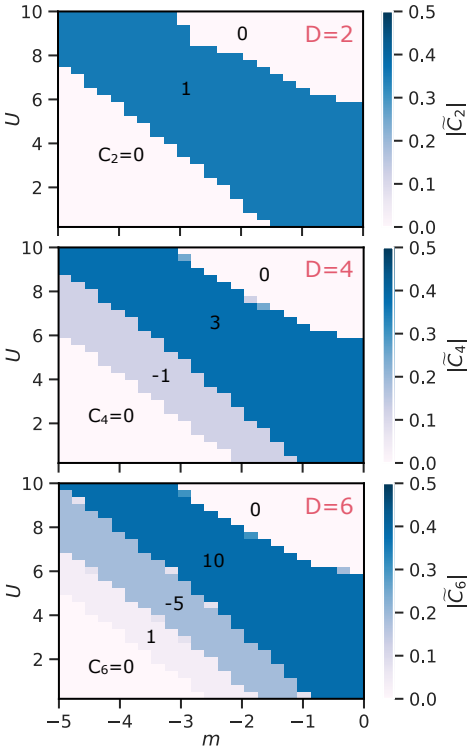


Fig. 7: *Topological m-U phase diagram of the interacting QWZ+U model for dimensions $D = 2, 4, 6$ as obtained by DMFT. The $D/2$-th Chern number C_D for the different phases is directly indicated in the plots. For $D = 2$, the first Chern number $C = C_2$. The color scale codes the modulus of the normalized Chern number defined as $\widetilde{C}_D = C_D/2^D$ (see text). Energy units: $t = t^*/\sqrt{D}$ with $t^*=1$. Figure taken from Ref. [5].*

acquires a nontrivial additional (staggered) orbital structure due to the second term, which adds to the term in Eq. (46) proportional to $\boldsymbol{\gamma}^{(0)}_{D=2}$. We thus expect a nontrivial modification of the $U = 0$ topological phase diagram.

The actual calculation by means of DMFT is conceptually simple: The lattice model is self-consistently mapped onto two Anderson impurity models, one for the A orbitals and another one the B orbitals, which are related to each other by A-B sublattice symmetry. As our intention is to cover the entire m-U parameter space, a simplified DMFT scheme, the two-site DMFT [22], is helpful. This has successfully been cross-checked against accurate DMFT results obtained with the standard Lanczos solver [23].

The top panel of Fig. 7 displays the topological phase diagram in the m-U control-parameter space for dimension $D = 2$. The phase diagram is shown for negative m only. The $m > 0$ range is obtained by reflection at the $m = 0$ axis with a simultaneous sign change $C \mapsto -C$. At $U = 0$ and when increasing m, we thus pass from $C = 0$ over $C = +1$ (for $m > -\sqrt{2}$) and $C = -1$ (for $m > 0$) to $C = 0$ (for $m > \sqrt{2}$), i.e., from a band insulator over two topologically different Chern insulators back to a band insulator again.

Right at $m = m_c = 0$, i.e., at the topological phase transition, the bulk system is a so-called semi-metal. The semi-metal state at this critical mass parameter does not have a (one-dimensional) Fermi surface. The gap is rather closed at isolated $\boldsymbol{k}$ points in the 1BZ, namely at $\boldsymbol{k}_c = (\pi, 0)$ and $\boldsymbol{k}_c = (0, \pi)$.

For $U > 0$, the Chern number is computed from the topological Hamiltonian, Eq. (52), and is thus affected by the self-energy at $\omega = 0$. As $\boldsymbol{\Sigma}(\omega)$ is diagonal in orbital space, this merely amounts to a renormalization of the chemical potential, $\mu \to \mu + \Sigma_+(\omega{=}0)$, and of the mass parameter, $m \to m + \Sigma_-(\omega{=}0)$. Since m acts as the strength of a staggered orbital field, $m = 0$ means that the orbital polarization $n_A - n_B$ vanishes and that $\Sigma_A(\omega) = \Sigma_B(\omega)$. Hence, $\Sigma_-(\omega) = 0$ and $\Sigma_+(\omega) = \Sigma_A(\omega) = \Sigma_B(\omega)$. Furthermore, the $\omega = 0$ value of $\Sigma_+(\omega) = U/2$ is exactly compensated by the chemical potential $\mu = U/2$. Hence, for finite but not too large U and at $m = 0$, the topological Hamiltonian equals the noninteracting one, and thus the system remains in a (correlated) semi-metal state.

For $U > U_c \approx 6$, the self-consistently determined DMFT self-energy develops a pole at $\omega = 0$. This implies that the concept of the topological Hamiltonian breaks down. In fact, at U_c there is a transition from the correlated semi-metal to a topologically trivial Mott insulator with $C = 0$, which is somewhat reminiscent of the paradigmatic metal-to-Mott-insulator transition predicted by DMFT [18].

For $m \neq 0$, the system becomes orbitally polarized. At constant $U < U_c$ and for decreasing m, the Chern-insulator phase with $C = 1$ extends up to a U-dependent critical mass parameter $m_c(U)$, where a topological phase transition to the band insulator takes place. Upon decreasing m for $U > U_c$ the Mott insulator is stable up to a second U-dependent critical mass parameter $m_{\text{Mott}}(U)$, where a transition to the correlated $C{=}1$ Chern-insulator takes place. The phase diagram shown in Fig. 7 suggests that the Mott insulator and the band insulator (both with $C{=}0$) cannot be connected continuously, as they are separated in parameter space by topologically nontrivial phases with $C \neq 0$.

Arbitrary even dimensions D

Since DMFT is a (rather questionable) approximation in case of a two-dimensional lattice model, it is worthwhile to ask if and how these results change for the same model on a higher-dimensional lattice. Eventually, in the $D \to \infty$ limit, the DMFT becomes exact, so that one might expect exact results for a strongly correlated and topologically nontrivial model at the same time. This extension to lattices of arbitrary dimensions D requires some effort. We concentrate on even dimensions, for odd D a somewhat different but similar approach is necessary. Let us mention that the $D \to \infty$ limit is not unique and that approaching it via even or via odd D yields different results.

The QWZ model on the hypercubic lattice with arbitrary even D, defined via Eq. (46), requires an explicit representation for the Hermitian and traceless generators $\gamma_D^{(\mu)}$ with $\mu = 0, 1, ..., D$ of the Clifford algebra Eq. (47). There is in fact a general recursive prescription [12], namely

$$\begin{aligned} \boldsymbol{\gamma}_{D+2}^{(0)} &= \boldsymbol{\tau}_z \otimes \mathbf{1} \quad , & \boldsymbol{\gamma}_{D+2}^{(r)} &= \boldsymbol{\tau}_x \otimes \boldsymbol{\gamma}_D^{(r)} \quad \text{for } r = 1, ..., D \\ \boldsymbol{\gamma}_{D+2}^{(D+1)} &= \boldsymbol{\tau}_x \otimes \boldsymbol{\gamma}_D^{(0)}, & \boldsymbol{\gamma}_{D+2}^{(D+2)} &= \boldsymbol{\tau}_y \otimes \mathbf{1} \, , \end{aligned} \tag{53}$$

where $\mathbf{1}$ is the $2^{D/2}$-dimensional unity. Explicitly, $\boldsymbol{\gamma}^{(0)} = \text{diag}(+1, +1, ..., -1, -1, ...)$. The number of orbitals $M = 2^{D/2}$ scales exponentially with D. Since a lower-dimensional faithful

representation is not possible [12, 24], the model becomes rather artificial. This is the price we have to pay to get a correlated and topologically nontrivial model solvable by DMFT.

Noninteracting electronic structure

At $U = 0$ and arbitrary even D, the band structure is easily obtained by squaring $\boldsymbol{t}(\boldsymbol{k})$, as shown above, and we get the dispersions of the two $M/2$-fold degenerate bands $\epsilon_\pm(\boldsymbol{k}) = \pm\big(t^2 \sum_r \sin^2 k_r + (m + t\sum_r \cos k_r)^2\big)^{1/2}$. We set $t = t^*/\sqrt{D}$ and $t^* = 1$. Due to the point-group symmetries, band closures are found at the high-symmetry points (HSPs) $\boldsymbol{k}_\mathrm{c} = \boldsymbol{k}_{n_0} \equiv (0, ..., 0, \pi, ..., \pi)$ in the 1BZ, and for the $\binom{D}{n_0}$ inequivalent permutations of the components, where n_0 counts the number of vanishing entries k_r. To get a band closure, the condition $m = (D-2n_0)\, t$ must be met or, equivalently,

$$m/t^* = (D-2n_0)/\sqrt{D}\,. \tag{54}$$

Thanks to the Clifford algebra, we can easily compute the moments $M_\alpha^{(n)} = \int d\omega\, \rho_\alpha^n(\omega)$ of the local partial density of states (DOS) of the orbital α for small n. We find the normalization condition $M_\alpha^{(0)} = 1$, the center of gravity $M_\alpha^{(1)} = m\gamma_{\alpha\alpha}^{(0)} = \pm m$, and the α-independent variance $M_\alpha^{(2)} - \big(M_\alpha^{(1)}\big)^2 = t^2 D = 1$. Hence, the scaling of the hopping leads to a proper balance between H_0 and H_1 for $D\to\infty$. The mass parameter m must not be scaled to maintain a nontrivial model.

The dispersions for the $D = 4$ model are shown in Fig. 8 (left). For $D = 4$ we have $M = 2^{D/2} = 4$, and thus the degeneracy of each band is $M/2 = 2$. Comparing results for different m (see color code), we find gap closures for $m = -2, -1, 0, 1, 2$ at the high-symmetry points (HSPs) $\boldsymbol{k}_\mathrm{c}$ with $n_0 = 4, 3, 2, 1, 0$, respectively. This is consistent with Eq. (54), which yields $m = (4-2n_0)/2 = 2-n_0$ for $D = 4$. The respective number of equivalent HSPs is 1, 4, 6, 4, 1, and the total number of wave vectors getting critical when varying m from $m = -\infty$ to $m = +\infty$ is $2^D = 16$.

The right panel in Fig. 8 displays the A-orbital DOS in the $D \to \infty$ limit for mass parameter $m = -1.5$. It can be calculated analytically (see Ref. [5])

$$\rho_\alpha(\omega) = \frac{1}{2}\frac{1}{t^*\sqrt{\pi}}\Theta\big(|\omega| - t^*/\sqrt{2}\big)\,\mathrm{sign}\,\omega \sum_{s=\pm}\left(\frac{\omega}{\sqrt{\omega^2 - t^{*2}/2}} + s z_\alpha\right) e^{-\frac{\left(s\sqrt{\omega^2 - t^{*2}/2} - m\right)^2}{t^{*2}}}. \tag{55}$$

Recall the definition $z_\alpha = +1, -1$ for A, B orbitals. Interestingly, the DOS has an m-independent gap given by $\Delta = \sqrt{2}t^*$, opposed to any finite-D DOS. This point requires discussion, see below.

Topological phase diagrams at $U = 0$

The middle panel of Fig. 8 displays the topological phase diagrams of the noninteracting QWZ model for $D = 2, 4, ...$, when using m as a control parameter. For each even dimension D, the red points mark the critical mass parameters given by Eq. (54). In fact they separate topologically inequivalent states of the system characterized by different Chern numbers. Actually, for arbitrary even D, the relevant invariant is the $(D/2)$-th Chern number C_D, which is obtained

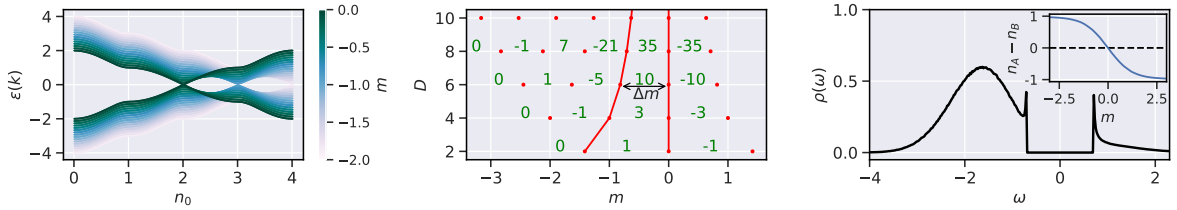


Fig. 8: *Dispersions of the two degenerate bands $\epsilon_\pm(\boldsymbol{k})$ of the $D = 4$ model* (left) *along straight shortest lines in the 1BZ connecting HSPs characterized by n_0. Results for different m, see color code.* Middle: *Different topological phases with Chern numbers $C_D(n_0)$ (green), separated by critical m-values (red dots) for different D.* Right: *$U = 0$ DOS on the A-orbitals at $m = -1.5$ for $D = \infty$. Inset: orbital polarization as function of m for $D = \infty$. Nearest-neighbor hopping: $t = t^*/\sqrt{D}$, $t^* = 1$ sets the energy scale. Figure taken from Ref. [5].*

as a D-dimensional integration of the D-form, which has been discussed above, over the 1BZ. (Note that here we depart from the conventional notation $C_{D/2}$ for the $(D/2)$-th Chern number.) The $(D/2)$-th Chern number of a topological phase for an m that satisfies, for some $n_0 \in \{0, ..., D-1\}$, the condition

$$\frac{D-2n_0-2}{\sqrt{D}} < m/t^* < \frac{D-2n_0}{\sqrt{D}} \,, \tag{56}$$

i.e., that lies between the critical mass parameters with band closures at $\boldsymbol{k}_{2n_0-2}$ and $\boldsymbol{k}_{n_0}$, respectively, is given by [12, 25, 26]

$$C_D(n_0) = (-1)^{n_0+\frac{D}{2}} \binom{D-1}{n_0} \,. \tag{57}$$

For the extreme cases

$$m/t^* < -\frac{D}{\sqrt{D}} \quad \text{and for} \quad \frac{D}{\sqrt{D}} < m/t^* \tag{58}$$

we have band insulators with $C_{D/2}(n_0 = D-1) = 0$ and $C_{D/2}(n_0{=}0) = 0$, respectively.

Fig. 8 (middle) shows that the m-distance between two neighboring transitions shrinks to zero for $D \to \infty$. Eq. (56) gives us $\Delta m = 2t^*/\sqrt{D}$. Hence, for high D the system is arbitrarily close to criticality for *any* m, and in the limit $D \to \infty$, the set of critical m's becomes dense in any finite m-interval. The phase diagram exhibits a *continuum* of topologically different phases. This observation has a couple of implications.

First, it means that the definition of a critical point in the 1BZ becomes elusive for $D \to \infty$, since $\epsilon_\pm(\boldsymbol{k}) = \epsilon_\pm(\boldsymbol{k}')$ if $\|\boldsymbol{k}-\boldsymbol{k}'\| = 0$, where we have defined $\|\boldsymbol{k}\|^2 \equiv \lim_{D\to\infty} D^{-1} \sum_{r=1}^{D} k_r^2$. $\|\cdot\|$ is a semi-norm, i.e., $\|\boldsymbol{k}\| = 0$ does not imply $\boldsymbol{k} = 0$. Hence, the concept of band closures at isolated points in the 1BZ breaks down. For a given m, we have $\epsilon_\pm(\boldsymbol{k}) = 0$ at a critical wave vector $\boldsymbol{k} = \boldsymbol{k}_c(m)$ but also at all wave vectors $\boldsymbol{k}$ with $\|\boldsymbol{k}-\boldsymbol{k}_c(m)\| = 0$. Furthermore, while the number $\binom{D}{n_0}$ of equivalent critical HSPs at a given critical m and the total number 2^D of HSPs in the 1BZ diverge, their ratio approaches a constant when $D \to \infty$.

Second, since all values of the mass parameter are critical, one would expect the absence of a gap in the DOS for any m. To understand the fact that the DOS displayed in Fig. 8 (right) has

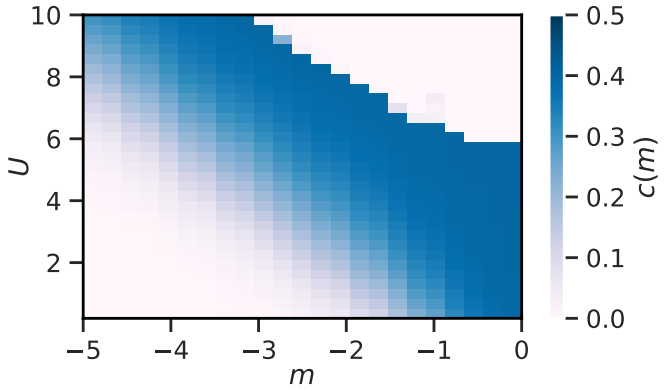


Fig. 9: *The topological phase diagram of the QWZ+U model in the $D\to\infty$ limit. The color codes the topologically invariant Chern density $c(m,U)$ in the m-U plane. Only the $m\leq 0$ range is plotted as the phase diagram is symmetric under $m\to -m$. Figure taken from Ref. [5].*

an m-independent gap $\Delta = 2t^*$, one can analytically analyze the DOS at a critical m at any finite D. For sufficiently small frequencies (see Ref. [5]),

$$\rho_\alpha(\omega) = c(D,n_0)|\omega|^{D-1}/t^{*D} \tag{59}$$

with a coefficient $c(D,n_0)$ which, for any n_0 and thus for any critical m, tends to zero exponentially fast with $D\to\infty$. This demonstrates that the contribution of wave vectors with $\epsilon_\pm(\boldsymbol{k})=0$ to the DOS vanishes in the $D\to\infty$ limit. For $D\to\infty$, we conclude that the DOS does not indicate whether the system is a (Chern) insulator or a semi-metal. The distinction between insulator and semi-metal becomes meaningless in this limit.

A third important observation is related to the Chern number. Eq. (57) shows that for $D\to\infty$ only the modulus of the Chern number, and only after proper normalization, has a well-defined limit. Since $\sum_{n_0=0}^{D-1}|C_D(n_0)| = 2^{D-1}$, we thus introduce a *Chern density* as $c(n_0) = \lim_{D\to\infty}|C_D(n_0)|/2^{D-1}$. To express the Chern density as a function of m, we employ $n_0 = \big(D-m\sqrt{D}/t^*\big)/2$, see Eq. (54). Furthermore, since $\Delta m\mapsto 0$ we write $dm\equiv 2t^*/\sqrt{D}$. With the Moivre-Laplace theorem, which states that the binomial distribution converges to the normal distribution, we then find

$$c(n_0) = \lim_{D\to\infty}\sqrt{\frac{2}{\pi D}}e^{-2\frac{(D/2-n_0)^2}{D}} = c(m)\,dm \tag{60}$$

with

$$c(m) = \frac{1}{t^*\sqrt{2\pi}}e^{-\frac{1}{2}\frac{m^2}{t^{*2}}}\,. \tag{61}$$

The Chern density is a smooth Gaussian density. It is normalized by construction and turns out to have unit variance. This is important, as it shows that not only dynamic correlation effects but also nontrivial topological properties survive the $D\to\infty$ limit when using the standard scaling of the hopping.

Topological phase diagram of the interacting system for $D\to\infty$

Fig. 9 displays the topological phase diagram of the QWZ+U model in the limit $D\to\infty$. The control parameters are m and U, and the topological invariant, the Chern density, is indicated by

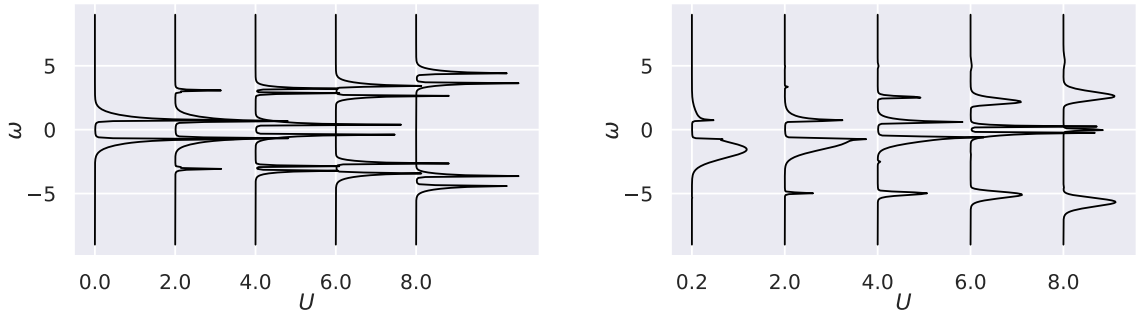


Fig. 10: *Local spectral function on $\alpha = A$-orbitals for $m = 0$ (left) and $m = -1.5$ (right) for various U in the $D \to \infty$ limit as obtained by two-site DMFT. Figure taken from Ref. [5].*

the color code. The results have been obtained numerically employing the simplified two-site DMFT scheme [22, 5] and have been cross-checked against full DMFT employing the Lanczos solver [23]. The Chern density is obtained from the topological Hamiltonian, where we merely have to take into account a renormalization of the chemical potential, $\mu \mapsto \mu + \Sigma_+(\omega{=}0)$, and of the mass parameter, $m \mapsto m + \Sigma_-(\omega{=}0)$.

We first discuss the $U = 0$ case. As a function of m, the Chern density is a Gaussian, see Eq. (61). This means that the system smoothly evolves from a conventional band insulator, with $c(m, 0) \to 0$ in the limit $m \to -\infty$, to a Chern insulator (or semimetal, this cannot be distinguished). The invariant increases and takes the maximum value $c(m, 0) = 1/\sqrt{2\pi}$ at $m = 0$. The phase diagram for $m > 0$ is simply obtained by reflection at the $m = 0$ axis (also for $U > 0$). Already the $U = 0$ phase diagram is quite unconventional, since we have an infinitely fine classification, where *each* value of the (modulus of the) control parameter m defines a separate topological phase. At the same time, each value for m is also critical.

With increasing U at $m = 0$, the Chern density $c(0, U)$ stays at its maximum, until at $U = U_\mathrm{c} \approx 6t^*$ the system undergoes a correlation-driven transition to a Mott-insulating phase. The Mott phase is topologically trivial with $c = 0$. Approaching U_c either from above or from below, the transition is characterized by a continuously vanishing renormalization factor $z \mapsto 0$, where $z \equiv 1/\big(1 - \partial\Sigma_\alpha(\omega{=}0)/\partial\omega\big)$ is independent of the orbital type α. The quantity z plays the role of a band-gap renormalization [27]. Examples for the U-dependent interacting spectral function on the A orbitals are given in Fig. 10 for $m = 0$ and for $m < 0$.

For $U > U_\mathrm{c}$, the Mott phase extends to $m < 0$ and is bounded for all m by a line of critical interactions $U_\mathrm{c}(m)$. For $m \to -\infty$ we observe that $U_\mathrm{c}(m)$ develops into a linearly increasing function of $|m|$. This is due the fact that the system becomes fully orbital-polarized, see the inset in Fig. 8 (right). Hence, the self-energy becomes static and approaches constants $\Sigma_A \to U$, $\Sigma_B \to 0$, such that the renormalization of m is trivial: $m \to m + \Sigma_-(\omega{=}0) \to m + U/2$. As a consequence, the band insulator with $c = 0$ cannot be smoothly connected to the Mott insulator with $c = 0$ without passing topologically nontrivial states with $c > 0$.

The whole phase diagram Fig. 9 can be understood as being the $D \to \infty$ limit of m-U phase diagrams at finite D, see Fig. 7, where we have used the same color coding as in Fig. 9 for the

Chern density, and where the (signed and unnormalized) Chern numbers are given additionally and directly label the topological phases in the figure. Note the alternating sign and the monotonic increase of the modulus of the Chern number along any straight path from the band to the Mott insulator. With increasing D, the number of topologically nontrivial phases $C_D(m) \neq 0$ increases and they cover ever narrower regions in the m-U plane (recall $\Delta m = 2t^*/\sqrt{D}$ for $U = 0$), until they shrink to one-dimensional curves of constant color given by $c(m, U) = \text{const}$ in Fig. 9. In the limit $D \to \infty$, systems on these iso-Chern curves are topologically equivalent, while on paths crossing iso-Cherns one passes through a continuum of topologically different phases.

6 Concluding discussion

We have demonstrated that DMFT can be used to study the interplay between topology and correlations in a model for a Chern insulator on an infinite-dimensional lattice. While the approach is numerically exact, the underlying model necessarily involves, besides the $D \to \infty$ limit, an infinite number of orbitals and is thus somewhat artificial. There are three main lines along which the discussion could proceed:

The considered example belongs (at $U = 0$) to the so-called class A of the complete topological classification of noninteracting lattice-fermion systems. The other nine classes of the "ten-fold way" have not yet been considered (a simple modification of the theory covers class A III). One important question is, whether interacting lattice-fermion models can be constructed that remain well-defined and nontrivial in the $D \to \infty$ limit for *all* classes. In the positive case, this would demonstrate that the ten classes "survive" if one switches on a Hubbard-type interaction, and we would have explicit models as prototypes, at least in the (comparatively simple) $D \to \infty$ limit. Even in this limit, where DMFT applies, however, we would not be sure that the classification was complete – another problem to be tackled.

Complexity rises dramatically, if one attempts a topological classification of all interacting systems on finite-dimensional lattice. Clearly, already the correlation problem is hard. On the other hand, we are allowed to consider continuous deformations, preferably to models, which are more simple from the correlation point of view. This might help to proceed. Up to now a general classification has not been achieved and theory is still at the very beginning, see Ref. [28] for a discussion. Some important questions are: Is the topological phase diagram of a noninteracting system stable against weak interactions? Vice versa, can topologically nontrivial phases be generated by interactions? In the $D \to \infty$ limit, the answer is "yes", see Fig. 9. More important, are there topologically nontrivial phases in interacting models that have no analog to those of noninteracting ones, i.e., cannot be connected to a phase classified by the ten-fold way? Is a topologically nontrivial Mott insulator possible? What are the consequences of the bulk-boundary correspondence in case of interacting systems?

On a more pragmatic level and with the focus to concrete materials, one may attempt to compute topological invariants of two- and three-dimensional interacting systems by approximate numerical tools. DMFT itself and its various extensions to include nonlocal correlations to

some extent, suggest themselves and appear attractive since the Chern number, for example, can be computed by means of a formula based on the single-particle Green function only [29, 30]. However, this expression rather is inconvenient for numerical studies as it involves an additional frequency integration. Invoking the tool of continuous deformations, it has been shown [31] that a strong simplification is possible and that it is sufficient to consider the self-energy at $\omega = 0$ only, i.e., that the invariant can be computed from the topological Hamiltonian. A different concept is that of the so-called pole expansion [32, 33]. Here, the self-energy is represented by a hybridization function involving auxiliary orbitals. For a given self-energy, this allows us to map the interacting to a noninteracting problem such that the standard methods of topological classification apply. However, it has been questioned [34] that these two concepts are equivalent in case of a $\boldsymbol{k}$-dependent self-energy. Another more fundamental problem is that any scheme, which is solely based on the single-particle Green function, disregards two-particle correlations to some extent. The two-particle excitation spectra may undergo a gap closure while the one-particle gap stays open, so that a complete topological classification cannot be achieved in this way. Ref. [35] provides an example along this line of reasoning.

Acknowledgements

The author would like to thank Lara Heyl, Christian Jöns, David Krüger and Robin Quade (University of Hamburg) for numerous discussions and acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) through the Cluster of Excellence “Advanced Imaging of Matter”-EXC 2056, Project ID No. 390715994, and through the Research Unit FOR 5249-449872909 (Project P8) is gratefully acknowledged.

References

[1] B.A. Bernevig: *Topological insulators and topological superconductors* (Princeton University Press, 2013)

[2] J. Asbóth, L. Oroszlány, and A. Pályi: *A Short Course on Topological Insulators* Lecture Notes in Physics, vol. 919 (Springer, Cham, 2016)

[3] M. Nakahara: *Geometry, topology, and physics* (IOP Publishing, Bristol, 1998)

[4] S.M. Bhattacharjee, M. Mj, and A. Bandyopadhyay (eds.): *Topology and Condensed Matter Physics* (Springer, Singapore, 2017)

[5] D. Krüger and M. Potthoff, Phys. Rev. Lett. **126**, 196401 (2021)

[6] L. Heyl: *Oberflächenzustände von Chern-Isolatoren in hohen Raumdimensionen* (B.Sc. thesis, University of Hamburg, 2021)

[7] T. Kato, J. Phys. Soc. Jpn. **5**, 435 (1950)

[8] A. Messiah: *Quantenmechanik*, Vol. II (de Gruyter, Berlin, 1979)

[9] M.V. Berry, Proc. R. Soc. London A **392**, 45 (1984)

[10] P.A.M. Dirac, Proc. R. Soc. London A **133**, 60 (1931)

[11] J.E. Avron, L. Sadun, J. Segert, and B. Simon, Commun. Math. Phys. **124** (1989)

[12] E. Prodan and H. Schulz-Baldes: *Bulk and Boundary Invariants for Complex Topological Insulators: From K-Theory to Physics* (Springer, 2016)

[13] C.-K. Chiu, J.C.Y. Teo, A.P. Schnyder, and S. Ryu, Rev. Mod. Phys. **88**, 035005 (2016)

[14] X.-L. Qi, Y.-S. Wu, and S.-C. Zhang, Phys. Rev. B **74**, 085308 (2006)

[15] K. von Klitzing, Rev. Mod. Phys. **58**, 519 (1986)

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

[17] F.D.M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988)

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

[19] Z. Wang and B. Yan, J. Phys.: Condens. Matter **25**, 155601 (2013)

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

[21] E. Müller-Hartmann, Int. J. Mod. Phys. B **3**, 2169 (1989)

[22] M. Potthoff, Phys. Rev. B **64**, 165114 (2001)

[23] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)

[24] S. Lang: *Algebra* (Springer, New York, 2002)

[25] M.F.L. Golterman, K. Jansen, and D.B. Kaplan, Phys. Lett. B **301**, 219 (1993)

[26] X.-L. Qi, T.L. Hughes, and S.-C.Zhang, Phys. Rev. B **78**, 195424 (2008)

[27] M. Sentef, J. Kuneš, P. Werner, and A.P. Kampf, Phys. Rev. B **80**, 155116 (2009)

[28] S. Rachel, Rep. Prog. Phys. **81**, 116501 (2018)

[29] Z. Wang, X.-L. Qi, and S.-C. Zhang, Phys. Rev. Lett. **105**, 256803 (2010)

[30] K. Ishikawa and T. Matsuyama, Z. Phys. C **33**, 41 (1986)

[31] Z. Wang and S.-C. Zhang, Phys. Rev. X **2**, 031008 (2012)

[32] S.Y. Savrasov, K. Haule, and G. Kotliar, Phys. Rev. Lett. **96**, 036404 (2006)

[33] L. Wang, H. Jiang, X. Dai, and X.C. Xie, Phys. Rev. B **85**, 235135 (2012)

[34] P. Thunström and K. Held, arXiv:1907.03899

[35] Y.-Y. He, H.-Q. Wu, Z.Y. Meng, and Z.-Y. Lu, Phys. Rev. B **93**, 195164 (2016)

11 Beyond DMFT: Spin Fluctuations, Pseudogaps and Superconductivity

Karsten Held
Institute for Solid State Physics, TU Wien
1040 Wien, Austria

Contents

1 Introduction **2**

2 Dynamical vertex approximation **6**

3 Ladder dynamical vertex approximation **10**

4 Hubbard model, cuprates and nickelates **12**

5 Spin fluctuations **16**

6 Pseudogap **19**

7 Superconductivity **24**

8 Conclusion and outlook **27**

1 Introduction

With the success of dynamical mean-field theory (DMFT) [1–3] in calculating strongly correlated electron systems, there have been attempts from the very beginning [4] to systematically extend DMFT. The aim is here to keep the good description of DMFT for local electronic correlations, but also to capture non-local correlations beyond.

Indeed, the local DMFT correlations are doing an excellent job in describing a quasiparticle renormalization with a weight Z that is uniform in momentum space – a surprisingly good approximation for many transition metal-oxides and heavy fermion systems; as well as the Mott-Hubbard metal-insulator transition that emerges for $Z \to 0$ [2]. Furthermore, all kinds of orders (magnetic, orbital, charge density wave ...) are quite well captured in three-dimensional systems up to the vicinity of the phase transition. Here, close to the phase transition, the mean-field nature of DMFT surfaces, among others, in form of mean-field critical exponents. Non-local correlations are here essential for describing the proper critical behavior. With a diverging correlation length, long-range correlations feed back to the self-energy which thus becomes non-local.

The DMFT approach has been covered already in various other chapters of this Autumn School [3], and the present chapter will thus focus instead on non-local electronic correlations beyond DMFT. When are such non-local correlations important?

A lot of our insight into physical phenomena stems from weak coupling perturbation theory. Even though such an approach is certainly not applicable to strongly correlated electron systems, it often nonetheless provides for some qualitative understanding. One example, where non-local correlations enter the self-energy are spin-fluctuations. These can be calculated at weak coupling in the random-phase approximation (RPA) which is discussed in quantum field theory textbooks. The RPA is the geometric series of all ladder diagrams with no, one, two etc. Coulomb interaction lines, as illustrated in Fig. 1 (top). The figure just shows one term of the sum, where "..." indicates that all orders in U are included. The RPA can be used to calculate the magnetic susceptibility, Fig. 1 (top, without dashed line). As we will see later, the poles of this susceptibility constitute bosonic quasiparticle excitations coined magnons or, in the paramagnetic phase, paramagnons.

Now, when we add the dashed black line in Fig. 1 (top), we obtain a self-energy Feynman diagram. It describes the coupling of the electron to the spin-fluctuations. We can also call it the electron-(para)magnon interaction. As spin-fluctuations are very non-local we get – whenever spin-fluctuations become important – contributions from diagrams with lattice sites $i \neq j \neq k$ in Fig. 1 (top). Such self-energy diagrams are certainly not contained in DMFT, which sums up all *local* contributions of Feynman diagrams for the self-energy. Hence, whenever spin fluctuations become large, we must expect important corrections to the DMFT self-energy. Indeed this is not restricted to spin fluctuations, but depending on what kind of spin combination the particle and the hole in the RPA ladder have, one can also obtain the coupling of electrons to charge, orbital etc. fluctuations.

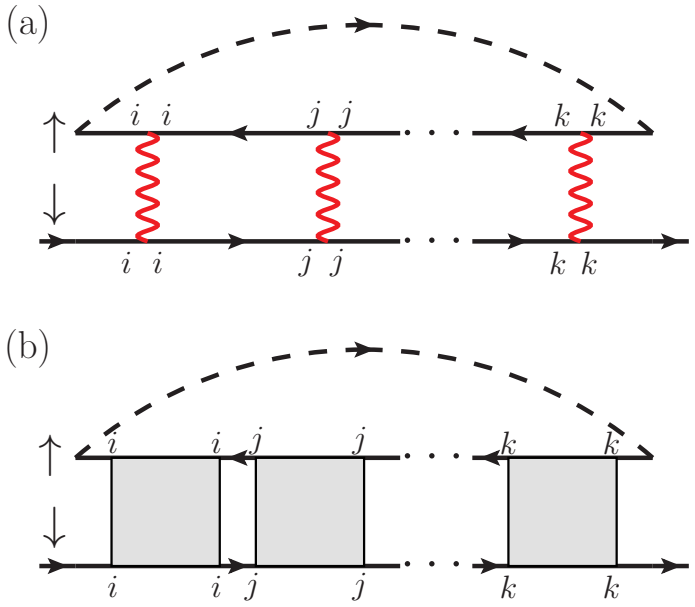


Fig. 1: *Top: Feedback of spin fluctuations (paramagnons) calculated in RPA to the self-energy. Bottom: In diagrammatic extensions of DMFT such physics is taken into account but with the bare interaction U (red wiggled line) replaced by a local vertex calculated from an impurity model (light gray box). Depending on the flavor of the diagrammatic extension of DMFT different local vertices are employed. Reproduced from [5].*

That means non-local correlations are certainly relevant whenever spin, charge etc. fluctuations are important. An obvious regime where this is the case is the vicinity of a second-order phase transition as already mentioned. Here the magnetic, charge etc. susceptibility diverges and significant changes to the DMFT solution are thus to be expected. For low dimensional systems we will get corrections also further away from the phase transition. The Mermin-Wagner theorem prohibits long-range order with a continuous symmetry breaking in two-dimensions at finite temperature. Hence, antiferromagnetic order is restricted to zero temperature. However, above this zero-temperature antiferromagnetic phase, we have now strong antiferromagnetic fluctuations in a wide temperature range, even with exponentially large correlation lengths. DMFT has been developed with the limit of dimension $d = \infty$ in mind, see [1] and Chapter "Why calculate in infinite dimensions?" by D. Vollhardt [3]. Hence, also from this perspective it is not surprising that we need to expect larger corrections to DMFT for low dimensional systems.

On the other hand, we would like to keep the success of DMFT in describing a major part of electronic correlations rather well: the local correlations. Two major routes to do so have been developed to this end, see Fig. 2 for an illustration. Cluster extensions of DMFT [6] put the DMFT concept of locality onto a cluster (instead of a single site) that is embedded in a DMFT bath. For a single site cluster this is just DMFT. Illustrated in Fig. 2 is a two-site cluster where thus non-local correlations between the two red sites of the cluster are captured. Such two site clusters can, e.g., describe the formation of a spin singlet between two electrons on the two sites. For a four site cluster also d-wave superconductivity can be described. However such

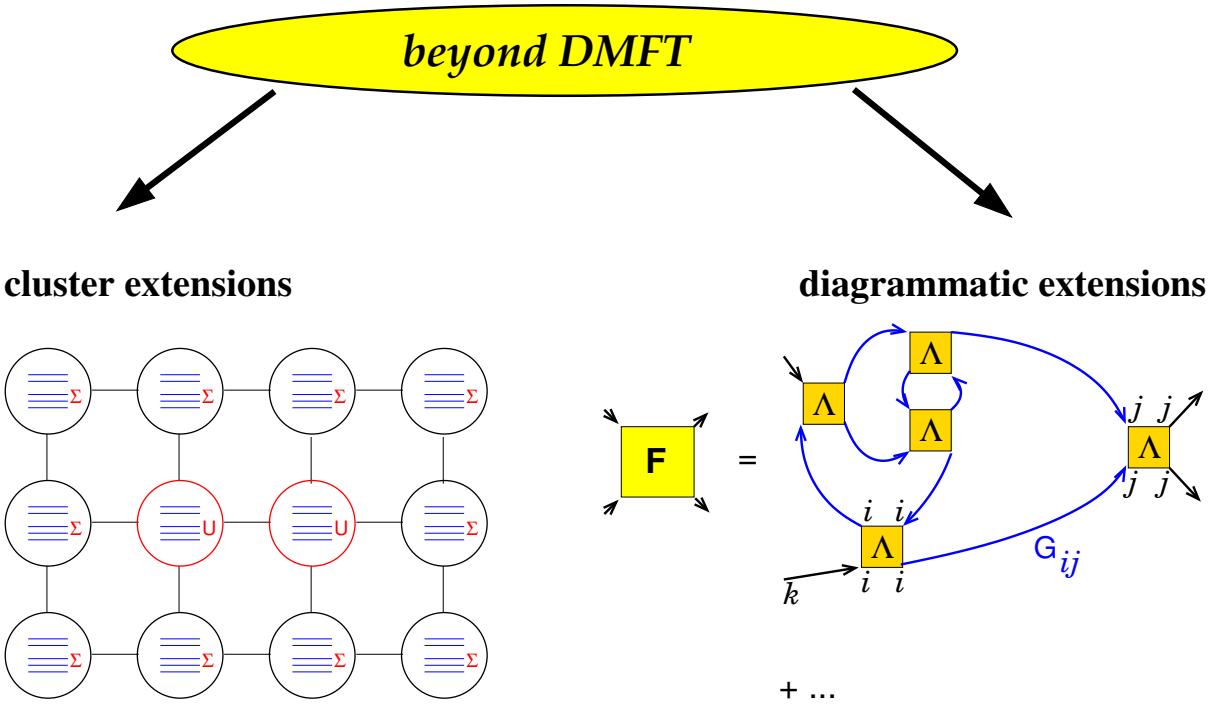


Fig. 2: *Left: In cluster extensions, a couple of sites (here two marked in red) are embedded in a dynamical mean field which can be formulated through the self energy Σ. Restricted to a single site, DMFT is obtained; extended to the full lattice the exact solution is recovered. Right: In diagrammatic extensions, the locality of DMFT is put to the next level, the two-particle level. Local two-particle building blocks Λ are connected by non-local Green function lines G_{ij}, resulting in a non-local full vertex F or susceptibility χ. When restricting G_{ij} to its local contribution G_{ii}, DMFT is recovered. Adapted from [7].*

small clusters tend, in practice, to largely overestimate the physics that is compatible with the cluster such as the spin singlet formation and d-wave superconductivity for a two and four site cluster, respectively. While one can go to clusters of size 10×10, a proper finite size scaling remains challenging. This is even more true for realistic multi-orbital calculations that are restricted to a handful of sites.

The other route extends DMFT [5] Feynman diagrammatically. Here, the concept of locality is not extended to a cluster but instead to the n-particle vertex. For $n = 1$ we have the one-particle vertex which is nothing but the self-energy. And a local self-energy is just the DMFT approximation. For $n = 2$, i.e., the two-particle level, we start instead with a local two-particle vertex and construct from it the non-local full vertex F, see Fig. 2 (right), as well as the non-local self-energy and Green function. This is the level most commonly applied nowadays. Similarly as for the cluster extensions there are different flavors depending on which local vertex and which connecting Green functions are taken. At the end of the day, most of these different flavors are very similar, at least as long as they take the two-particle, four-point vertex as a starting point. Other approaches start from a three-point local vertex which is a more severe approximation. For an overview and comparison, see the review [5].

The originary method, called dynamical vertex approximation (DΓA) [8], considers the vertex in terms of real fermions as local. A second widely applied approach is the dual fermion (DF) approach [9]. Also within DΓA different flavors are used. In its most complete form, the fully irreducible vertex Λ is approximated to be local. Then the parquet equations are needed to construct F and the self-energy Σ. A local Λ is an excellent approximation, even for the two-dimensional Hubbard model in the superconducting doping regime [10]. In the ladder DΓA variant, the irreducible vertex Γ_ℓ in a certain channel ℓ such as the particle-hole (ph) channel is considered to be local. In this case, the Bethe-Salpeter equation is sufficient to calculate F along the same lines as in the RPA, only now with Γ_ℓ instead of U as a building block, i.e the light gray box in Fig. 1 (bottom) is Γ_ℓ in this case. The ladder variant is numerically much less expensive. It is sufficient if a certain channel dominates, but does not capture the coupling of different channels into each other.

One can also extend the concept of locality to higher n-particle vertices. The $n = 3$-particle vertex level, for example, has been employed for estimating the error of the $n = 2$-particle calculations [11]. For $n \to \infty$ the exact solution is recovered.

Turning back to Fig. 1, we see that such diagrammatic extensions are well suited to describe spin-fluctuations and their feedback to the fermionic self-energy. The same physics as is qualitatively described in RPA, is now captured for strongly correlated electrons since the local Λ already encodes non-perturbatively all DMFT correlations. The Bethe-Salpeter ladder of Fig. 1 is precisely the same as is also used to calculate the DMFT susceptibility, cf. the Chapter "DMFT for linear response functions" by E. Pavarini [3]. What is not covered in DMFT is how these spin-fluctuations impact the self-energy as well as self-consistency effects, i.e., how the changed Σ modifies G or that the local Γ_ℓ itself becomes different from DMFT. These self-consistency effects lead to a dampening of the spin-fluctuations compared to the mean-field DMFT solution, up to the point of fulfilling the Mermin-Wagner theorem in two-dimensions.

Diagrammatic extensions of DMFT have been highly successful: (i) The critical behavior in the vicinity of (quantum) phase transition could be described for the first time in electronic models, a topic which was covered already excessively in the Autumn School 2018 [12], see also the review [5]. (ii) It was realized that the two-dimensional square lattice Hubbard model with perfect nesting is insulating all the way down to zero interaction [13], correcting earlier cluster DMFT results. (iii) The pseudogap and d-wave superconductivity can be described in the two-dimensional Hubbard model, a topic which we will cover in the following chapters. (iv) A new polariton, the π-ton has been discovered in model calculations [14]. (v) Realistic materials calculations are possible and have been pursued, e.g., for $SrVO_3$ [15].

In the following, we will first introduce one of the diagrammatic extensions of DMFT, the DΓA, in Sec. 2. Simplifications when using the ladder variant of DΓA are outlined in Sec. 3. The Hubbard model is introduced in Sec. 4 and its justification for cuprates and nickelates is discussed. Physical results regarding spin fluctuations, the pseudogap and superconductivity are discussed in Sec. 5, Sec. 6, and Sec. 7, respectively. Finally, Sec. 8 provides a conclusion and outlook.

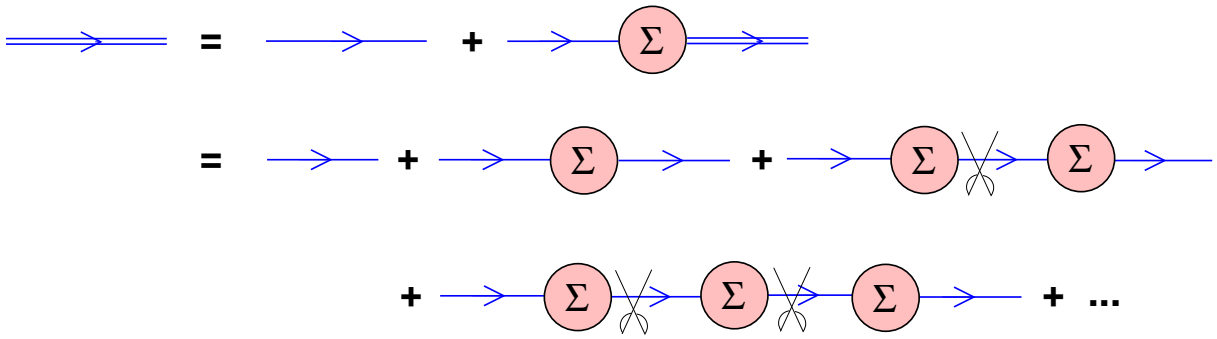


Fig. 3: *Dyson equation connecting the Green function and self energy Σ (single blue line: non-interacting G_0; double blue line: interacting G). The pair of scissors indicates that these diagrams are one-particle reducible (i.e., cutting one G_0 line separates the Feynman diagram into two parts). From [12].*

2 Dynamical vertex approximation

The aim of the present section is to provide the reader with the basic idea of the dynamical vertex approximation. It builds upon a similar chapter of a preceding Jülich Autumn School [12]. A more detailed description for a second reading can be found in the review [5]. Further information on how to calculate the superconducting critical temperature T_c and how to include the asymptotic form of the vertex, can be found in [16].

The basic idea of the dynamical vertex approximation (DΓA) is a resummation of Feynman diagrams, not order by order of the Coulomb interaction as in conventional perturbation theory, but in terms of their locality. That is, we assume the fully irreducible n-particle vertex to be local and from this building block we construct further diagrams and non-local correlations.

The first level ($n = 1$) is then just the DMFT which corresponds to all local Feynman diagrams for the self-energy Σ. Note that Σ is nothing but the fully irreducible $n = 1$-particle vertex. One particle-irreducibility here means that cutting one Green function line does not separate the Feynman diagram into two pieces. Indeed such reducible diagrams must not be included in the self-energy since it is exactly these diagrams that are generated from the Dyson equation (Fig. 3) which, resolved for G, reads

$$G_{\mathbf{k}\nu} = \left(1/G^0_{\mathbf{k}\nu} - \Sigma_{\mathbf{k}\nu}\right)^{-1} \tag{1}$$

for momentum $\mathbf{k}$, Matsubara frequency ν and non-interacting Green function $G^0_{\mathbf{k}\nu}$. Here and in the following, single blue lines denote non-interacting Green functions G_0 and double blue lines indicate interacting Green functions G. Fig. 3 further shows how one-particle reducible diagrams are generated through the Dyson equation. Hence these must not be contained in the Feynman diagrams that constitute Σ, to avoid a double counting. That means, Σ must be one-particle reducible in terms of G^0.[1]

On the next level, for $n = 2$, we assume the locality of the two-particle fully irreducible vertex Λ. For the two-particle vertex, fully irreducible means that cutting two Green function lines

[1] In terms of G the skeleton diagrams for Σ are also two-particle reducible. But that is another story that is connected with the way how Σ enters G.

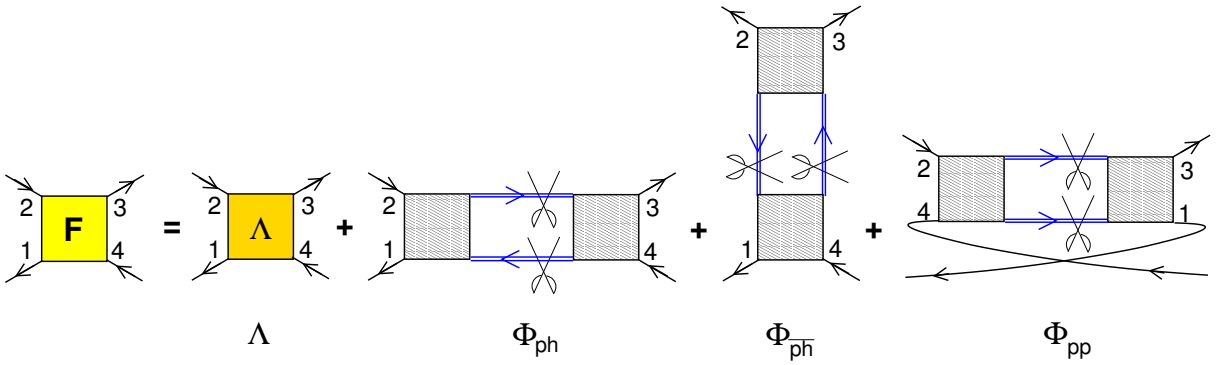


Fig. 4: *Parquet decomposition of the full (reducible) vertex F into the fully irreducible vertex Λ and two-particle reducible diagrams Φ_ℓ in the three channels. The two pairs of scissors indicate the reducibility of the three Φ_ℓ's. Each two-fermion Feynman diagram belongs to one and only one of the three Φ_ℓ's or to Λ.*

does not separate the diagram into two pieces. There are three different kinds (channels) ℓ of reducible vertices Φ_ℓ and a fourth, Λ, that is fully irreducible. Most importantly each diagram falls in one and only one of those four subgroups. Thus the full vertex F, containing all diagrams, can be written as the sum those. This decomposition of the vertex is called *parquet decomposition* and is graphically displayed in Fig. 4.

The reason why there are three distinct reducible parts Φ_ℓ is that say leg 1 may stay connected with leg 2, 3, or 4 when cutting two Green function lines as indicated in Fig. 4. These three possible channels ℓ are denoted as particle-hole (ph), transversal particle-hole ($\overline{ph}$) and particle-particle (pp). The irreducible vertex of each channel is just the complement: $\Gamma_\ell = F - \Phi_\ell$. It is important to note that each reducible diagram is contained in one and only one of these channels. One can show this by contradiction: otherwise cutting lines in two channels would result in a diagram with one incoming and two outgoing lines, which is not possible because of the conservation of (fermionic) particles.

There is a set of six exact equations, also called the "parquet equations" [17,5] to the confusion of the common student, that allows us to calculate from a given Λ the six quantities: full vertex F, self-energy Σ, Green function G and the three reducible vertices Φ_ℓ.

(1) The first equation is the actual parquet equation [Fig. 4, Eq. (2)]

$$F^{\nu\nu'\omega}_{r,\mathbf{k}\mathbf{k}'\mathbf{q}} = \Lambda^{\nu\nu'\omega}_{r,\mathbf{k}\mathbf{k}'\mathbf{q}} + \Phi^{\nu\nu'\omega}_{ph,r,\mathbf{k}\mathbf{k}'\mathbf{q}} + \Phi^{\nu\nu'\omega}_{\overline{ph},r,\mathbf{k}\mathbf{k}'\mathbf{q}} + \Phi^{\nu\nu'\omega}_{pp,r,\mathbf{k}\mathbf{k}'\mathbf{q}} \tag{2}$$

where $r \in \{c, s\}$ is the symmetric/antisymmetric spin combination, i.e., $F_{c/s} = F_{\uparrow\uparrow} \pm F_{\uparrow\downarrow}$ and similarly for other vertices.[2] The name indicates that F_c and F_s give rise to the charge and spin fluctuations, respectively.

(2-4) Three of the six "parquet" equations are the Bethe-Salpeter equation in the three channels ℓ. In [Fig. 5, Eq. (3)] we here only reproduce the $\ell = ph$ channel; again with $r \in \{c, s\}$ for

[2]This assumes SU(2) symmetry, if this is broken altogether four spin combinations need to be taken into account.

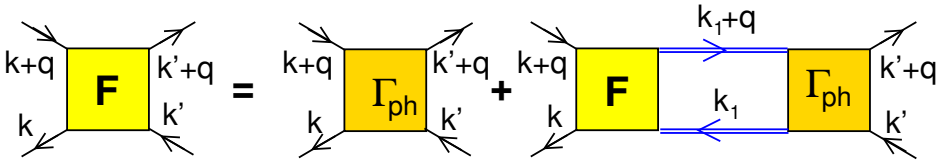


Fig. 5: *Bethe-Salpeter equation in the particle-hole channel, $\ell = ph$, allowing us to calculate F from the irreducible vertex in Γ_{ph} in this channel. The rightmost term produces the particle-hole reducible diagrams denoted as Φ_{ph} in Fig. 4; $\Gamma_{ph} = \Lambda + \Phi_{\overline{ph}} + \Phi_{pp}$.*

a symmetric/antisymmetric spin combination[3]

$$F^{\nu\nu'\omega}_{r,\mathbf{k}\mathbf{k}'\mathbf{q}} = \Gamma^{\nu\nu'\omega}_{r,ph,\mathbf{k}\mathbf{k}'\mathbf{q}} + \sum_{\mathbf{k_1}\nu_1} F^{\nu\nu_1\omega}_{r,\mathbf{k}\mathbf{k_1}\mathbf{q}}\, G_{\mathbf{k_1}\nu_1}\, G_{(\mathbf{k_1}+\mathbf{q})(\nu_1+\omega)}\, \Gamma^{\nu_1\nu'\omega}_{ph,r,\mathbf{k_1}\mathbf{k}'\mathbf{q}}\,; \tag{3}$$

One often combines frequency ν and momentum $\mathbf{k}$ to a four vector $k = (\mathbf{k}, \nu)$. We do not do so in this Chapter for the equations, but in the figures k represents $(\mathbf{k}, \nu)$. Further it is important to choose a momentum-frequency convention for the vertices and stick to it. Because of energy and momentum conservation we only need three momenta and frequencies for the four points $1, 2, 3, 4$ of the two-particle vertices in Fig. 4. Unless noted otherwise, we use the ph convention which is the natural one for the ph channel and already used in Fig. 5. That is, the $1, 2, 3, 4$ frequency-momenta of Fig. 4 are related to Fig. 5 as follows: $k_1 = k$, $k_4 = k'$, $k_2 = k+q$, and—because of energy-momentum conservation— $k_3 = k'+q$. The Bethe-Salpeter equations in the other channels have the same structure just with another Γ_ℓ, $\ell = \overline{ph}$ or pp, with another way to connect the building blocks (see Fig. 4), and with another natural (diagonal) frequency-momentum q.

(5) The fifth equation is the Dyson equation that we already introduced [Fig. 3, Eq. (1)].

(6) Finally, the sixth equation is the Schwinger-Dyson equation [Fig. 6, Eq. (4)] which reads

$$\Sigma_{\mathbf{k}\nu} = \frac{Un}{2} - \frac{U}{2} \sum_{k',q} \sum_{\nu',\omega} \left(F^{\nu\nu'\omega}_{c,\mathbf{k}\mathbf{k}'\mathbf{q}} - F^{\nu\nu'\omega}_{s,\mathbf{k}\mathbf{k}'\mathbf{q}} \right) G_{(\mathbf{k}+\mathbf{q})(\nu+\omega)}\, G_{\mathbf{k}'\nu'}\, G_{(\mathbf{k}'+\mathbf{q})(\nu'+\omega)} \tag{4}$$

and connects Σ and F. Here we consider a single-orbital and a local interaction U that connects opposite spins. For this reason only $F_{\uparrow\downarrow} = (F_c - F_s)/2$ enters. The first term is simply the Hartree(-Fock) term, which is not included in Fig. 6 and where n is the average number of electrons per site in the paramagnetic phase.

This set of six exact parquet equations allows us to determine the six quantities F, Φ_ℓ, Σ, and G, as well as $\Gamma_\ell = F - \Phi_\ell$. If we knew the exact Λ, we could determine all of these quantities exactly, as well as associated one- and two-particle physics. Unfortunately, we do not know the exact Λ. Hence we need some approximation. If we approximate Λ by the bare Coulomb interaction U, we obtain the parquet approximation [17]. We can do better than this, and in

[3] We implicitly assume a proper normalization of the momentum and Matsubara frequency sums, i.e., $\sum_{\mathbf{k}} \widehat{=} \frac{1}{N_K} \sum_{\mathbf{k}}$ and $\sum_{\nu} \widehat{=} \frac{1}{\beta} \sum_{\nu}$, where N_k is the number of momentum points and $\beta = 1/T$ the inverse temperature.

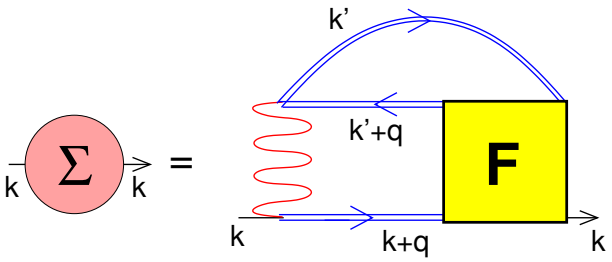


Fig. 6: *Schwinger-Dyson equation connecting the full vertex F and the self-energy Σ.*

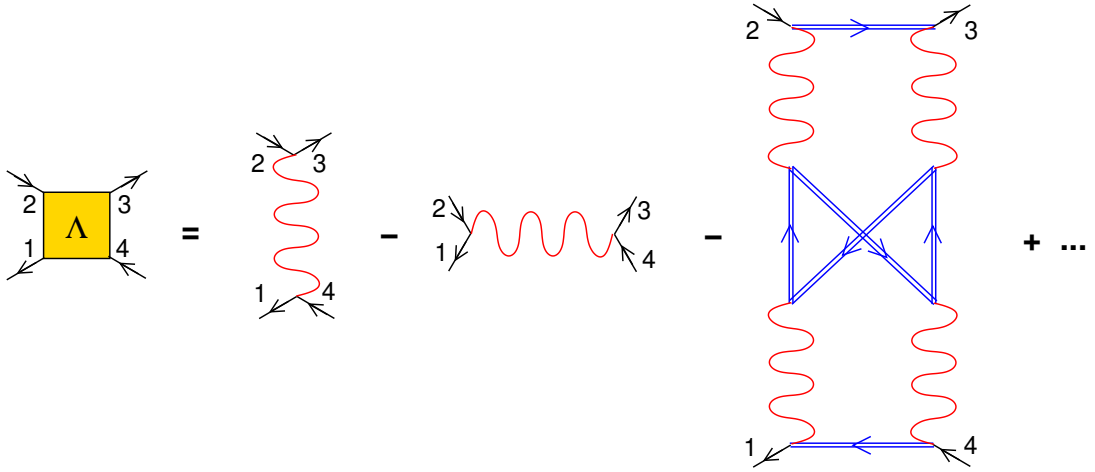


Fig. 7: *Lowest order Feynman diagrams that contribute to the fully irreducible vertex Λ.*

DΓA we approximate Λ by all local Feynman diagrams. Quantum Monte Carlo simulations show that this is an excellent approximation [10]. Indeed Λ is very compact. All two-particle reducible diagrams are generated from it and the first irreducible diagram that enters Λ besides the bare Coulomb interaction is of fourth order, see Fig. 7. So to speak the irreducible diagrams Λ form a skeleton from which many more diagrams are generated. Because of this, it is more local than Γ_ℓ and, in particular, much more local than F, Σ or G.

This local irreducible Λ is "summed up" in practice by solving an Anderson impurity model, similar as in DMFT but now we calculate the two-particle Green function of the impurity model and from this Λ. For example, we can use continuous-time quantum Monte Carlo simulations in the hybridization expansion see "QMC impurity solvers" by P. Werner [3], e.g., using the `w2dynamics` package [18] which allows us to calculate all two-particle responses using worm sampling [19].

In principle, one can then further turn to the $n = 3$-particle level etc.; and, for the $n \to \infty$-particle fully irreducible local vertex DΓA one recovers the exact solution. As a matter of course determining the $n = 3$-particle vertex becomes already cumbersome. But it may serve at least as an error estimate [11] if one is truncating the scheme at the two-particle vertex level. Also completely new physics, that is hitherto not understood, may be hiding behind diagrams generated from the $n = 3$-particle irreducible vertex. Some physical processes such as Raman scattering naturally require three-particle vertex functions.

In the present section we have learned about the fully fledged parquet DΓA. It is unbiased with respect to all three channels ℓ and thus treats antiferromagnetic or charge fluctuations in the ph-channel on a par with e.g., superconducting fluctuations or weak localization corrections in the pp channel. It also mixes the different channels. For two channels this mixing generates, with some fantasy, a traditional parquet floor like pattern with ladder rungs in one direction intermixed by ladder rungs in the orthogonal direction etc., hence the name. Being unbiased definitely is a great advantage. For example, in [14] we were looking for weak localization corrections to the optical conductivity in the pp channel and excitons that live in the ph channel. Instead we found that for strongly correlated systems that host strong alternating spin or charge fluctuations, the $\overline{ph}$ channel is actually dominant for the optical conductivity, against all expectations.

The drawback of the full parquet solution is that we deal with three momenta and three frequencies. Even on a rather coarse discretization grid, we thus easily end up with Terabytes of data. In the Bethe-Salpeter equation (3) the bosonic momentum $\mathbf{q}$ and frequency ω decouple, so that it can be well distributed on several computer cores without the need to communicate. However, this natural $\mathbf{q}$ and ω is channel dependent (see [5]), i.e., different for the three channels of Fig. 4. Hence, when we add the three channels in the parquet equation (2), we mix different momenta, requiring a lot of communication between different cores, which –say– have previously solved the Bathe-Salpeter equation in different channels. Network traffic thus becomes the computational bottleneck on a supercomputer for parquet DΓA. Efforts to mitigate this problem are the truncated unity basis for momentum space [20], the intermediate representation (IR) for frequency space [21], and the single boson exchange decomposition [22].

3 Ladder dynamical vertex approximation

If we know that a certain channel is dominating, we can restrict ourselves to this particular channel and neglect the others and the mixing of different channels through the parquet equation. Since Φ_{ph} and $\Phi_{\overline{ph}}$ are related by crossing symmetry (invariance of exchanging the two incoming lines [5]) both of these channels contribute to F in the same way, albeit with a crossing-symmetry exchanged frequency and momentum combination. Hence, both channels must be included in general.[4] In this case, we hence only need to solve the Bethe-Salpeter equation (3) and can consider a local Γ_{ph} as input. For a bare interaction U, the Bethe-Salpeter equation (3) yields diagrams as in Fig. 8(a). If we improve on this and use a local Γ_{ph} we get the ladder DΓA approach and the diagrams of Fig. 8(b). From these diagrams (plus the contribution from the crossing-symmetrically related $\overline{ph}$ channel) we get F, and from F in turn through the Schwinger-Dyson equation (4) the self-energy.

[4] When we eventually connect F with altogether four Green functions to a susceptibility with a single bosonic momentum and frequency [see Eq. (6) and Fig. 12 below)], the $\mathbf{q} = (\pi, \pi, \ldots)$ susceptibility is dominated in case of antiferromagnetic spin fluctuations by the ph channel, whereas the $\mathbf{q} = (0, 0, \ldots)$ optical susceptibility (conductivity) is dominated by the transversal particle-hole channel (π-tons [14]).

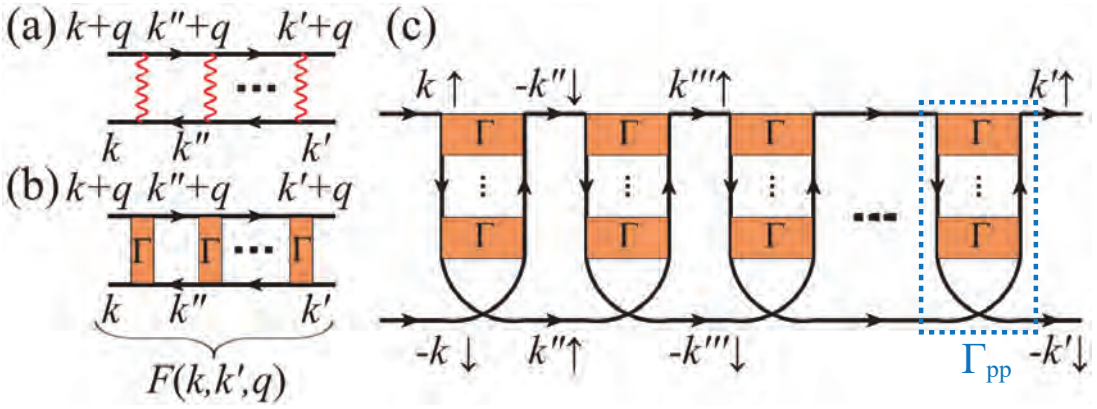


Fig. 8: *(a) Spin fluctuations as calculated from the Bethe-Salpeter ladder diagrams in terms of the bare Coulomb interaction U in RPA. (b) Same as (a) but now calculated in ladder DΓA from the local irreducible vertex in the ph channel: $\Gamma \equiv \Gamma_{ph}$. (c) These antiferromagnetic spin fluctuations enter as Γ_{pp} (via the parquet equation) in the particle-particle channel. This leads to the binding of two electrons into a Cooper pair, and if this bosonic Cooper pair Bose-Einstein condensates to superconductivity. From [25].*

If we want to calculate superconducting fluctuations in ladder DΓA we can plug the antiferromagnetic spin fluctuations as a superconducting pairing glue (which is nothing but Γ_{pp}) into the pp channel. This is so-to-speak a poor man's one-step parquet calculation. We get the ph and $\overline{ph}$ spin fluctuations into the pp channel, but we do not feed back the pp fluctuations to the ph and $\overline{ph}$ channel. For details on the superconducting calculations and also regarding the treatment of high frequencies, see [16].

A self-consistency with respect to the recalculation of the Green function entering the Bethe-Salpeter equation (3) is possible [23]. Different schemes have been proposed as well for a self-consistency with respect to Γ (or Λ) starting with [24], for an overview see [5]. A simpler and widely used approach is to do, instead, a so-called Moriyaesque λ-correction [5]. It essentially adds a mass to the paramagnons (dampens the antiferromagnetic spin fluctuations). This mass is fixed by a sum-rule for the susceptibility, and automatically warranties the correct high-frequency asymptotics of the self-energy. This λ-correction has been introduced to mimic the self-consistency and represents a considerable simplification of the calculations. For a further-going discussion see [5, 25].

The advantage of the ladder DΓA is that –as long as we do not couple the ladders through the parquet equations– the ladder only depends on a single frequency-momentum q instead of three (q,k,k'). Hence numerically much lower temperatures, larger frequency grids and finer momentum grids are feasible. Also realistic multi-orbital *ab initio* DΓA calculations are possible with the ladder DΓA version [15]. The results presented in the following have been obtained by ladder DΓA with a Moriyaesque λ-correction.

4 Hubbard model, cuprates and nickelates

In this Chapter, we consider the one-band Hubbard model in two or three-dimensions

$$\mathcal{H} = -\sum_{ij,\sigma} t_{ij}\, c^\dagger_{i\sigma} c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}. \tag{5}$$

It consists of two terms: (i) a hopping term t_{ij} between sites i and j, which we restrict in the following to a nearest neighbor t, next-nearest neighbor t' and next-next-nearest neighbor t''; and (ii) a local Coulomb repulsion U. Here $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) creates (annihilates) an electron on site i with spin σ in second quantization and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$.

The Hubbard model is the quintessential model for strongly correlated electron systems, similar as the Ising model for statistical physics or the Drosophila fly for genetics. For $t_{ij} = 0$, we have just a collection of atoms and a spectra with peaks at $\pm U/2$ for half-filling. For $U = 0$, there is no interaction and we can solve the tight-binding Hamiltonian by Fourier transforming t_{ij} to $\varepsilon_{\mathbf{k}}$ in momentum space. Then, just occupying all single-particle states up to the Fermi energy gives the ground state. But, if we switch on U the electrons become correlated, the expectation value $\langle n_{i\sigma} n_{j\sigma'}\rangle \neq \langle n_{i\sigma}\rangle\langle n_{j\sigma'}\rangle$. In particular local double occupations $\langle n_{i\uparrow} n_{i\downarrow}\rangle$ are heavily reduced compared to the non-interacting, uncorrelated value.

While a screened interaction can –to a good approximation– often be replaced by the purely local interaction U, materials require typically the consideration of more bands than the single-band Hubbard model. This is even the case when we restrict ourselves to the low-energy orbitals around the Fermi energy. An important exception are cuprate and nickelate superconductors.

The arguably simplest cuprate and nickelate crystal structure is the "infinite-layer"[5] structure displayed in Fig. 9 (A) and (B). Since the valence of the spacer cations is Ca^{+2} and $Nd(La)^{+3}$, the formal oxidation state is Cu^{+2} and Ni^{+1}, respectively, so that both cuprates and nickelates are in a formal $3d^9$ configuration. After such basic chemistry considerations, the first step to get an idea of the relevant orbitals is doing a density functional theory (DFT) calculation.

Fig. 10 shows that for the cuprates a single $3d_{x^2-y^2}$-orbital is crossing the Fermi level. It gives rise to the single hole-like Fermi surface sheet displayed in the lower left panel of Fig. 10. This Cu $3d_{x^2-y^2}$-band and Fermi surface can well be described with proper hopping parameters t, t' and t''. But now we need to include the effect of electronic correlations. These are indicated by the dashed arrows and the side panels of Fig. 10(A): the Cu $3d_{x^2-y^2}$-band splits into an upper and lower Hubbard band. Since the oxygen orbitals are just a few eV below the Fermi energy, these oxygen orbitals end up above the lower Hubbard band. As a consequence we have a charge transfer insulator according to the scheme of Zaanen-Sawatzky-Allen and not a Mott insulator, which one might have expected from the splitting into Hubbard bands. That is, if we dope cuprates, and we need to do so to have a superconductor, the holes go into the oxygen bands. Hence, in the case of cuprates, an Emery model description which incorporates the copper $3d_{x^2-y^2}$-band and oxygen p_x and p_y bands as visualized in Fig. 9(C) is the most appropriate low-energy model. Nonetheless, the majority of theoretical papers studying the

[5] The two distinct layers displayed in Fig. 9 are repeated on top of each other ad infinitum.

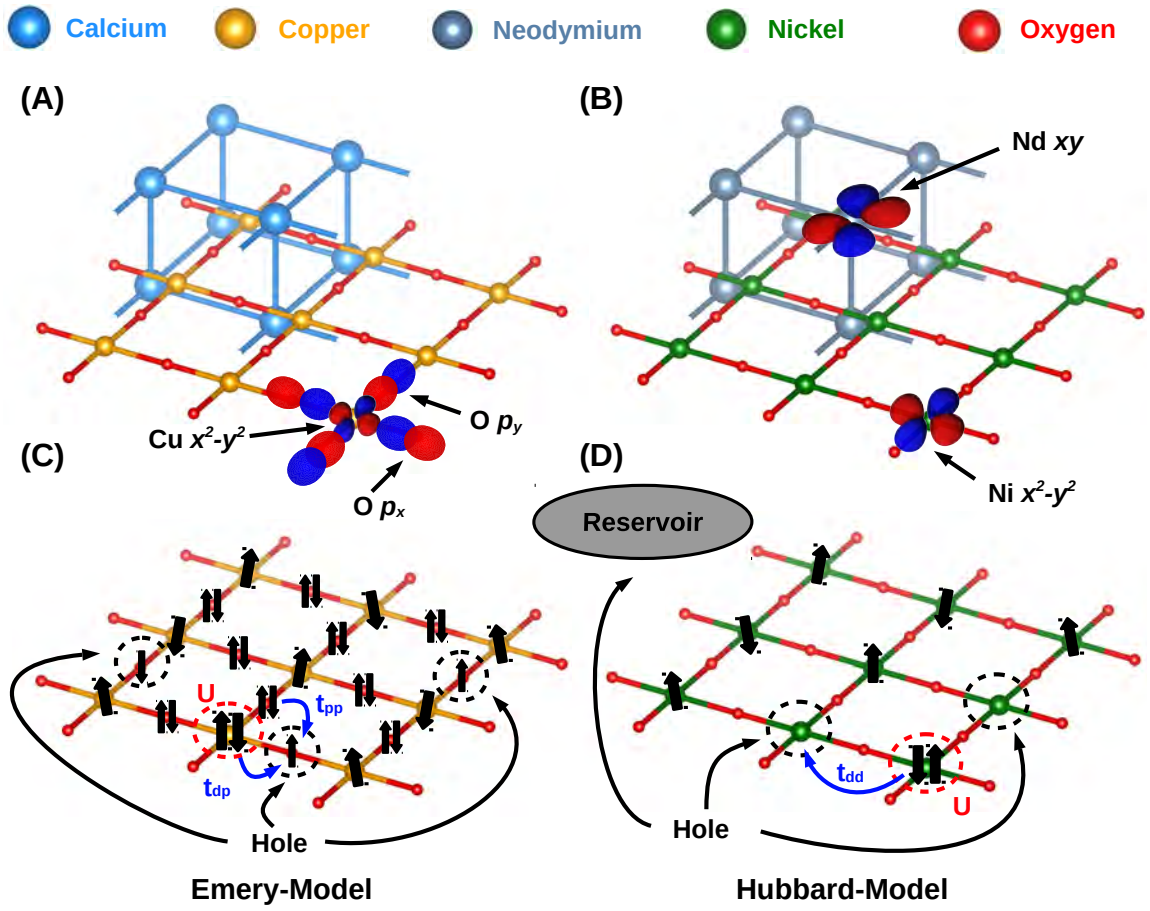


Fig. 9: *Crystal structure of (A) a characteristic cuprate, $CaCuO_2$, and (B) the nickelate $NdNiO_2$. The $Cu(Ni)O_2$ layers are separated by Ca(Nd) spacing layers in this "infinite-layer" structure. Also shown are the relevant low-energy orbitals in both cases. (C) Cuprates are charge transfer insulators so that the oxygen orbitals need to be included for an accurate microscopic description. The simplest such model is the* **Emery model** *with copper $3d_{x^2-y^2}$, oxygen p_x and p_y orbitals. As indicated there is a hopping t_{pd} between Cu and O sites, and t_{pp} between O sites; double occupations are suppressed by the interaction U on the Cu sites just as in the Hubbard model. (D) For nickelates we have a Ni-$3d_{x^2-y^2}$-band and a A-pocket derived from the Nd $5d_{xy}$ band crossing the Fermi energy. Both are largely decoupled but we need to calculate, e.g., by DFT+DMFT, how many holes go into the Ni-$3d_{x^2-y^2}$-band* **Hubbard model** *and how many go into the A-pocket* **reservoir**. *From [26].*

cuprates use the single-band Hubbard model. This is to some extend justified by the fact that oxygen-hole spin and copper spin form a Zhang-Rice singlet, which can be described effectively by the Hubbard model. Also in experiment, there is a single Fermi surface as in Fig. 10(C) of mixed oxygen and copper character.

In case of the nickelates, these oxygen bands are instead at a considerably lower energy. Hence the lower Hubbard band is now above the oxygen band and we would have a Mott insulator – if it was not for the Nd(La) $4d$-bands.[6] These Nd(La) $4d$ bands are closer to the Ni $3d$-bands

[6]We here show $LaNiO_2$ since it is slightly easier to calculate than $NdNiO_2$ in DFT because it has no $4f$-electrons. Experimentally, $La_{1-x}Sr(Ca)_xNiO_2$ is also a nickelate superconductor with a T_c comparable to $NdNiO_2$.

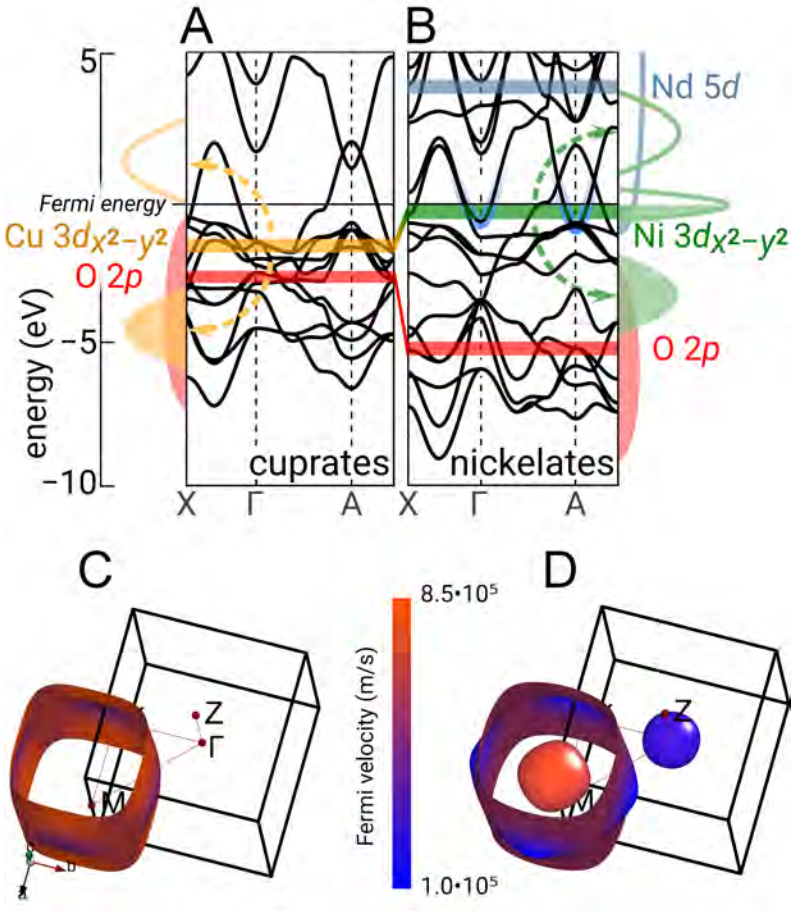


Fig. 10: *(A,B) Electronic structures of (A) $CaCuO_2$ and (B) $LaNiO_2$ as calculated in DFT. The side panels indicate the effect of electronic correlations and the color bars the center of energy of the most important bands. (C) and (D) DFT Fermi surface corresponding to (A) and (B), respectively. From [26].*

than the unoccupied Ca and Cu s bands are to the Cu bands. Hence they briefly cross the Fermi level around the Γ and A momentum, which leads to the formation of pockets around these momenta, see Fig. 10(D). The charge transferred to the Nd(La) pockets also leads to a self-doping of the Ni $3d_{x^2-y^2}$-band, even for the parent compound Nd(La)NiO_2 which has about 0.05 holes per Ni. In such a situation, a quasi-particle peak develops as indicated in the side panel of Fig. 10 (B) with a quasiparticle mass enhancement $m^*/m = 1/Z$ calculated to be about five [27]. A further correlation effect is that the Γ-pocket is shifted above the Fermi level, at least for larger Sr-doping and for $LaNiO_2$.

Altogether this leaves us with two relevant bands as displayed in Fig. 9(B): the Ni $3d_{x^2-y^2}$-band and the Nd(La)-derived A-pocket. Both bands do not hybridize and can hence, to a first approximation, be considered as decoupled. The expectation is that the more strongly correlated Ni $3d_{x^2-y^2}$-band is responsible for the superconductivity. Using Occam's razor, i.e., if we try to identify the most simple model, we end up with a one-band Hubbard model for the Ni $3d_{x^2-y^2}$-band and a decoupled reservoir (A-pocket) that must be taken into account for translating the

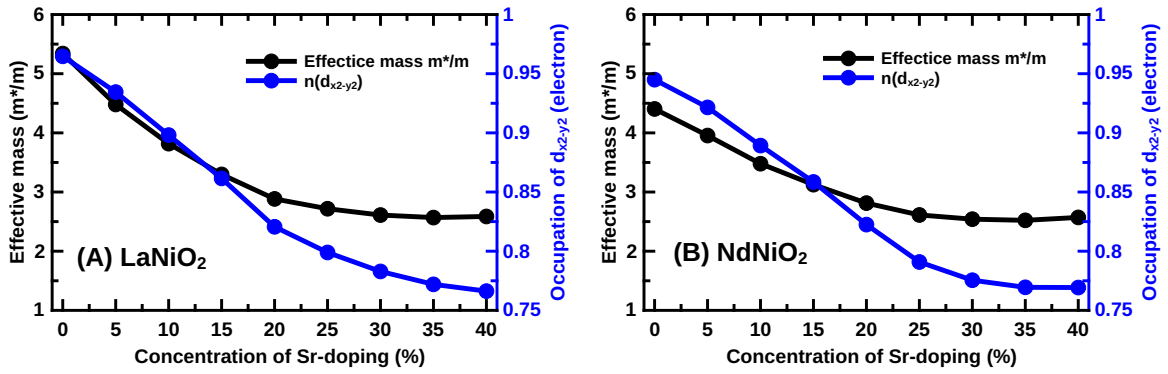


Fig. 11: *Translation of Sr-doping to the occupation of the Ni $3d_{x^2-y^2}$-band (blue line; right y-axis) and corresponding effective mass enhancement m^*/m (black line; left y-axis) as calculated by DFT+DMFT for (A) $La_{1-x}Sr_xNiO_2$ and (B) $Nd_{1-x}Sr_xNiO_2$. From [27] (Supplemental Material).*

Sr-doping of Nd(La)$_{1-x}$Sr$_x$NiO$_2$ into the actual hole doping of this Hubbard model. Otherwise the A-pocket is merely a passive bystander. This simple model is illustrated in Fig. 9(D).

The hopping parameters of the nickelate Hubbard model can be obtained from a Wannier function projection onto the Ni $3d_{x^2-y^2}$-band: $t = 0.395\,\text{eV}$, $t'/t = -0.25$, $t''/t = 0.12$ [27]. Further, the interaction strength U can be calculated by constrained random phase approximation (cRPA, see Chapter "The GW+EDMFT method" by F. Aryasetiawan [3]). Considering the fact that U is frequency(ω)-dependent in cRPA and –within the relevant energy range– on average slightly larger than $U(\omega{=}0)$, one obtains $U \approx 8t$ from cRPA. For further details see [27].

The translation from Sr-doping to the occupation of the $3d_{x^2-y^2}$-band has been calculated in DFT+DMFT[7] including all five Nd and all five Ni d bands in DMFT. Fig. 11 (blue curve) shows the results: One sees that roughly 50% of the holes (there is one hole per Sr) go into the $3d_{x^2-y^2}$-band and the remaining 50% go into the pockets. For larger Sr-dopings, at around 25%, the curve flattens because here the Ni $3d_{3z^2-r^2}$ orbital approaches the Fermi level and accommodates holes as well. The one-band Hubbard model DMFT calculation gives a very similar spectrum as the full DFT+DMFT calculations with $5{+}5$ Nd+Ni orbitals. Also the effective mass plotted in Fig. 11 (black curve) agrees [27]. Altogether this hints that the simple Hubbard model is a good approximation for nickelates; experimental results are also consistent with this picture so-far.

[7]For the DFT+DMFT method, see the Chapter "LDA+DMFT for strongly correlated materials" by A. Lichtenstein [3]

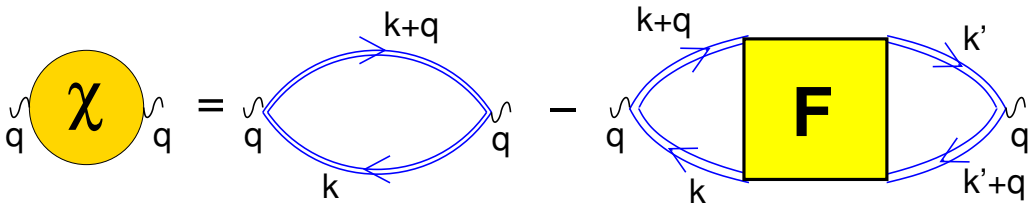


Fig. 12: *The susceptibility χ consists of the bare bubble χ^0 (first term on the right hand side) and vertex corrections that can be calculated from the full vertex F.*

5 Spin fluctuations

Spin fluctuations as visualized in Fig. 8(a,b) enter the full vertex F but subsequently also the susceptibility which is connected to F as (Fig.12)

$$\chi_{r,\mathbf{q}\omega} = \underbrace{\sum_{\mathbf{k}\nu} G_{\mathbf{k}\nu}\, G_{(\mathbf{k}+\mathbf{q})(\nu+\omega)}}_{\equiv\chi^0_{\mathbf{q}\omega}} - \sum_{\mathbf{k}\mathbf{k}\prime\nu\nu'} G_{\mathbf{k}\nu}\, G_{(\mathbf{k}+\mathbf{q})(\nu+\omega)}\, F^{\nu\nu'\omega}_{r,\mathbf{k}+\mathbf{k}'\mathbf{q}}\, G_{\mathbf{k}'\nu}\, G_{(\mathbf{k}'+\mathbf{q})(\nu+\omega)} \qquad (6)$$

for $r = c/s$ i.e., the spin and charge susceptibility, respectively. As before, we restrict ourselves to the paramagnetic phase for simplicity. The first term on the right hand side of Eq. (6) is just the bare bubble contribution χ^0, obtained directly from two (interacting) Green functions; the second term are vertex corrections calculated form F. The minus sign is a matter of definition of F. When F is calculated in RPA as in Fig. 8(a), we have $\Gamma_r = \pm U$ for $r = c/s$ and obtain a geometric sum, which eventually yields

$$\chi^{\mathrm{RPA}}_{r,\mathbf{q}\omega} = \frac{\chi^0_{\mathbf{q}\omega}}{1 \pm U\chi^0_{\mathbf{q}\omega}} \qquad (7)$$

where the +/- is for $r = c/s$.

For certain ω's and $\mathbf{q}$'s Eq. (7) [and Eq. (6)] develops poles. We call these bosonic excitations (para)magnons. The ω-$\mathbf{q}$ energy-momentum dispersion relation of these quasiparticles follows the position of the poles.

The minus sign in Eq. (7) already indicates that spin fluctuations are generally stronger in a Hubbard model. This can change if additional non-local interactions V are included. These trigger a transition from an antiferromagnetic spin to a charge density wave order for $\mathcal{Z}V > U$ in mean field ($\mathcal{Z}$: number of neighbors; V: nearest neighbor Coulomb repulsion). Which magnetic order dominates in RPA, solely depends on the $\mathbf{q}$ for which $\chi^0_{\mathbf{q}\,\omega=0}$ is strongest. Close to half-filling, we often have Fermi surfaces where for a $\mathbf{k}$ on the Fermi surface ($\nu = 0$) also $\mathbf{k}+\mathbf{q} = (\pi, \pi, \ldots)$ is at or close to the Fermi surface. Then both $G_{\mathbf{k}\nu=0}$ and $G_{(\mathbf{k}+(\pi,\pi,\ldots))(\nu+\omega)=0}$ are large in Eq. (6), and the antiferromagnetic susceptibility $\chi^0_{\mathbf{q}=(\pi,\pi,\ldots)\omega=0}$ is maximal. Of course, this is just the weak coupling picture. For larger Coulomb interactions we form large magnetic moments and the change of physics is reflected in vertex F corrections beyond RPA. More recently it could be demonstrated [25] that the local vertex $\Gamma_{r=m}$ is suppressed compared to the RPA $\Gamma_s = -U$ value. This is shown in Fig. 13 (left). This suppression, in particular

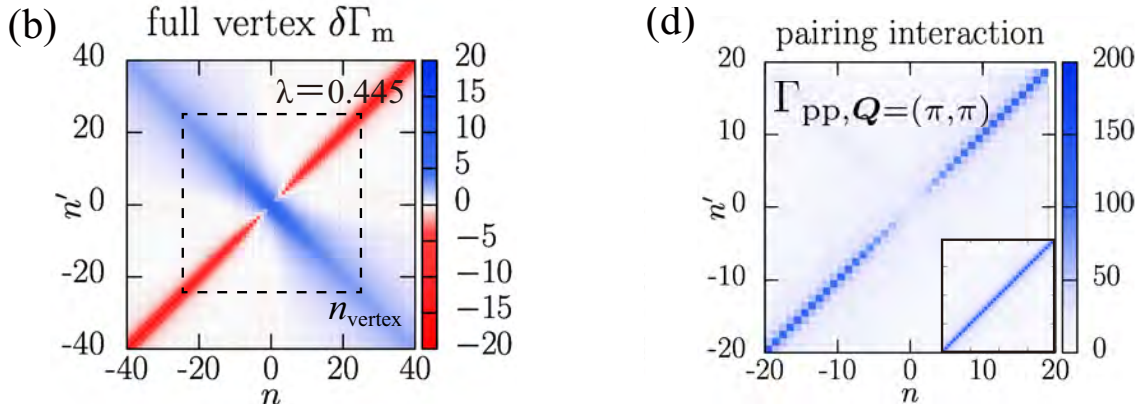


Fig. 13: *Left: Difference of the local ph vertex in the magnetic channel from $-U$, i.e., $\delta\Gamma_m = \Gamma_s - (-U)$ vs. the two fermionic Matsubara frequencies $\nu_n = i(2n+1)\pi T$ and $\nu'_{n'}$ at zero bosonic frequency $\omega = 0$. At small frequencies, $\Gamma_{s,ph}$ and associated spin fluctuations are suppressed. Right: This suppression also leads to a suppression of the non-local pp vertex Γ_{pp}, i.e., the superconducting pairing glue, at small frequencies. It is obtained from antiferromagnetic spin fluctuations (i.e., from the local Γ_s corresponding to the left panel and the Bethe-Salpeter equation, see Sec. 7). Right inset: Γ_{pp} as calculated in RPA. From [25] where further details and parameters can be found.*

that at the relevant small frequencies, leads to reduced spin fluctuations. Consequently, the superconducting pairing as calculated along the line of Fig. 8(b,c) is suppressed as well, see Fig. 13 (right). Its origin are local particle-particle excitations that enter Γ_s and reduce it along the side diagonal frequencies $\nu = -\nu'$.

This reduction of the DΓA spin susceptibility is key for a good description of antiferromagnetic spin fluctuations, which are grossly overestimated in RPA for somewhat larger Coulomb interactions U. In fact, it was shown that qualitatively and quantitatively the DΓA susceptibility excellently agrees with other recent numerical calculations [28].

From the susceptibility one can obtain the correlation length ξ. Fourier-transformed to real space $\mathbf{r}$ and time, the (equal-time) magnetic susceptibility behaves as

$$\chi_{s\mathbf{r}} = \langle \mathbf{S}(\mathbf{r})\mathbf{S}(\mathbf{0}) \rangle \sim \left(\frac{|\mathbf{r}|}{\xi}\right)^{-1/2} e^{-\frac{|\mathbf{r}|}{\xi}} \tag{8}$$

at large distances $\mathbf{r}$. Here, $\mathbf{S}(\mathbf{r})$ denotes the spin operator of the electrons at position (lattice site) $\mathbf{r}$. Fig. 14 shows this spin-spin correlations or susceptibility of the Hubbard model. Clearly, an alternating (antiferromagnetic) correlation is visible. At high temperatures (left panel), the correlation length is short and correlations quickly decay. In two dimensions we get, however, an exponential increase of the correlation length with $1/T$ in DΓA [13] which is the reason behind the very large correlation length of Fig. 14 (right panel). While a correlation length of $\xi = 4$ lattice sites (left panel) can still be covered in cluster extensions of DMFT and lattice quantum Monte Carlo methods, the rapidly increasing correlation length at lower temperatures quickly puts a numerical limit to such cluster approaches.

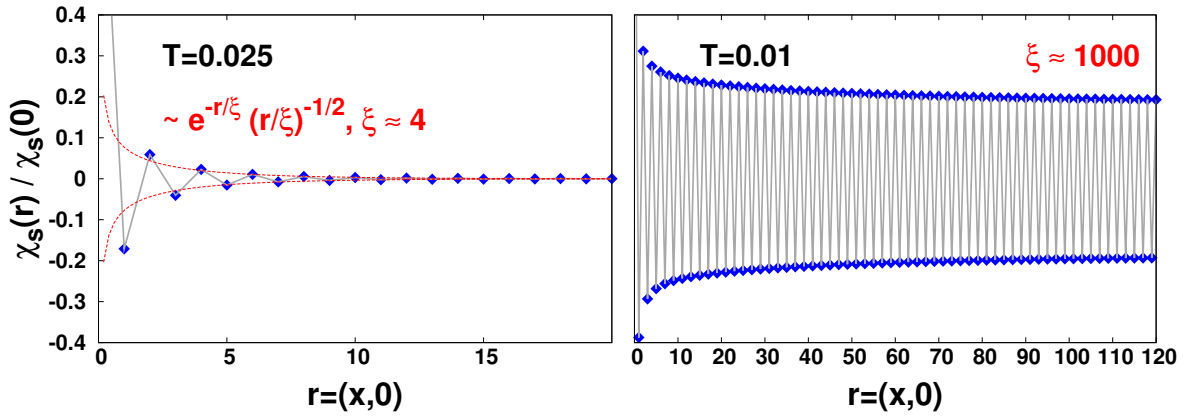


Fig. 14: *Antiferromagnetic correlation function* $\chi_{s,\mathbf{r}}$ *normalized to its* $\mathbf{r} = 0$ *value vs. distance* $\mathbf{r}$ *for the two dimensional Hubbard model along the* x*-direction* ($t' = t'' = 0$, $U = 0.5(4t)$; *all energies are here in units of* $4t$). *The left and right panel show two different temperatures,* $T = 0.025(4t)$ *(left) and* $T = 0.025(4t)$ *(right), where largely different correlation lengths* (ξ) *are observed. From [13].*

At a three dimensional phase transition, the susceptibility diverges with critical exponent ν: $\xi \sim (T-T_c)^{-\nu}$, and similarly $\chi_{r,\mathbf{Q}\omega=0} \sim (T-T_c)^{-\gamma}$ with a critical exponent γ. These critical exponents could be calculated in DΓA and DF for the first time for correlated electronic models, a topic that has been covered in a preceding Jülich Autumn School [12].

In practice, the correlation length is not calculated from fitting $\chi_{s\mathbf{r}}$. Instead one fits it Fourier-transform $\chi_{s,\mathbf{q}\omega=0}$ in momentum space. For large correlation lengths and in the vicinity of its maximum at the dominating wave vector $\mathbf{Q}$,[8] the susceptibility is of the Ornstein-Zernike form

$$\chi_{s,\mathbf{q}\,\omega=0} \sim \frac{1}{(\mathbf{q}-\mathbf{Q})^2 + \xi^{-2}}. \tag{9}$$

That is, the inverse correlation length ξ^{-1} corresponds to the width of the susceptibility around its peak at $\mathbf{q} = \mathbf{Q}$.

In two dimensions, the Mermin-Wagner theorem is fulfilled and there is no long range antiferromagnetic order for finite temperature in the DΓA and with the diagrammatic extension of DMFT [5]. In three-dimensions the antiferromagnetic phase transition temperature T_c is reduced compared to DMFT, as shown in Fig. 15. Also the critical behavior in the vicinity of the phase transition is not any longer of mean-field type as in DMFT. Instead the critical exponents well agree[9] with those of the Heisenberg model [5, 12] for half-filling, as to be expected from universality.

[8] For example, $\mathbf{Q} = (\pi, \pi)$ for the two-dimensional Hubbard model at half-filling.

[9] within the numerical error bars in the studied critical temperature regime

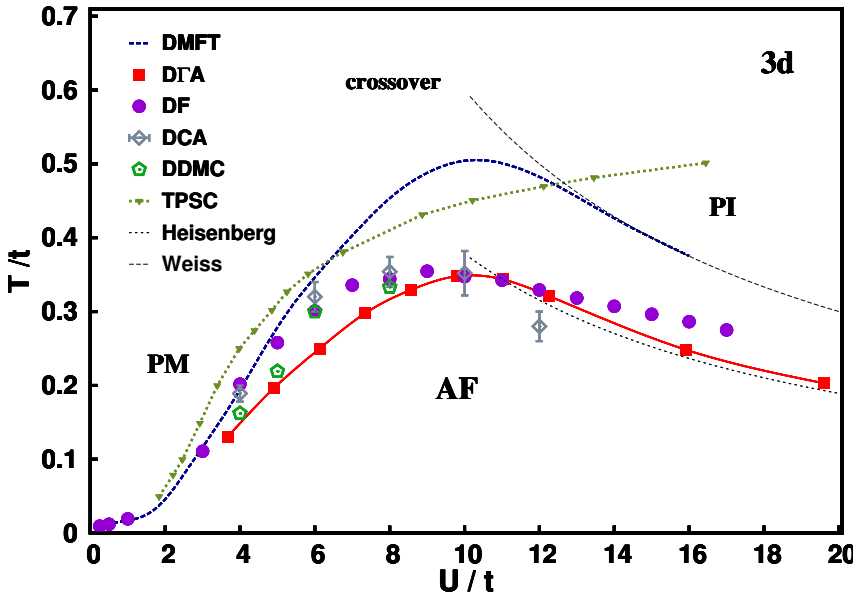


Fig. 15: *Phase diagram of the half-filled three-dimensional Hubbard model on a cubic lattice with nearest-neighbor hopping t comparing various methods for the phase transition from the paramagnetic metallic (PM) and insulating (PI) phase to the antiferromagnetic insulating phase (AF). From [5] where further details and the abbreviation of the various methods can be found.*

6 Pseudogap

Besides the susceptibility, spin fluctuations also impact the electronic self-energy, as already indicated in Fig. 1. The effect of F on the self-energy Σ can be directly calculated via the Schwinger-Dyson Eq. (4) [Fig. 6]. Physically, this corresponds to electron-(para)magnon scattering which can reduce the lifetimes of the quasiparticles. An alternative way to write the Schwinger-Dyson equation and to express the fermion-boson interaction is shown in Fig. 16. Here, simply the F and two Green functions from Fig. 6 have been combined to γ, i.e., mathematically

$$\gamma^{\nu\omega}_{r,\mathbf{kq}} = \sum_{\mathbf{k}'\nu'} F_{r,\mathbf{kk}'\mathbf{q}}\, G_{\mathbf{k}'\nu'}\, G_{(\mathbf{k}'+\mathbf{q})(\nu'+\omega)} \tag{10}$$

so that the Schwinger-Dyson equation becomes

$$\Sigma_{\mathbf{k}\nu} = \frac{Un}{2} - \frac{U}{2} \sum_{\mathbf{q}\,\omega} \left(\gamma^{\nu\omega}_{c,\mathbf{kq}} - \gamma^{\nu\omega}_{s,\mathbf{kq}}\right) G_{(\mathbf{k}+\mathbf{q})(\nu+\omega)} \ . \tag{11}$$

Now, we can rewrite this further by merging U-reducible diagrams into an effective interaction W_r, in the spirit of Hedin's GW method (we will not go into further details here and refer the reader to the Chapter " The GW+EDMFT method" by F. Aryasetiawan [3], and to [22]). Beyond GW, here also spin fluctuations are included in W_s. This is displayed in Fig. 16 (right). The remaining U-irreducible $\tilde{\gamma}$ can be interpreted as the fermion-boson interaction and W as a boson propagator. In the case of GW these bosons are plasmons (charge fluctuations), while for the one-band Hubbard model (para)magnons (spin fluctuations) dominate. Such a reformulation of the Feynman diagrams is also the idea of the single boson exchange (SBE) approach [22], which

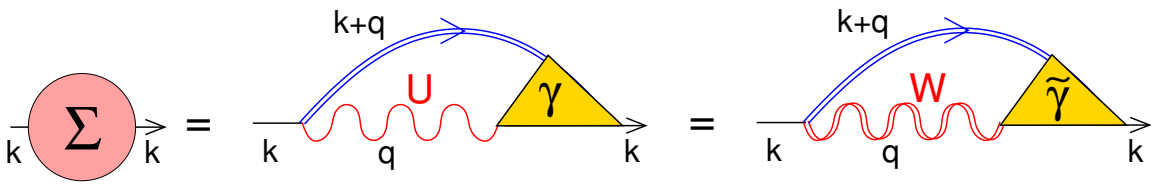


Fig. 16: *Schwinger-Dyson equation to calculate the self-energy Σ rewritten in terms of U and γ or W and $\tilde{\gamma}$.*

has been used recently to rewrite the DF and DΓA parquet equations. Much of the physics is already contained in the boson propagators W so that the remaining fully irreducible parts of the SBE parquet formalism decay much faster in frequency and momentum.

Particularly strong spin fluctuations, as they occur for the Hubbard model in two dimensions, can give rise to a physical phenomenon called pseudogap. Theoretically, this pseudogap has been established in numerical simulations and in weak coupling calculations. Experimentally, it has been observed for cuprate superconductors, both in the spectrum as well as in transport measurements. This kind of physics is naturally included in DΓA and other diagrammatic extensions of DMFT.

Fig. 17 shows the Fermi surface of the two-dimensional Hubbard model for typical hopping parameters of Cu(Ni) $3d_{x^2-y^2}$ orbitals in cuprates (nickelates). If we want to calculate the self-energy for a momentum $\mathbf{k}$ on the Fermi surface, it will be effected by spin fluctuations. These effects can be calculated via Eq. (11) [Fig. 16]. In Fig. 17 three particular points of the Fermi surface are displayed: The antinodal momentum on the Fermi surface (PG) is close to $(\pi, 0)$. Here, the pseudogap first opens in experiment and in numerical calculations for the Hubbard model at strong coupling. The nodal momentum (ARC) along the diagonal where an ARC-like part of the Fermi surface survives after the opening of the pseudogap around PG [see Fig. 19 (rightmost panel) below]. Finally, the hot spot (HS) which is the point of the Fermi surface where $\mathbf{k}+(\pi,\pi)$ lies on the Fermi surface as well. At weak coupling, the pseudogap opens first at the HS.

The spin-fermion coupling of Fig. 16 leads to an additional contribution to the self-energy on top of the local DMFT self-energy. In the Bethe-Salpeter equation Eq. (4) or equivalently Eq. (11), the vertex F respectively γ – or in Fig. 16 (right) W– are dominated by spin-fluctuations for the two-dimensional Hubbard model. These spin fluctuations are strongly peaked around the antiferromagnetic weave vector $\mathbf{Q} = (\pi, \pi)$ and can be modeled approximately by the Ornstein-Zernicke form Eq. (9) around $\mathbf{Q}$. Hence the spin-fermion contribution to the self-energy can approximately written as (see [29])

$$\Sigma_{\mathbf{k}\,\nu=i\delta} \propto \tilde{\gamma} T \sum_{\mathbf{q}} \frac{G_{(\mathbf{k}+\mathbf{q})\,\nu=i\delta}}{(\mathbf{q}-\mathbf{Q})^2 + \xi^{-2}}\,, \tag{12}$$

employing the rapid decay of $\tilde{\gamma}$ with bosonic frequency ω and restricting ourselves to the Fermi energy ($\nu = i\delta$, $\delta \to 0$)[10]. Further one can replace G by G^0 so that Eq. (12) can also be

[10]One can approximate this, at the cost of some additional broadening, by the lowest Matsubara frequency $\nu_0 = i\pi T$.

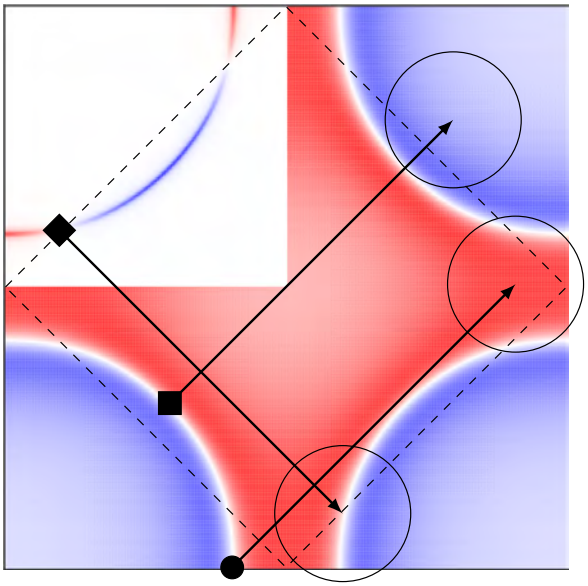


Fig. 17: *Typical Fermi surface of the t-t'-t'' Hubbard model. The blue and red colors denote the occupied and unoccupied part of the Brillouin zone, respectively. The circle (PG), square (ARC) and diamond (HS) denote specific momenta* $\mathbf{k}$ *that are discussed in the text. The corresponding momenta* $\mathbf{k} + (\pm\pi, \pm\pi)$ *connected to these by the antiferromagnetic wave vector* $\mathbf{Q} = (\pi, \pi)$ *are also plotted, as circles with typical inverse correlation lengths* ξ^{-1} *as radius. The upper left quadrant shows the parts of the Fermi surface that are "heated" (red) and "cooled" (blue), i.e., become less and more coherent because of* $\mathrm{Im}\tilde{\gamma}$ *(see text). From [29].*

employed as an ansatz for the spin-fermion self-energy or for fitting the pseudogap with only two fee parameters [$\tilde{\gamma}$ and ξ; if $\mathbf{Q}$ is known].

The sum of Eq. (12) is dominated by $\mathbf{q} \approx \mathbf{Q}$ where the antiferromagnetic spin fluctuations are strongest. In weak coupling theory $\tilde{\gamma} = 1$, and we get a damping, i.e., an imaginary part of the self-energy whenever $G_{(\mathbf{k}+\mathbf{Q})0}$ has a sizable imaginary part. This is the case whenever we have spectral weight $A_{(\mathbf{k}+\mathbf{Q})\nu} = -\frac{1}{\pi}\mathrm{Im}G_{(\mathbf{k}+\mathbf{Q})\nu}$ for (real frequency) $\nu = 0$. For $G = G^0$, i.e., without dampening, this is only possible if $\mathbf{k} + \mathbf{Q}$ is on the Fermi surface, too. Deviations by about the inverse correlation length (circles in Fig. 17) are possible, as then antiferromagnetic correlations are still sizable (which is the essence of the Ornstein-Zernicke form).

The condition that, for a considered $\mathbf{k}$ on the Fermi surface, also $\mathbf{k} + \mathbf{Q}$ is on the Fermi surface is exactly fulfilled for the hot spots (HS) in Fig. 17. At weak coupling, the prefactors of Eq. (12) are small and we need very long correlation lengths ξ to get a sizable imaginary part of the self-energy. Hence, here the pseudogap opens first at the hot spots.

At strong coupling the typical inverse correlation lengths are as displayed in Fig. 17 for the onset of the pseudogap. Numerical calculations (see Figs. 18, 19 below) and experiment show that the pseudogap opens first at the part of the Fermi surface marked as "PG" in Fig. 17. One reason for this might be that with the shorter correlation length, i.e., larger ξ^{-1}, the large spectral contribution of the van Hove singularity at $(\pi, 0)$[11] becomes relevant.

[11] and cubic-symmetrically related momenta

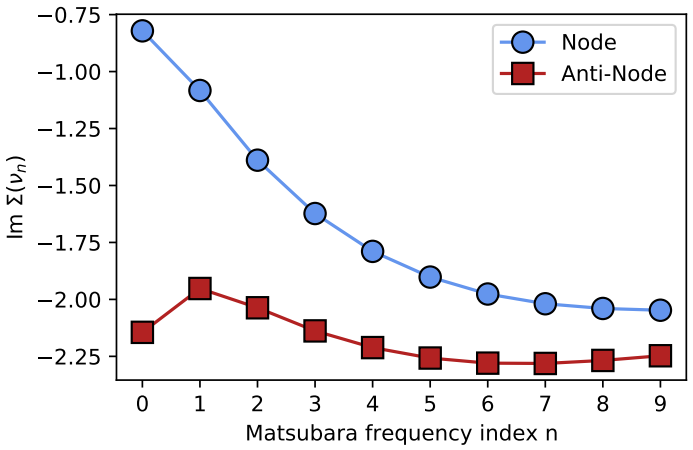


Fig. 18: *Imaginary part of the DΓA self-energy vs. index n of the Matsubara frequency $\nu_n = i(2n+1)\pi T$, in the pseudogap region for a typical Hubbard model. Parameters: $U = 8$, $t' = -0.2$, $t'' = 0.1$, $T = 0.05$ (all in units of t), and 10% hole doping. Clearly the self-energy is very different for the two momenta plotted, i..e, in the nodal direction of the Fermi surface and in the anti-nodal direction. The downturn of the anti-nodal self-energy indicates the development of a pole and thus a splitting of the spectral function away from the Fermi energy. Figure by courtesy of Paul Worm.*

A second mechanism has been identified in [29]: the spin-fermion interaction $\tilde{\gamma}$ develops an imaginary part because of particle-hole asymmetry at strong U. Hence we also get an imaginary part for Σ in Eq. (12) from the *real part* of the Green function $G_{(\mathbf{k}+\mathbf{q})\,\nu=0}$,

$$\mathrm{Re}G_{\mathbf{k}\,\nu=i\delta} = \frac{\mu - \varepsilon_{\mathbf{k}} - \mathrm{Re}\Sigma_{\mathbf{k}\,\nu=i\delta}}{\left(\delta - \mathrm{Im}\Sigma_{(\mathbf{k})\,\nu=i\delta}\right)^2 + \left(\mu - \varepsilon_{\mathbf{k}} - \mathrm{Re}\Sigma_{\mathbf{k}\,\nu=i\delta}\right)^2} \,. \tag{13}$$

This real part is displayed as false color in Fig. 17 for the non-interacting Green function (Σ=0). It has opposite sign (blue vs. red in Fig. 17) for the occupied (unoccupied) part of the Brillouin zone, where $\varepsilon_{\mathbf{k}} + \mathrm{Re}\Sigma_{\mathbf{k}\,\nu=i\delta} < (>)\mu$; μ is the chemical potential.

Hence, for the PG momentum where we scatter into the blue occupied part in Fig. 17 the imaginary part of the DMFT self-energy is enhanced by non-local correlations. This part of the Fermi surface is hence strongly dampened and eventually develops a pseudogap. In contrast, for the ARC momentum the imaginary part of the DMFT self-energy is even reduced (in absolute terms). The ARC quasiparticles are so-to-speak "cooled", i.e., become even more coherent because of spin fluctuations. This dichotomy is another mechanism that opens the pseudogap first in the nodal (PG) region and not at the hot spot (HS) if we are at strong coupling.

Fig. 18 shows the typical momentum differentiation that we obtain for the self-energy in the pseudogap region. The self-energy in the nodal and anti-nodal region is largely different. In the nodal direction, $\mathrm{Im}\Sigma_{\mathbf{k},\nu} \sim -\nu$ with the slope corresponding to the quasiparticle weight Z at this momentum. At the opening of the pseudogap, for momenta in the PG region, the self-energy first develops a large imaginary part, $\lim_{\nu\to 0} \mathrm{Im}\Sigma_{\mathbf{k},\nu}$. This corresponds to a strong dampening or extremely short life times of quasiparticles in this region. In Fig. 18 this would mean a flat curve for small frequencies. Eventually, the self-energy develops a pole, and we

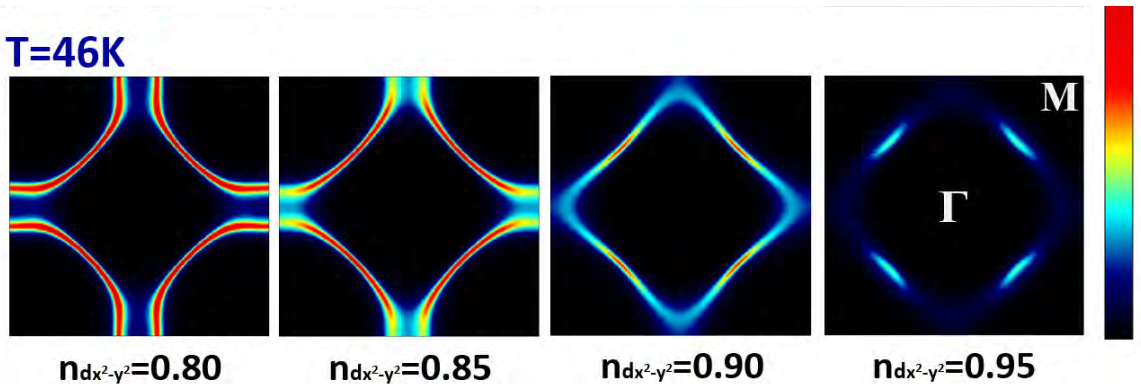


Fig. 19: *DΓA spectrum at the Fermi energy for different dopings of nickelate superconductors throughout the Brillouin zone (i.e, x- and y-axis are the k_x and k_y momentum ranging from $-\pi$ to π, and the false color show the spectral function at the Fermi energy $\nu = 0$). From [27].*

can see in Fig. 18 the onset of such a pole as the downturn of $\mathrm{Im}\Sigma_{\mathrm{k},\nu}$ for small Matsubara frequencies.

This is akin to the Mott insulator where the self-energy behaves as

$$\Sigma_{\mathbf{k},\nu} = \frac{U^2}{4\nu} \tag{14}$$

However, now this pole only occurs in the PG momentum regime, and the origin are spin fluctuations not Mott-Hubbard physics. Consequently the prefect or, being given by the spin fluctuation strength, is smaller than $\frac{U^2}{4}$. In the PG regime, the spectrum is hence gapped because of this pole structure (or strongly suppressed if Im is only large), whereas the ARC region still shows spectral weight.

Let us now turn to the nickelate superconductors. Fig. 19 shows the spectrum of the Ni $3d_{x^2-y^2}$ band[12] for nickelates as calculated by DΓA for the effective Hubbard model (see Sec. 4 where also the parameters can be found). The parent compound $NdNiO_2$, corresponding to $n_{3d_{x^2-y^2}} = 0.95$ (rightmost panel) has a pseudogap in the antinodal (PG) region, indicated by missing spectral weight in Fig. 19. This DΓA prediction still awaits an experimental confirmation; angular resolved photoemission experiments are urgently needed but have not been successfully done yet.

For making nickelates superconducting, one needs to dope this $3d_{x^2-y^2}$ band, e.g., to $n_{3d_{x^2-y^2}} = 0.80$ or 0.85. For these superconducting dopings, we have a clear Fermi surface throughout the Brillouin zone in Fig. 19 and no pseudogap.[13]

[12] There are additional Nd pockets, see Sec. 4.

[13] As a technical remark, the spectrum has been calculated from $A_{\mathbf{k}\nu_0} = -1/\pi\ G_{\mathbf{k}\nu_0}$ at the lowest Matsubara frequency $\nu_0 = i\pi T$, which leads to some additional broadening (smearing) but avoids the error-prone and cumbersome analytical continuation.

7 Superconductivity

Next we turn to the task of calculating the superconducting order and critical temperature T_c from the antiferromagnetic spin fluctuations. The general procedure has already been indicated in Fig. 8(c). More specifically, the first step is the calculation of the full vertex F from Γ_{ph} (plus the crossing symmetrically related $\overline{ph}$ channel), by summing the Bethe-Salpeter ladder diagrams in this channel(s). This calculation includes spin and charge fluctuations, but in the Hubbard model spin fluctuations prevail. In the second step, we insert these spin fluctuations into the pp channel. That is, we calculate $\Gamma_{pp} = \Lambda + \Phi_{ph} + \Phi_{\overline{ph}}$. Since no pp reducible diagrams have yet been included (except for the local ones), we can also rewrite this as $\Gamma_{pp} = F - \Phi_{pp}^{\rm local}$. From this Γ_{pp} we can next recalculate F including superconducting fluctuations via the Bethe-Salpeter equation in the pp channel. This is akin to the Schwinger-Dyson Eq. (3) in the ph channel, except that we have to properly rotate the momenta and to use Γ_{pp} instead of Γ_{ph}.
Similarly as in RPA [Eq. (7)], we get a superconducting (SC) susceptibility of the form

$$\chi_{SC,\mathbf{q}=\mathbf{0}\,\omega=0} = \sum_{kk'\nu\nu'} \underline{\underline{\chi^0}} \Big(1 + \underline{\underline{\Gamma_{pp,\mathbf{q}=\mathbf{0}\omega=0}}}\ \underline{\underline{\chi^0}}\Big)^{-1} \tag{15}$$

Here, we are interested in the instability at $\omega = 0$ and the coupling of two fermions with momentum $\mathbf{k}$ and $-\mathbf{k}$ into a Cooper pair as in Fig. 8(c), i.e., a total momentum $\mathbf{q} = 0$ in the pp channel[14]. The generalized[15] bare bubble susceptibility at $\omega = 0$ and $\mathbf{q} = 0$ is

$$\chi^0_{\mathbf{k}\mathbf{k}'\nu\nu'} = G_{\mathbf{k}'\nu'}\, G_{-\mathbf{k}-\nu}\, \delta(\mathbf{k}-\mathbf{k}')\, \delta_{\nu,\nu'} \tag{16}$$

and the double underlines denote matrices with respect to $\mathbf{k}\nu$ and $\mathbf{k}'\nu'$.
In order to obtain superconductivity, i.e., a diverging $\chi_{SC,\mathbf{q}=\mathbf{0}\,\omega=0}$ one of the eigenvalues λ of the matrix $-\underline{\underline{\Gamma_{pp,\mathbf{q}=\mathbf{0}\omega=0}}}\ \underline{\underline{\chi^0}}$ has to approach $\lambda = 1$. For electron-phonon mediated superconductivity this is simple since the electron-phonon coupling gives rise to an attractive (negative) Γ_{pp}. For the repulsive Hubbard model, this is much more difficult to achieve since Γ_{pp} also includes the repulsive (positive) Coulomb interaction $+U$. Nonetheless, an attraction can be mediated by antiferromagnetic spin fluctuations through retardation (ν, ν') and non-locality (k, k'). The k, k' structure of the diverging ($\lambda \to 1$) eigenvector (or the corresponding real space structure) determines the symmetry of the superconducting symmetry breaking and gap (s-wave, d-wave etc.). For phonons with an across the board attraction, one gets s-wave superconductivity. For the spin fluctuations in the Hubbard model d-wave is more favorable to avoid the large local repulsion $+U$.[16]
The Mermin-Wagner theorem also applies for superconductivity, where in two dimensions and at finite temperatures only the Nobel-prize-winning topological Berezinskii-Kosterlitz-Thouless

[14] This $\mathbf{q}$ is not to be confused with that in the ph channel which is peaked around $\mathbf{Q} = (\pi, \pi)$ and corresponds to another combination of the external legs of the four-point vertex F.

[15] without $kk'\nu\nu'$ summation

[16] For the d-wave, Γ_{pp} mediates between momenta $\mathbf{k}$ and $\mathbf{k}'$ on the Fermi surface, for which the superconducting eigenvalue has opposite sign.

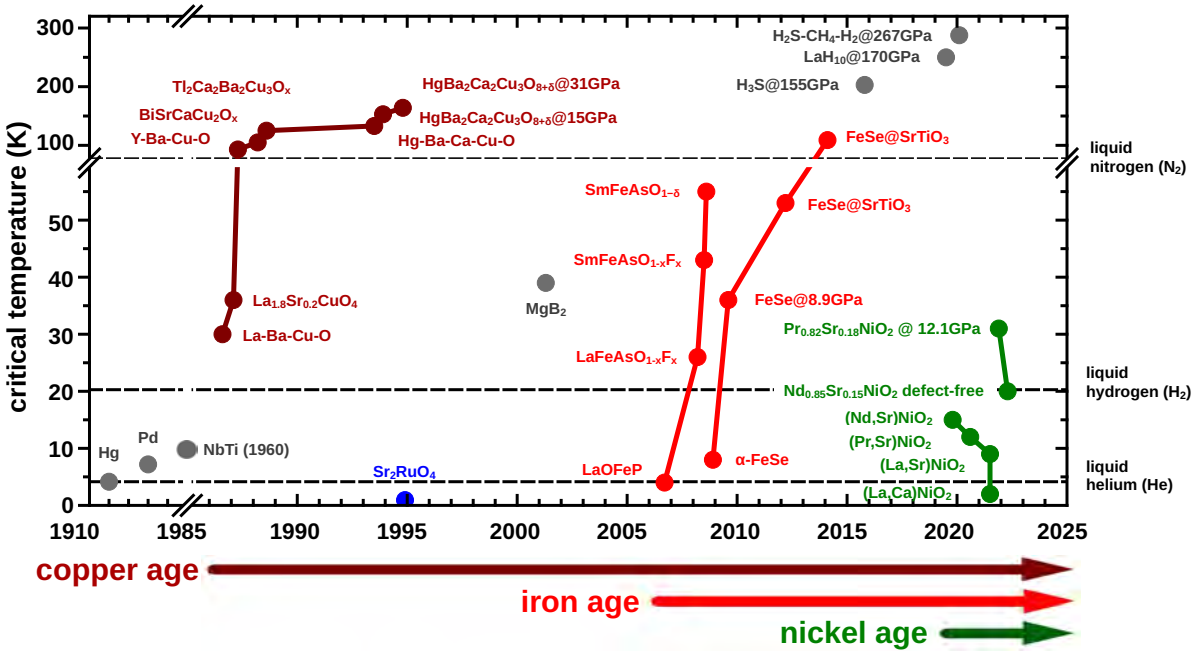


Fig. 20: *Superconducting T_c vs. year of discovery for various important superconductors. After the copper age, there has been an iron age, and now we are in the nickel age. From [31].*

(BKT) transition is possible. At first glance, this is a contradiction to the observed superconductivity in cuprates and nickelates. However, these materials are not perfectly two-dimensional but layered quasi-two-dimensional materials. In this situation, and with the strongly increasing correlation length around the BKT transition, even a tiny coupling in the inter-layer direction will trigger superconductivity, a little bit below the BKT transition. In the DΓA calculation, a self-consistency is necessary to (possibly) get a BKT transition. Without the self-consistent feedback of the superconducting fluctuations, we eventually get a mean-field kind of superconducting transition and a finite T_c which is closer to experiment than the ideal two-dimensional calculation with BKT transition.

As for superconducting materials, the three years ago discovered nickelate superconductors [30] have led to enormous theoretical and experimental efforts. One therefore also speaks of the nickel age for superconductivity, see Fig. 20. Also shown are hydrogen-based superconductors that are phonon-mediated and have a T_c above room temperature, but only under the enormous pressure of a diamond anvil cell. Hence the challenge is to increase T_c for the unconventional (i.e., not phonon-mediated) correlated superconductors or to reduce the pressure for the hydrogen-based superconductors. Here, the nickelates have a T_c that is still quite substantially below that of the cuprates. While one can expect T_c to further increase somewhat with new and better synthesized nickelate films, one should not expect a room temperature nickelate superconductor. The high hope is instead that nickelates and cuprates are very similar but also decisively distinct, an ideal situation to discriminate the essentials from the incidentals for high-temperature superconductivity. The iron pnictides are instead pretty far away from the cuprate or nickelate physics.

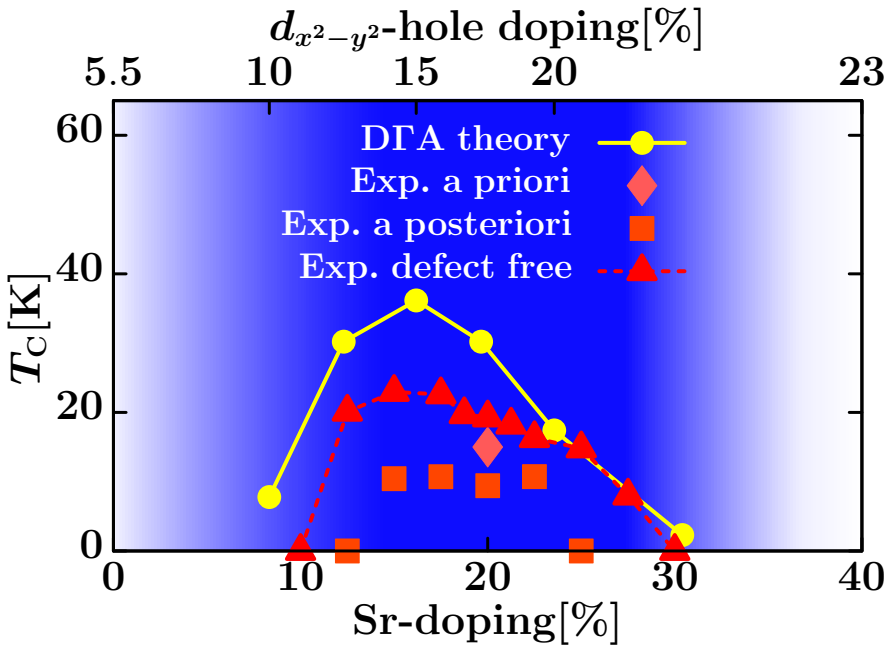


Fig. 21: *Superconducting phase diagram T_c vs. Sr-doping (lower x-axis) and vs. doping of the $3d_{x^2-y^2}$-orbital (upper x-axis) as calculated for $Nd_{1-x}Sr_xNiO_2$ in DΓA [27]. At the time of the calculation a single experimental data point ("a priori" [30]) was available. The first experimental superconducting dome ("a posteriori" [32]) showed already reasonable agreement. With cleaner films ("defect free" [33]) the agreement has become breathtaking, in particular if one considers how challenging it is to calculate T_c reliably. Also the pentalayer nickelate ($T_c \approx 13\,K$, 20% holes in the $3d_{x^2-y^2}$ orbital [34]) well matches the T_c of DΓA and the shown infinite layer experiment. Adapted from [27].*

In Sec. 4, we have already pointed out that the nickelates can be described by a one-band Hubbard model with an appropriately adjusted doping; and in Fig. 19 we have shown the thus calculated spectrum with a pseudogap for the (non-superconducting) parent compound $NdNiO_2$.

For this nickelate Hubbard model, we find [27] (not surprisingly) d-wave superconductivity in ladder DΓA. The superconducting T_c vs. doping is plotted in Fig. 21 as a function of Sr-doping. Actually the theoretical calculation was a prediction here since, when three years ago superconductivity in nickelates has been discovered [30], only a single T_c at 20% doping was available at first because of the difficulties to synthesize $Nd_{1-x}Sr_xNiO_2$ in the low oxidation state Ni^{+1}. With the recent progress to synthesize clean superconducting nickelate films [33], the theoretical predicted phase diagram has been spectacularly confirmed in experiment, see Fig. 21.

The physical reason that we see in Fig. 21 a superconducting dome is two-fold: The downturn at large doping is because antiferromagnetic spin-fluctuations get weaker further and further away from half-filling. The down-turn at small doping, towards half-filling, on the other hand is a consequence of the pseudogap which develops in this doping region, see Fig. 19. Thus, the electron propagators in Fig. 8(c) loose coherence in a larger part of the Fermi surface, and the superconducting susceptibility is suppressed despite considerable spin fluctuations.

Recently also a pentalayer nickel$_{12}$ has been synthesized [35], which has very similar hopping parameters [34]. Its $T_c = 14\,\mathrm{K}$ agrees with DΓA and with the T_c of the infinite layer nickelates shown in Fig. 21. Here, DMFT indicates that pentalayer nickelates have no pockets [34] and thus a doping of 0.2 holes per site in the $3d_{x^2-y^2}$-orbital. The absence of pockets in pentalayer nickelates further corroborates the picture of a decoupled reservoir that is not relevant for superconductivity as advocated in Fig. 9. Altogether the DFT+DMFT and DΓA calculations for nickelates and the experiments for the different nickelates provide for a consistent picture [34]. This gives us some hope that we might finally be able to actually calculate and predict superconducting T_c's, the arguably biggest challenge of solid state theory.

8 Conclusion and outlook

Diagrammatic extensions of DMFT are very appealing in several ways: They combine the good description of DMFT with the, to the best of our knowledge, most important non-local physics. In this Chapter we have focused on spin fluctuations and how they mediate superconductivity, but other fluctuations such as charge fluctuations, weak localization corrections, excitons –you name it– are treated on an equal footing. Also (quantum) criticality can be described, a topic that has been discussed in an earlier Jülich Autumn School [12]. Diagrammatic extensions of DMFT are also very appealing since they merry quantum field theoretical with its qualitative understanding and numerics which is unavoidable for a quantitative description of strongly correlated electrons systems.

Different variants of diagrammatic extensions of DMFT exist and, for the sake of brevity, we have concentrated here on the first (and widely employed) variant: the dynamical vertex approximation. All variants have in common that they calculate a local vertex and construct non-local correlations from this vertex diagrammatically. In regions of the phase diagram where non-local correlations are short range, results are similar as for cluster extensions of DMFT. However, the diagrammatic extensions also offer the opportunity to study long-range correlations, which is key for (quantum) criticality but also important in other situations, as well as to calculate materials with many orbitals [15].

There is still plenty of room for improvement: starting from (i) various self-consistencies, using (ii) a cluster instead of a single site as a starting point, (iii) compactifying the vertex with the intermediate representation (IR) in frequency space [21], the truncated unity in momentum space [20] and the single-boson exchange [22] in Feynman diagram space, and thus allowing for parquet DΓA calculations at lower temperatures. Another development has been (iv) the calculation of the underlying two-particle vertices directly for real frequencies, which is possible using the numerical renormalization group (NRG) method, see Chapter "The physics of quantum impurity models" by J. von Delft [3].

In this Chapter, we have concentrated on antiferromagnetic spin fluctuations, how they open a pseudogap and how they mediate superconductivity. The T_c predicted for nickelates agrees well with experiment – actually much better than what we dared to hope for. This gives us some confidence that we are on the right track to better model and understand superconductivity, that

we eventually have the tools to predict T_c for new materials. While many theoreticians in the many-body community consider antiferromagnetic spin fluctuations to be at the origin of high temperature superconductivity, the mechanism for unconventional superconductivity remains hotly debated. Maybe through a careful analysis and predictions we can now prove that this is indeed the microscopic mechanism for high-temperature superconductivity.
Diagrammatic extensions of DMFT such as the DΓA also offer the opportunity to study many other phenomena and to do materials calculations. Phenomena such as changes of the topology in strongly correlated are hitherto hardly understood, a Berezinskii-Kosterlitz-Thouless transition in two-dimensions could possibly be described, or Luttinger or spin Peierls physicists in one dimension. All of this leaves plenty of opportunities for the next generation of physicists.

Acknowledgment

First of all, I would like to thank my co-workers on the research presented here, Oleg Janson, Andrey Katanin, Anna Kauch, Motoharu Kitatani, Freidrich Krien, Jan M. Tomczak, Georg Rohringer, Thomas Schäfer, Liang Si, Alessandro Toschi, and Paul Worm. Without them, the success of diagrammatic extensions of DMFT would not have been possible. Further I would like to thank Motoharu Kitatani and Paul Worm for carefully reading the manuscript. Finally, I gratefully acknowledge financially support by the Austrian Science Fund (FWF) through project P32044 and the Research Unit QUAST of German Science Foundation (DFG) (DFG FOR5249; Austrian part financed through FWF project I5868).

References

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

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

[3] E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.):
Dynamical Mean-Field Theory of Correlated Electrons,
Modeling and Simulation, Vol. 12 (Forschungszentrum Jülich, 2022)
`http://www.cond-mat.de/events/correl22`

[4] A. Schiller and K. Ingersent, Phys. Rev. Lett. **75**, 113 (1995)

[5] G. Rohringer, H. Hafermann, A. Toschi, A.A. Katanin, A.E. Antipov, M.I. Katsnelson, A.I. Lichtenstein, A.N. Rubtsov, and K. Held, Rev. Mod. Phys. **90**, 025003 (2018)

[6] T. Maier, M. Jarrell, T. Pruschke, and M.H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005)

[7] K. Held, A. Katanin, and A. Toschi, Prog. Theor. Phys. (Suppl.) **176**, 117 (2008)

[8] A. Toschi, A.A. Katanin, and K. Held, Phys Rev. B **75**, 045118 (2007)

[9] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **77**, 033101 (2008)

[10] T.A. Maier, M.S. Jarrell, and D.J. Scalapino, Phys. Rev. Lett. **96**, 047005 (2006)

[11] T. Ribic, P. Gunacker, S. Iskakov, M. Wallerberger, G. Rohringer, A.N. Rubtsov, E. Gull, and K. Held, Phys. Rev. B **96**, 235127 (2017)

[12] K. Held: *Quantum criticality and superconductivity in diagrammatic extensions of DMFT*,
in E. Pavarini and E. Koch and D. Vollhardt and A.I. Lichtenstein (eds.):
DMFT: From Infinite Dimensions to Real Materials,
Reihe: Modeling and Simulations, vol. 8 (Forschungszentrum Jülich, 2018)
`http://www.cond-mat.de/events/correl18`

[13] T. Schäfer, F. Geles, D. Rost, G. Rohringer, E. Arrigoni, K. Held, N. Blümer, M. Aichhorn, and A. Toschi, Phys. Rev. B **91**, 125109 (2015)

[14] A. Kauch, P. Pudleiner, K. Astleithner, P. Thunström, T. Ribic, and K. Held,
Phys. Rev. Lett. **124**, 047401 (2020)

[15] A. Galler, P. Thunström, P. Gunacker, J.M. Tomczak, and K. Held,
Phys. Rev. B **95**, 115107 (2017)

[16] M. Kitatani, R. Arita, T. Schäfer, and K. Held, arXiv:2203.12844 (2022)

[17] N.E. Bickers: *Theoretical Methods for Strongly Correlated Electrons*
(Springer, 2004), chap. 6, pp. 237–296

[18] M. Wallerberger, A. Hausoel, P. Gunacker, A. Kowalski, N. Parragh, F. Goth, K. Held, and G. Sangiovanni, Comput. Phys. Commun. **235**, 388 (2019)

[19] P. Gunacker, M. Wallerberger, E. Gull, A. Hausoel, G. Sangiovanni, and K. Held, Phys. Rev. B **92**, 155102 (2015)

[20] C.J. Eckhardt, C. Honerkamp, K. Held, and A. Kauch, Phys. Rev. B **101**, 155104 (2020)

[21] M. Wallerberger, H. Shinaoka, and A. Kauch, Phys. Rev. Research **3**, 033168 (2021)

[22] F. Krien, A. Valli, and M. Capone, Phys. Rev. B **100**, 155149 (2019)

[23] J. Kaufmann, C. Eckhardt, M. Pickem, M. Kitatani, A. Kauch, and K. Held, Phys. Rev. B **103**, 035120 (2021)

[24] T. Ayral and O. Parcollet, Phys. Rev. B **94**, 075159 (2016)

[25] M. Kitatani, T. Schäfer, H. Aoki, and K. Held, Phys. Rev. B **99**, 041115 (2019)

[26] K. Held, L. Si, P. Worm, O. Janson, R. Arita, Z. Zhong, J.M. Tomczak, and M. Kitatani, Front. Phys. **9** (2022)

[27] M. Kitatani, L. Si, O. Janson, R. Arita, Z. Zhong, and K. Held, npj Quantum Mater. **5**, 59 (2020)

[28] T. Schäfer *et al.*, Phys. Rev. X **11**, 011058 (2021)

[29] F. Krien, P. Worm, P. Chalupa, A. Toschi, and K. Held, arXiv:2107.06529 (2021)

[30] D. Li, K. Lee, B.Y. Wang, M. Osada, S. Crossley, H.R. Lee, Y. Cui, Y. Hikita, and H.Y. Hwang, Nature **572**, 624 (2019)

[31] L. Si, P. Worm, and K. Held, Crystals **12** (2022)

[32] D. Li, B.Y. Wang, K. Lee, S.P. Harvey, M. Osada, B.H. Goodge, L.F. Kourkoutis, and H.Y. Hwang, Phys. Rev. Lett. **125**, 027001 (2020)

[33] K. Lee, B.Y. Wang, M. Osada, B.H. Goodge, T.C. Wang, Y. Lee, S. Harvey, W.J. Kim, Y. Yu, C. Murthy, S. Raghu, L.F. Kourkoutis, and H.Y. Hwang, arXiv:2203.02580 (2022)

[34] P. Worm, L. Si, M. Kitatani, R. Arita, J.M. Tomczak, and K. Held, arXiv:2111.12697 (2021)

[35] G.A. Pan *et al.*, Nat. Mater. **21**, 160 (2022)

12 The GW+EDMFT Method

Ferdi Aryasetiawan
Mathematical Physics, Lund University
Professorsgatan 1, 223 57 Lund, Sweden

Contents

1 Introduction **2**

2 GW method **3**
2.1 Screened interaction W . . . 3
2.2 Random-phase approximation . . . 5
2.3 GW self-energy . . . 6

3 GW+EDMFT method **7**
3.1 G^0W^0+DMFT . . . 7
3.2 Extended DMFT . . . 9
3.3 GW+EDMFT . . . 11
3.4 Multitier GW+EDMFT scheme . . . 12
3.5 Applications to cubic perovskites $SrMoO_3$ and $SrVO_3$ . . . 15

A Derivation of the GW approximation **18**
A.1 Equation of motion . . . 18
A.2 Self-energy . . . 19
A.3 Self-energy expansion in the screened interaction . . . 22
A.4 GW approximation . . . 23

B Functional derivation of GW+EDMFT **24**

1 Introduction

Dynamical mean-field theory (DMFT) [1] has been combined with density functional theory (DFT), commonly referred to as LDA+DMFT [2], which is regarded as a major breakthrough in the description of strongly correlated materials since it allows for a realistic description of the electronic structure of materials with strong dynamical correlations ranging from strongly correlated metals to Mott-Hubbard insulators. Starting from a DFT electronic structure as a background, a model Hamiltonian containing the strongly correlated manifold is constructed by making a tight-binding fit to the DFT bandstructure. This is usually done by using maximally localized Wannier functions [3]. The single-particle Hamiltonian is then complemented by a Hubbard-U term describing the onsite interaction between the strongly correlated states and a double-counting (DC) term is subtracted. This yields the DFT+DMFT Hamiltonian

$$\hat{H} = \sum_{\mathbf{R}i,\mathbf{R}'j,\sigma} \hat{c}^{\dagger}_{\mathbf{R}i\sigma}\big(H^{\rm DFT}_{\mathbf{R}i,\mathbf{R}'j,\sigma} - H^{\rm DC}_{\mathbf{R}i,\mathbf{R}'j,\sigma}\big)\hat{c}_{\mathbf{R}'j\sigma} + \frac{1}{2}\sum_{\mathbf{R},ijkl,\sigma\sigma'} \hat{c}^{\dagger}_{\mathbf{R}i\sigma}\hat{c}^{\dagger}_{\mathbf{R}j\sigma'}\, U^{\sigma\sigma'}_{ijkl}\, \hat{c}_{\mathbf{R}l\sigma'}\hat{c}_{\mathbf{R}k\sigma}, \qquad (1)$$

where $\mathbf{R}i$ label an atomic site $\mathbf{R}$ and an orbital i whereas σ labels the spin. The onsite effective Coulomb interaction U is defined as

$$U^{\sigma\sigma'}_{ijkl} = \int d^3r d^3r'\ \phi^{\sigma *}_i(\mathbf{r})\phi^{\sigma' *}_j(\mathbf{r}')\, U(\mathbf{r},\mathbf{r}')\, \phi^{\sigma}_k(\mathbf{r})\phi^{\sigma'}_l(\mathbf{r}'). \qquad (2)$$

The problem defined by the Hamiltonian (1) can then be solved by using the DMFT machinery. By mapping the lattice problem to an impurity problem, this method captures the essential physics of local correlations responsible for many phenomena in strongly correlated systems, not accessible within the one-particle description. Notable among these are the Mott metal-to-insulator transition and Kondo effects, which are beyond static theories. The combination of DMFT and DFT opens up wider possibilities of investigating electronic properties of strongly correlated materials than is possible within earlier static theories such as the DFT+U scheme. It provides a coherent description of metal and insulator on equal footing.

The basic assumption of local self-energy, however, hinders applications of this scheme to phenomena involving long-range correlations such as charge disproportionation and superconductivity, as well as phenomena involving coupling to long-range charge and spin fluctuations. Even in iron, a seemingly simple system, it has been found that its Curie transition temperature $T_{\rm C}$ from paramagnetic to ferromagnetic ordering is greatly overestimated when computed within DMFT. The predicted value of $T_{\rm C}$ of iron by DMFT is 1900 K, almost twice as large as the experimental value of 1043 K [4]. It is likely that this large discrepancy has its origin in the long-range part of the self-energy. Also, there is evidence from photoemission experiments [5] for a strong momentum dependence of the mass renormalization of iron. Another phenomenon for which a nonlocal self-energy is expected to play an important role is the non-Fermi-liquid behavior in two dimensions arising from the van Hove singularity [6].

This local self-energy assumption is one of the fundamental problems inherent in DFT+DMFT. The other is the issue of double counting, inherited from the DFT+U scheme. There is also the problem of determining the Hubbard U, which is often treated as adjustable, making the

method less predictive when applied to real materials. In recent years, however, it has become possible to compute the Hubbard U for real materials using a number of methods such as the constrained LDA (cLDA) [7] or the constrained random-phase approximation (cRPA) [8].
There are thus two main aspects of DFT+DMFT which should be improved: the assumption of local self-energy and the problem of double counting. The assumption of the self-energy being local is expected to be valid when the correlated electrons are highly localized such that the local Coulomb interaction far exceeds the nonlocal one. As the correlated electrons become more itinerant, one anticipates that the nonlocal interaction becomes more significant and accordingly the importance of the nonlocal components of the self-energy increases. The second aspect, the double-counting problem, is a long-standing one, which was already recognized in the earlier LDA+U method. The problem arises from the nonlinear dependence on the density of the Kohn-Sham exchange-correlation energy, which renders it difficult, if not impossible, to remove the contribution from the density of the correlated electrons only.
One approach to address the above two issues is the combination of the GW method and extended DMFT (GW+EDMFT) [9, 10], the former replaces the Kohn-Sham DFT. Being a Green function method, it is possible to remove the double-counting correction precisely. Moreover, the GW method was originally conceived to treat extended systems such as the electron gas so that it naturally incorporates nonlocal or long-range self-energies. Seen from the point of view of GW, the GW+EDMFT method can be regarded as a route of going beyond RPA by including vertex corrections in the form of local correlations.

2 GW method

In this section, the GW approximation (GWA) [11–13], which may be regarded as the background electronic structure in the GW+EDMFT method, is elaborated. The screened interaction, which is an essential part of the GW+EDMFT method, is discussed first in detail. The detailed derivations of the self-energy and response functions are given in Appendix A. In our notation, r is a combined variable for space and spin: $r = (\mathbf{r}, \sigma)$, and an integer represents space, spin, and time: $n = (r_n, t_n)$. Atomic units are used throughout where

$$m = \hbar = e = a_0 = 1. \tag{3}$$

2.1 Screened interaction W

The screened interaction is an essential part of GWA as well as in the GW+EDMFT scheme. Due to the Coulomb repulsion and the exchange interaction, electrons tend to avoid each other, resulting in the formation of a screening hole around each electron. This leads to an effective screened interaction between an electron and the other electrons in the system, which is much smaller than the bare interaction, especially in metals. It is the density response functions that determine the screened interaction.
Within GWA, only the linear density response function is needed. Higher-order density response functions correspond to vertex corrections. The linear response theory can be estab-

lished by applying an external time-dependent field $V_{\rm ext}$ to a system of electrons. The induced density to linear order in the applied field is given by

$$\rho_{\rm ind}(1) = \int d2\ R(1,2)\, V_{\rm ext}(2)\,, \tag{4}$$

where R is the linear density response function. This induced density in turns generates an induced back potential, with v the bare interaction,

$$V_{\rm ind}(1) = \int d2\ v(1{-}2)\, \rho_{\rm ind}(2)\,, \tag{5}$$

which screens the applied potential

$$V_{\rm tot}(1) = V_{\rm ext}(1) + V_{\rm ind}(1)\,. \tag{6}$$

The screened or total potential in terms of the linear response function is then given by

$$\begin{aligned} V_{\rm tot}(1) &= V_{\rm ext}(1) + \int d2\ v(1{-}2)\, \rho_{\rm ind}(2) \\ &= V_{\rm ext}(1) + \int d2\, d3\ v(1{-}2)\, R(2,3)\, V_{\rm ext}(3) \\ &= \int d3 \Big[\delta(1{-}3) + \int d2\, v(1{-}2)\, R(2,3)\Big] V_{\rm ext}(3)\,. \end{aligned} \tag{7}$$

The quantity in the square bracket can be identified as the inverse dielectric matrix

$$\varepsilon^{-1}(1,3) = \frac{\delta V_{\rm tot}(1)}{\delta V_{\rm ext}(3)} = \delta(1{-}3) + \int d2\, v(1{-}2)\, R(2,3)\,. \tag{8}$$

Applying the above formula to the instantaneous Coulomb potential

$$v(1{-}2) = v(r_1{-}r_2)\, \delta(t_1{-}t_2) \tag{9}$$

as the applied perturbing field yields the screened Coulomb interaction

$$W(1,2) = \int d3\ \varepsilon^{-1}(1,3)\, v(3{-}2) = v(1{-}2) + \int d3\, d4\ v(1{-}4)\, R(4,3)\, v(3{-}2)\,, \tag{10}$$

or in matrix form

$$W = \varepsilon^{-1} v = v + vRv\,. \tag{11}$$

Thus, $W(1,2)$ may be interpreted as a screened potential at space-time point 1 of a test charge located at space-time point 2. Fourier transforming Eq. (10) with respect to $t = t_1{-}t_2$ yields

$$W(r_1, r_2; \omega) = v(r_1{-}r_2) + \int dr_3 dr_4\ v(r_1{-}r_4) R(r_4, r_3; \omega) v(r_3{-}r_2)\,. \tag{12}$$

There are two limiting cases of interest. The static screened interaction $W(r_1, r_2; \omega{=}0)$ can be seen as the time-average of $W(r_1, r_2; t)$, as follows from the definition of the Fourier transform. Equivalently, from Eq. (12), it is the screened interaction of the static bare interaction $v(r_1{-}r_2)$

(without $\delta(t_1-t_2)$). At the other end of the limit, $W(r_1, r_2; \omega\to\infty)$ approaches the bare interaction $v(r_1-r_2)$ since $R(\omega)$ decays asymptotically as $\sim 1/\omega^2$ as can be seen from Eq. (90). This simply means that at very high frequency the electrons cannot react sufficiently fast to the rapid variation of the applied potential so that screening is ineffective.
The screened interaction via the linear density response encodes the charge excitation spectrum of the system, as can be seen in Eq. (12). Thus, when the applied frequency matches a collective charge excitation (plasmon) energy of the system, a resonance peak in the imaginary part of W is expected to appear.
The correlation part of the screened interaction is the second term on the right-hand side of (12),

$$W^c(\omega) = v\, R(\omega)\, v\,. \tag{13}$$

Since the Coulomb interaction v is frequency independent, the spectral representation of W^c is identical in form to that of $R(\omega)$ as in Eq. (91) of Appendix A

$$W^c(r_1, r_2; \omega) = \int_{-\infty}^{0} d\omega'\, \frac{B(r_1, r_2; \omega')}{\omega - \omega' - i\eta} + \int_{0}^{\infty} d\omega'\, \frac{B(r_1, r_2; \omega')}{\omega - \omega' + i\eta}\,, \tag{14}$$

where

$$B(\omega) = v\, S(\omega)\, v = -\frac{1}{\pi}\mathrm{sign}(\omega)\,\mathrm{Im}\, W(\omega)\,, \tag{15}$$

with S given in Eq. (92). Since $S(\omega)$ is anti-symmetric, so is $B(\omega)$:

$$B(-\omega) = -B(\omega)\,. \tag{16}$$

2.2 Random-phase approximation

It is unrealistic to calculate the exact response functions for real systems since it would require computation of the exact excited states of the many-electron system. This necessitates approximations, as commonly the case in many-electron theory. A practical and highly successful approximation is the random-phase approximation (RPA) [14]. As discussed later, it can be understood as the time-dependent Hartree approximation since it takes into account only the change in the Hartree potential with respect to the time-dependent perturbing field.
In the Green function formalism, the response function is expressed as in Eq. (88), and using the identity $\delta G = -G\,\delta G^{-1}\,G$ as in Eq. (84), we find

$$R(1,2) = i\int d3\, d4\; G(1,3)\, \frac{\delta G^{-1}(3,4)}{\delta V_{\mathrm{ext}}(2)} G(4,1^+)\,. \tag{17}$$

Using $\delta G^{-1}/\delta V_{\mathrm{ext}}$ as in Eq. (85) we obtain

$$R(1,2) = -i\int d3\, d4\; G(1,3)\left(\left[\delta(3{-}2) + \frac{\delta V_{\mathrm{H}}(3)}{\delta V_{\mathrm{ext}}(2)}\right]\delta(3{-}4) + \frac{\delta\Sigma(3,4)}{\delta V_{\mathrm{ext}}(2)}\right) G(4,1^+)\,. \tag{18}$$

RPA corresponds to neglecting the term $\delta\Sigma/\delta V_{\mathrm{ext}}$ in (18) yielding

$$\begin{aligned} R(1,2) &= -i\int d3\; G(1,3)\left[\delta(3{-}2) + \frac{\delta V_{\mathrm{H}}(3)}{\delta V_{\mathrm{ext}}(2)}\right] G(3,1^+) \\ &= P(2,1) + \int d3\, d4\; P(1,3)\, v(3{-}4)\, R(4,2)\,. \end{aligned} \tag{19}$$

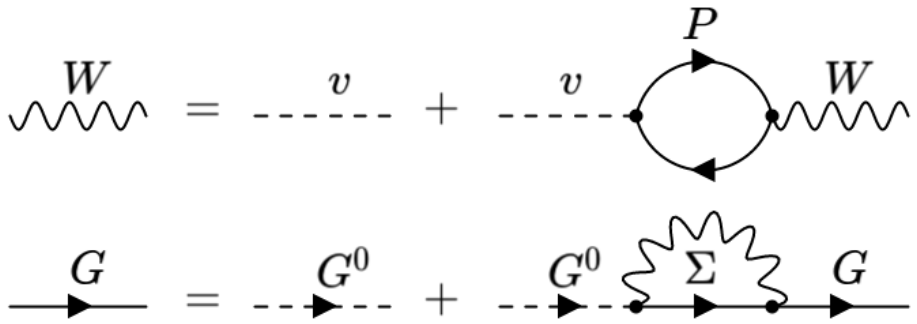


Fig. 1: *Feynman diagrams for the screened interaction in the random-phase approximation and the self-energy within the GW approximation. Figures courtesy of Fredrik Nilsson.*

where

$$P(1,2) = -iG(1,2)\,G(2,1^+). \tag{20}$$

Thus, in RPA only the change in the Hartree potential is taken into account and for this reason it can be regarded as the time-dependent Hartree approximation.
In frequency space, the polarization P in RPA is given by

$$P(r_1,r_2;\omega) = -i\int \frac{d\omega'}{2\pi}\, G(r_1,r_2;\omega{+}\omega')\,G(r_2,r_1;\omega')\,. \tag{21}$$

Replacing G by G^0 given in Eq. (109) yields

$$P^0(r_1,r_2;\omega) = \sum_{n}^{\rm occ}\sum_{n'}^{\rm unocc}\left(\frac{\phi_n(r_2)\phi^*_{n'}(r_2)\phi^*_n(r_1)\phi_{n'}(r_1)}{\omega-(\varepsilon_{n'}{-}\varepsilon_n{-}i\delta)} - \frac{\phi_n(r_1)\phi^*_{n'}(r_1)\phi^*_n(r_2)\phi_{n'}(r_2)}{\omega+(\varepsilon_{n'}{-}\varepsilon_n{-}i\delta)}\right). \tag{22}$$

2.3 GW self-energy

Within the GW approximation [11–13], the self-energy is given by

$$\Sigma(1,2) = iG(1,2)\,W(2,1)\,, \tag{23}$$

where W is the screened Coulomb interaction, usually calculated within RPA. Using the convolution theorem, the Fourier transform into frequency space yields

$$\Sigma(r_1,r_2;\omega) = i\int \frac{d\omega'}{2\pi}\, G(r_1,r_2;\omega{+}\omega')\,W(r_2,r_1;\omega')\,{\rm e}^{i\eta\omega'}. \tag{24}$$

The screened interaction and the corresponding GW self-energy within RPA are shown in Fig. 1.
GWA has been found to be very successful in providing an accurate electronic structure for weakly to moderately correlated systems, whose valence electrons usually originate from s- or p-orbitals. Being based on perturbation theory, GWA is not expected to be able to describe the Mott metal-to-insulator transition in strongly correlated systems. To see the problem with the

GWA consider the expression for the imaginary part of the Green function, which is proportional to the spectral function:

$$\mathrm{Im}\, G(\omega) = \frac{\mathrm{Im}\, \Sigma(\omega)}{\big(\omega - \varepsilon - \mathrm{Re}\, \Sigma(\omega)\big)^2 + \big(\mathrm{Im}\, \Sigma(\omega)\big)^2}\,. \tag{25}$$

In order to open up a gap from a metallic state, $\mathrm{Im}\, \Sigma(\omega)$, which is inversely proportional to the quasiparticle life-time, should become infinite around the Fermi level so that the quasiparticle weight present in the metal at the Fermi level is removed. The transition from a metal to a Mott insulator poses a fundamental difficulty because the GW self-energy varies as ω^2 around the Fermi level, characteristic of a Fermi liquid, as shown below. From Eq. (14) and using (16)

$$\begin{aligned}\frac{\partial}{\partial\omega} W^{\mathrm{c}}(r_1, r_2; \omega)\Big|_{\omega=0} &= -\int_{-\infty}^{0} d\omega' \, \frac{B(r_1, r_2; \omega')}{(-\omega' - i\eta)^2} - \int_{0}^{\infty} d\omega' \, \frac{B(r_1, r_2; \omega')}{(-\omega' + i\eta)^2} \\ &= +\int_{0}^{\infty} d\omega' \, \frac{B(r_1, r_2; \omega')}{(\omega' - i\eta)^2} - \int_{0}^{\infty} d\omega' \, \frac{B(r_1, r_2; \omega')}{(-\omega' + i\eta)^2} \;\; = \;\; 0, \end{aligned} \tag{26}$$

Since $B(\omega)$ is proportional to $\mathrm{Im}\, W(\omega)$ and recognizing that $B(0) = 0$, it follows from Eqs. (113) and (114) that for small ω the imaginary part of the self-energy behaves as

$$\Gamma(\omega) \propto \omega^2. \tag{27}$$

This implies that the GW self-energy cannot not open up a Mott gap starting from a metallic band structure.

3 GW+EDMFT method

The problem with GWA in treating strongly correlated systems such as Mott insulators motivates the combination of GWA and extended DMFT (EDMFT). On the other hand, from the point of view of DMFT, GWA provides a route for taking into account nonlocal self-energy, missing in DMFT or EDMFT.

3.1 G^0W^0+DMFT

It is instructive to first consider the combination of the one-shot GWA, $\Sigma = iG^0W^0$, with the standard DMFT. This combination may be seen as a simple extension of the DFT+DMFT scheme, in which the background DFT electronic structure is replaced by the GW one. Apart from improving the band structure, the double-counting problem inherent in the DFT+DMFT scheme is avoided. However, there are different ways of removing the double-counting term when self-consistency is not imposed. Either one removes the local projection of the lattice GW self-energy or one removes the impurity GW self-energy. These two double-counting forms are not strictly the same.

In the G^0W^0+DMFT scheme the self-energy is given by

$$\hat{\Sigma}(\omega) = \sum_{\mathbf{k}nn'} |\psi_{\mathbf{k}n}\rangle \, \Sigma^{GW}_{nn'}(\mathbf{k}, \omega) \, \langle\psi_{\mathbf{k}n'}| + \sum_{mm'} |\phi_m\rangle \, \big(\Sigma^{\mathrm{imp}}_{mm'}(\omega) - \Sigma^{\mathrm{DC}}_{mm'}(\omega)\big) \langle\phi_{m'}|\,. \tag{28}$$

Here $\psi_{\mathbf{k}n}$ are the DFT eigenstates and ϕ_m the localized orbitals, e.g., the maximally localized Wannier orbitals constructed from the correlated states. $\Sigma^{GW}_{nn'}$ is the GW self-energy obtained from a one-shot GW calculation based on the DFT band structure and $\Sigma^{\rm imp}$ the impurity self-energy calculated with a dynamical U computed using the constrained RPA (cRPA) method. $\Sigma^{\rm DC}_{mm'}$ is the double-counting correction, defined as

$$\Sigma^{\rm DC}_{mm'}(\omega) = i \sum_{m_1 m_2} \int \frac{d\omega'}{2\pi}\, G^{\rm loc}_{m_1 m_2}(\omega+\omega')\, W^{\rm loc}_{m m_1 m_2 m'}(\omega')\,, \tag{29}$$

where $G^{\rm loc}_{m_1 m_2}$ is the local projection of the DFT Green function expressed in the localized orbitals. In Matsubara frequencies,

$$\Sigma^{\rm DC}_{mm'}(i\omega_n) = -\sum_{m_1 m_2} \frac{1}{\beta} \sum_k G^{\rm loc}_{m_1 m_2}(i\omega_n+i\nu_k)\, W^{\rm loc}_{m m_1 m_2 m'}(i\nu_k)\,, \tag{30}$$

where ω_n and ν_k are, respectively, the fermionic and bosonic Matsubara frequencies, and β is the inverse temperature. The two different ways of subtracting the double-counting term arise from the ambiguity in $W^{\rm loc}$ when it is not determined self-consistently. If $W^{\rm loc}$ is calculated from the impurity problem one has

$$W^{\rm loc}(\omega) = \big(1 - U(\omega)P^{\rm loc}(\omega)\big)^{-1} U(\omega), \tag{31}$$

in which $P^{\rm loc}$ is the local projection of the RPA polarization within the correlated subspace and $U(\omega)$ is obtained from cRPA. This definition of $W^{\rm loc}$ corresponds to the fully screened interaction of the impurity model computed within GWA. On the other hand, $W^{\rm loc}$ can also be defined to be the local projection of the lattice W onto the correlated subspace

$$W^{\rm loc}(\omega) = \sum_{\mathbf{k}} W(\omega, \mathbf{k})\,, \tag{32}$$

which is not necessarily the same as the one in Eq. (29), when self-consistency is not imposed. In Fig. 2 the quasiparticle bandstructure and the k-integrated spectra of a prototypical correlated metal $SrVO_3$ within the G^0W^0+DMFT scheme are shown and compared with the results obtained from the one-shot G^0W^0 and LDA+DMFT. The one-shot G^0W^0 quasiparticle dispersion shows a band narrowing compared with that of LDA, as expected. The G^0W^0 occupied band width of 0.8 eV, while a significant improvement over the LDA band width, is still too large compared with the value measured in photoemission experiment, which is about 0.6 eV [16]. In contrast, the LDA+DMFT calculations, which have been performed using a frequency-dependent U, result in an occupied band width of only 0.4 eV, which is significantly smaller than the experimental value as can be seen in Fig. 2. There is a very large band narrowing compared with the starting LDA bandstructure. This band narrowing is significantly stronger than in the standard LDA+DMFT with a static U [17]. One may then conclude that the frequency dependence of U effectively enhances correlations, which would correspond to a larger static U. The G^0W^0+DMFT scheme, on the other hand, brings the quasiparticle dispersion in close agreement with experiment with an occupied band width of slightly smaller than 0.6 eV. It is interesting to observe that a nonlocal self-energy tends to increase the band width while a local self-energy narrows the band width.

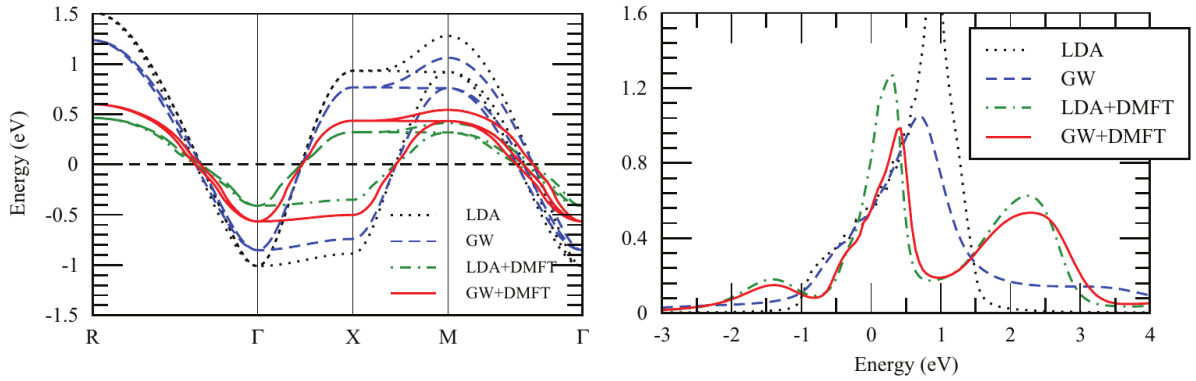


Fig. 2: *Quasiparticle bandstructure (left) and k-integrated spectral function (right) of* $SrVO_3$ *within the* G^0W^0+DMFT *scheme, compared with the results based on* G^0W^0 *and* LDA + DMFT *with dynamic U. The figures are taken from Sakuma et al. [15].*

3.2 Extended DMFT

In order to develop the full GW+EDMFT scheme it is necessary to extend DMFT in order to take into account long-range interactions. The G^0W^0+DMFT scheme described in the previous section does not take into account self-consistency in the local screened interaction. As in the original DMFT, self-consistency is only imposed on the local Green function, which may be regarded as a one-particle correlation function. The concept of DMFT, however, is more general and can be extended to higher-order correlation functions. A natural step beyond the original DMFT is to impose self-consistency on the local screened interaction $W_{\rm loc}$ or equivalently the local (charge) response function, which is a two-particle correlation function. This extended DMFT (EDMFT) [9] furnishes a means to solve an extended Hubbard model with nonlocal interactions using the framework of DMFT. As in the original DMFT, the auxiliary system is the Anderson impurity model but in addition to the hybridization function (Weiss field) $\Delta(\tau)$, which determines the local Green function, another "Weiss field", $\mathcal{U}(\tau)$, is needed to determine the local screened interaction $W_{\rm loc}$. In this case $\mathcal{U}(\tau)$ plays the role of a dynamic effective impurity interaction in such a way that when it is screened by the impurity polarization reproduces at self-consistency the local projection of the lattice screened interaction. Thus, $\mathcal{U}(\tau)$ is analogous to $\mathcal{G}(\tau)$, which when "screened" by the impurity self-energy reproduces at self-consistency the local projection of the lattice Green function. It can then be seen that the impurity polarization is analogous to the impurity self-energy.

The effective impurity action is given by

$$S_{\rm imp} = -\int_0^\beta d\tau d\tau' \sum_\sigma d^*_\sigma(\tau)\, \mathcal{G}^{-1}(\tau-\tau')\, d_\sigma(\tau') + \int_0^\beta d\tau d\tau'\, n_\uparrow(\tau)\, \mathcal{U}(\tau-\tau')\, n_\downarrow(\tau')\,, \qquad (33)$$

where

$$-\left(\frac{\partial}{\partial\tau} - \mu\right) \mathcal{G}(\tau-\tau') - \int_0^\beta d\tau''\, \Delta(\tau-\tau'')\, \mathcal{G}(\tau''-\tau') = \delta(\tau-\tau'), \qquad (34)$$

or in frequency space

$$\mathcal{G}^{-1}(i\omega_n) = i\omega_n + \mu - \Delta(i\omega_n). \tag{35}$$

The self-consistency condition in EDMFT requires self-consistency in both the local Green function and the local screened interaction:

$$G_{\rm loc} = G_{\rm imp}, \quad W_{\rm loc} = W_{\rm imp}, \tag{36}$$

where

$$W_{\rm imp}(\omega) = \mathcal{U}(\omega) + \mathcal{U}(\omega) P_{\rm imp}(\omega) W_{\rm imp}(\omega), \tag{37}$$

and $W_{\rm loc}$ is the local projection of the lattice screened interaction W obtained from the equation

$$W(\omega) = U + UP(\omega)W(\omega), \tag{38}$$

where P is the lattice polarization. U is the bare interaction which may contain long-range (offsite) components. Comparison with the Dyson equation for the Green function,

$$G(\omega) = G^0(\omega) + G^0(\omega)\Sigma(\omega)G(\omega), \tag{39}$$

shows that the polarization P is the analog of the self-energy whereas W is the analog of the Green function, whereas U plays the role of the bare propagator G^0. For the Hubbard model with onsite interactions we have $U_{ij} = U\delta_{ij}$ which yields $\mathcal{U} = U$ by construction and in this case EDMFT reduces to the regular DMFT. For a system with offsite interactions this is not the case. The effects of the long-range interactions are taken into account in the dynamically screened effective impurity interaction $\mathcal{U}(\omega)$. In real materials U will be the bare Coulomb interaction $v(\mathbf{r}-\mathbf{r}')$.

3.2.1 Self-consistency loop in EDMFT

To initiate the self-consistency loop in EDMFT, the local polarization and the local self-energy are, respectively, set to $P_{\rm imp}$ and $\Sigma_{\rm imp}$

$$P_{\rm loc} = P_{\rm imp}, \quad \Sigma_{\rm loc} = \Sigma_{\rm imp}. \tag{40}$$

The basic assumption of EDMFT, in which the lattice self-energy and the lattice polarization are approximated by their local values, is applied

$$P_{\mathbf{k}} = P_{\rm loc}, \quad \Sigma_{\mathbf{k}} = \Sigma_{\rm loc}. \tag{41}$$

From the Dyson equation the lattice Green function $G_{\mathbf{k}}$ and the lattice screened interaction $W_{\mathbf{k}}$ are computed

$$G_{\mathbf{k}} = G^0_{\mathbf{k}} + G^0_{\mathbf{k}}\, \Sigma_{\mathbf{k}}\, G_{\mathbf{k}}\,, \tag{42}$$

$$W_{\mathbf{k}} = U_{\mathbf{k}} + U_{\mathbf{k}}\, P_{\mathbf{k}}\, W_{\mathbf{k}}\,, \tag{43}$$

from which the local Green function and the local screened interaction are obtained

$$G_{\rm loc} = \sum_{\mathbf{k}} G_{\mathbf{k}}, \quad W_{\rm loc} = \sum_{\mathbf{k}} W_{\mathbf{k}}. \tag{44}$$

Applying the EDMFT self-consistency condition,

$$G_{\rm imp} = G_{\rm loc}, \quad W_{\rm imp} = W_{\rm loc}, \tag{45}$$

allows us to construct the fermionic Weiss field $\mathcal{G}$ and the effective impurity interaction (bosonic Weiss field) $\mathcal{U}$ from the Dyson equations

$$G_{\rm imp} = \mathcal{G} + \mathcal{G}\,\Sigma_{\rm imp}\,G_{\rm imp} \quad \rightarrow \quad \mathcal{G} = G_{\rm imp}\big(1 + \Sigma_{\rm imp}\,G_{\rm imp}\big)^{-1}, \tag{46}$$

$$W_{\rm imp} = \mathcal{U} + \mathcal{U}\,P_{\rm imp}\,W_{\rm imp} \quad \rightarrow \quad \mathcal{U} = W_{\rm imp}\big(1 + P_{\rm imp}\,W_{\rm imp}\big)^{-1}. \tag{47}$$

$\mathcal{G}$ and $\mathcal{U}$ provide an input for solving the impurity problem with the action defined in Eq. (33), yielding a new $G_{\rm imp}$ and the impurity charge susceptibility $\chi_{\rm imp}$. A new impurity self-energy and a new impurity polarization are extracted as

$$\Sigma_{\rm imp} = \mathcal{G}^{-1} - G_{\rm imp}^{-1}, \tag{48}$$

$$P_{\rm imp} = \chi_{\rm imp}\big(1 + \mathcal{U}\,\chi_{\rm imp}\big)^{-1}. \tag{49}$$

The last equation is formally equivalent to

$$P_{\rm imp} = \mathcal{U}^{-1} - W_{\rm imp}^{-1} \tag{50}$$

but it is numerically preferable since it avoids possible singularities in $\mathcal{U}^{-1}$ and $W_{\rm imp}^{-1}$.
The loop is now continued until the EDMFT self-consistency condition in Eq. (36) is fulfilled.
The concept of DMFT is quite general and it is possible to extend the self-consistency condition on local quantities to a three-particle correlation function, such as the one formulated in the TRILEX approach [18], and to higher-order correlation functions.

3.3 *GW*+EDMFT

We are now in the position to construct the full GW+EDMFT scheme [9, 10]. The formal functional derivation is given in Appendix B. As in EDMFT, two propagators are now at disposal: the fermionic one-particle Green function G and the bosonic screened interaction W. These two propagators constitute the basic variables in the free-energy functional Ψ (see Appendix B). The lattice self-energy Σ and polarization P are no longer approximated by the corresponding local impurity quantities but they now in addition contain nonlocal components from the GW self-energy. The self-energy is then given by

$$\Sigma(\mathbf{k}, i\omega_n) = \Sigma_{\rm imp}(i\omega_n) + \Sigma^{GW}_{\rm nonloc}(\mathbf{k}, i\omega_n). \tag{51}$$

Similarly, the polarization is given by

$$P(\mathbf{k}, i\omega_n) = P_{\rm imp}(i\omega_n) + P^{GW}_{\rm nonloc}(\mathbf{k}, i\omega_n). \tag{52}$$

$\Sigma_{\rm imp}$ and $P_{\rm imp}$ are, respectively, the EDMFT impurity self-energy and polarization. The above two expressions are free from double counting since both $\Sigma^{GW}_{\rm nonloc}$ and $P^{GW}_{\rm nonloc}$ contain no local components, which have been subtracted out from the corresponding lattice quantities

$$\Sigma^{GW}_{\rm nonloc}(\mathbf{k}, i\omega_n) = \Sigma^{GW}(\mathbf{k}, i\omega_n) - \Sigma^{GW}_{\rm loc}(i\omega_n) \tag{53}$$

$$P^{GW}_{\rm nonloc}(\mathbf{k}, i\omega_n) = P^{GW}(\mathbf{k}, i\omega_n) - P^{GW}_{\rm loc}(i\omega_n)\,. \tag{54}$$

The lattice Green function and screened interaction are calculated from Σ and P using the Dyson equations

$$G = G^0 + G^0\Sigma G, \tag{55}$$

$$W = v + vPW\,, \tag{56}$$

where G^0 is the Hartree Green function and v is the bare interaction.

3.4 Multitier *GW*+EDMFT scheme

To apply the full GW+EDMFT scheme described in the previous sections to real materials, due to computational complexity it is inevitable to adopt approximations. Fortunately, in most correlated materials, for which the GW+EDMFT scheme is designed, it is usually the case that the electronic structure can be separated into correlated and noncorrelated bands. Although the separation is arbitrary, physical consideration imposes a strong constraint on the choice of correlated bands, which are typically partially filled narrow bands across or around the chemical potential originating from localized states of $3d$- or $4f$-orbitals.

In the multitier GW+EDMFT scheme [19, 20], the Hilbert space is decomposed into a correlated subspace and the rest. A concrete example is provided by the prototypical correlated metal $SrVO_3$ in which the narrow partially filled t_{2g} band is naturally chosen as the correlated subspace. GW+EDMFT self-consistency is performed only in the correlated subspace whereas contribution to the self-energy from the rest of the Hilbert space is accounted for by a one-shot GW. There is a physical motivation for this strategy. It is known that fully self-consistent GW alone tends to worsen the one-shot results (in comparison with experiment). Within the correlated subspace, however, the detrimental effects of self-consistency on the GW self-energy are expected to be counteracted by vertex corrections in the form of the EDMFT self-energy.

It is possible to introduce an intermediate subspace $\mathcal{I}$, containing an additional few bands around the Fermi level considered to be important for the low-energy physics. The intermediate subspace then includes a set of correlated bands crossing the Fermi energy and, in addition, a few more weakly correlated bands which hybridize with the correlated bands. The correlated subspace $\mathcal{C}$ is spanned by the narrow bands that cannot be accurately described by GWA. In this case, the GW+EDMFT self-consistency is performed within the intermediate subspace $\mathcal{I}$. In Fig. 3 the different subspaces and tiers are illustrated. Together with the three subspaces there are three different tiers that work on the respective subspace. In tier III the full space is treated within the one-shot GWA (G^0W^0-approximation).

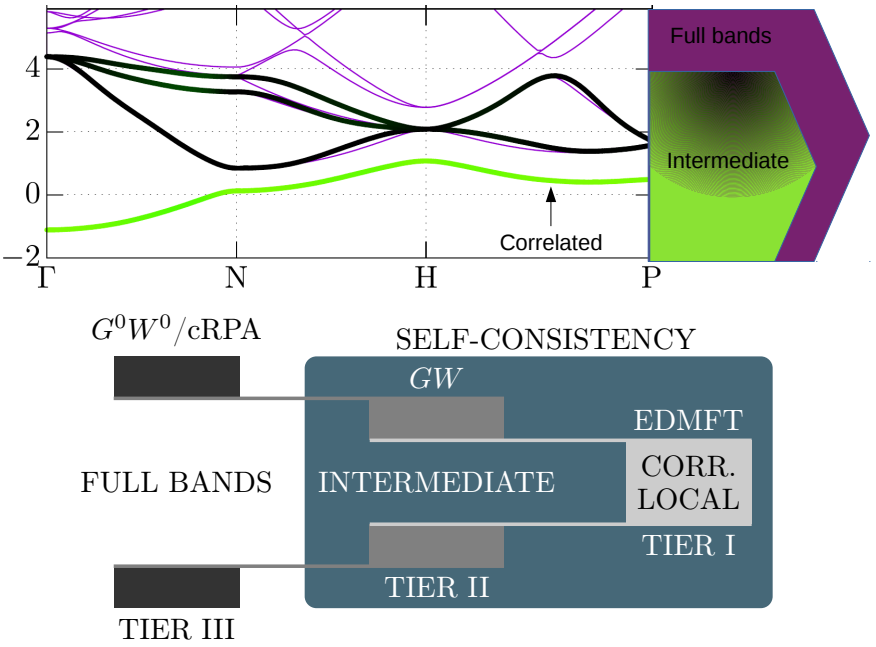


Fig. 3: *Upper panel: Schematic figure of the subspaces in multitier GW+EDMFT, reproduced from Nilsson et al. [19]. Lower panel: Schematic figure of how the tiers work on the different subspaces, reproduced from Boehnke et al. [20].*

In practice, one starts with a DFT bandstructure and performs a one-shot GW calculation on the entire Hilbert space (tier III) yielding a self-energy $\Sigma^{G^0W^0}$. A calculation of the bare effective interaction, $U(\omega)$, within the intermediate subspace is also performed using the constrained random-phase approximation (cRPA) method. The effective bare propagator ($G^0_{\mathbf{k}}$) and interaction ($U_{\mathbf{q}}$) on the intermediate subspace are then computed as

$$G^{0\,-1}_{\mathbf{k}} = \underbrace{\mathrm{i}\omega_n + \mu - \varepsilon^{\mathrm{KS}}_{\mathbf{k}}}_{G^{-1}_{\mathrm{KS},\mathbf{k}}} + V_{\mathrm{XC},\mathbf{k}} - \underbrace{\left(\Sigma^{G^0W^0}_{\mathbf{k}} - \Sigma^{G^0W^0}_{\mathbf{k}}\big|_I\right)}_{\Sigma^{\mathrm{r}}_{\mathbf{k}}\big|_I} \tag{57}$$

$$U^{-1}_{\mathbf{q}} = v^{-1}_{\mathbf{q}} - \underbrace{\left(P^{G^0G^0}_{\mathbf{q}} - P^{G^0G^0}_{\mathbf{q}}\big|_I\right)}_{P^r_{\mathbf{q}}}. \tag{58}$$

Here $\varepsilon^{\mathrm{KS}}_{\mathbf{k}}$ are the Kohn-Sham (KS) DFT eigenenergies, $V_{\mathrm{XC},\mathbf{k}}$ the exchange-correlation potential, $\Sigma^{G^0W^0}_{\mathbf{k}}$ the one-shot GW self-energy, $v_{\mathbf{q}}$ the bare Coulomb interaction and $P^{G^0G^0}_{\mathbf{q}}$ the RPA polarization computed from the Kohn-Sham bandstructure. The notation $A\big|_I$ means that all internal sums when computing the quantity A are restricted to the intermediate subspace. U is the effective Coulomb interaction on the intermediate subspace computed using cRPA and G^0 is the analogous downfolding for the single-particle Green function. The exchange-correlation potential is removed from the Kohn-Sham Green function and replaced by the G^0W^0 self-energy. The exact double counting $\Sigma^{G^0W^0}|_I$ is known and thus the scheme is free from the double-counting problem.

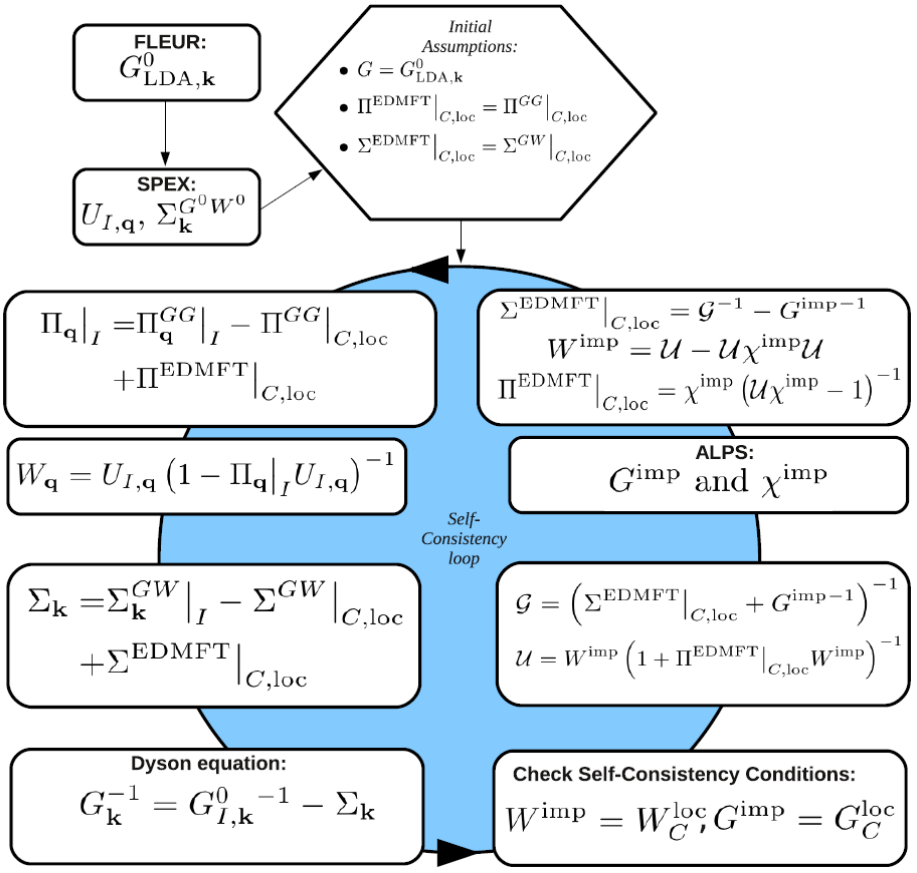


Fig. 4: *Self-consistency flow chart of the multitier GW+EDMFT scheme. Reproduced from Nilsson et al. [19].*

The bare propagators G^0 and U are computed only once and used in subsequent self-consistent GW+EDMFT calculation within the intermediate subspace. At each iteration the GW self-energy within the intermediate subspace (tier II) is computed, while the local part of the self-energy within the correlated subspace $\mathcal{C}$ is computed using the EDMFT impurity solver (tier I). The two self-energies are then summed with the double counting removed. The screened interaction W is computed in a similar fashion in which the polarization plays the role of the self-energy. This yields the final equations for the single particle Green function G and screened interaction W

$$G^{-1}_{\mathbf{k}} = \underbrace{G^{0\,-1}_{\mathbf{k}}}_{\text{TIER III}} - \underbrace{\left(\Sigma^{GW}_{\mathbf{k}}|_I - \Sigma^{GW}|_{C,\mathrm{loc}} + \Delta V_{\mathrm{H}}|_I\right)}_{\text{TIER II}} - \underbrace{\Sigma^{\mathrm{imp}}|_{C,\mathrm{loc}}}_{\text{TIER I}} . \tag{59}$$

$$W^{-1}_{\mathbf{q}} = \underbrace{U^{-1}_{\mathbf{q}}}_{\text{TIER III}} - \underbrace{\left(P^{GG}_{\mathbf{q}}|_I - P^{GG}|_{C,\mathrm{loc}}\right)}_{\text{TIER II}} - \underbrace{P^{\mathrm{imp}}|_{C,\mathrm{loc}}}_{\text{TIER I}} . \tag{60}$$

$\Sigma^{GW}_{\mathbf{k}}$ is the self-energy computed within the self-consistent GWA and Σ^{imp} the local impurity self-energy computed within EDMFT. $\Delta V_{\mathrm{H}}|_I$ is the change of the Hartree potential within the intermediate subspace. $\Sigma^{GW}|_{C,\mathrm{loc}}$ and $P^{GG}|_{C,\mathrm{loc}}$ are, respectively, the double counting for the self-energy and the polarization in tier II.

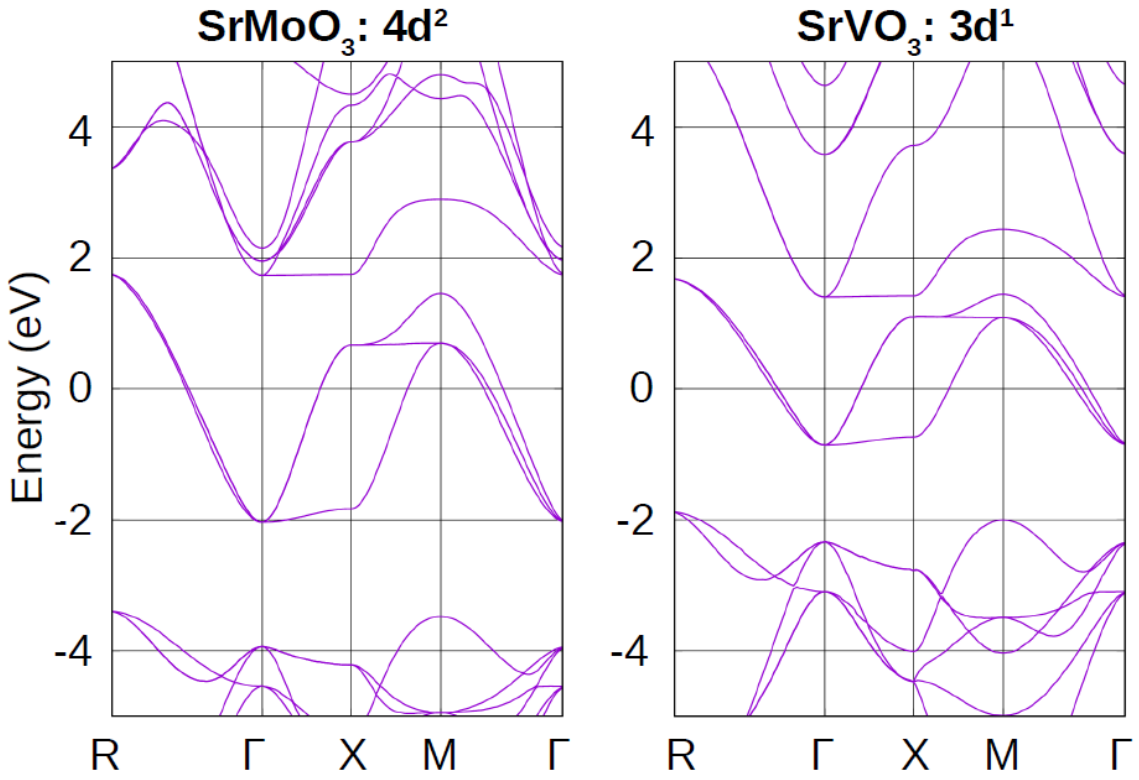


Fig. 5: *LDA bandstructures of $SrMoO_3$ and $SrVO_3$ with t_{2g} bands crossing the Fermi level. Figures courtesy of Fredrik Nilsson.*

The impurity EDMFT self-energy and polarization provide local vertex corrections within the correlated subspace through, respectively, $\Sigma^{\rm imp}\big|_{C,\rm loc}$ and $P^{\rm imp}\big|_{C,\rm loc}$ at each iteration in the self-consistency cycle in tier II. This feature, a feedback between local and nonlocal correlations both on the single-particle and two particle level, is not accounted for in one-shot GW+DMFT. The self-consistency cycle is illustrated in Fig. 4.

3.5 Applications to cubic perovskites $SrMoO_3$ and $SrVO_3$

Application of the GW+EDMFT method to real materials is illustrated for the case of cubic perovskites $SrMoO_3$ and $SrVO_3$. In Fig. 5 the LDA bandstructure for the two compounds are shown. Both compounds have an open d-shell, split into the t_{2g} and e_g components by the crystal field, with the e_g bands lying higher in energy and unoccupied. As can be seen in the figure, the t_{2g}-band of $SrMoO_3$ is wider than that of $SrVO_3$ since the $4d$-orbitals of Mo are more spread than the $3d$-orbitals of V. It is then expected that $SrMoO_3$ is less correlated than $SrVO_3$ and as such these compounds provide an interesting testing ground for GW+EDMFT.

LDA+DMFT calculations cannot reproduce the satellite structures of $SrMoO_3$ unless U is chosen to be unrealistically large. This suggests that the satellites may be of different nature from the commonly assumed Hubbard bands and it was proposed that they could be plasmons arising from long-range charge fluctuations [21].

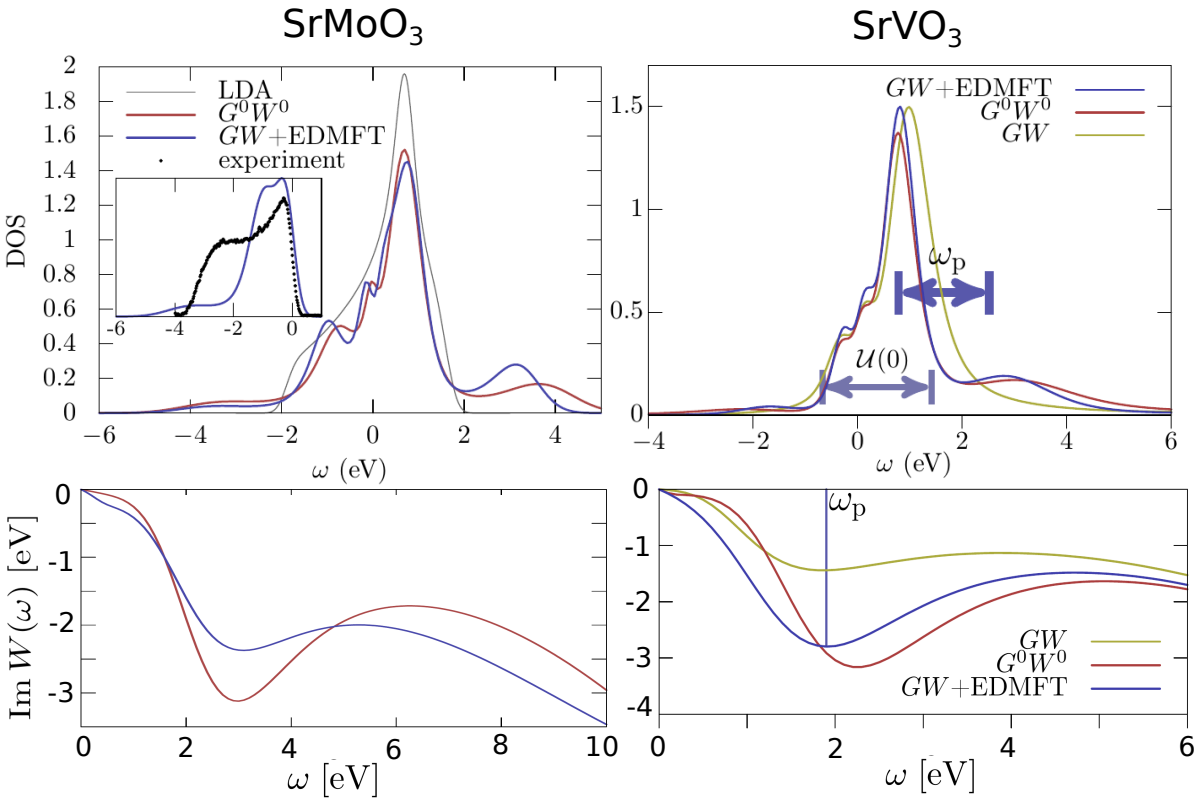


Fig. 6: *Spectral functions of $SrMoO_3$ and $SrVO_3$ calculated using GW+EDMFT. For $SrMoO_3$ comparison is made with LDA and one-shot GW as well as with experiment. For $SrVO_3$ comparison is made with one-shot and fully self-consistent GW. The lower figures show the imaginary part of the screened interaction. The figures are reproduced from Boehnke et al. [20] and Nilsson et al. [19]. The experimental spectra for $SrMoO_3$ is taken from Wadati et al. [21].*

Contrary to $SrMoO_3$ the electronic structure of $SrVO_3$ can be reproduced relatively well within LDA+DMFT with a suitable choice of U [17]. Experimentally, a significantly renormalized quasiparticle peak compared with the LDA bandwidth as well as upper and lower satellite structures can be observed. $SrVO_3$ has therefore been commonly thought of as a strongly correlated metal with upper and lower Hubbard bands.

It has however been found that the spectral function of $SrVO_3$ can be well reproduced within the cumulant expansion [22]. The cumulant expansion describes long-range charge fluctuations (plasmons) but cannot account for the strong local correlations giving rise to Hubbard bands. This raises the question concerning the physical origin of the satellites in $SrVO_3$, whether they are collective excitations or atomic in origin. This issue can neither be addressed by DMFT nor the cumulant method since the former accounts only local correlations but ignores long-range charge fluctuations whereas the latter is a theory constructed to describe a coupling between electrons and bosonic excitations such as plasmons and phonons. GW+EDMFT, on the other hand, accounts for these two aspects of local and nonlocal correlations and is therefore well suited to investigate the nature of the satellites in $SrVO_3$.

In Fig. 6 the t_{2g}-spectral functions for $SrMoO_3$ and $SrVO_3$ and $\mathrm{Im}\, W$ computed using the multitier GW+EDMFT method are shown. In these calculations the intermediate subspace (and hence the self-consistency) was restricted to the t_{2g}-bands. Information about collective charge excitations in $\mathrm{Im}\, W$ is manifested as peaks at the excitation energies. From Fig. 6 it is quite evident that these excitations are present in both $SrMoO_3$ and $SrVO_3$. In $SrMoO_3$ these long-range charge fluctuations give rise to a shoulder structure in the spectral function. This conforms with the suggestion by Wadati *et al.* [21] that the shoulder structure is of plasmonic origin although the satellite position and weight are lower compared with experiment. However, oxygen vacancies can lead to an overestimation of the satellite weight [23] in these materials.

More revealing is the results for $SrVO_3$. The feedback from long-range screening included in the self-consistency cycle reduces the effective impurity interaction $\mathcal{U}$ substantially. It is clear from Fig. 6 that the self-consistently computed $\mathcal{U}$ for $SrVO_3$ is much too small to account for the satellites as Hubbard bands, whose separation is approximately given by $\mathcal{U}(0)$. Similar to $SrMoO_3$, long-range charge fluctuations give rise to a peak in $\mathrm{Im}\, W$ which in turn yields plasmon satellites in the spectral function. Thus, these calculations strongly suggest that $SrVO_3$ is a moderately correlated metal with plasmon satellites rather than Hubbard bands of atomic origin. It is worth noting that the quasiparticle renormalization in GW+EDMFT is slightly underestimated compared with experiment, which points to the important role of nonlocal vertex corrections beyond GW+EDMFT.

GW+EDMFT has also been applied to investigate screening from e_g-states and antiferromagnetic correlations in $d^{(1,2,3)}$-perovskites [24] as well as normal state of $Nd_{1-x}Sr_xNiO_2$ [25].

Appendices

A Derivation of the GW approximation

In this Appendix, the GW approximation is derived in detail. The derivations have been carried out for zero temperature, but extension to finite temperature is quite straightforward.
In the occupation number representation, the electronic part of the Hamiltonian is given by

$$\hat{H} = \int dr\, \hat{\psi}^\dagger(r)\, h_0(r)\, \hat{\psi}(r) + \frac{1}{2}\int dr dr'\, \hat{\psi}^\dagger(r)\, \hat{\psi}^\dagger(r')\, v(r{-}r')\, \hat{\psi}(r')\, \hat{\psi}(r)\,, \tag{61}$$

where

$$h_0 = -\frac{1}{2}\nabla^2 + V(r) \text{ and } \; v(r{-}r') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{62}$$

$\hat{\psi}(r)$ is the field operator which annihilates an electron at r and $\hat{\psi}^\dagger(r)$ is its conjugate which creates an electron at r. In our notation, r is a combined variable for space and spin: $r = (\mathbf{r}, \sigma)$ and an integer represents space, spin, and time: $n = (r_n, t_n)$. Atomic units are used throughout, where

$$m = \hbar = e = a_0 = 1. \tag{63}$$

The atomic unit of energy is Hartree and one Hartree is equal to 2 Rydberg or about 27.2 eV.

A.1 Equation of motion

We will define and work out the self-energy using the equation of motion approach. From the equation of motion of the field operator

$$i\frac{\partial \hat{\psi}_{\rm H}(1)}{\partial t_1} = \Big[\hat{\psi}_{\rm H}(1), \hat{H}\Big]\,, \tag{64}$$

one obtains

$$\left(i\frac{\partial}{\partial t_1} - h_0(1)\right) G(1,2) + i\int d3\; v(1{-}3)\, G^{(2)}(1,2,3,3^+) = \delta(1{-}2), \tag{65}$$

where

$$v(1{-}2) = v(r_1{-}r_2)\, \delta(t_1{-}t_2)\,, \tag{66}$$

and $G^{(2)}$ is the two-particle Green function defined as

$$G^{(2)}(1,2,3,4) = (-i)^2 \big\langle \Psi_0 \big| T\; \hat{\psi}_{\rm H}(1)\, \hat{\psi}^\dagger_{\rm H}(2)\, \hat{\psi}_{\rm H}(3)\, \hat{\psi}^\dagger_{\rm H}(4) \big| \Psi_0 \big\rangle. \tag{67}$$

The notation 3^+ means (r_3, t_3^+).
The equation of motion of the two-particle Green function will in turn generate the three-particle Green function and so forth resulting in a hierarchy of ever increasing order of Green functions. The mass operator M is introduced to truncate the hierarchy:

$$i\int d3\, v(1{-}3)\, G^{(2)}(1,2,3,3^+) = -\int d3\, M(1,3)\, G(3,2)\,. \tag{68}$$

The self-energy Σ is defined to be the mass operator with the mean-field Hartree potential $V_{\rm H}$ removed

$$\Sigma(1,2) = M(1,2) - V_{\rm H}(1)\,\delta(1{-}2)\,, \tag{69}$$

where

$$V_{\rm H}(1) = \int d3\, v(1{-}3)\rho(3)\,. \tag{70}$$

The equation of motion of the Green function then becomes

$$\left(i\frac{\partial}{\partial t_1} - h(1)\right) G(1,2) - \int d3\, \Sigma(1,3)\, G(3,2) = \delta(1{-}2)\,, \tag{71}$$

where h includes the Hartree potential

$$h = h_0 + V_{\rm H}. \tag{72}$$

From the equation of motion in Eq. (71) the inverse Green function is given by

$$G^{-1}(1,2) = \left(i\frac{\partial}{\partial t_1} - h(1)\right)\delta(1{-}2) - \Sigma(1,2)\,. \tag{73}$$

This expression will be used to construct an iterative equation for the self-energy.

A.2 Self-energy

Instead of following the conventional diagrammatic approach of many-body perturbation theory (see, e.g., Fetter and Walecka [26]), the self-energy can be conveniently derived using the functional derivative method [11, 12]. By applying a time-dependent probing field $\varphi(rt)$ the response of the Green function with respect to this probing field can be related to the self-energy. Once derivatives are taken, the field is set to zero. In the presence of an external field it is suitable to work in the interaction representation and define the Green function as

$$iG(1,2) = \frac{\big\langle \Psi_0 \big| T\; \hat{S}\hat{\psi}_{\rm D}(1)\,\hat{\psi}^\dagger_{\rm D}(2) \big| \Psi_0 \big\rangle}{\big\langle \Psi_0 \big| \hat{S} \big| \Psi_0 \big\rangle} \tag{74}$$

where $\hat{\psi}_{\rm D}(1)$ is the field operator in the interaction (Dirac) picture

$$\hat{\psi}_{\rm D}(rt) = {\rm e}^{i\hat{H}t}\,\hat{\psi}(r)\,e^{-i\hat{H}t}. \tag{75}$$

It is to be noted that the above field operators are independent of the probing field $\varphi(rt)$. The operator $\hat{S}$ is defined as

$$\hat{S} = \hat{U}_{\rm D}(\infty, -\infty), \tag{76}$$

where

$$\hat{U}_{\rm D}(t,t') = T\,{\rm e}^{-i\int_{t'}^{t} d\tau \hat{\phi}(\tau)} \tag{77}$$

is the time-evolution operator in the interaction picture, with

$$\hat{\phi}(t) = \int d^3r\, \hat{\rho}_{\rm D}(rt)\,\varphi(rt)\,, \tag{78}$$

where $\hat{\rho}_{\rm D}(rt)$ is the density operator. The definition of G in Eq. (74) reduces to the usual definition when $\varphi = 0$.

We will now take the functional derivative of G with respect to φ. In order to do this it is only necessary to work out the functional derivative of $\hat{S}$ with respect to φ since the field operators in the Dirac picture do not depend on φ:

$$\frac{\delta \hat{S}}{\delta\varphi(3)} = \frac{\delta}{\delta\varphi(3)} T\,{\rm e}^{-i\int d4\,\hat{\rho}_{\rm D}(4)\,\varphi(4)} = -i\,T\,\hat{S}\,\hat{\rho}_{\rm D}(3)\,. \tag{79}$$

We obtain from Eq. (74)

$$\begin{aligned} i\frac{\delta G(1,2)}{\delta\varphi(3)} = &- \frac{i\big\langle\Psi_0\big|T\,\hat{S}\,\hat{\rho}_{\rm D}(3)\,\hat{\psi}_{\rm D}(1)\,\hat{\psi}^\dagger_{\rm D}(2)\big|\Psi_0\big\rangle}{\big\langle\Psi_0\big|\hat{S}\big|\Psi_0\big\rangle} \\ &+ \frac{i\big\langle\Psi_0\big|T\,\hat{S}\,\hat{\psi}_{\rm D}(1)\,\hat{\psi}^\dagger_{\rm D}(2)\big|\Psi_0\big\rangle\big\langle\Psi_0\big|T\,\hat{S}\,\hat{\rho}_{\rm D}(3)]\big|\Psi_0\big\rangle}{\big\langle\Psi_0\big|\hat{S}\big|\Psi_0\big\rangle^2}\,. \end{aligned} \tag{80}$$

The two-particle Green function in the interaction picture is defined according to

$$G^{(2)}(1,2,3,4) = (-i)^2\frac{\big\langle\Psi_0\big|T\,\hat{S}\,\hat{\psi}_{\rm D}(1)\,\hat{\psi}^\dagger_{\rm D}(2)\,\hat{\psi}_{\rm D}(3)\,\hat{\psi}^\dagger_{\rm D}(4)\big|\Psi_0\big\rangle}{\big\langle\Psi_0\big|\hat{S}\big|\Psi_0\big\rangle}\,, \tag{81}$$

and obtain

$$\frac{\delta G(1,2)}{\delta\varphi(3)} = iG(1,2)\rho(3) - G^{(2)}(1,2,3,3^+). \tag{82}$$

This relation is fundamental because it expresses the two-particle Green function, and hence the self-energy through Eqs. (68) and (69), as a linear response of the one-particle Green function with respect to the probing field.

A formal expression for the self-energy can now be written down by substituting $G^{(2)}$ in Eq. (68) with the expression in Eq. (82)

$$\begin{aligned} \Sigma(1,2) = &\quad i\int d3\,d4\;v(1{-}3)\frac{\delta G(1,4)}{\delta\varphi(3)}G^{-1}(4,2) \\ = &-i\int d3\,d4\;v(1{-}3)\,G(1,4)\frac{\delta G^{-1}(4,2)}{\delta\varphi(3)}\,, \end{aligned} \tag{83}$$

where in the second line the following identity has been used

$$G\,G^{-1} = 1 \to \delta G\,G^{-1} + G\,\delta G^{-1} = 0 \to \delta G = -G\,\delta G^{-1}\,G\,. \tag{84}$$

From the equation of motion in (71) one finds, keeping in mind that the probing field φ is present in h,

$$\frac{\delta G^{-1}(4,2)}{\delta\varphi(3)} = -\left(\delta(4{-}3) + \frac{\delta V_{\rm H}(4)}{\delta\varphi(3)}\right)\delta(4{-}2) - \frac{\delta\Sigma(4,2)}{\delta\varphi(3)}\,. \tag{85}$$

This expression together with Eq. (83) yields an *iterative* equation for the self-energy, which facilitates an expansion of the self-energy in powers of the Coulomb interaction. Since

$$V_{\rm H}(4) = \int d5\,v(4{-}5)\rho(5)\,, \tag{86}$$

we have

$$\frac{\delta V_{\rm H}(4)}{\delta\varphi(3)} = \int d5\, v(4{-}5)\frac{\delta\rho(5)}{\delta\varphi(3)} = \int d5\, v(4{-}5)\, R(5,3)\,, \tag{87}$$

where R is the time-ordered linear density response function

$$R(1,2) = \frac{\delta\rho(1)}{\delta\varphi(2)} = -i\frac{\delta G(1,1^+)}{\delta\varphi(2)}\,. \tag{88}$$

From Eq. (80) we find, with $2 \to 1^+$ and after setting $\varphi = 0 \to \hat{S} = 1$, the time-ordered linear density response function is given by

$$iR(1,2) = \big\langle\Psi_0\big|T\,\hat\rho(1)\,\hat\rho(2)\big|\Psi_0\big\rangle - \rho(1)\,\rho(2)\,. \tag{89}$$

The spectral representation can be obtained by inserting a complete set of N-electron eigenstates of the Hamiltonian and performing a Fourier transformation yielding

$$R(r_1,r_2;\omega) = \sum_n \left(\frac{\big\langle\Psi_0\big|\Delta\hat\rho(r_1)\big|\Psi_n\big\rangle\big\langle\Psi_n\big|\Delta\hat\rho(r_2)\big|\Psi_0\big\rangle}{\omega - (E_n{-}E_0) + i\eta} - \frac{\big\langle\Psi_0\big|\Delta\hat\rho(r_2)\big|\Psi_n\big\rangle\big\langle\Psi_n\big|\Delta\hat\rho(r_1)\big|\Psi_0\big\rangle}{\omega + (E_n{-}E_0) - i\eta} \right), \tag{90}$$

where

$$\Delta\hat\rho(r) = \hat\rho(r) - \rho(r)\,.$$

For analytical treatment it is useful to write the density response function in terms of its spectral representation as

$$R(r_1,r_2;\omega) = \int_{-\infty}^{0} d\omega'\,\frac{S(r_1,r_2;\omega')}{\omega-\omega'-i\eta} + \int_0^{\infty} d\omega'\,\frac{S(r_1,r_2;\omega')}{\omega-\omega'+i\eta}\,, \tag{91}$$

where S is the spectral function given by

$$S(r_1,r_2;\omega) = \sum_n \big\langle\Psi_0\big|\Delta\hat\rho(r_1)\big|\Psi_n\big\rangle\big\langle\Psi_n\big|\Delta\hat\rho(r_2)\big|\Psi_0\big\rangle\,\delta(\omega{-}E_n{+}E_0)\,, \tag{92}$$

defined to be anti-symmetric in frequency,

$$S(-\omega) = -S(\omega)\,.$$

The above spectral function is valid for systems with time-reversal symmetry (no magnetic field) for which

$$\big\langle\Psi_0\big|\Delta\hat\rho(r_1)\big|\Psi_n\big\rangle\big\langle\Psi_n\big|\Delta\hat\rho(r_2)\big|\Psi_0\big\rangle = \big\langle\Psi_0\big|\Delta\hat\rho(r_2)\big|\Psi_n\big\rangle\big\langle\Psi_n\big|\Delta\hat\rho(r_1)\big|\Psi_0\big\rangle. \tag{93}$$

The density response functions are central quantities in the calculation of the self-energy. Iteration of Eq. (83) together with (85) generates increasingly higher order density response functions, which determine the self-energy. The iteration starts by setting $\delta\Sigma/\delta\varphi = 0$ yielding the well-known GW approximation.

A.3 Self-energy expansion in the screened interaction

It is possible to reformulate the iterative equation for the self-energy in Eqs. (83) and (85) in terms of the screened interaction W instead of the bare interaction v [11,12]. Instead of varying the bare probing field φ, one varies the total screened potential $V_{\rm tot}$ in Eq. (6).
The linear density response function is reformulated by introducing the polarization function P, defined as the linear response function with respect to the *total* potential $V_{\rm tot}$

$$P(1,2) = \frac{\delta\rho(1)}{\delta V_{\rm tot}(2)}\,. \tag{94}$$

In terms of the polarization function, the induced density becomes

$$\rho_{\rm ind}(1) = \int d2\; P(1,2)\, V_{\rm tot}(2)\,. \tag{95}$$

Equating Eqs. (4) and (95) and using Eq. (6) for $V_{\rm tot}$ and Eq. (5) for $V_{\rm ind}$ one finds

$$R(1,2) = P(1,2) + \int d3\, d4\; P(1,4)\, v(4{-}3)\, R(3,2)\,. \tag{96}$$

The response function R is related to the polarization function as

$$R(1,2) = \frac{\delta\rho(1)}{\delta\varphi(2)} = \int d3\, \frac{\delta\rho(1)}{\delta V_{\rm tot}(3)} \frac{\delta V_{\rm tot}(3)}{\delta\varphi(2)} = \int d3\; P(1,3)\, \varepsilon^{-1}(3,2) \tag{97}$$

and Eq. (56) for the screened interaction can be written as

$$W(1,2) = v(1{-}2) + \int d3\, d4\; v(1{-}3)\, P(3,4)\, W(4,2)\,. \tag{98}$$

Using the chain rule for functional derivatives, the iterative equation for the self-energy can now be written in terms of the screened interaction W

$$\frac{\delta G^{-1}(4,2)}{\delta\varphi(3)} = \int d5\, \frac{\delta G^{-1}(4,2)}{\delta V_{\rm tot}(5)} \frac{\delta V_{\rm tot}(5)}{\delta\varphi(3)} = -\int d5 \left(\delta(5{-}4)\, \delta(4{-}2) + \frac{\delta\Sigma(4,2)}{\delta V_{\rm tot}(5)} \right) \varepsilon^{-1}(5,3), \tag{99}$$

where the second line has been obtained from the definition of ε^{-1} in Eq. (8) and from Eq. (73). Substituting the above expression for $\delta G^{-1}/\delta\varphi$ into Eq. (83) yields

$$\Sigma(1,2) = iG(1,2)\, W(2,1) + i \int d4\, d5\; G(1,4)\, W(5,1)\, \frac{\delta\Sigma(4,2)}{\delta V_{\rm tot}(5)}\,, \tag{100}$$

where W is given in Eq. (56). This iterative equation clearly displays the expansion of the self-energy in powers of the screened interaction W as the equation is iterated. The first term is the well-known GW approximation.

A.4 GW approximation

Fourier transformation of

$$\Sigma(1,2) = iG(1,2)\,W(2,1) \tag{101}$$

yields, using convolution theorem,

$$\Sigma(r_1,r_2;\omega) = i\int\frac{d\omega'}{2\pi}\,G(r_1,r_2;\omega+\omega')\,W(r_2,r_1;\omega')\,\mathrm{e}^{i\eta\omega'}. \tag{102}$$

Using the spectral representation of G,

$$G(r_1,r_2;\omega) = \int_{-\infty}^{\mu} d\omega'\,\frac{A^{\mathrm{h}}(r_1,r_2\,;\omega')}{\omega-\omega'-i\delta} + \int_{\mu}^{\infty} d\omega'\,\frac{A^{\mathrm{e}}(r_1,r_2;\,\omega')}{\omega-\omega'+i\delta}\,, \tag{103}$$

one finds for the exchange part

$$\Sigma^{\mathrm{x}}(r_1,r_2) = iv(r_1{-}r_2)\int\frac{d\omega'}{2\pi}\int_{-\infty}^{\mu}d\omega_1\,\frac{A^{\mathrm{h}}(r_1,r_2;\omega_1)\,\mathrm{e}^{i\eta\omega'}}{\omega+\omega'-\omega_1-i\delta} = -v(r_1{-}r_2)\int_{-\infty}^{\mu}d\omega_1\,A^{\mathrm{h}}(r_1,r_2;\omega_1)\,. \tag{104}$$

The factor $\exp(i\eta\omega')$ ensures that the contour is closed along a semicircle above the real axis so that only poles below the chemical potential corresponding to occupied states are picked up.
For the correlation part we find, using the spectral representation of W^c in Eq. (14), dropping the space variables for clarity,

$$\Sigma^{\mathrm{c}}(\omega) = \int_{-\infty}^{\mu}d\omega_1\int_0^{\infty}d\omega_2\frac{A^{\mathrm{h}}(\omega_1)B(\omega_2)}{\omega+\omega_2-\omega_1-i\delta} - \int_{\mu}^{\infty}d\omega_1\int_{-\infty}^{0}d\omega_2\frac{A^{\mathrm{e}}(\,\omega_1)B(\omega_2)}{\omega+\omega_2-\omega_1+i\delta}\,. \tag{105}$$

The spectral representation of the correlation part of the self-energy has the same form as that of the Green function and can be written as

$$\Sigma^{\mathrm{c}}(r_1,r_2;\omega) = \int_{-\infty}^{\mu}d\omega'\,\frac{\Gamma^{\mathrm{h}}(r_1,r_2;\omega')}{\omega-\omega'-i\delta} + \int_{\mu}^{\infty}d\omega'\,\frac{\Gamma^{\mathrm{e}}(r_1,r_2;\omega')}{\omega-\omega'+i\delta}, \tag{106}$$

where

$$\Gamma^{\mathrm{h}}(\omega) = \quad\frac{1}{\pi}\,\mathrm{Im}\,\Sigma^{\mathrm{c}}(\omega)\,\Theta(\mu{-}\omega)\,, \tag{107}$$

$$\Gamma^{\mathrm{e}}(\omega) = -\frac{1}{\pi}\,\mathrm{Im}\,\Sigma^{\mathrm{c}}(\omega)\,\Theta(\omega{-}\mu)\,. \tag{108}$$

If we use a noninteracting Green function

$$G^0(r_1,r_2;\omega) = \sum_n^{\mathrm{occ}}\frac{\phi_n(r_1)\phi_n^*(r_2)}{\omega-\varepsilon_n-i\delta} + \sum_n^{\mathrm{unocc}}\frac{\phi_n(r_1)\phi_n^*(r_2)}{\omega-\varepsilon_n+i\delta}, \tag{109}$$

corresponding to

$$A^{\mathrm{h}}(r_1,r_2;\omega) = \sum_n^{\mathrm{occ}}\phi_n(r_1)\,\phi_n^*(r_2)\,\delta(\omega{-}\varepsilon_n)\,, \tag{110}$$

$$A^{\mathrm{e}}(r_1,r_2;\omega) = \sum_n^{\mathrm{unocc}}\phi_n(r_1)\,\phi_n^*(r_2)\,\delta(\omega{-}\varepsilon_n)\,, \tag{111}$$

where the one-particle states $\{\phi_n\}$ with eigenvalues $\{\varepsilon_n\}$ are usually chosen to be the Kohn-Sham orbitals or Bloch states, then the exchange potential reduces to the familiar expression

$$\Sigma^{\rm x}(r_1,r_2) = -v(r_2{-}r_1)\sum_n^{\rm occ}\phi_n(r_1)\,\phi_n^*(r_2). \tag{112}$$

The spectral functions of the correlation part of the self-energy become

$$I^{\rm h}(r_1,r_2;\omega\le\mu) = \sum_n^{\rm occ}\phi_n(r_1)\,\phi_n^*(r_2)\,B(r_2,r_1;\varepsilon_n{-}\omega)\,\Theta(\varepsilon_n{-}\omega)\,, \tag{113}$$

$$I^{\rm e}(r_1,r_2;\omega>\mu) = \sum_n^{\rm unocc}\phi_n(r_1)\,\phi_n^*(r_2)\,B(r_2,r_1;\omega{-}\varepsilon_n)\,\Theta(\omega{-}\varepsilon_n)\,. \tag{114}$$

The correlation part of the self-energy can then be obtained from the spectral representation in Eq. (106).

The above expression shows that collective charge excitations (plasmons) embodied in $B(\omega)$, proportional to $\mathrm{Im}\,W$, are transferred to $\mathrm{Im}\,\Sigma$. Through the Dyson equation these plasmon excitations are coupled to the Green function and appear as satellites in the spectral function. This explicitly describes a clear picture of the coupling between the electrons and the collective charge excitations in the system, which is illustrated in the Feynman diagram in Fig. 1. An added electron or hole represented by the Green function line interacts with the system and induces a collective charge excitation or a plasmon which is reabsorbed at a later time. This coupling to the plasmon excitation renormalizes the noninteracting Green function via the self-energy, resulting in a heavier effective mass and the transfer of quasiparticle weight to the satellite region.

B Functional derivation of *GW*+EDMFT

The Hamiltonian of a many-electron system moving in some external potential $v_{\rm ext}$ is given by

$$\hat H = \hat H_0 + \hat V, \tag{115}$$

where

$$\hat H_0 = \int dr\,\hat\psi^\dagger(r)\,h_0(r)\,\hat\psi(r)\,, \tag{116}$$

$$h_0(r) = -\frac{1}{2}\bigtriangledown^2 + v_{\rm ext}(r), \tag{117}$$

$$\hat V = \frac{1}{2}\int dr dr'\,\hat\psi^\dagger(r)\,\hat\psi^\dagger(r')\,v(r{-}r')\,\hat\psi(r')\,\hat\psi(r)\,. \tag{118}$$

r labels position and spin variables: $r=(\mathbf{r},\sigma)$ and $v(r{-}r') = 1/|\mathbf{r}{-}\mathbf{r}'|$. The interaction term of the Hamiltonian can be rewritten as

$$\hat V = \frac{1}{2}\int dr dr'\,\hat\rho(r)\,v(r{-}r')\big(\hat\rho(r') - \delta(r{-}r')\big), \tag{119}$$

where the second term is the unphysical self-interaction term. $\hat{V}$ can be expressed in terms of the density fluctuation operator,

$$\Delta\hat{\rho}(r) = \hat{\rho}(r) - \rho(r), \tag{120}$$

as

$$\hat{V} = \Delta\hat{V} + \int dr\, V_{\rm H}(r)\hat{\rho}(r) - E_{\rm H} - \frac{1}{2}\int drdr'\, \delta(r-r')\, v(r-r')\, \hat{\rho}(r), \tag{121}$$

where

$$\Delta\hat{V} = \frac{1}{2}\int drdr'\, \Delta\hat{\rho}(r)\, v(r-r')\, \Delta\hat{\rho}(r'), \tag{122}$$

$$V_{\rm H}(r) = \int dr'\, v(r-r')\, \rho(r'), \tag{123}$$

and

$$E_{\rm H} = \frac{1}{2}\int drdr'\, \rho(r)\, v(r-r')\, \rho(r'). \tag{124}$$

The Hartree potential $V_{\rm H}$ can be incorporated into the one-particle part of the Hamiltonian $\hat{H}_0$ and the Hartree energy $E_{\rm H}$ can be absorbed into the chemical potential. The self-interaction term can also be included in $\hat{H}_0$ and it will eventually cancel out of physical observables.
In the functional integral formalism, one evaluates at each time slice $\epsilon = \beta/M$ [27]

$$\mathrm{e}^{-\epsilon(\hat{H}-\mu N)} = \mathrm{e}^{-\epsilon(\hat{H}_0-\mu\hat{N}+\hat{V})} \sim \mathrm{e}^{-\epsilon(\hat{H}_0'-\mu\hat{N})}\mathrm{e}^{-\epsilon\Delta\hat{V}}, \tag{125}$$

where $\hat{H}_0'$ includes the one-particle residue $\hat{V}-\Delta\hat{V}$. Using the Hubbard-Stratonovich transformation, which is essentially a Gaussian integral formula for operators, one has

$$\int\prod_i\left(\frac{d\theta_i}{\sqrt{2\pi}}\right)\mathrm{e}^{-\frac{1}{2}\theta A\theta+\theta\hat{J}} = \frac{1}{\sqrt{\det A}}\,\mathrm{e}^{\frac{1}{2}\hat{J}A^{-1}\hat{J}}. \tag{126}$$

Applying this operator identity one obtains

$$\mathrm{e}^{-\epsilon\Delta\hat{V}} = \mathrm{e}^{-\epsilon\frac{1}{2}\int drdr'\Delta\hat{\rho}(r)v(r-r')\Delta\hat{\rho}(r')}. \tag{127}$$

Associating $\hat{J}\to i\Delta\hat{\rho}$, $A^{-1}\to\epsilon v$ and $\theta\to\epsilon\phi$, one finds

$$\mathrm{e}^{-\epsilon\Delta\hat{V}} = [\det(\epsilon v^{-1})]^{1/2}\int\frac{d\phi}{\sqrt{2\pi}}\,\mathrm{e}^{-\epsilon\frac{1}{2}\int drdr'\phi(r)v^{-1}(r-r')\phi(r')+i\epsilon\int dr\phi(r)\Delta\hat{\rho}(r)}, \tag{128}$$

where ϕ is an auxiliary bosonic field which is real (not a Grassmann variable).
The partition function is given by

$$Z = \int\mathcal{D}[\psi^*\psi]\,\mathrm{e}^{-S(\psi^*,\psi)}, \tag{129}$$

where a correspondence between the annihilation and creation field operators and the Grassmann variables has been made as follows:

$$\hat{\psi}(x)\to\psi(x),\ \ \hat{\psi}^\dagger(x)\to\psi^*(x). \tag{130}$$

The label x represents position, spin and imaginary time variables: $x = (r, \tau) = (\mathbf{r}, \sigma, \tau)$. The Coulomb interaction is written as

$$v(x{-}x') = v(r{-}r')\,\delta(\tau{-}\tau'). \tag{131}$$

The action is given by

$$S(\psi^*,\psi) = \int dx\,\psi^*(x)\left(\frac{\partial}{\partial\tau} - \mu + h_0(x)\right)\psi(x) + \frac{1}{2}\int dx dx'\,\psi^*(x)\,\psi^*(x')\,v(x{-}x')\,\psi(x')\,\psi(x). \tag{132}$$

Using the Hubbard-Stratonovich transformation described above the action can be naturally decomposed as a sum of three terms

$$S(\psi^*,\psi,\phi) = S_{\rm F}(\psi^*,\psi) + S_{\rm B}(\phi) + S_{\rm BF}(\psi^*,\psi,\phi), \tag{133}$$

where

$$S_{\rm F}(\psi^*,\psi) = \int dx\,\psi^*(x)\left(\frac{\partial}{\partial\tau} - \mu + h_0(x) + V_H(x)\right)\psi(x), \tag{134}$$

$$S_{\rm B}(\phi) = \frac{1}{2}\int dx dx'\,\phi(x)\,v^{-1}(x{-}x')\,\phi(x'), \tag{135}$$

$$S_{\rm BF}(\psi^*,\psi,\phi) = -i\int dx\,\phi(x)\big(\psi^*(x)\psi(x) - \rho(x)\big). \tag{136}$$

$S_{\rm F}$ and $S_{\rm B}$ correspond, respectively, to the fermion and boson mean-field actions. $S_{\rm BF}$ is the interaction term between the bosonic and fermionic fields. Implicit in this last term is the presence of the Coulomb interaction, which can be brought out by making a transformation $\phi = v\widetilde{\phi}$, yielding a more intuitive expression

$$S_{\rm BF}(\psi^*,\psi,\widetilde{\phi}) = \int dx dx'\,\widetilde{\phi}(x')\,v(x'{-}x)\big(\psi^*(x)\psi(x) - \rho(x)\big). \tag{137}$$

The partition function after the Hubbard-Stratonovich transformation becomes

$$Z = \int \mathcal{D}[\psi^*,\psi,\phi]\;\mathrm{e}^{-S_{\rm F}(\psi^*,\psi)-S_{\rm B}(\phi)}\mathrm{e}^{-S_{\rm BF}(\psi^*,\psi,\phi)}. \tag{138}$$

Using the coupling constant integration technique, the contribution to the thermodynamic potential $\Omega = -\ln Z$ from the $S_{\rm BF}$ coupling term can be calculated

$$Z_\alpha = \int \mathcal{D}[\psi^*,\psi,\phi]\;\mathrm{e}^{-S_{\rm F}(\psi^*,\psi)-S_{\rm B}(\phi)}\;\mathrm{e}^{-\alpha S_{\rm BF}(\psi^*,\psi,\phi)}. \tag{139}$$

$$\frac{\partial\Omega_\alpha}{\partial\alpha} = -\frac{\partial\ln Z_\alpha}{\partial\alpha} = -\frac{1}{Z_\alpha}\frac{\partial Z_\alpha}{\partial\alpha} = \big\langle S_{\rm BF}(\psi^*,\psi,\phi)\big\rangle_\alpha, \tag{140}$$

$$\Omega - \Omega_0 = \int_0^1 d\alpha\frac{\partial\Omega_\alpha}{\partial\alpha} = \int_0^1 d\alpha\big\langle S_{\rm BF}(\psi^*,\psi,\phi)\big\rangle_\alpha. \tag{141}$$

Introducing source potentials $J_{\rm F}$ and $J_{\rm B}$ which couple, respectively, to the fermion and boson fields one finds

$$Z[J_{\rm F},J_{\rm B}] = \int \mathcal{D}[\psi^*,\psi,\phi]\;\mathrm{e}^{-S_{\rm F}(\psi^*,\psi)-S_{\rm B}(\phi)-\psi^*J_{\rm F}\psi+\frac{1}{2}\phi J_{\rm B}\phi}\;\mathrm{e}^{-S_{\rm BF}(\psi^*,\psi,\phi)}. \tag{142}$$

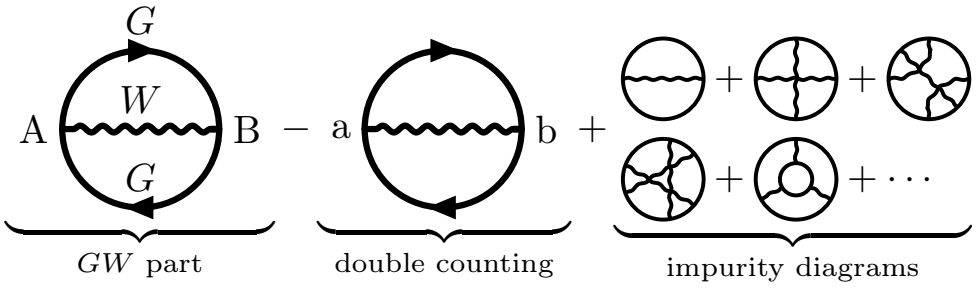


Fig. 7: *The GW+EDMFT approximation to the Ψ-functional. The first diagram corresponds to GWA and the second diagram the double counting. Lower-case indexes are restricted to the same unit cell, and thus correspond to diagrams accounted for in the impurity problem. The figure is taken from Boehnke et al. [20].*

Consider first the case when $S_{\rm BF} = 0$, i.e., no coupling between the bosonic and fermionic fields. $S_{\rm F}$ can be associated with a thermodynamic potential $\Omega_{\rm F}$ and its Legendre transform

$$\Gamma_F[G] = \Omega_{\rm F}[J_{\rm F}] - J_{\rm F}G = \mathrm{Tr}\ \ln G - \mathrm{Tr}\,(G_0^{-1}G - 1), \tag{143}$$

where

$$G_0^{-1}(x,x') = -\left(\frac{\partial}{\partial\tau} - \mu + v_{\rm ext}(x) + V_{\rm H}(x)\right)\delta(x-x'). \tag{144}$$

For the bosonic part,

$$Z_{\rm B}[J_{\rm B}] = \int \mathcal{D}[\phi]\ \mathrm{e}^{-\frac{1}{2}\phi(v^{-1}-J_{\rm B})\phi} = \left(\det(v^{-1} - J_{\rm B})\right)^{-1/2}, \tag{145}$$

where the Gaussian integral formula for real variables has been used. Defining the boson propagator

$$W(x,y) = \big\langle\phi(x)\phi(y)\big\rangle = \frac{2}{Z_{\rm B}}\frac{\delta Z_{\rm B}}{\delta J_{\rm B}(x,y)} = 2\frac{\delta \ln Z_{\rm B}}{\delta J_{\rm B}(x,y)}. \tag{146}$$

and using the identity

$$\ln\det M = \mathrm{Tr}\ \ln M, \tag{147}$$

one finds

$$\ln Z_{\rm B}[J_{\rm B}] = -\frac{1}{2}\ln\det\left(v^{-1} - J_{\rm B}\right) = -\frac{1}{2}\mathrm{Tr}\ \ln\left(v^{-1} - J_{\rm B}\right), \tag{148}$$

$$W[J_{\rm B}] = 2\frac{\delta \ln Z_{\rm B}[JB]}{\delta J_{\rm B}} = \left(v^{-1} - J_{\rm B}\right)^{-1} \ \rightarrow J_{\rm B}[W] = v^{-1} - W^{-1}, \tag{149}$$

$$\Omega_{\rm B}[J_{\rm B}] = -\ln Z_{\rm B}[J_{\rm B}] = \frac{1}{2}\mathrm{Tr}\ \ln W^{-1} = -\frac{1}{2}\mathrm{Tr}\ \ln W. \tag{150}$$

One then obtains

$$\Gamma_{\rm B}[W] = \Omega_{\rm B}[J_{\rm B}] + \frac{1}{2}\mathrm{Tr}\,(J_{\rm B}W) = -\frac{1}{2}\mathrm{Tr}\ \ln W + \frac{1}{2}\mathrm{Tr}\left(v^{-1}W - 1\right), \tag{151}$$

which shows that $\Gamma_{\rm B}$, the Legendre transform of $\Omega_{\rm B}$, is indeed a functional of W.

Taking into account the coupling term $S_{\rm BF}$ yields

$$\Gamma[G,W] = \Gamma_{\rm F}[G] + \Gamma_{\rm B}[W] + \Psi[G,W], \tag{152}$$

where from Eq. (141)

$$\Psi[G,W] = \int_0^1 d\alpha \, \langle S_{\rm BF}(\psi^*,\psi,\phi)\rangle_\alpha \, . \tag{153}$$

Functional derivative with respect to G yields the Dyson equation for G

$$\frac{\delta\Gamma[G,W]}{\delta G} = \frac{\delta\Gamma_{\rm F}[G]}{\delta G} + \frac{\delta\Psi[G,W]}{\delta G} \tag{154}$$

$$-J_{\rm F} = G^{-1} - G_0^{-1} + \frac{\delta\Psi[G,W]}{\delta G}, \tag{155}$$

with the fermion self-energy

$$\Sigma[G,W] = \frac{\delta\Psi[G,W]}{\delta G} + J_{\rm F}. \tag{156}$$

Similarly, functional derivative with respect to the boson propagator W yields

$$\frac{\delta\Gamma[G,W]}{\delta W} = \frac{\delta\Gamma_{\rm B}[W]}{\delta W} + \frac{\delta\Psi[G,W]}{\delta W} \tag{157}$$

$$\frac{1}{2}J_{\rm B} = -\frac{1}{2}W^{-1} + \frac{1}{2}v^{-1} + \frac{\delta\Psi[G,W]}{\delta W}, \tag{158}$$

with the boson self-energy or polarization

$$P[G,W] = -2\frac{\delta\Psi[G,W]}{\delta W} + J_{\rm B}. \tag{159}$$

For $J_{\rm F} = J_{\rm B} = 0$, which correspond to the extrema of $\Gamma[G,W]$ we find the Dyson equations for G and W

$$\frac{\delta\Gamma[G,W]}{\delta G} = 0 \rightarrow G \, = G_0 + G_0\Sigma G, \tag{160}$$

$$\frac{\delta\Gamma[G,W]}{\delta W} = 0 \rightarrow W = v + v\,P\,W. \tag{161}$$

The EDMFT approximation corresponds to an onsite approximation to G and W

$$\Psi^{\rm EDMFT} = \Psi[G_{\mathbf{RR}}, W_{\mathbf{RRRR}}]. \tag{162}$$

GWA, on the other hand, corresponds to the first-order diagram in Ψ

$$\Psi^{GW} = -\frac{1}{2}{\rm Tr}\, GWG \, . \tag{163}$$

This yields the total GW+EDMFT functional in Fig. 7

$$\Psi^{GW+{\rm EDMFT}} = \Psi^{GW} + \Psi^{\rm EDMFT} + \frac{1}{2}{\rm Tr}(G_{\mathbf{RR}}W_{\mathbf{RRRR}}G_{\mathbf{RR}}), \tag{164}$$

where the last term is the double-counting term.

References

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

[2] For a review see, e.g., K. Held, Adv. Phys. **56**, 829 (2007)

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

[4] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, Phys. Rev. Lett. **87**, 067205 (2001)

[5] J. Schäfer, M. Hoinkis, E. Rotenberg, P. Blaha, and R. Claessen, Phys. Rev. B **72**, 155115 (2005)

[6] V. Yu. Irkhin, A.A. Katanin, and M.I. Katsnelson, Phys. Rev. B **64**, 165107 (2001)

[7] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005)

[8] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A.I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004)

[9] P. Sun and G. Kotliar, Phys. Rev. B **66**, 085120 (2002)

[10] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003)

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

[12] L. Hedin and S. Lundqvist, Solid State Physics **23**, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, New York, 1969)

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

[14] D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953)

[15] R. Sakuma, P. Werner, and F. Aryasetiawan, Phys. Rev. B **88**, 235110 (2013)

[16] T. Yoshida *et al.*, Phys. Rev. B **82**, 085119 (2010)

[17] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and O.K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004)

[18] T. Ayral and O. Parcollet, Phys. Rev. B **92**, 115109 (2015)

[19] F. Nilsson, L. Boehnke, P. Werner, and F. Aryasetiawan, Phys. Rev. Mater. **1**, 043803 (2017)

[20] L. Boehnke, F. Nilsson, F. Aryasetiawan, and P. Werner, Phys. Rev. B **94**, 201106 (2016)

[21] H. Wadati, J. Mravlje, K. Yoshimatsu, H. Kumigashira, M. Oshima, T. Sugiyama, E. Ikenaga, A. Fujimori, A. Georges, A. Radetinac, K.S. Takahashi, M. Kawasaki, and Y. Tokura, Phys. Rev. B **90**, 205131 (2014)

[22] M. Gatti and M. Guzzo, Phys. Rev. B **87**, 155147 (2013)

[23] S. Backes, T.C. Rödel, F. Fortuna *et al.*, Phys. Rev. B **94**, 241110 (2016)

[24] F. Petocchi, F. Nilsson, F. Aryasetiawan, and P. Werner, Phys. Rev. Res. **2**, 013191 (2020)

[25] F. Petocchi, V. Christiansson, F. Nilsson, F. Aryasetiawan, and P. Werner, Phys. Rev. X **10**, 041047 (2020)

[26] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems*, (Dover, Mineola, NY, 2003)

[27] J.W. Negele and H. Orland, *Quantum Many-Particle Systems*, (Westview Press, Boulder, CO, 1998)

13 DMFT and GW+DMFT for Systems out of Equilibrium

Martin Eckstein
University of Erlangen-Nürnberg
91058 Erlangen, Germany

Contents

1 Introduction **2**

2 Keldysh formalism **3**
2.1 The L-shaped time contour 3
2.2 Contour-ordered Green functions 7
2.3 Perturbation theory 10

3 Non-equilibrium DMFT and beyond **16**
3.1 Non-equilibrium DMFT 16
3.2 Impurity solvers 17
3.3 Multi-band DMFT+GW 19

4 Photodoping in Mott and charge-transfer insulators **23**
4.1 Overview 23
4.2 Electronic structure in a photo-excited charge-transfer insulator 24

5 Outlook **27**

1 Introduction

Pump-probe experiments with short laser pulses allow to excite condensed matter systems and to spectroscopically analyze the subsequent dynamics of their microscopic degrees of freedom over a wide range of timescales, from fast electronic processes below one femtosecond to the picosecond dynamics of the crystal lattice and of collective orders like superconductivity. Femtosecond pulses are nowadays available in a large spectrum of photon energies, and many complementary probing techniques have been brought to the ultrafast time domain: For example, time- and angular-resolved photoemission (trARPES) can probe the momentum-dependent electronic structure, time-resolved Xray scattering can reveal the evolution of the crystal lattice, and resonant inelastic Xray scattering (RIXS) based on intense Xray pulses from free electron laser sources can give access to low energy magnetic and orbital excitations. Such experiments have opened up a field of research in condensed matter physics with many facets [1–3]: On the one hand, the hope is that a real-time observation of dynamical processes can reveal the interaction between microscopic degrees of freedom more directly than conventional equilibrium probes. On the other hand, many experiments have shown that an excitation of correlated solids far out of equilibrium can give rise to states that are not simply related to equilibrium phase diagram, which suggests the possibility for a controlled engineering of the collective behavior in solids on ultrafast timescales.

In general, we can distinguish two classes of pathways used to establish new equilibrium states. First, the redistributions of electron populations between different bands and orbitals can trigger a nontrivial dynamics. This redistribution is often loosely referred to as photo-doping, although it does not imply a change of the total electron count (like in chemical doping), but only a change of the electron count in a certain subset of bands. Such photo-excitation processes have been shown to, e.g., induce a metal-insulator transition in VO_2 on timescales below 100 fs [4], lead to the enhancement of excitonic condensates [5], or trigger photo-induced non-thermal phase transitions like the photo-induced melting of antiferromagnetic order [6]. In strongly correlated electron systems, photo-doping can have a direct impact on the electronic structure, due to a shift of the bands, the formation or destruction of coherent quasiparticles, and the ultrafast modification of relevant interaction parameters due to screening, such as the Hubbard U (see Sec. 4.1). The second pathway relies on a direct effect of the laser field while a system is driven. This includes the idea of nonlinear phononics [7], where the anharmonic coupling to strongly driven phonon modes is used to steer the crystal lattice, but also the magnetic and electronic properties of solids. More generally, the period-averaged dynamics of a periodically driven system can be understood in terms of a so-called Floquet Hamiltonian, which can differ significantly from the un-driven Hamiltonian in the presence of nonlinearities. A prominent example of this so-called Floquet engineering, which is more routinely used in cold atom experiments [10], is the generation of topologically nontrivial bands by circularly polarized light in graphene [8, 9], and there are many theoretical proposals towards Floquet engineering of interacting quantum systems, such as a manipulation of superconducting pairing or spin-exchange interactions [2, 3]. In a solid, however, heating processes from the drive are usually inevitable,

so that properties of the driven state rely both on the field-induced modification of the Hamiltonian, and on the effect of non-thermal electron populations similar to photo-doping.
The theoretical description of such non-equilibrium processes must therefore capture the joint evolution of the spectrum (density of states) and the distribution functions, which mutually depend on each other in a correlated system. The Keldysh formalism provides the framework to discuss many-body physics for transient and steady-state non-equilibrium situations. Technically, many-body approaches which are rooted in a diagrammatic formalism can be generalized to the time-domain by replacing imaginary-time arguments by a time variable on a more general real-time contour. In particular, this holds for dynamical mean-field theory (DMFT) [11–14] and its extensions, which in equilibrium present a versatile approach to obtain the electronic structure of correlated electrons. The numerical evaluation of the resulting equations in real-time is however far more challenging, as discussed below. In the context of DMFT, the main challenge is the non-perturbative solution of the quantum impurity problem: Real-time Quantum Monte Carlo techniques have to cope with the notorious dynamical sign problem, while wave-function based techniques such as matrix product state evolutions suffer from the full exponentially large dimension of the relevant Hilbert space (see Sec. 3.2). In spite of that, non-equilibrium DMFT is applied to an increasing set of topics in particular related to the study of photo-induced dynamics in Mott insulators, e.g., to understand light-induced phases such as hidden states with spin and orbital order [15] or superconductivity [16], or the investigation of strong-field phenomena such as strong-field localization and high-harmonic generation [17,18].
The purpose of this lecture is to explain the foundations of non-equilibrium DMFT and some of its extensions, more than giving a summary of recent applications. As such, the lecture has a significant overlap with a previous lecture in this school [19]. Nevertheless, in order to provide a self-contained set of notes we will repeat some of the basic concepts. Compared to the previous lecture, the present notes do not contain details of the solution of integral equations for real-time non-equilibrium Green functions (there is now a detailed technical report on this in the literature [20]), but instead include a chapter on the GW+DMFT formalism out of equilibrium, which is a promising route towards first principles simulations of correlated electron systems out of equilibrium.

2 Keldysh formalism

This chapter recapitulates basis aspects of the Keldysh formalism. For a more detailed review, the reader might consider the textbooks by Kamenev [21], Stefanucci and van Leeuwen [22], or Ref. [14].

2.1 The L-shaped time contour

Time evolution operator

We consider a system which at initial time t_0 is in a given quantum state, or more generally a statistical mixture $\rho = \sum_n w_n |\psi_n\rangle\langle\psi_n|$ of states $|\psi_n\rangle$ with probabilities w_n. In the follow-

ing, we will assume the latter to be the thermal Gibbs ensemble for a given Hamiltonian, $\rho = e^{-\beta H(t_0)}/Z$. For times $t > t_0$, the system then evolves according to a time-dependent Hamiltonian $H(t)$. Taking the expectation value of an observable O in the time-evolved state and averaging over initial states gives

$$\langle O(t)\rangle = \sum_n w_n \langle\psi_n|\,\mathcal{U}(t_0,t)\,O\,\mathcal{U}(t,t_0)|\psi_n\rangle = \mathrm{Tr}\,\rho\mathcal{U}(t_0,t)\,O\,\mathcal{U}(t,t_0)\,, \tag{1}$$

with the time-evolution operator $\mathcal{U}(t,t')$ from time t' to t. For $t > t'$ (forward evolution), the latter is given by ($\hbar = 1$ throughout these notes)

$$\mathcal{U}(t,t') = T_t\, e^{-i\int_{t'}^t d\bar{t}\, H(\bar{t})}, \tag{2}$$

where T_t is the time-ordering operator which brings operators at later time to the left. The action of the time-ordering can be read in two ways: First, by splitting the time-interval $[t',t]$ in $N\to\infty$ infinitesimal time-steps ($t_0 \equiv t', t_1, ..., t_N = t$), over which the Hamiltonian can be taken to be constant, we have

$$T_t\, e^{-i\int_{t'}^t d\bar{t} H(\bar{t})} = T_t \lim_{N\to\infty}\prod_{j=1}^{N} e^{-i(t_j-t_{j-1})H(t_j)} = \lim_{N\to\infty} e^{-i(t_N-t_{N-1})H(t_N)}\cdots e^{-i(t_1-t_0)H(t_1)}, \tag{3}$$

i.e., the time-ordering operator brings the evolution over the infinitesimal time-steps $[t_j, t_{j+1}]$ in the correct order. Alternatively, we can expand the exponent in Eq. (2) in a Taylor series and apply the time-ordering to the individual terms,

$$T_t\, e^{-i\int_{t'}^t d\bar{t} H(\bar{t})} = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{t'}^t dt_1\cdots\int_{t'}^t dt_n\, T_t\Big(H(t_1)\cdots H(t_n)\Big) \tag{4}$$

$$= 1 + \sum_{n=1}^{\infty}(-i)^n\int_{t'}^t dt_1\int_{t'}^{t_1} dt_2\cdots\int_{t'}^{t_{n-1}} dt_n\, H(t_1)\cdots H(t_n). \tag{5}$$

This expression is recognized as iterative solution of the integral equation for the evolution operator, $\mathcal{U}(t,t') = 1 + \int_{t'}^t dt_1 H(t_1)\mathcal{U}(t_1,t')$.

L-shaped time-contour and contour-ordered correlation functions

With an analogous notation, the backward time-evolution can be expressed using the anti time-ordering operator $T_{\bar{t}}$ that brings operators at later time to the right, $\mathcal{U}(t',t) = T_{\bar{t}}\, e^{-i\int_{t'}^t d\bar{t} H(\bar{t})}$ for $t' < t$. Finally, writing the density matrix ρ in terms of the imaginary time evolution with $H(\tau) \equiv H(0)$, the expectation value Eq. (1) becomes

$$\mathrm{Tr}\,\rho\mathcal{U}(t_0,t)\,O\,\mathcal{U}(t,t_0) \;= \frac{1}{Z}\,\mathrm{Tr}\left(T_\tau e^{-\int_0^\beta d\tau H(\tau)}\right)\left(T_{\bar{t}} e^{-i\int_t^0 dt H(t)}\right)O\left(T_t e^{-i\int_0^t dt H(t)}\right). \tag{6}$$

(Here and in the following we set $t_0 = 0$ without loss of generality). The three time-ordered exponentials in this expression can be combined into a single time-ordering along the L-shaped

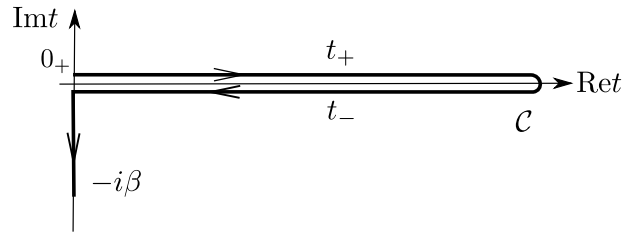


Fig. 1: *The Keldysh contour $\mathcal{C}$, ranging from time 0 to a maximum time t_{max}, back to time 0, and finally to $-i\beta$ on the imaginary-time branch. Times on the upper and lower real-time branch are denoted by t_+ and t_-, respectively. Both t_+ and t_- are real, and the index $\pm$ serves only to distinguish backward and forward time-evolution. The arrows denote the time-ordering along $\mathcal{C}$ from "earlier" to "later" contour times.*

time contour $\mathcal{C}$, which extends from 0 to some maximum time $t_{\max}$ in forward direction, back to 0, and finally to $-i\beta$ in the imaginary direction (see Fig. 1). Hence, Eq. (6) becomes

$$\langle \mathcal{O}(t) \rangle = \frac{1}{Z} \operatorname{Tr} T_{\mathcal{C}}\, e^{-i\int_{\mathcal{C}} dt H(t)} O(t), \tag{7}$$

with the contour-ordering operator

$$T_{\mathcal{C}} A(t)B(t') = \begin{cases} A(t)B(t') & t \text{ later than } t' \text{ on } \mathcal{C} \\ \xi B(t')A(t) & t' \text{ later than } t \text{ on } \mathcal{C}. \end{cases} \tag{8}$$

The sign ξ is -1 if the permutation of A and B involves an odd number of permutations of fermion creation of annihilation operators, and $+1$ otherwise. The sign is not relevant in Eq. (7), but only for the definition of correlation functions below. Equations (7) and (3) are then understood in the same way: splitting the contour in $N \to \infty$ infinitesimal steps ($t_0 \equiv 0_+, t_1, ..., t_N = -i\beta$), with $t_o = t_+$ being the time argument of the operator O, the operator $T_{\mathcal{C}}$ in the expression

$$T_{\mathcal{C}}\Big(e^{-i\int_{\mathcal{C}} dt H(t)} O(t)\Big) = T_{\mathcal{C}}\Big(\prod_{j=1}^{N} e^{-i(t_j - t_{j-1})H(t_j)} O(t_o)\Big) \tag{9}$$

orders the evolution over the infinitesimal timesteps $[t_j, t_{j-1}]$ and O such that Eq. (6) is recovered.

Analogous to Eq. (7), we naturally define multi-point contour-ordered correlation functions, such as the two-point function $\frac{1}{Z} \operatorname{Tr} T_{\mathcal{C}}\, e^{-i\int_{\mathcal{C}} dt H(t)} A(t)B(t')$. In short, we will use the notation

$$\frac{1}{Z} \operatorname{Tr} T_{\mathcal{C}}\, e^{-i\int_{\mathcal{C}} dt H(t)} A(t)B(t')\cdots = \langle T_{\mathcal{C}} \hat{A}(t)\hat{B}(t')\cdots \rangle_H, \tag{10}$$

where $\langle \cdots \rangle_H = \frac{1}{Z} \operatorname{Tr} e^{-\beta H(0)} \cdots$ is the initial state expectation value, and an operator $\hat{O}(t)$ (with a hat) is understood in the Heisenberg picture, $\hat{O}(t) = \mathcal{U}(0,t) O \mathcal{U}(t,0)$. When all time arguments are on the imaginary branch of $\mathcal{C}$, contour-ordered correlation functions are identical to the imaginary-time correlation functions considered in the Matsubara formalism

$$\big\langle T_{\mathcal{C}} \hat{A}(-i\tau)\hat{B}(-i\tau')\cdots \big\rangle_H = \frac{1}{Z} \operatorname{Tr} T_\tau e^{-i\int_0^\beta d\bar{\tau} H(\bar{\tau})} A(\tau)B(\tau'). \tag{11}$$

Keldysh action

The discrete time formulation (9) shows that contour-ordered expectation values can be represented as a path integral, in full in analogy to imaginary-time ordered expectation values. In particular, for a normal-ordered Hamiltonian $H[c^\dagger, c]$ in terms of creation and annihilation operators c and $c^\dagger$ (suppressing for simplicity single-particle labels for spin, orbital, momentum), the coherent state path integral for the partition function (or rather, any generating function for contour-ordered expectation values) is given by

$$Z = \mathrm{Tr}\, T_{\mathcal{C}}\, e^{-i\int_{\mathcal{C}} dt H(t)} = \int \mathcal{D}[\bar{c}, c] e^{iS[\bar{c}(t), c(t)]}, \quad S = \int_{\mathcal{C}} dt \Big(\bar{c}(t) i\partial_t c(t) - H[\bar{c}, c] \Big), \tag{12}$$

with anti-periodic or periodic boundary conditions

$$c(0_+) = \pm c(-i\beta) \tag{13}$$

for bosons or fermions, respectively. This expression is understood in the usual way as the $N \to \infty$ continuum limit of the path integral on a discretized time contour with timesteps $t_0 = 0_+, ..., t_N = -i\beta$, and action

$$S_N = \sum_{n=1}^{N} \delta t_n \Big(i\bar{c}_n \frac{\bar{c}_n - c_{n-1}}{\delta t_n} - H[\bar{c}_n, c_{n-1}] \Big), \quad c_n \equiv c(t_n), \;\; \delta t_n = t_n - t_{n-1}. \tag{14}$$

One should check that the exponent iS reduces to the usual imaginary time action

$$iS = -\int_0^\beta d\tau \left(\bar{c}(\tau) \partial_\tau c(\tau) + H[\bar{c}(\tau), c(\tau)] \right) \tag{15}$$

when $t = -i\tau$ is restricted to the vertical branch of $\mathcal{C}$.
With the action (12), contour-ordered expectation values have a path integral representation

$$\mathrm{Tr}\, T_{\mathcal{C}}\, e^{-i\int_{\mathcal{C}} d\bar{t} H(\bar{t})} A(t) B(t') \cdots = \int \mathcal{D}[\bar{c}, c] e^{iS} A(t) B(t') \cdots, \tag{16}$$

where on the right-hand side operators $O(t) \equiv O[\bar{c}(t), c(t)]$ are understood in their coherent state representation. In the following, we will use the notation

$$\big\langle A(t) B(t') \cdots \big\rangle_S = \frac{1}{Z} \int \mathcal{D}[\bar{c}, c] e^{iS} A(t) B(t') \cdots \tag{17}$$

for contour-ordered expectation values related to an action S.
It should be stressed that in spite of the imaginary contribution from the real-time branches of $\mathcal{C}$, the path integral is convergent for both bosons and fermions. For example, the action $iS = i\int_{\mathcal{C}} dt\, \bar{c}(t)(i\partial_t - \epsilon)c(t)$ for a single bosonic degree of freedom ($H = \epsilon c^\dagger c$) defines a convergent Gaussian integral, other than the path integral for a pure real-time path which would be marginally convergent. With this, field theoretical techniques like the formulation of effective actions and the Hubbard Stratonovich decoupling, as well as Wick's theorem and the derivation of perturbation theory carry over from the imaginary-time Matsubara formalism to the Keldysh formalism on the L-shaped contour $\mathcal{C}$ by a simple replacement of the time variable.

2.2 Contour-ordered Green functions

Definition and relation to other representations

The most important two-point correlation function for the formulation of many-body methods is the Green function

$$G_{jj'}(t,t') = -i\big\langle c_j(t)\bar{c}_{j'}(t')\big\rangle_S = -i\big\langle T_{\mathcal{C}}\,\hat{c}_j(t)\hat{c}^\dagger_{j'}(t')\big\rangle_H. \tag{18}$$

Here j denote single particle labels, such as orbital, spin, and lattice site. The contour-ordered Green function does not only naturally appear in the context of diagrammatic perturbation theory, but it also contains all information on experimentally accessible single-particle observables. Before formulating many-body theory, we therefore explain how to relate the contour-ordered Green functions to the Green functions which are conventionally used for the description of systems in equilibrium.
(i) When both time arguments are on the vertical branch, we have

$$G_{j,j'}(-i\tau,-i\tau') = iG^M_{j,j'}(\tau-\tau'), \tag{19}$$

where $G^M_{j,j'}(\tau-\tau') = -\big\langle T_\tau\hat{c}_j(\tau)\hat{c}^\dagger_{j'}(\tau')\big\rangle_H$ is the imaginary time-ordered Matsubara Green function. This relation will be in particular important below to relate expressions in the Matsubara and Keldysh formalism. (ii) When one time argument is on the lower and one the other on the upper real-time contour, the contour ordering implies a fixed operator ordering. The two possible orderings are

$$G_{j,j'}(t_-,t'_+) = -i\left\langle c_j(t)c^\dagger_{j'}(t')\right\rangle \equiv G^>_{j,j'}(t,t'), \tag{20}$$
$$G_{j,j'}(t_+,t'_-) = -i\xi\big\langle c^\dagger_{j'}(t')c_j(t)\big\rangle \equiv G^<_{j,j'}(t,t'), \tag{21}$$

with $\xi = 1$ ($\xi = -1$) for Bosons (Fermions). These two functions therefore represent the propagation of an additional particle ($G^>$) or a hole ($G^<$) in the time-evolving system. Finally, (iii), of relevance below are the retarded and advanced Greens functions

$$G^R_{j,j'}(t,t') = \Theta(t-t')\big(G^>_{j,j'}(t,t') - G^<_{j,j'}(t,t')\big), \tag{22}$$
$$G^A_{j,j'}(t,t') = \Theta(t'-t)\big(G^<_{j,j'}(t,t') - G^>_{j,j'}(t,t')\big). \tag{23}$$

Spectroscopic interpretation

The Green functions $G^>$ and $G^<$ have a direct interpretation in terms of spectroscopic probes that add or remove electrons, such as tunnelling experiments and (inverse) time-resolved photoemission spectroscopy [23]. In an idealized description of such an experiment, we couple a probe orbital f to a given orbital j of the system, by adding the term $H' = s(t-t_p)e^{i\omega(t-t_p)}f^\dagger c_j + h.c.$ to the Hamiltonian, where the function $s(t-t_p)e^{i\omega(t-t_p)}$ is a time-dependent probe field with probe frequency ω and probe envelope $s(t-t_p)$ centered around some probe time t_p. For electron addition (removal) the probe orbital is assumed to be filled (empty) at time $t = -\infty$, and

the signal $I^>$ ($I^<$) is given by the change $\left|\langle f^\dagger f\rangle_{t=\infty} - \langle f^\dagger f\rangle_{t=-\infty}\right|$ after and before the probe. Straightforward time-dependent perturbation theory leads to the expression

$$I_{jj}^{>(<)}(\omega, t_p) = \int dt\, dt'\; e^{i\omega(t-t')}(\pm i) G_{j,j}^{>(<)}(t_p+t, t_p+t')\; s(t)s(t')^* + \mathcal{O}(s^4). \tag{24}$$

The upper (lower) sign refers to electron addition (removal); Eq. (24) is stated for fermions. This equation will be used below for the interpretation of non-equilibrium Green functions within a spectral representation.

Relation to equilibrium Green functions

In equilibrium, translational invariance in time implies that real-time propagators $X(t,t')$ depend only on the time-difference, and can be represented using the Fourier transform $X(\omega) = \int dt\, e^{i\omega t} X(t,0)$. Within the spectroscopic interpretation with an infinitely long probe pulse ($s(t) = $ const.), Eq. (24) gives $I_j^{>(<)}(\omega) \propto \pm i G_{jj}^{>(<)}(\omega)$, so that the electron and hole propagator can be interpreted in terms of the electron addition and removal spectrum, respectively. A straightforward expansion in (many-body) energy eigenstates (Lehmann representation) shows that the Green functions (20) and (21) are related to a single spectral density $A_{j,j'}(\omega) = -\frac{1}{\pi}\operatorname{Im} G^R_{j,j'}(\omega+i0)$ and the Fermi function $f(\omega)$ (again, the relations are stated for fermions)

$$G^>_{j,j'}(\omega) = \quad 2\pi i A_{j,j'}(\omega) f(-\omega), \tag{25}$$

$$G^<_{j,j'}(\omega) = -2\pi i A_{j,j'}(\omega) f(\quad \omega). \tag{26}$$

The electron removal spectrum is therefore given by the spectral density $A(\omega)$ multiplied with an occupation function $f(\omega)$ that gives the probability to find a state at ω to be occupied, while the electron addition spectrum is given by the same spectrum and the probability to find the state unoccupied, $f(-\omega) = 1 - f(\omega)$. It should be emphasized that this relation holds for any system in thermal equilibrium, also an interacting one. The distribution function $f(\omega)$ is entirely universal, i.e., it does not depend on any details of the Hamiltonian but only on temperature. This universal relation is analogous to the fluctuation-dissipation theorem, which gives a universal relation between response functions and time-dependent autocorrelation functions which is valid for any system in thermal equilibrium.

Moreover, the Lehmann representation expansion shows that the spectral function is related to the Matsubara Green function by the analytical relation

$$G^M_{j,j'}(\tau) = -\int d\omega A_{j,j'}(\omega) e^{-\omega\tau} f(-\omega). \tag{27}$$

This highlights how, in equilibrium, it is in principle sufficient to solve the system on the imaginary branch of the contour. The knowledge of $G(\tau)$ would suffice to determine the spectrum, so that the equilibrium theory is complete in terms of the imaginary time propagators. In practice, however, extracting real frequency information from imaginary time Green functions is an ill-conditioned problem (analytical continuation).

Green functions in a time-evolving state: Wigner representation

In a time-evolving state, time-translational invariance is lost, so that both spectral and occupation functions depend on two time-arguments separately. It is still often convenient to introduce a partial Fourier representation: A symmetric choice is the Wigner transform for a function $X(t,t')$, with average time $t_{av} = (t+t')/2$ and a relative time $t_{\rm rel} = t-t'$, and a Fourier transform with respect to $t_{\rm rel}$,

$$X(t_{av},\omega) = \int dt_{\rm rel}\; e^{i\omega t_{\rm rel}}\, X(t_{av}+t_{\rm rel}/2, t_{av}-t_{\rm rel}/2). \tag{28}$$

For example, the Wigner transforms $G^{<,>}(t,\omega)$ are then naturally related to time-resolved electron addition and removal experiments: With a Gaussian probe profile $S(t) = \exp(-t^2/2\tau^2)$ with duration τ, Eq. (24) gives

$$I_j^{>(<)}(\omega,t) \propto \int d\omega'\, dt'\; (\pm i) G_{jj}^{>(<)}(t+t',\omega+\omega')\; e^{-t'^2/\tau^2} e^{-\omega'^2\tau^2}, \tag{29}$$

i.e., an average of $G_{jj}^{>(<)}(t,\omega)$ over a window of width $\delta t = \tau$ in time and $\delta\omega = 1/\tau$ in frequency which satisfies energy-time uncertainty.[1]
In analogy to Eqs. (25) and (26), we can parametrize $G^{>(<)}(t,\omega)$ in terms of a time-dependent spectral function and a time-dependent occupation function

$$G^{<}(\omega,t) = \quad 2\pi i A(\omega,t)\; F(\omega,t), \tag{30}$$

$$G^{>}(\omega,t) = -2\pi i A(\omega,t)\big(1-F(\omega,t)\big), \tag{31}$$

omitting orbital indices for simplicity (A and F are matrices in orbital space). This representation emphasizes that the real-time Keldysh formalism provides a set of equations which describe the joint evolution of the spectral and distribution functions, as outlined in the introduction. This fact can also be turned into an approximate time-evolution scheme in terms of a quantum Boltzmann equation [24] which is an approximate equation of motion for a time-dependent occupation function, supplemented with certain approximations for the spectrum. One should note, however, that while $\pm iG^{>(<)}(t,\omega)$ is real (hermitian), it is not necessarily positive and cannot directly be interpreted in terms of a phase space probability in time and frequency. Only after averaging over an energy-time uncertainty window we obtain a positive quantity, see Eq. (29) above. This is analogous to the Wigner phase space density $W(x,p)$ as function of position and momentum, which becomes a phase space probability distribution only after suitably averaging over the conjugate variables x and p. Although the Wigner representation is often convenient to represent the time evolution, for numerical evaluation the (t,t')

[1] Note that this does not mean that $G(t,t')$ would be measurable only up to time-frequency uncertainty. Instead, Eq. (24) shows that in principle the full time dependence can be retrieved from experiment. For example, to measure $G^<(t,t')$ in a given time window, we choose an orthonormal basis $\phi_n(t)$ for time-dependent functions in that interval, and expand $-iG^<(t,t') = \sum_{n,n'} \phi_n^*(t) g_{n,n'} \phi_{n'}(t')$. The matrix $g_{n,n'}$ is hermitian and positive definite. A probe pulse $S(t) = \phi_n(t)$ then measures the diagonal components, $I^< = g_{n,n}$. A probe pulse $S(t) = \phi_n(t) + e^{i\varphi}\phi_m(t)$ gives $I^< = g_{n,n} + g_{m,m} + e^{-i\varphi} g_{n,m} + e^{i\varphi} g_{n,m}$, so that off-diagonal components $g_{n,m}$ can be obtained by scanning the phase difference φ [25].

representation is therefore often more suitable, because it allows for a causal time-propagation algorithm even on short timescales where A and F cannot be understood as positive distribution functions.

Causal property of contour-ordered functions

Finally, we note an important property of contour-ordered correlation functions which will be used below. The values of the contour-ordered function $G(t,t')$ with t and t' on different branches of $\mathcal{C}$ are not all independent, because the largest real-time argument can be shifted between the upper to the lower contour branch. For example, for $t' < t$,

$$G(t_+,t'_+) = \frac{1}{Z}\,\mathrm{Tr}\, e^{-\beta H(0)} \underbrace{\mathcal{U}(0,t)\mathcal{U}(t,t_{\max})}_{\mathcal{C}_-}\underbrace{\mathcal{U}(t_{\max},t)c\mathcal{U}(t,t')c^\dagger\mathcal{U}(t',0)}_{\mathcal{C}_+} = G(t_-,t'_+). \quad (32)$$

The brackets indicate the part of the contour-ordered operator $T_\mathcal{C} e^{-i\int_\mathcal{C} dt H(t)}$ along the upper and lower branch, respectively. Because the time-evolution between t and $t_{\max}$ along the upper and lower branch cancel, c can be shifted between the two branches. The redundancy which follows from Eq. (32) can be used to represent the contour-ordered Green functions in terms of fewer components; in particular, real time components can always be represented in terms of $G^<$ and $G^>$, hence the knowledge of the contour-ordered Green function is precisely equivalent to the knowledge of time-dependent spectral and distribution functions.

2.3 Perturbation theory

Equation of motion and inverse Green functions

The action (12) for the noninteracting Hamiltonian $H(t) = \epsilon(t)c^\dagger c$ can be written as a quadratic form $S = \int dtdt'\,\bar{c}(t)\delta_\mathcal{C}(t,t')(i\partial_t - \epsilon)c(t')$. Here $\delta_\mathcal{C}(t,t')$ is the delta-function consistent with the contour integral, i.e., $\int_\mathcal{C} dt'\,\delta_\mathcal{C}(t,t')g(t') = g(t)$ for any function $g(t)$ on $\mathcal{C}$. This quadratic action is to be understood as the continuum limit of a discrete form $S = \sum_{a,a'}\bar{c}_a A_{aa'}c_{a'}$ where a, a' label all orbital and discrete time indices. Gaussian integration for the discrete action yields the moments

$$\langle c_a\bar{c}_{a'}\rangle_S = \frac{1}{Z}\int \mathcal{D}[\bar{c},c]e^{iS}c_a\bar{c}_{a'} = \frac{1}{Z}\int \mathcal{D}[\bar{c},c]e^{-\sum_{b,b'}\bar{c}_b(-iA_{bb'})c_{b'}}c_a\bar{c}_{a'} = (iA^{-1})_{a,a'}. \quad (33)$$

i.e., the Greens function $G_{aa'} = -i\langle c_a\bar{c}_{a'}\rangle_S$ and A are inverse matrices in time. Reinstating the continuum limit, the equation $A\cdot G = 1$ yields the equation of motion for the free Green function G,

$$\int_\mathcal{C} d\bar{t}\,\delta_\mathcal{C}(t,\bar{t})(i\partial_{\bar{t}} - \epsilon)G(\bar{t},t') = \big(i\partial_t - \epsilon(t)\big)G(t,t') = \delta_\mathcal{C}(t,t'). \quad (34)$$

In the derivation of this equation one should note that in contrast to the inverse of the discrete matrix A, the differential equation (34) does not have a unique solution unless a proper boundary condition is specified. This boundary condition is provided by the relation

$$G(0_+,t) = \xi G(-i\beta,t), \quad G(t,0_+) = \xi G(t,-i\beta), \quad (35)$$

with the Bose (Fermi) sign ξ, which follows from the (anti)-periodic boundary condition (13) of c and $\bar{c}$ in the path integral.
From now on we use a continuum notation assuming that all operations take place within the space of (anti)-periodic functions. Multiplication of two contour functions corresponds to convolution along $\mathcal{C}$,

$$[A * B](t,t') = \int_{\mathcal{C}} d\bar{t}\, A(t,\bar{t})B(\bar{t},t'), \tag{36}$$

and the inverse $A^{-1}(t,t')$ of a function $A(t,t')$ is understood as the differential or integral equation $\int_{\mathcal{C}} d\bar{t}\, A^{-1}(t,\bar{t})A(\bar{t},t') = \int_{\mathcal{C}} d\bar{t}\, A(t,\bar{t})A^{-1}(\bar{t},t') = \delta_{\mathcal{C}}(t,t')$ with the boundary condition (35) (additional matrix multiplication in orbital indices implied).

Wick's theorem

Wick's theorem is a consequence of Gaussian integrals, and it therefore holds for contour-ordered functions as well as for imaginary time-ordered functions: For a quadratic action $iS_0 = -\sum_{a,a'} \bar{c}_a(-iA_{aa'})c_{a'}$, the n-point expectation values factorize in two-point functions,

$$\langle c_1 \cdots c_n \bar{c}_{n'} \cdots \bar{c}_{1'} \rangle_{S_0} = \sum_{\pi} \xi^{\pi} \langle c_1 \bar{c}_{\pi(1)'} \rangle_{S_0} \cdots \langle c_n \bar{c}_{\pi(n)'} \rangle_{S_0}. \tag{37}$$

where π runs over all permutations of the numbers $(1,...,n)$, and ξ^{π} is the sign of the permutation for fermions and 1 for bosons, i.e., the right hand side of the expression is the determinant (permanent) of the matrix $M_{ij} = \langle c_j \bar{c}_{j'} \rangle_S$ for fermions (bosons). Wick's theorem therefore directly implies a factorization of contour-ordered correlation functions for a noninteracting Hamiltonian H_0,

$$(-i)^n \big\langle T_{\mathcal{C}} \hat{c}(t_1) \cdots \hat{c}(t_n) \hat{c}^{\dagger}(t'_n) \cdots \hat{c}^{\dagger}(t'_1) \big\rangle_{H_0} = \sum_{\pi} \xi^{\pi} G_0(t_1, t'_{\pi(1)}) \cdots G_0(t_n, t'_{\pi(n)}) \tag{38}$$

in terms of the Green function $G_0(t,t') = -i\big\langle T_{\mathcal{C}} \hat{c}(t) \hat{c}^{\dagger}(t') \big\rangle_{H_0}$.
For illustration, let use the factorization of contour-ordered correlation functions to analyze the function

$$\chi^R(t,t') = -i\Theta(t-t') \left\langle [\hat{O}(t), \hat{O}(t')] \right\rangle_H, \tag{39}$$

which gives the response of the operator O to a time dependent field coupling to O (Kubo relation). In the Matsubara formalism, the response is obtained from the analytical continuation of the imaginary-time ordered correlation function $\chi(\tau) = -\big(\big\langle T_\tau \hat{O}(\tau)\hat{O}(0) \big\rangle - \langle O \rangle^2\big)$; taking a noninteracting fermionic system $H = \sum_a \epsilon_a c_a^{\dagger} c_a$ and a single particle observable $O = \sum_{ab} o_{ab} c_a^{\dagger} c_b$, one would, after some algebra, obtain the convolution of the unoccupied density of states $A_a^{>}(\omega) = A_a(\omega) f(-\omega)$ and the occupied density of states $A_a^{<}(\omega) = A_a(\omega) f(\omega)$,

$$-\frac{1}{\pi} \operatorname{Im} \chi^R(\Omega^+) = \sum_{a,b} |o_{ab}|^2 \int d\omega \left(A_b^{>}(\omega+\Omega) A_a^{<}(\omega) - A_b^{>}(\omega-\Omega) A_a^{<}(\omega) \right). \tag{40}$$

For $\Omega > 0$, the two parts of the expression naturally describe energy absorption and emission when going from an occupied state a to and unoccupied state b. The same expression can be obtained in a relatively straightforward manner from the Keldysh formalism. We start from the contour-ordered correlation function $\chi(t,t') = -i\big(\big\langle T_{\mathcal{C}}\hat{O}(t)\hat{O}(t')\big\rangle - \big\langle\hat{O}(t)\big\rangle\big\langle\hat{O}(t')\big\rangle\big)$, which is related to the response function by

$$\chi^R(t,t') = \Theta(t{-}t')\,\big(\chi(t_-,t'_+) - \chi(t_+,t'_-)\big). \tag{41}$$

Application of Wick's theorem for general contour arguments (t,t') directly gives

$$\chi(t,t') = -i\sum_{ab,cd} o_{ab}o_{cd}\,\big\langle T_{\mathcal{C}}\hat{c}^\dagger_a(t)\hat{c}_d(t')\big\rangle_{H_0}\big\langle T_{\mathcal{C}}\hat{c}_b(t)\hat{c}^\dagger_c(t')\big\rangle_{H_0} = -i\sum_{ab}|o_{ab}|^2\,G_{aa}(t,t')\,G_{bb}(t',t).$$

where in the second step we used that $G_{ab} = \delta_{ab}G_{aa}$. Using $G(t_+,t'_-) = G^<(t,t')$ and $G(t_-,t'_+) = G^>(t,t')$, we can read off

$$\chi(t_-,t'_+) = -i\sum_{ab}|o_{ab}|^2\,G^>_{aa}(t,t')\,G^<_{bb}(t',t)\,,\quad \chi(t_+,t'_-) = -i\sum_{ab}|o_{ab}|^2\,G^<_{aa}(t,t')\,G^>_{bb}(t',t),$$

after which the spectral representation (30) and (31) directly yields Eq. (40) by inverse Fourier transform. This exercise should highlight how working with contour-ordered Green functions in equilibrium is equivalent to and working with imaginary time Green functions and doing an analytical continuation.

Diagrammatic perturbation theory

Because Wick's theorem applies for contour-ordered correlation functions as well as for imaginary time-ordered functions, the construction of diagrammatic perturbation theory does not depend on the time contour, and is formally the same for Matsubara Green functions and contour-ordered Green functions. Green functions are represented in terms of connected diagrams G representing products of the noninteracting Green function G_0 and the interaction V. Because the Keldysh formalism is obtained from the Matsubara formalism by just replacing the time contour, one can translate a diagram for $G^M(\tau{-}\tau')$ in terms of the noninteracting Green function $G_0^M(\tau)$ and the interaction $V(\tau)$ to the Keldysh formalism by using the following rules:

(42)

(1) For all correlation functions ($X \equiv G,\ G_0,\ V$), replace $X^M(\tau_i{-}\tau_j)$ by $-iX(t_i,t_j)$, where the arguments represent either external arguments τ and τ' or internal arguments, which are later integrated over [cf. Eq. (19)].

(2) Replace internal integrals $\displaystyle\int_0^\beta d\tau_j$ by $\displaystyle i\int_{\mathcal{C}} dt_j$.

(3) Replace $\delta(\tau{-}\tau')$ by the contour delta function $-i\delta_{\mathcal{C}}(t,t')$.

For example, let us illustrate these rules for the retarded interaction (omitting spin and site indices for simplicity)

$$S_V^M = \int_0^\beta d\tau d\tau'\,\bar{c}(\tau^+)c(\tau)\,V(\tau{-}\tau')\,\bar{c}(\tau'^+)c(\tau'), \tag{43}$$

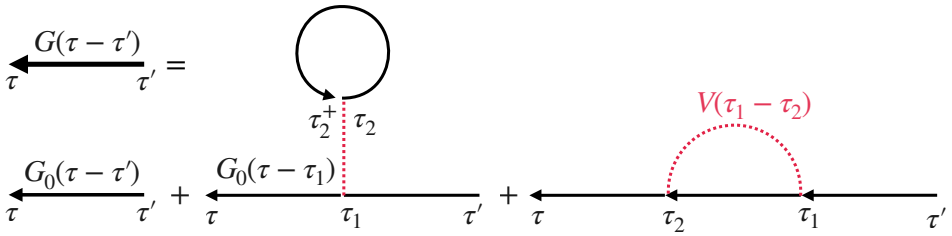


Fig. 2: *Leading diagrams for the expansion of the Green function* $G(\tau) = -\langle T_\tau c(\tau) c^\dagger(0)\rangle$ *in terms of the interaction terms* (43)*, corresponding to the Hartree (left diagram) and Fock (right diagram) self energy.*

in the imaginary-time formalism, where the superscript $+$ means that the time argument is infinitesimally later. The last two diagrams in Fig. 2, obtained by expanding $e^{-S_V^M}$ to leading order in the expression $G(\tau) = -\langle T_\tau c(\tau) c^\dagger\rangle$, reads in the Matsubara formalism

$$\begin{aligned} G(\tau-\tau') = &- \int d\tau_1 d\tau_2 \; G_0(\tau-\tau_1)\, G_0(\tau_1-\tau_2)\, G_0(\tau_2-\tau')\, V(\tau_1-\tau_2) \\ &+ \int d\tau_1 d\tau_2 \; G_0(\tau-\tau_1)\, G_0(\tau_2-\tau_2^+)\, G_0(\tau_1-\tau')\, V(\tau_1-\tau_2), \end{aligned} \tag{44}$$

Hence in the Keldysh formalism we have

$$\begin{aligned} G(t,t') = &\, i \int_{\mathcal{C}} dt_1 dt_2 \; G_0(t,t_1) G_0(t_1,t_2) G_0(t_2,t') V(t_1,t_2) \\ &- i \int_{\mathcal{C}} dt_1 dt_2 \; G_0(t,t_1) G_0(t_2,t_2^+) G_0(t_1,t') V(t_1,t_2), \end{aligned}$$

where, using rules (i) and (ii) above, the overall sign is determined from a factor $(-i)$ on the left-hand side (one propagator) and a factor $(-i)^4 i^2$ (four propagators, two integrals) on the right-hand side. An instantaneous interaction is $V(\tau-\tau') = \delta(\tau-\tau')\, U$ is replaced by $-iV(t,t') = -i\delta_{\mathcal{C}}(t,t')\, U(t)$ using rules (i) and (iii).

As in the Matsubara formalism, we introduce the self-energy $\Sigma(t,t')$, which is the sum of all one-particle irreducible diagrams without external Green functions legs, from which the Green function is obtained by the Dyson equation. The latter is replaced from the Matsubara contour,

$$G(\tau-\tau') = G_0(\tau-\tau') + \int_0^\beta d\tau_1 d\tau_2\, G_0(\tau-\tau_1)\, \Sigma(\tau_1-\tau_2)\, G(\tau_2-\tau') \tag{45}$$

using again rules (i) and (ii), to the integral equation on $\mathcal{C}$,

$$G(t,t') = G_0(t,t') + \int_{\mathcal{C}} dt_1 dt_2\, G_0(t,t_1)\, \Sigma(t_1,t_2)\, G(t_2,t'). \tag{46}$$

For example, the first order self energy (Hartree and Fock) can be read off Eq. (44),

$$\Sigma_H(\tau-\tau') = \delta(\tau-\tau') \int d\tau_1 V(\tau-\tau_1)\, G(\tau_1-\tau_1^+) \tag{47}$$

$$\Sigma_F(\tau-\tau') = -V(\tau-\tau')\, G(\tau,\tau'), \tag{48}$$

where in the first line the argument τ_1^+ is infinitesimally later than τ_1, so that $G(\tau_1-\tau_1^+) = \langle c^\dagger c\rangle \equiv n$ is the density. In the Keldysh formalism, the above rules give

$$\Sigma_H(t,t') = -i\delta_{\mathcal{C}}(t,t')\int_{\mathcal{C}} dt_1 V(t,t_1)\, G(t_1,t_1^+) \tag{49}$$

$$\Sigma_F(t,t') = iV(t,t')\, G(t,t'). \tag{50}$$

In the first line, $-iG(t_1,t_1^+) = n(t)$ is again the density. For an instantaneous interaction $V(\tau-\tau') = \delta(\tau-\tau')\, U$, the Hartree diagram is thus $\Sigma_H(t,t') = \delta(t,t')\, Un(t)$. A more interesting exercise is to rewrite the contribution for a retarded interaction. In this case, for t on the real-branch of $\mathcal{C}$ the contour integral can be rewritten as

$$\int_{\mathcal{C}} dt_1 V(t,t_1)\,(-i)\, G(t_1,t_1^+) = \int_{\mathcal{C}} dt_1 V(t,t_1)\, n(t_1) = \int_{-\infty}^{\infty} dt_1 \big(V(t,t_{1,+})-V(t,t_{1,-})\big) n(t_1),$$

where we have shifted the imaginary branch of $\mathcal{C}$ to $t=-\infty$, and assumed that $V(t,t')$ decays sufficiently fast; in the last step, the contour integral is written explicitly in terms of integrals over the upper and lower branch, with opposite integration direction. Finally, using the causality property (32), which allows to shift the larger of the two arguments t and t_1 freely between the upper/lower contour, the integral is transformed to

$$\int dt_1 n(t_1)\,\Theta(t-t_1)\big(V(t_{1,+},t_-)-V(t_{1,-},t_+)\big) = \int dt_1 V^R(t,t_1)\, n(t_1). \tag{51}$$

The Hartree self energy is therefore given by

$$\Sigma_H(t,t') = \delta_{\mathcal{C}}(t,t')\int dt_1 V^R(t,t_1)\, n(t_1) \equiv \delta_{\mathcal{C}}(t,t')\, V_{\text{mf}}(t), \tag{52}$$

which is a time-dependent mean-field potential $V_{\text{mf}}(t)$; one can see that the retarded interaction is the response function that determines the mean-field potential at time t due to the density at time t_1.

Numerical solution of the Dyson equation

For a given self-energy, the Dyson equation (46) must be solved in order to obtain the Green function. Writing Eq. (46) in short as $G = G_0 + G_0 * \Sigma * G$, it can be reformulated in the form $G^{-1} = G_0^{-1} - \Sigma$, with the inverse operator in time as introduced below Eq. (36). For the noninteracting problem $H_0 = \sum_{ab} c_a^\dagger h_{ab}(t) c_b$, the inverse Green function is $G_0^{-1} = \delta_{\mathcal{C}}(t,t')\big(i\partial_t - h(t)\big)$ [cf. Eq. (34)], so that the Dyson equation can be formulated as an integral differential equation on $\mathcal{C}$,

$$\big(i\partial_t - h(t)\big)\, G(t,t') - \int_{\mathcal{C}} d\bar{t}\, \Sigma(t,\bar{t})\, G(\bar{t},t') = \delta_{\mathcal{C}}(t,t') \tag{53}$$

(omitting orbital and spin indices, i.e., G, h, Σ are understood as matrices in orbital space). Equation (53) and the equivalent Eq. (46) are integral and integro-differential equations on $\mathcal{C}$, which in most cases must be solved numerically (Kadanoff-Baym equations). It is important

to note that one can obtain the solution in a time-stepping manner: With an equidistant time grid ($t_n = n\delta t$), let us denote the nth timeslice of the two-time function $X(t,t')$ to include all elements where one argument is $n\delta t$ and the other is either $m\delta t$ with $m \leq n$ or on the imaginary time branch; timeslice $n = -1$ is simply the Matsubara component of X (both time arguments on the imaginary time branch). Causality then implies that G on timeslice n can be computed from G on all timeslices $m < n$ and Σ on all timeslices $m \leq n$ as an input. Moreover, the self-energy $\Sigma[G]$ must be a causal functional of G, which implies that Σ on timeslice n can be evaluated in terms of G (or G_0) on timeslices $m \leq n$. Hence the typical evolution algorithm will proceed in two steps:

(1) Initial state simulation: Solve the problem on the imaginary time axis, in order to determine the Matsubara components of Σ and G.

(2) Timestepping: For each $n \geq 0$, do the following: (i) Extrapolate Σ from timeslice $n-1$ to timeslice n. (ii) Compute G on timesline n by solving the Dyson equation. (iii) Compute Σ on timeslice n using G on timesline $m \leq n$. Iterate (ii) and (iii) until convergence.

Having in mind that G basically contains the information on spectral and distribution function, one can again see that the Keldysh formalism provides a joint equation of motion for spectral and distribution functions, which is nonlinear (because Σ is a nonlinear functional of G) and non-Markovian, i.e., the propagation depends on all the history. The actual implementation depends on the parametrization of the Green functions. Detailed implementations are described, e.g., in Refs. [26] and [20]. The latter is the basis for the open source code NESSi `http://www.nessi.tuxfamily.org`, which provides a high-order accurate solution of various integral equations on $\mathcal{C}$, as well as basic routines to construct diagrams from contour-ordered Green functions.

In spite of being just a linear equation, the solution of the Dyson equation can provide a severe numerical bottleneck, in particular for multi-orbital simulations. Determination of G for each element of a timeslice n requires a convolution integral over all previous times. Hence the numerical effort to solve the Dyson equation up to time n scales like $\mathcal{O}(n^3)$. More serious is the requirement to keep G and/or Σ at all previous timeslices in memory: For example, for a realistic simulation with $L = 10$ bands in an energy window of $10\,\mathrm{eV}$ over a time window of $1\,\mathrm{ps}$=$1000\,\mathrm{fs}$, we can expect that the timestep δt should resolve the inverse of the largest energy scale, $\delta t \ll \hbar/10\,\mathrm{eV} \approx 0.1\,\mathrm{fs}$, so that the simulation would extend over at least $n = 10^4$ timesteps (probably more). With each element $G(t,t')$ being an $L\times L$ matrix in orbital space, this corresponds to 10^{10} complex numbers or, more than $100\,\mathrm{GB}$. As one of such objects must be kept for each momentum in a suitably discretized Brillouin zone, realistic materials simulations would quickly reach the limit of current day computational capabilities.

Clearly, the equidistant discretization is far from optimal, and there are several directions discussed in the present literature to overcome these limitations. A simple possibility is to implement a systematic truncation of the memory kernel $\Sigma(t,t')$ within the time-stepping procedure [27]. In this way, simulations up to $n = 10^6$ timeslices could be performed within DMFT

simulations [28,29] for certain parameters. In this case, the numerical effort scales linearly with n, and the memory is constant. Another promising route is to explore different compact representations of the two-time functions, such as hierarchical storage formats [30]. Beyond that, also approximate solutions of the integral equations, such as the Generalized Kadanoff-Baym Ansatz [31] or Quantum Boltzmann equations [32, 33] have the potential to extend simulations over a large number of timesteps n at a linear cost $\mathcal{O}(n^1)$ in computation and a constant cost $\mathcal{O}(n^0)$ in memory.

3 Non-equilibrium DMFT and beyond

3.1 Non-equilibrium DMFT

Non-equilibrium DMFT formalism

DMFT approximates only the spatial correlations in a mean-field manner, but accurately treats local temporal fluctuations. The main approximation is the locality of the self-energy, which becomes exact in the limit of infinite coordination number [34]. The formulation of DMFT within the Keldysh and the Matsubara framework differs only by the choice of the time contour, and all arguments regarding the derivation of DMFT, such as the cavity method [11] and power counting arguments for the locality of the self-energy, can be transferred one-to-one from imaginary time to $\mathcal{C}$. We therefore only briefly summarize the formalism. For clarity, the DMFT equations in this section are all stated for the single-band Hubbard model in the spin-symmetric phase; orbital indices can easily be added.
With a spatially local self energy $\Sigma_{ij} = \delta_{ij}\Sigma_{jj}$, lattice Green functions are obtained by solving the Dyson equation

$$G^{-1}_{jj'}(t,t') = \big(i\partial_t + \mu\big)\delta_{\mathcal{C}}(t,t')\delta_{jj'} - \delta_{jj'}\Sigma_{jj}(t,t') - \delta_{\mathcal{C}}(t,t')v_{ij}(t), \tag{54}$$

where $v_{ij}(t)$ are the hopping matrix elements, which contain the external laser fields. To obtain a best local approximation to the self-energy, one can imagine to start from the skeleton expansion $\hat{\Sigma}[G]$ of the self energy, i.e., a diagrammatic expansion in terms of the fully dressed propagator G, and then restrict the expansion for Σ_{jj} to the terms which contain only the local propagator G_{jj}, but are otherwise summed up to all orders in the interaction. Due to the locality of the interaction, these local contributions to the skeleton expansion can be generated by solving an impurity model with a single interacting site and an arbitrary noninteracting environment (bath). Because the functional form of the skeleton expansion does not depend on the noninteracting part of the action, the functional dependence of the impurity self-energy $\Sigma_{\text{imp}}[G_{\text{imp}}]$ on the impurity Green function is the same as the functional dependence of the local lattice self energy $\Sigma_{jj}[G_{jj}]$ on the local Green function. Hence, the impurity self-energy serves as an approximation for the lattice self-energy,

$$\Sigma_{jj'}(t,t') = \delta_{jj'}\Sigma_{\text{imp}}(t,t'), \tag{55}$$

as long as the bath is designed such that the self-consistency condition

$$G_{\text{imp}}(t,t') \stackrel{!}{=} G_{jj}(t,t'). \tag{56}$$

is satisfied. (For simplicity of notation, we also assume translational invariance in space.) For this construction, the impurity model must have the same local interaction as the lattice model, and a general quadratic contribution to the action,

$$S_{\text{imp}} = \int_{\mathcal{C}} dt dt' \Big(\sum_{\sigma} \bar{c}_\sigma(t) \mathcal{G}_\sigma^{-1}(t,t') c_\sigma(t) - U(t)\, \bar{c}_\uparrow(t) c_\uparrow(t) \bar{c}_\downarrow(t) c_\downarrow(t) \delta_{\mathcal{C}}(t,t') \Big), \tag{57}$$

$$\mathcal{G}^{-1}(t,t') = \big(i\partial_t + \mu\big)\delta_{\mathcal{C}}(t,t') - \Delta(t,t'), \tag{58}$$

which describes one site of the lattice embedded in an environment with hybridization function $\Delta(t,t')$. This is the action of a time-dependent Anderson Impurity Hamiltonian. From the action one obtains the interacting impurity Green function

$$G_{\text{imp}}(t,t') = -i\big\langle c(t)\bar{c}(t')\big\rangle_{S_{\text{imp}}}, \tag{59}$$

and the impurity self-energy is set by the impurity Dyson equation

$$G_{\text{imp}}^{-1}(t,t') = \mathcal{G}^{-1}(t,t') - \Sigma_{\text{imp}}(t,t'). \tag{60}$$

Equations (54) through (60) provide the closed set of equations for non-equilibrium DMFT, and the auxiliary quantity $\Delta(t,t')$ can be eliminated when the local lattice Green function G_{jj} equals the corresponding impurity quantity.

The actual implementation of the self-consistency depends on the algorithm used to solve the impurity model. Keeping in mind that the self-consistent equations are integral equations in time, one should bring them to a form that is numerically most stable. For example, G^{-1} is an integral operator which has a singular contribution $\delta_C(t,t')i\partial_t$ on the time diagonal [cf. Eq. (34)], Eq. (60) cannot be directly solved for Σ_{imp} on a given time grid after G_{imp} has been calculated. Instead, one can e.g., extract the function $Z = (i\partial_t + \mu - \Sigma_{\text{imp}})^{-1}$ by solving the linear equation

$$G_{\text{imp}} = Z * (1 + \Delta * G_{\text{imp}}) \tag{61}$$

for Z (this equation is equivalent to (60) and (58)), and then use Z to solve Eq. (54) for the lattice Green function in its equivalent integral form $(1 - Z * v) * G = Z$.

3.2 Impurity solvers

The most challenging part of the DMFT equations is the solution of the auxiliary problem, i.e., the determination of the Green function (59) from the action (57) with a given hybridization function $\Delta(t,t')$. Like in equilibrium DMFT, the action (57) can be mapped to a single impurity Anderson impurity model (SIAM)

$$H_{\text{SIAM}} = U c_\uparrow^\dagger c_\uparrow c_\downarrow^\dagger c_\downarrow + \sum_{p,\sigma} \epsilon_{p,\sigma}\, a_{p\sigma}^\dagger a_{p\sigma} + \sum_{p,\sigma} \big(V_p(t) a_{p,\sigma}^\dagger c_\sigma + h.c. \big), \tag{62}$$

with time-dependent hybridization $V_p(t)$ to bath orbitals labelled by p. In fact, one can see that the hybridization function Δ_{SIAM} representing the discrete bath in the SIAM is given by

$$\Delta_{\text{SIAM}}(t,t') = \sum_p V_p(t) g_p(t,t') V_p^*(t'), \tag{63}$$

with the isolated Green function g_p of a single site of the bath [35]. Hence the parameters V_p can be determined by a simple fit of the analytical expression $\Delta_{\text{SIAM}}(t,t')$ to the hybridization obtained from the DMFT self-consistency. The most efficient way to make use of this discrete representation has been to compute the time-dependent Green functions using a many-body wave function $|\Psi(t)\rangle$ in a matrix product state (MPS) representation [36]. However, while a MPS representation is efficient for ground states which satisfy an area-law entanglement, the time-propagated state can often not be represented efficiently, leading to an exponential increase of computational resources with the simulated time. This has restricted simulations so far to relatively short times [37]. Moreover, it turns out to be necessary to increase the number of bath orbitals in the representation (63) for large time in order to ensure that $\Delta_{\text{SIAM}}(t,t')$ decays as a function of relative time, as required for a continuum bath. To ensure such a decay, one could couple the bath orbitals themselves to simple reservoirs that can be treated within a Lindblad equation. This would increase the computational cost, as it requires a propagation of the many-body density-matrix instead of the wave function, but it would allow to keep the number of auxiliary bath sites fixed for long time simulations. Such an open system representation for the non-equilibrium SIAM has been used so far only to solve DMFT for non-equilibrium steady states [38].

An alternative to the discrete bath representation are Quantum Monte Carlo (QMC) techniques, which can give numerically exact results in equilibrium [39]. A direct extension of QMC algorithms by extending the time contour from the imaginary to the L-shaped contour suffers from a severe phase problem and is restricted to very short times. There are interesting and fundamental problems related to in the short-time dynamics, such as dynamical phase transitions [40], but in order to study the photo-induced dynamics in most materials few hopping times are often not yet sufficient. More promising real-time QMC algorithms are formulated in the spirit of diagrammatic Monte Carlo methods, such as the so called inchworm algorithm [41] which samples the self-energy in the strong coupling expansion, or is vertex generalization, the recently proposed slime-mold algorithm [42]. So far the high numerical cost has prevented an application of these approaches within real-time DMFT simulations, but recent developments in particular towards non-equilibrium steady state DMFT promise that real-time Quantum Monte Carlo may soon be used to simulate the photo-induced dynamics of correlated solids.

Most applications of non-equilibrium DMFT to describe photo-induced dynamics in solids have been based on a perturbative solution of the impurity model, such as adaptations of the iterated perturbation theory [11], or alternatively a systematic expansion in $\Delta(t,t')$ [43]. In this so-called strong-coupling expansion one splits the action S into the local part $S_{at} = \int_{\mathcal{C}} dt H_{at}(t)$, with the atomic Hamiltonian H_{at} that contains the local interaction, and the time-nonlocal terms

S_{n-l}. The latter can have a very general form, such as

$$S_{n-l} = \sum_{\gamma} \int_{\mathcal{C}} dt_1 dt_2 \, \bar{\phi}_\gamma(t) \Delta_\gamma(t,t') \psi_\gamma(t'), \tag{64}$$

where γ sums over all hybridization and interaction channels, and ψ and $\bar{\phi}$ are general operators. In the single impurity Anderson model, e.g., $\gamma \equiv \sigma$, $\bar{\phi}_\gamma \equiv \bar{c}_\sigma$, $\psi_\gamma \equiv c_\sigma$, but the general formulation also allows for inter-orbital hybridizations in multi-band systems, anomalous hybridizations in superconducting systems, $\int_{\mathcal{C}} dt_1 dt_2 \big(c_\uparrow(t)\Delta_{cc}(t,t')c_\downarrow(t') + \bar{c}_\downarrow(t)\Delta_{\bar{c}\bar{c}}(t,t')\bar{c}_\uparrow(t')\big)$, retarded density-density interactions $\sum_{\sigma\sigma'} \int_{\mathcal{C}} dt_1 dt_2 \, n_\sigma(t) V_{\sigma,\sigma'}(t,t') n_\sigma(t')$ (here $\bar{\phi}_{\sigma,\sigma'} \equiv \bar{c}_\sigma c_\sigma$ and $\psi_{\sigma,\sigma'} \equiv \bar{c}_{\sigma'} c_{\sigma'}$), or electron-phonon interactions ($\psi = c_\sigma b, c_\sigma \bar{b}$ etc., acting in a local space of electrons and phonons). The strong-coupling expansion is expected to work well in the Mott phase, as it can be formulated for an arbitrary local part of the Hamiltonian, and it defines a conserving approximation (which respects conservation laws of energy and particle number also in the approximate solution of the time-evolution). In particular the lowest order (the so called non-crossing approximation, NCA) has been used extensively to study the dynamics of Mott-insulators within DMFT. The technical details of this expansion on the Keldysh contour are explained in Refs. [43] and [14] as well as in earlier notes of this lecture series [19].

3.3 Multi-band DMFT+GW

Extensions and Limitations of DMFT

The main limitation of DMFT is the local approximation to the self-energy. As in equilibrium DMFT, non-equilibrium variants have been formulated for both cluster extensions and diagrammatic extensions of DMFT. Cluster extensions are in particular important to keep short range correlations of spin and charge on the scale of few lattice sites, which can have a pronounced influence on the short time dynamics. For example, photo-excited carriers in a Mott insulator, as described in a one band Hubbard model, can loose kinetic energy on the timescale of few inverse hopping times due to the interaction with short-range antiferromagnetic order [44], i.e., the background of spin fluctuations serves a an efficient heat bath for the electronic quasiparticles, which can act much faster than the cooling of photo-excited carriers by coupling to phonons.

Another limitation of plain DMFT is the restriction to local interactions. An important consequence of non-local interactions in the solid is the feedback of the long-range Coulomb interactions on the parameters of the model Hamiltonians due to dynamical screening. For large excitation densities, or interactions involving many bands, one can expect a sizeable renormalization of the Hubbard U via screening, which may even close a Mott gap. The question of screening is also closely related to the determination of parameters for time-dependent models from ab-initio theory. The combination of a non-equilibrium Green function approach such as DMFT with time-dependent density functional theory encounters the double counting problem, which is even harder to solve out of equilibrium than in equilibrium simulations. An interesting perspective is therefore the combination of non-equilibrium GW with DMFT [45, 46], which

can be formulated in a consistent functional language. This formalism is presented in the next chapter, following mainly Ref. [47].

The GW+DMFT formalism

For simplicity we consider a system with several orbitals per unit cell, but only density-density interactions. The most general Hamiltonian is

$$H = \sum_{j,j'} \sum_{a,b} J_{j,a;j',a'} c^\dagger_{j,a} c_{j',a'} + \frac{1}{2} \sum_{j,j'} \sum_{a,a'} v_{j,a;j',a'} n_{j,a} n_{j',a'} \equiv H_0 + H_V \tag{65}$$

where j and j' label the unit cells, and a, a' are combined spin/orbital indices, J is the hopping matrix, and v the interaction. The action is therefore

$$S = \int d1 d2 \; \bar{c}(1)\, G_0^{-1}(1,2)\, c(2) - \frac{1}{2} \int d1 d2 \; n(1)\, V(1,2)\, n(2) \equiv S_0 + S_V, \tag{66}$$

where we have introduced a combined notation $1 \equiv (t_1, j_1, a_1)$ for time, space and orbital indices, with $\int d1 = \int_{\mathcal{C}} dt_1 \sum_{j_1,a_1}$, and $\delta(1,2) = \delta_{\mathcal{C}}(t_1,t_2)\delta_{j_1,j_2}\delta_{a_1,a_2}$, as well as the interaction matrix $V(1,2) = v_{j_1,a_1;j_2,a_2}\, \delta_{\mathcal{C}}(t_1,t_2)$, and the noninteracting Green function $G_0^{-1}(1,2) = (i\partial_{t_1} + \mu)\, \delta(1,2) - J(1,2)$. The first step is to decouple the interaction using a Hubbard-Stratonovich transformation using a real field $\varphi(1)$ and the Gaussian identity

$$e^{iS_V} = \frac{1}{Z_\varphi} \int \mathcal{D}[\varphi]\, e^{iS_\varphi} e^{-\int d1 \varphi(1) n(1)}, \quad S_\varphi = -\frac{1}{2} \int d1 d2 \; \varphi(1)\, V^{-1}(1,2)\, \varphi(2). \tag{67}$$

Here V^{-1} is inverse in site, orbital, and time, $\int d\bar{1}\, V^{-1}(1,\bar{1})\, V(\bar{1},1') = \delta(1,1')$. It is worthwhile to note that the Gaussian integral over φ in Eq. (67) is convergent: Using the rules Eq. (42), the action iS_φ reduces to $iS_\varphi = -\frac{1}{2}\int_0^\beta d\tau \sum_{j,a,j',a'} \varphi_{j,a}(\tau)(v^{-1})_{j,a;j',a'}\varphi_{j',a'}(\tau)$ on the Matsubara branch, which defines a convergent integral if the matrix v is positive definite (repulsive interaction).

After the Hubbard-Stratonovich decoupling, the system is described by electrons interacting with a fluctuating bosonic field φ, with the effective action

$$S_{\text{eff}}[c,\bar{c},\varphi] = S_\varphi[\varphi] + S_0[\bar{c},c] + i \int d1 \; \varphi(1)\, n(1), \tag{68}$$

in the sense that the partition function is $Z_{\text{eff}} = \int \mathcal{D}[\bar{c},c] \int \mathcal{D}[\varphi] e^{iS_{\text{eff}}[c,\bar{c},\varphi]}$. In the new representation, we then introduce the propagators W of the interaction as well as the electronic density correlation functions χ,

$$W(1,1') = \quad i\big\langle \varphi(1)\varphi(1') \big\rangle^{\text{con}}_{S_{\text{eff}}} \tag{69}$$

$$\chi(1,1') = -i\big\langle n(1)n(1') \big\rangle^{\text{con}}_{S_{\text{eff}}} \tag{70}$$

with the connected correlation function $\langle AB\rangle^{\text{con}} = \langle AB\rangle - \langle A\rangle\langle B\rangle$. The Hubbard Stratonovich transformation implies the exact relation

$$\langle \varphi(1) \rangle = i \int d\bar{1} \; V(1,\bar{1}) \langle n(\bar{1}) \rangle \equiv i V_{\text{mf}}(1) \tag{71}$$

$$W(1,1') = V(1,1') + \int d\bar{1} d\bar{2} \; V(1,\bar{1})\, \chi(\bar{1},\bar{2})\, V(\bar{2},1'), \tag{72}$$

which shows that $\varphi(1)$ takes the role of a fluctuating mean field, and its propagator W is the fully screened interaction. We can now use the action (68) and formulate the diagrammatic perturbation theory for the electronic (G) and bosonic (W) propagators in terms of the electron-boson interaction $e^{iS_{n-\varphi}} = e^{-\int d1 n(1)\varphi(1)}$, where the noninteracting propagators are the noninteracting Green function G_0 and the bare interaction V. We introduce self-energies $\Sigma[G,W]$ and $\Pi[G,W]$ for the electrons and bosons,

$$W^{-1} = V^{-1} - \Pi[W,G], \quad G^{-1} = G_0^{-1} - \Sigma[W,G]. \tag{73}$$

(Π is called the polarization function). The perturbation theory is formulated in terms of a self-consistent skeleton expansion, i.e., the propagators in the self-energies do not contain self-energy insertions. From the action, the lowest-order approximation would be [cf. Eqs. (47) and (48)]

$$\Sigma(1,1') = \quad iG(1,1')W(1,1') - i\langle\varphi(1)\rangle\,\delta(1,1') \equiv \Sigma_{GW}(1,1') - i\langle\varphi(1)\rangle\,\delta(1,1'), \tag{74}$$

$$\Pi(1,1') = -iG(1,1')\,G(1',1) \equiv \Pi_{GW}(1,1'). \tag{75}$$

The second term in Σ, together with the exact relation (71), is just the Hartree mean-field potential $-i\langle\varphi(1)\rangle\,\delta(1,1') = V_{\text{mf}}(1)\,\delta(1,1')$. The terms Σ_{GW} and Π_{GW} define the celebrated GW approximation, which was developed by Hedin [48]: In this approximation the electronic self-energy is expanded to leading order in the fully screened interaction, and the polarization is approximated by the Lindhard function expressed in terms of the renormalized electron propagators. While this approach is understood to be an accurate approximation in electronic structure calculations of weakly correlated materials, (superior to standard density functional approximations), it fails to describe the effect of strong local correlations, such as the Mott transition. To overcome this limitation, one can therefore combine the idea of DMFT with the GW approximation [45, 46]:

Let us assume that we can identify a certain subset C of orbitals which are considered to be strongly correlated, so that within that subset we would like to use a more accurate approximation for the self-energy and the polarization. In the spirit of DMFT, one can aim to add on top of the self-energy and polarization defined in Eqs. (74) and (75) a nonperturbative self-energy Σ_{corr} and polarization Π_{corr} which (i), act only on the correlated orbitals, (ii) are local in space, and (iii) are given by all contributions to the skeleton expansion which contain only the space-local propagators of the correlated manifold at the same site. In other words, we set

$$(\Sigma_{\text{corr}})_{j,a';j,a'}(t,t') = \begin{cases} \delta_{j,j'}\Sigma_{\text{loc,cor}}[G_{\text{loc,cor}}, W_{\text{loc,cor}}]_{a';a'}(t,t') & a,a' \in C \\ 0 & \text{otherwise} \end{cases}, \tag{76}$$

and analogous for Π_{corr}, where here we define the local propagators in the correlated subspace $a, a' \in C$ as

$$(G_{\text{loc,cor}})_{a,a'}(t,t') = G_{j,a;ja'}(t,t'), \quad (W_{\text{loc,cor}})_{a,a'}(t,t') = W_{j,a;ja'}(t,t'). \tag{77}$$

Like in conventional DMFT, we then introduce an electron-boson impurity model as an auxiliary device to evaluate these local contributions. The impurity action is

$$S_{\text{imp}} = \int_{\mathcal{C}} dtdt' \sum_{a,a'\in C} \bar{c}_a(t)\mathcal{G}^{-1}_{a,a'}(t,t')c_{a'}(t') - \frac{1}{2}\sum_{a,a'\in C}\int_{\mathcal{C}} dtdt'\, \varphi_a(t)\mathcal{W}^{-1}_{a,a'}(t,t')\varphi_{a'}(t') + i\sum_{a\in C}\int_{\mathcal{C}} dt\, \varphi_a(t)h_a(t)(t) + i\sum_{a\in C}\int_{\mathcal{C}} dt\, \varphi_a(t)n_a(t). \tag{78}$$

Here $\mathcal{G}$, $\mathcal{W}$ and h essentially define the noninteracting propagators of electrons and bosons in the impurity model; they are free parameters, which will be fixed by self-consistency conditions below. The last term is the electron-boson interaction, which is the same as in the lattice action (68). From the impurity model we measure the impurity correlation functions $\big\langle\varphi_a(t)\big\rangle_{\text{imp}}$, $(W_{\text{imp}})_{a,a'}(t,t') = i\big\langle\varphi_a(t)\varphi_{a'}(t')\big\rangle^{\text{con}}_{\text{imp}}$ and $(G_{\text{imp}})_{a,a'}(t,t') = -i\big\langle c_a(t)\bar{c}_{a'}(t')\big\rangle_{\text{imp}}$, and use the Dyson equation (understood as matrix equation in time and in the space C)

$$W^{-1}_{\text{imp}} = \mathcal{W}^{-1} - \Pi_{\text{imp}}, \tag{79}$$

$$G^{-1}_{\text{imp}} = \mathcal{G}^{-1} - \Big(\Sigma_{\text{imp}}(t,t') - i\delta(1,1')\langle\varphi(1)\rangle_{\text{imp}}\Big). \tag{80}$$

Here, we have treated the Hartree self-energy separately. If then we request a self-consistency between impurity quantities and local lattice quantities on the correlated subspace,

$$G_{\text{imp}} \stackrel{!}{=} G_{\text{loc,cor}}, \quad W_{\text{imp}} \stackrel{!}{=} W_{\text{loc,cor}}, \quad \big\langle\varphi_a(t)\big\rangle_{\text{imp}} = \big\langle\varphi_{j,a}(t)\big\rangle_{\text{imp}} \text{ for } a \in C, \tag{81}$$

it is guarantied that Σ_{imp} and Π_{imp} provide the sum of all local diagrams. Finally adding the correlated and GW self-energies gives the full approximation to the lattice self-energy,

$$\Sigma(1,1') = \Big[iG(1,1')W(1,1') - iG_{\text{loc,cor}}(1,1')W_{\text{loc,cor}}(1,1')\Big] + \Sigma_{\text{imp}}(1,1') - i\delta(1,1')\langle\varphi(1)\rangle, \tag{82}$$

$$\Pi(1,1') = \Big[-iG(1,1')G(1,1') + iG_{\text{loc,cor}}(1,1')G_{\text{loc,cor}}(1,1')\Big] + \Pi_{\text{imp}}. \tag{83}$$

Here in the GW self-energy and polarization we have subtracted the local contribution, which is already contained in Σ_{imp} and Π_{imp} respectively. Moreover, from the impurity model we take only the contribution beyond the Hartree mean-field term, because $\langle\varphi(1)\rangle$ is the same in the impurity and lattice due to the self-consistency.

This completes the description of the theory: One can start with a guess for the free parameters $\mathcal{W}$, $\mathcal{G}$ [or Δ], and h, solve the impurity model to extract Σ_{imp} and Π_{imp}, calculate the lattice self energies and polarization from Eqs. (82) and (83), from that calculate the lattice G and W, extract their local contributions, use the self-consistency condition to obtain a new guess for $\mathcal{W}$, $\mathcal{G}$, and h. In the actual implementation [47], there are some differences: First, the impurity model is transformed to a purely electronic model by integrating out the phonons,

$$S_{\text{imp}} = \int_{\mathcal{C}} dtdt' \sum_{a,a'\in C} \bar{c}_a(t)\tilde{\mathcal{G}}^{-1}_{a,a'}(t,t')c_{a'}(t') - \frac{1}{2}\sum_{a,a'\in C}\int_{\mathcal{C}} dtdt'\, n_a(t)\mathcal{W}_{a,a'}(t,t')n_{a'}(t'),$$

with the exact relations (cf. Eqs. (71) and (72))

$$W_{\text{imp}}(t,t') = \mathcal{W} + \mathcal{W}\chi_{\text{imp}}\mathcal{W}, \quad \langle\varphi\rangle_{\text{imp}} = i\int d1'\,\mathcal{W}(1,1')\langle n(1')\rangle. \tag{84}$$

Hence $\langle\varphi\rangle_{\text{imp}}$ and W_{imp} can be obtained by measuring χ_{imp} and $\langle n\rangle_{\text{imp}}$. This can be done, e.g., with the strong coupling equation as described around Eq. (64). Moreover, in the real-time Keldysh formalism, we again avoid directly solving for the inverse operators in discrete time (as in Eq. (79)), but transform all equations to well-behaved integral equations (see also the discussion around Eq. (61)). Finally, the bosonic propagators have time-local contributions, such as an instantaneous interaction $V\delta_C(t,t')$, which must be kept separately in the real time formalism. For a detailed discussion of these technical points, the reader is referred to Ref. [47].

4 Photodoping in Mott and charge-transfer insulators

4.1 Overview

In a Mott or charge-transfer insulator, electrons in a partially filled band get localized due to the Coulomb interaction. They still keep active spin and orbital degrees of freedom, leading to a large variety of magnetically and orbitally ordered low temperature phases, which can turn into unconventional metallic states or superconductors upon doping. DMFT and its extensions have been instrumental in the understanding of this physics, and non-equilibrium DMFT is therefore a natural starting point to investigate the non-equilibrium phenomena induced in Mott insulators driven by strong laser fields. The most straightforward way to excite the Mott insulator is to use a short laser pulse that is resonant to the charge gap. This will impulsively create mobile charge carriers, such as doubly occupied sites ("doublons") and holes in a single band Mott insulator, and will be followed by a sequence of dynamical processes:

Immediate response of the electronic structure: In a correlated electron system, the redistribution of charges between different orbitals leads to an almost immediate response of the electronic spectrum. Since the kinetic energy of photo-excited charges after an impulsive excitation is often high, the electrons will not form coherent quasiparticles on the shortest times, but the metallic character of the photo-excited state will become manifest in the formation of incoherent spectral weight within the Mott or charge-transfer gap. In a multi-band system, the inter-band Coulomb interaction can give rise to band shifts, and moreover the mobile charges will change the dynamic screening environment and therefore also interactions like the Hubbard U in the valence band. Below, we will discuss the change of the Hubbard U in a photo-excited charge-transfer insulator, which has been predicted theoretically [49,50] and recently measured using time-resolved Xray absorption [51].

Thermalization: Thermalization of a system implies that the properties of the system eventually approach the properties of a system in equilibrium, at a temperature T_f such that the total energy equals the thermal energy expectation value $E_{\text{th}}(T_f)$ at temperature T_f, $\langle H(t)\rangle \equiv E_{\text{tot}} \stackrel{!}{=} E_{\text{th}}(T_f)$. In a correlated electron system, one can expect that the electronic system can thermalize before passing on a substantial fraction of its energy to other degrees of freedom (phonons),

leading to the formation of a correlated electron liquid at high temperature, which subsequently cools down by coupling to the lattice. While such a "hot electron picture" is quite obvious and has been employed ever since in the investigation of photo-induced solids, the correlated nature of the state brings in few nontrivial aspects: The high-temperature state, with temperature of the order of the electronic band width, can have peculiar spectral and transport properties, such as bad metallic behavior [52, 28]. In equilibrium, the same state would not be accessible because also the lattice would have to be heated, and such high temperature correlated fermion liquids are therefore rather accessible in cold atom quantum simulators [53]. The understanding of their dynamics at short times and high energies offers the opportunity for a detailed and systematic comparison of theory and experiment.

Prethermal photo-doped states: In the presence of the Mott or charge-transfer gap, full thermalization can be inhibited for relatively long times. The thermalization of a single-band photo-doped Mott insulator has been analyzed within DMFT in Ref. [54], by calculating the double occupancy and the spectral functions after the pulse. One finds that the double occupancy increases during the pulse, and subsequently shows an exponential relaxation to a new final value, from which the thermalization time can be extracted. This timescale strongly depends on U, and ranges from few hopping times (few femtoseconds for a typical eV bandwidth) in the correlated metal, to thousands of hopping times in the Mott phase. The fast thermalization in a small gap Mott insulator has been experimentally investigated in 1T-TaS_2 [55]. At large U, instead, the double occupancy D can be viewed as an almost conserved quantity, so that one can expect the electronic state to be described as a quasi-thermal state with a thermodynamic variable D in addition to the total particle number N and the total energy E. In fact, simulations for the one-band Hubbard model show that the system quickly establishes a distribution function with a universal form, given by a Fermi function with separate chemical potentials in the upper and lower Hubbard band [28]. This universal form hints at a description in terms of few additional slow variables. Upon energy transfer to the lattice (or to spins [44] and other degrees of freedom), the state may acquire a low temperature but still have a substantial fraction of additional doubly occupied sites and holes, i.e, it is a cold correlated liquid of spins (singly occupied sites), holes and doublons. The quasi-steady properties of such photo-doped states may support phases different from those in the equilibrium phase diagram. They can be explored by taking the equilibrium states of a suitable model in which D is turned into an exactly conserved quantity (for the Hubbard model, this is a generalized the t-J model, in which charge recombination processes are projected out [57, 16]). Alternatively, one can try to work within the Hubbard model and establish the quasi-steady state as a true steady state under the application of suitable reservoirs [56]. In both cases, for the Hubbard model, η-pairing superconductivity is found for strong photo-excitation [57, 16].

4.2 Electronic structure in a photo-excited charge-transfer insulator

Finally, we briefly review the first application of the non-equilibrium GW+DMFT to a realistic material simulation [49]. We focus on a three-band model for a charge-transfer insulator. It

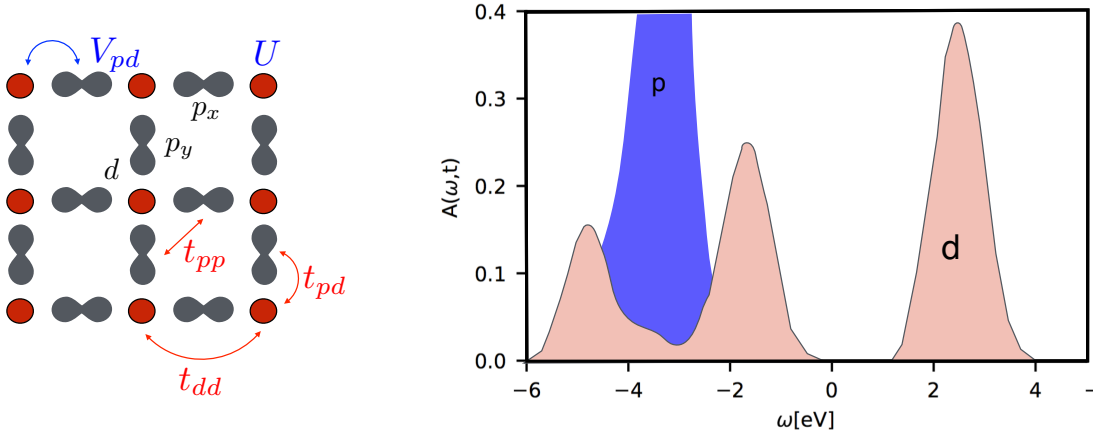


Fig. 3: *Left: 3-band Emery model for a charge-transfer insulator. Right: p and d spectral functions in equilibrium, as obtained with GW+DMFT. Adapted from Ref. [49]*

.

describes a two-dimensional cubic lattice with two ligand orbitals p_x and p_y and one d orbital per unit cell (see Fig. 3, left). We take into account a nearest neighbor p-d hopping $t_{pd} = 0.4\,\text{eV}$, a direct d-d hopping $t_{dd} = -0.1\,\text{eV}$, a charge-transfer energy (energy difference $\epsilon_p - \epsilon_d = -2\,\text{eV}$), a inter-site p-d density-density interaction $V_{pd} = 2\,\text{eV}$, and a local Hubbard interaction on the d orbitals, $U_{dd} = 8\,\text{eV}$. The parameters are determined to best match the spectrum for the cuprate LSCO. With 5 electrons per unit cell, the system has a nominally half-filled d band, which is split into upper and lower Hubbard band due to the large U_{dd}. The model is solved in and out of equilibrium with the GW+DMFT formalism (see Sec. 3.2), using the non-crossing approximation as an impurity solver (Sec. 3.3). The local spectral function in equilibrium ($\beta = 5$) is shown in Fig. 3, right panel; one can clearly distinguish the upper and lower Hubbard band in the d-related spectrum, where the lower Hubbard band hybridizes with the p bands.

We now simulate the time-dependent spectral and occupation functions during and after the excitation of the system with a short electric field pulse. The pulse couples to the model via both a Peierls phase and dipolar matrix elements (for details see [47]). The electric field has the form

$$E(t) = E_0 e^{-4.6(t-t_0)^2/t_0^2} \sin(\Omega(t-t_0)), \tag{85}$$

with a frequency Ω that is varied below mostly resonant to the charge-transfer gap, and a duration of roughly two cycles ($t_0 = 4\pi/\Omega$). It is polarized along the (11) direction, and the amplitude is adapted to achieve a certain excitation density.

Figure 4a) shows the spectrum shortly after an excitation with about 5% photo-doping (measured by the change of the double occupancy after the pulse). Time-resolved spectra are obtained from the Wigner transform $A_\alpha(\omega, t) = -\frac{1}{\pi} \operatorname{Im} G^R_\alpha(t, \omega + i0)$ for $\alpha = p, d$. One can clearly see a shift and broadening of all bands. To further elucidate the origin of these band structure changes, one can switch off the nonlocal GW self-energies and polarizations, and keep only the Hartree and Fock self-energies (HF+DMFT) in the simulation. In this case, the band shifts on the d band can mainly be understood as a consequence of the mean-field p-d interaction, which

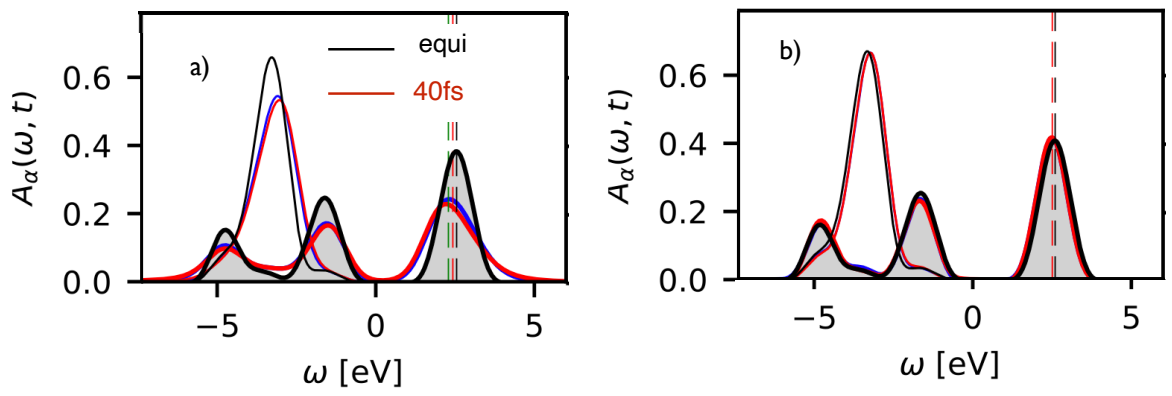


Fig. 4: *a) Spectra after photo-excitation of the Emery model in equilibrium (black lines) and shortly after the excitation ($40\,fs$, red). The pulse frequency $\Omega = 6\,eV$ is resonant to an excitation across the gap, and the excitation density (change in the doubly occupancy on the d-sites) is about 5%. One finds a significant broadening and shift of the bands. b) Same parameters, but taking only the static Hartree and Fock diagrams beyond DMFT into account. Adapted from Ref. [49]*

.

implies a shift of the d level given by $\delta\epsilon_d = V_{pd}\,\delta\langle n_d\rangle$ depending on the change $\delta\langle n_d\rangle$ of the p occupation. The result is shown in Fig. 4b). The center of mass shift of the upper Hubbard band turns out to be about a factor two larger in the GW+DMFT simulation as compared to the HF+DMFT simulation. This shows a significant quantitative role of the dynamic screening processes due to the long-range interaction, which are captured by the GW formalism. In the language of GW+DMFT, one can understand these screening processes in turns of a reduction of the onsite interaction U in the photo-excited state. Similar band shifts after photo-excitation have been reported for LSCO using XAS from a core level [51], which have also been interpreted as a dynamic screening of the Hubbard U. Because of the strong core-valence excitonic character of the final state in XAS, XAS cannot directly be linked to the single-particle spectral function presented in the present study. To get a quantitative interpretation of the experiment it would be interesting to compute the XAS within GW+DMFT, along the lines presented in Ref. [58].

In addition to the quantitative relevance of the dynamic screening processes, we observe that the spectrum is strongly broadened in the GW+DMFT simulation, while band shifts are more or less rigid within HF+DMFT. The reason is that the interaction of electrons with the nonlocal dynamic charge fluctuations, which is captured by the GW diagrams, opens a new scattering channel in the photo-doped state. For the same reason, also the dynamics of the occupation functions is very different in the two cases (not shown here, see Ref. [49]): The interaction of electrons with charge fluctuations leads to a rapid relaxation of the photo-doped electron and hole distributions to the bottom of the upper Hubbard band and the top of the hybrid p/d bands, respectively, whereas this relaxation would require the coupling to an additional bath of phonons in the HF+DMFT simulation. Hence, we see that keeping the dynamical nonlocal fluctuations is essential for an accurate description of photo-doped states.

5 Outlook

In these notes we have reviewed the theoretical basis for a description of correlated electron systems out of equilibrium using the non-equilibrium extension of DMFT and GW+DMFT. While such simulations have been used to understand many aspects of photo-exited states in solids, a desirable future development will be to bring the non-equilibrium formalism to a similar level of quantitative predictive power as the equilibrium DMFT formalism. The main challenge in this direction is the solution of the impurity problem. There are currently no non-perturbative impurity solvers which work at long times. While low-order variants of the strong-coupling expansion, such as NCA, can be applied very flexibly within the Mott phase, they become increasingly inaccurate for correlated metallic phases. An interesting perspective are given by the application of the non-perturbative methods (QMC [41] and Hamiltonian-based solvers [35]). In the near future, it might in particular become feasible to use these non-perturbative techniques for a study of non-equilibrium steady states, which then can provide an effective description of long-lived photo-excited phases [16, 56], or be elevated to study slow time-evolution using quantum Boltzmann equations based on steady-state DMFT [32].

Acknowledgements

I acknowledge funding from the ERC via starting grant No. 716648.

References

[1] C. Giannetti, M. Capone, D. Fausti, M. Fabrizio, F. Parmigiani, and D. Mihailovic, Adv. in Phys. **65**, 58 (2016)

[2] D. Basov, R.D. Averitt, and D. Hsieh, Nat. Mater. **16**, 1077 (2017)

[3] A. de la Torre, D.M. Kennes, M. Claassen, S. Gerber, J.W. McIver, and M.A. Sentef, Rev. Mod. Phys. **93**, 041002 (2021)

[4] D. Wegkamp *et al.*, Phys. Rev. Lett. **113**, 216401 (2014)

[5] S. Mor *et al.*, Phys. Rev. Lett. **119**, 086401 (2017)

[6] A. de la Torre *et al.*, Comm. Phys. **5**, 35 (2022)

[7] M. Först, C. Manzoni, S. Kaiser, Y. Tomioka, Y. Tokura, R. Merlin, and A. Cavalleri, Nat. Phys. **7**, 854 (2011)

[8] T. Oka and H. Aoki, Phys. Rev. B **79**, 081406(R) (2009)

[9] J.W. McIver, B. Schulte, F.-U. Stein, T. Matsuyama, G. Jotzu, G. Meier, and A. Cavalleri, Nat. Phys. **16**, 38 (2020)

[10] T. Oka and S. Kitamura, Annu. Rev. Condens. Matter Phys. **10**, 387 (2019)

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

[12] P. Schmidt, and H. Monien, arXiv:cond-mat/0202046

[13] J.K. Freericks, V.M. Turkowski, and V. Zlatić, Phys. Rev. Lett. **97**, 266408 (2006)

[14] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, and Ph. Werner, Rev. Mod. Phys. **86**, 779 (2014)

[15] J. Li, H.U.R. Strand, P. Werner, and M. Eckstein, Nat. Comm. **9**, 4581 (2018)

[16] J. Li, D. Golez, Ph. Werner, and M. Eckstein, Phys. Rev. B **102**, 165136 (2020)

[17] Y. Murakami, M. Eckstein, and Ph. Werner, Phys. Rev. Lett. **121**, 057405 (2018)

[18] N. Dasari, J. Li, Ph. Werner and M. Eckstein, Phys. Rev. B **101**, 161107 (2020)

[19] M. Eckstein: Ch. 15 in E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.): *DMFT: From Infinite Dimensions to Real Materials*
Modeling and Simulation vol. 8 (Forschungszentrum Jülich, 2018)
`http://www.cond-mat.de/events/correl18`

[20] M. Schüler, D. Golez, Y. Murakami, N. Bittner, A. Hermann, H.U.R. Strand, Ph. Werner and M. Eckstein, Comp. Phys. Commun. **257**, 107484 (2020)

[21] A. Kamenev: *Field Theory of Non-Equilibrium Systems* (Cambridge University Press, 2011)

[22] G. Stefaniccu and R. v. Leeuwen: *Nonequilibrium Many-Body Theory of Quantum Systems* (Cambridge University Press, 2013)

[23] J.K. Freericks, H.R. Krishnamurthy, and Th. Pruschke, Phys. Rev. Lett. **102**, 136401 (2009)

[24] H. Haug, and A. Jauho: *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, 2008)

[25] F. Randi, D. Fausti, and M. Eckstein, Phys. Rev. B **95**, 115132 (2017)

[26] K. Balzer and M. Bonitz: *Nonequilibrium Green's Functions Approach to Inhomogeneous Systems* (Springer, 2008)

[27] Ch. Stahl, N. Dasari, J. Li, A. Picano, Ph. Werner, M. Eckstein, Phys. Rev. B **105**, 115146 (2022)

[28] N. Dasari, J. Li, Ph. Werner, and M. Eckstein, Phys. Rev. B **103**, 201116 (2021)

[29] A. Picano and M. Eckstein, Phys. Rev. B **103**, 165118 (2021)

[30] J. Kaye and D. Golež, SciPost Phys. **10**, 91 (2021)

[31] N. Schlünzen, J.-P. Joost, and M. Bonitz, Phys. Rev. Lett. **124**, 076601 (2020)

[32] A. Picano, J. Li, and M. Eckstein, Phys. Rev. B **104**, 085108 (2021)

[33] M. Wais, M. Eckstein, R. Fischer, P. Werner, M. Battiato, and K. Held, arXiv:1806.02570

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

[35] Ch. Gramsch, K. Balzer, M. Eckstein, and M. Kollar, Phys. Rev. B **88**, 235106 (2013)

[36] F.A. Wolf, I.P. McCulloch, and U. Schollwöck, Phys. Rev. B **90**, 235131 (2014)

[37] K. Balzer, F.A. Wolf, I.P. McCulloch, Ph. Werner, and M. Eckstein, Phys. Rev. X **5**, 031039 (2015)

[38] E. Arrigoni, M. Knap, and W. von der Linden, Phys. Rev. Lett. **110**, 086403 (2013)

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

[40] M. Eckstein, M. Kollar, and Ph. Werner, Phys. Rev. Lett. **103**, 056403 (2009)

[41] G. Cohen, E. Gull, D.R. Reichman, and A.J. Millis, Phys. Rev. Lett. **115**, 266802 (2015)

[42] A.J. Kim, J. Li, M. Eckstein, and Ph. Werner, arXiv:2204.13562

[43] M. Eckstein and Ph. Werner, Phys. Rev. B **82**, 115115 (2010)

[44] M. Eckstein and Ph. Werner, Sci. Rep. **6**, 21235 (2016)

[45] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003)

[46] K. Held, Ch. 13 in E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.):
The LDA+DMFT Approach to Strongly Correlated Materials
Modeling and Simulation vol. 1, (Forschungszentrum Jülich, 2011)
`http://www.cond-mat.de/events/correl11`

[47] D. Golez, M. Ecksteoin, Ph. Werner, Phys. Rev. B **100**, 235117 (2019)

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

[49] D. Golež, L. Boehnke, M. Eckstein, and P. Werner, Phys. Rev. B **100**, 041111 (2019)

[50] N. Tancogne-Dejean, M.A. Sentef, and A. Rubio, Phys. Rev. Lett. **121**, 097402 (2018)

[51] D.R. Baykusheva *et al.*, Phys. Rev. X **12**, 011013 (2022)

[52] X. Deng, J. Mravlje, R. Zitko, M. Ferrero, G. Kotliar, and A. Georges,
Phys. Rev. Lett. **110**, 086401 (2013)

[53] P.T. Brown *et al.*, Science **363**, 379 (2018)

[54] M. Eckstein and Ph. Werner, Phys. Rev. B **84**, 035122 (2011)

[55] M. Ligges *et al.*, Phys. Rev. Lett. **120**, 166401 (2018)

[56] J. Li and M. Eckstein, Phys. Rev. B **103**, 045133 (2021)

[57] Y. Murakami, Sh. Takayoshi, T. Kaneko, Zh. Sun, D. Golež, A.J. Millis, and Ph. Werner,
Comm. Phys. **5**, 23 (2022)

[58] Ph. Werner, D. Golež and M. Eckstein, arXiv:2204.06762

Index

Symbols
Ψ-functional, 12.28
δ-doping, 9.11
g-factor in a solid-state environment, 2.10
t-t' Hubbard model, 8.41
2DES, 9.2

A
adiabatic theorem, 10.4
analytic continuation, 6.3
Anderson impurity model, 3.3, 5.2, 5.4
Anderson molecule, 8.6, 8.48
Anderson-Holstein impurity model, 5.23
antiferromagnetism, 11.19
atomic-sphere truncation, 8.27
average spectrum method, 6.20

B
band theory, 10.13
Bayes' theorem, 6.10
Bechgaard salts, 2.34
Bernoulli polynomials, 6.7
Berry
  connection, 10.5
  curvature, 10.9
  phase, 10.6
Bethe-Salpeter equation, 8.30, 11.8
bit, 6.29
Bloch Hamiltonian, 10.14
Bohr's atomic model, 1.4
boson distribution function, 5.27
bulk-boundary correspondence, 10.18, 10.22

C
carbon nanotubes, 2.34
central limit theorem, 6.8
Chern
  density, 10.28
  insulator, 10.13
  number, 10.12
chiral Luttinger liquid, 2.38
Clifford algebra, 10.21
cluster dynamical mean field theory, 11.3, 11.4
cluster extensions of DMFT, 1.33
coherent potential approximation (CPA), 1.18, 1.21, 1.25, 1.28, 1.32
cold atoms in optical lattices, 1.13, 1.31
composite fermions, 2.36
  Fermi liquid of, 2.37
continuous-time quantum Monte Carlo, 8.14
  impurity solvers, 5.6
    determinant ratio, 5.10
    expansion order, 5.20
corrections to mean-field theory, 1.31
correlated subspace, 9.5
correlation-induced semimetal, 9.20
cRPA, 12.13
cuprates, 11.12

D
decay of a quasiparticle, 2.15
default model, 6.17
density functional theory (DFT), 7.3, 8.2, 11.12
density of states
  effect of transverse electromagnetic interactions, 2.34
DFT+DMFT, 1.30, 9.5
discrepancy principle, 6.16
disorder, 1.12
double-counting correction, 8.2, 8.24
downfolding, 8.24
dual fermions, 7.3, 11.5
dual functions, 6.6
dumbbell position, 9.12
dynamical mean-field theory (DMFT), 1.24, 1.32, 1.33, 5.4, 8.19, 10.19, 11.2, 12.2
  impurity solver, 8.14
  multi-site and multi-orbital problems, 8.21
  approximation, 5.5
  approximation errors, 4.17
  error correction, 4.24

exact diagonalization solver, 4.14
setting a database, 4.16
dynamical susceptibility, 3.13, 8.41
dynamical vertex approximation, 1.33, 11.4–11.6, 11.10
Dyson equation, 8.6, 11.6

E
electron loitering, 2.33
electronic correlations, 1.11, 1.24, 1.30
energy level flow diagram, 3.11
entropy, 6.29
Euler polynomials, 6.6
exact diagonalization, 4.14
extended DMFT (EDMFT), 12.9

F
Falicov-Kimball model, 1.22
Fermi golden rule, 2.16
Fermi liquid, 1.17, 1.22, 1.27
Fermi-liquid theory, 2.2, 3.8
ferromagnetism, 1.6, 1.11, 1.15, 1.29
filling factor, 2.36
Floquet-engineering, 13.2
fractionally charged excitations, 2.38
frequency-dependent interaction, 5.26
functional integral, 6.20
functional renormalization group (fRG), 1.33

G
gamma process, 6.22
gauge transformations, 10.6
Green function
Matsubara, 6.4
retarded, 6.4
Gutzwiller approximation, 1.16, 1.19
Gutzwiller wave function, 1.16, 1.18, 1.20, 1.32
GW
approximation, 12.6
method, 12.3
GW+DMFT, 1.30, 13.20
GW+EDMFT, 12.7
functional derivation, 12.24

H
hairy-ball theorem, 10.8
hard-sphere fluid, 1.10
Hartree-Fock, 8.20
heavy fermion systems, 2.5
hidden Mott insulator, 9.17
high-temperature superconducting cuprates, 8.41
Holstein-Hubbard model, 5.23
honeycomb lattice, 9.22
Hubbard bands, 12.16
Hubbard dimer, 8.4, 8.46
Hubbard model, 1.11, 1.13–1.15, 1.20, 1.25, 4.28, 8.19, 11.12
Mott transition, 4.28
Hubbard-Stratonovich transformation, 7.10, 12.25
hybridization expansion, 13.18
continuous-time impurity solver, 5.12
acceptance probability, 5.15
Green function measurement, 5.15
Krylov algorithm, 5.19
Legendre measurement, 5.17
matrix formalism, 5.17
multi-orbital impurity models, 5.18
proposal probability, 5.15
segment representation, 5.14
hybridization function, 5.3, 5.13, 7.9, 8.5
hydrodynamic sound velocity, 2.9

I
ill-posed problem, 6.2
impurity entropy, 3.7, 3.13
impurity problem, 7.9
in-gap state, 9.7
infinite dimensions, 1.3, 1.6, 1.18
Ising model, 1.4, 1.6, 1.8, 1.11, 1.32

K
Kadanoff-Baym equations, 13.14
kagomé lattice, 9.22
Karush-Kuhn-Tucker conditions, 6.27
Keldysh
action, 13.6
contour, 13.5
Green function, 13.7
kinetic energy, 5.22
Kondo
cloud, 3.6

interaction, 3.5
model, 3.5
temperature
Anderson model, 3.12
Kondo model, 3.6
Krylov space, 5.20
Kullback-Leibler divergence, 6.31

L
Landau levels, 2.36
degeneracy of, 2.36
Landau theory of Fermi liquids, 2.2
effective mass of ^{3} He, 2.5
adiabatic "switching-on" of the interaction, 2.2
chemical potential, 2.6
classical Hamiltonian for quasiparticles, 2.23
collective modes, 2.24
definition of Landau parameters, 2.8
energy functional, 2.6
heat capacity, 2.8
inhomogeneous systems, 2.33
kinetic equation for the quasiparticle distribution function, 2.22
Landau interaction function, 2.6
Landau parameters, 2.8, 2.13
linearized kinetic equation, 2.24
local quasiparticle energy, 2.7
Luttinger's theorem, 2.3
quasiparticle effective mass, 2.7, 2.28
quasiparticle energy, 2.6
quasiparticles
decay processes, 2.4
phase space argument for decay rates, 2.3
quasiparticles and quasiholes, 2.3
spin mass of a quasiparticle, 2.13
Lang-Firsov transformation, 5.24
LDA+DMFT, 8.19, 12.2
least-squares (LS)
non-negative (NNLS), 6.27
Legendre fitting, 4.14
Legendre transform, 12.27
Lehmann representation, 8.50
linear density response function, 12.3
linear response, 8.34
Liouville's theorem, 2.23
local Dyson equation, 8.6
local moment formation, 3.3
local-vertex approximation, 8.32
localization and electron-electron interactions, 2.34
localization of the basis, 8.27
Luttinger liquid, 2.5
local spectral function, 2.35
plane-wave occupation number, 2.35
spectral function, 2.34
transport properties, 2.35

M
machine learning, 4.4
activation function, 4.10
back-propagation, 4.11
data transformation, 4.23
hyper parameters, 4.20
learning rule, 4.8
neural networks, 4.9, 4.10
objective function, 4.8
Perceptron model, 4.7
training, 4.20
magnetic monopole, 10.11
magnetic susceptibility, 8.28, 8.34
atomic limit, 8.37
DMFT, 8.38
dynamical, 8.41
longitudinal and transverse, 8.41
Mott-insulators, 8.38
non-interacting, 8.35
Stoner model, 8.36
magnon, 11.16
maximum entropy method, 6.17
Bryan, 6.19
classic, 6.19
historic, 6.18
maximum likelihood estimator, 6.10
mean field, 1.7, 1.20, 1.26
mean-field theory, 1.4, 1.6, 1.15, 1.20, 1.23, 1.32
metal-insulator transition, 1.12, 1.17, 1.18, 1.21, 1.25, 1.27, 1.28
model building, 8.24
Moiré heterostructures, 1.12
momentum occupation number

discontinuity at k_F, 2.3
Mott transition, 1.17, 1.27, 1.31, 8.19
multitier GW+EDMFT, 12.12

N
natural heterostructure, 9.13
nickelates, 11.12, 11.23, 11.26
non-crossing approximation, 13.19
non-equilibrium DMFT, 13.16
non-Fermi liquid, 9.12
non-negative least-squares, 6.12
nonequilibrium DMFT, 1.29
normal distribution, truncated, 6.32
Nozières Fermi-liquid theory, 3.8
numerical renormalization group, 3.8

O
orthogonality catastrophe, 2.34
oxygen vacancy, 9.7

P
parallel transport, 10.9
parquet equation, 11.7, 11.8
partition function, 7.4, 12.25
path integral, 6.20
formalism, 7.2
Pauli blocking effect, 2.3
photo-doping, 13.2
Picard plot, 6.14
plasmon, 12.16
Poincaré lemma, 10.9
polar catastrophe, 9.6
polarization
impurity, 12.9
lattice, 12.10
RPA, 12.6
polaron operators, 5.24
potential energy, 5.21
prethermalization, 13.24
principle of detailed balance, 2.24
projection scheme, 7.15
pseudogap, 11.19, 11.23

Q
quantum database, 4.13
quantum Hall effect, 10.17
quantum Hall liquid, 2.35
quantum impurity model, 3.2, 5.2
action, 5.3
bath Green function, 5.4
Green function, 5.3
hybridization function, 5.3
Weiss Green function, 5.4, 5.5
quantum Monte Carlo, 7.2
continuous-time, 8.14
quasi one-dimensional metals
TTF-TCNQ, 2.34
quasiparticle decay rate
effect of disorder, 2.33
quasiparticle lifetime
approximate expressions in 2D, 2.21
approximate formula in three dimensions, 2.19
calculation by Fermi golden rule, 2.16
Giuliani and Quinn formula, 2.21
quenching of the kinetic energy at high magnetic field, 2.36
QWZ model, 10.14

R
random-phase approximation (RPA), 11.2, 12.5
regularization, 6.13
renormalization, 2.26
constant, 2.34
group, 6.25
resistance minimum, 3.2

S
Schrieffer-Wolff transformation, 3.5
Schwinger-Dyson equation, 11.8, 11.9, 11.20
screened interaction, 12.3
self-consistency, 8.19, 9.5
self-consistent equation, 1.8, 1.20, 1.24, 1.26
self-energy, 1.20–1.22, 1.24–1.27, 5.5
Hubbard dimer, 8.5
local, 8.5, 8.8
Shannon entropy, 6.29
single-impurity Anderson model, 1.13, 1.23, 1.24, 3.3
singular value decomposition (SVD), 6.10
Sommerfeld model for electrons in metals, 2.2
spectral function, 6.3

experimental measurement of, 2.21
spectral representation
Green function, 12.23
response function, 12.21
screened interaction, 12.23
self-energy, 12.23
spin fluctuations, 11.3, 11.16, 11.18
spin glass, 1.8
superconductivity, 11.24, 11.26
surface states, 10.22
susceptibility, 8.34, 11.16

T
Tikhonov regularization, 6.14
time-reversal symmetry, 10.16
topological
Hamiltonian, 10.20
invariants, 10.3
phase diagram, 10.28
topological properties, 1.27–1.29
tunneling
between identical quantum wells, 2.21
experiments at high magnetic field, 2.37
twistronics, 1.12
two-channel Kondo model, 3.17

V
vertex, 11.3, 11.9, 11.17

W
Wannier functions, 8.25
weak-coupling continuous-time impurity solver, 5.7
acceptance probability, 5.9
Green function measurement, 5.11
proposal probability, 5.9
Weiss field, 8.20
Wick rotation, 6.3
Wick theorem, 13.11
Wigner representation, 13.9
Wigner's theorem, 10.16
Wilson chain, 3.9
Wolff model, 1.13
wrapping number, 10.16

Z
zeroes of the Laughlin wave function, 2.36

Schriften des Forschungszentrums Jülich
Modeling and Simulation

1. **The LDA+DMFT approach to strongly correlated materials**
   Lecture Notes of the Autumn School 2011 Hands-on LDA+DMFT
   edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2011), 420 pages
   ISBN 978-3-89336-734-4

2. **Correlated Electrons: From Models to Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2012
   edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (2012), 450 pages
   ISBN 978-3-89336-796-2

3. **Emergent Phenomena in Correlated Matter**
   Lecture Notes of the Autumn School on Correlated Electrons 2013
   edited by E. Pavarini, E. Koch, and U. Schollwöck (2013), 520 pages
   ISBN 978-3-89336-884-6

4. **DMFT at 25: Infinite Dimensions**
   Lecture Notes of the Autumn School on Correlated Electrons 2014
   edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2014), 450 pages
   ISBN 978-3-89336-953-9

5. **Many-Body Physics: From Kondo to Hubbard**
   Lecture Notes of the Autumn School on Correlated Electrons 2015
   edited by E. Pavarini, E. Koch, and P. Coleman (2015), 500 pages
   ISBN 978-3-95806-074-6

6. **Quantum Materials: Experiments and Theory**
   Lecture Notes of the Autumn School on Correlated Electrons 2016
   edited by E. Pavarini, E. Koch, J. van den Brink, G. Sawatzky (2016), 420 pages
   ISBN 978-3-95806-159-0

7. **The Physics of Correlated Insulators, Metals, and Superconductors**
   Lecture Notes of the Autumn School on Correlated Electrons 2017
   edited by E. Pavarini, E. Koch, R. Scalettar, and R. Martin (2017), 450 pages
   ISBN 978-3-95806-224-5

8. **DMFT: From Infinite Dimensions to Real Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2018
   edited by E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (2018), 480 pages
   ISBN 978-3-95806-313-6

9. **Many-Body Methods for Real Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2019
   edited by E. Pavarini, E. Koch, and S. Zhang (2019), 520 pages
   ISBN 978-3-95806-400-3

Schriften des Forschungszentrums Jülich
Modeling and Simulation

10. **Topology, Entanglement, and Strong Correlations**
    Lecture Notes of the Autumn School on Correlated Electrons 2020
    edited by E. Pavarini and E. Koch (2020), 500 pages
    ISBN 978-3-95806-466-9

11. **Simulating Correlations with Computers**
    Lecture Notes of the Autumn School on Correlated Electrons 2021
    edited by E. Pavarini and E. Koch (2021), 420 pages
    ISBN 978-3-95806-529-1

12. **Dynamical Mean-Field Theory of Correlated Electrons**
    Lecture Notes of the Autumn School on Correlated Electrons 2022
    edited by E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (2022), 450 pages
    ISBN 978-3-95806-619-9