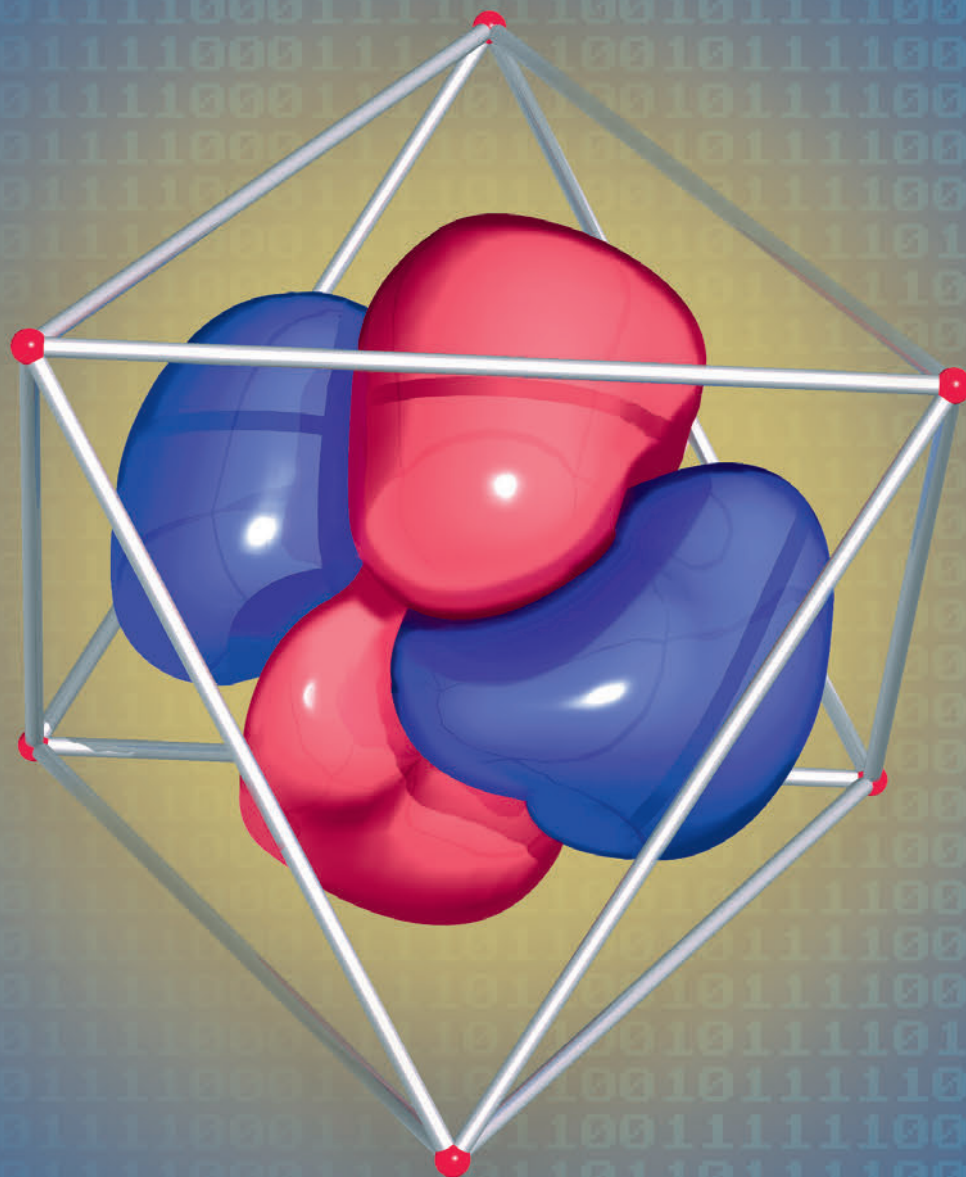




Correlated Materials: Methods and Applications

Eva Pavarini and Erik Koch (Eds.)

Forschungszentrum Jülich GmbH
Peter Grünberg Institute
Jülich Supercomputing Centre

**Lecture Notes of the Autumn School on
Correlated Electrons 2026**

Eva Pavarini and Erik Koch (Eds.)

Correlated Materials: Methods and Applications

Autumn School organized by
the Peter Grünberg Institute
at Forschungszentrum Jülich
21 – 25 September 2026

Schriften des Forschungszentrums Jülich
Modeling and Simulation

Band / Volume 16

ISSN 2192-8525

ISBN 978-3-95806-909-1

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
Jülich Supercomputing Centre

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Schloemer & Partner GmbH, Düren

Copyright: Forschungszentrum Jülich 2026

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 16

ISSN 2192-8525

ISBN 978-3-95806-908-4 (Print)

ISBN 978-3-95806-909-1 (E-Book)

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](https://creativecommons.org/licenses/by/4.0/),
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Contents

Preface

1. Dynamical Mean-Field Theory for Materials
Eva Pavarini
2. Constrained Density-Functional Theory: Formalism, Algorithms and Applications
Xavier Gonze
3. Green Functions and Self-Energy Functionals
Robert Eder
4. Exact Diagonalization and Lanczos Method
Erik Koch
5. DMRG as Impurity Solver
Ulrich Schollwöck
6. Diagrammatic Monte Carlo for the Hubbard Model
Riccardo Rossi
7. Mean-Field Theory of Spin and Charge Order in the Two-Dimensional Hubbard Model
Walter Metzner
8. Introduction to Topological Phases in Condensed Matter
Cécile Repellin
9. Ginzburg–Landau Theory of Superconductivity
Carsten Timm
10. Experimental Characterization of Phase Transitions
Anna Böhrer
11. From Mott to Cluster Mott: Probing Correlations with RIXS
Markus Grüninger
12. Light Control of Many-Body Phases
Matteo Mitrano
13. Variational Monte Carlo for the Hubbard Model
Luca Tocchio
14. Neural Quantum States for Strongly Correlated Systems
Annabelle Bohrdt
15. Quantum Many-Body Simulations on Digital Quantum Computers
Benedikt Fauseweh

Index

Preface

Solving the many-electron problem is the grand challenge of condensed-matter physics. Interacting electrons form emergent co-operative states, whose nature is difficult to unravel. Progress in understanding correlated states of matter requires the combination of several complementary approaches, theoretical and experimental. Only in this way one can pinpoint the microscopic mechanisms driving the co-operative phenomena characterizing strongly-correlated systems.

The goal of this year's school is to provide students with an overview of modern many-body methods and their application to materials, with an outlook to the future of many-body simulations. The program will start with introducing the fundamentals: density-functional theory, the many-body problem and its complexity, emergent phenomena, generalized Hubbard models and their physics. More advanced lectures will cover many-body methods: static and dynamical mean-field theories, cluster methods, density-matrix renormalization-group, tensor networks, and machine learning. Additional lectures will address more explorative approaches, such as variational methods suitable for quantum computers and many-body solvers exploiting artificial neural networks. The lectures will show how the approaches can be used to unravel the mechanism of paradigmatic emergent phenomena in materials: non-conventional superconductivity, Mott phases, orbital ordering, topological phases of matter, the quantum Hall effect, and quantum spin-liquid phenomena. The topics will be treated with a focus on explaining key experiments in a realistic setting and an outlook on questions of materials design. Dedicated experimental lectures will cover experimental methods for characterizing many-body phases as well as experimental equilibrium and out-of-equilibrium probes of many-body states.

A school of this size and scope requires backing from many sources. We are very grateful for all the practical and financial support we have received. The Peter Grünberg Institute at the Forschungszentrum Jülich and the Jülich Supercomputing 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) continued also this year to support the school and supplied additional funds.

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. J. Timmer 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, Amit Chauhan, Qiwei Li, Christopher Mertens 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 the practical issues.

Eva Pavarini and Erik Koch

August 2026

1 Dynamical Mean-Field Theory for Materials

Eva Pavarini
Peter Grünberg Institute
Forschungszentrum Jülich

Contents

1	Introduction	2
2	From DMFT to LDA+DMFT	5
2.1	DMFT for a toy model: The Hubbard dimer	5
2.2	Non-local Coulomb interaction	13
2.3	Quantum-impurity solvers: Continuous-time quantum Monte Carlo	15
2.4	Hartree-Fock (LDA+ U) versus DMFT approximation	20
2.5	DMFT for the one-band Hubbard model	25
2.6	DMFT for multi-orbital models	28
3	Building materials-specific many-body models	30
3.1	Model construction	30
3.2	Localization of the basis and range of the Coulomb interaction	33
3.3	Hubbard Hamiltonians for t_{2g} and e_g systems	34
3.4	Spin-orbit interaction and effects of the basis choice	35
3.5	Non-spherical Coulomb terms and double-counting correction	38
4	Understanding correlated materials with DMFT	40
5	Conclusion and outlook	42
A	Hubbard dimer and Anderson molecule	43
B	Lehmann representation of the local Green function	47

1 Introduction

In 1965, when the two founding papers [1] of density-functional theory (DFT) had just been published in the *Physical Review*, very few could recognize the revolution in the making. Efficient techniques for materials science applications were missing and computers were not that powerful. And yet in 50 years, thanks to remarkable ideas, novel algorithms and steady advance in computational power, brilliant minds brought DFT to its present splendor, making it the *standard model* of condensed matter physics. In 1998, in his Nobel lecture [2], Walter Kohn, described the most important contributions of DFT to science with these words

The first is in the area of fundamental understanding. Theoretical chemists and physicists, following the path of the Schrödinger equation, have become accustomed to think in a truncated Hilbert space of single particle orbitals. The spectacular advances achieved in this way attest to the fruitfulness of this perspective. However, when very high accuracy is required, so many Slater determinants are required (in some calculations up to $\sim 10^9$!) that comprehension becomes difficult.

DFT changed the focus from the N -electron ground-state wavefunction $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ to a three-dimensional variable, the electronic ground-state density $n(\mathbf{r})$, or other directly measurable quantities, such as response functions. These, within a given approximation of the DFT exchange-correlation functional, could be calculated *from first principles*, i.e., using as input the type of atoms involved and, at most, their positions. In practice, $n(\mathbf{r})$ is obtained by mapping the actual many-body problem onto an auxiliary single-electron Hamiltonian with the same ground-state electron density, the Kohn-Sham Hamiltonian. The associated Kohn-Sham eigenvalues are thus in principle merely Lagrange multipliers. Remarkably, however, the big success of DFT came, in part, from bold applications of the theory beyond its actual realm of validity. To general surprise, the Kohn-Sham eigenvalues turned out to be in many cases excellent approximations to the actual elementary excitations of a given material. Early on it became clear, however, that this Ansatz fails qualitatively for a whole category of systems, those in which local Coulomb repulsion effects are large, also known as *strongly-correlated materials*. Paradoxically, in describing strong-correlation phenomena, simple models describing generic features of the microscopic mechanism are much more effective than DFT-based materials-specific calculations. This happens, e.g., for the Kondo effect, heavy-fermion behavior, or the metal-insulator transition. Thus criticisms arose. Particularly outspoken in this context was another Nobel laureate, P.W. Anderson, who emphasized the emergent nature of a true many-body phenomenon [3]. Emergent states typically elude a simple mapping to an effective non-interacting system. For a while, two apparently incompatible philosophies thus coexisted. The first-principles school identified in the materials dependence the essential ingredient for understanding the real world, and tried to correct the failures of the practical implementations of DFT by corrections, often ad hoc, of the exchange-correlation potential. Instead, the many-body models school identified canonical models which explain specific emergent phenomena, dismissing the materials dependence as non-relevant, non-generic detail. In the last three decades these different world-views started to merge, and it became apparent that both sides were right and wrong at the same time. While the

local Coulomb repulsion is indeed key, materials aspects turn out to be essential for understanding real correlated materials. In hindsight, we can now put this debate in a different perspective. The electronic many-body problem, in the non-relativistic limit and in the Born-Oppenheimer approximation, is described by the Hamiltonian

$$\hat{H}_e = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha>\alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|}, \quad (1)$$

where $\{\mathbf{r}_i\}$ are electron coordinates, $\{\mathbf{R}_\alpha\}$ nuclear coordinates and Z_α the nuclear charges. Using a complete one-electron basis, for example the basis $\{\varphi_a(\mathbf{r})\}$, where $\{a\}$ are the quantum numbers, we can write this Hamiltonian in second quantization as

$$\hat{H}_e = \underbrace{-\sum_{ab} t_{ab} c_a^\dagger c_b}_{\hat{H}_0} + \underbrace{\frac{1}{2} \sum_{aa'bb'} U_{aa'bb'} c_a^\dagger c_{a'}^\dagger c_{b'} c_b}_{\hat{H}_U}. \quad (2)$$

Here the hopping integrals are given by

$$t_{ab} = - \int d\mathbf{r} \overline{\varphi_a}(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 - \underbrace{\sum_\alpha \frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|}}_{v_{\text{en}}(\mathbf{r})} \right) \varphi_b(\mathbf{r}), \quad (3)$$

while the elements of the Coulomb tensor are

$$U_{aa'bb'} = \int d\mathbf{r}_2 \int d\mathbf{r}_1 \overline{\varphi_a}(\mathbf{r}_1) \overline{\varphi_{a'}}(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \varphi_{b'}(\mathbf{r}_2) \varphi_b(\mathbf{r}_1). \quad (4)$$

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, $\{\varphi_a^{\text{KS}}(\mathbf{r})\}$, obtained, e.g., in the local density approximation (LDA) or its simple extensions.¹ In this case, it is useful to replace the electron-nuclei interaction $v_{\text{en}}(\mathbf{r})$ with the DFT potential $v_{\text{R}}(\mathbf{r})$, which includes in addition the Hartree term $v_{\text{H}}(\mathbf{r})$ and the (approximate) exchange-correlation potential $v_{\text{xc}}(\mathbf{r})$

$$v_{\text{R}}(\mathbf{r}) = v_{\text{en}}(\mathbf{r}) + \underbrace{\int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}}_{v_{\text{H}}(\mathbf{r})} + v_{\text{xc}}(\mathbf{r}) \quad (5)$$

so that

$$\tilde{t}_{ab} = - \int d\mathbf{r} \overline{\varphi_a^{\text{KS}}}(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 + v_{\text{R}}(\mathbf{r}) \right) \varphi_b^{\text{KS}}(\mathbf{r}). \quad (6)$$

To avoid double counting (DC), we have, however, to subtract from $\hat{H}_U$ the term $\hat{H}_{\text{DC}}$, which describes the Coulomb terms already included in the hopping integrals

$$\hat{H}_e = \underbrace{-\sum_{ab} \tilde{t}_{ab} c_a^\dagger c_b}_{\hat{H}_0 = \hat{H}_e^{\text{LDA}}} + \underbrace{\frac{1}{2} \sum_{aba'b'} \tilde{U}_{aa'bb'} c_a^\dagger c_{a'}^\dagger c_{b'} c_b}_{\Delta \hat{H}_U} - \hat{H}_{\text{DC}}. \quad (7)$$

¹For the purpose of many-body calculations the differences between LDA, GGA or their simple 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 $\hat{H}_e^{\text{LDA}} \sim \hat{H}_{\text{eff}}$, where $\hat{H}_{\text{eff}}$ 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 $\hat{H}_{\text{eff}}$ is obtained via a canonical transformation, so that $\hat{H}_{\text{eff}} \sim \hat{S}^{-1} \hat{H}_e \hat{S}$, 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, however. A model of form $\hat{H}_e^{\text{LDA}}$ does not describe correctly the Mott metal-insulator transition, no matter what the specific values of the parameters $\tilde{t}_{ab}$ are.² Thus for strongly-correlated systems the low-energy effective model must have a different form. For Mott systems a canonical Hamiltonian is the Hubbard model

$$\hat{H} = - \sum_{\sigma} \sum_{ii'} t^{i,i'} c_{i\sigma}^{\dagger} c_{i'\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (8)$$

which includes, in addition to a single-electron term, the on-site Coulomb repulsion. This Hamiltonian 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 $\hat{H}_e^{\text{LDA}}$, 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 (8) in one 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 [4–7]. 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 [4, 5].

DMFT was initially applied 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 thirty 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 impressive, 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 [8–10]. 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.

²One 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; it is also the worst case for DMFT, since the coordination number is the lowest possible. It is nonetheless very useful in order to explain how the method works. 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. In Sec. 3 we describe state-of-the-art 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

$$\hat{H} = \varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_{\sigma} \left(c_{1\sigma}^\dagger c_{2\sigma} + c_{2\sigma}^\dagger c_{1\sigma} \right) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (9)$$

with $i = 1, 2$. The ground state for $N = 2$ electrons (half filling) is the singlet³

$$|G\rangle_H = \frac{a_2(t, U)}{\sqrt{2}} (c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger - c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger) |0\rangle + \frac{a_1(t, U)}{\sqrt{2}} (c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger) |0\rangle \quad (10)$$

with

$$a_1^2(t, U) = \frac{1}{\Delta(t, U)} \frac{\Delta(t, U) - U}{2}, \quad a_2^2(t, U) = \frac{4t^2}{\Delta(t, U)} \frac{2}{\Delta(t, U) - U}, \quad (11)$$

and

$$\Delta(t, U) = \sqrt{U^2 + 16t^2}. \quad (12)$$

The energy of this state is

$$E_0(2) = 2\varepsilon_d + \frac{1}{2}(U - \Delta(t, U)). \quad (13)$$

In the $T \rightarrow 0$ limit, using the Lehmann representation (see Appendix B), one can show that the local Matsubara Green function for spin σ takes then the form

$$G_{i,i}^\sigma(i\nu_n) = \frac{1}{4} \left(\frac{1 + w(t, U)}{i\nu_n - \underbrace{(E_0(2) - \varepsilon_d + t - \mu)}_{E_0(2) - E_-(1) - \mu}} + \frac{1 - w(t, U)}{i\nu_n - \underbrace{(-E_0(2) + U + 3\varepsilon_d + t - \mu)}_{E_+(3) - E_0(2) - \mu}} \right. \\ \left. + \frac{1 - w(t, U)}{i\nu_n - \underbrace{(E_0(2) - \varepsilon_d - t - \mu)}_{E_0(2) - E_+(1) - \mu}} + \frac{1 + w(t, U)}{i\nu_n - \underbrace{(-E_0(2) + U + 3\varepsilon_d - t - \mu)}_{E_-(3) - E_0(2) - \mu}} \right), \quad (14)$$

³Eigenstates 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

$$w_{\pm}(t, U) = \frac{1}{4}(1 \pm w(t, U)), \quad (15)$$

$$w(t, U) = 2a_1(t, U)a_2(t, U) = \frac{4t}{\Delta(t, U)}. \quad (16)$$

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.,

$$G_{i,i}^{\sigma}(i\nu_n) = \frac{1}{2} \left(\underbrace{\frac{1}{i\nu_n + \mu - \varepsilon_d + t - \Sigma^{\sigma}(0; i\nu_n)}}_{G^{\sigma}(0; i\nu_n)} + \underbrace{\frac{1}{i\nu_n + \mu - \varepsilon_d - t - \Sigma^{\sigma}(\pi; i\nu_n)}}_{G^{\sigma}(\pi; i\nu_n)} \right). \quad (17)$$

The self-energy is given by

$$\Sigma^{\sigma}(k; i\nu_n) = \frac{U}{2} + \frac{U^2}{4} \frac{1}{i\nu_n + \mu - \varepsilon_d - \frac{U}{2} - e^{ik} 3t}. \quad (18)$$

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

$$\begin{aligned} \Sigma_l^{\sigma}(i\nu_n) &= \frac{1}{2} \left(\Sigma^{\sigma}(\pi; i\nu_n) + \Sigma^{\sigma}(0; i\nu_n) \right) = \frac{U}{2} + \frac{U^2}{4} \frac{i\nu_n + \mu - \varepsilon_d - \frac{U}{2}}{(i\nu_n + \mu - \varepsilon_d - \frac{U}{2})^2 - (3t)^2} \\ &= \frac{U}{2} + \frac{U^2}{4} \frac{i\nu_n}{(i\nu_n)^2 - (3t)^2} \end{aligned} \quad (19)$$

The difference

$$\begin{aligned} \Delta \Sigma_l^{\sigma}(i\nu_n) &= \frac{1}{2} \left(\Sigma^{\sigma}(\pi; i\nu_n) - \Sigma^{\sigma}(0; i\nu_n) \right) = \frac{U^2}{4} \frac{3t}{(i\nu_n + \mu - \varepsilon_d - \frac{U}{2})^2 - (3t)^2} \\ &= \frac{U^2}{4} \frac{3t}{(i\nu_n)^2 - (3t)^2}, \end{aligned} \quad (20)$$

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

$$F^{\sigma}(i\nu_n) = \frac{(t + \Delta \Sigma_l^{\sigma}(i\nu_n))^2}{i\nu_n + \mu - \varepsilon_d - \Sigma_l^{\sigma}(i\nu_n)} \quad (21)$$

which for $U = 0$ becomes

$$F_0^{\sigma}(i\nu_n) = \frac{t^2}{i\nu_n}. \quad (22)$$

By using these definitions, we can rewrite the local Green function as

$$G_{i,i}^{\sigma}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - F^{\sigma}(i\nu_n) - \Sigma_l^{\sigma}(i\nu_n)}. \quad (23)$$

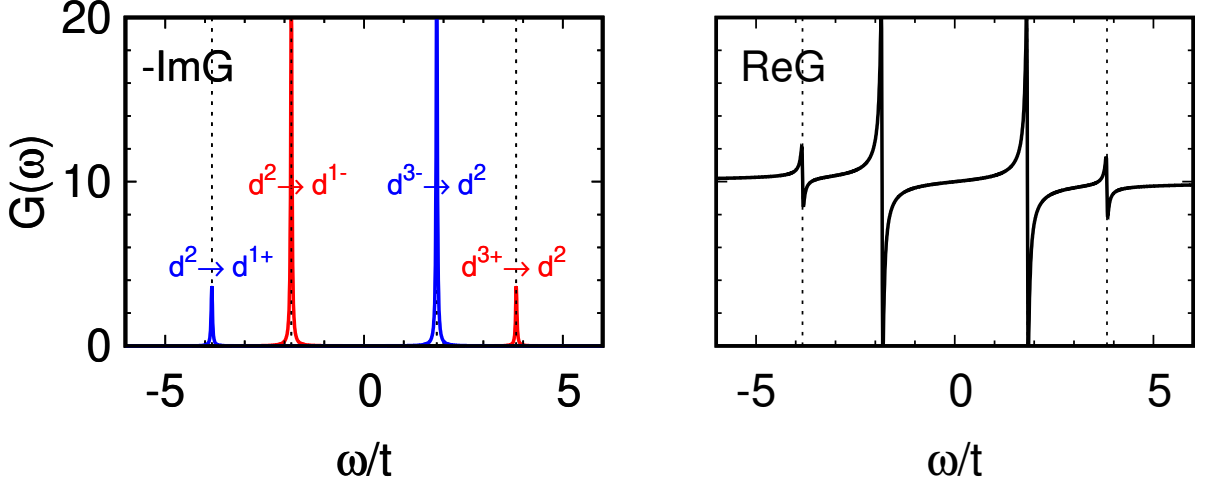


Fig. 1: *Hubbard dimer: Imaginary (left) and real (right) part of the retarded Green function, obtained setting $i\nu_n \rightarrow \omega + i\delta$ (analytic continuation) in Eq. (14). Red lines: $k = 0$ contribution. Blue lines: $k = \pi$ contribution. Dashed lines: Poles of the retarded Green function. Parameters: $t = 1, U = 4$. The weight of the poles yielding the smaller peaks, $w_-(t, U)$, defined in Eq. (15), goes to zero for $U \rightarrow 0$. 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 ($i\nu_n \rightarrow \omega + i\delta$), 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*

$$\Sigma_l^\sigma(i\nu_n) = \frac{1}{\mathfrak{G}_{i,i}^\sigma(i\nu_n)} - \frac{1}{G_{i,i}^\sigma(i\nu_n)}, \quad (24)$$

where $\mathfrak{G}_{i,i}^\sigma(i\nu_n)$ is given by

$$\mathfrak{G}_{i,i}^\sigma(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - F^\sigma(i\nu_n)}. \quad (25)$$

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

$$\hat{H}^A = \varepsilon_s \sum_\sigma \hat{n}_{s\sigma} - t \sum_\sigma \left(c_{d\sigma}^\dagger c_{s\sigma} + c_{s\sigma}^\dagger c_{d\sigma} \right) + \varepsilon_d \sum_\sigma \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow}, \quad (26)$$

where s labels the uncorrelated bath site and d the correlated quantum-impurity site. The first constraint would be that Hamiltonian (26) 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

$$\begin{aligned}
 |1\rangle &= c_{1\uparrow}^\dagger c_{2\uparrow}^\dagger |0\rangle, & |4\rangle &= \frac{1}{\sqrt{2}} (c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger - c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger) |0\rangle, \\
 |2\rangle &= c_{1\downarrow}^\dagger c_{2\downarrow}^\dagger |0\rangle, & |5\rangle &= c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger |0\rangle, \\
 |3\rangle &= \frac{1}{\sqrt{2}} (c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger) |0\rangle, & |6\rangle &= c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger |0\rangle.
 \end{aligned} \tag{27}$$

In this basis the Hamiltonian of the Hubbard dimer has the matrix form

$$\hat{H}_2(\varepsilon_d, U, t) = \begin{pmatrix} 2\varepsilon_d & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d+U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_d+U \end{pmatrix}. \tag{28}$$

The ground state, the singlet given in Eq. (10), can be obtained by diagonalizing the lower 3×3 block. For the Anderson molecule, ordering the basis in the same way ($1 \rightarrow d, 2 \rightarrow s$), this Hamiltonian becomes

$$\hat{H}_2^A(\varepsilon_d, U, t; \varepsilon_s) = \begin{pmatrix} \varepsilon_d + \varepsilon_s & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_d + \varepsilon_s & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_d + \varepsilon_s & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_d + \varepsilon_s & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d+U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_s \end{pmatrix}. \tag{29}$$

Comparing the lower 3×3 block of $\hat{H}_2^A(\varepsilon_d, U, t; \varepsilon_s)$ 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

$$\hat{H}_2^A(\varepsilon_d, U, t; \mu) = \hat{H}_2(\varepsilon_d + \frac{U}{4}, \frac{U}{2}, t). \tag{30}$$

The $N = 2$ ground state of $\hat{H}_2^A(\varepsilon_d, U, t; \mu)$ has thus the form of the ground-state for the Hubbard dimer

$$|G\rangle_A = \frac{a_2(t, U/2)}{\sqrt{2}} (c_{d\uparrow}^\dagger c_{s\downarrow}^\dagger - c_{d\downarrow}^\dagger c_{s\uparrow}^\dagger) |0\rangle + \frac{a_1(t, U/2)}{\sqrt{2}} (c_{d\uparrow}^\dagger c_{d\downarrow}^\dagger + c_{s\uparrow}^\dagger c_{s\downarrow}^\dagger) |0\rangle, \tag{31}$$

and the condition $n_s = n_d = 1$ is satisfied. Since $\varepsilon_s \neq \varepsilon_d$, however, the eigenstates of $\hat{H}^A$ 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

$$G_{d,d}^{\sigma}(i\nu_n) = \frac{1}{4} \left(\frac{1 + w'(t, U)}{i\nu_n - (E_0(2) - E_-(1) - \mu)} + \frac{1 - w'(t, U)}{i\nu_n - (E_+(3) - E_0(2) - \mu)} \right. \\ \left. + \frac{1 - w'(t, U)}{i\nu_n - (E_0(2) - E_+(1) - \mu)} + \frac{1 + w'(t, U)}{i\nu_n - (E_-(3) - E_0(2) - \mu)} \right), \quad (32)$$

where

$$E_0(2) - E_{\pm}(1) - \mu = -\left(E_{\pm}(3) - E_0(2) - \mu\right) = -\frac{1}{4} \left(2\Delta(t, U/2) \pm \Delta(t, U)\right) \quad (33)$$

and

$$w'(t, U) = \frac{1}{2} \frac{32t^2 - U^2}{\Delta(t, U)\Delta(t, U/2)}. \quad (34)$$

After some rearrangement we obtain a much simpler expression

$$G_{d,d}^{\sigma}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}_0^{\sigma}(i\nu_n) - \Sigma_A^{\sigma}(i\nu_n)}. \quad (35)$$

The impurity self-energy equals the local self-energy of the Hubbard dimer

$$\Sigma_A^{\sigma}(i\nu_n) = \frac{U}{2} + \frac{U^2}{4} \frac{i\nu_n}{(i\nu_n)^2 - (3t)^2}, \quad (36)$$

as one may see comparing it to equation (19). The hybridization function is given by

$$\mathcal{F}_0^{\sigma}(i\nu_n) = \frac{t^2}{i\nu_n}, \quad (37)$$

as for the non-interacting Hubbard dimer, Eq. (22). For $U = 0$, $G_{d,d}^{\sigma}(i\nu_n)$ equals the non-interacting impurity Green function

$$G_{d,d}^{0\sigma}(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}_0^{\sigma}(i\nu_n)}. \quad (38)$$

The impurity Green function thus satisfies the *impurity Dyson equation*

$$\Sigma_A^{\sigma}(i\nu_n) = \frac{1}{G_{d,d}^{0\sigma}(i\nu_n)} - \frac{1}{G_{d,d}^{\sigma}(i\nu_n)}. \quad (39)$$

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 $\Delta\Sigma_l^{\sigma}(\omega) = 0$ yields large errors. The right panels demonstrate, however, that the spectral function of the Anderson molecule is quite similar to the

⁴The 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.

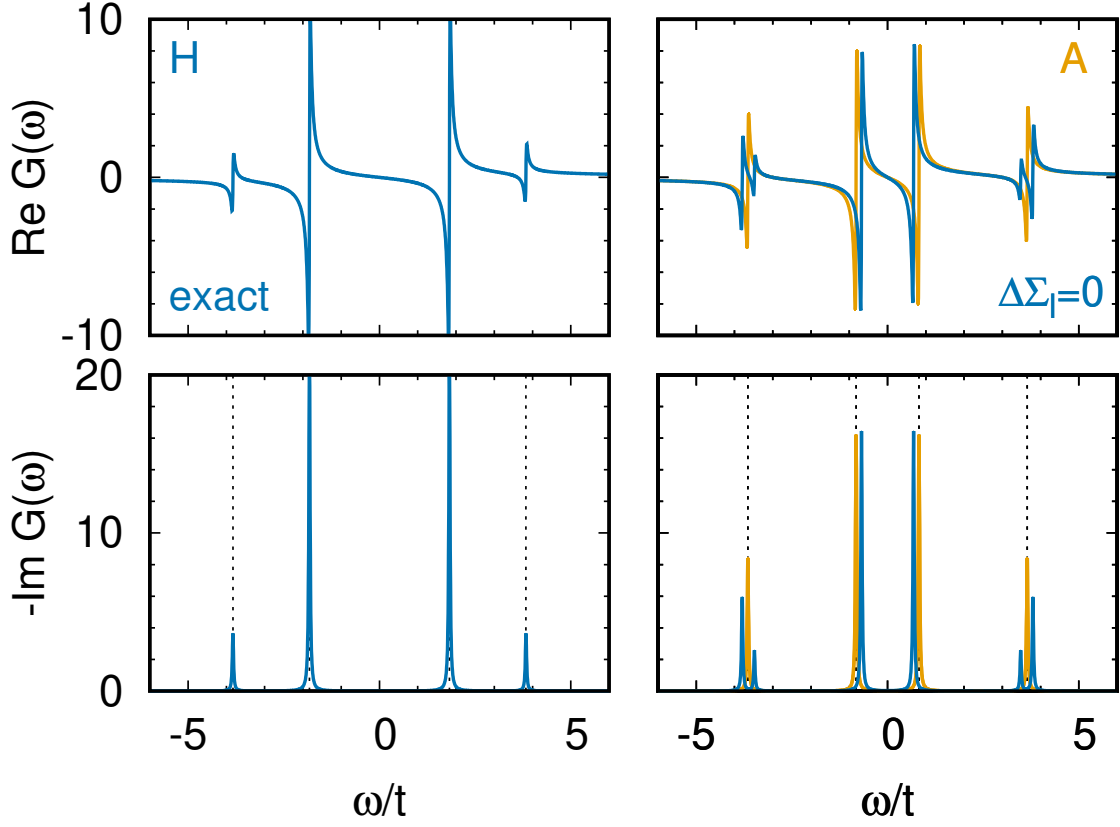


Fig. 2: Retarded Green function of the Hubbard dimer ($t = 1, U = 4$) and of the Anderson molecule ($\varepsilon_s = \varepsilon_d + U/2$) in the zero temperature limit. Left panels: Hubbard dimer, exact Green function. Right panels, blue: Hubbard dimer in the local self-energy approximation, i.e., with $\Delta\Sigma_l^\sigma(\omega) = 0$. Right panels, orange: Anderson molecule. Dashed lines: Poles for the Hubbard dimer (left) or the Anderson molecule (right).

one of the Hubbard dimer with $\Delta\Sigma_l^\sigma(\omega) = 0$. The small remaining deviations come from the fact that, for the Hubbard dimer, in the impurity Dyson equation, the non-interacting impurity Green function is replaced by $\mathcal{G}_{i,i}^\sigma(i\nu_n)$ in the local self-energy approximation, i.e., by the *bath* Green function

$$\mathcal{G}_{i,i}^\sigma(i\nu_n) = \frac{1}{i\nu_n + \mu - \varepsilon_d - \mathcal{F}_l^\sigma(i\nu_n)}, \quad (40)$$

where

$$\mathcal{F}_l^\sigma(i\nu_n) = \frac{t^2}{i\nu_n + \mu - \varepsilon_d - \Sigma_A^\sigma(i\nu_n)}. \quad (41)$$

We are now in the position of explaining how DMFT works for the Hamiltonian of the Hubbard dimer, choosing the Anderson molecule Hamiltonian (26) as the auxiliary quantum-impurity model. The procedure can be split in the following steps

1. Build the initial quantum impurity model with $G_{d,d}^{0\sigma}(i\nu_n) = G_{i,i}^{0\sigma}(i\nu_n)$. The initial bath is thus defined by energy $\varepsilon_s = \varepsilon_d$ and hopping t .

2. Calculate the local Green function $G_{d,d}^\sigma(i\nu_n)$ 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

$$\Delta\mathcal{F}_l(i\nu_n) = \mathcal{F}_l^\sigma(i\nu_n) - \mathcal{F}_0^\sigma(i\nu_n) = t^2 p^2 \left(-\frac{2}{i\nu_n} + \frac{1}{i\nu_n - \varepsilon_a} + \frac{1}{i\nu_n + \varepsilon_a} \right) \quad (42)$$

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 is very good for any U value. This indicates that when the local-self-energy approximation (hence **DMFT**) 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

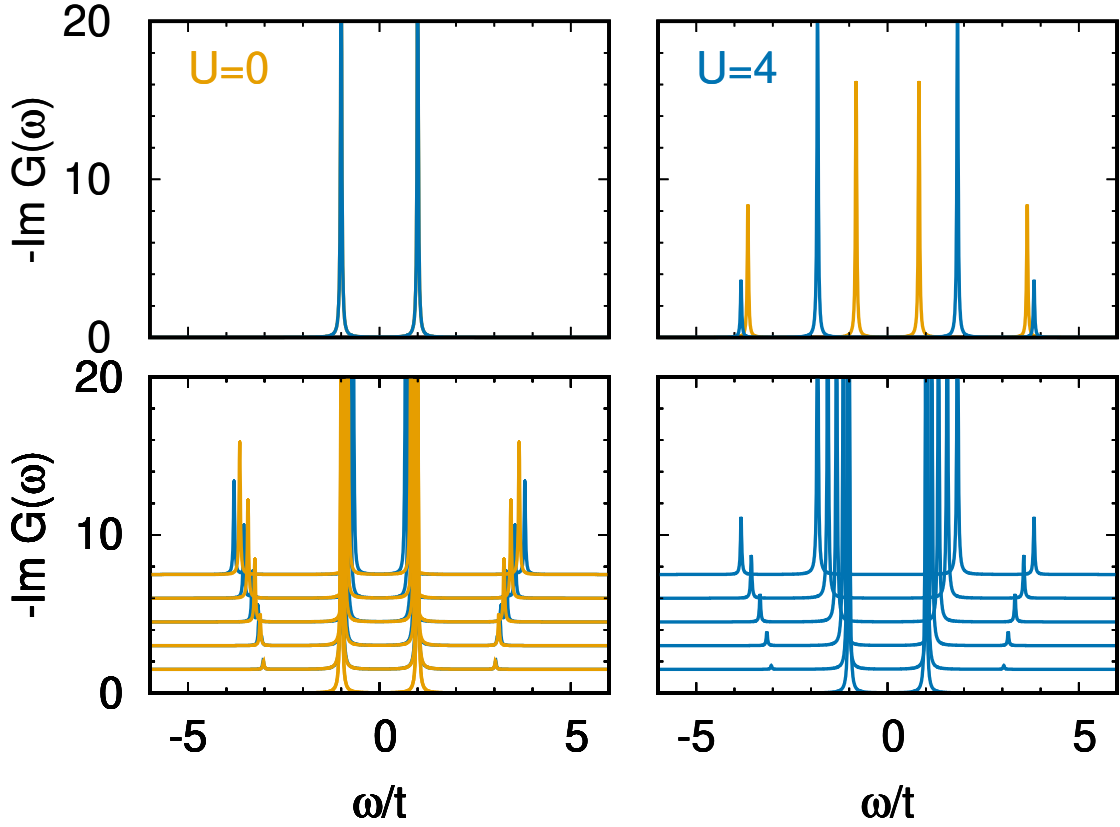


Fig. 3: Imaginary part of the retarded Green function of the Anderson molecule (orange) and Hubbard dimer (blue) in the zero temperature limit. In the bottom left panel the local self-energy approximation is adopted for the Hubbard dimer; in all other cases the exact Green function of the Hubbard dimer 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 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 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. This also points to the possible shortcomings of calculations in which the DMFT quantum-impurity model for the *lattice* Hubbard model is solved via exact diagonalization, however using a single bath site or very few; this is sufficient in the limit of large gap,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.

⁵For 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. [11].

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

$$\begin{aligned} \hat{H} = & \varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_{\sigma} (c_{1\sigma}^\dagger c_{2\sigma} + c_{2\sigma}^\dagger c_{1\sigma}) + U \sum_{i=1,2} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \\ & + \sum_{\sigma\sigma'} (V - 2J_V - J_V \delta_{\sigma\sigma'}) \hat{n}_{1\sigma} \hat{n}_{2\sigma'} - J_V \sum_{i \neq i'} (c_{i\uparrow}^\dagger c_{i\downarrow} c_{i'\downarrow}^\dagger c_{i'\uparrow} + c_{i'\uparrow}^\dagger c_{i'\downarrow} c_{i\uparrow} c_{i\downarrow}), \end{aligned} \quad (43)$$

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

$$\hat{H}_2^{\text{NL}} = \begin{pmatrix} 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d + V - J_V & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d + U & -J_V \\ 0 & 0 & 0 & -\sqrt{2}t & -J_V & 2\varepsilon_d + U \end{pmatrix},$$

where the basis is defined in Eq. (27). 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 $\hat{H}_2^{\text{NL}}$ 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 $\hat{H}_2^{\text{NL}} = \hat{H}_2(\varepsilon'_d, U', t)$, where the right-hand-side term is the $N=2$ -electron Hamiltonian of the $J_V = V = 0$ Hubbard dimer, Eq. (28), with parameters $\varepsilon'_d = \varepsilon_d + V/2$ and $U' = U - V$. The $N=2$ ground state is thus still given by Eq. (10), provided that we replace U with U' in the coefficients. Eventually, in the limiting case $U = V$, $\hat{H}_2^{\text{NL}}$ 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. (14) setting

$$\begin{aligned} E_{\pm}(N=1, U; V) &= E_{\pm}(N=1, U; 0) &= \varepsilon_d \pm t \\ E_{\pm}(N=3, U; V) &= E_{\pm}(N=3, U; 0) + 2V &= 3\varepsilon_d \pm t + U + 2V \\ E_0(N=2, U; V) &= E_0(N=2, U-V; 0) + V &= E_0(2, U-V) + V \\ \mu(U; V) &= \mu(U) + V &= \mu + V \\ w_{\pm}(t, U; V) &= w_{\pm}(t, U-V; 0). \end{aligned}$$

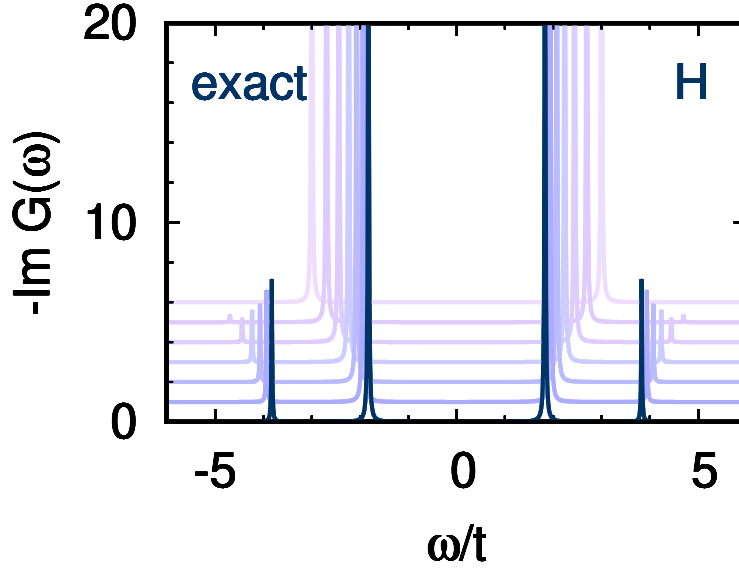


Fig. 4: Imaginary part of the retarded Green function of the Hubbard dimer in the zero temperature limit ($U = 4$, $t = 1$), increasing the intersite Coulomb repulsion V from 0 to $V = U = 4$ in equal steps; we have set $J_V = 0$. The dark blue line corresponds to $V = 0$.

Thus we have, recalling that $U' = U - V$,

$$G_{i,i}^{\sigma}(i\nu_n) = \frac{w_+(t, U')}{i\nu_n - (E'_0(2) - \frac{1}{2}V - \varepsilon_d + t - \mu')} + \frac{w_-(t, U')}{i\nu_n - (-E'_0(2) + U' + \frac{1}{2}V + 3\varepsilon_d + t - \mu')} \\ + \frac{w_-(t, U')}{i\nu_n - (E'_0(2) - \frac{1}{2}V - \varepsilon_d - t - \mu')} + \frac{w_+(t, U')}{i\nu_n - (-E'_0(2) + U' + \frac{1}{2}V + 3\varepsilon_d - t - \mu')}, \quad (44)$$

where we set $\mu' = \mu - V/2 = 2\varepsilon_d + U'/2$ 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, $t \rightarrow t + V/2$. 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 [12], a very successful QMC-based quantum-impurity solver. In this approach, the first step consists in splitting the Hamiltonian into bath ($\hat{H}_{\text{bath}}$), hybridization ($\hat{H}_{\text{hyb}}$), and local ($\hat{H}_{\text{loc}}$) terms

$$\hat{H}^A = \underbrace{\varepsilon_s \sum_{\sigma} \hat{n}_{s\sigma}}_{\hat{H}_{\text{bath}}} - t \underbrace{\sum_{\sigma} (c_{d\sigma}^{\dagger} c_{s\sigma} + c_{s\sigma}^{\dagger} c_{d\sigma})}_{\hat{H}_{\text{hyb}}} + \underbrace{\varepsilon_d \sum_{\sigma} \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow}}_{\hat{H}_{\text{loc}}}. \quad (45)$$

Next, we write the partition function Z as a perturbation series in the hybridization. To this end, we define $\hat{H}_0 = \hat{H}_{\text{bath}} + \hat{H}_{\text{loc}}$ and rewrite the partition function as

$$Z = \text{Tr} \left(e^{-\beta(\hat{H}_0 - \mu \hat{N})} \hat{V}(\beta) \right) \quad (46)$$

where the operator $\hat{V}(\beta)$ is given by

$$\hat{V}(\beta) = e^{\beta(\hat{H}_0 - \mu \hat{N})} e^{-\beta(\hat{H}_0 + \hat{H}_{\text{hyb}} - \mu \hat{N})} = \sum_m \underbrace{\int_0^{\beta} d\tau_1 \cdots \int_{\tau_{m-1}}^{\beta} d\tau_m}_{\int d\tau^m} \underbrace{(-1)^m \prod_{l=m}^1 \hat{H}_{\text{hyb}}(\tau_l)}_{\hat{O}^m(\tau)}, \quad (47)$$

and

$$\hat{H}_{\text{hyb}}(\tau_l) = e^{\tau_l(\hat{H}_0 - \mu \hat{N})} \hat{H}_{\text{hyb}} e^{-\tau_l(\hat{H}_0 - \mu \hat{N})} = -t \sum_{\sigma} \left(c_{d\sigma_l}^{\dagger}(\tau_l) c_{s\sigma_l}(\tau_l) + c_{s\sigma_l}^{\dagger}(\tau_l) c_{d\sigma_l}(\tau_l) \right). \quad (48)$$

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

$$\int d\tau^{2k} \longrightarrow \int d\tau^k \int d\bar{\tau}^k \quad \text{and} \quad \hat{O}^{2k}(\tau) \longrightarrow \sum_{\sigma, \bar{\sigma}} \hat{O}_{\sigma, \bar{\sigma}}^{2k}(\tau, \bar{\tau}) \quad (49)$$

where

$$\hat{O}_{\sigma, \bar{\sigma}}^{2k}(\tau, \bar{\tau}) = (t)^{2k} \prod_{i=1}^k \left(c_{d\bar{\sigma}_i}^{\dagger}(\bar{\tau}_i) c_{s\bar{\sigma}_i}(\bar{\tau}_i) c_{s\sigma_i}^{\dagger}(\tau_i) c_{d\sigma_i}(\tau_i) \right). \quad (50)$$

The vector $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_k)$ gives the spins $\{\sigma_i\}$ associated with the k impurity annihilators at imaginary times $\{\tau_i\}$, while $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \dots, \bar{\sigma}_k)$ 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

$$\frac{Z}{Z_{\text{bath}}} = \sum_k \int d\tau^k \int d\bar{\tau}^k \sum_{\sigma, \bar{\sigma}} d_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau}) t_{\sigma, \bar{\sigma}}^k(\tau, \bar{\tau}) \quad (51)$$

$$d_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau}) = \frac{t^{2k}}{Z_{\text{bath}}} \text{Tr}_{\text{bath}} \left(e^{-\beta(\hat{H}_{\text{bath}} - \mu \hat{N}_s)} \mathcal{T} \prod_{i=k}^1 c_{s\sigma_i}^\dagger(\tau_i) c_{s\bar{\sigma}_i}(\bar{\tau}_i) \right) \quad (52)$$

$$t_{\sigma, \bar{\sigma}}^k(\tau, \bar{\tau}) = \text{Tr}_{\text{loc}} \left(e^{-\beta(\hat{H}_{\text{loc}} - \mu \hat{N}_d)} \mathcal{T} \prod_{i=k}^1 c_{d\sigma_i}(\tau_i) c_{d\bar{\sigma}_i}^\dagger(\bar{\tau}_i) \right), \quad (53)$$

where $Z_{\text{bath}} = 1 + 2e^{-\beta(\varepsilon_s - \mu)} + e^{-2\beta(\varepsilon_s - \mu)}$ and

$$c_{d\sigma}(\tau) = e^{\tau(\hat{H}_{\text{loc}} - \mu \hat{N}_d)} c_{d\sigma} e^{-\tau(\hat{H}_{\text{loc}} - \mu \hat{N}_d)}, \quad c_{s\sigma}(\tau) = e^{\tau(\hat{H}_{\text{bath}} - \mu \hat{N}_s)} c_{s\sigma} e^{-\tau(\hat{H}_{\text{bath}} - \mu \hat{N}_s)}.$$

The trace involving only bath operators is simple to calculate, since $\hat{H}_{\text{bath}}$ describes an independent-electron problem for which Wick's theorem holds. It is given by the determinant

$$d_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau}) = \det(\mathcal{F}_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau})) \quad (54)$$

of the $k \times k$ non-interacting hybridization-function matrix, with elements

$$(\mathcal{F}_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau}))_{i', i} = \mathcal{F}_{\bar{\sigma}_{i'}, \sigma_i}^0(\bar{\tau}_{i'} - \tau_i) \quad (55)$$

where

$$\mathcal{F}_{\bar{\sigma}, \sigma}^0(\tau) = \delta_{\bar{\sigma}, \sigma} \frac{t^2}{1 + e^{-\beta(\varepsilon_s - \mu)}} \times \begin{cases} -e^{-\tau(\varepsilon_s - \mu)} & \tau > 0, \\ +e^{-(\beta + \tau)(\varepsilon_s - \mu)} & \tau < 0. \end{cases} \quad (56)$$

This is the imaginary time Fourier transform of the hybridization function introduced previously

$$\mathcal{F}_{\bar{\sigma}, \sigma}^0(i\nu_n) = \frac{t^2}{i\nu_n - (\varepsilon_s - \mu)} \delta_{\bar{\sigma}, \sigma}. \quad (57)$$

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 [13]

$$\frac{Z}{Z_{\text{bath}}} = \left(\prod_\sigma \sum_{k_\sigma=0}^\infty \int d\tau_\sigma^{k_\sigma} \int d\bar{\tau}_\sigma^{k_\sigma} \right) d_{\bar{\sigma}, \sigma}^k(\tau, \bar{\tau}) t_{\sigma, \bar{\sigma}}^k(\tau, \bar{\tau}) \quad (58)$$

where the vectors $\sigma = (\sigma_\uparrow, \sigma_\downarrow)$ and $\bar{\sigma} = (\bar{\sigma}_\uparrow, \bar{\sigma}_\downarrow)$ have $(k_\uparrow, k_\downarrow)$ components, and for each k_σ component $\sigma_i = \bar{\sigma}_i = \sigma$. Thus

$$t_{\sigma, \bar{\sigma}}^k(\tau, \bar{\tau}) = \text{Tr}_{\text{loc}} \left(e^{-\beta(\hat{H}_{\text{loc}} - \mu \hat{N}_d)} \mathcal{T} \prod_\sigma \prod_{i=k_\sigma}^1 c_{d\sigma_i}(\tau_{\sigma_i}) c_{d\sigma_i}^\dagger(\bar{\tau}_{\bar{\sigma}_i}) \right). \quad (59)$$

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

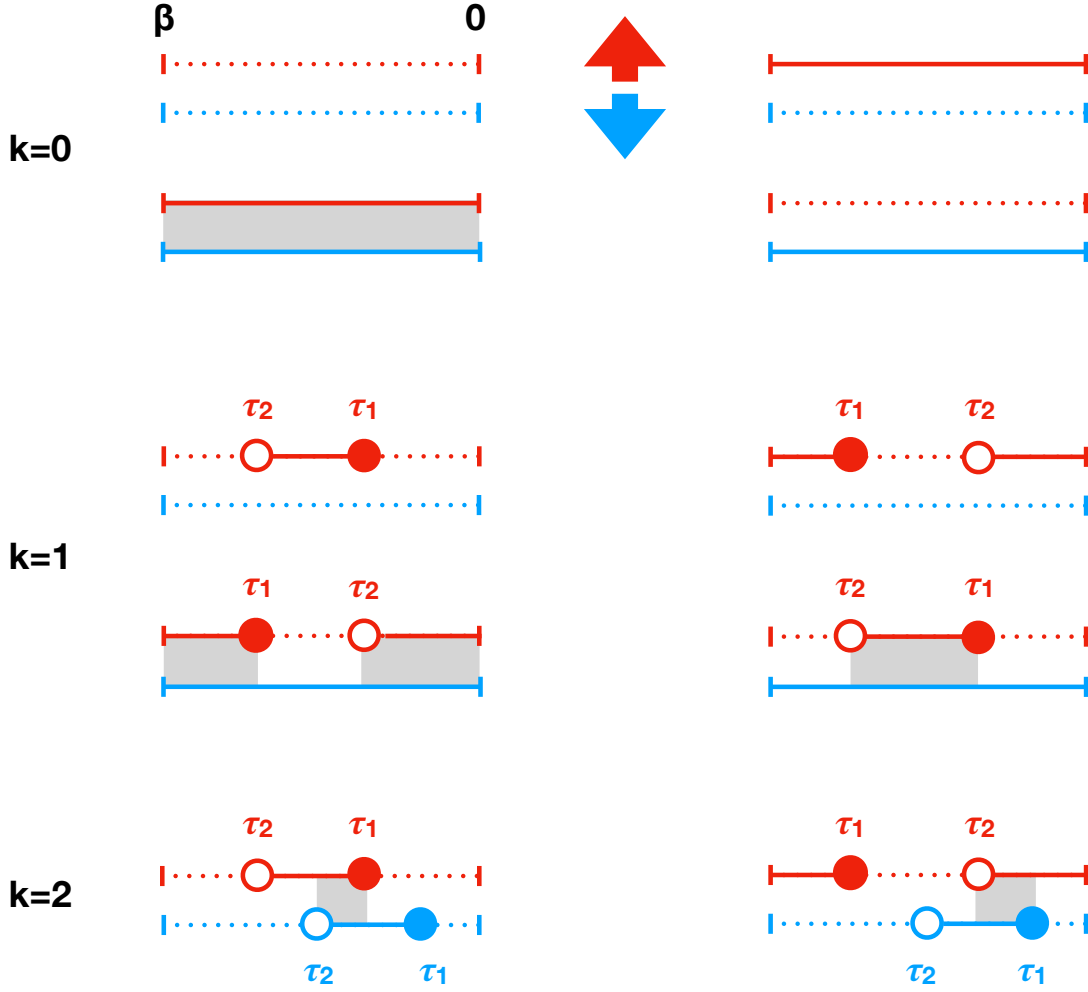


Fig. 5: Representative configurations contributing to the local trace at zeroth, first and second order. The timelines for spin up are red and those for spin down are blue. The filled circles correspond to the insertion of a creator (time τ_1), and the empty circles to the insertion of an annihilator (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 $c_{d\sigma}^\dagger|0\rangle$. A given configuration yields, at order $k = k_\uparrow + k_\downarrow$

$$t_{\sigma,\bar{\sigma}}^k(\tau, \bar{\tau}) = \left(\prod_{\sigma} s_{\sigma}^{k_{\sigma}} \right) e^{-\sum_{\sigma\sigma'} ((\varepsilon_d - \mu)\delta_{\sigma\sigma'} + \frac{U}{2}(1 - \delta_{\sigma,\sigma'})) l_{\sigma,\sigma'}} \quad (60)$$

where $l_{\sigma,\sigma'}$ is the length of the overlap of the τ segments for spins σ and σ' , respectively, while $s_{\sigma} = \text{sign}(\tau_{\sigma_1} - \bar{\tau}_{\sigma_1})$ 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 $Z_{\text{loc}} = 1 + 2e^{-\beta(\varepsilon_d - \mu)} + e^{-\beta(2(\varepsilon_d - \mu) + U)}$. 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

$$d_{\sigma\sigma}^1(\tau_1, \tau_2) = \mathcal{F}_{\sigma\sigma}^0(\tau_1, \tau_2) = -\frac{t^2}{2} \delta_{\sigma,\bar{\sigma}} \text{sign}(\tau_1 - \tau_2). \quad (61)$$

The local trace at the same order is instead given by

$$t_{\sigma\bar{\sigma}}^1(\tau_2, \tau_1) = \text{Tr}_{\text{loc}} \left(e^{-\beta(\hat{H}_{\text{loc}} - \mu \hat{N}_d)} \mathcal{T} c_{d\sigma}(\tau_2) c_{d\sigma}^\dagger(\tau_1) \right). \quad (62)$$

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

$$t_{\uparrow\uparrow}^1(\tau_2, \tau_1) = \begin{cases} +e^{-(\tau_2 - \tau_1)(\varepsilon_d - \mu)} & = +e^{+\tau_{21}U/2} \\ -e^{-(\beta - (\tau_1 - \tau_2))(\varepsilon_d - \mu)} & = -e^{(\beta + \tau_{21})U/2} \\ -e^{-\beta(2(\varepsilon_d - \mu) + U) + (\tau_1 - \tau_2)(\varepsilon_d - \mu + U)} & = -e^{-\tau_{21}U/2} \\ +e^{-(\tau_2 - \tau_1)(\varepsilon_d - \mu + U) - \beta(\varepsilon_d - \mu)} & = +e^{(\beta - \tau_{21})U/2} \end{cases} \quad (63)$$

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

$$\begin{aligned} \frac{Z}{Z_{\text{bath}}} &\sim Z_{\text{loc}} + \sum_{\sigma} \int_0^{\beta} d\tau_2 \int_0^{\beta} d\tau_1 d_{\sigma\sigma}^1(\tau_1, \tau_2) t_{\sigma\sigma}^1(\tau_2, \tau_1) \\ &\sim Z_{\text{loc}} \left(1 - \beta \frac{1 - e^{\frac{\beta U}{2}}}{1 + e^{\frac{\beta U}{2}}} \frac{2t^2}{U} \right). \end{aligned} \quad (64)$$

The exact formula of the partition function can be obtained from the eigenvalues and eigenvectors in Appendix A

$$\frac{Z}{Z_{\text{bath}}} = Z_{\text{loc}} \frac{3(1 + e^{\frac{\beta U}{2}}) + e^{\frac{\beta U}{4}} \left(4e^{-\frac{\beta \Delta(t, U)}{4}} + 4e^{+\frac{\beta \Delta(t, U)}{4}} + e^{+\frac{\beta \Delta(t, U/2)}{2}} + e^{-\frac{\beta \Delta(t, U/2)}{2}} \right)}{8(1 + e^{\frac{\beta U}{2}})}. \quad (65)$$

Its Taylor expansion in powers of t/U yields, at second order, the expression above. Going back to Eq. (62), one can observe that, for $k = 1$, the local trace is proportional to the local Green function, $G_{d,d}^{\sigma}(\tau)$. Indeed, $G_{d,d}^{\sigma}(\tau)$ 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

$$G_{d,d}^{\sigma}(\tau) \sim -\frac{1}{\beta} \int_0^{\beta} \int_0^{\beta} \underbrace{d\tau_2 d\tau_1 d_{\sigma\sigma}^1(\tau_1, \tau_2) t_{\sigma\sigma}^1(\tau_2, \tau_1)}_{w^1} \delta(\tau - (\tau_2 - \tau_1)) \frac{1}{\mathcal{F}_{\sigma\sigma}^0(\tau_1 - \tau_2)}. \quad (66)$$

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

$$Z = \sum_c w_c = \sum_c |w_c| \text{sign } w_c. \quad (67)$$

In a compact form, we can write $w_c = d\tau_c d_c t_c$ where $d\tau_c = \prod_\sigma \prod_i^{k_\sigma} d\tau_{\sigma i} d\bar{\tau}_{\bar{\sigma} i}$, 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

$$\langle \hat{O} \rangle = \frac{\sum_c \langle \hat{O} \rangle_c |w_c| \text{sign } w_c}{\sum_c |w_c| \text{sign } w_c} = \frac{\sum_c \text{sign } w_c \langle \hat{O} \rangle_c |w_c| / \sum_c |w_c|}{\sum_c \text{sign } w_c |w_c| / \sum_c |w_c|} \approx \frac{\frac{1}{N_c} \sum_c \langle \hat{O} \rangle_c \text{sign } w_c}{\frac{1}{N_c} \sum_c \text{sign } w_c}. \quad (68)$$

The term $\frac{1}{N_c} \sum_c \text{sign } w_c$ 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 $c \rightarrow c'$. Within the Metropolis algorithm, the acceptance ratio is

$$R_{c \rightarrow c'} = \min \left(1, \frac{p_{c' \rightarrow c} |w_{c'}|}{p_{c \rightarrow c'} |w_c|} \right) \quad (69)$$

where $p_{c \rightarrow c'}$ is the proposal probability for the update $c \rightarrow c'$. 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 τ_{in} ; the move is rejected if τ_{in} is in a region where a segment exists. If created, the segment can have at most length l_{max} , given by the distance between τ_{in} and the time at which the next creator is, hence

$$p_{c \rightarrow c'} = \frac{d\bar{\tau}}{\beta} \frac{d\tau}{l_{\text{max}}}. \quad (70)$$

The proposal probability of the reverse move (removing a segment) is instead given by the inverse of the number of existing segments

$$p_{c' \rightarrow c} = \frac{1}{k_\sigma + 1}. \quad (71)$$

The acceptance ratio for the insertion of a segment becomes then

$$R_{c \rightarrow c'} = \min \left(1, \frac{\beta l_{\text{max}}}{k_\sigma + 1} \left| \frac{d_{c'}}{d_c} \frac{t_{c'}}{t_c} \right| \right). \quad (72)$$

For the impurity Green function, here the most important observable, the direct average yields

$$\langle \hat{O} \rangle_c = \langle G_{d,d}^\sigma \rangle_c = \sum_{\sigma'} \sum_{i=1}^{k_\sigma} \sum_{j=1}^{k_\sigma} \Delta(\tau, \tau_{\sigma' j} - \bar{\tau}_{\sigma' j}) (M^{k'_\sigma})_{\sigma' j, \sigma' i} \delta_{\sigma, \sigma' j} \delta_{\sigma, \bar{\sigma}' i} \quad (73)$$

where $M^k = (\mathcal{F}^k)^{-1}$ is the inverse of the hybridization matrix and

$$\Delta(\tau, \tau') = -\frac{1}{\beta} \begin{cases} \delta(\tau - \tau') & \tau' > 0 \\ -\delta(\tau - (\tau' + \beta)) & \tau' < 0 \end{cases}. \quad (74)$$

One can verify that at order $k = 1$ this indeed returns Eq. (66).

2.4 Hartree-Fock (LDA+ U) versus DMFT approximation

Let us now compare the exact solution of the Hubbard dimer with the result of the Hartree-Fock (HF) approximation, which is the basis of the LDA+ U method. It consists in replacing

$$\hat{H}_U = U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rightarrow \hat{H}_U^{\text{HF}} = U \sum_i (\hat{n}_{i\uparrow} \bar{n}_{i\downarrow} + \hat{n}_{i\downarrow} \bar{n}_{i\uparrow} - \bar{n}_{i\uparrow} \bar{n}_{i\downarrow}), \quad (75)$$

where $\bar{n}_{i\sigma}$ is the HF expectation value of the operator $\hat{n}_{i\sigma}$. It is convenient to define

$$\begin{aligned} n_i &= \bar{n}_{i\uparrow} + \bar{n}_{i\downarrow} & n &= \frac{1}{2}(n_1 + n_2) & \delta n &= \frac{1}{2}(n_1 - n_2) \\ m_i &= \frac{1}{2}(\bar{n}_{i\uparrow} - \bar{n}_{i\downarrow}) & m_+ &= \frac{1}{2}(m_1 + m_2) & m_- &= \frac{1}{2}(m_1 - m_2) \end{aligned}$$

Inverting these relations, in the absence of charge disproportionation ($\delta n = 0$), we find

$$\bar{n}_{i\uparrow} = (m_+ + (-1)^{i-1} m_-) + n/2 \quad \bar{n}_{i\downarrow} = -(m_+ + (-1)^{i-1} m_-) + n/2.$$

The Hartree-Fock version of the Hubbard dimer Hamiltonian equals the non-interacting Hamiltonian plus a shift of the on-site level. This shift depends on the site and the spin

$$\hat{H}_{\text{HF}} = \sum_{i\sigma} (\varepsilon_d + \Delta_{i\sigma}) \hat{n}_{i\sigma} - t \sum_{\sigma} (c_{1\sigma}^\dagger c_{2\sigma} + c_{2\sigma}^\dagger c_{1\sigma}) - \Delta_0 \quad (76)$$

$$\Delta_0 = 2U \left(\frac{n^2}{4} - m_+^2 - m_-^2 \right) \quad (77)$$

$$\Delta_{i\sigma} = U \left(\frac{1}{2}n - \sigma (m_+ + (-1)^{i-1} m_-) \right), \quad (78)$$

where $\sigma = +1$ for spin up and $\sigma = -1$ for spin down. Thus we can write immediately the local Green function matrix for site i . It is convenient to use the site basis, hence, to calculate the matrix $G_{i,i'}^\sigma(i\nu_n)$. Then we have

$$G_{i,i}^\sigma(i\nu_n) = \begin{pmatrix} i\nu_n - (\varepsilon_d - \mu + \Sigma_{1,1}^\sigma(i\nu_n)) & t \\ t & i\nu_n - (\varepsilon_d - \mu + \Sigma_{2,2}^\sigma(i\nu_n)) \end{pmatrix}_{i,i}^{-1} \quad (79)$$

where we introduced the diagonal self-energy matrix $\Sigma_{i,i'}^\sigma(i\nu_n) = \Delta_{i\sigma} \delta_{i,i'}$. This shows that the Hartree-Fock self-energy is not dependent on the frequency, i.e., Hartree-Fock is a *static* mean-field approach. The value of the parameters m_+ and m_- have to be found solving the *self-consistent equations*

$$\bar{n}_{i\sigma} = \langle \hat{n}_{i\sigma} \rangle = \frac{1}{\beta} \sum_n e^{-i\nu_n 0^-} G_{i,i}^\sigma(i\nu_n) = G_{i,i}^\sigma(0^-). \quad (80)$$

Since the exact ground state of the Hubbard dimer is a singlet, let us consider first the anti-ferromagnetic Hartree-Fock solution ($m_+ = 0$ and $m_- \neq 0$). In this case, the Hartree-Fock self-energy depends on the site and the interaction couples states with different k . This becomes

clear rewriting the Green-function matrix in the basis of the bonding ($k = 0$) and anti-bonding ($k = \pi$) creation/annihilation operators

$$G^\sigma(i\nu_n) = \frac{1}{2} \begin{pmatrix} i\nu_n - (\varepsilon_d - t - \mu + \frac{1}{2} \sum_i \Sigma_{i\sigma}(i\nu_n)) & \frac{1}{2} \sum_i (-1)^{i-1} \Sigma_{i\sigma}(i\nu_n) \\ \frac{1}{2} \sum_i (-1)^{i-1} \Sigma_{i\sigma}(i\nu_n) & i\nu_n - (\varepsilon_d + t - \mu + \frac{1}{2} \sum_i \Sigma_{i\sigma}(i\nu_n)) \end{pmatrix}^{-1}. \quad (81)$$

The diagonal terms of the matrix $\Sigma_{k,k'}^\sigma(i\nu_n)$ are thus identical

$$\Sigma_{0,0}^\sigma(i\nu_n) = \Sigma_{\pi,\pi}^\sigma(i\nu_n) = \frac{1}{2} (\Sigma_{1,1}^\sigma(i\nu_n) + \Sigma_{2,2}^\sigma(i\nu_n)). \quad (82)$$

The off-diagonal terms $\Sigma_{0,\pi}^\sigma(i\nu_n)$ and $\Sigma_{\pi,0}^\sigma(i\nu_n)$ are not zero, however. This tells us that, by introducing the antiferromagnetic HF correction, we lower the symmetry of the system. Let us now calculate explicitly the eigenstates for different fillings. It is sufficient to diagonalize $\hat{H}_1$, the Hamiltonian in the 1-electron sector; the many-electron states can be obtained by filling the one-electron states respecting the Pauli principle. The Hamiltonian $\hat{H}_1$ can be written as $\hat{H}_1 = \hat{H}'_1 + \varepsilon_d \hat{N} - \Delta_0$, and, in the antiferromagnetic case we then have

$$\hat{H}'_1 = \begin{pmatrix} U(\frac{1}{2}n - m_-) & -t & 0 & 0 \\ -t & U(\frac{1}{2}n + m_-) & 0 & 0 \\ 0 & 0 & U(\frac{1}{2}n + m_-) & -t \\ 0 & 0 & -t & U(\frac{1}{2}n - m_-) \end{pmatrix}. \quad (83)$$

This leads to the (normalized) states

$ 1\rangle_l$	$E_l(1)$
$ 1\rangle_3 = a_2 1, 1/2, \uparrow\rangle_1 - a_1 1, 1/2, \uparrow\rangle_2$	$\varepsilon_0(1) + \Delta_1(t, U)$
$ 1\rangle_2 = a_1 1, 1/2, \downarrow\rangle_1 - a_2 1, 1/2, \downarrow\rangle_2$	$\varepsilon_0(1) + \Delta_1(t, U)$
$ 1\rangle_1 = a_1 1, 1/2, \uparrow\rangle_1 + a_2 1, 1/2, \uparrow\rangle_2$	$\varepsilon_0(1) - \Delta_1(t, U)$
$ 1\rangle_0 = a_2 1, 1/2, \downarrow\rangle_1 + a_1 1, 1/2, \downarrow\rangle_2$	$\varepsilon_0(1) - \Delta_1(t, U)$

where $\varepsilon_0(1) = \varepsilon_d + U(1/2 + 2m_-^2 - n^2/2)$ and $a_1^2 = \frac{1}{2} \left(1 + \frac{Um_-}{\Delta_1(t, U)}\right)$, while $\Delta_1(t, U) = \sqrt{(m_-U)^2 + t^2}$. At half filling, if we assume that only the ground doublet is occupied, solving the self-consistent equations (80) yields

$$m_- = 0 \quad \text{or} \quad m_- = \frac{1}{2} \sqrt{1 - \frac{4t^2}{U^2}}. \quad (84)$$

As a result, for the non-trivial solution ($m_- \neq 0$), the gap at half filling is $E_g^{\text{HF}} = 2\Delta_1(t, U) = U$. In Fig. 6, Hartree-Fock and exact spectral function are compared for a specific parameter choice. Looking at the size of the gap only, one could naively infer that, for the case shown, Hartree-Fock is an excellent approximation, better than DMFT (see Fig. 3, top-right panel). This would

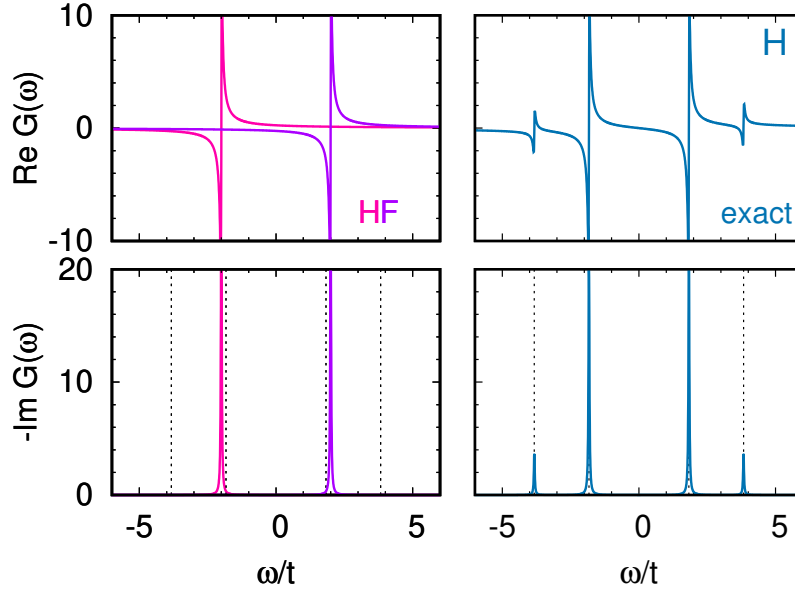


Fig. 6: Local retarded Green function $G^\sigma(\omega)$ of the Hubbard dimer ($t = 1, U = 4$) at half filling and in the zero temperature limit. Left: Antiferromagnetic Hartree-Fock (HF) approximation; green and violet distinguish $G^\uparrow(\omega)$ and $G^\downarrow(\omega)$ at a given site. Right: Exact local Green function, for which $G^\uparrow(\omega) = G^\downarrow(\omega)$. Dashed lines: Poles of the exact local Green function.

be, however, the wrong conclusion. Although, due to the small coordination number, in this specific case, DMFT indeed sizably underestimates the gap, it nevertheless captures the essential nature of the ground state. Instead, while HF, in this specific case,⁶ gives an almost exact gap, it does it via a qualitatively incorrect description (antiferromagnetic instead of singlet ground state). Increasing the coordination number, DMFT approaches the exact solution both in the paramagnetic and magnetic phase, while HF does not. Let us analyze in more detail the differences between the HF approximation and the exact solution of the Hubbard dimer. The HF Hamiltonian for two electrons is given by $\hat{H}_2 = \hat{H}'_2 + \varepsilon_d \hat{N} - \Delta_0$, and

$$\hat{H}'_2 = \begin{pmatrix} U & 0 & 0 & -2Um_- & 0 & 0 \\ 0 & U(1-2m_+) & 0 & 0 & 0 & 0 \\ 0 & 0 & U(1+2m_+) & 0 & 0 & 0 \\ -2Um_- & 0 & 0 & U & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & U \end{pmatrix}. \quad (85)$$

⁶In real materials the Hartree-Fock approximation often overestimates the size of the gap, however.

For $m_+=0$ and $m_- \neq 0$ (antiferromagnetic solution) the eigenvalues and eigenvectors are

$ 2\rangle_l$	$E_l(2)$
$ 2\rangle_5 = \frac{1}{\sqrt{2}} \left(2, 0, 0\rangle_0 + a_2 2, 1, 0\rangle - \frac{a_1}{\sqrt{2}} (2, 0, 0\rangle_1 + 2, 0, 0\rangle_2) \right)$	$\varepsilon_0(2) + 2\Delta_1(t, U)$
$ 2\rangle_4 = \frac{1}{\sqrt{2}} (2, 0, 0\rangle_1 - 2, 0, 0\rangle_2)$	$\varepsilon_0(2)$
$ 2\rangle_3 = 2, 1, 1\rangle$	$\varepsilon_0(2)$
$ 2\rangle_2 = 2, 1, -1\rangle$	$\varepsilon_0(2)$
$ 2\rangle_1 = a_1 2, 1, 0\rangle + a_2 \frac{1}{\sqrt{2}} (2, 0, 0\rangle_1 + 2, 0, 0\rangle_2)$	$\varepsilon_0(2)$
$ 2\rangle_0 = \frac{1}{\sqrt{2}} \left(2, 0, 0\rangle_0 - a_2 2, 1, 0\rangle + \frac{a_1}{\sqrt{2}} (2, 0, 0\rangle_1 + 2, 0, 0\rangle_2) \right)$	$\varepsilon_0(2) - 2\Delta_1(t, U)$

where $\varepsilon_0(2)=2\varepsilon_d + U(1 + 2m_-^2 - n^2/2)$, and $a_1^2 = t^2/\Delta_1^2(t, U)$. The antiferromagnetic Hartree-Fock ground state has an overlap with the correct ground state, however incorrectly mixes triplet and singlet states, thus breaking the rotational symmetry of the model. For this reason, its energy, in the large U limit, is $2\varepsilon_d - 2t^2/U$ and not $2\varepsilon_d - 4t^2/U$ as in the exact case.

For the ferromagnetic solution of the Hartree-Fock equations ($m_-=0$ and $m_+ \neq 0$) the eigenvalues and eigenvectors are instead

$ 2\rangle_l$	$E_l(2)$
$ 2\rangle_5 = 2, 1, -1\rangle$	$\varepsilon_0^+(2) + 2Um_+$
$ 2\rangle_4 = \frac{1}{\sqrt{2}} \left(2, 0, 0\rangle_0 - \frac{1}{\sqrt{2}} (2, 0, 0\rangle_1 + 2, 0, 0\rangle_2) \right)$	$\varepsilon_0^+(2) + 2t$
$ 2\rangle_3 = \frac{1}{\sqrt{2}} (2, 0, 0\rangle_1 - 2, 0, 0\rangle_2)$	$\varepsilon_0^+(2)$
$ 2\rangle_2 = 2, 1, 0\rangle$	$\varepsilon_0^+(2)$
$ 2\rangle_1 = \frac{1}{\sqrt{2}} \left(2, 0, 0\rangle_0 + \frac{1}{\sqrt{2}} (2, 0, 0\rangle_1 + 2, 0, 0\rangle_2) \right)$	$\varepsilon_0^+(2) - 2t$
$ 2\rangle_0 = 2, 1, 1\rangle$	$\varepsilon_0^+(2) - 2Um_+$

where $\varepsilon_0^+(2)=2\varepsilon_d + U(1 + 2m_+^2 - n^2/2)$. The ferromagnetic Hartree-Fock correction thus yields an incorrect sequence of levels; the ground state for large U/t , indicated as $|2\rangle_0$ in the table, has no overlap with the exact ground state of the Hubbard dimer. It is, instead, one of the states of the first excited triplet. An important observation is that, despite the errors, the energy difference between ferro- and antiferro-magnetic ground state is

$$E_F - E_{AF} \sim \frac{2t^2}{U}, \quad (86)$$

which is indeed the correct value in the small t/U limit. It does not correspond, however, to the actual singlet-triplet excitation energy, $\Gamma \sim 4t^2/U$.

We can now directly compare the **Hartree-Fock** and the **dynamical mean-field approximation** for the two-site Hubbard model. Both approaches are based on the solution of self-consistent mean-field-type equations, and therefore they are both mean-field methods. In Hartree-Fock the self-energy is frequency-independent (static), while in DMFT depends on the frequency (dynamical). In Hartree-Fock the self-energy can be site-dependent, as we have seen in the antiferromagnetic case. In DMFT, allowing for a site-dependent self-energy leads to cluster or cluster-like extensions of the method. If the same unit cell is used in DMFT and Hartree-Fock, we can identify another relation between the two methods. In the infinite-frequency limit, the DMFT self-energy equals the Hartree-Fock self-energy. This can be shown analytically in a simple way. For the case of the two-site Hubbard model, the expansion of the local lattice Green function at high frequency in the local-self-energy approximation is

$$G_{i,i}^{\sigma}(i\nu_n) = \frac{1}{i\nu_n} - \frac{(\mu - \varepsilon_d - \Sigma_l^{\sigma}(\infty))}{(i\nu_n)^2} + \mathcal{O}\left(\frac{1}{(i\nu_n)^3}\right). \quad (87)$$

Instead, the expansion of the impurity Green-function can be written as

$$\begin{aligned} G_{d,d}^{\sigma}(i\nu_n) &= \int_0^{\beta} d\tau e^{i\nu_n\tau} G_{d,d}^{\sigma}(\tau) \\ &= \frac{1}{i\nu_n} \int_0^{\beta} d\tau \frac{d e^{i\nu_n\tau}}{d\tau} G_{d,d}^{\sigma}(\tau) = \frac{1}{i\nu_n} \left(-G_{d,d}^{\sigma}(\beta^-) - G_{d,d}^{\sigma}(0^+) - \int_0^{\beta} d\tau e^{i\nu_n\tau} \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right) \\ &= \frac{1}{i\nu_n} + \frac{1}{(i\nu_n)^2} \left(\left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{\beta^-} + \left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{0^+} + \int_0^{\beta} d\tau e^{i\nu_n\tau} \frac{d^2 G_{d,d}^{\sigma}(\tau)}{d\tau^2} \right) \\ &= \frac{1}{i\nu_n} + \frac{\left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{\beta^-} + \left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{0^+}}{(i\nu_n)^2} + \mathcal{O}\left(\frac{1}{(i\nu_n)^3}\right) \end{aligned} \quad (88)$$

where

$$\left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{\beta^-} + \left. \frac{d G_{d,d}^{\sigma}(\tau)}{d\tau} \right|_{0^+} = -\left\langle \left\{ [\hat{H}_A - \mu \hat{N}, c_{d\sigma}], c_{d\sigma}^{\dagger} \right\} \right\rangle = -\mu + \varepsilon_d + U \langle \hat{n}_{-\sigma} \rangle. \quad (89)$$

From this result we can conclude that, assuming self-consistency has been reached,

$$\Sigma_l^{\sigma}(\infty) = U \langle \hat{n}_{-\sigma} \rangle. \quad (90)$$

This is exactly the Hartree-Fock expression of the self-energy that we found earlier, and indeed equals the **infinite-frequency limit** of the DMFT self-energy we previously calculated; however, the occupations $\langle \hat{n}_{-\sigma} \rangle$ in DMFT and Hartree-Fock calculations are typically not the same. In the case of the dimer, DMFT yields $\langle \hat{n}_{-\sigma} \rangle = 1/2$; this would correspond in Hartree-Fock to a (trivial) solution in which the self-energy merely shifts all the energy levels by the same amount, and has therefore no relevant effects. These conclusions also apply to the LDA+ U method, which is based on the Hartree-Fock treatment of the local Coulomb interaction. A more detailed discussion of differences and relation between the LDA+ U and LDA+DMFT method can be found in my chapter in Ref. [14].

2.5 DMFT for the one-band Hubbard model

The Hubbard Hamiltonian (8) is in principle the simplest model for the description of the Mott metal-insulator transition. In the tight-binding approximation it becomes

$$\hat{H} = \varepsilon_d \sum_{\sigma i} \hat{n}_{i\sigma} - t \sum_{\sigma \langle ii' \rangle} c_{i\sigma}^\dagger c_{i'\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (91)$$

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 (91) 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

$$\hat{H}^A = \underbrace{\sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}}^s \hat{n}_{\mathbf{k}\sigma}}_{\hat{H}_{\text{bath}}} + \underbrace{\sum_{\mathbf{k}\sigma} \left(V_{\mathbf{k}}^s c_{\mathbf{k}\sigma}^\dagger c_{d\sigma} + \text{h.c.} \right)}_{\hat{H}_{\text{hyb}}} + \underbrace{\varepsilon_d \sum_{\sigma} \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow}}_{\hat{H}_{\text{imp}}}. \quad (92)$$

In this model the on-site Coulomb repulsion U appears only in the impurity Hamiltonian, $\hat{H}_{\text{imp}}$, while the terms $\hat{H}_{\text{bath}}$ and $\hat{H}_{\text{hyb}}$, 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 $G_{d,d}^\sigma(i\nu_n)$. From the impurity Dyson equation we can calculate the impurity self-energy

$$\Sigma_A^\sigma(i\nu_n) = \left(G_{d,d}^{0\sigma}(i\nu_n) \right)^{-1} - \left(G_{d,d}^\sigma(i\nu_n) \right)^{-1}. \quad (93)$$

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

$$G_{i_c, i_c}^\sigma(i\nu_n) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \frac{1}{i\nu_n + \mu - \varepsilon_{\mathbf{k}} - \Sigma_A^\sigma(i\nu_n)}, \quad (94)$$

where $N_{\mathbf{k}}$ is the number of $\mathbf{k}$ 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

$$G_{i_c, i_c}^\sigma(i\nu_n) \sim G_{d,d}^\sigma(i\nu_n). \quad (95)$$

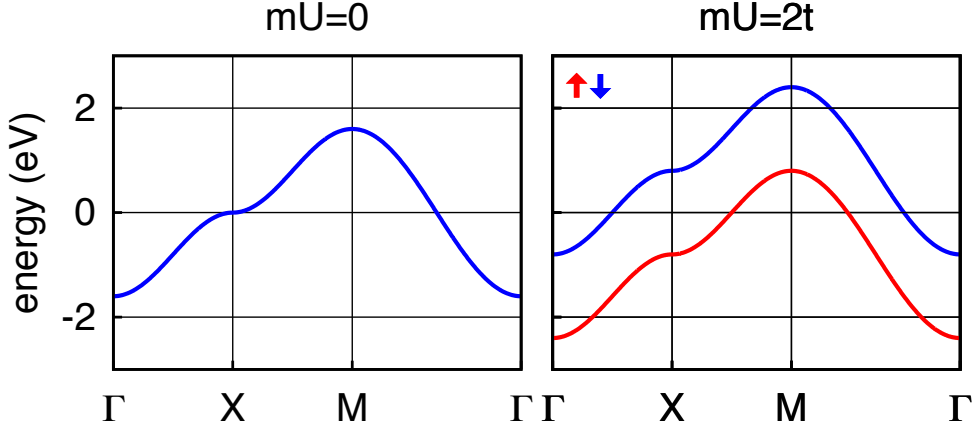


Fig. 7: The metal-insulator transition in ferromagnetic Hartree-Fock. The calculation is for a square lattice tight-binding model with dispersion $\varepsilon_{\mathbf{k}} = -2t(\cos k_x + \cos k_y)$.

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.⁷ If we chose the same primitive cell as in DMFT, the Hartree-Fock self-energy matrix is

$$\Sigma_{i,i'}^{\sigma}(i\nu_n) = U \left(\frac{n}{2} - \sigma m \right) \delta_{i,i'}, \quad (96)$$

where $\sigma = +1$ for spin up and $\sigma = -1$ 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

$$\hat{H}_{\text{HF}} = \sum_{\mathbf{k}\sigma} \left[\varepsilon_{\mathbf{k}} + U \left(\frac{1}{2} - \sigma m \right) \right] \hat{n}_{\mathbf{k}\sigma}. \quad (97)$$

This shows that $h_{\text{eff}} = 2Um$ plays the role of an effective magnetic field (Weiss field). The self-consistency criterion is

$$\bar{n}_{\sigma} = \bar{n}_{i\sigma} = \langle \hat{n}_{i\sigma} \rangle_{\text{HF}}, \quad (98)$$

where the expectation value $\langle \hat{n}_{i\sigma} \rangle_{\text{HF}}$ is calculated using the Hamiltonian $\hat{H}_{\text{HF}}$, which in turn depends on $\bar{n}_{\sigma}$ via m . This gives the self-consistency equation

$$m = \frac{1}{2} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}\sigma} \frac{\sigma e^{-\beta(\varepsilon_{\mathbf{k}} + U(\frac{1}{2} - \sigma m) - \mu)}}{1 + e^{-\beta(\varepsilon_{\mathbf{k}} + U(\frac{1}{2} - \sigma m) - \mu)}}. \quad (99)$$

If we set $m=0$ the equation is satisfied; for such a trivial solution the static mean-field correction in Eq. (97) 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. 7. The static mean-field correction in Eq. (97) equals the contribution of the Hartree diagram to the self-energy, $\Sigma_{\text{H}}^{\sigma}(i\nu_n) = U\bar{n}_{-\sigma}$. In many-body perturbation theory, however, $\bar{n}_{\sigma} = 1/2$, i.e., $m=0$.

⁷Keeping in mind that many self-consistent solutions obtained with the Hartree-Fock method are spurious.

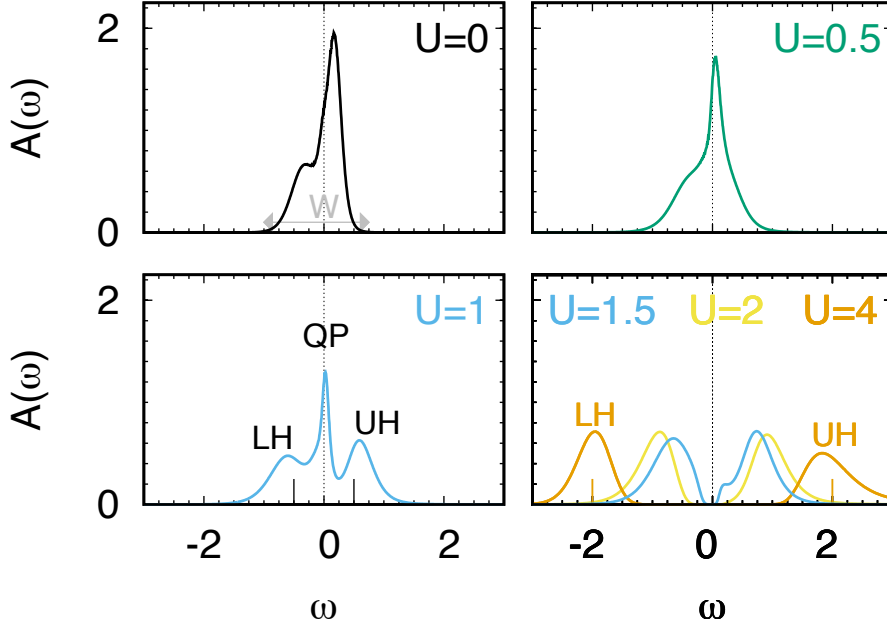


Fig. 8: *VOMoO₄: LDA+DMFT spectral function at finite temperature for $0 \leq U \leq 4$. Energies are in eV and spectral functions in states/spin/eV. The calculations have been done using a continuous-time hybridization-expansion QMC solver [13]. A detailed LDA+DMFT study of the electronic and magnetic properties of VOMoO₄ can be found in Ref. [15].*

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 mU 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 $\mathcal{G}_{i,i}^{\sigma}(i\nu_n)$. The emerging picture of the Mott transition is described in Fig. 8 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 $\Sigma_A^{\sigma}(\omega)=0$ 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. 8. 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. 8. 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

$$\Sigma_A^{\sigma}(\omega + i0^+) \sim \frac{U}{2} + \frac{U^2}{4} \frac{a(t, U)}{\omega + i0^+}. \quad (100)$$

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.

2.6 DMFT for multi-orbital models

The multi-orbital Hubbard-like Hamiltonian has the form

$$\hat{H} = \hat{H}_0 + \hat{H}_U \quad (101)$$

$$\hat{H}_0 = - \sum_{ii'} \sum_{\sigma\sigma'} \sum_{mm'} t_{m\sigma,m'\sigma'}^{i,i'} c_{im\sigma}^\dagger c_{i'm'\sigma'} \quad (102)$$

$$\hat{H}_U = \frac{1}{2} \sum_i \sum_{\sigma\sigma'} \sum_{mm'} \sum_{pp'} U_{mpm'p'} c_{im\sigma}^\dagger c_{ip\sigma'}^\dagger c_{ip'\sigma'} c_{im'\sigma}, \quad (103)$$

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 form, 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

$$\begin{aligned} \varepsilon_{\mathbf{k}} &\rightarrow (H_{\mathbf{k}})_{m\sigma,m'\sigma'} & (i\nu_n + \mu) &\rightarrow (i\nu_n + \mu) \hat{1}_{m\sigma,m'\sigma'} \\ t_{m\sigma,m'\sigma'}^{i,i'} &\rightarrow t_{m\sigma,m'\sigma'}^{i,i'} & \varepsilon_d &\rightarrow \varepsilon_{m\sigma,m'\sigma'}^{i,i'} = -t_{m\sigma,m'\sigma'}^{i,i} \end{aligned}$$

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

$$\mathcal{G}^\sigma(i\nu_n) \rightarrow \mathcal{G}_{m,m'}^{\sigma,\sigma'}(i\nu_n) \quad G^\sigma(i\nu_n) \rightarrow G_{m,m'}^{\sigma,\sigma'}(i\nu_n) \quad \Sigma^\sigma(i\nu_n) \rightarrow \Sigma_{m,m'}^{\sigma,\sigma'}(i\nu_n).$$

The corresponding generalization of the self-consistency loop is shown schematically in Fig. 9. 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, typically, going beyond a limited 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 [13]. 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 case and in absence of spin-orbit interaction or external fields, an obvious symmetry to exploit

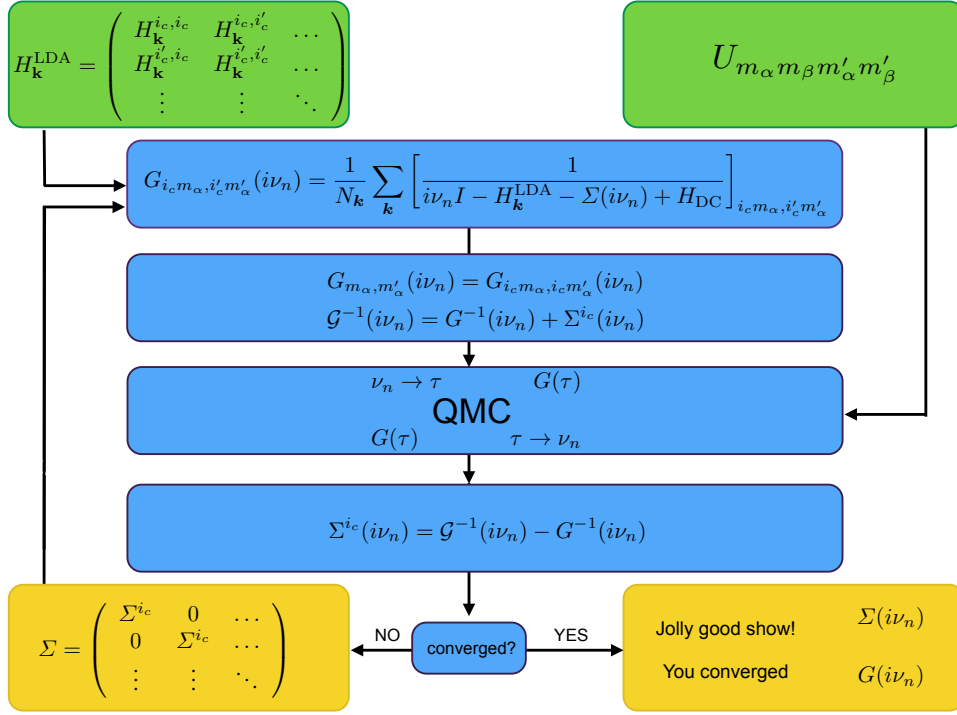


Fig. 9: LDA+DMFT self-consistency loop. The one-electron Hamiltonian is built in the basis of Bloch states obtained from localized Wannier functions, for example in the local-density approximation (LDA); this gives $H_{\mathbf{k}}^{\text{LDA}}$. The set $\{i_c\}$ labels the equivalent correlated sites 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.

is the rotational invariance of spins, from which follows

$$A_{m, m'}^{\sigma, \sigma'}(i\nu_n) = \delta_{\sigma, \sigma'} A_{m, m'}(i\nu_n), \quad (104)$$

where $A = \mathcal{G}, G, \Sigma$. In addition, if we use a basis of real functions, the local Green-function matrices are real and symmetric in imaginary time τ , hence

$$A_{m, m'}^{\sigma, \sigma'}(i\nu_n) = \delta_{\sigma, \sigma'} A_{m, m'}(i\nu_n) = \delta_{\sigma, \sigma'} A_{m', m}(i\nu_n). \quad (105)$$

Finally, often the unit cell contains several equivalent correlated sites, indicated as $\{i_c\}$ in Fig. 9. 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

$$A_{m, m'}^{\sigma, \sigma'}(i\nu_n) = \delta_{\sigma, \sigma'} A_{m, m}(i\nu_n) \delta_{m, m'}. \quad (106)$$

3 Building materials-specific many-body models

3.1 Model construction

How do we build realistic Hubbard-like models for correlated materials? The state-of-the-art approach relies on constructing, for a given system, *materials-specific* Kohn-Sham Wannier functions $\varphi_{im\sigma}^{\text{KS}}(\mathbf{r})$. These can be obtained via electronic structure calculations based on density-functional theory [8–10], e.g., in the LDA approximation.⁸ After we constructed the complete one-electron basis, the first steps in model-building are those described in the introduction. We recall here the essential points and then discuss the next stage. The many-body Hamiltonian can be expressed as $\hat{H} = \hat{H}_0 + \hat{H}_U - \hat{H}_{\text{DC}}$, with $\hat{H}_0 = \hat{H}^{\text{LDA}}$ and

$$\hat{H}_0 = - \sum_{ii'} \sum_{\sigma\sigma'} \sum_{mm'} t_{m\sigma, m'\sigma'}^{i, i'} c_{im\sigma}^\dagger c_{i'm'\sigma'}, \quad (107)$$

$$\hat{H}_U = \frac{1}{2} \sum_{ii'jj'} \sum_{\sigma\sigma'} \sum_{mm'pp'} U_{mp\ m'p'}^{ij\ i'j'} c_{im\sigma}^\dagger c_{jp\sigma'}^\dagger c_{j'p'\sigma'} c_{i'm'\sigma}. \quad (108)$$

The hopping integrals in $\hat{H}_0$ are calculated replacing the electron-nuclei interaction, $v_{\text{en}}(\mathbf{r})$, with the self-consistent DFT reference potential, $v_{\text{R}}(\mathbf{r}) = v_{\text{en}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r})$, defined in Eq. (5). The latter includes the long-range Hartree term, $v_{\text{H}}(\mathbf{r})$, and the exchange-correlation contribution, $v_{\text{xc}}(\mathbf{r})$; to avoid including the *effects* of these terms twice, we thus introduce the double-counting correction, $\hat{H}_{\text{DC}}$, so that

$$\hat{H}_U \rightarrow \Delta\hat{H}_U = \hat{H}_U - \hat{H}_{\text{DC}}.$$

Unfortunately we do not know which important correlation effects are included in $\hat{H}_0$ via $v_{\text{R}}(\mathbf{r})$, 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 two categories](#): *uncorrelated* (or light) and *correlated* (or heavy); 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 [8], 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 transition-metal oxides or f for heavy-fermion systems) and label their states with the atomic quantum numbers l and $m = -l, \dots, l$ of that shell. Thus

$$U_{mp\ m'p'}^{ij\ i'j'} \sim \begin{cases} U_{mp\ m'p'}^l & ij\ i'j' = iiii \quad \wedge \quad mp, m'p' \in l \\ 0 & ij\ i'j' \neq iiii \quad \vee \quad mp, m'p' \notin l. \end{cases}$$

⁸Using GGA or similar functionals in place of LDA yields minor differences in the many-body Hamiltonian; on the other hand, using LDA+ U or similar approximations yields Hartree-Fock-like effects that would have to be subtracted via the double-counting correction.

Within this approximation $\Delta\hat{H}_U$ is replaced by $\Delta\hat{H}_U^l = \hat{H}_U^l - \hat{H}_{\text{DC}}^l$, where $\hat{H}_{\text{DC}}^l$ is, e.g., given by the static mean-field contribution of $\hat{H}_U^l$. There is a drawback in this procedure, however. By splitting electrons into light and heavy, 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) [8–10]. 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, in order to build the final Hamiltonian for DMFT calculations, 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., keeping 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

$$\hat{H}^{\text{LDA}} = - \sum_{ii'} \sum_{\sigma\sigma'} \sum_{m_\alpha m'_\alpha} t_{m_\alpha\sigma, m'_\alpha\sigma'}^{i,i'} c_{im_\alpha\sigma}^\dagger c_{i'm'_\alpha\sigma'}. \quad (109)$$

The local *screened* Coulomb interaction for this set of orbitals is the on-site tensor

$$\hat{H}_U^l = \frac{1}{2} \sum_i \sum_{\sigma\sigma'} \sum_{m_\alpha m'_\alpha} \sum_{m_\beta m'_\beta} U_{m_\alpha m_\beta m'_\alpha m'_\beta} c_{im_\alpha\sigma}^\dagger c_{im_\beta\sigma'}^\dagger c_{im'_\beta\sigma'} c_{im'_\alpha\sigma}. \quad (110)$$

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.⁹

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

⁹The choice might influence how severe the QMC sign problem is, however.

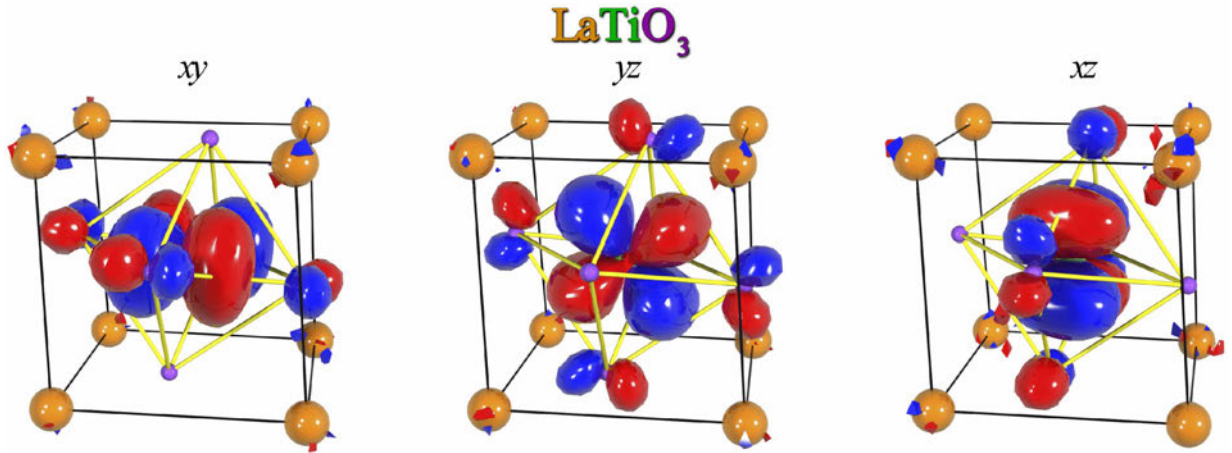


Fig. 10: NMTO Wannier-like orbitals for t_{2g} states in LaTiO_3 obtained via massive downfolding to the t_{2g} bands. The t_{2g} -like orbitals have O p tails at the neighboring O sites reflecting the distortions of the lattice. The figure has been taken from Ref. [16].

necessary to include non-local Coulomb terms. The effect of downfolding on the localization of Wannier functions is illustrated for example in Fig. 10. Finally, another disadvantage of massive downfolding is that the energy window in which the model is valid is more narrow.

Considering all advantages and disadvantages, 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. 10, 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-nuclei interaction $v_{en}(\mathbf{r})$. 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 to describe 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 a 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. 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 [16], the maximal-localization scheme [17], or projectors. Fig. 10 shows examples of NMTO-based Wannier functions.

3.2 Localization of the basis and range of the Coulomb interaction

No matter what construction procedure is used, a common characteristic of *ab-initio* Wannier functions is that they are site-centered and localized.¹⁰ 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 couple 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 $\psi_{im\sigma}(\mathbf{r})\overline{\psi_{i'm'\sigma'}(\mathbf{r})} \sim \delta_{i,i'}\delta(\mathbf{r}-\mathbf{T}_i)$. Even for such a basis choice, the unscreened Coulomb interaction is *not* local. It is given by

$$U_{mp\ m'p'}^{ijj'} \propto \frac{\delta_{i,i'}\delta_{j,j'}}{|\mathbf{T}_i - \mathbf{T}_j|}, \quad (111)$$

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 $\hat{H}_{\text{DC}}$. The extreme case of the $\delta(\mathbf{r}-\mathbf{T}_i)$ 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 down-folding 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. 11 for a simple example of two quantum well potentials.¹¹ 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.

¹⁰Differences in localizations between the various construction procedures are actually small for the purpose of many-body calculations, provided that the same bands are spanned with the same accuracy.

¹¹Another, 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.

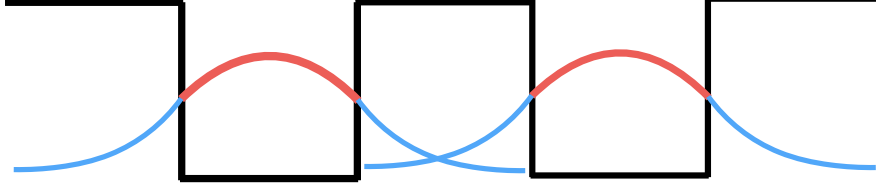


Fig. 11: *The problem of two quantum wells. The figure shows (schematically) for each well the wavefunction of a bound state. If we consider only the part of the wavefunction inside its own well (red in the figure), the differential overlap (and hence the hopping integral) between functions centered on different wells would be zero.*

3.3 Hubbard Hamiltonians for t_{2g} and e_g systems

Several strongly-correlated transition-metal oxides can be described via minimal materials-specific Hubbard-like models which involve only t_{2g} or only e_g bands. A representative system of this kind is the layered perovskite Sr_2RuO_4 with the $4d\ t_{2g}^4$ electronic configuration. Its crystal structure is shown in Fig. 12 (left side), together with the associated LDA bands crossing the Fermi level (top-right panel). Due to the layered structure the xz and yz bands are quasi one-dimensional and the xy band is quasi two-dimensional. Thus, the t_{2g} bands give rise, in first approximation, to a Fermi surface made of four crossing lines (from the xz, yz bands) and a circle (from the xy band), shown schematically in the bottom-right panel of Fig. 12. Experimentally, Sr_2RuO_4 is a correlated metal down to 1.5 K; below this temperature it becomes an anomalous superconductor. Other representative cases of t_{2g} systems are the Mott insulating perovskites LaTiO_3 and YTiO_3 , with the electronic configuration $3d\ t_{2g}^1$. A paradigmatic e_g system is instead the orbitally ordered insulator KCuF_3 , with the electronic configuration $t_{2g}^6 e_g^3$. For all these materials, if we massively downfold all LDA bands but the t_{2g} (or the e_g), the resulting 3-band (or 2-band) generalized Hubbard model takes the form

$$\begin{aligned} \hat{H} = & \sum_{i\sigma} \sum_{mm'} \varepsilon_{m,m'} c_{im\sigma}^\dagger c_{im'\sigma} - \sum_{\sigma} \sum_{i \neq i'} \sum_{mm'} t_{m,m'}^{i,i'} c_{im\sigma}^\dagger c_{i'm'\sigma} \\ & + U \sum_{i m} \hat{n}_{im\uparrow} \hat{n}_{im\downarrow} + \frac{1}{2} \sum_{\substack{i\sigma\sigma' \\ m \neq m'}} (U - 2J - J\delta_{\sigma,\sigma'}) \hat{n}_{im\sigma} \hat{n}_{im'\sigma'} \\ & - J \sum_{i m \neq m'} \left(c_{im\uparrow}^\dagger c_{im\downarrow}^\dagger c_{im'\uparrow} c_{im'\downarrow} + c_{im\uparrow}^\dagger c_{im\downarrow} c_{im'\downarrow}^\dagger c_{im'\uparrow} \right), \end{aligned} \quad (112)$$

where $m, m' = xy, yz, xz$ for t_{2g} and $m, m' = 3z^2 - r^2, x^2 - y^2$ for e_g . The parameters U and J are the direct and exchange screened Coulomb integrals. The Coulomb interaction $\hat{H}_U$ is here assumed to have full $O(3)$ rotational symmetry, as in the atomic limit.¹² The first two terms of

¹²A derivation of the Coulomb interaction tensor for the free atom can be found in my chapter in Ref. [8]. There the difference in the values of U and J for the t_{2g} and e_g orbitals is also discussed.

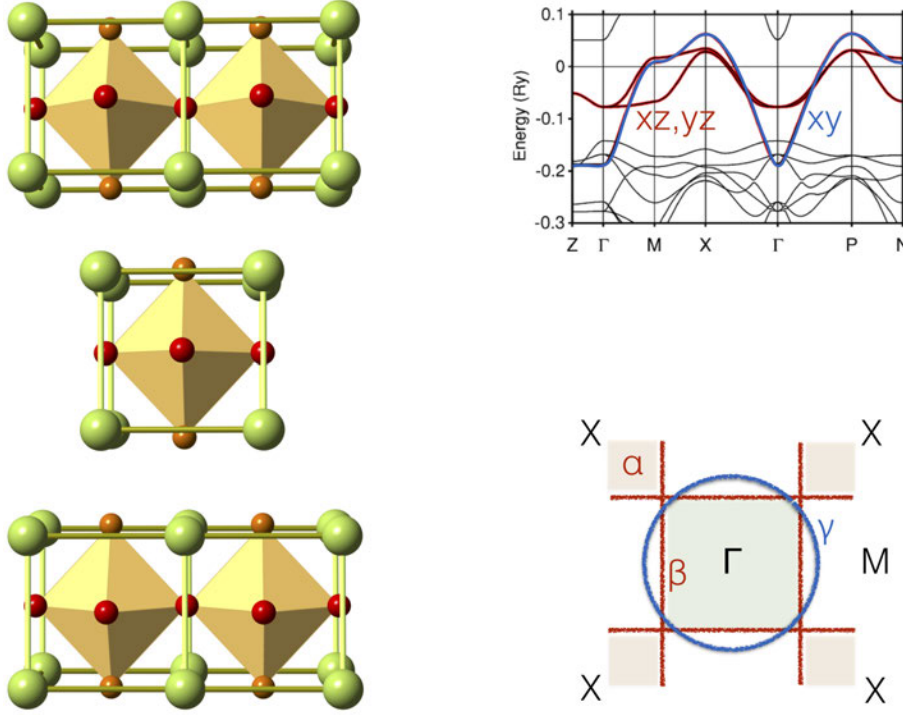


Fig. 12: Left: Crystal structure of the tetragonal layered perovskite Sr_2RuO_4 . Right: Low-energy LDA band structure (top) and schematic representation of the associated Fermi surface (bottom). The band structure was calculated using the Nth-Order Muffin-Tin Orbital (NMTO) method. The figure is rearranged from Ref. [18].

$\hat{H}_U$ are the so-called density-density terms, and the last two are the pair-hopping and spin-flip interactions. We dropped the double-counting correction $\hat{H}_{\text{DC}}$, which in this case is a mere shift of the chemical potential. The energies $\varepsilon_{m,m'}$ are the elements of the crystal-field matrix. In the case of cubic symmetry, the crystal-field matrix, the self-energy, the Green function and the spectral function are all diagonal in orbital space. For low-symmetry systems, however, this is not true. It can be seen in Fig. 13, which shows the diagonal and off-diagonal elements of the spectral-function matrix for the orthorhombic Mott insulator YTiO_3 .

3.4 Spin-orbit interaction and effects of the basis choice

In many interesting systems the spin-orbit interaction $\hat{H}_{\text{SO}}$ plays an important role. In the atomic limit, for the d shells the spin-orbit interaction is

$$\hat{H}_{\text{SO}} = \sum_{\mu} \lambda_{\mu} \sum_{mm'} \sum_{\sigma\sigma'} \epsilon_{m\sigma,m'\sigma'}^{\mu} c_{m\sigma}^{\dagger} c_{m'\sigma'}, \quad \epsilon_{m\sigma,m'\sigma'}^{\mu} = \langle m\sigma | l_{\mu} s_{\mu} | m'\sigma' \rangle, \quad (113)$$

where $\mu = x, y, z$, and λ_{μ} are the spin-orbit couplings, with $\lambda_{\mu} = \lambda$ in $O(3)$ symmetry, and

$$\lambda \sim g\mu_B^2 \left\langle \frac{1}{r} \frac{d}{dr} v_R(r) \right\rangle. \quad (114)$$

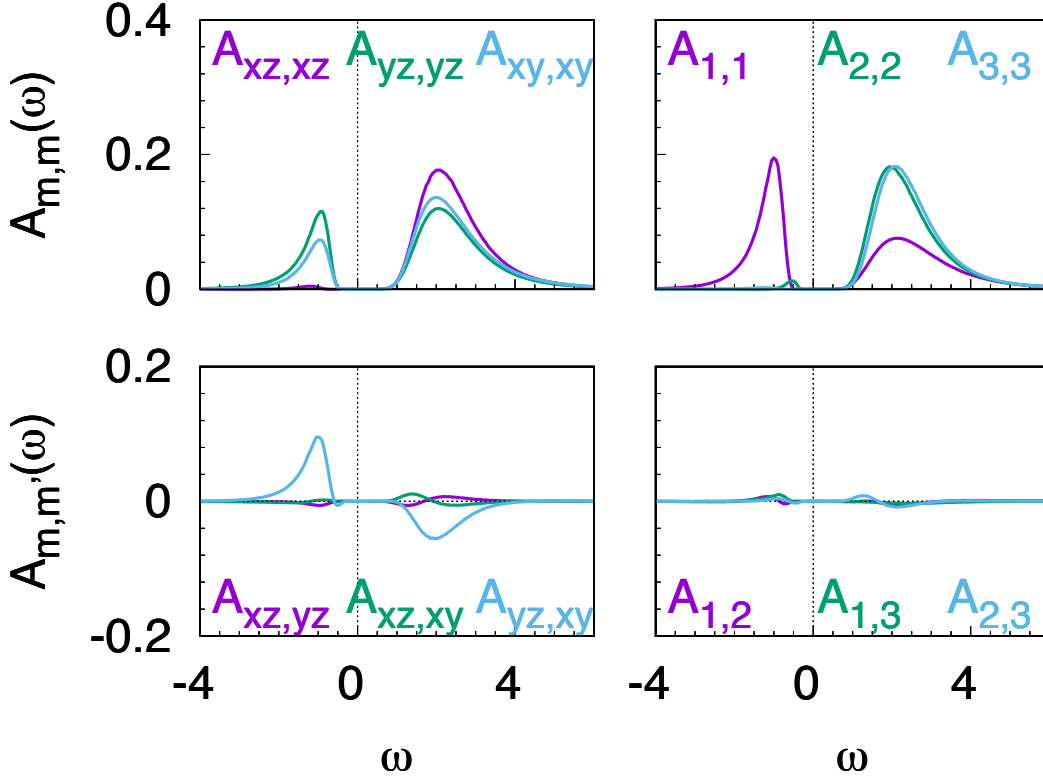


Fig. 13: The LDA+DMFT spectral function matrix of the orbitally-ordered t_{2g}^1 system YTiO_3 , in the (xz, yz, xy) basis (left panels) and in the crystal-field basis (right panels) [16, 19].

For a given material, the on-site one-electron Hamiltonian is thus the matrix $\hat{\varepsilon}$ with elements $\hat{\varepsilon}_{m\sigma, m'\sigma'} = \varepsilon_{m, m'} + \sum_{\mu} \lambda_{\mu} \epsilon_{m\sigma, m'\sigma'}^{\mu}$, where the terms $\varepsilon_{m, m'}$ arise from the crystal field. More specifically, for a t_{2g} system with tetragonal site symmetry we have

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_{xy} & 0 & 0 & 0 & \frac{\lambda_y}{2} & -\frac{i\lambda_x}{2} \\ 0 & \varepsilon_{yz} & \frac{i\lambda_z}{2} & -\frac{\lambda_y}{2} & 0 & 0 \\ 0 & -\frac{i\lambda_z}{2} & \varepsilon_{xz} & \frac{i\lambda_x}{2} & 0 & 0 \\ 0 & -\frac{\lambda_y}{2} & -\frac{i\lambda_x}{2} & \varepsilon_{xy} & 0 & 0 \\ \frac{\lambda_y}{2} & 0 & 0 & 0 & \varepsilon_{yz} & -\frac{i\lambda_z}{2} \\ \frac{i\lambda_x}{2} & 0 & 0 & 0 & \frac{i\lambda_z}{2} & \varepsilon_{xz} \end{pmatrix}. \quad (115)$$

Although $\hat{H}_{\text{SO}}$ looks like an innocent one-body term, it turns out that, for materials, simulations including this term are more difficult. This has two reasons: (i) QMC calculations involve Green function matrices of larger size, e.g., 6×6 as in the case just discussed, hence they are from the start computationally more demanding; (ii) QMC calculations are often hampered by a much stronger sign problem; even when it can be tamed, this often happens at the price of further increasing the computational time. Thus, specific basis choices and approximations are used. A possible approach consists in working in the basis that diagonalizes the non-interacting local Green function or the non-interacting local Hamiltonian; such a choice is known to reduce

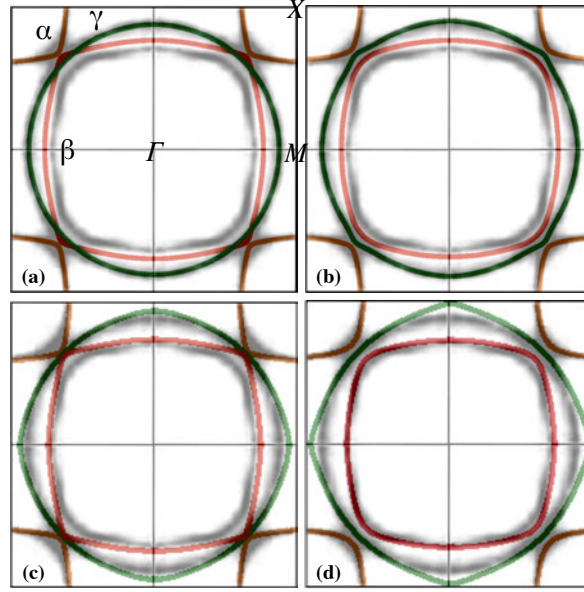


Fig. 14: Fermi surface of Sr_2RuO_4 calculated with LDA (a), LDA+SO (b), LDA+DMFT (c) and LDA+SO+DMFT (d). Figure from [20]. The grey maps are experimental results from Ref. [23].

the sign problem, as was first shown in Ref. [13] for the case without spin-orbit interaction. For a system with tetragonal symmetry, the states that diagonalize the local Green function belong either to the Γ_6 or to the Γ_7 irreducible representations, both 2-dimensional. There are two (coupled) Γ_7 representations, defining the space $\Gamma'_7 \oplus \Gamma''_7$. The analytic expression of these states can be found in Refs. [18, 20, 21]. The transformation to the $\Gamma_6 \oplus \Gamma'_7 \oplus \Gamma''_7$ basis is, of course, in principle, a mere basis change. Approximations are made, however, if all off-diagonal elements of the Green function are set to zero or the Coulomb tensor is truncated to further tame the sign problem or to reduce the computational time.

It has to be noticed that the Coulomb interaction of the t_{2g} Hubbard Hamiltonian is *only invariant under orthogonal transformations* of the basis. Thus if we change basis to the Γ_6 - Γ'_7 - Γ''_7 representation, the form of the interaction tensor changes. The exact expression of the **Coulomb tensor in the total angular momentum basis** (i.e., the one relevant when the t_{2g} crystal-field splitting is zero) can be found in Ref. [22]. For the Coulomb interaction tensor in the basis of **spherical harmonics** you can instead look the derivation in Ref. [8].

In the presence of crystal-field splitting, if the spin-orbit interaction is small in comparison, it is often preferable to perform the calculations in the t_{2g} basis. To this end, it is key to make QMC codes very efficient in order to reduce as much as possible statistical errors and increase the average sign. Exact LDA+SO+DMFT calculations in the t_{2g} basis have been successfully performed, e.g., for Sr_2RuO_4 , using an interaction-expansion continuous-time quantum Monte Carlo solver, and an orbital-dependent phase which makes the Green function matrix real [20, 21]. This approach allowed us to study, for example, the effects of the spin-orbit interaction on the Fermi surface without approximations. The results are shown in Fig. 14 in comparison with experimental data; we will discuss them in the next section.

3.5 Non-spherical Coulomb terms and double-counting correction

While for a free atom the Coulomb tensor is spherical (symmetry $O(3)$), in a material the screened Coulomb tensor has, in general, the symmetry of the lattice. Taking into account non-spherical Coulomb terms is, in general hard, both because they make QMC calculations more difficult and can worsen the sign problem, and because in their presence the double-counting correction has to be explicitly accounted for, even when massive downfolding is used. For these reasons they are typically neglected. Recently it was shown that they can play, however, a very important role for the Fermi surface [20]. Let us therefore discuss how the double-counting correction can be treated with and without such terms, following the approach of Ref. [20]. One of the classical approximations for the double-counting correction is the so called “around mean-field” approximation. The idea is that LDA describes well the average Coulomb term, in the absence of orbital polarization. This is equivalent to using as double-counting correction the Hartree term of the Coulomb interaction tensor, i.e., the operator

$$\begin{aligned} \hat{H}_U^{\text{DC}} = & U \sum_m (\hat{n}_{m\uparrow} \bar{n}_{m\downarrow} + \bar{n}_{m\uparrow} \hat{n}_{m\downarrow}) + (U-2J) \sum_{m \neq m'} (\hat{n}_{m\uparrow} \bar{n}_{m'\downarrow} + \bar{n}_{m\uparrow} \hat{n}_{m'\downarrow}) \\ & + (U-3J) \sum_{\sigma} \sum_{m > m'} (\hat{n}_{m\sigma} \bar{n}_{m'\sigma} + \bar{n}_{m\sigma} \hat{n}_{m'\sigma}) - \mu \hat{N}_d \\ & - U \sum_m \bar{n}_{m\uparrow} \bar{n}_{m\downarrow} + (U-2J) \sum_{m \neq m'} \bar{n}_{m\uparrow} \bar{n}_{m'\downarrow} + (U-3J) \sum_{\sigma} \sum_{m > m'} \bar{n}_{m\sigma} \bar{n}_{m'\sigma} \end{aligned} \quad (116)$$

where $\bar{n}_{m\sigma} = n/d$, if n is the number of the correlated electrons per site and d the orbital degeneracy. Within this approximation we have, after collecting all terms,

$$\begin{aligned} \hat{H}_U^{\text{DC}} = & (\delta\mu - \mu) \hat{N}_d - \frac{n^2}{d} (U(2d-1) - 5(d-1)) \\ & \delta\mu = \frac{n}{d} (U(2d-1) - 5J(d-1)). \end{aligned}$$

If we perform massive downfolding to the correlated bands, as previously mentioned, this is merely a shift of the chemical potential and can therefore be neglected. Let us now consider the case in which the Coulomb interaction has an additional term that does not change the average U but has tetragonal symmetry

$$\Delta \hat{H}_U = \frac{\Delta U}{3} (2\hat{n}_{xy\uparrow} \hat{n}_{xy\downarrow} - \hat{n}_{xz\uparrow} \hat{n}_{xz\downarrow} - \hat{n}_{yz\uparrow} \hat{n}_{yz\downarrow})$$

We can now use the around mean-field approximation for this term as well. We find

$$\begin{aligned} \Delta \hat{H}_U^{\text{DC}} = & \frac{n}{6} \frac{\Delta U}{3} \sum_{\sigma} (2\hat{n}_{xy\sigma} - \hat{n}_{xz\sigma} - \hat{n}_{yz\sigma}) = \frac{n}{6} \Delta U \sum_{\sigma} \hat{n}_{xy\sigma} - \delta\mu' \hat{N} \\ \delta\mu' = & \frac{n}{6} \frac{\Delta U}{3}. \end{aligned}$$

This term, in addition to a shift of the chemical potential, yields an effective change of the crystal-field splitting ε_{CF} , and has therefore to be accounted for explicitly.

How does ΔU change the Fermi surface of Sr_2RuO_4 ? The Fermi surface is determined by the poles of the Green function at zero frequency. These depend on the non-interacting Hamiltonian and the self-energy matrix at zero frequency. In the Fermi-liquid regime, and within the DMFT approximation, the effect of the self-energy is merely to modify the on-site part of the Hamiltonian, i.e., the crystal-field splitting $\varepsilon_{\text{CF}} = \varepsilon_{xz/yz} - \varepsilon_{xy}$ and the spin-orbit couplings λ_μ

$$\varepsilon_{\text{CF}} \rightarrow \varepsilon_{\text{CF}} + \Delta\varepsilon_{\text{CF}}(0),$$

$$\lambda_\mu \rightarrow \lambda_\mu + \Delta\lambda_\mu(0).$$

Both $\Delta\varepsilon_{\text{CF}}(0)$ and $\Delta\lambda_\mu(0)$ are positive for Sr_2RuO_4 , and lead to an almost doubling of the LDA parameters. The positive $\Delta\varepsilon_{\text{CF}}(0)$ shrinks the β sheet (xz/yz bands) and enlarges the γ (xy band) sheet. This can be understood from the schematic Fermi surface and the LDA band structure in Fig. 12. Enhancing the crystal-field splitting corresponds to slightly moving the xy band downwards and the xz/yz bands upwards with respect to the Fermi level. The enhancement of the spin-orbit couplings has a large Hartree-Fock component [21], since the spin-orbit interaction yields a small but finite off-diagonal occupation matrix. For an $O(3)$ -symmetric Coulomb tensor, the Hartree-Fock enhancement of the spin-orbit coupling is thus

$$\begin{aligned} \frac{\Delta\lambda_z}{2} &= i(U-3J) n_{xz,yz}^{\uparrow\uparrow} \\ \frac{\Delta\lambda_y}{2} &= -(U-3J) n_{xy,yz}^{\uparrow\downarrow}, \\ \frac{\Delta\lambda_x}{2} &= -i(U-3J) n_{xz,xy}^{\uparrow\downarrow}, \end{aligned}$$

where $n_{m,m'}^{\sigma\sigma'}$ are the off-diagonal elements of the density matrix. The Coulomb-enhanced spin-orbit coupling improves the agreement with the experimental Fermi surface at the degeneracy points (e.g., along the Γ - X direction). The agreement with ARPES data, however, further deteriorates for the γ sheet. This can be seen in Fig. 14, in which the LDA and LDA+DMFT Fermi surface are shown on top of ARPES data from Ref. [23].

Including correlation effects has thus two opposite effects: on the one hand, the agreement with experiments improves (with respect to LDA) for the β sheet; on the other hand, it deteriorates for the γ sheet. This can be seen comparing either panels (a) and (c) or panels (b) and (d) in Fig. 14. Introducing tetragonal Coulomb terms, and in particular the term ΔU , however, reduces the crystal-field enhancement to

$$\varepsilon_{\text{CF}} \rightarrow \varepsilon_{\text{CF}} + \Delta\varepsilon'_{\text{CF}}(0),$$

where $\Delta\varepsilon'_{\text{CF}}(0)$ becomes almost zero for cRPA-based estimates of ΔU . This leads to an almost perfect Fermi surface, as shown in Fig. 15. Non-spherical Coulomb terms turn out to be more important for properties that reflect the point symmetry of the lattice, like the Fermi surface, than for properties that average over orbitals, like the total spectral function [21].

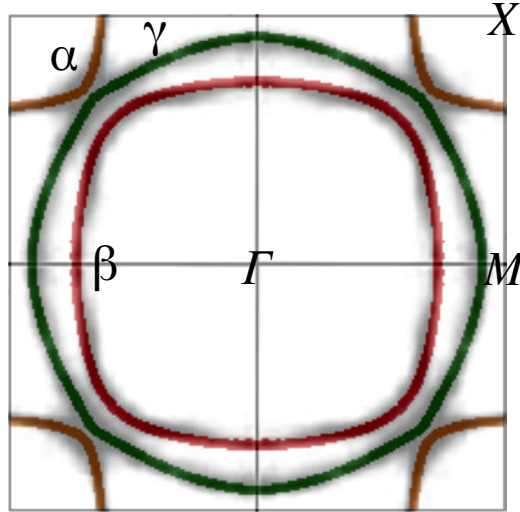


Fig. 15: The LDA+SO+DMFT Fermi surface of Sr_2RuO_4 calculated including the effects of the non-spherical Coulomb term ΔU . Figure from [20].

4 Understanding correlated materials with DMFT

The ultimate test for a method is the actual advance it brings in solving outstanding problems. The LDA+DMFT approach has proven very successful, advancing qualitatively our understanding of the physics of strongly-correlated materials. This has two reasons. From the [many-body side](#), the success of the method is rooted in the fact that dynamical mean-field theory captures exactly key limits: non-interacting limit, atomic limit, single-impurity limit, infinite dimension limit. It also describes very well phases with different characteristics: paramagnetism, long-range order, metallic and insulating phase. Thus DMFT can be used to explore transitions/crossovers from one corner to another of the phase diagram, building a coherent picture as a function of both internal and external parameters (e.g., U , J , doping and temperature). This differs from static mean-field approaches, such as the LDA+ U method; the latter only captures the behavior in a specific corner of the phase diagram (and typically merely the ground state), e.g., long-range magnetic order¹³ or, if one uses large supercells, spin-glass-like phases. This limitation makes static mean-field approaches unsuitable to study, e.g., the effects of temperature or external parameters, and for identifying a mechanism behind a certain phenomenon, or finding the onset of a Mott phase. From the [materials-dependence side](#), by constructing many-body models from Kohn-Sham Wannier functions, the LDA+DMFT technique builds on the shoulders of the most successful theory for materials, DFT. In order to illustrate the success of LDA+DMFT, here I will briefly review two paradigmatic examples.

¹³It should be noted that LDA+ U is less accurate than DMFT also in capturing long-range order. This is because LDA+ U is based on the Hartree-Fock approximation; as we have seen for the Hubbard dimer, it is, therefore, the static (high-frequency) limit of LDA+DMFT. All results from LDA+ U are thus by construction contained in LDA+DMFT, provided that, of course, calculations are performed for the same Hamiltonian, the same basis and the same unit cell. Unlike LDA+ U , however, LDA+DMFT also captures correctly dynamical fluctuations and actual phase transitions. See Ref. [14] for more details. Long-range order is well described in DMFT via a local cluster with the size of the (enlarged) broken-symmetry unit cell, without additional computational cost.

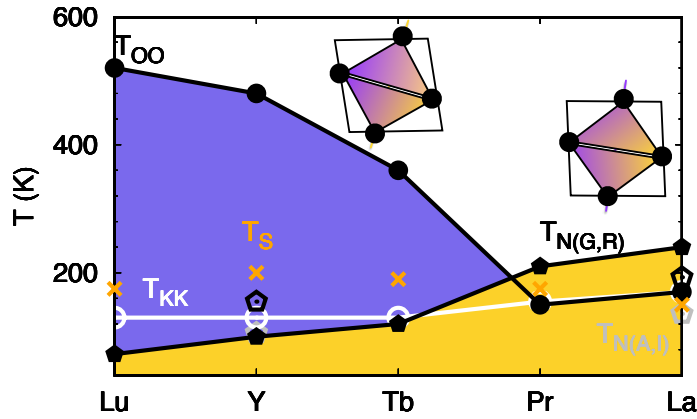


Fig. 16: Inversion of magnetic and orbital transition in the ReVO_3 series with increasing rare-earth (Re) ionic radius. While in most orbitally ordered systems the magnetic transition occurs at lower temperature than the orbital transition ($T_N < T_{OO}$), in the ReVO_3 series $T_N \gtrsim T_{OO}$ for large rare-earth (Re) radius. Figure from [35].

The first is unravelling the importance of **crystal-field splitting** in determining low-energy properties of strongly-correlated systems. The metal-insulator Mott transition arises, in the simplest picture, from the competition between the bandwidth, W , and the on-site Coulomb repulsion, U . This competition is well captured by the single-band Hubbard model; however, the latter is typically insufficient for understanding the complex behavior displayed by actual correlated materials. One reason is the importance of multi-orbital effects. Pioneering works based on lattice QMC calculations for the multi-band Hubbard model have indeed shown that orbital degeneracy strongly influence the size of the Mott gap [24, 25]. Accounting for multi-orbital effects strongly complicates the matter from the viewpoint of theory: realistic models have many unknown independent parameters and are considerably more difficult to solve. The LDA+DMFT approach provided the first framework to study multi-orbital phenomena systematically in a realistic setting. One of the early success of the method was to show that, contrarily to what was often assumed, small details do matter. A crystal-field splitting much smaller than the bandwidth can favor the Mott metal-insulator transition. In fact, everything else remaining the same, the crystal-field splitting reduces orbital degeneracy, thereby modifying the correlation strength. This insight explained the metal-insulator transition in the series of $3d^1$ perovskites: SrVO_3 (metal), CaVO_3 (metal), LaTiO_3 (small gap insulator), YTiO_3 (insulator) [19]. Important crystal-field effects are now recognized in most correlated materials. They have been key to settle several hot debates, e.g., to understand the complex phase diagram of layered $\text{Ca}_{2-x}\text{Sr}_x\text{RuO}_4$ ruthenates [26]. More recently, it was shown that the crystal-field splitting and its subtle interplay with the **spin-orbit interaction** is crucial for determining the correlation strength, the shape of the Fermi surface and the effective masses [20–22, 27].

The second success story is unravelling the **origin of orbital ordering** in materials. More than fifty years ago, thanks to seminal works of Kugel and Khomskii [28], it became clear that orbital degrees of freedom play a key role in the properties of strongly-correlated Mott insulators; in particular, they can give rise to orbital ordering via many-body super-exchange interactions [29,

30]. For a long time it was however hotly debated if, in actual materials, it is Kugel-Khomskii (KK) super-exchange or rather the Jahn-Teller effect that control the onset of orbital ordering. This is a classical case of a chicken-and-egg problem, with two theoretical mechanisms leading to the same final state. The systems for which super-exchange-driven orbital ordering was initially proposed were the e_g perovskites KCuF_3 and LaMnO_3 . Combining DMFT with a novel disentanglement scheme [31], it became possible to solve the problem. It was shown that, in KCuF_3 and LaMnO_3 , the KK super-exchange interaction, although strong, is not sufficiently large to drive alone orbital ordering at experimental temperatures [31, 32]. More recently, it was shown that t_{2g} systems are more likely than e_g materials to exhibit super-exchange-driven orbital ordering; LaVO_3 was then identified as the first clear-cut case of a Kugel-Khomskii material [33, 34]. This led to the solution of another long-standing problem: the origin of the puzzling inversion of magnetic and orbital transition in the ReVO_3 series with increasing rare-earth (Re) ionic radius, see Fig. 16 and Refs. [34, 35]. Furthermore, as a byproduct, it was shown that DMFT and super-exchange theory can be successfully combined to pinpoint the nature of spin-orbital phases. This can be done via a general scheme for constructing **exact materials-specific super-exchange Hamiltonians** in an irreducible tensor basis [33–35]; the latter provides at the same time analytic expressions and numerical results. This scheme can be used for the realistic description of all KK-type super-exchange phenomena, static (e.g., orbital ordering) and dynamical (e.g., spin-orbital excitations, orbital and spin liquid behavior).

5 Conclusion and outlook

The LDA+DMFT approach and its extension has proven very successful for describing correlated materials. In the last decades, it also became very versatile. It is now possible, e.g., to study multi-orbital models including the full Coulomb vertex and the spin-orbit interaction, and to calculate response functions or bosonic excitations. Still, several challenges remain. Models with more than a few orbitals or sites and out-of-equilibrium phenomena are still, e.g., hard to study. Furthermore, since the birth of DFT, Moore’s law constantly helped as much as new algorithms in extending the frontier; that has come now slowly to a halt.

Quantum computers and artificial intelligence, or new, not yet foreseeable, technological advances might help us in the future. The present deceleration, however, urges us to think to what we should strive for. It is perhaps easier to discuss first what we *do not* need. We do not need a magical calculating machine that gives us answers with no explanations. Reproducing the data is, of course, the aim of theory, but not the only goal. As scientists we need to know why. The danger of giving this question up is evident if we look back at the historical controversy between copernican and ptolemaic system. At the time, the ptolemaic system was the one which agreed better with experiments. Indeed, it had been fine-tuned along the years, via a series of ad hoc assumptions. The copernican system was initially not performing so well, among others because it was based on circular orbits. But, at the end, it is the copernican system that made a contribution to science, by identifying the correct picture.

Appendices

A Hubbard dimer and Anderson molecule

The Hamiltonian of the **Hubbard dimer** is given by

$$\hat{H} = \varepsilon_d \sum_{\sigma} \sum_{i=1,2} n_{i\sigma} - t \sum_{\sigma} (c_{1\sigma}^{\dagger} c_{2\sigma} + c_{2\sigma}^{\dagger} c_{1\sigma}) + U \sum_{i=1,2} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}.$$

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, \frac{1}{2}, \sigma\rangle_1 =$	$c_{1\sigma}^{\dagger} 0\rangle$	1	$\frac{1}{2}$	ε_d
$ 1, \frac{1}{2}, \sigma\rangle_2 =$	$c_{2\sigma}^{\dagger} 0\rangle$	1	$\frac{1}{2}$	ε_d
$ 2, 1, 1\rangle =$	$c_{2\uparrow}^{\dagger} c_{1\uparrow}^{\dagger} 0\rangle$	2	1	$2\varepsilon_d$
$ 2, 1, -1\rangle =$	$c_{2\downarrow}^{\dagger} c_{1\downarrow}^{\dagger} 0\rangle$	2	1	$2\varepsilon_d$
$ 2, 1, 0\rangle =$	$\frac{1}{\sqrt{2}} (c_{1\uparrow}^{\dagger} c_{2\downarrow}^{\dagger} + c_{1\downarrow}^{\dagger} c_{2\uparrow}^{\dagger}) 0\rangle$	2	1	$2\varepsilon_d$
$ 2, 0, 0\rangle_0 =$	$\frac{1}{\sqrt{2}} (c_{1\uparrow}^{\dagger} c_{2\downarrow}^{\dagger} - c_{1\downarrow}^{\dagger} c_{2\uparrow}^{\dagger}) 0\rangle$	2	0	$2\varepsilon_d$
$ 2, 0, 0\rangle_1 =$	$c_{1\uparrow}^{\dagger} c_{1\downarrow}^{\dagger} 0\rangle$	2	0	$2\varepsilon_d + U$
$ 2, 0, 0\rangle_2 =$	$c_{2\uparrow}^{\dagger} c_{2\downarrow}^{\dagger} 0\rangle$	2	0	$2\varepsilon_d + U$
$ 3, \frac{1}{2}, \sigma\rangle_1 =$	$c_{1\sigma}^{\dagger} c_{2\uparrow}^{\dagger} c_{2\downarrow}^{\dagger} 0\rangle$	3	$\frac{1}{2}$	$3\varepsilon_d + U$
$ 3, \frac{1}{2}, \sigma\rangle_2 =$	$c_{2\sigma}^{\dagger} c_{1\uparrow}^{\dagger} c_{1\downarrow}^{\dagger} 0\rangle$	3	$\frac{1}{2}$	$3\varepsilon_d + U$
$ 4, 0, 0\rangle =$	$c_{1\uparrow}^{\dagger} c_{1\downarrow}^{\dagger} c_{2\uparrow}^{\dagger} c_{2\downarrow}^{\dagger} 0\rangle$	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

$ 1, S, S_z\rangle_\alpha$	$E_\alpha(1, S)$	$d_\alpha(1, S)$
$ 1, \frac{1}{2}, \sigma\rangle_+ = \frac{1}{\sqrt{2}}(1, \frac{1}{2}, \sigma\rangle_1 - 1, \frac{1}{2}, \sigma\rangle_2)$	$\varepsilon_d + t$	2
$ 1, \frac{1}{2}, \sigma\rangle_- = \frac{1}{\sqrt{2}}(1, \frac{1}{2}, \sigma\rangle_1 + 1, \frac{1}{2}, \sigma\rangle_2)$	$\varepsilon_d - t$	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

$ 2, S, S_z\rangle_\alpha$	$E_\alpha(2, S)$	$d_\alpha(2, S)$
$ 2, 0, 0\rangle_+ = a_1 2, 0, 0\rangle_0 - \frac{a_2}{\sqrt{2}}(2, 0, 0\rangle_1 + 2, 0, 0\rangle_2)$	$2\varepsilon_d + \frac{1}{2}(U + \Delta(t, U))$	1
$ 2, 0, 0\rangle_o = \frac{1}{\sqrt{2}}(2, 0, 0\rangle_1 - 2, 0, 0\rangle_2)$	$2\varepsilon_d + U$	1
$ 2, 1, m\rangle_o = 2, 1, m\rangle$	$2\varepsilon_d$	3
$ 2, 0, 0\rangle_- = a_2 2, 0, 0\rangle_0 + \frac{a_1}{\sqrt{2}}(2, 0, 0\rangle_1 + 2, 0, 0\rangle_2)$	$2\varepsilon_d + \frac{1}{2}(U - \Delta(t, U))$	1

where

$$\Delta(t, U) = \sqrt{U^2 + 16t^2},$$

and

$$a_1^2 = a_1^2(t, U) = \frac{1}{\Delta(t, U)} \frac{\Delta(t, U) - U}{2} \quad a_2^2 = a_2^2(t, U) = \frac{4t^2}{\Delta(t, U)} \frac{2}{\Delta(t, U) - U},$$

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, \frac{1}{2}, \sigma\rangle_+ = \frac{1}{\sqrt{2}}(1, \frac{1}{2}, \sigma\rangle_1 + 1, \frac{1}{2}, \sigma\rangle_2)$	$3\varepsilon_d + U + t$	2
$ 3, \frac{1}{2}, \sigma\rangle_- = \frac{1}{\sqrt{2}}(1, \frac{1}{2}, \sigma\rangle_1 - 1, \frac{1}{2}, \sigma\rangle_2)$	$3\varepsilon_d + U - t$	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

$$\hat{H} = \varepsilon_s \sum_{\sigma} \hat{n}_{2\sigma} - t \sum_{\sigma} (c_{1\sigma}^\dagger c_{2\sigma} + c_{2\sigma}^\dagger c_{1\sigma}) + \varepsilon_d \sum_{\sigma} \hat{n}_{1\sigma} + U \hat{n}_{1\uparrow} \hat{n}_{1\downarrow}.$$

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, \frac{1}{2}, \sigma\rangle_1 =$	$c_{1\sigma}^\dagger 0\rangle$	1	$\frac{1}{2}$	ε_d
$ 1, \frac{1}{2}, \sigma\rangle_2 =$	$c_{2\sigma}^\dagger 0\rangle$	1	$\frac{1}{2}$	ε_s
$ 2, 1, 1\rangle =$	$c_{2\uparrow}^\dagger c_{1\uparrow}^\dagger 0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$ 2, 1, -1\rangle =$	$c_{2\downarrow}^\dagger c_{1\downarrow}^\dagger 0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$ 2, 1, 0\rangle =$	$\frac{1}{\sqrt{2}}(c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger) 0\rangle$	2	1	$\varepsilon_d + \varepsilon_s$
$ 2, 0, 0\rangle_0 =$	$\frac{1}{\sqrt{2}}(c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger - c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger) 0\rangle$	2	0	$\varepsilon_d + \varepsilon_s$
$ 2, 0, 0\rangle_1 =$	$c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger 0\rangle$	2	0	$2\varepsilon_d + U$
$ 2, 0, 0\rangle_2 =$	$c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger 0\rangle$	2	0	$2\varepsilon_s$
$ 3, \frac{1}{2}, \sigma\rangle_1 =$	$c_{1\sigma}^\dagger c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger 0\rangle$	3	$\frac{1}{2}$	$\varepsilon_d + 2\varepsilon_s$
$ 3, \frac{1}{2}, \sigma\rangle_2 =$	$c_{2\sigma}^\dagger c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger 0\rangle$	3	$\frac{1}{2}$	$2\varepsilon_d + \varepsilon_s + U$
$ 4, 0, 0\rangle =$	$c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger 0\rangle$	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

$ 1, S, S_z\rangle_\alpha$	$E_\alpha(1, S)$	$d_\alpha(1, S)$
$ 1, \frac{1}{2}, \sigma\rangle_+ = \alpha_1 1, \frac{1}{2}, \sigma\rangle_1 - \alpha_2 1, \frac{1}{2}, \sigma\rangle_2$	$\frac{1}{2}(\varepsilon_d + \varepsilon_s + \sqrt{(\varepsilon_d - \varepsilon_s)^2 + 4t^2})$	2
$ 1, \frac{1}{2}, \sigma\rangle_- = \alpha_2 1, \frac{1}{2}, \sigma\rangle_1 + \alpha_1 1, \frac{1}{2}, \sigma\rangle_2$	$\frac{1}{2}(\varepsilon_d + \varepsilon_s - \sqrt{(\varepsilon_d - \varepsilon_s)^2 + 4t^2})$	2

where $d_\alpha(N)$ is the spin degeneracy of the α manifold. For $\varepsilon_s = \varepsilon_d + U/2$ the eigenvalues are

$$E_\pm(1, S) = \varepsilon_d + \frac{1}{4}(U \pm \Delta(U, t)),$$

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

$ 2, S, S_z\rangle_\alpha$	$E_\alpha(2, S)$	$d_\alpha(2, S)$
$ 2, 0, 0\rangle_+ = b_1 2, 0, 0\rangle_0 - \frac{b_2}{\sqrt{2}}(2, 0, 0\rangle_1 + 2, 0, 0\rangle_2)$	$2\varepsilon_d + \frac{U}{2} + \frac{1}{4}(U + 2\Delta(t, \frac{U}{2}))$	1
$ 2, 0, 0\rangle_o = \frac{1}{\sqrt{2}}(2, 0, 0\rangle_1 - 2, 0, 0\rangle_2)$	$2\varepsilon_d + U$	1
$ 2, 1, m\rangle_o = 2, 1, m\rangle$	$2\varepsilon_d + \frac{U}{2}$	3
$ 2, 0, 0\rangle_- = b_2 2, 0, 0\rangle_0 + \frac{b_1}{\sqrt{2}}(2, 0, 0\rangle_1 + 2, 0, 0\rangle_2)$	$2\varepsilon_d + \frac{U}{2} + \frac{1}{4}(U - 2\Delta(t, \frac{U}{2}))$	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 $|2, 0, 0\rangle_-$ differ, however. Finally, for $N = 3$ electrons, the eigenstates are

$ 3, S, S_z\rangle_\alpha$	$E_\alpha(3, S)$	$d_\alpha(3, S)$
$ 3, \frac{1}{2}, \sigma\rangle_+ = \alpha_2 1, \frac{1}{2}, \sigma\rangle_1 + \alpha_1 1, \frac{1}{2}, \sigma\rangle_2$	$3\varepsilon_d + U + \frac{1}{4}(U + \Delta(t, U))$	2
$ 3, \frac{1}{2}, \sigma\rangle_- = \alpha_1 1, \frac{1}{2}, \sigma\rangle_1 - \alpha_2 1, \frac{1}{2}, \sigma\rangle_2$	$3\varepsilon_d + U + \frac{1}{4}(U - \Delta(t, U))$	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

$$G_{i,i}^{\sigma}(i\nu_n) = - \int_0^{\beta} d\tau e^{i\nu_n\tau} \langle \mathcal{T} c_{i\sigma}(\tau) c_{i\sigma}^{\dagger}(0) \rangle, \quad (117)$$

where $\mathcal{T}$ is the time-ordering operator, $\beta = 1/k_B T$, and ν_n a fermionic Matsubara frequency. Let us assume to know all eigenstates $|N_l\rangle$ and their energy $E_l(N)$, for arbitrary number of electrons N . Thus, formally

$$G_{i,i}^{\sigma}(i\nu_n) = - \frac{1}{Z} \sum_{Nl} \int_0^{\beta} d\tau e^{i\nu_n\tau} e^{-\Delta E_l(N)\beta} \langle N_l | c_{i\sigma}(\tau) c_{i\sigma}^{\dagger}(0) | N_l \rangle, \quad (118)$$

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 $c_{i\sigma}^{\dagger}(0) = c_{i\sigma}^{\dagger}$. We now insert a complete set of states, obtaining

$$\begin{aligned} G_{i,i}^{\sigma}(i\nu_n) &= - \frac{1}{Z} \sum_{l'NN'} \int_0^{\beta} d\tau e^{i\nu_n\tau} e^{-\Delta E_l(N)\beta} \langle N_l | c_{i\sigma}(\tau) | N'_{l'} \rangle \langle N'_{l'} | c_{i\sigma}^{\dagger} | N_l \rangle \\ &= - \frac{1}{Z} \sum_{l'NN'} \int_0^{\beta} d\tau e^{-\Delta E_l(N)\beta} e^{(i\nu_n + \Delta E_l(N) - \Delta E_{l'}(N'))\tau} |\langle N'_{l'} | c_{i\sigma}^{\dagger} | N_l \rangle|^2 \\ &= \frac{1}{Z} \sum_{l'NN'} \frac{e^{-\Delta E_{l'}(N')\beta} + e^{-\Delta E_l(N)\beta}}{i\nu_n + \Delta E_l(N) - \Delta E_{l'}(N')} |\langle N'_{l'} | c_{i\sigma}^{\dagger} | N_l \rangle|^2. \end{aligned} \quad (119)$$

Due to the weight $|\langle N'_{l'} | c_{i\sigma}^{\dagger}(0) | N_l \rangle|^2$ 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

$$G_{i,i}^{\sigma}(i\nu_n) = \sum_{l'N} \frac{e^{-\beta\Delta E_l(N)}}{Z} \left(\frac{|\langle (N-1)_{l'} | c_{i\sigma} | N_l \rangle|^2}{i\nu_n - \Delta E_l(N) + \Delta E_{l'}(N-1)} + \frac{|\langle (N+1)_{l'} | c_{i\sigma}^{\dagger} | N_l \rangle|^2}{i\nu_n - \Delta E_{l'}(N+1) + \Delta E_l(N)} \right).$$

Let us consider as example the atomic limit of the Hubbard model at half filling (for which $\mu = \varepsilon_d + U/2$). In this case all sites are decoupled; there are four eigenstates per site, the vacuum $|0\rangle$, with $\Delta E(0) = 0$, the doublet $|1_{\sigma}\rangle = c_{i\sigma}^{\dagger}|0\rangle$, with $\Delta E_{\sigma}(1) = -U/2$, and the doubly-occupied singlet $|2\rangle = c_{i\uparrow}^{\dagger}c_{i\downarrow}^{\dagger}|0\rangle$, with $\Delta E(2) = 0$. Furthermore, $Z = 2(1 + e^{\beta U/2})$ and

$$|\langle (N-1)_{l'} | c_{i\sigma} | N_l \rangle|^2 = \begin{cases} 1 & \text{if } |N_l\rangle = |2\rangle \vee |1_{\sigma}\rangle \\ 0 & \text{otherwise} \end{cases} \quad |\langle (N+1)_{l'} | c_{i\sigma}^{\dagger} | N_l \rangle|^2 = \begin{cases} 1 & \text{if } |N_l\rangle = |0\rangle \vee |1_{-\sigma}\rangle \\ 0 & \text{otherwise} \end{cases}$$

Thus, after summing up the four non-zero contributions, we find

$$G_{i,i}^{\sigma}(\nu_n) = \frac{1}{2} \left(\frac{1}{i\nu_n + U/2} + \frac{1}{i\nu_n - U/2} \right). \quad (120)$$

References

- [1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964);
W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)
- [2] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999)
- [3] P.W. Anderson: *More and different – Notes from a thoughtful curmudgeon*
(World Scientific, Singapore, 2011)
- [4] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [5] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989);
Z. Phys. B **76**, 211 (1989); Int. J. Mod. Phys. B **3**, 2169 (1989)
- [6] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [7] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992)
- [8] 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>
- [9] 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>
- [10] 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>
- [11] E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008)
- [12] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner,
Rev. Mod. Phys. **83**, 349 (2011)
- [13] A. Flesch, E. Gorelov, E. Koch, E. Pavarini, Phys. Rev. B **87**, 195141 (2013)
- [14] E. Pavarini, E. Koch, R. Scalettar, and R.M. Martin (eds.):
The physics of correlated insulators, metals, and superconductors
Reihe Modeling and Simulation, Vol. 7 (Forschungszentrum Jülich, 2017)
<http://www.cond-mat.de/events/correl17>
- [15] A. Kiani and E. Pavarini, Phys. Rev. B **94**, 075112 (2016)

- [16] E. Pavarini, A. Yamasaki, J. Nuss, and O.K. Andersen, New. J. Phys **7**, 188 (2005)
- [17] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012)
- [18] G. Zhang and E. Pavarini, Phys. Status Solidi RRL **12**, 1800211 (2018)
- [19] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and O.K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004)
- [20] G. Zhang, E. Gorelov, E. Sarvestani, and E. Pavarini, Phys. Rev. Lett. **116**, 106402 (2016)
- [21] E. Sarvestani, G. Zhang, E. Gorelov, and E. Pavarini, Phys. Rev. B **97**, 085141 (2018)
- [22] G. Zhang and E. Pavarini, Phys. Rev. B **99**, 125102 (2019)
- [23] A. Damascelli, D.H. Lu, K.M. Shen, N.P. Armitage, F. Ronning, D.L. Feng, C. Kim, Z.-X. Shen, T. Kimura, Y. Tokura, Z.Q. Mao, and Y. Maeno, Phys. Rev. Lett. **85**, 5194 (2000)
- [24] O. Gunnarsson, E. Koch, and R.M. Martin, Phys. Rev. B **56**, 1146 (1996)
- [25] E. Koch, O. Gunnarsson, and R.M. Martin, Phys. Rev. B **60**, 15714 (1999)
- [26] E. Gorelov, M. Karolak, T.O. Wehling, F. Lechermann, A.I. Lichtenstein, E. Pavarini, Phys. Rev. Lett. **104**, 226401 (2010)
- [27] N. Samani, G. Zhang and E. Pavarini, Phys. Rev. Lett. **132**, 236505 (2024)
- [28] K.I. Kugel and D.I. Khomskii, Zh. Eksp. Teor. Fiz. **64**, 1429 (1973)
[Sov. Phys. JETP **37**, 725 (1973)]
- [29] D.I. Khomskii, *Transition metal compounds*, Cambridge University Press (2014)
- [30] E. Pavarini and E. Koch (eds.): *Orbital physics in correlated matter*, Reihe Modeling and Simulation, Vol. 13 (Forschungszentrum Jülich, 2023)
<http://www.cond-mat.de/events/correl23>
- [31] E. Pavarini, E. Koch, and A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008)
- [32] E. Pavarini and E. Koch, Phys. Rev. Lett. **104**, 086402 (2010)
- [33] X-J. Zhang, E. Koch and E. Pavarini, Phys. Rev. B **105**, 115104 (2022)
- [34] X-J. Zhang, E. Koch and E. Pavarini, Phys. Rev. B **106**, 115110 (2022)
- [35] X-J. Zhang, E. Koch and E. Pavarini, Phys. Rev. Lett. **135**, 026508 (2025)

2 **Constrained Density-Functional Theory: Formalism, Algorithms and Applications**

Xavier Gonze

IMCN, Université catholique de Louvain

8, chemin de l'observatoire, B-1348 Louvain-la-neuve

Contents

1	Introduction	2
2	Background: density-functional theory	4
2.1	Hohenberg-Kohn and Kohn-Sham theorems	4
2.2	Spin-dependent generalization of the basic theorems	7
2.3	Approximate exchange-correlation functionals	10
2.4	Algorithms for the self consistency	11
3	Constrained density-functional theory: formalism and approximations	13
3.1	Formal constrained density-functional theory	14
3.2	Constraints at the Hohenberg-Kohn level	15
3.3	Constrained density-functional theory at the Kohn-Sham level	17
4	Constrained density-functional theory algorithms	18
4.1	Microiterations	18
4.2	Penalty function approaches	19
4.3	Lagrange multipliers again: a potential-based SCF approach	20
5	Some applications of constrained density-functional theory	22
6	Conclusion	28

1 Introduction

The vast majority of first-principles simulations of ground-state properties of molecules, condensed matter and nanosystems relies on density-functional theory (DFT) [1, 2]. Over the last decades, DFT has achieved remarkable success, providing a broadly applicable and computationally efficient framework to describe the structural, electronic, magnetic, and vibrational properties of a wide variety of systems. It has enabled predictive simulations of complex materials, accurate total energies, forces and response functions, and has become the workhorse of modern materials science and quantum chemistry [3].

Despite these achievements, DFT also exhibits well-known limitations [4, 5]. These limitations become particularly severe in systems where electron-electron interactions play a dominant role, such as transition-metal oxides, heavy fermion compounds, or strongly correlated molecular systems. In these cases, standard approximations to the exchange-correlation functional may fail to capture key physical phenomena, including Mott insulating behavior, charge localization, or dynamical spectral properties [5]. More generally, the static and mean-field nature of conventional DFT functionals makes them ill-suited for describing strong correlation effects and the associated many-body excitations.

To overcome these difficulties, more advanced theoretical frameworks have been developed. Among them, dynamical mean-field theory (DMFT) is a powerful approach to treat local electronic correlations beyond the mean-field level, by mapping the lattice problem onto an effective quantum impurity model embedded in a self-consistently determined bath. [6, 5]. The combination of DFT and DMFT (DFT+DMFT) has proven particularly successful, allowing one to retain the realistic band-structure description provided by DFT while capturing the dynamical correlations characteristic of strongly correlated materials. One extracts from DFT localized orbitals, determines hopping parameters, and estimates interaction strengths (the Hubbard- U) entering DMFT calculations [7]. For example, the parameters of the Hubbard model, here written for the single-band case, can all be determined

$$\hat{H} = - \sum_{\langle ij \rangle, \sigma} t_{ij} \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (1)$$

where i, j label lattice sites, $\langle i, j \rangle$ denotes pairs of sites coupled by hopping, $\sigma = \uparrow, \downarrow$ is the spin index, $\hat{c}_{i\sigma}^\dagger$ and $\hat{c}_{i\sigma}$ are fermionic creation and annihilation operators, t_{ij} is the hopping amplitude between sites i and j , $\hat{n}_{i\sigma}$ is the number operator at site i for spin σ , and U is the on-site Coulomb interaction.

In the general context of the simulation of real materials based on model many-body Hamiltonians, DFT plays an essential, albeit auxiliary role, as it provides the underlying one-electron Hamiltonian and effective interactions and enables the construction of low-energy effective models. It also allows one to identify relevant degrees of freedom and to construct corresponding many-body models.

Even if not explicitly mentioned, such usage of DFT beyond electronic ground state calculations is often based on constrained density-functional theory (cDFT), introduced in 1984 by

Dederichs and coworkers [8]. cDFT provides a versatile framework to access states and observables beyond the ground state, and to define physically meaningful quantities such as local charges, magnetizations, or interaction parameters [4].

Beyond finding the parameters for DMFT calculations, cDFT is routinely applied in several other fields of research. Constraining the charge on some molecular fragments allows one to explore the gradual transfer of an electron from one fragment to another, and provide parameters for Marcus theory [9, 10]. Constraining the spin magnetization in the neighborhood of an atom inside a solid allows one to obtain the energy of the system as a function of the local magnetization [8].

In this respect, and as a second example of a model Hamiltonian amenable to definition from cDFT, all the parameters of a Heisenberg model including isotropic exchange (J_{ij}), Dzyaloshinskii-Moriya interaction ($\mathbf{D}_{ij}$), and single-ion anisotropy (A_i), for a set of atoms labelled with i and j , with local spin magnetization vectors $\mathbf{S}_i$ and $\mathbf{S}_j$ [11–14]

$$H = - \sum_{\langle ij \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j - \sum_{\langle ij \rangle} \mathbf{D}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j) - \sum_i A_i (S_i^z)^2, \quad (2)$$

can, in principle, be extracted using constrained DFT for a given real magnetic material.

The specification of local magnetization or local charge can be combined with more usual variables governing the energy in first-principle calculations, like the atomic positions or cell parameters. Thus cDFT can provide parameters for models of magnetic state of matter (including Heisenberg model), with the associated description of magnons [8, 15, 16], but also their coupling of phonons. Similarly, parameters can be obtained for second-principles models [17–20], or for constituting training sets for the fitting of machine-learning interatomic potentials [21–24].

The implementations and applications of cDFT over the years have been numerous, and, for chemistry, have been reviewed by Kaduk and Van Voorhis in 2012 [4]. In 2016, a list of existing implementations has been collected by Melander and coworkers [25]. Then, one further implementation has been described by Hegde and coworkers [26]. The field being enormous, it will not be the purpose of the present lecture to cover all implementations or applications of cDFT, but to provide relevant entry points in the literature, and provide some examples of applications.

Several methods have been proposed to impose the constraints. In the first one [27, 4], an inner “micro”-self-consistency loop is added to the usual DFT self-consistency loop. In this inner loop, the potential (or local magnetic field) is varied to impose the constraint. In another one [28, 29], a penalty function is added to the energy functional. Recently, a potential-based self-consistency approach capable of placing the local magnetization, fragment charge, atomic positions and stresses on a par has been proposed and applied [30].

The present lecture is organized as follows. In Sec. 2, I recall the fundamental aspects of density-functional theory, including its spin-dependent generalization, approximations to the exchange-correlation functional, and the algorithms used to achieve self-consistency. This background is essential to understand both the formal and practical extensions introduced by

constrained approaches. This section might obviously be skipped by those who have already a good command of DFT.

In Sec. 3, the formal framework of constrained density-functional theory is presented, emphasizing its foundation within the Hohenberg-Kohn variational principle and the Levy constrained-search formulation. The types of constraints that can be rigorously formulated at the many-body level are discussed, as well as those defined at the Kohn-Sham level. Then, the approximations that arise in practical implementations are mentioned.

In Sec. 4, I review the main algorithmic strategies used to enforce constraints in practice, including Lagrange-multiplier formulations, penalty-function approaches, and recent developments based on potential-based self-consistent schemes. Their respective advantages, limitations, and interconnections, are highlighted, while recent progress in unifying these approaches are mentioned.

Finally, in Sec. 5, some capabilities of constrained DFT are shown through representative applications. These include the determination of model Hamiltonian parameters such as the Hubbard U , the exploration of magnetic energy landscapes and magnetoelastic coupling, and recent advances in dynamical response theory. In particular, I show how constrained density-functional perturbation theory provides access to nonadiabatic spin dynamics and reveals novel physical effects such as the emergence of a finite magnon mass.

The paper concludes in Sec. 6 with a summary of the main concepts and an outlook on future developments and applications of constrained DFT.

Atomic units ($|e| = 1$, $m_e=1$, $\hbar = 1$) are used throughout.

2 Background: density-functional theory

Density-functional theory (DFT) provides a formally exact framework to determine the ground-state properties of interacting electrons in an external potential v_{ext} . Its practical implementation relies on an effective single-particle description and on iterative self-consistency algorithms. In this section, I recall the main concepts that will be used in the later sections, including its spin generalizations, both collinear and non-collinear. Many readers, already knowledgeable of DFT, might skip this section, to dive directly in constrained density-functional theory specificities.

2.1 Hohenberg-Kohn and Kohn-Sham theorems

Hohenberg-Kohn theorems. The theoretical foundation of DFT is provided by the Hohenberg-Kohn (HK) theorems [31]. For a system of N interacting electrons in a local external one-body potential $v_{\text{ext}}(\mathbf{r})$, where the interaction $\hat{V}_{ee}$ is considered fixed while the external potential is allowed to vary, the first theorem establishes a one-to-one correspondence between the ground-state density $\rho(\mathbf{r})$ and the external potential (up to an additive constant). As a consequence, the ground-state density determines the Hamiltonian,

$$\hat{H}[\rho] = \hat{T} + \hat{V}_{ee} + \hat{v}_{\text{ext}}[\rho], \quad (3)$$

where $\hat{T}$ is the kinetic energy operator, and all properties (not only ground-state ones) are (formally) unique functionals of the ground-state density, except for the possible dependence on the above-mentioned additive constant.

The second HK theorem introduces a variational principle: for a given interaction $\hat{V}_{ee}$, there exists an universal functional $F[\rho]$, independent of v_{ext} , such that the ground-state energy is obtained by minimizing

$$E_{v_{\text{ext}}}[\rho] = F[\rho] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r}. \quad (4)$$

The exact ground-state energy is thus

$$E_{v_{\text{ext}}}^{\text{GS}} = \min_{\rho \rightarrow N} E_{v_{\text{ext}}}[\rho], \quad (5)$$

where $\rho \rightarrow N$ indicates that the electronic density integrates to N in the whole space.

The functional $F[\rho]$ is formally defined as

$$F[\rho] = \langle \Psi[\rho] | \hat{T} + \hat{V}_{ee} | \Psi[\rho] \rangle, \quad (6)$$

where $\Psi[\rho]$ is the interacting many-electron ground-state wavefunction obtained from the first HK theorem. This approach has the caveat that the set of densities over which the minimization has to be performed only includes those that are v -representable, that is, those that correspond to the ground state of a Hamiltonian for the defined interaction $\hat{V}_{ee}$ with some external local potential.

Levy constrained search approach. A rigorous formulation avoiding the notion of v -representability has been provided by Levy [32]. In his constrained-search approach, the functional $F[\rho]$ is determined from

$$F[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle, \quad (7)$$

the minimization being performed over all antisymmetric wavefunctions yielding the density ρ . This formulation is particularly useful conceptually, as it emphasizes that DFT is fundamentally a variational theory over densities. $F[\rho]$ contains all the many-body effects (kinetic + interaction). However, its explicit form is unknown. This expression serves primarily as an existence proof.

By the same token, the demonstration of the second HK theorem now becomes transparent. Indeed, the well-known extremal principle of quantum mechanics delivers

$$E_{v_{\text{ext}}} = \min_{\Psi \rightarrow N} \langle \Psi | \hat{H}_{v_{\text{ext}}} | \Psi \rangle. \quad (8)$$

In order to prove the second HK theorem, an intermediate minimization over charge densities is introduced,

$$E_{v_{\text{ext}}} = \min_{\rho \rightarrow N} \left\{ \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{H}_{v_{\text{ext}}} | \Psi \rangle \right\}, \quad (9)$$

as above, the Hamiltonian is decomposed,

$$\hat{H}_{v_{\text{ext}}} = \hat{T} + \hat{V}_{ee} + \hat{v}_{\text{ext}}, \quad (10)$$

so that, using the one-body characteristic of the external potential operator, one obtains:

$$E_{v_{\text{ext}}} = \min_{\rho \rightarrow N} \left\{ \min_{\Psi \rightarrow \rho} \left[\langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} \right] \right\}. \quad (11)$$

With the definition of the universal functional $F[\rho]$, Eq. (7) the variational problem becomes:

$$E_{v_{\text{ext}}} = \min_{\rho \rightarrow N} \left\{ F[\rho] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} \right\} = \min_{\rho \rightarrow N} E_{v_{\text{ext}}}[\rho]. \quad (12)$$

With such a relation, in mathematical terms, $E_{v_{\text{ext}}}$ and $F[\rho]$ are related to each other by a Legendre transform.

Kohn-Sham mapping to non-interacting fermions. The Kohn-Sham (KS) approach [33] provides a practical scheme to evaluate the ground-state density by mapping the interacting many-electron system onto an auxiliary system of non-interacting fermions reproducing the same density.

Kohn and Sham observed that the Hohenberg-Kohn theorems can be applied to interacting electrons, but also to non-interacting electrons. In the latter case, the ground-state of the many-body problem is known: a set of one-particle wavefunctions forming a Slater determinant (non-interacting electrons) to fulfill the antisymmetry requirement. They subtracted from the $F[\rho]$ functional the non-interacting electron kinetic energy $T_s[\rho]$ and the classical electrostatic energy of the electronic density (the so-called Hartree energy), $E_H[\rho]$, to define the exchange-correlation (xc) energy functional, incorporating all many-body effects beyond the classical electrostatic interaction

$$E_{\text{xc}}[\rho] = F[\rho] - T_s[\rho] - E_H[\rho], \quad (13)$$

with

$$T_s[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} | \Psi \rangle, \quad (14)$$

and

$$E_H[\rho] = \frac{1}{2} \int \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (15)$$

Matching the variational principles for the interacting and non-interacting electron cases, they derived a variational principle for the energy of a set of N “fictitious” non-interacting electrons as a functional of orthonormal one-electron occupied wavefunctions $\{\phi_i\}$

$$E_{v_{\text{ext}}}^{\phi_i}[\{\phi_i\}] = T_s[\{\phi_i\}] + \int \rho[\{\phi_i\}](\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} + E_{\text{Hxc}}[\rho[\{\phi_i\}]], \quad (16)$$

with non-interacting kinetic energy and electronic density given now in terms of the non-interacting independent wavefunctions by

$$T_s[\{\phi_i\}] = \sum_i f_i \langle \phi_i | \hat{T} | \phi_i \rangle, \quad (17)$$

and

$$\rho[\{\phi_i\}](\mathbf{r}) = \sum_i f_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}), \quad (18)$$

f_i being occupation numbers (e.g., $f_i = 2$ for doubly occupied wavefunctions, spin up and spin down, 0 otherwise).

The stationary condition of the functional $E_{v_{\text{ext}}}^{\phi_i}[\{\phi_i\}]$ under orthonormality constraints leads to the following set of equations (an unitary transformation is performed to obtain eigenfunctions)

$$(\hat{T} + \hat{v})|\phi_i\rangle = \varepsilon_i|\phi_i\rangle, \quad (19)$$

with the “Kohn-Sham” potential (the external potential screened by the potential created by the non-interacting electrons) defined as

$$v[\rho](\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{Hxc}}[\rho](\mathbf{r}), \quad (20)$$

with

$$v_{\text{Hxc}}[\rho](\mathbf{r}) = \frac{\delta E_{\text{Hxc}}}{\delta \rho(\mathbf{r})}, \quad (21)$$

or equivalently,

$$v_{\text{Hxc}}[\rho](\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{\text{xc}}}{\delta \rho(\mathbf{r})}, \quad (22)$$

and the density being obtained from the wavefunctions following Eq. (18).

The Kohn-Sham equations must be solved self-consistently, since the potential depends on the density, itself constructed from the one-electron wavefunctions. This will be examined further in Sec. 2.4.

2.2 Spin-dependent generalization of the basic theorems

The Hohenberg-Kohn and Kohn-Sham frameworks can be generalized to systems with spin degrees of freedom, as established by von Barth and Hedin [34]. In the general (non-collinear) case, the electronic state is described by a spinorial wavefunction $|\Psi\rangle$ (each electron spatial argument is complemented with a spinor argument), and the basic variable is the spin-density matrix $\rho_{\alpha\beta}(\mathbf{r})$. The values of α and β can be $\uparrow$ or $\downarrow$. The spin-density matrix can be decomposed as

$$\rho_{\alpha\beta}(\mathbf{r}) = \frac{1}{2}\rho(\mathbf{r})\delta_{\alpha\beta} + \frac{1}{2}\mathbf{m}(\mathbf{r}) \cdot \boldsymbol{\sigma}_{\alpha\beta}, \quad (23)$$

where $\mathbf{m}(\mathbf{r})$ is the local magnetization density, $\rho(\mathbf{r})$ is the scalar electronic density (the trace of the density matrix), and $\boldsymbol{\sigma}$ is the vector of Pauli matrices. The external potential becomes spin dependent and can be represented as a Hermitian matrix $v_{\text{ext},\alpha\beta}(\mathbf{r})$, which may equivalently be decomposed into a scalar potential and a magnetic field. In the collinear case, the spin-density matrix reduces to the pair of diagonal spin densities $\rho^\uparrow(\mathbf{r})$ and $\rho^\downarrow(\mathbf{r})$, and similarly for the external potential.

Spin-dependent Hohenberg-Kohn theorem. In the spin-dependent case, a naive extension of the first Hohenberg-Kohn theorem is not sufficient, since spin-dependent potentials with different longitudinal magnetic field may yield the same spin density (see the Appendix of Ref. [34]), referred to as the magnetic field gauge freedom. However, it can still be shown that,

for a non-degenerate ground state, the many-body wavefunction is uniquely determined by the spin-density matrix $\rho_{\alpha\beta}(\mathbf{r})$, and therefore all ground-state properties that do not depend on the magnetic field gauge are functionals of this quantity.

The corresponding variational principle generalizes straightforwardly

$$E_{v_{\text{ext}}}[\rho_{\alpha\beta}] = F[\rho_{\alpha\beta}] + \int \rho_{\alpha\beta}(\mathbf{r}) v_{\text{ext},\beta\alpha}(\mathbf{r}) d\mathbf{r}, \quad (24)$$

with

$$E_{v_{\text{ext}}}^{\text{GS}} = \min_{\rho_{\alpha\beta} \rightarrow N} E_{v_{\text{ext}}}[\rho_{\alpha\beta}], \quad (25)$$

and the universal functional F defined through the constrained search is

$$F[\rho_{\alpha\beta}] = \min_{\Psi \rightarrow \rho_{\alpha\beta}} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle. \quad (26)$$

Spin-dependent Kohn-Sham equations (non-collinear case). Following the same strategy as in the spin-independent case, one introduces a non-interacting system reproducing the spin-density matrix. The exchange-correlation functional is formally introduced as

$$E_{\text{xc}}[\rho_{\alpha\beta}] = F[\rho_{\alpha\beta}] - T_{\text{s}}[\rho_{\alpha\beta}] - E_{\text{H}}[\rho]. \quad (27)$$

The energy functional is then written as

$$E_{v_{\text{ext}}}[\rho_{\alpha\beta}] = T_{\text{s}}[\rho_{\alpha\beta}] + \int \rho_{\alpha\beta}(\mathbf{r}) v_{\text{ext},\beta\alpha}(\mathbf{r}) d\mathbf{r} + E_{\text{H}}[\rho] + E_{\text{xc}}[\rho_{\alpha\beta}], \quad (28)$$

where $\rho(\mathbf{r}) = \sum_{\alpha} \rho_{\alpha\alpha}(\mathbf{r})$ is the scalar electronic density.

The non-interacting kinetic energy is expressed in terms of one-electron spinorial wavefunctions $|\Phi_i\rangle$, with spinor components $|\phi_{i\alpha}\rangle$,

$$T_{\text{s}}[\{\Phi_i\}] = \sum_{i\alpha} f_i \langle \phi_{i\alpha} | \hat{T} | \phi_{i\alpha} \rangle, \quad (29)$$

and the spin-density matrix is obtained from the occupied wavefunctions as

$$\rho_{\alpha\beta}(\mathbf{r}) = \sum_i f_i \phi_{i\alpha}^*(\mathbf{r}) \phi_{i\beta}(\mathbf{r}). \quad (30)$$

Occupation numbers f_i are $f_i = 1$ for occupied wavefunctions, 0 otherwise.

Minimization of the functional under orthonormality constraints leads to a set of coupled one-particle equations

$$(\hat{T} + \hat{v})|\Phi_i\rangle = \varepsilon_i |\Phi_i\rangle, \quad (31)$$

where the effective potential is now a 2×2 Hermitian operator in spin space,

$$v_{\alpha\beta}(\mathbf{r}) = v_{\text{ext},\alpha\beta}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) \delta_{\alpha\beta} + v_{\text{xc},\alpha\beta}(\mathbf{r}), \quad (32)$$

with the Hartree potential depending only on the scalar electronic density,

$$v_{\text{H}}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (33)$$

and the exchange-correlation potential defined as the functional derivative

$$v_{xc,\alpha\beta}(\mathbf{r}) = \frac{\delta E_{xc}[\rho_{\alpha\beta}]}{\delta \rho_{\beta\alpha}(\mathbf{r})}. \quad (34)$$

The Kohn-Sham equations thus take the form of matrix Schrödinger equations in spin space, and must be solved self-consistently since the potential depends on the spin-density matrix constructed from the spinor wavefunctions.

Similarly to the representation of the spin-density matrix, the effective potential can be written alternatively as

$$v_{\alpha\beta}(\mathbf{r}) = v_{\text{ext},\alpha\beta}(\mathbf{r}) + v_{\text{Hsxc}}(\mathbf{r}) \delta_{\alpha\beta} + \frac{1}{2} (\mathbf{B}_{\text{xc}}(\mathbf{r}) \cdot \boldsymbol{\sigma})_{\alpha\beta}, \quad (35)$$

where v_{Hsxc} is a scalar potential (the Hartree potential plus half the trace of the xc potential) and $\mathbf{B}_{\text{xc}}$ acts as an effective magnetic field.

In the spin collinear case, spin-flip components (off-diagonal) of the Hamiltonian vanish, the two spin channels separate, and the eigenfunctions are pure spin states (either up or down). Accordingly, the spin-density matrix is diagonal, there is no dependence of the exchange-correlation energy functional on the off-diagonal components of the density matrix, which cause (self-consistently) the spin-flip components of the Hamiltonian to vanish. Explicitly, the Kohn-Sham equations become spin dependent

$$(\hat{T} + \hat{v}_{\sigma})|\phi_{i\sigma}\rangle = \varepsilon_{i\sigma}|\phi_{i\sigma}\rangle, \quad (36)$$

with $\sigma = \uparrow, \downarrow$, and effective potentials

$$v_{\sigma}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + v_{xc,\sigma}(\mathbf{r}), \quad (37)$$

where

$$v_{xc,\sigma}(\mathbf{r}) = \frac{\delta E_{xc}[\rho^{\uparrow}, \rho^{\downarrow}]}{\delta \rho^{\sigma}(\mathbf{r})}. \quad (38)$$

The spin densities are obtained from the spin-resolved KS wavefunctions as

$$\rho^{\sigma}(\mathbf{r}) = \sum_i f_{i\sigma} |\phi_{i\sigma}(\mathbf{r})|^2. \quad (39)$$

The scalar charge density used in the Hartree potential is

$$\rho(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) + \rho^{\downarrow}(\mathbf{r}). \quad (40)$$

While the collinear case covers ferromagnetic and antiferromagnetic materials, the non-collinear case leads to a fully coupled problem in spin space, which is essential for describing spin textures, spin spirals, spin-orbit coupling, and systems subject to spatially varying magnetic fields. In both collinear and non-collinear cases, the role of the exchange-correlation functional becomes more important than in the non-magnetic case, as the Hartree contribution does not encode any spin polarization effect, while such information is essential for the description of magnetic materials.

2.3 Approximate exchange-correlation functionals

In practice, the exact functional $E_{xc}[\rho]$ is unknown and must be approximated. The simplest and historically most important approximation is the local density approximation (LDA) [33], in which

$$E_{xc}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r}, \quad (41)$$

where ε_{xc} is the exchange-correlation energy density of the homogeneous electron gas.

The LDA can be generalized to spin-polarized systems, leading to the local spin-density approximation (LSDA) [34]. The key idea is to use, at each point in space, the exchange-correlation energy density of a homogeneous electron gas with the same local spin polarization.

Collinear case. In the collinear case, the basic variables are the spin densities $\rho^\uparrow(\mathbf{r})$ and $\rho^\downarrow(\mathbf{r})$, or equivalently the total density $\rho(\mathbf{r})$ and the magnetization density $m(\mathbf{r}) = \rho^\uparrow(\mathbf{r}) - \rho^\downarrow(\mathbf{r})$. The LSDA functional is written as

$$E_{xc}^{\text{LSDA}}[\rho^\uparrow, \rho^\downarrow] = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r})) d\mathbf{r}, \quad (42)$$

where

$$\rho(\mathbf{r}) = \rho^\uparrow(\mathbf{r}) + \rho^\downarrow(\mathbf{r}), \quad \zeta(\mathbf{r}) = \frac{\rho^\uparrow(\mathbf{r}) - \rho^\downarrow(\mathbf{r})}{\rho(\mathbf{r})} = \frac{m(\mathbf{r})}{\rho(\mathbf{r})}, \quad (43)$$

is the local spin polarization, and $\varepsilon_{xc}(\rho, \zeta)$ is the exchange-correlation energy per particle of a homogeneous spin-polarized electron gas.

The corresponding exchange-correlation potentials are obtained as functional derivatives with respect to the spin densities,

$$v_{xc,\sigma}(\mathbf{r}) = \frac{\delta E_{xc}^{\text{LSDA}}}{\delta \rho^\sigma(\mathbf{r})}, \quad \sigma = \uparrow, \downarrow, \quad (44)$$

which, using the chain rule, can be expressed in terms of derivatives of $\varepsilon_{xc}(\rho, \zeta)$ with respect to ρ and ζ .

Non-collinear case. In the general non-collinear case, the exchange-correlation functional depends on the spin-density matrix $\rho_{\alpha\beta}(\mathbf{r})$, or equivalently on the pair $(\rho(\mathbf{r}), \mathbf{m}(\mathbf{r}))$. The LSDA is constructed by assuming that, locally, the system behaves as a homogeneous electron gas with density $\rho(\mathbf{r})$ and magnetization magnitude $|\mathbf{m}(\mathbf{r})|$. One thus defines

$$E_{xc}^{\text{LSDA}}[\rho_{\alpha\beta}] = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho(\mathbf{r}), \zeta(\mathbf{r})) d\mathbf{r}, \quad (45)$$

with

$$\zeta(\mathbf{r}) = \frac{|\mathbf{m}(\mathbf{r})|}{\rho(\mathbf{r})}, \quad (46)$$

so that the functional depends only on rotationally invariant quantities in spin space. In particular, it is invariant under global spin rotations, as required for a proper non-collinear formulation.

The exchange-correlation potential is then obtained as a functional derivative with respect to the spin-density matrix,

$$v_{xc,\alpha\beta}(\mathbf{r}) = \frac{\delta E_{xc}^{\text{LSDA}}}{\delta \rho_{\beta\alpha}(\mathbf{r})}. \quad (47)$$

Using the decomposition in scalar and vector components, it can be written as

$$v_{xc,\alpha\beta}(\mathbf{r}) = v_{\text{sxc}}(\mathbf{r}) \delta_{\alpha\beta} + (\mathbf{B}_{\text{xc}}(\mathbf{r}) \cdot \boldsymbol{\sigma})_{\alpha\beta}, \quad (48)$$

where

$$v_{\text{sxc}}(\mathbf{r}) = \frac{\partial(\rho \varepsilon_{\text{xc}})}{\partial \rho}, \quad \mathbf{B}_{\text{xc}}(\mathbf{r}) = \frac{\partial(\rho \varepsilon_{\text{xc}})}{\partial \mathbf{m}} = \frac{\partial(\rho \varepsilon_{\text{xc}})}{\partial |\mathbf{m}|} \frac{\mathbf{m}(\mathbf{r})}{|\mathbf{m}(\mathbf{r})|}. \quad (49)$$

Within LSDA, the effective exchange-correlation magnetic field $\mathbf{B}_{\text{xc}}(\mathbf{r})$ is therefore locally parallel to the magnetization density $\mathbf{m}(\mathbf{r})$. This property reflects the fact that the approximation is based on a homogeneous reference system without spatial spin texture.

Generalized gradient approximations (GGA), such as PBE [35], improve upon LDA by including gradient corrections

$$E_{\text{xc}}^{\text{GGA}}[\rho] = \int f(\rho(\mathbf{r}), \nabla \rho(\mathbf{r})) d\mathbf{r}. \quad (50)$$

GGAs can be generalized to the spin-dependent case, both collinear and non-collinear. More elaborate approximations include meta-GGAs and hybrid functionals [36], which incorporate additional non-local information or exact exchange contributions, with also spin-dependent generalizations.

Despite their successes, approximate functionals exhibit well-known limitations, including self-interaction errors and difficulties with strongly correlated systems. These limitations motivate extensions of DFT such as DFT+ U [37, 1, 5] or constrained DFT, discussed in this lecture.

2.4 Algorithms for the self consistency

In this section, the discussion of self consistency applies generically to the non-spin-polarized, collinear spin-polarized and non-collinear spin-polarized cases. For sake of simplicity I have used the notation corresponding to the non-spin-polarized case. The adaptation to the other spin cases is rather easy.

The Kohn-Sham equations, Eqs. (18)–(22) define a non-linear self-consistent problem that is typically solved iteratively.

Eq. (20) delivers the effective Kohn-Sham potential (often denoted “screened” potential in the context of self-consistency) as a functional of the density. Similarly, one finds the density as a functional of the screened potential as follows. For a given trial screened potential u , associated with the local operator $\hat{u}$, one solves the corresponding Schrödinger equation

$$(\hat{T} + \hat{u})|\phi_{i,u}\rangle = \varepsilon_{i,u}|\phi_{i,u}\rangle, \quad (51)$$

which yields a set of Kohn-Sham wavefunctions $|\phi_{i,u}\rangle$. These wavefunctions are then used in Eq. (18) to construct the electronic density, thereby defining the density as a functional of the potential, denoted ρ^v

$$\rho^v[u] \triangleq \rho[\phi_{i,u}]. \quad (52)$$

The self-consistent density ρ^* is thus obtained as the fixed point of this mapping, satisfying

$$\rho^* = \rho^v[v[\rho^*]]. \quad (53)$$

In these expressions, the density, the potential, and the wavefunctions all depend on the spatial coordinate. For conciseness, this explicit dependence, e.g., as in Eq. (18), is omitted here and in most of the following expressions.

Many iterative techniques [38–41] have been developed over the years to tackle the self-consistency problem, Eqs. (18)–(22), of its brief formulation Eq. (53). The analysis of the self-consistency algorithms starts with some definitions. The following operator K associates to a trial input density, ρ^{in} , an output density, ρ_n^{out} ,

$$\rho^{\text{out}} = K[\rho^{\text{in}}] \triangleq \rho^v[v[\rho^{\text{in}}]]. \quad (54)$$

The discrepancy between the output and input densities,

$$R[\rho^{\text{in}}] \triangleq \rho^{\text{out}} - \rho^{\text{in}} = K[\rho^{\text{in}}] - \rho^{\text{in}}, \quad (55)$$

is usually referred as the density residual. Such density residual vanish at the fixed point. Iteratively, one starts from a guess input density, computes the output density and the corresponding residual, then goes to the next step with a better input density, etc. The vast majority of algorithms to solve this self-consistency problem rely on the knowledge of pairs of trial density and corresponding density residual, to infer the next trial density. The easiest algorithm to implement, simple mixing, is defined at step n by

$$\rho_{n+1}^{\text{in}} = \rho_n^{\text{in}} + \lambda R[\rho_n^{\text{in}}], \quad (56)$$

with λ being a tunable parameter, describing the incorporation of a fraction of the residual in the input density. If $\lambda = 1$, the output density is simply taken as the new input density. Most sophisticated algorithms take advantage of the history (at least the most recent part of it), and possibly include some preconditioning operator P , even varying at each step,

$$\rho_{n+1}^{\text{in}} = \rho_n^{\text{in}} + P_n \sum_{j=1}^n \lambda_j R[\rho_j^{\text{in}}], \quad (57)$$

where the set of λ_j parameters and the possible preconditioners P_n are computed on the flight from the history, and differ for different algorithms.

Instead of such density-based mixing approaches, potential-based mixing approaches have also been proposed [39]. In order to distinguish the (non-linear) operators appearing in this potential-based approach from those appearing in the density-based approach (namely K , R and P), they are labeled with a “v” superscript. In the potential-based approaches, instead of Eqs. (53)–(57), one relies on

$$v^* = v[\rho^v[v^*]], \quad (58)$$

$$v_n^{\text{out}} = K^v[v_n^{\text{in}}] \triangleq v[\rho^v[v_n^{\text{in}}]], \quad (59)$$

$$R^v[v_n^{\text{in}}] \triangleq v_n^{\text{out}} - v_n^{\text{in}} = K^v[v_n^{\text{in}}] - v_n^{\text{in}}, \quad (60)$$

$$v_{n+1}^{\text{in}} = v_n^{\text{in}} + P_n^v \sum_{j=1}^n \lambda_j R^v[v_j^{\text{in}}]. \quad (61)$$

The density- and potential-mixing approaches are dual of each other: in the case of usual (unconstrained) DFT, for each density-based mixing algorithm, there exists an equivalent potential-based mixing algorithm in which the pairs of density and corresponding density residual are replaced by pairs of potential and corresponding potential residual.

Focusing on the potential-based approach, the self-consistent electronic energy expression, is now formulated as a minimization problem in the space of trial screened potentials, as follows,

$$E_{v_{\text{ext}}}^{\text{SC}} = \min_u E_{v_{\text{ext}}}^v[u] \quad (62)$$

with

$$E_{v_{\text{ext}}}^v[u] = T[\{\phi_{i,u}\}] + \int \rho^v[u](\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} + E_{\text{Hxc}}[\rho^v[u]]. \quad (63)$$

The gradient of this functional of the potential has been computed in Ref. [42],

$$\frac{\delta E_{v_{\text{ext}}}^v[u]}{\delta u(\mathbf{r})} = \int \frac{\delta \rho^v[u](\mathbf{r}')}{\delta u(\mathbf{r})} (K^v[u](\mathbf{r}') - u(\mathbf{r}')) d\mathbf{r}' = \int \chi_0(\mathbf{r}, \mathbf{r}') R^v[u](\mathbf{r}') d\mathbf{r}'. \quad (64)$$

where the independent-particle susceptibility $\chi_0(\mathbf{r}, \mathbf{r}')$ is to be evaluated at the screened potential u . This gradient obviously vanishes at the minimum, since $R^v[v^*]$ vanishes. In practice, multiplication by χ_0^{-1} , like in Ref. [42], delivers a preconditioned gradient, which is nothing else than the residual $R^v[u]$, so that $\chi_0(\mathbf{r}, \mathbf{r}')$ does not even have to be computed. Hence, this approach shows that the usual potential-based self-consistency algorithms, Eq. (61) can be understood as mixing the preconditioned potential gradients of the electronic energy Eq. (63) from the current and previous steps.

3 Constrained density-functional theory: formalism and approximations

Ordinary DFT has the aim to compute the ground-state total energy and corresponding density/magnetization of a system of electrons placed in the potential of nuclei with known, fixed positions. Constrained DFT can enlarge the scope of ordinary DFT: for the same set of nuclei with known positions, the characterization of low-lying excited states or a superposition thereof is usually of great interest, and accessible by cDFT. With constraints, the system will usually not be in its ground state, but in a physically relevant state anyhow, or in a mixture of physically relevant states. Moreover, there is considerable freedom in the type of constraint that can be applied to explore such low-lying excited states. I will distinguish between the treatment of constraints that can be addressed at the Hohenberg-Kohn level, namely, those that are formulated in terms of the density and magnetization, or in terms of the correlated wavefunction (e.g. symmetry-constrained wavefunction), from those that are formulated in terms of quantities that are defined only at the Kohn-Sham level, e.g., that depend on the KS eigenenergies or the KS wavefunctions. For the first type of constraints, there exist analogs of the Hohenberg-Kohn theorems, based on the constrained-search strategy: the knowledge of an adequate exchange-correlation functional would deliver the exact energy under constraint.

3.1 Formal constrained density-functional theory

At the most fundamental level, constrained DFT can be formulated within the Hohenberg-Kohn framework by restricting the domain of the variational problem. This amounts to extending the Levy constrained-search formulation by imposing additional conditions on the admissible wavefunctions or densities.

Let $\hat{O}_I$ denote a set of observables (for example, symmetry operators, fragment charges, or local magnetizations, see the next subsection), and let N_I be the target values of their expectation values. The subscript I runs from 1 to the total number of constraints $I_{\max}$. One defines a constrained universal functional as

$$F^{\{N_I\}}[\rho_{\alpha\beta}] = \min_{\substack{\Psi \rightarrow \rho_{\alpha\beta} \\ \langle \Psi | \hat{O}_I | \Psi \rangle = N_I}} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle. \quad (65)$$

The corresponding variational principle becomes

$$E_{v_{\text{ext}}}^{\{N_I\}}[\rho_{\alpha\beta}] = F^{\{N_I\}}[\rho_{\alpha\beta}] + \int \rho_{\alpha\beta}(\mathbf{r}) v_{\text{ext},\beta\alpha}(\mathbf{r}) d\mathbf{r}, \quad (66)$$

with

$$E_{v_{\text{ext}}}^{\{N_I\}} = \min_{\rho_{\alpha\beta} \rightarrow N} E_{v_{\text{ext}}}^{\{N_I\}}[\rho_{\alpha\beta}]. \quad (67)$$

This formulation is a direct generalization of the Levy constrained-search approach, Eq. (7), in which the minimization domain is further restricted. Eq.(67) delivers the lowest possible energy of the many-body Hamiltonian for a wavefunction that fulfills the constraints. It provides a formal justification for constrained DFT whenever the constraints can be expressed as expectation values of operators defined at the many-body level.

One can then pursue with the definition of an exchange-correlation functional, then the variational principle for the Kohn-Sham energy:

$$E_{\text{xc}}^{\{N_I\}}[\rho_{\alpha\beta}] = F^{\{N_I\}}[\rho_{\alpha\beta}] - T_s^{\{N_I\}}[\rho_{\alpha\beta}] - E_{\text{H}}[\rho]. \quad (68)$$

The energy functional is written as

$$E_{v_{\text{ext}}}^{\{N_I\}}[\rho_{\alpha\beta}] = T_s^{\{N_I\}}[\rho_{\alpha\beta}] + \int \rho_{\alpha\beta}(\mathbf{r}) v_{\text{ext},\beta\alpha}(\mathbf{r}) d\mathbf{r} + E_{\text{H}}[\rho] + E_{\text{xc}}^{\{N_I\}}[\rho_{\alpha\beta}]. \quad (69)$$

Note that the different functionals F , T_s and E_{xc} formally acquire a dependence on the constraints. They are not identical to the DFT functionals. Indeed, for example,

$$T_s^{\{N_I\}}[\rho_{\alpha\beta}] = \min_{\substack{\Psi \rightarrow \rho_{\alpha\beta} \\ \langle \Psi | \hat{O}_I | \Psi \rangle = N_I}} \langle \Psi | \hat{T} | \Psi \rangle, \quad (70)$$

compare with Eq. (14).

In principle, if the exact constrained functional $F^{\{N_I\}}[\rho_{\alpha\beta}]$ (or equivalently the associated exchange-correlation functional) were known, the corresponding constrained energy would be obtained exactly. In practice, however, one uses the same approximate functionals as in ordinary DFT, which makes cDFT an approximate albeit powerful method to access low-lying excited states.

3.2 Constraints at the Hohenberg-Kohn level

Constrained DFT was introduced in 1984 by Dederichs *et al.* [8]. Their first example of a constraint was the charge contained in a given volume V ,

$$C_V[\rho] = \int_V \rho(\mathbf{r}) d\mathbf{r}. \quad (71)$$

Many examples of constraints of this type can be constructed in a unified way using suitable weight functions in real space. In general, such constraints are written as

$$C_I[\rho_{\alpha\beta}] = \int w_{I,\alpha\beta}(\mathbf{r}) \rho_{\beta\alpha}(\mathbf{r}) d\mathbf{r}, \quad (72)$$

where the weights $w_{I,\alpha\beta}(\mathbf{r})$ define the observable of interest.

A natural generalization of Eq. (71) is obtained by partitioning the system into fragments $A, B, \dots$, each associated with a weight function $w_I(\mathbf{r})$. The corresponding fragment charge is defined as

$$C_I[\rho] = \int w_I(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r}. \quad (73)$$

Such constraints allow one to enforce a given charge distribution for each fragment, with the sum of the charges over all fragments being the total number of electrons N , which is essential for the description of charge-transfer states or diabatic states.

Several families of weight functions are commonly used in practice [8, 4, 28, 30].

- *Characteristic (step) functions.* The simplest choice consists in partitioning space into non-overlapping regions Ω_A , and defining

$$w_A(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in \Omega_A, \\ 0 & \text{otherwise.} \end{cases} \quad (74)$$

This choice directly corresponds to the original formulation of Dederichs *et al.*, and is conceptually simple. However, the resulting potentials are discontinuous at the boundaries, which may lead to numerical difficulties.

- *Smooth spherical functions.* In this case, each weight function $w_{\text{rad}}(|\mathbf{r}_\kappa|)$ is centered around an atom κ , and depends only on the distance with respect to that atom, $\mathbf{r}_\kappa \triangleq \mathbf{r} - \boldsymbol{\tau}_\kappa$. The radial weight function $w_{\text{rad}}(r)$ is equal to 1 for r smaller than some cut-off radius r_{c1} , decreases smoothly beyond that radius, and becomes exactly zero beyond some other cut-off radius r_{c2} . An alternative formulation, more convenient for numerical evaluation uses

$$w_A(\mathbf{r}) = w_{\text{rad}2}(\mathbf{r}_\kappa^2), \quad (75)$$

with the following obvious relation $w_{\text{rad}}(t^{1/2}) = w_{\text{rad}2}(t)$.

However, using such smooth spherical functions, one cannot partition the whole space, and hence the total charge in fragments each attached to one atom. In the case of magnetization, still, this is not a big problem, since most of the relevant magnetization is localized in small atom-centered regions, while the magnetization is small in the interatomic zone.

- *Smooth real-space partitioning functions.* In order to avoid discontinuities while partitioning the space, one often introduces smooth weight functions satisfying for every point in space

$$\sum_I w_I(\mathbf{r}) = 1, \quad (76)$$

with the weight functions $w_I(\mathbf{r})$ varying continuously in space. The charge at every point in space is partitioned among the close atoms. Examples include partitionings based on atomic-centered functions leading to smooth interpolation between regions.

- *Density-based (promolecule) weights.* As a subclass of the latter, weights can be defined using reference atomic (spherical) or fragment densities $\rho_A^{(0)}(\mathbf{r})$, called “promolecules” as

$$w_A(\mathbf{r}) = \frac{\rho_A^{(0)}(\mathbf{r})}{\sum_I \rho_I^{(0)}(\mathbf{r})}. \quad (77)$$

Such definitions ensure that the partitioning follows the chemical environment and adapts to the spatial distribution of the density.

In spin-dependent DFT, one can similarly constrain the magnetization associated with a fragment A , using in the collinear case,

$$C_A[\rho_{\alpha\beta}] = \int w_A(\mathbf{r}) (\rho^\uparrow(\mathbf{r}) - \rho^\downarrow(\mathbf{r})) d\mathbf{r}, \quad (78)$$

or, in the non-collinear case,

$$\mathbf{C}_A[\rho_{\alpha\beta}] = \int w_A(\mathbf{r}) \mathbf{m}(\mathbf{r}) d\mathbf{r}. \quad (79)$$

These constraints are widely used to fix local magnetic moments or their orientation, and to extract magnetic coupling parameters.

Beyond fixing the magnitude of the local magnetization, it is often desirable to only constrain its direction, while its magnitude is allowed to vary. Let $\mathbf{n}_A$ be a unit vector specifying the desired direction of the magnetization. A constraint fixing the direction of $\mathbf{C}_A$ can then be formulated by imposing that its component orthogonal to $\mathbf{n}_A$ vanishes

$$\mathbf{n}_A \times \mathbf{C}_A[\rho_{\alpha\beta}] = \mathbf{0}. \quad (80)$$

Equivalently, one may require that

$$\mathbf{C}_A[\rho_{\alpha\beta}] = C_A \mathbf{n}_A, \quad (81)$$

where C_A is left free.

More generally, any observable that can be expressed as a linear functional of the density or spin-density matrix can serve as a constraint. This includes, for example, spatial multipole moments of the density,

$$C_A[\rho] = \int w_A(\mathbf{r}) r^l Y_{lm}(\hat{\mathbf{r}}) \rho(\mathbf{r}) d\mathbf{r}, \quad (82)$$

for the multipoles centered on the origin.

All these constraints share a common structure: they are linear functionals of $\rho_{\alpha\beta}$ and can therefore be incorporated within the Hohenberg-Kohn framework without leaving the density-based formalism. They define a restricted variational space corresponding to fixed values of selected collective variables, and thus provide access to a family of states beyond the ground state.

In addition to the constraints based on the electronic charge and on the magnetization, the Hohenberg and Kohn results (followed by the Kohn and Sham ones) can be extended to the lowest-energy state within a given symmetry sector, as shown by Gunnarsson and Lundqvist [43]. In their approach, the admissible wavefunctions are restricted to those compatible with a set of commuting operators $\{\hat{O}_I\}$

$$[\hat{H}, \hat{O}_I] = 0. \quad (83)$$

In this case, the wavefunction and the energy of the lowest state for the specified symmetry are functionals of the density (or spin-density), and a variational principle analogous to DFT holds (with formally modified F and E_{xc} functionals).

One can also introduce constraints targeting specific electronic subspaces, for example localized atomic-like orbitals (such as d or f manifolds). Provided such subspace is defined at the many-body wavefunction level (as well as the Kohn-Sham level), a cDFT formalism ensues. Denoting by $\rho_d(\mathbf{r})$ the contribution of a given subspace to the density, and relying on one of the weighting schemes previously defined, the associated constraint can be formulated as

$$C_d[\rho] = \int w(\mathbf{r}) \rho_d(\mathbf{r}) d\mathbf{r}. \quad (84)$$

Such constraints are closely related to the construction of effective Hubbard-like models and to the definition of interaction parameters.

3.3 Constrained density-functional theory at the Kohn-Sham level

One should emphasize that not all constraints can be formulated at the Hohenberg-Kohn level. Constraints involving explicitly Kohn-Sham quantities, such as orbital occupations or eigenvalues, or Kohn-Sham wavefunctions, have no direct many-body counterpart and are inherently tied to the auxiliary non-interacting system. These constraints fall outside the rigorous DFT framework and should be regarded as density-functional-inspired schemes rather than strict extensions of DFT.

By the way, constraining the occupation numbers of the Kohn-Sham wavefunctions gives a method identical to the well-known Δ SCF approach for computing the excited states of molecules. See Ref. [44] for an attempt to formulate the Δ SCF on a firm footing, this reference being also an entry point in the history of the Δ SCF method and its applications.

4 Constrained density-functional theory algorithms

The formal framework of constrained DFT, discussed in Sec. 3, does not by itself provide a practical prescription to enforce the constraints during the self-consistent solution of the Kohn-Sham equations. Algorithmic strategies must therefore be devised to impose the constraints while preserving numerical stability and convergence.

In this section, after framing the cDFT minimization problem in terms of Lagrange multipliers, I review the main approaches used in practice, emphasizing their underlying principles and their connection with standard self-consistent field (SCF) techniques.

Formally, the most direct way to impose constraints is through the introduction of Lagrange multipliers, following the original formulation of Dederichs *et al.* [8]. For a set of constraints $C_I[\rho_{\alpha\beta}] = N_I$, one defines the augmented functional

$$E_{v_{\text{ext}}}^{\{N_I\}+}[\rho_{\alpha\beta}, \{\lambda_I\}] = E_{v_{\text{ext}}}[\rho_{\alpha\beta}] + \sum_I \lambda_I (C_I[\rho_{\alpha\beta}] - N_I), \quad (85)$$

where λ_I are Lagrange multipliers. At the stationary point of $E_{v_{\text{ext}}}^{\{N_I\}+}$, one has

$$\frac{\delta E_{v_{\text{ext}}}^{\{N_I\}+}}{\delta \rho_{\alpha\beta}} = 0, \quad C_I[\rho_{\alpha\beta}] = N_I, \quad (86)$$

which ensures that the constraints are exactly satisfied.

In the case of generic constraints of the type Eq. (72), the Euler-Lagrange equations lead to modified Kohn-Sham equations where the effective potential is augmented by a constraining contribution,

$$v_{\alpha\beta}(\mathbf{r}) \longrightarrow v_{\alpha\beta}(\mathbf{r}) + \sum_I \lambda_I w_{I,\beta\alpha}(\mathbf{r}). \quad (87)$$

Within this formulation, the constrained ground-state energy corresponds to a saddle point (minimax problem),

$$E_{v_{\text{ext}}}^{\{N_I\}+} = \min_{\rho_{\alpha\beta}} \max_{\{\lambda_I\}} E_{v_{\text{ext}}}^{\{N_I\}+}[\rho_{\alpha\beta}, \{\lambda_I\}], \quad (88)$$

i.e., a minimization with respect to the density and a maximization with respect to the Lagrange multipliers [4].

4.1 Microiterations

A practical implementation of the Lagrange-multiplier method requires the simultaneous determination of the density (or wavefunctions) and of the multipliers $\{\lambda_I\}$. A possible strategy is based on nested iterative loops, often referred to as *microiterations* [4].

In this approach, one distinguishes: (1) an *outer loop*, in which the Kohn-Sham equations are solved self-consistently; (2) an *inner loop* (microiterations), in which the multipliers are updated so as to enforce the constraints.

The outer loop closely resembles a normal DFT calculation, with SCF iterations being performed to optimize the KS wavefunctions. Within each step of the outer loop, a second loop of

microiterations is performed to determine the Lagrange multipliers $\{\lambda_I\}$ that make the density satisfy the charge and spin constraints. After the first few iterations of the outer loop, it is common for the inner loop to converge after only two or three microiterations. While conceptually simple, this nested-loop structure may be computationally demanding. By the way, the reverse ordering might be possible: having an inner loop to determine density and wavefunctions for fixed Lagrange multipliers $\{\lambda_I\}$, while the outer loop involves a multi-dimensional search in the space of Lagrange parameters.

4.2 Penalty function approaches

An alternative to the Lagrange multiplier method consists in replacing the linear constraint term by a quadratic penalty function. The constrained energy functional is then written as

$$E_{v_{\text{ext}}}^{\{N_I\}\text{pen}}[\rho_{\alpha\beta}] = E_{v_{\text{ext}}}[\rho_{\alpha\beta}] + \sum_I \frac{\kappa_I}{2} (C_I[\rho_{\alpha\beta}] - N_I)^2, \quad (89)$$

where κ_I are penalty parameters.

This approach eliminates the need for auxiliary variables (the λ_I), and reduces the constrained optimization problem to a standard minimization. However, the constraints are only satisfied approximately, and enforcement becomes exact only in the limit $\kappa_I \rightarrow \infty$.

Penalty-based formulations are used in several practical implementations, including constrained non-collinear magnetism [28]. Indeed, in that work, the additional term added to the energy is quadratic in the deviation from the target configuration (see, e.g., Eq. (A.2) of Ref. [28]), which corresponds to a penalty functional rather than a Lagrange-multiplier formulation. Indeed, the added term scales as $(C_I - N_I)^2$, not linearly in $(C_I - N_I)$, and therefore does not enforce the constraint exactly at finite parameter values.

An important property of such schemes is that the penalty contribution to the energy vanishes as the constraint is satisfied. In particular, it can be shown that, near convergence,

$$E_{v_{\text{ext}}}^{\{N_I\}\text{pen}} - E_{v_{\text{ext}}}^{\{N_I\}+} \propto \frac{1}{\kappa_I}, \quad (90)$$

so that increasing κ_I improves the accuracy of the constraint at the expense of numerical stiffness, with associated difficulties to converge.

More recently, Royo and Stengel [29] have shown that the apparent limitations of penalty-function approaches can be overcome by exploiting exact Legendre-transform relations between different functionals. In their framework, the penalty functional Eq. (89) is interpreted as a convenient auxiliary functional, whose constrained variables (e.g. target moments) play the role of independent parameters. The physically relevant quantities, corresponding to the exact Lagrange-multiplier formulation, can then be recovered *a posteriori* from the penalty-based results via simple linear-algebra operations involving response matrices. In particular, second derivatives of the penalty functional yield a “stiffened” response (with improved numerical conditioning), which can be exactly transformed into the true constrained response associated with the variational principle of Eq. (85). This procedure establishes a one-to-one mapping between

penalty-function and Lagrange-multiplier approaches, demonstrating their formal equivalence at the level of observables. In addition to providing a practical route to accurate constrained calculations, this result justifies the use of penalty functionals as efficient preconditioners, combining numerical robustness with exact enforcement of the constraints after post-processing.

4.3 Lagrange multipliers again: a potential-based SCF approach

The potential-based SCF formulation introduced in Sec. 2.4, see Eqs. (58)–(64), can be extended to constrained problems by incorporating Lagrange multipliers directly into the potential-dependent energy functional [30]. A key advantage of this formulation is that it preserves the structure of the SCF problem, allowing one to reuse without modifications the algorithms developed for unconstrained DFT. In the present subsection, this approach is described for an electronic density single constraint; its extension to multiple spin-dependent constraints is discussed in Ref. [30].

For a single constraint $C_A[\rho] = N_A$, the augmented functional takes the form

$$E_{v_{\text{ext}}, N_A}^{v+}[u, \Lambda] = E_{v_{\text{ext}}}^v[u] + \Lambda(\rho_A^v[u] - N_A), \quad (91)$$

where I introduce the shorthand notation

$$\rho_A^v[u] \equiv C_A[\rho^v[u]], \quad (92)$$

and Λ is the associated Lagrange multiplier.

For a fixed value of Λ , the constrained problem reduces to a standard minimization with respect to the trial potential,

$$E_{v_{\text{ext}}, N_A}^+[\Lambda] = \min_u E_{v_{\text{ext}}, N_A}^{v+}[u, \Lambda]. \quad (93)$$

The gradient of $E_{v_{\text{ext}}}^v[u]$ with respect to the potential is given by Eq. (64). Applying the same derivation to the augmented functional leads to

$$\frac{\delta E_{v_{\text{ext}}, N_A}^{v+}[u, \Lambda]}{\delta u(\mathbf{r})} = \int \chi_0(\mathbf{r}, \mathbf{r}') R^{v+}(\mathbf{r}') d\mathbf{r}', \quad (94)$$

where the modified residual is defined as

$$R^{v+}(\mathbf{r}) = R^v(\mathbf{r}) + \Lambda w_A(\mathbf{r}). \quad (95)$$

From Eqs. (93) and (94), the self-consistent solution corresponds to a potential $u = v^*$ satisfying

$$0 = R^{v+}(\mathbf{r}) = v_{\text{out}}(\mathbf{r}) - v^*(\mathbf{r}) + \Lambda w_A(\mathbf{r}), \quad (96)$$

for all $\mathbf{r}$. Thus, the self-consistent condition can be interpreted as the requirement that the difference between output and input potentials lies entirely in the subspace spanned by the weight function w_A , with proportionality factor Λ .

An explicit expression for the Lagrange multiplier is obtained by projecting Eq. (96) onto w_A ,

$$\Lambda = -R_A^v[v^*] (W_{AA})^{-1}, \quad (97)$$

where

$$R_A^v[u] = \int R^v(\mathbf{r}) w_A(\mathbf{r}) d\mathbf{r}, \quad (98)$$

and

$$W_{AA} = \int w_A(\mathbf{r}) w_A(\mathbf{r}) d\mathbf{r}. \quad (99)$$

This result provides a direct and explicit determination of the Lagrange multiplier from the potential residual. This feature is specific to the potential-based formulation and contrasts with wavefunction- or density-based approaches, where the multiplier must typically be determined through an additional iterative procedure.

Going one step further, the dependence on Λ can be eliminated by inserting Eq. (97) back into the functional. Evaluating Λ at a general u leads to the definition of the functional

$$E_{v_{\text{ext}}, N_A}^{\text{cDFT}}[u] = E_{v_{\text{ext}}}^v[u] - R_A^v[u] (W_{AA})^{-1} (\rho_A^v[u] - N_A). \quad (100)$$

This functional depends only on the potential u , while both v_{ext} and N_A appear as external parameters on an equal footing. By construction, it yields the constrained ground-state energy when evaluated at the self-consistent potential,

$$E_{v_{\text{ext}}, N_A}^{\text{SC}} = E_{v_{\text{ext}}, N_A}^{\text{cDFT}}[v^*], \quad (101)$$

and is stationary with respect to variations of u around v^* ,

$$E_{v_{\text{ext}}, N_A}^{\text{cDFT}}[u] = E_{v_{\text{ext}}, N_A}^{\text{cDFT}}[v^*] + \mathcal{O}((u - v^*)^2). \quad (102)$$

The corresponding functional derivative reads

$$\frac{\delta E_{v_{\text{ext}}, N_A}^{\text{cDFT}}[u]}{\delta u(\mathbf{r})} = \int \chi_0(\mathbf{r}, \mathbf{r}') R^{v+}(\mathbf{r}') d\mathbf{r}' + \left(\int \epsilon_e(\mathbf{r}, \mathbf{r}') w_A(\mathbf{r}') d\mathbf{r}' \right) (W_{AA})^{-1} (\rho_A^v[u] - N_A), \quad (103)$$

with

$$\Lambda_A[u] = -R_A^v[u] (W_{AA})^{-1}, \quad (104)$$

and where $\epsilon_e(\mathbf{r}, \mathbf{r}')$ is the electron dielectric response function [42, 30].

At self-consistency, both $R^{v+}[v^*, \Lambda_A[v^*]]$ and $\rho_A^v[v^*] - N_A$ vanish, which confirms the stationary character of the functional.

An important property of $\Lambda_A[u]$ is that it enforces the orthogonality condition

$$\int R^{v+}(\mathbf{r}) w_A(\mathbf{r}) d\mathbf{r} = 0, \quad (105)$$

i.e., it removes the component of the residual along the constraint direction.

This naturally motivates the introduction of a modified residual adapted to constrained DFT,

$$R^{\text{cDFT}}(\mathbf{r}) = R^{v+}(\mathbf{r}) + c w_A(\mathbf{r}) (\rho_A^v[u] - N_A),$$

where c is a numerical parameter of order unity.

Since the two contributions to this modified residual belong to orthogonal subspaces, the condition $R^{\text{cDFT}}[u] = 0$ ensures simultaneously the fulfillment of the SCF condition and of the constraint. This formulation provides a direct route to extend standard potential-based SCF algorithms to constrained DFT calculations.

5 Some applications of constrained density-functional theory

The computation of the Hubbard U for use in DMFT. Following Cococcioni and de Gironcoli [7], an important application of constrained density-functional theory is the determination of effective interaction parameters entering low-energy many-body models, such as the U parameter in the Hubbard model or extensions thereof, see e.g., Ref. [5, 45], used in DFT+DMFT approaches. In this formulation based on constrained DFT and linear-response theory, U is computed in a fully first-principles and internally consistent manner. The key idea is to evaluate the curvature of the total energy with respect to variations of the occupation of a localized orbital subspace associated with a given atomic site.

In this framework, one considers the total energy as a functional of constrained occupations $\{q_I\}$ of localized orbitals,

$$E[\{q_I\}] = \min_{n(\mathbf{r}), \{\Lambda_I\}} \left\{ E[n(\mathbf{r})] + \sum_I \Lambda_I (n_I - q_I) \right\}, \quad (106)$$

where n_I denotes the occupation of the localized manifold on site I , and Λ_I are Lagrange multipliers enforcing the constraint.

The Hubbard interaction is then defined, with respect to the non-self-consistent non-interacting Hamiltonian, as the curvature of the interacting energy with respect to the occupation of a given site,

$$U = \frac{\partial^2 E[\{q_I\}]}{\partial q_I^2}, \quad (107)$$

evaluated at the ground-state occupation.

However, in the DFT case, this curvature contains mean-field contributions that are not related to explicit electron-electron interactions, but that appear in the Kohn-Sham treatment. To isolate the physically meaningful interaction, the same construction is applied to the Kohn-Sham energy,

$$E_{\text{KS}}[\{q_I\}] = \min_{n(\mathbf{r}), \{\Lambda_I\}} \left\{ E_{\text{KS}}[n(\mathbf{r})] + \sum_I \Lambda_I^{\text{KS}} (n_I - q_I) \right\}, \quad (108)$$

and the effective Hubbard parameter is defined as the difference between the interacting and non-interacting curvatures,

$$U = \frac{\partial^2 E[\{q_I\}]}{\partial q_I^2} - \frac{\partial^2 E_{\text{KS}}[\{q_I\}]}{\partial q_I^2}. \quad (109)$$

A key simplification arises from the observation that the derivatives of the energy with respect to occupations can be expressed in terms of the Lagrange multipliers,

$$\frac{\partial E}{\partial q_I} = -\Lambda_I, \quad \frac{\partial^2 E}{\partial q_I^2} = -\frac{\partial \Lambda_I}{\partial q_I}, \quad (110)$$

and similarly for the Kohn-Sham system. This allows one to rewrite Eq. (109) as

$$U = \frac{\partial \Lambda_I^{\text{KS}}}{\partial q_I} - \frac{\partial \Lambda_I}{\partial q_I}. \quad (111)$$

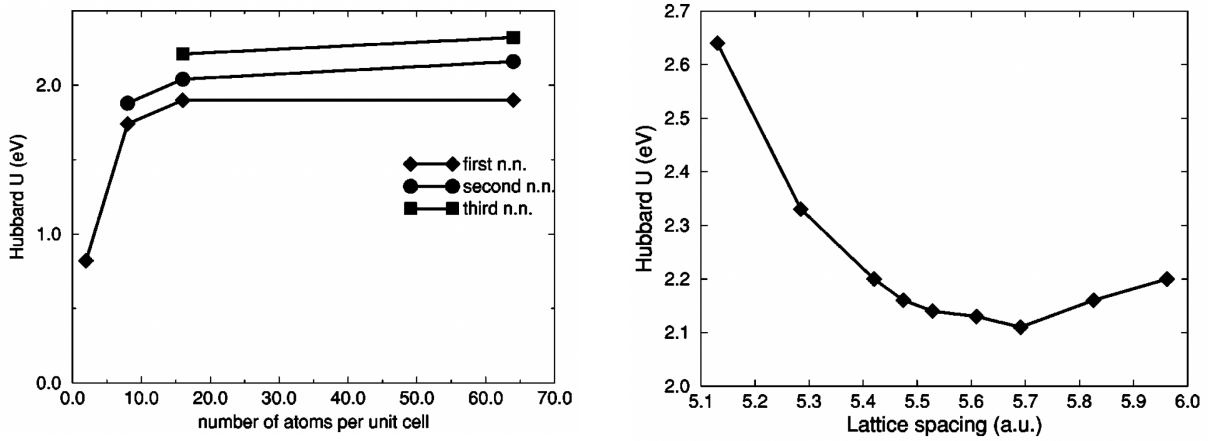


Fig. 1: (Left) Calculated Hubbard U in metallic iron for different supercells. Lines connect results from the cell-extrapolation procedure and different symbols correspond to inclusion of screening contributions up to the indicated shell of (nearest-)neighbors of the perturbed atom. (Right) Lattice spacing dependence of the calculated Hubbard U parameter for iron. Reprinted figures from Ref. [7] with permission. Copyright ©2005 The American Physical Society.

In practice, it is more convenient to perform a Legendre transform and work with the Lagrange multipliers as independent variables. In this representation, the perturbation induced by Λ_I corresponds to a shift of the local potential acting on the selected subspace,

$$\Delta v(\mathbf{r}) = \sum_I \Lambda_I P_I, \quad (112)$$

where P_I is the projector onto the localized orbitals of site I .

One then introduces the interacting and non-interacting response matrices

$$\chi_{IJ} = \frac{\partial n_I}{\partial \Lambda_J}, \quad \chi_{IJ}^0 = \frac{\partial n_I}{\partial \Lambda_J^{\text{KS}}}, \quad (113)$$

which describe how localized occupations respond to local potential shifts.

Within this linear-response framework, the effective Hubbard interaction is obtained as

$$U = (\chi_0^{-1} - \chi^{-1})_{II}, \quad (114)$$

which highlights the screening effects due to the electronic environment. This expression is formally analogous to a screened interaction in linear-response theory, where χ^0 represents the bare (non-interacting) response and χ the fully screened one.

The resulting values of U can then be used in DFT+ U or DFT+DMFT calculations, establishing a rigorous connection between first-principles electronic structure methods and many-body model Hamiltonians.

In practical calculations, the response functions are evaluated numerically by applying small local perturbations to the Kohn-Sham potential and monitoring the changes in the occupations of the localized orbitals. Importantly, the screening is computed within supercells of increasing size, without artificially decoupling the localized subspace, ensuring a realistic description of the environment.

The convergence of the computed Hubbard interaction with respect to the spatial extent of the screening can be illustrated using the case of metallic iron, as shown in the left side of Fig. 1. The effective interaction parameter U is obtained from the linear-response expression Eq. (114), and depends on the range of the induced charge redistribution following a local perturbation. By including progressively larger coordination shells in the response functions, one observes that the value of U converges, reflecting the degree of spatial delocalization of the screening charge. In metallic systems such as iron, screening is relatively long ranged, and contributions from several coordination shells are required to obtain a quantitatively accurate value of U . On the right side of Fig. 1, the dependence of U on the lattice parameter of Fe (e.g. for different pressures) is shown. This behavior highlights the importance of treating screening within the full crystal environment, rather than within an artificially localized or atomic picture.

Constraining the direction of local magnetization in BCC Fe. I now turn to a representative application of cDFT to explore the energy landscape as a function of local spin magnetizations in addition to the more usual DFT variables like atomic positions and lattice parameters. In this application, I consider body-centered cubic (bcc) iron, using a two-atom conventional cell, in which the directions of local magnetization vectors on each sublattice are independently controlled [30]. In this setup, the angle between the magnetization vectors of the two sublattices is continuously varied from the ferromagnetic configuration (0°) to the antiferromagnetic configuration (180°), without constraining the magnetization vector length, that is determined self-consistently to minimize the energy.

The resulting dependence of the total energy and local magnetic moment on the spin angle is shown in the left column of Fig. 2. The calculated energy variation from different methodologies, including the potential-based self-consistent algorithm, agree reasonably well, for both LDA and GGA(PBE) functionals. The magnitude of the local magnetization decreases as the angle between neighboring spins increases, reflecting the gradual destabilization of the ferromagnetic alignment. Interestingly, the evolution of the magnetic moment as a function of the spin angle is not perfectly smooth, but instead exhibits small irregularities despite the high numerical precision achieved in the calculations. This behavior is attributed to non-smooth changes in the electronic density of states, in particular the presence of Van Hove critical points, that shift as the spin configuration evolves.

Beyond such validation step, the potential-based SCF method enables the investigation of magnetoelastic coupling by analyzing the response of the stress tensor to changes in the magnetic configuration. When the lattice parameter is kept fixed, the rotation of the magnetic moments induces significant variations in the stress, which can be conveniently characterized through the pressure (the trace of the stress tensor). As shown in the top right panel of Fig. 2, the pressure varies over a wide range, reaching amplitudes of the order of 8–12 GPa depending on the exchange-correlation functional used. This variation is substantial when compared to the bulk modulus of iron, highlighting the strong coupling between magnetic order and mechanical properties.

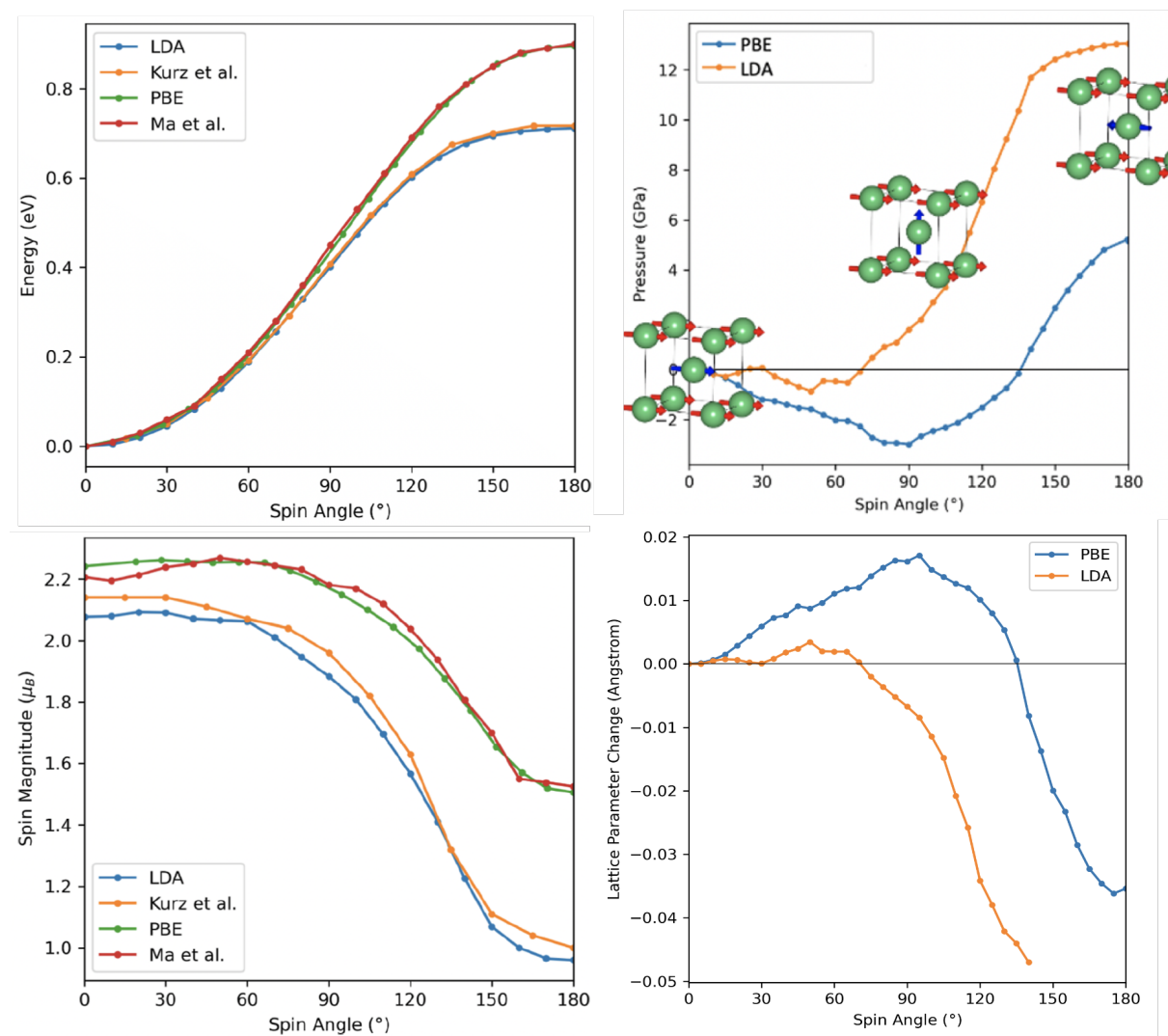


Fig. 2: Energy (top, left), pressure (top, right), spin magnitude (bottom, left) and relaxed lattice parameter (bottom, right), as a function of the spin angle between the two sublattices in a Fe BCC conventional cell. The top right figure also illustrates the relative angle between the two spin sublattices. LDA data (blue) from Ref. [30] are compared to those from Kurz et al. [46] (orange), while GGA(PBE) data (green) from Ref. [30] are compared to those of Ma and Dudarev [28] (red). Reproduced from Ref. [30] with permission. Copyright ©2022 The Authors. Published by the American Chemical Society.

Allowing for structural relaxation at fixed spin angle further reveals the consequences of this coupling at the level of equilibrium geometry. For each imposed magnetic configuration, the lattice parameter can be optimized, yielding the results shown in the bottom right panel of Fig. 2. The evolution of the equilibrium lattice parameter closely follows the trends observed in the pressure, demonstrating consistency between the stress response and total-energy minimization. The magnitude of the lattice parameter variation reaches approximately 0.06 Å, corresponding to about 2% of the total lattice parameter, which is a sizable structural effect. As in the case of the magnetization, a slightly irregular dependence on the spin angle is observed, again reflecting subtle features in the electronic structure.

Altogether, this example demonstrates that the constrained DFT formalism allows one to treat magnetic degrees of freedom, structural parameters, and mechanical response on an equal footing. It provides direct access not only to the energy landscape associated with noncollinear magnetic configurations, but also to the associated stresses and equilibrium structures, thereby enabling a comprehensive characterization of magnetoelastic couplings within a unified first-principles approach. Such capabilities are essential for the construction of effective model Hamiltonians, second-principles methods, or machine-learning potentials that explicitly incorporate coupled spin and lattice degrees of freedom.

Electron-inertia-induced magnon mass within constrained DFPT. The perturbative version of cDFT gives the possibility to access, in a controlled and numerically stable manner, the nonadiabatic dynamics of spin degrees of freedom, including their coupling to lattice and electronic excitations. In this context, the recent work of Royo and Stengel (RS) [29] provides a remarkable advance by establishing a unified framework to compute the frequency-dependent response of noncollinear magnets under constrained magnetic moments, and by demonstrating the emergence of finite magnon inertia. In what follows, I summarize the main formal ingredients of this result, emphasize its connection to the constrained functional approach, and discuss its physical implications for spin dynamics.

A key conceptual step of the theory consists in performing a systematic expansion of the constrained functional in powers of frequency. The spin-phonon Hessian can be expressed as

$$U(\omega) = K + i\omega G - \omega^2 M + \dots, \quad (115)$$

where K is the static (adiabatic) Hessian with respect to atomic displacements and local magnetization (that is, the matrix of interatomic force constants, spin susceptibilities and their adiabatic couplings), G is the Berry-curvature tensor governing precessional dynamics of spins, and M is a mass tensor arising at second order in frequency.

While the first two terms are familiar from adiabatic phonon and the semiclassical theory of magnons, the third term represents a genuine nonadiabatic correction. Crucially, the tensor M is not limited to ionic contributions in the phonon sector: it also contains an electronic component stemming from the frequency dispersion of the response functions.

When restricting the coupled atomic displacement – magnetization dynamical equation to the spin-spin sector, one obtains a generalized equation of motion for the local spin amplitudes

$$\sum_l M_{jl} \ddot{s}_l - G_{jl} \dot{s}_l + K_{jl} s_l = 0, \quad (116)$$

where j and l label atomic sites. This result represents a fundamental modification of the conventional Landau-Lifshitz dynamics [47]. Indeed, the RS formalism introduces a second-order inertial term proportional to $M\ddot{s}$. From a physical standpoint, this implies that magnons behave as particles with finite effective mass, rather than purely massless precessional modes.

The microscopic origin of this mass term is explicitly identified in the theory [29]: the additional contribution arises from “the inertia of the electrons that are dragged by a given degree

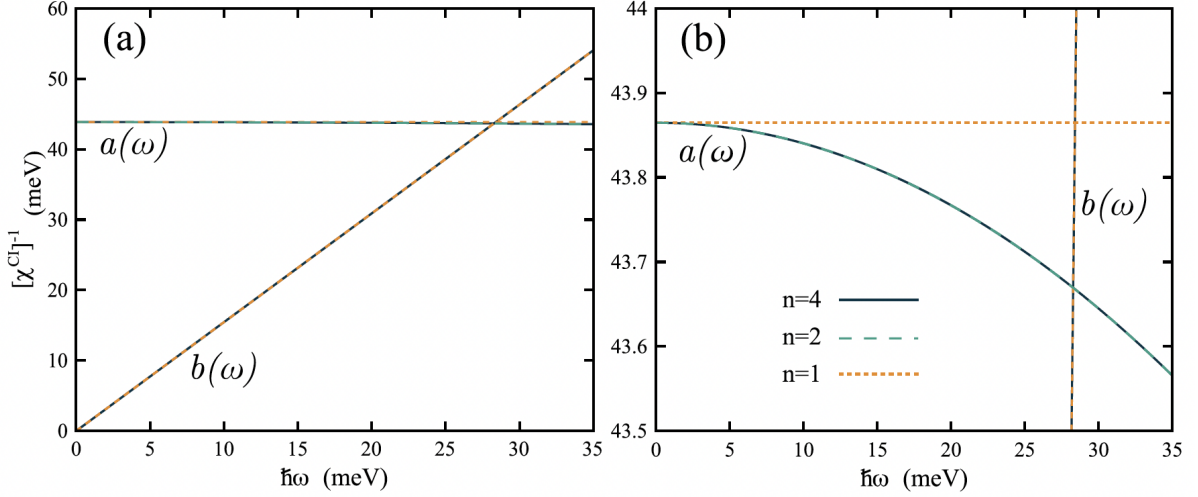


Fig. 3: (a) Frequency dependence of the optical-magnon coefficients of Eq. (117), as obtained from polynomial frequency interpolation at different degrees: $n=4$ (dark solid line), $n=2$ (turquoise dashed line), and $n=1$ (orange dashed line). (b) shows an amplified view of the abscissa axis around the intersection of the curves shown in (a). Reproduced from Royo and Stengel [29] under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

of freedom along its trajectory”. In other words, during spin precession, the electronic cloud cannot instantaneously follow the rotation of the local magnetization. The resulting delay produces an effective kinetic energy term, which manifests itself at second order in the frequency expansion.

Royo and Stengel examine the size of this correction in the case of ferromagnetic CrI_3 , a van der Waals ferromagnet that remains spin-ordered down to the monolayer limit [48]. Its properties arise from the interplay of electronic, vibrational, and magnetic degrees of freedom. Magnon-phonon coupling has been treated with model Hamiltonians for this material [49].

Although small, the effect can be seen in the analysis of the optical magnon response. Royo and Stengel define

$$u_{\pm}(\omega) = a(\omega) \mp b(\omega) \approx K \mp \omega G - \omega^2 M + \dots \quad (117)$$

The zeros of $u_{\pm}$ correspond to the magnon resonances. The “+” and “−” modes correspond to clockwise (+) and counterclockwise (−) spin precession.

Fig. 3 shows the frequency dependence of the coefficients entering the magnon propagator. In particular: the coefficient $b(\omega)$ is nearly linear in ω , consistent with the Berry-curvature term; the coefficient $a(\omega)$, however, exhibits a very small quadratic curvature which cannot be captured within the standard adiabatic approximation. At finite frequencies (e.g. THz regime), inertial effects may become significant and modify the propagation and damping of spin waves.

6 Conclusion

In this lecture, I have presented the formal foundations, algorithmic strategies, and selected applications of constrained density-functional theory (cDFT). Starting from the Hohenberg-Kohn framework and the Levy constrained-search formulation, I have shown that cDFT can be understood as a natural extension of density-functional theory in which the variational space is restricted to enforce physically meaningful conditions, such as fragment charges, local magnetizations, or symmetry constraints. This perspective provides a rigorous grounding for a wide range of practical methods that go beyond the standard ground-state description.

At the formal level, I have emphasized that constraints formulated in terms of (spin-)density functionals retain a strict connection with many-body quantum mechanics, whereas constraints defined at the Kohn-Sham level introduce additional approximations. In practice, however, both classes of approaches play an essential role in modern electronic-structure calculations, enabling access to excited-state properties, diabatic states, and effective model parameters.

On the algorithmic side, I have reviewed the main strategies used to enforce constraints, including the Lagrange-multiplier formulation with micro-iterations, penalty-function approaches, and the more recent development based on potential-based self-consistent schemes. The latter method exhibit the most straightforward connection between constrained and unconstrained DFT calculations. Recent advances have further shown that penalty-function methods can be improved, and considered as preconditioned versions of the constrained problem, with exact results recovered through appropriate transformations.

The applications discussed in this work illustrate several aspects of constrained DFT. In particular, I have highlighted its capability in the construction of effective Hamiltonians, such as the determination of Hubbard interaction parameters or magnetic exchange couplings, as well as its use in exploring complex energy landscapes involving coupled spin, charge, and structural degrees of freedom. The ability of cDFT to treat these variables on an equal footing makes it a powerful tool for bridging first-principles calculations and model-based descriptions of condensed matter systems.

A particularly interesting recent development in constrained density-functional theory, is the ability of cDFT to provide parameters to study the dynamical regime of models. In this framework, constraints not only improve the numerical stability of response calculations but also enable the systematic investigation of nonadiabatic effects. The emergence of an effective magnon mass associated with electron inertia provides a striking example of how constrained approaches can uncover subtle physical phenomena that are beyond the reach of traditional adiabatic theories, and calls for a refinement of established models such as the Landau-Lifshitz description of spin dynamics.

Looking forward, constrained DFT is expected to continue to play an important role in several areas of computational materials science and chemistry. These include the construction of accurate low-energy Hamiltonians for strongly correlated systems, the development of second-principles and machine-learning models incorporating electronic and magnetic degrees of freedom, and the investigation of ultrafast and non-equilibrium phenomena where coupled electronic, spin, and lattice dynamics are essential.

References

- [1] R.M. Martin: *Electronic structure. Theory and practical methods.* (Cambridge University Press, Cambridge, 2004)
- [2] F. Giustino: *Materials Modelling using Density Functional Theory* (Oxford University Press, 2014)
- [3] A.M. Teale, T. Helgaker, and A. Savin, *Phys. Chem. Chem. Phys.* **24**, 28700 (2022)
- [4] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, *Chem. Rev.* **112**, 321 (2012)
- [5] R.M. Martin, L. Reining, and D.M. Ceperley: *Interacting Electrons. Theory and Computational Approaches.* (Cambridge University Press, Cambridge, 2016)
- [6] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [7] M. Cococcioni and S. de Gironcoli, *Phys. Rev. B* **71**, 035105 (2005)
- [8] P.H. Dederichs, S. Blügel, R. Zeller, and H. Akai, *Phys. Rev. Lett.* **53**, 2512 (1984)
- [9] R.A. Marcus, *J. Chem. Phys.* **24**, 966 (1956)
- [10] R.A. Marcus and N. Sutin, *Biochim. Biophys. Acta – Rev. Bioenerg.* **811**, 265 (1985)
- [11] S. Blundell: *Magnetism in Condensed Matter* (Oxford University Press, 2001)
- [12] I. Dzyaloshinskii, *J. Phys. Chem. Solids* **4**, 241 (1958)
- [13] T. Moriya, *Phys. Rev.* **120**, 91 (1960)
- [14] H. Komatsu, Y. Nonomura, and M. Nishino, *Phys. Rev. B* **103**, 214404 (2021)
- [15] R. Gebauer and S. Baroni, *Phys. Rev. B* **61**, R6459 (2000)
- [16] D.M. Bylander, Q. Niu, and L. Kleinman, *Phys. Rev. B* **61**, R11875 (2000)
- [17] P. Garcia-Fernandez, J.C. Wojdeł, J. Iniguez, and J. Junquera, *Phys. Rev. B* **93**, 195137 (2016)
- [18] A.R. Damodaran, J.D. Clarkson, Z. Hong, H. Liu, A.K. Yadav, C.T. Nelson, S.-L. Hsu, M.R. McCarter, K.-D. Park, V. Kravtsov, A. Farhan, Y. Dong, Z. Cai, H. Zhou, P. Aguado-Puente, P. García-Fernández, J. Íñiguez, J. Junquera, A. Scholl, M.B. Raschke, L.-Q. Chen, D.D. Fong, R. Ramesh, and L. Martin, *Nat. Mater.* **16**, 1003 (2017)
- [19] C. Escorihuela-Sayalero, J.C. Wojdeł, and J. Iniguez, *Phys. Rev. B* **95**, 094115 (2017)
- [20] P. Shafer, P. García-Fernández, P. Aguado-Puente, A.R. Damodaran, A.K. Yadav, C.T. Nelson, S.-L. Hsu, J.C. Wojdeł, J. Íñiguez, L.W. Martin, E. Arenholz, J. Junquera, and R. Ramesh, *Proc. Natl. Acad. Sci. USA* **115**, 915 (2018)

- [21] I. Novikov, B. Grabowski, F. Körmann, and A. Shapeev, *npj Comput. Mater.* **8**, 1 (2022)
- [22] S. Nikolov, M. Wood, A. Cangi, J. Maillet, M. Marinica, A. Thompson, M. Desjarlais, and J. Tranchida, *npj Comput. Mater.* **7**, 1 (2021)
- [23] M. Eckhoff and J. Behler, *npj Comput. Mater.* **7**, 1 (2021)
- [24] R. Drautz, *Phys. Rev. B* **102**, 024104 (2020)
- [25] M. Melander, E. Ö. Jónsson, J.J. Mortensen, T. Vegge, and J.M.G. Lastra, *J. Chem. Theory Comput.* **12**, 5367 (2016)
- [26] O. Hegde, M. Grabowski, X. Zhang, O. Waseda, T. Hickel, C. Freysoldt, and J. Neugebauer, *Phys. Rev. B* **102**, 144101 (2020)
- [27] Q. Wu and T. Van Voorhis, *J. Chem. Theory Comput.* **2**, 765 (2006)
- [28] P.-W. Ma and S.L. Dudarev, *Phys. Rev. B* **91**, 054420 (2015)
- [29] M. Royo and M. Stengel, *Phys. Rev. X* **16**, 011049 (2026)
- [30] X. Gonze, B. Seddon, J. Elliott, C. Tantardini, and A. Shapeev, *J. Chem. Theory Comput.* **18**, 6099 (2022)
- [31] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964)
- [32] M. Levy, *Proc. Natl. Acad. Sci. USA* **76**, 6062 (1979)
- [33] W. Kohn and L.J. Sham, *Phys. Rev.* **140**, A1133 (1965)
- [34] U.V. Barth and L. Hedin, *J. Phys. C: Solid State Phys.* **5**, 1629 (1972)
- [35] J.P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996)
- [36] J. Heyd, G.E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003)
- [37] V.I. Anisimov, J. Zaanen, and O.K. Andersen, *Phys. Rev. B* **44**, 943 (1991)
- [38] P. Pulay, *J. Comp. Chem.* **3** (1982)
- [39] D. Vanderbilt and S.G. Louie, *Phys. Rev. B* **30**, 6118 (1984)
- [40] L.D. Marks and D.R. Luke, *Phys. Rev. B* **78**, 075114 (2008)
- [41] N.D. Woods, M.C. Payne, and P.J. Hasnip, *J. Phys.: Condens. Matter* **31**, 453001 (2019)
- [42] X. Gonze, *Phys. Rev. B* **54**, 4383 (1996)
- [43] O. Gunnarsson and B. Lundqvist, *Phys. Rev. B* **13**, 4274 (1976)

- [44] W. Yang and P.W. Ayers, arXiv:2403.04604
- [45] S. Kundu and D. Sénéchal, SciPost Phys. Core **7**, 033 (2024)
- [46] P. Kurz, F. Förster, L. Nordström, G. Bihlmayer, and S. Blügel, Phys. Rev. B **69**, 024415 (2004)
- [47] L.D. Landau and E.M. Lifshitz, Phys. Z. Sowjetunion **8**, 153 (1935)
- [48] B. Huang, G. Clark, E. Navarro-Moratalla, D.R. Klein, R. Cheng, K.L. Seyler, D. Zhong, E. Schmidgall, M.A. McGuire, D.H. Cobden, W. Yao, D. Xiao, P. Jarillo-Herrero, and X. Xu, Nature **546**, 270 (2017)
- [49] P. Delugas, O. Baseggio, I. Timrov, S. Baroni, and T. Gorni, Phys. Rev. B **107**, 214452 (2023)

3 Green Functions and Self-Energy Functionals

Robert Eder
Institute for Quantum Materials and Technologies
Karlsruhe Institute of Technology

Contents

1	Introduction	2
2	Green functions and the self-energy	3
2.1	The Green function and its analytical properties	3
2.2	The self-energy and its analytical properties	8
2.3	Physical significance of the self-energy	10
3	The Luttinger-Ward functional	12
3.1	The Green function as a functional integral	12
3.2	Construction of the Luttinger-Ward functional	17
4	Self-energy functional theories	20
4.1	Dynamical mean-field theory	20
4.2	Metal-insulator-transition in a dimer	22
5	Summary and conclusion	26
	Appendices	27

1 Introduction

Green functions are quantities of central importance in the theory of interacting many-particle systems. In the simplest terms they may be viewed as describing the propagation of some kind of ‘perturbation’ through a system of interacting particles. On the one hand, by their very definition Green functions contain information about the system which can be directly compared to experiments, on the other hand they are a useful mathematical tool because there is a wide variety of techniques to calculate them approximately, and thus overcome the, in general, insoluble nature of the many-particle problem. Of particular importance is the single-particle Green function $G(\mathbf{k}, \omega)$, which describes the propagation of an added particle with wave number $\mathbf{k}$ and energy $\hbar\omega$ through the system and which is directly related to the experimentally measurable photoemission and inverse photoemission spectrum. The famous Dyson equation then expresses the single-particle Green function in terms of another central quantity in many-body physics, the self-energy $\Sigma(\mathbf{k}, \omega)$. This may be viewed as a $\mathbf{k}$ - and ω -dependent potential which mimics the effect of the other particles on the propagation of the added particle (it is therefore hardly a surprise that to simplest approximation the self-energy equals the Hartree-Fock potential). However, the physical significance of the self-energy is considerably wider. Namely, in 1961 Luttinger and Ward (LW) published a seminal paper [1] which became the foundation of many important developments in the quantum theory of many-particle systems. They showed that the grand canonical potential Ω of an interacting Fermion system can be expressed as a functional of the single-particle Green function $G(\mathbf{k}, \omega)$ and the self-energy $\Sigma(\mathbf{k}, \omega)$. A key step thereby was the construction of the Luttinger-Ward functional $\Phi[G]$, a functional of the Green function which essentially describes the deviation of Ω from a non-interacting system with the same band structure as the interacting one. Luttinger and Ward gave an explicit expression for $\Phi[G]$ as a sum over infinitely many Feynman diagrams and also showed that Σ is the functional derivative of $\Phi[G]$ with respect to G and that Ω is stationary under variations of Σ . The expression for Ω and the properties of $\Phi[G]$ became the basis for many important results, such as the Luttinger theorem [2] which states that interactions between electrons do not change the volume of the Fermi surface, or the construction of conserving approximations by Baym and Kadanoff [3, 4], where $\Phi[G]$ is approximated by keeping only a subclass of Feynman diagrams. While the representation of Green functions in terms of Feynman diagrams was highly successful for systems such as the electron gas, a new challenge arose with increasing interest in strongly correlated electron systems such as the Hubbard model. Namely any Feynman diagram expansion implicitly treats the interaction between the particles as a perturbation and assumes a continuous evolution from the noninteracting system, which is questionable for systems such as Mott-insulators. It is the purpose of the present lecture to sketch a further development of the ideas of Luttinger and Ward which is mainly due to Potthoff, the so-called self-energy functional theory. We present Potthoff’s non-perturbative re-derivation of the results of Luttinger and Ward [5, 6] and the application of their theorems to strongly correlated electron systems by combining them with numerical methods. As an application, we discuss the metal-insulator transition in the single-band Hubbard model.

2 Green functions and the self-energy

2.1 The Green function and its analytical properties

We start with a discussion of Green functions and their properties and use. We will not cover the representation of Green functions in terms of Feynman diagrams, since excellent introductions to this subject can be found in various textbooks [7–9]. In the present notes we try to be consistent with Fetter/Walecka (FW) [8].

We consider a system of interacting fermions and assume that there is some complete basis of single electron states $\varphi_\alpha(x)$. Here $x = (\mathbf{r}, \sigma)$ is the combined real-space and spin coordinate, whereas α is a shorthand for a set of quantum numbers. For example, in an LCAO-type description of a solid, which is what we mostly have in mind, we would have $\alpha = (i, n, \nu, \sigma)$ where $i \in \{1, \dots, N\}$ denotes the unit cell, $n \in \{1, \dots, n_{Atom}\}$ the number of the atoms in the basis, $\nu \in \{s, p_x, p_y, p_z, d_{xy}, \dots\}$ the type of orbital on the respective atom, and σ the z -component of spin. We denote the number of orbitals in a unit cell by n_{orb} , the total number of α 's thus is $N_\alpha = 2Nn_{orb}$. Upon Fourier transformation one would replace $i \rightarrow \mathbf{k}$, the wave vector.

In all that follows we consider a grand canonical ensemble with inverse temperature $\beta = 1/k_B T$ and chemical potential μ . Introducing fermionic creation/annihilation operators $c_\alpha^\dagger/c_\alpha$ for electrons in the states $\varphi_\alpha(x)$, the grand canonical Hamiltonian $K = H - \mu \hat{N}$ (with $\hat{N}$ the operator for the number of electrons) can be written as $K = K_0 + K_1$ with [7–9]

$$K_0 = \sum_{\alpha, \beta} (t_{\alpha, \beta} - \mu \delta_{\alpha, \beta}) c_\alpha^\dagger c_\beta \quad \text{and} \quad K_1 = \frac{1}{2} \sum_{\alpha, \beta, \gamma, \delta} V_{\alpha, \beta, \delta, \gamma} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta. \quad (1)$$

The matrix elements in this Hamiltonian are given by

$$t_{\alpha, \beta} = \int dx \varphi_\alpha^*(x) \left(-\frac{\hbar^2}{2m} \nabla^2 + V_0(\mathbf{r}) \right) \varphi_\beta(x),$$

$$V_{\alpha, \beta, \delta, \gamma} = \int dx \int dx' \varphi_\alpha^*(x) \varphi_\beta^*(x') V_1(x-x') \varphi_\gamma(x') \varphi_\delta(x).$$

with the external (interparticle) potential $V_0(\mathbf{r})$ ($V_1(x-x')$). For later reference we calculate

$$[c_\alpha, K] = \sum_\beta (t_{\alpha, \beta} - \mu \delta_{\alpha, \beta}) c_\beta + \sum_{\beta, \gamma, \delta} V_{\alpha, \beta, \delta, \gamma} c_\beta^\dagger c_\gamma c_\delta, \quad (2)$$

where we have used the identity $V_{\alpha, \beta, \gamma, \delta} = V_{\beta, \alpha, \delta, \gamma}$ obtained by exchanging the integration variables $x \leftrightarrow x'$ in the definition of $V_{\alpha\beta\gamma\delta}$. It obviously holds that $[H, \hat{N}] = 0$ so that eigenstates of H have a fixed particle number. The thermal average of any operator $\hat{O}$ is

$$\langle \hat{O} \rangle_{th} = \frac{1}{Z} \text{Tr} \left(e^{-\beta K} \hat{O} \right) = \frac{1}{Z} \sum_i e^{-\beta K_i} \langle i | \hat{O} | i \rangle. \quad (3)$$

Here $|i\rangle$ are the eigenstates of H , $H|i\rangle = E_i|i\rangle$, and $K_i = E_i - \mu N_i$ the corresponding eigenvalues of K . Also,

$$Z = \text{Tr} \left(e^{-\beta K} \right) = \sum_i e^{-\beta K_i} \quad (4)$$

is the grand partition function, which is a quantity of interest by itself, because the grand canonical potential of the system is given by $\Omega = -k_B T \log(Z)$.

We proceed to the definition of *Green functions* which basically describe the following *gedanken experiment*: the system is initially in thermal equilibrium, but at a certain time t we do something to the system by acting with an operator $\hat{A}$. $\hat{A}$ could add or remove an electron, flip a spin or similar. We then let the system evolve for some time and undo the change at a time t' by acting with an operator $\hat{B}$, which often is $\hat{A}^\dagger$. We then form the overlap with the state we would have obtained had the system evolved without perturbation. In other words, the Green function describes how the ‘perturbation’ created by $\hat{A}$ propagates in the interval $t \rightarrow t'$ before it is removed by $\hat{B}$.

We now formalize this idea, but for reasons which will become clear only later we make a digression and first introduce *imaginary time Green functions*. Moreover, we specialize to the *single particle Green function*, where the perturbation is removing or adding a particle, i.e., $\hat{A} = c_\alpha$ and $\hat{B} = c_\beta^\dagger$, or vice versa. For any operator $\hat{O}$ the imaginary time Heisenberg operator is $\hat{O}(\tau) = e^{\tau K/\hbar} \hat{O} e^{-\tau K/\hbar}$ and the imaginary time Green function is

$$\begin{aligned} G_{\alpha,\beta}(\tau, \tau') &= -\langle T c_\alpha(\tau) c_\beta^\dagger(\tau') \rangle_{th} \\ &= -\Theta(\tau - \tau') \langle c_\alpha(\tau) c_\beta^\dagger(\tau') \rangle_{th} + \Theta(\tau' - \tau) \langle c_\beta^\dagger(\tau') c_\alpha(\tau) \rangle_{th} \\ &= \frac{1}{Z} \left(-\Theta(\tau - \tau') \sum_{i,j} e^{-\beta K_i} e^{\frac{\tau - \tau'}{\hbar} (K_i - K_j)} \langle i | c_\alpha | j \rangle \langle j | c_\beta^\dagger | i \rangle \right. \\ &\quad \left. + \Theta(\tau' - \tau) \sum_{i,j} e^{-\beta K_i} e^{\frac{\tau' - \tau}{\hbar} (K_j - K_i)} \langle i | c_\beta^\dagger | j \rangle \langle j | c_\alpha | i \rangle \right). \end{aligned} \quad (5)$$

T in the first line is the *time ordering operator*, which reorders the following Heisenberg operators such that their times decrease from left to right and multiplies by (-1) for each exchange of two Fermion operators, as can be seen in the second line. In going from the second to the third line we have used (3), inserted a resolution of unity $\sum_j |j\rangle \langle j| = 1$, and used $\langle i | \hat{O}(\tau) | j \rangle = e^{\frac{\tau}{\hbar} (K_i - K_j)} \langle i | \hat{O} | j \rangle$. Eq. (5) shows that $G_{\alpha,\beta}$ really is a function of $\tau - \tau'$ only: $G_{\alpha,\beta}(\tau, \tau') = G_{\alpha,\beta}(\tau - \tau')$. To simplify the notation we henceforth replace $\tau - \tau' \rightarrow \tau$, which is equivalent to choosing $\tau' = 0$. The τ -dependence of both terms in (5) is $e^{(-\beta + \frac{|\tau|}{\hbar}) K_i} e^{-\frac{|\tau|}{\hbar} K_j}$. Since the K_i are bounded from below, namely by K_0 for the ground state with a given μ , but unbounded from above in the thermodynamical limit, G is therefore well-defined only for $\tau \in [-\beta\hbar, \beta\hbar]$ [10]. It therefore can be written as a Fourier series with frequencies $n\pi/\hbar\beta$. However, it is easy to see from (5) that for $\tau \in [-\beta\hbar, 0]$ one has $G(\tau + \beta\hbar) = -G(\tau)$, so that only odd n can contribute. All in all we find the Fourier expansion

$$G(\tau) = \frac{1}{\beta\hbar} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu \tau} G(i\omega_\nu), \quad (6)$$

$$G(i\omega_\nu) = \int_0^{\beta\hbar} d\tau e^{i\omega_\nu \tau} G(\tau), \quad (7)$$

with $\omega_\nu = (2\nu+1)\pi/\beta\hbar$ and integer ν . The ω_ν are the Fermionic *Matsubara frequencies*.

Performing the integral (7) with (5) for $G_{\alpha,\beta}(\tau)$, thereby using $e^{i\omega_\nu\beta\hbar} = -1$ for any ν , we obtain

$$G_{\alpha,\beta}(i\omega_\nu) = \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta K_i} + e^{-\beta K_j}}{i\omega_\nu - \frac{1}{\hbar}(K_j - K_i)} \langle i|c_\alpha|j\rangle \langle j|c_\beta^\dagger|i\rangle. \quad (8)$$

This so-called *Lehmann representation* of the Green function is exact but not really helpful for practical computations because it requires knowledge of all eigenstates and energies of H . It can be used, however, for establishing certain properties of the Green function as will be seen in the following.

We proceed to the definition of the real-time Green function. Defining the real-time Heisenberg operator $\hat{O}(\tau) = e^{itK/\hbar} \hat{O} e^{-itK/\hbar}$ the *retarded real-time Green function* is

$$\begin{aligned} G_{\alpha,\beta}^R(t, t') &= -i\Theta(t-t') \left(\langle c_\alpha(t) c_\beta^\dagger(t') \rangle_{th} + \langle c_\beta^\dagger(t') c_\alpha(t) \rangle_{th} \right), \\ &= -i\Theta(t-t') \frac{1}{Z} \left(\sum_{i,j} e^{-\beta K_i} e^{i\frac{t-t'}{\hbar}(K_i - K_j)} \langle i|c_\alpha|j\rangle \langle j|c_\beta^\dagger|i\rangle \right. \\ &\quad \left. + \sum_{i,j} e^{-\beta K_i} e^{i\frac{t-t'}{\hbar}(K_j - K_i)} \langle i|c_\beta^\dagger|j\rangle \langle j|c_\alpha|i\rangle \right). \end{aligned} \quad (9)$$

Again this is a function of $t-t'$ only and, using the formula derived in Appendix A

$$-i\Theta(t) e^{-iEt} = \lim_{\eta \rightarrow 0^+} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega t}}{\omega - E + i\eta}$$

we find its Fourier transform

$$G_{\alpha,\beta}^R(\omega) = \frac{1}{Z} \lim_{\eta \rightarrow 0^+} \sum_{i,j} \frac{e^{-\beta K_i} + e^{-\beta K_j}}{\omega + i\eta - \frac{1}{\hbar}(K_j - K_i)} \langle i|c_\alpha|j\rangle \langle j|c_\beta^\dagger|i\rangle. \quad (10)$$

Comparison with Eq. (8) shows that $G_{\alpha,\beta}^R(\omega)$ can be obtained from $G_{\alpha,\beta}(i\omega_\nu)$ by replacing $i\omega_\nu \rightarrow \omega + i\eta$. In other words, there is one function $G_{\alpha,\beta}(z)$ of the complex variable z , often called *the Green function*, which gives $G_{\alpha,\beta}(i\omega_\nu)$ when evaluated for the Matsubara frequencies, and $G_{\alpha,\beta}^R(\omega)$ when evaluated on a line infinitesimally above the real axis. The existence of such a function is the very reason why the imaginary time Green function is introduced in the first place. In principle, the quantities of interest, of course, are the real-time Green functions. For example, the single-particle Green function (10) contains information about the photoemission and inverse photoemission spectrum of the system. On the other hand, the imaginary-time Green function (8) can be evaluated approximately by using the powerful technique of expansion in Feynman diagrams [7–9], which is *not* possible for the real-time Green function. The standard way to obtain the real-time Green functions, which is used again and again in the literature, is to first obtain an approximate $G_{\alpha,\beta}(i\omega_\nu)$ by doing an expansion in Feynman diagrams and then obtain the real-time Green function by continuing it analytically to a line infinitesimally above the real axis. If one has an analytical expression for $G_{\alpha,\beta}(i\omega_\nu)$ this can be done by simply replacing $i\omega_\nu \rightarrow \omega + i\eta$.

We now discuss some properties of *the* Green function. In Eq. (10) the product of matrix elements $\langle i|c_\alpha|j\rangle\langle j|c_\beta^\dagger|i\rangle$ can differ from zero only if $N_j = N_i+1$, whence $K_j-K_i = E_j-E_i-\mu$. $G_{\alpha,\beta}(z)$ therefore has poles on the real axis, which correspond to differences of energies of states whose electron number differ by one. Next, we rewrite the Green function as

$$\mathbf{G}(z) = \int_{-\infty}^{\infty} d\omega \frac{\boldsymbol{\rho}(\omega)}{z-\omega},$$

where the elements of the *spectral density matrix* $\boldsymbol{\rho}(\omega)$ can be read off from (8)

$$\rho_{\alpha,\beta}(\omega) = \frac{1}{Z} \sum_{i,j} (e^{-\beta K_i} + e^{-\beta K_j}) \langle i|c_\alpha|j\rangle\langle j|c_\beta^\dagger|i\rangle \delta\left(\omega - \frac{K_i-K_j}{\hbar}\right). \quad (11)$$

Since ω is real we find $\rho_{\alpha,\beta}^*(\omega) = \rho_{\beta,\alpha}(\omega)$, i.e., $\boldsymbol{\rho}(\omega)$ is Hermitian for any ω . Next consider any vector $\mathbf{v}$ of length N_α and define the linear combination $c_\mathbf{v}^\dagger = \sum v_\alpha c_\alpha^\dagger$. Then

$$\sum_{\alpha,\beta} v_\alpha^* \rho_{\alpha,\beta}(\omega) v_\beta = \frac{1}{Z} \sum_{i,j} (e^{-\beta K_i} + e^{-\beta K_j}) |\langle j|c_\mathbf{v}^\dagger|i\rangle|^2 \delta\left(\omega - \frac{K_i-K_j}{\hbar}\right) > 0$$

i.e., $\boldsymbol{\rho}(\omega)$ is positive definite for each ω . It then follows from (11) that $(\mathbf{G}(z))^\dagger = \mathbf{G}(z^*)$ so that $\mathbf{G}(z)$ is *not* Hermitian for complex z . Next consider

$$f(z) = \sum_{\alpha,\beta} v_\alpha^* G_{\alpha,\beta}(z) v_\beta = \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta K_i} + e^{-\beta K_j}}{z - \frac{K_i-K_j}{\hbar}} |\langle j|c_\mathbf{v}^\dagger|i\rangle|^2. \quad (12)$$

We write $z = x+iy$ and find the imaginary part of this to be

$$\text{Im } f(z) = -\frac{y}{Z} \sum_{i,j} \frac{e^{-\beta K_i} + e^{-\beta K_j}}{\left(x - \frac{K_i-K_j}{\hbar}\right)^2 + y^2} |\langle j|c_\mathbf{v}^\dagger|i\rangle|^2. \quad (13)$$

This expression cannot vanish unless $y = 0$, so that for z away from the real axis all eigenvalues of $\mathbf{G}(z)$ must have a nonvanishing imaginary part, otherwise we could choose $\mathbf{v}$ to be the normalized right-hand eigenvector belonging to a purely real eigenvalue λ and find that $f(z)=\lambda$, in contradiction to (13). This implies in particular that for z away from the real axis all eigenvalues of $\mathbf{G}(z)$ are different from zero so that the determinant of $\mathbf{G}(z)$ is different from zero [11] and its inverse $\mathbf{G}^{-1}(z)$ does exist. Using Cramer's rule we find the elements of the inverse Green function

$$G_{\alpha,\beta}^{-1}(z) = \frac{(-1)^{\alpha+\beta} \det \mathbf{M}_{\alpha,\beta}(z)}{\det \mathbf{G}(z)}$$

where $\mathbf{M}_{\alpha,\beta}(z)$ is the respective minor of $\mathbf{G}(z)$, i.e., the matrix $\mathbf{G}(z)$ with line α and column β discarded. Since for z away from the real axis all elements of $\mathbf{G}(z)$ are finite (see Eq. (8)) and the determinant of $\mathbf{G}(z)$ is different from zero it moreover follows that away from the real axis all elements of $\mathbf{G}^{-1}(z)$ are analytical functions of z .

We proceed to a discussion of the behavior of $\mathbf{G}(z)$ for large $|z|$. We assume that the range of ω where the elements of $\boldsymbol{\rho}(\omega)$ are different from zero is finite, which simply means that the change

in energy upon adding or removing an electron is bounded. To discuss the limit $|z| \rightarrow \infty$ we rewrite $\mathbf{G}(z)$ as

$$G_{\alpha,\beta}(z) = \frac{1}{Z} \sum_{i,j} e^{-\beta K_i} \left(\frac{\langle i|c_\alpha|j\rangle\langle j|c_\beta^\dagger|i\rangle}{z - \frac{K_j - K_i}{\hbar}} + \frac{\langle i|c_\beta^\dagger|j\rangle\langle j|c_\alpha|i\rangle}{z + \frac{K_j - K_i}{\hbar}} \right), \quad (14)$$

and expand

$$\frac{1}{z \pm \frac{K_j - K_i}{\hbar}} \rightarrow \frac{1}{z} \mp \frac{K_j - K_i}{\hbar z^2} + \mathcal{O}\left(\frac{1}{z^3}\right).$$

Next we use $(K_j - K_i)\langle j|c_\alpha|i\rangle = \langle j|Kc_\alpha|i\rangle - \langle j|c_\alpha K|i\rangle = \langle j|[K, c_\alpha]|i\rangle$ and find

$$G_{\alpha,\beta}(z) \rightarrow \frac{\delta_{\alpha,\beta}}{z} + \frac{\langle \{c_\beta^\dagger, [c_\alpha, K]\} \rangle_{th}}{\hbar z^2} + \mathcal{O}\left(\frac{1}{z^3}\right).$$

Using (2) we obtain

$$\langle \{c_\beta^\dagger, [c_\alpha, K]\} \rangle_{th} = t_{\alpha,\beta} - \mu \delta_{\alpha,\beta} + \sum_{\gamma,\delta} (V_{\alpha,\gamma,\beta,\delta} - V_{\alpha,\gamma,\delta,\beta}) \langle c_\gamma^\dagger c_\delta \rangle_{th}. \quad (15)$$

The term involving V looks like the Hartree-Fock potential $V_{\alpha,\beta}^{(HF)}$, however, whereas for the true Hartree-Fock potential the thermal average has to be taken using the Hartree-Fock wave functions and energies, the thermal average in (15) has to be taken using the fully interacting eigenstates and energies. Keeping this subtle difference in mind we still call the third term the Hartree-Fock potential $V_{\alpha,\beta}^{(HF)}$ so that

$$\mathbf{G}(z) \rightarrow \frac{1}{z} + \frac{\mathbf{t} - \mu + \mathbf{V}^{(HF)}}{\hbar z^2} + \mathcal{O}\left(\frac{1}{z^3}\right). \quad (16)$$

Eq. (14) also highlights the physical content of the Green function. We consider a single band of *non*-interacting electrons where $\alpha = (\mathbf{k}, \sigma)$, and $K = \sum_{\mathbf{k}} (\varepsilon_{\mathbf{k}} - \mu) c_{\mathbf{k},\sigma}^\dagger c_{\mathbf{k},\sigma}$. All eigenstates can then be characterized by the occupation numbers of the states $(\mathbf{k}, \sigma)$ being either 0 or 1. In

$$G_{\mathbf{k},\mathbf{k}}(z) = \frac{1}{Z} \sum_{i,j} e^{-\beta K_i} \left(\frac{\langle i|c_{\mathbf{k},\sigma}^\dagger|j\rangle\langle j|c_{\mathbf{k},\sigma}|i\rangle}{z + \frac{K_j - K_i}{\hbar}} + \frac{\langle i|c_{\mathbf{k},\sigma}|j\rangle\langle j|c_{\mathbf{k},\sigma}^\dagger|i\rangle}{z - \frac{K_j - K_i}{\hbar}} \right)$$

the matrix element $\langle j|c_{\mathbf{k},\sigma}|i\rangle$ in the first term is then different from zero only if $|i\rangle$ has an electron in the single-electron state $(\mathbf{k}, \sigma)$, whereas $|j\rangle$ has none, while all other momenta have identical occupation. It follows that $E_i = E_j + \varepsilon_{\mathbf{k}}$, whence $K_j - K_i = -\varepsilon_{\mathbf{k}} + \mu$. The first term therefore describes the removal of an electron with momentum $\mathbf{k}$ and spin σ . By analogous reasoning, in the second term $K_j - K_i = \varepsilon_{\mathbf{k}} - \mu$ and this term describes addition of an electron. In both terms $K_j - K_i$ is independent of both, $|i\rangle$ and $|j\rangle$, so that we can use $\sum_j |j\rangle\langle j| = 1$ and find

$$G_{\mathbf{k},\mathbf{k}}(z) = \frac{1}{Z} \sum_i e^{-\beta K_i} \left(\frac{\langle i|c_{\mathbf{k},\sigma}^\dagger c_{\mathbf{k},\sigma}|i\rangle}{z - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar}} + \frac{\langle i|c_{\mathbf{k},\sigma} c_{\mathbf{k},\sigma}^\dagger|i\rangle}{z - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar}} \right) = \frac{\langle n_{\mathbf{k},\sigma} \rangle_{th}}{z - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar}} + \frac{\langle 1 - n_{\mathbf{k},\sigma} \rangle_{th}}{z - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar}}.$$

Both terms have poles which trace the electrons' dispersion relation $\varepsilon_{\mathbf{k}} - \mu$, but the first term describes only the occupied part of the band, the second term the unoccupied part. The Green function is the combined photoemission and inverse photoemission spectrum and this holds true also for the interacting case.

2.2 The self-energy and its analytical properties

We proceed to the definition of another key quantity, the *self-energy*, and start by discussing the *equation of motion* of the imaginary time Green function Eq. (5). We set $G_{\alpha,\beta}(\tau, 0) \rightarrow G_{\alpha,\beta}(\tau)$. It follows readily from the definition of the imaginary-time Heisenberg operator $\hat{O}(\tau)$ that $-\hbar\partial_\tau\hat{O}(\tau) = [\hat{O}(\tau), K]$. Moreover, using, e.g., the representation $\Theta(\tau) = \int_{-\infty}^{\tau} dx \delta(x)$ one finds $\partial_\tau\Theta(\pm\tau) = \pm\delta(\tau)$. Combining everything we get the equation of motion of the Green function

$$\begin{aligned} -\hbar\partial_\tau G_{\alpha,\beta}(\tau) &= \hbar\delta(\tau) \left\langle \{c_\alpha, c_\beta^\dagger\} \right\rangle_{th} + \left\langle T [c_\alpha(\tau), K] c_\beta^\dagger(0) \right\rangle_{th} \\ &= \hbar\delta(\tau) \delta_{\alpha,\beta} + \sum_\nu (t_{\alpha,\nu} - \mu\delta_{\alpha,\nu}) G_{\nu,\beta}(\tau) + F_{\alpha,\beta}(\tau), \end{aligned} \quad (17)$$

$$F_{\alpha,\beta}(\tau) = - \sum_{\nu,\kappa,\lambda} V_{\alpha,\nu,\kappa,\lambda} \left\langle T [(c_\nu^\dagger c_\lambda c_\kappa)(\tau) c_\beta^\dagger(0)] \right\rangle_{th}, \quad (18)$$

where (2) was used. In the *noninteracting case*, i.e., all $V_{\alpha,\nu,\kappa,\lambda}=0$, we have $F_{\alpha,\beta}(\tau)=0$ and the equation closes. Using the representation of the δ -function in terms of Matsubara-frequencies [8]

$$\delta(\tau) = \frac{1}{\hbar\beta} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu\tau}$$

straightforward Fourier transform gives the noninteracting Green function

$$\left(i\omega_\nu - \frac{t-\mu}{\hbar} \right) \mathbf{G}_0(i\omega_\nu) = 1. \quad (19)$$

We return to the case $V \neq 0$. We recall that $G(\tau)$ fulfills $G(\tau+\beta\hbar) = -G(\tau)$ for $\tau \in [-\beta\hbar, 0]$, which established its Fourier expansion (6). It follows that $\partial_\tau G(\tau+\beta\hbar) = -\partial_\tau G(\tau)$ whence $F(\tau)$ must obey the same boundary condition: $F(\tau+\beta\hbar) = -F(\tau)$. Accordingly, $F(\tau)$ has the same Fourier expansion (6) as $G(\tau)$ itself and we find

$$\left(i\omega_\nu - \frac{t-\mu}{\hbar} \right) \mathbf{G}(i\omega_\nu) - \frac{1}{\hbar} \mathbf{F}(i\omega_\nu) = 1.$$

Now we define the *self-energy* $\Sigma(i\omega_\nu)$ by $\mathbf{F}(i\omega_\nu) = \hbar\Sigma(i\omega_\nu) \mathbf{G}(i\omega_\nu)$ (recall that $\mathbf{G}^{-1}(z)$ exists for all z away from the real axis so that $\Sigma(i\omega_\nu)$ is well defined) whence

$$\left(i\omega_\nu - \frac{t-\mu}{\hbar} - \Sigma(i\omega_\nu) \right) \mathbf{G}(i\omega_\nu) = 1. \quad (20)$$

By analytical continuation, $i\omega_\nu \rightarrow z$, this equation defines $\Sigma(z)$ for all z away from the real axis. Next, notice that the brackets on the left-hand side of (19) and (20) are $\mathbf{G}_0^{-1}(z)$ and $\mathbf{G}^{-1}(z)$, respectively, so that we immediately read off the famous *Dyson equation*

$$\mathbf{G}^{-1}(z) = \mathbf{G}_0^{-1}(z) - \Sigma(z) = z - \frac{t-\mu}{\hbar} - \Sigma(z). \quad (21)$$

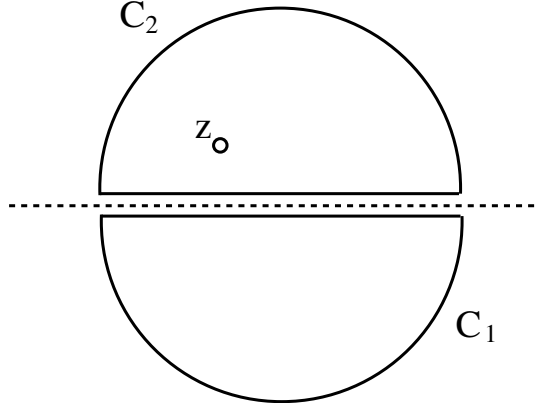


Fig. 1: Integration contours for the proof of the spectral representation of Σ . The dashed line is the real z' axis.

We discuss some properties of $\Sigma(z)$ which follow from its definition. Since both $\mathbf{G}^{-1}(z)$ and $\mathbf{G}_0^{-1}(z)$ are analytical in the complex z -plane except for the real axis, the same must hold true for $\Sigma(z)$. Next, it follows from (16) that for $|z| \rightarrow \infty$

$$\mathbf{G}^{-1}(z) \rightarrow z - \frac{\mathbf{t} - \mu}{\hbar} - \frac{\mathbf{V}^{(HF)}}{\hbar} + \mathcal{O}\left(\frac{1}{z}\right) \Rightarrow \Sigma(z) \rightarrow \frac{\mathbf{V}^{(HF)}}{\hbar} + \mathcal{O}\left(\frac{1}{z}\right).$$

The shifted function $\bar{\Sigma}(z) = \Sigma(z) - \frac{\mathbf{V}^{(HF)}}{\hbar}$ therefore vanishes as $1/z$ for large $|z|$. Now consider the equation $\mathbf{G}^{-1}(z) \mathbf{G}(z) = 1$, take the Hermitian conjugate and use $[\mathbf{G}(z)]^\dagger = \mathbf{G}(z^*)$. It follows that $[\mathbf{G}^{-1}(z)]^\dagger = \mathbf{G}^{-1}(z^*)$. Since trivially $[\mathbf{G}_0^{-1}(z)]^\dagger = \mathbf{G}_0^{-1}(z^*)$ and $\mathbf{V}^{HF}$ is Hermitian it follows that $\bar{\Sigma}(z^*) = \bar{\Sigma}^\dagger(z)$. Next, for real ω , define the real matrices $\mathbf{K}(\omega)$ and $\mathbf{J}(\omega)$ by

$$\bar{\Sigma}(\omega + i0^+) = \mathbf{K}(\omega) + i\mathbf{J}(\omega) \quad \Rightarrow \quad \bar{\Sigma}(\omega - i0^+) = \mathbf{K}^T(\omega) - i\mathbf{J}^T(\omega). \quad (22)$$

We introduce the symmetrized/antisymmetrized linear combinations $\mathbf{K}^{(\pm)} = (\mathbf{K} \pm \mathbf{K}^T)/2$ and $\mathbf{J}^{(\pm)} = (\mathbf{J} \pm \mathbf{J}^T)/2$ whence

$$\bar{\Sigma}^{(+)}(z) = \frac{1}{2} \left(\bar{\Sigma}(z) + \bar{\Sigma}^T(z) \right), \quad \bar{\Sigma}^{(-)}(z) = \frac{i}{2} \left(\bar{\Sigma}(z) - \bar{\Sigma}^T(z) \right). \quad (23)$$

The latter can be expressed in terms of $\mathbf{K}^{(\pm)}(z)$ and $\mathbf{J}^{(\pm)}(z)$ as

$$\begin{aligned} \bar{\Sigma}^{(+)}(\omega \pm i0^+) &= \mathbf{K}^{(+)}(\omega) \pm i\mathbf{J}^{(+)}(\omega), \\ \bar{\Sigma}^{(-)}(\omega \pm i0^+) &= -\mathbf{J}^{(-)}(\omega) \pm i\mathbf{K}^{(-)}(\omega). \end{aligned} \quad (24)$$

Now consider the integration contours in Figure 1 which consist of lines infinitesimally above and below the real axis and semicircles at infinity. Since $\bar{\Sigma}(z)$ is analytic away from the real axis we have for any z in the upper half-plane

$$\oint_{C_1} dz' \frac{\bar{\Sigma}^{(\pm)}(z')}{z' - z} = 0 \quad \Rightarrow \quad \int_{-\infty}^{\infty} d\omega \frac{\bar{\Sigma}^{(\pm)}(\omega - i0^+)}{\omega - z} = 0. \quad (25)$$

The second equation follows because the integrand is $\propto 1/z'^2$ for large $|z'|$ so that the contribution from the arc vanishes and only the line infinitesimally below the real axis contributes,

where $z = \omega - i0^+$, and $\omega - i0^+ - z = \omega - z$ whenever z has a finite imaginary part. We can now use (24) and obtain

$$\int_{-\infty}^{\infty} d\omega \frac{\mathbf{K}^{(\pm)}(\omega)}{\omega - z} = i \int_{-\infty}^{\infty} d\omega \frac{\mathbf{J}^{(\pm)}(\omega)}{\omega - z}. \quad (26)$$

Next we apply Cauchy's theorem to the contour C_2 in Figure 1 and find

$$\bar{\Sigma}^{(\pm)}(z) = \frac{1}{2\pi i} \oint_{C_2} dz' \frac{\bar{\Sigma}^{(\pm)}(z')}{z' - z}.$$

Again the contribution from the arc vanishes. We use (26) to eliminate the integrals over $\mathbf{K}^{(+)}$ and $\mathbf{J}^{(-)}$, revert (23) to obtain $\bar{\Sigma}(z)$ and arrive at the spectral representation of the self-energy, as derived by Luttinger [12]

$$\Sigma(z) = \frac{\mathbf{V}^{(HF)}}{\hbar} + \int_{-\infty}^{\infty} d\omega \frac{\boldsymbol{\sigma}(\omega)}{\omega - z} \quad (27)$$

with the Hermitian matrix $\boldsymbol{\sigma}(\omega) = \mathbf{J}^{(+)}(\omega) - i\mathbf{K}^{(-)}(\omega)$.

Lastly, we notice a property of the self-energy that will be of some importance later on: we will often be concerned with systems which contain both, ‘correlated orbitals’ and ‘uncorrelated orbitals’. A given orbital α is ‘uncorrelated’ if all interaction matrix elements $V_{\kappa\lambda\mu\nu} = 0$ when at least one of the four indices equals α . In other words, electrons in an uncorrelated orbital do not interact with the other electrons. An example is the well-known Anderson model where one has correlated ‘ f -orbitals’ hybridizing with uncorrelated ‘conduction electrons’. It then follows from the definition (18) that $F_{\alpha,\beta}(z) = 0$ whenever α is uncorrelated, and since $\hbar\Sigma = FG^{-1}$ the same holds true for $\Sigma_{\alpha,\beta}(z)$. Since we have seen above that $\Sigma_{\beta,\alpha}^*(z) = \Sigma_{\alpha,\beta}(z^*)$, we find $\Sigma_{\alpha,\beta}(z) = 0$ also if β is uncorrelated: *matrix elements of the self-energy involving uncorrelated orbitals are zero!*

2.3 Physical significance of the self-energy

We briefly recall the key results of our discussion so far: The Green function and self-energy are related by the Dyson equation (21), they are analytical functions in the complex frequency plane except for the real axis, and they have the spectral representations (11) and (27). The poles of the Green function on the real axis give the ionization and affinity energies of the system, i.e., the energies it takes to remove or add an electron. We now want to discuss the physical consequences of this formal structure. For the remainder of this subsection we consider the case of a single band, set $\hbar = 1$ and call $t_{\mathbf{k}} - V_{\mathbf{k}}^{(HF)} - \mu = \varepsilon_{\mathbf{k}}$. Moreover, we drop the dependence on $\mathbf{k}$ and replace $\varepsilon_{\mathbf{k}} \rightarrow \varepsilon$, $G(\mathbf{k}, \omega) \rightarrow G(\omega)$ and so on. To begin with, we consider a system with finite N , where the spacing between eigenvalues K_i and hence between the poles of $G(\omega)$ is finite. We have

$$G(\omega) = \frac{1}{\omega - \varepsilon - \bar{\Sigma}(\omega)} = \sum_{i=1}^{n+1} \frac{Z_i}{\omega - \omega_i} \quad (28)$$

$$\bar{\Sigma}(\omega) = \sum_{i=1}^n \frac{\sigma_i}{\omega - \zeta_i} \quad (29)$$

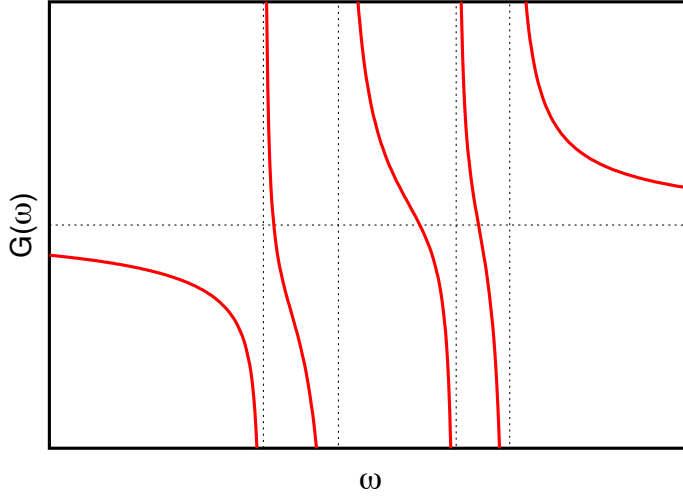


Fig. 2: The Green function $G(\omega)$ for real ω . The dashed vertical lines indicate the poles, ω_i .

where $Z_i > 0$ and ω_i, ζ_i real. On the real axis, $G(\omega)$ crosses zero precisely once in each interval $[\omega_i, \omega_{i+1}]$ with negative slope, see Figure 2. It follows from (28) that $G(\zeta) = 0$ implies $\Sigma(\zeta) \rightarrow \infty$ so that the ζ_i in (29) must be the zeroes of $G(\omega)$. Moreover, the ζ_i are ‘sandwiched’ in between the poles ω_i and their number is indeed one less than the ω_i , see Figure 2. Since $G^{-1}(\omega_i) = 0$ we find for ω close to ω_i

$$G^{-1}(\omega) = \left(1 - \frac{\partial \Sigma}{\partial \omega} \Big|_{\omega_i}\right) (\omega - \omega_i) \Rightarrow Z_i = \left(1 - \frac{\partial \Sigma}{\partial \omega} \Big|_{\omega_i}\right)^{-1}. \quad (30)$$

The residue Z_i , which equals the weight observed in photoemission/inverse photoemission, therefore is related to the slope of the self-energy, which is a well-known result.

We consider a situation where the self-energy has a single pole, $\sigma(\omega) = \sigma_0 \delta(\omega - \zeta)$, and reinsert the $\mathbf{k}$ -dependence

$$G^{-1}(\mathbf{k}, z) = z - \varepsilon_{\mathbf{k}} - \frac{\sigma_0}{z - \zeta} \Rightarrow G(\mathbf{k}, z) = \frac{Z_+}{z - \omega_{+, \mathbf{k}}} + \frac{Z_-}{z - \omega_{-, \mathbf{k}}} \quad (31)$$

where we have introduced

$$\omega_{\pm, \mathbf{k}} = \frac{1}{2} \left((\varepsilon_{\mathbf{k}} + \zeta) \pm \sqrt{(\varepsilon_{\mathbf{k}} - \zeta)^2 + 4\sigma_0} \right), \quad Z_{\pm} = \pm \frac{\omega_{\pm} - \zeta}{\omega_{+} - \omega_{-}}.$$

It is easy to see that $Z_{\pm} > 0$ and $Z_+ + Z_- = 1$ as it has to be. We can see from this that the single pole of the noninteracting Green function at $\omega = \varepsilon_{\mathbf{k}}$ is split into two poles at $\omega_{\pm, \mathbf{k}}$, i.e., the single band with dispersion $\varepsilon_{\mathbf{k}}$ becomes two bands with dispersions $\omega_{\pm, \mathbf{k}}$. Note that irrespective of the value of $\varepsilon_{\mathbf{k}}$ we find $\omega_{+, \mathbf{k}} - \omega_{-, \mathbf{k}} \geq 2\sqrt{\sigma_0}$ so that the two resulting bands are separated by a gap of width $\geq 2\sqrt{\sigma_0}$. Thus, a single pole of the self-energy with macroscopic residuum σ_0 ‘pushes open’ a gap in the band structure around its position ζ . In fact, this is exactly the situation encountered in a Mott insulator, where the self-energy has a single dominant pole of strength $\approx U^2/4$ which ‘pushes open’ the Mott gap of width $\approx U$.

3 The Luttinger-Ward functional

3.1 The Green function as a functional integral

We start from the representation of the Green function in terms of a functional integral over *Grassmann variables*. An excellent introduction to Grassmann variables and their use in quantum field theory can be found in the textbook by Negele/Orland [9]. We give a brief summary for readers unfamiliar with Grassmann variables but it should be noted that this is nowhere near sufficient to fully understand their use. Grassmann variables are objects (here we write them as φ_i^* or φ_j , where i and j distinguish different Grassmann variables) which anticommute: $\varphi_i^* \varphi_j^* = -\varphi_j^* \varphi_i^*$, $\varphi_i \varphi_j^* = -\varphi_j^* \varphi_i$, and $\varphi_i \varphi_j = -\varphi_j \varphi_i$. It follows immediately that the square of any Grassmann variable is zero. While we will be using Grassmann variables with (φ_i^*) and without (φ_j) an asterisk it should be noted that there is no such thing as the complex conjugate of a Grassmann variable: just as the index, the asterisk is simply a part of the name of the Grassmann variable. Once we have defined a set of Grassmann variables, the corresponding *Grassmann algebra* consists of all possible sums of products of Grassmann variables with complex coefficients. For example, if we have the two Grassmann variables φ^* and φ , the corresponding algebra consists of all expressions $a_0 + a_1\varphi + a_2\varphi^* + a_3\varphi\varphi^*$ with complex a_0, a_1, a_2 and a_3 (note that all higher powers of Grassmann variables vanish). The key property of the Grassmann variables, which is used in quantum field theory, is the rule for ‘integration’ over Grassmann variables which is

$$\int d\varphi \varphi = 1, \quad \int d\varphi 1 = 0.$$

In other words, integration over Grassmann variables is the same as differentiation for ordinary numbers, the only difference being that there are just two ‘functions’ of a Grassmann variable. When we integrate over several Grassmann variables it has to be kept in mind that the differentials $d\varphi_i^*$ and $d\varphi_j$ anticommute as well. Transcendental functions of elements of the Grassmann variables are defined via their power series expansion. To illustrate the above, we evaluate

$$\begin{aligned} & \int d\varphi^* d\varphi e^{a_1\varphi + a_2\varphi^* + a_3\varphi\varphi^*} \\ &= \int d\varphi^* d\varphi \left(1 + (a_1\varphi + a_2\varphi^* + a_3\varphi\varphi^*) + \frac{a_1a_2}{2}(\varphi\varphi^* + \varphi^*\varphi) \right) = a_3. \end{aligned} \quad (32)$$

Note that the power series expansion of e^x terminates very quickly due to the fact that all powers of Grassmann variables higher than the first vanish. Moreover, there is a nonvanishing contribution to the integral only from those terms in the integrand where each Grassmann variable to be integrated over appears precisely once.

After these preliminaries, we turn to the representation of the imaginary-time Green function as functional integral over Grassmann variables. The full derivation is too lengthy to present here but it is explained in a very understandable way by Negele/Orland [9]. Let the imaginary time interval $[0, \hbar\beta]$ be divided into M intervals of length $\epsilon = \hbar\beta/M$, where we assume, for convenience, that M is even. Define the imaginary time grid points $\tau_k = k\epsilon$, $k = 1, \dots, M$.

Then introduce Grassmann variables $\varphi_{\gamma,k}^*$ and $\varphi_{\gamma,k}$ where γ is the ‘compound index’ that labels the Fermion operators $c_{\gamma}^{\dagger}$ and c_{γ} , whereas $k = 1, \dots, M$ denotes the position in the imaginary time grid. Then, the partition function can be written as $Z = \lim_{M \rightarrow \infty} Z_M$ where [9]

$$Z_M = \prod_{\gamma} \int \prod_{k=1}^M d\varphi_{\gamma,k}^* d\varphi_{\gamma,k} e^{-S(\varphi^*, \varphi)}$$

$$S = \epsilon \sum_{k=1}^M \left(\sum_{\gamma} \varphi_{\gamma,k}^* \frac{\varphi_{\gamma,k} - \varphi_{\gamma,k-1}}{\epsilon} + \frac{1}{\hbar} K(\varphi_{\gamma,k}^*, \varphi_{\gamma,k-1}) \right). \quad (33)$$

Here the ‘Grassmann Hamiltonian’ $K(\varphi_{\gamma,k}^*, \varphi_{\gamma,k-1})$ is obtained from the Fermionic Hamiltonian (1) by replacing $c_{\gamma}^{\dagger} \rightarrow \varphi_{\gamma,k}^*$ $c_{\gamma} \rightarrow \varphi_{\gamma,k-1}$. Note that in the definition of S the term with $k = 1$ also involves the variable $\varphi_{\gamma,0}$, which is *not* included in the set of integration variables. Rather one has to set $\varphi_{\gamma,0} = -\varphi_{\gamma,M}$ to account for Fermi statistics [9]. If we let $M \rightarrow \infty$, $\epsilon \rightarrow 0$, $\epsilon \sum_k \dots$ appears to become an integral whereas $(\varphi_{\gamma,k} - \varphi_{\gamma,k-1})/\epsilon$ appears to become a derivative whence, apparently,

$$S \rightarrow \int_0^{\hbar\beta} d\tau \left(\sum_{\gamma} \varphi_{\gamma}^* \frac{\partial \varphi_{\gamma}}{\partial \tau} + \frac{1}{\hbar} K(\varphi^*, \varphi) \right).$$

This notation can often be found in the literature but as stressed by Negele/Orland it can be misleading because the ‘trajectories’ $\varphi(\tau)$ may not at all be expected to be ‘smooth’.

Similarly, the imaginary time Green function (5) can be shown to be [9]

$$G_{\alpha,\beta}(\tau, \tau') = - \lim_{M \rightarrow \infty} \frac{1}{Z_M} \prod_{\gamma} \int \prod_{k=1}^M d\varphi_{\gamma,k}^* d\varphi_{\gamma,k} \varphi_{\alpha,k(\tau)}^* \varphi_{\beta,k(\tau')} e^{-S(\varphi^*, \varphi)}. \quad (34)$$

For each M the symbols $k(\tau)$ and $k(\tau')$ denote those points on the imaginary-time grid which are closest to τ and τ' so that $\epsilon k(\tau) \rightarrow \tau$ and $\epsilon k(\tau') \rightarrow \tau'$ for $M \rightarrow \infty$.

To illustrate the meaning of the functional integral let us calculate the partition function for non-interacting electrons where $K = \sum_{\gamma} (E_{\gamma} - \mu) c_{\gamma}^{\dagger} c_{\gamma}$. Constructing the Grassmann Hamiltonian as described above, introducing $\zeta_{\gamma} = (E_{\gamma} - \mu)/\hbar$ and substituting $\varphi_{\gamma,0} = -\varphi_{\gamma,M}$ we find

$$S(\varphi^*, \varphi) = \sum_{\gamma} \sum_{k=2}^M \left[(\varphi_{\gamma,k}^* \varphi_{\gamma,k} - \varphi_{\gamma,k}^* \varphi_{\gamma,k-1}) + \epsilon \zeta_{\gamma} \varphi_{\gamma,k}^* \varphi_{\gamma,k-1} \right]$$

$$+ \sum_{\gamma} \left[(\varphi_{\gamma,1}^* \varphi_{\gamma,1} + \varphi_{\gamma,1}^* \varphi_{\gamma,M}) - \epsilon \zeta_{\gamma} \varphi_{\gamma,1}^* \varphi_{\gamma,M} \right].$$

This has the form $\sum_{\gamma} S_{\gamma}$ where S_{γ} contains only $\varphi_{\gamma,k}^*$ and $\varphi_{\gamma,k}$ and the different S_{γ} are sums of products of two Grassmann variables and therefore commute with each other. It follows that $e^{-S} = \prod_{\gamma} e^{-S_{\gamma}}$ and we can compute Z_M by integrating each $e^{-S_{\gamma}}$ separately und multiplying the results. Accordingly, we drop the eigenstate index γ for the time being.

We introduce some notation

$$S(\varphi^*, \varphi) = \sum_{k=1}^M (S_{1,k} + S_{2,k}) = \sum_{k=1}^M S_k$$

with

$$S_{1,k} = -\varphi_k \varphi_k^* \quad S_{2,k} = (\epsilon \zeta - 1) \begin{cases} \varphi_k^* \varphi_{k-1} & k=2, \dots, M, \\ -\varphi_1^* \varphi_M & k=1, \end{cases}$$

and proceed to exponentiate $-S(\varphi^*, \varphi)$. Since $S(\varphi^*, \varphi)$ is a sum of products of two Grassmann variables, S^L is a sum of products of $2L$ Grassmann variables. In total there are $2M$ different Grassmann variables (namely φ_k and φ_k^* for $k = 1, \dots, M$) so that any product of more than $2M$ Grassmann variables necessarily contains at least two Grassmann variables more than once and thus is zero. Hence, $S^L = 0$ for $L > M$. On the other hand, the integration measure is $\prod_{k=1}^M d\varphi_k^* d\varphi_k$ so that $2M$ Grassmann variables are integrated over. Therefore, any product of less than $2M$ Grassmann variables necessarily gives zero when integrated over, so that in the end only the term $(-S)^M$ in the expansion of e^{-S} gives a nonvanishing contribution

$$\begin{aligned} \int \prod_{k=1}^M d\varphi_k^* d\varphi_k e^{-S(\varphi^*, \varphi)} &= \frac{(-1)^M}{M!} \int \prod_{k=1}^M d\varphi_k^* d\varphi_k S^M(\varphi^*, \varphi) \\ &= (-1)^M \int \prod_{k=1}^M d\varphi_k^* d\varphi_k S_M S_{M-1} \dots S_2 S_1. \end{aligned}$$

In going to the second line we have used the fact that $S_k^2 = 0$ whence S^M is the sum over those $M!$ terms where all M different S_k are multiplied in different sequence. However, all of these $M!$ products are equal because the S_k commute. Next, since $S_k = S_{1,k} + S_{2,k}$, the product $S_M S_{M-1} \dots S_2 S_1$ is the sum of 2^M terms obtained by picking either $S_{1,k}$ or $S_{2,k}$ for each k . To see which terms give nonvanishing result upon integration, we switch to a graphical representation, see Figure 3. Let us pick $S_{1,k}$ for $k=1$, see Figure 3b). Then, if we pick $S_{2,k}$ for $k=2$ we get a factor of $\varphi_1^2 = 0$. We therefore *have to* pick $S_{1,k}$ also for $k=2$ and this repeats at every k . Accordingly, if we start with $S_{1,1}$ we need to continue with $S_{1,k}$ until $k=M$. Next, let us pick $S_{2,k}$ for $k=1$, see Figure 3c). Then, if we picked $S_{1,k}$ for $k=2$ we would not get a factor of φ_1 in the product and integration over $d\varphi_1$ would give zero. We therefore have to pick $S_{2,2}$. And again this repeats at every k , so if we start with $S_{2,1}$ we have to continue with $S_{2,k}$ for every k so that in the end we find

$$\int \prod_{k=1}^M d\varphi_k^* d\varphi_k S_M S_{M-1} \dots S_2 S_1 = \int \prod_{k=1}^M d\varphi_k^* d\varphi_k \left(\prod_{k=1}^M S_{1,k} + \prod_{k=1}^M S_{2,k} \right).$$

When we start with $S_{2,1}$, however, we still have to bring φ_M to the first position of the product, see Figure 3c), to obtain the form $\prod_{k=1}^M \varphi_k \varphi_k^*$. To do so we have to commute φ_M through $2M-1$ Grassmann variables, resulting in an extra factor of (-1) . Reinserting the γ -labels and

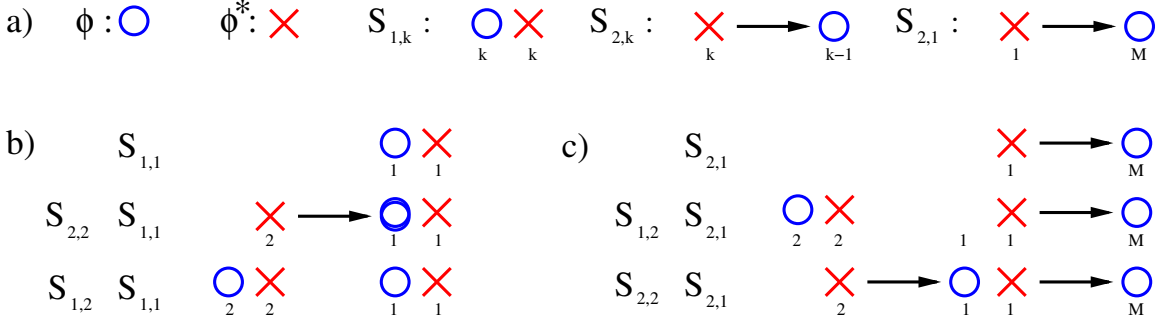


Fig. 3: a) Explanation of symbols, b) starting with $S_{1,1}$, c) starting with $S_{2,1}$.

substituting $\epsilon = \hbar\beta/M$ and $\zeta_\gamma = (E_\gamma - \mu)/\hbar$ we finally obtain

$$\begin{aligned}
 Z &= \prod_{\gamma} \lim_{M \rightarrow \infty} (-1)^M [(-1)^M + (\epsilon\zeta_\gamma - 1)^M] \\
 &= \prod_{\gamma} \left[1 + \lim_{M \rightarrow \infty} \left(1 - \frac{\beta(E_\gamma - \mu)}{M} \right)^M \right] = \prod_{\gamma} (1 + e^{-\beta(E_\gamma - \mu)}),
 \end{aligned}$$

which is indeed the partition function for noninteracting fermions. Doing all calculations at finite M and taking the limit $M \rightarrow \infty$ in the end is unwieldy but this can be circumvented by switching to the Fourier transforms $\tilde{\varphi}_{\gamma,\nu}$ and $\tilde{\varphi}_{\gamma,\nu}^*$, defined as

$$\tilde{\varphi}_{\gamma,\nu}^* = \frac{1}{\sqrt{M}} \sum_{k=1}^M e^{-i\omega_\nu \tau_k} \varphi_{\gamma,k}^*, \quad \tilde{\varphi}_{\gamma,\nu} = \frac{1}{\sqrt{M}} \sum_{k=1}^M e^{i\omega_\nu \tau_k} \varphi_{\gamma,k}. \quad (35)$$

The frequencies $\omega_\nu = (2\nu+1)\pi/\hbar\beta$ are the Fermionic Matsubara frequencies defined above. The second equation looks like the Hermitean conjugate of the first one but recall that there is no such thing as the Hermitean conjugate of a Grassmann variable so this is actually a *definition*. If we restrict ν to $-M/2 \leq \nu \leq M/2 - 1$, so that the number of ω_ν equals the number of τ_k , the above transformations are unitary as shown in Appendix B, and can be reverted to give

$$\varphi_{\gamma,k}^* = \frac{1}{\sqrt{M}} \sum_{\nu} e^{i\omega_\nu \tau_k} \tilde{\varphi}_{\gamma,\nu}^*, \quad \varphi_{\gamma,k} = \frac{1}{\sqrt{M}} \sum_{\nu} e^{-i\omega_\nu \tau_k} \tilde{\varphi}_{\gamma,\nu}. \quad (36)$$

Note that $e^{i\omega_\nu \tau_M} = e^{i(2\nu+1)\pi} = -1$. Accordingly, if we *define* $\varphi_{\gamma,0}$ to be (36) evaluated for $k=0$ we automatically have $\varphi_{\gamma,0} = -\varphi_{\gamma,M}$ which was implicit in the definition of S . Since the transformation from the $\varphi_{\gamma,k}$ to the $\tilde{\varphi}_{\gamma,\nu}$ is unitary the Jacobian is unity and

$$\prod_{k=1}^M \prod_{\gamma} \int d\varphi_{\gamma,k}^* d\varphi_{\gamma,k} \rightarrow \prod_{\nu=-M/2}^{M/2-1} \prod_{\gamma} \int d\tilde{\varphi}_{\gamma,\nu}^* d\tilde{\varphi}_{\gamma,\nu}$$

We proceed to rewrite S in terms of the $\tilde{\varphi}^*$ and $\tilde{\varphi}$. First, consider the ‘derivative term’

$$\epsilon \sum_{k=1}^M \varphi_{\gamma,k}^* \frac{\varphi_{\gamma,k} - \varphi_{\gamma,k-1}}{\epsilon} = \sum_{\nu,\nu'} \frac{\epsilon}{M} \sum_{k=1}^M e^{i(\omega_\nu - \omega_{\nu'})\tau_k} \frac{1 - e^{i\omega_{\nu'}\epsilon}}{\epsilon} \tilde{\varphi}_{\nu}^* \tilde{\varphi}_{\nu'} = \epsilon \sum_{\nu} \frac{1 - e^{i\omega_\nu\epsilon}}{\epsilon} \tilde{\varphi}_{\nu}^* \tilde{\varphi}_{\nu}$$

In the limit of small ϵ (i.e. $M \rightarrow \infty$) we approximate the square bracket as

$$-e^{\frac{i\omega_\nu\epsilon}{2}} \frac{e^{\frac{i\omega_\nu\epsilon}{2}} - e^{-\frac{i\omega_\nu\epsilon}{2}}}{\epsilon} = -e^{\frac{i\omega_\nu\epsilon}{2}} \left(\frac{de^{i\omega_\nu x}}{dx} \Big|_{x=0} + \mathcal{O}(\epsilon^2) \right) \rightarrow -i\omega_\nu e^{\frac{i\omega_\nu\epsilon}{2}} \approx -i\omega_\nu e^{i\omega_\nu\epsilon} \quad (37)$$

and arrive at the final expression

$$\epsilon \sum_{k=1}^M \varphi_{\gamma,k}^* \frac{\varphi_{\gamma,k} - \varphi_{\gamma,k-1}}{\epsilon} = -\epsilon \sum_{\nu} i\omega_\nu e^{i\omega_\nu\epsilon} \tilde{\varphi}_\nu^* \tilde{\varphi}_\nu.$$

Readers may find the approximation (37) and in fact even keeping the factors of $e^{i\omega_\nu\epsilon}$ when $\epsilon \rightarrow 0$ questionable, but indeed both make sense as shown in Appendix C. Next, a typical term in K_0 becomes

$$\epsilon \sum_{k=1}^M (t_{\gamma,\beta} - \mu \delta_{\gamma,\beta}) \varphi_{\gamma,k}^* \varphi_{\beta,k-1} = \epsilon \sum_{\nu} e^{i\omega_\nu\epsilon} (t_{\gamma,\beta} - \mu \delta_{\gamma,\beta}) \tilde{\varphi}_{\gamma,\nu}^* \tilde{\varphi}_{\beta,\nu}$$

whereas the products of 4 Grassmann variables in K_1 become

$$\epsilon \sum_{k=1}^M \varphi_{\alpha,k}^* \varphi_{\beta,k}^* \varphi_{\gamma,k-1} \varphi_{\delta,k-1} = \frac{\epsilon^2}{\hbar\beta} \sum_{\nu_1, \nu_2, \nu_3, \nu_4} \delta_{\nu_1+\nu_2, \nu_3+\nu_4} e^{i(\omega_{\nu_3}+\omega_{\nu_4})\epsilon} \tilde{\varphi}_{\alpha,\nu_1}^* \tilde{\varphi}_{\beta,\nu_2}^* \tilde{\varphi}_{\gamma,\nu_3} \tilde{\varphi}_{\delta,\nu_4}.$$

The various powers of ϵ which show up as prefactors can be eliminated by introducing the rescaled Fourier transform $\varphi_{\gamma,\nu} = \sqrt{\epsilon} \tilde{\varphi}_{\gamma,\nu}$, which is actually quite natural when going over to the limit of continuous imaginary time, see Appendix B. The latter transformation is no longer unitary, rather the Jacobian is ϵ^M which does not have a well-defined limit as $M \rightarrow \infty$. However, for any finite M the Jacobian is the same in both numerator and denominator in (34) and simply cancels. Having rewritten S in terms of the (rescaled) Fourier amplitudes the limit $M \rightarrow \infty$ becomes trivial: since we had $-M/2 \leq \nu \leq M/2 - 1$, we simply let each ν range over all integers and our final result reads

$$-S[\varphi^*, \varphi] = \sum_{\nu} \varphi_{\alpha,\nu}^* e^{i\omega_\nu\epsilon} G_{0,\alpha,\beta}^{-1}(i\omega_\nu) \varphi_{\beta,\nu} - \frac{1}{2\hbar^2\beta} \sum_{\substack{\alpha,\beta,\gamma,\delta \\ \nu_1,\nu_2,\nu_3,\nu_4}} V_{\alpha\beta,\delta,\gamma} \delta_{\nu_1+\nu_2, \nu_3+\nu_4} e^{2i(\omega_{\nu_3}+\omega_{\nu_4})\epsilon} \varphi_{\alpha,\nu_1}^* \varphi_{\beta,\nu_2}^* \varphi_{\gamma,\nu_3} \varphi_{\delta,\nu_4}. \quad (38)$$

Here the limit $\epsilon \rightarrow 0^+$ has to be taken and the matrix $\mathbf{G}_0^{-1}(i\omega_\nu) = i\omega_\nu - (\mathbf{t} - \mu)/\hbar$ is the inverse noninteracting Green function, see (19). Next, using Eqs. (6), (7) it is easy to show that

$$\frac{1}{\hbar\beta} \int_0^{\hbar\beta} d\tau \int_0^{\hbar\beta} d\tau' e^{i\omega_\nu\tau} e^{-i\omega_{\nu'}\tau'} G(\tau, \tau') = \delta_{\nu,\nu'} G(i\omega_\nu),$$

the ‘finite- M -version’ of which is

$$\begin{aligned} G_{\alpha,\beta}(i\omega_\nu) &= \lim_{M \rightarrow \infty} \frac{\epsilon^2}{\hbar\beta} \sum_{k,k'=1}^M e^{i\omega_\nu\tau_k} e^{-i\omega_{\nu'}\tau_{k'}} G_{\alpha,\beta}(\tau_k, \tau_{k'}) \\ &= - \frac{\prod_{\mu=-\infty}^{\infty} \prod_{\gamma} \int d\varphi_{\gamma,\mu}^* d\varphi_{\gamma,\mu} \varphi_{\alpha,\nu} \varphi_{\beta,\nu}^* e^{-S(\varphi^*, \varphi)}}{\prod_{\mu=-\infty}^{\infty} \prod_{\gamma} \int d\varphi_{\gamma,\mu}^* d\varphi_{\gamma,\mu} e^{-S(\varphi^*, \varphi)}}, \end{aligned} \quad (39)$$

because $\frac{\epsilon^2}{\hbar\beta} = \left(\frac{\sqrt{\epsilon}}{\sqrt{M}}\right)^2$. This expression readily lends itself to a perturbation expansion in K_1 which leads to the known representation of the Green function in terms of Feynman diagrams [9]. However, we do not pursue this here but switch to our central objective, namely the construction of the Luttinger-Ward functional, thereby closely following Potthoff [5, 6].

3.2 Construction of the Luttinger-Ward functional

We start by slightly changing our point of view. We note that the various objects that we are concerned with, the noninteracting Green function $\mathbf{G}_0$, the full Green function $\mathbf{G}$ and the self-energy Σ , all are just sets of complex numbers: $F_{\alpha,\beta}(i\omega_\nu)$, $F \in \{G_0, G, \Sigma\}$. Then, we may consider (39) with (38) as the definition of a functional $\mathcal{G}[\mathbf{G}_0^{-1}]$, which maps one such set, $\mathbf{G}_0^{-1}(i\omega_\nu)$, to another one: $\mathbf{G}(i\omega_\nu)$. If the initial set happens to be a noninteracting Green function of the form $\mathbf{G}_0^{-1}(i\omega_\nu) = i\omega_\nu - \frac{1}{\hbar}(\mathbf{t} - \mu)$ with some set of physical single particle elements $\mathbf{t}$, $\mathcal{G}[\mathbf{G}_0^{-1}]$ yields the corresponding interacting Green function, with interaction part K_1 . Accordingly, K_1 in (38) may be viewed as an implicit parameter of the functional and we consider K_1 fixed from now on. One may wonder if the functional $\mathcal{G}[\mathbf{G}_0^{-1}]$ gives a well-defined result for any choice of $\mathbf{G}_0^{-1}$. This is probably not the case but we simply restrict its domain to those $\mathbf{G}_0^{-1}$ which give a well-defined result. Next consider the functional

$$\Omega[\mathbf{G}_0^{-1}] = -\frac{1}{\beta} \ln \left(\prod_{\nu=-\infty}^{\infty} \prod_{\gamma} \int d\varphi_{\gamma,\nu}^* d\varphi_{\gamma,\nu} e^{-S(\varphi^*, \varphi)} \right),$$

which is the grand canonical potential for a physical $\mathbf{G}_0^{-1}$ because the bracket is Z in this case, see (33). Its variation with respect to $\mathbf{G}_0^{-1}$ is

$$\beta \frac{\partial \Omega[\mathbf{G}_0^{-1}]}{\partial G_{0,\alpha,\beta}^{-1}(i\omega_\nu)} = -e^{i\omega_\nu \epsilon} \mathcal{G}[\mathbf{G}_0^{-1}]_{\beta,\alpha}(i\omega_\nu). \quad (40)$$

Note the ‘inverted’ order of indices of G on the right hand side, which follows by a comparison of (38) and (39). Next, for a given ‘self energy’ Σ and a given ‘Green function’ $\mathbf{G}$ consider

$$\mathbf{D}[\mathbf{G}, \Sigma] = \mathcal{G}[\mathbf{G}^{-1} + \Sigma] - \mathbf{G}.$$

Assume Σ is a true self-energy corresponding to some $\mathbf{G}_0^{-1}(i\omega_\nu) = i\omega_\nu - \frac{1}{\hbar}(\mathbf{t} - \mu)$ (and interaction K_1) and $\mathbf{G}$ the corresponding interacting Green function. By the Dyson equation $\mathbf{G}^{-1} + \Sigma = \mathbf{G}_0^{-1}$, and since $\mathcal{G}[\mathbf{G}_0^{-1}] = \mathbf{G}$ we find $\mathbf{D}[\mathbf{G}, \Sigma] = 0$. Next, for a given ‘Green function’ $\mathbf{G}$ define the functional $\mathcal{S}[\mathbf{G}]$ to be the ‘self energy’ which minimizes the norm of $\mathbf{D}[\mathbf{G}, \Sigma]$, defined as

$$|\mathbf{D}| = \frac{1}{N} \sum_{\alpha,\beta} \sum_{\nu} |D_{\alpha,\beta}(i\omega_\nu)|^2.$$

For a physical Green function $\mathbf{G}$ and the corresponding self-energy Σ we found $\mathbf{D} = 0$, which gives the minimal value of $|\mathbf{D}|$, whence $\mathcal{S}[\mathbf{G}] = \Sigma$ and $\mathcal{S}$ simply maps the Green function to

the self-energy. If $\mathbf{G}$ is not a physical Green function, but sufficiently ‘reasonable’ such that the above functionals can be defined, we expect

$$\mathcal{G} [\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]] = \mathbf{G} + \delta \mathbf{G} \quad (41)$$

where $\delta \mathbf{G} \rightarrow 0$ whenever $\mathbf{G}$ approaches a physical Green function. After these preliminaries we define the *Luttinger-Ward functional*, which is the following complex valued functional of a Green function

$$\Phi[\mathbf{G}] = \Omega [\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]] + \frac{1}{\beta} \sum_{\lambda} e^{i\omega_{\lambda}\epsilon} \left(-\ln \det \mathbf{G}(i\omega_{\lambda}) + \text{Tr} \mathbf{G}(i\omega_{\lambda}) \mathcal{S}[\mathbf{G}](i\omega_{\lambda}) \right). \quad (42)$$

The logarithm of the determinant is the sum of the logarithms of the $2Nn_{orb}$ eigenvalues of $\mathbf{G}(i\omega_{\lambda})$ and the limit $\epsilon \rightarrow 0^+$ is understood (to shorten the notation we henceforth omit factors like $e^{i\omega_{\lambda}\epsilon}$ and reinsert them only in the final result). Consider the variation of this functional under a change of $\mathbf{G}$. Using (40), (41) and the chain rule we find the variation of the first term

$$\begin{aligned} \beta \frac{\partial \Omega [\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]]}{\partial G_{\alpha,\beta}(i\omega_{\nu})} &= - \sum_{\lambda} \sum_{\delta,\gamma} \mathcal{G} [\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]]_{\delta,\gamma}(i\omega_{\lambda}) \frac{\partial (G^{-1} + \mathcal{S}[\mathbf{G}])_{\gamma,\delta}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} \\ &= - \sum_{\lambda} \sum_{\delta,\gamma} (G_{\delta,\gamma} + \delta G_{\delta,\gamma})(i\omega_{\lambda}) \frac{\partial (G^{-1} + \mathcal{S}[\mathbf{G}])_{\gamma,\delta}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} \end{aligned} \quad (43)$$

Next, notice that for each λ

$$\sum_{\gamma,\delta} G_{\delta,\gamma}(i\omega_{\lambda}) G_{\gamma,\delta}^{-1}(i\omega_{\lambda}) = 2Nn_{orb}, \quad (44)$$

and since the $G_{\alpha,\beta}(i\omega_{\nu})$ for different α, β and ν must be considered as independent variables

$$\frac{\partial G_{\delta,\gamma}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} = \delta_{\nu,\lambda} \delta_{\alpha,\delta} \delta_{\beta,\gamma}.$$

Differentiating (44) with respect to $G_{\alpha,\beta}(i\omega_{\nu})$ we therefore obtain

$$\begin{aligned} \sum_{\gamma,\delta} \left(\delta_{\nu,\lambda} \delta_{\alpha,\delta} \delta_{\beta,\gamma} G_{\gamma,\delta}^{-1}(i\omega_{\nu}) + G_{\delta,\gamma}(i\omega_{\lambda}) \frac{\partial G_{\gamma,\delta}^{-1}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} \right) &= 0 \\ \Rightarrow - \sum_{\gamma,\delta} G_{\delta,\gamma}(i\omega_{\lambda}) \frac{\partial G_{\gamma,\delta}^{-1}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} &= \delta_{\nu,\lambda} G_{\beta,\alpha}^{-1}(i\omega_{\nu}), \end{aligned}$$

and inserting this into (43) we find

$$\beta \frac{\partial \Omega [\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]]}{\partial G_{\alpha,\beta}(i\omega_{\nu})} = G_{\beta,\alpha}^{-1}(i\omega_{\nu}) - \sum_{\lambda} \sum_{\delta,\gamma} G_{\delta,\gamma}(i\omega_{\lambda}) \frac{\partial \mathcal{S}[\mathbf{G}]_{\gamma,\delta}(i\omega_{\lambda})}{\partial G_{\alpha,\beta}(i\omega_{\nu})} + \mathcal{O}(\delta G). \quad (45)$$

Using Appendix D the derivative of the second term on the right hand side of (42) becomes

$$\beta \frac{\partial}{\partial G_{\alpha,\beta}(i\omega_{\nu})} \left(-\frac{1}{\beta} \sum_{\lambda} \ln \det \mathbf{G}(i\omega_{\lambda}) \right) = -G_{\beta,\alpha}^{-1}(i\omega_{\nu}), \quad (46)$$

whereas the derivative of the third term (multiplied by β) is

$$\mathcal{S}[\mathbf{G}]_{\beta,\alpha}(i\omega_\nu) + \sum_{\lambda} \sum_{\delta,\gamma} G_{\delta,\gamma}(i\omega_\lambda) \frac{\partial \mathcal{S}[\mathbf{G}]_{\gamma,\delta}(i\omega_\lambda)}{\partial G_{\alpha,\beta}(i\omega_\nu)}. \quad (47)$$

Adding up (45), (46) and (47) and reinserting the exponential of $e^{i\omega_\nu\epsilon}$ we obtain

$$\beta \frac{\partial \Phi[\mathbf{G}]}{\partial G_{\alpha,\beta}(i\omega_\nu)} = e^{i\omega_\nu\epsilon} \mathcal{S}[\mathbf{G}]_{\beta,\alpha}(i\omega_\nu) + \mathcal{O}(\delta G). \quad (48)$$

If we now let $\mathbf{G}$ become the physical Green function for some noninteracting Green function (and the fixed interaction K_1) $\mathcal{S}[\mathbf{G}]$ becomes the physical self-energy and δG vanishes. We arrive at the key result that the self-energy is the functional derivative of the Luttinger-Ward functional Φ with respect to the Green function (but note the inverted matrix indices on the right hand side). For a physical $\mathbf{G}$, $\Omega[\mathbf{G}^{-1} + \mathcal{S}[\mathbf{G}]]$ becomes the grand canonical potential and we arrive at the expression for Ω first derived by Luttinger and Ward

$$\Omega = -\lim_{\epsilon \rightarrow 0} \frac{1}{\beta} \sum_{\lambda} e^{i\omega_\lambda\epsilon} (\ln \det \mathbf{G}^{-1}(i\omega_\lambda) + \text{Tr} \mathbf{G}(i\omega_\lambda) \mathbf{\Sigma}(i\omega_\lambda)) + \Phi[\mathbf{G}], \quad (49)$$

whereas Eq. (48) becomes

$$\frac{1}{\beta} e^{i\omega_\nu\epsilon} \mathbf{\Sigma}_{\beta,\alpha}(i\omega_\nu) = \frac{\partial \Phi[\mathbf{G}]}{\partial G_{\alpha,\beta}(i\omega_\nu)}. \quad (50)$$

As it stands Eq. (49) is not really helpful for evaluating Ω , because if one wanted to use the definition (42) to evaluate $\Phi[\mathbf{G}]$ one would have to know Ω in the first place. However, Luttinger and Ward have given a representation of $\Phi[\mathbf{G}]$ as a sum over infinitely many Feynman diagrams [1]. Moreover, a highly nontrivial result is Eq. (50), which shows that the self-energy is the derivative of the functional $\Phi[\mathbf{G}]$ with respect to the Green function.

Next, the expression (49) gives Ω as a functional of $\mathbf{G}$ (the self-energy $\mathbf{\Sigma}$ can be eliminated using the Dyson equation) and in the next step we want to re-write Ω as a functional of $\mathbf{\Sigma}$. Since the self-energy is the derivative of Φ with respect to $\mathbf{G}$, see Eq. (50) this can be done by Legendre transform. Assuming that the functional $\mathcal{S}[\mathbf{G}]$ can be inverted to yield $\mathbf{G}[\mathbf{\Sigma}]$ define the Legendre transform

$$\begin{aligned} F[\mathbf{\Sigma}] &= \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \sum_{\lambda} \sum_{\gamma,\delta} G[\mathbf{\Sigma}]_{\delta,\gamma}(i\omega_\lambda) \frac{\partial \Phi[\mathbf{G}]}{\partial G_{\delta,\gamma}(i\omega_\lambda)} \\ &= \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \frac{1}{\beta} \sum_{\lambda} \sum_{\gamma,\delta} e^{i\omega_\lambda\epsilon} G[\mathbf{\Sigma}]_{\delta,\gamma}(i\omega_\lambda) \mathbf{\Sigma}_{\gamma,\delta}(i\omega_\lambda) \end{aligned}$$

and inserting this into (49) we find

$$\Omega = -\lim_{\epsilon \rightarrow 0} \frac{1}{\beta} \sum_{\lambda} e^{i\omega_\lambda\epsilon} \ln \det (\mathbf{G}_0^{-1}(i\omega_\lambda) - \mathbf{\Sigma}(i\omega_\lambda)) + F[\mathbf{\Sigma}]. \quad (51)$$

By virtue of being a Legendre transform this functional obeys

$$\beta \frac{\partial F[\mathbf{\Sigma}]}{\partial \Sigma_{\alpha,\beta}(i\omega_\nu)} = -e^{i\omega_\nu\epsilon} G_{\beta,\alpha}(i\omega_\nu), \quad (52)$$

which is also easily verified directly. Using this and again the theorem from Appendix D we arrive at the key result:

$$\frac{\partial \Omega}{\partial \Sigma_{\alpha,\beta}(i\omega_\nu)} = 0. \quad (53)$$

Let us summarize what we found. First, there exists a functional $\Phi[\mathbf{G}]$ such that the grand potential Ω is given by (49). By itself this is not really spectacular because one could always define $\Phi[\mathbf{G}]$ via (49). What is highly nontrivial, however, is the fact that the self-energy can be obtained as the functional derivative of $\Phi[\mathbf{G}]$ at the exact $\mathbf{G}$, see (50). Similarly, there exists a functional $F[\Sigma]$ such that Ω is given by (51) and the Green function is obtained as the derivative of $F[\Sigma]$, see (52). Moreover, when expressed as a functional of the self-energy, Ω is stationary with respect to variations of Σ , see (53), *at the exact self-energy*. Moreover, the functionals $\Phi[\mathbf{G}]$ and $F[\Sigma]$ have an additional important property: they depend only on the interaction part K_1 in the Hamiltonian, but are completely independent of the noninteracting term K_0 or, equivalently, the noninteracting Green function $\mathbf{G}_0$. To see this, let us carefully go through our above construction of the Luttinger-Ward functional. The functionals $\mathcal{G}[\mathbf{G}_0^{-1}]$ and $\Omega[\mathbf{G}_0^{-1}]$ have the noninteracting Green function as their argument while there is no further ‘intrinsic’ dependence on $\mathbf{G}_0$ (these functionals do depend on K_1 , however!). It follows that the functional $\mathcal{S}[\mathbf{G}]$, which is defined using only the functional $\mathcal{G}$, does not have any implicit dependence on $\mathbf{G}_0$ either, i.e., it can in principle be evaluated without knowledge of $\mathbf{G}_0$. Accordingly, $\Phi[\mathbf{G}]$ is a functional only of the $G_{\alpha,\beta}(i\omega_\lambda)$ with no implicit dependence on $\mathbf{t}$ and μ , so that the same holds true for its Legendre-transform $F[\Sigma]$.

4 Self-energy functional theories

4.1 Dynamical mean-field theory

In the preceding section we have seen that the grand potential Ω of an interacting Fermi system can be represented as a functional of its self-energy, see Eq. (51), and that this functional is stationary at the true self-energy, see Eq. (53). This is a promising result, in that such stationarity principles are put to use in almost every area of physics. In fact, one might come up with the following variational determination of the self-energy: choose a trial self-energy of the form (27), whence Ω ultimately becomes a functional of the spectral density $\sigma(\omega)$ ($\mathbf{V}^{(HF)}$ is itself a functional of the Green function and hence the self-energy) and then use (53) to derive an Euler-Lagrange equation for $\sigma(\omega)$. The problem is that the definition of the functional $\Phi[\mathbf{G}]$, Eq. (42), allowed to prove some of its properties, such as (50), but that it is completely impossible to actually evaluate it for a given ‘trial $\mathbf{G}$ ’ (this would for example necessitate knowledge of Ω). Accordingly, it is equally impossible to give an explicit expression for its Legendre transform $F[\Sigma]$ and evaluate this for a given ‘trial Σ ’.

However, we can circumvent this problem by making use of the fact that the functional $F[\Sigma]$ does not depend on the noninteracting part of the Hamiltonian K_0 , but only on the interaction part K_1 . For concreteness we consider the well-known Hubbard model on a lattice with N sites

and periodic boundary conditions

$$H = \sum_{i,j} \sum_{\sigma} t_{i,j} c_{i,\sigma}^{\dagger} c_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} = \sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k},\sigma}^{\dagger} c_{\mathbf{k},\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow}.$$

Thereby $n_{i,\sigma} = c_{i,\sigma}^{\dagger} c_{i,\sigma}$ and $\varepsilon_{\mathbf{k}}$ is the Fourier transform of $t_{i,j}$. In addition we consider an artificial system, called the *reference system*, with Hamiltonian $\tilde{H} = \sum_i \tilde{H}_i$, where all H_i have the identical form of an Anderson impurity model

$$\tilde{H}_i = \sum_{\nu} \varepsilon_{\nu} l_{i,\nu,\sigma}^{\dagger} l_{i,\nu,\sigma} + \sum_{\nu} (V_{\nu} l_{i,\nu,\sigma}^{\dagger} c_{i,\sigma} + H.c.) + U n_{i,\uparrow} n_{i,\downarrow}, \quad (54)$$

with again $n_{i,\sigma} = c_{i,\sigma}^{\dagger} c_{i,\sigma}$. This describes a single ‘Hubbard-site’ (described by $c_{i,\sigma}^{\dagger}$ and $c_{i,\sigma}$) coupled to a number of ‘ligand’ sites (described by $l_{i,\nu,\sigma}^{\dagger}$ and $l_{i,\nu,\sigma}$). We assume that the impurity system can be solved and denote the Green function and self-energy for this system by $\tilde{G}$ and $\tilde{\Sigma}$. We have seen above that all matrix elements of the self-energy involving uncorrelated sites vanish. Since the ligands in the Hamiltonian (54) are uncorrelated, the only nonvanishing element is $\tilde{\Sigma}_{c,c}(z)$.

After these preliminaries we come to the crucial point: since H and $\tilde{H}$ have the same interaction, namely $K_1 = U \sum_i n_{i,\uparrow} n_{i,\downarrow}$, the functional $F[\tilde{\Sigma}]$ is the same for the Hubbard model and the reference system. We now make use of this fact by taking the self-energy $\tilde{\Sigma}$ of the impurity model to be the ‘trial self-energy’ for the lattice system. Since the reference system consists of disconnected clusters its self-energy is diagonal in real space: $\tilde{\Sigma}_{(i,c),(j,c)}(z) = \delta_{i,j} \tilde{\Sigma}_{c,c}(z)$, whence its Fourier transform is $\mathbf{k}$ -independent: $\tilde{\Sigma}(\mathbf{k}, z) = \tilde{\Sigma}_{c,c}(z)$. The resulting Green function of the lattice therefore is $G^{-1}(\mathbf{k}, z) = z - (\varepsilon_{\mathbf{k}} - \mu)/\hbar - \tilde{\Sigma}_{c,c}(z)$. Using (51) we thus obtain an *approximate* Ω_{latt} for the Hubbard model (the factor of 2 is for spin)

$$\Omega_{latt} = F[\tilde{\Sigma}] - \lim_{\varepsilon \rightarrow 0} \frac{2}{\beta} \sum_{\mathbf{k}} \sum_{\lambda} e^{i\omega_{\lambda}\varepsilon} \ln \left(i\omega_{\lambda} - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar} - \tilde{\Sigma}_{c,c}(i\omega_{\lambda}) \right). \quad (55)$$

(since there is only one site/unit cell the Green function is diagonalized by Fourier transform and its determinant of its inverse is just $\prod_{\mathbf{k}} G_{bfk}^{-1}$). Now we know that Ω is stationary under variations of the self-energy, see Eq. (53). We cannot perform a variation of the self-energy over *all* possible self-energies but we may restrict the domain to ‘impurity model representable’ ones, that means those self-energies, which can be obtained by solution of the impurity model for some set of parameters $\{\varepsilon_{\nu}, V_{\nu}\}$. The variation of $\tilde{\Sigma}$ then amounts to a variation of the parameters ε_{ν} and V_{ν} in (54). Let us discuss how the value of $F[\tilde{\Sigma}]$ changes with the parameters of the impurity model, $t \in \{\varepsilon_{\nu}, V_{\nu}\}$. Using the chain rule and (52) we find

$$\begin{aligned} \frac{\partial F[\tilde{\Sigma}]}{\partial t} &= \sum_{\alpha,\beta} \sum_{\lambda} \frac{\partial F[\tilde{\Sigma}]}{\partial \tilde{\Sigma}_{\alpha,\beta}(i\omega_{\lambda})} \frac{\partial \tilde{\Sigma}_{\alpha,\beta}(i\omega_{\lambda})}{\partial t} \\ &= -\frac{1}{\beta} \sum_{\alpha,\beta} \sum_{\lambda} e^{i\omega_{\lambda}\varepsilon} \tilde{G}_{\beta,\alpha}(i\omega_{\lambda}) \frac{\partial \tilde{\Sigma}_{\alpha,\beta}(i\omega_{\lambda})}{\partial t} \\ &= -\frac{2N}{\beta} \sum_{\lambda} e^{i\omega_{\lambda}\varepsilon} \tilde{G}_{c,c}(i\omega_{\lambda}) \frac{\partial \tilde{\Sigma}_{c,c}(i\omega_{\lambda})}{\partial t}. \end{aligned}$$

The last equation follows from the fact that there are N lattice sites i and the two spin directions for each site. It follows that

$$\frac{\partial \Omega_{latt}}{\partial t} = -\frac{2}{\beta} \sum_{\mathbf{k}, \lambda} e^{i\omega_\lambda \varepsilon} \left[\tilde{G}_{c,c}(i\omega_\lambda) - G(\mathbf{k}, i\omega_\lambda) \right] \frac{\partial \tilde{\Sigma}_{c,c}(i\omega_\lambda)}{\partial t} = 0. \quad (56)$$

We may now perform a limiting procedure by sending the number of ligands $\rightarrow \infty$ that means considering a continuum of ligands which hybridize with the correlated c -orbital. Then the number of parameters t for which (56) must be fulfilled becomes a continuum. The simplest way to fulfill all of these conditions at once is to set the square bracket equal to zero for each λ that means

$$\tilde{G}_{c,c}(i\omega_\lambda) = \frac{1}{N} \sum_{\mathbf{k}} G(\mathbf{k}, i\omega_\lambda) = \frac{1}{N} \sum_{\mathbf{k}} \frac{1}{i\omega_\lambda - \frac{\varepsilon_{\mathbf{k}} - \mu}{\hbar} - \tilde{\Sigma}_{c,c}(i\omega_\lambda)}. \quad (57)$$

This equation must be solved simultaneously for each ω_λ by adjusting the hybridization function $V(E)$ and the density of states $\rho(E)$ for the ligands in the impurity model, and in fact is nothing but the well-known self-consistency equation for dynamical mean-field theory (DMFT) [13]. The derivation above shows that DMFT can be understood as approximate minimization of the grand potential as functional of the self-energy. The impurity model thereby acts as the ‘self-energy preparation lab’ to generate trial Green functions and the associated exact self-energies. So far we have partitioned the lattice into single sites and each site was ‘decorated’ with a ‘bath’ of ligands. However, we also could have partitioned the lattice into small but finite clusters and decorated each cluster with a bath of noninteracting ligand sites. Going through similar steps as above one can derive the self-consistency equations for various cluster-generalizations of DMFT [6]. The variational principle for the grand potential as a functional of the self-energy thus provides a unifying principle to derive all of these theories.

4.2 Metal-insulator-transition in a dimer

We have seen that DMFT can be derived from the variational principle for the grand potential by restricting the domain of self-energies to those which are derivable from an impurity model. However, self-energy functional theory as outlined above allows for even simpler approaches and as an example we now study Potthoff’s re-derivation of the phase diagram for the metal-insulator transition in the Hubbard model [14]. We again consider the single-band Hubbard model on a simple cubic N -site lattice with periodic boundary condition and in addition use *particle-hole symmetry* to simplify the calculation. We first note that the simple cubic lattice can be divided into two *sublattices*, A and B , such that all neighbors of a site on the A sublattice are on the B -sublattice and vice versa. For particle-hole symmetry to exist it is necessary that the hopping integral $t_{i,j} \neq 0$ only if the sites i and j belong to different sublattices and that all $t_{i,j}$ are real so $t_{i,j} = t_{j,i}$. Introducing the operator of total electron number at site i , $n_i = n_{i,\uparrow} + n_{i,\downarrow}$ and using $(n_{i,\sigma})^2 = n_{i,\sigma}$, the Hubbard Hamiltonian can be rewritten as

$$K = \sum_{i,j,\sigma} t_{i,j} c_{i,\sigma}^\dagger c_{j,\sigma} + \frac{U}{2} \sum_{i=1}^N (n_i - 1)(n_i - 1) + \left(\frac{U}{2} - \mu \right) \sum_{i=1}^N n_i - N \frac{U}{2}. \quad (58)$$

Next, we introduce the new operators $\bar{c}_{i,\sigma} = \sigma_i c_{i,\sigma}^\dagger$ and $\bar{c}_{i,\sigma}^\dagger = \sigma_i c_{i,\sigma}$ where $\sigma_i = 1$ for $i \in A$ and $\sigma_i = -1$ for $i \in B$. This is called particle-hole transformation because for example $\bar{c}_{i,\sigma}$ creates a hole rather than an electron. Accordingly the ‘vacuum state’ for the new operators is

$$|\bar{0}\rangle = \left(\prod_{i=1}^N c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger \right) |0\rangle,$$

which has $2N$ electrons. We write $\bar{n}_{i,\sigma} = \bar{c}_{i,\uparrow}^\dagger \bar{c}_{i,\uparrow}$ and similarly for all operators formed from the original operators $c_{i,\sigma}^\dagger$ and $c_{i,\sigma}$. Next, we have

$$\sum_{i,j} t_{i,j} c_{i,\sigma}^\dagger c_{j,\sigma} = \sum_{i,j} \sigma_i \sigma_j t_{i,j} \bar{c}_{i,\sigma} \bar{c}_{j,\sigma}^\dagger = \sum_{i,j} t_{j,i} \bar{c}_{j,\sigma}^\dagger \bar{c}_{i,\sigma},$$

where we have used $\sigma_i \sigma_j = -1$ for $t_{i,j} \neq 0$ and $t_{i,j} = t_{j,i}$ to arrive at the last equality. We see that the first term on the right hand side of (58) takes exactly the same form when written in terms of the $\bar{c}$ -operators as when written in terms of the c -operators. Since $n_i - 1 = 1 - \bar{n}_i$ the same holds true for the second term. This does not hold true for the third term, but this term can be simply eliminated by choosing $\mu = \frac{U}{2}$. With this choice, the entire Hamiltonian (58) is has exactly the same form when expressed in terms of either the c -operators or the $\bar{c}$ -operators. It follows that for each eigenstate $|i\rangle$ of K with n electrons there is an eigenstate $|\bar{i}\rangle$ with n holes (i.e. $2N - n$ electrons) which has exactly the same eigenvalue as $|i\rangle$. An immediate consequence is that the thermal expectation value of $\langle \hat{N} \rangle = N$.

Following Potthoff we now choose a reference system that consists of N dimers with one ‘Hubbard-site’ hybridizing with one bath-site, i.e., the Hamiltonian for one dimer reads

$$\tilde{K} = -V \sum_{\sigma} (c_{\sigma}^\dagger l_{\sigma} + l_{\sigma}^\dagger c_{\sigma}) + \left(\varepsilon_b - \frac{U}{2} \right) \sum_{\sigma} l_{\sigma}^\dagger l_{\sigma} + \frac{U}{2} (n_c - 1)(n_c - 1) - \frac{U}{2}. \quad (59)$$

Here $l_{\sigma}^\dagger$ creates an electron in the ligand site and $n_c = c_{\uparrow}^\dagger c_{\uparrow} + c_{\downarrow}^\dagger c_{\downarrow}$. Compared to the reference system (54) for DMFT which was discussed in the previous section, this amounts to retaining only a single ligand. The prefactor of the second term is $\varepsilon_b - U/2$ because $\mu = U/2$. Knowing that the Hubbard model has particle-hole symmetry we demand this also for the reference system. The transformation $c^\dagger \leftrightarrow c$, $b^\dagger \leftrightarrow -b$ then indeed converts the Hamiltonian into itself *except* for the second term. Setting $\varepsilon_b = U/2$, however, eliminates this term and particle-hole symmetry is obtained. The Hamiltonian thus becomes

$$\tilde{K} = -V \sum_{\sigma} (c_{\sigma}^\dagger l_{\sigma} + l_{\sigma}^\dagger c_{\sigma}) + \frac{U}{2} (n_c - 1)(n_c - 1) - \frac{U}{2} \quad (60)$$

and the only remaining parameter to be varied is V . The Fock space of the dimer has a dimension of $2^4 = 16$ so that all eigenstates $|i\rangle$ and corresponding eigenenergies K_i can be obtained by numerical diagonalization of the matrix for $\tilde{K}$. If we construct basis functions with fixed particle number, spin and z -component of the spin, the problem in fact can be broken down to diagonalizing 2×2 matrices. We therefore can calculate the grand potential $\tilde{\Omega}$ by numerical

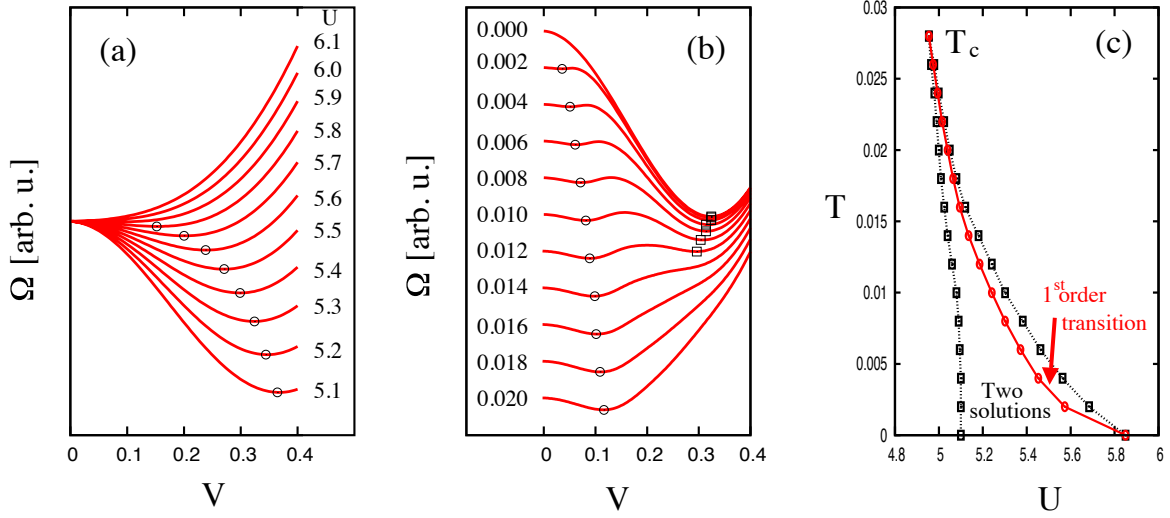


Fig. 4: (a): $\Omega(V)$ versus V at $T = 0$, variation with U , (b): Ω versus V at $U = 5.2$, variation with T , (c): the resulting phase diagram.

evaluation of the grand partition function (4), the Green function $\tilde{G}_{\alpha,\beta}(z)$ by using the Lehmann representation (8) and the self-energy $\tilde{\Sigma}_{\alpha,\beta}(z)$ from the Dyson equation (21). Again, only the entry $\tilde{\Sigma}_{c,c}(z)$ may differ from zero for any z , which is a good check for the computer program. For concreteness, at $T = 0$ the self-energy of the dimer reads [15]

$$\tilde{\Sigma}(z) = \frac{U}{2} + \frac{U^2}{8} \left(\frac{1}{z + 3V} + \frac{1}{z - 3V} \right). \quad (61)$$

Note that this has exactly the form (27), in particular the first term is indeed the Hartree-Fock potential $\langle n_\sigma \rangle U$ because $\langle n_\sigma \rangle = 1/2$ due to particle-hole symmetry. Reverting Eq. (51) we can now calculate the numerical value of the functional $F[\tilde{\Sigma}]$

$$F[\tilde{\Sigma}] = \tilde{\Omega} + \lim_{\varepsilon \rightarrow 0} \frac{1}{\beta} \sum_{\lambda} e^{i\omega_{\lambda}\varepsilon} \ln \det \tilde{\mathbf{G}}^{-1}(i\omega_{\lambda}).$$

Notice that this is possible *only* because we could obtain the numerical value of $\tilde{\Omega}$ by full exact diagonalization of $\tilde{K}$. This procedure gives us a self-energy $\tilde{\Sigma}(z)$ and the corresponding numerical value of the functional $F[\tilde{\Sigma}]$ and we recall that both of these are functions of the hybridization matrix element V in (60). Next, we proceed as above: we use the self-energy $\tilde{\Sigma}(z)$ as a trial self-energy for the lattice model and write

$$\Omega_{\text{latt}} = -\frac{2}{\beta} \sum_{\lambda} e^{i\omega_{\lambda}\varepsilon} \int d\varepsilon \rho_0(\varepsilon) \ln (i\omega_{\lambda} - \varepsilon + \mu - \tilde{\Sigma}_{c,c}(i\omega_{\lambda})) + F[\tilde{\Sigma}],$$

where $\rho_0(\varepsilon)$ is the density of states for the conduction band. Following Potthoff [14] we use a semielliptical density of states of width $W=4$

$$\rho_0(\varepsilon) = \frac{1}{2\pi} \sqrt{4 - \varepsilon^2}.$$

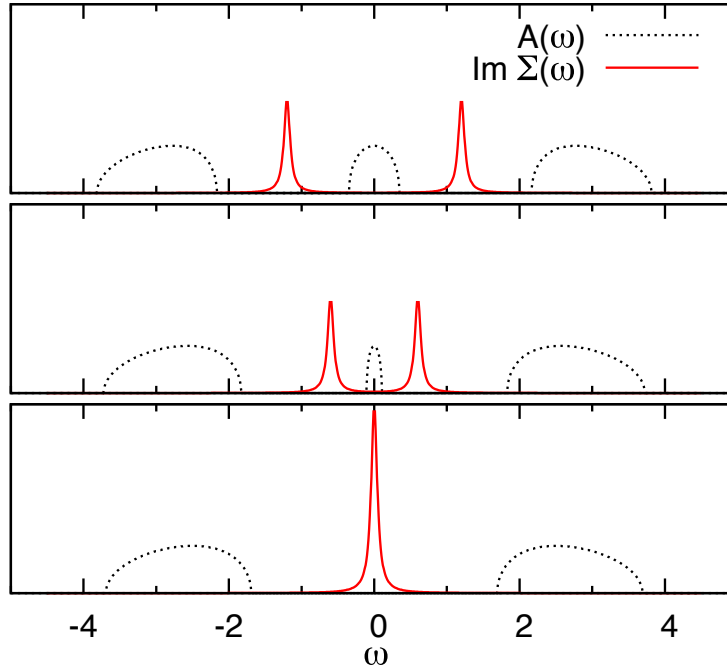


Fig. 5: Single particle spectral function and imaginary part of the self-energy (calculated with an imaginary part of 0.05 for the frequency) obtained from the angle integrated Green function (62) with the self-energy (61). Parameter values are $U = 5$ and $V = 0.4$ (top), $V = 0.2$ (center) and $V = 0$ (bottom).

In this approximation Ω becomes a function of V and Figure 4(a) shows $\Omega(V)$ at $T = 0$ for different U . For smaller U , say $U = 5.4$, there are two stationary points: a maximum at $V = 0$ and a minimum at finite V , which is the physical solution. At $U_c \approx 5.82$, however, the two extrema coalesce into a single minimum at $V = 0$, which is the only stationary point for larger U . This change from finite V to $V = 0$ precisely corresponds to the metal-insulator transition. To see this we recall equation (61) for $\tilde{\Sigma}(z)$ and insert this into the $\mathbf{k}$ -integrated Green function

$$G(z) = 2 \int_{-2}^2 d\varepsilon \frac{\rho_0(\varepsilon)}{z + \frac{U}{2} - \varepsilon - \tilde{\Sigma}(z)}. \quad (62)$$

This is shown in Figure (5) together with $\text{Im } \tilde{\Sigma}(z)$ for $z = \omega - i\eta$. The self-energy has two poles at $\zeta_1 = -3V$ and $\zeta_2 = 3V$ which give the two Lorentzian peaks in $\text{Im } \tilde{\Sigma}(z)$. As discussed above these two poles ‘push open’ gaps in the density of states around themselves, see the top panel of Figure 5. There is still a finite spectral weight at the Fermi level, $\omega = 0$, i.e., the system is a metal. As V becomes smaller, see the center panel of Figure 5, the slope of $\tilde{\Sigma}(z)$ in the energy range between ζ_1 and ζ_2 (which is $\frac{\partial \Sigma}{\partial \omega} = -\left(\frac{U}{6V}\right)^2$ at $\omega = 0$) becomes steeper and steeper so that according to (30) the spectral weight at the Fermi level becomes smaller and smaller. Finally, as $V \rightarrow 0$ the two poles coalesce, see the bottom panel of Figure 5, and there remains only a single pole at $\omega = 0$ which pushes open a gap of order U and the spectral density at $\omega = 0$ vanishes – the system has turned into a Mott-insulator! This is precisely the scenario predicted by DMFT [13], whereby DMFT calculations find $U_c \approx 5.84$ [13] as compared to $U_c \approx 5.82$ from the simple dimer calculation.

Next, Figure 4(b) shows $\Omega(V)$ for the fixed value of $U = 5.2$ and different temperatures T . For most temperatures there are three stationary points whereby the local maximum can be discarded. It follows, that there are actually two possible solutions for each temperature which cross in between $T = 0.10$ and $T = 0.12$. This implies that there is a 1st order phase transition between these two temperatures. Repeating the procedure for various U gives the phase diagram in Figure 4(c). There is only a metallic solution for small U , at a first U_{c1} a second insulating solution starts to appear, at U_c there is a first order metal-insulator transition and from U_{c2} on there is only an insulating solution. The results obtained in this way by the solution of a dimer are qualitatively very similar to those obtained by extensive numerical renormalization group [16] and quantum Monte Carlo [17] calculations in the framework of DMFT. The main deficiency of the dimer calculation is the underestimation of the critical temperature T_c in Figure 4(c) by a factor of ≈ 2 .

This application of self-energy functional theory, whereby trial self-energies $\Sigma(z)$ and the numerical value of Ω and hence $F[\Sigma]$ are obtained by exact diagonalization of a suitably chosen reference system is called the *Variational Cluster Approximation*. This was first proposed by Potthoff and successfully applied to a wide variety of problems.

5 Summary and conclusion

We have discussed the single-particle Green function $G(z)$ which describes the propagation of a particle added to a system of interacting fermions. It is a function of the complex frequency variable z which is analytical throughout the complex z -plane except for the real axis. On the real axis it has poles which give the energies needed to add or remove a particle. Via the Dyson equation it can be related to the self-energy $\Sigma(z)$ which describes the effects of the interaction with the propagating particle with the other particles in the system. We have derived the spectral representations Eq. (11) and Eq. (27) for these quantities. We have then given a nonperturbative derivation of the results of Luttinger and Ward: the grand canonical potential Ω is a functional of the Green function and self-energy (see Eq. (49), the self-energy is the derivative of the Luttinger-Ward functional with respect to the Green function (see Eq. (50)) and Ω is stationary with respect to variations of the self-energy (see Eq. (53)). We have then seen that Ω can also be expressed as a functional of the self-energy alone (see Eq. (51)) which is stationary with respect to variations of the self-energy. This then is the basis of self-energy functional theory which provides a simple unifying framework for Dynamical Mean-Field Theory and its cluster generalizations and which also can be applied directly in the framework of the Variational Cluster Approximation.

Appendix A

We want to prove the formula

$$-i\Theta(t) e^{-i\omega_0 t} = \lim_{\eta \rightarrow 0^+} \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega t}}{\omega - \omega_0 + i\eta}. \quad (63)$$

We consider the integral on the right-hand-side as a contour integral in the complex ω -plane along the real axis. We initially limit the range of integration to be $\int_{-R}^R d\omega \dots$ and take the limit $R \rightarrow \infty$ in the end. We augment this contour by a semicircle of radius R around the origin so that the resulting contour is closed. For any ω on the semicircle we have

$$e^{-i\omega t} = e^{\sin(\varphi)Rt} e^{-i\cos(\varphi)Rt}$$

where φ is the phase of ω . Assume that $t \sin(\varphi) < 0$. Then, for any $\epsilon > 0$ we can find an R such that $R e^{\sin(\varphi)Rt} < \epsilon$. It follows that the contour integral over the part of the semicircle where $t \sin(\varphi) < 0$ vanishes in the limit $R \rightarrow \infty$. For $t > 0$ this means that the integral over the semicircle in the lower half-plane $\pi < \varphi < 2\pi$ vanishes, whereas it is the semicircle in the upper half-plane $0 < \varphi < \pi$ which gives vanishing contribution for $t < 0$. We can thus replace the integral on the right-hand-side by an integral along a *closed* contour which consists of the real axis and an infinitely large semicircle in the lower (upper) half-plane for $t > 0$ ($t < 0$). But this means we can invoke the *theorem of residues* which tells us that the integral is $\pm 2\pi i$ times the sum of residues of all poles within the integration contour. The integrand has one pole at $\omega_1 = \omega_0 - i\eta$ with residuum $e^{-i\omega_1 t}$. This is in the lower half-plane so that we know immediately that for $t < 0$ the integral vanishes i.e., it is $\propto \Theta(t)$. For $t > 0$ we readily find that the integral is $-2\pi i e^{-i\omega_1 t}$ and inserting this into (63) and taking the limit $\eta \rightarrow 0$ proves (63), which shows that the Fourier transform of the left-hand side is $(\omega - \omega_0 + i\eta)^{-1}$.

Appendix B

Call $\lambda = e^{i\frac{2m\pi}{M}}$ where m and M are integers. Then

$$\sum_{k=1}^M \lambda^k = \lambda \sum_{k=0}^{M-1} \lambda^k = \lambda \frac{\lambda^M - 1}{\lambda - 1},$$

but since $\lambda^M = (e^{2\pi i})^m$ this is zero. Defining M -component vectors $\vec{\varphi}_\gamma^* = (\varphi_{\gamma,1}^*, \varphi_{\gamma,2}^*, \dots, \varphi_{\gamma,M}^*)$ and $\vec{v}_\nu = (e^{-i\omega_\nu \tau_1}, e^{-i\omega_\nu \tau_2}, \dots, e^{-i\omega_\nu \tau_M}) / \sqrt{M}$ the first of Equations (35) can be written as $\varphi_{\gamma,\nu}^* = \vec{v}_\nu^* \cdot \vec{\varphi}_\gamma^*$. Introducing $\lambda = e^{2i\frac{\mu-\nu}{M}\pi}$ one has, using $\omega_\nu = (2\nu+1)\pi/\hbar\beta$ and $\tau_k = \hbar\beta k/M$,

$$\vec{v}_\nu^* \cdot \vec{v}_\mu = \frac{1}{M} \sum_{k=1}^M e^{i(\omega_\mu - \omega_\nu)\tau_k} = \frac{1}{M} \sum_{k=1}^M \lambda^k = \begin{cases} 1 & \text{for } \mu - \nu = nM, \\ \lambda \frac{\lambda^M - 1}{\lambda - 1} = 0 & \text{for } \mu - \nu \neq nM. \end{cases} \quad (64)$$

with integer n . The M vectors $\vec{v}_\nu$ with $-M/2 \leq \nu \leq M/2 - 1$ therefore are orthonormal. The matrix T , which transforms the M variables φ_k^* into the M variables $\tilde{\varphi}_\nu^*$, $\vec{\tilde{\varphi}}_\nu^* = T \vec{\varphi}_k^*$, therefore

has the $\vec{v}_\mu$ as its rows, whence $T^\dagger = T^{-1}$ and the transformation is unitary. Eq. (36) follows immediately.

Let us now consider the limit of continuous time, i.e., $M \rightarrow \infty$, so that $\epsilon = \hbar\beta/M \rightarrow 0$ and $\tau_k = k\epsilon$ become continuous. Then Eqs. (35) can be written as

$$\begin{aligned}\varphi_{\gamma,\nu}^* &= \sqrt{\epsilon} \tilde{\varphi}_{\gamma,\nu}^* = \sqrt{\frac{1}{\hbar\beta}} \sum_{k=1}^M e^{-i\omega_\nu \tau_k} \varphi_{\gamma,k}^* \epsilon \rightarrow \sqrt{\frac{1}{\hbar\beta}} \int_0^{\hbar\beta} e^{-i\omega_\nu \tau} \varphi_\gamma^*(\tau) d\tau \\ \varphi_\gamma^*(\tau) &= \frac{1}{\sqrt{\hbar\beta}} \sum_{\nu=-\infty}^{\infty} e^{i\omega_\nu \tau} \varphi_{\gamma,\nu}^*\end{aligned}\quad (65)$$

If $\varphi_\gamma^*(\tau)$ were a function of order unity the integral would be of order $\hbar\beta$, i.e., independent of M . Accordingly we need to modify the definitions as above when taking the continuum limit, $M \rightarrow \infty$.

Appendix C

To get a better understanding of the factors of $e^{i\omega_\nu \epsilon}$ we calculate the noninteracting grand potential from the Fourier transformed version of S (again for only a single γ)

$$\begin{aligned}Z_M &= \int \prod_{\nu=-M/2}^{M/2-1} d\tilde{\varphi}_\nu^* d\tilde{\varphi}_\nu e^{-S[\tilde{\varphi}_\nu^*, \tilde{\varphi}_\nu]}, \\ -S[\tilde{\varphi}_\nu^*, \tilde{\varphi}_\nu] &= \sum_{\nu=-M/2}^{M/2-1} (1 - e^{i\omega_\nu \epsilon} (1 - \epsilon\zeta)) \tilde{\varphi}_\nu \tilde{\varphi}_\nu^*.\end{aligned}$$

The Grassmann integration is trivial and we get the partition function and the grand potential

$$\begin{aligned}Z_M &= \prod_{\nu=-M/2}^{M/2-1} (1 - e^{i\omega_\nu \epsilon} (1 - \epsilon\zeta)) \\ \Omega &= - \lim_{\epsilon \rightarrow 0^+} \frac{1}{\beta} \sum_{\nu=-\infty}^{\infty} \log (1 - e^{i\omega_\nu \epsilon} (1 - \epsilon\zeta)) = - \lim_{\epsilon \rightarrow 0^+} \frac{1}{\beta} \sum_{\nu=-\infty}^{\infty} g(i\omega_\nu),\end{aligned}$$

where $g(z) = \log(1 - e^{\epsilon z} (1 - \epsilon\zeta))$. When viewed as a function of the complex variable z , $g(z)$ is analytical in the whole complex plane except for the real axis. This is because $\log(z)$ is analytical except for the negative real axis where it has a branch cut. Now we use a standard trick from quantum field theory. Since $i\beta\hbar\omega_\nu = (2\nu+1)\pi$, the Fermi function $f(z) = (e^{\beta\hbar z} + 1)^{-1}$ has poles at all $z_\nu = i\omega_\nu$. Setting $z = z_\nu + \delta$ we find $f(z) = -\frac{1}{\beta\hbar\delta} + \dots$ so that the residuum is $-\frac{1}{\beta\hbar}$ at each z_ν . Invoking the theorem of residues we can therefore rewrite

$$\Omega = \lim_{\epsilon \rightarrow 0^+} \frac{\hbar}{2\pi i} \oint_{\mathcal{C}} dz f(z) g(z),$$

where $\mathcal{C}$ is the contour shown in Figure 6(a), which encircles all z_ν in counterclockwise fashion. Next, the integrals along each of the two clover-shaped contours in Figure 6(b) are zero, *provided* the integrand is analytic in the interior of the two curves. This is certainly true because the

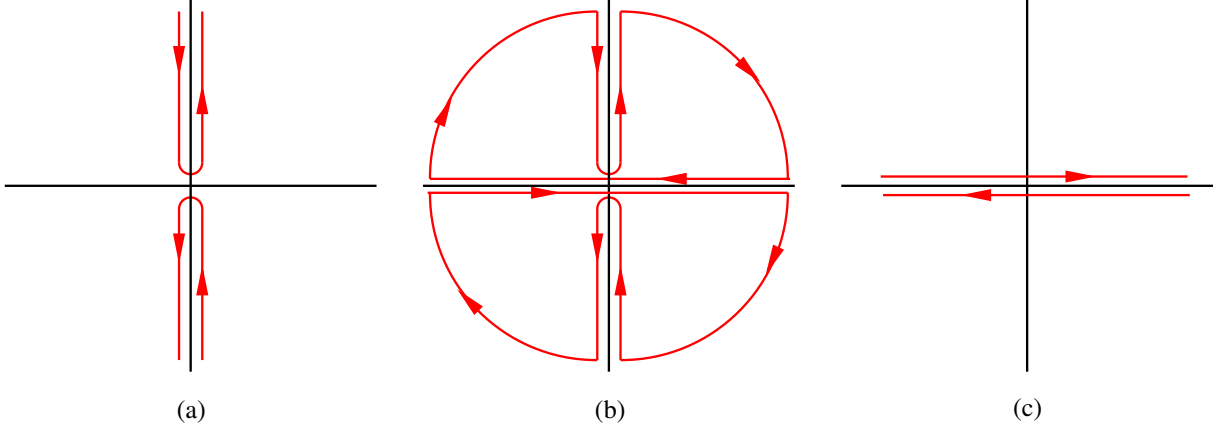


Fig. 6: Integration contours used in Appendix C.

poles of $f(z)$ are on the imaginary axis, whereas $g(z)$ is singular only on the real axis and both are outside the curves in Figure 6(b). For large positive $\text{Re}(z)$ we have $f(z) \propto e^{-\beta\hbar\text{Re}(z)}$, so that the contribution from the large semicircle with $\text{Re}(z) > 0$ vanishes. For $\text{Re}(z) < 0$, on the other hand, $e^{\epsilon z} \propto e^{-\epsilon|\text{Re}(z)|} \rightarrow 0$ so that $g(z) \propto -e^{-\epsilon|\text{Re}(z)|} (1-\epsilon\zeta)$ and the contribution from the large semicircle with $\text{Re}(z) < 0$ also vanishes. It follows that the integral along the contour $\mathcal{C}$ in Figure 6(a) is equal to that along the contour $\mathcal{C}'$ in Figure 6(c) (note the inverted direction of the curves in Figure 6(c) as compared to Figure 6(b)!). It is therefore precisely the factor of $e^{i\omega\nu\epsilon}$ which allows to convert the sum over Matsubara frequencies to an integral around the real axis. Parameterizing the two parts of the contour in Figure 6(c) by $z = \omega \pm i\eta$ we find

$$\Omega = \lim_{\eta \rightarrow 0^+} \lim_{\epsilon \rightarrow 0^+} \frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega [f(\omega+i\eta)g(\omega+i\eta) - f(\omega-i\eta)g(\omega-i\eta)].$$

The square bracket is the discontinuity of the function $f(z)g(z)$ across the real axis. The Fermi function is continuous, but the function $g(z)$ has a discontinuity of $2\pi i$ across the real axis when the argument of the logarithm is negative, the well-known branch cut of the logarithm along the negative real axis. Consider $g(z)$ for small ϵ

$$g(z) = \log(1 - e^{\epsilon z}(1-\epsilon\zeta)) \approx \log(1 - (1+\epsilon z)(1-\epsilon\zeta)) = \log \epsilon + \log(\zeta - z).$$

In forming $g(\omega+i\eta) - g(\omega-i\eta)$ the term $\log \epsilon$ drops out so

$$\begin{aligned} \Omega &= \lim_{\eta \rightarrow 0^+} \frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega f(\omega) (\log(\zeta - \omega - i\eta) - \log(\zeta - \omega + i\eta)) \\ &= -\hbar \int_{-\infty}^{\infty} d\omega f(\omega) \Theta(\omega - \zeta) = -\hbar \int_{\zeta}^{\infty} d\omega f(\omega). \end{aligned}$$

Using the antiderivative of the Fermi function

$$f(\omega) = -\frac{1}{\beta\hbar} \frac{d}{d\omega} \log(1 + e^{-\beta\hbar\omega}),$$

reinserting the state label γ and recalling $\zeta_\gamma = (E_\gamma - \mu)/\hbar$ we thus obtain the correct result

$$\Omega = -\frac{1}{\beta} \sum_{\gamma} \log(1 + e^{-\beta\hbar\zeta_\gamma}) = -\frac{1}{\beta} \sum_{\gamma} \log(1 + e^{-\beta(E_\gamma - \mu)}).$$

Now let us use the expression (38) for S which was obtained by the somewhat questionable approximation (37). Setting $V_{\alpha\beta,\delta,\gamma} = 0$ and again dropping the state label γ we now find

$$-S[\tilde{\varphi}_\nu^*, \tilde{\varphi}_\nu] = \sum_{\nu=-M/2}^{M/2-1} \tilde{\varphi}_\nu^* e^{i\omega_\nu \epsilon} (-G_0^{-1}(i\omega_\nu)) \tilde{\varphi}_\nu \tilde{\varphi}_\nu^*$$

with $G_0^{-1}(i\omega_\nu) = i\omega_\nu - \zeta$. Again converting the sum over Matsubara frequencies to a contour integral we obtain

$$\Omega = \lim_{\epsilon \rightarrow 0^+} \frac{\hbar}{2\pi i} \oint_{\mathcal{C}} dz f(z) e^{\epsilon z} \log(\zeta - z).$$

Considering again the contour in Figure 6(b), the Fermi function guarantees that the arcs with $\text{Re}(z) > 0$ give no contribution, whereas now the factor of $e^{\epsilon z}$ now does the same for the arcs with $\text{Re}(z) < 0$. We can again change the integration contour to that in Figure 6(c) and arrive at the correct for Ω . Bearing in mind the role played by the factors of $e^{\epsilon z}$, namely to determine how integration contours can be deformed, the approximation (37) thus is obviously perfectly reasonable.

Appendix D

Here we prove the identity

$$\frac{\partial \ln \det A}{\partial A_{ij}} = A_{ji}^{-1}$$

We use Laplace's formula and expand $\det A$ in terms of minors

$$\det A = \sum_{l=1,n} (-1)^{i+l} A_{il} M_{il}$$

Since none of the minors M_{il} contains the element A_{ij} we find

$$\frac{\partial \ln \det A}{\partial A_{ij}} = \frac{(-1)^{i+j} M_{ij}}{\det A}$$

Next, the i^{th} column of A^{-1} is the solution of the system of equations $Ac = E_i$ where E_i is the i^{th} column of the unit matrix. It has all elements equal to zero, except for the i^{th} , which is one. We use Cramer's rule and find for the j^{th} element of the i^{th} column

$$A_{ji}^{-1} = \det \bar{A}_j / \det A$$

where $\bar{A}_j$ is the matrix where the j^{th} column has been replaced by E_i . Now we use again Laplace's formula for $\det \bar{A}_j$ and obtain

$$A_{ji}^{-1} = (-1)^{i+j} M_{ij} / \det A$$

which proves the theorem. As an application we assume that the matrix elements of A are functions of some parameter α . Then we find

$$\frac{\partial \ln \det A}{\partial \alpha} = \sum_{i,j} \frac{\partial \ln \det A}{\partial A_{ij}} \frac{\partial A_{ij}}{\partial \alpha} = \sum_{i,j} A_{ji}^{-1} \frac{\partial A_{ij}}{\partial \alpha} = \text{Tr } A^{-1} \frac{\partial A}{\partial \alpha}.$$

References

- [1] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960)
- [2] J.M. Luttinger, Phys. Rev. **119**, 1153 (1960)
- [3] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)
- [4] G. Baym, Phys. Rev. **127**, 1391 (1962)
- [5] M. Potthoff, Condens. Mat. Phys. **9**, 557 (2006)
- [6] M. Potthoff in A. Avella and F. Mancini (eds.): *Theoretical Methods for Strongly Correlated Systems* (Springer, 2011); see also arXiv:11082183
- [7] A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinski: *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey 1964)
- [8] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems* (McGraw-Hill, San Francisco, 1971)
- [9] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Addison-Wesley, Redwood, 1988)
- [10] G. Baym and N.D. Mermin, J. Math. Phys. **2**, 232 (1961)
- [11] I. Dzyaloshinskii, Phys. Rev. B **68**, 085113 (2003)
- [12] J.M. Luttinger, Phys. Rev. **121**, 942 (1961)
- [13] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [14] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003)
- [15] E. Lange, Mod. Phys. Lett. B **12**, 915 (1998)
- [16] R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64** 045103 (2001)
- [17] J. Joo and V. Oudovenko, Phys. Rev. B **64** 103102 (2001)

4 Exact Diagonalization and Lanczos Method

Erik Koch

Jülich Supercomputing Centre

Forschungszentrum Jülich

Contents

1	Many-electron states	2
2	Second quantization	5
2.1	Creation and annihilation operators	5
2.2	Representation of Slater determinants	6
2.3	Representation of n -body operators	7
2.4	Transforming the orbital basis	11
3	Exact diagonalization	12
3.1	Variational principles	12
3.2	Matrix eigenvalue problem	14
3.3	Dimension of the Hilbert space and sparseness	15
4	Lanczos Method	17
4.1	Krylov space	19
4.2	Spectral functions	23
A	Appendices	25
A.1	Non-orthonormal basis	25
A.2	Useful commutation relations	26
A.3	Downfolding	27

1 Many-electron states

One of the great surprises of quantum mechanics is the existence of indistinguishable objects. Classically this is not possible: objects can always be distinguished at least by their position in space, meaning that indistinguishable objects must be identical. This is Leibniz' Principle of the Identity of Indiscernibles [1]. For quantum objects, however, the uncertainty principle makes the distinction of particles by their position impossible. This allows for the existence of elementary particles. They form the basic units of all matter. So, quite remarkably, all the different objects we know are made of indistinguishable building blocks.

In the formalism of quantum mechanics, indistinguishability means that no observable lets us distinguish one of these particles from the other. This means that every observable for, e.g., electrons, must treat each electron in the same way. Thus, in principle, observables must act on *all* electrons in the universe. In practice we can, of course, distinguish electrons localized on the moon from those in our lab to an excellent approximation. Thus, for all practical purposes, we can restrict our description to the electrons in the system under consideration, assuming that the differential overlap with all other electrons vanishes. Any observable $M(x_1, \dots, x_N)$ for the N electrons in our system must then be symmetric under permutations of the variables x_i .

The consequences are straightforward: An observable $M(x)$ acting on a single-particle degree of freedom x must act on all indistinguishable particles in the same way, i.e., $\sum_i M(x_i)$. Likewise, a two-body observable $M(x, x')$ must act on all pairs in the same way, $\sum_{i,j} M(x_i, x_j)$ with $M(x, x') = M(x', x)$. We can thus write any observable in the form

$$M(\mathbf{x}) = M^{(0)} + \sum_i M^{(1)}(x_i) + \frac{1}{2!} \sum_{i \neq j} M^{(2)}(x_i, x_j) + \frac{1}{3!} \sum_{i \neq j \neq k} M^{(3)}(x_i, x_j, x_k) + \dots \quad (1)$$

$$= M^{(0)} + \sum_i M^{(1)}(x_i) + \sum_{i < j} M^{(2)}(x_i, x_j) + \sum_{i < j < k} M^{(3)}(x_i, x_j, x_k) + \dots, \quad (2)$$

where the summations can be restricted since the operators must be symmetric in their arguments, while for two or more identical coordinates the operator is really one of lower order: $M^{(2)}(x_i, x_i)$, e.g., only acts on a single coordinate and should be included in $M^{(1)}$.

For the many-body wavefunctions $\Psi(x_1, x_2, \dots)$ the situation is slightly more complex. Since the probability density $|\Psi(x_1, x_2, \dots)|^2$ is an observable, the wavefunction should transform as one-dimensional (irreducible) representations of the permutation group. Which irreducible representation applies to a given type of elementary particle is determined by the spin-statistics theorem [2,3]: The wavefunctions of particles with integer spin are symmetric, those of particles with half-integer spin change sign when two arguments are exchanged. From an arbitrary N -particle wavefunction we thus obtain a many-electron wavefunction by antisymmetrizing

$$\mathcal{A}\Psi(x_1, \dots, x_N) := \frac{1}{\sqrt{N!}} \sum_P (-1)^P \Psi(x_{p(1)}, \dots, x_{p(N)}) , \quad (3)$$

where $(-1)^P$ is the parity of the permutation P that maps $n \rightarrow p(n)$. Since there are $N!$ different permutations, this quickly becomes an extremely costly operation. Remarkably, a product of

N single-electron states φ_α can be antisymmetrized much more efficiently (in $\mathcal{O}(N^3)$ steps) by writing it in the form of a determinant

$$\Phi_{\alpha_1, \dots, \alpha_N}(x_1, \dots, x_N) := \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_{\alpha_1}(x_1) & \varphi_{\alpha_2}(x_1) & \cdots & \varphi_{\alpha_N}(x_1) \\ \varphi_{\alpha_1}(x_2) & \varphi_{\alpha_2}(x_2) & \cdots & \varphi_{\alpha_N}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{\alpha_1}(x_N) & \varphi_{\alpha_2}(x_N) & \cdots & \varphi_{\alpha_N}(x_N) \end{vmatrix}. \quad (4)$$

For $N=1$ the Slater determinant is simply the one-electron orbital $\Phi_\alpha(x) = \varphi_\alpha(x)$ while for $N=2$ we get the familiar expression $\Phi_{\alpha, \alpha'}(x, x') = (\varphi_\alpha(x)\varphi_{\alpha'}(x') - \varphi_{\alpha'}(x)\varphi_\alpha(x'))/\sqrt{2}$ for the two-electron Slater determinant.

Slater determinants are important because they can be used to build a basis of the many-electron Hilbert space. To see how, we consider a complete set of orthonormal single-electron states

$$\sum_n \overline{\varphi_n(x)} \varphi_n(x') = \delta(x-x') \text{ (complete)} \quad \int dx \overline{\varphi_n(x)} \varphi_m(x) = \delta_{n,m} \text{ (orthonormal)}. \quad (5)$$

To expand an arbitrary N -particle function $a(x_1, \dots, x_N)$, we start by considering it as a function of x_1 with $x_2, \dots, x_N$ kept fixed. We can then expand it in the complete set $\{\varphi_n\}$ as

$$a(x_1, \dots, x_N) = \sum_{n_1} a_{n_1}(x_2, \dots, x_N) \varphi_{n_1}(x_1)$$

with expansion coefficients that depend on the remaining coordinates

$$a_{n_1}(x_2, \dots, x_N) = \int dx_1 \overline{\varphi_{n_1}(x_1)} a(x_1, x_2, \dots, x_N).$$

These, in turn, can be expanded as a functions of x_2

$$a_{n_1}(x_2, \dots, x_N) = \sum_{n_2} a_{n_1, n_2}(x_3, \dots, x_N) \varphi_{n_2}(x_2).$$

Repeating this, we obtain the expansion of a in product states

$$a(x_1, \dots, x_N) = \sum_{n_1, \dots, n_N} a_{n_1, \dots, n_N} \varphi_{n_1}(x_1) \cdots \varphi_{n_N}(x_N)$$

with

$$a_{n_1, \dots, n_N} = \int dx_1 \cdots \int dx_N \overline{\varphi_{n_1}(x_1)} \cdots \overline{\varphi_{n_N}(x_N)} a(x_1, \dots, x_N).$$

For an antisymmetric N -particle function Ψ the expansion coefficients are thus antisymmetric under permutation of the indices: $a_{n_{p(1)}, \dots, n_{p(N)}} = (-1)^P a_{n_1, \dots, n_N}$. Fixing some particular order of the indices, e.g., $n_1 < n_2 < \dots < n_N$, we thus get an expansion in Slater determinants

$$\Psi(x_1, \dots, x_N) = \sum_{n_1 < \dots < n_N} a_{n_1, \dots, n_N} \sqrt{N!} \Phi_{n_1, \dots, n_N}(x_1, \dots, x_N).$$

Since we can write any antisymmetric function as such a configuration-interaction expansion, the set of Slater determinants

$$\left\{ \Phi_{n_1, \dots, n_N}(x_1, \dots, x_N) \mid n_1 < n_2 < \dots < n_N \right\} \quad (6)$$

forms a basis of the N -electron Hilbert space. Since the overlap of two Slater determinants

$$\begin{aligned} \int d\mathbf{x} \overline{\Phi_{\alpha_1, \dots, \alpha_N}(\mathbf{x})} \Phi_{\beta_1, \dots, \beta_N}(\mathbf{x}) &= \frac{1}{N!} \sum_{P, P'} (-1)^{P+P'} \prod_n \int dx_n \overline{\varphi_{\alpha_{p(n)}}(x_n)} \varphi_{\alpha_{p'(n)}}(x_n) \\ &= \begin{vmatrix} \langle \varphi_{\alpha_1} | \varphi_{\beta_1} \rangle & \cdots & \langle \varphi_{\alpha_1} | \varphi_{\beta_N} \rangle \\ \vdots & \ddots & \vdots \\ \langle \varphi_{\alpha_N} | \varphi_{\beta_1} \rangle & \cdots & \langle \varphi_{\alpha_N} | \varphi_{\beta_N} \rangle \end{vmatrix} \end{aligned} \quad (7)$$

is the determinant of the overlap of the constituent orbitals, the Slater determinants (6) form a complete orthonormal basis of the N -electron Hilbert space when the orbitals $\varphi_n(x)$ are a complete orthonormal basis of the one-electron Hilbert space.

While we use a set of N one-electron orbitals $\varphi_n(x)$ to define an N -electron Slater determinant $\Phi_{\alpha_1, \dots, \alpha_N}(\mathbf{x})$, this representation is not unique: Any unitary transformation among the N occupied orbitals will not change the determinant. Thus, strictly, a Slater determinant is not determined by the set of indices we usually give, but, up to a phase, by the N -dimensional subspace spanned by the orbitals $\varphi_1, \dots, \varphi_N$ in the single-electron Hilbert space. The projector to this space is the one-body density matrix

$$\Gamma^{(1)}(x, x') = N \int dx_2 \cdots dx_N \overline{\Phi(x, x_2, \dots, x_N)} \Phi(x', x_2, \dots, x_N). \quad (8)$$

To see this, we expand the Slater determinant along its first row

$$\Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N) = \frac{1}{\sqrt{N}} \sum_{n=1}^N (-1)^{1+n} \varphi_{\alpha_n}(x_1) \Phi_{\alpha_{i \neq n}}(x_2, \dots, x_N), \quad (9)$$

where $\Phi_{\alpha_{i \neq n}}(x_2, \dots, x_N)$ is the determinant with the first row and the n -th column removed, which can be written as $N-1$ -electron Slater determinants with orbital α_n removed. Inserting this into (8) we find

$$\Gamma_{\Phi}^{(1)}(x, x') = \sum_{n=1}^N \overline{\varphi_{\alpha_n}(x)} \varphi_{\alpha_n}(x'), \quad (10)$$

which is the expansion of the one-body density matrix in its eigenfunctions (natural orbitals), with eigenvalues (natural occupation numbers) either one or zero. Any many-electron wavefunction $\Psi(\mathbf{x})$ with the same one-body density matrix $\Gamma_{\Phi}^{(1)}$ equals the Slater determinant $\Phi(\mathbf{x})$ up to a phase, i.e., $|\langle \Psi | \Phi \rangle| = 1$.

We can generalize this procedure and calculate higher order density matrices by introducing the generalized Laplace expansion

$$\Phi_{\alpha_1 \dots \alpha_N}(\mathbf{x}) = \binom{N}{p}^{-1/2} \sum_{n_1 < \dots < n_p} (-1)^{1+\sum_i n_i} \Phi_{\alpha_{n_1} \dots \alpha_{n_p}}(x_1, \dots, x_p) \Phi_{\alpha_{i \notin \{n_1, \dots, n_p\}}}(x_{p+1}, \dots, x_N),$$

which is obtained by writing the permutation of all N indices as a permutation of $N-p$ indices and the remaining p indices separately, summing over all distinct sets of p indices. This allows us to evaluate arbitrary matrix elements and higher order density matrices [4]. But as can be seen from the above expansion, the expressions very quickly get quite cumbersome. Fortunately there is a representation that is much better suited to handling antisymmetric wavefunctions. It is called second quantization.

2 Second quantization

Second quantization for fermions [5, 6] can be viewed as a generalization of the Dirac notation to many-electron states. The idea is to split the wavefunction into coordinate and state, $\varphi_\alpha(x) = \langle x | \alpha \rangle$, and absorb the coordinate in the operators, rewriting expectation values as

$$\int dx \overline{\varphi_\alpha(x)} M(x) \varphi_\beta(x') = \langle \alpha | \int dx | x \rangle M(x) \langle x | \beta \rangle =: \langle \alpha | \hat{M} | \beta \rangle.$$

For many-electron systems the coordinates are represented by the field-operators $\hat{\Psi}(x)$. To implement antisymmetry, they need to anticommute $\hat{\Psi}(x)\hat{\Psi}(x') = -\hat{\Psi}(x')\hat{\Psi}(x)$, which implies the Pauli principle $\hat{\Psi}(x)^2 = 0$. The operators $\hat{\Psi}(x)$ can be thought of as removing an electron with coordinates $x=(\mathbf{r}, \sigma)$ from the system. Applying them on a state with no electrons thus gives zero. The *entire* formalism then follows from the following set of properties of the vacuum state $|0\rangle$ and the anticommutation relations of the field operators and their adjoints:

$$\boxed{\begin{array}{ll} \hat{\Psi}(x)|0\rangle = 0 & \{\hat{\Psi}(x), \hat{\Psi}(x')\} = 0 \\ \langle 0|0\rangle = 1 & \{\hat{\Psi}(x), \hat{\Psi}^\dagger(x')\} = \delta(x-x') \end{array}} \quad (11)$$

where the curly braces denote the anti-commutator $\{A, B\} := AB + BA$.

Defining the (Hermitian) particle number operator

$$\hat{N} := \int dx \hat{\Psi}^\dagger(x) \hat{\Psi}(x) \quad (12)$$

we can verify that $\hat{N}|0\rangle = 0$, i.e., the vacuum contains no particle, and

$$\hat{N}\hat{\Psi}(x) = -\int dx' \hat{\Psi}^\dagger(x') \hat{\Psi}(x) \hat{\Psi}(x') = \int dx' \left(\hat{\Psi}(x) \hat{\Psi}^\dagger(x') - \delta(x-x') \right) \hat{\Psi}(x') = \hat{\Psi}(x) (\hat{N}-1), \quad (13)$$

i.e., applying $\hat{\Psi}(x)$ on a state reduces the particle number by one. Taking the adjoint of (13) it then follows that $\hat{\Psi}^\dagger(x)$ creates a particle.

2.1 Creation and annihilation operators

For a single-electron state $\varphi_\alpha(x)$ we can define the operator

$$c_\alpha^\dagger := \int dx \varphi_\alpha(x) \hat{\Psi}^\dagger(x), \quad (14)$$

that inserts an electron at position x weighted with the amplitude $\varphi_\alpha(x)$, hence called creation operator for state $|\alpha\rangle$. The anticommutator with the field annihilator just gives back the single-electron wavefunction

$$\{\hat{\Psi}(x), c_\alpha^\dagger\} = \int dx' \varphi_\alpha(x') \{\hat{\Psi}(x), \hat{\Psi}^\dagger(x')\} = \varphi_\alpha(x). \quad (15)$$

Thus, together with their adjoints, the annihilation operators,

$$c_\alpha := \int dx \overline{\varphi_\alpha(x)} \hat{\Psi}(x), \quad (16)$$

they form the algebra

$$\begin{aligned} c_\alpha |0\rangle &= 0 & \{c_\alpha, c_\beta\} &= 0 = \{c_\alpha^\dagger, c_\beta^\dagger\} \\ \langle 0|0\rangle &= 1 & \{c_\alpha, c_\beta^\dagger\} &= \langle \alpha|\beta\rangle \end{aligned} \quad (17)$$

For simplicity one usually works with orthonormal orbitals, $\langle \varphi_n | \varphi_m \rangle = \delta_{n,m}$

2.2 Representation of Slater determinants

We have now all the tools in place to write the Slater determinant (4) in second quantization, using the creation operators to specify the occupied orbitals and the field operators to give the coordinates for the real-space representation:

$$\Phi_{\alpha_1 \alpha_2 \dots \alpha_N}(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{N!}} \langle 0 | \hat{\Psi}(x_1) \hat{\Psi}(x_2) \dots \hat{\Psi}(x_N) c_{\alpha_N}^\dagger \dots c_{\alpha_2}^\dagger c_{\alpha_1}^\dagger | 0 \rangle. \quad (18)$$

Note how writing the Slater determinant as an expectation value of annihilation and creation operators nicely separates the coordinates on the left from the orbitals on the right. This is just the desired generalization of the Dirac notation $\varphi(x) = \langle x | \varphi \rangle$.

Not surprisingly, the proof of (18) is by induction. As a warm-up we consider the case of a single-electron wavefunction ($N = 1$). Using the anticommutation relation (15), we see that

$$\langle 0 | \hat{\Psi}(x_1) c_{\alpha_1}^\dagger | 0 \rangle = \langle 0 | \varphi_{\alpha_1}(x_1) - c_{\alpha_1}^\dagger \hat{\Psi}(x_1) | 0 \rangle = \varphi_{\alpha_1}(x_1). \quad (19)$$

For the two-electron state $N = 2$, we anticommute $\hat{\Psi}(x_2)$ in two steps to the right

$$\begin{aligned} \langle 0 | \hat{\Psi}(x_1) \hat{\Psi}(x_2) c_{\alpha_2}^\dagger c_{\alpha_1}^\dagger | 0 \rangle &= \langle 0 | \hat{\Psi}(x_1) (\varphi_{\alpha_2}(x_2) - c_{\alpha_2}^\dagger \hat{\Psi}(x_2)) c_{\alpha_1}^\dagger | 0 \rangle \\ &= \langle 0 | \hat{\Psi}(x_1) c_{\alpha_1}^\dagger | 0 \rangle \varphi_{\alpha_2}(x_2) - \langle 0 | \hat{\Psi}(x_1) c_{\alpha_2}^\dagger \hat{\Psi}(x_2) c_{\alpha_1}^\dagger | 0 \rangle \\ &= \varphi_{\alpha_1}(x_1) \varphi_{\alpha_2}(x_2) - \varphi_{\alpha_2}(x_1) \varphi_{\alpha_1}(x_2). \end{aligned} \quad (20)$$

We see how anticommuting automatically produces the appropriate signs for the antisymmetric wavefunction. Dividing by $\sqrt{2}$, we obtain the desired two-electron Slater determinant.

The general case of an N -electron state works just the same. Anti-commuting $\hat{\Psi}(x_N)$ all the way to the right produces $N-1$ terms with alternating sign

$$\begin{aligned} \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) \hat{\Psi}(x_N) c_{\alpha_N}^\dagger c_{\alpha_{N-1}}^\dagger \dots c_{\alpha_1}^\dagger | 0 \rangle &= \\ &+ \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) c_{\alpha_{N-1}}^\dagger \dots c_{\alpha_1}^\dagger | 0 \rangle \varphi_{\alpha_N}(x_N) \\ &- \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) \prod_{n \neq N-1} c_{\alpha_n}^\dagger | 0 \rangle \varphi_{\alpha_{N-1}}(x_N) \\ &\vdots \\ &(-1)^{N-1} \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) c_{\alpha_N}^\dagger \dots c_{\alpha_2}^\dagger | 0 \rangle \varphi_{\alpha_1}(x_N). \end{aligned}$$

Using (18) for the $N-1$ -electron states, this is nothing but the Laplace expansion of

$$\begin{vmatrix} \varphi_{\alpha_1}(x_1) & \varphi_{\alpha_2}(x_1) & \cdots & \varphi_{\alpha_N}(x_1) \\ \varphi_{\alpha_1}(x_2) & \varphi_{\alpha_2}(x_2) & \cdots & \varphi_{\alpha_N}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{\alpha_1}(x_N) & \varphi_{\alpha_2}(x_N) & \cdots & \varphi_{\alpha_N}(x_N) \end{vmatrix}$$

along the N th row. Dividing by $\sqrt{N!}$ we see that we have shown (18) for N -electron states, completing the proof by induction.

Given this representation of Slater determinants it is easy to eliminate the coordinates so we can work with N -electron states rather than N -electron wavefunctions—just as in Dirac notation. In particular we can rewrite the basis of Slater determinants (6) into a basis of product *states*

$$\{c_{n_N}^\dagger \cdots c_{n_1}^\dagger |0\rangle \mid n_1 < \cdots < n_N\}, \quad (21)$$

which allows us to express any N -electron state as

$$|\Psi\rangle = \sum_{n_1 < \cdots < n_N} a_{n_1, \dots, n_N} c_{n_N}^\dagger \cdots c_{n_1}^\dagger |0\rangle. \quad (22)$$

2.3 Representation of n -body operators

To work with N -electron states rather than Slater determinants, we also have to rewrite the N -electron operators $M(x)$ appropriately. This is easily done by incorporating the coordinates that we have separated from the Slater determinants into the operators such that the expectation values remain unchanged. This is, again, analogous to the Dirac formalism:

$$\int dx \overline{\varphi_n(x)} M(x) \varphi_m(x) = \langle \varphi_n | \underbrace{\int dx |x\rangle M(x) \langle x|}_{=:\hat{M}} \varphi_m \rangle = \langle \varphi_n | \hat{M} | \varphi_m \rangle. \quad (23)$$

For N -electron Slater determinants it becomes

$$\begin{aligned} & \int dx_1 \cdots dx_N \overline{\Phi_{\beta_1 \dots \beta_N}(x_1, \dots, x_N)} M(x_1, \dots, x_N) \Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N) \\ &= \int dx_1 \cdots dx_N \langle 0 | c_{\beta_1} \cdots c_{\beta_N} \hat{\Psi}^\dagger(x_N) \cdots \hat{\Psi}^\dagger(x_1) | 0 \rangle M(x_1, \dots, x_N) \langle 0 | \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N) c_{\alpha_N}^\dagger \cdots c_{\alpha_1}^\dagger | 0 \rangle \\ &= \langle 0 | c_{\beta_1} \cdots c_{\beta_N} \hat{M} c_{\alpha_N}^\dagger \cdots c_{\alpha_1}^\dagger | 0 \rangle \end{aligned}$$

with the representation of the n -body operator in terms of field operators

$$\hat{M} := \frac{1}{N!} \int dx_1 \cdots dx_N \hat{\Psi}^\dagger(x_N) \cdots \hat{\Psi}^\dagger(x_1) M(x_1, \dots, x_N) \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N). \quad (24)$$

Note that this particular form of the operator is only valid when applied to N -electron states, since we have used that the N annihilation operators bring us to the zero-electron space, where $|0\rangle\langle 0| = \mathbb{1}$. Keeping this in mind, we can work entirely in terms of our algebra (11).

To see what (24) means, we look, in turn, at the different n -body parts of $M(\mathbf{x})$, (2):

$$M(\mathbf{x}) = M^{(0)} + \sum_i M^{(1)}(x_i) + \sum_{i < j} M^{(2)}(x_i, x_j) + \sum_{i < j < k} M^{(3)}(x_i, x_j, x_k) + \dots \quad (25)$$

We start with the simplest case, the zero-body operator, which is just a number, $M^{(0)}$. Operating on an N -electron state $|\Psi\rangle$ with $\hat{N}|\Psi\rangle = N|\Psi\rangle$, it gives

$$\begin{aligned} \hat{M}^{(0)} &= \frac{M^{(0)}}{N!} \int dx_1 dx_2 \dots x_N \hat{\Psi}^\dagger(x_N) \dots \hat{\Psi}^\dagger(x_2) \hat{\Psi}^\dagger(x_1) \hat{\Psi}(x_1) \hat{\Psi}(x_2) \dots \hat{\Psi}(x_N) \\ &= \frac{M^{(0)}}{N!} \int dx_2 \dots x_N \hat{\Psi}^\dagger(x_N) \dots \hat{\Psi}^\dagger(x_2) \hat{N} \hat{\Psi}(x_2) \dots \hat{\Psi}(x_N) \\ &= \frac{M^{(0)}}{N!} \int dx_2 \dots x_N \hat{\Psi}^\dagger(x_N) \dots \hat{\Psi}^\dagger(x_2) 1 \hat{\Psi}(x_2) \dots \hat{\Psi}(x_N) \\ &\vdots \\ &= \frac{M^{(0)}}{N!} 1 \cdot 2 \dots N = M^{(0)}. \end{aligned} \quad (26)$$

using that by (13) every annihilator reduces the expectation value of $\hat{N}$ by one.

Remarkably, while we started from an operator acting on N -electron states, the result in second quantized form is independent of the number of electrons. We will see that this is a general feature of operators in second quantization which makes working in Fock spaces simple.

We note that (26) just means that the overlap of two Slater determinants (7) is equal to that of the corresponding product states

$$\int d\mathbf{x} \overline{\Phi_{\alpha_1, \dots, \alpha_N}(\mathbf{x})} \Phi_{\beta_1, \dots, \beta_N}(\mathbf{x}) = \langle 0 | c_{\alpha_1} \dots c_{\alpha_N} c_{\beta_N}^\dagger \dots c_{\beta_1}^\dagger | 0 \rangle. \quad (27)$$

2.3.1 One-body operators

Next we consider one-body operators $\sum_j M^{(1)}(x_j)$

$$\begin{aligned} \hat{M}^{(1)} &= \frac{1}{N!} \int dx_1 \dots dx_N \hat{\Psi}^\dagger(x_N) \dots \hat{\Psi}^\dagger(x_1) \sum_j M^{(1)}(x_j) \hat{\Psi}(x_1) \dots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \sum_j \int dx_j \hat{\Psi}^\dagger(x_j) M^{(1)}(x_j) (N-1)! \hat{\Psi}(x_j) \\ &= \frac{1}{N} \sum_j \int dx_j \hat{\Psi}^\dagger(x_j) M^{(1)}(x_j) \hat{\Psi}(x_j) \\ &= \int dx \hat{\Psi}^\dagger(x) M^{(1)}(x) \hat{\Psi}(x) \end{aligned} \quad (28)$$

Here we have first anticommutated $\hat{\Psi}^\dagger(x_j)$ all the way to the left and $\hat{\Psi}(x_j)$ to the right. Since these take the same numbers of anticommutations, there is no sign involved. In between these field operators we are left with a zero-body operator for $N-1$ electrons, producing, when $\hat{M}^{(1)}$

acts on an N -electron state, a factor of $(N-1)!$. Again we notice that we obtain an operator that no longer depends on the number of electrons, i.e., that is valid in the entire Fock space.

Expanding the field-operators in a complete orthonormal set $\hat{\Psi}(x) = \sum_n \varphi_n(x) c_n$ gives

$$\hat{M}^{(1)} = \sum_{n,m} \int dx \overline{\varphi_n(x)} M(x) \varphi_m(x) c_n^\dagger c_m = \sum_{n,m} \langle \varphi_n | M^{(1)} | \varphi_m \rangle c_n^\dagger c_m = \sum_{n,m} c_n^\dagger M_{nm}^{(1)} c_m. \quad (29)$$

The matrix elements $M_{nm}^{(1)} = \langle \varphi_n | M^{(1)} | \varphi_m \rangle$ transform like a single-electron matrix $\mathbf{M}^{(1)}$: With $|\tilde{\varphi}_\mu\rangle = \sum_m |\varphi_m\rangle \langle \varphi_m | \tilde{\varphi}_\mu \rangle = \sum_m |\varphi_m\rangle U_{m,\mu}$ and writing the annihilation operators as a column vector $\mathbf{c}$ we see that

$$\hat{M}^{(1)} = \mathbf{c}^\dagger \mathbf{M}^{(1)} \mathbf{c} = \mathbf{c}^\dagger \mathbf{U} \mathbf{U}^\dagger \mathbf{M}^{(1)} \mathbf{U} \mathbf{U}^\dagger \mathbf{c} = \tilde{\mathbf{c}}^\dagger \tilde{\mathbf{M}}^{(1)} \tilde{\mathbf{c}}. \quad (30)$$

Once we have arrived at the representation in terms of orbitals, we can restrict the orbital basis to a non-complete set. This simply gives the operator in the variational (Fock) subspace spanned by the orbitals.

We note that the expression (29) not only works for local operators but also for differential operators like the momentum or kinetic energy: we have taken care not to exchange the order of $M^{(1)}$ and one of its field operators. We can write truly non-local operators in a similar way. As an example, the one-body density operator is given by

$$\hat{I}^{(1)}(x; x') = \hat{\Psi}^\dagger(x) \hat{\Psi}(x') \quad (31)$$

so that one coordinate is not integrated over, rather setting it to x in the bra and x' in the ket. In an orthonormal basis it becomes

$$\hat{I}^{(1)}(x; x') = \sum_{n,m} \overline{\varphi_n(x)} \varphi_m(x') c_n^\dagger c_m. \quad (32)$$

2.3.2 Two-body operators

For the two-body operators $\sum_{i<j} M^{(2)}(x_i, x_j)$ we proceed in the familiar way, anti-commuting first the operators with the coordinates involved in $M^{(2)}$ all the way to the left and right, respectively. This time we are left with a zero-body operator for $N-2$ electrons:

$$\begin{aligned} \hat{M}^{(2)} &= \frac{1}{N!} \int dx_1 \cdots dx_N \hat{\Psi}^\dagger(x_N) \cdots \hat{\Psi}^\dagger(x_1) \sum_{i<j} M^{(2)}(x_i, x_j) \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \sum_{i<j} \int dx_i dx_j \hat{\Psi}^\dagger(x_j) \hat{\Psi}^\dagger(x_i) M^{(2)}(x_i, x_j) (N-2)! \hat{\Psi}(x_i) \hat{\Psi}(x_j) \\ &= \frac{1}{N(N-1)} \sum_{i<j} \int dx_i dx_j \hat{\Psi}^\dagger(x_j) \hat{\Psi}^\dagger(x_i) M^{(2)}(x_i, x_j) \hat{\Psi}(x_i) \hat{\Psi}(x_j) \\ &= \frac{1}{2} \int dx dx' \hat{\Psi}^\dagger(x') \hat{\Psi}^\dagger(x) M^{(2)}(x, x') \hat{\Psi}(x) \hat{\Psi}(x') \end{aligned}$$

Expanding in an orthonormal basis, we get

$$\begin{aligned}\hat{M}^{(2)} &= \frac{1}{2} \sum_{n,n',m,m'} \int dx dx' \overline{\varphi_{n'}(x') \varphi_n(x)} M^{(2)}(x, x') \varphi_m(x) \varphi_{m'}(x') c_n^\dagger c_{n'}^\dagger c_m c_{m'} \\ &= \frac{1}{2} \sum_{n,n',m,m'} \langle \varphi_n \varphi_{n'} | M^{(2)} | \varphi_m \varphi_{m'} \rangle c_n^\dagger c_{n'}^\dagger c_m c_{m'}\end{aligned}\quad (33)$$

where the exchange of the indices in the second line is a consequence of the way the Dirac state for two electrons is usually written: first index for the first coordinate, second index for the second, while taking the adjoint of the operators changes their order. $M_{nn',mm'} = \langle \varphi_n \varphi_{n'} | M^{(2)} | \varphi_m \varphi_{m'} \rangle$ transforms like a fourth-order tensor: Transforming to a different basis, $|\tilde{\varphi}_\mu\rangle = \sum_m |\varphi_m\rangle \langle \varphi_m | \tilde{\varphi}_\mu\rangle$ gives

$$\tilde{M}_{\nu\nu',\mu\mu'}^{(2)} = \sum_{n,n',m,m'} \langle \tilde{\varphi}_\nu | \varphi_n \rangle \langle \tilde{\varphi}_{\nu'} | \varphi_{n'} \rangle M_{nn',mm'} \langle \varphi_m | \tilde{\varphi}_\mu \rangle \langle \varphi_{m'} | \tilde{\varphi}_{\mu'} \rangle. \quad (34)$$

Form the symmetry of the two-body operator $M^{(2)}(x, x') = M^{(2)}(x', x)$ follows $M_{nn',mm'} = M_{n'n,mm'}$, while $M_{nn,mm'}$ will not contribute to $\hat{M}^{(2)}$ since $(c_n^\dagger)^2 = 0$, and likewise for $M_{nn',mm}$. Note that the representation (33) is not quite as efficient as it could be: The terms with n and n' and/or m and m' exchanged connect the same basis states. Collecting these terms by introducing an ordering of the operators and using the symmetry of the matrix elements we obtain

$$\hat{M}^{(2)} = \sum_{n' > n, m' > m} c_{n'}^\dagger c_n^\dagger \underbrace{\left(M_{nn',mm'}^{(2)} - M_{n'n,mm'}^{(2)} \right)}_{=: \check{M}_{nn',mm'}^{(2)}} c_m c_{m'}. \quad (35)$$

Since the states $\{c_{n'}^\dagger c_n^\dagger | 0\rangle \mid n' > n\}$ form a basis of the two-electron Hilbert space, considering nn' as the index of a basis state, the $\check{M}_{nn',mm'}^{(2)}$ form a two-electron matrix $\check{\mathbf{M}}^{(2)}$ and the basis transform becomes the matrix $\check{\mathbf{U}}^{(2)}$

$$\tilde{c}_\nu^\dagger \tilde{c}_\nu^\dagger = \sum_{n' > n} c_{n'}^\dagger c_n^\dagger \left(\langle \varphi_n | \tilde{\varphi}_\nu \rangle \langle \varphi_{n'} | \tilde{\varphi}_{\nu'} \rangle - \langle \varphi_{n'} | \tilde{\varphi}_\nu \rangle \langle \varphi_n | \tilde{\varphi}_{\nu'} \rangle \right). \quad (36)$$

The two-body density operator is

$$\hat{\Gamma}^{(2)}(x, y; y', x') = \hat{\Psi}^\dagger(x) \hat{\Psi}^\dagger(y) \hat{\Psi}(y') \hat{\Psi}(x') \quad (37)$$

from which the corresponding one-body operator is obtained by integration

$$\int dy \hat{\Gamma}^{(2)}(x, y; y, x) = \hat{\Gamma}^{(1)}(x; x') (\hat{N} - 1). \quad (38)$$

The procedure of rewriting operators in second quantization obviously generalizes to observables acting on more than two electrons in the natural way.

We note that, while we started from a form of the operators (25) that was explicitly formulated in an N -electron Hilbert space, the results (26), (29), and (33) are of the same form no matter what value N takes. Thus these operators are valid not just on some N -electron Hilbert space, but on the entire Fock space. Moreover, it is impossible to write states which are not antisymmetric or unphysical operators which are not treating all particles in the same way, eliminating many opportunities for making mistakes. These are the particular strengths of the second-quantized formalism.

2.4 Transforming the orbital basis

It is worth to note that the creators transform in the same way as the orbitals they represent

$$|\beta_i\rangle = \sum_{\mu} |\alpha_{\mu}\rangle \underbrace{\langle\alpha_{\mu}|\beta_i\rangle}_{=:U_{\mu i}} \rightsquigarrow c_{\beta_i}^{\dagger} = \sum_{\mu} c_{\alpha_{\mu}}^{\dagger} U_{\mu i}, \quad (39)$$

so the “operators” really transform like states. Writing the transformation matrix as $U = e^M$, where M is anti-Hermitian, $M^{\dagger} = -M$ when U is unitary, but can be any matrix when U is merely invertible, we can write the basis transformation in a form appropriate for operators:

$$c_{\beta_i}^{\dagger} = e^{c^{\dagger} M c} c_{\alpha_{\mu}}^{\dagger} e^{-c^{\dagger} M c}. \quad (40)$$

To see this, we use the Baker-Campbell-Hausdorff formula in the form

$$e^{\lambda A} B e^{-\lambda A} = B + \lambda [A, B] + \frac{\lambda^2}{2!} [A, [A, B]] + \frac{\lambda^3}{3!} [A, [A, [A, B]]] + \dots, \quad (41)$$

where the expansion coefficients follow by taking the derivatives of the left hand side at $\lambda = 0$, together with the commutator (using $[AB, C] = A\{B, C\} - \{A, C\}B$)

$$[c_{\alpha_{\mu}}^{\dagger} c_{\alpha_{\nu}}, c_{\alpha_{\kappa}}^{\dagger}] = c_{\alpha_{\mu}}^{\dagger} \delta_{\nu, \kappa} \quad (42)$$

from which we obtain for the repeated commutators

$$\left[\sum_{\mu, \nu} M_{\mu\nu} c_{\alpha_{\mu}}^{\dagger} c_{\alpha_{\nu}}, \sum_{\kappa} c_{\alpha_{\kappa}}^{\dagger} (M^n)_{\kappa i} \right] = \sum_{\mu\nu\kappa} c_{\alpha_{\mu}}^{\dagger} M_{\mu\nu} \delta_{\nu, \kappa} (M^n)_{\kappa i} = \sum_{\mu} c_{\alpha_{\mu}}^{\dagger} (M^{n+1})_{\mu i}. \quad (43)$$

To keep the derivation simple, we have chosen to transform an operator from the orthonormal basis that we also used to write the exponential operator. Being linear, the transform works, of course, the same for an arbitrary creation operator.

Using this form of the basis transformation and noticing that $e^{-c^{\dagger} M c} |0\rangle = |0\rangle$, we immediately see that acting with the exponential of a one-body operator on a product state results in another product state

$$e^{c^{\dagger} M c} \prod c_{\alpha_n}^{\dagger} |0\rangle = \prod e^{c^{\dagger} M c} c_{\alpha_n}^{\dagger} e^{-c^{\dagger} M c} |0\rangle = \prod c_{\beta_n}^{\dagger} |0\rangle. \quad (44)$$

This is, e.g., used when working in the interaction picture. Anticommutators with transformed operators, (39), are simply $\{c_{\alpha_j}, e^{c^{\dagger} M c} c_{\alpha_i}^{\dagger} e^{-c^{\dagger} M c}\} = \langle\alpha_j|e^M|\alpha_i\rangle$.

Annihilation operators, being the adjoint of the creators, transform in just the expected way

$$c_{\beta_i} = e^{-c^{\dagger} M^{\dagger} c} c_{\alpha_{\mu}} e^{c^{\dagger} M^{\dagger} c}, \quad (45)$$

which means that for unitary transformations, where M is anti-Hermitian, creators and annihilators transform in the same way. Note that in the imaginary-time formalism the annihilators are, via analytic continuation, chosen to transform in the same way as the creators, making them different from the adjoint of the creators.

3 Exact diagonalization

We have worked, so far, with complete, i.e., infinite bases. This is, of course, not possible in actual computer simulations, where we have to confine ourselves to finite basis sets. Such calculations on subspaces are based on the variational principle.

3.1 Variational principles

The variational principle and the Schrödinger equation are equivalent. Consider the energy expectation value as a wave-function functional

$$E[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (46)$$

Its variation is

$$E[\Psi + \delta\Psi] = E[\Psi] + \frac{\langle \delta\Psi | H | \Psi \rangle + \langle \Psi | H | \delta\Psi \rangle}{\langle \Psi | \Psi \rangle} - \langle \Psi | H | \Psi \rangle \frac{\langle \delta\Psi | \Psi \rangle + \langle \Psi | \delta\Psi \rangle}{\langle \Psi | \Psi \rangle^2} + \mathcal{O}^2. \quad (47)$$

The first-order term vanishes for $H|\Psi\rangle = E[\Psi]|\Psi\rangle$, which is the Schrödinger equation. Since the eigenfunctions

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle, \quad (48)$$

can be chosen to form an orthonormal basis, we can expand any wavefunction as

$$|\Psi\rangle = \sum_n |\Psi_n\rangle \langle \Psi_n | \Psi \rangle \quad (49)$$

and determine, as long as $\langle \Psi | \Psi \rangle \neq 0$, its energy expectation value

$$\frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_{m,n} \langle \Psi | \Psi_m \rangle \langle \Psi_m | H | \Psi_n \rangle \langle \Psi_n | \Psi \rangle}{\sum_{m,n} \langle \Psi | \Psi_m \rangle \langle \Psi_m | \Psi_n \rangle \langle \Psi_n | \Psi \rangle} = \frac{\sum_n E_n |\langle \Psi_n | \Psi \rangle|^2}{\sum_n |\langle \Psi_n | \Psi \rangle|^2}. \quad (50)$$

Since by definition no eigenenergy can be lower than the ground state energy E_0 , we immediately see that the energy expectation value can never drop below the ground state energy

$$\frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_n E_n |\langle \Psi_n | \Psi \rangle|^2}{\sum_n |\langle \Psi_n | \Psi \rangle|^2} \geq \frac{\sum_n E_0 |\langle \Psi_n | \Psi \rangle|^2}{\sum_n |\langle \Psi_n | \Psi \rangle|^2} = E_0. \quad (51)$$

We can use the same argument to generalize this variational principle: Assume we have arranged the eigenenergies in ascending order, $E_0 \leq E_1 \leq \dots$, then the energy expectation value for a wavefunction that is orthogonal to the n lowest eigenstates, can not drop below E_n

$$\frac{\langle \Psi_{\perp n} | H | \Psi_{\perp n} \rangle}{\langle \Psi_{\perp n} | \Psi_{\perp n} \rangle} \geq E_n \quad \text{if } \langle \Psi_i | \Psi_{\perp n} \rangle = 0 \text{ for } i = 0, \dots, n-1. \quad (52)$$

This generalized variational principle is, of course, only of practical use if we know something about the eigenstates, e.g., when we can use symmetries to ensure orthogonality.

For an ab-initio Hamiltonian of N electrons in the field of nuclei of charge Z_α at position $\mathbf{R}_\alpha$,

$$H = -\frac{1}{2} \sum_i \Delta_i - \sum_{i,\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}, \quad (53)$$

the Schrödinger equation is a partial differential equation. In second quantization it becomes a linear-algebra problem: We introduce an orbital basis set $\{\varphi_k | k\}$, which for simplicity we assume here to be orthonormal, from which we construct an orthonormal basis of N -electron product states, $\{\Phi_{k_1, \dots, k_N} | k_1 < \dots < k_N\}$. To simplify the notation we sort the basis states, e.g., lexicographically in the orbital indices $\mathbf{k} = (k_1, \dots, k_N)$ and define the row vector of basis states $|\Phi\rangle := (|\Phi_1\rangle, |\Phi_2\rangle, \dots)$. The expansion of a state $|\Psi\rangle$ in this basis can then be written as

$$|\Psi\rangle = \sum_{k_1 < \dots < k_N} |\Phi_{k_1, \dots, k_N}\rangle \langle \Phi_{k_1, \dots, k_N} | \Psi \rangle = \sum_i |\Phi_i\rangle a_i = |\Phi\rangle \mathbf{a}, \quad (54)$$

where $\mathbf{a}$ is the vector of expansion coefficients. Likewise we can write the Schrödinger equation as a matrix eigenvalue problem

$$\mathbf{H}\mathbf{a} = \langle \Phi | \hat{H} | \Phi \rangle \mathbf{a} = \begin{pmatrix} \langle \Phi_1 | \hat{H} | \Phi_1 \rangle & \langle \Phi_1 | \hat{H} | \Phi_2 \rangle & \dots \\ \langle \Phi_2 | \hat{H} | \Phi_1 \rangle & \langle \Phi_2 | \hat{H} | \Phi_2 \rangle & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = E \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} = E\mathbf{a}. \quad (55)$$

From the eigenvectors of the matrix $\mathbf{H}$ we easily recover the eigenstates of the Hamiltonian

$$\mathbf{H}\mathbf{a}_n = E_n \mathbf{a}_n \quad \rightsquigarrow \quad \hat{H}|\Psi_n\rangle = E_n |\Psi_n\rangle \quad \text{with } |\Psi_n\rangle = |\Phi\rangle \mathbf{a}_n. \quad (56)$$

Unfortunately, for an ab-initio Hamiltonian like (53) we need an infinite orbital basis set, so that the Hamiltonian matrix $\mathbf{H}$ is infinite dimensional. A pragmatic approach to allow for computer simulations is to simply restrict the calculation to a finite basis $|\tilde{\Phi}\rangle := (|\tilde{\Phi}_1\rangle, \dots, |\tilde{\Phi}_{\tilde{L}}\rangle)$, i.e., work with a finite matrix $\tilde{\mathbf{H}} := \langle \tilde{\Phi} | \hat{H} | \tilde{\Phi} \rangle$ of dimension $\tilde{L}$. The crucial question is then how the eigenvectors

$$\tilde{\mathbf{H}}\tilde{\mathbf{a}}_n = \tilde{E}_n \tilde{\mathbf{a}}_n \quad \rightsquigarrow \quad |\tilde{\Psi}_n\rangle := |\tilde{\Phi}\rangle \tilde{\mathbf{a}}_n \quad (57)$$

relate to those of H . The answer is surprisingly simple [7]: The eigenvalues of $\tilde{\mathbf{H}}$, ordered as $\tilde{E}_0 \leq \tilde{E}_1 \leq \dots \leq \tilde{E}_{\tilde{L}-1}$, are variational with respect to those of H ,

$$E_n \leq \tilde{E}_n \quad \text{for } n \in \{0, \dots, \tilde{L}-1\}. \quad (58)$$

To show this, we construct a state in $\text{span}(|\tilde{\Psi}_0\rangle, \dots, |\tilde{\Psi}_n\rangle)$, which by construction has an energy expectation value $\leq \tilde{E}_n$, that is orthogonal to the exact eigenstates $|\Psi_0\rangle, \dots, |\Psi_{n-1}\rangle$, so that by the generalized variational principle its expectation value is $\geq E_n$. This being a homogeneous system of n linear equations with $n+1$ variables, such a state certainly exists, hence $E_n \leq \tilde{E}_n$. To get reliable results, we simply have to systematically increase the basis until the change in the desired eigenvalues becomes smaller than the accuracy required by the physical problem. The art is, of course, to devise clever basis sets such that this is achieved already for bases of low dimensions.

The convergence of the matrix eigenvalues with increasing basis size is surprisingly regular. Let us extend our original basis of $\tilde{L}$ states by an additional $L - \tilde{L}$ states. Then, repeating the above argument with the L -dimensional problem taking the role of $\hat{H}$, we obtain (58) with E_n being the eigenvalues of the L -dimensional Hamiltonian matrix $\mathbf{H}$. Since $\mathbf{H}$ now is finite, we can use the same argument for $-\mathbf{H}$, obtaining

$$-E_{L-i} \leq -\tilde{E}_{\tilde{L}-i} \quad \text{for } i \in \{1, \dots, \tilde{L}\}. \quad (59)$$

Taking the two inequalities together we obtain

$$E_n \leq \tilde{E}_n \leq E_{n+(L-\tilde{L})} \quad \text{for } n \in \{0, \dots, \tilde{L}-1\}. \quad (60)$$

For the special case $L = \tilde{L} + 1$ of adding a single basis state, this is the Hylleraas-Undheim/MacDonald nesting property for eigenvalues in successive approximations

$$E_1 \leq \tilde{E}_1 \leq E_2 \leq \tilde{E}_2 \leq \dots \leq \tilde{E}_L \leq E_{L+1}. \quad (61)$$

3.2 Matrix eigenvalue problem

For practical calculations we have to set up the Hamiltonian matrix $\tilde{\mathbf{H}} = \langle \tilde{\Phi} | \hat{H} | \tilde{\Phi} \rangle$ and the state vectors $\tilde{\mathbf{a}}$ for the chosen basis. This is particularly easy for a basis of Slater determinants constructed from a basis set of K orbitals $\{\varphi_k \mid k = 0, \dots, K-1\}$. The basis states are then the N -electron product states of $|\Phi_{k_1, \dots, k_N}\rangle = c_{k_N}^\dagger \dots c_{k_1}^\dagger |0\rangle$ with $k_1 < \dots < k_N$. We can write them more computer friendly as

$$|n_{K-1}, \dots, n_1, n_0\rangle = (c_{K-1}^\dagger)^{n_{K-1}} \dots (c_1^\dagger)^{n_1} (c_0^\dagger)^{n_0} |0\rangle \quad (62)$$

which is the occupation number representation with $n_k \in \{0, 1\}$ and $\sum n_k = N$. It is natural to interpret the vector of occupation numbers as the binary representation of the integer $\sum_k 2^{n_k}$. This implies a natural ordering of the basis functions $|\Phi_l\rangle$. For the simple case of $K=4$ orbitals and $N=2$ electrons we obtain

i	(n_3, n_2, n_1, n_0)	state	l
0	0000		
1	0001		
2	0010		
3	0011	$c_1^\dagger c_0^\dagger 0\rangle = \Phi_1\rangle$	1
4	0100		
5	0101	$c_2^\dagger c_0^\dagger 0\rangle = \Phi_2\rangle$	2
6	0110	$c_2^\dagger c_1^\dagger 0\rangle = \Phi_3\rangle$	3
7	0111		
8	1000		
9	1001	$c_3^\dagger c_0^\dagger 0\rangle = \Phi_4\rangle$	4
10	1010	$c_3^\dagger c_1^\dagger 0\rangle = \Phi_5\rangle$	5
11	1011		
12	1100	$c_3^\dagger c_2^\dagger 0\rangle = \Phi_6\rangle$	6
13	1101		
14	1110		
15	1111		

The bit representation of the basis states also simplifies setting up the Hamiltonian matrix. Given the Hamiltonian in second quantization

$$\hat{H} = \sum_{n,m} T_{nm} c_n^\dagger c_m + \sum_{n'>n, m'>m} \underbrace{(U_{nn',mm'} - U_{n'n,mm'})}_{=\check{U}_{nn',mm'}} c_{n'}^\dagger c_n^\dagger c_m c_{m'} \quad (63)$$

the matrix element $\langle \Phi_l | \hat{H} | \Phi_{l'} \rangle$, with $|\Phi_{l'}\rangle = c_{l'_N}^\dagger \cdots c_{l'_1}^\dagger |0\rangle$, is given by

$$\sum_{n,m} T_{nm} \langle 0 | c_{l_1} \cdots c_{l_N} c_n^\dagger c_m c_{l'_N}^\dagger \cdots c_{l'_1}^\dagger |0\rangle + \sum_{\substack{n'>n \\ m'>m}} \check{U}_{nn',mm'} \langle 0 | c_{l_1} \cdots c_{l_N} c_{n'}^\dagger c_n^\dagger c_m c_{m'} c_{l'_N}^\dagger \cdots c_{l'_1}^\dagger |0\rangle.$$

Anticommuting the operators coming from the Hamiltonian, the matrix elements become overlaps of $N+1$ and $N+2$ -electron product states, which, by (27) and (7), are just the determinants of the overlap matrices of the corresponding orbitals. When $\hat{H}$ is written in the same orbitals as these $|\Phi_l\rangle$, the overlap matrices simplify to permutation matrices with determinant ± 1 . In the occupation number representation, calculating this Fermi sign reduces to counting set bits. As an example we consider a simple hopping of an electron:

$$\begin{aligned} c_6^\dagger c_2 |\Phi_{l(181)}\rangle &= c_6^\dagger c_2 c_7^\dagger c_5^\dagger c_4^\dagger c_2^\dagger c_0^\dagger |0\rangle \\ &= (-1)^3 c_6^\dagger c_7^\dagger c_5^\dagger c_4^\dagger c_2 c_2^\dagger c_0^\dagger |0\rangle \\ &= (-1)^3 c_6^\dagger c_7^\dagger c_5^\dagger c_4^\dagger (1 - c_2^\dagger c_2) c_0^\dagger |0\rangle \\ &= (-1)^3 c_6^\dagger c_7^\dagger c_5^\dagger c_4^\dagger c_0^\dagger |0\rangle \\ &= + |\Phi_{l(241)}\rangle = (-1)^2 c_7^\dagger c_6^\dagger c_5^\dagger c_4^\dagger c_0^\dagger |0\rangle \end{aligned}$$

In the occupation number representation this becomes

$$\begin{array}{ccccccc} & \downarrow & & & & & \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{array} = (-1)^c \begin{array}{ccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}$$

where c is the count of set bits between the orbitals of the electron hop. Note that practically all processors have a dedicated machine instruction for counting set bits, see also [8].

3.3 Dimension of the Hilbert space and sparseness

Setting up basis states and Hamiltonian matrix in this way, we can easily solve the many-body problem on our variational space by using any linear algebra library. This is the exact diagonalization approach. As discussed above, it gives us variational estimates of the ground and excited states. But there is a serious practical problem: the dimension of the many-body Hilbert space. For an N -electron problem with a basis set of K orbitals there are $K(K-1)(K-2)\cdots(K-(N-1))$ ways of picking N occupied orbitals out of K . Since we only use one specific ordering of these indices, we have to divide by $N!$ to obtain the number of such determinants:

$$\dim \mathcal{H}_K^{(N)} = \frac{K!}{N!(K-N)!} = \binom{K}{N}. \quad (64)$$

Using Stirling's formula we see that for an N -electron problem this increases faster than exponentially with the size K of the basis set. This is the problem we face when converging the

basis set for a finite system, e.g., a molecule. For solids we usually keep the number of orbitals per lattice site fixed, but scale to the thermodynamic limit, increasing the system size M while keeping the electron density N/M fixed. Also here the Hilbert space increases faster than exponentially. To give an impression of the problem we note that for $N = 25$ electrons and $K = 100$ orbitals the dimension already exceeds 10^{23} .

For exact diagonalization the problem gets even worse. Assuming we have a machine with 1 TeraBytes = 2^{40} Bytes of RAM available. Using single precision (4 bytes) for the matrix elements, storing a matrix of dimension $(2^{40}/4)^{1/2} = 524\,288$ would already use up all memory. The dimension problem can be somewhat mitigated by exploiting symmetries: When the Hamiltonian commutes with the projection of the total spin, the number of up- and down-spin electrons is conserved separately. The N -electron Hamiltonian is then block diagonal in the sectors with fixed $N_\uparrow$ and $N_\downarrow$. The dimension of these blocks is significantly smaller than that of the full N -electron Hilbert space. Using the same orbital basis for each spin

$$\dim \mathcal{H}_{2K}^{(N_\uparrow, N_\downarrow)} = \binom{K}{N_\uparrow} \times \binom{K}{N_\downarrow}. \quad (65)$$

The S_z symmetry can be very easily implemented using the same ideas as introduced for the general case: just use bit representations for the up- and down-spin electrons separately. In fact, when the total spin projection is conserved, we can distinguish electrons of different spin. Still, the Hilbert space of the single-band, half-filled Hubbard model with just 12 sites has dimension 853 776. Using further symmetries, if they exist, we could bring down the dimension somewhat further, however at the expense of considerable and problem-specific effort.

The key to going to larger systems is the realization that the vast majority of the elements of the Hamiltonian matrix is zero. This is quite easy to see. For the ab-initio Hamiltonian (53) with electron-electron repulsion, matrix elements between configurations that differ in more than two electron occupations vanish. Thus, for each configuration there may only be the diagonal element, $N \times (K - N)$ hopping terms, and $N(N-1)/2 \times (K-N)(K-N-1)/2$ pair-hopping terms. Thus the fraction of non-zero matrix elements of $\tilde{H}$ to the total number is

$$\left(1 + N \left(1 + \frac{N-1}{2} \frac{(K-N-1)}{2} \right) (K-N) \right) / \binom{K}{N} \quad (66)$$

which, with increasing problem size, rapidly approaches zero. For the example of $N = 25$ electrons in $K = 100$ orbitals only 834 376 of the (over 10^{23}) matrix elements per row can be non-zero. This is the worst case. The sparsity of many-body Hamiltonians is even more pronounced when working in a tight-binding basis with short-ranged hopping and local electron-electron repulsion. Thus, many-body Hamiltonians are exceedingly sparse and the more so the larger the problem. They are therefore ideally suited for approaches like the Lanczos method, that are based on matrix-vector products, which for the sparse matrices scale close to linearly in the matrix dimension.

4 Lanczos Method

As we have seen, we can find the ground-state $|\Psi_0\rangle$ and its energy E_0 for a Hamiltonian H from the variational principle. The wavefunction-functional

$$E[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (67)$$

is minimized for $\Psi = \Psi_0$, with $E[\Psi_0] = E_0$. The functional gradient

$$\frac{\delta E[\Psi]}{\delta \langle \Psi |} = \frac{H|\Psi\rangle - E[\Psi]|\Psi\rangle}{\langle \Psi | \Psi \rangle} = |\Psi_a\rangle \quad (68)$$

gives the direction of steepest-ascent of the functional from the point $|\Psi\rangle$. Moving in the opposite direction will thus result in a wavefunction with lower energy expectation value: $E[\Psi - \alpha \Psi_a] < E[\Psi]$ for small, positive α .

To find the optimum value of α , we minimize the quadratic form $E[\Psi - \alpha \Psi_a]$. For this, it is convenient to introduce an orthogonal basis in the space spanned by the two vectors $|\Psi\rangle$ and $|\Psi_a\rangle$. From (68) we see that $\text{span}(|\Psi\rangle, |\Psi_a\rangle) = \text{span}(|\Psi\rangle, H|\Psi\rangle)$. As first basis vector, we normalize $|\Psi\rangle$

$$|v_0\rangle = |\Psi\rangle / \sqrt{\langle \Psi | \Psi \rangle},$$

for the second vector we orthogonalize $H|v_0\rangle$ to $|v_0\rangle$

$$|\tilde{v}_1\rangle = H|v_0\rangle - |v_0\rangle \langle v_0 | H | v_0 \rangle \quad (69)$$

and normalize to obtain $|v_1\rangle$. With $a_n := \langle v_n | H | v_n \rangle$ and $b_1^2 := \langle \tilde{v}_1 | \tilde{v}_1 \rangle$, eq. (69) becomes

$$H|v_0\rangle = b_1|v_1\rangle + a_0|v_0\rangle \quad (70)$$

from which we see that $\langle v_1 | H | v_0 \rangle = b_1$.

We can then write any normalized wavefunction in $\text{span}(|\Psi\rangle, H|\Psi\rangle) = \text{span}(|v_0\rangle, |v_1\rangle)$ as

$$|v\rangle = \cos(\theta)|v_0\rangle + \sin(\theta)|v_1\rangle. \quad (71)$$

Minimizing the expectation value

$$\langle v | H | v \rangle = a_0 \cos^2(\theta) + 2b_1 \sin(\theta) \cos(\theta) + a_1 \sin^2(\theta), \quad (72)$$

with respect to θ , we obtain, dividing by $\cos^2(\theta)$, the quadratic equation

$$b_1 \tan^2(\theta) + (a_0 - a_1) \tan(\theta) - b_1 = 0. \quad (73)$$

Solving for θ we find the lowest-energy state on the subspace spanned by $|v_0\rangle$ and $H|v_0\rangle$. Alternatively, we can diagonalize the Hamiltonian matrix on the two-dimensional subspace, which in the basis $|v_0\rangle, |v_1\rangle$ is given by

$$H_{\text{span}(|\Psi\rangle, H|\Psi\rangle)} = \begin{pmatrix} a_0 & b_1 \\ b_1 & a_1 \end{pmatrix}. \quad (74)$$

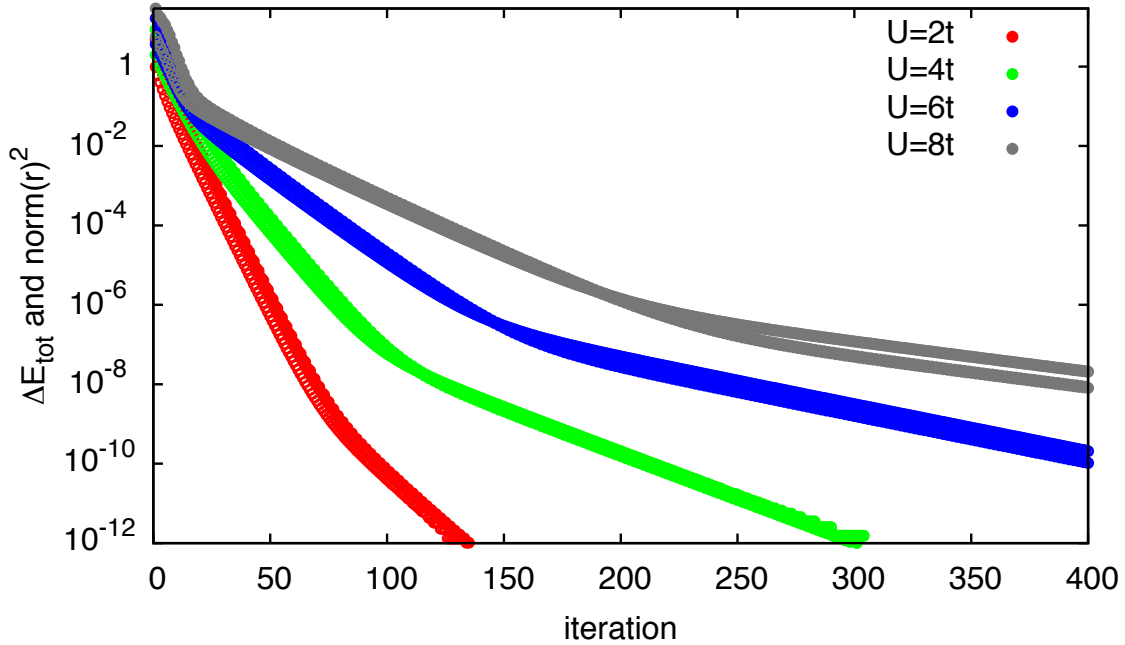


Fig. 1: Convergence of the residual (filled circles) and the corresponding lowest eigenvalue (open circles) for a steepest-descent minimization of a Hubbard-chain of 10 sites at half-filling, starting from a random initial vector.

Naturally, we can use the variational state of lowest energy

$$|\Psi^{(2)}\rangle = \cos(\theta_{\min})|v_0\rangle + \sin(\theta_{\min})|v_1\rangle \quad (75)$$

as the starting point for another steepest-descent minimization. Doing this repeatedly, we obtain a series of vectors with decreasing energy expectation value, which rapidly converge to a minimum. For a generic functional such a steepest-descent minimization would usually end up in a local, not the global minimum, which makes the optimization of high-dimensional functions such a hard problem. The energy functional (67), however, has only minima for the ground-states, all other stationary points are saddle points. We can thus expect rapid convergence to the ground state, examples are given in figure 1, except when the starting vector $|v_0\rangle$ is orthogonal to the ground state. In this case also $\langle\Psi_0|H|v_0\rangle = 0$ so that we only converge to the lowest state that overlaps with $|v_0\rangle$. Should $|v_0\rangle$ happen to be an exact eigenvector, $H|v_n\rangle$ does not add a new dimension so that the algorithm terminates with $b_1 = 0$.

For checking convergence of this steepest-descent method, introduced by Kantorovich [9] and, independently, by Hestenes and Karush [10], we can monitor the change in the energy expectation value or determine when the residual

$$r[\Psi] = \|(H - E[\Psi])|\Psi\rangle\|^2 = \langle\Psi|H^2|\Psi\rangle - E[\Psi]^2, \quad (76)$$

which measures the quality of the eigenstate, becomes sufficiently small. As shown in Fig. 1, both are closely related.

4.1 Krylov space

If we apply the method of steepest-descent L times, starting from a vector $|v_0\rangle$, the resulting vector will lie in

$$\mathcal{K}^L(|v_0\rangle) = \text{span}(|v_0\rangle, H|v_0\rangle, H^2|v_0\rangle, \dots, H^L|v_0\rangle), \quad (77)$$

the $L+1$ -dimensional *Krylov space* [11] of H over $|v_0\rangle$. Instead of repeatedly minimizing the energy in two-dimensional subspaces, we could directly find the state of lowest energy in $\mathcal{K}^L(|v_0\rangle)$. Having more degrees of freedom for the minimization will lead to even faster convergence.

To implement this idea, we construct an orthonormal basis $|v_n\rangle$ of the Krylov space. We start with the normalized vector $|v_0\rangle$. The second basis vector $|v_1\rangle$ is constructed as in the steepest-descent method (69):

$$b_1|v_1\rangle = |\tilde{v}_1\rangle = H|v_0\rangle - a_0|v_0\rangle. \quad (78)$$

The next basis vector is likewise constructed as $H|v_1\rangle$ orthogonalized to all previous vectors, and normalized

$$b_2|v_2\rangle = |\tilde{v}_2\rangle = H|v_1\rangle - \sum_{i=0}^1 |v_i\rangle\langle v_i|H|v_1\rangle = H|v_1\rangle - a_1|v_1\rangle - b_1|v_0\rangle. \quad (79)$$

where $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. The fourth basis vector is

$$b_3|v_3\rangle = |\tilde{v}_3\rangle = H|v_2\rangle - \sum_{i=0}^2 |v_i\rangle\langle v_i|H|v_2\rangle = H|v_2\rangle - a_2|v_2\rangle - b_2|v_1\rangle. \quad (80)$$

The last term in the orthogonalization vanishes, because (78) together with the orthogonality of the already constructed basis vectors for $n < 3$ implies $\langle v_2|H|v_0\rangle = b_1\langle v_2|v_1\rangle + a_0\langle v_2|v_0\rangle = 0$. The construction of the further basis vectors follows the same scheme

$$b_{n+1}|v_{n+1}\rangle = |\tilde{v}_{n+1}\rangle = H|v_n\rangle - \sum_{i=0}^n |v_i\rangle\langle v_i|H|v_n\rangle = H|v_n\rangle - a_n|v_n\rangle - b_n|v_{n-1}\rangle$$

with $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. Rearranging shows that H is tridiagonalized

$$H|v_n\rangle = b_n|v_{n-1}\rangle + a_n|v_n\rangle + b_{n+1}|v_{n+1}\rangle$$

which in turn implies that $H|v_i\rangle$ is orthogonal to all basis states, except $|v_i\rangle$ and $|v_{i\pm 1}\rangle$. This tridiagonalization of H is the essence of the *Lanczos method* [12].

After L steps the Hamiltonian on the $L+1$ -dimensional Krylov space is given by

$$H_{\mathcal{K}^L(|v_0\rangle)} = \begin{pmatrix} a_0 & b_1 & 0 & 0 & & 0 & 0 \\ b_1 & a_1 & b_2 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & a_2 & b_3 & & 0 & 0 \\ 0 & 0 & b_3 & a_3 & & 0 & 0 \\ & \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & & a_{L-1} & b_L \\ 0 & 0 & 0 & 0 & \cdots & b_L & a_L \end{pmatrix} \quad (81)$$

v=init	
b0=norm2 (v)	not part of tridiagonal matrix
scal (1/b0, v)	$v = v_0\rangle$
w=0	
w=w+H*v	$w = H v_0\rangle$
a[0]=dot (v, w)	
axpy (-a[0], v, w)	$w = \tilde{v}_1\rangle = H v_0\rangle - a_0 v_0\rangle$
b[1]=norm2 (w)	
for n=1, 2, ...	
if abs(b[n])<eps then exit	invariant subspace
scal (1/b[n], w)	$w = v_n\rangle$
scal (-b[n], v)	$v = -b_n v_{n-1}\rangle$
swap (v, w)	
w=w+H*v	$w = H v_n\rangle - b_n v_{n-1}\rangle$
a[n]=dot (v, w)	$a[n] = \langle v_n H v_n\rangle - b_n\langle v_n v_{n-1}\rangle$
axpy (-a[n], v, w)	$w = \tilde{v}_{n+1}\rangle$
b[n+1]=norm2 (w)	
diag(a[0]..a[n], b[1]..b[n])	getting a_{n+1} needs another $H v\rangle$
if converged then exit	
end	

Table 1: *The implementation of the Lanczos iteration requires only two N -dimensional vectors for tridiagonalizing H and thus for calculating the ground-state energy. Constructing the Lanczos-approximation of the ground-state vector requires a second iteration and one additional N -dimensional vector. The by far most expensive operation is the matrix-vector product, which requires a problem specific implementation, while the vector operations use the BLAS.*

If we do not normalize the basis vectors, we obtain an iteration of the form

$$|\Phi_{n+1}\rangle = H |\Phi_n\rangle - \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle} |\Phi_n\rangle - \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle} |\Phi_{n-1}\rangle \quad (82)$$

where $|\Phi_n\rangle = \prod_{i=1}^n b_i |v_n\rangle$ in terms of which we have

$$a_n = \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle}, \quad b_n^2 = \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle}. \quad (83)$$

In this unnormalized basis the Hamiltonian appears non-Hermitian

$$H |\Phi_n\rangle = b_n^2 |\Phi_{n-1}\rangle + a_n |\Phi_n\rangle + |\Phi_{n+1}\rangle, \quad (84)$$

but, of course, it actually is

$$\langle\Phi_{n+1}|H|\Phi_n\rangle = \langle\Phi_{n+1}|\Phi_{n+1}\rangle = b_{n+1}^2 \langle\Phi_n|\Phi_n\rangle = \langle\Phi_n|H|\Phi_{n+1}\rangle. \quad (85)$$

The numerical implementation only requires keeping two N -dimensional vectors in memory. It is shown in table 1.

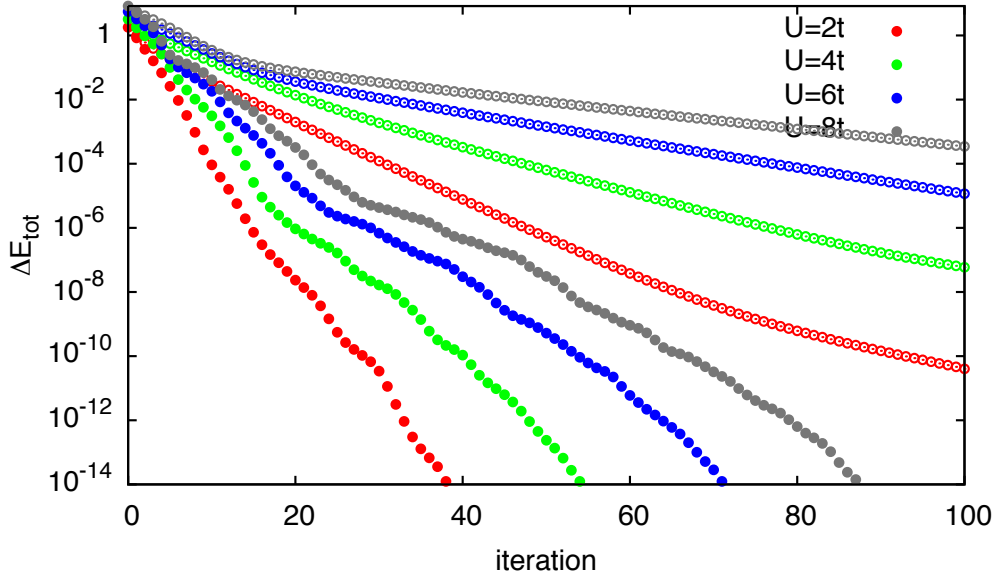


Fig. 2: Convergence of the lowest eigenvalue for a Lanczos iteration (full circles) compared to steepest-descent minimization (open circles) of a 10-site Hubbard-chain at half-filling, starting from a random initial vector. Due to the additional variational degrees of freedom, Lanczos converges significantly faster. Overall, convergence for the half-filled system gets harder for larger U , as the distance to the lowest excited states is reduced ($\sim t^2/U$) and the spectrum widens ($\sim U$). In all cases, convergence is reached after less than $L \approx 100$ Lanczos iterations, to be compared to the dimension $N=63\,504$ of the Hilbert space.

Diagonalizing (81), after a few tens to hundred iterations, the lowest eigenvalue of the tridiagonal representation of H on the Krylov space gives an excellent approximation to the ground-state energy of H in the full Hilbert space (Fig. 2). A formal estimate of the convergence was given by Kaniel and Paige [13]: For an $N+1$ -dimensional, symmetric matrix H with eigenvalues E_n , the lowest eigenvalue $\tilde{E}_0$ of the tridiagonal representation of H on the $(L+1)$ -dimensional Krylov space over $|v_0\rangle$ fulfills

$$\frac{\tilde{E}_0 - E_0}{E_N - E_0} \leq \left(\frac{\tan(\arccos(\langle \tilde{\Psi}_0 | \Psi_0 \rangle))}{T_L \left(1 + 2 \frac{E_1 - E_0}{E_N - E_1} \right)} \right)^2 \quad (86)$$

where $T_L(x)$ is the Chebyshev polynomial of order L and $\langle \tilde{\Psi}_0 | \Psi_0 \rangle$ the overlap of the Lanczos approximation to the ground-state $\tilde{\Psi}_0$ with the ground-state of H . Thus, if the initial state $|v_0\rangle$ is not orthogonal to the non-degenerate ground-state, convergence is exponential with a rate roughly increasing with the square root of the gap to the first excited state measured in units of the width of the spectrum.

The approximate ground-state vector is given by the linear combination

$$|\tilde{\Psi}_0\rangle = \sum_{n=0}^L \tilde{\psi}_{0,n} |v_n\rangle, \quad (87)$$

where $\tilde{\psi}_0$ is the ground-state vector of the $L+1$ -dimensional tridiagonal matrix (81). Instead of storing all $L+1$ basis vectors $|v_n\rangle$, we can restart the Lanczos iteration from the same $|v_0\rangle$,

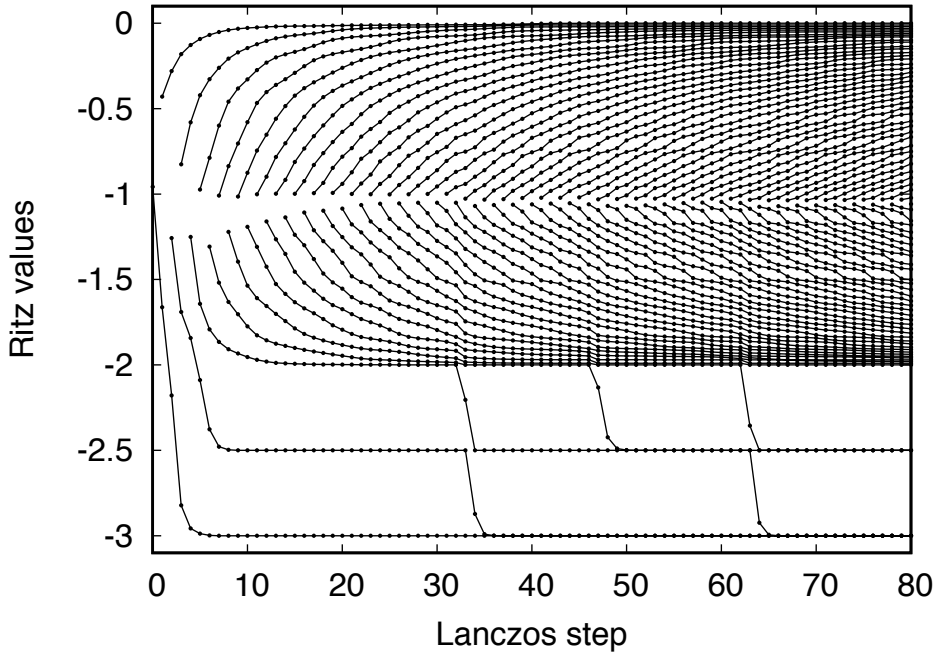


Fig. 3: *Lanczos method for a matrix with eigenvalues $-3, -3, -2.5, -2, -1.99, -1.98, \dots, -0.01, 0$. For clarity the n -th lowest/highest Lanczos eigenvalues are connected by lines. Note how the eigenvalues of successive steps are nested as described by (61). Because of the gap the lowest eigenvalue is reached much faster than the highest. For the degenerate eigenspace of value -3 only the state proportional to the projection of $|v_0\rangle$ on that space is found. Well after convergence of the low eigenvalues orthogonality is lost and ghost states appear.*

accumulating the sum (87) iteration by iteration. This only requires keeping one additional N -dimensional vector in memory.

So far we have tacitly assumed that the Krylov vectors $H^n|v_0\rangle$ are linearly independent. If not, there will be a vector $H|\tilde{v}_m\rangle$ that vanishes when orthogonalized to the previous states, i.e., $b_n = 0$. This means that the Krylov space $\text{span}(|v_0\rangle, |v_1\rangle, \dots, |v_m\rangle)$ is invariant under H , i.e., we have found an exact eigenspace of H . For a large matrix H it is quite unlikely to be that lucky. Still, as the Lanczos iteration approaches the ground-state, we encounter a similar situation: Close to an eigenstate, the functional (67) becomes almost stationary, i.e., the coefficients b_n almost vanish. Normalization of the very short vector $|\tilde{v}_n\rangle$ then amplifies numerical noise in that vector. This makes the numerical $|v_n\rangle$, which in theory should automatically be orthogonal to all $|v_m\rangle$ with $m < n-2$, actually have finite overlaps with these vectors. This loss of orthogonality manifests itself in the appearance of multiple copies of eigenvectors (ghost states) which are unrelated to the actual multiplicities of the eigenvalues. This makes the Lanczos method unpractical for tridiagonalizing dense matrices. For the ground-state the variational principle prevents severe problems from the loss of orthogonality. An example of the appearance of ghost states is shown in figure 3.

If we want to reliably obtain excited states, we need to explicitly orthogonalize to the previous basis states. This leads to the Lanczos method with complete reorthogonalization [13]. A similar orthogonalization is performed in the Arnoldi method [14], which, however, is devised for unsymmetric matrices.

4.2 Spectral functions

The Lanczos method gives excellent approximations to the largest and smallest eigenvalues, but, as seen in Fig. 3, not for the bulk of the spectrum. It therefore looks ill suited for determining functions that depend on the excited states, like the Lehmann representation

$$G_c(z) = \left\langle \Psi_c \left| \frac{1}{z - H} \right| \Psi_c \right\rangle = \sum_{n=0}^N \frac{\langle \Psi_c | \Psi_n \rangle \langle \Psi_n | \Psi_c \rangle}{z - E_n} \quad (88)$$

which, in terms of the eigenstates on the Krylov space $\mathcal{K}^L(|\Psi_c\rangle)$, would be written as

$$\check{G}_c(z) = \left\langle \Psi_c \left| \frac{1}{z - \check{H}_c} \right| \Psi_c \right\rangle = \sum_{n=0}^L \frac{\langle \Psi_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \Psi_c \rangle}{z - \check{E}_n}. \quad (89)$$

This is straightforward to calculate: We run L Lanczos iterations, starting from the (normalized) vector $|\Psi_c\rangle$, to create the tridiagonal $\check{H}_c$. The matrix element of the resolvent is the top left matrix element of the inverse of

$$z - \check{H}_c = \begin{pmatrix} z-a_0 & -b_1 & 0 & 0 & \cdots & 0 & 0 \\ -b_1 & z-a_1 & -b_2 & 0 & \cdots & 0 & 0 \\ 0 & -b_2 & z-a_2 & -b_3 & \cdots & 0 & 0 \\ 0 & 0 & -b_3 & z-a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & z-a_{L-1} & -b_L \\ 0 & 0 & 0 & 0 & \cdots & -b_L & z-a_L \end{pmatrix}. \quad (90)$$

This is easily determined, partitioning the matrix as indicated

$$z - \check{H}_c = \begin{pmatrix} z-a_0 & B^{(1)T} \\ B^{(1)} & z-\check{H}_c^{(1)} \end{pmatrix} \quad (91)$$

and inverting the block-matrix, giving, see appendix A.3,

$$[(z-\check{H}_c)^{-1}]_{00} = \left(z - a_0 - B^{(1)T} (z-\check{H}_c^{(1)})^{-1} B^{(1)} \right)^{-1} = \left(z - a_0 - b_1^2 [(z-\check{H}_c^{(1)})^{-1}]_{00} \right)^{-1}.$$

Repeating inversion by partitioning for the submatrices $\check{H}^{(n)}$ we obtain the continued fraction

$$\check{G}_c(z) = [(z - \check{H}_c)^{-1}]_{00} = \frac{1}{z - a_0 - \frac{b_1^2}{z - a_1 - \frac{b_2^2}{z - a_2 - \cdots}}}, \quad (92)$$

which terminates with $-b_L^2/(z-a_L)$. We find the spectral representation (89) by diagonalizing the Lanczos matrix $\check{H}_c$ giving us the $L+1$ eigenvalues $\check{E}_n$ and eigenvectors $\check{\psi}_n$. Since

$$|\check{\Psi}_n\rangle = \sum_{l=0}^L \check{\psi}_{n,l} |v_l\rangle \quad (93)$$

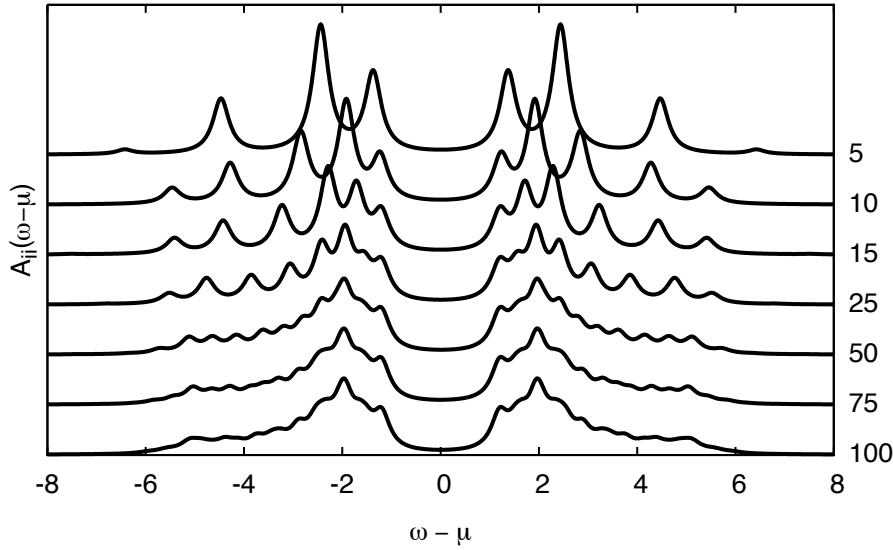


Fig. 4: Convergence of the spectral function with increasing number of Lanczos steps, $L=5, 10, 15, 25, 50, 75$, and 100 , for a 14-site Hubbard chain with $U = 5t$ at half filling. With increasing L , more and more moments of the photoemission and inverse photoemission part of the spectrum are reproduced correctly.

the matrix elements are given by $\langle \check{\Psi}_n | \Psi_c \rangle = \check{\psi}_{n,0}$. Thus

$$\check{G}_c(z) = \sum_{n=0}^L \frac{|\check{\psi}_{n,0}|^2}{z - \check{E}_n}. \quad (94)$$

The spectral function $\check{A}(\omega \pm i\eta) = \mp \frac{1}{\pi} \text{Im } \check{G}(\omega \pm i\eta)$ obtained this way, surprisingly, converges very quickly. An example is shown in figure 4.

To understand how the $L+1$ eigenstates of $\check{H}$ can represent the full spectrum so well, we consider the moments of the spectral function

$$\int_{-\infty}^{\infty} d\omega \omega^m \check{A}(\omega) = \sum_{n=0}^L |\check{\psi}_{n,0}|^2 \check{E}_n^m = \sum_{n=0}^L \langle \Psi_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \Psi_c \rangle \check{E}_n^m = \langle \Psi_c | \check{H}^m | \Psi_c \rangle. \quad (95)$$

Since $\check{H}$ is the projection of H onto the Krylov space $\mathcal{K}^L(|\Psi_c\rangle)$, we have $\check{H}^m |\Psi_c\rangle = H^m |\Psi_c\rangle$ for $m \leq L$. Thus the Lanczos representation $\check{A}(z)$ correctly reproduces the first $2L+1$ moments of the spectral function $A(z)$. A further Lanczos step adds one new level to the continued fraction (92), leaving all previous terms unchanged. $b_m^2 = 0$ then implies that the continued fraction terminates, and all moments are given correctly. A near vanishing $b_m^2 \approx 0$, which gives rise to the loss of orthogonality of the Lanczos vectors, for the spectral function merely means that further terms in the continued fraction hardly contribute any more.

So far we have considered diagonal elements of the resolvent. Off-diagonal matrix elements

$$G_{c_1, c_2}(z) = \left\langle \Psi_{c_2} \left| \frac{1}{z - H} \right| \Psi_{c_1} \right\rangle \quad (96)$$

are easily obtained by considering the diagonal elements for the linear combinations

$$\left\langle \Psi_{c_1} \pm \Psi_{c_2} \left| \frac{1}{z - H} \right| \Psi_{c_1} \pm \Psi_{c_2} \right\rangle = G_{c_1, c_1}(z) \pm G_{c_1, c_2}(z) \pm G_{c_2, c_1}(z) + G_{c_2, c_2}(z). \quad (97)$$

A Appendices

A.1 Non-orthonormal basis

A general one-electron basis of functions $|\chi_n\rangle$ will have an overlap matrix $S_{nm} = \langle\chi_n|\chi_m\rangle$ that is positive definite (and hence invertible) and hermitian. The completeness relation is

$$\mathbb{1} = \sum_{k,l} |\chi_k\rangle (S^{-1})_{kl} \langle\chi_l|. \quad (98)$$

With it we can easily write the Schrödinger equation $\hat{H}|v\rangle = \varepsilon|v\rangle$ in matrix form

$$\sum_k \underbrace{\langle\chi_i|H|\chi_k\rangle}_{=:H_{ik}} \underbrace{\sum_l (S^{-1})_{kl} \langle\chi_l|v\rangle \langle\chi_i|\hat{H}|v\rangle}_{=:v_k} = \varepsilon \langle\chi_i|v\rangle = \varepsilon \sum_k \underbrace{\langle\chi_i|\chi_k\rangle}_{=:S_{ik}} \underbrace{\sum_l (S^{-1})_{kl} \langle\chi_l|v\rangle}_{=:v_k}. \quad (99)$$

Collecting all components this becomes the generalized eigenvalue problem $\mathbf{H}\mathbf{v} = \varepsilon\mathbf{S}\mathbf{v}$. From the solution $\mathbf{v}$ we can easily construct $|v\rangle = \sum v_k |\chi_k\rangle$ [15]. It is, however, often more convenient to have an orthonormal basis, so that we do not have to deal with the overlap matrices in the definition of the second quantized operators or the generalized eigenvalue problem.

To orthonormalize the basis $\{|\chi_n\rangle\}$, we need to find a basis transformation $\mathbf{T}$ such that

$$|\varphi_n\rangle := \sum_m |\chi_m\rangle T_{mn} \quad \text{with} \quad \langle\varphi_n|\varphi_m\rangle = \delta_{mn}. \quad (100)$$

This implies that $\mathbf{T}^\dagger \mathbf{S} \mathbf{T} = \mathbb{1}$, or equivalently $\mathbf{S}^{-1} = \mathbf{T} \mathbf{T}^\dagger$. This condition does not uniquely determine $\mathbf{T}$. In fact there are many orthonormalization techniques, e.g., Gram-Schmidt orthonormalization or Cholesky decomposition.

Usually we will have chosen the basis functions $|\chi_n\rangle$ for a physical reason, e.g., atomic orbitals, so that we would like the orthonormal basis functions to be as close to the original basis as possible, i.e., we ask for the basis transformation $\mathbf{T}$ that minimizes

$$\begin{aligned} \sum_n \left\| |\varphi_n\rangle - |\chi_n\rangle \right\|^2 &= \sum_n \left\| \sum_m |\chi_m\rangle (T_{mn} - \delta_{mn}) \right\|^2 \\ &= \text{Tr} (\mathbf{T}^\dagger - \mathbb{1}) \mathbf{S} (\mathbf{T} - \mathbb{1}) = \text{Tr} (\underbrace{\mathbf{T}^\dagger \mathbf{S} \mathbf{T}}_{=\mathbb{1}} - \mathbf{T}^\dagger \mathbf{S} - \mathbf{S} \mathbf{T} + \mathbf{S}). \end{aligned} \quad (101)$$

Given an orthonormalization $\mathbf{T}$, we can obtain any other orthonormalization $\tilde{\mathbf{T}}$ by performing a unitary transformation, i.e., $\tilde{\mathbf{T}} = \mathbf{T} \mathbf{U}$. Writing $\mathbf{U} = \exp(i\lambda \mathbf{M})$ with $\mathbf{M}$ a Hermitian matrix, we obtain the variational condition

$$0 \stackrel{!}{=} \text{Tr} (+i\mathbf{M} \mathbf{T}^\dagger \mathbf{S} - i\mathbf{S} \mathbf{T} \mathbf{M}) = i \text{Tr} (\mathbf{T}^\dagger \mathbf{S} - \mathbf{S} \mathbf{T}) \mathbf{M}, \quad (102)$$

which is fulfilled for $\mathbf{S} \mathbf{T} = \mathbf{T}^\dagger \mathbf{S}$, i.e., $\mathbf{S} \mathbf{T}^2 = \mathbf{T}^\dagger \mathbf{S} \mathbf{T} = \mathbb{1}$. The second variation at $\mathbf{T} = \mathbf{S}^{-1/2}$

$$\frac{1}{2} \text{Tr} (\mathbf{M}^2 \mathbf{S}^{1/2} + \mathbf{S}^{1/2} \mathbf{M}^2) > 0 \quad (103)$$

is positive, since $\mathbf{S}$ and the square of the hermitian matrix $\mathbf{M}$ are both positive definite. Hence the Löwdin symmetric orthogonalization [16]

$$\mathbf{T}_{\text{Löwdin}} = \mathbf{S}^{-1/2} \quad (104)$$

minimizes the modification of the basis vectors due to orthogonalization.

A.2 Useful commutation relations

Expressions of commutators of products of operators can be derived by adding and subtracting terms that differ only in the position of one operator, e.g.,

$$\begin{aligned}
 [A_1 A_2 \cdots A_N, B] &= A_1 A_2 \cdots A_N B - B A_1 A_2 \cdots A_N \\
 &= A_1 A_2 \cdots A_N B - A_1 A_2 \cdots B A_N \\
 &\quad + A_1 A_2 \cdots B A_N - A_1 \cdots B A_{N-1} A_N \\
 &\quad + \cdots \\
 &\quad + A_1 B A_2 \cdots A_N - B A_1 A_2 \cdots A_N \\
 &= \sum_i A_1 \cdots A_{i-1} [A_i, B] A_{i+1} \cdots A_N
 \end{aligned}$$

The following special cases are particularly useful

$$\begin{aligned}
 [AB, C] &= A [B, C] + [A, C] B \\
 &= A \{B, C\} - \{A, C\} B
 \end{aligned}$$

$$\begin{aligned}
 [A, BC] &= B [A, C] + [A, B] C \\
 &= [A, B] C + B [A, C] \\
 &= \{A, B\} C - B \{A, C\}
 \end{aligned}$$

$$\begin{aligned}
 [AB, CD] &= A [B, C] D + AC [B, D] + [A, C] DB + C [A, D] B \\
 &= A \{B, C\} D - AC \{B, D\} + \{A, C\} DB - C \{A, D\} B
 \end{aligned}$$

Important examples are

$$\begin{aligned}
 [c_i^\dagger c_j, c_\gamma^\dagger] &= \langle j | \gamma \rangle c_i^\dagger \\
 [c_i^\dagger c_j, c_\gamma] &= -\langle i | \gamma \rangle c_j
 \end{aligned}$$

For the commutator of products of creation and annihilation operators appearing in one- and two-body operators we find

$$[c_i^\dagger c_j, c_\alpha^\dagger c_\beta] = [c_i^\dagger c_j, c_\alpha^\dagger] c_\beta + c_\alpha^\dagger [c_i^\dagger c_j, c_\beta] = \langle j | \alpha \rangle c_i^\dagger c_\beta - \langle \beta | i \rangle c_\alpha^\dagger c_j$$

and

$$[c_i^\dagger c_j^\dagger c_k c_l, c_\alpha^\dagger c_\beta] = \langle l | \alpha \rangle c_i^\dagger c_j^\dagger c_k c_\beta + \langle k | \alpha \rangle c_i^\dagger c_j^\dagger c_\beta c_l - \langle \beta | j \rangle c_i^\dagger c_\alpha^\dagger c_k c_l - \langle \beta | i \rangle c_\alpha^\dagger c_j^\dagger c_k c_l$$

A.3 Downfolding

To integrate-out high-energy degrees of freedom, we partition the Hilbert space of the full system into states of interest (low-energy states) and ‘other’ states, which will be integrated out. The Hamiltonian is then written in blocks

$$H = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix}, \quad (105)$$

where H_{00} is the Hamiltonian restricted to the states of interest (reduced Hilbert space), H_{11} the Hamiltonian for the ‘other’ states, and the T matrices describe transitions between the two subspaces. The resolvent is partitioned likewise

$$G(\omega) = (\omega - H)^{-1} = \begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix}^{-1}. \quad (106)$$

Its elements are easily determined by solving the system of two linear matrix equations

$$\begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix} \begin{pmatrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{pmatrix}, \quad (107)$$

keeping track of the order of the sub-matrix products. The resolvent on the reduced Hilbert space is thus given by

$$G_{00}(\omega) = \left(\omega - \underbrace{(H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10})}_{=H_{\text{eff}}(\omega)} \right)^{-1}. \quad (108)$$

This expression looks just like the resolvent for a Hamiltonian H_{eff} on the reduced Hilbert space. This effective Hamiltonian describes the physics of the full system, but operates only on the small reduced Hilbert space: For an eigenvector $H|\Psi\rangle = E|\Psi\rangle$ on the full Hilbert space

$$H|\Psi\rangle = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix} \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} = E \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} \quad (109)$$

its projection $|\Psi_0\rangle$ onto the reduced Hilbert space is an eigenstate of $H_{\text{eff}}(E)$. On the other hand, we can construct the full eigenstate from a solution $H_{\text{eff}}(E)|\Psi_0\rangle = E|\Psi_0\rangle$ on the reduced Hilbert space by upfolding $|\Psi\rangle \propto (\mathbb{1} + (E - H_{11})^{-1} T_{10})|\Psi_0\rangle$.

Of course, this drastic simplification comes at a price: the effective Hamiltonian is energy dependent. If the hopping matrix elements in T_{01} are small, and/or the states in the part of the Hilbert space that has been integrated out are energetically well-separated from the states that are explicitly considered, this energy dependence can, to a good approximation, be neglected. We can then replace ω by some characteristic energy ε_0 for the states in the reduced Hilbert space to obtain an energy-independent Hamiltonian

$$H_{\text{eff}}(\omega) = H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10} \approx H_{00} + T_{01}(\varepsilon_0 - H_{11})^{-1} T_{10} = H_{\text{eff}}(\varepsilon_0) \quad (110)$$

that gives a good description of the electrons in the reduced Hilbert space, i.e., the states with an energy close to ε_0 . Expanding $(\omega - H_{11})^{-1}$ about ε_0 , we can systematically improve the approximation (linear and higher-order methods).

References

- [1] P. Forrest: *The Identity of Indiscernibles*, in The Stanford Encyclopedia of Philosophy
<http://plato.stanford.edu/entries/identity-indiscernible/>
- [2] W. Pauli: *The connection between spin and statistics*, Phys. Rev. **58**, 715 (1940)
- [3] R.F. Streater and A.S. Wightman: *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964)
- [4] P.-O. Löwdin: *Quantum Theory of Many-Particle Systems I*, Phys. Rev. **97**, 1474 (1955)
- [5] P. Jordan and O. Klein: *Zum Mehrkörperproblem in der Quantenmechanik*, Z. Physik **45**, 751 (1927)
- [6] P. Jordan and E. Wigner: *Über das Paulische Äquivalenzverbot*, Z. Physik **47**, 631 (1928)
- [7] R. Peierls: *Surprises in Theoretical Physics* (Princeton University Press, 1979)
- [8] H.S. Warren: *Hacker's Delight*, 2nd ed. (Addison-Wesley, 2013)
- [9] L.V. Kantorovich: *Functional analysis and applied mathematics*, Uspekhi Mat. Nauk, **3**, 89 (1948)
- [10] M.R. Hestenes and W. Karush: *A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix*, J. Res. Nat. Bureau Standards **47**, 45 (1951)
- [11] A.N. Krylov: *On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined*, Izv. Akad. Nauk SSSR, Otd. Mat. Estest. **7**, 491 (1931)
- [12] C. Lanczos: *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators*, J. Res. Nat. Bur. Stand. **49**, 255 (1950)
- [13] G.H. Golub and C.F. van Loan: *Matrix Computations* (Johns Hopkins Univ. Press, 1996)
- [14] W.E. Arnoldi: *The Principle of Minimized Iteration in the Solution of the Matrix Eigenproblem*, Quarterly of Applied Mathematics **9**, 17 (1951)
- [15] A. Szabo and N.S. Ostlund: *Modern Quantum Chemistry* (McGraw Hill, New York, 1989)
- [16] P.-O. Löwdin: *On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals*, J. Chem. Phys. **18**, 365 (1950)

5 DMRG as Impurity Solver

Ulrich Schollwöck

Department of Physics, University of Munich

Theresienstrasse 37, 80333 München

Contents

1	Tensor networks and embedding techniques	2
2	Tensor networks in a nutshell	4
3	DMFT in a nutshell	8
4	Bath representations: chains, stars, forks and trees	11
5	Time evolution: Finite temperatures, linear prediction, time steppers	14
6	Test case: LiV_2O_4	18
7	Real time and complex time evolution	23
8	Outlook	27

1 Tensor networks and embedding techniques

In this lecture, I will bring together two widely used approaches to the quantum many-body problem which take quite unrelated perspectives. After a long prehistory, tensor networks firmly entered the realm of strongly correlated systems as described by the Heisenberg and Hubbard model Hamiltonians in the form of White's density-matrix renormalization-group (DMRG) in 1992 [1]. This method for determining ground states and low-lying excited states was quickly recognized to be, ignoring some subtle details, nothing but an efficient way of variationally determining the lowest-lying eigenstates of a Hamiltonian within a very special class of ansatz states, so-called matrix-product states (MPS). The Hamiltonian is treated exactly, the state space is strongly restricted. Powerful generalizations and improvements have been brought forward over the last 30 years [2, 3].

At the same time, earlier insights into simplifications of the quantum many-body problem in large dimensions (more precisely, in systems with large coordination numbers) [4] led Georges and Kotliar to the formulation of dynamical mean-field theory (DMFT) as an embedding approach to the many-body problem [5–9]. Unlike tensor network methods, it approximates the interaction part of the Hamiltonian by treating a manageable subsystem exactly while modeling the rest of the system by an effective non-interacting “bath” and computes physical observables without explicitly calculating the many-body wave function. DMFT crucially relies on a self-consistency condition for the determination of the bath. Its earliest predecessor is the well-known Weiss mean-field theory of ferromagnetism, where self-consistency is achieved at the level of spin magnetization: The interaction of one particular spin to the others is modeled by a coupling to an effective magnetic field constant in space and time, and the self-consistency condition is that the magnetization of that spin in the effective field must be the same as the magnetization leading to that effective field; after all, the chosen spin was not special. In DMFT for electronic problems, one electronic orbital (site) is picked out and couples through electronic hopping to other orbitals (sites). The “bath” is now time-dependent (dynamical) as opposed to the effective field in Weiss theory, and it turns out that for many problems this is the decisive step forward. Just like the Weiss field, which implies that we consider the other spins to be non-interacting, the bath consists of non-interacting electrons, making the problem more treatable. The self-consistency condition of DMFT, to be detailed below, is that the local self-energy of the chosen orbital is the same as the local self-energy of the lattice; again, the chosen orbital was not special. As in the case of DMRG, original DMFT has been the first step towards highly useful generalizations of the idea of dynamical embedding, and in both cases these generalizations are the subject of intense ongoing research [7–9]. What is particularly interesting about DMFT is that the approach also works for real systems, with the band structure of the material under study calculated by density functional theory (DFT). The DFT band structure has issues with strong correlations due to local interactions, but this part is dealt with by DMFT. The DFT+DMFT method, as we may call it, has emerged as a very powerful method to study real strongly correlated materials beyond simple Hubbard and Heisenberg Hamiltonians [9].

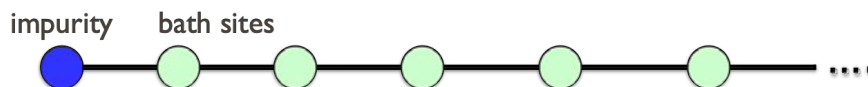


Fig. 1: *Impurity and bath in the NRG framework: The interacting impurity (blue) sits at the end of a one-dimensional chain. All other sites correspond to orbitals with non-interacting electrons, which have an on-site energy and hop to the nearest-neighbor; they form the bath.*

These seemingly unrelated approaches to the quantum many-body problem come together because DMFT (and generalizations) requires for the self-consistency condition the calculation of the local Green function of the chosen interacting orbital, embedded in the non-interacting bath. While this is a much simpler problem than the full many-body problem, this calculation still represents the core numerical challenge and limitation of DMFT, where actual DMFT calculations spend most of their time. A substantial fraction of methodological progress has been related to improving schemes for calculating that Green function. As the setup resembles that of an interacting impurity in a solid with bands of (non-interacting) electrons, well-known from the famous Kondo problem, one speaks colloquially of the “impurity problem” that has to be solved within DMFT and of the “impurity Green function”; the methods are called “impurity solvers”. We will use DMRG and more generally tensor network techniques, of which it is one, as impurity solver, hence the title of the lecture.

Most numerical techniques for solving quantum many-body problems have been tried and are used regularly for the impurity problem. Arguably the most versatile technique at the moment is continuous-time quantum Monte Carlo, which comes in two main incarnations, treating kinetic terms as a correction to the local interaction, or the local interaction as a correction to kinetic terms (CT-HYB and CT-INT) [10, 11]. We will speak of QMC-DMFT. Other methods are ED-DMFT and NRG-DMFT. In the former [12], the bath is represented by a short chain of non-interacting electronic sites, with dynamical quantities expressed through Matsubara (imaginary) frequencies (for details, see below); calculations are based on exact diagonalization (ED). The use of DMRG replacing ED is suggested by the fact that DMRG is able to treat larger systems than ED. Alternatively, one can express bath properties by using real frequencies and use the framework of the numerical renormalization group (NRG), which was specifically developed for the Kondo problem, to solve the impurity problem [13]. The use of DMRG replacing NRG is suggested by the fact that NRG leads to an effectively one-dimensional problem (Figure 1), where DMRG excels. The computational scaling of DMRG is also more advantageous for more complicated (e.g. multi-orbital) problems.

The idea of using tensor network (TN) algorithms, in particular DMRG, as an impurity solver, is therefore unsurprisingly not recent, but goes back to the early 2000s (see e.g. [14] for an early example; the next bigger steps came around 10 years later, see [15–18]). The two main approaches to TN-DMFT (real-time DMFT [15–17] and imaginary-time TN-DMFT [18]) are direct adaptations of NRG-DMFT and ED-DMFT. Despite this conceptual simplicity, TN-DMFT has faced numerous challenges and it is only recently that it has emerged not just as a serious competitor to other approaches, but even having a certain lead, in particular when used in cooperation with other DMFT impurity solvers.

I will set out by discussing some of the essentials of DMRG, MPS and tensor networks, to set the nomenclature and also in order to develop an understanding where the methodological bottlenecks lie. I will do the same for DMFT, but in a more compressed form as the progress I want to describe is essentially on the tensor network side; in fact, in our calculations we use the highly developed and refined TRIQS library [19,20] more or less as a black box for DMFT which in a self-consistency loop both generates the precise impurity problem to be solved and then processes the solution we provide. There is no change to DMFT involved except for the impurity problem module.

We will then discuss crucial improvements to earlier tensor network setups for impurity problems and more generally the calculation of spectral (Green) functions. They are at the heart of this lecture, and in fact topics of ongoing research.

Next, I will set these improvements to work, using the example of LiV_2O_4 [21]. The low-temperature properties of this material have puzzled the community for almost three decades, and numerous explanations have been put forward. All were hampered by the fact that these properties have no high-temperature counterpart and only emerge after a low-temperature crossover where several very small temperature scales are involved. In the past, DMFT (or rather, the impurity solvers) were not able to access the crossover regime and lower temperatures. This is now possible using TN solvers; at the higher temperatures also accessible to QMC solvers, agreement is excellent. Most importantly, we now have an understanding of the low-temperature behavior of LiV_2O_4 , and a methodology which should be able to solve similar problems quite easily.

I conclude by ongoing research which has not yet fully been integrated into TN-DMFT, but will be so very soon: can we make calculations of spectral functions and Green functions using real frequencies numerically less limited in reachable times, hence frequency resolution, by a smart use of time-evolution contours in the complex plane? Current results seem to be very promising [22].

2 Tensor networks in a nutshell

Before we dive into details, the key messages which will be required for the rest of the lecture, are: tensor networks yield low-entanglement approximations of quantum states; ground states and mixed thermal states are mostly weakly entangled, hence the power of tensor networks; real-time evolutions take us quickly out of this cosy corner, making long time evolutions, hence fine spectral resolution (which we require), a key challenge in the field.

Let us illustrate tensor networks by starting with the simplest such network, a matrix product state (MPS): let us consider the state space of a (one-dimensional) lattice with L sites, where the local Hilbert space of site i is spanned by site-local states $\{|\sigma_i\rangle\}$. The simplest local Hilbert spaces are the two degrees of a spin-1/2, $\{|\uparrow\rangle, |\downarrow\rangle\}$, or the four degrees of freedom of a local electronic orbital, $\{|\emptyset\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$. Let us call the number of local degrees of freedom d , here 2 or 4. The Hilbert space of the entire system, given by the tensor product of L local

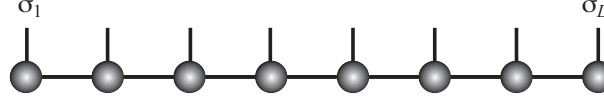


Fig. 2: A matrix product state (MPS), the simplest tensor network state (TNS): each site (orbital) is represented by a tensor (spherical blob), here with three indices, $M_{\alpha\beta}^{\sigma}$. The physical leg, indicated by a vertical line, represents the local state $|\sigma\rangle$; the two auxiliary legs, indicated by horizontal lines, are contracted over when linked between tensors. The remaining scalar is the expansion coefficient $c^{\sigma_1\sigma_2\ldots\sigma_L}$ in Eq. 1.

spaces, then has dimension d^L , and the most generic pure quantum state reads

$$|\Psi\rangle = \sum_{\sigma_1, \dots, \sigma_L} c^{\sigma_1\sigma_2\ldots\sigma_L} |\sigma_1\sigma_2\ldots\sigma_L\rangle \quad (1)$$

with complex coefficients $c^{\sigma_1\sigma_2\ldots\sigma_L}$. In bulk systems we are interested in the thermodynamic limit $L \rightarrow \infty$, making an exact treatment unfeasible in that limit, as d^L diverges exponentially. Matrix product states (MPS) form a subset of the full Hilbert space, where the expansion coefficients are given by the product of matrices M^{σ_i} : at each site i , there is one matrix for each local state $|\sigma_i\rangle$. The most general MPS then reads

$$|\Psi\rangle = \sum_{\sigma_1, \dots, \sigma_L} M^{\sigma_1} M^{\sigma_2} \ldots M^{\sigma_L} |\sigma_1\sigma_2\ldots\sigma_L\rangle. \quad (2)$$

As the matrices are multiplied, their dimensions must match between neighboring matrices; the first and last matrix must be row and column vectors, such that the total product is a scalar $c^{\sigma_1\sigma_2\ldots\sigma_L}$. We can immediately think about generalizations: we read the local objects as (three-legged) tensors $M_{\alpha\beta}^{\sigma}$ where the “auxiliary indices” α and β are contracted with neighboring tensors. We can have, e.g., five-legged tensors $M_{\alpha\beta\gamma\delta}^{\sigma}$, with four auxiliary indices, which link to four other tensors, sitting on the nearest-neighbor sites of a square lattice; ultimately the state coefficients result from contracting over all of the auxiliary indices. An MPS is shown in Figure 2.

It can be shown easily [3] by applying singular value decompositions (SVDs) to coefficients $c^{\sigma_1\sigma_2\ldots\sigma_L}$ that each state can be expressed exactly as an MPS. The price to pay is that towards the center, the matrices become exponentially large, with the largest dimension being $d^{L/2}$, so nothing seems to be gained. The interesting observation is that many quantum states can be expressed almost exactly by much smaller matrices that are limited to some dimension of at most $D \times D$, where D is somewhere in $O(100)$ to $O(1000)$, for which matrix multiplications are fast and efficient, as they scale as D^3 . The effective number of coefficients is then $O(dD^2L)$, no longer exponential in system size.

Why is this? It turns out (we only need the result) that if the matrices are chosen to obey certain normalization conditions (always possible), each chain can be partitioned between site ℓ and $\ell+1$ into parts A and B and the MPS be written as

$$|\Psi\rangle = \sum_{\alpha} s_{\alpha} |\alpha\rangle_A |\alpha\rangle_B \quad (3)$$

where $s_\alpha \geq 0$, $\sum_\alpha s_\alpha^2 = 1$ and the states

$$|\alpha\rangle_A = \sum_{\sigma_1 \dots \sigma_\ell} (M^{\sigma_1} M^{\sigma_2} \dots M^{\sigma_\ell})_{1,\alpha} |\sigma_1 \dots \sigma_\ell\rangle \quad (4)$$

$$|\alpha\rangle_B = \sum_{\sigma_{\ell+1} \dots \sigma_L} (M^{\sigma_{\ell+1}} M^{\sigma_{\ell+2}} \dots M^{\sigma_L})_{\alpha,1} |\sigma_{\ell+1} \dots \sigma_L\rangle \quad (5)$$

form orthonormal sets within A and B. This is a so-called Schmidt decomposition. Imagine you have an exact MPS, then the sum will run over an exponential number of values α . It turns out, however, that for many states of interest (ground states and low excitations of Hamiltonians with short-ranged interactions) the values of the (ordered) s_α decay very quickly to exponentially small values, and only a minuscule amount actually contributes to the state. In fact, the reduced density matrix for A reads

$$\hat{\rho}_A = \sum_\alpha s_\alpha^2 |\alpha\rangle_{AA} \langle\alpha|, \quad (6)$$

such that errors in observables become negligible if we restrict ourselves to the D largest values s_α , which imposes a compression condition on the matrices M^σ with very little actual loss. As the von Neumann entropy of entanglement of A and B is given by

$$S_{\text{vN}} = -\text{tr}_A \hat{\rho}_A \ln \hat{\rho}_A = -\sum_\alpha s_\alpha^2 \ln s_\alpha^2, \quad (7)$$

a quickly decaying spectrum of the s_α corresponds to weak entanglement between A and B. This link is not mathematically rigorous (it is easy to construct counterexamples [23]), but effectively holds for all states we are encountering in actual systems. The quickly decaying spectrum then reflects the so-called area law: while thermal entropy scales as the volume of a system, the von Neumann entropy of low-lying states (of gapful Hamiltonians) scales only as the surface area between A and B [24]. In one dimension, the surface is a point, and entanglement and the computational complexity of the state (which is roughly exponential in entanglement) do not increase, hence the extreme power of MPS for such systems. In two dimensions, entanglement scales as L for a system of size $L \times L$; the required matrix dimensions then grow exponentially in L , strongly limiting performance of MPS in two dimensions (other TNs do not have this issue, but are otherwise computationally much more demanding).

In this lecture, we are concerned with time-dependent Green functions, which require a time-evolution of a quantum state. It is also known that the states obeying the area law form a (non-empty!) set of measure zero in Hilbert space in the thermodynamic limit, which happens to contain the most interesting states for us. Time evolution will take us inevitably out of this comfortable, but minuscule corner of Hilbert space. In the worst case, entanglement will grow linearly in time, hence required matrix dimensions D grow exponentially in time. This very strongly limits the reachable times. In the case of spectral functions, the entanglement growth is more benign (logarithmic), but still dimensions D will grow with some power law in time, imposing a strong restriction on reachable times. This limitation, of course, limits the resolution we may achieve in frequency space and is therefore the central bottleneck of the following discussion.

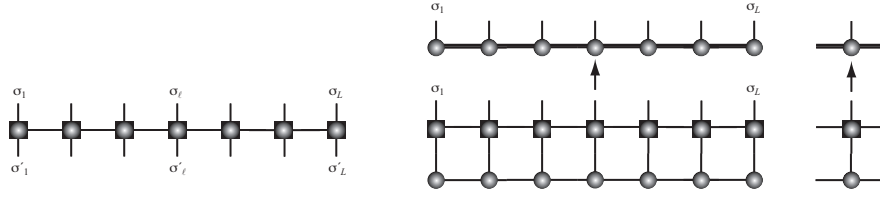


Fig. 3: A matrix product operator (MPO): on each site (orbital), we have a tensor (square blob), now with four indices, $W_{\alpha\beta}^{\sigma\sigma'}$. The physical legs, indicated by vertical lines, represents the ingoing and outgoing local states $|\sigma^{(\prime)}\rangle$; the two auxiliary legs, indicated by horizontal lines, are contracted over when linked between tensors. When applying an MPO (auxiliary leg dimension D_O) to an MPS, a new MPS results, but with auxiliary leg dimension D_OD .

Finding ground states within the ansatz class of MPS for a given Hamiltonian is what DMRG does, and we can treat this as a black box here. Similarly, there is a large number of methods for the time evolution of MPS [25], which are contained as black boxes in available tensor network codes. As we will push those to the limit, at least a minimal understanding is useful. We want to calculate $|\Psi(t)\rangle = e^{-i\hat{H}t}|\Psi(0)\rangle$ with results also at intermediate times.

The earliest algorithms [26–29] all built on the idea of the Trotter decomposition. In its simplest incarnation, it runs as follows: we decompose the time evolution into small timesteps Δt . If the Hamiltonian decomposes into local Hamiltonians acting on a few neighboring sites as $\hat{H} = \sum_j \hat{h}_j$, at least the neighboring local Hamiltonians will not commute in non-trivial problems, $[\hat{h}_j, \hat{h}_{j+1}] \neq 0$. A decomposition of the exponential is therefore not exact,

$$e^{-i\hat{H}\Delta t} = e^{-i\sum_j \hat{h}_j\Delta t} \approx \prod_j e^{-i\hat{h}_j\Delta t}, \quad (8)$$

but the error is of order $(\Delta t)^2$ and can be made arbitrarily small (and smarter Trotter decompositions exist).

The more relevant error, however, comes from the increasing entanglement of the time-evolving state. Technically, this manifests itself as follows (for details, see [3, 25]): the factorized time-evolution operator can be expressed as a matrix product operator (MPO), which looks like an MPS, but with two physical legs, one for the ingoing physical state and one for the outgoing state. It can be applied to an MPS by contracting over physical legs, and a new MPS results with matrix dimension D_OD , where D_O is the matrix dimension of the MPO. This state (its MPS representation) has to be compressed back to some matrix dimension D' , at an acceptable loss of accuracy, to keep it numerically tractable. Growing entanglement manifests itself by growing D' if we want to keep a certain accuracy, or declining accuracy if we limit D' to some maximum value. Either way, this is the key bottleneck and it represents physics, not some algorithmic short-coming.

When using more recent time-evolution methods such as TDVP (time-dependent variational principle), which has become pretty much the standard by now [25], both the time-stepping error and the entanglement growth problem persists, but TDVP is much better at treating longer-ranged interactions and overall substantially faster. Another approach, the Krylov space method,

which we will discuss below, will allow us to circumvent certain unexpected problems arising from time-stepping; the entanglement issue persists. We will discuss one way out of this generic problem, linear prediction, further below; it is quite specific to Green functions and spectral functions, but this is what we want to calculate.

3 DMFT in a nutshell

Let us now turn to a similarly concise presentation of DMFT, as it has been presented numerous times in this lecture series over the years (most recently in [8, 9]). Even more importantly, the fine details do not matter to us and are best taken care of by the mentioned standard open-source packages. We mainly discuss where and how TNS (mostly MPS) come into play.

Imaginary-time TN-DMFT. This approach simply adapts the approach of [12]. There, frequency-dependent quantities are expressed as a discrete set of values at the imaginary Matsubara frequencies $i\omega_n$, where $\omega_n = \pi(2n+1)/\beta$ with $\beta = (k_B T)^{-1}$; they become dense as we approach $T = 0$.

Let us start from the Green function of the non-interacting part of the Hubbard Hamiltonian which has a dispersion relation $\varepsilon_{\mathbf{k}}$ (which we may have obtained from band structure, as given by DFT). The chemical potential μ is chosen to give the correct filling. Then $G_{\text{lattice}}^0(\mathbf{k}, i\omega_n) = (i\omega_n + \mu - \varepsilon_{\mathbf{k}})^{-1}$, a standard result. When switching on the interaction, this modifies the Green function through the self-energy $\Sigma_{\text{lattice}}(\mathbf{k}, i\omega_n)$. Let us assume we know it, then $G_{\text{lattice}}(\mathbf{k}, i\omega_n) = (i\omega_n + \mu - \varepsilon_{\mathbf{k}} - \Sigma_{\text{lattice}}(\mathbf{k}, i\omega_n))^{-1}$. By a Fourier transformation, we can find the local lattice Green function at position 0 (which we take to be the location of the impurity), which we call G_{loc} , as

$$G_{\text{loc}}(i\omega_n) = \int d\mathbf{k} \frac{1}{i\omega_n + \mu - \varepsilon_{\mathbf{k}} - \Sigma_{\text{lattice}}(\mathbf{k}, i\omega_n)}. \quad (9)$$

The central approximation of DMFT is that Σ_{lattice} does *not* depend on $\mathbf{k}$, $\Sigma_{\text{lattice}} = \Sigma_{\text{lattice}}(i\omega_n)$, and that the momentum dependence of the Green function only enters through the non-interacting dispersion relation $\varepsilon_{\mathbf{k}}$. Then

$$G_{\text{loc}}(i\omega_n) = \int d\mathbf{k} \frac{1}{i\omega_n + \mu - \varepsilon_{\mathbf{k}} - \Sigma_{\text{lattice}}(i\omega_n)}, \quad (10)$$

where we still have to find $\Sigma_{\text{lattice}}(i\omega_n)$, which is done through self-consistency: the local self-energy of the lattice must be the same as the local self-energy of the interacting impurity whose electrons are exchanged with the bath,

$$\Sigma_{\text{lattice}}(i\omega_n) \stackrel{!}{=} \Sigma_{\text{imp}}(i\omega_n) \quad (11)$$

(Of course, we have to have $\Sigma_{\text{imp}}(i\omega_n)$; at some point, we must start with an educated guess.)

Now the local lattice Green function must be close to the local Green function of the impurity, which is no special site, just one we picked, $G_{\text{imp}}(i\omega_n) \stackrel{!}{=} G_{\text{loc}}(i\omega_n)$. The impurity Green function reads

$$G_{\text{imp}}(i\omega_n) = \frac{1}{i\omega_n + \mu - \Lambda(i\omega_n) - \Sigma_{\text{imp}}(i\omega_n)}, \quad (12)$$

where the hybridization $\Lambda(i\omega_n)$ encodes the hopping of electrons between impurity and bath. As we have both $G_{\text{imp}}(i\omega_n)$ and $\Sigma_{\text{imp}}(i\omega_n)$, we can determine $\Lambda(i\omega_n)$ from Equation (12). It can be shown that for L_b non-interacting bath sites with on-site energies ε_l and hopping amplitude V_l between impurity and bath site l , the hybridization function is given by

$$\Lambda_{\text{discr}}(i\omega_n) = \sum_{l=1}^{L_b} \frac{|V_l|^2}{i\omega_n - \varepsilon_l}. \quad (13)$$

We turn this relationship around and find for given L_b (typically small, $O(10)$ or so) V_l and ε_l such that $\Lambda_{\text{discr}}(i\omega_n)$ approximates $\Lambda(i\omega_n)$ as closely as possible, i.e., minimizes

$$\chi = \frac{1}{N} \sum_{n=1}^N |\Lambda(i\omega_n) - \Lambda_{\text{discr}}(i\omega_n)|^2 \quad (14)$$

or some similar cost function (one may place different emphasis on the Matsubara frequencies, typically only the smaller Matsubara frequencies are considered, as the large-frequency behavior is universal).

Now comes the moment of the TN impurity solver. We consider the Hamiltonian $\hat{H}$ formed from the impurity site with its local microscopic interaction at position 0 and the bath sites at positions 1 through L_b , with on-site energies ε_l and hopping V_l between 0 and l . For this Hamiltonian, we use DMRG to find either the ground state $|\Psi_{\text{GS}}\rangle$ (for a $T = 0$ calculation) or the thermal density operator (for $T > 0$) by an imaginary time evolution, as described below.

Starting from this (mixed) state, we determine the retarded Green function of the impurity in imaginary time as (here written for the ground state; the thermal expression is a simple adaptation, because we will represent mixed states as pure states, see below; we also suppress spin indices)

$$G^p(\tau) = -\Theta(\tau) \langle \Psi_{\text{GS}} | \hat{c}_0 e^{-(\hat{H}-E_0)\tau} \hat{c}_0^\dagger | \Psi_{\text{GS}} \rangle, \quad G^h(\tau) = \Theta(-\tau) \langle \Psi_{\text{GS}} | \hat{c}_0^\dagger e^{(\hat{H}-E_0)\tau} \hat{c}_0 | \Psi_{\text{GS}} \rangle \quad (15)$$

such that

$$G_{\text{imp}}(\tau) = G^p(\tau) + G^h(\tau), \quad (16)$$

where $\Theta(\tau)$ is the Heaviside step function. At this moment, the use of the TN impurity solver is complete. We obtain the impurity Green function in Matsubara frequencies as

$$G_{\text{imp}}(i\omega_n) = \int_{-\infty}^{\infty} e^{-i\omega_n \tau} G_{\text{imp}}(\tau). \quad (17)$$

We now obtain the self-energy $\Sigma_{\text{imp}}(i\omega_n)$ from Dyson's equation

$$\Sigma_{\text{imp}}(i\omega_n) = (G_{\text{imp}}^0)^{-1}(i\omega_n) - G_{\text{imp}}^{-1}(i\omega_n), \quad (18)$$

where $G_{\text{imp}}^0(i\omega_n)$ is the impurity Green function if we switch the interaction on the impurity site off. We have closed the self-consistency loop! It turns out that we need $O(10)$ iterations of the loop to converge the result. Imaginary-time TN-DMFT was pioneered in [18], and has been

applied in various cases, culminating in [21], which will be detailed as a test case below. The key advantage of imaginary-time TN-DMFT is that imaginary time-evolution is much more benign than real-time evolution and that the bath sizes are quite small. The downside is that any quantity on the real-frequency axis can only be obtained by analytical continuation from the imaginary to the real axis, which is numerically unstable and also the key issue with QMC-DMFT, in particular for very low frequencies.

Real-time TN-DMFT. Historically this approach, which is in fact numerically more challenging, predates imaginary-time TN-DMFT. It was inspired by NRG-DMFT, which works on the real-frequency (energy) axis and does not suffer from the drawbacks of analytical continuation. NRG-DMFT also excels at the lowest temperatures; however, it is handicapped by structural constraints that lead to an explosion of computation time for the study of interesting materials and is not good at larger frequencies. As NRG can be formulated as a very special MPS algorithm, the use of real-time TN-DMFT promises better performance.

Compared to imaginary-time TN-DMFT, the main changes are that local bath and impurity Green functions, the local bath and impurity self-energies, and the hybridization function are functions of ω instead of $i\omega_n$. The hybridization function is represented by a bath in a different way: let us take an electronic band of some bandwidth $2D$ (D not to be confused with the matrix dimension D), and discretize it into N_b intervals I_l which in the simplest case have equal size, but different choices are possible (in NRG for instance, intervals become exponentially small close to the Fermi edge, a key ingredient for the success of NRG). The local site energy and the hopping between impurity and bath energy interval is given by

$$|V_l|^2 = - \int_{I_l} d\omega \frac{1}{\pi} \text{Im} \Delta(\omega) \quad (19)$$

$$\varepsilon_l = - \frac{1}{|V_l|^2} \int_{I_l} d\omega \frac{1}{\pi} \omega \text{Im} \Delta(\omega) \quad (20)$$

The required bath sizes N_b are *much* larger, of the order of a few hundred sites, because we have to discretize energies on a very fine grid. We cannot really expect information lost by coarse-graining in energy (frequency) space to miraculously reappear. Also, the spectral function $A(\omega)$, essentially the imaginary part of the Green function, contains a lot of fine structure to be resolved: in principle, it consists of a dense set of δ -peaks for $T \rightarrow 0$. (All this information is smoothened out and hidden across all imaginary frequencies when analytically continuing from the real to the imaginary axis, which makes the reverse procedure so difficult.)

From the pure TN perspective, the key difference is that time evolutions are now in real time, e.g.,

$$G^p(t) = \langle \Psi_{\text{GS}} | \hat{c}_0 e^{-i(\hat{H}-E_0)t} \hat{c}_0^\dagger | \Psi_{\text{GS}} \rangle. \quad (21)$$

Numerically, this is very challenging due to strong entanglement growth, hence achievable simulation times are quite limited. Nevertheless, the method has been successfully applied to two-orbital problems [15–17] and materials with three and more orbitals, starting with [30], albeit with limited frequency resolution.

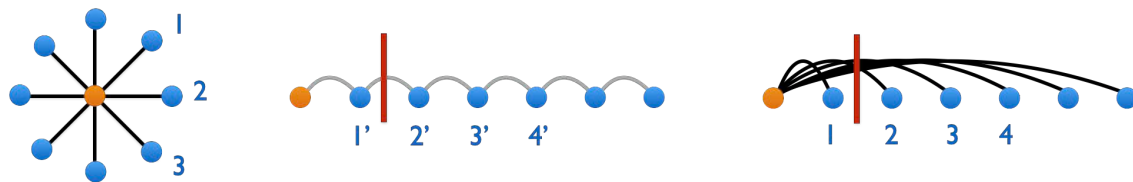


Fig. 4: Mapping the bath: On the left is the natural star-like structure of the problem: an impurity (orange) is linked by hopping to bath orbitals (blue). This can be mapped on to a chain with nearest-neighbor hopping by tridiagonalization (center); we call this the “chain geometry”. It can also be simply translated into long-ranged hoppings by flattening the star (right); we call this the “star geometry”.

4 Bath representations: chains, stars, forks and trees

The label “tensor network” reflects that we consider states on graphs to whose vertices we attach local Hilbert spaces; they can be site or orbital, momentum or energy local; the first is the most frequent case. This leaves a large amount of freedom: what graphs do we consider? How do we arrange the orbitals on the graph? In some cases, like when we are interested in a one-dimensional nearest-neighbor hopping Hubbard model, the choice is pretty obvious: we choose a one-dimensional chain, and put adjacent orbitals on adjacent vertices; then we have on-site interactions and hopping to the neighboring vertices. In many other cases, like ours, the answer is not so clear; rigorous results hardly exist.

Tensor network algorithms like DMRG exploit gauge degrees of freedom for the local tensors which allow us to give them very convenient (generalized) orthonormality properties [3] which both accelerate and numerically stabilize algorithms: many calculations can be avoided and replaced by exact analytical results. These orthonormality properties rely on the absence of loops in the graphs. Hence, quite generally, we prefer to work with chain-like tensor networks, which we already discussed (Figure 2), and tree tensor networks to be discussed below (Figure 6 shows two cases important for us). This lecture will entirely focus on loop-free networks.

Single-orbital setup. For a single-orbital problem (one relevant electronic band), the choice is between a chain or a tree-like structure. Almost all experience so far is on chains in this case, and we will only consider those. There is then the question how we arrange the bath orbitals (sites). The simple impurity problem is given by one impurity orbital with interacting electrons; these electrons can hop back and forth between the impurity orbital and one of many bath orbitals; there is no hopping between the bath orbitals themselves and the electrons on the bath orbitals are non-interacting (these are highly delocalized bath states), so the physics suggests a star-like structure as shown on the left of Figure 4.

How do we map this to a linear tensor network? We leave the impurity orbital untouched and put it on one end. Standard wisdom is that hopping entangles parts of a system, so it should be convenient to map the bath to a linear object with short-ranged hopping. In fact, any non-interacting problem can be brought into tridiagonal form, where the diagonal represents on-site

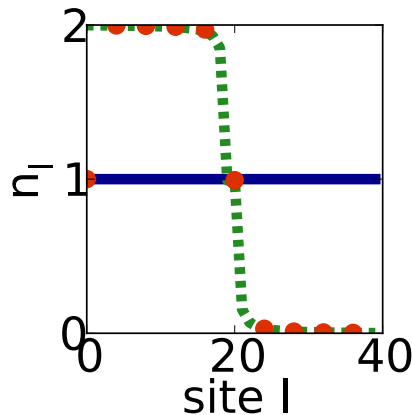


Fig. 5: Occupation numbers in a $T = 0$ bath with an impurity.

energies and the two off-diagonals a nearest-neighbor hopping. This tridiagonalization step is routinely employed in NRG [31], and the resulting Hamiltonian formally resembles a Hubbard chain with on-site energies and nearest-neighbor hopping (shown in the center of Figure 4). On the other hand, we could simply place the bath orbitals one after another on the chain, leaving the star-structure unaffected (shown on the right of Figure 4). The price to pay are long-ranged hopping terms, which are harder to implement and should lead to faster entanglement. Or so it seems. In fact, it turns out that the entanglement growth is much faster in the chain geometry, hence the required computational resources are much larger.

Why is this? Figure 5 has the answer. Unlike in a typical Hubbard model problem, on which most intuition has been developed, and in fact unlike the chain geometry setup, where there is the same (or no) on-site energy on each site after tridiagonalization, in our case, the on-site energies are spread widely. In a real-frequency TN-DMFT, one encounters at $T = 0$ occupation numbers as shown in Figure 5 if we order the sites l from small to big on-site energy ε_l : the low-lying orbitals contain 2 electrons, the high-lying orbitals none, and the edge would be sharp in the absence of the impurity. Coupling to the impurity broadens the step. This blocks the dynamics on almost all sites except close to the step: there are either no electrons that can hop or electrons cannot hop into the full orbital. This dynamical blocking without counterpart in the “normal” Hubbard model strongly suppresses entanglement growth.

Multi-orbital setups. In many cases of interest, the band structure around the Fermi level is dominated by multiple bands. For instance, in transition metals, there are 5 $3d$ -bands; effects like crystal symmetry breaking often separate groups of $3d$ -bands by energies such that only one group needs to be considered, often leading to 3-band or 2-band situations. In such cases, each of the orbitals of the “impurity” (n orbitals of interacting electrons for a n -band setup) connects to one (sometimes more, e.g., in the presence of spin-orbit coupling) of the n bands (represented by one bath each). Of course, one may represent n baths by n linear tensor networks which are joined into one huge linear structure. This means, however, that cumbersome, extremely long-ranged hoppings appear, in particular in real-time TN-DMFT with its large baths.

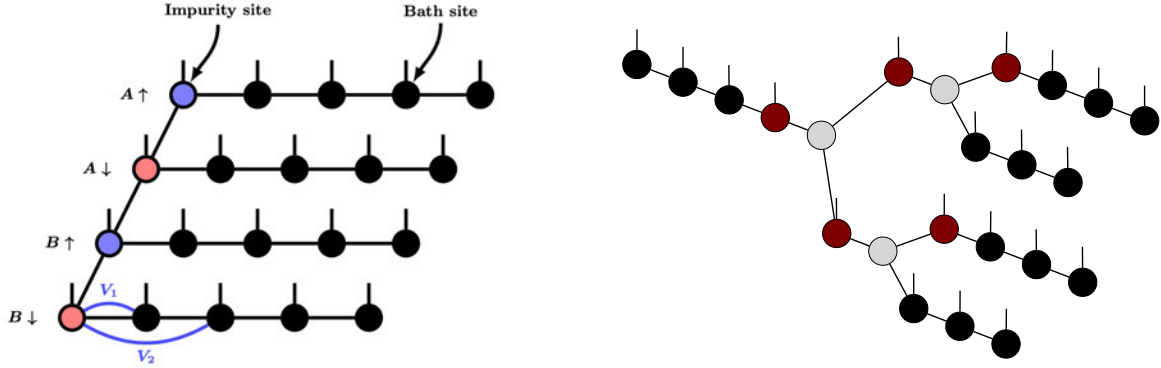


Fig. 6: Fork tensor network (left; from [30]) and T3NS (three-legged tree tensor network; right). Note that the fork shows a two-orbital problem where the spin-up and spin-down electrons have been put into separate baths for each band, which is often numerically advantageous. In the T3NS we distinguish two types of tensors, “physical” (red or black), with two auxiliary and one physical leg, as in an MPS, and “geometrical” (grey), which have three auxiliary and no physical leg, carrying the structure of the tree. The red tensors show where the impurities could sit in a five-band problem.

A natural answer is provided by fork tensor networks [30]: One arranges the 3 or 5 interacting orbitals “vertically”, and the bath orbitals which represent the 3 or 5 bands fan out “horizontally”, which resembles a fork (left part of Figure 6). The big advantage of this setup is that it contains no loops and places corresponding bath and impurity sites in proximity; it is a special case of a tree tensor network. This setup has allowed important progress in DMFT calculations for transition metal oxides. There is, however, one big disadvantage: while all bath sites and the first and last impurity orbitals are represented by three-legged tensors, as in an MPS, with computational cost scaling as D^3 , the central impurity orbital(s) yield four-legged tensors (one physical and three auxiliary legs), such that computation cost scales as D^4 . Given that D can be pretty large, $O(100)$ or even $O(1000)$, the calculation cost for this single tensor will become dominant at some point.

A way out, pioneered by [32] (and first used in [33], slightly modified, in the present context), are T3NS (three-legged tree tensor network states); right part of Figure 6. They sit on a tree-tensor network where *every* tensor is three-legged. This is achieved by introducing two different types of vertices on a graph, physical vertices which connect to two other vertices and have one physical leg, and “geometrical” vertices which carry the branching structure: they connect to three other vertices and have no physical leg. In the worst case of extreme branching, the number of vertices is doubled, but this factor 2 in computational cost vastly outweighs the disadvantage of four-legged tensors, which actually increases with the number of bands. In a T3NS one has the additional advantage that one can represent the bath sites efficiently by trees instead of chains; for very large baths in real-time TN-DMFT this may hold considerable advantage.

5 Time evolution: Finite temperatures, linear prediction, time steppers

Finite temperatures. Most tensor network calculations are concerned with ground states or the time-evolution of pure states prepared in, say, cold atom experiments. On the other hand, many materials present interesting cross-overs in their properties, as there are competing energy scales; the smaller scales, which often correspond to interesting collective behavior, become more prominent as temperature is lowered and thermal effects no longer “smear out” these scales. It is therefore relevant to be able to calculate spectral functions at any non-zero temperature. There are several techniques for that. The historically oldest technique is called purification [29] and transforms any mixed state into a pure state, to which conventional tensor network techniques can be applied. Consider some density operator $\hat{\rho}$ on a system S . Its eigendecomposition reads

$$\hat{\rho}_S = \sum_{\alpha} p_{\alpha} |\alpha\rangle_S \langle\alpha|, \quad (22)$$

where the (real) p_{α} sum to 1, and where the states $\{|\alpha\rangle_S\}$ live in the Hilbert space $\mathcal{H}_S$ of the physical system S . We now “double” the physical system by creating an environment E which has the same Hilbert space as S , $\mathcal{H}_E = \mathcal{H}_S$, so we can identify 1-to-1 states $|\alpha\rangle_S$ and $|\alpha\rangle_E$ in S and E . In this larger Universe SE , we introduce a pure state

$$|\Psi\rangle_{SE} := \sum_{\alpha} \sqrt{p_{\alpha}} |\alpha\rangle_S |\alpha\rangle_E; \quad (23)$$

then the mixed state of S is given by the partial trace over E of the pure state projector given by $|\Psi\rangle_{SE}$:

$$\hat{\rho}_S = \text{tr}_E(|\Psi\rangle_{SESE}\langle\Psi|). \quad (24)$$

This allows us to replace *all* calculations using mixed states by familiar pure-state calculations, at the cost of working in a universe double the size of the system, for instance

$$\langle\hat{O}_S\rangle_S = \text{tr}_S \hat{O}_S \hat{\rho}_S = \text{tr}_S \hat{O}_S \text{tr}_E(|\Psi\rangle_{SESE}\langle\Psi|) = \text{tr}_{SE} \hat{O}_S |\Psi\rangle_{SESE}\langle\Psi| = {}_{SE}\langle\Psi| \hat{O}_S |\Psi\rangle_{SE}. \quad (25)$$

Similarly, and relevant in our case, the real-time evolution of a density operator,

$$\hat{\rho}_S(t) = e^{-i\hat{H}_S t} \hat{\rho}_S(0) e^{+i\hat{H}_S t}, \quad (26)$$

becomes a pure-state real-time evolution of $|\Psi\rangle_{SE}$,

$$|\Psi\rangle_{SE}(t) = e^{-i\hat{H}_S t} e^{-i\hat{H}_E t} |\Psi(0)\rangle_{SE}, \quad (27)$$

where $\hat{H}_E$ can be any hermitian operator. It turns out that the choice of $\hat{H}_E$ strongly influences the amount of entanglement, hence the dimensions of the MPS and the computational cost. Without a rigorous mathematical proof, the best choice is to set $\hat{H}_E = -\hat{H}_S$. A hand-waving argument is to consider the time-evolution of a purified projector on an eigenstate of $\hat{H}_S$: it

is exactly for this choice of $\hat{H}_E$ that the purification stays invariant, hence maintains constant entanglement as it ideally should for an energy eigenstate evolving in time.

The problem is of course that we usually don't know the eigenrepresentation of $\hat{\rho}_S$; if we did, we could solve everything directly. In the case of a thermal density operator $\hat{\rho}_S(\beta) = e^{-\beta\hat{H}_S}$ there is a simple constructive procedure: we can write

$$\hat{\rho}_S(\beta) = e^{-\beta\hat{H}_S/2} \cdot \hat{I} \cdot e^{-\beta\hat{H}_S/2} \quad (28)$$

where (up to normalization!) the identity operator $\hat{I}$ is the infinite-temperature density operator $\hat{\rho}_S(0)$. An imaginary time-evolution up to times $-i\beta/2$ can be carried out by standard time-evolution tools and is in fact numerically advantageous: all numerical errors are damped out exponentially. It therefore remains to purify the identity. But this is very simple: assuming that $\mathcal{H}_S$ is a tensor product of local state spaces $\mathcal{H}_S^1$ of dimension d (works in our use cases), the identity, normalized to 1, is purified by

$$|\Psi(0)\rangle_{SE}^1 = \frac{1}{\sqrt{d}} \sum_{\sigma} |\sigma\rangle_S |\sigma\rangle_E, \quad (29)$$

where we couple states in S and E pairwise with equal weight. This can be represented as a MPS of dimension $D = 1$ (build this as an exercise; or check for instance [3]) and can be written down by hand. In fact, smart choices of pairs enables the introduction of good quantum numbers for a drastic speed-up: for instance, for a spin-1/2,

$$|\Psi(0)\rangle_{SE}^1 = \frac{1}{\sqrt{2}} (|\uparrow\rangle_S |\downarrow\rangle_E - |\downarrow\rangle_S |\uparrow\rangle_E) \quad (30)$$

is a purification which conserves both magnetization and total spin. The purification of the normalized identity on all of S is then simply the tensor product of these states on all sites.

Linear prediction. Either due to entanglement growth (as the dominant problem) meeting limited computational resources or due to accumulating numerical errors, high-accuracy calculations are strongly limited in time. This is a crucial problem: take a correlated metal. The main energy scale is set by the bandwidth, which is often of the order of 1 eV (corresponding to a temperature of 11 600 K). At the same time, many correlation effects are on energy scales of the order of 1 meV (or about 10 K). This means that we have to resolve frequencies down to 0.001 or less of the typical frequency set by the bandwidth, or times that are about 1000 times the typical time scale (like the inverse hopping in the Hubbard model). Similarly, we have to access temperatures being a similar fraction of the dominant temperature scale.

At the moment, no tensor network time-evolution for non-trivial problems gets anywhere near that. A powerful, but also quite simple technique that allows to extend the time range by a factor of 10 or so for specific time-dependent quantities is provided by linear prediction [34]. Let us assume that we have a time series of values x_n at times $n\Delta t$, $n = 1, \dots, N$. We make the bold assumption that the values x_n are for any n a *fixed* linear combination of p previous values,

$$x_n \stackrel{!}{=} - \sum_{k=1}^p a_k x_{n-k}. \quad (31)$$

We obtain the $\{a_k\}$ by using available data x_n and choosing the $\{a_k\}$ to provide the optimal fit over some training set of $\{x_n\}$, typically excluding short-time behavior. Once we have the $\{a_k\}$, we can predict $x_{N+1} = \sum_{k=1}^p a_k x_{N+1-k}$ from data up to x_N , and extend this prediction iteratively as long as we wish. Formally, we can introduce a column vector with $\mathbf{x}_n = [x_{n-1} \dots x_{n-p}]^T$ and a matrix

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_p \\ 1 & 0 & & \\ 0 & 1 & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & 1 \end{bmatrix} \quad (32)$$

such that $\mathbf{x}_{n+1} = \mathbf{A} \cdot \mathbf{x}_n$ and $\mathbf{x}_{n+m} = \mathbf{A}^m \cdot \mathbf{x}_n$. The last equation indicates that the time series is given by a superposition of p oscillating exponential decays,

$$x_{n+m} = \sum_{\nu=1}^p c_\nu e^{-i(\omega_\nu - i\eta_\nu)m} x_n. \quad (33)$$

But this is ideal for the calculation of time-dependent Green functions: they usually have the form

$$G(k, \omega) = \frac{1}{\omega - \varepsilon_k - \Sigma(k, \omega)} \quad (34)$$

where the complex self-energy $\Sigma(k, \omega)$ determines where the poles of $G(k, \omega)$ sit. Assuming we have a single pole at $\omega_1 - i\eta_1$, the Fourier transform from frequency to time yields

$$G(k, t) = c_1 e^{-i(\omega_1 - i\eta_1)t}, \quad (35)$$

and a superposition of such terms for multiple poles. But this is exactly the form of the time series of linear prediction; it is hence perfectly suited for our purpose.

Time steppers, beware! Perhaps the biggest surprise (which should not have been one) encountered in imaginary-time TN-DMFT was the problematic use of time-steppers: as discussed, most time-evolution methods rely on partitioning the object $e^{-\hat{H}\tau}$ into small timesteps $e^{-\hat{H}\Delta\tau}$ and factorizing (“Trotterizing”) this expression. The Trotter error is given by some power in $\Delta\tau$, which we therefore control very well; it can be made very small.

The Fourier transformation of the Matsubara Green function $G(\omega) = -\langle \mathcal{T} \hat{c}(0) \hat{c}^\dagger(\tau) \rangle$, where $\mathcal{T}$ is the time-ordering operator, reads for large frequencies *exactly* $1/i|\omega|$. This reflects the jump of 1 at $t=0$ due to time-ordering and the fact that fermionic operators anticommute, hence a fundamental quantum property. Yet, using time stepper methods naively, one finds that the numerator deviates substantially from 1. This is *not* a consequence of the Trotter errors as such. Assume that the value of $G(\tau)$ is exact at each timestep $\tau_n = n\Delta\tau$. How do we interpolate between these values? This is important because the statement that “short times correspond to large frequencies” in a Fourier transformation is not precise: large frequencies correspond to short time *intervals* in sampling; in fact, this is the origin of the wrong large- ω behavior.

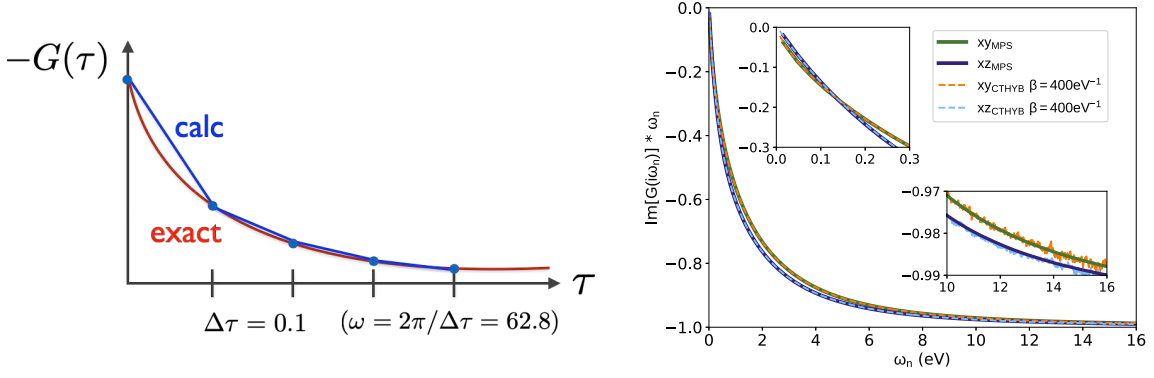


Fig. 7: Time stepper problem (left): even if we know $G(\tau)$ exactly at certain points in time, how do we interpolate? After proper time evolution (right): for some Green function calculated both by MPS and QMC, we find that $\text{Im } G(i\omega_n) \omega_n$ approaches -1 correctly for large ω_n ; the QMC results exhibit statistical noise around the MPS results (see second inset).

The changes in $G(\tau)$ are largest for small times, and therefore reducing timesteps is important for small t . Naive reductions of $\Delta\tau$ in time-stepping methods increase computation time drastically and numerical errors somewhat. An alternative approach, which we use for short times before switching to time-steppers is the Krylov method, which we formulate here for real-times (so that there is no misunderstanding that it applies to real and imaginary time-evolutions): assume we want the time evolution of some state, $|\psi(\delta t)\rangle = e^{-i\hat{H}\delta t}|\psi(0)\rangle$; inspired by a Taylor series expansion of the evolution operator (which is not advisable), we iteratively build a Krylov-basis of Hilbert space by performing an iterative Gram-Schmidt orthogonalization of the sequence $\{|\psi(0)\rangle, \hat{H}|\psi(0)\rangle, \hat{H}^2|\psi(0)\rangle, \dots\}$, leading to Krylov basis vectors $|q_0\rangle, \dots, |q_{m-1}\rangle$. This becomes computationally very expensive with increasing powers of $\hat{H}$, but m less than 10 easily suffices. We then get highly exact results by evaluating $e^{-i\hat{H}\delta t}|\psi(0)\rangle$ in the subspace spanned by the (incomplete) Krylov basis: by the Gram-Schmidt construction, the Hamiltonian is tridiagonal in the Krylov basis, $\alpha_n = \langle q_n|\hat{H}|q_n\rangle$ and $\beta_{n-1} = \langle q_n|\hat{H}|q_{n-1}\rangle$. If we introduce the matrix $Q_m = [|q_0\rangle|q_1\rangle \dots |q_{m-1}\rangle]$ with the Krylov vectors as columns, the projected Hamiltonian matrix $T_m = Q_m^\dagger H Q_m$ has dimension $m \times m$ and is tridiagonal. It can then be shown that

$$|\psi(\delta t)\rangle \approx Q_m \exp(-iT_m \delta t)|e_0\rangle, \quad (36)$$

where $|e_0\rangle = [1 \ 0 \ 0 \ \dots]^T$. Now, as m is small and T_m almost trivial to diagonalize numerically, we can calculate at essentially no cost any time-evolution not just for δt , but also smaller times, at the same or even higher accuracy; in other words, we get arbitrarily small time steps at almost no extra cost. There is then no need to guess an interpolation of $G(t)$, and results become excellent.

It turns out that the recipe of using Krylov time evolution for short times before switching to faster methods for longer times immediately resolves the large-frequency problem in $G(i\omega_n)$.

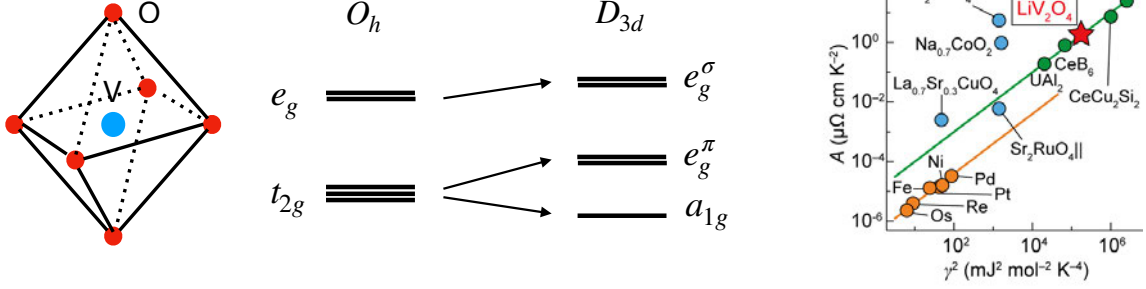


Fig. 8: Lithium vanadate LiV_2O_4 : (left) Due to its crystalline structure, the degeneracy of the 5 3d orbitals of Vanadium is lifted; the Fermi surface crosses two degenerate e_g^π and one a_{1g} orbital. (right) The surprising feature of lithium vanadate is that in a plot of the Kadowaki-Wood ratio of A , the T^2 -coefficient in electrical resistivity, and γ^2 , the (square of the) Sommerfeld coefficient in the electronic specific heat capacity, it sits right on the line defined by heavy fermion materials (from [35]).

6 Test case: LiV_2O_4

Let us illustrate the previous discussion by a material, where in my opinion, tensor networks have for the first time decisively outperformed competing impurity solvers and been crucial in elucidating (new) physics [21], as opposed to previous calculations, where quantitative improvements in performance and accuracy were achieved without making a real qualitative difference. At the same time, the cooperative concurrent use of competing impurity solvers served as an excellent mutual benchmark in regimes where several methods work, providing confidence in the validity of the approaches.

The material in question is LiV_2O_4 , which we will call lithium vanadate in the following, although there are further lithium vanadates with other stoichiometric coefficients. The relevant orbitals around the Fermi edge are the 5 d -orbitals of vanadium. Due to the local symmetry provided by the crystalline structure (Figure 8), these 5 orbitals are not degenerate, but split into 2 e_g^σ orbitals, 2 e_g^π orbitals and one a_{1g} orbital. The resulting bands are such that the 2 bands from the e_g^σ orbitals are sufficiently high above the Fermi energy that they can be neglected here. The lower three bands are all cut by the Fermi energy, and we have an effective three-band problem. Furthermore, an electron counting yields that there are 1.5 electrons per V in these 3 orbitals, which are hence on average quarter-filled. Also, the lattice structure is such that there is geometric frustration of (antiferro)magnetic interactions, such that this material does not show magnetic order down to the lowest temperatures. It was also observed that the Fermi liquid scale, where $\rho \propto T^2$, is only reached well below 10 K.

The reason why this material has created a lot of interest since around 1997 [36–38] is an anomalously large effective mass enhancement below about 25 K; experimentally, one finds $m^*/m \sim 25$, pointing to very strong collective effects. A further reason of interest was that in a Kadowaki-Wood plot of A , the T^2 -coefficient in electrical resistivity, vs. γ^2 , the (squared) Sommerfeld coefficient in the electronic specific heat capacity (Figure 8), lithium vanadate is

right on a line populated by “heavy fermion” materials, where large effective masses are found in rare earth compounds. If one claimed that lithium vanadate was a heavy fermion material, it were the first of its kind in a transition metal oxide.

The problem is that the standard explanation of heavy fermion behavior cannot work: in rare earths, there are very narrow f -bands, corresponding to almost localized electrons, and d -bands with itinerant electrons, which carry conductivity. Similar to the Kondo effect, where localized impurity electrons form a singlet with free electrons in an antiferromagnetic interaction, the localized f -band electrons form singlets with the d -band itinerant electrons, which are therefore much less mobile, yielding a very large effective mass. In lithium vanadate there are no f -band electrons. Yet, one finds that the e_g^π bands have a width of about 2 eV, and the a_{1g} band is much narrower at about 0.5 eV or so, so an early attempt at an explanation was that the e_g^π electrons play the role of the itinerant electrons, and the a_{1g} electrons that of the localized electrons [39]. Yet, this explanation does not hold up to closer inspection: the dominant magnetic interaction is a ferromagnetic Hund coupling, various experimental details do not match this explanation, and finally a band width of 0.5 eV is not excessively small.

Many other explanations were attempted, and it seems that most people were ultimately satisfied with an explanation based on extensive numerics [40], where it was found that the interactions in lithium vanadate repopulate bands such that there is a density of $n \approx 0.98$ in the a_{1g} band, in close proximity of a density 1, with an associated transition to a Mott insulator. The closeness of the transition should induce the observed strong mass renormalization; the e_g^π bands are then mere bystanders.

The problem of all explanations was, to the extent that numerics were used, that the typical temperature scale set by the bands is around 11 000 K (1 eV), and that starting from there, numerical methods could not get below around 300 K, whereas the physics of interest is on the scale of 10 K or so. This requires methods that can reach such temperatures and provide energetic resolution better than 1/1000 of the typical energy scale! This is now possible.

In our calculations, we modeled the interactions on the Vanadium atoms by a standard three-band Hubbard-Kanamori model given by

$$\begin{aligned} \hat{H}_K = & U \sum_m \hat{n}_{m\uparrow} \hat{n}_{m\downarrow} + U' \sum_{m \neq m'} \hat{n}_{m\uparrow} \hat{n}_{m'\downarrow} + (U' - J) \sum_{m < m', \sigma} \hat{n}_{m\sigma} \hat{n}_{m'\sigma} - J \sum_{m \neq m'} \hat{d}_{m\uparrow}^\dagger \hat{d}_{m\downarrow} \hat{d}_{m'\downarrow}^\dagger \hat{d}_{m'\uparrow} \\ & + J \sum_{m \neq m'} \hat{d}_{m\uparrow}^\dagger \hat{d}_{m\downarrow}^\dagger \hat{d}_{m'\downarrow} \hat{d}_{m'\uparrow}. \end{aligned}$$

Microscopic calculations yield values $U = 3.94$ eV, $U' = 2.83$ eV, and $J = 0.56$ eV, where $U' = U - 2J$, which are generally agreed on.

We applied (i) our TN-DMFT solver using purified MPS at non-zero temperatures down to 2.9 K ($\beta = 4000$), and (ii) a high-performance QMC-DMFT solver down to the lowest temperatures possible, and (iii) an NRG-DMFT solver, which is very reliable at temperatures close to zero, but is only able to treat a simplified version of the Hubbard-Kanamori model, the Dvorin-Narath model, where the last line of the Hubbard-Kanamori model containing pair hopping terms is dropped.

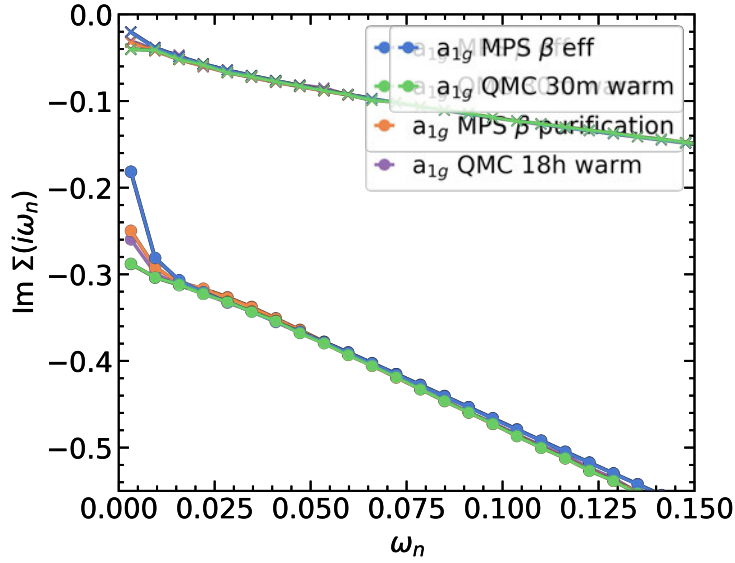


Fig. 9: *Limits of accuracy:* We show the most challenging quantity calculated, the imaginary part of the self-energy. $\beta = 1000$, the lowest temperature reached by QMC. On the top are results for the e_g^π bands without specific issues. On the bottom are results for the a_{1g} band. Blue: a $T = 0$ calculation for TN-DMFT; orange: the $\beta = 1000$ result for TN-DMFT. QMC agrees well even at the difficult smallest Matsubara frequency, provided that the warm-up time is ramped up from more typical 30 minutes to extreme 18 hours.

Let us be clear up front regarding the limitations. Figure 9 shows that QMC-DMFT recognizes the bend of the difficult $\text{Im } \Sigma_{a_{1g}}$ towards zero at the lowest frequencies, required by Fermi liquid physics, only after unusually long warm-up times; acceptance rates are extremely low. It should be noted, though, that here QMC-DMFT has been pushed to $\beta=1000$ (11.6 K), beyond the usual state of the art. TN-DMFT maps $\text{Im } \Sigma$ properly, but runs into difficulties at its lowest temperature $\beta=4000$ (2.9 K), where we observe an odd-even effect in bath length for the 1–2 lowest Matsubara frequencies. Physics dictates the correct choice, but a question mark remains.

Let us now turn to the physics. In a DFT calculation, where interactions are treated in a very approximate manner, the (low-temperature) occupations are found to be $n(a_{1g}) = 0.427$ and $n(e_g^\pi) = 0.537$. Proper consideration of the local interactions shifts them to $n(a_{1g}) \approx 0.9$ and $n(e_g^\pi) \approx 0.3$. This already excludes the explanation based on a vicinity to a Mott transition; $n(a_{1g}) \approx 0.9$ is too small for that.

A crucial observation is now, on the other hand, given by the spectral function $A(\mathbf{k}, \omega)$. Figure 10 shows in the left panel as lines the DFT bands around the Fermi energy; the colors indicate the spectral function $A(\mathbf{k}, \omega)$ at $T=11.6$ K. The crucial feature is the very narrow horizontal structure, a flat band, right at the Fermi energy. The physics is easier to see if we consider the momentum-integrated $A(\omega)$ at various temperatures (right panel), focusing on the contribution from the a_{1g} band. As we lower temperature, a very sharp peak emerges at temperatures of a few 10 K, still rounded out at around 40 K, very pronounced and peaked at around 10 K. This very narrow peak (not visible in previous calculations due to temperature) corresponds to an emerging flat band with quite localized, hence very heavy electrons, and is therefore at the bottom of the experimental phenomenon to be explained.

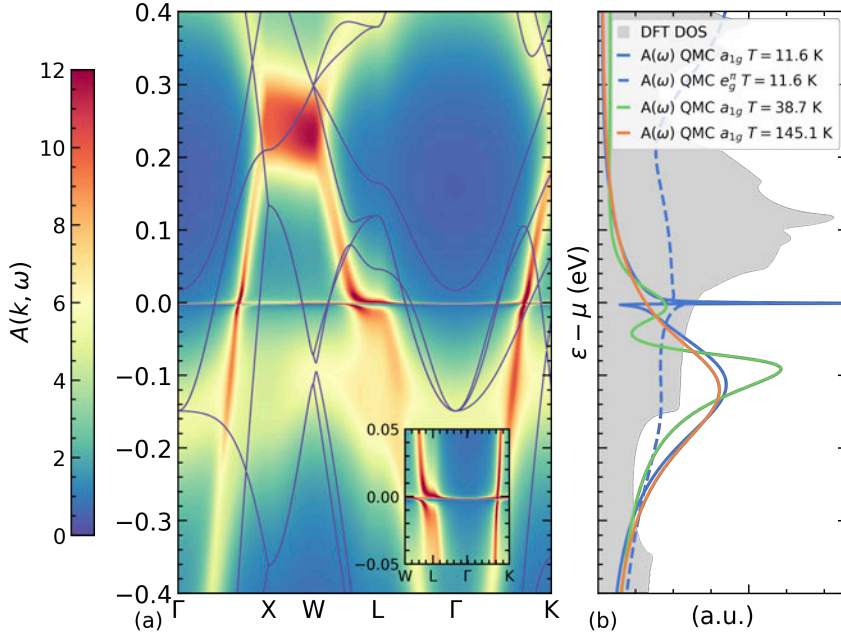


Fig. 10: Spectral functions $A(\mathbf{k}, \omega)$ at $T=11.6$ K (left) and $A(\omega)$ at various T (right). The lines on the left are the DFT band structure. On the right, one sees that the a_{1g} part of $A(\omega)$ develops a very sharp peak at very low T , previously unreachable. From [21].

It turns out that the effect is due to Hund's coupling J , involving all three bands: the two e_g^π are *not* bystanders. We can show that for fixed U , U' the splitting of the levels (hence the redistribution of the densities to the values reported above) is driven by J , but for $J \gtrsim 0.3$ eV essentially independent of J . The Hund coupling at the same time leads to the formation of high-spin states between electrons in all three bands (almost the entire weight of states with 2 or 3 electrons on the V d -orbitals is in $S=1$ and $S=3/2$ states). This is the underlying mechanism of the low-energy flat band, because this inhibits the low-energy motion in the a_{1g} band, which we can also read as a high effective mass. We might call the effect ‘‘Hund-driven orbital-selective Mott physics’’.

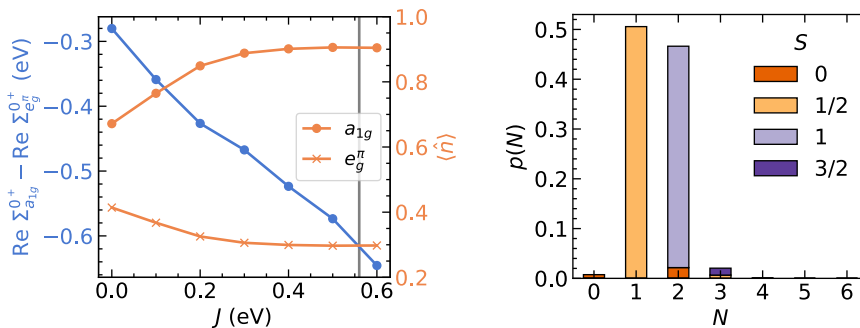


Fig. 11: Hund's physics in lithium vanadate. Left panel: difference in the real parts of the zero-frequency self-energy of the e_g^π and a_{1g} bands, showing their relative shift due to Hund's coupling J , and the resulting changes in electronic densities. Above a certain threshold, densities are robust under changes of J . Right panel: total spin vs. number of electrons on the impurity orbitals at $J=0.56$ eV. High-spin states dominate. From [21].

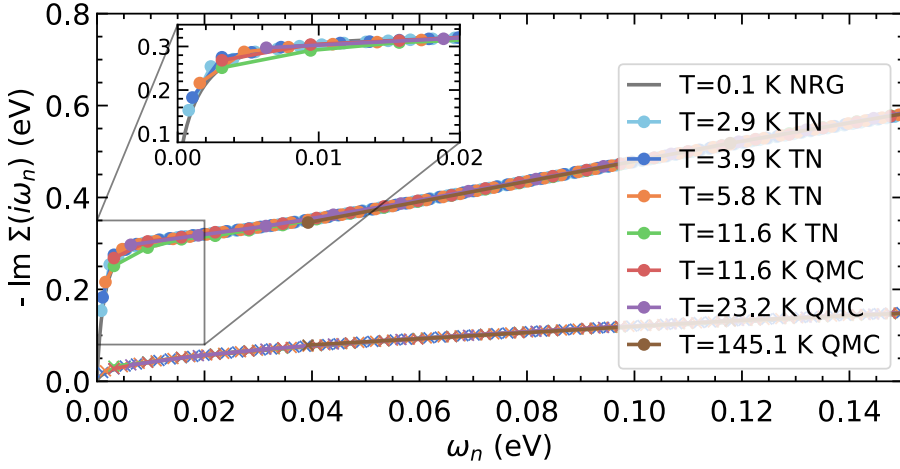


Fig. 12: *Imaginary part of the low-frequency self-energy of the e_g^π bands (bottom) and the a_{1g} band. The results of all impurity solvers at various T collapse on the same curve; in particular, the sharp bend appearing for the a_{1g} band self-energy is properly captured. From [21].*

Let us return to the question of the reliability of these simulations. It can be assessed from the imaginary part of the self-energy on the Matsubara axis $\Sigma(i\omega_n)$, which is not the physically most intuitive quantity, but numerically particularly demanding. As one can see, all three methods – NRG at the lowest T , tensor networks at very low T , and QMC at higher T seamlessly overlap (Figure 12). In particular, the strong bend emerging at the lowest ω_n for very low T is resolved consistently. A numerically easier, but physically more obvious quantity is the spectral weight right at the Fermi edge. It should become large for the a_{1g} band at the lowest temperatures, and linear in T in the Fermi liquid regime. This is observed, but more importantly, the results of NRG, TN and QMC impurity solvers connect with high accuracy (Figure 13). In fact, the narrow flat band observed in theory was at the same time independently seen in ARPES measurements of lithium vanadate at 6 K [35].

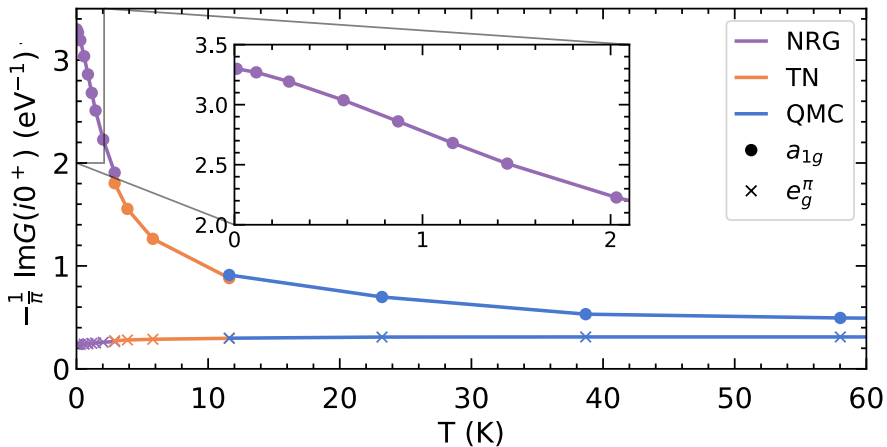


Fig. 13: *Zero-frequency limit of the spectral function of the a_{1g} band of lithium vanadate: NRG, TN and QMC solvers match seamlessly across the entire temperature range. From [21].*

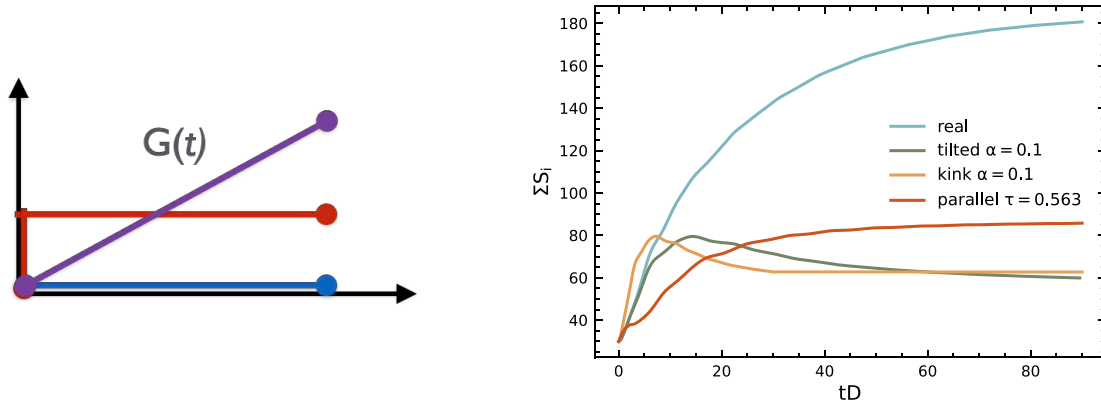


Fig. 14: Left: Contours in the complex plane for the same final real time: real time (blue), parallel (red) and tilted (violet). Right: entanglement growth with time for various contours. All contours in the complex plane have largely lower entanglement growth than the real-time contour, saving substantial resources and allowing larger times. From [22].

7 Real time and complex time evolution

The central weakness of TN-DMFT on the imaginary axis is, as in the case of QMC-DMFT, the necessity to perform some kind of analytical continuation from the imaginary to the real axis, an ill-conditioned problem. While a plethora of techniques has been developed (most prominently, maximum-entropy methods and the Padé approximation), none of them is fully convincing. On the other hand, real-frequency information is experimentally crucial, for instance for scattering, ARPES, and transport experiments. As we have already seen, we often require extremely high resolution, better than $1/1000$. For this, analytic continuation methods are particularly bad.

The problem with real-time calculations is that they are limited in time due to entanglement growth, hence yield limited frequency resolution. One way forward is the use of time-evolution contours in the complex plane, where going into the complex plane suppresses high-frequency information, which we are usually less interested in, and strongly suppresses entanglement growth. The complex plane allows myriads of contours, and it is hard to name an “optimal” one. For instance, one contour might allow for faster time-evolutions because of less entanglement, but loses some accuracy. In the first self-consistency loops of a DMFT calculation, this may be an acceptable trade-off. In any case, the chosen contour must come with a numerically robust continuation scheme to the real axis.

Figure 14 (left) shows possible contours, in particular two that have been studied in depth: (i) a contour tilted by an angle α with respect to the real axis, and (ii) a contour which follows the imaginary axis up to some imaginary time τ and then continues parallel to the real axis (“parallel contour”). In all cases, entanglement, hence required tensor sizes, are much smaller than for a real time evolution (right panel); actual calculation times, which of course depend on α and τ , cannot easily be compared between contours as it is hard to define final identical accuracy on the real axis. In general, the parallel contour is slower, but ultimately yields the more accurate results for the moment. (All is of course much faster than on the real axis.)

Let us therefore consider the parallel contour in more detail. We introduce a generalized spectral function

$$\mathcal{A}_\tau(\omega) := \frac{1}{2\pi} \lim_{\eta \rightarrow 0^+} \int dt \, e^{i(\omega + i\eta)t} \langle \{ \hat{c}_0(t + i\tau), \hat{c}_0^\dagger \} \rangle_{\psi_0} \quad (37)$$

such that we can connect to the usual spectral function and the retarded Green function as

$$\mathcal{A}_0(\omega) = \mathcal{A}(\omega) = -\frac{1}{\pi} \text{Im } G^R(\omega). \quad (38)$$

This is already the complete information required for $G^R(\omega)$: we can determine $\text{Re } G^R(\omega)$ from $\text{Im } G^R(\omega)$ by invoking the Kramers-Kronig relation, whose numerical implementation does not introduce unexpected problems.

The usefulness of this approach comes from the following analytically exact inversion formula

$$\mathcal{A}_\tau(\omega) = \mathcal{A}(\omega) e^{-\tau|\omega|} \quad (\eta \rightarrow 0). \quad (39)$$

This simple result, which only holds for the spectral function, not Green functions in general, can be derived easily by expressing $\mathcal{A}_\tau(\omega)$ in the Lehmann representation. We can therefore directly calculate $\mathcal{A}(\omega)$ from $\mathcal{A}_\tau(\omega)$!

Yet, a naive application of this formula runs into the following problem: because of $|\omega|$, $e^{-\tau|\omega|}$ has a kink right at $\omega=0$, in the low-frequency regime that interests us most. As $\mathcal{A}(\omega)$ does *not* have such a kink, there must be one in $\mathcal{A}_\tau(\omega)$ to exactly compensate for that kink in the exponential. Numerically, one finds (not surprisingly) that this kink is rounded out due to finite times reached! There is, however, a surprisingly simple and cheap workaround: we can eliminate the kink to some desired order in τ by a superposition of several $\mathcal{A}_{\tau_k}(\omega)$, where the displacement from the real axis is by various τ_k . In fact, it is not even necessary to redo the costly time-evolutions; one may add small imaginary time-evolutions at the origin which cover the difference between the various τ_k . Hence, the additional cost is negligible.

In the superposition of n contours, we choose the weights a_k (which by construction must be both positive and negative) such that

$$\sum_{k=1}^n a_k = 1 \quad \sum_{k=1}^n a_k \tau_k^l = 0 \quad (40)$$

where $l = 1, \dots, n-1$. For $n=3$, then, we eliminate the kink up to ω^2 in

$$h(\omega) = \sum_{k=1}^n a_k e^{-\tau_k|\omega|}. \quad (41)$$

Summing the inversion formula over the choices of τ_k weighted by a_k , we find

$$\mathcal{A}(\omega) h(\omega) = \sum_{k=1}^n a_k \mathcal{A}_{\tau_k}(\omega) \quad (42)$$

which we can very stably solve for $\mathcal{A}(\omega)$ also for small ω . A robust choice is for instance $n=3$ with $\tau_k = 1, 1.15, 1.3$.

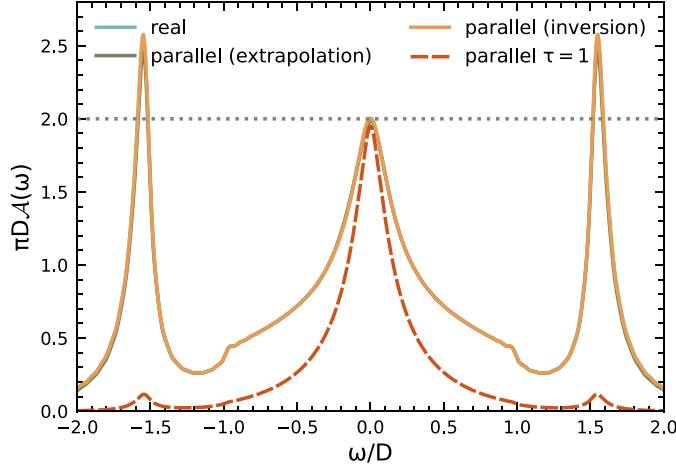


Fig. 15: Spectral function of the SIAM for $U/D=2$: here, a quite precise real-time calculation is feasible. The raw data from a parallel contour is far away, but various schemes to bring us to the real axis work very well (here, we discuss only the “inversion” scheme). The sum rule $\pi D\mathcal{A}(0) = 2$ is well obeyed. From [22].

As a simple test case, consider the single-impurity Anderson model $\hat{H} = \hat{H}_{\text{imp}} + \hat{H}_{\text{bath}}$ where

$$\hat{H}_{\text{imp}} = U\hat{n}_{0\uparrow}\hat{n}_{0\downarrow} + \varepsilon_0(\hat{n}_{0\uparrow} + \hat{n}_{0\downarrow}) \quad (43)$$

and

$$\hat{H}_{\text{bath}} = \sum_{k=1}^{N_b} \sum_{\sigma} (v_{0,k\sigma} \hat{c}_{k\sigma}^{\dagger} \hat{c}_{0\sigma} + v_{0,k\sigma}^* \hat{c}_{0\sigma}^{\dagger} \hat{c}_{k\sigma}) + \sum_{k=1}^{N_b} \sum_{\sigma} \varepsilon_{k\sigma} \hat{n}_{k\sigma} \quad (44)$$

with a semi-circular density of states

$$-\frac{1}{\pi} \text{Im} \Delta(\omega) = \frac{D}{2\pi} \sqrt{1 - \left(\frac{\omega}{D}\right)^2} \quad (45)$$

at half-filling. This is a well-understood problem, hence a good benchmark. The gold-standard results here are provided by NRG. The bath contained 299 sites, the density of states was discretized linearly on the bandwidth $2D$. Figure 15 shows the results for the spectral function for $U/D=2$ also from a real-time calculation which is possible here. One recognizes that the “raw” parallel contour data is quite far from the real-frequency data, but that the inversion scheme (and also some extrapolation scheme, which we will not pursue here) very reliably put us onto the real-frequency data.

Yet, the spectral function is comparatively benign, and in this plot possible issues with the precise match of the sum rule $\pi D\mathcal{A}(\omega=0) = 2$ are obscured. Let us therefore turn to the imaginary part of the self-energy, which is numerically challenging, in the limit $\omega/D \rightarrow 0$ for various U/D . We observe (Figure 16) that a real-time calculation yields very bad results at low frequencies, but that the complex parallel-contour calculation, combined with the inversion method, is in excellent agreement with NRG results down to very low frequencies. The remaining discrepancies come from the discretization which limits resolution and also to some extent a quite naive use of Dyson’s formula $\Sigma(\omega) = G_0^{-1}(\omega) - G^{-1}(\omega)$. Both can be improved (the extreme quality of the NRG results rests on such an improvement [41]).

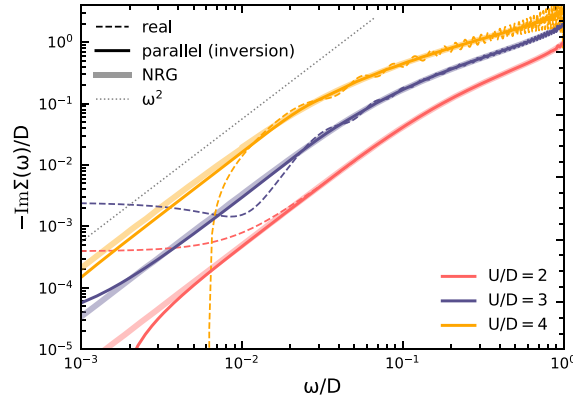


Fig. 16: Low-frequency imaginary part of the self-energy of the SIAM: NRG vs. TN on the real axis and a complex parallel contour, combined with our inversion method. Pure real-time calculations fail; complex-time calculations agree very well with NRG down to very low frequencies. Remaining discrepancies are mainly due to the limited bath size. From [22].

All methods are increasingly challenged for multi-orbital (multi-band) problems. For NRG, for instance, three bands are the limit under the proviso that the Hubbard-Kanamori Hamiltonian is simplified to the Dvorin-Narath Hamiltonian, where the pair-hopping term is eliminated. This increases the orbital symmetry from $SO(3)$ to $SU(3)$, making NRG almost exact. For the Hubbard-Kanamori Hamiltonian, results can be obtained, but are less accurate; still, they present the best currently available benchmark (in particular when combined with the methods in [41]). We again use a semi-circular DOS for the baths.

In both cases, calculations on the real-time axis are simply not sufficient. In the case of the Dvorin-Narath model, complex contours perform significantly better, but the parallel contour combined with inversion emerges as the clear winner, agreeing with NRG down to very low frequencies. As in the case of the SIAM, obvious improvements are still possible. For the Hubbard-Kanamori model, a similar picture emerges. NRG and the results from the parallel contour plus inversion do not agree as well, but this is probably rather due to the limited accuracy of NRG for this model. It can be shown, however, that improvements would shift the NRG curve downwards, closer to the TN result.

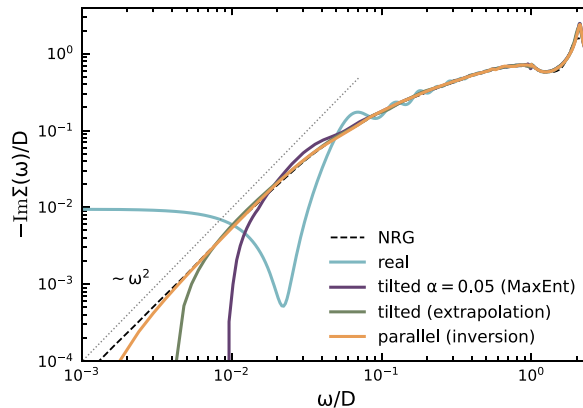


Fig. 17: Low-frequency imaginary part of the self-energy of the Dvorin-Narath model: NRG vs. TN on the real axis and various complex contours, using different continuation methods. From [22].

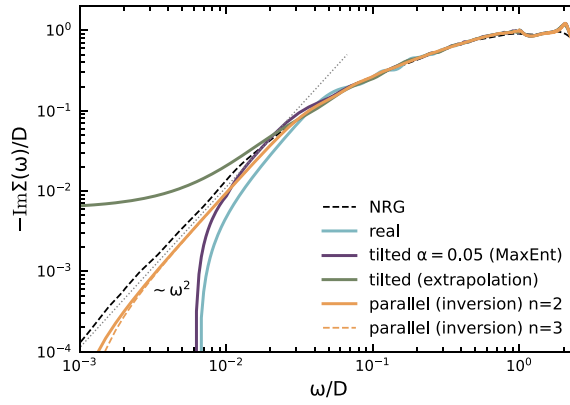


Fig. 18: Low-frequency imaginary part of the self-energy of the HK model: NRG vs TN on the real axis and various complex contours, using different continuation methods. Note the robustness of the inversion method: the superposition of 2 and 3 contours respectively hardly affects the outcome. From [22].

8 Outlook

Quo vadis, TN-DMFT? What is the future potential? On the one hand, we have by now a very reliable TN-DMFT working on the imaginary axis. It has shown its power in solving a long-debated problem, the “heavy fermion” behavior of lithium vanadate, by accessing very low temperatures reliably in a three-band problem. The setup of the problem and the quite short run-times indicate that 5-band and maybe even 7-band problems are within reach. Of course, all will depend on the size of the relevant temperature and energy scales to be resolved.

For TN-DMFT working on the real-time axis the jury is still out to some extent, but it is clear already that time evolution on complex contours is perfectly suited to obtain much more precise spectral functions and self-energies. What is still ongoing research is to go beyond “one-shot” calculations and to integrate the results into the self-consistency loops of DMFT. There is, however, no reason why this should present conceptual difficulties beyond the subtleties of a proper implementation. To be followed!

References

- [1] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
- [2] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)
- [3] U. Schollwöck, Ann. Phys. **326**, 96 (2011)
- [4] W. Metzner, D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [5] A. Georges, G. Kotliar, Phys. Rev. **B 45**, 6479 (1992)
- [6] A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [7] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)
- [8] E. Pavarini, E. Koch, A. Lichtenstein, and D. Vollhardt (Eds.):
Dynamical Mean-Field Theory of Correlated Electrons,
Reihe Modeling and Simulation, Vol. 12 (Forschungszentrum Jülich, 2022)
<http://www.cond-mat.de/events/correl22>
- [9] E. Pavarini, E. Koch, A. Lichtenstein, and D. Vollhardt (Eds.):
Understanding Correlated Materials with DMFT,
Reihe Modeling and Simulation, Vol. 15 (Forschungszentrum Jülich, 2025)
<http://www.cond-mat.de/events/correl25>
- [10] P. Werner, A. Comanac, L. De'Medici, M. Troyer, A.J. Millis,
Phys. Rev. Lett. **97**, 076405 (2006)
- [11] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, P. Werner,
Rev. Mod. Phys. **83**, 349 (2011)
- [12] M. Caffarel, W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
- [13] R. Bulla, T.A. Costi, D. Vollhardt, Phys. Rev. **B 64**, 045103 (2001)
- [14] C. Raas, G.S. Uhrig, Eur. Phys. J. **B 45**, 293 (2005)
- [15] M. Ganahl, P. Thunström, F. Verstraete, K. Held, H.G. Evertz,
Phys. Rev. **B 90**, 045144 (2014)
- [16] F.A. Wolf, I.P. McCulloch, O. Parcollet, U. Schollwöck, Phys. Rev. **B 90**, 115124 (2014)
- [17] M. Ganahl, M. Aichhorn, H.G. Evertz, P. Thunström, K. Held, F. Verstraete,
Phys. Rev. **B 92**, 155132 (2015)
- [18] F.A. Wolf, A. Go, I.P. McCulloch, A.J. Millis, U. Schollwöck,
Phys. Rev. **X 5**, 041032 (2015)

- [19] O. Parcollet, M. Ferrero, T. Ayrat, H. Hafermann, I. Krivenko, L. Messio, P. Seth, *Comput. Phys. Commun.* **196**, 398 (2015)
- [20] M. Aichhorn *et al.*, *Comput. Phys. Commun.* **204**, 200 (2016)
- [21] M. Grundner, F.B. Kugler, O. Parcollet, U. Schollwöck, A. Georges, A. Hampel, *Phys. Rev. B* **112**, L041106 (2025)
- [22] M. Grundner, P. Westhoff, F.B. Kugler, O. Parcollet, U. Schollwöck, *Phys. Rev. B* **109**, 155124 (2024)
- [23] N. Schuch, M.M. Wolf, F. Verstraete, J.I. Cirac, *Phys. Rev. Lett.* **100**, 030504 (2008)
- [24] J. Eisert, M. Cramer, M.B. Plenio, *Rev. Mod. Phys.* **82**, 277 (2010)
- [25] S. Paeckel, T. Köhler, A. Swoboda, S.R. Manmana, U. Schollwöck, C. Hubig, *Ann. Phys.* **411**, 167998 (2019)
- [26] G. Vidal, *Phys. Rev. Lett.* **93**, 040502 (2004)
- [27] A.J. Daley, C. Kollath, U. Schollwöck, G. Vidal, *J. Stat. Mech.: Theory Exp.* P04005 (2004)
- [28] S.R. White, A. Feiguin, *Phys. Rev. Lett.* **93**, 076401 (2004)
- [29] F. Verstraete, J.J. Garica-Ripoll, J.I. Cirac, *Phys. Rev. Lett.* **93**, 207204 (2004)
- [30] D. Bauernfeind, M. Zingl, R. Triebl, M. Aichhorn, H.G. Evertz, *Phys. Rev. X* **7**, 031013 (2017)
- [31] K.G. Wilson, *Rev. Mod. Phys.* **47**, 773 (1975)
- [32] K. Gunst, F. Verstraete, S. Wouters, O. Legeza, D. Van Neck, *J. Chem. Theory Comput.* **14**, 2026 (2018)
- [33] X. Cao, Y. Lu, P. Hansmann, M.W. Haverkort, *Phys. Rev. B* **104**, 115119 (2021)
- [34] T. Barthel, U. Schollwöck, S.R. White, *Phys. Rev. B* **79**, 245101 (2009)
- [35] D. Oh *et al.*, *Proc. Nat. Acad. Sci.* **122**, e2518213122 (2025)
- [36] S. Kondo *et al.*, *Phys. Rev. Lett.* **78**, 3729 (1997)
- [37] H. Takagi, C. Urano, S. Kondo, M. Nohara, Y. Ueda, T. Shiraki, T. Okubo, *Mater. Sci. Eng.: B* **63**, 147 (1999)
- [38] C. Urano, M. Nohara, S. Kondo, F. Sakai, H. Takagi, T. Shiraki, T. Okubo, *Phys. Rev. Lett.* **85**, 1052 (2000)

-
- [39] V.I. Anisimov, M.A. Korotin, M. Zöhl, T. Pruschke, K. Le Hur, T.M. Rice, Phys. Rev. Lett. **83**, 364 (1999)
- [40] R. Arita, K. Held, A.V. Lukoyanov, V.I. Anisimov, Phys. Rev. Lett. **98**, 166402 (2007)
- [41] F.B. Kugler, Phys. Rev. B **105**, 245132 (2022)

6 Diagrammatic Monte Carlo for the Hubbard Model

Riccardo Rossi
Institute of Physics
École Polytechnique Fédérale de Lausanne (EPFL)
Route Cantonale, 1015 Lausanne, Switzerland

Contents

1	Introduction	2
2	Notation and theoretical setup	4
3	Choosing the expansion: diagrammatic renormalization and beyond	6
3.1	Bare-expansion functionals	7
3.2	Shifted-action expansions	8
3.3	Why full one-particle renormalization can converge to the wrong answer	11
4	Computing the expansion: summing all diagrams to high orders	13
4.1	Sampling Feynman diagrams	14
4.2	Why summing all symmetrized Feynman diagrams is important	14
4.3	How to sum all diagrams	17
5	Summing the expansion: convergence, singularities and resummation	30
5.1	Finite radius of convergence and polynomial complexity	30
5.2	Qualitative structure of the singularities and analytic continuation	30
6	Conclusions	32

1 Introduction

The main obstacle to a reliable determination of the properties of an interacting quantum many-body system is the exponential scaling of the Hilbert space with system size. The thermodynamic properties of a macroscopic system do not change when considering a system with a few additional particles, yet the computational cost of the direct exact solution increases many times. This means that the exponential cost of the exact solution is not justified by an intrinsic exponential complexity of the thermodynamic problem, and that most of the exponentially-large wave function does not lead to any observable physical effect.

Diagrammatic Monte Carlo avoids this exponential scaling issue by directly evaluating physical observables, which are typically few-body correlation functions. This is done by writing an expansion in terms of “connected” correlations, which have a well-defined thermodynamic limit. Loosely speaking, in many situations a moderate number of connected correlations is able to combinatorially generate the exponential amount of information contained in the wave function, at least for few-body observables.

We emphasize here the distinction between standard perturbative approaches and diagrammatic Monte Carlo. In conventional perturbation theory, one typically selects a limited subset of diagrams, which may then be resummed to infinite order. This approach can be effective for weakly correlated systems, or in special cases where a particular class of diagrams dominates. In diagrammatic Monte Carlo, by contrast, all diagrammatic contributions to a given expansion are sampled up to a chosen order, without distinguishing between different classes of diagrams, except for the hierarchy imposed by the choice of expansion itself. The convergence, or lack thereof, is then assessed by comparing results obtained at different truncation orders, providing a systematic way to control the reliability of the final result.

We also want to stress the difference with respect to other quantum Monte Carlo techniques, which can also be used to compute the expectation value of observables as the ratio of two partition-function-like objects

$$\langle O \rangle = \frac{A}{Z}, \quad (1)$$

where A and Z are the exponential of extensive quantities, $A, Z \sim e^{\# \text{Volume} / \text{Temperature}}$. Typical quantum Monte Carlo approaches sample both A and Z , which results generically, for fermionic systems, in an exponential scaling with system size and temperature, a fact known as the sign problem, which has its origin in the fact that A and Z are exponentially-large quantities. In diagrammatic Monte Carlo, instead, one writes expansions of the form

$$\langle O \rangle = O_0 + O_1 + O_2 + \dots, \quad (2)$$

where each of the O_j coefficients is an intensive, or at most extensive, quantity. This means that the fermionic sign problem cannot apply by definition, and we can take the thermodynamic limit directly. By contrast, cancellations in the calculation of O_j are instead welcomed, as this just means that the contribution of the O_j term is smaller. In this document we review that, for fermionic systems, there are indeed such strong cancellations between different Feynman

diagrams that the resulting series has a non-zero radius of convergence, in contrast to what happens for bosonic theories, which have generically zero radius of convergence. This fact is important from a theoretical point of view as it allows one to show that fermionic systems on the lattice can be simulated on a classical computer with polynomial complexity, at least in some regions of the phase diagram.

Document organization. The purpose of this document is to present an introduction to the techniques that are used for diagrammatic Monte Carlo calculations in the Hubbard model. The method consists of three main parts or tasks involving our central object, the expansion:

1. **Choosing the expansion: diagrammatic renormalization and beyond (Sec. 3).** Choosing the right expansion for the problem studied is crucial. For the expectation value of a physical observable, O_{phys} , we write

$$O_{\text{phys}} = O_{\text{expan}}(z=1), \quad (3)$$

where z is some effective coupling constant, and where we have introduced the expansion function $O_{\text{expan}}(z)$, which has a $z=0$ Taylor expansion of the form

$$O_{\text{expan}}(z) = \sum_{n=0}^N z^n O_{\text{expan},n} + \mathcal{O}(z^{N+1}). \quad (4)$$

The expansion function $O_{\text{expan}}(z)$ is completely arbitrary and in general unphysical apart from $z=1$.

We are going to review that fully-self-consistent one-particle renormalization expansions can converge to a wrong physical answer. It turns out that the naive bare expansion and non-diagrammatic schemes based on the general formalism of shifted-actions are instead particularly effective and safe choices.

2. **Computing the expansion: summing all diagrams to high orders (Sec. 4).** Next, we need to discuss the evaluation of the coefficients of the expansion, $O_{\text{expan},n}$, which can be written as a sum of Feynman diagrams.

We mentioned the fact that there are strong cancellations between fermionic diagrams, which means that diagram enumeration is highly inefficient. We are going to describe techniques that make it possible to sum exactly over *all* Feynman-diagram topologies, for various expansions and physical quantities. As we show that computing the sum of all Feynman diagrams at fixed expansion order is NP-hard, the best we can reasonably achieve in general are exponential-in-order algorithms, which are still a massive improvement from the factorially-scaling enumeration technique.

Once all the diagrams of the expansion are summed over, it remains to perform an integration over the internal variables, which can be handled with a specifically-designed Markov-chain Monte Carlo procedure over sets of variables.

3. **Summing the expansion: convergence, singularities and resummation (Sec. 5).** The cancellation between the factorial number of diagrams for fermionic systems on a lattice is almost perfect, resulting in a finite radius of convergence at non-zero temperature.

When the radius of convergence is larger than one, the naive truncated sum of the z -expansion evaluated at $z=1$ gives an exponentially-convergent result. As the computational cost scales exponentially with order, this results in a polynomial-time algorithm.

When the radius of convergence is smaller than one, we know that there is a singularity somewhere at $|z|<1$. This singularity is determined by the physical properties of the system and it can be approximately identified with a calculation done at large-enough orders. If there is a singularity in the segment $z\in[0, 1]$, this corresponds to a physical phase transition, which means that the expansion is unable to enter that particular phase and another expansion should be used. If there is no indication of a singularity for $z\in[0, 1]$, analytical continuation to the physical point $z=1$ is, in principle, possible, and we discuss techniques to achieve this, which are sometimes referred to as “resummation”. The consistency of the procedure is then checked also by comparing the results at different expansion orders and for various resummation techniques.

2 Notation and theoretical setup

The Hubbard model. In this document we consider the Hubbard model. However, the formalism and the discussions presented here generalize in a straightforward way to any lattice system. For continuous-space or long-range systems, after a more careful analysis of the applicable expansions, most considerations presented in this document still apply.

We work at thermal equilibrium at inverse temperature β . The Hubbard model [1–3] is defined by the grand-canonical Hamiltonian

$$\hat{H}_{\text{phys}} = -\mu \sum_{\sigma, \mathbf{r}} \hat{n}_{\sigma}(\mathbf{r}) - \sum_{\sigma, \mathbf{r}, \mathbf{r}'} t_{\mathbf{r}, \mathbf{r}'} \hat{c}_{\sigma}^{\dagger}(\mathbf{r}) \hat{c}_{\sigma}(\mathbf{r}') + U \sum_{\mathbf{r}} \hat{n}_{\uparrow}(\mathbf{r}) \hat{n}_{\downarrow}(\mathbf{r}), \quad (5)$$

where $\sigma \in \{\downarrow, \uparrow\}$ denotes a spin index, $\mathbf{r}$ a lattice coordinate, $\hat{c}_{\sigma}(\mathbf{r})$ ($\hat{c}_{\sigma}^{\dagger}(\mathbf{r})$) the fermionic annihilation (creation) operators, $\hat{n}_{\sigma}(\mathbf{r}) = \hat{c}_{\sigma}^{\dagger}(\mathbf{r}) \hat{c}_{\sigma}(\mathbf{r})$, μ the chemical potential, t the hopping matrix, U the onsite Coulomb repulsion, and “phys” means that we are interested in a fixed, given, physical value of the Hubbard-model parameters μ , t and U .

The partition function and the grand potential. From the grand-canonical Hamiltonian $\hat{H}$, we introduce the grand-canonical partition function

$$Z = \text{Tr} e^{-\beta \hat{H}_{\text{phys}}}. \quad (6)$$

The partition function is the *exponential* of an extensive quantity.

In order to obtain an extensive quantity, we take the logarithm of the partition function, which gives the grand potential Ω

$$\Omega_{\text{phys}} = -\frac{1}{\beta} \log Z, \quad (7)$$

where “phys” hereafter is used to mean the physical value of the quantity in the Hubbard model at fixed chemical potential μ and interaction strength U . We stress that this notation is useful to avoid confusion when introducing expansion functions

$$\Omega_{\text{phys}} = \Omega_{\text{expn}}(z=1), \quad (8)$$

where z is an effective coupling constant and $\Omega_{\text{expn}}(z)$ is an expansion function, several examples of which will be given later in the document.

Imaginary-time formalism and notation for spacetime and collective variables. For the theoretical description of the thermal properties of a quantum system, it is convenient to use the imaginary-time formalism. For simplicity, we set the value of the reduced Planck constant $\hbar$ to one. We denote by capital roman letters like $X = (\mathbf{r}, \tau)$ spacetime variables consisting of a lattice position $\mathbf{r}$ and an imaginary time $\tau \in [0, \beta]$. We use calligraphic capital roman letters like $\mathcal{X} = (\mathbf{r}, \tau, \sigma)$ for spin-spacetime collective coordinates, where $\sigma \in \{\uparrow, \downarrow\}$. We also use shorthand integration symbols

$$\int dX f(X) = \int_0^\beta d\tau \sum_{\mathbf{r}} f(\mathbf{r}, \tau), \quad \int d\mathcal{X} f(\mathcal{X}) = \int_0^\beta d\tau \sum_{\mathbf{r}, \sigma} f(\mathbf{r}, \tau, \sigma). \quad (9)$$

For an operator $\hat{O}$, we define the imaginary-time Heisenberg picture as

$$\hat{O}(\tau) = e^{\tau \hat{H}_{\text{phys}}} \hat{O} e^{-\tau \hat{H}_{\text{phys}}}. \quad (10)$$

Green function. Using the time-ordered operation $\hat{T}$, which orders Heisenberg-picture fermionic operators in terms of the imaginary-time coordinate, we define the Green function as

$$G_{\text{phys}; \mathcal{Y}, \mathcal{Y}'} = -\frac{1}{Z} \text{Tr} \left[e^{-\beta \hat{H}_{\text{phys}}} \hat{T} (\hat{c}(\mathcal{Y}) \hat{c}^\dagger(\mathcal{Y}')) \right]. \quad (11)$$

The $U=0$ value of the Green function defines the physical non-interacting Green function of the Hubbard model

$$G_{0;\text{phys}} = G_{\text{phys}}|_{U=0}. \quad (12)$$

Functional integral formalism. We review here the standard Grassmann-valued functional integration technique for fermions, which is useful to derive diagrammatic identities and for renormalization. The action of the Hubbard model can be written in the form

$$S_{\text{phys}} = -\int d\mathcal{X} d\mathcal{Y} \bar{\varphi}(\mathcal{X}) G_{0;\text{phys}}^{-1}(\mathcal{X}, \mathcal{Y}) \varphi(\mathcal{Y}) + U \int dX (\bar{\varphi}_\uparrow \bar{\varphi}_\downarrow \varphi_\downarrow \varphi_\uparrow)(X), \quad (13)$$

where $\varphi(\mathcal{X})$ and $\bar{\varphi}(\mathcal{X})$ are Grassmann fields, and where $G_{0;\text{phys}}$ is the non-interacting Green function of the Hubbard model (see Eq. (12)). It is convenient to use a shorthand notation for the kinetic part of the action

$$\int d\mathcal{X} d\mathcal{Y} \bar{\varphi}(\mathcal{X}) G_{0;\text{phys}}^{-1}(\mathcal{X}, \mathcal{Y}) \varphi(\mathcal{Y}) = \langle \varphi | G_{0;\text{phys}}^{-1} | \varphi \rangle. \quad (14)$$

The functional integral formalism is useful as it avoids the use of quantum operators and of the time-ordering operation. For instance, the grand potential corresponds to a “simple” integral over Grassmann fields

$$\Omega_{\text{phys}} = -\frac{1}{\beta} \log \int \mathcal{D}[\varphi, \bar{\varphi}] e^{-S_{\text{phys}}}. \quad (15)$$

More generally, we can write any time-ordered correlation function as an expectation value over the Grassmann-valued path-integral measure $\mathcal{D}[\varphi, \bar{\varphi}] e^{-S_{\text{phys}}}$.

3 Choosing the expansion: Diagrammatic renormalization and beyond

In this document, we use a purely-functional formalism to describe renormalization. The advantages with respect to the standard diagrammatic formalism are: (i) it provides a clearer view of the assumptions made in the renormalization procedure; (ii) it provides more efficient ways to perform renormalization; (iii) it can be used to build non-diagrammatic renormalization procedures. Some of the functional expansions we consider in this document have a natural diagrammatic interpretation, which we briefly mention.

At the center of this algebraic formulation stand bare-expansion functionals, discussed in Sec. 3.1. Bare-expansion functionals are not only used to define the bare expansion, but, most importantly, they are also used to define renormalization.

A general formalism for renormalization, which can be applied to traditional diagrammatic expansions and beyond, is based on so-called shifted actions (see Sec. 3.2 and Ref. [4, 5]). Some applications are discussed, in particular the study of broken-symmetry phases and the full renormalization of the chemical potential.

Finally, as a cautionary tale on the dangers of a cavalier attitude in renormalization, in Sec. 3.3 we discuss the issue of the misleading convergence of the popular fully-self-consistent one-particle renormalization, which has been shown to happen in the Hubbard model at strong-enough coupling. This is due to the non-invertibility of the mapping between the non-interacting Green function and the interacting one, and shows up, in the shifted-action, as an unphysical singularity, an artifact of the renormalization scheme that does not correspond to a physical phase transition. This shows that the choice of the expansion is crucial: careful attention should be devoted not only to optimizing convergence, but also to guaranteeing that the convergent result corresponds to the physical one. We also discuss a large class of renormalization procedures that do not suffer from the misleading convergence found in the full one-particle renormalization, and which can therefore be safely used without conditions.

3.1 Bare-expansion functionals

3.1.1 The bare-expansion action functional

We define the bare-expansion action functional S_{bare} as

$$S_{\text{bare}}[G_0, V] := -\langle \varphi | G_0^{-1} | \varphi \rangle + \int dX V(X) (\bar{\varphi}_\uparrow \bar{\varphi}_\downarrow \varphi_\downarrow \varphi_\uparrow)(X), \quad (16)$$

which is a generalization of S_{phys} introduced in Eq. (13) where we have added, for later use, a spacetime-dependent onsite interaction $V(X)$.

3.1.2 Retrieving the physical action from the bare-expansion action

The most important property of the bare action S_{bare} is that it can reproduce the physical Hubbard-model action S_{phys} (defined in Eq. (13)) when the non-interacting Green function G_0 is equal to the physical Hubbard-model Green function $G_{0;\text{phys}}$ (see Eq. (12) for a definition) and when the interaction $V(X)$ is uniform and equal to the Hubbard U parameter (introduced in Eq. (5))

$$S_{\text{bare}}[G_0=G_{0;\text{phys}}, V(X)=U] = S_{\text{phys}}. \quad (17)$$

3.1.3 Bare-expansion functionals for other quantities

Partition function. Using S_{bare} , we can define any other bare-expansion functionals by using the Grassmann functional integral. The bare-expansion functional for the zero-interaction-normalized partition function, $\tilde{Z}_{\text{bare}}$, is

$$\tilde{Z}_{\text{bare}}[G_0, V] := \frac{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, V]}}{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, 0]}}. \quad (18)$$

Grand potential. The bare-expansion grand-potential functional Ω_{bare} is similarly defined as

$$\Omega_{\text{bare}}[G_0, V] - \Omega_{\text{bare}}[G_0, 0] := -\frac{1}{\beta} \log \tilde{Z}_{\text{bare}}[G_0, V]. \quad (19)$$

Green function. The bare-expansion functional for the Green function, G_{bare} , is defined as

$$G_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, V] = -\frac{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, V]} \varphi(\mathcal{Y}) \bar{\varphi}(\mathcal{Y}')}{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, V]}} = \frac{A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, V]}{\tilde{Z}_{\text{bare}}[G_0, V]}, \quad (20)$$

where we have introduced the auxiliary quantity $A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, V] = -\frac{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, V]} \varphi(\mathcal{Y}) \bar{\varphi}(\mathcal{Y}')}{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{bare}}[G_0, 0]}}$.

Self-energy. When $G_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, V]$ is viewed as a function of $\mathcal{Y}$ and $\mathcal{Y}'$, it is a formally invertible operator, and we denote its inverse by $G_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}^{-1}[G_0, V]$. The bare-expansion self-energy functional Σ_{bare} is defined as

$$\Sigma_{\text{bare}}[G_0, V] = G_0^{-1} - G_{\text{bare}}^{-1}[G_0, V]. \quad (21)$$

3.2 Shifted-action expansions

3.2.1 General formalism for fermionic renormalization

The bare-expansion functionals, defined in Sec. 3.1, form the basis of the shifted-action formalism for renormalization.

Shifted action. The most general fermionic renormalization scheme can be obtained by a z -dependent *shifted action* $S_{\text{expan}}(z)$

$$S_{\text{expan}}(z) := S_{\text{bare}}[G_0(z), V(X)=zU], \quad (22)$$

where $V(X) = zU$ is uniform in spacetime and proportional to the Hubbard-model U , and where $G_0(z)$ is any function of z with the property

$$G_0(z=1) = G_{0;\text{phys}}. \quad (23)$$

Retrieving the physical action. The shifted-action definition immediately implies

$$S_{\text{expan}}(z=1) = S_{\text{phys}}. \quad (24)$$

Retrieving physical quantities. The definition of the shifted action is such that any quantity computed with the $S_{\text{expan}}(z)$ action, by construction, coincides with the physical one when the coupling constant z is equal to one, $z=1$. As an explicit example, consider the shifted-action Green function

$$G_{\text{expan}}(z) = - \frac{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{expan}}(z)} \varphi(\mathcal{Y}) \bar{\varphi}(\mathcal{Y}')}{\int D[\bar{\varphi}, \varphi] e^{-S_{\text{expan}}(z)}} = G_{\text{bare}}[G_0(z), V(X)=zU]. \quad (25)$$

Thanks to Eq. (24), we immediately obtain $G_{\text{expan}}(z=1) = G_{\text{phys}}$.

Expansion coefficients. The expansion coefficients are defined from the Taylor expansion in z

$$G_{\text{expan}}(z) = G_{\text{bare}}[G_0(z), V(X)=zU] = \sum_{n=0}^N z^n G_{\text{expan},n} + \mathcal{O}(z^{N+1}). \quad (26)$$

3.2.2 The bare expansion

The bare expansion is simply obtained by considering a shifted action where the non-interacting Green function is a constant function of the coupling constant z : $G_0(z) = G_{0;\text{phys}}$

$$S_{\text{bare-ex}}(z) := S_{\text{bare}}[G_{0;\text{phys}}, V(X)=zU]. \quad (27)$$

For instance, the bare expansion function for the grand potential is

$$\Omega_{\text{bare-ex}}(z) := \Omega_{\text{bare}}[G_{0;\text{phys}}, V(X)=zU]. \quad (28)$$

3.2.3 Linearly-shifted actions I: the α -shift expansion

In this document we discuss some of the choices of shifted actions that are relevant for the Hubbard model. The simplest type of shifted action is just a linear function of z .

From the shifted Hamiltonian to the shifted action. The minimal choice for a linear function is just a chemical potential shift, which is operationally referred to as “ α -shift”, and can be defined by the shifted Hamiltonian

$$\hat{H}_{\alpha\text{-sh}}(z) = \hat{H}_{\text{phys}} + (1-z) \alpha U \sum_{\mathbf{r}, \sigma} \hat{n}_{\sigma}(\mathbf{r}) - (1-z) U \sum_{\mathbf{r}} \hat{n}_{\uparrow}(\mathbf{r}) \hat{n}_{\downarrow}(\mathbf{r}), \quad (29)$$

where $\alpha \in \mathbb{R}$ is a free parameter.

We note that the $z=0$ shifted Hamiltonian, $\hat{H}_{\alpha\text{-sh}}(z=0)$, is quadratic. This means that the corresponding shifted-action $S_{\alpha\text{-sh}}(z)$ is also quadratic for $z=0$, and can therefore be written in the form of Eq. (22) with $G_0^{-1}(z) = G_{01}^{-1} + z \Sigma_{01}$, for some G_{01} and Σ_{01} .

Freedom in the shifted action. We also note the following important property

$$\hat{H}_{\alpha\text{-sh}}(z=1) \equiv \hat{H}_{\text{phys}}, \quad (30)$$

which means that the associated shifted action satisfies the same property: $S_{\alpha\text{-sh}}(z=1) = S_{\text{phys}}$. The α -shift expansion coefficients depend on α , but the final result of the expansion, when evaluated for $z=1$, must be α -independent.

Diagrammatic interpretation. The α -shift expansion corresponds to a linear renormalization of the chemical potential as a function of the effective interaction strength $z U$. The appearance of such a linear term can be justified by the fact that the mean-field choice $\alpha = n_0/2$, where n_0 is the non-interacting spin- σ density, cancels the first-order (aka Hartree) contribution to the particle number, and eliminates all “tadpole” diagrams from the expansion.

3.2.4 Linearly-shifted actions II: symmetry breaking terms

Breaking the symmetry in the expansion. A bare expansion cannot converge in a symmetry-broken phase as the corresponding order parameter is trivially zero at all orders. To study these phases, we must consider an expansion starting point that breaks the symmetry [6, 7]. In analogy with mean-field theory, this can be achieved in a straightforward way within the linear shifted-action formalism.

Example: an antiferromagnetic phase expansion. To be concrete, let us consider the case of the antiferromagnetic phase in the Hubbard model on a bipartite lattice made of sublattices A and B . We introduce

$$v_{\sigma}(\mathbf{r}) = \alpha U + \gamma U s_{\sigma}(\mathbf{r}), \quad (31)$$

where $\alpha, \gamma \in \mathbb{R}$ are free parameters, $\sigma \in \{\downarrow, \uparrow\}$ and $s_\uparrow(\mathbf{r})=1$ for $\mathbf{r} \in A$, $s_\uparrow(\mathbf{r})=-1$ otherwise, and $s_\downarrow(\mathbf{r})=-s_\uparrow(\mathbf{r})$. The shifted Hamiltonian is

$$\hat{H}_{\alpha\gamma\text{-sh}}(z) = \hat{H}_{\text{phys}} + (1-z) \sum_{\mathbf{r}, \sigma} v_\sigma(\mathbf{r}) \hat{n}_\sigma(\mathbf{r}) - (1-z) U \sum_{\mathbf{r}} \hat{n}_\uparrow(\mathbf{r}) \hat{n}_\downarrow(\mathbf{r}). \quad (32)$$

where $\hat{H}_{\alpha\gamma\text{-sh}}(z=0)$ is quadratic, and the choice of the local potentials $v_\sigma(\mathbf{r})$ guarantees that $\hat{H}_{\alpha\gamma\text{-sh}}(z=1) = \hat{H}_{\text{phys}}$. This means that we can retrieve the physical answer by evaluating the expansion function at $z=1$. If this procedure is performed in the infinite-system-size limit, one can effectively compute quantities in the antiferromagnetic phase of the original physical Hubbard model.

We also note that the linear shifted-action formalism for symmetry breaking can be straightforwardly extended to superfluid phases using the Nambu formalism. We can also consider adding non-linear symmetry-breaking terms, which can be particularly useful if there is no mean-field theory that breaks that symmetry.

3.2.5 Linearly-shifted actions III: bosonic shifts

A straightforward way to treat renormalization beyond the one-particle channel is to use Hubbard-Stratonovich fields to decompose the Hubbard interaction.

Example: particle-particle channel renormalization. To illustrate this procedure, we focus on the particle-particle channel. We can rewrite the physical-system action in terms of the “bosonic” field $\eta(X)$

$$S_{\text{expan}}(z) = -\langle \varphi | G_0^{-1}(z) | \varphi \rangle - \int_{X,Y} \bar{\eta}(X) \eta(Y) W_0^{-1}(X, Y)(z) + \sqrt{z} \int_X (\bar{\eta} \varphi_\uparrow \varphi_\downarrow + \text{c.c.})(X). \quad (33)$$

If $G_0(z=1) = G_{0,\text{phys}}$ and $W_0^{-1}(X, Y)(z=1) = \delta(X, Y)/U$, then $S_{\text{expan}}(z=1) = S_{\text{phys}}$.

If $W_0(z)$ is not a constant function of z , the shifted action formalism effectively achieves a renormalization in the particle-particle channel. For instance, if we consider the linear shifted-action choice

$$W_0^{-1}(z) = \Gamma_0^{-1} + z\Pi_0, \quad (34)$$

where Γ_0 is the sum of all particle-particle bubble diagrams, and Π_0 is the particle-particle bubble, then this shifted-action expansion is equivalent to a diagrammatic expansion in which all particle-particle bubble diagrams are eliminated. There is a corresponding functional definition of this expansion in terms of the elimination of the first term in the expansion of the particle-particle vertex. This particle-particle-renormalized expansion can be generalized to higher orders and to various degrees of self-consistency.

3.3 Why full one-particle renormalization can converge to the wrong answer

The shifted-action formalism is very general, and can describe general diagrammatic and non-diagrammatic expansion schemes with various degrees of self-consistency. As we briefly describe here, fully-self-consistent schemes, while powerful and physically motivated, can, in some situations, artificially create unphysical singularities in the shifted action that are not associated with physical phase transitions, which lead in some cases to a misleading convergence towards an incorrect unphysical result (see Ref. [8]). This means that they should be used with care. Their applicability to a given situation can be checked by using the following sufficient criterion: if we are strictly inside the radius of convergence for the shifted action when viewed as a function of the coupling constant z , then a misleading convergence to an unphysical result can be ruled out.

Here, we limit the discussion to two expansions: (i) the full renormalization of the chemical potential to fix the density, which does not present any such issues and can always be used without checking any applicability condition; (ii) the full one-particle renormalization, whose shifted-action has unphysical branches that make the renormalized expansion converge to a wrong result at strong coupling.

3.3.1 Fully-self-consistent expansions I: full chemical-potential renormalization

As a generalization of the α -shift expansion above, we can consider the problem of finding the chemical potential $\mu(z)$ such that the density $n(z)$ is a constant function of the coupling constant z

$$\mu(z) \quad \text{such that} \quad n(z) = n. \quad (35)$$

This expansion, while working in the grand-canonical ensemble where the number of particles is not fixed, shares features with the canonical ensemble. It is useful, in particular, to disentangle the trivial density-changing effect of the change of the coupling constant z from genuine interaction effects. We emphasize that this expansion, despite its simple functional definition, has no diagrammatic interpretation.

The parametrization $\mu(z)$ does not typically create unphysical singularities: for all $z \in [0, 1]$, $\mu(z)$ is a real chemical potential of some physical Hubbard model, and a singularity in the mapping $\mu \rightarrow n$ would mean a physical phase transition involving an incompressible phase of matter at that particular density and interaction strength.

3.3.2 Fully-self-consistent expansions II: one-particle renormalization and misleading convergence

Extending the logic of the full chemical-potential renormalization introduced above, we can consider a non-interacting z -dependent Green function $G_0(z)$ that makes the interacting Green function $G(z)$ a z -independent function

$$G_0(z) \quad \text{such that} \quad G(z) = G. \quad (36)$$

This expansion is often referred to as the one-particle bold(-line) expansion, where the “bold” terminology comes from its Feynman-diagram interpretation, where self-energy diagram insertions can be replaced by a bold line representing the sum of all such insertions.

Functional definition. We now introduce the bold expansion more carefully by considering the full V -potential dependence. We consider the self-consistency condition for the non-interacting “bold” Green function $G_{0;\text{bold}}[G, V]$, which is a functional of the Green function G and the spacetime-dependent interaction $V(X)$, defined such that the interacting Green function G is independent of the interaction strength field V

$$G^{-1} = (G_{0;\text{bold}}[G, V])^{-1} - \Sigma_{\text{bare}}[G_{0;\text{bold}}[G, V], V]. \quad (37)$$

It is also useful to express the self-energy directly in terms of the interacting Green function by introducing the self-energy bold functional $\Sigma_{\text{bold}}[G, V]$

$$\Sigma_{\text{bold}}[G, V] := \Sigma_{\text{bare}}[G_{0;\text{bold}}[G, V], V], \quad (38)$$

where $G_{0;\text{bold}}[G, V]$ is a solution of the self-consistent Eq. (37). The functional formalism we have used here is fully non-perturbative.

Perturbative bold functionals. When the self-consistent equation for the bold formalism is solved order-by-order in powers of V , the coefficients of the Σ_{bold} expansion can be interpreted in terms of so-called skeleton Feynman diagrams, i.e., diagrams that cannot be made disconnected after removing one or two G lines.

The perturbative solution for $G_{0;\text{bold}}[G, V]$ need not be the unique solution of the non-linear equation: importantly, this non-uniqueness is the basis of a crucial issue that we now discuss. The non-invertibility of the mapping creates singularities, and, most importantly, these singularities do not allow the analytic continuation of the perturbative branch to the physical point. This general mathematical fact is understood already at the level of a simple toy model introduced in Ref. [9], a zero-dimensional version of the Hubbard model, whose Green function is

$$G_{\text{bare}}(G_0, V) = \frac{G_0}{1 - V G_0^2}. \quad (39)$$

The relation $G_0 \rightarrow G$ is not one-to-one: this creates a singularity in bold functionals. For instance, the self-energy bold functional is

$$\Sigma_{\text{bold}}(G, V) = \frac{-1 \pm \sqrt{1 + 4 V G^2}}{2G}, \quad (40)$$

where the $+$ choice corresponds to the perturbative branch of the self-energy

$$\Sigma_{\text{bold,perturbative}}(G, V) = \frac{-1 + \sqrt{1 + 4 V G^2}}{2G}, \quad (41)$$

and the $-$ to a non-perturbative branch, inaccessible from analytic continuation of the perturbative branch. We stress that the non-invertibility of the mapping $G_0 \rightarrow G$ does not have any physical consequence, it does not represent a physical phase transition. However, it has the unfortunate effect of creating an artificial singularity when using the fully-self-consistent diagrammatic formalism at large-enough U .

Misleading convergence. The singularity in the bold self-energy functional can be shown to lead to a misleading convergence of the truncated bold expansion in the zero-dimensional model, where the scheme can converge, albeit slowly, to a wrong solution. More precisely, if we define the Green function sequence G_N

$$G_N^{-1} = G_0^{-1} - \Sigma_{\text{bold}, \leq N}(G_N, U), \quad (42)$$

where $\Sigma_{\text{bold}, \leq N}(G, V)$ is the N -th order truncated expansion of Σ_{bold} around $V = 0$

$$\Sigma_{\text{bold}, \leq N}(G, V) = \sum_{n=1}^N V^n \Sigma_{\text{bold}, n}(G), \quad (43)$$

we find that, due to the presence of this artificial singularity, for large-enough U , G_N converges to the wrong physical answer: $G_N \rightarrow \tilde{G} \neq G$, where G is the physical interacting Green function of the Hubbard model corresponding to the non-interacting Green function G_0 .

The misleading convergence issue can be avoided by imposing the condition that one is strictly inside the radius of convergence of the shifted action. Indeed, it can be shown that a self-consistent solution of the truncated bold equations does not lead to this misleading convergence issue if this condition is verified.

It has been shown in Ref. [8] that the two-dimensional Hubbard model, at large-enough U , presents this issue. This limits the applicability of the bold formalism in the Hubbard model to weak coupling strengths, where the reliability of the converged result can be verified with this criterion.

4 Computing the expansion: Summing all diagrams to high orders

The main goal of this section is the discussion of the efficient computation of the expansion coefficients $O_{\text{expan}, n}$ for a few schemes and quantities. There are three main strategies to evaluate these coefficients: (i) One can enumerate all Feynman diagrams and compute them one-by-one, which is not optimal at high orders and accordingly not discussed in this document; (ii) One can sample different Feynman-diagram topologies, which we rapidly review in Sec. 4.1; (iii) Finally, one can sum over all diagram topologies, symmetrized over the vertex positions, which is justified in Sec. 4.2 and detailed in Sec. 4.3.

More specifically, in Sec. 4.2 we discuss, based both on explicit calculations and general considerations, the computational advantage of summing over all symmetrized diagrams. We also show that summing all Feynman diagrams is an NP-hard problem, which means that we can only expect to solve it in exponential time. Still, exponential algorithms provide a substantial qualitative and quantitative advantage over the enumeration of a factorial number of Feynman diagrams, allowing one to derive the overall polynomial complexity of Sec. 5.1.

Finally, in Sec. 4.3 we show how we can achieve the sum over all symmetrized diagram topologies for various expansions and physical quantities. We end by showing how to perform the remaining integration over the internal vertices thanks to a Markov-chain algorithm over sets.

4.1 Sampling Feynman diagrams

The original form of diagrammatic Monte Carlo consists of sampling diagram topologies. While sampling Feynman diagrams is very flexible as it allows one to directly select the diagram topologies to include in the expansion and is straightforward to formulate in any representation for the edges (notably in the momentum representation), it does not use the strong cancellations between Feynman graphs. For this reason, it generally scales factorially with the expansion order. For a discussion of this sampling algorithm, arguably the “truly diagrammatic” Monte Carlo, see, e.g., Ref. [10–13].

4.2 Why summing all symmetrized Feynman diagrams is important

4.2.1 Spacetime representation and symmetrized sum of all Feynman diagrams

Let O_{phys} be some observable we want to estimate with our diagrammatic Monte Carlo algorithm. We have seen that, in the shifted-action formalism, the observable is obtained by evaluating an expansion function $O_{\text{expan}}(z)$ of an effective coupling constant z for $z=1$

$$O_{\text{phys}} = O_{\text{expan}}(z=1). \quad (44)$$

The expansion function $O_{\text{expan}}(z)$ is generally known from a power series

$$O_{\text{expan}}(z) = \sum_{n=0}^N z^n O_{\text{expan},n} + \mathcal{O}(z^{N+1}). \quad (45)$$

It turns out that, in most expansion schemes (we will briefly mention later how to generalize this discussion in other cases), $O_{\text{expan}}(z)$ can be written as a functional of the spacetime-dependent interaction strength, $O_{\text{expan}}[V]$, evaluated for $V(X) = zU$

$$O_{\text{expan}}[V(X)=zU] = O_{\text{expan}}(z), \quad (46)$$

which means that we can write

$$O_{\text{expan},n} = \frac{1}{n!} \int dX_1 \cdots dX_n V(X_1) \cdots V(X_n) O_{\text{expan}}(\{X_1, \dots, X_n\}) \Big|_{V(X)=U}, \quad (47)$$

where $O_{\text{expan}}(\{X_1, \dots, X_n\})$ is the functional derivative with respect to $V(X_1) \cdots V(X_n)$, and it is therefore a function of the *set* of the interaction vertices $X_1, \dots, X_n$, i.e., it is completely symmetrized with respect to $X_1, \dots, X_n$.

As an example, we consider the α -shift expansion of Sec. 3.2.3, in which we generalize the chemical-potential shift to a functional of the interaction strength, resulting in a shifted action with quadratic propagator of the form

$$\langle \varphi | G_0^{-1}(z) | \varphi \rangle = \langle \varphi | G_{0;\text{phys}}^{-1} | \varphi \rangle + \alpha U \langle \varphi | \varphi \rangle - \alpha \int d\mathcal{X} V(X) (\bar{\varphi} \varphi)(\mathcal{X}) \Big|_{V(X)=zU}, \quad (48)$$

which means that $G_0^{-1}(z)$ can be retrieved from a functional of V evaluated at $V(X) = zU$, and the same is true for the interaction term, by construction.

The main object of interest in this part of the document is $O_{\text{expn}}(\{X_1, \dots, X_n\})$, which can be interpreted as the sum of *all* Feynman diagrams in the given expansion scheme with fixed space-time positions of the interaction vertices $X_1, \dots, X_n$, *symmetrized* over the vertex positions. We justify its importance in the efficient calculation of the expansion in examples below.

4.2.2 Symmetrizing diagrams over vertex positions is important

We consider here an example that shows that the symmetrization over spacetime-vertex positions can be important. We consider the one-site single-species fermionic model

$$\hat{H} = -\mu\hat{N}, \quad (49)$$

where $\hat{N}$ is the particle number, which is either 1 or 0. The partition function is $1+e^{\beta\mu}$ and the average particle number is $(1+e^{-\beta\mu})^{-1}$. We decide to build the following chemical-potential expansion from half-filling

$$\mu(z) = z. \quad (50)$$

We consider the free energy

$$F(z) = -\frac{1}{\beta} \log Z(z) = -\frac{1}{\beta} \log(1+e^{\beta z}) = \sum_{n=0}^{\infty} z^n F_n. \quad (51)$$

One notices that $F(z) = -\frac{1}{\beta} \log 2 - z/2 - \frac{1}{\beta} \log \cosh(\beta z/2)$, which implies $F_{2n+1} = 0$ for $n \geq 1$. As the closest singularities of $F(z)$ are at $z = \pm i\pi/\beta$, we have $F_{2n} \sim -\frac{1}{\beta} \frac{(-\beta^2)^n}{n\pi^{2n}}$ at large n . We now consider the Feynman-diagram expansion for the free energy in the imaginary-time representation: at order n , there is only one Feynman diagram, with weight proportional to

$$G_0(\tau_1 - \tau_2) G_0(\tau_2 - \tau_3) \cdots G_0(\tau_n - \tau_1), \quad (52)$$

where $\tau_j \in [0, \beta]$ are the imaginary-time positions of the two-point interaction vertices, and $G_0(\tau) = -1/2$ for $\tau \in [0, \beta]$, $G_0(\tau + \beta) = -G_0(\tau)$. This shows that

$$|G_0(\tau_1 - \tau_2) G_0(\tau_2 - \tau_3) \cdots G_0(\tau_n - \tau_1)| = \frac{1}{2^n}. \quad (53)$$

This means that cancellations inside this single Feynman diagram when symmetrizing over vertex positions are perfect for odd n , and exponential for even n , as the sum over symmetrized vertex positions is asymptotically $(\frac{\pi}{2})^n$ smaller than the absolute value of the weight of the unsymmetrized Feynman diagram (see Fig. 1 for an example at third order).

4.2.3 Summing over all Feynman-diagram topologies is important

We show here that summing over all Feynman-diagram topologies can be important in a specific example. We consider the one-site Hubbard model

$$\hat{H}(z) = -\sum_{\sigma} \mu_{\sigma} \hat{n}_{\sigma} + z U \hat{n}_{\uparrow} \hat{n}_{\downarrow}. \quad (54)$$



Fig. 1: Vertex symmetrization of Feynman diagrams is important. We show the single Feynman diagram contributing to the chemical potential expansion for the grand potential of the one-site Hubbard model at third order, augmented with all possible (factorial) choices for vertex symmetrization. At half filling, the two contributions shown here cancel exactly.

The partition function is

$$Z(z) = 1 + \sum_{\sigma} e^{\beta\mu_{\sigma}} + e^{\beta(\mu_{\uparrow} + \mu_{\downarrow} - zU)}. \quad (55)$$

We consider the limit $\beta\mu_{\downarrow} \rightarrow -\infty$, in which the free energy $F(z) = -\frac{1}{\beta} \log Z(z)$ becomes

$$F(z) = -\frac{1}{\beta} \log(1 + e^{\beta\mu_{\uparrow}}) - \frac{1}{\beta} e^{\beta\mu_{\downarrow}} \frac{1 + e^{\beta\mu_{\uparrow} - \beta zU}}{1 + e^{\beta\mu_{\uparrow}}} + \mathcal{O}(e^{2\beta\mu_{\downarrow}}), \quad (56)$$

which implies

$$F_n = -\frac{(-\beta U)^n}{\beta n!} \frac{e^{\beta\mu_{\downarrow}}}{1 + e^{-\beta\mu_{\uparrow}}} + \mathcal{O}(e^{2\beta\mu_{\downarrow}}), \quad \text{for } n \geq 1. \quad (57)$$

We now consider the Feynman diagrams contributing to the grand potential for a given ordering of the imaginary-time positions of the vertices $\tau_1 < \tau_2 < \dots < \tau_n$: in the low-density limit for the spin-down particles, the $\downarrow$ lines must connect τ_j with τ_{j+1} to avoid creating subleading negative-time edges, while the up lines can connect vertices without any constraint. This means that the symmetrization over vertex positions does nothing in this limit, and it is therefore an ideal case to study the effect of diagram-topology cancellations. The contribution of a given connected Feynman diagram at fixed imaginary-time positions of the vertices is

$$\left(\prod_{j=1}^n G_{0;\downarrow}(\tau_{j+1} - \tau_j) \right) \text{sign } \eta \prod_{j=1}^n G_{0;\uparrow}(\tau_j - \tau_{\eta_j}), \quad (58)$$

where $\eta \in S_n$ is a permutation of the n vertices and $\tau_{n+1} \equiv \tau_1$. In order to further simplify calculations, we suppose $\mu_{\uparrow} = 0$, which allows one to write the contribution of a given connected Feynman diagram as

$$\frac{e^{\beta\mu_{\downarrow}}}{2^n} \text{sign } \eta \prod_{j=1}^n \text{sign}(\tau_j - \tau_{\eta_j}), \quad (59)$$

whose absolute value is $e^{\beta\mu_{\downarrow}}/2^n$ (see Fig. 2 for an example at third order). The time integration contributes $\beta^n/n!$, which needs to be multiplied by the $n!$ Feynman diagrams. Comparing with the exact expression of F_n , $F_n \sim \frac{1}{n!}$, we conclude that there are factorial cancellations between Feynman-diagram topologies. This means that sampling individual Feynman-diagram topologies scales factorially with n in this case.

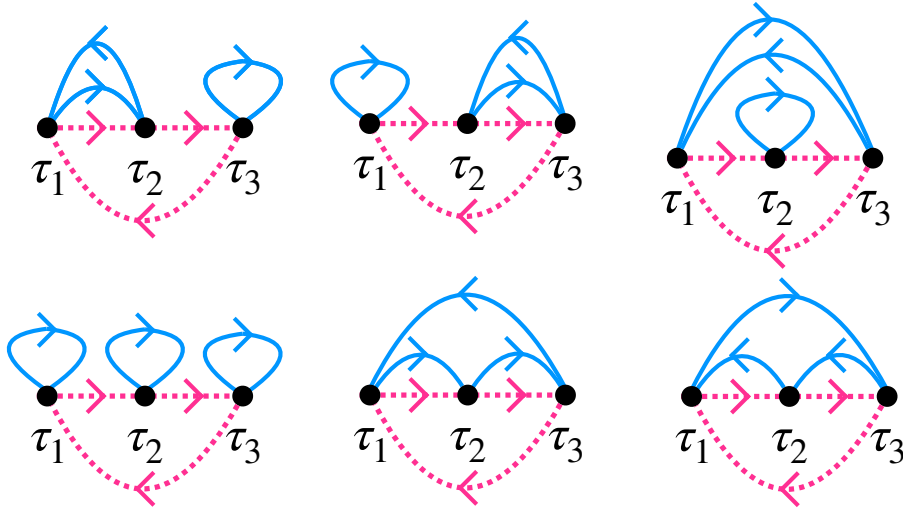


Fig. 2: Summing over all Feynman-diagram topologies is important. We show all order-three Feynman diagrams contributing to the grand potential of the one-site Hubbard model. We are in the limit where the spin-down chemical potential goes to $-\infty$. We choose $\tau_1 < \tau_2 < \tau_3$, so that vertex symmetrization gives no additional contribution thanks to the low density of spin-down particles. Spin-up propagators are denoted with blue full lines, while spin-down edges are red dashed lines. At half filling, all the factorial Feynman diagrams have the same absolute contribution, but they have different signs, which leads to an almost-perfect cancellation to give an exponential contribution in the order.

One might additionally argue that the reason for the almost-perfect cancellation really comes from different diagram topologies, and not from the imaginary-time sign of the Green function. Indeed, removing $\text{sign}(\tau_j - \tau_{\eta_j})$ and performing the sum over the Feynman diagrams η gives a zero contribution.

4.3 How to sum all diagrams

We have shown that evaluating the symmetrized sum of all Feynman diagrams is crucial to take advantage of the almost-perfect cancellation of fermionic diagrams. Here we show how to efficiently evaluate the sum of all Feynman diagrams for various quantities and expansions.

4.3.1 Wick's theorem for the sum of all connected and disconnected diagrams

In this section we show that Wick's theorem makes it possible to obtain the sum of all symmetrized Feynman diagrams (see Ref. [14, 15]). We stress that Wick's theorem does not distinguish between connected and disconnected diagrams.

We focus on the calculation of the normalized partition function $\tilde{Z}$ in the bare expansion

$$\tilde{Z}_{\text{bare}}[G_0, zV] = \frac{Z_{\text{bare}}[G_0, zV]}{Z_{\text{bare}}[G_0, 0]} = \frac{\int D[\varphi, \bar{\varphi}] e^{-S_{\text{bare}}[G_0, zV]}}{\int D[\varphi, \bar{\varphi}] e^{-S_{\text{bare}}[G_0, 0]}} \stackrel{*}{=} \sum_{n=0}^{\infty} z^n \tilde{Z}_{\text{bare}, n}[G_0, V], \quad (60)$$

where $\stackrel{*}{=}$ is used to denote a formal power series, which is just a notation that means that we do not want to discuss here the convergence properties of the series, even if, in this case, the series

is always absolutely convergent. The coefficients of the expansion have the explicit form

$$\tilde{Z}_{\text{bare},n}[G_0, V] = \frac{1}{n!} \int dX_1 \cdots dX_n V(X_1) \cdots V(X_n) \tilde{Z}_{\text{bare}}[G_0](\{X_1, \dots, X_n\}), \quad (61)$$

where we have introduced the n -th functional derivative with respect to V of the order- n bare-expansion contribution

$$\tilde{Z}_{\text{bare}}[G_0](\{X_1, \dots, X_n\}) := \langle (\bar{\varphi}_\uparrow \varphi_\uparrow \bar{\varphi}_\downarrow \varphi_\downarrow)(X_1) \cdots (\bar{\varphi}_\uparrow \varphi_\uparrow \bar{\varphi}_\downarrow \varphi_\downarrow)(X_n) \rangle_{S_{\text{bare}}[G_0, 0]}. \quad (62)$$

$\tilde{Z}_{\text{bare}}[G_0](\{X_1, \dots, X_n\})$ is the sum of all Feynman diagrams with fixed spacetime vertex positions $X_1, \dots, X_n$, symmetrized over permutations of $X_1, \dots, X_n$. For this reason, we use the set notation $\{X_1, \dots, X_n\}$ instead of a sequence notation like $(X_1, \dots, X_n)$.

The action $S_{\text{bare}}[G_0, 0]$ is quadratic in φ and $\bar{\varphi}$: this means that we can compute moments exactly with Wick's theorem: any Gaussian average of monomials of the Grassmann fields is equal to the sum over all possible products of contractions of $G_0(\mathcal{X}, \mathcal{Y}) = -\langle \varphi(\mathcal{X}) \bar{\varphi}(\mathcal{Y}) \rangle$, with a proper sign coming from fermionic exchange

$$(-1)^m \langle \varphi(\mathcal{X}_1) \bar{\varphi}(\mathcal{Y}_1) \cdots \varphi(\mathcal{X}_m) \bar{\varphi}(\mathcal{Y}_m) \rangle_{S_{\text{bare}}[G_0, 0]} = \sum_{\sigma \in S_m} \text{sign } \sigma \prod_{j=1}^m G_0(\mathcal{X}_j, \mathcal{Y}_{\sigma_j}) = \det \mathbb{G}_0, \quad (63)$$

where the $m \times m$ matrix $\mathbb{G}_0$ is defined by

$$[\mathbb{G}_0]_{jk} = G_0(\mathcal{X}_j, \mathcal{Y}_k). \quad (64)$$

This determinantal form of Wick's theorem, which is well known and used in various quantum Monte Carlo algorithms, shows that we can compute the symmetrized sum of a factorial number of Feynman diagrams with an effort that only increases polynomially with the order of the expansion. For the particular case of the Hubbard model, we can compute the $n!$ diagrams symmetrized over $n!$ permutations of the vertices, which define $\tilde{Z}_{\text{bare}}[G_0](\{X_1, \dots, X_n\})$, in $\mathcal{O}(n^3)$ time, which is a remarkable result given the large number of symmetrized Feynman diagrams. However, the diagrams we can compute with Wick's theorem are partition function diagrams, and not all of them are well-defined in the thermodynamic limit. In order to obtain diagrammatic expansions for physical observables such as the free energy density, which has a well-defined thermodynamic limit, we need to remove disconnected diagrams from the partition function, as shown in the rest of this section.

4.3.2 Computing the contribution of connected Feynman diagrams is NP-hard

We have seen in the specific case of the bare expansion of the Hubbard model that we can evaluate the sum of all diagrams, connected and disconnected, in a time increasing polynomially with the order of the expansion. One might expect that the task of removing disconnected diagrams from this sum is rather easy, and maybe hope for a polynomial-time algorithm. Instead, we show here that evaluating the contribution of order- n connected Feynman diagrams is NP-hard.

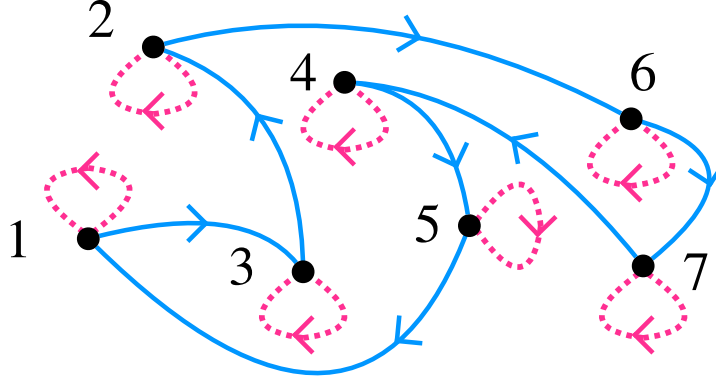


Fig. 3: Computing the contribution of connected Feynman diagrams is NP-hard. *If a general algorithm to compute the symmetrized sum of all connected Feynman diagrams in polynomial time for the Hubbard model existed, then it could be used to solve all NP problems in polynomial time. Shown here is a connected Feynman diagram of a Hubbard-like model with a diverging chemical potential for the spin-down particle. The sum of all such connected Feynman diagrams gives the Hamiltonian cycle polynomial of the graph with the $\mathbb{G}_{0;\uparrow}$ matrix.*

More precisely, we are going to show that evaluating the n -th coefficient of the bare expansion, built from connected Feynman diagrams, for a general G_0 , is NP-hard. For this proof, it is sufficient to consider a discretized spacetime with indices j , and the Grassmann action

$$S_{\text{bare}}[G_0](z) = - \sum_{\sigma \in \{\uparrow, \downarrow\}; j, k=1}^n \bar{\varphi}_{\sigma; j} [G_0^{-1}]_{\sigma; jk} \varphi_{\sigma; k} + z \sum_{j=1}^n \bar{\varphi}_{\uparrow; j} \varphi_{\uparrow; j} \bar{\varphi}_{\downarrow; j} \varphi_{\downarrow; j}. \quad (65)$$

For this proof, we consider

$$G_{0;\downarrow, jk} = \delta_{jk}, \quad (66)$$

while $G_{0;\uparrow, jk}$ is arbitrary. This choice makes the interaction effectively one-body. Accordingly, there is only one Feynman diagram contributing at order n , which is built from $G_{0;\uparrow}$ lines and trivial tadpole lines for $G_{0;\downarrow}$ (see Fig. 3). This means that the total order- n integrated contribution is, by construction, given by the symmetrization of this single Feynman diagram

$$\left(\prod_{j=1}^n G_{0;\downarrow; jj} \right) \sum_{\eta \in H_n} \prod_{j=1}^n G_{0;\uparrow; j\eta_j}, \quad (67)$$

where H_n is the set of permutations of the vertices $1, \dots, n$ having exactly one cycle. The term $\sum_{\eta \in H_n} \prod_{j=1}^n G_{0;\uparrow; j\eta_j}$ is the Hamiltonian cycle polynomial with connectivity matrix $G_{0;\uparrow; jk}$. This means that the existence of a Hamiltonian cycle of a general graph, which is an NP-complete problem (see Ref. [16–18]), could be detected by the evaluation of this polynomial for a special choice of the matrix $G_{0;\uparrow; jk}$. This means that computing the order n contribution, either by sampling Feynman diagrams or with any other algorithm to compute the sum, is a hard, almost-surely exponential, problem.

In practice, this means that we should not expect the existence of generic polynomial-time algorithms to evaluate connected symmetrized Feynman-diagram contributions. For instance,

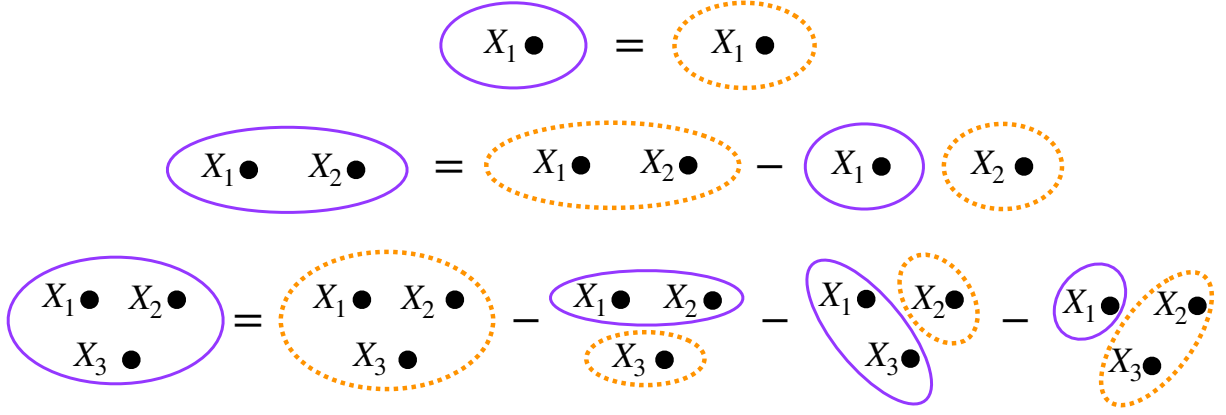


Fig. 4: Recursive formula to remove disconnected Feynman diagrams. *Dashed-line orange sets represent $\tilde{Z}(S)$, the sum of all Feynman diagrams (connected+disconnected) contributing to the partition function, while full-line violet sets correspond to $\tilde{\Omega}(S)$. Shown is the recursive formula to obtain $\tilde{\Omega}(S)$ from the knowledge of $\tilde{Z}(S')$ for $S' \subseteq S$.*

the Held-Karp algorithm can be adapted to compute the Hamiltonian cycle polynomial relevant to the one-body expansion we have discussed in the proof. It has a run-time of $\mathcal{O}(n^2 2^n)$ for an n -th vertex graph. For a k -body interaction, with $k > 1$, we certainly should not expect a meaningfully faster runtime.

4.3.3 The connected determinant approach

Diagrams contributing to physical observables with a proper thermodynamic limit must be connected. We have shown that computing connected Feynman diagrams is NP-hard, so we should not expect a faster-than-exponential algorithm to compute them. The purpose of this section is to show a general k -body algorithm to compute symmetrized connected Feynman diagrams contributing to the bare expansion. The runtime of this so-called “connected-determinant” method (see Ref. [19]) scales as $\mathcal{O}(3^n)$, which can be asymptotically improved to $\mathcal{O}(n^2 2^n)$ using fast subset convolution techniques (introduced in Ref. [20]), even if not practical for the typical n values encountered.

We write the bare-expansion grand potential

$$\Omega_{\text{bare}}[G_0, zV] = -\frac{1}{\beta} \log Z_{\text{bare}}[G_0, zV] = \Omega_{\text{bare}}[G_0, 0] - \frac{1}{\beta} \log \tilde{Z}_{\text{bare}}[G_0, zV]. \quad (68)$$

To simplify the formulas, it is useful to introduce

$$\tilde{\Omega}_{\text{bare}}[G_0, V] := \log \tilde{Z}_{\text{bare}}[G_0, V], \quad (69)$$

which we expand in formal power series in z

$$\tilde{\Omega}_{\text{bare}}[G_0, zV] \stackrel{*}{=} \sum_{n=0}^{\infty} z^n \tilde{\Omega}_{\text{bare},n}[G_0, V], \quad (70)$$

where

$$\tilde{\Omega}_{\text{bare},n}[G_0, V] = \frac{1}{n!} \int dX_1 \cdots dX_n V(X_1) \cdots V(X_n) \tilde{\Omega}_{\text{bare}}[G_0](\{X_1, \dots, X_n\}). \quad (71)$$

$\tilde{\Omega}_{\text{bare}}[G_0](\{X_1, \dots, X_n\})$ is the sum of all connected Feynman diagrams with fixed spacetime vertex positions $X_1, \dots, X_n$, symmetrized with respect to the vertex positions. We simplify here the notation by dropping “bare” and the explicit G_0 dependence:

$$\begin{aligned}\tilde{\Omega}_{\text{bare}}[G_0](\{X_1, \dots, X_n\}) &\mapsto \tilde{\Omega}(\{X_1, \dots, X_n\}), \\ \tilde{Z}_{\text{bare}}[G_0](\{X_1, \dots, X_n\}) &\mapsto \tilde{Z}(\{X_1, \dots, X_n\}).\end{aligned}\tag{72}$$

By expanding the logarithm, we can express $\tilde{\Omega}$ in terms of the easy-to-compute $\tilde{Z}(S)$. Explicitly

$$\tilde{\Omega}(\{X_1\}) = \tilde{Z}(\{X_1\}),\tag{73}$$

$$\tilde{\Omega}(\{X_1, X_2\}) = \tilde{Z}(\{X_1, X_2\}) - \tilde{Z}(\{X_1\}) \tilde{Z}(\{X_2\}),\tag{74}$$

$$\begin{aligned}\tilde{\Omega}(\{X_1, X_2, X_3\}) &= \tilde{Z}(\{X_1, X_2, X_3\}) - \tilde{Z}(\{X_1\}) \tilde{Z}(\{X_2, X_3\}) - \tilde{Z}(\{X_2\}) \tilde{Z}(\{X_1, X_3\}) \\ &\quad - \tilde{Z}(\{X_3\}) \tilde{Z}(\{X_1, X_2\}) + 2 \tilde{Z}(\{X_1\}) \tilde{Z}(\{X_2\}) \tilde{Z}(\{X_3\}),\end{aligned}\tag{75}$$

and so on. These expressions for computing $\tilde{\Omega}(S)$ given $\tilde{Z}(S')$ for all $S' \subseteq S$ are not very efficient as the number of terms is equal to the number of possible partitions of a set of n elements, the Bell number, which grows factorially with n .

Using the fact that $\tilde{\Omega}(S)$ represents the sum of all connected Feynman diagrams with spacetime vertex positions at S , we can derive the more efficient recursive formula

$$\tilde{\Omega}(S) = \tilde{Z}(S) - \sum_{S' \subsetneq S \setminus \{X_1\}} \tilde{\Omega}(\{X_1\} \cup S') \tilde{Z}(S \setminus S' \setminus \{X_1\}),\tag{76}$$

which has the interpretation of removing from $\tilde{Z}(S)$ the diagrams that can be split in a part connected to the vertex $\{X_1\}$, represented by $\tilde{\Omega}(\{X_1\} \cup S')$, and another part of the diagram that is disconnected from it, $\tilde{Z}(S \setminus S' \setminus \{X_1\})$. See Fig. 4 for a graphical interpretation of the recursive formula.

The number of arithmetic operations needed to obtain $\tilde{\Omega}(S)$ from $\tilde{Z}$ is $\mathcal{O}(3^n)$, as we now show. Let $\tilde{\Omega}_{X_1}(S) = \tilde{\Omega}(\{X_1\} \cup S)$. Assuming $\tilde{\Omega}_{X_1}(S')$ and $\tilde{Z}(S')$ are known for all $S' \subsetneq S$: one can apply the recursive formula to obtain $\tilde{\Omega}_{X_1}(S)$ with a number of multiplications equal to $2^{|S|} - 1$ and a number of additions equal to $2^{|S|}$. Ignoring the additions, the total number of multiplications to obtain $\tilde{\Omega}_{X_1}(S)$ is

$$\sum_{S' \subsetneq S} (2^{|S'|} - 1) = \sum_{k=0}^{|S|} \binom{|S|}{k} 2^k - 2^{|S|+1} + 1 = 3^{n-1} - 2^n + 1,\tag{77}$$

where we have used the fact that $|S| = n - 1$.

4.3.4 Algebraic formulation: nilpotent polynomials

In order to allow for generalization of the formalism we have developed, it is convenient to introduce so-called nilpotent polynomials.

Nilpotent polynomials are an automatic tool to compute functional derivatives of functionals. They are based on nilpotent symbols η_j with the following properties

$$\eta_j^2 = 0, \quad \eta_j \eta_k = \eta_k \eta_j, \quad (78)$$

where $j, k \in \{1, \dots, n\}$. Let $F[V]$ be a functional of the variable $V(X)$

$$F[V] \stackrel{*}{=} \sum_{k=0}^{\infty} \frac{1}{k!} \int dX_1 \cdots dX_k V(X_1) \cdots V(X_k) F(\{X_1, \dots, X_k\}). \quad (79)$$

If the functional is evaluated for $V_\eta(X) = \sum_{j=1}^n \delta(X - X_j) \eta_j$, we have

$$F[V=V_\eta] = \sum_{S \subseteq \{X_1, \dots, X_n\}} F(S) \eta(S), \quad (80)$$

where we have introduced $\eta(S) := \prod_{X_j \in S} \eta_j$. This means that, if our goal is to compute high-order functional derivatives with respect to the field $V(X)$, we can look at the coefficients of the functional when evaluated with V_η . While this does not sound very useful when only one functional $F[V]$ is involved, it becomes extremely convenient when considering combinations of functionals. For example, assume that we know the functional derivatives of two functionals $F_1[V]$ and $F_2[V]$, and suppose we want to compute the functional derivatives of their product

$$F_3[V] = F_1[V] F_2[V]. \quad (81)$$

For this, we can evaluate $F_j[V]$ for $V = V_\eta$

$$F_j[V] = \sum_{S \subseteq \{X_1, \dots, X_n\}} F_j(S) \eta(S), \quad (82)$$

which allows one to write

$$\sum_{S_3 \subseteq \{X_1, \dots, X_n\}} F_3(S_3) \eta(S_3) = \sum_{S_1, S_2 \subseteq \{X_1, \dots, X_n\}} F_1(S_1) F_2(S_2) \eta(S_1) \eta(S_2). \quad (83)$$

Using

$$\eta(S_1) \eta(S_2) = \begin{cases} \eta(S_1 \cup S_2) & \text{if } S_1 \cap S_2 = \emptyset \\ 0 & \text{otherwise} \end{cases}, \quad (84)$$

we obtain

$$F_3(S_3) = \sum_{S_1 \subseteq S_3} F_1(S_1) F_2(S_3 \setminus S_1), \quad (85)$$

which allows one to express the functional derivatives of F_3 in terms of functional derivatives of F_1 and F_2 . We can also interpret the previous formulation in terms of operations between nilpotent polynomials, i.e., polynomials in the nilpotent variables $\eta_1, \dots, \eta_n$, which we denote by $F_j[[\boldsymbol{\eta}]]$

$$F_3[[\boldsymbol{\eta}]] = F_1[[\boldsymbol{\eta}]] F_2[[\boldsymbol{\eta}]]. \quad (86)$$

We have already mentioned that the sum of all Feynman diagrams symmetrized over the fixed positions of spacetime vertices corresponds to a functional derivative of the expansion functional with respect to the interaction (see Eq. (47)). This means that we can use nilpotent polynomials to efficiently manipulate such functional derivatives. Moreover, they can be useful for generalizing the formalism to shifted-action renormalization.

Let us reproduce the connected-determinant derivation from this purely-algebraic nilpotent-polynomial formulation. We evaluate the functionals $\tilde{\Omega}_{\text{bare}}[G_0, V]$ and $\tilde{Z}_{\text{bare}}[G_0, V]$ for $V = V_\eta$

$$\sum_{S \subseteq \{X_1, \dots, X_n\}} \tilde{\Omega}(S) \eta(S) = \log \left(\sum_{S \subseteq \{X_1, \dots, X_n\}} \tilde{Z}(S) \eta(S) \right). \quad (87)$$

Equivalently, in terms of nilpotent polynomials

$$\tilde{\Omega}[[\boldsymbol{\eta}]] = \log \tilde{Z}[[\boldsymbol{\eta}]], \quad (88)$$

which means that computing all grand-potential connected diagrams is equivalent to taking the logarithm of a nilpotent polynomial.

In order to efficiently evaluate the logarithm, we take the derivative of both sides with respect to η_1 , and we set η_1 to zero afterwards (this is to avoid null terms like $\eta_1^2 = 0$ having a non-zero contribution $2\eta_1 \neq 0$ to the derivative)

$$\sum_{S \subseteq \{X_2, \dots, X_n\}} \tilde{\Omega}(\{X_1\} \cup S) \eta(S) = \frac{\sum_{S \subseteq \{X_2, \dots, X_n\}} \tilde{Z}(\{X_1\} \cup S) \eta(S)}{\sum_{S \subseteq \{X_2, \dots, X_n\}} \tilde{Z}(S) \eta(S)}, \quad (89)$$

which we rewrite as

$$\sum_{S', S'' \subseteq \{X_2, \dots, X_n\}} \tilde{\Omega}(\{X_1\} \cup S') \tilde{Z}(S'') \eta(S') \eta(S'') = \sum_{S \subseteq \{X_2, \dots, X_n\}} \tilde{Z}(\{X_1\} \cup S) \eta(S). \quad (90)$$

Using Eq. (84), we immediately recognize the previous equation as equivalent to Eq. (76).

4.3.5 Bare-expansion functional for the Green function

The connected determinant formalism can be applied to correlation functions as well. Here we use the nilpotent-polynomial algebraic formulation. We illustrate the method for the case of the bare expansion for the one-particle Green function

$$G_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, zV] = - \frac{\int \mathcal{D}[\varphi, \bar{\varphi}] e^{-S_{\text{bare}}[G_0, zV]} \varphi(\mathcal{Y}) \bar{\varphi}(\mathcal{Y}')}{\int \mathcal{D}[\varphi, \bar{\varphi}] e^{-S_{\text{bare}}[G_0, zV]}} = \frac{A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, zV]}{\tilde{Z}_{\text{bare}}[G_0, zV]}. \quad (91)$$

In analogy to what was done before for $\tilde{Z}$, we expand A in a formal series in z

$$A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0, zV] \stackrel{*}{=} \sum_{n=0}^{\infty} \frac{z^n}{n!} \int dX_1 \cdots dX_n V(X_1) \cdots V(X_n) A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0](\{X_1, \dots, X_n\}). \quad (92)$$

Using Wick's theorem, Eq. (63), it is possible to compute $A_{\text{bare}; \mathcal{Y}, \mathcal{Y}'}[G_0](S)$ and $\tilde{Z}_{\text{bare}}[G_0](S)$ in polynomial-in- n time for any set of vertex positions S in terms of determinants involving

matrices built from the non-interacting Green function G_0 . In terms of Feynman diagrams, $A_{\text{bare};\mathcal{Y},\mathcal{Y}'}(S)$ is the sum of all (i.e. connected and disconnected) symmetrized diagrams with S as internal vertices and having $\mathcal{Y}$ and $\mathcal{Y}'$ as external vertices.

Using the nilpotent polynomial formalism, it is immediate to derive $G_{\text{bare};\mathcal{Y},\mathcal{Y}'}[G_0](S)$: as before, we evaluate the functional $G_{\text{bare};\mathcal{Y},\mathcal{Y}'}[G_0, V]$ for $V(X) = V_\eta(X) = \sum_{j=1}^n \delta(X - X_j) \eta_j$

$$G_{\text{bare};\mathcal{Y},\mathcal{Y}'}[G_0, V_\eta] = \frac{A_{\text{bare};\mathcal{Y},\mathcal{Y}'}[G_0, V_\eta]}{\tilde{Z}_{\text{bare}}[G_0, V_\eta]}, \quad (93)$$

which is equivalent, by definition, to

$$G_{\text{bare};\mathcal{Y},\mathcal{Y}'}[[\boldsymbol{\eta}]] = \frac{A_{\text{bare};\mathcal{Y},\mathcal{Y}'}[[\boldsymbol{\eta}]]}{\tilde{Z}_{\text{bare}}[[\boldsymbol{\eta}]]}, \quad (94)$$

which means that computing connected diagrams for the Green function is equivalent to a nilpotent polynomial division.

In order to efficiently evaluate the nilpotent polynomial division, we write

$$G_{\text{bare};\mathcal{Y},\mathcal{Y}'}[[\boldsymbol{\eta}]] \tilde{Z}_{\text{bare}}[[\boldsymbol{\eta}]] = A_{\text{bare};\mathcal{Y},\mathcal{Y}'}[[\boldsymbol{\eta}]]. \quad (95)$$

The latter equation is a multiplication between nilpotent polynomials, which we know how to perform. Specifically, after removing the “bare” label and the explicit G_0 dependence

$$\sum_{S' \subseteq \{X_1, \dots, X_n\}} G_{\mathcal{Y},\mathcal{Y}'}(S') \sum_{S'' \subseteq \{X_1, \dots, X_n\}} \tilde{Z}(S'') \eta(S') \eta(S'') = \sum_{S \subseteq \{X_1, \dots, X_n\}} A_{\mathcal{Y},\mathcal{Y}'}(S) \eta(S), \quad (96)$$

which means that we can obtain a recursive relation for the Green function bare-expansion functional

$$G_{\mathcal{Y},\mathcal{Y}'}(S) = A_{\mathcal{Y},\mathcal{Y}'}(S) - \sum_{S' \subsetneq S} G_{\mathcal{Y},\mathcal{Y}'}(S') \tilde{Z}(S \setminus S'), \quad (97)$$

for all $S \subseteq \{X_1, \dots, X_n\}$, where we have used $\tilde{Z}(\emptyset) = 1$.

4.3.6 Self-energy

In order to compute one-particle properties, it is often advantageous to compute the self-energy instead of the Green function. In this section we describe the algorithm to build the bare-expansion self-energy functional used, e.g., in Ref. [21].

We specialize here the discussion to the bare spin-balanced normal-phase expansion for a translation-invariant system

$$G_0(\mathcal{Y}, \mathcal{Y}') = \delta_{\sigma,\sigma'} G_{0;\sigma}(Y - Y'), \quad (98)$$

where $\mathcal{Y} = (\sigma, Y)$, $\mathcal{Y}' = (\sigma', Y')$. In this case, Wick’s theorem implies

$$\tilde{Z}_{\text{bare}}[G_0](S) = (-1)^{|S|} \det \mathbb{G}_{0;\uparrow}(S) \det \mathbb{G}_{0;\downarrow}(S), \quad (99)$$

where $S = \{X_1, \dots, X_{|S|}\}$, and where we have introduced the matrix $[\mathbb{G}_{0;\sigma}(S)]^{jk} = G_{0;\sigma}(X_j, X_k)$ for $X_j, X_k \in S$.

We define the auxiliary quantity $\mathcal{A}$

$$\mathcal{A}_{\text{bare};\sigma;jk}[G_0](S) = \frac{\partial}{\partial G_{0;\sigma}(X_k, X_j)} \tilde{Z}_{\text{bare}}[G_0](S) = (-1)^{|S|} [\text{adj } \mathbb{G}_{0;\sigma}(S)]_{jk} \det \mathbb{G}_{0;-\sigma}(S), \quad (100)$$

where adj means the adjugate of a matrix. The diagrammatic interpretation of the function $\mathcal{A}_{\text{bare};\sigma,X_j,X_k}[G_0](S)$ is the sum of all diagrams for the partition function with fixed spacetime vertices S in which we remove a $G_{0;\sigma}(X_k, X_j)$ line.

From now on we impose, for simplicity, $G_{0;\uparrow} = G_{0;\downarrow}$. In order to remove disconnected diagrams, we can remove all diagrams that factorize in a partition function diagram and the connected part

$$\Xi_{jk}(S) = \mathcal{A}_{jk}(S) - \sum_{S' \subsetneq S} \Xi_{jk}(S') \tilde{Z}(S \setminus S'), \quad (101)$$

where we have identified vertices like X_k with their index k , and where we have used a shorthand notation for $\mathcal{A}_{\text{bare}}$, $\tilde{Z}_{\text{bare}}$ and Ξ_{bare} , the latter being an auxiliary quantity equal to the sum of all connected Feynman diagrams contributing to $\mathcal{A}_{\text{bare}}$, which is zero for $S = \emptyset$.

We introduce another auxiliary quantity ρ , defined by

$$\rho_k^j(S) = \sum_{l: X_l \in S} G_0(X_j, X_l) \Xi_{lk}(S), \quad (102)$$

for $X_j \notin S$.

We now have all the ingredients to evaluate the bare self-energy. The Hartree contribution Σ_{jj} is simply

$$\Sigma_{jj}(S) = \Xi_{jj}(S), \quad (103)$$

where $X_j \in S$. The non-Hartree contribution can be obtained from

$$\Sigma_{jk}(S) = \Xi_{jk}(S) - \sum_{l, S' \subsetneq S} \Sigma_{jl}(S') \rho_k^l(S \setminus S'), \quad (104)$$

where $X_j, X_k \in S$, $X_j \neq X_k$. Diagrammatically, Eq. (104) means that we are removing from Ξ all diagrams that can be split in a one-particle-irreducible part, $\Sigma_{jl}(S')$, and the convolution of a G_0 line with another piece, $\rho_k^l(S \setminus S')$. Algebraically, we are writing the self-energy as

$$\Sigma G G_0^{-1} = G_0^{-1} G G_0^{-1} - G_0^{-1}, \quad (105)$$

where $\rho = G G_0^{-1}$ and $\Xi = G_0^{-1} G G_0^{-1} - G_0^{-1}$. In order to derive Eq. (104) and Eq. (103), we could have then used

$$\Sigma[[\eta]] \rho[[\eta]] = \Xi[[\eta]], \quad (106)$$

where $\Sigma[[\eta]]$, $\rho[[\eta]]$ and $\Xi[[\eta]]$ are nilpotent-polynomial-valued matrices, and where we have defined $\rho(\emptyset) = \mathbb{I}$.

For a translation-invariant system, it is convenient to measure the self-energy directly in momentum space. This can be achieved by defining the Fourier transform of the non-Hartree part of the self-energy for a given set of interaction vertices S

$$\Sigma_K^{NH}(S) = \frac{1}{|S|(|S|-1)} \sum_{j \neq k} \Sigma_{jk}(S) e^{-iK \cdot (X_j - X_k)}, \quad (107)$$

where $K = (i\omega_n, \mathbf{k})$ is a collective momentum variable made of a fermionic Matsubara frequency ω_n and a spatial momentum $\mathbf{k}$. The Hartree part of the self-energy gives a constant contribution to the momentum self-energy.

In terms of computational cost, to compute the sum of all one-particle-irreducible diagrams with n vertices, symmetrized over the vertex positions and over the n^2 possible choices of pairs for the external points, the algorithm requires computing 2^n determinants and adjugates, which can be done in $\mathcal{O}(n^3 2^n)$ or faster, and $\mathcal{O}(n^2 3^n)$ for the elimination of disconnected and one-particle reducible diagrams.

4.3.7 One-particle renormalization

We have seen that within the shifted-action formalism, one-particle renormalization can be achieved by letting the non-interacting one-particle propagator be a function of the coupling constant z

$$S_{\text{bare}}[G_0(z), zV]. \quad (108)$$

Diagrammatically, it is natural to consider expansion schemes such that subdiagrams consisting of low-order self-energy insertions do not contribute at higher orders. This procedure can be applied at various levels of self-consistency, either at the functional level or at the z -function level, the latter being beyond the scope of this document.

We decide here to focus on the most challenging-to-derive algorithm for one-particle renormalization: the full renormalization of the one-particle propagator. Most of the other possible expansion schemes are specializations or adaptations of this algorithm in simpler settings, which justifies the choice of including the bold expansion in this document. The bold expansion scheme is defined by (see also Eq. (37) and Eq. (38))

$$G_{0;\text{bold}}[G, V]^{-1} = G^{-1} + \Sigma_{\text{bold}}[G, V]. \quad (109)$$

We fix the order of the expansion to n , and then fix a given set of n vertices, $E = \{X_1, \dots, X_n\}$. We write Eq. (109) as a nilpotent polynomial equation by evaluating the functionals for $V = V_\eta$, and we drop the G dependence

$$\mathbb{G}_{0;\text{bold}}[[\boldsymbol{\eta}]] = \mathbb{G} - \mathbb{G}_{0;\text{bold}}[[\boldsymbol{\eta}]] \boldsymbol{\Sigma}_{\text{bold}}[[\boldsymbol{\eta}]] \mathbb{G}, \quad (110)$$

where $\mathbb{G}_{0;\text{bold}}$, $\boldsymbol{\Sigma}$ and $\mathbb{G}$ are $n \times n$ matrices. The difficulty of solving the previous equation for $\mathbb{G}_{0;\text{bold}}[[\boldsymbol{\eta}]]$ is that $\boldsymbol{\Sigma}_{\text{bold}}[[\boldsymbol{\eta}]]$ is an unknown functional of G to be determined from the equation

$$\boldsymbol{\Sigma}_{\text{bare}}[\mathbb{G}_{0;\text{bold}}[[\boldsymbol{\eta}]]][[\boldsymbol{\eta}]] = \boldsymbol{\Sigma}_{\text{bold}}[[\boldsymbol{\eta}]]. \quad (111)$$

Suppose that we have determined $\Sigma_{\text{bold}}(S)$ for all $S \subseteq E$ with $|S| \leq k$. We can then first obtain $\Sigma_{\text{bold}}(S) \mathbb{G}$, and then compute $\mathbb{G}_{0;\text{bold}}(S)$ for $|S| \leq k$ from Eq. (110). The crucial observation is that, in order to obtain $\Sigma_{\text{bold}}(S')$ for $|S'| = k+1$ we only need to know $\mathbb{G}_{0;\text{bold}}(S)$ for $|S| \leq k$, which means that we can solve Eq. (111) recursively. More specifically, in order to build the nilpotent polynomial $\Sigma_{\text{bare}}[\mathbb{G}_{0;\text{bold}}[[\eta]]][[\eta]]$ for $|S| = k+1$ we need to compute $\mathcal{A}$ by using $G_{0;\text{bold}}[[\eta]]$, computing Ξ , ρ , and finally Σ_{bare} . Formally, the algorithm is identical to the bare one, the only difference being that the operations must be done in the ring of nilpotent polynomials.

In terms of computational cost, the computation of $\mathcal{A}$ is the most expensive part. $\mathcal{A}(S)$ involves computing the adjugate and the determinant of a matrix containing nilpotent polynomials in $|S|$ variables, whose cost can be bounded assuming the use of the naive Gaussian elimination by $\mathcal{O}(|S|^3 3^{|S|})$. This means that the total cost to compute the sum of all symmetrized skeleton diagrams contributing to the self-energy for all pairs of vertices as external points is bounded by $\mathcal{O}(n^3 4^n)$, much better than the $\mathcal{O}(n^2 (n!)^2)$ naive enumeration cost. This derivation is useful as a similar methodology can be used for many other one-particle renormalization schemes. Ref. [22] presents a method based on the combinatorial summation of Feynman diagrams that is able to reduce the scaling of the evaluation of bold functionals to $\mathcal{O}(n^3 3^n)$.

4.3.8 Sampling vertex positions with Many-Configuration Markov-Chain Monte Carlo

Integrating over vertex positions. So far we have discussed how to compute the sum of all symmetrized Feynman diagrams for a given quantity O in a given expansion scheme at fixed spacetime positions $X_1, \dots, X_n$ of the interaction vertices, which we call $O_{\text{exp}}(\{X_1, \dots, X_n\})$. The purpose of this section is to show how to perform the integration over the positions of the internal ensemble of vertices $\{X_1, \dots, X_n\}$

$$O_{\text{exp},n} = \frac{1}{n!} \int dX_1 \cdots dX_n V(X_1) \cdots V(X_n) O_{\text{exp}}(\{X_1, \dots, X_n\}), \quad (112)$$

to obtain the final estimate for the order- n coefficient of the expansion, $O_{\text{exp},n}$. The integration over $X_1, \dots, X_n$ does not present many difficulties. Even for an infinite system, as we are computing connected quantities, $X_1, \dots, X_n$ cluster together and the integral is finite (a trivial exception being the grand potential, where one has to fix the position of one of the vertices to compute the well-defined grand-potential density).

For the d -dimensional Hubbard model, the integration runs over d spatial components and one imaginary time component, making the total number of integration variables equal to $(d+1)n$ for an order n calculation. As $n = \mathcal{O}(10)$, the integration is high-dimensional and not amenable to most deterministic techniques. For this reason, the use of Monte Carlo sampling is essential to obtain a statistical estimate of the value of the integral. A simple order-changing Markov-chain Monte Carlo algorithm that includes an order-changing update is sufficient to obtain an estimate of the integral. However, as a byproduct of the diagram summing procedure we have calculated $O_{\text{exp}}(S)$ for all $S \subseteq \{X_1, \dots, X_n\}$, and therefore it is natural to look for more efficient ways to use that information. Essentially, at fixed n , we would like to consider all configurations S

as possible updates for the Markov-chain Monte Carlo procedure. This is achieved with the Many-configuration Markov-chain Monte Carlo algorithm (MC³) (see Ref. [23]), which we briefly sketch below.

Markov-chain of sets of configurations. In the following, we use for simplicity of presentation a lighter notation. Suppose the goal is to obtain a Monte Carlo estimate of integrals of this form

$$O_k = \sum_{S \subseteq \mathcal{E}: |S|=k} O(S), \quad (113)$$

for $k \in \{0, 1, \dots, n\}$, where $\mathcal{E}$ is an abstract set (aka ensemble) defining the configuration space. Suppose that, given an ensemble E contained in $\mathcal{E}$, $E \subseteq \mathcal{E}$, one has automatically obtained in some way all $O(S)$ for each subset S of E , $S \subseteq E$. The challenge is to find an efficient Markov-chain Monte Carlo algorithm that maximally takes advantage of this information. One could consider taking an equilibrium probability distribution for the Markov chain proportional to $|O(S)|$, and perform a Markov chain over sets of different cardinality. The problem is that, with this formulation, only the information on one set S is used at each Monte Carlo step. We are therefore led to consider some sort of all-subset probability distribution of the form $\sum_{S \subseteq E} |O(S)|$. The problem with the latter attempt is that we are summing over functions with a different number of variables, and therefore we cannot even define the configuration space for the Markov chain. This can be solved in the following way: for a given subset S of the set E , we generate the missing elements $E \setminus S$ with a certain probability distribution $p_{\text{gen}}(E \setminus S | S)$, such that $p_{\text{gen}}(E \setminus S | S) |O(S)|$ is now proportional to a proper probability distribution in the configuration space of sets with $|E| = n$ elements. More precisely, given an ensemble E of cardinality n , we define our target equilibrium probability distribution for the Markov chain as

$$p_{\text{eq}}(E) = \frac{\sum_{S \subseteq E} p_{\text{gen}}(E \setminus S | S) \lambda_{|S|} |O(S)|}{\mathcal{N}_n}, \quad (114)$$

where we have added a weight $\lambda_{|S|} \geq 0$ for later convenience, and $\mathcal{N}_n$ is an unknown normalization factor. With the ensemble probability distribution $p_{\text{eq}}(E)$, it is possible to obtain a Monte Carlo estimator for O_k (see Eq. (113)), for $k \in \{0, 1, \dots, n\}$

$$O_k = \mathcal{N}_n \left\langle \sum_{S \subseteq E: |S|=k} \frac{p_{\text{gen}}(E \setminus S | S) O(S)}{\sum_{S' \subseteq E} p_{\text{gen}}(E \setminus S' | S') \lambda_{|S'|} |O(S')|} \right\rangle_{p_{\text{eq}}}. \quad (115)$$

The normalization $\mathcal{N}_n$ can be obtained from the zeroth order estimator, which plays the role of an effective normalization sector for the Markov chain

$$\frac{O_0}{O(\emptyset)} = 1 = \mathcal{N}_n \left\langle \frac{p_{\text{gen}}(E | \emptyset)}{\sum_{S' \subseteq E} p_{\text{gen}}(E \setminus S' | S') \lambda_{|S'|} |O(S')|} \right\rangle_{p_{\text{eq}}}. \quad (116)$$

Detailed balance condition. Our goal is to build a Markov chain of ensembles $E \rightarrow E' \rightarrow \dots$, where $|E| = |E'| = n$, that has $p_{\text{eq}}(E)$ as equilibrium probability distribution. We focus

on the following natural update choice for the set E : given the ensemble E , we pick one of the subsets $S \subseteq E$ with probability $\frac{p_{\text{gen}}(E \setminus S | S) \lambda_{|S|} |O(S)|}{\sum_{S' \subseteq E} p_{\text{gen}}(E \setminus S' | S') \lambda_{|S'|} |O(S')|}$, and we generate E' with the probability distribution $p_{\text{gen}}(E' \setminus S | S)$. This stochastic process defines the probability of update $p_{\text{update}}(E' | E)$ from the ensemble E to the ensemble E' of the same cardinality ($|E| = |E'| = n$)

$$p_{\text{update}}(E' | E) = \frac{\sum_{S \subseteq E \cap E'} p_{\text{gen}}(E \setminus S | S) \lambda_{|S|} |O(S)| p_{\text{gen}}(E' \setminus S | S)}{\sum_{S' \subseteq E} p_{\text{gen}}(E \setminus S' | S') \lambda_{|S'|} |O(S')|}. \quad (117)$$

We write the detailed balance condition

$$p_{\text{update}}(E' | E) p_{\text{eq}}(E) = p_{\text{update}}(E | E') p_{\text{eq}}(E'), \quad (118)$$

and we see that it is verified independently even at the level of each intermediate subset S (hyperdetailed balance).

This update strategy also motivates reasonable choices for λ_k : we choose it in such a way that each order is “chosen” to build the new ensemble a certain fraction of the time. It is particularly important to choose the empty subset $\emptyset$ a non-negligible fraction of the time to have a reasonably well-estimated normalization factor.

Generating sets of configurations. We discuss how to obtain a natural form of $p_{\text{gen}}(E \setminus S | S)$, i.e., the probability of generating the subset $E \setminus S$ starting from the subset S . A natural choice for this distribution can be built starting from the single-vertex probability $p_{\text{gen}}(\{X\} | S)$, which is a probability distribution function that adds a vertex X given the subset of interaction vertices S . A simple and natural choice for $p_{\text{gen}}(\{X\} | S)$ is

$$p_{\text{gen}}(\{X\} | S) = \frac{1}{|S|} \sum_{X_j \in S} p_{\text{gen}}(\{X\} | \{X_j\}), \quad (119)$$

where $p_{\text{gen}}(\{X\} | \{X_j\})$ is the probability distribution of generating the vertex X from the vertex X_j . We also need to define $p_{\text{gen}}(\{X\} | \emptyset)$. Assuming for simplicity that we are considering the grand potential, which is translation invariant, a natural choice can be a uniform probability distribution over spacetime for $p_{\text{gen}}(\{X\} | \emptyset)$.

This formulation has the operational interpretation of uniformly sampling a vertex belonging to the subset S , $X_j \in S$, and generating the vertex X with the probability distribution $p_{\text{gen}}(\{X\} | \{X_j\})$. If we want to generate more than one element, following this stochastic procedure, there is more than one possibility. For instance, to generate the subset $\{X_1, X_2\}$, we can first generate X_1 and then X_2 , or in the opposite order. We decide, as a natural choice, to average over all these possibilities. For every subset $S \subseteq \{X_1, \dots, X_n\}$, we can then write down the probability $p_{\text{gen}}(S' | S)$ of generating the set S' , $S' \cap S = \emptyset$, from the set S , as

$$p_{\text{gen}}(S' | S) = \frac{1}{|S'|} \sum_{X \in S'} p_{\text{gen}}(\{X\} | S \cup S' \setminus \{X\}) p_{\text{gen}}(S' \setminus \{X\} | S), \quad (120)$$

which has a runtime cost of $\mathcal{O}(n 2^n)$.

5 Summing the expansion: Convergence, singularities and resummation

5.1 Finite radius of convergence and polynomial complexity

It is possible to show that (see e.g. [24]), for all non-zero temperatures, the radius of convergence of the bare expansion is non-zero. More precisely, for quantities such as the grand potential Ω , there exists $R > 0$ such that for $|z| < R$

$$\left| \Omega_{\text{bare}}(z) - \sum_{n=0}^N z^n \Omega_{\text{bare},n} \right| \leq A_\epsilon \left(\epsilon + \frac{|z|}{R} \right)^{N+1}, \quad (121)$$

for every $\epsilon > 0$. The behavior of the radius of convergence R as a function of β depends on the instabilities of the $U=0$ ground state with respect to small U perturbations: if the non-interacting system is stable at zero temperature for infinitesimal U , which is the case when the system is gapped, then the radius of convergence will be non-zero even at zero temperature. Typically, if a Fermi surface is present, it is unstable at zero temperature towards superconducting, spin, or other ordered states for either $U > 0$ or $U < 0$, or both. This means that the zero-temperature radius of convergence is zero. However, for a fixed U , these Fermi-surface instabilities happen at exponentially low temperature, which means that the inverse of the radius of convergence increases only logarithmically with inverse temperature.

If the radius of convergence is larger than 1, truncating the $z=1$ series at the truncation order n converges exponentially with n . As we have shown that evaluating the coefficient of the series has a cost that increases exponentially with expansion order, the total computational cost $t(\epsilon)$ to obtain a precision of ϵ on a given quantity is polynomial inside the convergence radius (see Ref. [25])

$$t(\epsilon) \sim \epsilon^{-\alpha}, \quad (122)$$

where we have defined $\alpha = \log a / \log R$, and we have assumed that the computational cost to obtain the order n scales as a^n . This result is important as it shows that interacting fermionic quantum matter on the lattice at finite temperature can be simulated with a polynomial-time algorithm, at least in a finite region of the phase diagram. Renormalization allows one to change the region in which this polynomial-time simulation can be achieved by enlarging the radius of convergence. Moreover, using appropriate shifted actions we are able to enter different phases of matter by changing the perturbative starting point.

5.2 Qualitative structure of the singularities and analytic continuation

We now sketch the qualitative structure of the singularities of the bare, non-symmetry-breaking expansion of the two-dimensional Hubbard model, viewed as a function of the complex interaction strength $U \in \mathbb{C}$. This perspective is useful for understanding how the expansion can, in some cases, be analytically continued beyond its radius of convergence.

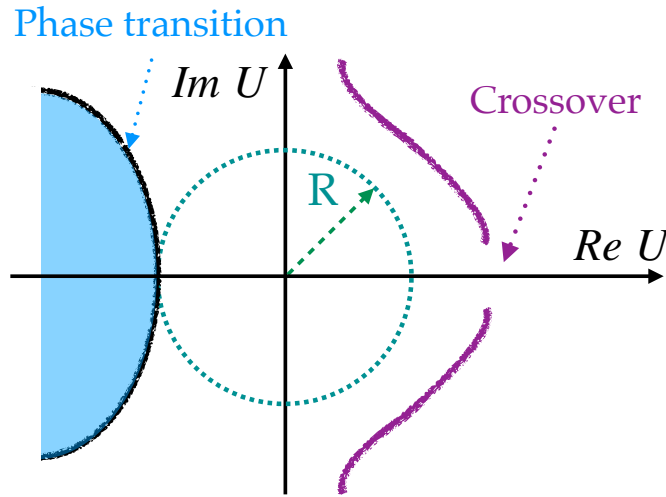


Fig. 5: Qualitative sketch of the complex- U analytic structure of the Hubbard model.

For $\text{Re } U < 0$, the interaction is attractive, and the system typically becomes unstable toward an s -wave superfluid phase. This singularity often sets the radius of convergence of the bare expansion in the doped finite-temperature two-dimensional Hubbard model. The circle $|U| = R$ denotes this radius of convergence; for $\text{Re } U > 0$, however, nothing physical necessarily occurs on this circle. At zero temperature and for repulsive interactions, $\text{Re } U > 0$, the system is unstable toward antiferromagnetism sufficiently close to half filling. At finite temperature, the corresponding antiferromagnetic singularities do not reach the real axis. As a result, analytic continuation remains possible in principle, while these nearby singularities give rise physically to a crossover regime with strong antiferromagnetic correlations.

Figure 5 illustrates the typical qualitative structure of the singularities in the complex U plane in the finite-temperature, doped regime of the two-dimensional Hubbard model. For $\text{Re } U < 0$, the interaction is attractive. When $\text{Re } U$ is sufficiently negative, the system typically becomes an s -wave superfluid. The corresponding singularity often determines the radius of convergence of the expansion, and manifests itself in perturbative coefficients that alternate in sign as a function of expansion order.

The circle $|U| = R$, where R denotes the radius of convergence, is only a mathematical boundary of the Taylor expansion. Nothing special happens on this circle except at the singularity, typically located in the attractive region $\text{Re } U < 0$. Therefore, in principle, there is no obstruction to analytically continuing the function beyond this circle in other directions of the complex U plane.

For positive real U , several scenarios are possible. First, there may be no additional singularity beyond the one associated with the attractive side. Second, there may be a singularity on the positive real axis, corresponding to a physical finite-temperature phase transition. In this case, analytic continuation from weak coupling cannot reach the point of interest, since it lies in a different phase of matter. Third, there may be no singularity on the positive real axis, but there may be nearby complex singularities associated with transitions occurring at lower temperature, or possibly at zero temperature. Such singularities do not prevent analytic continuation in principle, but they can generate strong crossover behavior.

Thus, in the first and third scenarios, analytic continuation beyond the radius of convergence is possible in principle and can be attempted using resummation techniques. In practice, however, the difficulty increases as singularities approach the positive real axis. Since one usually has access only to a finite number of expansion orders, typically of order $\mathcal{O}(10)$, there is a practical limit to the regions of the phase diagram, in temperature and interaction strength, that can be reliably reached by resummation.

6 Conclusions

We have introduced Diagrammatic Monte Carlo as a way to exchange the exponential scaling of the many-body system with a series summation problem. Focusing on the Hubbard model, we have discussed the choice of the diagrammatic expansion, how to efficiently evaluate Feynman diagrams, and how to sum the resulting series based on the analytic properties of the expansion function. We have shown that recent algorithms for summing all Feynman diagrams make it possible to achieve a polynomial-time solution of the many-electron problem in some parameter regimes. This has made it possible to characterize the finite-temperature regime of the two-dimensional doped Hubbard model, notably the pseudogap and the magnetic crossovers, in a controlled way. Future work in the Hubbard model should be devoted to extending the range of applicability of the technique with the help of new renormalization schemes and strong-coupling descriptions, and to entering broken-symmetry phases including stripes and the long-conjectured d -wave superconducting state.

References

- [1] J. Hubbard, Proc. R. Soc. Lond. A **276**, 238 (1963)
- [2] D.P. Arovas, E. Berg, S.A. Kivelson, and S. Raghu, Annu. Rev. Condens. Matter Phys. **13**, 239 (2022)
- [3] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, Annu. Rev. Condens. Matter Phys. **13**, 275 (2022)
- [4] R. Rossi, F. Werner, N. Prokof'ev, and B. Svistunov, Phys. Rev. B **93**, 161102(R) (2016)
- [5] R. Rossi, F. Šimkovic, and M. Ferrero, EPL **132**, 11001 (2020)
- [6] G. Spada, R. Rossi, F. Šimkovic, R. Garioud, M. Ferrero, K. Van Houcke, and F. Werner, arXiv:2103.12038
- [7] R. Garioud, F. Šimkovic, R. Rossi, G. Spada, T. Schäfer, F. Werner, and M. Ferrero, Phys. Rev. Lett. **132**, 246505 (2024)
- [8] E. Kozik, M. Ferrero, and A. Georges, Phys. Rev. Lett. **114**, 156402 (2015)
- [9] R. Rossi and F. Werner, J. Phys. A: Math. Theor. **48**, 485202 (2015)
- [10] N.V. Prokof'ev and B.V. Svistunov, Phys. Rev. Lett. **99**, 250201 (2007)
- [11] K. Van Houcke, E. Kozik, N. Prokof'ev, and B. Svistunov, Phys. Procedia **6**, 95 (2010)
- [12] E. Kozik, K. Van Houcke, E. Gull, L. Pollet, N. Prokof'ev, B. Svistunov, and M. Troyer, EPL **90**, 10004 (2010)
- [13] N. Prokof'ev: *Diagrammatic Monte Carlo*, in E. Pavarini, E. Koch, and S. Zhang (eds.): *Many-Body Methods for Real Materials, Modeling and Simulation Vol. 9* (Forschungszentrum Jülich, 2019)
- [14] A.N. Rubtsov, V.V. Savkin, and A.I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005)
- [15] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)
- [16] R.M. Karp, in R.E. Miller and J.W. Thatcher (eds.): *Complexity of Computer Computations* (Plenum, New York, 1972), pp. 85–103
- [17] M. Held and R.M. Karp, J. Soc. Indust. Appl. Math. **10**, 196 (1962)
- [18] R. Bellman, J. ACM **9**, 61 (1962)
- [19] R. Rossi, Phys. Rev. Lett. **119**, 045701 (2017)

- [20] A. Björklund, T. Husfeldt, P. Kaski, and M. Koivisto, in *Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (STOC 2007), pp. 67–74
- [21] F. Šimkovic, R. Rossi, A. Georges, and M. Ferrero, *Science* **385**, ade9194 (2024)
- [22] E. Kozik, *Nat. Commun.* **15**, 7916 (2024)
- [23] F. Šimkovic and R. Rossi, arXiv:2102.05613
- [24] G. Benfatto, A. Giuliani, and V. Mastropietro, *Ann. Henri Poincaré* **7**, 809 (2006)
- [25] R. Rossi, N. Prokof'ev, B. Svistunov, K. Van Houcke, and F. Werner, *EPL* **118**, 10004 (2017)

7 Mean-Field Theory of Spin and Charge Order in the Two-Dimensional Hubbard Model

Walter Metzner

Max Planck Institute for Solid State Research

D-70569 Stuttgart, Germany

Contents

1	Introduction	2
2	Unrestricted Hartree-Fock theory	3
2.1	Hubbard model	3
2.2	Hartree-Fock equations in real space	4
2.3	Classification of magnetic states	5
2.4	Results from unrestricted real space calculations	9
3	Mean-field solutions in the thermodynamic limit	11
3.1	Néel and spiral states	12
3.2	Commensurate coplanar states	13
3.3	Landau theory	17
3.4	Spiral to stripe transition and comprehensive phase diagrams	21
4	Renormalized mean-field theory	22
A	Landau coefficients	24

1 Introduction

The two-dimensional Hubbard model is a prototype for competing ordering tendencies in interacting electron systems. It captures key features of the valence electrons in high- T_c cuprates such as antiferromagnetism and d -wave superconductivity [1]. In spite of substantial progress in the development of quantum many body methods, only fragments of the phase diagram of this important model have been firmly established [2, 3].

There is no doubt that at half-filling the ground state of the two-dimensional Hubbard model with pure nearest neighbor hopping is a Néel ordered antiferromagnet for any repulsive Hubbard interaction $U > 0$, and beyond a certain critical repulsion if hopping amplitudes beyond nearest neighbors are present. However, for electron densities away from half-filling, the Néel state easily becomes unstable [4–7], leading to other magnetic states or possibly phase separation. Moreover, the Hubbard model away from half-filling is usually metallic and thus prone to pairing instabilities.

There is a whole zoo of possible magnetic states. Unlike the Néel state, most of them are characterized by one or several ordering wave vectors which are generically incommensurate with the lattice periodicity. Many approximate calculations point toward planar circular spirals [4–16] and collinear spin density waves [17–25] as the most important candidates. The latter entail charge order with regions of reduced charge density arranged along one-dimensional lines, and are therefore also known as “spin-charge stripes” [26]. By contrast, the charge distribution of circular spiral states remains uniform.

Recently, exact numerical calculations on finite lattices have provided strong evidence for stripe order in the ground state of the Hubbard model with pure nearest neighbor hopping and large U for special hole-doping fractions: $1/10$ and $1/8$, corresponding to electron densities 0.9 and 0.875 , respectively [27, 28]. No superconductivity was found in these cases. However, allowing for next-nearest neighbor hopping, superconductivity reemerges for a range of densities away from half-filling [29]. Exact numerical evaluations at finite temperatures yield peaks in the spin susceptibility at the Néel wave vector (π, π) near half-filling, and peaks at incommensurate wave vectors in the spin and charge susceptibility for sufficiently large doping [30, 31].

Unlike superconductivity, magnetic order in the (repulsive) Hubbard model can be captured already within a static mean-field approximation, also known as Hartree-Fock theory. While the size of the order parameter and the regime of ordered states in the phase diagram are grossly overestimated by mean-field theory, finite size effects can be studied in more detail because much larger lattices can be dealt with. For Néel, circular spiral, and stripe states, mean-field calculations can be performed directly in the thermodynamic limit, albeit only for wave vectors commensurate with the lattice in the case of stripes. Indeed, numerous Hartree-Fock studies of the two-dimensional Hubbard model have been carried out. However, in many of them the magnetic states were restricted to ferromagnetic and Néel [32–34], or spiral order [12]. The latter includes ferromagnetic and Néel states as the special cases with wave vectors $(0, 0)$ and (π, π) , respectively. For small hole doping, this restriction often leads to phase separation in paramagnetic, ferromagnetic, and antiferromagnetic regions [12, 34]. Allowing for arbitrary

collinear spin order or even for completely arbitrary spin configurations – combined with charge order – spin-charge stripes have been found [17–21].

At finite temperatures, magnetic long-range order in two-dimensional systems with SU(2) spin symmetry is prohibited by thermal fluctuations, as is rigorously established by the Mermin-Wagner theorem [35]. Magnetic states obtained in Hartree-Fock approximation at finite temperatures obviously violate this theorem. However, such states become meaningful as an ingredient for theories of fluctuating magnetic order, where the electron is fractionalized into a “chargon” with a magnetically ordered pseudospin, and a fluctuating spin rotation matrix, which conserves the SU(2) spin symmetry of the electronic state [36–38].

A comprehensive and completely unrestricted Hartree-Fock study of the two-dimensional Hubbard model at zero and finite temperatures has been performed only recently [39, 40]. Unrestricted real space calculations on large lattices (from 20×20 to 48×48) have been complemented by momentum space calculations for Néel, spiral, and commensurate stripe states in the thermodynamic limit. In this lecture I describe the various approaches for solving the Hartree-Fock equations, and I review the results for the two-dimensional Hubbard model with a moderate interaction strength, in particular its magnetic phase diagram as a function of density and temperature. Both the “pure” Hubbard model, with pure nearest neighbor hopping t , as well as its physically more relevant extension with a next-nearest neighbor hopping t' will be considered.

2 Unrestricted Hartree-Fock theory

2.1 Hubbard model

The Hubbard Hamiltonian for spin- $\frac{1}{2}$ fermions with intersite hopping amplitudes $t_{jj'}$ and a local interaction U reads

$$H = H_0 + H_{\text{int}} = \sum_{j,j',\sigma} t_{jj'} c_{j\sigma}^\dagger c_{j'\sigma} + U \sum_j n_{j\uparrow} n_{j\downarrow}, \quad (1)$$

where $c_{j\sigma}$ ($c_{j\sigma}^\dagger$) annihilates (creates) an electron on lattice site j with spin orientation σ ($\uparrow$ or $\downarrow$), and $n_{j\sigma} = c_{j\sigma}^\dagger c_{j\sigma}$. We consider exclusively the Hubbard model on a square lattice. The hopping matrix $t_{jj'}$ depends only on the distance between the sites j and j' . We restrict it to nearest neighbor contributions with an amplitude $-t$, and to next-nearest neighbor contributions with an amplitude $-t'$. The bare dispersion relation is then given by

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

We focus on the repulsive Hubbard model where U is positive, corresponding to a screened Coulomb repulsion between electrons. The repulsive two-dimensional Hubbard model is the most intensively studied model for the electrons moving in the copper-oxide planes of cuprate high temperature superconductors [1]. We use the nearest neighbor hopping amplitude t as our energy unit, that is, we set $t = 1$ in numerical results.

2.2 Hartree-Fock equations in real space

In the mean-field (Hartree-Fock) approximation, the interaction part of the Hamiltonian H_{int} is replaced by effective quadratic (in $c, c^\dagger$) terms. Following Zaanen and Gunnarsson [18], we decouple H_{int} as

$$H_{\text{int}}^{\text{MF}} = \sum_{j\sigma} \Delta_{j\bar{\sigma}} n_{j\sigma} - \sum_j (\Delta_{j-} c_{j\uparrow}^\dagger c_{j\downarrow} + \Delta_{j+} c_{j\downarrow}^\dagger c_{j\uparrow}) - \frac{1}{U} \sum_j (\Delta_{j\uparrow} \Delta_{j\downarrow} - \Delta_{j-} \Delta_{j+}), \quad (3)$$

with the convention $\bar{\uparrow} = \downarrow$ and $\bar{\downarrow} = \uparrow$. The parameters $\Delta_{j\alpha}$ are self-consistently determined by the “gap equations”

$$\Delta_{j\sigma} = U \langle n_{j\sigma} \rangle, \quad (4a)$$

$$\Delta_{j+} = \Delta_{j-}^* = U \langle c_{j\uparrow}^\dagger c_{j\downarrow} \rangle. \quad (4b)$$

The decoupling in Eq. (3) allows for arbitrary spin and charge patterns. In most of the previous mean-field studies of the Hubbard model, only the first (Hartree) or the second (Fock) term on the right hand side of Eq. (3) has been considered, restricting the spin order to collinear or spiral order, respectively.

The mean-field equations can be solved iteratively on finite lattices. We denote the total number of lattice sites by $L = L_x L_y$, where L_x and L_y is the number of sites in x and y direction, respectively. The mean-field Hamiltonian can be written in matrix form as

$$H^{\text{MF}} = H_0 + H_{\text{int}}^{\text{MF}} = \sum_{j,j'} \sum_{\sigma,\sigma'} c_{j\sigma}^\dagger \mathcal{H}_{jj'}^{\sigma\sigma'} c_{j'\sigma'} + \text{const.}, \quad (5)$$

where $\mathcal{H}_{jj'}^{\sigma\sigma'} \in \mathbb{C}^{2L \times 2L}$ is a square matrix with $4L$ real parameters to be determined self-consistently ($\Delta_{j\uparrow}$, $\Delta_{j\downarrow}$, $\text{Re } \Delta_{j+}$, and $\text{Im } \Delta_{j+}$ for each lattice site). These parameters are related to the expectation values of the charge and spin order parameters by

$$\Delta_{j\uparrow} + \Delta_{j\downarrow} = U \langle n_j \rangle, \quad (6a)$$

$$\Delta_{j\uparrow} - \Delta_{j\downarrow} = 2U \langle S_j^z \rangle, \quad (6b)$$

$$\Delta_{j+} + \Delta_{j-} = 2U \langle S_j^x \rangle, \quad (6c)$$

$$\Delta_{j+} - \Delta_{j-} = 2iU \langle S_j^y \rangle, \quad (6d)$$

where $n_j = n_{j\uparrow} + n_{j\downarrow}$ and $\vec{S}_j = \frac{1}{2} c_j^\dagger \vec{\sigma} c_j$, with the Pauli matrices $\vec{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$.

The calculations can be preformed at fixed density n by enforcing the constraint $L^{-1} \sum_{j\sigma} \langle n_{j\sigma} \rangle = n$. In the ground state this is achieved by calculating the expectation values on the right-hand side of Eq. (4) as

$$\langle c_{j\sigma}^\dagger c_{j\sigma'} \rangle = \sum_{\ell=1}^N (v_{j\sigma}^\ell)^* v_{j\sigma'}^\ell, \quad (7)$$

where N is the total particle number, ℓ labels the $2L$ eigenvalues ε_ℓ of $\mathcal{H}_{jj'}^{\sigma\sigma'}$ in ascending order, and $v_{j\sigma}^\ell$ is the normalized eigenvector corresponding to the ℓ -th eigenvalue. At finite temperatures $T > 0$, one needs to introduce a chemical potential μ , and Eq. (7) is replaced by

$$\langle c_{j\sigma}^\dagger c_{j\sigma'} \rangle = \sum_{\ell=1}^{2L} (v_{j\sigma}^\ell)^* v_{j\sigma'}^\ell f(\varepsilon_\ell - \mu), \quad (8)$$

with $f(\varepsilon) = 1/(1+e^{\varepsilon/T})$ the Fermi distribution function. The chemical potential is adjusted to enforce the chosen value for the average density

$$L^{-1} \sum_j \langle n_j \rangle = L^{-1} \sum_{\ell=1}^{2L} f(\varepsilon_{\ell} - \mu) = n. \quad (9)$$

This relation is obtained via the orthonormality property of the eigenvectors $\sum_{j,\sigma} (v_{j\sigma}^{\ell})^* v_{j\sigma}^{\ell} = 1$. The system of mean-field equations can be solved in the following iterative manner: A random initial set of parameters $\Delta_{j\alpha}$ with $\alpha \in \{\uparrow, \downarrow, +, -\}$ and $\Delta_{j-} = \Delta_{j+}^*$, is plugged into $\mathcal{H}_{jj'}^{\sigma\sigma'}$, from which the eigenvectors and the expectation values in Eq. (4) are computed to update the values of $\Delta_{j\alpha}$. This procedure is repeated until convergence is reached. At finite temperatures, the chemical potential is computed at every iteration from Eq. (9).

Within the mean-field approximation, the free energy is given by

$$F = -T \sum_{\ell} \log [1 + e^{-\beta(\varepsilon_{\ell} - \mu)}] - \frac{1}{U} \sum_j (\Delta_{j\uparrow} \Delta_{j\downarrow} - \Delta_{j-} \Delta_{j+}) + \mu N. \quad (10)$$

2.3 Classification of magnetic states

The zoo of distinct magnetic states obtained from the unrestricted mean-field calculations can be classified solely according to their spin pattern [41]. The modulation of the charge density is roughly proportional to the square of the spin amplitudes, that is,

$$\langle n_j \rangle \approx \text{const.} \langle \vec{S}_j^2 \rangle + \text{const.}, \quad (11)$$

as can be deduced from the lowest order coupling between the charge and spin density wave order parameters in a Landau theory [42].

To identify a state from a (converged) real space spin pattern defined by the spin expectation values $\langle \vec{S}_j \rangle$, it is convenient to use a Fourier representation,

$$\langle \vec{S}_j \rangle = \sum_{\mathbf{q}} \vec{S}_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{r}_j}, \quad (12)$$

where $\vec{S}_{\mathbf{q}}$ with $\vec{S}_{\mathbf{q}} = \vec{S}_{-\mathbf{q}}^*$ is the Fourier component of the spins with wave vector $\mathbf{q}$, and $\mathbf{r}_j$ represents the real space coordinate of lattice site j . The sum runs over the $L = L_x L_y$ momenta allowed by the periodic boundary conditions on the finite lattice. Restricting L_x and L_y to even numbers, the allowed momenta can be written as

$$\mathbf{q} \in \{ (\pi - 2\pi\nu_x/L_x, \pi - 2\pi\nu_y/L_y) \mid \nu_{\alpha} \in \{0, 1, \dots, L_{\alpha}-1\} \}. \quad (13)$$

In the majority of cases only one or two modes with fixed wave vectors $\mathbf{q}$ (together with their partner $-\mathbf{q}$) contribute significantly to $\langle \vec{S}_j \rangle$. Usually, these wave vectors have the form $\mathbf{Q}_x = (\pi - 2\pi\eta_x, \pi)$ and/or $\mathbf{Q}_y = (\pi, \pi - 2\pi\eta_y)$. The parameters η_{α} are frequently referred to as “incommensurabilities” in the literature. In a finite system, η_{α} is restricted to integer multiples of $1/L_{\alpha}$. For lattices with $L_x = L_y$, one always finds $\eta_x = \eta_y \equiv \eta$. States with diagonal wave

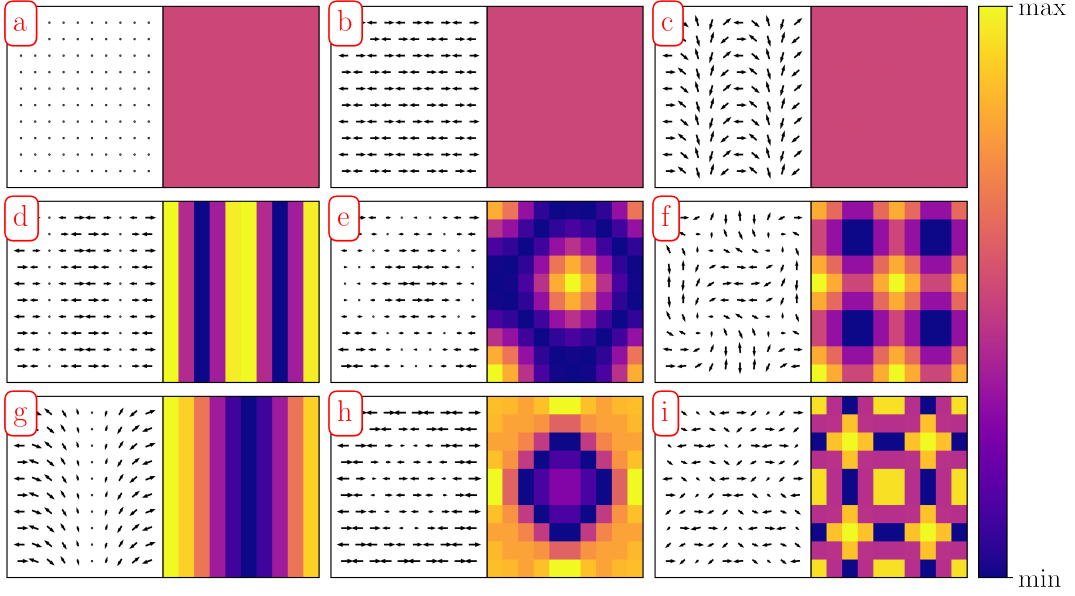


Fig. 1: Distinct orders found in unrestricted mean-field calculations, schematically shown on a 10×10 lattice. In each panel, the square on the left shows the relative spin orientations and amplitudes (length of the arrows) of each phase, choosing a frame such that the spins lie in the xy plane and the bottom left spin points along the x direction. The right plot in each panel shows the corresponding charge modulation, using a color code defined on the right edge of the figure. The various panels exemplify the following magnetic orders: (a) paramagnetism, (b) Néel antiferromagnetism, (c) spiral order, (d) stripe order, (e) collinear bidirectional stripe order, (f) coplanar bidirectional stripe order, (g) beat order, (h) other collinear orders, (i) “strange” order. Only one example of strange order is shown.

vectors of the form $\mathbf{Q}_{xy} \equiv (\pi - 2\pi\eta, \pi - 2\pi\eta)$ as dominant contributions are rarely found, and typically only for relatively small lattices.

In cases with only one mode, one can express $\vec{S}_{\mathbf{q}}$ as

$$\vec{S}_{\mathbf{q}} = \frac{1}{2} \left(\vec{S} \delta_{\mathbf{q}, \mathbf{Q}} + \vec{S}^* \delta_{\mathbf{q}, -\mathbf{Q}} \right), \quad (14)$$

where $\mathbf{Q}$ has the form $\mathbf{Q}_x$ or $\mathbf{Q}_y$. In states with two modes, one has

$$\vec{S}_{\mathbf{q}} = \frac{1}{2} \left(\vec{S}_x \delta_{\mathbf{q}, \mathbf{Q}_x} + \vec{S}_x^* \delta_{\mathbf{q}, -\mathbf{Q}_x} \right) + \frac{1}{2} \left(\vec{S}_y \delta_{\mathbf{q}, \mathbf{Q}_y} + \vec{S}_y^* \delta_{\mathbf{q}, -\mathbf{Q}_y} \right). \quad (15)$$

Depending on the model parameters, temperature, and lattice size, the unrestricted mean-field calculations yield 9 distinct states [39]. The spin and charge order pattern for a representative of each state is sketched in Fig. 1. States that can be transformed into each other by a point group symmetry operation of the lattice (rotations and reflections), by a spatial translation (changing the phases of $\vec{S}_x$ and $\vec{S}_y$), or by a global $SU(2)$ rotation of the spin frame, belong to the same class. In the following, each class is described by one of its representatives. In principle, there might be other possible spin configurations, for example the phases F and G in the classification of Ref. [41], but they never appear in the fairly broad parameter regime explored in Ref. [39].

2.3.1 Paramagnetism

In the paramagnetic phase the expectation value of the spin on each site vanishes, i.e., $\langle \vec{S}_j \rangle = \vec{0}$, and the charge distribution is uniform, $\langle n_j \rangle = n$ (see panel (a) of Fig. 1).

2.3.2 Néel antiferromagnetism

In a Néel antiferromagnet only $\vec{S}_{(\pi,\pi)}$ is nonzero. In real space, adjacent spins point in opposite directions, all spins have the same amplitude, and the charge is homogeneously distributed, as shown in panel (b) of Fig. 1.

2.3.3 Spiral

Planar circular spiral states are briefly referred to as “spiral” states. Such a state is characterized by a single mode with a wave vector $\mathbf{Q}$, and (as a representative)

$$\vec{S} = S_0 \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}, \quad (16)$$

where S_0 is a real number. In real space, the corresponding spin pattern has the form

$$\langle \vec{S}_j \rangle = S_0 \begin{pmatrix} \cos(\mathbf{Q} \cdot \mathbf{r}_j) \\ \sin(\mathbf{Q} \cdot \mathbf{r}_j) \\ 0 \end{pmatrix}. \quad (17)$$

The spin magnitude $|\langle \vec{S}_j \rangle|$ is constant (independent of site j), and the charge distribution is therefore homogeneous, as shown in panel (c) of Fig. 1. The wave vector $\mathbf{Q}$ has the form $\mathbf{Q}_x = (\pi - 2\pi\eta, \pi)$ or $\mathbf{Q}_y = (\pi, \pi - 2\pi\eta)$, that is, only one component deviates from π . The Néel state can be viewed as the special case of a spiral with $\mathbf{Q} = (\pi, \pi)$.

2.3.4 Collinear unidirectional stripes

Stripe states are generally a superposition of several Fourier modes. However, usually one of them is much larger than all others. For a collinear unidirectional stripe state described by a single $\mathbf{q}$ -mode, the amplitude has the form

$$\vec{S} = S_0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \quad (18)$$

In this state the spins are ordered collinearly, and the spin modulation in real space is given by

$$\langle \vec{S}_j \rangle = S_0 \begin{pmatrix} \cos(\mathbf{Q} \cdot \mathbf{r}_j) \\ 0 \\ 0 \end{pmatrix}. \quad (19)$$

Once again, the wave vector $\mathbf{Q}$ has either the form $\mathbf{Q}_x$ or $\mathbf{Q}_y$. The charge order of this state is a unidirectional charge density wave with a wave vector $2\mathbf{Q}$, that is, $\langle n_j \rangle - n \propto \cos(2\mathbf{Q} \cdot \mathbf{r}_j)$. The minima in the amplitude of the spin coincide with the minima of the charge modulation, as shown in panel (d) of Fig. 1.

2.3.5 Collinear bidirectional stripes (CIBS)

A collinear bidirectional stripe consists of two orthogonal stripes with parallel spin orientations, that is, the (dominant) components of $\vec{S}_q$ in Eq. (15) are given by

$$\vec{S}_x = \vec{S}_y = S_0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad (20)$$

where S_0 is a real constant. In real space, this corresponds to

$$\langle \vec{S}_j \rangle = S_0 \begin{pmatrix} \cos(\mathbf{Q}_x \cdot \mathbf{r}_j) + \cos(\mathbf{Q}_y \cdot \mathbf{r}_j) \\ 0 \\ 0 \end{pmatrix}. \quad (21)$$

This phase has charge modulations in both x and y directions, proportional to $[\cos(\mathbf{Q}_x \cdot \mathbf{r}_j) + \cos(\mathbf{Q}_y \cdot \mathbf{r}_j)]^2$, as displayed in panel (e) of Fig. 1.

2.3.6 Coplanar bidirectional stripes (CpBS)

A coplanar bidirectional stripe state consists of two orthogonal stripes with orthogonal spin orientations, that is, with two $\mathbf{q}$ -modes of the form

$$\vec{S}_x = S_0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{S}_y = S_0 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad (22)$$

with S_0 a real constant. In real space this corresponds to

$$\langle \vec{S}_j \rangle = S_0 \begin{pmatrix} \cos(\mathbf{Q}_x \cdot \mathbf{r}_j) \\ \cos(\mathbf{Q}_y \cdot \mathbf{r}_j) \\ 0 \end{pmatrix}. \quad (23)$$

This phase also has charge modulations in both directions, proportional to $[\cos(\mathbf{Q}_x \cdot \mathbf{r}_j)]^2 + [\cos(\mathbf{Q}_y \cdot \mathbf{r}_j)]^2$, as shown in panel (f) of Fig. 1.

2.3.7 Beat states

Sometimes one finds states whose dominant $\mathbf{q}$ -components take the form (up to a lattice rotation) $\mathbf{Q}_x = (\pi - 2\pi\eta, \pi)$ and $\mathbf{Q}'_x = (\pi - 2\pi\eta', \pi)$ with $\eta \neq \eta'$. As shown in panel (g) of Fig. 1, in

these states there are unidirectional charge modulations and the spin patterns usually show beat-like modulations, since the spin components are sums of cosines with different wave numbers and phases. Such states are finite size artifacts. They arise when the “optimal” incommensurability η_{opt} lies in between η and η' , which is however not allowed by the finite size of the system (cf. Eq. (13)). In the thermodynamic limit, where any real value of η is permitted, these states converge to a pure stripe or spiral phase.

2.3.8 Other collinear orders

In some cases all the spins are ordered collinearly, but not in the shape of Néel order, stripe order or collinear bidirectional stripes. As noted already by Zaanen and Gunnarsson [18], these states will likely be replaced by stripe states in the thermodynamic limit. For example, the lattice sizes treated in our calculations are too small to resolve stripes with wave vectors $\mathbf{Q}$ close to (π, π) , so that the system prefers to order in a more complex manner. Such a state, sketched in panel (h) of Fig. 1, can be viewed as a kind of stripe order where the lines of constant density, instead of lying parallel to the x (or y) direction, form a ring. In these cases, one finds several nonzero components of $\vec{S}_{\mathbf{q}}$.

2.3.9 Strange orders

In some cases one cannot classify a state by any of the previously described orders. Such states do not exhibit a simple structure in momentum space and often show complex charge patterns. These states will be referred to as “strange order”. In the vast majority of cases, their spin order is coplanar. An example is sketched in panel (i) of Fig. 1. The strange states are also expected to disappear in the thermodynamic limit [39].

2.4 Results from unrestricted real space calculations

As a typical result from an unrestricted Hartree-Fock calculation on a finite lattice, a (n, T) phase diagram obtained for $t' = -0.15t$ and $U = 3t$ on a 20×20 lattice [39] is shown in Fig. 2. The various states classified according to the above scheme are indicated by distinct colors. The black transition lines have been obtained from calculations in the thermodynamic limit, which will be described in the next section.

The transition temperature T^* obtained in the thermodynamic limit (solid line) is very close to the transition temperature found for the finite system. For densities at and close to half-filling, the system is Néel ordered. With few exceptions, Néel order is also obtained in the entire electron-doped region (for $n > 1$). Not only T^* , but also the boundary of the Néel ordered region inside the magnetic regime obtained from the calculation in the thermodynamic limit (dashed line) agrees very well with the real-space finite size calculation.

On the hole-doped side, the calculation in the thermodynamic limit (described in the next section) yields a regime of stable spiral order with a continuously varying incommensurability η between the dashed and the dotted black lines in Fig. 2. In the real space calculation on the

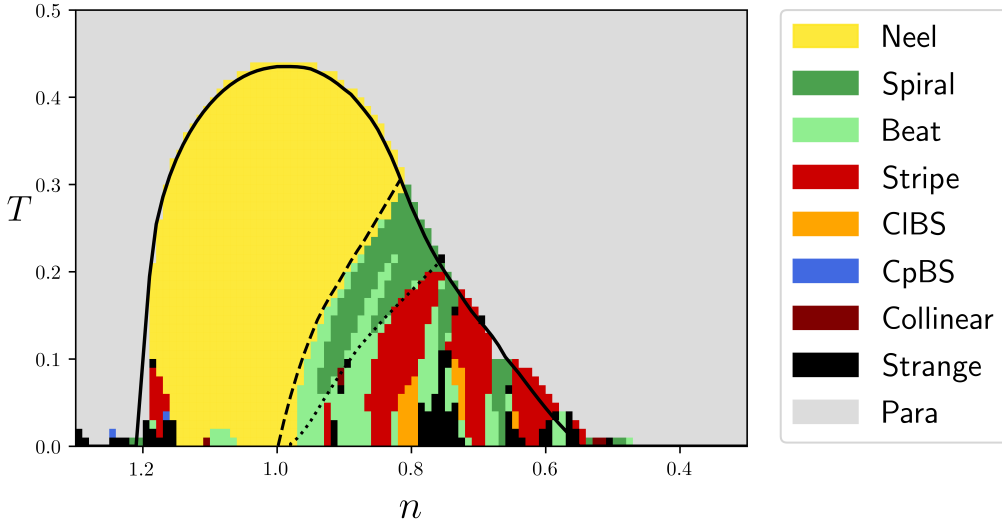


Fig. 2: Phase diagram for $U = 3t$ and $t' = -0.15t$. The colors label the states resulting from the real space calculation on a 20×20 lattice. The black lines were obtained from calculations in momentum space in the thermodynamic limit. The solid black line indicates the transition temperature T^* separating the paramagnetic from the magnetically ordered regime, the dashed black line the transition between Néel and spiral order, and the dotted black line the divergence of the charge susceptibility in the spiral state. Figure taken from Ref. [39].

20×20 lattice, the spiral regime is split into spiral and beat order regions, where the incommensurabilities in the spiral regions are either $\eta = 0.05$ (close to the Néel region) or $\eta = 0.1$. At densities and temperatures where one finds beat states it is energetically more costly for the system to order as a spiral state with the discrete incommensurabilities 0.05 or 0.1 allowed on the 20×20 lattice. Since intermediate ordering vectors are prohibited due to the periodic boundary conditions, the system orders in a beat ordered pattern. Looking at the wavevectors of the Fourier modes $\vec{S}_q$ in these states, one finds $\mathbf{Q}' = (\pi - 2\pi\frac{1}{20}, \pi)$ and $\mathbf{Q}'' = (\pi - 2\pi\frac{1}{10}, \pi)$. Solving the mean-field equations in the beat region between the spiral states on a 40×40 lattice one finds that beat order is indeed replaced by spiral order with an intermediate wavevector $\mathbf{Q} = (\pi - 2\pi\frac{3}{40}, \pi)$. Hence, the real space calculations are consistent with a spiral phase with a smoothly varying η .

Below the dotted line in Fig. 2, and for any temperature at densities below $n \approx 0.75$, spiral states exhibit negative (for some wave vectors) spin and charge susceptibilities, and are thus unstable. In this region stripes, other collinear states, and strange states appear. The stripe region is split into several domains with distinct incommensurabilities: $\eta = 0.05$ for densities $n \approx 0.93$, $\eta = 0.1$ for densities close to $n \approx 0.82$, $\eta = 0.15$ for densities around 0.7, and $\eta = 0.2$ near $n \approx 0.62$. These stripe regimes are separated by intermediate regions of beat ordered and strange phases, where the predominant wavevectors are the ones of the two adjacent stripe regimes. Repeating the calculation on a larger 40×40 lattice for a few (n, T) pixels in this intermediate regime, one finds stripes with intermediate values of η . This confirms the hypothesis that beat and strange states are just artifacts of the finite system size.

A particular point in parameter space has received special attention in recent years: the ground state ($T=0$) at a density $n = \frac{7}{8}$ for the two-dimensional Hubbard model with pure nearest

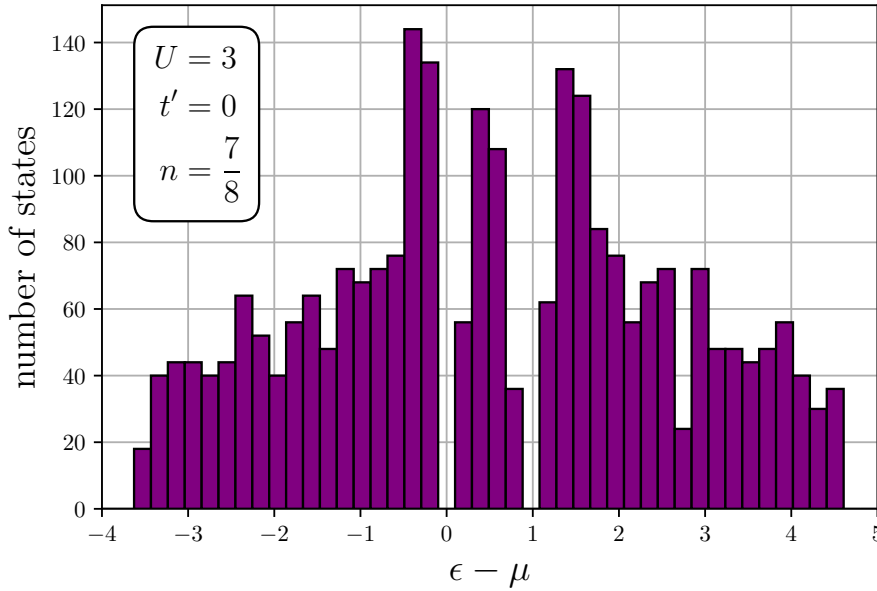


Fig. 3: Density of states histogram for $n = \frac{7}{8}$, $T = 0$, $t' = 0$ and $U = 3t$, calculated on a 32×40 lattice. The states are counted in equidistant bins in the energy range between -4 and 5 . Figure taken from Ref. [39].

neighbor hopping. Exact numerical ground state techniques have provided fairly convincing evidence that for strong coupling the ground state is an insulating stripe state in this case [27–29]. Hence, it is interesting to investigate this special point in more detail, keeping however a moderate interaction strength $U = 3t$, in line with the Hartree-Fock approximation.

The solution of the real-space mean-field equations unambiguously yields unidirectional stripes with an incommensurability $\eta = \frac{1}{16}$ in this case. To see whether this stripe ground state is an insulator, one may look at the distribution of energy levels, which is a finite size proxy to the density of states. In Fig. 3, a histogram of the distribution of the energy levels ε_ℓ is shown, shifted by the chemical potential (placed half-way between the highest occupied and the lowest unoccupied state). One can see that there are no energy levels close to the chemical potential – there is a sizable band gap ($\Delta \approx 0.25t$). This stripe state is therefore indeed an insulator. There are other special choices of parameters leading to an insulating ground state, but metallic behavior is much more common.

3 Mean-field solutions in the thermodynamic limit

For several specific types of magnetic order, the Hartree-Fock equations can be solved directly in the thermodynamic limit. Among these, spiral states are particularly easy to treat, for any ordering wave vector $\mathbf{Q}$. The Néel state can be viewed as a special case of spiral order. More generally, any magnetic order can be treated in the thermodynamic limit if it is commensurate with the lattice, that is, if some (reduced) translation invariance remains in the ordered state. Finally, for temperatures slightly below the critical temperature T^* , an expansion in powers of the order parameter can help to discriminate distinct magnetic states.

3.1 Néel and spiral states

For circular spiral states, the mean-field equations assume a particularly simple form in momentum space, for arbitrary (also incommensurate) ordering wave vectors $\mathbf{Q}$ [4–13]. This is due to a residual translation symmetry of the ordered state, consisting of a lattice translation combined with a rotation in spin space. In the special case $\mathbf{Q} = (\pi, \pi)$, spiral order is just Néel order. Choosing the spin order to lie in the xy plane, the expectation value $\langle \vec{S}_j \rangle$ has the form Eq. (17). A Fourier transformation reveals that this type of order is associated with anomalous expectation values $\langle c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}+\mathbf{Q}\downarrow} \rangle$ in momentum space, and the mean-field interaction term (3) simplifies to

$$H_{\text{int}}^{\text{MF}} = - \sum_{\mathbf{k}} (\Delta^* c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}+\mathbf{Q}\downarrow} + \Delta c_{\mathbf{k}+\mathbf{Q}\downarrow}^\dagger c_{\mathbf{k}\uparrow}) + L|\Delta|^2/U, \quad (24)$$

with the self-consistency condition $\Delta = UL^{-1} \sum_{\mathbf{k}} \langle c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}+\mathbf{Q}\downarrow} \rangle$. Defining the two-component spinor $\Psi_{\mathbf{k}} = (c_{\mathbf{k}\uparrow}, c_{\mathbf{k}+\mathbf{Q}\downarrow})$, the mean-field Hamiltonian can be written in the concise form

$$H^{\text{MF}} = \sum_{\mathbf{k}} \Psi_{\mathbf{k}}^\dagger \begin{pmatrix} \varepsilon_{\mathbf{k}} & -\Delta^* \\ -\Delta & \varepsilon_{\mathbf{k}+\mathbf{Q}} \end{pmatrix} \Psi_{\mathbf{k}} + L|\Delta|^2/U. \quad (25)$$

Its diagonalization yields two quasi-particle bands with dispersion relations

$$E_{\mathbf{k}}^\pm = \varepsilon_{\mathbf{k}}^\pm \pm \sqrt{(\varepsilon_{\mathbf{k}}^-)^2 + |\Delta|^2}, \quad (26)$$

where $\varepsilon_{\mathbf{k}}^\pm = \frac{1}{2}(\varepsilon_{\mathbf{k}} \pm \varepsilon_{\mathbf{k}+\mathbf{Q}})$. The free energy is then obtained in the form

$$F = -T \sum_{\mathbf{k}} \sum_{\ell=\pm} \ln \left(1 + e^{-\beta(E_{\mathbf{k}}^\ell - \mu)} \right) + L|\Delta|^2/U + \mu N. \quad (27)$$

The magnetic gap Δ and the ordering wave vector $\mathbf{Q}$ are obtained by minimizing F . Minimization with respect to Δ yields the gap equation

$$\Delta = \frac{U\Delta}{L} \sum_{\mathbf{k}} \frac{f(E_{\mathbf{k}}^- - \mu) - f(E_{\mathbf{k}}^+ - \mu)}{2\sqrt{(\varepsilon_{\mathbf{k}}^-)^2 + |\Delta|^2}}, \quad (28)$$

where f is the Fermi distribution. The free energy does not depend on the phase of Δ , which can therefore be chosen arbitrarily. The optimal wave vector has usually the “axial” form $\mathbf{Q} = (Q, \pi)$, degenerate with (π, Q) , or sometimes the “diagonal” form $\mathbf{Q} = (Q, \pm Q)$.

To probe the stability of the spiral state obtained from solving the mean-field equations with a spiral ansatz, one can check whether the static (zero frequency) spin and charge susceptibilities in the spiral state are positive for all wave vectors $\mathbf{q}$. The combined spin and charge susceptibility is a 4×4 matrix with matrix elements χ^{ab} , where the indices $a, b = 0, 1, 2, 3$ label charge (0) and spin (1, 2, 3) components [10, 43]. In a rotated spin frame, where the spiral state looks translation invariant, its Fourier transform $\tilde{\chi}^{ab}(\mathbf{q}, \omega)$ depends only on a single wave vector. The conserving approximation [44] for susceptibilities corresponding to mean-field theory for the free energy is given by the so-called random phase approximation (RPA). In the rotated spin frame, the RPA susceptibility can be written as [10, 43],

$$\tilde{\chi}(\mathbf{q}, \omega) = \tilde{\chi}_0(\mathbf{q}, \omega) [\mathbf{1}_4 - \Gamma_0 \tilde{\chi}_0(\mathbf{q}, \omega)]^{-1}, \quad (29)$$

where $\Gamma_0 = 2U \text{diag}(-1, 1, 1, 1)$, and the bare susceptibility $\tilde{\chi}_0(\mathbf{q}, \omega)$, is given by

$$\tilde{\chi}_0^{ab}(\mathbf{q}, \omega) = -\frac{T}{4L} \sum_{\mathbf{k}} \sum_{k_0} \text{Tr} [\sigma^a G(\mathbf{k}+\mathbf{q}, ik_0+iq_0) \sigma^b G(\mathbf{k}, ik_0)] \Big|_{iq_0 \rightarrow \omega+i0^+}, \quad (30)$$

where $G(\mathbf{k}, ik_0)$ is the mean-field single-particle Green function in the spinor basis defined by $\Psi_{\mathbf{k}}$, the sum over k_0 is a sum over fermionic Matsubara frequencies, and σ^0 is the two-dimensional unit matrix, while $\sigma^1, \sigma^2, \sigma^3$ are the Pauli matrices.

3.2 Commensurate coplanar states

For *commensurate* order, the system retains some (reduced) translation invariance. Defining a suitable unit cell in the ordered state, the mean-field equations can then be solved in the thermodynamic limit. With very few exceptions, which are probably finite size artifacts, the spin order found in the unrestricted real space calculations is *coplanar*, that is, all the spins align in a common plane. Except for very large hole-doping in the presence of a large $|t'/t|$, the order is also *unidirectional*, that is, at most one (for Néel order none) of the two components of the contributing wave vectors is $\neq \pi$, such that spins order antiferromagnetically in one of the directions x or y , and the electron density is homogeneous in this direction. In the present context it is therefore worthwhile to formulate and evaluate the mean-field equations for commensurate, coplanar, and unidirectional order in a momentum space representation [40]. Unidirectional collinear stripe and circular spiral order are special cases thereof.

Due to the SU(2) symmetry of the Hubbard model, the coplanar order can lie in any (fixed) plane in spin space. We select the xy plane as a convenient choice, and we choose the y direction in real space as the direction of simple antiferromagnetic order. In x direction, we assume that the ordered state is periodic with a periodicity P . The magnetization and charge density profiles can then be expanded in Fourier modes as follows

$$\vec{S}_j = \sum_{n \text{ odd}} (M_n^x \vec{e}_x + M_n^y \vec{e}_y) e^{i\mathbf{Q}_n \cdot \mathbf{r}_j}, \quad (31a)$$

$$\rho_j = \sum_{n \text{ even}} \varrho_n e^{i\mathbf{Q}_n \cdot \mathbf{r}_j}, \quad (31b)$$

where $\mathbf{Q}_n = (2\pi n/P, \pi n)$ with $n = 0, \dots, \bar{P}-1$, and $\vec{e}_x$ and $\vec{e}_y$ are unit vectors in the x and y direction, respectively. The first sum is running only over odd integers n because the spin order is antiferromagnetic in y -direction, while the second sum is restricted to even integers since ρ_j is translation invariant in y -direction. We define $\bar{P}$ as the smallest positive integer satisfying $\mathbf{Q}_{\bar{P}} = (0, 0)$ modulo reciprocal lattice vectors. If P is even, one has $\bar{P} = P$, and $\bar{P} = 2P$ if P is odd. The summations in Eq. (31) run over a finite number ($\bar{P}/2$) of terms. Since the spin and charge densities are real, the coefficients M_n^x, M_n^y , and ϱ_n must obey

$$M_n^x = (M_{-n}^x)^* = (M_{\bar{P}-n}^x)^*, \quad (32a)$$

$$M_n^y = (M_{-n}^y)^* = (M_{\bar{P}-n}^y)^*, \quad (32b)$$

$$\varrho_n = (\varrho_{-n})^* = (\varrho_{\bar{P}-n})^*. \quad (32c)$$

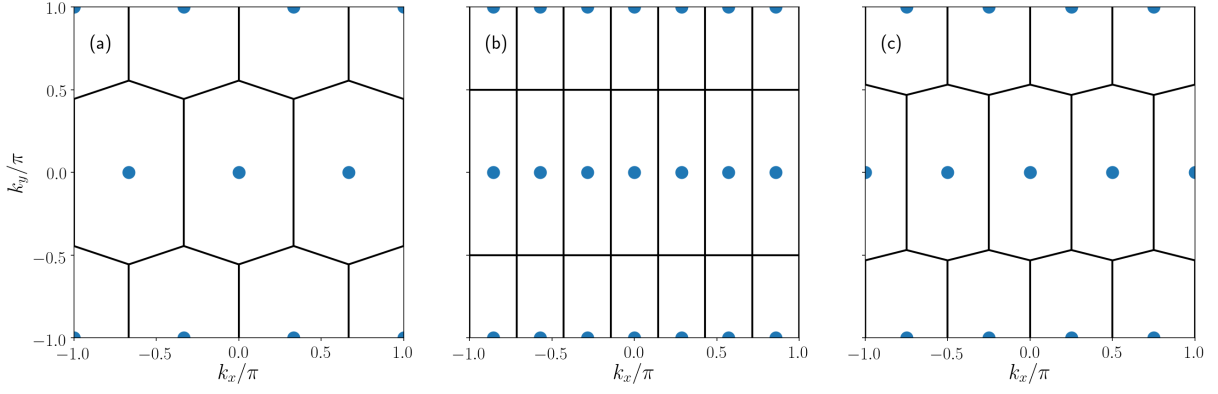


Fig. 4: Construction of the reduced Brillouin zone for $P = 6$ (a), $P = 7$ (b), and $P = 8$ (c). The blue dots represent the momenta $\mathbf{Q}_n$ with $n = 0, \dots, \bar{P}-1$ modulo vectors of the original reciprocal lattice. The black lines separate distinct reduced Brillouin zones. The original Brillouin zone is divided into $\bar{P}$ equivalent reduced Brillouin zones. Figure taken from [40].

The translational symmetry breaking of the original lattice leads to an enlarged unit cell containing $\bar{P}$ inequivalent sites with distinct expectation values. In momentum space, the size of the original Brillouin zone is reduced by a factor $1/\bar{P}$. The new reciprocal lattice can be constructed by adding all vectors of the form $\mathbf{Q}_n$ with $n = 0, \dots, \bar{P}-1$ to the original reciprocal lattice vectors (see Fig. 4). The reduced Brillouin zone can then be defined as the set of all points which are closer to a given vector of the new reciprocal lattice than to any other (Wigner-Seitz construction). The reduced Brillouin zones for $P = 6, 7, 8$ are plotted in Fig. 4.

The mean-field interaction term Eq. (3) can be written in the following manifestly SU(2) symmetric form [40]

$$H_{\text{int}}^{\text{MF}} = \sum_j \sum_{a=0}^3 s_a \left[\Delta_j^a c_j^\dagger \sigma^a c_j - (\Delta_j^a)^2 / U \right], \quad (33)$$

where c_j is the two-component spinor composed of $c_{j\uparrow}$ and $c_{j\downarrow}$. The sign s_a is plus one in the charge channel ($a = 0$) and minus one in the spin channel ($a = 1, 2, 3$). The gap parameters Δ_j^a are defined by charge and spin expectation values as $\Delta_j^a = \frac{1}{2}U \langle c_j^\dagger \sigma^a c_j \rangle$. They are related to the previously defined gap parameters via $\Delta_j^0 = \frac{1}{2}(\Delta_{j+} + \Delta_{j-})$, $\Delta_j^1 = \frac{1}{2}(\Delta_{j+} - \Delta_{j-})$, $\Delta_j^2 = \frac{1}{2i}(\Delta_{j+} - \Delta_{j-})$, and $\Delta_j^3 = \frac{1}{2}(\Delta_{j\uparrow} - \Delta_{j\downarrow})$. Inserting Eq. (31) into Eq. (33), adding the kinetic term, and Fourier transforming, we obtain

$$H_{\text{MF}} = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + \sum_{\mathbf{k}} \sum_{a=0}^3 \sum_{n=0}^{\bar{P}-1} s_a \Delta_n^a c_{\mathbf{k}}^\dagger \sigma^a c_{\mathbf{k}+\mathbf{Q}_n} - L \sum_{a=0}^3 \sum_{n=0}^{\bar{P}-1} s_a \frac{|\Delta_n^a|^2}{U}, \quad (34)$$

with

$$\Delta_n^0 = \begin{cases} \frac{1}{2}U \varrho_n & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd} \end{cases}, \quad \Delta_n^{1,2} = \begin{cases} 0 & \text{for } n \text{ even} \\ U M_n^{x,y} & \text{for } n \text{ odd} \end{cases}. \quad (35)$$

Defining the $\bar{P}$ -component spinor

$$\Psi_{\mathbf{k},\sigma} = (c_{\mathbf{k},\sigma}, c_{\mathbf{k}+\mathbf{Q}_1,\bar{\sigma}}, c_{\mathbf{k}+\mathbf{Q}_2,\sigma}, \dots, c_{\mathbf{k}+\mathbf{Q}_{\bar{P}-2},\sigma}, c_{\mathbf{k}+\mathbf{Q}_{\bar{P}-1},\bar{\sigma}}) \quad (36)$$

with the convention $\bar{\uparrow} = \downarrow$ and $\bar{\downarrow} = \uparrow$, one can cast the Hamiltonian (34) in the form

$$H_{\text{MF}} = \sum_{\mathbf{k} \in \text{BZ}'} \sum_{\sigma} \Psi_{\mathbf{k},\sigma}^{\dagger} \mathcal{H}_{\mathbf{k},\sigma}^{(P)} \Psi_{\mathbf{k},\sigma}, \quad (37)$$

where the constant term in Eq. (34) has been dropped, and the momentum sum is restricted to the reduced Brillouin zone BZ' . The matrix $\mathcal{H}_{\mathbf{k},\sigma}^{(P)}$ is composed of the elements

$$[\mathcal{H}_{\mathbf{k},\sigma}^{(P)}]_{\ell\ell'} = \begin{cases} \varepsilon_{\mathbf{k}+\mathbf{Q}_{\ell}} & \text{if } \ell = \ell' \\ \Delta_{\sigma,n}^{p(\ell)} & \text{if } \ell' = (\ell+n) \bmod \bar{P}, n \text{ odd} \\ \Delta_n^0 & \text{if } \ell' = (\ell+n) \bmod \bar{P}, n \text{ even} \end{cases}, \quad (38)$$

where $p(\ell) = +$ if ℓ is even and $p(\ell) = -$ if ℓ is odd, and

$$\begin{aligned} \Delta_{\uparrow,n}^{\pm} &= -(\Delta_n^1 \pm i\Delta_n^2), \\ \Delta_{\downarrow,n}^{\pm} &= -(\Delta_n^1 \mp i\Delta_n^2). \end{aligned} \quad (39)$$

Since Δ_0^0 only shifts the chemical potential μ , one may redefine μ as $\mu - \Delta_0^0$ and set $\Delta_0^0 = 0$. For example, for $P = 3$ and $P = 6$, the matrix $\mathcal{H}_{\mathbf{k},\sigma}^{(P)}$ has the form

$$\mathcal{H}_{\mathbf{k},\sigma}^{(6) \text{ or } (3)} = \begin{pmatrix} \varepsilon_{\mathbf{k},0} & \Delta_{\sigma,1}^+ & \Delta_2^0 & \Delta_{\sigma,3}^+ & \Delta_4^0 & \Delta_{\sigma,5}^+ \\ \Delta_{\sigma,5}^- & \varepsilon_{\mathbf{k},1} & \Delta_{\sigma,1}^- & \Delta_2^0 & \Delta_{\sigma,3}^- & \Delta_4^0 \\ \Delta_4^0 & \Delta_{\sigma,5}^+ & \varepsilon_{\mathbf{k},2} & \Delta_{\sigma,1}^+ & \Delta_2^0 & \Delta_{\sigma,3}^+ \\ \Delta_{\sigma,3}^- & \Delta_4^0 & \Delta_{\sigma,5}^- & \varepsilon_{\mathbf{k},3} & \Delta_{\sigma,1}^- & \Delta_2^0 \\ \Delta_2^0 & \Delta_{\sigma,3}^+ & \Delta_4^0 & \Delta_{\sigma,5}^+ & \varepsilon_{\mathbf{k},4} & \Delta_{\sigma,1}^+ \\ \Delta_{\sigma,1}^- & \Delta_2^0 & \Delta_{\sigma,3}^- & \Delta_4^0 & \Delta_{\sigma,5}^- & \varepsilon_{\mathbf{k},5} \end{pmatrix}. \quad (40)$$

The parameters Δ_n^a are self-consistently determined as

$$\Delta_n^a = \frac{U}{2L} \sum_{\mathbf{k}} \langle c_{\mathbf{k}}^{\dagger} \sigma^a c_{\mathbf{k}+\mathbf{Q}_n} \rangle = \frac{U}{2L} \sum_{\mathbf{k} \in \text{BZ}'} \sum_{\sigma} \langle \Psi_{\mathbf{k},\sigma}^{\dagger} \Gamma_{n,\sigma}^a \Psi_{\mathbf{k},\sigma} \rangle, \quad (41)$$

where the matrices $\Gamma_{\sigma}^{a,n}$ have been defined as

$$[\Gamma_{n,\sigma}^0]_{\ell\ell'} = \begin{cases} 1 & \text{if } \ell' = (\ell+n) \bmod \bar{P}, n \text{ even} \\ 0 & \text{otherwise} \end{cases} \quad (42a)$$

$$[\Gamma_{n,\sigma}^1]_{\ell\ell'} = \begin{cases} 1 & \text{if } \ell' = (\ell+n) \bmod \bar{P}, n \text{ odd} \\ 0 & \text{otherwise} \end{cases} \quad (42b)$$

$$[\Gamma_{n,\sigma}^2]_{\ell\ell'} = \begin{cases} i(-1)^{\sigma+p(\ell)} & \text{if } \ell' = (\ell+n) \bmod \bar{P}, n \text{ odd} \\ 0 & \text{otherwise} \end{cases} \quad (42c)$$

with $(-1)^{\uparrow} = +1$ and $(-1)^{\downarrow} = -1$.

Since $\mathcal{H}_{\mathbf{k},\uparrow}^{(P)}$ and $\mathcal{H}_{\mathbf{k},\downarrow}^{(P)}$ are related to each other by an inversion of the sign of Δ_n^2 , the expectation values on the right hand side of Eq. (41) take the same value for each of the two spin projections. Hence, one can simplify Eq. (41) to

$$\Delta_n^a = UL^{-1} \sum_{\mathbf{k} \in \text{BZ}'} \langle \Psi_{\mathbf{k}}^\dagger \Gamma_n^a \Psi_{\mathbf{k}} \rangle, \quad (43)$$

where $\Psi_{\mathbf{k}} = \Psi_{\mathbf{k},\uparrow}$ and $\Gamma_n^a = \Gamma_{n,\uparrow}^a$. In other words, one can solve the mean-field equations using only the matrix $\mathcal{H}_{\mathbf{k},\uparrow}^{(P)}$. The right hand side of Eq. (43) is computed from Eq. (38) making an initial random assumption on the mean-field parameters Δ_n^a , which are then updated using again Eq. (43). The procedure is repeated until convergence is reached.

To find the energetically best state, one needs to solve Eq. (43) for different values of P and retain the state with the lowest mean-field free-energy. In Ref. [40], the original Brillouin zone BZ was discretized with N_k^2 equally spaced points and, for a fixed N_k , only divisors of N_k were admitted as possible values for P . For every fixed set of parameters, over 90 integer values of P ranging from 2 to 220 were offered, each of them with a suitably adjusted N_k such that N_k/P is integer.

It is instructive to see how Néel, spiral, and collinear stripe states are captured as special cases of the above general formalism.

3.2.1 Néel order

In the case of Néel order, one has

$$\vec{S}_j = M(-1)^j (\cos \varphi \vec{e}_x + \sin \varphi \vec{e}_y), \quad (44a)$$

$$\rho_j = \text{const.}, \quad (44b)$$

where φ parametrizes the orientation of the spin order in the xy plane, and M represents its amplitude. Néel order has the period $P = 2$, so that the matrix $\mathcal{H}_{\mathbf{k},\sigma}^{(P)}$ in Eq. (38) is two-dimensional and only the two parameters Δ_1^1 and Δ_1^2 contribute, where

$$\Delta_1^1 = \Delta \cos \varphi, \quad \Delta_1^2 = \Delta \sin \varphi, \quad (45)$$

with $\Delta = UM$.

3.2.2 Circular spiral order

Circular spiral order in the xy plane has the general form

$$\vec{S}_j = M(\cos(\mathbf{Q} \cdot \mathbf{r}_j + \varphi) \vec{e}_x \pm \sin(\mathbf{Q} \cdot \mathbf{r}_j + \varphi) \vec{e}_y), \quad (46a)$$

$$\rho_j = \text{const.}, \quad (46b)$$

where, as in the case of Néel order, φ parametrizes the orientation of the spin order in the xy plane and M its (constant) amplitude. $\mathbf{Q}$ is a generic wave vector of the form $(\pi - 2\pi\eta, \pi)$ with $\eta > 0$. The “+” or “−” sign distinguishes between spirals rotating anti-clockwise and clockwise.

This type of order emerges as a special case of our general formalism if Δ_n^a is non-zero only for one mode n_0 and its conjugate $\bar{P}-n_0$, with n_0 odd and such that $\mathbf{Q}$ can be approximated by $\mathbf{Q}_{n_0}$, with a suitably chosen P . Spiral order as in Eq. (46) is then described by

$$\Delta_{n_0}^1 = \frac{1}{2}\Delta e^{i\varphi}, \quad \Delta_{\bar{P}-n_0}^1 = \frac{1}{2}\Delta e^{-i\varphi}, \quad (47a)$$

$$\Delta_{n_0}^2 = \mp \frac{i}{2}\Delta e^{i\varphi}, \quad \Delta_{\bar{P}-n_0}^2 = \pm \frac{i}{2}\Delta e^{-i\varphi}, \quad (47b)$$

with $\Delta = UM$. All other Δ_n^a are zero. The matrix $\mathcal{H}_{\mathbf{k},\sigma}^{(P)}$ thus simplifies to a diagonal block matrix form with $\bar{P}/2$ matrices of size two on the diagonal. Indeed spiral order can be described by a simpler 2×2 mean-field Hamiltonian for each $\mathbf{k}$ -point, as described above. Eqs. (47) apply only if $n_0 \neq \bar{P}-n_0$, which is fulfilled for any spiral state which is not a Néel state.

3.2.3 Stripe order

Defining stripe order as any type of collinear order that differs from Néel antiferromagnetism, the magnetization and charge densities have the form

$$\vec{S}_j = f_S(\mathbf{r}_j) (\cos \varphi \vec{e}_x + \sin \varphi \vec{e}_y), \quad (48a)$$

$$\rho_j = f_\rho(\mathbf{r}_j), \quad (48b)$$

where once again φ is an angle parameterizing the orientation of the spins in the xy plane, while $f_S(\mathbf{r}_j)$ and $f_\rho(\mathbf{r}_j)$ are two functions defining the spatial modulation of the magnetization amplitude and charge density, respectively. They can be expressed in terms of their Fourier coefficients as

$$f_S(\mathbf{r}_j) = \sum_{n \text{ odd}} M_n e^{i\mathbf{Q}_n \cdot \mathbf{r}_j}, \quad (49a)$$

$$f_\rho(\mathbf{r}_j) = \sum_{n \text{ even}} \rho_n e^{i\mathbf{Q}_n \cdot \mathbf{r}_j}. \quad (49b)$$

Stripe order can be obtained as a particular case of the general formalism, with Δ_n^a fulfilling

$$\Delta_n^0 = U\rho_n, \quad \Delta_n^1 = UM_n \cos \varphi, \quad \Delta_n^2 = UM_n \sin \varphi. \quad (50)$$

Depending on the coefficients M_n and ρ_n , the spin and charge profiles can be sinusoidal or sharp (like domain walls) or anything in between.

3.3 Landau theory

Slightly below the critical temperature T^* , the type of magnetic order can be determined from the Landau theory of the phase transition, which is valid for small order parameters. The Landau theory can be derived from the microscopic (mean-field) theory by decoupling the Hubbard interaction with spin and charge order parameter fields via a Hubbard-Stratonovich transformation, and subsequently expanding the resulting effective action in powers of the order parameters. In this section we need to use the path-integral formalism for quantum many-body systems, as described, for example, in the textbook by Negele and Orland [45].

3.3.1 Derivation of the effective action

The Hubbard interaction can be written in the SU(2) symmetric form [46–48]

$$U n_{j\uparrow} n_{j\downarrow} = \frac{U}{4} (c_j^\dagger c_j)^2 - \frac{U}{4} (c_j^\dagger \vec{\sigma} \cdot \vec{n}_j c_j)^2, \quad (51)$$

where $\vec{n}_j$ is an arbitrary unit vector. In the path integral formalism, the operators c_j and $c_j^\dagger$ are replaced by time dependent anticommuting (Grassmann) fields $c_j(\tau)$ and $c_j^*(\tau)$, respectively. Because $\vec{n}_j$ can be arbitrarily chosen, we take the average over all possible $\vec{n}_j$ with a properly defined measure such that $\int \mathcal{D}\vec{n} = 1$.

Performing a Hubbard-Stratonovich transformation [45], the interaction terms in Eq. (51) can be decoupled by means of two fields, $\rho_j(\tau)$ and $\rho_j^S(\tau)$, representing space and (imaginary) time dependent fluctuations of the charge and spin amplitude, respectively. Defining a spin field as $\vec{S}_j(\tau) = \rho_j^S(\tau) \vec{n}_j(\tau)$, the Hubbard interaction can be represented as

$$e^{-U \int_0^\beta d\tau \sum_j n_{j\uparrow} n_{j\downarrow}} = \int \mathcal{D}\rho \mathcal{D}\vec{S} e^{-(\mathcal{S}_\rho + \mathcal{S}_S + \mathcal{S}_{\text{int}})}, \quad (52)$$

where $\beta = 1/T$ is the inverse temperature, and

$$\mathcal{S}_\rho = \frac{1}{U} \int_0^\beta d\tau \sum_j \rho_j^2, \quad (53a)$$

$$\mathcal{S}_S = \frac{1}{U} \int_0^\beta d\tau \sum_j |\vec{S}_j|^2, \quad (53b)$$

$$\mathcal{S}_{\text{int}} = \int_0^\beta d\tau \sum_j c_j^* (i\rho_j + \vec{S}_j \cdot \vec{\sigma}) c_j. \quad (53c)$$

To keep the notation light, the time dependence of the bosonic (ρ_j , $\vec{S}_j$) and fermionic (c_j , c_j^*) fields is not written.

An effective action for the bosonic fields ρ_j and $\vec{S}_j$ can be derived by integrating out the fermions. Except for a field-independent term, one obtains

$$\mathcal{S}^{\text{eff}}[\rho, \vec{S}] = \frac{1}{U} \int_0^\beta d\tau \sum_j (\rho_j^2 + |\vec{S}_j|^2) - \text{tr} \ln \left[1_{\text{st}} - G_0 \cdot (i\rho + \vec{\sigma} \cdot \vec{S}) \right], \quad (54)$$

where G_0 is the Fourier transform to real space and imaginary time of the bare Matsubara Green function $G_0(k) = (i\nu + \mu - \varepsilon_{\mathbf{k}})^{-1}$, while ρ and $\vec{S}$ are diagonal matrices in space and time defined as $\rho_{jj'}(\tau, \tau') = \rho_j(\tau) \delta_{jj'} \delta(\tau - \tau')$ and $\vec{S}_{jj'}(\tau, \tau') = \vec{S}_j(\tau) \delta_{jj'} \delta(\tau - \tau')$. The trace is summing over space and time indices, $1_{\text{st}} = \delta_{jj'} \delta(\tau - \tau')$ is the space-time unit matrix, and $G_0 A$ is the space-time matrix product $\sum_{j''} \int_0^\beta d\tau'' G_{0,jj''}(\tau - \tau'') A_{j''j'}(\tau'', \tau')$.

3.3.2 Taylor expansion of the effective action

We now expand the logarithm in Eq. (54) in powers of ρ_j and $\vec{S}_j$. Such an expansion is justified in the vicinity of the critical temperature T^* , where the magnetic and charge order parameters

are small. To this end we write

$$\text{tr} \ln [1_{\text{st}} - G_0 A] = - \sum_{n=1}^{\infty} \frac{1}{n} \text{tr} [(G_0 A)^n], \quad (55)$$

with $A = i\rho + \vec{\sigma} \cdot \vec{S}$. The trace does not depend on the representation. In the following, all calculations are performed in momentum and frequency space.

The quadratic term in $\vec{S}_q$, the (spatio-temporal) Fourier transform of $\vec{S}_j(\tau)$, takes the form

$$\int_q [U^{-1} - \Pi_0(q)] \vec{S}_{-q} \cdot \vec{S}_q, \quad (56a)$$

$$\Pi_0(q) = - \int_k G_0(k) G_0(k+q). \quad (56b)$$

Here and in the following, $k = (\mathbf{k}, ik_0)$ and $q = (\mathbf{q}, iq_0)$ are collective variables comprising lattice momenta and Matsubara frequencies, and $\int_k = T \sum_{k_0} \int \frac{d^2 \mathbf{k}}{(2\pi)^2}$ is a shorthand notation for a Matsubara frequency sum and a momentum integration. We define T^* as the critical temperature where $\min_{\mathbf{q}} [U^{-1} - \Pi_0(\mathbf{q}, 0)] = 0$, signaling an instability towards the formation of magnetic order. This condition is first met, in the most general case, at four symmetry related wave vectors in the Brillouin zone of the form $\mathbf{q} = \pm \mathbf{Q}_x$ or $\mathbf{q} = \pm \mathbf{Q}_y$, with $\mathbf{Q}_x = (\pi - 2\pi\eta, \pi)$ and $\mathbf{Q}_y = (\pi, \pi - 2\pi\eta)$. This means that the magnetic order forming right below T^* can be characterized by these wave vectors. Note that, because of the opposite sign between the two terms on the right hand side of Eq. (51), the coefficient of the term quadratic in ρ_q is $U^{-1} + \Pi_0(q)$, which is always positive. Thus, within mean-field theory, an instability towards charge order alone can never occur in the Hubbard model.

For a mean-field study of the Taylor-expanded effective action (54), we can thus assume that $\vec{S}_q$ possesses solely modes at $\mathbf{q} = \pm \mathbf{Q}_x$ and $\mathbf{q} = \pm \mathbf{Q}_y$ close to $T = T^*$, corresponding to the ansatz

$$\vec{S}_q = \left[\vec{M}_x \delta(\mathbf{q} - \mathbf{Q}_x) + \vec{M}_x^* \delta(\mathbf{q} + \mathbf{Q}_x) + \vec{M}_y \delta(\mathbf{q} - \mathbf{Q}_y) + \vec{M}_y^* \delta(\mathbf{q} + \mathbf{Q}_y) \right] \delta_{q_0, 0}, \quad (57)$$

where $\vec{M}_x$ and $\vec{M}_y$ are constant complex vectors. We assume static fields, consistent with our mean-field treatment.

Third order terms involving only spin fields vanish due to time-reversal symmetry. The third order term involving two $\vec{S}_q$ fields and one ρ_q field takes the form

$$\int_{q, q'} i\lambda(q, q') \vec{S}_q \cdot \vec{S}_{-q'} \rho_{q'-q}, \quad (58)$$

with a coupling function $\lambda(q, q')$. Inserting Eq. (57) into this equation, we see that, within mean-field theory, the only charge modes that couple to $\vec{M}_x$ and $\vec{M}_y$ are those where $\mathbf{q} = 0$, $\mathbf{q} = \pm 2\mathbf{Q}_x$, $\mathbf{q} = \pm 2\mathbf{Q}_y$ and $\mathbf{q} = \pm(\mathbf{Q}_x \pm \mathbf{Q}_y)$. Neglecting the $\mathbf{q} = 0$ mode, which does not lead to any symmetry breaking, and higher order spin-charge interactions (this approximation will be justified below), we can write

$$i\rho_q = \left[\phi_x \delta(\mathbf{q} - 2\mathbf{Q}_x) + \phi_x^* \delta(\mathbf{q} + 2\mathbf{Q}_x) + \phi_y \delta(\mathbf{q} - 2\mathbf{Q}_y) + \phi_y^* \delta(\mathbf{q} + 2\mathbf{Q}_y) \right. \\ \left. + \phi_+ \delta(\mathbf{q} - \mathbf{Q}_+) + \phi_+^* \delta(\mathbf{q} + \mathbf{Q}_+) + \phi_- \delta(\mathbf{q} - \mathbf{Q}_-) + \phi_-^* \delta(\mathbf{q} + \mathbf{Q}_-) \right] \delta_{q_0, 0}, \quad (59)$$

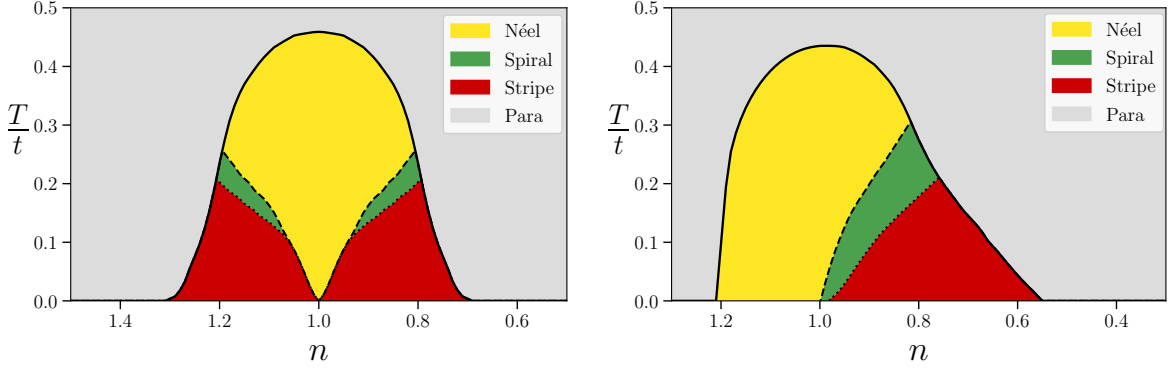


Fig. 5: Mean field phase diagrams in the thermodynamic limit for $U = 3t$ with $t' = 0$ (left) and $t' = -0.15t$ (right). All transition lines have been computed in the thermodynamic limit. The magnetic order in the “stripe” regime is always coplanar, but not necessarily collinear. Figures taken from Ref. [39].

where ϕ_x, ϕ_y , and $\phi_{\pm}$ are complex constants, and $\mathbf{Q}_{\pm} = \mathbf{Q}_x \pm \mathbf{Q}_y$.

Inserting Eq. (57) and (59) into Eq. (54), and expanding up to quartic order in $\vec{M}_x$ and $\vec{M}_y$, to quadratic order in ϕ_x, ϕ_y , and $\phi_{\pm}$, and to third order in the mixed terms, we obtain the effective potential

$$\begin{aligned}
 V(\vec{M}_x, \vec{M}_y, \phi_x, \phi_y, \phi_{\pm}) = & s \left(|\vec{M}_x|^2 + |\vec{M}_y|^2 \right) \\
 & + u_0 \left(|\vec{M}_x|^2 + |\vec{M}_y|^2 \right)^2 + u_1 \left(|\vec{M}_x|^2 - |\vec{M}_y|^2 \right)^2 \\
 & + u_2 \left(|\vec{M}_x \cdot \vec{M}_x|^2 + |\vec{M}_y \cdot \vec{M}_y|^2 \right) + u_3 \left(|\vec{M}_x \cdot \vec{M}_y|^2 + |\vec{M}_x \cdot \vec{M}_y^*|^2 \right) \\
 & - r_1 (|\phi_x|^2 + |\phi_y|^2) - r_2 (|\phi_+|^2 + |\phi_-|^2) \\
 & + b_1 \left(\phi_x \vec{M}_x^* \cdot \vec{M}_x^* + \phi_y \vec{M}_y^* \cdot \vec{M}_y^* + \text{c.c.} \right) \\
 & + b_2 \left(\phi_+ \vec{M}_x^* \cdot \vec{M}_y^* + \phi_- \vec{M}_x^* \cdot \vec{M}_y + \text{c.c.} \right). \quad (60)
 \end{aligned}$$

The coefficients $s, u_0, u_1, u_2, u_3, r_1, r_2, b_1$, and b_2 are determined by frequency and momentum integrals of products of bare propagators G_0 . The concrete expressions are presented in the Appendix.

The charge degrees of freedom can be eliminated from the theory by imposing $\partial V / \partial \phi_{\alpha} = 0$, for $\alpha = x, y, \pm$, which yields

$$\phi_x = \frac{b_1}{r_1} \vec{M}_x \cdot \vec{M}_x, \quad \phi_y = \frac{b_1}{r_1} \vec{M}_y \cdot \vec{M}_y, \quad (61a)$$

$$\phi_+ = \frac{b_2}{r_2} \vec{M}_x \cdot \vec{M}_y, \quad \phi_- = \frac{b_2}{r_2} \vec{M}_x \cdot \vec{M}_y^*. \quad (61b)$$

From the above equations we see that at the extremal points of the potential V the charge order parameter is a bilinear of the spin order parameter. Therefore, in order to get an effective theory that is at most quartic in $\vec{M}_x$ and $\vec{M}_y$, one has to retain all and only the terms in Eq. (60) in the

expansion of the bosonic action (54). Inserting Eqs. (61) into (60), one gets

$$\begin{aligned}
 V_{\text{eff}}(\vec{M}_x, \vec{M}_y) = & s \left(|\vec{M}_x|^2 + |\vec{M}_y|^2 \right) \\
 & + u_0 \left(|\vec{M}_x|^2 + |\vec{M}_y|^2 \right)^2 + u_1 \left(|\vec{M}_x|^2 - |\vec{M}_y|^2 \right)^2 \\
 & + \tilde{u}_2 \left(|\vec{M}_x \cdot \vec{M}_x|^2 + |\vec{M}_y \cdot \vec{M}_y|^2 \right) + \tilde{u}_3 \left(|\vec{M}_x \cdot \vec{M}_y|^2 + |\vec{M}_x \cdot \vec{M}_y^*|^2 \right),
 \end{aligned} \tag{62}$$

with $\tilde{u}_2 = u_2 + b_1^2/r_1$ and $\tilde{u}_3 = u_3 + b_2^2/r_2$. Minimizing $V_{\text{eff}}(\vec{M}_x, \vec{M}_y)$ with respect to $\vec{M}_x$ and $\vec{M}_y$ one can determine the magnetic state at temperatures right below T^* . A Landau theory with an effective potential of the form (62) has previously been derived from general symmetry arguments [41, 49, 50]. The form of the Landau theory restricted to the case of a single mode ($\vec{M}_x$ or $\vec{M}_y$) was derived earlier in Ref. [42].

3.4 Spiral to stripe transition and comprehensive phase diagrams

All calculations show that essentially three types of magnetic order can occur in the two-dimensional Hubbard model: Néel, spiral, and (in a broad sense) stripe. The stripe order is associated with charge order. Its spin order is usually collinear, but not always. The phase boundaries of these orders can be computed directly in the thermodynamic limit, as described above. In Fig. 5 the corresponding phase diagrams are shown for $U = 3t$ and two choices of t' . The phase diagram for $t' = 0$ is particle-hole symmetric. The ground state is Néel ordered only at half-filling in this case, and stripe ordered otherwise, while spiral order occurs only in a small region at finite temperatures. For $t' = -0.15t$, the entire electron doped regime is Néel ordered, and for small hole-doping there is now a spiral regime even in the ground state. As a general trend, the spiral regime becomes larger for a larger $|t'|$.

The Landau theory is consistent with the phase diagrams in Fig. 5. Minimizing the potential $V_{\text{eff}}(\vec{M}_x, \vec{M}_y)$ in Eq. (62), one obtains the same sequence of Néel, spiral, and stripe states close to T^* . For $t' = -0.3t$, the Landau theory yields a coplanar bidirectional stripe phase for very large hole doping ($n < 0.6$) [40].

For $T < T^*$, the regime labeled as “stripe” in Fig. 5 hosts actually two distinct types of order: a non-collinear phase near the spiral regime, and a collinear stripe phase at larger hole doping [40]. Both are stripe phases in the sense that they are associated with unidirectional charge order. The structure of the non-collinear phase is rather complex. It is a superposition of several spiral modulations with three or four distinct wave vectors. Two of them are determined by the wave vector at which the spin-charge susceptibility diverges upon approaching the boundary of the (single $\mathbf{Q}$) spiral regime. The “multi-spiral” regime between the conventional spiral and the collinear stripe phase is quite narrow, and its width shrinks to zero upon approaching T^* . A complete phase diagram showing all the phase transitions is provided in Fig. 6. The transition points at $T = 0$ and T^* have been computed for $U = 3t$ and $t' = -0.3t$. The coplanar bidirectional stripe phase at very large hole doping (small n) does not exist for the smaller values of $|t'|$ in Fig. 5. The nature of the phase between the collinear unidirectional stripe phase and the coplanar bidirectional stripe phase (marked by a question mark) has not yet been clarified.

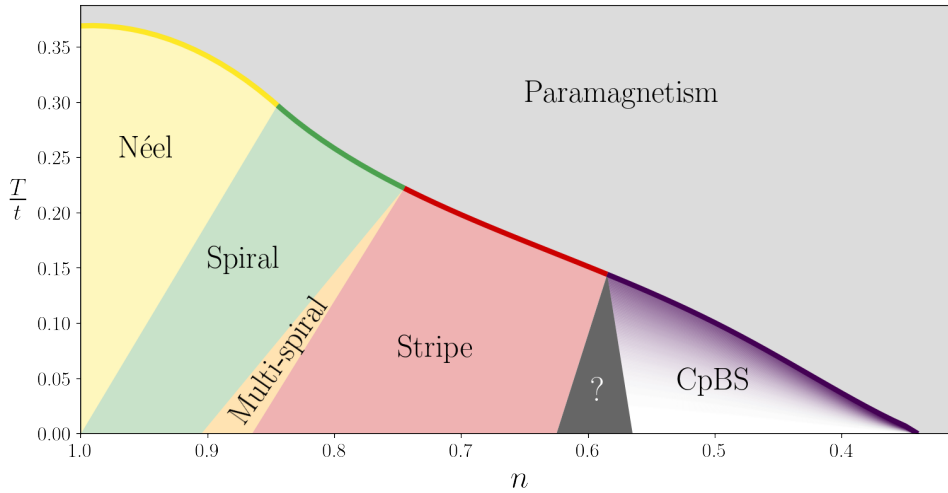


Fig. 6: Mean field phase diagram for $U = 3t$ and $t' = -0.3t$. The various types of magnetic instabilities at T^* were determined from the Landau theory, the states at $T = 0$ by solving the Hartree-Fock equations in the thermodynamic limit. The phase transitions for $0 < T < T^*$ are indicated only schematically by straight lines connecting the calculated transition points at $T = 0$ and $T = T^*$. Figure taken from Ref. [40].

4 Renormalized mean-field theory

The Hartree-Fock theory discussed in the preceding sections yields valuable hints on possible magnetic ordering tendencies. However, it grossly overestimates the size of the ordered regimes, even in the ground state, and even for weak interaction strengths. Moreover, it completely misses the pairing instability, since the attractive d -wave pairing in the two-dimensional Hubbard model is purely fluctuation driven.

At weak or moderate interactions, these limitations can be overcome by a *renormalized* mean-field theory, where the mean-field equations are evaluated with renormalized (effective) interactions, which can be computed from a functional renormalization group (fRG) flow [51]. This fusion of the fRG with mean-field theory can be formulated quite systematically, and without any bias on a specific ordering channel [52].

A detailed discussion of the fRG+MF method is beyond the scope of these lectures. However, it is instructive to compare the resulting phase diagrams to those obtained from the plain Hartree-Fock approximation. A phase diagram obtained recently [53] from the renormalized mean-field theory with an unrestricted evaluation of the corresponding mean-field equations is shown in Fig. 7. The size of the magnetic regime is much smaller compared to the Hartree-Fock result, in spite of the larger interaction $U = 4t$. Indeed, the *effective* magnetic interaction is reduced by fluctuations to $U^m \approx 2t$. Moreover, a sizable effective d -wave pairing interaction U^p leads to extended superconducting regimes, often coexisting with the magnetic order. Looking at details one can see that the presence of superconductivity favors spiral order over stripe order, and it suppresses magnetic order completely at the van Hove filling $n \approx 0.83$, where the Fermi level is situated at the van Hove singularity of the bare density of states.

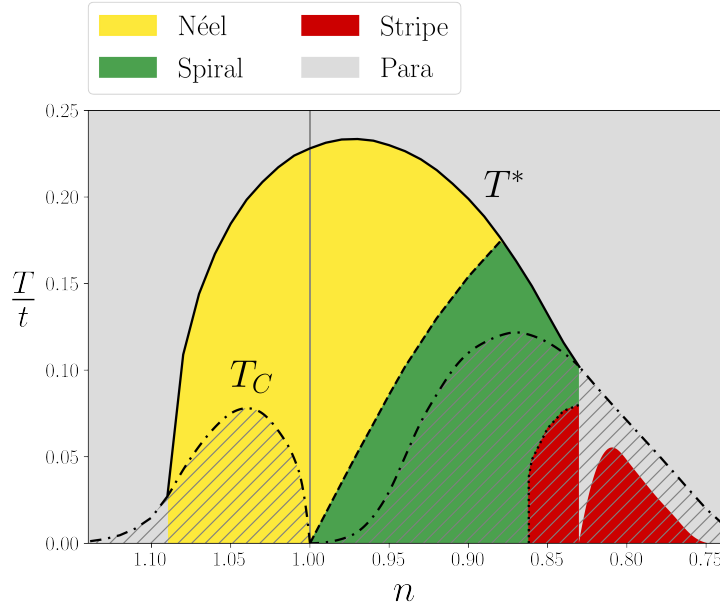


Fig. 7: (n, T) phase diagram of the two-dimensional Hubbard model for $U = 4t$ and $t' = -0.2t$ as obtained from renormalized mean-field theory. All the black lines have been computed directly in the thermodynamic limit. The solid line indicates the onset of magnetic order. The dash-dotted lines show the onset of superconductivity, and the shaded areas the superconducting regions. The dotted line indicates the instability of the spiral states, as signaled by a diverging charge susceptibility. Near the spiral region, the region labeled as “stripe” may also contain a non-collinear charge ordered state. The boundary of the small stripe-ordered dome at low densities could only be estimated from finite size calculations. Figure taken from Ref. [53].

The Mermin-Wagner theorem prohibiting the spontaneous breaking of continuous symmetries at finite temperatures in two dimensions is violated also by the renormalized mean-field theory (as in plain Hartree-Fock). Hence, T^* in Fig. 7 indicates only the onset of strong magnetic correlations, and T_C the temperature scale for the formation of Cooper pairs. A complete theory needs to take order parameter fluctuations into account, but the renormalized mean-field theory is a useful starting point, and it identifies the key players in the intriguing interplay of competing correlations in the two-dimensional Hubbard model.

Acknowledgement: The material reviewed in these lecture notes is based on a wonderful collaboration with Robin Scholle, Pietro Bonetti, and Demetrio Vilardi [39, 40, 53].

Appendix

A Landau coefficients

In this section, the microscopic expressions for the coefficients of the effective potential in Eq. (60) are presented. The coefficients of the quadratic terms are given by

$$s = \frac{2}{U} - 2\Pi_0(Q_x) = \frac{2}{U} - 2\Pi_0(Q_y), \quad (63a)$$

$$r_1 = \frac{2}{U} + 2\Pi_0(2Q_x) = \frac{2}{U} + 2\Pi_0(2Q_y), \quad (63b)$$

$$r_2 = \frac{2}{U} + 2\Pi_0(Q_x+Q_y) = \frac{2}{U} + 2\Pi_0(Q_x-Q_y), \quad (63c)$$

where the bare bubble $\Pi_0(q)$ has been defined in Eq. (56), and $Q_\alpha = (\mathbf{Q}_\alpha, 0)$ for $\alpha = x, y$.

The third order coefficients b_1 and b_2 are given by

$$b_1 = 2 \int_k G_0(k) G_0(k+Q_x) G_0(k+2Q_x) = \{x \leftrightarrow y\}, \quad (64a)$$

$$b_2 = 4 \int_k G_0(k) G_0(k+Q_x) G_0(k+Q_x+Q_y) = \{x \leftrightarrow y\}. \quad (64b)$$

The fourth-order coefficients can be conveniently expressed in terms of the integrals

$$\begin{aligned} E_1 &= \int_k G_0(k) G_0(k+Q_x) G_0(k+2Q_x) G_0(k+Q_x), \\ E_2 &= \int_k G_0(k) G_0(k+Q_x) G_0(k) G_0(k+Q_x), \\ E_3 &= \int_k G_0(k) G_0(k+Q_x) G_0(k) G_0(k+Q_y), \\ E_4 &= \int_k G_0(k) G_0(k+Q_x) G_0(k+Q_x+Q_y) G_0(k+Q_y), \end{aligned} \quad (65)$$

in the form $u_0 = E_2+2E_3-E_4$, $u_1 = E_2-2E_3+E_4$, $u_2 = 2E_1-E_2$, and $u_3 = 4E_4$.

References

- [1] D.J. Scalapino, *Rev. Mod. Phys.* **84**, 1383 (2012)
- [2] D.P. Arovas, E. Berg, S.A. Kivelson, and S. Raghu, *Annu. Rev. Condens. Matter Phys.* **13** 239 (2022)
- [3] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, *Annu. Rev. Condens. Matter Phys.* **13** 275 (2022)
- [4] B.I. Shraiman and E.D. Siggia, *Phys. Rev. Lett.* **62**, 1564 (1989)
- [5] B.I. Shraiman and E.D. Siggia, *Phys. Rev. B* **46**, 8305 (1992)
- [6] A.V. Chubukov and D.M. Frenkel, *Phys. Rev. B* **46**, 11884 (1992)
- [7] A.V. Chubukov and K.A. Musaelian, *Phys. Rev. B* **51**, 12605 (1995)
- [8] T. Dombre, *J. Phys. France* **51**, 847 (1990)
- [9] R. Frésard, M. Dzierzawa, and P. Wölfle, *Europhys. Lett.* **15**, 325 (1991)
- [10] A.P. Kampf, *Phys. Rev. B* **53**, 747 (1996)
- [11] V.N. Kotov and O.P. Sushkov, *Phys. Rev. B* **70**, 195105 (2004)
- [12] P.A. Igoshev, M.A. Timirgazin, A.A. Katanin, A.K. Arzhnikov, and V.Y. Irkhin, *Phys. Rev. B* **81**, 094407 (2010)
- [13] H. Yamase, A. Eberlein, and W. Metzner, *Phys. Rev. Lett.* **116**, 096402 (2016)
- [14] A. Eberlein, W. Metzner, S. Sachdev, and H. Yamase, *Phys. Rev. Lett.* **117**, 187001 (2016)
- [15] J. Mitscherling and W. Metzner, *Phys. Rev. B* **98**, 195126 (2018)
- [16] P.M. Bonetti, J. Mitscherling, D. Vilardi, and W. Metzner, *Phys. Rev. B* **101**, 165142 (2020)
- [17] H.J. Schulz, *J. Phys. France* **50**, 2833 (1989)
- [18] J. Zaanen and O. Gunnarsson, *Phys. Rev. B* **40**, 7391 (1989)
- [19] D. Poilblanc and T.M. Rice, *Phys. Rev. B* **39**, 9749 (1989)
- [20] H.J. Schulz, *Phys. Rev. Lett.* **64**, 1445 (1990)
- [21] M. Kato, K. Machida, H. Nakanishi, and M. Fujita, *J. Phys. Soc. Jpn.* **59**, 1047 (1990)
- [22] G. Seibold, C. Castellani, C. Di Castro, and M. Grilli, *Phys. Rev. B* **58**, 13506 (1998)

- [23] M. Fleck, A.I. Lichtenstein, E. Pavarini, and A.M. Oleś, Phys. Rev. Lett. **84**, 4962 (2000)
- [24] M. Raczkowski and F.F. Assaad, Phys. Rev. B **82**, 233101 (2010)
- [25] R. Peters and N. Kawakami, Phys. Rev. B **89**, 155134 (2014)
- [26] S.A. Kivelson, I.P. Bindloss, E. Fradkin, V. Oganessian, J.M. Tranquada, A. Kapitulnik, and C. Howald, Rev. Mod. Phys. **75**, 1201 (2003)
- [27] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R.M. Noack, H. Shi, S.R. White, S. Zhang, and G.K.-L. Chan, Science **358**, 1155 (2017)
- [28] M. Qin, C.M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S.R. White, and S. Zhang, Phys. Rev. X **10**, 031016 (2020)
- [29] H.-C. Jiang and T.P. Devereaux, Science **365**, 1424 (2019)
- [30] F. Simkovic, R. Rossi, and M. Ferrero, Phys. Rev. Res. **4** 043201 (2022)
- [31] B. Xiao, Y.-Y. He, A. Georges, and S. Zhang, Phys. Rev. X **13**, 011007 (2023)
- [32] H.Q. Lin and J.E. Hirsch, Phys. Rev. B **35**, 3359 (1987)
- [33] W. Hofstetter and D. Vollhardt, Ann. Phys. (Berlin) **510**, 48 (1998)
- [34] E. Langmann and M. Wallin, J. Stat. Phys. **127**, 825 (2007)
- [35] N.D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966)
- [36] M.S. Scheurer, S. Chatterjee, W. Wu, M. Ferrero, A. Georges, and S. Sachdev, Proc. Natl. Acad. Sci. USA **115**, E3665 (2018)
- [37] S. Sachdev, Rep. Prog. Phys. **82**, 014001 (2019)
- [38] P.M. Bonetti and W. Metzner, Phys. Rev. B **106**, 205152 (2022)
- [39] R. Scholle, P.M. Bonetti, D. Vilardi, and W. Metzner, Phys. Rev. B **108**, 035139 (2023)
- [40] R. Scholle, W. Metzner, D. Vilardi, and P.M. Bonetti, Phys. Rev. B **109**, 235149 (2024)
- [41] S. Sachdev, H.D. Scammell, M.S. Scheurer, and G. Tarnopolsky, Phys. Rev. B **99**, 054516 (2019)
- [42] O. Zachar, S.A. Kivelson, and V.J. Emery, Phys. Rev. B **57**, 1422 (1998)
- [43] P.M. Bonetti and W. Metzner, Phys. Rev. B **105**, 134426 (2022)
- [44] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)
- [45] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Addison-Wesley, 1987)

- [46] Z.Y. Weng, C.S. Ting, and T.K. Lee, Phys. Rev. B **43**, 3790 (1991)
- [47] H. Schulz, in D. Baeriswyl (ed.): *The Hubbard Model* (Plenum, New York, 1995)
- [48] K. Borejsza and N. Dupuis, Phys. Rev. B **69**, 085119 (2004)
- [49] Y. Zhang, E. Demler, and S. Sachdev, Phys. Rev. B **66**, 094501 (2002)
- [50] M. De Prato, A. Pelissetto, and E. Vicari, Phys. Rev. B **74**, 144507 (2006)
- [51] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer, Rev. Mod. Phys. **84**, 299 (2012)
- [52] J. Wang, A. Eberlein, and W. Metzner, Phys. Rev. B **89**, 121116 (2014)
- [53] R. Scholle, P.M. Bonetti, W. Metzner, and D. Vilardi, arXiv:2602.20073

8 Introduction to Topological Phases in Condensed Matter

Cécile Repellin
LPMMC, CNRS, Université Grenoble Alpes
25 avenue des Martyrs, 38000 Grenoble, France

Contents

1	Organizing principles in condensed matter	2
1.1	Symmetry breaking	2
1.2	The quantum Hall effect	3
1.3	Topological invariants and topological equivalence	4
1.4	Different types of topology	4
2	Berry phase in quantum mechanics	5
2.1	Definition of the Berry phase	5
2.2	Gauge invariance, Berry connection and Berry curvature	7
2.3	Analogy to electromagnetism	8
3	Chern number in a condensed matter system	8
3.1	Bloch states and Chern number	8
3.2	Chern number without translation invariance	9
3.3	Many-body Chern number	9
3.4	Explicitly gauge invariant expression of the Chern number	10
3.5	Chern number and Hall conductance	11
3.6	The Chern number is an integer	11
3.7	Edge states	13
4	Chern insulator models	15
4.1	Harper-Hofstadter model	15
4.2	Haldane model	18
4.3	Chern number and spontaneous symmetry breaking in moiré materials	20

Topological phases of matter are an extremely rich and active field of research. The first topological phase of matter ever observed was the integer quantum Hall effect, which was first observed by von Klitzing in 1980 [1]. Yet, it is only with the theoretical understanding of topology as an organizing principle of phases of matter in the 2000s that the field really started to blossom. Along with theoretical progress came the discovery of new materials, and the realization that many known materials are in fact topological. Topological phases are comprised of phases as varied as insulators, semimetals, and superconductors, and are observed in any dimension. They reach many experimental fields, from bulk 3D materials to two-dimensional heterostructures, thin films, as well as synthetic quantum matter, such as cold atoms in optical lattices, photonic systems, Josephson junction arrays ... As they apply to all wave phenomena, the same topological unifying principles also describe non-quantum systems, from oceanic waves to active matter, or mechanical resonators. These lectures notes focus on non-interacting topological insulators in two-dimensional quantum systems, and more specifically on the example of the Chern insulator. In Sec. 1, we give a very qualitative introduction to topological phases of matter by presenting it as an organizing principle in condensed matter. In Sec. 2–3, we define core mathematical concepts of topology, such as the Berry phase, Berry curvature and Chern number, and highlight their relevance to some physical quantities. Finally, in Sec. 4, we detail a few paradigmatic models of Chern insulators, introducing some of the microscopic ingredients allowing topology to emerge, and highlight their relevance to solid state and cold atom experiments.

1 Organizing principles in condensed matter

Understanding the world through the lens of physics often means looking for universality: unifying principles, which describe many phenomena regardless of the physical system, its geometry, constituents, or scale. The emergence of universality is particularly striking in condensed matter, where phases of matter arise from the complex interactions between macroscopically many constituents, and similar phenomena can be observed in extremely varied materials.

1.1 Symmetry breaking

The concept of symmetry breaking is an immensely helpful organizing principle of phases of matter, describing crystals, magnets, superfluids, and more. It can be summarized like so: the low-temperature phase of a system is less symmetric than the Hamiltonian describing it. For example, the position of atoms in a crystalline solid breaks the continuous translation and rotation symmetry, which is restored at higher temperature, in the liquid and gas phases. Likewise, magnets form when the magnetization of a solid selects a preferential orientation, thus breaking rotational symmetry.

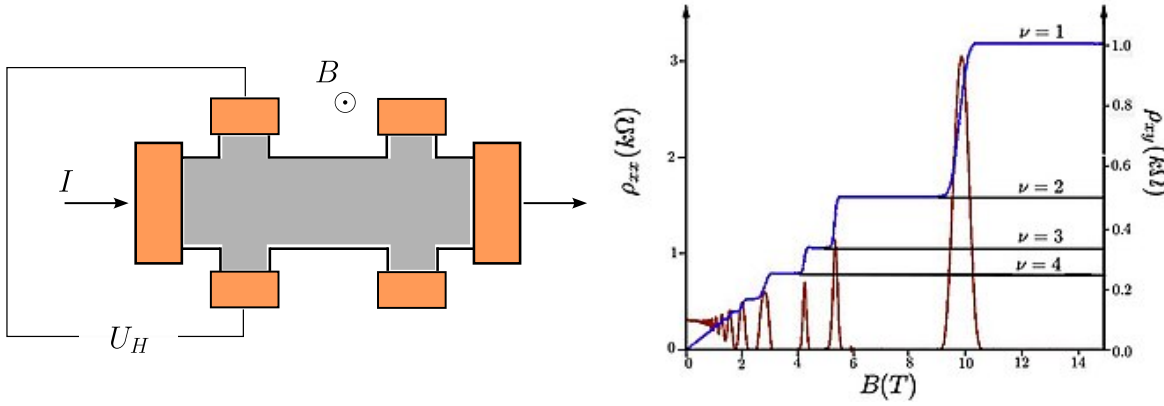


Fig. 1: Integer quantum Hall effect. (Left) Set-up of a Hall experiment. A sample is placed in a magnetic field B , and a current I is imposed. A transverse or Hall voltage U_H is measured. (Right) Longitudinal and transverse (Hall) resistivity ρ_{xx} and $\rho_{xy} = U_H/I$, of a two-dimensional electron gas as a function of magnetic field. Both quantities were divided by the quantum unit of conductance e^2/h . Whenever $\rho_{xx} = 0$, the longitudinal conductance σ_{xx} vanishes as well, and the transverse resistance $\rho_{xy} = 1/\sigma_{xy}$ is quantized and forms a plateau. The integer ν is the ratio between electronic density and magnetic field (in units of the quantum of magnetic flux), and is known as the filling factor. It is equal to the Hall conductance in units of e^2/h . Figure reproduced from Wikipedia, *The Quantum Hall Effect*.

1.2 The quantum Hall effect

While the concept of symmetry breaking describes a very large number of phases of matter, it fails (at least in its traditional formulation) to explain several phenomena. The integer quantum Hall effect, discovered in 1980 by Klaus von Klitzing [1] is the most emblematic, and historically the first such phenomenon. It describes the quantization of the transverse (or Hall) conductance σ_{xy} of a two-dimensional electron gas (2DEG) subjected to a perpendicular magnetic field (see Fig. 1)

$$\sigma_{xy} = n \frac{e^2}{h} = n \frac{1}{25812.807 \Omega} \quad (1)$$

where n is an integer, and e and h are universal constants, respectively, the charge of the electron, and the Planck constant. Originally observed in silicon-based transistors at low temperature, it is often studied in gallium arsenide heterostructures, and is also observed in graphene at room temperature [2]. In spite of the important differences between these materials, and the inherent presence of disorder, which differs from sample to sample, the quantization occurring in the quantum Hall effect is expressed purely in terms of universal constants, and does not require any factor accounting for sample variety. It is also remarkably precise (up to the 9th digit), which has made the integer quantum Hall effect the standard of metrology for measuring the quantum unit of conductance e^2/h .

1.3 Topological invariants and topological equivalence

The remarkable quantization of the integer quantum Hall effect is of course not by chance, but rather the result of the mathematical nature of the Hall conductance: when the longitudinal conductance vanishes (which occurs in the QHE), it is equal, in units of e^2/h , to a topological invariant which can only take integer values, the Chern number. This peculiar situation, where the longitudinal conductance vanishes, but the transverse one does not, is associated with a state which is insulating in the bulk of the system, with one-dimensional conducting channels at the edge. The edge channels are chiral (the direction of the magnetic field, or sign of the Chern number determines the direction of chirality), which protects them from any backscattering-induced gap opening. Loosely speaking, a topological invariant is a mathematical property of a space, which remains invariant under continuous transformations (homeomorphisms). In the case of the Chern number as applied to condensed matter problems, continuous transformations are those which do not close the bulk gap of the system, i.e., the band gap, as Chern numbers describe materials well described by band theory. This gives an additional hint to explain the universality of the quantization of the Hall conductance across many materials and samples: there is a vast parameter space in which a band gap may remain open.

The robustness of topological phases is associated with the concept of topological equivalence. Two phases of matter are considered to be topologically equivalent if their topological invariant is equal, for example two insulators with equal Chern number. In other words, there is a smooth transformation connecting the two corresponding ground states. For insulators, this means, as already explained in the case of Chern insulators, that there exists a transformation connecting the two states, which does not close the gap. The notion of equivalence generalizes to metals: there exists a finite-depth quantum circuit connecting the two states, i.e., a product of local operators whose number remains finite in the thermodynamic limit. When a symmetry protects the topology (for example time-reversal symmetry in the case of $\mathbb{Z}_2$ topological insulators), one considers only local operators preserving the given symmetry to define topological equivalence.

1.4 Different types of topology

The realization that some topological invariants are connected to physical properties was a milestone in classifying phases of matter, and goes much beyond the connection between the Chern number and the Hall conductance. The Chern number is a property of insulators in 2 dimensions not preserving time-reversal symmetry. Two-dimensional insulators preserving time-reversal symmetry are characterized by a topological invariant taking only two possible values (0 or 1), a $\mathbb{Z}_2$ invariant, as was clarified by C. Kane and E. Mele [3]. The classification of topological phases extends to any dimension, and any symmetry in a table called the tenfold way [4], which holds for weakly interacting quantum matter. This is the context where band theory applies, and the topological invariant then characterizes the individual bands of a system. This type of topological phase of matter – *topological bands* – is to be clearly distinguished from the vastly different concept of *topological order*. A state with topological order is a gapped many-body state with a complex entanglement structure, which cannot be adiabati-

cally connected to a product of single-particle states (unlike, e.g., a Chern insulator). In turn, the elementary excitations are fractionalized: they are called anyons, and may carry a fractional charge and a fractional statistics (beyond bosons and fermions). Non-abelian anyons have an especially complex exchange statistics, which makes them promising for quantum computing technologies: they carry an intrinsic degeneracy, permitting logical gate operations, and the realization of qubits topologically immune to decoherence, in principle. The fractional quantum Hall effect is an example of a phase of matter with topological order (abelian or non-abelian depending on the fraction). In the context of frustrated magnetism, gapped spin liquids also carry topological order.

For a review of topological phases of matter, including the above examples, as well as semimetals, and topological defects in symmetry-breaking phases, we recommend the book by R. Moessner and J.E. Moore [5]. For a review of topological insulators and superconductors, see also the book by B.A. Bernevig and T. Hughes [6]. For a review of topological order and its link to quantum computation, see the review by Nayak *et al.* [7].

In the rest of these lecture notes, we will focus on one topological phase of matter, the Chern insulator (a two-dimensional insulator characterized by a Chern number). We will define the concepts of Berry phase, Berry connection, and Berry curvature necessary for its definition, and explain why it must be an integer. We will also go over some paradigmatic Chern insulator models, and highlight one very consequential conclusion from the connection between quantized Hall effect and Chern number: the possibility to have a quantized Hall conductance in the absence of a magnetic field.

2 Berry phase in quantum mechanics

The topological invariants characterizing topological phases of matter are geometric properties of the underlying quantum state. Here, we will start by defining the simplest geometric property of a quantum state, the Berry phase [8], which is the phase acquired by a quantum state after undergoing adiabatic evolution, independently of the energy of the quantum state, and is a simple consequence of the adiabatic theorem of quantum mechanics.

2.1 Definition of the Berry phase

Let us consider a quantum system described by a Hamiltonian H_λ , where λ is a control parameter, which may be multidimensional. Initially ($t=0$), the system is prepared in the ground state¹ of $H_{\lambda(t=0)}$, the initial Hamiltonian

$$\Psi(t=0) = \Psi_{\lambda(t=0)}^0. \quad (2)$$

From $t = 0$ to $t = t_f$, the system undergoes an adiabatic evolution. We choose a path in parameter space, such that the ground state Ψ_λ^0 of H_λ is non-degenerate at each point, so that it can be

¹In all generality, the Berry phase is defined for any non-degenerate eigenstate. We use the ground state to simplify notation, but the generalization is straightforward.

differentiated along this path.² The adiabatic theorem states that for an adiabatic evolution, the system must remain in the ground state of the Hamiltonian H_λ at all times, which is defined up to an overall phase. In other words, at zeroth order of the adiabatic approximation

$$\Psi(t) = e^{i\phi(t)} \Psi_{\lambda(t)}^0. \quad (3)$$

Using this assumption, we can derive the value of $\phi(t)$ using the Schrödinger equation

$$i\hbar \frac{d|\Psi(t)\rangle}{dt} = H_{\lambda(t)} |\Psi(t)\rangle. \quad (4)$$

For practical reasons, we define $c^0(t) = e^{i\phi(t)}$. Inserting the adiabatic approximation into the Schrödinger equation, we obtain

$$i\hbar \dot{c}^0 |\Psi_{\lambda(t)}^0\rangle + i\hbar c^0 \frac{d|\Psi_{\lambda(t)}^0\rangle}{dt} = c^0 H_{\lambda(t)} |\Psi_{\lambda(t)}^0\rangle = c^0 E_{\lambda(t)}^0 |\Psi_{\lambda(t)}^0\rangle, \quad (5)$$

where we have used the notation $\dot{c}^0$ for the time derivative of c^0 , and $E_{\lambda(t)}^0$ is the lowest eigenvalue of $H_{\lambda(t)}$. This leads to the following differential equation for c^0

$$\dot{c}^0 = c^0 \left(-i \frac{E_{\lambda(t)}^0}{\hbar} - \langle \Psi_{\lambda(t)}^0 | \frac{d|\Psi_{\lambda(t)}^0\rangle}{dt} \right). \quad (6)$$

Upon integration, using the initial condition $c^0(t=0) = 1$, or $\Phi(t=0) = 0$, we find

$$\Phi(t_f) = -\frac{1}{\hbar} \int_0^{t_f} E_{\lambda(t)}^0 dt + i \int_0^{t_f} \langle \Psi_{\lambda(t)}^0 | \frac{d|\Psi_{\lambda(t)}^0\rangle}{dt} dt. \quad (7)$$

There are thus two contributions to the phase associated with an adiabatic evolution, which are the two terms in the above expression:

- $\Phi_{\text{dyn}} = -\frac{1}{\hbar} \int_0^{t_f} E_{\lambda(t)}^0 dt$ is the dynamical phase. It depends on the eigenenergies of the Hamiltonian, and on the time it takes to go through the adiabatic evolution.
- $\Phi_{\text{Berry}} = i \int_0^{t_f} \langle \Psi_{\lambda(t)}^0 | \frac{d|\Psi_{\lambda(t)}^0\rangle}{dt} dt$ is the geometric or Berry phase.

Note that the integrand in the expression of the Berry phase is purely imaginary, such that Φ_{Berry} is a real number – it is a phase, not a decay of the wave function. It can easily be seen by noting that $\frac{d}{dt} \langle \Psi_{\lambda(t)}^0 | \Psi_{\lambda(t)}^0 \rangle = 0$ due to the normalization of the wavefunction.

During an adiabatic evolution, the wavefunction of a quantum system thus acquires a phase, the Berry phase, which only depends on the eigenstates visited by the evolution. It is important to note that the Berry phase is actually independent of time: it does not depend on how long it takes to go through the adiabatic path (assuming adiabaticity is preserved, of course). Indeed,

²We are defining here the abelian Berry phase. In case of a degenerate manifold, one can define a non-abelian Berry phase, as long as the manifold itself is separated by a gap from other states.

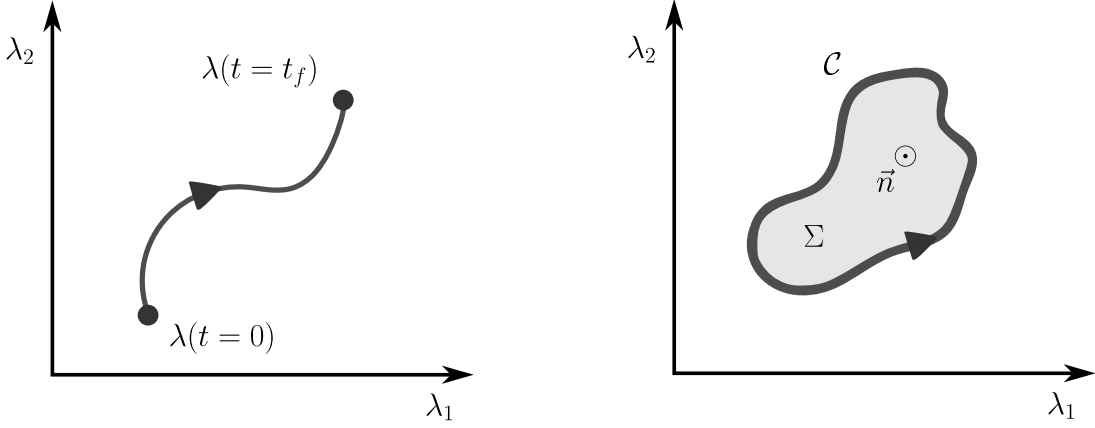


Fig. 2: Open (left) and closed (right) paths in the multidimensional parameter space λ . We have used a two-dimensional parameter space for ease of representation.

time can be removed from the expression of the Berry phase, by replacing the integral over time by an integral over the path in parameter space

$$\Phi_{\text{Berry}} = i \int_0^{t_f} \frac{d\lambda}{dt} \langle \Psi_{\lambda(t)}^0 | \nabla_{\lambda} | \Psi_{\lambda(t)}^0 \rangle dt = i \int_{\lambda(0)}^{\lambda(t_f)} \langle \Psi_{\lambda}^0 | \nabla_{\lambda} | \Psi_{\lambda}^0 \rangle \cdot d\lambda, \quad (8)$$

where we have used the notation $\nabla_{\lambda} = \left(\frac{\partial}{\partial \lambda_1}, \frac{\partial}{\partial \lambda_2}, \dots \right)$ to express the gradient of the multidimensional variable λ .

2.2 Gauge invariance, Berry connection and Berry curvature

We start by defining the Berry connection (or Berry potential) $\mathcal{A}_{\lambda}$ as the real-valued vector field defined at every point on the parameter path, which, once integrated, is equal to the Berry phase

$$\mathcal{A}_{\lambda} = i \langle \Psi_{\lambda}^0 | \nabla_{\lambda} | \Psi_{\lambda}^0 \rangle. \quad (9)$$

The Berry connection is a gauge dependent quantity. Under a $U(1)$ gauge transformation of the wave function $\Psi_{\lambda}^0 \rightarrow e^{i\chi_{\lambda}} \Psi_{\lambda}^0$, it transforms as

$$\mathcal{A}_{\lambda} \rightarrow \mathcal{A}_{\lambda} - \nabla \chi_{\lambda}. \quad (10)$$

For a generic path, the Berry phase is gauge dependent as well, and transforms as

$$\Phi_{\text{Berry}} \rightarrow \Phi_{\text{Berry}} + \chi_{\lambda(0)} - \chi_{\lambda(t_f)} \quad (11)$$

On the other hand, if the Berry phase is calculated along a closed path, it is clearly gauge independent. We define the notations for the contour $\mathcal{C}$, the enclosed surface Σ , and the unit vector $\vec{n}$ perpendicular to the surface in the right hand side of Fig. 2. If $\mathcal{A}_{\lambda}$ has no singularity within the contour, and the surface Σ is orientable, the contour integral in the expression of the Berry phase may be replaced by a surface integral of the flux of the curl of $\mathcal{A}_{\lambda}$ inside the contour, using Stokes' theorem

$$\Phi_{\text{Berry}} = \oint_{\mathcal{C}} \mathcal{A}_{\lambda} \cdot d\lambda = \iint_{\Sigma} (\nabla_{\lambda} \wedge \mathcal{A}_{\lambda}) \cdot \vec{n} d\sigma. \quad (12)$$

It is convenient, and customary, to define the Berry curvature, as the amplitude of the curl of the Berry connection

$$\Omega_\lambda = (\nabla_\lambda \wedge \mathcal{A}_\lambda) \cdot \vec{n}_z. \quad (13)$$

The Berry curvature is explicitly gauge independent, since the curl of a gradient is 0. If the manifold over which the Berry curvature is defined is compact, its integral over this entire manifold is an integer topological invariant called the Chern number.

2.3 Analogy to electromagnetism

The Berry curvature, Berry connection, and Berry phase, as defined in the previous paragraph, are reminiscent of similar concepts in the physics of electromagnetism. The gauge independent Berry curvature Ω_λ is analogous to a magnetic field, while the gauge dependent Berry connection corresponds to the vector potential $\vec{A}(\vec{r})$, whose curl is equal to the magnetic field. Additionally, a particle of charge q in a magnetic field undergoes the Aharonov-Bohm effect: upon traveling along a path $\mathcal{C}$, it acquires a Berry phase Φ_{AB} , which is called an Aharonov-Bohm phase and has the following expression

$$\Phi_{AB} = \frac{q}{\hbar} \int_{\mathcal{C}} \vec{A}(\vec{r}) \cdot d\vec{r} = \frac{2\pi}{\Phi_0} \int_{\mathcal{C}} \vec{A}(\vec{r}) \cdot d\vec{r}, \quad (14)$$

where Φ_0 is a unit quantum of magnetic flux. For a closed loop, the Aharonov-Bohm phase is gauge independent, and is equal to the flux of the magnetic field through the loop.

3 Chern number in a condensed matter system

With the concepts of Berry curvature, Berry connection, and Berry phase defined in the previous section, we are now equipped to define the Chern number in the context of a condensed matter system. In this context, the Chern number characterizes the way the single-particle wave functions of a system change under variations of momentum. We will define it, explain why it must be an integer, and highlight its connection to the Hall conductance.

3.1 Bloch states and Chern number

In the previous section, we have used a generic parameter space λ ; we will now focus on a two-dimensional space with discrete translation symmetry, such that the momentum along either space direction k_x and k_y is preserved, and defined in a compact space, the Brillouin zone. This is the familiar situation of a two-dimensional lattice system with periodic boundary conditions, and $(k_x, k_y) \in [-\pi/a, \pi/a]^2$, where a is the lattice constant.

In a system with discrete translation invariance, Bloch's theorem states that single-particle states may be expressed as the product of a plane wave $e^{i\vec{k} \cdot \vec{r}}$ and a periodic function $u_{\vec{k}}(r)$, called a Bloch state. We further assume a quantum system described by band theory, with a well isolated band. Note that in principle, one may calculate the Chern number of a group of bands, through

the non-abelian Berry connection, as long as this group is well isolated energetically from other bands. Here, we consider the Chern number of a single band, for simplicity. This condition is to ensure that single-particle eigenstates remain non-degenerate across the Brillouin zone. We can now define the Berry connection and Berry curvature associated with variations of momentum in terms of the Bloch states of said band

$$\mathcal{A}_{\vec{k}}^{x,y} = i \langle u_{\vec{k}} | \partial_{k_{x,y}} u_{\vec{k}} \rangle \quad (15)$$

$$\Omega_{\vec{k}} = \partial_{k_x} \mathcal{A}_{\vec{k}}^y - \partial_{k_y} \mathcal{A}_{\vec{k}}^x. \quad (16)$$

The Chern number of the band is defined as the integral of the Berry curvature over the Brillouin zone

$$C = \frac{1}{2\pi} \int_{BZ} \Omega_{\vec{k}} dk_x dk_y. \quad (17)$$

3.2 Chern number without translation invariance

Let us comment on the assumption of translational invariance: it is a convenient assumption, which allows us to use k_x and k_y , but not a necessary one. For a system without translational invariance (such as a system with edges, or disorder, or an amorphous lattice), one may define a local Chern marker [9]

$$C_{xy}(\mathbf{r}) = -2\pi i \text{Tr} [Q \hat{x}, P \hat{y}], \quad (18)$$

where $\hat{x}, \hat{y}$ are position operators, and $P, Q = \mathbb{1} - P$ are projectors onto occupied and unoccupied states. The local Chern marker may not be quantized, but its average in the bulk of an insulator is.

3.3 Many-body Chern number

The second assumption we made was that of a system described by band theory, such that the approximation of independent electrons can be used, and the Chern number may be calculated from single-particle states. More generally, one may define the many-body Chern number of a correlated insulator, based on its many-body wavefunction Ψ_{MB} [10]. The parameter space over which the evolution of Ψ_{MB} is examined is that of the twist angles θ_x, θ_y of periodic boundary conditions, i.e., the wavefunctions satisfy $\phi(x, y=L_y) = e^{i\theta_y} \phi(x, y=0)$ and $\phi(x=L_x, y) = e^{i\theta_x} \phi(x=0, y)$. For a non-degenerate insulator, which remains non-degenerate at all twist angles, the many-body Berry connection writes

$$\mathcal{A}_{\vec{\theta}}^{x,y} = \langle \Psi_{MB} | \partial_{\theta_{x,y}} \Psi_{MB} \rangle. \quad (19)$$

The many-body Berry curvature and Chern number are then defined very similarly to their band theory counterparts. The many-body Berry curvature writes

$$\Omega_{\vec{\theta}}^{MB} = \partial_{\theta_x} \mathcal{A}_{\vec{\theta}}^y - \partial_{\theta_y} \mathcal{A}_{\vec{\theta}}^x \quad (20)$$

and integration over the twist angles yields the many-body Chern number

$$C^{MB} = \frac{1}{2\pi} \iint \Omega_{\vec{\theta}}^{MB} \quad (21)$$

Note that for large enough systems, the integral over the twist angle may be omitted, and the many-body Chern number may be evaluated from the many-body Berry curvature at any angle, as self-averaging occurs within the many-body wave function in the thermodynamic limit [11].

3.4 Explicitly gauge invariant expression of the Chern number

The Chern number, as expressed in Eq. (17) is clearly gauge invariant, as the integral of the Berry curvature, even though the Bloch states are not. It is convenient to introduce an expression of the Chern number, which is explicitly gauge invariant [12]. This is especially practical when calculating the Chern number in a concrete model: it fixes any gauge discrepancy introduced by the determination of Bloch states through numerical diagonalization without having to introduce a smooth gauge explicitly.

To obtain the gauge invariant expression of the Chern number, we will decompose it into the sum of Berry phases associated with the change of momentum when going around the infinitesimal square loop $l_{\vec{k}}$

$$l_{\vec{k}} : \vec{k} \rightarrow \vec{k} + d\vec{k}_x \rightarrow \vec{k} + d\vec{k}_x + d\vec{k}_y \rightarrow \vec{k} + d\vec{k}_y \rightarrow \vec{k}. \quad (22)$$

In terms of the Bloch states, this Berry phase may be written as

$$\begin{aligned} \gamma_{\vec{k}} &= \oint_{l_{\vec{k}}} \mathcal{A} \cdot d\vec{l} \\ &= idk_x \langle u_{\vec{k}} | \partial_{k_x} u_{\vec{k}} \rangle + idk_y \langle u_{\vec{k}+d\vec{k}_x} | \partial_{k_y} u_{\vec{k}+d\vec{k}_x} \rangle \\ &\quad - idk_x \langle u_{\vec{k}+d\vec{k}_x+d\vec{k}_y} | \partial_{k_x} u_{\vec{k}+d\vec{k}_x+d\vec{k}_y} \rangle - idk_y \langle u_{\vec{k}+d\vec{k}_y} | \partial_{k_y} u_{\vec{k}+d\vec{k}_y} \rangle. \end{aligned} \quad (23)$$

Using the following identity (first order expansion of the exponential)

$$idt \langle u | \frac{d}{dt} u \rangle = -\text{Im} \log \langle u(t) | u(t+dt) \rangle + \mathcal{O}(dt^2) \quad (24)$$

the infinitesimal Berry phase is expressed as an overlap of Bloch wavefunctions, which is explicitly gauge invariant

$$\gamma_{\vec{k}} = -\text{Im} \log \langle u_{\vec{k}} | u_{\vec{k}+d\vec{k}_x} \rangle \langle u_{\vec{k}+d\vec{k}_x} | u_{\vec{k}+d\vec{k}_x+d\vec{k}_y} \rangle \langle u_{\vec{k}+d\vec{k}_x+d\vec{k}_y} | u_{\vec{k}+d\vec{k}_y} \rangle \langle u_{\vec{k}+d\vec{k}_y} | u_{\vec{k}} \rangle. \quad (25)$$

The Chern number may thus be evaluated through an explicitly gauge invariant product of overlaps of Bloch states, as

$$C = \frac{1}{2\pi} \int_{BZ} \gamma_{\vec{k}} dk_x dk_y. \quad (26)$$

3.5 Chern number and Hall conductance

The remarkable discovery by Thouless, Khomoto, Nightingale, den Nijs (TKNN) [13] was to connect the Hall or transverse conductance of insulators to the Chern number

$$\sigma_{xy} = \frac{e^2}{h} \sum_{i \in \text{occupied}} C_i \quad (27)$$

In units of the quantum of conductance e^2/h , the transverse conductance is given by the sum of the Chern numbers of all occupied bands. Because of this discovery, the Chern number in the context of condensed matter is sometimes called the TKNN integer.

The goal of TKNN was to explain the quantization of σ_{xy} in the integer quantum Hall effect (so far explained through Laughlin's charge pumping argument, as we will explain later), by connecting it mathematically to a quantized quantity. To do so, they introduced a periodic potential, and calculated σ_{xy} within linear response through the Kubo formula, yielding the Chern number. As already mentioned in the first section of these lecture notes, this relationship between the Hall conductance of insulators and the Chern number explains the remarkable accuracy (metrologically precise) of the quantization of σ_{xy} , regardless of the specific details of the disorder landscape in the sample.

3.6 The Chern number is an integer

Here, we expose two arguments explaining why the Hall conductance of insulators must be quantized. The first one is Laughlin's historical argument for the quantization of σ_{xy} in the integer quantum Hall effect. The second one is a more mathematical argument, based on the definition of different gauges throughout the Brillouin zone to calculate the Chern number.

3.6.1 Laughlin's argument: the charge pump

Soon after the experimental discovery of the integer quantum Hall effect, Laughlin proposed the first explanation of the quantization of the Hall conductance [14]. Laughlin's argument is based on a topological charge pump, and applies to any system with vanishing longitudinal conductance (any insulator). We summarize it here in simple terms.

Let us picture a sample in the cylinder geometry, i.e., a rectangle with periodic boundary conditions along the y direction (see Fig. 3). Threading a magnetic flux along the cylinder axis does not change the overall magnetic field going through the sample (no additional Hall effect). However, due to Faraday's law of induction, such magnetic flux creates a current along the x direction, from one edge of the cylinder to the other. For an adiabatic flux insertion, this current may be approximated as the adiabatic derivative of the energy with respect to the inserted flux

$$J = c \frac{\partial E_\Phi}{\partial \Phi}. \quad (28)$$

For a large enough cylinder perimeter, Aharonov-Bohm oscillations are suppressed, and we can replace the derivative by a finite difference

$$J = c \frac{\Delta E}{\Delta \Phi}. \quad (29)$$

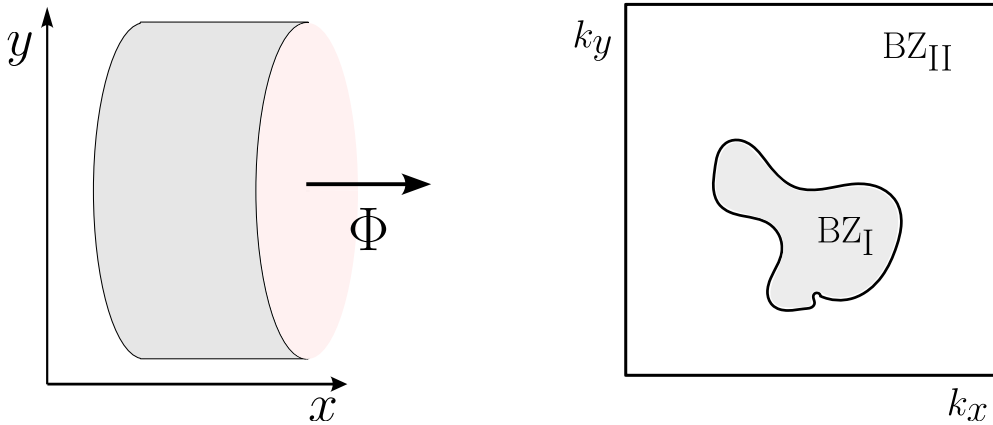


Fig. 3: Arguments for the quantization of σ_{xy} and C . (Left) Laughlin charge pump. As a quantum of magnetic flux is adiabatically inserted along the axis of a cylinder, an integer number of electrons is transferred from one edge to the other. (Right) When the Chern number is $C \neq 0$, it is impossible to define a smooth gauge over the whole Brillouin zone. We define at least two regions I and II, and the Chern number may be calculated from the singularity of wave functions at their boundary.

We consider the insertion of a quantum of flux $\Phi_0 = hc/e$. After the insertion, the Hamiltonian must go back to its initial form up to a gauge transform, such that all energy levels remain the same. The difference of energy thus originates only from the difference of occupation of these states, i.e., the transfer of n electrons from one edge to the other, where p is an integer, which results in a difference of potential V between both edges. The Hall conductance (for vanishing longitudinal conductance) then writes

$$\sigma_{xy} = \frac{J}{V} = pc \frac{e}{\Phi_0} = n \frac{e^2}{h}. \quad (30)$$

As expected, we find that the Hall conductance is quantized, and equal to an integer in units of the quantum of conductance.

3.6.2 The Chern number as an obstruction to the definition of a smooth gauge for Bloch states

There are several ways to gain an intuitive understanding of what is the Chern number. Here, we will explain one such possible interpretation: when the Chern number is non-zero, there must be some gauge discontinuity for the Bloch states – in other words, it is not possible to use the same gauge definition throughout the Brillouin zone.

Let us first reason by contradiction, and assume that there exists a smooth gauge throughout the entire Brillouin zone. The gauge defined for the Bloch states also applies to the Berry connection, such that we may use Stokes' theorem to express the Chern number as an integral of the Berry connection over the surface boundary $\mathcal{C}$. The surface being the Brillouin zone, it does not have a boundary: the contour $\mathcal{C}$ is reduced to a point, and the Chern number is 0.

$$C = \frac{1}{2\pi} \int_{BZ} \Omega_{\vec{k}} dk_x dk_y = \oint_{\mathcal{C}} \mathcal{A} \cdot d\vec{l} = 0. \quad (31)$$

To gain further insight, let us take the example of a generic two-band model, such that there are two sites A and B per unit cell. The Bloch states are 2-dimensional spinors parametrized by the complex numbers α_k and β_k

$$|u_k\rangle = \alpha_k |A_k\rangle + \beta_k |B_k\rangle. \quad (32)$$

A possible gauge choice, which we call gauge I, is to fix that α_k is a real positive number, which is a valid gauge choice, as long as $\alpha_k \neq 0$. If $\alpha_k = 0$, we need a different gauge; we choose gauge II such that β_k is a real positive number. Since α_k and β_k cannot vanish at the same time, we can always choose one gauge or the other, and we can divide the Brillouin zone into two zones, one with gauge I (BZ_I), one with gauge II (BZ_{II}). See Fig. 3.

On the boundary between the two zones BZ_I and BZ_{II}, $\alpha_k \neq 0$ and $\beta_k \neq 0$, and the Bloch states and Berry connection are related by the following gauge transform

$$|u_k\rangle_{\text{I}} = e^{i\chi(k)} |u_k\rangle_{\text{II}} \quad (33)$$

$$\mathcal{A}_{\text{I}} = \mathcal{A}_{\text{II}} + \nabla\chi(k). \quad (34)$$

We are now equipped to calculate the Chern number, by using Stokes' theorem separately in BZ_I and BZ_{II}. $\mathcal{B}$ denotes the oriented boundary between these two zones, which may be continuous or made of several parts.

$$C = \frac{1}{2\pi} \int_{\text{BZ}_{\text{I}}} \omega_k dk_x dk_y + \frac{1}{2\pi} \int_{\text{BZ}_{\text{II}}} \omega_k dk_x dk_y = \frac{1}{2\pi} \oint_{\mathcal{B}} (\mathcal{A}_{\text{I}} - \mathcal{A}_{\text{II}}) \cdot dl = \frac{1}{2\pi} \oint_{\mathcal{B}} \nabla\chi(k) \cdot dl \quad (35)$$

The last contour integral evaluates to $\chi(k_+^*) - \chi(k_-^*)$, where k^* is any chosen point on the contour, and the index $\pm$ corresponding to its limits when approaching from either side on the contour. It must be equal to $2\pi n$, where n is an integer, since the Bloch state $|u_k^*\rangle$ is single valued. As a result

$$C = n \quad (36)$$

and the Chern number is an integer, which accounts for the gauge discontinuities within the Brillouin zone.

3.7 Edge states

Shortly after Laughlin's explanation of the quantization of σ_{xy} in the quantum Hall effect through a charge pump [14], Halperin realized [15] that there must be a chiral mode conducting current at the edge of the system, and that this mode also explains the quantization. Halperin's argument may be generalized to all band insulators with a Chern number $C \neq 0$: there must be C one-dimensional chiral modes at the edge. It is an example of the bulk-boundary correspondence of topological insulators: the existence of a topological invariant in the bulk implies the existence of robust gapless modes at the edge. Here, we give a simple image for this argument, and highlight some consequences.

The crux of Laughlin's argument is that charges are being transferred adiabatically from one edge of the sample to the other following the insertion of a quantum of flux. This charge transfer

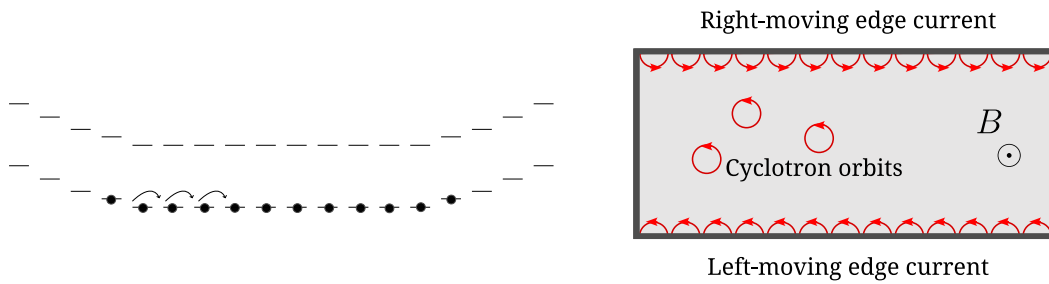


Fig. 4: *Chiral edge states in the quantum Hall effect. (Left) Effect of a confining potential on the Landau levels. The Fermi level crosses a Landau level where its group velocity is non-zero, hence the chirality of edge modes. During a Laughlin charge pump, transport occurs from one edge to the other because the inserted magnetic flux moves the Landau level states from one side to the other. (Right) Semi-classical picture of the chiral edge modes, with skipping orbits at the edge.*

implies the existence of a reservoir of charges (electrons or holes) at each edge. Moreover, the reservoirs must be gapless to ensure the adiabaticity of transport; therefore, there must be gapless edge modes in order to have $\sigma_{xy} \neq 0$.

In the quantum Hall effect, the chirality of edge modes is easily seen by considering the effect of a confining potential on the single-particle energy spectrum (see Fig. 4 left). Independent electrons in 2D in a magnetic field occupy degenerate energy levels called Landau levels, which are perfectly flat unless one introduces a confining potential. This confining potential creates a curve on either ends of the Landau levels, ensuring that the energy of electrons becomes infinite far away from the sample. Edge modes correspond to the crossing between the Fermi level and a curved Landau level: their chirality is given by the sign of the group velocity $\partial E / \partial r$ at the crossing.

The semiclassical picture of electrons in a magnetic field also provides an intuitive understanding of the origin of chiral edge modes (see Fig. 4 right). Due to the Lorentz force $-e\vec{v} \wedge \vec{B}$, electrons move in a circular fashion, with an orientation determined by the sign of the magnetic field. In the bulk of the sample, these orbits are closed, but the presence of the edge prevents the orbit from closing, forcing a pattern of skipping orbits, where the current only goes in one direction on one edge, and in the other direction on the opposite edge.

Whether in a quantum Hall system, or more generally in a Chern insulator, the edge modes are dissipationless, and protected against any local perturbation, particularly disorder. Because of their chirality, backscattering into the same channel is not allowed, and backscattering into the edge mode with the opposite chirality is exponentially suppressed by the presence of a gap in the bulk, as long as the edges are sufficiently far apart. We say that the gapless edge modes of Chern insulators are topologically protected. The existence of a dissipationless chiral mode would be impossible in a purely one-dimensional system, which is consistent with the fact that all 1D systems are Anderson localized for any disorder strength. Here, it is the presence of the 2D topological bulk, which makes it possible. One may also note that a Chern number C implies the existence of C dissipationless chiral edge modes, through the Landauer-Buttiker formalism. There could be more edge modes, but they would not be protected against disorder.

4 Chern insulator models

As we have explained in the previous section, a Chern insulator is a band insulator, such that the total Chern number of bands below the gap is non-zero. We have also defined the necessary concepts, and highlighted some very general consequences, such as a non-zero Hall conductance and the presence of robust chiral edge modes. To make things more concrete, we will now detail some explicit Chern insulator models. This will help us obtain an intuition of the microscopic ingredients that make topology possible in a model. We will provide some examples of their realizations, focusing on solid state and cold atom implementations. For a more complete review of topological phases of cold atoms, we refer to the J. Dalibard' lecture series [16].

4.1 Harper-Hofstadter model

In the previous sections, we have made many references to the quantum Hall effect, which arises for electrons confined to two dimensions in a transverse magnetic field. Our first model is a lattice version of the (continuum) quantum Hall effect, introduced by Harper in 1955 [17], and solved by Hofstadter in 1976 [18].

4.1.1 Minimal coupling to a magnetic field

We consider a tight-binding model of non-interacting particles hopping between neighboring sites of a square lattice

$$H = - \sum_{\langle i,j \rangle} t (c_i^\dagger c_j + c_j^\dagger c_i) \quad (37)$$

If we add a magnetic field, the hopping amplitude on the lattice should acquire a phase equal to the Aharonov-Bohm phase associated with the coupling of a charged particle to the electromagnetic vector potential $\vec{A}$

$$-t c_i^\dagger c_j \rightarrow -t e^{i \frac{2\pi}{\Phi_0} \int_{r_i}^{r_j} \vec{A}(\vec{r}) \cdot d\vec{r}} c_i^\dagger c_j, \quad (38)$$

where Φ_0 is a unit quantum of magnetic flux. To make the model more explicit, we choose the Landau gauge

$$\vec{A} = (0, Bx, 0). \quad (39)$$

Within this gauge choice, our tight-binding model has complex hopping amplitudes on the vertical bonds, with a phase $e^{i\Phi x}$, which ensures a total flux $\Phi = Ba^2$ through each minimal square plaquette, where a is the lattice constant (see Fig. 5).

The Hofstadter model is very rich, with an energy spectrum as a function of flux per lattice plaquette Φ , which has a fractal structure called the Hofstadter butterfly. For our purpose of introducing a topological model, we focus on the simpler case where Φ is a rational number p/q in units of the quantum of magnetic flux Φ_0 . In this case, one notices that the Hofstadter model has an enlarged unit cell compared to the original square lattice with no magnetic field. Indeed, there is a translation symmetry by q lattice sites in the x direction, and by 1 lattice site in the

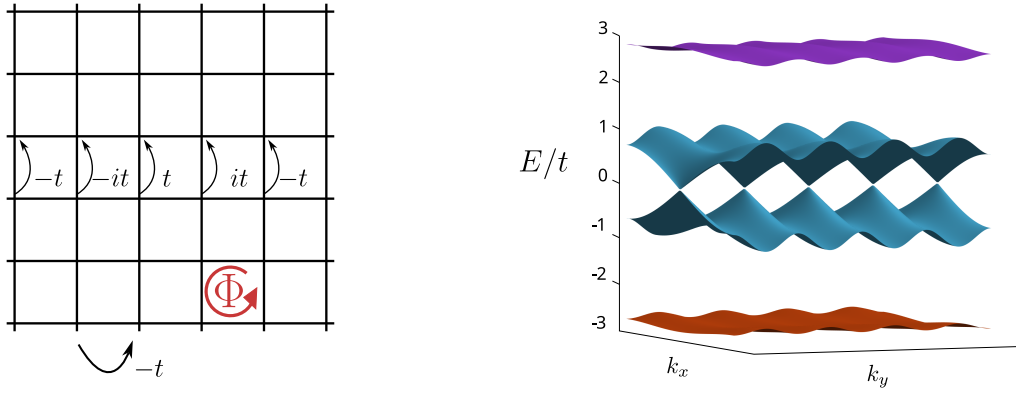


Fig. 5: The Harper-Hofstadter model on a square lattice for a flux per plaquette $\Phi = \pi/2$. (Left) Sketch of the model: a square lattice with nearest-neighbor hopping of amplitude t on horizontal bonds, and $te^{i\Phi x}$ (respectively $te^{-i\Phi x}$) on upward (respectively downward) vertical bonds. (Right) Band structure of the Harper-Hofstadter model with flux $\pi/2$ per plaquette. The lowest band is well separated from other bands by a gap, and carries a Chern number 1.

y direction. As a result, there are q bands in this model. The lowest band of the Hofstadter model carries a Chern number $C = 1$. Other bands also carry a non-zero Chern number, and the sum over all bands yields a total Chern number 0, as for any lattice model. We will not demonstrate this result, which may be obtained by applying the Chern number formula onto the Bloch wavefunctions of the lowest band of the model. Of course, this is not a surprising result: since we explicitly added a magnetic field in the model, we expect to recover the physics of the quantum Hall effect. In the continuum limit $\Phi \ll \Phi_0$, the bands are the non-dispersive Landau levels of the quantum Hall effect. For larger Φ , lattice effects are more important, and the bands carry some dispersion.

4.1.2 Physical relevance of the Hofstadter model – solid state

In typical crystalline solids, such as graphene, or the semiconductor quantum wells typically used in quantum Hall experiments, lattice effects are irrelevant, and a continuum model is perfectly sufficient. Indeed, for a lattice unit cell with a size of a few Ångström, it would take a magnetic field of 10^5 Tesla to achieve a unit quantum of flux Φ_0 per unit cell, which is several orders of magnitude beyond current possibilities.

Lattice effects may become more important in solid state systems if the unit cell is significantly enlarged, for example by creating a moiré superlattice (see Fig. 6). By stacking two atomically thin two-dimensional materials with a slight lattice constant mismatch, or two or more layers of the same 2D material with a small angle misalignment, one obtains an interference pattern of the two lattices, with a much larger periodicity than the original lattices. In these moiré heterostructures, the superlattice scale can be of order 10 nm, which allows the emergence of Hofstadter physics at accessible magnetic fields of the order of 10 Tesla. This effect was first well demonstrated in an experiment [19] where a (well aligned, or Bernal stacked) bilayer of graphene was stacked on top of a substrate of hexagonal boron nitride, a two-dimensional

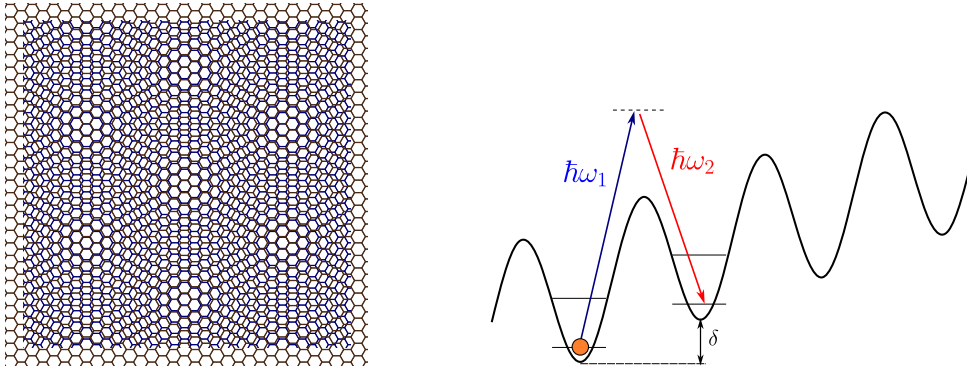


Fig. 6: Implementation of the Harper-Hofstadter model in solid state (left) and cold atoms in optical lattices (right). (Left) Moiré superlattice resulting from the alignment of a graphene and an hexagonal boron nitride (hBN) layer, which have a 10% lattice constant mismatch. There, a sizable fraction of a magnetic flux per superlattice triangular plaquette is achievable with magnetic fields of 1–10 Tesla. (Right) Laser-assisted tunneling in an optical lattice: an atom in the ground state of a given lattice site is excited to a high energy state, followed by stimulated emission to the ground state of the neighboring site. Through this process, a complex factor $e^{i(k_1 - k_2) \cdot r}$ is imprinted onto its wave function, achieving the desired Aharonov-Bohm phase.

insulator with a lattice constant very close to that of graphene. The Harper-Hofstadter model, with the square lattice replaced by a triangular lattice to fit the crystal structure of these two-dimensional solids, then describes very well the physics of moiré systems under a magnetic field.

4.1.3 Realization of the Harper-Hofstadter model in cold atoms

Tight-binding models are realized in ultracold atomic gases by using an optical lattice, i.e., an array of standing waves, which create a periodic potential for the neutral atoms. Such a set-up provides the original hopping model, without any complex amplitude corresponding to the Aharonov-Bohm effect. To simulate the effect of the magnetic field on the neutral atoms, Jaksch and Zoller [20] proposed a method of Raman-assisted tunneling, which imprints the Aharonov-Bohm phase onto specified tunneling terms, and realizes the Harper-Hofstadter model.

The basic idea of laser-assisted tunneling is summarized in Fig. 6. To imprint an artificial gauge field onto the optical lattice, one first needs to inhibit the direct (real amplitude) transition. Using the gauge choice defined in Eq. (39), we leave the periodic potential unaffected in the x direction, and inhibit transitions along the vertical bonds by detuning the potential through an energy mismatch δ between neighboring sites. The tunneling is then restored by adding two Raman lasers at resonance with this mismatch: $\delta = \hbar\omega_1 - \hbar\omega_2$. To tunnel, an atom thus follows a two-step process, summarized in Fig. 6: absorption of a photon of frequency ω_1 , followed by the stimulated emission of a photon of frequency ω_2 . The tunneling occurs with a phase, which corresponds to the wavevector mismatch between the two lasers: $e^{i(\vec{k}_1 - \vec{k}_2) \cdot \vec{r}}$. This method was implemented by two independent groups in 2013, with flux per plaquette $\Phi/\Phi_0 = \pi/2$ [21] and π [22], respectively.

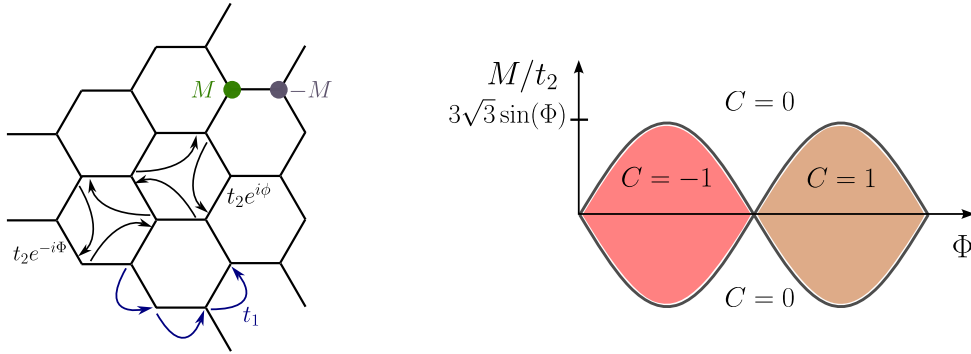


Fig. 7: Haldane model of a Chern insulator [23]. (Left) The model consists of a honeycomb lattice with nearest-neighbor hopping of real amplitude t_1 , second-neighbor hopping of amplitude $t_2 e^{i\Phi}$, and a staggered on-site mass M for sites A and B . (Right) Phase diagram of the Haldane model. Depending on the values of the parameter M/t_2 , the lowest band has Chern number 0, 1, or -1 . The Chern number is opposite in the other band.

4.2 Haldane model

The Harper-Hofstadter model described in the previous section is simply a lattice model in a magnetic field. In 1988, Haldane [23] realized that there is no need for a net field to obtain a non-zero Chern number, and hence a quantized Hall effect. He developed a model based on a hexagonal lattice, with a staggered flux, such that the overall magnetic flux per unit cell is zero. Beyond the absence of a magnetic field, Haldane showed that there are distinct ways to open a gap in a gapless system, which respectively result in trivial or topological bands. We now go into the details of the model and its different terms, and discuss its relevance to solid state systems, and its implementation in cold atoms in optical lattices. Fig. 7 summarizes the model.

4.2.1 Graphene tight-binding model, and different gap opening terms

We start from graphene's tight-binding model, which consists of nearest-neighbor hopping on a honeycomb lattice. There are two inequivalent lattice sites A and B , thus we can write the Hamiltonian in the following form

$$H = -t_1 \sum_i c_A^\dagger(\vec{r}_i) c_B(\vec{r}_i) + c_A^\dagger(\vec{r}_i) c_B(\vec{r}_i + \vec{a}_1) + c_A^\dagger(\vec{r}_i) c_B(\vec{r}_i + \vec{a}_2) + h.c. \quad (40)$$

where $\vec{a}_1$ and $\vec{a}_2$ are the Bravais lattice vectors of the triangular lattice. Going into momentum space, one may rewrite the Hamiltonian in Bloch form

$$H = \sum_{\vec{k}} \begin{pmatrix} A_{\vec{k}} & B_{\vec{k}} \end{pmatrix} h(\vec{k}) \begin{pmatrix} A_{\vec{k}} \\ B_{\vec{k}} \end{pmatrix}. \quad (41)$$

where the Bloch Hamiltonian is a 2×2 matrix, and writes

$$h(\vec{k}) = -t_1 \begin{pmatrix} 0 & 1 + e^{-i\vec{k} \cdot \vec{a}_1} + e^{-i\vec{k} \cdot \vec{a}_2} \\ 1 + e^{i\vec{k} \cdot \vec{a}_1} + e^{i\vec{k} \cdot \vec{a}_2} & 0 \end{pmatrix}. \quad (42)$$

Diagonalizing this Hamiltonian yields the two energy bands of graphene

$$E_{\pm}(\vec{k}) = \pm t_1 \sqrt{\left(1 + \cos(\vec{k} \cdot \vec{a}_1) + \cos(\vec{k} \cdot \vec{a}_2)\right)^2 + \left(\sin(\vec{k} \cdot \vec{a}_1) + \sin(\vec{k} \cdot \vec{a}_2)\right)^2}. \quad (43)$$

The two bands touch at the two inequivalent corners of the hexagonal Brillouin zone, $K^{\pm}$. In the vicinity of these points, the dispersion relation is linear, describing relativistic particles, and the Hamiltonian takes the Dirac form

$$h(q = \vec{k} - \vec{K}^{\pm}) = \frac{3}{2} t_1 a (\pm q_x \sigma_x + q_y \sigma_y) \quad (44)$$

where $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices.

The Dirac crossings in the honeycomb lattice model are protected by the product of two symmetry operators, inversion $\mathcal{C}_2$ and time-reversal symmetry $\mathcal{T}$. This means that no gap opening can occur, as long as these symmetries are preserved. Let us examine the effect of a gap opening via breaking of $\mathcal{C}_2$ symmetry. This can be done by introducing an on-site mass term, which gives a different energy to sites A and B . The extra term in the Bloch Hamiltonian writes

$$M \sigma_z \quad (45)$$

and the new energy spectrum is

$$E_{\pm}(\vec{k}) = \sqrt{M^2 \pm t_1^2 \left(\left(1 + \cos(\vec{k} \cdot \vec{a}_1) + \cos(\vec{k} \cdot \vec{a}_2)\right)^2 + \left(\sin(\vec{k} \cdot \vec{a}_1) + \sin(\vec{k} \cdot \vec{a}_2)\right)^2 \right)} \quad (46)$$

For this new band structure, the only gap closing occurs at $M = 0$. Note that taking the limit $M \rightarrow \pm\infty$ yields two decoupled copies of a triangular model with one site per unit cell. Such a model is necessarily topologically trivial, since there is only one band.³ As a result, adding a staggered mass in the Haldane model results in two topologically trivial $C = 0$ bands.

We now add a different term to the honeycomb lattice model, in the form of a next-nearest-neighbor hopping of amplitude $t_2 e^{i\Phi}$ (see Fig. 7). This term breaks time-reversal symmetry, since a time-reversal conjugation transforms $e^{i\Phi}$ into $e^{-i\Phi}$, but unlike the Harper-Hofstadter model, the unit cell is unchanged, and the total flux per hexagon is zero. This adds a diagonal term to the Bloch Hamiltonian, of the form

$$2t_2 \sin(\Phi) (-\sin(ka_1) + \sin(ka_2) - \sin(ka_1 + ka_2)) \sigma_z \quad (47)$$

$$+ 2t_2 \cos(\Phi) (\cos(ka_1) + \cos(ka_2) + \cos(ka_1 + ka_2)) \mathbb{1} \quad (48)$$

This term also opens a gap in the spectrum of the graphene model, but unlike the mass term, it has opposite signs at the K^+ and K^- points. This allows the bands to acquire a non-zero Chern number whenever the time-reversal breaking term dominates over the mass term, i.e., $|M/t_2| < 3\sqrt{3} \sin \Phi$, as can be verified through a direct calculation of the Chern number. See Fig. 7 for a phase diagram of the lowest band of the Haldane model.

³To convince ourselves that a one band model necessarily has Chern number 0, one can recall the obstruction to defining a smooth gauge in the entire Brillouin zone: the one-dimensional wavefunction cannot vanish anywhere in the entire Brillouin zone because it must remain normalized.

4.2.2 Physical relevance of the Haldane model in the solid state – spin-orbit coupling

In a seminal paper [3], Kane and Mele demonstrated that the phase-carrying hopping terms of the Haldane model could actually be physically relevant to graphene, when introducing the appropriate spin-orbit coupling term. Yet, graphene preserves time-reversal symmetry, and the Haldane model does not. This issue is resolved by taking two time-reversed copies of the graphene's honeycomb model, one for each spin. Kane and Mele showed that spin-orbit coupling introduces a complex tunneling next-nearest neighbor term, with opposite phases for each spin, such that time-reversal symmetry is preserved. This realizes a model for what is known as the quantum spin Hall effect, a topological insulator, which can be thought of as a product of two Chern insulators with opposite Chern numbers. It turns out that the actual spin-orbit coupling term is much too weak to produce a measurable gap in graphene, and that physical realizations of topology have to be searched for elsewhere.

The first quantum spin Hall insulator was realized in quantum wells of HgTe / HgCdTe [24]. However, no time-reversal symmetry breaking (or spin polarization) was achieved there. The first Chern insulator was realized [25–27] in thin films of three-dimensional topological insulators, doped with a magnetic element, following a theoretical prediction [28]. The 3D topological insulators (e.g. $(\text{Bi,Sb})_2\text{Te}_3$) are materials with a large spin-orbit coupling, which plays the role of creating non-trivial topology in the band structure. The magnetic doping element (e.g. Cr or Fe) breaks time-reversal symmetry, by allowing the spin polarization of the ground state.

4.2.3 Realization of the Haldane model in cold atoms

The Haldane model was realized using ultracold fermions in an optical lattice [29]. The time-reversal breaking next-nearest neighbor hopping terms are implemented through a circular modulation of the lattice positions, i.e., by ‘shaking’ the lattice in a circular motion. Similarly to the laser-assisted tunneling process described in the realization of the Harper-Hofstadter model, the periodic modulation of the lattice may be treated using Floquet theory, in order to obtain the effective Hamiltonian, which is here the Haldane model.

4.3 Chern number and spontaneous symmetry breaking in moiré materials

The latest observations of a solid state Chern insulator occurred in two-dimensional moiré materials, namely twisted bilayer graphene aligned with hexagonal boron nitride [30], twisted bilayer transition metal dichalcogenites (MoTe_2) [31], as well as other graphene systems aligned with hexagonal boron nitride. In these materials, time-reversal symmetry is broken spontaneously, and not explicitly, as a result of strong interactions, while the topology has a different origin depending on the specific moiré material.

Integer quantum Hall states at integer filling are spin polarized. While this makes sense from the point of view of single-particle physics, since the Zeeman interaction lifts the degeneracy between spin up and down within a Landau level, the main contribution to this ferromagnetism

comes from electronic interactions. Essentially, a spin-polarized wavefunction must be spatially antisymmetric, which lowers its Coulomb interaction compared to other wavefunctions, through a process called exchange. A similar process can occur in any flat band, i.e., a band with a low enough dispersion, such that interactions dominate over kinetic energy, as long as the single-particle states of the flat band have a sufficient overlap with one another (i.e. the flat band is the result of a quantum interference between different states, rather than orbitals localized on individual sites). Moiré materials are an example of such material, where the band can be flattened, for example by tuning the rotation angle between two layers to a ‘magic angle’ condition.

Twisted bilayer graphene consists of two layers of graphene, misaligned with an angle θ . A superlattice (or moiré) potential results from this misalignment, which folds the bands of graphene onto themselves, inside a mini Brillouin zone. At each band crossing (resulting either from this folding, or from two bands from two different layers), there is a gap opening due to level repulsion and interlayer tunneling. If we focus around the middle of the spectrum, which is relevant at low energy because this is typically where the Fermi energy lies, there are the Dirac cones originating from each graphene layer. Similarly to graphene, the Dirac cones are protected by the product of inversion $\mathcal{C}_2$ and time-reversal symmetry $\mathcal{T}$. Similarly to our discussion of the Haldane model, this gives us different ways to open a gap. Yet, twisted bilayer graphene presents a very important difference compared to graphene: the chirality of the Dirac cones is the same at opposite corners of the Brillouin zones for a given valley, with the other valley restoring time-reversal symmetry by having the opposite chirality. In contrast, the chirality of the Dirac cones is opposite for the opposite corners of the Brillouin zone in graphene. This peculiar configuration, which is only observed through the interference of two graphene layers, leads to the fact that a simple staggered mass term (equivalent to that of the Haldane model) may in fact open a topological gap in twisted bilayer graphene’s band structure [32], i.e., lead to a model with two bands with opposite Chern numbers. Experimentally, the $\mathcal{C}_2$ breaking is provided through the alignment of one of the graphene layers with its substrate of hexagonal boron nitride, which, at leading order, creates a staggered mass term Δ_{AB} .

The band structure of twisted bilayer graphene aligned with hexagonal boron nitride thus involves two almost flat bands, separated by a small gap of order Δ_{AB} , which have Chern numbers $C = \pm 1$. These bands have a fourfold degeneracy, due to the spin and valley degrees of freedom. Chern numbers are equal in opposite spin copies, but opposite in opposite valleys, since the two valleys are connected by time-reversal conjugation. Combining the topology with the flatness of the bands leads to the valley and spin polarization of the system at integer fillings of either band through quantum Hall ferromagnetism, as explained above. Experimentally, one observes a quantized Hall conductance, even at zero magnetic field [30]. In twisted bilayer MoTe₂, there are no Dirac cones to start with, so the origin of topology is different, and rather comes from the non-trivial coupling between the two layers, which introduces a winding of the Bloch states [33]. Likewise, there is a spontaneous spin-valley polarization due to quantum Hall ferromagnetism, thanks to the flatness of the band if the twist angle is tuned properly [31].

References

- [1] K. v. Klitzing, G. Dorda, and M. Pepper, *Phys. Rev. Lett.* **45**, 494 (1980)
- [2] K.S. Novoselov, Z. Jiang, Y. Zhang, S.V. Morozov, H.L. Stormer, U. Zeitler, J.C. Maan, G.S. Boebinger, P. Kim, and A.K. Geim, *Science* **315**, 1379 (2007)
- [3] C.L. Kane and E.J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005)
- [4] A. Altland and M.R. Zirnbauer, *Phys. Rev. B* **55**, 1142 (1997)
- [5] M.J. Moessner R.: *Topological Phases of Matter* (Cambridge University Press, 2021)
- [6] T.H.B.A. Bernevig: *Topological Insulators and Topological Superconductors* (Princeton University Press, 2013)
- [7] C. Nayak, S.H. Simon, A. Stern, M. Freedman, and S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083 (2008)
- [8] M.V. Berry, *Proc. Roy. Soc. A* **392**, 45 (1984)
- [9] R. Bianco and R. Resta, *Phys. Rev. B* **84**, 241106(R) (2011)
- [10] Q. Niu, D.J. Thouless, and Y.-S. Wu, *Phys. Rev. B* **31**, 3372 (1985)
- [11] M.B. Hastings and S. Michalakis, *Commun. Math. Phys.* **334**, 433 (2014)
- [12] T. Fukui, Y. Hatsugai, and H. Suzuki, *J. Phys. Soc. Jpn.* **74**, 1674 (2005)
- [13] D.J. Thouless, M. Kohmoto, M.P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982)
- [14] R.B. Laughlin, *Phys. Rev. B* **23**, 5632(R) (1981)
- [15] B.I. Halperin, *Phys. Rev. B* **25**, 2185 (1982)
- [16] J. Dalibard, Collège de France Lecture Series (2017–2018)
- [17] P.G. Harper, *Proc. Phys. Soc. A* **68**, 874 (1955)
- [18] D.R. Hofstadter, *Phys. Rev. B* **14**, 2239 (1976)
- [19] C.R. Dean, L. Wang, P. Maher, C. Forsythe, F. Ghahari, Y. Gao, J. Katoch, M. Ishigami, P. Moon, M. Koshino, T. Taniguchi, K. Watanabe, K.L. Shepard, J. Hone, and P. Kim, *Nature* **497**, 598 (2013)
- [20] D. Jaksch and P. Zoller, *New J. Phys.* **5**, 56 (2003)
- [21] M. Aidelsburger, M. Atala, M. Lohse, J.T. Barreiro, B. Paredes, and I. Bloch, *Phys. Rev. Lett.* **111**, 185301 (2013)

- [22] H. Miyake, G.A. Siviloglou, C.J. Kennedy, W.C. Burton, and W. Ketterle, *Phys. Rev. Lett.* **111**, 185302 (2013)
- [23] F.D.M. Haldane, *Phys. Rev. Lett.* **61**, 2015 (1988)
- [24] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L.W. Molenkamp, X.-L. Qi, and S.-C. Zhang, *Science* **318**, 766 (2007)
- [25] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang, Z.-Q. Ji, Y. Feng, S. Ji, X. Chen, J. Jia, X. Dai, Z. Fang, S.-C. Zhang, K. He, Y. Wang, L. Lu, X.-C. Ma, and Q.-K. Xue, *Science* **340**, 167 (2013)
- [26] J.G. Checkelsky, R. Yoshimi, A. Tsukazaki, K.S. Takahashi, Y. Kozuka, J. Falson, M. Kawasaki, and Y. Tokura, *Nat. Phys.* **10**, 731 (2014)
- [27] X. Kou, S.-T. Guo, Y. Fan, L. Pan, M. Lang, Y. Jiang, Q. Shao, T. Nie, K. Murata, J. Tang, Y. Wang, L. He, T.-K. Lee, W.-L. Lee, and K.L. Wang, *Phys. Rev. Lett.* **113**, 137201 (2014)
- [28] R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai, and Z. Fang, *Science* **329**, 61 (2010)
- [29] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger, *Nature* **515**, 237 (2014)
- [30] M. Serlin, C.L. Tschirhart, H. Polshyn, Y. Zhang, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, and A.F. Young, *Science* **367**, 900 (2020)
- [31] F. Xu, Z. Sun, T. Jia, C. Liu, C. Xu, C. Li, Y. Gu, K. Watanabe, T. Taniguchi, B. Tong, J. Jia, Z. Shi, S. Jiang, Y. Zhang, X. Liu, and T. Li, *Phys. Rev. X* **13** (2023)
- [32] N. Bultinck, S. Chatterjee, and M.P. Zaletel, *Phys. Rev. Lett.* **124** (2020)
- [33] F. Wu, T. Lovorn, E. Tutuc, I. Martin, and A.H. MacDonald, *Phys. Rev. Lett.* **122**, 086402 (2019)

9 Ginzburg–Landau Theory of Superconductivity

Carsten Timm

Institute of Theoretical Physics

TUD Dresden University of Technology

Contents

1	Introduction	2
2	Landau theory of phase transitions	2
3	Ginzburg–Landau theory for neutral superfluids	5
3.1	Fluctuations for $T > T_c$	8
3.2	Fluctuations for $T < T_c$	9
3.3	Time-dependent Ginzburg–Landau theory	10
4	Ginzburg–Landau theory for superconductors	11
4.1	Type-I and type-II superconductors	14
4.2	The Anderson–Higgs mechanism	17
4.3	Ginzburg–Landau theory as an effective field theory	21
4.4	Ginzburg–Landau theory for nontrivial symmetries	22

1 Introduction

Historically, the first major step in the theoretical description of superconductivity was taken by the brothers F. and H. London [1]. They proposed a theory for the electromagnetic properties of superconductors—the ideal conductivity and the Meißner–Ochsenfeld effect—based on a two-fluid picture. Specifically, they proposed that part of the electrons in a superconductor behave like a dissipationless superfluid, whereas the rest behave like a normal fluid showing Ohmic dissipation. Conceptually, London theory augments the Maxwell equations of electromagnetism by constitutive equations describing superconducting materials. The density of the superfluid, n_s , is treated as given and there is no way to understand the dependence of n_s on, for example, temperature or magnetic field. Moreover, n_s is assumed to be constant in time and uniform in space—an assumption that is expected to fail close to the surface of a superconductor.

These deficiencies were cured by the Ginzburg–Landau theory [2] put forward in 1950. The starting point is the much more general Landau theory of phase transitions [3]. Both these theories are phenomenological in the sense that they do not rely on a particular microscopic mechanism but only on quite general symmetry principles.

The next large step was the microscopic mean-field theory of superconductivity developed by Bardeen, Cooper, and Schrieffer (BCS) [4] in 1957. Soon thereafter, Gor’kov [5] derived Ginzburg–Landau theory from BCS theory and thus clarified the microscopic interpretation of the quantities appearing in Ginzburg–Landau theory. The relation between Ginzburg–Landau theory and BCS theory is somewhat similar to the one between phenomenological thermodynamics and statistical physics. In particular, the development of the latter does not diminish the usefulness of the former. For example, predictions derived in the framework of Ginzburg–Landau theory do not suffer from lack of knowledge of the underlying microscopic mechanisms.

2 Landau theory of phase transitions

Landau [3] introduced the concept of the *order parameter* to describe phase transitions. In this context, an order parameter is a thermodynamic variable that is zero on one side of the transition and nonzero on the other. The theory neglects fluctuations, which means that the order parameter is assumed to be constant in space and time. The appropriate thermodynamic potential is written as a function of the order parameter Δ and certain other thermodynamic variables such as pressure or volume, magnetic field, etc. We will always call the potential the *free energy* F , but whether it really is a free energy, a free enthalpy, or something else depends on which quantities are given (pressure vs. volume etc.). Hence, we write $F = F(\Delta, T)$, where T is the temperature, and further variables have been suppressed. The equilibrium state at temperature T is the ensemble state that minimizes the free energy $F(\Delta, T)$.

Landau’s idea was to expand F in terms of Δ , including only those terms that are allowed by the symmetry of the system and keeping the minimum number of the lowest-order terms required to get nontrivial results. The strength of Landau theory is that we do not need to know the underlying microscopic description, e.g., the Hamiltonian, but only the character of the order parameter and the symmetries of the system.

As an example, we consider the ferromagnetic Heisenberg model. The order parameter is the magnetization $\mathbf{M} \in \mathbb{R}^3$. The underlying Hamiltonian and thus also the free energy F are invariant under all rotations of $\mathbf{M}$. This is a $\text{SO}(3)$ symmetry since $\text{SO}(3)$ is the group of all proper rotation matrices in three-dimensional space. Furthermore, it is invariant under time reversal, which flips the magnetization, $\mathbf{M} \mapsto -\mathbf{M}$. This time-reversal symmetry is redundant in this case but not in general.

Following Landau, we assume F to be a smooth function at $\mathbf{M} = 0$ so that we can expand F in a Taylor series in $\mathbf{M}$ about $\mathbf{M} = 0$, with nonzero radius of convergence. The leading terms are

$$F = F_0 + \alpha \mathbf{M} \cdot \mathbf{M} + \frac{\beta}{2} (\mathbf{M} \cdot \mathbf{M})^2 + \mathcal{O}((\mathbf{M} \cdot \mathbf{M})^3), \quad (1)$$

where $\mathcal{O}((\mathbf{M} \cdot \mathbf{M})^3)$ represents terms of higher order, which we suppress from now on. It is by no means necessary to truncate the expansion at this order but this truncation is sufficient to describe second-order phase transitions. We must include enough terms to make F bounded from below. For example, the above truncation would be invalid for $\beta < 0$. Including an explicit factor of $1/2$ in the β term is a convention. F_0 , α , and β are functions of temperature (and possibly pressure etc.).

The zero-order term F_0 is the free energy of the nonmagnetic state. While it may have a strong temperature dependence, it is irrelevant for the magnetic phase transition. One therefore commonly drops this term, which means that F is taken to be the *difference* of the free energies in the magnetic and the nonmagnetic states. We will follow this practice.

What is the corresponding expansion for a superfluid or superconductor? Lacking a microscopic theory but suspecting an analogy to the well-understood Bose–Einstein condensation, Ginzburg and Landau [2] assumed that a superfluid is described by a single one-particle wave function $\Psi_s(\mathbf{r})$. They imposed the plausible normalization

$$\int d^3r |\Psi_s(\mathbf{r})|^2 = N_s = n_s V, \quad (2)$$

where N_s is the total number of particles in the condensate and V is the volume of the system. They then chose the complex amplitude ψ of $\Psi_s(\mathbf{r})$ as the order parameter, with the normalization $|\psi|^2 \propto n_s$. They thereby neglected the spatial variation of $\Psi_s(\mathbf{r})$ on an atomic scale. The order parameter is a complex number.

The free energy must not depend on the global phase of ψ since the global phase of quantum states is not observable. The relevant symmetry is thus a global $\text{U}(1)$ symmetry under phase change. ($\text{U}(1)$ is the group of 1×1 unitary matrices, i.e., numbers on the unit circle in the complex plane.) Thus we obtain the expansion

$$F = \alpha \psi^* \psi + \frac{\beta}{2} (\psi^* \psi)^2 = \alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4. \quad (3)$$

Odd powers are excluded since they are not differentiable at $\psi = 0$. If $\beta > 0$, which is not guaranteed by symmetry but is true for superconductors and superfluids, $F(\psi)$ is bounded from below. Now there are two cases, which are illustrated in Fig. 1:

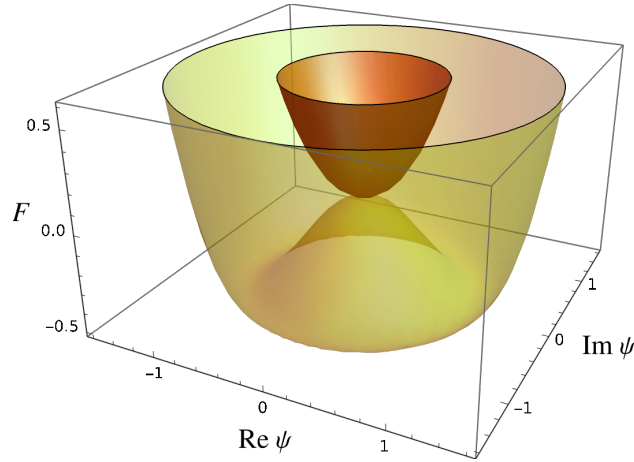


Fig. 1: Landau free energy $F(\psi)$ for a superfluid or superconductor with $\alpha/\beta = 1$ (orange) and $\alpha/\beta = -1$ (yellow).

- For $\alpha \geq 0$, $F(\psi)$ has a single minimum at $\psi = 0$. Thus the equilibrium state has $n_s = 0$. This is clearly a normal fluid or normal metal.
- For $\alpha < 0$, $F(\psi)$ has a ring of minima with equal modulus $|\psi|$ but arbitrary phase. This shape is often called the wine-bottle or Mexican-hat potential. We easily see that

$$\frac{\partial F}{\partial |\psi|} = 2\alpha |\psi| + 2\beta |\psi|^3 \stackrel{!}{=} 0 \quad (4)$$

has the solutions $|\psi| = 0$, which describes a maximum, and

$$|\psi| = \sqrt{-\frac{\alpha}{\beta}}, \quad (5)$$

which describes the ring of minima. Note that the radicand is positive.

In Landau theory, the phase transition clearly occurs when $\alpha = 0$. Since then $T = T_c$ by definition, it is useful to expand α and β to leading order in T around T_c . Hence,

$$\alpha \cong \alpha' (T - T_c), \quad \alpha' > 0, \quad (6)$$

$$\beta \cong \text{const.} \quad (7)$$

Then the order parameter satisfies

$$|\psi| \cong \sqrt{-\frac{\alpha' (T - T_c)}{\beta}} = \sqrt{\frac{\alpha'}{\beta}} \sqrt{T_c - T} \quad \text{for } T \leq T_c. \quad (8)$$

The scaling $|\psi| \propto (T_c - T)^{1/2}$ with exponent $1/2$ is characteristic for mean-field theories. Mean-field theories do not predict the true value of the critical exponent since they neglect fluctuations. However, this is not a dramatic problem for superconductors in three dimensions since deviations from mean-field behavior are limited to a very narrow temperature interval around T_c and are typically unobservable [6].

All solutions with this value of $|\psi|$ minimize the free energy. In a given experiment, only one of them is realized. This specific equilibrium state has an order parameter $\psi = |\psi| e^{i\phi}$ with some fixed phase ϕ and is thus not invariant under changes of the phase. We say that the global U(1) symmetry is *spontaneously broken* since the free energy F has it but the specific equilibrium state does not.

The expansion of F up to fourth order is only justified as long as ψ is small. Its quantitative validity is thus limited to temperatures not too far below T_c . However, one can often gain insight into the qualitative behavior of a system by applying Landau theory outside of this range.

Note that the true thermodynamic free energy as a function of given variables (here temperature) is F evaluated at the equilibrium solution for ψ . From this, we can calculate further thermodynamic variables. In particular, the entropy is $S_s = -\partial F/\partial T$. Since the expression for F above only includes the contributions of the superfluid the entropy calculated from it will also only contain these contributions, which is here indicated by the subscript “s”. The missing contribution from the normal state is generally large but smooth at T_c . For $T \geq T_c$, the free energy is $F(\psi=0) = 0$ and thus we find $S_s = 0$. For $T < T_c$ we instead obtain

$$\begin{aligned} S_s &= -\frac{\partial}{\partial T} F \left(\sqrt{-\frac{\alpha}{\beta}} \right) = -\frac{\partial}{\partial T} \left(-\frac{\alpha^2}{\beta} + \frac{1}{2} \frac{\alpha^2}{\beta} \right) = \frac{\partial}{\partial T} \frac{\alpha^2}{2\beta} \\ &\cong \frac{\partial}{\partial T} \frac{(\alpha')^2}{2\beta} (T-T_c)^2 = \frac{(\alpha')^2}{\beta} (T-T_c) = -\frac{(\alpha')^2}{\beta} (T_c-T) < 0. \end{aligned} \quad (9)$$

The entropy is continuous at $T = T_c$. The corresponding heat-capacity contribution is

$$C_s = T \frac{\partial S_s}{\partial T} = \begin{cases} \frac{(\alpha')^2}{\beta} T & \text{for } T < T_c, \\ 0 & \text{for } T \geq T_c. \end{cases} \quad (10)$$

Thus, the heat capacity has a jump discontinuity of

$$\Delta C_s = -\frac{(\alpha')^2}{\beta} T_c \quad (11)$$

at T_c . The phase transition is of *second* order since a second derivative of the thermodynamic potential F with respect to its natural variable T is the lowest discontinuous one.

3 Ginzburg–Landau theory for neutral superfluids

To be able to describe spatially nonuniform situations, Ginzburg and Landau had to go beyond the Landau description for a constant order parameter [2]. To do so, they included gradient terms in the free energy. We will first discuss the case of a superfluid of electrically neutral particles (think of ^4He). We define a macroscopic condensate wave function $\psi(\mathbf{r})$, which is essentially given by $\Psi_s(\mathbf{r})$ averaged over length scales large compared to atomic distances. In the spirit of Landau theory, Ginzburg and Landau included the simplest term containing gradients of ψ and allowed by symmetry into the free energy:

$$F[\psi] = \int d^3r \left[\alpha |\psi(\mathbf{r})|^2 + \frac{\beta}{2} |\psi(\mathbf{r})|^4 + \gamma (\nabla \psi(\mathbf{r}))^* \cdot \nabla \psi(\mathbf{r}) \right], \quad (12)$$

where α and β are now defined as densities. We expect spatial changes of $\psi(\mathbf{r})$ to cost energy—this is analogous to the energy of domain walls in magnetic systems—which implies that $\gamma > 0$. The free energy is now a *functional* of ψ , also called the Ginzburg–Landau functional. Functionals are usually written with their argument (a function) in square brackets.

If we interpret $\psi(\mathbf{r})$ as a coarse-grained condensate wave function it is natural to identify the gradient term as a kinetic energy by writing

$$F[\psi] = \int d^3r \left[\alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m^*} \left| \frac{\hbar}{i} \nabla \psi \right|^2 \right], \quad (13)$$

where m^* is an effective mass and $\gamma = \hbar^2/2m^*$. We have suppressed the spatial argument for ease of writing.

The free energy is minimal in equilibrium for the canonical ensemble. So, in principle, we have to find the distribution of microstates that minimizes F . Following Ginzburg and Landau, we use a mean-field approximation, where this distribution is described only in terms of the spatially dependent order-parameter field $\psi(\mathbf{r})$. Again, $F[\psi]$ is the true free energy in the thermodynamic sense only for the equilibrium solution for ψ . From $F[\psi]$, we can derive a differential equation for the function $\psi(\mathbf{r})$ that minimizes F . The derivation is very similar to the derivation of the Lagrange equation of the second kind from Hamilton’s principle known from classical mechanics. We write the condition in terms of a functional derivative,

$$\frac{\delta F}{\delta \psi^*(\mathbf{r})} = 0 \quad (14)$$

and recall that ψ and ψ^* have to be treated as independent since $\text{Re } \psi$ and $\text{Im } \psi$ are independent. We perform an integration by parts for the kinetic-energy term, ignore an irrelevant boundary term, and obtain

$$\begin{aligned} \frac{\delta F}{\delta \psi^*(\mathbf{r})} &= \frac{\delta}{\delta \psi^*(\mathbf{r})} \int d^3r' \left[\alpha \psi^*(\mathbf{r}') \psi(\mathbf{r}') + \frac{\beta}{2} [\psi^*(\mathbf{r}')]^2 \psi^2(\mathbf{r}') - \frac{\hbar^2}{2m^*} \psi^*(\mathbf{r}') [\nabla']^2 \psi(\mathbf{r}') \right] \\ &= \int d^3r' \left[\alpha \delta(\mathbf{r}' - \mathbf{r}) \psi(\mathbf{r}') + \beta \delta(\mathbf{r}' - \mathbf{r}) \psi^*(\mathbf{r}') \psi^2(\mathbf{r}') - \frac{\hbar^2}{2m^*} \delta(\mathbf{r}' - \mathbf{r}) [\nabla']^2 \psi(\mathbf{r}') \right] \\ &= \alpha \psi(\mathbf{r}) + \beta \psi^*(\mathbf{r}) \psi^2(\mathbf{r}) - \frac{\hbar^2}{2m^*} \nabla^2 \psi(\mathbf{r}) = 0. \end{aligned} \quad (15)$$

Rearranging terms and suppressing the spatial argument yields

$$-\frac{\hbar^2}{2m^*} \nabla^2 \psi + \alpha \psi + \beta |\psi|^2 \psi = 0. \quad (16)$$

This equation is very similar to the time-independent Schrödinger equation but, interestingly, contains a nonlinear term. It is also called the (time-independent) Gross–Pitaevskii equation, in particular when derived from a microscopic description of an interacting superfluid [7, 8]. The nonlinear term describes a self-interaction of the condensate.

As an example, we apply this equation to find the variation of $\psi(\mathbf{r})$ close to the surface of a superfluid filling the half space $x > 0$. We impose the boundary condition $\psi(x=0) = 0$ and

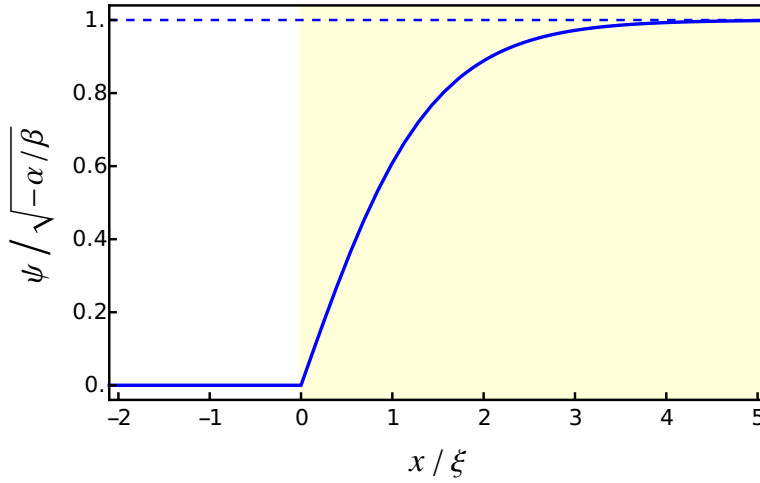


Fig. 2: Superfluid order parameter ψ as a function of spatial coordinate x , from Ginzburg–Landau theory. The superfluid fills the half space $x > 0$.

assume that the solution only depends on x . Then

$$-\frac{\hbar^2}{2m^*} \psi''(x) + \alpha \psi(x) + \beta |\psi(x)|^2 \psi(x) = 0. \quad (17)$$

For $x \rightarrow \infty$, we should obtain the uniform Landau solution

$$\lim_{x \rightarrow \infty} \psi(x) = \sqrt{-\frac{\alpha}{\beta}}. \quad (18)$$

Writing

$$\psi(x) = \sqrt{-\frac{\alpha}{\beta}} f(x), \quad (19)$$

we obtain

$$\underbrace{-\frac{\hbar^2}{2m^*} f''(x)}_{>0} + f(x) - f(x)^3 = \xi^2 f''(x) + f(x) - f(x)^3 = 0. \quad (20)$$

This equation contains a characteristic length scale, the *Ginzburg–Landau coherence length*

$$\xi \equiv \sqrt{-\frac{\hbar^2}{2m^* \alpha}} \cong \sqrt{\frac{\hbar^2}{2m^* \alpha' (T_c - T)}}. \quad (21)$$

It has a strong temperature dependence and actually diverges for $T \rightarrow T_c^-$. Although it is nonlinear, Eq. (20) can be solved analytically. It is easy to check that

$$f(x) = \tanh \frac{x}{\sqrt{2} \xi} \quad (22)$$

is a solution satisfying the boundary conditions at $x = 0$ and $x \rightarrow \infty$. The solution is sketched in Fig. 2.

3.1 Fluctuations for $T > T_c$

So far, we have only considered the state ψ_0 that minimizes the functional $F[\psi]$. At nonzero temperatures, the system will fluctuate about ψ_0 . For a uniform system we have

$$\psi(\mathbf{r}) = \psi_0 + \delta\psi(\mathbf{r}), \quad (23)$$

with $\psi_0 = \text{const.}$ We consider the cases $T > T_c$ and $T < T_c$ separately.

For $T > T_c$, the mean-field solution is $\psi_0 = 0$ so that $\psi(\mathbf{r}) = \delta\psi(\mathbf{r})$. We now write the partition function Z as a sum of Boltzmann factors containing the Ginzburg–Landau functional of all possible states,

$$Z = \int D^2\psi e^{-F[\psi]/k_B T}. \quad (24)$$

This equation requires some discussion. We are obviously using $F[\psi]$ as if it were the Hamiltonian of the field $\psi(\mathbf{r})$, although we have introduced it as a generalized free energy. Recall that $\psi(\mathbf{r})$ is the condensate wave function, averaged over atomic length scales. More precisely, fluctuations of the wave function on shorter length scales have been integrated out so that $F[\psi]$ can be understood as a thermodynamic free energy with respect to these small-scale fluctuations but it is the *energy* for the longer-scale fluctuations we are considering here. This has been explained nicely by Ma in Ref. [9].

Equation (24) contains a functional integral over infinitely many complex variables, namely the values of $\psi(\mathbf{r}) \in \mathbb{C}$ for all $\mathbf{r}$. The mathematical details go beyond the scope of this lecture but are not essential to understand the physics. A good introduction can be found in Ref. [10]. The integral is difficult to evaluate. In the common *Gaussian approximation*, we restrict $F[\psi]$ to second order in $\delta\psi = \psi$. This allows to evaluate Z by Fourier transformation: We start from

$$F[\psi] \cong \int d^3r \left[\alpha \psi^*(\mathbf{r})\psi(\mathbf{r}) + \frac{\hbar^2}{2m^*} \left(\frac{1}{i} \nabla\psi(\mathbf{r}) \right)^* \cdot \frac{1}{i} \nabla\psi(\mathbf{r}) \right] \quad (25)$$

and insert

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \psi_{\mathbf{k}}, \quad (26)$$

with the large volume V of the system, which gives

$$F[\psi] \cong \frac{1}{V} \sum_{\mathbf{k}\mathbf{k}'} \underbrace{\int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}+i\mathbf{k}'\cdot\mathbf{r}}}_{V\delta_{\mathbf{k}\mathbf{k}'}} \left[\alpha \psi_{\mathbf{k}}^* \psi_{\mathbf{k}'} + \frac{\hbar^2}{2m^*} (\mathbf{k}\psi_{\mathbf{k}})^* \cdot \mathbf{k}'\psi_{\mathbf{k}'} \right] = \sum_{\mathbf{k}} \left(\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \psi_{\mathbf{k}}^* \psi_{\mathbf{k}}. \quad (27)$$

Thus we obtain a Gaussian integral, which is easily evaluated:

$$Z \cong \prod_{\mathbf{k}} \left\{ \int_{\mathbb{C}} d^2\psi \exp \left(-\frac{1}{k_B T} \left(\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \psi^* \psi \right) \right\} = \prod_{\mathbf{k}} \frac{\pi k_B T}{\alpha + \frac{\hbar^2 k^2}{2m^*}}. \quad (28)$$

From this, we can obtain thermodynamic quantities. For example, the superfluid contribution to the heat capacity is

$$C_s = -T \frac{\partial^2 F}{\partial T^2} = T \frac{\partial^2}{\partial T^2} k_B T \ln Z \cong k_B T \frac{\partial^2}{\partial T^2} T \sum_{\mathbf{k}} \ln \frac{\pi k_B T}{\alpha + \frac{\hbar^2 k^2}{2m^*}}. \quad (29)$$

We only consider the term that is singular at T_c . It stems from the temperature dependence of $\alpha \cong \alpha' (T - T_c)$. This term is

$$C_{\text{crit}} \cong -k_B T^2 \frac{\partial^2}{\partial T^2} \sum_{\mathbf{k}} \ln \left(\alpha + \frac{\hbar^2 k^2}{2m^*} \right) = k_B T^2 \sum_{\mathbf{k}} \frac{(\alpha')^2}{\left(\alpha + \frac{\hbar^2 k^2}{2m^*} \right)^2}. \quad (30)$$

Going over to an integral over $\mathbf{k}$, corresponding to the thermodynamic limit $V \rightarrow \infty$, we obtain

$$C_{\text{crit}} \cong k_B T^2 \left(\frac{2m^*}{\hbar^2} \right)^2 (\alpha')^2 V \int \frac{d^3 k}{(2\pi)^3} \frac{1}{\left(k^2 + \frac{2m^* \alpha}{\hbar^2} \right)^2} = \frac{k_B T^2}{2\sqrt{2} \pi} \frac{(m^*)^{3/2}}{\hbar^3} (\alpha')^2 V \frac{1}{\sqrt{\alpha}}. \quad (31)$$

We obtain $C_{\text{crit}} \propto (T - T_c)^{-1/2}$. The exponent $-1/2$ of the specific heat is characteristic for Gaussian fluctuations. Recall that at the mean-field level, C_s just showed a step at T_c . We find that superfluid fluctuations in the normal state lead to a divergence for $T \rightarrow T_c^+$.

The derivation shows that for any $\mathbf{k}$ we have two fluctuation modes with dispersion $\epsilon_{\mathbf{k}} = \alpha + (\hbar^2/2m^*) k^2$, corresponding to the real and imaginary parts of $\psi_{\mathbf{k}}$. The dispersion has an *energy gap* $\alpha > 0$. In the language of field theory, one says that the superfluid fluctuations above T_c form two degenerate *massive* branches.

3.2 Fluctuations for $T < T_c$

Below T_c , the situation is more complex due to the wine-bottle potential from Fig. 1: All states with amplitude $|\psi_0| = \sqrt{-\alpha/\beta}$ are equally good mean-field solutions. It is plausible that fluctuations of the phase have low energies since global changes of the phase do not change F . To see this, we write $\psi(\mathbf{r}) = [\psi_0 + \delta\psi(\mathbf{r})] e^{i\phi(\mathbf{r})}$, where ψ_0 and $\delta\psi$ are now real and $\psi_0 = \sqrt{-\alpha/\beta} > 0$. The Ginzburg–Landau functional becomes

$$F[\psi] \cong \int d^3 r \left[\alpha (\psi_0 + \delta\psi)^2 + \frac{\beta}{2} (\psi_0 + \delta\psi)^4 + \frac{\hbar^2}{2m^*} \left(\frac{1}{i} (\nabla \delta\psi) e^{i\phi} + (\psi_0 + \delta\psi) (\nabla \phi) e^{i\phi} \right)^* \cdot \left(\frac{1}{i} (\nabla \delta\psi) e^{i\phi} + (\psi_0 + \delta\psi) (\nabla \phi) e^{i\phi} \right) \right]. \quad (32)$$

Keeping only terms up to second order in the fluctuations (Gaussian approximation), we obtain

$$\begin{aligned} F[\delta\psi, \phi] &\cong \int d^3 r \left[\alpha \delta\psi^2 + 3\beta \psi_0^3 \delta\psi^2 + \frac{\hbar^2}{2m^*} \left(\frac{1}{i} \nabla \delta\psi \right)^* \cdot \frac{1}{i} \nabla \delta\psi \right. \\ &\quad \left. + \frac{\hbar^2}{2m^*} \underbrace{\left(\left(\frac{1}{i} \nabla \delta\psi \right)^* \cdot \psi_0 \nabla \phi + (\psi_0 \nabla \phi)^* \cdot \frac{1}{i} \nabla \delta\psi \right)}_{=0} + \frac{\hbar^2}{2m^*} (\psi_0 \nabla \phi)^* \cdot \psi_0 \nabla \phi \right] \\ &= \int d^3 r \left[-2\alpha \delta\psi^2 + \frac{\hbar^2}{2m^*} \left(\frac{1}{i} \nabla \delta\psi \right)^* \cdot \frac{1}{i} \nabla \delta\psi - \frac{\hbar^2}{2m^*} \frac{\alpha}{\beta} (\nabla \phi)^* \cdot \nabla \phi \right], \end{aligned} \quad (33)$$

up to an irrelevant constant. First-order terms have to cancel. The functional can be simplified by Fourier transformation, in analogy to the case $T > T_c$:

$$F[\delta\psi, \phi] \cong \sum_{\mathbf{k}} \left[\left(-2\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \delta\psi_{\mathbf{k}}^* \delta\psi_{\mathbf{k}} - \frac{\alpha}{\beta} \frac{\hbar^2 k^2}{2m^*} \phi_{\mathbf{k}}^* \phi_{\mathbf{k}} \right]. \quad (34)$$

Up to second order, there are no terms mixing amplitude fluctuations $\delta\psi$ and phase fluctuations ϕ . The amplitude fluctuations are massive with an energy gap $-2\alpha = 2\alpha' (T_c - T)$. They are not degenerate. On the other hand, phase fluctuations are *massless* with energies

$$\epsilon_{\mathbf{k}}^{\phi} = -\frac{\alpha}{\beta} \frac{\hbar^2}{2m^*} k^2. \quad (35)$$

The appearance of these ungapped so-called *Goldstone modes* is characteristic for systems with spontaneously broken continuous symmetries. We note without derivation that the heat capacity diverges like $C \sim (T_c - T)^{-1/2}$ for $T \rightarrow T_c^-$, analogous to the case $T > T_c$.

3.3 Time-dependent Ginzburg–Landau theory

The Ginzburg–Landau approach can be extended to describe dynamics. The coarse-grained wave function $\psi(\mathbf{r}, t)$ is then time dependent. The relevant quantity is the Lagrange-density functional $\mathcal{L}[\psi, \nabla\psi, \dot{\psi}]$, from which one obtains the action

$$\mathcal{S} = \int_{t_i}^{t_f} dt L = \int_{t_i}^{t_f} dt \int d^3r \mathcal{L}, \quad (36)$$

where t_i and t_f are the initial and final time, respectively, and L is the Lagrange functional. Equations of motion for $\psi(\mathbf{r}, t)$ follow from the variational principle $\delta\mathcal{S} = 0$, in analogy to the Maxwell equations in electrodynamics [10].

We only give the main expressions for the case of a neutral superfluid, without derivations. The Lagrange density is

$$\mathcal{L} = i\hbar\psi^* \frac{\partial\psi}{\partial t} - \alpha|\psi|^2 - \frac{\beta}{2}|\psi|^4 - \frac{\hbar^2}{2m^*} |\nabla\psi|^2 \quad (37)$$

so that the Lagrange functional reads

$$L = \int d^3r i\hbar\psi^* \frac{\partial\psi}{\partial t} - F[\psi], \quad (38)$$

with the Ginzburg–Landau functional F . From $\delta\mathcal{S} = 0$, we then obtain, in analogy to the derivation of Eq. (16),

$$i\hbar \frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m^*} \nabla^2\psi + \alpha\psi + \beta|\psi|^2\psi. \quad (39)$$

This time-dependent Ginzburg–Landau or Gross–Pitaevskii equation [7,8] is clearly a nonlinear generalization of the time-dependent Schrödinger equation. It allows us to calculate the time evolution of the macroscopic wave function.

For illustration, we revisit the superfluid fluctuations for $T < T_c$ starting from time-dependent Ginzburg–Landau theory. Making the time-derivative term more symmetric by means of integration by parts and writing $\psi(\mathbf{r}, t) = [\psi_0 + \delta\psi(\mathbf{r}, t)] e^{i\phi(\mathbf{r}, t)}$ in analogy to Sec. 3.2, we obtain

$$\frac{i\hbar}{2} \psi^* \frac{\partial\psi}{\partial t} - \frac{i\hbar}{2} \frac{\partial\psi^*}{\partial t} \psi = \frac{i\hbar}{2} (\psi_0 + \delta\psi) \frac{\partial\delta\psi}{\partial t} + \frac{i\hbar}{2} (\psi_0 + \delta\psi)^2 i \frac{\partial\phi}{\partial t}$$

$$\begin{aligned}
& -\frac{i\hbar}{2} \frac{\partial \delta\psi}{\partial t} (\psi_0 + \delta\psi) - \frac{i\hbar}{2} (\psi_0 + \delta\psi)^2 (-i) \frac{\partial \phi}{\partial t} \\
& = -\hbar (\psi_0 + \delta\psi)^2 \frac{\partial \phi}{\partial t}
\end{aligned} \tag{40}$$

and thus, to second order,

$$\mathcal{L} \cong -2\hbar \sqrt{-\frac{\alpha}{\beta}} \delta\psi \frac{\partial \phi}{\partial t} + 2\alpha \delta\psi^2 - \frac{\hbar^2}{2m^*} \left(\frac{1}{i} \nabla \delta\psi \right)^* \cdot \frac{1}{i} \nabla \delta\psi + \frac{\hbar^2}{2m^*} \frac{\alpha}{\beta} (\nabla \phi)^* \cdot \nabla \phi. \tag{41}$$

First-order terms cancel due to Eq. (39). This form shows that the *canonically conjugate field* to the phase ϕ is

$$\pi \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = -2\hbar \sqrt{-\frac{\alpha}{\beta}} \delta\psi, \tag{42}$$

i.e., the amplitude modes. Here, $\dot{\phi} \equiv \partial \phi / \partial t$. The phase and amplitude fluctuations thus stand in the same relationship as coordinates and momenta in classical mechanics.

Like in classical mechanics, one can go from the Lagrangian description to a Hamiltonian description by a Legendre transformation. The Hamiltonian density is

$$\mathcal{H} = \pi \dot{\phi} - \mathcal{L} = -2\alpha \delta\psi^2 + \frac{\hbar^2}{2m^*} \left(\frac{1}{i} \nabla \delta\psi \right)^* \cdot \frac{1}{i} \nabla \delta\psi - \frac{\hbar^2}{2m^*} \frac{\alpha}{\beta} (\nabla \phi)^* \cdot \nabla \phi. \tag{43}$$

We thus recover Eq. (33).

4 Ginzburg–Landau theory for superconductors

To describe superconductors, one has to take the charge q of the particles forming the condensate into account. We allow for the possibility that q is not the electron charge $-e$. There are two additional terms in the Ginzburg–Landau functional [2], where we use Gaussian units:

- The canonical momentum has to be replaced by the kinetic momentum or, equivalently, the gradient has to be replaced by the gauge-covariant derivative by way of minimal coupling

$$\frac{\hbar}{i} \nabla \rightarrow \frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A}(\mathbf{r}), \tag{44}$$

where $\mathbf{A}(\mathbf{r})$ is the vector potential.

- The energy density of the magnetic field, $B^2/8\pi$, has to be included.

The result is [2]

$$F = \int d^3r \left[\alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m^*} \left| \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right) \psi \right|^2 + \frac{B^2}{8\pi} \right], \tag{45}$$

where $\mathbf{B} = \nabla \times \mathbf{A}$. This is the *Helmholtz* free energy $F[\mathbf{B}]$ with natural variable $\mathbf{B}$, at least with respect to the fluctuations on microscopic scales, as noted above. If we understand F as the difference of the free energies in the superconducting and the normal state, we should

subtract the latter, which introduces an additional term $-B_n^2/8\pi$ under the integral, where $\mathbf{B}_n$ is the magnetic induction in the normal state, which we take to be uniform.

Minimizing the free-energy functional F with respect to the coarse-grained wave function ψ , we obtain, in analogy to Sec. 3,

$$\frac{1}{2m^*} \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right)^2 \psi + \alpha \psi + \beta |\psi|^2 \psi = 0. \quad (46)$$

Other than for neutral superfluids, the mean-field state of the system is now characterized not only by $\psi(\mathbf{r})$ but also by the vector potential $\mathbf{A}(\mathbf{r})$. Hence, we also have to minimize F with respect to $\mathbf{A}$,

$$\begin{aligned} \frac{\delta F}{\delta A_n(\mathbf{r})} &= \frac{\delta}{\delta A_n(\mathbf{r})} \int d^3 r' \left[\frac{1}{2m^*} \left(\left[-\frac{\hbar}{i} \nabla' - \frac{q}{c} \mathbf{A}(\mathbf{r}') \right] \psi^*(\mathbf{r}') \right) \cdot \left(\frac{\hbar}{i} \nabla' - \frac{q}{c} \mathbf{A}(\mathbf{r}') \right) \psi(\mathbf{r}') \right. \\ &\quad \left. + \frac{1}{8\pi} (\nabla' \times \mathbf{A}(\mathbf{r}'))^2 \right] \\ &= \int d^3 r' \left[-\frac{q}{2m^*c} \delta(\mathbf{r}' - \mathbf{r}) \psi^*(\mathbf{r}') \left(\frac{\hbar}{i} \frac{\partial}{\partial r'_n} - \frac{q}{c} A_n(\mathbf{r}') \right) \psi(\mathbf{r}') \right. \\ &\quad \left. - \frac{q}{2m^*c} \left(\left[-\frac{\hbar}{i} \frac{\partial}{\partial r'_n} - \frac{q}{c} A_n(\mathbf{r}') \right] \psi^*(\mathbf{r}') \right) \delta(\mathbf{r}' - \mathbf{r}) \psi(\mathbf{r}') \right. \\ &\quad \left. + \frac{1}{4\pi} \sum_{ij} B_i(\mathbf{r}') \epsilon_{ijn} \frac{\partial}{\partial r'_j} \delta(\mathbf{r}' - \mathbf{r}) \right] \\ &= i \frac{q\hbar}{2m^*c} \psi^*(\mathbf{r}) \frac{\partial \psi}{\partial r_n} - i \frac{q\hbar}{2m^*c} \frac{\partial \psi^*}{\partial r_n} \psi(\mathbf{r}) + \frac{q^2}{m^*c^2} \psi^*(\mathbf{r}) \psi(\mathbf{r}) A_n(\mathbf{r}) \\ &\quad - \frac{1}{4\pi} \sum_{ij} \epsilon_{ijn} \frac{\partial}{\partial r_j} B_i(\mathbf{r}) = 0. \end{aligned} \quad (47)$$

Collecting the three equations for $n = x, y, z$, we obtain

$$-i \frac{q\hbar}{2m^*c} ([\nabla \psi^*] \psi - \psi^* \nabla \psi) + \frac{q^2}{m^*c^2} |\psi|^2 \mathbf{A} + \frac{1}{4\pi} \nabla \times \mathbf{B} = 0. \quad (48)$$

With Ampère's law

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j}, \quad (49)$$

we find

$$\mathbf{j} = \frac{c}{4\pi} \nabla \times \mathbf{B} = i \frac{q\hbar}{2m^*} ([\nabla \psi^*] \psi - \psi^* \nabla \psi) - \frac{q^2}{m^*c} |\psi|^2 \mathbf{A}. \quad (50)$$

In the limit of uniform ψ , Eq. (50) simplifies to

$$\mathbf{j} = -\frac{q^2 |\psi|^2}{m^*c} \mathbf{A}. \quad (51)$$

This should reproduce the London equation [1] for the current density, $\mathbf{j} = -(e^2 n_s / mc) \mathbf{A}$, where n_s is the superfluid density. This obviously requires that

$$\frac{q^2 |\psi|^2}{m^*} = \frac{e^2 n_s}{m}. \quad (52)$$

Equation (51) is evidently not gauge invariant since a gauge transformation $\mathbf{A} \rightarrow \mathbf{A} + \nabla\chi$ changes the current. Charge conservation requires $\nabla \cdot \mathbf{j} = 0$ and thus the vector potential must satisfy $\nabla \cdot \mathbf{A} = 0$. This choice of gauge is called *London gauge* in this context and *Coulomb gauge* in general.

Based on flux-quantization experiments, it is natural to set $q = -2e$. For m^* one might then insert twice the effective electron (band) mass of the metal. However, it turns out to be difficult to measure m^* independently of the superfluid density n_s and it is therefore common to set m^* to twice the bare electron mass, $m^* = 2m \equiv 2m_e$. All system-specific properties are thus absorbed into

$$n_s = \frac{m}{e^2} \frac{q^2 |\psi|^2}{m^*} = \frac{m}{e^2} \frac{4e^2 |\psi|^2}{2m} = 2 |\psi|^2. \quad (53)$$

With this normalization of $\psi(\mathbf{r})$, London theory follows as a special case from Ginzburg–Landau theory.

Taking the curl of Eq. (51), we obtain the *second London equation* [1]

$$\nabla \times \mathbf{j} = -\frac{q^2 |\psi|^2}{m^* c} \nabla \times \mathbf{A} = -\frac{q^2 |\psi|^2}{m^* c} \mathbf{B}. \quad (54)$$

The curl of Ampère’s law Eq. (49) now gives

$$\nabla \times (\nabla \times \mathbf{B}) = \frac{4\pi}{c} \nabla \times \mathbf{j} = -\frac{4\pi q^2 |\psi|^2}{m^* c^2} \mathbf{B}. \quad (55)$$

Using the so-called BAC-CAB rule from vector calculus, we find

$$\nabla^2 \mathbf{B} - \nabla(\nabla \cdot \mathbf{B}) = \nabla^2 \mathbf{B} = \frac{4\pi q^2 |\psi|^2}{m^* c^2} \mathbf{B}. \quad (56)$$

After introducing the *magnetic penetration depth*

$$\lambda \equiv \sqrt{\frac{m^* c^2}{4\pi q^2 |\psi|^2}}, \quad (57)$$

this equation assumes the simple form

$$\nabla^2 \mathbf{B} = \frac{1}{\lambda^2} \mathbf{B}. \quad (58)$$

Let us consider a semi-infinite superconductor filling the half space $x > 0$. An induction field $\mathbf{B}_{\text{appl}} = B_{\text{appl}} \hat{\mathbf{y}}$ is applied parallel to the surface. One can immediately see that the equation is solved by

$$\mathbf{B}(x) = B_{\text{appl}} \hat{\mathbf{y}} e^{-x/\lambda} \quad \text{for } x \geq 0. \quad (59)$$

The magnetic field thus decreases exponentially with the distance from the surface of the superconductor, with the decay length given by the penetration depth λ . In the bulk we indeed find $\mathbf{B} \rightarrow 0$, and thus perfect diamagnetism, which is the Meißner–Ochsenfeld effect.

The penetration depth can be written as

$$\lambda = \sqrt{\frac{mc^2}{8\pi e^2 |\psi|^2}} \cong \sqrt{\frac{mc^2}{8\pi e^2 \left(-\frac{\alpha}{\beta}\right)}}. \quad (60)$$

We have found two characteristic length scales: λ describing the penetration of the magnetic field and ξ describing the variation of the coarse-grained superconducting wave function (order parameter). Within the mean-field theories we have so far employed, both quantities scale as

$$\lambda, \xi \propto \frac{1}{\sqrt{-\alpha}} \propto \frac{1}{\sqrt{T_c - T}} \quad (61)$$

close to T_c . Thus their dimensionless ratio $\kappa \equiv \lambda/\xi$ is roughly temperature independent. It is called the *Ginzburg–Landau parameter*.

4.1 Type-I and type-II superconductors

The Ginzburg–Landau parameter κ determines the behavior of superconductors in an applied magnetic field. Superconductors with small κ are said to be of *type I*. The exact condition is $\kappa < 1/\sqrt{2}$. This means that spatial changes of the magnetic induction $\mathbf{B}$ are less costly than changes of the wave function. In other words, the magnetic field is less stiff than the wave function. Conventional superconductors are usually of type I.

A system will only become superconducting if this reduces its free energy. The experimental situation we have in mind is that the sample is placed in a large solenoid, which maintains a constant magnetic field. When the sample becomes superconducting, the source that drives a current through the solenoid has to do work to keep the field constant. This work has to be taken into account in the energy balance between the superconducting and the normal-conducting states but the Helmholtz free energy F does not contain it [6]. The calculation of this additional energy is a standard problem in macroscopic electrodynamics. (The simplest setup for an explicit derivation involves an infinitely long cylindrical body concentric within an infinitely long solenoid. In this setup, the magnetic field $\mathbf{H}$ is continuous and given in terms of the sheet current density $\mathbf{K}$ carried by the solenoid.) The result for the additional energy is

$$-\int d^3r \frac{\mathbf{H} \cdot \mathbf{B}}{4\pi}, \quad (62)$$

where the magnetic field $\mathbf{H}$ is given by

$$\mathbf{H}(\mathbf{r}) = 4\pi \frac{\delta F}{\delta \mathbf{B}(\mathbf{r})}. \quad (63)$$

Adding this contribution to the Ginzburg–Landau functional yields

$$G[\mathbf{H}] = F[\mathbf{B}] - \int d^3r \frac{\mathbf{H} \cdot \mathbf{B}}{4\pi}. \quad (64)$$

Thermodynamically speaking, this equation is the Legendre transformation from the *Helmholtz* free energy F with natural variable $\mathbf{B}$ to the *Gibbs* free energy G with natural variable $\mathbf{H}$.

The Helmholtz free energy F is given in Eq. (45). For the evaluation of Eq. (63), we require the functional derivative of $\mathbf{A}$ with respect to $\mathbf{B}$. To find it, we note that the relation $\mathbf{B} = \nabla \times \mathbf{A}$ can be explicitly inverted:

$$\mathbf{A}(\mathbf{r}) = \frac{1}{4\pi} \int d^3r' \mathbf{B}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (65)$$

This expression respects the Coulomb gauge, i.e., $\nabla \cdot \mathbf{A} = 0$. The derivation is formally analogous to the one of the Biot–Savart law for $\mathbf{B}$ for given current density $\mathbf{j}$ [11]. From Eq. (65), we obtain

$$\frac{\delta A_i(\mathbf{r})}{\delta B_j(\mathbf{r}')} = \frac{1}{4\pi} \sum_k \epsilon_{ijk} \frac{r_k - r'_k}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (66)$$

We can then evaluate Eq. (63). A lengthy calculation gives

$$\begin{aligned} \mathbf{H}(\mathbf{r}) &= \frac{1}{c} \int d^3 r' \left[i \frac{q\hbar}{2m^*} ([\nabla' \psi^*(\mathbf{r}')] \psi(\mathbf{r}') - \psi^*(\mathbf{r}') \nabla' \psi(\mathbf{r}')) - \frac{q^2}{m^* c} |\psi(\mathbf{r}')|^2 \mathbf{A}(\mathbf{r}') \right] \\ &\quad \times \frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} + \mathbf{B}(\mathbf{r}) = -\frac{1}{c} \int d^3 r' \mathbf{j}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} + \mathbf{B}(\mathbf{r}), \end{aligned} \quad (67)$$

where the current density $\mathbf{j}$ is the one appearing in the second Ginzburg–Landau equation (50). One can write the result as $\mathbf{H} = -4\pi\mathbf{M} + \mathbf{B}$ with the magnetization due to the screening currents,

$$\mathbf{M}(\mathbf{r}) = \frac{1}{4\pi c} \int d^3 r' \mathbf{j}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}. \quad (68)$$

Evidently, the thermodynamic relation (63) forces us to treat $\mathbf{j}$ as an *induced*, not free, current density. This is also reasonable from the point of view of electrodynamics: In macroscopic magnetostatics as presented in textbooks [11], we are used to divide the current into bound and free currents, where notionally the former are restricted to the atomic scale. However, the correct distinction is between currents that are induced by the applied field and currents imposed by the experimenter. The screening currents in superconductors are free currents in the sense that charge carriers are flowing over macroscopic distances—famously so—but they are nevertheless induced by the applied magnetic field, expressed by the perfect diamagnetism. With Eqs. (45), (64), and (67), the Gibbs free energy becomes

$$G = \int d^3 r \left[\alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m^*} \left| \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right) \psi \right|^2 - \frac{H^2}{8\pi} + 2\pi M^2 \right]. \quad (69)$$

The awkward factor of 2π instead of $1/8\pi$ in front of M^2 is due to using Gaussian units.

In the normal state, we have $\psi = 0$ and $\mathbf{H} = \mathbf{H}_{\text{appl}}$, the applied field. Now taking G to be the difference of the Gibbs free energies of the superconducting and normal states, we obtain

$$G = G_s - G_n = \int d^3 r \left[\alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m^*} \left| \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right) \psi \right|^2 - \frac{H^2}{8\pi} + 2\pi M^2 + \frac{H_{\text{appl}}^2}{8\pi} \right]. \quad (70)$$

The evaluation becomes simple in the case of a uniform system: In the superconducting state, the uniform solution of the Ginzburg–Landau equation (46) is $|\psi|^2 = -\alpha/\beta > 0$ and the $\mathbf{B}$ field in the bulk vanishes due to the Meißner–Ochsenfeld effect. Hence, we can, on the one hand, choose a gauge with $\mathbf{A} = 0$ and, on the other, we get

$$-\frac{H^2}{8\pi} + 2\pi M^2 = -2\pi M^2 + 2\pi M^2 = 0. \quad (71)$$

This leads to the simple result

$$G = \int d^3r \left(-\frac{\alpha^2}{2\beta} + \frac{H_{\text{appl}}^2}{8\pi} \right). \quad (72)$$

Superconductivity occurs for $G < 0$. Not surprisingly, the applied magnetic field hinders superconductivity. The *critical field* H_c at which superconductivity is destroyed is obtained from the condition $G = 0$, which yields

$$H_c = \sqrt{4\pi \frac{\alpha^2}{\beta}} \cong \sqrt{\frac{4\pi}{\beta}} \alpha' (T_c - T). \quad (73)$$

We can now express the phenomenological parameters α and β in terms of the measurable quantities λ and H_c as

$$\alpha = -\frac{2e^2}{mc^2} \lambda^2 H_c^2, \quad (74)$$

$$\beta = 4\pi \left(\frac{2e^2}{mc^2} \right)^2 \lambda^4 H_c^2. \quad (75)$$

One can also show that the free energy of domain walls between normal-conducting and superconducting regions is positive, which explains why type-I superconductors favor uniform states [6].

Conversely, type-II superconductors are defined by a large Ginzburg–Landau parameter $\kappa > 1/\sqrt{2}$. The free energy of domain walls is then negative [6]. Hence, the system tends to maximize the total area of domain walls. This tendency is counterbalanced by flux quantization—the magnetic flux in a normal region cannot be split into portions smaller than the flux quantum $\Phi_0 \equiv hc/2e$, as we show in the following.

Consider an arbitrary closed path ∂S forming the edge of a surface S . The magnetic flux Φ through the surface S is

$$\Phi = \int_S d\mathbf{a} \cdot \mathbf{B} = \int_S d\mathbf{a} \cdot (\nabla \times \mathbf{A}) = \oint_{\partial S} d\mathbf{s} \cdot \mathbf{A}, \quad (76)$$

by Stokes' theorem. We solve the second Ginzburg–Landau equation (50) for $\mathbf{A}$ and insert the result,

$$\begin{aligned} \Phi &= \oint_{\partial S} d\mathbf{s} \cdot \left[-\frac{m^*c}{q^2|\psi|^2} \mathbf{j} + i \frac{\hbar c}{2q|\psi|^2} ([\nabla\psi^*]\psi - \psi^*\nabla\psi) \right] \\ &= \oint_{\partial S} d\mathbf{s} \cdot \left[-\frac{mc}{e^2n_s} \mathbf{j} + i \frac{\hbar c}{2q} (-2i)\nabla\phi \right] = \frac{mc}{e} \oint_{\partial S} d\mathbf{s} \cdot \mathbf{v}_s - \frac{\hbar c}{2e} \oint_{\partial S} d\mathbf{s} \cdot \nabla\phi, \end{aligned} \quad (77)$$

where $\mathbf{v}_s \equiv -\mathbf{j}/en_s$ is the superfluid velocity. The last term contains the change of the phase of the macroscopic wave function along the whole path, which must be an integer multiple of 2π for $\psi(\mathbf{r})$ to be continuous:

$$\oint_{\partial S} d\mathbf{s} \cdot \nabla\phi = -2\pi n, \quad n \in \mathbb{Z}. \quad (78)$$

The minus sign is a convention, which is useful for negatively charged condensates. Thus we find for the so-called *fluxoid*

$$\Phi' \equiv \Phi - \frac{mc}{e} \oint_{\partial S} d\mathbf{s} \cdot \mathbf{v}_s = \frac{\hbar c}{2e} 2\pi n = \frac{\hbar c}{2e} n = n\Phi_0, \quad n \in \mathbb{Z}. \quad (79)$$

Here n is the *winding number* enclosed by the path ∂S . We see that the fluxoid is quantized. Moreover, if the path ∂S everywhere lies deep inside a superconductor the current vanishes, Φ' equals Φ , and the flux Φ itself is quantized.

The smallest amount of fluxoid that can penetrate a superconductor is one flux quantum Φ_0 . It does so as a *vortex line*. Following the above arguments, the phase ϕ of ψ changes by -2π as one circles a vortex in the positive direction, where, by convention, the direction of a vortex is the direction of the magnetic induction $\mathbf{B}$. Thus at the center of the vortex line (the *vortex core*), the phase is undefined. This is only consistent with a continuous ψ if $\psi = 0$ holds there. Because of the phase winding, a vortex cannot end anywhere in the interior of a superconductor. Vortices can thus only terminate at a surface or form closed loops. One can understand the winding of the phase as a *topological* property of the vortex since it cannot be changed by any continuous deformation of $\psi(\mathbf{r})$ unless ψ is tuned to zero. In this sense, vortices are *topological defects* of the condensate.

To obtain the coarse-grained wave function $\psi(\mathbf{r})$ and the magnetic induction $\mathbf{B}(\mathbf{r})$ for a vortex line, one has to solve the Ginzburg–Landau equations (46) and (50) together with Ampère’s law Eq. (49) [6]. This can only be done numerically. One finds that $|\psi|$ starts to grow linearly with the distance from the vortex core. The characteristic length scale of this rise is, not surprisingly, the coherence length ξ . On the other hand, $\mathbf{B}$ falls off with distance on the scale of the magnetic penetration depth λ .

One also finds a more complicated phase diagram than for type-I superconductors. At weak applied fields $H < H_{c1}$, the flux is expelled completely. In an intermediate field range $H_{c1} < H < H_{c2}$, the equilibrium state is a regular lattice of vortex lines [12], specifically a triangular lattice of parallel straight lines. The properties of the vortex lattice and the critical fields H_{c1} and H_{c2} can be evaluated within Ginzburg–Landau theory. Details can be found in Refs. [6, 13].

4.2 The Anderson–Higgs mechanism

When discussing fluctuations about the mean-field state of a superconductor, we have to take the coupling to the electromagnetic field into account. The suitable thermodynamic potential in this case is the Helmholtz free energy

$$F = \int d^3r \left[\alpha |\psi|^2 + \frac{\beta}{2} |\psi|^4 + \frac{1}{2m^*} \left| \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right) \psi \right|^2 + \frac{B^2}{8\pi} - \frac{B_{\text{appl}}^2}{8\pi} \right], \quad (80)$$

which expresses that fluctuations of the magnetic field *increase* the energy. We proceed analogously to the case of a neutral superfluid and employ the Gaussian approximation throughout, i.e., we expand F up to second order in fluctuations.

For $T > T_c$, we note that the kinetic term

$$\frac{1}{2m^*} \left| \left(\frac{\hbar}{i} \nabla - \frac{q}{c} \mathbf{A} \right) \psi \right|^2 \quad (81)$$

is explicitly of second order in the fluctuations $\delta\psi = \psi$ so that electromagnetic-field fluctuations only appear in even higher orders. Thus to second order, the order-parameter fluctuations decouple from the electromagnetic fluctuations,

$$F \cong \int d^3r \left[\alpha \psi^* \psi + \frac{1}{2m^*} \left| \frac{\hbar}{i} \nabla \psi \right|^2 + \frac{B^2}{8\pi} \right] = \sum_{\mathbf{k}} \left(\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \psi_{\mathbf{k}}^* \psi_{\mathbf{k}} + \sum_{\mathbf{k} \neq 0} \frac{\mathbf{B}_{\mathbf{k}}^* \cdot \mathbf{B}_{\mathbf{k}}}{8\pi}. \quad (82)$$

The superconducting fluctuations consist of two degenerate massive branches with dispersion $\epsilon_{\mathbf{k}} = \alpha + \hbar^2 k^2 / 2m^*$, like for the neutral superfluid. The magnetic-field fluctuations decouple and are those of a free field. The final sum excludes the uniform magnetic-field mode, which is the applied field subtracted in Eq. (80).

The case of $T < T_c$ is more interesting. Writing $\psi = (\psi_0 + \delta\psi) e^{i\phi}$ as above, we obtain, to second order,

$$\begin{aligned} F \cong & \int d^3r \left[\alpha \delta\psi^2 + 3\beta\psi_0^2 \delta\psi^2 + \frac{1}{2m^*} \left(\frac{\hbar}{i} \nabla \delta\psi \right)^* \cdot \frac{\hbar}{i} \nabla \delta\psi \right. \\ & + \frac{1}{2m^*} \underbrace{\left(\left(\frac{\hbar}{i} \nabla \delta\psi \right)^* \cdot \hbar\psi_0 \nabla \phi + (\hbar\psi_0 \nabla \phi)^* \cdot \frac{\hbar}{i} \nabla \delta\psi \right)}_{=0} + \frac{1}{2m^*} (\hbar\psi_0 \nabla \phi)^* \cdot \hbar\psi_0 \nabla \phi \\ & + \frac{1}{2m^*} \underbrace{\left(\left(\frac{\hbar}{i} \nabla \delta\psi \right)^* \cdot \left(-\frac{q}{c} \right) \mathbf{A} \psi_0 + \left(-\frac{q}{c} \mathbf{A} \psi_0 \right)^* \cdot \frac{\hbar}{i} \nabla \delta\psi \right)}_{=0} \\ & + \frac{1}{2m^*} \left((\hbar\psi_0 \nabla \phi)^* \cdot \left(-\frac{q}{c} \right) \mathbf{A} \psi_0 + \left(-\frac{q}{c} \mathbf{A} \psi_0 \right)^* \cdot \hbar\psi_0 \nabla \phi \right) \\ & \left. + \frac{1}{2m^*} \left(-\frac{q}{c} \mathbf{A} \psi_0 \right)^* \cdot \left(-\frac{q}{c} \right) \mathbf{A} \psi_0 + \frac{1}{8\pi} (\nabla \times \mathbf{A})^* \cdot (\nabla \times \mathbf{A}) \right] \\ = & \int d^3r \left[-2\alpha \delta\psi^2 + \frac{1}{2m^*} \left(\frac{\hbar}{i} \nabla \delta\psi \right)^* \cdot \frac{\hbar}{i} \nabla \delta\psi \right. \\ & \left. - \frac{\hbar^2}{2m^*} \frac{\alpha}{\beta} \left(\nabla \phi - \frac{q}{\hbar c} \mathbf{A} \right)^* \cdot \left(\nabla \phi - \frac{q}{\hbar c} \mathbf{A} \right) + \frac{1}{8\pi} (\nabla \times \mathbf{A})^* \cdot (\nabla \times \mathbf{A}) \right], \quad (83) \end{aligned}$$

where constant terms have been dropped and first-order terms again cancel. We see that the phase ϕ of the macroscopic wave function and the vector potential appear in the combination $\nabla \phi - (q/\hbar c) \mathbf{A}$. This property is a consequence of gauge invariance: Observable physical properties are invariant under the simultaneous gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla \chi, \quad (84)$$

$$U' = U - \frac{1}{c} \frac{\partial \chi}{\partial t}, \quad (85)$$

$$\psi' = e^{iq\chi/\hbar c} \psi, \quad (86)$$

where we here use U for the scalar electric potential and $\chi(\mathbf{r}, t)$ is an arbitrary scalar field. Equation (86) means that the phase of ψ transforms as

$$\phi' = \phi + \frac{q}{\hbar c} \chi. \quad (87)$$

Besides $\nabla\phi - (q/\hbar c)\mathbf{A}$, $\mathbf{B} = \nabla \times \mathbf{A}$ is also gauge invariant since $\nabla \times \nabla\chi = 0$.

Gauge invariance is not a symmetry of the physical system but a redundancy in its description. Hence, variations of ϕ and $\mathbf{A}$ that take the form of Eqs. (84) and (87) are not physical fluctuations. One way to deal with this is to choose a particular gauge from the manifold of equivalent descriptions. A useful choice in this case is to make the new collective wave function real, i.e., to require $\phi' = 0$. Equations (84) and (87) then imply that the new vector potential is

$$\mathbf{A}' = \mathbf{A} - \frac{\hbar c}{q} \nabla\phi. \quad (88)$$

The Ginzburg–Landau functional can be expressed in terms of the new vector potential and the amplitude fluctuations $\delta\psi$ as

$$F[\delta\psi, \mathbf{A}'] \cong \int d^3r \left[-2\alpha \delta\psi^2 + \frac{1}{2m^*} \left(\frac{\hbar}{i} \nabla\delta\psi \right)^* \cdot \frac{\hbar}{i} \nabla\delta\psi - \frac{\alpha}{\beta} \frac{q^2}{2m^*c^2} \mathbf{A}'^* \cdot \mathbf{A}' + \frac{1}{8\pi} (\nabla \times \mathbf{A}')^* \cdot (\nabla \times \mathbf{A}') \right]. \quad (89)$$

The procedure is often described by saying that the phase is absorbed into the vector potential by a gauge transformation.

Finally, we drop the prime and insert Fourier-transformed fields,

$$F[\delta\psi, \mathbf{A}] \cong \sum_{\mathbf{k}} \left[\left(-2\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \delta\psi_{\mathbf{k}}^* \delta\psi_{\mathbf{k}} - \frac{\alpha}{\beta} \frac{q^2}{2m^*c^2} \mathbf{A}_{\mathbf{k}}^* \cdot \mathbf{A}_{\mathbf{k}} + \frac{1}{8\pi} (\mathbf{k} \times \mathbf{A}_{\mathbf{k}})^* \cdot (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) \right]. \quad (90)$$

Obviously, the massive amplitude fluctuations decouple from electromagnetic fluctuations to second order and behave like for a neutral superfluid. The amplitude modes $\delta\psi_{\mathbf{k}}$ are called *Higgs modes* [14] in this context.

For the interpretation of the last expression, it is useful to decompose the vector potential $\mathbf{A}$ into longitudinal components $\mathbf{A}_{\mathbf{k}}^{\parallel}$ parallel to the wave vector $\mathbf{k}$ and transverse components $\mathbf{A}_{\mathbf{k}}^{\perp}$ perpendicular to it:

$$\mathbf{A}_{\mathbf{k}} = \underbrace{\hat{\mathbf{k}} (\hat{\mathbf{k}} \cdot \mathbf{A}_{\mathbf{k}})}_{\equiv \mathbf{A}_{\mathbf{k}}^{\parallel}} + \underbrace{\mathbf{A}_{\mathbf{k}} - \hat{\mathbf{k}} (\hat{\mathbf{k}} \cdot \mathbf{A}_{\mathbf{k}})}_{\equiv \mathbf{A}_{\mathbf{k}}^{\perp}}, \quad (91)$$

with the unit vector $\hat{\mathbf{k}} = \mathbf{k}/k$. The magnetic-field energy density can now be written as

$$\begin{aligned} \frac{1}{8\pi} (\mathbf{k} \times \mathbf{A}_{\mathbf{k}})^* \cdot (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}) &= \frac{1}{8\pi} (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}^{\perp})^* \cdot (\mathbf{k} \times \mathbf{A}_{\mathbf{k}}^{\perp}) \\ &= \frac{1}{8\pi} [k^2 \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp} - (\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}^{\perp*})(\mathbf{k} \cdot \mathbf{A}_{\mathbf{k}}^{\perp})] = \frac{1}{8\pi} k^2 \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp} \end{aligned} \quad (92)$$

and the Ginzburg–Landau functional becomes

$$F[\delta\psi, \mathbf{A}] \cong \sum_{\mathbf{k}} \left[\left(-2\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \delta\psi_{\mathbf{k}}^* \delta\psi_{\mathbf{k}} - \frac{\alpha}{\beta} \frac{q^2}{2m^* c^2} \mathbf{A}_{\mathbf{k}}^{\parallel*} \cdot \mathbf{A}_{\mathbf{k}}^{\parallel} - \frac{\alpha}{\beta} \frac{q^2}{2m^* c^2} \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp} + \frac{1}{8\pi} k^2 \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp} \right]. \quad (93)$$

The terms proportional to $-\alpha/\beta = |\psi_0|^2$ are due to superconductivity. Without it, we would have the free-field functional

$$F_{\text{free}}[\mathbf{A}] = \sum_{\mathbf{k}} \frac{1}{8\pi} k^2 \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp}, \quad (94)$$

which only contains the two transverse components. This result persists in quantum electrodynamics. The transverse modes do not have an energy gap—photons are massless.

In the presence of superconductivity, all three components of $\mathbf{A}$ appear. Moreover, all components obtain a *mass term*. This is the famous *Anderson–Higgs mechanism* first proposed for superconductors by P.W. Anderson [15, 16]. The same general idea also explains (most of) the masses of elementary particles, although in a more complicated case since the broken symmetry is more complicated than the U(1) symmetry of superconductivity [17–19]. The “Higgs bosons” in our case are the quanta of the amplitude fluctuations [14].

To extract the dispersion from Eq. (93), one has to go beyond our static picture and include the time or frequency dependence of the fields. The frequency-dependent term for the vector potential is of the form

$$-\frac{1}{8\pi} \frac{\omega^2}{c^2} \mathbf{A}_{\mathbf{k}}^* \cdot \mathbf{A}_{\mathbf{k}}. \quad (95)$$

This is plausible based on Eq. (94) and Lorentz invariance. Equation (93) then predicts that the longitudinal field components $\mathbf{A}_{\parallel}$ are dispersionless,

$$\omega^2 = -\frac{\alpha}{\beta} \frac{4\pi q^2}{m^*} = \frac{4\pi n_s e^2}{m} \equiv \omega_p^2. \quad (96)$$

Here, ω_p is the *plasma frequency*. This is the frequency with which charge inhomogeneities oscillate under the influence of the Coulomb interaction. These plasma oscillations are not restricted to superconductors and, in particular, also occur in normal metals, where n_s is replaced by the density $n = N/V$ of all conduction electrons. The dispersion of plasmons is encoded in the dielectric function $\epsilon(\mathbf{k}, \omega)$ in the limit of small wave vectors $\mathbf{k}$, which can be calculated in the random-phase approximation (RPA) [10]. Such a calculation reveals that there are momentum-dependent corrections, i.e., the dispersion is not perfectly flat.

On the other hand, the transverse components are dispersive,

$$\omega^2 = \omega_p^2 + k^2 c^2. \quad (97)$$

This is because transverse plasma oscillations mix with the transverse electromagnetic field modes. They become indistinguishable from longitudinal plasma oscillations in the long-wavelength ($\mathbf{k} \rightarrow 0$) limit, giving the same frequency ω_p but quickly transform into electromagnetic-field-dominated modes with optical dispersion $\omega = kc$ for larger $\mathbf{k}$. The approximate

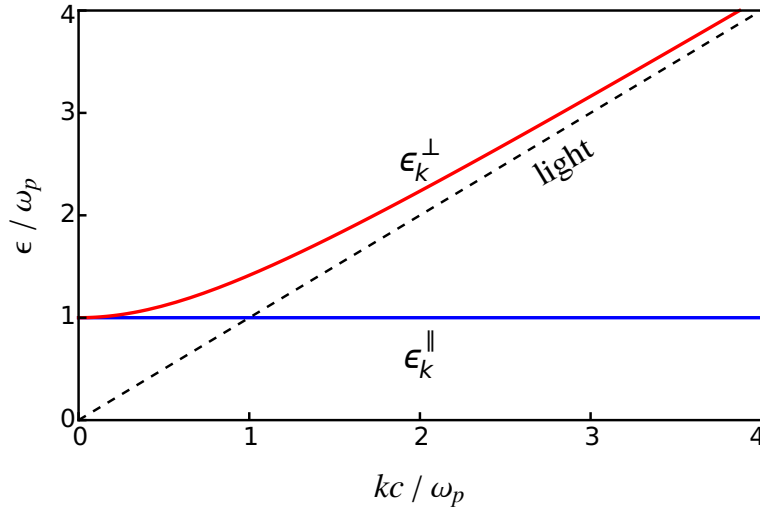


Fig. 3: Dispersion $\epsilon_k^{\parallel}$ of longitudinal plasma modes and $\epsilon_k^{\perp}$ of transverse plasma modes. For comparison, the dispersion of light is shown as the dashed straight line.

dispersions are sketched in Fig. 3. For completeness, we note that the RPA corrections to the dispersion of longitudinal plasma modes are on the order of $k^2 v_F^2$, where $v_F \ll c$ is the electronic Fermi velocity.

The dispersion of the transverse modes also generates the Meißner–Ochsenfeld effect. To see this, we write the functional as

$$F \cong \sum_{\mathbf{k}} \left(-2\alpha + \frac{\hbar^2 k^2}{2m^*} \right) \delta\psi_{\mathbf{k}}^* \delta\psi_{\mathbf{k}} + \sum_{\mathbf{k}} \frac{1}{8\pi} \frac{1}{\lambda^2} \mathbf{A}_{\mathbf{k}}^{\parallel*} \cdot \mathbf{A}_{\mathbf{k}}^{\parallel} + \sum_{\mathbf{k}} \frac{1}{8\pi} \left(\frac{1}{\lambda^2} + k^2 \right) \mathbf{A}_{\mathbf{k}}^{\perp*} \cdot \mathbf{A}_{\mathbf{k}}^{\perp}, \quad (98)$$

where λ is the magnetic penetration depth defined in Eq. (57). The photon mass is thus proportional to $1/\lambda$. We minimize the free energy with respect to the transverse components of the vector potential,

$$\frac{\partial F}{\partial \mathbf{A}_{\mathbf{k}}^{\perp*}} = \frac{1}{8\pi} \left(\frac{1}{\lambda^2} + k^2 \right) \mathbf{A}_{\mathbf{k}}^{\perp} \stackrel{!}{=} 0. \quad (99)$$

Fourier transformation back to real space gives

$$\frac{1}{\lambda^2} \mathbf{A}^{\perp} - \nabla^2 \mathbf{A}^{\perp} = 0. \quad (100)$$

Taking the curl, we reobtain Eq. (58),

$$\frac{1}{\lambda^2} \mathbf{B} - \nabla^2 \mathbf{B} = 0. \quad (101)$$

4.3 Ginzburg–Landau theory as an effective field theory

So far, Ginzburg–Landau theory has been introduced as a phenomenological theory based on Landau theory and the idea that superconductivity involves the condensation of charged particles, which happen to have charge $-2e$. Two years after the publication of the microscopic BCS theory [4], Lev Gor’kov [5] managed to derive Ginzburg–Landau theory from BCS theory. The correspondence is perfect in the limit that the gap is sufficiently small, i.e., T is close to T_c ,

and the electromagnetic field varies sufficiently slowly. These are indeed the conditions under which Ginzburg and Landau expected their theory to be valid. Gor'kov's derivation links the parameters appearing in the Ginzburg–Landau functional to microscopic quantities. In particular, he found that $q = -2e$ and $m^* = 2m$.

Gor'kov [5] used equations of motion for electronic Green functions, which he decoupled using a mean-field-like approximation that allowed for spatial variations of the mean-field decoupling term. A more modern exposition using functional integrals is given by Altland and Simons [10]. While we cannot discuss this derivation here, the main ideas are recounted for students who have studied functional integrals: The partition function of the interacting electrons is written as a functional integral

$$Z = \int D\bar{\psi} D\psi e^{-S[\bar{\psi}, \psi]}, \quad (102)$$

where ψ is a Grassmann field, i.e., a field of anticommuting variables, required for a functional integral describing fermions, and S is the action. The evaluation of the functional integral would be simple even for Grassmann fields if the action S were quadratic. However, it is not, due to the presence of a quartic interaction term. This interaction is decoupled with a Hubbard–Stratonovich transformation, which means that unity is introduced in the form of a Gaussian functional integral over a bosonic auxiliary field. This Gaussian integral is constructed in such a way that the interaction term is canceled. Physically speaking, the four-fermion interaction is replaced by the exchange of an auxiliary boson.

Replacing the auxiliary bosonic field by a constant that makes the action stationary corresponds to the mean-field approximation of BCS theory. Next, the action is expanded in the auxiliary field about this saddle-point solution and the, now Gaussian, functional integral over the *fermionic* field is performed. The result is the partition function expressed as a functional integral over the deviation of the bosonic field from the saddle-point value. The action in the integral, truncated at a suitable order, is essentially the time-dependent Ginzburg–Landau functional.

4.4 Ginzburg–Landau theory for nontrivial symmetries

Recall that the idea behind Landau and also Ginzburg–Landau theory is to expand the functional of interest in order parameters and external fields. This expansion should include all symmetry-allowed terms up to the order we find to be necessary. The general procedure is discussed in Ref. [20]. The relevant symmetries are crystal symmetries, in particular those described by the crystallographic point group, and time-reversal symmetry. In addition, all terms must be gauge invariant. Group theory and in particular representation theory of groups provides powerful methods for the analysis of symmetry and the construction of allowed terms, which cannot be derived here. The book Ref. [21] contains a good introduction from a condensed-matter-physics point of view. Many useful tables can be found on the Bilbao Crystallographic Server [22] and in Ref. [23]. Recent application of the ideas outlined here to specific superconductors can be found in Refs. [24, 25].

Above, we have assumed that only a single complex superconducting order parameter ψ is required to describe the physics. However, multiple complex order parameters can be relevant for two reasons:

- There is a single relevant order parameter but it is intrinsically of higher dimension—technically, this is the case for order parameters that transform according to higher-dimensional irreducible representations of the point group—or
- multiple order parameters of different symmetries are accidentally close in energy.

The Ginzburg–Landau functional can thus depend on several complex order-parameter fields $\psi_n(\mathbf{r})$ (belonging to different components of the same irreducible representation or to different irreducible representations) and on the local magnetic induction $\mathbf{B}(\mathbf{r})$. We here consider the Helmholtz free energy, see Eq. (45). The functional might also depend on other fields, such as magnetic order parameters or distortions, which can be treated analogously. To construct symmetry-allowed terms, we first have to determine how these fields transform under point-group operations and under time reversal.

The magnetic field transforms like a *pseudovector* (or axial vector) under point-group operations, i.e., $\mathbf{B}$ behaves like the spatial vector $\mathbf{r}$ under proper rotations but is even under inversion. Furthermore, $\mathbf{B}$ is odd under time reversal. The vector potential $\mathbf{A}$ transforms like a vector but is also odd under time reversal.

Moreover, ψ_n is mapped to ψ_n^* under time reversal. This is plausible in view of Eq. (50): Under time reversal, $\psi \mapsto \psi^*$ and $\mathbf{A} \mapsto -\mathbf{A}$ so that the current density $\mathbf{j}$ changes sign. In addition, the functional contains gauge-covariant derivatives

$$\mathbf{D} \equiv \nabla - i \frac{q}{\hbar c} \mathbf{A}(\mathbf{r}), \quad (103)$$

see Eq. (44). $\mathbf{D}$ transforms like a vector under point-group operations since $\nabla = \partial/\partial\mathbf{r}$ and $\mathbf{A}$ do. Moreover, under time reversal, $\mathbf{D}$ maps to $\mathbf{D}^*$ since $\mathbf{A}$ is odd. For later, we note that under the spatial integral, we have the equality

$$\begin{aligned} \int d^3r f^* \mathbf{D} g &= \int d^3r \left(f^* \nabla g - f^* i \frac{q}{\hbar c} \mathbf{A} g \right) = \int d^3r \left[-(\nabla f)^* g + \left(i \frac{q}{\hbar c} f \right)^* g \right] \\ &= - \int d^3r (\mathbf{D} f)^* g \end{aligned} \quad (104)$$

for any two differentiable functions $f(\mathbf{r})$ and $g(\mathbf{r})$ so that the integral converges. We have used integration by parts and dropped an irrelevant boundary term in the infinite.

So far, we have implicitly assumed *conventional* superconductivity in the sense that the superconducting order parameter ψ does not break lattice symmetries. In other words, ψ was taken to be invariant under all point-group operations, i.e., technically speaking, to transform according to the trivial irreducible representation. In the general case, superconducting order parameters ψ_n may transform nontrivially under point-group operations. How they transform is completely encoded in the irreducible representation to which they belong. A well-known example is the

cuprate high-temperature superconductors showing the famous $d_{x^2-y^2}$ -wave pairing. The z -axis is a fourfold rotation axis of the crystal but the superconducting order parameter changes sign under rotation by $\pi/2$ about this axis, like the basis function $\mathbf{r} \mapsto x^2 - y^2$.

The next step is to construct all combinations of the fields and the gauge-covariant derivative that are invariant under all point-group operations and under time reversal, are gauge invariant, and are real. Gauge invariance requires all terms to contain the same number of factors ψ_n as of factors ψ_n^* . This can already be seen by considering global phase rotations $\psi_n(\mathbf{r}) \mapsto e^{i\alpha} \psi_n(\mathbf{r})$, which is a special gauge transformation. (Time-reversal symmetry is not sufficient for this since a term proportional to $\psi_n + \psi_n^*$ is invariant under time reversal but forbidden by gauge invariance.)

As an example, we consider a cubic superconductor with point group O_h and a three-component order parameter $\boldsymbol{\psi} = (\psi_x, \psi_y, \psi_z)$ that transforms like a vector. This means that it transforms according to the irreducible representation (irrep) T_{1u} of O_h . All terms in the Ginzburg–Landau functional must transform according to the trivial irrep A_{1g} .

The second-order term must involve products of one component of $\boldsymbol{\psi}^*$ and one component of $\boldsymbol{\psi}$. It is very useful to consider the reduction of products of irreps, in this case [21–23]

$$T_{1u} \otimes T_{1u} = A_{1g} \oplus E_g \oplus T_{1g} \oplus T_{2g}. \quad (105)$$

Since A_{1g} appears exactly once in the above reduction there is one way to construct a fully symmetric term out of the two vectors $\boldsymbol{\psi}^*$ and $\boldsymbol{\psi}$. This term is $\boldsymbol{\psi}^* \cdot \boldsymbol{\psi} = \psi_x^* \psi_x + \psi_y^* \psi_y + \psi_z^* \psi_z$. To fourth order, we have to consider [21–23]

$$T_{1u} \otimes T_{1u} \otimes T_{1u} \otimes T_{1u} = 4 A_{1g} \oplus 3 A_{2g} \oplus 7 E_g \oplus 10 T_{1g} \oplus 10 T_{2g}. \quad (106)$$

As A_{1g} occurs four times we could expect that there are four distinct terms of fourth order. This would indeed be correct for a product of four distinct vectors. However, here we have two factors of $\boldsymbol{\psi}^*$ and two of $\boldsymbol{\psi}$ and we will see shortly that one contribution drops out by being trivial. All these terms can be constructed from second-order products $\psi_m \psi_n$ and their complex conjugates. Equation (105) implies that these products can, in principle, be organized into terms transforming according to the four irreps A_{1g} , E_g , T_{1g} , and T_{2g} . The product of any of these irreps with itself contains the trivial A_{1g} exactly once. We can thus obtain the maximum number of four independent fourth-order terms in this way. The four possibilities are:

- The A_{1g} product of second order is $\boldsymbol{\psi} \cdot \boldsymbol{\psi} = \psi_x^2 + \psi_y^2 + \psi_z^2$, which gives the fourth-order term

$$(\boldsymbol{\psi}^* \cdot \boldsymbol{\psi}^*)(\boldsymbol{\psi} \cdot \boldsymbol{\psi}). \quad (107)$$

- The E_g combination of second order is [21–23] $(\psi_x^2 - \psi_y^2, (1/\sqrt{3})(2\psi_z^2 - \psi_x^2 - \psi_y^2))$, which gives the fully symmetric fourth-order term

$$\left(\psi_x^2 - \psi_y^2, \frac{2\psi_z^2 - \psi_x^2 - \psi_y^2}{\sqrt{3}} \right)^* \cdot \left(\psi_x^2 - \psi_y^2, \frac{2\psi_z^2 - \psi_x^2 - \psi_y^2}{\sqrt{3}} \right)$$

$$= \frac{4}{3} [(\psi_x^*)^2 \psi_x^2 + (\psi_y^*)^2 \psi_y^2 + (\psi_z^*)^2 \psi_z^2] - \frac{2}{3} [(\psi_y^*)^2 \psi_z^2 + (\psi_z^*)^2 \psi_x^2 + (\psi_x^*)^2 \psi_y^2 + \text{c.c.}], \quad (108)$$

where “c.c.” denotes the complex conjugate.

- The T_{1g} combination of second order would be [21–23]

$$(\psi_y \psi_z - \psi_z \psi_y, \psi_z \psi_x - \psi_x \psi_z, \psi_x \psi_y - \psi_y \psi_x) = (0, 0, 0). \quad (109)$$

It vanishes because T_{1g} is the antisymmetric part of $T_{1u} \otimes T_{1u}$. Hence, we cannot construct an independent fourth-order term in this way.

- The T_{2g} combination of second order is [21–23]

$$(\psi_y \psi_z + \psi_z \psi_y, \psi_z \psi_x + \psi_x \psi_z, \psi_x \psi_y + \psi_y \psi_x) \propto (\psi_y \psi_z, \psi_z \psi_x, \psi_x \psi_y), \quad (110)$$

which gives the fourth-order term

$$(\psi_y \psi_z, \psi_z \psi_x, \psi_x \psi_y)^* \cdot (\psi_y \psi_z, \psi_z \psi_x, \psi_x \psi_y) = |\psi_y|^2 |\psi_z|^2 + |\psi_z|^2 |\psi_x|^2 + |\psi_x|^2 |\psi_y|^2. \quad (111)$$

One can find three linear combinations of the three obtained terms that are more compact and more useful for discussing physics:

$$(\psi^* \cdot \psi)^2, \quad (112)$$

$$(\psi^* \cdot \psi^*)(\psi \cdot \psi), \quad (113)$$

$$\boldsymbol{\mu} \cdot \boldsymbol{\mu}, \quad (114)$$

where

$$\boldsymbol{\mu} \equiv i(\psi^* \times \psi) = i \begin{pmatrix} \psi_y^* \psi_z - \psi_z^* \psi_y \\ \psi_z^* \psi_x - \psi_x^* \psi_z \\ \psi_x^* \psi_y - \psi_y^* \psi_x \end{pmatrix}. \quad (115)$$

This quantity is odd under time reversal (since $\psi_n \mapsto \psi_n^*$) and transforms like a pseudovector (T_{1g}). The factor i has been inserted to make $\boldsymbol{\mu}$ real. We return to the meaning of $\boldsymbol{\mu}$ below.

Similarly, we can find the terms of second order in the order parameter and containing two derivatives. In this case, we require products containing one component each of $\mathbf{D}^*$, $\mathbf{D}$, ψ^* , and ψ . Now we have four distinct vectors and Eq. (106) suggests that there are four terms of this type. But $\mathbf{D}^*$ and $\mathbf{D}$ are operators so that the order of factors matters. Due to gauge invariance, we must be able to express the functional in terms of the derivatives $D_m \psi_n$ and their complex conjugates only. Equation (105) implies that these derivatives can be organized into terms transforming according to the four irreps A_{1g} , E_g , T_{1g} , and T_{2g} :

- The A_{1g} combination is $\mathbf{D} \cdot \psi$, which gives the fourth-order term $(\mathbf{D} \cdot \psi)^* (\mathbf{D} \cdot \psi)$.

- The E_g combination is $(D_x\psi_x - D_y\psi_y, (1/\sqrt{3})(2D_z\psi_z - D_x\psi_x - D_y\psi_y))$, which gives the fully symmetric term

$$\begin{aligned} & \left(D_x\psi_x - D_y\psi_y, \frac{2D_z\psi_z - D_x\psi_x - D_y\psi_y}{\sqrt{3}} \right)^* \cdot \left(D_x\psi_x - D_y\psi_y, \frac{2D_z\psi_z - D_x\psi_x - D_y\psi_y}{\sqrt{3}} \right) \\ &= \frac{4}{3} [(D_x\psi_x)^*(D_x\psi_x) + (D_y\psi_y)^*(D_y\psi_y) + (D_z\psi_z)^*(D_z\psi_z)] \\ & \quad - \frac{2}{3} [(D_y\psi_y)^*(D_z\psi_z) + (D_z\psi_z)^*(D_x\psi_x) + (D_x\psi_x)^*(D_y\psi_y) + \text{c.c.}]. \end{aligned} \quad (116)$$

- The T_{1g} combination is $(D_y\psi_z - D_z\psi_y, D_z\psi_x - D_x\psi_z, D_x\psi_y - D_y\psi_x)$, which does not vanish since $\mathbf{D}$ and $\boldsymbol{\psi}$ are distinct. This combination results in the fully symmetric term

$$\begin{aligned} & (D_y\psi_z)^*(D_y\psi_z) + (D_z\psi_y)^*(D_z\psi_y) + (D_z\psi_x)^*(D_z\psi_x) + (D_x\psi_z)^*(D_x\psi_z) \\ & + (D_x\psi_y)^*(D_x\psi_y) + (D_y\psi_x)^*(D_y\psi_x) - (D_y\psi_z)^*(D_z\psi_y) - (D_z\psi_y)^*(D_y\psi_z) \\ & - (D_z\psi_x)^*(D_x\psi_z) - (D_x\psi_z)^*(D_z\psi_x) - (D_x\psi_y)^*(D_y\psi_x) - (D_y\psi_x)^*(D_x\psi_y). \end{aligned} \quad (117)$$

- The T_{2g} combination is $(D_y\psi_z + D_z\psi_y, D_z\psi_x + D_x\psi_z, D_x\psi_y + D_y\psi_x)$, which gives the term

$$\begin{aligned} & (D_y\psi_z)^*(D_y\psi_z) + (D_z\psi_y)^*(D_z\psi_y) + (D_z\psi_x)^*(D_z\psi_x) + (D_x\psi_z)^*(D_x\psi_z) \\ & + (D_x\psi_y)^*(D_x\psi_y) + (D_y\psi_x)^*(D_y\psi_x) + (D_y\psi_z)^*(D_z\psi_y) + (D_z\psi_y)^*(D_y\psi_z) \\ & + (D_z\psi_x)^*(D_x\psi_z) + (D_x\psi_z)^*(D_z\psi_x) + (D_x\psi_y)^*(D_y\psi_x) + (D_y\psi_x)^*(D_x\psi_y). \end{aligned} \quad (118)$$

However, the four terms we have constructed are not independent because of Eq. (104), which implies that $(D_i\psi_j)^*(D_j\psi_i) = (D_j\psi_j)^*(D_i\psi_i)$ up to an irrelevant boundary term. With that, we only obtain three independent terms, which can be written as

$$(\mathbf{D} \cdot \boldsymbol{\psi})^* (\mathbf{D} \cdot \boldsymbol{\psi}) = (D_x\psi_x + D_y\psi_y + D_z\psi_z)^* (D_x\psi_x + D_y\psi_y + D_z\psi_z), \quad (119)$$

$$(D_x\psi_x)^*(D_x\psi_x) + (D_y\psi_y)^*(D_y\psi_y) + (D_z\psi_z)^*(D_z\psi_z), \quad (120)$$

$$\begin{aligned} & (D_y\psi_z)^*(D_y\psi_z) + (D_z\psi_y)^*(D_z\psi_y) + (D_z\psi_x)^*(D_z\psi_x) \\ & + (D_x\psi_z)^*(D_x\psi_z) + (D_x\psi_y)^*(D_x\psi_y) + (D_y\psi_x)^*(D_y\psi_x). \end{aligned} \quad (121)$$

More interestingly, a term linear in the magnetic induction appears. Recall that $\boldsymbol{\mu}$ defined in Eq. (115) is a time-reversal-odd pseudovector (T_{1g}). Since $\mathbf{B}$ is also a time-reversal-odd pseudovector a term proportional to

$$\mathbf{B} \cdot \boldsymbol{\mu} = \mathbf{B} \cdot i(\boldsymbol{\psi}^* \times \boldsymbol{\psi}) = i(B_x\psi_y^*\psi_z + B_y\psi_z^*\psi_x + B_z\psi_x^*\psi_y) + \text{c.c.} \quad (122)$$

is allowed. This means that the superconducting state carries a magnetization proportional to $\boldsymbol{\mu}$. It is nonzero if at least two of the components of $\boldsymbol{\psi}$ are nonzero and have a nontrivial phase difference. This is *exactly* the condition for the superconducting order parameter to break time-reversal symmetry. Microscopically, the presence of a magnetization is easy to understand since

the most natural way to realize a multi-component order parameter is by means of spin-triplet (spin-1) Cooper pairs, which couple to $\mathbf{B}$ through the Zeeman interaction.

Finally, does the same mechanism also produce a term of linear order in derivatives? We try to construct terms that contain one factor each of $\mathbf{D}$, ψ , and ψ^* . $\mathbf{D} \cdot \boldsymbol{\mu}$ is not a good idea since

$$\mathbf{D} \cdot \boldsymbol{\mu} + \text{c.c.} = \nabla \cdot \boldsymbol{\mu} - i \frac{q}{\hbar c} \mathbf{A} \cdot \boldsymbol{\mu} + \nabla \cdot \boldsymbol{\mu} + i \frac{q}{\hbar c} \mathbf{A} \cdot \boldsymbol{\mu} = 2 \nabla \cdot \boldsymbol{\mu} \quad (123)$$

is a total derivative and thus reduces to an irrelevant boundary term under the spatial integral. Hence, we instead consider

$$J \equiv \psi^* \cdot (\mathbf{D} \times \psi). \quad (124)$$

Using integration by parts, we find

$$\begin{aligned} J &= \psi_x^* D_y \psi_z + \psi_y^* D_z \psi_x + \psi_z^* D_x \psi_y + \text{c.c.} \\ &= \psi_x^* D_y \psi_z + \psi_y^* D_z \psi_x + \psi_z^* D_x \psi_y + \psi_x D_y^* \psi_z^* + \psi_y D_z^* \psi_x^* + \psi_z D_x^* \psi_y^*, \end{aligned} \quad (125)$$

up to an irrelevant boundary term. J is even under time reversal. Under proper rotations, J transforms like a scalar. However, it is the product of three vectors and is thus *odd* under inversion, i.e., it is a pseudoscalar transforming according to A_{1u} . Thus such a term is forbidden by inversion symmetry.

On the other hand, if the crystal does not possess inversion symmetry, i.e., if we are dealing with a noncentrosymmetric superconductor, the behavior under inversion is immaterial and the above term is allowed [26, 27]. Such terms of first order in derivatives are called *Lifshitz invariants*. In our example, the point group becomes O . The irreps A_{1g} and A_{1u} restricted to the subgroup O of O_h are identical and are called A_1 as an irrep of O .

Lifshitz invariants have interesting consequences: Let us consider the situation deep inside a noncentrosymmetric superconductor, where the $\mathbf{B}$ field vanishes and $\mathbf{A}$ can be gauged to zero. An ansatz $\psi(\mathbf{r}) = \psi_0 e^{i\mathbf{q} \cdot \mathbf{r}}$ generates a term proportional to

$$J = \left(\mathbf{q} - \frac{q}{\hbar c} \mathbf{A} \right) \cdot i (\psi_0^* \times \psi_0) = \mathbf{q} \cdot \boldsymbol{\mu}_0, \quad (126)$$

with $\boldsymbol{\mu}_0 \equiv i (\psi_0^* \times \psi_0)$, in the free energy, which is linear in the wave vector $\mathbf{q}$. It is only present if $\boldsymbol{\mu}_0 \neq 0$, i.e., if time-reversal symmetry is broken. In addition, there are three positive terms of second order in derivatives, Eqs. (119)–(121), which generate terms of second order in $\mathbf{q}$. This means that the free energy is minimized by a solution with $\mathbf{q} \neq 0$ and spatially rotating phase factor $e^{i\mathbf{q} \cdot \mathbf{r}}$. The conclusion is that time-reversal-symmetry-breaking states in noncentrosymmetric superconductors are *helical states* [27].

References

- [1] F. London and H. London, Proc. Roy. Soc. (London) **A149**, 71 (1935);
F. London and H. London, Physica **2**, 341 (1935)
- [2] V.L. Ginzburg and L.D. Landau, Zh. Eksp. Teor. Fiz. **20**, 1064 (1950)
- [3] L.D. Landau, Zh. Eksp. Teor. Fiz. **7**, 19 (1937)
- [4] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)
- [5] L.P. Gor'kov, Soviet Phys. JETP **9**, 1364 (1959) [Zh. Eksp. Teor. Fiz. **36**, 1918 (1959)]
- [6] M. Tinkham: *Introduction to Superconductivity*, 2nd ed. (McGraw–Hill, New York, 1996)
- [7] E.P. Gross, Il Nuovo Cimento **20**, 454 (1961)
- [8] L.P. Pitaevskii, Sov. Phys. JETP **13**, 451 (1961)
- [9] S.-K. Ma: *Modern Theory of Critical Phenomena* (Benjamin, Reading, 1976)
- [10] A. Altland and B. Simons: *Condensed Matter Field Theory*, 2nd ed. (Cambridge University Press, Cambridge, 2010)
- [11] J.D. Jackson: *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999)
- [12] A.A. Abrikosov, Soviet Phys. JETP **5**, 1174 (1957) [Zh. Eksp. Teor. Fiz. **32**, 1442 (1957)]
- [13] C. Timm: *Theory of Superconductivity* (Lecture Notes, TU Dresden, 2026)
https://tu-dresden.de/mn/physik/itp/cmt/ressourcen/dateien/skripte/Skript_Supra.pdf
- [14] R. Shimano and N. Tsuji, Ann. Rev. Condens. Matter Phys. **11**, 103 (2020)
- [15] P.W. Anderson, Phys. Rev. **110**, 827 (1958)
- [16] P.W. Anderson, Phys. Rev. **112**, 1900 (1958)
- [17] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964)
- [18] P.W. Higgs, Phys. Lett. **12**, 132 (1964)
- [19] P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
- [20] M. Sgrist and K. Ueda, Rev. Mod. Phys. **63**, 239 (1991)
- [21] M.S. Dresselhaus, G. Dresselhaus, and A. Jorio: *Group Theory—Application to the Physics of Condensed Matter* (Springer, Berlin, 2008)

- [22] M.I. Aroyo, J.M. Perez-Mato, C. Capillas, E. Kroumova, S. Ivantchev, G. Madariaga, A. Kirov, and H. Wondratschek, *Zeitschrift für Kristallographie* **221**, 15 (2006); M.I. Aroyo, A. Kirov, C. Capillas, J.M. Perez-Mato, and H. Wondratschek, *Acta Cryst.* **A62**, 115 (2006)
- [23] G. Katzer: *Character tables for point groups used in chemistry*
https://gernot-katzers-spice-pages.com/character_tables/
- [24] H. Waje, F. Jakubczyk, J. van den Brink, and C. Timm, *Phys. Rev. B* **112**, 144519 (2025)
- [25] F. Jakubczyk, J.M. Link, and C. Timm, *Phys. Rev. Research* **8**, 023363 (2026)
- [26] V.P. Mineev and K.V. Samokhin, *Sov. Phys. JETP* **78**, 401 (1994)
- [27] M. Smidman, M.B. Salamon, H.Q. Yuan, and D.F. Agterberg, *Rep. Prog. Phys.* **80**, 036501 (2017)

10 Experimental Characterization of Phase Transitions

Anna E. Böhmer
Experimental Physics IV
Faculty for Physics and Astronomy
Ruhr University Bochum, 44801 Bochum

Contents

1	Introduction	2
2	Background	2
3	Experimental identification of a phase transition	5
3.1	General considerations	5
3.2	The phase diagram of iron-based superconductors	8
4	Determination of the nature of a phase	10
4.1	Experimental techniques for identifying ordered phases	10
4.2	Determination of magnetic and structural order parameters in iron-based superconductors	15
5	Concluding remarks	18

1 Introduction

Phase transitions are an everyday phenomenon. Whether it is the melting of ice cubes, the boiling of water, or the hardening of candle wax, we all experience the sudden change of properties of a material when external conditions (such as the temperature) change. A story on the impact of a structural phase transition concerns the two modifications of Sn: β -Sn is a shiny malleable metal, perfect for making uniform buttons. α -Sn is a gray brittle semiconductor of little use. As the story goes, Napoleon's troops in 1812 marched into freezing Russian winter with temperatures as low as -40°C , where their tin uniform buttons began to disintegrate due to the transition from β -Sn to α -Sn. Without buttons to keep their coats closed, the soldiers could not keep warm and in the end Napoleon lost the campaign.

The phase transitions of correlated materials go far beyond the solid-liquid-gas aggregate states or the different structural modifications of elements and compounds. Typically, we are interested in transitions that arise from electronic or magnetic interactions. One finds diverse subtle structural transitions, but also a rich variety of magnetic phase transitions, transitions to superconductivity and even more exotic phase transitions. Identifying the presence of phase transitions and how they evolve with external “tuning parameters” (control parameters) such as pressure, magnetic field or material composition is a crucial step in understanding a material. Often, only discrete points in a phase diagram can be probed, so establishing a complete phase diagram with areas of different phases bordered by phase lines involves knowledge of constraints for phase diagrams, interpretation, and, sometimes, speculation. In addition to establishing the topology of a phase diagram, the nature of the different phases needs to be determined. This may be straightforward, but can also seem like an impossible problem in some cases. The whole arsenal of solid-state physics is often employed to understand all aspects of the nature of a phase diagram and the individual phases.

In these lecture notes, I will present aspects of the experimental characterization of phase transitions from a personal point of view. I will start with a brief description of what constitutes a phase transition and how phase transitions are classified. I will touch on the important matter of symmetries and constraints for phase diagrams. Second, I will explain how phase transitions can be identified experimentally. As an example, I will discuss the establishing of the temperature-substitution phase diagram of the iron-based material $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$ and the temperature-pressure phase diagram of FeSe. In the third part of these lecture notes, I will explain some aspects of how the nature of a phase can be determined, with a focus on structural, magnetic, and superconducting phase transitions. The iron-based materials provide again rich examples for the determination of structural and magnetic orders.

2 Background

A phase transition is a consequence of the general tendency of nature to maximize entropy. This is quantified in the general principle of minimization of free energy $F = U - TS$ (or $G = U + pV - TS$). The entropy term dominates at high temperatures and a “disordered” high-

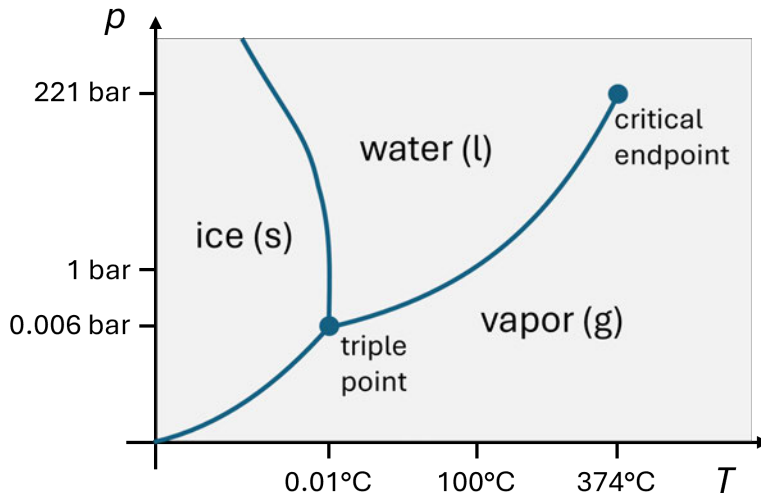


Fig. 1: The solid (s) - liquid (l) - gas (g) p - T phase diagram of H_2O . Along phase lines, two phases coexist in equilibrium. A phase transition occurs when a line is crossed. At atmospheric pressure of $p = 1$ bar, water vapor condenses into liquid water at 100°C and water freezes into ice at 0°C . At the “triple point”, all three phases coexist in equilibrium. The liquid-gas phase line terminates in a “critical endpoint”. For pressure/temperature values beyond the critical endpoint, the distinction between liquid and gas vanishes. Figure reproduced from Ref. [2].

entropy state is preferred. However, if there are states with a lower energy U these will eventually be transitioned into on lowering the temperature. At zero temperature, the entropy term no longer contributes and the ground state of a system is the one with lowest energy.

A phase is defined as a homogeneous region in a material, where intensive variables such as the specific volume, V , specific entropy, S , or particle density, and all of their derivatives vary continuously. However, under certain conditions, the equilibrium state of a material may be the coexistence of different phases. The pressure, p , temperature, T , and chemical potential, μ , of the two phases that are in contact with each other need to be equal. For given p and T , the condition for phase equilibrium is therefore $\mu_1(T, p) = \mu_2(T, p)$. In a p - T diagram (see Fig. 1 for an example), there will be a line along which the two phases may exist in equilibrium. When temperature or pressure are varied such that this line is crossed, a phase transition occurs [1]. During a phase transition, the properties of the material change abruptly. This sudden change of properties is experimentally seen as an “anomaly” in a measured curve (e.g., $V(T)$). Here, we will not discuss situations in which multiple phases coexist in equilibrium, though such situations are very important as well. Indeed, rich effects are observed in situations where the number of phases changes (such as when a solution of table salt in hot water is cooled and salt crystals form).

The Gibbs phase rule governs the possibility of phase coexistence. It says that the number of freely varying intensive variables, F , equals $F = C - P + N + 1$, where C is the number of components in a system, P the number of phases, and N the number of different ways of performing work on the system. The phase diagram in Fig. 1 is an example of a single-component system ($C=1$) with only one way of performing work (pressure-volume work) ($N=1$), therefore $F = 3 - P$. That means the areas in the phase diagram (both pressure and temperature vary

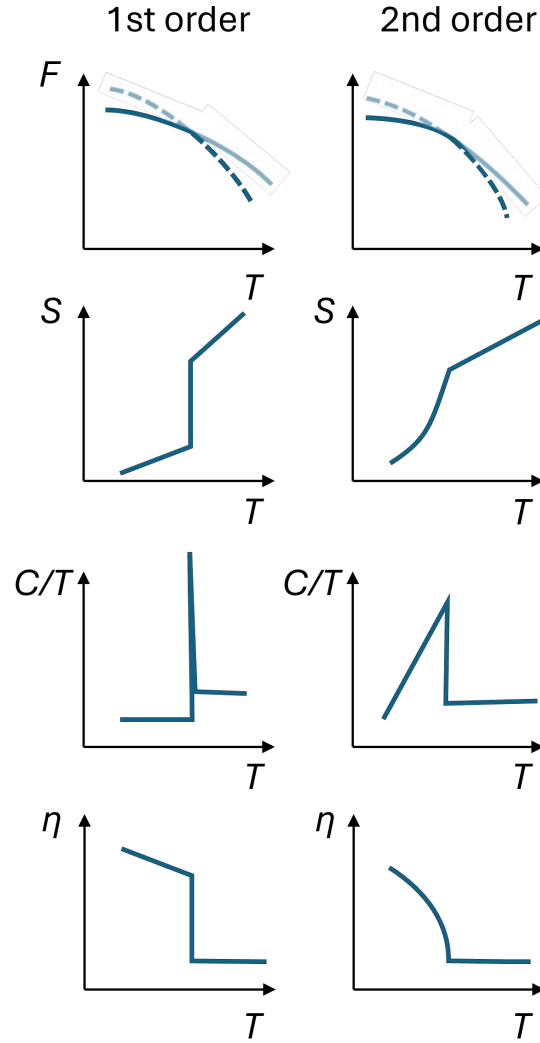


Fig. 2: Differences between first-order (discontinuous) and second-order (continuous) phase transitions. A first-order phase transition occurs, when the minimum free energy (dark blue) has a slope change, e.g., when the free energies of two phases (indicated as a dashed and a continuous line) cross. The entropy $S = \frac{\partial F}{\partial T}$ shows a discontinuity and the heat capacity $C/T = \frac{\partial S}{\partial T}$ a delta-type peak. The order parameter assumes a finite value in a discontinuous fashion on entering the ordered phase. At a second-order phase transition, the entropy shows a slope change only, and the heat capacity a discontinuity. The order parameter increases in a continuous fashion.

freely, $F=2$) have to be single-phase regions ($P=1$). Along the phase lines, temperature and pressure are linked, so that $F=1$ and two phases coexist ($P=2$). Three phases can coexist, but pressure and temperature cannot vary freely in that case ($F=0$). This is the triple point.

Phase transitions are classified according to Ehrenfest into different types. A system is described by its free energy F (or G). A phase transition is said to be of first/second order if the first/second derivatives of the free energy show a discontinuity (“a jump” or “step-like anomaly”), respectively. Thus, a phase transition is a mathematical singularity. Any derivative of a thermodynamic potential (such as F or G) is a thermodynamic quantity, and the character of a phase transition is therefore determined by measuring the evolution of thermodynamic quanti-

ties on crossing a phase line (see Fig. 2). First derivatives of the free energy are, e.g., the volume ($V = \partial G/\partial p$), the entropy ($S = -\partial F/\partial T$), or the magnetization ($M = -\partial F/\partial B$, with the appropriate variables held constant in each case). Typically considered second derivatives are the heat capacity $C_p/T = \partial S/\partial T$ or the thermal expansion coefficient $\beta = 1/V \partial V/\partial T$.

Figure 2 presents examples for the anomalies of thermodynamic quantities at first- and second-order phase transitions. A first-order phase transition results in a step in the entropy, which means a delta-like anomaly in the heat-capacity, whereas a second-order phase transition results in a step in the heat capacity, which means a slope change (“kink”) in the entropy. The finite entropy change ΔS at a first-order phase transition corresponds to a latent heat $Q = T\Delta S$. It naturally causes a hysteresis, which means that the transitions occurs at a lower temperature on cooling than on heating (or equivalently for other control parameters).

Famously, phase transitions are described in Landau theory, where the free energy is expanded in terms of an “order parameter” that describes the ordered phase. The order parameter is finite only in the ordered phase and reflects all of its symmetry properties. An example is the magnetization of a ferromagnet. In a continuous phase transition, the order parameter increases continuously from zero, whereas in a discontinuous phase transition (only first-order transition), the order parameter abruptly assumes a finite value.

It is important to determine the nature of phase transition (continuous or discontinuous), as many constraints depend on it. A phase transition can only be continuous if it breaks one symmetry [3]. Examples are a loss of a C_2 rotation in a tetragonal-to-orthorhombic structural transition, or the loss of time-reversal for a paramagnetic-to-ferromagnetic phase transition. Discontinuous phase transitions may also occur between phases of the same symmetry (e.g., liquid-gas), or change more than one symmetry element. A phase line can terminate in a critical endpoint only if there is no symmetry change between the two phases. In addition, there are multiple constraints on how different phase lines can meet in a phase diagram, see Refs. [4] or [5] for just two examples. Such constraints depend on the type of the phase transitions involved. The constraints can be extremely helpful to complete and understand a phase diagram as, e.g., in Ref. [6].

3 Experimental identification of a phase transition

3.1 General considerations

Many techniques are used to study phase transitions. Typically, a property of the material is measured as a function of an extensive variable, such as electrical resistance as a function of temperature or magnetization as a function of magnetic field. In principle, *all* properties of a system change suddenly when it undergoes a phase transition, which can be seen as an “anomaly” in such data. Therefore, the measurement of any property can indicate a phase transition. Nevertheless, some techniques are better suited to identify the presence of a phase transition than others. This depends on accessibility of the technique, resolution of the technique, ease of use, effort for sample preparation and, of course, whether the anomaly at the

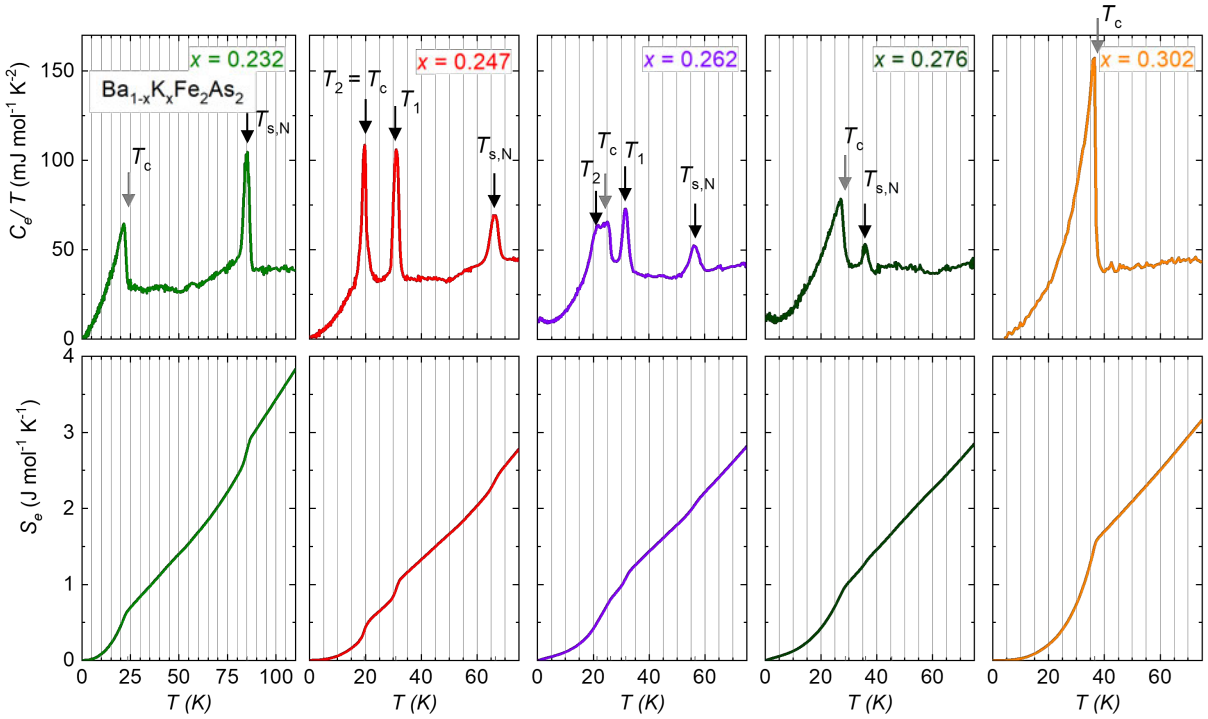


Fig. 3: Example of the electronic heat capacity of $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$ as a function of temperature. A phonon background has been subtracted from the experimental data and the electronic entropy has been calculated. Various first-order (black arrows) and second-order (grey arrows) phase transitions are observed. All transitions are broadened from an ideal shape to different degrees. Signatures in the entropy can be very subtle. Figure reproduced from [7].

phase transition is sizable in that specific case. Frequently, the electrical resistance is one of the first quantities measured when one searches for phase transitions. Other typical properties include magnetization, heat capacity, or thermal expansion. For thermodynamic quantities, a second-order phase transition should appear as a kink or step (depending on whether the quantity is a first or second derivative of the free energy). Note that the electrical resistance is a transport property and not a thermodynamic quantity. Therefore, it cannot be used in the same way to make statements about the type of a phase transition. Microscopic or local probes can, of course, also clearly show the onset of some type of order, which is associated with a phase transition. Examples are extremely diverse, such as changes of the diffraction pattern indicating a modification of the crystal structure, signs of the onset of magnetic order in muon spin resonance (μSR), nuclear magnetic resonance (NMR) or Kerr effect, or signatures of inversion symmetry breaking in second harmonic generation in optical systems.

Usually, a phase transition is identified experimentally as a sudden change in the measured property. Figure 3 shows examples of multiple first- and second-order phase transitions as seen in a heat-capacity measurement. In general, phase transitions are never infinitely sharp in experimental data, so what is theoretically a mathematical singularity will be “washed out”. Reasons can be inhomogeneities in the chemical composition of a sample, internal stresses in the sample, or inhomogeneities in the external parameters (e.g., pressure inhomogeneity for a pressure-driven phase transition). Therefore, what is ideally a discontinuity (“a step”) in a

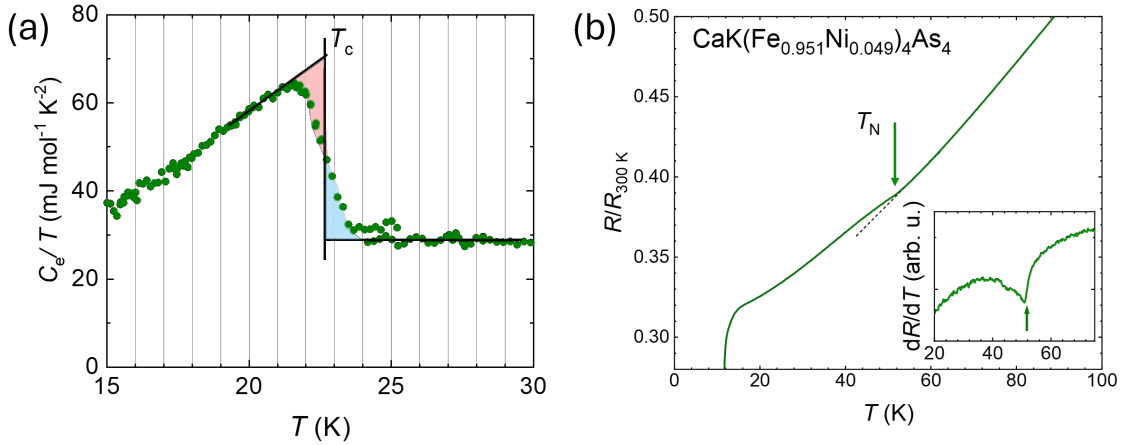


Fig. 4: (a) The entropy-conserving construction, which is a common criterion for determining the temperature of a second-order phase transition. The areas above and below the idealized curve (red and blue) should be equal. (b) Example of a second-order phase transition in a resistance measurement. In such a transport probe, defining a criterion for a phase transition is not unambiguous. The transition is seen as a deviation from the extrapolated slope of the high-temperature phase (dashed line). A sharp feature is seen in the temperature derivative. (a) based on data in Ref. [7], (b) based on data in Ref. [8].

measured quantity will become rounded, and a delta-function will become a broadened peak. As a consequence, it is not always easy to distinguish a first-order phase transition from a second-order phase transition in practice. A broadened peak may look not very different from a washed-out step. Typically, if an anomaly looks “symmetric” in rising and falling slope, one would consider the transition to be of first-order. Observation of a hysteresis is a tell-tale sign of a first-order phase transition. However, the absence of an observed hysteresis does not mean that a transition is not first-order. The latent heat may just be too small to cause a significant effect, and most experimental setups cause small differences on heating and cooling anyways, due to the instrumentation.

An additional issue is that, sometimes, a change of properties would occur gradually even in perfect condition and is therefore by definition not a phase transition. It may be called a “crossover” instead. Crossovers may, e.g., be observed beyond the critical endpoint of a first-order phase transition [9]. Due to the presence of intrinsic broadening, it can be difficult to distinguish a true phase transition from a crossover.

A phase transition temperature, T_{crit} , is uniquely defined in the case of an ideally sharp phase transition. In real data, however, one often needs to define a criterion for the T_{crit} . For thermodynamic quantities that show a peak at the phase transition, this is naturally the peak maximum. For step-like anomalies in a thermodynamic quantity (such as C/T at a second-order phase transition), one often defines the mid-point of the step in an “entropy conserving” construction (see Fig. 4). However, if the measured quantity is not a thermodynamic quantity (such as electrical resistance), the criterion can be ambiguous. The more washed out a transition is, the harder T_{crit} is to define. As a general rule, the transition temperature should be taken at the “sharpest feature” in the data. In any case, the criterion should always be defined and stated clearly, and checked for consistency between different publications.

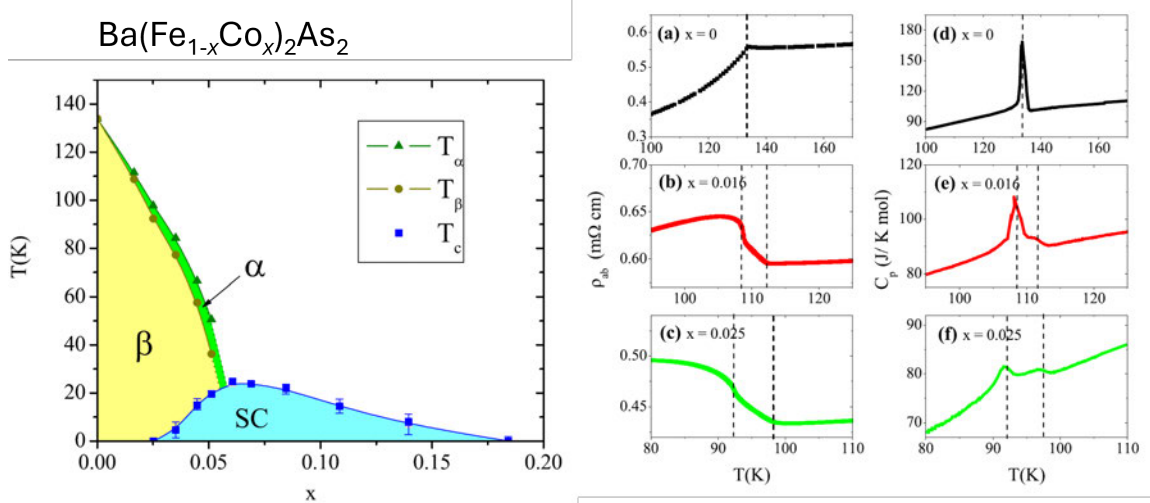


Fig. 5: *Left side: The phase diagram of $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$. The points are determined from resistivity and heat capacity measurements, examples of which are shown on the right side. For $x = 0$ a single first-order phase transition is observed. For $x = 0.016$ a higher-temperature second-order phase transition is followed by a first-order phase transition at a slightly lower temperature. Figure reproduced from Ref. [12].*

3.2 The phase diagram of iron-based superconductors

High-temperature superconductivity in iron-based compounds was discovered in 2008 [10]. Like many unconventional superconductors, iron-based materials undergo other phase transitions besides superconductivity. Superconductivity arises when these other phase transitions are suppressed to lower temperatures by applied pressure or chemical substitution.

One of the most studied iron-based systems is $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$. Anomalies in resistance, magnetization, heat capacity and thermal expansion clearly show its phase transitions [11–13] (see Fig. 5). With the thermodynamic probes, the order of the phase transitions can be determined, as explained in the previous section. Indeed, it was found that a first-order transition of the parent compound BaFe_2As_2 splits into a second-order phase transition and a lower-temperature first-order phase transition upon Co substitution. For high Co-substitution levels, this lower temperature first-order phase transition changes its character and becomes a second-order phase transition as well. These findings have consequences for theoretical modeling, which was employed, e.g., in Ref. [14, 15].

Identifying a phase transition may be challenging under certain conditions. For example, the available methods are restricted when a transition occurs under applied physical pressure or at extremely low temperatures. As an example, heat capacity measurements under applied pressure are extremely challenging [20]. Also, the sample quality can strongly affect whether a phase transition is visible in a dataset; they may be extremely smeared out so that they are invisible above a background. Some transitions are also genuinely suppressed by disorder in a low-quality sample. For example, superconductivity can be completely destroyed by disorder. The case of FeSe provides a useful example of the determination of a phase diagram using electrical resistance measurements under pressure (see Fig. 6). Figure 6(e) shows the temperature-

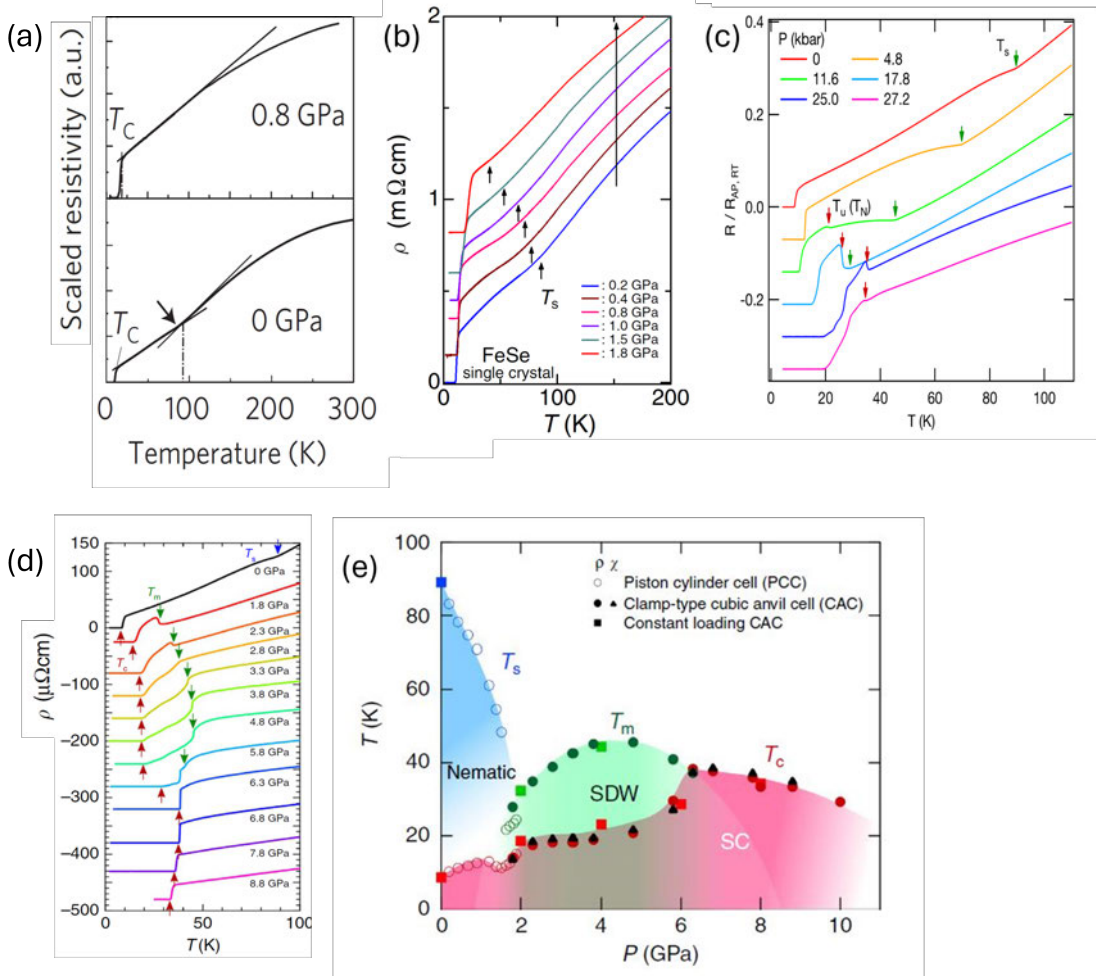


Fig. 6: Determination of the temperature-pressure phase diagram of FeSe using resistance measurements. In the first dataset in (a), the structural transition is indicated by a slope change slightly below 100 K, but not resolved under applied pressure. In the second dataset in (b), this transition could be followed to higher pressures. In the third dataset in (c), another transition at $T_u(T_N)$ could also be resolved. In (d), data to higher pressures was obtained [T_N is labelled T_m here in (d)], from which the phase diagram in (e) was compiled. (a) reproduced from [16], (b) reproduced from [17], (c) reproduced from [18], (d,e) reproduced [19]

pressure phase diagram of FeSe. FeSe [21] is an iron-based superconductor whose superconductivity at $T_c = 8$ K was discovered in 2008. FeSe also shows another phase transition at $T_s = 90$ K, whose signature is a subtle kink in resistance. This transition is identified as a tetragonal-to-orthorhombic structural transition via x-ray diffraction and considered an electronically driven nematic transition. Early resistance measurements under pressure showed that the superconducting transition temperature of FeSe increases dramatically up to 40 K at a pressure of ~ 8 GPa. However, the fate of the structural phase transition under pressure was not clear [16]. Surprisingly, μ SR first detected the emergence of magnetic order with increasing transition temperature T_N under increasing pressure, where no magnetic order is found at ambient pressure [22]. Several years later, higher-quality samples permitted to follow the resistance anomaly at T_s under pressure, revealing that the structural phase transition was suppressed [17]. But these resistance measurements did not detect any sign of the magnetic phase transition

at T_N . Thus, it was thought possible that the structural and magnetic phase transitions do not occur in the same sample. Only when signatures of both phase transitions were observed in the same measurement [18], it was confirmed that the phase diagram of FeSe under pressure was qualitatively different from the phase diagram of $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$. Finally, the data were extended to higher pressure ranges [19]. These data revealed that the signature of the magnetic phase transition in resistance measurements changes from an upturn on decreasing temperature to a downturn, respectively (see Fig. 6(e)). Such a change in shape of the resistance anomaly is indeed very much possible as the signature of a transition in electrical transport measurements depends on many microscopic details.

4 Determination of the nature of a phase

Determining the nature of a phase is very much synonymous with identifying its order parameter (or the symmetry changes at the transition). This can be straightforward, or prove to be an almost impossible problem. Some typical questions that need to be answered for the determination of the nature of a phase are:

- Does the ordered phase change the crystal structure, e.g., decrease rotational symmetries of the lattice or reduce translation symmetry (e.g., by doubling a lattice parameter)?
- Is the ordered phase magnetic in nature (thus, break time-reversal symmetry)?
- Is the ordered phase superconducting?

There are cases where the answer is 'none of the above', and a more exotic type of order is realized. Less specific, yet important questions for the characterization of a phase are:

- Does the order parameter correspond to a periodic modulation of a quantity, e.g., a charge-density wave or antiferromagnetism, and which wave-vector $\mathbf{Q}$ characterizes it? Or is the order parameter homogeneous ($\mathbf{Q} = 0$)?
- Is the order even long-range (i.e., the entire sample is well ordered), or does the order only form in "patches" with a limited spatial extent ("correlation length")?

4.1 Experimental techniques for identifying ordered phases

Different techniques are suitable to answer the salient questions on the nature of a phase. Changes in the crystal structure are mainly addressed with diffraction techniques (see Fig. 7). A (slight) distortion of the unit cell, which reduces its symmetry, results in the splitting of a Bragg peak. For example, a tetragonal crystal has only a single in-plane lattice parameter, such that the (200) and (020) Bragg peaks occur at the same diffraction angle. However, in an orthorhombic phase with two distinct in-plane lattice parameters, (200) and (020) occur at different diffraction angles. A change of periodicity, such as the doubling of the unit cell along a specific direction, will lead to the emergence of new "superstructure" diffraction peaks. Diffraction techniques

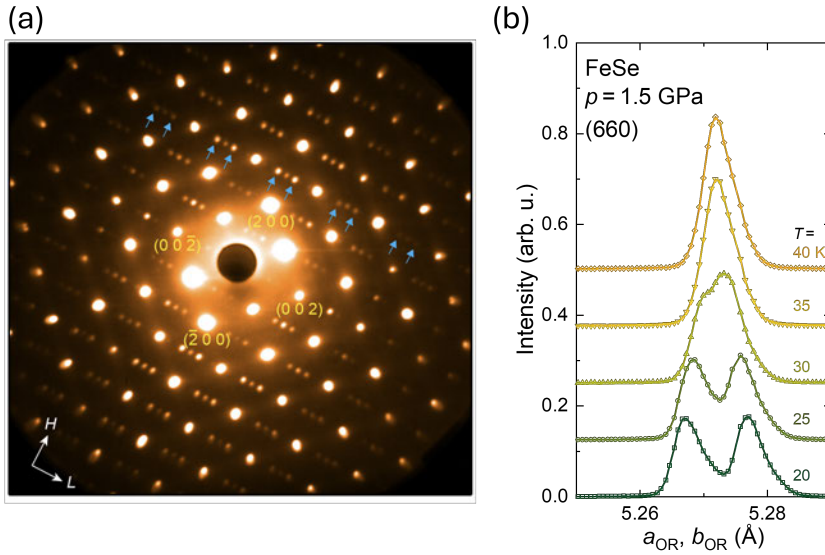


Fig. 7: Examples of diffraction data revealing a structural phase transition. (a) Diffraction pattern of LaTe_3 . The small peaks marked by blue arrows (superstructure) result from a charge-density wave transition which leads to an approximate tripling of the c lattice parameter. (b) A distortion of the unit cell can be revealed by a peak splitting. In this case of FeSe at 1.5 GPa of applied pressure, the distortion sets in at approximately 35 K. (a) Reproduced from Ref. [24], (b) based on data in Ref [25].

can thus inform on the $\mathbf{Q}$ -vector of an order parameter, as it leads to a periodic modulation, i.e., a superstructure. For example, magnetic diffraction can clarify the wavevector $\mathbf{Q}$ of an antiferromagnetic order. Diffraction techniques can also inform on a finite correlation length of the order from the peak width, see Ref. [23] for an example.

For magnetic systems, a variety of specialized probes exist. Some techniques are particularly suitable to detect time-reversal-symmetry breaking with very high sensitivity, namely μSR or the magneto-optical Kerr effect. Other techniques probe the periodic internal magnetic fields that arise from the ordered magnetic moments in a crystal (“hyperfine fields”). In techniques such as μSR , NMR, Mössbauer spectroscopy, or ESR, the Zeeman effect of these magnetic fields on a “probe” magnetic moment (implanted muon, nuclear magnetic moment, electron spin) is measured.

Nevertheless, macroscopic techniques that average the entire sample, momentum, and energy are often in the first line when the nature of a phase is studied. Magnetization measurements as a function of temperature and field are extremely useful for characterizing magnetic phases [26, 27], see Fig. 8. A ferromagnetic transition is most clearly revealed by the emergence of finite magnetization on decreasing temperature in a small applied field, whereas the magnetic susceptibility is expected to follow the Curie-Weiss law $\chi = C/(T - T_C)$, with T_C the Curie temperature. An “Arrott-plot” [28] is employed to accurately determine the Curie temperature from magnetization measurements. The ferromagnetic phase is also clearly characterized by its hysteretic response to applied magnetic fields. An antiferromagnetic transition has a classical signature of a cusp in the susceptibility-vs-temperature data and an anisotropic response below this temperature. However, one needs to distinguish materials with localized

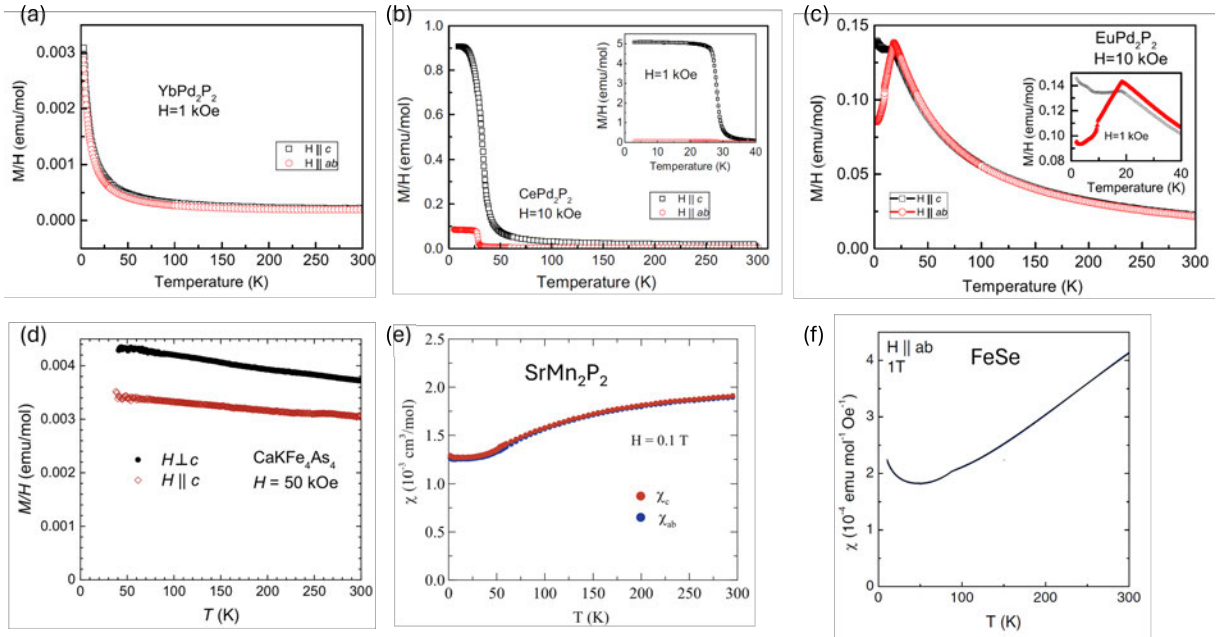


Fig. 8: Examples of magnetization data. The top row shows magnetic compounds with localized magnetic moments, (a) a paramagnet (b) a ferromagnetic transition at approximately 30 K, (c) an antiferromagnetic transition around 20 K (followed by another transition at lower temperatures). The bottom row shows itinerant magnets, (d) a paramagnet (data below 35 K is missing because the material becomes superconducting below that temperature) and (e) an antiferromagnetic transition. The material in (f) remains paramagnetic at all temperatures and the kink is related to a structural transition. (a-c) reproduced from [29], (d) reproduced from [30], (e) reproduced from [31], (f) reproduced from [32]

magnetic moments from itinerant magnets (where the magnetic moments are carried by the conduction electrons). In the latter, the magnetic susceptibility can differ dramatically from the localized (textbook-like) behavior and an antiferromagnetic transition can have very different or even extremely subtle signatures. Localized-moment paramagnets, that do not order magnetically, exhibit the characteristic Curie law of the susceptibility $\chi = C/T$, whereas the Pauli paramagnetic susceptibility of conduction electrons is temperature independent in simple cases. Identifying superconductivity is notorious. Frequently and famously, superconductivity has been reported for materials that turned out not to be superconducting in the end. These include spectacular claims of exceptional high-temperature superconductivity [33], but also other circumstances [34]. The majority of such claims were solid research contributions and the scientific community worked well to confirm or disprove them. On the other hand, superconductivity has also been missed in some instances and rediscovered only years later. The fact that superconductivity has so often been misidentified is intrinsic to its unusual properties. There is a traditional 'triad' of macroscopic experiments, which consists of resistance, magnetic susceptibility and heat capacity measurements, see Fig. 9. Only if the characteristic signatures of superconductivity have been observed in all three experiments, while taking care of the pitfalls of each of the techniques, a well-founded claim that a material is a superconductor can be made.

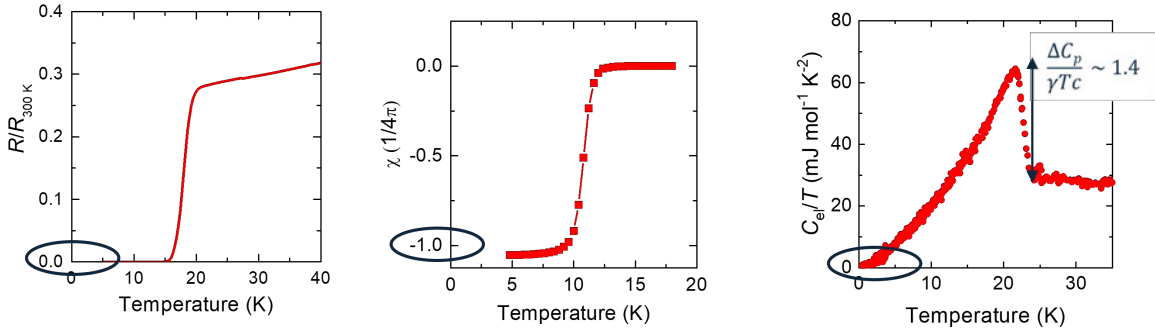


Fig. 9: The classical three macroscopic experimental results that help identify a superconducting transition. (a) zero resistance, (b) full diamagnetic shielding, (c) sizable heat capacity jump and zero residual heat capacity. (a,b) based on data in Ref. [35], (c) based on data in Ref. [7].

The observation of zero resistance marked the discovery of superconductivity. Very often, zero resistance (or even a drop of resistance) at a transition temperature T_c that is suppressed in magnetic fields is taken as a signature of superconductivity. In most cases where a new superconducting material is found, this is indeed the starting point. However, different effects play a role. First of all, zero resistance only means that there is one current path which is superconducting. This could be indeed a path along grain boundaries, a distributed minority phase, or even just a sample's surface. Also, a careful four-probe measurement is necessary, to avoid misleading effects when a sample is not homogeneous. However, superconductivity could be missed in resistance measurements in case a material has a very low critical current density that is exceeded in the experiment.

Second, the Meissner effect (a magnetic susceptibility of -1 , irrespective of how the superconducting state is reached) is the defining property of a superconductor. Experiments are performed in two manners: 'field cooled', which means that a magnetic field is applied before the sample is cooled through its superconducting transition, or 'zero-field cooled', where a magnetic field is only applied after the sample has been cooled into the superconducting state (see Fig. 10). There are many reasons why a superconductor may not show a susceptibility of -1 experimentally. In many cases, vortex pinning prevents finding a susceptibility of -1 in a field-cooled experiment on a type-II superconductor. Namely, the expulsion of field lines (the epitome of the Meissner effect) does not occur because vortices cannot move freely. For the zero-field cooled experiment, the applied magnetic field needs to be lower than the lower critical field H_{c1} (or the field of first flux penetration) for the susceptibility to reach -1 . In new materials, H_{c1} may not be accurately known. On the other hand, demagnetization effects may lead to the observation of a value much smaller than -1 . Demagnetization effects are only negligible when a thin plate-like or needle-like sample is measured with field applied along its long dimension. Finally and naturally, the sample mass needs to be sufficiently well known to determine the absolute value of the susceptibility. However, there are cases when this is not possible.

A fundamental effect that limits the conclusions to be drawn from a zero-field cooled magnetization experiment is that only the sample surface is probed. A 'glass ball wrapped in Aluminum'

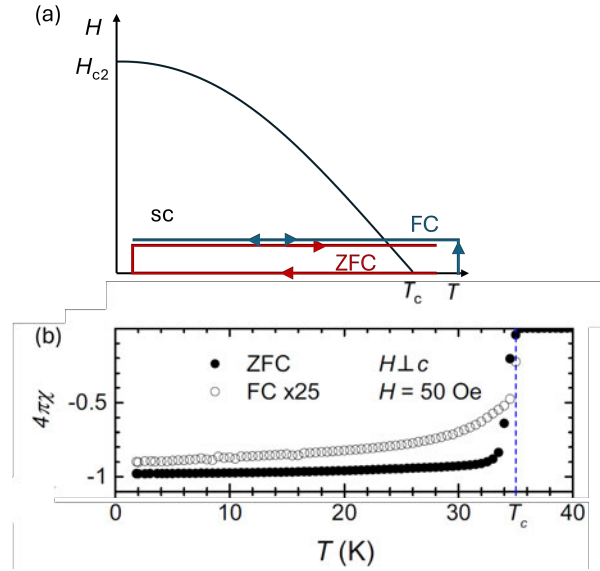


Fig. 10: (a) Paths in the temperature-field plane that correspond to a field-cooled (FC) and zero-field cooled (ZFC) magnetization measurement for a superconductor. (b) Example of a FC and ZFC measurement. Note that the FC data is multiplied by a factor of 25. (b) Adapted from Ref. [30].

would show a susceptibility of -1 , as the superconducting outer layer already shields the inside. Therefore, the value of the susceptibility in such an experiment is more appropriately called superconducting shielding fraction and should not be called superconducting volume fraction (as is often the case). In the end, observation of full magnetic flux expulsion is often intrinsically hampered by vortex pinning, and a well-measured shielding fraction of -1 is an indication for superconductivity, but not a definite proof.

The third important quantity to experimentally prove superconductivity in a material is the heat capacity. It is intrinsically a probe of the sample volume. A heat capacity curve similar to the classic curve derived in BCS theory is a tell-tale sign of a homogeneous superconductor. Two quantities are particularly important: First, the size of the step in heat capacity should be of the order of γT_c (BCS value: $\Delta C_p = 1.43\gamma T_c$, though 'strong-coupling' can increase this value and multi-band superconductivity can decrease it). If the Sommerfeld coefficient γ can be determined, this is an extremely useful sign for true bulk superconductivity. Second, the low-temperature limit of the heat capacity C/T should approach zero, which proves a fully gapped superconducting state. While there are cases when a superconductor intrinsically has a finite C/T in the zero temperature limit, more often this is a sign of sample inhomogeneity. However, a pitfall here is that the electronic heat capacity is added to a phonon background which is often too large to subtract accurately, so that the comparison with the BCS curve is difficult.

Overall, only a combination of different techniques can unambiguously prove superconductivity (zero-resistance, strong diamagnetism and BCS-like heat capacity). Some of these are difficult to implement in specific cases (e.g., under applied pressure) so that this classic trio may not be applied and a clear identification of superconductivity can be hard work, though not impossible.

There are many more 'exotic' order parameters, whose identification needs creative and knowledgeable usage of the wide range of techniques of solid-state physics. To give just one example, 'ferroaxial order' was recently proposed to be identified using specific elastotransport coefficients [36–38]

On concluding this section, I want to mention a famous case where the identification of the order parameter of a phase has eluded the community for almost 40 years and up to date, even though the presence of a phase transition is crystal clear [39]. In URu_2Si_2 , a second-order phase transition is clearly observed at $T_{\text{HO}} = 17 \text{ K}$ [40]. The heat capacity shows a large step, resistance and magnetization show clear anomalies and phase lines can be clearly followed on applied pressure or magnetic field. However, this phase is shown to neither magnetic order, nor superconductivity (the material becomes superconducting at a much lower temperature) and also not a structural phase transition. Many other order parameters have been proposed but could not be experimentally confirmed [39]. Indeed, it seems that all imaginable techniques of solid-state physics have been thrown at this problem, and the phase still remains unidentified. It is aptly named "hidden order".

4.2 Determination of magnetic and structural order parameters in iron-based superconductors

The family of iron-based superconductors provides interesting examples for the determination of structural and antiferromagnetic order parameters. After the identification of higher-temperature phase transitions in the iron-based superconductors, some signatures of which were noted in the discovery paper in 2008 [10], a concerted effort led to the determination of the nature of the order parameters. As LaFeAsO was the first material in the family to be studied, here the identification was attempted first and proceeded quickly. Some evidence for a stripe-type antiferromagnetic order was presented only months after the discovery [41]. And slightly later only, neutron diffraction confirmed the stripe-type antiferromagnetic order ("spin density wave") in LaFeAsO [42]. The same work also identified a tetragonal-to-orthorhombic structural transition about 15 K higher than the antiferromagnetic transition. Similarly, diffraction was employed to determine the nature of the ordered phases in $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$ [43], which turned out to be analogous. Figure 11 presents schematically this stripe-type antiferromagnetic phase and the related tetragonal-to-orthorhombic structural distortion. Yet, the determination of ordered phases proved to be more complex in other iron-based systems.

In FeSe under pressure, two phase transitions (apart from superconductivity) were also identified. The tetragonal-to-orthorhombic structural transition is also present at ambient pressure. The electrical resistance data showed that this transition is suppressed by applied pressure and, eventually, x-ray diffraction could determine the evolution of the structural distortion as a function of temperature and pressure [45, 25]. A second phase transition was shown to emerge under applied pressure. Local magnetic probes such as NMR [46, 47], μSR [22] and time-domain Mössbauer spectroscopy [45] clearly showed that this phase is antiferromagnetic in nature. They also showed indications for stripe-type magnetic order (e.g., [48, 49]). However,

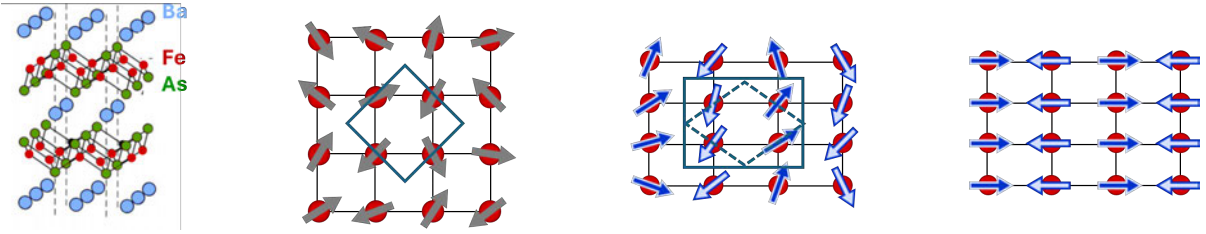


Fig. 11: Depiction of the characteristic phase transitions in iron-based materials. (a) Sketch of the crystal structure with the Fe atoms indicated in red. (b) View of the Fe plane in the tetragonal paramagnetic state. The unit cell is indicated by the turquoise square, Fe-Fe bonds by black lines. (c) The orthorhombic state with arises from a distortion of the tetragonal unit cell. The “old” unit cell is indicated as the dashed line. The orthorhombic unit cell is twice as large and indicated by the continuous line. (d) Sketch of the stripe-type antiferromagnetic state. Figure adapted from [44].

unambiguous determination of the magnetic order and its $\mathbf{Q}$ -vector from a diffraction technique could not be obtained yet. Magnetic neutron diffraction could not resolve the antiferromagnetic phase, most likely due to a combination of small available samples and small ordered moments. FeSe also shows a phase transition to a completely different crystal structure, which is unrelated to electronic correlations but “terminates” the phase diagram [25].

Other interesting magnetic phases were also identified in the iron-based materials. In the determination of all of these phases, theoretical predictions guided the experiments by proposing likely order parameters. Yet, the experiment always has the final say. A small phase pocket of a tetragonal magnetic phase was found in hole-doped iron-based superconductors, such as $\text{Sr}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$ [50] (after transport measurements indicated the same feature in pressurized samples [51]). Even though this phase is antiferromagnetically ordered, it turned out to be an example where neutron diffraction alone was not able to uniquely identify the order parameter. In this case, Mössbauer spectroscopy, a local magnetic probe, was the decisive experiment to identify the phase [52]. The new phase corresponds to a superposition of two magnetic order parameters with different $\mathbf{Q}$ -vectors (see Fig. 12). Neutron diffraction cannot easily differentiate between such a phase and a phase in which the sample splits into different domains, in each of which the order parameter only has a single $\mathbf{Q}$ -vector.

Finally, we will discuss the identification of another type of magnetic order. It occurs in Co and Ni substituted $\text{CaKFe}_4\text{As}_4$ [8]. In this material, a phase transition was identified by a subtle kink in resistance at around 50 K. Heat capacity showed a tiny second-order jump. No anomaly in magnetization could be resolved. So, there was clear proof for a well-defined phase transition but the order parameter remained to be identified. Meier et al. [8], used a variety of techniques alongside guidance from theoretical modeling (see Fig. 13).

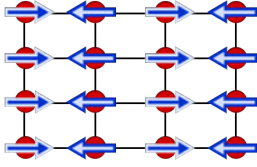
- Single-crystal x-ray diffraction showed neither the emergence of superstructure peaks (so no doubling, etc., of the unit cell), nor splitting or change in peak profiles which would indicate a distortion of the unit cell. Thus, a structural phase transition was excluded.
- Mössbauer spectroscopy proved the presence of an in-plane hyperfine field at the Fe sites, and indicated that all Fe atoms feel the same value of the hyperfine field.

Magnetic order:

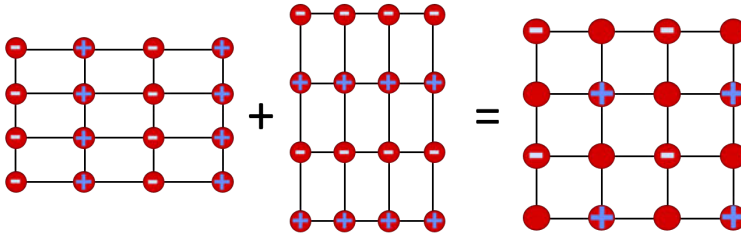
$$\mathbf{M}(\mathbf{r}) = M_1 \cos(Q_x \mathbf{r}) + M_2 \cos(Q_y \mathbf{r})$$

$$\text{with } Q_x = (\pi, 0), Q_y = (0, \pi)$$

(a)

Stripe-type magnetic order: $M_2 = 0$ 

(b)

Spin-charge density wave: $M_1 = \pm M_2$ 

(c)

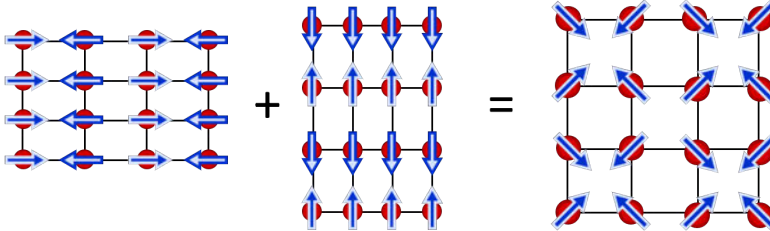
Spin vortex crystal: $|M_1| = |M_2|$ and $M_1 \perp M_2$ 

Fig. 12: Complex magnetic orders realized in iron-based materials, the generally “multi- Q ” order parameter is noted. (a) Stripe-type magnetic state with a “single- Q ” order parameter. (b) Spin-charge density wave state, which can be seen as a superposition of two stripe-type orders with different Q . (c) Depiction of the so-called spin-vortex crystal state realized in substituted $\text{CaKFe}_4\text{As}_4$.

- Nuclear magnetic resonance showed that half of the As nuclei experience no hyperfine field, while the other half experiences a hyperfine field along the c -axis with alternating direction.

From the six likely magnetic orders proposed by theoretical modeling, only one is consistent with these observations. Thus, the so-called spin-vortex crystal magnetic order [see Fig. 12(c)] was determined in Ni-substituted $\text{CaKFe}_4\text{As}_4$ [8]. Note that an unbiased symmetry analysis is possible based on the fact that the transition is of second-order nature. Therefore, the symmetry

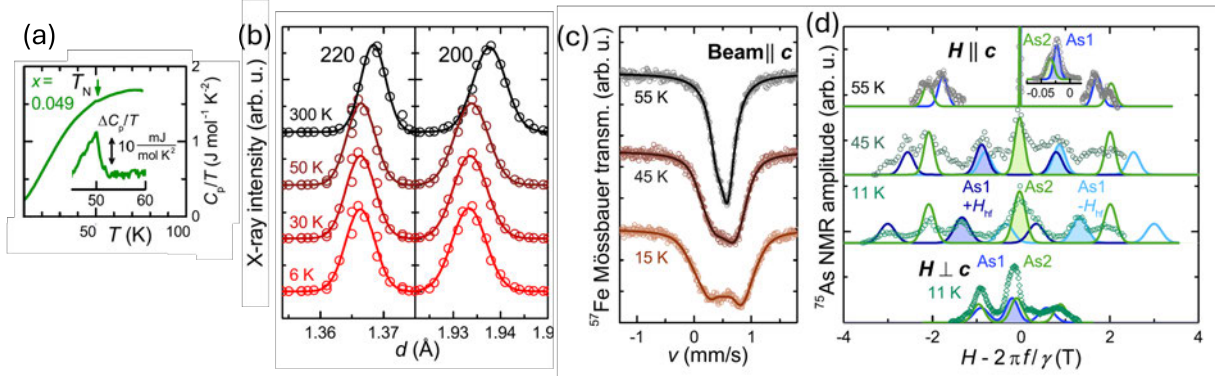


Fig. 13: Data permitting the identification of the spin-vortex crystal magnetic phase in $\text{Ca}(\text{Fe}_{0.951}\text{Ni}_{0.049})_4\text{As}_4$. (a) Heat capacity (b) X-ray diffraction (c) Mössbauer spectroscopy and (d) nuclear magnetic resonance. Figure adapted from Ref. [53].

group of the low-temperature phase has to differ from the symmetry group of the parent phase by the loss of a single symmetry element. Then, all possible magnetic groups that correspond to the reduction of the parent symmetry group by one element are examined and checked for their consistency with the experimental results [54]. This approach also leads to the conclusion of the spin-vortex crystal magnetic phase.

Finally, neutron scattering was performed and confirmed the ordered state to be the hedgehog spin-vortex crystal centered on the As1 atom, with alternate stacking along the c -axis, completely in agreement with the previous conclusions [55].

5 Concluding remarks

The identification of phase transitions and the determination of the corresponding order parameters is a central topic in condensed matter physics. Electronic correlations can lead to structural, magnetic, superconducting, or other “exotic” phase transitions, and a rich phenomenology. There are many experimental and theoretical techniques to experimentally characterize phase transitions and ordered phases, all with their own strengths and pitfalls. In some cases, long-standing and niche techniques may provide a breakthrough while new techniques are continuously being developed and optimized. It is an important skill to find the right technique for the problem at hand, given restrictions, e.g., on available samples and necessary experimental conditions. Such an effort is of high value as the knowledge of its phase diagram is the basis for understanding a correlated-electron material.

References

- [1] L.D. Landau and E.M. Lifschitz: *Lehrbuch der Theoretischen Physik V: Statistische Physik* (Akademie Verlag, Berlin, 1975) Chapter VIII
- [2] Wikipedia: *Phasendiagramm* Online; accessed 01-July-2026
<https://de.wikipedia.org/wiki/Phasendiagramm>
- [3] L.D. Landau and E.M. Lifschitz: *Lehrbuch der Theoretischen Physik V: Statistische Physik* (Akademie Verlag, Berlin, 1975) Chapter XIV
- [4] S.K. Yip, T. Li, and P. Kumar, Phys. Rev. B **43** (1991)
- [5] V. Taufour, U.S. Kaluarachchi, and V.G. Kogan, Phys. Rev. B **94**, 060410(R) (2016)
- [6] K. Semeniuk, D. Hafner, P. Khanenko, T. Lühmann, J. Banda, J.F. Landaeta, C. Geibel, S. Khim, E. Hassinger, and M. Brando, Phys. Rev. B **107**, L220504 (2023)
- [7] A.E. Böhmer, F. Hardy, L. Wang, T. Wolf, P. Schweiss, and C. Meingast, Nat. Commun. **6**, 8911 (2015)
- [8] W.R. Meier, Q.-P. Ding, A. Kreyssig, S.L. Bud'ko, A. Sapkota, K. Kothapalli, V. Borisov, R. Valentí, C.D. Batista, P.P. Orth, R.M. Fernandes, A.I. Goldman, Y. Furukawa, A.E. Böhmer, and P.C. Canfield, npj Quantum Mater. **3**, 5 (2018)
- [9] E. Gati, M. Garst, R.S. Manna, U. Tutsch, B. Wolf, L. Bartosch, H. Schubert, T. Sasaki, J.A. Schlueter, and M. Lang, Sci. Adv. **2**, e1601646 (2016)
- [10] Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, J. Am. Chem. Soc. **130**, 3296 (2008)
- [11] N. Ni, M.E. Tillman, J.-Q. Yan, A. Kracher, S.T. Hannahs, S.L. Bud'ko, and P.C. Canfield, Phys. Rev. B **78**, 214515 (2008)
- [12] J.-H. Chu, J.G. Analytis, C. Kucharczyk, and I.R. Fisher, Phys. Rev. B **79**, 014506 (2009)
- [13] C. Meingast, F. Hardy, R. Heid, P. Adelman, A. Böhmer, P. Burger, D. Ernst, R. Fromknecht, P. Schweiss, and T. Wolf, Phys. Rev. Lett. **108**, 177004 (2012)
- [14] R.M. Fernandes, D.K. Pratt, W. Tian, J. Zarestky, A. Kreyssig, S. Nandi, M.G. Kim, A. Thaler, N. Ni, P.C. Canfield, R.J. McQueeney, J. Schmalian, and A.I. Goldman, Phys. Rev. B **81**, 140501 (2010)
- [15] R.M. Fernandes, A.V. Chubukov, J. Knolle, I. Eremin, and J. Schmalian, Phys. Rev. B **85**, 024534 (2012)

- [16] S. Medvedev, T.M. McQueen, I.A. Troyan, T. Palasyuk, M.I. Eremets, R.J. Cava, S. Naghavi, F. Casper, V. Ksenofontov, G. Wortmann, and C. Felser, *Nat. Mater.* **8**, 630 (2009)
- [17] K. Miyoshi, K. Morishita, E. Mutou, M. Kondo, O. Seida, K. Fujiwara, J. Takeuchi, and S. Nishigori, *J. Phys. Soc. Jpn.* **83**, 013702 (2014)
- [18] T. Terashima, N. Kikugawa, S. Kasahara, T. Watashige, T. Shibauchi, Y. Matsuda, T. Wolf, A.E. Böhmer, F. Hardy, C. Meingast, H. v. Löhneysen, and S. Uji, *J. Phys. Soc. Jpn.* **84**, 063701 (2015)
- [19] J.P. Sun, K. Matsuura, G.Z. Ye, Y. Mizukami, M. Shimozawa, K. Matsubayashi, M. Yamashita, T. Watashige, S. Kasahara, Y. Matsuda, J.Q. Yan, B.C. Sales, Y. Uwatoko, J.G. Cheng, and T. Shibauchi, *Nat. Commun.* **7**, 12146 (2016)
- [20] Y. Zheng, Y. Wang, F. Hardy, A.E. Böhmer, T. Wolf, C. Meingast, and R. Lortz, *Phys. Rev. B* **89**, 054514 (2014)
- [21] A.E. Böhmer and A. Kreisel, *J. Phys.: Condens. Matter* **30**, 023001 (2018)
- [22] M. Bendele, A. Amato, K. Conder, M. Elender, H. Keller, H.-H. Klauss, H. Luetkens, E. Pomjakushina, A. Raselli, and R. Khasanov, *Phys. Rev. Lett.* **104**, 087003 (2010)
- [23] K. Cho, M. Kończykowski, S. Teknowijoyo, M.A. Tanatar, J. Guss, P.B. Gartin, J.M. Wilde, A. Kreyssig, R.J. McQueeney, A.I. Goldman, V. Mishra, P.J. Hirschfeld, and R. Prozorov, *Nat. Commun.* **9**, 2796 (2018)
- [24] A. Kogar, A. Zong, P.E. Dolgirev, X. Shen, J. Straquadine, Y.-Q. Bie, X. Wang, T. Rohwer, I.-C. Tung, Y. Yang, R. Li, J. Yang, S. Weathersby, S. Park, M.E. Kozina, E.J. Sie, H. Wen, P. Jarillo-Herrero, I.R. Fisher, X. Wang, and N. Gedik, *Nat. Phys.* **16**, 159 (2020)
- [25] A.E. Böhmer, K. Kothapalli, W.T. Jayasekara, J.M. Wilde, B. Li, A. Sapkota, B.G. Ueland, P. Das, Y. Xiao, W. Bi, J. Zhao, E.E. Alp, S.L. Bud'ko, P.C. Canfield, A.I. Goldman, and A. Kreyssig, *Phys. Rev. B* **100**, 064515 (2019)
- [26] S. Mugiraneza and A.M. Hallas, *Commun. Phys.* **5**, 95 (2022)
- [27] D.C. Johnston, *Phys. Rev. B* **91**, 064427 (2015)
- [28] A. Arrott, *Phys. Rev.* **108**, 1394 (1957)
- [29] G. Drachuck, A.E. Böhmer, S.L. Bud'ko, and P.C. Canfield, *J. Magn. Magn. Mater.* **417**, 420 (2016)
- [30] W.R. Meier, T. Kong, U.S. Kaluarachchi, V. Taufour, N.H. Jo, G. Drachuck, A.E. Böhmer, S.M. Saunders, A. Sapkota, A. Kreyssig, M.A. Tanatar, R. Prozorov, A.I. Goldman, F.F. Balakirev, A. Gurevich, S.L. Bud'ko, and P.C. Canfield, *Phys. Rev. B* **94**, 064501 (2016)

- [31] N.S. Sangeetha, S. Pakhira, Q.-P. Ding, L. Krause, H.-C. Lee, V. Smetana, A.-V. Mudring, B.B. Iversen, Y. Furukawa, and D.C. Johnston, *Proc. Natl. Acad. Sci. USA* **118**, e2108724118 (2021)
- [32] A.E. Böhmer, F. Hardy, F. Eilers, D. Ernst, P. Adelman, P. Schweiss, T. Wolf, and C. Meingast, *Phys. Rev. B* **87**, 180505 (2013)
- [33] S. Lee, J.-H. Kim, and Y.-W. Kwon: *The First Room-Temperature Ambient-Pressure Superconductor*, arXiv:2307.12008 (2023)
- [34] E.A. Yelland, S.M. Hayden, S.J.C. Yates, C. Pfleiderer, M. Uhlarz, R. Vollmer, H. v. Löhneysen, N.R. Bernhoeft, R.P. Smith, S.S. Saxena, and N. Kimura, *Phys. Rev. B* **72**, 214523 (2005)
- [35] A.E. Böhmer, A. Sapkota, A. Kreyssig, S.L. Bud'ko, G. Drachuck, S.M. Saunders, A.I. Goldman, and P.C. Canfield, *Phys. Rev. Lett.* **118**, 107002 (2017)
- [36] E. Day-Roberts, R.M. Fernandes, and T. Birol, *Phys. Rev. Lett.* **134**, 016401 (2025)
- [37] Q. Jiang, E. Day-Roberts, B. Gonzalez, A. Das, D.H. Torchinsky, T. Birol, R.M. Fernandes, and I.R. Fisher: *Evidence for ferroaxial order in 1T-TiSe₂ via elastoresistivity measurements* arXiv:2603.14613 (2026)
- [38] S. Edwards, E. Rosenberg, I. Maccari, J. Wen, C. Hu, X. Xu, J.-W. Kim, P.J. Ryan, R.M. Fernandes, F. de Juan, M.N. Gastiasoro, and J.-H. Chu: *Ferroaxial and nematic transitions in the charge density wave phase of 1T-TiSe₂*, arXiv:2603.14614 (2026)
- [39] J. Mydosh and P. Oppeneer, *Phil. Mag.* **94**, 3642 (2014)
- [40] T.T.M. Palstra, A.A. Menovsky, J. v. d. Berg, A.J. Dirkmaat, P.H. Kes, G.J. Nieuwenhuys, and J.A. Mydosh, *Phys. Rev. Lett.* **55**, 2727 (1985)
- [41] J. Dong, H.J. Zhang, G. Xu, Z. Li, G. Li, W.Z. Hu, D. Wu, G.F. Chen, X. Dai, J.L. Luo, Z. Fang, and N.L. Wang, *Europhys. Lett.* **83**, 27006 (2008)
- [42] C. de la Cruz, Q. Huang, J.W. Lynn, J. Li, W.R. II, J.L. Zarestky, H.A. Mook, G.F. Chen, J.L. Luo, N.L. Wang, and P. Dai, *Nature* **453**, 899 (2008)
- [43] D.K. Pratt, W. Tian, A. Kreyssig, J.L. Zarestky, S. Nandi, N. Ni, S.L. Bud'ko, P.C. Canfield, A.I. Goldman, and R.J. McQueeney, *Phys. Rev. Lett.* **103**, 087001 (2009)
- [44] A. Böhmer and A. Kreyssig, *Physik in unserer Zeit* **48**, 70 (2017)
- [45] K. Kothapalli, A.E. Böhmer, W.T. Jayasekara, B.G. Ueland, P. Das, A. Sapkota, V. Taufour, Y. Xiao, E.E. Alp, S.L. Bud'ko, P.C. Canfield, A. Kreyssig, and A.I. Goldman, *Nat. Commun.* **7**, 12728 (2016)

- [46] T. Imai, K. Ahilan, F.L. Ning, T.M. McQueen, and R.J. Cava, Phys. Rev. Lett. **102**, 177005 (2009)
- [47] K. Rana and Y. Furukawa, Front. Phys. **10**, 849284 (2022)
- [48] P. Wiecki, M. Nandi, A.E. Böhmer, S.L. Bud'ko, P.C. Canfield, and Y. Furukawa, Phys. Rev. B **96**, 180502 (2017)
- [49] R. Khasanov, Z. Guguchia, A. Amato, E. Morenzoni, X. Dong, F. Zhou, and Z. Zhao, Phys. Rev. B **95**, 180504(R) (2017)
- [50] S. Avci, O. Chmaissem, J. Allred, S. Rosenkranz, I. Eremin, A. Chubukov, D. Bugaris, D. Chung, M. Kanatzidis, J.-P. Castellán, J. Schlueter, H. Claus, D. Khalyavin, P. Manuel, A. Daoud-Aladine, and R. Osborn, Nat. Commun. **5**, 3845 (2014)
- [51] E. Hassinger, G. Gredat, F. Valade, S.R. de Cotret, A. Juneau-Fecteau, J.-P. Reid, H. Kim, M.A. Tanatar, R. Prozorov, B. Shen, H.-H. Wen, N. Doiron-Leyraud, and L. Taillefer, Phys. Rev. B **86**, 140502 (2012)
- [52] J.M. Allred, K.M. Taddei, D.E. Bugaris, M.J. Krogstad, S.H. Lapidus, D.Y. Chung, H. Claus, M. Kanatzidis, D.E. Brown, J. Kang, R.M. Fernandes, I. Eremin, S. Rosenkranz, O. Chmaissem, and R. Osborn, Nat. Phys. **12** (2016)
- [53] W.R. Meier, T. Kong, S.L. Bud'ko, and P.C. Canfield, Phys. Rev. Mater. **1**, 013401 (2017)
- [54] W.R. Meier: *Growth, properties and magnetism of CaKFe₄As₄* (PhD thesis, Iowa State University, 2018)
- [55] A. Kreyssig, J.M. Wilde, A.E. Böhmer, W. Tian, W.R. Meier, B. Li, B.G. Ueland, M. Xu, S.L. Bud'ko, P.C. Canfield, R.J. McQueeney, and A.I. Goldman, Phys. Rev. B **97**, 224521 (2018)

11 From Mott to Cluster Mott: Probing Correlations with RIXS

Markus Grüninger
Institute of Physics II
University of Cologne

Contents

1	Introduction	2
2	RIXS on correlated transition-metal compounds	4
3	Resonance behavior	6
4	Studying local multiplets of Mott insulators with RIXS	7
4.1	Cubic multiplets, Hund’s coupling, and strong spin-orbit coupling	7
4.2	$5d^4$ configuration: Non-magnetic $J=0$ K_2OsCl_6	9
4.3	$5d^3$ K_2ReCl_6 : spin-orbit coupling in a half-filled shell	11
5	Cluster Mott insulators	15
5.1	RIXS interferometry	16
5.2	Face-sharing $5d$ iridates: dimers, trimers, and disorder phenomena	18
5.3	Between Mott and cluster Mott: face-sharing $4d$ Ru dimers	26

1 Introduction

Correlated electron systems offer fundamentally interesting physics in exotic quantum phases of matter, and their characteristic excitations weave the rich tapestry of their fascinating properties. To develop a comprehensive understanding of a given compound or phase, it is crucial to examine it from the distinct perspectives offered by various spectroscopic tools. Prominent examples include inelastic scattering of neutrons or of photons, angle-resolved photoemission (ARPES), and optical spectroscopy. Resonant inelastic x-ray scattering (RIXS) is a most powerful experimental technique with unique advantages for studying charge-neutral low-energy excitations [1]. This covers a wide field of magnetic and orbital excitations, phonons (and electron-phonon coupling), density-wave excitations, plasmons, electron-hole continua and excitons, and charge-transfer excitations.¹ To underscore the versatility of RIXS, we focus on the example of spin/orbital excitations in Mott insulators. RIXS has been used to, e.g., determine the dispersions of magnons, paramagnons, orbitons, spin-orbit excitons, and of a two-triplon bound state [7–13], probe the continua of bi-magnons, spinons, and triplons [14–17], detect spin excitations with ΔS^z up to 5 [18, 19], and demonstrate the bond-directional character of magnetic excitations in a proximate Kitaev spin liquid [20], to name but a few.

RIXS is a relatively new addition to the field. This is mainly due to enormous advances in synchrotron light sources and instrumentation, improving, e.g., energy resolution and count rates for soft, tender, and hard x-rays [21–24], but also in theoretical description [1–5]. We use an example from the hard x-ray range. In 2011, RIXS at the Ir L_3 edge ($2p_{3/2}$ to $5d$, 11.2 keV) has been predicted to be promising for probing low-energy excitations of spin-orbit entangled $j = 1/2$ moments in $5d^5$ compounds [25]. First results in 2011 used a resolution $\delta E \approx 1$ eV [26]. In 2012–14, this has been improved to $\delta E = 130$ meV [8], 36 meV [22], 30 meV [9], and 25 meV [10], which is widely used until today. The current record is 10 meV but comes with reduced intensity [27]. Still, there are only two user facilities for high-resolution hard x-ray RIXS.²

In these lecture notes, we discuss RIXS on Mott-insulating $4d$ and $5d$ transition-metal compounds with strong spin-orbit coupling. We focus on cluster Mott insulators [28], a new class of materials in which electrons are delocalized over a small cluster, e.g., dimers or tetrahedra, while inter-cluster charge fluctuations are suppressed by Coulomb interactions. Such cluster Mott insulators offer the possibility of tailoring and controlling unconventional quasimolecular magnetic moments and exchange interactions, paving the way for novel phases of quantum matter. RIXS is the ideal tool to study the quasimolecular electronic structure and magnetic moments that are governed by the competition of correlation, intra-cluster hopping, spin-orbit coupling, and crystal field. The excitation energy $\hbar\omega_0$ of local intra-cluster excitations does not show dispersion. Still, the $\mathbf{q}$ dependence is a key feature: Coherent resonant scattering on the cluster sites yields an interference pattern in the $\mathbf{q}$ -dependent RIXS intensity that reflects the dynamical structure factor $S(\mathbf{q}, \omega_0)$ of the intra-cluster excitations. This interference pattern provides valuable information on the symmetry and character of the electronic states.

¹For reviews of RIXS, see [1–5]. RIXS has also been addressed within the Jülich Autumn School [6].

²European Synchrotron Radiation Facility (ESRF, Grenoble) and Advanced Photon Source (APS, ANL, USA).

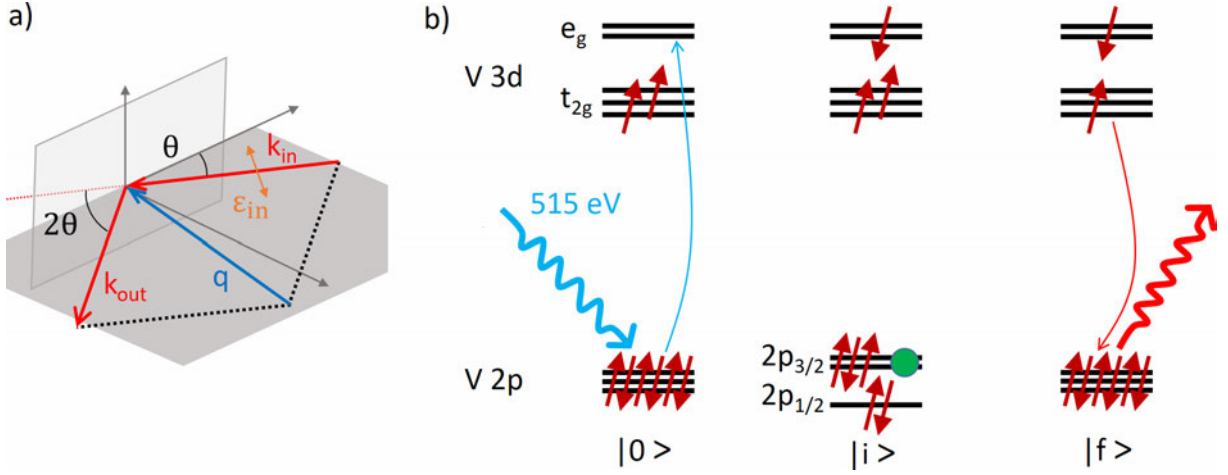


Fig. 1: Resonant inelastic x-ray scattering. *a)* Sketch of the scattering geometry with transferred momentum $\mathbf{q} = \mathbf{k}_{in} - \mathbf{k}_{out}$, where 2θ (not to be confused with $2 \cdot \theta$) is the scattering angle and θ the angle of incidence. The sketch uses linear incident polarization ε_{in} lying in the horizontal scattering plane, while the polarization of the scattered photons is not resolved. *b)* Example for a direct RIXS process at the V L_3 edge at about 515 eV, as observed in YVO_3 [29]. In the ground state $|0\rangle$, a V^{3+} d^2 site hosts two t_{2g} electrons with $S = 1$. Upon absorption of an x-ray photon with suitable energy E_{in} , a $2p$ core electron is promoted to the $3d$ shell (in this example, we have chosen an e_g orbital). At the L_3 edge, the intermediate state $|i\rangle$ shows a $2p_{3/2}$ core hole. Within the $2p$ level, spin-orbit coupling is strong, splitting the $2p^5$ configuration ($S = 1/2$, $L = 1$) into $2p_{1/2}$ and $2p_{3/2}$ states with $J = 1/2$ and $3/2$, respectively. Finally, an electron relaxes from the $3d$ shell to the $2p$ core shell and an x-ray photon with energy E_{out} is emitted. In the example, the final state $|f\rangle$ with excitation energy $E_{loss} = E_{in} - E_{out}$ shows a $t_{2g}^1 e_g^1$ state with $S = 0$, i.e., a different $3d^2$ multiplet. The excitation affects both the orbital occupation and the spin. The change of spin from $S = 1$ to 0 is possible since spin is not a good quantum number in the intermediate state with a $2p_{3/2}$ core hole. For simplicity, we focused on the transferred energy E_{loss} but also the transferred momentum $\mathbf{q}$ and the polarizations of incident and scattered photons are relevant.

The basic principles of RIXS will be introduced in Sec. 2, while Sec. 3 presents an example of resonance behavior. To contextualize the results on cluster Mott compounds, Sec. 4 discusses local cubic multiplets in Mott insulators with strong spin-orbit coupling. We address $5d^3$ K_2ReCl_6 and $5d^4$ K_2OsCl_6 , model systems for the competition of Hund's coupling and spin-orbit coupling. In Sec. 5 on cluster Mott insulators, we introduce the concept of RIXS interferometry and demonstrate its power for the study of cluster Mott insulators but also disorder phenomena. In face-sharing $5d$ iridate dimers and trimers, large hopping yields quasimolecular behavior. In an intuitive picture, the quasimolecular orbitals are built from spin-orbit entangled j states. In $4d$ Ru dimers, however, the dichotomy of conventional single-site Mott behavior and quasimolecular cluster Mott insulators breaks down. Hopping, Hund's coupling, and trigonal crystal-field splitting are of comparable size, resulting in an electronic structure with properties that lie between Mott behavior and cluster Mott behavior. Altogether, these compounds provide a new perspective on the competition of hopping and correlations.

2 RIXS on correlated transition-metal compounds

In (R)IXS, an incident x-ray photon with energy E_{in} and wave vector $\mathbf{k}_{\text{in}}$ is absorbed and a photon with E_{out} and $\mathbf{k}_{\text{out}}$ is emitted, see Fig. 1a). This creates³ an excitation with energy $E_{\text{in}} - E_{\text{out}}$ and $\mathbf{q} = \mathbf{k}_{\text{in}} - \mathbf{k}_{\text{out}}$. For the large x-ray energies such as 11.2 keV at the Ir L_3 edge, the phase space for energy and momentum transfer is huge. With $|q| = (2E_{\text{in}}/\hbar c) \sin(2\theta/2)$, the transferred momentum is as large as, e.g., $8 \text{ \AA}^{-1}$ for $2\theta = 90^\circ$ and $E_{\text{in}} = 11.2 \text{ keV}$. For soft x-rays, $|q|$ is smaller, e.g., $0.4 \text{ \AA}^{-1}$ at the V L_3 edge at 0.51 keV, again for $2\theta = 90^\circ$.

For comparison, we first consider inelastic x-ray scattering (IXS) [30,31]. In *non-resonant* scattering, the x-rays couple to the electronic charge density ρ_c and directly probe the corresponding dynamical structure factor $S_c(\mathbf{q}, \omega)$, measuring the fluctuations of ρ_c . IXS is a versatile tool for the study of electron dynamics and phonons, for instance up to 200 meV in graphite [32].^{4,5}

In contrast, RIXS is a resonant technique. The incident energy E_{in} is tuned to an x-ray absorption edge, i.e., an atomic transition. Examples are transition-metal K edges (e.g., Cu $1s$ to $4p$), transition-metal L edges (e.g., Cu $2p$ to $3d$), and ligand K edges (e.g., O $1s$ to $2p$).⁶ The choice of the absorption edge is decisive for the selection rules, opening entirely different excitation channels and turning RIXS into a most versatile tool. We illustrate this for RIXS at a transition-metal L_3 edge, involving two dipole transitions between $2p$ and d states, see Fig. 1b). The initial absorption process promotes a $2p$ core electron to an empty state in the partially filled d shell, i.e., right into the states relevant to the low-energy properties. Consequently, L -edge RIXS is particularly sensitive to spin and orbital excitations. We choose the case of V $3d^2$ in YVO_3 [29]. The intermediate state shows a $2p^5 3d^3$ configuration with a $2p$ core hole. The strong potential of the core hole attracts the excited electron, which in the intermediate state does not delocalize in the lattice but is bound to the core-hole site. The intermediate state has a very short life time. It decays via the relaxation of an electron from the $3d$ valence shell to the $2p$ core level, emitting a photon with E_{out} . This leaves the d shell in an excited state $3d^{2*}$. Our example in Fig. 1b) uses an $S = 1$ t_{2g}^2 ground state and an $S = 0$ $t_{2g}^1 e_g^1$ final state. Compared to the ground state, the RIXS process created an e_g electron and a t_{2g} hole on the same site, i.e., an on-site exciton, a charge-neutral excitation. In this example, the exciton corresponds to a multiplet excitation in which both the orbital occupation and the total spin have been changed. The change in total angular momentum is balanced by the polarization of the scattered photon.

Thus far, we only referred to local properties, describing a *local* excitation. In fact, the dispersion of orbital excitations is often comparable to the peak width, so that they can be approximated as being local, see Sec. 4.2. However, this is related to the, in many cases, small effective hopping of orbital excitations but is not caused by the RIXS process. Based on the translational invariance of the lattice, the exciton lives in $\mathbf{q}$ space and may delocalize. Propagating orbital excitations with a sizable dispersion are called orbitons and can be observed in RIXS [12]. We

³Also annihilation processes with $E_{\text{out}} > E_{\text{in}}$ are possible, equivalent to anti-Stokes Raman scattering.

⁴See [33] for lecture notes of L.H. Tjeng on imaging of orbitals via non-resonant IXS.

⁵In non-resonant IXS, sub-meV resolution has been achieved [34], but RIXS is more challenging, since high resolution has to be realized at the particular energy dictated by the absorption edge of the compound under study.

⁶The K , L , and M edges correspond to core shells with main quantum number $n = 1, 2$, and 3 , respectively.

consider the more common case of magnons, propagating spin waves. In a system with long-range magnetic order, a local spin flip is not an eigenstate, but magnons are. A magnon can be viewed as a superposition of local spin flips, and the excitation of a magnon with a given $\mathbf{q}$ corresponds to a coherent sum over spin-flip RIXS processes on all sites involved in the magnon. Since x-rays carry large $\mathbf{q}$, the dispersions of magnons and orbitons can be mapped out.

Spin-flip excitations are allowed in our example of RIXS at a transition-metal L_3 edge [7,25,35]. A spin flip within the d shell is possible due to the strong spin-orbit coupling within the $2p$ shell that splits the core hole states into $2p_{1/2}$ and $2p_{3/2}$, see Fig. 1. The splitting amounts to several eV in $3d$ transition metals such as V or Cu and is as large as 1.6 keV in Ir. Based on this strong spin-orbit coupling, spin is not a good quantum number in the intermediate state. In comparison, spin-orbit coupling vanishes for the $1s$ shell with $L = 0$. Therefore, K -edge RIXS with a $1s$ core hole does not allow for spin-flip excitations [29]. This demonstrates the decisive role of the absorption edge for the selection rules.

All of the RIXS results discussed in the following have been obtained at a transition-metal L edge, using so-called *direct* RIXS as shown in Fig. 1b). Indirect RIXS becomes important when direct RIXS is forbidden by selection rules. As an example for indirect RIXS, we consider the Cu K edge. We start with a dipole transition from the $1s$ core orbital to the *empty* $4p$ valence shell far above the Fermi level, and later the same electron relaxes. The excitation within the partially filled $3d$ shell arises in a shake-up process via the interaction with the core hole in the intermediate state, where the short-lived $4p$ electron can be viewed as a spectator of the physics of the $3d$ shell. For instance, the core hole may be screened by the excitation of a phonon. Thus, phonons can be studied in indirect RIXS, but they are suppressed in direct RIXS [36].

The resonant character of RIXS comes with a series of advantages such as resonant enhancement of the scattering cross section, element and orbital specificity, and control over selection rules. When a solid absorbs an x-ray photon, there are many possible decay channels, and resonant enhancement at an x-ray absorption edge provides a decent signal for low-energy inelastic scattering. At the same time, the choice of a particular absorption edge yields element specificity. Measuring at an absorption edge of a given transition metal, RIXS probes in particular the electronic, orbital, and spin excitations in which ions of this transition-metal element are involved. The dipole selection rules add orbital selectivity that can be controlled via the polarization. Furthermore, the choice of the absorption edge also selects the shell of valence orbitals involved in the intermediate state as well as the selection rules, as discussed above. RIXS at different edges provides different information on the excitations. One example is the comprehensive picture of the crystal-field (or local $d-d$) and magnetic excitations of the paradigmatic antiferromagnet NiO that has been obtained by studying RIXS at the Ni M , L , and K edges as well as at the O K edge [37–44].

Note that RIXS is a bulk sensitive technique. The x-ray penetration depth depends strongly on the energy and the compound, but it is of the order of 0.1 μm to a few micrometer. Still it is possible to investigate, e.g., thin films, which may require employing a small angle of incidence. Modern synchrotrons are highly brilliant light sources that offer spot sizes down to a few micrometers, permitting the study of small single crystals (order of 100 μm).

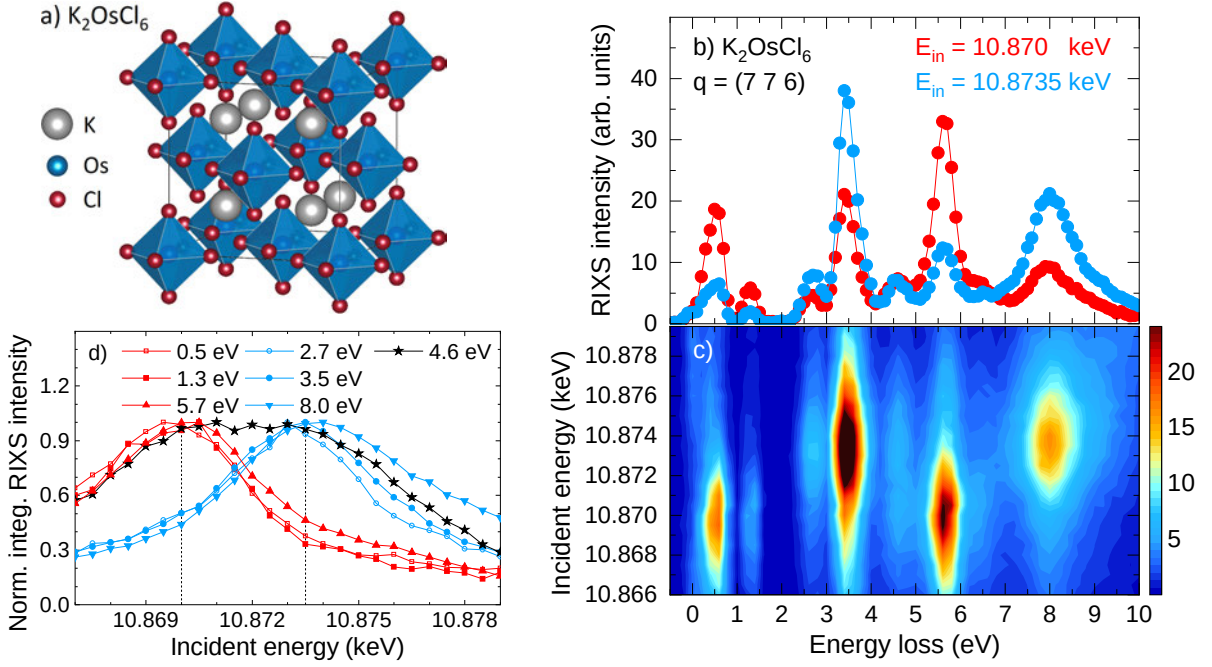


Fig. 2: Resonance behavior of the RIXS intensity at a transition-metal L edge. We use the $5d^4$ Mott insulator K_2OsCl_6 as an example [45]. Resonant enhancement is revealed by RIXS spectra for a series of incident energies E_{in} across the Os L_3 edge. The energy loss of RIXS peaks is independent of E_{in} . The peak intensity shows resonant enhancement for initial x-ray absorption to either t_{2g} or e_g orbitals at 10.870 keV or 10.8735 keV, respectively. This allows us to distinguish intra- t_{2g} excitations, excitations to e_g orbitals, and charge-transfer excitations. *a)* Cubic crystal structure of K_2OsCl_6 (space group $Fm\bar{3}m$ at 300 K). *b)* RIXS spectra measured at t_{2g} and e_g resonance at 20 K. The energy loss axis is the same as for the resonance map in *c)*. *d)* Normalized cuts through the resonance map at constant loss. The behavior at 4.6 eV indicates two distinct excitations. As shown in *a)*, the $OsCl_6$ octahedra are well separated and form an fcc lattice. The structure can be derived from the widely studied double perovskites $A_2BB'O_6$ by placing ordered vacancies on the B' sites.

3 Resonance behavior

The resonance behavior is a key feature, as discussed above, and we want to highlight a further aspect: The resonance behavior supports a first peak assignment in the RIXS spectra, a central task in spectroscopy. The Os L_3 edge of the Mott-insulating $5d^4$ antiferrofluorite halide K_2OsCl_6 provides a transparent example [45]. The crystal structure hosts well-separated $OsCl_6$ octahedra, see Fig. 2a), giving rise to well-localized excitations with narrow line widths and thus a particularly clear picture. Figure 2b) shows RIXS spectra for two incident energies E_{in} . These spectra are equivalent to cuts through the resonance map shown in Fig. 2c), a color plot being built from RIXS spectra for a series of incident energies across the Os L_3 edge. For all observed peaks, the energy loss $E_{loss} = E_{in} - E_{out}$ is independent of E_{in} , demonstrating the excitonic RIXS character of these features with a well-defined excitation energy. This is analogous to Raman scattering with visible light, where the matrix elements depend on the laser frequency but the energy of the Raman peaks does not. This must be distinguished from a fluorescence process in

which the energy loss increases with E_{in} . Consider, e.g., the excitation of an unbound electron-hole pair in a scenario of non-interacting electrons. The core electron may be promoted into a delocalized band state, while the core hole is filled by the decay of an electron from the bottom of the band. In this case, an increase of E_{in} puts the excited electron in a higher-lying band state, while the emission process from the bottom of the band always yields the same energy for the outgoing photons, i.e., the energy loss *increases* with E_{in} . In contrast, in a RIXS process, the energy E_{out} of the emitted photon increases along with E_{in} , reflecting the *constant energy loss* $E_{\text{loss}} = E_{\text{in}} - E_{\text{out}}$ independent of the incident photon energy for a well-defined excitonic excitation such as an on-site orbital d - d excitation or a spin flip.

The data in Fig. 2 reveal a dramatic enhancement of the RIXS signal for incident energies around 10.870 to 10.874 keV. For E_{in} as large as roughly 10^4 eV, it is stunning that the enhancement is restricted to a window of E_{in} of only a few eV, see Fig. 2d). The width of the resonance profile reflects the life time of the intermediate state. For $5d$ L edges in the hard x-ray regime, this life time is of the order of a few femtoseconds only. Note that $3d$ L edges in the soft x-ray range below 1 keV show longer life times and thus even narrower resonance profiles.

Typical for $5d$ transition-metal L_3 edges, K_2OsCl_6 shows two distinct resonance energies that correspond to t_{2g} and e_g resonance. In the absorption step, the $2p$ core electron can be promoted to a t_{2g} or an e_g orbital of the valence d shell, assuming that both subshells contain empty states. In K_2OsCl_6 , the two resonance energies equal 10.870 keV and 10.8735 keV. The difference of 3.5 eV provides a rough estimate of the cubic crystal-field splitting $10 Dq$. We emphasize that the resonance behavior allows for a first peak assignment, i.e., it allows us to distinguish intra- t_{2g} excitations from d - d excitations to e_g orbitals and charge-transfer excitations.

The RIXS peaks below 2 eV show t_{2g} resonance, identifying them as intra- t_{2g} excitations $|t_{2g}^4\rangle \rightarrow |t_{2g}^{4*}\rangle$, see Sec. 4. In cubic symmetry, they are governed by Hund's coupling J_H and spin-orbit coupling ζ , see Sec. 4. Crystal-field or d - d excitations to e_g orbitals are described by $|t_{2g}^4\rangle \rightarrow |t_{2g}^3 e_g^1\rangle$ in K_2OsCl_6 . These additionally involve the cubic crystal-field splitting $10 Dq$, which above has been estimated as 3.5 eV. Accordingly, excitations to e_g orbitals resonate at a higher E_{in} and show larger energy loss, here 2.7 to 4.6 eV, see Fig. 2. Remarkably, t_{2g} resonance is also observed at higher energy loss, e.g., at 4.6 and 5.7 eV. This marks the onset of charge-transfer excitations from Cl $3p$ to Os $5d$ states, $|5d_{\text{Os}}^4 3p_{\text{Cl}}^6\rangle \rightarrow |5d_{\text{Os}}^5 3p_{\text{Cl}}^5\rangle$. The resonance behavior allows us to distinguish $|t_{2g}^5 3p_{\text{Cl}}^5\rangle$ final states from $|t_{2g}^4 e_g^1 3p_{\text{Cl}}^5\rangle$ excitations showing e_g resonance, e.g., at 8.0 eV loss. The narrow line widths of the RIXS peaks of K_2OsCl_6 reveal the distinct t_{2g} and e_g resonance behavior of the charge-transfer excitations particularly well.

4 Studying local multiplets of Mott insulators with RIXS

4.1 Cubic multiplets, Hund's coupling, and strong spin-orbit coupling

In Mott insulators, electronic correlations dominate over the kinetic energy gained upon band formation and delocalization. This yields local spin and orbital degrees of freedom, a vast playground for quantum magnetism. In transition-metal compounds with a partially filled d

t_{2g}^n	# states	multiplets	multiplet energies	excitations
1	6	2T_2	0	–
2	15 = 9+3+2+1	${}^3T_1, {}^1T_2/{}^1E, {}^1A_1$	$U - xJ_H$ ($x = -3, -1, 2$)	$2J_H, 5J_H$
3	20 = 4+6+4+6	${}^4A_2, {}^2T_1/{}^2E, {}^2T_2$	$3U - xJ_H$ ($x = -9, -6, -4$)	$3J_H, 5J_H$
4	15 = 9+3+2+1	${}^3T_1, {}^1T_2/{}^1E, {}^1A_1$	$6U - xJ_H$ ($x = -13, -11, -8$)	$2J_H, 5J_H$
5	6	2T_2	$10U - 20J_H$	–

Table 1: Overview of local t_{2g}^n multiplets for $n = 1$ to 5 in cubic symmetry [46–48]. The second and third columns give the number of states and the multiplet terms. The dash in, e.g., ${}^1T_2/{}^1E$ denotes degenerate states. Multiplet energies are given in the same order as the terms. The prefactor of U counts the number of electron pairs, while the prefactor of J_H depends on the orbital occupation. The last column gives the corresponding energies of on-site excitations.

shell, the character of the local magnetic moment depends on the electron count but also on Coulomb repulsion, spin-orbit coupling, and crystal-field effects. RIXS directly measures the local multiplet structure, offering a quantitative determination of the electronic parameters and thus the character of the local moments.

To warm up, we consider a single $5d$ transition-metal site and focus on Coulomb interactions in cubic symmetry, while we neglect spin-orbit coupling, non-cubic crystal field, and hybridization with the ligands. The large spatial extent of the $5d$ orbitals produces a large cubic crystal-field splitting, $10 Dq > 3 \text{ eV}$, as shown above. This motivates the useful and often employed starting point that assumes $10 Dq \approx \infty$ (Kanamori model). It allows us to neglect e_g orbitals, restricting the discussion to intra- t_{2g} excitations. For the t_{2g} shell, one can quantify Coulomb interactions by two parameters, Hubbard U and Hund’s coupling J_H [46–48]. In terms of Coulomb repulsion, it is particularly costly to put two electrons into the same orbital, while it is beneficial if two electrons occupy different orbitals with parallel spins. The corresponding on-site energies are given in Table 1. The configurations t_{2g}^n with $n = 2, 3$, and 4 allow for on-site intra- t_{2g} excitations between Hund’s multiplets. The excitation energies depend on J_H but not on U since an on-site excitation changes the orbital occupation but not the local electron count. We will focus on the competition of Hund’s coupling and strong-spin-orbit coupling in cubic $5d^3$ and $5d^4$ compounds.

Before discussing the results, we add a note on the value of J_H . For the entire d shell, Coulomb interactions are often parameterized by the Slater-Condon parameters F^0 , F^2 , and F^4 that correspond to monopole, dipole, and quadrupole contributions.⁷ In the t_{2g} -only Kanamori model discussed above, U and J_H are defined as (a comprehensive discussion can be found in [47])

$$U = F^0 + \frac{4}{49}F^2 + \frac{36}{441}F^4, \quad J_H = \frac{3}{49}F^2 + \frac{20}{441}F^4. \quad (1)$$

However, if one considers the entire d shell including the e_g orbitals in the model and the analysis, i.e., for finite $10 Dq$, one finds different expressions,

$$U_d = F^0, \quad J_{H,d} = \frac{1}{14}(F^2 + F^4). \quad (2)$$

⁷Equivalently, one may employ the Racah parameters $A = F^0 - \frac{49}{441}F^4$, $B = \frac{1}{49}F^2 - \frac{5}{441}F^4$, $C = \frac{35}{441}F^4$.

Typically, a fixed ratio F^4/F^2 is considered,⁸ often $F^4/F^2 \approx 5/8$. This yields

$$J_H \approx 0.77 J_{H,d}. \quad (3)$$

The numerical value of Hund's coupling depends on the model used. A comparison of both models has been reported for $5d^4$ K_2OsCl_6 and $5d^3$ K_2ReCl_6 [45, 49]. An analysis of the RIXS data of K_2OsCl_6 that considers the entire $5d$ shell yields $F^2 = 3.7$ eV, $F^4 = 2.2$ eV, and $J_{H,d} = 0.43$ eV. Equation (3) predicts $J_H \approx 0.33$ eV, in reasonable agreement with an analysis restricted to the t_{2g} states, giving $J_H \approx 0.28$ eV [45] (see below). The t_{2g} -only model may yield reasonable results, but it is important to indicate the model to which one refers.

In recent years, the rich physics originating from strong spin-orbit coupling combined with electronic correlations has attracted enormous interest [28, 50–54]. In $5d$ transition-metal compounds, spin-orbit coupling is comparable to Hund's coupling. Therefore, spin and orbital degrees of freedom are strongly entangled, and the resulting effective total angular momentum states differ qualitatively from conventional spin states. This is the basis for multipolar order in $5d^1$ and $5d^2$ compounds [55], excitonic magnetism in $5d^4$ and $4d^4$ materials [56], and experimental realizations of Kitaev-type bond-directional exchange interactions in, e.g., $5d^5$ $j=1/2$ compounds [20, 57, 58]. In passing, we note that $5d^5$ $j=1/2$ K_2IrCl_6 has been discussed as a nodal-line spin liquid [59] and as an example for the entanglement not only between spins and orbitals but also with lattice degrees of freedom [36, 60]. Below, we address $5d^4$ K_2OsCl_6 , providing an exceptionally clean realization of the non-magnetic $J=0$ state induced by strong spin-orbit coupling, and the effect of spin-orbit coupling on a half-filled shell in $5d^3$ K_2ReCl_6 [45, 49].

4.2 $5d^4$ configuration: Non-magnetic $J=0$ K_2OsCl_6

Excellent model systems for spin-orbit-entangled quantum states can be found in the Mott-insulating antiferrofluorite transition-metal halides A_2MX_6 with K_2PtCl_6 -type structure. These compounds exist for a wide variety of tetravalent metal ions ($M = \text{Pt}, \text{Re}, \text{Os}, \text{and Ir}$, etc.) for $A = \text{K}, \text{Rb}, \text{Cs}$, or $(\text{NH})_4$, and $X = \text{F}, \text{Cl}, \text{Br}$, and I . The isolated MX_6 octahedra (see Fig. 2a)) suppress direct metal-metal hopping and thereby favor localized behavior. This structural simplicity facilitates a direct comparison between experimental spectra and atomic multiplet calculations. Furthermore, at room temperature these compounds show a cubic crystal structure [61–63], allowing us to focus on the interplay of Hund's coupling J_H and spin-orbit coupling ζ for the case where the two coupling constants are comparable in size. This intermediate regime is located between the LS coupling limit realized for $\zeta/J_H \rightarrow 0$ and the jj coupling regime⁹ that corresponds to $\zeta/J_H \rightarrow \infty$.

The t_{2g}^4 configuration has attracted considerable interest in the context of *excitonic magnetism* [56, 64]. Starting from a local non-magnetic $J=0$ ground state, strong exchange interactions between neighboring sites may cause a sizable dispersion of the $J=1$ excited state such that this

⁸Another common choice is Racah $C/B=4$, equivalent to $F^4/F^2=36/55$.

⁹In LS coupling, the total angular momentum J of an ion is calculated by coupling the total spin S and the total angular momentum L , where both S and L are obtained by summing over the electrons. In contrast, in jj coupling, l and s of each electron are first coupled to j before J is calculated by summing over the electrons.

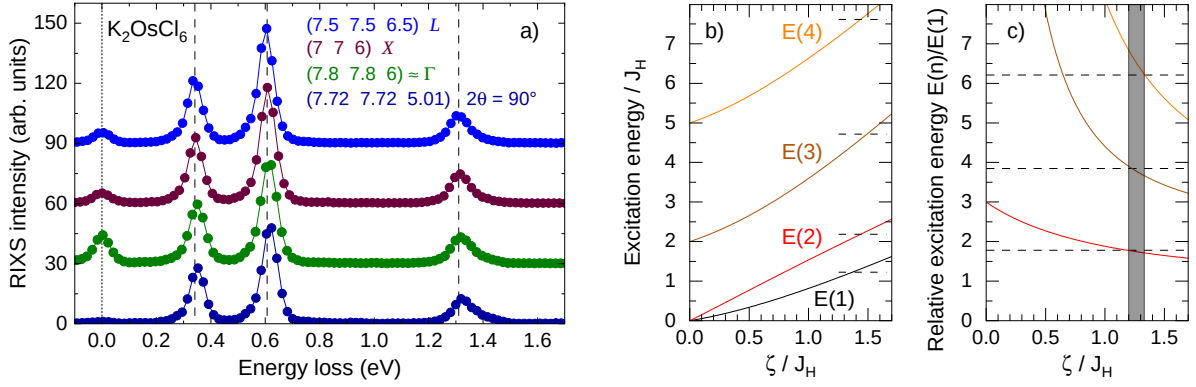


Fig. 3: Intra- t_{2g} excitations of K_2OsCl_6 . *a)* RIXS spectra at selected q points show local excitations at 0.34, 0.61, and 1.31 eV with negligible q dependence [45]. Dashed: peak energies at the X point. A fourth peak has been detected at 2.1 eV [45]. The elastic line is weak since a scattering angle 2θ close to 90° has been used along with incident π polarization. Full suppression of the elastic line is achieved for $2\theta = 90^\circ$ (dark blue), in which case the polarization of the scattered x-rays is perpendicular to the incident π polarization. The data, plotted with a vertical offset, have been collected at 20 K in the tetragonal phase. The energy resolution of 63 meV is not sufficient to resolve the non-cubic splitting of 4 meV observed in infrared data [45]. *b)* Effect of spin-orbit coupling on the excitation energies $E(1)$ to $E(4)$ in the t_{2g} -only Kanamori model. Dashed: experimental results assuming $J_H = 0.28$ eV. *c)* Ratios of excitation energies $E(n)/E(1)$ as a function of ζ/J_H . Dashed: experimental values. Grey: $\zeta/J_H = 1.20$ – 1.33 .

magnetic state may condense, yielding a magnetic ground state. The key signature of excitonic magnetism is a magnetic amplitude mode equivalent to a Higgs mode that has been discussed for antiferromagnetic Ca_2RuO_4 [56,65]. However, layered Ca_2RuO_4 shows a sizable non-cubic crystal-field splitting Δ_{CF} larger than ζ . Therefore, the magnetism of t_{2g}^4 Ca_2RuO_4 can also be described using a low-energy $S = 1$ model with anisotropy, and the pros and cons of the two approaches, starting from $J = 0$ or $S = 1$, have been discussed extensively [48,65–70].

For K_2OsCl_6 , Fig. 3a) shows the low-energy RIXS spectra for selected q points [45]. An energy resolution $\delta E = 63$ meV has been achieved at the Os L_3 edge, i.e., $\delta E/E_{in} \approx 6 \cdot 10^{-6}$. The spectra show intra- t_{2g} excitations at about 0.34, 0.61, and 1.31 eV. Note that the error bar of the energy of a well-separated peak is smaller than δE . Within the experimental resolution, the peak energies do not depend on the transferred momentum q . The absence of dispersion underpins the local character of these excitations. The excitation energies agree very well with the results obtained from infrared spectroscopy for $q \approx 0$ [45]. A further RIXS peak has been observed at 2.1 eV [45]. This intra- t_{2g} mode is another example of the important role of selection rules [71]. Its RIXS intensity vanishes for a scattering angle of 90° , which is often chosen experimentally to suppress the contribution of the elastic line at zero energy loss, see Fig. 3a).

Above we have discussed the limit of $10 Dq = \infty$ with spin-orbit coupling $\zeta = 0$. The neglect of the e_g orbitals can be motivated by the large value of $10 Dq \approx 3.5$ eV and by the resonance behavior, see Sec. 3. For the t_{2g}^4 configuration, Coulomb interactions yield a nine-fold degenerate 3T_1 ground state and excited states at $2J_H$ and $5J_H$, see Table 1. However, the experimental data show more than two peaks as spin-orbit coupling splits the 3T_1 multiplet into the $J = 0$

ground state, the $J = 1$ triplet, and the $J = 2$ quintuplet. The five t_{2g}^4 energies are given by [72]

$$E_{J=0/{}^1A_1} = \frac{5}{2}J_H \left(1 \mp \sqrt{1 + \frac{2}{5}\frac{\zeta}{J_H} + \left(\frac{3}{5}\frac{\zeta}{J_H}\right)^2} \right) - \frac{\zeta}{2} \quad (4)$$

$$E_{J=1} = -\frac{\zeta}{2} \quad (5)$$

$$E_{J=2/({}^1T_2/{}^1E)} = \frac{2}{2}J_H \left(1 \mp \sqrt{1 - \frac{1}{2}\frac{\zeta}{J_H} + \left(\frac{3}{4}\frac{\zeta}{J_H}\right)^2} \right) + \frac{\zeta}{4}. \quad (6)$$

The corresponding excitation energies $E(1)$ to $E(4)$, relative to $E_{J=0}$, are depicted in Fig. 3b) as a function of ζ/J_H . Fitting the four intra- t_{2g} excitation energies observed in RIXS yields $J_H \approx 0.28$ eV and $\zeta/J_H \approx 1.45$. The dashed lines in Fig. 3b) indicate the experimental peak energies for $J_H = 0.28$ eV. These results firmly establish that K_2OsCl_6 is located in the intermediate regime between LS coupling and jj coupling.

This claim is supported by the ratios $E(n)/E(1)$ for $n = 2$ to 4, which yield $\zeta/J_H = 1.20$ – 1.33 , see Fig. 3c). A discussion of relative excitation energies may be favorable for comparison with other d^4 compounds. We focus on the ratio $E(2)/E(1)$. For $\zeta/J_H \rightarrow 0$ (LS coupling), one finds $E(1) \approx 0.5\zeta$ and $E(2) \approx 1.5\zeta$, see Eqs. (4) to (6). For $\zeta/J_H \rightarrow \infty$ (jj coupling), $E(1) \approx E(2) \approx 1.5\zeta$. The ratio $E(2)/E(1)$ changes from 3 to 1 between the two limits, see Fig. 3c). In K_2OsCl_6 , $E(2)/E(1) \approx 1.8$ marks the intermediate regime. RIXS results on $5d^4$ iridates [73–76] show very similar ratios $E(2)/E(1)$, pointing to similar values of ζ/J_H .

The t_{2g} -only Kanamori model offers a reasonable description of the intra- t_{2g} excitation energies. With the assumption $10 Dq = \infty$, it is not surprising that the agreement with the experimental data is not perfect. Taking into account the e_g orbitals with finite $10 Dq$ has clear consequences: it lifts the degeneracy of the $J = 2$ quintuplet as well as the degeneracy of the 1T_2 and 1E multiplets [45]. For $J = 2$, the effect is too small to be resolved in our RIXS data. Regarding 1T_2 and 1E , the RIXS peak at 1.31 eV exhibits an asymmetric line shape with a high-energy shoulder, see Fig. 3a). Infrared data of K_2OsCl_6 confirm a peak splitting [45]. A combined analysis of RIXS and infrared data considering the entire d shell yields $\zeta_d = 0.34$ eV and $J_{H,d} = 0.43$ eV, equivalent to $\zeta_d/J_{H,d} \approx 0.8$ [45]. The different definitions of J_H and $J_{H,d}$ have been discussed above, see Eqs. (1) and (2). We emphasize that the choice of the model also affects the value of the spin-orbit coupling constant, $\zeta \neq \zeta_d$. Beyond the degeneracies discussed above, the Kanamori model and a model considering the entire d shell achieve a similar level of agreement with the intra- t_{2g} excitation energies observed in $5d^4$ K_2OsCl_6 [45]. However, the analytic expressions of the energies in the Kanamori model, see Eqs. (4)–(6), simplify the analysis considerably.

4.3 $5d^3$ K_2ReCl_6 : spin-orbit coupling in a half-filled shell

The role of spin-orbit coupling in the case of a half-filled t_{2g} shell is an interesting conceptual question. In the LS coupling limit, the three electrons maximize the total spin according to Hund's rule, giving rise to a spin-only $S = 3/2$ scenario with quenched orbital momentum. Based

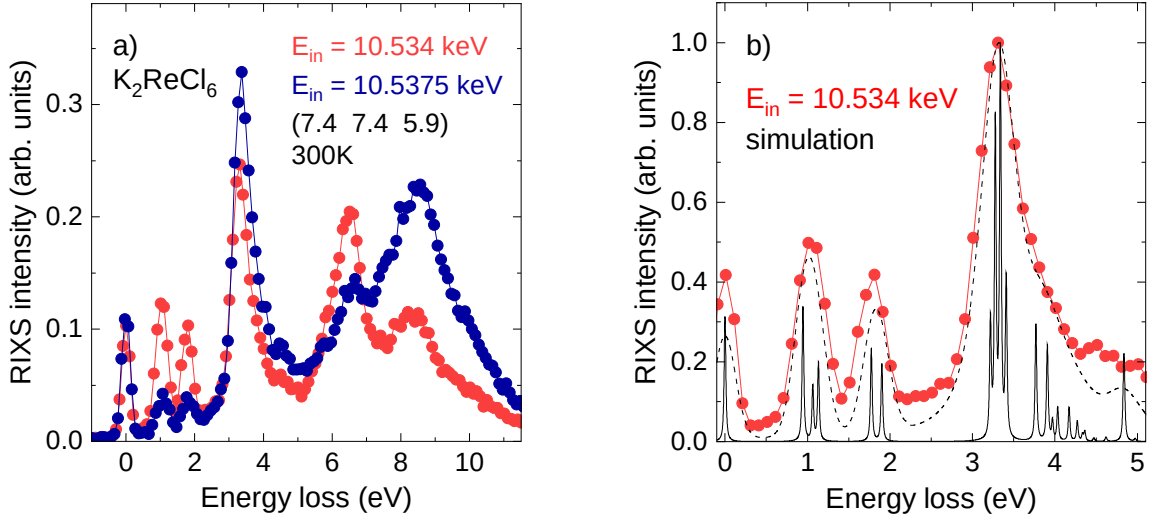


Fig. 4: RIXS spectra of cubic $5d^3$ K_2ReCl_6 [49]. *a)* The data have been measured at 300 K with 295 meV resolution for incident energy $E_{\text{in}} = 10.534$ keV (t_{2g} resonance) and 10.5375 keV (e_g resonance). Intra- t_{2g} peaks are observed at 1.0 and 1.8 eV, while excitations to e_g states are peaking at 3.3 eV. *b)* Comparison of RIXS data (red) with a simulated spectrum (black) calculated for $10 Dq = 3.25$ eV, $\zeta_d = 290$ meV, and $J_{H,d} = 464$ meV. Solid: Voigt oscillators with 10 meV line width. Dashed: same data with larger line width.

on the equal occupation of the three t_{2g} orbitals, a distortion does not lower the energy. This spin-only scenario describes, e.g., the gapless magnetic excitations reported for $4d^3$ $\text{Ca}_3\text{LiRuO}_6$ [77]. Streltsov and Khomskii [53] pointed out that this well-known scenario breaks down for large ζ/J_H in the jj coupling limit. For $\zeta/J_H \rightarrow \infty$, the electrons occupy states with $j = 3/2$ and $1/2$ character. The half-filled t_{2g} shell corresponds to a $3/4$ -filled $j = 3/2$ quartet, allowing for a Jahn-Teller splitting of the t_{2g}^3 configuration [53], in stark contrast to the spin-only $S = 3/2$ scenario. It is intriguing to consider how the many-body states evolve as a function of ζ . In the intermediate-coupling regime, the entanglement of spin and orbital angular momenta gives rise to anisotropy and the opening of a spin gap. In fact, sizable spin gaps have been reported for, e.g., $5d^3$ $\text{Ca}_3\text{LiOsO}_6$ and Ba_2YOsO_6 [78–81], which have been discussed in terms of a spin-orbit-entangled $J = 3/2$ state [82]. Antifluorite K_2ReCl_6 is an ideal model system for examining the competition of strong spin-orbit coupling and Hund’s coupling in a cubic crystal field with only small interactions between adjacent sites.

Thus far, RIXS at the Re L_3 edge has hardly been studied. The resonant character of RIXS is a major experimental challenge because the setup has to be optimized for every absorption edge to be studied. For hard x-rays, this affects in particular the energy resolution δE as one has to choose an analyzer crystal with an appropriate Bragg peak for energy discrimination. Figure 4 shows RIXS spectra of cubic $5d^3$ K_2ReCl_6 with $\delta E = 295$ meV for incident energies that correspond to t_{2g} resonance and e_g resonance, respectively [49].

Equivalent to the case of $5d^4$ K_2OsCl_6 , cf. Fig. 2, the resonance behavior of $5d^3$ K_2ReCl_6 reveals the intra- t_{2g} character of the RIXS peaks at 1.0 and 1.8 eV, strong excitations from $|t_{2g}^3\rangle$ to $|t_{2g}^2 e_g^1\rangle$ around 3.5 eV, and charge-transfer excitations from the ligand Cl p shell to either Re

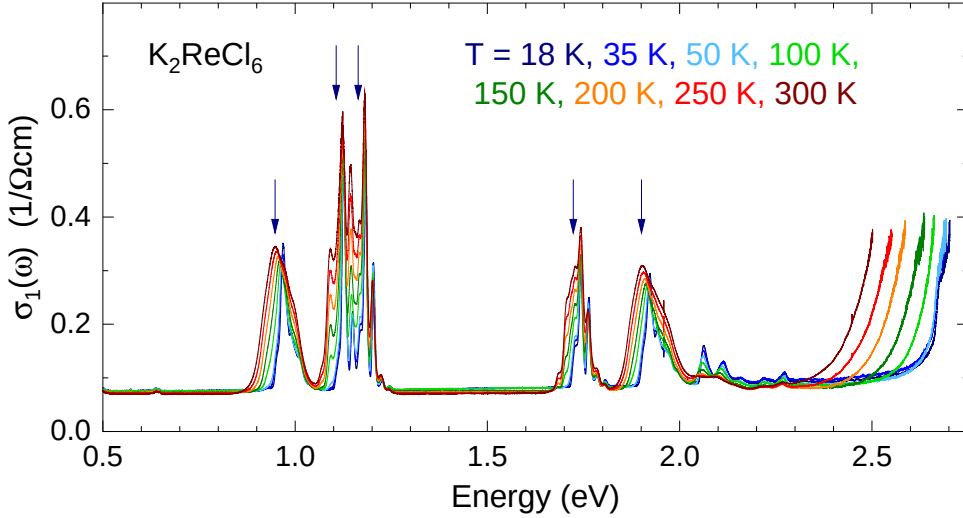


Fig. 5: Intra- t_{2g} excitations in the optical conductivity $\sigma_1(\omega)$ of $5d^3$ K_2ReCl_6 . RIXS shows two peaks at 1.0 and 1.8 eV, cf. Fig. 4. The infrared data resolve all five on-site $d-d$ excitations between 0.9 and 2 eV (arrows). In $\sigma_1(\omega)$, the $d-d$ excitations are both parity forbidden and spin forbidden but acquire a small spectral weight in a phonon-assisted process with a characteristic temperature dependence. The rich line shape is caused by vibronic sidebands, i.e., electron-phonon coupling. These weak absorption features can only be detected below the Mott gap, the onset of which is observed at 2.7 eV at low temperature. The even weaker peaks between 2.0 and 2.3 eV correspond to pair excitations of intra- t_{2g} excitations on neighboring sites [49].

t_{2g} or e_g orbitals above 5 eV. Based on Coulomb interactions, assuming $10 Dq = \infty$ and $\zeta = 0$, one expects a $S = 3/2$ 4A_2 ground state and intra- t_{2g} excitations at $3J_H$ and $5J_H$, see Table 1. Assuming $\zeta = 0$, the excitation energies 1.0 and 1.8 eV suggest $J_H \approx 0.35$ eV. Both spin-orbit coupling and finite $10 Dq$ yield a peak splitting that we cannot resolve in RIXS with 295 meV resolution, see Fig. 4b). Four RIXS peaks have been reported in $5d^3$ $\text{Ca}_3\text{LiOsO}_6$ and Ba_2YO_6 with $\delta E = 150$ meV [82]. Infrared data of K_2ReCl_6 resolve all five intra- t_{2g} excitations [49].

For the study of local $d-d$ excitations, RIXS and infrared spectroscopy are complementary in terms of selection rules and energy resolution [49]. On-site $d-d$ excitations are allowed in L -edge RIXS [1], see Fig. 1. In contrast, infrared absorption is predominantly sensitive to electric dipole excitations such as inter-site excitations across the Mott gap. In the presence of inversion symmetry, $d-d$ excitations are parity forbidden in optical spectroscopy. However, a finite matrix element arises in a phonon-assisted process, where inversion symmetry is broken by the simultaneous creation or annihilation of an odd-symmetry phonon. This gives rise to weak phonon-assisted $d-d$ excitations [36, 45, 49, 83, 84]. Note that the peak energies in RIXS and optical data are shifted by the phonon energy, and that the phonon-assisted nature of the process leaves its fingerprints in a characteristic temperature dependence of the spectral weight and line shape of the optical data. In K_2ReCl_6 , the spectral weight in infrared data is further suppressed by the spin-selection rule, which can be relaxed by spin-orbit coupling.

Weak phonon-assisted $d-d$ excitations can be detected in infrared transmittance measurements on single crystals. This requires that the weak excitations are located in a transparent frequency

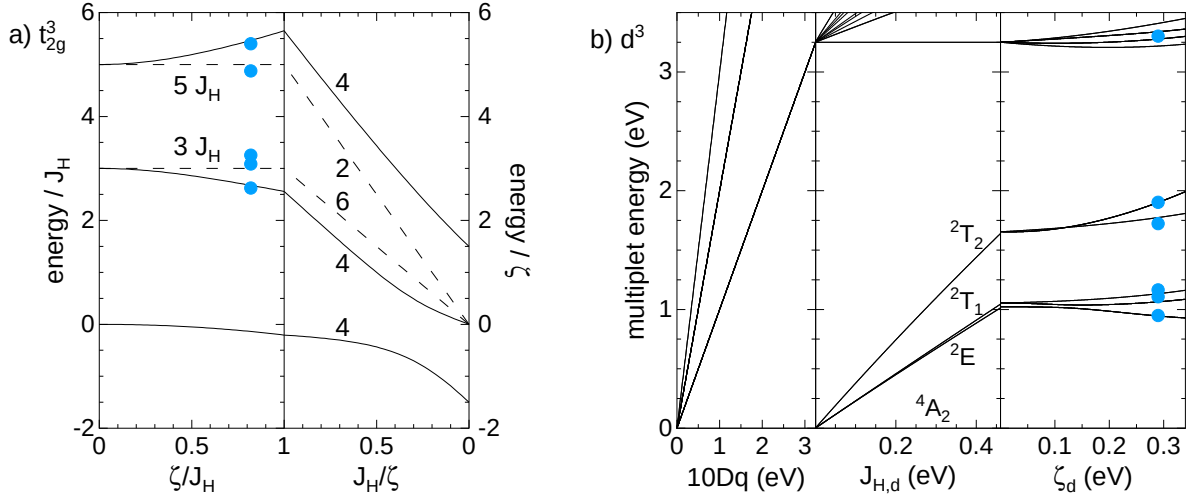


Fig. 6: Energies for the d^3 configuration in cubic symmetry [49]. a) Absolute energies of the t_{2g}^3 states in the Kanamori model for $10Dq = \infty$, see Eq. (7). Spin-orbit coupling mixes the states plotted with solid lines, giving rise to in total four excitation energies relative to the ground state. The energies are shown from $\zeta/J_H = 0 \rightarrow \infty$. Note the change of scale of both axes. Blue symbols show the experimental result for $\zeta/J_H = 0.82$ with $J_H = 343$ meV. Numbers in the right panel denote the degeneracy. b) Energy level diagram including e_g states for finite $10Dq$. The panels show the excitation energies upon consecutively switching on the cubic crystal-field splitting $10Dq$, Hund's coupling $J_{H,d}$, and spin-orbit coupling ζ_d . The energies have been calculated using QUANTY [88, 89]. Blue symbols depict the experimental results for the five intra- t_{2g} excitation energies and the t_{2g} -to- e_g RIXS peak at 3.3 eV.

range with no strong absorption features, e.g., between the phonons and the Mott gap. In this range, infrared data of $K_2\text{ReCl}_6$ reveal five phonon-assisted absorption bands with a multitude of vibronic phonon sidebands,¹⁰ see Fig. 5. These phonon sidebands must be distinguished from the symmetry-breaking phonon mode of the phonon-assisted process. For $K_2\text{ReCl}_6$, a thorough analysis yields the five electronic excitation energies 948, 1107, 1164, 1723, and 1901 meV [49], resolving the splitting of the RIXS peaks observed at 1.0 and 1.8 eV.

We start from the t_{2g} -only Kanamori model with $10Dq = \infty$ in cubic symmetry, which describes four of the five excitation energies. The Kanamori model fails to capture the distinct energies of the two states at 1107 and 1164 meV. The eigenstates for finite ζ are described by comparably simple analytic expressions [49, 85, 86]. The absolute energies $E(i)$ for $i = 0$ to 4 are plotted in Fig. 6a) and in ascending order are given by (for $n \in \{0, 1, 2\}$) [49]

$$E(0) = E_0, \quad E(1) = E_2, \quad E(2) = 3J_H, \quad E(3) = E_1, \quad E(4) = 5J_H, \quad (7)$$

$$\frac{E_n}{J_H} = \frac{8}{3} - \frac{\tau}{3} \cdot \cos \left[\frac{1}{3} \arccos \left(\frac{4(284 - 3\tau^2)}{\tau^3} \right) + 2\pi \frac{n}{3} \right], \quad \tau = 2\sqrt{19 + 3 \left(\frac{3}{2} \frac{\zeta}{J_H} \right)^2}.$$

The four excitation energies result after subtracting the ground-state energy $E(0)$. Finite ζ splits the excitations at $3J_H$ and $5J_H$. For intermediate coupling with $\zeta/J_H \approx 1$, this splitting is of the

¹⁰In L -edge RIXS, such vibronic sidebands have been resolved in $5d^5$ $K_2\text{IrCl}_6$ [36].

order of $\zeta/2$. In the limit of jj coupling, the ground state shows all three electrons in $j = 3/2$ states, and one can excite one or two electrons to $j = 1/2$ with respective excitation energies $3\zeta/2$ and 3ζ . Reasonable agreement between experiment and Kanamori model is obtained for $\zeta/J_H \approx 0.8$ with $J_H = 343$ meV [49], see blue symbols in Fig. 6a). We conclude that K_2ReCl_6 is in the intermediate-coupling regime, similar to K_2OsCl_6 . Spin-orbit coupling ζ is sizable but not large enough for a description in terms of jj coupling.

Spin-orbit coupling mainly mixes the Jahn-Teller-active 2T_2 multiplet into the ground state. However, the admixture increases only $\propto (\zeta/J_H)^2$. A significant Jahn-Teller effect requires strong spin-orbit coupling $\zeta/J_H \gtrsim 1.5$ [53]. In K_2ReCl_6 with $\zeta/J_H \approx 0.8$, the expectation value $\langle L_z \rangle \approx (1/25)(\zeta/J_H)^2 \approx 0.027$ is still very small and more than 97 % of the ground-state weight is contributed by the $S = 3/2$ 4A_2 multiplet [49]. This agrees with a Curie-Weiss analysis of the magnetic susceptibility [87], which finds an effective magnetic moment $\mu_{\text{eff}} \approx 3.81 \mu_B$ that is close to the value $2\sqrt{S(S+1)} \mu_B \approx 3.87 \mu_B$ expected for a $S = 3/2$ compound.

For the excitation energies, the agreement between experiment and theory can be improved by considering the entire $5d$ shell, including the e_g orbitals. This in particular explains the peak splitting around 1.1 eV. In K_2ReCl_6 , infrared spectroscopy detects the intra- t_{2g} excitations with high accuracy, but it cannot resolve the on-site $d-d$ excitations to e_g states with energies above the Mott gap, see Fig. 5. However, the t_{2g} -to- e_g excitations give a strong peak at 3.3 eV in RIXS. The combination of RIXS and optical spectroscopy is a most powerful tool for studying multiplets in Mott insulators. Together, these techniques provide a comprehensive picture of the electronic structure, enabling quantitative extraction of microscopic parameters. Figure 6b) gives the on-site d^3 excitation energies for a model for the entire d shell. The first panel depicts the effect of the cubic crystal field, $10Dq$, and the second panel shows the splitting of the cubic multiplets by Coulomb interactions. The 2E and 2T_1 multiplets are degenerate in the Kanamori model, but this degeneracy is lifted by finite $10Dq$ even for $\zeta = 0$. The third panel shows the effect of spin-orbit coupling, which finally yields five intra- t_{2g} excitation energies. A fit of the five experimental intra- t_{2g} excitation energies and of the energy of the RIXS peak at 3.3 eV yields $10Dq = 3.25$ eV, $\zeta_d = 290$ meV, and $J_{H,d} = 464$ meV (with $F^2 = 3.93$ eV and $F^4/F^2 = 36/55$), see symbols in Fig. 6b). For these parameters, a RIXS spectrum calculated with QUANTY [88, 89], using the experimental resolution 295 meV, agrees very well with the experimental data, see Fig. 4b).

5 Cluster Mott insulators

We now switch from conventional Mott insulators to Mott-insulating materials with lattices containing small clusters such as dimers, trimers, or tetramers. Cluster Mott insulators are a novel class of transition-metal compounds in which electrons are delocalized over, e.g., individual dimers due to strong *intra*-cluster hopping t [28, 90–97], cf. Figs. 7a) and 8a). Intra-cluster hopping dominates over Coulomb repulsion, but Coulomb wins against *inter*-cluster hopping. A set of dimers with one orbital per site is the most simple example. In the limit of large intra-dimer hopping, quasimolecular bonding and antibonding orbitals are formed. With, e.g., one

electron per dimer, the lattice of dimers hosts a half-filled band of bonding orbitals. A cluster Mott insulator may arise for sufficiently small inter-dimer hopping such that Coulomb repulsion suppresses inter-cluster charge fluctuations.

In Sec. 4, we addressed conventional Mott insulators and discussed how the electronic multiplets depend on the competition of Hund's coupling and spin-orbit coupling. In a Mott insulator, virtual inter-site hopping processes yield exchange interactions. In a cluster Mott insulator, the "local" physics is much richer. One may hope that this is also valid for the exchange couplings generated by inter-cluster hopping. In, e.g., iridate dimers, the magnetic moments were predicted to be anisotropic and temperature dependent due to the existence of low-energy electronic states [98]. In general, the quasimolecular magnetic moments are determined by Coulomb interactions, spin-orbit coupling, crystal fields, and intra-cluster hopping. This yields a multitude of cluster states with the possibility of tuning the character of the magnetic moments [91, 92, 94], as discussed below. The ultimate goal of such tuning is to engineer the interactions between quasimolecular moments to realize novel magnetic phases such as quantum spin liquids.

For a dimer in the limit of strong hopping, the electrons are delocalized in (anti-)bonding orbitals. However, also the conventional Mott limit can be realized on a dimer, with electrons that are localized on individual sites experiencing exchange coupling. This limit requires strong electronic correlations dominating over intra-cluster hopping. RIXS is well suited to distinguish these two limits and demonstrate the quasimolecular electronic structure of cluster Mott insulators [90–94, 99–101]. Moreover, RIXS can quantify the competition of hopping, correlations, spin-orbit coupling, and crystal-field splitting, even in the intriguing crossover regime when both hopping and correlations are relevant [94].

5.1 RIXS interferometry

For an intuitive picture, we start from *non-resonant elastic* x-ray scattering, a standard tool for studying the crystal structure. Bragg peaks result from interference of the scattered x-rays, i.e., from a coherent sum over scattering events on all lattice sites. Now consider *resonant* elastic scattering. For an incident photon energy at, e.g., the Ir L edge, the signal mainly stems from scattering on Ir ions and, to good approximation, can be derived by coherently summing over Ir sites only. In RIXS, *resonant inelastic* scattering, we have to coherently sum all the paths that lead from the ground state $|0\rangle$ to a given excited state $|f\rangle$. If this excited state is delocalized over – and restricted to – a dimer, the corresponding scattering events can occur only on the two Ir sites of this dimer. More precisely, at the Ir L edge, the intermediate state shows a $2p$ core hole (cf. Fig. 1) on either of the two Ir sites. Summing coherently over the corresponding scattering events yields two-beam interference of the scattered x-rays, see Fig. 7c), causing a sinusoidal intensity modulation $I(\mathbf{q})$ that is equivalent to an inelastic incarnation of Young's double-slit experiment. For dimers, sinusoidal modulation of the RIXS intensity was predicted in the mid 90's [99, 100] and observed in 2019 [90]. At the L edge, the two "slits" effectively correspond to Ir $2p$ core levels that are involved in the intermediate state of the RIXS process. With a size of a few picometers, they can be considered as nearly point-like, providing a nearly constant

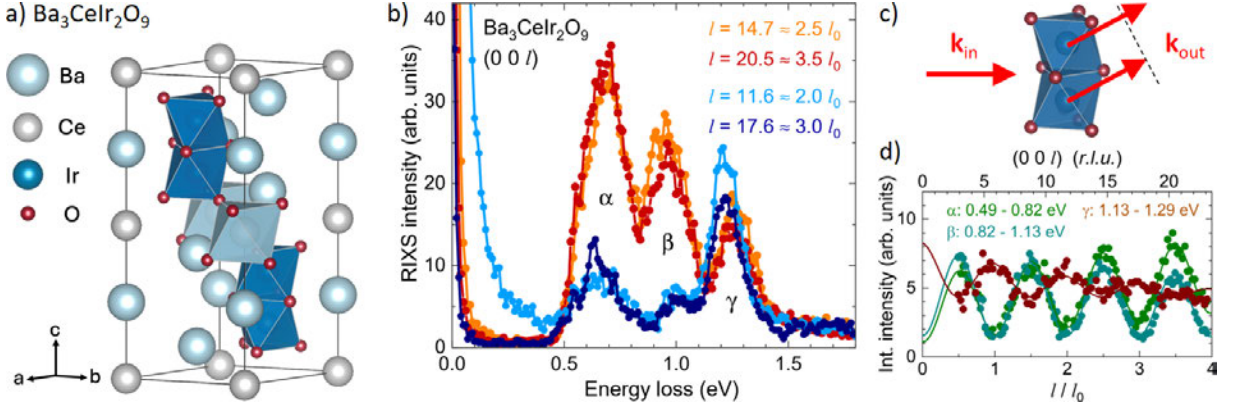


Fig. 7: RIXS on the two-hole dimer $\text{Ba}_3\text{CeIr}_2\text{O}_9$ [90]. *a)* Crystal structure of $\text{Ba}_3\text{CeIr}_2\text{O}_9$ with dimers built from face-sharing IrO_6 octahedra. The intra-dimer Ir-Ir distance is given by d . *b)* L_3 -edge RIXS spectra at selected $\mathbf{q}$ points, measured at 20 K with 27 meV resolution. The peak intensities vary strongly with $\mathbf{q}$. The strong intensity of the elastic line at (0 0 11.6) (light blue) reflects a scattering angle far away from 90° . *c)* Two-beam interference for RIXS on a dimer. *d)* Integrated intensities of the three RIXS peaks for a wide range of q_c , the component of $\mathbf{q}$ parallel to the dimer axis. The top (bottom) axis shows q_c in units of $2\pi/c$ ($l_0 = 2\pi/d$). The two peaks below 1.1 eV agree with $\sin^2(q_c d/2)$ behavior, the third peak predominantly shows $\cos^2(q_c d/2)$ modulation, cf. Eq. (10).

envelope for the intensity modulation.¹¹ This is important since it allows us to study the RIXS interference pattern over a very large range of $\mathbf{q}$, e.g., 20 Brillouin zones in Fig. 7d). We call this approach RIXS interferometry.

For a single dimer, the RIXS intensity for excitation from the ground state $|0\rangle$ to an excited state $|f\rangle$ with the excitation energy $\hbar\omega_f$ is given by

$$I(\mathbf{q}, \omega) = |A_f(\mathbf{q})|^2 \delta(\omega - \omega_f), \quad (8)$$

where¹² $A_f(\mathbf{q})$ is the coherent sum of the amplitudes of all possible scattering processes on the two dimer sites giving rise to the excitation $|f\rangle$. This sum is the source of the interference. In a crystal, any dimer may be excited. In the total RIXS intensity, this yields an *incoherent* sum over all dimers, summing the intensities (not the amplitudes) of all dimers. A cluster shows discrete excitation energies without dispersion. Nevertheless, the RIXS intensity strongly depends on $\mathbf{q}$. Calculating the amplitude $A_f(\mathbf{q})$ is significantly simplified in the fast-collision approximation [1], which in the hard x-ray range is often appropriate for direct RIXS processes due to the very short core-hole lifetime of only a few femtoseconds. In this case, the amplitude can be split into a resonant prefactor and the dynamical structure factor $S(\mathbf{q}, \omega_f)$. For constant incident energy, the former can be neglected. In the dipole approximation, this yields [1]

$$A_f(\mathbf{q}) \sim \langle f | \sum_{\mathbf{R}} e^{i\mathbf{q}\mathbf{R}} [D^\dagger(\epsilon_{\text{out}}^*) D(\epsilon_{\text{in}})]_{\mathbf{R}} | 0 \rangle, \quad (9)$$

¹¹This is similar to neutron scattering on nuclei. In contrast, the form factor for magnetic neutron scattering depends on the more extended magnetization density, so that the intensity is suppressed with increasing $\mathbf{q}$.

¹²The amplitude also depends on the incident energy and the polarizations of incident and scattered photons [1].

where $\mathbf{R}$ runs over all sites over which the final state $|f\rangle$ is delocalized, D denotes the local dipole transition operator, and ϵ_{in} and ϵ_{out} give the polarization of incident and scattered photons, respectively. For simplicity, we consider a dimer with $\mathbf{R}_{1,2} = (0, 0, \pm d/2)$, i.e., Ir-Ir distance d . In the presence of inversion symmetry, the eigenstates show either even or odd parity, and the dipole matrix elements on the two sites may only differ by sign. The coherent sum over resonant scattering processes on the two dimer sites yields

$$I_+(\mathbf{q}) \sim (e^{i\mathbf{q}\mathbf{R}_1} + e^{i\mathbf{q}\mathbf{R}_2})^2 \sim \cos^2(q_c d/2), \quad I_-(\mathbf{q}) \sim (e^{i\mathbf{q}\mathbf{R}_1} - e^{i\mathbf{q}\mathbf{R}_2})^2 \sim \sin^2(q_c d/2), \quad (10)$$

where q_c denotes the z component of $\mathbf{q}$ parallel to the dimer axis, while I_+ (I_-) applies to excitations that preserve (flip) parity. The $\mathbf{q}$ -dependent RIXS intensity thus contains a structural fingerprint of the cluster, here the intra-dimer distance d . Moreover, important information on the symmetry and character of the electronic states is encoded in the $\cos^2(q_c d/2)$ vs. $\sin^2(q_c d/2)$ behavior. The original elastic version of Young's double-slit experiment shows the $I_+(\mathbf{q})$ modulation since the final state and initial state are identical. In contrast, in RIXS an excitation from, e.g., an even-parity bonding state to an odd-parity anti-bonding state is described by $I_-(\mathbf{q})$, essentially providing information on the relative phase of the wavefunction on the two sites.

In summary, the interference pattern directly probes the wave function of cluster Mott insulators and carries information on the symmetry and character of the quasimolecular electronic states [90–95]. It is straightforward to extend the concept to larger clusters such as trimers [93] (see below) or tetrahedra [92]. Related studies of modulated RIXS intensities have been reported for, e.g., homonuclear diatomic molecules such as O_2 [5, 102], the magnetic excitations of the van der Waals ferromagnet $\text{Fe}_{4.75}\text{GeTe}_2$ [103], and in the context of witnessing entanglement [104, 105]. A sinusoidal intensity modulation can also be observed for magnetic excitations when magnetic correlations along a given direction are restricted to pairs of nearest neighbors, e.g., in bilayers [106, 107] or Kitaev materials [20].

5.2 Face-sharing $5d$ iridates: dimers, trimers, and disorder phenomena

A cluster Mott insulator is based on strong intra-cluster hopping. This is realized in, e.g., $5d$ transition-metal compounds with MO_6 octahedra in face-sharing geometry found in the large family of hexagonal perovskites [108]. In, e.g., $\text{Ba}_3\text{CeIr}_2\text{O}_9$, the intra-dimer Ir-Ir distance d is about $2.5 \text{ \AA}$, see Fig. 7a), shorter than in Ir metal. The analysis of our RIXS data yields direct hopping of the order of 1 eV [90]. An attractive feature of the family of $6H$ -perovskites $\text{Ba}_3M\text{Ir}_2\text{O}_9$ is that it allows for M ions with different valence, e.g., Ta^{5+} , Ce^{4+} , In^{3+} , or Ca^{2+} . Thus, iridate dimers can be realized with one to four t_{2g} holes and different magnetic moments. Related dimers with even larger t_{2g} hole counts exist, e.g., in the $4d$ ruthenates $\text{Ba}_3M\text{Ru}_2\text{O}_9$. The non-magnetic dimer compound $\text{Ba}_3\text{CeIr}_2\text{O}_9$ hosts Ir^{4+} ions. The $\text{Ir}^{4+} t_{2g}^5$ configuration with a single t_{2g} hole per site is well known for the formation of local spin-orbit-entangled $j = 1/2$ moments. In a Mott picture, the non-magnetic character would result from strong antiferromagnetic exchange within a dimer, causing a singlet ground state. In a cluster Mott scenario, the two holes fill a bonding orbital, see Fig. 8b). Spectroscopic measurements can tell the difference. In

L_3 -edge RIXS, the $j = 1/2$ moments have a clear fingerprint: the excitation to $j = 3/2$ at 1.5ζ , about 0.6 to 0.7 eV. The three-peak structure of the RIXS spectra of $\text{Ba}_3\text{CeIr}_2\text{O}_9$ is clearly different, see Fig. 7b), pointing to a cluster Mott picture. Indeed, the $\mathbf{q}$ dependence firmly establishes the quasimolecular character. The RIXS peaks do not show a dispersion, but their intensity strongly depends on $\mathbf{q}$. Figure 7d) shows the integrated intensities $I(q_c)$ of the three RIXS peaks over a very wide range of q_c , spanning roughly 20 Brillouin zones. The period of the interference pattern equals $2\pi/d$, being incommensurate with the size of the Brillouin zone $2\pi/c$. The two peaks below about 1.1 eV reveal a $\sin^2(q_c d/2)$ behavior, while the third peak at 1.2 eV is of $\cos^2(q_c d/2)$ type. This pinpoints the electrons (or holes) being delocalized over a given dimer and provides valuable information on the character of the excitations.

A dimer built from face-sharing IrO_6 octahedra shows mirror symmetry but breaks inversion symmetry. This mixes even- and odd-parity states such that the RIXS intensity for a single dimer follows $I(q_c) \propto [\alpha \sin(q_c d/2) + \beta \cos(q_c d/2)]^2$. However, the crystal structure shows two dimer orientations, rotated by π around the dimer axis, see Fig. 7a). Adding the RIXS intensities of the two orientations yields $I_{\text{tot}}(q_c) \propto [\alpha \sin(q_c d/2)]^2 + [\beta \cos(q_c d/2)]^2$, a sinusoidal modulation with a finite offset [90]. The data demonstrate sinusoidal behavior over a very wide range of $\mathbf{q}$ with only a small change of amplitude, i.e., with a slowly varying envelope. This reflects the nearly point-like nature of the $2p$ core hole, as mentioned above. The envelope is hardly affected by the negligible “slit width” but reflects polarization effects, since a change of q requires a change of the scattering geometry, see Fig. 1a). Thus, α and β effectively vary with $\mathbf{q}$.

Now we address the quantitative analysis. The dimer structure with a 3-fold rotation axis splits the t_{2g} orbitals into the a_{1g} and e_g^π orbitals, see Fig. 8a), quantified by Δ_{trig} . Hopping is diagonal in the a_{1g} - e_g^π basis and forms the corresponding (anti-)bonding states, see top panel of Fig. 8b). Assuming that strong hopping suppresses the role of spin-orbit coupling, the quasimolecular ground state is approximated as a spin singlet with the two holes in the bonding a_{1g} orbital. In the opposite limit of strong spin-orbit coupling, we start from local $j = 1/2$ and $3/2$ states. In this picture, hopping yields (anti-)bonding j states, see bottom panel of Fig. 8b). Note that hopping is diagonal in the j basis for $t_{e_g^\pi}/t_{a_{1g}} = -1$ but not for realistic ratios $t_{e_g^\pi}/t_{a_{1g}} \approx -1/2$. Still, the ground state in this picture hosts both holes in a spin-orbit-entangled bonding orbital, $J = 0$ instead of $S = 0$. The electronic parameters allow us to tell which of the two limits is closer to reality. A thorough comparison of the RIXS data with model calculations – considering only the t_{2g} states – yields $t_{a_{1g}} = 1.1$ eV, $t_{e_g^\pi} = -0.5$ eV, $U = 1.0$ eV, $J_H = 0.3$ eV, spin-orbit coupling $\zeta = 0.45$ eV, and the trigonal crystal-field splitting $\Delta_{\text{trig}} = -0.2$ eV [90]. This parameter set quantifies the quasimolecular character. In the Mott limit with dominant Coulomb repulsion, each dimer site carries one hole, and virtual hopping processes mediate exchange couplings. With increasing hopping, states with the two holes on the same site are more and more mixed into the ground state. In the quasimolecular limit of dominant hopping, the two holes occupy a bonding orbital. In this limit, the ground state shows equal weights of states with one hole per site or with double occupancy. For the parameters determined for $\text{Ba}_3\text{CeIr}_2\text{O}_9$, these weights are plotted in the left panel of Fig. 8c) as a function of $t_{a_{1g}}$ (for fixed $t_{e_g^\pi}/t_{a_{1g}} = -0.45$). The weight of states with double occupancy already increases strongly for small $t_{a_{1g}}$ and reaches more than

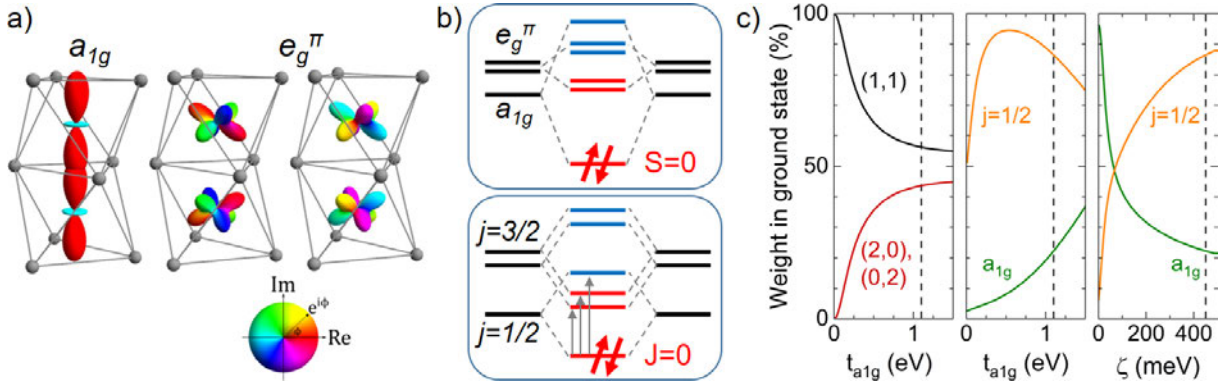


Fig. 8: Quasimolecular electronic structure of a two-hole dimer [90]. *a)* t_{2g} orbitals in trigonal symmetry: a_{1g} and e_g^π orbitals. The color code depicts the phase. For face-sharing dimers, hopping is diagonal in the a_{1g} - e_g^π basis. *b)* In the limit of large hopping, a non-interacting picture is appropriate, where hopping forms bonding (red) and anti-bonding orbitals (blue). Top: Neglecting spin-orbit coupling ζ , two holes occupy the bonding a_{1g} orbital in the ground state. Bottom: For strong ζ , the quasimolecular (anti-)bonding orbitals originate from spin-orbit-entangled j states. This picture correctly describes the $\sin^2$ vs. $\cos^2$ interference patterns of the three RIXS peaks (grey arrows; cf. Fig. 7), providing an intuitive picture of the physics. *c)* Weights of contributions to the ground state wavefunction. Left: In the Mott limit for $t_{a1g} \rightarrow 0$, each site hosts one hole (black). With increasing hopping, the contribution of states with double occupancy increases (red). Both curves extrapolate to 50 % for $t_{a1g} \rightarrow \infty$. Dashed: $t_{a1g} = 1.1$ eV, highlighting the quasimolecular character. Middle and right: Weights for the two holes being either in the bonding $j = 1/2$ orbital or in the bonding a_{1g} orbital. The latter dominates for $t_{a1g} \gtrsim 2$ eV or for small ζ . For realistic parameters, the ground state is well approximated by two holes occupying the bonding $j = 1/2$ orbital. The plots use the parameters found for $\text{Ba}_3\text{CeIr}_2\text{O}_9$ ($t_{a1g} = 1.1$ eV, $t_{e_g^\pi}/t_{a1g} = -0.45$, $U = 1.0$ eV, $J_H = 0.3$ eV, $\zeta = 0.45$ eV, $\Delta_{\text{trig}} = -0.2$ eV), scanning either t_{a1g} or ζ .

40 % for realistic hopping. This provides a quantitative foundation for the quasimolecular description of $\text{Ba}_3\text{CeIr}_2\text{O}_9$ and for using the single-particle-like pictures of (anti-)bonding orbitals in Fig. 8b). Hopping is so large that correlations do not play a prominent role for a single dimer. However, correlations still suppress inter-cluster charge fluctuations.

Comparing quasimolecular states of a_{1g} and spin-orbit entangled j character, the latter provide a better basis for an intuitive understanding. With $\zeta \approx 0.4$ – 0.5 eV, spin-orbit coupling is a decisive factor even for large hopping. For the ground state, the middle and right panels of Fig. 8c) plot the weights of states with double occupancy of the bonding a_{1g} or bonding $j = 1/2$ orbitals. In the limit of extreme intra-cluster hopping, the compound is a band insulator with a filled band of bonding a_{1g} states. However, for $t_{a1g} = 1.1$ eV, the occupation of the bonding a_{1g} orbital is reduced to roughly 20 %. In contrast, in the j basis, double occupancy of the bonding $j = 1/2$ orbital describes about 85 % of the ground state. Interestingly, the quasimolecular j states already dominate for ζ larger than 0.1 eV, see right panel of Fig. 8c). The scenario of quasimolecular j states also provides an intuitive description of the observed q dependence. The two lowest RIXS peaks are assigned to excitations to bonding orbitals originating from $j = 3/2$ states. These are split by Δ_{trig} , J_H , and the hopping ratio $t_{e_g^\pi}/t_{a1g} \neq -1$. In contrast, the third

peak corresponds to the excitation to an anti-bonding state, see vertical grey arrows in Fig. 8b). This agrees with the $\sin^2$ modulation of the lower two peaks and the $\cos^2$ behavior of the third.

Dimers with an odd number of holes: An odd number of holes yields a magnetic moment, underscoring the important role of spin-orbit coupling. For different hole counts, phase diagrams have been discussed as a function of ζ and U (for fixed J_H/U) [98]. We restrict ourselves to a brief discussion of the three-hole dimer compound $\text{Ba}_3\text{InIr}_2\text{O}_9$. It exhibits persistent spin dynamics down to 20 mK [109] and thus has been discussed as a promising spin-liquid candidate. A quantitative understanding of the magnetic moments is a prerequisite for discussing magnetism. Using RIXS interferometry, the quasimolecular nature has been established for $\text{Ba}_3\text{InIr}_2\text{O}_9$ [91]. The central aspect is that, with three holes per dimer, the actual character of the magnetic moments can be tuned via the size of hopping. In the ground state, two holes fill the bonding $j = 1/2$ orbital but the orbital occupied by the third hole depends on the electronic parameters. It occupies the anti-bonding $j = 1/2$ orbital for small hopping, but the bonding $j = 3/2$ orbital for larger hopping. The value of hopping thus determines the character of the quasimolecular magnetic moment. The compound $\text{Ba}_3\text{InIr}_2\text{O}_9$ is found to be very close to this phase transition, on the large-hopping side [91, 98]. This is a simple example for a highly promising feature of cluster Mott insulators: in general, the magnetic moments may be tunable via electronic parameters.

Disorder: Disorder may have an unusual effect on magnetism in cluster Mott insulators. Substituting a magnetic ion in, e.g., a dimer by a nonmagnetic one does not create a simple vacancy but changes the magnetic moment of this dimer from quasimolecular to single-site type, which also affects the exchange interactions. RIXS interferometry is very well suited to disentangle the different local moments in a disordered system and to assign the excitations to different transition-metal sites, as shown for $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$ [95], see Fig. 9. The sister compound $\text{Ba}_3\text{CeIr}_2\text{O}_9$, discussed above, shows perfect cation ordering, with the Ir ions occupying the $4f$ sites and the much larger Ce ions the $2a$ sites. In $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$, a strong Ti-Ir site disorder is caused by the comparable ionic radii of the Ir^{4+} and Ti^{4+} ions [110]. This allows for x different from 2 with Ir and Ti ions randomly distributed over the $4f$ and $2a$ sites [95]. With strong site disorder, the discussions of $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$ as a spin-liquid candidate [111–113] must be reconsidered.

Counterintuitively, the average magnetic moment of $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$ *decreases* as the concentration x of magnetic Ir^{4+} ions increases [112]. This can be rationalized via the coexistence of non-magnetic quasimolecular Ir_2 dimers (see $\text{Ba}_3\text{CeIr}_2\text{O}_9$) and single-site Ir^{4+} $j = 1/2$ moments, where the relative contribution of non-magnetic dimers increases with increasing x . For small x , the Ir ions are diluted throughout the lattice. For $x = 0.3$, RIXS data show a narrow peak at 0.57 eV, roughly 1.5ζ , revealing the presence of individual $5d^5$ Ir^{4+} sites, see Fig. 9a). This so-called spin-orbit exciton corresponds to the local excitation from the $j = 1/2$ ground state to a $j = 3/2$ excited state, which has also been observed in K_2IrCl_6 [36] and many other $5d^5$ iridates. In $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$, the shoulder at 0.7 eV is caused by the non-cubic crystal field. For large $x = 1.5$ or 1.8 , the presence of dimers yields further RIXS features, but the narrow peak

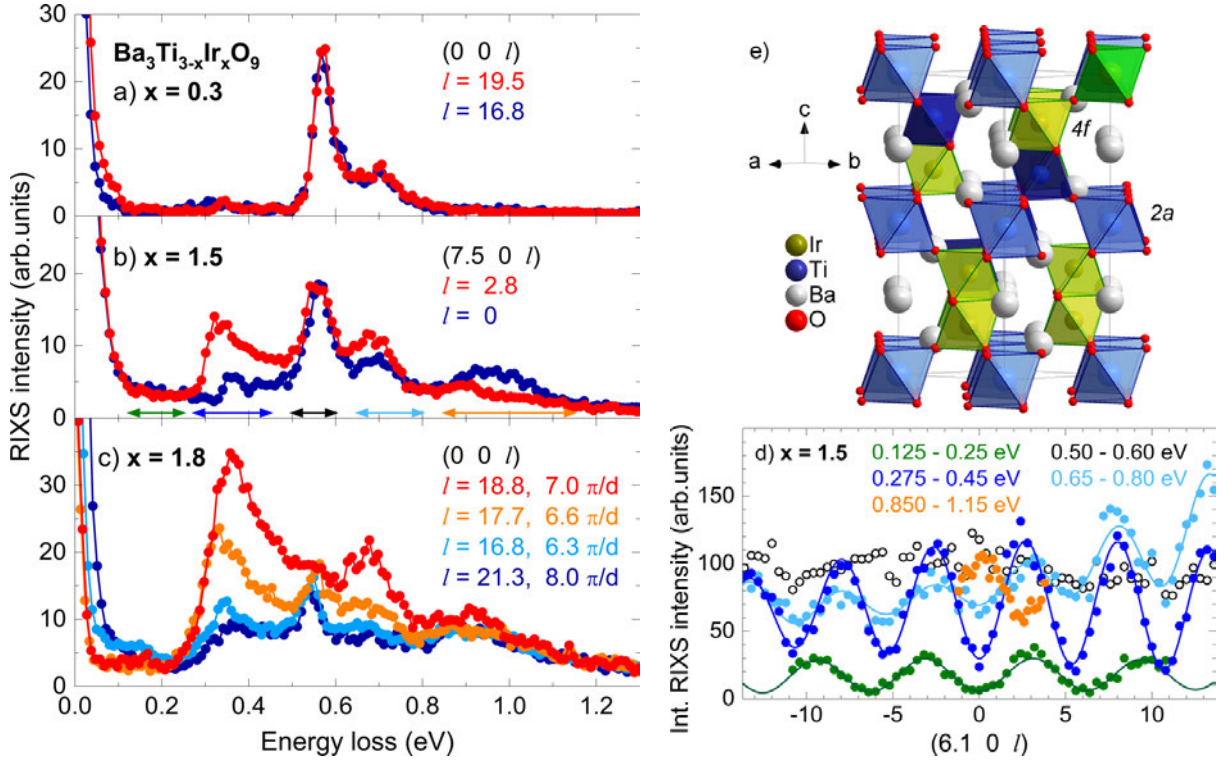


Fig. 9: Strong Ti-Ir site disorder in $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$ [95]. a)-c) RIXS spectra for $x = 0.3, 1.5$, and 1.8 at 20 K . d) Integrated RIXS intensity for $x = 1.5$. The corresponding energy ranges are marked by arrows in b). We find three types of excitations: (i) The narrow peak of the spin-orbit exciton is the hallmark feature for individual $j = 1/2$ sites. The RIXS intensity of this single-site excitation does not depend on q but may show some polarization dependence (black dots in d)). (ii) Quasimolecular face-sharing Ir_2 dimers ($4f$ sites) contribute broad peaks above 0.2 eV with sinusoidal intensity modulation. (iii) Exchange-coupled $\text{Ir } 2a\text{-}4f$ pairs yield a magnetic excitation at 0.15 eV , which is revealed by the larger period of the interference pattern. e) Sketch of the crystal structure, cf. Fig. 7a). Ti-Ir site disorder yields Ir ions on $2a$ sites (green) and Ti ions on $4f$ dimer sites (dark blue). This produces $2a\text{-}4f$ pairs and single Ir^{4+} $j = 1/2$ moments.

of the spin-orbit exciton prevails, demonstrating that individual $j = 1/2$ sites are still present. For all x , the 0.57 eV peak does not show any q dependence, neither in excitation energy nor in intensity, see Fig. 9a)-d), clearly establishing its local, single-site character.

The RIXS spectra reveal the crossover from predominant $j = 1/2$ sites for small x to mainly dimers for x closer to 2. For $x = 1.5$ and 1.8 , the three broad RIXS peaks at about $0.35, 0.7$, and 0.95 eV can be assigned to excitations on dimers. This is firmly established by the sinusoidal q dependence of the integrated RIXS intensity, see Fig. 9d). The lower two peaks show $\sin^2(q_c d/2)$ behavior (blue and light blue), while a $\cos^2(q_c d/2)$ response is observed around 0.95 eV (orange). This is fully equivalent with the results discussed for $\text{Ba}_3\text{CeIr}_2\text{O}_9$ with RIXS peaks at $0.7, 1.0$, and 1.2 eV . The reduced energy scale in $\text{Ba}_3\text{Ti}_{3-x}\text{Ir}_x\text{O}_9$ roughly can be attributed to the larger intra-dimer Ir-Ir distance $d \approx 2.65\text{ \AA}$ [110] compared to $2.54\text{ \AA}$ for $M = \text{Ce}$. A larger d yields a smaller bonding-antibonding splitting. Note that $d = 2.65\text{ \AA}$ agrees with the observed period of approximately $5.3 \times 2\pi/c$, see Fig. 9d).

The power of RIXS interferometry to unravel disorder phenomena is most evident from the analysis of the weak feature at 0.15 eV found for large x . Its integrated RIXS intensity (green symbols in Fig. 9d)) shows a sinusoidal two-site interference pattern. However, the period is about 17 % larger, corresponding to a distance of (2.26 ± 0.03) Å projected onto the c axis. This agrees with the projected distance of 2.22 Å between the $2a$ and $4f$ sites reported in x-ray diffraction at 300 K [110]. The interference pattern thus pinpoints the nature of the 0.15 eV peak. Induced by Ti-Ir site disorder, it corresponds to a magnetic excitation of Ir pairs with one Ir ion on a $2a$ site and one on a $4f$ dimer site. The $2a$ and $4f$ sites are also displaced *perpendicular* to c , and the assignment is corroborated by the observation of the corresponding intensity modulation for q within the ab plane [95]. The excitation energy of 0.15 eV can be explained by considering two $j = 1/2$ moments with strong Heisenberg exchange via a 180° Ir-O-Ir bond, equivalent to the case of square-lattice Sr_2IrO_4 with magnon energies up to 0.2 eV [8, 10]. Note the difference between a pair of exchange-coupled local moments and quasimolecular orbitals. Compared to face-sharing octahedra, hopping t is much smaller for a 180° bond (corner-sharing geometry). In this case, Coulomb repulsion U suppresses a quasimolecular character, giving rise to two $j = 1/2$ moments with exchange coupling $J \propto t^2/U$. In RIXS, the magnetic excitation of such a spin pair also yields a two-site interference pattern. A stunning example of this has been found for the magnetic excitations of the Kitaev-compound Na_2IrO_3 with spin-spin correlations that are restricted to two nearest-neighbor sites [20].

Trimers: Stacks of face-sharing MO_6 octahedra also yield linear trimers with large hopping, see Fig. 10c). An example is $\text{Ba}_4\text{AM}_3\text{O}_{12}$ with $M = \text{Mn, Ru, Rh, and Ir}$, where the valence of the A ions controls the hole count [108, 114–117]. Based on *ab initio* calculations, a transition from localized electrons for $3d$ Mn to quasimolecular behavior for $5d$ Ir has been suggested [118]. However, the results reported for different compounds paint a vivid picture. Unconventional behavior is found in $4d$ $\text{Ba}_4\text{Nb}_{1-x}\text{Ru}_{3+x}\text{O}_{12}$, covering the range from a quantum spin liquid to a heavy-fermion strange metal [114]. The three-hole trimer $\text{Ba}_4\text{Ir}_3\text{O}_{10}$ has been claimed to host a quantum spin liquid down to 0.2 K with a very large frustration parameter and spinon-like excitations observed in RIXS [119, 120]. Interestingly, the discussion favors exchange-coupled $j = 1/2$ moments over a quasimolecular scenario. In contrast, for the three-hole trimer $\text{Ba}_5\text{CuIr}_3\text{O}_{12}$ a quasimolecular picture with covalency-driven collapse of spin-orbit coupling has been claimed [121]. In fact, it has often been assumed that the role of spin-orbit coupling is suppressed for large hopping. Also $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$ with nonmagnetic Nb^{5+} ions and two t_{2g} holes per Ir trimer has been discussed as a spin-liquid candidate [115–117]. With two holes per three sites, it is hard to motivate a conventional Mott insulator. In a quasimolecular picture, however, the two holes occupy a bonding orbital, giving rise to a non-magnetic ground state.

Trimers of face-sharing IrO_6 octahedra exhibit inversion symmetry. In a single-particle picture, hopping yields bonding, non-bonding, and antibonding states, see Fig. 10a). Bonding and antibonding states show even parity, but non-bonding states acquire a minus sign upon inversion. The wavefunction of a non-bonding state has no weight on the central trimer site and opposite phases on the two outer sites, see Fig. 10c), such that the energy of non-bonding orbitals is

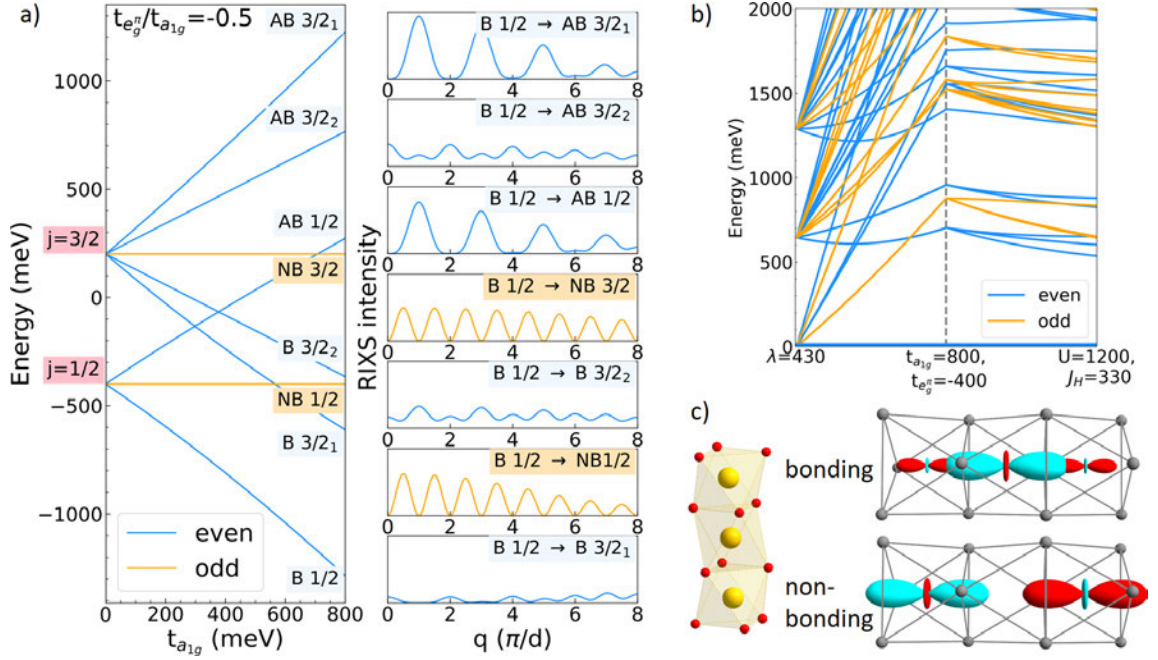


Fig. 10: Quasimolecular states of a linear trimer [93]. *a) Absolute energies of electronic states in a single-particle picture, calculated by exact diagonalization. The energies are given as a function of t_{a1g} for $t_{eg}/t_{a1g} = -1/2$ in the idealized case of cubic crystal field. For $t_{a1g} = 0$, spin-orbit coupling yields $j = 1/2$ and $3/2$ states split by 1.5ζ . Hopping yields even-parity (anti-) bonding states (blue) and odd-parity non-bonding states (orange). The finite splitting of the bonding $j = 3/2$ states reflects that hopping is not diagonal in the j basis (also the antibonding $j = 3/2$ states are split). Still, a description in terms of $j = 1/2$ and $3/2$ states provides a valid, intuitive picture. Right panel: Calculated RIXS intensity for excitations from the bonding $j = 1/2$ state to all other levels. The interference patterns reveal the symmetry of the involved states, see Eq. (11). b) Excitation energies for a two-hole trimer for $\zeta = 0.43$ eV, assuming octahedral site symmetry. In the left panel, t_{a1g} increases to 0.8 eV, as in a), such that the ground state shows two holes in the bonding $j = 1/2$ orbital. For strong hopping, switching on the correlations has only a minor effect (right panel). c) Left: Linear trimer of face-sharing IrO_6 octahedra. The central Ir site is a center of inversion. Right: Simple example for (non-)bonding states in the case of a_{1g} orbitals, highlighting the essential symmetry properties.*

independent of hopping. Evaluating Eq. (9) for a trimer, we find that the modulation pattern of the RIXS intensity of a given excitation is sensitive to the parity of the states involved [93],

$$I_{eo}(q_c) \sim |\alpha|^2 \sin^2(q_c d), \quad I_{ee}(q_c) \sim |\beta|^2 \cos^2(q_c d) + \gamma \cos(q_c d) + |\delta|^2, \quad (11)$$

where I_{eo} refers to excitations from even to odd or vice versa, flipping the inversion eigenvalue, while I_{ee} corresponds to excitations from even to even or from odd to odd, and q_c denotes the component of $\mathbf{q}$ parallel to the trimer axis. These expressions can be motivated intuitively and also apply to the many-body case. An excitation from an even-parity bonding orbital to an odd-parity non-bonding state has vanishing matrix elements on the central site. The RIXS process has to be summed over the two outer sites with distance $2d$. Together with the flip of parity, this results in $\sin^2(q_c(2d)/2)$ behavior. Similarly, the $|\beta|^2 \cos^2(q_c d)$ term in I_{ee} arises from the two

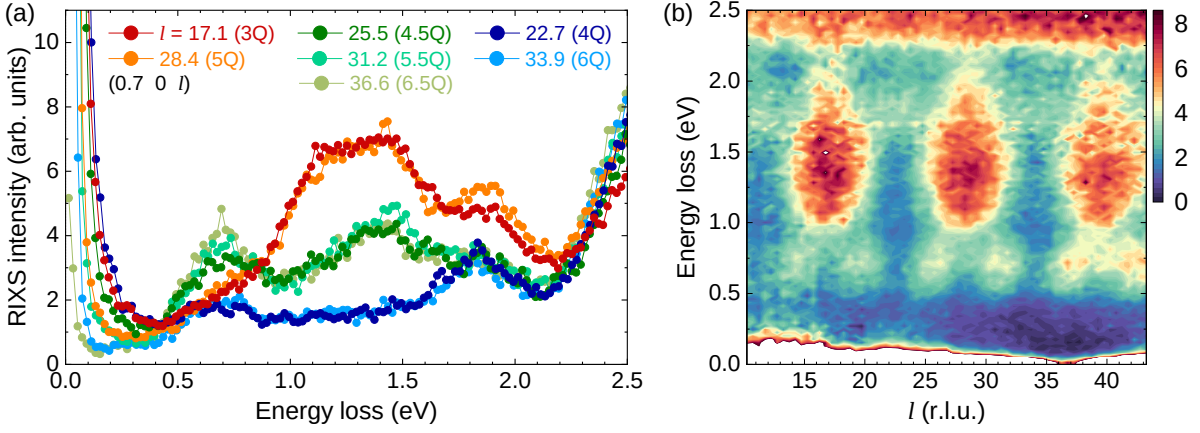


Fig. 11: Quasimolecular intra- t_{2g} excitations of the two-hole trimer $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$ [93]. *a)* RIXS spectra at selected $\mathbf{q}$ points, measured at 300 K at the Ir L_3 edge with 63 meV resolution. The overall behavior is equivalent to the case of Ir dimers, see Fig. 7: Peak intensities depend on q_c (along the trimer axis, parallel to the c axis) but peak energies do not. The elastic line is weak for $q_c = 36.6 \cdot 2\pi/c$ where the scattering angle is close to 90° but increases with decreasing q_c , i.e., decreasing l . *b)* RIXS intensity as a function of q_c and energy loss. The data show a modulation period of $2\pi/(2d) \approx 5.67 \cdot 2\pi/c$ around 0.7 eV and $2\pi/d$ around 1.5 eV. This reflects the Ir-Ir distances d and $2d$ and demonstrates the quasimolecular character.

outer sites, now contributing with the same phase. The $|\delta|^2$ term stems from the scattering at the central site, for which the matrix elements may differ in size compared to the outer sites. Finally, the term proportional to $\gamma = (\beta^* \delta + \delta^* \beta)$ highlights the interference between scattering events at the central and outer trimer sites. Through the prefactors α , β , γ , and δ , the RIXS intensity provides a wealth of information about the quasimolecular wavefunction, see Fig. 10a).

Figure 11a) shows Ir L_3 -edge RIXS spectra of $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$ [93]. Equivalent to the dimer case, the RIXS spectra are rich and the peak intensities depend on q_c while the peak energies do not, pointing to a quasimolecular character. This is confirmed by the $\mathbf{q}$ -dependent RIXS intensity, see Fig. 11b). A linear trimer exhibits two intra-trimer Ir-Ir distances, d and $2d$. The existence of two modulation periods $2\pi/d$ and $2\pi/(2d)$, see Eq. (11), is thus the hallmark signature of a quasimolecular electronic structure. In $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$, the fast modulation of $I_{eo} \propto \sin^2(q_c d)$ with period π/d is observed around 0.7 eV, while the features above 1 eV roughly show $\sin^2(q_c d/2)$ behavior with a period of $2\pi/d$. The latter arises for $\gamma \ll -|\beta|^2$, see Eq. (11). This implies a dominant interference between central and outer sites, requiring a large weight of the wavefunction on the central site and opposite phases on the central site compared to the outer sites. The analysis roughly yields one hole on the central site and half a hole per outer site [93].

The rich physics of trimers offers an excellent opportunity to study the competition of spin-orbit coupling and hopping. Remarkably, a non-interacting picture that combines strong ζ with large hopping provides an intuitive explanation of the RIXS data. At the same time, the key RIXS features strongly restrict the relevant parameter range. Large $\zeta = 0.4\text{--}0.5$ eV is well established in iridates. From this starting point, the absence of RIXS features below 0.4 eV requires large hopping, see Fig. 10a). The RIXS intensity at low energy, around 0.7 eV, shows $\sin^2(q_c d)$ mod-

ulation, implying the excitation to an odd-parity non-bonding state. This requires that hopping does not exceed about 0.8 eV, see Fig. 10b). In this range, a dominant $\sin^2(q_c d)$ response at low energy is possible as the RIXS intensity of the excitation to the lowest (even) bonding $j = 3/2$ orbital nearly vanishes, see bottom right panel of Fig. 10a). Finally, the strong $\sin^2(q_c d/2)$ modulation above 1 eV mainly comes from excitations to antibonding $j = 1/2$ states.

In a non-interacting scenario, the ground state hosts two holes in the bonding $j = 1/2$ orbital, causing non-magnetic behavior. For finite hopping, the labeling as $j = 1/2$ or $3/2$ states is not fully correct since hopping is not diagonal in the j states for $t_{e\pi}/t_{a1g} \neq -1$. Hopping splits the local $j = 3/2$ quadruplets into two bonding doublets, a non-bonding quartet, and two antibonding doublets, see Fig. 10a). The splitting into doublets is caused by a finite mixing of $j = 1/2$ and $3/2$ character. Due to this mixing, the energy of the bonding $j = 1/2$ state is not perfectly linear in hopping. However, the deviations are small so that a discussion in terms of j states is still meaningful. This remains valid even upon considering electronic interactions with $U = 1.2$ eV and $J_H \approx 0.3$ eV, see Fig. 10b). Coulomb interactions slightly affect the excitation energies and lift degeneracies of multiplets. However, this has only a minor effect on the two-hole ground state and the overall behavior. On the trimer, large hopping wins against correlations. This in particular applies to $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$ with fewer holes than Ir sites. The main conclusion is very similar to the case of face-sharing Ir dimers: The two-hole trimer compound $\text{Ba}_4\text{NbIr}_3\text{O}_{12}$ exhibits a quasimolecular electronic structure that, to a large extent, can be understood in a non-interacting picture. Both large spin-orbit coupling and large hopping are decisive. In the ground state, two holes occupy a spin-orbit-entangled bonding $j = 1/2$ orbital.

5.3 Between Mott and cluster Mott: face-sharing 4d Ru dimers

Face-sharing $4d^4$ Ru dimers form a non-trivial bridge between d^4 Mott insulators such as K_2OsCl_6 (Sec. 4.1) and the cluster Mott behavior found in $5d$ Ir dimers (Sec. 5.2). In the latter, large hopping yields quasimolecular behavior, but $4d$ ruthenates realize a different part of phase space. The $4d^4$ dimer $\text{Ba}_3\text{CeRu}_2\text{O}_9$ is located in the intricate but hardly explored crossover regime *between* the Mott limit and the quasimolecular limit [94]. Like the iridate sister compounds, the ruthenates $\text{Ba}_3M\text{Ru}_2\text{O}_9$ exist for M ions with different valences [122–124]. Using Ru L_3 -edge RIXS for $M = \text{Ce}^{4+}$ with four-hole Ru dimers [94], we find that the electronic structure is governed by a delicate balance of competing interactions. Hopping, Hund's coupling, and trigonal crystal-field splitting Δ_{trig} are of comparable size, giving rise to a rich phase diagram [28, 94, 98]. Spin-orbit coupling $\zeta = 0.15$ eV is the smallest electronic parameter but decisive for the spin-orbit-entangled singlet character of the ground state.

The $4d$ Ru L_3 edge is at about 2.84 keV in the tender x-ray range, a challenging regime. RIXS with a resolution of ≈ 60 meV can be measured at PETRA III [24]. Compared to the Ir L edge, the smaller incident energy restricts the range of q that can be studied. Figure 12a) shows RIXS spectra for fixed modulus $|q|$, measured over a wide range of angle of incidence θ . Compared to the d^4 Mott limit, see Fig. 3, the data are very rich. The competition of hopping, J_H , Δ_{trig} , and ζ produces a large number of excitations, preventing a simple peak assignment [94].

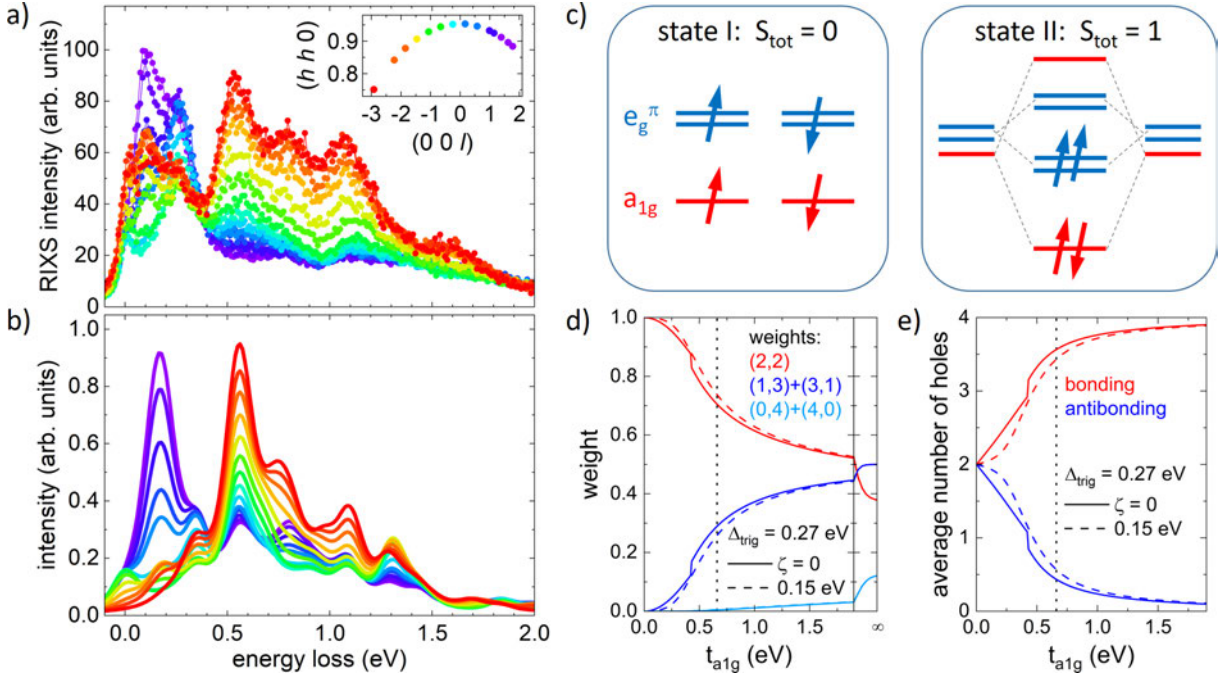


Fig. 12: Between Mott and cluster Mott: the $4d^4$ dimer $\text{Ba}_3\text{CeRu}_2\text{O}_9$ [94]. a) Ru L_3 -edge RIXS spectra measured at 20 K for different angles of incidence. Inset: corresponding $\mathbf{q}$ points. b) Numerical results for $U = 2$ eV, $J_H = 0.26$ eV, $\zeta = 0.15$ eV, $\Delta_{\text{trig}} = 0.27$ eV, $t_{a_{1g}} = 0.66$ eV, and $t_{e_g}/t_{a_{1g}} = -0.45$. c) Sketches of the electronic structure in the Mott limit and the quasimolecular limit for $\Delta_{\text{trig}} > 0$ and $\zeta = 0$. d) Cumulative weights w_{nm} for n holes on one site in the ground state as a function of $t_{a_{1g}}$ for fixed $t_{e_g}/t_{a_{1g}} = -0.45$. Vertical dotted line highlights $t_{a_{1g}} = 0.66$ eV. e) Average number of holes in bonding and antibonding orbitals in the ground state.

The RIXS intensity of all peaks strongly varies with θ . The angle of incidence sets $\mathbf{q}$ as well as the polarization. Unfortunately, it is not trivial to disentangle the $\mathbf{q}$ -dependent interference pattern from polarization effects. As discussed above, the RIXS intensity is the squared sum of several terms, where each term carries its own dependence on $\mathbf{q}$ and polarization. For hard x-ray RIXS on $5d$ compounds, covering several periods of the interference pattern, the polarization determines the envelope (see, e.g., Fig. 9d)). At the Ru L_3 edge, however, RIXS can explore only about one period in $\text{Ba}_3\text{CeRu}_2\text{O}_9$. In this case, a discussion of the modulation requires a careful analysis [94], which we skip here for simplicity.

We determine the electronic parameters by numerical simulation of the RIXS spectra, see Fig. 12b). Best agreement is found for $U = 2$ eV, $J_H = 0.26$ eV, $\zeta = 0.15$ eV, $\Delta_{\text{trig}} = 0.27$ eV, $t_{a_{1g}} = 0.66$ eV, and $t_{e_g}/t_{a_{1g}} = -0.45$. The values for U , J_H , and ζ are well within the expected range for ruthenates, while Δ_{trig} and the hoppings are more material-specific parameters. The numerical results reproduce the key features of the RIXS data, describing the overall peak structure and the pronounced dependence on θ . For small θ , i.e., positive l (purple), the intensity is maximized below about 0.4 eV but suppressed above 0.4 eV.

To understand the puzzling ground state, we first discuss the limiting cases of small and large hopping, considering $\zeta = 0$ for simplicity. In the Mott limit, the on-site Coulomb repulsion U suppresses charge fluctuations. Each Ru site hosts two holes that form spin $S = 1$ due to

Hund's coupling (cubic 3T_1 multiplet, see Tab. 1). For positive Δ_{trig} , there is one hole in the a_{1g} orbital and one in an e_g^π orbital, see state I in Fig. 12c). Hopping is diagonal in the a_{1g} - e_g^π basis and yields exchange interactions. The degenerate e_g^π orbitals exhibit a Kugel-Khomskii-type exchange, favoring parallel spins on the two sites. Nevertheless, the ground state shows $S_{\text{tot}}=0$ since the larger antiferromagnetic coupling driven by $t_{a_{1g}}$ wins. In the opposite quasimolecular limit of large hopping, the lowest bonding orbital has a_{1g} character since $|t_{a_{1g}}| > |t_{e_g^\pi}|$, see state II in Fig. 12c). This yields two holes in the bonding a_{1g} orbital and two holes within the four bonding e_g^π states. The two e_g^π holes form a triplet with $S_{\text{tot}}=1$ due to Hund's coupling.

The intermediate regime, between Mott and cluster Mott, needs a more quantitative discussion. Concerning the character of the ground state, we consider the cumulative weights w_{nm} of terms showing n holes on one site and $m=4-n$ on the other, see Fig. 12d). The Mott limit is characterized by $w_{22} \approx 1$, i.e., roughly two holes per site. With increasing hopping, the weight of states with $n \neq 2$ increases. The quasimolecular limit, reached for infinite hopping, exhibits $w_{22} = 3/8$ and $w_{13}+w_{31} = 1/2$. For the parameters appropriate for $\text{Ba}_3\text{CeRu}_2\text{O}_9$, the weights w_{nm} are plotted in Fig. 12d) as a function of $t_{a_{1g}}$, both for $\zeta = 0$ (solid) and 0.15 eV (dashed). For $\zeta = 0$, the jump of w_{22} close to $t_{a_{1g}} = 0.4$ eV marks the phase transition between the more localized state I with $S_{\text{tot}}=0$ and state II with $S_{\text{tot}}=1$. In the discussion above, state II has been motivated from a quasimolecular perspective, while state I follows naturally for the Mott limit. We emphasize that state II violates Hund's rule (not showing $S=1$ per site) and cannot be motivated by exchange coupling between d^4 sites. Nevertheless, state II is already realized for moderate $t_{a_{1g}} \approx 0.4$ eV, where $w_{22} > 0.8$ still indicates a localized character. With $t_{a_{1g}} = 0.66$ eV (vertical dotted line), $\text{Ba}_3\text{CeRu}_2\text{O}_9$ falls within the regime of state II, but the degree of Mott localization is still sizable. We find $w_{22} = 0.70$ (0.74) for $\zeta = 0$ (0.15) eV. These values are within the intermediate regime but closer to the Mott limit, $w_{22} = 1$, than to the quasimolecular limit, $w_{22} = 3/8$.

The intermediate character of $\text{Ba}_3\text{CeRu}_2\text{O}_9$ is supported by the average number of holes in bonding and antibonding states, n_B and n_{AB} , see Fig. 12e). In the quasimolecular limit, all four holes occupy bonding orbitals, $n_B = 4$. In the Mott limit, the four holes are distributed equally over the (anti-)bonding orbitals, $n_B = n_{AB} = 2$. For the parameters of $\text{Ba}_3\text{CeRu}_2\text{O}_9$, the phase transition between states I and II occurs around $t_{a_{1g}} = 0.4$ eV where $n_B \approx 3$, halfway between the two limits. For $t_{a_{1g}} = 0.66$ eV, we find $n_B = 3.6$ (3.4) for $\zeta = 0$ (0.15) eV. This provides a quantitative measure of the intermediate character of $\text{Ba}_3\text{CeRu}_2\text{O}_9$. With $w_{22} \gtrsim 0.70$, the charge distribution tends to the Mott limit with predominantly two holes per Ru site, minimizing the Coulomb energy. At the same time, the average number of holes in bonding orbitals is large, $n_B \gtrsim 3.4$, pointing to the quasimolecular limit. In the $4d^4$ ruthenate dimer compound $\text{Ba}_3\text{CeRu}_2\text{O}_9$, the dichotomy between localized states in the Mott limit and quasimolecular states in the cluster Mott case breaks down.

In $4d$ ruthenates, compared to iridates, the smaller hopping boosts the role of correlations for the cluster moments. The physics of the intermediate regime has hardly been explored. For $4d^4$ Ru dimers, the vivid competition of Hund's coupling, hopping, and crystal field yields a

series of distinct ground states (for $\zeta = 0$) [94].¹³ Hopping and crystal field can be tuned by, e.g., external pressure, potentially allowing for control of the local physics. Spin-orbit coupling turns the phase transitions into crossovers, but the electronic structure and the low-energy excitations are still highly sensitive to a change of the electronic parameters. In $\text{Ba}_3\text{MRu}_2\text{O}_9$ with M^{3+} ions and five holes per dimer, such a strong sensitivity has been reported [125].

Acknowledgement

It is a pleasure to acknowledge the (former) PhD students most closely involved in this project: A. Revelli, M. Magnaterra, P. Warzanowski, L. Pätzold, E. Bergamasco, J. Attig, and A. Sandberg. The results discussed here are based on several fruitful collaborations. We would like to recognize the long-standing members of our team: M. Moretti Sala, G. Monaco, P. van Loosdrecht, H. Gretarsson, and C. Sahle on the experimental side and M. Hermanns, S. Trebst, and J. van den Brink on the theory side. All of the experimental data shown here were measured on single crystals provided by the group of P. Becker and L. Bohatý. We acknowledge the European Synchrotron Radiation Facility (ESRF) for providing beam time at beamline ID20. We acknowledge DESY – a member of the Helmholtz Association HGF – for access to beam time. Furthermore, we acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Project No. 277146847 – CRC 1238 (project B03).

¹³For $\text{Ba}_3\text{CeRu}_2\text{O}_9$, we focused on $\Delta_{\text{trig}} > 0$, but more states can be realized if we skip this restriction [94].

References

- [1] L.J.P. Ament, M. van Veenendaal, T.P. Devereaux, J.P. Hill, and J. van den Brink, *Rev. Mod. Phys.* **83**, 705 (2011)
- [2] T. Tohyama, *J. Electron Spectrosc. Relat. Phenom.* **200**, 209 (2015)
- [3] M. Mitrano, S. Johnston, Y.-J. Kim, and M. Dean, *Phys. Rev. X* **14**, 040501 (2024)
- [4] F.M. de Groot, M.W. Haverkort, H. Elnaggar, A. Juhin, K.-J. Zhou, and P. Glatzel, *Nat. Rev. Methods Primers* **4**, 45 (2024)
- [5] F. Gel'mukhanov, M. Odelius, S.P. Polyutov, A. Föhlisch, and V. Kimberg, *Rev. Mod. Phys.* **93**, 035001 (2021)
- [6] J. van den Brink: *Resonant Inelastic X-ray Scattering on Elementary Excitations*, in E. Pavarini, E. Koch, J. van den Brink, and G. Sawatzky (Eds.): *Quantum Materials: Experiments and Theory*, Modeling and Simulation, Vol. 6 (FZ Jülich, 2016) <http://www.cond-mat.de/events/correl16>
- [7] L. Braicovich, J. van den Brink, V. Bisogni, M. Moretti Sala, L.J.P. Ament, N.B. Brookes, G.M. De Luca, M. Salluzzo, T. Schmitt, V.N. Strocov, and G. Ghiringhelli, *Phys. Rev. Lett.* **104**, 077002 (2010)
- [8] J. Kim, D. Casa, M.H. Upton, T. Gog, Y.-J. Kim, J.F. Mitchell, M. van Veenendaal, M. Daghofer, J. van den Brink, G. Khaliullin, and B.J. Kim, *Phys. Rev. Lett.* **108**, 177003 (2012)
- [9] J. Kim, A.H. Said, D. Casa, M.H. Upton, T. Gog, M. Daghofer, G. Jackeli, J. van den Brink, G. Khaliullin, and B.J. Kim, *Phys. Rev. Lett.* **109**, 157402 (2012)
- [10] J. Kim, M. Daghofer, A.H. Said, T. Gog, J. van den Brink, G. Khaliullin, and B.J. Kim, *Nat. Commun.* **5**, 4453 (2014)
- [11] M. Le Tacon, G. Ghiringhelli, J. Chaloupka, M. Moretti Sala, V. Hinkov, M.W. Haverkort, M. Minola, M. Bakr, K.J. Zhou, S. Blanco-Canosa, C. Monney, Y.T. Song, G.L. Sun, C.T. Lin, G.M. De Luca, M. Salluzzo, G. Khaliullin, T. Schmitt, L. Braicovich, and B. Keimer, *Nat. Phys.* **7**, 725 (2011)
- [12] J. Schlappa, K. Wohlfeld, K.J. Zhou, M. Mourigal, M.W. Haverkort, V.N. Strocov, L. Hozoi, C. Monney, S. Nishimoto, S. Singh, A. Revcolevschi, J.-S. Caux, L. Patthey, H.M. Rønnow, J. van den Brink, and T. Schmitt, *Nature* **485**, 82 (2012)
- [13] Y. Tseng, E. Paris, K.P. Schmidt, W. Zhang, T.C. Asmara, R. Bag, V.N. Strocov, S. Singh, J. Schlappa, H.M. Rønnow, and T. Schmitt, *Commun. Phys.* **6**, 138 (2023)

- [14] V. Bisogni, L. Simonelli, L.J.P. Ament, F. Forte, M. Moretti Sala, M. Minola, S. Huotari, J. van den Brink, G. Ghiringhelli, N.B. Brookes, and L. Braicovich, *Phys. Rev. B* **85**, 214527 (2012)
- [15] M. Rossi, P. Marabotti, Y. Hirata, G. Monaco, M. Krisch, K. Ohgushi, K. Wohlfeld, J. van den Brink, and M. Moretti Sala, *Eur. Phys. J. Plus* **135**, 676 (2020)
- [16] J. Schlappa, U. Kumar, K.J. Zhou, S. Singh, M. Mourigal, V.N. Strocov, A. Revcolevschi, L. Patthey, H.M. Rønnow, S. Johnston, and T. Schmitt, *Nat. Commun.* **9**, 5394 (2018)
- [17] J. Schlappa, T. Schmitt, F. Vernay, V.N. Strocov, V. Ilakovac, B. Thielemann, H.M. Rønnow, S. Vanishri, A. Piazzalunga, X. Wang, L. Braicovich, G. Ghiringhelli, C. Marin, J. Mesot, B. Delley, and L. Patthey, *Phys. Rev. Lett.* **103**, 047401 (2009)
- [18] J. Li, Y. Gu, Y. Takahashi, K. Higashi, T. Kim, Y. Cheng, F. Yang, J. Kuneš, J. Pelliciari, A. Hariki, and V. Bisogni, *Phys. Rev. X* **13**, 011012 (2023)
- [19] H. Elnaggar, A. Nag, M.W. Haverkort, M. Garcia-Fernandez, A. Walters, R.-P. Wang, K.-J. Zhou, and F. de Groot, *Nat. Commun.* **14**, 2749 (2023)
- [20] M. Magnaterra, K. Hopfer, Ch.J. Sahle, M. Moretti Sala, G. Monaco, J. Attig, C. Hickey, I.-M. Pietsch, F. Breitner, P. Gegenwart, M.H. Upton, Jungho Kim, S. Trebst, P.H.M. van Loosdrecht, J. van den Brink, and M. Grüninger, arXiv:2301.08340 (2023)
- [21] N.B. Brookes, F. Yakhou-Harris, K. Kummer, A. Fondacaro, J.C. Cezar, D. Betto, E. Velez-Fort, A. Amorese, G. Ghiringhelli, L. Braicovich, R. Barrett, G. Berruyer, F. Cianciosi, L. Eybert, P. Marion, P. van der Linden, and L. Zhang, *Nucl. Instrum. Meth. A* **903**, 175 (2018)
- [22] M. Moretti Sala, C. Henriquet, L. Simonelli, R. Verbeni, and G. Monaco, *J. Electron Spectrosc. Relat. Phenom.* **188**, 150 (2013)
- [23] M. Moretti Sala, K. Martel, C. Henriquet, A. Al Zein, L. Simonelli, C.J. Sahle, H. Gonzalez, M.-C. Lagier, C. Ponchut, S. Huotari, R. Verbeni, M. Krisch, and G. Monaco, *J. Synchrotron Rad.* **25**, 580 (2018)
- [24] H. Gretarsson, D. Ketenoglu, M. Harder, S. Mayer, F.-U. Dill, M. Spiwek, H. Schulte-Schrepping, M. Tischer, H.-C. Wille, B. Keimer, and H. Yavaş, *J. Synchrotron Rad.* **27**, 538 (2020)
- [25] L.J.P. Ament, G. Khaliullin, and J. van den Brink, *Phys. Rev. B* **84**, 020403(R) (2011)
- [26] K. Ishii, I. Jarrige, M. Yoshida, K. Ikeuchi, J. Mizuki, K. Ohashi, T. Takayama, J. Matsuno, and H. Takagi, *Phys. Rev. B* **83**, 115121 (2011)
- [27] J. Kim, D. Casa, A. Said, R. Krakora, B.J. Kim, E. Kasman, X. Huang, and T. Gog, *Sci. Rep.* **8**, 1958 (2018)

- [28] D.I. Khomskii and S.V. Streltsov, Chem. Rev. **121**, 2992 (2021)
- [29] E. Benckiser, L. Fels, G. Ghiringhelli, M. Moretti Sala, T. Schmitt, J. Schlappa, V.N. Strocov, N. Mufti, G.R. Blake, A.A. Nugroho, T.T.M. Palstra, M.W. Haverkort, K. Wohlfeld, and M. Grüninger, Phys. Rev. B **88**, 205115 (2013)
- [30] W. Schülke: *Electron Dynamics by Inelastic X-Ray Scattering* (Oxford University Press, New York, 2007)
- [31] M. Krisch and F. Sette: *Inelastic X-Ray Scattering from Phonons*, in M. Cardona, R. Merlin (Eds.): *Light Scattering in Solids IX. Topics in Applied Physics 108* (Springer, Berlin, Heidelberg, 2006)
- [32] M. Mohr, J. Maultzsch, E. Dobardžić, S. Reich, I. Milošević, M. Damnjanović, A. Bosak, M. Krisch, and C. Thomsen, Phys. Rev. B **76**, 035439 (2007)
- [33] L.H. Tjeng: *Imaging Orbitals with X-rays*, in E. Pavarini and E. Koch (Eds.): *Orbital Physics in Correlated Matter, Modeling and Simulation*, Vol. 13 (FZ Jülich, 2023) <http://www.cond-mat.de/events/correl23>
- [34] Y. Shvyd'ko, S. Stoupin, D. Shu, S.P. Collins, K. Mundboth, J. Sutter, and M. Tolkiehn, Nat. Commun. **5**, 4219 (2014)
- [35] M.W. Haverkort, Phys. Rev. Lett. **105**, 167404 (2010)
- [36] P. Warzanowski, M. Magnaterra, Ch.J. Sahle, M. Moretti Sala, P. Becker, L. Bohatý, I. Císařová, G. Monaco, T. Lorenz, P.H.M. van Loosdrecht, J. van den Brink, and M. Grüninger, Phys. Rev. B **110**, 195120 (2024)
- [37] S.G. Chiuzbăian, G. Ghiringhelli, C. Dallera, M. Grioni, P. Amann, X. Wang, L. Braicovich, and L. Patthey, Phys. Rev. Lett. **95**, 197402 (2005)
- [38] G. Ghiringhelli, M. Matsubara, C. Dallera, F. Fracassi, R. Gusmeroli, A. Piazzalunga, A. Tagliaferri, N.B. Brookes, A. Kotani, and L. Braicovich, J. Phys.: Condens. Matter **17**, 5397 (2005)
- [39] S. Huotari, T. Pylkkänen, G. Vankó, R. Verbeni, P. Glatzel, and G. Monaco, Phys. Rev. B **78**, 041102(R) (2008)
- [40] L.-C. Duda, T. Schmitt, M. Magnuson, J. Forsberg, A. Olsson, J. Nordgren, K. Okada, and A. Kotani, Phys. Rev. Lett. **96**, 067402 (2006)
- [41] C.-Y. Liu, K. Ruotsalainen, K. Bauer, R. Decker, A. Pietzsch, and A. Föhlisch, Phys. Rev. B **106**, 035104 (2022)
- [42] G. Ghiringhelli, A. Piazzalunga, C. Dallera, T. Schmitt, V.N. Strocov, J. Schlappa, L. Patthey, X. Wang, H. Berger, and M. Grioni, Phys. Rev. Lett. **102**, 027401 (2009)

- [43] D. Betto, Y.Y. Peng, S.B. Porter, G. Berti, A. Calloni, G. Ghiringhelli, and N.B. Brookes, *Phys. Rev. B* **96**, 020409(R) (2017)
- [44] A. Nag, H.C. Robarts, F. Wenzel, J. Li, H. Elnaggar, R.-P. Wang, A.C. Walters, M. García-Fernández, F.M.F. de Groot, M.W. Haverkort, and K.-J. Zhou, *Phys. Rev. Lett.* **124**, 067202 (2020)
- [45] P. Warzanowski, M. Magnaterra, P. Stein, G. Schlicht, Q. Faure, Ch.J. Sahle, T. Lorenz, P. Becker, L. Bohatý, M. Moretti Sala, G. Monaco, P.H.M. van Loosdrecht, and M. Grüninger, *Phys. Rev. B* **108**, 125120 (2023)
- [46] S. Sugano, Y. Tanabe, and H. Kamimura: *Multiplets of Transition-Metal Ions in Crystals* (Academic Press, New York, 1970)
- [47] A. Georges, L. de' Medici, and J. Mravlje, *Ann. Rev. Condens. Matter Phys.* **4**, 137 (2013)
- [48] G. Zhang and E. Pavarini, *Phys. Rev. B* **95**, 075145 (2017)
- [49] P. Warzanowski, M. Magnaterra, G. Schlicht, Q. Faure, Ch.J. Sahle, P. Becker, L. Bohatý, M. Moretti Sala, G. Monaco, M. Hermanns, P.H.M. van Loosdrecht, and M. Grüninger, *Phys. Rev. B* **109**, 155149 (2024)
- [50] W. Witczak-Krempa, G. Chen, Y.B. Kim, and L. Balents, *Annu. Rev. Condens. Matter Phys.* **5**, 57 (2014)
- [51] J.G. Rau, E.K.-H. Lee, and H.-Y. Kee, *Annu. Rev. Condens. Matter Phys.* **7**, 195 (2016)
- [52] R. Schaffer, E.K.-H. Lee, B.-J. Yang, and Y.B. Kim, *Rep. Prog. Phys.* **79**, 094504 (2016)
- [53] S.V. Streltsov and D.I. Khomskii, *Phys. Rev. X* **10**, 031043 (2020)
- [54] T. Takayama, J. Chaloupka, A. Smerald, G. Khaliullin, and H. Takagi, *J. Phys. Soc. Jpn.* **90**, 062001 (2021)
- [55] L.V. Pourovskii, D. Fiore Mosca, L. Celiberti, S. Khmelevskiy, A. Paramekanti, and C. Franchini, *Nat. Rev. Mater.* **10**, 674 (2025)
- [56] G. Khaliullin, *Phys. Rev. Lett.* **111**, 197201 (2013)
- [57] G. Jackeli and G. Khaliullin, *Phys. Rev. Lett.* **102**, 017205 (2009)
- [58] S.H. Chun, J.-W. Kim, J. Kim, H. Zheng, C.C. Stoumpos, C.D. Malliakas, J.F. Mitchell, K. Mehlawat, Y. Singh, Y. Choi, T. Gog, A. Al-Zein, M. Moretti Sala, M. Krisch, J. Chaloupka, G. Jackeli, G. Khaliullin, and B.J. Kim, *Nat. Phys.* **11**, 462 (2015)
- [59] Q. Wang, A. de la Torre, J.A. Rodriguez-Rivera, A.A. Podlesnyak, W. Tian, A. A. Aczel, M. Matsuda, P.J. Ryan, J.-W. Kim, J.G. Rau, and K.W. Plumb, *Phys. Rev. X* **15**, 021021 (2025)

- [60] N. Iwahara and W. Furukawa, Phys. Rev. B **108**, 075136 (2023)
- [61] J.E. Fergusson and P.F. Heveldt, Aust. J. Chem. **27**, 661 (1974)
- [62] R.L. Armstrong, D. Mintz, B.M. Powell, and W.J.L. Buyers, Phys. Rev. B **17**, 1260 (1978)
- [63] M. Saura-Múzquiz, M. Avdeev, H.E.A. Brand, and B.J. Kennedy, Inorg. Chem. **61**, 15961 (2022)
- [64] D.I. Khomskii: *Transition metal compounds* (Cambridge University Press, 2014)
- [65] A. Jain, M. Krautloher, J. Porras, G. H. Ryu, D. P. Chen, D. L. Abernathy, J. T. Park, A. Ivanov, J. Chaloupka, G. Khaliullin, B. Keimer, and B.J. Kim, Nat. Phys. **13**, 633 (2017)
- [66] S. Kunkemöller, D. Khomskii, P. Steffens, A. Piovano, A.A. Nugroho, and M. Braden, Phys. Rev. Lett. **115**, 247201 (2015)
- [67] H. Gretarsson, H. Suzuki, Hoon Kim, K. Ueda, M. Krautloher, B.J. Kim, H. Yavaş, G. Khaliullin, and B. Keimer, Phys. Rev. B **100**, 045123 (2019)
- [68] G. Zhang and E. Pavarini, Phys. Rev. B **101**, 205128 (2020)
- [69] T. Feldmaier, P. Strobel, M. Schmid, P. Hansmann, and M. Daghofer, Phys. Rev. Res. **2**, 033201 (2020)
- [70] I. Vergara, M. Magnaterra, P. Warzanowski, J. Attig, S. Kunkemöller, D.I. Khomskii, M. Braden, M. Hermanns, and M. Grüninger, Phys. Rev. B **106**, 085103 (2022)
- [71] B.J. Kim and G. Khaliullin, Phys. Rev. B **96**, 085108 (2017)
- [72] N. Iwahara and S. Shikano, Phys. Rev. Res. **5**, 023051 (2023)
- [73] B. Yuan, J.P. Clancy, A.M. Cook, C.M. Thompson, J. Greedan, G. Cao, B.C. Jeon, T.W. Noh, M.H. Upton, D. Casa, T. Gog, A. Paramekanti, and Y.-J. Kim, Phys. Rev. B **95**, 235114 (2017)
- [74] M. Kusch, V.M. Katukuri, N.A. Bogdanov, B. Büchner, T. Dey, D.V. Efremov, J.E. Hamann-Borrero, B.H. Kim, M. Krisch, A. Maljuk, M. Moretti Sala, S. Wurmehl, G. Aslan-Cansever, M. Sturza, L. Hozoi, J. van den Brink, and J. Geck, Phys. Rev. B **97**, 064421 (2018)
- [75] A. Nag, S. Bhowal, A. Chakraborty, M.M. Sala, A. Efimenko, F. Bert, P.K. Biswas, A.D. Hillier, M. Itoh, S.D. Kaushik, V. Siruguri, C. Meneghini, I. Dasgupta, and S. Ray, Phys. Rev. B **98**, 014431 (2018)

- [76] A.A. Aczel, Q. Chen, J.P. Clancy, C. dela Cruz, D. Reig-i-Plessis, G.J. MacDougall, C.J. Pollock, M.H. Upton, T.J. Williams, N. LaManna, J.P. Carlo, J. Beare, G.M. Luke, and H.D. Zhou, *Phys. Rev. Mater.* **6**, 094409 (2022)
- [77] S. Calder, D.J. Singh, V.O. Garlea, M.D. Lumsden, Y.G. Shi, K. Yamaura, and A.D. Christianson, *Phys. Rev. B* **96**, 184426 (2017)
- [78] E. Kermarrec, C.A. Marjerrison, C.M. Thompson, D.D. Maharaj, K. Levin, S. Kroecker, G.E. Granroth, R. Flacau, Z. Yamani, J.E. Greedan, and B.D. Gaulin, *Phys. Rev. B* **91**, 075133 (2015)
- [79] S. Calder, J.G. Vale, N.A. Bogdanov, X. Liu, C. Donnerer, M.H. Upton, D. Casa, A.H. Said, M.D. Lumsden, Z. Zhao, J.-Q. Yan, D. Mandrus, S. Nishimoto, J. van den Brink, J.P. Hill, D.F. McMorrow, and A.D. Christianson, *Nat. Commun.* **7**, 11651 (2016)
- [80] A.E. Taylor, R. Morrow, R.S. Fishman, S. Calder, A.I. Kolesnikov, M.D. Lumsden, P.M. Woodward, and A.D. Christianson, *Phys. Rev. B* **93**, 220408(R) (2016)
- [81] A.E. Taylor, R. Morrow, M.D. Lumsden, S. Calder, M.H. Upton, A.I. Kolesnikov, M.B. Stone, R.S. Fishman, A. Paramakanti, P.M. Woodward, and A.D. Christianson, *Phys. Rev. B* **98**, 214422 (2018)
- [82] A.E. Taylor, S. Calder, R. Morrow, H.L. Feng, M.H. Upton, M.D. Lumsden, K. Yamaura, P.M. Woodward, and A.D. Christianson, *Phys. Rev. Lett.* **118**, 207202 (2017)
- [83] B. Henderson and G.F. Imbusch: *Optical spectroscopy of inorganic solids* (Oxford, 1989)
- [84] B.N. Figgis and M.A. Hitchman: *Ligand Field Theory and its Applications* (Wiley, 1999)
- [85] P.E. Hoggard, *Z. Naturforsch.* **36a**, 1276 (1981)
- [86] H. Kamimura, S. Koide, H. Sekiyama, and S. Sugano, *J. Phys. Soc. Jpn.* **15**, 1264 (1960)
- [87] A. Bertin, T. Dey, D. Brüning, D. Gorkov, K. Jenni, A. Krause, P. Becker, L. Bohatý, D. Khomskii, V. Pomjakushin, L. Keller, M. Braden, and T. Lorenz, *Phys. Rev. B* **109**, 094409 (2024)
- [88] M.W. Haverkort, M. Zwierzcki, and O.K. Andersen, *Phys. Rev. B* **85**, 165113 (2012)
- [89] M.W. Haverkort, *J. Phys.: Conf. Ser.* **712**, 012001 (2016)
- [90] A. Revelli, M. Moretti Sala, G. Monaco, P. Becker, L. Bohatý, M. Hermanns, T.C. Koethe, T. Fröhlich, P. Warzanowski, T. Lorenz, S. V. Streltsov, P.H.M. van Loosdrecht, D.I. Khomskii, J. van den Brink, and M. Grüninger, *Sci. Adv.* **5**, eaav4020 (2019)

- [91] A. Revelli, M. Moretti Sala, G. Monaco, M. Magnaterra, J. Attig, L. Peterlini, T. Dey, A.A. Tsirlin, P. Gegenwart, T. Fröhlich, M. Braden, C. Grams, J. Hemberger, P. Becker, P.H.M. van Loosdrecht, D.I. Khomskii, J. van den Brink, M. Hermanns, and M. Grüninger, *Phys. Rev. B* **106**, 155107 (2022)
- [92] M. Magnaterra, J. Attig, L. Peterlini, M. Hermanns, M.H. Upton, J. Kim, L. Prodan, V. Tsurkan, I. Kézsmárki, P.H.M. van Loosdrecht, and M. Grüninger, *Phys. Rev. Lett.* **133**, 046501 (2024)
- [93] M. Magnaterra, A. Sandberg, H. Schilling, P. Warzanowski, L. Pätzold, E. Bergamasco, Ch.J. Sahle, B. Detlefs, K. Ruotsalainen, M. Moretti Sala, G. Monaco, P. Becker, Q. Faure, G.S. Thakur, M. Songvilay, C. Felser, P.H.M. van Loosdrecht, J. van den Brink, M. Hermanns, and M. Grüninger, *Phys. Rev. B* **111**, 085122 (2025)
- [94] L. Pätzold, A. Sandberg, H. Schilling, H. Gretarsson, E. Bergamasco, M. Magnaterra, P. Becker, P.H.M. van Loosdrecht, J. van den Brink, M. Hermanns, and M. Grüninger, *npj Quantum Mater.* **11**, 48 (2026)
- [95] M. Magnaterra, M. Moretti Sala, G. Monaco, P. Becker, M. Hermanns, P. Warzanowski, T. Lorenz, D.I. Khomskii, P.H.M. van Loosdrecht, J. van den Brink, and M. Grüninger, *Phys. Rev. Res.* **5**, 01316 (2023)
- [96] G.V. Chen and C. Wu, *npj Quantum Mater.* **9**, 1 (2024)
- [97] V. Jayakumar and C. Hickey, *Phys. Rev. B* **113**, 035151 (2026)
- [98] Y. Li, A.A. Tsirlin, T. Dey, P. Gegenwart, R. Valentí, and S.M. Winter, *Phys. Rev. B* **102**, 235142 (2020)
- [99] Y. Ma and M. Blume, *Rev. Sci. Instrum.* **66**, 1543 (1995)
- [100] F. Gel'mukhanov and H. Ågren, *Phys. Rev. A* **49**, 4378 (1994)
- [101] V.M. Katukuri, X. Lu, D.E. McNally, M. Dantz, V.N. Strocov, M. Moretti Sala, M.H. Upton, J. Terzic, G. Cao, O.V. Yazyev, and T. Schmitt, *Phys. Rev. B* **105**, 075114 (2022)
- [102] J. Söderström, A. Ghosh, L. Kjellsson, V. Ekholm, T. Tokushima, C. Sâthe, N. Velasquez, M. Simon, O. Björneholm, L. Duda, A.N. de Brito, M. Odelius, J.-C. Liu, J. Wang, V. Kimberg, M. Agåker, J.-E. Rubensson, and F. Gel'mukhanov, *Sci. Adv.* **10**, eadk3114 (2024)
- [103] V.K. Bhartiya, T. Kim, J. Li, T.P. Darlington, D.J. Rizzo, Y. Gu, S. Fan, C. Nelson, J.W. Freeland, X. Xu, D.N. Basov, J. Pelliciari, A.F. May, C. Mazzoli, and V. Bisogni, *npj Quantum Mater.* **10**, 80 (2025)

- [104] T. Ren, Y. Shen, S.F.R. TenHuisen, J. Sears, W. He, M.H. Upton, D. Casa, P. Becker, M. Mitrano, M.P.M. Dean, and R.M. Konik, arXiv:2404.05850 (2024)
- [105] J. Kwon, J. Kim, G. Oh, S. Jin, K. Kim, H. Kim, S. Ha, H.-W.J. Kim, G. Sim, B. Wehinger, G. Garbarino, N. Maraytta, M. Merz, M. Le Tacon, C.J. Sahle, A. Longo, J. Kim, A. Go, G.Y. Cho, B.H. Kim, and B.J. Kim, *Nat. Mater.* **25**, 610 (2026)
- [106] M. Moretti Sala, V. Schnells, S. Boseggia, L. Simonelli, A. Al-Zein, J.G. Vale, L. Paolasini, E.C. Hunter, R.S. Perry, D. Prabhakaran, A.T. Boothroyd, M. Krisch, G. Monaco, H.M. Rønnow, D.F. McMorro, and F. Mila, *Phys. Rev. B* **92**, 024405 (2015)
- [107] Z. Porter, P.M. Sarte, T. Petersen, M.H. Upton, L. Hozoi, and S.D. Wilson, *Phys. Rev. B* **106**, 115140 (2022)
- [108] L.T. Nguyen and R.J. Cava, *Chem. Rev.* **121**, 2935 (2021)
- [109] T. Dey, M. Majumder, J.C. Orain, A. Senyshyn, M. Prinz-Zwick, S. Bachus, Y. Tokiwa, F. Bert, P. Khuntia, N. Büttgen, A.A. Tsirlin, and P. Gegenwart, *Phys. Rev. B* **96**, 174411 (2017)
- [110] T. Sakamoto, Y. Doi, and Y. Hinatsu, *J. Solid State Chem.* **179**, 2595 (2006)
- [111] T. Dey, A.V. Mahajan, P. Khuntia, M. Baenitz, B. Koteswararao, and F.C. Chou, *Phys. Rev. B* **86**, 140405(R) (2012)
- [112] R. Kumar, D. Sheptyakov, P. Khuntia, K. Rolfs, P.G. Freeman, H.M. Rønnow, T. Dey, M. Baenitz, and A.V. Mahajan, *Phys. Rev. B* **94**, 174410 (2016)
- [113] W.-J. Lee, S.-H. Do, Sungwon Yoon, S. Lee, Y.S. Choi, D.J. Jang, M. Brando, M. Lee, E.S. Choi, S. Ji, Z.H. Jang, B.J. Suh, and K.-Y. Choi, *Phys. Rev. B* **96**, 014432 (2017)
- [114] H. Zhao, Y. Zhang, P. Schlottmann, R. Nandkishore, L.E. DeLong, and G. Cao, *Phys. Rev. Lett.* **132**, 226503 (2024)
- [115] L.T. Nguyen and R.J. Cava, *Phys. Rev. Mater.* **3**, 014412 (2019)
- [116] G.S. Thakur, S. Chattopadhyay, T. Doert, T. Herrmannsdörfer, and C. Felser, *Cryst. Growth Des.* **20**, 2871 (2020)
- [117] A. Bandyopadhyay, S. Lee, D.T. Adroja, M.R. Lees, G.B.G. Stenning, P. Aich, L. Tortora, C. Meneghini, G. Cibir, A. Berlie, R.A. Saha, D. Takegami, A. Meléndez-Sans, G. Poelchen, M. Yoshimura, K.D. Tsuei, Z. Hu, T.-S. Chan, S. Chattopadhyay, G.S. Thakur, and K.-Y. Choi, *Phys. Rev. Mater.* **8**, 074405 (2024)
- [118] E.V. Komleva, D.I. Khomskii, and S.V. Streltsov, *Phys. Rev. B* **102**, 174448 (2020)
- [119] G. Cao, H. Zheng, H. Zhao, Y. Ni, C.A. Pocs, Y. Zhang, F. Ye, C. Hoffmann, X. Wang, M. Lee, M. Hermele, and I. Kimchi, *npj Quantum Mater.* **5**, 26 (2020)

- [120] Y. Shen, J. Sears, G. Fabbri, A. Weichselbaum, W. Yin, H. Zhao, D.G. Mazzone, H. Miao, M.H. Upton, D. Casa, R. Acevedo-Esteves, C. Nelson, A.M. Barbour, C. Mazzoli, G. Cao, and M.P.M. Dean, *Phys. Rev. Lett.* **129**, 207201 (2022)
- [121] M. Ye, H.-S. Kim, J.-W. Kim, C.-J. Won, K. Haule, D. Vanderbilt, S.-W. Cheong, and G. Blumberg, *Phys. Rev. B* **98**, 201105(R) (2018)
- [122] Y. Doi, M. Wakeshima, Y. Hinatsu, A. Tobo, K. Ohoyama, and Y. Yamaguchi, *J. Mater. Chem.* **11**, 3135 (2001)
- [123] Y. Doi, K. Matsuhira, and Y. Hinatsu, *J. Solid State Chem.* **165**, 317 (2002)
- [124] Y. Hinatsu and Y. Doi, *Bull. Chem. Soc. Jpn.* **76**, 1093 (2003)
- [125] B. Yuan, B.H. Kim, Q. Chen, D. Dobrowolski, M. Azmańska, G.M. Luke, S. Fan, V. Bisogni, J. Pellicciari, and J.P. Clancy, *Phys. Rev. Lett.* **134**, 106504 (2025)

12 Light Control of Many-Body Phases

Matteo Mitrano
Harvard University, Department of Physics
17 Oxford Street, 02138 Cambridge (MA)

Contents

1	Introduction	2
2	Experimental techniques	2
3	Photoinduced insulator-to-metal transitions	4
3.1	Mott-Hubbard and charge-transfer insulators	5
3.2	Incoherent excitation processes and photodoping	6
3.3	Coherent control of the lattice and nonlinear phononics	8
4	Light control of ordered states	10
4.1	Charge-ordered phases	11
4.2	Magnetism	14
4.3	Ferroic orders	16
5	Floquet engineering	18
5.1	Control of band structure and functional responses	20
5.2	Towards quantum control of correlated states	22
5.3	Electronic metastability via coherent dressing	24
6	Light-induced superconductivity	27
6.1	Experimental reports of light-induced superconductivity	27
6.2	Theoretical scenarios	33
7	Outlook	35

1 Introduction

This chapter is devoted to the use of the light-matter interaction to coherently engineer many-body states in quantum materials. Ultrashort optical fields have enabled advances in the manipulation of quantum states in solids and in the discovery of emergent phenomena out of equilibrium [1–5]. Light now stands alongside doping, gating, strain, and pressure as a tuning knob for material properties, and is increasingly used as a tool to synthesize novel states of matter.

The field of *photoinduced phase transitions* has been shaped by the recognition that light can do more than heat a material or incoherently excite quasiparticles. In systems governed by a subtle balance of competing electronic interactions, optical excitation can trigger colossal cooperative responses. Early demonstrations established the central idea that prompt photoexcitations can nucleate macroscopic changes of electronic and structural order [6]. Driven by advances in tunable laser sources, the field has moved from the study of melting, thermalization, and structural responses [7] toward the implementation of coherent control protocols in which tailored optical fields are used to modify Hamiltonians, coherently manipulate ordered states, and access nonequilibrium states without direct equilibrium analogs.

In the following, we focus primarily on experiments aimed at *controlling* material properties. There is a vast literature in which nonequilibrium techniques are used to perturb solids and study near-equilibrium properties. Our primary interest is instead in emergent phenomena and functionalities induced by light in strongly interacting electron systems [5, 8], although we will occasionally discuss systems better described as weakly interacting electron ensembles. Experiments involving optical dressing in cavities will be discussed only in the outlook, as this exciting and rapidly developing direction lies beyond the scope of this chapter.

To keep the chapter focused, we discuss only selected key experimental achievements, together with high-level aspects of the theory, and refer the reader to the literature where appropriate.

2 Experimental techniques

Progress in the light control of quantum materials is closely tied to the development of new laser sources and experimental probes. Most light-control experiments rely on a pump-probe protocol: an intense optical field, the *pump*, drives a material out of equilibrium and toward a desired quantum state, while a delayed and typically weaker pulse, the *probe*, interrogates the properties of the transient state. The measurement is repeated stroboscopically when the system relaxes back to equilibrium between successive pump pulses, or performed in single-shot mode when materials undergo irreversible transitions.

Early experiments relied on a comparatively limited set of optical excitation conditions. The development of ultrafast light sources that reliably generate intense pulses from the terahertz to the extreme ultraviolet, with tunable pulse duration, polarization, bandwidth, and repetition rate [9–13], has opened a much more selective approach to driving quantum materials. Optical fields can now be matched to specific electronic transitions, coupled directly to collective modes, or tuned into transparency regions to minimize incoherent excitation processes. It is this selectivity that turns light from an external perturbation into a control field.

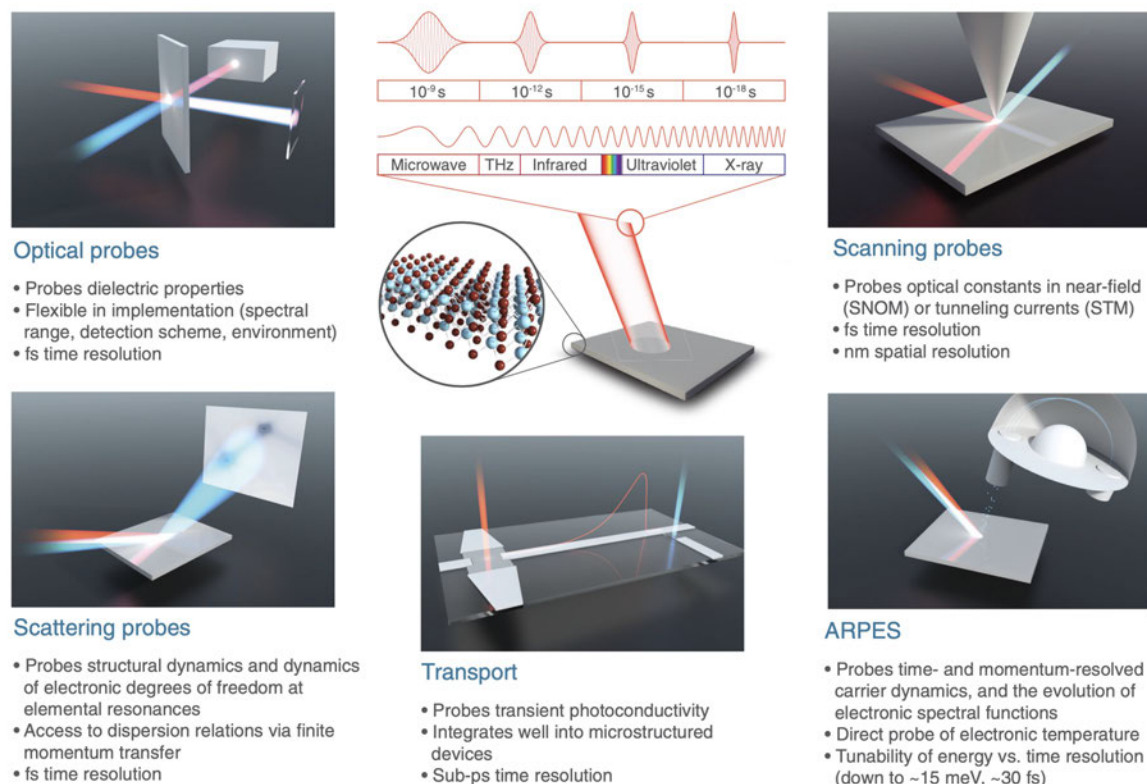


Fig. 1: Overview of experimental techniques used to study optically driven quantum materials. The upper central panel shows the spectral regions and timescales of ultrafast pulses commonly used to drive materials out of equilibrium. The surrounding panels show time-resolved probes with complementary spatial, energy, and momentum resolution. Reprinted figure with permission from Ref. [4]. Copyright (2022) by the American Physical Society.

These developments have occurred in tandem with a gradual expansion of the experimental probe toolkit. This section introduces several experimental techniques that will facilitate the discussion in the rest of the chapter (Fig. 1). The first, and most widely used, is *time-resolved optical spectroscopy*. Since part of the laser output is used to generate the pump pulse, another fraction can often be used to interrogate the time-dependent optical properties of the material [14]. This broad family of experiments encompasses many techniques and observables, ranging from transient absorption and reflectivity measurements to track the redistribution of spectral weight that accompanies a change of electronic state [15–19] to polarization-rotation experiments measuring the magnetization in magnetic systems [20]. Extending time-resolved optical experiments from the visible into the terahertz (THz) range gives the most direct access to the low-energy electrodynamics of correlated systems, charge-density-wave states, and superconductors [21], where the relevant gaps and collective modes reside. Second- and third-harmonic polarimetry, in turn, is sensitive to the local symmetry of the lattice [22, 23] and has been used to probe ordered states in and out of equilibrium [24–26].

A second major experimental tool is time- and angle-resolved photoemission spectroscopy (trARPES) [27]. In this technique, ultrafast XUV photons (generated either by cascaded up-

conversion of the fundamental laser wavelength or by high-harmonic generation) impinge on a material and trigger electron emission through the photoelectric effect. Detection of the emitted electrons provides momentum- and energy-resolved access to the single-particle spectral function. This technique is essential for tracking band structure changes in optically-driven materials and detecting nonthermal regimes far from equilibrium [28]. It has been used to study decoherence [29], the coupling of electrons to other degrees of freedom [30], and the dynamics of gaps and quasiparticles in high-temperature superconductors [31] and charge-density-wave states [32, 33]. Remarkably, trARPES has also provided some of the clearest visualizations of Floquet-Bloch states to date [34–37]. Finally, momentum microscopes are enabling photoemission studies of two-dimensional materials and devices over multiple Brillouin zones.

A third major class of experimental methods are time-resolved x-ray and electron scattering. The development of femtosecond x-ray sources, particularly x-ray free-electron lasers (XFELs) [38], together with ultrafast electron sources operating from the keV to MeV range [39], has enabled scattering experiments in which the probe wavelength is comparable to interatomic distances. Time-resolved x-ray and electron diffraction provide direct access to structural dynamics [40, 41] and are especially sensitive to phase transitions in which electronic degrees of freedom are coupled to lattice distortions [42–47]. In parallel, resonant x-ray techniques have extended this approach beyond structural probes. By tuning the photon energy to elemental absorption edges, time-resolved x-ray absorption (trXAS) and resonant inelastic x-ray scattering (trRIXS) can selectively track charge, spin, orbital, and lattice degrees of freedom in optically driven materials [48, 49]. With the introduction of high-resolution spectrometers at XFELs, trRIXS is becoming an increasingly effective discovery tool for novel states of matter.

Finally, the increasing relevance of two-dimensional materials, heterostructures, and devices has motivated the development of local and device-integrated time-resolved probes. Time-resolved scanning tunneling microscopy (STM) accesses ultrafast tunneling dynamics with atomic-scale spatial sensitivity [50, 51], while time-resolved scanning near-field optical microscopy (trSNOM) probes transient optical constants with nanometer resolution [52]. These techniques are especially valuable when photoinduced states are spatially inhomogeneous, confined to surfaces or interfaces, or embedded in mesoscopic structures. In parallel, ultrafast transport measurements have emerged as a complementary route to nonequilibrium dynamics, providing direct access to transient photoconductivity, photocurrents, and light-induced changes in device response [53]. These local methods are decisive when the photoinduced state is spatially inhomogeneous, and together with the optical, photoemission, and scattering probes above they form the toolkit used throughout this chapter.

3 Photoinduced insulator-to-metal transitions

Photoinduced insulator-to-metal transitions are a natural starting point for understanding nonequilibrium phenomena in correlated solids [54, 55]. The equilibrium state is characterized by a single, well-defined quantity, the charge gap, and decades of equilibrium work provide a firm benchmark for experiments far from equilibrium. We first introduce Mott-Hubbard and charge-

transfer insulators, and then examine the incoherent and coherent pathways by which light turns an insulator into a metal.

Every photoinduced transition shares a common sequence of events: light couples to a fast subsystem, usually the electrons, which in turn drives the slow variables that define a phase, such as a lattice distortion, an order parameter, or a magnetization [6]. What separates the two pathways is whether this drive preserves phase information or merely deposits energy into the target. In an incoherent process the absorbed energy dephases among the fast carriers faster than the slow variables can respond. The excitation then acts as a sudden reshaping of the free-energy landscape, and the material reaches a new state through relaxation, dissipation, and nucleation. A coherent process instead channels energy into a specific collective coordinate, driving it with a well-defined phase and with damping slower than its period, so that the system evolves along a controlled, often oscillatory, trajectory rather than by relaxation. The distinction is properly defined with respect to the slow variables, since an incoherent electronic excitation can itself set a lattice mode ringing coherently. Because they inject little entropy, coherent excitation processes are more selective, energy-efficient and tunable, and they make light a genuine control field rather than a source of heat.

3.1 Mott-Hubbard and charge-transfer insulators

The physics of strongly correlated materials is governed by the competition between electronic delocalization and Coulomb-driven localization [56, 57]. When electron-electron interactions become comparable to the kinetic energy, the motion of each carrier is no longer independent and becomes constrained by the configuration of the other valence electrons. A minimal Hamiltonian capturing this competition is the extended Hubbard model,

$$H = -t \sum_{i\sigma} \left(c_{i\sigma}^\dagger c_{i+1\sigma} + H.c. \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} + V_1 \sum_i n_i n_{i+1} + V_2 \sum_i n_i n_{i+2} + \dots, \quad (1)$$

where $c_{i\sigma}^\dagger$ and $c_{i\sigma}$ create and annihilate an electron with spin σ on lattice site i , $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$, and $n_i = n_{i\uparrow} + n_{i\downarrow}$. The hopping amplitude t sets the kinetic energy scale, while U describes the on-site Coulomb repulsion between two electrons occupying the same site. The terms $V_1, V_2, \dots$ account for longer-range Coulomb interactions between electrons on nearest-neighbor, next-nearest-neighbor, and more distant sites. The corresponding electronic bandwidth is $W = 2zt$, where z is the number of nearest neighbors.

When only the on-site Coulomb interaction is retained, the resulting simple Hubbard model is governed primarily by two parameters, the number of electrons per site (*filling*), and the interaction ratio U/W (*Mottness*). The essential physics of the half-filled Hubbard model is already contained in its two exactly solvable limits. For $U = 0$ the model reduces to a tight-binding band which, being half filled, describes a metal. In the opposite atomic limit $t = 0$ every site is singly occupied, intersite transport is blocked by the Coulomb penalty U , and the ground state is a Mott insulator (see Fig. 2). Tuning U/W between these limits drives a metal-to-insulator transition. The spectrum splits into a lower and an upper Hubbard band separated by a charge gap of order U , which measures the energy cost of creating an empty site (*holon*) and a

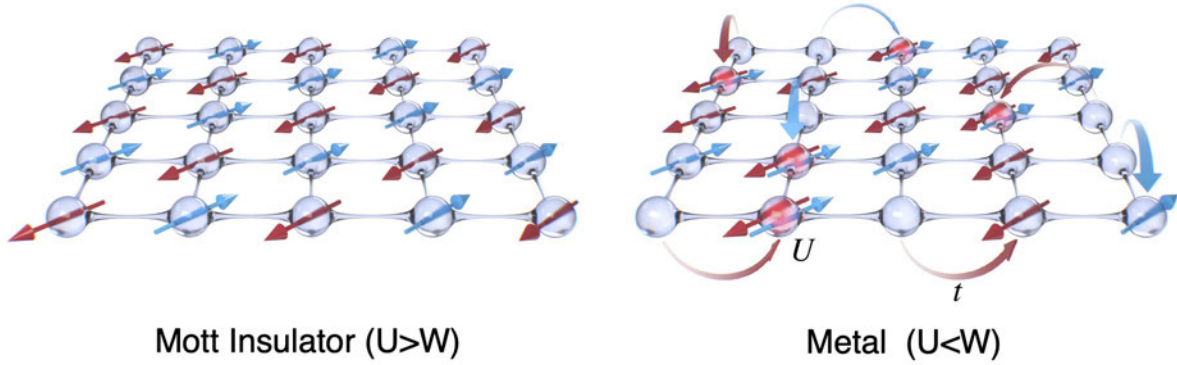


Fig. 2: *Correlated electron dynamics in a square lattice at half filling. When the on-site Coulomb repulsion U is larger than the electron bandwidth W , the electrons localize and form a Mott insulating state. In the opposite limit ($U < W$), they transition into a metallic state.*

doubly occupied one (*doublon*). Spin degrees of freedom introduce an additional energy scale. Virtual hopping onto a neighboring occupied site and back lowers the energy of antiparallel spins, generating an antiferromagnetic superexchange $J = 4t^2/U$. Charge and spin are thus controlled by two widely separated scales, $U \gg J$.

In real materials the gap is not always set by the Hubbard U . Following Zaanen, Sawatzky, and Allen [58], one distinguishes Mott-Hubbard insulators, where the gap separates the lower and upper Hubbard bands ($\sim U$), from charge-transfer insulators, where the lowest charge excitation moves an electron from a filled ligand band onto the correlated site, so that the gap is set by the charge-transfer energy $\Delta < U$. Most transition-metal oxides of interest, including cuprates and nickelates, belong to the charge-transfer class, a distinction that is relevant for determining which excitations are addressed by the optical pump.

3.2 Incoherent excitation processes and photodoping

The simplest incoherent process for inducing an insulator-to-metal transition with light is the injection of mobile charge carriers across the gap. In a half-filled Mott insulator, photons with energies comparable to or larger than the on-site interaction U can delocalize charges between neighboring sites, creating holon-doublon pairs and thereby driving a transient insulator-to-metal transition. Although this mechanism (often referred to as *photodoping*) resembles chemical doping in the introduction of mobile charges, photoinduced metallic states may host multiple types of carriers, nonthermal distributions, and emergent dynamic phenomena.

A compelling example is the one-dimensional Mott insulator (BEDT-TTF)-F₂TCNQ. This molecular solid hosts chains of BEDT-TTF molecules and is characterized by weak electron-lattice coupling and the absence of a Peierls distortion. Okamoto *et al.* investigated its response by pumping along the BEDT-TTF chains with 1.55 eV photons [59], well above the 700 meV Mott gap, and probing the optical response over the 0.08-2.5 eV range (Fig. 3). Immediately after photoexcitation, the reflectivity associated with the Mott gap is depleted, while the low-

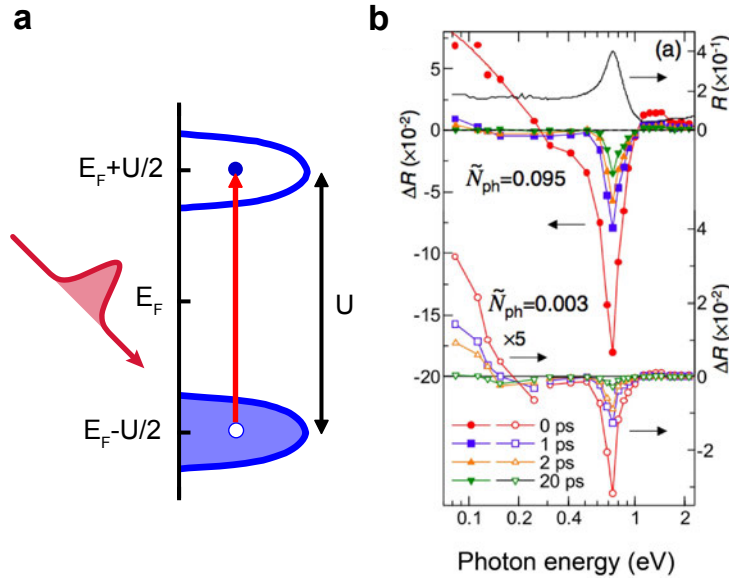


Fig. 3: Photoinduced insulator-to-metal transition in (BEDT-TTF)-F₂TCNQ. **a** Sketch of photodoping of a Mott insulator. The pump excites electrons from the lower Hubbard band at $E_F - U/2$ (E_F , Fermi energy) into the upper Hubbard band at $E_F + U/2$. **b** Transient reflectivity changes $\Delta R = R(t) - R_{eq}$ along the a -axis after photoexcitation with 1.55 eV photons with 0.095 photons/ET molecule and 0.003 photons/ET molecule excitation densities. Adapted with permission from Ref. [59]. Copyright (2007) by the American Physical Society.

energy reflectivity below roughly 200 meV increases, signaling gap melting and transient metallization. The transient reflectivity change $\Delta R(t)$ follows a double-exponential relaxation and recovers within approximately 2 ps. Subsequent experiments showed that holon-doublon pairs exhibit quantum interference between bound and unbound states at early times [60], and that their relaxation times depend sensitively on the degree of electronic correlation [61]. This example illustrates a largely electronic route to transient metallization, in which the insulating state is destabilized by photocarrier injection without evidence for a structural phase transition. Similar organics also support the reverse process. In metallic α -(BEDT-TTF)₂I₃, a 1.5-cycle near-infrared pump pulse (7 fs, 9.3 MV cm⁻¹, 0.6–0.95 eV) has been shown to drive a transient metal-to-insulator transition, tracked by broadband reflectivity resolved in energy and delay [62]. The field lowers the hopping t by roughly 10% through dynamical localization, opening a gap associated with charge order. Photoinduced metallization has also been observed in other charge-transfer insulators, including two-dimensional copper oxides [63, 64].

While the discussion so far has emphasized electronic dynamics, lattice degrees of freedom can play several important roles as relaxation channels for holon-doublon pairs, as drivers of polaronic resonances, or as cooperative variables in the photoinduced response. In some materials, electronic and structural degrees of freedom are so strongly intertwined that the photoinduced insulator-to-metal transition is accompanied by a proper structural phase transition. A paradigmatic case is that of vanadium dioxide, VO₂.

At equilibrium, VO₂ exhibits a sharp insulator-to-metal transition accompanied by a structural change from a low-temperature monoclinic phase (M1), characterized by dimerized V–V

chains, to a high-temperature rutile phase (R), with undimerized V–V bonds [65]. Ultrafast optical pulses can similarly switch VO₂ from an insulating to a metallic state, and this photoinduced transition is accompanied by a monoclinic-to-rutile structural transition. A series of pump-probe experiments using 1.55 eV excitation have examined the relative roles of electronic and lattice degrees of freedom in this process [66,42,67,68,45]. Beyond identifying the primary driver of metallization, an important question is whether the accompanying structural transition is coherent or incoherent.

In a solid-solid transition where the crystal symmetry increases with temperature, as in the M1-to-R transition of VO₂, the structural response can occur either through a displacive or order-disorder mechanism. The VO₂ transition has long been interpreted as displacive, corresponding to a soft phonon mode of the high-symmetry metallic phase, based on symmetry arguments, diffraction measurements, and lattice-dynamics simulations. In this picture, the atoms collectively rearrange their positions along one or a few coherent vibrational coordinates that connect the low- and high-symmetry structures. Recent total x-ray scattering measurements, however, revealed diffuse scattering that is not expected for a purely displacive transition, but is characteristic of an order-disorder process [45]. These results indicate that the photoinduced insulator-to-metal transition in VO₂ is incoherent not only in its electronic dynamics, but also in the accompanying structural response.

3.3 Coherent control of the lattice and nonlinear phononics

We now turn to coherent pathways and discuss examples in which light is used to drive the lattice directly, rather than to inject carriers across the gap. Ultrashort pulses can coherently excite phonons through displacive excitation of modes whose equilibrium position shifts upon photoexcitation, or through impulsive stimulated Raman scattering [69,70]. These mechanisms, however, either rely on the excitation of electronic states or reach only a restricted set of Raman modes, and therefore offer limited tunability.

The development of intense, tunable sources in the mid-infrared and THz range has opened new excitation mechanisms, including the resonant driving of infrared-active vibrations to amplitudes large enough that lattice anharmonicity becomes essential (see Fig. 4a).

This connection is especially important in correlated oxides, where the crystal structure directly shapes electronic behavior. In ABO₃ perovskites, the transition-metal ion is coordinated by an oxygen octahedron, and even small rotations or tilts of these octahedra can reshape the electronic ground state. In manganites, for example, rotations of the MnO₆ octahedra modify the Mn–O–Mn bond angles, controlling the orbital overlap along the bonds and thereby the one-electron bandwidth W as well as the sign and magnitude of the exchange interactions. Hopping is maximal for a straight 180° bond and decreases as the bond bends. A vibration that periodically distorts these bonds therefore provides, in the language of the previous sections, a direct handle on U/W , and hence on the balance between insulating and metallic behavior.

This mechanism underlies the first vibrationally driven insulator-to-metal transition, reported by Rini *et al.* in the magnetoresistive manganite Pr_{0.7}Ca_{0.3}MnO₃ [72]. At low temperature, this

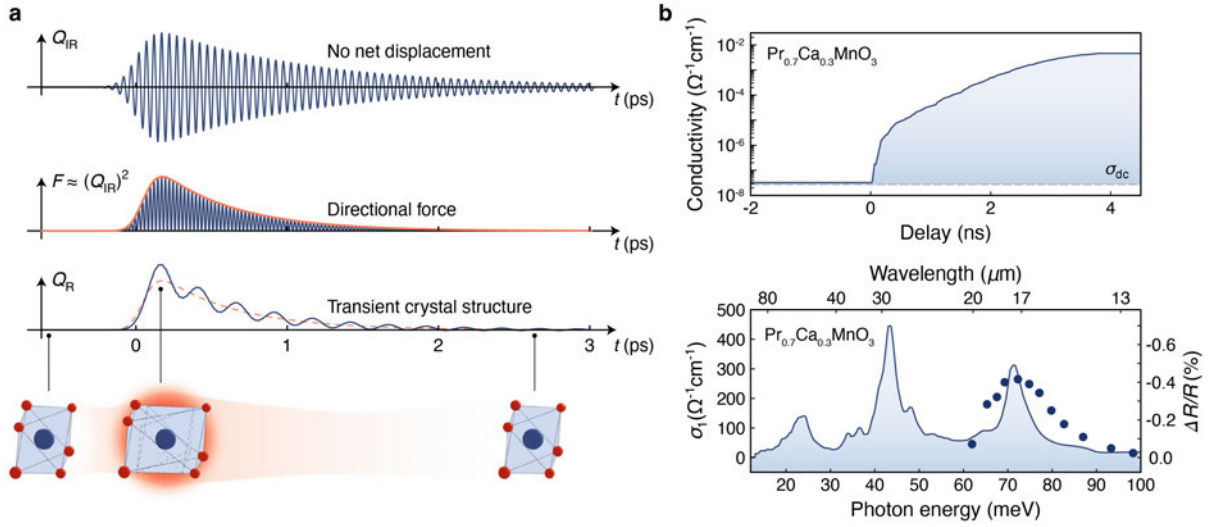


Fig. 4: *Nonlinear phononics and photoinduced insulator-to-metal transitions. (a) Nonlinear phononic response for an anharmonic coupling $V_{anh} \propto Q_{IR}^2 Q_R$, coherently driven infrared phonon amplitude (top), induced lattice force (middle), and resulting structural distortion along the Raman coordinate (bottom), together with a schematic distortion of a TMO_6 octahedron. (b) Transient resistivity increase in resonantly excited $Pr_{0.7}Ca_{0.3}MnO_3$ (top), and comparison of its equilibrium vibrational spectrum and resonant transient reflectivity response (bottom). Panels adapted with permission from Ref. [71]. Copyright (2021) by Springer Nature Limited. Data reproduced with permission from Ref. [72]. Copyright (2007) by Springer Nature Limited.*

compound is a charge- and orbital-ordered insulator. Resonant excitation of an infrared-active Mn–O stretching mode at 71 meV (17 THz) modulates the Mn–O–Mn bond angle, transiently broadening the electronic bandwidth and collapsing the gap. The resistivity drops by five orders of magnitude into a metastable metallic state (Fig. 4b). Unlike photodoping, no carriers are promoted across the gap and the lattice becomes the tuning knob of the electronic transition. Driving a lattice mode to large amplitude is a general strategy whose systematic formulation is known as *nonlinear phononics* [73, 74, 71]. The essential physics can be captured by two anharmonically-coupled oscillators. An intense mid-infrared field resonantly drives an infrared-active mode Q_{IR} through its Born effective charge Z^* , while the lowest-order lattice anharmonicity, $V_{anh} = g Q_{IR}^2 Q_R$, couples this motion to a Raman-active mode Q_R

$$\ddot{Q}_{IR} + \gamma_{IR} \dot{Q}_{IR} + (\omega_{IR}^2 + 2g Q_R) Q_{IR} = Z^* E(t), \quad (2)$$

$$\ddot{Q}_R + \gamma_R \dot{Q}_R + \omega_R^2 Q_R = -g Q_{IR}^2. \quad (3)$$

Because the driven infrared mode oscillates about zero, it does not by itself produce a net structural distortion. The quadratic term Q_{IR}^2 , however, generates a rectified force on Q_R , displacing it toward a new quasi-equilibrium position and producing a coherent structural distortion along a selected Raman coordinate. This construction extends well beyond pairs of phonons, and in recent years it has been shown that the driven mode can couple anharmonically to magnetization, exchange and hopping integrals, or macroscopic strain, providing a symmetry-selective route to control both structural and electronic degrees of freedom (Table 1).

Table 1: *Nonlinear phononic processes and their effect on the structural and functional properties of a material, after Ref. [71]. Q_{driven} , resonantly driven mode; Q_c , coupled phonon; M , magnetization; $\mathbf{S}_i$, local spin on site i ; U , on-site Coulomb repulsion; t , hopping integral; ε , macroscopic strain.*

Nonlinearity	Effect
$Q_{\text{driven}}^2 \cdot Q_c$, $Q_{\text{driven},1} \cdot Q_{\text{driven},2} \cdot Q_c$	Quasi-static lattice distortion
$(\mathbf{Q}_{\text{driven}} \times \mathbf{Q}_{\text{driven}}) \cdot \mathbf{M}$, $Q_{\text{driven}}^2 \mathbf{S}_i \cdot \mathbf{S}_j$	Change of magnetism
$Q_{\text{driven}}^2 \cdot U$, $Q_{\text{driven}}^2 \cdot t$	Change of electronic structure
$Q_{\text{driven}}^2 \cdot \varepsilon$	Strain
Q_{driven}^4 , $Q_{\text{driven}}^2 \cdot Q_c^2$, ...	Phonon amplification, phonon squeezing

The reach of this approach extends beyond bulk excitation. A clear demonstration is provided by epitaxial thin films of the antiferromagnetic insulator NdNiO_3 , whose metal-to-insulator transition is governed, as in the manganites, by the Ni–O–Ni bond angle and hence by the bandwidth [75]. In this experiment, a 33 nm film was grown on a LaAlO_3 substrate, and a 150 fs mid-infrared pulse was tuned not to a vibration of the film, but to the $\sim 15 \mu\text{m}$ ($\sim 80 \text{ meV}$) Al–O stretching mode of the *substrate*. Below the transition temperature, the driven substrate mode launches a lattice distortion that propagates across the interface and strains the nickelate, thus launching an insulator-to-metal transition. Spectral weight is transferred from high to low energy, and the quasi-dc conductivity of the film increases by five orders of magnitude within 3 ps. The metallic state is long-lived, with a lifetime exceeding 1 ns, while an estimated heating below 5 K rules out a thermal origin. The decisive evidence that the substrate is the active element is that the photosusceptibility of the transition follows the infrared absorption of LaAlO_3 , rather than that of NdNiO_3 . The heterostructure itself therefore becomes a design parameter, as changes in the substrate can be exploited to shift both resonance and magnitude of the response.

4 Light control of ordered states

In the previous section, we discussed transitions described by a single scalar quantity, the charge gap. We now turn to phenomena involving symmetry breaking and described by an *order parameter*, a quantity that vanishes in the symmetric state and acquires well-defined magnitude and phase once the symmetry is lowered. At equilibrium, lowering the temperature deforms the free energy until the symmetric minimum gives way to equivalent broken-symmetry minima. The system selects one of them, and transitions into a new ordered phase (Fig. 5a). Charge-density waves, antiferromagnets, and ferroelectrics are all manifestations of this physics, differing in the symmetry that is broken and in the microscopic nature of the order parameter.

Light can similarly manipulate ordered states in quantum materials. Intense pulses can drive the nonthermal melting of ordered phases, but can also, more interestingly, induce order far from equilibrium by coherently driving a quantum coordinate of the material (Fig. 5b). The frontier of the field lies in realizing nontrivial ordered states, possibly without equilibrium counterparts, and making them metastable after the excitation has decayed. In the following, we discuss examples of light control of ordered phases, such as charge order, magnetism, and ferroic orders.

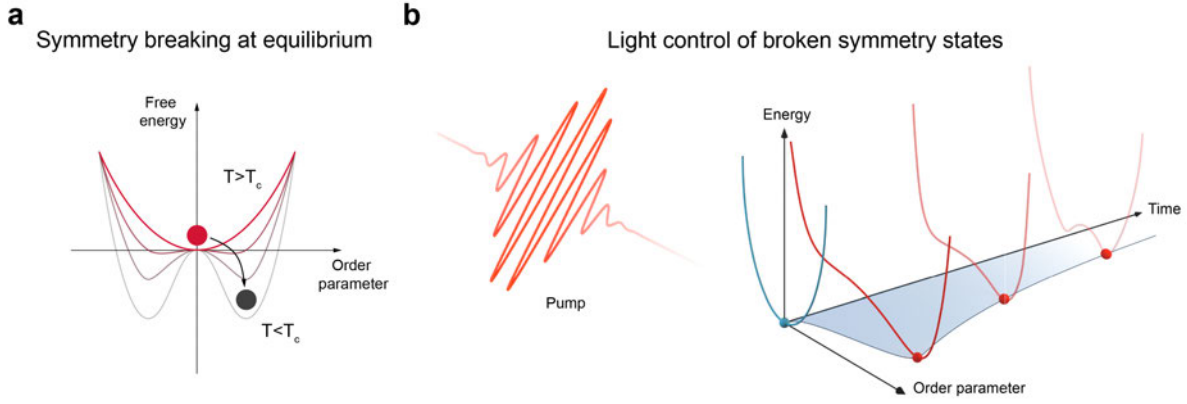


Fig. 5: Symmetry breaking and its control with light. (a) At equilibrium, order emerges through spontaneous symmetry breaking. Lowering the temperature reshapes the free energy, creating equivalent broken-symmetry minima into which the system evolves from the high-symmetry configuration. (b) Pump pulses can reshape the free energy landscape in time and induce more ordered states. The evolving depth and position of the minima track the order parameter dynamics, with the goal of steering the system toward nontrivial, possibly metastable states.

4.1 Charge-ordered phases

Charge-ordered phases form when valence electrons spontaneously self-organize into a spatially modulated state with a well-defined periodicity, often accompanied by a commensurate or incommensurate distortion of the lattice [76]. The order parameter is the complex amplitude $\eta = |\eta|e^{i\phi}$ of the density modulation, and its magnitude measures the strength of the order, while its phase ϕ fixes the position of the wave relative to the lattice. Such order parameter structure generally supports two characteristic collective excitations: amplitude, or amplitudon, fluctuations that modulate $|\eta|$, and phase, or phason, fluctuations of the modulation.

Charge-ordered states can be broadly grouped into two classes. In one, the order is driven by a structural Peierls-type instability, namely a periodic lattice distortion opens a gap at the Fermi surface, electronic and lattice modulations are locked together, and the amplitude mode can be identified, to a good approximation, with a lattice phonon. In the other, the modulation is primarily electronic, driven by Coulomb interactions and the kinetic frustration of charge carriers, with lattice distortions only playing a minor role. This distinction determines whether light excitation of the charge-ordered phase results in coherent lattice motion and structural rearrangements, or dynamics of the sole electronic subsystem. In practice, assigning a charge-ordered phase to one of these limits is often nontrivial [77], and the relative roles of electronic and lattice degrees of freedom remain widely debated. The classification used here should therefore be understood as an organizing principle rather than a sharp boundary.

Because charge order modulates the density on microscopic, often atomic, length scales, it is most directly accessed with momentum-resolved probes. trARPES tracks the opening and closing of the charge-order gap and the reconstruction of the single-particle spectral function, while time-resolved x-ray and electron scattering measure the superlattice reflections that directly fingerprint the modulation, giving access to its amplitude, periodicity, and correlation length.

As a first example, we return to charge- and orbital-ordered manganites, where charge order is intertwined with orbital and magnetic degrees of freedom. In $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$, localization of the Mn 3d electrons produces a coupled charge-orbital superlattice modulation, accompanied by a Jahn-Teller distortion of the MnO_6 octahedra. Beaud *et al.* used femtosecond x-ray diffraction to track the photoinduced melting of this order [43]. Although the structural transition involves many electronic and vibrational coordinates, they found that the photoinduced charge order melting is captured by a single time-dependent order parameter governed by the electronic excitation. Owing to the coupling to the lattice, photoexcitation also induces a strong coherent oscillation at ~ 2.45 THz around the equilibrium positions of the higher-symmetry lattice.

The structural response is similarly strong in the quasi-1D Peierls insulator $\text{K}_{0.3}\text{MoO}_3$, the *blue bronze*, where the coherent oscillations represent the amplitude mode itself. Here the charge-density wave is a textbook periodic lattice distortion that opens a gap on a nested Fermi surface below $T_c = 183$ K. Huber *et al.* tracked the structural dynamics across the photoinduced charge-density-wave-to-metal transition with femtosecond x-ray diffraction [78]. At low excitation they followed coherent oscillations of the $(1, (4-q_b), 0.5)$ superlattice peak with a 1.7 THz frequency along the amplitude-mode coordinate, while above the melting threshold the periodic lattice distortion first collapsed and then transiently recovered according to the properties of the high-symmetry phase rather than those of the low-temperature phase. A photoemission study by Liu *et al.* resolved the same dynamics in the electronic structure, observing a 1.7 THz modulation of the bonding band as the signature of the amplitude mode, together with coherent phasons generated by the parametric amplification of phase fluctuations [79].

Similar cooperative responses of electronic and structural degrees of freedom are observed in layered transition-metal dichalcogenides. In these materials, the electronic modulation is accompanied by a commensurate structural supermodulation, with a wavelength that depends on the compound. In 1T-TiSe₂, Rohwer *et al.* used extreme-ultraviolet photoemission to image the band dispersion at large momenta and found that long-range charge order collapses within ~ 20 fs, much faster than lattice motion, a speed attributed to transient screening of the electronic interaction by photogenerated carriers [32]. In the related compound 1T-TaS₂, a two-dimensional Mott insulator with commensurate charge order, Petersen *et al.* showed that the charge order, identified by the splitting of occupied subbands at the Brillouin-zone boundary, melts before the lattice responds [33]. Together, these results challenge a purely structural, Fermi-surface-nesting picture and point to a leading electronic role. Photoexcited 1T-TaS₂ can also enter a hidden metallic state with metastable character [80], making it one of the clearest examples of a light-induced quantum state inaccessible by equilibrium means.

In cuprate superconductors, charge order is largely electronic and tied to the formation of alternating stripes of holes and electrons in the CuO_2 planes [82, 83]. It is widely regarded as competing with superconductivity. In $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ at $x = 1/8$ doping, this competition is the sharpest, as static stripe order suppresses bulk superconductivity, and possibly gives rise to a modulated phase known as pair density wave [84]. Early resonant x-ray scattering experiments showed a pump-induced suppression of the stripes under resonant vibrational excitation [85]. Mitrano *et al.* used time-resolved resonant soft x-ray scattering (elastic and in-

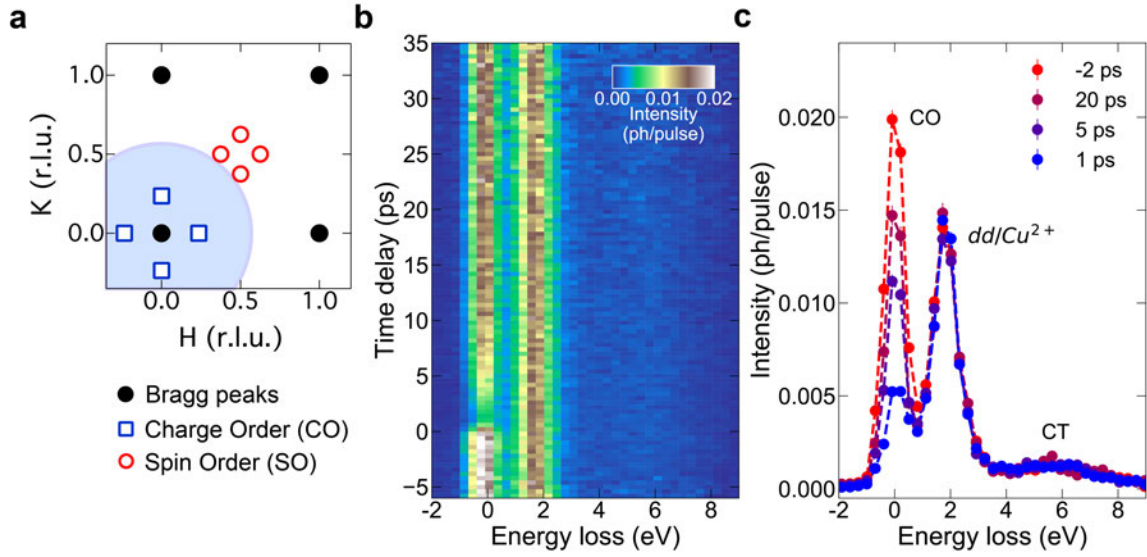


Fig. 6: Ultrafast charge order melting in a cuprate superconductor. *a.* Reciprocal space map of structural, charge order (CO) and spin order (SO) peaks in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$. The blue area represents the maximum momentum transfer achievable at the Cu L-edge (931 eV). *b.* Time-resolved resonant inelastic x-ray scattering (trRIXS) spectra at the Cu L-edge of photoexcited $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$ at $Q_{\text{CO}} = (0.236, 0, 1.5)$ r.l.u. for variable pump-probe time delay. *c.* trRIXS spectra for a selection of time delays showing a prompt melting of the CO, whereas higher energy features of the inelastic spectrum, e.g., d-d excitations/ Cu^{2+} emission and charge transfer (CT) excitations, remain unaffected. Error bars represent Poisson counting uncertainties. Adapted from Ref. [81] under CC BY-NC 4.0.

elastic) at a free-electron laser to measure the charge-order dynamics in this material after a 1.55 eV pump [81, 86]. The superlattice reflection was suppressed by up to 75% and recovered through a purely exponential relaxation, at odds with the coherent amplitude-mode oscillations observed in the Peierls systems discussed above (Fig. 6). The relaxation rate was found to be strongly momentum dependent and consistent with the presence of low-energy diffusive modes. At the lowest energies, the recovery obeys dynamic scaling laws governed by the annihilation of topological defects in the charge-order pattern. The competition between charge order and superconductivity has also been leveraged to observe the opposite effect, as quenching superconductivity with high-energy photons can enhance charge-order correlations. Wandel and co-workers demonstrated this effect in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ using time-resolved resonant soft x-ray scattering [87]. After an infrared pulse quenched superconductivity, the charge-order correlation length nearly doubled within roughly 1 ps in a nonthermal response. This enhancement was naturally understood as the annealing of charge-order topological defects, such as discommensurations, that are stabilized by superconductivity at equilibrium. Once the condensate is removed, the charge order becomes more coherent. In the language of Fig. 5a, removing one order deepens the free-energy minimum of the other.

Finally, optical excitation can stabilize charge-ordered states absent at equilibrium. LaTe_3 hosts a unidirectional charge-density wave along the c axis, while the nearly equivalent perpendicular a axis remains disordered due to a small structural anisotropy. Using ultrafast electron diffraction, Kogar *et al.* showed that photoexcitation weakens the equilibrium c -axis modulation and

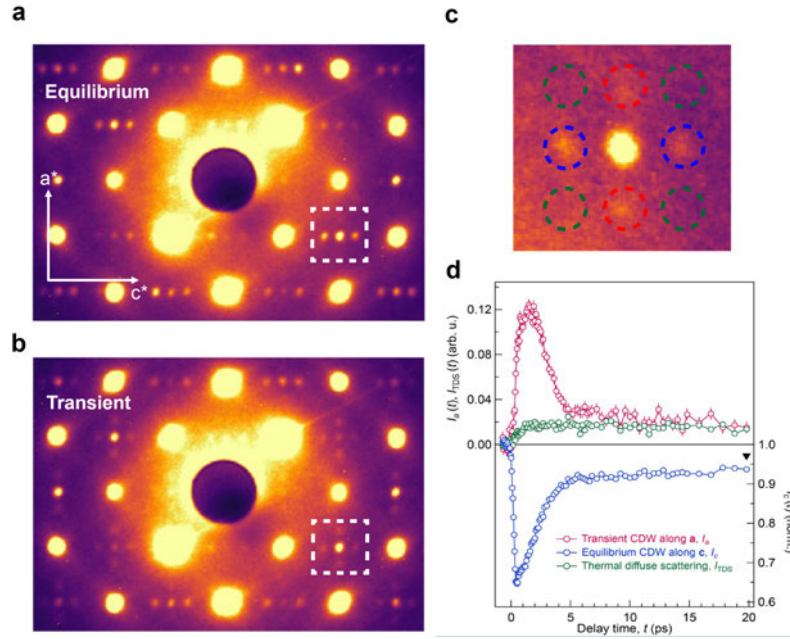


Fig. 7: Light-induced charge-density wave in LaTe_3 . (a)–(b) Ultrafast electron diffraction patterns before and 1.8 ps after photoexcitation with an 80 fs, 800 nm pulse. Dashed lines mark the regions where charge-order superlattice peaks appear. (c) Equilibrium c -axis satellites (blue) are suppressed while new a -axis satellites (red) emerge. Diffuse scattering is circled in green. (d) Integrated satellite intensities along the two axes as a function of time delay. The light-induced a -axis order grows as the c -axis order is quenched, and both recover on a common timescale of ~ 1.7 ps, revealing their mutual competition. Courtesy of Dr. A. Kogar [47].

transiently nucleates a new charge-density wave along the a axis (Fig. 7) [47]. The induced order rises within about 1 ps and decays as the original order recovers. Their nearly identical relaxation times show that the two orders draw on a common reservoir of charge carriers and are locked in direct competition. The transient a -axis wave is a genuine nonequilibrium state without counterparts in the equilibrium phase diagram, establishing light-induced excitation as a route to otherwise inaccessible states, a theme we return to in the discussion of hidden orders.

4.2 Magnetism

Magnetism is one of the most traditional arenas for the optical control of ordered states, and provides a clear example of how light-control strategies have evolved from controlling macroscopic properties (e.g., magnetization) to microscopic interactions (e.g., magnetic superexchange). Following the observation of ultrafast demagnetization in ferromagnetic nickel [88], ultrafast lasers enabled the detection and control of ferromagnetic and antiferromagnetic resonances [89,90], all-optical magnetic switching [91,92], and correlated spin dynamics across different magnetic sublattices [93]. These phenomena mostly involve changes in the macroscopic magnetization and arise from a combination of laser-induced heating, effective opto-magnetic fields, and optical perturbations of the magnetic anisotropy [20]. In this class of experiments, light acts on the spins through the electronic and optical response of the material, while the lattice enters mainly as a heat bath or passive degree of freedom.

With the advent of intense mid-infrared and THz sources, resonant protocols such as nonlinear phononics opened a new route to coherent magnetic control. In this case, coherently driven phonons, rather than photoexcited electrons, act on the spins, and the key nonlinearity is the coupling between the lattice and magnetic order, as summarized in Table 1. The lattice excitation strategy depends on the material, and rare-earth orthoferrites $R\text{FeO}_3$ have served as a model platform for these experiments. As canted antiferromagnets, their Dzyaloshinskii–Moriya interaction produces a small, magneto-optically readable net moment, while their proximity to anisotropy-driven spin-reorientation transitions amplifies weak perturbations of both THz spin resonances and rare-earth–iron f - d exchange. Magnetic states are often read out with time-resolved polarization rotation or birefringence measurements or with x-ray measurements.

In DyFeO_3 , resonant excitation of an infrared-active B_u phonon transiently renormalizes the magnetic superexchange due to a rectified distortion along an A_g mode coordinate and coherently steers the spin-reorientation transition, demonstrating direct control of the interaction J , rather than of the magnetization alone [94]. In ErFeO_3 , driving two near-degenerate B_u phonons with a controlled relative phase generates circular ionic motion, which mimics an effective magnetic field along the rotation axis and launches coherent spin precession [95]. Together with other experiments beyond the scope of this review, these results helped establish the field of *magnetophononics*, where lattice anharmonicity provides a symmetry-selective handle on magnetic interactions, and motivated theoretical descriptions of the effective fields, orbital moments, and exchange modulations carried by driven phonons [96,97]. In rare-earth halides, recent theoretical [98] and experimental [99] work has shown that coherent chiral phonons, driven by circularly polarized terahertz pulses, can polarize the paramagnetic spins of cerium fluoride in a manner analogous to a quasi-static magnetic field of order 1 T. In CoF_2 , three-phonon mixing processes of the form $Q_{IR,1}Q_{IR,2}Q_R$ distort the crystal field and decompensate the two spin sublattices, polarizing the antiferromagnet like a static piezomagnetic strain, but without changing the volume, and generating a ferrimagnetic moment of $0.2 \mu_B$ per unit cell [100].

More recent advances have pushed these ideas from transient modulation toward magnetic states that outlive the coherent drive, connecting magnetophononics to the broader theme of light-induced metastable order. In the titanate YTiO_3 , a ferromagnet whose order is suppressed by orbital fluctuations, Disa *et al.* showed that resonant excitation of a 9 THz oxygen rotation mode with B_{2u} symmetry enhances and stabilizes ferromagnetism, saturating the moment at low temperature and shifting the onset of order from $T_c = 27$ K to above 80 K [101]. Here, the driven phonon does not simply perturb the magnetization, but creates a more ordered state, the magnetic analog of the light-induced orders discussed for charge degrees of freedom. Such states can also be remarkably long lived. In the van der Waals antiferromagnet FePS_3 , THz pulses tuned to a phonon and a magnon induce a metastable magnetization that persists for more than two milliseconds, with a robustness that increases as the Néel transition is approached and order-parameter fluctuations diverge [102]. In this case, proximity to criticality, rather than a structural or thermal bottleneck, stabilizes the state, a mechanism we will encounter again when discussing hidden and metastable orders.

4.3 Ferroic orders

Ferroic orders form another broad class of ordered states with fundamental and technological relevance, and light provides an effective tuning knob for many of them. Although many types of ferroic order exist, here we focus on three classes: ferroelectrics, multiferroics, and ferroaxial systems [57]. A *ferroelectric* is a material with a nonzero electric polarization that can be switched between equivalent directions by an applied electric field. A *multiferroic* hosts two or more ferroic orders, most commonly ferroelectricity and magnetism, coupled through the magnetoelectric interaction, which enables electric-field control of magnetism. Finally, *ferroaxial* systems are characterized by a rotational texture that can host vortices of electric dipoles. Ferroaxial order breaks neither spatial inversion nor time-reversal symmetry, and is considered a promising platform for nonvolatile data storage.

Some of the most compelling examples of optical control of ferroelectrics build directly on nonlinear phononics. In LiNbO_3 , Mankowsky *et al.* demonstrated transient reversal of the ferroelectric polarization not by driving the soft polar mode directly, but by resonantly exciting a high-frequency auxiliary phonon that, through anharmonic coupling, displaces the ions along the polar coordinate [103]. The reversal could be induced in both directions, a prerequisite for switching, but remained short lived because depolarizing fields rapidly restored the original state after the drive. Overcoming this volatility requires a host whose polar order is intrinsically close to an instability, and SrTiO_3 is a paradigmatic example of such a system. Although its physics is rich and largely beyond the scope of this chapter, SrTiO_3 is a *quantum paraelectric* and lies on the disordered side of a quantum phase transition between paraelectric and ferroelectric states (Fig. 8a), where zero-point fluctuations of the polar mode suppress long-range order down to zero temperature. This proximity to a quantum critical point makes the polar order exceptionally susceptible to external perturbations.

Ultrafast experiments exploited this susceptibility directly. Nova *et al.* used mid-infrared pulses to drive the lattice of SrTiO_3 and reported an emergent *metastable* ferroelectric state, identified by an optical second-harmonic signal that indicates broken inversion symmetry and builds up over many pump cycles [25]. The likely mechanism behind these observations is a rectified, long-lived strain generated by the nonlinear drive, which biases the soft mode toward the polar minimum and is stabilized by the softness of the lattice near the transition. In parallel, Li and co-workers used an intense THz field tuned to the soft mode itself to drive the ions directly toward their ferroelectric positions, inducing a transient hidden ferroelectric phase [105]. Together, these indirect nonlinear phononic and direct soft-mode routes established SrTiO_3 as a testbed for light-induced ferroelectricity.

Disentangling the structural origin of these states requires structural probes such as time-resolved x-ray scattering. Using femtosecond x-ray diffuse scattering, Fechner *et al.* showed that resonant excitation of the 17 THz Ti–O stretching mode quenches lattice fluctuations and tips the balance between the polar instability and the competing antiferrodistortive oxygen rotations that frustrate ferroelectricity in equilibrium [106]. This picture was later refined by measurements resolving how the drive dynamically suppresses these rotations [107]. A complementary ques-

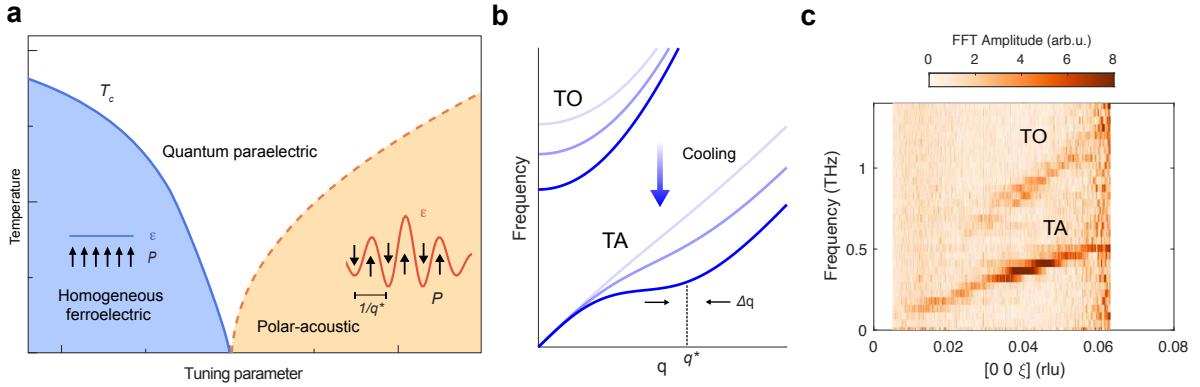


Fig. 8: Phase diagram and low-energy modes of SrTiO_3 . **a.** Phase diagram showing the nearby ferroelectric (blue) and polar acoustic (orange) phases around the quantum paraelectric regime. The tuning parameter here is the flexoelectric coupling strength. **b.** Dispersion of transverse phonon modes. When lowering temperature, the TO branch softens and the TA branch deviates from linear dispersion developing a kink at finite wavevector k^* , indicative of an instability. These excitations are polar and modulate the polarization spatially with a wavelength $\sim 2\pi/k^*$. **c.** Magnitude of the Fourier transform of time-domain x-ray scattering data at 20 K. Adapted with permission from Ref. [104]. Copyright (2025) by Springer Nature Limited.

tion is whether the “failed” ferroelectric transition in SrTiO_3 is frustrated only by fluctuations of the uniform soft mode, or also by polar correlations at finite wavevector. To address this, Orenstein *et al.* combined single-cycle THz excitation, resonant with the soft modes of SrTiO_3 , with femtosecond hard x-ray diffuse scattering around the $(3, 3, 3)$ Bragg peak [104]. A crucial element of this experiment was the ability to reverse the polarity of the THz field. The diffuse scattering signal at finite momentum reversed with the field, demonstrating that the observed acoustic-like oscillations carried polar character and therefore broke local inversion symmetry. Fourier-transform inelastic x-ray scattering then reconstructed the dispersion and spatial texture of these modes (Fig. 8b-c), revealing nanometer-scale *polarization density waves*, namely mesoscopic modulations of the polarization that intensify near the quantum critical point. These experiments underscore the power of x-rays to make the structural degrees of freedom underlying the ferroelectric transition directly observable.

Ultrafast experiments on multiferroics are extensive, and we use here one representative example of light control. In the spin-spiral multiferroic TbMnO_3 , magnetoelectric coupling produces *electromagnons*, spin excitations that are active to the electric field of light rather than only to its magnetic field. Below 27 K, TbMnO_3 develops an incommensurate cycloidal spin order that also generates a ferroelectric polarization, making it a natural platform to test whether an optical electric field can act directly on magnetic order. Kubacka *et al.* drove the strongest electromagnon with an intense few-cycle THz pulse centered at 1.8 THz, with the electric field oriented along the crystallographic direction that activates this mode [108]. They then followed the magnetic response using femtosecond resonant soft x-ray diffraction at the Mn L_2 edge, measuring the intensity of the $(0, q, 0)$ cycloidal magnetic reflection. The diffracted intensity oscillated at the electromagnon frequency, reversed sign when the THz electric field was in-

verted, and was strongly suppressed outside the multiferroic phase. These observations showed that the electric-field channel opened by magnetoelectric coupling can coherently rotate the spin cycloid, providing a route to magnetic control on sub-picosecond timescales.

Recent experiments extend these control protocols to ferroaxial order. In ferroaxial systems, the order parameter is an axial vector associated with a rotational distortion, so switched domains need not be driven back by a restoring stray field. In $\text{RbFe}(\text{MoO}_4)_2$, Zeng *et al.* engineered the conjugate field to this axial order using circularly polarized THz pulses resonant with a doubly degenerate infrared-active phonon [109]. The driven phonon displacement and the THz electric field rotate together, and their cross product acts as an effective axial field whose sign is set by the light helicity. Probed by circular-dichroic second-harmonic generation, a single pulse switched the ferroaxial domain, while a pulse of opposite helicity switched it back. Remarkably, the switched state persisted for many hours, demonstrating photoinduced, nonvolatile, and rewritable ferroaxial switching.

5 Floquet engineering

We now turn to a conceptually distinct regime, in which the optical field coherently *dresses* the electronic states for as long as it is present. When a nonresonant field oscillates faster than the electrons can follow, it does not primarily deposit energy, but instead renormalizes the Hamiltonian itself, shifting and reshaping bands much as the ac Stark effect splits atomic levels [110]. *Floquet engineering* provides the natural language for this physics. It recasts a time-periodic problem in terms of an effective, time-independent Hamiltonian, turning coherent optical dressing into a quasistationary band-structure problem and making light a direct handle on the electronic Hamiltonians and topology of quantum materials [111, 112].

Following Ref. [4], the essential idea is captured by a periodically driven atomic chain (Fig. 9a). Consider noninteracting spinless fermions on a one-dimensional lattice,

$$H = -t_0 \sum_n (c_n^\dagger c_{n+1} + \text{H.c.}) = \sum_k \varepsilon(k) c_k^\dagger c_k, \quad \varepsilon(k) = -2t_0 \cos(ka), \quad (4)$$

with hopping t_0 , lattice constant a , and a cosine band of width $4t_0$. A longitudinal field $E(t) = E_0 \sin(\omega t)$ enters through the Peierls substitution $t_0 \rightarrow t_h(t) = t_0 e^{ieaA(t)/\hbar}$, where $A(t) = (E_0/\omega) \cos(\omega t)$ is the vector potential. When the drive is fast compared with the electronic timescales, the rapid phase oscillations average out, giving the time-averaged hopping

$$t_{\text{eff}} = \langle t_h(t) \rangle = t_0 J_0(F), \quad F = \frac{aeE_0}{\hbar\omega}, \quad (5)$$

where J_0 is the zeroth-order Bessel function and F is the dimensionless drive strength. Since $|J_0(F)| < 1$, the effective hopping is reduced. This process results in the emergence of *dynamical localization* [113], as the band narrows, carriers localize, and transport is fully suppressed at the zeros of J_0 . Equivalently, in momentum space the field shakes the dispersion through the minimal coupling $k \rightarrow k - eA(t)/\hbar$, so that the period-averaged band is flatter than its equilibrium counterpart (Fig. 9a). In correlated systems, the same construction can renormalize

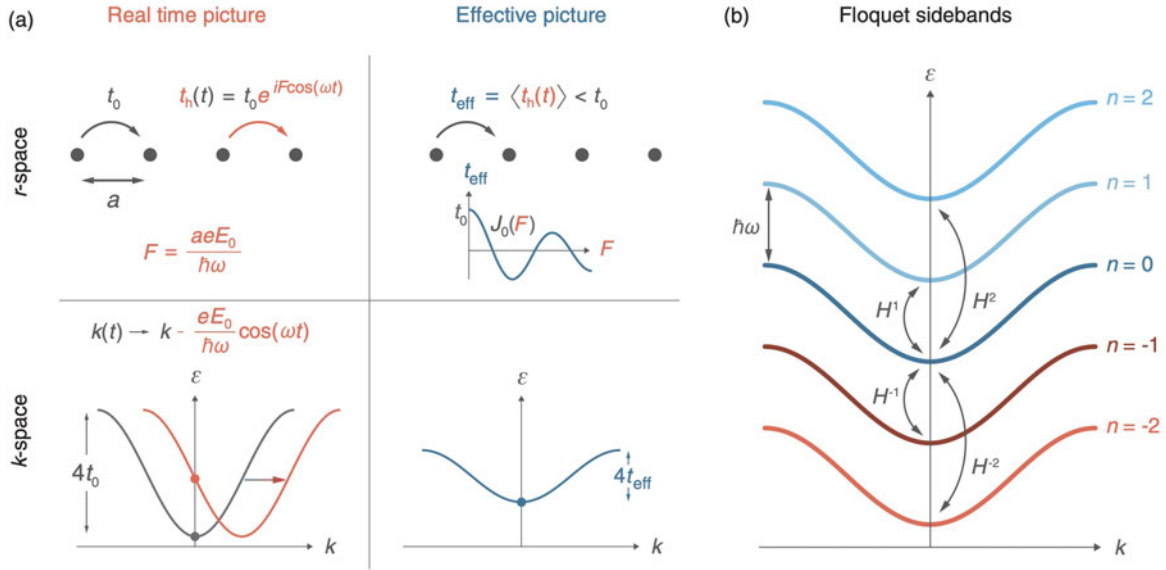


Fig. 9: Floquet engineering of a band structure. (a) Dynamical localization in a driven atomic chain. In real time, the optical field periodically modulates the hopping and shakes the electronic band in momentum space. When time-averaged, this motion reduces the bandwidth and leads to a smaller effective hopping amplitude, $t_{\text{eff}} = t_0 J_0(F)$. (b) Floquet representation of a periodically driven system. The time-dependent Hamiltonian is mapped onto replicas of the equilibrium band structure separated by integer multiples of the drive frequency ω . Reprinted figure with permission from Ref. [4]. Copyright (2022) by the American Physical Society.

superexchange and other electronic interactions, providing a coherent route to control strongly correlated phases [114].

This time-averaged description of coherent dressing allows the driven material to be viewed in terms of a reconstructed band structure in the presence of the electromagnetic field. For a Hamiltonian with period $T = 2\pi/\omega$, the Floquet theorem [115] allows one to write the solutions of the Schrödinger equation as $|\Psi_l(t)\rangle = e^{-i\varepsilon_l t/\hbar} |\Phi_l(t)\rangle$, where $|\Phi_l(t)\rangle$ are T -periodic states and ε_l are quasienergies, the temporal analog of crystal momentum in Bloch's theorem. Expanding the Hamiltonian in Fourier harmonics, $H_n = (1/T) \int_0^T dt e^{in\omega t} H(t)$, maps the time-dependent problem onto a static eigenvalue problem in an extended space

$$\mathcal{H} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \\ \cdots & H_0 + \hbar\omega & H_{-1} & H_{-2} & \cdots \\ \cdots & H_1 & H_0 & H_{-1} & \cdots \\ \cdots & H_2 & H_1 & H_0 - \hbar\omega & \cdots \\ & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (6)$$

given by an infinite ladder of replicas of the time-averaged Hamiltonian H_0, H_1, H_2 , shifted by integer multiples of $\hbar\omega$ and coupled by harmonics of the drive. In a solid, these replicas form *Floquet-Bloch bands*, periodic in both momentum and energy and separated by the photon energy $\hbar\omega$ (Fig. 9b). In the high-frequency limit, only H_0 remains and one recovers the dressed

band of Eq. (5). At finite frequency, however, the sidebands hybridize. The higher-order terms, to leading order $H_{\text{eff}} \simeq H_0 + [H_{-1}, H_1]/\hbar\omega$ [116], open gaps and can imprint properties with no equilibrium analog, including nontrivial band topology [117–119].

Applying this elegant picture to ultrafast solid-state spectroscopy is demanding. Reaching $F \sim 1$ requires large fields, typically 10^7 – 10^8 Vm $^{-1}$, which confines experiments to short, low-frequency pulses and places coherent dressing in direct competition with heating and decoherence. Nevertheless, over the past decade, experimental advances have enabled the observation of Floquet states in a range of solid-state platforms, which we review below.

5.1 Control of band structure and functional responses

The observation of Floquet-engineered states in solids has been driven by a set of theoretical predictions. As an example, Oka and Aoki predicted that a circularly polarized field could open a gap at the Dirac point and induce a photoinduced dc Hall current [117], while Lindner, Refael, and Galitski showed that periodic driving could turn a trivial semiconductor quantum well into a *Floquet topological insulator* hosting helical edge states [118]. These proposals established that optical fields could be used to actively manipulate the band topology.

Resolving coherent band-structure dressing requires a momentum-resolved probe, and trARPES quickly became the technique of choice for these experiments. In a landmark work, Wang *et al.* excited the topological surface states of Bi₂Se₃ with mid-infrared pulses (≈ 120 meV) and used trARPES to image the emergence of Floquet-Bloch replicas of the Dirac cone [34]. These replicas only appeared upon hybridization between the light-field and the electrons at overlap with the pump and were separated by the photon energy, as discussed in the previous section. Further, they developed gaps at crossing points between bands of different Floquet order, thus providing direct evidence that light could coherently reconstruct the band structures.

The interpretation of these spectra, however, requires some care, as addressed in subsequent work by Mahmood *et al.* [35]. Floquet-Bloch states are dressed states of the *Bloch electrons inside the crystal*, as they inherit the properties of the equilibrium bands of a material and disperse as replicas of those bands. Volkov states, by contrast, are dressed states of the *outgoing photoelectron in vacuum*. They are also referred to as laser-assisted photoemission (LAPE), and their sidebands are tied to the free-electron final state rather than to the band structure of the material. Since both effects generate replicas spaced by $\hbar\omega$, and since the two channels can interfere at the surface, distinguishing them is essential. This can be achieved by eliminating the out-of-plane electric field of the pump. By tuning the pump polarization and exploiting the different momentum dependence of the two couplings, Mahmood *et al.* showed that the Volkov contribution can be suppressed, allowing the genuine Floquet-Bloch response to be isolated.

How rapidly this dressing is established was addressed by Ito *et al.*, who used subcycle trARPES to follow the topological surface state of Bi₂Te₃ under a strong, phase-stable mid-infrared drive [36]. The pump was an in-plane, *s*-polarized field centered at 25 THz ($\hbar\Omega \simeq 0.1$ eV), with a surface field of ~ 0.8 MV cm $^{-1}$, while the probe was a 17 fs ultraviolet beam. They first observed a pump-induced intraband current that appeared as an asymmetry of the Dirac cone,

roughly one half-cycle before the field maximum. Then, Floquet-Bloch sidebands emerged within a single optical cycle and gained strong weight near the field crest. Thus, in a sufficiently coherent band, Floquet dressing can form on the timescale of one optical period, linking band-structure engineering directly to the intraband acceleration and interband transitions in solids. Floquet dressing was demonstrated at nearly the same time in a semiconductor, where Zhou *et al.* hybridized the valence band of black phosphorus with the $n = -1$ replica of the conduction band, opening a dynamical gap that decayed within ~ 130 fs [120].

Graphene, the system for which Floquet topological gaps were first predicted, has also been driven into Floquet-engineered states with mid-infrared light. In a picosecond transport experiment, McIver *et al.* drove monolayer graphene with circularly polarized mid-infrared pulses (46 THz, 191 meV) and sampled the transverse current on-chip with a laser-triggered photoconductive switch, observing a light-induced anomalous Hall effect [53]. The gate dependence revealed a ~ 60 meV conductance plateau around the Dirac point, where a Haldane-like Floquet gap is expected, with an anomalous Hall conductance approaching $1.8 e^2/h$. Direct spectroscopic evidence arrived only recently, because graphene poses additional constraints for trARPES experiments: its Dirac cone lies at the K point, requiring high-harmonic probe photons, while Volkov sidebands and fast scattering can obscure weak Floquet replicas. Two back-to-back recent trARPES studies overcame these obstacles and successfully reported evidence for Floquet replicas in graphene (Fig. 10). Choi *et al.* used a $5 \mu\text{m}$ (246 meV) mid-infrared drive and a 26.4 eV probe to resolve replica Dirac cones in monolayer epitaxial graphene. By rotating the pump polarization, they showed that the replicas arise from the scattering between Floquet-Bloch and Volkov states [121]. Merboldt *et al.* reached the same conclusion with time-resolved momentum microscopy, using a 0.65 eV infrared drive and a 26.5 eV extreme-ultraviolet probe. The momentum asymmetry of the sidebands, together with calculations including ultrafast dephasing, showed that Floquet sidebands survive even when scattering broadens the light-induced gaps [122]. Similar experiments have also yielded Floquet-engineered states and gap opening beyond Dirac semimetals. Bielinski *et al.* used helicity-dependent, circularly polarized mid-infrared driving and trARPES to control the Dirac mass gap of the topological antiferromagnet MnBi_2Te_4 , with opposite pump helicities producing different gaps [37].

Because Floquet engineering reshapes electronic structure, it also appears in optical and functional responses. A clear example is that of valleytronics in monolayer transition-metal dichalcogenides. Sie *et al.* used intense circularly polarized light tuned below the gap of monolayer WS_2 and transient exciton reflectance to resolve a valley-selective *optical (ac) Stark shift* up to 18 meV [123], and later isolated the much smaller *Bloch-Siegert shift* [124]. Valley-selective circular dichroism was key to separate the two contributions, providing a textbook example of Floquet dressing in an effective two-level system. These experiments showed that the co-rotating field component produces the optical Stark shift, whereas the usually neglected counter-rotating component produces the Bloch-Siegert shift.

The examples discussed so far largely involve weakly interacting band electrons. A complementary limit appears when the field dresses localized single-ion multiplets. Shan *et al.* drove the magnetic insulator MnPS_3 far off resonance from the on-site Mn d - d transition, using fields

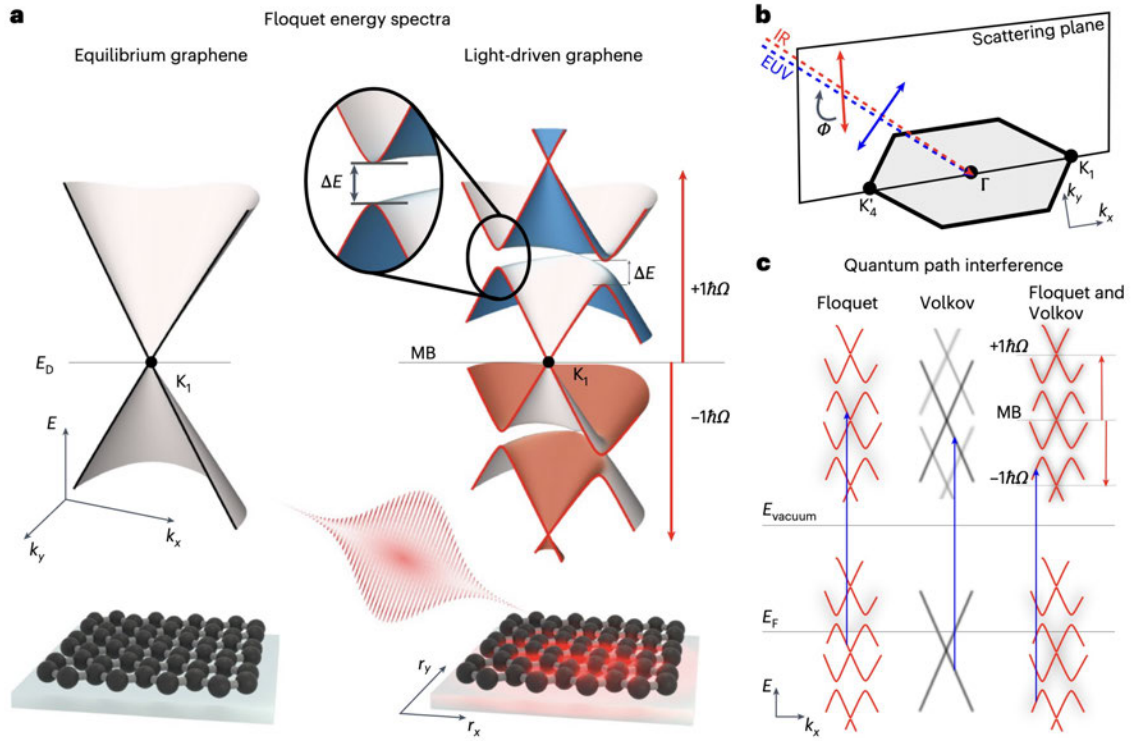


Fig. 10: Floquet engineering of electronic bands in graphene. (a) Energy- and in-plane-momentum-resolved sketch of the graphene bands near the K point in equilibrium (black) and under periodic driving (red). Coherent dressing produces sidebands of the main Dirac band, spaced by the drive photon energy $\hbar\Omega$, and energy gaps ΔE that open where sidebands of different photon order cross and hybridize. (b) Experimental geometry of the pump probe experiment. (c) Resulting Floquet and Volkov states, as well as their quantum interference. Reprinted from Ref. [122], licensed under a Creative Commons Attribution 4.0 International License. Copyright (2025) by the authors.

of order 10^9 Vm^{-1} , and observed coherent on-off switching of the second-harmonic efficiency within 100 fs, with an on-off ratio exceeding ten and no measurable dissipation [125]. This response corresponds to a giant Floquet renormalization of the optical nonlinearity. In a different experiment, Zhang *et al.* used off-resonant linearly polarized light and time-resolved second-harmonic generation to induce an electronic polarization in antiferromagnetic Cr_2O_3 [26]. These studies broaden Floquet engineering from the reshaping of itinerant bands to the control of atomic-like states and single-ion responses.

5.2 Towards quantum control of correlated states

When periodic driving addresses only a few isolated electronic levels, Floquet engineering becomes a form of quantum control. This limit has a long history in weakly interacting systems, where ultrafast optical fields have produced coherent Rabi rotations, picosecond population control, and complete state manipulation in atoms, semiconductor excitons, and quantum dots [126–129]. Extending this control to strongly correlated electrons is a central frontier [4,5], both because these systems host the most dramatic nonequilibrium responses and because pe-

periodic driving is predicted to renormalize interactions and stabilize exotic electronic phases. The key challenge is to identify correlated states that can be coherently addressed and read out before dissipation and decoherence destroy the coherent Floquet dressing.

Hubbard excitons in one-dimensional Mott insulators are a compelling platform for demonstrating Floquet engineering of many-body states. Unlike conventional electron-hole excitons, they are bound holon-doublon pairs stabilized by Coulomb repulsion [130, 131]. Together with spin-charge separation, this binding leads to nearly degenerate even- and odd-parity states. They therefore realize a many-body, two-level system with strong, well-characterized linear and non-linear optical responses [132, 133]. Their few-level structure gives direct access to the coherent evolution of the many-body wavefunction, which is otherwise difficult to isolate from band renormalization and symmetry breaking in more complex correlated materials.

Driving the excitons with an intense, nonresonant mid-infrared field $\vec{E}_{\text{MIR}}(t) = \vec{E}_{\text{MIR}}^0 \cos(\Omega t)$ couples the bright odd state $|u\rangle$ and the dark even state $|g\rangle$ through $\hat{H}_{\text{int}}(t) = \vec{\mu}_{ug} \cdot \vec{E}_{\text{MIR}} |u\rangle\langle g| + \text{h.c.}$ The optical Stark and Bloch-Siegert shifts, the usual fingerprints of Floquet driving, carry opposite signs and nearly cancel for a degenerate pair [123–125], so that the dominant effect is a coherent mixing of the two parities within the optical cycle. The Floquet states retain their quasienergy but acquire a time-dependent composition,

$$|u(t)\rangle = e^{iE_{\text{exc}}t/\hbar} \left[\cos \frac{\vartheta(t)}{2} |u\rangle + e^{i\varphi(t)} \sin \frac{\vartheta(t)}{2} |g\rangle \right] \quad (7)$$

$$|g(t)\rangle = e^{iE_{\text{exc}}t/\hbar} \left[\cos \frac{\vartheta(t)}{2} |g\rangle - e^{-i\varphi(t)} \sin \frac{\vartheta(t)}{2} |u\rangle \right] \quad (8)$$

which describes a coherent rotation of the many-body wavefunction on a Bloch sphere spanned by the even and odd excitons, with polar and azimuthal angles $\vartheta(t)$ and $\varphi(t)$ evolving at twice the drive frequency. Cycle-averaged, this rotation redistributes spectral weight into Floquet sidebands at $E_{\text{exc}} \pm 2n\Omega$ for $|u(t)\rangle$ and $E_{\text{exc}} \pm (2n+1)\Omega$ for $|g(t)\rangle$. Although the drive is far off-resonance, at strong fields the coupling $g = \vec{\mu}_{ug} \cdot \vec{E}_{\text{MIR}}$ becomes comparable to Ω , so the Rabi frequency locks to the drive frequency and g sets only the amplitude of the rotation. This ultrastrong-coupling regime, made possible jointly by the near-degeneracy of the excitons and the large field, departs from conventional resonant two-level driving and from the quantum-dot experiments above, where the coupling is smaller than both level splitting and drive frequency. Baykusheva *et al.* implemented this protocol in the chain cuprate Sr_2CuO_3 [134]. Here, Hubbard excitons form two near-degenerate levels around 1.72 eV. A mid-infrared pump at 0.12 eV (28 THz), far below the 1.8 eV Mott gap, dresses the exciton, while a near-infrared probe generates resonant third-harmonic emission at the exciton energy. Because $|g\rangle$ is dark, rotation toward the even state suppresses the third-harmonic yield by up to 70% at 1.8 MV cm^{-1} (Fig. 11). The polarization and field dependences follow the J_0 Bessel form expected for Floquet dressing, and probe detuning reveals a genuine sub-gap Floquet sideband in $\chi^{(3)}$, showing that coherence survives several drive cycles. At higher fields, the wavefunction undergoes continuous oscillatory rotations, reaching an instantaneous angle of $\pi/2$ at 1.7 MV cm^{-1} . This uniaxial $\pi/2$ rotation is the minimal building block of programmable quantum control sequences based on pulse shaping techniques and opens the route to Ramsey interferometry, dynamical decoupling, and ultimately full $\text{SU}(2)$ control of correlated many-body wavefunctions [135–137].

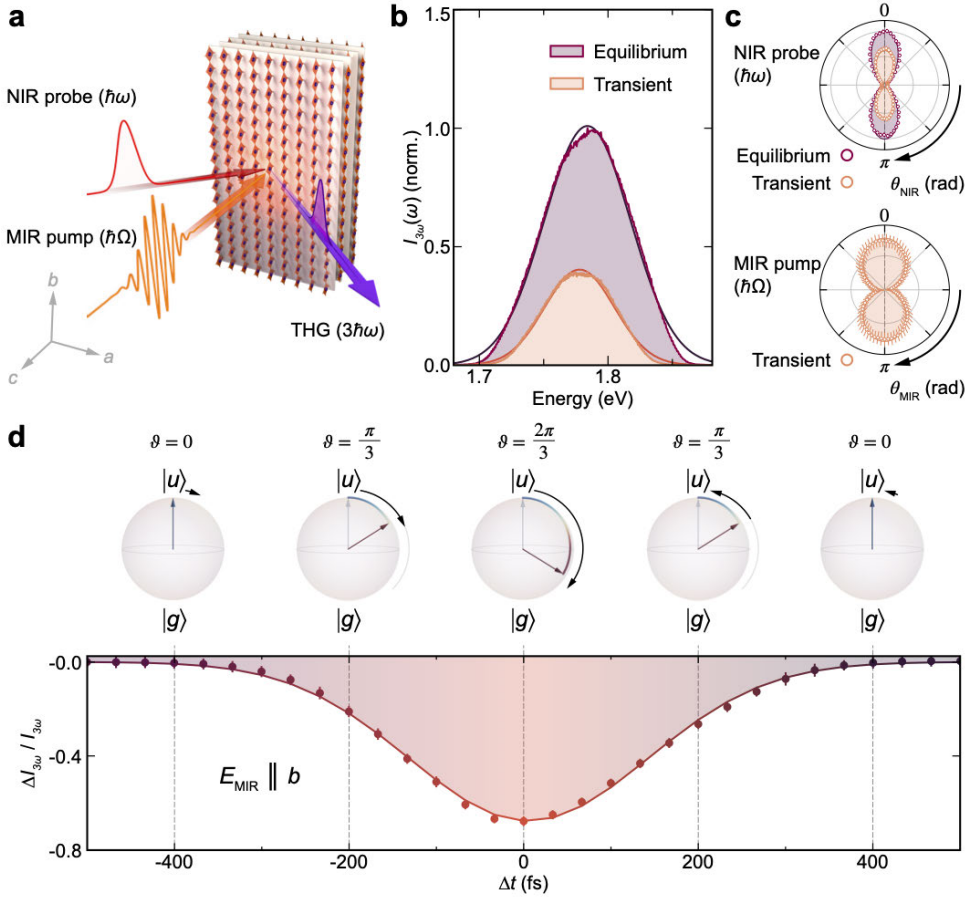


Fig. 11: *Quantum control of Hubbard excitons in Sr_2CuO_3 . (a) Sketch of the mid-infrared pump ($\hbar\Omega = 0.12$ eV) third-harmonic emission (THG) probe experiment. (b) Normalized THG spectra of a 0.59 eV probe at equilibrium (purple) and at zero delay after a 1.8 MV cm^{-1} pump (orange). (c) THG intensity as a function of probe (top) and pump (bottom) polarization angle. The suppression vanishes for a pump polarized normal to the chains and follows the $1 - J_0(A_1 \cos \theta)^4$ dependence (solid line) expected for Floquet dressing. (d) Time-resolved differential THG $\Delta I_{3\omega}/I_{3\omega}$ for $E_{\text{MIR}} = 1.8 \text{ MV cm}^{-1}$. The Bloch spheres mark the instantaneous rotation angle ϑ at selected delays. Reprinted with permission from Ref. [134]. Copyright (2026) by Springer Nature Limited.*

5.3 Electronic metastability via coherent dressing

The Floquet-engineered states discussed so far exist only while the drive is present. Because the dressed bands and gaps arise from hybridization with the photon field, they decay once pump coherence dissipates, and the renormalized Hamiltonian returns to its equilibrium form. This short lifetime is the main obstacle to converting Floquet-engineered states into persistent, functional states of matter. One alternative route is to use longer drives, but this requires careful control of energy dissipation to avoid runaway heating. Another is to use dressing not as an end in itself, but as a transient *gate*. While present, the field can lift a symmetry that otherwise forbids a quantum process, allowing the system to cross a barrier into a configuration that survives after the drive is gone. Once the symmetry is restored, the reverse pathway is again blocked, leaving the system trapped in a long-lived metastable state.

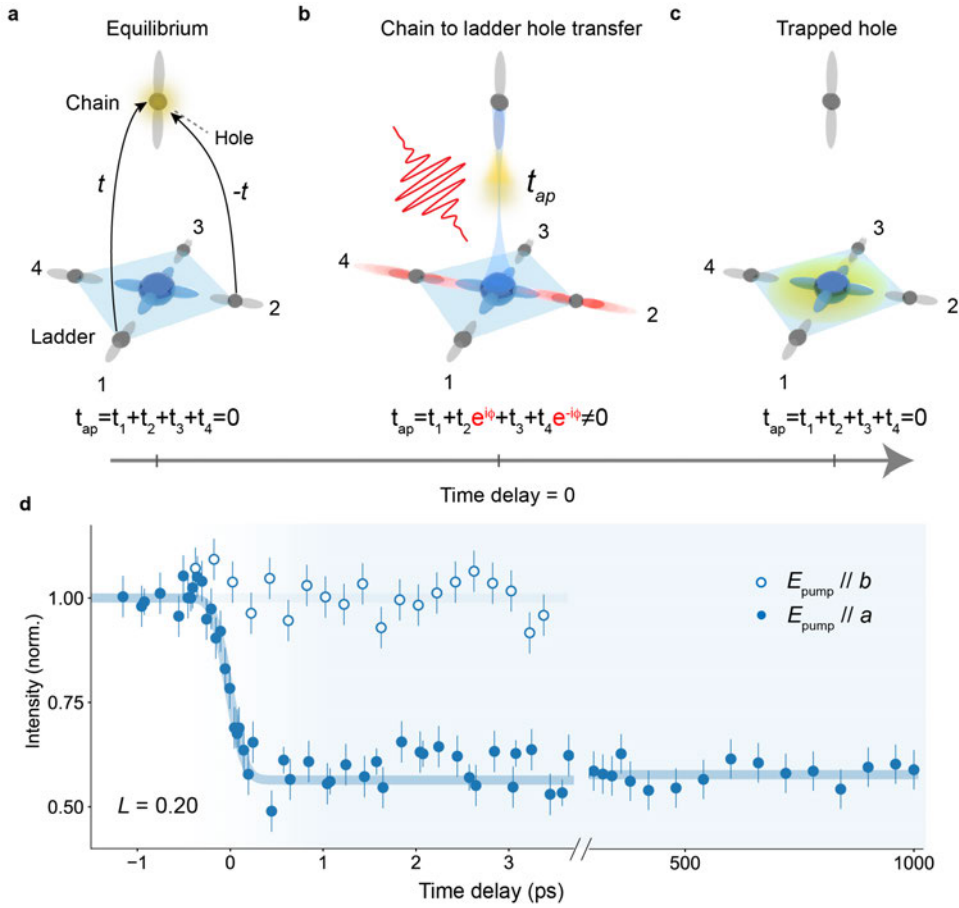


Fig. 12: Electronic metastability in $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$. **a.** Each plaquette on the ladder features O 2p (grey) and Cu 3d (blue) orbitals constituting a Zhang-Rice singlet. At equilibrium, holes mostly reside on the chains and the approximate D_{4h} symmetry makes the hopping contributions to the apical chain oxygen alternate in sign and cancel, resulting in vanishing hopping t_{ap} between chains and ladders. **b.** However, this symmetry is broken when the Hamiltonian is dressed by an intense in-plane pump field E_{pump} (time delay = 0), resulting in a nonzero t_{ap} . The transient activation of the symmetry-forbidden hopping t_{ap} leads to a chain-to-ladder hole transfer. **c.** Once the field is removed, t_{ap} vanishes, trapping the transferred holes in the ladder. **d.** Time-dependent charge-order diffraction intensity at $L = 0.20$. For an in-plane pump ($E_{\text{pump}} \parallel a$, filled circles) the charge order is suppressed within the pulse and remains so up to nanosecond delays, whereas an out-of-plane pump ($E_{\text{pump}} \parallel b$, open circles) produces no suppression, as expected for the symmetry-activated mechanism. Adapted with permission from Ref. [49]. Copyright (2025) by Springer Nature Limited.

A recent realization of this idea was demonstrated in the optically driven cuprate ladder [49]. The cuprate material $\text{Sr}_{14}\text{Cu}_{24}\text{O}_{41}$ is a self-doped charge-transfer insulator consisting of two nearly decoupled subsystems: hole-rich chains that act as charge reservoirs, and two-leg ladders that host charge order below 250 K. Padma *et al.* excited single crystals with intense near-infrared pulses (1.55 eV, just below the charge-transfer gap, up to 7.7 MV cm^{-1}) polarized in the ladder plane, and tracked the response with time-resolved THz reflectivity, trXRD, trXAS, and trRIXS. From the charge-ordered state, the 100-fs pump induces a nonequilibrium

electronic state that persists for tens of nanoseconds, far beyond the pulse duration (Fig. 12d). trXAS reveals a transfer of roughly 3% holes per Cu from chains to ladders, trXRD shows the concomitant suppression of charge-order correlations, and the low-energy response develops emergent collective modes [138]. Together, these probes identify a metastable state stabilized by the transient activation of an otherwise forbidden chain-to-ladder hopping channel.

Why this channel is closed at equilibrium, and how light opens it, follows from a transparent symmetry argument (Fig. 12a-c). The doped holes in the ladder occupy Zhang-Rice singlets built from the Cu $3d_{x^2-y^2}$ orbital and the surrounding in-plane O $2p$ orbitals. The relevant ligand combination is the symmetric superposition

$$\psi_p(\mathbf{r}; \mathbf{R}) = \frac{1}{2} \left[\psi_{p_x}(\mathbf{r}; \mathbf{R} - \frac{a_0}{2} \hat{x}) - \psi_{p_x}(\mathbf{r}; \mathbf{R} + \frac{a_0}{2} \hat{x}) + \psi_{p_y}(\mathbf{r}; \mathbf{R} - \frac{a_0}{2} \hat{y}) - \psi_{p_y}(\mathbf{r}; \mathbf{R} + \frac{a_0}{2} \hat{y}) \right], \quad (9)$$

which, like the $3d_{x^2-y^2}$ orbital it accompanies, is antisymmetric under a quarter rotation about the Cu-apical-oxygen axis,

$$\hat{R}(\frac{\pi}{2}) \psi_p(\mathbf{r}; \mathbf{R}) = -\psi_p(\mathbf{r}; \mathbf{R}), \quad (10)$$

whereas the apical-oxygen $2p_z$ orbital that bridges the ladder to the chain is invariant under this symmetry operation. The hopping integral between the ladder state and the apical oxygen is then a sum of four rotated contributions whose alternating signs cancel exactly,

$$t_{ppz} = \frac{1}{4} \left[\int \psi_{2p_z}^* \mathcal{H} \psi_p d\mathbf{r}^3 - \int \psi_{2p_z}^* \mathcal{H} \psi_p d\mathbf{r}^3 + \int \psi_{2p_z}^* \mathcal{H} \psi_p d\mathbf{r}^3 - \int \psi_{2p_z}^* \mathcal{H} \psi_p d\mathbf{r}^3 \right] = 0, \quad (11)$$

so the chains and ladders are decoupled by symmetry. An intense in-plane field can lift this symmetry and produce a transient tunneling matrix element. Introducing the field through the minimal coupling

$$\mathcal{H}(\mathbf{A}) = \mathcal{H}(\mathbf{A}=0) - \frac{e}{2m_e c} \sum_j [i\hbar \nabla_j \cdot \mathbf{A}(\mathbf{r}_j) + \text{h.c.}] + O(\mathbf{A}^2), \quad (12)$$

and incorporating it as a Peierls phase on the hopping, $t(\mathbf{r}_i, \mathbf{r}_j) \rightarrow t(\mathbf{r}_i, \mathbf{r}_j) \exp\left(\frac{ie}{\hbar c} \int_{\mathbf{r}_i}^{\mathbf{r}_j} \mathbf{A}(\mathbf{r}) \cdot d\mathbf{r}\right)$, unbalances the four contributions according to their orientation relative to the field. The cancellation is now incomplete, and the apical hopping becomes finite,

$$t_{ppz} = \frac{1}{2} \left(t_{pp} + t_{pp} - t_{pp} e^{-i\frac{eA_0 a_0}{2\hbar c}} - t_{pp} e^{i\frac{eA_0 a_0}{2\hbar c}} \right) = t_{pp} \left[1 - \cos\left(\frac{eA_0 a_0}{2\hbar c}\right) \right] \simeq \frac{1}{2} \left(\frac{eA_0 a_0}{2\hbar c} \right)^2 t_{pp} \propto A_0^2, \quad (13)$$

scaling as the square of the field amplitude. Averaging over the drive period recovers the steady-state effective hopping that characterizes the coherently-dressed ladder,

$$t_{ppz}^{(\text{eff})} = \frac{\Omega}{2\pi} \int_0^{2\pi/\Omega} t_{pp} \left[1 - \cos\left(\frac{ea_0 A_0 \cos(\Omega\tau)}{2\hbar c}\right) \right] d\tau = \left[1 - \mathcal{J}_0\left(\frac{ea_0 A_0}{2\hbar c}\right) \right] t_{pp}, \quad (14)$$

where $\mathcal{J}_0$ is the zeroth-order Bessel function. In the weak-field limit, this activated hopping also grows as A_0^2 . It acts as a transient gate: while the pulse is present, holes flow from the chains

into the lower-density ladders; once coherence is lost, symmetry is restored, $t_{ppz} \rightarrow 0$, and the transferred holes are trapped. The same symmetry argument applies to the Cu orbital matrix elements, as discussed in Ref. [49]. The observed E^2 scaling and the absence of a response for out-of-plane polarization are therefore direct fingerprints of symmetry-activated charge transfer. Metastable photoinduced states are otherwise rare, and typically rely on material-specific cooperative effects, often involving the lattice, that trap the system behind a structural barrier. The significance of this experiment is that it isolates a purely electronic, symmetry-based route to metastability without structural transitions. Coherent optical dressing transiently switches on a forbidden term in the Hamiltonian, redistributes charge, and then leaves the system trapped once the symmetry is restored. This establishes transient Hamiltonian engineering as a general design principle for metastable quantum states, applicable wherever symmetry separates a material into weakly coupled sectors.

6 Light-induced superconductivity

As seen in the previous sections, coherent control can drive materials into ordered states. A major challenge is whether light can also engineer macroscopic quantum coherence and produce superconducting responses with enhanced or emergent properties at temperatures far above those accessible in equilibrium. This section reviews the experimental evidence for *light-induced superconductivity*, and the theoretical scenarios proposed to explain it.

6.1 Experimental reports of light-induced superconductivity

The relationship between light and superconductivity is often antagonistic. Superconducting condensates are among the most fragile electronic ground states, and eV photons have been shown to readily destroy them by breaking Cooper pairs, populating the gap with hot quasiparticles, and quenching macroscopic phase coherence. Pump–probe experiments have long exploited this effect to track the suppression and recovery of superconductivity, following pair-breaking and re-condensation dynamics in cuprates and conventional superconductors through their transient THz response [139–142].

However, light can in some cases *enhance* superconducting coherence rather than degrade it. In conventional superconductors near T_c , weak sub-gap microwave irradiation can sharpen the gap and modestly raise T_c , an effect first observed by Wyatt and Dayem and later described by Eliashberg [143–145]. This microwave-stimulated superconductivity is small, confined to a narrow window around T_c , and arises because light redistributes quasiparticles away from the gap edge near the Fermi level, enhancing the coherence of an already formed condensate. The more radical idea pursued over the past fifteen years is that intense mid-infrared driving may instead stimulate condensation far from equilibrium, starting from a normal metal and transiently producing superconducting correlations well above the equilibrium T_c .

Optical signatures of superconductivity. The optical properties of a superconductor are remarkable and provide a clear fingerprint for pump–probe experiments [146, 147]. Since pairing gaps typically lie in the meV range, the natural probe is time-resolved THz spectroscopy, which accesses these energy scales and measures the complex conductivity $\tilde{\sigma}(\omega, \Delta t) = \sigma_1(\omega, \Delta t) + i\sigma_2(\omega, \Delta t)$ as a function of pump–probe delay Δt .¹ At equilibrium, a normal metal is described by the Drude model,

$$\tilde{\sigma}(\omega) = \frac{\sigma_{\text{dc}}}{1 - i\omega\tau}, \quad \text{where} \quad \sigma_1(\omega) = \frac{\sigma_{\text{dc}}}{1 + \omega^2\tau^2}, \quad \text{and} \quad \sigma_2(\omega) = \frac{\sigma_{\text{dc}}\omega\tau}{1 + \omega^2\tau^2}. \quad (15)$$

A Drude metal displays a dissipative peak in σ_1 centered at zero frequency, while the reactive part vanishes, $\sigma_2(\omega) \rightarrow 0$ as $\omega \rightarrow 0$. As the superconducting state develops, the coherent part of this Drude weight collapses into a zero-frequency δ function in the real conductivity,

$$\sigma_1(\omega) = \pi \frac{n_s e^2}{m} \delta(\omega) + \sigma_1^{\text{reg}}(\omega), \quad (16)$$

where n_s is the superfluid density and the regular part σ_1^{reg} is gapped, with dissipative absorption suppressed below the pair-breaking edge 2Δ . The spectral weight removed from finite frequencies is conserved and transferred entirely to the zero-frequency δ function, as required by the Ferrell–Glover–Tinkham sum rule [147]. Through the Kramers–Kronig relations, this δ function in σ_1 enforces a purely inductive, divergent imaginary conductivity,

$$\sigma_2(\omega) = \frac{n_s e^2}{m\omega} = \frac{1}{\mu_0 \lambda_L^2 \omega} \propto \frac{1}{\omega}, \quad \lambda_L = \left(\frac{m}{\mu_0 n_s e^2} \right)^{1/2}, \quad (17)$$

where λ_L is the London penetration depth. The essential point is that this $1/\omega$ inductive response follows from the condensate itself, namely from the zero-frequency δ function in σ_1 , and not from the detailed gap structure. It is therefore universal and survives even for strongly anisotropic or nodal gaps, as in the d -wave cuprates, where σ_1^{reg} lacks a sharp gap edge. The experiments reviewed below search for this transient above- T_c signature in the THz response after resonant mid-infrared excitation, possibly accompanied by a gap-like suppression of σ_1 .

Cuprates. Building on structural control through nonlinear phononics (Section 3.3) and on the competition between charge order and pairing (Section 4.1), the first experiment in this direction sought to enhance superconductivity by suppressing a competing order in the driven state. Fausti *et al.* studied the stripe-ordered cuprate $\text{La}_{1.675}\text{Eu}_{0.2}\text{Sr}_{0.125}\text{CuO}_4$, a 1/8-hole-doped compound in which static charge stripes suppress bulk superconductivity [149]. Resonant excitation of an in-plane Cu–O stretching mode transiently induced an inductive response in $\sigma_2(\omega, \Delta t)$, together with a Josephson plasma edge in the c -axis reflectivity, the hallmark of three-dimensional interlayer superconducting coupling. A Josephson plasma resonance is the collective oscillation of the interlayer phase of the superconducting condensate, sustained by Cooper-pair tunneling between the CuO_2 planes. The results were interpreted as the revival of

¹There are subtleties in interpreting the instantaneous THz conductivity as a quasi-equilibrium optical conductivity, but they go beyond the scope of these lecture notes.

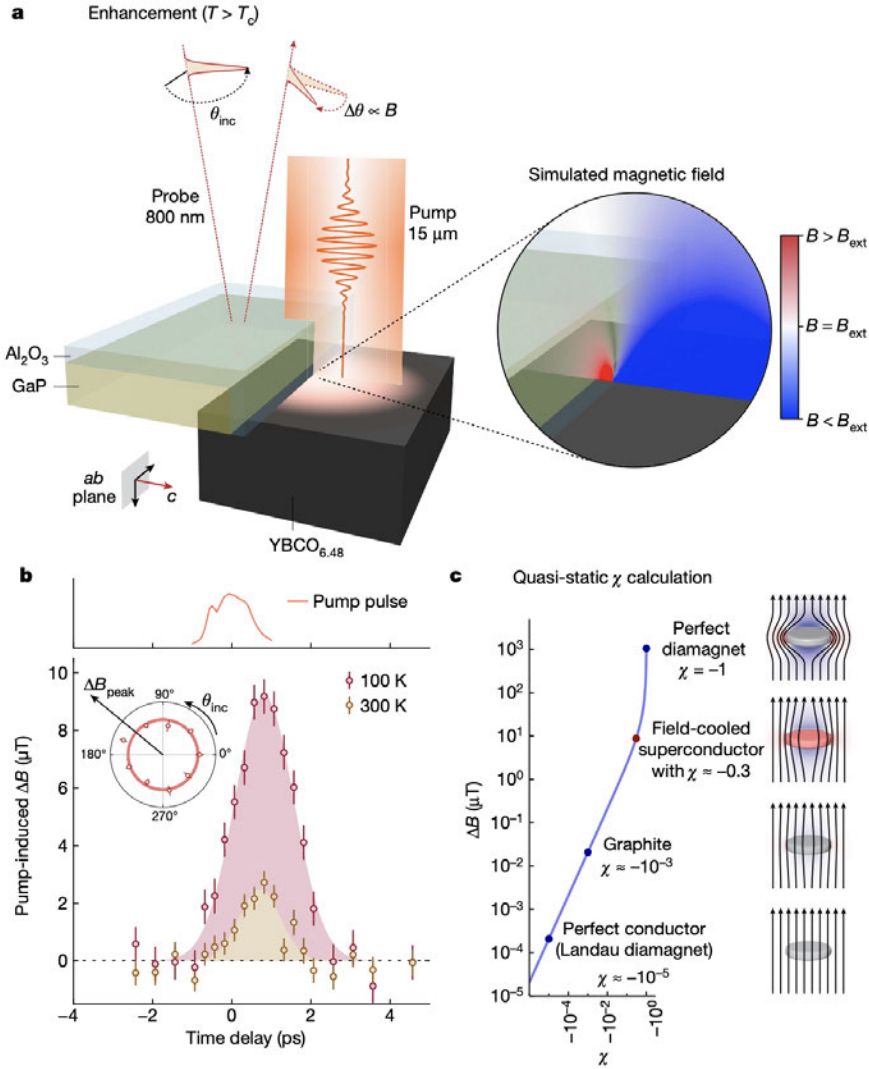


Fig. 13: Meissner effect in an optically-driven cuprate. **a.** Transient expulsion of an applied magnetic field from optically driven $\text{YBa}_2\text{Cu}_3\text{O}_{6.48}$, measured by ultrafast Faraday-rotation magnetometry after resonant excitation of the apical-oxygen phonon. **b.-c.** The diamagnetic screening response, observed far above the equilibrium T_c , corresponds to an inferred volume susceptibility $\chi \simeq -0.3$ and constitutes a direct magnetic fingerprint of a transient superconductor. Reprinted from Ref. [148], licensed under a Creative Commons Attribution 4.0 International License. Copyright (2024) by the authors.

latent superconducting fluctuations following pump-induced melting of the competing stripe-ordered state, in close analogy with the charge-order dynamics discussed in Section 4.1.

Following this early result, experiments then focused on bilayer cuprates $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ (YBCO), generally regarded as cleaner and less inhomogeneous copper oxides. These materials can be idealized as a stack of inequivalent Josephson junctions. The thin *intra*-bilayer and thick *inter*-bilayer junctions produce two longitudinal Josephson plasmons, a low-frequency inter-bilayer mode near 30 cm^{-1} and a high-frequency intra-bilayer mode near 475 cm^{-1} , together with a transverse plasmon in σ_1 whose weight tracks the c -axis superfluid density. Hu *et al.* reso-

nantly drove the apical-oxygen B_{1u} phonon at ~ 17 THz (~ 15 μm), a mode tied to a structural parameter that correlates with T_c [150], and found that light induces (1) a transient $1/\omega$ inductive response and (2) a *redistribution* of coherence between the two Josephson plasmons. The inter-bilayer mode blue-shifts and strengthens, while the intra-bilayer mode red-shifts, with the total coherent spectral weight approximately conserved [151, 152]. Notably, this enhancement persists far above T_c , up to temperatures coincident with the pseudogap line rather than with T_c , and appears even in underdoped samples with weaker charge-order correlations. These results suggest a connection to the pseudogap state and, possibly, a response of *preformed pairs* already present in the normal state. Femtosecond x-ray diffraction further constrained the accompanying lattice motion. By reconstructing the transient structure of driven YBCO from a subset of diffraction peaks, Mankowsky *et al.* identified a genuine lattice distortion induced via nonlinear-phononic coupling, but with displacements of only a picometer, or a few tenths of a percent of the relevant bond lengths [44]. Moreover, this distortion is smaller than, and opposite in sign to, the static structural difference between higher- and lower- T_c cuprates. These observations imply that structural rearrangements alone cannot explain the full enhancement of superconducting coherence, and suggest that the origin of this effect is mainly electronic.

In recent years, it has become clear that optical signatures, however suggestive, may not by themselves distinguish a genuine superconductor from a metal with anomalously high mobility. To resolve this fundamental ambiguity, it is necessary to perform ultrafast measurements of magnetic expulsion. Using ultrafast optical magnetometry, Fava *et al.* detected expulsion of an applied magnetic field from optically driven $\text{YBa}_2\text{Cu}_3\text{O}_{6.48}$, a transient Meissner effect with inferred volume susceptibility $\chi \simeq -0.3$ that persists up to room temperature [148] (Fig. 13). Because diamagnetic screening is the hallmark of a phase-coherent condensate and does not rely on model-dependent inversions of reflectivity data, this result provides the strongest evidence to date that light enhances superconducting coherence in these materials.

Fullerides. Signatures of light-enhanced superconductivity are not unique to the cuprates, and have been detected also in different families of organic metals. The most extensively characterized case is that of the alkali-doped fulleride K_3C_{60} , where the effect appears far from any obvious competing order and in a material with a more conventional equilibrium superconductivity. At equilibrium, K_3C_{60} is a superconductor ($T_c = 20$ K) with three narrow, half-filled bands derived from the C_{60} t_{1u} orbitals close to the Fermi level [153]. In this system, pairing is believed to arise from the cooperation of strong electronic correlations and coupling to intramolecular Jahn–Teller vibrations. In the simplest BCS mean-field form, the equilibrium transition temperature is given by $k_B T_c = 1.13 \hbar \omega_D e^{-\frac{1}{N(\varepsilon_F)V}}$, where ω_D is the phonon Debye frequency, $N(\varepsilon_F)$ is the density of states at the Fermi level, and V is the effective pairing interaction. T_c increases with both the characteristic phonon energy and, more dramatically, with the density of states and pairing interaction. A natural question is whether nonlinear phonon excitation can be conducive to an enhanced superconducting state by modulating any of these parameters. The infrared-active T_{1u} modes can be driven directly by a mid-infrared field but do not couple, by symmetry, to the t_{1u} electrons. The pairing modes are instead the Raman-

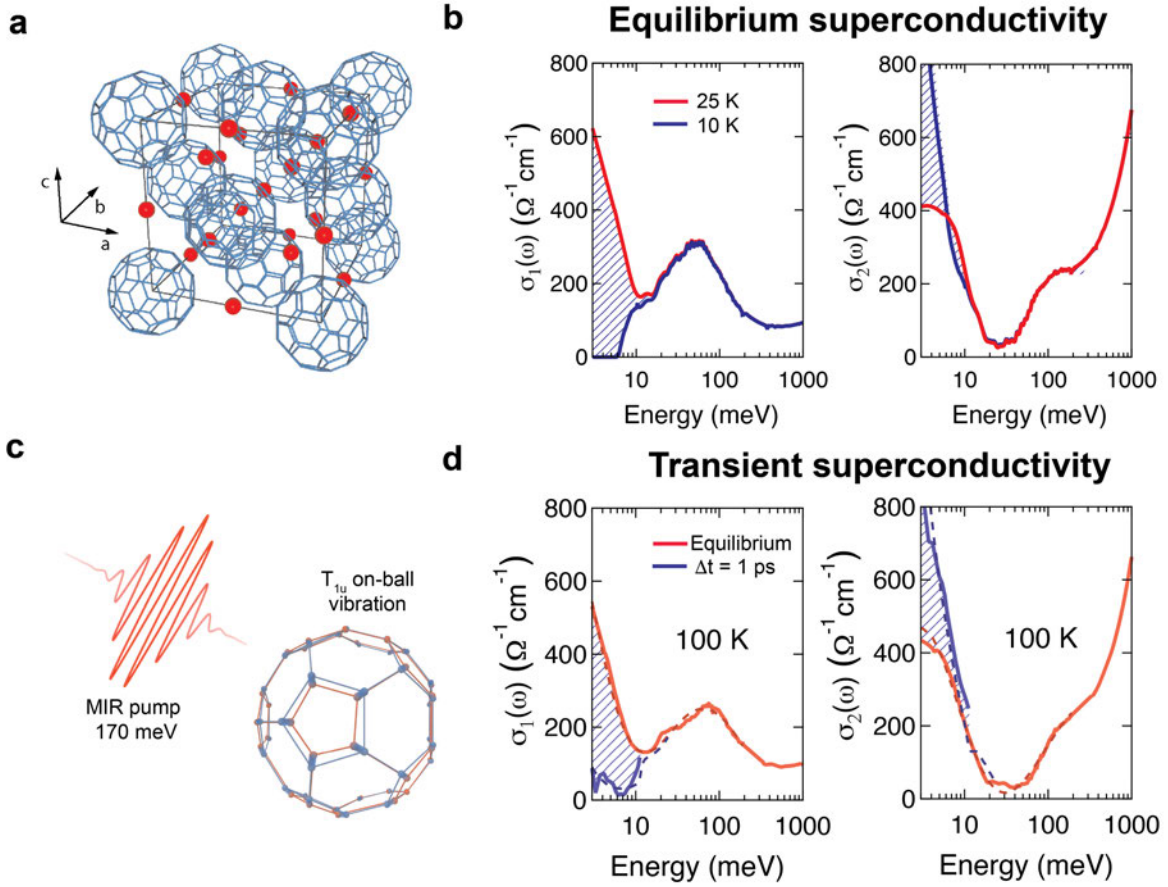


Fig. 14: Light-induced superconductivity in K_3C_{60} . **a.** K_3C_{60} unit cell. **b.** Equilibrium optical conductivity above and below the superconducting transition ($T_c = 20$ K). **c.** Resonant mid-infrared excitation of the T_{1u} molecular vibration. **d.** The terahertz response develops a gapped $\sigma_1(\omega)$ and a $1/\omega$ divergence in $\sigma_2(\omega)$ far above the equilibrium $T_c = 20$ K. Adapted with permission from Ref. [154]. Copyright (2016) by Springer Nature Limited.

active H_g Jahn–Teller distortions, which lift the orbital degeneracy of the molecular states and carry the strongest coupling. Mitrano *et al.* therefore excited the $T_{1u}(4)$ mode at $7 \mu\text{m}$ to drive quasi-static H_g distortion via lowest-order anharmonic coupling, $Q_{T_{1u}}^2 Q_{H_g}$ [154]. The transient THz response reproduces the gapped σ_1 and divergent σ_2 of the equilibrium superconductor at base temperatures up to ~ 100 K (Fig. 14), several times T_c , with the photoinduced response fading away at higher temperatures. While the photoinduced gap was slightly larger than that at equilibrium, this driven state showed both the key electrodynamic signatures of a BCS-like superconductor, suggesting a rather homogeneous response.

A series of measurements then tested whether this state follows the expected phenomenology of a BCS superconductor. Cantaluppi *et al.* applied hydrostatic pressure and found that the light-induced gap weakens under compression [155], mirroring the equilibrium superconductor, whose T_c decreases as the bandwidth grows and the density of states falls, and opposite to an ordinary metal, whose conductivity improves under pressure. Wang *et al.* probed the transient state with on-chip ultrafast transport and observed a current-dependent, critical-current-

like nonlinearity, characteristic of a superconductor and absent both in the equilibrium metal and under visible-light excitation [156]. Budden *et al.* showed that sustained driving converts the picosecond transient into a metastable state lasting more than ten nanoseconds, with a transport response compatible with vanishing resistance [157], echoing the metastability theme of Sections 4 and 5.3. Finally, Rowe *et al.* mapped the response versus drive frequency and found a pronounced resonant enhancement near 10 THz [158], identifying the spectral region where coupling to the pairing modes is most effective.

Organics. Similar optical responses have also been observed in layered BEDT-TTF-based organic conductors. Unlike the molecular fullerenes, the κ -(BEDT-TTF)₂X salts are quasi-two-dimensional and share several defining features with the cuprates. They derive from a half-filled band of (BEDT-TTF)₂ dimers on an anisotropic triangular lattice, lie near the boundary between an antiferromagnetic Mott insulator and a superconductor, and show evidence for a $d_{x^2-y^2}$ pairing gap, as in the cuprates [159]. Crucially, their position in the phase diagram can be tuned continuously. In κ -(BEDT-TTF)₂Cu[N(CN)₂]Br_xCl_{1-x}, substituting Cl with Br acts as a bandwidth, or chemical-pressure, knob that drives the system across a bandwidth-controlled Mott transition, from an antiferromagnetic insulator at $x = 0$ to a metallic superconductor at higher Br content [160]. This combination of two-dimensionality, cuprate-like competition between magnetism and pairing, and fine control of correlation strength makes the κ -organics a highly-tunable platform for testing light-induced superconductivity.

Buzzi *et al.* drove molecular C=C vibrations in κ -(BEDT-TTF)₂Cu[N(CN)₂]Br, the member of the series just on the metallic side of the Mott transition ($T_c = 12.5$ K), and observed a divergent $1/\omega$ response in σ_2 , and a gap opening in σ_1 above T_c [161]. The light-induced gap is much larger than, and qualitatively different from, the equilibrium superconducting gap, showing that the driven state is not simply the equilibrium condensate revived above T_c , but a distinct nonequilibrium state. By mapping the entire phase diagram of the family, Buzzi *et al.* also showed that the effect appears only near the Mott boundary, in the metallic compound with strong normal-state superconducting fluctuations, and is absent both in the Mott insulator and in the more itinerant metal farther from the transition [162]. Proximity to the Mott transition, and to the fluctuating precursor it sustains, therefore appears to be a prerequisite for the onset of light-induced superconducting coherence.

A general phenomenon. Taken together, these experiments show that resonant vibrational driving can induce transient superconducting-like responses in very different systems, ranging from layered cuprates to alkali fullerenes and organic charge-transfer salts. This phenomenon appears to be general across platforms, and a recurring pattern emerges. Light-induced or enhanced superconductivity is usually found to emerge from strongly fluctuating states near an equilibrium instability, including the stripe-ordered and pseudogap regimes of the cuprates, the Jahn–Teller-correlated metal of K₃C₆₀, and the Mott-proximate organics. This observation suggests that the drive amplifies or reorganizes pre-existing pairing tendencies, rather than creating superconductivity from an inert normal state or relying on structural distortions unique

to a given material. Uemura framed this viewpoint in terms of the dynamical response of a phase-incoherent bosonic fluid [163]. The next section reviews the main theoretical scenarios proposed to explain the microscopic origin of this phenomenon.

6.2 Theoretical scenarios

The microscopic mechanism of light-induced superconductivity, and even the precise nature of the transient state, remain open questions. No single theory accounts for all the experimental observations, but the existing proposals fall into a few broad classes.

Lattice-mediated mechanisms. The largest class of theories builds on the nonlinear phononics framework of Section 3.3. The starting point is a resonantly-driven infrared phonon Q_{IR} which rectifies into a quasi-static Raman displacement, $\langle Q_R \rangle \simeq -g \langle Q_{\text{IR}}^2 \rangle / \Omega_R^2$, thereby shifting the structural parameters that control the pairing scale [44]. Pairing can then be enhanced in different ways. In the picture of Sentef *et al.*, the distortion narrows the electronic band, reducing the hopping t , increasing the density of states $\bar{N}(E_F)$, and thereby enhancing the dimensionless electron–phonon coupling

$$\lambda = 2 \bar{N}(E_F) \int_0^\infty \frac{\alpha^2 F(\Omega)}{\Omega} d\Omega, \quad (18)$$

to which the BCS pairing scale is exponentially sensitive [164]. Kennes *et al.* considered instead a *quadratic* electron-phonon coupling, $gK \sum_i \hat{n}_i \hat{x}_i^2$, for which the local phonon frequency depends on the electronic occupation; exciting the phonon to a mean population n_B then renormalizes the effective on-site interaction,

$$U^* = U - (n_B + \tfrac{1}{2}) \omega_0 \left(2\sqrt{1+2g} - 1 - \sqrt{1+4g} \right), \quad (19)$$

where the bracket is positive for either sign of g , so that driving the phonon ($n_B > 0$) lowers U^* and strengthens the attraction, a purely electronic “squeezing” of the interaction [165]. A third route relies instead on a parametric process. When a coupling is modulated periodically in time, the Cooper-pair fluctuation amplitude obeys, schematically, a Mathieu-type equation

$$\ddot{\Delta}_{\mathbf{q}} + \omega_{\mathbf{q}}^2 [1 + \alpha \cos(\Omega t)] \Delta_{\mathbf{q}} = 0, \quad (20)$$

so that for $\Omega \simeq 2\omega_{\mathbf{q}}$ the pairing susceptibility is parametrically amplified, just as periodically stiffening a pendulum amplifies its swing, without a true equilibrium transition [166]. Related treatments trace the enhancement to nonequilibrium phonon populations that increase the effective coupling [167] or to the transient cooling of the electronic subsystem by mid-infrared absorption [168].

Electronic scenarios. A second class of mechanisms dispenses with the lattice altogether. In the Hubbard model the staggered pair operator

$$\hat{\eta}^+ = \sum_j (-1)^j \hat{c}_{j\uparrow}^\dagger \hat{c}_{j\downarrow}^\dagger \quad (21)$$

is an eigenoperator of the Hamiltonian on a bipartite lattice, $[\hat{H}, \hat{\eta}^+] = U \hat{\eta}^+$, so the states $(\hat{\eta}^+)^N |0\rangle$ are exact eigenstates with off-diagonal long-range order,

$$\langle \hat{\Delta}_i^\dagger \hat{\Delta}_j \rangle \xrightarrow{|i-j| \rightarrow \infty} \text{const} \times e^{i\pi(R_i - R_j)}, \quad \hat{\Delta}_i = \hat{c}_{i\uparrow} \hat{c}_{i\downarrow}, \quad (22)$$

a condensate of pairs carrying center-of-mass momentum $Q = \pi$ [169]. The open question is how to populate these states out of equilibrium. Kaneko *et al.* showed that an optical pulse, entering as a Peierls phase on the hopping, connects η -sectors through the current operator with the selection rule $\Delta\eta = \pm 1$ at an energy cost $\sim U$, so that an intrinsically nonlinear, multi-photon process tuned to $\omega_{\text{pump}} \sim U$ climbs a ladder of η -pairing states [170]. Tindall *et al.* found a complementary route in which heating the spin sector toward infinite temperature, while $\hat{\eta}^+ \hat{\eta}^-$ is conserved, forces the charge sector into a steady state with long-range η order [171, 172]. These mechanisms are appealing because they are purely electronic and symmetry-based, in the spirit of the Hamiltonian-engineering route of Section 5.3. However, we need to note that the generated pairs are finite-momentum, a $Q = \pi$ pair-density wave rather than a uniform condensate, and the results rest on the idealized single-band Hubbard model, so their connection to the measured THz response of real solids remains indirect.

Amplification of bosonic excitations or superconducting fluctuations. A particularly compelling proposal, due to Dai and Lee, occupies the middle ground between creating pairs and amplifying an existing condensate [173, 174]. It supposes that charge- $2e$ bosons made of electron pairs are already present in the normal state but remain *gapped* and uncondensed, with dispersion $E_k = \sqrt{\Delta^2 + v^2 k^2}$. These bosons could be quantum-fluctuating pairs of the cuprate pseudogap, or the doublon-holon excitations of certain nearly Mott-localized metals. A drive that periodically modulates the boson gap,

$$\mathcal{L} = \frac{1}{2} |\partial_t \phi|^2 - \frac{v^2}{2} |\nabla \phi|^2 - \frac{1}{2} \Delta^2 |\phi|^2 - \frac{U}{8} |\phi|^4 - \lambda \cos(\Omega t) |\phi|^2, \quad (23)$$

then acts as a parametric pump on these bosons. Diagonalizing the driven problem gives complex eigenfrequencies on a resonant ring of momenta,

$$\omega = \frac{\Omega}{2} \pm i \theta_k, \quad \theta_k = \sqrt{\left(\frac{\lambda}{2E_k} \right)^2 - \left(\frac{\Omega}{2} - E_k \right)^2}, \quad (24)$$

so that when the drive satisfies $\Omega \simeq 2E_k$ (hence $\Omega \gtrsim 2\Delta$), boson particle-hole pairs at $\pm k$ are created out of the zero-point fluctuations and grow exponentially. Because of Bose statistics the instability opens not in energy, as for driven fermions, but along a ring in momentum space, building a coherent, two-mode-squeezed “particle-hole condensate.” Its oppositely charged components drift in opposite directions under a field, producing precisely the inductive response $\sigma(\omega) = i e^{*2} \rho_s / \omega$ of a superconductor, with a transient superfluid density ρ_s that grows and then saturates, even though no equilibrium condensate exists, and, because condensation occurs on a ring rather than at a point, without a Meissner effect. In the variant aimed at the organics, the relevant bosons are the fractionalized charge- e chargons of a metal close to a

Mott transition, and the drive transiently tips the system across that transition; the light-induced “gap” is then reinterpreted as a photoinduced *Mott* gap, a picture that rationalizes why the response in κ -(BEDT-TTF)₂Cu[N(CN)₂]Br appears only near the transition and is larger than, and distinct from, the equilibrium superconducting gap. Other theoretical scenarios propose amplifying the collective response of pre-existing superconducting fluctuations. A relevant example is the parametric driving of the interlayer Josephson plasmon in a bilayer cuprate, analyzed by Michael *et al.* [175]. The plasma frequency is set by the interlayer Josephson coupling, $\omega_{\text{JP}} \simeq \sqrt{j_c d^2 (e^*)^2 / \epsilon}$, and the resonantly driven apical-oxygen phonon modulates the in-plane superfluid stiffness, introducing a time-periodic term that acts as a parametric drive on the plasmon, exactly the modulation in the Mathieu equation. When the drive frequency matches the sum of two plasmon branches, $\omega_{\text{ph}} = \omega_1 + \omega_2$, the mode amplitudes grow exponentially at a rate $\text{Im } \omega = |g \xi Q_0|$, and an overdamped plasma edge that is otherwise invisible in equilibrium re-emerges in the transient reflectivity. The result mimics an enhanced superfluid response in $\sigma_2(\omega)$, and reproduces the measured THz signatures, without an increase in T_c or in the number of Cooper pairs.

7 Outlook

The preceding sections offer a necessarily incomplete view of how light control of many-body states has evolved over the past two decades. We close by highlighting three frontiers likely to shape its next stage.

The first is the convergence of ultrafast coherent control with the language of quantum control developed in quantum information science. The quantum control of Hubbard excitons discussed in Section 5.2 illustrates that a single uniaxial $\pi/2$ rotation driven by a mid-infrared field already constitutes the elementary building block of a programmable control sequence [134]. Pulse shaping should extend this approach to Ramsey interferometry, dynamical decoupling, and ultimately full SU(2) control of correlated states [135, 137]. The broader ambition is to treat selected collective excitations of a quantum material as qubit-like degrees of freedom that can be initialized, rotated, entangled, and read out coherently before dissipation takes over. In valleytronics, the coherent encoding and switching of valley information in a WS₂ monolayer at room temperature has already shown that logic operations can be written into a solid on ultrafast timescales [176].

A second frontier is the move from bulk crystals to mesoscale structures and devices. Ultrafast transport, as illustrated by the on-chip detection of a light-induced anomalous Hall effect in graphene [53], measures nonequilibrium currents directly and probes the functional response of a driven material without relying only on the inversion of optical spectra. The next step is to bring this approach to patterned and gated structures, where lithographically defined channels can be used to contact, address, and integrate light-engineered states. Van der Waals materials and heterostructures are especially well suited to this program. They are atomically thin, electrostatically tunable, and freely stackable, allowing driving, gating, and transport readout to be combined in a single device, while interactions are tuned *in situ* across phase boundaries.

The third, and conceptually most far-reaching, frontier is the control of quantum materials through their electromagnetic environment, by embedding them in optical cavities [177]. Cavity-mediated control of the metal-to-insulator transition in 1T-TaS₂ [178], and cavity-induced, exciton-like attraction observed in bilayer graphene ultrastrongly coupled to a terahertz resonator [179] show that the electromagnetic environment alone can reorganize electronic response. Extending this idea to superconductivity is a compelling emergent direction, with first reports of a cavity-enhanced superconducting response in an underdoped cuprate [180] and of an enhanced transition temperature in few-layer NbSe₂ tuned to a phononic cavity mode [181]. These results connect cavity engineering directly to the light-induced superconducting responses discussed in Section 6.

A crucial conceptual distinction in the possible implementations of cavity experiments is that between *driven* and *dark* cavities. In the former, light still injects energy into the material, even if the cavity reshapes the electromagnetic modes through which this occurs. In the latter, no external photons are involved, and the material is solely modified by the vacuum fluctuations of the confined field, enhanced at the cavity resonances. The breakdown of topological protection in the integer quantum Hall effect induced by cavity vacuum fields in a Hall-bar geometry [182], together with the recent observation of a cavity-altered superconducting ground state at a hexagonal boron nitride interface [183], shows that an electromagnetic environment can reshape a many-body state without a drive. This development addresses a central limitation of many driven protocols reviewed in this chapter, namely the heating and dissipation that accompany continuous excitation and limit the lifetime of the engineered state. A dark cavity does not push the system into a transient excited configuration, rather it modifies the *ground state* itself. Whether vacuum fields can be made strong enough to engineer correlated ground states at will remains an open question. But the possibility of equilibrium, dissipation-free control of many-body states is, in our view, one of the most exciting prospects for the field.

Acknowledgments

Part of the work described in the manuscript is the result of the author's collaboration with many researchers listed in the references below. The author is especially grateful to Michele Buzzi, Ankit Disa, Michael Först, Anshul Kogar, Roman Mankowsky, Michael A. Sentef, and Mariano Trigo for discussions and input on the manuscript, as well as to Brad Baxley for assistance with graphics.

References

- [1] J. Zhang and R. Averitt, *Ann. Rev. Mater. Res.* **44**, 19 (2014)
- [2] C. Giannetti, M. Capone, D. Fausti, M. Fabrizio, F. Parmigiani, and D. Mihailovic, *Adv. Phys.* **65**, 58 (2016)
- [3] D.N. Basov, R.D. Averitt, and D. Hsieh, *Nat. Mater.* **16**, 1077 (2017)
- [4] A. de la Torre, D.M. Kennes, M. Claassen, S. Gerber, J.W. McIver, and M.A. Sentef, *Rev. Mod. Phys.* **93**, 041002 (2021)
- [5] J. Bloch, A. Cavalleri, V. Galitski, M. Hafezi, and A. Rubio, *Nature* **606**, 41 (2022)
- [6] K. Nasu: *Photoinduced Phase Transitions* (World Scientific, 2004)
- [7] R.D. Averitt and A.J. Taylor, *J. Phys.: Condens. Matter* **14**, R1357 (2002)
- [8] B. Keimer and J.E. Moore, *Nat. Phys.* **13**, 1045 (2017)
- [9] G. Steinmeyer, D.H. Sutter, L. Gallmann, N. Matuschek, and U. Keller, *Science* **286**, 1507 (1999)
- [10] R.A. Bartels, A. Paul, H. Green, H.C. Kapteyn, M.M. Murnane, S. Backus, I.P. Christov, Y. Liu, D. Attwood, and C. Jacobsen, *Science* **297**, 376 (2002)
- [11] G. Cerullo and S. De Silvestri, *Rev. Sci. Instr.* **74**, 1 (2003)
- [12] K. Reimann, *Rep. Prog. Phys.* **70**, 1597 (2007)
- [13] B. Liu, H. Bromberger, A. Cartella, T. Gebert, M. Först, and A. Cavalleri, *Opt. Lett.* **42**, 129 (2017)
- [14] J. Orenstein, *Physics Today* **65**, 44 (2012)
- [15] R.W. Schoenlein, W.Z. Lin, J.G. Fujimoto, and G.L. Eesley, *Phys. Rev. Lett.* **58**, 1680 (1987)
- [16] S. Koshihara, Y. Tokura, T. Mitani, G. Saito, and T. Koda, *Phys. Rev. B* **42**, 6853(R) (1990)
- [17] S.D. Brorson, A. Kazeroonian, J.S. Moodera, D.W. Face, T.K. Cheng, E.P. Ippen, M.S. Dresselhaus, and G. Dresselhaus, *Phys. Rev. Lett.* **64**, 2172 (1990)
- [18] J. Demsar, B. Podobnik, V.V. Kabanov, T. Wolf, and D. Mihailovic, *Phys. Rev. Lett.* **82**, 4918 (1999)
- [19] D.S. Chemla and J. Shah, *Proc. Natl. Acad. Sci. USA* **97**, 2437 (2000)

- [20] A. Kirilyuk, A.V. Kimel, and T. Rasing, *Rev. Mod. Phys.* **82**, 2731 (2010)
- [21] S.L. Dexheimer (Ed.): *Terahertz Spectroscopy: Principles and Applications* (CRC Press, 2008), 1st ed.
- [22] S. Ghimire and D.A. Reis, *Nat. Phys.* **15**, 10 (2019)
- [23] L. Zhao, D. Torchinsky, J. Harter, A. de la Torre, and D. Hsieh, in B.D. Guenther and D.G. Steel (eds.): *Encyclopedia of Modern Optics (2nd Ed.)* (Elsevier, 2018), p. 207
- [24] M. Fiebig, K. Miyano, Y. Tomioka, and Y. Tokura, *Science* **280**, 1925 (1998)
- [25] T.F. Nova, A.S. Disa, M. Fechner, and A. Cavalleri, *Science* **364**, 1075 (2019)
- [26] X. Zhang, T. Carbin, A.B. Culver, K. Du, K. Wang, S.-W. Cheong, R. Roy, and A. Kogar, *Nat. Mater.* **23**, 790 (2024)
- [27] F. Boschini, M. Zonno, and A. Damascelli, *Rev. Mod. Phys.* **96**, 015003 (2024)
- [28] I. Gierz, J.C. Petersen, M. Mitrano, C. Cacho, I.C.E. Turcu, E. Springate, A. Stöhr, A. Köhler, U. Starke, and A. Cavalleri, *Nat. Mater.* **12**, 1119 (2013)
- [29] U. Höfer, I.L. Shumay, C. Reuß, U. Thomann, W. Wallauer, and T. Fauster, *Science* **277**, 1480 (1997)
- [30] S. Gerber, S.-L. Yang, D. Zhu, H. Soifer, J.A. Sobota, S. Rebec, J.J. Lee, T. Jia, B. Moritz, C. Jia, A. Gauthier, Y. Li, D. Leuenberger, Y. Zhang, L. Chaix, W. Li, H. Jang, J.-S. Lee, M. Yi, G.L. Dakovski, S. Song, J.M. Glowina, S. Nelson, K.W. Kim, Y.-D. Chuang, Z. Hussain, R.G. Moore, T.P. Devereaux, W.-S. Lee, P.S. Kirchmann, and Z.-X. Shen, *Science* **357**, 71 (2017)
- [31] C.L. Smallwood, J.P. Hinton, C. Jozwiak, W. Zhang, J.D. Koralek, H. Eisaki, D.-H. Lee, J. Orenstein, and A. Lanzara, *Science* **336**, 1137 (2012)
- [32] T. Rohwer, S. Hellmann, M. Wiesenmayer, C. Sohrt, A. Stange, B. Slomski, A. Carr, Y. Liu, L.M. Avila, M. Kalläne, S. Mathias, L. Kipp, K. Rossnagel, and M. Bauer, *Nature* **471**, 490 (2011)
- [33] J.C. Petersen, S. Kaiser, N. Dean, A. Simoncig, H.Y. Liu, A.L. Cavalieri, C. Cacho, I.C.E. Turcu, E. Springate, F. Frassetto, L. Poletto, S.S. Dhesi, H. Berger, and A. Cavalleri, *Phys. Rev. Lett.* **107**, 177402 (2011)
- [34] Y.H. Wang, H. Steinberg, P. Jarillo-Herrero, and N. Gedik, *Science* **342**, 453 (2013)
- [35] F. Mahmood, C. Chan, Z. Alpichshev, D. Gardner, Y. Lee, and N. Gedik, *Nat. Phys.* **12**, 306 (2016)

- [36] S. Ito, M. Schüler, M. Meierhofer, S. Schlauderer, J. Freudenstein, J. Reimann, D. Afanasiev, K.A. Kokh, O.E. Tereshchenko, J. Gädde, M.A. Sentef, U. Höfer, and R. Huber, *Nature* **616**, 696 (2023)
- [37] N. Bielinski, R. Chari, J. May-Mann, S. Kim, J. Zwettler, Y. Deng, A. Aishwarya, S. Roychowdhury, C. Shekhar, M. Hashimoto, D. Lu, J. Yan, C. Felser, V. Madhavan, Z.-X. Shen, T.L. Hughes, and F. Mahmood, *Nat. Phys.* **21**, 458 (2025)
- [38] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H.-D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, and J. Galayda, *Nat. Photonics* **4**, 641 (2010)
- [39] S.P. Weathersby, G. Brown, M. Centurion, T.F. Chase, R. Coffee, J. Corbett, J.P. Eichner, J.C. Frisch, A.R. Fry, M. Gühr, N. Hartmann, C. Hast, R. Hettel, R.K. Jobe, E.N. Jongewaard, J.R. Lewandowski, R.K. Li, A.M. Lindenberg, I. Makasyuk, J.E. May, D. McCormick, M.N. Nguyen, A.H. Reid, X. Shen, K. Sokolowski-Tinten, T. Vecchione, S.L. Vetter, J. Wu, J. Yang, H.A. Dürr, and X.J. Wang, *Rev. Sci. Instr.* **86**, 073702 (2015)
- [40] K. Sokolowski-Tinten, C. Blome, J. Blums, A. Cavalleri, C. Dietrich, A. Tarasevitch, I. Uschmann, E. Förster, M. Kammler, M. Horn-von Hoegen, and D. von der Linde, *Nature* **422**, 287 (2003)
- [41] D.M. Fritz, D.A. Reis, B. Adams, R.A. Akre, J. Arthur, C. Blome, P.H. Bucksbaum, A.L. Cavalieri, S. Engemann, S. Fahy, R.W. Falcone, P.H. Fuoss, K.J. Gaffney, M.J. George, J. Hajdu, M.P. Hertlein, P.B. Hillyard, M.H. von Hoegen, M. Kammler, J. Kaspar, R. Kienberger, P. Krejcik, S.H. Lee, A.M. Lindenberg, B. McFarland, D. Meyer, T. Montagne, E.D. Murray, A.J. Nelson, M. Nicoul, R. Pahl, J. Rudati, H. Schlarb, D.P. Siddons, K. Sokolowski-Tinten, T. Tschentscher, D. von der Linde, and J.B. Hastings, *Science* **315**, 633 (2007)
- [42] P. Baum, D.-S. Yang, and A.H. Zewail, *Science* **318**, 788 (2007)
- [43] P. Beaud, A. Caviezel, S.O. Mariager, L. Rettig, G. Ingold, C. Dornes, S.-W. Huang, J.A. Johnson, M. Radovic, T. Huber, T. Kubacka, A. Ferrer, H.T. Lemke, M. Chollet, D. Zhu, J.M. Glownia, M. Sikorski, A. Robert, H. Wadati, M. Nakamura, M. Kawasaki, Y. Tokura, S.L. Johnson, and U. Staub, *Nat. Mater.* **13**, 923 (2014)
- [44] R. Mankowsky, A. Subedi, M. Först, S.O. Mariager, M. Chollet, H.T. Lemke, J.S. Robinson, J.M. Glownia, M.P. Minitti, A. Frano, M. Fechner, N.A. Spaldin, T. Loew, B. Keimer, A. Georges, and A. Cavalleri, *Nature* **516**, 71 (2014)

- [45] S. Wall, S. Yang, L. Vidas, M. Chollet, J.M. Glowina, M. Kozina, T. Katayama, T. Henighan, M. Jiang, T.A. Miller, D.A. Reis, L.A. Boatner, O. Delaire, and M. Trigo, *Science* **362**, 572 (2018)
- [46] S. Vogelgesang, G. Storeck, J.G. Horstmann, T. Diekmann, M. Siviš, S. Schramm, K. Rossnagel, S. Schäfer, and C. Ropers, *Nat. Phys.* **14**, 184 (2018)
- [47] A. Kogar, A. Zong, P.E. Dolgirev, X. Shen, J. Straquadine, Y.-Q. Bie, X. Wang, T. Rohwer, I.-C. Tung, Y. Yang, R. Li, J. Yang, S. Weathersby, S. Park, M.E. Kozina, E.J. Sie, H. Wen, P. Jarillo-Herrero, I.R. Fisher, X. Wang, and N. Gedik, *Nat. Phys.* **16**, 159 (2020)
- [48] M.P.M. Dean, Y. Cao, X. Liu, S. Wall, D. Zhu, R. Mankowsky, V. Thampy, X.M. Chen, J.G. Vale, D. Casa, J. Kim, A.H. Said, P. Juhas, R. Alonso-Mori, J.M. Glowina, A. Robert, J. Robinson, M. Sikorski, S. Song, M. Kozina, H. Lemke, L. Patthey, S. Owada, T. Katayama, M. Yabashi, Y. Tanaka, T. Togashi, J. Liu, C. Rayan Serrao, B.J. Kim, L. Huber, C.-L. Chang, D.F. McMorro, M. Först, and J.P. Hill, *Nat. Mater.* **15**, 601 (2016)
- [49] H. Padma, F. Glerean, S.F.R. TenHuisen, Z. Shen, H. Wang, L. Xu, J.D. Elliott, C.C. Homes, E. Skoropata, H. Ueda, B. Liu, E. Paris, A. Romaguera, B. Lee, W. He, Y. Wang, S.H. Lee, H. Choi, S.-Y. Park, Z. Mao, M. Calandra, H. Jang, E. Razzoli, M.P.M. Dean, Y. Wang, and M. Mitrano, *Nat. Mater.* **24**, 1584 (2025)
- [50] T.L. Cocker, V. Jelic, M. Gupta, S.J. Molesky, J.A.J. Burgess, G. De Los Reyes, L.V. Titova, Y.Y. Tsui, M.R. Freeman, and F.A. Hegmann, *Nat. Photonics* **7**, 620 (2013)
- [51] T.L. Cocker, D. Peller, P. Yu, J. Repp, and R. Huber, *Nature* **539**, 263 (2016)
- [52] M. Wagner, Z. Fei, A.S. McLeod, A.S. Rodin, W. Bao, E.G. Iwinski, Z. Zhao, M. Goldflam, M. Liu, G. Dominguez, M. Thiemens, M.M. Fogler, A.H. Castro Neto, C.N. Lau, S. Amarie, F. Keilmann, and D.N. Basov, *Nano Lett.* **14**, 894 (2014)
- [53] J.W. McIver, B. Schulte, F.-U. Stein, T. Matsuyama, G. Jotzu, G. Meier, and A. Cavalleri, *Nat. Phys.* **16**, 38 (2020)
- [54] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, and P. Werner, *Rev. Mod. Phys.* **86**, 779 (2014)
- [55] Y. Murakami, D. Golež, M. Eckstein, and P. Werner, *Rev. Mod. Phys.* **97**, 035001 (2025)
- [56] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [57] D.I. Khomskii: *Transition Metal Compounds* (Cambridge University Press, 2014)
- [58] J. Zaanen, G.A. Sawatzky, and J.W. Allen, *Phys. Rev. Lett.* **55**, 418 (1985)

- [59] H. Okamoto, H. Matsuzaki, T. Wakabayashi, Y. Takahashi, and T. Hasegawa, *Phys. Rev. Lett.* **98**, 037401 (2007)
- [60] S. Wall, D. Brida, S.R. Clark, H.P. Ehrke, D. Jaksch, A. Ardavan, S. Bonora, H. Uemura, Y. Takahashi, T. Hasegawa, H. Okamoto, G. Cerullo, and A. Cavalleri, *Nat. Phys.* **7**, 114 (2011)
- [61] M. Mitrano, G. Cotugno, S.R. Clark, R. Singla, S. Kaiser, J. Stähler, R. Beyer, M. Dressel, L. Baldassarre, D. Nicoletti, A. Perucchi, T. Hasegawa, H. Okamoto, D. Jaksch, and A. Cavalleri, *Phys. Rev. Lett.* **112**, 117801 (2014)
- [62] T. Ishikawa, Y. Sagae, Y. Naitoh, Y. Kawakami, H. Itoh, K. Yamamoto, K. Yakushi, H. Kishida, T. Sasaki, S. Ishihara, Y. Tanaka, K. Yonemitsu, and S. Iwai, *Nat. Commun.* **5**, 5528 (2014)
- [63] H. Okamoto, T. Miyagoe, K. Kobayashi, H. Uemura, H. Nishioka, H. Matsuzaki, A. Sawa, and Y. Tokura, *Phys. Rev. B* **83**, 125102 (2011)
- [64] E. Baldini, M.A. Sentef, S. Acharya, T. Brumme, E. Sheveleva, F. Lyzwa, E. Pomjakushina, C. Bernhard, M. van Schilfgaarde, F. Carbone, A. Rubio, and C. Weber, *Proc. Natl. Acad. Sci. USA* **117**, 6409 (2020)
- [65] F.J. Morin, *Phys. Rev. Lett.* **3**, 34 (1959)
- [66] A. Cavalleri, C. Tóth, C.W. Siders, J.A. Squier, F. Ráksi, P. Forget, and J.C. Kieffer, *Phys. Rev. Lett.* **87**, 237401 (2001)
- [67] S. Wall, D. Wegkamp, L. Foglia, K. Appavoo, J. Nag, J. Haglund, Jr., J. Stähler, and M. Wolf, *Nat. Commun.* **3**, 721 (2012)
- [68] V.R. Morrison, R.P. Chatelain, K.L. Tiwari, A. Hendaoui, A. Bruhács, M. Chaker, and B.J. Siwick, *Science* **346**, 445 (2014)
- [69] H.J. Zeiger, J. Vidal, T.K. Cheng, E.P. Ippen, G. Dresselhaus, and M.S. Dresselhaus, *Phys. Rev. B* **45**, 768 (1992)
- [70] R. Merlin, *Solid State Commun.* **102**, 207 (1997)
- [71] A.S. Disa, T.F. Nova, and A. Cavalleri, *Nat. Phys.* **17**, 1087 (2021)
- [72] M. Rini, R. Tobey, N. Dean, J. Itatani, Y. Tomioka, Y. Tokura, R.W. Schoenlein, and A. Cavalleri, *Nature* **449**, 72 (2007)
- [73] M. Först, C. Manzoni, S. Kaiser, Y. Tomioka, Y. Tokura, R. Merlin, and A. Cavalleri, *Nat. Phys.* **7**, 854 (2011)
- [74] A. Subedi, A. Cavalleri, and A. Georges, *Phys. Rev. B* **89**, 220301 (2014)

- [75] A.D. Caviglia, R. Scherwitzl, P. Popovich, W. Hu, H. Bromberger, R. Singla, M. Mitrano, M.C. Hoffmann, S. Kaiser, P. Zubko, S. Gariglio, J.-M. Triscone, M. Först, and A. Cavalleri, *Phys. Rev. Lett.* **108**, 136801 (2012)
- [76] G. Grüner: *Density Waves in Solids* (CRC Press, 1994), 1st ed.
- [77] S. Hellmann, T. Rohwer, M. Kalläne, K. Hanff, C. Sohrt, A. Stange, A. Carr, M.M. Murnane, H.C. Kapteyn, L. Kipp, M. Bauer, and K. Rossnagel, *Nat. Commun.* **3**, 1069 (2012)
- [78] T. Huber, S.O. Mariager, A. Ferrer, H. Schäfer, J.A. Johnson, S. Grübel, A. Lübcke, L. Huber, T. Kubacka, C. Dornes, C. Laulhe, S. Ravy, G. Ingold, P. Beaud, J. Demsar, and S.L. Johnson, *Phys. Rev. Lett.* **113**, 026401 (2014)
- [79] H.Y. Liu, I. Gierz, J.C. Petersen, S. Kaiser, A. Simoncig, A.L. Cavalieri, C. Cacho, I.C.E. Turcu, E. Springate, F. Frassetto, L. Poletto, S.S. Dhesi, Z.-A. Xu, T. Cuk, R. Merlin, and A. Cavalleri, *Phys. Rev. B* **88**, 045104 (2013)
- [80] L. Stojchevska, I. Vaskivskyi, T. Mertelj, P. Kusar, D. Svetin, S. Brazovskii, and D. Mihailovic, *Science* **344**, 177 (2014)
- [81] M. Mitrano, S. Lee, A.A. Husain, L. Delacretaz, M. Zhu, G. de la Peña Munoz, S.X.-L. Sun, Y.I. Joe, A.H. Reid, S.F. Wandel, G. Coslovich, W. Schlotter, T. van Driel, J. Schneeloch, G.D. Gu, S. Hartnoll, N. Goldenfeld, and P. Abbamonte, *Sci. Adv.* **5**, eaax3346 (2019)
- [82] B. Keimer, S.A. Kivelson, M.R. Norman, S. Uchida, and J. Zaanen, *Nature* **518**, 179 (2015)
- [83] R. Comin and A. Damascelli, *Ann. Rev. Condens. Matter Phys.* **7**, 369 (2016)
- [84] D.F. Agterberg, J.S. Davis, S.D. Edkins, E. Fradkin, D.J. Van Harlingen, S.A. Kivelson, P.A. Lee, L. Radzihovsky, J.M. Tranquada, and Y. Wang, *Ann. Rev. Condens. Matter Phys.* **11**, 231 (2020)
- [85] M. Först, R.I. Tobey, H. Bromberger, S.B. Wilkins, V. Khanna, A.D. Caviglia, Y.-D. Chuang, W.S. Lee, W.F. Schlotter, J.J. Turner, M.P. Minitti, O. Krupin, Z.J. Xu, J.S. Wen, G.D. Gu, S.S. Dhesi, A. Cavalleri, and J.P. Hill, *Phys. Rev. Lett.* **112**, 157002 (2014)
- [86] M. Mitrano, S. Lee, A.A. Husain, M. Zhu, G. de la Peña Muñoz, S.X.-L. Sun, Y.I. Joe, A.H. Reid, S.F. Wandel, G. Coslovich, W. Schlotter, T. van Driel, J. Schneeloch, G.D. Gu, N. Goldenfeld, and P. Abbamonte, *Phys. Rev. B* **100**, 205125 (2019)
- [87] S. Wandel, F. Boschini, E.H. da Silva Neto, L. Shen, M.X. Na, S. Zohar, Y. Wang, S.B. Welch, M.H. Seaberg, J.D. Koralek, G.L. Dakovski, W. Hettel, M.-F. Lin, S.P. Moeller, W.F. Schlotter, A.H. Reid, M.P. Minitti, T. Boyle, F. He, R. Sutarto,

- R. Liang, D. Bonn, W. Hardy, R.A. Kaindl, D.G. Hawthorn, J.-S. Lee, A.F. Kemper, A. Damascelli, C. Giannetti, J.J. Turner, and G. Coslovich, *Science* **376**, 860 (2022)
- [88] E. Beaurepaire, J.-C. Merle, A. Daunois, and J.-Y. Bigot, *Phys. Rev. Lett.* **76**, 4250 (1996)
- [89] A.V. Kimel, A. Kirilyuk, P.A. Usachev, R.V. Pisarev, A.M. Balbashov, and T. Rasing, *Nature* **435**, 655 (2005)
- [90] T. Kampfrath, A. Sell, G. Klatt, A. Pashkin, S. Mährlein, T. Dekorsy, M. Wolf, M. Fiebig, A. Leitenstorfer, and R. Huber, *Nat. Photonics* **5**, 31 (2011)
- [91] C.D. Stanciu, F. Hansteen, A.V. Kimel, A. Kirilyuk, A. Tsukamoto, A. Itoh, and T. Rasing, *Phys. Rev. Lett.* **99**, 047601 (2007)
- [92] A. Stupakiewicz, C.S. Davies, K. Szerenos, D. Afanasiev, K.S. Rabinovich, A.V. Boris, A.D. Caviglia, A.V. Kimel, and A. Kirilyuk, *Nat. Phys.* **17**, 489 (2021)
- [93] I. Radu, K. Vahaplar, C. Stamm, T. Kachel, N. Pontius, H.A. Dürr, T.A. Ostler, J. Barker, R.F.L. Evans, R.W. Chantrell, A. Tsukamoto, A. Itoh, A. Kirilyuk, T. Rasing, and A.V. Kimel, *Nature* **472**, 205 (2011)
- [94] D. Afanasiev, J.R. Hortensius, B.A. Ivanov, A. Sasani, E. Bousquet, Y.M. Blanter, R.V. Mikhaylovskiy, A.V. Kimel, and A.D. Caviglia, *Nat. Mater.* **20**, 607 (2021)
- [95] T.F. Nova, A. Cartella, A. Cantaluppi, M. Först, D. Bossini, R.V. Mikhaylovskiy, A.V. Kimel, R. Merlin, and A. Cavalleri, *Nat. Phys.* **13**, 132 (2017)
- [96] D.M. Juraschek, M. Fechner, A.V. Balatsky, and N.A. Spaldin, *Phys. Rev. Mater.* **1**, 014401 (2017)
- [97] D.M. Juraschek, P. Narang, and N.A. Spaldin, *Phys. Rev. Res.* **2**, 043035 (2020)
- [98] D.M. Juraschek, T. Neuman, and P. Narang, *Phys. Rev. Res.* **4**, 013129 (2022)
- [99] J. Luo, T. Lin, J. Zhang, X. Chen, E.R. Blackert, R. Xu, B.I. Yakobson, and H. Zhu, *Science* **382**, 698 (2023)
- [100] A.S. Disa, M. Fechner, T.F. Nova, B. Liu, M. Först, D. Prabhakaran, P.G. Radaelli, and A. Cavalleri, *Nat. Phys.* **16**, 937 (2020)
- [101] A.S. Disa, J. Curtis, M. Fechner, A. Liu, A. von Hoegen, M. Först, T.F. Nova, P. Narang, A. Maljuk, A.V. Boris, B. Keimer, and A. Cavalleri, *Nature* **617**, 73 (2023)
- [102] B. Ilyas, T. Luo, A. von Hoegen, E. Viñas Boström, Z. Zhang, J. Park, J. Kim, J.-G. Park, K.A. Nelson, A. Rubio, and N. Gedik, *Nature* **636**, 609 (2024)
- [103] R. Mankowsky, A. von Hoegen, M. Först, and A. Cavalleri, *Phys. Rev. Lett.* **118**, 197601 (2017)

- [104] G. Orenstein, V. Krapivin, Y. Huang, Z. Zhang, G. de la Peña Muñoz, R.A. Duncan, Q. Nguyen, J. Stanton, S. Teitelbaum, H. Yavas, T. Sato, M.C. Hoffmann, P. Kramer, J. Zhang, A. Cavalleri, R. Comin, M.P.M. Dean, A.S. Disa, M. Först, S.L. Johnson, M. Mitrano, A.M. Rappe, D.A. Reis, D. Zhu, K.A. Nelson, and M. Trigo, *Nat. Phys.* **21**, 961 (2025)
- [105] X. Li, T. Qiu, J. Zhang, E. Baldini, J. Lu, A.M. Rappe, and K.A. Nelson, *Science* **364**, 1079 (2019)
- [106] M. Fechner, M. Först, G. Orenstein, V. Krapivin, A.S. Disa, M. Buzzi, A. von Hoegen, G. de la Pena, Q.L. Nguyen, R. Mankowsky, M. Sander, H. Lemke, Y. Deng, M. Trigo, and A. Cavalleri, *Nat. Mater.* **23**, 363 (2024)
- [107] H.H. Wang, M. Fechner, G. De Vecchi, S.L. Griffitt, G. Orenstein, J. Stanton, V. Krapivin, M.T. Wong, Z. Zhang, M. Bionta, V. Esposito, M. Henstridge, M.C. Hoffmann, P.L. Kramer, Z. Porter, R.A. Duncan, T. Sato, S.K. Kim, H. Yavas, S. Teitelbaum, K. Nelson, A.S. Disa, M. Först, M. Trigo, and A. Cavalleri, *arXiv:2606.16267* (2026)
- [108] T. Kubacka, J.A. Johnson, M.C. Hoffmann, C. Vicario, S. de Jong, P. Beaud, S. Grübel, S.-W. Huang, L. Huber, L. Patthey, Y.-D. Chuang, J.J. Turner, G.L. Dakovski, W.-S. Lee, M.P. Minitti, W. Schlotter, R.G. Moore, C.P. Hauri, S.M. Koohpayeh, V. Scagnoli, G. Ingold, S.L. Johnson, and U. Staub, *Science* **343**, 1333 (2014)
- [109] Z. Zeng, M. Först, M. Fechner, D. Prabhakaran, P.G. Radaelli, and A. Cavalleri, *Science* **390**, 195 (2025)
- [110] S.H. Autler and C.H. Townes, *Phys. Rev.* **100**, 703 (1955)
- [111] T. Oka and S. Kitamura, *Ann. Rev. Condens. Matter Phys.* **10**, 387 (2019)
- [112] M.S. Rudner and N.H. Lindner, *Nat. Rev. Phys.* **2**, 229 (2020)
- [113] D.H. Dunlap and V.M. Kenkre, *Phys. Rev. B* **34**, 3625 (1986)
- [114] J.H. Mentink, K. Balzer, and M. Eckstein, *Nat. Commun.* **6**, 6708 (2015)
- [115] G. Floquet, *Annales scientifiques de l'École Normale Supérieure* **12**, 47 (1883)
- [116] M. Bukov, L. D'Alessio, and A. Polkovnikov, *Adv. Phys.* **64**, 139 (2015)
- [117] T. Oka and H. Aoki, *Phys. Rev. B* **79**, 081406(R) (2009)
- [118] N.H. Lindner, G. Refael, and V. Galitski, *Nat. Phys.* **7**, 490 (2011)
- [119] T. Kitagawa, T. Oka, A. Brataas, L. Fu, and E. Demler, *Phys. Rev. B* **84**, 235108 (2011)
- [120] S. Zhou, C. Bao, B. Fan, H. Zhou, Q. Gao, H. Zhong, T. Lin, H. Liu, P. Yu, P. Tang, S. Meng, W. Duan, and S. Zhou, *Nature* **614**, 75 (2023)

- [121] D. Choi, M. Mogi, U. De Giovannini, D. Azoury, B. Lv, Y. Su, H. Hübener, A. Rubio, and N. Gedik, *Nat. Phys.* **21**, 1100 (2025)
- [122] M. Merboldt, M. Schüler, D. Schmitt, J.P. Bange, W. Bennecke, K. Gadge, K. Pierz, H.W. Schumacher, D. Momeni, D. Steil, S.R. Manmana, M.A. Sentef, M. Reutz, and S. Mathias, *Nat. Phys.* **21**, 1093 (2025)
- [123] E.J. Sie, J.W. McIver, Y.-H. Lee, L. Fu, J. Kong, and N. Gedik, *Nat. Mater.* **14**, 290 (2015)
- [124] E.J. Sie, C.H. Lui, Y.-H. Lee, L. Fu, J. Kong, and N. Gedik, *Science* **355**, 1066 (2017)
- [125] J.-Y. Shan, M. Ye, H. Chu, S. Lee, J.-G. Park, L. Balents, and D. Hsieh, *Nature* **600**, 235 (2021)
- [126] S.T. Cundiff, A. Knorr, J. Feldmann, S.W. Koch, E.O. Göbel, and H. Nickel, *Phys. Rev. Lett.* **73**, 1178 (1994)
- [127] B.E. Cole, J.B. Williams, B.T. King, M.S. Sherwin, and C.R. Stanley, *Nature* **410**, 60 (2001)
- [128] D. Press, T.D. Ladd, B. Zhang, and Y. Yamamoto, *Nature* **456**, 218 (2008)
- [129] J. Berezovsky, M.H. Mikkelsen, N.G. Stoltz, L.A. Coldren, and D.D. Awschalom, *Science* **320**, 349 (2008)
- [130] D.G. Clarke, *Phys. Rev. B* **48**, 7520 (1993)
- [131] F.H.L. Essler, F. Gebhard, and E. Jeckelmann, *Phys. Rev. B* **64**, 125119 (2001)
- [132] H. Kishida, H. Matsuzaki, H. Okamoto, T. Manabe, M. Yamashita, Y. Taguchi, and Y. Tokura, *Nature* **405**, 929 (2000)
- [133] M. Ono, K. Miura, A. Maeda, H. Matsuzaki, H. Kishida, Y. Taguchi, Y. Tokura, M. Yamashita, and H. Okamoto, *Phys. Rev. B* **70**, 085101 (2004)
- [134] D.R. Baykusheva, D. Carmichael, C.S. Weber, I.-T. Lu, F. Glerean, T. Meng, P.B.M. De Oliveira, C.C. Homes, I.A. Zaliznyak, G.D. Gu, M.P.M. Dean, A. Rubio, D.M. Kennes, M. Claassen, and M. Mitrano, *Nat. Mater.* **25**, 937 (2026)
- [135] H.A. Rabitz, M.M. Hsieh, and C.M. Rosenthal, *Science* **303**, 1998 (2004)
- [136] L.M.K. Vandersypen and I.L. Chuang, *Rev. Mod. Phys.* **76**, 1037 (2005)
- [137] J. Choi, H. Zhou, H.S. Knowles, R. Landig, S. Choi, and M.D. Lukin, *Phys. Rev. X* **10**, 031002 (2020)

- [138] H. Padma, P. Sharma, S.F.R. TenHuisen, F. Glerean, A. Roll, P. Zhou, S. Kundu, A. Romaguera, E. Skoropata, H. Ueda, B. Liu, E. Paris, Y. Wang, S.H. Lee, Z. Mao, M.P.M. Dean, E.W. Huang, E. Razzoli, Y. Wang, and M. Mitrano, *Phys. Rev. Lett.* **136**, 196501 (2026)
- [139] R.D. Averitt, G. Rodriguez, A.I. Lobad, J.L.W. Siders, S.A. Trugman, and A.J. Taylor, *Phys. Rev. B* **63**, 140502 (2001)
- [140] R.A. Kaindl, M. Woerner, T. Elsaesser, D.C. Smith, J.F. Ryan, G.A. Farnan, M.P. McCurry, and D.G. Walmsley, *Science* **287**, 470 (2000)
- [141] A. Pashkin, M. Porer, M. Beyer, K.W. Kim, A. Dubroka, C. Bernhard, X. Yao, Y. Dagan, R. Hackl, A. Erb, J. Demsar, R. Huber, and A. Leitenstorfer, *Phys. Rev. Lett.* **105**, 067001 (2010)
- [142] J. Demsar, *J. Low Temp. Phys.* **201**, 676 (2020)
- [143] A.F.G. Wyatt, V.M. Dmitriev, W.S. Moore, and F.W. Sheard, *Phys. Rev. Lett.* **16**, 1166 (1966)
- [144] A.H. Dayem and J.J. Wiegand, *Phys. Rev.* **155**, 419 (1967)
- [145] G.M. Eliashberg, *JETP Lett.* **11**, 114 (1970) [*Pis'ma Zh. Eksp. Teor. Fiz.* **11**, 186 (1970)]
- [146] M. Dressel and G. Grüner: *Electrodynamics of Solids: Optical Properties of Electrons in Matter* (Cambridge University Press, 2002)
- [147] M. Tinkham: *Introduction to Superconductivity* (Dover Publications, 2004), 2nd ed.
- [148] S. Fava, G. De Vecchi, G. Jotzu, M. Buzzi, T. Gebert, Y. Liu, B. Keimer, and A. Cavalleri, *Nature* **632**, 75 (2024)
- [149] D. Fausti, R. Tobey, N. Dean, S. Kaiser, A. Dienst, M.C. Hoffmann, S. Pyon, T. Takayama, H. Takagi, and A. Cavalleri, *Science* **331**, 189 (2011)
- [150] E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, and O.K. Andersen, *Phys. Rev. Lett.* **87**, 047003 (2001)
- [151] W. Hu, S. Kaiser, D. Nicoletti, C.R. Hunt, I. Gierz, M.C. Hoffmann, M. Le Tacon, T. Loew, B. Keimer, and A. Cavalleri, *Nat. Mater.* **13**, 705 (2014)
- [152] S. Kaiser, C.R. Hunt, D. Nicoletti, W. Hu, I. Gierz, H.Y. Liu, M. Le Tacon, T. Loew, D. Haug, B. Keimer, and A. Cavalleri, *Phys. Rev. B* **89**, 184516 (2014)
- [153] O. Gunnarsson, *Rev. Mod. Phys.* **69**, 575 (1997)
- [154] M. Mitrano, A. Cantaluppi, D. Nicoletti, S. Kaiser, A. Perucchi, S. Lupi, P. Di Pietro, D. Pontiroli, M. Riccò, S.R. Clark, D. Jaksch, and A. Cavalleri, *Nature* **530**, 461 (2016)

- [155] A. Cantaluppi, M. Buzzi, G. Jotzu, D. Nicoletti, M. Mitrano, D. Pontiroli, M. Riccò, A. Perucchi, P. Di Pietro, and A. Cavalleri, *Nat. Phys.* **14**, 837 (2018)
- [156] E. Wang, J.D. Adelinia, M. Chavez-Cervantes, T. Matsuyama, M. Fechner, M. Buzzi, G. Meier, and A. Cavalleri, *Nat. Commun.* **14**, 7233 (2023)
- [157] M. Budden, T. Gebert, M. Buzzi, G. Jotzu, E. Wang, T. Matsuyama, G. Meier, Y. Laplace, D. Pontiroli, M. Riccò, F. Schlawin, D. Jaksch, and A. Cavalleri, *Nat. Phys.* **17**, 611 (2021)
- [158] E. Rowe, B. Yuan, M. Buzzi, G. Jotzu, Y. Zhu, M. Fechner, M. Först, B. Liu, D. Pontiroli, M. Riccò, and A. Cavalleri, *Nat. Phys.* **19**, 1821 (2023)
- [159] B.J. Powell and R.H. McKenzie, *J. Phys.: Condens. Matter* **18**, R827 (2006)
- [160] D. Faltermeier, J. Barz, M. Dumm, M. Dressel, N. Drichko, B. Petrov, V. Semkin, R. Vlasova, C. Mézière, and P. Batail, *Phys. Rev. B* **76**, 165113 (2007)
- [161] M. Buzzi, D. Nicoletti, M. Fechner, N. Tancogne-Dejean, M.A. Sentef, A. Georges, T. Biesner, E. Uykur, M. Dressel, A. Henderson, T. Siegrist, J.A. Schlueter, K. Miyagawa, K. Kanoda, M.-S. Nam, A. Ardavan, J. Coulthard, J. Tindall, F. Schlawin, D. Jaksch, and A. Cavalleri, *Phys. Rev. X* **10**, 031028 (2020)
- [162] M. Buzzi, D. Nicoletti, S. Fava, G. Jotzu, K. Miyagawa, K. Kanoda, A. Henderson, T. Siegrist, J.A. Schlueter, M.-S. Nam, A. Ardavan, and A. Cavalleri, *Phys. Rev. Lett.* **127**, 197002 (2021)
- [163] Y.J. Uemura, *Phys. Rev. Mater.* **3**, 104801 (2019)
- [164] M.A. Sentef, A.F. Kemper, A. Georges, and C. Kollath, *Phys. Rev. B* **93**, 144506 (2016)
- [165] D.M. Kennes, E.Y. Wilner, D.R. Reichman, and A.J. Millis, *Nat. Phys.* **13**, 479 (2017)
- [166] M. Knap, M. Babadi, G. Refael, I. Martin, and E. Demler, *Phys. Rev. B* **94**, 214504 (2016)
- [167] M. Babadi, M. Knap, I. Martin, G. Refael, and E. Demler, *Phys. Rev. B* **96**, 014512 (2017)
- [168] A. Nava, C. Giannetti, A. Georges, E. Tosatti, and M. Fabrizio, *Nat. Phys.* **14**, 154 (2018)
- [169] C.N. Yang, *Phys. Rev. Lett.* **63**, 2144 (1989)
- [170] T. Kaneko, T. Shirakawa, S. Sorella, and S. Yunoki, *Phys. Rev. Lett.* **122**, 077002 (2019)
- [171] J. Tindall, B. Buča, J.R. Coulthard, and D. Jaksch, *Phys. Rev. Lett.* **123**, 030603 (2019)

- [172] J. Tindall, F. Schlawin, M. Buzzi, D. Nicoletti, J.R. Coulthard, H. Gao, A. Cavalleri, M.A. Sentef, and D. Jaksch, *Phys. Rev. Lett.* **125**, 137001 (2020)
- [173] Z. Dai and P.A. Lee, *Phys. Rev. B* **104**, 054512 (2021)
- [174] Z. Dai and P.A. Lee, *Phys. Rev. B* **104**, L241112 (2021)
- [175] M.H. Michael, A. von Hoegen, M. Fechner, M. Först, A. Cavalleri, and E. Demler, *Phys. Rev. B* **102**, 174505 (2020)
- [176] F. Gucci, E.B. Molinero, M. Russo, P. San-Jose, F.V.A. Camargo, M. Maiuri, M. Ivanov, Á. Jiménez-Galán, R.E.F. Silva, S. Dal Conte, and G. Cerullo, *Nat. Photonics* **20**, 266 (2026)
- [177] F. Schlawin, D.M. Kennes, and M.A. Sentef, *Appl. Phys. Rev.* **9**, 011312 (2022)
- [178] G. Jarc, S.Y. Mathengattil, A. Montanaro, F. Giusti, E.M. Rigoni, R. Sergo, F. Fassioli, S. Winnerl, S. Dal Zilio, D. Mihailovic, P. Prelovšek, M. Eckstein, and D. Fausti, *Nature* **622**, 487 (2023)
- [179] F. Helmrich, H.S. Adlong, M. Kroner, I. Khanonkin, G. Scalari, J. Faist, A. İmamoğlu, and T.F. Nova, *Nature* **654**, 361 (2026)
- [180] A. Montanaro, V. Plastovets, N. Khatiwada, J. Fiore, G. Jarc, A. Alabbadi, A. Mastropasqua, E.M. Rigoni, S.Y. Mathengattil, S.D. Zilio, F.F. Olsen, F. Novelli, S. Winnerl, M.A. Sentef, D.M. Kennes, A.J. Millis, F. Piazza, and D. Fausti, *arXiv:2606.18084* (2026)
- [181] H. Zhang, Z. Feng, I.-T. Lu, Z. Li, S. Guo, Q. Shang, D. Luan, M. Panmai, K. Watanabe, T. Taniguchi, A. Rubio, and W. Gao, *arXiv:2606.19171* (2026)
- [182] F. Appugliese, J. Enkner, G.L. Paravicini-Bagliani, M. Beck, C. Reichl, W. Wegscheider, G. Scalari, C. Ciuti, and J. Faist, *Science* **375**, 1030 (2022)
- [183] I. Keren, T.A. Webb, S. Zhang, J. Xu, D. Sun, B.S.Y. Kim, D. Shin, S.S. Zhang, J. Zhang, G. Pereira, J. Yao, T. Okugawa, M.H. Michael, E. Viñas Boström, J.H. Edgar, S. Wolf, M. Julian, R.P. Prasankumar, K. Miyagawa, K. Kanoda, G. Gu, M. Cothrine, D. Mandrus, M. Buzzi, A. Cavalleri, C.R. Dean, D.M. Kennes, A.J. Millis, Q. Li, M.A. Sentef, A. Rubio, A.N. Pasupathy, and D.N. Basov, *Nature* **650**, 864 (2026)

13 Variational Monte Carlo for the Hubbard Model

Luca F. Tocchio
Politecnico di Torino
Corso Duca degli Abruzzi, 24 10129-Turin, Italy

Contents

1	Introduction	2
2	The Hubbard model	3
2.1	The single-orbital Hubbard model	3
2.2	The multi-orbital Hubbard model	3
3	Variational Monte Carlo	4
3.1	The Metropolis algorithm	6
4	Variational wave functions	8
4.1	Jastrow-Slater wave functions for the single-orbital Hubbard model	8
4.2	Computing ratios for the Jastrow-Slater wave function	10
4.3	The backflow terms	13
4.4	Jastrow-Slater wave functions for the multi-orbital Hubbard model	16
5	Optimizing the variational state: Stochastic reconfiguration	17
6	Correlation functions in variational Monte Carlo: Definitions and examples	19
7	Final remarks	24

1 Introduction

One of the reasons why the study of correlated Hamiltonians in two dimensions is one of the most active and debated fields in modern condensed matter physics is the absence of a clear-cut method for tackling non-exactly solvable Hamiltonians, like the Hubbard model. Variational Monte Carlo (VMC) is one of the few methods that allows us to address Hubbard models, including its multi-orbital extensions, directly in two dimensions, describing non-local correlations and ordering, including non-local superconductivity [1]. We remark that VMC is a nonperturbative approach for which performance is not significantly affected by the dimensionality of the system or by the size of the local Hilbert space. In this respect, VMC might be more efficient than other methods, like density-matrix renormalization group or tensor-network methods, in which the complexity increases with both the range of the interaction and the dimension of the local Hilbert space [2]. It can also provide complementary information with respect to methods like dynamical mean-field theory [3], which can capture accurate dynamical local correlations, but does not include nonlocal spatial correlations beyond mean field.

The VMC methods rely on two main ingredients. The first one is the construction of a variational wave function that approximates the correct ground-state. This approach relies on the variational principle, stating that the energy of a given quantum state is always bounded from below by the exact ground-state energy, thus giving us guidance to obtain the best solution to the problem. Moreover, a suitable parametrization of the ground state allows us to consider a wide range of different quantum phases (superconductivity, magnetic and charge order, Mott insulators...). By contrast, the main limitation of our approach is the fact that it is based on a given Ansatz, which may contain a relevant bias.

The second one is the Monte Carlo sampling: Computing expectation values of observables over the variational estimate of the ground-state wave function requires a numerical Monte Carlo sampling since the wave function is correlated. Moreover, the variational state contains many parameters to be optimized at the same time. This is a highly non-trivial and computationally demanding task that is made possible thanks to the Stochastic Reconfiguration method, which allows us to find the optimal set for hundreds of independent parameters [4]. Once the optimal variational state is obtained, the presence of, for instance, spin order, charge order, and superconductivity can be detected with the computation of appropriate correlation functions. These physical observables are computed on finite lattice sizes with periodic boundary conditions and might require size scaling to the thermodynamic limit.

In Sec. 2, we briefly introduce the single-orbital Hubbard model and its multi-orbital extension. In this lecture, we will focus only on electronic models, even though the Hubbard model can also be extended to bosonic particles. In Sec. 3, we introduce the general scheme under which variational Monte Carlo works. Then, in Sec. 4, we present the variational wave functions that can properly describe superconductive and magnetic states in single-orbital and multi-orbital Hubbard models. We also detail how to perform Monte Carlo calculations for the given wave functions. In Sec. 5, we describe how to optimize a variational wave function with the stochastic reconfiguration method. In Sec. 6, we present a few results for correlation functions that can be computed in variational Monte Carlo. Finally, in Sec 7, we draw the conclusions.

2 The Hubbard model

2.1 The single-orbital Hubbard model

In this section, we introduce the single-orbital Hubbard model on the square lattice, a prototype model for studying the role of electronic correlations [5, 6]. In particular, this model is the simplest model able to capture the essential characteristics of the cuprates' phase diagram. This Hamiltonian comprises an interaction term $\mathcal{H}_{\text{int}}$ that contains the on-site electron-electron repulsion U and a kinetic part $\mathcal{H}_0$ that includes the nearest neighbor hopping integral t , and the next-nearest neighbor hopping integral t' . This last term is directly connected with the critical temperature for superconductivity [7]. The Hamiltonian is defined as

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{int}} = -t \sum_{\langle R, R' \rangle, \sigma} c_{R\sigma}^\dagger c_{R'\sigma} - t' \sum_{\langle\langle R, R' \rangle\rangle, \sigma} c_{R\sigma}^\dagger c_{R'\sigma} + \text{H.c} + U \sum_R n_{R\uparrow} n_{R\downarrow}, \quad (1)$$

where $c_{R\sigma}^\dagger$ and $c_{R\sigma}$ denote the creation and annihilation operators of an electron with spin σ on site R , respectively, while $n_{R\sigma} = c_{R\sigma}^\dagger c_{R\sigma}$ is the electron density per spin σ on site R . The symbols $\langle \dots \rangle$ and $\langle\langle \dots \rangle\rangle$ represent the summation over nearest- and next-nearest-neighbor sites, respectively. By calling N_e the total number of electrons and L the total number of sites, the electron density is $n = N_e/L$. At half filling, i.e., when $n = 1$, the model interpolates between a metal at $U = 0$ and an insulator driven by electronic correlations, the so-called Mott insulator, at $t = 0$. A metal-insulator transition will occur at a critical value of U/t that depends on the lattice geometry.

Despite its simplicity, this model cannot be exactly solved, except in a few limiting cases, like in one spatial dimension with nearest-neighbor hopping, where the so-called Bethe Ansatz provides us with exact results [8], and in the half-filled case for any value of U/t at $t' = 0$ on the square lattice, where the auxiliary-field Monte Carlo approach becomes numerically exact, with no sign problem [9]. In all other cases, we need to rely on approximate numerical methods.

2.2 The multi-orbital Hubbard model

When extending the Hubbard model from a single relevant orbital to a multi-orbital scheme, we need to include new interaction terms. In addition to the intra-orbital on-site Coulomb interaction U , we shall also include the inter-orbital on-site Coulomb repulsion U' and the Hund's coupling J . The latter term tends to maximize the total spin on a single site, as a consequence of the intra-atomic exchange energy. The interaction term of the Hubbard Hamiltonian assumes then a more complicated form, also known as the Kanamori Hamiltonian [10, 11]

$$\begin{aligned} \mathcal{H}_{\text{int}} = & U \sum_R \sum_\alpha n_{R,\alpha,\uparrow} n_{R,\alpha,\downarrow} + U' \sum_R \sum_{\alpha \neq \beta} n_{R,\alpha,\uparrow} n_{R,\beta,\downarrow} + (U' - J) \sum_R \sum_{\alpha > \beta} \sum_\sigma n_{R,\alpha,\sigma} n_{R,\beta,\sigma} \\ & - J \sum_R \sum_{\alpha \neq \beta} c_{R,\alpha,\uparrow}^\dagger c_{R,\alpha,\downarrow}^\dagger c_{R,\beta,\downarrow} c_{R,\beta,\uparrow} + J \sum_R \sum_{\alpha \neq \beta} c_{R,\alpha,\uparrow}^\dagger c_{R,\alpha,\downarrow}^\dagger c_{R,\beta,\downarrow} c_{R,\beta,\uparrow}, \end{aligned} \quad (2)$$

where α and β are orbital indices, $c_{R,\alpha,\sigma}^\dagger$ ($c_{R,\alpha,\sigma}$) is the fermionic operator that creates (annihilates) an electron in orbital α , site R and spin σ , and $n_{R,\alpha,\sigma} = c_{R,\alpha,\sigma}^\dagger c_{R,\alpha,\sigma}$ is the density operator on site R , orbital α and spin σ .

The first three terms refer to the on-site density-density interaction between electrons with opposite spins in the same orbital (U), opposite spins in different orbitals (U'), and parallel spins in different orbitals ($U' - J < U'$). In the latter term, the interaction constant is lower because of the spin alignment effect. The fourth is the spin-flip process (J), and the last term describes pair hopping, i.e., an up-down spin pair of electrons that moves from one orbital to another one (J). The preservation of rotational symmetry is guaranteed for $U' = U - 2J$, which will be used throughout the discussion.

The kinetic part of the multi-orbital Hubbard model depends on the specific hopping scheme that characterizes the model under investigation. Here, we consider iron-based superconductors as an example. In this case, a three Fe orbital (d_{xz} , d_{yz} , and d_{xy}) kinetic Hamiltonian $\mathcal{H}_0$ [12, 13] is able to reproduce qualitatively the generic shape and the orbital content of the Fermi surfaces of iron-based superconductors, while remaining sufficiently light to allow for accurate numerical simulations. The tight-binding Hamiltonian is then defined as

$$\mathcal{H}_0 = \sum_k \sum_{\alpha,\beta} \sum_{\sigma} c_{k,\alpha,\sigma}^\dagger T_{\alpha,\beta}(k) c_{k,\beta,\sigma}, \quad (3)$$

where α and β are orbital indices ($1=xz$, $2=yz$, and $3=xy$), while $c_{k,\alpha,\sigma}^\dagger$ ($c_{k,\alpha,\sigma}$) is the fermionic operator that creates (annihilates) an electron in orbital α , with momentum k and spin σ . The intra- and inter-orbital hoppings are given as

$$\begin{aligned} T_{1,1}(k) &= 2t_2 \cos k_x + 2t_1 \cos k_y + 4t_3 \cos k_x \cos k_y, \\ T_{2,2}(k) &= 2t_1 \cos k_x + 2t_2 \cos k_y + 4t_3 \cos k_x \cos k_y, \\ T_{3,3}(k) &= 2t_5(\cos k_x + \cos k_y) + 4t_6 \cos k_x \cos k_y + \varepsilon_{xy}, \\ T_{1,2} &= T_{2,1}^* = 4t_4 \sin k_x \sin k_y, \\ T_{1,3} &= T_{3,1}^* = 2it_7 \sin k_x + 4it_8 \sin k_x \cos k_y, \\ T_{2,3} &= T_{3,2}^* = 2it_7 \sin k_y + 4it_8 \sin k_y \cos k_x. \end{aligned}$$

The hopping parameters (in units of eV) are set as: $t_1 = 0.02$, $t_2 = 0.06$, $t_3 = 0.03$, $t_4 = -0.01$, $t_5 = 0.1$, $t_6 = 0.15$, $t_7 = -0.1$, $t_8 = -t_7/2$ [13]. The two orbitals d_{xz} and d_{yz} are degenerate, while the orbital d_{xy} has a crystal field $\varepsilon_{xy} = 0.2$.

3 Variational Monte Carlo

Let us now introduce the general scheme under which variational Monte Carlo is defined. Let us first fix a complete orthonormalized basis in the Hilbert space $\{|x\rangle\}$

$$\sum_x |x\rangle\langle x| = \mathbb{I}. \quad (4)$$

In order to clarify this statement with an example, an element of the basis for the single-orbital Hubbard model is a vector that contains the electronic occupation of each lattice site.

Given the aforementioned basis, each quantum state can be written as:

$$|\Psi\rangle = \sum_x |x\rangle \langle x|\Psi\rangle = \sum_x \Psi(x) |x\rangle. \quad (5)$$

In this way, the expectation value of a generic operator $\mathcal{O}$ over a given variational state can be written as

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_x \langle \Psi | x \rangle \langle x | \mathcal{O} | \Psi \rangle}{\sum_{x'} \langle \Psi | x' \rangle \langle x' | \Psi \rangle}. \quad (6)$$

The main problem in evaluating this expectation value lies in the number of configurations in the sum, which becomes exponentially large with the number of sites and particles. Therefore, for large sizes, it is impossible to perform an exact enumeration of the configurations to compute $\mathcal{O}$ exactly. However, Eq. (6) can be recast into a form that can be easily treated by standard Monte Carlo methods

$$\langle \mathcal{O} \rangle = \frac{\sum_x |\langle \Psi | x \rangle|^2 \frac{\langle x | \mathcal{O} | \Psi \rangle}{\langle x | \Psi \rangle}}{\sum_{x'} |\langle \Psi | x' \rangle|^2} = \frac{\sum_x |\Psi(x)|^2 O_L(x)}{\sum_{x'} |\Psi(x')|^2}, \quad (7)$$

where we have defined the local estimator

$$O_L(x) = \frac{\langle x | \mathcal{O} | \Psi \rangle}{\langle x | \Psi \rangle} = \sum_{x'} \langle x | \mathcal{O} | x' \rangle \frac{\langle x' | \Psi \rangle}{\langle x | \Psi \rangle}. \quad (8)$$

At first sight, the computation of the local estimator looks like a hard task, since it requires a summation over all the states of the many-body Hilbert space; however, when considering operators like the Hamiltonian, only a few terms contribute to the sum, thanks to the locality of the Hubbard Hamiltonian. Indeed, if we use the local basis, whose elements are the electronic configurations in the lattice, an element $|x\rangle$ is connected only to a few other configurations that differ in the hopping of one electron from a given site to one of its neighbors; then, the maximum number of such processes is equal to the number of sites L times the number of bonds times the spin degeneracy. Therefore, the computation of the local estimator requires only a limited number of operations, usually proportional to the number of sites or particles in the system.

The quantity

$$P(x) = \frac{|\Psi(x)|^2}{\sum_{x'} |\Psi(x')|^2} \quad (9)$$

can be interpreted as a probability distribution, since it is non-negative and normalized. This property guarantees that variational Monte Carlo is not affected by the sign problem that is common to other Monte Carlo approaches.

Thus, the problem of computing a quantum average of the operator $\mathcal{O}$ can be rewritten as the average of the random variable $O_L(x)$, see Eq. (8), over the probability distribution $P(x)$, see Eq. (9). In particular, if we consider the expectation value of the Hamiltonian, the local estimator is named local energy and is defined by

$$e_L(x) = \frac{\langle x | \mathcal{H} | \Psi \rangle}{\langle x | \Psi \rangle}. \quad (10)$$

It is then possible to define a stochastic algorithm (e.g., a Markov process) in which a sequence of electronic configurations $\{|x_n\rangle\}$ is generated. One can use, for example, the Metropolis algorithm, which is a customary algorithm in Monte Carlo simulations, for generating states according to the given probability distribution. With the Metropolis algorithm, after an initial equilibration time, configurations will be distributed according to the desired probability $P(x)$. Then, the quantum expectation value $\langle O \rangle$ is evaluated as the mean value of the random variable $O_L(x)$ over the visited configurations

$$\langle O \rangle \approx \frac{1}{N} \sum_{n=1}^N O_L(x_n), \quad (11)$$

where N is the total number of visited configurations, after equilibration.

An important feature of VMC is the zero-variance property: if the variational state $|\Psi\rangle$ coincides with an eigenstate of the Hamiltonian with eigenvalue E , the local energy $e_L(x)$ is a constant

$$e_L(x) = \frac{\langle x | \mathcal{H} | \Psi \rangle}{\langle x | \Psi \rangle} = E \frac{\langle x | \Psi \rangle}{\langle x | \Psi \rangle} = E. \quad (12)$$

Therefore, the variance of the random variable $e_L(x)$ is zero. Even if this is an extreme case, it teaches us that the variance of $e_L(x)$ decreases when the variational state approaches an exact eigenstate. This fact is important to reduce statistical fluctuations and improve the numerical efficiency. Notice that the zero-variance property is a feature that exists only for quantum expectation values, while it is absent in classical calculations, where observables have thermal fluctuations.

3.1 The Metropolis algorithm

In this lecture, we do not discuss the details of Markov chains and the conditions that guarantee that the configurations $|x_n\rangle$ generated along the Markov chain are distributed according to the probability distribution $P(x)$ of Eq. (9), for large values of n . More details can be found, for instance, in Ref. [1]. Instead, we just recall the Metropolis algorithm that defines a transition probability for generating a new electronic configuration $|x'\rangle$, starting from an old one $|x\rangle$. Once a move is proposed to generate the new configuration $|x'\rangle$ starting from the old configuration $|x\rangle$, the acceptance probability $A(x'|x)$ is defined as

$$A(x'|x) = \min \left\{ 1, \frac{T(x|x')P(x')}{T(x'|x)P(x)} \right\}, \quad (13)$$

where $T(x'|x)$ is the trial probability for generating $|x'\rangle$ from $|x\rangle$. If the trial probability is symmetric ($T(x'|x) = T(x|x')$), Eq. (13) further simplifies to:

$$A(x'|x) = \min \left\{ 1, \frac{P(x')}{P(x)} \right\}, \quad (14)$$

where

$$\frac{P(x')}{P(x)} = \frac{|\Psi(x')|^2}{|\Psi(x)|^2}, \quad (15)$$

thus making it unnecessary to know the normalization of the wave function.

Recalling that each configuration $|x\rangle$ is the electron occupation of each lattice site (with a fixed z component of the spin), we set in the following a generic procedure to run a Metropolis Monte Carlo algorithm for fermions:

- Generate a random number to identify the sites R and R' among which the electron (with fixed spin σ) will hop, defining the trial probability $T(x'|x)$. In strongly correlated configurations (i.e., large values of U/t at half filling), an additional trial move can be proposed, as a spin flip between two electrons with opposite spin on singly occupied sites.
- If the electron can perform the move (specifically, it must not violate the Pauli principle), the algorithm computes the acceptance ratio as $A(x'|x)$; otherwise, the move is rejected.
- Generate a random uniform number $\eta \in (0, 1]$ to compare with the acceptance ratio. The new configuration $|x'\rangle$ is accepted for $A(x'|x) \geq \eta$; otherwise, the move is rejected.
- Once thermalization is reached, i.e., when configurations are generated according to the desired probability distribution, observables, such as the local energy $E_L(x)$, can be computed every $O(L)$ steps, where L is the lattice size, to have uncorrelated configurations. As a rule of thumb, thermalization time is similar to correlation time, that is, the necessary time to run a simulation from a starting configuration $|x\rangle$ to obtain an independent configuration $|x'\rangle$.

If the configurations are uncorrelated, the error bar associated with the variational estimate of an observable is given by

$$\sigma_{\langle \mathcal{O} \rangle} = \sqrt{\frac{1}{N(N-1)} \left(\sum_{i=1}^N O_L^2(x_i) - \frac{1}{N} \left(\sum_{i=1}^N O_L(x_i) \right)^2 \right)}. \quad (16)$$

However, the usual acceptance probability for VMC in the Hubbard model is quite small, implying that many moves are rejected, and the electronic configurations stored during the Monte Carlo sampling could be correlated. Consequently, the error bar will be underestimated if we use Eq. (16). A simple procedure used to solve the problem of successive correlated configuration, in the evaluation of the error bar, is the *binning technique*, in which L_b measures are grouped in a bin, in order to compute a partial average within the bin b :

$$\bar{O}_b = \frac{1}{L_b} \sum_{i=(b-1)L_b+1}^{bL_b} O_L(x_i); \quad (17)$$

here, $b = 1, \dots, N_{\text{bin}}$ such that $N = \sum_{b=1}^{N_{\text{bin}}} L_b$. Whenever L_b is larger than the correlation time, the configurations are almost uncorrelated. Based on the binning technique, the variational energy and its error bar are given by

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{bin}}} \sum_{b=1}^{N_{\text{bin}}} \bar{O}_b \quad \text{and} \quad \sigma_{\langle \mathcal{O} \rangle} = \sqrt{\frac{1}{N_{\text{bin}}(N_{\text{bin}}-1)} \sum_{b=1}^{N_{\text{bin}}} (\bar{O}_b - \langle \mathcal{O} \rangle)^2}. \quad (18)$$

4 Variational wave functions

4.1 Jastrow-Slater wave functions for the single-orbital Hubbard model

The Jastrow-Slater wave function can be generically written as

$$|\Psi_J\rangle = \mathcal{J}|\Phi_0\rangle, \quad (19)$$

where $\mathcal{J}$ is a Jastrow factor that embeds correlation in the wave function and $|\Phi_0\rangle$ is the ground state of a noninteracting quadratic auxiliary Hamiltonian $\mathcal{H}_{\text{aux}}$, according to the definition

$$|\Phi_0\rangle = \prod_{k=1}^{N_e} \varphi_k^\dagger |0\rangle, \quad (20)$$

where N_e is the total number of electrons and φ_k are the eigenstates of $\mathcal{H}_{\text{aux}}$. Indeed, we can easily diagonalize any Hamiltonian that can be written in a simple quadratic form as

$$\mathcal{H}_{\text{aux}} = \mathbf{d}^\dagger \mathbf{T} \mathbf{d}, \quad (21)$$

where $\mathbf{d}^\dagger = (\mathbf{d}_\uparrow^\dagger, \mathbf{d}_\downarrow^\dagger)$ is a vector with $2L$ components ($\mathbf{d}_\sigma^\dagger = d_{1,\sigma}^\dagger, \dots, d_{L,\sigma}^\dagger$), $\mathbf{T}$ is a $2L \times 2L$ matrix, and L is the number of lattice sites. By introducing a unitary matrix $\mathbf{U}$ (that preserves the anticommutation rules of fermions), we can diagonalize $\mathcal{H}_{\text{aux}}$

$$\mathcal{H}_{\text{aux}} = \mathbf{d}^\dagger \mathbf{U} \mathbf{U}^\dagger \mathbf{T} \mathbf{U} \mathbf{U}^\dagger \mathbf{d} = \mathbf{\Phi}^\dagger \mathbf{E} \mathbf{\Phi} = \sum_k \varepsilon_k \varphi_k^\dagger \varphi_k, \quad (22)$$

where $\mathbf{E}$ is a diagonal matrix that contains the eigenvalues and $\mathbf{\Phi} = (\varphi_1, \dots, \varphi_{2L})$ is defined in terms of the eigenvectors.

The choice of the noninteracting quadratic Hamiltonian is a careful aspect, since we need to include in it the relevant physical terms that we want to study in our model. For instance, if one would be interested in studying the emergence of superconductivity in the Hubbard model, when varying the electronic density, and its competition with antiferromagnetic Néel order, one could define the following auxiliary Hamiltonian

$$\mathcal{H}_{\text{aux}} = \mathcal{H}_0 + \mathcal{H}_{AF} + \mathcal{H}_{BCS}. \quad (23)$$

The first term consists of the *kinetic energy* of the electrons in the Hubbard model

$$\mathcal{H}_0 = -t \sum_{\langle R, R' \rangle, \sigma} c_{R\sigma}^\dagger c_{R'\sigma} - \tilde{t}' \sum_{\langle\langle R, R' \rangle\rangle, \sigma} c_{R\sigma}^\dagger c_{R'\sigma} + \text{H.c.} \quad (24)$$

The second term includes antiferromagnetism by *standard Néel order*

$$\mathcal{H}_{AF} = \Delta_{AF} \sum_R (-1)^{x+y} \left(c_{R\uparrow}^\dagger c_{R\uparrow} - c_{R\downarrow}^\dagger c_{R\downarrow} \right), \quad (25)$$

while the last one introduces the *BCS electron pairing*

$$\mathcal{H}_{BCS} = \sum_{R, \eta=x, y} \Delta_\eta \left(c_{R,\uparrow}^\dagger c_{R+\eta,\downarrow}^\dagger - c_{R,\downarrow}^\dagger c_{R+\eta,\uparrow}^\dagger \right) + \text{H.c.} - \mu \sum_{R, \sigma} c_{R\sigma}^\dagger c_{R\sigma}, \quad (26)$$

where μ is a chemical potential and $\Delta_x = -\Delta_y$ represents the usual d -wave symmetry that is realized in the single-band Hubbard model. Further developments can be included to describe the striped modulations of charge and spin that are a common feature in the ground state of the single-orbital Hubbard model on the square lattice [14–16]. The parameters $\tilde{t}'$, Δ_{AF} , $\Delta\eta$, μ are variational parameters that can be optimized, as described in Sec. 5, while we keep $t = 1$ to fix the energy scale.

The reader could now wonder how a BCS state can be written as a Slater determinant. This is indeed possible, via a convenient particle-hole canonical transformation of the annihilation and creation operators for down spins, i.e.

$$d_{R,\uparrow}^\dagger \equiv c_{R,\uparrow}^\dagger \quad d_{R,\downarrow} \equiv c_{R,\downarrow}^\dagger. \quad (27)$$

The products of creation and annihilation operators that build up the auxiliary Hamiltonian are then rewritten as

$$\begin{aligned} c_{R\uparrow}^\dagger c_{R'\uparrow} &= d_{R\uparrow}^\dagger d_{R'\uparrow} & c_{R\downarrow}^\dagger c_{R'\downarrow} &= -d_{R'\downarrow}^\dagger d_{R\downarrow} \\ c_{R,\uparrow}^\dagger c_{R+\eta,\downarrow}^\dagger &= c_{R,\uparrow}^\dagger c_{R+\eta,\downarrow} & c_{R,\downarrow}^\dagger c_{R+\eta,\uparrow}^\dagger &= -c_{R+\eta,\uparrow}^\dagger c_{R,\downarrow}. \end{aligned} \quad (28)$$

In this way, the eigenstates of the BCS Hamiltonian can be expressed as “orbitals”, similarly to the ones that can be obtained by diagonalizing the kinetic term.

It should be noted that the particle-hole transformation changes the total number of electrons per spin species N_σ^c , according to the following relations

$$\begin{aligned} N_\uparrow^c &= \sum_R c_{R,\uparrow}^\dagger c_{R,\uparrow} = \sum_R d_{R,\uparrow}^\dagger d_{R,\uparrow} = N_\uparrow^d \\ N_\downarrow^c &= \sum_R c_{R,\downarrow}^\dagger c_{R,\downarrow} = L - \sum_R d_{R,\downarrow}^\dagger d_{R,\downarrow} = L - N_\downarrow^d, \end{aligned} \quad (29)$$

where L is the total number of sites.

Whenever the total magnetization is zero, i.e., $N_\uparrow^c = N_\downarrow^c$, the total number of electrons after the particle-hole transformation, given by $N_\uparrow^d + N_\downarrow^d$, is equal to L , that is, the system, after the particle-hole transformation, is always half-filled.

The *Jastrow factor* $\mathcal{J}$ is a long-range correlator that, on the lattice, takes the following form

$$\mathcal{J} = \exp \left[-\frac{1}{2} \sum_{R,R'} v_{R,R'} (n_R - n)(n_{R'} - n) \right], \quad (30)$$

where $v_{R,R'}$ is a pseudo-potential for density-density correlations in the variational state, n_R is the electronic density on site R , and n is the average electronic density. For translationally invariant models, like the Hubbard Hamiltonian of Eq. (1), $v_{R,R'}$ only depends on the relative distance of the two sites $|\mathbf{R} - \mathbf{R}'|$. The Jastrow pseudo-potential can be optimized for all possible distances, which are $O(L)$ in translationally invariant systems, via the stochastic reconfiguration method of Sec. 5. The role of the long-range tail of the Jastrow factor is to create a bound state between doubly occupied sites and empty sites, thus not allowing for conduction, but still

allowing local density fluctuations. Indeed, it has been shown that the Jastrow term may turn a non-interacting metallic state $|\Phi_0\rangle$ at half filling into a Mott insulator, that is, an insulator driven by electronic correlation, even in the absence of magnetic order [17]. We mention the existence of a deep connection between a Mott insulator at half-filling and a superconductor at finite doping in the Resonant Valence Bond (RVB) theory of high-temperature superconductivity proposed by Anderson [18]. He suggested that a superconducting phase may emerge when doping a Mott insulator with “preformed” singlet pairs. In fact, the RVB state describes an insulator that is a liquid of spin singlets, which become mobile when the system is doped, thus forming actual superconducting pairs.

We remark that replacing the Jastrow factor with a much simpler term that penalizes the formation of double occupancies is not able to reproduce a Mott insulator unless $U = \infty$. This simplified interaction term, called the Gutzwiller factor, is defined as [19]

$$|\Psi_G\rangle = P_G|\Phi_0\rangle \quad P_G = \exp \left[-\frac{g}{2} \sum_R (n_R - n)^2 \right] \quad (31)$$

and corresponds to the $R = R'$ case of the Jastrow factor. The effect of the Gutzwiller factor becomes clear once the variational state is expanded in a basis set whose elements $\{|x\rangle\}$ represent configurations with electron occupations of each lattice site. Since the Gutzwiller factor is diagonal in this basis, we have that:

$$\langle x|\Psi_G\rangle = P_G(x)\langle x|\Phi_0\rangle, \quad (32)$$

where $P_G(x) \leq 1$ is a number that depends on how many doubly-occupied sites are present in a given electronic configuration. Therefore, the amplitude of the non-interacting state $|\Phi_0\rangle$ is renormalized by $P_G(x)$. At half filling, density excitations are represented by doublons (doubly-occupied sites) and holons (empty sites). In the absence of electronic correlation, these objects are free to move and are then responsible for conductivity (for example, in the fermionic model, a doublon is negatively charged with respect to the average background, while the holon is positively charged). The effect of the Gutzwiller factor is to penalize the formation of such excitations; however, once created, doublons and holons are no longer correlated, thus being free to move independently. Only when the energetic penalty is infinite, an insulator is obtained. On the contrary, the long-range tail of the Jastrow factor allows to correlate holons and doublons and keeps them bound together. This fact can be understood via a classical mapping to a model of charged particles, where holons and doublons represent positive and negative charges [1].

4.2 Computing ratios for the Jastrow-Slater wave function

Both in the Metropolis algorithm of Eq. (15) and in the evaluation of the local observables of Eq. (8), we need to compute ratios of amplitudes of the variational state with respect to a generic element of the basis set, i.e., $\langle x'|\Psi\rangle/\langle x|\Psi\rangle$.

The advantage of considering Jastrow-Slater wave functions in the variational Monte Carlo technique is that the calculations can be efficient and fast. Indeed, since the Jastrow factor is

diagonal in the chosen basis, we have that

$$\langle x|\Psi\rangle = \mathcal{J}(x)\langle x|\Phi_0\rangle, \quad (33)$$

where $\mathcal{J}(x)$ is the value of the Jastrow operator computed for the configuration $|x\rangle$, i.e., $\mathcal{J}|x\rangle = \mathcal{J}(x)|x\rangle$. Therefore, given the electronic configuration, $\mathcal{J}(x)$ is a number that can be evaluated in $O(L^2)$ operations. Moreover, $\langle x|\Phi_0\rangle$ can be evaluated in $O(L^3)$ operations since it is a Slater determinant.

Let us now briefly discuss how to compute ratios for the two cases. First, let us focus on the ratio of Jastrow factors $\mathcal{J}(x')/\mathcal{J}(x)$ and show that it requires only $O(L)$ operations. We focus on a simplified form for the Jastrow factor

$$\mathcal{J} = \exp \left[-\frac{1}{2} \sum_{R,R'} v_{R,R'} n_R n_{R'} \right], \quad (34)$$

which is equivalent to the one of Eq. (30) when the particle number is conserved (apart from a multiplicative factor).

When $|x'\rangle$ is close to $|x\rangle$, i.e.

$$|x\rangle \equiv |n_1, \dots, n_k, \dots, n_l, \dots, n_L\rangle \quad |x'\rangle \equiv |n_1, \dots, n_k+1, \dots, n_l-1, \dots, n_L\rangle, \quad (35)$$

it is possible to define a fast computation

$$\frac{\mathcal{J}(x')}{\mathcal{J}(x)} = \frac{\mathcal{J}(n_1, \dots, n_k+1, \dots, n_l-1, \dots, n_L)}{\mathcal{J}(n_1, \dots, n_k, \dots, n_l, \dots, n_L)} = \frac{\exp(-\sum_R v_{R,k} n_R)}{\exp(-\sum_R v_{R,l} n_R)} e^{v_{k,l}-v_{k,k}}, \quad (36)$$

where we used the fact that the pseudo-potential is symmetric and translationally invariant. The term $e^{v_{k,l}-v_{k,k}}$ in Eq. (36) does not depend upon the electronic configuration and can be computed at the beginning of the simulation, once for all. Instead, the ratio of the exponentials depends on the fermionic configuration, which is sampled along the Markov chain. However, the computation of the ratio can be done in $O(1)$ operations once we compute and store a vector of dimension L that depends on $|x\rangle$, defined as

$$T_{\text{Jastrow}}(R') = \sum_R v_{R,R'} n_R. \quad (37)$$

Then, updating $T_{\text{Jastrow}}(R')$ to the new configuration $|x'\rangle$ costs $O(L)$:

$$n'_R = n_R + \delta_{k,R} - \delta_{l,R} \Rightarrow T'_{\text{Jastrow}}(R') = T_{\text{Jastrow}}(R') + v_{k,R'} - v_{l,R'}. \quad (38)$$

Therefore, we can compute from scratch $T_{\text{Jastrow}}(R')$ at the beginning of the Markov chain for all the sites $R' = 1, \dots, L$ and then update it every time a new configuration is accepted along the Markov chain. From time to time, it is recommended to recompute Eq. (37) from scratch, since the rounding error of the update can accumulate and give rise to numerical errors.

Let us now move to the $\langle x|\Phi_0\rangle$ term, and show that it indeed corresponds to a Slater determinant that can be computed in $O(L^3)$ operations with standard linear algebra routines. First of all, let

us contract the site and spin indices of the auxiliary quadratic Hamiltonian $\mathcal{H}_{\text{aux}}$ into a single index I from 1 to $2L$: $c_{R,\uparrow} \equiv d_R$, $c_{R,\downarrow} \equiv d_{R+L}$. Then, we recall that $\mathcal{H}_{\text{aux}}$ can be diagonalized and its eigenstates are the single-particle orbitals $\varphi_k^\dagger|0\rangle$

$$\mathcal{H}_{\text{aux}} = \sum_k \varepsilon_k \varphi_k^\dagger \varphi_k, \quad (39)$$

with ε_k the single-particle energies and $\varphi_k^\dagger = \sum_I U_{I,k} d_I^\dagger$ (where $\mathbf{U}$ is the unitary matrix that diagonalizes $\mathcal{H}_{\text{aux}}$). Then, the many-particle state $|\Phi_0\rangle$ can be constructed by occupying the N_e lowest single-particle orbitals, where N_e is the number of electrons

$$|\Phi_0\rangle = \prod_{k=1}^{N_e} \varphi_k^\dagger |0\rangle = \prod_{k=1}^{N_e} \left(\sum_I U_{I,k} d_I^\dagger \right) |0\rangle. \quad (40)$$

A generic electronic configuration for N_e electrons in the lattice, visited along the Markov chain, can be written as

$$|x\rangle = d_{R_1}^\dagger \dots d_{R_{N_e}}^\dagger |0\rangle, \quad (41)$$

where a generic R_j assumes values from 1 to $2L$: the positions of spin-up electrons coincide with the site number, while the positions of spin-down electrons must be shifted by L . The overlap $\langle x|\Phi_0\rangle$ can then be written as

$$\begin{aligned} \langle x|\Phi_0\rangle &= \langle 0|d_{R_{N_e}} \dots d_{R_1} \left(\sum_I U_{I,1} d_I^\dagger \right) \dots \left(\sum_I U_{I,N_e} d_I^\dagger \right) |0\rangle = \\ &= \langle 0|d_{R_{N_e}} \dots d_{R_1} \left[\sum_p (-1)^p \prod_{k=1}^{N_e} U_{p\{R_j\},k} \right] d_{R_1}^\dagger \dots d_{R_{N_e}}^\dagger |0\rangle. \end{aligned} \quad (42)$$

The sum in $[\dots]$ is over all the permutations of the $\{R_j\}$ in $|x\rangle$ (with the sign from anticommutations), thus implying that $\langle x|\Phi_0\rangle = \det\{U_{R_j,k}\}$. In practice, the diagonalization of the auxiliary Hamiltonian can be done at the beginning of the simulation, obtaining the matrix of the eigenvectors $\mathbf{U}$, of which we store only the N_e columns corresponding to the occupied orbitals. Then, the overlap $\langle x|\Phi_0\rangle$ with the generic configuration $|x\rangle$ is given by the determinant of the matrix $\tilde{\mathbf{U}}$

$$\tilde{\mathbf{U}} = \begin{pmatrix} U_{R_1,1} & \dots & U_{R_1,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_l,1} & \dots & U_{R_l,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_{N_e},1} & \dots & U_{R_{N_e},N_e} \end{pmatrix}, \quad (43)$$

obtained by taking only the rows corresponding to the occupied sites of the matrix $\mathbf{U}$. Now, if we consider a new electronic configuration $|x'\rangle$, obtained from $|x\rangle$ after the hopping of the l -th electron from position R_l to position $R_{l'}$, i.e., $|x'\rangle = d_{R_{l'}}^\dagger d_{R_l} |x\rangle$, we obtain a new matrix $\tilde{\mathbf{U}}'$ that

will differ from $\tilde{\mathbf{U}}$ for the l -th row

$$\tilde{\mathbf{U}}' = \begin{pmatrix} U_{R_1,1} & \cdots & U_{R_1,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_l',1} & \cdots & U_{R_l',N_e} \\ \vdots & \ddots & \vdots \\ U_{R_{N_e},1} & \cdots & U_{R_{N_e},N_e} \end{pmatrix}. \quad (44)$$

Then, the ratio between two configurations that differ only by the hopping of a single electron is given by

$$\frac{\langle x' | \Phi_0 \rangle}{\langle x | \Phi_0 \rangle} = \frac{\langle x | d_{R_l}^\dagger d_{R_l'} | \Phi_0 \rangle}{\langle x | \Phi_0 \rangle} = \frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}}. \quad (45)$$

It is then possible to show that this ratio of determinants can be computed in only $O(L^2)$ operations, thus speeding up the simulation with respect to a full calculation of a determinant from scratch [1].

4.3 The backflow terms

While the Jastrow-Slater wave function is able to capture Mott insulating and superconductive states, it has a strong intrinsic limitation: Electronic correlations are introduced only via a multiplicative factor, while the nodes of the variational state are the ones set by the uncorrelated state $|\Phi_0\rangle$. In order to overcome this limitation, we have introduced a novel variational state that redefines the single-particle orbitals of the uncorrelated state, taking into account the electronic repulsion, the so-called backflow correlations [20, 21].

Let us consider the single-particle eigenstates $\varphi_k(R, \sigma)$ of the auxiliary Hamiltonian $\mathcal{H}_{\text{aux}}$, written in the basis of lattice sites and z component of the spins. Backflow correlations modify them, according to the electronic configuration on the lattice

$$\varphi_k^b(R, \sigma) \equiv \tilde{\varepsilon} \varphi_k(R, \sigma) + \eta_1 \sum_{R' \text{ n.n. } R} D_R H_{R'} \varphi_k(R', \sigma) + \eta_2 \sum_{R' \text{ n.n.n. } R} D_R H_{R'} \varphi_k(R', \sigma), \quad (46)$$

where $\tilde{\varepsilon} = \varepsilon$ if the site R is doubly occupied and is surrounded by at least one empty site, while $\tilde{\varepsilon} = 1$ in all the other cases. Here ε, η_l , ($l = 1, 2$) are variational parameters to be optimized, according to the method defined in Sec. 5. The operators $D_R = n_{R,\uparrow} n_{R,\downarrow}$ and $H_R = h_{R,\uparrow} h_{R,\downarrow}$, with $h_{R,\sigma} = 1 - n_{R,\sigma}$, are non-zero only if the site R is doubly occupied or empty, respectively. Thus, they are the operators that allow us to include many-body correlations inside the single-particle orbitals. Moreover, the shorthand notations n.n. and n.n.n. indicate nearest- and next-nearest-neighbor sites, respectively. In this way, already the determinant part of the wave function includes correlation effects, strongly improving the accuracy of the many-body state. This is a substantial improvement with respect to Jastrow factors, where electron-electron correlation is included only via a multiplicative term.

The terms in Eq. (46) are the dominant contributions to the backflow operator, having a qualitative influence on the properties of the correlated wave functions. In addition, we can also take

into account further terms that are useful in the intermediate-coupling regime and correspond to all possible hopping processes

$$\begin{aligned}
\varphi_k^b(R, \sigma) \equiv & \tilde{\varepsilon} \varphi_k(R, \sigma) + \eta_1 \sum_{R' \text{ n.n. } R} D_R H_{R'} \varphi_k(R', \sigma) \\
& + \eta_2 \sum_{R' \text{ n.n.n. } R} D_R H_{R'} \varphi_k(R', \sigma) \\
& + \beta_1 \sum_{R' \text{ n.n. } R} n_{R, \sigma} h_{R, -\sigma} n_{R', -\sigma} h_{R', \sigma} \varphi_k(R', \sigma) \\
& + \beta_2 \sum_{R' \text{ n.n.n. } R} n_{R, \sigma} h_{R, -\sigma} n_{R', -\sigma} h_{R', \sigma} \varphi_k(R', \sigma) \\
& + \gamma_1 \sum_{R' \text{ n.n. } R} (D_R n_{R', -\sigma} h_{R', \sigma} + n_{R, \sigma} h_{R, -\sigma} H_{R'}) \varphi_k(R', \sigma) \\
& + \gamma_2 \sum_{R' \text{ n.n.n. } R} (D_R n_{R', -\sigma} h_{R', \sigma} + n_{R, \sigma} h_{R, -\sigma} H_{R'}) \varphi_k(R', \sigma), \quad (47)
\end{aligned}$$

where, in addition to ε and $\eta_l, \beta_l, \gamma_l$ ($l=1, 2$) are variational parameters to be optimized. We mention that the computation of $\langle x | \Phi_0 \rangle$ and of the determinant ratios introduced in the previous subsection can be extended to the presence of backflow correlations with a similar degree of complexity [1].

Let us now present a few numerical results that show the relevance of backflow correlations in describing electronic correlations. We focus on the half-filled case, where the effect of electronic correlations is maximal, and we consider the simpler single-orbital Hubbard model. Moreover, we work in the so-called “nonmagnetic sector”, that is, neglecting any possible magnetic order in the variational wave function. In this way, we focus on the appearance of Mott insulating states that are driven solely by electronic correlations.

First, let us discuss the range of backflow correlations. We present the results in Fig. 1, left panels, for a 2D square lattice with both nearest neighbor hopping t and next-nearest neighbor hopping t' . We report that backflow parameters are irrelevant for all distances larger than the ones connected by the hopping amplitude. Most importantly, a remarkable gain in energy is obtained only by a suitable optimization of the backflow parameters up to second neighbors, while poor results are obtained when there is no backflow (or its range is smaller than the one of the hopping). Moreover, in Fig. 1, right panel, we present the effect of backflow correlations in the large- U limit. We consider a one-dimensional Hubbard model, which is the same as the one introduced in Eq. (1), just denoting the nearest-neighbor hopping with t_1 and the next-nearest-neighbor hopping with t_2 . Only the presence of backflow correlations in the wave function allows for a proper extrapolation to the infinite- U limit of the Hubbard model, that is, the Heisenberg model. Here, the energy of the Heisenberg model is provided by density-matrix renormalization group (DMRG) calculations, which are numerically exact for 1D models. We also compare the variational results with the exact ones on the 18-site cluster. In Fig. 2, we show the accuracy of the variational state (with and without backflow correlations) and the overlap with the exact ground state for two hopping values, i.e., $t'/t = 0$ and $t'/t = 0.7$. The backflow term is able to highly improve the accuracy both for weak and strong couplings. Moreover,

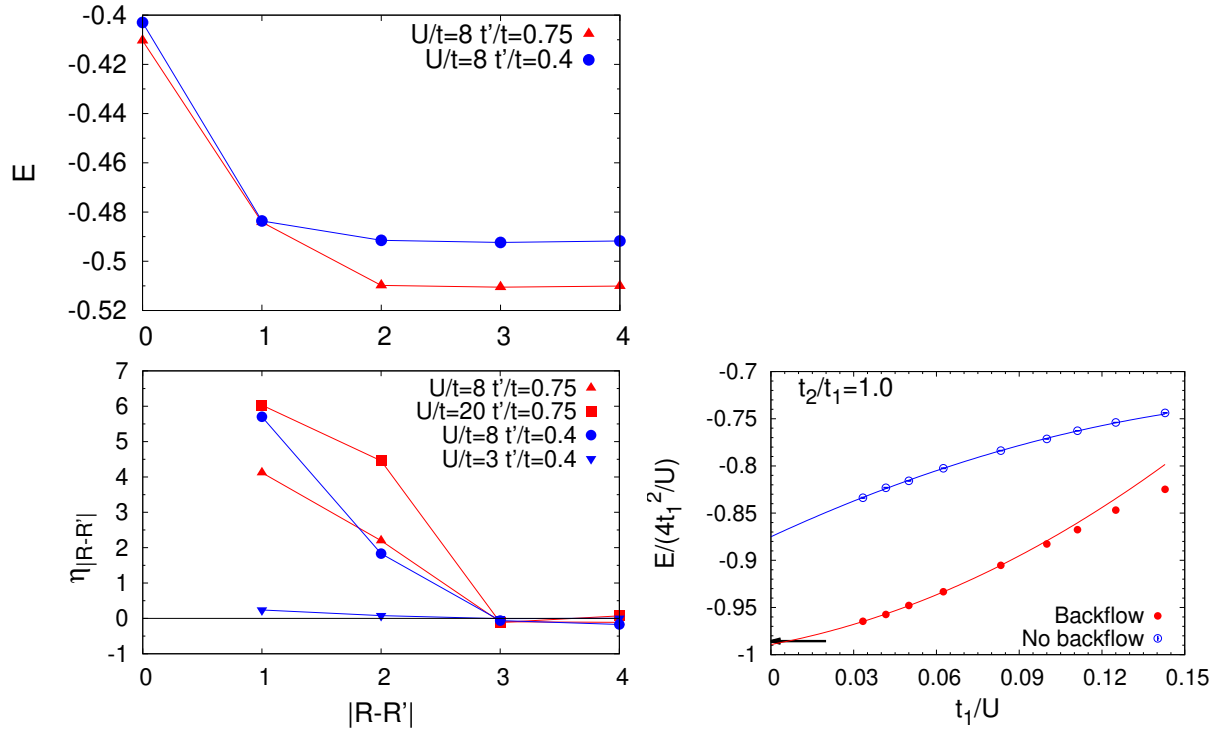


Fig. 1: Upper left panel: Ground-state energy as a function of the range of backflow parameters on the 2D square lattice for $t'/t = 0.4$ and 0.75 at $U/t = 8$. A range equal to zero corresponds to no backflow correlations in the wave function. Lower left panel: Optimized backflow parameters $\eta|_{R-R'}$ as a function of the distance $|R-R'|$ between lattice sites. Results for $U/t = 3$ and $t'/t = 0.4$ correspond to a metallic state, the others to Mott insulating states. Data are shown for a lattice with $L = 98$ sites [21]. Right panel: Energy (in units of $4t_1^2/U$), as a function of t_1/U , for the 1D Hubbard model with $t_2/t_1 = 1$. Filled (empty) circles denote the results with (without) backflow correlations. The DMRG energy of the corresponding Heisenberg model with $J_2/J_1 = 1$ is shown by an arrow [22]. Data are presented for an $L = 120$ chain [21].

the overlap between the exact ground state and the backflow state remains very high, even for large values of U/t , and the improvement with respect to the state without backflow is crucial, especially when a finite value of t' is included.

These results suggest that backflow correlations are indeed necessary to properly describe electronic correlations. In particular, backflow correlations turned out to be particularly useful in the presence of frustration [20, 23]. We do not discuss frustration in this lecture, but let us just say that it is related to the presence of competing hopping parameters in Mott insulators, like the case of square lattices with both nearest neighbor and next-nearest neighbor hopping parameters. The presence of frustration may destroy magnetic orders and lead to the formation of spin liquid phases, i.e., Mott insulating states without long-range magnetic order. Moreover, the backflow wave function has been the starting point for constructing a new class of wave functions, named neural network backflow, which uses a neural network to learn the optimal transformation via variational Monte Carlo calculations [24].

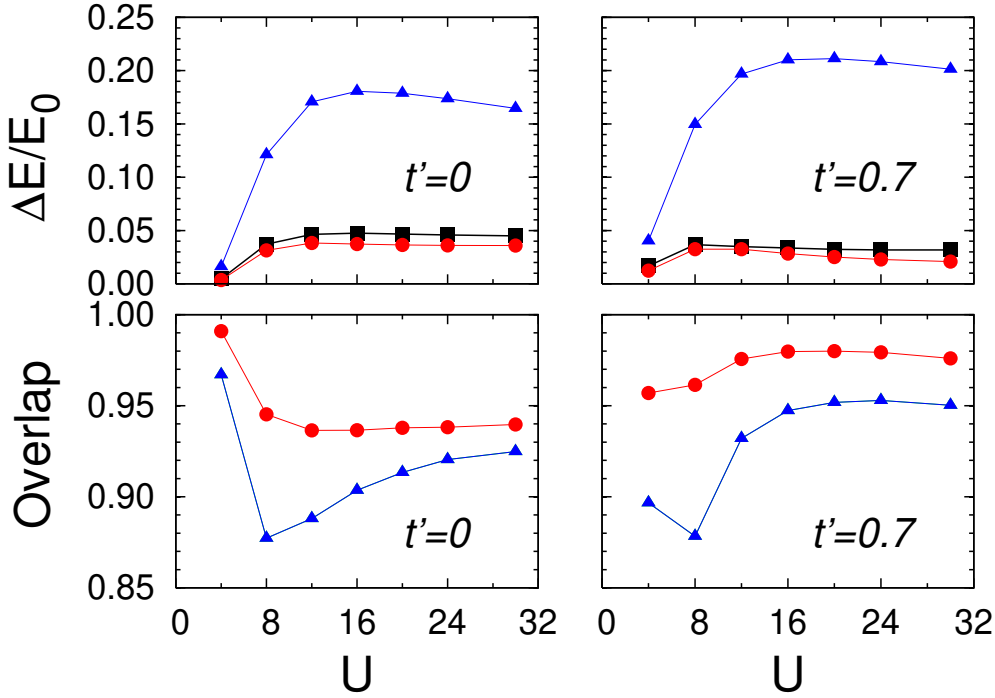


Fig. 2: Results for 18 electrons on 18 sites as a function of U/t . Upper panels: Accuracy of energy $\Delta E = (E_v - E_0)$, E_v and E_0 being the variational and the exact values, respectively. Lower panels: Overlap between the exact ground state and the variational wave functions. The variational state without backflow is denoted by blue triangles, the variational state with backflow correlations is denoted by red circles [20].

4.4 Jastrow-Slater wave functions for the multi-orbital Hubbard model

The Jastrow-Slater wave function for a multi-orbital Hubbard model can be generalized in the following form [25, 26]

$$|\Psi\rangle = \mathcal{J}_U \mathcal{J}_{U'} \mathcal{J}_J |\Phi_0\rangle. \quad (48)$$

The Jastrow factor has been generalized as

$$\begin{aligned} \mathcal{J}_U &= \exp \left(-\frac{1}{2} \sum_{\alpha} \sum_{R,R'} v_{R,R'}^{\alpha,\alpha} n_{\alpha,R} n_{\alpha,R'} \right) \\ \mathcal{J}_{U'} &= \exp \left(-\frac{1}{2} \sum_{\alpha \neq \beta} \sum_{R,R'} v_{R,R'}^{\alpha,\beta} n_{\alpha,R} n_{\beta,R'} \right). \end{aligned} \quad (49)$$

When $\alpha = \beta$, the Jastrow factor is intra-orbital with a pseudopotential $v_{R,R'}^{\alpha,\alpha}$, whereas for $\alpha \neq \beta$ the electronic correlations are inter-orbital and its pseudopotential is written as $v_{R,R'}^{\alpha,\beta}$. The Jastrow factors mimic the Coulomb interaction present in the Hubbard-Kanamori Hamiltonian of Eq. (2). Specifically, $\mathcal{J}_U$ and $\mathcal{J}_{U'}$ take into account the presence of the intra-orbital repulsion U and the inter-orbital repulsion U' , respectively. In order to take into account the presence of the Hund's coupling in the wave function, we define a spin-spin Jastrow factor that favors configu-

rations where electron spins in different orbitals (and same site) are aligned

$$\mathcal{J}_J = \exp \left(-\frac{1}{2} \sum_{\alpha \neq \beta} u^{\alpha, \beta} \sum_R S_{\alpha, R}^z S_{\beta, R}^z \right). \quad (50)$$

Here, the correlator involves the z component of the spin $S_{\alpha, R}^z$ only for different orbitals on site R , as suggested by the definition of the Hund's coupling.

The uncorrelated state $|\Phi_0\rangle$ is now the ground state of a new quadratic auxiliary Hamiltonian, which is appropriate for the multi-orbital model. Focusing only on the investigation of superconductivity in the model and neglecting any possible magnetic order, the auxiliary Hamiltonian can be defined as

$$\mathcal{H}_{\text{aux}} = \mathcal{H}_0 + \mathcal{H}_{BCS}, \quad (51)$$

where $\mathcal{H}_0$ is the same as the tight-binding Hamiltonian of Eq. (3), while $\mathcal{H}_{BCS}$ is now defined as

$$\mathcal{H}_{BCS} = - \sum_{R, \alpha, \sigma} \mu_\alpha c_{R, \alpha, \sigma}^\dagger c_{R, \alpha, \sigma} + \sum_{R, \alpha, \delta} [\Delta_{\alpha, \delta} (c_{R, \alpha, \uparrow} c_{R+\delta, \alpha, \downarrow} - c_{R, \alpha, \downarrow} c_{R+\delta, \alpha, \uparrow}) + \text{h.c.}], \quad (52)$$

where $\delta = x, y, x+y$, and $x-y$ indicates nearest and next-nearest neighbors of the site R and μ_α defines the chemical potential of orbital α . The (singlet) pairing amplitudes $\Delta_{\alpha, \delta}$ and μ_α are optimized, while the hopping parameters in $\mathcal{H}_0$ are usually kept fixed to the bare Hamiltonian ones. Inter-orbital pairing amplitudes are found to be negligible in the optimal wave function, in agreement with [27] and therefore not included in the auxiliary Hamiltonian.

5 Optimizing the variational state: Stochastic reconfiguration

We have introduced the functional form of the wave function, which depends on many variational parameters, ranging from the pseudopotentials in the Jastrow factor to the BCS coupling or the hopping parameters in the auxiliary Hamiltonian. This freedom in the choice of the variational wave function is meaningful only if we can construct a procedure that allows the optimization of many variational parameters at the same time, in order to find the best approximation of the ground state. One possible way to optimize the set of parameters is to use the zero-variance property, minimizing the variance of the energy in the variational state [28]. In contrast, the *stochastic reconfiguration* exploits the minimization of the variational energy; unlike the minimization of variance, where the minimum value is zero, the ground state energy is unknown a priori [4]. However, for a given form of the trial wave function, energy-minimized wave functions yield more accurate values of expectation values than wave functions whose variance is minimized [29].

Let us explain the stochastic reconfiguration assuming a generic functional form for the wave function $|\Psi(\{p_k\})\rangle$ with $k = 1, \dots, N_p$, which depends on a set of N_p variational parameters

$\{p_k\}$. The variational energy of the state can be decreased by the application of the projection operator $(\mathcal{H}-\Lambda)$, where Λ is an appropriate constant energy shift

$$|\Psi_\Lambda\rangle = (\mathcal{H}-\Lambda) |\Psi^0\rangle, \quad (53)$$

where $|\Psi^0\rangle = |\Psi(\{p_k^0\})\rangle$ is the starting state, depending on an initial set of variational parameters $\{p_k^0\}$. However, in general, the new state $|\Psi_\Lambda\rangle$ will not be in the Jastrow-Slater form. Therefore, the idea is to find the quantum state $|\Psi'\rangle = |\Psi(\{p'_k\})\rangle$, which gives the best possible approximation of $|\Psi_\Lambda\rangle$. If the change of the variational parameters is taken to be small, we can approximate the new parameters as $p'_k = p_k^0 + \delta p_k$, and the new state $|\Psi'\rangle$ can be expanded at first order in $\{\delta p_k\}$

$$|\Psi'\rangle = \delta p_0 |\Psi^0\rangle + \sum_{k=1}^{N_p} \delta p_k \frac{\partial |\Psi^0\rangle}{\partial p_k} + O(\delta p_k^2). \quad (54)$$

Here, the series is truncated to the second order, and δp_0 is considered to match the normalization condition. Then, we define the local operator $\mathcal{O}_k$ (with $k = 1, \dots, N_p$) on the configuration $|x\rangle$ as the logarithmic derivative with respect to the variational parameter p_k

$$\mathcal{O}_k = \frac{1}{\Psi^0(x)} \frac{\partial \Psi^0(x)}{\partial p_k} = \frac{\partial \ln \langle x | \Psi^0 \rangle}{\partial p_k}. \quad (55)$$

Imposing $\mathcal{O}_0 = 1$, $|\Psi'\rangle$ can be expressed by using the logarithmic derivative operators

$$|\Psi'\rangle = \sum_{k=0}^{N_p} \delta p_k \mathcal{O}_k |\Psi^0\rangle. \quad (56)$$

Then, the stochastic reconfiguration method imposes that the projection of the shifted state $|\Psi_\Lambda\rangle$ and of the new state $|\Psi'\rangle$ on the subspace $\mathcal{O}_j |\Psi^0\rangle$ is equal for $j = 0 \dots N_p$

$$\langle \Psi^0 | \mathcal{O}_j | \Psi' \rangle = \langle \Psi^0 | \mathcal{O}_j | \Psi_\Lambda \rangle. \quad (57)$$

This ensures that the energy computed on $|\Psi'\rangle$ is lower than the one computed on $|\Psi^0\rangle$. Then, by inserting the definition of $|\Psi_\Lambda\rangle$ from Eq. (53) and the one of $|\Psi'\rangle$ from Eq. (56), the stochastic reconfiguration condition becomes

$$\sum_{k=0}^{N_p} \delta p_k \langle \mathcal{O}_j \mathcal{O}_k \rangle = \langle \mathcal{O}_j (\Lambda - \mathcal{H}) \rangle. \quad (58)$$

The symbol $\langle \dots \rangle$ stands for $\langle \Psi^0 | \dots | \Psi^0 \rangle$. Subsequently, the conditions for $j=0$ and $j \neq 0$ are separated, and the $k=0$ term is extracted from the summation, leading to the following relations

$$\delta p_0 = - \sum_{k=1}^{N_p} \delta p_k \langle \mathcal{O}_k \rangle + \Lambda - \langle \mathcal{H} \rangle \quad \text{and} \quad \delta p_0 \langle \mathcal{O}_j \rangle + \sum_{k=1}^{N_p} \delta p_k \langle \mathcal{O}_j \mathcal{O}_k \rangle = \Lambda \langle \mathcal{O}_j \rangle - \langle \mathcal{O}_j \mathcal{H} \rangle. \quad (59)$$

The parameter δp_0 , related to normalization, does not affect any physical observable; indeed, the two conditions can be unified in a compact form, in the unknowns $\{\delta p_k\}$ as

$$\sum_{k=1}^{N_p} \delta p_k S_{jk} = f_j, \quad (60)$$

where $f_j = \langle \mathcal{O}_j \rangle \langle \mathcal{H} \rangle - \langle \mathcal{O}_j \mathcal{H} \rangle$ is the generalized force and $S_{jk} = \langle \mathcal{O}_j \mathcal{O}_k \rangle - \langle \mathcal{O}_j \rangle \langle \mathcal{O}_k \rangle$ is the covariance matrix of logarithmic derivatives of $N_p \times N_p$ dimension (symmetric and semi-positive definite). During the Monte Carlo procedure, the expectation values are computed by stochastic sampling, the linear system is numerically solved, and, successively, parameters are updated according to the following scheme

$$\delta p_k = \sum_{j=1}^{N_p} S_{jk}^{-1} f_j, \quad p'_k = p_k^0 + \tau \delta p_k. \quad (61)$$

τ is the acceleration parameter that determines the convergence of the algorithm (in principle, each parameter could have a different τ_k). A large τ guarantees rapid convergence, although it could threaten the stability of the algorithm. Hence, τ is chosen sufficiently small to generate a variational energy that can be expanded in series

$$E(\Psi') = E(\Psi_0) + \tau \sum_{k=0}^{N_p} \frac{\partial E(\Psi_0)}{\partial p_k} \delta p_k + O(\tau^2). \quad (62)$$

Therefore, by using the definition of δp_k in Eq. (61) and the fact that the derivative of the energy with respect to a parameter is equal to minus the generalized force ($\frac{\partial E(\Psi_0)}{\partial p_k} = -f_k$), we can rewrite the change in the variational energy as

$$\delta E = E(\Psi') - E(\Psi_0) = -\tau \sum_{j,k=1}^{N_p} S_{jk}^{-1} f_j f_k + O(\tau^2). \quad (63)$$

Here, the decrease of the variational energy ($\delta E \leq 0$) is ensured by the semi-positive definition of the covariance matrix of the logarithmic derivatives S_{jk} .

Examples of calculations of logarithmic derivative operators for Jastrow-Slater wave functions can be found in Refs. [1, 30]

6 Correlation functions in variational Monte Carlo: Definitions and examples

In this section, we introduce a few relevant correlators, which expectation values over the optimal variational state allow us to describe the properties of our variational approximation of the ground-state wave function.

The first one is the *static structure factor* $N(\mathbf{q})$ defined in reciprocal space as:

$$N(\mathbf{q}) = \langle n_{\mathbf{q}} n_{-\mathbf{q}} \rangle = \frac{1}{L} \sum_{R,R'} \langle n_R n_{R'} \rangle e^{i\mathbf{q} \cdot (\mathbf{R} - \mathbf{R}')}, \quad (64)$$

where $\langle \dots \rangle$ indicates the expectation value over the variational wave function. In particular, the presence of a peak at a given $\mathbf{q}$ vector, diverging with the system size, denotes the presence of true charge order in the system. A non-diverging peak signals the presence of short-range charge correlations that do not lead to long-range order.

Moreover, the static structure factor allows us to assess the metallic or insulating nature of the ground state by considering just ground-state expectation values, without directly calculating energy differences. Indeed, the behavior of the static structure factor at small momenta (or, equivalently, at large wavelengths) reveals the nature of the state associated with that wave function. To show this, let us recall the Feynman single-mode approximation, originally formulated in the context of collective excitations in liquid Helium [31] and then applied to fermionic systems [32]. In this context, a variational Ansatz of an excited state $|\Psi_{\mathbf{q}}\rangle$, with a given momentum $\mathbf{q}$, can be obtained by applying $n_{\mathbf{q}}$ to the ground state wave function $|\Psi\rangle$ (or one approximation for it), namely

$$|\Psi_{\mathbf{q}}\rangle = n_{\mathbf{q}}|\Psi\rangle. \quad (65)$$

The variational estimator of the excitation energy is then given by

$$E_{\mathbf{q}} - E_0 = \frac{\langle\Psi_{\mathbf{q}}|\mathcal{H} - E_0|\Psi_{\mathbf{q}}\rangle}{\langle\Psi_{\mathbf{q}}|\Psi_{\mathbf{q}}\rangle} = \frac{\langle\Psi|n_{-\mathbf{q}}[\mathcal{H}, n_{\mathbf{q}}]|\Psi\rangle}{\langle\Psi_{\mathbf{q}}|\Psi_{\mathbf{q}}\rangle} = \frac{\langle\Psi|[n_{-\mathbf{q}}\mathcal{H}], n_{\mathbf{q}}|\Psi\rangle}{\langle\Psi_{\mathbf{q}}|\Psi_{\mathbf{q}}\rangle}, \quad (66)$$

where $\mathcal{H}$ is the Hubbard Hamiltonian. The sum of both commutators $(n_{-\mathbf{q}}\mathcal{H}n_{\mathbf{q}} - n_{-\mathbf{q}}n_{\mathbf{q}}\mathcal{H}) + (n_{-\mathbf{q}}\mathcal{H}n_{\mathbf{q}} - \mathcal{H}n_{-\mathbf{q}}n_{\mathbf{q}})$ is equivalent to the double commutator

$$[n_{-\mathbf{q}}, [\mathcal{H}, n_{\mathbf{q}}]] = n_{-\mathbf{q}}(\mathcal{H}n_{\mathbf{q}} - n_{\mathbf{q}}\mathcal{H}) - (\mathcal{H}n_{\mathbf{q}} - n_{\mathbf{q}}\mathcal{H})n_{-\mathbf{q}} \quad (67)$$

due to the inversion symmetry $\mathbf{q} \leftrightarrow -\mathbf{q}$. Therefore, the excitation energy is given by

$$E_{\mathbf{q}} - E_0 = \frac{1}{2} \frac{\langle\Psi|[n_{-\mathbf{q}}, [\mathcal{H}, n_{\mathbf{q}}]]|\Psi\rangle}{N_{\mathbf{q}}}, \quad (68)$$

where $N(\mathbf{q}) = \langle\Psi|n_{-\mathbf{q}}n_{\mathbf{q}}|\Psi\rangle$ is the static structure factor for the ground state. The double commutator $[n_{-\mathbf{q}}, [\mathcal{H}, n_{\mathbf{q}}]]$ can be evaluated and involves the kinetic term only, since the potential term of the Hamiltonian contains density-density interactions that commute with $n_{\mathbf{q}}$ [21]

$$[n_{-\mathbf{q}}, [\mathcal{H}, n_{\mathbf{q}}]] = \frac{1}{L} \sum_{\mathbf{k}, \sigma} (\varepsilon_{\mathbf{k}+\mathbf{q}} + \varepsilon_{\mathbf{k}-\mathbf{q}} - 2\varepsilon_{\mathbf{k}}) c_{\mathbf{k}, \sigma}^{\dagger} c_{\mathbf{k}, \sigma}, \quad (69)$$

where $\varepsilon_{\mathbf{k}}$ is the dispersion relation of the kinetic part of the Hubbard Hamiltonian. Then, in the limit of small momenta, one gets

$$E_{\mathbf{q}} - E_0 \propto \lim_{q \rightarrow 0} \frac{|\mathbf{q}|^2}{N(\mathbf{q})}. \quad (70)$$

Then, whenever $N(\mathbf{q})$ is linear in $\mathbf{q}$, the energy spectrum is gapless, suggesting that the system is metallic; by contrast, if $N(\mathbf{q}) \sim \mathbf{q}^2$, this construction gives a finite gap, leading to an insulator. As an example, we show in Fig. 3 the behavior of $N(\mathbf{q})$ across the metal-insulator transition that occurs in the single-orbital Hubbard model as a function of U/t at $t'/t = 0.75$. In order to highlight the location of the metal-insulator transition we plot $N(\mathbf{q})/|\mathbf{q}|$: If the system is metallic, $N(\mathbf{q})/|\mathbf{q}|$ extrapolates to a finite value for $|\mathbf{q}| \rightarrow 0$; conversely, if the system is insulating, $N(\mathbf{q})/|\mathbf{q}|$ extrapolates to zero for $|\mathbf{q}| \rightarrow 0$; see Eq. (70).

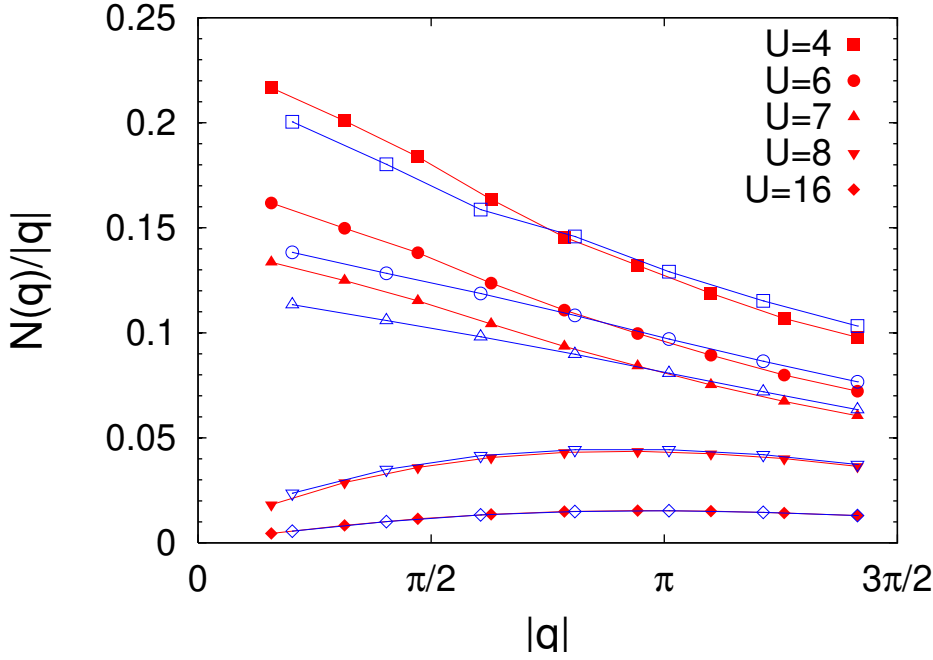


Fig. 3: Variational results for $N(\mathbf{q})$ divided by $|\mathbf{q}|$, as a function of $|\mathbf{q}|$, for $L = 98$ (empty symbols) and for $L = 162$ (full symbols) at $t'/t = 0.75$. The metal-insulator transition takes place between $U/t = 7$ and $U/t = 8$, that is, where $N(\mathbf{q})$ changes its behavior from linear to quadratic for $|\mathbf{q}| \rightarrow 0$ [20].

The second one is the spin-spin correlator. Analogously to the case of charge order, spin order in the system is associated to a peak in the *spin-spin correlations* defined as

$$S(\mathbf{q}) = \frac{1}{L} \sum_{R,R'} \langle S_R^z S_{R'}^z \rangle e^{i\mathbf{q} \cdot (\mathbf{R} - \mathbf{R}')}, \quad (71)$$

where S_R^z is the spin operator along the z direction, i.e., $S_R^z = 1/2(c_{R,\uparrow}^\dagger c_{R,\uparrow} - c_{R,\downarrow}^\dagger c_{R,\downarrow})$. Also, the presence of a spin gap can be detected by looking at the behavior of $S(\mathbf{q})$ at small momenta. As an example, we show the behavior of $S(\mathbf{q})$ for the one dimensional Hubbard model. The model is the same as the one we introduced in Eq. (1), just denoting nearest neighbor hopping as t_1 and next-nearest neighbor hopping as t_2 . Even if in one dimension there is no long-range order, short-range spin correlations are present in the insulating phase. They can be detected as a non-diverging peak in $S(\mathbf{q})$. The peak remains commensurate at $Q = \pi$ up to $t_2/t_1 \sim 0.75$ and then shifts, as a function of the ratio t_2/t_1 , to the value of $Q = \pi/2$ for large values of t_2/t_1 . Results can be seen in Fig. 4.

Then, another crucial task in ground-state simulations is the detection of superconductivity in the model. The presence of finite BCS parameters in the variational state after optimization is a necessary condition for the presence of superconductivity. It is not a sufficient condition because either the system can be an insulator, induced by the long-range tail of the Jastrow factor in the variational state, or finite BCS parameters can be a finite-size effect, disappearing in the thermodynamic limit, as it occurs in the $U = 0$ limit of the Hubbard model. Hence, to reveal the presence of true superconductivity in the state, we can compute the so-called *singlet*

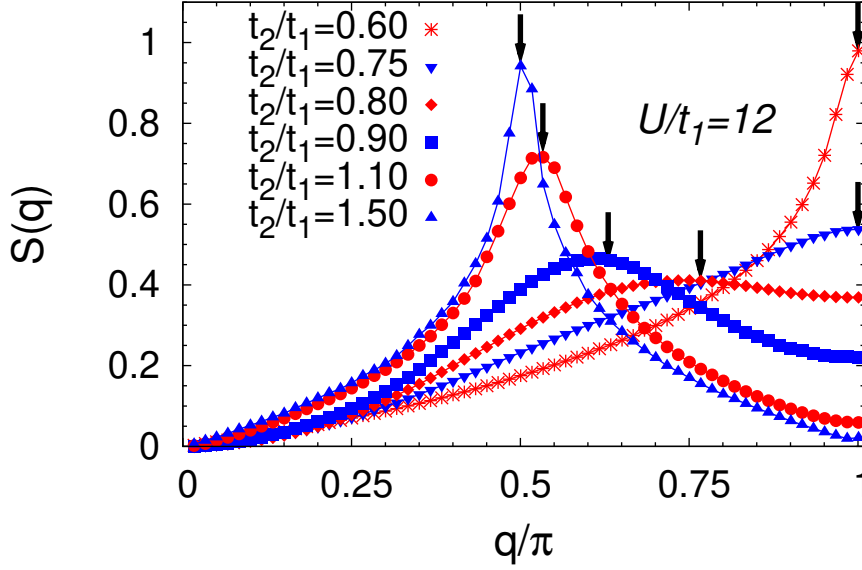


Fig. 4: Spin-spin correlations $S(\mathbf{q})$ at $U/t_1 = 12$ for different values of the hopping ratio t_2/t_1 . Arrows indicate the location of the peaks. Data are shown on a $L = 120$ lattice size [33].

pairing correlation function in real space, that we restrict here to be intra-orbital

$$D_\alpha(r) = \frac{1}{L} \sum_R \langle P_{\alpha,R} P_{\alpha,R+r}^\dagger \rangle, \quad (72)$$

where $\langle \dots \rangle$ means the usual expectation value on the variational ground state, α is the orbital index, and the singlet operator is defined as: $P_{\alpha,R} = c_{\alpha,R+\eta,\downarrow} c_{\alpha,R,\uparrow} - c_{\alpha,R+\eta,\uparrow} c_{\alpha,R,\downarrow}$. In practice, the $P_{\alpha,R}$ operator destroys a singlet oriented along the η direction (for instance, $\eta = x$ or $\eta = y$) in position R , while $P_{\alpha,R+r}^\dagger$ creates a singlet with the same orientation in position $R+r$. Then, the system is superconductive when the following condition is met:

$$\phi_\alpha^2 = \lim_{r \rightarrow \infty} D_\alpha(r) \text{ is finite.} \quad (73)$$

The symbol ϕ_α denotes the superconductive order parameter, that, for $|\phi_\alpha| \neq 0$ indicates that the system is superconductive.

From a practical point of view, on a finite lattice, we evaluate $D_\alpha(r)$ at the maximum distance, i.e., $r = l/2$ (if the lattice size is defined as $L = l \times l$) because of the periodic boundary conditions. Then, we consider different lattice sizes and make an extrapolation of ϕ_α to the thermodynamic limit ($l \rightarrow \infty$), in order to assess the actual presence of superconductivity. Moreover, it might be useful to compare the superconductive correlation functions for finite values of U with the ones obtained at $U = 0$. In the latter case, finite-size kinetic contributions from the hopping are present. Hence, superconductivity emerges only when we observe an increase of ϕ_α in the presence of interaction, with respect to the $U = 0$ case.

As an example, we show in Fig. 5, left panels, the results for ϕ_α^2 as a function of U in the three-orbital Hubbard-Kanamori Hamiltonian of Eqs. (2) and (3). Results are presented for two relevant electronic densities of $n = 3 + 1/3$ and $n = 3 + 2/3$. In both cases, superconductivity

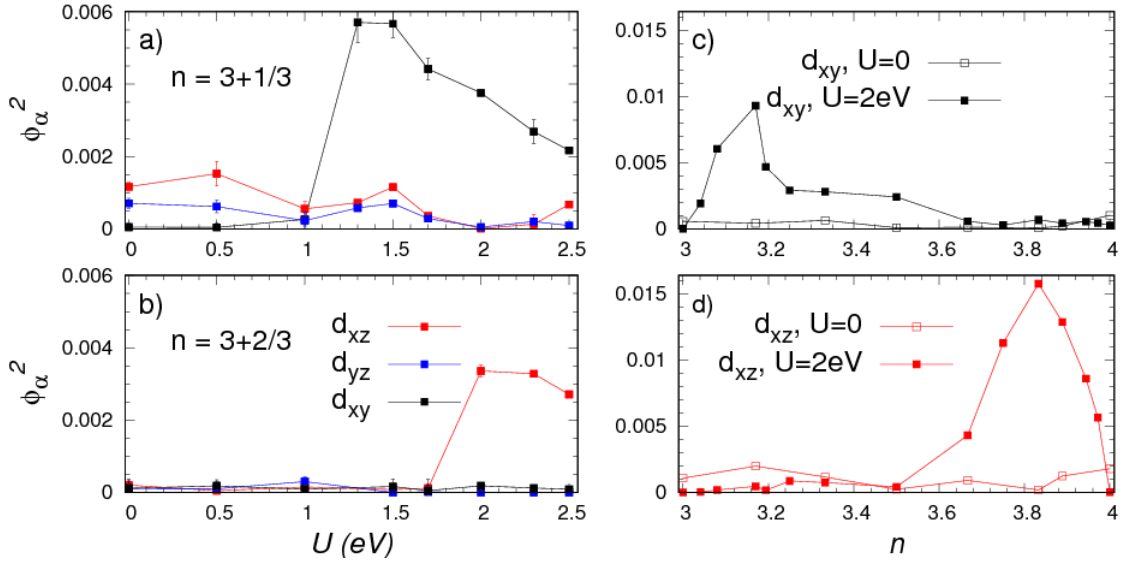


Fig. 5: Square of the superconducting order parameter ϕ_α^2 in each orbital as a function of U at $n = 3 + 1/3$ [panel a)] and $n = 3 + 2/3$ [panel b)]. Data are for the $L = 18 \times 18$ cluster. Square of the superconducting order parameter in the d_{xy} [panel c)] and d_{xz} [panel d)] orbitals as a function of n at $U = 2$ eV (full squares) and at $U = 0$ (empty squares). Data are for the $L = 12 \times 12$ cluster. All data are computed at $J = 0.2 U$ [25].

develops only in one of the three orbitals, while the other two orbitals, instead, do not show any enhancement of the superconducting correlations with respect to the $U = 0$ case. We also show the behavior of ϕ_α^2 as a function of density n in Fig. 5, right panels, for a value of U where superconductivity is present in one of the orbitals. Here, we also report the non-interacting values at $U = 0$, to emphasize the relevance of the Hubbard interaction to generate the electron pairing. Even if not shown here, we mention that superconductivity is not present when the Hund's coupling is set to zero.

As a comparison, in Fig. 6, we report the intra-orbital pairing correlations $D_\alpha(r)$ at densities $n = 3 + 1/3$ and $n = 3 + 2/3$, both within the orbital-selective phase and at $U = 0$. In both cases, there is one orbital where pairing correlations at large distances are strongly enhanced going from $U = 0$ to large values of U . On the contrary, in the other two orbitals, pairing correlations remain similar or are even suppressed when electron-electron correlation is switched on.

We mention, that in addition to the correlators introduced above, one can also compute expectation values of simpler single-particle observables. For instance, in a multi-orbital model, one could compute the electronic density per orbital, defined as $n_\alpha = \frac{1}{L} \langle \sum_{R,\sigma} n_{R,\alpha,\sigma} \rangle$. This observable allows us to identify orbital selective phenomena, that are related to superconductivity. In Fig. 7, we observe a differentiation in the orbital occupation of the two degenerate orbitals (d_{xz} and d_{yz}) above a critical value of U that corresponds to the onset of superconductivity in the most occupied orbital. We remark that the onset of superconductivity and orbital selectivity is related to the presence of a long-range tail in the Jastrow term of the variational wave function, that allows for a proper treatment of spatial correlations.

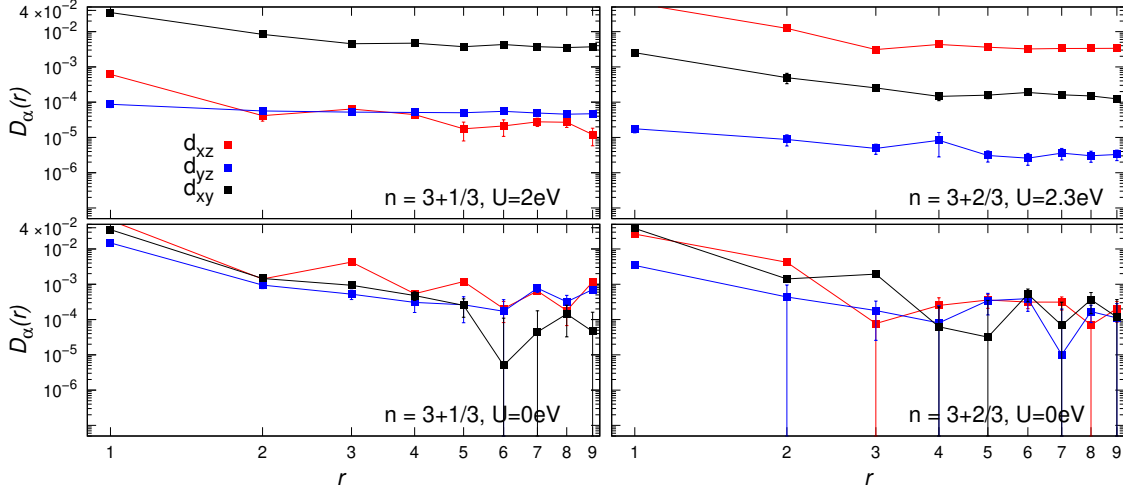


Fig. 6: Upper panels: $D_\alpha(r)$ in all orbitals, for a value of U where superconductivity is present. Data are shown at electronic densities $n = 3+1/3$ (left column) and $n = 3+2/3$ (right column). Lower panels: The same as above, but for $U = 0$. Data are computed on a $L = 18 \times 18$ cluster at $J = 0.2U$ [25].

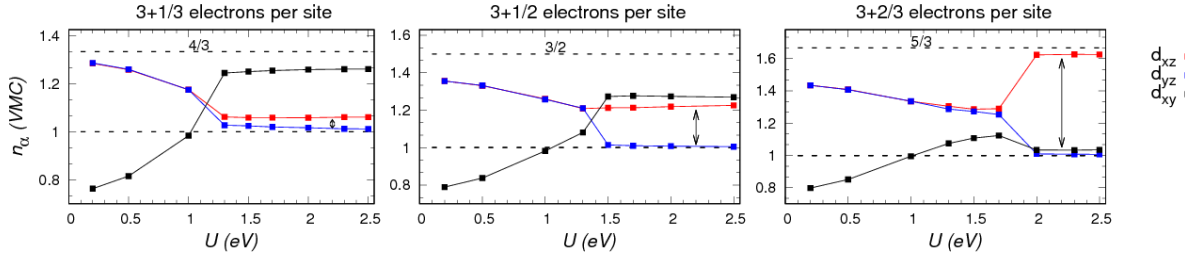


Fig. 7: Electronic density per orbital n_α as a function of U for three different total electronic densities. We fix $J/U = 0.2$ and compute data for the $L = 12 \times 12$ cluster. The arrows highlight the size of the differentiation between the d_{xz} and the d_{yz} orbitals [25].

7 Final remarks

In this lecture, we introduced the basics of the variational Monte Carlo method in general and in relation with the Hubbard model, both in its simpler single-orbital version and in its multi-orbital extension.

Variational Monte Carlo is a conceptually simple approach, where the computation of physical observables over a variational approximation of the ground state wave function is rewritten as averages of local observables over the usual quantum mechanical probability distribution. This definition ensures that VMC has no sign problem, since we sample a nonnegative quantity.

The core of the variational Monte Carlo approach is to define accurate variational states that can reproduce the correct ground state physics; for this reason, we devoted Sec. 4 to the definition of different variational states. We described the Jastrow-Slater wave function for the single-orbital Hubbard model and how it can be improved with the backflow terms. We also presented an extension of the Jastrow-Slater wave function to the multi-orbital case.

Then, once a functional form of the variational wave function is built, it is crucial to be able to optimize the variational parameters that are contained within the variational state. The concomitant optimization of many variational parameters is made possible with the stochastic re-configuration method, described in Sec. 5.

Finally, we introduced several many-body correlators that allow us to characterize the physics of the optimal variational approximation of the ground state. Among the other, we discussed how to discriminate whether the system is metallic or insulating, detect the presence of charge or spin order and the emergence of superconductivity. All these features lead to a complete understanding of the ground-state properties of our correlated Hubbard Hamiltonian.

We close this section by commenting on which paths might variational Monte Carlo take in the future.

A first one is based on the application of relatively simple variational states, of the Jastrow-Slater form, to complicated multi-orbital Hamiltonians where approaches that are nonperturbative in electronic correlations and can address spatial correlations are lacking. These ideas have been already discussed in these lecture notes, in the field of iron-based superconductors, and can have a relevant impact, for instance, in the field of Kagome metals. Here, electronic correlations are a necessary ingredient, while the complicated kinetic electronic model is a challenge for all the contemporary many-body methods [34].

A second one is the improvement of variational wave functions with respect to the scheme introduced in this lecture. In this respect, neural-network quantum states, based on the transformer architecture, show excellent accuracy and appear as promising tools for solving the intricacies of competing orders in the Hubbard model. However, the domain of applicability of these more evolved variational wave functions is, at the moment, limited to the single-orbital Hubbard model on lattices with a simple geometry [35–37].

Finally, a third one that goes away from the target of these notes is to consider not only ground-state properties, but to provide VMC calculations of dynamics and finite-temperature properties, that involve also excited states. This is a wide and growing field, based on the time-dependent variational principle, which might be the subject of another set of lecture notes.

References

- [1] F. Becca and S. Sorella: *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017)
- [2] U. Schollwöck, *Ann. Phys.* **326**, 96 (2011)
- [3] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [4] S. Sorella, *Phys. Rev. B* **71**, 241103 (2005)
- [5] J.P.F. LeBlanc *et al.*, *Phys. Rev. X* **5**, 041041 (2015)
- [6] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, *Annu. Rev. Condens. Matter Phys.* **13**, 275 (2022)
- [7] E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, and O. Andersen, *Phys. Rev. Lett.* **87**, 047003 (2001)
- [8] E.H. Lieb, and F.Y. Wu, *Phys. Rev. Lett.* **20**, 1445 (1968)
- [9] S. Zhang, J. Carlson, and J.E. Gubernatis, *Phys. Rev. Lett.* **74**, 3652 (1995)
- [10] J. Kanamori, *Progr. Theor. Phys.* **30**, 275 (1963)
- [11] L. Georges, L. de Medici and J. Mravlje, *Annu. Rev. Condens. Matter Phys.* **4**, 137 (2013)
- [12] M. Daghofer, A. Nicholson, A. Moreo, and E. Dagotto, *Phys. Rev. B* **81**, 014511 (2010)
- [13] L. Fanfarillo, A. Valli, and M. Capone, *Phys. Rev. Lett.* **125**, 177001 (2020)
- [14] B.-X. Zheng *et al.*, *Science* **358**, 1155 (2017)
- [15] A.S. Darmawan, Y. Nomura, Y. Yamaji, and M. Imada, *Phys. Rev. B* **98**, 205132 (2018)
- [16] V. Marino, F. Becca, and L.F. Tocchio, *SciPost Phys.* **12**, 180 (2022)
- [17] M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti, *Phys. Rev. Lett.* **94**, 026406 (2005)
- [18] P.W. Anderson, *Science* **235**, 1196 (1987)
- [19] M.C. Gutzwiller, *Phys. Rev. Lett.* **10**, 159 (1963)
- [20] L.F. Tocchio, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. B* **78**, 041101(R) (2008)
- [21] L.F. Tocchio, F. Becca, and C. Gros, *Phys. Rev. B* **83**, 195138 (2011)
- [22] R. Chitra, S. Pati, H.R. Krishnamurty, D. Sen, and S. Ramasesha, *Phys. Rev. B* **52**, 6581 (1995)

- [23] L.F. Tocchio, H. Feldner, F. Becca, R. Valenti, and C. Gros, Phys. Rev. B **87**, 035143 (2013)
- [24] D. Luo and B.K. Clark, Phys. Rev. Lett. **122**, 226401 (2019)
- [25] V. Marino, A. Scazzola, F. Becca, M. Capone, and L.F. Tocchio, Phys. Rev. Lett. **134**, 196502 (2025)
- [26] V. Marino: *Superconductivity and Orbital Selectivity in a three-orbital Hubbard model for the iron-based superconductors – Variational Monte Carlo analysis* (Doctoral Thesis, 2025) <http://iris.polito.it/handle/11583/3001278>
- [27] T. Misawa and M. Imada, Nat. Comm. **5**, 5738 (2014)
- [28] C.J. Umrigar, K.G. Wilson, and J.W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988)
- [29] M. Snajdr and S.M. Rothstein, J. Chem. Phys. **112**, 4935(2000)
- [30] S. Yunoki and S. Sorella, Phys. Rev. B **74**, 014408 (2006)
- [31] R.P. Feynman, Phys. Rev. **94**, 262 (1954)
- [32] A.W. Overhauser, Phys. Rev. B **3**, 1888 (1971)
- [33] L.F. Tocchio, F. Becca, and C. Gros, Phys. Rev. B **81**, 205109 (2010)
- [34] D. Di Sante, T. Neupert, G. Sangiovanni, R. Thomale, R. Comin, I. Zeljkovic, J.G. Checkelsky, and S.D. Wilson, Rev. Mod. Phys. **98**, 015002 (2026)
- [35] J.R. Moreno, G. Carleo, A. Georges, and J. Stokes, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022)
- [36] R. Rende, A. Nikolaenko, L.L. Viteritti, S. Sachdev, and Y. -Hui Zhang, arXiv:2603.02316 (2026)
- [37] L.L. Viteritti, R. Rende, C. Roth, A. Sengupta, G. Carleo, and A. Georges, arXiv:2604.21978 (2026)

14 Neural Quantum States for Strongly Correlated Systems

Annabelle Bohrdt
Ludwig-Maximilians-Universität München
Faculty of Physics and
Arnold Sommerfeld Center for Theoretical Physics
Theresienstr. 37, 80333 München

Contents

1	Introduction	2
2	Expressivity of neural network quantum states	4
2.1	Exactly representable states	5
2.2	Entanglement properties	13
3	Training neural quantum states	17
3.1	Variational Monte Carlo	17
3.2	Trainability and design choices	22
4	Concluding remarks	26

1 Introduction

Numerically simulating strongly correlated quantum systems is in general a challenging task: as the Hilbert space dimension grows exponentially, an exact representation is no longer feasible. Analytical approximations are often based on weak interactions or correlations and cannot be applied in the strongly correlated regime. In one dimension, matrix product states (MPS) constitute the state of the art to represent quantum many-body states efficiently, i.e., with a number of parameters that scales polynomially with the system size; see Ref. [1] for a review. Tensor network states in general make use of locality in order to efficiently represent states. For MPS and their one-dimensional structure, which renders the optimization through the density matrix renormalization group algorithm highly efficient, this however poses limitations on the class of states that can be represented with a number of parameters that scales polynomially and not exponentially with the system size. In particular the simulation of two-dimensional systems becomes highly challenging for MPS. A conceptually different approach has been taken in variational Monte Carlo: Before the advent of neural quantum states, variational Monte Carlo was commonly based on physically motivated trial wavefunctions. Typical examples include Slater–Jastrow states in continuum electronic-structure calculations and Gutzwiller-projected Fermi-sea or BCS states for lattice models. In these approaches, the qualitative structure of the wavefunction is chosen from physical intuition, while a comparatively small or structured set of variational parameters is optimized stochastically.

Ten years ago, a seminal paper proposed a new class of variational states: neural quantum states [2]. The basic idea is that a neural network is used to approximate the quantum many-body state

$$|\psi_\theta\rangle = \sum_{\mathbf{s}} \psi_\theta(\mathbf{s}) |\mathbf{s}\rangle, \quad (1)$$

where the network parameters¹ θ are the variational parameters of the ansatz. The sum runs over all possible configurations spanning a basis of the Hilbert space, e.g., for a spin-1/2 system all possible spin configurations in the z -basis, or for itinerant systems like a Fermi-Hubbard model all Fock space configurations. The input to the network is then a given basis configuration $\mathbf{s}$, and the output is the corresponding wavefunction coefficient $\psi_\theta(\mathbf{s})$. The same construction can be employed also in first quantization, where the input is the position of all N particles $\{\mathbf{r}_i\}_{i=1,\dots,N}$. We will in the following restrict ourselves to pure states, but note that density matrices can also be represented by neural quantum states with results for finite temperature states as well as open system dynamics. Neural quantum states generalize the philosophy of more conventional variational Monte Carlo approaches by replacing part or all of the hand-designed correlation structure by a highly flexible parametrization with many trainable parameters, thus to some degree trading interpretability against expressivity of the ansatz. We will discuss in more detail below in Sec. 2.1 how physically motivated design choices can be combined with neural networks.

¹typically the weights and biases, see below for more details.

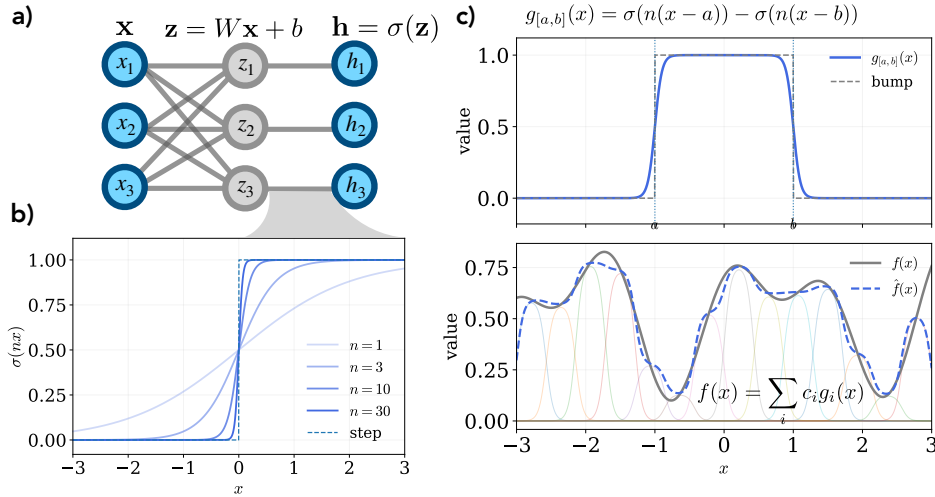


Fig. 1: a) Basic building blocks of a feed-forward neural network. The input vector $\mathbf{x}$ is transformed by an affine map $\mathbf{z} = W\mathbf{x} + b$, followed by a non-linear activation $\mathbf{h} = \sigma(\mathbf{z})$, depicted in b). c) a sigmoid activation can be combined to form approximate window functions, which in turn can be summed to approximate a target function.

Contrary to typical machine learning problems, in general there exists no training data in this approach. The neural network represents the quantum state of interest, and therefore encodes the probability distribution from which one obtains samples according to the state following the Born rule. These samples are distributed as projective measurement outcomes of the current state in the chosen reference or computational basis. The data, i.e., the samples, is thus generated by the network itself, and each training step corresponds to a newly generated dataset. The updates to the network parameters are calculated based on these samples, and ideally, upon updating the network parameters many times, the quantum state represented by the neural network is very close to the desired state, e.g., the ground state of a given Hamiltonian. This optimization procedure, discussed in more detail in Sec. 3, follows the standard variational Monte Carlo approach: sample from the state, estimate expectation values from those samples, and update the variational parameters accordingly.

Compared to tensor networks, locality is not automatically built into a generic neural network architecture. It can, however, be incorporated through architectural choices such as local connectivity, convolutional layers, or equivariant constructions. Given the great expressivity of neural networks (to be discussed more below), the hope behind neural quantum states is to be able to represent a large class of interesting quantum many-body states efficiently.

Neural networks most generally consist of linear parts, i.e., matrix multiplications, and non-linear activation functions, see Fig. 1. These non-linear activation functions essentially give the neural network their power. Neural networks with already one hidden layer are universal function approximators: they can represent any continuous function on a bounded domain to arbitrary accuracy, given enough neurons [3, 4]. More formally, let $K \subset \mathbb{R}^n$ be compact and let $f \in C(K)$. Let $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ be a non-polynomial, continuous activation function. Then for every $\varepsilon > 0$, there exists an integer $N \in \mathbb{N}$ and parameters $a_i \in \mathbb{R}$, $w_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ such that

the function

$$\hat{f}(x) = \sum_{i=1}^N a_i \sigma(w_i \cdot x + b_i) \quad (2)$$

satisfies

$$\sup_{x \in K} |f(x) - \hat{f}(x)| < \varepsilon. \quad (3)$$

The basic intuition behind this theorem is as follows: a small network with a non-linear activation like a sigmoid, defined as

$$\sigma(x) = 1/(1 + e^{-x}), \quad (4)$$

can approximate a 'bump'. Consider an interval $[a, b]$, then the function

$$g(x) = \sigma(n(x-a)) - \sigma(n(x-b)) \quad (5)$$

approximates a function that is 1 on the interval $[a, b]$ and 0 else in the limit of large n , see Fig. 1c). Any function f that is piece-wise constant can now be approximated by sums of such small neural networks. A neural network essentially is a finite sum of functions of the form $c\sigma(ax+b)$, and the number of neurons determines how many such functions we sum over. In the case of neural quantum states, the neural network often maps from discrete numbers, like ± 1 on each site for a spin-1/2 system, to a wavefunction coefficient. Restricted Boltzmann machines have been shown to approximate probability distributions over discrete variables arbitrarily well [5]. Note that in practice, there are two potential problems with this theorem: (i) it does not guarantee an *efficient* representation, such that potentially prohibitively many parameters are needed; and (ii) it only makes a statement about expressivity, not about trainability: finding the set of parameters that corresponds for example to a desired many-body state is not guaranteed. In the remainder of these notes, we will first discuss the question of expressivity in Section 2 and then move to the optimization in Section 3. The focus of these notes is on variational pure-state NQS for strongly correlated lattice systems, with short comments on continuum fermions and time evolution. Other important directions, such as density-matrix parametrizations, open-system dynamics, and neural representations of operators, are mentioned only briefly. These lecture notes are intended as an introduction and rough overview, not as a comprehensive and exhaustive review. For the latter, see e.g. Refs. [6] and [7] for general reviews and [8] for a review on time evolution methods with neural quantum states.

2 Expressivity of neural network quantum states

The main promise, and at the same time the main mystery, of neural quantum states is their expressivity: per the universal function approximator theorem, neural networks can in principle represent any smooth function with arbitrary accuracy. In practice, we are however limited to a finite number of variational (read: network) parameters. This poses the question: how expressive is a neural network with a fixed number of parameters? A closely related question pertains to how the network makes use of these parameters to represent a desired quantum state, and

how we should choose the network architecture for a given problem. While the original work in Ref. [2] focused on restricted Boltzmann machines, many different network architectures are now being used as neural quantum states, ranging from convolutional neural networks to Transformers. Additionally, standard neural network architectures can be combined with physically motivated constructions, such as determinants. Depending on the problem under consideration, different representations can be chosen, e.g., regarding the reference basis, reference sign structure, complex versus real output, or representation of a density matrix for mixed states. Early work has shown that some interesting quantum many-body states can be represented exactly by a neural network, sometimes by default, sometimes per clever choice of the network architecture. We will discuss some of these instances in section 2.1 to gain a feeling for how the specific architecture and/or non-linearities can be used to enforce physical properties of the state under consideration. Through physically inspired architectures, we can gain not only an advantage in expressivity for the problem under consideration, but also some sense of interpretability. However, the guiding principle for what neural quantum states can represent efficiently remains a subject of ongoing research. This is to be contrasted with one of the state-of-the-art methods to numerically capture quantum many-body systems, namely matrix product states (MPS). In the case of MPS, the entanglement constitutes a guiding principle, which limits which states can be efficiently represented. Entanglement properties of neural quantum states depend, once again, on architectural choices, such as number and choice of non-linearities, but at least some volume law entangled states can be represented efficiently using neural quantum states [9, 10]. We will discuss the relation between matrix product states and neural quantum states as well as the latter's entanglement properties in more detail in section 2.2

2.1 Exactly representable states

Historically, restricted Boltzmann machines (RBMs), see Fig. 2a) have been used as the first neural quantum states [2]. While other architectures may exhibit more expressivity, RBMs often enable a better understanding. We can write an RBM wavefunction for spins $s_i = \pm 1$ with M hidden units as

$$\psi_{\text{RBM}}(\mathbf{s}) = \exp \left(\sum_i a_i s_i \right) \prod_{\mu=1}^M 2 \cosh \left(b_{\mu} + \sum_i W_{i\mu} s_i \right). \quad (6)$$

In the following, we want to use this RBM representation to demonstrate how different types of quantum states can be represented. We start by the N -spin GHZ state,

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}} (|\uparrow\uparrow\cdots\uparrow\rangle + |\downarrow\downarrow\cdots\downarrow\rangle), \quad (7)$$

whose computational-basis amplitudes are nonzero only if $s_1 = s_2 = \cdots = s_N$. This global condition can be enforced by nearest-neighbor equality constraints,

$$\psi_{\text{GHZ}}(\mathbf{s}) \propto \prod_{i=1}^{N-1} \delta_{s_i, s_{i+1}}. \quad (8)$$

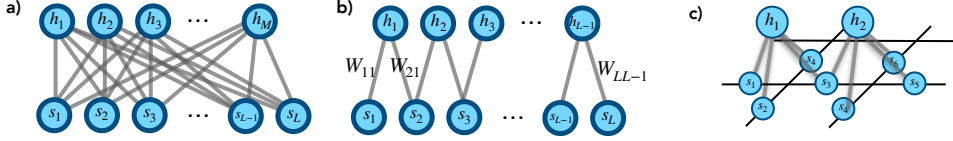


Fig. 2: a) Restricted Boltzmann machine architecture; b) Connectivity for the complex restricted Boltzmann machine to represent the GHZ state, where alignment of each pair of nearest neighboring spins is mediated through a hidden unit; c) Connectivity for a complex restricted Boltzmann machine to represent the Toric code ground state with one hidden unit per vertex.

Each equality constraint can be implemented by one complex hidden unit², see Fig. 2b)

$$2 \cosh \left[\frac{i\pi}{4} (s_i - s_{i+1}) \right] = \begin{cases} 2, & s_i = s_{i+1}, \\ 0, & s_i \neq s_{i+1}. \end{cases} \quad (9)$$

Thus, choose $M = N-1$ hidden units, no visible or hidden biases, $a_i = 0$, $b_\mu = 0$, and weights

$$W_{i\mu} = \frac{i\pi}{4} \delta_{i,\mu} - \frac{i\pi}{4} \delta_{i,\mu+1}, \quad \mu = 1, \dots, N-1. \quad (10)$$

The resulting RBM wavefunction is

$$\psi_{\text{RBM}}(\mathbf{s}) = \prod_{\mu=1}^{N-1} 2 \cosh \left[\frac{i\pi}{4} (s_\mu - s_{\mu+1}) \right]. \quad (11)$$

Therefore,

$$\psi_{\text{RBM}}(\mathbf{s}) = \begin{cases} 2^{N-1}, & s_1 = s_2 = \dots = s_N, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

After normalization, this gives exactly

$$|\psi_{\text{RBM}}\rangle = \frac{1}{\sqrt{2}} (|\uparrow\uparrow \dots \uparrow\rangle + |\downarrow\downarrow \dots \downarrow\rangle). \quad (13)$$

This example demonstrates how hidden units can be used to enforce local constraints, in this case $s_i = s_{i+1}$.

Similarly, hidden units can be used to enforce local parity constraints, see also [9]. Our next example is the Toric-code Hamiltonian [11] with spins on the links (edges) of a square lattice, indicated by the index e ,

$$H_{\text{TC}} = - \sum_v A_v - \sum_p B_p, \quad A_v = \prod_{e \in +_v} \sigma_e^x, \quad B_p = \prod_{e \in \partial p} \sigma_e^z. \quad (14)$$

We work in the σ^x -basis,

$$\sigma_e^x |s_e\rangle_x = s_e |s_e\rangle_x, \quad s_e = \pm 1. \quad (15)$$

²For real finite RBM parameters, these hard zeros cannot be enforced exactly; see the discussion below.

In this basis, the vertex constraint $A_v = 1$ becomes

$$\prod_{e \in +_v} s_e = 1. \quad (16)$$

Thus a toric-code ground state may be written as an equal-weight superposition over closed-loop configurations,

$$|\psi_{\text{TC}}\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{\mathbf{s}: \prod_{e \in +_v} s_e = 1 \ \forall v} |\mathbf{s}\rangle_x. \quad (17)$$

Using an RBM with binary hidden variables $h_v \in \{0, 1\}$,

$$\psi_{\text{RBM}}(\mathbf{s}) = \prod_v \left[1 + \exp \left(b_v + \sum_e W_{ev} s_e \right) \right], \quad (18)$$

one hidden unit per vertex, i.e., connecting to four visible units, enforces the local parity constraint. Choose

$$b_v = \frac{i\pi d_v}{2}, \quad W_{ev} = \begin{cases} -\frac{i\pi}{2}, & e \in +_v, \\ 0, & \text{otherwise,} \end{cases} \quad (19)$$

where d_v is the degree of vertex v . Then

$$1 + \exp \left(b_v + \sum_e W_{ev} s_e \right) = 1 + \prod_{e \in +_v} s_e. \quad (20)$$

Therefore

$$\psi_{\text{RBM}}(\mathbf{s}) = \prod_v \left(1 + \prod_{e \in +_v} s_e \right) = \begin{cases} 2^{N_v}, & \prod_{e \in +_v} s_e = 1 \ \forall v, \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

After normalization, this is exactly the toric-code ground state in the σ^x -basis. The plaquette terms B_p flip closed loops, and the equal-weight superposition is invariant under these flips.

Apart from constraints, hidden units can also be viewed as effectively mediating correlations between visible units. Consider a spin Jastrow wavefunction

$$\psi_{\text{J}}(\mathbf{s}) = \exp \left(\sum_i a_i s_i + \sum_{i < j} J_{ij} s_i s_j \right), \quad s_i = \pm 1. \quad (22)$$

This state can be represented exactly by an RBM with one hidden unit for each nonzero pair coupling J_{ij} [12].

For a single pair (i, j) , introduce a hidden unit coupled only to s_i and s_j ,

$$\psi_{ij}(s_i, s_j) = 2 \cosh [w_{ij} (s_i + \eta_{ij} s_j)], \quad (23)$$

where

$$\eta_{ij} = \text{sgn}(J_{ij}), \quad w_{ij} = \frac{1}{2} \text{arcosh} (e^{2|J_{ij}|}). \quad (24)$$

If $s_i s_j = \eta_{ij}$, then

$$\psi_{ij}(s_i, s_j) = 2 \cosh(2w_{ij}) = 2e^{2|J_{ij}|}, \quad (25)$$

whereas if $s_i s_j = -\eta_{ij}$, then

$$\psi_{ij}(s_i, s_j) = 2. \quad (26)$$

Thus,

$$\psi_{ij}(s_i, s_j) = C_{ij} \exp(J_{ij} s_i s_j), \quad (27)$$

with the spin-independent constant $C_{ij} = 2e^{|J_{ij}|}$. Constants of this form only affect the overall normalization. Therefore, choosing one hidden unit for every pair (i, j) gives

$$\psi_{\text{RBM}}(\mathbf{s}) = \exp\left(\sum_i a_i s_i\right) \prod_{i < j} 2 \cosh[w_{ij}(s_i + \text{sgn}(J_{ij})s_j)]. \quad (28)$$

Using the identity above,

$$\psi_{\text{RBM}}(\mathbf{s}) \propto \exp\left(\sum_i a_i s_i + \sum_{i < j} J_{ij} s_i s_j\right) = \psi_{\text{J}}(\mathbf{s}). \quad (29)$$

Hence a spin Jastrow state is exactly representable by an RBM with one hidden unit per nonzero pair interaction.

These examples demonstrate how hidden units can encode correlations as well as enforce constraints between the visible units. Along the same lines of reasoning, it has been shown that an exact representation of a variety of states is efficiently possible using neural quantum states, e.g., the Laughlin state [13–15] or Toric code ground states with independently tunable topological-sector amplitudes [16]. Using deeper networks and more advanced architectures, the same intuition applies in principle.

The exact RBM constructions above rely on complex-valued parameters. This is important: with real finite parameters in the standard RBM form, the factors $2 \cosh(b_\mu + \sum_i W_{i\mu} s_i)$ are nonzero for generic configurations, such that hard zeros in the wavefunction are not obtained exactly. Complex parameters, on the other hand, allow one to use zeros of the hyperbolic cosine to enforce constraints exactly. Thus, hidden units can be viewed not only as mediating correlations, but also as enforcing hard constraints through destructive interference.

Matrix Product States.— A different angle is provided by relating matrix product states to neural quantum states: here, we do not explicitly construct the weights and biases of the neural network to achieve certain conditions on $\psi(\mathbf{s})$ as above, but instead use the non-linearities of the neural network to mimic the tensor contractions within the matrix product state. Different approaches can be used to represent a given MPS as an NQS: in Ref. [17], a tensor network contraction is expressed as a sum over discrete auxiliary variables, i.e., the hidden variable marginalization of a Boltzmann machine. This can be viewed as the virtual indices of the matrices/tensors in the MPS become the hidden spins in the network. This Boltzmann-machine representation³ of the state is efficient in the sense that it has an almost-linear overhead in the number of nonzero MPS parameters. In a related work [18], explicit mappings between tensor

³Note that this is not a restricted Boltzmann machine.

networks and restricted Boltzmann machines are constructed, with the important consequence that a non-local restricted Boltzmann machine can correspond to an exponentially growing bond dimension (i.e. not an efficient representation) for the local tensor network. Beyond marginalizing over hidden units, one can also view the network itself, i.e., the computation of hidden variables throughout the layers, as a form of contraction. In [19], a neural network is constructed such that the layers emulate the tensor contraction. The network here is viewed as a computational circuit that evaluates the contraction. Using this algorithm, efficiently contractible tensor networks can be represented by polynomially sized neural networks, with network size controlled by the contraction complexity. A conceptually slightly different approach is taken in Ref. [20], where the mapping from an MPS to a recurrent neural network (RNN) is shown. In this case, one can make use of the fact that the left (or right) environment of an MPS can be defined recursively, such that the hidden state of the RNN can be identified with the current MPS boundary matrix. The RNN then updates its memory exactly as the MPS contraction would update the boundary as one moves through the system. The squared MPS bond dimension χ^2 in this case corresponds to the RNN memory size.

In general, a tensor network represents the state $\psi(\mathbf{s})$ by contracting virtual indices (i.e. performing tensor products), while neural quantum states represent $\psi(\mathbf{s})$ by computing or marginalizing over hidden variables. This enables mappings between the two different representations since contracting over virtual indices can be viewed as a hidden variable computation. These results show that, if a quantum many-body state can be expressed as a matrix product state, then there exists an efficient neural quantum state representation of this state.

Fermions.— Among the most challenging systems for numerical simulations are interacting fermions in two or three spatial dimensions. In order to gain intuition how to construct neural networks that can represent quantum states of interacting fermions, we will start from non-interacting fermions on a lattice, following Ref. [21], described by the Hamiltonian

$$\hat{H}_0 = \sum_{i,j} \sum_{\alpha,\beta} h_{i\alpha,j\beta} \hat{c}_{i\alpha}^\dagger \hat{c}_{j\beta}, \quad (30)$$

where i, j label lattice sites and α, β label spin. It is useful to combine site and spin into a single index $I = (i, \alpha)$.

Diagonalizing the single-particle Hamiltonian gives orbitals $\Phi_{I\epsilon}$ and operators

$$\hat{\gamma}_\epsilon^\dagger = \sum_I \Phi_{I\epsilon} \hat{c}_I^\dagger, \quad \hat{H}_0 = \sum_\epsilon \epsilon_\epsilon \hat{\gamma}_\epsilon^\dagger \hat{\gamma}_\epsilon. \quad (31)$$

For N fermions, the many-body ground state is obtained by occupying the N lowest single-particle orbitals,

$$|\psi_0\rangle = \prod_{\epsilon=1}^N \hat{\gamma}_\epsilon^\dagger |0\rangle = \prod_{\epsilon=1}^N \left(\sum_I \Phi_{I\epsilon} \hat{c}_I^\dagger \right) |0\rangle. \quad (32)$$

Now consider a Fock basis state $|\mathbf{s}\rangle$ with occupied spin-orbitals $I_1, \dots, I_N$, written in a fixed canonical order,

$$|\mathbf{s}\rangle = \hat{c}_{I_1}^\dagger \hat{c}_{I_2}^\dagger \cdots \hat{c}_{I_N}^\dagger |0\rangle. \quad (33)$$

Its overlap with the free-fermion ground state is

$$\langle \mathbf{s} | \psi_0 \rangle = \langle 0 | \hat{c}_{I_N} \cdots \hat{c}_{I_1} \prod_{\epsilon=1}^N \left(\sum_I \Phi_{I\epsilon} \hat{c}_I^\dagger \right) | 0 \rangle. \quad (34)$$

Only those terms survive in which the N occupied orbitals $\epsilon = 1, \dots, N$ are assigned to the occupied spin-orbitals $I_1, \dots, I_N$. Thus,

$$\langle \mathbf{s} | \psi_0 \rangle = \sum_{p \in S_N} \left[\prod_{\epsilon=1}^N \Phi_{I_{p(\epsilon)}, \epsilon} \right] \langle 0 | \hat{c}_{I_N} \cdots \hat{c}_{I_1} \hat{c}_{I_{p(1)}}^\dagger \cdots \hat{c}_{I_{p(N)}}^\dagger | 0 \rangle. \quad (35)$$

Using the fermionic anticommutation relations, the creation operators are reordered into the canonical order,

$$\hat{c}_{I_{p(1)}}^\dagger \cdots \hat{c}_{I_{p(N)}}^\dagger = \text{sgn}(p) \hat{c}_{I_1}^\dagger \cdots \hat{c}_{I_N}^\dagger. \quad (36)$$

Therefore,

$$\langle 0 | \hat{c}_{I_N} \cdots \hat{c}_{I_1} \hat{c}_{I_{p(1)}}^\dagger \cdots \hat{c}_{I_{p(N)}}^\dagger | 0 \rangle = \text{sgn}(p), \quad (37)$$

and hence

$$\langle \mathbf{s} | \psi_0 \rangle = \sum_{p \in S_N} \text{sgn}(p) \prod_{\epsilon=1}^N \Phi_{I_{p(\epsilon)}, \epsilon}. \quad (38)$$

This is precisely the Leibniz formula for the determinant of the matrix obtained by selecting from Φ the rows corresponding to the occupied spin-orbitals in $\mathbf{s}$

$$\langle \mathbf{s} | \psi_0 \rangle = \det \begin{pmatrix} \Phi_{I_1,1} & \Phi_{I_1,2} & \cdots & \Phi_{I_1,N} \\ \Phi_{I_2,1} & \Phi_{I_2,2} & \cdots & \Phi_{I_2,N} \\ \vdots & \vdots & & \vdots \\ \Phi_{I_N,1} & \Phi_{I_N,2} & \cdots & \Phi_{I_N,N} \end{pmatrix}. \quad (39)$$

Equivalently, if $\mathbf{s} \circ \Phi$ denotes the submatrix of Φ obtained by selecting the occupied rows specified by $\mathbf{s}$, then $\langle \mathbf{s} | \psi_0 \rangle = \det(\mathbf{s} \circ \Phi)$. At this point, we have an exact expression for the wavefunction coefficients $\psi_0(\mathbf{s})$ of the ground state of a fermionic system without interactions and without neural network.

In order to include interactions, we have to go beyond the single-particle orbitals used so far. As a concrete example, we consider the prototypical model for interacting spin-1/2 fermions on a lattice, see also [22]: the Fermi-Hubbard model, given by

$$\hat{H} = -t \sum_{\langle i,j \rangle, \sigma} \left(\hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + \text{h.c.} \right) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}. \quad (40)$$

Here $\hat{c}_{i\sigma}^\dagger$ and $\hat{c}_{i\sigma}$ create and annihilate a fermion with spin $\sigma \in \{\uparrow, \downarrow\}$ on lattice site i . The first term describes hopping between nearest-neighbor sites $\langle i, j \rangle$ with amplitude t , while the second term describes an on-site interaction of strength U between opposite spins.

As a minimal example, we consider the two-site Fermi–Hubbard model in the $N_v = 2$ particle sector with $S_v^z = 0$. The exact ground state can be written as

$$|\psi_0\rangle = \left[\frac{\cos \theta}{\sqrt{2}} \left(\hat{c}_{1\uparrow}^\dagger \hat{c}_{1\downarrow}^\dagger + \hat{c}_{2\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger \right) + \frac{\sin \theta}{\sqrt{2}} \left(\hat{c}_{1\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger + \hat{c}_{2\uparrow}^\dagger \hat{c}_{1\downarrow}^\dagger \right) \right] |0\rangle. \quad (41)$$

The angle θ is determined by

$$\frac{1}{\tan \theta} = u \left(\sqrt{1 + \frac{1}{u^2}} - 1 \right), \quad u = \frac{U}{4t}. \quad (42)$$

For $U = 0$, one has $\theta = \pi/4$, such that all four configurations have equal weight. In this case the state is a single Slater determinant, $\langle \mathbf{s} | \psi_{\text{SD}} \rangle = \det(\mathbf{s} \circ \Phi)$, with

$$\Phi = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix}. \quad (43)$$

Here the rows correspond to the physical spin-orbitals ($1 \uparrow, 2 \uparrow, 1 \downarrow, 2 \downarrow$), and the two columns correspond to the occupied spin-up and spin-down orbitals.

For $U > 0$, the relative weight of doubly occupied configurations is reduced, while inter-site singlet configurations become enhanced. This can no longer be represented by a single Slater determinant of the visible fermions. In order to express this state, we introduce a single (spinless) hidden fermion, that can occupy two possible hidden states, denoted by S and D . The hidden orbital is chosen as $\tilde{\Phi} = (c \tan \theta, c)^T$, where c is a normalization constant. The hidden occupation $\tilde{s}$ is made explicitly dependent on the configuration of the visible fermions,

$$\tilde{s}(\mathbf{s}) = \begin{cases} S, & \text{if there is no double occupation in } \mathbf{s}, \\ D, & \text{otherwise.} \end{cases} \quad (44)$$

Thus the hidden fermion selects different hidden amplitudes depending on whether the visible configuration is singly occupied on both sites or contains a doublon. In order to obtain the wave function of the physical fermions, we project back onto the physical Hilbert space by fixing the hidden occupation according to $\tilde{s}(\mathbf{s})$, expressed as $\hat{P}$ below

$$|\psi_{\text{HFDS}}\rangle = \hat{P} \left[\frac{c}{2} \left(\hat{c}_{1\uparrow}^\dagger \hat{c}_{1\downarrow}^\dagger + \hat{c}_{2\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger \right) |0\rangle_v \otimes |D\rangle_h + \frac{c \tan \theta}{2} \left(\hat{c}_{1\uparrow}^\dagger \hat{c}_{2\downarrow}^\dagger + \hat{c}_{2\uparrow}^\dagger \hat{c}_{1\downarrow}^\dagger \right) |0\rangle_v \otimes |S\rangle_h \right]. \quad (45)$$

The hidden fermion thus, similar to the hidden nodes in the RBM examples above, mediates correlations between the physical or visible fermions and allow us to express correlated fermionic states as the determinant of an enlarged matrix.

The two-site example illustrates the basic idea behind hidden fermion determinant states, originally introduced in Ref. [23]: one augments the physical Hilbert space by auxiliary fermionic

degrees of freedom and then projects back to the physical Hilbert space. The hidden degrees of freedom allow the effective determinant to depend on the visible configuration.

We will now introduce this approach a bit more generally, i.e., going beyond the toy example of two visible and a single hidden fermion. We denote a Fock-space configuration in the enlarged Hilbert space by $|\mathbf{s}, \tilde{\mathbf{s}}\rangle$, where $\mathbf{s}$ labels the physical fermion occupations and $\tilde{\mathbf{s}}$ labels the hidden fermion occupations. Let Φ_v and Φ_h denote the orbital matrices associated with the visible and hidden fermions, respectively. Before projection, the wavefunction amplitude in the enlarged Hilbert space is given by a determinant,

$$\langle \mathbf{s}, \tilde{\mathbf{s}} | \psi_{\text{HFDS}} \rangle = \det \begin{pmatrix} \mathbf{s} \circ \Phi_v \\ \tilde{\mathbf{s}} \circ \Phi_h \end{pmatrix}. \quad (46)$$

Here $\mathbf{s} \circ \Phi_v$ denotes the rows of Φ_v selected by the occupied visible spin-orbitals in $\mathbf{s}$, and $\tilde{\mathbf{s}} \circ \Phi_h$ denotes the rows of Φ_h selected by the occupied hidden orbitals. To obtain a wavefunction in the physical Hilbert space, the hidden configuration is made dependent on the visible one, $\tilde{\mathbf{s}} = \tilde{\mathbf{s}}(\mathbf{s})$. This gives the projected amplitude

$$\langle \mathbf{s} | \psi_{\text{HFDS}} \rangle = \det \begin{pmatrix} \mathbf{s} \circ \Phi_v \\ \tilde{\mathbf{s}}(\mathbf{s}) \circ \Phi_h \end{pmatrix}. \quad (47)$$

The lower blocks are configuration-dependent because the hidden occupation $\tilde{\mathbf{s}}(\mathbf{s})$ is chosen as a function of the visible configuration. Thus, hidden fermions turn a fixed Slater determinant in an enlarged Hilbert space into a configuration-dependent determinant in the physical Hilbert space,

$$\psi_{\text{HFDS}}(\mathbf{s}) = \det M(\mathbf{s}). \quad (48)$$

In practical hidden-fermion determinant states, the lower blocks of the matrix are parametrized by a neural network. A different view on the entire hidden fermion construction is thus that the final non-linear activation function is a determinant [24]. From a standard machine-learning perspective, a determinant is an unusual choice of non-linear activation. However, as we have seen above, from a physical perspective, this choice comes very naturally, as free fermions are captured exactly, and correlations can be encoded efficiently. An additional angle on why the determinant can be a highly efficient non-linear activation function is that it naturally leads to high-order correlations between the different inputs and thus visible nodes, which can, e.g., easily lead to a favorable entanglement scaling as discussed in more detail in Sec. 2.2.

This expressive determinant activation comes with an additional numerical cost: each wavefunction evaluation requires a determinant of the enlarged visible-plus-hidden matrix, scaling naively as $O(N_{\text{tot}}^3)$. During Monte Carlo sampling, however, successive configurations often differ only locally, so determinant ratios and inverse matrices can be updated using low-rank formulas rather than recomputed from scratch. Recent fermionic NQS architectures exploit this idea explicitly, trading off expressivity and efficient local updates to reach a favorable accuracy-cost Pareto frontier [23, 25, 26].

Apart from the hidden fermion determinant states, there are other neural quantum state constructions in similar spirit to describe interacting fermions:

- multiplying a fermionic reference state, such as a Slater determinant or pair-product/geminal state, by a correlation factor. This factor can be a conventional Jastrow factor or a neural-network correlator, for example an RBM or autoregressive network [27, 28].
- neural backflow, where a neural network is used to make the single particle orbitals configuration dependent [29]. It has been shown that the hidden fermion determinant can be expressed as a neural backflow state with a Jastrow factor [25].
- replacing the determinant in the above constructions by a Pfaffian, which is naturally equipped to capture pairing [30].

The Fermi–Hubbard model discussed above is a minimal setting in which the competition between kinetic energy and local interactions is present. The fermionic constructions discussed above, however, are not restricted to this model. They apply more generally to lattice fermion Hamiltonians, including longer-range hopping [31] and interactions, multi-orbital models [32, 33], disordered systems, and all-to-all interacting fermionic models [34, 10].

The determinant- and Pfaffian-based ideas discussed above for lattice fermions also have natural analogues in continuum systems. In this setting, one often works in a first-quantized representation, where a configuration is given by the particle coordinates $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$, possibly together with spin labels, and the wavefunction is represented directly as $\psi_\theta(\mathbf{R})$. Fermionic antisymmetry is then imposed at the level of the coordinate-space wavefunction, rather than through anticommuting creation and annihilation operators. A common strategy is to combine neural-network correlation factors or backflow transformations with antisymmetric structures such as Slater determinants or Pfaffians. This is the basic idea behind continuum fermionic neural-network wavefunctions such as FermiNet and PauliNet, as well as more recent Pfaffian–Jastrow–backflow ansätze for continuum Fermi gases [35–37]. These continuum constructions are conceptually close to the lattice determinant-based NQS discussed above: neural networks make the orbitals, pair functions, or correlation factors configuration-dependent, thereby introducing correlations beyond a fixed mean-field determinant or Pfaffian. Recent work also investigates the expressivity of fermionic neural-network architectures and asks under which conditions antisymmetric wavefunctions can be approximated systematically by neural ansätze [38, 39].

2.2 Entanglement properties

In order to quantify the expressivity of a variational state, some measure has to be chosen. A popular choice is the entanglement, as it provides a resource in quantum computing, plays a crucial role in many quantum systems, and is known to be the guiding principle for matrix product states. Typically, we consider the bipartite entanglement, i.e., we split our full system into a subsystem A and B, trace out part B, and consider properties of the resulting reduced density matrix ρ_A of part A in order to quantify how entangled A and B are. Upon writing the

full state as a Schmidt decomposition

$$|\psi\rangle = \sum_j \sqrt{\lambda_j} |\psi_j^A\rangle \otimes |\psi_j^B\rangle \quad (49)$$

with Schmidt coefficients λ_j , we can write the reduced density matrix as

$$\rho_A = \sum_j e^{-\xi_j} |\psi_j^A\rangle \langle \psi_j^A| \quad (50)$$

with entanglement energies $\xi_j = -\log \lambda_j$. The von Neumann entanglement entropy is then

$$S_A = -\text{Tr } \rho_A \log \rho_A = -\sum_i \lambda_i \log \lambda_i = \sum_i \lambda_i \xi_i. \quad (51)$$

Note that the von Neumann entanglement entropy is just one scalar summary of the entire spectrum, and therefore naturally more information is contained in the full entanglement spectrum. Often the 'amount' of entanglement, and in particular the scaling of S_A with the area (boundary) and volume of the region A is considered. Interesting entanglement properties can however go beyond this scaling, e.g., the topological entanglement entropy. For neural quantum states on system sizes beyond exact diagonalization, the von Neumann entanglement entropy cannot be evaluated efficiently and one resorts typically to the second Rényi entropy, which can be evaluated based on Monte Carlo samples through the SWAP trick [40] as

$$S_{2,A} = -\log \text{Tr } \rho_A^2 = -\log \sum_i \lambda_i^2 = -\log \langle \text{SWAP}_A \rangle. \quad (52)$$

Similar to the von Neumann entanglement entropy, the second Rényi entropy is also a scalar summary of the full entanglement spectrum. The latter is however more sensitive to the largest eigenvalues, while the von Neumann entropy weights the tail of the distribution in comparison more.

Matrix product states are constructed such that the bond dimension χ directly corresponds to the number of Schmidt coefficients λ_j that is taken into account at each bond along the one-dimensional path through the system. Schmidt values λ_j with $j > \chi$ are set to zero, effectively truncating the entanglement spectrum and thereby limiting how much entanglement S_A can be represented. For neural quantum states, the natural question arises: what determines how much entanglement can be represented by a given architecture? In order to tackle this question, we have to be even more precise and distinguish: (i) capability – what can the architecture represent if the best choice of parameters is found, which may provide a highly non-trivial task; (ii) trainability – (how) can we determine these optimal parameters; and (iii) typicality – how much entanglement⁴ does this architecture feature upon random initialization, i.e., without any training. We discuss some results on the capability and typicality of neural quantum states with regard to entanglement below, and refer to section 3 for more details on trainability in general.

⁴Or whatever other measure we want to apply to quantify expressivity.

2.2.1 Capability

A number of research papers following different strategies has by now found that there are (at least) two ingredients with a strong influence on the capabilities of a given architecture:

Locality.— The connectivity within the network has – at least in the context of RBMs – been shown to determine the network’s capability to represent area versus volume law entanglement: locality, i.e., restricting connections to a few nodes, leads to the capability to represent area but not volume law entanglement [9]. This is conceptually similar to the case of MPS, where locality is exploited to efficiently optimize the variational state, but also leads to the discussed limitations regarding entanglement behavior. An RBM with long-range connectivity, which is the standard case of each hidden node being connected to each visible node, can represent volume law entanglement.

Non-linearity.— We have seen in Sec. 2.1 at the example of the RBM how the hidden units effectively can mediate higher-order correlation function. Beyond the concrete examples shown above, one can also see how (higher-order) correlations between the visible units mediated by the hidden units emerge through a Taylor expansion, see [41] for more details. Similarly, Ref. [42] employs a Taylor expansion of the non-linear activation function to prove that an extensive number of non-linearities is necessary in a feed-forward neural network to represent a volume law entangled state. This proof relies on the fact that only a small number of sites is coupled in each non-linearity, such that only a finite amount of entanglement can be generated. The results of [42] are in line with earlier work emphasizing the role of the depth [43]. In Ref. [24], the standard non-linear activation function is fully replaced by an expansion in orders of products of the visible units or correlations. This correlator transformer approach provides a particularly clear picture on the role of the non-linearity: it is shown in [24] that in order to represent volume law entanglement, correlations of the order of the system size are required. Typical non-linear activation functions however have very quickly decaying pre-factors of such high-order correlations of the visible units. The required large pre-factors can be restored through deeper networks, where the repeated application of the corresponding non-linearity enables higher-order correlation between the visible units.

As pointed out in [24] and [42], the determinant (and similarly the Pfaffian) in the fermionic ansätze described above can effectively be viewed as a special non-linear activation function, which immediately enables access to high order polynomials of the inputs. These determinant and Pfaffian based constructions therefore more naturally represent states featuring high amounts of entanglement. While this can be naturally seen through the arguments given in [24] in terms of high order polynomials of the inputs which encode the entanglement, one can also see this from a more physical point of view, as it has been shown that states with a Fermi surface naturally exhibit an entanglement scaling of $L^{d-1} \log L$, so a multiplicative logarithmic correction to an area law scaling [44]. The determinant construction naturally encompasses states featuring a Fermi surface and thus the corresponding entanglement scaling.

This favorable behavior of the determinant has also been observed in practice: Ref. [45] found that representing the ground state of the SYK model, which features volume law entanglement, requires an exponential in system size number of parameters if the state is represented by a feedforward neural network. At the time, the authors concluded that neural quantum states cannot represent volume law entangled states. A comment on this paper, Ref. [10], argued that the failure to efficiently represent the state is due to the attempt to represent a fermionic ground state with an architecture that is not tailored towards fermions. A complementary perspective is provided by the role of the non-linearity: determinant-based ansätze provide high-order polynomial structure from the start, and therefore offer a much more suitable inductive bias for fermionic states with large entanglement than generic feed-forward architectures.

2.2.2 Typicality

In order to determine how much entanglement a given architecture typically features, the network parameters are randomly initialized, and the entanglement entropy is evaluated without any training. Ref. [9] demonstrates that the von Neumann entanglement entropy for randomly initialized RBMs exhibits volume law behavior for the small system sizes considered, as long as the ratio of hidden to visible units does not become too large. Many hidden units, and therefore many variational parameters, may seem favorable for the expressivity of the ansatz. However, at least in the case of random initialization considered here, more hidden units lead effectively to a state closer to a product state because there is less dependence on the visible units [9].

In [46], different autoregressive architectures, which commonly feature a softmax as their output activation function, are considered. Autoregressive architectures are designed using a chain rule of probabilities, such that each local probability can be normalized, yielding a fully normalized probability. The output activation function is therefore commonly chosen to be a softmax. The normalized probabilities enable perfect sampling and avoid Markov-chain Monte Carlo sampling, i.e., there are no autocorrelations. The authors of Ref. [46] evaluate the second Rényi entropy $S_{2,A}$ for different widths of the distribution from which the random network parameters are sampled and examine the scaling of $S_{2,A}$ with the size of the region A . The two main findings of this work are: (i) the non-linear (output) activation function matters: the standard softmax normalization can suppress fluctuations and entanglement in randomly initialized autoregressive wavefunctions, especially when the logits become large. They proposed a square-modulus normalization as an alternative output map, which restores stronger fluctuations and larger entanglement in the random-state ensemble; and (ii) the width of the random initialization distribution can strongly affect both the entanglement of autoregressive NQS and their subsequent VMC performance. For Gaussian-initialized RNN and autoregressive transformer wavefunctions, the average entanglement is typically maximized at an intermediate width: too small widths give nearly product states, while too large widths can lead, in softmax-based architectures, to localized output distributions and suppressed entanglement. The Gaussian width that minimizes VMC convergence time is also found at intermediate values and shifts to smaller values as the network size increases. This suggests that initialization should be chosen with the entanglement structure of the ansatz in mind.

In Ref. [47], the second Rényi entropy is evaluated for randomly initialized non-autoregressive vision transformers as well as Gutzwiller projected hidden fermion determinant states. For various widths of the distribution from which the network weights are sampled, the hidden fermion determinant state exhibits a higher value of the second Rényi entropy than the vision transformer. Simultaneously, a larger width of the distribution of the network weights consistently leads to a higher entanglement entropy for the vision transformer, as predicted by the arguments in [24]. From the arguments discussed in Sec. 2.2.1, a high entanglement content is expected for the determinant based construction.

3 Training neural quantum states

In order to represent a desired quantum state, we need to find the best variational parameters θ for the problem at hand. The procedure of optimizing these parameters is often referred to as training, and typically consists of an iterative procedure, where the parameters are changed by some small amount $\delta\theta$ in each training step. Given the representation we have (configuration in, wavefunction coefficient out), everything is based on samples from the quantum many-body state.

Below, we first outline how to evaluate observables using samples from the probability distribution defined by our neural quantum state. This constitutes the main tool required to implement the variational Monte Carlo (VMC) method. The core objective of VMC is to optimize the parameters θ of a trial wavefunction $|\psi_\theta\rangle$ in order to find the best possible approximation to our target state, e.g., the ground state of a given Hamiltonian. The same variational framework can be used for several tasks. In ground-state searches, one minimizes the variational energy. In real- or imaginary-time evolution, one updates the parameters such that the variational state follows the projected time-dependent Schrödinger equation. If a target state is known, one may instead minimize the infidelity, and if measurement data are available, one can train the NQS by state reconstruction. In all cases, the central computational ingredients are wavefunction ratios, Monte Carlo samples, and parameter derivatives of the log-amplitude.

3.1 Variational Monte Carlo

The expectation value of a general quantum-mechanical observable $\hat{O}$ with respect to the trial wavefunction $|\psi_\theta\rangle$ is

$$\langle\hat{O}\rangle_\theta = \frac{\langle\psi_\theta|\hat{O}|\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle}. \quad (53)$$

By inserting a resolution of the identity,

$$\sum_{\mathbf{s}} |\mathbf{s}\rangle\langle\mathbf{s}| = \mathbb{I}, \quad (54)$$

and using the Born-rule probability distribution

$$p_\theta(\mathbf{s}) = \frac{|\psi_\theta(\mathbf{s})|^2}{\sum_{\mathbf{s}'} |\psi_\theta(\mathbf{s}')|^2}, \quad (55)$$

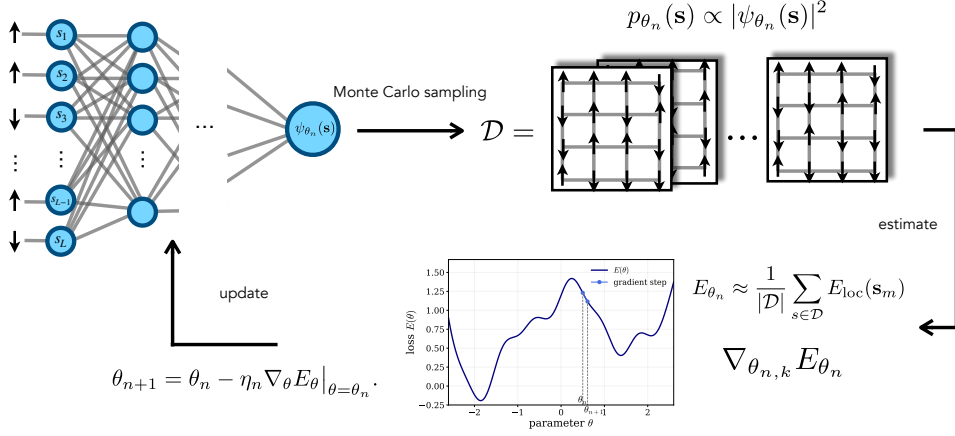


Fig. 3: Updating the neural network parameters according to the approximative gradient of the energy obtained via Monte Carlo sampling.

we can rewrite the expectation value as an average over classical configurations:

$$\langle \hat{O} \rangle_{\theta} = \sum_{\mathbf{s}} p_{\theta}(\mathbf{s}) O_{\text{loc}}(\mathbf{s}) = \mathbb{E}_{p_{\theta}} [O_{\text{loc}}(\mathbf{s})]. \quad (56)$$

Here we have defined the local observable

$$O_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \frac{\psi_{\theta}(\mathbf{s}')}{\psi_{\theta}(\mathbf{s})}. \quad (57)$$

This formulation allows us to estimate the quantum expectation value using a classical Monte Carlo average. We first generate M samples $\{\mathbf{s}_1, \dots, \mathbf{s}_M\}$ from the distribution $p_{\theta}(\mathbf{s})$, for example using the Metropolis algorithm:

Our quantum state $|\psi_{\theta}\rangle$ encodes a probability distribution over samples $\mathbf{s}$ given by

$$p_{\theta}(\mathbf{s}) = \frac{|\langle \mathbf{s} | \psi_{\theta} \rangle|^2}{\langle \psi_{\theta} | \psi_{\theta} \rangle} = \frac{|\psi_{\theta}(\mathbf{s})|^2}{\sum_{\mathbf{s}'} |\psi_{\theta}(\mathbf{s}')|^2}. \quad (58)$$

This expression defines a probability distribution over the entire computational basis. For our Neural Quantum State ansatz, $|\psi_{\theta}\rangle$, the probability of a configuration $\mathbf{s}$ is therefore proportional to $|\psi_{\theta}(\mathbf{s})|^2$. We can now employ Markov-chain Monte Carlo to draw samples from this quantum mechanical probability distribution. Starting from a configuration $\mathbf{s}$, one proposes a local or nonlocal update $\mathbf{s} \rightarrow \mathbf{s}'$ and accepts it with probability

$$A(\mathbf{s} \rightarrow \mathbf{s}') = \min \left[1, \frac{|\psi_{\theta}(\mathbf{s}')|^2 T(\mathbf{s}' \rightarrow \mathbf{s})}{|\psi_{\theta}(\mathbf{s})|^2 T(\mathbf{s} \rightarrow \mathbf{s}')} \right], \quad (59)$$

where T denotes the proposal probability. After sufficiently many such Monte Carlo proposals and updates, the samples follow the distribution $p_{\theta}(\mathbf{s})$ as desired. In contrast, autoregressive neural quantum states factorize the probability distribution as

$$p_{\theta}(\mathbf{s}) = \prod_{i=1}^N p_{\theta}(s_i | s_1, \dots, s_{i-1}), \quad (60)$$

such that configurations can be sampled directly from the normalized conditional probabilities, without a Markov-chain. This avoids autocorrelation between samples, but imposes an ordering of the degrees of freedom and requires an architecture compatible with the autoregressive factorization. This is often achieved through an explicit masking.

Upon sampling M samples from the distribution $p_\theta(\mathbf{s})$, we compute the sample mean of the local observable

$$\langle \hat{O} \rangle_\theta \approx \frac{1}{M} \sum_{m=1}^M O_{\text{loc}}(\mathbf{s}_m). \quad (61)$$

In order to find the ground state, we need to evaluate the expectation value of the Hamiltonian $\hat{H}$, which defines the variational energy, $E_\theta = \langle \hat{H} \rangle_\theta$. The corresponding local observable is the local energy,

$$E_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} H_{\mathbf{s},\mathbf{s}'} \frac{\psi_\theta(\mathbf{s}')}{\psi_\theta(\mathbf{s})}, \quad (62)$$

where

$$H_{\mathbf{s},\mathbf{s}'} = \langle \mathbf{s} | \hat{H} | \mathbf{s}' \rangle. \quad (63)$$

The variational energy can therefore be estimated as

$$E_\theta = \mathbb{E}_{p_\theta} [E_{\text{loc}}(\mathbf{s})] \approx \frac{1}{M} \sum_{m=1}^M E_{\text{loc}}(\mathbf{s}_m), \quad \mathbf{s}_m \sim p_\theta(\mathbf{s}). \quad (64)$$

To evaluate the local energy for a given sample $\mathbf{s}$, one only needs the configurations $\mathbf{s}'$ connected to $\mathbf{s}$ by nonzero Hamiltonian matrix elements $H_{\mathbf{s},\mathbf{s}'}$, together with the corresponding wavefunction ratios

$$\frac{\psi_\theta(\mathbf{s}')}{\psi_\theta(\mathbf{s})}. \quad (65)$$

3.1.1 Simple ground state search

For a ground state search, we want to minimize the energy, i.e., our cost function is E_θ . In order to optimize the network parameters θ , we must compute its gradient with respect to each variational parameter θ_k . The analytic expression for this gradient, often derived using the log-derivative trick, is given by a covariance

$$\nabla_{\theta_k} E_\theta = 2\text{Re} (\mathbb{E}_{p_\theta} [E_{\text{loc}} \mathcal{O}_k^*] - \mathbb{E}_{p_\theta} [E_{\text{loc}}] \mathbb{E}_{p_\theta} [\mathcal{O}_k^*]). \quad (66)$$

This introduces a new quantity, $\mathcal{O}_k(\mathbf{s})$, which is the logarithmic derivative of the wavefunction

$$\mathcal{O}_k(\mathbf{s}) = \frac{1}{\psi_\theta(\mathbf{s})} \frac{\partial \psi_\theta(\mathbf{s})}{\partial \theta_k} = \partial_{\theta_k} \log \psi_\theta(\mathbf{s}). \quad (67)$$

For a neural network, this derivative is readily available via backpropagation. It measures how sensitively the log-amplitude changes under a variation of parameter θ_k .

Crucially, the gradient can be estimated from the same set of M samples used to compute the energy

$$\nabla_{\theta_k} E_\theta \approx 2\text{Re} \left(\frac{1}{M} \sum_{i=1}^M E_{\text{loc}}(\mathbf{s}_i) \mathcal{O}_k^*(\mathbf{s}_i) - \left(\frac{1}{M} \sum_{i=1}^M E_{\text{loc}}(\mathbf{s}_i) \right) \left(\frac{1}{M} \sum_{i=1}^M \mathcal{O}_k^*(\mathbf{s}_i) \right) \right) \quad (68)$$

With the ability to compute this stochastic gradient, we can employ standard gradient-based optimization algorithms, such as Stochastic Gradient Descent (SGD) or Adam, to iteratively update the network parameters and minimize the energy

$$\theta \leftarrow \theta - \eta \nabla_\theta E_\theta, \quad (69)$$

where η is the learning rate. This iterative process of sampling, calculating the energy and its gradient, and updating the parameters constitutes the complete VMC optimization loop.

3.1.2 Stochastic reconfiguration

Simple gradient descent updates the parameters according to the Euclidean gradient in parameter space. However, the physically relevant distance is not the Euclidean distance between parameter vectors, but the distance between the corresponding quantum states. Stochastic reconfiguration (SR), an established tool in the variational Monte Carlo community, takes this geometry into account. The basic idea of SR is to perform an approximate imaginary time evolution [48] and thereby ultimately projecting onto the ground state: Let $\{|n\rangle\}$ be the eigenstates of the Hamiltonian,

$$\hat{H}|n\rangle = E_n|n\rangle, \quad E_0 \leq E_1 \leq E_2 \leq \dots \quad (70)$$

Starting from an initial state with nonzero overlap with the ground state,

$$|\psi(0)\rangle = \sum_n c_n |n\rangle, \quad c_0 \neq 0, \quad (71)$$

imaginary-time evolution gives

$$|\psi(\tau)\rangle = e^{-\tau \hat{H}} |\psi(0)\rangle = \sum_n c_n e^{-\tau E_n} |n\rangle. \quad (72)$$

After factoring out the ground-state contribution,

$$|\psi(\tau)\rangle = e^{-\tau E_0} \left[c_0 |0\rangle + \sum_{n>0} c_n e^{-\tau(E_n - E_0)} |n\rangle \right]. \quad (73)$$

Thus all excited-state contributions are exponentially suppressed relative to the ground state. After normalization,

$$\frac{|\psi(\tau)\rangle}{\| |\psi(\tau)\rangle \|} \xrightarrow{\tau \rightarrow \infty} |0\rangle. \quad (74)$$

Stochastic reconfiguration can be understood as projecting this imaginary-time evolution onto the variational manifold spanned by the neural quantum state ansatz.

For a small parameter update $\delta\theta$, we linearize the variational state as

$$|\psi_{\theta+\delta\theta}\rangle \approx |\psi_\theta\rangle + \sum_k \delta\theta_k \partial_{\theta_k} |\psi_\theta\rangle. \quad (75)$$

Imaginary-time evolution, on the other hand, changes the state in linear approximation as

$$|\psi_\theta\rangle \longrightarrow (1 - \delta\tau \hat{H}) |\psi_\theta\rangle. \quad (76)$$

Stochastic reconfiguration chooses $\delta\theta$ such that the variational change of the wavefunction best approximates this imaginary-time step. This leads to the linear system⁵

$$\sum_l S_{kl} \delta\theta_l = -\delta\tau F_k, \quad (77)$$

where

$$S_{kl} = \langle O_k^* O_l \rangle_c = \langle O_k^* O_l \rangle - \langle O_k^* \rangle \langle O_l \rangle \quad (78)$$

is the quantum geometric tensor, or covariance matrix of logarithmic derivatives, and

$$F_k = \langle O_k^* E_{\text{loc}} \rangle_c = \langle O_k^* E_{\text{loc}} \rangle - \langle O_k^* \rangle \langle E_{\text{loc}} \rangle. \quad (79)$$

Here $O_k(s)$ is the logarithmic derivative as before. The SR update therefore reads

$$\delta\theta = -\delta\tau S^{-1} F. \quad (80)$$

Equivalently, SR can be viewed as a natural-gradient method on the variational manifold. The quantum geometric tensor or S -matrix measures how much the physical quantum state changes when one changes the parameters θ_k . Compared to ordinary gradient descent, the SR update therefore takes into account how sensitively the quantum state changes under variations of the parameters. Note that, as SR for ground state search is based on imaginary time evolution, one can use the same derivation to simulate a real time evolution by simply inserting the complex i in the SR equation.

Historically, SR was developed in the variational Monte Carlo community before the use of neural quantum states, where typically only a few variational parameters are used [48]. Neural networks on the other hand often operate with orders of magnitude more variational parameters, such that the inversion of the S matrix can become computationally prohibitive. This has motivated approximate or iterative variants of SR, including minimum-step stochastic reconfiguration (minSR) [49] and related linear-algebra reformulations of SR [51], which aim to retain the geometric advantages of SR while scaling to modern deep NQS architectures.

Building on these scalable formulations of stochastic reconfiguration, several recent works have adapted ideas from modern optimization to the SR update. The SPRING optimizer (Subsampled Projected-Increment Natural Gradient Descent) combines the minimum-step formulation of SR with randomized Kaczmarz-type updates, using subsampled projected increments to accelerate the solution of the SR linear system [52]. A related direction is the MARCH optimizer

⁵For a detailed derivation, see e.g. Refs. [48–50].

(Moment-Adaptive ReConfiguration Heuristic), which further incorporates first- and second-moment information, in analogy with the way Adam extends momentum-based gradient descent. In practice, MARCH can be viewed as an adaptive-momentum variant of SR/MinSR designed to improve stability and convergence in large-scale NQS optimizations [53]. These scalable formulations address the cost of solving the SR equations, but they do not remove the need for regularization: in practice, small singular values or noisy directions are still damped or truncated to stabilize the update.

3.2 Trainability and design choices

The fact that a given network architecture has the capability to represent a certain class of quantum states, such as volume law entangled states, does not necessarily imply the conclusion that a specific or even any volume law entangled state can be found upon training the neural quantum state. As a concrete example, for restricted Boltzmann machines, Ref. [5] showed that any probability distribution over discrete variables, and therefore the wavefunction amplitudes for any quantum state, can be approximated arbitrarily well. However, such an approximation might require the network weights to become extremely large, which can be very hard to optimize to [24].

In general, beyond theoretical considerations as discussed above, these questions can be extremely hard to separate: if we perform variational Monte Carlo for example to find the ground state of a system and do not converge to the true ground state, is the reason the limited capability of the ansatz, or the challenging optimization landscape which keeps us from finding the network parameters encoding the desired quantum state? Below we discuss some design choices as well as possible routes towards convergence.

Reference basis.— First, one must choose a reference basis. Since the network represents amplitudes $\psi_\theta(\mathbf{s})$ in this basis, the complexity of the amplitudes and phases, as well as the optimization problem, can depend strongly on the basis choice. A basis close to an approximate eigenbasis can make the wavefunction compact, with most weight concentrated on a reference configuration and physically interpretable excitations [54]. This compactness, however, is not automatically advantageous for a generic NQS. A very peaked distribution can make sampling difficult, since rare configurations may still contribute substantially to the local energy or gradient. This sampling issue is closely related to the support-mismatch problem discussed in Ref. [55]: even if a wavefunction is compact in a given basis, standard VMC estimators can become unstable or biased when the sampling distribution fails to cover configurations that are needed to evaluate local energies, gradients, or time-evolution updates. A peaked distribution can also make optimization difficult: representing near-zero amplitudes over a large part of Hilbert space often requires large logits, weights, or biases, which may be hard to reach in practice and can lead to ill-conditioned updates. A flatter basis distribution can therefore be preferable, even if it is less compact, because the required correlations can be encoded with lower-order terms and more moderate parameter values. This viewpoint is consistent with anal-

yses of NQS basis dependence in terms of correlator, cluster, or cumulant expansions, where the relevant question is not only how localized the probability distribution is, but how simple the amplitude and phase structure are in the chosen basis [24, 56].

A further complication is that basis choice can affect trainability even when the target state remains representable. Kožić et al. showed in a rotated Ising model that local basis rotations can relocate the exact wavefunction within the variational loss landscape, increasing its geometric distance from typical initializations and making optimization more likely to become trapped in saddle-point or high-curvature regions. Thus, basis dependence is not only a question of expressivity or sampling, but also of the geometry of the optimization landscape [57]. The best basis is therefore not universally the most localized or the flattest one, but the one in which the target state is simple for the chosen architecture, sampling strategy, and optimizer.

Reference sign or phase structure.— Second, one may build in a reference sign or phase structure. For real wavefunctions it is often useful to separate amplitudes and signs,

$$\psi(\mathbf{s}) = |\psi(\mathbf{s})| S(\mathbf{s}), \quad S(\mathbf{s}) \in \{-1, +1\}. \quad (81)$$

More generally, for complex wavefunctions one may separate amplitude and phase. The sign or phase structure can be the most difficult part of the wavefunction to learn. Supervised studies of frustrated magnets found that learning the signs generalizes much worse than learning the amplitudes [58], and variational studies of the frustrated J_1 - J_2 model showed that optimization can get stuck in low-energy states with an incorrect or overly simple sign structure [59, 60]. A common strategy is therefore to build in part of the sign structure explicitly. For bipartite antiferromagnets, the Marshall sign rule provides such a reference sign, and it has often been used either as an explicit gauge transformation or as a prior to stabilize NQS optimizations [61, 51]. This strategy is powerful when the reference sign is accurate, but it can also bias the optimization if the true ground-state sign structure deviates from it.

For doped or fermionic lattice models, the sign structure becomes even more subtle. In Fock space, the basis states already contain the fermionic operator ordering. The neural network represents the coefficients in this fixed Fock basis. Thus, in a second-quantized formulation, a determinant-based ansatz should not be described as “enforcing antisymmetry” of the coefficient function under particle exchange. Rather, the determinant provides a structured sign and nodal pattern for the Fock-space amplitudes, inherited from the underlying single-particle orbitals or from their configuration-dependent generalizations. This can be a useful physical prior for fermionic systems, as seen in determinant-based NQS for the t - J model [31]. At the same time, studies of autoregressive RNNs for doped antiferromagnets show that the difficulty is not only the fermionic sign: doping also enlarges the Hilbert space and can make both amplitude and phase distributions substantially more complex [62]. More recently, Boolean Fourier analysis has provided another perspective on sign structures by expanding $S(\mathbf{s})$ in parity functions,

$$S(\mathbf{s}) = \sum_{I \subseteq \{1, \dots, N\}} a_I \prod_{i \in I} s_i. \quad (82)$$

In this language, sign learning is related to identifying the relevant high-order parity features. This suggests that difficult sign structures may not be structureless, but may require architectures or input features adapted to the appropriate correlators [63].

Symmetries.— Third, symmetries can be imposed either by the architecture or by explicit projection. Enforcing particle-number conservation, spin symmetries, translation invariance, point-group symmetries, or gauge constraints reduces the effective variational space and can improve both expressivity and trainability. Autoregressive models can, e.g., be constrained to sample only configurations with the desired quantum numbers. In particular, $U(1)$ -symmetric RNNs preserve the autoregressive structure while restricting the samples to a fixed magnetization or particle-number sector [64–66]. Spatial symmetries can be incorporated by symmetrizing the wavefunction over the symmetry group, by targeting fixed quantum-number sectors, or by using equivariant architectures such as group convolutional networks [67, 68]. The practical tradeoff can be computational: explicit projection over a symmetry group often requires evaluating the ansatz on all symmetry-related configurations, increasing the cost by the size of the group. For determinant- or Pfaffian-based fermionic states this overhead can be particularly relevant, since each symmetry image may require an additional determinant or Pfaffian evaluation. Built-in equivariance can reduce the overhead of explicit symmetry projection by enforcing the symmetry at the architectural level. For example, group convolutional neural networks implement lattice symmetries directly through equivariant layers and have been shown to improve NQS accuracy without increasing memory use [68]. In practice, the best-performing NQS are often not generic neural networks, but those in which physical structure is built in from the start.

Initialization.— In order to ease the optimization problem, one can initialize the network parameters in a favorable place in the optimization landscape. This can be done by setting the network weights according to the entries of a matrix product state [69], which can be efficiently optimized within the capacities of MPS using the DMRG. While this is in general an intriguing option, it can in practice limit the architectural choices of the neural quantum state. Another option is to first perform a state reconstruction task by using data from a quantum simulator. If the available data are projective measurements in one basis, that basis can be chosen as the reference basis of the NQS and the Kullback-Leibler divergence or Wasserstein metric can be minimized to match the probability distribution given by the NQS with the measured data distribution [70–72]. Going beyond snapshots in one basis is possible by either considering a few local rotations, which can still be computed [73], or by additionally performing a minimization to match global correlations in one or more bases beyond the reference basis [72]. These works have demonstrated that improved convergence of the subsequent ground state search can be achieved if the NQS is initialized using data from a state that is close to the ground state. This advantage remains when using experimental data, where state preparation and measurement imperfections can yield a data set significantly different from the ideal ground state. In the absence of numerical or experimental data, one can also choose the random distribution and its width in a favorable manner as discussed in Sec. 2.2, see also e.g. Ref. [46].

Quantum geometric tensor.— One perspective on the complexity of the optimization landscape is provided by the quantum geometric tensor of the variational manifold (the S -matrix) introduced in the context of stochastic reconfiguration above. Its spectrum gives local information about which parameter directions correspond to physically distinct changes of the quantum state. Large eigenvalues correspond to directions in which the state changes strongly, while small eigenvalues correspond to flat, redundant, or poorly sampled directions. In the SR update, Eq. (80), the energy gradient is preconditioned with the inverse of S , so directions in which a small parameter change produces a large change of the quantum state are correspondingly rescaled.

The effective rank of the quantum geometric tensor gives an estimate of the number of linearly independent directions in Hilbert space that are accessible by infinitesimal changes of the variational parameters and can thus be used as a diagnostic of the practical representational power of an ansatz. If the rank saturates as the network is widened, additional parameters are largely redundant for the target state. Note however that the relevant geometry of the variational manifold evolves during training: The spectrum of the QGT can also change substantially during training and across physical regimes, as observed for complex RBM wavefunctions in Ref. [74] and for vision transformers and hidden fermion determinant states in [47]. The rank of the quantum geometric tensor should not be interpreted as a global measure of the expressivity of the ansatz. Rather, it measures the dimension of the local tangent space of physically distinct wavefunction variations at the current parameter point. Since this tangent space changes during training, the QGT rank is a local and trajectory-dependent diagnostic. It is nevertheless useful in practice: if increasing the number of parameters no longer increases the effective QGT rank near the optimized state, and the variational accuracy saturates at the same time, then the additional parameters are largely redundant for the state being represented [75]. Conversely, a growing QGT rank indicates that the enlarged ansatz provides additional independent directions in Hilbert space that the optimizer can use. In Ref. [75], the width of the network is increased at fixed depth to increase the number of parameters. From the QGT point of view, increasing width and increasing depth explore different extensions of the variational manifold. Width adds more tangent vectors of the same structural form, whereas depth changes the non-linear map itself, as discussed in Sec. 2.2.1, and can therefore modify the tangent space qualitatively. A saturation of the QGT rank under widening indicates local redundancy within that family, but it does not rule out an increase of the effective rank, or an improvement in accuracy, under a change of architecture such as adding depth.

The matrix S describes the geometry of the variational manifold, but it does not by itself determine the update. The second ingredient is the force vector F_k , which measures the overlap between the tangent vector $\partial_{\theta_k}|\psi_\theta\rangle$ and the imaginary-time direction $-(\hat{H}-E)|\psi_\theta\rangle$. Thus F tells us which directions in the tangent space are energetically useful, while S tells us how these tangent vectors overlap with one another. The SR equation

$$S \delta\theta = -\eta F \quad (83)$$

therefore finds the linear combination of tangent vectors that best approximates the imaginary-time update within the variational manifold. Convergence depends not only on the rank of

S , but also on whether the force vector has support in well-conditioned directions of S . If F points mostly into small-eigenvalue or noisy directions, the SR update becomes unstable and regularization is required.

In combination with the variational error and the stochastic-reconfiguration force vector, the QGT spectrum provides a useful diagnostic of whether convergence is limited by insufficient local expressivity, ill-conditioning, or by the global optimization landscape [74, 75].

4 Concluding remarks

Neural quantum states provide a flexible variational ansatz for strongly correlated quantum systems. Their appeal lies in the possibility of combining neural-network expressivity with physical structure, such as locality, symmetries, reference signs, Jastrow factors, determinants, Pfaffians, or autoregressive factorizations. At the same time, expressivity alone is not sufficient: an architecture may be capable of representing a target state, but the corresponding parameters may still be difficult to find in practice. More broadly, understanding how expressive different NQS architectures really are, if and when they work well in practice, and why particular design choices succeed or fail remains an active area of research. These notes aim to provide some intuition for these questions by separating capability, typicality, and trainability, and by illustrating how architectural ingredients such as nonlinearities, connectivity, symmetries, basis choice, and fermionic determinant structures affect the states that can be represented and optimized. The usefulness of an NQS therefore depends on the interplay between capability, typicality, trainability, sampling, and the chosen reference basis. This is both a challenge and an opportunity. Rather than searching for a universally optimal neural network ansatz, current progress suggests that the most successful approaches are those in which machine-learning flexibility is combined with problem-specific physical insight. Neural quantum states should therefore be viewed not as a replacement for traditional variational ansätze, but as a framework that can extend them in a highly expressive form.

Acknowledgments

I am grateful to the many people from whom I learned and with whom I discussed all neural quantum state related aspects. I want to in particular thank (in alphabetical order) Julian Blasek, Annika Böhrer, Francesco Conoscenti, Fabian Döschl, Hannah Lange, Roger Melko, Balint Soczo, and Anka van de Walle, with whom I gained a deeper understanding of many of the things presented in these notes. I gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy-EXC-2111-390814868, by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme ERC Starting Grant QuaQuaMA (Grant No. 101217531), and by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

References

- [1] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)
- [2] G. Carleo and M. Troyer, Science **355**, 602 (2017)
- [3] G. Cybenko, Math. Control Signals Syst. **2**, 303 (1989)
- [4] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2**, 359 (1989)
- [5] N.L. Roux and Y. Bengio, Neural Comput. **20**, 1631 (2008)
- [6] M. Medvidović and J.R. Moreno, Eur. Phys. J. Plus **139** (2024)
- [7] H. Lange, A.V. de Walle, A. Abedinnia, and A. Bohrdt, Quantum Sci. Technol. **9**, 040501 (2024)
- [8] M. Schmitt and M. Heyl, arXiv 2506.03124 (2025)
- [9] D.-L. Deng, X. Li, and S. Das Sarma, Phys. Rev. X **7** (2017)
- [10] Z. Denis, A. Sinibaldi, and G. Carleo, Phys. Rev. Lett. **134** (2025)
- [11] A.Y. Kitaev, Ann. Phys. **303**, 2 (2003)
- [12] M.Y. Pei and S.R. Clark, J. Phys. A: Math. Theor. **54**, 405304 (2021)
- [13] S.R. Clark, J. Phys. A: Math. Theor. **51**, 135301 (2018)
- [14] R. Kaubruegger, L. Pastori, and J.C. Budich, Phys. Rev. B **97**, 195136 (2018)
- [15] I. Glasser, N. Pancotti, M. August, I.D. Rodriguez, and J.I. Cirac, Phys. Rev. X **8** (2018)
- [16] P. Chen, B. Yan, and S.X. Cui, Phys. Rev. B **111** (2025)
- [17] Y. Huang and J.E. Moore, Phys. Rev. Lett. **127**, 170601 (2021)
- [18] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Phys. Rev. B **97**, 085104 (2018)
- [19] O. Sharir, A. Shashua, and G. Carleo, Phys. Rev. B **106**, 205136 (2022)
- [20] D. Wu, R. Rossi, F. Vicentini, and G. Carleo, Phys. Rev. Res. **5** (2023)
- [21] F. Becca and S. Sorella: *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017)
- [22] H. Lange: *New approaches to strongly correlated electron systems* (PhD Thesis, LMU München, 2026)
- [23] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Proc. Natl. Acad. Sci. USA **119** (2022)

- [24] F. Döschl and A. Bohrdt, arXiv:2508.14152 (2025)
- [25] Z. Liu and B.K. Clark, Phys. Rev. B **110**, 115124 (2024)
- [26] Y. Gu, Z. Han, W. Li, Z. Xiao, T. Xiang, M. Qin, L. Wang, and D. Lv, arXiv (2026)
- [27] Y. Nomura, A.S. Darmawan, Y. Yamaji, and M. Imada, Phys. Rev. B **96**, 205152 (2017)
- [28] S. Humeniuk, Y. Wan, and L. Wang, SciPost Phys. **14**, 171 (2023)
- [29] D. Luo and B.K. Clark, Phys. Rev. Lett. **122**, 226401 (2019)
- [30] A. Chen, Z.-Q. Wan, A. Sengupta, A. Georges, and C. Roth, arXiv (2025)
- [31] H. Lange, A. Böhler, C. Roth, and A. Bohrdt, Phys. Rev. Lett. **135**, 136504 (2025)
- [32] H. Lange, A. Chen, A. Georges, F. Grusdt, A. Bohrdt, and C. Roth, arXiv:2602.10091
- [33] H. Lange, J. Tirpitz, and A. Bohrdt, in preparation (2026)
- [34] C. Gauvin-Ndiaye, J. Tindall, J.R. Moreno, and A. Georges, Phys. Rev. Lett. **134**, 076502 (2025)
- [35] D. Pfau, J.S. Spencer, A.G. d.G. Matthews, and W.M.C. Foulkes, Phys. Rev. Res. **2**, 033429 (2020)
- [36] J. Hermann, Z. Schätzle, and F. Noé, Nat. Chem. **12**, 891 (2020)
- [37] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, Commun. Phys. **7**, 148 (2024)
- [38] L. Fu: *Fermi Sets: Universal and interpretable neural architectures for fermions* (2026)
- [39] L. Zhang, G. Duan, and D. Luo, arXiv:2605.29683 (2026)
- [40] M.B. Hastings, I. González, A.B. Kallin, and R.G. Melko, Phys. Rev. Lett. **104**, 157201 (2010)
- [41] P. Mehta, M. Bukov, C.-H. Wang, A.G. Day, C. Richardson, C.K. Fisher, and D.J. Schwab, Phys. Rep. **810**, 1 (2019)
- [42] N. Paul, Phys. Rev. Lett. **136**, 120403 (2026)
- [43] X. Gao and L.-M. Duan, Nat. Commun. **8** (2017)
- [44] B. Swingle, Phys. Rev. Lett. **105**, 050502 (2010)
- [45] G. Passetti, D. Hofmann, P. Neiteimeier, L. Grunwald, M.A. Sentef, and D.M. Kennes, Phys. Rev. Lett. **131** (2023)

- [46] A. Jreissaty, H. Zhang, J.C. Quijano, J. Carrasquilla, and R. Wiersema, *Phys. Rev. Res.* **8** (2026)
- [47] F. Conoscenti, H. Lange, F. Döschl, C. Kuhlenkamp, J. Carrasquilla, and A. Bohrdt, in preparation (2026)
- [48] S. Sorella, *Phys. Rev. Lett.* **80**, 4558 (1998)
- [49] A. Chen and M. Heyl, *Nat. Phys.* **20**, 1476 (2024)
- [50] A. Chen: *Large-scale simulation of deep neural quantum states* (PhD Thesis, Universität Augsburg, 2025)
- [51] R. Rende, L.L. Viteritti, L. Bardone, F. Becca, and S. Goldt, *Commun. Phys.* **7**, 260 (2024)
- [52] G. Goldshlager, N. Abrahamsen, and L. Lin, arXiv (2024)
- [53] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, L. Wang, and D. Lv, arXiv (2025)
- [54] J.A. Sobral, M. Perle, and M.S. Scheurer, *Nat. Commun.* **16** (2025)
- [55] Z.-Q. Wan, R. Wiersema, and S. Zhang, arXiv:2603.18148 (2026)
- [56] R.S. Cortes Santamaria, A.S. Shankar, M. Dalmonte, R. Verdel, and N. Niggemann, *SciPost Phys. Core* **9** (2026)
- [57] S.B. Kožić, V. Zlatić, F. Franchini, and S.M. Giampaolo, arXiv:2512.17893 (2025)
- [58] T. Westerhout, N. Astrakhantsev, K.S. Tikhonov, M.I. Katsnelson, and A.A. Bagrov, *Nat. Commun.* **11**, 1593 (2020)
- [59] M. Bukov, M. Schmitt, and M. Dupont, *SciPost Phys.* **10**, 147 (2021)
- [60] A. Szabó and C. Castelnovo, *Phys. Rev. Res.* **2**, 033075 (2020)
- [61] W. Marshall, *Proc. Roy. Soc. A* **232**, 48 (1955)
- [62] H. Lange, F. Döschl, J. Carrasquilla, and A. Bohrdt, *Commun. Phys.* **7**, 187 (2024)
- [63] I. Schurov, A. Kravchenko, M.I. Katsnelson, A.A. Bagrov, and T. Westerhout, arXiv:2508.09870 (2025)
- [64] M. Hibat-Allah, M. Ganahl, L.E. Hayward, R.G. Melko, and J. Carrasquilla, *Phys. Rev. Res.* **2**, 023358 (2020)
- [65] S. Morawetz, I.J.S. De Vlugt, J. Carrasquilla, and R.G. Melko, *Phys. Rev. A* **104**, 012401 (2021)

- [66] A. Malyshev, J.M. Arrazola, and A.I. Lvovsky, arXiv (2023)
- [67] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Phys. Rev. Lett. **121**, 167204 (2018)
- [68] C. Roth and A.H. MacDonald, arXiv:2104.05085 (2021)
- [69] R. Kaneko and S. Goto, Phys. Rev. B **112**, 155163 (2025)
- [70] S. Czischek, M.S. Moss, M. Radzihovsky, E. Merali, and R.G. Melko, Phys. Rev. B **105** (2022)
- [71] M.S. Moss, S. Ebadi, T.T. Wang, G. Semeghini, A. Bohrdt, M.D. Lukin, and R.G. Melko, Phys. Rev. A **109**, 032410 (2024)
- [72] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, Quantum **9**, 1675 (2025)
- [73] E.R. Bennewitz, F. Hopfmueller, B. Kulchytskyy, J. Carrasquilla, and P. Ronagh, Nat. Mach. Intell. **4**, 618 (2022)
- [74] C.-Y. Park and M.J. Kastoryano, Phys. Rev. Res. **2**, 023232 (2020)
- [75] S. Dash, L. Gravina, F. Vicentini, M. Ferrero, and A. Georges, Commun. Phys. **8**, 92 (2025)

15 Quantum Many-Body Simulations on Digital Quantum Computers

Benedikt Fauseweh
TU Dortmund University
Otto-Hahn-Str 4, 44227 Dortmund

Contents

1	Introduction	2
1.1	A computational interpretation of quantum mechanics	2
1.2	Quantum mechanics as a generalized probability theory	3
1.3	The formal task of quantum simulation	4
1.4	Physics as a quantum program	5
2	Quantum bits and quantum circuits	6
2.1	Quantum bit and unitary gates	6
2.2	Entanglement and controlled-NOT	8
2.3	Quantum circuit diagrams	9
2.4	Universality of single-qubit gates and the CNOT	10
3	Simulating strongly correlated systems on quantum computers	11
3.1	Non-equilibrium dynamics of spin systems	12
3.2	Simulating fermions	14
3.3	Variational quantum eigensolver	18
4	Quantum error correction	21
4.1	Introduction to error correction	21
4.2	Bit and phase flip errors	23
4.3	Partial error correction for non-equilibrium phases of matter	25

1 Introduction

Solving the quantum many-body problem is notoriously difficult, due to the exponential scaling of the total Hilbert space dimension with the number of particles. We can understand this problem by starting from the arguably simplest quantum mechanical system: a two-level atom. We denote the orthonormal states $|0\rangle$ and $|1\rangle$ as a basis in the two-dimensional Hilbert space. A quantum system built from N coupled two-level systems defines a Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes N}$ of dimension 2^N . A generic pure state

$$|\psi\rangle = \sum_{x \in \{0,1\}^N} c_x |x\rangle, \quad \sum_x |c_x|^2 = 1, \quad (1)$$

is determined by 2^N complex amplitudes c_x and hence the number of parameters needed merely to write down the state grows exponentially with N . In contrast the resources of the laboratory, e.g., the number of atoms, grow typically linearly with N . To follow the time evolution of such a state on a classical computer one must, in the absence of other special structures like symmetries, store and update this exponentially long list. Recently such a time evolution was implemented numerically on classical computers for $N = 50$ [1], using the JUPITER supercomputer at Forschungszentrum Jülich. It is probably safe to say, that a vector with $N = 300$ and 128 bits of precision will not be used in a classical simulation in our lifetime, as the number of amplitudes is larger than the number of atoms in the visible universe. This single observation, that the description of correlated quantum matter is generically intractable by classical means, is the reason we want to think more about what it means to compute a quantum system.

A possible response is to turn the difficulty into a tool. If a classical computer struggles because it must track an exponential amount of quantum information, then a controllable quantum system, which carries that information natively, should not. The task of using one quantum system to imitate the dynamics of another is *quantum simulation*, and the realization of that task on a programmable, gate-based quantum computer is *digital quantum simulation*.

1.1 A computational interpretation of quantum mechanics

The idea that quantum mechanics can be used computationally on its own emerged around 1980. Manin remarked that the dimension of a many-particle state space multiplies under composition, whereas classical configuration spaces only add, and suggested that “quantum automata” might therefore compute in ways no classical device could [2]. Benioff constructed an explicit quantum-mechanical Hamiltonian whose unitary evolution implements an arbitrary Turing machine, showing that classical computation is contained within quantum dynamics [3].

The proposal of a physical quantum computer is due to Feynman [4] in 1982. Feynman asked whether a classical computer can simulate quantum physics, and argued that it cannot do so efficiently. A faithful classical model of the correlations measured on an entangled pair of photons, would have to reproduce the joint statistics of incompatible measurements, and any local assignment of probabilities to the underlying variables is forced to take negative values. Since a classical probabilistic machine samples from non-negative distributions, it cannot reproduce

these correlations without exponential overhead. Feynman then proposed building the simulator out of quantum elements that obey the same laws as the target, conjecturing that such a system could serve as a quantum computer.

In 1985, Deutsch placed this idea on a sharper footing [5] and defined the universal quantum computer as a model of computation in its own right. He introduced an artificial task on which it outperforms any classical machine, introducing the concept of “quantum parallelism”. The computational and the physical notions of simulation were thereby unified.

In 1994 Shor developed a quantum algorithm that factors integers and computes discrete logarithms exponentially faster than the best known classical method [6, 7]. Shor’s result detached quantum computation from its origin in physics simulation and established it as a general computational paradigm.

The idea begun by Feynman was further developed by Lloyd in 1996, who proved the conjecture that a quantum computer can efficiently simulate the time evolution of any quantum system governed by local interactions [8].

1.2 Quantum mechanics as a generalized probability theory

It is worth asking what separates a quantum device from a classical randomized one. The cleanest answer, due in this form to Aaronson [9], is that quantum mechanics is the theory one obtains by replacing classical probability with the ℓ^1 norm by the ℓ^2 norm.

A classical state over N outcomes is a probability vector $p = (p_1, \dots, p_N)$ with $p_i \geq 0$ and

$$\|p\|_1 = \sum_i p_i = 1. \quad (2)$$

The maps that send probability vectors to probability vectors are exactly the (column-)stochastic matrices, that is, the linear maps preserving the ℓ^1 norm for non-negative probabilities. A quantum state is instead a vector $|\psi\rangle \in \mathbb{C}^N$ normalized in the ℓ^2 sense,

$$\|\psi\|_2 = \left(\sum_i |\psi_i|^2 \right)^{1/2} = 1, \quad (3)$$

and the linear maps preserving this norm are exactly the unitaries $U \in \mathbf{U}(N)$. The Born rule $p_i = |\psi_i|^2$ is the probability distribution determined from the amplitudes at the moment of measurement, and Gleason’s theorem [10] singles it out as the only consistent choice in dimension three and above.

That the exponents 1 and 2 are the only ones that yield a non-trivial theory is not an accident. Demanding that a linear map preserve the ℓ^p norm imposes a system of polynomial constraints on its entries, and one can show that the resulting transformations are only permutations and sign flips unless $p = 1$, which gives classical probability, or $p = 2$, which gives quantum mechanics [11]. The use of complex rather than real amplitudes is fixed by a comparable demand: requiring that every allowed evolution possess a square root, so that it may be applied over half a time step, forces the field to be $\mathbb{C}$, although alternative formulations exist [12].

The single feature that distinguishes the ℓ^2 theory from the ℓ^1 theory is that amplitudes may be negative or complex, and so may cancel. Consider the Hadamard gate in the standard basis

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (4)$$

Acting on $|0\rangle$ it produces the superposition $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, which on measurement looks exactly like a fair coin. A classical fair coin tossed twice leaves one no more certain of the outcome than after one toss. The quantum coin behaves differently:

$$H^2 |0\rangle = |0\rangle. \quad (5)$$

The amplitude to reach $|1\rangle$ accumulates over two paths, through the intermediate $|0\rangle$ and through the intermediate $|1\rangle$, and these carry opposite signs,

$$\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \left(-\frac{1}{\sqrt{2}}\right) = 0, \quad (6)$$

so the outcome $|1\rangle$ disappears through destructive interference. No dynamics on non-negative probabilities can produce this cancellation. This is how a quantum computer differs uniquely from a classical computer.

1.3 The formal task of quantum simulation

In this section we want to formalize the task of quantum simulation mathematically. Throughout we set $\hbar = 1$. We start from the time-dependent Schrödinger equation

$$i \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle \quad (7)$$

which governs the evolution of a closed system under a Hamiltonian H . For time-independent H its solution is generated by the unitary propagator

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle, \quad U(t) = e^{-iHt}. \quad (8)$$

This defines the first and central problem, *Hamiltonian* or *dynamical simulation*: given a description of a local Hamiltonian H on n qubits, a time t , an accuracy ε , and a means of preparing an input $|\psi_0\rangle$, produce a state $|\phi\rangle$ with

$$\| |\phi\rangle - U(t) |\psi_0\rangle \| \leq \varepsilon. \quad (9)$$

In this context, *local* does not mean geometrically local, but it refers to the terms in the Hamiltonian that act nontrivially only on a small number of qubits. For example the term k -local, refers to a Hamiltonian in which every single term acts at most onto k qubits.

In contrast the time-independent Schrödinger equation poses a different problem,

$$H |n\rangle = E_n |n\rangle, \quad (10)$$

the determination of the spectrum and eigenstates of H , most often its ground state energy $E_0 \leq E_n$ and ground state $|0\rangle$ or equivalently its ground state manifold.

The two tasks are not of equal difficulty. Dynamical simulation (9) is tractable on a quantum computer, whereas estimating the ground-state energy of a generic local Hamiltonian is Quantum Merlin Arthur (QMA)-complete [13, 14]. A problem that is QMA-complete is believed to be hard even with a perfect quantum computer available. Thus simulating how a system evolves is one thing, and finding the ground state is another.

The tractability of the dynamical problem is the content of Lloyd's theorem [8]. Suppose the Hamiltonian is a sum of local terms,

$$H = \sum_{l=1}^L H_l, \quad (11)$$

each H_l is k -local. The exponential e^{-iHt} does not factor into a product over the H_l because the terms do not commute. It can, however, be approximated by such a product over a short time. Writing r for the number of time slices, the first-order Lie-Trotter formula [15] reads

$$e^{-iHt} = \left(\prod_{l=1}^L e^{-iH_l t/r} \right)^r + \mathcal{O}\left(\frac{t^2}{r} \sum_{l < j} \|[H_l, H_j]\|\right), \quad (12)$$

where the error term is the leading contribution of the Baker-Campbell-Hausdorff expansion and is controlled by the commutators of H_l [16]. To reach accuracy ε over time t it therefore suffices to take

$$r = \mathcal{O}(t^2/\varepsilon) \quad (13)$$

slices, each built from L elementary exponentials, every one of which acts on at most k constituents. The total count of exponentials grows polynomially in n , t and $1/\varepsilon$. For a quantum device the application of $e^{-iH_l t/r}$ is assumed to be the elementary operation, while for classical computers this step scales with the dimension of the Hilbert space $\dim \mathcal{H} = 2^N$. This is the precise sense in which a quantum computer simulates local quantum dynamics *efficiently*, while the direct classical evolution of (1) does not [8]. Higher-order product formulas [17] and methods based on sparse-Hamiltonian techniques and quantum signal processing [18, 19] improve the dependence on t and on ε .

1.4 Physics as a quantum program

Equation (12) allows us to interpret quantum dynamics from the perspective of computer science. The Hamiltonian is no longer only an object of study but a *program*: the physics of interest is specified by the list of local terms H_l , and the job of the simulator is to run the propagator $U(t)$ they generate. Two strategies exist for doing so, and divide the field of quantum simulation [20, 21].

In *analog* quantum simulation one engineers a controllable system whose native Hamiltonian H_{sim} reproduces, under a known mapping, the target H_{tgt} , and one simply lets it evolve. The

continuous evolution under H_{sim} is itself the computation and no decomposition into elementary steps occurs. Ultracold atoms in optical lattices realizing Hubbard models [22, 23] and trapped-ion spin chains [24] are examples of such analog simulations. Analog quantum simulations reach large particle numbers and incur no Trotter error in the ideal limit, but each is tied to a particular family of Hamiltonians and offers no route to error correction.

In *digital* quantum simulation, one instead compiles the propagator into a sequence of gates drawn from a fixed universal set, exactly as in Eq. (12). The price is a discretization error and a circuit depth that grows with t and $1/\varepsilon$. The reward is universality, since any local Hamiltonian can be simulated by the same hardware, and compatibility with quantum error correction, which is what makes accurate long-time simulation conceivable in principle.

2 Quantum bits and quantum circuits

In the first section we introduced the concept of quantum computers and the formal task of quantum simulation. In this section we want to make this approach more concrete. We define the quantum bit, characterize the operations one may perform on it, introduce the concept of entanglement and how to induce it on a quantum register, and prove that a small set of operations suffices to realize any arbitrary unitary transformation.

2.1 Quantum bit and unitary gates

A quantum bit, or qubit, is a normalized vector in a two-dimensional complex Hilbert space equipped with a orthonormal basis $\{|0\rangle, |1\rangle\}$,

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \quad |\alpha|^2 + |\beta|^2 = 1, \quad \alpha, \beta \in \mathbb{C}. \quad (14)$$

Of the four real parameters in α and β , normalization removes one and the irrelevance of a global phase removes another. Thus, without loss of generality, we can write a single qubit state as,

$$|\psi\rangle = \cos \frac{\vartheta}{2} |0\rangle + e^{i\varphi} \sin \frac{\vartheta}{2} |1\rangle, \quad \vartheta \in [0, \pi], \quad \varphi \in [0, 2\pi). \quad (15)$$

One interprets the angles (ϑ, φ) as the polar and azimuthal angles of a point on the unit sphere, the Bloch sphere. The basis states $|0\rangle$ and $|1\rangle$ sit at the poles, and orthogonal states are antipodal rather than separated by a right angle. In contrast to a classical bit, which can only occupy the poles, a qubit state can exist over the entire surface.

Between preparation and measurement a qubit evolves under the Schrödinger equation, whose solution is the propagator $U(t) = e^{-iHt}$ of Section 1.3. Thus, whenever we control a qubit by applying a pulse, turning on an interaction, or letting it evolve for a prescribed time, the corresponding operation on the state is a unitary transformation. A quantum gate is such a controlled unitary time evolution. Instead of describing the detailed Hamiltonian, pulse shape, and duration that realize an operation in a particular device, we describe the net effect on the

qubits by a unitary matrix U . For example, a gate may represent the evolution generated by some control Hamiltonian H_{gate} for a fixed time τ ,

$$U_{\text{gate}} = e^{-iH_{\text{gate}}\tau}. \quad (16)$$

In this sense, gates are not fundamentally different from time-evolution operators. They are time-evolution operators viewed at selected times and used as elementary building blocks of a computation. The defining property of a unitary gate is that its adjoint is its inverse,

$$U^\dagger U = U U^\dagger = I. \quad (17)$$

Hence every quantum gate is reversible. This reversibility is in contrast to many ordinary classical logic gates such as AND and OR. These gates map two input bits to one output bit and therefore discard information: from the output alone one cannot reconstruct the input. A unitary gate, by contrast, maps input states to output states with the same dimension. No information is lost during the coherent part of the quantum circuit. Another consequence of unitary gates is that they preserve the ℓ^2 norm of a state vector,

$$\|U|\psi\rangle\|^2 = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle \quad (18)$$

and therefore preserve the total probability discussed in Section 1.2. The action of a gate may redistribute probability amplitudes among basis states, and it may change their relative phases, but it cannot change the total probability.

The operators of one qubit can be expressed through the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (19)$$

Each is Hermitian, unitary and traceless, squares to the identity, and has eigenvalues ± 1 . They satisfy the algebra

$$\sigma_j \sigma_k = \delta_{jk} I + i \epsilon_{jkl} \sigma_l, \quad (20)$$

and together with the identity I they span the complex 2×2 matrices, so any single-qubit Hamiltonian takes the form $H = h_0 I + \mathbf{h} \cdot \boldsymbol{\sigma}$ with real h_0 and $\mathbf{h}$. Exponentiating a Pauli operator generates a rotation. For a unit vector $\mathbf{n}$, the identity $(\mathbf{n} \cdot \boldsymbol{\sigma})^2 = I$ gives

$$R_{\mathbf{n}}(\vartheta) = e^{-i\frac{\vartheta}{2} \mathbf{n} \cdot \boldsymbol{\sigma}} = \cos \frac{\vartheta}{2} I - i \sin \frac{\vartheta}{2} \mathbf{n} \cdot \boldsymbol{\sigma}, \quad (21)$$

which rotates the Bloch vector through angle ϑ about the axis $\mathbf{n}$. These rotations exhaust the single-qubit gates up to a phase: every $U \in U(2)$ admits the decomposition

$$U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta) \quad (22)$$

for real angles $\alpha, \beta, \gamma, \delta$, see also [25, 26]. Three rotations about two fixed axes therefore reach any single-qubit operation.

Several of these unitaries recur often enough to carry their own names. The three Pauli operators act as gates in their own right, written $\sigma_x = X$, $\sigma_y = Y$ and $\sigma_z = Z$ in this context. X is the quantum bit flip $|0\rangle \leftrightarrow |1\rangle$, Z leaves $|0\rangle$ fixed while sending $|1\rangle \mapsto -|1\rangle$, i.e., Z is a phase flip, and Y applies both a bit flip and a phase flip. The Hadamard gate, already met in Section 1.2, maps the basis onto the states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ and obeys $HXH = Z$ and $HZH = X$. Two more diagonal gates are important for error correction, the phase gate S and the T gate,

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}, \quad (23)$$

which up to a global phase are the rotations $R_z(\pi/2)$ and $R_z(\pi/4)$ and satisfy $S = T^2$ and $Z = S^2$. The T gate is also called the $\pi/8$ gate, after the half-angles $\pm\pi/8$ that appear on the diagonal of $R_z(\pi/4)$.

Having introduced the Pauli matrices allows us to also consider mixed quantum states. A density operator ρ of a single qubit is a Hermitian, positive operator on $\mathbb{C}^2$ with unit trace. The identity I and the three Pauli matrices $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ form a basis of the Hermitian 2×2 matrices, so ρ can always be expanded in them, leading to $\rho = \frac{1}{2}(I + \mathbf{r} \cdot \boldsymbol{\sigma})$. The vector $|\mathbf{r}| \leq 1$ occupies the interior of the Bloch sphere.

2.2 Entanglement and controlled-NOT

Single-qubit gates alone, however freely composed, cannot generate every state. Acting on a register they appear as a tensor product, and such an operator sends products to products,

$$(U_1 \otimes U_2)(|\psi\rangle \otimes |\phi\rangle) = (U_1|\psi\rangle) \otimes (U_2|\phi\rangle). \quad (24)$$

A state that cannot be written as a single tensor product is called entangled. The Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$ is the canonical example, and no product of single-qubit gates produces it from an unentangled input. This is a consequence of the counting argument of Section 1, i.e., a product state of n qubits is fixed by $\mathcal{O}(n)$ parameters, a general state by 2^n , and only a gate that couples qubits can carry the register between the two regimes.

For two subsystems (A) and (B), this distinction can be made precise by the Schmidt decomposition. Any pure state can be written as

$$|\Psi\rangle = \sum_{\mu=1}^{\chi} \lambda_{\mu} |\mu\rangle_A \otimes |\mu\rangle_B, \quad (25)$$

with non-negative Schmidt coefficients λ_{μ} satisfying $\sum_{\mu} \lambda_{\mu}^2 = 1$. A product state has Schmidt rank $\chi = 1$. Entanglement is present when more than one Schmidt coefficient is nonzero.

The same information is contained in the reduced density matrix of either subsystem. Tracing out (B) gives

$$\rho_A = \text{tr}_B |\Psi\rangle \langle\Psi| = \sum_{\mu=1}^{\chi} \lambda_{\mu}^2 |\mu\rangle_A \langle\mu|_A. \quad (26)$$

For a product state, ρ_A is again a pure-state projector. For an entangled state, ρ_A is mixed, even though the full state is pure. The corresponding entanglement entropy is

$$S_A = -\text{tr}_A \rho_A \log_2 \rho_A = -\sum_{\mu=1}^{\chi} \lambda_{\mu}^2 \log_2 \lambda_{\mu}^2. \quad (27)$$

It vanishes for product states and is positive for entangled states. Thus the entanglement entropy is a measure for entanglement between two subsystems A and B. For the Bell state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad (28)$$

the two Schmidt coefficients are both $1/\sqrt{2}$, so that

$$\rho_A = \frac{I}{2}, \quad S_A = 1. \quad (29)$$

Thus the Bell state contains one bit of entanglement between the two qubits. This is also called an *ebit*.

A gate that can create entanglement out of a product state is the controlled-NOT or, for short, CNOT gate [25]. It is defined by its action on the two qubits,

$$\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x, \quad (30)$$

which flips the target qubit when the control is $|1\rangle$, that is $|c, t\rangle \mapsto |c, t \oplus c\rangle$, where $\oplus$ is binary addition. Acting on a control that is prepared in a superposition entangles the pair of control and target qubit,

$$\text{CNOT} |+\rangle, 0\rangle = \frac{1}{\sqrt{2}}(|0, 0\rangle + |1, 1\rangle), \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \quad (31)$$

2.3 Quantum circuit diagrams

It is convenient to display a computation as a diagram. Each qubit is drawn as a horizontal wire, time runs from left to right, and a gate is a labelled box placed on the wire or wires it acts on. The controlled-NOT has its own symbol, a filled dot on the control wire joined by a vertical line to an $\oplus$ sign on the target, and a measurement in the computational basis is drawn as a meter in a box. The entangling operation of the previous subsection, which prepares a Bell state from the input $|00\rangle$ and then reads out the second qubit, is



The diagram is read from left to right: the Hadamard is applied first on the upper qubit, and the controlled-NOT follows. The unitary represented by a circuit is the product of its gate operators

in the *reverse* order, because the gate drawn first acts first and therefore stands rightmost in the product. A single qubit with H , then S , then T applied,

$$|\psi\rangle \text{---} \boxed{H} \text{---} \boxed{S} \text{---} \boxed{T} \text{---} \quad (33)$$

implements the operator $U = T S H$, in the same convention by which $T S H |\psi\rangle$ has H act first. Diagrams also state identities between circuits. The relation $\text{CNOT} = (I \otimes H) \text{CZ} (I \otimes H)$, with CZ the controlled- Z drawn as two linked dots, becomes the statement that the circuits

$$\begin{array}{c} \bullet \\ | \\ \oplus \end{array} = \begin{array}{c} \text{---} \bullet \text{---} \\ | \\ \boxed{H} \text{---} \bullet \text{---} \boxed{H} \end{array} \quad (34)$$

represent the same unitary. We use such diagrams throughout to specify simulation algorithms compactly. A single Trotter step, for instance, will appear in Section 3 as a fixed pattern of gates that is then repeated along the time axis.

Connectivity constraints make circuit identities important. On many quantum processors, such as superconducting qubits, two-qubit gates cannot be applied between arbitrary pairs of qubits, but only along the edges of a typically sparse hardware graph. To bring two qubits next to one another, one can exchange their quantum states using a SWAP gate. In terms of CNOT gates, this operation is implemented as

$$\begin{array}{c} \text{---} \times \text{---} \\ \text{---} \end{array} = \begin{array}{c} \bullet \oplus \bullet \\ | \quad | \quad | \\ \oplus \bullet \oplus \end{array} \quad (35)$$

where the three CNOTs exchange the computational basis states of the two qubits. SWAP gates are therefore often inserted by a compiler to route qubits through the hardware graph until the desired two-qubit interaction becomes local.

2.4 Universality of single-qubit gates and the CNOT

A small set of elementary gates is useful only if it is sufficiently expressive. The central structural result behind the circuit model is that arbitrary unitaries U can be built from local gates. More precisely, arbitrary single-qubit gates together with one entangling two-qubit gate, conventionally chosen as the CNOT, form a universal gate set [26, 27, 25]. This means that every unitary operation on an n -qubit register can be decomposed into a finite sequence of one-qubit gates and CNOT gates.

This statement should be understood in close analogy with universality in classical computation. In classical circuits, gates such as NAND are universal because arbitrary Boolean functions can be constructed from them. In quantum circuits, the objects to be constructed are not Boolean functions but unitary matrices. Universality therefore says that a small set of experimentally available operations is enough, in principle, to realize any arbitrary unitary matrix.

A general unitary on n qubits is a $2^n \times 2^n$ matrix, and therefore acts on the full Hilbert space at once. The core idea is to decompose this large unitary into *two-level unitaries*. Let $|x\rangle$

denote the computational basis. A two-level unitary G_{pq} acts non-trivially only on the subspace spanned by two basis states $|p\rangle$ and $|q\rangle$,

$$G_{pq}(\vartheta, \varphi, \chi) = \sum_{x \neq p, q} |x\rangle \langle x| + a |p\rangle \langle p| + b |p\rangle \langle q| + c |q\rangle \langle p| + d |q\rangle \langle q|, \quad (36)$$

with a complex Givens rotation matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} e^{i\varphi} \cos \vartheta & e^{i\chi} \sin \vartheta \\ -e^{-i\chi} \sin \vartheta & e^{-i\varphi} \cos \vartheta \end{pmatrix}. \quad (37)$$

It therefore mixes only two amplitudes of a state $|\psi\rangle = \sum_x c_x |x\rangle$,

$$\begin{pmatrix} c_p \\ c_q \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} c_p \\ c_q \end{pmatrix}, \quad c_x \mapsto c_x \quad (x \neq p, q). \quad (38)$$

By choosing the angles one can represent a selected matrix element of U . These elementary transformations can be decomposed into one-qubit gates and CNOTs. Note that this decomposition is universal, but not necessarily efficient, i.e., the decomposition of a $2^n \times 2^n$ matrix would require an exponential number of such two-level unitaries.

This *universality*, as stated above, requires a continuous set of single qubit gates. Physically, however, this continuum of operations is not what one wants to use in quantum computer. The microscopic control fields are analog and therefore cannot be set with infinite precision. Moreover, small calibration errors would then translate directly into small, continuously varying logical errors. Thus it makes sense to further reduce this set to a discrete gate set. It turns out, that only having H, S, T is sufficient to efficiently realize any arbitrary single qubit unitary to arbitrary precision [28, 29]. Together with the CNOT, this finite set is thus universal. This limitation to a discrete set of gates is required for the application of quantum error correction.

3 Simulating strongly correlated systems on quantum computers

We have so far introduced quantum circuits as a model of computation and quantum simulation as an abstract task. We now apply this knowledge as a computational tool to tackle strongly correlated quantum systems.

The examples in this section are chosen with this purpose in mind. We start with real-time evolution of spin-1/2 models, because local Pauli operators are already native objects of the qubit register. Fermionic lattice models require an additional encoding step, and we will use them to discuss how fermionic creation and annihilation operators can be mapped to qubit operators. Finally we tackle the ground-state problem. This brings us to variational algorithms and their limitations.

Throughout this section we use $S_i^\alpha = \sigma_i^\alpha/2$ for spin-1/2 operators. The elementary gate set is assumed to contain arbitrary single-qubit rotations and the controlled-NOT. This is a convenient compilation target rather than a hardware statement. On a specific device the native entangling gate may be different, but it can be converted to the same language by local basis changes.

3.1 Non-equilibrium dynamics of spin systems

We consider a simple spin Hamiltonian in a time-dependent magnetic field [30]. We aim to describe the time evolution of an initial state through quantum circuits. The time dependent Hamiltonian is given by,

$$H(t) = \sum_{\langle i,j \rangle} J_{\perp} (S_i^x S_j^x + S_i^y S_j^y) + J_z S_i^z S_j^z + \sum_i \mathbf{H}_i(t) \cdot \mathbf{S}_i. \quad (39)$$

The notation $\langle i, j \rangle$ denotes couplings on an arbitrary but sparse interaction graph, e.g., a two-dimensional square lattice. The first two terms describe an XXZ exchange interaction. The last term is a local magnetic field, which can depend on time and on the site. In the Pauli language this is

$$H(t) = \sum_{\langle i,j \rangle} \frac{J_{\perp}}{4} (X_i X_j + Y_i Y_j) + \frac{J_z}{4} Z_i Z_j + \frac{1}{2} \sum_i (h_i^x(t) X_i + h_i^y(t) Y_i + h_i^z(t) Z_i). \quad (40)$$

Note that each term in the sum is a Pauli string, i.e., a product of Pauli operators. The time evolution task is to approximate

$$U(t, 0) = \mathcal{T} \exp \left[-i \int_0^t H(s) ds \right] \quad (41)$$

for an initial state that may be a product state, a prepared ground state or some other state that was prepared beforehand. We divide the interval into r pieces, $\delta t = t/r$, and approximate the Hamiltonian on the m th interval by its midpoint value $H_m = H(t_m)$, $t_m = (m+1/2)\delta t$. A first-order product formula gives

$$U(t, 0) = \prod_{m=r-1}^0 \left[\prod_i U_i^{\text{field}}(t_m, \delta t) \prod_{\langle i,j \rangle} U_{ij}^{XX}(\delta t) U_{ij}^{YY}(\delta t) U_{ij}^{ZZ}(\delta t) \right] + \mathcal{O}(\delta t). \quad (42)$$

The ordering inside one step is a convention. A symmetric second-order step replaces $\prod_l e^{-iH_l \delta t}$ by half steps around the last factor and reduces the local error from $\mathcal{O}(\delta t^2)$ to $\mathcal{O}(\delta t^3)$, at the cost of a longer circuit.

The local field terms are single-qubit gates. For example

$$\exp \left[-i \frac{h_i^z(t_m)}{2} Z_i \delta t \right] = R_z(h_i^z(t_m) \delta t), \quad (43)$$

and similarly for x and y . For a general field one may either implement a single rotation about the axis $\mathbf{h}_i/h$, with $h = |\mathbf{h}_i|$, as,

$$U_i^{\text{field}}(t_m, \delta t) = \exp \left[-i \frac{\delta t}{2} \mathbf{h}_i(t_m) \cdot \boldsymbol{\sigma}_i \right] = R_{\hat{\mathbf{h}}_i}(|\mathbf{h}_i(t_m)| \delta t). \quad (44)$$

As long as a universal gate set is available this rotation can always be implemented. Let us take a concrete example: A quantum computer only allows R_z and R_x rotations. Then a generic

We decompose the two-qubit terms into single qubit gates and CNOT. A useful primitive is the exponential of a Pauli string. For the ZZ interaction,

The circuit is given by

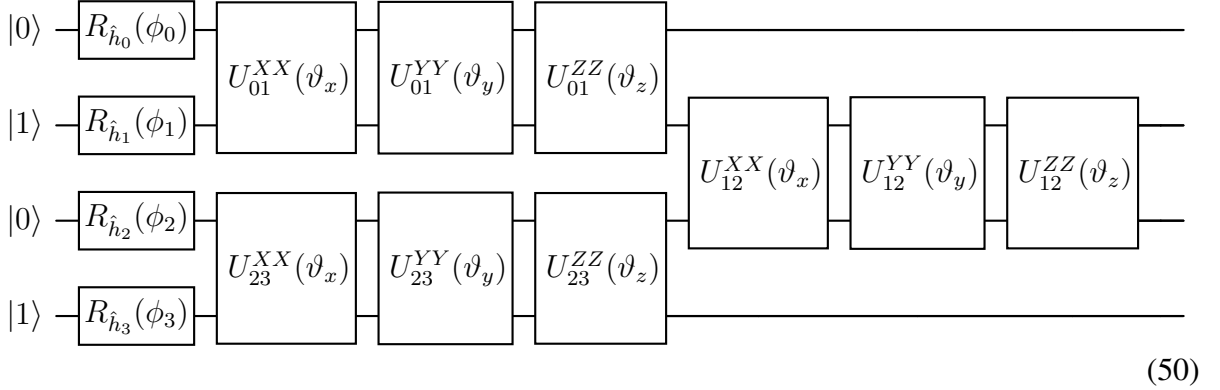
The XX and YY interactions are the same construction in a different basis. Since $HXH = Z$, the XX exponential is obtained by rotating both qubits from the z basis to the x basis, applying the ZZ construction, and rotating back

with $\vartheta_x = J_\perp \delta t / 2$. For the YY term one may use $S^\dagger H Y H S = Z$, or equivalently rotate the y axis to the z axis before applying the same entangling phase

with $\vartheta_y = J_\perp \delta t / 2$. For one pair (i, j) the three Pauli products $X_i X_j$, $Y_i Y_j$ and $Z_i Z_j$ commute. Their product therefore gives the exact exponential of the pair Hamiltonian,

The Trotter error in Eq. (42) is not caused by these three factors on the same bond. It comes from non-commuting terms that share a *single* site, and from the time dependence of $\mathbf{H}_i(t)$.

A complete first-order Trotter step for a four site chain with open boundary conditions therefore has the structure shown below.



Here $\phi_i = |\mathbf{h}_i(t_m)| \delta t$. The first entangling layer acts on bonds $(0, 1)$ and $(2, 3)$, and the second layer acts on bond $(1, 2)$. Longer chains repeat the same odd-even pattern. We can estimate the required number of gates from this construction. Note that typically the entangling gates, e.g., the CNOTs, are the most noisy gates on current quantum computing platforms. It thus makes sense to count the entangling gates to estimate the reliability of a quantum circuit. A single Pauli two-qubit exponential costs two CNOTs. An XXZ bond therefore costs six CNOTs if implemented by the three separate circuits. For a chain with N sites and open boundary conditions, one first-order step costs $6(N-1)$ CNOTs for the exchange layers and N one-qubit field rotations. For r steps the number of gates thus scales as $\mathcal{O}(rN)$ for a one-dimensional local Hamiltonian. As a result the circuit depth grows with the spacetime volume of the local problem, not with the Hilbert-space dimension.

In Fig. 1 we see an example of this time evolution for a spin plaquette driven by a short magnetic field pulse, see inset. The experiment was realized on an IBM superconducting quantum processor [30]. Shown is the staggered magnetization of the system. Since only antiferromagnetic ZZ interactions are present in this model, the system starts at $t = 0$ from a ground state configuration. Here six time steps are sufficient to realize an accurate time evolution. The deviation of the experimental results from the Trotterized time evolution is due to noisy gates and noisy readout.

3.2 Simulating fermions

Since spin systems are already qubit systems no additional mapping is required. For fermions the situation is different and we require a mapping to the Pauli algebra. Fermionic creation and annihilation operators obey

$$\{c_p, c_q^\dagger\} = \delta_{pq}, \quad \{c_p, c_q\} = \{c_p^\dagger, c_q^\dagger\} = 0. \quad (51)$$

In contrast Pauli operators on different qubits commute. A fermion-to-qubit mapping must encode these signs. Historically this problem was solved through the Jordan-Wigner transfor-

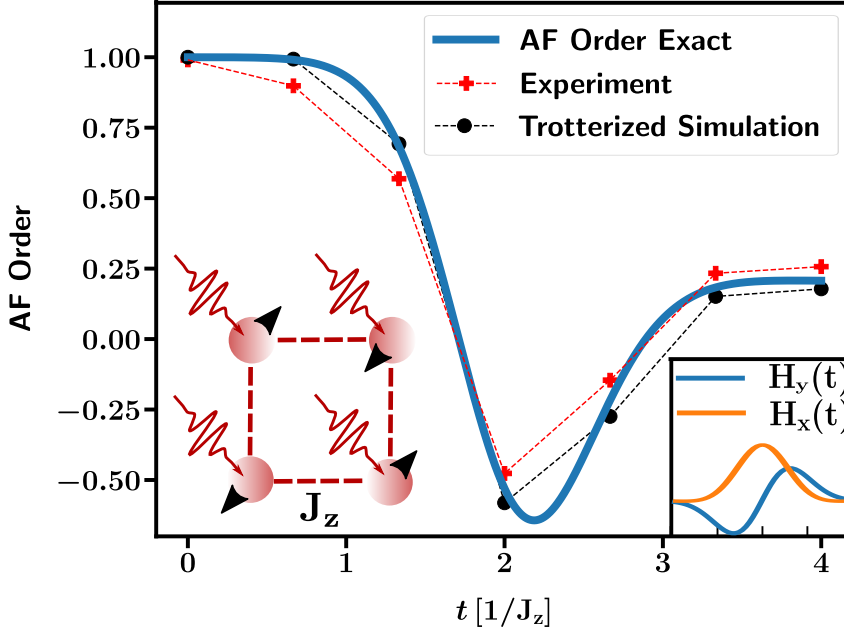


Fig. 1: Nonequilibrium quantum simulation of a spin plaquette subject to a magnetic field pulse. Only the non-vanishing spin-spin coupling is shown. The circular polarized pulse is parameterized as $H_x(t) = h_0 \exp(-(t-t_0)/(2\tau^2)) \cos(\omega(t-t_0))$ and $H_y(t) = h_0 \exp(-(t-t_0)/(2\tau^2)) \sin(\omega(t-t_0))$ with parameters $h_0=2$, $\omega=1$, $\tau=0.7$, $t_0=2$, see inset. Displayed is the staggered magnetization $S_{\text{AF}} = \sum_{i=1}^4 (-1)^i S_z^i$. Comparison between the experimental results, the ideal Trotterized time evolution and the exact time evolution. Adopted from [30].

mation, which is also the most pedagogical introduction to the topic. We define

$$c_p^\dagger = \left(\prod_{q < p} Z_q \right) \sigma_p^-, \quad c_p = \left(\prod_{q < p} Z_q \right) \sigma_p^+, \quad (52)$$

where $\sigma^+ = (X + iY)/2$ and $\sigma^- = (X - iY)/2$ in the convention $|0\rangle$ empty, $|1\rangle$ occupied. The string of Z operators implements the parity of all modes to the left of p . This string produces the minus sign when two fermionic operators are interchanged.

For example, the number operator is local,

$$n_p = c_p^\dagger c_p = (1 - Z_p)/2, \quad (53)$$

whereas a hopping term between $p < q$ becomes

$$c_p^\dagger c_q + c_q^\dagger c_p = \frac{1}{2} (X_p Z_{p+1} \cdots Z_{q-1} X_q + Y_p Z_{p+1} \cdots Z_{q-1} Y_q). \quad (54)$$

A density-density interaction is again local in the occupation basis,

$$n_p n_q = \frac{1}{4} (1 - Z_p - Z_q + Z_p Z_q). \quad (55)$$

Thus the cost of Jordan-Wigner depends on the ordering of the fermionic operators. From this it follows straightforwardly that one-dimensional fermionic systems with short range interactions

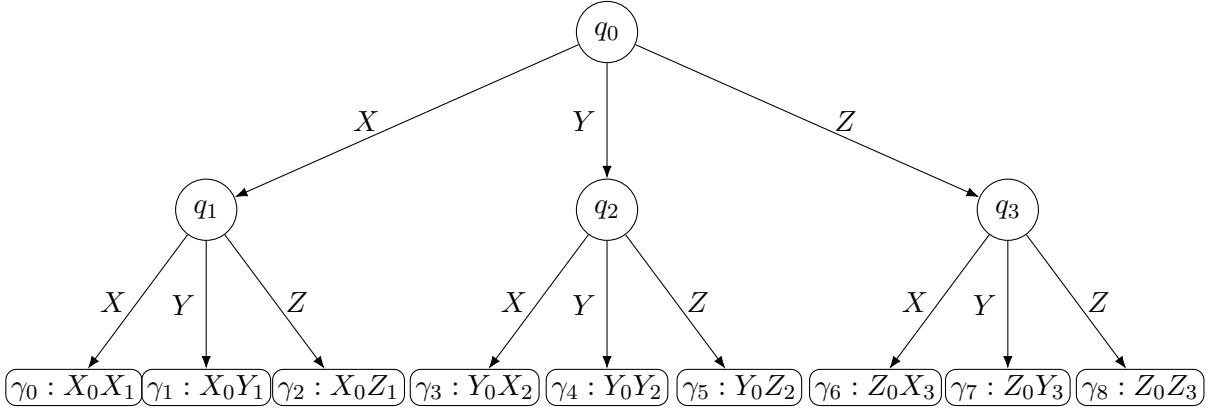


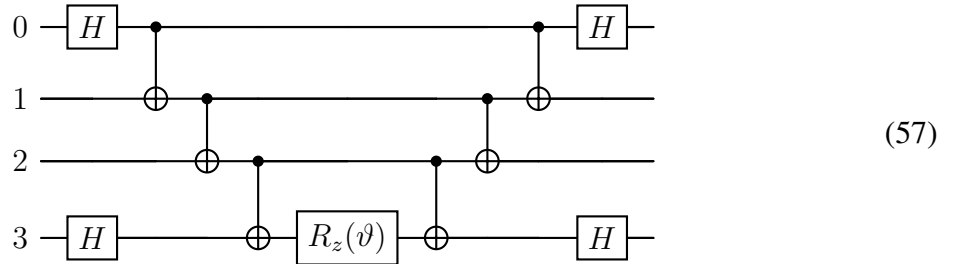
Fig. 2: Schematic ternary-tree assignment of Majorana operators to Pauli strings [31]. The labels along a root-to-leaf path specify the Pauli operators in the string. In a full mapping one selects $2M$ of these leaves for M fermionic modes and pairs them into creation and annihilation operators as in Eq. (58).

can be implemented with short string operators, but two and three-dimensional lattices or general quantum-chemistry Hamiltonians will contain long parity strings. This affects the number of entangling gates in the circuit.

For a general Pauli string the circuits can be implemented in the following fashion. Suppose

$$P = X_p Z_{p+1} \cdots Z_{q-1} X_q, \quad U_P(\vartheta) = \exp(-i\vartheta P/2). \quad (56)$$

First rotate every non- Z factor to Z . Then compute the parity onto one qubit with a ladder of CNOTs, apply $R_z(\vartheta)$, and uncompute. For the string $X_0 Z_1 Z_2 X_3$ this gives



For a length- ℓ string this costs $2(\ell-1)$ CNOTs and is hence undesirable.

To overcome this issue in higher dimensions we use tree-based decompositions to improve the mapping by distributing parity information over a tree rather than putting it into one long line. A particularly symmetric construction is the ternary-tree mapping of Ref. [31], see Fig. 2. Each edge of a ternary tree is labelled by one of the Pauli matrices X, Y, Z . A leaf ℓ is associated with the Pauli string obtained by multiplying the edge labels along the path from the root to that leaf. With a suitable assignment of leaves, these strings form a set of pairwise anticommuting Majorana operators.

The Majorana operators can then be easily translated to fermionic operators via

$$\gamma_{2p} = c_p + c_p^\dagger, \quad \gamma_{2p+1} = -i(c_p - c_p^\dagger), \quad \{\gamma_a, \gamma_b\} = 2\delta_{ab}. \quad (58)$$

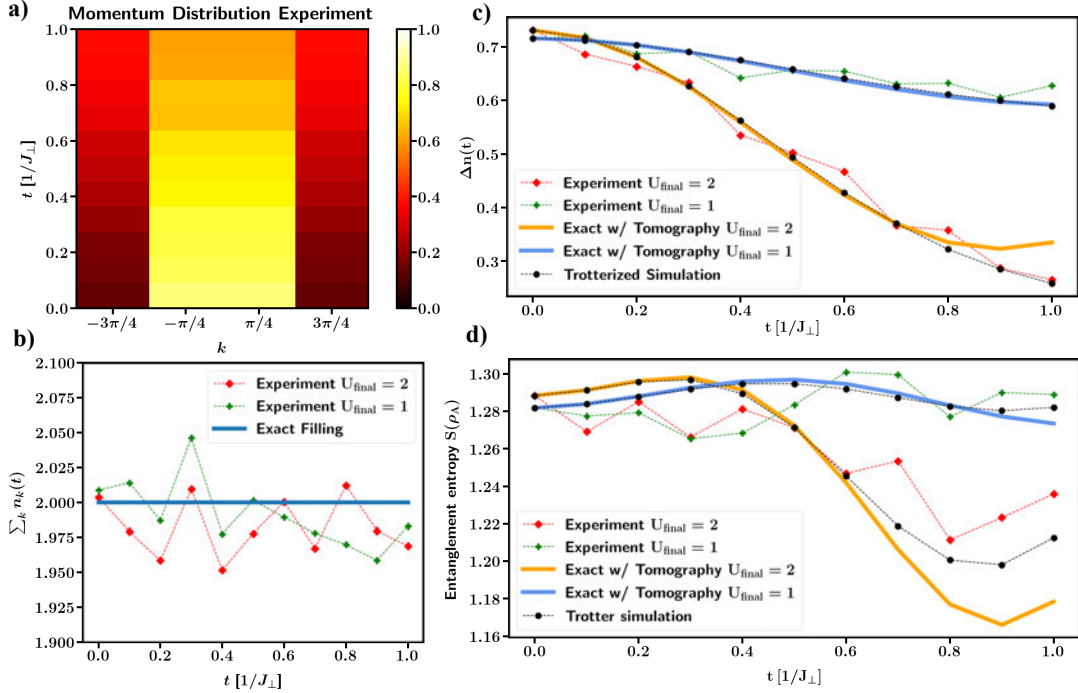


Fig. 3: *Quantum simulation of an interaction quench in a fermion system. a) Time and momentum dependence of the fermionic distribution function on a periodic four site chain after an interaction quench to $U_{\text{final}} = 2$ at $t=0$. b) Time evolution of the filling factor. Comparison between the exact filling, the weak quench $U_{\text{final}} = 1$ and the strong quench $U_{\text{final}} = 2$. c) Time evolution of the jump in the Fermi distribution. Comparison between the experimental results for weak and strong quenches and classical simulations. d) Entanglement entropy of a bipartition of the system. Adopted from [30].*

A valid mapping sends the $2M$ Majoranas of M fermionic modes to $2M$ qubit Pauli strings that square to identity and anticommute pairwise. In the Jordan-Wigner transformation the anticommutation of all Fermions is enforced by a linear parity string. In the tree construction it is enforced by the branching structure: two paths that first separate at a node carry different Pauli labels on that qubit, and hence the corresponding strings anticommute on only one site. A complete ternary tree of height h has 3^h leaves, so it can host order 3^h Majoranas with Pauli weight h . The Pauli weight, i.e. the number of Pauli operators, is therefore logarithmic in the number of fermionic modes and so is the necessary number of CNOTs to implement such an operator.

As an example of fermionic dynamics we investigate a four-site chain of interacting spinless fermions, equivalent to an XXZ chain,

$$H = \sum_i J_{\perp} (S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + U(t) S_i^z S_{i+1}^z. \quad (59)$$

Using the Jordan-Wigner transformation this is equal to the nearest-neighbor model

$$H = \frac{J_{\perp}}{2} \sum_i (c_i^{\dagger} c_{i+1} + c_{i+1}^{\dagger} c_i) + U(t) \sum_i \left(n_i - \frac{1}{2} \right) \left(n_{i+1} - \frac{1}{2} \right), \quad (60)$$

up to a constant. We want to investigate how the non-interacting ground state for $U = 0$ evolves if the interaction is suddenly switched on at $t = 0$, i.e., $U(t) = U_{\text{final}}\Theta(t)$. The subsequent unitary dynamics allows us to investigate how many-body systems relax after excitation. On a quantum computer the initial Fermi sea is not a computational basis state as it is already entangled in the site basis. We here skip over this encoding problem of the initial state and focus on the time evolution. We can apply the same $XX + YY$ circuits used in the spin section implement the kinetic term. The interaction term is diagonal and uses the ZZ circuit. In Fig. 3 we see the results from an IBM Quantum Computer for the momentum distribution $n_k(t)$ and the entanglement entropy for a 4-site system. Notably the quantum computation shows a significant amount of noise, which is due to the added encoding circuit not shown here.

3.3 Variational quantum eigensolver

So far we focused on the time evolution problem. In this section we tackle the ground state problem. As already laid out in the introduction, the ground state problem is QMA-complete and we do not expect a general quantum speedup in contrast to the time evolution problem. However such exact statements may affect practical approaches only marginally. Here we will show that the variational principle can also be applied for states prepared on quantum computers. We want to find a state $|\psi\rangle$ with low energy

$$E[\psi] = \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}. \quad (61)$$

The variational principle gives as an upper bound $E[\psi] \geq E_0$ for every normalized $|\psi\rangle$, with equality only in the ground-state subspace. This property is used in an algorithm called the *Variational Quantum Eigensolver* (VQE). The VQE uses a parameterized circuit $U(\boldsymbol{\vartheta})$ to prepare a quantum state $|\psi(\boldsymbol{\vartheta})\rangle$ and determine a cost function $C(\boldsymbol{\vartheta})$ that is the energy expectation value of the state,

$$|\psi(\boldsymbol{\vartheta})\rangle = U(\boldsymbol{\vartheta})|0\rangle^{\otimes n}, \quad C(\boldsymbol{\vartheta}) = \langle\psi(\boldsymbol{\vartheta})|H|\psi(\boldsymbol{\vartheta})\rangle. \quad (62)$$

The idea is now to minimize $C(\boldsymbol{\vartheta})$ using the parameters $\boldsymbol{\vartheta}$ in the unitary $U(\boldsymbol{\vartheta})$. One might ask, how to realize such a complicated subroutine like minimization on a quantum computer. The answer is: one does not. Instead we rather use the quantum computer to sample the energy expectation value and then give that task to a classical computer, which executes a minimization algorithm, such as steepest descent. More precisely, the quantum processor estimates $C(\boldsymbol{\vartheta})$ by measuring Pauli strings in a decomposition $H = \sum_a h_a P_a$. The classical optimizer then proposes new parameters which are again used by the quantum processor to get a new estimate of the Pauli strings. This hybrid quantum-classical loop is closed until the cost no longer decreases, see Fig. 4.

The main open question we have not yet addressed is the choice of the Ansatz, i.e., the parameterized circuit $U(\boldsymbol{\vartheta})$. Various strategies have been proposed to design different ansatz schemes [32]. So called *Hardware-efficient ansätze* [33] are tailored to the native gate sets and

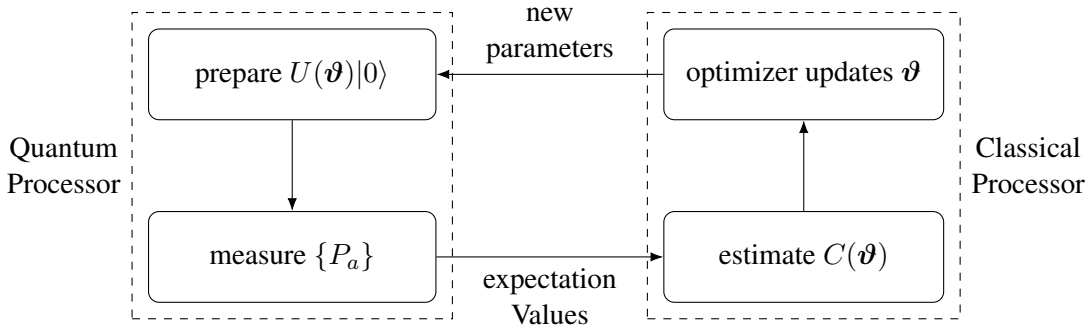


Fig. 4: Hybrid structure of the Variational Quantum Eigensolver. The quantum device prepares and measures states. The classical computer constructs the cost function and chooses the next parameters.

connectivity of quantum devices, thereby reducing circuit depth, but they often suffer from limited expressivity. In contrast, problem-inspired approaches such as the *Variational Hamiltonian Ansatz* [34] aim to mimic adiabatic state preparation, potentially improving convergence, although at the cost of reduced flexibility or increased circuit depth. More recently, approaches such as the *Quantum-optimal-control-inspired ansatz* [35] have explored symmetry breaking and control-theoretic concepts to enhance expressivity and optimization performance. Despite these advances, identifying ansätze that simultaneously balance expressivity, trainability, and noise resilience remains an open problem [36, 37].

We want to explore what is currently possible to achieve with a useful example, the one-dimensional axial next nearest neighbor Ising model, or ANNNI model [38]. The Hamiltonian is given by

$$H(J_2, B_x) = -J_1 \sum_i Z_i Z_{i+1} + J_2 \sum_i Z_i Z_{i+2} + B_x \sum_i X_i. \quad (63)$$

Here $J_1 > 0$ favors ferromagnetic order, while $J_2 > 0$ favors antiferromagnetic alignment at distance two. The transverse field B_x introduces quantum fluctuations. We set $J_1 = 1$ as the unit of energy and use the ratio J_2/J_1 as the control parameter. For $B_x = 0$ the Hamiltonian is classical. If $J_2/J_1 < 1/2$, the ferromagnetic term dominates and the two classical states $|00 \cdots 0\rangle$ and $|11 \cdots 1\rangle$ span the ground state manifold. If $J_2/J_1 > 1/2$, the next nearest neighbor term enforces a period four pattern such as

$$|\psi_{22}\rangle \simeq |00110011 \cdots\rangle, \quad (64)$$

up to translations and global spin flip. The point $J_2/J_1 = 1/2$ is highly degenerate at zero field. A small transverse field lifts this degeneracy and reveals a richer finite size phase diagram. Besides the ferromagnetic and period four regions one finds modulated states and a paramagnetic region at sufficiently large field. Strictly speaking these are true phases only in the thermodynamic limit. On a finite quantum processor they appear as phaselike regions, i.e., regions in which the optimized ground state has a characteristic structure and changes rapidly at special parameter values. This makes the model a good benchmark. The Hamiltonian is simple in terms of Pauli operators, but its ground states are sensitive to frustration and to quantum fluctuations.

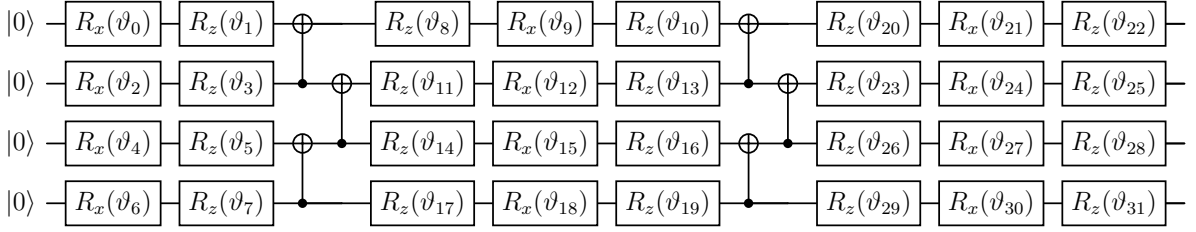


Fig. 5: One layer of the VQE ansatz for the ANNNI model for four qubits. Each single qubit rotation carries an independent variational parameter. The layer starts with rotations around the x and z axes, followed by a staggered pattern of CNOT gates. The same entangling pattern is then repeated after a second block of single qubit rotations. More layers repeat this structure with new variational parameters.

For this model we use a hardware efficient circuit assuming a one-dimensional device connectivity. One layer for four qubits is shown in Fig. 5. It starts from local rotations, then applies controlled not gates. Local rotations between and after the entangling gates provide the variational parameters. The same pattern can be tiled to larger systems. For L layers the circuit is obtained by repeating this block L times with independent angles. Increasing L enlarges the set of states that can be represented. In particular, correlations can spread over longer distances and the circuit can better interpolate between the ferromagnetic state, the period four state, and states in between. At the same time, every extra layer adds two qubit gates, and these gates usually dominate the noise. The useful depth is therefore not set only by expressivity but by a compromise between the variational error of a shallow circuit and the hardware error of a deep circuit.

The energy alone is not always the most reliable observable for detecting changes in the wave function on noisy hardware. A more sensitive quantity is the fidelity susceptibility. Let $p = J_2/J_1$ and let $|\psi(p)\rangle$ denote the ground state or the optimized VQE state. The fidelity between neighboring points is

$$F(p, \delta) = |\langle \psi(p-\delta) | \psi(p+\delta) \rangle|. \quad (65)$$

For small δ one defines the fidelity susceptibility through

$$F(p, \delta) = 1 - \chi_F(p) \delta^2 + \mathcal{O}(\delta^4). \quad (66)$$

It measures how quickly the ground state changes when the Hamiltonian parameter is varied. Away from a transition the state changes smoothly and χ_F is small. Near a level crossing or an avoided crossing the state changes strongly and χ_F develops a peak. We can determine the fidelity susceptibility also from perturbation theory. If $H(p) = H_0 + p H_A$ and $|n(p)\rangle$ are exact eigenstates, then

$$\chi_F(p) = 2 \sum_{n \neq 0} \frac{|\langle n(p) | H_A | 0(p) \rangle|^2}{(E_n(p) - E_0(p))^2}. \quad (67)$$

The spectral gap therefore enters in the denominator. This is why the fidelity susceptibility can capture a gap closure. It is also largely agnostic to the order parameter, which is useful for frustrated models where the relevant order may be modulated or hard to guess.

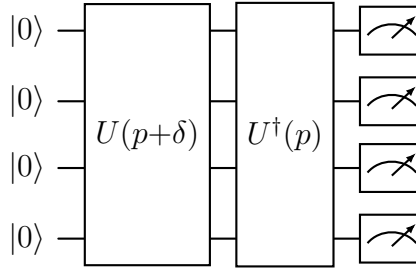


Fig. 6: *Overlap circuit for the fidelity susceptibility. The probability of measuring 0000 gives the squared overlap between two variational states. Repeating the circuit for nearby values of $p = J_2/J_1$ yields the finite difference estimate of χ_F .*

On a quantum computer the overlap in Eq. (65) can be measured with a simple inverse circuit. Suppose that $U(p)$ prepares the optimized state at p and that $U(p+\delta)$ prepares the optimized state at the nearby point. The circuit in Fig. 6 applies $U(p+\delta)$ and then the inverse of $U(p)$. The probability to measure the all zero bit string is

$$P_0 = \left| \langle 0 |^{\otimes N} U^\dagger(p) U(p+\delta) | 0 \rangle^{\otimes N} \right|^2 = \left| \langle \psi(p) | \psi(p+\delta) \rangle \right|^2. \quad (68)$$

Thus the fidelity is obtained as $F = \sqrt{P_0}$. In a VQE calculation there is one further subtlety. The optimizer can choose symmetry broken representatives of the same degenerate subspace at neighboring values of p . Then the direct overlap may be small even though both states describe the same phase. One can partially correct this by maximizing the overlap over symmetry operations of the Hamiltonian, for example the global bit flip $X^{\otimes N}$ and, for periodic boundary conditions, lattice translations. This gives a fidelity measure that is less sensitive to arbitrary choices made by the variational optimizer.

In Fig. 7 we see the results of the VQE approach for the ANNNI model. On the top the correlation functions demonstrate the transitions of the ground state due to spatial fluctuations, e.g. the change of the oscillation frequency in the intermediate J_2/J_1 ratio. In the bottom panel the ground state energy is compared to an ideal VQE simulation and to the VQE obtained from the quantum computer. The large noise in the energy prohibits any statements on phase transitions in the system. In contrast the fidelity susceptibility clearly identifies strong changes in the ground state wave function. Thus the VQE should not be judged only by the absolute accuracy of the energy. The same measured states can contain robust information about changes in the wave function, and this information can be easier to extract on noisy devices.

4 Quantum error correction

4.1 Introduction to error correction

Physical qubits are never perfectly isolated from their environment. Typical errors are qubit relaxation, dephasing and coherent errors of the gates, e.g., overrotations. If such errors accumulate with circuit depth, then the simulated time evolution can become dominated by errors

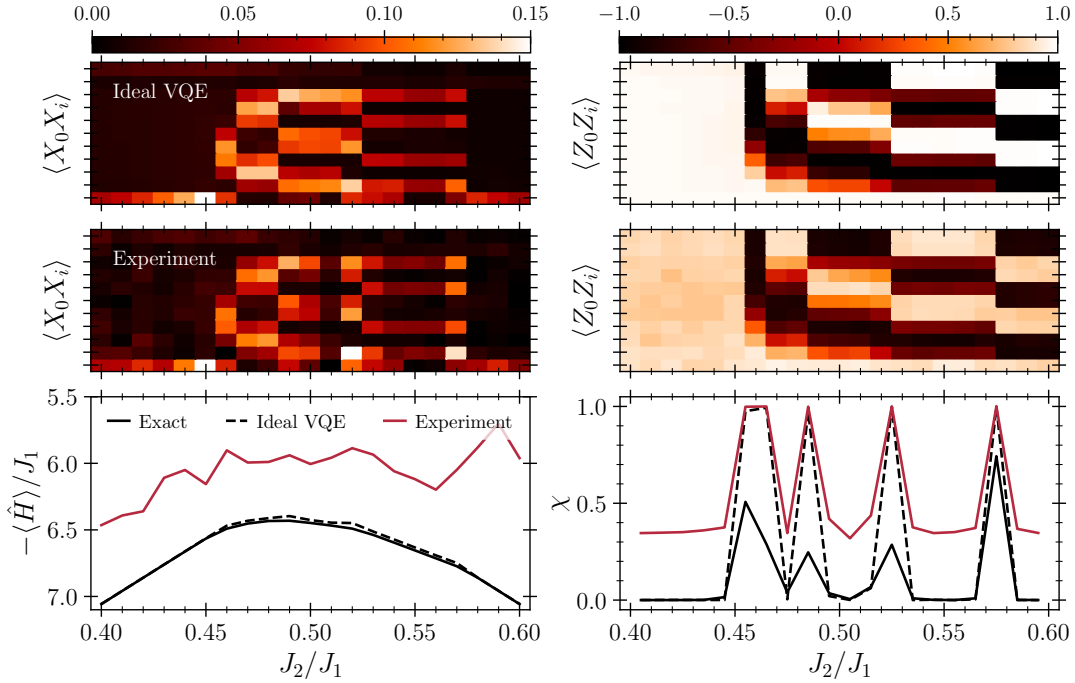


Fig. 7: Variational Quantum Eigensolver solution of the ANNNI model. The correlation functions $\langle X_0 X_i \rangle$ and $\langle Z_0 Z_i \rangle$ alongside energy and fidelity susceptibility χ for $N = 12$ at $B_x = 0.1$ with open boundary conditions. The top row shows the ideal noise-free VQE correlation functions, while the middle row shows the experimental results on *ibm.kyoto* quantum processor, starting from site $i = 1$ on the y -axis. The J_2/J_1 axes are shared with the energy and fidelity susceptibility plotted on the bottom rows.

rather than by the Hamiltonian we wanted to study. Quantum error correction tries to overcome this problem. The core idea is that quantum information can be stored nonlocally in a larger Hilbert space and we can use targeted measurements, that do not perturb this information, to diagnose errors without measuring the encoded state itself.

We begin with a classical example. Suppose that we want to protect one classical bit against a bit flip. The simplest code stores the logical bit redundantly in three physical bits,

$$0 \mapsto 000, \quad 1 \mapsto 111. \quad (69)$$

If at most one bit flips, the corrupted strings are still closer to the intended codeword than to the other one. For example, the string 010 is interpreted as 000, while 101 is interpreted as 111. This is the majority vote principle. The redundancy has not removed errors, but it has converted a small number of errors into a detectable pattern.

The states 000 and 111 are then what define the *code space*.

The same approach can be expressed in terms of parity checks. For three bits b_1, b_2, b_3 one measures whether neighboring bits agree. The two parity checks are

$$s_1 = b_1 \oplus b_2, \quad s_2 = b_2 \oplus b_3. \quad (70)$$

For the codewords 000 and 111 both parities vanish. If the first bit flips, one obtains $s_1 = 1$ and $s_2 = 0$. If the second bit flips, both parities are one. If the third bit flips, one obtains $s_1 = 0$ and

$s_2 = 1$. Thus the two parity checks reveal the position of the error. They do not reveal whether the stored bit was zero or one. This distinction is essential. The measurement extracts only the error *syndrome*, not the logical information.

The quantum version of the same idea encodes one logical qubit into three physical qubits as

$$|0_L\rangle = |000\rangle, \quad |1_L\rangle = |111\rangle, \quad (71)$$

so that an arbitrary logical state is

$$|\psi_L\rangle = \alpha |000\rangle + \beta |111\rangle. \quad (72)$$

Note that this is different from copying a quantum state, e.g., $|\phi\rangle \rightarrow |\phi\rangle |\phi\rangle$, which the no cloning theorem forbids. The state is encoded into a two dimensional subspace of the eight dimensional Hilbert space of three qubits. A bit flip on the first, second, or third physical qubit gives

$$X_1 |\psi_L\rangle = \alpha |100\rangle + \beta |011\rangle, \quad (73)$$

$$X_2 |\psi_L\rangle = \alpha |010\rangle + \beta |101\rangle, \quad (74)$$

$$X_3 |\psi_L\rangle = \alpha |001\rangle + \beta |110\rangle. \quad (75)$$

These three error spaces are orthogonal to the code space and to each other. We can therefore identify which error occurred without collapsing the phases α and β .

The quantum parity checks are measurements of the two commuting operators

$$S_1 = Z_1 Z_2, \quad S_2 = Z_2 Z_3. \quad (76)$$

Both logical codewords are eigenstates of S_1 and S_2 with eigenvalue $+1$. This remains true for every superposition in the code space. If a physical bit flip X_i occurs, it anticommutes with the parity checks that contain qubit i . The measured eigenvalues of S_1 and S_2 therefore provide the syndrome. The outcomes $(+1, +1)$ mean no bit flip, $(-1, +1)$ means a flip on qubit one, $(-1, -1)$ means a flip on qubit two, and $(+1, -1)$ means a flip on qubit three. After this diagnosis one applies X_i to the identified qubit. The correction maps the state back to $|\psi_L\rangle$.

The operators S_1 and S_2 are called *stabilizers* because they leave the code space invariant. The code space is the common $+1$ eigenspace of all stabilizers. Error correction is therefore a repeated sequence of syndrome measurements and conditional operations. The syndrome measurements are allowed because they commute with the logical information. They ask whether the state is still inside the code space, and if not, in which orthogonal error space it lies.

4.2 Bit and phase flip errors

A qubit has more possible errors than a classical bit. The Pauli matrix X exchanges $|0\rangle$ and $|1\rangle$ and is therefore the quantum analogue of a bit flip. The Pauli matrix Z leaves $|0\rangle$ unchanged and changes the sign of $|1\rangle$,

$$Z |0\rangle = |0\rangle, \quad Z |1\rangle = -|1\rangle. \quad (77)$$

This is a relative phase flip. It has no classical analogue for a single computational basis state, but it changes coherent superpositions and is therefore just as damaging for a quantum computation. The Pauli matrix Y is, up to a global phase, the product of both errors, i.e., $Y = iZX$. For this reason it is enough to understand how to detect and correct X and Z errors.

The three qubit code above corrects one bit flip. It does not correct one phase flip, since

$$Z_1 |\psi_L\rangle = \alpha |000\rangle - \beta |111\rangle. \quad (78)$$

This state is still in the common $+1$ eigenspace of Z_1Z_2 and Z_2Z_3 . The parity checks therefore cannot see the error. To correct phase flips one can use the same code after a change of basis from the Z to the X basis. This basis transformation is realized through the Hadamard gate,

$$HZH = X. \quad (79)$$

Thus a phase flip in the computational basis is a bit flip in the X basis. A three qubit phase flip code is obtained from the bit flip code by applying Hadamard gates to all qubits. With

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}, \quad (80)$$

one may encode

$$|0_L\rangle = |+++\rangle, \quad |1_L\rangle = |--\rangle. \quad (81)$$

The stabilizers are now

$$S_1 = X_1X_2, \quad S_2 = X_2X_3. \quad (82)$$

Their measurement detects a single Z error in exactly the same way that Eq. (76) detects a single X error.

A full quantum error correcting code combines both ideas. The smallest historical example is the nine qubit code developed by Shor [39]. Modern stabilizer codes use the same principles but try to minimize the use large stabilizer measurements and large number of qubits.

There is one important conceptual point that allows quantum error correction to be also applied to continuous errors. For example, a small unwanted rotation around the x axis on the first qubit has the form

$$R_x(\epsilon) = \exp\left(-i\frac{\epsilon}{2}X_1\right) = \cos\left(\frac{\epsilon}{2}\right)I - i\sin\left(\frac{\epsilon}{2}\right)X_1. \quad (83)$$

Acting on an encoded state this gives a coherent superposition of no error and a full bit flip,

$$R_x(\epsilon) |\psi_L\rangle = \cos\left(\frac{\epsilon}{2}\right) |\psi_L\rangle - i\sin\left(\frac{\epsilon}{2}\right) X_1 |\psi_L\rangle. \quad (84)$$

Before the syndrome measurement this is not a classical random error. It is a quantum superposition of two orthogonal syndrome sectors. The parity measurement projects the state into one of them. With probability $\cos^2(\epsilon/2)$ the syndrome is trivial and the state is projected back to the code space, thus no error occurred. With probability $\sin^2(\epsilon/2)$ the syndrome is that of a full X_1 error. In that branch we apply X_1 and again recover $|\psi_L\rangle$. The measurement has discretized the continuous error into a finite set of correctable alternatives.

The same reasoning applies to general errors, since any operator can be expanded in the Pauli string basis. Thus stabilizer error correction focuses on a discrete set of Pauli errors even though the microscopic dynamics of noise is continuous.

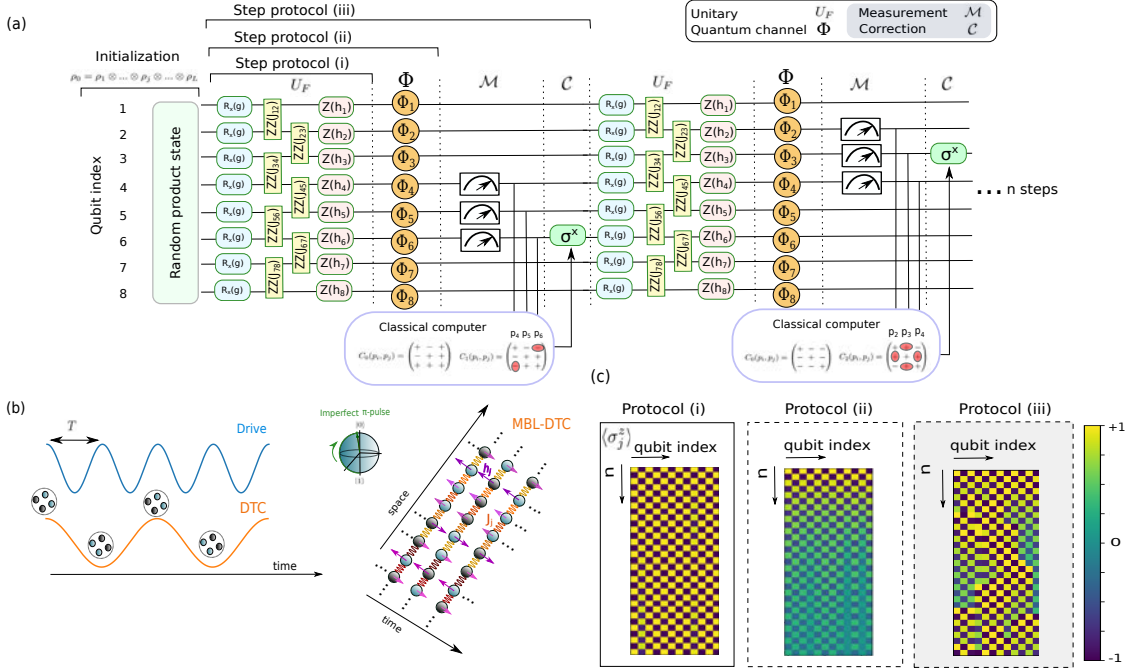


Fig. 8: Quantum-classical feedback scheme for the stabilization of a discrete time crystal on a quantum computer. *a)* The time evolution of the system is subject to a noise channel Φ_i . Afterwards a parity measurements determines possibly flipped spins and corrects them. *b)* Shows the mechanism behind the many-body localized time crystal phase. *c)* Demonstrates the effect of exact unitarity, noise and noise plus correction scheme. Adopted from [41].

4.3 Partial error correction for non-equilibrium phases of matter

Quantum error correction aims to preserve an arbitrary logical state for an arbitrarily long circuit. This is a very demanding goal. For many quantum simulation tasks this is not necessary, as we do not need to protect all quantum information equally well. Here we discuss an example that protects only the part of the dynamics that defines a phase of matter. One measures and corrects selected stabilizers or selected observables that are tied to the order of interest, while accepting that other information is still lost.

We focus on a many body localized discrete time crystal (MBL-DTC). Such a system is characterized by a robust subharmonic response to a periodic drive. That means if the Hamiltonian has period T , local observables oscillate with period nT , $n > 1$, in our case $n = 2$. In an ideal closed system this period doubling can persist indefinitely in the thermodynamic limit because disorder and interactions prevent the system from heating to a featureless state. On a noisy quantum processor the situation is different since the oscillations decay due to decoherence after a limited number of Floquet cycles [40].

The most studied model is a one dimensional periodically kicked Ising chain with Floquet unitary

$$U_F = \exp \left(-i \frac{T}{4} \sum_{j=1}^{L-1} J_j Z_j Z_{j+1} \right) \exp \left(-i \frac{T}{2} \sum_{j=1}^L h_j Z_j \right) \exp \left(-i \frac{\pi g}{2} \sum_{j=1}^L X_j \right). \quad (85)$$

The couplings J_j and fields h_j are disordered. The last factor is close to a global spin flip when g is close to one. Therefore a product state in the Z basis is approximately mapped to its spin reversed state after one period, and back again after two periods. The interactions and disorder stabilize this response against small imperfections in the pulse, i.e., $g \approx 1$. This is the many body localized discrete time crystal regime.

In Fig. 8 the circuit for the MBL-DTC is shown, including a noise channel Φ_i and a correction scheme [41]. Here the logical space are the classical bit strings that we want to protect from local flips. The error correction code measures a small contiguous region of M qubits in the Z basis after each Floquet unitary. For a product state in the computational basis the relevant local information is contained in the local Z parity data. This parity is constant in the ideal MBL-DTC phase. Thus, if one physical qubit has flipped, all correlations involving that qubit change sign. The qubit with the largest number of such sign changes is identified as the most likely error location. A corrective X operation is then applied to that qubit. We see the effect of the stabilization in comparison to a perfect quantum computer (unitary) and a noisy one in Fig. 8 c).

Such partial correction schemes can stabilize a phase even in the absence of fully error corrected quantum processor. For quantum simulation this is a promising viewpoint. Many target problems do not require preserving an unknown quantum state with arbitrary accuracy. They require preserving a physically meaningful structure, such as a symmetry sector, a conserved quantity, a topological constraint, or a pattern of spatiotemporal order.

References

- [1] H. De Raedt, J. Kraus, A. Hertel, V. Mehta, M. Bode, M. Hrywniak, K. Michielsen, and T. Lippert, *Future Gener. Comput. Syst.* **183**, 108592 (2026)
- [2] Y.I. Manin: *Vychislimoe i Nevychislimoe [Computable and Uncomputable]* (Sovetskoye Radio, Moscow, 1980) In Russian; English translation in *Mathematics as Metaphor: Selected Essays* (AMS, 2007)
- [3] P. Benioff, *J. Stat. Phys.* **22**, 563 (1980)
- [4] R.P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982)
- [5] D. Deutsch, *Proceedings of the Royal Society of London A* **400**, 97 (1985)
- [6] P.W. Shor, pp. 124–134 in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society Press, 1994)
- [7] P.W. Shor, *SIAM J. Comput.* **26**, 1484 (1997)
- [8] S. Lloyd, *Science* **273**, 1073 (1996)
- [9] S. Aaronson: *Quantum Computing Since Democritus* (Cambridge University Press, 2013)
- [10] A.M. Gleason, *J. Math. Mech.* **6**, 885 (1957)
- [11] S. Aaronson, arXiv:quant-ph/0401062 (2004)
- [12] P. Barrios Hita, A. Trushechkin, H. Kampermann, M. Epping, and D. Bruß, *Phys. Rev. Lett.* **136**, 240202 (2026)
- [13] A.Y. Kitaev, A.H. Shen, and M.N. Vyalyi: *Classical and Quantum Computation*, Graduate Studies in Mathematics, Vol. 47 (American Mathematical Society, 2002)
- [14] J. Kempe, A. Kitaev, and O. Regev, *SIAM J. Comput.* **35**, 1070 (2006)
- [15] H.F. Trotter, *Proc. Am. Math. Soc.* **10**, 545 (1959)
- [16] A.M. Childs, Y. Su, M.C. Tran, N. Wiebe, and S. Zhu, *Phys. Rev. X* **11**, 011020 (2021)
- [17] M. Suzuki, *J. Math. Phys.* **32**, 400 (1991)
- [18] D.W. Berry, G. Ahokas, R. Cleve, and B.C. Sanders, *Commun. Math. Phys.* **270**, 359 (2007)
- [19] G.H. Low and I.L. Chuang, *Phys. Rev. Lett.* **118**, 010501 (2017)
- [20] I.M. Georgescu, S. Ashhab, and F. Nori, *Rev. Mod. Phys.* **86**, 153 (2014)
- [21] J.I. Cirac and P. Zoller, *Nat. Phys.* **8**, 264 (2012)

- [22] M. Greiner, O. Mandel, T. Esslinger, T.W. Hänsch, and I. Bloch, *Nature* **415**, 39 (2002)
- [23] I. Bloch, J. Dalibard, and S. Nascimbène, *Nat. Phys.* **8**, 267 (2012)
- [24] R. Blatt and C.F. Roos, *Nat. Phys.* **8**, 277 (2012)
- [25] M.A. Nielsen and I.L. Chuang: *Quantum Computation and Quantum Information* (Cambridge University Press, 2010), 10th anniversary ed.
- [26] A. Barenco, C.H. Bennett, R. Cleve, D.P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J.A. Smolin, and H. Weinfurter, *Phys. Rev. A* **52**, 3457 (1995)
- [27] D.P. DiVincenzo, *Phys. Rev. A* **51**, 1015 (1995)
- [28] A.Y. Kitaev, *Russ. Math. Surv.* **52**, 1191 (1997)
- [29] C.M. Dawson and M.A. Nielsen, *Quantum Inf. Comput.* **6**, 81 (2006)
- [30] B. Fauseweh and J.-X. Zhu, *Quantum Inf. Process.* **20**, 138 (2021)
- [31] Z. Jiang, A. Kalev, W. Mruzekiewicz, and H. Neven, *Quantum* **4**, 276 (2020)
- [32] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G.H. Booth, and J. Tennyson, *Phys. Rep.* **986**, 1 (2022)
- [33] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J.M. Chow, and J.M. Gambetta, *Nature* **549**, 242 (2017)
- [34] D. Wecker, M.B. Hastings, and M. Troyer, *Phys. Rev. A* **92**, 042303 (2015)
- [35] A. Choquette, A. Di Paolo, P.K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, *Phys. Rev. Res.* **3**, 023092 (2021)
- [36] S. Liu, S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, *Phys. Rev. B* **107**, 024204 (2023)
- [37] C. Cao, Y. Zhou, S. Tannu, N. Shannon, and R. Joynt, *Quantum* **9**, 1942 (2025)
- [38] K. Lively, T. Bode, J. Szangolies, J.-X. Zhu, and B. Fauseweh, *Phys. Rev. Res.* **6**, 043254 (2024)
- [39] A.R. Calderbank and P.W. Shor, *Phys. Rev. A* **54**, 1098 (1996)
- [40] X. Mi, *et al.*, *Nature* **601**, 531 (2022)
- [41] G. Camacho and B. Fauseweh, *Phys. Rev. Res.* **6**, 033092 (2024)

Index

A

Aharonov-Bohm effect, 8.8, 8.15
alpha-shift expansion, 6.9
Anderson model, 5.25
Anderson molecule, 1.7, 1.45
Anderson-Higgs mechanism, 9.17
anticommutation relations, 4.5
atomic-sphere truncation, 1.33
auxiliary Hamiltonian, 13.8

B

$\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$, 10.8, 10.15
backflow, 13.13
bare expansion, 6.8
 functional, 6.7
BCS theory, 9.21
Bell state, 15.8
Berry connection, 8.7
Berry phase, 8.5
Bloch sphere, 15.6
bold expansion, 6.12

C

$\text{CaKFe}_4\text{As}_4$, 10.16
cavity quantum materials, 12.36
charge pump, 8.11
charge-density wave, 12.11
charge-transfer insulator, 12.6
Chern number, 8.4, 8.8
cluster Mott insulator, 11.15, 11.18, 11.28
 quasimolecular electronic structure, 11.15, 11.20
 quasimolecular magnetic moment, 11.16
 site disorder, 11.21
code space, 15.22
coherence length, 9.7
complex time-evolution, 5.23
configuration interaction, 4.3, 4.13
constrained density-functional theory
 formal, 2.14
 Lagrange multipliers, 2.18
 penalty functions, 2.19
continued fraction, 4.23
continuous-time quantum Monte Carlo, 1.15
crystal-field effects, 1.40

D

d-d excitation, 11.7
 phonon-assisted, 11.13
density functional theory, 1.2
density matrix, 4.4
DFT+DMFT, 1.28
diagrammatic Monte Carlo, 6.2
diffraction, 10.10
dimer, 11.15, 11.17
discrete time crystal, 15.25
double counting correction, 1.3, 1.30, 1.38
downfolding, 1.30, 4.27
Dvorin-Narath model, 5.26
dynamical localization, 12.18
dynamical mean-field theory, 1.25, 3.22
 impurity solver, 1.15
 multi-site and multi-orbital problems, 1.28
Dyson equation, 1.7, 3.8

E

edge states, 8.13
Ehrenfest classification, 10.4
entanglement, 5.6, 14.13, 15.8
 ebit, 15.9
 entropy, 15.9
error bars, 13.7

F

fermionic sign problem, 6.2
fermionic states, 14.9
fermionic ternary-tree, 15.16
ferroelectricity, 12.16
FeSe, 10.8
Feynman diagrams
 sampling, 6.14
 symmetrized sum, 6.14
 topology cancellations, 6.15
fidelity susceptibility, 15.20
field operators, 4.5
Floquet engineering, 12.18
Floquet-Bloch states, 12.19
fluctuation, 9.8
flux quantum, 9.16
fork tensor network, 5.13

G

gauge dependence, 8.7, 8.12
 gauge invariance, 9.18
 Gaussian approximation, 9.8, 9.17
 ghost states, 4.22
 Gibbs phase rule, 10.3
 Ginzburg-Landau theory, 9.2
 superconductors, 9.11
 superfluids, 9.5
 Goldstone modes, 9.10
 grand canonical potential, 3.4
 Grassmann variables, 3.12
 Green function, 3.5
 as functional integral, 3.13
 Gutzwiller factor, 13.10

H

Hadamard gate, 15.4
 Haldane model, 8.18
 Hartree-Fock approximation, 1.20, 1.26, 7.4
 Heisenberg model, 2.3
 Higgs mode, 9.19
 Hofstadter model, 8.15
 Hohenberg-Kohn theorems, 2.4
 Hubbard dimer, 1.5, 1.43
 Hubbard exciton, 12.23
 Hubbard model, 1.25, 2.2, 6.3, 7.3, 12.5
 multi-orbital, 1.28, 13.3
 single-orbital, 13.3
 Hubbard models for t_{2g} and e_g systems, 1.34
 Hubbard-Kanamori model, 5.19
 Hund's coupling, 5.21, 11.8, 11.14
 hybridization function, 1.6

I

imaginary time Green function, 3.4
 indistinguishability, 4.2
 intra- t_{2g} excitations, 11.7, 11.10, 11.13, 11.25
 inversion by partitioning, 4.27
 iridates, 11.18

J

Jastrow factor, 13.9
 Jordan-Wigner transformation, 15.15

K

Kohn-Sham approach, 2.6
 Krylov space, 4.19

Krylov time-evolution, 5.17

L

Löwdin orthonormalization, 4.25
 Lanczos method, 4.19
 Landau theory, 7.17, 9.2
 LDA+DMFT, 1.25, 1.28
 LDA+U vs LDA+DMFT, 1.20, 1.24, 1.40
 Lehmann representation, 1.47, 3.5
 Levy constrained-search approach, 2.5
 Lifshitz invariant, 9.27
 light-induced
 order, 12.10
 superconductivity, 12.27
 linear prediction, 5.15
 lithium vanadate, 5.18
 local Dyson equation, 1.7
 local spin-density approximation, 2.10
 localization of the basis, 1.33
 London theory, 9.12
 Luttinger-Ward functional, 3.18

M

magnetization measurements, 10.11
 magnetoelastic couplings, 2.26
 magnon mass, 2.26
 many-body Chern number, 8.9
 many-body thermalization, 15.18
 Markov-chain Monte Carlo
 many-configuration, 6.28
 matrix product state, 5.4, 14.8
 Matsubara frequencies, 3.4
 metal insulator transition, 3.22
 metastable states, 12.27
 Metropolis algorithm, 13.6
 misleading convergence, 6.13
 model building, 1.30
 moment expansion, 4.24
 Mott insulator, 12.5
 Mott transition, 1.25
 multi-spiral order, 7.21
 multiplets, 11.8

N

Néel order, 7.7, 7.12, 7.16
 nilpotent polynomials, 6.21
 non-spherical Coulomb terms, 1.38
 nonlinear phononics, 12.9

O

occupation number representation, 4.14
 one-body operators, 4.8
 orbital ordering, 1.41
 orbital selectivity, 13.23
 order parameter, 9.2, 12.10
 orthonormalization, 4.25

P

pairing correlation function, 13.21
 parameterized quantum circuit, 15.18
 particle-particle channel, 6.10
 Pauli matrices, 15.7
 Pauli string, 15.12
 phase diagram, 10.3
 phase gate, 15.8
 phase transition, 9.3
 experimental signatures, 10.6
 magnetic, 10.11
 second order, 9.5
 structural, 10.10
 superconducting, 10.12
 photodoping, 12.6
 photoinduced phase transitions, 12.2
 plasma frequency, 9.20
 potential-based SCF, 2.20
 pump-probe spectroscopy, 12.2
 purification, 5.14

Q

quantum circuit diagram, 15.9
 quantum control, 12.22
 quantum error correction, 15.22
 quantum geometric tensor, 14.21, 14.25
 quantum Hall effect, 8.3, 8.11, 8.14
 quantum Merlin Arthur completeness, 15.5
 quantum Monte Carlo
 continuous-time, 1.15
 variational, 13.4
 quantum simulation, 15.2
 analog, 15.5
 digital, 15.2
 qubit, 15.6

R

Rényi entropy, 14.14
 radius of convergence, 6.3
 reference system, 3.21

renormalization

functional formalism, 6.6

renormalized mean-field theory, 7.22

RIXS, 11.4

resonant enhancement, 11.6
 RIXS interferometry, 11.16, 11.23
 scattering geometry, 11.3
 transition-metal L edge, 11.4

S

Schmidt decomposition, 5.5, 14.14

coefficients, 15.8

second quantization, 4.5

self-consistency, 1.25

self-energy, 3.8

Hubbard dimer, 1.6

local, 1.6, 1.9

spectral representation, 3.10

shifted action, 6.8

singularities

complex interaction strength, 6.31

Slater determinant, 4.3

spectral function, 4.23

spin density-functional theory, 2.7

spin-orbit coupling, 11.5, 11.20

$J=0$ singlet, 11.10

half-filled shell, 11.14

intermediate regime, 11.9, 11.11, 11.15

spin-orbit entangled j states, 11.20

spin-orbit effects, 1.35, 1.40

spin-orbit interaction, 1.35

spin-spin correlations, 13.20

spin-statistics connection, 4.2

spiral order, 7.7, 7.12, 7.16

spontaneous symmetry breaking, 12.10

$\text{Sr}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$, 10.16

stabilizer, 15.23

static structure factor, 13.19

steepest descent, 4.17

stochastic reconfiguration, 13.17, 14.20

strange order, 7.9

stripe order

collinear, 7.17

bidirectional, 7.8

unidirectional, 7.7

coplanar bidirectional, 7.8

super-exchange, 12.6

Hamiltonians, 1.41
SWAP gate, 15.10
symmetry breaking, 6.9, 8.2
syndrome measurement, 15.23

T

T gate, 15.8
tensor network-DMFT, 5.8
topological
 equivalence, 8.4
 invariant, 8.4
 order, 8.4
tree tensor network, 5.13
trimer, 11.23
Trotter decomposition, 5.7
Trotter time evolution, 15.14
two-body operators, 4.9

U

unitary gate, 15.7
universality, 15.10

V

vacuum state, 4.5
variational cluster approximation, 3.26
variational Monte Carlo, 13.4, 13.10, 14.17
variational quantum eigensolver, 15.18
vortex, 9.17

W

Wannier functions, 1.31
wave function
 backflow, 13.13
 Jastrow-Slater, 13.8, 13.16
Weiss field, 1.26
Wick's theorem, 6.17

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
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
13. **Orbital Physics in Correlated Matter**
Lecture Notes of the Autumn School on Correlated Electrons 2023
edited by E. Pavarini and E. Koch (2023), 520 pages
ISBN 978-3-95806-689-2
14. **Correlations and Phase Transitions**
Lecture Notes of the Autumn School on Correlated Electrons 2024
edited by E. Pavarini and E. Koch (2024), 480 pages
ISBN 978-3-95806-751-6
15. **Understanding Correlated Materials with DMFT**
Lecture Notes of the Autumn School on Correlated Electrons 2025
edited by E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (2025), 480 pages
ISBN 978-3-95806-813-1
16. **Correlated Materials: Methods and Applications**
Lecture Notes of the Autumn School on Correlated Electrons 2026
edited by E. Pavarini and E. Koch (2026), 480 pages
ISBN 978-3-95806-908-4 (Print) 978-3-95806-909-1 (E-Book)