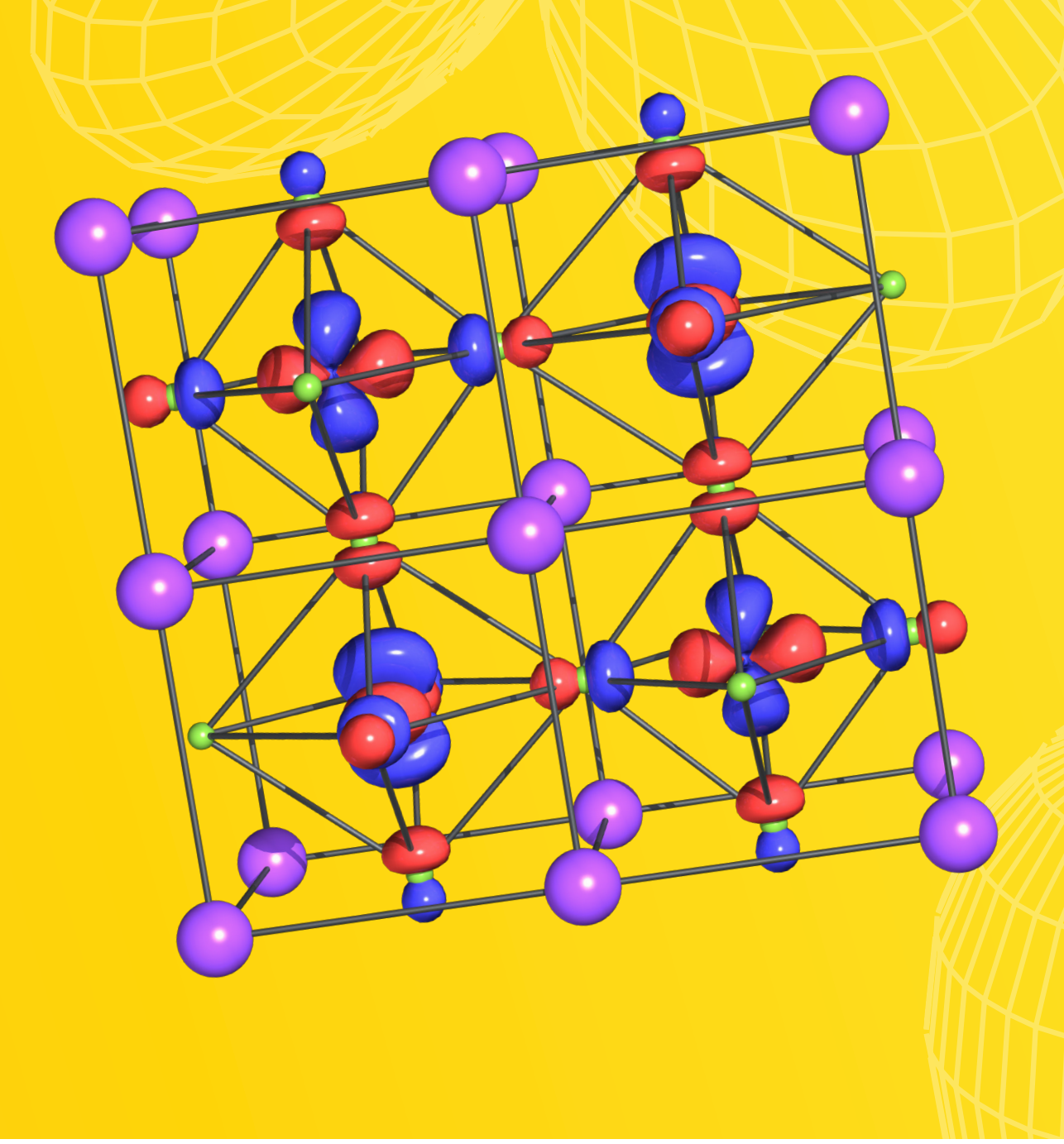


Orbital Physics in Correlated Matter

Eva Pavarini and Erik Koch (Eds.)

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 13

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

Lecture Notes of the Autumn School on Correlated Electrons 2023

Eva Pavarini and Erik Koch (Eds.)

Orbital Physics in Correlated Matter

Autumn School organized by
the Institute for Advanced Simulation
at Forschungszentrum Jülich
18 – 22 September 2023

Schriften des Forschungszentrums Jülich
Modeling and Simulation Band / Volume 13

ISSN 2192-8525 ISBN 978-3-95806-689-2

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

Publisher:	Forschungszentrum Jülich GmbH Institute for Advanced Simulation
Cover Design:	Grafische Medien, Forschungszentrum Jülich GmbH
Printer:	Kern GmbH, Bexbach
Copyright:	Forschungszentrum Jülich 2023
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 13

ISSN 2192-8525
ISBN 978-3-95806-689-2

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



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

Contents

Preface

1. Orbital Ordering in Materials
   Eva Pavarini
2. The Jahn-Teller Effect
   Arnout Ceulemans
3. Orbitally Induced Peierls Mechanism for Charge-Orbital Orderings in Transition-Metal Compounds
   Takashi Mizokawa
4. Multiplets in Transition Metal Ions and Introduction to Multiband Hubbard Models
   Robert Eder
5. Exchange Mechanisms
   Erik Koch
6. Spin-Orbital Entanglement in Mott Insulators
   Andrzej Oleś
7. Imaging Orbitals with X-rays
   Hao Tjeng
8. Probing Spin, Charge and Orbital Degrees of Freedom by X-Ray Spectroscopy
   Eva Benckiser
9. Strong Correlations at Oxide Interfaces: What is Hidden in a Plane View?
   Jak Chakhalian
10. Orbitals, Frustration and Quantum Criticality
    Matthias Vojta
11. Quantum Compass and Kitaev Models
    Jeroen van den Brink
12. Kitaev Magnets
    Simon Trebst
13. Self Interaction Corrections to Density Functional Theory
    Mark Pederson
14. Coupled-Cluster Theory for Materials Science
    Andreas Grüneis
15. Slave-Boson Theories of Multi-Orbital Correlated Systems
    Nicola Lanatà
16. DMFT for f-Electron Systems
    Bernard Amadon
17. Super-QMC: Strong Coupling Perturbation for Lattice Models
    Alexander Lichtenstein

Index

Preface

In a classic paper 50 years ago, Kugel and Khomskii demonstrated that in strongly-correlated systems orbital ordering can arise from a purely electronic super-exchange mechanism and not just the conventional co-operative Jahn-Teller effect. This work opened the field of orbital physics which, since then, is undergoing continuous growth. It was understood that, besides orbital ordering, super-exchange can give rise to the orbital analogue of spin-liquid states. It was shown that the directional character of the orbitals can introduce anisotropic super-exchange interactions, which, in a simplified setting, are described by compass models, a prototype for the Kitaev model. More surprising phenomena arise from the entanglement of spin and orbital degrees of freedom. New developments aim at tuning orbital occupations by pushing the system out of equilibrium, as well as at orbital-controlled electronics.

The goal of this year's school is to provide students with an overview of the state-of-the art in the field of orbital physics and the techniques used to investigate strongly-correlated systems hosting phenomena stemming from orbital degrees of freedom. After introducing fundamental models and effects, lectures will focus on their realizations in materials. Advanced lectures will address orbital phases and Kitaev systems, as well as theoretical approaches and experimental probes of spin, orbital, and charge degrees of freedom.

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 Institute for Advanced Simulation at the Forschungszentrum Jülich and the Jülich Supercomputer Centre provided the major part of the funding and were vital for the organization of the school as well as for the production of this book. The Institute for Complex Adaptive Matter (ICAM) continued also this year to endorse 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. H. Lexis of the Verlag des Forschungszentrum Jülich and to Mrs. D. Mans of the Grafische Betriebe for providing their expert support in producing the present volume on a tight schedule. We heartily thank our students and postdocs who helped with proof-reading the manuscripts, often on quite short notice: Elaheh Adibi, Qiwei Li, Neda Samani, and Xue-Jing Zhang.

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

Eva Pavarini and Erik Koch

August 2023

1 Orbital Ordering in Materials

Eva Pavarini
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Cubic crystal-field splitting **6**

3 Tight-binding e_g and t_{2g} bands of cubic perovskites **13**

4 Jahn-Teller effect **19**

5 Kugel-Khomskii superexchange **28**

6 The origin of orbital ordering in materials **34**

7 Conclusion **39**

A Constants and units **40**

B Atomic orbitals **40**
B.1 Radial functions 40
B.2 Real harmonics 41
B.3 Slater-Koster integrals 42

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

The term *orbital ordering* (OO) indicates the emergence of a broken symmetry state in which localized occupied orbitals form a regular pattern, in a similar way as spins do in magnetically ordered structures. Orbital ordering phenomena typically occur in Mott insulators with *orbital degrees of freedom*; for transition-metal compounds, the main focus of this lecture, the latter stem from the partially filled d shells of the transition metal. The perhaps most representative case is the perovskites $KCuF_3$, shown in Fig. 1. In first approximation $KCuF_3$ is cubic (O_h point group) with Cu^{2+} at the center of a regular octahedron of F^- ions (anions), enclosed in a cage of K^+ (cations). Due to O_h symmetry at the Cu site, the d manifold, 5-fold degenerate for free Cu^{2+}, splits into a t_{2g} triplet (xz, yz, xy), lower in energy, and a e_g doublet (x^2-y^2 and $3z^2-r^2$); the electronic configuration of the Cu^{2+} ion is thus $t_{2g}^6 e_g^3$ (one $3d$ hole). The t_{2g} states are completely filled and do not play any active role in OO; instead, electrons in the e_g^3 configuration have orbital degeneracy d=2. Making an analogy with spin degrees of freedom, they behave as an effective τ=1/2 *pseudospin*; in this view, one of the two e_g states, say $|x^2-y^2\rangle$, plays the role of the pseudospin up, $|\nearrow\rangle$, and the other one, $|3z^2-r^2\rangle$, of the pseudospin down, $|\searrow\rangle$. The two pseudospin states are degenerate and, by symmetry, one could expect them to be equally occupied. In reality the symmetry is broken and $KCuF_3$ is orbitally ordered with the orbital structure shown in Fig. 1; depicted are the empty (hole) e_g states at each Cu site. Furthermore, the system exhibits a *co-operative Jahn-Teller (JT) distortion*, also shown in Fig. 1, with long and short Cu-F bonds alternating in the $\boldsymbol{ab}$ plane. Indeed, the two phenomena – electronic OO and structural JT distortion – are concurrent; it is therefore difficult to say which one is the cause and which one is, instead, the effect. This is a classical case of a chicken-and-egg problem. The second paradigmatic system showing OO is $LaMnO_3$ (ion Mn^{3+}, configuration $3d^4$), the mother compound of colossal magnetoresistance manganites, also a perovskite. Due to the Hund's rule coupling J, the actual electronic configuration of Mn^{3+} is $t_{2g}^3 e_g^1$. The half-filled t_{2g}^3 state has no orbital degeneracy; the only orbital degrees of freedom are, as for $KCuF_3$, those associated with e_g electrons. Again, the system is orbitally ordered and OO goes hand in hand with the co-operative JT distortion. Among t_{2g} systems, i.e., materials with partially filled t_{2g} shells, classical examples of orbitally-ordered crystals are the perovskites $LaTiO_3$ and $YTiO_3$ (configuration t_{2g}^1), $LaVO_3$ and YVO_3 (t_{2g}^2), and Ca_2RuO_4 (t_{2g}^4); in these cases the t_{2g} electrons behave as an orbital pseudospin τ=1. Although this is not a prerequisite for orbital ordering, as we have seen, many orbitally-ordered materials are perovskites; for this reason in the present lecture we will use the perovskite structure as representative.

The origin of orbital ordering has been investigated for decades. One of the problems in clarifying its nature is that, while magnetic order can be directly probed, e.g., via neutron scattering experiments, orbital ordering is typically only indirectly observed. Indeed, its principal hallmark is the presence of the co-operative Jahn-Teller distortion itself. Identifying the origin of orbital ordering is thus intimately related to finding the cause of the co-operative Jahn-Teller distortion. In this lecture I will first illustrate the two main mechanisms [1, 2] which have been proposed as possible explanation for OO phenomena, the classical Jahn-Teller effect [1],

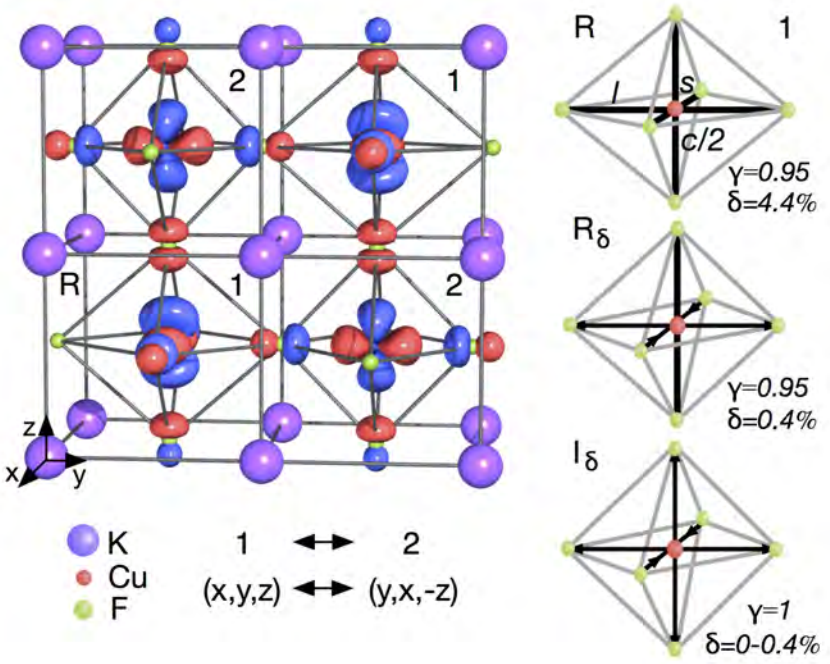


Fig. 1: *Crystal structure, distortions, and orbital ordering in $KCuF_3$. Cu is at the center of F octahedra enclosed in a K cage. The conventional cell is tetragonal with axes* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. *The pseudocubic axes* $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ *pointing towards neighboring Cu, are shown in the corner. Short (s) and long (l) CuF bonds alternate between* $\mathbf{x}$ *and* $\mathbf{y}$ *along all pseudocubic axes (co-operative Jahn-Teller distortion). The distortions are measured by $\delta=(l-s)/(l+s)/2$ and $\gamma=c/a\sqrt{2}$. R is the experimental structure ($\gamma=0.95$, $\delta=4.4\%$), R_δ ($\gamma=0.95$) and I_δ ($\gamma=1$) two ideal structures with reduced distortions. In the I_0 structure the cubic crystal-field at the Cu site splits the $3d$ manifold into a t_{2g} triplet and a e_g doublet. In the R structure, site symmetry is lowered further by the tetragonal compression ($\gamma<1$) and the Jahn-Teller distortion ($\delta \neq 0$). The figure shows the highest-energy $3d$ orbital. From Ref. [3].*

perhaps enhanced by Coulomb repulsion [4], and Kugel-Khomskii (KK) superexchange [2]. Kanamori well illustrated the first mechanism in an influential work [1] in 1960; the main idea is that electron-phonon coupling yields a static Jahn-Teller distortion, which lowers the symmetry of the system and produces a crystal-field splitting. As a consequence, electrons preferably occupy the lower energy states, giving rise to a periodic pattern of occupied orbitals. This is self-evident in the limit in which the crystal-field splitting is very large, let us say, larger than the bandwidth; the lower-energy states at each site will be clearly the first ones to be occupied. If, however, the bandwidth is large in comparison with the crystal-field splitting, the hopping integrals can strongly reduce such a tendency to orbital ordering. A natural question thus arises at this point. How large should the crystal-field splitting be to give rise to a orbitally-ordered state? To answer this question we have to remind ourselves that transition-metal systems with partially filled d shells are also typical examples of strongly-correlated materials. Their low-energy properties are believed to be well described by a generalized multi-band *Hubbard model*

$$\hat{H} = \hat{H}_0 + \hat{H}_U,$$

the sum over a one-electron term $\hat{H}_0$ describing the transition-metal d bands and a Coulomb electron-electron repulsion term $\hat{H}_U$. The one-electron term is

$$\hat{H}_0 = -\sum_{ii'}\sum_{\sigma}\sum_{mm'} t^{i,i'}_{mm'}\, c^{\dagger}_{im\sigma} c_{im'\sigma}\,,$$

where $c^{\dagger}_{im\sigma}$ creates an electron at site i with spin σ and orbital quantum number m, and the parameter $t^{i,i'}_{mm'}$ are the hopping integrals ($i{\neq}i'$) or the crystal-field splittings ($i{=}i'$). The Coulomb repulsion can be written as

$$\hat{H}_U = \frac{1}{2}\sum_{i}\sum_{\sigma\sigma'}\sum_{m_\alpha m'_\alpha}\sum_{m_\beta m'_\beta} U_{m_\alpha m_\beta m'_\alpha m'_\beta}\, c^{\dagger}_{im_\alpha\sigma} c^{\dagger}_{im_\beta\sigma'} c_{im'_\beta\sigma'} c_{im'_\alpha\sigma}\,.$$

The elements the Coulomb interaction tensor, $U_{m_\alpha m_\beta m'_\alpha m'_\beta}$, can be expressed in terms of the Slater integrals.[1] Here we will restrict the discussion to the e_g or t_{2g} manifolds only. In this case, in the basis of real harmonics, the Hubbard model takes the form

$$\begin{aligned}\hat{H}= &-\sum_{ii'}\sum_{\sigma}\sum_{mm'} t^{i,i'}_{mm'}\, c^{\dagger}_{im\sigma} c_{im'\sigma} + U\sum_{i}\sum_{m}\hat{n}_{im\uparrow}\hat{n}_{im\downarrow}\\ &+\frac{1}{2}\sum_{i}\sum_{\sigma\sigma'}\sum_{m\neq m'}(U{-}2J{-}J\delta_{\sigma,\sigma'})\,\hat{n}_{im\sigma}\hat{n}_{im'\sigma'}\\ &-J\sum_{i}\sum_{m\neq m'}\Big[c^{\dagger}_{im\uparrow}c^{\dagger}_{im\downarrow}c_{im'\uparrow}c_{im'\downarrow}{+}c^{\dagger}_{im\uparrow}c_{im\downarrow}c^{\dagger}_{im'\downarrow}c_{im'\uparrow}\Big]\,, \end{aligned} \qquad (1)$$

where m, m' are here either t_{2g} or e_g states, $U_{mm'mm'}{=}U_{m,m'}{=}U{-}2J(1{-}\delta_{m,m'})$ and, for $m \neq m'$, $U_{mm'm'm}{=}J_{m,m'}{=}J$. The last two terms describe the pair-hopping and spin-flip processes ($U_{mmm'm'}{=}J_{m,m'}$ if we use a basis of real harmonics, while for spherical harmonics $U_{mmm'm'}{=}0$). Finally, $U{=}U_0$ and $J{=}J_1$ (t_{2g} electrons) or $J{=}J_2$ (e_g electrons), with

$$\begin{aligned} U_0 =& F_0 + \frac{8}{5}\mathcal{J}_{\rm avg}, & \mathcal{J}_{\rm avg}=&\frac{5}{7}\frac{1}{14}(F_2+F_4)\\ J_1 =& \frac{3}{49}F_2 + \frac{20}{9}\frac{1}{49}F_4, & J_2 =& -2\mathcal{J}_{\rm avg} + 3J_1\,. \end{aligned}$$

In strongly correlated systems described by a Hamiltonian of type (1), it turns out that a small crystal-field splitting, a fraction of the bandwidth, is sufficient to produce orbital order even at high temperature. This happens because the Coulomb repulsion effectively enhances it, while suppressing orbital fluctuations [4]. Hence, the mechanism illustrated by Kanamori becomes very efficient in the presence of strong correlations (small t/U limit, the typical limit for Mott insulators; here t is an average hopping integral). This is, however, not the end of the story: Coulomb electron-electron interaction provides, in addition, an alternative explanation of the origin of orbital ordering. In a seminal work, Kugel and Khomskii [2] have shown in 1973 that, in the presence of orbital degeneracy, many-body effects can produce orbital ordering

[1] For a pedagogical introduction see, e.g, Ref. [5].

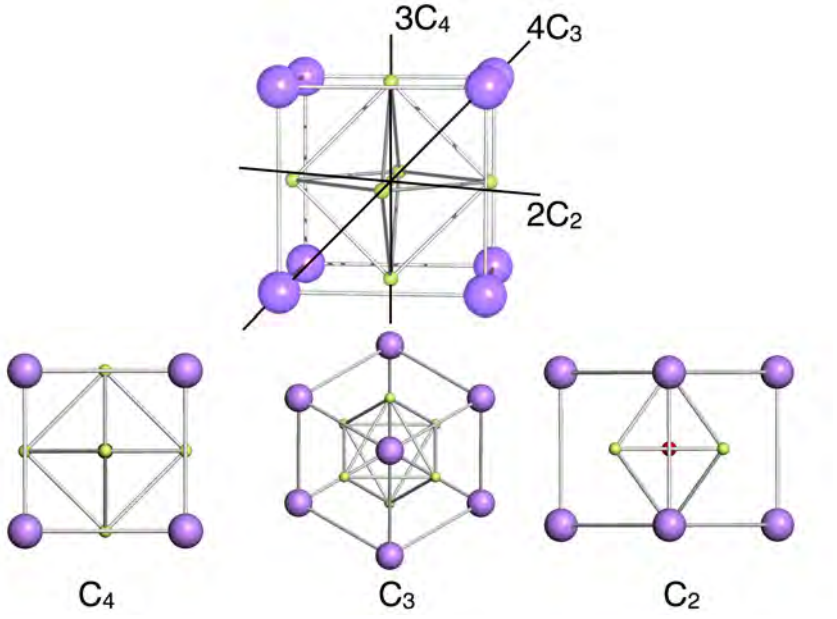


Fig. 2: *The unit cell of a cubic perovskite ABC_3 and its symmetry axes; the lattice constant is a. The transition metal B (red) is at $(0,0,0)$; the ligands C (green) are located at $(\pm a/2,0,0)$, $(0,\pm a/2,0)$,$(0,0,\pm a/2)$ and form an octahedron; the cations A are located at $(\pm a/2,\pm a/2,\pm a/2)$, $(\pm a/2,\mp a/2,\pm a/2)$, $(\mp a/2,\pm a/2,\pm a/2)$, $(\pm a/2,\pm a/2,\mp a/2)$ and form a cube. The bottom figures illustrate the rotational symmetries of the cell.*

even in the absence of a static distortion, i.e., of a crystal-field splitting. This happens via electronic spin-orbital superexchange, the effective low-energy interaction which emerges, in the small t/U limit, from the orbitally-degenerate Hubbard model. In this picture, the cooperative Jahn-Teller distortion is rather the consequence than the cause of orbital order. As I already mentioned, the predictions of the two theories for the final broken-symmetry structure are basically identical for most systems; thus it is very hard to determine which of the two mechanisms, Jahn-Teller effect or Kugel-Khomskii superexchange, dominates. In the last part of the lecture we will see how the problem was recently solved in representative cases [3, 6] by using a new theoretical approach based on the local-density-approximation + dynamical mean-field theory (LDA+DMFT) [7–10] method. For the paradigmatic systems $KCuF_3$ and $LaMnO_3$, it was shown that Kugel-Khomskii superexchange alone, although strong, cannot explain the presence of the Jahn-Teller distortion above 350 K ($KCuF_3$) [3] and 650 K ($LaMnO_3$) [6]; experimentally, however, the distortion persists in both systems basically up to the melting temperature. This leads to the conclusion that a mechanism directly generating a static crystal-field splitting, such as the standard Jahn-Teller effect, is necessary to explain the experimental findings. In fact, for $KCuF_3$ and other ionic systems, it turns out that even the classical JT picture fails. A new mechanism, based on Born-Mayer repulsion, has to be invoked to describe the actual experimental structure and the associated ordering at high temperature [11]. A true Kugel-Khomskii system was recently identified in $LaVO_3$ [12].

2 Cubic crystal-field splitting

Let us consider a system with the ideal cubic perovskite structure ABC_3, shown in Fig. 2. In this structure, B is the transition metal with partially filled d shell. The site symmetry at site B is cubic; thus, as we mentioned before, d states split into e_g and t_{2g}. Let us understand how exactly this happens. For a free ion, the potential $v_{\rm R}(\boldsymbol{r})$ which determines the single-electron energies is rotationally invariant, i.e., it has symmetry $O(3)$. This means that all single-electron states within a given l shell are degenerate, as it happens in the case of hydrogen-like atoms. When the same ion is inside a molecule or a solid, $v_{\rm R}(\boldsymbol{r})$ has in general lower symmetry, corresponding to a finite point group.[2] Thus one-electron states within a given shell l, degenerate for the free atom, can split. The symmetry reduction arises from the *crystal field*; the latter has two components, the Coulomb potential generated by the surrounding charged ions, dominant in ionic crystals, and the *ligand field* due to the bonding neighbors. In this section we will analyze the first contribution; the covalent contribution to the crystal-field splitting is discussed in the next section. Both effects give rise to a similar splitting of levels; which contribution dominates depends on the system.

Let us thus assume that the crystal is perfectly ionic and that the ions can be treated as point charges q_α (point-charge model). Then, the one-electron potential can be written as

$$v_{\rm R}(\boldsymbol{r}) = \sum_\alpha \frac{q_\alpha}{|\boldsymbol{R}_\alpha - \boldsymbol{r}|} = v_0(r) + \sum_{\alpha\neq 0} \frac{q_\alpha}{|\boldsymbol{R}_\alpha - \boldsymbol{r}|} = v_0(r) + v_c(\boldsymbol{r}), \tag{2}$$

where $\boldsymbol{R}_\alpha$ are the positions of the ions and q_α their charges. The term $v_0(r)$ is the ionic central potential at site $\boldsymbol{R}_0$, with spherical symmetry. The term $v_c(\boldsymbol{r})$ is the electric field generated at a given site $\boldsymbol{R}_0$ by all the surrounding ions in the crystal and it is called *crystal-field potential*.

For the perovskite structure ABC_3 we are interested in the crystal-field potential at the site of the transition metal, B. Let us first assume that only the contribution of nearest neighbors (the negative C ions, typically oxygens or fluorines) is relevant. The six C ions are located at positions $(\pm d_C, 0, 0), (0, \pm d_C, 0), (0, 0, \pm d_C)$ and have all the same charge q_C, while the B ion is at $(0, 0, 0)$; in terms of a, the cubic lattice constant, $d_C{=}a/2$. Then we can write the potential around ion B as

$$v_{\rm R}(\boldsymbol{r}) = \frac{q_B}{r} + \frac{q_C}{d_C}\left[\Delta v\left(\frac{x}{d_C};\frac{r}{d_C}\right) + \Delta v\left(\frac{y}{d_C};\frac{r}{d_C}\right) + \Delta v\left(\frac{z}{d_C};\frac{r}{d_C}\right)\right]$$

where

$$\Delta v(\xi;\rho) = \frac{1}{\sqrt{1+\rho^2}}\left[\frac{1}{\sqrt{1+\frac{2\xi}{1+\rho^2}}} + \frac{1}{\sqrt{1-\frac{2\xi}{1+\rho^2}}}\right].$$

Via the Taylor expansion

$$\frac{1}{\sqrt{1+\eta}} \sim 1 - \frac{1}{2}\eta + \frac{3}{8}\eta^2 - \frac{5}{16}\eta^3 + \frac{35}{128}\eta^4 + \dots$$

[2] For a concise introduction to group theory see, e.g., Ref. [13], chapter 6.

we can find an approximate expression of $\Delta v(\xi;\rho)$ for small ξ, i.e., close to ion B; the first contribution with less than spherical symmetry is

$$v_{\rm oct}(\boldsymbol{r}) = \frac{35}{4}\frac{q_C}{d_C^5}\left(x^4+y^4+z^4-\frac{3}{5}r^4\right) = D\left(x^4+y^4+z^4-\frac{3}{5}r^4\right).$$

We can rewrite this potential as

$$v_{\rm oct}(\boldsymbol{r}) = \frac{7\sqrt{\pi}}{3}\frac{q_C}{d_C^5}r^4\left[Y_0^4(\vartheta,\varphi)+\sqrt{\frac{5}{14}}\left(Y_4^4(\vartheta,\varphi)+Y_{-4}^4(\vartheta,\varphi)\right)\right], \tag{3}$$

where

$$\begin{aligned} Y_0^4(\vartheta,\varphi) &= \frac{3}{16}\frac{1}{\sqrt{\pi}}\left(35\cos^4\vartheta-30\cos^2\vartheta+3\right) = \frac{3}{16}\frac{1}{\sqrt{\pi}}\frac{35z^4-30z^2r^2+3r^4}{r^4}, \\ Y_{\pm4}^4(\vartheta,\varphi) &= \frac{3}{16}\sqrt{\frac{35}{2\pi}}\sin^4\vartheta e^{\pm4i\varphi} = \frac{3}{16}\sqrt{\frac{35}{2\pi}}\frac{(x\pm iy)^4}{r^4}. \end{aligned}$$

To obtain the crystal field due to the cubic cage of cations A (with charge q_A), shown in Fig. 2 we repeat the same calculation; the main difference is that there are eight A ions, located at positions $(\pm d_C,\pm d_C,\pm d_C)$,$(\mp d_C,\pm d_C,\pm d_C)$, $(\pm d_C,\mp d_C,\pm d_C)$, $(\pm d_C,\pm d_C,\mp d_C)$, with the distance from the origin being $d_A{=}\sqrt{3}a/2$. By following the same procedure that we used for B octahedron, one can show that

$$v_{\rm cube}(\boldsymbol{r}) = -\frac{8}{9}\frac{q_A}{q_C}\left(\frac{d_C}{d_A}\right)^5 v_{\rm oct}(\boldsymbol{r}),$$

i.e., $v_{\rm cube}(\boldsymbol{r})$ has the same form as $v_{\rm oct}(\boldsymbol{r})$; this happens because a cube and an octahedron are dual polyhedra[3] and have therefore the same symmetry properties. If $q_A/q_C>0$, $v_{\rm cube}(\boldsymbol{r})$ has opposite sign than $v_{\rm oct}(\boldsymbol{r})$; in the case of a perovskite, however, A positions are occupied by cations, i.e., positive ions; thus the crystal field due to the A cage has the same sign of the crystal field generated by the B octahedron.

The crystal-field potential $v_{\rm c}(\boldsymbol{r})$ lowers the site symmetry and can therefore split the $(2l{+}1)$-fold degeneracy of the atomic levels. To calculate how the l manifold splits, we use two approaches. The first is exact and based on group theory. We assume for simplicity that the symmetry is only O (group of the proper rotations which leave a cube invariant); using the full symmetry group of the cube, $O_h = O\otimes C_i$ (where C_i is the group made by the identity and the inversion) does not change the result, because the spherical harmonics are all either even or odd. The character table of group O is given by

partner	functions	O	E	$8C_3$	$3C_2$	$6C_2'$	$6C_4$
	$(x^2+y^2+z^2)$	A_1	1	1	1	1	1
		A_2	1	1	1	-1	-1
	$(x^2-y^2, 3z^2-r^2)$	E	2	-1	2	0	0
(R_x,R_y,R_z)	(x,y,z)	T_1	3	0	-1	-1	1
	(xy,xz,yz)	T_2	3	0	-1	1	-1

(4)

[3]Every polyhedron has a dual which can be obtained by exchanging the location of faces and vertices.

Let us explain this table. The first line yields the group, here O, and the symmetry operations of the group, collected in classes $\mathcal{C}_k$, here $\{E\}, \{C_3\}, \{C_2\}, \{C_2'\}, \{C_4\}$. For each class only a representative element is given and the number N_k in front of this element yields the number of operations in the class; for example $8C_3$ indicates 8 symmetry operations in class $\{C_3\}$. The symmetry operation C_n is an anticlockwise rotation of an angle $\alpha = 2\pi/n$. For a finite group with h elements, the h group operations $\{O(g)\}$ can be expressed as h matrices $\{\Gamma(g)\}$ acting on an invariant linear space; the basis of this space, $\{|m\rangle\}$, can be, for example, a finite set of linearly independent functions, such as the spherical harmonics with angular quantum number l. The collection of matrices $\{\Gamma(g)\}$ is a *representation* of the group; the dimension of the invariant linear space yields also the dimension of the matrices, i.e., the *dimensionality* of the representation. Each group has infinitely many possible representations, but some sets are special and play the role of an orthonormal basis in a space of vectors; they are called *irreducible*. If G is the group of operations which leave the Hamiltonian invariant, the irreducible representations of G can be used to classify all eigenstates of the Hamiltonian; eigenstates which build a basis for different irreducible representations are mutually orthogonal and have typically (leaving the cases of accidental degeneracy and hidden symmetry aside) different energies. The irreducible representations Γ_i of group O are listed in the first column of Table 4, below the group name; they are A_1 (trivial representation, made of 1-dimensional identity matrices), A_2, also 1-dimensional, E, two-dimensional, and T_1 and T_2, both three-dimensional. The numbers appearing in Table 4 are the characters $\chi_i(g)$, defined as

$$\chi_i(g) = \mathrm{Tr}\ \Gamma_i(g) = \sum_m \langle m|\Gamma_i(g)|m\rangle = \sum_m \Gamma_i^{mm}(g)\,.$$

For a given representation (corresponding to a line of Table 4) the character for a specific element can be found below the corresponding class label (columns of Table 4); all elements in the same class have the same character. Thus the second column of the character table, showing the character of the identity, yields also the dimensionality d_i of the representation itself. Next we calculate the characters of the matrix representation Γ^l constructed using spherical harmonics with angular quantum number l as a basis. An easy way to do this is to assume that the rotation axis is also the axis of quantization, i.e., $\hat{z}$; the characters do not depend on the actual direction of the quantization axis but only on the angle α of rotation. Thus for $O(g) = C_\alpha$ we have

$$\begin{aligned} C_\alpha\ Y_m^l(\vartheta,\varphi) =& Y_m^l(\vartheta,\varphi-\alpha) = e^{-im\alpha}\ Y_m^l(\vartheta,\varphi) \\ \Gamma_{mm'}^l(C_\alpha) =& \delta_{mm'} e^{-im\alpha}. \end{aligned}$$

This yields the following expression for the character

$$\chi^l(C_\alpha) = \sum_{m=-l}^{l} e^{-im\alpha} = \frac{\sin(l+\frac{1}{2})\alpha}{\sin\frac{\alpha}{2}}\,.$$

The characters for representations Γ^l are therefore

O	E	$8C_3$	$3C_2$	$6C_2$	$6C_4$
$\Gamma^0 = \Gamma^s$	1	1	1	1	1
$\Gamma^1 = \Gamma^p$	3	0	-1	-1	1
$\Gamma^2 = \Gamma^d$	5	-1	1	1	-1
$\Gamma^3 = \Gamma^f$	7	1	-1	-1	-1

In spherical symmetry (group $O(3)$) representations Γ^l are irreducible. In cubic symmetry (group O), instead, the Γ^l can be reducible, i.e., they can be written as the tensorial sum $\oplus$ of irreducible representations of the group O. The various components can be found by using the orthogonality properties of irreducible representations, which lead to the decomposition formula

$$\Gamma^l = \bigoplus_i a_i \Gamma_i \qquad \text{with} \quad a_i = \langle \Gamma_i | \Gamma^l \rangle = \frac{1}{h} \sum_g \left[\chi_i(g)\right]^* \chi^l(g)\,, \tag{5}$$

where h, the number of elements in the group, is 24 for group O. Hereafter the symmetry representations of electronic states are written in lower case to distinguish them from capital letters which we will use later for labeling vibrational modes. We find

$$\begin{aligned} \Gamma^s &= a_1 \\ \Gamma^p &= t_1 \\ \Gamma^d &= e \oplus t_2 \\ \Gamma^f &= a_2 \oplus t_1 \oplus t_2\,. \end{aligned}$$

Thus, in cubic symmetry, the s- and the p-functions do not split, because the a_1 irreducible representation is one-dimensional and the t_1 irreducible representation is 3-dimensional. Instead, d-functions split into a doublet and a triplet, and f-functions into a singlet and two triplets. To determine which functions $\{|m\rangle_i\}$ form a basis (a so-called set of *partner functions*) for a specific irreducible representation Γ_i we can, e.g., use the projector for that representation

$$\hat{\mathcal{P}}_i = \frac{d_i}{h} \sum_g \left[\chi_i(g)\right]^* O(g). \tag{6}$$

In our case, we can read directly the partner functions $\{|m\rangle_i\}$ for a given irreducible representation of the group O in the first column of Table 4, on the left. In short, for representation e possible partner functions are ($x^2{-}y^2$, $3z^2{-}r^2$) and for representation t_2 we can instead use (xy, xz, yz). A small step is still missing: As we already mentioned, the full symmetry of the B site is O_h, and the group O_h can be obtained as direct product, $O_h = O \otimes C_i$; with respect to O, group O_h has twice the number of elements and classes, and thus twice the number of irreducible representations. The latter split into even ($a_{1g}, a_{2g}, e_g, t_{1g}, t_{2g}$) and odd ($a_{1u}, a_{2u}, e_u, t_{1u}, t_{2u}$). All d-functions are even, and therefore $x^2{-}y^2$ and $3z^2{-}r^2$ are partners functions for the e_g irreducible representation, while xy, xz, yz are partner functions for the

t_{2g} irreducible representation. Summarizing, t_{2g} states (xy, xz, and yz) and e_g states (x^2-y^2 and $3z^2-r^2$) have in general (again excluding the cases of accidental degeneracy and hidden symmetry) different energy.

Group theory tells us *if* the degenerate $2l+1$ levels split at a given site in a lattice, but not of *how much* they do split, and which orbitals are higher in energy. We can, however, calculate the crystal-field splitting approximately using the potential (3) as a perturbation. This is the second approach previously mentioned; differently from group theory, it is not exact, but it gives us an estimate of the size of the effect and the sign of the splitting. For d states we can calculate the elements of the octahedral potential $v_{\rm oct}(\boldsymbol{r})$ in the basis of atomic functions $\psi_{nlm}(\rho,\vartheta,\varphi)=R_{nl}(\rho)Y^l_m(\vartheta,\varphi)$, where $R_{nl}(\rho)$ is the radial part, $\rho=Zr$, Z is the atomic number, $Y^m_l(\vartheta,\varphi)$ a spherical harmonic, and n the principal quantum number (Appendix B). We obtain

$$\langle\psi_{n20}\ \ |\hat{v}_{\rm oct}|\psi_{n20}\ \ \rangle = +6Dq \qquad\qquad \langle\psi_{n2\pm1}|\hat{v}_{\rm oct}|\psi_{n2\pm1}\rangle = -4Dq$$

$$\langle\psi_{n2\pm2}|\hat{v}_{\rm oct}|\psi_{n2\pm2}\rangle = +\ \ Dq \qquad\qquad \langle\psi_{n2\pm2}|\hat{v}_{\rm oct}|\psi_{n2\mp2}\rangle = +5Dq$$

where $Dq=q_C\langle r^4\rangle/6d_C^5$ and $\langle r^k\rangle=\int r^2dr\ r^k\ R^2_{n2}(Zr)$. The crystal-field splitting between e_g and t_{2g}-states can be then obtained by diagonalizing the crystal-field matrix

$$H_{\rm CF} = \begin{pmatrix} Dq & 0 & 0 & 0 & 5Dq \\ 0 & -4Dq & 0 & 0 & 0 \\ 0 & 0 & 6Dq & 0 & 0 \\ 0 & 0 & 0 & -4Dq & 0 \\ 5Dq & 0 & 0 & 0 & Dq \end{pmatrix}.$$

We find two degenerate e_g eigenvectors with energy $6Dq$

$$|\psi_{n20}\rangle = |3z^2-r^2\rangle,$$
$$\frac{1}{\sqrt{2}}\big(|\psi_{n2-2}\rangle+|\psi_{n22}\rangle\big) = |x^2-y^2\rangle,$$

and three degenerate t_{2g} eigenvectors with energy $-4Dq$

$$\frac{i}{\sqrt{2}}\big(|\psi_{n2-2}\rangle-|\psi_{n22}\rangle\big) = |xy\rangle,$$
$$\frac{1}{\sqrt{2}}\big(|\psi_{n2-1}\rangle-|\psi_{n21}\rangle\big) = |xz\rangle,$$
$$\frac{i}{\sqrt{2}}\big(|\psi_{n2-1}\rangle+|\psi_{n21}\rangle\big) = |yz\rangle.$$

The total splitting is

$$\Delta_{\rm CF} = E_{e_g}-E_{t_{2g}} = 10Dq.$$

Thus the e_g-states are actually higher in energy than the t_{2g}-states. This happens because e_g electrons point towards the negative C ions (see Fig. 3), and will therefore feel a larger Coulomb repulsion than t_{2g} electrons, which have the lobes directed between two negative C ions.

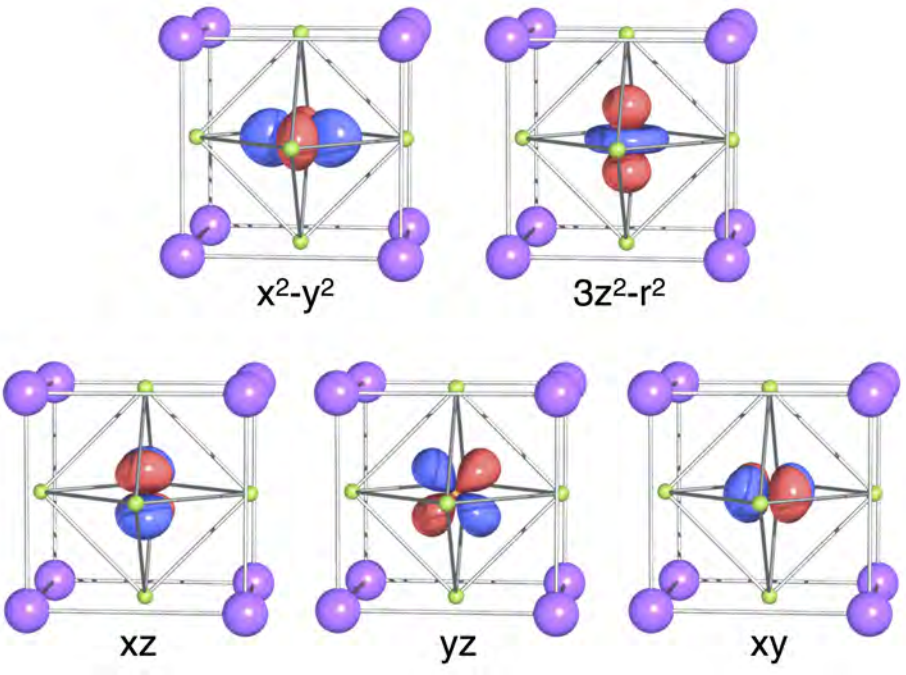


Fig. 3: *The Cu e_g and t_{2g} Wannier orbitals for the cubic perovskite $KCuF_3$, obtained from first principles calculations, using a Wannier basis that spans all bands.*

How general is, however, this result? We obtained it via a truncated Taylor expansion of the potential close to the nucleus. Does this mean that we have perhaps neglected important higher-order terms? For a generic lattice, we can expand the crystal-field potential (2) in spherical harmonics using the exact formula

$$\frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} = \sum_{k=0}^{\infty} \frac{r_<^k}{r_>^{k+1}} \frac{4\pi}{2k+1} \sum_{q=-k}^{k} Y_q^k(\vartheta_2, \varphi_2) \overline{Y}_q^k(\vartheta_1, \varphi_1),$$

where $r_<$ ($r_>$) is the smaller (larger) of r_1 and r_2. The crystal-field potential takes the form

$$v_c(\boldsymbol{r}) = \sum_{k=0}^{\infty} \sum_{q=-k}^{k} B_q^k Y_q^k, \tag{7}$$

where $B_q^k = (-1)^q \bar{B}_{-q}^k$. Although the series in (7) is in principle infinite, one can terminate it by specifying the wavefunctions, since

$$\langle Y_m^l | Y_q^k | Y_{m'}^l \rangle = 0 \quad \text{if} \quad k > 2l.$$

For example, for p electrons $k \leq 2$, for d-electrons, $k \leq 4$, and f electrons $k \leq 6$. Thus, for d-electrons and O_h symmetry, the terms that appear in the potential (3) are actually also the only ones to be taken into account, because all other terms yield an expectation value equal to zero. Finally, the derivation of both equations (3) and (7) presented here might let us think that the first-nearest neighbors are those that determine the crystal field. This is, however, not always the case, because Coulomb repulsion is a long-range interaction; for example, in some systems

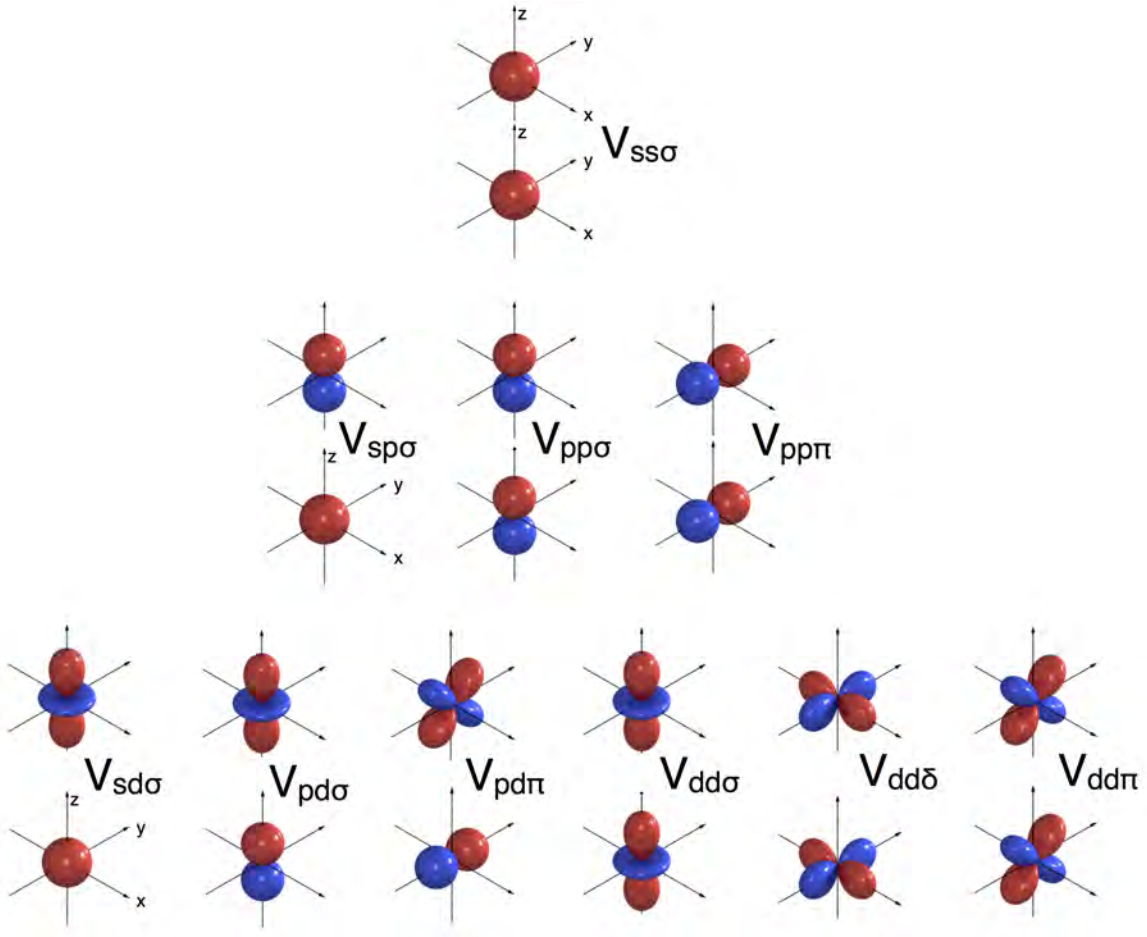


Fig. 4: *Independent Slater-Koster two-center integrals for s, p, and d atomic orbitals (Appendix B). The label σ indicates that the bonding state is symmetrical with respect to rotations about the bond axis; the label π that the bond axis lies in a nodal plane; the label δ that the bond axis lies in two nodal planes.*

the first-nearest neighbors yield cubic symmetry at a given site but further neighbors lower the symmetry.[4] Furthermore, the point-charge model discussed in this section is useful to explain the relation between crystal field and site symmetry, however yields unsatisfactory results for the crystal-field splitting in real materials. Corrections beyond the point-charge approximation turn out to be important. In addition, as we will see in the next section, in many systems the crystal field has a large, sometimes dominant, covalent contribution, the ligand field. The modern approach to calculate crystal-field splittings including the ligand-field contribution is based on material-specific potentials obtained *ab-initio* via density-functional-theory (DFT) and the associated DFT localized Wannier functions. Nevertheless, it is worth to point out the remarkable success of the point-charge model in giving qualitatively correct d crystal-field states in cubic perovskites; such a success relies on the fact that this approach, even if approximate, yields the exact symmetry of final states, i.e., the same obtained via group theory, and does not neglect any relevant (e.g., high-order) term.

[4] This means that, of course, O_h is not the actual symmetry of the site.

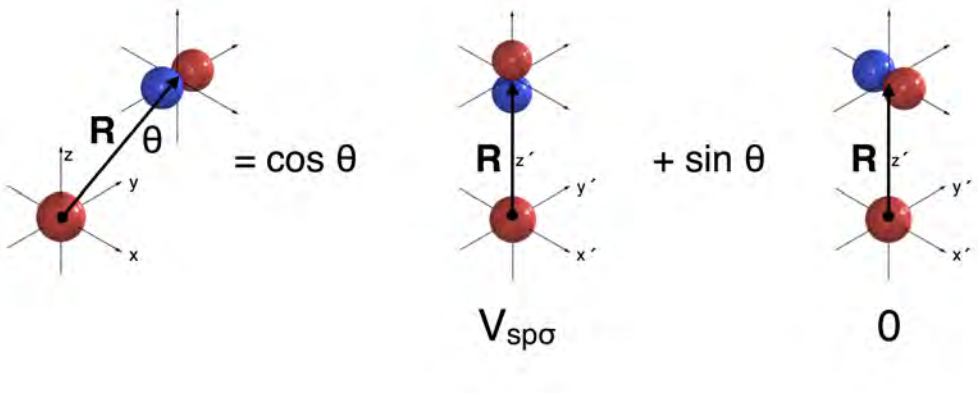


Fig. 5: *Illustration of the decomposition of a general s-p two-center integral in terms of $V_{sp\sigma}$.*

3 Tight-binding e_g and t_{2g} bands of cubic perovskites

In this section we will construct the bands of $KCuF_3$ in the cubic limit using tight-binding theory. Let us first remind ourselves of the crucial steps of this approach. The one-electron Hamiltonian can be written as

$$\hat{h}_e(\boldsymbol{r}) = -\frac{1}{2}\nabla^2 + \sum_{i\alpha} v(\boldsymbol{r}-\boldsymbol{T}_i-\boldsymbol{R}_\alpha) = -\frac{1}{2}\nabla^2 + v_R(\boldsymbol{r}),$$

where $\boldsymbol{R}_\alpha$ are the positions of the basis $\{\alpha\}$ atoms in the unit cell and $\boldsymbol{T}_i$ the lattice vectors. We take as a basis atomic orbitals with quantum numbers lm (we drop here the principal quantum number for convenience). For each atomic orbital we construct a Bloch state

$$\psi^\alpha_{lm}(\boldsymbol{k},\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_i e^{i\boldsymbol{T}_i\cdot\boldsymbol{k}}\ \psi_{lm}(\boldsymbol{r}-\boldsymbol{T}_i-\boldsymbol{R}_\alpha), \tag{8}$$

where N is the number of lattice sites. In the Bloch basis (8), the Hamiltonian and the overlap matrix are given by

$$H^{\alpha,\alpha'}_{lm,l'm'}(\boldsymbol{k}) = \langle \psi^\alpha_{lm}(\boldsymbol{k}) | \hat{h}_e | \psi^{\alpha'}_{l'm'}(\boldsymbol{k}) \rangle,$$
$$O^{\alpha,\alpha'}_{lm,l'm'}(\boldsymbol{k}) = \langle \psi^\alpha_{lm}(\boldsymbol{k}) | \psi^{\alpha'}_{l'm'}(\boldsymbol{k}) \rangle.$$

These matrices define a generalized eigenvalue problem, the solution of which yields the band structure. The Hamiltonian matrix is given by

$$H^{\alpha,\alpha'}_{lm,l'm'}(\boldsymbol{k}) = \varepsilon^0_{l'\alpha'}\, O^{\alpha,\alpha'}_{lm,l'm'}(\boldsymbol{k}) + \Delta\varepsilon^\alpha_{lm,l'm'}\,\delta_{\alpha,\alpha'} - \frac{1}{N}\sum_{i\alpha\neq i'\alpha'} e^{i(\boldsymbol{T}_{i'}-\boldsymbol{T}_i)\cdot\boldsymbol{k}}\ t^{i\alpha,i'\alpha'}_{lm,l'm'}\,.$$

Here $\varepsilon^0_{l\alpha}$ are the atomic levels, and $\Delta\varepsilon^\alpha_{lm,l'm'}$ the crystal-field matrix elements

$$\Delta\varepsilon^\alpha_{lm,l'm'} = \int d\boldsymbol{r}\ \overline{\psi_{lm}}(\boldsymbol{r}-\boldsymbol{R}_\alpha)\Big[v_R(\boldsymbol{r}) - v(\boldsymbol{r}-\boldsymbol{R}_\alpha)\Big]\psi_{l'm'}(\boldsymbol{r}-\boldsymbol{R}_\alpha)\,, \tag{9}$$

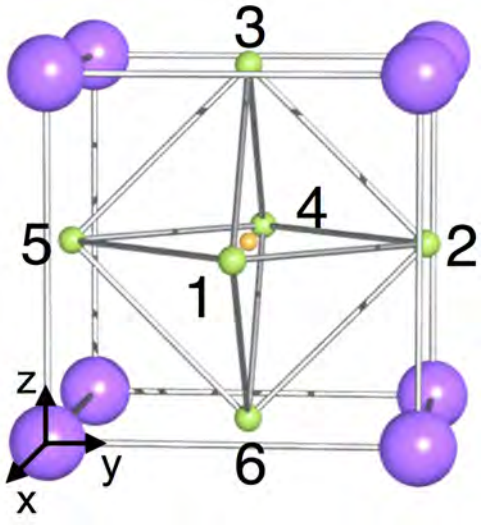


Fig. 6: *Unit cell of idealized cubic $KCuF_3$ with cubic axes in the left corner.*

which are two-center integrals. Finally,

$$t^{i\alpha,i'\alpha'}_{lm,l'm'} = -\int d\boldsymbol{r}\ \overline{\psi_{lm}}(\boldsymbol{r}-\boldsymbol{R}_\alpha-\boldsymbol{T}_i)\Big[v_R(\boldsymbol{r})-v(\boldsymbol{r}-\boldsymbol{R}_{\alpha'}-\boldsymbol{T}_{i'})\Big]\psi_{l'm'}(\boldsymbol{r}-\boldsymbol{R}_{\alpha'}-\boldsymbol{T}_{i'}). \qquad (10)$$

The hopping integrals (10) contain two- and three-center terms; if the basis is sufficiently localized we can, however, neglect the three-center contributions and assume $t^{i\alpha,i'\alpha'}_{lm,l'm'} \sim -V^{i\alpha,i'\alpha'}_{lm,l'm'}$, where

$$V^{i\alpha,i'\alpha'}_{lm,l'm'} = \int d\boldsymbol{r}\ \overline{\psi_{lm}}(\boldsymbol{r}-\boldsymbol{R}_\alpha-\boldsymbol{T}_i)v(\boldsymbol{r}-\boldsymbol{R}_\alpha-\boldsymbol{T}_i)\psi_{l'm'}(\boldsymbol{r}-\boldsymbol{R}_{\alpha'}-\boldsymbol{T}_{i'})$$

is a Slater-Koster two-center integral (Appendix B). A generic Slater-Koster two-center integral can be expressed as a function of a few independent two-center integrals, shown in Fig. 4 for s, p, and d-functions. Apart from the σ bond, which is the strongest, other bonds are possible; the π bonds are made of orbitals which share a nodal plane to which the bond axis belongs, and the δ bond, for which two nodal planes intersect in the bond axis connecting the two ions. Fig. 5 shows how to obtain a generic two-center integral involving p and s orbitals.[5] Let us now consider the case of the e_g and t_{2g} bands of $KCuF_3$; here we assume for simplicity that the system is an ideal cubic perovskite, shown in Fig. 6. The primitive cell contains one formula unit (a single K cube in Fig. 1). The cubic axes are $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, and the lattice constant is a. A Cu atom at site $\boldsymbol{R}_i$ is surrounded by two apical F atoms, F_3 at $\boldsymbol{R}_i + \frac{1}{2}\boldsymbol{z}$ and F_6 at $\boldsymbol{R}_i - \frac{1}{2}\boldsymbol{z}$, and four planar F atoms, F_1 and F_4 at $\boldsymbol{R}_i \pm \frac{1}{2}\boldsymbol{x}$ and F_2 and F_5 at $\boldsymbol{R}_i \pm \frac{1}{2}\boldsymbol{y}$. In Fig. 7 one can see the effects of the cubic approximation on the e_g bands: the crystal-field splitting of the e_g states is zero, the band width slightly reduced, gaps disappear, and the dispersion relations is sizably modified. The cubic band structure in Fig. 7 was obtained with a unit cell containing two formula units, in order to compare it with the band structure of the experimental (Jahn-Teller distorted) structure of $KCuF_3$; hence we see four (instead of two) e_g bands. The band-structure of cubic $KCuF_3$ for

[5] More details on the tight-binding approach can be found in Ref. [13].

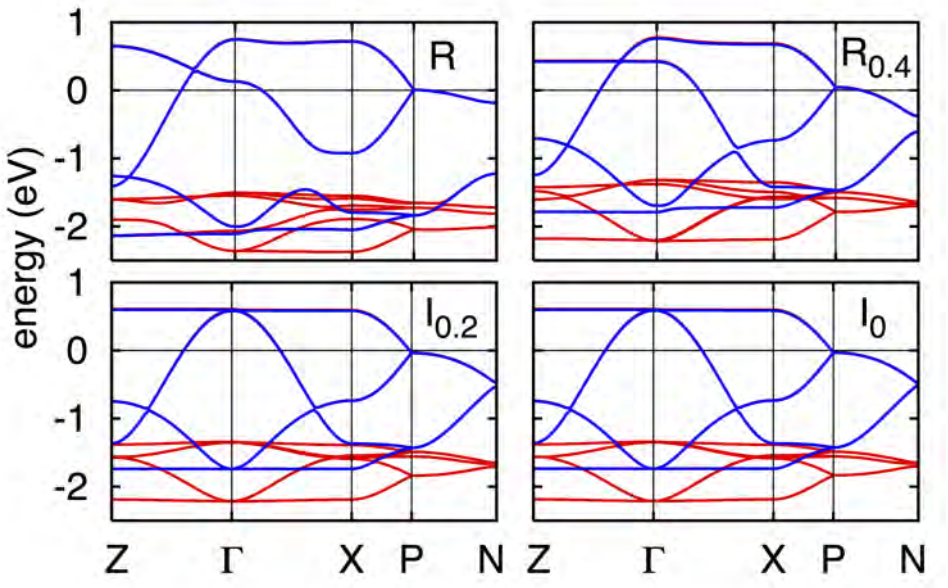


Fig. 7: *LDA e_g (blue) and t_{2g}(red) band structure of $KCuF_3$ for the experimental structure (R) and ideal structures with progressively reduced distortions (see Fig. 1). I_0: simple cubic. The unit cell used in this calculation contains two formula units. The figure is from Ref. [3].*

a cell with one formula unit is shown in Fig. 8; in the following we will refer for comparison to that figure only. Let us take as tight-binding basis the atomic $3d$ e_g orbitals for Cu and the $2p$ orbitals for F; we neglect for convenience the overlap integrals (i.e., we assume that our atomic functions are, approximately, localized *Wannier functions*). For such a tight-binding basis the only relevant Slater-Koster parameter is $V_{pd\sigma}$. The $|3z^2-r^2\rangle_i$ and $|x^2-y^2\rangle_i$ states of the Cu at $\boldsymbol{R}_i$ can couple via $V_{pd\sigma}$ to $|z^c\rangle_i$, the p_z orbitals of F_3 and F_6, to $|x^a\rangle_i$, the p_x orbitals of F_1 and F_4 and to $|y^b\rangle_i$, the p_y orbitals of F_2 and F_5. From the basis $|\alpha\rangle_i$ of localized atomic functions we construct the Bloch states $|\boldsymbol{k}\alpha\rangle = \frac{1}{\sqrt{N}}\sum_i e^{i\boldsymbol{k}\cdot\boldsymbol{R}_i}|\alpha\rangle_i$, and obtain the tight-binding Hamiltonian

$H^{\rm TB}_{e_g}$	$\lvert\boldsymbol{k}\ z^c\rangle$	$\lvert\boldsymbol{k}\ x^a\rangle$	$\lvert\boldsymbol{k}\ y^b\rangle$	$\lvert\boldsymbol{k}\ 3z^2-r^2\rangle$	$\lvert\boldsymbol{k}\ x^2-y^2\rangle$
$\lvert\boldsymbol{k}\ z^c\rangle$	ε_p	0	0	$-2V_{pd\sigma}s_z$	0
$\lvert\boldsymbol{k}\ x^a\rangle$	0	ε_p	0	$V_{pd\sigma}s_x$	$-\sqrt{3}V_{pd\sigma}s_x$
$\lvert\boldsymbol{k}\ y^b\rangle$	0	0	ε_p	$V_{pd\sigma}s_y$	$\sqrt{3}V_{pd\sigma}s_y$
$\lvert\boldsymbol{k}\ 3z^2-r^2\rangle$	$-2V_{pd\sigma}\overline{s}_z$	$V_{pd\sigma}\overline{s}_x$	$V_{pd\sigma}\overline{s}_y$	ε_d	0
$\lvert\boldsymbol{k}\ x^2-y^2\rangle$	0	$-\sqrt{3}V_{pd\sigma}\overline{s}_x$	$\sqrt{3}V_{pd\sigma}\overline{s}_y$	0	ε_d

(11)

where $s_\alpha = ie^{-ik_\alpha a/2}\sin k_\alpha a/2$, $\alpha = x, y, z$, $\varepsilon_p < \varepsilon_d = \varepsilon_p + \Delta_{pd}$, and $V_{pd\sigma} < 0$. If $|V_{pd\sigma}|/\Delta_{pd}$ is small, the occupied bands are F p-like, while the partially filled bands Cu e_g-like. We now calculate the bands along high-symmetry lines.[6] Along Γ-Z, the eigenvalues ε_i ($\varepsilon_i \le \varepsilon_{i+1}$) of

[6]Special points: Γ $= (0,0,0)$, Z$= (0,0,\pi/a)$, X$= (\pi/a,0,0)$, M$= (\pi/a,\pi/a,0)$, R$= (\pi/a,\pi/a,\pi/a)$.

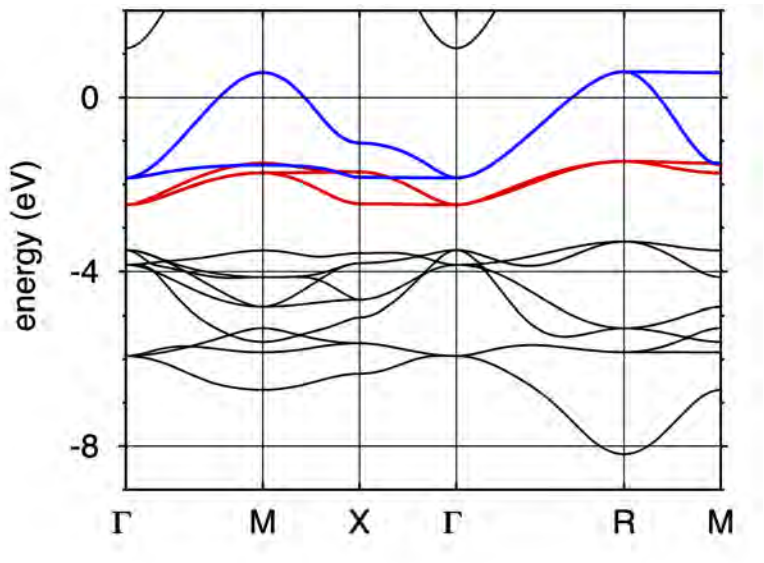


Fig. 8: *LDA band structure of cubic $KCuF_3$. The t_{2g} bands are in red and the e_g bands in blue.*

$H^{\rm TB}_{e_g}$ are

$$\begin{aligned}
\varepsilon_2 &= \varepsilon_p \\
\varepsilon_3 &= \varepsilon_p \\
\varepsilon_4 &= \varepsilon_d \\
\varepsilon_{1,5} &= \varepsilon_p + \tfrac{1}{2}\Delta_{pd} \pm \tfrac{1}{2}\sqrt{\Delta^2_{pd} + 16V^2_{pd\sigma}|s_z|^2}
\end{aligned}$$

where ε_1 (sign $-$) is bonding and F z-like, while ε_5 (sign $+$) anti-bonding and Cu $3z^2-r^2$-like. Along Γ-X, we have instead the dispersion relations

$$\begin{aligned}
\varepsilon_2 &= \varepsilon_p \\
\varepsilon_3 &= \varepsilon_p \\
\varepsilon_4 &= \varepsilon_d \\
\varepsilon_{1,5} &= \varepsilon_p + \tfrac{1}{2}\Delta_{pd} \pm \tfrac{1}{2}\sqrt{\Delta^2_{pd} + 16V^2_{pd\sigma}|s_x|^2}
\end{aligned}$$

where ε_1 is bonding and F x-like, while ε_5 anti-bonding and Cu x^2-y^2-like. To obtain the e_g-like bands, instead of diagonalizing $H^{\rm TB}_{e_g}$ as we have done above, we can also use the *down-folding* procedure, which, for non-interacting electrons, can be carried out exactly. This method works as follows. We divide the orbitals in passive (F p) and active (Cu d), and write the eigenvalues equation as

$$\begin{bmatrix} H_{pp} & H_{pd} \\ H_{dp} & H_{dd} \end{bmatrix} \begin{bmatrix} |\boldsymbol{k}\, p\rangle \\ |\boldsymbol{k}\, d\rangle \end{bmatrix} = \varepsilon \begin{bmatrix} I_{pp} & 0 \\ 0 & I_{dd} \end{bmatrix} \begin{bmatrix} |\boldsymbol{k}\, p\rangle \\ |\boldsymbol{k}\, d\rangle \end{bmatrix},$$

where H_{pp} (I_{pp}) is the Hamiltonian (identity matrix) in the p-electron space (3×3), and H_{dd} (I_{dd}) the Hamiltonian (identity matrix) in the d-electron space (2×2). By downfolding to the d sector we obtain the energy-dependent operator H^{ε}_{dd}, which acts in the d space only

$$H^{\varepsilon}_{dd} = H_{dd} - H_{dp}(H_{pp} - \varepsilon I_{pp})^{-1} H_{pd},$$

and a correspondingly transformed and energy-dependent basis set for the active space, $|\mathrm{k}\ d\rangle_\varepsilon$. The operator H^ε_{dd} has the same eigenvalues and eigenvectors as the original Hamiltonian. In the case of the e_g bands ($H^\varepsilon_{dd} = H^\varepsilon_{e_g}$) of $KCuF_3$

$H^\varepsilon_{e_g}$	$\lvert\boldsymbol{k}\ 3z^2{-}r^2\rangle_\varepsilon$	$\lvert\boldsymbol{k}\ x^2{-}y^2\rangle_\varepsilon$
$\lvert\boldsymbol{k}\ 3z^2{-}r^2\rangle_\varepsilon$	$\varepsilon'_d{-}2t^\sigma_\varepsilon[\frac{1}{4}(\cos k_x a{+}\cos k_y a){+}\cos k_z a]$	$2t^\sigma_\varepsilon[\frac{\sqrt{3}}{4}(\cos k_x a{-}\cos k_y a)]$
$\lvert\boldsymbol{k}\ x^2{-}y^2\rangle_\varepsilon$	$2t^\sigma_\varepsilon[\frac{\sqrt{3}}{4}(\cos k_x a{-}\cos k_y a)]$	$\varepsilon'_d{-}2t^\sigma_\varepsilon[\frac{3}{4}(\cos k_x a{+}\cos k_y a)]$

(12)

where the effective parameters are

$$t^\sigma_\varepsilon = \frac{V^2_{pd\sigma}}{\varepsilon - \varepsilon_p}, \quad \varepsilon'_d = \varepsilon_d + 3t^\sigma_\varepsilon .$$

The downfolding procedure has *renormalized* the parameters ε_d of the original model (11), but also introduced a new interaction: inter-orbital coupling. Furthermore, H^ε_{dd} and the Bloch basis are now energy dependent. Along ΓZ, the eigenvalues of (12) are given implicitly by the equations $\varepsilon{=}\varepsilon_d{+}2t^\sigma_\varepsilon{-}2t_\varepsilon\cos k_z a$ (band ε_5) and $\varepsilon{=}\varepsilon_d$ (band ε_4); in second-order perturbation theory we find

$$t^\sigma_\varepsilon \sim t^\sigma_{\varepsilon_d} = \frac{V^2_{pd\sigma}}{\Delta_{pd}},$$

$$\varepsilon_5 \sim \varepsilon_d + 2t^\sigma_{\varepsilon_d} - 2t^\sigma_{\varepsilon_d}\cos k_z a\,.$$

From Hamiltonian (12) it is relatively easy to see that the e_g bands are 2-fold degenerate along direction Γ-R, to find the dispersion along Γ-M and R-M, and to obtain the e_g-like bands in Fig. 8. By Fourier transforming the Bloch states $|\boldsymbol{k}\ 3z^2{-}r^2\rangle_\varepsilon$ and $|\boldsymbol{k}\ x^2{-}y^2\rangle_\varepsilon$ we can build a set of Wannier functions. They have $3z^2{-}r^2$ or $x^2{-}y^2$ symmetry as the atomic orbitals, and, additionally, they span, to arbitrary accuracy, the e_g bands. These Wannier functions are by construction longer range than atomic orbitals, since they have p tails on the downfolded neighboring F sites.

We can now repeat the same calculation for the t_{2g} bands. The minimal tight-binding basis is of course different with respect to the case of e_g bands. The states $|xy\rangle_i$ of the Cu ion located at $\boldsymbol{R}_i$ are coupled via $V_{pd\pi}$ to the $|y^a\rangle_i$, the p_y orbitals of F_1 and F_4 and to $|x^b\rangle_i$, the p_x orbitals of F_2 and F_5; in a similar way, $|xz\rangle_i$ is coupled via $V_{pd\pi}$ to the $|z^a\rangle_i$, the p_z orbitals of F_1 and F_4, and to the $|x^c\rangle_i$, the p_x orbitals of F_3 and F_6; finally $|yz\rangle_i$ is coupled via $V_{pd\pi}$ to the $|z^b\rangle_i$, the p_z orbitals of F_2 and F_5, and to the $|y^c\rangle_i$, the p_y orbitals of F_3 and F_6. After constructing for each $|\alpha\rangle_i$ the corresponding Bloch state, we obtain the tight-binding Hamiltonian. The latter splits into three decoupled blocks,

$H^{\mathrm{TB}}_{t_{2g}}$	$\lvert\boldsymbol{k}\ y^a\rangle$	$\lvert\boldsymbol{k}\ x^b\rangle$	$\lvert\boldsymbol{k}\ xy\rangle$
$\lvert\boldsymbol{k}\ y^a\rangle$	ε_p	0	$2V_{pd\pi}s_x$
$\lvert\boldsymbol{k}\ x^b\rangle$	0	ε_p	$2V_{pd\pi}s_y$
$\lvert\boldsymbol{k}\ xy\rangle$	$2V_{pd\pi}\overline{s}_x$	$2V_{pd\pi}\overline{s}_y$	ε_d

and cyclic permutations of x, y, and z (and, correspondingly, of a, b, and c). In the Γ-X direction we thus find

$$\begin{aligned}\varepsilon_{2'}(\boldsymbol{k}) =& \varepsilon_d \\ \varepsilon_5(\boldsymbol{k}) =& \varepsilon_p + \frac{\Delta_{pd}}{2} + \frac{\sqrt{\Delta_{pd}^2 + 16 V_{pd\pi}^2 |s_x|^2}}{2} \\ \sim& \varepsilon_d + 2t^\pi_{\varepsilon_d} - 2t^\pi_{\varepsilon_d} \cos k_x a\end{aligned}$$

where $t^\pi_{\varepsilon_d} = V^2_{pd\pi}/\Delta_{pd}$. By downfolding the oxygen states we obtain

$H^\varepsilon_{t_{2g}}$	$\lvert \boldsymbol{k}\ yz\rangle_\varepsilon$	$\lvert \boldsymbol{k}\ xz\rangle_\varepsilon$	$\lvert \boldsymbol{k}\ xy\rangle_\varepsilon$
$\lvert \boldsymbol{k}\ yz\rangle_\varepsilon$	$\varepsilon''_d - 2t^\pi_\varepsilon\,(\cos k_x a + \cos k_y a)$	0	0
$\lvert \boldsymbol{k}\ xz\rangle_\varepsilon$	0	$\varepsilon''_d - 2t^\pi_\varepsilon\,(\cos k_x a + \cos k_z a)$	0
$\lvert \boldsymbol{k}\ yz\rangle_\varepsilon$	0	0	$\varepsilon''_d - 2t^\pi_\varepsilon\,(\cos k_y a + \cos k_z a)$

where the parameters in the matrix are

$$\begin{aligned}\varepsilon''_d =& \varepsilon_d + 4t^\pi_\varepsilon, \\ t^\pi_\varepsilon =& \frac{|V_{pd\pi}|^2}{\varepsilon - \varepsilon_p}.\end{aligned}$$

As in the case of the e_g bands, we find renormalized energy levels and effective band dispersions; since different Cu t_{2g} states couple to different F p states, and we neglected hopping integral between oxygens, the xy, xz, and yz bands are totally decoupled in our model. We are now in the position of calculating the (approximate) expression of the covalent contribution to the e_g-t_{2g} crystal-field splitting, i.e., the energy difference

$$\Delta_{\rm CF} \sim \varepsilon'_d - \varepsilon''_d = 3\frac{|V_{pd\sigma}|^2}{\Delta_{pd}} - 4\frac{|V_{pd\pi}|^2}{\Delta_{pd}} > 0. \tag{13}$$

As we can see, the sign of the covalent crystal-field splitting is the same as that of the ionic contribution. This happens for two reasons. First, the so-called d bands are the anti-bonding states of the p-d Hamiltonian, hence both the energy of the e_g and t_{2g} states moves upwards due to the interaction with the p orbitals. Second, σ bonds are stronger than π bonds, hence e_g states shift to sizably higher energy than t_{2g} states.

The tight-binding model we have used so far is oversimplified, but it already qualitatively well describes the e_g and t_{2g} bands in Fig. 8. A more accurate description can be obtained including other Slater-Koster integrals, such as the hopping to apical F s states, or between neighboring F p states. With increasing number of parameters, it becomes progressively harder to estimate them, e.g., from comparison with experiments; furthermore a large number of fitting parameters makes it impossible to put a theory to a test. Modern techniques allow us, however, to calculate hopping integrals and crystal-field splittings *ab-initio*, using localized Wannier functions as the basis and the Kohn-Sham potential $v_R(\boldsymbol{r})$ as the one-electron potential; because Wannier functions are orthogonal, the corresponding overlap matrix is by construction diagonal.

4 Jahn-Teller effect

In order to introduce the Jahn-Teller effect we have to take a step backwards and start from the central equation of solid-state physics, the eigenvalue problem $\hat{H}\Psi = E\Psi$, defined (in the non-relativistic limit) by the many-body Hamiltonian

$$\hat{H} = \underbrace{-\frac{1}{2}\sum_i \nabla_i^2}_{\hat{T}_e} + \underbrace{\frac{1}{2}\sum_{i\neq i'} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|}}_{\hat{V}_{ee}} \underbrace{-\sum_{i\alpha} \frac{Z_\alpha}{|\boldsymbol{r}_i - \boldsymbol{R}_\alpha|}}_{\hat{V}_{en}} \underbrace{-\sum_\alpha \frac{1}{2M_\alpha}\nabla_\alpha^2}_{\hat{T}_n} + \underbrace{\frac{1}{2}\sum_{\alpha\neq\alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_{\alpha'}|}}_{\hat{V}_{nn}} .$$

Here $\{\boldsymbol{r}_i\}$ are the coordinates of the N_e electrons, $\{\boldsymbol{R}_\alpha\}$ those of the N_n nuclei, Z_α the atomic numbers, and M_α the nuclear masses. The Born-Oppenheimer Ansatz

$$\Psi(\{\boldsymbol{r}_i\}, \{\boldsymbol{R}_\alpha\}) = \psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\})\, \Phi(\{\boldsymbol{R}_\alpha\}) , \tag{14}$$

splits the Schrödinger equation $\hat{H}\Psi = E\Psi$ into the system

$$\begin{cases} \hat{H}_e\psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\}) &= \varepsilon(\{\boldsymbol{R}_\alpha\})\psi(\{\boldsymbol{r}_i\}; \{\boldsymbol{R}_\alpha\}), \\ \\ \hat{H}_n\Phi(\{\boldsymbol{R}_\alpha\}) &= E\Phi(\{\boldsymbol{R}_\alpha\}), \end{cases} \tag{15}$$

where the Hamilton operators for the electrons ($\hat{H}_e$) and that for the lattice ($\hat{H}_n$) are

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}, \tag{16}$$

$$\hat{H}_n = \hat{T}_n + \varepsilon(\{\boldsymbol{R}_\alpha\}) = \hat{T}_n + \hat{U}_n, \tag{17}$$

and where in (17) we neglect non-adiabatic corrections.[7] In the electronic Hamiltonian (16) the atomic positions $\{\boldsymbol{R}_\alpha\}$ are simple parameters. The electronic eigenvalue $\varepsilon(\{\boldsymbol{R}_\alpha\})$ acts as potential for the nuclei and defines a Born-Oppenheimer (BO) energy surface. While (16) describes the electronic structure, (17) yields the equilibrium crystal structure and the vibrational modes. These equations are impossible to solve in the general case. The first difficulty is that Hamiltonian (16) describes the electronic quantum many-body problem. The latter can be solved only approximately, for example the energy of the ground state can be obtained via density-functional theory using one of the known approximations to the universal functional. For strongly-correlated systems, advanced methods combine density-functional theory with many-body approaches such as the dynamical mean-field theory [7, 8]. The second issue is the very high number of atoms, and therefore of $\{\boldsymbol{R}_\alpha\}$ parameters to explore; finally, even if we solve the electronic many-body problem exactly, we still have to deal with the nuclear many-body problem, Hamiltonian (17). Despite all these obstacles, let us assume for a moment that, for a given system, we did solve the electronic problem for general values of $\{\boldsymbol{R}_\alpha\}$. Let us also assume that the set of positions $\{\boldsymbol{R}_\alpha\} = \{\boldsymbol{R}_\alpha^0\}$ defines a specific crystal structure, whose

[7] We neglect the operator $\hat{\Lambda}_n$, with elements $\langle m|\hat{\Lambda}_n|m'\rangle = -\sum_\alpha \frac{1}{M_\alpha}\left[\frac{1}{2}\langle\psi_m|\nabla_\alpha^2\psi_{m'}\rangle + \langle\psi_m|\nabla_\alpha\psi_{m'}\rangle\cdot\nabla_\alpha\right]$

electronic ground state (i.e., the lower energy BO surface) has degeneracy $d > 1$. We can at this point ask ourself the question: Is structure $\{\boldsymbol{R}^0_\alpha\}$ actually stable?
The Jahn-Teller theorem states that any electronically degenerate system can lower its energy by undergoing some structural distortions, and therefore is unstable.[8] This is due to the coupling between electrons and lattice. In order to better understand the microscopic origin of this phenomenon, let us consider a system in a high-symmetry structure, $\{\boldsymbol{R}^0_\alpha\}$, for which the electronic ground state has energy $\varepsilon(\{\boldsymbol{R}^0_\alpha\})$ with degeneracy $d > 1$. This means that there are d Born-Oppenheimer surfaces degenerate for $\{\boldsymbol{R}_\alpha\} = \{\boldsymbol{R}^0_\alpha\}$,

$$\varepsilon_m(\{\boldsymbol{R}^0_\alpha\}) = \varepsilon(\{\boldsymbol{R}^0_\alpha\}).$$

In the rest of the chapter we will take $\varepsilon(\{\boldsymbol{R}^0_\alpha\})$ as the energy zero. The corresponding degenerate electronic wavefunctions are $\psi_m(\{\boldsymbol{r}_i\};\{\boldsymbol{R}^0_\alpha\})$. Let us expand the nuclear potential $\hat{U}_n$ for one of these surfaces around the symmetric structure $\{\boldsymbol{R}^0_\alpha\}$. This leads to the Taylor series

$$\hat{H}_n = \hat{T}_n + \sum_{\alpha\mu}\left[\frac{\partial \hat{U}_n}{\partial u_{\alpha\mu}}\right]_{\{\boldsymbol{R}^0_\alpha\}} u_{\alpha\mu} + \frac{1}{2}\sum_{\alpha\mu}\sum_{\alpha'\mu'}\left[\frac{\partial^2 \hat{U}_n}{\partial_{\alpha\mu}\partial_{\alpha'\mu'}}\right]_{\{\boldsymbol{R}^0_\alpha\}} u_{\alpha\mu}\, u_{\alpha'\mu'} + \ldots,$$

where $\boldsymbol{u}_\alpha = \boldsymbol{R}_\alpha - \boldsymbol{R}^0_\alpha$ are displacement vectors with respect to the equilibrium position, and $\mu = x, y, z$. If $\{\boldsymbol{R}^0_\alpha\}$ is an equilibrium structure, the gradient is zero and

$$\hat{H}_n \sim \hat{T}_n + \frac{1}{2}\sum_{\alpha\mu}\sum_{\alpha'\mu'}\left[\frac{\partial^2 \hat{U}_n}{\partial_{\alpha\mu}\partial_{\alpha'\mu'}}\right]_{\{\boldsymbol{R}^0_\alpha\}} u_{\alpha\mu}u_{\alpha'\mu'} + \cdots = \hat{T}_n + \hat{U}^{\rm PH}_n(\{\boldsymbol{R}^0_\alpha\}) + \ldots, \qquad (18)$$

The standard procedure to diagonalize (18) consists of two steps. First we change coordinates

$$\tilde{u}_{\alpha\mu} = u_{\alpha\mu}\sqrt{M_\alpha}.$$

Second we introduce the dynamical matrix

$$D_{\alpha\mu,\alpha\mu'} = \frac{1}{\sqrt{M_\alpha}}\frac{1}{\sqrt{M_{\alpha'}}}\left[\frac{\partial^2 \hat{U}_n}{\partial_{\alpha\mu}\partial_{\alpha'\mu'}}\right]_{\{\boldsymbol{R}^0_\alpha\}},$$

and diagonalize it. Its N_m eigenvectors are the normal modes $\boldsymbol{Q}_\eta$,

$$D\boldsymbol{Q}_\eta = \omega^2_\eta \boldsymbol{Q}_\eta,$$

$$Q_{\eta\nu} = \sum_{\alpha=1}^{N_n}\sum_{\mu=x,y,z} a_{\eta\nu,\alpha\mu}u_{\alpha\mu},$$

with $\eta = 1, \ldots N_m$, and $\nu = x, y, z$. The *normal coordinates* $\{Q_{n\nu}\}$, together with the associated canonically-conjugated momenta $\{P_{n\nu}\}$, bring (18) in the form

$$\hat{H}_n \sim \frac{1}{2}\sum_{\eta\nu}\left[P^2_{\eta\nu} + \omega^2_\eta Q^2_{\eta\nu}\right]. \qquad (19)$$

[8] The only exceptions are linear molecules and Kramers degeneracy.

In a crystal, this Hamiltonian yields the phonon energy levels. Let us now determine the possible N_m normal modes for a cubic perovskite. For simplicity we consider here only a single octahedron and the modes associated with the vibrations of its atoms. Given that each atom can move in three directions, and there are 6 atoms of type C and 1 atom of type B, in principle such a system has 21 degrees of freedom; eliminating global translations (3 degrees of freedom) and global rotations (3 degrees of freedom), i.e., displacements which are not vibrations, 15 degrees of freedom are left, hence the system has 15 possible normal modes. In group theory language, assuming again for simplicity that the group is O instead of O_h, one can show that these modes can be labeled as belonging to irreducible representations A_1, E, T_1 or T_2. To obtain this result we first build a matrix representation of the group in the linear space of all possible displacements; this space is 21-dimensional, and so is the associated matrix representation $\Gamma_{\rm tot}$. The latter can be expressed as the direct product $\Gamma_{\rm tot} = \Gamma_{\rm a.s.} \otimes \Gamma_{\rm vector}$, where $\Gamma_{\rm a.s.}$ is the so-called atomic-site representation. $\Gamma_{\rm a.s.}$ has as a basis the original atomic positions (without displacements); in our case it is has therefore dimensionality 7. The character of $\Gamma_{\rm a.s.}$ for a given operation is simply the number of sites left invariant by that operation. Finally, in group O the irreducible representation for a vector is $\Gamma_{vector} = T_1$; this can be seen from the partner functions (x, y, z) in Table 4. Summarizing all this in a character table, we have

O	E	$8C_3$	$3C_2$	$6C_2$	$6C_4$
$\Gamma^{\rm a.s.}$	7	1	3	1	3
$\Gamma^{\rm tot} = \Gamma^{\rm a.s.} \otimes \Gamma_{\rm vector}$	21	0	-3	-1	3

Once we know the characters for representation $\Gamma_{\rm tot}$, we can split the latter into irreducible representations of group O via the decomposition formula Eq. (5). After subtracting (tensor subtraction $\ominus$) the representations for mere translations (T_1) and mere rotations (T_1) of the octahedron,[9] we arrive at the final decomposition of the vibrational-modes representation $\Gamma_{\rm vibrations} = \Gamma_{\rm tot} \ominus \Gamma_{\rm vector} \ominus \Gamma_{\rm rotation} = A_1 \oplus E \oplus 2T_1 \oplus 2T_2$. Normal modes which are a basis for different irreducible representations have in general different energies. Let us focus on modes A_1 and E. We can obtain mode A_1 by using the projector, Eq. (6), for irreducible representation A_1. As a matter of fact, if we assume that atom F_1 (Fig. 9) is displaced by $\boldsymbol{u}_1$, by applying the projector $\hat{P}^{A_1}$ to $\boldsymbol{u}_1$ we generate automatically the linear combination of atomic displacements (all having the same length) forming the mode of symmetry A_1. This leads to

$$\boldsymbol{Q}_0 = \boldsymbol{u}_1(q_0) + \boldsymbol{u}_2(q_0) + \boldsymbol{u}_3(q_0) + \boldsymbol{u}_4(q_0) + \boldsymbol{u}_5(q_0) + \boldsymbol{u}_6(q_0).$$

[9] The representation for an improper vector (rotation) is $\Gamma_{rotation} = T_1$, as can be seen from the corresponding partner functions (R_x, R_y, R_z) in Table 4.

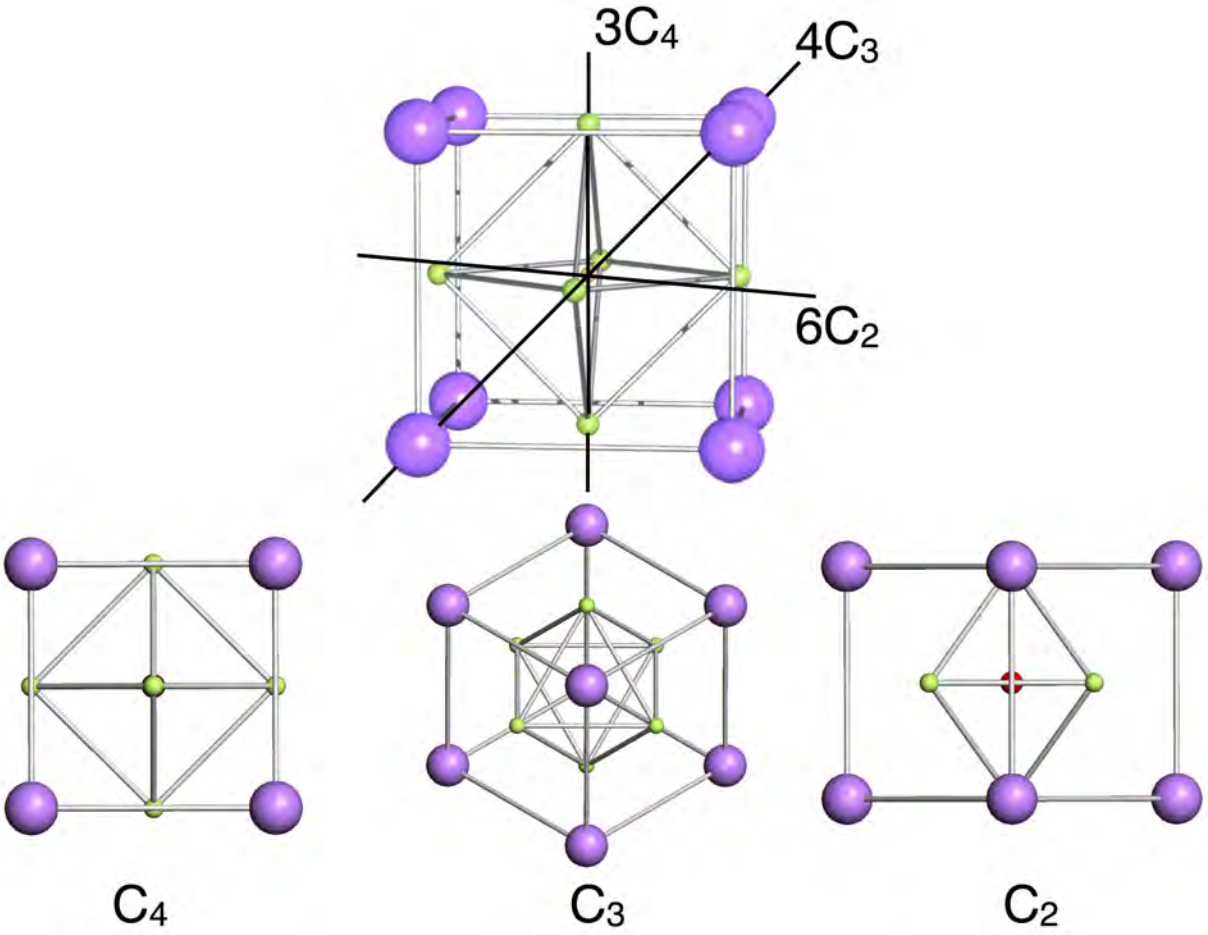


Fig. 9: *Unit cell (top) and vibrational modes Q_0, Q_1, and Q_2 of cubic $KCuF_3$.*

Here $\boldsymbol{u}_i$ are the (normalized) displacements for the C_i atom (see Fig. 9) which we rewrite as

$$\begin{aligned}
\boldsymbol{u}_1(q_0) &= \tfrac{1}{\sqrt{6}} q_0 (1,0,0) \\
\boldsymbol{u}_2(q_0) &= \tfrac{1}{\sqrt{6}} q_0 (0,1,0) \\
\boldsymbol{u}_3(q_0) &= \tfrac{1}{\sqrt{6}} q_0 (0,0,1) \\
\boldsymbol{u}_4(q_0) &= -\tfrac{1}{\sqrt{6}} q_0 (1,0,0) \\
\boldsymbol{u}_5(q_0) &= -\tfrac{1}{\sqrt{6}} q_0 (0,1,0) \\
\boldsymbol{u}_6(q_0) &= -\tfrac{1}{\sqrt{6}} q_0 (0,0,1)
\end{aligned}$$

The potential energy of such a *breathing mode* is

$$U_n^{\rm PH} = \frac{1}{2} C_{A_1} q_0^2 \,.$$

The $\boldsymbol{Q}_0$ mode expands or compresses the unit cell, but does not change its symmetry which remains cubic. Hence, this mode has no influence on the stability of the structure, at most it can affect the actual value of the lattice constant. More interesting are the two degenerate modes of type E. These modes can be obtained in a similar way as we have done for $\boldsymbol{Q}_0$, this time using the projector for irreducible representation E; within the resulting 2-dimensional space,

we choose as basis the mutually orthogonal modes that transform as the $l = 2$ partner functions of E, x^2-y^2 and $3z^2-r^2$. These are $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, shown in Fig. 9. They are defined as

$$\boldsymbol{Q}_1 = \boldsymbol{u}_1(q_1) + \boldsymbol{u}_2(q_1) + \boldsymbol{u}_4(q_1) + \boldsymbol{u}_5(q_1),$$
$$\boldsymbol{Q}_2 = \boldsymbol{u}_1(q_2) + \boldsymbol{u}_2(q_2) + \boldsymbol{u}_3(q_2) + \boldsymbol{u}_4(q_2) + \boldsymbol{u}_5(q_2) + \boldsymbol{u}_6(q_2),$$

where the displacements are

$$\begin{array}{rclrcl}
\boldsymbol{u}_1(q_1) &=& \frac{1}{\sqrt{4}}q_1(1,0,0) & \boldsymbol{u}_1(q_2) &=& -\frac{1}{\sqrt{12}}q_2(1,0,0)\\
\boldsymbol{u}_2(q_1) &=& -\frac{1}{\sqrt{4}}q_1(0,1,0) & \boldsymbol{u}_2(q_2) &=& -\frac{1}{\sqrt{12}}q_2(0,1,0)\\
\boldsymbol{u}_3(q_1) &=& (0,0,0) & \boldsymbol{u}_3(q_2) &=& \frac{2}{\sqrt{12}}q_2(0,0,1)\\
\boldsymbol{u}_4(q_1) &=& -\frac{1}{\sqrt{4}}q_1(1,0,0) & \boldsymbol{u}_4(q_2) &=& \frac{1}{\sqrt{12}}q_2(1,0,0)\\
\boldsymbol{u}_5(q_1) &=& \frac{1}{\sqrt{4}}q_1(0,1,0) & \boldsymbol{u}_5(q_2) &=& \frac{1}{\sqrt{12}}q_2(0,1,0)\\
\boldsymbol{u}_6(q_1) &=& (0,0,0) & \boldsymbol{u}_6(q_2) &=& -\frac{2}{\sqrt{12}}q_2(0,0,1)
\end{array}$$

The corresponding quadratic potential has the form

$$\hat{U}_n^{\mathrm{PH}} = \frac{1}{2}C_E(q_1^2 + q_2^2).$$

The normal modes T_1 and T_2 can be obtained in a similar way; since they are not relevant for structure stability in the example considered here we do not provide their form explicitly.
Up to now we have assumed that the hypothetical high-symmetry structure $\{\boldsymbol{R}_\alpha^0\}$ is a stationary point. In general, however, this might or might not be true. The behavior of the BO energy surfaces close to the point in which they are degenerate allows us to separate them into two classes, the first one in which $\{\boldsymbol{R}_\alpha^0\}$ is a stationary point for all degenerate electronic states m (Renner-Teller intersection), and the second in which the surface is not a stationary point at least for some of the surfaces (Jahn-Teller intersection). The classical Jahn-Teller systems are those for which $\nabla\hat{U}_n(\{\boldsymbol{R}_\alpha^0\}) \neq 0$ at least in some direction (see, e.g., Fig. 10). Let us now calculate the first-order correction to the m degenerate eigenvalues due to a small distortion around $\{\boldsymbol{R}_\alpha^0\}$. The electronic Hamiltonian (16) has matrix elements

$$\langle\psi_m|\hat{H}_e(\{\boldsymbol{R}_\alpha\})|\psi_{m'}\rangle = \underbrace{\sum_{\alpha\mu}\langle\psi_m|\left[\frac{\partial\hat{H}_e}{\partial u_{\alpha\mu}}\right]_{\{\boldsymbol{R}_\alpha^0\}}|\psi_{m'}\rangle u_{\alpha\mu}}_{\hat{U}^{\mathrm{JT}}_{m,m'}} + \cdots = \hat{U}^{\mathrm{JT}}_{m,m'} + \ldots.$$

The perturbation $\hat{U}^{\mathrm{JT}}$, the Jahn-Teller potential, couples the degenerate BO energy surfaces; it also couples electrons and lattice vibrations, as we can see from the coordinates $u_{\alpha\mu}$ appearing in the expression above. Thus, if there are modes for which $\hat{U}^{\mathrm{JT}} \neq C\,\hat{I}$ where $\hat{I}$ is the identity matrix and C a constant, the system gains energy at linear order via a distortion which lowers the symmetry; the Jahn-Teller theorem states that such modes always exist for electronically degenerate systems (with the exceptions of Kramers degeneracy and linear molecules).

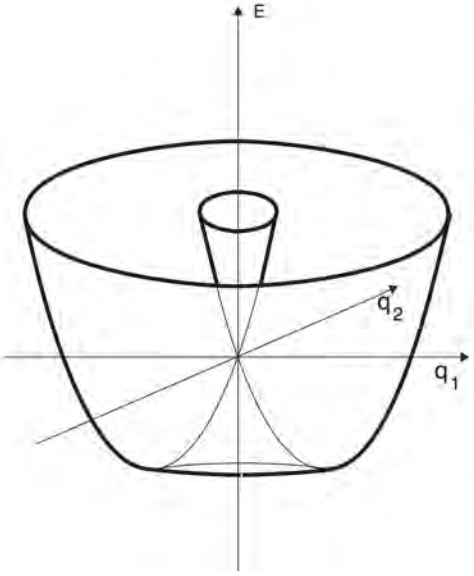


Fig. 10: *Born-Oppenheimer potential-energy surface exhibiting the form of a mexican hat. The slope of the curve at small distortions q_1, q_2 yields the Jahn-Teller coupling constant λ.*

In order to better understand the effect of the electron-lattice coupling, we generalize the Born-Oppenheimer Ansatz as follows

$$\Psi(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_\alpha\}) \;=\; \sum_m \psi_m(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_\alpha\})\,\Phi_m(\{\boldsymbol{R}_\alpha\}).$$

To find the equations for the functions $\{\Phi_m\}$, we write the Schrödinger equation $H\Psi = E\Psi$, multiply on the left by $\overline{\psi_m}$, and integrate over the coordinates of the electrons. We obtain

$$\hat{H}_n\Phi_m(\{\boldsymbol{R}_\alpha\}) = \left[\hat{T}_n + \hat{U}_n^{\rm PH}\right]\Phi_m(\{\boldsymbol{R}_\alpha\}) + \sum_{m'}\hat{U}^{\rm JT}_{m,m'}\Phi_{m'}(\{\boldsymbol{R}_\alpha\}) = E\Phi_m(\{\boldsymbol{R}_\alpha\}). \qquad (20)$$

The dynamics of the system close to the degeneracy point is determined by all degenerate sheets. The minimization of the new potential energy yields a new structure $\{\tilde{\boldsymbol{R}}^0_\alpha\}$ in which the electronic states are not any more degenerate. The modes that can produce such an instability should satisfy the condition

$$A_1 \in ([\Gamma_m \otimes \Gamma_m] \otimes (\Gamma_{\rm vibrations} \neq A_1)),$$

where Γ_m is the irreducible representation to which the electronic degenerate states belong, and $[\Gamma_m \otimes \Gamma_m]$ is the symmetric direct product; for e_g states, $[e_g \otimes e_g]{=}a_1 \oplus e_g$. The trivial representation A_1 has to be excluded from $\Gamma_{\rm vibrations}$ because, as already discussed, it does not lower the symmetry. In the case cubic $KCuF_3$, the relevant normal modes coupling to the degenerate e_g electronic states are thus the E modes; as for the electronic states, if the group $O \to O_h$, then $E \to E_g$. Thus we can say that $KCuF_3$ is an example of a $e_g \otimes E_g$ Jahn-Teller system, a system in which an electronic doublet (e_g) is coupled to a doublet of normal modes (E_g). The form of the Jahn-Teller potential $\hat{U}^{\rm JT}$ can be obtained from the effect of perturbations of type $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ on the crystal-field matrix. As for the crystal field, there are both a ionic and

a covalent contribution. For the ionic contribution, we can use once more perturbation theory. In this case, we have to take into account that the Cu-F distance d_C depends on the direction, i.e,

$$d_C \to d_C + \delta d_C^{\mu},$$

where $\mu = x, y, z$; the specific δd_C^{μ} values for each atom are given by the specific vibrational mode. After summing up all contribution, the first non-cubic correction due to E_g modes is

$$\Delta v_{\rm JT} = \lambda \begin{pmatrix} q_2 & q_1 \\ q_1 & -q_2 \end{pmatrix}.$$

It is, at this point, useful to introduce pseudo-spin operators acting on the e_g states, i.e., operators $\hat{\tau}_\mu$ with $\mu = x, y, z$ and

$$\hat{\tau}_z|\searrow\rangle = -|\searrow\rangle, \qquad \hat{\tau}_x|\searrow\rangle = +|\nearrow\rangle, \qquad \hat{\tau}_y|\searrow\rangle = -i|\nearrow\rangle$$

$$\hat{\tau}_z|\nearrow\rangle = +|\nearrow\rangle, \qquad \hat{\tau}_x|\nearrow\rangle = +|\searrow\rangle, \qquad \hat{\tau}_y|\nearrow\rangle = +i|\searrow\rangle$$

where $|\nearrow\rangle = |x^2 - y^2\rangle$ and $|\searrow\rangle = |3z^2 - r^2\rangle$. In matrix form these operators can be written as pseudo-Pauli matrices

$$\hat{\tau}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \hat{\tau}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \hat{\tau}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{21}$$

We can then rewrite the Jahn-Teller potential as

$$\Delta v_{\rm JT} = \lambda \Big(q_1 \hat{\tau}_x + q_2 \hat{\tau}_z \Big),$$

where $\lambda \sim (q_C/d_C^4)\ (36/7\sqrt{3}) > 0$. This potential expresses both the essence of the Jahn-Teller theorem and its relation with orbital order; the systems gains energy at linear order by making a distortion; the latter produces a crystal-field splitting, which leads to preferential occupation of the lower energy level. For example, if $q_1 = 0$ and $q_2 < 0$ (tetragonal compression) the $3z^2 - r^2$ state is higher in energy. Let us now calculate the covalent contribution to the Jahn-Teller potential. In this case the linear-order correction is

$$\Delta\varepsilon_{lm,l'm'}(\mathbf{0}, \boldsymbol{R}_\alpha + \boldsymbol{u}) - \Delta\varepsilon_{lm,l'm'}(\mathbf{0}, \boldsymbol{R}_\alpha) \sim \nabla \Delta\varepsilon_{lm,l'm'}(\mathbf{0}, \boldsymbol{R}_\alpha) \cdot \boldsymbol{u}$$

For e_g-states we use for simplicity the following approximations[10]

$$\begin{aligned}
\Delta\varepsilon_{3z^2-r^2,3z^2-r^2} &\sim \left[n^2 - \frac{1}{2}(l^2{+}m^2) \right]^2 \tilde{V}_{dd\sigma}, \\
\Delta\varepsilon_{3z^2-r^2,x^2-y^2} &\sim \frac{\sqrt{3}}{2}(l^2{-}m^2) \left[n^2 - \frac{1}{2}(l^2{+}m^2) \right] \tilde{V}_{dd\sigma}, \\
\Delta\varepsilon_{x^2-y^2,x^2-y^2} &\sim \frac{3}{4}(l^2{-}m^2)^2 \tilde{V}_{dd\sigma}.
\end{aligned}$$

[10]The crystal-field integrals are also two-center integrals; the table of Slater-Koster integrals in Appendix B is thus still valid, provided that $V_{ll'\alpha}$ are replaced by the corresponding crystal-field terms, which we indicate as $\tilde{V}_{ll'\alpha}$.

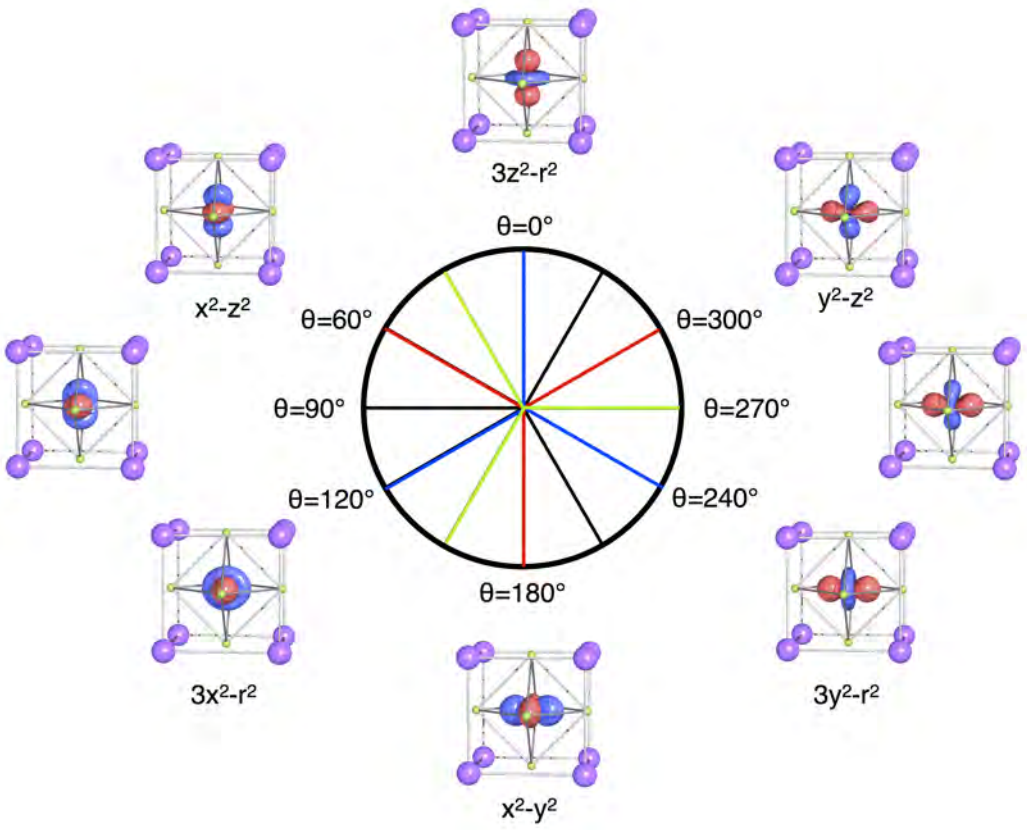


Fig. 11: *Linear combinations of e_g-states, $|\vartheta\rangle = -\sin\frac{\vartheta}{2}|x^2-y^2\rangle + \cos\frac{\vartheta}{2}|3z^2-r^2\rangle$. The $\vartheta = 0°$ orbital is the excited state in the presence of a tetragonal compression along the z axis, while $\vartheta = \pm 2\pi/3$ are excited states for a tetragonal compression along x or y. This three-fold degeneracy (rotation by $\pm 2\pi/3$) is due to cubic symmetry.*

By summing up the contributions from all C ions for each mode, we obtain

$$\Delta\varepsilon_{\rm JT}(q_1,q_2) = \lambda \begin{pmatrix} q_2 & q_1 \\ q_1 & -q_2 \end{pmatrix} = \lambda\Big(q_1\hat{\tau}_x + q_2\hat{\tau}_z\Big),$$

where $\lambda \sim -\frac{\sqrt{3}}{2}\tilde{V}'_{dd\sigma} > 0$. This is the same form of potential that we have obtained for the ionic contribution. Again, if $q_1 = 0$ and $q_2 < 0$ (tetragonal compression) the $3z^2-r^2$ is higher in energy. In conclusion, if we neglect the kinetic energy of the nuclei (limit $M_\alpha/m_e \to \infty$), the ground state of the system can be calculated by minimizing a potential energy of the form

$$\hat{U}(q_1,q_2) = \hat{U}^{\rm JT} + \hat{U}^{\rm PH}_n = \lambda \begin{pmatrix} q_2 & q_1 \\ q_1 & -q_2 \end{pmatrix} + \frac{1}{2}C_E\,(q_1^2+q_2^2)\,\hat{I}, \qquad (22)$$

where $\hat{I}$ is the 2×2 identity matrix. To find the minimum of (22), it is convenient to introduce polar coordinates, which we define as $q_2 = -q\cos\vartheta, q_1 = -q\sin\vartheta$, so that for $0 < \vartheta < \pi/2$ we have $q_1 \le 0$ (compression of $\hat{x}$ axis) and $q_2 \le 0$ (compression of $\hat{z}$ axis); this corresponds

to the distortion of the octahedron labeled with number 1 in Fig. 1. In these coordinates

$$\hat{U}^{\rm JT} = -\lambda q \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix}.$$

The diagonalization of matrix (22) yields two eigenvalues; the lower energy branch

$$E_-(q) = -\lambda q + \frac{C_E}{2} q^2$$

takes the form of a mexican hat, shown in Fig. 10. The minimum of $E_-(q)$ is obtained for $q = q_0 = \lambda/C$ and has value

$$E_{\rm JT} = -\lambda^2/2C_E;$$

the quantity $E_{\rm JT}$ is defined as the Jahn-Teller energy of the system. The electronic ground state can be written as

$$|\vartheta\rangle_G = -\sin\frac{\vartheta-\pi}{2}|x^2{-}y^2\rangle + \cos\frac{\vartheta-\pi}{2}|3z^2{-}r^2\rangle.$$

The excited state (hole orbital), with energy

$$E_+(q) = \lambda q + \frac{C_E}{2} q^2,$$

is then given by

$$|\vartheta\rangle_E = -\sin\frac{\vartheta}{2}|x^2{-}y^2\rangle + \cos\frac{\vartheta}{2}|3z^2{-}r^2\rangle.$$

The states $|\vartheta\rangle_E$ with different ϑ are shown in Fig. 11. In the simple model discussed so far, all states $|\vartheta\rangle_G$ have the same Jahn-Teller energy. Cubic symmetry, however, only requires states

$$|\vartheta\rangle,\ |\vartheta{+}2\pi/3\rangle,\ |\vartheta{-}2\pi/3\rangle$$

to be degenerate. The additional (accidental) degeneracy is removed when we take into account anharmonic terms, the lowest order of which has the form

$$U^{\rm anh}(q_1,q_2) = A\big(q_2^3 - 3q_2q_1^2\big) = Aq^3\big(\cos^3\vartheta - 3\cos\vartheta\sin^2\vartheta\big) = -Aq^3\cos 3\vartheta$$

and yields the tetragonal distortion as a ground state, with $\vartheta = 0, \pm 2\pi/3$ for positive A and with $\vartheta = \pi, \pi \pm 2\pi/3$ for negative A. Higher-order terms can make the $\boldsymbol{Q}_1$ Jahn-Teller distortion ($\vartheta = \pi/2, \pi/2 \pm 2\pi/3$) more stable [1]. For a periodic lattice, mode $\boldsymbol{Q}_1$ leads to a co-operative distortion where long and short bonds alternate in the x and y direction; in such a case, the hole orbital rotates by $\pi/2$ if we move from a Cu site to its Cu first-nearest neighbors in the $\boldsymbol{ab}$ plane. Let us now analyze the different electronic configurations that can occur in perovskites. For the electronic configuration $3d^1{=}3t_{2g}^1$, the procedure is as the one illustrated above, except that t_{2g} states are 3-fold degenerate and form π bonds, which are weaker, therefore the splitting introduced by the Jahn-Teller effect is smaller than for e_g states. In the case of electronic configurations $3d^n$ with $n > 1$, to determine if the ion is Jahn-Teller active one has to consider the degeneracies of the many-body state, including Coulomb repulsion. Weak Jahn-Teller states are $3d^1$ (Ti^{3+} in $LaTiO_3$) and $3d^2$ (V^{3+} in $LaVO_3$), as also $3t_{2g}^4$, $3t_{2g}^5$, $3t_{2g}^4e_g^2$, $3t_{2g}^5e_g^2$; strong Jahn-Teller configurations are, e.g., $3d^9$ (Cu^{2+} in $KCuF_3$) and $3t_{2g}^3e_g^1$ (Mn^{3+} in $LaMnO_3$); the configurations $3t_{2g}^3$ and $3t_{2g}^3e_g^2$ are not degenerate and therefore usually not Jahn-Teller active.

5 Kugel-Khomskii superexchange

Let us now start from a totally different perspective, from the Hubbard model for a cubic perovskite with partially filled e_g or t_{2g} shell. The Hamiltonian takes the form $\hat{H} = \hat{H}_0 + \hat{H}_T + \hat{H}_U$ where

$$\begin{aligned}
\hat{H}_0 =& \varepsilon_d \sum_i \sum_\sigma \sum_m \hat{n}_{im\sigma} \\
\hat{H}_T =& -\sum_{i\neq i'} \sum_\sigma \sum_{mm'} t^{i,i'}_{mm'}\, c^\dagger_{im\sigma} c_{im'\sigma} \\
\hat{H}_U =& U \sum_i \sum_m \hat{n}_{im\uparrow}\hat{n}_{im\downarrow} + \frac{1}{2}\sum_i \sum_{\sigma\sigma'} \sum_{m\neq m'} (U - 2J - J\delta_{\sigma,\sigma'})\, \hat{n}_{im\sigma}\hat{n}_{im'\sigma'} \\
& -J \sum_i \sum_{m\neq m'} \left[c^\dagger_{im\uparrow} c^\dagger_{im\downarrow} c_{im'\uparrow} c_{im'\downarrow} + c^\dagger_{im\uparrow} c_{im\downarrow} c^\dagger_{im'\downarrow} c_{im'\uparrow} \right],
\end{aligned}$$

and where m labels either the e_g or the t_{2g} orbitals. Kugel and Khomskii have shown that, in the large t/U limit, this Hamiltonian can be mapped onto an effective generalized superexchange Hamiltonian with an orbitally-ordered ground state. The actual general super-exchange Hamiltonian can be found in Ref. [20], for possible the e_g^n and t_{2g}^n configurations.
Here, to understand the origin of the orbital super-exchange interaction, let us simplify the problem and consider first a system with only two atoms ($i = A, B$) and two orbitals ($\varepsilon_d = \varepsilon_{e_g}$), and for which the hopping matrix is diagonal in the orbitals

$$\hat{H}_T = -t \sum_\sigma \sum_m \left[c^\dagger_{Am} c_{Bm} + c^\dagger_{Bm} c_{Am} \right].$$

Furthermore, let us simplify the Coulomb interaction and neglect the spin-flip and pair-hopping terms, so that

$$\hat{H}_U \longrightarrow \hat{H}'_U = U \sum_{i=AB} \sum_m \hat{n}_{im\uparrow}\hat{n}_{im\downarrow} + \frac{1}{2} \sum_{i=AB} \sum_{\sigma\sigma'} \sum_{m\neq m'} (U - 2J - J\delta_{\sigma,\sigma'})\, \hat{n}_{im\sigma}\hat{n}_{im'\sigma'}.$$

Finally, we assume that the systems has one electron per atom (quarter filling, e_g^1 configuration). In the $t = 0$ or atomic limit there are two types of possible states for this system, those in which each atom is occupied by one electron, $|1,1\rangle_\alpha$, and those in which one atom has two electrons and the other zero, $|2,0\rangle_{\alpha'}$. The 16 states of type $|1,1\rangle_\alpha$, all degenerate with energy $E_\alpha(1,1) = 2\varepsilon_{e_g}$, can be written as $c^\dagger_{Am_A\sigma_A} c^\dagger_{Bm_B\sigma_B}|0\rangle$ with $\alpha = (m_A\sigma_A, m_B\sigma_B)$; here $m_i\sigma_i$ are the quantum numbers for the electron at site $i = A, B$. There are 12 states $|2,0\rangle_\alpha$ with one atom occupied by two electrons; they are listed below together with their energies

$\|2,0\rangle_{\alpha'}$			$E_{\alpha'}(2,0)$	
$\|2,0\rangle_{i1m}$	$=$	$c^\dagger_{im\uparrow} c^\dagger_{im\downarrow}\|0\rangle$	$2\varepsilon_{e_g} + U$	
$\|2,0\rangle_{i2m}$	$=$	$c^\dagger_{im\uparrow} c^\dagger_{im'\downarrow}\|0\rangle$	$2\varepsilon_{e_g} + U - 2J$	$m' \neq m$
$\|2,0\rangle_{i3\sigma}$	$=$	$c^\dagger_{im\sigma} c^\dagger_{im'\sigma}\|0\rangle$	$2\varepsilon_{e_g} + U - 3J$	$m' \neq m$

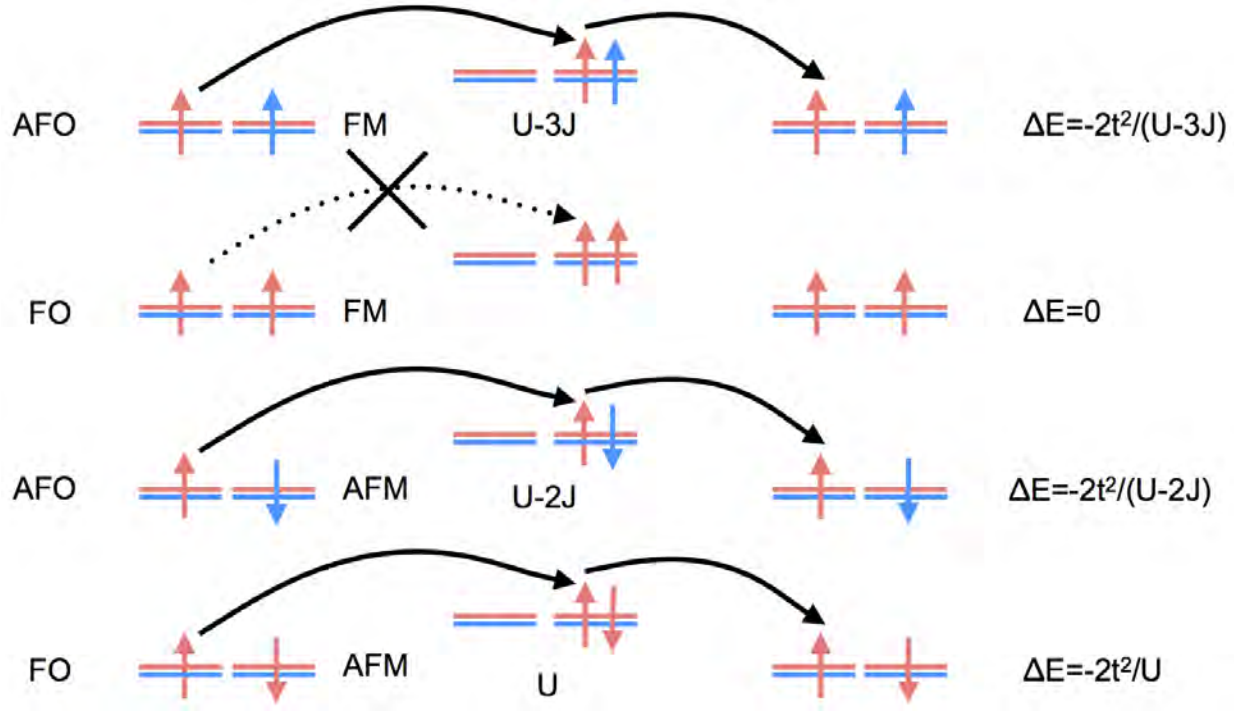


Fig. 12: *Superexchange energy gain for possible quarter-filling ground states of a two-site 2-fold degenerate Hubbard model with orbital- and spin-diagonal hopping matrices.*

The Coulomb repulsion U is positive and J is small with respect to U; therefore the $|1,1\rangle_\alpha$ states define the ground-state manifold. If t is finite but small ($t/U \ll 1$), we can treat $\hat{H}_T$ as a perturbation, and calculate the second-order correction to the energy of states $|1,1\rangle_\alpha$. This correction is always negative (energy gain) and it is given by the matrix

$$\Delta E_{\alpha_1,\alpha_2}(1,1) = -\sum_{\alpha'} {}_{\alpha_1}\langle 1,1|\hat{H}_T|2,0\rangle_{\alpha'} \frac{1}{E_{\alpha'}(2,0) - E_\alpha(1,1)} {}_{\alpha'}\langle 2,0|\hat{H}_T|1,1\rangle_{\alpha_2}$$

There are four interesting cases, depicted in Fig. 12. The first is the ferro-magnetic (same spin) and antiferro-orbital (different orbitals) state, first line of the figure. The corresponding second order energy gain ($\alpha_1 = \alpha_2 = m\sigma, m'\sigma$) is

$$\Delta E_{\alpha_1,\alpha_1}(1,1) = -\frac{2t^2}{U - 3J}.$$

For the ferro-magnetic (same spin) and ferro-orbital (same orbital) state (second line in the figure, $\alpha_1 = \alpha_2 = m\sigma, m\sigma$) the energy gain is, instead, zero

$$\Delta E_{\alpha_1,\alpha_1}(1,1) = 0.$$

The reason is that no hopping is possible due to the Pauli principle. For the antiferro-magnetic antiferro-orbital state (third line, $\alpha_1 = \alpha_2 = m\sigma, m' - \sigma$), we have

$$\Delta E_{\alpha_1,\alpha_1}(1,1) = -\frac{2t^2}{U - 2J},$$

and finally for the antiferro-magnetic ferro-orbital state ($\alpha_1 = \alpha_2 = m\sigma, m-\sigma$) we find

$$\Delta E_{\alpha_1,\alpha_1}(1,1) = -\frac{2t^2}{U}.$$

Among these four states, the ferro-magnetic antiferro-orbital state is thus the lowest in energy. The main message is that the system gains superexchange energy by occupying preferentially different orbitals with the same spin, although the orbitals are by themselves degenerate. The 16×16 matrix of the second-order energy corrections $\Delta E_{\alpha_1,\alpha_2}(1,1)$ can be rewritten as the effective superexchange Hamiltonian

$$\begin{aligned}\hat{H}_{\rm SE} =& 2\Gamma_{-+}\left[\boldsymbol{S}^A\cdot\boldsymbol{S}^B-\frac{1}{4}\right]\left[O_z^A\,O_z^B\,+\,\frac{1}{4}\right]+2\Gamma_{+-}\left[\frac{1}{4}\,+\,S_z^A\,S_z^B\right]\left[\boldsymbol{O}^A\cdot\boldsymbol{O}^B-\frac{1}{4}\right]\\ &+2\Gamma_{--}\left[\left(\boldsymbol{S}^A\cdot\boldsymbol{S}^B-S_z^A\,S_z^B\right)\left(\boldsymbol{O}^A\cdot\boldsymbol{O}^B-O_z^A\,O_z^B\right)-\left(S_z^A\,S_z^B-\frac{1}{4}\right)\left(O_z^A\,O_z^B-\frac{1}{4}\right)\right]\end{aligned}$$

where $\mathbf{O}_i = \boldsymbol{\tau}_i/2$ are operators acting only on orbital degrees of freedom and $\boldsymbol{\tau}$ are the pseudo-spin operators introduced in the previous section, Eq. (21), and

$$\Gamma_{-+} = \frac{4t^2}{U} \qquad \Gamma_{+-} = \frac{4t^2}{U-3J} \qquad \Gamma_{--} = -\frac{4t^2}{U-2J}.$$

When the second-order Hamiltonian is written in this form it is immediately clear that, among the four states we considered, the ferro-magnetic antiferro-orbital state is lower in energy. This happens because the superexchange coupling Γ_{+-} is the largest. If the orbital degeneracy is one, we can replace the terms $\boldsymbol{O}^A\cdot\boldsymbol{O}^B$ and $O_z^A\,O_z^B$ with the ferro-orbital value 1/4; then, the terms proportional to Γ_{+-} and Γ_{--} drop out and we recover the Heisenberg superexchange Hamiltonian, as expected for the one-band Hubbard model.

What about $KCuF_3$ and $LaMnO_3$? If we consider only hopping integrals between neighboring B sites in the cubic perovskite structure, the hopping integral matrices take the simple form

$$t_{mm'}^{i,i\pm\hat{z}} = t_\varepsilon\begin{pmatrix}0&0\\0&1\end{pmatrix} \qquad t_{mm'}^{i,i\pm\hat{x}} = t_\varepsilon\begin{pmatrix}\frac{3}{4}&\frac{\sqrt{3}}{4}\\\frac{\sqrt{3}}{4}&\frac{1}{4}\end{pmatrix} \qquad t_{mm'}^{i,i\pm\hat{y}} = t_\varepsilon\begin{pmatrix}\frac{3}{4}&-\frac{\sqrt{3}}{4}\\-\frac{\sqrt{3}}{4}&\frac{1}{4}\end{pmatrix}. \tag{23}$$

The structure of these matrices can be obtained by using Slater-Koster two-center integrals. The only non-zero hopping integral in the $\hat{z}$ direction is the one between $|3z^2-r^2\rangle$ states. As we have previously seen by using the downfolding approach, it is given by $t_\varepsilon = V_{pd\sigma}^2/(\varepsilon-\varepsilon_p)$.

As in the case of the two-site molecule, for integer filling (n electrons per atom) and in the large t_ε/U limit the lattice Hubbard model can be mapped onto an effective superexchange Hamiltonian by downfolding high-energy states in which some of the atoms have an electron number larger than n. Only two electronic configurations are relevant for orbital ordering, e_g^1 ($LaMnO_3$) and e_g^3 ($KCuF_3$). The remaining partially filled state, e_g^2, is magnetic with $S=1$ but, due to Hund's rule coupling J, it exhibits no orbital degeneracy ($L=0$). After excluding e_g^2 we can, for simplicity, set $J=0$. Let us now construct all atomic states $|N_e\rangle_\alpha$ with N_e electrons.

For a single atom they are

$\|N_e\rangle_\alpha$	$E_{\alpha'}(N_e)$	$d(N_e)$
$\|0\rangle$	$E(0)=0$	$d(0)=1$
$\|1\rangle = c^\dagger_{m\sigma}\|0\rangle$	$E(1)=\varepsilon_{e_g}$	$d(0)=4$
$\|2\rangle = c^\dagger_{m\sigma}c^\dagger_{m'\sigma'}\|0\rangle$	$E(2)=2\varepsilon_{e_g}+U$	$d(0)=6$
$\|3\rangle = c^\dagger_{m\sigma}c^\dagger_{m'\uparrow}c^\dagger_{m'\downarrow}\|0\rangle$	$E(3)=3\varepsilon_{e_g}+3U$	$d(0)=4$
$\|4\rangle = c^\dagger_{m\uparrow}c^\dagger_{m\downarrow}c^\dagger_{m'\uparrow}c^\dagger_{m'\downarrow}\|0\rangle$	$E(4)=4\varepsilon_{e_g}+6U$	$d(0)=1$

The total (spin and orbital) degeneracy of the n-electron sector, $d(N_e)$, is given in the third column. Let us consider two neighboring sites i and i' and their states $|N_e\rangle^i_\alpha$ and $|N'_e\rangle^{i'}_{\alpha'}$, where α and α' run over all degenerate states in the N_e-electron sector. We define the collective state of such a two-site system as $|N_e\rangle^i_\alpha|N'_e\rangle^{i'}_{\alpha'}$. Let us start from an e_g^1 configuration. In the large-U limit, at quarter filling (n=1) the ground state will be within the N_e=N'_e=1 manifold, $|\mathrm{G}\rangle=\{|1\rangle^i_\alpha|1\rangle^{i'}_{\alpha'}\}$. The latter has a degeneracy 4^N, where N is the number of sites, here N=2; this degeneracy can be partially lifted via virtual excitations to the doubly occupied states $|\mathrm{E}\rangle=\{|2\rangle^i_\alpha|0\rangle^{i'}\}, \{|0\rangle^i|2\rangle^{i'}_{\alpha'}\}$, which in turn generate an effective low-energy Hamiltonian $\hat{H}_{\mathrm{SE}}$. We can again calculate $\hat{H}_{\mathrm{SE}}$ by treating $\hat{H}_T$ as a perturbation.

Let us consider at first only pairs of sites along the $\hat{z}$ axis. In second-order perturbation theory in $\hat{H}_T$, we obtain for the lattice the following effective Hamiltonian

$$\begin{aligned}
\hat{H}^{\hat{z}}_{\mathrm{SE}} &\sim -\frac{1}{U}\sum_E \hat{H}_T|E\rangle\langle E|\hat{H}^\dagger_T \\
&= -\frac{t^2}{U}\frac{1}{2}\sum_{ii'}\sum_{\sigma\sigma'}\sum_{\alpha}\left\{c^\dagger_{i\tau\sigma}|0\rangle^{i\,i}\langle 0|c_{i\tau\sigma'}\left(c_{i'\tau\sigma}|2\rangle^{i'\,i'}_{\alpha\,\alpha}\langle 2|c^\dagger_{i'\tau\sigma'}\right) + (i\longleftrightarrow i')\right\}\delta_{\tau,\searrow} \\
&= -\frac{2t^2}{U}\frac{1}{2}\sum_{ii'}\sum_{\sigma\sigma'}\left\{(-1)^{-\sigma'-\sigma}P^i_{\tau\sigma-\sigma'}P^{i'}_{\tau\sigma'-\sigma} + \frac{1}{2}\left(P^i_{\tau\sigma\sigma}P^{i'}_{-\tau\sigma'\sigma'} + P^i_{-\tau\sigma\sigma}P^{i'}_{\tau\sigma'\sigma'}\right)\right\}\delta_{\tau,\searrow},
\end{aligned}$$

where we already replaced in the denominator $\Delta E = E(2)+E(0)-2E(1)$ with its value, U, and where, once more, $|\searrow\rangle = |3z^2{-}r^2\rangle$, $|\nearrow\rangle = |x^2{-}y^2\rangle$. In Hamiltonian $\hat{H}^{\hat{z}}_{\mathrm{SE}}$ we introduced the operators $P^i_{\tau\sigma\sigma'}$, which are given by

$$P^i_{\tau\sigma\sigma'} = c^\dagger_{i\tau\sigma}|0\rangle\langle 0|c_{i\tau\sigma'} = \hat{o}^z_{\tau\tau}\ \left(\hat{s}^z_{\sigma\sigma'}+\hat{s}^+_{\sigma\sigma'}+\hat{s}^-_{\sigma\sigma'}\right).$$

In this expression on the right-hand side we rewrote $P^i_{\tau\sigma\sigma'}$ as product of an orbital and a spin term, defined as follows:

$$\begin{aligned}
\hat{o}^z_{\tau\tau'} &= \left(\frac{n_i}{2}\,\hat{I}+(-1)^\tau O^i_z\right)\delta_{\tau\tau'} & \hat{s}^z_{\sigma\sigma} &= \left(\frac{n_i}{2}\,\hat{I}+(-1)^\sigma S^i_z\right)\delta_{\sigma\sigma'} \\
\hat{o}^+_{\tau\tau'} &= O^i_+(1-\delta_{\tau\tau'}) & \hat{s}^+_{\sigma\sigma'} &= S^i_+(1-\delta_{\sigma\sigma'}) \\
\hat{o}^-_{\tau\tau'} &= O^i_-(1-\delta_{\tau\tau'}) & \hat{s}^-_{\sigma\sigma'} &= S^i_-(1-\delta_{\sigma\sigma'}),
\end{aligned}$$

where $(-1)^\sigma = +1$ for spin (pseudospin) up and -1 otherwise; the operator $\hat{I}$ is the identity matrix. Hence, we can express the effective Hamiltonian as

$$\hat{H}^{\hat{z}}_{\rm SE} = \frac{\Gamma}{2}\sum_{ii'}\left[\boldsymbol{S}^i\cdot\boldsymbol{S}^{i'} - \frac{n_i n_{i'}}{4}\right]\left[O^i_z - \frac{n_i}{2}\right]\left[O^{i'}_z - \frac{n_{i'}}{2}\right] + \frac{1}{2}\left[O^i_z O^{i'}_z - \frac{n_i n_{i'}}{4}\right],$$

where $\Gamma = 4t^2/U > 0$. If we drop all processes involving orbital $|\nearrow\rangle$ we recover the usual superexchange Heisenberg Hamiltonian for the one-band Hubbard model

$$\hat{H}^{\hat{z}}_{\rm SE} = \frac{\Gamma}{2}\sum_{ii'}\left[\boldsymbol{S}^i\cdot\boldsymbol{S}^{i'} - \frac{n_i n_{i'}}{4}\right].$$

Let us now consider two neighboring sites and the energy of some possible states $|{\rm G}\rangle = \{|1\rangle^i_\alpha|1\rangle^{i'}_{\alpha'}\}$. A ferro-magnetic spin configuration has energy

$$\Delta E_{\tau\uparrow,\tau'\uparrow} = -\frac{\Gamma}{4}(1-\delta_{\tau,\tau'}),$$

hence, there is an energy gain if the electrons occupy different orbitals, i.e., if the systems has antiferro-orbital arrangement. Let us consider now a antiferro-magnetic spin arrangement. The corresponding energy is

$$\Delta E_{\tau\uparrow,\tau'\downarrow} = -\frac{\Gamma}{2}\delta_{\tau,\tau'}\delta_{\tau,\searrow} - \frac{\Gamma}{4}(1-\delta_{\tau,\tau'})$$

The expression above shows that in the antiferro-magnetic case the system gains more energy if the occupied state is $|\searrow\rangle$ at both sites. Up to now we considered magnetically ordered states. In $LaMnO_3$ and $KCuF_3$, however, orbital order takes place well above the magnetic transition. Let us then assume that the system is orbitally ordered but paramagnetic, with occupied state

$$|\vartheta\rangle_i = -\sin\frac{\vartheta-\pi}{2}|x^2{-}y^2\rangle + \cos\frac{\vartheta-\pi}{2}|3z^2{-}r^2\rangle$$

at site i and $|\vartheta\rangle_{i\pm\hat{z}} = |\vartheta\rangle_i$ at the neighboring site $i' = i \pm \hat{z}$. This choice corresponds to ferro-orbital order along $\hat{z}$, the type of stacking realized in $LaMnO_3$ (see Fig. 13). What is the value of ϑ than minimizes the energy? We can calculate it using the variational method. The superexchange energy gain with respect to a paramagnetic paraorbital state is given by

$$\Delta E(\vartheta) = \frac{\Gamma}{16}\Big(\cos^2(\vartheta{-}\pi) + 2\cos(\vartheta{-}\pi)\Big).$$

This function is minimized for $\vartheta = 0$, an angle corresponding to a tetragonal compression. To determine the optimal angle for the three-dimensional system we have in addition to take into account the effective Hamiltonian stemming from virtual hoppings in the remaining directions. Due to cubic symmetry, if we rotate the quantization axis, the superexchange Hamiltonian has the same form in all directions; to sum up all terms we have merely to rotate back the quantization axis to $\hat{z}$. Hence, we have to make the replacements

$$O^i_z \underbrace{\to}_{\hat{z}\to\hat{x}} -\frac{1}{2}O^i_z - \frac{\sqrt{3}}{2}O^i_x$$

$$O^i_z \underbrace{\to}_{\hat{z}\to\hat{y}} -\frac{1}{2}O^i_z + \frac{\sqrt{3}}{2}O^i_x$$

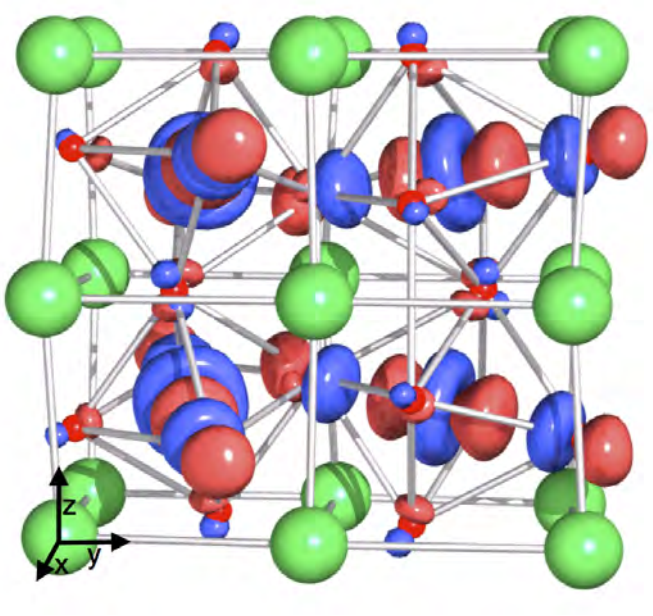


Fig. 13: *Orbital order (LDA+DMFT calculations) in the rare-earth perovskite $TbMnO_3$ with the $GdFeO_3$-type structure. From Ref. [6]. This system has the same structure of $LaMnO_3$.*

Let us assume antiferro-orbital order in the plane, again as in the case of $LaMnO_3$, shown in Fig. 13. This means that, for $i' = i \pm \hat{x}$ or $i' = i \pm \hat{y}$, the occupied state is

$$|\vartheta\rangle_{i'} = +\sin\frac{\vartheta-\pi}{2}|x^2-y^2\rangle + \cos\frac{\vartheta-\pi}{2}|3z^2-r^2\rangle.$$

We can easily verify that $|\vartheta\rangle_{i'}=|-\vartheta+2\pi\rangle_i$. This is state $|\vartheta\rangle_i$ rotated by $\pi/2$ ($x \to y$, $y \to -x$). The total superexchange energy gain with respect to a paramagnetic paraorbital state is then given by[11]

$$\Delta E(\vartheta) = \frac{\Gamma}{16}\left(3\cos^2(\vartheta-\pi)-\frac{3}{2}\right).$$

This expression has a minimum for $\vartheta = \pi/2$ (Jahn-Teller-like $\boldsymbol{Q}_1$ distortion). For the e_g^3 configuration ($KCuF_3$), due to particle-hole symmetry, we obtain the same result. This can be verified by observing, first of all, that the e_g bands obtained from the hopping-integrals matrices (23) –the bands which we have discussed in detail in Sec. 3– are symmetric with respect to the Fermi level for half filling. In addition, the energy difference entering in the denominator of the superexchange Hamiltonian for an e_g^3 ground state, $\Delta E = E(4) + E(2) - 2E(3)$, has the same value ($\Delta E = U$) as in the case of an e_g^1 ground state. The main difference between $LaMnO_3$ (e_g^1) and $KCuF_3$ (e_g^3), for what concerns the results presented in this section, is that the stacking along $\hat{z}$, ferro-orbital for $LaMnO_3$, can be either antiferro- or ferro-orbital for $KCuF_3$; Fig. 1 shows the case of antiferro-orbital arrangement. Remarkably, the variational energy gain $\Delta E(\vartheta)$ is the same for both types of stacking along $\hat{z}$, i.e., for $|\vartheta\rangle_{i\pm\hat{z}} = |\vartheta\rangle_i$ and for $|\vartheta\rangle_{i\pm\hat{z}} = |-\vartheta+2\pi\rangle_i$. The conclusions of this section are thus identical for $LaMnO_3$ and $KCuF_3$.

[11] For the application of this approach to the general super-exchange Hamiltonian see Ref. [20].

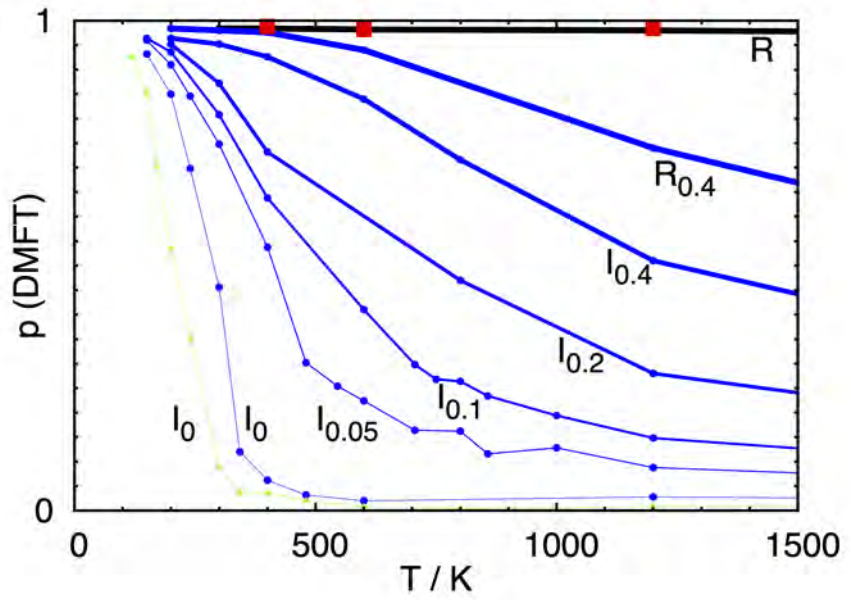


Fig. 14: *Orbital order transition in $KCuF_3$. Orbital polarization p as a function of temperature calculated in LDA+DMFT. R: experimental structure. Circles: idealized structures R_δ and I_δ with decreasing crystal-field and U=7 eV. Green/Triangles: U= 9 eV, I_0 only. Red/Squares: two-sites CDMFT. From Ref. [3].*

6 The origin of orbital ordering in materials

As we discussed in the introduction, the hallmark of orbital order is the co-operative Jahn-Teller distortion. This static distortion gives rise to a crystal field, which splits the otherwise degenerate e_g doublet or t_{2g} triplet. Due to Coulomb repulsion, it turns out that even a crystal-field splitting much smaller than the band width can lead to orbital ordering. The importance of this effect for real materials has been realized first for $LaTiO_3$ and $YTiO_3$ [4]. This reduction of orbital fluctuation is dynamical, but it can be already understood from the static Hartree-Fock contribution to the self-energy; the latter yields an effective enhancement of the crystal-field proportional to orbital polarization p. For an e_g system p is defined as the difference in occupation between the most and the least occupied orbital, $|1\rangle$ and $|2\rangle$, the so-called natural orbitals. Thus $p{=}n_1{-}n_2$, and the Hartree-Fock self-energy correction to the crystal-field splitting is

$$\Delta\varepsilon_{\rm CF} = \Sigma_2(\omega_n{\to}\infty) - \Sigma_1(\omega_n{\to}\infty) \sim \frac{1}{2}(U{-}5J)p.$$

If $p > 0$, as it happens in the presence of a crystal-field $\varepsilon_{\rm CF}{=}\varepsilon_2{-}\varepsilon_1{>}0$, this term effectively increases the crystal-field splitting. This effect is at work not only in $LaTiO_3$ and $YTiO_3$, but also in several other systems with different electronic structure and even smaller crystal-field splittings. The case of $3d^9$ $KCuF_3$ and $3d^4$ $LaMnO_3$ is extreme: the e_g crystal-field splitting is $\sim 0.5{-}1$ eV; with such a large splitting, orbital fluctuations are suppressed up to the melting temperature. Thus, Coulomb repulsion makes the Jahn-Teller mechanism discussed in the article of Kanamori very efficient. This result, however, does not clarify which of the two mech-

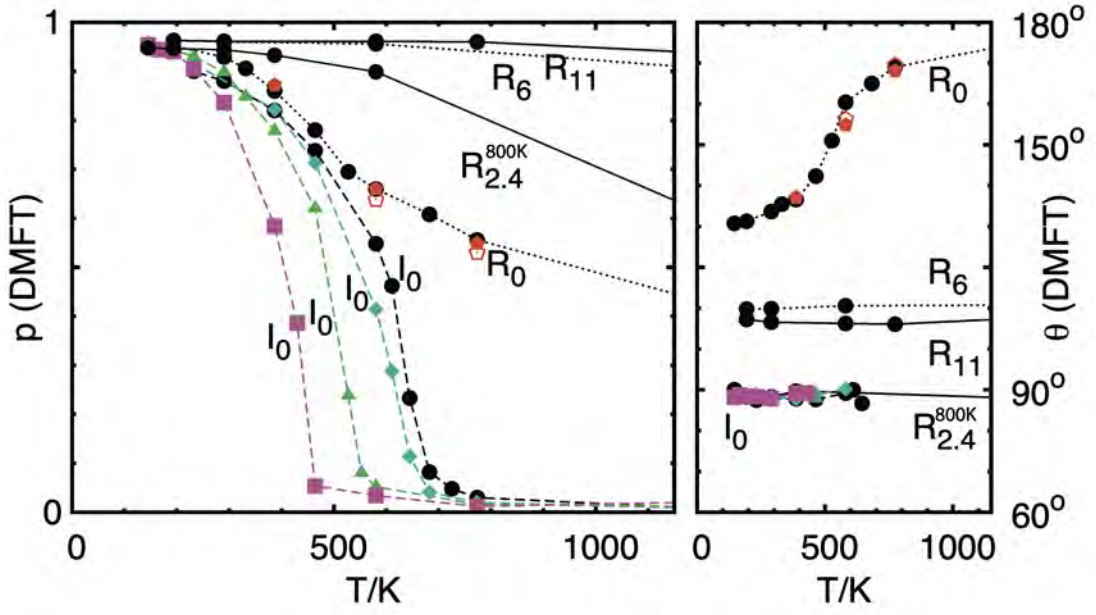


Fig. 15: *Orbital-ordering transition in $LaMnO_3$. Orbital polarization p (left) and (right) occupied state $|\vartheta\rangle = \cos\frac{\vartheta}{2}|3z^2-r^2\rangle + \sin\frac{\vartheta}{2}|x^2-y^2\rangle$ as a function of temperature. Solid lines: 300 K experimental structure (R_{11}) and 800 K experimental structure. Dots: orthorhombic structures with half (R_6) or no (R_0) Jahn-Teller distortion. Pentagons: 2 (full) and 4 (empty) site CDMFT. Dashes: ideal cubic structure (I_0). Circles: $U = 5$ eV. Diamonds: U=5.5 eV. Triangles: $U = 6$ eV. Squares: U=7 eV. Crystal field splittings (meV): 840 (R_{11}), 495 (R_6), 168 ($R_{2.4}^{800\ K}$), and 0 (I_0). From Ref. [6].*

anisms, Kugel-Khomskii superexchange or conventional electron-phonon coupling, plays the major role in causing orbital order and stabilizing the distortion. Remarkably, in fact, Coulomb repulsion has also an important effect on structure stabilization. LDA+U total energy calculations have early on shown that the co-operative Jahn-Teller distortion is stabilized by U [14,15], a result confirmed recently by LDA+DMFT [16]. This could be – and initially was – taken as an indication that superexchange is the driving mechanism. If this is the case, it is, however, hard to explain why the magnetic transition temperature ($T_N \sim 40$ K for $KCuF_3$ and $T_N \sim 140$ K for $LaMnO_3$), also determined by superexchange, is relatively low while the co-operative Jahn-Teller distortion persists up to the melting temperature. On the other hand, if Kugel-Khomskii superexchange is not the driving mechanism, the associated energy gain should be small with respect to the total energy gain due to the Jahn-Teller distortion.

To clarify the nature of the dominant mechanism, we disentangled electron-phonon and superexchange effects. To this end we performed LDA+DMFT (single-site and cluster) calculations for a series of hypothetical structures, in which the distortions (and thus the crystal-field splitting) are progressively reduced. In the case of $KCuF_3$, these hypothetical structures are shown in Fig. 1, and the corresponding e_g bands are shown in Fig. 7. For each structure we calculate the order parameter, the orbital polarization p. In Fig. 14 we show p as a function of temperature. For the experimental structure (R in the figure), we find that $p \sim 1$ up to the melting temperature. The empty orbitals on different sites make the pattern shown in Fig. 1.

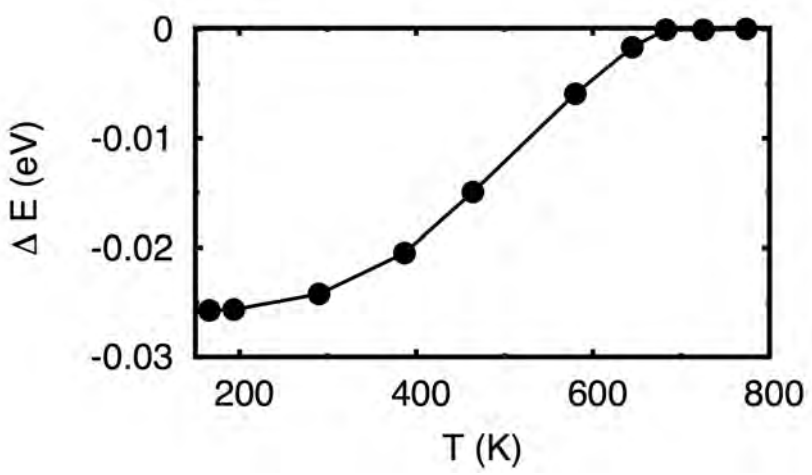


Fig. 16: *Superexchange energy gain for $LaMnO_3$, $\Delta E \sim -T_{\rm KK}/2$. From Ref. [6].*

For the ideal cubic structure I_0, we find that p=0 at high temperature, but a transition occurs at $T_{\rm KK} \sim 350$ K. This $T_{\rm KK}$ is the critical temperature in the absence of electron-phonon coupling, i.e., the superexchange critical temperature. Our results show that around 350 K superexchange alone could indeed drive the co-operative Jahn-Teller distortion; it cannot, however, explain the presence of a co-operative Jahn-Teller distortion above 350 K. We performed a similar study fo $LaMnO_3$. For this $t_{2g}^3 e_g^1$ system we have to take into account the Hund's rule coupling between e_g electrons and t_{2g} spins, $\boldsymbol{S}_{t_{2g}}$. Thus the minimal model to understand orbital order is the modified Hubbard model [17]

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

Here the local magnetic field h=$JS_{t_{2g}}$ describes the Hund's rule coupling to t_{2g} electrons, and $u_{i\sigma,i'\sigma'}$=$2/3(1-\delta_{i,i'})$ accounts for the disorder in orientation of the t_{2g} spins. By performing the same type of analysis as for $KCuF_3$, we find the impressively large $T_{\rm KK}$∼700 K (Fig. 15). There is a small point neglected so far; besides the co-operative Jahn-Teller distortion and tetragonal compression, $LaMnO_3$ exhibits a $GdFeO_3$-type distortion (Fig. 13), which tends to reduce the e_g band width [4]. To account for this we studied the orbital-order transition for the ideal structure R_0, which retains all distortions except for the Jahn-Teller one. For structure R_0 we cannot obtain $T_{\rm KK}$ from $p(T)$, because, due to the ∼200 meV crystal-field splitting, Coulomb repulsion strongly suppress orbital fluctuations even at 1500 K. We can, however, study the evolution with temperature of the occupied orbital, here defined as $|\vartheta\rangle = \cos\frac{\vartheta}{2}|3z^2-r^2\rangle + \sin\frac{\vartheta}{2}|x^2-y^2\rangle$. For the experimental structure (R_{11}) we find $\vartheta \sim 108°$, in agreement with experiments, while for the I_0 structure we obtain $\vartheta = 90°$. For the R_0 structure we find two regimes: At high temper-

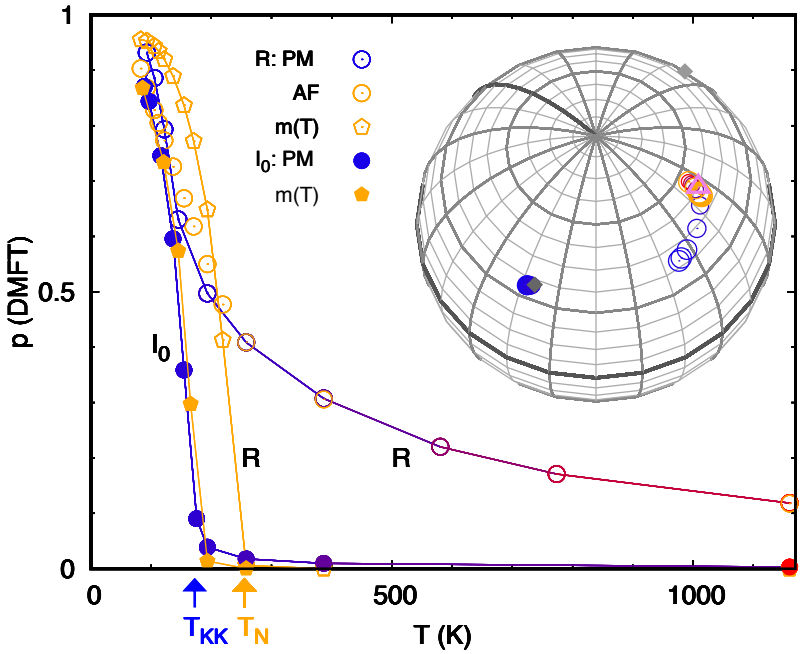


Fig. 17: *Orbital ordering in the t_{2g}^2 system $LaVO_3$. Filled circles: ideal case without crystal-field splitting. Empty circles: experimental structure. For each temperature, the associated hole orbital is shown on the Bloch sphere. At high temperature it coincides with the predictions of crystal-field theory (triangle). At the Kugel-Khomskii transition temperature, $T_{\rm KK}$, it starts to move towards the ideal Kugel-Khomskii result (filled blue circle). From Ref. [12].*

ature the occupied orbital is the lower-energy crystal-field orbital (ϑ=180°). At $T_{\rm KK} \sim 550$ K superexchange rotates this ϑ towards 90°, reaching 130° in the zero-temperature limit; this is the actual superexchange transition temperature for $LaMnO_3$. Such $T_{\rm KK}$ is still remarkably large, however not sufficient to explain the persistence of the Jahn-Teller distortion in nanoclusters up to basically melting temperature [18]. Furthermore, the superexchange energy gain associated with orbital order (Fig. 16) is small compared to the total energy gain due to the Jahn-Teller distortion, calculated via LDA+U [14, 15] or LDA+DMFT [16]. Thus, as in the case of $KCuF_3$, the conclusion is that a static crystal-field splitting, as the one generated by the electron-lattice coupling, is essential to explain orbital ordering at high temperature. We obtained a similar conclusion for various families of compounds, including t_{2g} systems, an indication that pure Kugel-Khomskii materials are actually rare. The first clear-cut case in which the super-exchange interaction controls orbital ordering, turning the hole orbital way from the state expected from crystal-field theory, was recently identified in the t_{2g}^2 system $LaVO_3$ [12]. This is shown in Fig. 17, where the changes in the hole orbital on lowering the temperature can be followed on the Bloch sphere (empty circles). Decreasing the temperature the color of the empty circles changes from red to blue, while the associated polarization increases towards its maximum vale. One may see that at high temperature they overlap with the pink triangle, representing the state expected from crystal-field theory. Decreasing the temperature they move towards the filled blue circle, representing the Kugel-Khomskii ideal value.

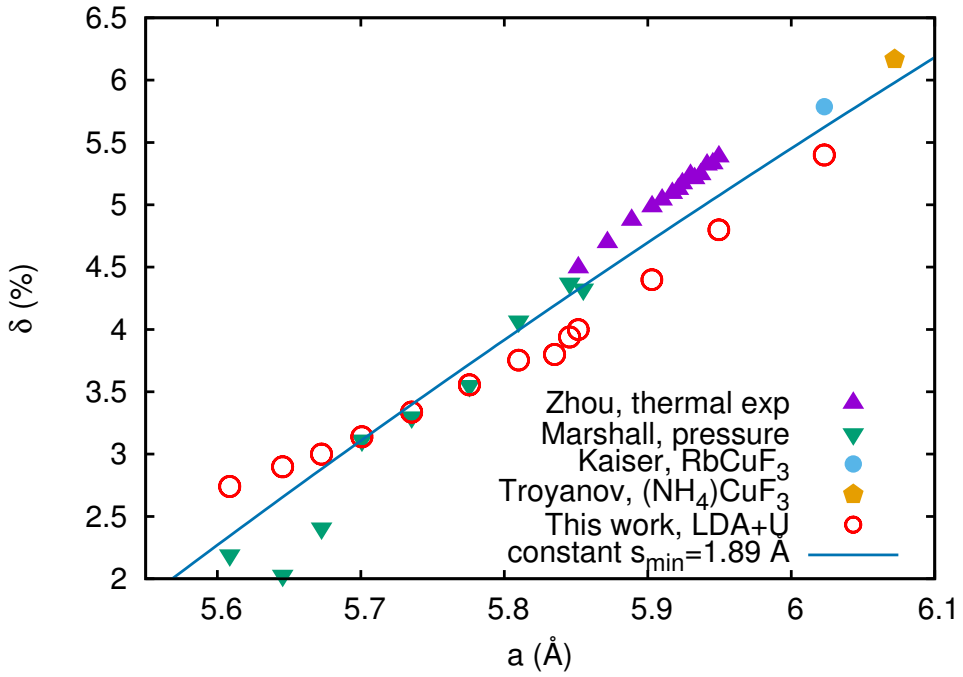


Fig. 18: *Distortion parameter δ as a function of lattice constant a in thermally expanding $KCuF_3$ [21], under hydrostatic pressure [22], for $RbCuF_3$ [23] and $(NH_4)CuF_3$ [24], compared to our calculations and the values obtained assuming a constant short Cu-F distance ($s_{\min}$). From Ref. [11].*

Let us now return to $KCuF_3$, the case we have examined in greater detail. The main conclusion we had reached is that a static distortion is necessary to explain the presence of orbital ordering at high temperature. Based on the discussion so far, one could at this point conclude that the latter is determined by the Jahn-Teller effect. However, it turns out that the reality is even more complex. Indeed, in a second-order transition one would expect that the order parameter goes to zero at the transition temperature, $T_{\rm OO}$. In the case of electron-phonon-coupling driven orbital ordering, the order parameter is the Jahn-Teller distortion. If $T_{\rm OO}$ is not yet reached at the melting temperature, the order parameter should at least decrease with temperature. In $KCuF_3$, however, it has been found that this simple picture fails to describe experiments. This is shown in Fig. 18. Increasing the temperature the lattice constant increases by thermal expansion. At the same time the (dimensionless) Jahn-Teller distortion parameter δ also increases. This surprising behavior is due to the fact that the short Cu-F bond remains almost constant while the long Cu-F bond becomes longer [11], instead of the two changing coherently together as expected via the Jahn-Teller Q_1 mode. Going to the microscopical origin of this behavior, it turns out that the Jahn-Teller mode is so soft that the distortion is actually determined by the Born-Mayer repulsion of the ions. Thus the distortion increases with the lattice constant, and, via thermal expansion, the order parameter increases with temperature. This new ordering mechanism was identified in Ref. [11] for the first time. It can operate even in closed-shell systems and would result in an inverted Landau transition, with symmetry breaking *above* a critical temperature.

7 Conclusion

In this lecture we have studied two mechanisms that can lead to orbital ordering phenomena in Mott insulators. The first is well illustrated in the influential paper of Kanamori, Ref. [1]. In this picture, a co-operative Jahn-Teller distortion generates a static crystal-field, which in turn splits orbitals otherwise degenerate. This mechanism is made more efficient by Coulomb repulsion; the latter enhances the orbital polarization, leading to a orbitally-ordered state even if the crystal-field splitting is a mere fraction of the bandwidth [4]. The second mechanism, proposed by Kugel and Khomskii [2] in 1973, predicts orbital ordering even in the absence of a static crystal field; in this picture, orbital ordering is due to the superexchange interaction, the effective interaction emerging from the orbitally-degenerate Hubbard model in the large U limit. The general super-exchange Hamiltonians for e_g and t_{2g} systems can be found in Ref. [20], where the interaction is decomposed in its irreducible tensor components. In paradigmatic materials, both the Jahn-Teller and super-exchange coupling predict a similar type of order. Thus identifying which interaction dominates is very difficult. For this reason, the riddle of the origin of orbital ordering in materials can be viewed an example of a chicken-and-egg problem – and has been accordingly a matter of debate for decades.

In the last section we saw how this problem was solved in representative cases. This was done by disentangling the superexchange Kugel-Khomskii interaction from the rest. For the two classical text-book examples of orbitally-ordered systems, $KCuF_3$ and $LaMnO_3$, it was shown via this approach that, although Kugel-Khomskii superexchange is very efficient, it cannot alone explain the presence of a co-operative Jahn-Teller distortion up to the melting temperature. The conclusion is that an interaction giving directly rise to a static crystal-field splitting, e.g., electron-phonon coupling, is necessary to explain experimental findings [3, 6]. The same result was obtained for many other materials, with either e_g or t_{2g} partially filled shells. This shows that purely super-exchange driven ordering is rare in nature. A clear cut case of Kugel-Khomskii material was nevertheless recently identified, $LaVO_3$ [12]. Finally, to complicate the matter, for $KCuF_3$ it was shown that not even the Jahn-Teller effect alone does explain the evolution of distortions with temperature. A new ordering mechanism in which the Born-Mayer repulsion of the ions plays a key role had to be introduced [11]. Only then it could be understood why the order parameter experimentally increases (instead of decreasing) with temperature.

Appendices

A Constants and units

In this lecture, formulas are given in atomic units. The unit of mass m_0 is the electron mass ($m_0 = m_e$), the unit of charge e_0 is the electron charge ($e_0 = e$), the unit of length a_0 is the Bohr radius ($a_0 = a_B \sim 0.52918$ Å), and the unit of time is $t_0 = 4\pi\varepsilon_0\hbar a_0/e^2$. In these units, m_e, a_B, e and $1/4\pi\varepsilon_0$ have the numerical value 1, the speed of light is $c = 1/\alpha \sim 137$, and the unit of energy is $1\mathrm{Ha} = e^2/4\pi\varepsilon_0 a_0 \sim 27.211$ eV.

B Atomic orbitals

B.1 Radial functions

The nlm hydrogen-like atomic orbital is given by

$$\psi_{nlm}(\rho,\vartheta,\varphi) = R_{nl}(\rho)Y^l_m(\vartheta,\varphi),$$

where $R_{nl}(\rho)$ is the radial function and $Y^l_m(\vartheta,\varphi)$ a spherical harmonic, $\rho = Zr$ and Z the atomic number. In atomic units, the radial functions are

$$R_{nl}(\rho) = \sqrt{\left(\frac{2Z}{n}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}}\, e^{-\rho/n} \left(\frac{2\rho}{n}\right)^l L^{2l+1}_{n-l-1}\left(\frac{2\rho}{n}\right),$$

where L^{2l+1}_{n-l-1} are generalized Laguerre polynomials of degree $n-l-1$.
The radial function for $n = 1, 2, 3$ are

$$\begin{aligned}
R_{1s}(\rho) &= 2 \quad Z^{3/2} && e^{-\rho} \\
R_{2s}(\rho) &= \tfrac{1}{2\sqrt{2}} \quad Z^{3/2} && (2-\rho)\, e^{-\rho/2} \\
R_{2p}(\rho) &= \tfrac{1}{2\sqrt{6}} \quad Z^{3/2} && \rho\, e^{-\rho/2} \\
R_{3s}(\rho) &= \tfrac{2}{3\sqrt{3}} \quad Z^{3/2} && (1-2\rho/3+2\rho^2/27)\, e^{-\rho/3} \\
R_{3p}(\rho) &= \tfrac{4\sqrt{2}}{9\sqrt{3}} \quad Z^{3/2} && \rho(1-\rho/6)\, e^{-\rho/3} \\
R_{3d}(\rho) &= \tfrac{2\sqrt{2}}{81\sqrt{15}} \quad Z^{3/2} && \rho^2\, e^{-\rho/3}
\end{aligned}$$

where we used the standard notation s for l=0, p for l=1 and d for l=2. The spherical Harmonics, using the Condon-Shortley convention, are given by

$$Y^\ell_m(\vartheta,\varphi) = (-1)^m \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}\, P^\ell_m(\cos\vartheta)\, e^{im\varphi} \tag{24}$$

where $P^\ell_m(\cos\vartheta)$ in an associated Legendre polynomial.

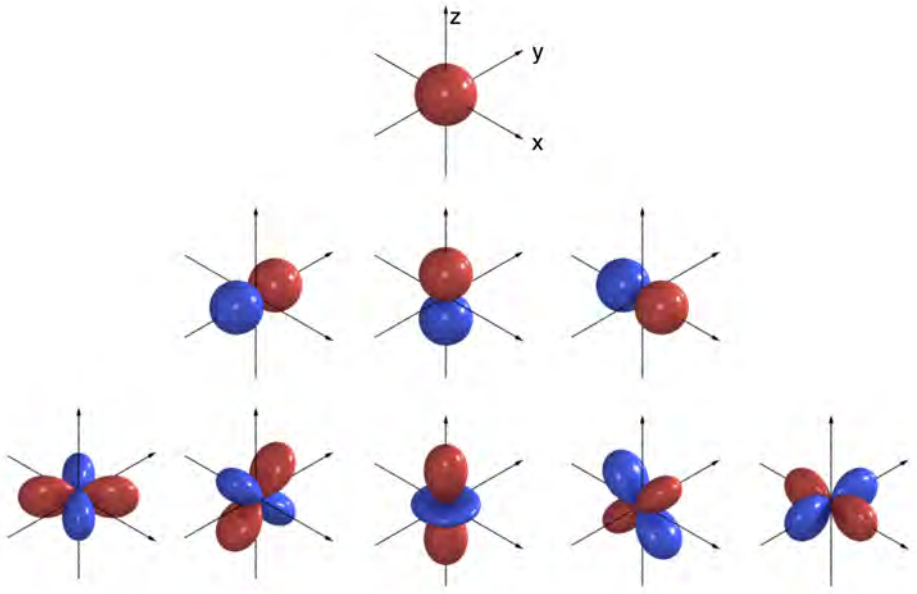


Fig. 19: *The s (first row), p_y, p_z, p_x (second row), and d_{xy}, d_{yz}, $d_{3z^2-r^2}$, d_{xz}, $d_{x^2-y^2}$ (last row) real harmonics.*

B.2 Real harmonics

To study solids, it is usually convenient to work in the basis of real harmonics. The latter are defined in terms of the spherical harmonics as follows:

$$y_{l0} = Y^l_0, \quad y_{lm} = \frac{1}{\sqrt{2}}\big(Y^l_{-m} + (-1)^m Y^l_m\big), \quad y_{l-m} = \frac{i}{\sqrt{2}}\big(Y^l_{-m} - (-1)^m Y^l_m\big), \quad m > 0.$$

Using the definitions $x = r\sin\vartheta\cos\varphi$, $y = r\sin\vartheta\sin\varphi$, $z = r\cos\vartheta$, so that

$$\cos\vartheta = \frac{z}{r}, \qquad e^{\pm i\varphi}\sin\vartheta = \frac{(x \pm iy)}{r}, \tag{25}$$

we can express the $l = 0, 1, 2$ real harmonics (Fig. 19) as

$$\begin{aligned}
s \quad &= y_{00} \quad = Y^0_0 \qquad\qquad = \sqrt{\tfrac{1}{4\pi}} \\
p_y \quad &= y_{1-1} = \tfrac{i}{\sqrt{2}}(Y^1_{-1} + Y^1_1) = \sqrt{\tfrac{3}{4\pi}} \quad y/r \\
p_z \quad &= y_{10} \quad = Y^0_2 \qquad\qquad = \sqrt{\tfrac{3}{4\pi}} \quad z/r \\
p_x \quad &= y_{11} \quad = \tfrac{1}{\sqrt{2}}(Y^1_{-1} - Y^1_1) = \sqrt{\tfrac{3}{4\pi}} \quad x/r \\
d_{xy} \quad &= y_{2-2} = \tfrac{i}{\sqrt{2}}(Y^2_{-2} - Y^2_2) = \sqrt{\tfrac{15}{4\pi}} \quad xy/r^2 \\
d_{yz} \quad &= y_{2-1} = \tfrac{i}{\sqrt{2}}(Y^2_{-1} + Y^2_1) = \sqrt{\tfrac{15}{4\pi}} \quad yz/r^2 \\
d_{3z^2-r^2} \quad &= y_{20} \quad = Y^0_2 \qquad\qquad = \sqrt{\tfrac{15}{4\pi}}\tfrac{1}{2\sqrt{3}}\,(3z^2-r^2)/r^2 \\
d_{xz} \quad &= y_{21} \quad = \tfrac{1}{\sqrt{2}}(Y^2_{-1} - Y^2_1) = \sqrt{\tfrac{15}{4\pi}} \quad xz/r^2 \\
d_{x^2-y^2} \quad &= y_{22} \quad = \tfrac{1}{\sqrt{2}}(Y^2_{-2} + Y^2_2) = \sqrt{\tfrac{15}{4\pi}}\tfrac{1}{2} \quad (x^2-y^2)/r^2
\end{aligned}$$

B.3 Slater-Koster integrals

The interatomic Slater-Koster two-center integrals are defined as

$$E_{lm,l'm'} = \int d\boldsymbol{r}\ \overline{\psi_{lm}}(\boldsymbol{r}-\boldsymbol{d})V(\boldsymbol{r}-\boldsymbol{d})\psi_{l'm'}(\boldsymbol{r}).$$

They can be expressed as a function of radial integrals $V_{ll'\alpha}$, which scale with the distance d roughly as $d^{-(l+l'+1)}$ [19], and direction cosines, defined as

$$l{=}\boldsymbol{d}\cdot\hat{x}/d, \quad m{=}\boldsymbol{d}\cdot\hat{y}/d, \quad n{=}\boldsymbol{d}\cdot\hat{z}/d.$$

The Slater-Koster integrals for s-, p-, and d-orbitals [19] are listed below.

$$
\begin{array}{lcrrr}
E_{s,s} & = & V_{ss\sigma} & & \\
E_{s,x} & = & lV_{sp\sigma} & & \\
E_{x,x} & = & l^2V_{pp\sigma} & +(1{-}l^2)V_{pp\pi} & \\
E_{x,y} & = & lmV_{pp\sigma} & -lmV_{pp\pi} & \\
E_{x,z} & = & lnV_{pp\sigma} & -lnV_{pp\pi} & \\
E_{s,xy} & = & \sqrt{3}lmV_{sd\sigma} & & \\
E_{s,x^2-y^2} & = & \frac{1}{2}\sqrt{3}(l^2{-}m^2)V_{sd\sigma} & & \\
E_{s,3z^2-r^2} & = & [n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{sd\sigma} & & \\
E_{x,xy} & = & \sqrt{3}l^2mV_{pd\sigma} & +m(1{-}2l^2)V_{pd\pi} & \\
E_{x,yz} & = & \sqrt{3}lmnV_{pd\sigma} & -2lmnV_{pd\pi} & \\
E_{x,zx} & = & \sqrt{3}l^2nV_{pd\sigma} & +n(1{-}2l^2)V_{pd\pi} & \\
E_{x,x^2-y^2} & = & \frac{\sqrt{3}}{2}l[(l^2{-}m^2)]V_{pd\sigma} & +l(1{-}l^2{+}m^2)V_{pd\pi} & \\
E_{y,x^2-y^2} & = & \frac{\sqrt{3}}{2}m[(l^2{-}m^2)]V_{pd\sigma} & -m(1{+}l^2{-}m^2)V_{pd\pi} & \\
E_{z,x^2-y^2} & = & \frac{\sqrt{3}}{2}n[(l^2{-}m^2)]V_{pd\sigma} & -n(l^2{-}m^2)V_{pd\pi} & \\
E_{x,3z^2-r^2} & = & l[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{pd\sigma} & -\sqrt{3}ln^2V_{pd\pi} & \\
E_{y,3z^2-r^2} & = & m[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{pd\sigma} & -\sqrt{3}mn^2V_{pd\pi} & \\
E_{z,3z^2-r^2} & = & n[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{pd\sigma} & +\sqrt{3}n(l^2{+}m^2)V_{pd\pi} & \\
E_{xy,xy} & = & 3l^2m^2V_{dd\sigma} & +(l^2{+}m^2{-}4l^2m^2)V_{dd\pi} & +(n^2{+}l^2m^2)V_{dd\delta} \\
E_{xy,yz} & = & 3lm^2nV_{dd\sigma} & +ln(1{-}4m^2)V_{dd\pi} & +ln(m^2{-}1)V_{dd\delta} \\
E_{xy,zx} & = & 3l^2mnV_{dd\sigma} & +mn(1{-}4l^2)V_{dd\pi} & +mn(l^2{-}1)V_{dd\delta} \\
E_{xy,x^2-y^2} & = & \frac{3}{2}lm(l^2{-}m^2)V_{dd\sigma} & 2lm(m^2{-}l^2)V_{dd\pi} & \frac{1}{2}lm(l^2{-}m^2)V_{dd\delta} \\
E_{yz,x^2-y^2} & = & \frac{3}{2}mn(l^2{-}m^2)V_{dd\sigma} & -mn[1{+}2(l^2{-}m^2)]V_{dd\pi} & +mn[1{+}\frac{1}{2}(l^2{-}m^2)]V_{dd\delta} \\
E_{zx,x^2-y^2} & = & \frac{3}{2}nl(l^2{-}m^2)V_{dd\sigma} & +nl[1{-}2(l^2{-}m^2)]V_{dd\pi} & -nl[1{-}\frac{1}{2}(l^2{-}m^2)]V_{dd\delta} \\
E_{xy,3z^2-r^2} & = & \sqrt{3}lm[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{dd\sigma} & -2\sqrt{3}lmn^2V_{dd\pi} & \frac{\sqrt{3}}{2}lm(1{+}n^2)V_{dd\delta} \\
E_{yz,3z^2-r^2} & = & \sqrt{3}mn[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{dd\sigma} & +\sqrt{3}mn(l^2{+}m^2{-}n^2)V_{dd\pi} & -\frac{\sqrt{3}}{2}mn(l^2{+}m^2)V_{dd\delta} \\
E_{zx,3z^2-r^2} & = & \sqrt{3}ln[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{dd\sigma} & +\sqrt{3}ln(l^2{+}m^2{-}n^2)V_{dd\pi} & -\frac{\sqrt{3}}{2}ln(l^2{+}m^2)V_{dd\delta} \\
E_{x^2-y^2,x^2-y^2} & = & \frac{3}{4}(l^2{-}m^2)^2V_{dd\sigma} & +[l^2{+}m^2{-}(l^2{-}m^2)^2]V_{dd\pi} & +[n^2{+}\frac{1}{4}(l^2{-}m^2)^2]V_{dd\delta} \\
E_{x^2-y^2,3z^2-r^2} & = & \frac{\sqrt{3}}{2}(l^2{-}m^2)[n^2{-}\frac{1}{2}(l^2{+}m^2)]V_{dd\sigma} & +\sqrt{3}n^2(m^2{-}l^2)V_{dd\pi} & +\frac{\sqrt{3}}{4}(1{+}n^2)(l^2{-}m^2)V_{dd\delta} \\
E_{3z^2-r^2,3z^2-r^2} & = & [n^2{-}\frac{1}{2}(l^2{+}m^2)]^2V_{dd\sigma} & +3n^2(l^2{+}m^2)V_{dd\pi} & \frac{3}{4}(l^2{+}m^2)^2V_{dd\delta}
\end{array}
$$

References

[1] J. Kanamori, J. Appl. Phys. **31**, S14 (1960)

[2] K.I. Kugel and D.I. Khomskii, Zh. Eksp. Teor. Fiz. **64**, 1429 (1973)

[3] E. Pavarini, E. Koch, and A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008)

[4] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, O.K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004)
E. Pavarini A. Yamasaki, J. Nuss and O.K. Andersen, New J. Phys. **7**, 188 (2005)

[5] E. Pavarini: *The LDA+DMFT Approach*, in [7]

[6] E. Pavarini and E. Koch, Phys. Rev. Lett. **104**, 086402 (2010)
A. Flesch, G. Zhang, E. Koch, and E. Pavarini, Phys. Rev. B **85**, 035124 (2012)

[7] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (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`

[8] 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`

[9] 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`

[10] E. Pavarini, E. Koch, A. Lichtenstein, 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`

[11] H. Sims, E. Pavarini, and E. Koch, Phys. Rev. B **96**, 054107 (2017)

[12] X-J. Zhang, E. Koch and E. Pavarini, Phys. Rev. B **106**, 115110 (2022)

[13] E. Pavarini, E. Koch, F. Anders, M. Jarrell (eds.):
Correlated Electrons: From Models to Materials,
Reihe Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
`http://www.cond-mat.de/events/correl12`

[14] V.I. Anisimov, F. Aryasetiawan and A.I. Lichtenstein,
J. Phys. Condens. Matter **9**, 767 (1997)

[15] W.G. Yin, D. Volja, and W. Ku, Phys. Rev. Lett. **96**, 116405 (2006)

[16] I. Leonov, N. Binggeli, Dm. Korotin, V.I. Anisimov, and D. Vollhardt,
Phys. Rev. Lett. **101**, 096405 (2008)
I. Leonov, Dm. Korotin, N. Binggeli, V.I. Anisimov, and D. Vollhardt,
Phys. Rev. B **81**, 075109 (2010)

[17] K.H. Ahn, and A.J. Millis, Phys. Rev. B **61**, 13545 (2000)

[18] M.C. Sánchez, G. Subías, J. García, and J. Blasco, Phys. Rev. Lett. **90**, 045503 (2003)

[19] W.A. Harrison: *Electronic Structure and The Properties of Solids* (Dover, 1989)

[20] X-J. Zhang, E. Koch and E. Pavarini, Phys. Rev. B **105**, 115104 (2022)

[21] L.G. Marshall, J. Zhou, J. Zhang, J. Han, S.C. Vogel, X.Yu, Y. Zhao, M. Fernández-Díaz,
J. Cheng, and J.B. Goodenough, Phys. Rev. B **87**, 014109 (2013)

[22] J.-S. Zhou, J.A. Alonso, J.T. Han, M.T. Fernández-Díaz, J.-G. Cheng, and
J.B. Goodenough, J. Fluorine Chem. **132**, 1117 (2011)

[23] V. Kaiser, M. Otto, F. Binder, and D. Babel, Z. Anorg. Allg. Chem. 585, **93** (1990)

[24] S.I. Troyanov, I.V. Morozov, and Y.M. Korenev,
Zhurnal Neorganicheskoi Khimii **38**, 984 (1993) [Russian J. Inorg. Chem. **38**, 909 (1993)]

2 The Jahn-Teller Effect

Arnout Ceulemans
Department of Chemistry, KU Leuven
Celestijnenlaan 200F, B-3001 Leuven, Belgium

Contents

1 The Jahn-Teller theorem **2**
1.1 The distorted rutile structure 2
1.2 Origin of orbital instability 3
1.3 The Jahn-Teller Hamiltonian 4
1.4 The pseudo-Jahn-Teller effect 7

2 The doublet $E \times e$ Paradigm **9**
2.1 The potential energy surface 9
2.2 The dynamic system 11
2.3 Berry phase 15

3 The triplet $T \times (e + t_2)$ Jahn-Teller system **18**
3.1 The Hamiltonian 18
3.2 Dynamics 20

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 The Jahn-Teller theorem

In 1937 Jahn and Teller wrote:

Theorem 1. *All non-linear nuclear configurations for an orbitally degenerate electronic state are unstable.*

This statement was the beginning of a fruitful line of research both in physics and chemistry. Over the years, it has provided deep theoretical insights as well as important practical applications, in spectroscopy, magnetism, superconductivity and chemical reactivity. First and foremost, the theorem is a particular example of the more general physical principle of *symmetry breaking*. As Pierre Curie once enounced: *c'est la dissymmétrie qui crée le phénomène* (it is the lack of symmetry that creates the phenomenon). The world appears where the initial symmetry is broken, and the phenomena start to abound. This is accompanied by a decrease of temperature, which suggest that the high symmetry state is also highly energetic, and the spontaneous breaking of symmetry is driven by a decrease in energy. In this presentation, the focus will be on the theoretical aspects of the theorem, in particular group theory and topology.[1]

1.1 The distorted rutile structure

At the molecular level degeneracies are usually linked to the presence of symmetry, described by the molecular point groups. A textbook case from structural inorganic chemistry concerns the crystal structures of divalent transition-metal difluorides from CrF_2 to ZnF_2 [2]. These difluorides crystallize according to the rutile structure. Rutile is the mineral of TiO_2. In this structure the metal ions are surrounded by a regular octahedron of six ligands, at equal distances from the central atom. Cr(II) and Cu(II) ions are notable exceptions in the series. For these two metal ions the rutile lattice is distorted, forming a tetragonal coordination, with four equatorial ligands at short distance and two axial ones at longer distances, as indicated in Fig. 1. The figure also shows the crystal field configuration of the d-electrons, with one electron in the e_g shell for Cr(II) and one hole for Cu(II).

The *mean* value of these distances agrees with the expected trends for the d-metal contraction, but clearly some force is distorting the ligand sphere around the ion. What distinguishes these ions form the rest? These are the only two ions in the series for which the ground state configuration is characterized by an odd number of electrons in the e_g shell. The resulting ground states are ${}^2E_g$ multiplets, hence states that are orbitally twofold degenerate. They thus would exemplify the Jahn-Teller (JT) theorem, which states that such ground states are unstable, and will spontaneously distort to lower symmetries. The distortion will lift the degeneracy, and thus remove the cause of instability. Indeed symmetry breaking from O_h to D_{4h} will split the multiplets as follows

$$ {}^2E_g \rightarrow {}^2A_{1g} + {}^2B_{1g}. \tag{1}$$

Further interesting additional observations can also be made:

[1] The presentation, including several figures and formulas, is based on the recent monograph [1].

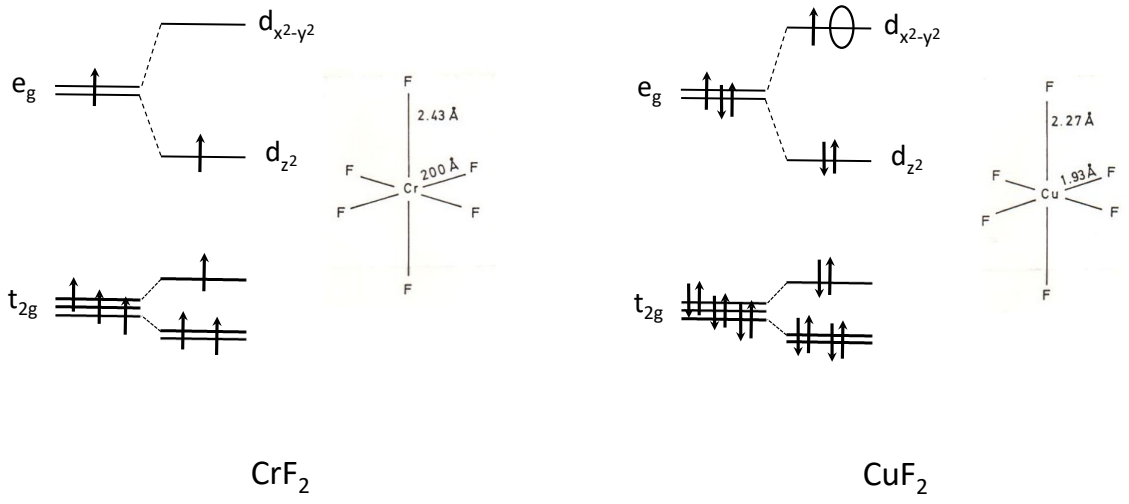


Fig. 1: *Jahn-Teller distortions in CrF_2 and CuF_2*

- The origin of the JT effect is clearly attributed to a local on-site orbital characteristic, which apparently is strong enough to distort the lattice structure. This inscribes the JT theorem in the broad theme of the lecture course.

- It could be argued that similar considerations would apply to the Fe(II) and Co(II) ions which have open t_{2g} shells giving rise to threefold degenerate ground states. These ions are indeed also exemplifying JT instabilities, but the instability is much smaller than in the case of the instabilities caused by the e_g shells. A further distinction is thus in order: the JT force can give rise to molecular structures which are frozen in a particular distorted geometry, or can be weaker and give rise to a vibronic ground state, with dynamic fluctuations. Such fluctuations show up as large anisotropic thermal structure factors in X-ray analysis. We will identify these two regimes as the static versus the dynamic JT effect. In reality however, systems will adopt all sorts of intermediate stages.

- Finally, the symmetry breaking itself is not complete, but rather tries to conserve as much symmetry as possible. Indeed the tetragonal subgroup is the maximal subgroup of O_h, for which d_{z^2} and $d_{x^2-y^2}$ are no longer degenerate. It removes the threefold axes that cause the degeneracy, but keeps all other symmetry elements. This economic principle is known as the epikernel principle.

1.2 Origin of orbital instability

Why is symmetry breaking a spontaneous process in degenerate states? The standard answer to this is that in these states there is always an imbalance between the symmetry of the nuclear charge distribution and the symmetry of the electron density. So electron densities of the individual e_g orbitals have only tetragonal symmetry, while the nuclear distribution is octahedral. The result is a force which acts on the nuclei and displaces them to a new equilibrium position

with D_{4h} symmetry. This argument is based on the fact that for non-degenerate states the electron density always adopts the symmetry of the nuclear frame. For a non-degenerate state $|\Psi\rangle$, a distortion force along a nuclear coordinate Q, is given by

$$F = \frac{\partial}{\partial Q}\langle\Psi|\mathcal{H}|\Psi\rangle|_{Q=0} = \left\langle\Psi\right| \frac{\partial\mathcal{H}}{\partial Q} \left|\Psi\right\rangle\Big|_{Q=0}. \tag{2}$$

If Ψ is non-degenerate, the density $\Psi^*\Psi$, is totally symmetric and the force matrix element can only differ from zero if the Hamiltonian part is likewise totally symmetric, i.e., if the Q-coordinate conserves the symmetry. When extending this argument to degenerate states, it is argued that the average density still is totally symmetric, but that this is no longer true for the density associated with individual components. The sum of the densities of the d_{z^2} and $d_{x^2-y^2}$ states is indeed equal along the three coordinate axes of an octahedron, but the separate densities of the two components is not: it is axial for the d_{z^2} orbital and equatorial for the $d_{x^2-y^2}$ counterpart. In this argument the assumption is made that the electron densities for individual components of degenerate states cannot possibly have the symmetry of the nuclear frame. In fact this is not true. For the twofold degenerate component it suffices to rewrite the components in complex conjugate form, to obtain for both an electron cloud with perfect octahedral symmetry.

$$|\Psi_\pm\rangle = \frac{1}{\sqrt{2}}\left(d_{z^2} \pm i d_{x^2-y^2}\right) \tag{3}$$

Indeed the densities of both these components are equal to the average density of d_{z^2} and $d_{x^-y^2}$, and thus totally symmetric. The real difference between degenerate and non-degenerate states is that in the case of degenerate states, the calculation of the distortion force requires to set up and diagonalize a matrix equation, operating in the degeneracy basis of the state manifold. If for instance we use the $\{\Psi_+, \Psi_-\}$ basis, the JT force will entirely be 'demoted' to the off-diagonal entries of the Hamiltonian matrix.

1.3 The Jahn-Teller Hamiltonian

The potential energy surface in the neighborhood of a JT instability is described by a Taylor series expansion of the Hamiltonian in the coordinate space of active nuclear distortions. The very first and essential terms of the expansion are the first-order force term and the harmonic second-order restoring potential

$$\mathcal{H} = H_0 + \sum_{\Lambda\lambda}\left(\frac{\partial H}{\partial Q_{\Lambda\lambda}}\right)_0 Q_{\Lambda\lambda} + \frac{1}{2}\sum_{\Lambda\lambda} K_\Lambda\, Q^2_{\Lambda\lambda}. \tag{4}$$

Here the distortion Q-coordinates are labeled by an irreducible representation Λ of the high-symmetry molecular point group, and its component or subrepresentation, λ. H_0 is the electronic Hamiltonian in the high-symmetry origin of the coordinate system, relaxed with respect to symmetry-preserving totally symmetric coordinates. Its eigenfunctions are the states of the degenerate manifold. The second-order term is the standard harmonic restoring force, with K_Λ being the harmonic force constant. This term holds the molecular frame together and attracts

the nuclei towards the coordinate origin. The force constant can be obtained from the IR and Raman spectra. The essential term is the linear term, which describes the interaction between the electronic states and the nuclear distortion modes. This linear interaction is the force which pulls the nuclei away from their original symmetry positions. The derivative in this term represents the slope of the energy as a function of the coordinate displacement, evaluated at the high-symmetry point. As a derivative of the Hamiltonian with respect to nuclear positions, this term affects the electron-nuclei Coulomb attraction term, and as a result it is a one-electron operator. This is an important property, which ultimately explains why the JT phenomenon is so tightly linked to orbital properties.

At this point a proper definition of the symmetry properties is in order. The coordinates have already been labeled as $Q_{\Lambda\lambda}$. Likewise the degenerate manifold will be labeled by the degenerate irreducible representation Γ, and its components accordingly by a subrepresentation label γ as $|\Psi^{\Gamma}_{\gamma}\rangle$. The symmetry labels incorporate the entire action of a symmetry element of the point group, $\hat{R} \in G$, on these quantities:

$$\hat{R}Q_{\Lambda\lambda} = \sum_{\lambda'} D^{\Lambda}_{\lambda'\lambda}(R)\, Q_{\Lambda\lambda'} \quad \text{and} \quad \hat{R}|\Psi^{\Gamma}_{\gamma}\rangle = \sum_{\gamma'} D^{\Gamma}_{\gamma'\gamma}(R)\, |\Psi^{\Gamma}_{\gamma'}\rangle. \tag{5}$$

Here the D-matrix elements refer to the irreducible representation (irrep) matrices $\mathbb{D}(R)$ which describe the transformation of the basis functions under all the elements of the symmetry group. What makes the JT Hamiltonian tractable, and in fact extremely attractive, is that instead of working in the entire Hilbert space, it operates in an extremely confined space, comprising at first only the degenerate manifold. Matrix elements of the linear interaction term in this manifold may be factorized according to the Wigner-Eckart theorem as a reduced force element, denoted by the constant F_{Λ}, and a Clebsch-Gordan coupling coefficient, which contains the entire group-theoretical knowledge of the interaction

$$\left\langle \Psi^{\Gamma}_{\gamma_a} \left| \left(\frac{\partial H}{\partial Q_{\Lambda\lambda}} \right)_0 \right| \Psi^{\Gamma}_{\gamma_b} \right\rangle = F_{\Lambda} \left\langle \Gamma\gamma_a | \Lambda\lambda\Gamma\gamma_b \right\rangle. \tag{6}$$

In the second-quantization formalism, we now introduce creation and annihilation operators for the electronic states. Since these are fermionic in nature we label them respectively as $f^{\dagger}$ and f. A normalized N-electron determinant is obtained as a sequence of particles being created from the vacuum state

$$f^{\dagger}_{\alpha} f^{\dagger}_{\beta} ... f^{\dagger}_{\nu} \left|0\right\rangle \equiv \left|\alpha\beta...\nu\right|. \tag{7}$$

The adjoint of this expression is then

$$\left\langle 0\right| f_{\nu} ... f_{\beta} f_{\alpha} = \overline{\left|\alpha\beta...\nu\right|}. \tag{8}$$

Since the linear part of the Hamiltonian involves a one-electron operator, we can express the coupling in operator form as

$$\left(\frac{\partial H}{\partial Q_{\Lambda\lambda}} \right)_0 = \sum_{\gamma_a\gamma_b} f^{\dagger}_{\gamma_a} \left\langle \Psi^{\Gamma}_{\gamma_a} \left| \left(\frac{\partial H}{\partial Q_{\Lambda\lambda}} \right)_0 \right| \Psi^{\Gamma}_{\gamma_b} \right\rangle f_{\gamma_b}. \tag{9}$$

An alternative view point of the interaction involves a recoupling, where the fermion parts are first coupled to an excitation operator with symmetry Λ. This recoupling is carried out by transferring the $\Gamma_b\gamma_b$ irrep of the ket part to the bra, and corresponds to a basic symmetry property of the coupling coefficients [1]. One has, apart from an overall Λ-dependent phase factor which can be incorporated into the force-parameter,

$$\langle \Gamma\gamma_a|\Lambda\lambda\Gamma\gamma_b\rangle = \left(\frac{\dim\Gamma}{\dim\Lambda}\right)^{1/2}\langle\Gamma\gamma_a\Gamma\overline{\gamma_b}|\Lambda\lambda\rangle. \tag{10}$$

Note that the symmetry properties of the annihilation operator, $\Gamma\gamma_b$, appear in the coupling operator as the complex conjugate component, in view of the transfer from ket to bra. Then, these results are inserted into the operator expression, yielding

$$\left(\frac{\partial H}{\partial Q_{\Lambda\lambda}}\right)_0 = k_\Lambda \sum_{\gamma_a\gamma_b} f^\dagger_{\gamma_a}\langle\Gamma\gamma_a\Gamma\overline{\gamma_b}|\Lambda\lambda\rangle f_{\gamma_b}, \tag{11}$$

where k_Λ takes over the role of the F_Λ force elements, by incorporating the dimensional factor

$$k_\Lambda = \left(\frac{\dim\Gamma}{\dim\Lambda}\right)^{1/2} F_\Lambda \tag{12}$$

Vice-versa, since this is a summation over all components, one could as well remove the complex conjugate bar from the coupling coefficient and replace the annihilation operator, f_{γ_b} by its time reversed form, which is denoted by the tilde operator as $\tilde{f}_{\gamma_b}$. The tilde indicates that the annihilation operator $\tilde{f}_{\gamma_b}$ transform in exactly the same way as the corresponding creation operator $f^\dagger_{\gamma_b}$, and as the time reversed of the annihilation operator f_{γ_b}. The operator expression then finally becomes

$$H = k_\Lambda \sum_{\gamma_a\gamma_b}\langle\Gamma\gamma_a\Gamma\gamma_b|\Lambda\lambda\rangle f^\dagger_{\gamma_a}\tilde{f}_{\gamma_b} = k_\Lambda\left(\mathbf{f}^\dagger\mathbf{f}\right)^\Lambda_\lambda. \tag{13}$$

The bracket in the final line of this equation symbolizes the coupling of the fermion creation and annihilation operators to the symmetry of the boson. In this formalism the slope parameter is usually represented as k_Λ. In second quantization we now also add the vibrational mode, expressed in boson creation and annihilation operators

$$Q_{\Lambda\lambda} = \frac{1}{\sqrt{2}}\left(b^\dagger_{\Lambda\lambda} + \tilde{b}_{\Lambda\lambda}\right). \tag{14}$$

Again note the tilde over the annihilation operator. Indeed, both creation and annihilation parts must share the $\Lambda\lambda$ symmetry properties of $Q_{\Lambda\lambda}$. In order to combine the fermionic and bosonic parts it must be taken into account that this involves a scalar product over the Λ tensor, as a fermionic variable and an associated bosonic derivative. Since derivatives and variables transform in conjugate ways, one must write

$$\sum_{\Lambda\lambda}\left(\frac{\partial H}{\partial Q_{\Lambda\lambda}}\right)_0 Q_{\Lambda\lambda} = \sum_\Lambda \kappa_\Lambda\left(\mathbf{f}^\dagger\mathbf{f}\right)^\Lambda \odot \left(\mathbf{b}^\dagger{+}\mathbf{b}\right)_\Lambda. \tag{15}$$

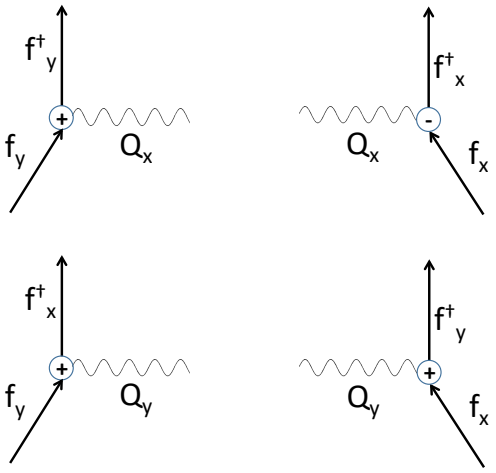


Fig. 2: *Diagram of coupling schemes for the JT matrix elements in the $E \times e$ case (*vide infra*); $f_x^\dagger$ and $f_y^\dagger$ create an electron in resp. d_{z^2} and $d_{x^2-y^2}$ orbitals.*

The dot refers to the scalar product of boson and fermion part which guarantees the total symmetry of the Hamiltonian, due to compensating symmetries in both ingredients. When components follow the spherical (l, m) quantization, the dot product is defined as

$$\left(\mathbf{f}^\dagger\mathbf{f}\right)^l \odot \left(\mathbf{b}^\dagger+\mathbf{b}\right)_l = \sum_m (-1)^m \left(\mathbf{f}^\dagger\mathbf{f}\right)^l_m \left(b^\dagger_{l,-m}+\tilde{b}_{l,-m}\right). \tag{16}$$

The concise second-quantization formalism in Eq. (15) says it all! The fermion creation-annihilation double operator is exactly an excitation operator which requires a field of symmetry $\Lambda\lambda$. This is symbolized for the E-case in Fig. 1.3.
The difference with a proper excitation is that instead of a photon the excitation is brought about by a vibration. To this interaction element one finally adds the harmonic part of the active vibrations. This complements the potential energy of the JT surface with the kinetic energy of the nuclei. The harmonic potential is now replaced by the harmonic oscillator

$$\sum_{\Lambda\lambda} \hbar\omega_\Lambda \Big(b^\dagger_{\Lambda\lambda} b_{\Lambda\lambda} + \frac{1}{2}\Big). \tag{17}$$

The result is a genuine vibronic operator where bosons and fermions meet

$$\mathcal{H} = \sum_\Lambda \kappa_\Lambda \left(\mathbf{f}^\dagger\mathbf{f}\right)^\Lambda \odot \left(\mathbf{b}^\dagger+\mathbf{b}\right)_\Lambda + \sum_{\Lambda\lambda} \hbar\omega_\Lambda \Big(b^\dagger_{\Lambda\lambda} b_{\Lambda\lambda} + \frac{1}{2}\Big). \tag{18}$$

1.4 The pseudo-Jahn-Teller effect

When two electronic states are not strictly degenerate but close together in energy, it should be very surprising that the symmetry breaking mechanism would suddenly be completely quenched. Instead a non-totally symmetric matrix element between both states is symmetry allowed and

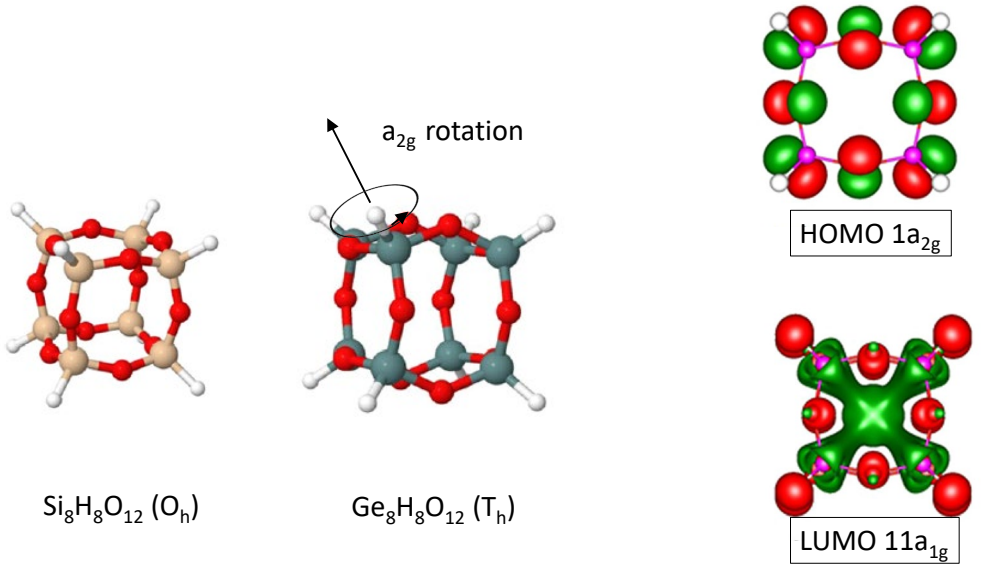


Fig. 3: *Structural comparison between Si and Ge POSS (left), and, HOMO and LUMO for the Ge cluster (right).*

may perfectly well induce a distortion, providing the relaxation term outweighs the harmonic force constant. This is the so-called pseudo-JT effect. Following the formalism of Bersuker [3], let two states be separated by a splitting 2Δ and with an off-diagonal force element FQ, where Q is a non-totally symmetric distortion coordinate. Assume further that the two states share the same force constant K_0. In that case matrix diagonalization leads to the roots

$$E_{\pm} = \frac{1}{2}K_0Q^2 \pm \left(\Delta^2+F^2Q^2\right)^{1/2} = \frac{1}{2}\left(K_0 \pm \frac{F^2}{\Delta}\right)Q^2 \pm \Delta \mp \left(\frac{F^4}{\Delta^3}\right)Q^4 \pm \ldots \qquad (19)$$

If $|\Delta| < F^2/K_0$, then the curvature of the lower energy root becomes negative, and the system will be unstable with respect to Q. An exceptional illustration of this effect is seen in the $O_h \rightarrow T_h$ symmetry breaking in the polyhedral oligomeric sesquioxane (POSS), $Ge_8H_8O_{12}$. While the Silicon isomer has cubic symmetry O_h, it is found by DFT calculations that the Germanium isomer is distorted to the rare tetrahedral symmetry group T_h [4]. In Fig. 3 we display both structures, as well as the HOMO ($1a_{2g}$) and LUMO ($11a_{1g}$) of the Germanium isomer. The off-diagonal matrix element between both orbitals transforms as the direct product: $a_{1g}\times a_{2g} = a_{2g}$. The pseudo-JT effect thus promotes a distortion along the a_{2g} mode. This corresponds precisely to a rotation of the oxygen bridges in between the Germanium atoms. Neighboring vertices on the cube will thereby rotate in opposite directions lowering the symmetry to T_h.

An important caveat is in order here. In principle, for any symmetry breaking it will always be possible to find a pair of interacting states with the right combination of irreducible representations. So the predictive power of the effect is rather limited. A detailed examination of the composition of the relevant orbitals, and a demonstration of the overlap of their off-diagonal density and an observed distortion is required.

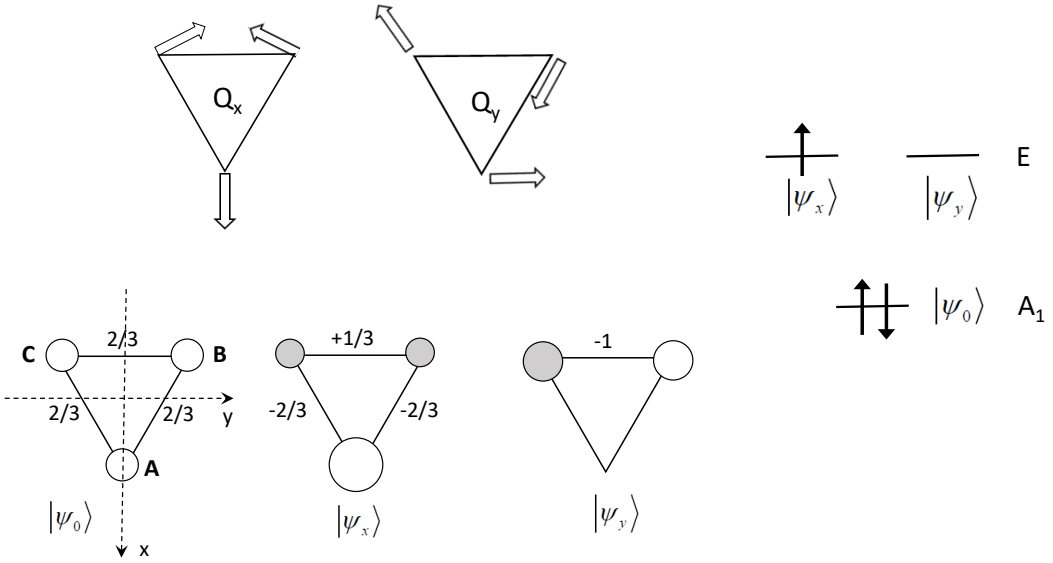


Fig. 4: *Trigonal Na_3 cluster, with doublet ground level; orbitals and distortion modes.*

2 The doublet $E\times e$ Paradigm

The icon of the JT theorem is the Mexican hat potential, corresponding to a twofold degenerate E state, coupled to a twofold degenerate e vibration. This occurs both in cubic and in trigonal or pentagonal symmetry groups. We examine in some detail the standard case of a triangular instability.

2.1 The potential energy surface

The system considered is a tri-atomic molecule in an E state, with components E_x and E_y. The symmetry at the origin is D_{3h}, but since three atoms are coplanar, we could as well work in C_{3v} symmetry. The components are represented schematically in Fig. 4. Their symmetry behavior under the generators of C_{3v} (with right-handed threefold axis) are given by

$$\begin{aligned}
\hat{C}_3 \begin{pmatrix} |Ex\rangle & |Ey\rangle \end{pmatrix} &= \begin{pmatrix} |Ex\rangle & |Ey\rangle \end{pmatrix} \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \\
\hat{\sigma}_x \begin{pmatrix} |Ex\rangle & |Ey\rangle \end{pmatrix} &= \begin{pmatrix} |Ex\rangle & |Ey\rangle \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (20)
\end{aligned}$$

Here an active view of symmetry operations is adopted: they displace the functions itself, be it orbitals or distortions, while leaving the nuclei in place.
The direct product of the orbital state reads

$$E \times E = \left[a_1{+}e\right] + a_2. \qquad (21)$$

According to the JT selection rule the activity resides in the non-totally symmetric part of the symmetrized product, $[a_1{+}e]$, being the e-vibration. The components of this vibration are also

shown in the figure. They are labeled as Q_x and Q_y. Using local (x, y) coordinates for the individual atoms, the expressions for these vibrations are given by:

$$\begin{aligned} Q_x &= \frac{1}{\sqrt{3}}\left[Q_A^x + \left(-\frac{\sqrt{3}}{2}Q_B^y - \frac{1}{2}Q_B^x\right) + \left(\frac{\sqrt{3}}{2}Q_C^y - \frac{1}{2}Q_C^x\right)\right] \\ Q_y &= \frac{1}{\sqrt{3}}\left[Q_A^y + \left(-\frac{1}{2}Q_B^y + \frac{\sqrt{3}}{2}Q_B^x\right) + \left(-\frac{1}{2}Q_C^y - \frac{\sqrt{3}}{2}Q_C^x\right)\right]. \end{aligned} \tag{22}$$

The action of the group generators on these functions is given by

$$\begin{aligned} \hat{C}_3\begin{pmatrix} Q_x & Q_y \end{pmatrix} &= \begin{pmatrix} Q_x & Q_y \end{pmatrix}\begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \\ \hat{\sigma}_x\begin{pmatrix} Q_x & Q_y \end{pmatrix} &= \begin{pmatrix} Q_x & Q_y \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{23}$$

Note the sign change here as compared to Eq. (20). This is based on the convention that these modes were chosen to mimic the behavior of central quadrupolar harmonics x^2-y^2 and xy, as opposed to the fermion states which follow the dipolar harmonics x and y. With K_e the force constant of the boson mode, and F_e the linear force element, in a fermion basis $\{|x\rangle|y\rangle\}$ the Hamiltonian reads

$$\mathcal{H} = \frac{K}{2}\left(Q_x^2 + Q_y^2\right) + \frac{F_e}{\sqrt{2}}\begin{pmatrix} Q_x & Q_y \\ Q_y & -Q_x \end{pmatrix}. \tag{24}$$

Diagonalization of this Hamiltonian then yields the familiar Mexican hat surface, consisting of two parabolic sheets, with rotational symmetry along the threefold axis:

$$E_\pm = \frac{K}{2}\left(Q_x^2 + Q_y^2\right) \pm \frac{F_e}{\sqrt{2}}\sqrt{Q_x^2 + Q_y^2}\,. \tag{25}$$

The central C_{3v} point of the diagram is unstable, and the energy gain by distortion into the trough is the so-called JT energy, given by

$$E_{JT} = -\frac{1}{2}\frac{F_e^2}{K}\,. \tag{26}$$

If the system rotates around in the trough the nuclei perform circular motions around the trigonal equilibrium positions. This motion is an internal rotation or *libration* as shown in Fig. 5. Additional higher-order terms to the Hamiltonian will essentially maintain the shape of the surface, but introduce warping. As an example, the second-order terms in Q_xQ_y and $Q_x^2 - Q_y^2$ will warp the potential energy surface, giving rise to local hill tops, alternating with local minima. The stationary points correspond to isosceles triangles. Detailed calculations by Cocchini *et al.* [5] for the sodium trimer yield a JT stabilization energy in the order of 670 cm^{-1}, and a rotational barrier of 130 cm^{-1}.

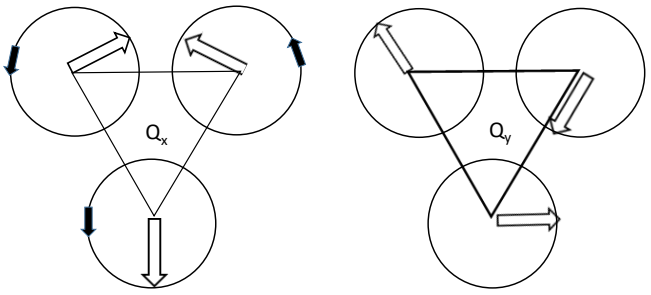


Fig. 5: *Internal rotation along the trough of the Mexican hat; a 90° anti-clockwise rotation takes the Q_x distortion to Q_y.*

2.2 The dynamic system

A fascinating aspect of the Mexican hat potential is certainly its obvious rotational symmetry. This symmetry ultimately goes back to the unitary symmetry of the diabolical degeneracy point at the origin. For a full grasp of this symmetry, we now rewrite the Hamiltonian in its dynamic form, including the nuclear kinetic energy term. According to the standard boson-fermion formalism, the ket functions are generated by the $f_x^\dagger, f_y^\dagger$ operators, and the boson modes are created by $b_x^\dagger, b_y^\dagger$, with coordinate and momentum operators as

$$Q_x = \frac{1}{\sqrt{2}}\left(b_x^\dagger + b_x\right) \quad \text{and} \quad P_x = \frac{i}{\sqrt{2}}\left(b_x^\dagger - b_x\right). \tag{27}$$

A unit of length is defined as $\sqrt{\hbar/m\omega}$ and the oscillator quantum $\hbar\omega$ is taken as the unit of energy. This rescaling absorbs all fundamental constants:

$$\mathcal{H} = \begin{pmatrix} b_x^\dagger b_x + b_y^\dagger b_y + 1 + \kappa\left(b_x^\dagger + b_x\right) & \kappa\left(b_y^\dagger + b_y\right) \\ \kappa\left(b_y^\dagger + b_y\right) & b_x^\dagger b_y + b_y^\dagger b_y + 1 - \kappa\left(b_x^\dagger + b_x\right) \end{pmatrix}. \tag{28}$$

Here, κ is the linear coupling parameter, and 1 is the zero-point energy. *Subsequently this is taken out as the zero of the energy scale.* The angular momentum associated with a rotation in (Q_x, Q_y) is given by

$$\hat{\mathcal{L}}_z = Q_x P_y - Q_y P_x = \frac{i}{2}\Big(\left(b_x^\dagger + b_x\right)\left(b_y^\dagger - b_y\right) - \left(b_y^\dagger + b_y\right)\left(b_x^\dagger - b_x\right)\Big) = i\left(b_y^\dagger b_x - b_x^\dagger b_y\right). \tag{29}$$

To find out the rotational symmetry of the Hamiltonian we calculate the commutator with $\mathcal{L}_z$

$$\left[\hat{\mathcal{L}}_z, \mathcal{H}\right] = i \begin{pmatrix} \kappa\left(b_y^\dagger + b_y\right) & -\kappa\left(b_x^\dagger + b_x\right) \\ -\kappa\left(b_x^\dagger + b_x\right) & -\kappa\left(b_y^\dagger + b_y\right) \end{pmatrix}. \tag{30}$$

Unexpectedly perhaps, the two operators do not commute! However we should be aware that the Hamiltonian describes a coupled situation where both boson and fermion fields are affected. To this aim, we introduce an angular coordinate ϕ in distortion space, with $Q_x = Q\cos\phi$

and $Q_y = Q\sin\phi$. The ground state wavefunction (with $\kappa<0$) of the static Hamiltonian as a function of ϕ is given by

$$|\psi_-\rangle = \cos\frac{\phi}{2}\,|x\rangle + \sin\frac{\phi}{2}\,|y\rangle. \tag{31}$$

This shows the rotation of the wavefunction along the trough, but at half speed as compared to the coordinate change. The wavefunction provides a connection between a base space, providing the real distortions of the system, and a function space, which for every point in the base space, gives a fermion vector. As the boson vector is a direct product of the fermion vector (remember $e \in [E\times E]$), we can qualify the fermion space as a fundamental spin space, and the boson space on top of that as a coupled vector space. The geometry of this connection will be examined in the next section. Here it suffices to define a rotation operator for the fermion states in analogy with the pseudo-spin operator $\mathcal{S}_z$

$$\hat{\mathcal{S}}_z = \frac{i}{2}\big(f_y^\dagger f_x - f_x^\dagger f_y\big). \tag{32}$$

Pursuing this analogy with spin-orbit coupling further, we can define the total momentum operator as $\hat{\mathcal{J}}_z$ by

$$\hat{\mathcal{J}}_z = \hat{\mathcal{L}}_z + \hat{\mathcal{S}}_z\,. \tag{33}$$

This sum operator commutes with the Hamiltonian, as the sum of the commutator of $\hat{\mathcal{S}}_z$ and the commutator with the boson part cancels out: $[\mathcal{S}_z,\mathcal{H}] = -[\mathcal{L}_z,\mathcal{H}]$. In order to take advantage of the conservation of angular momentum, we now impose symmetry adapted combinations of bosons and fermions. One has

$$b_\pm^\dagger = \frac{1}{\sqrt{2}}\left(b_x^\dagger \pm i b_y^\dagger\right) \quad \text{and} \quad b_\pm = \frac{1}{\sqrt{2}}\left(b_x \mp i b_y\right). \tag{34}$$

These operators are eigenoperators of $\hat{\mathcal{L}}_z$ with opposite eigenvalues

$$\left[\hat{\mathcal{L}}_z, b_\pm^\dagger\right] = \pm b_\pm^\dagger \quad \text{and} \quad \left[\hat{\mathcal{L}}_z, b_\pm\right] = \mp b_\pm\,. \tag{35}$$

Analogous symmetry adaptation of the fermion operators yields

$$|\uparrow\rangle = \frac{1}{\sqrt{2}}\big(|x\rangle + i|y\rangle\big) \quad \text{and} \quad |\downarrow\rangle = \frac{1}{\sqrt{2}}\big(|x\rangle - i|y\rangle\big). \tag{36}$$

As eigenfunctions of the $\hat{\mathcal{S}}_z$ operator, these combinations are like α and β spins

$$\hat{\mathcal{S}}_z|\uparrow\rangle = +\frac{1}{2}|\uparrow\rangle \quad \text{and} \quad \hat{\mathcal{S}}_z|\downarrow\rangle = -\frac{1}{2}|\downarrow\rangle. \tag{37}$$

The total symmetry-adapted Hamiltonian is now expressed in the transformed fermion basis

$$\begin{pmatrix} |\uparrow\rangle \\ |\downarrow\rangle \end{pmatrix}: \qquad \mathcal{H} = \begin{pmatrix} b_+^\dagger b_+ + b_-^\dagger b_- & \kappa\sqrt{2}\big(b_-^\dagger + b_+\big) \\ \kappa\sqrt{2}\big(b_+^\dagger + b_-\big) & b_+^\dagger b_+ + b_-^\dagger b_- \end{pmatrix}. \tag{38}$$

To solve this Hamiltonian equation it is of paramount importance to define an Ansatz. An Ansatz is a general expression of the form of the solution, which holds the symmetry of the

system, and expresses the coupling scheme between the boson and fermion degrees of freedom. The Ansatz reads

$$|\Psi\rangle_{l+1/2} = \left(b_+^\dagger\right)^l \Phi_1(\xi)\,|\uparrow\rangle + \left(b_+^\dagger\right)^{l+1} \Phi_2(\xi)\,|\downarrow\rangle. \tag{39}$$

Here the variable ξ is defined as

$$\xi = b_+^\dagger b_-^\dagger. \tag{40}$$

This variable thus corresponds to a two-photon boson excitation, combing two excitations with opposite angular momentum. The total angular momentum of this variable is thus equal to zero, and it can be considered as a double purely radial excitation. The Ansatz shows that in order to obtain a vibronic state with angular momentum $l+1/2$, on top of an arbitrary number of radial excitations we can either excite l quanta of $b_+^\dagger$ and couple this to a spin-up fermion state, or excite $l+1$ quanta of $b_+^\dagger$ and couple this with a spin-down fermion. These are the only two channels to arrive at a state with the desired momentum. This state will always be degenerate with a time-reversed counterpart, which is given by

$$|\Psi\rangle_{-l-1/2} = \left(b_-^\dagger\right)^l \Phi_1(\xi)\,|\downarrow\rangle + \left(b_-^\dagger\right)^{l+1} \Phi_2(\xi)\,|\uparrow\rangle. \tag{41}$$

The Ansatz clearly shows that the vibronic wavefunction cannot be factorized as a product of a fermion and a boson part: we have definitely taken leave from the Born-Oppenheimer approximation. In summary the JT equations to be solved read in matrix form

$$\mathcal{H}|\Psi\rangle = E|\Psi\rangle = \begin{pmatrix} b_x^\dagger b_x + b_y^\dagger b_y + \kappa\left(b_x^\dagger + b_x\right) & \kappa\left(b_y^\dagger + b_y\right) \\ \kappa\left(b_y^\dagger + b_y\right) & b_x^\dagger b_y + b_y^\dagger b_y - \kappa\left(b_x^\dagger + b_x\right) \end{pmatrix} \begin{pmatrix} \left(b_+^\dagger\right)^l & \Phi_1(\xi) \\ \left(b_+^\dagger\right)^{l+1} & \Phi_2(\xi) \end{pmatrix}. \tag{42}$$

We refer to [1] for a detailed discussion of the solution of these equations. Interestingly the equations can ultimately be turned into a form of Heun's differential equation. Closed solutions of this equation do not seem to exist, except for some special values of κ. Eigenvalues are characterized by half integral values of j and are plotted as a function of the coupling parameter, in close-up in Fig. 6.

At the left of the diagram, for $\kappa = 0$, are found the oscillator levels of the e-vibration. When the coupling sets in, the trough develops, and ultimately – in the strong coupling limit – the spectrum reduces to a rotational spectrum with a regular sequence of half-integral j-values, superimposed on a transversal oscillator. Looking in detail at the lowest vibronic levels, in the limit of zero coupling strength the ground level with $j = 1/2$ reduces to the product of the electronic degeneracy and the totally-symmetric zero phonon state. The excited oscillator state at $E = 1$ corresponds to the vector addition of an $l = 1$ vibrational level to the fermion spin, yielding $j = 1/2,\, 3/2$. As the coupling is turned on, the $j = 1/2$ excited state is raised due to its interaction with the equisymmetric ground state, while the $j = 3/2$ level is expected to descent in energy, as seen in the figure.

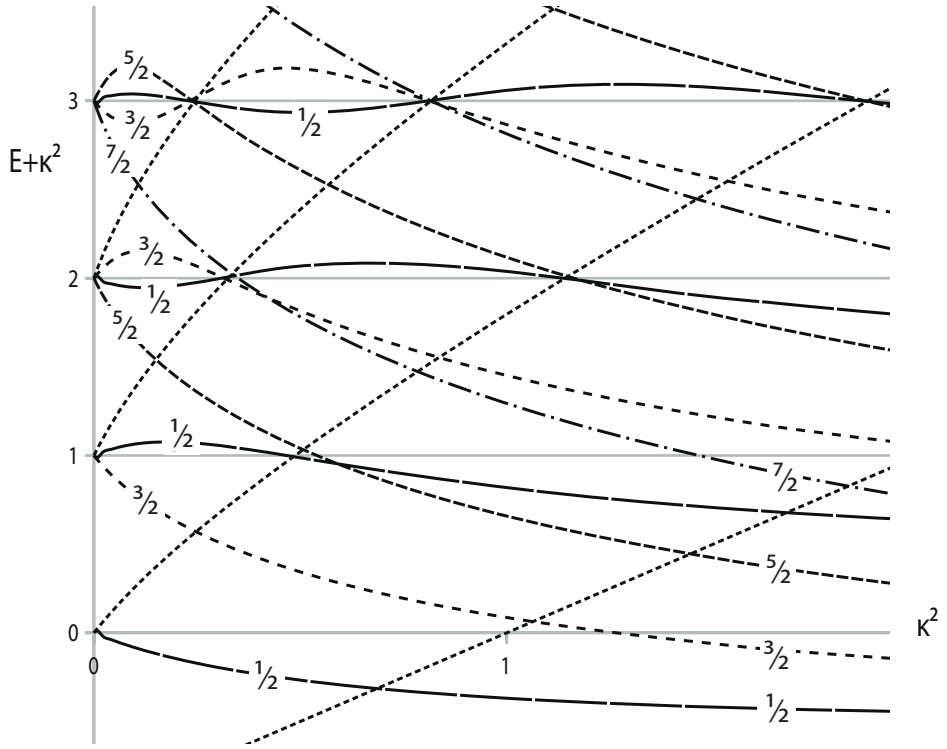


Fig. 6: *$E \times e$ JT Hamiltonian: solutions of the dynamic equation as a function of κ^2; individual lines are characterized by angular momentum j; ascending dotted lines represent extra mathematical solutions that are unphysical.*

If quadratic warping terms are introduced, the rotational symmetry is broken to C_{3v}. Accordingly, the j states subduce trigonal levels as indicated below

$$\begin{aligned} j = 1/2 &\rightarrow E \\ j = 3/2 &\rightarrow A_1 + A_2 \,. \end{aligned} \tag{43}$$

In a strong coupling regime with extensive trigonal warping, the vibronic regime in essence reduces to local oscillations in three localized wells. Depending on the signs of the warping parameters, the minima are either at $\phi = 0°, 120°, 240°$ with saddle points in between, or vice-versa. Small vibrational overlap between these wells opens the possibility of tunneling.The lowest tunneling states are obtained by setting up a 3×3 hopping matrix between the wells. The matrix element between the wells essentially is a Huang-Rhys overlap factor, with a positive sign. The result is a two state diagram, with an E ground state, and an A excited state, which are separated by a tunneling splitting. Here the lower E state correlates with the ground $j = 1/2$ level of the diagram. The upper A state correlates likewise with the $j = 3/2$ parentage. As Eq. (43) shows, it can be either A_1 or A_2. For wells located at turning points $0°, 120°, 240°$, the A level is identified as an A_1 level. When the surface is turned upside down, with minima now at $60°, 180°, 300°$, the A level has A_2 symmetry. These results follow from the electronic part of the wavefunction in Eq. (31), as the vibrational overlap is symmetric with respect to reflections in C_{3v} [6].

2.3 Berry phase

We already drew attention to the sign change of the wavefunction after revolving around the conical intersection. The acquired phase is a geometric phase, which is generally known as a Berry phase, following the seminal work of Michael Berry [7]. Berry's phase was identified with the concept of holonomy in geometry. To present this concept, two ingredients are required: the base space, and the fiber. In the JT case the base space is the coordinate space $\{Q_x, Q_y\}$ formed by the two distortion modes. With each point in the base space a wavefunction can be associated. The phase of this wavefunction may vary over a range $0, 2\pi$. The phase variable forms a so-called fiber, associated with a particular point on the base space. The collection of all these fibers over the entire base space forms a fiber bundle. Now the holonomy is what we observe in the fiber bundle when a closed loop is performed in the base space. Clearly, in order to be meaningful, a connection must exist which controls the change of the phase in consecutive fibers, corresponding to adjacent points on the base space. Berry showed how this connection is provided by the time-dependent Schrödinger equation, under adiabatic constraints. This is fulfilled in the case of a circuit driven by slowly moving nuclei along the trough of the potential, with instantaneous adaptation of the wavefunction, not involving excitations. Ideally we may think of a slow rotation which is hindered by the surface warping along the circuit. The treatment proceeds as follows: let $|n(\mathbf{R})\rangle$ represent the non-degenerate quantum state of a system, dependent on external parameters $\mathbf{R}$, which corresponds to a particular nuclear configuration along the low-energy trough. The eigenvalue is given by

$$\mathcal{H}(\mathbf{R})|n(\mathbf{R})\rangle = E_n(\mathbf{R})|n(\mathbf{R})\rangle. \tag{44}$$

The wavefunction $|n(\mathbf{R})\rangle$ must be *single valued* in the relevant parameter domain, and be *differentiable*. The wavefunction which solves the time-dependent Schrödinger equation in the adiabatic regime is then given by

$$|\Psi\rangle = \exp\left(-\frac{iE_n}{\hbar}t\right)|n(\mathbf{R})\rangle. \tag{45}$$

Here a time-dependent phase factor, the so-called dynamical phase, is added. This factor measures the passage of time. In the JT application we consider a closed circuit, C, in a space defined by nuclear displacements, $\mathbf{R}(t)$, where the distortion varies smoothly and slowly in time, as the nuclei evolve on a minimal energy path. Since the adiabatic state depends on the coordinates, it will change accordingly, but continuously, i.e., the Hamiltonian does not change rapidly enough to allow excitations to other states with energy $E_m(\mathbf{R})$. Nonetheless, by slowly driving the state around in the distortion space, an extra time dependence will appear, as is evident from the notation $|n(\mathbf{R}(t))\rangle$. In order to keep satisfying the time dependent equation, we must include a further compensatory phase factor. Eq. (45) is thus rewritten as

$$|\Psi\rangle = \exp\left(-\frac{iE_n}{\hbar}t\right)\exp\Big(i\gamma_n(t)\Big)|n(\mathbf{R}(t))\rangle. \tag{46}$$

Here the second exponential represents the geometric phase that is at the core of Berry's treatment. Applying the time dependent equation yields

$$i\hbar\frac{d}{dt}\,|\Psi\rangle = E_n|\Psi\rangle - \hbar\frac{d\gamma_n}{dt}\,|\Psi\rangle + i\hbar\exp\left(-\frac{iE_n}{\hbar}t\right)\exp\Big(i\gamma_n(t)\Big)\frac{d}{dt}|n(\mathbf{R}(t))\rangle. \tag{47}$$

In order to satisfy the Schrödinger equation, one must require that the sum of the second and third terms cancel

$$-\frac{d\gamma_n}{dt}|\Psi\rangle + i\exp\left(-\frac{iE_n}{\hbar}t\right)\exp\Big(i\gamma_n(t)\Big)\frac{d}{dt}|n(\mathbf{R}(t))\rangle = 0\,. \tag{48}$$

This can be rewritten as

$$d\gamma_n = i\langle n(\mathbf{R})|dn(\mathbf{R})\rangle = i\langle n(\mathbf{R})|\nabla_R|n(\mathbf{R})\rangle\cdot\mathrm{d}\mathbf{R} \tag{49}$$

When completing a closed loop, the total build-up of the phase is measured by the line integral along the path

$$\gamma_n(C) = \oint_C d\gamma_n = i\oint\langle n|dn\rangle, \tag{50}$$

with $|dn\rangle = \nabla_R|n\rangle\cdot\mathrm{d}\mathbf{R}$. Furthermore since the ket function is normalized, one has

$$d\langle n|n\rangle = \langle dn|n\rangle + \langle n|dn\rangle = \overline{\langle n|dn\rangle} + \langle n|dn\rangle = 0\,. \tag{51}$$

This implies that the matrix element $\langle n|dn\rangle$ is purely imaginary, and thus that $\gamma_n(C)$ will be real. This integral is the famous Berry phase. If the path is defined on a curved surface this phase will be non-trivial. In order to apply this treatment to the JT system, it is first of all noted that the electronic wavefunction $|\psi_-\rangle$ given in Eq. (31) is not single-valued, since

$$|\psi_-(2\pi)\rangle = \exp(i\pi)|\psi_-(0)\rangle. \tag{52}$$

So $|\psi_-\rangle$ does not correspond to $|n(\mathbf{R})\rangle$. However, by gradually removing the phase of π during the circuit, we obtain the required single-valued function

$$|n(\mathbf{R})\rangle = \exp\left(-\frac{i\phi}{2}\right)|\psi_-(\phi)\rangle = \exp\left(-\frac{i\phi}{2}\right)\Big(\cos\phi/2|E_x\rangle + \sin\phi/2|E_y\rangle\Big). \tag{53}$$

And thus

$$\begin{aligned} d|n(\mathbf{R})\rangle &= \exp\left(-\frac{i\phi}{2}\right)\left(-\frac{id\phi}{2}|\psi_-(\phi)\rangle + d|\psi_-(\phi)\rangle\right) \\ \langle n(\mathbf{R})|dn(\mathbf{R})\rangle &= -\frac{i\,d\phi}{2}\,. \end{aligned} \tag{54}$$

Here we made use of the fact that $|\psi_-\rangle$ is real, and hence

$$d\langle\psi|\psi\rangle = 2\langle\psi|d\psi\rangle = 0\,. \tag{55}$$

Inserting the result in Eq. (50) yields

$$\gamma_n(C) = \oint_C d\gamma_n = i\oint\langle n|dn\rangle = i\oint\Big(-\frac{i}{2}\Big)d\phi = \pi\,. \tag{56}$$

As Berry writes, *one might say that the dynamical phase factors in Eq.* (45) *and* γ_n *in Eq.* (46) *give the system's best answers to two questions about its adiabatic circuit. For the dynamical phase the question is: how long did your journey take? For* $\gamma_n(C)$ *it is: where did you go to?* Here we open a brief parenthesis: as the integral $\langle\psi|d\psi\rangle$ is zero, the function $|\psi\rangle$ is said to follow the law of parallel transport. It means that the change of the function is orthogonal to the function itself. This implies that the function accumulates during its path the torsion that is forced upon the system by the path, and as a result its end state after a full circuit will end up with a net phase difference. Following a function under parallel transport and detecting the phase change after a full circuit is thus a direct way to obtain the Berry phase.
Now what are the implications of the Berry phase for the JT treatment? In Eq. (49) it is noted that the gradient element adds an extra phase to the wavefunction, exactly as the vector potential $\mathbf{A}$ does to a charged particle in magnetism. In view of this analogy, we may introduce a vector field terminology, and write

$$\mathbf{A} = i\langle n(\mathbf{R})|\nabla_R|n(\mathbf{R})\rangle \tag{57}$$

and

$$\gamma_n(C) = \oint_C \mathbf{A}\cdot \mathrm{d}\mathbf{R}\,. \tag{58}$$

$\mathbf{A}$ lives in parameter space, and emanates from the topology of this space. As it is dependent on the phase of the basis vectors, it is not unique, and when applying the formula in Eq. (57), one must make sure that the basis vector is locally single-valued. In the JT case working out these expressions yields

$$A_\phi = i\big\langle n\big|\,\frac{\delta}{\delta\phi}\,\big|n\big\rangle = \frac{1}{2}\,. \tag{59}$$

The form of this vector potential is analogous to the field created by a Dirac monopole of strength $1/2$. The source of this monopole is nothing else than the conical intersection itself. The question thus arises if the dynamic calculations which we performed indeed include the vector potential associated with the conical intersection, or if an additional field term in the Hamiltonian is required. The short answer is that the dynamic treatment, which we have presented, does indeed contain the Berry phase from the start, so there was no need to invoke it afterwards. This being said, the literature hardly offers explicit demonstrations of this correspondence. An exception is the treatment by Johnsson and Stedman [8]. To spell out the angular momentum of the nuclear motion in the presence of a vector field, we must include the term $-q\mathbf{A}$ in the vector product $\mathbf{R}\wedge\mathbf{P}$

$$\mathbf{R}\wedge(\mathbf{P}-q\mathbf{A}) = \mathcal{L} - q\frac{1}{2} = \mathcal{J}_z\,, \tag{60}$$

where $q = \pm 1$ is the charge of the particle. The fact that we recover the angular momentum operator of the dynamic JT treatment indicates that this treatment indeed fully incorporates the Berry phase. The angular momentum thus truly reflects the dual boson-fermion characteristic of the JT Hamiltonian.

3 The triplet $T \times (e + t_2)$ Jahn-Teller system

Triple degeneracies occur in cubic and icosahedral symmetries. The symmetries of the JT modes are generated as

$$\begin{aligned} O_h &: [T \times T] - A_{1g} = e_g + t_{2g} \\ I_h &: [T \times T] - A_g = h_g \end{aligned} \tag{61}$$

T-terms have a frequent occurrence in coordination compounds and metal clusters, often with important implications for magnetism. The strength of the coupling is usually less pronounced than for E-terms, giving rise to all sorts of dynamic properties. A notable example of an icosahedral T-system is the ground state of the fulleride anion, C_{60}^{-}, due to the occupation of the t_{1u} LUMO of Buckminsterfullerene by a single electron.

3.1 The Hamiltonian

As before two spaces are to be considered: the fermion basis defines a three-dimensional sphere, with unit vectors $|T_x\rangle, |T_y\rangle, |T_z\rangle$, and the boson space, forming a five-dimensional Euclidean space, with unit vectors Q_θ, Q_ϵ for the e_g-modes and Q_ξ, Q_η, Q_ζ for the t_{2g}-modes.
In the linear coupling regime, the Hamiltonian is given by

$$\mathcal{H} = \frac{1}{2}K_E\left(Q_\theta^2 + Q_\epsilon^2\right) + \frac{1}{2}K_{T_2}\left(Q_\xi^2 + Q_\eta^2 + Q_\zeta^2\right) + \mathcal{H}' \tag{62}$$

with

$$\mathcal{H}' = F_E \begin{pmatrix} -\frac{1}{2}Q_\theta + \frac{\sqrt{3}}{2}Q_\epsilon & 0 & 0 \\ 0 & -\frac{1}{2}Q_\theta - \frac{\sqrt{3}}{2}Q_\epsilon & 0 \\ 0 & 0 & Q_\theta \end{pmatrix} + \frac{F_{T_2}}{\sqrt{2}} \begin{pmatrix} 0 & -Q_\zeta & -Q_\eta \\ -Q_\zeta & 0 & -Q_\xi \\ -Q_\eta & -Q_\xi & 0 \end{pmatrix}. \tag{63}$$

The potential energy surface is defined in 5D coordinate space. However a concise view of the topography of the surface can be achieved by projection in 3D fermion space. The procedure is as follows: consider an electronic eigenvector (x, y, z), normalized to unity

$$|T(\mathbf{r})\rangle = x|T_x\rangle + y|T_y\rangle + z|T_z\rangle. \tag{64}$$

Antipodal points (x, y, z) and $(-x, -y, -z)$ describe the same solution, hence the electronic space is restricted to a hemisphere. This surface has the topology of a projective plane. Now minimize the energy for every direction on this sphere

$$\frac{\delta}{\delta Q_{\Lambda\lambda}}\left(\mathbf{r}^\dagger \mathcal{H} \mathbf{r}\right) = \mathbf{r}^\dagger \frac{\delta \mathcal{H}}{\delta Q_{\Lambda\lambda}} \mathbf{r} = 0 \qquad \forall\ Q_{\Lambda\lambda} \in [\Gamma]^2 - \Gamma_0\,. \tag{65}$$

This yields a set of equations from which we may determine the stationary coordinates, denoted as $\|Q_{\Lambda\lambda}\|_r$. Reinserting these coordinates in the energy expressions yields the function $\|E\|_r$. This function does not represent eigenenergies, except in the stationary points where it is indeed

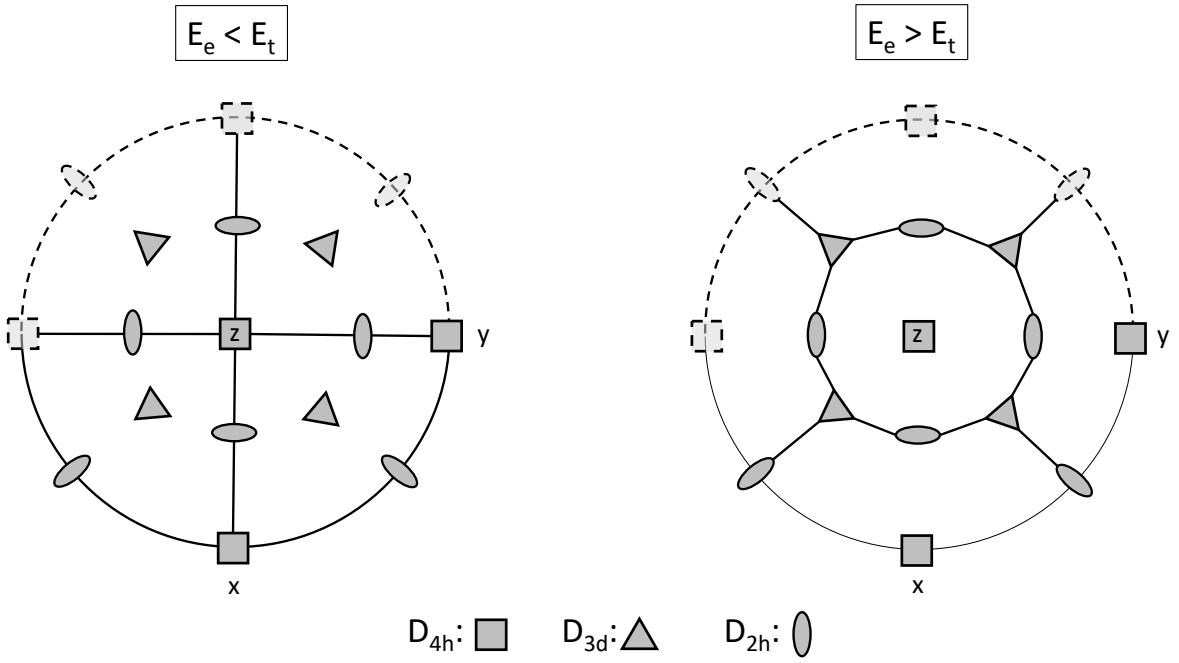


Fig. 7: *Jahn-Teller distortions for a cubic triplet, projected on a hemisphere; the z-axis is the upright axis.*

a root of the matrix equation! Hence this function is *isostationary*, i.e., it coincides with the hypersurface in the stationary points. As an example, for the cubic T-terms, the isostationary function is given by

$$\langle \|E\| \rangle_{\mathbf{r}} = \frac{1}{5}\big(2E_E^{JT}+3E_{T_2}^{JT}\big) + \frac{6}{5}\big(E_E^{JT}-E_{T_2}^{JT}\big) f_4 \tag{66}$$

with

$$E_E^{JT} = -\frac{1}{2}\frac{F_E^2}{K_E} \qquad E_{T_2}^{JT} = -\frac{1}{3}\frac{F_{T_2}^2}{K_{T_2}} \qquad f_4 = \frac{1}{2}\Big(x^4+y^4+z^4 - 3\left(x^2y^2+x^2z^2+y^2z^2\right)\Big). \tag{67}$$

The f_4 polynomial, which controls the topography, is recognized here as the cubic invariant of the fourth-order spherical harmonics, which also provides the crystal field potential in octahedral symmetry. The term preceding f_4 involves the difference of the JT stabilization energies. If the stabilization along e-modes is more pronounced, the surface is characterized by tetragonal minima, with orthorhombic saddle points in between. The trigonal extrema in this case are hill tops on the surface. In contrast if the t_2-modes prevail, the surface will be turned upside down, as shown in Fig. 7. Additional second-order terms in the Hamiltonian will produce a further warping of the surface. As an example, when both e and t_2 modes are active, and there is a strong second-order interaction term between them, the next cubic invariant of rank 6 will take over control, and produce a surface with six orthorhombic D_{2h} minima and twelve C_{2h} saddle points in between.

In the icosahedral case the linear JT Hamiltonian is isotropic and the hypersurface corresponds to a 3D sphere. However upon introduction of second-order warping terms minima and maxima

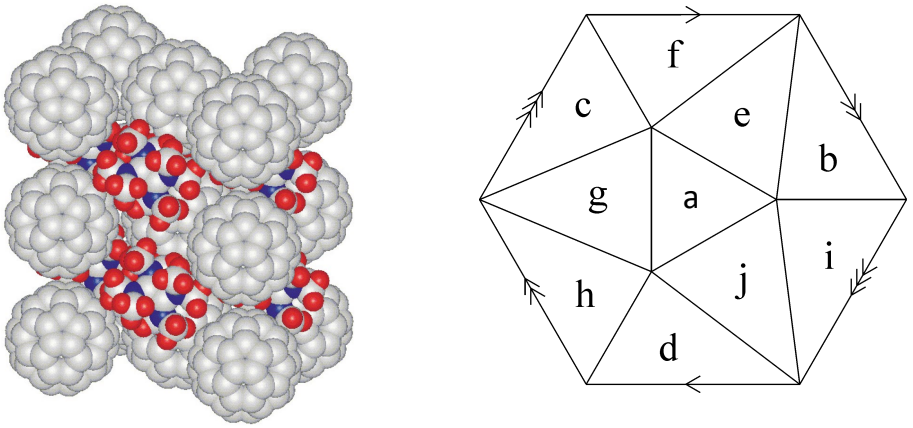


Fig. 8: *Jahn-Teller distortions for an isosahedral triplet, in a projective plane, consisting of six vertices and ten triangular faces: the vertices correspond to D_{5d} points, and the faces to D_{3d} points; left is shown the crystal structure of TDAE$^{+}C_{60}^{-}$ (taken from [9]).*

appear. These are governed by an icosahedral invariant of rank 6. Again two regimes are possible: either ten trigonal minima, and six pentagonal hill tops, or vice-versa. The pentagonal points are all equidistant and form the complete graph of six nodes. The nine neighbors of each trigonal point split into two orbits of orders three and six. A case in point is the anion of C_{60}. Fulleride ions may be formed by reduction with alkali metals, or electron donors such as tetrakis(dimethylamino)ethylene [9] TDAE, see Fig. 8.

3.2 Dynamics

For an understanding of the dynamics we turn to the high-symmetry case, where the Hamiltonian is limited to the linear force elements, and with – in case of cubic symmetry – degenerate coupling between e and t_2 modes: $E_e^{JT} = E_{t_2}^{JT}$. As the isostationary function demonstrates, the potential energy minimum in this case forms a continuum. This corresponds to a 3D spherical trough in the 5D coordinate space. In analogy to the circular motion of atoms in the JT-trough for a triangle shown before in Fig. 5, in the present case of a spherical trough the loci of displacements of individual atoms form a sphere, centered at their high-symmetry positions. Judd has provided a detailed analysis of this internal rotation in the case of a T-type JT surface in an octahedron [10]. The motions of equivalent atoms are locked and concerted so that the total degree of freedom corresponds to the symmetry group of a 3D sphere, which is the special orthogonal group in 3D: $SO(3)$.

The Hamiltonian for the highly symmetric limit, also known as the $P \times d$ Hamiltonian, describes the coupling between a dipolar fermion part and its symmetrized square which corresponds to a quadrupolar tensor. It consists of a harmonic part, $\mathcal{H}_0$, and the standard linear coupling term, $\mathcal{H}'$, which is the scalar product of the fermion tensor and the coordinate tensor

$$\mathcal{H}' = k \sum_q (-1)^q (\mathbf{f}^\dagger \mathbf{f})_q^2 Q_{-q} \,. \tag{68}$$

Here k is the coupling parameter. The coordinates are written in their complex format, defined

as

$$Q_0 = Q_\theta\ , \quad Q_{\pm1} = \mp\frac{1}{\sqrt{2}}Q_\eta - \frac{i}{\sqrt{2}}Q_\xi \quad \text{and} \quad Q_{\pm2} = \frac{1}{\sqrt{2}}Q_\epsilon \pm \frac{i}{\sqrt{2}}Q_\zeta\,. \tag{69}$$

In matrix form, acting in the space of the fermions, ordered as $|{+1}\rangle, |\ 0\ \rangle, |{-1}\rangle$, the coupling Hamiltonian reads

$$\mathcal{H}' = k\begin{pmatrix} \frac{1}{\sqrt{6}}Q_0 & \frac{1}{\sqrt{2}}Q_{-1} & Q_{-2} \\ -\frac{1}{\sqrt{2}}Q_{+1} & -\frac{2}{\sqrt{6}}Q_0 & -\frac{1}{\sqrt{2}}Q_{-1} \\ Q_{+2} & \frac{1}{\sqrt{2}}Q_{+1} & \frac{1}{\sqrt{6}}Q_0 \end{pmatrix}. \tag{70}$$

As this is a scalar product of spherical tensors, the coupling Hamiltonian will be $SO(3)$ invariant. The secular equation of $\mathcal{H}'$ reduces to

$$E^3 - \frac{E}{2}Q^2 + \frac{1}{3\sqrt{6}}I_3^3 = 0 \tag{71}$$

where

$$\begin{aligned} Q^2 =& Q_0^2 - 2Q_{+1}Q_{-1} + 2Q_{+2}Q_{-2} \\ I_3 =& Q_0\big(Q_0^2 - 6Q_{+2}Q_{-2} - 3Q_{+1}Q_{-1}\big) + \frac{3\sqrt{3}}{\sqrt{2}}\big(Q_{+2}Q_{-1}^2 + Q_{-2}Q_{+1}^2\big). \end{aligned} \tag{72}$$

The interesting aspect of this secular equation is that it contains two $SO(3)$ invariants: Q is the squared norm of the distortion space and thus measures the extent of the distortion, while I_3 is a third-order invariant, proportional to the determinant of the JT Hamiltonian. The roots of the eigenvalue equation can be expressed using the angle representation. Rewrite I_3 as:

$$I_3 = Q^3 \cos 3\gamma\,. \tag{73}$$

The equation can then easily be solved by the trigonometric expressions for the three roots

$$E_k = -kQ\frac{2}{\sqrt{6}}\cos\left(\gamma - \frac{2n\pi}{3}\right)\ n = 0, 1, 2. \tag{74}$$

What is the meaning of the angle γ which appears when solving the secular equation? The answer to this question takes us to the 5D oscillator formed by the quadrupolar Q_{m_l} JT modes. The parent symmetry of the 5D oscillator is the special unitary group $SU(5)$ which allows for all possible unitary transformations of the five quadrupolar modes. This group can conveniently be restricted to its subgroup of orthogonal transformations, $SO(5)$. However, when considering the JT Hamiltonian, symmetry is restricted to an even lower subgroup, corresponding to the sphere in 3D, with symmetry group $SO(3)$. A clear understanding of the embedding of $SO(3)$ in $SO(5)$ is offered by the surface oscillations of a vibrating sphere, which have been studied extensively in nuclear physics as a model for the vibrating nucleus. Low-energy nuclear vibrations indeed match the five quadrupolar modes (and likewise for the tidal waves on earth), dipoles being removed as they correspond to spurious translations. On the other hand the totally symmetric scalar mode, which corresponds to a breathing mode, is way higher in energy, since it stretches the surface everywhere. Besides, in the JT context it is inactive as it doesn't break

the symmetry. The model of the vibrating hollow sphere thus provides a complete description of the quadrupolar modes. Since the quadrupolar modes have the smallest allowed L value, they can only introduce a minimal symmetry breaking: they distort a sphere into an ellipsoid. An ellipsoid is a surface characterized by three orthogonal axes of different length. The sum of these lengths must be constant in time, in order to avoid any admixture of the radial breathing mode. Hence proper ellipsoidal distortions have only two degrees of freedom. These correspond to the tetragonal Q_θ mode and the orthorhombic Q_ϵ mode. The tetragonal mode leads to a prolate ($Q_\theta > 0$) or oblate ($Q_\theta < 0$) ellipse, which still has cylindrical symmetry along the z-axis. The radius of this ellipse is thus described as

$$d(\theta, \phi) = R\big(1 + c(3\cos^2\theta - 1)\big), \tag{75}$$

where R is the radius of the sphere, and c is scaling constant which oscillates in time with the vibration. The orthorhombic mode will further break this axial symmetry, by repartitioning the distortion between the x- and y- directions. A general ellipsoidal distortion with principal axes along the Cartesian directions is thus described by a vector in the space formed by these two coordinates. This is a bimodal distortion [11]. Turning to polar coordinates, the parametric description of this distortion reads

$$\begin{pmatrix} Q_\theta \\ Q_\epsilon \\ Q_\xi \\ Q_\eta \\ Q_\zeta \end{pmatrix} = Q \begin{pmatrix} \cos\gamma \\ \sin\gamma \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{76}$$

The angle which appears here refers to the balance between tetragonal and orthorhombic modes, and this is precisely the angle γ which appeared in the secular equation. The ellipsoid which is obtained by this bimodal distortion is still aligned with the Cartesian reference frame. Spherical symmetry of course requires that the ellipsoid is free to rotate in 3D space. This is where the three remaining quadrupolar modes come in. The general orientation of the ellipsoidal distortion can be performed by the Euler rotation matrix in the full space of the five L=2 modes. In summary a symmetry adaptation has been performed of the five degrees of freedom: three angles describe the orientation in 3D space and present the spherical $SO(3)$ symmetry of the vibrating sphere. A radius and an extra angle γ define the ellipsoidal distortion. As this extra angular degree of freedom appears in the secular equation, the total Hamiltonian has only $SO(3)$ symmetry, and does not form a spherical surface in 5D. The T-term JT problem is thus at its core a symmetry breaking chain $SU(5) \downarrow SO(5) \downarrow SO(3)$. The relevant irreducible representations (irreps) for each of these groups are as follows:

- Excitations of n oscillator quanta in $SU(5)$ correspond to irrep $[n]$.
- Relevant irreps in $SO(5)$ are denoted as $(\nu, 0)$. Here ν is a whole number, which is known as the Racah seniority number.
- Irreps in $SO(3)$ are characterized by the angular momentum quantum number L.

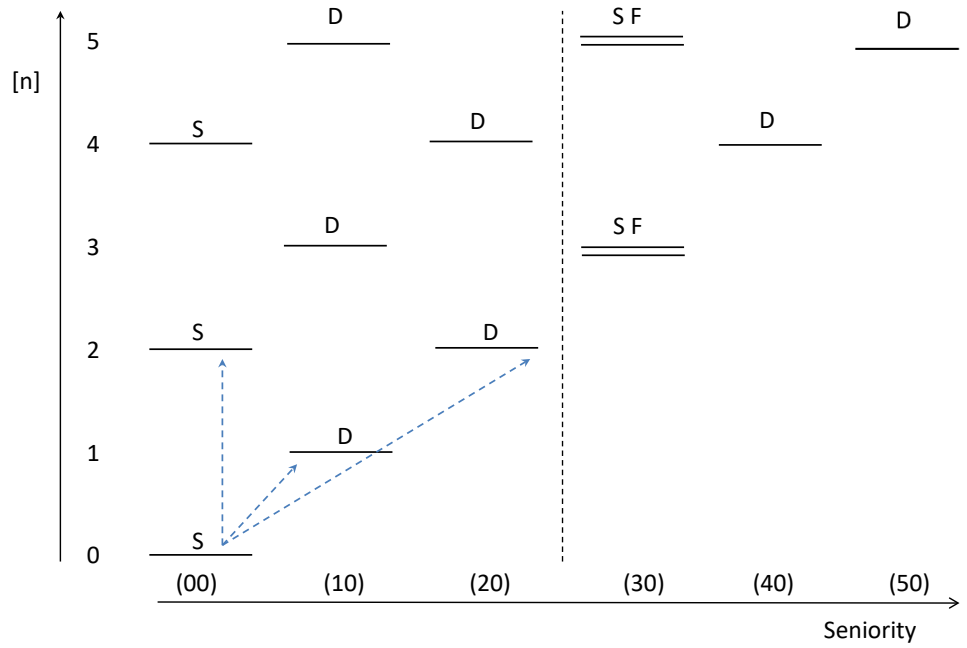


Fig. 9: *3D assignments of the surface vibrations of a sphere; the vertical axis represents the oscillator states of a 5D oscillator; the horizontal axis decomposes these states in $SO(5)$ seniorities; spherical components are indicated by S,D,F,... angular momentum; note the repetition of the SDD pattern, with a period of three.*

The branching scheme along the symmetry lowering chain is shown in Fig. 9. It represents the energy of the boson spectrum at zero JT coupling. As the coupling sets in, the L levels are coupled with the P fermion level, according to the usual vector addition rules. As an example, in order to realize a vibronic state with P symmetry, only S and D levels are involved: $S{\times}P = P$ and $D{\times}P = P{+}D{+}F$. Interestingly the diagram shows that the initial SDD pattern shows a perfect repetition with a period of three. This observation allows to construct an Ansatz, with spherical symmetry and the two extra-spherical degrees of freedom which appear in the secular equation: the radius Q which measures the vertical radial excitations in the diagram, and the angle γ which takes us through the horizontal seniority period in the diagram. We conclude by providing the Ansatz

$$\Psi_0 = \begin{pmatrix} \left(\quad \sqrt{3}b^\dagger_{-1}\mathcal{F}_1 \quad +(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{-1}\mathcal{F}_2\right) f^\dagger_{+1} \\ \left(\mathcal{F}_0 - 2b^\dagger_0\,\mathcal{F}_1 \quad -2(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_0\,\mathcal{F}_2\right) f^\dagger_0 \\ \left(\quad \sqrt{3}b^\dagger_{+1}\mathcal{F}_1 + \sqrt{6}(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{+1}\mathcal{F}_2\right) f^\dagger_{-1} \end{pmatrix} \qquad \text{and}$$

$$\Psi_{+1} = \begin{pmatrix} \left(\mathcal{F}_0 \quad +b^\dagger_0\,\mathcal{F}_1 \quad +(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_0\,\mathcal{F}_2\right) f^\dagger_{+1} \\ \left(\quad -\sqrt{3}b^\dagger_{+1}\mathcal{F}_1 - \sqrt{3}(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{+1}\mathcal{F}_2\right) f^\dagger_0 \\ \left(\quad \sqrt{6}b^\dagger_{+2}\mathcal{F}_1 + \sqrt{6}(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{+2}\mathcal{F}_2\right) f^\dagger_{-1} \end{pmatrix}, \; \Psi_{-1} = \begin{pmatrix} \left(\quad \sqrt{6}b^\dagger_{-2}\mathcal{F}_1 + \sqrt{6}(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{-2}\mathcal{F}_2\right) f^\dagger_{+1} \\ \left(\quad -\sqrt{3}b^\dagger_{-1}\mathcal{F}_1 - \sqrt{3}(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_{-1}\mathcal{F}_2\right) f^\dagger_0 \\ \left(\mathcal{F}_0 \quad +b^\dagger_0\,\mathcal{F}_1 \quad +(\mathbf{b}^\dagger\mathbf{b}^\dagger)^2_0\,\mathcal{F}_2\right) f^\dagger_{-1} \end{pmatrix}.$$

Here the $\mathcal{F}$ functions depend only on the $SO(5)$ constants Q and γ. $\mathcal{F}_0$ provides the coupling with the S states with seniority $(3\nu, 0)$, $\mathcal{F}_1$ takes care of coupling with the D states with seniority $(3\nu{+}1, 0)$, and $\mathcal{F}_2$ runs over the D states with seniority $(3\nu{+}2, 0)$.

References

[1] A. Ceulemans: *The theory of the Jahn-Teller effect* (Springer Nature Switzerland, 2022)

[2] A.F. Wells: *Structural inorganic chemistry* (Clarendon Press, Oxford, 1945)

[3] I.B. Bersuker: *The Jahn-Teller effect* (Cambridge University Press, 2006)

[4] J.T. Muya, J.D. Kelling, A. Ceulemans, C. Parish, J. Chem. Phys. **154**, 164305 (2021)

[5] F. Cocchini, T.H. Upton, J. Chem. Phys. **88**, 6068 (1988)

[6] T. Weike, D.M.G. Williams, A. Viel, W. Eisfeld, J. Chem. Phys. **151**, 074302 (2019)

[7] M.V. Berry: *The quantum phase, five years after* in: A. Shapere and F. Wilczek (eds.): *Geometric Phases in Physics*, Advanced Series in Mathematical Physics, Vol. 5, p. 7 (World Scientific, Singapore, 1989)

[8] M.T. Johnsson, G.E. Stedman, J. Phys.: Condens. Matter **11**, 787 (1999)

[9] A. Schilder, W. Bietsch, M. Schwoerer, New J. Phys. **1**, 5 (1999)

[10] B.R. Judd: *The theory of the Jahn-Teller effect* in C.D. Flint (ed.): *Vibronic processes in inorganic chemistry*, NATO ASI Series C: Mathematical and Physical Sciences, vol. 288, pp. 79-101 (Kluwer, Dordrecht, 1989)

[11] A. Auerbach, N. Manini, E. Tossatti, Phys. Rev. B **49**, 12998 (1994)

3 Orbitally Induced Peierls Mechanism for Charge-Orbital Orderings in Transition-Metal Compounds

Takashi Mizokawa
Department of Applied Physics, Waseda University
3-4-1 Okubo, Shinjuku, 169-8555 Tokyo, Japan

Contents

1 Introduction **2**
1.1 Orbital degrees of freedom in transition-metal compounds 2
1.2 Metal-insulator transitions in transition-metal compounds 3

2 Orbitally induced Peierls mechanism **4**
2.1 Case study on $CuIr_2S_4$ 4
2.2 Square lattice models for orbitally induced Peierls transition 7
2.3 Triangular lattice models for orbitally induced Peierls transition 10
2.4 Effect of strong coupling 12

3 Application of the orbitally induced Peierls mechanism to transition-metal compounds **12**
3.1 Spinel systems 12
3.2 Pyrochlore systems 15
3.3 Triangular lattice systems 16
3.4 Honeycomb lattice systems 19
3.5 Kagome lattice systems 21

4 Conclusion **22**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

1.1 Orbital degrees of freedom in transition-metal compounds

Transition-metal compounds exhibit rich physical properties which have been attracting overwhelming research activities in the field of condensed matter physics [1–4]. Their electrical and magnetic properties are governed by the transition-metal d electrons. The d orbitals have five-fold degeneracy in the atomic limit and are split into three-fold degenerate t_{2g} (xy, yz, and zx) and two-fold degenerate e_g ($3z^2-r^2$ and x^2-y^2) orbitals under the cubic ligand field. When the d electrons become itinerant in transition-metal compounds with octahedral or tetrahedral coordination, they are accommodated by the t_{2g} and/or e_g bands creating multiple Fermi surfaces. In quasi one-dimensional materials such as TaS_3, a charge-density wave or modulation of charge density is induced by a Peierls transition in which the Fermi-surface nesting and the electron-lattice interaction play essential roles [5]. Also spin-density wave or modulation of spin density and direction in metallic transition-metal compounds such as LaFeAsO can be explained by nesting of the Fermi surfaces and the weak or moderate d-d Coulomb interaction [6]. On the other hand, the d electrons can be localized due to the strong d-d Coulomb interaction (Mott insulators and Wigner crystal states). In the localized case, the valence electrons are accommodated by the atomic like t_{2g} and e_g orbitals. The partially occupied t_{2g} or e_g orbitals exhibit orbital orderings due to Jahn-Teller effect [7], Kugel-Khomskii mechanism [8], and spin-orbit interaction [9]. For example, one of the e_g orbitals is occupied in multiferroic $TbMnO_3$ providing the $3x^2-r^2/3y^2-r^2$ orbital ordering which is stabilized by the strong Jahn-Teller effect, breaking the e_g orbital degeneracy through elongation of the MnO_6 octahedra [10]. As for the t_{2g} orbitals, the Jahn-Teller effect is relatively weak and the Kugel-Khomskii mechanism plays more important roles in determining the spin and orbital states via the superexchange interaction [4]. When one or two of the t_{2g} orbitals are occupied/unoccupied in $3d$ and $4d$ oxides, the Jahn-Teller effect, the Kugel-Khomskii mechanism, and the spin-orbit interaction may compete to determine the orbital ordering [11]. In $YTiO_3$, $LaVO_3$, and YVO_3 with one or two of the t_{2g} orbitals occupied, the orbital ordering is governed by the Kugel-Khomskii mechanism with possible interplay with the Jahn-Teller distortion and the tilting of the MO_6 octahedron (M=transition metal). In $LaTiO_3$ and Ca_2RuO_4, the spin-orbit interaction can play an important role with the Kugel-Khomskii effect and the Jahn-Teller distortion [12]. When one or two of the t_{2g} orbitals are occupied/unoccupied in insulating $5d$ oxides such as Sr_2IrO_4, the orbital degeneracy is lifted by the strong spin-orbit interaction [13].

Transition-metal compounds with MO_6 octahedra harbor various crystal structures such as perovskite, rocksalt, rutile, corundum, and spinel structures. In the perovskite (or layered perovskite) structure, the MO_6 octahedra share their corners and the transition-metal sites form a simple cubic lattice (or a square lattice). The transfer terms between t_{2g} orbitals at the neighboring sites are effectively given by the M-ligand-M hybridization governed by the transfer integral ($pd\pi$) and the ligand-to-d charge-transfer energy Δ [4]. In the spinel systems (or triangular lattice systems such as CdI_2-type and $NaFeO_2$-type structure), the MO_6 octahedra share

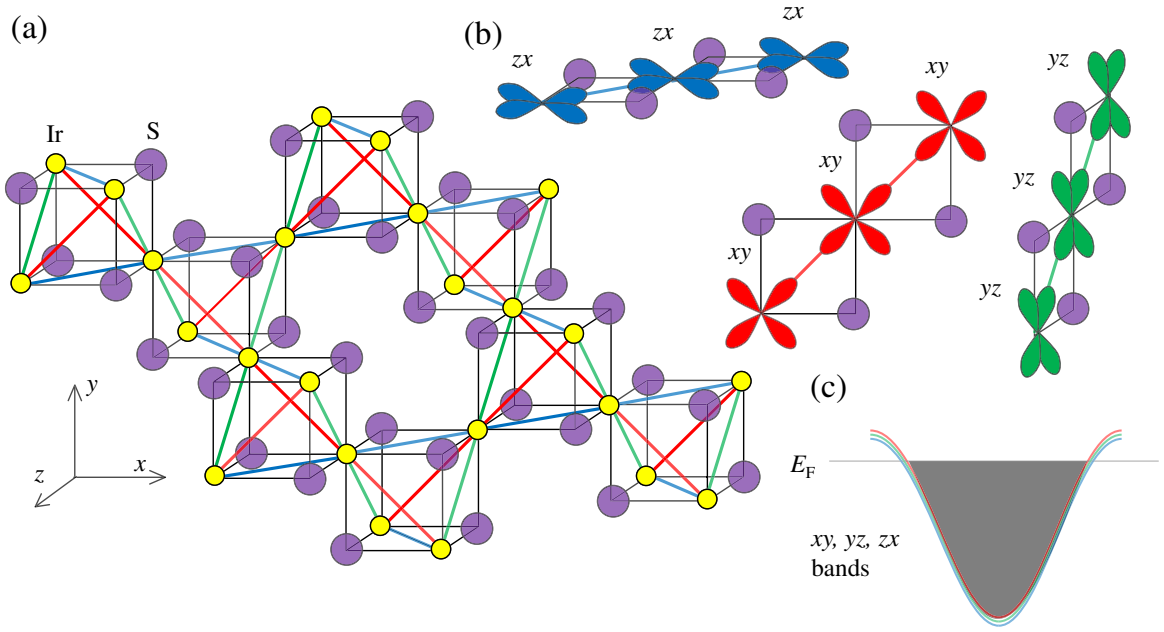


Fig. 1: *(a) Spinel structure of $CuIr_2S_4$ and its pyrochlore lattice of Ir. (b) Ir $5d$ yz, zx, and xy orbitals form one-dimensional bands along $(0,1,1)$ or $(0,1,-1)$ direction, $(1,0,1)$ or $(1,0,-1)$ direction, and $(1,1,0)$ or $(1,-1,0)$ direction, respectively. (c) 5.5 electrons (0.5 holes) per Ir are distributed to the three one-dimensional bands of yz, zx, and xy orbitals.*

their edges and the transition-metal sites form a pyrochlore lattice (or a triangular lattice) as shown in Fig. 1. The transfer integrals between t_{2g} orbitals at the neighboring sites are given by the direct M-M transfer integral ($dd\sigma$) or the indirect M-ligand-M transfer governed by ($pd\pi$) and Δ.

1.2 Metal-insulator transitions in transition-metal compounds

Both in the corner-sharing and edge-sharing systems, once the d electrons are localized (Mott insulators and Wigner crystal states), their magnetic properties are primarily described by models made up from localized spins and orbitals. The inter-site charge excitation across the band gap provides the superexchange interaction between the localized spins and orbitals. In this situation, the interplay between spin ordering and orbital ordering is described by the Kugel-Khomskii model [8]. Several systems including V_2O_3, Ca_2RuO_4, $YNiO_3$, and NiS_2 exhibit metal-insulator transitions with cooling which can be viewed as Mott transitions [1–3]. The d electrons are localized below the transition temperature. Interestingly, the transitions of V_2O_3, Ca_2RuO_4, and $YNiO_3$ are accompanied by strong lattice distortions associated with orbital ordering (V_2O_3 and Ca_2RuO_4) or charge disproportionation ($YNiO_3$). In general, local-density approximation (LDA) band structure calculations fail to explain the magnitude of band gap (Mott gap).

When the d electrons are itinerant, the edge-sharing systems often exhibit more complicated behaviors than the corner-sharing ones due to the direct M-M transfer terms. Most of the metallic perovskites such as $SrVO_3$, $SrCoO_3$, $SrRuO_3$, Sr_2RuO_4, ReO_3, and $SrIrO_3$ are paramagnetic or

ferromagnetic metals without any metal-insulator transitions. On the other hand, several spinel or triangular lattice systems such as $MgTi_2O_4$, $LiRh_2O_4$, $CuIr_2S_4$, and $LiVS_2$ are metallic at room temperature and undergo exotic metal-insulator transitions with cooling. Their metal-insulator transitions are accompanied by strong lattice distortion which will be discussed in the following chapters. The insulating phases are nonmagnetic and can be described by LDA band structure calculations if the lattice distortion are properly taken into account (There are some exceptions: It is still difficult to describe the nonmagnetic insulating states of $MgTi_2O_4$ and $LiVO_2$ by LDA). The lattice distortion is often characterized by the metal-metal dimerization or trimerization which is driven by the direct M-M transfer. It is expected that the geometry of Fermi surfaces plays an important role in the metal-insulator transitions of the relatively itinerant systems.

2 Orbitally induced Peierls mechanism

2.1 Case study on $CuIr_2S_4$

In the itinerant case, the moderate Coulomb interaction between the d electrons (d-d Coulomb interaction) remains and, therefore, a possible instability towards spin and/or charge ordering may occur due to Fermi-surface nesting. In addition, even without the d-d Coulomb interaction, the electron-lattice interaction can provide a lattice modulation and consequent charge modulation with a periodicity corresponding to the Fermi-surface nesting. Such charge and/or spin ordering (charge-density wave and/or spin-density wave) as well as the lattice modulation can open an energy gap at the Fermi level. Especially when the Fermi surfaces are purely one-dimensional (planes in the k-space), Fermi-surface nesting is always realized and the system inevitably undergoes a Peierls transition in which an energy gap is formed at the Fermi level due to the charge and/or spin and/or lattice modulation. Even though the band structure near the Fermi level is changed by the modulation to open the energy gap, the d orbital occupation is not drastically changed by the charge/spin/lattice modulation. Usually, there is no interplay between spin/charge/lattice modulation and the d-orbital degrees of freedom in the itinerant transition-metal compounds. Yet there are several transition-metal compounds such as $CuIr_2S_4$ and $LiRh_2O_4$ which exhibit a charge/lattice modulation and band Jahn-Teller like distortion, suggesting an interaction between the charge ordering and the orbital ordering. Since the t_{2g} bands are degenerate near the Fermi level in the cubic systems, the tetragonal distortion can be stabilized by removing the band degeneracy due to band Jahn-Teller effect. Motivated by these peculiar phase transitions, the idea of an orbitally induced Peierls mechanism has been introduced [14, 15]. Historically, it has been applied to explain the complicated charge and orbital ordering in spinel-type t_{2g} transition-metal compounds in which the transition-metal sites form a pyrochlore lattice. Among them, $CuIr_2S_4$ with a pyrochlore lattice structure shows a metal-insulator transition around 230 K [16] with tetragonal distortion and octamer charge ordering [17]. In this subsection, we consider the orbitally induced Peierls description of $CuIr_2S_4$. Under octahedral coordination, the lobes of the e_g (t_{2g}) orbitals are directed towards (between)

the ligands. Therefore, in an edge-sharing octahedron, the t_{2g} orbitals have substantial transfer integrals between the neighboring transition-metal sites. Let us neglect the indirect transfer between the t_{2g} and e_g orbitals via the ligand orbitals and consider the direct ones between the t_{2g} orbitals on the pyrochlore lattice. Here, the transfer integrals between the t_{2g} orbitals are restricted by the cubic symmetry. As shown in Fig. 1(a), the Ir sites form a pyrochlore lattice in which Ir_4 tetrahedra share their corners. Along the $(1, \pm 1, 0)$ direction of the pyrochlore lattice, only the transfer integrals between the xy orbitals survive by symmetry. Also along the $(1, 0, \pm 1)$ and $(0, 1, \pm 1)$ directions, only the zx and yz orbitals have non-zero transfer integrals, respectively. Consequently, the xy, yz, and zx orbitals form one-dimensional bands along the $(1, \pm 1, 0)$, $(1, 0, \pm 1)$, or $(0, 1, \pm 1)$ directions, respectively. In $CuIr_2S_4$ there are 5.5 electrons (0.5 holes) in the t_{2g} bands. If the 0.5 holes are distributed in the three one-dimensional bands, each band accommodates $1/6$ holes. The six pairs of one-dimensional Fermi surfaces (six pairs of parallel planes in the three-dimensional k-space) have a Peierls instability due to electron-lattice or electron-electron interaction. Since the paired parallel planes are spanned by the wave vectors along $(1, \pm 1, 0)$, $(1, 0, \pm 1)$, $(0, 1, \pm 1)$ with a magnitude of $\pi/6$ (the unit of the wave vectors is $1/a$ where a is the Ir-Ir distance.), the pyrochlore lattice undergoes a distortion with 12 times periodicity along the $(1, \pm 1, 0)$, $(1, 0, \pm 1)$, and $(0, 1, \pm 1)$ directions. Such a Peierls instability of the pyrochlore lattice has two disadvantages from a theoretical point of view. First, the 12-times lattice modulation along the six directions is very complicated and the elastic-energy loss is expected to be large. Second, the indirect Ir-Ir transfers via the ligands, which are neglected in the above argument, are not negligibly small in $CuIr_2S_4$. There are finite transfer integrals between the different t_{2g} orbitals. Since the one-dimensional xy, yz, zx bands are degenerate in energy, the one-dimensional Fermi surfaces can be strongly deformed by the xy-yz, yz-zx, and zx-xy indirect transfer terms. Indeed, the calculated Fermi surfaces for the cubic $CuIr_2S_4$ have no Fermi-surface nesting [18–20]. In addition, the predicted charge ordering is inconsistent with the octamer charge-ordering observed in $CuIr_2S_4$. Here, it should be noted that the calculated band structure (and the Fermi surfaces) for the high temperature metallic phase are roughly consistent with the yz, zx, and xy bands mixed with the xy-yz, yz-zx, and zx-xy indirect transfer terms [18]. For the low temperature insulating phase, the band gap opening by the lattice distortion is partially explained by band structure calculations [19, 20]. In addition, the unique electronic and lattice properties of $CuIr_2S_4$ harbor the metastable disordered state induced by light or x-ray illumination [21–24]. In the metastable state, the crystal symmetry is at least partially recovered indicating that the long-range charge-order is destructed [21]. However, the electrical conductivity is much smaller than that of the high temperature metallic phase, and the energy gap at the Fermi level remains [22]. In the weak coupling limit, the energy gap should be closed once the long-range order of the charge and lattice modulation is destroyed. The observation of the metastable disordered state with bad conductivity suggests that the Ir-Ir dimers can survive without long-range order and that the conductivity is derived from a kind of bi-polaron hopping in the valence-bond liquid state.

The multi-orbital Fermi surfaces of $CuIr_2S_4$ can be reorganized in a band Jahn-Teller manner to enhance their nesting character for spin- and/or charge-density wave formation. When the

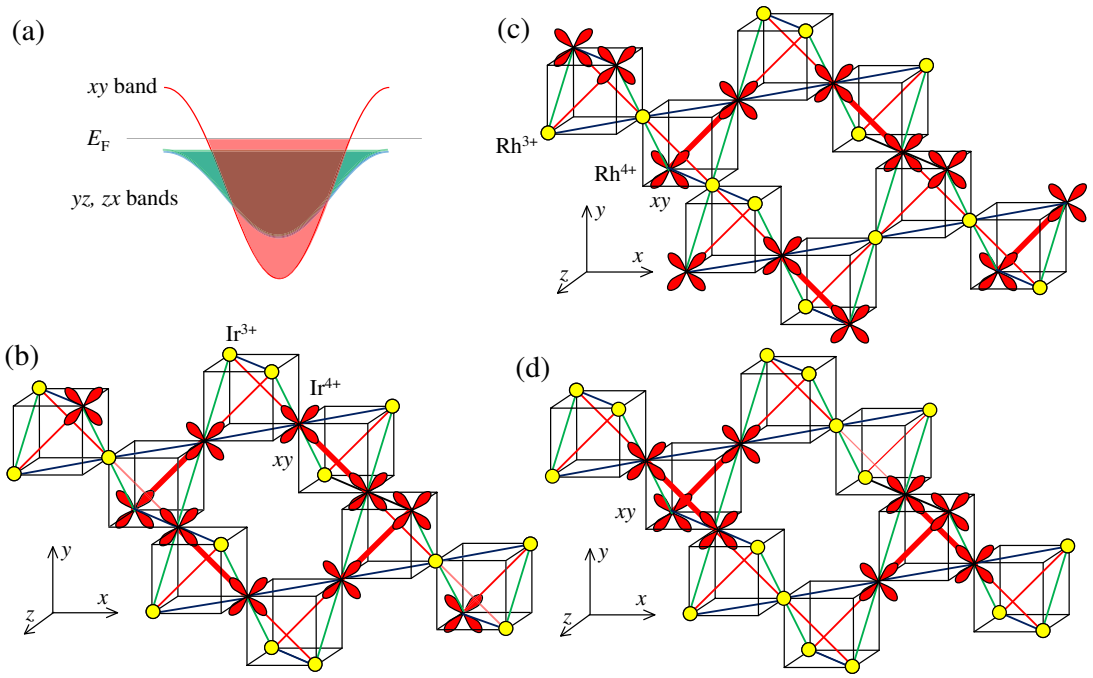


Fig. 2: *(a) Sketch of the yz, zx, and xy bands under a band Jahn-Teller distortion (elongated along the z-axis) for $CuIr_2S_4$. (b) Charge and orbital ordering with the octamer on the Ir pyrochlore lattice in $CuIr_2S_4$. The Ir $5d$ xy orbitals of the Ir^{4+} site form the strong Ir-Ir bonds along the $(1,1,0)$ and $(1,-1,0)$ directions. (c) Charge and orbital ordering on the Rh pyrochlore lattice in $LiRh_2O_4$. (d) Ir^{4+} or Rh^{4+} tetramer model for $CuIr_2S_4$ or $LiRh_2O_4$.*

cubic lattice is elongated along the z direction (c axis) while keeping its volume, the transfer integral between the xy orbitals becomes larger than that between the yz orbitals (and that between the zx orbitals) as illustrated in Fig. 2(a). Consequently, the width of xy band is much larger than that of the yz and zx bands. When the 0.5 holes are accommodated by the xy band, one-dimensional Fermi surfaces are formed along the $(1,\pm1,0)$ directions which are spanned by $Q=\pi/2$. Then the pyrochlore lattice shows modulation with four-times periodicity along $(1,\pm1,0)$ which can stabilize the Ir^{3+}-Ir^{3+}-Ir^{4+}-Ir^{4+} charge ordering. The Ir^{4+}-Ir^{4+} bond length is shortened due to the molecular orbital formation of the xy orbitals. The four Ir^{4+}-Ir^{4+} dimers created along the $(1,\pm1,0)$ directions form the Ir^{4+} octamer as shown in Fig. 2(b). The predicted charge ordering is consistent with the x-ray diffraction result reported by Radaelli *et al.* [17] Also the charge disproportionation between Ir^{3+} and Ir^{4+} is confirmed by core-level x-ray photoemission spectroscopy [22]. In this sense, the charge contrast along the $(1,\pm1,0)$ chains is much stronger than that of the weak coupling charge-density wave. The observed Ir $5d$ band width is comparable to that predicted by LDA [22], indicating weak electronic correlation. Therefore, the localized Ir^{4+}-Ir^{4+} dimers are stabilized by the strong electron-lattice interaction rather than the electronic correlation. The fluctuations of the Ir-Ir dimers are observed even above T_c in $CuIr_2S_4$ by means of pair distribution function measurements [25].

2.2 Square lattice models for orbitally induced Peierls transition

In the present subsection, we extend the idea of the orbitally induced Peierls mechanism and apply it to simplified two- and three-band models of transition-metal compounds with a square lattice. In order to introduce the basic idea of the orbitally induced Peierls mechanism, we start from a yz/zx two-band model on a square lattice [see Fig. 3(a)]. The transfer term along the x (y) direction for the neighboring zx (yz) orbitals is given by t_{zx} (t_{yz}). Note that the transfer integrals between the neighboring zx (or yz) orbitals are positive. The zx and yz orbitals form one-dimensional bands with energy dispersions of $E = 2t_{zx}\cos(k_x)$ and $E = 2t_{yz}\cos(k_y)$, respectively. Here, the unit for k_x and k_y is $1/a$ where a is the lattice constant of the square lattice. When the two one-dimensional bands accommodate some holes, one-dimensional Fermi surfaces are created as displayed in the middle panel of Fig. 3(b). In a realistic system, however, the two bands are mixed by the transfer term between the yz and zx orbitals ($t_{yz\text{-}zx}$) and consequently the Fermi surfaces become two-dimensional as shown in the right panel of Fig. 3(b). Such two-dimensional Fermi surfaces have a hidden instability by the combination of band Jahn-Teller and Peierls effect. Let us assume that the square lattice is compressed along the x-axis and elongated along the y-axis. Then the magnitude of t_{zx} becomes larger than that of t_{yz} under the distortion. If the ratio of t_{zx}/t_{yz} is large enough, only the wider zx band can accommodate the holes (band Jahn-Teller effect). Without $t_{yz\text{-}zx}$, the zx band forms a one-dimensional band with nesting vector Q [see the middle panel of Fig. 3(c)]. The orbitally induced Peierls state is robust against the indirect transfer term $t_{yz\text{-}zx}$ due to the energy splitting between the yz and zx orbitals. Under the effect of $t_{yz\text{-}zx}$, a quasi one-dimensional Fermi surface dominated by zx character is formed as shown in the right panel of Fig. 3(c) and gives nesting vector Q' rather than Q. In addition to the charge and/or spin modulation along the x-axis, the unit cell is doubled along the y-axis. In this scenario, by introducing the band Jahn-Teller distortion and ferro-type orbital ordering, charge- and/or spin-density waves are realized due to the Fermi-surface nesting. When the band Jahn-Teller distortion alone is enough to lower the energy, the orbital is restricted by it and then the Peierls instability follows (orbital restrictive case). On the other hand, the energy gain by the band Jahn-Teller distortion is not a necessary condition for the orbitally induced Peierls transition. Depending on the band width ratio t_{zx}/t_{yz} and the elastic-energy loss, the band Jahn-Teller distortion alone may not be enough to lower the energy since the system remains metallic. A Fermi-surface change by virtual Jahn-Teller distortion can be followed by a Peierls transition with wave vector Q or Q' to stabilize an insulating state with charge and lattice modulation. If the energy gain by the Peierls gap opening is large enough to compensate the elastic-energy loss by the band Jahn-Teller and Peierls distortion, one of the yz and zx orbitals is selected and the square lattice is distorted in a band Jahn-Teller manner (orbital selective case).

The orbitally induced Peierls state is robust against the interlayer transfer term t_z. With the t_z term illustrated in the right panel of Fig. 3(a), the zx and yz band dispersions are $E = 2t_{zx}\cos(k_x) + 2t_z\cos(k_z)$ and $E = 2t_{yz}\cos(k_y) + 2t_z\cos(k_z)$ providing the warped Fermi surfaces as shown in Fig. 4(a). Here, the unit for k_z is $1/c$ where c is the lattice constant along

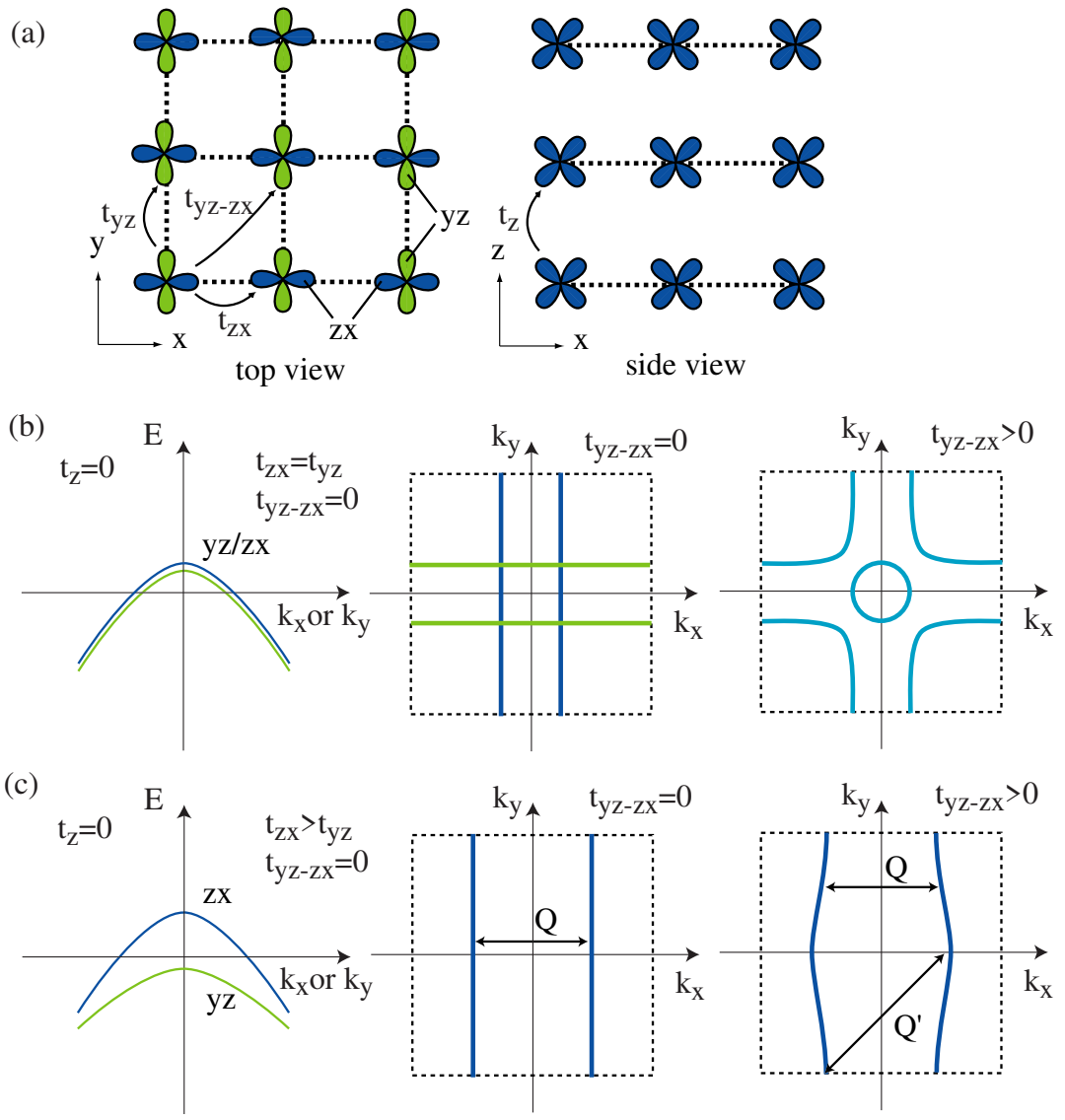


Fig. 3: *(a) Two-band model with yz/zx orbitals on a square lattice. t_{zx} (t_{yz}) is a transfer integral along the x (y) between zx (yz) orbitals. $t_{yz\text{-}zx}$ is a transfer integral along the diagonal direction between zx and yz orbitals. (b) Left and middle: Band structure and Fermi surface without yz-zx mixing by t_{zx} and band Jahn-Teller effect. Right: Fermi surface with yz-zx mixing by $t_{yz\text{-}zx}$ and without band Jahn-Teller effect. (c) Left and middle: Band structure and Fermi surface without yz-zx mixing by $t_{yz\text{-}zx}$ and with band Jahn-Teller effect. Right: Fermi surface with yz-zx mixing by $t_{yz\text{-}zx}$ and band Jahn-Teller effect.*

the z direction. With the band Jahn-Teller effect, the yz orbital is more stabilized and has smaller band width to be fully occupied. The holes are taken by the zx orbitals and their Fermi surfaces have the nesting vector Q'' as illustrated in Fig. 4(b). The yz band may reach the Fermi level around $k_z = \pi$ creating a small hole pocket as shown in the right panel of in Fig. 4(b). Under the Peierls distortion along the x and z directions due to the zx Fermi surface, the zx Fermi surface disappears while the yz Fermi pocket can remain. Such a metallic state with charge-orbital modulation would be relevant for several layered materials including $IrTe_2$.

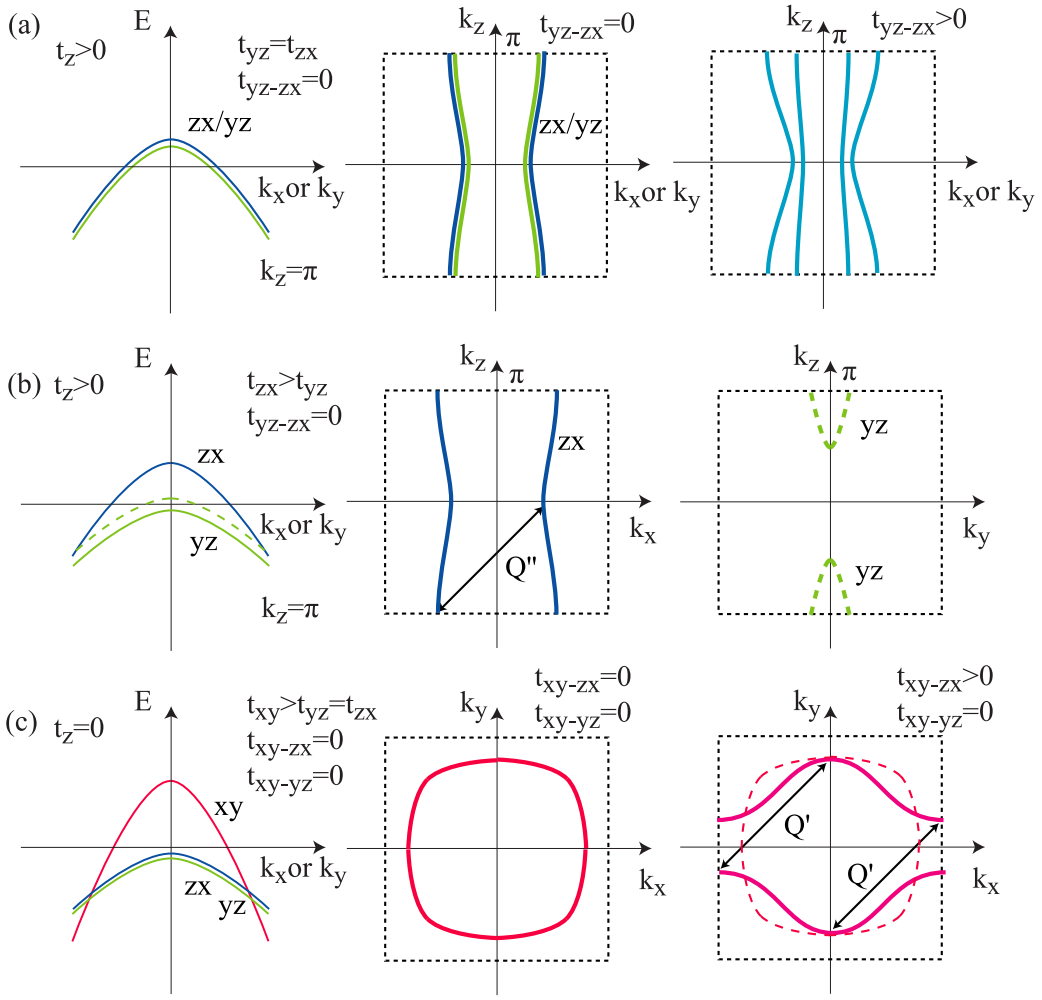


Fig. 4: *(a) Fermi surfaces of a yz/zx two-band model with interlayer transfer integral t_z and without band Jahn-Teller effect. (b) Fermi surfaces of a yz/zx two-band model with interlayer transfer integral t_z and band Jahn-Teller effect. (c) Left and middle: Band dispersion and Fermi surface of a $yz/zx/xy$ three-band model with yz and zx fully occupied. Right: Fermi surface of a $yz/zx/xy$ three-band model with anisotropic xy-zx mixing.*

In the next step, let us extend the idea of orbitally induced Peierls mechanism to a $yz/zx/xy$ three-band model with a xy Fermi surface. Let us assume that the yz and zx bands are fully occupied by electrons and the remaining xy band with $E = 2t_{xy}\cos(k_x) + 2t_{xy}\cos(k_y) + 2t_{xy\text{-}xy}\cos(k_x{+}k_y)$ forms a closed Fermi surface. Note that t_{xy} and $t_{xy\text{-}xy}$ are positive and negative, respectively. With such a circular Fermi surface without orbital degeneracy, an orbital instability or band Jahn-Teller effect is not expected. However, if a large energy gain by Peierls gap-opening is expected after a geometrical change of the Fermi surface (Lifshitz transition) by a sort of orbital anisotropy, the system may find its way to lower the symmetry for better Fermi-surface nesting. In the present model, a slight rhombic distortion lifts the degeneracy between the yz/zx bands and the xy-yz and xy-zx couplings become nonequivalent. Consequently, the Fermi surface of the xy band undergoes a Lifshitz transition and obtains good nesting character as schematically shown in Fig. 4(c).

2.3 Triangular lattice models for orbitally induced Peierls transition

Let us consider a yz/zx/xy three-band model on a triangular lattice. The transfer terms along the $(1,0,0)$, $(1/2,\sqrt{3}/2,0)$, and $(-1/2,\sqrt{3}/2,0)$ directions (under the X-, Y-, and Z-coordinates with $X=1/\sqrt{2}(x+y)$, $Y=1/\sqrt{6}(-x+y+2z)$, $Z=1/\sqrt{3}(x-y+z)$) are given by t_{xy}, t_{yz}, and t_{zx} for the neighboring xy, yz, zx orbitals. The xy, zx and yz orbitals form one-dimensional bands with energy dispersions with $E=2t_{xy}\cos(k_x)$, $E=2t_{zx}\cos(k_x/2+\sqrt{3}k_y/2)$, and $E=2t_{yz}\cos(-k_x/2+\sqrt{3}k_y/2)$, respectively, as shown in the left panel of Fig. 5(a). Here, the unit for k_x and k_y is $1/a$ where a is the lattice constant of the triangular lattice. When the three one-dimensional bands accommodate one hole as shown in Fig. 5(a), one-dimensional Fermi surfaces with 5/6 filling are created as illustrated in the middle panel of Fig. 5(a). In a realistic system, the three bands are mixed by the transfer terms between them ($t_{yz\text{-}zx}$, $t_{zx\text{-}xy}$, and $t_{xy\text{-}yz}$) and consequently, the Fermi surfaces become two-dimensional as shown in the right panel of Fig. 5(a). When the three one-dimensional bands accommodate one electron as shown in Fig. 5(b), one-dimensional Fermi surfaces with 1/6 filling are created as illustrated in the middle panel of Fig. 5(b). With the transfer terms $t_{yz\text{-}zx}$, $t_{zx\text{-}xy}$, and $t_{xy\text{-}yz}$, six electron pockets around the M point and one hole pocket at the zone center are created as shown in the right panel of Fig. 5(b). Similarly to the square lattice model, the triangular lattice can by deformed along one of the bond directions (The equilateral triangle is deformed into an isosceles one). Let us assume that the triangle is compressed along the horizontal axis (X-axis). Since $t_{xy}>t_{yz}=t_{zx}$ under the lattice distortion, only the wider xy band can accommodate the electron as illustrated in the left panel of Fig. 5(c). The xy band forms a one-dimensional band with nesting vector Q [see the middle panel of Fig. 5(c)]. The orbitally induced Peierls state is robust against the indirect transfer terms $t_{xy\text{-}yz}$ and $t_{xy\text{-}yz}$ due to the Jahn-Teller energy splitting between the xy and yz/zx orbitals. Under the effect of $t_{xy\text{-}yz}$ and $t_{xy\text{-}yz}$, a quasi one-dimensional Fermi surface dominated by xy character still keeps the nesting condition. Since the xy band accommodates one electron and is half-filled, the periodicity along the X-axis is doubled with xy-xy dimers as shown in the right panel of Fig. 5(c).

Under the strong trigonal ligand field, the yz, zx, and xy orbitals are reconstructed to be the a_{1g} and e_g^π orbitals. This situation is relevant for the corundum system where the $M\mathrm{O}_6$ octahedra share their faces along the c-axis. Yet, in several triangular lattice systems such as $\mathrm{Na}_x\mathrm{CoO}_2$, the a_{1g} and e_g^π orbitals are also more convenient to describe their Fermi surfaces. For example, when x is larger than 0.5 in $\mathrm{Na}_x\mathrm{CoO}_2$, the pudding-mold like a_{1g} bands form isotropic Fermi surfaces and the e_g^π orbitals are fully occupied [see the left and middle panels of Fig. 5(d)] [26, 27]. If the anisotropic hybridization between the a_{1g} and one of the e_g^π orbitals deforms the Fermi surface for better nesting with Q' as illustrated in the right panel of Fig. 5(d), then the orbitally induced Peierls mechanism is triggered to provide charge and orbital ordering. Such an anisotropic a_{1g}-e_g^π coupling would be induced, for example, by the anisotropic Na ion arrangement in $\mathrm{Na}_x\mathrm{CoO}_2$ [28, 29].

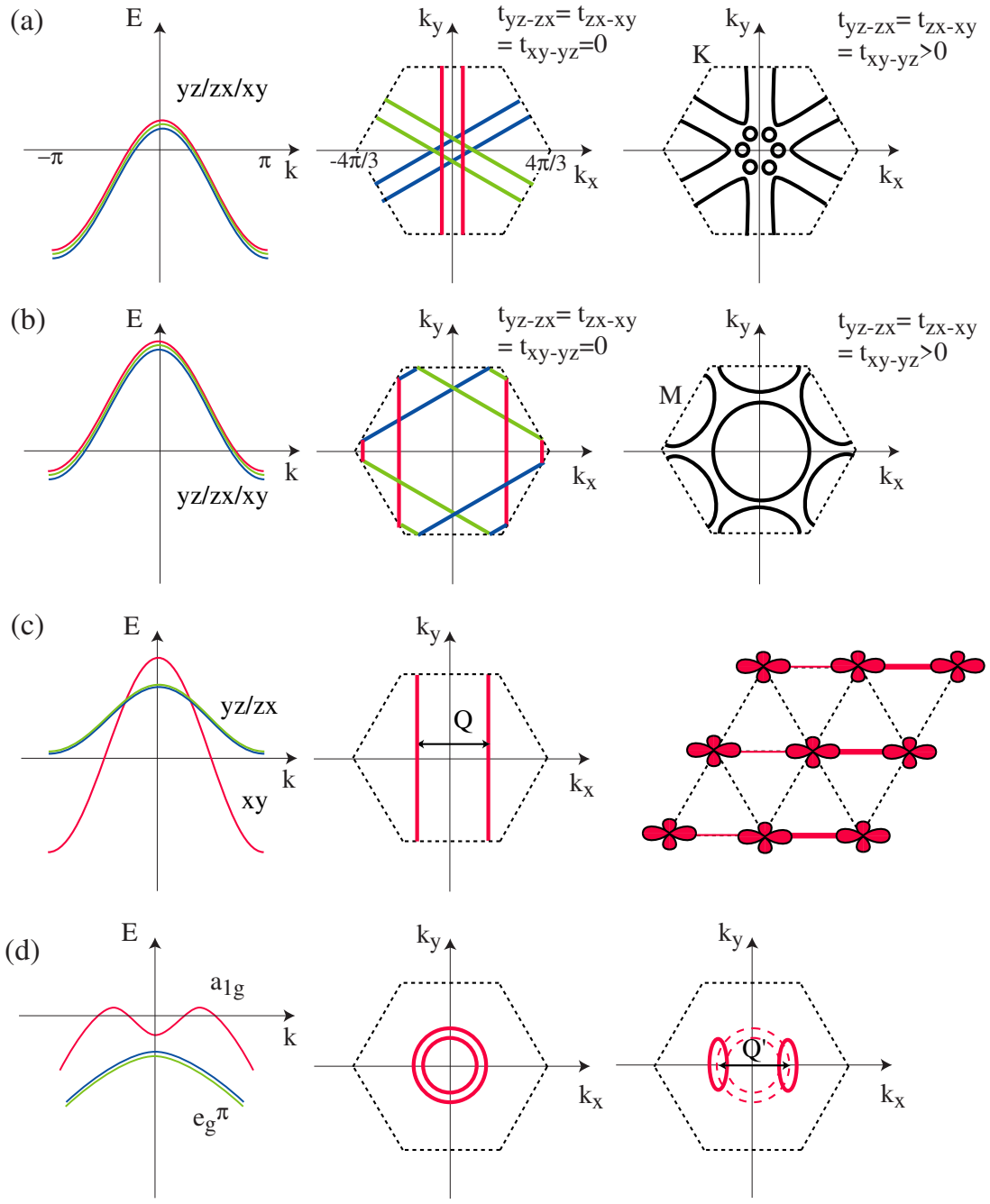


Fig. 5: *(a) Left: Band dispersion for a triangular lattice three-band model ($yz/zx/xy$ orbitals) with one hole per site. k represents k_x or $k_x/2+\sqrt{3}k_y/2$ or $k_x/2-\sqrt{3}k_y/2$. $t_{yz\text{-}zx}$, $t_{zx\text{-}xy}$, and $t_{xy\text{-}yz}$ are the inter-orbital transfer terms between the $yz/zx/xy$ orbitals. Middle and right: Fermi surfaces without and with interorbital transfer terms. (b) Band dispersion, Fermi surfaces without and with inter-orbital transfer terms for the three-band model with one electron per site. (c) Left and middle: Band structure and Fermi surfaces with band Jahn-Teller effect. Right: Orbital ordering by the orbitally induced Peierls effect on the triangular lattice (half filled case). (d) Left and middle: Band dispersion and a_{1g} Fermi surface with fully occupied e_g^π. Right: Fermi surfaces deformed by anisotropic a_{1g}-e_g^π mixing. Q' indicates a possible nesting vector.*

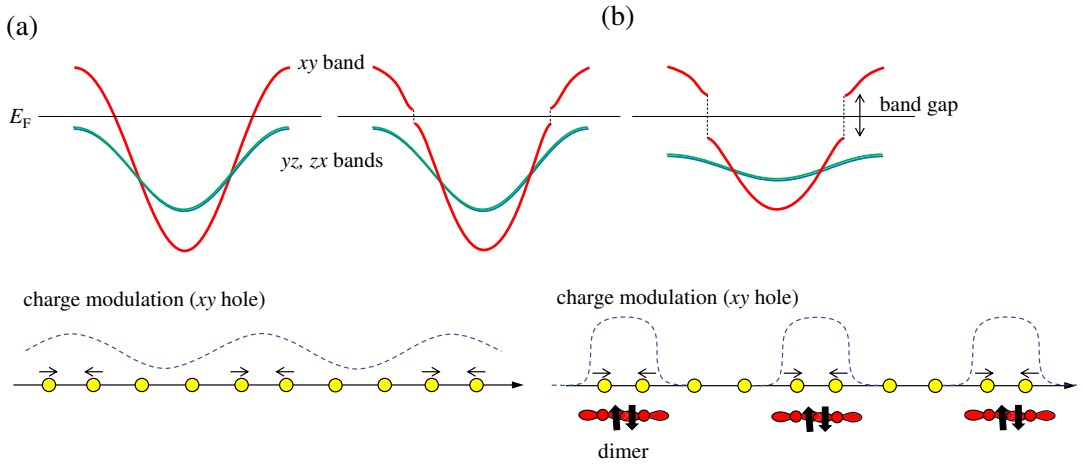


Fig. 6: *(a) Band gap opening and charge modulation for the weak coupling case. (b) Band gap opening and charge modulation for the strong coupling case.*

2.4 Effect of strong coupling

In the previous arguments, it is implicitly assumed that the electron lattice interaction (or the d-d Coulomb interaction) responsible for the Peierls transition is assumed to be a weak perturbation to the metallic state. Therefore, perfect Fermi-surface nesting is a necessary condition for the Peierls transition. In such a weak-coupling case, the charge and lattice modulation is sinusoidal with relatively small amplitude as shown in Fig. 6(a). The mean field treatment works well to describe the transition, and the energy gap E_G and the transition temperature T_c should satisfy $E_G/k_BT_c \sim 3.5$. However, even in one of the most weakly correlated systems $CuIr_2S_4$, E_G/k_BT_c is about 10 (E_G is about 0.2 eV and T_c is 230 K). This indicates that the electron-lattice interaction is rather strong and that local charge- and lattice-fluctuations can survive even in the disordered phase. Under strong coupling, the charge is localized at the strongly distorted bond as shown in Fig. 6(a), and its fluctuation remains above T_c. Indeed, Ir-Ir dimers are observed in the cubic phase above T_c for $CuIr_2S_4$ [25]. Most of the transition-metal compounds discussed in the next chapter fall in the strong coupling regime.

3 Application of the orbitally induced Peierls mechanism to transition-metal compounds

3.1 Spinel systems

Apart from the metal-insulator transition in $CuIr_2S_4$, the orbitally induced Peierls mechanism can be applied to that of $LiRh_2O_4$ around 170 K accompanied by the Rh^{3+}/Rh^{4+} charge ordering [30, 31]. The charge ordering with four-times periodicity of Rh^{3+}-Rh^{3+}-Rh^{4+}-Rh^{4+} occurs along the $(1, 1, 0)$ and $(1, -1, 0)$ chains of the Rh pyrochlore lattice. This is similar to that of CuI_2S_4 and can be explained by the orbitally induced Peierls instability. However, the experimentally observed charge-ordering pattern of $LiRh_2O_4$ is different from the octamer structure

of $CuIr_2S_4$ and satisfies the Anderson condition where each Rh_4 tetrahedron exactly accommodates two Rh^{4+} to minimize the Coulomb repulsion energy between them [31]. As a result, the Rh^{3+}-Rh^{3+}-Rh^{4+}-Rh^{4+} arrangement along the $(1, 1, 0)$ chain is shifted relative to that along the neighboring $(1, -1, 0)$ chain as shown in Fig. 2(c). Shiomi *et al*. proposed that the electronic correlation, which is expected to be stronger in the $4d$ oxides than the $5d$ sulfides, is responsible for destabilization of the octamer structure which does not satisfy the Anderson condition [31]. If the electronic correlation is negligibly weak, the Rh_4 tetrahedron can be fully occupied by Rh^{4+} keeping the Rh^{3+}-Rh^{3+}-Rh^{4+}-Rh^{4+} arrangement along the $(1, \pm 1, 0)$ chains as illustrated in Fig. 2(d).

$LiRh_2O_4$ undergoes the cubic to tetragonal transition ($c > a$) at 220 K which is followed by a Rh^{3+}/Rh^{4+} charge ordering and the Rh^{4+}-Rh^{4+} dimerization. The Rh^{3+}-Rh^{4+} charge fluctuations [32] and the local Rh^{4+}-Rh^{4+} dimers [31] are observed between 220 K and 170 K, indicating that orbital symmetry-breaking plays a more important role in $LiRh_2O_4$. In the case of $CuIr_2S_4$, the local Ir-Ir dimers are observed even in the cubic phase [25], while the Rh-Rh dimers are observed only in the tetragonal phase. Most probably, the bond directions of the Rh-Rh dimers tend to be aligned through the stronger electronic correlation. While the electron-lattice interaction (or the dimerization) drives the distortion with the four-times periodicity of the chains in $CuIr_2S_4$, the inter-site electron-electron interaction plays a primary role in $LiRh_2O_4$.

$MgTi_2O_4$ exhibits a metal-insulator transition around 260 K which is accompanied by a structural transition from cubic to tetragonal ($c < a$) [33]. The metal-insulator transition is accompanied by the Ti^{3+}-Ti^{3+} dimerization with suppression of magnetic susceptibility. The nearest-neighbor Ti-Ti distances become nonequivalent and spirals of long and short Ti-Ti bonds are formed [34]. In the itinerant picture, the yz and zx bands get wider than the xy band under the tetragonal distortion and accommodate the Ti $3d$ electrons. Consequently, the one-dimensional yz and zx bands are formed along the $(0, 1, \pm 1)$ and $(1, 0, \pm 1)$ directions which respectively accommodate 0.5 electron per Ti site. The quarter filled yz and zx bands are stabilized by the orbital ordering with four-times periodicity of yz-yz and zx-zx dimers as shown in Fig. 7(a). The orbital ordering is consistent with the spirals of long and short Ti-Ti bonds. It is also possible to explain the orbital ordering based on the Kugel-Khomskii mechanism in the localized picture [35] although it is difficult to describe the metal-insulator transition. Starting from the itinerant picture, Heitler-London like correlation effects can be included to provide the Ti-Ti dimer with the spin-singlet bond character. The localized nature of the Ti-Ti dimer has experimentally been suggested from the survival of the local Ti-Ti dimers above the transition temperature [36] and the multiplet structure of the Ti $2p$ x-ray absorption spectrum [37]. Interestingly, both the itinerant model and the localized model provide the same conclusion on the orbital ordering of $MgTi_2O_4$. Also it should be noted that the LDA+U like approach cannot describe the spin-singlet ground state of $MgTi_2O_4$.

AlV_2O_4 harbors V^{2+}/V^{3+} charge ordering along the $(1, 1, 1)$ direction [38, 39] and exhibits peculiar V^{3+} trimerization and V^{2+} tetramerization [40]. (The combination of the trimer and the tetramer can be viewed as a heptamer [39].) Along the $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$ chains

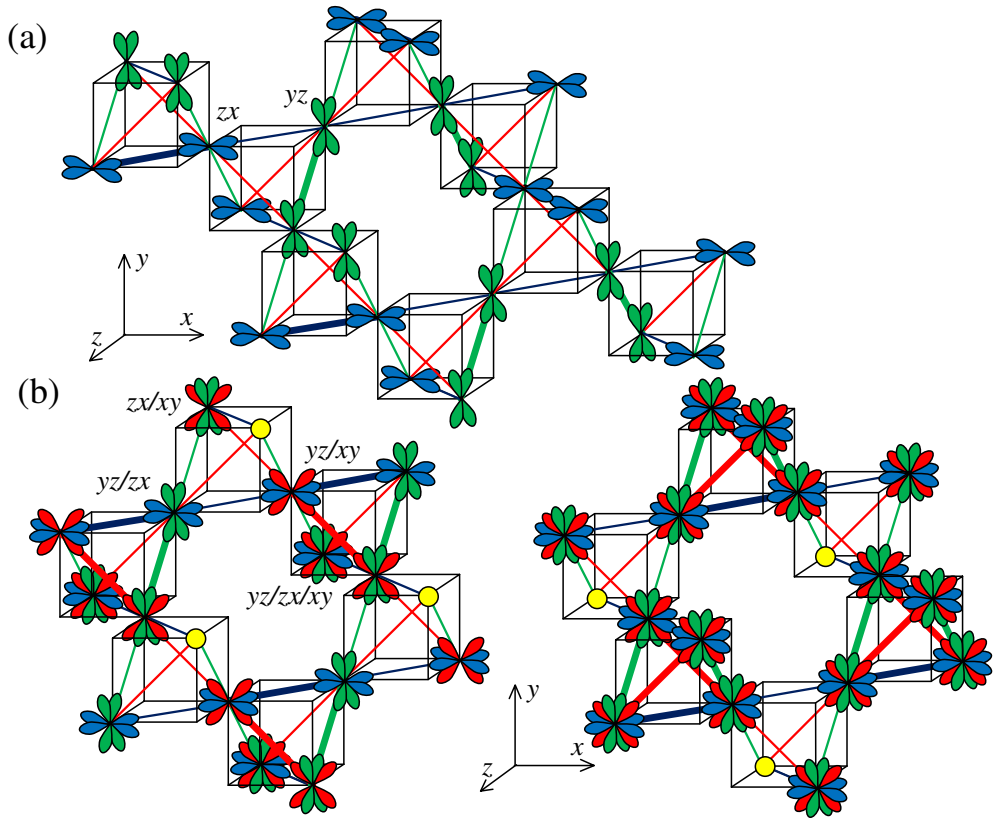


Fig. 7: *(a) yz-yz-zx-zx orbital ordering for $MgTi_2O_4$. (b) V^{2+}/V^{3+} charge ordering for AlV_2O_4. The yellow circles indicate the orbitally disordered V^{3+} sites. In the left panel for the V^{3+} Kagome layer, the V^{3+} trimers are formed by the yz/zx, zx/xy, and yz/xy orbital ordering. In the right panel for the V^{2+} Kagome layer, the tetramers are formed with the V^{2+} sites ($yz/zx/xy$) in the neighboring triangular lattice layer.*

which cross the $(1,1,1)$ Kagome and triangular lattice layers, the xy, zx, and yz bands are quarter filled, respectively. Similarly to $CuIr_2S_4$ and $MgTi_2O_4$, each chain is expected to have a lattice distortion with four-times periodicity due to V^{2+}(Kagome)-V^{2+}(tri)-V^{3+}(Kagome)-V^{3+}(tri) charge ordering. Along the $(1,-1,0)$, $(1,0,-1)$, and $(0,1,-1)$ chains in the Kagome layers, the xy, zx, and yz bands are half filled respectively. The Peierls instability gives lattice distortion with two-times periodicity (alternating short and long bonds) to each chain, resulting in the trimer formation. As a result, while the V^{3+} sites harbor the trimers confined in the Kagome layer, the V^{2+} sites form the tetramers between the Kagome and triangular lattice layers. We speculate that the quarter-filled chains trigger this peculiar charge ordering and multimer formation.

The magnetite Fe_3O_4 harbors Fe^{3+}-Fe^{2+}-Fe^{3+} trimers below the Verwey transition temperature at 125 K [41–43]. In Fe_3O_4, since the minority spin t_{2g} electrons are tied to the majority spin t_{2g} and e_g electrons via Hund coupling at the Fe^{2+} site, it is not straightforward to apply the orbitally induced Peierls mechanism to describe the trimer. However, it is still possible to discuss the Fe^{3+}-Fe^{2+}-Fe^{3+} arrangement along the chains based on a half filled spin-polarized t_{2g} band. For example, when the yz band is selected by the minority spin electron, it provides site-centered charge modulation along the $(0,1,\pm1)$ chains forming an Fe^{3+}-Fe^{2+}-Fe^{3+} arrangement.

3.2 Pyrochlore systems

Several pyrochlore systems exhibit interesting metal-insulator transitions with possible charge or orbital ordering. However, none of them exhibit band Jahn-Teller like distortions. Since $M\mathrm{O}_6$ octahedra share their conners in the pyrochlore structure, their t_{2g} band dispersions are strongly affected by the indirect t_{2g}-ligand-t_{2g} transfer terms and become more complicated than those of the spinel structure. Therefore, the energy gain by the band Jahn-Teller distortion, if it may exist, tends to be reduced.

Among the transition-metal oxides with pyrochlore structure, $\mathrm{CsW_2O_6}$ exhibits a unique metal-insulator transition around 215 K which is accompanied by W trimers [44–46]. The crystal symmetry of the insulating phase just below the transition is still cubic although the tetrahedron of W sites in the pyrochlore lattice is strongly distorted by the trimerization [46]. While $\mathrm{Ir^{3+}/Ir^{4+}}$ ($\mathrm{Rh^{3+}/Rh^{4+}}$) charge disproportionation is observed in $\mathrm{CuIr_2S_4}$ ($\mathrm{LiRh_2O_4}$), $\mathrm{W^{5+}/\ W^{6+}}$ charge disproportionation is absent in $\mathrm{CsW_2O_6}$ [47]. The absence of the band Jahn-Teller effect and the charge disproportionation is consistent with the complicated band structure. Streltsov *et al*. proposed that the Fermi surfaces calculated with the W $5d$ spin-orbit interaction satisfy nesting conditions with $(\pi, 0, 0)$, $(0, \pi, 0)$, and $(0, 0, \pi)$ and the metal-insulator transition can be explained by the simple Peierls mechanism [48]. However, it is still difficult to explain the trimer formation. Nakai and Hotta emphasized the electronic correlation effect due to a flat band created by the pyrochlore lattice geometry and the spin-orbit interaction [49]. Under the strong spin-orbit interaction, there are $\frac{1}{\sqrt{2}}(y'z'+\mathrm{i}z'x')\uparrow$, $\frac{1}{\sqrt{2}}(y'z'-\mathrm{i}z'x')\downarrow$, $\frac{1}{\sqrt{6}}[2x'y'\uparrow-(y'z'+\mathrm{i}z'x')\downarrow]$, and $\frac{1}{\sqrt{6}}[2x'y'\downarrow+(y'z'-\mathrm{i}z'x')\uparrow]$ in the $j = 3/2$ branch. (Here, the x'-, y'-, and z'-axes are along the M-O bonds of the octahedron. In the spinel case, they are identical to the x-, y-, and z-axes in the figures.) The former two orbitals have substantial transfer integrals along the z'-axis which is roughly along the M-ligand-M bond or the chain direction. If such orbitals are selectively occupied, they can form a quasi one-dimensional band along one of the chains. The virtual one-dimensional bands are quarter filled and provide the lattice distortion with four-times periodicity along the chain. Once the short W-W bond is created, the two $\mathrm{WO_6}$ octahedra with the shortened W-W bond are rotated and the other $\mathrm{WO_6}$ octahedron connected to the two octahedra gets closer to the W-W bond. As a result, W trimers can be formed as shown in Fig. 8(a). In addition, such orbitals would be consistent with the zigzag chain order which was proposed by Hirai *et al*. for the low temperature phase [45]. If $z'x' \pm \mathrm{i}x'y'$ and $(x'y' \pm \mathrm{i}y'z')$ orbitals are occupied in the $(0, 1, 1)$ and $(1, 0, 1)$ chains (or $(0, 1, -1)$ and $(1, 0, -1)$ chains), the zigzag structure of the short W-W bonds can be formed through the distortion with the four-times periodicity along the chains.

There is no clear understanding for the difference between the multimer case and the zigzag chain case. Most probably, the rotation and Jahn-Teller distortion of the $M\mathrm{O}_6$ octahedra should be analyzed more carefully. The zigzag chains are also formed in the nonmagnetic insulating phase of $\mathrm{Tl_2Ru_2O_7}$ [50].

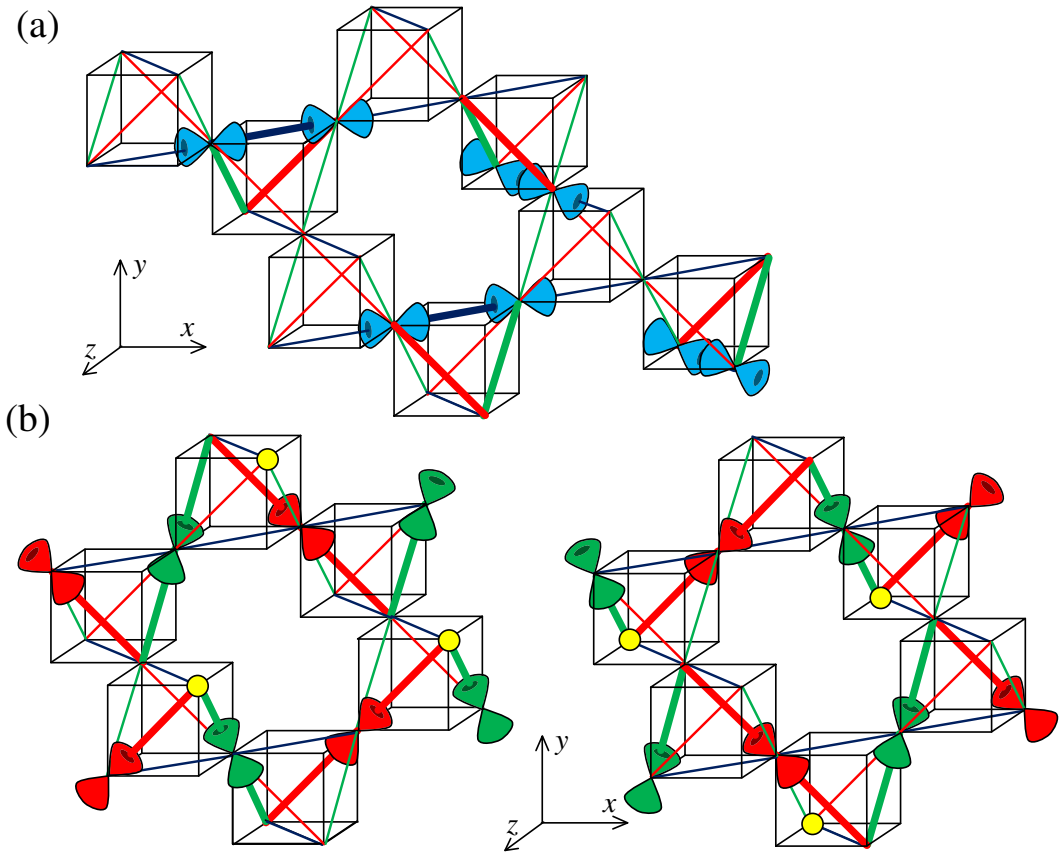


Fig. 8: *(a) W trimers and possible orbital ordering for CsW_2O_6. (b) Zigzag chains of short W-W bonds. The thick solid lines indicate the short bonds.*

3.3 Triangular lattice systems

The MO_6 octahedra share their corners and form a triangular lattice of M sites as shown in Fig. 9(a). Among the triangular lattice systems, $NaTiO_2$ exhibits a transition to the nonmagnetic insulating state with lattice distortion from trigonal to monoclinic [51, 52]. The direct t_{2g}-t_{2g} transfer term would be important just like in the spinel systems. However, in band structure calculations for $NaTiO_2$ [53] as well as CoO_2, the t_{2g} band dispersion is upwardly convex around the Γ point indicating the indirect M-O-M transfer is dominant. In addition, the trigonal ligand field can break the three-fold degeneracy and the t_{2g} orbitals are split into the a_{1g} $[\frac{1}{\sqrt{3}}(xy+yz+zx)]$ and e_g^π $[\frac{1}{\sqrt{3}}(xy+e^{\pm 2\pi i/3}yz+e^{\pm 4\pi i/3}zx)]$ orbitals. As shown in the left panel of Fig. 5(b), the xy, yz, and zx orbitals may form one-dimensional bands along the $(1,0,0)$, $(1/2,\sqrt{3}/2,0)$, and $(-1/2,\sqrt{3}/2,0)$ directions of the X-, Y-, and Z-coordinates. Inclusion of the mixing between the xy, yz, and zx orbitals provide the one hole pocket at the zone center (with a_{1g} character) and the six hole pockets around the M points (with e_g^π character) for the t_{2g}^1 system as shown in the right panel of Fig. 5(b). The Fermi surfaces calculated for the trigonal phase by Subedi [53] are roughly consistent with this simple picture except the hole pocket at the zone center. The absence of the hole pocket can be assigned to the trigonal ligand field which stabilizes the a_{1g} orbital. If the xy orbital has larger band width due to the contraction

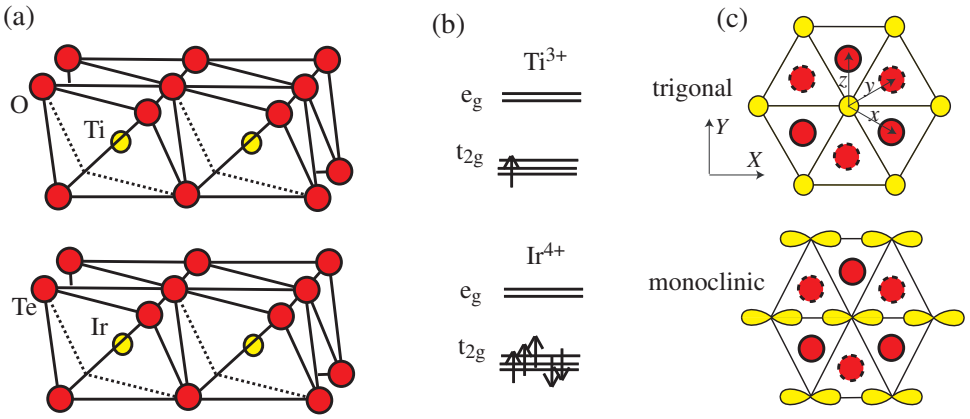


Fig. 9: *(a)Triangular lattice layers of edge-sharing MO_6 octahedra. (b) Electronic configurations for Ti^{3+} and Ir^{4+}. (c) Orbital ordering for the monoclinic phase.*

along the X direction [$X = (x-y)/\sqrt{2}$], the Ti $3d$ electron is accommodated by the quasi one-dimensional xy band. Since the xy band is half-filled [Fig. 9(b)], the Ti-Ti dimerization with the two-times lattice modulation is realized by the Peierls transition [Fig. 9(c)]. However, the triangular lattice of $NaTiO_2$ is not clearly deformed from equilateral to isosceles. Instead, the TiO_6 octahedron is compressed to stabilize the xy orbital [52].

$IrTe_2$ with the Ir triangular lattice exhibits a structural phase transition at ~ 270 K from trigonal to monoclinic, accompanied by anomalies in electrical resistivity and magnetic susceptibility [54–57] An electron diffraction study by Yang *et al.* shows that the structural transition is accompanied by a superstructure with wave vector of $Q = (1, 0, -1)/5$ [57]. Such a superstructure can be explained by a charge-density wave driven by a perfect or partial nesting of multi-band Fermi surfaces [57]. However, the monoclinic distortion can be attributed to a band Jahn-Teller like instability, suggesting the orbitally induced Peierls mechanism. In addition, a charge modulation of Ir charge disproportionation is indicated by an Ir $4f$ x-ray photoemission study [58]. Although the formal valence of Ir is +4, partial charge transfer from Te to Ir induces a Ir^{3+}/Ir^{4+} mixed valence and the Ir^{4+}-Ir^{4+} dimers are arranged as illustrated in Fig. 10(b).

$LiVO_2$ is a classical system with V^{3+} (t_{2g}^2) which exhibits a magnetic-nonmagnetic transition around 500 K with V trimerization [1, 59–61]. Pen *et al.* pointed out that the xy, yz, and zx orbitals are occupied at the V sites connected by the $(1, 0, 0)$, $(1/2, \sqrt{3}/2, 0)$, and $(-1/2, \sqrt{3}/2, 0)$ bonds of the X-, Y-, and Z-coordinates (see Fig. 5) [62]. When yz/zx, zx/xy, and xy/yz orbitals are occupied, respectively, at the three V sites in the trimer, the three V^{3+} form three singlet bonds resulting in the nonmagnetic ground state with the trimer. Although the multiplet structure of V $2p$ x-ray absorption spectroscopy of $LiVO_2$ indicates localized V $3d$ electrons with Hund coupling, it is still possible to make a spin-singlet ground-state from three V^{3+} in the trimer [63]. Interestingly, such trimerization can be described by the orbitally induced Peierls mechanism or the orbital selective Peierls mechanism [64] in a manner similar to $MgTi_2O_4$. The triangular lattice can be decomposed into chains running along $(1, 0, 0)$, $(1/2, \sqrt{3}/2, 0)$, and $(-1/2, \sqrt{3}/2, 0)$ directions. Therefore, the xy, zx, and yz orbitals can form quasi one-dimensional bands along the three directions which are 2/3 filled and can induce orbital order-

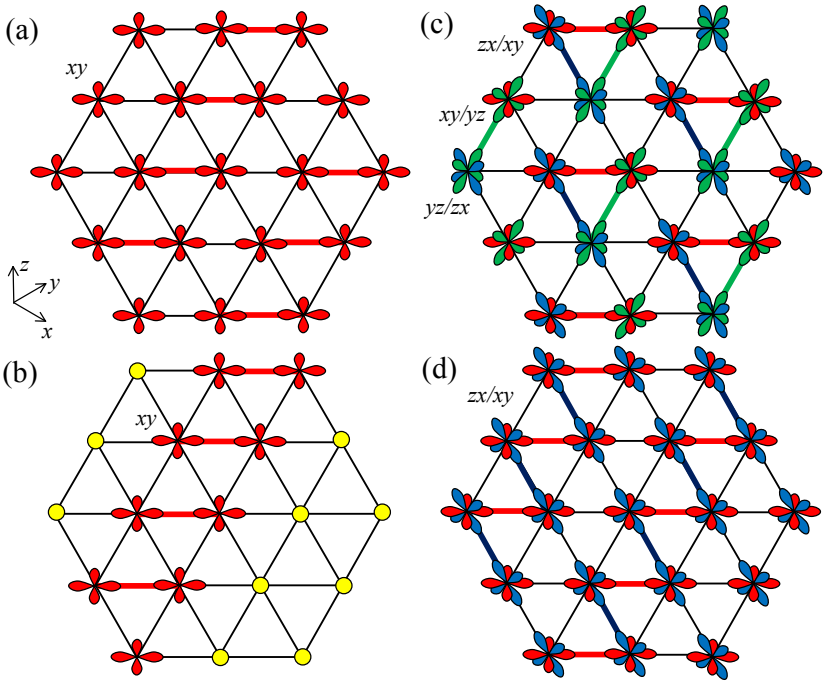


Fig. 10: *(a) Orbital ordering in $NaTiO_2$ (b) Charge and orbital ordering in $IrTe_2$. (c and d) Orbital ordering in $LiVO_2$ and $LiVS_2$. The thick solid lines indicate the short bonds.*

ing and dimerization with three-times periodicity. The xy-xy, zx-zx, and yz-yz dimers are formed along the $(1,0,0)$, $(1/2,\sqrt{3}/2,0)$, and $(-1/2,\sqrt{3}/2,0)$ chains, respectively. Here, the degenerate xy, yz, and zx bands are strongly modified by the inter-orbital transfer terms ($t_{yz\text{-}zx}$, $t_{zx\text{-}xy}$, and $t_{xy\text{-}yz}$,) and lose the nesting condition as illustrated in Fig. 5(a) or (b). Therefore, the electron-lattice interaction should be strong enough to stabilize the dimer bonds even without perfect Fermi-surface nesting.

The dimer bonds are indicated by the thick lines in Fig. 10(c). The xy-xy and zx-zx dimers overlap at the intersection site of the $(1,0,0)$ and $(1/2,\sqrt{3}/2,0)$ chains where both of the xy and zx orbitals are occupied. The $(-1/2,\sqrt{3}/2,0)$ chain goes through the other site of the xy-xy (zx-zx) dimer, and the xy and yz (zx and yz) orbitals are occupied at the site. As a result, the three V^{3+} sites are connected by the xy-xy, zx-zx, and yz-yz singlet bonds as shown in Fig. 10(c). Here, it should be noted that the singlet bond picture is inconsistent with the Hund coupling of the t_{2g}^2 configuration which may weaken the V-V bond. It is still an open question whether the trimer is sustainable against the electronic correlation although its stability was suggested by cluster model calculations [63]. $Li_{2-x}MoO_3$ with Mo^{4+} (t_{2g}^2) on a triangular lattice also exhibits Mo trimers similar to $LiVO_2$ [65]. The Mo trimers can survive up to the highest temperature indicating the trimer due to the orbitally induced Peierls mechanism or the orbital selective Peierls mechanism is more stable in the more itinerant systems. Orbitally induced Peierls transitions can be classified into two categories: orbital restrictive or orbital selective. In case of $CuIr_2S_4$ and $LiRh_2O_4$, the t_{2g} orbital is restricted by band Jahn-Teller effect. As for

AlV_2O_4 and $LiVO_2$, one of the t_{2g} orbitals is selected depending on the bond direction.
Katayama *et al*. found that $LiVS_2$ undergoes a metal-insulator transition at 314 K which is accompanied by V trimerization similar to $LiVO_2$ [66]. Since the V $3d$ electrons are more itinerant in $LiVS_2$ than $LiVO_2$, the orbitally induced Peierls description is more suitable. Also in $LiVS_2$, the xy-xy, zx-zx, and yz-yz singlet bonds are stabilized by the strong lattice distortion. E_G/k_BT_c (E_G is the magnitude of the band gap, and T_c is the transition temperature.) is about 6 indicating strong coupling [67]. Indeed, short range order of the zigzag chain structure and lattice fluctuations are observed above the transition temperature [68]. Interestingly, the zigzag chain structure can be stabilized by the ferro-type orbital order as shown in Fig. 10(d). In the figure, the xy and zx orbitals are occupied at every V site and the xy-xy and zx-zx dimers form the zigzag chains. There are two other domains where the zx and yz orbitals or yz and xy orbitals are occupied at every site. Since the d electrons in $LiVS_2$ are rather itinerant, the band Jahn-Teller effect and the Fermi-surface nesting can collaborate to provide the ferro-type orbital order and the lattice modulation with three-times periodicity.
$BaV_{10}O_{15}$ with V^{2+}/V^{3+} mixed valence exhibits a structural transition at 123 K driven by V trimerization [69] and V $3d$ orbital order similar to $LiVO_2$ [70]. At the three V^{3+} sites in the trimer, the xy, yz, and zx orbitals are unoccupied respectively. It is possible to create molecular orbitals from them and to put an extra electron in the bonding molecular orbital. Therefore, the trimer is expected to be stable against electron doping (up to the doping level of 1/3 per V site). Indeed, the V trimer in $BaV_{10}O_{15}$ is likely to accommodate one extra electron. The extra electron is shared by the three V sites. In addition, V^{2+}/V^{3+} charge fluctuation is observed by x-ray photoemission spectroscopy [71] indicating that the ground state is a superposition of V^{2+}-V^{3+}-V^{3+}, V^{3+}-V^{2+}-V^{3+}, and V^{3+}-V^{3+}-V^{2+} configurations.

3.4 Honeycomb lattice systems

Since the honeycomb lattice cannot be decomposed into chains, it is not straightforward to apply the orbitally induced Peierls description to its charge-orbital ordering. The dimerization on the honeycomb lattice can be more exactly described by the theory of Jackeli and Khomskii [72] in which the orbital degeneracy is broken by the bond selection. However, in this subsection, following the idea of orbitally induced Peierls mechanisms, we would like to discuss dimerization induced by possible orbital order.
Assuming ferro-type orbital order of xy, local xy-xy bonds can be created for t_{2g}^1 or t_{2g}^5 systems as shown in Fig. 11(a). Since the number of xy-xy bonds is limited, the energy gain by the singlet bond formation may not be enough for a metal-insulator transition by orbitally induced Peierls mechanism. However, several insulating systems undergo magnetic-nonmagnetic transitions by dimerization. $TiCl_3$ and $TiBr_3$ with Ti^{3+} (t_{2g}^1) become nonmagnetic with Ti-Ti dimerization below 217 K [73] and below 178 K [74], respectively . The dimerization can be described by ab-initio band structure calculations [75,76]. Ilmenite $MgVO_3$ with V^{4+} (t_{2g}^1) also exhibits V-V dimers on the honeycomb lattice below 500 K [77]. Similar to $TiCl_3$ and $TiBr_3$, ferro-type orbital order is expected as shown in Fig. 11(a).

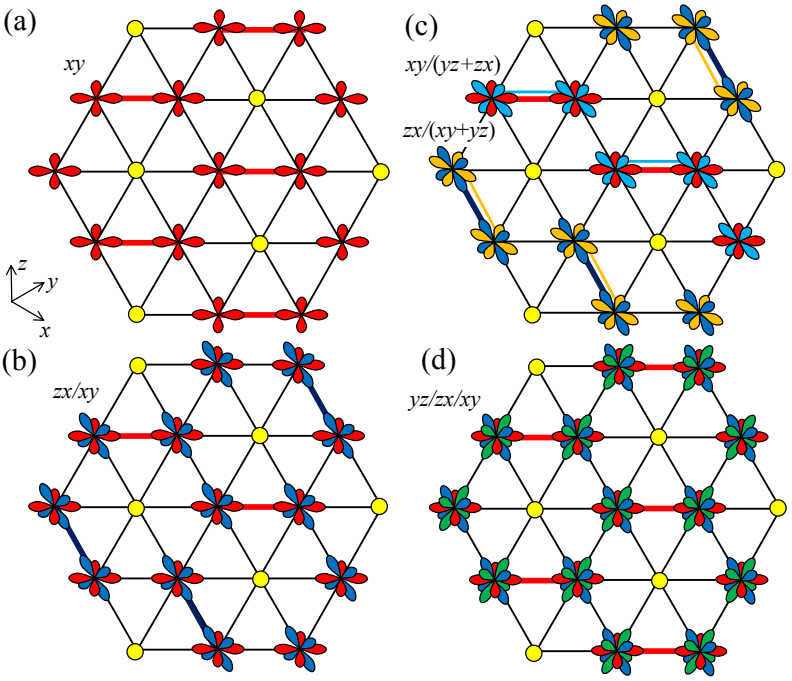


Fig. 11: *(a) Orbital ordering for t_{2g}^1 on a honeycomb lattice. (b and c) Orbital ordering for t_{2g}^2 on a Kagome lattice in Li_2RuO_3. (d) Orbital ordering for t_{2g}^3 on a Kagome lattice in $MoCl_3$. The thick solid lines indicate the short bonds.*

$4d$ and $5d$ honeycomb systems with t_{2g}^1 and t_{2g}^5 configurations (such as Na_2IrO_3 and $RuCl_3$) do not show dimerization at ambient pressure due to the strong spin-orbit interaction [78, 79]. In this context, it is striking that Li_2RuO_3 exhibits a metal-insulator transition about 540 K due to Ru-Ru dimerization on the honeycomb lattice [80]. In Li_2RuO_3, the Ru^{4+} (t_{2g}^4) ion has two holes in the t_{2g} orbitals. Assuming ferro-type orbital order of xy and yz holes, xy-xy and yz-yz singlet bonds can be created as illustrated in Fig. 11(b). Although the arrangement of the short Ru-Ru bonds (indicated by the thick lines in the figure) is consistent with the experimental result [80], the number of these bonds is limited, and the energy gain by the singlet bonds may not be enough. Kimber *et al*. pointed out that the $\frac{1}{\sqrt{2}}(yz+zx)$ orbitals form π-bonds in addition to the xy-xy σ-bonds [81]. Under the orbital order with yz, $\frac{1}{\sqrt{2}}(zx+xy)$ and xy, $\frac{1}{\sqrt{2}}(yz+zx)$, the short Ru-Ru bonds are stabilized by the double bonding in the nonmagnetic phase of Li_2RuO_3. Interestingly, a recent experimental work reports that the xy-xy σ-bonding is more robust than the π-bonding by the yz and zx orbitals and that the partially disordered phase can be created by optical breaking of the π bonding [82]. $TcCl_3$ with Tc^{3+} (t_{2g}^4) has the Tc-Tc dimers on the honeycomb lattice [83]. The arrangement of the dimers is the type of Fig. 11(a) rather than that of Fig. 11(c).

Another striking system is $MoCl_3$ which undergoes a magnetic-nonmagnetic transition around 585 K with the strong Mo-Mo dimerization below 585 K [84–86]. The Mo-Mo dimerization in the honeycomb lattice is illustrated in Fig. 11(d). The Mo^{3+} ion has three holes in the t_{2g} orbitals in $MoCl_3$. Therefore, in addition to the xy-xy σ-bonds and the $\frac{1}{\sqrt{2}}(yz+zx)$ π-bonds, the $\frac{1}{\sqrt{2}}(yz-zx)$ orbitals may form δ-bonds although δ-bonding is usually weak.

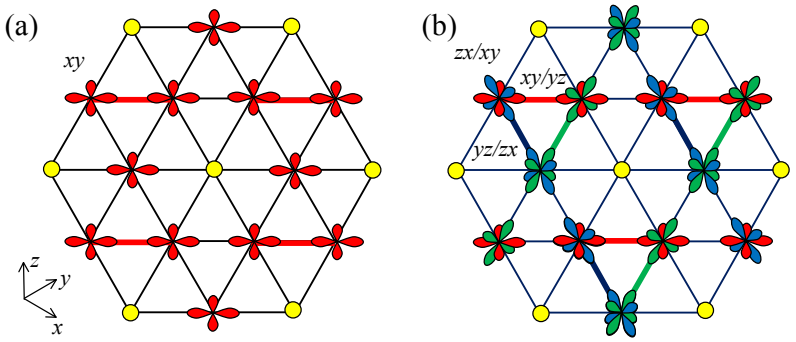


Fig. 12: *(a) Orbital ordering for t_{2g}^1 on a Kagome lattice. (b) Orbital ordering for $Na_2Ti_3Cl_8$. The thick solid lines indicate the short bonds.*

3.5 Kagome lattice systems

$Na_2Ti_3Cl_8$ consists of a Ti^{2+} (t_{2g}^2) Kagome lattice and undergoes a two step structural phase transitions around 210 K (partial trimerization) and 190 K (full trimerization) on cooling [87–89]. It is possible to describe the spin singlet trimer based on localized spins [90]. On the other hand, the substantial lattice distortion of the trimerization by itself suggests that a strong Ti-Ti interaction plays a vital role beyond the localized picture. Therefore, it is also useful to describe the trimerization based on the itinerant picture: an orbitally induced Peierls mechanism [91]. In contrast to the honeycomb lattice, the Kagome lattice can be decomposed into chains running along the $(1, 0, 0)$, $(1/2, \sqrt{3}/2, 0)$, and $(-1/2, \sqrt{3}/2, 0)$ directions. Therefore, the yz, zx, and xy orbitals can form quasi one-dimensional bands along the three directions. As illustrated in Fig. 12(a), each one-dimensional band is half filled and induces a lattice modulation with two-times periodicity (The short Ti-Ti bonds are indicated by the thick lines). The short Ti-Ti bonds along the three directions form the trimers in agreement with the experimental observation.
The Mo^{4+} (t_{2g}^2) Kagome lattice in $Zn_2Mo_3O_8$ hosts Mo trimers similar to $Na_2Ti_3Cl_8$ [92, 93]. The Mo trimers with six t_{2g} electrons are stable up to the highest temperature available indicating that the trimer is more stable in the more itinerant system. At the three Mo sites in the trimer, one can construct a molecular orbital with bonding character from the unoccupied t_{2g} orbitals to accommodate an extra electron. Indeed, the Mo trimers can survive in $LiZn_2Mo_3O_8$ with seven Mo $4d$ electrons in the Mo trimer [92, 93]. The Mo trimers have localized spin-1/2 due to the extra electron and form a geometrically frustrated spin-1/2 triangular lattice. This can be viewed as a cluster Mott insulating state [94, 95] providing exotic spin-liquid behaviors [92, 93]. Very recently, Mo^{3+}/Mo^{4+} charge fluctuation were observed by x-ray photoemission spectroscopy [96]. This situation is similar to the V trimer in $BaV_{10}O_{15}$ and is different from the W trimer in CsW_2O_6. A possible relationship between the charge-fluctuation and the spin-liquid behavior should be examined by further studies. Nb_3Cl_8 has a Kagome lattice with Nb trimers. Since Nb^{2+}: Nb^{3+} = 1 : 2 in Nb_3Cl_8, the Nb trimer accommodates seven t_{2g} electrons and hosts localized spin-1/2 [97–99]. The spin-1/2 triangular lattice of the cluster Mott insulating state is a new playground to study spin liquids and exotic superconductivity [100]. Interestingly, Nb_3Cl_8 exhibits a further magnetic-nonmagnetic transition around 90 K although

orbital degeneracy is already lifted. Haraguchi *et al*. proposed charge disproportionation between the trimers [99]. Apart from the fabrication of atomic layer systems, the origin of this exotic phase transition will be a target of future work.

4 Conclusion

By combining band Jahn-Teller effect (or local orbital polarization) and Peierls instability (or local singlet bond formation), the orbitally induced Peierls mechanism can explain charge-orbital ordering in a wide range of nonmagnetic (spin singlet) transition-metal compounds with edge-sharing octahedra and t_{2g} orbital degrees of freedom. With the ferro-type t_{2g} orbital ordering mostly due to the band Jahn-Teller effect, quasi one-dimensional bands are created by one of the t_{2g} orbitals. Consequently, charge ordering and/or dimerization are induced by Fermi-surface nesting. In some cases, intervening quasi one-dimensional bands provide trimers or tetramers. Without ferro-type orbital ordering, collaboration between local orbital polarization and singlet bond formation can stabilize nonmagnetic ground states with multimers. In most of the cases, both the itinerant model and the localized model predict the same charge-orbital order, probably due to the approximate one-to-one correspondence between the Fermi surface geometry and the bond direction. However, there are still several unsolved questions even in the most studied $CuIr_2S_4$. The energy landscape for various lattice distortion should be elucidated in order to fully understand the space and time fluctuation and evolution of the Ir-Ir dimers and the orbital polarization. Since the electronic correlation is stronger in the Ti oxides ($NaTiO_2$ and $MgTi_2O_4$) than the Ir chalcogenides, the Heitler-London like wave function should be taken into accounts to describe their electronic properties. Such a theoretical approach on the multiband lattice model is highly challenging. As for the t_{2g}^2 and t_{2g}^4 systems such as $LiVO_2$ and Li_2RuO_3, the effect of Hund coupling should be clarified theoretically and experimentally.

Acknowledgement

The authors would like to thank Prof. D.I. Khomskii and Prof. S.V. Streltsov for long term collaborations and their theoretical supports. Also the authors would like to thank Prof. L.H. Tjeng, Prof. G.A. Sawatzky, Dr. K. Takubo, Dr. M. Okawa, and Dr. D. Ootsuki for long term collaborations on synchrotron spectroscopy studies of various transition-metal compounds. This work was partially supported by Grants-in-Aid from the Japan Society of the Promotion of Science (JSPS) (No. JP22H01172).

References

[1] J.B. Goodenough: *Magnetism and the Chemical Bond*, (Interscience, New York, 1963)

[2] N.F. Mott: *Metal-insulator transitions* (Taylor-Francis, London, 1974)

[3] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039-1263 (1998)

[4] D.I. Khomskii: *Transition Metal Compounds* (Cambridge University Press, 2014)

[5] P. Monceau, Adv. Phys. **61**, 325 (2012)

[6] D.J. Singh and M.-H. Du, Phys. Rev. Lett. **100**, 237003 (2008)

[7] H.A. Jahn and E. Teller, Proc. Royal Soc. A **161**, 220 (1937)

[8] K.I. Kugel and D.I. Khomskii, Usp. Fiz. Nauk. **136**, 621 (1982)
[Sov. Phys. Usp. **231**, 25 (1982)]

[9] J. Kanamori, Prog. Theor. Phys. **17**, 177 (1957)

[10] T. Kimura, T. Goto, H. Shintani, K. Ishizaka, T. Arima, and Y. Tokura,
Nature **426**, 55 (2003)

[11] T. Mizokawa and A. Fujimori, Phys. Rev. B **54**, 5368 (1996)

[12] T. Mizokawa, L.H. Tjeng, G.A. Sawatzky, G. Ghiringhelli, O. Tjernberg, N.B. Brookes,
H. Fukazawa, S. Nakatsuji, and Y. Maeno, Phys. Rev. Lett. **87**, 077202 (2001)

[13] B.J. Kim, H. Jin, S.J. Moon, J.-Y. Kim, B.-G. Park, C.S. Leem, J. Yu, T.W. Noh,
C. Kim, S.-J. Oh, J.-H. Park, V. Durairaj, G. Cao, and E. Rotenberg,
Phys. Rev. Lett. **101**, 076402 (2008)

[14] D.I. Khomskii and T. Mizokawa, Phys. Rev. Lett. **94**, 156402 (2005)

[15] D.I. Khomskii and S.V. Streltsov, Chem. Rev. **121**, 2992 (2021)

[16] S. Nagata, T. Hagino, Y. Seki, and T. Bitoh, Physica B **194-196**, 1077 (1994)

[17] P.G. Radaelli, Y. Horibe, M.J. Gutmann, H. Ishibashi, C.H. Chen, R.M. Ibberson,
Y. Koyama, Y.S. Hor, V. Kirykhin, and S.-W. Cheong, Nature **416**, 155 (2002)

[18] T. Oda, M. Shirai, N. Suzuki and K. Motizuki, J. Phys.: Condens. Matter **7**, 4433 (1995)

[19] T. Sasaki, M. Arai, T. Furubayashi, and T. Matsumoto, J. Phys. Soc. Jpn. **73**, 1875 (2004)

[20] S. Sarkar, M. De Raychaudhury, and T. Saha-Dasgupta, Phys. Rev. B **79**, 113104 (2009)

[21] H. Ishibashi, T.Y. Koo, Y.S. Hor, A. Borissov, P.G. Radaelli, Y. Horibe, S.-W. Cheong, and
V. Kiryukhin, Phys. Rev. B **66**, 144424 (2002)

[22] K. Takubo, S. Hirata, J.-Y. Son, J.W. Quilty, T. Mizokawa, N. Matsumoto, and S. Nagata,
Phys. Rev. Lett. **95**, 246401 (2005)

[23] V. Kiryukhin, Y. Horibe, Y.S. Hor, H.J. Noh, S.-W. Cheong, and C.H. Chen, Phys. Rev. Lett. **97**, 225503 (2006)

[24] M. Naseska, P. Sutar, Y. Vaskivskyi, I. Vaskivskyi, D. Vengust, D. Svetin, V.V. Kabanov, D. Mihailovic, and T. Mertelj, New J. Phys. **23**, 053023 (2021)

[25] E.S. Bozin, W.G. Yin, R.J. Koch, M. Abeykoon, Y.S. Hor, H. Zheng, H.C. Lei, C. Petrovic, J.F. Mitchell, and S.J.L. Billinge, Nat. Commun. **10**, 3638 (2019)

[26] K. Kuroki and R. Arita, J. Phys. Soc. Jpn. **76**, 083707 (2007)

[27] K. Ikedo, Y. Wakisaka, S. Hirata, K. Takubo, and T. Mizokawa, J. Phys. Soc. Jpn. **78**, 063707 (2009)

[28] D.N. Argyriou, O. Prokhnenko, K. Kiefer, and C.J. Milne, Phys. Rev. B **76**, 134506 (2007)

[29] B. Raveau and Md.M. Seikh, Z. Anorg. Allg. Chem. **641**, 1385 (2015)

[30] Y. Okamoto, S. Niitaka, M. Uchida, T. Waki, M. Takigawa, Y. Nakatsu, A. Sekiyama, S. Suga, R. Arita, and H. Takagi, Phys. Rev. Lett. **101**, 086404 (2008)

[31] M. Shiomi, K. Kojima, N. Katayama, S. Maeda, J.A. Schneeloch, S. Yamamoto, K. Sugimoto, Y. Ohta, D. Louca, Y. Okamoto, and H. Sawa, Phys. Rev. B **105**, L041103 (2022)

[32] Y. Nakatsu, A. Sekiyama, S. Imada, Y. Okamoto, S. Niitaka, H. Takagi, A. Higashiya, M. Yabashi, K. Tamasaku, T. Ishikawa, and S. Suga, Phys. Rev. B **83**, 115120 (2011)

[33] M. Isobe and Y. Ueda, J. Phys. Soc. Jpn. **71**, 1848 (2002)

[34] M. Schmidt, W. Ratcliff, P.G. Radaelli, K. Refson, N.M. Harrison, and S.-W. Cheong, Phys. Rev. Lett. **92**, 056402 (2004)

[35] S. Di Matteo, G. Jackeli, C. Lacroix, and N.B. Perkins, Phys. Rev. Lett. **93**, 077208 (2004)

[36] L. Yang, R.J. Koch, H. Zheng, J.F. Mitchell, W. Yin, M.G. Tucker, S.J.L. Billinge, and E.S. Bozin, Phys. Rev. B **102**, 235128 (2020)

[37] T. Yamaguchi, M. Okawa, H. Wadati, T.Z. Regier, T. Saitoh, Y. Takagi, A. Yasui, M. Isobe, Y. Ueda, and T. Mizokawa, J. Phys. Soc. Jpn. **91**, 074704 (2022)

[38] K. Matsuno, T. Katsufuji, S. Mori, M. Nohara, A. Machida, Y. Moritomo, K. Kato, E. Nishibori, M. Takata, M. Sakata, K. Kitazawa, and H. Takagi, Phys. Rev. Lett. **90**, 096404 (2003)

[39] Y. Horibe, M. Shingu, K. Kurushima, H. Ishibashi, N. Ikeda, K. Kato, Y. Motome, N. Furukawa, S. Mori, and T. Katsufuji, Phys. Rev. Lett. **96**, 086406 (2006)

[40] A.J. Browne, S.A.J. Kimber, and J.P. Attfield, Phys. Rev. Materials **1**, 052003(R) (2017)

[41] J.P. Wright, J.P. Attfield, and P.G. Radaelli, Phys. Rev. Lett. **87**, 266401 (2001)

[42] M.S. Senn, J.P. Wright, and J.P. Attfield, Nature **481**, 173 (2012)

[43] M. Taguchi, A. Chainani, S. Ueda, M. Matsunami, Y. Ishida, R. Eguchi, S. Tsuda, Y. Takata, M. Yabashi, K. Tamasaku, Y. Nishino, T. Ishikawa, H. Daimon, S. Todo, H. Tanaka, M. Oura, Y. Senba, H. Ohashi, and S. Shin, Phys. Rev. Lett. **115**, 256405 (2015)

[44] R.J. Cava, R.S. Roth, T. Siegrist, B. Hessen, J.J. Krajewski, and W.F. Peck, Jr., J. Solid State Chem. **103**, 359 (1993)

[45] D. Hirai, M. Bremholm, J.M. Allred, J. Krizan, L.M. Schoop, Q. Huang, J. Tao, and R.J. Cava, Phys. Rev. Lett. **110**, 166402 (2013)

[46] Y. Okamoto, H. Amano, N. Katayama, H. Sawa, K. Niki, R. Mitoka, H. Harima, T. Hasegawa, N. Ogita, Y. Tanaka, M. Takigawa, Y. Yokoyama, K. Takehana, Y. Imanaka, Y. Nakamura, H. Kishida, and K. Takenaka, Nat. Commun. **11**, 3144 (2020)

[47] R. Nakamura, D. Takegami, A. Melendez-Sans, L.H. Tjeng, M. Okawa, T. Miyoshino, N.L. Saini, M. Kitamura, D. Shiga, H. Kumigashira, M. Yoshimura, K.-D. Tsuei, Y. Okamoto, and T. Mizokawa, Phys. Rev. B **106**, 195104 (2022)

[48] S.V. Streltsov, I.I. Mazin, R. Heid, and K.-P. Bohnen, Phys. Rev. B **94**, 241101(R) (2016)

[49] H. Nakai and C. Hotta, Nat. Commun. **13**, 579 (2022)

[50] S. Lee, J.-G. Park, D.T. Adroja, D. Khomskii, S. Streltsov, K.A. McEwen, H. Sakai, K. Yoshimura, V.I. Anisimov, D. Mori, R. Kanno, and R. Ibberson, Nat. Mater. **5**, 471 (2006)

[51] K. Takeda, K. Miyake, K. Takeda, and K. Hirakawa, J. Phys. Soc. Jpn. **61**, 2156 (1992)

[52] S.J. Clarke, A.J. Fowkes, A. Harrison, R.M. Ibberson, and M.J. Rosseinsky, Chem. Mater. **10**, 372 (1998)

[53] A. Subedi, Phys. Rev. B **95**, 195149 (2017)

[54] S. Jobic, P. Deniard, R. Brec, J. Rouxel, A. Jouanneaux, and A.N. Fitch, Z. Anorg. Allg. Chem. **598**, 199 (1991)

[55] N. Matsumoto, K. Taniguchi, R. Endoh, H. Takano, and S. Nagata, J. Low Temp. Phys. **117**, 1129 (1999)

[56] S. Pyon, K. Kudo, and M. Nohara, J. Phys. Soc. Jpn. **81**, 053701 (2012)

[57] J.J. Yang, Y.J. Choi, Y.S. Oh, A. Hogan, Y. Horibe, K. Kim, B.I. Min, and S.-W. Cheong, Phys. Rev. Lett. **108**, 116402 (2012)

[58] D. Ootsuki, Y. Wakisaka, S. Pyon, K. Kudo, M. Nohara, M. Arita, H. Anzai, H. Namatame, M. Taniguchi, N.L. Saini, and T. Mizokawa, Phys. Rev. B **86**, 014519 (2012)

[59] W. Tian, M.F. Chisholm, P.G. Khalifah, R. Jin, B.C. Sales, S.E. Nagler, and D. Mandrus, Mater. Res. Bull. **39**, 1319 (2004)

[60] T. Jin-no, Y. Shimizu, M. Itoh, S. Niitaka, and H. Takagi, Phys. Rev. B **87**, 075135 (2013)

[61] K. Kojima, N. Katayama, S. Tamura, M. Shiomi, and H. Sawa,
Phys. Rev. B **100**, 235120 (2019)

[62] H.F. Pen, J. van den Brink, D.I. Khomskii, and G.A. Sawatzky,
Phys. Rev. Lett. **78**, 1323 (1997)

[63] H.F. Pen, L.H. Tjeng, E. Pellegrin, F.M.F. de Groot, G.A. Sawatzky, M.A. van Veenendaal, and C.T. Chen, Phys. Rev. B **55**, 15500 (1997)

[64] S.V. Streltsov and D.I. Khomskii, Phys. Rev. B **89**, 161112(R) (2014)

[65] A.C.W.P. James and J.B. Goodenough, J. Solid State Chem. **76**, 87 (1988)

[66] N. Katayama, M. Uchida, D. Hashizume, S. Niitaka, J. Matsuno, D. Matsumura, Y. Nishihata, J. Mizuki, N. Takeshita, A. Gauzzi, M. Nohara, and H. Takagi,
Phys. Rev. Lett. **103**, 146405 (2009)

[67] T. Tanaka, Y. Kawasaki, S. Endou, S. Kimura, Y. Ideta, Y. Kishimoto, T. Ohno, N. Katayama, M. Nohara, and H. Takagi, J. Phys. Soc. Jpn. **78**, 054709 (2009)

[68] N. Katayama, K. Kojima, T. Yamaguchi, S. Hattori, S. Tamura, K. Ohara, S. Kobayashi, K. Sugimoto, Y. Ohta, K. Saitoh, and H. Sawa, npj Quantum Mater. **6**, 16 (2021)

[69] T. Kajita, T. Kanzaki, T. Suzuki, J.E. Kim, K. Kato, M. Takata, and T. Katsufuji,
Phys. Rev. B **81**, 060405(R) (2010)

[70] K. Takubo, T. Kanzaki, Y. Yamasaki, H. Nakao, Y. Murakami, T. Oguchi, and T. Katsufuji,
Phys. Rev. B **86**, 085141 (2012)

[71] T. Yoshino, M. Okawa, T. Kajita, S. Dash, R. Shimoyama, K. Takahashi, Y. Takahashi, R. Takayanagi, T. Saitoh, D. Ootsuki, T. Yoshida, E. Ikenaga, N.L. Saini, T. Katsufuji, and T. Mizokawa, Phys. Rev. B **95**, 075151 (2017)

[72] G. Jackeli and D.I. Khomskii, Phys. Rev. Lett. **100**, 147203 (2008)

[73] S. Ogawa, J. Phys. Soc. Jpn. **15**, 1901 (1960)

[74] S. Pei, J. Tang, C. Liu, J. Mei, Z. Guo, B. Lyu, N. Zhang, Q. Huang, D. Yu, L. Huang, J. Lin.L. Wang, and M. Huang, Appl. Phys. Lett. **117**, 133103 (2020)

[75] K. Motizuki, S. Miyata, and N. Suzuki, J. Phys. Soc. Jpn. **45**, 1613 (1978)

[76] V.V. Gapontsev, D.D. Gazizova, and S.V. Streltsov,
J. Phys.: Condens. Matter **33**, 495803 (2021)

[77] H. Yamamoto, S. Kamiyama, I. Yamada, and H. Kimura,
J. Am. Chem. Soc. **144**, 1082 (2022)

[78] K. Foyevtsova, H.O. Jeschke, I.I. Mazin, D.I. Khomskii, and R. Valenti,
Phys Rev B **88**, 035107 (2013)

[79] S.V. Streltsov and D.I. Khomskii, PNAS. **113**, 10491 (2016)

[80] Y. Miura, Y. Yasui, M. Sato, N. Igawa, and K. Kakurai,
J. Phys. Soc. Jpn. **76**, 033705 (2007)

[81] S.A.J. Kimber, I.I. Mazin, J. Shen, H.O. Jeschke, S.V. Streltsov, D.N. Argyriou, R. Valenti, and D.I. Khomskii, Phys. Rev. B **89**, 081408(R) (2014)

[82] P. McArdle, F.-T. Huang, J. Yang, M.-W. Chu, S.-W. Cheong, and M.M. Qazilbash, Phys. Rev. B **105**, 245148 (2022)

[83] F. Poineau, E.V. Johnstone, P.F. Weck, P.M. Forster, E. Kim, K.R. Czerwinski, and A.P. Sattelberger, Inorg. Chem. **51**, 4915 (2012)

[84] H. Schäfer, H.G. von Schnering, J.V. Tillack, F. Kuhnen, H. Wörle, and H. Baumann, Z. Anorg. Allgem. Chem. **353**, 281 (1967)

[85] H. Hillebrecht, P.J. Schmidt, H.W. Rotter, G. Thiele, P. Zönnchen, H. Bengel, H.J. Cantow, S.N. Magonov, and M.H. Whangbo, J. Alloys. Compd. **246**, 70 (1997)

[86] M.A. McGuire, J. Yan, P. Lampen-Kelley, A.F. May, V.R. Cooper, L. Lindsay, A. Puretzky, L. Liang, Santosh KC, E. Cakmak, S. Calder, and B.C. Sales,
Phys. Rev. Materials **1**, 064001 (2017)

[87] D.J. Hinz, G. Meyer, T. Dedecke, and W. Urland,
Angew. Chem., Int. Ed. Engl. **34**, 71 (1995)

[88] N. Hanni, M.D. Frontzek, J. Hauser, D. Cheptiakov, and K. Kramer,
Z. Anorg. Allg. Chem. **643**, 2063 (2017)

[89] Z.A. Kelly, T.T. Tran, and T.M. McQueen, Inorg. Chem. **58**, 11941 (2019)

[90] A. Paul, C.M. Chung, T. Birol, and H.J. Changlani, Phys. Rev. Lett. **124**, 167203 (2020)

[91] D.I. Khomskii, T. Mizokawa, and S.V. Streltsov, Phys. Rev. Lett. **127**, 049701 (2021)

[92] J.P. Sheckelton, J.R. Neilson, D.G. Soltan, and T.M. McQueen, Nat. Mater. **11**, 493 (2012)

[93] M. Mourigal, W.T. Fuhrman, J.P. Sheckelton, A. Wartelle, J.A. Rodriguez-Rivera, D.L. Abernathy, T.M. McQueen, and C.L. Broholm, Phys. Rev. Lett. **112**, 027202 (2014)

[94] G. Chen, H.-Y. Kee, and Y.B. Kim, Phys. Rev. B **93**, 245134 (2016)

[95] S.A. Nikolaev, I.V. Solovyev, and S.V. Streltsov, npj Quantum Materials **6**, 25 (2021)

[96] R. Nakamura *at al*., unpublished.

[97] F.A. Cotton, M.P. Diebold, and W.J. Roth, J. Am. Chem. Soc. **109**, 2833 (1987)

[98] M. Ströbele, J. Glaser, A. Lachgar, and H.-J. Meyer,
Z. Anorg. Allg. Chem. **627**, 2002 (2001)

[99] Y. Haraguchi, C. Michioka, M. Ishikawa, Y. Nakano, H. Yamochi, H. Ueda, and K. Yoshimura, Inorg. Chem. **56**, 3483 (2017)

[100] Y. Zhang, Y. Gu, H. Weng, K. Jiang, and J. Hu, Phys. Rev. B **107**, 035126 (2023)

4 Multiplets in Transition Metal Ions and Introduction to Multiband Hubbard Models

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

Contents

1 Introduction **2**

2 Multiplets of a free ion **2**
2.1 General considerations 2
2.2 Matrix elements of the Coulomb interaction 5
2.3 Diagonal matrix elements 7
2.4 Analytical calculation of multiplet energies by the diagonal sum-rule 9
2.5 Solution of the Coulomb problem by exact diagonalization 11
2.6 Spin-orbit coupling 13

3 Effects of the environment in the crystal **14**
3.1 Crystalline electric field 14
3.2 Charge transfer 19

4 Multiband Hubbard models **23**

5 Conclusion **25**

A Gaunt coefficients **26**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

Compounds containing $3d$ or $4f$ transition-metal or rare-earth ions have been intriguing solid state physicists ever since the appearance of solid state physics as a field of research. In fact, already in the 1930's NiO became the first known example of a correlated insulator in that it was cited by deBoer and Verwey as a counterexample to the then newly invented Bloch theory of electron bands in solids [1]. During the last 25 years $3d$ and $4f$ compounds have become one of the central fields of solid state physics following the discovery of heavy fermion compounds, cuprate superconductors, the colossal magnetoresistance phenomenon in the manganites and, most recently, the iron-pnictide superconductors.
It was conjectured early on that the reason for the special behavior of these compounds is the strong Coulomb interaction between electrons in their partially filled $3d$ or $4f$ shells. The $3d$ wave functions are orthogonal to those of the inner-shells, such as $1s$, $2s$, $2p$, $3s$ and $3p$, solely due to their angular part $Y_{2,m}(\vartheta, \varphi)$. Their radial part $R_{3,2}(r)$ therefore is not pushed out to regions far from the nucleus by the requirement to be orthogonal to the inner shell wave functions and therefore is concentrated close to the nucleus (the situation is exactly the same for the $4f$ wave functions). Any two electrons in the $3d$ shell thus are forced to be close to each other on average so that their mutual Coulomb repulsion is strong (the Coulomb repulsion between two $3d$ electrons is small, however, when compared to the Coulomb force due to the nucleus and the inner shells so that the electrons *have to* stay close to one another!). For clarity let us mention that the Coulomb repulsion between electrons in the inner shells of heavier elements is actually much stronger than that in the $3d$ shell of transition metals or the $4f$ shell of rare earths. This, however, is irrelevant because these inner shells are several $100-1000$ eV below the Fermi energy so that they are simply completely filled and inert. On the other hand, the $3d$ orbitals in transition metal compounds and the $4f$ orbitals in rare earth compounds participate in the bands at the Fermi level so that the strong Coulomb interaction in these orbitals directly influences the conduction electrons. The conduction bands in such compounds therefore form dense many-body-systems of strongly interacting electrons, where the average energy of interaction is large compared to the average kinetic energy. This dominance of the interaction energy implies a propensity to show ordering phenomena and the ensuing quantum phase transitions and superconducting domes. It is therefore ultimately the Coulomb repulsion in the partially filled $3d$ shells of the transition metals and the $4f$ shells of the rare earths which gives rise to the wide variety of spectacular phenomena observed in compounds containing these elements. Let us therefore discuss this Coulomb interaction in more detail.

2 Multiplets of a free ion

2.1 General considerations

In the following we restrict ourselves to $3d$ transition metal ions for definiteness, but the theory is easily adapted to other atomic shells. We consider a Ni^{2+} ion in vacuum which has the

Term	J	E (eV)
^3F	4	0.000
	3	0.169
	2	0.281
^1D	2	1.740
^3P	2	2.066
	1	2.105
	0	2.137
^1G	4	2.865
^1S	0	6.514

Table 1: *Energies of the multiplets of Ni^{2+} from Ref. [2]. J is the total angular momentum quantum number and the $J = 4$ member of ^3F has been taken as the zero of energy.*

electron configuration [Ar] $3d^8$. It is a standard exercise in textbooks of atomic physics to show that the d^8 configuration has the following multiplets or terms: ^3F, ^3P, ^1G, ^1D and ^1S, whereby according to the first two Hund's rules ^3F is the ground state. 'Multiplets' thereby is simply another word for 'eigenstates of 8 electrons in the electric field of the Ni nucleus and the Ar core' (the electrons in the shells below $3d$ may be considered as inert due to the large binding energies of these shells). The energies of the multiplets can be deduced experimentally for example by analyzing the optical spectrum of Ni vapor and are listed in Table 1. They span a range of several eV whereby multiplets with nonzero spin are in addition split by spin-orbit coupling which results in intervals of order 0.1 eV. All of these eigenstates correspond to the same electron configuration, namely [Ar] $3d^8$, so that the fact that, say, ^3P has a higher energy than ^3F is not due to an electron having been promoted from a state with low energy to one with high energy as in an optical transition. Rather, the excited multiplets — ^3P, ^1G, ^1D and ^1S — should be viewed as *collective excitations* of the 8-electron system, similar in nature to a plasmon in an electron gas. And just as a plasmon can exist only due to the Coulomb interaction between electrons, the multiplet splitting in atomic shells also originates from the Coulomb interaction between electrons. This is what we discuss next.

As a first step we introduce Fermionic creation and annihilation operators $c^\dagger_{n,l,m,\sigma}$ which create an electron with z-component of spin σ in the orbital with principal quantum number n, orbital angular momentum l, and z-component of orbital angular momentum m. In the case of a partly filled $3d$ shell all $n_i = 3$ and all $l_i = 2$ identically, so that these two indices could be omitted, but we will keep them for the sake of generality. In the following we will often contract (n, l, m, σ) to the 'compound index' ν for brevity, so that, e.g., $c^\dagger_{\nu_i} = c^\dagger_{n_i,l_i,m_i,\sigma_i}$.

The procedure we follow is degenerate first-order perturbation theory as discussed in practically any textbook of quantum mechanics. The unperturbed Hamiltonian H_0 thereby corresponds to the energies of the different atomic shells

$$H_0 = \sum_{n,l} \varepsilon_{n,l} \sum_{m,\sigma} c^\dagger_{n,l,m,\sigma} c_{n,l,m,\sigma}$$

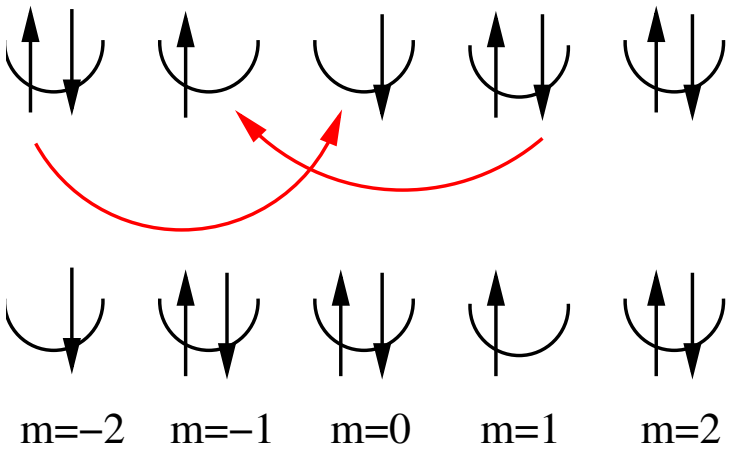


Fig. 1: *Coulomb scattering of two electrons in the d-shell. In the initial state $|\nu\rangle$ (top) the electrons are distributed over the five d-orbitals which are labeled by their m-values. Due to their Coulomb interaction two electrons scatter from each other and are simultaneously transferred to different orbitals, resulting in the state $|\mu\rangle$ (bottom).*

whereas the Coulomb interaction is considered as the perturbation H_1 (we ignore spin-orbit coupling for the time being). The d^n configuration comprises all states which are obtained by distributing n electrons over the $2 \cdot 5 = 10$ spin-orbitals:

$$|\nu\rangle = |\nu_1, \nu_2 \ldots \nu_n\rangle = c^\dagger_{\nu_1} c^\dagger_{\nu_2} \ldots c^\dagger_{\nu_n} |0\rangle, \tag{1}$$

and the number of these states obviously is $n_c = 10!/(n!\,(10-n)!)$. In writing the basis states as in (1) we need to specify an ordering convention for the creation operators on the right hand side. For example, only states are taken into account where $m_1 \leq m_2 \leq m_3 \cdots \leq m_n$. Moreover, if two m_i are equal the $c^\dagger_{m_i,\downarrow}$-operator is assumed to be to the left of the $c^\dagger_{m_i,\uparrow}$-operator. If we adopt this convention, every possible state obtained by distributing the n electrons over the 10 spin-orbitals is included exactly once in the basis. If the n_i and l_i were to take different values we could generalize this, e.g., by demanding that the (n_i, l_i, m_i)-triples be ordered lexicographically. As will be seen later, strict application of an ordering convention for the Fermi operators is necessary to determine the correct Fermi signs for the matrix elements.

If only H_0 were present all states (1) would be degenerate with energy $E = E[Ar] + n \cdot \varepsilon_{3,2}$, where $E[Ar]$ is the energy of the Argon core. The Coulomb interaction H_1 between the electrons (partially) lifts this degeneracy and this is the physical reason for the multiplet splitting. The standard procedure in degenerate first order perturbation theory is to set up the secular matrix $h_{\mu,\nu} = \langle\mu|H_1|\nu\rangle$ and diagonalize it to obtain the first order energies and wave functions [3]. The diagonal matrix elements $\langle\nu|H_1|\nu\rangle$ describe the fact that the Coulomb repulsion between two electrons in different orbitals depends on the spatial character of these orbitals, whereas the off-diagonal matrix elements $\langle\mu|H_1|\nu\rangle$ describe the scattering of two electrons 'within the $3d$ shell' as shown in Figure 1.

In second quantization the Coulomb Hamiltonian H_1 takes the form

$$\begin{aligned} H_1 &= \frac{1}{2} \sum_{\nu_1,\nu_2,\nu_3,\nu_4} V(\nu_1,\nu_2,\nu_3,\nu_4)\; c^\dagger_{\nu_1} c^\dagger_{\nu_2} c_{\nu_3} c_{\nu_4}, \\ V(\nu_1,\nu_2,\nu_3,\nu_4) &= \int dx \int dx'\; \psi^*_{\nu_1}(x)\, \psi^*_{\nu_2}(x')\; V_c(x,x')\; \psi_{\nu_4}(x)\, \psi_{\nu_3}(x'), \\ V_c(x,x') &= \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}. \end{aligned} \tag{2}$$

Here $x = (\boldsymbol{r}, \sigma)$ is the combined position and spin coordinate with $\int dx \cdots = \sum_\sigma \int d\boldsymbol{r} \ldots$ and V_c is the Coulomb interaction between electrons. Note the factor of $1/2$ in front of H_1 and the correspondence of indices and integration variables $\nu_4 \leftrightarrow x$ and $\nu_3 \leftrightarrow x'$ in the Coulomb matrix element, see textbooks of many-particle physics such as Fetter-Walecka [6].

2.2 Matrix elements of the Coulomb interaction

Our single-particle basis consists of atomic spin-orbitals so if we switch to spherical coordinates (r, ϑ, φ) for $\boldsymbol{r}$ the wave functions in (2) are

$$\psi_{\nu_i}(x) = R_{n_i,l_i}(r)\; Y_{l_i,m_i}(\vartheta,\varphi)\; \delta_{\sigma,\sigma_i}. \tag{3}$$

For a table of spherical harmonics $Y_{l,m}$ see Ref. [4]. The radial wave functions R_{n_i,l_i} are assumed to be real — as is the case for the true radial wave function of bound states in a central potential. Apart from that we do not really specify them. It will turn out that these radial wave functions enter the Coulomb matrix elements only via a discrete and rather limited set of real numbers which are often obtained by a fit to experiment.
In addition to (3), we use the familiar multipole expansion of the Coulomb interaction [5]

$$\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} Y^*_{k,m}(\vartheta',\varphi')\; \frac{4\pi}{2k+1}\; \frac{r_<^k}{r_>^{k+1}}\; Y_{k,m}(\vartheta,\varphi). \tag{4}$$

We now insert (3) and (4) into (2). We recall that $\int dx \cdots = \sum_\sigma \int d\boldsymbol{r} \ldots$ and first carry out the sums over spin variables:

$$\sum_{\sigma,\sigma'} \delta_{\sigma,\sigma_1}\delta_{\sigma',\sigma_2}\delta_{\sigma,\sigma_4}\delta_{\sigma',\sigma_3} \;\; = \;\; \delta_{\sigma_1,\sigma_4}\; \delta_{\sigma_2,\sigma_3}.$$

This reflects the fact that since the Coulomb interaction does not depend on spin, the spins of the two electrons are conserved in the Coulomb scattering. Next, we pick one term with given k and m from the multipole expansion (4) and proceed to the integration over the spatial variables (r, ϑ, φ) and $(r', \vartheta', \varphi')$. Let us first consider (ϑ, φ) and adopt the compact notation $(\vartheta, \varphi) = \Omega$. These variables always come as arguments of spherical harmonics and there is one from $\psi^*_{\nu_1}(x)$, one from the multipole expansion (4), and one from $\psi_{\nu_4}(x)$. We obtain the integral

$$\int d\Omega\; Y^*_{l_1,m_1}(\Omega)\, Y_{k,m}(\Omega)\, Y_{l_4,m_4}(\Omega), \tag{5}$$

where $\int d\Omega \cdots = \int_0^{2\pi} d\varphi \int_{-1}^{1} dcos(\vartheta) \ldots$. Such a dimensionless integral over three spherical harmonics is called a Gaunt coefficient and it follows from the Wigner-Eckart theorem that it is proportional to a Clebsch-Gordan coefficient [7, 8].
Next we recall $Y_{l,m}(\vartheta,\varphi) = P_{l,m}(\vartheta)\, e^{im\varphi}$ [3] whence the integral (5) is proportional to

$$\int_0^{2\pi} d\varphi\; e^{-i(m_1-m-m_4)\varphi} = \delta_{m,m_1-m_4}.$$

We introduce the following notation for nonvanishing Gaunt coefficients

$$c^k(lm;l'm') = \sqrt{\frac{4\pi}{2k+1}} \int d\Omega\; Y^*_{l,m}(\Omega)\; Y_{k,m-m'}(\Omega)\; Y_{l',m'}(\Omega),$$

where we have also included 'half of the factor $\frac{4\pi}{2k+1}$' from (4). Then, (5) becomes

$$\sqrt{\frac{4\pi}{2k+1}} \int d\Omega\; Y^*_{l_1,m_1}(\Omega)\, Y_{k,m}(\Omega)\, Y_{l_4,m_4}(\Omega) = \delta_{m,m_1-m_4}\; c^k(l_1m_1;l_4m_4). \tag{6}$$

Since the remaining ϑ-dependent factors $P_{l,m}(\vartheta)$ are real [3] it follows that all Gaunt coefficients are real as well. Using this property the integral over (ϑ',φ') becomes

$$\sqrt{\frac{4\pi}{2k+1}} \int d\Omega'\; Y^*_{l_2,m_2}(\Omega')\, Y^*_{k,m}(\Omega')\, Y_{l_3,m_3}(\Omega') = \delta_{m,m_3-m_2}\; c^k(l_3m_3;l_2m_2). \tag{7}$$

Since both (6) *and* (7) must be different from zero for the *same* m in order to obtain a nonvanishing contribution, we must have $m_1-m_4 = m_3-m_2$ or $m_1+m_2 = m_3+m_4$, i.e., the total L^z is conserved in the scattering process. This could have been expected from the very beginning and our formalism incorporates this.
It remains to do the integral over the two radial variables r and r'. These two integrations cannot be disentangled so we find a factor of

$$R^k(n_1l_1,n_2l_2,n_3l_3,n_4l_4) = \int_0^\infty dr\, r^2 \int_0^\infty dr'\, r'^2\; R_{n_1,l_1}(r)R_{n_2,l_2}(r')\; \frac{r_<^k}{r_>^{k+1}}\; R_{n_4,l_4}(r)R_{n_3,l_3}(r'). \tag{8}$$

These integrals have the same dimension as V_c, i.e., energy. Collecting everything we find

$$\begin{aligned} V(\nu_1,\nu_2,\nu_3,\nu_4) = \sum_{k=0}^{\infty} & c^k(l_1m_1;l_4m_4)\, c^k(l_3m_3;l_2m_2)\, R^k(n_1l_1,n_2l_2,n_3l_3,n_4l_4) \\ & \times \delta_{\sigma_1,\sigma_4}\; \delta_{\sigma_2,\sigma_3}\; \delta_{m_1+m_2,m_3+m_4}\,. \end{aligned} \tag{9}$$

The number of relevant multipole orders k in this sum is severely limited by the properties of the Gaunt coefficients $c^k(lm;l'm')$. First, since these are proportional to Clebsch-Gordan coefficients the three l-values appearing in them have to obey the so-called *triangular condition* [3] $k \leq \min(l,l')$ whence $k \leq \min(l_1+l_4, l_2+l_3)$. For Coulomb scattering in a d shell all $l_i = 2$ whence $k \leq 4$. Second, the parity of the spherical harmonic Y_{lm} is $(-1)^l$. For Coulomb scattering within a given atomic shell all l_i are equal and for integrals such as (5) or (7) to be different from zero the spherical harmonic $Y_{k,m}$ from the multipole expansion must have

positive parity whence k must be even. For Coulomb scattering within a d shell therefore only R^0, R^2 and R^4 are relevant. This shows that the sloppy definition of the radial wave function $R_{n_i,l_i}(r)$ is not a real problem because details of this wave function are irrelevant anyway. In a way, these three parameters may be viewed as a generalization of the Hubbard-U in that R^k is something like the 'the Hubbard-U for k-pole interaction'. Lastly we note that the $c^k(lm;l'm')$ are tabulated in Appendix 20a of the textbook by Slater [7] or Table 4.4 of the textbook by Griffith [8], and also in the Appendices I and II of the present note.

2.3 Diagonal matrix elements

The expression (9) is exact but somewhat complicated so let us try to elucidate its physical content and thereby also make contact with various approximate ways to describe the Coulomb interaction which can be found in the literature. We recall

$$H_1 = \frac{1}{2} \sum_{\nu_1,\nu_2,\nu_3,\nu_4} V(\nu_1,\nu_2,\nu_3,\nu_4)\; c^\dagger_{\nu_1} c^\dagger_{\nu_2} c_{\nu_3} c_{\nu_4},$$

and pick those terms from H_1 where either $\nu_4 = \nu_1$ and $\nu_3 = \nu_2$ (case 1) or $\nu_3 = \nu_1$ and $\nu_4 = \nu_2$ (case 2). Notice that the Pauli principle requires $\nu_1 \neq \nu_2$ — otherwise H_1 contains the product $c^\dagger_{\nu_1} c^\dagger_{\nu_1} = 0$. In both cases the four Fermion operators can be permuted to give the product of number operators $n_{\nu_1} n_{\nu_2}$ (with $n_\nu = c^\dagger_\nu c_\nu$) whereby in case 2 an odd number of interchanges of Fermion operators is necessary so that an additional factor of (-1) appears. Since $\nu_1 \neq \nu_2$ no nonvanishing anticommutators arise in this permutation of operators. Whereas for case 1 the product $\delta_{\sigma_1,\sigma_4}\; \delta_{\sigma_2,\sigma_3}$ in (9) always is 1, it vanishes for case 2 unless $\sigma_1{=}\sigma_2$. We had $\nu_1 \neq \nu_2$ so that for case 1 the two orbitals may have the same orbital quantum numbers n, l, m but then must differ in their spin, whereas in case 2 the spins have to be equal so that the orbital quantum numbers definitely must be different. Using (9) the respective matrix elements are

$$\begin{aligned} V(\nu_1,\nu_2,\nu_2,\nu_1) &= \sum_{k=0}^{\infty} c^k(l_1m_1;l_1,m_1)\, c^k(l_2m_2;l_2,m_2) R^k(n_1l_1,n_2l_2,n_2l_2,n_1l_1), \\ V(\nu_1,\nu_2,\nu_1,\nu_2) &= \delta_{\sigma_1,\sigma_2} \sum_{k=0}^{\infty} c^k(l_1m_1;l_2,m_2)\, c^k(l_1m_1;l_2,m_2) R^k(n_1l_1,n_2l_2,n_1l_1,n_2l_2). \end{aligned} \quad (10)$$

It is customary to introduce the abbreviations

$$\begin{aligned} a^k(lm;l'm') &= c^k(lm;lm)\;\; c^k(l'm';l'm') \\ b^k(lm;l'm') &= c^k(lm;l'm')\; c^k(lm;l'm') \\ F^k(nl;n'l') &= R^k(nl,n'l',n'l',nl) \\ G^k(nl;n'l') &= R^k(nl,n'l',nl,n'l') \end{aligned} \quad (11)$$

The F^k and G^k are called Slater-Condon parameters. The a^k and b^k are listed in Appendix 20a of Slater's textbook [7] and also in the Appendix of the present note.

We want to bring these diagonal matrix elements to a more familiar form and continue to specialize to a partly filled $3d$ shell. In this case all $n_i = 3$ and $l_i = 2$ so that for each k there

is only one F^k and one G^k and, in fact, $G^k = F^k$. For brevity we omit the n- and l quantum numbers in the rest of the paragraph so that, e.g., the electron operators become $c^\dagger_{m,\sigma}$ where m is the z-component of $\boldsymbol{L}$. The sum of all diagonal matrix elements then becomes

$$H_{1,diag} = \sum_m U_{m,m}\, n_{m,\uparrow} n_{m,\downarrow} + \frac{1}{2} \sum_{m \neq m'} \left(U_{m,m'} \sum_{\sigma,\sigma'} n_{m,\sigma} n_{m',\sigma'} - J_{m,m'} \sum_{\sigma} n_{m,\sigma} n_{m',\sigma} \right),$$

$$U_{m,m'} = \sum_{k \in \{0,2,4\}} a^k(m,m')\, F^k, \quad J_{m,m'} = \sum_{k \in \{0,2,4\}} b^k(m,m')\, F^k. \tag{12}$$

The first term on the r.h.s. originates from case 1 with $m_1 = m_2$ and the factor of $\frac{1}{2}$ in front of this term is cancelled because there are two identical terms of this type with either $\nu_1 = (m,\uparrow)$ and $\nu_2 = (m,\downarrow)$ or $\nu_1 = (m,\downarrow)$ and $\nu_2 = (m,\uparrow)$. We introduce the operators of electron density $n_m = n_{m,\uparrow} + n_{m,\downarrow}$ and electron spin $S^z_m = \frac{1}{2}(n_{m,\uparrow} - n_{m,\downarrow})$ and rewrite

$$\sum_{\sigma,\sigma'} n_{m,\sigma}\, n_{m',\sigma'} = n_m\, n_{m'} \qquad \sum_{\sigma} n_{m,\sigma}\, n_{m',\sigma} = 2 \left(S^z_m\, S^z_{m'} + \frac{n_m n_{m'}}{4} \right),$$

so that

$$H_{1,diag} = \sum_m U_{m,m}\, n_{m,\uparrow}\, n_{m,\downarrow} + \frac{1}{2} \sum_{m \neq m'} \left(\big(U_{m,m'} - \frac{1}{2}\, J_{m,m'}\big) n_m n_{m'} - 2 J_{m,m'}\, S^z_m S^z_{m'} \right). \tag{13}$$

This is the sum of a density-density interaction $\propto U_{m,m'}$ and an Ising-like spin interaction $\propto J_{m,m'}$. The interaction parameters depend on the orbitals and can be expressed in terms of the Slater-Condon parameters F^k and the products of Gaunt coefficients a^k and b^k. It is obvious from (11) and (12) that $J_{m,m'} > 0$ so that the spin interaction is ferromagnetic — this is in fact the physical origin of the first Hund's rule.

To complete the Hund's rule term we pick those terms in H_1 where $\nu_1 = (m,\sigma)$, $\nu_2 = (m',\bar\sigma)$, $\nu_3 = (m,\bar\sigma)$ and $\nu_4 = (m',\sigma)$. In these terms the product $\delta_{\sigma_1,\sigma_4}\delta_{\sigma_2,\sigma_3}$ is non-vanishing as well and for both values of σ the matrix element (2) is

$$\sum_{k \in \{0,2,4\}} c^k(m,m')\, c^k(m,m') F^k = \sum_{k \in \{0,2,4\}} b^k(m,m') F^k = J_{m,m'}$$

The Fermion operators are $c^\dagger_{m,\uparrow} c^\dagger_{m',\downarrow} c_{m,\downarrow} c_{m',\uparrow} + c^\dagger_{m,\downarrow} c^\dagger_{m',\uparrow} c_{m,\uparrow} c_{m',\downarrow} = -(S^+_m S^-_{m'} + S^-_m S^+_{m'})$, i.e., the transverse part of the Heisenberg exchange. Combining these terms with the Ising-like spin exchange term we obtain

$$H_{1,H} = \sum_m U_{m,m}\, n_{m,\uparrow} n_{m,\downarrow} + \frac{1}{2} \sum_{m \neq m'} \left(\big(U_{m,m'} - \frac{1}{2} J_{m,m'}\big) n_m n_{m'} - 2 J_{m,m'}\, \boldsymbol{S}_m \cdot \boldsymbol{S}_{m'} \right). \tag{14}$$

This is now the sum of a density-density interaction and a spin-rotation invariant ferromagnetic spin exchange. It has to be kept in mind that this Hamiltonian has been obtained by retaining only a relatively small subset of matrix elements in the original Coulomb Hamiltonian. A

further simplification which is often used is to replace $U_{m,m'}$ and $J_{m,m'}$ by their averages over all corresponding pairs (m, m'). Using the a^k and b^k in the Appendix one readily obtains

$$U = \frac{1}{25}\sum_{m,m'} U_{m,m'} \qquad\qquad = F^0,$$
$$U{-}J = \frac{1}{20}\sum_{m\neq m'} \big(U_{m,m'} - J_{m,m'}\big) = F^0 - \frac{1}{14}\big(F^2{+}F^4\big),$$

so that $J = (F^2{+}F^4)/14$.

To conclude the discussion we consider the diagonal matrix elements $\langle\nu|H_1|\nu\rangle$ in the basis of n-electron states $|\nu\rangle$ defined in (1). Since ν_1 and ν_2 in (10) can be any two out of the n occupied orbitals in $|\nu\rangle$ the total diagonal matrix element of H_1 is obtained by summing over all $n(n{-}1)/2$ pairs (i, j) formed from the occupied orbitals

$$\langle\nu|H_1|\nu\rangle = \sum_{i<j}\sum_{k}\Big(a^k(l_i m_i, l_j, m_j)F^k(n_i l_i, n_j l_j) - \delta_{\sigma_i\sigma_j} b^k(l_i m_i, l_j, m_j)G^k(n_i l_i, n_j l_j)\Big). \quad (15)$$

As will be seen in the next paragraph, this formula is actually sufficient to calculate the multiplet energies.

2.4 Analytical calculation of multiplet energies by the diagonal sum-rule

We now show that the theory developed so far is in fact sufficient to give analytical formulas for the energies of the multiplets which can be compared to experiment. The first ingredient is the so-called diagonal sum-rule. This is simply the well-known theorem that the sum of the eigenvalues of a Hermitean matrix H is equal to its trace $\mathrm{tr}(H) = \sum_i H_{ii}$. It follows immediately by noting that the trace of a matrix is invariant under basis transformations, i.e., $\mathrm{tr}(H) = \mathrm{tr}(UHU^{-1})$ for any unitary matrix U. By choosing U to be the matrix which transforms to the basis of eigenvectors of H the diagonal sum-rule follows immediately.

Next, one uses the fact that the Hamilton matrix is block-diagonal, with blocks defined by their values of L^z and S^z — this is the consequence of the δ-functions in (9). The diagonal sum-rule then can be applied separately for each of these blocks. In addition, the dimension of the blocks decreases as L^z and S^z approach their maximum possible values so that the number of multiplets contained in a given block decreases and the multiplet energies are easy to read off.

As an example for the procedure let us consider a p^2 configuration (by particle-hole symmetry this is equivalent to a p^4 configuration). We write the Fermion operators in the form $c^\dagger_{l,m,\sigma}$, i.e., we suppress the principal quantum number n. Since we have 6 possible states for a single p-electron - three m-values and two spin directions per m-value — we have 15 states for two electrons. The triangular condition for the Gaunt coefficients now restricts the multipole order k to be ≤ 2. Again, only even k contribute, so that we have two Slater-Condon parameters, F^0 and F^2 (and $G^k = F^k$). Table 2, which is taken from Slater's textbook [7], gives the values of the coefficients $a^k(1, m; 1, m')$ and $b^k(1, m; 1, m')$.

m	m'	a^0	$25a^2$	b^0	$25b^2$
±1	±1	1	1	1	1
±1	0	1	-2	0	3
0	0	1	4	1	4
±1	∓1	1	1	0	6

Table 2: *The coefficients a^k and b^k for two p-electrons.*

We first consider the sector with $S^z = 1$. The highest possible L^z is $L^z = 1$ which is realized only for a single state, $|1\rangle = c^\dagger_{1,0,\uparrow} c^\dagger_{1,1,\uparrow}|0\rangle$. We can conclude that one of the multiplets is ${}^3P$ and its energy is equal to the diagonal matrix element of $|1\rangle$ which by (15) is

$$E({}^3P) = \sum_{k\in\{0,2\}} \Big(a^k(1,1;1,0) - b^k(1,1;1,0)\Big) F^k = F^0 - \frac{5}{25}F^2.$$

We proceed to the sector $S^z = 0$. Here the highest possible L^z is $L^z = 2$ again obtained for only single state namely $c^\dagger_{1,1,\downarrow} c^\dagger_{1,1,\uparrow}|0\rangle$. We conclude that we also have ${}^1D$ with energy

$$E({}^1D) = \sum_{k\in\{0,2\}} a^k(1,1;1,1)\, F^k = F^0 + \frac{1}{25}F^2.$$

The two multiplets that we found so far, ${}^1D$ and ${}^3P$, comprise $5 + 9 = 14$ states; we thus have just one state missing, which can only be ${}^1S$. To find its energy, we need to consider the sector $S^z = 0$ and $L^z = 0$. There are three states in this sector: $c^\dagger_{1,0,\downarrow} c^\dagger_{1,0,\uparrow}|0\rangle$, $c^\dagger_{1,-1,\uparrow} c^\dagger_{1,1,\downarrow}|0\rangle$ and $c^\dagger_{1,-1,\downarrow} c^\dagger_{1,1,\uparrow}|0\rangle$. Two out of the three eigenvalues of the 3×3 Hamiltonian in the basis spanned by these states must be $E({}^3P)$ and $E({}^1D)$, because these multiplets also have members with $S^z = 0$ and $L^z = 0$. To obtain $E({}^1S)$ we accordingly compute the sum of the diagonal elements of the 3×3 matrix using (15) and set

$$E({}^3P) + E({}^1D) + E({}^1S) = \sum_{k\in\{0,2\}} \Big(a^k(1,0;1,0) + 2\; a^k(1,-1;1,1)\Big) F^k,$$

$$\rightarrow\; E({}^1S) = F^0 + \frac{10}{25}F^2.$$

This example shows the way of approach for multiplet calculations using the diagonal sum-rule: one starts out with a state with maximum L^z or S^z for which there is usually only a single basis state. This basis state belongs to some multiplet whose energy simply equals the 'diagonal element' of the 1×1 Hamiltonian. Then one proceeds to lower S^z and/or L^z and obtains energies of additional multiplets by calculating the trace of the respective block of the Hamilton matrix and using the known energies of multiplets with higher L^z or S^z. It turns out that in this way the energies of *all* multiplets involving s, p, d or f electrons can be expressed in terms of the Slater-Condon parameters by analytical formulas. A rather complete list can be found for example in the Appendices 21a and 21 of the textbook by Slater [7].

Multiplet theory was originally developed to discuss the spectra of atoms or ions in the gas phase. The question then arises, as to what are the values of the Slater-Condon parameters.

	Si	P^+	S^{2+}	S	Cl^+
^3P	0.0000	0.0000	0.0000	0.0000	0.0000
^1D	0.7809	1.1013	1.4038	1.1454	1.4449
^1S	1.9087	2.6750	3.3675	2.7500	3.4564
r	1.4442	1.4289	1.3988	1.4010	1.3921

Table 3: *Energies (in eV) of multiplets for different atoms and ions with p^2 or p^4 configurations outside a closed shell (taken from the NIST data base [2]) and the resulting values of r in (16).*

Of course one might attempt to compute these parameters using, e.g., Hartree-Fock wave functions in the expression (8). It turns out, however, that very frequently the number of multiplets considerably exceeds the number of relevant Slater-Condon parameters. In the case of the p^2 configuration we had three multiplets, ^3P, ^1D and ^1S, but only two Slater-Condon parameters F^0 and F^2. This would suggest to obtain the values of the Slater-Condon parameters by fit to the spectroscopic data and the textbook by Slater [7] contains a vast amount of experimental data which are analyzed in this way. For the p^2 configuration we restrict ourselves to a simple cross check. Using the above expressions we find

$$r = \frac{E(^1S) - E(^1D)}{E(^1D) - E(^3P)} = \frac{3}{2}, \tag{16}$$

independently of the values of F^0 and F^2. This relation therefore should be obeyed by all ions with two p-electrons outside filled shells, such as the series Si, P^{1+}, S^{2+}, or two holes in a filled p-shell such as the series S, Cl^+. The energies of the multiplets of these atoms/ions are available in the database [2] and Table 3 shows the energies and the resulting values of r.
They are in fact quite close $3/2$. Notice that the width of the multiplet spectrum increases considerable when going to the positively charge ions. This is because in positively charged ions the radial wave functions are more contracted, whence the values of the Slater-Condon parameters increase. Despite this, the ratio r is quite constant and in good agreement with multiplet theory.

2.5 Solution of the Coulomb problem by exact diagonalization

Using the diagonal sum rule one can derive analytical formulae for the energies of the multiplets. For further applications of multiplet theory, however, it is often useful to solve the problem numerically, using the method of exact diagonalization which will be outlined in the following. The basis states (1) correspond to all possible ways of distributing n electrons over the 10 spin-orbitals of the $3d$-shell (two spin directions for each $m \in \{-2, -1, \ldots, 2\}$). As illustrated in Figure 2 we can code each of these basis states by an integer $0 \leq i \leq 2^{10}$. If we really use all of these integers we are actually treating all states with $0 \leq n \leq 10$ simultaneously but this will be convenient for generalizations of the theory. Next, for a given initial state $|\nu_1, \nu_2, \ldots \nu_n\rangle$ we can let the computer search for all possible transitions of the type shown in Figure 2 and compute the corresponding matrix elements from (9) using, say, the $c^k(lm; l'm')$ copied from

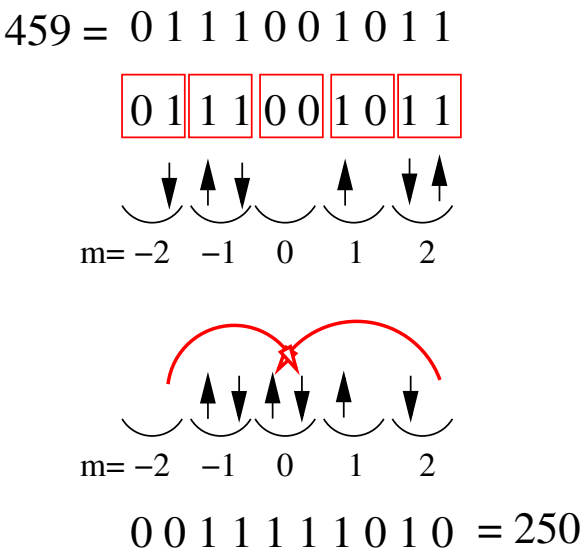


Fig. 2: *The coding of basis states by integers and a scattering process.*

Slater's textbook and some given R^0, R^2 and R^4. Let us consider the following matrix element of a term in H_1 between two states with n electrons:

$$\langle\mu|\ V(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\ c^\dagger_{\lambda_1}c^\dagger_{\lambda_2}c_{\lambda_3}c_{\lambda_4}\ |\nu\rangle =$$

$$\langle 0|c_{\mu_n}\dots c_{\mu_1}\ V(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\ c^\dagger_{\lambda_1}c^\dagger_{\lambda_2}c_{\lambda_3}c_{\lambda_4}\ c^\dagger_{\nu_1}c^\dagger_{\nu_2}\dots c^\dagger_{\nu_n}|0\rangle.$$

For this to be nonzero, the operators $c^\dagger_{\lambda_3}$ and $c^\dagger_{\lambda_4}$ must be amongst the $c^\dagger_{\nu_i}$, otherwise the annihilation operators in the Hamiltonian, c_{λ_3} and c_{λ_4}, could be commuted to the right where they annihilate $|0\rangle$. In order for c_{λ_4} to 'cancel' $c^\dagger_{\lambda_4}$ it must first be commuted to the position right in front of $c^\dagger_{\lambda_4}$. If this takes n_4 interchanges of Fermion operators we get a Fermi sign of $(-1)^{n_4}$. Bringing next c_{λ_3} right in front of $c^\dagger_{\lambda_3}$ by n_3 interchanges of Fermion operators gives a sign of $(-1)^{n_3}$. Analogously, c_{λ_1} and c_{λ_2} must be amongst the c_{μ_i} and the creation operators $c^\dagger_{\lambda_1}$ and $c^\dagger_{\lambda_2}$ in the Hamiltonian have to be commuted to the left to stand to the immediate right of their respective 'partner annihilation operator' so as to cancel it. If this requires an additional number of Fermion interchanges n_1 for $c^\dagger_{\lambda_1}$ and n_2 for $c^\dagger_{\lambda_2}$ there is an additional Fermi sign of $(-1)^{n_1+n_2}$. The total matrix element therefore is $(-1)^{n_1+n_2+n_3+n_4}V(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$. The correct Fermi sign is crucial for obtaining correct results and must be evaluated by keeping track of all necessary interchanges of Fermion operators. The necessity to determine the Fermi sign is the very reason why we have to adopt an ordering convention and strictly adhere to it.

Once the matrix $\langle\mu|H_1|\nu\rangle$ has been set up it can be diagonalized numerically. The following Table 4 gives the resulting multiplet energies for d^8 and d^7, the values of L and S for each multiplet and the degeneracy n. The values of the R^k parameters have been calculated [9] by using Hartree-Fock wave functions $R_{3,2}$ for Ni^{2+} and Co^{2+} in (8). The energy of the lowest multiplet is taken as the zero of energy and it turns out that all *energy differences* depend only on R^2 and R^4. Note the increasing complexity of the level schemes with increasing number of holes in the d-shell. Comparing the energies of the multiplets for d^8 with the experimental values in

E	S	L	n	Term	E	S	L	n	Term
0.0000	1	3	21	^3F	0.0000	3/2	3	28	^4F
1.8420	0	2	5	^1D	1.8000	3/2	1	12	^4P
1.9200	1	1	9	^3P	2.1540	1/2	4	18	^2G
2.7380	0	4	9	^1G	2.7540	1/2	5	22	^2H
13.2440	0	0	1	^1S	2.7540	1/2	1	8	^2P
					3.0545	1/2	2	10	^2D
					4.5540	1/2	3	14	^2F
					9.9774	1/2	2	10	^2D

Table 4: *Energies of the d^8 multiplets calculated with $R^2 = 10.479$ eV, $R^4 = 7.5726$ eV (Left), and energies of the d^7 multiplets calculated with $R^2 = 9.7860$ eV, $R^4 = 7.0308$ eV (Right).*

Table 1 one can see good agreement with deviations of order 0.1 eV. The only exception is ^1S. This is hardly a surprise because here the theoretical energy is $\approx$13 eV which is comparable to the difference in energy between the $3d$ and the $4s$ shell in Ni (which is $\approx$10 eV). It follows that the basic assumption of the calculation, namely that the separation between atomic shells is large compared to the multiplet splitting, is not fulfilled for this special multiplet. To treat ^1S more quantitatively it would likely be necessary to include basis states with configurations like $3d^7\,4s^1$, or, put another way, to consider the screening of the Coulomb interaction by particle-hole excitations from the $3d$ into the $4s$ shell.
Finally, the Table shows that the ground states indeed comply with the two first Hund's rules: they have maximum spin and maximum orbital angular momentum for this spin. It can be shown that this is indeed always the case as long as one uses Coulomb and exchange integrals with the correct, i.e. positive, sign [7, 8].

2.6 Spin-orbit coupling

So far we have neglected spin-orbit coupling but this can be included easily into the formalism. The corresponding Hamiltonian is

$$H_{SO} = \lambda_{SO} \sum_{i=1}^{n} \boldsymbol{l}_i \cdot \boldsymbol{S}_i = \lambda_{SO} \sum_{i=1}^{n} \left(l_i^z S_i^z + \frac{1}{2}\Big(l_i^+ S_i^- + l_i^- S_i^+ \Big) \right).$$

where $\boldsymbol{l}_i$ ($\boldsymbol{S}_i$) are the operators of orbital (spin) angular momentum of the ith electron. The spin-orbit coupling constant λ_{SO} can be written as [3]

$$\lambda_{SO} = \frac{\hbar^2}{2m_e^2 c^2 r_{orb}} \left. \frac{dV_{at}}{dr} \right|_{r=r_{orb}}$$

where m_e is the electron mass, c the velocity of light, V_{at} is the atomic potential acting on the electron and r_{orb} the spatial extent of the radial wave function.
The first term on the right hand side can be translated into second quantized form easily

$$H_{SO}^{\parallel} = \lambda_{SO} \sum_{m=-l}^{l} \frac{m}{2} \Big(c_{l,m,\uparrow}^{\dagger} c_{l,m,\uparrow} - c_{l,m,\downarrow}^{\dagger} c_{l,m,\downarrow} \Big). \tag{17}$$

As regards the transverse part, we note the matrix elements of the orbital angular momentum raising/lowering operator [3]: $\langle l, m \pm 1 | l^{\pm} | l, m \rangle = \sqrt{(l \mp m)(l \pm m + 1)}$ whence

$$H_{SO}^{\perp} = \frac{\lambda_{SO}}{2} \sum_{m=-l}^{l-1} \sqrt{(l-m)(l+m+1)} \Big(c_{l,m+1,\downarrow}^{\dagger} c_{l,m,\uparrow} + c_{l,m,\uparrow}^{\dagger} c_{l,m+1,\downarrow} \Big). \tag{18}$$

Spin-orbit coupling can be implemented rather easily into the exact diagonalization formalism discussed above, the main difficulty again is keeping track of the Fermi sign. Due to the fact that neither L^z nor S^z are conserved anymore the corresponding reduction of the Hilbert space is no longer possible. In $3d$ transition-metal compounds the spin-orbit coupling constant λ_{SO} for the $3d$ shell is rather small, of order $\lambda_{SO} \approx 0.05$ eV and can be neglected for many purposes. In the rare-earth elements spin-orbit coupling in the $4f$ shell is quite strong, $\lambda_{SO} \approx 0.5$ eV, and spin-orbit coupling must be taken into account.

3 Effects of the environment in the crystal

So far we have considered a single ion in vacuum. Next, we discuss how the results must be modified if the ion is embedded in a solid. We will see that the small spatial extent of the $3d$ or $4f$ radial wave functions $R_{n,l}(r)$ suppresses the effects of the environment in a solid, so that in many cases the main effect of embedding the ion into a solid is the partial splitting of the multiplets of the free ion. As in the preceding chapter we write down everything explicitly for a $3d$ shell but the theory is easily transferred to other shells.

In many transition-metal compounds the $3d$ ions are surrounded by an approximately octahedral or tetrahedral 'cage' of non-metal ions such as oxygen, sulphur, arsenic. These nearest neighbor ions, which will be called 'ligands' in the following, have a twofold effect: first, they produce a static electric field, the so-called *crystalline electric field* or CEF, and second there may be *charge transfer* that means an electron can tunnel back and forth between a ligand orbital and a $3d$-orbital of the transition metal ion due to the overlap of the respective wave functions. We discuss these effects one by one.

3.1 Crystalline electric field

Let us first consider the crystalline electric field, whereby we model the ligands by n_c point charges $Z_n e$ at the positions $\boldsymbol{R}_n$. The corresponding term in the Hamiltonian for the electrons on the ion in question is (recall that the electron charge is negative)

$$-V_{CEF}(\boldsymbol{r}) = -\sum_{n=1}^{n_c} \frac{Z_n}{|\mathbf{r} - \mathbf{R}_n|} = -\frac{Z_{av}}{R_{av}} \sum_{k=0}^{\infty} \sum_{m=-k}^{k} \gamma_{k,m} \left(\frac{r}{R_{av}} \right)^k \sqrt{\frac{4\pi}{2k+1}}\, Y_{k,m}(\vartheta, \varphi),$$

$$\gamma_{k,m} = \sqrt{\frac{4\pi}{2k+1}} \sum_{n=1}^{n_c} \frac{Z_n}{Z_{av}} \left(\frac{R_{av}}{R_n} \right)^{k+1} Y_{k,m}^{*}(\vartheta_n, \varphi_n). \tag{19}$$

Here we have again used multipole expansion (4) of the Coulomb potential and introduced the average distance and charge of the ligands, R_{av} and Z_{av}. Going over to 2nd quantization the Hamiltonian becomes [6]

$$H_{CEF} = \sum_{i,j} V_{CEF}(\nu_i, \nu_j)\; c^\dagger_{\nu_i} c_{\nu_j},$$
$$V_{CEF}(\nu_1, \nu_2) = \int dx\; \psi^*_{\nu_1}(x)\, V_{CEF}(\boldsymbol{r})\, \psi_{\nu_2}(x), \tag{20}$$

where the wave functions $\psi_\nu(x)$ are again given by (3). In calculating $V_{CEF}(\nu_1, \nu_2)$ we start with the sum over σ and find a factor of $\delta_{\sigma_1,\sigma_2}$. The integral over the polar angles (ϑ, φ) again gives a factor of $\delta_{m_1,m+m_2}$ and a Gaunt coefficient. As for the integral over r we note that the radial dependence of the wave functions $\psi_\nu(x)$ is given by $R_{3,2}(r)$, which differs appreciably from zero only in a narrow range $r \leq r_{3d}$. Then we find

$$V_{CEF}(\nu_1, \nu_2) = \delta_{\sigma_1,\sigma_2} \sum_k \gamma_{k,m_1-m_2}\, c^k(2, m_1; 2, m_2)\, I_k,$$
$$I_k = -\frac{Z_{av} e^2}{R_{av}} \left(\frac{r_{3d}}{R_{av}}\right)^k \int_0^\infty d\rho\; \rho^{k+2}\; \tilde{R}^2_{nl}(\rho). \tag{21}$$

Here we have introduced the dimensionless variable $\rho = r/r_{3d}$, and the dimensionless wave function $\tilde{R}_{nl}(\rho) = r_{3d}^{3/2} R_{nl}(\rho r_{3d})$. Since this has a range of unity and

$$\int_0^\infty d\rho\; \rho^2\; \tilde{R}^2_{nl}(\rho) = 1$$

we expect that the dimensionless radial integral in I_k is of order unity so that $I_k \propto \left(\frac{r_{3d}}{R_{av}}\right)^k$. As expected, a small $r_{3d} \ll R_{av}$ suppresses the effect of the environment and the sum over k usually can be terminated after the lowest $k > 0$ for which $\gamma_{k,m}$ does not vanish for some m. Moreover, for a d-shell it again follows from the triangular condition for the Gaunt coefficients that $k \leq 4$ and from parity that k only be even. The term with $k = 0$ gives merely a constant shift and can be omitted so that only $k = 2$ and $k = 4$ need to be considered. As was the case for the Coulomb interaction, the CEF can be described by very few – in fact only one if only the lowest order in r_{3d}/R_{av} is kept – parameters I_k which depend on the radial wave function $R_{3,2}(r)$. These parameters again are frequently fitted to experiment. The actual form of the matrix elements then depends on the geometry of the 'cage' of ligands via the sums $\gamma_{k,m}$.

As an example let us consider the case of an ideal octahedron of identical charges. More precisely, let the nucleus of the transition-metal ion be the origin of the coordinate system, and six identical charges eZ be located at $(\pm R, 0, 0)$, $(0, \pm R, 0)$ and $(0, 0, \pm R)$. This means that $R_n = R = R_{av}$ and $Z_n = Z = Z_{av}$, whence

$$\gamma_{k,m} \;=\; \sqrt{\frac{4\pi}{2k+1}} \sum_{n=1}^{6} Y^*_{k,m}(\vartheta_n, \varphi_n). \tag{22}$$

We divide the six charges into two groups: group 1 comprises the four charges in the x-y plane at $(\pm R, 0, 0)$ and $(0, \pm R, 0)$. These have $\vartheta_n = \frac{\pi}{2}$ and $\varphi_n = \frac{n\pi}{2}$ with $n = 0, 1, 2, 3$. Since

$Y_{l,m}(\vartheta,\varphi) = P_{l,m}(\vartheta)\, e^{im\varphi}$, we find that the contribution of group 1 to $\gamma_{k,m}$ is proportional to

$$\sum_{n=0}^{3} \left(e^{\frac{im\pi}{2}}\right)^n = \begin{cases} \dfrac{\left(e^{2\pi i}\right)^m - 1}{e^{\frac{im\pi}{2}} - 1} = 0 & e^{\frac{im\pi}{2}} \neq 1, \\ 4 & e^{\frac{im\pi}{2}} = 1. \end{cases}$$

The four charges of group 1 therefore give a nonvanishing contribution only for $m = 0, 4$. Group 2 comprises the two charges at $(0, 0, \pm R)$. Inspection of tables of spherical harmonics [4] shows that always

$$Y_{lm}(\vartheta,\varphi) \propto \sin^m(\vartheta)\, e^{im\varphi} = \left(\frac{x+iy}{r}\right)^m,$$

so that the charges of group 2 contribute only for $m = 0$.
Combining everything we see that for the ideal octahedron we need to actually evaluate the sum (22) only for $Y_{2,0}$, $Y_{4,0}$ and $Y_{4,\pm4}$ whereby for the last case only the charges in the x-y plane need to be considered. We start with $Y_{2,0}$ and note that $Y_{2,0}(\vartheta,\varphi) \propto 3\cos^2(\vartheta) - 1$ [4]. It follows that $\sum_{n=1}^{6}\ Y_{2,0}(\vartheta_n,\varphi_n) \propto 4\cdot(-1) + 2\cdot 2 = 0$, so that $Y_{2,0}$ does not contribute. Using the expressions [4]

$$Y_{4,0}(\vartheta,\varphi) = \frac{3}{16}\sqrt{\frac{1}{\pi}}\cdot\left(35\cos^4\vartheta - 30\cos^2\vartheta + 3\right)$$
$$Y_{4,4}(\vartheta,\varphi) = \frac{3}{16}\sqrt{\frac{35}{2\pi}}\cdot\sin^4\vartheta\cdot e^{4i\varphi}$$

we then find after straightforward calculation

$$\gamma_{4,0} = \sqrt{\frac{49}{4}} \qquad \text{and} \qquad \gamma_{4,4} = \sqrt{\frac{35}{8}}, \tag{23}$$

as well as $\gamma_{4,-4} = \gamma_{4,4}$. Using the tabulated values of the $c^4(2,m;2,m')$ (see Appendix), $V_{CEF}(\nu_1,\nu_2)$ can be written as $\delta_{\sigma_1,\sigma_2}$ times a matrix in the indices m_1 and m_2

$$V_{CEF}(m_1,m_2) \;=\; \frac{I_4}{6}\begin{pmatrix} 1 & 0 & 0 & 0 & 5 \\ 0 & -4 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & -4 & 0 \\ 5 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{24}$$

This matrix has the eigenvalues I_4 (twofold degenerate) with corresponding eigenfunctions

$$d_{x^2-y^2}(\Omega) = \frac{1}{\sqrt{2}}\big(Y_{2,-2}(\Omega) + Y_{2,2}(\Omega)\big) = \sqrt{\frac{15}{16\pi}}\frac{x^2-y^2}{r^2},$$
$$d_{3z^2-r^2}(\Omega) = \qquad Y_{2,0}(\Omega) \qquad = \sqrt{\frac{5}{16\pi}}\frac{3z^2-r^2}{r^2}, \tag{25}$$

and $-2I_4/3$ (threefold degenerate) with eigenfunctions

$$
\begin{aligned}
d_{xy}(\Omega) &= \frac{i}{\sqrt{2}}\big(Y_{2,-2}(\Omega) - Y_{2,2}(\Omega)\big) = \sqrt{\frac{15}{4\pi}}\frac{xy}{r^2},\\
d_{yz}(\Omega) &= \frac{i}{\sqrt{2}}\big(Y_{2,-1}(\Omega) + Y_{2,1}(\Omega)\big) = \sqrt{\frac{15}{4\pi}}\frac{yz}{r^2},\\
d_{xz}(\Omega) &= \frac{1}{\sqrt{2}}\big(Y_{2,-1}(\Omega) - Y_{2,1}(\Omega)\big) = \sqrt{\frac{15}{4\pi}}\frac{xz}{r^2}.
\end{aligned}
\tag{26}
$$

The two eigenfunctions for eigenvalue I_4 are called e_g orbitals, whereas the three eigenfunctions for eigenvalue $-2I_4/3$ are called t_{2g} orbitals. If the ligands are O^{2-} ions, $Z{=}-2$ whence $I_4 > 0$, i.e., the e_g orbitals are higher in energy than the t_{2g} orbitals. This can be readily understood by comparing the d_{xy} and the $d_{x^2-y^2}$ orbital. In the x-y plane the lobes of $d_{x^2-y^2}$ are along the axes and point directly towards the negative charges at $(\pm R, 0, 0)$ and $(0, \pm R, 0)$, whereas the lobes of the d_{xy} orbital point along the diagonals and thus optimally avoid these negative charges. For the negatively charged electron, it is therefore energetically advantageous to be in the d_{xy} orbital. The splitting between the eigenvalues is frequently called $10Dq = E(e_g) - E(t_{2g})$, so that in our point-charge model $Dq = I_4/6$.

Note that the five functions $d_\alpha(\Omega)$ in (25) and (25) are pairwise orthogonal. This means that they are obtained by a unitary transformation from the five original spherical harmonics $Y_{2,m}(\Omega)$ and can be used as basis functions. These functions are of utmost importance in the theoretical discussion of elements with partially filled d-shells and are can be found again and again in the literature. Polar plots of these functions also can be found in the literature [4].

We see that for octahedral coordination the effect of the CEF on a $3d$ level can be summarized in a single parameter $10Dq$, which may for example be obtained by a fit to experiment. This way of dealing with the CEF is very similar in spirit to our treatment of the Coulomb interaction, in that details of the radial wave functions $R_{n,l}(r)$ are absorbed into numerical parameters which can be adjusted to experiment. Alternatively, the numerical value of $10Dq$ for a given solid may also be obtained from a fit to a density functional band structure.

By adding H_{CEF}, which is a quadratic form in the operators $c_\nu^\dagger/c_\nu$, to the Hamiltonian for the intra-atomic Coulomb interaction discussed above we can now discuss the splitting of the original multiplets of the free ion under the influence of the electrostatic potential of the environment. The following should be noted: the above discussion refers to the wave function of a *single electron*. The multiplets, however, are collective eigenstates of all n electrons in an atomic shell which are created by the Coulomb interaction between electrons. The question of how these collective states split in a cubic environment is not at all easy to answer. One way would be exact diagonalization including the term H_{CEF}.

Plots of the energies of the resulting 'crystal-field multiplets' versus $10Dq$ are called Tanabe-Sugano diagrams [10]. An example is shown in Figure 3.1 which shows the eigenenergies of the d^8 and d^7 configuration with Coulomb interaction and increasing cubic CEF, $10Dq$. One realizes that the highly degenerate multiplets of the free ion are split into several levels of lower degeneracy by the CEF, which is to be expected for a perturbation which lowers the symmetry.

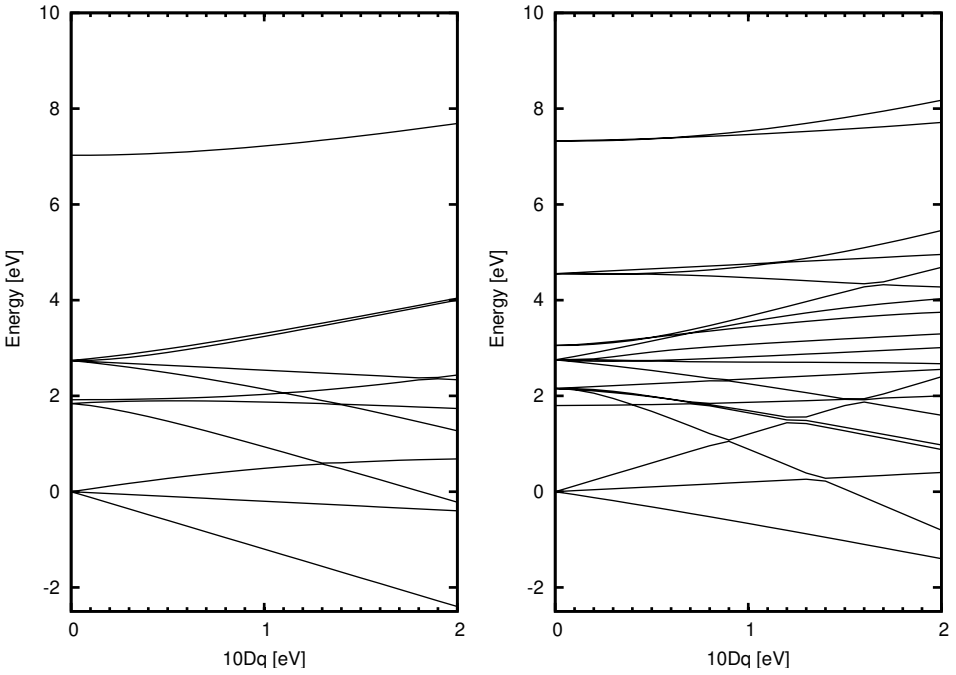


Fig. 3: *Examples for Tanabe-Sugano diagrams: the splitting of multiplets of d^8 (left) and d^7 (right) for increasing $10Dq$. The Slater-Condon parameters have the values given in Table 2.*

Note that the components into which a given multiplet splits all have the same spin as the multiplet itself. This is because the spin of an electron does not 'feel' an electrostatic potential; or, more precisely, because the operator of total spin commutes with any operator which acts only on the real-space coordinates $\boldsymbol{r}_i$ of the electrons.

An interesting example for the application of the Tanabe-Sugano diagrams are transition-metal ions in aqueous solution. In fact, the preference of transition-metal ions for an environment with cubic symmetry is so strong that such immersed ions often surround themselves with an octahedron of water molecules. Thereby the dipole moments of these six molecules all point away from the ion and thus create an electric field which cubic symmetry which again gives rise to an e_g-t_{2g} splitting. Optical transitions between the CEF-split multiplets, which are possible only due to slight distortions of the octahedron or the generation/annihilation of vibrational quanta during the transition, correspond to frequencies in the visible range and result in the characteristic colors of such solutions. The Tanabe-Sugano diagrams have proved to be a powerful tool to understand the absorption spectra of such solutions [8]. By matching the energies of the observed transitions to energy differences in the Tanabe-Sugano diagrams one can extract estimates for the Slater-Condon parameters and for $10Dq$. The values of the Slater-Condon parameters turn out to be somewhat smaller than those for ions in vacuum due to dielectric screening in the solution. An independent estimate for $10Dq$ can also be extracted from measured heats of hydration – this is because both $10Dq$ and the electrostatic energy of the system 'ion plus octahedron' depend on the distance between the transition-metal ion and the water molecules – and compared to the estimate from the absorption spectrum whereby good agreement is usually obtained [11].

3.2 Charge transfer

We continue our discussion of a transition-metal ion at the origin of the coordinate system surrounded by a 'cage' of n ligands at $\mathbf{R}_n$. The second mechanism by which the ligands may influence the energy levels of the transition metal ion is *charge transfer*. This means that the $3d$ levels of the transition metal ion hybridize with atomic orbitals on the ligands which shifts the energies of the $3d$ levels. To understand how this happens, let us consider a toy Hamiltonian which describes just a single 'd-orbital' $|\psi_1\rangle$ with energy ε_1 coupled to a single 'ligand orbital' $|\psi_2\rangle$ with energy ε_2

$$H = \sum_{i=1}^{2} \varepsilon_i \, c_i^\dagger c_i - \left(t \; c_1^\dagger c_2 + H.c.\right)$$

We have suppressed the spin index and the meaning of the creation/annihilation operators should be self-evident). The hybridization matrix element $-t = \langle\psi_1|H|\psi_2\rangle$ thereby originates from the overlap of the atomic wave functions and facilitates the transfer of an electron between the two orbitals. The ansatz $|\psi\rangle = u|\psi_1\rangle + v|\psi_2\rangle$ for an eigenstate readily leads to the 2×2 matrix

$$h = \begin{pmatrix} \varepsilon_1 & -t \\ -t & \varepsilon_2 \end{pmatrix}, \tag{27}$$

whose eigenvalues are

$$E_\pm = \frac{\varepsilon_1+\varepsilon_2}{2} \pm \sqrt{\left(\frac{\varepsilon_1-\varepsilon_2}{2}\right)^2 + t^2}.$$

We may assume without loss of generality that $\varepsilon_1 > \varepsilon_2$, whence $\sqrt{\left(\frac{\varepsilon_1-\varepsilon_2}{2}\right)^2 + t^2} = \frac{\varepsilon_1-\varepsilon_2}{2} + \Delta$, with some $\Delta > 0$. It follows that $E_- = \varepsilon_2 - \Delta < \varepsilon_2$ and $E_+ = \varepsilon_1 + \Delta > \varepsilon_1$. This means that the lower level is shifted downwards by Δ, whereas the upper level is shifted upwards by the same amount, an effect known as *level repulsion*. This mechanism can split the degeneracy of the $3d$-level because, depending on the geometry of the cage, different $3d$ orbitals can have different hybridization matrix elements with the ligand orbitals.

Note that the eigenstates now are a mixture of the two orbitals. For $t \ll \varepsilon_1 - \varepsilon_2$, however, the weight of $|\psi_2\rangle$ in the eigenstate for E_- is $\left(\frac{t}{\varepsilon_1-\varepsilon_2}\right)^2$ which means the state still has predominant $|\psi_1\rangle$ character.

To describe charge transfer quantitatively we need to enlarge our set of Fermion operators $c_\nu^\dagger/c_\nu$ by operators $l_\mu^\dagger/l_\mu$ which create/annihilate electrons in orbitals centered on the ligands. We simplify matters by assuming that only $2p$ orbitals are relevant for the ligands, as would be the case for oxygen ligands. For the rest of this paragraph on charge transfer we switch to a new set of basis functions which is more suitable for the discussion of hybridization. First, we use $3d$ wave functions whose angular part is given by the real-valued spherical harmonics (25) and (26)

$$\psi_{\nu_i}(x) \;\; = \;\; R_{3,2}(r) \; d_\alpha(\Omega) \; \delta_{\sigma,\sigma_i}, \tag{28}$$

with $\alpha \in \{xy, xz, yz, x^2-y^2, 3z^2-r^2\}$, so that now $\nu_i = (\alpha, \sigma)$.

For the ligand orbitals we use wave functions whose angular part is given by the real-valued p-like spherical harmonics

$$p_x(\Omega) = \frac{1}{\sqrt{2}}\big(-Y_{1,1}(\Omega) + Y_{1,-1}(\Omega)\big) = \sqrt{\frac{3}{4\pi}}\,\frac{x}{r}, \tag{29}$$

$$p_y(\Omega) = \frac{i}{\sqrt{2}}\big(\;\; Y_{1,1}(\Omega) + Y_{1,-1}(\Omega)\big) = \sqrt{\frac{3}{4\pi}}\,\frac{y}{r}, \tag{30}$$

$$p_z(\Omega) = \qquad\quad Y_{1,0}(\Omega) \qquad\quad = \sqrt{\frac{3}{4\pi}}\,\frac{z}{r}, \tag{31}$$

and are centered on the ligands

$$\psi_{\mu_j}(x) \;\; = \;\; R_{2,1}(r_{n_j})\; p_{\beta_j}(\Omega_n)\; \delta_{\sigma,\sigma_j}. \tag{32}$$

Here, $\mathbf{r}_n = \mathbf{r} - \mathbf{R}_n$ and $\beta \in \{x, y, z\}$ so that $\mu_j = (n_j, \beta_j, \sigma_j)$. The obvious generalization of the toy Hamiltonian then is

$$H_{CT} = \sum_i \varepsilon_{\nu_i} c^\dagger_{\nu_i} c_{\nu_i} + \sum_j \varepsilon_{\mu_j} l^\dagger_{\mu_j} l_{\mu_j} - \sum_{i,j} \left(t_{\nu_i,\mu_j}\; c^\dagger_{\nu_i} l_{\mu_j} + H.c. \right). \tag{33}$$

This would still not be very useful because it contains a large number of parameters, in particular the *hybridization integrals* $-t_{\nu_i,\mu_j}$. The crucial simplification comes about because these hybridization integrals can be expressed in terms of very few parameters by using the celebrated *Slater-Koster tables* [12]. For example, for the present case where only the p orbitals of the ligands are taken into account there are just two relevant parameters: $V_{pd\sigma}$ and $V_{pd\pi}$. More precisely, a typical entry in the Slater-Koster tables looks like

$$-t_{1x,2xy} = \sqrt{3}\, l^2 m\, V_{pd\sigma} + m\,(1-2l^2)\, V_{pd\pi}.$$

This gives the hopping integral $-t_{1x,2xy}$ between a p_x orbital on atom 1 and a d_{xy} orbital on atom 2 as a function of the components of the unit vector (l, m, n) pointing from atom 1 to atom 2. Thereby the parameters $V_{pd\sigma}$ and $V_{pd\pi}$ depend only on the distance between the two atoms. It is obvious from this that the hopping orbitals $-t_{\nu_i,\mu_j}$ in Eq. (33) depend on the geometry of the 'cage' of ligands. By inserting the unitary transformation (25) and (26) as well as (31), H_{CT} now could be transformed to the original complex spherical harmonics $Y_{2,m}(\Omega)$ and then be easily included into exact diagonalization formalism discussed above. The main problem is that the number of orbitals in the cluster and hence the dimension of the Hilbert space increases considerably so that one has to resort to numerical methods such as the Lanczos algorithm [13]. To illustrate the procedure and thereby show how to alleviate the problem of the increase of the Hilbert space dimension, we specialize again to the case where the ligands form an ideal octahedron, with the transition metal ion in the center of gravity. In other words, the ligands again are located at $(\pm R, 0, 0)$, $(0, \pm R, 0)$ and $(0, 0, \pm R)$. We want to solve the Hamiltonian (33) for this cluster of seven ions assuming that the parameters $V_{pd\sigma}$, $V_{pd\pi}$, ε_{ν_i} and ε_{μ_j} are given. For simplicity we set the energies ε_{ν_i} of the $3d$ orbitals equal to zero and assume that $\varepsilon_{\mu_j} = \varepsilon > 0$

for all ligand orbitals. $V_{pd\sigma}$ and $V_{pd\pi}$ depend only on the distance between ligand and transition-metal ion and therefore are the same for all six ligands. Since we are retaining three p-orbitals on each ligand and the five d-orbitals on the transition-metal ion, the total number of orbitals in the cluster would be $5 + 6 \cdot 3 = 23$. What we would have to do is to go through all six ligands, determine (l, m, n) for each of them, set up the hopping integral between each of the five $3d$ orbitals and each of the three $2p$ orbitals on the respective ligand using the Slater-Koster tables. This would give us a 23×23 matrix instead of the 2×2 matrix (27), the eigenvalues of which would tell us how the $3d$ orbitals are shifted by the hybridization. Fortunately enough, the high symmetry of the octahedral cluster allows us to bring the Hamiltonian to block-diagonal form and obtain analytical expressions for the energies. The key simplification comes about by constructing *hybridizing combinations* of $2p$ orbitals on the six ligands. Consider the d_{xy} orbital in Figure 4. Using symmetry arguments or the Slater-Koster tables one can show that out of the 18 p orbitals on the ligands only the four p-orbitals shown in the Figure have a nonvanishing hybridization integral with the d_{xy} orbital. These four orbitals moreover hybridize with no other d orbital. Then, we form the following linear combinations of these four orbitals:

$$
\begin{aligned}
|1\rangle &= \frac{1}{2}\Big(\psi_{1,y}(x) + \psi_{2,x}(x) - \psi_{3,y}(x) - \psi_{4,x}(x)\Big),\\
|2\rangle &= \frac{1}{2}\Big(\psi_{1,y}(x) + \psi_{2,x}(x) + \psi_{3,y}(x) + \psi_{4,x}(x)\Big),\\
|3\rangle &= \frac{1}{2}\Big(\psi_{1,y}(x) - \psi_{2,x}(x) - \psi_{3,y}(x) + \psi_{4,x}(x)\Big),\\
|4\rangle &= \frac{1}{2}\Big(\psi_{1,y}(x) - \psi_{2,x}(x) + \psi_{3,y}(x) - \psi_{4,x}(x)\Big),
\end{aligned}
$$

where we have dropped the spin index of the $\psi_{\mu_j}(x)$ for brevity. If p orbitals on different ligands are orthogonal to each other, $\langle\psi_{i,\alpha}|\psi_{j,\beta}\rangle = \delta_{i,j}\delta_{\alpha,\beta}$, these four combinations are orthonormal, that means we can use them as new basis functions. Next, using the matrix elements of H indicated in Figure 4, which can be easily verified using the Slater-Koster tables, we see that

$$\langle d_{xy}|H_{CT}|i\rangle = -2V_{pd\pi}\ \delta_{i,1}.$$

This means that the states $|2\rangle$, $|3\rangle$ and $|4\rangle$ do not mix with d_{xy} and since they also do not mix with any other of the five d-orbitals, they are eigenstates of H_{CT} with energy ε by construction. We thus need to keep only $|d_{xy}\rangle$ and $|1\rangle$ and thus arrive at exactly the same 2×2 matrix Eq. (27) as for the toy model

$$h = \begin{pmatrix} 0 & -2V_{pd\pi} \\ -2V_{pd\pi} & \varepsilon \end{pmatrix}. \tag{34}$$

with eigenvalues $E_\pm = \frac{\varepsilon}{2} \pm \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + 4V_{pd\pi}^2}$. To simplify our expressions we assume weak hybridization, $V_{pd\pi} \ll \varepsilon$, whence the energy of the lower eigenstate, which for $\varepsilon > 0$ has predominantly $|d_{xy}\rangle$ character, becomes

$$E(t_{2g}) \approx -\frac{4V_{pd\pi}^2}{\varepsilon}.$$

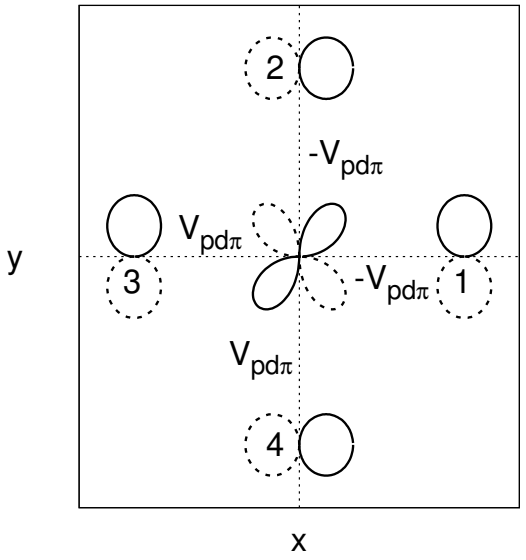


Fig. 4: *p orbitals on ligands with nonvanishing hybridization with the d_{xy} orbital in the center. The figure shows the x-y plane, lobes with positive (negative) sign are drawn by full (dashed) lines. The labels of the ligands are given next to the p orbitals, the hybridization integrals obtained from the Slater-Koster tables are indicated for each bond.*

The upper eigenstate, which predominantly has ligand-p character, has energy $\varepsilon + 4V_{pd\pi}^2/\varepsilon$. We could have proceeded in exactly the same way if instead of the x-y plane we would have considered the x-z or y-z plane and the d_{xz} or d_{yz} orbitals. Therefore, all of the three t_{2g} orbitals are shifted by the same energy and remain degenerate in the presence of hybridization!
In a similar but slightly more complicated way one finds that the e_g-orbitals $d_{x^2-y^2}$ and $d_{3z^2-r^2}$ also remain degenerate and are shifted to

$$E(e_g) \approx -\frac{4V_{pd\sigma}^2}{\varepsilon}.$$

We have thus found the energy levels of the Hamiltonian (33) for the octahedral cluster with only p orbitals on the ligands: there are five states with predominant $3d$ character and energies $-4V_{pd\pi}^2/\varepsilon$ (t_{2g}, 3-fold degenerate) or $-4V_{pd\sigma}^2/\varepsilon$ (e_g, 2-fold degenerate). We also have five corresponding states with predominant p-character and energies $\varepsilon + 4V_{pd\pi}^2/\varepsilon$ (3-fold degenerate) or $\varepsilon + 4V_{pd\sigma}^2/\varepsilon$ (2-fold degenerate). And finally we have the non-bonding combinations which have pure p character and retain their energy of ε. Obviously there must be 13 of these.
We see that charge transfer results in the same splitting into t_{2g} and e_g orbitals as the electrostatic potential due to the charges on the ligands. (In fact, it follows from the theory of irreducible representations of symmetry groups [3, 8, 10, 11] that this holds true for *any* perturbation with cubic symmetry). Therefore, if we are only interested in the energies of the eigenstates we may as well drop the ligand orbitals from the Hamiltonian and describe the splitting due to charge transfer by an 'effective $10Dq$' given by

$$10Dq_{CT} = \frac{4}{\varepsilon}\big(V_{pd\pi}^2 - V_{pd\sigma}^2\big).$$

This would have to be added to the 'electrostatic $10Dq$' discussed earlier.

To conclude this section, we mention that using the octahedron-shaped cluster discussed in the preceding section by the exact diagonalization method has been an extraordinarily successful method for the simulation of valence band photoemission spectra, X-ray absorption spectra, and core-level photoemission spectra of $3d$ transition-metal compounds [14–20]. In many cases, the spectra calculated in a mere octahedron can be compared peak-by-peak to experimental spectra. This also provides unambiguous evidence that the multiplets of the free ion, slightly modified by CEF and charge transfer, do persist in the solid.

4 Multiband Hubbard models

We have now discussed all necessary parts of the Hamiltonian to describe transition-metal and rare-earth compounds, i.e., multiband Hubbard models. We view the solid as an array of ions with a certain number of atomic orbitals on each of them and assume that these orbitals are labeled by some index i. The position of the ion on which orbital i is centered is $\mathbf{R}_i$. Then, we split the orbitals in the solid into two groups: the correlated and the uncorrelated orbitals. The correlated orbitals have radial wave functions with small spatial extent and the Coulomb interaction between electrons in these orbitals is strong. The uncorrelated orbitals are more extended and the Coulomb interaction between electrons in these orbitals is weak enough to be neglected. Of course, this division of the orbitals is arbitrary to some extent. In principle, one might also include Coulomb interaction between electrons in orbitals on different ions but we neglect this because it will in general be much weaker than the interaction between electrons on the same ion.

Then, the problem arises how to choose these orbitals. For example for a d shell we could choose orbitals whose radial part is given by the spherical harmonics $Y_{2,m}(\Omega)$ but we might as well choose the real-valued spherical harmonics $d_\alpha(\Omega)$ in (25) and (26). The preceding discussion has shown that the $Y_{l,m}(\Omega)$ are convenient for the discussion of 'purely atomic' aspects of the problem, such as the Coulomb interaction within atomic shells and the spin-orbit coupling, whereas the real-valued spherical harmonics $d_\alpha(\Omega)$ are more convenient for 'solid related' aspects such as inter-ion hopping and CEF splitting. Since the $Y_{l,m}(\Omega)$ and the $d_\alpha(\Omega)$ are related by a unitary transformation this is more of a notational problem. Next, we introduce creation/annihilation operators $c^\dagger_{\nu_i}/c_{\nu_i}$ for electrons in these orbitals. Thereby we choose the compound index $\nu_i = (\mathbf{R}_i, n_i, l_i, m_i, \sigma_i)$ for Coulomb interaction and spin-orbit coupling and $\nu_i = (\mathbf{R}_i, n_i, l_i, \alpha_i, \sigma_i)$ with $\alpha_i \in \{s, p_x, p_y, p_z, d_{xy}, \dots\}$ for the inter-ion hopping and CEF. The inter-ion hopping is obtained be generalizing (33)

$$H_0 = \sum_i \varepsilon_{\nu_i} c^\dagger_{\nu_i} c_{\nu_i} - \sum_{i \neq j} \left(t_{\nu_i,\nu_j}\ c^\dagger_{\nu_i} c_{\nu_j} + H.c. \right). \tag{35}$$

The hopping integrals $-t_{\nu_i,\nu_j}$ again can be expressed in terms of relatively few V-parameters via the Slater-Koster tables, the numerical values of the V-parameters and the energies ε_{ν_i} can be obtained by fit to a density functional band structure. For the correlated orbitals thereby extra care is necessary due to the 'double counting problem' (see, e.g., Ref. [21]).

Moreover, we add for each ion the electrostatic part of the CEF, (20). So far the Hamiltonian is a quadratic form in Fermion operators and can always be solved after Fourier transform.
Next, for the correlated orbitals we add the Coulomb interaction (2) with the matrix elements (9). Since we are considering only the Coulomb interaction within a given atomic shell, all four $c^\dagger_{\nu_i}/c_{\nu_i}$ operators in each term of (2) must have the same $\mathbf{R}_i$. The Hamiltonian now is quartic in Fermion operators and thus not solvable anymore. Rather, we have to resort to one of the many approximation schemes known so far for correlated electrons. Finally we may also add the spin-orbit coupling. Whether this is necessary depends on the magnitude of the spin-orbit coupling constant λ_{SO}. Since spin-orbit coupling is a relativistic effect, λ_{SO} is larger for heavy elements. It is more or less negligible for $3d$ ions, but important for $5d$ transition metals or $4f$ rare earths.
It is obvious that the resulting Hamiltonian is quite complicated and it is highly desirable to simplify it. There are several possible ways to do so.

1. 'Integrating out' uncorrelated orbitals which act only to connect correlated orbitals.

   To see what this means, consider the toy Hamiltonian for three orbitals $|d_1\rangle$, $|d_2\rangle$ and $|l\rangle$:

   $$H = \Delta\, l^\dagger l - t\left(d_1^\dagger l + l^\dagger d_1 + d_2^\dagger l + l^\dagger d_2\right),$$

   where we have dropped the spin index for simplicity and the meaning of the Fermion operators should be obvious. It may be viewed as describing a 'bond' connecting the two 'd-orbitals' $|d_1\rangle$ and $|d_2\rangle$ (which have an energy of zero) via the 'bridging orbital' $|l\rangle$ which has energy Δ. We introduce the bonding/antibonding combinations $d^\dagger_\pm = \frac{1}{\sqrt{2}}\,(d_1^\dagger \pm d_2^\dagger)$ whence the Hamiltonian becomes

   $$H = \Delta\, l^\dagger l - \sqrt{2}t\left(\, d_+^\dagger l + l^\dagger d_+\right).$$

   the *ansatz* $|\psi\rangle = (A_+ d_+^\dagger + A_l l^\dagger + A_- d_-^\dagger)|0\rangle$ then leads to the 3×3 Hamilton matrix

   $$h = \begin{pmatrix} 0 & -\sqrt{2}t & 0 \\ -\sqrt{2}t & \Delta & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

   which has eigenvalues $E = 0, (\Delta \pm \sqrt{\Delta^2+8t^2})/2$. For the sake of simplicity we consider the limit $\Delta \gg t$ whence the energies become $E = 0, \Delta{+}2t^2/\Delta, -t^2/\Delta$. The eigenfunction for $E_1 = 0$ is $|\psi_1\rangle = d_-^\dagger|0\rangle$, the one for $E_2 = -2t^2/\Delta$ is $|\psi_2\rangle \approx d_+^\dagger|0\rangle$ and the one for $E_3 = \Delta{+}2t^2/\Delta$ is $|\psi_3\rangle \approx l^\dagger|0\rangle$. In other words, the wave function for the high energy state E_3 has mainly 'bridging orbital' character, whereas those of the two low energy states E_1 and E_2 have predominant d-character. Now consider the effective Hamiltonian

   $$H_{\mathit{eff}} = -\frac{t^2}{\Delta}\sum_{i=1,2} d_i^\dagger d_i - \frac{t^2}{\Delta}\Big(d_1^\dagger d_2 + H.c.\Big).$$

   It is obvious that the eigenenergies and corresponding eigenstates of H_{eff} are the same as the two low energy eigenstates of the original Hamiltonian. In other words, H_{eff} describes

the *low energy sector* of the full Hamiltonian and the high-energy bridging orbital has disappeared.

With this reasoning, one is often omitting uncorrelated 'bridging orbitals' from the Hamiltonian H_0 and uses an effective $\tilde{H}_0$ that comprises only the correlated orbitals and 'effective hopping integrals'. The latter can again be obtained by a fit to the band structure, whereby however only bands with predominant d-character must be taken into account. Clearly, this reduces the number of orbitals which is important if one uses numerical methods.

2. Taking the limit of large CEF or, in the simplest case where the correlated electrons are in octahedral coordination, the limit of large $10Dq$. Then, one may restrict the basis to states where the numbers of electrons in the t_{2g} and e_g orbitals are fixed. For example, for Ni^{2+} (i.e. d^8) in cubic symmetry one may assume in the limit of large $10Dq$ that the six t_{2g}-orbitals always are completely filled. Then, one needs to consider only the two electrons in the partially filled e_g level, resulting in a significant reduction of the number of possible basis states. Similarly, for compounds containing early transition metals such as Scandium, Titanium or Vanadium, one often assumes that the e_g orbitals are so high in energy that only the t_{2g} orbitals need to be taken into account.

3. Finally, one may use the simplified form of the Coulomb interaction as in Eq. (14).

An example for this 'reduction process' can be found in the paper by Craco *et al.* [22] where the authors discuss the photoemission and inverse photoemission spectrum of $SmO_{1-x}F_xFeAs$ thereby using a Hamiltonian which contains only the five Fe $3d$ orbitals and a Coulomb interaction of precisely the form (14) where $U_{m,m'}$ and $J_{m,m'}$ are replaced by average values.

5 Conclusion

We have seen that the Coulomb repulsion between electrons in partially filled atomic shells leads to multiplet splitting. The multiplets may be viewed as collective excitations of the 'not-so-many-body-system' formed by the electrons in the shell. We have seen that a relatively simple theory—essentially degenerate first order perturbation theory—describes the energies of the multiplets quite well and gives a good description of the line spectra of free atoms. If transition metal atoms are embedded into a solid the collective excitations of the electrons in their partly filled $3d$ shells are modified by the crystalline electric field of their environment and by hybridization with orbitals on neighboring atoms. If these effects are taken into account, which is relatively easy if one uses the exact diagonalization method, the resulting 'extended multiplet theory' turns out to be quite successful in reproducing a wide variety of experimental results for transition metal compounds. While this 'extended multiplet theory' refers to a single transition metal ion, we have also seen that there are simplifications and extensions of this theory to lattice systems, i.e., the multiband Hubbard models. These then are the appropriate models to describe compounds containing $3d$ or $4d$ transition metal ions.

A Gaunt coefficients

m	m'	c^0	$7\,c^2$	$21\,c^4$	a^0	$49\,a^2$	$441\,a^4$	b^0	$49\,b^2$	$441\,b^4$
±2	±2	1	-2	1	1	4	1	1	4	1
±2	±1	0	$\sqrt{6}$	$-\sqrt{5}$	1	-2	-4	0	6	5
±2	0	0	-2	$\sqrt{15}$	1	-4	6	0	4	15
±1	±1	1	1	-4	1	1	16	1	1	16
±1	0	0	1	$\sqrt{30}$	1	2	-24	0	1	30
0	0	1	2	6	1	4	26	1	4	36
±2	∓2	0	0	$\sqrt{70}$	1	4	1	0	0	70
±2	∓1	0	0	$-\sqrt{35}$	1	-2	-4	0	0	35
±1	∓1	0	$-\sqrt{6}$	$-\sqrt{40}$	1	1	16	0	6	40

Table 5: *Gaunt coefficients* $c^k(2,m;2,m')$*, and the* $a^k(2,m;2,m')$ *and* $b^k(2,m;2,m')$

m	m'	c^0	$15\,c^2$	$33\,c^4$	$\frac{429}{5}\,c^6$
±3	±3	1	-5	3	-1
±3	±2	0	5	$-\sqrt{30}$	$\sqrt{7}$
±3	±1	0	$\sqrt{10}$	$\sqrt{54}$	$-\sqrt{28}$
±3	0	0	0	$-\sqrt{63}$	$\sqrt{84}$
±2	±2	1	0	-7	6
±2	±1	0	$\sqrt{15}$	$\sqrt{32}$	$-\sqrt{105}$
±2	0	0	$-\sqrt{20}$	$-\sqrt{3}$	$4\sqrt{14}$
±1	±1	1	3	1	-15
±1	0	0	$\sqrt{2}$	$\sqrt{15}$	$5\sqrt{14}$
0	0	1	4	6	20
±3	∓3	0	0	0	$-\sqrt{924}$
±3	∓2	0	0	0	$\sqrt{462}$
±3	∓1	0	0	$\sqrt{42}$	$-\sqrt{210}$
±2	∓2	0	0	$\sqrt{70}$	$\sqrt{504}$
±2	∓1	0	0	$-\sqrt{14}$	$-\sqrt{378}$
±1	∓1	0	$-\sqrt{24}$	$-\sqrt{40}$	$-\sqrt{420}$

Table 6: *The Gaunt coefficients* $c^k(3,m;3,m')$

References

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

[2] Yu. Ralchenko, A.E. Kramida, J. Reader, and NIST ASD Team
NIST Atomic Spectra Database (ver. 4.1.0), (2011)
`https://physics.nist.gov/asd`

[3] L.D. Landau and E.M. Lifshitz: *Course of Theoretical Physics*
(Pergamon Press, Oxford New York, 1977)

[4] *Table of spherical harmonics* (Wikipedia)
`https://en.wikipedia.org/wiki/Table_of_spherical_harmonics`

[5] J.D. Jackson: *Classical Electrodynamics* (Wiley, New York, 1999)

[6] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many Particle Systems*
(McGraw-Hill, San Francisco, 1971)

[7] J.C. Slater: *Quantum Theory of Atomic Structure*
(McGraw-Hill, New York, 1960)

[8] J.S. Griffith: *The Theory of Transition Metal Ions*
(Cambridge University Press, Cambridge, 1961)

[9] M. Haverkort, PhD Thesis (Universität zu Köln, 2005)

[10] S. Sugano, Y. Tanabe, and H. Kitamura: *Multiplets of Transition Metal Ions*
(Academic Press, New York 1970)

[11] B.N. Figgis: *Introduction to Ligand Fields*
(Interscience Publishers, New York London Sydney, 1966)

[12] J.C. Slater and G.F. Koster, Phys. Rev. **94**, 1498 (1054)
For an online version of the tables see *Tight binding* (Wikipedia)
`https://en.wikipedia.org/wiki/Tight_binding`

[13] E. Koch: *The Lanczos Method* in E. Pavarini, E. Koch, and P. Coleman (eds.):
Many-Body Physics: From Kondo to Hubbard
Modeling and Simulation, Vol. 5 (Forschungszentrum Jülich, 2015)

[14] A. Fujimori and F. Minami, Phys. Rev. B **30**, 957 (1984)

[15] J. van Elp, J.L. Wieland, H. Eskes, P. Kuiper, G.A. Sawatzky, F.M.F. de Groot, and
T.S. Turner, Phys. Rev. B **44**, 6090 (1991)

[16] A. Fujimori, N. Kimizuka, T. Akahane, T. Chiba, S. Kimura, F. Minami,
K. Siratori, M. Taniguchi, S. Ogawa, and S. Suga, Phys. Rev. B **42**, 7580 (1990)

[17] F.M.F. de Groot, Journal of Electron Spectroscopy and Related Phenomena, **67** 525 (1994)

[18] M. Finazzi, N.B. Brookes, and F.M.F. de Groot, Phys. Rev. B **59**, 9933 (1999)

[19] F.M.F. de Groot, Coordination Chemistry Reviews, **249** 31 (2005)

[20] F.M.F. de Groot and A. Kotani: *Core Level Spectroscopy of Solids* (Taylor & Francis, Abingdon on Thames, 2008)

[21] I.A. Nekrasov, N.S. Pavlov, and M.V. Sadovskii, JETP Lett. **95**, 581 (2012)

[22] L. Craco, M.S. Laad, S. Leoni, and H. Rosner, Phys. Rev. B **78**, 134511 (2008)

5 Exchange Mechanisms

Erik Koch
Jülich Supercomputer Centre
Forschungszentrum Jülich

Contents

1 Introduction **2**

2 Coulomb exchange **4**

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

4 Double exchange **18**

5 Orbital-ordering **21**

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

7 Conclusion **27**

A Pauli matrices **28**

B Downfolding **29**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

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

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

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

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

depends on their distance and relative orientation. This can, however, not be the origin of the magnetism found in actual materials: In a classical system charges cannot flow in thermodynamic equilibrium, the celebrated Bohr-van Leeuwen theorem, and hence there are no magnetic moments to begin with [2].
In quantum mechanics, however, non-vanishing charge currents in the ground state are not uncommon: An electron in state $\Psi(\vec{r})$ corresponds to a current density

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

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

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

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

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

The constant of proportionality between spin and magnetic moment differs from that between orbital momentum and moment by the gyromagnetic ratio g_0. Dirac theory gives $g_e = 2$, which is changed to $g_e \approx 2.0023\ldots$ by QED corrections.
Atomic moments are thus of the order of μ_B. For two such moments at a distance of 1 Å the magnetostatic energy (2) is of the order of 0.05 meV, corresponding to a temperature of less than 1 K. Therefore, magnetic ordering which, e.g., in magnetite (Fe_3O_4), persists till about 860 K, must originate from an interaction other than the magnetostatic interaction of dipoles. Indeed, it is the interplay of electronic properties which are apparently unrelated to magnetism, the Pauli principle in combination with the Coulomb repulsion (Coulomb exchange) as well

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

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

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

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

2 Coulomb exchange

The Coulomb repulsion between electrons,

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

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

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

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

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

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

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

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

and obtain

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

where the direct terms are the Coulomb integral

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

while the cross terms give the exchange integral

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

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

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

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

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

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

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

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

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

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

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

with energy

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

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

$$\begin{aligned} J_{ab} &= \int d^3r_1\, \overline{\Phi(\vec{r}_1)} \underbrace{\int d^3r_2\, \frac{1}{|\vec{r}_1 - \vec{r}_2|}\, \Phi(\vec{r}_2)}_{=(2\pi)^{-3}\int dk\, \Phi(k) e^{ikr_1} 4\pi/k^2} && (18) \\ &= \frac{1}{(2\pi)^3}\int d^3k \underbrace{\int d^3r_1\, e^{i\vec{k}\cdot\vec{r}_1} \overline{\Phi(\vec{r}_1)}}_{=\overline{\Phi(k)}}\, \Phi(\vec{k}\,)\, \frac{4\pi}{k^2} && (19) \\ &= \frac{1}{(2\pi)^3}\int d^3k\, |\Phi(\vec{k}\,)|^2\, \frac{4\pi}{k^2} > 0 && (20) \end{aligned}$$

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

Since H_U only contains interactions within the system of electrons, it commutes with the total orbital momentum $[H_U, \vec{L}_{\rm tot}] = 0$. Obviously it also commutes with the total spin $\vec{S}_{\rm tot}$. The

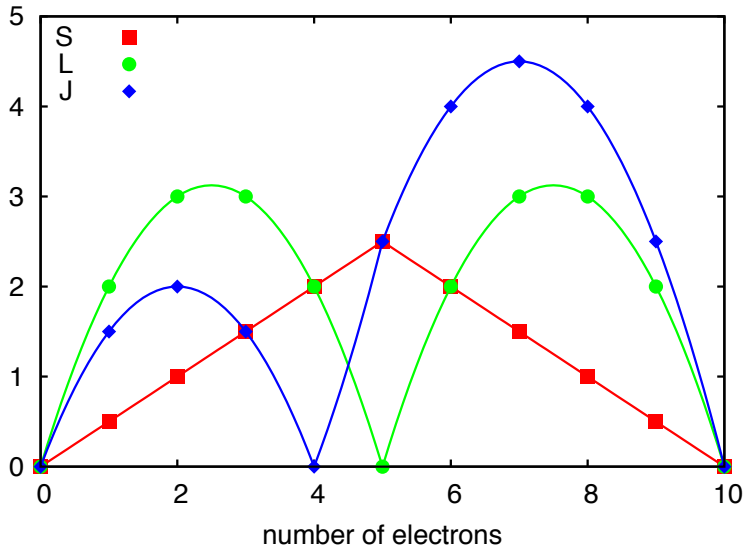


Fig. 1: *Angular momenta of the Hund's rules ground state* $^{2S+1}L_J$ *for* d*-shells.*

eigenstates of $H_0 + H_U$ can thus be classified by their quantum numbers L and S. These terms are written as ^{2S+1}L. For p- and d-shells they are listed in table 1. Hund's rules give the multiplet term with the lowest energy: For a given shell, this lowest state has the largest possible spin (Hund's first rule). If there are several terms of maximum multiplicity, the one with lowest energy has the largest total orbital momentum (Hund's second rule). There is a third Hund's rule, which, however, is not related with the electron-electron repulsion but with spin-orbit coupling: Within L-S coupling H_{SO} splits the atomic orbitals into eigenstates of the total angular momentum $\vec{J} = \vec{L}_{\text{tot}} + \vec{S}_{\text{tot}}$. The multiplets ^{2S+1}L thus split into $^{2S+1}L_J$. The term with the lowest energy is the one with smallest J if the shell is less than half-filled and largest J if it is more than half-filled (Hund's third rule). These rules are illustrated for d-shells in Fig. 2. A more detailed discussion of multiplet effects and the Coulomb interaction in atomic-like systems can be found in [5, 6], calculations of multiplets including spin-orbit coupling can be performed online, at `http://www.cond-mat.de/sims/multiplet`.

s					^{2}**S**									
p^1 or p^5					^{2}**P**									
p^2 or p^4		^1S	^1D						^{3}**P**					
p^3					^2P	^2D					^{4}**S**			
d^1 or d^9					^{2}**D**									
d^2 or d^8		^1S	^1D	^1G					^3P	^{3}**F**				
d^3 or d^7					^2P	$^{2\times}{}^2D$	^2F	^2G	^2H			^4P	^{4}**F**	
d^4 or d^6	$^{2\times}{}^1S$	$^{2\times}{}^1D$	^1F	$^{2\times}{}^1G$	^1I				$^{2\times}{}^3P$	^3D	$^{2\times}{}^3F$	^3G	^3H	^{5}**D**
d^5					^2S	^2P	$^{3\times}{}^2D$	$^{2\times}{}^2F$	$^{2\times}{}^2G$	^2H	^2I	^4P	^4D ^4F ^4G	^{6}**S**

Table 1: *Atomic multiplets for open* s*-,* p*-, and* d*-shells. For terms that appear multiple times the number of distinct terms is indicated. The Hund's rules ground state is indicated in bold.*

3 Kinetic exchange

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

3.1 A toy model

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

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

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

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

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

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

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

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

and two states where the electrons are in the same orbital

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

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

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

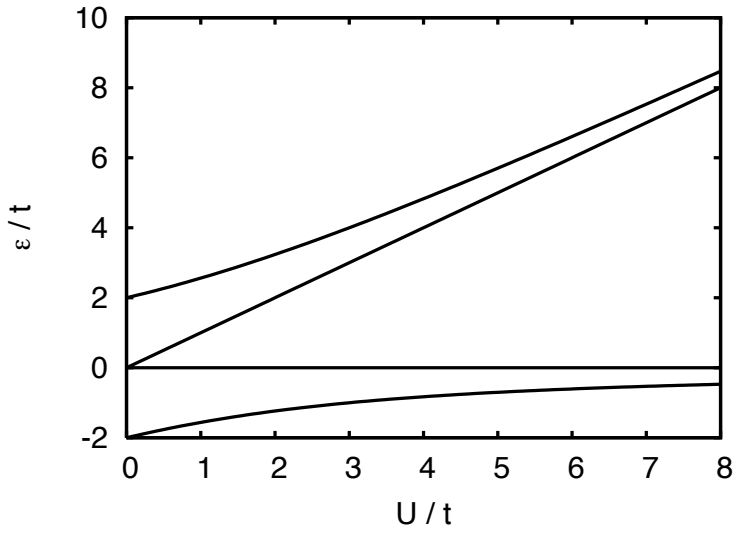


Fig. 2: *Spectrum of the two-site Hubbard model as a function of U. For large U there are two levels with energy close to zero. Their energy difference corresponds to the exchange energy. The remaining two states with ionic character have an energy roughly proportional to U.*

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

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

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

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

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

3.2 Direct exchange

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

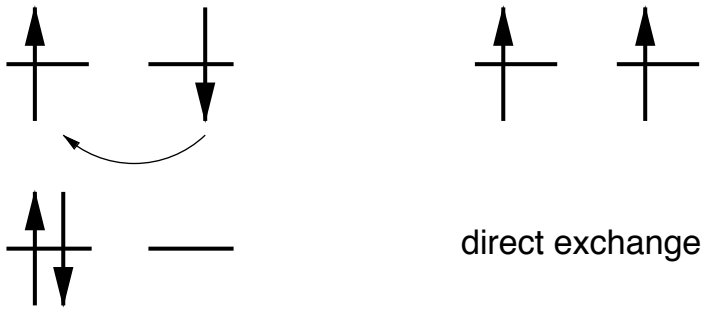


Fig. 3: *Simple picture of direct exchange: The antiparallel alignment of the spins (left) is favored, since it allows the electrons to hop to the neighboring site. For parallel spins (right) hopping is suppressed by the Pauli principle.*

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

An instructive method to analyze the large-U limit, which can readily be generalized to more complex situations, where we can no longer diagonalize the full Hamiltonian, is the downfolding technique. The mathematical background is explained in the appendix. The idea of downfolding is to partition the Hilbert space into parts that are of interest, here the low-energy covalent type states, and states that should be projected out, here the high-energy ionic states. With this partitioning we can view the Hamiltonian matrix (24) as built of 2×2 submatrices. Calculating the inverse on the space of covalent states (see Eqn. (89) in the appendix) we find an *effective Hamiltonian* which now operates on the covalent states only:

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

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

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

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

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

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

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

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

3.3 Second quantization for pedestrians

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

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

for the single-electron states, and

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

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

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

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

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

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

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

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

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

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

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

Looking at equation (82) in the appendix we see that the spin operators are given in second quantization by

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

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

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

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

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

3.4 Mean-field treatment

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

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

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

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

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

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

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

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

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

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

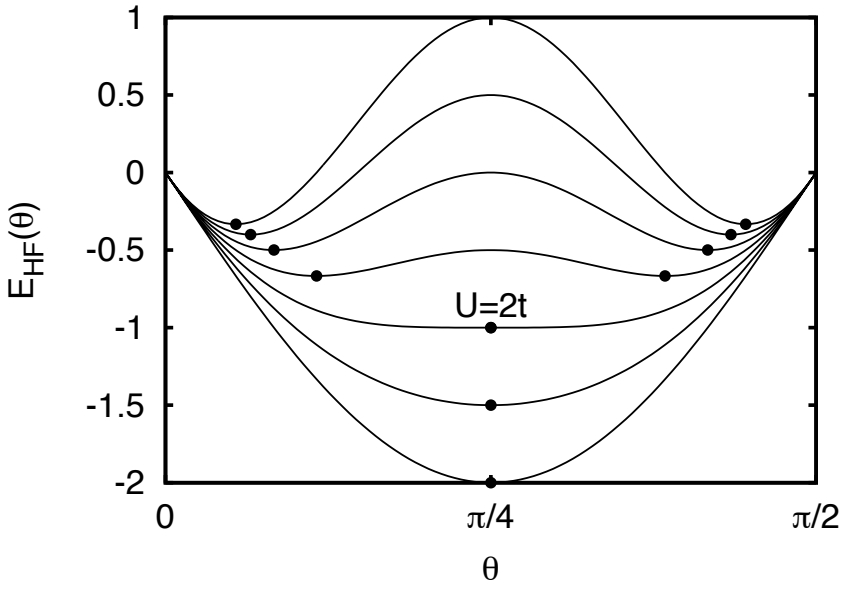


Fig. 4: *Energy expectation value for a Slater determinant $\Psi(\theta, \pi/2-\theta)$ for U=0, t, $2t$, ..., $6t$. When $U \leq 2t$ the minimum is at $\theta = \pi/4$. This is the Hartree-Fock solution with the bonding orbitals φ_+ occupied. For $U \geq 2t$, $\theta = \pi/4$ is still an extremal point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

3.5 Superexchange

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

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

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

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

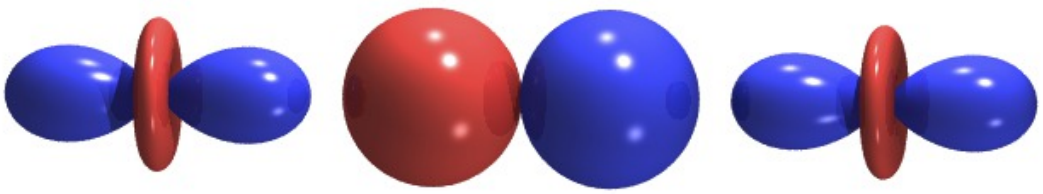


Fig. 5: *In superexchange an oxygen p-orbital mediates the exchange interaction between two transition-metal d-orbitals.*

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

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

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

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

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

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

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

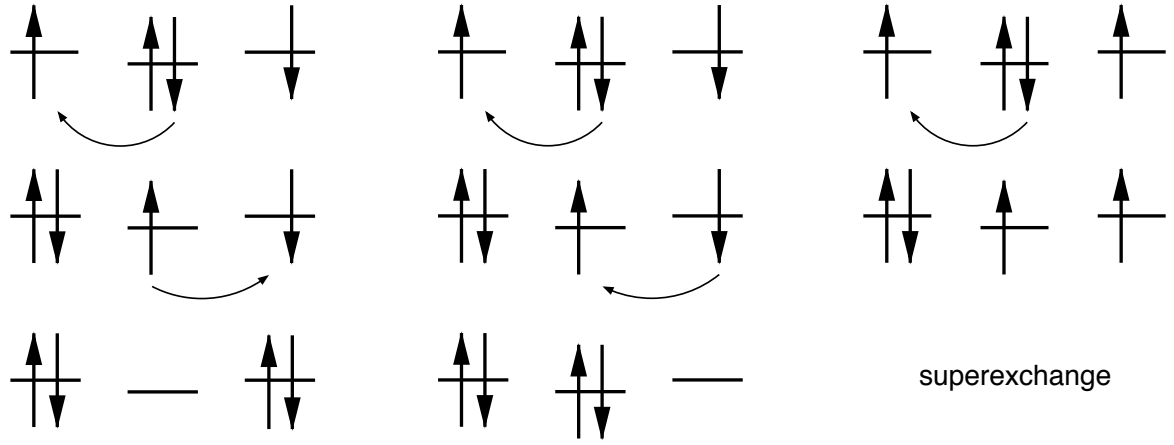


Fig. 6: *Simple picture of superexchange. Here the orbital on the central site is different from the orbitals on the sides. Typically, in the center there is an oxygen p-orbital coupling two d-orbitals. This situation is illustrated in Fig. 5. For antiparallel spins on the d-orbitals there are two ways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

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

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

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

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

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

3.6 Ferromagnetic superexchange

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

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

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

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

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

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

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

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

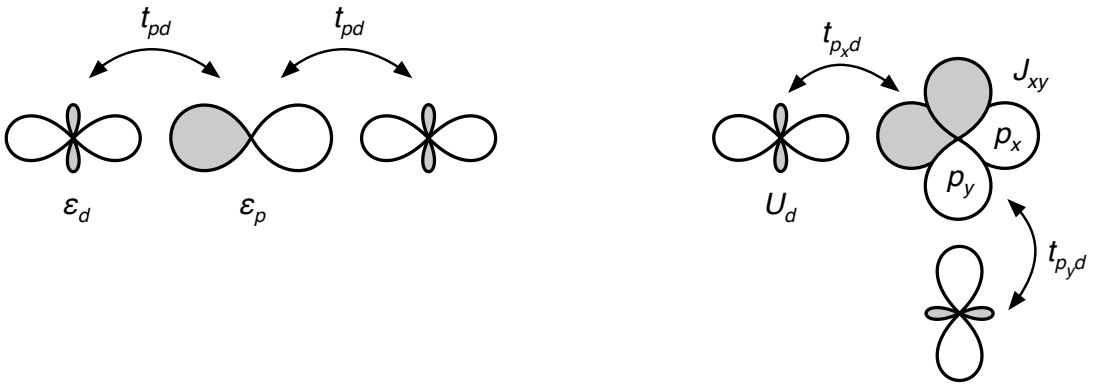


Fig. 7: *Dependence of superexchange on geometry: When the d-orbitals interact via an oxygen in-between (the 180° geometry shown on the left), both d-orbitals couple to the same p-orbital, while the hopping to the two other p-orbitals vanishes by symmetry. The result is antiferromagnetic superexchange.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

giving the effective Hamiltonian

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

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

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

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

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

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

4 Double exchange

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

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

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

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

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

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

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

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

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

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

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

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

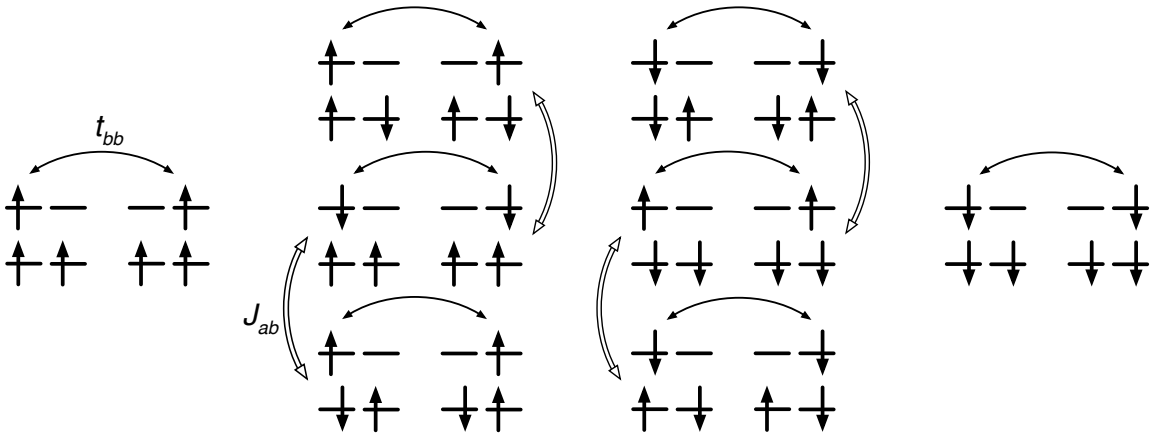


Fig. 8: *Matrix elements entering the double-exchange Hamiltonian. Hopping matrix elements t_{bb} are indicated as double arrows, Coulomb-exchange matrix elements J_{ab} as double lines. Note that the right half of the states are obtained from the left by flipping all spins.*

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

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

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

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

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

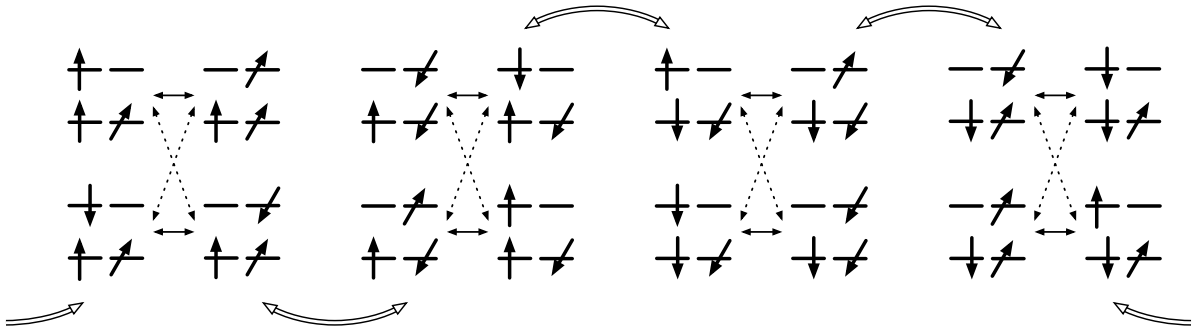


Fig. 9: *With quantization axes tilted between the sites, all states couple. Matrix elements are indicated by arrows: Hopping only couples sites with the same occupation of the a-orbitals. Full lines stand for $t_{bb}\cos(\vartheta/2)$, dotted lines for matrix elements proportional to $t_{bb}\sin(\vartheta/2)$.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

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

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

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

so the hopping becomes

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

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

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

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

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

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

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

5 Orbital-ordering

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

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

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

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

Downfolding to the states without doubly occupied sites, we obtain

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

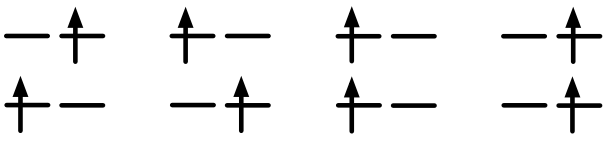
Thus we find that there is an interaction between the states with exchanged orbital-occupation, i.e., an orbital-exchange. For the present case of ferromagnetically aligned spins, the direct orbital exchange coupling favors the orbital singlet, when the hopping matrix elements are of the same sign. In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

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

so that we can write

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

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



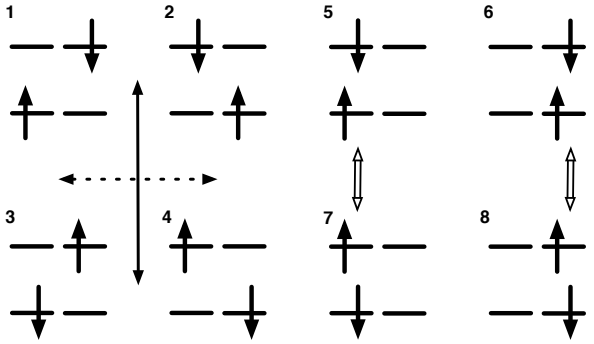


Fig. 11: *Basis states for electrons of opposite spin. The numbering used for the matrix (63) is indicated. Spin exchange is indicated by the full, orbital exchange by the dotted arrow. The states with both electrons on the same site are coupled via Coulomb exchange (double arrows).*

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

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

from which we obtain

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

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

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

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

6 Extended systems

6.1 Hubbard model

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

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

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

6.2 Mott transition

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

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

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

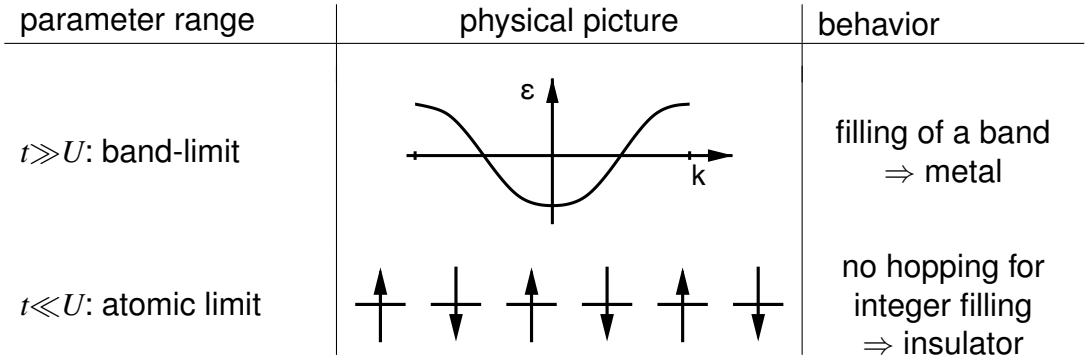


Fig. 12: *Metal-insulator transition for half-filling, i.e., one electron per site.*

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

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

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

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

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

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

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

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

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

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

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

6.3 Heisenberg model

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

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

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

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

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

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

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

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

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

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

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

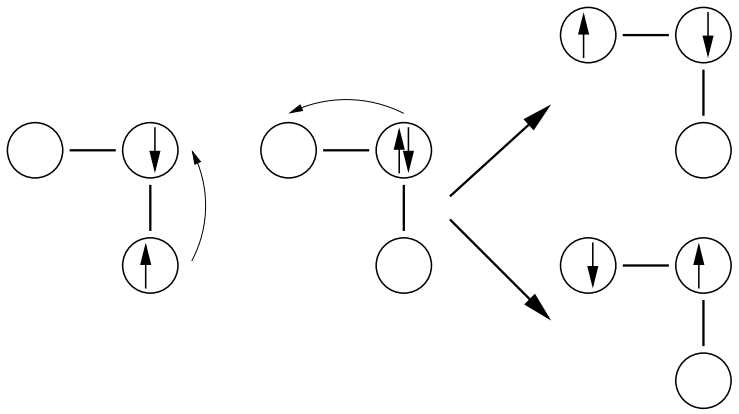


Fig. 13: *Processes contained in the three-site term T': indirect hopping processes to a second-nearest neighbor site with an intermediate (virtual) doubly occupied state. In the first process the two hopping processes are performed by the same electron, in the second process each electron hops once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

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

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

with

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

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

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

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

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

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

7 Conclusion

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

Appendices

A Pauli matrices

Here we collect the most important properties of the Pauli matrices. The Pauli or spin matrices are defined as

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{80}$$

They are hermitian, i.e. $\sigma_i^\dagger = \sigma_i$, and $\sigma_i^2 = 1$. Therefore their eigenvalues are ± 1. The eigenvectors of σ_z are $|m_z\rangle$, $m_z = \pm 1$:

$$|+1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{81}$$

For these vectors we find

$$\sigma_x|\,m_z\rangle = |\,-m_z\rangle \quad \sigma_y|\,m_z\rangle = i m_z|\,-m_z\rangle \quad \sigma_z|\,m_z\rangle = m_z|\,m_z\rangle. \tag{82}$$

The products of the Pauli matrices are $\sigma_x\,\sigma_y = i\sigma_z$, where the indices can be permuted cyclically. From this follows for the commutator

$$[\sigma_x, \sigma_y] = 2i\sigma_z, \tag{83}$$

while the anticommutator vanishes:

$$\{\sigma_x, \sigma_y\} = 0 \tag{84}$$

Finally a rotation by an angle φ about the axis $\hat{n}$ changes the spin matrices

$$R_{\hat{n}}(\varphi) = e^{-i\hat{n}\cdot\vec{\sigma}\,\varphi/2} = \cos(\varphi/2) - i\sin(\varphi/2)\,\hat{n}\cdot\vec{\sigma}\,. \tag{85}$$

B Downfolding

To integrate-out high-energy degrees of freedom, we partition the Hilbert space of the full system into states of interest (low-energy states) and 'other' states, which will be integrated out. The Hamiltonian is then written in blocks

$$H = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix}, \tag{86}$$

where H_{00} is the Hamiltonian restricted to the states of interest (reduced Hilbert space), H_{11} the Hamiltonian for the 'other' states, and the T matrices describe transitions between the two subspaces. The resolvent is partitioned likewise

$$G(\omega) = (\omega - H)^{-1} = \begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix}^{-1}. \tag{87}$$

Its elements are easily determined by solving the system of two linear matrix equations

$$\begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix} \begin{pmatrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{pmatrix}, \tag{88}$$

keeping track of the order of the sub-matrix products. The resolvent on the reduced Hilbert space is thus given by

$$G_{00}(\omega) = \left(\omega - [\underbrace{H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10}}_{=H_{\text{eff}}(\omega)}] \right)^{-1}. \tag{89}$$

This expression looks just like the resolvent for a Hamiltonian H_{eff} on the reduced Hilbert space. This effective Hamiltonian describes the physics of the full system, but operates only on the small reduced Hilbert space: For an eigenvector $H|\Psi\rangle = E|\Psi\rangle$ on the full Hilbert space

$$H|\Psi\rangle = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix} \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} = E \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} \tag{90}$$

its projection $|\Psi_0\rangle$ onto the reduced Hilbert space is an eigenstate of $H_{\text{eff}}(E)$. On the other hand, we can construct the full eigenstate from a solution $H_{\text{eff}}(E)|\Psi_0\rangle = E|\Psi_0\rangle$ on the reduced Hilbert space by upfolding $|\Psi\rangle \propto (\mathbb{1} + (E - H_{11})^{-1} T_{10})|\Psi_0\rangle$.

Of course, this drastic simplification comes at a price: the effective Hamiltonian is energy dependent. If the hopping matrix elements in T_{01} are small, and/or if the states in the part of the Hilbert space that has been integrated out are energetically well-separated from the states that are explicitly considered, this energy dependence can, to a good approximation, be neglected. We can then replace ω by some characteristic energy ε_0 for the states in the reduced Hilbert space to obtain an energy-independent Hamiltonian

$$H_{\text{eff}}(\omega) = H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10} \approx H_{00} + T_{01}(\varepsilon_0 - H_{11})^{-1} T_{10} = H_{\text{eff}}(\varepsilon_0) \tag{91}$$

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] R. Peierls: *Surprises in Theoretical Physics* (Princeton University Press, 1979)

[2] J.H. van Vleck: *Quantum Mechanics: The Key to Understanding Magnetism*, in S. Lundqvist (ed.): *Nobel Lectures, Physics 1971-1980* (World Scientific, Singapore, 1992) http://www.nobelprize.org/nobel_prizes/physics/laureates/1977

[3] Patrik Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)

[4] A. Auerbach: *Interacting Electrons and Quantum Magnetism* (Springer, Heidelberg, 1994)

[5] R. Eder: *Multiplets in Transition-Metal Ions and Introduction to Multiband Hubbard Models*, in [18]

[6] M. Weissbluth: *Atoms and Molecules* (Academic Press, San Diego, 1978)

[7] P.W. Anderson: *Theory of Magnetic Exchange Interactions*, in F. Seitz and D. Turnbull (eds.): *Solid State Physics* **14**, 99 (1963)

[8] D.C. Mattis: *The Theory of Magnetism Made Simple* (World Scientific, Singapore, 2006)

[9] S. Blundell: *Magnetism in Condensed Matter* (Oxford University Press, 2001)

[10] E. Koch: *Many-Electron States*, in [19]

[11] A. Szabo and N.S. Ostlund: *Modern Quantum Chemistry* (Dover Publications, 1996)

[12] J.B. Goodenough: *Goodenough-Kanamori rule*, Scholarpedia **3**, 7382 (2008) http://scholarpedia.org/article/Goodenough-Kanamori_rule

[13] K.I. Kugel and D.I. Khomskii, Zh. Eksp. Teor. Fiz. **64**, 1429 (1973) [Sov. Phys. JETP **37**, 725 (1973)]

[14] A. Montorsi (Ed.): *The Hubbard Model*, (World Scientific, Singapore, 1992)

[15] F. Gebhard: *The Mott Metal-Insulator Transition* (Springer, Heidelberg, 1997)

[16] A. Messiah: *Méchanique Quantique* (Dunond, Paris, 1964)

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

[18] E. Pavarini, E. Koch, J. van den Brink, and G. Sawatzky:
Quantum Materials: Experiments and Theory
Reihe Modeling and Simulation, Vol. 6 (Forschungszentrum Jülich, 2016)
http://www.cond-mat.de/events/correl16

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

6 Spin-Orbital Entanglement in Mott Insulators

Andrzej M. Oleś
Institute of Theoretical Physics
Jagiellonian University
Prof. Stanisława Łojasiewicza 11, Kraków, Poland

Contents

1 Entangled superexchange: SU(2) ⊗ SU(2) model **2**
1.1 Spin-orbital Hilbert space in a Mott insulator 2
1.2 Modifications due to finite spin-orbit coupling λ 3

2 Orbital physics for partly filled e_g orbitals **5**

3 Coulomb interactions in spin-orbital Hilbert space **9**
3.1 Kanamori parameters: Coulomb U and Hund's exchange J 9
3.2 Goodenough-Kanamori rules 14

4 Kugel-Khomskii model for Mott insulators **16**
4.1 Kugel-Khomskii model: 3D for $KCuF_3$ and 2D for K_2CuF_4 16
4.2 Entanglement in the ferromagnetic excitations of K_2CuF_4 20
4.3 Weak spin-orbital entanglement for large spins $S{=}2$ in $LaMnO_3$ 22

5 Spin-orbital entanglement in t_{2g} electron models **24**
5.1 Entangled phases of $LaVO_3$ and YVO_3 24
5.2 Spin-orbital entanglement on a triangular lattice 27

6 Experimental consequences of spin-orbital entanglement **30**

7 Summary **31**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Entangled superexchange: SU(2) ⊗ SU(2) model

1.1 Spin-orbital Hilbert space in a Mott insulator

At large on-site Coulomb repulsion U, electrons in a transition metal oxide localize and have no kinetic energy. The new state of electronic matter which emerges under strong Coulomb repulsion is a Mott insulator. Then the electron state is given by the spin component and the orbital occupied by this electron. It was one of the great achievements of Kugel and Khomskii [1] to realize that in the case of two orbitals available at each site i, the Hilbert space of a Mott insulator is spanned by spin-orbital states, i.e., it suffices to specify a spin and a pseudospin (orbital) component of each electron to define its quantum state at site i. Such localized electrons in the absence of kinetic energy interact by superexchange [2, 3].

To illustrate these concepts, we begin with a study of a one-dimensional (1D) spin-orbital superexchange model $\mathcal{H}_{\rm SE}$ defined in a Mott insulator with on-site repulsion U by the spin-orbital Hilbert space spanned by the eigenstates $\{|\uparrow\rangle, |\downarrow\rangle\}$, of spin $S = 1/2$, and orbital (pseudospin) operator $T = 1/2$, with the eigenstates $\{|+\rangle, |-\rangle\}$. Such states at two neighboring sites i and $i+1$ are coupled by 1D spin-orbital ('Kugel-Khomskii') superexchange [4–6],

$$\mathcal{H}_{\rm SE} = J \sum_i \left[\left(\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \alpha \right) \left(\mathbf{T}_i \cdot \mathbf{T}_{i+1} + \beta \right) - \alpha\beta + \varepsilon_z \sum_i \tau_i^{(c)} \right], \tag{1}$$

where $\tau_i^{(c)} = T_i^{(c)} = \sigma_i^z/2$, and we take the orbital splitting $J\varepsilon_z = E_z = 0$. Here a constant term $\alpha\beta$ is eliminated and the Hamiltonian $\mathcal{H}_{\rm SE}$ includes only operator terms. The 1D kinetic energy is given by the orbital-flavor (α=+, −) conserving hopping $\propto t$, and the interaction energy is given by either $U n_{i\alpha\uparrow} n_{i\alpha\downarrow}$ or $U n_{i\alpha\sigma} n_{i\alpha\bar{\sigma}}$, which both cost the Coulomb repulsion energy U. Of particular interest is the strongly correlated regime $U \gg t$, where electrons localize and interact by antiferromagnetic (AF) superexchange [2],

$$J = \frac{4t^2}{U}. \tag{2}$$

For two degenerate orbitals one needs to introduce a doubly-degenerate Hubbard model [7]. One finds then again the same exchange constant J (2) as in the derivation of the t-J model from the Hubbard model in the limit $U \gg t$ [8].

It was a great achievement to realize that spin and orbital states are entangled and are parts of the same Hilbert space [9, 10]. Thus the superexchange $\mathcal{H}_{\rm SE}$ in Eq. (1) is not just a scalar product of two involved subspaces, spin and orbital, but describes joint quantum fluctuations of these two operators [9–12]. The superexchange model (1) depends on two parameters $\{\alpha, \beta\}$, and they decide about the type of order. It describes a competition between four spin-orbital phases, where the order in each sector can be either ferro- or antiferro-. The phases where quantum fluctuations exist in the ground state only in at most one sector (spin or orbital) are disentangled, as the phases I-III, see the phase diagram in Fig. 1. Otherwise, we recognize one entangled phase IV-VI which has three different regimes. Here spin-orbital entanglement increases when the quantum critical point (QCP) $(-1/4, -1/4)$ is approached along the diagonal

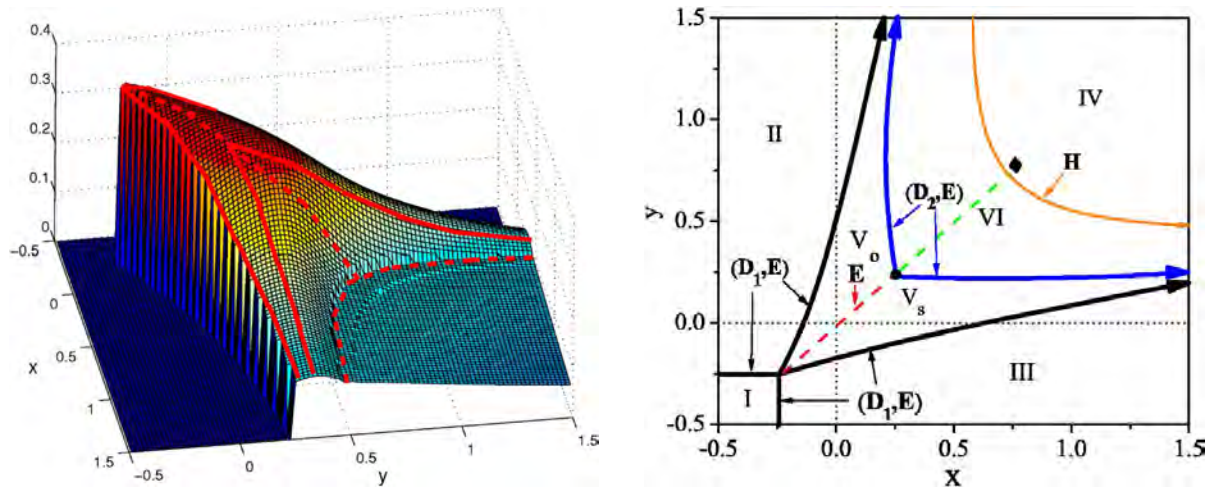


Fig. 1: *Spin-orbital entanglement in the 1D SU(2) ⊗ SU(2) model (1) at $E_z = 0$. Left—The von Neumann entropy per site $\mathcal{S}_{vN}/L$ (3) for the ground state at $L = 8$ as a function of x and y. The phase boundaries (solid and dashed lines) are drawn to guide the eye. Right—Phase diagram of a coupled 1D spin-orbital chain. The diamond point is located at $(3/4, 3/4)$. Quantum phases are distinguished by entanglement: I, II, and III are disentangled, IV is weakly, and V& VI stronger entangled. The parameters (x, y) are the same as (α, β) in Fig. 2. Images after Ref. [6].*

line $x = y$. At the QCP itself, spin-orbital entanglement is maximal within the phase V, changes to a plateau in IV, and next drops towards zero in IV beyond the QCP ($\alpha = \beta < -1/4$).

A standard measure of entanglement between two subsystems A and B in the ground state $|\mathrm{GS}\rangle$ of a system of size L is the von Neumann entropy [13]: $\mathcal{S}_{\mathrm{vN}} = -\mathrm{Tr}_A\{\rho_A \ln \rho_A\}/L$, see Fig. 1. Here our two subsystems are spins and orbitals and the entanglement concerns the entire system (in other applications the system would frequently be separated into A and B by cutting one bond). The von Neumann entropy is obtained by integrating the density matrix, $\rho_A = \mathrm{Tr}_B|\mathrm{GS}\rangle\langle\mathrm{GS}|$ over subsystem B. Consequently, we use here the following definition of the von Neumann spin-orbital entanglement entropy:

$$\mathcal{S}_{\mathrm{vN}} = -\frac{1}{L}\mathrm{Tr}_S\{\rho_S \ln \rho_S\}, \tag{3}$$

where

$$\rho_S = \mathrm{Tr}_T|\mathrm{GS}\rangle\langle\mathrm{GS}| \tag{4}$$

is the reduced spin-only density matrix (4), with the orbital $\{T\}$ degrees of freedom being integrated out.

1.2 Modifications due to finite spin-orbit coupling λ

Spin and orbital operators may also couple directly on-site via the spin-orbit interaction [14]. It is in general quantum but we present the Ising coupling here for more clarity. Then the 1D model Hamiltonian consists of two qualitatively distinct terms [15],

$$\mathcal{H} = \mathcal{H}_{\mathrm{SE}} + \mathcal{H}_{\mathrm{SOC}}, \tag{5}$$

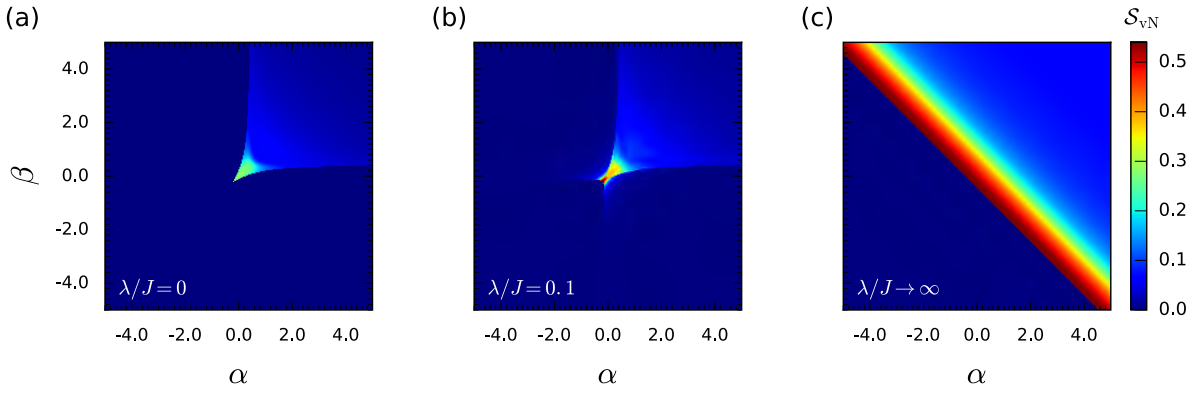


Fig. 2: *The von Neumann spin-orbital entanglement entropy, $\mathcal{S}_{\rm vN}$ (3), calculated using ED on an L=12-site periodic chain for the spin-orbital model Eq. (5) and for the increasing value of the spin-orbit coupling λ [15]: (a) $\lambda/J=0$, (b) $\lambda/J=0.1$, and (c) $\lambda/J\to\infty$.*
Image by courtesy of Dorota Gotfryd.

and includes next to the superexchange $\mathcal{H}_{\rm SE}$ (1), the spin-orbit Ising interaction $\propto\lambda$. The model Hamiltonian (5) depends then on three parameters $\{\alpha,\beta,\lambda\}$. Altogether, our choice means that the spin-orbital exchange interaction has the simplest possible form that can, nevertheless, simulate a realistic situation found in the transition metal oxides. We note that the spin-orbital exchange (1) would have in general a more complex form. For instance, this would be the case if, e.g. three instead of two active orbitals were taken into account, and the corrections from finite Hund's exchange were included (as relevant for the $5d$ iridates).
The second term in Eq. (5) stands for the on-site spin-orbit coupling (SOC) and reads,

$$\mathcal{H}_{\rm SOC}=2\lambda\sum_i S_i^z T_i^z. \tag{6}$$

Here the parameter λ measures the strength of the on-site spin-orbit coupling (of relativistic origin). The above Ising form of the spin-orbit coupling was chosen as the simplest possible and yet nontrivial term. Moreover, exactly such a form of the spin-orbit coupling is *typically* realized in systems with *two* active orbitals. This is the case, for instance, of the active t_{2g} doublets in YVO_3 [16, 17] and Sr_2VO_4 [18], or in optical lattices [19]. In fact, such a highly anisotropic form of spin-orbit coupling is valid for any system with an active orbital doublet, either two directional p (p_x and p_y) or two planar t_{2g} (xz and yz) orbitals.
The line $\beta{=}{-}\alpha$ plays a special role in the phase diagram of Fig. 1. In order to better understand the physical consequences of increasing λ, we display the onset of the spin-orbital entanglement once $\beta{=}{-}\alpha$. As shown in Fig. 2 for increasing λ, the region of finite spin-orbital entanglement increases dramatically and includes both previously disentangled phases, II and III [15]. In fact, the largest entanglement is found in the vicinity of the line $\beta{=}{-}\alpha$, when $\alpha{+}\beta>-1/4$. At $\alpha{+}\beta=-1/4$ the spin-orbital entanglement entropy jumps from $\mathcal{S}_{\rm vN}=0$ to a maximal value and that happens at the QCP. Thus, the qualitative result of Fig. 1 breaks down. We conclude that finite spin-orbit coupling transfers on-site entanglement to on-bond entanglement in the phases antiferromagnetic/ferro-orbital (AF/FO, phase II) and ferromagnetic/alternating-orbital (FM/AO, phase III) which are initially disentangled (at $\lambda{=}0$, see Fig. 3.)

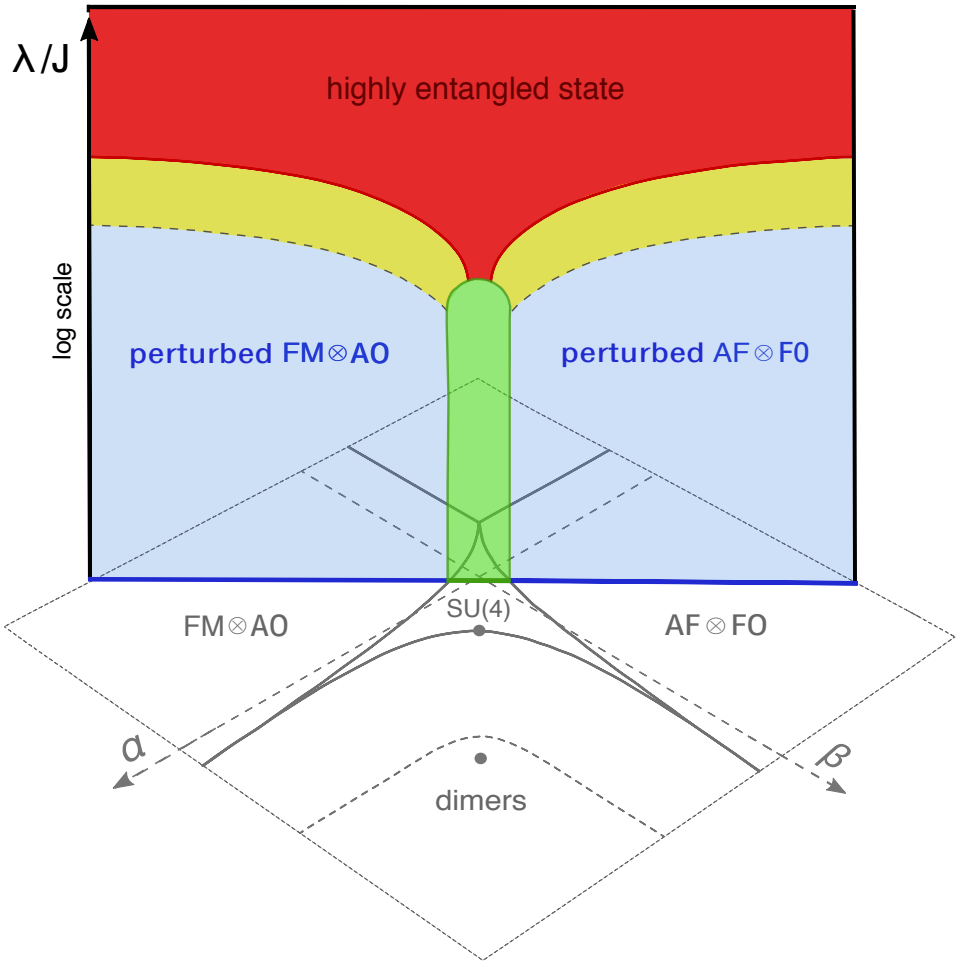


Fig. 3: *Schematic quantum phase diagram of Hamiltonian (1) in the* (α, β) *plane, see Fig. 1. The colorful vertical plane shows how spin-orbital entanglement extends to the highly entangled state, with on-bond entanglement in two disentangled phases: FM⊗AO and AF⊗FO, see Fig. 2. Increasing spin-orbit coupling* λ *generates on-bond entangled states (the green region marks entangled states; note that the vertical scale is logarithmic). Image reproduced from Ref. [15].*

2 Orbital physics for partly filled e_g orbitals

It is important to realize that modeling of transition-metal oxides [3] can be performed on different levels of sophistication. We shall present here some effective orbital-only and spin-orbital superexchange models for correlated $3d$ orbitals. In a perovskite lattice they are coupled by hopping t between nearest neighbor ions, while the hopping to more distant neighbors and for other lattices may be generated using the general rules formulated by Slater and Koster [20]. The orbitals have particular shapes and belong to two irreducible representations of the O_h cubic point group:
(i) a two-dimensional (2D) representation of e_g-orbitals $\{|3z^2{-}r^2\rangle/\sqrt{6},\ |x^2{-}y^2\rangle/\sqrt{2}\}$, and
(ii) a three-dimensional (3D) representation of t_{2g}-orbitals $\{|xy\rangle, |yz\rangle, |zx\rangle\}$.
In the absence of any tetragonal distortion or crystal-field (CF) due to surrounding oxygens, the $3d$-orbitals are degenerate within each irreducible representation of the O_h point group and

have typically a large splitting $\propto 10D_q \approx 2.0$ eV between them. When some of such degenerate orbitals are partly filled, electrons (or holes) have both spin and orbital degree of freedom. The kinetic energy H_t in a perovskite follows from the hybridization between $3d$- and $2p$-orbitals. In an effective d-orbital model, the oxygen $2p$-orbitals are not included explicitly and we define the largest hopping element t obtained between two orbitals of the same type, which both belong to the nearest neighbor $3d$ ions in a lattice.

We begin with conceptually simpler t_{2g} orbitals where finite hopping t results from the d-p hybridization along π-bonds and each element couples a pair of identical orbitals active along a given bond. Each t_{2g} orbital is active along two cubic axes, while the hopping is forbidden along the axis perpendicular to the plane of this orbital, e.g. the hopping between two xy-orbitals vanishes along the c axis (due to the cancellations caused by orbital phases). It is therefore convenient to introduce the following short-hand notation for the orbital degree of freedom [12],

$$|a\rangle \equiv |yz\rangle, \qquad |b\rangle \equiv |zx\rangle, \qquad |c\rangle \equiv |xy\rangle. \tag{7}$$

The labels $\gamma = a, b, c$ refer here to the cubic axes where the hopping is absent between two orbitals of a given type,

$$H_t(t_{2g}) = -t \sum_{\alpha} \sum_{\langle ij\rangle \| \gamma \neq \alpha} a^{\dagger}_{i\alpha\sigma} a_{j\alpha\sigma}, \tag{8}$$

Here $a^{\dagger}_{i\alpha\sigma}$ is an electron creation operator in a t_{2g}-orbital $\alpha \in \{yz, zx, xy\}$ with spin $\sigma = \uparrow, \downarrow$ at site i, and the local electron density operator for a spin-orbital state is $n_{i\alpha\sigma} = a^{\dagger}_{i\alpha\sigma} a_{i\alpha\sigma}$. For t_{2g} electrons not only spin but also orbital flavor is conserved in each hopping process $\propto t$.

The hopping Hamiltonian for e_g electrons concerns σ-bands and couples here two directional e_g-orbitals $\{|i\zeta_\gamma\rangle, |i\zeta_\gamma\rangle\}$ along a bond $\langle ij\rangle \parallel \gamma$ (we use again the same notation t) [21],

$$H_t(e_g) = -t \sum_{\alpha} \sum_{\langle ij\rangle \| \alpha, \sigma} a^{\dagger}_{i\zeta_\alpha\sigma} a_{j\zeta_\alpha\sigma}. \tag{9}$$

Indeed, a hopping with amplitude $-t$ between two sites i and j occurs only when an electron with spin σ transfers between the two directional orbitals $|\zeta_\gamma\rangle$ oriented along the bond $\langle ij\rangle$ direction, i.e., $|\zeta_\gamma\rangle \propto |3x^2{-}r^2\rangle$, $|3y^2{-}r^2\rangle$, or $|3z^2{-}r^2\rangle$, along the cubic axis $\gamma = \{a, b, c\}$. We will similarly denote by $|\xi_\gamma\rangle$ an orthogonal orbital to $|\zeta_\gamma\rangle$. It is perpendicular to the bond $\langle ij\rangle$ direction, i.e., $|\xi_\gamma\rangle \propto |y^2{-}z^2\rangle$, $|z^2{-}x^2\rangle$, and $|x^2{-}y^2\rangle$ along the cubic axis $\gamma \in \{a, b, c\}$, and $\langle \xi_\gamma | \zeta_\gamma \rangle = 0$. For the moment we consider only electrons with one spin component, $\sigma = \uparrow$, to focus on the orbital problem. While such a choice of an over-complete basis $\{\zeta_a, \zeta_b, \zeta_c\}$ is convenient, for writing down the kinetic energy a particular orthogonal basis is needed.

The usual choice is to take

$$|z\rangle \equiv \tfrac{1}{\sqrt{6}}|3z^2{-}r^2\rangle, \qquad |\bar{z}\rangle \equiv \tfrac{1}{\sqrt{2}}|x^2{-}y^2\rangle, \tag{10}$$

i.e., the basis of *real* e_g *orbitals* [21]. However, this basis is the natural one only for the bonds parallel to the c-axis but not for those within the (a, b) plane, and for $\uparrow$-spin electrons the hopping reads (here for clarity we omit spin index σ),

$$H^{\uparrow}_t(e_g) = -\frac{1}{4}t \sum_{\langle mn\rangle \| ab} \left[3a^{\dagger}_{i\bar{z}} a_{j\bar{z}} + a^{\dagger}_{iz} a_{jz} \mp \sqrt{3} \left(a^{\dagger}_{i\bar{z}} a_{jz} + a^{\dagger}_{iz} a_{j\bar{z}} \right) \right] - t \sum_{\langle ij\rangle \| c} a^{\dagger}_{iz} a_{jz}. \tag{11}$$

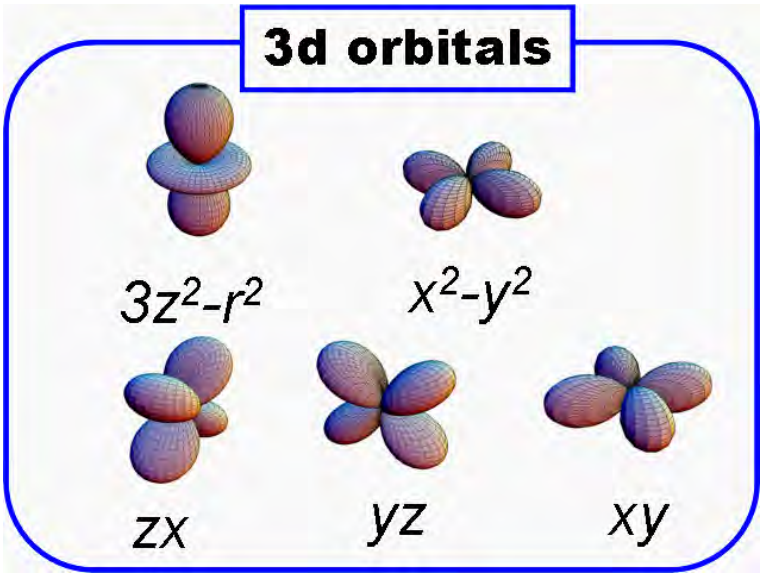


Fig. 4: *Schematic representation of* $3d$ *orbitals: Top—*e_g *orbital basis* $\{|3z^2{-}r^2\rangle, |x^2{-}y^2\rangle\}$. *Bottom—three* t_{2g} *orbital states* $\{|zx\rangle, |yz\rangle, |xy\rangle\} \equiv \{|b\rangle, |a\rangle, |c\rangle\}$. *These representations are split in a regular octahedron [3]. Image by courtesy of Yoshinori Tokura.*

Although this expression is of course cubic invariant, it does not manifest this symmetry but takes a very different appearance depending on the bond direction.
However, the symmetry is better visible using the basis of *complex* e_g *orbitals* at each site i [21],

$$|i{+}\rangle = \tfrac{1}{\sqrt{2}}\big(|iz\rangle - \mathrm{i}\,|i\bar{z}\rangle\big), \qquad |i{-}\rangle = \tfrac{1}{\sqrt{2}}\big(|iz\rangle + \mathrm{i}\,|i\bar{z}\rangle\big), \tag{12}$$

standing for "up" and"down" pseudospin flavors, with the local pseudospin operators being defined as follows,

$$\tau_i^{+} \equiv c_{i+}^{\dagger}c_{i-}, \qquad \tau_i^{-} \equiv c_{i-}^{\dagger}c_{i+}, \qquad \tau_i^{z} \equiv \tfrac{1}{2}(c_{i+}^{\dagger}c_{i+} - c_{i-}^{\dagger}c_{i-}) = \tfrac{1}{2}(n_{i+}{-}n_{i-}). \tag{13}$$

The three directional $\{|i\zeta_\gamma\rangle\}$ and three planar $\{|i\xi_\gamma\rangle\}$ orbitals at site i, associated with the three cubic axes ($\gamma \in \{a, b, c\}$), are the real orbitals,

$$|i\zeta_\gamma\rangle = \tfrac{1}{\sqrt{2}}\left[e^{-\mathrm{i}\vartheta_\alpha/2}|i{+}\rangle + e^{+\mathrm{i}\vartheta_\alpha/2}|i{-}\rangle\right] = \cos(\vartheta_\alpha/2)|iz\rangle - \sin(\vartheta_\alpha/2)|i\bar{z}\rangle, \tag{14}$$

$$|i\xi_\gamma\rangle = \tfrac{1}{\sqrt{2}}\left[e^{-\mathrm{i}\vartheta_\alpha/2}|i{+}\rangle - e^{+\mathrm{i}\vartheta_\alpha/2}|i{-}\rangle\right] = \sin(\vartheta_\alpha/2)|iz\rangle + \cos(\vartheta_\alpha/2)|i\bar{z}\rangle, \tag{15}$$

with the phase factors $\vartheta_{ia} = -4\pi/3$, $\vartheta_{ib} = +4\pi/3$, and $\vartheta_{ic} = 0$, and thus correspond to the pseudospin lying in the equatorial plane and pointing in one of the three equilateral "cubic" directions defined by the angles $\{\vartheta_{i\alpha}\}$.
Using the above complex-orbital representation (12), we can write the *orbital Hubbard model* for e_g electrons with only one spin flavor $\sigma =\uparrow$ in a form similar to the spin Hubbard model,

$$\mathcal{H}_{e_g}^{\uparrow} = -\frac{t}{2}\sum_{\gamma}\sum_{\langle ij\rangle\|\gamma}\Big[\Big(a_{i+}^{\dagger}a_{j+}{+}a_{i-}^{\dagger}a_{j-}\Big){+}\gamma\Big(e^{-\mathrm{i}\chi_\gamma}a_{i+}^{\dagger}a_{j-}{+}e^{+\mathrm{i}\chi_\gamma}a_{i-}^{\dagger}a_{j+}\Big)\Big]{+}\bar{U}\sum_{m}n_{i+}n_{i-}, \tag{16}$$

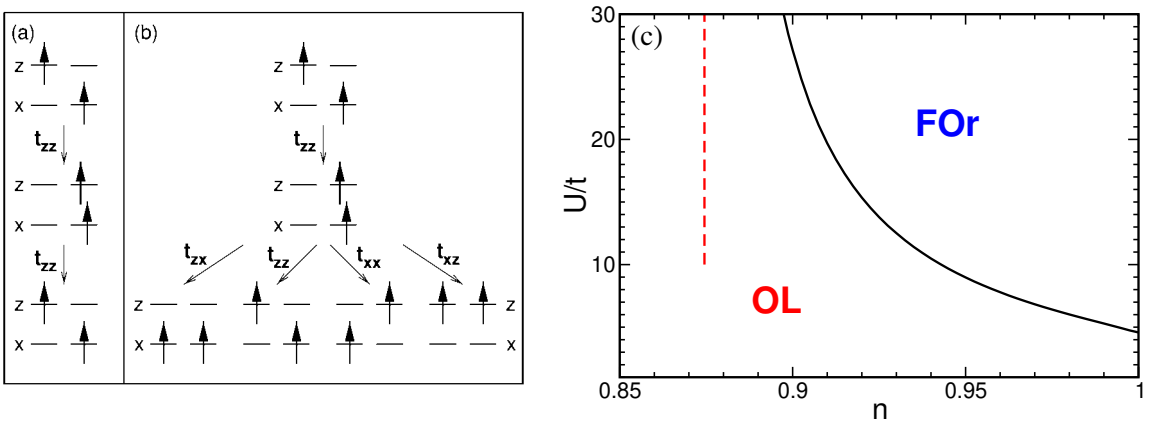


Fig. 5: *Virtual charge excitations leading to the e_g-orbital superexchange model for a strongly correlated system with $|z\rangle$ and $|x\rangle \equiv |\bar{z}\rangle$ real e_g orbitals (10) in the subspace of $\uparrow$-spin states: (a) for a bond along the c axis $\langle ij\rangle \parallel c$; (b) for a bond in the ab plane $\langle ij\rangle \parallel ab$. In a FM plane of $KCuF_3$ ($LaMnO_3$) the superexchange favors AO state of $|\mathrm{AO}\pm\rangle$ orbitals (not shown). (c) The transition from FOr to OL found at $d=\infty$ at finite U, and at $U=\infty$ (dashed line). Images (a-b) are reproduced from Ref. [22]; image (c) is reproduced from Ref. [23].*

with $\chi_a = +2\pi/3$, $\chi_b = -2\pi/3$, and $\chi_c = 0$, and where the parameter γ, explained below, takes for e_g orbitals the value $\gamma = 1$. The appearance of the phase factors $e^{\pm i\chi_\gamma}$ is characteristic of the orbital problem—these factors occur because the orbitals have an actual shape in real space so that each hopping process depends on the bond direction and may change the orbital flavor. The inter-orbital Coulomb interaction $\propto \bar{U}$ [22] is then the only Coulomb term which couples the electron densities in two basis orbitals $n_{i\mu} = a^\dagger_{i\mu}a_{i\mu}$, with $\mu \in \{+,-\}$; its form in invariant under any local basis transformation to a pair of orthogonal orbitals, i.e., it gives energy $\bar{U}$ for a double occupancy, either when two real orbitals are simultaneously occupied $\bar{U}n_{iz}n_{i\bar{z}}$, or when two complex orbitals are both occupied, $\bar{U}\sum_i n_{i+}n_{i-}$.

A charge excitation between two transition metal ions with partly filled e_g-orbitals will arise by a hopping process between two active orbitals, $|i\zeta_\gamma\rangle$ and $|j\zeta_\gamma\rangle$. To capture such processes we introduce two projection operators on the orbital states for each bond,

$$\mathcal{P}^{(\gamma)}_{\langle ij\rangle} \equiv \left(\frac{1}{2}+\tau^{(\gamma)}_i\right)\left(\frac{1}{2}-\tau^{(\gamma)}_j\right)+\left(\frac{1}{2}-\tau^{(\gamma)}_i\right)\left(\frac{1}{2}+\tau^{(\gamma)}_j\right), \tag{17}$$

$$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle} \equiv 2\left(\frac{1}{2}-\tau^{(\gamma)}_i\right)\left(\frac{1}{2}-\tau^{(\gamma)}_j\right). \tag{18}$$

Recently a generalization of the e_g-orbital Hubbard model (16) was proposed to $d=\infty$ dimension [23]. Since the work of Metzner and Vollhardt [24] appeared, it is well known that the limit of $d=\infty$ is simpler for the correlation problems than any finite dimension as the diagrams addressing the correlations collapse to a single point and the Gutzwiller approximation to the variational ground state wave function [25] becomes exact [26]. The e_g orbital Hubbard model describes spinless fermions which propagate on a lattice and have two degenerate orbitals. Any double occupancy costs the same energy $\bar{U}$, exactly as in Eq. (16). A crucial observation is now that any orbital polarized state has no double occupancies, while the orbital liquid (OL) state has double occupancies and has to be renormalized.

The orbital Hubbard model (16) suggests that additional kinetic energy arises from the flavor-nonconserving hopping $\propto \gamma t$. Indeed, the only stable phase in the 3D e_g-orbital model is the orbital liquid (OL) phase [21]. In contrast, for the e_g-orbital model in $d = \infty$ dimensions [23], the OL dominates for most but not for all electron fillings. Indeed, close to half-filling $n = 1$, a FO phase is more stable. This phase has real e_g orbitals and is labeled as FOr in Fig. 5(c). Qualitatively this result is similar to the Nagaoka's theorem [27] for the spin Hubbard model, where a FM state is found close to half-filling. However, the mechanism is qualitatively different as the orbital-nonconserving hopping contributes and destabilizes the OL.
The resulting phase diagram of the e_g-orbital Hubbard model in the $(n, \bar{U})$ plane obtained in the Gutzwiller approximation contains mostly the OL phase, see Fig. 5(c). Here the FOr phase is more stable than the OL phase for $\bar{U} > U_c(n)$ if $n > n_c$. One finds the critical value $n_c = 0.8746$ of the electron density at which the energies of the OL and FOr are equal at $\bar{U} = \infty$, and below which the OL phase is therefore always stable.

3 Coulomb interactions in spin-orbital Hilbert space

3.1 Kanamori parameters: Coulomb U and Hund's exchange J

The full spin-orbital problem involves both degrees of freedom, as in Sec. 1. But in contrast to the simplified case of only one excitation energy U, one has to distinguish between different possible excitations, high-spin (HS) and low-spin (LS). Next to the Coulomb on-site repulsion U known from the Hubbard model, the degenerate Hubbard Hamiltonian [7] includes Hund's exchange J. In general, on-site Coulomb interactions between two electrons in $3d$ orbitals depend both on spin and orbital indices. Note that the electron interaction parameters in this model are effective ones, i.e., the $2p$-orbital parameters of O (or F) ions renormalize on-site Coulomb interactions between two electrons in $3d$ orbitals. A general form which includes only two-orbital interactions and the anisotropy of Coulomb and exchange elements is [28, 29]

$$\begin{aligned} H_{int} &= U \sum_{i\alpha} n_{i\alpha\uparrow} n_{i\alpha\downarrow} + \sum_{i,\alpha<\beta} \left(U'_{\alpha\beta} - \frac{1}{2} J_{\alpha\beta} \right) n_{i\alpha} n_{i\beta} - 2 \sum_{i,\alpha<\beta} J_{\alpha\beta}\, \vec{S}_{i\alpha} \cdot \vec{S}_{i\beta} \\ &+ \sum_{i,\alpha<\beta} J_{\alpha\beta} \left(a^{\dagger}_{i\alpha\uparrow} a^{\dagger}_{i\alpha\downarrow} a_{i\beta\downarrow} a_{i\beta\uparrow} + a^{\dagger}_{i\beta\uparrow} a^{\dagger}_{i\beta\downarrow} a_{i\alpha\downarrow} a_{i\alpha\uparrow} \right). \end{aligned} \tag{19}$$

Here $a^{\dagger}_{i\alpha\sigma}$ is an electron creation operator in any $3d$ orbital, $\alpha \in \{xy, yz, zx, 3z^2{-}r^2, x^2{-}y^2\}$, with spin states $\sigma = \uparrow, \downarrow$ at site i, and we shall use $\bar{\sigma} \equiv -\sigma$. The parameters $\{U, U'_{\alpha\beta}, J_{\alpha\beta}\}$ depend in general on the three Racah parameters $\{A, B, C\}$ [30], which may be derived from somewhat screened atomic values. While the intra-orbital Coulomb element is identical for all $3d$ orbitals,

$$U \equiv A + 4B + 3C, \tag{20}$$

the inter-orbital Coulomb $U'_{\alpha\beta}$ and exchange $J_{\alpha\beta}$ elements are anisotropic and depend on the involved pair of orbitals $\{\alpha, \beta\}$; the values of $J_{\alpha\beta}$ are given in Table 1. The inter-orbital Coulomb

Table 1: *On-site inter-orbital exchange elements $J_{\alpha\beta}$ for $3d$ orbitals as functions of the Racah parameters B and C (for more details see Ref. [30]).*

$3d$ orbital	xy	yz	zx	x^2-y^2	$3z^2-r^2$
xy	0	$3B+C$	$3B+C$	C	$4B+C$
yz	$3B+C$	0	$3B+C$	$3B+C$	$B+C$
zx	$3B+C$	$3B+C$	0	$3B+C$	$B+C$
x^2-y^2	C	$3B+C$	$3B+C$	0	$4B+C$
$3z^2-r^2$	$4B+C$	$B+C$	$B+C$	$4B+C$	0

$U'_{\alpha\beta}$ and Hund's exchange $J_{\alpha\beta}$ elements satisfy a relation with intra-orbital element U which guarantees the rotational invariance of interactions in the orbital space,

$$U = U'_{\alpha\beta} + 2J_{\alpha\beta}. \tag{21}$$

In all situations where only the orbitals belonging to a single irreducible representation of the cubic group (e_g or t_{2g}) are partly filled, e.g. in the titanates, vanadates, nickelates, or copper fluorides, the filled (empty) orbitals do not contribute to the dynamics, and the relevant exchange elements $J_{\alpha\beta}$ are all the same (see Table 1), i.e., either a pair of t_{2g} or for e_g orbitals,

$$J_H^t \equiv 3B+C, \tag{22}$$

$$J_H^e \equiv 4B+C. \tag{23}$$

Then one may use a simplified *degenerate* Hubbard model [7] with isotropic on-site interactions (for a given subset of $3d$ orbitals),

$$\begin{aligned} H_{int}^{(0)} &= U\sum_{i\alpha} n_{i\alpha\uparrow}n_{i\alpha\downarrow} + \left(U-\frac{5}{2}J_H\right)\sum_{i,\alpha<\beta} n_{i\alpha}n_{i\beta} - 2J_H\sum_{i,\alpha<\beta}\vec{S}_{i\alpha}\cdot\vec{S}_{i\beta} \\ &+J_H\sum_{i,\alpha<\beta}\left(a^\dagger_{i\alpha\uparrow}a^\dagger_{i\alpha\downarrow}a_{i\beta\downarrow}a_{i\beta\uparrow} + a^\dagger_{i\beta\uparrow}a^\dagger_{i\beta\downarrow}a_{i\alpha\downarrow}a_{i\alpha\uparrow}\right). \end{aligned} \tag{24}$$

Eq. (24) has two Kanamori parameters: the Coulomb intra-orbital element U (20) and Hund's exchange J_H, and we parametrize the interactions by

$$\eta = J_H/U \tag{25}$$

which stands either for J_H^t (22) or for J_H^e (23), depending on the electronic filling of $3d$ orbitals at site i. Here we also obtain the celebrated element $\bar{U} \equiv U-3J_H^e$, used before in Eq. (16) when only HS states occur. We emphasize that in a general case when both types of orbitals are partly filled, as in the colossal magnetoresistance (CMR) manganites [31], and both thus participate in charge excitations of Fig. 6, the above Hamiltonian with a single Hund's exchange element J_H^e is insufficient and the full anisotropy given in Eq. (24) has to be used instead to generate the correct charge excitation spectra for a given transition metal ion [30].

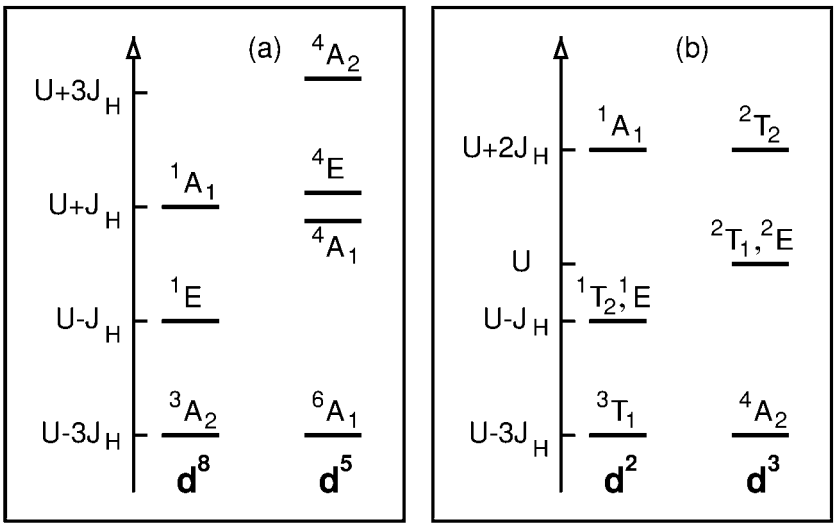


Fig. 6: *Energies of charge excitations ε_n (26) for selected cubic transition metal oxides, for: (a) e_g excitations to Cu^{3+} (d^8) and Mn^{2+} (d^5) ions; (b) t_{2g} excitations to Ti^{2+} (d^2) and V^{2+} (d^3) ions. The splittings between different states are due to Hund's exchange element J_H which refers to a pair of e_g and t_{2g} electrons in (a) and (b). Image reproduced from Ref. [28].*

In a strongly correlated regime $t \ll U$, we consider the case of partly filled degenerate $3d$ orbitals and large Hund's exchange J_H. This guarantees that electrons localize in high-spin ionic states, and effective low-energy superexchange interactions consist of all the contributions which originate from possible virtual charge excitations, $d_m^p d_n^p \rightleftharpoons d_m^{p+1} d_n^{p-1}$ —they take the form of a spin-orbital model. The charge excitation n costs the energy

$$\varepsilon_n = E_n(d^{p+1}) + E_0(d^{p-1}) - 2E_0(d^p), \tag{26}$$

where the d^p ions are in the initial high-spin ground states with spins $S = \frac{p}{2}$ and have the Coulomb interaction energy $E_0(d^p) = \binom{p}{2}(U - 3J_H^e)$ each if $p < 5$ (otherwise if $p > 5$ one has to consider p holes instead). The case of $p = 5$ electrons is special and will not be considered here as in the $t_{2g}^3 e_g^2$ configuration the orbital degree of freedom is quenched.

The same formula for the ground state energy applies as well to Mn^{3+} ions in d^4 configuration with $S = 2$ spin HS ground state, see Sec. 2. By construction, also the ion with less electrons (holes) for $p < 5$ ($p > 5$) is in the HS state and $E_0(d^{p-1}) = \binom{p-1}{2}(U - 3J_H)$. The excitation energies (26) are thus defined by the multiplet structure of an ion with more electrons (holes) in the configuration d^{p+1}, see Fig. 6. The lowest energy excitation is always given by $U - 3J_H$ —it is obtained from the HS state of the $3d^{p+1}$ ion with total spin $\mathcal{S} = S + 1/2$ and energy $E_1(d^{p+1}) = \binom{p+1}{2}(U - 3J_H)$, with J_H being Hund's exchange element for the electron (hole) involved in the charge excitation, either e_g or t_{2g}. Indeed, one recovers the lowest excitation energy in the HS subspace, ε_1. We emphasize that this lowest excitation energy ε_1 is universal and is found both in t_{2g} and e_g subspaces, i.e., it does not depend on the electron valence p, see Fig. 6. In contrast, the remaining energies $\{\varepsilon_n\}$ for $n > 1$ are all for LS excitations and are specific to a given valence p of the considered insulator with d^p ions. They have to be determined from the full local Coulomb interaction Hamiltonian (19), in general including also the anisotropy of the $\{U_{\alpha\beta}\}$ and $\{J_{\alpha\beta}\}$ elements.

Effective interactions in a Mott (or charge transfer) insulator with orbital degeneracy take the form of spin-orbital superexchange [1, 12]. Its general structure is given by the sum over all the nearest neighboring bonds $\langle ij\rangle \parallel \gamma$ connecting two transition metal ions and over the excitations n possible for each of them as,

$$\mathcal{H} = -\sum_n \frac{t^2}{\varepsilon_n} \sum_{\langle ij\rangle\parallel\gamma} P_{\langle ij\rangle}(\mathcal{S})\, \mathcal{O}^{\gamma}_{\langle ij\rangle}, \tag{27}$$

where $P_{\langle ij\rangle}(\mathcal{S})$ is the projection on the total spin $\mathcal{S} = S \pm 1/2$ and $\mathcal{O}^{\gamma}_{\langle ij\rangle}$ is the projection operator on the orbital state at the sites i and j of a given bond. Following this general procedure, one finds a spin-orbital model with Heisenberg spin interactions for spins $S = {}^{p}/_{2}$ of SU(2) symmetry coupled to the orbital operators which have much lower cubic symmetry, with the general structure of spin-orbital superexchange $\propto J$ (2) [28],

$$\mathcal{H}_J = J \sum_{\gamma} \sum_{\langle ij\rangle\parallel\gamma} \left\{ \hat{\mathcal{K}}^{(\gamma)}_{ij} \left(\vec{S}_i \cdot \vec{S}_j + S^2 \right) + \hat{\mathcal{N}}^{(\gamma)}_{ij} \right\}. \tag{28}$$

It connects ions at sites i and j along the bond $\langle ij\rangle \parallel \gamma$ and involves orbital operators, $\hat{\mathcal{K}}^{(\gamma)}_{ij}$ and $\hat{\mathcal{N}}^{(\gamma)}_{ij}$, which depend on the bond direction $\gamma = a, b, c$ for the three *a priori* equivalent directions in a cubic crystal. The spin scalar product, $\vec{S}_i \cdot \vec{S}_j$, is coupled to orbital operators $\hat{\mathcal{K}}^{(\gamma)}_{ij}$ which together with the other "decoupled" orbital operators, $\hat{\mathcal{N}}^{(\gamma)}_{ij}$, determine the orbital state in a Mott insulator. The form of these operators depends on the type of orbital degrees of freedom in a given model. They involve active orbitals on each bond $\langle ij\rangle \parallel \gamma$ along direction γ. Thus the orbital interactions are directional and have only the cubic symmetry of a (perovskite) lattice provided the symmetry in the orbital sector is not broken by other interactions, for instance by CF or Jahn-Teller (JT) terms.

The magnetic superexchange constants along each cubic axis J_{ab} and J_c in the effective spin model,

$$H = J_{ab} \sum_{\langle ij\rangle\parallel ab} \vec{S}_i \cdot \vec{S}_j + J_c \sum_{\langle ij\rangle\parallel c} \vec{S}_i \cdot \vec{S}_j, \tag{29}$$

are obtained from the spin-orbital model (28) by decoupling spin and orbital operators and next averaging the orbital operators over an underlying orbital (ordered or disordered) state. It gives effective magnetic exchange interactions: J_c for a bond along the c axis, and J_{ab} for bonds within the ab plane. The latter ones J_{ab}, could in principle still be different between the a and b axes in case of finite lattice distortions due to the JT effect or octahedra tilting, but we limit ourselves to idealized structures, with J_{ab} being the same for both planar directions. We show below that the spin-spin correlations along the c axis and within the ab planes,

$$s_c = \left\langle \vec{S}_i \cdot \vec{S}_j \right\rangle_c, \qquad s_{ab} = \left\langle \vec{S}_i \cdot \vec{S}_j \right\rangle_{ab}, \tag{30}$$

next to the orbital correlations, play an important role in the intensity distribution in optical spectroscopy.

In correlated insulators with partly occupied degenerate orbitals, not only the structure of the superexchange (28) is complex, but also the optical spectra exhibit strong anisotropy and temperature dependence near the magnetic transitions, as found, e.g., in $LaMnO_3$ [32, 33] or in the cubic vanadates, $LaVO_3$ and YVO_3 [28]. In such systems several excitations contribute to the excitation spectra, so one may ask how the spectral weight redistributes between individual subbands originating from these excitations. The spectral weight distribution is in general anisotropic already when orbital order (OO) [34] sets in and breaks the cubic symmetry, but even more so when A-type or C-type AF spin order occurs below the Néel temperature $T_{\rm N}$. At orbital degeneracy the superexchange consists of the terms $\propto H_n^{(\gamma)}(ij)$ as a superposition of individual contributions on each bond $\langle ij\rangle$ due to charge excitation n (26) [35],

$$\mathcal{H} = J \sum_n \sum_{\langle ij\rangle \| \gamma} H_n^{(\gamma)}(ij), \tag{31}$$

with the energy unit for each individual $H_n^{(\gamma)}(ij)$ term given by the superexchange constant J, see Eq. (2). It follows from d-d charge excitations with an effective hopping element t between neighboring transition metal ions and is the same as that obtained in a Mott insulator with nondegenerate orbitals in the regime of $U \gg t$. The spectral weight in optical spectroscopy is determined by the kinetic energy, and reflects the onset of spin order (SO) and/or OO [35]. In a correlated insulator, electrons are almost localized and the only kinetic energy which is left is associated with the same virtual charge excitations that contribute also to the superexchange. Therefore, the individual kinetic energy terms $K_n^{(\gamma)}$ may be directly determined from the superexchange (31) using the Hellmann-Feynman theorem,

$$K_n^{(\gamma)} = -2J \left\langle H_n^{(\gamma)}(ij) \right\rangle . \tag{32}$$

For convenience, we define here the $K_n^{(\gamma)}$ as positive quantities. Each term $K_n^{(\gamma)}$ (32) originates from a given charge excitation n along a bond direction $\langle ij\rangle \parallel \gamma$. These terms are straightforwardly related to the *partial optical sum rule* for individual Hubbard subbands, which reads [35]

$$\frac{a_0 \hbar^2}{e^2} \int_0^\infty \sigma_n^{(\gamma)}(\omega)\, d\omega = \frac{\pi}{2} K_n^{(\gamma)}, \tag{33}$$

where $\sigma_n^{(\gamma)}(\omega)$ is the contribution of excitation n to the optical conductivity for polarization along the γ axis, a_0 is the distance between transition metal ions, and the tight-binding model with nearest neighbor hopping is implied. Using Eq. (32) one finds that the intensity of each band is indeed determined by the underlying OO together with the spin-spin correlation along the direction corresponding to the polarization.

One has to distinguish the above partial sum rule (33) from the full sum rule for the total spectral weight in the optical spectroscopy for polarization along a cubic direction γ, involving

$$K^{(\gamma)} = -2J \sum_n \left\langle H_n^{(\gamma)}(ij) \right\rangle , \tag{34}$$

which stands for the total intensity in the optical d-d excitations. This quantity is usually of less interest as it does not allow for a direct insight into the nature of the electronic structure

being a sum over several excitations ε_n (26) and has a much weaker temperature dependence. In addition, it might also be more difficult to deduce the quantity from experiment.

3.2 Goodenough-Kanamori rules

While a rather advanced treatment of the quantum many-body physics is required in general for spin-orbital models, we want to present here certain principles which help to understand the heart of the problem and to give simple guidelines for interpreting experiments and for finding relevant physical parameters of the spin-orbital models of *undoped* cubic insulators. We will argue that such an approach based upon classical OO is well justified in many known cases, as quantum phenomena are often quenched by the JT coupling between orbitals and by lattice distortions, which are present below structural phase transitions and induce OO, either in spin-disordered, in spin-ordered phases, or in spin-liquid.

From the derivation of the Kugel-Khomskii (KK) model in Sec. 4.1, we observe that pairs of directional orbitals $\{|i\zeta_\gamma\rangle, |j\zeta_\gamma\rangle\}$ on neighboring ions favor AF SO, while pairs of orthogonal orbitals such as $\{|i\zeta_\gamma\rangle, |j\xi_\gamma\rangle\}$ favor FM SO. This is known as classical Goodenough-Kanamori rules (GKR) [36] predicting that the state with AF SO has simultaneously FO order, while FM SO is accompanied by AO order, see Figs. 7(a) and 7(b). Indeed, these rules emphasizing the complementarity of spin-orbital correlations are frequently employed to explain the observed spin-orbital order in several systems, particularly in those where spins are large, like in CMR manganites [31]. They agree with the general structure of spin-orbital superexchange in the KK model, where it is sufficient to consider the flavor-conserving hopping between pairs of directional orbitals $\{|i\zeta_\gamma\rangle, |j\zeta_\gamma\rangle\}$ [29, 37]. The excited states are then doubly occupied in one of the directional orbitals, while no effective interaction arises for two parallel spins (in triplet states), so the superexchange is AF. In contrast, for a pair of orthogonal orbitals, e.g. $\{|i\zeta_\gamma\rangle, |j\xi_\gamma\rangle\}$, two different orbitals are singly occupied and the FM term is stronger than the AF one as the excitation energy is lower. Therefore, configurations with AO order support FM SO.

The above complementarity of spin-orbital order is frustrated by inter-orbital hopping, or may be modified by spin-orbital entanglement [11], see below. In such cases the order in both channels could be the same, either FM/FO, see Fig. 7(c), or AF/AO, see Fig. 7(d). Again, when different orbitals are occupied in the excited state, the spin superexchange is weakly FM and when the same orbital is doubly occupied, the spin superexchange is stronger and AF. The latter AF exchange coupling dominates because antiferromagnetism, which is due to the Pauli principle, does not have to compete here with ferromagnetism. On the contrary, FM exchange is caused by the energy difference $\propto \eta$ between triplet and singlet excited states, with two different orbitals occupied.

This modification of the GKR is of importance in alkali RO_2 hyperoxides ($R =$ K, Rb, Cs) [38]. The JT effect is crucial for this generalization of the GKR—without it large inter-orbital hopping orders the T^x-orbital-mixing pseudospin component instead of the T^z component in a single plane. Altogether, such generalized GKR can arise whenever the OO on a bond is not solely stabilized by the same spin-orbital superexchange interaction that determines the spin exchange.

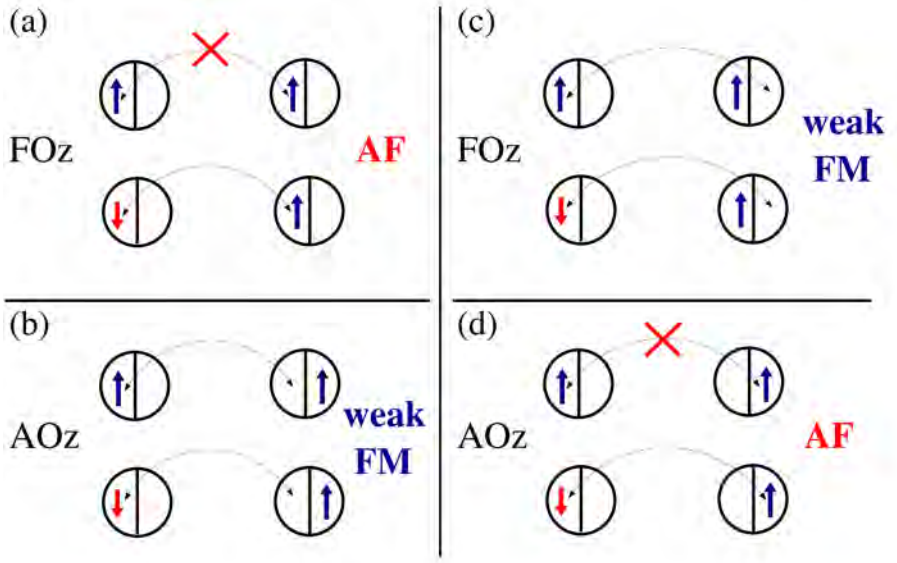


Fig. 7: *Artist's view of the GKR [36] for: (a) FOz and AF spin order and (b) AOz and FM spin order in a system with orbital flavor conserving hopping as alkali RO_2 hyperoxides (R = K, Rb, Cs) [38]. The charge excitations generated by inter-orbital hopping fully violate the GKR and support the states with the same spin-orbital order: (c) FOz and FM spin order and (d) AOz and AF spin order. Image reproduced from Ref. [38].*

On a geometrically frustrated lattice, for instance, another route to this behavior can occur when the ordered orbital component preferred by superexchange depends on the direction and the relative strengths fulfill certain criteria.

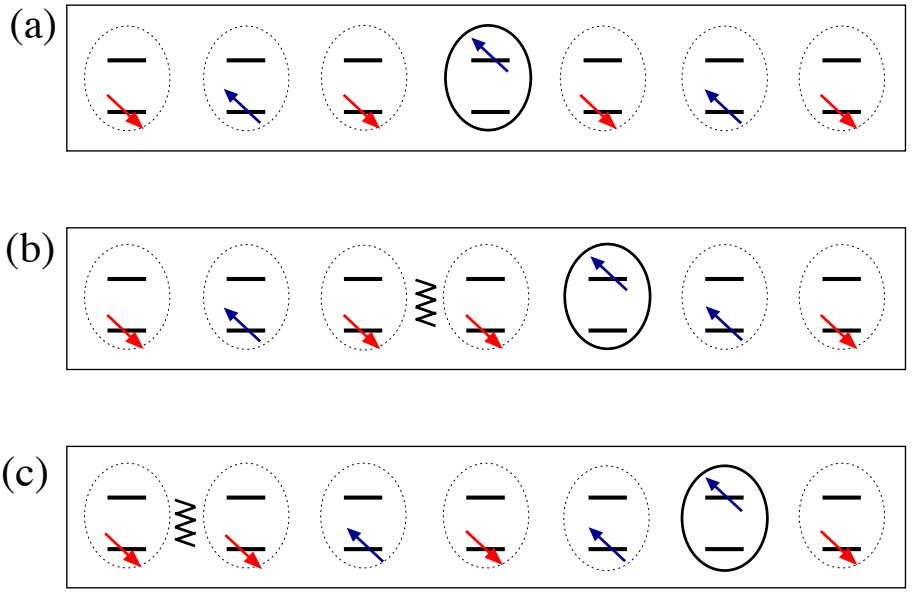


Fig. 8: *Schematic representation of the orbital motion and the spin-orbital separation in a 1D spin-orbital model. The first hop of the excited state (a)→(b) creates a spinon (wavy line) that moves via spin exchange $\propto J$. The next hop (b)→(c) gives an "orbiton" freely propagating as a "holon" with an effective hopping $t \sim J/2$. Image reproduced from Ref. [39].*

While a hole doped to the FM chain propagates freely, it creates a spinon and a holon in an AF background described by the t-J model. A similar situation occurs for an orbital excitation in an AF/FO spin-orbital chain [39]. An orbital excitation may propagate through the system only after creating a spinon in the first step, see Figs. 8(a) and 8(b). The spinon itself moves via spin flips $\propto J > t$, faster than the orbiton, and the two excitations get well separated, see Fig. 8(c). The orbital-wave picture of Sec. 2, on the other hand, would require the orbital excitation to move without creating the spinon in the first step. Note that this would be only possible for imperfect Néel AF SO. Thus, one concludes that the symmetry between spin and orbital sector is broken also for this reason and orbitals are so strongly coupled to spin excitations in realistic spin-orbital models with AF/FO order. In conclusion the mean field picture separating these two sectors of the Hilbert space breaks down.

4 Kugel-Khomskii model for Mott insulators

4.1 Kugel-Khomskii model: 3D for $KCuF_3$ and 2D for K_2CuF_4

The simplest and seminal spin-orbital model is obtained when a fermion has two flavors, spin and orbital, and both have two components, i.e., spin and pseudospin are $S = T = 1/2$. The physical realization is found in cuprates with degenerate e_g orbitals, such as $KCuF_3$ or K_2CuF_4 [1], where Cu^{2+} ions are in the d^9 electronic configuration, so charge excitations $d_i^9 d_j^9 \rightleftharpoons d_i^{10} d_j^8$ are made by holes. By considering the degenerate Hubbard model for two e_g orbitals one finds that d^8 ions have an equidistant multiplet structure, with three excitation energies which differ by $2J_H$ [here J_H stands for the J_H^e given by Eq. (23)], see Table 2. We emphasize that the correct spectrum has a doubly degenerate energy $U-J_H$, and the highest non-degenerate energy is $U+J_H$, see Fig. 6(a). Note that this result follows from the diagonalization of the local Coulomb interactions in the relevant subspaces—it reflects the fact that a double occupancy $\big(|z\uparrow z\downarrow\rangle$ or $|\bar{z}\uparrow\bar{z}\downarrow\rangle\big)$ in either orbital state ($|z\rangle$ or $|\bar{z}\rangle$) is not an eigenstate of the degenerate Hubbard model in the atomic limit (24), so the excitation energy U is absent in the spectrum, see Table 2.
The total spin state on the bond $\langle ij\rangle$ corresponds to $\mathcal{S}$=1 or 0, so the spin projection operators $P_{\langle ij\rangle}(1)$ and $P_{\langle ij\rangle}(0)$ are easily deduced, see Table 2. The orbital configuration on a bond $\langle ij\rangle$ is given by one of the orbital operators in Sec. 2, either $\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$ for the doubly occupied states involving different orbitals, or $\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$ for a double occupancy in a directional orbital at site i or j. This gives the rather transparent structure of one HS and three LS excitations in Table 2. The 3D KK model then follows from Eq. (27) [9,40]:

$$\mathcal{H}(d^9) = \sum_{\gamma}\sum_{\langle ij\rangle\parallel\gamma}\left\{-\frac{t^2}{U-3J_H}\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)\mathcal{P}^{(\gamma)}_{\langle ij\rangle}+\frac{t^2}{U-J_H}\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\mathcal{P}^{(\gamma)}_{\langle ij\rangle}\right.$$
$$\left.+\left(\frac{t^2}{U-J_H}+\frac{t^2}{U+J_H}\right)\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}\right\}+E_z\sum_i\tau_i^{(c)}. \quad (35)$$

The last term $\propto E_z$ is the CF which splits off the degenerate e_g orbitals when a JT lattice distortion occurs, and is together with Hund's exchange η, a second parameter to construct

Table 2: *Elements needed for the construction of the KK model from charge excitations on the bond $\langle ij\rangle$: excitation n, its type (HS or LS) and energy ε_n, total spin state (triplet or singlet) and the spin projection operator $P_{\langle ij\rangle}(\mathcal{S})$, and the orbital state and the corresponding orbital projection operator.*

charge excitation			spin state		orbital state	orbital
n	type	ε_n	$\mathcal{S}$	$P_{\langle ij\rangle}$	on a bond $\langle ij\rangle \parallel \gamma$	projection
1	HS	$U-3J_H$	1	$\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)$	$\vert i\zeta_\gamma\rangle\vert j\xi_\gamma\rangle$ $(\vert i\xi_\gamma\rangle\vert j\zeta_\gamma\rangle)$	$\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$
2	LS	$U-J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\vert i\zeta_\gamma\rangle\vert j\xi_\gamma\rangle$ $(\vert i\xi_\gamma\rangle\vert j\zeta_\gamma\rangle)$	$\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$
3	LS	$U-J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\vert i\zeta_\gamma\rangle\vert j\zeta_\gamma\rangle$	$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$
4	LS	$U+J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\vert i\zeta_\gamma\rangle\vert j\zeta_\gamma\rangle$	$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$

phase diagrams, see below. Here it refers to holes, i.e., large $E_z>0$ favors hole occupation in $|\bar{z}\rangle\equiv|x^2{-}y^2\rangle/\sqrt{2}$ orbitals, as in La_2CuO_4. On the other hand, while $E_z\simeq 0$, both orbitals have almost equal hole (electron) density.

Another form of the Hamiltonian (35) is obtained by introducing the coefficients,

$$r_1=\frac{1}{1{-}3\eta},\qquad r_2=r_3=\frac{1}{1{-}\eta},\qquad r_4=\frac{1}{1{+}\eta}, \tag{36}$$

and defining the superexchange constant J in the same way as in the t-J model Eq. (2). With the explicit representation of the orbital operators $\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$ and $\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$ in terms of $\left\{\tau_i^{(\gamma)}\right\}$ one finds,

$$\begin{aligned}\mathcal{H}(d^9) \;=\; \frac{1}{2}J\sum_\gamma\sum_{\langle ij\rangle\parallel\gamma}\Bigg\{&\left[-r_1\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)+r_2\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\right]\left(\frac{1}{4}-\tau_i^{(\gamma)}\tau_j^{(\gamma)}\right)\\&+(r_3{+}r_4)\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\left(\tau_i^{(\gamma)}+\frac{1}{2}\right)\left(\tau_j^{(\gamma)}+\frac{1}{2}\right)\Bigg\}+E_z\sum_i\tau_i^{(c)}.\end{aligned} \tag{37}$$

In the FM state spins are integrated out and one finds from the first term just the superexchange in the e_g orbital model analyzed above in Sec. 2.

The magnetic superexchange constants J_{ab} and J_c employed in the effective spin-orbital model (37) are obtained by decoupling spin and orbital operators and next averaging the orbital operators $\left\langle\hat{\mathcal{K}}_{ij}^{(\gamma)}\right\rangle$ over the classical state $|\Phi_0\rangle$ as given by Eq. (12). The relevant averages are given in Table 3, and they lead to the following expressions for the superexchange constants in Eq. (29),

$$J_c \;=\; \frac{1}{8}J\Big\{-r_1\sin^2\theta+(r_2{+}r_3)(1+\cos\theta)+r_4(1+\cos\theta)^2\Big\}, \tag{38}$$

$$J_{ab} \;=\; \frac{1}{8}J\left\{-r_1\left(\frac{3}{4}+\sin^2\theta\right)+(r_2{+}r_3)\left(1-\frac{1}{2}\cos\theta\right)+r_4\left(\frac{1}{2}-\cos\theta\right)^2\right\}, \tag{39}$$

which depend on two parameters: J from Eq. (2) and η as in Eq. (25), as well as on the OO of $|\pm\rangle$ orbitals specified by the orbital angle θ. It is clear that the FM term $\propto r_1$ competes

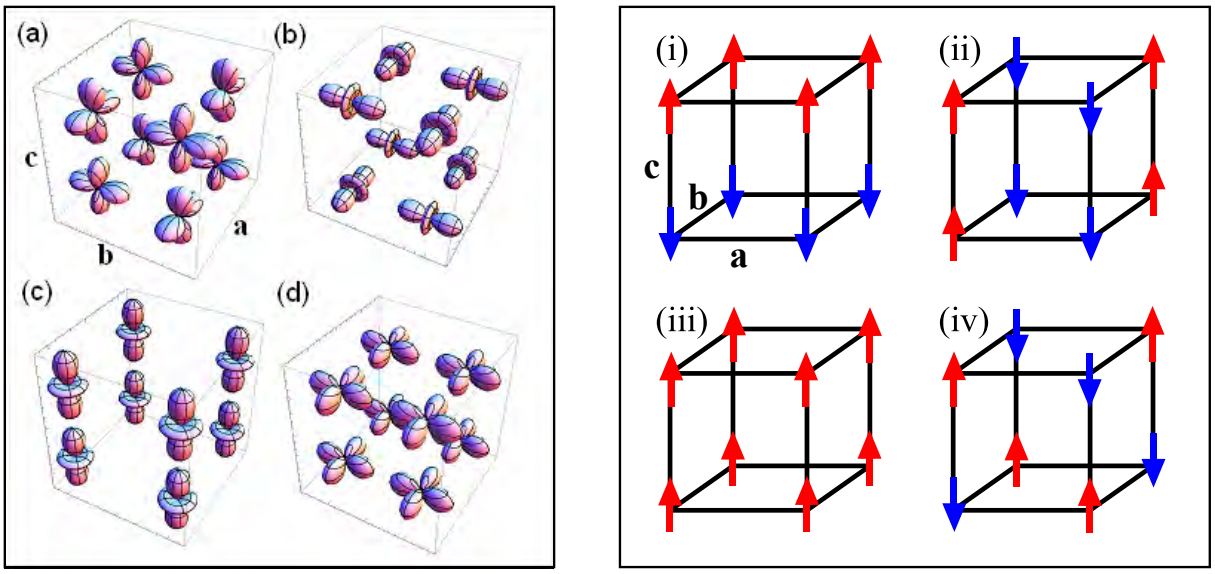


Fig. 9: *Spin-orbital entanglement in the KK model includes orbitals and spins: Left—schematic view of the four simplest e_g-orbital configurations on a representative cube of the 3D lattice: (a) alternating orbital (AO) order with $\langle\tau_i^{(a,b)}\rangle=\pm 1/2$ changing from site to site, and $\langle\tau_i^{c}\rangle=1/4$, obtained for $E_z<0$, (b) AO order with $\langle\tau_i^{(a,b)}\rangle=-1/2$, alternating between sites and $\langle\tau_i^{(c)}\rangle=-1/4$, obtained for $E_z>0$, (c) FO order with occupied z orbitals and $\langle\tau_i^{c}\rangle=1/2$ (cigar-shaped orbitals), and (d) FO order with occupied $\bar{z}$ orbitals and $\langle\tau_i^{c}\rangle=-1/2$ (clover-shaped orbitals). Right—schematic view of four spin configurations (arrows for up or down spins; $\{a,b,c\}$ are crystallographic directions) in phases with SO: (i) A-AF, (ii) C-AF, (iii) FM, and (iv) G-AF. Image by courtesy of Wojciech Brzezicki.*

with all the other AF LS terms. Nevertheless, in the ab planes, where the occupied hole e_g orbitals alternate, the larger FM contribution dominates and makes the magnetic superexchange J_{ab} weakly FM ($J_{ab}\lesssim 0$ when $\sin^2\theta\simeq 1$), while the stronger AF superexchange along the c axis ($J_c\gg|J_{ab}|$) favors quasi one-dimensional (1D) spin fluctuations. Thus $KCuF_3$ exhibits spinon excitations for $T>T_{\rm N}$.

Table 3: *Averages of the orbital projection operators standing in the spin-orbital interactions in the KK model (37) and determine the spin interactions in H_s (29) for the C-type OO of occupied e_g orbitals which alternate in the ab planes, as given by Eqs. (14). Nonequivalent cubic directions along the $\langle ij\rangle$ bonds are labeled by $\gamma=ab,c$.*

operator	average	ab	c
$\mathcal{Q}_{\langle ij\rangle}^{(\gamma)}$	$2\left\langle\left(\frac{1}{2}-\tau_i^{(\gamma)}\right)\left(\frac{1}{2}-\tau_j^{(\gamma)}\right)\right\rangle$	$\frac{1}{2}\left(\frac{1}{2}-\cos\theta\right)^2$	$\frac{1}{2}\left(1+\cos\theta\right)^2$
$\mathcal{P}_{\langle ij\rangle}^{(\gamma)}$	$\left\langle\frac{1}{4}-\tau_i^{(\gamma)}\tau_j^{(\gamma)}\right\rangle$	$\frac{1}{4}\left(\frac{3}{4}+\sin^2\theta\right)$	$\frac{1}{4}\sin^2\theta$
$\mathcal{R}_{\langle ij\rangle}^{(\gamma)}$	$2\left\langle\left(\frac{1}{2}+\tau_i^{(\gamma)}\right)\left(\frac{1}{2}+\tau_j^{(\gamma)}\right)\right\rangle$	$\frac{1}{2}\left(\frac{1}{2}+\cos\theta\right)^2$	$\frac{1}{2}\left(1-\cos\theta\right)^2$

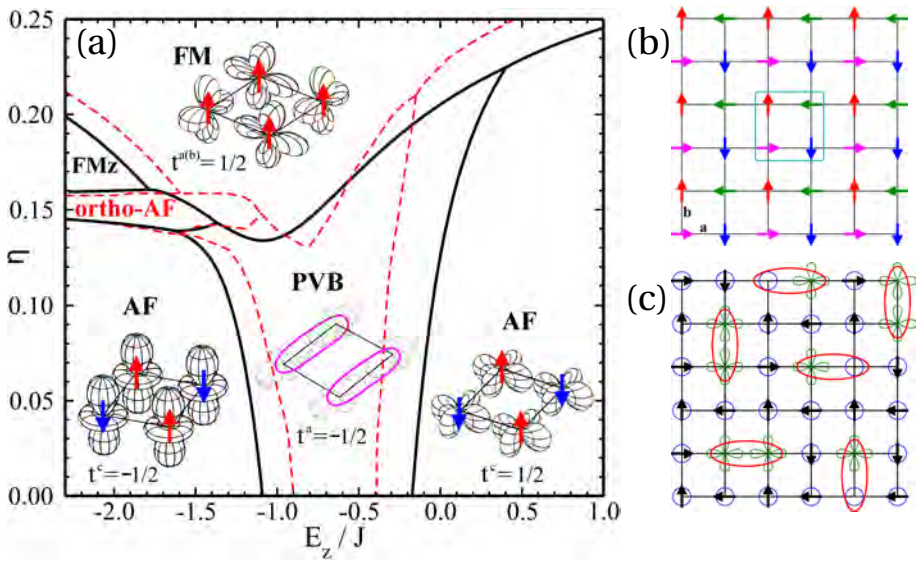


Fig. 10: *Spin-orbital phase diagram and entanglement in the 2D KK model: (a) phase diagram in the plaquette mean field (solid lines) and ERA (dashed lines) variational approximation, with insets showing representative spin and orbital configurations on a 2×2 plaquette—$\bar{z}$-like $\left(t^c{=}-\langle\tau_i^{(c)}\rangle{=}\frac{1}{2}\right)$ and z-like $\left(t^{a,c}{=}-\langle\tau_i^{(c,a)}\rangle{=}-\frac{1}{2}\right)$ orbitals are accompanied either by AF long range spin order (arrows) or by spin singlets on bonds in the PVB phase (ovals); (b) view of an exotic four-sublattice ortho-AF phase near the onset of FM (or FMz) phase; (c) artist's view of the ortho-AF phase—spin singlets (ovals) are entangled with either one or two orbital excitations $|z\rangle \rightarrow |\bar{z}\rangle$ (clovers). Images reproduced from Ref. [41].*

Consider first the 2D KK model on a square lattice, with $\gamma = a, b$ in Eq. (37), as in K_2CuF_4. In the absence of Hund's exchange, interactions between $S = {}^1/_2$ spins are AF. However, they are quite different depending on which of the two e_g orbitals are occupied by holes: $J^z_{ab} = \frac{1}{16}J$ for $|z\rangle$ and $J^{\bar{z}}_{ab} = \frac{9}{16}J$ for $|\bar{z}\rangle$ hole orbitals. As a result, the AF phases with SO in Fig. 9(iv) and the FO order shown in Figs. 9(c) and 9(d) are degenerate at finite CF $E_z = -\frac{1}{2}J$. This defines a quantum critical point (QCP) $Q_{2D} = (-{}^1/_2, 0)$ in the $(E_z/J, \eta)$ plane [while $Q_{3D} = (0, 0)$]. Actually, at this point also one more phase has the same energy—the FM spin phase of Fig. 9(i) with AO order of $|\pm\rangle$ orbitals, shown in Figs. 5(a&b) [40].

To capture the corrections due to quantum fluctuations, one may construct a plaquette mean field approximation or entanglement renormalization *ansatz* (ERA) [41]. One finds important corrections to the mean field phase diagram near the QCP Q_{2D}, and a plaquette valence bond (PVB) state is stable in between the above three phases accompanied by spin-orbital long range order, with spin singlets on the bonds $\parallel a$ ($\parallel b$) and stabilized by the directional orbitals $|\zeta_a\rangle$ ($|\zeta_b\rangle$). A novel ortho-AF phase appears as well when the magnetic interactions change from AF to FM ones due to increasing Hund's exchange η, and for $E_z/J < -1.5$, see Fig. 10(a). Since the nearest neighbor magnetic interactions are very weak, exotic four-sublattice ortho-AF SO emerges due to second and third nearest neighbor interactions, shown in Fig. 10(b). Such further neighbor interactions follow from spin-orbital excitations shown in Fig. 10(c). Note that both approximate methods employed in Ref. [41] (plaquette mean field approximation and ERA) give very similar range of stability of the ortho-AF phase.

4.2 Entanglement in the ferromagnetic excitations of K_2CuF_4

To investigate magnons (spin waves), we create a spin excitation at site $i = 0$ by decreasing the value of the order parameter $\langle S_0^z \rangle$ from S to $S-1$. In the simplest approach we disentangle [29] spin-orbital superexchange both in the ground and excited state, and use the same frozen AO order as in the initial state to determine spin exchange $J_\diamondsuit$. A spin excitation (a magnon) itself is best described by the transformation to Holstein-Primakoff (HP) bosons. In linear spin-wave theory, the magnon energy consists of two contributions and we introduce:

(i) Ising energy for a localized HP boson $I^{(0)} \equiv 4J_\diamondsuit S$, and

(ii) the propagating term $P^{(0)}(\vec{k}) \equiv -4J_\diamondsuit S\gamma_{\vec{k}}$.

The latter originates from quantum fluctuations $\propto -\frac{1}{2}J_\diamondsuit\big(\hat{S}_i^+\hat{S}_j^- + \hat{S}_i^-\hat{S}_j^+\big)$, where $\gamma_{\vec{k}} = \frac{1}{4}\sum_{\vec{\delta}} e^{i\vec{k}\cdot\vec{\delta}}$ determines the dispersion and depends on the 2D momentum $\vec{k} = (k_a, k_b)$ with $k_\alpha \in [-\pi, \pi)$. Here $\vec{\delta}$ stands for one of four nearest neighbors of the central site $i = 0$ shown in Fig. 11(a). The above two terms determine the magnon dispersion in a 2D ferromagnet,

$$\omega_{\vec{k}}^{(0)} = I^{(0)} + P^{(0)}(\vec{k}) = 4J_\diamondsuit S\,(1-\gamma_{\vec{k}}), \tag{40}$$

which serves as a reference below. The breaking of SU(2) symmetry is reflected by a Goldstone mode (at $\vec{k} = 0$), and $\omega_{\vec{k}} = J_\diamondsuit S k^2$ for $\vec{k} \to 0$ —we find that this result is insensitive to spin-orbital coupling. It is crucial that the above dispersion (40) is improved and the variational approximation (VA) is performed for each value of momentum $\vec{k}$ independently. One might expect that this reduces spin exchange, $J_\diamondsuit \to J_{\blacklozenge}$, and the magnon dispersion would soften. In this way we obtain the renormalized magnon dispersion which replaces Eq. (40),

$$\omega_{\vec{k}}(\{\theta_{iL}\}) = I(\{\theta_{iL}\};\vec{k}) + P(\{\theta_{iL}\};\vec{k}). \tag{41}$$

Note that the angles $\{\theta_{iL}\}$ are real and $L = A, B$ refers to the sublattice. If in addition it is assumed that orbital optimization for both sublattices is equivalent, we use the constraint $\theta_i \equiv \theta_{iA} = \theta_{iB}$ $(i = 1, 2, 3)$ which defines the Simplified Variational Approximation (SVA). Finally, we have verified the predictions of the VA by exact diagonalization employing a Numerical *Ansatz* (NA) with six states per sublattice: a spin defect with or without orbital excitation, and four spin-orbital states with spin excitation at the central site together with an orbital excitation at one of the nearest neighbors. The state with excitations within a shaded cluster depicted in Fig. 11(b) may be thus expressed in terms of these six states.

Taking as an example the K_2CuF_4 state at $E_z = -0.8\,J$ shown in Fig. 11(b), one finds that the orbital renormalization is appreciable—at the central site with spin excitation it is largely modified to $\sim (x^2-y^2)$, and the orbitals at the four neighboring sites are also changed. The latter orbitals found within the VA are only weakly changed as these latter sites have three neighbors belonging to the neighbors with undisturbed AO order in Fig. 11(a), but the one at the site of spin excitation itself is radically different. For this reason, we introduce a cutoff and assume that the orbitals at further neighbors of the excited spin are unchanged. One expects then

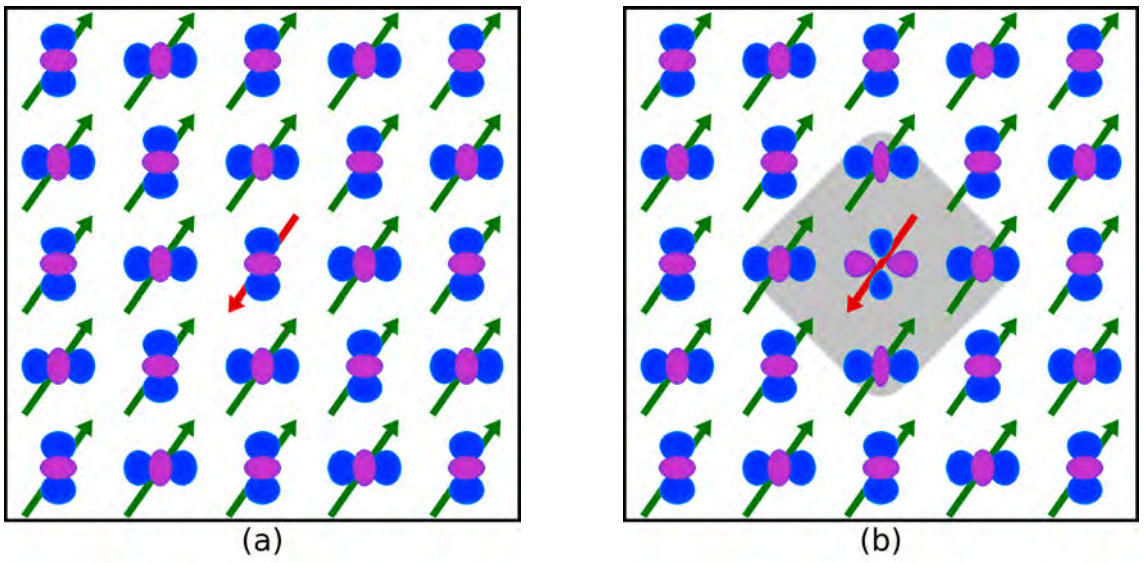


Fig. 11: *Artist's view of a spin excitation (inverted red arrow at the central site) in the FM plane of K_2CuF_4 (green arrows) and AO order of the orbitals occupied by holes at $E_z = -0.8J$, with: (a) frozen orbitals; (b) optimized orbitals at the central site and at four its neighboring sites in the square lattice, forming a quasiparticle (dressed magnon). The above value of E_z leads to the expected AO order in K_2CuF_4, with $\theta_{opt} \simeq 71^\circ$ in Eqs. (14). When the VA is used, case (a) is still realized at $\vec{k} \simeq 0$, while case (b) represents a dressed magnon with $\vec{k} \simeq M$ where orbital states in the shaded cluster are radically different from those shown for frozen orbitals in (a). Image reproduced from Ref. [42].*

a large dressing of the magnon, with the corresponding reduction of the effective FM interaction to $J_{\blacklozenge}$, particularly in the neighborhood of the M point. This is confirmed by the results shown in Fig. 12(a)—the magnon energy ω_M is reduced by $\sim 27\%$ from $\omega_M^{(0)}$. Internal consistency of the theory is confirmed by this reduction being nearly the same in all three methods used to treat spin-orbital coupling: VA, SVA, and NA.

At the X point we recognize the importance of independent optimization of orbitals on the two sublattices—the energy ω_X is reduced by $\sim 25\%$ from $\omega_X^{(0)}$ in the VA, while it stays almost unrenormalized in the SVA, see Fig. 12(a). The NA agrees very well with the results of the VA except for the points close to the M point along the M-Γ path. While the VA may underestimate somewhat the magnon dressing effect, altogether we find a comparison of the VA with the NA very encouraging indeed. The renormalization of the magnon energy increases fast when the orbital splitting $|E_z|$ is reduced, and one finds that the magnon energy reduction is large for $E_z = -0.3\,J$, e.g. by $\sim 60\%$ at the M point, see Fig. 12(b). The agreement between the VA and the NA is somewhat worse here, but still one may say that both methods qualitatively agree. Altogether, we suggest that the magnon softening may be very large for spin-orbital systems with low spin $S = 1/2$ as in K_2CuF_4. Note that similar softening is expected in the FM planes of $LaMnO_3$ and would represent an interesting future research topic.

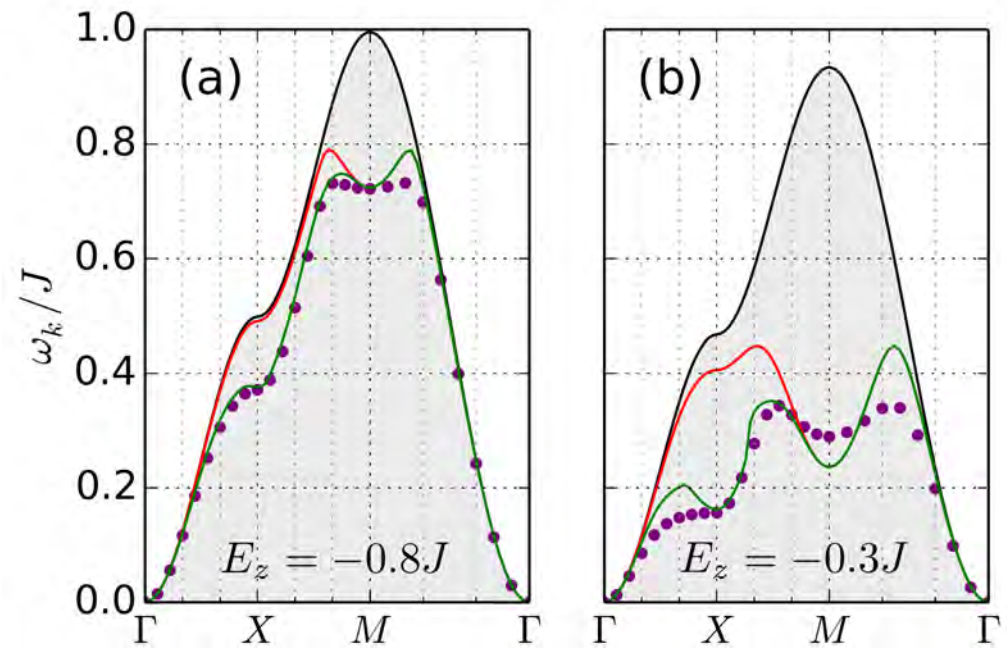


Fig. 12: *The magnon energy $\omega_{\vec{k}}/J$ obtained for the FM state of K_2CuF_4 at $J_H/U = 0.2$ and: (a) $E_z = -0.80J$ and (b) $E_z = -0.30J$. Results are presented for four approximations: frozen orbitals (black line and grey background), the VA (green line), the SVA (red line), and the 12-state NA (purple dots). The high symmetry points are: $\Gamma = (0,0)$, $X = (\pi,0)$, $M = (\pi,\pi)$. Image reproduced from Ref. [42].*

4.3 Weak spin-orbital entanglement for large spins S=2 in $LaMnO_3$

Electronic structure calculations predict A-AF SO, in agreement with experiment [32]. It follows from the spin-orbital superexchange between large spins $S = 2$ in $LaMnO_3$, due to the excitations involving e_g electrons. The energies of the five possible excited states [30] shown in Fig. 6(a) are: (i) the HS ($\mathcal{S} = \frac{5}{2}$) $^6A_1$ state, and (ii) the LS ($\mathcal{S} = \frac{3}{2}$) states: $^4A_1$, ^4E ($^4E_\epsilon$, $^4E_\theta$), and $^4A_2$. They are parameterized again by the intra-orbital Coulomb element U and by Hund's exchange J_H^e between a pair of e_g electrons at a Mn^{2+} (d^5) ion. The Racah parameters $B = 0.107$ eV and $C = 0.477$ eV justify an approximate relation $C \simeq 4B$, and we find the LS excitation spectrum: $\varepsilon(^4A_1) = U + \frac{3}{4}J_H$, $\varepsilon(^4E) = U + \frac{5}{4}J_H$ (twice), and $\varepsilon(^4A_2) = U + \frac{13}{4}J_H$. Using the spin algebra (Clebsch-Gordan coefficients) and considering again two possible e_g orbital configurations on the bonds, see Eqs. (17) and (18), and charge excitations by t_{2g} electrons, one finds a compact expression [43],

$$\mathcal{H}_e = \frac{1}{16}\sum_\gamma \sum_{\langle ij\rangle\parallel\gamma} \Big\{ -\frac{8}{5}\frac{t^2}{\varepsilon(^6A_1)}\Big(\vec{S}_i\cdot\vec{S}_j+6\Big)\mathcal{P}^{(\gamma)}_{\langle ij\rangle} + \Big[\frac{t^2}{\varepsilon(^4E)}+\frac{3}{5}\frac{t^2}{\varepsilon(^4A_1)}\Big]\Big(\vec{S}_i\cdot\vec{S}_j-4\Big)\mathcal{P}^{(\gamma)}_{\langle ij\rangle} + \Big[\frac{t^2}{\varepsilon(^4E)}+\frac{t^2}{\varepsilon(^4A_2)}\Big]\Big(\vec{S}_i\cdot\vec{S}_j-4\Big)\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}\Big\} + E_z\sum_i \tau_i^{(c)}. \quad (42)$$

In addition, t_{2g} electrons also contribute with $\mathcal{H}_t = \frac{1}{8}J\beta r_t\Big(\vec{S}_i\cdot\vec{S}_j-4\Big)$. Here $\beta = (t_\pi/t)^2$ follows from the difference between the effective d-d hopping elements along the σ and π

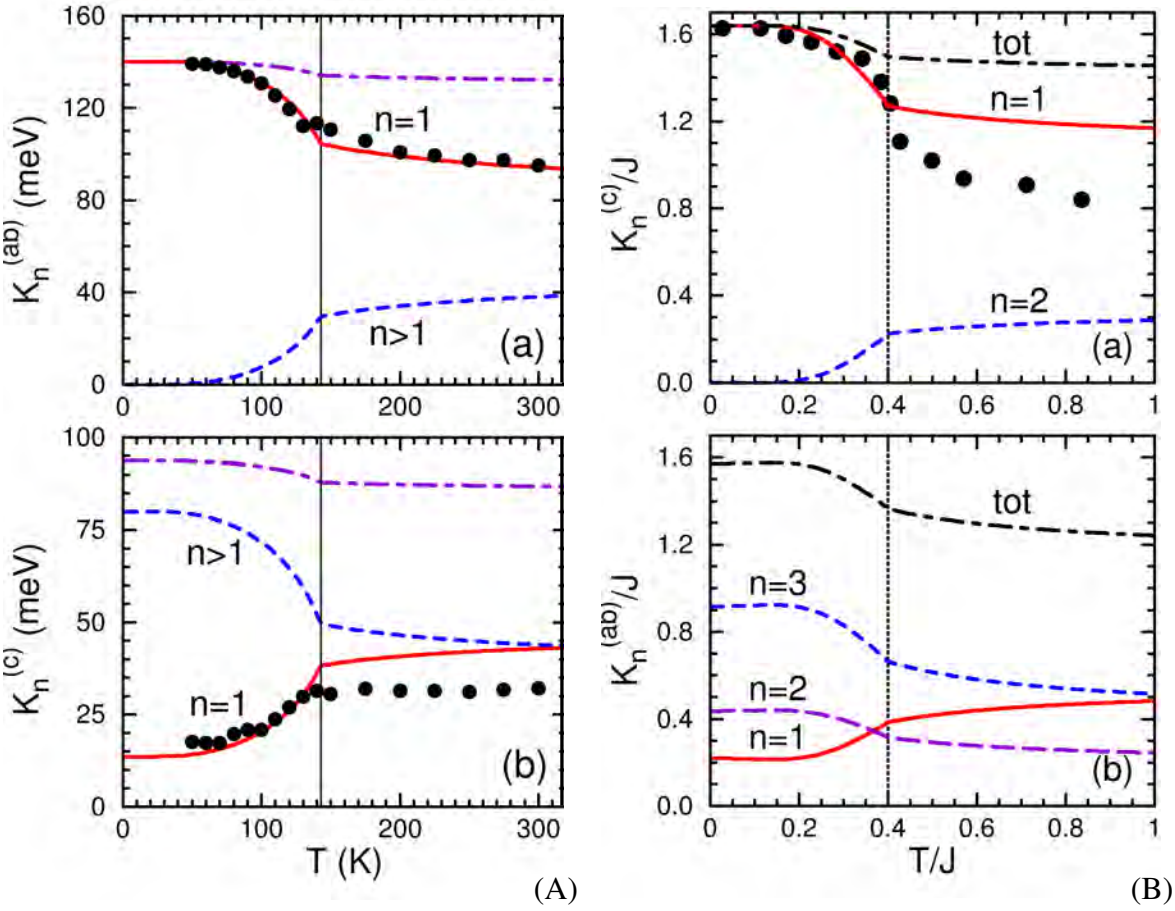


Fig. 13: *Kinetic energies per bond $K_n^{(\gamma)}$ Eq. (32) for increasing temperature T obtained from the respective spin-orbital models for FM (top) and AF (bottom) bonds along the axis γ: (A) LaMnO$_3$ (with $J = 150$ meV, $\eta \simeq 0.18$ [28], and experimental points [33]); (B) LaVO$_3$ with η=0.13 [35] and experimental points [44]. The kinetic energies in HS states ($n = 1$, red lines) are compared with the experiment (filled circles). Vertical dotted lines indicate values of $T_{\rm N}$. Images reproduced from Ref. [28].*

bonds, i.e., $\beta \simeq \frac{1}{9}$, while the coefficient r_t stands for a superposition of all t_{2g} excitations involved in the t_{2g} superexchange [28]. Note that spin-projection operators for high (low) total spin $\mathcal{S} = 2$ ($\mathcal{S} = 1$) cannot be used, but again the HS term stands for a FM contribution which dominates over the LS terms when $\left\langle \mathcal{P}_{\langle ij\rangle}^{(\gamma)} \right\rangle \simeq 1$. Charge excitations by t_{2g} electrons give double occupancies in active t_{2g} orbitals, so $\mathcal{H}_t$ is AF but this term is small—as a result FM interactions may dominate but again only along two spatial directions. Indeed, this happens for the realistic parameters of LaMnO$_3$ for the ab planes where SO is FM and coexists with AO order, while along the c axis SO is AF accompanied by FO order, in agreement with GKR, i.e., spin-orbital order is A-AF/C-AO. Indeed, this type of order is found both from the theory for realistic parameters and from electronic structure calculations [45]. The JT orbital interactions are responsible for the enhanced value of the orbital transition temperature [46].

Spin- and orbital-energy scale separately here, and the OO is mainly triggered by JT distortions [45]. The optical spectral weight due to HS states in LaMnO$_3$ may be easily derived from

the present model (42), following the general theory, see Eq. (32). One finds a very satisfactory agreement between the present theory and the experimental results [33], as shown in Fig. 13(A). We emphasize, that no fit is made here, i.e., the kinetic energies (32) are calculated using the same parameters as those used elsewhere for the magnetic exchange constants [28]. Therefore, such a good agreement with experiment suggests that indeed the spin-orbital superexchange may be disentangled for large $S = 2$ spins. Summarizing, we have found that spin-orbital entanglement is weak in this case [46]. *A posteriori*, this conclusion could be also drawn from a good agreement of spin excitations predicted by the theory with experimental data [47].

5 Spin-orbital entanglement in t_{2g} electron models

5.1 Entangled phases of $LaVO_3$ and YVO_3

In this case one uses the degenerate Hubbard model for three t_{2g} orbitals with J_H^t (22) [48]. Spin-orbital entanglement is stronger for t_{2g} than for e_g systems [29]. Due to large Coulomb interaction, the spin-orbital entangled state in RVO$_3$ ($R =$ La,. . . ,Lu) satisfies in a Mott insulator the local constraint at V^{3+} site i,

$$n_{ia} + n_{ib} + n_{ic} = 2, \tag{43}$$

and G-type OO competes with the spin-orbital entangled state. Rare earth site disorder favors the spin-orbital entanglement rather than a cooperative JT distortion [49]. The entanglement is best seen in the coupling between the spin and orbital phase transition [50]. Due to Hund's exchange J_H, one has here coupled $S = 1$ spins and $\tau = 1/2$ orbitals for three ($n = 1, 2, 3$) charge excitations ε_n arising from the transitions to [see Fig. 6(b)]:

(i) a high-spin state $^4A_2$ at energy $U-3J_H$,
(ii) two degenerate low-spin states $^2T_1$ and ^2E at U, and
(iii) a $^2T_2$ low-spin state at $U+2J_H$ [16].

Using η (25), we parametrize this multiplet structure by r_1, Eq. (36), and the top multiplet state,

$$r_5 = \frac{1}{1+2\eta}. \tag{44}$$

The cubic symmetry is broken and the CF induces orbital splitting in RVO$_3$, hence $\langle n_{ic}\rangle = 1$ and the orbital degrees of freedom are given by the doublet $\{a, b\}$, with $n_{ia} + n_{ib} = 1$, which defines the pseudospin operators $\vec{\tau}_i$ at site i. One derives a HS contribution $H_1^{(c)}(ij)$ for a bond $\langle ij\rangle$ along the c axis, and $H_1^{(ab)}(ij)$ for a bond in the ab plane:

$$H_1^{(c)}(ij) = -\tfrac{1}{3}Jr_1\big(\vec{S}_i\cdot\vec{S}_j + 2\big)\big(\tfrac{1}{4} - \vec{\tau}_i\cdot\vec{\tau}_j\big), \tag{45}$$

$$H_1^{(ab)}(ij) = -\tfrac{1}{6}Jr_1\big(\vec{S}_i\cdot\vec{S}_j + 2\big)\big(\tfrac{1}{4} - \tau_i^z\tau_j^z\big). \tag{46}$$

In Eq. (45) pseudospin operators $\vec{\tau}_i$ describe the low-energy dynamics of (initially degenerate) $\{xz, yz\}$ orbital doublet at site i; this dynamics is quenched in the plane, see $H_1^{(ab)}$ Eq. (46).

Here $\frac{1}{3}(\vec{S}_i\cdot\vec{S}_j+2)$ is the projection operator on the HS state for $S=1$ spins. The terms $H_n^{(c)}(ij)$ for LS excitations ($n=2,3$) contain instead the spin operator $(1-\vec{S}_i\cdot\vec{S}_j)$ (which guarantees that these terms cannot contribute for fully polarized spins $\langle\vec{S}_i\cdot\vec{S}_j\rangle=1$):

$$\begin{aligned} H_2^{(c)}(ij) &= -\tfrac{1}{12}\,J\,\big(1-\vec{S}_i\cdot\vec{S}_j\big)\big(\tfrac{7}{4}-\tau_i^z\tau_j^z-\tau_i^x\tau_j^x+5\tau_i^y\tau_j^y\big),\\ H_3^{(c)}(ij) &= -\tfrac{1}{4}Jr_5\big(1-\vec{S}_i\cdot\vec{S}_j\big)\big(\tfrac{1}{4}+\tau_i^z\tau_j^z+\tau_i^x\tau_j^x-\tau_i^y\tau_j^y\big). \end{aligned} \tag{47}$$

Again the terms $H_n^{(ab)}(ij)$ differ from $H_n^{(c)}(ij)$ only by the orbital operators,

$$\begin{aligned} H_2^{(ab)}(ij) &= -\tfrac{1}{8}\,J\,\big(1-\vec{S}_i\cdot\vec{S}_j\big)\big(\tfrac{19}{12}\mp\tfrac{1}{2}\tau_i^z\mp\tfrac{1}{2}\tau_j^z-\tfrac{1}{3}\tau_i^z\tau_j^z\big),\\ H_3^{(ab)}(ij) &= -\tfrac{1}{8}Jr_5\big(1-\vec{S}_i\cdot\vec{S}_j\big)\big(\tfrac{5}{4}\mp\tfrac{1}{2}\tau_i^z\mp\tfrac{1}{2}\tau_j^z+\tau_i^z\tau_j^z\big), \end{aligned} \tag{48}$$

where upper (lower) sign corresponds to bonds along the a (b) axis.
First, we present a mean field approximation for the spin and orbital bond correlations which are determined self-consistently after decoupling them from each other in $\mathcal{H}_J$ (28). Spin interactions in Eq. (29) are given by two exchange constants:

$$\begin{aligned} J_c &= \tfrac{1}{2}J\left\{\eta r_1-(r_1-\eta r_1-\eta r_5)(\tfrac{1}{4}+\langle\vec{\tau}_i\cdot\vec{\tau}_j\rangle)-2\eta r_5\langle\tau_i^y\tau_j^y\rangle\right\},\\ J_{ab} &= \tfrac{1}{4}J\left\{1-\eta r_1-\eta r_5+(r_1-\eta r_1-\eta r_5)(\tfrac{1}{4}+\langle\tau_i^z\tau_j^z\rangle)\right\}, \end{aligned} \tag{49}$$

determined by orbital correlations $\langle\vec{\tau}_i\cdot\vec{\tau}_j\rangle$ and $\langle\tau_i^\alpha\tau_j^\alpha\rangle$. By evaluating them one finds $J_c<0$ and $J_{ab}>0$ and the C-AF SO is supported.
In the orbital sector one finds at the same time,

$$H_\tau = \sum_{\langle ij\rangle_c}\left[J_c^\tau\vec{\tau}_i\cdot\vec{\tau}_j-J(1-s_c)\eta r_5\tau_i^y\tau_j^y\right]+J_{ab}^\tau\sum_{\langle ij\rangle_{ab}}\tau_i^z\tau_j^z, \tag{50}$$

$$J_c^\tau = \tfrac{1}{2}J\left[(1+s_c\,)r_1+(1-s_c\,)\eta(r_1{+}r_5)\right], \tag{51}$$

$$J_{ab}^\tau = \tfrac{1}{4}J\left[(1-s_{ab})r_1+(1+s_{ab})\eta(r_1{+}r_5)\right], \tag{52}$$

depending on spin correlations: $s_c=\langle\vec{S}_i\cdot\vec{S}_j\rangle_c$ and $s_{ab}=-\langle\vec{S}_i\cdot\vec{S}_j\rangle_{ab}$. In a classical C-AF state ($s_c=s_{ab}=1$) this mean field procedure becomes exact, and the orbital problem maps to Heisenberg pseudospin chains along the c axis, weakly coupled (as $\eta\ll1$) along a and b bonds,

$$H_\tau^{(0)}=Jr_1\left[\sum_{\langle ij\rangle_c}\vec{\tau}_i\cdot\vec{\tau}_j+\frac{1}{2}\eta\left(1+\frac{r_5}{r_1}\right)\sum_{\langle ij\rangle_{ab}}\tau_i^z\tau_j^z\right], \tag{53}$$

releasing large zero-point energy. Thus, spin C-AF and G-AO order with quasi-1D orbital quantum fluctuations support each other in RVO$_3$. Orbital fluctuations play here a prominent role and amplify the FM exchange J_c, making it even stronger than the AF exchange J_{ab} [16].
Having the individual terms $H_n^{(\gamma)}$ of the spin-orbital model, one may derive the spectral weights of optical spectra, see Eq. (32). The HS excitations have remarkable temperature dependence and the spectral weight decreases in the vicinity of the magnetic transition at $T_{\rm N}$, see Fig. 13(B). The observed behavior is reproduced in the theory only when spin-orbital interactions are treated in a cluster approach, i.e., they *cannot* be disentangled.

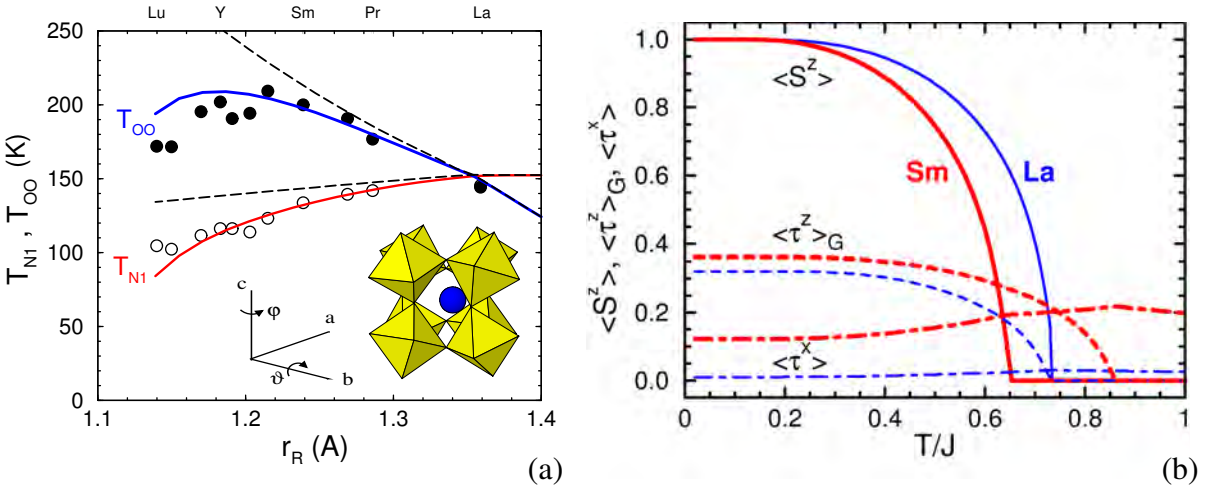


Fig. 14: *Phase transitions in the vanadium perovskites RVO_3: (a) phase diagram with the orbital $T_{\rm OO}$ and Néel $T_{\rm N1}$ transition temperatures obtained from the theory with and without orbital-lattice coupling (solid and dashed lines) [50], and from experiment (circles) [51]; (b) spin $\langle S_i^z\rangle$ (solid) and G-type orbital $\langle\tau_i^z\rangle_G$ (dashed) order parameters, vanishing at $T_{\rm N1}$ and $T_{\rm OO}$, respectively, and the transverse orbital polarization $\langle\tau_i^x\rangle$ (dashed-dotted lines) for $LaVO_3$ and $SmVO_3$ (thin and heavy lines). Images reproduced from Ref. [50].*

Unlike in $LaMnO_3$ where the spin and orbital phase transitions are well separated [46], in the RVO_3 ($R =$ Lu, Yb,. . . ,La) the two above transitions are close to each other [51]. It is not easy to reproduce the observed dependence of the transition temperatures $T_{\rm OO}$ and the Néel T_{N1} on the ionic radius r_R (in the RVO_3 compounds with small r_R there is also another magnetic transition at T_{N2} [52] which is not discussed here). The spin-orbital model was extended by the coupling to the lattice to unravel a nontrivial interplay between superexchange, the orbital-lattice coupling due to the $GdFeO_3$-like rotations of the VO_6 octahedra, and orthorhombic lattice distortions [50]. One finds that the lattice strain affects the onset of the magnetic and orbital order by partial suppression of orbital fluctuations, and the dependence of $T_{\rm OO}$ is non-monotonous, while T_{N1} is reduced, see Fig. 14(a). Thereby the orbital polarization $\propto \langle\tau^x\rangle$ increases with decreasing ionic radius r_R, see Fig. 14(b). The theoretical approach [50] demonstrates that orbital-lattice coupling is very important and reduces both $T_{\rm OO}$ and Néel T_{N1} for small ionic radii. Simultaneously, T_{N1} decreases to the left due to spin-orbital entanglement.

It has also been shown that the t_{2g} perovskite $LaVO_3$ is a unique case where the KK phase transition drives orbital order, in contrast to the usual case where the OO is controlled by the CF splitting enhanced by Coulomb interaction and both OO and SO transition are well separated [53]. As a consequence, the magnetic transition is close to (and even above) the superexchange driven OO order transition, and $T_{\rm N} > T_{\rm KK} \sim T_{OO}$, whereas typically magnetism arises at much lower temperatures than orbital ordering. In contrast, in YVO_3 the CF is sufficiently large to suppress the KK phase transition and spin-orbital interactions disentangle.

5.2 Spin-orbital entanglement on a triangular lattice

Finally, we wish to discuss here two rather interesting examples going also beyond the perovskite lattice, involving d^1 configurations. Two operators are entangled if their states at temperature $T = 0$ cannot be factorized into parts belonging to different subspaces. This happens precisely in some spin-orbital models and is the source of spin-orbital entanglement [29]. To verify whether entanglement occurs or not, it suffices to compute and analyze the spin, orbital and spin-orbital (four-operator) correlation functions for a bond $\langle ij\rangle$ along γ axis, given by

$$\begin{aligned}
S_{ij} &\equiv \tfrac{1}{d}\,\textstyle\sum_n \big\langle n\big|\vec{S}_i\cdot\vec{S}_j\big|n\big\rangle\,, &(54)\\
T_{ij} &\equiv \tfrac{1}{d}\,\textstyle\sum_n \big\langle n\big|(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\big|n\big\rangle\,, &(55)\\
C_{ij} &\equiv \tfrac{1}{d}\,\textstyle\sum_n \big\langle n\big|(\vec{S}_i\cdot\vec{S}_j - S_{ij})\,(\vec{T}_i\cdot\vec{T}_j - T_{ij})^{(\gamma)}\big|n\big\rangle &(56)\\
&= \tfrac{1}{d}\,\textstyle\sum_n \big\langle n\big|(\vec{S}_i\cdot\vec{S}_j)(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\big|n\big\rangle - \Big(\tfrac{1}{d}\textstyle\sum_n \big\langle n\big|\vec{S}_i\cdot\vec{S}_j\big|n\big\rangle\Big)\Big(\tfrac{1}{d}\textstyle\sum_m \big\langle m\big|(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\big|m\big\rangle\Big),
\end{aligned}$$

where d is the ground state degeneracy, and the pseudospin scalar product in Eqs. (55) and (56) is relevant for a model with active t_{2g} orbital degrees of freedom. As a representative example we evaluate here such correlations for a 2D spin-orbital model derived for a $NaTiO_2$ plane [54]; other situations with spin-orbital entanglement are discussed in Ref. [29].

To explain the physical origin of the spin-orbital model for $NaTiO_2$ [54], we consider a representative bond along the c axis shown in Fig. 15. For the realistic parameters of $NaTiO_2$ the t_{2g} electrons are almost localized in d^1 configurations of Ti^{3+} ions, hence their interactions with neighboring sites can be described by the effective superexchange and kinetic exchange processes. Virtual charge excitations between the neighboring sites, $d_i^1 d_j^1 \rightleftharpoons d_i^2 d_j^0$, generate magnetic interactions which arise from two different hopping processes for active t_{2g} orbitals: (i) the effective hopping $t = t_{pd}^2/\Delta$ which occurs via oxygen $2p_z$ orbitals with the charge transfer excitation energy Δ, in the present case along the 90° bonds, and (ii) direct hopping t' which couples the t_{2g} orbitals along the bond and gives a kinetic exchange interaction, as in the Hubbard model (2). Note that the latter processes couple orbitals with the same flavor, while the former ones couple different orbitals (for this geometry) so the occupied orbitals may be interchanged as a result of a virtual charge excitation—these processes are shown in Fig. 15.

The effective spin-orbital model considered here for $NaTiO_2$ reads [54],

$$\mathcal{H} = J\left((1-\alpha)\,\mathcal{H}_s + \sqrt{(1-\alpha)\alpha}\,\mathcal{H}_m + \alpha\,\mathcal{H}_d\right). \tag{57}$$

The parameter α in Eq. (57) is given by the hopping elements as follows,

$$\alpha = (t')^2/\left[t^2 + (t')^2\right]\,, \tag{58}$$

and interpolates between the superexchange $\mathcal{H}_s$ ($\alpha = 0$) and kinetic exchange $\mathcal{H}_d$ ($\alpha = 1$), while in between these two exchange elements and mixed exchange $\mathcal{H}_m$ contributes simultaneously; these terms are explained in Ref. [54]. This model is considered here in the absence of Hund's exchange η (25), i.e., at $\eta = 0$. One finds that all the orbitals contribute equally in the entire range of α, and each orbital state is occupied at two out of six sites in the entire regime of

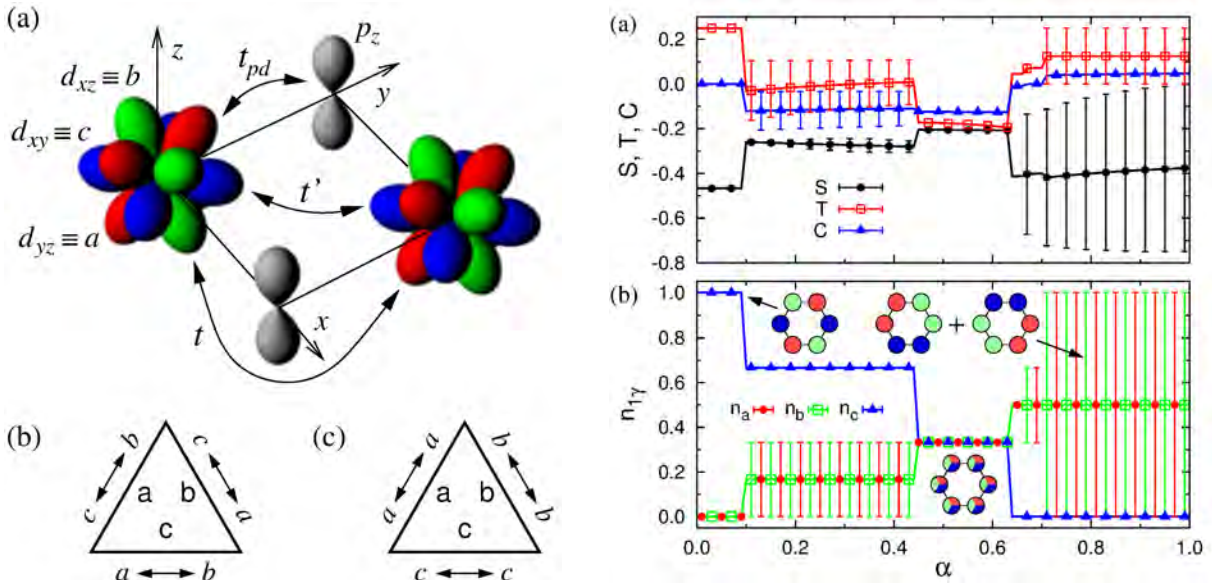


Fig. 15: *Left — (a) Hopping processes between t_{2g} orbitals along a bond parallel to the c axis in $NaTiO_2$: (i) t_{pd} between Ti(t_{2g}) and O($2p_z$) orbitals—two t_{pd} transitions define an effective hopping t, and (ii) direct d-d hopping t'. The t_{2g} orbitals (7) are shown by different color. The bottom part gives the hopping processes along the $\gamma = a, b, c$ axes that contribute to Eq. (57): (b) superexchange and (c) direct exchange. Right — Ground state for a free hexagon as a function of α: (a) bond correlations—spin S_{ij} Eq. (54) (circles), orbital T_{ij} Eq. (55) (squares), and spin–orbital C_{ij} Eq. (56) (triangles); (b) orbital electron densities $n_{1\gamma}$ at a representative site $i = 1$ (left-most site): n_{1a} (circles), n_{1b} (squares), n_{1c} (triangles). The insets indicate the orbital configurations favored by the superexchange $(\alpha = 0)$, by mixed interactions $0.44 < \alpha < 0.63$, and by the direct exchange $(\alpha = 1)$. The vertical lines indicate an exact range of configurations due to the degeneracy. Images reproduced from Ref. [55].*

α, see Fig. 15. The orbital state changes under increasing α and one finds as a result four distinct regimes, with abrupt transitions between them. In the superexchange model ($\alpha = 0$) there is precisely one orbital at each site which contributes, e.g. $n_{1c} = 1$ as the c orbital is active along both bonds. Having a frozen orbital configuration, the orbitals decouple from spins and the ground state is disentangled, with $C_{ij} = 0$, and one finds that the spin correlations $S_{ij} = -0.4671$, as for the AF Heisenberg ring of $L = 6$ sites. Orbital fluctuations increase gradually with increasing α and this results in finite spin-orbital entanglement $C_{ij} \simeq -0.12$ for $0.10 < \alpha < 0.44$; simultaneously spin correlations weaken to $S_{ij} \simeq -0.27$.

In agreement with intuition, when $\alpha = 0.5$ and all inter-orbital transitions shown in Fig. 15 have equal amplitude, there is large orbital mixing which is the most prominent feature in the intermediate regime of $0.44 < \alpha < 0.63$. Although spins are coupled by AF exchange, the orbitals fluctuate here strongly and reduce further spin correlations to $S_{ij} \simeq -0.21$. The orbital correlations are negative, $T_{ij} < 0$, the spin-orbital entanglement is finite, $C_{ij} \simeq -0.13$, and the ground state is unique ($d = 1$). Here all the orbitals contribute equally and $n_{1\gamma} = 1/3$ which may be seen as a precursor of the spin-orbital liquid state which dominates the behavior of the triangular lattice. The regime of larger values of $\alpha > 0.63$ is dominated by the kinetic exchange

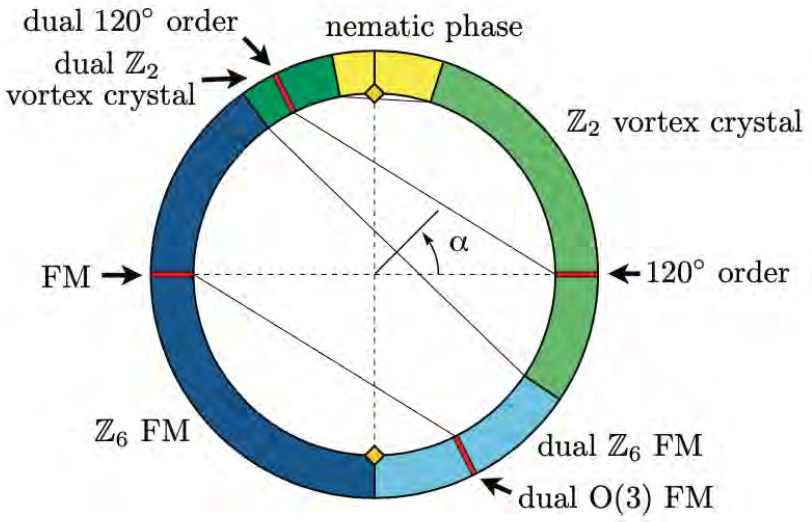


Fig. 16: *Phase diagram of the Kitaev-Heisenberg model Eq.* (59) *with parametrization* $(J, K) = (\cos\alpha, \sin\alpha)$ *as obtained from exact diagonalization data. Solid lines show the mapping between two Klein-dual points. Red lines mark the location of the four SU(2)—symmetric points. Yellow diamonds mark the two Kitaev points. Image reproduced from Ref. [56].*

in Eq. (57), and the ground state is degenerate with $d = 2$ [55], with strong scattering of possible electron densities $\{b_{i\gamma}\}$, see Fig. 15. Weak entanglement is found for $\alpha > 0.63$, where $C_{ij} \simeq 0$. Summarizing, except for the regimes of $\alpha < 0.09$ and $\alpha > 0.63$, the ground state of a single hexagon is strongly entangled, i.e., $C_{ij} < -0.10$, see Fig. 15.

As the last example we would like to highlight briefly the Heisenberg-Kitaev (HK) Hamiltonian on the triangular lattice [56]. Here spin-orbital entanglement is triggered by formation of effective $j = 1/2$ spins in a Mott insulator observed for the recently synthesized $Ba_2IrTi_2O_9$. The model is frustrated, both by its interactions and by geometry, see Fig. 16. The description of the microscopic physics is given here by a superposition of Heisenberg and Kitaev interaction,

$$\mathcal{H}_{HK} = J \sum_{\langle ij \rangle} \left(\vec{S}_i \cdot \vec{S}_j \right) + K \sum_{\gamma \| \langle ij \rangle} S_i^\gamma \, S_j^\gamma, \tag{59}$$

where $\vec{S}_i$ is a spin operator located on site i of the triangular lattice spanned by the lattice vectors $\vec{a}_x = (1, 0)^T$, $\vec{a}_y = (-1/2, \sqrt{3}/2)^T$, and $\vec{a}_z = -\vec{a}_x - \vec{a}_y$, for the lattice constant $a = 1$. The first term is the Heisenberg coupling $\propto J$, while the Kitaev term $\propto K$ explicitly breaks spin-rotation invariance and acts only between the same spin components $S_i^\gamma S_j^\gamma$ at nearest neighbor sites.

First, we observe that an infinitesimal Kitaev exchange removes the $120°$ order of the quantum Heisenberg model [56]. Second, the phase diagram of Fig. 16 is very rich and instead of $120°$ order, an extended $\mathbb{Z}_2$-vortex crystal phase arises which could be identified experimentally. Third, the phase diagram exhibits a duality, similar to the HK model on the honeycomb lattice [57]. This duality relates a pair of interactions on the right-hand side of the circle to a pair of interactions on the left-hand side, i.e., $J \to -J$ and $K \to 2J+K$. The corresponding dual states are related by a four-sublattice basis transformation. For more explanation see Ref. [56].

6 Experimental consequences of spin-orbital entanglement

The field of spin-orbital physics is recently growing and becoming richer due to new experiments. Understanding them is possible within simple models, mainly developed in low dimension. Also there entanglement is the strongest as quantum phenomena domininate in low dimension. We shall concentrate on them here as they uncover important principles of treating spin-orbital entanglement, both in the ground and in excited states. Spin-orbital entangled states occur in several quantum materials and the proper understanding of them becomes crucial for the quantitative analysis of the observed thermodynamic phase transitions.

Let us summarize briefly entanglement properties which could modify the experiment. First, I would like to recall the 1D chain, where each of the fractional quasiparticles carries both spin and orbital quantum numbers, and the two variables (spin and orbital) are always entangled in the collective excitations [58]. The spin-orbital chain with two orbital flavors has been carefully studied, both without the orbital spitting [e.g. in the SU(4) case] and for large splitting E_z.

Second, the 1D cuprate $CaCu_2O_3$ is a good example for the spin-orbital fractionalization along the chain direction, while at the same time no fractionalization is observed for the xy orbital which extends in both leg and rung direction [59]. Thus different degrees of dimensionality may be selected by orbitals and thus the entanglement depends on particular orbitals involved in the hopping. In a realistic 1D model for $CaCu_2O_3$ the splitting between the LS and HS terms occurs for finite Hund's exchange J_H. In fact, any spin-orbital model requires to include a finite, but realistic Hund's exchange to explain the experimental data [60].

Third, another important competitor to the spin-orbital physics and on-bond entanglement is the strong JT effect which is found as well in $KCuF_3$ [61]. It drives the orbital order and does change d-d excitations which become highly localized. At the same time, the low-energy excitations present clear dispersion. They match extremely well with the two-spinon continuum. So, we suggest that spin-orbital entanglement is a subtle property which is easily destroyed by some stronger local interaction, and the entanglement depends then strongly on the possible coupling to the lattice.

Fourth, spin-orbital coupling may lead to the disappearance of magnetic properties if it is sufficiently strong, removes the orbital degrees of freedom, and generates local singlets [62]. It is then challenging to study exchange interactions between these local singlets. In perovskites one finds a conventional Bose condensation of excitons into a magnetic state, while an unexpected 1D behavior supporting spin-liquid states emerges in honeycomb lattices. When spin and orbital channels are decoupled and orbital frustration induces then frustration in the spin channel, opening up the possibility of spin-orbital liquids with both spin and orbital entanglement [63].

Finally, the physics will change completely when spin-orbit interaction dominates over the superexchange. Then local singlets form and are only weakly coupled by inter-site terms [64]. As we have shown on the example of a 1D spin-orbital chain, entanglement is then strong but preferably limited to a single site, while the inter-site spin-orbital entanglement (which involves superexchange bonds) might be still induced as shown in Sec. 1.

7 Summary

Here we have focused on the interplay of spin and orbital degrees of freedom in realistic Mott insulators and have shown that even when the entanglement is absent, these two types of degrees of freedom (spin and orbital) decide about rather complex behavior, with competing tendencies to localize due to strong correlations and to delocalize to gain more kinetic energy. Quantum fluctuations are particularly well developed in the t_{2g} systems where they partly even destroy OO. As a result, a second order phase transition from the spin-orbital entangled state to a C-OO/G-AF ground state is induced in $LaVO_3$, where the long-range OO suppresses the spin-orbital entanglement [49]. On the other hand, entanglement may be easily removed by phase transitions, as a more fragile property of Mott insulators. It is near a phase transition that spin-orbital entanglement is quenched locally [41]. The detailed energy balance depends as well on the distribution of charge defects which also destroy spin-orbital entanglement locally.
Spin-orbital entanglement in the excited states is almost unexplored and awaits careful future studies. It may be responsible for the modified dispersion of spin (and orbital) excitations when either the spin or the orbital background is modified [42]. Yet, such modifications do not require that the local degrees of freedom factorize, but it suffices that their coupling is modified and causes measurable yet still unexplored properties.
Summarizing, spin-orbital entanglement is an important yet subtle property of Mott insulators. It is fragile and helps to understand how spins and orbitals complement each other and behave in the opposite way in ordered 3D materials. Perhaps the best example is the orbital liquid where orbital disorder coexists with FM order of spins in the ground state of CMR manganites. Then the spin-orbital entanglement is removed and the coherent spin FM order decouples from the orbitals [21]. Many properties of correlated insulators are still unexplored and hopefully will be investigated in the future. This concerns in particular the spin and orbital excitations. We should be prepared that this field has still some hidden surprises to discover and that some of them will be revealed gradually in the future.

Acknowledgments

I thank Wojciech Brzezicki, Louis Felix Feiner, Peter Horsch, Giniyat Khaliullin, Krzysztof Wohlfeld, Jan Zaanen, and Karol Życzkowski for numerous useful discussions and comments. I acknowledge National Science Centre (NCN, Poland) Project No. 2021/43/B/ST3/02166.

References

[1] K.I. Kugel and D.I. Khomskii, Sov. Phys. Usp. **25**, 231 (1982)

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

[3] D.I. Khomskii, *Transition Metal Compounds* (Cambridge University Press, Cambridge, 2014)

[4] Y. Yamashita, N. Shibata, and K. Ueda, Phys. Rev. B **58**, 9114 (1998)

[5] B. Frischmuth, F. Mila, and M. Troyer, Phys. Rev. Lett. **82**, 835 (1999)

[6] Y. Chen, Z.W. Wang, Y.Q. Li, and F.C. Zhang, Phys. Rev. B **75**, 195113 (2007)

[7] A.M. Oleś, Phys. Rev. B **28**, 327 (1983)

[8] K.A. Chao, J. Spałek, and A.M. Oleś, J. Phys. C **10**, L271 (1977)

[9] L.F. Feiner, A.M. Oleś, and J. Zaanen, Phys. Rev. Lett. **78**, 2799 (1997)

[10] L.F. Feiner, A.M. Oleś, and J. Zaanen, J. Phys.: Condens. Matter **10**, L555 (1998)

[11] A.M. Oleś, G. Khaliullin, P. Horsch, and L.F. Feiner, Phys. Rev. Lett. **96**, 147205 (2006)

[12] G. Khaliullin, Prog. Theor. Phys. Suppl. **160**, 155 (2005)

[13] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006)

[14] W. Witczak-Krempa, G. Chen, Y.B. Kim, and L. Balents, Annu. Rev. Condens. Matter Phys. **5**, 57 (2014)

[15] D. Gotfryd, E.M. Pärschke, J. Chaloupka, A.M. Oleś, and K. Wohlfeld, Phys. Rev. Research **2**, 013353 (2020)

[16] G. Khaliullin, P. Horsch, and A.M. Oleś, Phys. Rev. Lett. **86**, 3879 (2001)

[17] P. Horsch, G. Khaliullin, and A.M. Oleś, Phys. Rev. Lett. **91**, 257203 (2003)

[18] G. Jackeli and G. Khaliullin, Phys. Rev. Lett. **103**, 067205 (2009)

[19] N. Goldman and J. Dalibard, Phys. Rev. X **4**, 031027 (2014)

[20] C. Slater and G.F. Koster, Phys. Rev. **94**, 1498 (1954)

[21] L.F. Feiner and A.M. Oleś, Phys. Rev. B **71**, 144422 (2005)

[22] J. van den Brink, P. Horsch, F. Mack, and A.M. Oleś, Phys. Rev. B **59**, 6795 (1999)

[23] L.F. Feiner and A.M. Oleś, Phys. Rev. Research **4**, 043134 (2022)

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

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

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

[27] Y. Nagaoka, Phys. Rev. **147**, 392 (1966)

[28] A.M. Oleś, G. Khaliullin, P. Horsch, and L.F. Feiner, Phys. Rev. B **72**, 214431 (2005)

[29] A.M. Oleś, J. Phys.: Condens. Matter **24**, 313201 (2012)

[30] J.S. Griffith, *The Theory of Transition Metal Ions* (Cambridge University Press, Cambridge, 1971)

[31] E. Dagotto, T. Hotta, and A. Moreo, Physics Reports **344**, 1 (2001)

[32] Q. Huang, A. Santoro, J.W.Lynn, R.W. Erwin, J.A. Borchers, J.L. Peng, and R.L. Greene, Phys. Rev. B **55**, 14997 (1997)

[33] N.N. Kovaleva, A.M. Oleś, A.M. Balbashov, A. Maljuk, D.N. Argyriou, G. Khaliullin, and B. Keimer, Phys. Rev. B **81**, 235130 (2010)

[34] P. Czarnik, J. Dziarmaga, and A.M. Oleś, Phys. Rev. B **86**, 014420 (2017)

[35] G. Khaliullin, P. Horsch, and A.M. Oleś, Phys. Rev. B **70**, 195103 (2004)

[36] J.B. Goodenough, *Magnetism and the Chemical Bond* (Interscience, New York, 1963)

[37] W. Brzezicki, J. Phys.: Condens. Matter **32**, 313201 (2020)

[38] K. Wohlfeld, M. Daghofer, and A.M. Oleś, EPL—Europhys. Lett. **96**, 27001 (2011)

[39] K. Wohlfeld, M. Daghofer, S. Nishimoto, G. Khaliullin, and J. van den Brink, Phys. Rev. Lett. **107**, 147201 (2011)

[40] A.M. Oleś, L.F. Feiner, and J. Zaanen, Phys. Rev. B **61**, 6257 (2000)

[41] W. Brzezicki, J. Dziarmaga, and A.M. Oleś, Phys. Rev. Lett. **109**, 237201 (2012)

[42] M. Snamina and A.M. Oleś, New J. Phys. **21**, 023018 (2019)

[43] L.F. Feiner and A.M. Oleś, Phys. Rev. B **59**, 3295 (1999)

[44] S. Miyasaka, Y. Okimoto, and Y. Tokura, J. Phys. Soc. Jpn. **71**, 2086 (2002)

[45] E. Pavarini and E. Koch, Phys. Rev. Lett. **104**, 086402 (2010)

[46] M. Snamina and A.M. Oleś, Phys. Rev. B **94**, 214426 (2016)

[47] A.M. Oleś and L.F. Feiner, Phys. Rev. B **65**, 052414 (2002)

[48] N. Chikano, S. Hoshino, and H. Shinaoka, Phys. Rev. B **104**, 235125 (2021)

[49] J.-Q. Yan, W. Tian, H.B. Cao, S. Chi, F. Ye, A. Llobet, A. Puretzky, Q. Chen, J. Ma, Y. Ren, J.-G. Cheng, J.-S. Zhou, M.A. McGuire, and R.J. McQueeney, Phys. Rev. B **100**, 184423 (2019)

[50] P. Horsch, A.M. Oleś, L.F. Feiner, and G. Khaliullin, Phys. Rev. Lett. **100**, 167205 (2008)

[51] S. Miyasaka, Y. Okimoto, M. Iwama, and Y. Tokura, Phys. Rev. B **68**, 100406(R) (2002)

[52] J. Fujioka, T. Yasue, S. Miyasaka, Y. Yamasaki, T. Arima, H. Sagayama, T. Inami, K. Ishii, and Y. Tokura, Phys. Rev. B **82**, 144425 (2010)

[53] X.-J. Zhang, E. Koch, and E. Pavarini, Phys. Rev. B **106**, 115110 (2022)

[54] B. Normand and A.M. Oleś, Phys. Rev. B **78**, 094427 (2008)

[55] J. Chaloupka and A.M. Oleś, Phys. Rev. B **83**, 094406 (2011)

[56] M. Becker, M. Herrmanns, B. Bauer, M. Garst, and S. Trebst, Phys. Rev. B **91**, 155135 (2015)

[57] J. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **110**, 097204 (2013)

[58] C.-C. Chen, M. van Veenendaal, T.P. Devereaux, and K. Wohlfeld, Phys. Rev. B **91**, 165102 (2015)

[59] V. Bisogni, K. Wohlfeld, S. Nishimoto, C. Monney, J. Trinckauf, K. Zhou, R. Klaus, K. Koepernik, C. Sekar, V. Strocov, B. Büchner, T. Schmitt, J. van den Brink, and J. Geck, Phys. Rev. Lett. **114**, 096402 (2015)

[60] R. Fumagalli, J. Heverhagen, D. Betto, R. Arpaia, M. Rossi, D. Di Castro, N.B. Brookes, M. Moretti Sala, M. Daghofer, L. Braicovich, K. Wohlfeld, and G. Ghiringhelli, Phys. Rev. B **101**, 205117 (2020)

[61] J. Li, L. Xu, M. Garcia-Fernandez, A. Nag, H.C. Robarts, A.C. Walters, X. Liu, J. Zhou, K. Wohlfeld, J. van den Brink, H. Ding, and K.-J. Zhou, Phys. Rev. Lett. **126**, 106401 (2021)

[62] G. Khaliullin, Phys. Rev. Lett. **111**, 197201 (2013)

[63] C. Svoboda, M. Randeria, and N. Trivedi, Phys. Rev. B **95**, 014409 (2017)

[64] E.M. Plotnikova, M. Daghofer, J. van den Brink, and K. Wohlfeld, Phys. Rev. Lett. **116**, 106401 (2016)

7 Imaging Orbitals with X-rays

Liu Hao Tjeng
Max-Planck Institute for Chemical Physics of Solids
Nöthnitzer Str. 40, 01187 Dresden

Contents

1 Introduction **2**

2 Non-resonant inelastic scattering using an s core hole **2**

3 Experimental set-up **3**

4 Compton intensity and data treatment **4**

5 Orientational dependence of the s-NIXS intensities **5**

6 Ground state properties **7**

7 From ground state to excited states **8**

8 Analysis of excited states **10**

9 Many-body energy-level diagram **11**

10 Spectral lineshape **13**

11 Covalency **14**

12 Metallic systems: ground state properties **16**

13 Concluding remarks **17**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

Strongly correlated materials show a wide variety of physical phenomena which include unconventional superconductivity, heavy fermion behavior, various forms of magnetism and multiferroicity, as well as colossal magneto-resistance and metal-insulator transitions. These materials usually contain transition-metal, rare-earth or actinide elements: the presence of the atomic-like d or f orbitals provides the degrees of freedom to generate those intriguing phenomena. Here the intricate interplay of band formation with the local correlation and atomic multiplet effects leads to phases that are nearly iso-energetic, making the materials properties highly tunable by doping, temperature, pressure or magnetic field. Understanding the behavior of the d and f electrons is therefor essential. The orbitals that actively participate in the formation of the ground state and low energy excitations need to be identified. So far, these orbitals have mostly been deduced from optical, x-ray and neutron spectroscopies in which spectra must be analyzed using theory or modelling. This, however, is also a challenge in itself, since *ab-initio* calculations hit their limits due to the many-body nature of the problem.

Here we developed a new experimental method that circumvents the need for involved analysis and instead provides the information as measured [1–3]. With this technique, we can make a direct image of the active orbital and determine what the actual atomic-like object looks like in the solid of interest. The method is based on non-resonant inelastic x-ray scattering (NIXS, also known as x-ray Raman scattering). Essential is that we involve an s core hole (s-NIXS) in the experiment as we will explain below. The transitions $s \rightarrow d$ and $s \rightarrow f$ are, however, dipole-forbidden, and therefore we have to utilize high momentum transfers in the inelastic scattering process so that the beyond-dipole terms gain spectral weight. In this chapter we will present the basic principles of s-NIXS and its experimental implementation. To demonstrate the strength of the technique, we will show how we can obtain the image of a textbook example, the $3d(x^2-y^2)(3z^2-r^2)$ hole orbital of the Ni^{2+} ion in a NiO single crystal [1]. We will also illustrate how s-NIXS can be used to directly identify the orbital character of excited states so that the relevant energy parameters which determine the low energy excitations can be extracted [3]. We will explain how s-NIXS can unveil covalency effects [3] and show the power of the direct-imaging capability of s-NIXS for the study of metallic systems where strong inter-site charge fluctuations hamper a reliable quantitative analysis of spectroscopic data.

2 Non-resonant inelastic scattering using an s core hole

The theoretical description of inelastic x-ray scattering can be found in a number of publications, see e.g. [4–7]. The double differential cross-section is the product of the Thomson photon cross section $\left(\frac{d\sigma}{d\Omega}\right)_{Tho}$ and the dynamical structure factor $S(\vec{q},\omega)$

$$\frac{d^2\sigma}{d\Omega d\omega} = \left(\frac{d\sigma}{d\Omega}\right)_{Tho} S(\vec{q},\omega). \tag{1}$$

The dynamical structure factor is a function of the scattering vector $\vec{q} = \vec{k_i} - \vec{k_f}$ and the energy loss $\omega = \omega_i - \omega_f$

$$S(\vec{q},\omega) = \sum_f \left|\left\langle f\left|e^{i\vec{q}\cdot\vec{r}}\right|i\right\rangle\right|^2 \delta(\hbar\omega_i - \hbar\omega_f - \hbar\omega). \quad (2)$$

Here i and f are the initial and final states. The transition operator $e^{i\vec{q}\cdot\vec{r}}$ can be expressed by its Taylor terms $1 + i\vec{q}\cdot\vec{r} + (i\vec{q}\cdot\vec{r})^2 + \cdots$, thereby showing directly that for small values of $|\vec{q}|$ the monopole and dipole transitions will dominate the signal while for large $|\vec{q}|$ values the beyond-dipole transitions will also contribute significantly. In our experiment, we will utilize excitations involving a core level. It is therefore useful to expand the transition operator $e^{i\vec{q}\cdot\vec{r}}$ in terms of semi-normalized (Racah's normalization) spherical harmonics $C^{\hat{q}^*}_{km}$ and $C^{\hat{r}}_{km}$. This results in a sum over spherical Bessel functions $j_k(\vec{q}\cdot\vec{r})$ and the wave functions can be factorized into a radial and angular part so that $S(\vec{q},\omega)$ can be written as

$$S(\vec{q},\omega) = \sum_f \left| \sum_k^{\infty} i^k (2k+1) \left\langle R_f \left| j_k(\vec{q}\cdot\vec{r}) \right| R_i \right\rangle \sum_{m=-k}^{k} \left\langle \phi_f \left| C^{\hat{q}^*}_{km} C^{\hat{r}}_{km} \right| \phi_i \right\rangle \right|^2 \delta(\hbar\omega_i - \hbar\omega_f - \hbar\omega)\,. \quad (3)$$

Not all terms in the above equation contribute to the sum. In fact, the non-zero terms are subject to the so-called triangle condition and parity selection rules

$$\left|l_f - l_i\right| \le k \le l_f + l_i \quad \text{and} \quad \left|l_f + l_i + k\right| = \text{even}, \quad (4)$$

where $l_i \rightarrow l_f$ labels the angular momenta in the transition. For example, a $d \rightarrow f$ transition has contributions only from terms with $k = 1$ (dipole), 3 (octupole), and 5 (triakontadipole). An $s \rightarrow d$ or $s \rightarrow f$ transition, on the other hand, has only one term each, namely $k = 2$ (quadrupole) or $k = 3$ (hexapole).

In the following we focus our attention to transitions involving an s core hole. Our particular interest originates from the fact that the s core hole is spherically symmetric. With the scattering vector $\vec{q}$ being the only quantity that determines the quantization axis, we can expect that the intensity of the transition will be determined by the hole density of the atomic-like final state in the direction of the scattering vector. In other words, we may expect that by rotating the atomic-like object while keeping the measurement geometry fixed, we can obtain an intensity profile that reflects the shape of the hole density of that object. A mathematical proof that the intensity profile is proportional to the hole density is given by Sundermann and Haverkort [8].

3 Experimental set-up

The s-NIXS measurements were performed at the High-Resolution Dynamics Beamline P01 of the PETRA-III synchrotron facility in Hamburg, Germany. Fig. 1 illustrates the experimental setup, showing the incoming beam $(\vec{k}_i, \omega_i)$, single crystal sample, scattered beam $(\vec{k}_f, \omega_f)$, and the corresponding momentum transfer vector $(\vec{q})$. The energy of the x-ray photon beam incident on the sample was tuned with a Si(311) double-reflection crystal monochromator (DCM). The photons scattered from the sample were collected and energy-analyzed by an array of twelve

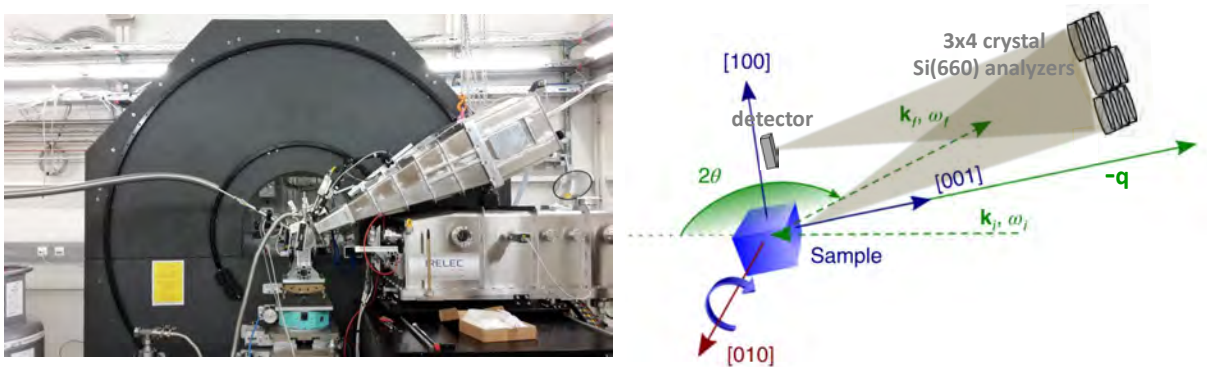


Fig. 1: *Experimental set-up and scattering geometry of the s-NIXS measurements at the P01 Beamline of the PETRA-III synchrotron facility in Hamburg, Germany. Schematic representation (right) from [1, 3].*

spherically bent Si(660) crystal analyzers. The analyzers are arranged in a 3×4 configuration. The energy of the analyzers ($\hbar\omega_f$) was fixed at 9690 eV; the energy loss spectra were measured by scanning the energy of the DCM ($\hbar\omega_i$). Each analyzer signal was individually recorded by a position-sensitive custom-made LAMBDA detector. The energy calibration was regularly checked by measuring the zero-energy-loss position of each spectrum. The best possible energy resolution was guaranteed by pixel-wise analysis of the detector recordings and measured as 0.7 eV (FWHM).

The positioning of the analyzer array determines the momentum transfer vector and the corresponding scattering triangle, which is defined by the incident and scattered photon momentum vectors, $\vec{k}_i$ and $\vec{k}_f$, respectively. The large scattering angle ($2\theta \approx 155°$) chosen for the current study assured a large momentum transfer of $|\vec{q}| = (9.6 \pm 0.1)\,\text{Å}^{-1}$ when averaged over all analyzers. $\vec{k}_f$ and 2θ were kept constant by fixing the energy and the position of the analyzer array. Since the energy transfer range of interest (100 to 120 eV) was small with respect to the incident and final energies ($\sim$9700 eV), variation of $\vec{k}_i$ during energy scanning was insignificant. This guaranteed that the scattering triangle was virtually unchanged throughout the course of the experiment with $|\vec{q}| \approx$ constant.

The core of the experimental work is the measurement of $S(\vec{q}, \omega)$ as a function of the orientation of the single crystal sample with respect to the fixed experimental geometry as shown in Fig. 1. Here we define the sample angle φ as the angle between the fixed momentum transfer vector $\vec{q}$ and the single crystal surface normal.

4 Compton intensity and data treatment

To demonstrate the feasibility and accuracy of the s-NIXS method, we used a single crystal of NiO as a model system. It crystallizes in the rocksalt structure and is an antiferromagnetic insulator, with a well-understood high-spin Ni d^8 configuration. The measurements were carried out at 20 K in a He gas cryostat. The left panel of Fig. 2 shows a collection of NIXS spectra of a NiO single crystal measured for many different sample angles. The spectra show the sharp

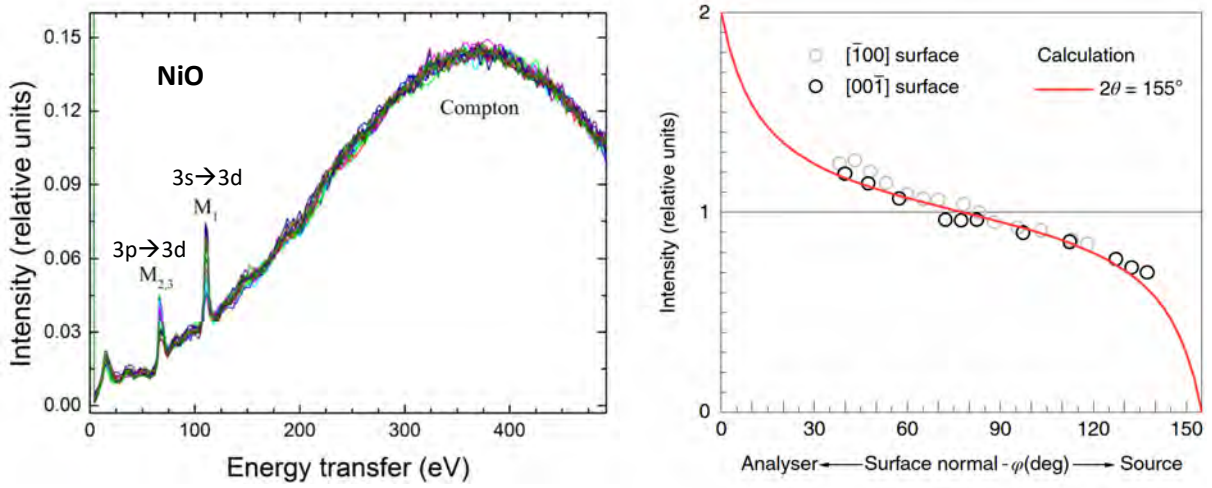


Fig. 2: *(left panel) Experimental NIXS spectra of NiO for a variety of crystal rotations φ with respect to the geometrically fixed momentum transfer vector $\vec{q}$. (right panel) Calculated and measured Compton intensity as function of sample angle φ with respect to specular geometry ($\varphi = 77.5°$) for a scattering angle $2\theta = 155°$. Reproduced from [1]*

$M_{2,3}$ edge ($3p \rightarrow 3d$) of nickel at around 70 eV and, important for the present study, the dipole-forbidden M_1 ($3s \rightarrow 3d$) excitations at around 110 eV. These features are on top of the broad Compton profile which peaks at about 350 eV energy transfer. The spectra are all normalized to this Compton signal.

We note that the line shape of the Compton profile does not change with angle. This is fully consistent with the fact that the scattering geometry is kept constant while rotating the sample. What does vary is the intensity of the Compton signal. This is related to how the x-rays are absorbed when entering the sample and when scattered out of the sample. This process can be modeled quantitatively since the NiO crystals used have well defined flat and shiny surfaces. The result of this modeling and the comparison with the experiment is displayed in the right panel of Fig. 2. The calculations were done for the actual scattering geometry of $2\theta = 155°$ so that for $\varphi = 77.5°$ specular geometry is fulfilled. Turning the sample towards (away from) the analyzer, increases (decreases) the intensity. One can observe a very good overall match between the experiment and the modeling. We thus can state that we understand fully the details of the scattering process, and in particular, that the intensity variations of the Compton profile is due to the absorption processes in the sample. We therefore can safely use the Compton profile to normalize our s-NIXS spectra.

To extract the M_1 edge intensities for further analysis, the Compton profile has been subtracted from each spectrum using a simple linear background.

5 Orientational dependence of the s-NIXS intensities

A close-up of the M_1 edge and its orientational dependence on $\vec{q} \parallel [001] - [101] - [100]$ is displayed in panel (a) of Fig. 3, and for $\vec{q} \parallel [001] - [111] - [110]$ in panels (b) and (c). The spectra consist of a single line and the line shape does not change with the angle. What does

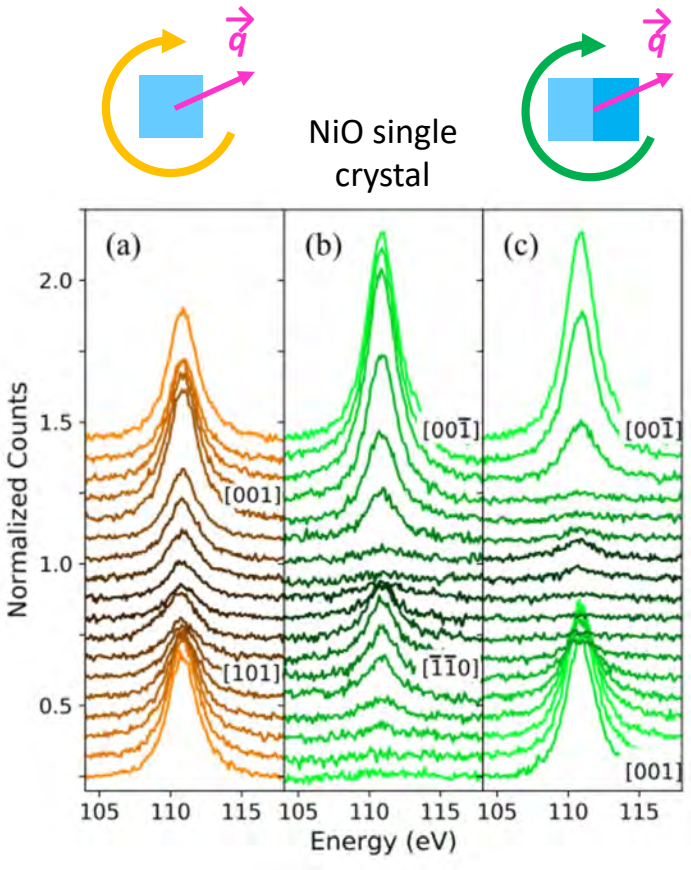


Fig. 3: *Stack of the Ni M_1 NIXS spectra acquired for $\vec{q}$ ∥ [001] – [101] – [100] (orange, panel a) and $\vec{q}$ ∥ [001] – [111] – [110] (green, panels b and c). Adapted from [1].*

change is the intensity. To quantitatively analyze these intensity variations, we determined the integrated intensity of each spectrum and placed it on a polar plot as displayed in the left and right panels of Fig. 4: the data points for $\vec{q}$ sweeping in the [001] – [101] – [100] plane are in orange, and for $\vec{q}$ in the [001] – [111] – [110] plane in green. They fall accurately on top of the orbital shapes which denote 'cuts' through the [001] – [101] – [100] (orange) and [001] – [111] – [110] (green) planes of the calculated three-dimensional orbital hole-density (square of the wave function) of the Ni high-spin $3d^8$ configuration in octahedral coordination, namely the $^3A_2$ $3d(x^2-y^2)3d(3z^2-r^2)$ as shown in the center panel of Fig. 4. This means that we indeed have generated a purely experimental method that can directly visualize the fundamental atomic-like quantum mechanical objects in solids. The information that we have obtained is extremely detailed; for example, we can clearly see the small lobes of the $3d(3z^2-r^2)$ contribution. We would like to remark that the $3d(x^2-y^2)$ contribution vanishes in the [001] – [111] – [110] plane, while both the $3d(x^2-y^2)$ and $3d(3z^2-r^2)$ contribute in the [001] – [101] – [100] plane. Here we note that we have used two orbital shapes in the inset in the right panel of Fig. 4: the blue dashed line is the $3d(x^2-y^2)3d(3z^2-r^2)$ function and the solid line is the same function convoluted with the angular acceptance of the 3×4 analyzers we used in our experiment. The near perfect agreement further demonstrates the accuracy of the method.

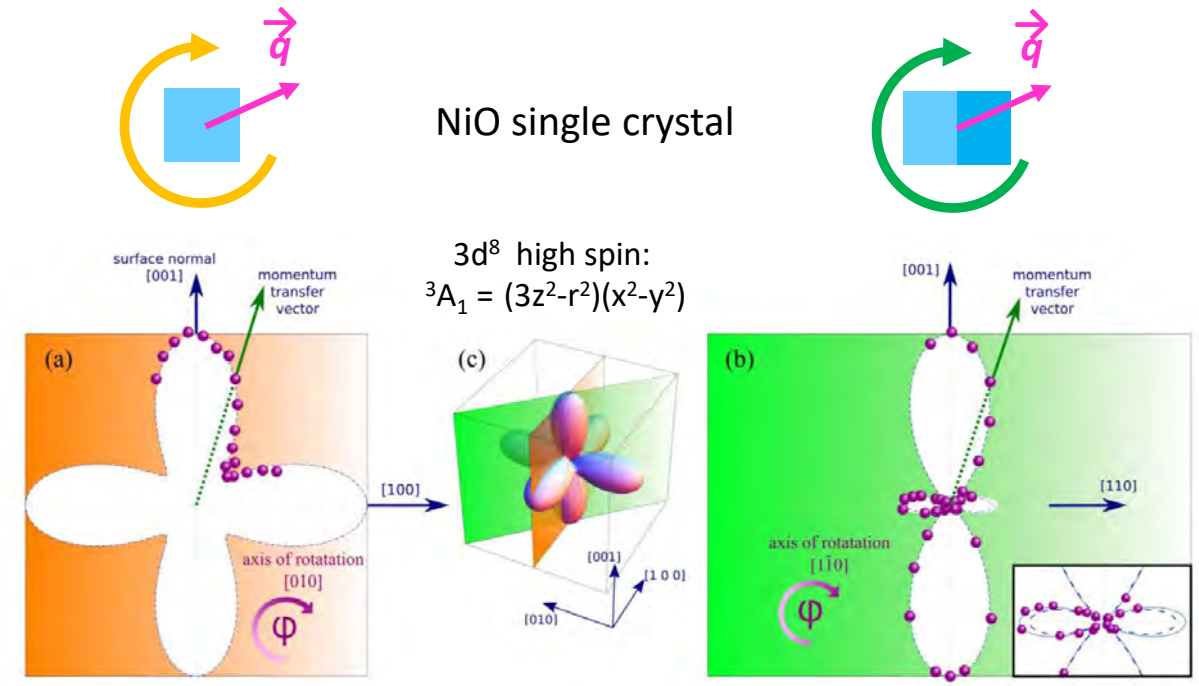


Fig. 4: *Left and right panels: polar plots of the integrated Ni M_1 intensities for $\vec{q}$ sweeping in the [001] – [101] – [100] plane (orange) and for $\vec{q}$ in the [001] – [111] – [110] plane (green). Central panel: theoretical three-dimensional orbital hole density of the Ni high-spin $3d^8$ configuration in octahedral coordination: $^3A_2$ $3d(x^2-y^2)3d(3z^2-r^2)$. Adapted from [1].*

6 Ground state properties

The orientational dependence of the integrated s-NIXS intensity at the Ni M_1 edge ($3s \rightarrow 3d$) directly maps the local orbital hole density of the ion in the ground state. There is no need to carry out a multiplet analysis of the spectral line shape to extract this information, in contrast to, for example, the non-s edges (e.g. $L_{2,3}(2p)$, $M_{2,3}$ $(3p)$, $M_{4,5}$ $(3d)$, $N_{4,5}$ $(4d)$, and $O_{4,5}$ $(5d)$) in both NIXS [5–7, 9–14] and XAS experiments [15–19]. The reason is fundamental: The M_1 $(3s \rightarrow 3d)$ quadrupolar excitation process involves a spherically symmetric s orbital, so the angular distribution of the intensity is solely determined by the hole charge distribution in the initial state with respect to the sole quantization axis which is given by the momentum transfer $\vec{q}$ [8]. This is similar to the dipole-allowed $s \rightarrow p$ transition in XAS, where an angular sweep of the polarization dependence maps out the orientation of the p hole directly.

We would like to emphasize that we do not need to perform complex configuration-interaction calculations to analyze or simulate the line shape of the spectra in order to obtain the local orbital hole density of the ion in the ground state. Details of the s-NIXS final states in fact do not matter because the information is extracted from the *integrated* intensity of the spectra (i.e., from the sum of the intensities of all final states). This can be understood as follows. Carrying out the energy integration of the dynamical structure factor $S(\vec{q},\omega)$ of Eq. (2), we obtain

$$\int S(\vec{q},\omega)\, d\omega = \sum_f \left|\langle f|e^{i\vec{q}\cdot\vec{r}}|i\rangle\right|^2 = \sum_f \langle i|e^{i\vec{q}\cdot\vec{r}}|f\rangle\langle f|e^{i\vec{q}\cdot\vec{r}}|i\rangle = \langle i|e^{i\vec{q}\cdot\vec{r}}e^{i\vec{q}\cdot\vec{r}}|i\rangle, \qquad (5)$$

where we have made use of the closure theorem $\sum_f |f\rangle\langle f| = 1$, assuming that the final states

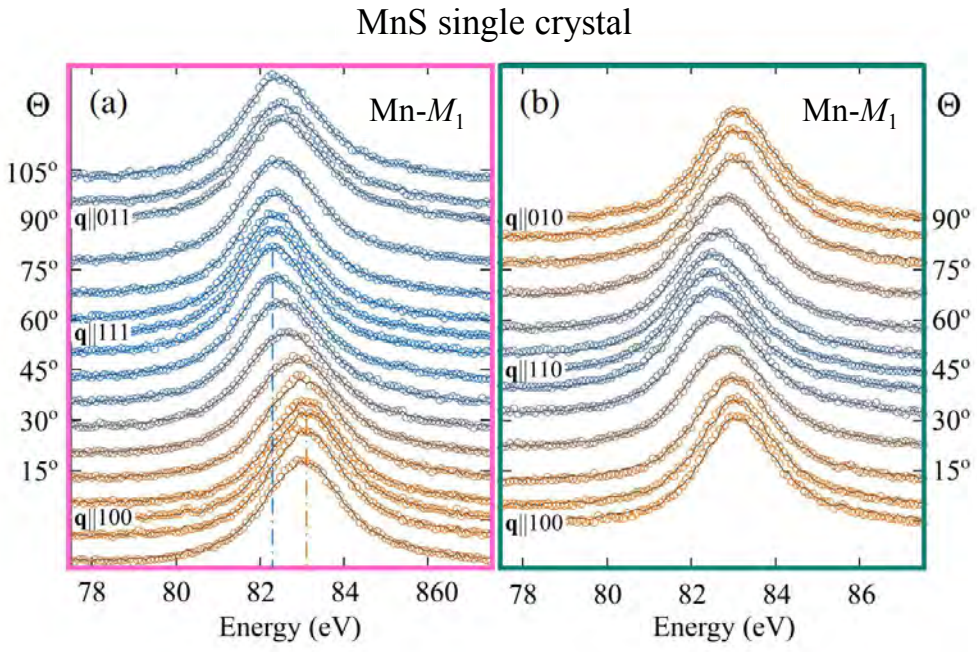


Fig. 5: *Stack of the Mn M_1 NIXS spectra acquired for $\vec{q}$ ∥ [100] – [111] – [011] (panel a) and [100] – [110] – [010] (panel b). The grey lines are the result of the fit procedure explained in the text. Reproduced from [3].*

$|f\rangle$ are complete. As a result, only the properties of the initial state are probed. Here we note that in this experiment we are considering only final states that can be reached in the $(s \rightarrow d)$ transition channel, so that what we have obtained is the d hole density in the initial state. This d hole density is a *locally* projected quantity since the final states carry the presence of the s core hole. The procedure is analogous to using linear and circular dichroism sum rules in x-ray absorption spectroscopies to extract expectation values of relevant quantum numbers of the system in the ground state [18, 20–22].

We would like to remark that the intensity distribution in s-NIXS is not what would be measured in an x-ray diffraction (XRD) experiment, even if such an experiment could be carried out with sufficient accuracy. In fact, s-NIXS provides information complementary to that from an XRD experiment by elucidating which local orbital or atomic wave function is active. The s-NIXS method presented here is not limited to ionic materials. In cases where configuration interaction effects play an important role due to covalency or itineracy, the image of the probed local orbital will reflect these effects directly as we will show below.

7 From ground state to excited states

Having established that s-NIXS is a powerful method for the investigation of the ground state, we will now explore the spectroscopy aspect of s-NIXS in order to study the excited states which are most often dominated by many-body atomic multiplet interactions in the d- and f-electron materials. In particular we aim to determine the orbital character of those states. The idea is that the use of an s core hole should simplify the analysis of the spectra in two significant

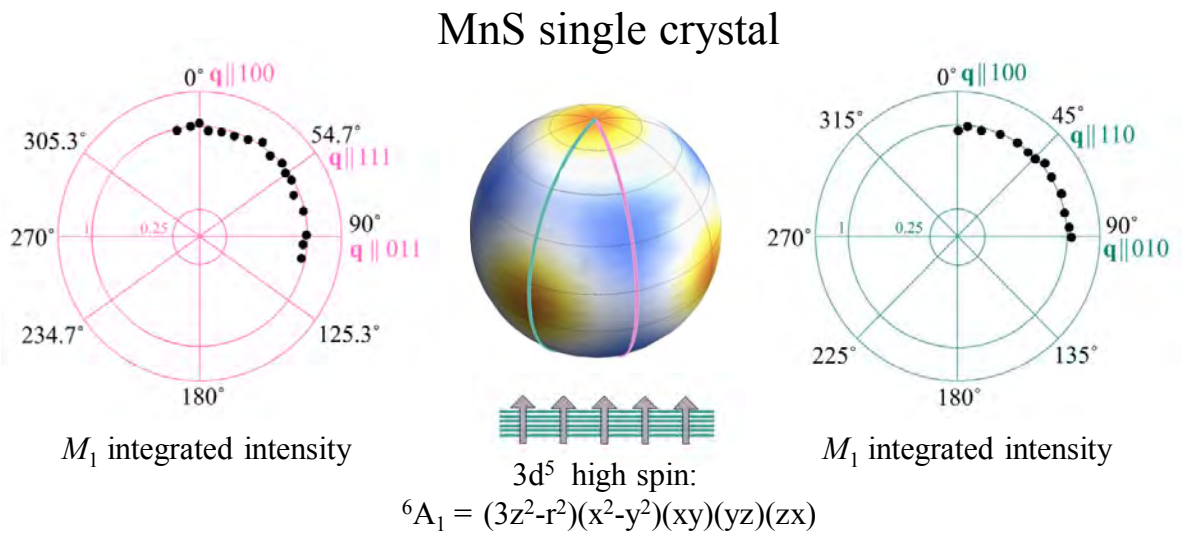


Fig. 6: *Left and right panels: polar plots of the integrated intensity of the Mn M_1 in the [100] – [111] – [011] plane (magenta) and in the [100] – [110] – [010] plane (dark cyan). Central panel: theoretical three-dimensional orbital hole density of the Mn high-spin $3d^5$ configuration in octahedral coordination: the spherical shaped ${}^6A_1$ $3d(x^2-y^2)(3z^2-r^2)(xy)(yz)(yz)$. Adapted from [3].*

aspects. First, the presence of the s hole does not add an extra orbital angular momentum so that use can be made of the well-established and readily available Sugano-Tanabe-Kamimura diagrams that depict the multiplet energy scheme of $3d$ ions for varying values of the crystal field [23]. Second, the identification of the orbital character of the multiplet state can be done in a direct manner by the imaging ability of the s-NIXS method. To exemplify these points, we have carried out the experiment on α-MnS, a rock salt type antiferromagnetic insulator with far from complete filling of the d shell so that orbital degrees of freedom are present in its excited states.

The α-MnS single crystal sample was grown by chemical vapor transport and during the measurements the sample was kept at a temperature of 50 K. The Mn M_1 edge spectra acquired at different angles are shown in panel (a) of Fig. 5 for the set $\vec{q} \parallel$ [100] – [111] – [011] and in panel (b) for $\vec{q} \parallel$ [100] – [110] – [010]. Also here the Compton contribution was used for normalization and has been subtracted from the spectra using a linear background. At first glance, there are no prominent variations in the peak intensities, but, as highlighted by the colors, the overall peak energy position varies as a function of angle, i.e., as a function the orientation of $\vec{q}$ with respect to the crystallographic axes.

Plotting first the integrated intensities as a function of angle in a polar plot, we indeed can observe from the left and right panels of Fig. 6 that there is essentially no orientational dependence, i.e., the hole charge density is constant in all directions. The spherical shape of the Mn $3d$ charge density as shown in the middle panel of Fig. 6 is fully consistent with the scenario in which all five spin-up or all five spin-down $3d$ orbitals are unoccupied, i.e., in which the Mn^{2+} $3d^5$ ion is in its Hund's rule high-spin ${}^6A_1$ ground state.

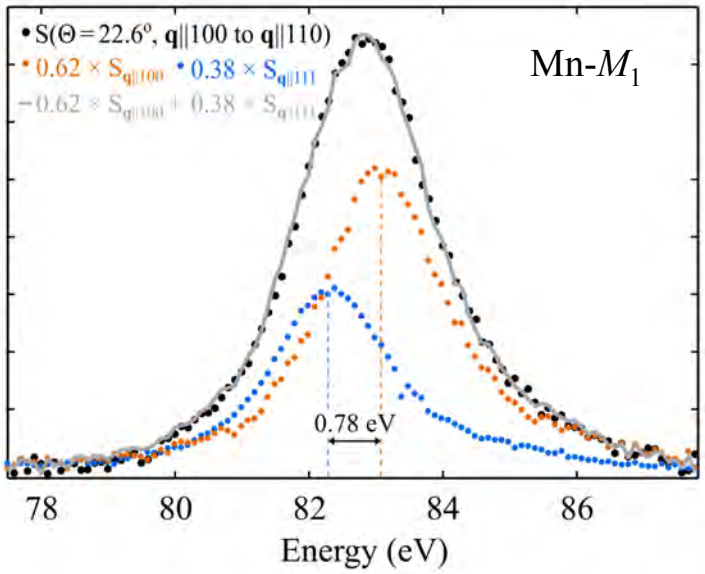


Fig. 7: *The experimental spectrum for a given angle Θ (black dots) is decomposed using the weighted sum of $\vec{q} \parallel$ [111] (blue dots) and $\vec{q} \parallel$ [100] (orange dots) spectra that provides the best fit (grey line). Reproduced from [3].*

8 Analysis of excited states

While the integrated intensity of the spectra is constant with angle, the energy position of the M_1 peak as displayed in panels (a) and (b) of Fig. 5 does vary with the orientation, between 82.37 eV (blue vertical line) for the $\vec{q} \parallel$ [111] spectrum and 83.15 eV (orange vertical line) for $\vec{q} \parallel$ [100]. Since no discernible dispersion can be expected for core-hole excitations, such a variation in energy indicates that the M_1 signal consists of features positioned at different energies whose relative intensities change with angle. With the Mn ion coordinated octahedrally by six S ions, we expect that the energy differences in the final states must be related to $10Dq$, the octahedral crystal field splitting between the t_{2g} and e_g orbitals.

We now assume that the two spectra acquired at $\vec{q} \parallel$ [111] and $\vec{q} \parallel$ [100], which show the extreme peak positions and narrowest line shapes, are the basic components making up the M_1 NIXS signal for every other orientation. We performed fits to all spectra using a linear combination of these two experimental spectra and determined their relative weights, so that each spectrum is described as $S(\Theta) = c_{111}(\Theta)S_{\mathbf{q}\parallel 111} + c_{100}(\Theta)S_{\mathbf{q}\parallel 100}$ where c_{111} and c_{100} are the free fitting parameters, as depicted in Fig. 7. The resulting fits are shown with gray lines in Fig. 5.

The weights $c_{111}(\Theta)$ and $c_{100}(\Theta)$ obtained by the fits are plotted in left panels Fig. 8 for the two sample orientations. We can observe that the angular dependence follows the shape of the t_{2g} and e_g states with great accuracy, allowing us to directly identify the orbitals reached in each excitation. In particular, the t_{2g} orbital shape is drawn by the angular dependence of $c_{111}(\Theta)$, the weight of the $S_{\vec{q}\parallel 111}$ component peaking at 82.37 eV. Likewise, the excitation into e_g orbitals is represented by $c_{100}(\Theta)$, the weight of the $S_{\vec{q}\parallel 100}$ component peaking at 83.15 eV. The difference between these two energies is due to the e_g-t_{2g} splitting, and it is therefore a direct measurement of the crystal field parameter $10Dq = 83.15\,\text{eV} - 82.37\,\text{eV} = 0.78\,\text{eV}$ as also illustrated in the right panel of Fig. 8.

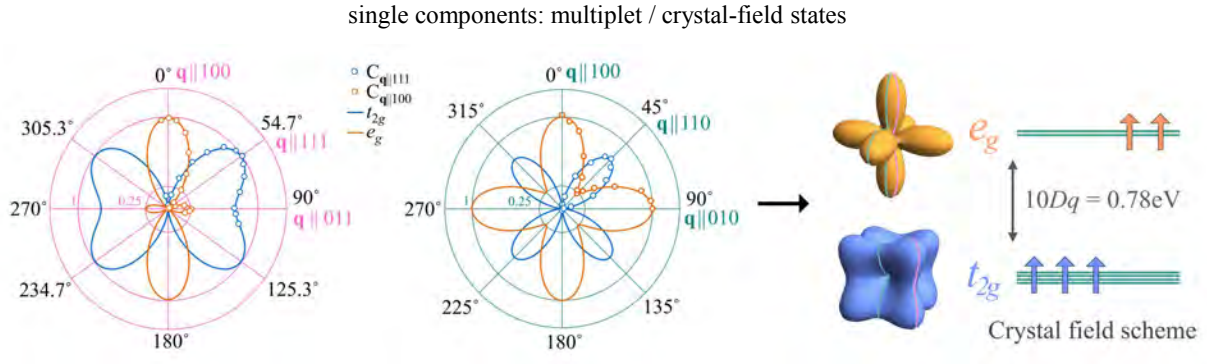


Fig. 8: *Left panels: the intensity of the $\vec{q} \parallel$ [100] (orange) and $\vec{q} \parallel$ [111] (blue) components follow the angular dependence of the e_g (orange) and t_{2g} (blue) orbitals, respectively. Right panels: The e_g-t_{2g} energy splitting is about $10Dq$ = 0.78 eV. Adapted from [3].*

9 Many-body energy-level diagram

A proper interpretation of the spectra must include the effects of Coulomb and exchange interactions between the electrons within the d shell, meaning that we need to put our results in a many body framework which takes into account both the full atomic multiplet theory and the local effect of the lattice. In particular, starting from the $3s^2 3d^5$ configuration for the ground state of the Mn ion in α-MnS, the s-NIXS process leads to a $3s^1 3d^6$ final state. Since the s core hole does not add an extra orbital angular momentum component, the multiplet structure of the final state will be closely related to the one of the $3d^6$ configuration. Therefore, one could make use of the well-established Sugano-Tanabe-Kamimura diagrams [23] for a quantitative analysis of s-NIXS spectra, after taking into account for some modifications due to the presence of the extra $3s$ spin. We exemplify this in Figs. 9 (a) and (b), where we reproduce the total energy diagrams for a Mn ion in $3d^6$ and $3s^1 3d^6$ configuration calculated with the *Quanty* code [24]. The corresponding Slater integrals F^2_{3d-3d} and F^4_{3d-3d} (and G^2_{3s-3d} for the $3s^1 3d^6$ configuration) have been obtained using the code by R.D. Cowan [25] and reduced to 80% of their Hartree-Fock value to account for configuration interaction effects not included in the calculation [16, 17].

To ease the comparison between the diagrams of the two configurations, the lowest energy state is not fixed at 0 eV for every $10Dq$ like it was done in the Sugano-Tanabe-Kamimura book [23]. Instead, the lowest state of the $3d^6$ configuration is put to zero only for $10Dq = 0$ eV, and serves as reference energy. In this way, the multiplets are split by the action of $10Dq$, but the average energy of the diagram is kept constant. The energy levels are labeled for zero crystal field ($10Dq = 0$ eV, spherical symmetry) with the ${}^{(2S_d+1)}L_d$ term symbols, where S_d and L_d represent the quantum numbers of the $3d$ shell and the orbital quantum numbers L_d are indicated with the usual letter notation (S, P, D, F, G...). The subscript in parenthesis indicates the degeneracy of each term symbol, also including the degeneracy given by the $3s$ hole. In total, the $3d^6$ configuration has $\binom{10}{6} = 210$ states, while $3s^1 3d^6$ has $2 \times \binom{10}{6} = 420$ states, due to the extra multiplicity of the $3s^1$ spin. The colors group the states in the diagrams according to their $3d$ spin multiplicity ($2S_d$+1): singlets are purple, triplets green, and quintets orange.

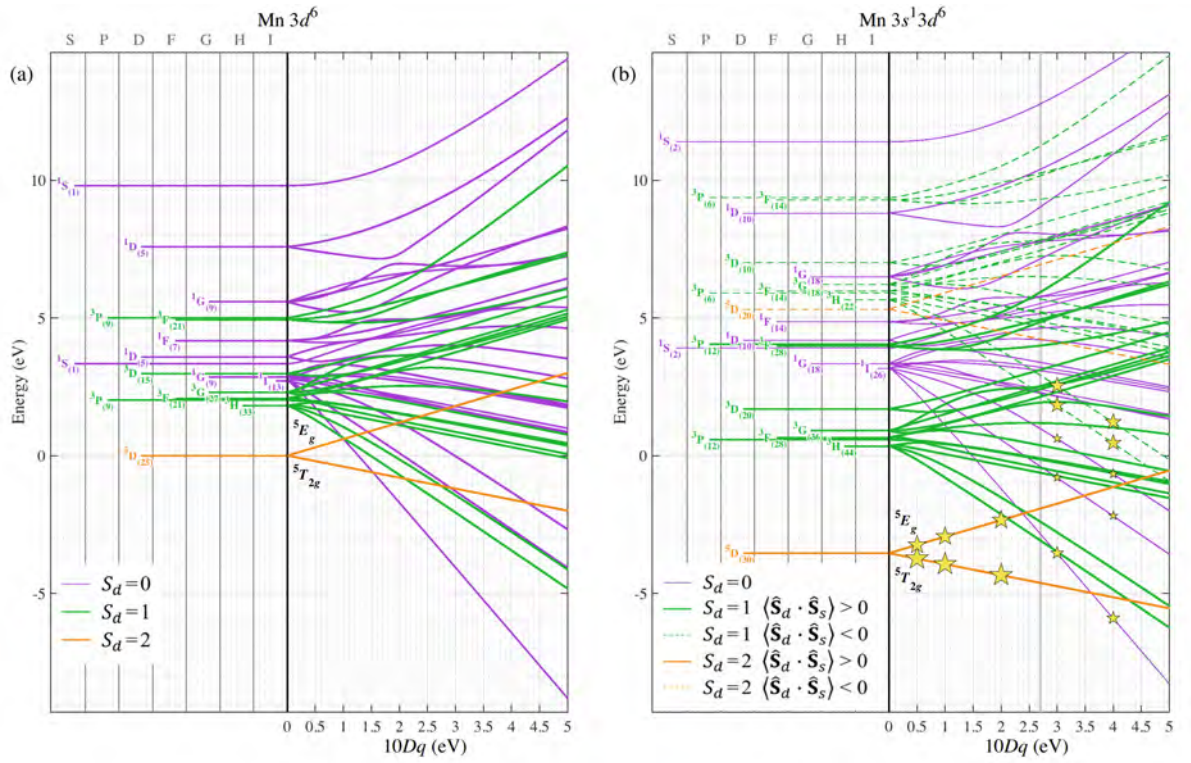


Fig. 9: *Modified Sugano-Tanabe-Kamimura diagrams showing the dependence on* $10Dq$ *of the energy levels of (a) Mn* $3d^6$ *(in absence of the core* $3s$ *hole) and of (b) Mn* $3s^1 3d^6$ *(the s-NIXS final state). Each multiplet is labeled with the corresponding term symbol at* $10Dq = 0$ *eV and in parenthesis its degeneracy. The colors indicate each state's spin multiplicity: purple for singlets, green for triplets and orange for quintets. In panel (b), the thick full lines represent the states having parallel* $\boldsymbol{S}_d$ *and* $\boldsymbol{S}_s$*, the thin dashed lines the states with* $\boldsymbol{S}_d$ *anti-parallel to* $\boldsymbol{S}_s$*, and the thin full lines the states not affected by* $\boldsymbol{S}_s$*. The stars indicate, for several* $10Dq$ *values, the possible final states of the s-NIXS process, and the vertical gray line at* $10Dq = 2.7$ *eV is the transition between the high and low spin configurations of the* $3d^5$ *ground state. Reproduced from [3].*

Starting from the singlets, it is straightforward to notice that the energy scheme within the set of singlets, as well as the dependence of each state on $10Dq$, is essentially the same between the two configurations $3d^6$ and $3s^1 3d^6$. This agrees with the fact that singlet states cannot be modified by an interaction with a $3s$ spin, given by $\langle \hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s \rangle$, which is zero for $S_d = 0$. The minor changes between the two groups ($\lesssim 10\%$ of the energy splittings) is due to the different values of the Slater integrals F^2_{3d-3d} and F^4_{3d-3d} between the two configurations. The sets of triplets (green lines) and quintets (orange) of the $3s^1 3d^6$ configuration, instead, are further divided in two subgroups depending on the relative alignment of $\boldsymbol{S}_d$ and $\boldsymbol{S}_s$. The states having the two spins parallel are represented with thick lines, while the thin dashed lines represent the states with opposite $\boldsymbol{S}_d$ and $\boldsymbol{S}_s$. These subgroups are each a replica of the corresponding $3d^6$ triplet or quintet set, rigidly shifted in energy due to the spin interaction.

To understand and predict these energy shifts due to the Coulomb interaction between the d electrons and the electron of the open $3s$ shell one can relate the Coulomb operator to the spin

		$S_{\text{tot}} = S_d + S_s$		$S_{\text{tot}} = S_d - S_s$	
S_d	$2S_d + 1$	$\langle \hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s \rangle$	$\Delta E\ [G^2_{3s-3d}]$	$\langle \hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s \rangle$	$\Delta E\ [G^2_{3s-3d}]$
0	1	0	0	0	0
$\frac{1}{2}$	2	$\frac{1}{4}$	$-\frac{1}{10}$	$-\frac{3}{4}$	$\frac{3}{10}$
1	3	$\frac{2}{4}$	$-\frac{2}{10}$	$-\frac{4}{4}$	$\frac{4}{10}$
$\frac{3}{2}$	4	$\frac{3}{4}$	$-\frac{3}{10}$	$-\frac{5}{4}$	$\frac{5}{10}$
2	5	$\frac{4}{4}$	$-\frac{4}{10}$	$-\frac{6}{4}$	$\frac{6}{10}$
$\frac{5}{2}$	6	$\frac{5}{4}$	$-\frac{5}{10}$	$-\frac{7}{4}$	$\frac{7}{10}$

Table 1: *Eigenvalues of $\hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s$, related to the energy splitting due to the Coulomb interaction of pure spin multiplets (i.e., neglecting spin-orbit interaction) in the d shell with the open $3s^1$ core shell for states with the d spin either parallel ($S_{tot} = S_d + S_s$) or anti-parallel ($S_{tot} = S_d - S_s$) to the core s spin. Reproduced from [3].*

operators of the d shell and s shell. We have

$$H^{\text{Coulomb}}_{3s-3d} = -\frac{2}{5}\hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_s\, G^2_{3s-3d} + n_d\big(F^0_{3s-3d} - \frac{1}{10}G^2_{3s-3d}\big). \tag{6}$$

The last term in the equation is constant for all multiplets within a $3s^1\,3d^n$ configuration and as such does not lead to a splitting between the different states. The eigenvalues of $\hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s$ can be obtained simply by inverting the formula [26] $(\hat{\boldsymbol{S}}_{\text{tot}})^2 = (\hat{\boldsymbol{S}}_d)^2 + (\hat{\boldsymbol{S}}_s)^2 + 2\hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s$, where S_{tot} can be $S_d + S_s = S_d + \frac{1}{2}$ or $|S_d - S_s| = |S_d - \frac{1}{2}|$, and remembering that the eigenvalue of $(\hat{\boldsymbol{S}})^2$ is $S(S+1)$. The resulting eigenvalues are listed in Table 1 for each possible value of S_d.
Typical values of G^2_{3s-3d} for the $3d$ series, after a reduction to 80 % of their Hartree-Fock values, range from 8.5 to 10.2 eV. In general, ΔE is larger for larger spin multiplicities. With these ingredients, one can easily build the Sugano-Tanabe-Kamimura diagrams of the NIXS final configuration from the ones without the $3s$ core hole, listed, for example, in Figs. 5.1–5.7 of the famous book of S. Sugano, Y. Tanabe and H. Kamimura [23]. This allows the diagram of the possible s-NIXS final states to be reproduced without the need for performing new calculations and, by comparing the diagrams to the spectra, quantitatively determine the value $10Dq$.

10 Spectral lineshape

The next step towards a complete understanding of the s-NIXS spectra is to realize that not all states depicted in Fig. 9 (b) can be reached starting from the ground state of Mn^{2+} in α-MnS. In our case, the addition of one extra $3d$ electron ($s = 1/2$) to the high-spin $^6A_1$ ($S_d = 5/2$) ground state can only lead to quintet final states ($S_d = 2$), with the $3s^1$ spin parallel to the majority spin of the $3d$. Therefore, $-\frac{2}{5}G^2_{3s-3d}\,\langle \hat{\boldsymbol{S}}_d \cdot \hat{\boldsymbol{S}}_s \rangle < 0$, and the low energy replica of the quintet set is reached. The possible s-NIXS final states for different values of $10Dq$ are

indicated with stars in Fig. 9 (b), where the values of $10Dq$ corresponding to the initial $3d^5$ high-spin configuration ${}^6A_1$ are to the left of the gray vertical line, and low-spin on the right. The size of each star is proportional to the intensity of the corresponding peak on the s-NIXS spectra, averaged over the two sets of directions presented here (i.e., $\mathbf{q} \parallel$ [100] – [111] – [011] and $\mathbf{q} \parallel$ [100] – [110] – [010]). There are only two states in the diagrams that can be reached with a s-NIXS excitation, namely the ${}^5T_2$ (the extra $3d$ electron occupying t_{2g} orbitals) and ${}^5E$ (the extra electron in the e_g). From the shape of the final state orbitals as imaged in Fig. 8, we can identify immediately that the lower energy peak belongs to the ${}^5T_2$ state and the higher to the ${}^5E$. It is then straightforward to understand, as predicted above, that in our case the experimental peak energy separation of 0.78 eV corresponds one-to-one to the $10Dq$ value.
It is worth looking into Fig. 9 (b) in more detail. For $10Dq$ values on the right of the gray vertical line, the ground state will no longer be the high-spin but the low-spin $3d^5$. The consequence for the s-NIXS spectrum is dramatic. It switches from a two-peak structure (two stars) into a five-peak features (three small stars, two large stars). This demonstrates that the line shape of the s-NIXS spectrum is an extremely sensitive indicator of the ground state symmetry. The value of $10Dq$ can be determined directly from the spread of the five peaks. Consequently, the ground state hole density will also change in going from high to low-spin, i.e., from spherical ($t_{2g}^3 e_g^2$-like) to highly non-spherical (t_{2g}^5-like), which can be revealed directly by the image obtained from the directional dependence of the integrated s-NIXS intensity.

11 Covalency

We now investigate the influence of covalency on the s-NIXS image of the local d hole density and the spectra. To this end, we have carried out configuration-interaction calculations [16, 17, 24] using an octahedral MnS_6 cluster which includes explicitly the hybridization between the Mn $3d$ and the S $3p$ orbitals. We have set the hopping integrals for the e_g orbital at 1.92 eV and for the t_{2g} at 1.15 eV [24] and varied the energy difference between the d^5 and $d^6\underline{L}$ configurations (charge transfer energy Δ). Here $\underline{L}$ denotes the S $3p$ ligand hole states. The results are shown in Fig. 10.
Starting with the ionic calculation, we have in the ground state an electron occupation of 3.000 for the t_{2g} orbital and 2.000 for the e_g (values listed in the central panel). The corresponding ground state hole density is spherical (Fig. 10, left panel, red line), and the directional dependence of the low and high energy peaks in the calculated s-NIXS spectra follow (Fig. 10, two right panels) the t_{2g} (blue) and e_g (orange) orbital shapes, respectively, as we have seen already in Fig. 8. Switching on the hybridization between the d^5 and $d^6\underline{L}$ configurations, we can see that the electron occupation in the ground state increases with lowering the Δ values. It increases faster for the e_g than for the t_{2g} (values listed in the central panel), consistent with the fact that the hopping integral with the ligand is larger for the e_g than for the t_{2g}. In the strongly covalent case of $\Delta = 1$ eV we have 3.136 in the t_{2g} and 2.287 for the e_g. Accordingly, the ground state hole density, proportional to the NIXS signal, decreases for lower Δ values and becomes strongly non-spherical (left panel). This also means that the amount of hybridization can be

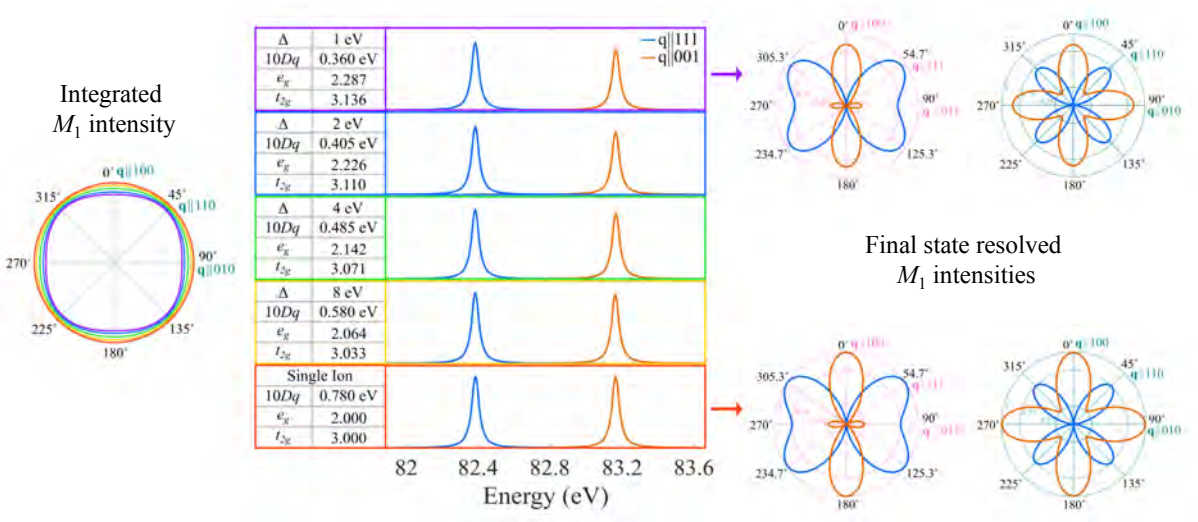


Fig. 10: *Mn M_1 simulations using a MnS_6 cluster calculated for various degrees of covalency. Left panel: Polar plot of the integrated intensity of the Mn M_1 in the [100] – [110] – [010] plane: from ionic (most outer circle) to strongly covalent (most inner distorted circle, Δ = 1 eV) case. Middle panel: Mn M_1 spectra along the $\vec{q}$ ∥ [100] (orange) and $\vec{q}$ ∥ [111] (blue) directions: from ionic (bottom curves) to strongly covalent (top curve, Δ = 1 eV). The Δ and corresponding ionic $10Dq$ values are indicated, together with the resulting e_g and t_{2g} occupation numbers. The energy splitting between the $\vec{q}$ ∥ [100] (orange) and $\vec{q}$ ∥ [111] (blue) peaks is fixed at 0.78 eV. Right panel: polar plot of the final state resolved Mn M_1 intensities (high energy peak in orange and low energy peak in blue), for the ionic (bottom) and strongly covalent (top, Δ = 1 eV) cases. Adapted from [3].*

extracted from the precise shape of the hole density as measured by s-NIXS.

Perhaps a surprising result is that the presence of hybridization does not have much influence on the s-NIXS spectrum, even for the $\Delta = 1\,\text{eV}$ case. It shows the same two peak structure, and the directional dependence of the low and high energy peaks still follows the t_{2g} (blue) and e_g (orange) orbital shapes, respectively. It may seem surprising that hybridization or covalency does little to the spectrum, but we can draw a parallel to x-ray absorption spectroscopy (XAS), which is also a core-level spectroscopy in which a core electron is excited into the valence shell. It is known that the $M_{4,5}$ edges of Ce and the $L_{2,3}$ of the $3d$ transition metal ions can be well reproduced using ionic calculations despite the fact that there is covalency. The reason is that the energy orderings of the electron configurations are identical in the initial state and in the XAS final state. Thus the spectral weights of the other local configurations are strongly suppressed due to quantum mechanical interference effects [16, 27]. Therefore, both XAS and NIXS, generally produce a spectrum that is very similar to the one that belongs to the main local configuration, e.g., the ionic configuration. Despite the fact that the overall s-NIXS spectrum is rather insensitive to hybridization, a closer look at the line shape and the intensities does reveal details that contain information about the hybridization strength. The middle panel of Fig. 10 show that the intensity of the high energy peak (e_g, orange) becomes smaller relative to that of the low energy peak (t_{2g}, blue) with decreasing Δ values. The s-NIXS spectrum can there-

fore be used to help determine quantitatively the parameter values describing the hybridization process.

In the ionic calculations, we have used a $10Dq$ value of 0.78 eV in order to get a separation of 0.78 eV between the two peaks in the Mn M_1 NIXS spectrum. We will name this the ionic-$10Dq$ parameter. Upon switching-on the hybridization, we must decrease the value of the ionic-$10Dq$ parameter in our calculations to maintain the 0.78 eV separation between the two peaks in the NIXS spectrum. A larger decrease is required when Δ gets smaller, i.e., when the hybridization gets stronger (see the values in the cental panel). This can be understood if one considers the fact that the hopping integral with the ligand is larger for e_g than for t_{2g}, and that the resulting difference in hybridization energy contributes to the energy splitting between the e_g bonding state and the t_{2g} equivalent. It is the combined effect of hybridization and the ionic-$10Dq$ that produces the 0.78 eV splitting in the NIXS spectrum, which we can define as the effective-$10Dq$. NIXS, like XAS, can thus provide direct access to the effective crystal field energy [16, 17, 28, 29]. The horizontal axis of the Sugano-Tanabe-Kamimura diagram presented in Fig. 5 can therefore be understood as the energy scale for the effective-$10Dq$ in covalent materials.

12 Metallic systems: ground state properties

The s-NIXS data presented so far concern insulating materials. We now will investigate the capabilities of this method for highly metallic systems. To this end, we have carried out measurements on elemental metals such as Cr, Fe, and Ni. For each metal we have used two single crystals, one having a surface with the (100) normal and the other the (111), in order to cover a sufficiently wide range of angles. All surfaces were epi-polished to mirror quality.

The results are shown in Fig. 11. The top panel displays the polar plot of the integrated M_1 intensities from scans collected from the (100) surface, and the middle panel from the (111) surface. The bottom panel summarizes the data in a 3-dimensional (3D) representation. Making use of the sum-rule argument explained above (since the quantities of concern are *integrated* intensities), we can assert that the images show the shape of the local $3d$ hole density of the metals in the ground state. We can clearly observe that there are significant differences between Cr, Fe, and Ni. These data can then be compared with *ab-initio* theoretical calculations in order to unveil to what extent electron correlation effects beyond mean field affect the ground state $3d$ charge density.

We would like to remark that information about the $3d$ charge density in highly metallic systems is in general very difficult to extract from spectroscopic data. In the presence of correlations, a quantitative analysis of the spectra is severely hampered by the large number of configurations that need to be included associated with the strong inter-site charge fluctuations. The use of *integrated* intensities in s-NIXS circumvents all these difficulties.

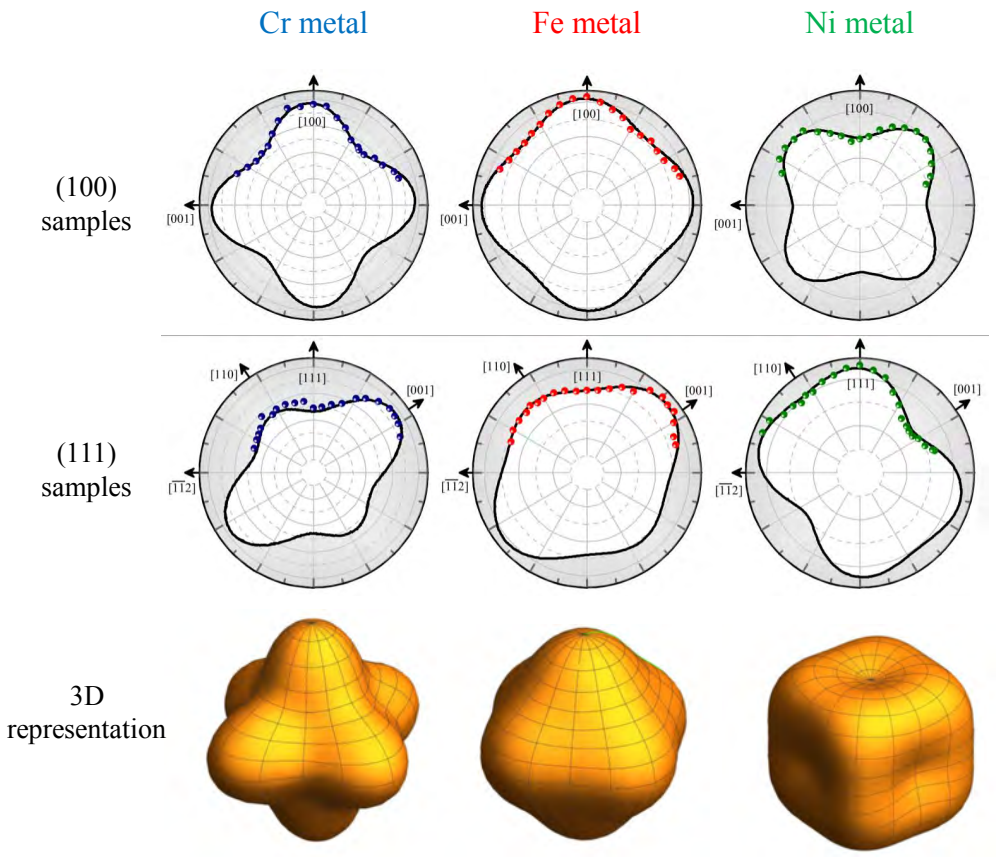


Fig. 11: *Shape of the hole density in the* 3d *shell of elemental Cr, Fe, and Ni metal. Top and middle panels: polar plots of the integrated intensity of the* M_1 *edge from the single crystal samples with the (100) and (111) surfaces, respectively. Bottom panel: a 3-dimensional representation of the measurement results.*

13 Concluding remarks

In conclusion, we have shown that s-NIXS has the unique ability to directly provide a spatial image of the local orbitals that are active in the formation of the ground state as well as of the excited states. This purely experimental method is especially valuable for tackling problems in complex systems where band formation and electron correlation effects are both important. s-NIXS can even provide direct information about the ground state properties of highly metallic systems, where spectroscopic techniques hit their limitations. The prospects of s-NIXS as a new experimental method need to be further explored. It is has, for example, the advantage that it is element-specific, and also allows for sophisticated sample environments, e.g., involving small samples, high pressures, and high/low temperatures.

Acknowledgements

The work presented in this chapter is the result of a fruitful and inspiring collaboration with M.W. Haverkort, M. Sundermann, H. Yavaş, H. Gretarsson, A. Amorese, B. Leedahl, K. Chen, M. Schmidt, H. Borrmann, Yu. Grin, and A. Severing.

References

[1] H. Yavaş, M. Sundermann, K. Chen, A. Amorese, A. Severing, H. Gretarsson, M.W. Haverkort, and L.H. Tjeng, Nat. Phys. **15**, 559 (2019)

[2] B. Leedahl, M. Sundermann, A. Amorese, A. Severing, H. Gretarsson, L. Zhang, A.C. Komarek, A. Maignan, M.W. Haverkort, and L.H. Tjeng, Nat. Commun. **10**, 5447 (2019)

[3] A. Amorese, B. Leedahl, M. Sundermann, H. Gretarsson, Z. Hu, H.-J. Lin, C.T. Chen, M. Schmidt, H. Borrmann, Yu. Grin, A. Severing, M.W. Haverkort, and L.H. Tjeng, Phys. Rev. X **11**, 011002 (2021)

[4] W. Schülke: *Electron Dynamics by Inelastic X-Ray Scattering* (Oxford University Press, 2007)

[5] M.W. Haverkort, A. Tanaka, L.H. Tjeng, and G.A. Sawatzky, Phys. Rev. Lett. **99**, 257401 (2007)

[6] R.A. Gordon, G.T. Seidler, T.T. Fister, M.W. Haverkort, G.A. Sawatzky, A. Tanaka, and T.K. Sham, EPL (Europhys. Lett.) **81**, 26004 (2008)

[7] R. Caciuffo, G. van der Laan, L. Simonelli, T. Vitova, C. Mazzoli, M.A. Denecke, and G.H. Lander, Phys. Rev. B **81**, 195104 (2010)

[8] M. Sundermann and M.W. Haverkort (unpublished)

[9] R.A. Gordon, M.W. Haverkort, S.S. Gupta, G.A. Sawatzky, J. Phys. Conf. Ser. **190**, 012047 (2009)

[10] J.A. Bradley, S.S. Gupta, G.T. Seidler, K.T. Moore, M.W. Haverkort, G.A. Sawatzky, S.D. Conradson, D.L. Clark, S.A. Kozimor, and K.S. Boland, Phys. Rev. B **81**, 193104 (2010)

[11] J.A. Bradley, K.T. Moore, G. van der Laan, J.P. Bradley, and R.A. Gordon, Phys. Rev. B **84**, 205105 (2011)

[12] T. Willers, F. Strigari, N. Hiraoka, Y.Q. Cai, M.W. Haverkort, K.-D. Tsuei, Y.F. Liao, S. Seiro, C. Geibel, F. Steglich, L.H. Tjeng, and A. Severing, Phys. Rev. Lett. **109**, 046401 (2012)

[13] M. Sundermann, M.W. Haverkort, S. Agrestini, A. Al-Zein, M. Moretti Sala, Y. Huang, M. Golden, A. de Visser, P. Thalmeier, L.H. Tjeng, and A. Severing, Proc. Nat. Acad. Science. U.S.A. **113**, 13989 (2016)

[14] M. Sundermann, H. Yavaş, K. Chen, D.J. Kim, Z. Fisk, D. Kasinathan, M.W. Haverkort, P. Thalmeier, A. Severing, and L.H. Tjeng, Phys. Rev. Lett. **120**, 016402 (2018)

[15] C.T. Chen, L.H. Tjeng, J. Kwo, H.L. Kao, P. Rudolf, F. Sette, and R.M. Fleming, Phys. Rev. Lett. **68**, 2543 (1992)

[16] F.M.F. de Groot, J. Electron Spectrosc. Relat. Phenom. **67**, 529 (1994)

[17] A. Tanaka and T. Jo, J. Phys. Soc. Jpn. textbf63, 2788 (1994)

[18] S.I. Csiszar, M.W. Haverkort, Z. Hu, A. Tanaka, H.H. Hsieh, H.-J. Lin, C.T. Chen, T. Hibma, and L.H. Tjeng, Phys. Rev. Lett. **95**, 187205 (2005)

[19] P. Hansmann, A. Severing, Z. Hu, M.W. Haverkort, C.F. Chang, S. Klein, A. Tanaka, H.H. Hsieh, H.-J. Lin, C.T. Chen, B. Fak, P. Lejay, and L.H. Tjeng, Phys. Rev. Lett. **100**, 066405 (2008)

[20] B.T. Thole, P. Carra, F. Sette, and G. van der Laan, Phys. Rev. Lett. **68**, 1943 (1992)

[21] P. Carra, B.T. Thole, M. Altarelli, and X. Wang, Phys. Rev. Lett. **70**, 694 (1993)

[22] C.T. Chen, Y.U. Idzerda, H.-J. Lin, N.V. Smith, G. Meigs, E. Chaban, G.H. Ho, E. Pellegrin, and F. Sette, Phys. Rev. Lett. **75**, 152 (1995)

[23] S. Sugano, Y. Tanabe, and H. Kamimura: *Multiplets of Transition-Metal Ions in Crystals* (Academic Press, New York, 1970)

[24] M.W. Haverkort, M. Zwierzycki, and O.K. Andersen, Phys. Rev. B **85**, 165113 (2012)

[25] R. Cowan: *The theory of atomic structure and spectra* (University of California, Berkeley, 1981)

[26] S. Blundell, *Magnetism in condensed matter* (Oxford University Press, 2001)

[27] O. Gunnarsson and K. Schönhammer, Phys. Rev. B **28**, 4315 (1983)

[28] S. Agrestini, Z. Hu, C.-Y. Kuo, M.W. Haverkort, K.-T. Ko, N. Hollmann, Q. Liu, E. Pellegrin, M. Valvidares, J. Herrero-Martin, P. Gargiani, P. Gegenwart, M. Schneider, S. Esser, A. Tanaka, A.C. Komarek, and L.H. Tjeng, Phys. Rev. B **91**, 075127 (2015)

[29] R.P. Wang, B. Liu, R.J. Green, M.U. Delgado-Jaime, M. Ghiasi, T. Schmitt, M.M. van Schooneveld, and F.M.F. de Groot, J. Phys. Chem. C **121**, 24919 (2017)

8 Probing Spin, Charge and Orbital Degrees of Freedom by X-Ray Spectroscopy

Eva Benckiser
Max-Planck-Institut für Festkörperforschung
Heisenbergstrasse 1, 70569 Stuttgart

Contents

1 Introduction **2**

2 Spin, charge, orbital, and lattice degrees of freedom in epitaxial multilayers **2**

3 Soft X-ray spectroscopy **5**
3.1 X-ray absorption and dichroism 7
3.2 Resonant X-ray scattering 12
3.3 X-ray resonant reflectometry 13

4 Case studies **16**
4.1 Interfacial doping in La_2CuO_4-$LaNiO_3$ hybrid structures 17
4.2 Orbital reflectometry of nickelate and vanadate superlattices 19
4.3 Noncollinear magnetic order in nickel oxide heterostructures 24

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

The properties of quantum materials, such as $3d$ transition-metal oxides, are largely governed by the collective behavior of their strongly interacting electrons and the material's response is subject to a complex interplay of the local spin, charge, and orbital quantum degrees of freedoms [1]. This manifests itself in various properties such as Mott metal-to-insulator transitions, different types of magnetic order (ferro-, ferri-, and anti-ferromagnetic), orbital order, and unconventional superconductivity [2, 3]. These diverse properties have been extensively studied in bulk materials over the past decades and are of great importance for both, fundamental and applied solid state research.

In recent years, technological advances have enabled the synthesis of heterostructures of quantum material oxides with ultra-thin layers and atomic layer precision, thereby providing different opportunities to manipulate correlated electron systems [4, 5]. Research is primarily aimed at rational materials design through the targeted realization of interfacial reconstructions. The investigation of model systems with specific reconstructions, the generation and understanding of new materials properties, in particular the stabilization of technologically interesting phases under ambient conditions are central motivations of the research area [6–8].

In heterostructures, electronic and magnetic phases that are inaccessible in the bulk can be stabilized and controlled by biaxial strain induced by epitaxy with a single-crystalline substrate, electronic confinement, interfacial doping, or magnetic interactions [9]. The success of these efforts depends on a detailed understanding of the interfacial interactions and reconstructions in these artificially layered materials, which often span only a few atomic layers. The presence of only a very small volume of the material of interest poses a problem for many solid-state spectroscopy methods, such as neutron scattering, muon spin relaxation, or optical spectroscopy, which are normally used to study bulk materials.

In this lecture I like to introduce soft X-ray spectroscopy, with special focus on X-ray absorption and resonant elastic X-ray scattering and reflectivity, as a non-destructive spectroscopic method, which provides important insights into the physics and chemistry of transition-metal oxide heterostructures. As these methods require soft X-rays with variable energy and polarization, it is a synchrotron technique. The rather shallow probing depth of soft X-ray spectroscopy compared to, e.g., neutron scattering, which is sometimes considered a disadvantage in the study of bulk materials, proves to be extremely advantageous for the study of epitaxial heterostructures that are typically not much thicker than 100 nm.

2 Spin, charge, orbital, and lattice degrees of freedom in epitaxial multilayers

A large number of $3d$ transition-metal oxides with strongly-correlated spin, charge, and orbital degrees of freedom crystallize in the perovskite structure with composition ABO_3, where B is a transition-metal ion, i.e., Sc – Zn, and A either an alkaline-earth (Mg – Ba), or a rare-earth ion (La – Lu) (Fig. 1(a)). This relatively simple, pseudo-cubic structure allows to combine different

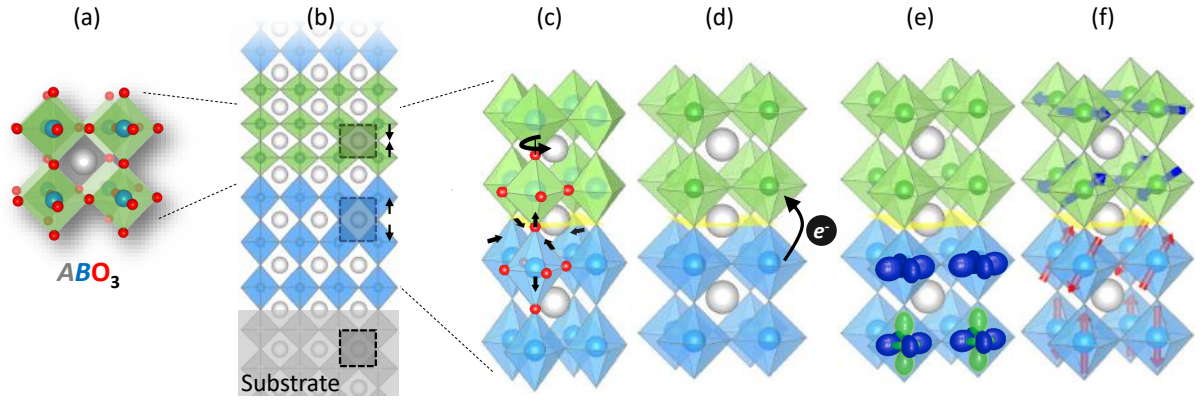


Fig. 1: *When combining different transition-metal oxide perovskites, ABO_3 (a), in an epitaxially strained heterostructure (b), different lattice (c), charge (d), orbital (e), and magnetic (f) reconstructions can occur at their interfaces.*

compounds in a cube-on-cube fashion in an epitaxial heterostructure (Fig. 1(b)). In this way, an artificially layered material can be created with emergent phases that are not present in the bulk phase diagrams of the individual components.

In fully strained heterostructures, the different B cation positions have a defined in-plane relationship resulting from the adaptation of the lattices at growth temperature. For cube-on-cube growth of the perovskite structure on a cubic substrate, exerting isotropic, biaxial strain (Fig. 1(b)), the simplest modification of the unit cell is a tetragonal distortion. This then leads to elongation (green material) or compression (blue material) of atomic distances along the growth direction, depending on the elastic properties of the material. However, the structural degrees of freedom in transition-metal oxide perovskites allow far more complex distortions, which can be understood in good approximation as tilts and rotations of rigid octahedra around the three cubic axes [10], as indicated in Fig. 1(c). These are determined by the lattice mismatch with the substrate, its crystal symmetry, the choice of the facet, and the connectivity conditions at the interfaces, and thus affect the lengths and angles of the B-O bonds. Therefore, it is important to consider structural distortions in the design, and to study them in detail in the grown heterostructures. *Ab-initio* theories such as density functional theory (DFT) have been shown to provide good predictions for structural modifications (see, e.g., Ref. [11]). Lower-energy scale electronic reconstructions, such as interfacial charge transfer or charge order can occur between multi-valence B ions (Fig. 1(d)). In addition, orbital polarization due to electronic confinement or reduced hopping across the interface due to a change in chemical bonding (Fig. 1(e)), as well as magnetic reconstruction, e.g., due to interfacial exchange coupling, can occur (Fig. 1(f)). The examples in Fig. 1 sketchily show different possible reconstructions, but in section 4 I will give concrete examples for different spin, charge, and orbital reconstruction mechanisms in heterostructures that our group has synthesized and studied in the past years.

The combination of theory, either by considering minimal models that capture the essential properties and *ab-initio* calculations that can investigate small, but relevant material differences, together with different experimental spectroscopy methods has proven to be a very powerful ap-

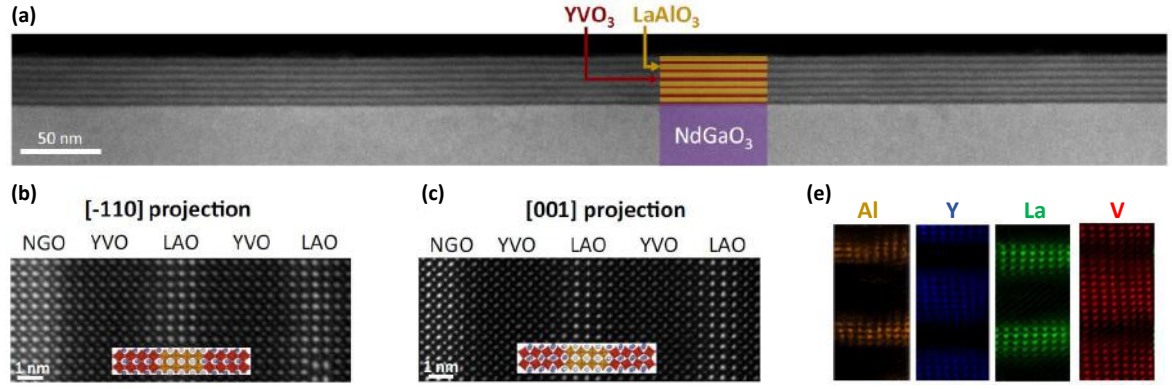


Fig. 2: *Scanning transmission electron microscopy (STEM) images of YVO_3-$LaAlO_3$ superlattices. (a) Low magnification image, (b,c) High-resolution images for two different crystallographic projections, revealing small structural distortions. (e) STEM-EELS maps show the elemental order in the layers. Reproduced from Refs. [19, 18].*

proach to predict and describe the properties of new material combinations. In addition, tremendous progress in the growth of epitaxial complex oxide heterostructures with atomic-layer precision by pulsed-laser deposition [4, 12, 9], magnetron sputtering [13, 14], and molecular-beam epitaxy [15–17] has been made. Nowadays it is possible to grow superlattices with altering layers of only a few unit cells (about 1 nm) of two materials with very high structural and chemical precision. As an example the scanning transmission electron microscopy (STEM) images of a YVO_3-$LaAlO_3$ superlattice, grown by pulsed-laser deposition on a $NdGaO_3$ substrate is shown in Fig. 2. The low-magnification annular dark-field STEM image, shown in Fig. 2(a) covers a lateral range of more than half a μm and the superlattice structure, i.e., the six-times repeated stacking of four unit cells of YVO_3 and eight unit cells of $LaAlO_3$, is highly ordered. The images with atomic resolution Fig. 2(b,c), which were taken in different projections with reference to the orthorhombic $NdGaO_3$ substrate structure, reveal structural distortions in form of tilts and rotations also in the superlattice layers. These structural modifications were reproduced by DFT+U calculations [18]. Based on the DFT-relaxed structures, the layer-resolved band structures were then calculated and brought into agreement with the orbital occupations determined in the experiment [19] (see section 4). In general, STEM is an important experimental method to study the structural distortions in heterostructures. In addition, electron energy loss spectroscopy (STEM-EELS) allows to study the element specific electronic structure with atomic resolution. However, typically only small spatial volumes are studied (often only a cross-section of 50 nm $\times$ 50 nm of a sample of 20 nm thickness is analyzed). Furthermore, the preparation of the electron-transparent TEM-lamella and high-energy electron beam damage can be critical issues for oxides, where the oxygen content is variable and often crucial for physical properties such as conductivity and magnetism.

Non-destructive X-ray spectroscopy, which includes X-ray absorption, resonant elastic and inelastic scattering, provides important additional and complementary experimental information, which I will discuss in detail in the following.

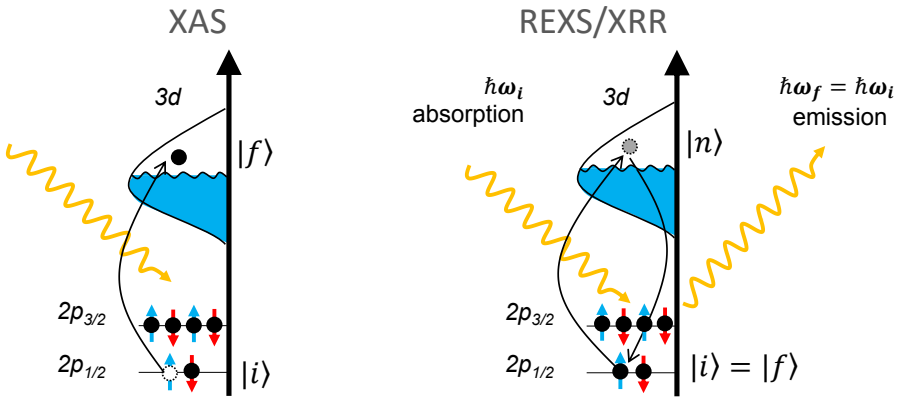


Fig. 3: *Schematic of X-ray absorption (XAS; left) and resonant elastic X-ray scattering processes (REXS; right).*

3 Soft X-ray spectroscopy

This lecture deals with the study of the spin, charge and orbital degrees of freedom of $3d$ transition-metal oxides. Therefore, we focus on the soft X-ray range as the relevant transition-metal $L_{3,2}$, oxygen K, and the rare-earth $M_{5,4}$ absorption edges fall into this energy window. The latter often occupy the A-cation sides in the ABO_3 perovskites and can lead to interesting exchange interactions between the localized Lanthanoide $4f$ and the transition-metal $3d$ moments (see section 4 and [20]). The soft X-ray range typically covers X-ray energies in the range from 50 to 1500 eV, corresponding to a wavelength range of 413–8.3 Å. We are interested in the spectroscopic information contained in the absorption fine structure that arises from dipole transitions of photo-exited core electrons to empty final states (Fig. 3). Fermi's golden rule provides the transition rate up to second order perturbation theory [21]

$$W = \frac{2\pi}{\hbar}\left|\langle f \mid H_{int} \mid i\rangle + \sum_n \frac{\langle f \mid H_{int} \mid n\rangle\langle n \mid H_{int} \mid i\rangle}{E_i - E_n}\right|^2 \delta(E_i - E_f), \tag{1}$$

where $\mid n\rangle$ denotes the intermediate (virtual) state and the sum runs over all possible intermediate states with energy E_n. The delta function reflects energy conservation and only depends on the energy of the initial and final state, and not on the energy of the virtual, intermediate states. The interaction Hamiltonian of X-rays with matter can be approximated by [22, 23]

$$H_{int} = \frac{e}{mc}\,\mathbf{p}\cdot\hat{\mathbf{A}} + \frac{e^2}{2mc^2}\,\hat{\mathbf{A}}\cdot\hat{\mathbf{A}}, \tag{2}$$

where m denotes the electron mass, $\mathbf{p}$ the electron momentum, and $\hat{\mathbf{A}}$ is the vector potential which characterizes the radiation field

$$\hat{\mathbf{A}}(r_k, t) \propto \sum_{i,k} \frac{1}{\sqrt{k}}\epsilon_i\left(a_{k,i}(t)e^{ikr} + a^\dagger_{k,i}(t)e^{-ikr}\right). \tag{3}$$

The first term in (2) is linear in creation $a_{k,i}$ (emission) and annihilation $a^\dagger_{k,i}$ (absorption) of a photon (with polarization vector ϵ), i.e., it describes processes with changes of ± 1 photons and

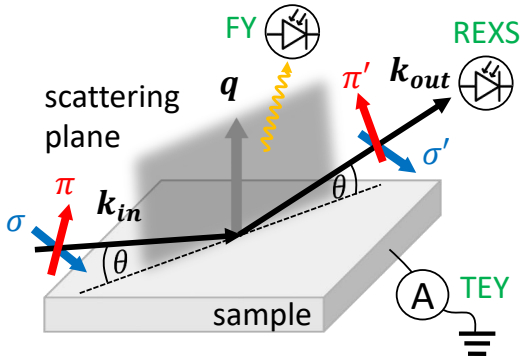


Fig. 4: *Schematic of the scattering geometry in a REXS experiment. θ is the scattering angle and k_{in} and k_{out} are the momentum vectors of incident and scattered waves, respectively. The scattering vector is defined by $\mathbf{q} = \mathbf{k}_{in} - \mathbf{k}_{out}$. The components of the polarization vectors are defined parallel π (π') and perpendicular σ (σ') to the scattering plane for the incident (scattered) beam. Absorption can be measured in fluorescence yield (FY) by a photodiode, or in total electron yield (TEY) by the drain current. REXS intensities are detected by a photodiode moving with the 2θ angle.*

therefore is relevant for the description of X-ray absorption spectroscopy (XAS). In the second term in (2) the vector potential is applied twice, i.e., it describes processes that change the number of photons by $\pm$ 2, or 0. That is the case in the scattering process (photon-in-photon-out). Resonant X-ray scattering thus is sensitive to the intermediate state (or the final state of the absorption). When the energy is conserved in the scattering process, i.e., $\hbar\omega_{\text{in}} = \hbar\omega_{\text{out}}$, we refer to it as being elastic and call it resonant elastic X-ray scattering (REXS) or in the case of $\hbar\omega_{\text{in}} \neq \hbar\omega_{\text{out}}$ to resonant inelastic X-ray scattering (RIXS). In the following we only consider elastic scattering. Restricting our self to electric dipole transitions ($E1$-$E1$ transitions in Ref. [24]) the interaction Hamiltonian simplifies to $H_{int} \propto \hat{\epsilon}'\mathbf{r}\hat{\epsilon}$, where $\mathbf{r}$ is the dipole operator and $\hat{\epsilon}$ and $\hat{\epsilon}'$ are again the polarization vectors of the incident and scattered light, respectively. They can be expressed by the orthogonal basis vectors perpendicular $\hat{\epsilon}_\sigma$ and parallel $\hat{\epsilon}_\pi$ to the scattering plane, $\hat{\epsilon} = \sigma\hat{\epsilon}_\sigma + \pi\hat{\epsilon}_\pi$, and accordingly for the one of the scattered photon (see sketch in Fig. 4).

In a crystal, each lattice site acts as a scattering center for the incident X-rays and is described by the atomic scattering amplitude

$$F(E, \mathbf{q}) = f_0(\mathbf{q}) + f_{\text{mag}}^{\text{non-res}} + f'(E) + if''(E). \tag{4}$$

Here $\mathbf{q} = \mathbf{k}_{\text{in}} - \mathbf{k}_{\text{out}}$ is the scattering vector. The first term in (4), $f_0(\mathbf{q}) \propto Zr_0$, is the so-called Thomson scattering, which is due to elastic scattering by a free charged particle. It arises from the scalar, isotropic polarizability of the scatterer, $f_0(\mathbf{q}) = f(q)\,\hat{\epsilon}' \cdot \hat{\epsilon}$, where $f(q)$ is the form factor, i.e., the Fourier transform of the particle's charge distribution. The second term is the non-resonant magnetic scattering amplitude $f_{\text{mag}}^{\text{non-res}}$ that contains the interaction between the magnetic field of the incoming wave with the spin of the electrons. As this term is usually small compared to the resonant terms, we will neglect it in the following. The last two terms in (4) are the so-called energy-dependent anomalous dispersion corrections $f'(E)$ and $f''(E)$.

The imaginary part $f''(E)$ is proportional to the XAS cross section, as we will see in the next section, and both terms, $f'(E)$ and $f''(E)$, determine the energy-dependent scattering studied in REXS. In a single-crystal diffraction experiment, photons are scattered with individual atomic scattering amplitudes $F_i(E, \mathbf{q})$ at different lattice sites i at position $\mathbf{r}_i$ in the lattice and interfere. The scattering cross-section then is proportional to $\left| \sum_i e^{i\mathbf{q}\cdot\mathbf{r}_i} F_i(E, \mathbf{q}) \right|^2$ and in the most general form $F(E, \mathbf{q})$ is a 3×3 tensor with complex entries

$$\hat{F} = \begin{pmatrix} F^{xx} & F^{xy} & F^{xz} \\ F^{yx} & F^{yy} & F^{yz} \\ F^{zx} & F^{zy} & F^{zz} \end{pmatrix}. \tag{5}$$

3.1 X-ray absorption and dichroism

The X-ray mass absorption coefficient $\mu(E)$ is defined by the Lambert-Beer law that describes the transmitted intensity through a material of density ρ and thickness d

$$I = I_0 e^{-\mu(E)\rho d}. \tag{6}$$

The atomic absorption coefficient is energy dependent and in the region between the absorption edges a continuous function, $\mu(E) \propto Z^4/E^3$, where Z is the atomic number. This means that X-rays with higher energy penetrate deeper. The inverse $1/\mu(E)$ is called the attenuation length as it describes the X-ray penetration depth into the material measured along the surface normal, where the intensity of the X-rays falls to $1/e$ of the value at the surface. Soft X-rays are strongly absorbed by matter, and typically the attenuation lengths in materials are less than 0.5 μm.[1] This means that the experiments have to be carried out in ultra-high vacuum and transmission measurements are not possible for most samples, except for powders or ultra-thin films on transparent membranes such as SiN. However, the decay products of the absorption process can be used to estimate the absorption. There are two principle decay processes of the created core hole: (i) emission of a fluorescence photon, and (ii) Auger decay, followed by secondary processes that emit electrons. The absorption cross section is then proportional to the (total) fluorescence yield (FY) that can be measured by using a photodiode that is placed at a position where elastic scattering is minimal (see Fig. 4). Alternatively, and optimally in addition to FY, the so-called total electron yield (TEY) can be measuring, e.g., by the drain current. When the photoelectrons leave the sample surface, a (small) current is generated by grounding the sample (Fig. 4). Both methods have advantages and disadvantages. Total FY measurements are bulk sensitive, but saturation effects in the vicinity of strong absorption lines can falsify the relative intensities in the fine structure. There are very interesting alternative measurement methods, which are, however, experimentally more complex [25]. The problem of saturation is less relevant in TEY measurements [26], but this type of detection is rather surface sensitive, because it depends on the effective escape depth of the photoelectrons, which is often less than 5 nm, and can vary strongly [27]. Since we focus on thin-film structures, which are usually thinner or about 5 nm thick, surface sensitivity is not a critical issue.

[1] The X-ray attenuation length for different solids can be looked up, for example, at https://henke.lbl.gov.

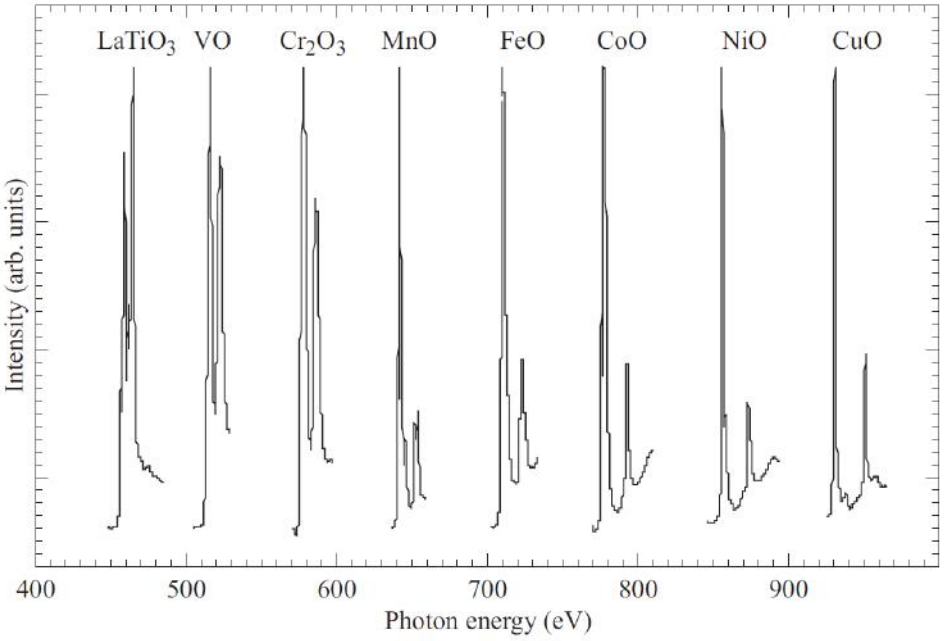


Fig. 5: *L-edge XAS spectra of a series of* $3d$ *transition-metal oxides. Taken from Ref. [29]*

By varying the energy, which can be done with high resolution (a typical energy band width is $\sim$100 meV) and over a wide energy range at a synchrotron beamline, one can observe the above discussed absorption edges, when the energy of the incoming X-rays match the intra-atomic transition energies (see the X-ray data booklet [28]). These transitions have well separated energies, which is the reason for the element sensitivity of the method. As mentioned above, we will focus on electric-dipole transitions, since they have the largest cross section. In forward direction $\mathbf{q} \approx 0$, the imaginary part of the scattering amplitude is related to the absorption cross section

$$I_{\text{XAS}} \propto -\frac{1}{E}\text{Im}\left[\hat{\epsilon}\cdot F(E)\right] \propto \left|\langle f \mid \hat{\epsilon}\cdot\mathbf{r} \mid i\rangle\right|^2 \delta(E_i - E_f - \hbar\omega). \tag{7}$$

Only transitions are allowed, which fulfill the dipole selection rules: $\Delta L = 1$, $\Delta m = \pm 1$. It is common to use spectroscopic notation to label specific transitions depending on the involved core level, i.e., the K-edge corresponds to $1s$ to, e.g., $2p$ or $4p$, the $L_{3,2}$ to $2p \rightarrow 3d, 4d, ...$, and the $M_{5,4}$ edge to $3d \rightarrow 4f, 5f$ orbital transitions, where the indices refer to the spin-orbit split core levels $p_{1/2}$, $p_{3/2}$, and $d_{5/2}$, $d_{3/2}$, respectively. For the energy of K-edges $E_K \approx Z(Z-1)\times 13.6$ eV provides a good approximation. Here Z is the atomic number of the element and the $Z-1$ term accounts for the screening of the nuclear charge by the second $1s$ electron. The L- and M-edges have a more pronounced fine structure. Here separated absorption lines are observed due to the spin-orbit splitting of the core levels, e.g., the two $L_{3,2}$ lines arise from the spin-orbit splitting of the $2p_{3/2}$ and $2p_{1/2}$ core levels (see Fig. 3). Their separation accordingly increases in the $3d$ transition-metal row from Sc to Zn as Z increases (Fig. 5). The XAS fine structure measured with higher energy resolution then provides detailed information about the valence state of a particular ion, its spin state, orbital occupation, as well as spin and orbital contribution to the magnetic moment and possible antiferromagnetism. For this purpose,

however, the fine structure must be examined closely, as its interpretation depends on whether the final states are more localized (like f-states) or delocalized (like p-states). The d-states, which are of particular interest to us here, lie somewhere in between, i.e., they are neither fully localized nor fully itinerant [1], which is precisely the cause of the strong correlation of local degrees of freedom that gives rise to their interesting physics.

While K-edge spectra corresponding to transitions from the single $1s$ core level to rather delocalized, empty p states, reflect the site- and symmetry projected unoccupied density of states and are often sufficiently described by mean-field approaches, e.g., local density approximation (LDA) or DFT, the M edges of the rare-earth ions have strong, sharp peaks near the edges that show strong atomic multiplet effects. These multiplet effects show when a core other than a $1s$ is present in the initial state and because of the strong $4f$ localization, since then there is significant overlap of core and valence wave functions in the ground state. This also applies for transition-metal L edges, where the multiplet structure, which is hardly screened in the solid as compared to the core potential, determines the spectral shape and influences the $L_{3,2}$ or $M_{5,4}$ branching ratio [30]. Different valence states show in a shift of spectral weight of the absorption lines. For anions (cations) with different valence state the absorption edge is shifted to lower (higher) photon energies, because of the lower (higher) ionization potential. Information on orbital occupations and magnetic moments can be obtained from the polarization-dependent fine structure. To describe the fine structure of L or M edges, many-body ligand-field cluster calculations have been shown to be particularly successful to determine important parameters, such as the crystal field splittings in NiO and spin states in cobaltates from the comparison with the experimental data [29]. When deriving a minimal tight-binding model from downfolding the DFT band structure to localized Wannier orbitals, important hopping parameters to the ligand ions can be derived in an *ab-initio* fashion from DFT($+U$) [31]. X-ray dichroism, that is the dependence of X-ray absorption on the polarization of the incident photons, occurs when the spherical symmetry at the site of the atom is broken by a magnetic or (crystalline) electric field [35]. Then the charge density around an atom becomes anisotropic. Depending on its origin and the light polarization used to detect it, one distinguishes X-ray natural linear dichroism (charge anisotropy due to crystal field), X-ray magnetic linear dichroism (charge anisotropy parallel and perpendicular to the magnetization axis) [36], and X-ray magnetic circular dichroism [35]. In the first case, for example, the scattering tensor of an atom whose $3d$ states are split by a tetragonal crystal field has unequal diagonal elements along the x and z direction

$$\hat{F} = \begin{pmatrix} F^{xx} & 0 & 0 \\ 0 & F^{xx} & 0 \\ 0 & 0 & F^{zz} \end{pmatrix}. \tag{8}$$

Therefore, the linear polarization dependence of transition-metal L edge XAS provides information on the $3d$ orbital occupation of the system. The intensity of polarized XAS along x, y and z direction is proportional to the number of holes (h) in xz, yz, xy, x^2-y^2, and $3z^2-r^2$ orbitals that have lobes along that direction. Sum rules allow to relate the d-orbital occupations (h to the total, integrated intensities I_i measured with $i = x$, y and z linear polarized X-rays

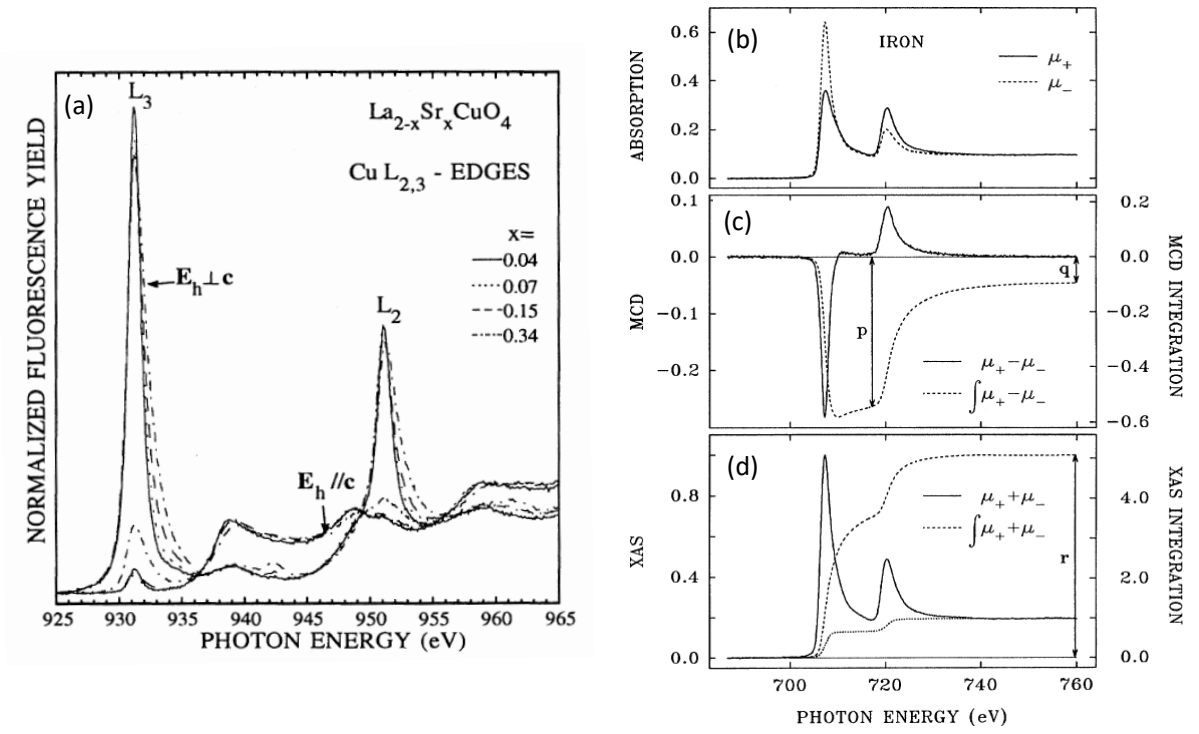


Fig. 6: *(a) Example for natural XLD in a $La_{2-x}Sr_xCuO_4$ thin film measured in FY (taken with permission from Ref. [32]), where the data were reproduced from Ref. [33]). (b-d) Example for circular magnetic dichroisms measured in a iron thin film (taken with permission from Ref. [34]).*

over the entire $L_{3,2}$ edge [32]

$$\begin{aligned} I_x &= \frac{1}{2}h_{xy} + \frac{1}{2}h_{xz} + \frac{2}{3}h_{x^2} \\ I_y &= \frac{1}{2}h_{xy} + \frac{1}{2}h_{yz} + \frac{2}{3}h_{y^2} \\ I_z &= \frac{1}{2}h_{xz} + \frac{1}{2}h_{yz} + \frac{2}{3}h_{z^2}. \end{aligned} \tag{9}$$

For clarity, we write these for symmetric e_g-orbitals x^2, y^2, z^2, which are related to the real wave functions as given in Ref. [29]. For $3d$ electron systems with fully filled t_{2g} and partially filled e_g orbitals, the sum rules simplify, and we can directly relate the ratio of e_g holes to the integrated XAS intensities for in-plane ($I_{x,y}$) and out-of-plane (I_z) polarization

$$X = \frac{h_{3z^2-\mathbf{r}^2}}{h_{x^2-y^2}} = \frac{3I_z}{4I_{x,y} - I_z}, \tag{10}$$

where $h_{x^2-z^2}$ and $h_{3z^2-\mathbf{r}^2}$ denote the number of holes in the $d_{x^2-z^2}$ and $d_{3z^2-\mathbf{r}^2}$ orbitals, respectively. Since for t_{2g}-systems the e_g-orbitals have finite hole occupations, the orbital occupations cannot be determined directly from the measured spectra, but cluster calculations can be used, as we will see in section 4.

The Cu-$L_{3,2}$ XAS data measured by Chen *et al.* [33], which are reproduced in Fig. 6(a), provide a very clear example for X-ray natural linear dichroism that originates from differences in

orbital occupation. In the parent compound ($x = 0$) of the high-temperature superconductor $La_{2-x}Sr_xCuO_4$ the valence configuration of Cu is $3d^9$. The Cu ions are square-planar coordinated by four oxygen ions, and the D_{4h} crystal field splitting results in one hole occupying the highest-energy $d_{x^2-y^2}$ orbital in the ground state. The $2p$ core electrons can be excited to this empty state with in-plane x or y polarization ($E \perp c$ in Fig. 6(a)), but no empty states are available for excitation with z polarization ($E \parallel c$ in Fig. 6(a)).

X-ray magnetic circular dichroism (XMCD) is the difference in absorption spectra measured with circular positive (σ^+) and negative (σ^-) polarization, and allows to study magnetism. Again, we consider the $3d$ transition metal L edges, since the magnetic properties are mainly determined by their d-valence electrons. In the absorption process, σ^+ and σ^- polarized photons transfer their angular momentum, $\pm\hbar$, respectively, to the excited photoelectron. Due to spin-orbit coupling in the core level, which is $l+s$ for $2p_{3/2}$ and $l-s$ for $2p_{1/2}$, i.e., opposite in sign, the angular momentum is in part transferred to spin momentum, and the different polarizations create photoelectrons with opposite spins at L_3 and L_2. The spin-split $3d$ final states can then only be reached by excited photoelectrons with the appropriate spin. The quantization axis is given by the magnetization direction, i.e., the maximum dichroism is measured, when magnetization axis and photon momentum are parallel.

Considering the scattering tensor for cubic, ferromagnetic materials with in-plane aligned magnetization in the scattering plane, the diagonal elements are equal, but two off-diagonal elements are non-zero and proportional to the XMCD signal measured in an absorption experiment [37]

$$\hat{F} = \begin{pmatrix} F^{xx} & iF^{xy} & 0 \\ -iF^{xy} & F^{xx} & 0 \\ 0 & 0 & F^{xx} \end{pmatrix}. \tag{11}$$

Important sum rules can also be derived for circular dichroism. The sum of the integrated intensities I_{L3} and I_{L2} of the polarization-averaged spectrum is again proportional to the total number of d-holes (charge sum rule). Following the notation in Ref. [35], we label energy integrals over the XMCD difference spectrum as A for the energy range of the L_3 edge and B for the L_2 edge, respectively. Then the sum rules allows to quantitatively determine the spin moment from the measured intensity $A-2B$, and the orbital moment is obtained from the dichroic intensity $A+B$ [35].

An example for X-ray magnetic circular dichroism in ferromagnetic iron is given in Ref. [34]. The spectra are shown Fig. 6(b), where intensities measured with right and left circular polarized light (labelled with μ^+ and μ^-, respectively) show a clear difference. In the XMCD spectrum (Fig. 6(c)) it can clearly be seen that this difference has an opposite sign for L_3 and L_2 edges. By integrating the spectrum over the respective energy ranges and integrating the polarization-averaged spectrum over the whole energy range (Fig. 6(d)), spin and orbital moments of $m_{\text{spin}} = 1.98$ and $m_{\text{orb}} = 0.085\ \mu_B$/Fe were determined from the sum rules [34].

3.2 Resonant X-ray scattering

Resonant scattering combines information on spatial modulation from diffraction with the spectroscopic information provided by X-ray absorption in a single experiment [38, 23]. Resonant scattering is element sensitive due to the strong enhancement of the cross section. Furthermore, the strong dependence of the intermediate state on the spin, orbital, and charge configuration of the resonant scattering centers provides access to local properties that I have already discussed in the context of XAS. This information is contained in the energy-dependent $f'(E)$ and $f''(E)$ terms, which are connected via the Kramers-Kronig relation. Both, real and imaginary part are required for the description of resonant elastic X-ray scattering (REXS). If the incident photon energy $\hbar\omega_i$ is very different from the resonance energy $E_i - E_f$ of the system, we say that we are in the non-resonant regime and assume that the scattering is largely independent of energy. Here, however, one must be careful when analyzing scattering intensities as a function of energy, since strong, sharp intensity variation across the resonances in the imaginary part $f''(E)$, (which is proportional to the absorption cross section as discussed above) lead to broader structures in the real part. Depending on how the real and imaginary part mix for a given momentum transfer, anomalous scattering signals can already occur for incident photon energies smaller than those in $f''(E)$. To make this point a little clearer we can look at the example in Fig. 7. Since the fine structure is not captured by tabulated theoretical values [39, 40], the real part is usually obtained via the Kramers-Kronig relation from the imaginary part by including measured or calculated resonance lines. The example shows the procedure for La-$M_{5,4}$. These correspond to transitions from the La $3d$ core electron to completely empty $4f$ states. Therefore, the $f''(E)$ fine structure consists essentially of two sharp Lorentzian lines (middle panel in Fig. 7). When using the Kramers-Kronig relation, sufficient extrapolation of the data outside the measurement range is important, since the integral runs over frequencies from zero to infinity. Therefore, the measured or calculated absorption fine structure data are scaled to tabulated data that are available over a wide energy range (up to 400 keV [40]). From the example in Fig. 7 it can be seen that if we choose an energy well before or in between the resonances in $f''(E)$, intensity variations in f' are still clearly visible, i.e., scattering signals measured even more than 100 eV away from the absorption resonance lines measured in XAS cannot generally be called "non-resonant". Since the REXS cross section

$$I_{\text{REXS}} \propto \sum_n \left| \frac{\langle f \mid \hat{\varepsilon} \cdot \mathbf{r} \mid n \rangle \langle n \mid \hat{\varepsilon}' \cdot \mathbf{r} \mid i \rangle}{E_i - E_n} \right|^2 \delta(E_i - E_f - \hbar\omega), \tag{12}$$

is sensitive to the intermediate state $|n\rangle$, which is the final state in XAS, all spectral information that can be gained from XAS, and was discussed above, is contained in the scattered intensity. Moreover, in the REXS experiment, one sees the spatial modulation of the respective properties due to the q dependence. To exploit this, different scan types are used to measure the moment-, energy- and polarization-dependent scattering intensity, which then need to be carefully analyzed, and simulations are often required to obtain quantitative information about modulations of the orbital occupation, charge order, or direction and magnitude of the magnetic moments [23]. However, element-specific electronic and magnetic properties can then be

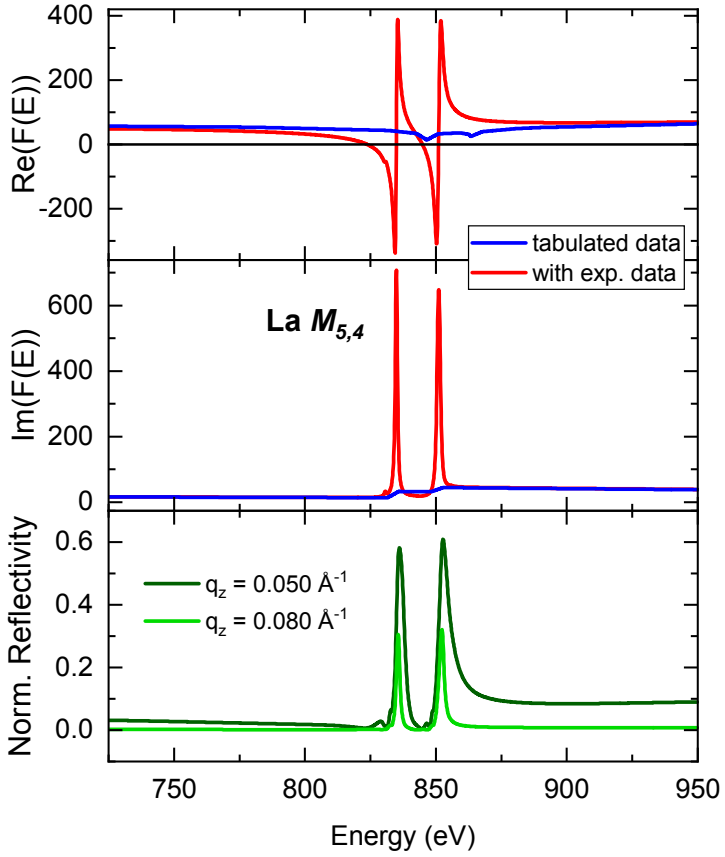


Fig. 7: *Real and imaginary part of the energy-dependent scattering factor for the compound $LaAlO_3$ in the energy range of the La-$M_{5,4}$ absorption edge. The real $f'(E)$ with resonance (red curve in the top panel) was obtained by Kramers-Kronig transformation of the experimentally determined XAS scaled and extrapolated with the tabulated data of $f''(E)$ (middle panel). The bottom panel shows the normalized X-ray reflectivity of a 30 nm thick film of $LaAlO_3$ on $SrTiO_3$ substrate at two different, fixed q_z values over the same energy range.*

determined, especially in transition-metal heterostructures, which are not accessible with other experimental techniques, as I will show in the examples in section 4. A disadvantage of REXS in the soft X-ray range is the relatively limited, accessible Ewald sphere, i.e., the limitation of momentum transfer by the wavelength of the incoming X-rays. For soft X-rays, the Ewald sphere is in the order of 0.1 Å^{-1}, which is sufficient to probe Bragg planes with separation of at least 10 Å.

3.3 X-ray resonant reflectometry

X-ray reflectometry, usually measured in the hard X-ray range far away from strong resonances, is an established method for the structural characterization of thin films and multilayers (see sketch in Fig. 8). It is based on the simple concept of multiple reflection and refraction of radiation at a surface and interfaces. In a typical experiment, the intensity of a scattered beam $R(\theta)$, normalized to the incoming intensity I_0, is measured as a function of the incident angle θ.

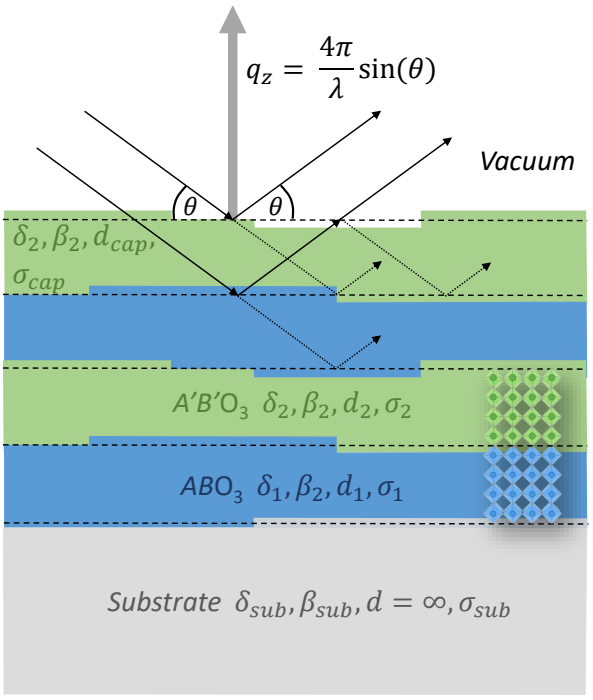


Fig. 8: *Specular scattering geometry ($\theta_{in} = \theta_{out}$, q_z parallel to the surface normal) and typical set of structural parameters used to simulated XRR data. The sketch shows an example of an ABO_3-$A'B'O_3$ superlattice with two repetitions of the bilayer, and with energy-dependent optical constants $\delta_{1,2}(E)$ and $\beta_{1,2}(E)$. The layer thicknesses d and root-mean-square roughnesses σ of different layers are usually fitting parameters.*

In such a scattering process, a momentum $q = 4\pi\sin(\theta)/\lambda$ is transferred.[2] The term reflectometry refers to specular scattering with $\theta_{\text{in}} = \theta_{\text{out}}$, i.e., q_z points along the surface normal z (Fig. 8). As is known from optics, when light emerges from an optically denser medium into an optically thinner one, there is a critical angle θ_c (depending on the wavelength λ), below which all incoming intensity is reflected on the surface (total internal reflection). Above θ_c, part of the radiation penetrates the material, interacts and parallel beams obtain a phase difference and interfere [41]. This results in characteristic features in the X-ray reflectivity of multilayers, such as Kiessig fringes, superlattice peaks, and changes in slope due to surface and interface roughness. By fitting the structures using calculated scattering factors (see Refs. [40,39]) structural parameters, such as the individual layer thicknesses and roughnesses (d_i and σ_i in Fig. 8), can be determined with high precision. When analyzing X-ray resonant reflectivity (XRR) data measured at energies near or at the resonances, and where one aims to determine layer-resolved changes in the energy-dependent scattering factors, the structural parameters are kept fixed. Since only an out-of-plane momentum transfer is considered, the layers can be treated as a homogeneous medium and the reflectivity can be described in the so-called optical approach. To do this, we introduce the complex refractive index in the X-ray range[3]

[2] It is practical to note the conversion between wavelength λ and energy E: $\lambda[\text{Å}] = 12398.4244/E[\text{eV}]$

[3] In some references the refractive index is defined by $n = 1-\delta-i\beta$, consistent with an opposite sign in the wave equation.

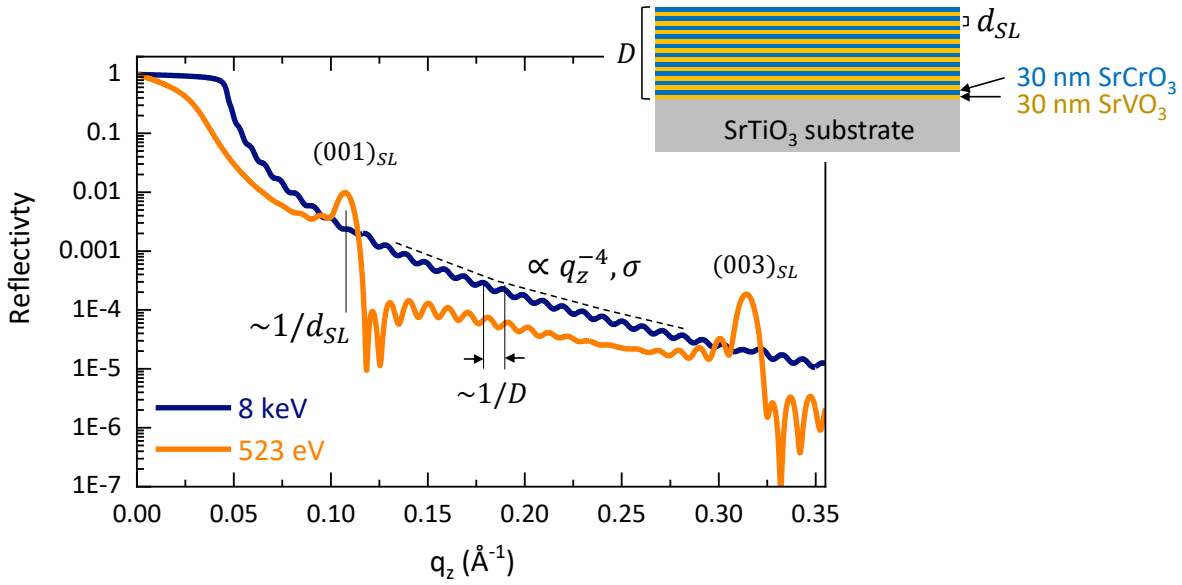


Fig. 9: *Example for the momentum-dependent XRR from a superlattice composed of ten repetitions of a [$SrVO_3$ (30 nm) / $SrCrO_3$ (30 nm)] bilayer on a $SrTiO_3$ substrate. The labels indicate characteristic features in the reflectivity curve, such as the total thickness D, the bilayer thickness d_{SL}, and possible surface roughness σ. The simulated XRR curves are shown for X-ray energies at 8 keV, i.e., far from resonances, and close to the V-L_2-resonance at 523 eV to demonstrate the effect of the energy-dependent change of contrast from the scattering factors. Since both materials have very similar electron densities the superlattice reflections $(00l)_{SL}$ are invisible in the non-resonant data, but clearly show in the V-L_2 data (see Fig. 12(a)). Note that the $(002)_{SL}$ reflection is not allowed in this specific superlattice structure with identical thicknesses of $SrCrO_3$ and $SrVO_3$ and no interface roughness.*

$$n = 1 - \delta + i\beta. \tag{13}$$

At energies (E) close to the resonance edges in a material, the atomic scattering factor (Eq. 4) shows strong variations in the energy-dependent real $f'(E)$ and imaginary $f''(E)$ part of the dispersion corrections. In case of forward scattering ($\mathbf{Q} \approx 0$) and negligible non-resonant magnetic scattering $f^{\text{mag}}_{\text{non-res}}$ [42], equation (4) reduces to

$$F(\mathbf{q} \approx 0, E) = Z^* + f'(E) + if''(E), \tag{14}$$

where $Z^* = Z - (Z/82.5)^{2.37}$ is the atomic number Z with a small relativistic correction [28]. The optical theorem connects the imaginary part of the scattering factor to the dielectric function $\varepsilon(E)$ by

$$f''(E) = -\frac{E^2}{2\pi(c\hbar)^2 N_p} \text{Im}\sqrt{\varepsilon(E)}, \tag{15}$$

where c is the speed of light, $\hbar$ the Planck constant, N_p the number of photons, and E the energy of the X-rays. Just like the $F(E)$ (Eq. 5), $\varepsilon(E)$ has the form of a 3×3 tensor

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon^{xx} & \varepsilon^{xy} & \varepsilon^{xz} \\ \varepsilon^{yx} & \varepsilon^{yy} & \varepsilon^{yz} \\ \varepsilon^{zx} & \varepsilon^{zy} & \varepsilon^{zz} \end{pmatrix}, \tag{16}$$

with complex entries $\varepsilon^{ij} = (\varepsilon_1)^{ij} + i(\varepsilon_2)^{ij}$ $(i, j = x, y, z)$ that are related to δ^{ij} and β^{ij} by $\varepsilon_1^{ij} = (1-\delta^{ij})^2 - (\beta^{ij})^2$ and $\varepsilon_2^{ij} = 2(1-\delta^{ij})\beta^{ij}$. In a compound consisting of N different atoms, δ and β are then given by the sums of atomic scattering factors weighted by the density ρ of the material

$$\delta(E) = \frac{2\pi\rho r_0(c\hbar)^2}{ME^2}\sum_{j=1}^{N} C_j\big(Z_j^* + f_j'(E)\big) \tag{17}$$

$$\beta(E) = \frac{2\pi\rho r_0(c\hbar)^2}{ME^2}\sum_{j=1}^{N} C_j f_j''(E). \tag{18}$$

Here C denotes the number of atoms of type j per formula unit, M is the molar weight of the compound, and r_0 is the Thompson scattering amplitude (classical electron radius). Furthermore, we know from equation (7) that $f''(E)$ is proportional to the X-ray absorption cross section I_{XAS}. Therefore, we can combine both relations to obtain reliable resonant tensor entries δ and β, for the different layer stacks, as shown in Fig. 7 for $LaAlO_3$. These can then be further modeled to obtain layer-specific optical constants, corresponding to the different reconstruction scenarios in a given material system. These can be, for example, different, layer-dependent orbital polarizations, as we have studied in the examples I will show in the next section 4.

Depending on the symmetry of the material, the dielectric tensor (Eq. 16), just like the scattering tensor takes on a simplified form [43]. For example, for materials with cubic, tetragonal and orthorhombic symmetries, all non-diagonal elements are zero. Going from orthorhombic to tetragonal and cubic symmetry, the tensor further simplifies with $\varepsilon_{xx} = \varepsilon_{yy}$ for tetragonal and $\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{zz}$ for cubic symmetries. Just as for $F(E)$, the dielectric tensor of ferromagnetic materials has specific, non-zero off-diagonal elements that are proportional to the X-ray magnetic circular dichroism (XMCD). Therefore, X-ray resonant magnetic reflectivity allows to measure magnetic moments of deeply buried atomic, magnetic layers in a multilayer [44].

4 Case studies

In the following, I will present some of our studies on complex oxide heterostructures, each with a different type of interface reconstruction as outlined in the introductory Fig. 1, and which we investigated using the different X-ray spectroscopy techniques presented in the previous section. In the first example, we show how the element sensitivity of XAS can be used to investigate an interfacial doping mechanism in a cuprate-nickelate hybrid structure. The second example shows how linearly polarized resonant X-ray reflectometry can be used to determine depth-resolved orbital polarization profiles in a quantitative manner. The last example shows how resonant elastic X-ray scattering provides unique information about noncollinear magnetic orderings and how such ordered moments can interact with other moments across interfaces.

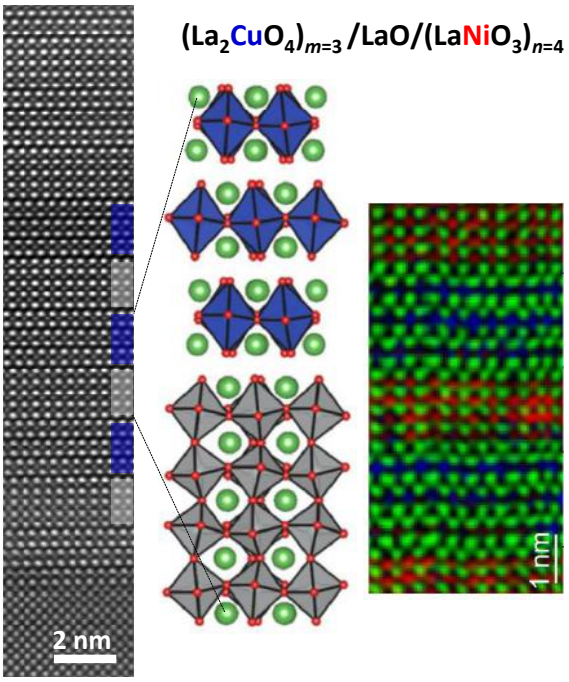


Fig. 10: *Left panel: High-angle annular dark field (HAADF) STEM image of the* $m = 3$, $n = 4$ *cuprate-nickelate hybrid structure grown on* $(LaAlO_3)_{0.3}(Sr_2AlTaO_6)_{0.7}$ *(LSAT) substrate. Middle panel: Sketch of the bilayer structure with the composition indicated in the above label. Right panel: Spatially-resolved elemental distribution extracted from the electron energy loss spectra (EELS) with color code: La–green, Cu–blue, and Ni–red, respectively. Reproduced with permission from Ref. [45].*

4.1 Interfacial doping in La_2CuO_4-$LaNiO_3$ hybrid structures

Layer-by-layer oxide molecular-beam epitaxy allows to grow complex oxides with atomic layer precision. We used this technique to synthesize a cuprate-nickelate multilayer structure (Fig. 10) and showed that these structures allow a clean separation of dopant and doped layers. The multilayer growth of La_2CuO_4 and $LaNiO_3$ can only be achieved with two LaO layers separating cuprate and nickelate blocks. This translates to an extra $La^{3+}O^{2-}$ atomic layer in the bilayer formula $(La_2CuO_4)_m$/LaO/$(LaNiO_3)_n$ (m, n integers), resulting in an additional charge of $+1$ at each interface, which we suspected to lead to a change in the nickel or copper electronic structure. We investigated this in detail using XAS and explicitly exploited element sensitivity. The spectra measured with soft X-rays across the Cu-$L_{3,2}$ and Ni-$L_{3,2}$ absorption edges are shown in Fig. 11(a,b). The measurements were performed with linearly polarized X-rays parallel (I_x) and perpendicular (I_z) to the interfaces. While the Cu-L spectra are characteristic of Cu^{2+} [46], the Ni-L edge spectrum shows signs of a mixture of Ni^{2+} and Ni^{3+}. The octahedral crystal field splits the Cu and Ni $3d$ levels into energetically lower t_{2g} and higher e_g orbitals. An additional elongation of the CuO_6 and NiO_6 octahedra along the $[001]$ direction, observed by STEM [45], leads to a further splitting of the e_g orbitals. This is reflected in the polarization dependence, where I_x (I_z) probes holes in the $d_{x^2-y^2}$ ($d_{3z^2-r^2}$) orbital. The resulting normalized linear dichroic difference spectrum $(I_x-I_z)/(2I_x+I_z)$ at the Cu-L edge is very

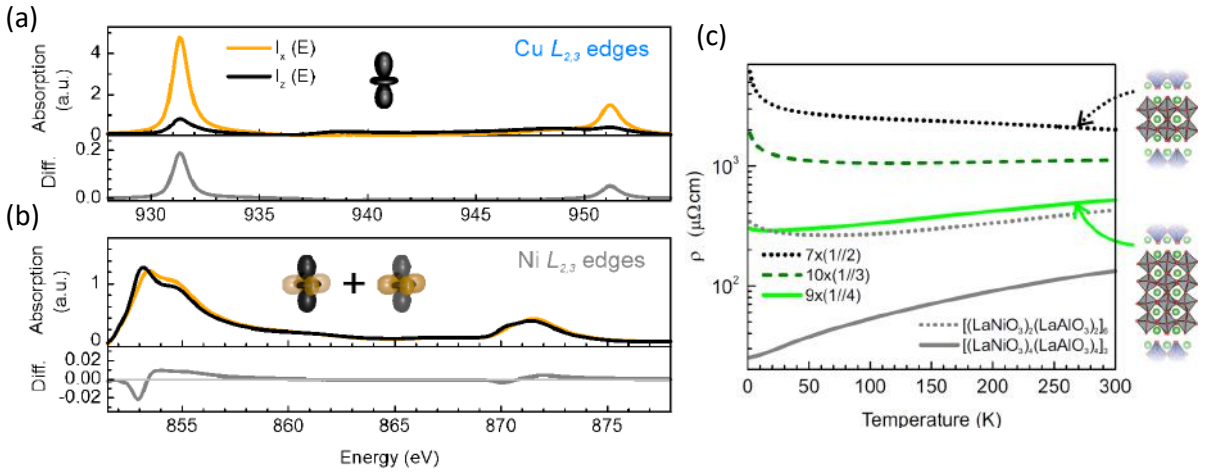


Fig. 11: *XAS spectra measured with the polarized X-rays parallel (orange curves) and perpendicular (black curves) to the sample surface across (a) the Cu-$L_{3,2}$ (b) the Ni-$L_{3,2}$ edges for the m=3, n=4 cuprate-nickelate hybrid structure shown in Fig. 10. In the bottom panels the normalized dichroic signals (grey curves) are shown. (c) Temperature-dependent resistivity of [(La_2CuO_4)$_m$/LaO/($LaNiO_3$)$_n$]$_l$ (m=1 n= 2, 3, 4, and l= 7, 10, 9) with average formal Ni valences of 2.5+, 2.67+, and 2.75+ for n= 2, 3, 4, respectively, compared to [($LaNiO_3$)$_n$($LaAlO_3$)$_n$]$_k$ superlattices with n= 2, 4 and k= 6, 3 with 3.0+ Ni valence. Taken with permission from Ref. [45].*

pronounced (Fig. 11(a) to be compared with Fig. 6(a)) and arises from a Jahn-Teller distortion that lowers the energy of the $d_{3z^2-r^2}$ orbital [32], leaving a hole in the $d_{x^2-y^2}$ orbital for the XAS final state. The Ni-L edge spectra also show linear dichroism, although less pronounced. To quantify this effect, we used the sum rules for e_g linear dichroism (Eq. 10). While DFT+U results show $X = 1$ for bulk $LaNiO_3$ with rhombohedral structure (space group $R\bar{3}c$, where all Ni-O distances are equal), we find a smaller $X_{\text{av}} = 0.94$ from XAS, corresponding to a higher $d_{3z^2-r^2}$ occupation on average in the $LaNiO_3$ stacks in the hybrid structures [45]. The layer-resolved DFT+U calculations show that the effect is most pronounced in the interface layers ($X_{IF} = 0.84$), which correlates with the stronger elongation of the interfacial NiO_6 octahedra in the [001] direction as seen by STEM [45]. The corresponding value in the central layers is $X_C = 0.91$.

In addition, DFT+U predicts a charge disproportionation between neighboring in-plane Ni sites that occurs predominantly in the interface layers and causes a band gap in the density of states of the interface layers. To test this prediction we preformed in-plane electronic transport measurements on different hybrid structures with m=1 and decreasing $LaNiO_3$ layer thickness n= 4, 3, 2 (Fig. 11 (c). The temperature-dependent resistivity shows metallic behavior for n=4, which we attribute to currents running through the inner, at most weakly disproportionated, metallic $LaNiO_3$ layers. When decreasing $LaNiO_3$ from four to three monolayers, we observe a metal-to-semiconductor transition, and finally, for n=2, a semiconducting behavior. Consistent with this observation, DFT+U results for n=2 indicate a band gap of 0.28 eV [45]. To distinguish confinement and doping effects, we compare the (La_2CuO_4)$_m$/LaO/($LaNiO_3$)$_n$ hybrid

structures with $[(LaNiO_3)_n(LaAlO_3)_n]_k$ (n=2, 4 and k=6, 3) superlattices, where the nickelate layers are confined to the same thickness, while remaining undoped (Ni^{3+}). The latter superlattices will be discussed in detail in the next section. Leaving differences in the lattice parameters of the two systems aside, it is interesting to note that the resistivity of the $LaNiO_3$-$LaAlO_3$ superlattice with two nickelate layers is comparable to the nickelate-cuprate hybrid structure with four nickelate layers. This implies that only the inner two layers are conducting, as predicted by the DFT calculations. In conclusion, our study on the cuprate-nickelate hybrid structures showed that doped electrons are accommodated primarily in the interfacial nickelate layers, where they induce a digital modulation of the Ni valence state and a rearrangement of the Ni-$3d$ orbital occupation.

4.2 Orbital reflectometry of nickelate and vanadate superlattices

The d orbital occupations and the strength of hybridization with the oxygen ligands determine the electronic transport properties and the magnetic exchange interactions and their anisotropy, via the Goodenough-Kanamori-Anderson (GKA) rules for superexchange [47, 48]. As pointed out in the introductory section a common effect created at interfaces is the confinement of electrons. We have studied heterostructures of two prototypical correlated oxides, Mott-Hubbard insulting YVO_3 and the negative charge-transfer system $RNiO_3$ (R = rare-earth ion) [49]. While the vanadates are a $3d$-t_{2g} electron system, in the nickelate the higher-lying e_g states are partially occupied. If we consider an interface of $LaNiO_3$ (YVO_3) with a wide band-gap insulator, such as $LaAlO_3$, the (virtual) hopping of electrons along the Ni(V)-O-Al bond is largely suppressed, due to the band-gap of $LaAlO_3$, which leads to the preferential occupation of the orbitals with lobes in the plane of the interface. In addition, the character of the Ni-O-Al or V-O-Al chemical bond changes, which is accompanied by a change in the oxygen hybridization of the d states. Last but not least, the modification of the local crystal fields impacts orbital polarization, as at the interface it is expected to be different from both, the bulk as well as in layers deeper in the stack that are further away from the interfaces. In superlattices of $LaNiO_3$ or YVO_3 with $LaAlO_3$, this interface is repeated several times periodically (see Fig. 2), which facilitates the investigation by means of XRR. The modulation in the electronic structure of interface and central layer in stacks of $LaNiO_3$ or YVO_3 is seen in so-called orbital reflectometry, where one measures the XRR with linear polarized light. A simple calculation of the structural factors shows this sensitivity (Fig. 12). If we consider a symmetric superlattice with the same thickness of stacks of two compounds ABO_3 and $AB'O_3$ (the example in Fig. 12(a) show a $(6/6)$ superlattice, with six pseudo-cubic unit cell of each material), without any symmetry breaking the even-order, $(00l)$, l= 2, 4, ... superlattice peak intensities vanish. If, however the electronic structure of interface (IF) and central (C) layers are different, resulting in different scattering factors f^{IF} and f^{C} due to interface reconstructions, this selection rule is broken and the intensity becomes proportional to the difference in the scattering factors (Fig. 12(b)). The same sensitivity for interface reconstructions arises in asymmetric $(8/4)$ and $(4/8)$ at the (003) reflection (Fig 12(c)). We take advantage of this in orbital reflectometry and simulate the polarization-

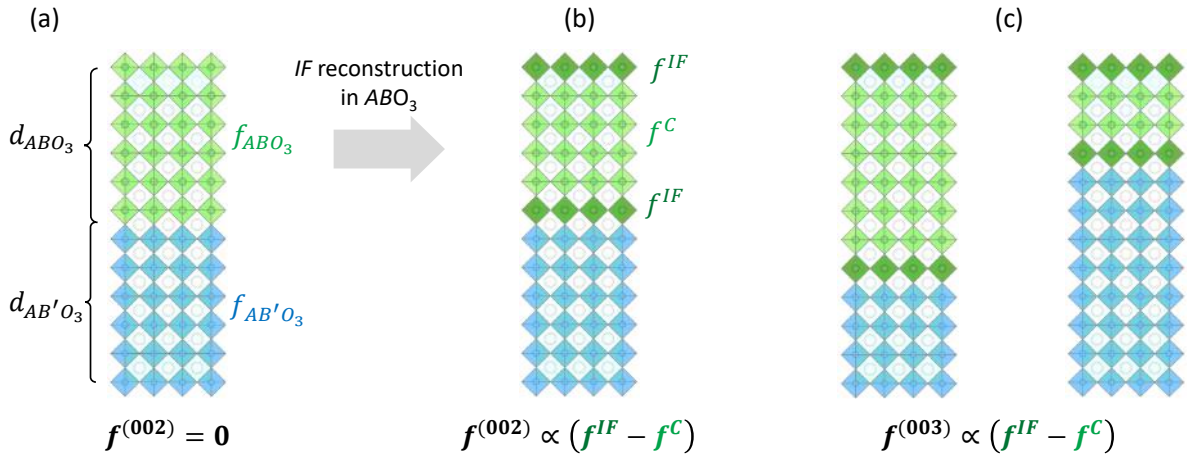


Fig. 12: *Scattering factors for different stacking of two materials ABO_3 (green) and $AB'O_3$ (blue) in a superlattice (the shown bilayer is repeated several times). (a) In case of a symmetric SL (d_{ABO_3}=$d_{AB'O_3}$) the (002) reflection of the SL structure vanishes. (b) When the ABO_3 interface layers reconstruct, resulting in different scattering factors in interface (f^{IF}) and central layers (f^{C}), the (002) reflection becomes allowed. (c) In the case of an asymmetric stacking sequence with ($2d_{ABO_3}$=$d_{AB'O_3}$ (left) or d_{ABO_3}=$2d_{AB'O_3}$ (right)), the (003) superlattice reflection is most sensitive to a difference of f^{IF} and f^{C}.*

dependent spectra, measured at fixed momentum transfer at (002) or (003) as a function of energy over the corresponding L edges of Ni or V. To ensure a unique fitting result, we only allow a redistribution of the dichroism in f^{IF} and f^{C}, so that the layer-weighted average, measured in XAS is fixed, i.e., $I_{\mathrm{XAS}} \propto n\,\mathrm{Im}(f^{IF}) + m\,\mathrm{Im}(f^{C})$ with n, m the number of unit cells of IF and C layers, respectively. The results of the simulations that best match the experimentally measured XRR then provide layer-specific linear dichroism spectra that can be quantitatively evaluated using the sum rules or cluster calculations to obtain orbital occupancies.
Since the t_{2g} orbital lobes point between the B-O bonds, while for e_g orbitals they point along the bonds, it is interesting to systematically compare reconstructions at YVO_3-$LaAlO_3$ and RNiO$_3$-$LaAlO_3$ interfaces. as I will discuss in the following. Our studies on $LaNiO_3$/$LaAlO_3$ superlattices showed that both, epitaxial strain and confinement effects at the interface lead to changes in the Ni-e_g orbital polarization depth profiles [50, 51]. In YVO_3-$LaAlO_3$ superlattices the interface effects produce an inverted orbital polarization in the layers next to $LaAlO_3$, compared to the central part of the YVO_3 layer stack [19].

4.2.1 Orbital polarization profiles in nickelate superlattices

The Ni^{3+} ion in RNiO$_3$ with R = rare-earth ion has nominally a $3d^7$ electron configuration and the octahedral crystal-field of the perovskite structure splits the atomic $3d$ orbital manifold into a lower-lying triply degenerate t_{2g} level that is fully occupied by six electrons, and a higher-lying doubly degenerate e_g level with a single electron. In bulk RNiO$_3$ the two Ni-e_g orbitals with $d_{x^2-y^2}$ and $d_{3z^3-r^2}$ symmetry are equally occupied. Model calculations have shown that the in-plane $d_{x^2-y^2}$ orbital occupation can be stabilized by epitaxial strain and confinement in

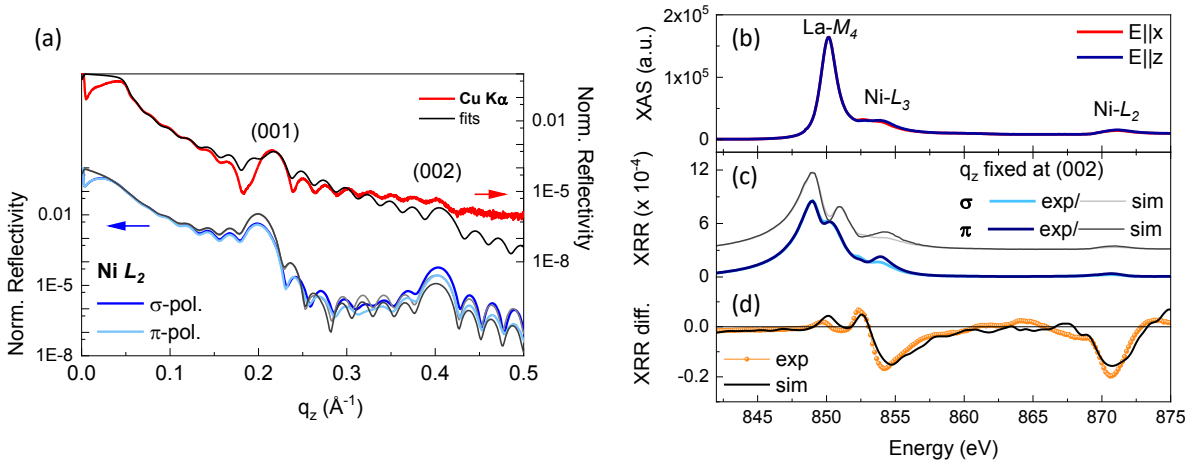


Fig. 13: *Experimental data and simulations of XRR data of a $LaNiO_3$-$DyScO_3$ (4/4) superlattice. (a) q_z-dependent reflectivity at fixed energy: non-resonant (Cu-K_α) and resonant to Ni-L_2. (b) Linear dichroism measured in XAS. (c,d) Energy-dependent linear-polarized reflectivity at (002) and the dichroic difference spectrum. Data reproduced from Ref. [50].*

a superlattice geometry, resulting in an electronic structure similar to that of the cuprate high-temperature superconductors [52–54]. In order to gain experimental insight on the relative effects of strain and confinement, we have grown superlattices with four-unit-cell-thick layers of metallic $LaNiO_3$ and layers of different band-insulating RXO_3 (R = La, Gd, Dy and X = Al, Ga, Sc) by pulsed-laser deposition on substrates that impose either compressive or tensile strain. Using such a symmetric superlattice geometry allows to determine depth-resolved orbital polarization profiles in a quantitative manner by exploiting the depth-dependence of reflectivity at momentum transfer q_z close to the (002) reflection. As introduced in section 3, the XRR analysis relies on optical constants and a structural model. For the analysis we used the software package ReMagX [55]. The structural parameters are obtained by fitting non-resonant, q_z-dependent hard X-ray reflectivity data (Fig. 13(a)), which are then fixed in the following analysis steps. To implement the energy-dependent fine structure across the relevant La-M and Ni-L absorption edges, we used the measured linear polarized XAS (Fig. 13(b)) to build the optical constants of $LaNiO_3$ in the way shown in Fig. 7.[4] Then we simulated the reflected intensity measured with fixed q_z as a function of energy (E) (Fig. 13 (c)) and its normalized dichroic difference spectrum (Fig. 13 (d)). To this end, we considered models with different tetragonal scattering tensors (Eq. 8) for f^{IF} in interface layers (B), and f^{C} in central layers (A) of the $LaNiO_3$ stacks, keeping the averaged value from XAS fixed. In this way, we were able to determine the redistribution of dichroism between the IF and C layers and the layer-dependent spectra [51]. Then we evaluated them using the sum rule for linear dichroism (Eq. 10), and defined the orbital polarization as

$$P = \left(\frac{4}{n_{e_g}} - 1\right)\frac{X-1}{X+1},$$

[4]For superlattices with the band insulator $LaAlO_3$ it is important to include the La-M resonances in their optical constants (see Fig. 7), as they are very close to the Ni-L_3 resonance.

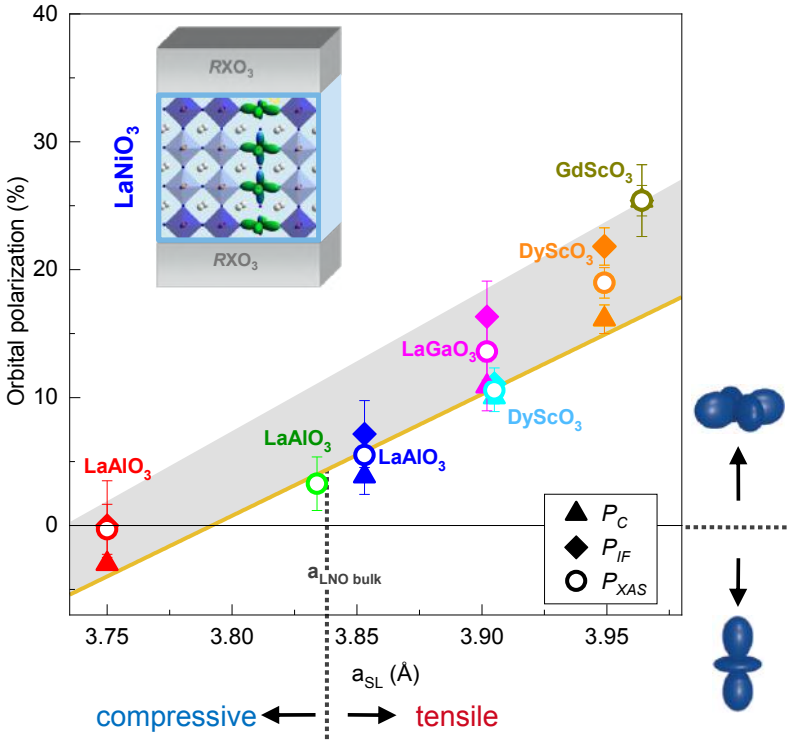


Fig. 14: *Layer-resolved orbital polarization, P_C and P_{IF}, as a function of in-plane lattice parameter a_{SL} for $LaNiO_3$-RXO_3 superlattices, grown on different strain-inducing substrates and with different composition of the buffer layers as indicated in the labels. The open symbols show the orbital polarization P_{XAS} obtained from the linear dichroism measured in XAS. Data reproduced from Refs. [50, 51].*

where $n_{e_g} = 4 - h_{e_g}$ is the sum of e_g electrons. The nickelates are negative charge-transfer insulators with a dominant $3d^8\underline{L}$ contribution in the ground state [56,49], where $\underline{L}$ denotes an oxygen ligand hole. Therefore the local, atomic Ni-e_g orbitals have rather $n_{e_g} \sim 2$. However, to compare superlattices with possibly different hybridization, i.e., possibly different n_{e_g}, we calculated orbital polarization with $n_{e_g} = 1$ for all different compositions. This means that the orbital polarizations we compare in Fig. 14 can be understood as those of the extended Wannier orbitals, which also have d-orbital symmetry. An illustration of the wave functions, obtained from DFT calculations, and further discussion can be found in Ref. [50]. The layer-resolved orbital polarizations P_C and P_{IF}, together with the layer-averaged values obtained from XAS,P_{XAS}, are shown in Fig. 14 as a function of their in-plane lattice parameters a_{SL}. The lattice parameters have been determined by X-ray diffraction for various $(4/4)$ $LaNiO_3$-RXO_3 superlattices grown on substrates with different lattice mismatch. We observed a tendency in all superlattices for the IF layers to have higher orbital polarization than the C layers with values up to 25%, which can be attributed to the confinement effect. However, it can be seen that the strain is the more effective control parameter (yellow line in Fig. 14), while the polarization attributed to the confinement effect from the band insulator layers falls into a comparatively narrow band of $\sim 5\%$ width (grey shaded area in Fig. 14).

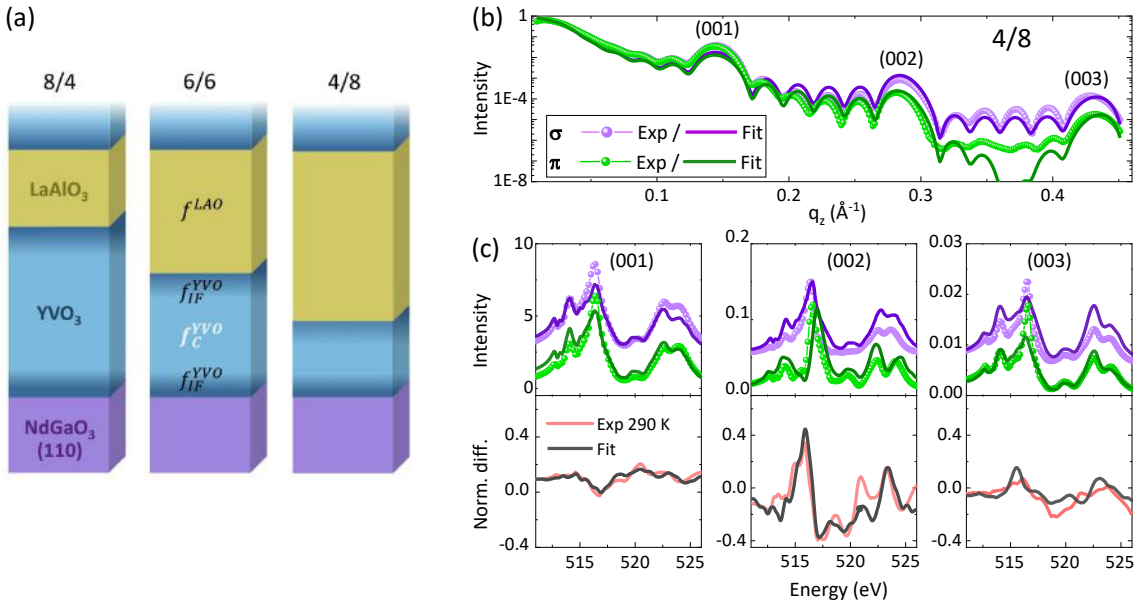


Fig. 15: *(a) YVO_3-$LaAlO_3$ superlattices with three different stacking sequences, with (8/4), (6/6) and (4/8) consecutive unit cells, were investigated to be as sensitive as possible to orbital reconstructions in interface layers (modelled with scattering factors f^{IF}) and central layers (f^C). (b) Representative q_z-dependent and (c) E-dependent scans at momenta fixed to the superlattice reflections $(00l)$ with $l = 1, 2, 3$ for the (4/8) superlattice at room temperature. Reproduced from Ref. [19].*

4.2.2 Vanadates - A t_{2g} system

To extend the methodology of orbital reflectometry to a t_{2g} system we have studied YVO_3-$LaAlO_3$ superlattices [19]. The compound YVO_3 is a strongly-correlated Mott-Hubbard insulator that shows no metal-insulator transition up to its melting point. The bulk crystallizes in an orthorhombic crystal structure (space group $Pbnm$ with lattice parameters a_o, b_o, and c_o) at room temperature with a V^{3+} electronic configurations shown in the left of Fig. 16(b). The low-temperature properties are governed by different orbital (OO) and spin (SO) ordered phases, which arise from competing crystal-field and superexchange interactions. Below 200 K C-type OO is observed, i.e., there is antiferro-orbital order in the orthorhombic ab_o plane, while along c_o there is ferro-orbital coupling. At 115 K the onset of corresponding G-type SO, and at 77 K a change to G-type OO and C-type SO phase was found [57]. In Ref. [19] we explored possible changes in the orbital occupations in YVO_3-$LaAlO_3$ superlattices. While the STEM(-EELS) images shown in Fig. 2 confirmed the high quality of the superlattice structure and the V^{3+} valence state, detailed X-ray diffraction characterization showed that $Pbnm$-type distortions are also present in the superlattice and that its structure follows the orientation of the substrate, i.e., the orthorhombic c_o axis lies in the interface planes. Accordingly, we rotate the coordinate system for the t_{2g} orbitals for better comparison with the bulk configuration (Fig. 16(b)). To obtain the depth-resolved information, we choose three superlattice structures, with $(YVO_3)_n$/$(LaAlO_3)_m$ bilayers with varying thicknesses of n= 4, 6, 8 and m= 8, 6, 4

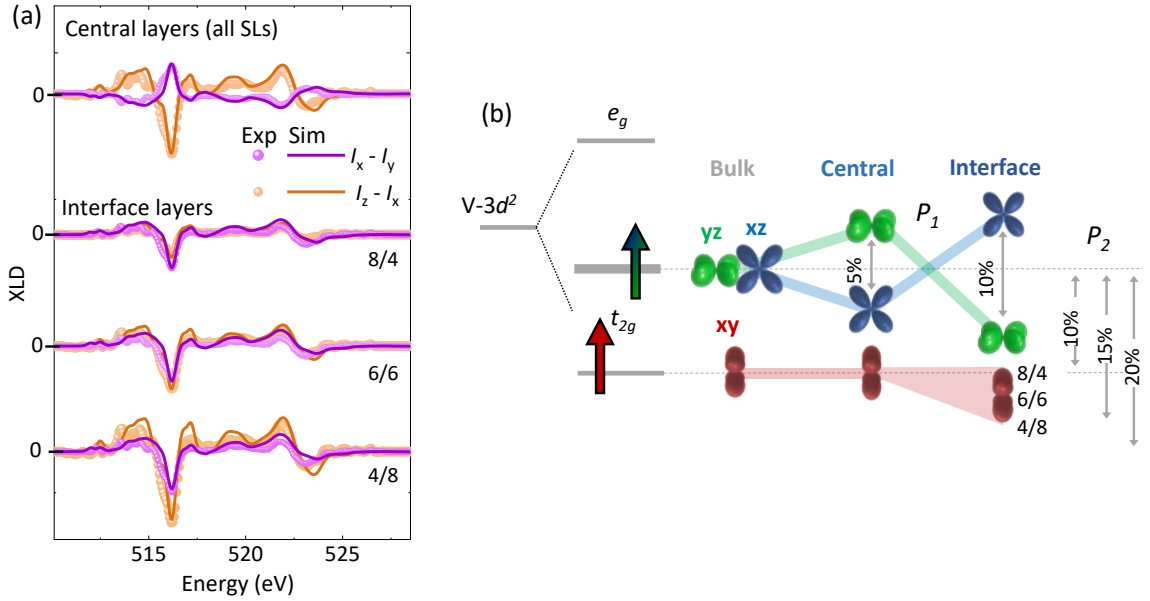


Fig. 16: *(a) Experimental (symbols) and fitted (lines) linear dichroism profiles between two polarization pairs for the central (C) layers of all three superlattices (top) and the interfacial (IF) layers of the* $(8/4)$*,* $(6/6)$*, and* $(4/8)$ *superlattices. (b) Schematic representation of the orbital polarization for the bulk, C and IF layers of the superlattices. Taken from [19].*

pseudo-cubic unit cells (Fig. 15(a)) to be maximally sensitive to interface reconstructions in the XRR measurements (Fig. 12). We simulated the linear dichroic reflectivity (q_z and E dependent, Fig. 15(b,c)) for different models of the heterostructure, again with f^{IF} and f^{C}, but now of orthorhombic symmetry, i.e., non-zero $F_{xx} \neq F_{yy} \neq F_{zz}$ in the scattering tensor (5). This is necessary because the t_{2g}, d_{xy}, d_{xz} and d_{yz} orbitals have pairwise lobes in the same spatial directions. In comparison, in the e_g system the $d_{x^2-y^2}$ orbital with z polarization is not accessible. By comparing the results, and then iteratively refining the model we obtained layer-resolved X-ray linear dichroism profiles (Fig. 16(a)) that were then compared with ligand-field cluster calculations to obtain the layer-dependent t_{2g}-orbital polarizations (P_1 and P_2 in Fig. 16(b)). As explained in section 3, the sum rules are not applicable to determine t_2g occupations. The results show that d_{xz} and d_{yz} orbital degeneracy is lifted in the superlattices unlike in bulk at room temperature, the d_{xz}-d_{yz} polarization is inverted between the C and IF layers, and the d_{xy} occupation in the IF layers depends on the number of YVO_3 layers (Fig. 16(b)). We also measured the temperature dependence of the spectra shown in Fig. 15(c) and found that the reconstructed orbital occupations are preserved down to 30 K [19].

4.3 Noncollinear magnetic order in nickel oxide heterostructures

As mentioned in the introduction, the accessible Ewald sphere in the soft X-ray region is often limiting for the study of perovskites with comparatively small lattice constants around 4 Å. For example, the ordering vector $q = (1/2\ 1/2\ 1/2)$ for the G-type antiferromagnetic order in YVO_3 is not accessible at energies near the V-L edge. The unusual magnetic order observed

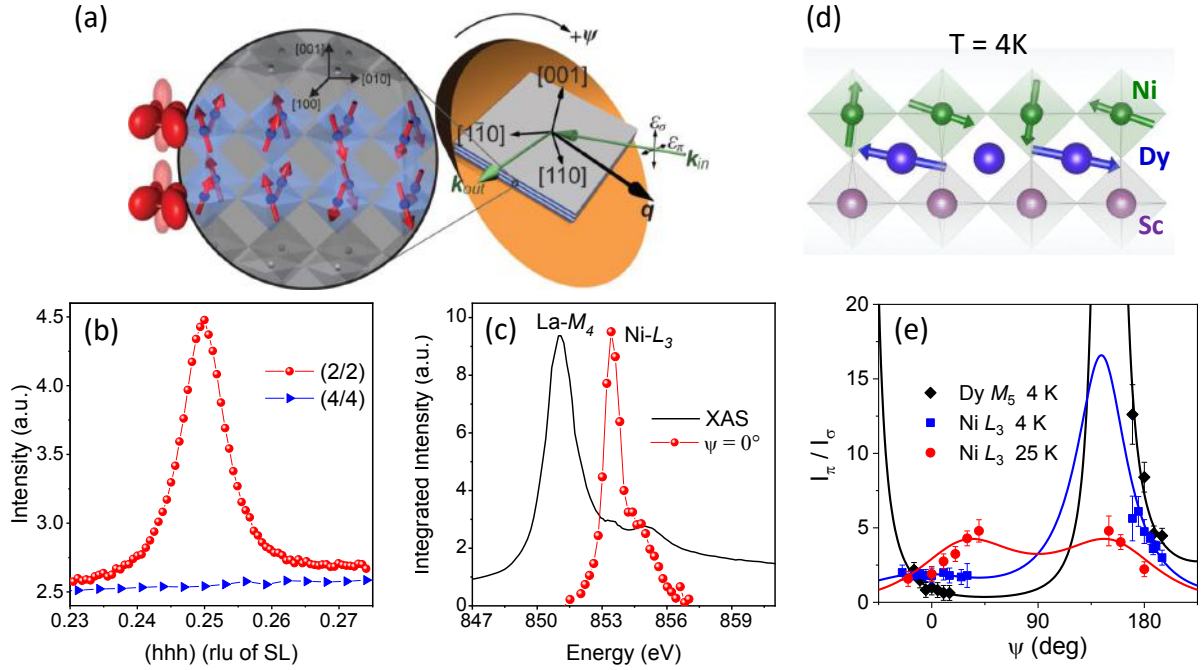


Fig. 17: *(a) Sketch of the magnetic order and REXS scattering geometry to study the magnetic order in $LaNiO_3$ superlattices. The sketch shows the wave vectors of the incoming and outgoing photons (light arrows), the corresponding momentum transfer q, the incoming photon polarization vectors ϵ_σ and ϵ_π, and the azimuthal angle ψ. (b) Scans around q^{mag} at $T = 10$ K and $E = 853.4$ eV for $LaNiO_3$-$LaAlO_3$ superlattices with $(2/2)$ and $(4/4)$ consecutive unit cells. (c) Comparison between XAS and energy dependence of the magnetic Bragg intensity around the Ni-L_3 edge at $\psi = 0°$. (d) Sketch of the Ni-Dy exchange interaction at the interface of a $LaNiO_3$-$DyScO_3$ superlattice derived from the azimuthal dependence (e) of the scattering at q^{mag}, measured resonant to the Dy-M_5 and Ni-L_3 edge at $T = 4$ K. The azimuthal dependence measured at Ni-L_3 at $T = 25$ K corresponds to that of $LaNiO_3$-$LaAlO_3$ $(2/2)$. Figures (a-c) and (d,e) were reproduced from data published in Ref. [58] and Ref. [20], respectively.*

in bulk rare-earth nickelates ($R \neq$ La) with an ordering vector $q^{\text{mag}} = (1/4\ 1/4\ 1/4)$ in cubic notation is a fortunate exception. It was first studied in $NdNiO_3$ by REXS at the Ni-L edge in Ref. [59]. We used REXS at the Ni-L edge to study the $LaNiO_3$-RXO_3 superlattices that I already introduced in the previous section. $LaNiO_3$ is the only bulk rare-earth nickelate that is paramagnetic and metallic down to the lowest temperatures [60]. We have shown that when this compound is grown epitaxially between RXO_3 layers in a superlattice, and the thickness is reduced to two unit cells, a magnetic order with q^{mag} is observed [58]. Figure 17(a) shows the scattering geometry used to access the magnetic Bragg peak at energies resonant to the Ni-L_3 edge. While a magnetic Bragg reflection is observed for the $(2/2)$ with two consecutive unit cells each of $LaNiO_3$ and $LaAlO_3$, this is not the case in $(4/4)$ superlattices with thicker layer stacks (Fig. 17(b)). The scattered intensity is strongly enhanced at the Ni-L_3 energy (Fig. 17(c)) and shows an azimuthal dependence (Fig. 17(e)) that is characteristic for a commensurate, non-collinear magnetic order with a $(\uparrow\rightarrow\downarrow\leftarrow\downarrow)$-type order of moments along the cubic perovskite $[111]$ direction (see sketch in Fig. 17(a)). In such an azimuthal scan, the Bragg condition is

preserved and the scattered intensity is measured, while the sample rotates around q_{mag}. During this, the projection of the polarization vectors ϵ_σ and ϵ_π onto the magnetization axis is changed. In the case of the noncollinear order in bulk nickelates, one therefore observes a π-periodic intensity modulation as a function of ψ [59]. For superlattices that are under biaxial strain from the substrate, the direction of the sublattice moments changes due to the changes in d-orbital occupation that controls the magneto-crystalline anisotropy via the spin-orbit coupling [58]. The precise direction of the sublattice magnetization can be determined by simulating the azimuthal dependence (solid lines in Fig. 17(e)). From this and the fact that we observed considerable conductivity in the magnetically ordered state, we conclude that a spin-density wave phase is stabilized in the epitaxial $LaNiO_3$ superlattice, which has no bulk analogue.
In a second REXS study we examined more closely the (2/2) $LaNiO_3$-$DyScO_3$ superlattice, which, as pointed out above, shows noncollinear q^{mag} order in the Ni spin system below 100 K [20]. Taking advantage of the element sensitivity of REXS by measuring the azimuthal dependence of the scattered intensity at q^{mag} and at energies resonant to Ni-L and Dy-M (Fig. 17(e)), we derived the following scenario. Upon cooling below 18 K, Dy-Ni exchange interactions at the $LaNiO_3$-$DyScO_3$ interfaces lead to a collinear magnetic ordering of the interface Dy moments (note the 2π-periodicity in intensity in the ψ scan) as well as a reorientation of the Ni spins in a direction dictated by the strong magnetocrystalline anisotropy of Dy (Fig. 17(d)). Such exchange interactions between local, paramagnetic rare-earth moments with the magnetic order of transition-metal ions is potentially interesting for manipulating spin structures in devices, as the large Dy moments provide anchoring points to external magnetic fields.

Acknowledgement

This lecture note is largely the result of the author's collaboration with many researchers listed in the references below. My special thanks go to Alex Frano, Meng Wu, Martin Bluschke, Friederike Wrobel and Padma Radhakrishnan, because without their excellent PhD work I would not have been able to show the nice examples in the last section. I would also like to thank Eberhard Goering, who taught me many experimental tricks and proofread these lecture notes.

References

[1] D.I. Khomskii: *Transition Metal Compounds* (Cambridge University Press, 2014)

[2] B. Keimer and J.E. Moore, Nat. Phys. **13**, 1045 (2017)

[3] B. Keimer, S.A. Kivelson, M.R. Norman, S. Uchida, and J. Zaanen, Nature **518**, 179 (2015)

[4] H.Y. Hwang, Y. Iwasa, M. Kawasaki, B. Keimer, N. Nagaosa, and Y. Tokura, Nat. Mater. **11**, 103 (2012)

[5] D.N. Basov, R.D. Averitt, and D. Hsieh, Nat. Mater. **16**, 1077 (2017)

[6] D.G. Schlom, L.-Q. Chen, C.-B. Eom, K.M. Rabe, S.K. Streiffer, and J.-M. Triscone, Annu. Rev. Mater. Res. **37**, 589 (2007)

[7] J. Mannhart and D.G. Schlom, Science **327**, 1607 (2010)

[8] H. Takagi and H.Y. Hwang, Science **327**, 1601 (2010)

[9] R. Ramesh and D.G. Schlom, Nat. Rev. Mater. **4**, 257 (2019)

[10] A.M. Glazer, Acta Crystallogr. A **31**, 756 (1975)

[11] J.M. Rondinelli and N.A. Spaldin, Adv. Mater. **23**, 3363 (2011)

[12] M. Bibes and A. Barthelemy, IEEE Trans. Electron Devices **54**, 1003 (2007)

[13] C. Weymann, C. Lichtensteiger, S. Fernandez-Peña, K. Cordero-Edwards, and P. Paruch, Appl. Surf. Sci. **516**, 146077 (2020)

[14] C.B. Eom, R.B. Van Dover, J.M. Phillips, D.J. Werder, J.H. Marshall, C.H. Chen, R.J. Cava, R.M. Fleming, and D.K. Fork, Appl. Phys. Lett. **63**, 2570 (1993)

[15] D.G. Schlom, L.-Q. Chen, X. Pan, A. Schmehl, and M.A. Zurbuchen, J. Am. Ceram. Soc. **91**, 2429 (2008)

[16] B. Jalan, R. Engel-Herbert, N.J. Wright, and S. Stemmer, J. Vac. Sci. Technol. A **27**, 461 (2009)

[17] J. Nordlander, M.A. Anderson, C.M. Brooks, M.E. Holtz, and J.A. Mundy, Appl. Phys. Rev. **9** (2022)

[18] P. Radhakrishnan, B. Geisler, K.Fürsich, D. Putzky, Y. Wang, G. Christiani, G. Logvenov, P. Wochner, P.A. van Aken, R. Pentcheva, and E. Benckiser, Phys. Rev. B **105**, 165117 (2022)

[19] P. Radhakrishnan, B. Geisler, K. Fürsich, D. Putzky, Y. Wang, S.E. Ilse, G. Christiani, G. Logvenov, P. Wochner, P.A. van Aken, E. Goering, R. Pentcheva, and E. Benckiser, Phys. Rev. B **104**, L121102 (2021)

[20] M. Bluschke, A. Frano, E. Schierle, M. Minola, M. Hepting, G. Christiani, G. Logvenov, E. Weschke, E. Benckiser, and B. Keimer, Phys. Rev. Lett. **118**, 207203 (2017)

[21] L.J.P. Ament, M. van Veenendaal, T.P. Devereaux, J.P. Hill, and J. van den Brink, Rev. Mod. Phys. **83**, 705 (2011)

[22] M. Blume in *Resonant Anomalous X-ray Scattering: Theory and Applications*, G. Materlik, J. Sparks, and K. Fischer (eds.) (North-Holland, Amsterdam, 1994)

[23] J. Fink, E. Schierle, E. Weschke, and J. Geck, Rep. Prog. Phys. **76**, 056502 (2013)

[24] J.P. Hannon, G.T. Trammell, M. Blume, and D. Gibbs, Phys. Rev. Lett. **61**, 1245 (1988)

[25] H. Wadati, A.J. Achkar, D.G. Hawthorn, T.Z. Regier, M.P. Singh, K.D. Truong, P. Fournier, G. Chen, T. Mizokawa, and G.A. Sawatzky, Appl. Phys. Lett. **100** 193906 (2012)

[26] R. Nakajima, J. Stöhr, and Y.U. Idzerda, Phys. Rev. B **59**, 6421 (1999)

[27] K. Zafar, P. Audehm, G. Schütz, E. Goering, M. Pathak, K. Chetry, P. LeClair, and A. Gupta, J. Electron Spectrosc. Relat. Phenom. **191**, 1 (2013)

[28] A. Thompson, D. Attwood, E. Gullikson, M. Howells, K.-J. Kim, J. Kirz, J. Kortright, I. Lindau, P. Pianetta, A. Robinson, J. Scofield, J. Underwood, D. Vaughan, G. Williams, and H. Winick: *X-ray Data Booklet* (Lawrence Berkeley National Laboratory, University of California, 2009)

[29] M. Haverkort (Ph.D. thesis, University of Cologne, 2005)

[30] B.T. Thole and G. van der Laan, Phys. Rev. B **38**, 3158 (1988)

[31] M.W. Haverkort, M. Zwierzycki, and O.K. Andersen, Phys. Rev. B **85**, 165113 (2012)

[32] M.D. Núñez Regueiro, M. Altarelli, and C.T. Chen, Phys. Rev. B **51**, 629 (1995)

[33] C.T. Chen, L.H. Tjeng, J. Kwo, H.L. Kao, P. Rudolf, F. Sette, and R.M. Fleming, Phys. Rev. Lett. **68**, 2543 (1992)

[34] C.T. Chen, Y.U. Idzerda, H.-J. Lin, N.V. Smith, G. Meigs, E. Chaban, G.H. Ho, E. Pellegrin, and F. Sette, Phys. Rev. Lett. **75**, 152 (1995)

[35] J. Stöhr, J. Magn. Mag. Mater. **200**, 470 (1999)

[36] D. Alders, L.H. Tjeng, F.C. Voogt, T. Hibma, G.A. Sawatzky, C.T. Chen, J. Vogel, M. Sacchi, and S. Iacobucci, Phys. Rev. B **57**, 11623 (1998)

[37] S. Macke and E. Goering, J. Phys.: Condens. Matter **26**, 363201 (2014)

[38] R.J. Green, R. Sutarto, F. He, M. Hepting, D.G. Hawthorn, and G.A. Sawatzky, Synchrotron Radiation News **33**, 20 (2020)

[39] B. Henke, E. Gullikson, and J. Davis, Atomic Data and Nuclear Data Tables **54**, 181 (1993)

[40] C.T. Chantler, J. Phys. Chem. Ref. Data **29**, 597 (2000)

[41] J. Als-Nielsen and D. McMorrow: *Elements of Modern X-ray Physics*, 2nd Ed. (John Wiley and Sons, Ltd, 2011)

[42] M. Blume, J. Appl. Phys. **57**, 3615 (1985)

[43] M.W. Haverkort, N. Hollmann, I.P. Krug, and A. Tanaka, Phys. Rev. B **82**, 094403 (2010)

[44] J.M. Tonnerre, M. De Santis, S. Grenier, H.C.N. Tolentino, V. Langlais, E. Bontempi, M. García-Fernández, and U. Staub, Phys. Rev. Lett. **100**, 157202 (2008)

[45] F. Wrobel, B. Geisler, Y. Wang, G. Christiani, G. Logvenov, M. Bluschke, E. Schierle, P.A. van Aken, B. Keimer, R. Pentcheva, and E. Benckiser, Phys. Rev. Mater. **2**, 035001 (2018)

[46] M. Grioni, J.B. Goedkoop, R. Schoorl, F.M.F. de Groot, J.C. Fuggle, F. Schäfers, E.E. Koch, G. Rossi, J.-M. Esteva, and R.C. Karnatak, Phys. Rev. B **39**, 1541 (1989)

[47] J.B. Goodenough, Phys. Rev. **100**, 564 (1955)

[48] J. Kanamori, J. Phys. Chem. Solids **10**, 87 (1959)

[49] G. Sawatzky and R. Green in E. Pavarini, E. Koch, J. van den Brink, G. Sawatzky (eds.): *Quantum Materials: Experiments and Theory*, Modeling and Simulation, Vol. 6 (Schriften des Forschungszentrums Jülich, Germany, 2016)

[50] M. Wu, E. Benckiser, M.W. Haverkort, A. Frano, Y. Lu, U. Nwankwo, S. Brück, P. Audehm, E. Goering, S. Macke, V. Hinkov, P. Wochner, G. Christiani, S. Heinze, G. Logvenov, H.-U. Habermeier, and B. Keimer, Phys. Rev. B **88**, 125124 (2013)

[51] E. Benckiser, M.W. Haverkort, S. Brück, E. Goering, S. Macke, A. Frañó, X. Yang, O.K. Andersen, G. Cristiani, H.-U. Habermeier, A.V. Boris, I. Zegkinoglou, P. Wochner, H.-J. Kim, V. Hinkov, and B. Keimer, Nat. Mater. **10**, 189 (2011)

[52] J. Chaloupka and G. Khaliullin, Phys. Rev. Lett. **100**, 016404 (2008)

[53] P. Hansmann, X. Yang, A. Toschi, G. Khaliullin, O.K. Andersen, and K. Held, Phys. Rev. Lett. **103**, 016401 (2009)

[54] M.J. Han, C.A. Marianetti, and A.J. Millis, Phys. Rev. B **82**, 134408 (2010)

[55] S. Macke, A. Radi, J.E. Hamann-Borrero, A. Verna, M. Bluschke, S. Brück, E. Goering, R. Sutarto, F. He, G. Cristiani, M. Wu, E. Benckiser, H.-U. Habermeier, G. Logvenov, N. Gauquelin, G.A. Botton, A.P. Kajdos, S. Stemmer, G.A. Sawatzky, M.W. Haverkort, B. Keimer, and V. Hinkov, Adv. Mater. **26**, 6554 (2014)

[56] R.J. Green, M.W. Haverkort, and G.A. Sawatzky, Phys. Rev. B **94**, 195127 (2016)

[57] S. Miyasaka, Y. Okimoto, M. Iwama, and Y. Tokura, Phys. Rev. B **68**, 100406 (2003)

[58] A. Frano, E. Schierle, M.W. Haverkort, Y. Lu, M. Wu, S. Blanco-Canosa, U. Nwankwo, A.V. Boris, P. Wochner, G. Cristiani, H.U. Habermeier, G. Logvenov, V. Hinkov, E. Benckiser, E. Weschke, and B. Keimer, Phys. Rev. Lett. **111**, 106804 (2013)

[59] V. Scagnoli, U. Staub, A.M. Mulders, M. Janousch, G.I. Meijer, G. Hammerl, J.M. Tonnerre, and N. Stojic, Phys. Rev. B **73**, 100409 (2006)

[60] J.B. Torrance, P. Lacorre, A.I. Nazzal, E.J. Ansaldo, and C. Niedermayer, Phys. Rev. B **45**, 8209 (1992)

9 Strong Correlations at Oxide Interfaces What is Hidden in a Plane View?

Jak Chakhalian
Rutgers University
136 Frelinghuysen Rd, Piscataway, NJ 08854

Contents

1 Primer on the physics of correlated oxides. 2

2 What are correlated oxide interfaces? 8

3 New quantum materials by geometrical lattice engineering 9

4 How can we grow perfect interfaces? 11

5 Examples of correlated oxide interfaces 12
5.1 Artificial ferroelectric metal 13
5.2 Orbital assisted Kondo lattice and spin-polarized 2D electron gas 15
5.3 New orbital order in graphene-like nickelates 19
5.4 Electronic structure of graphene-like nickelates 22
5.5 Artificial quantum spin liquid on lattices with extreme frustration 24

6 Problems to solve, ideas to try 28

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

A prime goal for this lecture is to provide you with a reasonably self-sufficient answer to the question of what interesting effects can happen if you join two dissimilar materials with correlated carriers to construct a 'sandwich' with the interface across those layers. Through the lens of physical phenomena, we will delve into the design ideas that lead to the creation of new synthetic quantum materials with properties primarily governed by the interface.
About the structure of the lecture: After introducing key concepts from the physics of correlated electrons, I switch to the guiding notions for building new synthetic materials with properties unattainable in bulk. Next, I briefly discuss the nucleation and growth of thin films based on the pulsed laser deposition method (PLD) or laser molecular beam epitaxy (MBE). After that, I illustrate those design ideas by several recent examples, ranging from a correlated polar metal to a quantum spin liquid. The lecture concludes with a list of ten currently *unsolved problems* that are worth further exploration.

1 Primer on the physics of correlated oxides.

Why transition metal oxides? Transition metal ions (TM) are commonly found in complex oxides, which make up the largest group of crystals on Earth. Besides oxygen, these compounds contain an element from the d-series in the periodic table, specifically $3d$, $4d$, or $5d$ TM oxides. In contemporary notation, many complex oxides with TM ions belong to the family of *quantum materials with correlated electrons* [1]. In general, it is the variation in the outermost d-shell configuration of these elements that gives rise to the great complexity in the crystal structures, electronic properties, and magnetic interactions in TMOs. There are many informative reviews on this topic, but as a one-stop source I recommend [2] as a comprehensive resource.
What crystal structures exist, and why are they formed? TMOs have numerous types of crystal structures, spanning all seven crystal systems! Although the most rigorous language of determining a crystal structure is to identify a conventional unit cell and the corresponding space group, going through all 230 space groups is certainly not the purpose of this lecture. Instead, as TMOs are predominately regarded as *ionic* crystals, the driving force of stabilizing a specific structure is the lattice energy: namely, for an ion located in a lattice, it experiences an overall electrostatic potential from the other ions (both cations and anions), which is the so-called Madelung potential V_I. The associated electrostatic energy of the ion is the product of its net charges with the Madelung potential. By summing over all ionic sites, we calculate the lattice energy. Broadly speaking, to determine the minimal value of the lattice energy we need to know the specific details of the crystal structure. However, since the Coulomb interaction between cation and anion is attractive, it is natural to assume that the lattice energy dramatically decreases as more anions surround a cation. On the other hand, as we place more and more anions near the cation, the repulsive interaction between anions *increases* the lattice energy. Moreover, if the anions are packed in such a way that the cation is rattling inside a void formed by the anions, this effect also increases the lattice energy. Collectively, to reach a balance among these competing effects and rationalize the ionic crystal structure, Linus Pauling proposed his famous five principles, also known as the 'Pauling rules' [3, 4].

Radius Ratio (r_C/r_A)	Coordination number	Type of geometry	
< 0.155	2	Linear	
0.155 - 0.225	3	Triangular	
0.225 - 0.414	4	Tetrahedral	
0.414 - 0.732	6	Octahedral	
0.732 - 1.000	8	Cubic	
1.000	12	Cuboctahedral	

Fig. 1: *Coordination number and type of polyhedral geometry determined by the cation-anion radius ratio in TMOs. The central cation is displayed with a small blue cycle, the oxygen ions with a big red cycle. The black rod between each cation and oxygen represents ionic bonding.*

Can we predict a crystal structure? According to the Pauling rules, the structure of a complex ionic crystal is mainly controlled by two factors: the local coordination number (CN) together with the polyhedral geometry of a cation and the network of the polyhedra spanning the crystal. First, Pauling's rule determines CN and its polyhedral geometry: this is the cation-anion radius ratio rule. Figure 1 summarizes the typical CN values and the corresponding polyhedrons.

Why is coordination so important? After you learned about the local polyhedra of cations, the next step is to understand how these polyhedra are interconnected. In real TMO solids, three common polyhedral networks can exist: *corner*-sharing polyhedra, *edge*-sharing polyhedra, and *face*-sharing polyhedra. An important principle pointed out by Pauling is that *sharing of edges and especially faces by two polyhedra cost more energy than sharing corners*. This is because in edge-sharing and face-sharing cases, the cations are located in closer proximity, increasing the electrostatic repulsion among them. In addition, for TMOs with multiple cations, those of high chemical valency and small coordination numbers tend not to share polyhedron elements, increasing their distance and thus reducing the repulsive interaction between them.

At this point, let me introduce two popular TM compounds to make the discussion more concrete. *Perovskites ABO_3.* The perovskite structure is relatively simple and common for compounds with the chemical formula ABO_3. Here we find two alternative combinations of A and B cations. If the A site is a *rare-earth* ion and the B site is a *transition metal* ion (e.g., $RENiO_3$ with RE = La to Yb), the charge state of each ion is $A^{3+}B^{3+}O_3^{2-}$. Alternatively, if the A site is an *alkaline-earth* ion (e.g., $ATiO_3$ with A = Mg to Ba) and the B site is a transition metal ion, the charge state is given by $A^{2+}B^{4+}O_3^{2-}$. No matter what combination, the A ion must be larger than the transition metal B ion, and it should be coordinated by *twelve* oxygens. At the same time, the B transition-metal ion is surrounded by *six* oxygens forming the octahedral coordination, and the network of corner-sharing B octahedra is the hallmark motif for perovskites (see Fig. 2c). Thus, a perovskite's ideal conventional unit cell is cubic with a B–O–B bond

angle of 90°. Of course, depending on the relative size of different ions, an actual unit cell can deviate from the cubic structure and it usually stabilizes in a lower-symmetry lattice. To predict if the structure deviates from the ideal cubic, in 1926, Goldschmidt introduced an index called *the tolerance factor* t_G, to quantify distortions in a perovskite crystal and predict the possible structure $t_G = (r_A{+}r_O)/\sqrt{2}(r_B{+}r_O)$, where r is an ionic radius.

Spinels AB_2O_4. Compared to perovskite, the spinel structure is more complicated (see Figs. 2c and 8b). The general chemical formula for the spinel structure is $A_{1-\delta}B_\delta[A_\delta B_{2-\delta}]O_4$. When δ is 0, it is known as the *normal spinel* AB_2O_4, in which all A cations are tetrahedrally coordinated, while all B cations are octahedrally coordinated. When δ equals to 1, the chemical formula is $B[AB]O_4$, and known as the *inverse* spinel. In this case, the tetrahedral sites are occupied by half of the B cations, while the other half of the B cations and all the A cations occupy the octahedral sites. Finally, when $0 \leq \delta \leq 1$, A and B cations mix up in both the tetrahedral and the octahedral sites. Unlike perovskite which assume many lattice structures and space groups, spinels usually stabilize into a *cubic structure* (with space group Fd3m). As for the charge state, assuming O^{2-} as is predominantly true in TMOs, there are two allowed ionic charge patterns for A and B ions: A^{2+}/B^{3+} (common in almost all cases) or A^{4+}/B^{4+} (rare but does exist, e.g., in $GeZn_2O_4$).

How do electrons behave inside TMOs? TMOs have a vast range of electronic behaviors, including those found in conventional metals and insulators, which are classified according to band theory. In addition, you can find various exotic phases such as high-temperature superconductivity, correlations-driven metal-insulator or Mott transitions, and topological states of quantum matter. The very diversity of TMOs makes them almost impossible to fit into a universal theory of their electronic properties. The usual approach in physics is to focus on the dominant term in the Hamiltonian while treating other terms as corrections or perturbations. Following this logic, in practice, the challenge of describing the electronic properties lies in selecting a starting point: whether the electrons are localized about corresponding ions or itinerant over the whole solid. Starting from those two extremes, several theoretical models have been developed, which I briefly introduced below (also see [5]).

Ionic Model. This simple but powerful model treats electrons from a local point of view. When we place a transition metal cation inside a solid, besides the Madelung potential, valence electrons of this cation experience additional Coulomb interactions stemming from surrounding oxygens, which we call the *crystal field*. Serving as a perturbation source, the degenerate energy levels obtained for isolated atoms are now split. Since the d sub-shell is the outmost shell of a transition metal ion, the crystal field can significantly affect its energy. As a result, the five spherical harmonics labeled by their quantum numbers (n, l, m) are no longer the eigenfunctions in the presence of a crystal field. Instead, we introduce new eigenfunctions that are linear combinations of those spherical harmonics.

These eigenfunctions' shapes (or electron density distributions) are plotted in Fig. 2a. Figure 2b displays a few distinct d orbital energy splitting patterns under different crystal field symmetry. Depending on the crystal field's local symmetry or the polyhedral coordination's geometry, the splitting sequences can be quite different. In this lecture, I will mainly discuss two types of

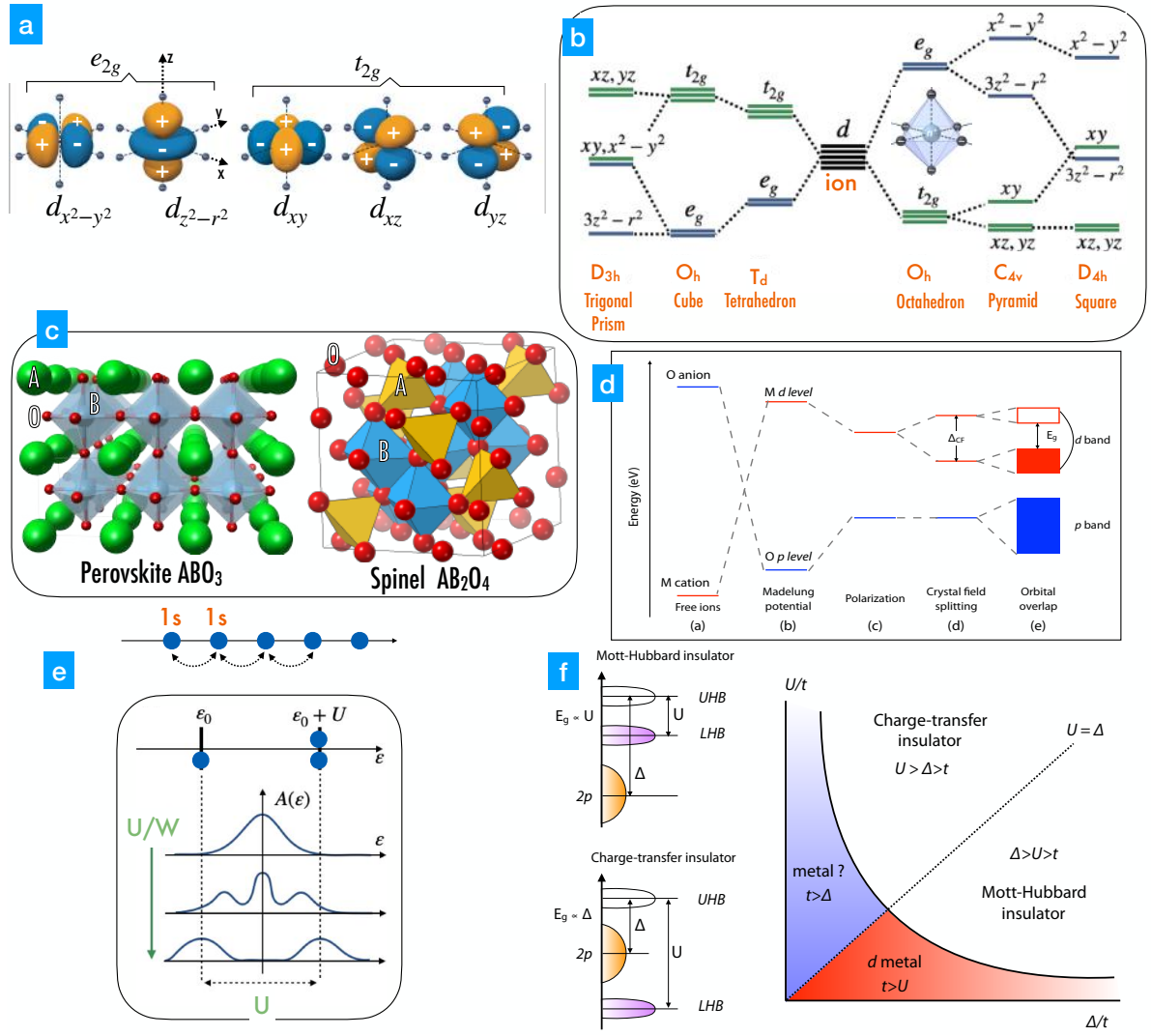


Fig. 2: *a: Shapes of the d orbitals. The notion of x, y, and z refer to the wave-function variables written in the Cartesian coordinate system. b: Energy splitting of the d orbitals under different crystal field symmetry. The notation for e_g, t_{2g} are defined according to group theory. c: Perovskite ABO_3 lattice and AB_2O_4 spinel unit cell. d: (a)–(e) Evolution of the electronic energy levels of TMOs in the ionic model. O and M refer to oxygen and transition metal ion, respectively. e: A schematic of the d band in the Hubbard model. For a half-filled band, the electron correlations (Hubbard U) are able to open a gap when its strength reaches the critical value, resulting in the lower (UHB) and upper Hubbard (LHB) bands separated by the correlated gap of U. f: (left top) Energy level diagram of a standard Mott-Hubbard insulator (MHI). In this case, the gap E_g is defined by U. (left bottom) Energy level diagram of the charge-transfer insulator (CTI). In this case, the gap E_g is defined by the charge-transfer Δ energy and intimately involves states on oxygen or anion in general. (right) Zaanen-Sawatzky-Allen phase diagram. Note, U and Δ represent the electron-electron correlation and charge-transfer energy, respectively. t is the electron hopping strength, a measure of the electronic bandwidth.*

crystal fields: octahedral and tetrahedral. In the octahedral field, the orbitals are split into an upper doublet e_g group including $d_{x^2-y^2}$ and $d_{3z^2-r^2}$ orbital states and a lower triplet t_{2g} group with d_{xy}, d_{yz}, and d_{xz} states.

How can we intuitively understand the origin of the energy splitting without relying on a group theory analysis? Recall that oxygen p orbitals are dumbbell-shaped and point along the Cartesian axes. In an octahedral environment, the lobes of e_g orbitals always point towards those of p orbitals, effectively increasing the Coulomb repulsions between e_g and p electrons and thus lifting up the energy levels. In contrast, t_{2g} orbitals have lobes pointing away from those of oxygen. Therefore, all t_{2g} levels are shifted downward. The energy gap between these two groups is denoted as Δ_{CF} (or 10Dq in chemistry). A similar analysis can be applied in the case of a tetrahedral environment, where the *splitting pattern is reversed* compared to that of octahedral coordination.

Hubbard model. The combination of the ionic model with the traditional band theory can successfully describe the electronic structure of many TMOs reasonably well. However, as early as 1937, it has been recognized that several transition metal compounds (e.g., CoO, NiO, Fe_2O_3), which are expected to be metallic, instead are wide-gap insulators. This failure in predicting the ground state in these materials signals that some critical factors are missing. The main reason is that in the ionic model, the electrons are considered independent and Coulomb repulsion is therefore omitted. This interaction, also known as electron correlation, is weak when electrons move in broad bands. However, in TMOs, the partially filled bands derived from d-electrons are usually very narrow, and the electrons appear more localized. Under this circumstance, the electron-electron correlations are inevitably amplified, exerting significant influence on the band structures and the overall physical properties of the materials. To quantitatively account for this observation, in 1963 Martin Gutzwiller, Junjiro Kanamori and John Hubbard independently proposed a new model Hamiltonian $\hat{H} = -t\sum_{\langle I,j\rangle,\sigma}(c^{\dagger}_{i,\sigma}c_{j,\sigma}+c^{\dagger}_{j,\sigma}c_{i,\sigma}) + U\sum_i n_{i,\uparrow}n_{i,\downarrow}$ [6–8]. In this Hubbard Hamiltonian, the first term describes the usual hopping effect of electrons from a site to its nearest-neighbors without spin-flip (so called kinetic term), whereas the second term accounts for the extra repulsive energy cost due to double occupation of the same lattice site. In this sense, the Hubbard model includes two competing processes (localization vs. delocalization), and the true ground state is determined by the relative strength between Hubbard U and hopping integral t, which is proportional to the electronic bandwidth W. The influence of Hubbard U on the electronic band structure is shown in Fig. 2e.

Mott insulators and Mott transitions. Now let us recap that in accordance with the original band theory, if a valence band with $2N$ capacity is half filled, the system is a metal. Nevertheless, now electron-electron correlations act to open a correlated or Coulomb gap in the valence band. If the correlation effect is weak compared to the bandwidth ($U/t \ll 1$), the band will not be split, and the material remains metallic. However, if the correlations are strong or $U/t \gg 1$, a correlated band-gap emerges and separates the $2N$ valence band into two new bands, called upper Hubbard band (UHB) and lower Hubbard band (LHB), each with N electron capacity [9]. For this reason, the material turned into an insulator, is collectively known as a Mott insulator or Mott-Hubbard insulator. I must stress that the Mott insulator is a highly non-trivial state, and many of the transition metal insulating compounds, which are predicted to be metallic, belong to this new class of quantum materials. It is also helpful to remember that for trivial band insulators or semiconductors, the energy gap is defined by the periodic potential of the crystal

lattice. In sharp contrast, in Mott crystals, the energy gap arises solely from electron-electron correlations.

Let me dig deeper into the excitation spectrum of Mott compounds. Naturally, starting from the noninteracting side ($U=0$) and gradually increasing U, we should sooner or later reach a critical point U_{cr} where a metal-to-insulator transition takes place. This transition is called the Mott transition. Among TMOs, a significant number of compounds undergo such a transition. It has been observed that several factors, such as temperature, pressure, and electric and magnetic field, can trigger the Mott transition. Another important fact is that even though the Hubbard Hamiltonian is not explicit about the underlying crystal structure, this Mott transition is usually accompanied by structural distortions and long-range spin orderings. Because of the entwined couplings, the question of a driving force behind the Mott transition, including the role of electronic correlations, is still not entirely understood.

Charge transfer insulators. Up to this point, I have considered only the effects of transition metal d bands. You should remember that anions and, specifically, oxygens are also very important. In fact, in many TMOs, the oxygen $2p$ bands are slightly lower in energy than the d bands. Here we can also ask, once the UHB and LHB are formed, what are their relative positions with respect to the oxygen p bands? As illustrated in Fig. 2f, there are mainly two cases of energy level diagrams expressing their relative positions. The oxygen p-band can either be lower than both of the Hubbard bands or in between these two bands. TMOs in the former case are a standard Mott insulator, whereas those in the latter case are given a new name, a change-transfer insulator (for an excellent discussion see [10, 11]).

To explain the difference, it is necessary to introduce a new energy scale, the so-called charge-transfer energy Δ_{CT}. As there are n electrons in the d levels, d^n configuration, two types of excitations exist. First, the electron can either hop onto another already occupied site in the same d level, say $d^n d^n \leftrightarrow d^{n-1}d^{n+1}$, or an electron from oxygen $2p$ band can hop onto the empty d-state, $d^n p^6 \leftrightarrow d^{n+1}p^5$. The first process costs us the energy U, while the second one cost a certain amount of charge-transfer energy, which is $\Delta_{CT} = E_d - E_p$.

The lowest charge excited state can be different depending on the ratio of U to Δ_{CT}. When $\Delta_{CT} \geq U$ [see Fig. 2f(left)], the oxygen p band lies lower than the LHB. The band gap is now determined by U. $d^n d^n \leftrightarrow d^{n-1}d^{n+1}$ costs less energy, as is typically the case for a conventional Mott insulator. However, when $\Delta_{CT} \leq U$, the electron hopping between oxygen and TM ion $d^n p^6 \leftrightarrow d^{n+1}p^5$ costs less energy and results in the lowest excited state. Here, the band gap is defined by the charge-transfer energy Δ_{CT} and not by U. The related materials are called charge-transfer insulators; in this case, the excited electrons are from the oxygen p levels. Moreover, it turns out that many of the physically interesting TMOs belong not to the Mott but to the charge-transfer family; the prototypical examples are the high-temperature superconducting cuprates, e.g., La_2CuO_4 or $YBa_2Cu_3O_{7-\delta}$.

Even more complicated situations occur if you consider the bandwidth, which is proportional to the electron hopping strength t. Specifically, what I considered so far was based on the assumption that $\Delta_{CT} \gg t$ and also $U \gg t$ so that one can expect the insulating ground state to emerge. However, you can imagine that if the bandwidth is large enough, the correlated

gap may fail to open, resulting in a metallic ground state. These ideas are rationalized in the so-called Zaanen-Sawatzky-Allen (ZSA) phase diagram, which is shown in Fig. 2f(right). For most transition metal compounds, their electronic structure can be qualitatively explained by this phase diagram. One significant effect is still missing from my discussion – the spin-orbit effect. The subject is so vast and important that instead, I refer the reader to the reviews [9, 12].

2 What are correlated oxide interfaces?

One of the prime goals of basic condensed matter physics is to seek out and explore new collective quantum states. Towards this goal, ultra-thin heterostructures composed of two or more structurally, chemically, and electronically dissimilar constituent oxides have been developed into a powerful approach over the past few decades [13–18].

Here, the main idea is that at the interface where the dissimilarities meet, the frustration caused by mismatches between the arrangement of atoms, charges, orbitals, or spins can trigger the emergence of phenomena with electronic and magnetic properties which non-trivially differ from the bulk compositions. For many research groups, the correlated interface engineering has opened a route to new materials behaviors using those mismatches as the control parameters. Paraphrasing the Nobel Prize winner Herbert Kroemer, 'The interface is a new material.'

A summary of the potential mismatches at oxide interfaces is shown in Fig. 3(left). As seen, at the oxide interfaces, the following degrees of freedom, can be rationally designed:

(1) *Epitaxial strain*. Strain results from lattice mismatches between the atomic arrangement of two different TMOs. By delicately applying strain, the M-O-M bond length and bond angle can be effectively tuned, which in turn may trigger electronic and magnetic phase transitions. For instance, using epitaxial strain as the control parameter, ultra-thin $NdNiO_3$ films have supported a remarkably enlarged phase diagram with several new states not observed in bulk [19];

(2) *Local symmetry*. As discussed in the previous section, TMOs have a variety of local symmetry or coordinate polyhedrons. Suppose we want two components with different local symmetries to grow as a heterostructure. To achieve growth, each subsystem needs to compromise, and the interfacial structure will deviate dramatically from their bulk counterparts. A robust example is the γ-Al_2O_3/$SrTiO_3$ interface: γ-Al_2O_3 has a spinel structure with tetrahedral and octahedral local symmetries, whereas $SrTiO_3$ is a perovskite with only octahedral symmetry. However, at the interface, an anomalous *square pyramid local symmetry* emerges for Ti ions, which strongly alters the electronic band structures;

(3) *Polar mismatch*. Even with the same lattice structure, the net charge of each atomic layer can be distinct from the naive ionic pictures. If a heterointerface consists of entities with different net charges per atomic plane, in this case the so-called 'polar mismatch' may occur, and the charges near the interface must rearrange to satisfy a condition of charge neutrality. This phenomenon has been found in many complex oxide heterostructures and for example was conjectured to be the source of the two-dimensional electron gases (2DEG) emerging between two *insulating* TMOs [14];

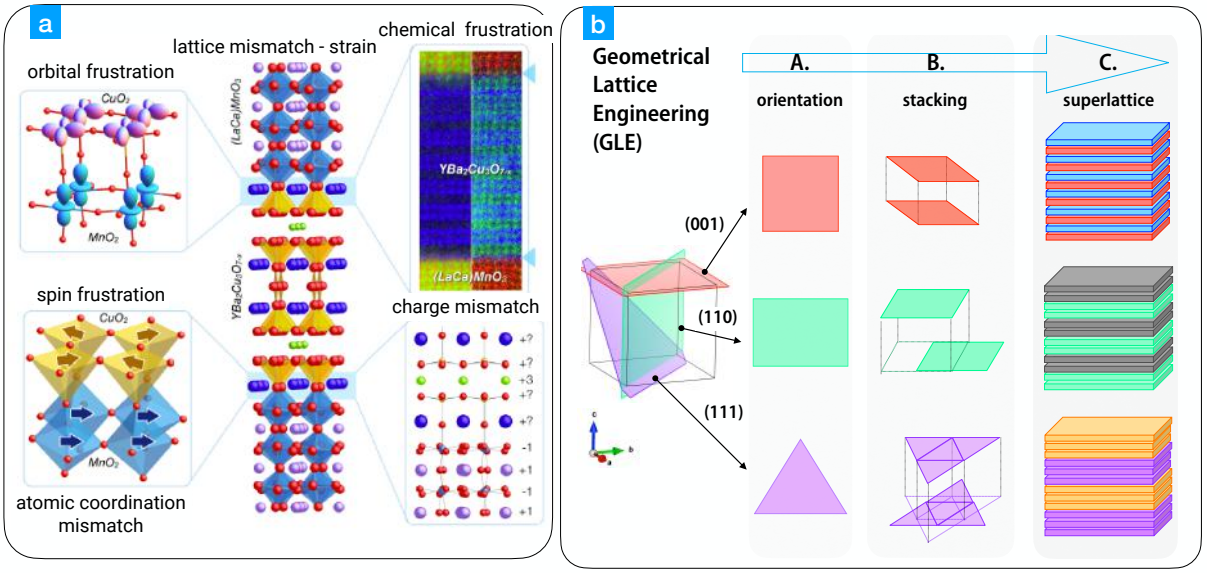


Fig. 3: *a: Mismatches at the correlated interface. b: Schematic illustration of the idea about geometrical lattice engineering, which has three control parameters named as orientation, stacking, and superlattice throughout its entire process.*

(4) *Orbital reconstruction*. Quite often, we find a variation of the orbital occupation at oxide interfaces that stems from the modulations of the atomic structure or charges due to the mismatches. As a result, unusual orbital configurations can be realized at the interface. To illustrate, in ultra-thin $LaNiO_3/LaAlO_3$ superlattices, it was found that unlike the bulk $LaNiO_3$ (Ni^{3+}, $3d^7$), where valence electron equally occupies the d_{x2-y2} and d_{z2} orbitals, at the superlattices interfaces d_{x2-y2} is preferentially occupied [20];
(5) *Magnetic coupling*. Oxide interfaces can be an effective tool to tune or even design magnetic interactions. You can imagine that if materials with a different kind of exchange coupling (for instance, FM vs. AFM) are attached in atomic proximity, their incompatible order parameters may eventually drive them to form a new magnetically balanced state (e.g., helical or canted spin arrangements). For example, an interesting phenomenon has been discovered at the interface of a high-T_c superconducting cuprate with a colossal magnetoresistance manganite, $YBa_2Cu_3O_7/La_{2/3}Ca_{1/3}MnO_3$, where surprisingly superconductivity and ferromagnetism coexist [21].

3 New quantum materials by geometrical lattice engineering

Inspired by the success of those interface engineering methods, recently another promising method, collectively known as "geometrical lattice engineering" (GLE), has been presented as a powerful tool to forge new topological and quantum many-body states. In close synergy with interface and strain engineering, where mismatches between layers induce unusual interactions, the key idea behind the GLE is to design fully epitaxial ultra-thin films and heterostructures with an *artificial lattice geometry generated by stacking of a precise number of atomic planes along a specific orientation* [22].

This concept can be illustrated by recognizing that the properties of a three-dimensional (3D) material can be drastically altered by changing parameters such as: the stacking of two-dimensional (2D) atomic planes, the specific arrangement of ions in those planes, their sequence, and the periodicity of layers fulfilling the charge neutrality condition. Conventionally, for thick bulk-like films, the effect of those variations is often negligible (maybe apart from anisotropy). In sharp contrast, for ultra-thin films, the directional stacking of atomic planes becomes dominant in defining the electronic and magnetic properties. Following this idea, many exciting materials systems have been theoretically proposed in pursuing exotic quantum states. At the same time, the experimental work on GLE has been primarily focused on the growth of cubic or pseudocubic (111)-oriented artificial lattices. In general, throughout the process of heteroepitaxial fabrication, to be able to design a new material by the GLE, you can follow three controllable routes. To explain this further, I will use a 3D simple cubic unit cell model to illustrate those control parameters (see Fig. 3b).

Growth orientation. Starting with the same bulk compound, its 3D crystal structure can be viewed as atomic layers stacking with different in-plane lattice geometries along different crystallographic directions. For example, as illustrated in Fig. 3 (right), while the (001) planes have square symmetry, the (110) and (111) planes have rectangular and triangular geometries, respectively. The required in-plane lattice geometry by design can be determined by selecting a proper structure and orientation of the substrate surface, which acts as a guiding template during the initial nucleation and growth stages. A typical example is the realization of a 2D magnetic lattice with extreme frustration derived from the ultra-thin (111)-oriented spinel-type structure AB_2O_4. This example I will describe in detail in Sec. 5.5.

Out-of-plane stacking sequence. In bulk crystals, the periodicity of the atomic planes can vary dramatically based on the choice of crystallographic direction to fulfill the requirement of translational symmetry and the relative atomic positions of neighboring lattice planes . For instance, the stacking of the adjacent layers can be either right on top of each other [the (001) stacking in Fig. 3], or shifted [the (110) stacking], or even entirely reversed [the (111) stacking]. This observation is at the heart of the design of artificial heterostructures since by controlling the number of stacking planes within that period you can forge unique quasi-2D lattices that do not exist in the naturally formed crystals. Among the prominent examples of GLE, I want to mention the generalized graphene lattice, which can be obtained by digitally tuning the number of atomic layers of (111)-oriented ABO_3 perovskite-type structures. I will present this case later in Sections 5.3 and 5.4.

Isostructural superlattices approach. Combining isostructural materials to establish superlattice structures with digital control over the individual number of layers adds another practical dimension to applying GLE. This approach can be very useful for achieving materials with complex chemical compositions or even thermodynamically unstable phases in the bulk form. A representative test case is the fabrication of (111)-oriented $1ABO_3/1AB'O_3$ superlattices [here "1" refers to a single cubic (or pseudocubic) unit cell] that gives rise to an A_2 $BB'O_6$ double perovskite [23].

4 How can we grow perfect interfaces?

In Section 1 we briefly discussed many theoretical concepts; now it is time to turn to something more applied. In this section, I want to focus on the question: 'How can we grow multi-layer structures with high-quality interfaces to match existing theoretical proposals?' Here, I describe one of the most popular methods for synthesizing such artificial complex oxide structures, called pulsed laser deposition or PLD. Despite its relatively young age, PLD has proven a versatile method for fabricating exceptionally high-quality epitaxial thin films and heterostructures during the last two decades (see [24] and the comprehensive references [25, 26]).

Compared with other popular physical vapor deposition (PDV) techniques, such as magnetron sputtering or all-oxide molecular beam epitaxy, several advantages make PLD particularly successful in growing complex oxide films. These include modestly priced instrumentation, stoichiometric transfer of ions from targets onto a substrate, an energetic forward-directed plume, and hyper-thermal interaction of the ablated species with the background gas (e.g., oxygen, nitrogen, argon). In other words, it is a PVD process performed in a high vacuum or low-pressure system using a pulsed laser as the heating source of ionic and molecular species.

What does a typical growth cycle look like? During the deposition, a pulsed laser with a pulse duration of $\sim$ 20 ns operating in the UV spectrum ($\lambda = 248$ nm) is focused on a small portion of a ceramic/polycrystalline target, which usually contains a stoichiometrically correct mixture of atoms to be synthesized on the substrate as the desired film. With a sufficiently high energy fluence of 1–2 J/cm^2, the ejected ions/molecules from the target vaporized by each laser pulse produce a directional plasma plume. Next, this highly forward-directed plume moves towards the substrate in a background gas atmosphere ranging from the ultra-high vacuum of 10^{-8} Torr and up to 1 Torr. This flux of oxidized and cooled to thermal energies ionic/molecular species rapidly propagate, reach the substrate, and eventually nucleate and crystallize into atomic layers of epitaxial solid films. To make a high-quality structure, you need to optimize several control parameters. Let me start with the targets. Since a complex oxide compound typically contains two or more kinds of atoms, a solid target suitable for PLD should be uniform and highly dense, possessing identical cation stoichiometry to the desired film. If the laser ablates a loose target, the resultant film will have a rough surface with microscopic molten droplets ejected from the target. The substrate crystal should closely match the lattice parameter and symmetry of the desired film. Finally, the crystallographic orientation of the substrate surface is critical as it determines the epitaxial orientation of the film; the substrate serves as the atomic template during the nucleation and the initial stage of film growth.

Once you select a substrate and the target, the remaining factors affecting the deposition process are laser energy and fluency, the distance between the target to the substrate, background gas pressure, and substrate temperature. As for the laser fluency, if we set it too low, the result of each laser pulse would be similar to thermal heating. In this case, the ejected flux of ionized species may deviate from the desired stoichiometry of the target due to the differences in vapor pressure among each constituent chemical element. To avoid this issue, we need to increase the laser fluence high enough to overcome the ablation threshold for a specific target, above which

the evaporation is independent of the vapor pressure, and the plume can maintain its proper stoichiometry. At the other extreme, running deposition with too high fluence would render the formation of macroscopic droplets or even damage the target. As a result, the typical laser fluence range is set around 1–4 J/cm^2. Next, to grow complex oxide films, molecular oxygen is often introduced into the chamber as the background gas for two reasons. First, the ejected low-mass molecular species needs to interact with O_2 to get the desired phase (remember, there is no solid oxygen!). Secondly, the background gas is required to reduce the high kinetic energies of the plume from several tens of eV down to meV; without a reduction of the kinetic energy, the complex ions of the plume would collide with the substrate's surface, potentially sputtering off the newly created island of crystalline phase and/or creating defects.

The substrate temperature is another critical factor in determining the quality of the films. However, temperature's role during the deposition process is rather complicated. On the one hand, a high substrate temperature is usually favored since it enhances the mobility of adatoms so that they can rearrange, forming a flat surface morphology. On the other hand, high temperatures can evaporate constituents with high vapor pressures out of the film, resulting in oxygen vacancy defects or missing cations. Another issue occurs when growing a superlattice structure composed of various oxide components. The thermally active atoms of each constituent layer can diffuse across the interface, destroying the atomic sharpness of the interface. In addition, for many complex oxides with low crystal symmetry, the epitaxial orientation of the film is very sensitive to the substrate temperature, often leaving a relatively narrow window for each phase.

Finally, under typical growth conditions, the deposition rate varies from a few tenths to one angstrom per laser pulse. This feature ensures precise control of growth on the sub-monolayer level and makes PLD a good choice for fabricating multilayers and superlattices of complex multi-element materials.

I only have discussed the most general trends for each essential control parameter. The specific values of those parameters are truly material- and growth chamber-dependent. This means that for every new material of interest and for each specific growth machine, finding and optimizing a comprehensive phase diagram for the best growth condition is necessary.

5 Examples of correlated oxide interfaces

In what follows there are some interesting examples of recent correlated oxide interfaces. The prime criterion for such a selection is to present the reader with the synthetic quantum materials that harbor emergent states or phases not seen in the bulk constituent layers. Also, unlike the previous sections, this section is more technical because it necessarily relies on the application of multiple state-of-the-art probes and advanced computational methods; as such, I suggest to treat those examples as a show-and-tell.

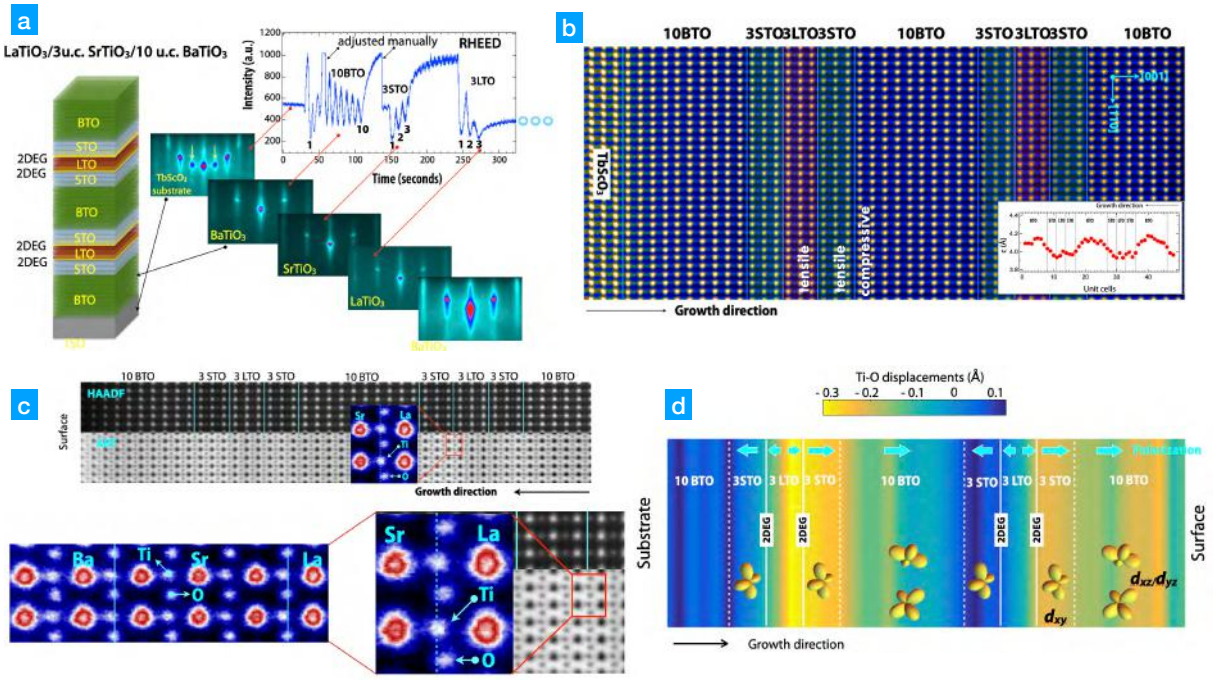


Fig. 4: *a: Diagram of a BTO/STO/LTO superlattice, where the yellow layer indicates the formation of a 2DEG. In-situ RHEED images confirm high crystallinity. b: STEM/EELS image reveals high-quality interfaces. Inset: extracted out-of-plane lattice parameters for each individual layer. c: High-resolution STEM imaging reveals significant Ti-O polar displacements, resulting in 8% enlargement of the lattice parameters. d: Summary of strength and direction of polarization.*

5.1 Artificial ferroelectric metal

Polar metals, commonly defined by the coexistence of polar crystal structure and metallicity, are thought to be scarce [27,28], because long-range electrostatic fields favoring the polar structure are expected to be fully screened by the conduction electrons of a metal. Generally, based on the type of atomic displacements, polar metals with perovskite structure fall into two main categories: A-site driven (e.g., positional shifts of Li, Nd, and Ca ions in $LiOsO_3$ [29], $NdNiO_3$ [30], and $CaTiO_{3-\delta}$ [28], respectively) or B-site driven (e.g., a shift of Ti ions in $BaTiO_{3-\delta}$ [31,32]) kinds. For the former category, recent theoretical work [28] has suggested the absence of a *fundamental incompatibility between the polarity and metallicity*, whereas, for the latter, polar displacements show a rapid decrease with increasing carrier concentration [32].

Moreover, in reducing from three to two dimensions, it remains an open question whether a polar metal can exist at all. Here, I describe the realization of a room temperature two-dimensional (2D) polar metal of the B-site type in *tri-color* (tri-layer) superlattices $BaTiO_3/SrTiO_3/LaTiO_3$. Such an explicitly non-centrosymmetric 2D metal provides a template to engineer an interesting quantum many-body state with three coexisting phases – ferroelectricity, ferromagnetism, and superconductivity.

Let us start with the design idea. First, we use tri-color rare-earth titanate heterostructures (see Fig. 4a) made of a layered arrangement of the ferroelectric alkaline-earth titanate $BaTiO_3$ (BTO, green), the paraelectric alkaline-earth titanate $SrTiO_3$ (STO, gray), and the Mott insulator rare-

earth titanate $LaTiO_3$ (LTO, red). Bulk $BaTi^{4+}O_3$ with $3d^0$ electron configuration is a well-known ferroelectric material, which can undergo complex structural and ferroelectric phase transitions on cooling, e.g., from cubic to tetragonal near 400 K, tetragonal to orthorhombic near 280 K, and orthorhombic to rhombohedral near 210 K [33] (ferroelectric properties are present in the latter three phases). Bulk $SrTi^{4+}O_3$ with $3d^0$ electron configuration is a band insulator with a charge gap size of $\sim$ 3.3 eV. In contrast, bulk $LaTi^{3+}O_3$ with $3d^1$ electron configuration is a Mott insulator and undergoes a G-type antiferromagnetic transition below 146 K [34]. In bulk, the lattice parameters are a = 3.905 Å for cubic $SrTiO_3$; 4.00 Å for cubic $BaTiO_3$; 3.958 Å for pseudocubic $TbScO_3$; and 3.956 Å for pseudocubic $LaTiO_3$. Based on these lattice parameters, the $SrTiO_3$ layers of BTO/STO/LTO superlattices grown on a TSO substrate are under tensile strain, whereas the $BaTiO_3$ layers are under compressive strain.

Another interesting feature of these designer superlattices is in the transfer of electrons from LTO ($3d^1$) into the STO ($3d^0$) layers leading to the formation of a two-dimensional electron gas (2DEG) [35] at the interfaces (yellow layer in Fig. 4a), which have a shared polar structure due to the presence of ferroelectric BTO. As clearly seen, this design has two inequivalent interfaces, BTO/STO and STO/LTO. What is interesting is the fact that both bi-color BTO/STO and BTO/LTO interfaces are insulating. Therefore, the metallicity in the 3-color BTO/STO/LTO structure comes from the 2DEG formed at the vicinity of the STO/LTO interface alone. Also, in contrast to itinerant electrons of the STO/LTO interface, the electrons at BTO/LTO interfaces are still localized, forming no 2DEG.

To monitor the crystal structures of the BTO/STO/LTO superlattice during growth, the measurements of in-situ reflection high-energy electron diffraction (RHEED) were performed. As seen in Fig.4a ultra-thin tri-color superlattices consisting of $(BTO)_{10}/(STO)_3/(LTO)_3$ (here the subscript refers to the number of unit cells) as well as reference samples of $(BTO)_{10}/(LTO)_3$ superlattice and BTO thin film were synthesized on $TbScO_3$ (110) single-crystal substrates by pulsed layer deposition in a layer-by-layer mode. During growth, the RHEED diffraction patterns for substrates, BTO, STO, and LTO layers stayed sharp, confirming the high crystallinity and epitaxy of BTO/STO/LTO superlattices.

Further, to determine the atomic-scale structures of the superlattices, their interfacial structure, and their chemical composition, the authors applied cross-sectional scanning transmission electron microscopy (STEM) with electron energy-loss spectroscopy (EELS). Figure 4b shows the tri-color superlattice's high-angle annular dark-field (HAADF) STEM image, revealing high-quality continuous and coherent interfaces without phase separation. In the Z-contrast HAADF image, the expected layer thickness and designed sequence of three layers are clearly distinguishable from the different intensities due to the difference in the scattering power of the layers. Additionally, as seen in the inset (red dots) in Fig. 4b, the periodicity of the growth sequence was determined from the periodic variation of the extracted out-of-plane lattice parameters for individual BTO, STO, and LTO layers. As engineered for the interfacial charge transfer, low-temperature electrical transport measurements of BTO/STO/LTO revealed the expected metallicity and large carrier density of conduction electrons, n_c in all tri-color samples ($n_c \sim 10^{14}$ cm^{-2}).

Next, I want to discuss the microscopic details of the centro-symmetric breaking of TiO_6 octahedra leading to the formation of a 2D polar metal in this structure. To address this, high-resolution HAADF- and ABF-STEM imaging were carried out, which allowed for the direct observation and extraction of precise atomic positions of all constituent atoms, including oxygen, across the interfaces. As shown in Fig. 4c, significant Ti-O polar displacements, i.e., relative shifts of Ti and O along the out-of-plane direction, are found in the BTO/STO/LTO tri-color structure. Additionally, a detailed quantitative analysis of the ABF-STEM image was performed to determine the amplitudes and directions of the *polar displacements*. The Ti-O polar displacements are found to be as large as 0.3 Å (!), which is an almost 8% enlargement of the lattice parameters. Moreover, these large Ti-O polar displacements not only exist in BTO but also extend deep into the STO and LTO layers, where the 2D metallic layer resides, thus inducing polar displacements into the metallic interface.

What about microscopic polarization and the connection to orbitals? Figure 4d summarizes the strength and direction of polarization labeled by the color map from blue to yellow. A striking feature is the periodic reversal of polar directions across the Mott LTO layers. It can be attributed to atomic displacements driven by local up-down symmetry breaking, typical of perovskite surfaces, at the STO/LTO interface [36]. More specifically, as seen in Fig. 4d, the authors find that around the LTO region, d_{xy} states are predominantly occupied. This orbital polarization decays exponentially. However, in the BTO region, d_{xz}/d_{yz} states are mainly occupied with the density shifted towards the left BTO/STO interface. The spatial separation of d_{xy} and d_{xz}/d_{yz} states is the combined effect of the electrostatic energy and the crystal field splitting [37]; namely, in the LTO region, the electrostatic potential from positively charged $(LaO)^{1+}$ layers dominates and is screened by d_{xy} electrons having in-plane dispersion. However, in the BTO region, the out-of-plane (or apical) Ti-O distances become substantially larger compared to the in-plane Ti-O distances due to the elongated c-lattice constant. This, in turn, lowers the on-site energy of d_{xz}/d_{yz} orbitals and results in the large increase in the d_{xz}/d_{yz} orbital occupancy compared to STO/LTO heterostructures.

5.2 Orbital assisted Kondo lattice and spin-polarized 2D electron gas

Magnetic interactions between the localized spins and conduction electrons are fundamental in many quantum many-body effects. Phenomenologically, in materials with localized spins coupled to conduction electrons, the Kondo interaction [38–40] competes with the magnetic Rudderman-Kittel-Katsuya-Yosida (RKKY) interaction [41], conceptualized in the so-called Doniach phase diagram [42] and Kondo lattice models [41]. In real transition metal crystals, however, the ground state depends on the strength of exchange interaction J, the electronic density ratio n_m/n_c of the localized magnetic moments n_m to conduction electrons n_c, and the orbital character of magnetically active electrons [43]. In the strong-coupling regime with large $|J|$, the Kondo interaction prevails and results in a Kondo singlet state [1], whereas on the weak-coupling side (small $|J|$), depending on the ratio n_m/n_c [44], the RKKY interaction may give rise to either a ferromagnetic (FM) or antiferromagnetic (AFM) order between the localized

spins. Notably, in the limiting case of $n_m/n_c \gg 1$, the localized spins tend to form ferromagnetic order by polarizing the conduction electrons [45]. In short, if we devise such a Kondo active structure with the specific set of control parameters described above, we should realize a highly desired artificial quantum material for spintronics with spin-polarized 2D metallicity.

In correlated d-electron heterointerfaces, the density ratio of n_m/n_c, the dimensionality, and the orbital polarization of the magnetic interactions are all vital components for the formation of a ground state [2]. Considering the splitting of the Ti t_{2g} band between d_{xy} and d_{xz}/d_{yz} subbands is a prime cause for the interesting emergent phenomena in the $SrTiO_3$-based heterostructures [46]. This raises an important question: What is an experimental phase diagram for the emerging magnetic interactions at the nanoscale in the tri-color system?

To answer this question, a set of tri-layer (tri-color) superlattices composed of [3 u.c. $LaTiO_3$/ $n$$SrTiO_3$/3 u.c. $YTiO_3$] ($n = 2, 3, 6$ unit cells or u.c.) and reference bi-layer samples [$m$$LaTiO_3$/ $n$$SrTiO_3$] (here $m = 3, 20$ u.c. and $n = 2, 3, 6$ u.c.) and [3 u.c. YTiO3/$n$$SrTiO_3$] ($n = 2$ and 6 u.c.) were epitaxially synthesized by PLD on $TbScO_3$ (110) substrates. Again, in this designer system, the interfacial charge transfer is used to create a two-dimensional electron gas at the interface between $LaTiO_3$ and $SrTiO_3$ (LTO/STO) connected to a spatially separated interface with localized magnetic moments at the $YTiO_3$/$SrTiO_3$ interface (YTO/STO) (see Fig. 5b).

Before I discuss the ground state of such a tri-color structure, let us investigate each component of the superlattice. Figure 5a shows the magnetic phase diagram of $R$$TiO_3$ (R=La, Pr, Nd, Sm), which exhibits an antiferromagnetic (AFM)-to-ferromagnetic (FM) phase transition across the series of rare-earth titanates [47]. For our tri-color design, we select $LaTiO_3$ ($\sim$0.2 eV gap, Ti $3d^1$) with the smallest distortion, which undergoes a G-type AFM phase transition below 146 K [34]. $SrTiO_3$ is a $3d^0$ system with no magnetism. And finally, for the third layer, the authors selected a significantly distorted $YTiO_3$ ($\sim$1.2 eV gap, Ti $3d^1$), where the FM order forms below 30 K [48].

To summarize the design idea: In this structure, the action is supposed to happen at the LTO/STO and STO/YTO interfaces. The STO layer acts as an active spacer which depending on its thickness, brings the 2DEG from LTO/STO closer or further away from the magnetic moments of the STO/YTO interface. That is why I name this unique tri-color design the structure to explore *interacting in space order parameters*.

Now, what about the tri-color system? Figure 5b shows a high-angle annular dark-field (HAADF) STEM image of the tricolor superlattice, revealing high-quality coherent interfaces without phase separation. The atomic positions of the elements La (large yellow dots), Sr (large white dots), Y (large blue dots), and Ti (small green dots) are labeled schematically. To check for the presence of the interfacial charge transfer across the two LTO/STO and STO/YTO interfaces, the layer-resolved electronic structure of 3LTO/6STO/3YTO was investigated by atomic scale STEM-EELS line scanning across the interfaces. As shown in Fig. 5b, by scanning atomic layer-resolved Ti $L_{2,3}$-edge spectra across LTO/STO and STO/YTO interfaces [along the white-dashed line in Fig. 5b] with a high energy resolution of 0.4 eV and a high spatial resolution of 0.8 Å, the evolution of the Ti electronic structure across the interfaces was mapped out. The bottom panel of Fig. 5b shows the reference spectra for Ti^{4+} and Ti^{3+} acquired from bulk-like

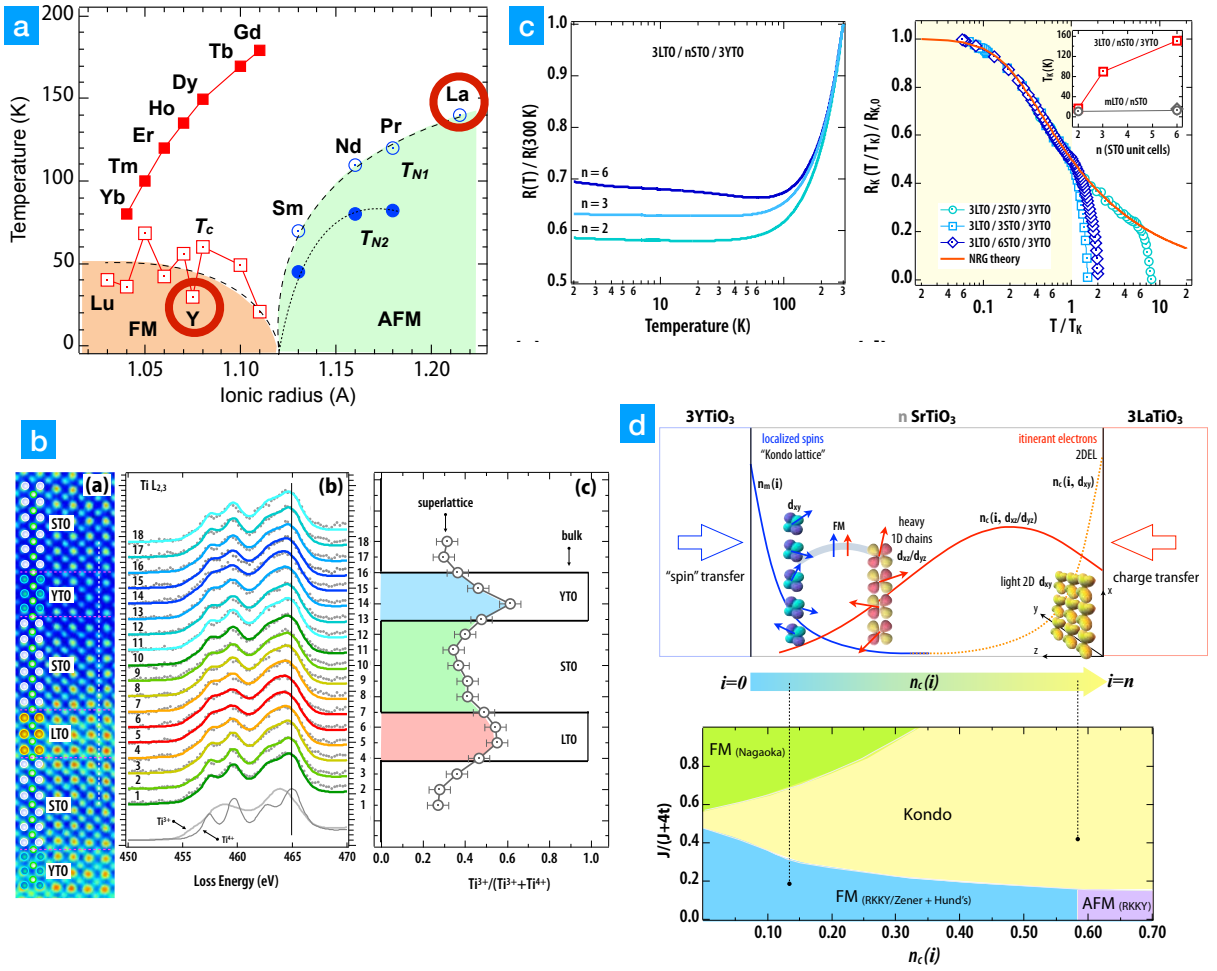


Fig. 5: *a: Magnetic phase diagram of $RTiO_3$. b: HAADF-STEM image of the tricolor superlattice reveals high-quality interfaces. STEM-EELS shows the layer-resolved chemical composition. c: All three samples display metallic behavior which agrees with theoretical fits. d: A view of the electron density distribution across the STO layer as dependent on layer thickness n and atomic plane i.*

$SrTiO_3$ and $LaTiO_3$ films. As seen in Fig. 5c, the Ti^{3+} spectra weight estimated from the fitting parameters of solid curves in Fig. 5b immediately reveal that in addition to the previously reported charge transfer from LTO into STO [49], there is an unexpected large charge transfer from YTO into STO which leads to a *localized electron layer at the YTO/STO interface* [50].

Next, the authors investigated the properties of interfacial electrons arising from the interfacial charge transfer by measuring temperature-dependent electrical transport. As clearly seen in Fig. 5c all three samples 3LTO/nSTO/3YTO (n=2,3,6) show characteristic metallic behavior with a weak upturn at a lower temperature. To rule out possible contributions from cation defects and oxygen vacancies, bilayer YTO/STO and LTO/STO samples were synthesized, and their transport properties were used as references [50, 51]. In sharp contrast to the highly insulating YTO/STO the sheet resistances of all the LTO/STO samples [51] show a *2D electron gas behavior*.

The other interesting feature seen in transport is the pronounced upturn in the sheet resistance at the lower temperature. Previous work on rare-earth titanate heterojunctions has attributed such an upturn to the Kondo lattice effect after carefully ruling out the contributions from weak localization [52] and electron-electron interactions [53]. One of the key features of the Kondo effect that immediately differentiates it from weak localization and electron-electron interactions is the *universal scaling behavior*. As shown in Fig. 5c, all tri-layer samples agree well with theoretical fits [solid red line, numerical renormalization group (NRG)] scaled Kondo resistances [$R_K(T/T_K)/R_{K,0}$]. The inset in Fig. 5c shows the extracted n-unit-cell dependent Kondo temperature T_K by fitting the experimental data of [mLTO/nSTO] and [3 u.c. LTO/nSTO/3 u.c. YTO] superlattice. Based on the fact that the YTO/STO interface is highly insulating [50], the observed Kondo scaling behavior lends support to the formation of the interfacial lattice of localized magnetic moments located at the YTO/STO interface of LTO/STO/YTO. So overall, in the tri-color LTO/STO/YTO structures, the authors created metallic carriers at the LTO/STO interface facing the lattice of magnetic moments formed at the YTO/STO interface.

To better understand the magnetic interactions in the tri-color LTO/STO/YTO, we can look at the STO layer thickness (n-dependent) and atomic plane (i-dependent) electronic density distribution plus d-orbital occupancy across the STO layer. In other words the question is what happens when we move the 2DEG closer towards the magnetic lattice. The conceptual picture is given in the top panel of Fig. 5d.

First, because we deal with a two-dimensional electron gas, the carriers have very specific orbital types to let 'light' electrons rapidly move along the interface in the x-y-plane. In other words, for a thicker STO layer, we got $n_c(d_{xy}) \gg n_m$ near the LTO/STO interface, resulting in the formation of a Kondo singlet state (fully screened magnetic moments) since light d_{xy} conduction electrons (dashed yellow line) with large carrier density are bound to the LTO/STO interface.

On the other hand, there is a low concentration of 'heavy' electrons (red arrows) with d_{xz}/d_{yz} character (here z is perpendicular to the interface), which slowly disperse away from the LTO/STO interface and appear near the magnetic STO/YTO interface (solid red line). Upon reaching the STO/YTO interface, these mobile heavy electrons interact with the localized magnetic moments $n_m(d_{xy})$. What is remarkable is that this *orbital-dependent ferromagnetic interaction* can proceed via two possible channels: (1) based on the Hund's rule, the interaction between the d_{xy} and d_{xz}/d_{yz} electrons results in the FM ground state [20] and (2) the Zener kinetic exchange, which can win the competition against the Kondo and RKKY interactions, again leading to the formation of a *localized ferromagnetic ground state with spin-polarized conduction electrons*. Finally, when the STO layer becomes ultra-thin, e.g., n=2, we have $n_c \sim n_m$ and the conduction carriers lose their distinct orbital character resulting in the mixed orbital state d_{xz}/d_{yz}/d_{xy}. In this case, in the ground state we have a direct competition between the Kondo screening, RKKY coupling, and Hund's energy. Based on this picture, the control of the STO thickness n indeed modulates the critical ratio between magnetic sites and mobile carriers n_m/n_c and their orbital character or orbital polarization to exert decisive control over the magnetic interactions.

5.3 New orbital order in graphene-like nickelates

In bulk, perovskite oxides have many exciting properties, including metal-insulator transitions, magnetism, superconductivity, charge and orbital orderings, and multi-ferroicity, to name a few. These infinite layer ABO_3 perovskite compounds consist of alternating AO/BO_2, ABO/O_2, and AO_3/B atomic planes along the pseudo-cubic $[001]$, $[110]$, and $[111]$ directions, respectively. Thus, the precise control during the growth of two or three pseudo-cubic (pc) unit cells of ABO_3 along the (111) direction leads to new lattice geometries with vertically shifted triangular planes of B sites and results in buckled honeycomb lattice as shown in Fig. 6a. The emergence of striking topological phases, including a quantum anomalous Hall state, was initially predicted for a honeycomb lattice by Haldane [54]. Recently, there was a spark of interest in the search for an *artificially stabilized graphene-like quasi-2D lattice* that can provide an ideal playground for interacting topological phases in complex oxides (for details, see [55]). Here I describe the case of rare-earth nickelates that illustrates the opportunities for designer topological phases by geometrical lattice engineering (GLE).

The first member of the rare earth nickelates series, $LaNiO_3$ (LN), is a paramagnetic metal. The other members of the family of nickelates ($RENiO_3$, *RE*=Pr, Nd,..., Lu, and Y) in bulk form exhibit metal-insulator transitions, E'-type antiferromagnetic ordering, charge ordering, and structural transitions with a strong dependence of the transition temperature on the size of the *RE* ion. Several theoretical works (see Refs. [57–60] for review) further emphasized the possibility of realizing interaction-driven topological phases without spin-orbit ions (e.g., Dirac half semimetal phase, anomalous quantum Hall insulator phase, or ferromagnetic nematic phase) in the weakly correlated limit on the buckled honeycomb lattice of $RENiO_3$ as shown in Fig. 6b-c. Moreover, in sharp contrast to bulk $LaNiO_3$, where orbital ordering is absent, theoretical modeling in the strongly correlated limit predicted the presence of an *orbitally ordered magnetic phase as the novel ground state* for the buckled honeycomb lattice of $LaNiO_3$.

How hard is it to grow (111) oriented films? Despite the conceptual simplicity, the growth of perovskites along the $[111]$-direction presents a formidable challenge. Contrary to the conventional $[001]$-direction, the epitaxial stabilization along the $[111]$ direction is far less understood due to unavoidable surface/chemical reconstruction effects. This can be seen as all perovskite substrates are strongly polar along this direction, e.g., $SrTiO_3$: $[SrO_3]^{4-}$, Ti^{4+}; $LaAlO_3$: $[LaO_3]^{3-}$, Al^{3+}, and so on. A possible solution to this polar catastrophe problem is to grow a thin metallic buffer layer at the beginning of the growth sequence to effectively screen the charge dipoles. However, one should pay particular attention as unwanted interfacial effects (between buffer layer and desired material) can significantly influence the buffered heterostructure.

As the reconstruction effect often appears only at the substrate and vacuum interfaces, choosing a substrate with the same sequence of charges per atomic plane as the desired material is another solution that does not require growing a metallic buffer layer. To investigate this, Middey *et al.* have grown $LaNiO_3$ films on (111) $SrTiO_3$ (with a polar jump at the film/substrate interface) and compared it to the growth of $LaNiO_3$ on (111) oriented $LaAlO_3$ (without any polar jump at the film/substrate interface) [56]. It was clearly demonstrated that while a thick bulk-like $LaNiO_3$

film is metallic and effectively screens charge dipoles, a thinner film becomes insulating. Using X-ray diffraction (XRD) and X-ray absorption spectroscopy (XAS), the authors uncovered massive amounts of oxygen vacancies within the thinner films. With the increased LNO thickness, the increased metallicity screens the polar jump, and the relative amount of proper Ni^{3+} ions rapidly increases. Finally, good stoichiometric $LaNiO_3$ along $[111]$ can be only obtained when the film thickness reaches 15 unit cells. In sharp contrast, Ni^{3+} was stabilized from the very initial stage of growth of LNO on the LAO (111) substrate and thus confirmed that the *absence of a polar jump at the film-substrate interface* is critically important for the epitaxial stabilization along $[111]$. As a result, the desired generalized *graphene-like crystal of RENiO_3* with *RE*=La to Nd has been successfully achieved for the first time on the LAO (111) substrate [60].

What about the ground state? According to theoretical calculations, the QAH topological phase should be accompanied by spontaneous ferromagnetism. However, X-ray resonant magnetic scattering measurements on (111) oriented [2u.c. $NdNiO_3$/4u.c. $LaAlO_3$] (2NNO/4LAO after that) superlattice ruled out the possibility of a long-range ferromagnetic ground state and instead established the presence of *antiferromagnetic* correlations (see Fig. 6d). In addition to magnetism, the orbital structure was investigated by the X-ray linear dichroism (XLD) technique. The XLD spectroscopy allows for uncovering different kinds of orbital ordering and the symmetry of a specific orbital state per chemical element of the film. In this resonant X-ray spectroscopy, one measures the difference in absorption with vertical polarization I_H vs. horizontal one I_V. If orbitals are preferentially aligned along one of the X-ray polarizations, the XLD signal becomes strongly enhanced. Conversely, for the orbitally disordered state, the X-ray linear dichroism is zero.

Despite its conceptual simplicity, the geometrical arrangement between the sample and X-ray polarization vector requires careful consideration for detecting orbital ordering. Specifically, as illustrated in Fig. 6e, all the e_g orbitals ($3d_{z2-r2}$, $3d_{x2-r2}$, and $3d_{y2-r2}$) are oriented at ϕ =54.7^o with respect to the $[111]$ growth axis of a NiO_6 octahedron. Because of this specific geometrical arrangement, the XLD signal is expected to be very small even for a ferro-orbital ordered (FOO) state with 100% orbital polarization.

To maximize the XLD signal, the samples can be mounted on a copper wedge (Fig. 6f), which 'mechanically' reorients the Ni-O bonds along the vertical polarization V and the in-plane horizontal polarization H. This specific sample orientation on the wedge aligns the $3d_{z2-r2}$ orbitals almost along V polarization, giving a finite dichroic signal. On the other hand, $3d_{x2-r2}$ orbitals are almost in the plane of H polarization with a small but finite angle with respect to the polarization vector H, resulting in an opposite and strongly reduced XLD signal. As a result, instead of perfect cancellation of linear dichroism, a finite XLD is expected to be present for the antiferro-orbital ordered (AFO) state. Fig. 6e-f shows resonant Ni L-edge X-ray absorption spectra (XAS) (I_V and I_H) and the XLD spectra (difference between I_V and I_H) obtained in the flat (ϕ = 0°) and the wedge (ϕ = 45°) geometries.

As anticipated from the discussion above, the XLD signal indeed strongly increased when the measurement was conducted for the ϕ = 45° geometry. As a control experiment, for the thick bulk-like (111) NNO film, the XLD measurement shows no significant orbital polarization.

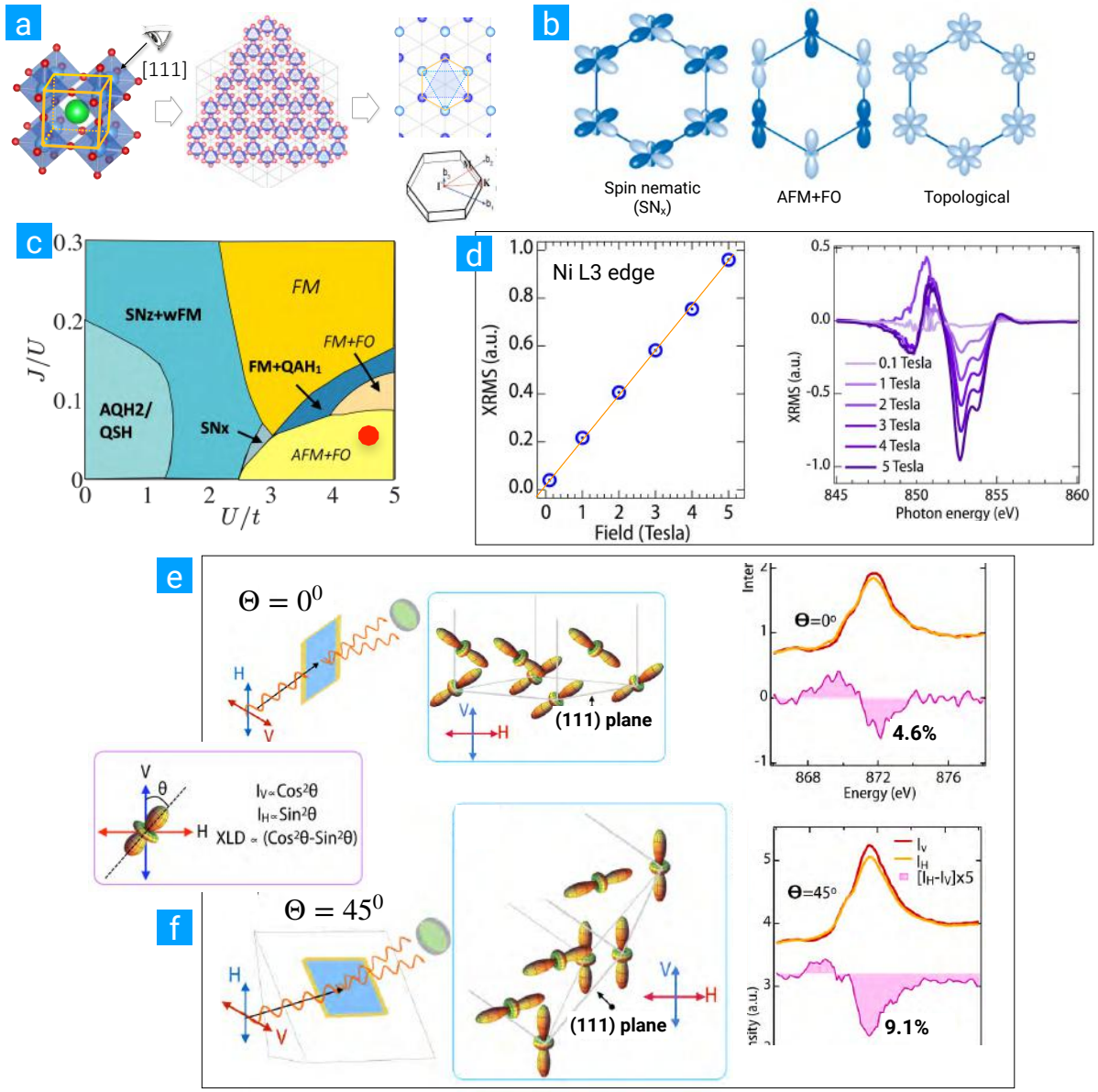


Fig. 6: *a: Growth of ABO_3 along $[111]$ leads to new lattice geometries. b-c: The weakly-correlated limit of $RENiO_3$ has been theorized to host interaction-driven topological phases such as a ferromagnetic nematic phase, Dirac half semimetal, and anomalous quantum Hall insulator. d: XRMS measurements on 2NNO/4LAO establish the presence of antiferromagnetic correlations. e-f: XLD spectra for $\phi = 0°$ (flat) and $\phi = 45°$ (wedge) geometries. The XLD signal was greatly increased for the wedge geometry, a sign of orbital polarization in 2NNO/4LAO.*

This control experiment emphasizes that the observed orbital polarization in the 2NNO/4LAO (111) superlattice is not a measurement artifact, and this buckled honeycomb lattice geometry engineered the orbitally polarized ground state. The obtained value of XLD around 9% is large as the finite bandwidth of the e_g bands and strong covalency strongly reduce the orbital polarization from the atomic limit. I need to mention, however, that by the nature of the spectroscopic probe, XLD can only establish the presence of orbital ordering or orbital polarization

but cannot resolve a specific type of orbital pattern present in the system. This can be done by using synchrotron-based resonant X-ray scattering on the Ni L-edge and by using density functional theory (DFT). Combined with the DFT prediction the experimental data revealed a novel kind of *anti-ferro-orbital ordering with staggered* $3d_{z2-r2}$ *orbitals rotated by 90° in subsequent layers*. This new quantum state is absent in the bulk nickelates.

5.4 Electronic structure of graphene-like nickelates

Probing buried graphene-like $[111]$ *perovskite layers using soft and hard X-ray photoemission.* As we have seen in the previous section, artificial heterostructures comprised of ultra-thin complex-oxide layers grown in the pseudo-cubic $[111]$ direction have been predicted to harbor a wide range of extraordinary quantum states stemming from the unique lattice geometry resembling graphene and the interplay between strong electronic correlations and band topology [60]. However, studying such atomic layers' electronic and structural properties remains a formidable challenge due to the limitations of conventional surface-sensitive techniques, which typically probe depths of only a few Angstroms.

In this section, I want to discuss a new experimental methodology that combines bulk-sensitive soft X-ray angle-resolved photoelectron spectroscopy (SX-ARPES), hard X-ray photoelectron spectroscopy (HAXPES), and first-principles DFT+U calculations. This powerful set is used to establish a direct and reliable method for *extracting momentum-resolved and angle-integrated valence-band electronic structures* of an ultra-thin buckled graphene-like layer of $NdNiO_3$-(111) sandwiched between two 4-unit-cell thick layers of insulating $LaAlO_3$-(111) (see Fig. 6a). Clearly, this is a challenging system for measurements as the active monolayer of $NdNiO_3$-(111) is buried under the capping layer of $LaAlO_3$-(111) [61].

First, soft X-ray angle-resolved photoemission spectroscopy (SX-ARPES) measurements shown in Figure 7b were performed using the high-resolution ADRESS beamline at the Swiss Light Source. To enhance the information depth and enable a momentum-resolved analysis of the hexagonal $NdNiO_3$, the measurements were conducted at high incoming photon energies ranging from 642 to 874 eV, far beyond conventional UV ARPES. Using such high photon energies effectively increases the inelastic mean-free paths of the photoelectrons within the superlattice by a factor of 3–5 compared to typical home ARPES investigations, substantially improving the probing depth. Theoretical calculations suggest the existence of two possible orbital arrangements in this material: one characterized by P1 symmetry in a 1×1 unit cell (P1 1×1) and another featuring P3 symmetry in a larger $\sqrt{3}\times\sqrt{3}R30^o$ lateral unit cell. Bulk-sensitive momentum-resolved measurements reveal excellent agreement with the band structure calculated using DFT+U for the Ni sites' antiferromagnetic (AFM) ordering with P1 1×1 symmetry shown in Figure 7c. These measurements provide direct evidence supporting the P1 1×1 symmetry, which perfectly aligns with the findings previously suggested by X-ray linear dichroism data (see Sec. 5.3) [61–63].

To investigate the entire sample depth, angle-integrated hard X-ray photoemission spectroscopy (HAXPES) measurements of the valence bands were carried out at a photon energy of 6.45 keV.

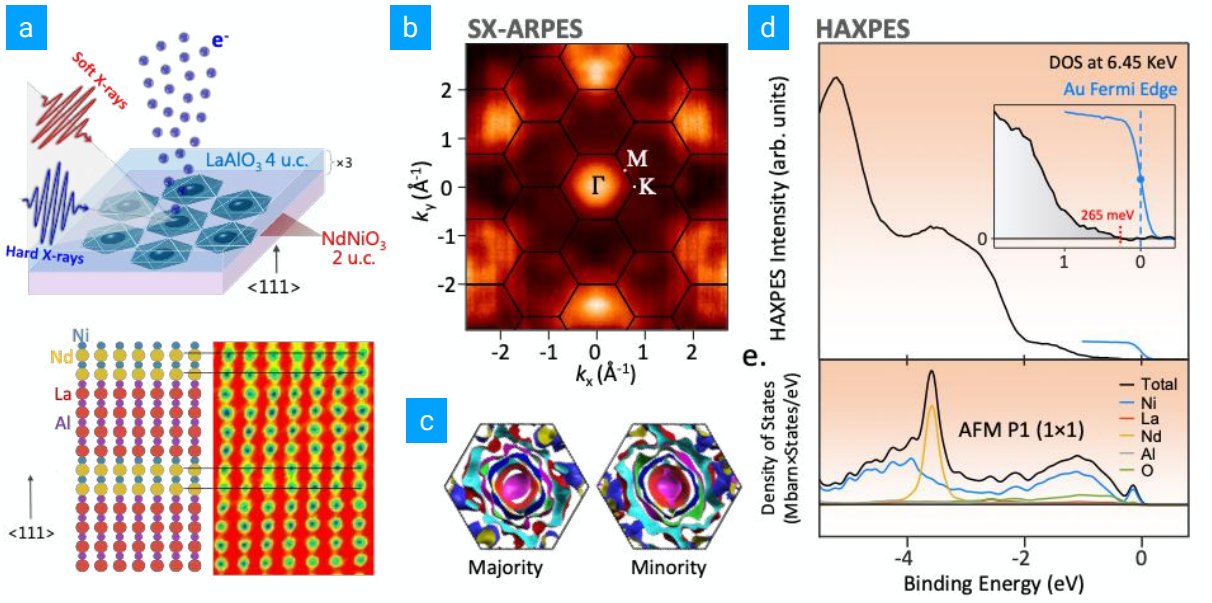


Fig. 7: *a: Schematic diagram of the sample and the SX-ARPES and HAXPES experimental geometries. b: Momentum-resolved SX-ARPES map of the Ni* $3d$ *states near the Fermi level. c: Isoenergetic cuts through the band structure in reciprocal space for the majority and minority bands. d: Bulk-sensitive HAXPES spectrum of the valence-band DOS. Inset shows a high-statistics spectrum of the valence-band maximum (at* $\sim$ 265 *meV), referenced to the Au Fermi edge. e: Cross-section-weighted element-projected and total DOS of the superlattice, calculated in the GGA+U framework of DFT and broadened to account for both experimental and hole lifetime broadening.*

At this hard X-ray energy, the estimated inelastic mean-free path is approximately 9 nanometers, enabling direct probing of the density of states across the entire sample. This approach facilitates a straightforward comparison between experimental data (Fig. 7d) and theoretical predictions (Fig. 7e). The experimental HAXPES valence-band spectrum shows remarkable agreement with the first-principles calculations regarding relative intensities and the positions of key features. Moreover, the experimental valence-band maximum is observed at a binding energy of $\sim$265 meV below the Fermi level. This value corresponds to the magnitude of the valence-band bandgap, indicating that the full bandgap of $NdNiO_3$-(111) is at least of this size. I remind the reader that in bulk (above 150 K), $NdNiO_3$ is a paramagnetic metal [61–63].

In conclusion, a combination of bulk-sensitive soft and hard X-ray photoemission techniques can be utilized to investigate the *momentum-resolved electronic band dispersion of a buried* two-dimensional $NdNiO_3$-(111) layer within a designed superlattice below a cap layer. Additionally, the density of states of this structure can be directly measured using HAXPES. Combined with first-principles DFT+U calculations, this new methodology provides direct and definitive evidence for an antiferro-orbital (AFO) order characterized by P1 symmetry within a 1×1 unit cell [64].

We learn from this example that we finally have a practical experimental approach to investigate the *momentum resolved* electronic structure of new quantum metals and semimetals as thin as a single atomic plane.

5.5 Artificial quantum spin liquid on lattices with extreme frustration

Can we make a new exotic state of matter? In what follows, I will discuss the, in my view, most enigmatic yet least experimentally understood state of matter called a spin liquid. To make things even more intriguing, we should look at its quantum version, a quantum spin liquid or QSL. If there is something that we do not fully understand microscopically, it would be the concept of liquid (e.g., water). These days, experimentalists are very good at finding and describing long-range order (LRO) by sharp Bragg intensities in reciprocal space. Conversely, a disordered spin gas phase (paramagnet) can be reliably detected. But the precise nanoscopic description of something which fluctuates in space and time with numerous short-range ordered (SRO) configurations of spins is still beyond our current computational capabilities.
Nevertheless, one can bravely embrace the idea and think, at least theoretically, if such a liquid of quantum spins is thermodynamically stable, what would be a Landau-kind order parameter if at all, and most importantly, how could we detect such a state experimentally (for a theory discussion see [65–70]).
As for the question of stability, we have one example of a true quantum spin liquid, ^{3}He, which exists only outside of the solid-state setting. Even theoretically, we still do not have complete answers to those questions. For example, we do know that a QSL cannot be described by the conventional Landau theory of phase transitions relying on spontaneously broken symmetry. Instead, one can introduce the idea of 'entanglement entropy' as a topological order parameter. Despite its novelty and usefulness for theory, experimentally, we do not have probes that couple to such a 'topological order parameter.' Thus, we mostly rely on negative statements about a quantum material in question for practical reasons. At best, we can verify in a magnetometer or via some sort of magnetic scattering that our magnetic crystal has no long-range spin order down to the lowest experimentally accessible temperature. This is hardly a satisfactory situation, but we need to say what would constitute a set of positive statements as an alternative. Here I list a few popular ones:

- Definition 1: a QSL is a state in which the spin-spin correlations decay to zero at large distances or $\langle S_i^\alpha S_j^\beta \rangle \to 0$ when $|r_i - r_j| \to \infty$. Objection: A classical liquid, spin nematic, or valence bond crystal also satisfies this definition.

- Definition 2: a QSL is a state without any spontaneously broken symmetry, but so is a valence-bond solid.

- Definition 3: a QSL is *a Mott insulator that possesses no long-range magnetic order, lacks any spontaneously broken symmetry, and carries a spectrum of fractional excitations*. At present, this is the definition most amicable for experimental verification. I recap that fractional excitations are quasi-particles, e.g., spinons, carrying a half-odd-integer spin, and fractionalized fermions are coupled to an emergent gauge field.

Why do we care about such exotics? Here is a short list of reasons: (1) Most SQLs are 'flat-band' systems; if doped, they may harbor high or potentially even room-temperature superconductivity, (2) for dimensionality greater than one, fractional excitations interact with each other

through emerging gauge fields, giving rise to string- and loop-like excitations akin to physics of quark-gluon plasma and (3) QSLs sustain a new type of topological non-local order and new spin excitations (anyons) which can be useful as an unconventional platform for quantum computing with topological qubits beyond silicon, aluminum, or ion traps.

How to make a QSL? As for the experimental realization of a QSL, the currently existing "recipes" are illuminating but very limited. On the one hand, a general guiding principle is that to reach a QSL, significant frustration resulting from the lattice geometry, multiple exchange terms, or bond conflict are the most essential prerequisites. After tremendous, decades-long efforts, promising *candidate materials* have been proposed and synthesized [71]. Interestingly, the underlying lattices of almost all known QSL candidates are bound to five types of geometries: triangular, pyrochlore, kagome, hyperkagome, and honeycomb lattices. This, in turn, limits the pursuit of new QSL materials and brings to the focus an open question of whether any additional lattice motifs can host a QSL and how it can be achieved experimentally?

Figure 8a illustrates the new approach for making such exotic phases called *geometrical lattice engineering* (GLE) (see Sec. 3 and Ref. [22]). GLE principally aims to design and fabricate lattices of artificial geometry by stacking on demand a specific number of atomic planes along unconventional crystallographic directions to facilitate unattainable in the bulk configuration of charges, orbitals, and spins. You had already seen this concept in action in subsection 5.3 where I described how to create graphene-like $NdNiO_3$ with a new anti-ferro orbital order.

Is there a real QSL based on the GLE? Here, I introduce a generic design of a new (quasi-2D) lattice derived from the spinel structure (chemical formula AB_2O_4) and demonstrate its feasibility for a QSL phase [72]. Concretely, I will use $CoCr_2O_4$ as a prototype; we fabricated a series of (111)-oriented ultra-thin films confined by non-magnetic Al_2O_3 layers into a quantum well geometry. Compared to its bulk counterpart, the onset of the ferrimagnetic transition decreases monotonically with reduced thickness and eventually shuts off in a single-unit slab of (111) $CoCr_2O_4$. In this quasi-2D limit, the degree of magnetic frustration becomes enhanced by almost 3 orders of magnitude with persisting spin fluctuations down to 30 mK. $CoCr_2O_4$ belongs to the normal spinel chromite family, $M Cr_2O_4$ (M=Mn, Fe, Co, and Ni) where the magnetically active M^{2+} ions occupy the tetrahedral A sites of the diamond sublattice and the Cr^{3+} ions occupy the octahedral B sites of the pyrochlore sublattice, possessing complex spin configuration of the ground state. Note, in bulk, $CoCr_2O_4$ has a collinear ferrimagnetic state first formed with the Curie temperature of $\sim$93 K, it transforms into an incommensurate spiral ferrimagnetic state at $\sim$26 K, and finally, an incommensurate to commensurate lock-in transition takes place at $\sim$14 K.

Now on to the GLE. As seen in Fig. 8a/b, when viewed along the $[111]$ direction, the structure is an intrinsic stacking of triangle (T) and Kagome (K) cation planes from Co and Cr ions embedded in the oxygen cubic close-packed frame. This leads to a sequence '-O-Cr(K)-O-Co(T)-Cr(T')-Co(T)-' in a single unit with four cation layers, which we denote as one quadruplet layer (1 QL). Based on this design idea, [n QL $CoCr_2O_4$/1.3 nm Al_2O_3] (n=1, 2, 4; 1 QL $\sim$ 0.48 nm) superlattices were fabricated by pulsed laser deposition on (0001)-oriented single crystal Al_2O_3 substrates.

What is the ground state of this new synthetic magnet? The investigation of the magnetic behavior of each sublattice was done by recording the resonant X-ray absorption spectroscopy (XAS) taken with left- and right-circularly polarized beams. The difference between those two spectra, called X-ray magnetic circular dichroism (XMCD), reflects the net magnetization of a specifically probed element. To make the statement even more contrasty, I will l compare the 4QL sample (bulk-like) to the most intriguing 1QL sample. As seen in Fig. 8c, the saturated magnetic signal is indeed observed in the thicker 4 QL $CoCr_2O_4$ and also it exhibits clear hysteresis loops at both Co and Cr L_3 edges. These findings signal for the long-range magnetic order even in 4QL thin samples. However, in sharp contrast, no hysteresis loop but a linear XMCD versus H relationship is found on both Co and Cr in 1 QL $CoCr_2O_4$, typical of a paramagnetic behavior!

To further examine if any long-range spin ordering emerges at extremely low temperatures, we performed the torque magnetometry measurements on 1 QL $CoCr_2O_4$ from 30 K down to 30 mK. This technique quantifies the magnetic torque response of a sample with respect to the applied magnetic field ($\tau \propto M \times H$). It is an exquisitely sensitive utility to probe vanishingly small magnetic signals from ultra-thin samples and interfaces. The result confirms that within the resolution of measurement and the entire temperature range, no hysteresis but a reversible parabolic $\tau \propto (\mu_0 H)^2$ relationship is observed for 1 QL, which implies a quantum paramagnetic behavior persisting down to 30 mK. Moreover, if we plot torque vs. T (see Fig. 8c (right)), we can immediately see that even at 30 mK, the ground state has large spin entropy, and its spectrum of excitations is *gapless* or spin-metal like. To make this story even more compelling, an extensive set of neutron reflectivity and mount spin resonance data affirms these conclusions.

Why does this quasi-2D artificial magnet enter a gapless QSL state? To obtain a microscopic insight into *how the designer lattice topology and quantum confinement alter the exchange interactions* and, consequently, the magnetic ground state, a set of DFT calculations and Monte-Carlo simulations were performed on bulk, 2, and 1 QL of (111) $CoCr_2O_4$. As seen in Fig. 8d (left), the theory revealed that the new emergent QSL ground state is a consequence of the markedly smaller exchange interaction along (111) because it is blocked by the vacuum-like spacer of Al_2O_3 with a gap of 4 eV in the ultrathin films compared to bulk. In fact, for 1 QL, this interaction is completely suppressed (!), while the in-plane interactions remain essentially bulk-like, all in contrast to the behavior in bulk. And now we understand that due to the entire blocked exchange along (111), our 1QL magnet turned into a 2D system with a pure Kagome-triangular motive of extremely large magnetic frustration. Fig. 8d (right) shows that the system reaches the frustration factor $f = T_{CW}/T_N$ of about 1000 (!) compared to the value in the bulk of 4–6.

Here is what we learned from GLE. The ground state is a well-defined long-range magnetically ordered state in bulk (QL $\to \infty$). As the number of QLs is reduced, it becomes more and more difficult to stabilize a conventional ordered state due to the enhancement of magnetic frustration. Eventually, the ground state becomes highly degenerate in 1 QL, unleashing dynamical spin fluctuations. I remind you that this is the regime where quantum effects play a pivotal role in bringing the system into a QSL (quantum paramagnet) state without a spin gap.

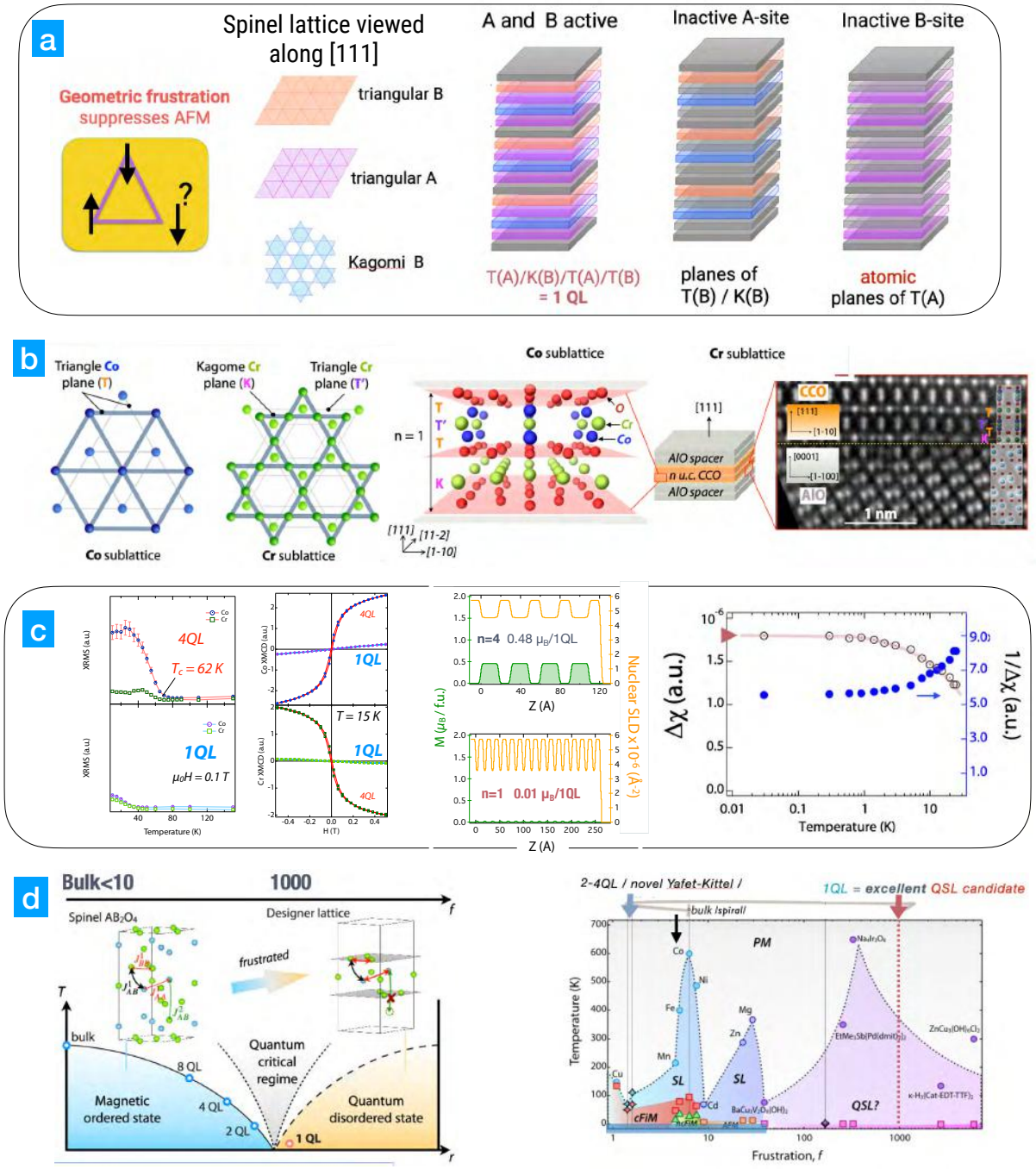


Fig. 8: *a: The spinel lattice is made of a stack of triangle and kagome planes when viewed along the* [111] *direction. b: View of cation triangle and kagome planes. c: XMCD measurements show long-range magnetic ordering even in 4QL CCO films, but paramagnetic behavior in 1QL CCO. Torque measurements reveal that even at 30 mK, the 1QL sample is gapless. d: Theory reveals the emergent QSL ground state is a consequence of the suppression of out-of-plane exchange interaction.*

6 Problems to solve, ideas to try

In this section, I want to point in ten directions, which is my challenge to you. Frankly, I do not have answers to any of those questions, so I leave it to you to seek possible solutions.

1. If you replace a conventional nano-second UV laser with a femtosecond one, what new synthesis regime can you reach? The intrigue here is that in the femtosecond regime, there is no time for laser heat dissipation as phonons are too slow (pico-second timescale).

2. How can we grow uniaxially strained structures?

3. Apart from GLE, how do we design structures with extreme frustration from interactions?

4. Can you think of a way to combine different classes of interesting quantum materials, such as TMO with TM dichalcogenides, TM oxyfluoride, and TM nitrites?

5. What happens if you combine different topological classes and antagonistic orders such as Dirac electrons-Cooper pairs or Cooper pairs and magnetic monopoles of a spin-ice?

6. Can you think of a design approach for structures that can 'zoom in' on a specific term of a Hamiltonian?

7. Can you create structures holding quantum chaos?

8. What about structures that reach a regime of quantum hydrodynamics?

9. How can you measure a spectrum of excitations (e.g., magnons, phonons, plasmons, orbitons) right at the interface?

10. What designer structures can directly reveal the quantum entanglement of fermions?

Acknowledgments

I dedicate this lecture to Prof. Daniel Khomskii, for whom I am indebted for many ideas described throughout the text and who was and still is a great source of influence and inspiration for what it means to be a physicist. Also, I express my deep gratitude to my postdoctoral advisor Prof. Bernhard Keimer who relentlessly shared his encyclopedic knowledge of condensed matter physics and for his guidance in my first steps into the physics of oxide interfaces. And finally, I offer my sincere thanks to all my Ph.D. students, postdocs, and collaborators. They contributed countless research hours and incredible talents in pushing the frontiers of this stimulating field. The introductory section 1 is derived in part from Ref. [73]. Work on this manuscript was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0022160.

References

[1] B. Keimer and J.E. Moore, Nat. Phys. **13** 1045 (2017)

[2] D.I. Khomskii: *Transition metal compounds* (Cambridge University Press, 2014)

[3] L. Pauling, J. Am. Chem. Soc. **51** 1010 (1929)

[4] L. Pauling: *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry* (Cornell University Press, 1960)

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

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

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

[8] J. Hubbard, Proc. Roy. Soc. London A **277** 237 (1964)

[9] D.I. Khomskii and S.V. Streltsov, Chem. Rev. **121** 2992 (2020)

[10] D.I. Khomskii: *Basic aspects of the quantum theory of solids: order and elementary excitations* (Cambridge University Press, 2010)

[11] E. Pavarini, E. Koch, J. van den Brink, and G.Sawatzky (eds.): *Quantum Materials: Experiments and Theory*, Modeling and Simulation, Vol. 6 (Forschungszentrum Jülich, 2016)

[12] W. Witczak-Krempa *et al.*, Annu. Rev. Condens. Matter Phys. **5** 57 (2014)

[13] J. Chakhalian *et al.*, Rev. Mod. Phys. **86** 1189 (2014)

[14] H.Y. Hwang *et al.*, Nat. Mater. **11** 103 (2012)

[15] P. Zubko *et al.*, Annu. Rev. Condens. Matter Phys. **2** 141 (2011)

[16] S. Stemmer and S.J. Allen, Annu. Rev. Mater. Res. **44** 151 (2014)

[17] K.R. Poeppelmeier and J.M. Rondinelli, Nat. Chem. **8** 292 (2016):

[18] J. Chakhalian, A.J. Millis, and J. Rondinelli, Nat. Mater. **11** 92 (2012)

[19] J. Liu *et al.*, Nat. Commun. **4** 2714 (2013)

[20] J.W. Freeland *et al.*, Europhys. Lett. **96** 57004 (2011)

[21] J. Chakhalian *et al.*, Nat. Phys. **2** 244 (2006)

[22] X. Liu *et al.*, MRS Commun. **6** 133 (2016)

[23] A.M. Cook and A. Paramekanti, Phys. Rev. Lett. **113** 077203 (2014)

[24] M. Kareev *et al.*, J. Appl. Phys. **109** 114303 (2011)

[25] G. Koster *et al.*: *Growth studies of heteroepitaxial oxide thin films using reflection high-energy electron diffraction (RHEED)* in G. Koster, M. Huijben, and G. Rijnders (eds.): *Epitaxial Growth of Complex Metal Oxides* (Woodhead Publishing, 2015) pp.3–29

[26] G. Koster: *Reflection high-energy electron diffraction (RHEED) for in situ characterization of thin film growth* in G. Koster and G. Rijnders (eds.): *In situ characterization of thin film growth* (Woodhead Publishing, 2011) pp.3–28

[27] P.W. Anderson and E.I. Blount, Phys. Rev. Lett. **14** 217 (1965)

[28] N.A. Benedek and T. Birol, J. Mater. Chem. C **4** 4000 (2016)

[29] Y. Shi *et al.*, Nat. Mater. **12** 1024 (2013)

[30] T.H. Kim *et al.*, Nature **533** 68 (2016)

[31] T. Kolodiazhnyi *et al.*, Phys. Rev. Lett. **104** 147602 (2010)

[32] J. Fujioka *et al.*, Sci. Rep. **5** 13207 (2015)

[33] T. Ishidate *et al.*, Phys. Rev. Lett. **78** 2397 (1997)

[34] M. Mochizuki and M. Imada, New J. Phys. **6** 154 (2004)

[35] Y. Cao *et al.*, Phys. Rev. Lett. **116** 076802 (2016)

[36] A. Ohtomo *et al.*, Nature **419** 378 (2002)

[37] Y.J. Chang *et al.*, Phys. Rev. Lett. **111** 126401 (2013)

[38] J. Kondo, Prog. Theor. Phys. **32** 37 (1964)

[39] A.C. Hewson: *The Kondo problem to heavy fermions* (Cambridge University Press, 1997)

[40] G.-Y. Guo, S. Maekawa, and N. Nagaosa, Phys. Rev. Lett. **102** 036401 (2009)

[41] H. Tsunetsugu, M. Sigrist, and K. Ueda, Rev. Mod. Phys. **69** 809 (1997)

[42] S. Doniach, physica B+C **91** 231 (1977)

[43] T. Jungwirth *et al.*, Rev. Mod. Phys. **78** 809 (2006)

[44] P. Fazekas and E. Müller-Hartmann, Z. Phys. B **85** 285 (1991)

[45] C. Zener, Phys. Rev. **81** 440 (1951)

[46] J. Mannhart and D.G. Schlom, Science **327** 1607 (2010)

[47] T. Katsufuji, Y. Taguchi, and Y. Tokura, Phys. Rev. B **56** 10145 (1997)

[48] Y. Cao *et al.*, Appl. Phys. Lett. **107** 112401 (2015)

[49] H.W. Jang *et al.*, Science **331** 886 (2011)

[50] Y. Cao *et al.*, Nat. Commun. **7** 10418 (2016)

[51] See Supplemental Material of [35] for details and additional data.

[52] S.-P. Chiu and J.-J. Lin, Phys. Rev. B **87** 035122 (2013)

[53] M. Gabay and J.-M. Triscone, Nat. Phys. **9** 610 (2013)

[54] D.N. Sheng *et al.*, Phys. Rev. Lett. **97** 036808 (2006)

[55] X. Liu *et al.*, MRS Commun. **6** 133 (2016)

[56] S. Middey *et al.*, Sci. Rep. **4** 6819 (2014)

[57] D. Xiao *et al.*, Nat. Commun. **2** 596 (2011)

[58] K.-Y. Yang *et al.*, Phys. Rev. B **84** 201104 (2011)

[59] D. Doennig, W.E. Pickett, and R. Pentcheva, Phys. Rev. B **89** 121110 (2014)

[60] G.A. Fiete and A. Rüegg, J. Appl. Phys. **117** 172602 (2015)

[61] A. Arab *et al.*, Nano Lett. **19** 8311 (2019)

[62] L. Plucinski *et al.*, Phys. Rev. B **78** 035108 (2008)

[63] S. Middey *et al.*, Phys. Rev. Lett. **116** 056801 (2016)

[64] A.X. Gray *et al.*, Nat. Mater. **10** 759 (2011)

[65] L. Savary and L. Balents, Rep. Prog. Phys. **80** 016502 (2016)

[66] J. Knolle and R. Moessner, Annu. Rev. Condens. Matter Phys. **10** 451 (2019)

[67] Y. Zhou, K. Kanoda, and T.-K. Ng, Rev. Mod. Phys. **89** 025003 (2017)

[68] S. Sachdev, Nat. Phys. **4** 173 (2008)

[69] L. Balents, Nature **464** 199 (2010)

[70] F. Mila, Eur. J. Phys. **21** 499 (2000)

[71] J.R. Chamorro, T.M. McQueen, and T.T. Tran, Chem. Rev. **121** 2898 (2020)

[72] X. Liu *et al.*, Nano Lett. **21** 2010 (2021): 2010

[73] X. Liu: *Artificial Quantum Many-Body States in Complex Oxide Heterostructures at Two-Dimensional Limit* (PhD Thesis, University of Arkansas, 2016)

10 Orbitals, Frustration and Quantum Criticality

Matthias Vojta
Institut für Theoretische Physik
Technische Universität Dresden

Contents

1 Introduction **2**

2 Quantum phase transitions **2**
2.1 Phenomenology and Landau-Ginzburg-Wilson theory 2
2.2 An example: Coupled dimers and $TlCuCl_3$ 4
2.3 Frustrated systems: What is different? 6

3 Frustration and novel states **7**
3.1 Classical spin liquids 8
3.2 Quantum spin liquids 9
3.3 Valence-bond solids 10
3.4 Order by disorder and unconventional types of order 11

4 More ingredients: Orbitals and spin-orbit coupling **11**
4.1 Magnetic anisotropies and novel forms of frustration 11
4.2 Orbitals and spin-orbital liquids 12

5 Conventional quantum criticality in frustrated systems **13**
5.1 Magnetic ordering transitions 13
5.2 Field-driven transitions and BEC phenomena 14

6 Transitions involving topological states **15**
6.1 Confinement transitions and fractionalized criticality 15
6.2 Deconfined quantum criticality 18

7 Mott and Kondo transitions **20**
7.1 Fermi liquids and non-Fermi liquids 21
7.2 Mott transitions 22
7.3 Kondo and orbital-selective Mott transitions 24

8 Summary **25**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

This chapter is devoted to phase transitions at zero temperature, usually called quantum phase transitions (QPT), their critical behavior, and its changes arising from frustration and the presence of orbital degrees of freedom [1].

QPT and quantum criticality define an active field of research which goes back to the work of Hertz in 1976 [2] who considered magnetic ordering transitions in metals. Much progress was made in the 1990s and 2000s [3], such that many classes of symmetry-breaking QPT in insulators are reasonably well understood by now, with agreement between experiment and theory. In contrast, transitions in metals remain only partially understood [3, 4]. Moreover, and most relevant to this chapter, recent developments in the field of frustrated and topological systems have brought into focus entirely new forms of quantum criticality which are under intense investigation today [5]. For some of them, microscopic ingredients beyond the simplest non-relativistic single-orbital picture are crucial, defining an extremely fruitful and rich avenue of research.

In the following, we will focus on interacting electrons in solids and thus on collective phenomena. In contrast, we will not cover transitions driven by the topology of band electrons; similarly, we will not be concerned with transitions driven by quenched disorder. Our primary interest is on thermodynamic and linear-response spectral properties of systems in the vicinity of a QPT. The non-equilibrium quantum dynamics near QPTs as well as genuine non-equilibrium phase transitions have become an intense research field on its own, but are beyond the scope of this chapter.

Given the complexity of material, we will mainly discuss conceptual ideas and qualitative aspects of theory; for concrete computations we refer the reader to the literature. Experimental results will be mentioned when appropriate.

2 Quantum phase transitions

Before turning to frustration, orbitals, and the like, we will summarize the main aspects of "conventional" quantum criticality. For reasons of space, this review can be nowhere close to complete. However, many extensive texts on this subject are available [3, 4, 6] which we refer the reader to for a more detailed exposure.

2.1 Phenomenology and Landau-Ginzburg-Wilson theory

A quantum phase transition (QPT) is a phase transition taking place at temperature $T = 0$ upon tuning a non-thermal control parameter like pressure or magnetic field. The finite-temperature properties near a continuous QPT are highly unusual: Due to the peculiar properties of the quantum ground state at the transition point, dubbed quantum critical point (QCP), the so-called quantum critical regime located at finite T above the QCP, Fig. 1, displays properties distinct from that of any stable phase of matter. These properties include power-law behavior with

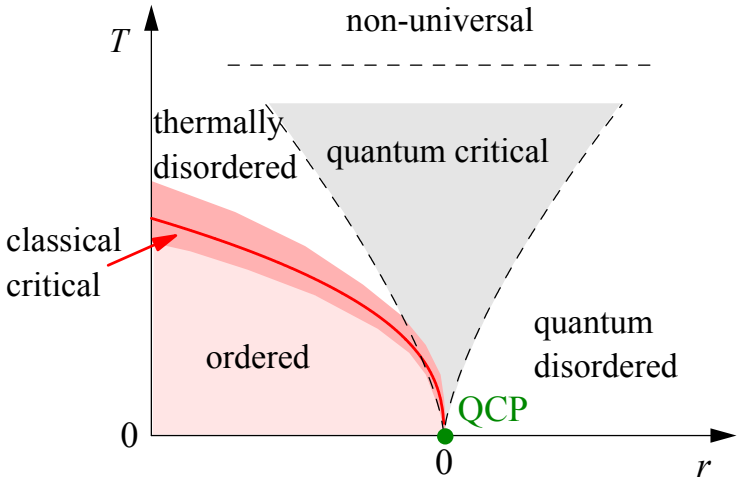


Fig. 1: *Generic phase diagram in the vicinity of a quantum critical point as function of a non-thermal control parameter r and temperature T. An ordered phase exists for $r < 0$ and low T, bounded by a line of classical phase transitions which terminates at the QCP at $r = 0$, $T = 0$. The quantum critical regime is defined by $k_B T \gg |r|^{\nu z}$, where ν and z are the correlation length and dynamical exponents.*

unconventional exponents of thermodynamic and transport quantities as function of absolute temperature as well as scaling behavior, where suitably rescaled observables depend only on dimensionless ratios of external parameters.

From a theoretical perspective, the universal properties of QPTs can often be described using a continuum quantum field theory for the transition's order parameter. The choice of the latter is dictated by the way in which symmetries of the Hamiltonian are spontaneously broken at the transition. This goes back to Landau who pioneered the ideas of symmetry breaking and local order parameters in the context of phase transitions. This concept was later extended to quantum phase transitions by taking into account temporal order-parameter fluctuations, i.e., quantum fluctuations – this leads to the so-called Landau-Ginzburg-Wilson (LGW) approach.

For Mott-insulating quantum magnets the LGW theory for a zero-temperature transition between a featureless paramagnet and, e.g., a collinear ordered antiferromagnet takes the form of a quantum φ^4 model with the action

$$\mathcal{S} = \int d^d x \int_0^\beta d\tau \left(\frac{c_0^2}{2} (\partial_i \vec{\varphi})^2 + \frac{1}{2} (\partial_\tau \vec{\varphi})^2 + \frac{\delta_0}{2} \vec{\varphi}^2 + \frac{u_0}{4!} (\vec{\varphi}^2)^2 \right) \tag{1}$$

where $\partial_i = \partial/\partial x_i$, and $\vec{\varphi}(\vec{x}, \tau)$ is a local N-component order-parameter field which is assumed to vary slowly in space and time and encodes the ordering tendency at a microscopic wavevector $\vec{Q}$. Further, τ is imaginary time, and c_0, δ_0, and u_0 are parameters. Decreasing the non-thermal control parameter δ_0 at low temperature tunes a transition between a disordered and an ordered phase, with the O(N) symmetry spontaneously broken in the latter; $N = 3$ for collinear Néel order in the presence of SU(2) spin symmetry. More precisely, δ_0 acquires a temperature-dependent renormalization, and the transition occurs at $\delta_0 = \delta_c$ where the renor-

malized δ vanishes. The distance to the QCP can be expressed as

$$r = \delta_0 - \delta_c(T{=}0) \tag{2}$$

and may be tuned by pressure or chemical composition. Eq. (1) can also describe non-magnetic ordering transitions, such as the onset of charge order accompanied by the breaking of lattice translation symmetry.

The thermodynamic properties of Eq. (1) are essentially understood, as they can be computed analytically using renormalization-group techniques as well as numerically. The critical exponents of the QPT are known to a good accuracy in all space dimensions. Similarly, dynamical and spectral properties have been considered, and a detailed exposition is given in Ref. [3].

In Eq. (1) space and time enter symmetrically, corresponding to a dynamical exponent $z = 1$. The time direction in the integral may be interpreted as an additional space direction, such that the quantum theory in d dimensions at $T = 0$ is equivalent to a classical theory in $D = d + z$ dimensions. While the local order-parameter description with $z = 1$ applies to many QPT in insulators, the situation in metals is more complicated due to the presence of low-energy fermionic excitations. Two additional remarks are in order: (i) QPTs into ferromagnetic or polarized phases in the presence of SU(2) spin symmetry follow a quantum dynamics different from that of the φ^4 model because a *conserved* density changes across the transition. (ii) Berry-phase terms, which are generically present in a field-theory description of spin systems, do not appear in Eq. (1) because they are irrelevant for the transition between featureless paramagnet and antiferromagnet. They are, however, responsible for much of the physics beyond LGW which will be described in Sec. 6.

For finite-temperature (i.e. classical) transitions, the upper critical dimension above which mean-field critical behavior is realized is $D_c^+ = 4$ for a standard φ^4 theory. In the quantum case, the presence of temporal fluctuations implies that the upper critical dimension for QPTs is given by $d_c^+ = 4{-}z$. For instance, continuous QPTs in $d{=}3$ with $z{=}1$ display mean-field behavior with logarithmic corrections. For phase transitions involving fermions the situation may be more complicated, though.

A last parenthetical remark here: Zero-temperature phase transitions, both continuous and discontinuous, can also occur in purely classical models. Obvious examples occur classical models of vector spins: For instance, the field-driven transition to saturation in a classical Heisenberg model is typically continuous.

2.2 An example: Coupled dimers and $TlCuCl_3$

A class of simple microscopic models displaying magnetic QPTs is given by coupled dimers, i.e., lattice systems with a crystallographic unit cell containing two spins $1/2$. Consider the Heisenberg Hamiltonian

$$H = J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j + \lambda J \sum_{\langle ij \rangle'} \vec{S}_i \cdot \vec{S}_j \tag{3}$$

where the first sum runs over all dimers, whereas the second sum covers all inter-dimer bonds. A square-lattice realization is shown in Fig. 2. The limit $\lambda = 0$ corresponds to disconnected

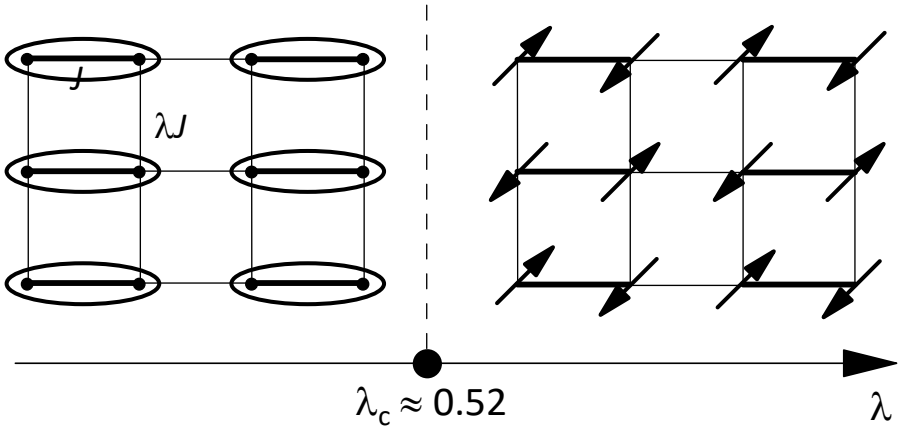


Fig. 2: *Square-lattice coupled-dimer model* (3) *with phase diagram: The two Heisenberg coupling J and λJ are shown as thick and thin lines, respectively. The ellipsoids represent singlet pairs of spins $1/2$. At a critical value λ_c the system transits from a gapped singlet paramagnet (left) to a Néel antiferromagnet (right).*

spin pairs, each of them having a singlet $S = 0$ ground state and a triplet $S = 1$ excited state, separated by an excitation energy J. The full lattice model has two distinct phases, which can be easily discussed:

Limit $\lambda \ll 1$: This implies weakly coupled dimers, leading to a disordered (i.e. quantum paramagnetic) phase with no broken symmetries and exponentially decaying spin correlations.

Limit $\lambda \sim 1$: Here the dimers are strongly coupled, and long-range antiferromagnetic order with broken SU(2) symmetry emerges. For the lattice shown in Fig. 2, $\lambda = 1$ represents a square lattice which is known to display long-range order.[1]

A quantum phase transition must occur at an intermediate value of λ, Fig. 2. As the order parameter is the staggered magnetization, the QPT is described by the LGW theory (1) with $N = 3$ components. The excitation gap of the quantum paramagnet closes upon approaching the QCP. The ordered phase displays two gapless Goldstone modes corresponding to the broken spin rotation symmetry as well as a gapped Higgs mode corresponding to amplitude fluctuations of the order parameter.

A paradigmatic experimental realization of coupled dimers, here in three space dimensions, is found in the Mott-insulating material $TlCuCl_3$ [7]. The magnetic Cu ions form dimers, and at ambient pressure and low temperature the material is in the quantum paramagnetic phase. Upon applying hydrostatic pressure, the inter-dimer interactions increase (i.e. λ in Eq. (3) increases) such that the system eventually reaches a state with antiferromagnetic long-range order. Ignoring the (weak) spin-orbit coupling, the QPT between the two states is described by the LGW theory (1) as above. Given that $D = d + z = 4$, the QPT is of mean-field character.

[1] The system in Fig. 2 becomes disordered again for $\lambda \gg 1$, as this limit corresponds to decoupled spin ladders.

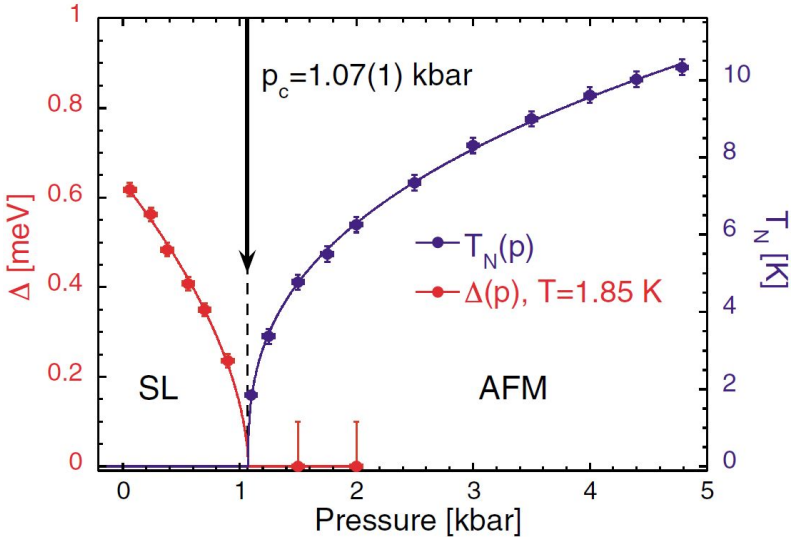


Fig. 3: *Experimental results for the coupled-dimer system $TlCuCl_3$, showing the magnetic excitation gap Δ and the Néel temperature T_N as function of applied pressure p. The magnetic QPT is located at $p_c = 1.07$ kbar. (Figure taken from Ref. [7])*

2.3 Frustrated systems: What is different?

The considerations so far cover "simple" symmetry-breaking transitions, i.e., transitions between a symmetric – and also otherwise featureless – state and a state which can be characterized by a local order parameter and spontaneously breaks one or more symmetries of the Hamiltonian. While such transitions can of course also occur in frustrated systems, more complicated situations frequently arise which cannot be captured by a simple LGW theory. Important cases are:

1. If a quantum paramagnetic phase is a fractionalized spin liquid, it is *not* featureless, because it is characterized by topological order.

2. The ordered-state manifold may be unconventional, i.e., not be characterized by a local order parameter or by a unique ordering wavevector. Long-range order may arise exclusively from fluctuation effects.

3. A transition might occur between states without spontaneous symmetry breaking.

4. The active quantum degrees of freedom can be different from the fluctuations of the order parameter, i.e., if a local order parameter exists, it might be a composite when expressed in the elementary degrees of freedom.

5. Frustration may enhance fluctuations such that the transition is rendered first order.

In Sec. 5 and 6 we will cover some of these cases in more detail.

3 Frustration and novel states

Frustration refers to the presence of multiple constraints which cannot be simultaneously satisfied. An important arena is frustrated magnetism where the constraints arise from the minimization of (pairwise) interaction energies: In a frustrated magnet, not all interactions can be simultaneously minimized. The perhaps simplest example is given by antiferromagnetically coupled Ising spins on a triangle. Frustration can arise from the geometry of the underlying lattice and/or from the nature of the interactions. The most obvious effect of frustration is to counteract the usual tendency towards symmetry-breaking order at low temperatures. As a result, a frustrated system may either have a strongly reduced ordering temperature or show no order at all, the latter often leading to exotic liquid-like phases. In addition, the suppression of conventional ordering phenomena can induce a competition of multiple less conventional phases, resulting in complex phase diagrams, non-trivial crossover phenomena, an accumulation of entropy at low temperature, and a large sensitivity to tuning parameters.

The past decade has seen a flurry of interest in frustrated systems [8–13], primarily driven by the search for novel states of matter. Prime examples are spin liquids with fractionalized degrees of freedom, skyrmion lattices with emergent artificial electrodynamics, fractionalized Fermi liquids, and their descendants. Many of these phases are characterized by non-trivial topological properties.

In this section, we introduce important concepts for frustrated magnets. The discussion here will focus on Mott insulators with local moments; frustrated metals define a large separate topic on its own, and we will only touch upon this in Sec. 7. We will consider lattice systems of local moments, i.e., quantum-mechanical spins transforming as SU(2) vectors, with a Hamiltonian containing two-spin interactions plus, perhaps, multi-spin exchange terms. The most generic model Hamiltonian is an antiferromagnetic Heisenberg model of spins S with nearest-neighbor interactions J,

$$\mathcal{H} = J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j \,. \tag{4}$$

The Heisenberg interaction in Eq. (4) favors antiparallel moments on neighboring lattice sites. Consequently, this interaction is non-frustrated on lattices where all closed loops of interaction paths have even length, such that an alternating up–down arrangement, corresponding to collinear magnetic order, can cover the lattice. This applies to the square and cubic lattices as well as, e.g., the honeycomb lattice. In contrast, frustration is induced on lattices with odd-length loops, e.g., the triangular, kagome, bcc, fcc, and pyrochlore lattices. On some of these lattices, a magnetically ordered ground state – often non-collinear – is realized for any S despite the existence of frustration, the triangular lattice with its 120° order being an established example, while in other cases order may be entirely absent.

In addition to the described *geometric frustration*, rooted in the geometry of the underlying lattice, incompatible constraints may be caused by the nature of the exchange interactions, leading to *exchange frustration*. A prominent case are so-called Kitaev interactions [14], to be described in Sec. 4 below.

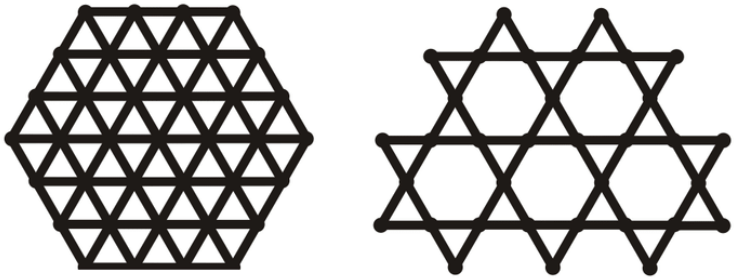


Fig. 4: *Two-dimensional lattices with geometric frustration: triangular (left), kagome (right).*

Given that frustration tends to suppress magnetic order, a popular experimental way to quantify frustration in a given system is the so-called frustration ratio, $f = |\Theta_{\rm CW}|/T_{\rm N}$, where $T_{\rm N}$ is the ordering temperature and $\Theta_{\rm CW}$ the Curie-Weiss temperature, the latter being a measure for the strength of exchange interactions [8]. Materials with $f > 5$ are commonly called "frustrated". The extreme case of no long-range order (LRO) down to $T = 0$, formally $f = \infty$, then corresponds to a ground state with only short-range correlations. A regime with highly correlated but fluctuating spins and no LRO at temperatures $T \ll |\Theta_{\rm CW}|$ is often dubbed "spin liquid" (although more precise definitions are available, see below).

3.1 Classical spin liquids

In the classical limit, formally obtained for spin size $S \to \infty$, spins can be viewed as unit vectors, and non-trivial commutators vanish. Frustration may lead to a classical ground state which is either unique up to global symmetry transformations – in this case the system is called "weakly frustrated" – or which has degeneracies scaling with the system size, rendering the system "strongly frustrated".[2] In the latter case, the resulting manifold of lowest-energy states defines a *classical spin liquid*. A celebrated example is spin ice, referring to moments with local Ising anisotropy and ferromagnetic interactions on a pyrochlore lattice, viz. a lattice of corner-sharing tetrahedra [15].

Often, a classical spin liquid can be characterized by a set of local conditions which define the ground-state manifold (but *not* a unique state up to global symmetry transformations, as explained above). Examples are the conditions "two in, two out" for the Ising configurations of individual tetrahedra of spin ice or the condition $\sum_\triangle \vec{S}_i = 0$ for the spin configurations of a kagome-lattice Heisenberg model. Hence, these conditions *underconstrain* the manifold of states; recall that the original problem of minimizing all Hamiltonian terms simultaneously *overconstrains* the manifold of states if frustration is present. Local constraints can often be formulated as an emergent lattice gauge theory. For instance, the "two in, two out" condition can be translated into $\mathrm{div}\, b = 0$ where b is an artificial magnetic field and div a suitably defined lattice divergence.

For Ising spins (i.e. with countable number of states) a classical spin liquid can be characterized by an extensive ground-state entropy S_0/N where N is the number of lattice sites. Typical

[2] Intermediate cases with sub-extensive degeneracies exist as well.

examples are the Ising model on a triangular lattice, with $S_0/(Nk_B) \approx 0.323$ [16], and classical spin ice, with $S_0/(Nk_B) \approx 1/2\ln(3/2) \approx 0.203$ [17]. For classical spin liquids made from XY or Heisenberg spins a residual entropy cannot be defined, but the degeneracy may be quantified via the difference between the number of continuous degrees of freedom and the number of local constraints.

Elementary excitations of classical spin liquids correspond to configurations which violate one (or more) of the local ground-state conditions; in the gauge-theory language these become elementary charges. For spin ice, the excitations are tetrahedra with "three-in, one-out" or "one-in, three-out" configurations; these have been shown to behave like magnetic monopoles upon including dipolar interactions [18].

3.2 Quantum spin liquids

With quantum fluctuations included, frustrated systems may realize local-moment states without symmetry breaking and only short-range order down to lowest temperatures. Such quantum spin liquids (QSLs) [9,11,12] display some differences compared to their classical counterparts: (i) Quantum fluctuations typically remove the extensive ground-state degeneracy of strongly frustrated systems by quantum tunnelling, resulting in unique ground states (up to global symmetry transformations or topological degeneracies). (ii) QSLs are thermodynamically stable phases of matter, characterized by emergent dynamic gauge fields and topological order. This implies the existence of fractionalized excitations which are coupled to the gauge field. Despite this coupling, the fractionalized excitations are asymptotically free, i.e., deconfined. (iii) The wavefunctions of QSLs can be characterized by long-range entanglement [19,20]. Importantly, QSLs need to be distinguished from "trivial" quantum paramagnets without topological order and fractionalization, like the coupled-dimer magnets of Sec. 2.2.

Different types of QSLs can be distinguished depending on the spectrum and statistics of the emergent excitations and on the gauge structure. Prominent examples are fully gapped $\mathbb{Z}_2$ spin liquids, for which topological order can be sharply defined, and algebraic U(1) spin liquids with gapless excitations. For an in-depth discussion of topological order and attempts of classifications we refer the reader to the literature [9, 11, 21]. Relevant to the existence of non-trivial many-body states is a theorem due to Lieb-Schulz-Mattis [22] and its higher-dimensional generalization by Hastings [23]. It states that in a system with half-odd-integer spin per unit cell and global U(1) symmetry, the excitation spectrum in the thermodynamic limit cannot simultaneously fulfill the two conditions: (a) the ground state is unique and (b) there is a finite gap to all excitations. This implies that a gapped symmetry-unbroken state must have a ground-state degeneracy which is topological in nature. We finally note that, conceptually, topological order and fractionalization may co-exist with spontaneous symmetry breaking: For instance, broken time-reversal symmetry on top of a spin liquid leads to a chiral spin liquid, while magnetic long-range order leads to a fractionalized ordered magnet.

An intuitive picture of a QSL with underlying SU(2) symmetry is provided by the resonating valence-bond (RVB) idea, Fig. 5, originally proposed by Anderson for the triangular-lattice

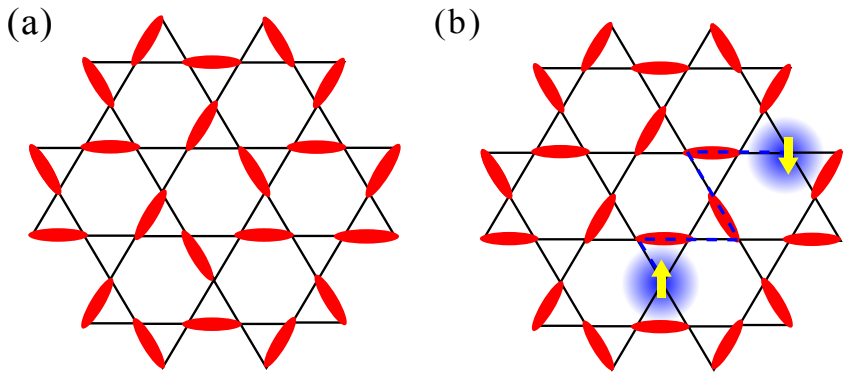


Fig. 5: *Illustration of an RVB state, here on the kagome lattice. (a) Nearest-neighbor dimer covering, with the ellipsoids representing singlet pairs of spins* $1/2$*. The RVB state is given by an equal-weighted superposition of different such coverings. (b) Pair of spinon excitations, each carrying spin* $1/2$*.*

Heisenberg model [24]. RVB refers to pairing spins on a lattice into singlets and then forming a quantum superposition of many different pairings, i.e., different dimer coverings of the lattice, such that the symmetries of the Hamiltonian are preserved.[3] This picture captures the aspect of fractionalized excitations, as the breaking of a dimer leads to two monomer excitations with independent dynamics: These monomers are objects carrying charge 0 and spin $1/2$, typically called *spinons*. In a $\mathbb{Z}_2$ spin liquid, they are coupled to an emergent $\mathbb{Z}_2$ gauge field, whose excitations are $\mathbb{Z}_2$ vortices (or fluxes) called *visons*.

A well-studied spin model with geometric frustration is the Heisenberg model on the kagome lattice. For quantum spins $1/2$, with antiferromagnetic interactions as in Eq. (4), there is strong numerical evidence that this realizes a fractionalized QSL. However, the nature of this QSL has not been conclusively clarified to date, as numerical results have been interpreted in favor of either a gapped $\mathbb{Z}_2$ spin liquid [26] or a U(1) spin liquid with a Dirac-cone spectrum [27, 28]. A candidate material realizing the kagome-lattice spin-$1/2$ Heisenberg model is Herbertsmithite, $ZnCu_3(OH)_6Cl_2$, which indeed displays spin-liquid-like behavior [29, 30]. However, the role of quenched disorder is debated [30]. Numerical evidence for QSL phases in Heisenberg models of spins $1/2$ has also been found for square [31] and triangular-lattice models [32] with first and second-neighbor interaction, so-called J_1-J_2 models. Close experimental realizations of the triangular-lattice J_1-J_2 model appear in the delafossite family $NaYbX_2$ ($X=$ S, Se, O) which show spin-liquid behavior at low T [33, 34].

3.3 Valence-bond solids

An alternative quantum paramagnetic state of spins $1/2$ that can be constructed from dimer coverings of the underlying lattice is a so-called valence-bond solid (VBS). In this state, the wavefunction is dominated by a single covering with a periodic arrangement of dimers. As a

[3] The first existence proof of a $\mathbb{Z}_2$ spin liquid was given for a triangular-lattice quantum dimer model which realizes an RVB phase [25].

result, the state spontaneously breaks translation and rotation symmetry of the lattice, hence the label *solid*. Excitations of VBS states carry integer spin, i.e., spinons are confined.
VBS variants can be constructed for larger constituent spins and/or from larger units, the common theme being that the state in the resulting magnetic unit cell represents a spin singlet. For instance, plaquette VBS with unit cells of four spins $1/2$ have been discussed for the square-lattice checkerboard and J_1-J_2 models.

3.4 Order by disorder and unconventional types of order

In addition to phases with unbroken spin symmetry, like spin liquids and valence-bond solids, frustrated spin systems can of course display phases with broken spin symmetry, both conventional and unconventional [13].
First, conventional magnetic order can emerge in an unconventional way. Most prominent is so-called "order by disorder" which refers to a situation where a frustration-induced degeneracy of the classical ground-state manifold is lifted by fluctuations, either thermal or quantum [35]. A well-studied example is the easy-plane pyrochlore antiferromagnet, where long-range order emerges due to fluctuations from a one-parameter manifold of classically degenerate states [36].
Second, less conventional magnetic order can appear as a result of large crystallographic unit cells or non-Heisenberg interactions. Among the possibilities are so-called multi-Q states where the ordering pattern results from the superposition of modulations with multiple inequivalent wavevectors, among which skyrmion lattices have attracted particular attention [37].
Third, ordered states may spontaneously break spin symmetry not by dipolar order, but by order in higher multipole channels. The simplest form is quadrupolar or spin-nematic order which breaks SU(2) symmetry and is described by a local rank-2 tensor order parameter [38,39]. Such order is known to be realized in certain spin-1 Heisenberg models with additional biquadratic interactions [40].

4 More ingredients: Orbitals and spin-orbit coupling

While the Heisenberg model provides a useful and rich arena for quantum magnetism, the description of real materials often requires to include physics beyond. Two important ingredients are spin-orbit coupling (SOC) and orbital degrees of freedom, which we discuss in turn.

4.1 Magnetic anisotropies and novel forms of frustration

In the non-relativistic limit, SOC is not present, implying that real space and spin space are entirely separate, with SU(2) spin rotation symmetry in the absence of a magnetic field. In contrast, non-vanishing SOC couples real space and spin space, such that symmetry transformations in general act on both position and spin. Consequently, the symmetry of spin–spin interactions in a solid is lowered (compared to Heisenberg) and is dictated by the lattice structure. The simplest forms of anisotropic interactions are (i) Ising, $S_i^z S_j^z$, (ii) XY, $S_i^x S_j^x + S_i^y S_j^y$,

and (iii) Dzyaloshinskii-Moriya $\vec{D}_{ij} \cdot (\vec{S}_i \times \vec{S}_j)$. The latter is antisymmetric under exchange $i \leftrightarrows j$ and hence requires broken inversion symmetry to exist; this does not apply to the former.

In addition to anisotropic interactions, magnetic anisotropies may also arise at the single-ion level. Such single-ion anisotropies are rooted in the orbital character of the magnetic state, combined with SOC. A common single-ion term in the Hamiltonian is $D{S_i^z}^2$, where $D > 0$ ($D < 0$) corresponds to an easy-plane (easy-axis) anisotropy, respectively. On a non-Bravais lattice, different sites can have distinct *local* anisotropy axes.

Importantly, magnetic anisotropies enable forms of frustration different from that of Heisenberg models. Some of those have already been mentioned above: An easy-axis anisotropy on the pyrochlore lattice, combined with ferromagnetic interactions, leads to spin-ice behavior. The corresponding easy-plane situation results in a U(1) classical degeneracy and order by disorder.

Anisotropic interactions may even induce frustration on lattices whose geometry is unfrustrated. Such exchange frustration is the key ingredient for the route to QSLs proposed by Kitaev [14]. The Kitaev honeycomb model features bond-dependent Ising interactions,

$$\mathcal{H} = -J^x \sum_{\langle ij \rangle_x} S_i^x S_j^x - J^y \sum_{\langle ij \rangle_y} S_i^y S_j^y - J^z \sum_{\langle ij \rangle_z} S_i^z S_j^z \tag{5}$$

where the bonds of the underlying honeycomb lattice have been divided into three sets of mutually parallel bonds, labeled $\alpha = x, y, z$, and $\langle ij \rangle_\alpha$ refers to a summation over the bonds of α type. The Kitaev model has attracted tremendous attention, as it realizes an exactly solvable $\mathbb{Z}_2$ spin liquid whose emergent excitations are Majorana fermions and static $\mathbb{Z}_2$ gauge fluxes. It has been subsequently generalized to other lattices and space dimensions [41]. Experimentally, strong Kitaev interactions on the honeycomb lattice have been deduced for the materials α-$RuCl_3$ [42, 43], Na_2IrO_3 [44, 45], and various polytypes of Li_2IrO_3 [46–48]; however, all of these materials display magnetic LRO at low temperatures due to the presence of additional interactions.

4.2 Orbitals and spin-orbital liquids

The ground state of ions with partially filled shells may contain, in addition to spin degrees of freedom, also orbital degrees of freedom. The latter arise from orbital degeneracies which themselves depend on the crystalline electric field arising from the potential of the surrounding ions. For example, Cu in octahedral coordination may realize a $3d^9$ configuration, with one hole in doubly degenerate e_g orbitals, leading to $S = 1/2$ spin and $\tau = 1/2$ orbital degrees of freedom. More complicated is V in a cubic environment with a $3d^2$ configuration in triply degenerate t_{2g} orbitals, resulting in $S = 1$ (by Hund's rule) and $\tau = 1$.

Insulators with orbital degrees of freedom require to write down spin-orbital exchange models [49]. While these are typically complicated and have low symmetry, reflecting the influence of both lattice and orbital structure on exchange processes, a qualitative understanding can often be gained by simpler more symmetric models. An example is the SU(4)-symmetric Kugel-

Khomskii model,

$$\mathcal{H} = J \sum_{\langle ij \rangle} (\vec{S}_i \cdot \vec{S}_j + \frac{1}{4})(\vec{\tau}_i \cdot \vec{\tau}_j + \frac{1}{4}), \tag{6}$$

with $\vec{\tau}_i$ representing the orbital degrees of freedom. While such spin-orbital models often exhibit phases with coexisting orbital and magnetic order, it has been suggested early on that, if combined with either geometric frustration or exchange frustration, they may also produce low-temperature states devoid of symmetry breaking in both the spin and orbital sector. Such states have consequently been dubbed spin-orbital liquids [50]. Indeed, the bond-dependent interactions of the Kitaev honeycomb model can be used to construct an exactly solvable model for a spin-orbital liquid [51].

5 Conventional quantum criticality in frustrated systems

Quantum phase transitions in frustrated magnetic insulators may be conventional in the sense that they involve symmetry breaking and local order parameters. Less conventional cases involving fractionalization and topology will be postponed to the next section.

5.1 Magnetic ordering transitions

The simplest case, a quantum transition from a featureless paramagnet to a symmetry-broken phase with antiferromagnetic or VBS order, is expected to be described by an LGW theory of φ^4 type, Eq. (1), with dynamical exponent $z = 1$. Symmetry and wavevector of the order parameter determine the effective number of order-parameter components and the structure of the interaction terms in the field theory.

Frustration enters in a non-trivial way, because the order-parameter structure of non-collinear or non-coplanar states is much richer than that of simple collinear magnets. Most straightforwardly, this translates into a larger number of components N in the corresponding φ^4 theory. This is not all: For instance, a non-collinear ordered state often breaks both SU(2) spin rotation symmetry and a $\mathbb{Z}_2$ chiral symmetry, and both symmetries can be broken either in a single or in two separate transitions. For the classical case, this has been studied for stacked triangular-lattice Heisenberg antiferromagnets: Monte Carlo simulations have observed a single transition with non-trivial critical exponents, different from that of standard O(N) universality, consistent with a proposal by Kawamura [52].[4] Numerical results for the quantum case are, to our knowledge, not available due to the notorious sign problem.

More seriously, frustration can render invalid the concept of discrete well-defined wavevector for critical fluctuations: Upon approaching an ordered state, fluctuations may become soft on a manifold of wavevectors, e.g., owing to frustration-induced degeneracies. Strong fluctuation effects may then cause the transition to be first order. Alternatively, exotic novel intermediate

[4] More recent theory works predict the transition in stacked triangular-lattice Heisenberg antiferromagnets to be weakly first order [53].

phases might emerge. An interesting open problem in this context constitutes the quantum melting of a skyrmion crystal [37]. Such a phase has been observed in a number of helical magnets. One prominent material is MnSi where the lack of inversion symmetry enables Dzyaloshinskii-Moriya interactions to produce long-wavelength helical order which in turn yields a skyrmion crystal in a small window of magnetic field and temperature [54]. In MnSi, long-range magnetic order can be suppressed by the application of pressure, giving way to an extended non-Fermi liquid phase at low temperature [55]. It has been speculated that this behavior is related to partial order, e.g., a skyrmion liquid, but a concise theory is not known.

A further complication, frequently present in strongly frustrated systems, arises due to order-by-disorder physics (Sec. 3.4): If the actual ordered state is selected by fluctuation effects from a larger (e.g. classically degenerate) manifold, then some or all properties of the transition may be determined by the larger symmetry of this manifold. This type of physics is known from $\mathbb{Z}_n$ clock models, or alternatively XY models with $\mathbb{Z}_n$ anisotropy. Here, anisotropies with $n \geq n_c$ are irrelevant at criticality, such that the critical behavior is that of the XY model. For $d = 2$ (or $D = 1 + 1$) this even changes the phase diagram, as an intermediate critical phase intervenes between the disordered and the $\mathbb{Z}_n$-ordered phases for $n \geq 5$ [56]. An example of recent interest are the finite-temperature intermediate phases present in the two-dimensional (2D) Heisenberg-Kitaev model [57] where the relevant ordered phases are sixfold degenerate as a result of Kitaev interactions reflecting spin-orbit coupling [58]. Theoretical results for the *quantum* phase transitions in this model indicate first-order behavior both on analytical [59] and numerical [60, 61] grounds, but the numerics has not reached conclusive accuracy yet.

Strong frustration may, in addition, lead to *dimensional reduction*: This refers to a situation where the effective spatial dimension of the order-parameter fluctuations is smaller than that expected from the microscopic model. For instance, a three-dimensional (3D) layered system with inter-layer frustration may display 2D critical behavior. Such dimensional reduction typically does not reach down to lowest energies and temperatures, due to residual higher-dimensional couplings, such that a dimensional crossover to fully 3D critical behavior at lowest temperatures occurs [62].

5.2 Field-driven transitions and BEC phenomena

Local-moment magnets can display a variety of QPTs as function of applied magnetic field. The simplest case is the transition at the saturation field of an SU(2)-symmetric Heisenberg magnet: Upon lowering the field, a high-field magnon becomes soft at a particular wavevector, and the transition can be understood as magnon Bose-Einstein condensation (BEC) which turns the fully polarized state into a canted antiferromagnet. The latter breaks the U(1) spin rotation symmetry about the field axis and is therefore also understood as a spin superfluid. The boson condensation nature of the QPT implies that this is in the universality class of the dilute Bose gas, with $z = 2$ [3]. A similar field-driven transition occurs between the low-field singlet and intermediate-field canted phases of the coupled-dimer magnets of Sec. 2.2 [63].

While these transitions involve only trivial magnetization plateaus at $M/M_{\rm sat} = 0$ and 1, frus-

trated magnets often display intermediate magnetization plateaus. The QPTs in and out of such a magnetization plateau may be of BEC type, but are more complicated if the plateau phase spontaneously breaks lattice translation symmetry. Then, the plateau phase and the adjacent canted phase break different symmetries, possibly resulting in two continuous transitions with an intermediate coexistence (i.e. supersolid) phase or a first-order transition [64]. Experimentally, such field-induced supersolidity has been discussed for the Shastry-Sutherland compound $SrCu_2(BO_3)_2$ [65] and for the spinel $MnCr_2S_4$ [66].
Strong frustration often renders the magnon bandwidth small, paving the way for more exotic field-driven transitions. As has been discussed for a variety of frustrated Heisenberg models, it is possible that the high-field phase displays multi-magnon bound states whose minimal energy lies below that of the single-magnon branch. Then, upon lowering the field, the first instability is in this multi-magnon sector, and the resulting ordered state can be understood as a condensate of magnon bound states [67]. The most important case is that of two-magnon bound states whose condensation induces a spin-nematic state: This is a state with quadrupolar order whose order parameter is a traceless rank-2 tensor. The QPT from the high-field state is either continuous of BEC type, with $z = 2$, or is of first order due to large fluctuations.
Last not least, we note that spin-orbit coupling drastically modifies the physics described above. First, magnetization is no longer conserved, such that the fully polarized state is not an eigenstate of the Hamiltonian. As a result, the magnetization in the high-field phase is not saturated even as $T \to 0$. Second, the lower symmetry typically implies that field-driven transitions break discrete symmetries only. The corresponding QPT are then of $\mathbb{Z}_n$ type, with dynamic exponent $z = 1$. For instance, this applies to the Kitaev material α-$RuCl_3$: Although its field-induced phases are not fully understood to date, it is clear that the magnetization in the asymptotic high-field phase receives substantial quantum corrections [68], and it can be expected that the QPT to the asymptotic high-field phase is either of Ising type or of first order.

6 Transitions involving topological states

Phase transitions in and out of topologically non-trivial states (more precisely, states with intrinsic topological order) can in general not be captured by LGW theory, as topology is associated with global instead of local properties. Nevertheless, topological states and their transitions can often be described by local quantum field theories which then involve novel emergent degrees of freedom coupled to gauge fields. We will discuss a few of such transitions in turn.

6.1 Confinement transitions and fractionalized criticality

QPTs in and out of topological liquid states are fundamentally different from the conventional transitions, as they necessarily involve the fractionalized degrees of freedom of the (spin) liquid. In many cases, these are spinons (i.e. fractionalized constituents of the microscopic spins) and excitations of the emergent gauge field in its deconfined phase. Continuous transitions out of a spin liquid can often be understood as a condensation transition of one of these particles (or

bound states thereof) [11, 69]. Physical spins are then composite objects in terms of the critical degrees of freedom. As a result, spin correlation functions display critical power laws with *large* anomalous exponents: While standard O(N) universality yields numerically small anomalous exponents, e.g. $\eta = 0.06$ for the 3D Heisenberg model, many of the exotic transitions discussed below have η values for physical correlators of order unity.

Starting from a fractionalized spin liquid, one can envision the following options for QPTs: (i) a confinement transition to a featureless paramagnet, (ii) a confinement transition with concomitant symmetry breaking, leading to e.g. magnetic or VBS order – typically these are Higgs-type transitions driven by the condensation of a particle with gauge charge, (iii) a condensation transition which leaves the deconfinement intact, which then leads to exotic fractionalized magnetic (AF*) or VBS states (VBS*), (iv) a transition to a different fractionalized spin liquid.

In the following, we list a few examples from the theory literature. The field theories are typically written down in terms of fractionalized particles coupled to gauge fields (a simple example being the CP^1 model (7) specified below); in some cases topological quantum field theories (most importantly, Chern-Simons theories) have also proven useful. Most considerations apply to two space dimensions; less work has been done for $d = 3$.

Transitions in group (i) require the presence of a featureless paramagnetic phase in addition to a topological spin liquid: The former can be realized, e.g., by application of a magnetic field or by the formation of singlet dimers as in bilayer models. A concrete example is the 2D toric-code model [70] in a longitudinal field [71]: It displays a continuous transition from a $\mathbb{Z}_2$ topological spin liquid to a featureless high-field phase. The transition has been shown to be in the Ising* universality class in $D = 2 + 1$ dimensions [72]. Here, Ising* refers the fact that the critical degrees of freedom have Ising symmetry, but are very different from a conventional order parameter, as they derive from the fractionalized excitations of the spin liquid. Hence, thermodynamic properties are that of Ising criticality in $D = 2 + 1$, but correlation functions of physical spins strongly differ from the conventional case as spins are composite objects here. This can be expected to generically apply to confinement transitions of $\mathbb{Z}_2$ spin liquids. A second example is the ferromagnetic honeycomb-lattice Kitaev model in a magnetic field [73, 74]: This displays a single transition between a $\mathbb{Z}_2$ spin liquid and a featureless high-field phase as well, Fig. 6. However, it is open whether this transition is weakly first order or continuous.

Transitions in group (ii) have been mainly discussed within effective field theories, and candidate models are known in many cases. A typical situation is that of vison condensation in a 2D $\mathbb{Z}_2$ spin liquid; if the vison has non-trivial transformation properties under lattice symmetries, its condensation generically breaks translation symmetry and induces VBS order. Such transitions have been argued to be of O(N)* type (where * again refers to the fact that the primary fields are fractionalized) – supplemented by lattice anisotropy terms which are irrelevant at criticality – where the number of components N of the vison-derived field depends on the lattice and the resulting VBS state. For example, the transition to a columnar VBS on both the square and honeycomb lattices is of 3D XY* type [25, 75, 76], while on the triangular lattice the transition to a columnar VBS is proposed to be of 3D O(6)* type [77]. In contrast, transitions to staggered VBS

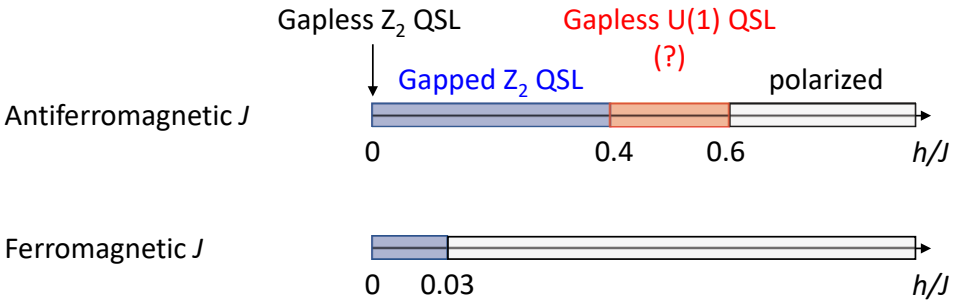


Fig. 6: *Schematic phase diagram of the isotropic Kitaev honeycomb model in a magnetic field along [111], i.e., with $h_x = h_y = h_z$. For antiferromagnetic coupling (top) the $\mathbb{Z}_2$ spin liquid is rather robust, and an intermediate second QSL emerges which is possibly of U(1) character. In contrast, for ferromagnetic coupling (bottom) a small field destroys the $\mathbb{Z}_2$ spin liquid in favor of a polarized phase [74].*

phases have been argued to be of first order [76]. Generally, liquid–VBS transitions may be realized in Heisenberg models with further-neighbor (e.g. J_1-J_2-J_3) exchange interactions. Instead of condensing visons one can consider condensing spinons in SU(2)-symmetric $\mathbb{Z}_2$ spin liquids. This produces a confined antiferromagnet with spiral order via an O(4)* transition [75, 78, 79] where the symmetry arises from a doublet of complex spinon fields. A resulting "global" phase diagram is shown in Fig. 7. Finally, condensing bound states of spinons and visons may induce conventional two-sublattice Néel order. At the latter transition, which is of more exotic type, both magnetic and VBS correlation functions acquire critical power laws [79]. In the absence of SU(2) symmetry, quantum numbers need to be reconsidered, but the general picture remains valid. One example here is the 2D toric-code model perturbed by an Ising interaction which has been shown to display a continuous transition of Ising* type from a $\mathbb{Z}_2$ liquid to a ferromagnetic phase driven by defect condensation [80]. A second example is the transition between a $\mathbb{Z}_2$ spin liquid and a superfluid phase in a Kagome-lattice XY model. This transition is in the XY* universality class and has been studied numerically in some detail in Ref. [81].

A transition in group (iii) is realized upon condensing objects which do *not* carry gauge charge, then leading to the coexistence of symmetry-breaking order and fractionalization. Hence, the transition involves the onset of symmetry breaking on the background of a fractionalized topological state – this has also been dubbed fractionalized quantum criticality. For instance, condensing a gauge-neutral Néel vector in a spin liquid yields an AF* phase, and a spin-Peierls instability of a spin liquid can result in a VBS* state. A nice example of the former has been proposed to occur in certain Kitaev-based spin-orbital liquids [82], while an example of the latter is the instability of Majorana Fermi surfaces in 3D Kitaev-based spin liquids [83].

Transitions between different spin-liquid phases, group (iv), have also been considered on the level of effective field theories. Ref. [84] has developed a theory for transitions between chiral and $\mathbb{Z}_2$ spin liquids in two space dimensions; such transitions have been argued to be equivalent to the condensation of an XY field coupled to a U(1) gauge field, where the critical XY field represents a singlet combination of spinons. A second case is the transition from a U(1) to a

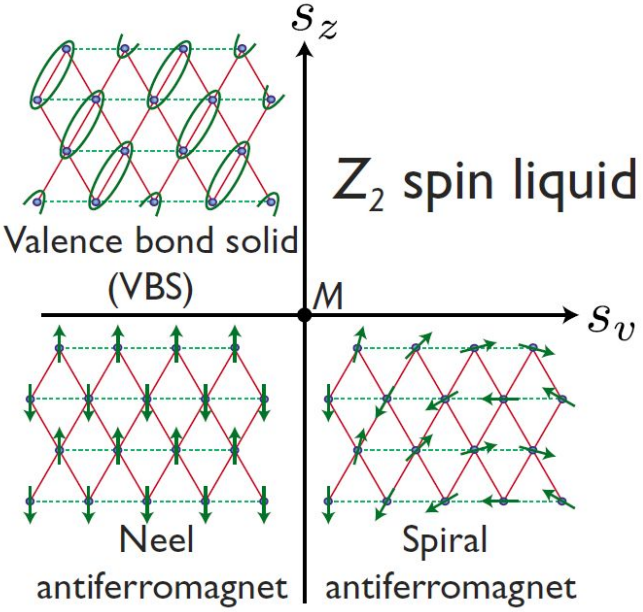


Fig. 7: *Global phase diagram for a 2D model of spinons with emergent $\mathbb{Z}_2$ gauge field, here shown for an anisotropic triangular lattice. The two parameters s_v and s_z represent masses of visons and spinons, respectively, in the doubled Chern-Simons theory considered in Ref. [75]. The spiral–$\mathbb{Z}_2$ spin liquid transition is described by a three-dimensional O(4)* theory, while the transition from VBS to $\mathbb{Z}_2$ spin liquid is of XY* type, see text. Further, the Néel–VBS transition is captured by a CP^1 theory (see Sec. 6.2 below), and the Néel–spiral transition is mean-field-like. (Figure taken from Ref. [75])*

$\mathbb{Z}_2$ spin liquid which is driven by the condensation of *pairs* of gauge-charged particles, akin to superconducting pairing. Such a transition has in fact been suggested to occur in the antiferromagnetic honeycomb Kitaev model in an applied magnetic field, Fig. 6: The small-field gapped $\mathbb{Z}_2$ spin liquid transits into a different spin liquid, suggested to be gapless and of U(1) type, before reaching the high-field phase [74, 85].

Among the few experimental examples of spin-liquid-related QPTs are field-driven transitions in suitable candidate materials, most notably in α-$RuCl_3$ and $NaYbX_2$ (X = S, Se, O). In α-$RuCl_3$ an intermediate-field spin liquid has been suggested to transit into the asymptotic high-field phase, with the transition being of first order [86]. In $NaYbX_2$, the zero-field spin liquid gives way to field-induced ordered states, but the nature of the quantum transition has not been probed in detail [34].

6.2 Deconfined quantum criticality

An interesting scenario for unconventional transitions between symmetry-broken states is that of deconfined quantum criticality [87]. It describes the possibility of a direct generic continuous QPT between two ordered states which break different symmetries. According to Landau theory and without fine-tuning, such a transition is forbidden, as it would be either of first order or split into two continuous transitions. At a deconfined quantum critical point, the critical degrees of

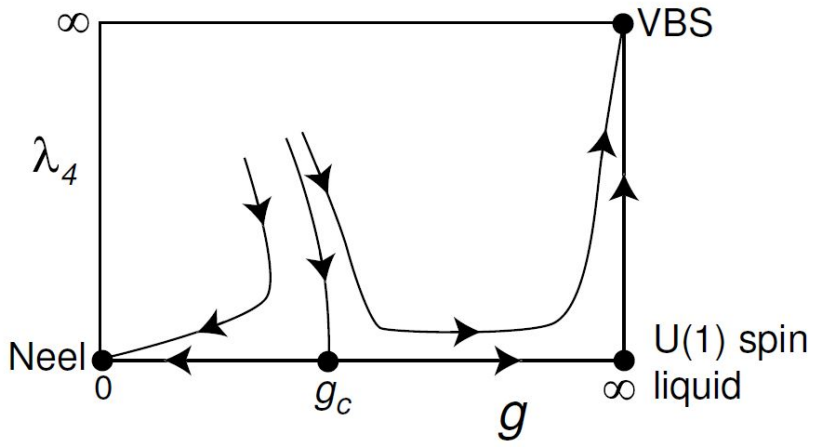


Fig. 8: *Schematic renormalization group flow proposed for the transition between a Néel antiferromagnet and a VBS in an SU(2)-symmetric magnet, as realized e.g., by the square-lattice spin-*$1/2$ *J-Q model. Increasing g destabilizes magnetic order; the parameter λ_4 represents the fugacity of monopoles in the U(1) gauge field. The horizontal axis $\lambda_4 = 0$ corresponds to the non-compact CP^1 theory of Eq.* (7)*, with g to be identified with the spinon mass s. (Figure taken from Ref. [89])*

freedom are fractionalized particles, and the order parameters of both phases are composites of these particles. This automatically leads to large anomalous exponents for order-parameter correlations.

The most thoroughly studied instance of deconfined quantum criticality is the transition between a Néel-ordered antiferromagnet and a valence-bond solid on the square lattice. The proposed field theory employs a CP^1 representation of spins, with deconfined bosonic spinons z_α and a U(1) gauge field A_μ, resulting in the action

$$\mathcal{S} = \int d^d x d\tau \left[\left| (\partial_\mu - i A_\mu) z_\alpha \right|^2 + s |z_\alpha|^2 + \frac{u}{2} \left(|z_\alpha|^2 \right)^2 + \frac{1}{2e^2} \left(\epsilon_{\mu\nu\lambda} \partial_\nu A_\lambda \right)^2 \right] \tag{7}$$

with the last term encoding the gauge-field dynamics. The primary transition, accessed by the variation of the mass parameter s, is that between a U(1) spin liquid and a Néel antiferromagnet. It is driven by the condensation of the z spinons which induces confinement via a Higgs mechanism; at this transition the gauge field can be assumed to be non-compact, as in the continuum limit of Eq. (7). However, the U(1) gauge field microscopically emerges from a complex-phase degree of freedom of the spinons and is therefore compact. This implies the existence of monopoles, and their condensation renders the U(1) spin liquid unstable towards a dimerized confined VBS phase, Fig. 8. Hence, deconfined spinons exist only at criticality [87–89].

The above proposal has been tested in detailed numerical simulations of the so-called J-Q model on the square lattice, where Q denotes the strength of a ring-exchange term [90]. While these simulations have verified a large part of the phenomenology of deconfined quantum criticality [90, 91], they have also found evidence for large logarithmic corrections to scaling which are not predicted by the field-theoretical framework [92]. We also note that direct numerical simulations of the proposed CP^1 field theory have found indications for the transition being weakly first order [93], a tendency which could not be confirmed in the J-Q model simulations.

The reasons for these discrepancies in numerical results are open, see Ref. [94] for a discussion.

Recent developments in the context of field-theoretical dualities have led to additional insights [95]. It has been conjectured that the non-compact CP^1 model is dual to a so-called QED_3 Gross-Neveu model at criticality, the latter describing Dirac fermions coupled to both a U(1) gauge field and local Ising degrees of freedom. This duality suggests that the deconfined QCP between a Néel antiferromagnet and a VBS displays an emergent SO(5) symmetry, which is supported by numerical results [94]. A weak first-order transition with quasi-universal behavior in its vicinity appears as a plausible scenario [96].

In addition to the Néel-VBS transition, various other Landau-forbidden transitions between two differently ordered phases have been discussed in the context of deconfined criticality. For instance, the transitions between a $\mathbb{Z}_2$ spin liquid and a VBS discussed in Sec. 6.1, as well as a transition between a $\mathbb{Z}_2$ spin liquid and a Néel state, also belong to this class, as a $\mathbb{Z}_2$ spin liquid displays topological order. Emergent higher symmetries, which can be rationalized via suitable dualities, appear to be common to many of the deconfined critical points [95].

Although a clear-cut experimental example realizing deconfined quantum criticality is lacking, a recently identified candidate is the frustrated Shastry-Sutherland magnet $SrCu_2(BO_3)_2$. Under applied pressure, it displays a transition from a plaquette VBS to an antiferromagnetic phase [97] which has been argued to be a deconfined QPT [98].

7 Mott and Kondo transitions

While all material presented so far was devoted to Mott insulators with local moments, we now turn to QPTs involving metallic phases [4]. For metals, the concepts of symmetry breaking and local order parameters apply equally, hence symmetry-breaking QPTs can be defined and characterized in analogy to insulators. However, the presence of low-energy particle–hole excitations and their coupling to order-parameter fluctuations complicates the theoretical analysis: Following the spirit of LGW theory requires to integrate out the particle–hole excitations to arrive at a theory for the order parameter alone; this approach has been developed in detail in the works of Hertz [2], Millis [99], and Moriya [100]. However, it was later realized that such an LGW theory is plagued with singularities. Consequently, more refined approaches keeping both order-parameter and fermionic fluctuations are required, and some progress has been made [101–104].

In this section, we will exclusively deal with even more intricate types of QPT, namely those involving the onset or loss of metallicity. Historically, the interaction-driven Mott transition has been discussed extensively. A younger topic is that of partial Mott transitions in multi-band or multi-orbital systems, with a subclass being transition where the Kondo effect breaks down. We will discuss these transitions – together with their relation to frustration – below.

7.1 Fermi liquids and non-Fermi liquids

Before diving into the physics of Mott transitions, we need to review some aspects of the low-energy physics of metals. The key concept is that of a Fermi liquid, which asserts a one-to-one correspondence of the low-energy many-body states between the interacting system under consideration and a hypothetical system of non-interacting electrons. This implies in particular the existence of quasiparticle excitations with charge $\pm e$ and spin $1/2$ (and forbids the existence of other low-energy excitations!). It also implies the existence of a Fermi surface, defined by the location of jumps in the momentum distribution $\langle n_{k\sigma}\rangle$ (or, equivalently, poles of the single-particle spectral function at $\omega = 0$). This Fermi surface then obeys Luttinger's theorem, i.e., has a momentum-space volume given by the total density of electrons n_{tot} (modulo filled bands):

$$\mathcal{V}_{\text{FL}} = K_d(n_{\text{tot}} \bmod 2) \tag{8}$$

where factors of 2 account for spin degeneracy, i.e., a full band corresponds to $n = 2$, and $K_d = (2\pi)^d/(2V_0)$ where V_0 is the unit-cell volume [105]. Under these conditions, the standard low-temperature Fermi-liquid properties $C(T) = \gamma T,\ \rho(T) = \rho_0 + AT^2$ etc., with γ, A being constants, follow immediately.[5]

Violations of Fermi-liquid behavior at low temperature, generically dubbed non-Fermi liquid, can have various sources. In clean systems, interaction effects can produce stable non-Fermi-liquid phases. One scenario is that the low-energy excitations display quantum numbers different that of from electron or holes, leading to distinct low-temperature properties. While such behavior is generic and well understood in $d = 1$, resulting in Luttinger liquids with spin-charge separation, similarly controlled descriptions in higher dimensions are scarce. A viable route to spin-charge-fractionalized metals is the doping of spin liquids [106].

Another scenario for stable non-Fermi liquids in $d \geq 2$ has been termed fractionalized Fermi liquid [107, 108]. In such a phase, charged excitations have conventional quantum numbers (charge $\pm e$ and spin $1/2$), but these coexist with additional deconfined fractionalized degrees of freedom. A generic construction starts from a fractionalized spin liquid and adds conventional carriers in a second band. If these subsystems remain weakly coupled, they realize a FL* phase (which has also been characterized as metallic spin liquid in the literature). Importantly, such a phase displays a Fermi surface with a volume violating Luttinger's theorem (8) in a quantized fashion, often [107]

$$\mathcal{V}_{\text{FL}^*} = K_d((n_{\text{tot}}-1) \bmod 2) \tag{9}$$

where the -1 accounts for the electrons forming the spin-liquid component. Low-temperature properties may or may not be Fermi-liquid-like, depending on whether the emergent excitations of the spin-liquid component are gapped or gapless. Fractionalized Fermi liquids may display a variety of instabilities driven by the strong correlations in the local-moment sector, including unconventional superconductivity [107, 109].

[5] A T^2 behavior of the resistivity requires the existence of Umklapp scattering processes, i.e., a sufficiently large Fermi surface.

Importantly, fractionalized Fermi liquids as well as other doped spin liquids are symmetric states (i.e. without spontaneously broken symmetries) with fractionalized excitations, much like insulating spin liquids. Given the insights into topological properties of fractionalized insulating phases, one may wonder about the topological characterization of non-Fermi-liquid metals. To our knowledge, relatively little work has been done in this direction. A sharp distinction between FL and FL* is the Fermi volume, and this can be considered a topological distinction. In contrast, some of the indicators established for insulators, like ground-state degeneracies and entanglement, cannot be easily applied because of the absence of an excitation gap [109], and more work is needed to clarify the topological nature of non-Fermi liquid metals.

7.2 Mott transitions

A Mott transition is an interaction-driven metal-to-insulator transition: It transforms a half-filled metallic band into an insulator of local moments. The most generic Hamiltonian for this physics is the Hubbard model of spinful electrons

$$\mathcal{H} = -t \sum_{\langle ij \rangle \sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + h.c. \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} \tag{10}$$

where the Mott insulator occurs for $U \gg t$. The Mott-insulating state is often accompanied by antiferromagnetic long-range order, and the quantum transition from a paramagnetic metal to an antiferromagnetic Mott insulator is generically of first order (or involves an intermediate antiferromagnetic metallic phase). This is different in the case of a spin-liquid Mott insulator: A "genuine" zero-temperature Mott transition from a paramagnetic metal to an insulating spin liquid can be continuous. As the existence of the spin liquid requires frustration, such transitions are expected to occur in half-filled Hubbard models on frustrated lattices upon varying U/t. In fact, a metal-to-spin liquid transition has been found in numerical simulations of the triangular-lattice Hubbard model which, however, appears to be first order [110, 111], Fig. 9, with superconductivity possibly appearing on the metallic side before the Mott transition [111]. A candidate experimental realization is in the organic compound κ-$(ET)_2Cu_2(CN)_3$ under pressure [112].

A defining criterion for a Mott transition is a quantized change in the Fermi volume: In a Fermi liquid, the momentum-space volume enclosed by the Fermi surface is given by the total number of electrons according to Luttinger's theorem (8). In a Mott insulator, there is no Fermi surface[6], and hence the Fermi volume changes at a single-band Mott transition by $K_d \times 1$. Such an abrupt change is nevertheless compatible with the QPT being continuous: Upon approaching a continuous Mott transition from the metallic side, the quasiparticle weight on the Fermi surface will vanish continuously, while the charge gap opens continuously on the insulating side. At criticality, one expects a *critical Fermi surface*, i.e., a well-defined $(d-1)$-dimensional mani-

[6] We do not consider the so-called Luttinger volume, $V_{\rm lutt} = \int_{G(k)>0} dk$, which accounts for both poles and zeroes of the Green's function. For an in-depth discussion on aspects of the Luttinger volume in Mott insulators see Ref. [113].

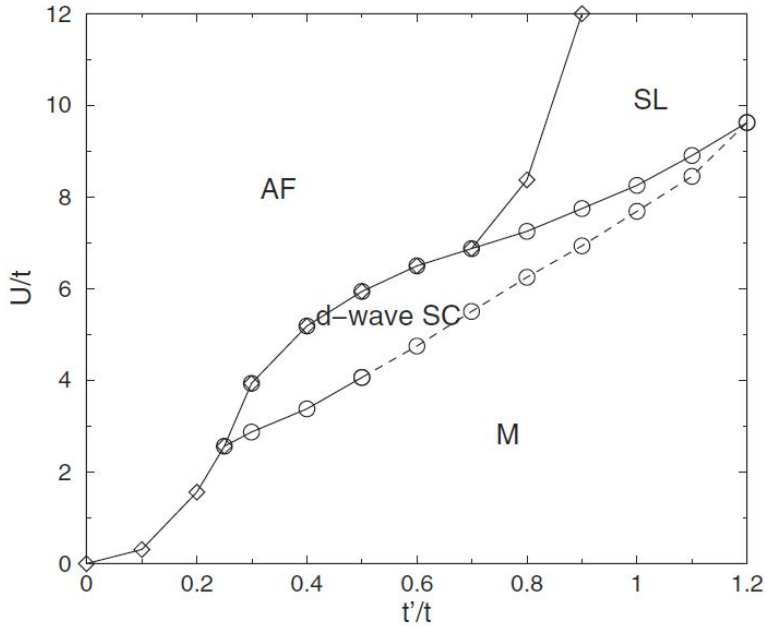


Fig. 9: *Cluster-DMFT phase diagram of the Hubbard model on an anisotropic triangular lattice as function of Hubbard interaction U and hopping ratio t'/t, where $t' = 0$ and $t' = t$ correspond to the square and triangular lattices, respectively. M, SC, AF, SL denote metal, superconductor, antiferromagnetic insulator, and spin-liquid phases, respectively. Solid (dashed) lines correspond to first-order (continuous) transitions. (Figure taken from Ref. [111])*

fold in momentum space where the electronic spectral function displays (possibly momentum dependent) power-law singularities [114].

A concise theoretical understanding of continuous zero-temperature Mott transitions is lacking to date. Most theoretical descriptions are based on slave-particle theories which involve separate degrees of freedom representing spin and charge of the electrons. Often, the charge degrees of freedom are encoded by bosons which are gapless and condensed in the metal, but gapped and disordered in the insulator. Hence, the insulator-to-metal transition becomes a BEC transition of charged bosons coupled to a gauge field [115]. However, such a description (at least in its simplest version) does not account for possible non-trivial momentum dependencies along the Fermi surface. Moreover, the fermionic character of the Mott phenomenon might require a formulation using non-bosonic critical degrees of freedom, but to our knowledge a successful theory of this type has not been formulated.

It is worth noting that apparent quantum critical behavior at elevated temperatures has been detected above the finite-temperature endpoint of a first-order Mott transition line. This remarkable observation, manifest, e.g., in scaling behavior of the resistivity, was first made in DMFT simulations of the single-band Hubbard model on a Bethe lattice [116], and later verified experimentally in three pressure-tuned organic compounds [117]. Subsequent work has linked this behavior to a $T = 0$ scale-invariant quantum critical insulator at the boundary of the metal–insulator phase coexistence regime [118].

7.3 Kondo and orbital-selective Mott transitions

In a multi-band or multi-orbital system, there is the possibility for a partial Mott transition. This is a transition between two *metallic* phases where the Fermi surface undergoes a quantized change. In the simplest case, one band (or orbital) changes its character from metallic to Mott-insulating while other bands remain metallic. Consequently, such a transition has also been dubbed orbital-selective Mott transition [119–121]. If stable in the low-temperature limit, the partial Mott phase violates Luttinger's theorem (8) and, hence, is a non-Fermi liquid metal. This is precisely the fractionalized Fermi-liquid phase (FL*) introduced in Sec. 7.1 above, and a transition between FL and FL* is an orbital-selective Mott transition (or a deconfinement transition in the language of the underlying gauge theory). Phenomenologically, such a transition can be expected to be accompanied by a jump in the Hall constant [122].

A natural territory for orbital-selective Mott physics are heavy-fermion metals [4, 123], as described by the Kondo-lattice Hamiltonian

$$\mathcal{H} = -t \sum_{\langle ij\rangle\sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + h.c. \right) + J \sum_i \vec{S}_i \cdot \vec{s}_i, \tag{11}$$

with J the Kondo coupling and $\vec{s}_i = \sum_{\sigma\sigma'} c^\dagger_{i\sigma} \vec{\tau}_{\sigma\sigma'} c_{i\sigma'}/2$ the conduction-electron spin density on site i. As pointed out early on by Doniach [124], the heavy-fermion phase diagram is governed by the competition between Kondo screening and RKKY interactions between local moments, leading to heavy Fermi liquids and ordered magnetic states, respectively. Later on, it has been suggested [108, 125–127] to consider, in addition to the ratio between Kondo temperature and RKKY interaction, a second tuning parameter which acts to suppress magnetic order in the local-moment subsystem – this is loosely labelled as "frustration" (alternatively: "quantum fluctuations"). This tuning parameter naturally enables access to fractionalized states. If RKKY interactions are sufficiently frustrated, then increasing them w.r.t. the Kondo scale leads to a breakdown of the Kondo effect without concomitant magnetic order, generically resulting in an FL* phase.

The resulting "global" phase diagram of heavy fermions is shown in Fig. 10. It features two transition lines, one involving the onset of antiferromagnetism and one involving the onset of deconfinement. Importantly, the onset of deconfinement in the paramagnetic metallic phase corresponds to an orbital-selective Mott transition into an FL* phase as advocated above, as FL* features deconfined fractionalized excitations in the local-moment sector. Such an orbital-selective Mott transition is easily driven by the reduction of Kondo screening in a frustrated regime, because it is Kondo screening which renders the local-moment electrons metallic. Hence, the onset of deconfinement also corresponds to a breakdown of the Kondo effect. The two transition lines define four phases: In addition to the paramagnetic phases FL and FL*, there are a conventional (AF) and a fractionalized (AF*) antiferromagnet. The Fermi volume is "large" in the FL phase, i.e., encloses both conduction and local-moment electrons, while it is "small" in FL* because it is determined by conduction electrons alone, hence violating Luttinger's theorem. In the metallic AF and AF* phase, translation symmetry breaking enlarges the unit cell,

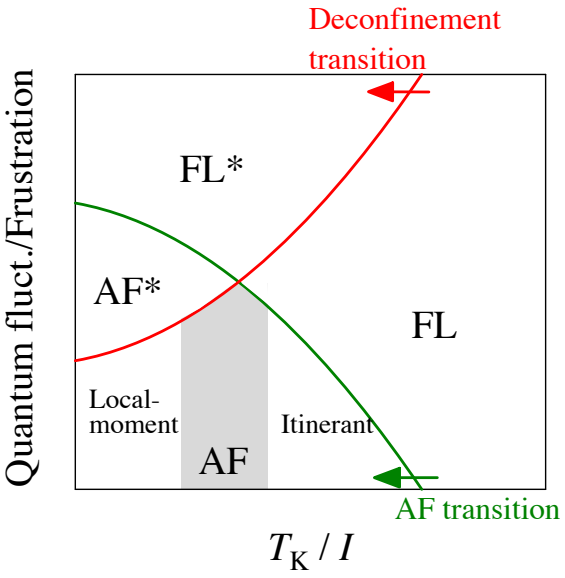


Fig. 10: *"Global" phase diagram for heavy-fermion metals (with one f electron per crystallographic unit cell), with two transitions for the onset of antiferromagnetism and for the breakdown of the Kondo effect (equivalently the onset of deconfinement). FL* is the fractionalized Fermi-liquid phase described in Sec. 7.1. Inside the AF phase, a crossover from more itinerant to more localized behavior occurs, which may be accompanied by one or more transitions where the Fermi-surface topology changes. Lastly, AF* refers to a fractionalized magnet, with magnetic LRO and fractionalized excitations coexisting. (Figure taken from Ref. [127])*

such that Luttinger's theorem is generically fulfilled. The transition from FL to AF is hence a conventional ordering transition, accompanied by the backfolding of bands.
A slightly different version of the global phase diagram has been put forward in Ref. [125], the main difference being that the coincidence of the Kondo-breakdown and magnetic transition lines is not considered accidental, but systematic. Ref. [128] has developed a corresponding extended DMFT description of a Kondo breakdown driven by magnetic criticality. Alternatively, this might be viewed as a case of deconfined criticality [129].

8 Summary

Frustrated magnetism and quantum criticality both constitute highly active fields of research in condensed matter physics, and both have received additional fuel in the last two decades by the improved understanding of topological phenomena in solids. This chapter aimed at an overview of the interplay of both, frustration and quantum criticality, with focus on theoretical ideas and concepts as well as links to current experiments in correlated-electron materials. While quantum criticality in clean insulators is mainly well understood, frustration brings in new ingredients – large degeneracies, order by disorder, and fractionalization – which often change the rules of the game, and we have discussed a few particularly fascinating outcomes. In metallic systems, the physics of quantum phase transitions is more complicated in general, due to the presence of low-energy fermions, with many open questions even without frustration. Clearly, this fascinating field invites more work, both theoretical and experimental.

References

[1] This chapter heavily draws from the review articles [5] and [6].

[2] J.A. Hertz, Phys. Rev. B **14**, 1165 (1976)

[3] S. Sachdev: *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, 2010)

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

[5] M. Vojta, Rep. Prog. Phys. **81**, 064501 (2018)

[6] M. Vojta, Rep. Prog. Phys. **66**, 2069 (2003)

[7] C. Rüegg *et al.*, Nature **423**, 62 (2003); Phys. Rev. Lett. **93**, 257201 (2004)

[8] A.P. Ramirez, Ann. Rev. Mat. Sci. **24**, 453 (1994)

[9] L. Balents, Nature **464**, 199 (2010)

[10] O.A. Starykh, Rep. Prog. Phys. **78**, 052502 (2015)

[11] L. Savary and L. Balents, Rep. Prog. Phys. **80**, 016502 (2017)

[12] Y. Zhou, K. Kanoda, and T.-K. Ng, Rev. Mod. Phys. **89**, 025003 (2017)

[13] C. Lacroix, P. Mendels, and F. Mila (Eds.): *Introduction to Frustrated Magnetism* (Springer, Heidelberg, 2011)

[14] A. Kitaev, Ann. Phys. (N.Y.) **321**, 2 (2006)

[15] S.T. Bramwell and M.J.P. Gingras, Science **294**, 1495 (2001)

[16] G.H. Wannier, Phys. Rev. **79**, 357 (1950)

[17] L. Pauling, J. Am. Chem. Soc. **57**, 2680 (1935)

[18] C. Castelnovo, R. Moessner, and S. L Sondhi, Nature **451**, 42 (2008)

[19] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006)

[20] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006)

[21] X.G. Wen, Phys. Rev. B **65**, 165113 (2002)

[22] E.H. Lieb, T.D. Schultz, and D.C. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961)

[23] M.B. Hastings, Phys. Rev. B **69**, 104431 (2004)

[24] P.W. Anderson, Mater. Res. Bull. **8**, 153 (1973)

[25] R. Moessner and S.L. Sondhi, Phys. Rev. Lett. **86**, 1881 (2001)

[26] S. Yan, D.A. Huse, and S.R. White, Science **332**, 1173 (2011)

[27] Y. Iqbal, F. Becca, S. Sorella, and D. Poilblanc, Phys. Rev. B **87**, 060405(R) (2013)

[28] Y.-C. He, M.P. Zaletel, M. Oshikawa, and F. Pollmann, Phys. Rev. X **7**, 031020 (2017)

[29] J.S. Helton *et al.*, Phys. Rev. Lett. **98**, 107204 (2007)

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

[31] H.-C. Jiang, H. Yao, and L. Balents, Phys. Rev. B **86**, 024424 (2012)

[32] Z. Zhu and S.R. White, Phys. Rev. B **92**, 041105(R) (2015)

[33] M. Baenitz *et al.*, Phys. Rev. B **98**, 220409(R) (2018)

[34] M.M. Bordelon *et al.*, Nat. Phys. **15**, 1058 (2019)

[35] J. Villain, R. Bidaux, J.-P. Carton, and R. Conte, J. Phys. France **41**, 1263 (1980)

[36] M.E. Zhitomirsky, M.V. Gvozdikova, P.C.W. Holdsworth, and R. Moessner, Phys. Rev. Lett. **109**, 077204 (2012)

[37] U.K. Rößler, A.N. Bogdanov, and C. Pfleiderer, Nature **442**, 797 (2006)

[38] M. Blume and Y. Hsieh, J. Appl. Phys. **40**, 1249 (1969)

[39] H. Chen and P. Levy, Phys. Rev. Lett. **27**, 1383 (1971)

[40] A. Läuchli, F. Mila, and K. Penc, Phys. Rev. Lett. **97**, 087205 (2006); *ibid.* **97**, 229901 (2006)

[41] K. O'Brien, M. Hermanns, and S. Trebst, Phys. Rev. B **93**, 085101 (2016)

[42] K.W. Plumb *et al.*, Phys. Rev. B **90**, 041112 (2014)

[43] A. Banerjee *et al.*, Science **356**, 1055 (2017)

[44] Y. Singh and P. Gegenwart, Phys. Rev. B **82**, 064412 (2010)

[45] S.H. Chun *et al.*, Nat. Phys. **11**, 462 (2015)

[46] Y. Singh, S. Manni, J. Reuther, T. Berlijn, R. Thomale, W. Ku, S. Trebst, and P. Gegenwart, Phys. Rev. Lett. **108**, 127203 (2012)

[47] T. Takayama *et al.*, Phys. Rev. Lett. **114**, 077202 (2015)

[48] K.A. Modic *et al.*, Nat. Commun. **5**, 4203 (2018)

[49] K.I. Kugel and D.I. Khomskii, Sov. Phys. Usp. **25**, 231 (1982)

[50] G. Khaliullin and S. Maekawa, Phys. Rev. Lett. **85**, 3950 (2000)

[51] H. Yao and D.-H. Lee, Phys. Rev. Lett. **107**, 087205 (2011)

[52] H. Kawamura, J. Phys. Soc. Jpn. **54**, 3220 (1985); **56**, 474(1987);
Phys. Rev. B **38**, 4916 (1988)

[53] M. Tissier, B. Delamotte, and D. Mouhanna, Phys. Rev. Lett. **84**, 5208 (2000)

[54] S. Mühlbauer *et al.*, Science **323**, 915 (2009)

[55] C. Pfleiderer, P. Böni, T. Keller, U.K. Rössler, and A. Rosch, Science **316**, 1871 (2007)

[56] J.V. José, L.P. Kadanoff, S. Kirkpatrick, and D.R. Nelson, Phys. Rev. B **16**, 1217 (1977)

[57] J. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **105**, 027204 (2010)

[58] C.C. Price and N.B. Perkins, Phys. Rev. B **88**, 024410 (2013)

[59] R. Schaffer, S. Bhattacharjee, and Y.-B. Kim, Phys. Rev. B **86**, 224417 (2012)

[60] J. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **110**, 097204 (2013)

[61] M. Gohlke, R. Verresen, R. Moessner, and F. Pollmann,
Phys. Rev. Lett. **119**, 157203 (2017)

[62] M. Garst, L. Fritz, A. Rosch, and M. Vojta, Phys. Rev. B **78**, 235118 (2008)

[63] T. Giamarchi, C. Rüegg, and O. Tchernyshyov, Nat. Phys. **4**, 198 (2008)

[64] N. Laflorencie and F. Mila, Phys. Rev. Lett. **99**, 027202 (2007)

[65] Y.H. Matsuda *et al.*, Phys. Rev. Lett. **111**, 137204 (2013)

[66] V. Tsurkan *et al.*, Sci. Adv. **3**, e1601982 (2017)

[67] N. Shannon, T. Momoi, and P. Sindzingre, Phys. Rev. Lett. **96**, 027213 (2006)

[68] L. Janssen, E.C. Andrade, and M. Vojta, Phys. Rev. B **96**, 064430 (2017)

[69] F.A. Bais and J.K. Slingerland, Phys. Rev. B **79**, 045316 (2009)

[70] A.Y. Kitaev, Ann. Phys. (N.Y.) **303**, 2 (2003)

[71] J. Vidal, S. Dusuel, and K.P. Schmidt, Phys. Rev. B **79**, 033109 (2009)

[72] M. Schuler, S. Whitsitt, L.-P. Henry, S. Sachdev, and A.M. Läuchli,
Phys. Rev. Lett. **117**, 210401 (2016)

[73] H.-C. Jiang, Z.-C. Gu, X.-L. Qi, and S. Trebst, Phys. Rev. B **83**, 245104 (2011)

[74] C. Hickey and S. Trebst, Nat. Commun. **10**, 530 (2019)

[75] C. Xu and S. Sachdev, Phys. Rev. B **79**, 064405 (2009)

[76] C. Xu and L. Balents, Phys. Rev. B **84**, 014402 (2011)

[77] K. Slagle and C. Xu, Phys. Rev. B **89**, 104418 (2014)

[78] A.V. Chubukov, T. Senthil, and S. Sachdev, Phys. Rev. Lett. **72**, 2089 (1994); A.V. Chubukov, S. Sachdev, and T. Senthil, Nucl. Phys. B **426**, 601 (1994)

[79] E.-G. Moon and C. Xu, Phys. Rev. B **86**, 214414 (2012)

[80] Y. Kamiya, Y. Kato, J. Nasu, and Y. Motome, Phys. Rev. B **92**, 100403(R) (2015); Phys. Rev. B **93**, 179902 (2016)

[81] S.V. Isakov, R.G. Melko, and M.B. Hastings, Science **335**, 193 (2012)

[82] U.F.P. Seifert, X.-Y. Dong, S. Chulliparambil, M. Vojta, H.-H. Tu, and L. Janssen, Phys. Rev. Lett. **125**, 257202 (2020)

[83] M. Hermanns, S. Trebst, and A. Rosch, Phys. Rev. Lett. **115**, 177205.

[84] M. Barkeshli, arXiv:1307.8194.

[85] N.D. Patel and N. Trivedi, PNAS **116**, 12199 (2019)

[86] S. Suetsugu *et al.*, J. Phys. Soc. Jpn. **91**, 124703 (2022)

[87] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, and M.P.A. Fisher, Science **303**, 1490 (2004)

[88] T. Senthil, L. Balents, S. Sachdev, A. Vishwanath, and M.P.A. Fisher, Phys. Rev. B **70**, 144407 (2004)

[89] T. Senthil, L. Balents, S. Sachdev, A. Vishwanath, and M.P.A. Fisher, J. Phys. Soc. Jpn. **74** Suppl. 1 (2005)

[90] A. Sandvik, Phys. Rev. Lett. **98**, 227202 (2007)

[91] R.G. Melko and R.K. Kaul, Phys. Rev. Lett. **100**, 017203 (2008)

[92] A. Sandvik, Phys. Rev. Lett. **104**, 177201 (2010)

[93] A.B. Kuklov, M. Matsumoto, N.V. Prokof'ev, B.V. Svistunov, and M. Troyer, Phys. Rev. Lett. **101**, 050405 (2008)

[94] A. Nahum, J.T. Chalker, P. Serna, M. Ortuno, and A.M. Somoza, Phys. Rev. X **5**, 041048 (2015)

[95] C. Wang, A. Nahum, M.A. Metlitski, C. Xu, and T. Senthil, Phys. Rev. X **7**, 031051 (2017)

[96] A. Nahum, Phys. Rev. B **102**, 201116(R) (2020)

[97] M.E. Zayed *et al.*, Nat. Phys. **13**, 962 (2017)

[98] J.Y. Lee, Y.-Z. You, S. Sachdev, and A. Vishwanath, Phys. Rev. X **9**, 041037 (2019)

[99] A.J. Millis, Phys. Rev. B **48**, 7183 (1993)

[100] T. Moriya: *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985)

[101] S.-S. Lee, Phys. Rev. B **80**, 165102 (2009)

[102] M.A. Metlitski and S. Sachdev, Phys. Rev. B **82**, 075127 (2010); *ibid.* **82**, 075128 (2010)

[103] A. Schlief, P. Lunts, and S.-S. Lee, Phys. Rev. X **7**, 021010 (2017)

[104] M.H. Gerlach, Y. Schattner, E. Berg, and S. Trebst, Phys. Rev. B **95**, 035124 (2017)

[105] M. Oshikawa, Phys. Rev. Lett. **84**, 3370 (2000)

[106] P.A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006)

[107] T. Senthil, S. Sachdev, and M. Vojta, Phys. Rev. Lett. **90**, 216403 (2003)

[108] T. Senthil, M. Vojta, and S. Sachdev, Phys. Rev. B **69**, 035111 (2004)

[109] U.F.P. Seifert, T. Meng, and M. Vojta, Phys. Rev. B **97**, 085118 (2018)

[110] H. Morita, S. Watanabe, and M. Imada, J. Phys. Soc. Jpn. **71**, 2109 (2002)

[111] B. Kyung and A.M.S. Tremblay, Phys. Rev. Lett. **97**, 046402 (2006)

[112] Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, Phys. Rev. Lett. **91**, 107001 (2003)

[113] A. Rosch, Eur. Phys. J. B **59**, 495 (2007)

[114] T. Senthil, Phys. Rev. B **78**, 035103 (2008)

[115] T. Senthil, Phys. Rev. B **78**, 045109 (2008)

[116] H. Terletska, J. Vucicevic, D. Tanasković, and V. Dobrosavljević, Phys. Rev. Lett. **107**, 026401 (2011)

[117] T. Furukawa, K. Miyagawa, H. Taniguchi, R. Kato, and K. Kanoda, Nat. Phys. **11**, 221 (2015)

[118] H. Eisenlohr, S.-S.B. Lee, and M. Vojta, Phys. Rev. B **100**, 155152 (2019)

[119] V. Anisimov, I. Nekrasov, D. Kondakov, T. Rice, and M. Sigrist, Eur. Phys. J. B **25**, 191 (2002)

[120] C. Pépin, Phys. Rev. Lett. **98**, 206401 (2007)

[121] M. Vojta, J. Low Temp. Phys. **161**, 203 (2010)

[122] P. Coleman, J.B. Marston, and A.J. Schofield, Phys. Rev. B **72**, 245111 (2005)

[123] P. Gegenwart, Q. Si, and F. Steglich, Nat. Phys. **4**, 186 (2008)

[124] S. Doniach, Physica B **91**, 231 (1977)

[125] Q. Si, Physica B **378-380**, 23 (2006); Phys. Stat. Sol. B **247**, 476 (2010)

[126] P. Coleman and A.H. Nevidomskyy, J. Low. Temp. Phys. **161**, 182 (2010)

[127] M. Vojta, Phys. Rev. B **78**, 125109 (2008)

[128] Q. Si, S. Rabello, K. Ingersent, and J.L. Smith, Nature **413**, 804 (2001); Phys. Rev. B **68**, 115103 (2003)

[129] T. Senthil, S. Sachdev, and M. Vojta, Physica B **359-361**, 9 (2005)

11 Quantum Compass and Kitaev Models

Jeroen van den Brink and Zohar Nussinov
Institute for Theoretical Solid State Physics
IFW Dresden
Department of Physics, Washington University
St. Louis, Missouri 63160, USA

Contents

1 Introduction to compass models **2**
1.1 Definition of compass models 3
1.2 90° compass models 4

2 Global, topological, and intermediate symmetries and invariances **7**
2.1 Exact and emergent symmetries 8
2.2 Consequences of intermediate symmetry 10
2.3 Symmetries of the 90° compass model 13

3 Kitaev's honeycomb model **16**
3.1 Features of Kitaev's honeycomb model 16
3.2 Majorana representation – Abelian phases 18
3.3 Braiding statistics 24
3.4 Broken time reversal symmetry – the non-Abelian phase 28

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction to compass models

Compass models are theories of matter in which the couplings between the internal spin (or other relevant field) components are inherently spatially (typically, direction) dependent. A simple illustrative example is furnished by the *90° compass model* on a square lattice in which only couplings of the form $\tau_i^x \tau_j^x$ (where $\{\tau_i^a\}_a$ denote Pauli operators at site i) are associated with nearest neighbor sites i and j separated along the x axis of the lattice while $\tau_i^y \tau_j^y$ couplings appear for sites separated by a lattice constant along the y axis. A very well-known compass model is the honeycomb Kitaev Hamiltonian. Such compass-type interactions can appear in diverse physical systems. This includes Mott insulators with orbital degrees of freedom where interactions sensitively depend on the spatial orientation of the orbitals involved, the low energy effective theories of frustrated quantum magnets, vacancy centers and cold atomic gases. Kitaev models, in particular the compass variant on the honeycomb lattice, realize basic notions of topological quantum computing. The fundamental inter-dependence between internal (spin, orbital, or other) and external (i.e. spatial) degrees of freedom which underlies compass models generally leads to very rich behaviors including the frustration of (semi-)classical ordered states on non-frustrated lattices and to enhanced quantum effects prompting, in certain cases, the appearance of zero temperature quantum spin liquids. As a consequence of these frustrations, new types of symmetries and their associated degeneracies may appear. In particular, these systems feature *intermediate* (more recently also referred to (especially in the high-energy and quantum information communities) and further classified as "higher form" or "subsystem") symmetries that lie midway between the extremes of global symmetries and local gauge symmetries and lead to effective dimensional reductions. We consider compass models in a unified manner, paying close attention to consequences of these symmetries, and to thermal and quantum fluctuations that stabilize orders via *order out of disorder* effects. We review non-trivial statistics and the appearance of *topological quantum orders* in compass systems in which, by virtue of their intermediate symmetry, standard orders do not arise.

Different physical contexts motivate compass models and they can emerge as low-energy effective models of systems with strongly interacting electrons. There are quite a few classes of materials where the microscopic interactions between electrons are described by an extended Hubbard model. Typically such materials contain transition-metal ions. Hubbard-type models incorporate both the hopping of electrons from lattice-site to lattice-site and the Coulomb interaction U between electrons that meet on the same site, typically the transition-metal ion. Particularly in the situation that electron-electron interactions are strong, effective low-energy models can be derived by expanding the Hubbard Hamiltonian in $1/U$, the inverse interaction strength. In such a low-energy model the interactions are only between the remaining *spin* and *orbital* degrees of freedom of the electrons. Compass model Hamiltonians arise when orbital degrees of freedom interact with each other.

In the situation that both orbital and spin degrees of freedom are present and their interactions are intertwined, the Kugel-Khomskii models arise [1]. Such models are relevant for strongly correlated electron systems such as transition metal (TM) oxides, when the low-energy electronic

behavior is dominated by the presence of very strong electron-electron interactions. The orbital degrees of freedom can be represented via pseudo-spins.
So-called e_g and t_{2g} orbital degrees of freedom that can emerge in transition metal compounds with electrons in partially filled TM d-shells, give rise to two-flavor compass models (for e_g) and to three-flavor compass models (for t_{2g}) [2–5]. Precisely these type of compass models also emerge in the study of systems of cold atoms in optical traps.

1.1 Definition of compass models

In order to define quantum compass models, we start by considering a lattice with sites on which quantum degrees of freedom live. Throughout this chapter the total number of lattice sites is denoted by N. Each lattice site has a vector pointing to it that is denoted by $\boldsymbol{r}$. When square (or cubic) lattices will be involved, these will be considered of dimension $N = L\times L$ (or $N = L\times L\times L$). On more general lattices, L denotes the typical linear dimension (i.e., linear extent along one of the crystal axis). We set the lattice constant to unity. The spatial dimensionality of the lattice is denoted by D (e.g., $D = 2$ for the square and honeycomb lattices, $D = 3$ in cubic and pyrochlore lattices etc.).
Depending on the problem at hand, we will refer to these degrees of freedom at the lattice sites as spins, pseudospins or orbitals. We denote these degrees of freedom by $\boldsymbol{\tau}_i$, where i labels the lattice sites and $\boldsymbol{\tau} \equiv \frac{1}{2}(\sigma^x, \sigma^y, \sigma^z)$, where σ^x, σ^y and σ^z are the Pauli matrices. In terms of the creation ($c^\dagger_\alpha$) and annihilation (c_α) operator for an electron in state α, the pseudospin operator $\boldsymbol{\tau}$ can be expressed as $\boldsymbol{\tau} = \frac{1}{2}\sum_{\alpha\beta} c^\dagger_\alpha \boldsymbol{\sigma}_{\alpha\beta} c_\beta$, where the sum is over the two different possibilities for each α and β. Here $\boldsymbol{\tau}$ is the fundamental $T = 1/2$ representation of SU(2), for $T > 1/2$ we use $\boldsymbol{T}$. A representation in terms of Pauli matrices is particularly useful for degrees of freedom that have two flavors, for instance two possible orientations of a spin (up or down) or two possible orbitals that an electron can occupy, as the Pauli matrices are generators of SU(2), the group of 2×2 matrices with determinant one. In most works in the field, it is common to explicitly label components of $T = 1/2$ degrees of freedom at different sites by $\sigma^\gamma_{\mathbf{r}}$. Following suite, we will at times (especially when discussing excitations in Kitaev's honeycomb model and its non-Abelian phase), interchangeably also use this more conventional notation. For degrees of freedom with n flavors, it makes sense to use a representation in terms of the generators of SU(n), which for the particular case of $n = 3$ are the eight Gell-Mann matrices λ_i, with $i = 1, 8$.
The name that one chooses to bestow upon the degree of freedom (whether *spin*, *pseudospin*, *color*, *flavor* or *orbital*) is of course mathematically irrelevant. For SU(2) quantum compass models it is important that the components of $\boldsymbol{\tau}$ obey the well-known commutation relation $[\tau^x, \tau^y]=i\tau^z$ and its cyclic permutations, and that $(\tau^\gamma)^2 = 1/4$ for any component $\gamma=x$, y or z. In the case of SU(3), in the fundamental representation $\boldsymbol{T}$ is the eight component vector $\boldsymbol{T}=\frac{1}{2}\sum_{\alpha\beta} c^\dagger_\alpha \boldsymbol{\lambda}_{\alpha\beta} c_\beta$, with commutation relations governed by those of the Gell-Mann matrices.
Compass models are characterized by the specific form that the interaction between the degrees of freedom assumes: (i) there is only an interaction between certain *vector components* of $\boldsymbol{\tau}$ and (ii) on different bonds in the lattice, different vector components interact. When, for instance,

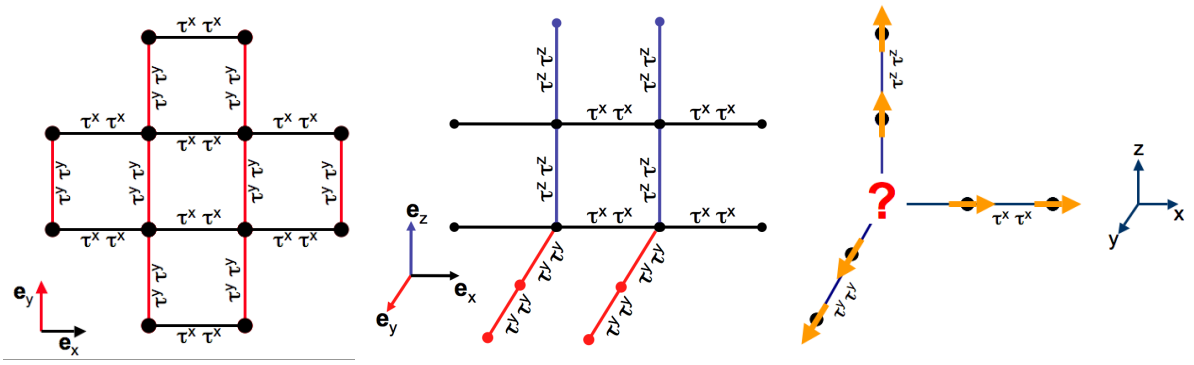


Fig. 1: *Left: The planar 90° compass model on a square lattice: the interaction of (pseudo-)spin degrees of freedom* $\boldsymbol{\tau} = (\tau^x, \tau^y)$ *along horizontal bonds that are connected by the unit vector* $\boldsymbol{e}_x$ *is* $\tau^x_{\boldsymbol{r}}\tau^x_{\boldsymbol{r}+\boldsymbol{e}_x}$. *Along vertical bonds* $\boldsymbol{e}_y$ *it is* $\tau^y_{\boldsymbol{r}}\tau^y_{\boldsymbol{r}+\boldsymbol{e}_y}$. *Middle: The 90° compass model on a cubic lattice: the interaction of (pseudo-)spin degrees of freedom* $\boldsymbol{\tau} = (\tau^x, \tau^y, \tau^z)$ *along horizontal bonds that are connected by the unit vector* $\boldsymbol{e}_x$ *is* $J\tau^x_i\tau^x_{i+\boldsymbol{e}_x}$. *On bonds connected by* $\boldsymbol{e}_y$ *it is* $J\tau^y_i\tau^y_{i+\boldsymbol{e}_y}$ *and along the vertical bonds it is* $J\tau^z_i\tau^z_{i+\boldsymbol{e}_z}$. *Right: Frustration in the 90° compass model on a cubic lattice. The interactions between pseudospins* $\boldsymbol{\tau}$ *are such that they tend to align their components* τ^x, τ^y *and* τ^z *along the* x, y *and* z*-axis, respectively. This causes mutually exclusive ordering patterns.*

a site i is linked to nearest neighbor sites j and k, the interaction along the lattice link $\langle ij\rangle$ can be of the type $\tau^x_i\tau^x_j$, whereas on the link $\langle ik\rangle$ it is $\tau^y_i\tau^y_k$. In the following sections specific Hamiltonians corresponding to various quantum compass models are introduced, in particular the 90° compass models, Kitaev's honeycomb model, 120° compass models and a number of generalizations thereof.

1.2 90° compass models

A basic realization of a quantum compass model can be set up on a two-dimensional square lattice, where every site has two horizontal and two vertical bonds. If one defines the interaction along horizontal lattice links $\langle ij\rangle_H$ to be $J\,\tau^x_i\tau^x_j$ and along the vertical links $\langle ij\rangle_V$ to be $J\,\tau^y_i\tau^y_j$, we have constructed the so-called *two-dimensional 90° quantum compass model* also known as the *planar 90° orbital compass model*, see Fig. 1. Its Hamiltonian is

$$H^{90^\circ}_{\square} = -J_x \sum_{\langle ij\rangle_H} \tau^x_i\tau^x_j - J_y \sum_{\langle ij\rangle_V} \tau^y_i\tau^y_j. \tag{1}$$

The isotropic variant of this system has equal couplings along the vertical and horizontal directions ($J_x = J_y = J$). The minus signs that appear in this Hamiltonian were chosen such that the interactions between the pseudospins $\boldsymbol{\tau}$ tend to stabilize uniform ground states with "ferro" pseudospin order. (In $D = 2$ the 90° compass models with "ferro" and "antiferro" interactions are directly related by symmetry). For clarity, we note that the isotropic two-dimensional compass model is very different from the two-dimensional Ising model

$$H^{Ising}_{\square} = -J \sum_{\langle ij\rangle_H} \tau^x_i\tau^x_j - J \sum_{\langle ij\rangle_V} \tau^x_i\tau^x_j = -J \sum_{\langle ij\rangle} \tau^x_i\tau^x_j,$$

where on each horizontal and vertical vertex of the square lattice the interaction is the same and of the form $\tau_i^x \tau_j^x$. It is also very different from the two-dimensional XY model

$$H_{\square}^{XY} = -J \sum_{\langle ij\rangle_H, \langle ij\rangle_V} (\tau_i^x \tau_j^x + \tau_i^y \tau_j^y),$$

because also in this case the interaction terms in the Hamiltonian are the same on all bonds.
One can rewrite the 90° compass Hamiltonian in a more compact form by introducing the unit vectors $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ that denote the bonds along the x- and y-direction in the 2D lattice, so that

$$H_{\square}^{90^\circ} = -J \sum_{\boldsymbol{r}} (\tau_{\boldsymbol{r}}^x \tau_{\boldsymbol{r}+\boldsymbol{e}_x}^x + \tau_{\boldsymbol{r}}^y \tau_{\boldsymbol{r}+\boldsymbol{e}_y}^y), \tag{2}$$

where the sum over $\boldsymbol{r}$ represents the sum over lattice sites and every is bond counted only once. With this notation the compass model Hamiltonian can be cast in the more general form

$$H_{\square}^{90^\circ} = -J \sum_{\boldsymbol{r},\gamma} \tau_{\boldsymbol{r}}^\gamma \tau_{\boldsymbol{r}+\boldsymbol{e}_\gamma}^\gamma, \tag{3}$$

where for the 90° square lattice compass model, $H_{\square}^{90^\circ}$, we have $\gamma = 1, 2$, $\{\tau^\gamma\} = \{\tau^1, \tau^2\} = \{\tau^x, \tau^y\}$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} = \{\boldsymbol{e}_x, \boldsymbol{e}_y\}$.
This generalized notation allows for different compass models and the more well-known models such as the Ising or Heisenberg model to be cast in the same form, see Table 1. For instance the two-dimensional square-lattice Ising model $H_{\square}^{Ising}$ corresponds to $\gamma = 1, 2$ with $\{\tau^\gamma\} = \{\tau^x, \tau^x\}$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_x, \boldsymbol{e}_y\}$. The Ising model on a three dimensional cubic lattice is then given by $\gamma = 1...3$, $\{\tau^\gamma\} = \{\tau^x, \tau^x, \tau^x\}$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$. The XY model on a square lattice $H_{\square}^{XY}$ corresponds to $\gamma = 1...4$, $\{\tau^\gamma\} = \{\tau^x, \tau^y, \tau^x, \tau^y\}$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_y\}$. Another example is the square lattice Heisenberg model, where we have $\gamma = 1...6$, $\{\tau^\gamma\} = \{\tau^x, \tau^y, \tau^z, \tau^x, \tau^y, \tau^z\}$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_y, \boldsymbol{e}_y\}$, so that in this case $\sum_\gamma \tau_{\boldsymbol{r}}^\gamma \tau_{\boldsymbol{r}+\boldsymbol{e}_\gamma}^\gamma$ is equal to $\sum_\gamma \boldsymbol{\tau}_{\boldsymbol{r}} \cdot \boldsymbol{\tau}_{\boldsymbol{r}+\boldsymbol{e}_\gamma}$.
This class of compass models can be further generalized in a straightforward manner by allowing for a coupling strength J_γ between the pseudospins τ^γ that depends on the direction of the bond γ (*anisotropic compass models* [6]) and by adding a field h_γ that couples to τ^γ linearly [7, 8]. This generalized class of compass models is then defined by the Hamiltonian

$$\mathcal{H}_{compass} = -\sum_{\boldsymbol{r},\gamma} \left(J_\gamma \tau_{\boldsymbol{r}}^\gamma \tau_{\boldsymbol{r}+\boldsymbol{e}_\gamma}^\gamma + h_\gamma \tau_{\boldsymbol{r}}^\gamma \right). \tag{4}$$

From a historical (as well as somewhat practical) viewpoint the *three dimensional 90° compass model* is particularly interesting. Denoted by $H_{3\square}^{90^\circ}$, it is customarily defined on a cubic lattice and given by $\mathcal{H}_{compass}$, Eq. (4), where γ spans three Cartesian directions: $\gamma = 1, 2, 3$ with $\{\tau^\gamma\} = \{\tau^x, \tau^y, \tau^y\}$, $J_\gamma = J = 1$, $h_\gamma = 0$ and $\{\boldsymbol{e}_\gamma\} = \{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$, so that

$$H_{3\square}^{90^\circ} = -J \sum_{\boldsymbol{r}} \left(\tau_{\boldsymbol{r}}^x \tau_{\boldsymbol{r}+\boldsymbol{e}_x}^x + \tau_{\boldsymbol{r}}^y \tau_{\boldsymbol{r}+\boldsymbol{e}_y}^y + \tau_{\boldsymbol{r}}^z \tau_{\boldsymbol{r}+\boldsymbol{e}_z}^z \right). \tag{5}$$

Thus, by allowing γ to assume values $\gamma = 1, 2, 3$, the square lattice 90 degree compass model of Eq. (3) is trivially extended to three spatial dimensions. Similarly, by allowing $\gamma = 1, 2, ..., D$,

Model Hamiltonian: $\mathcal{H} = -\sum_{\boldsymbol{r},\gamma} \tau^{\gamma}_{\boldsymbol{r}} \tau^{\gamma}_{\boldsymbol{r}+\boldsymbol{e}_\gamma}$

$\{\tau^\gamma\}$	$\{\boldsymbol{e}_\gamma\}$	model name	symbol	dim
$\{\tau^x\}$	$\{\boldsymbol{e}_x\}$	Ising chain	H_1^{Ising}	1
$\{\tau^x, \tau^y\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_x\}$	XY chain	H_1^{XY}	1
$\{\tau^x, \tau^y, \tau^z\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_x\}$	Heisenberg chain	H_1^{Heis}	1
$\{\tau^x, \tau^x\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y\}$	square Ising	$H_\square^{Ising}$	2
$\{\tau^x, \tau^x, \tau^x\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$	cubic Ising	$H_{3\square}^{Ising}$	3
$\{\tau^x, \tau^y, \tau^x, \tau^y\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_y\}$	square XY	$H_\square^{XY}$	2
$\{\tau^x, \tau^y, \tau^z, \tau^x, \tau^y, \tau^z\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_y, \boldsymbol{e}_y\}$	square Heisenberg	$H_\square^{Heis}$	2
$\{\tau^x, \tau^y\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y\}$	square 90° compass	$H_\square^{90^\circ}$	2
$\{\tau^x, \tau^y, \tau^z\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$	cubic 90° compass	$H_{3\square}^{90^\circ}$	3
$\{\frac{\tau^x+\sqrt{3}\tau^y}{2}, \frac{\tau^x-\sqrt{3}\tau^y}{2}\}$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y\}$	square 120° compass	$H_\square^{120^\circ}$	2
With $\{\theta_\gamma\} = \{0, 2\pi/3, 4\pi/3\}$:				
$\{\tau^x, \tau^x, \tau^x\}$	$\boldsymbol{e}_x \cos\theta_\gamma + \boldsymbol{e}_y \sin\theta_\gamma$	honeycomb Ising	$H_\hexagon^{Ising}$	2
$\{\tau^x, \tau^y, \tau^z\}$	$\boldsymbol{e}_x \cos\theta_\gamma + \boldsymbol{e}_y \sin\theta_\gamma$	honeycomb Kitaev	$H_\hexagon^{Kitaev}$	2
$\{\tau^x, \tau^x, \tau^z\}$	$\boldsymbol{e}_x \cos\theta_\gamma + \boldsymbol{e}_y \sin\theta_\gamma$	honeycomb XXZ	$H_\hexagon^{XXZ}$	2
$\pi^\gamma = \tau^x \cos\theta_\gamma + \tau^y \sin\theta_\gamma$	$\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$	cubic 120°	$H_{3\square}^{120^\circ}$	3
π^γ	$\boldsymbol{e}_x \cos\theta_\gamma + \boldsymbol{e}_y \sin\theta_\gamma$	honeycomb 120°	$H_\hexagon^{120^\circ}$	2
With $\{\theta_\gamma\} = \{0, 2\pi/3, 4\pi/3\}$ and $\eta = \pm 1$:				
$\{\tau^x, \tau^y, \tau^z\}$	$\eta\boldsymbol{e}_x \cos\frac{\theta_\gamma}{2} + \eta\boldsymbol{e}_y \sin\frac{\theta_\gamma}{2}$	triangular Kitaev	$H_\triangle^{Kitaev}$	2
π^γ	$\eta\boldsymbol{e}_x \cos\frac{\theta_\gamma}{2} + \eta\boldsymbol{e}_y \sin\frac{\theta_\gamma}{2}$	triangular 120°	$H_\triangle^{120}$	2

Table 1: *Generalized notation that casts compass models and the more well-known model Hamiltonians such as the Ising, XY or Heisenberg models in the same form. Additional spatial anisotropies can be introduced, for instance by coupling constants J_γ that depend on the bond direction $\boldsymbol{e}_\gamma$. Doing so would change the* strengths *of the interaction on different links, but not the form of those interactions: these are determined by how different vector components of $\boldsymbol{\tau}_{\boldsymbol{r}}$ and $\boldsymbol{\tau}_{\boldsymbol{r}+\boldsymbol{e}_\gamma}$ couple.*

it can be extended to arbitrary spatial dimension D (which we will return to in later sections). The structure of $H_{3\square}^{90^\circ}$ is schematically indicated in Fig. 1. This compass model is actually the one that was originally proposed by [1] in the context of orbital ordering. At that time it was noted that even if the interaction on each individual bond is Ising-like, the overall symmetry of the model is considerably more complicated.

It is typical for compass models that even the ground state structure is non-trivial. For a system governed by $H_{3\square}^{90^\circ}$, pairs of pseudospins on lattice links parallel to the x-axis, for instance, favor pointing their pseudospins $\boldsymbol{\tau}$ along x so that the expectation value $\langle \tau^x \rangle \neq 0$, see Fig. 1. Similarly,

on bonds parallel to the y-direction, it is advantageous for the pseudospins to align along the y direction, so that $\langle \tau^y \rangle \neq 0$. It is clear that at any given site the bonds along x, y and z cannot be satisfied at the same time. Therefore the interactions are strongly frustrated. This situation bears resemblance with the dipole-dipole interactions between magnetic needles that are positioned on a lattice, hence the name *compass* models.

Such a frustration of interactions is typical of compass models, but of course also appears in numerous other systems. Indeed, on a conceptual level, many of the ideas and results that will be discussed, such as renditions of thermal and quantum fluctuation-driven ordering effects, unusual symmetries and ground state sectors labeled by topological invariants, have similar incarnations in frustrated spin, charge, cold atom and Josephson junction array systems. Although these similarities are mostly conceptual there are also instances where there are exact correspondences. For instance, the two dimensional 90° compass model is, in fact, dual to the Xu-Moore model describing Josephson coupling between superconducting grains in a square lattice [9–11, 6, 12].

2 Global, topological, and intermediate symmetries and invariances

In terms of symmetries, compass systems are particularly rich. In what follows, we will discuss the invariances that these systems exhibit, but first recall the classification of orders and their relation to symmetry:

(i) *Global symmetry*. In many condensed matter systems (e.g. ferromagnets, liquids), there is an invariance of the basic interactions with respect to global symmetry operations (e.g., continuous rotations in the case of ferromagnets, uniform translations and rotations in liquids) that are to be simultaneously performed on all of the constituents of the system. At sufficiently low temperatures (or strong enough interactions), such symmetries might be *spontaneously* broken.

(ii) *Topological invariants and orders*. Topological orders have been the object of some fascination in more recent years [13]. In the condensed matter community, part of the activity in analyzing these types of order is stimulated by the prospects of fault-tolerant quantum computation. What lies at the crux of topological order is the observation is that even if, in some cases, global symmetry breaking cannot occur, systems may nevertheless still exhibit a robust order of a non-local, topological, type.

The most prominent examples of topological order – long studied by high energy theorists – are afforded by gauge theories [14, 15, 13]. Some of the current heavily studied quintessential models of topological quantum order in condensed matter and quantum information lattice theories, e.g., [16, 13] share much in common with the early pioneering lattice gauge theory concept along with the explicit simplest lattice gauge model first introduced by Franz Wegner [14].

Gauge theories display *local gauge symmetries* and indeed, in pure gauge theories – theories that have only gauge bosons yet no matter sources – the only measurable quantities pertain to correlators defined on loops, the so-called *Wilson loops*. Related products pertain to open contours in some cases when matter sources are present [15, 17, 18].

(iii) *Intermediate symmetry*. The crucial point is that many compass systems display symmetries which, generally, lie midway between the above two extremes of global symmetries and local gauge symmetries. These symmetries are sometimes known as "sliding" symmetries and aside from compass models are also present in numerous other systems. These include, amongst many others, arrays of Luttinger liquids [19, 20], quantum Hall smectic phases [21, 22], ring exchange models of frustrated models [23], and Kondo lattice systems [24]. In the past few years, there been an extremely intense resurgence of interest in such (in particular, "higher-form" type) symmetries that has been triggered anew by their study in the high energy community [25].

To clarify the distinction between these different symmetries, we can rephrase it in a formal way as it applies to general systems [26, 27]. Consider a theory with fields $\{\phi_{\boldsymbol{i}}\}$ that is characterized by a Hamiltonian H (or action $\mathcal{S}$).

Definition: A *d-dimensional gauge-like symmetry* of a theory is a group of symmetry transformations such that the minimal non-empty set of fields $\{\phi_{\boldsymbol{i}}\}$ changed by the group operations occupies a d-dimensional subset ($\mathcal{C}$) of the full D-dimensional region on which the theory is defined. In the following we will refer to such symmetries as *d-dimensional symmetries*.

To exercise this notion it is useful to make contact with known cases. Clearly local gauge symmetries correspond to symmetries of dimension $d = 0$. That is, gauge transformations can be applied locally at any point in space – a region of dimension $d = 0$. At the opposite extreme, e.g., in a nearest neighbor ferromagnet on a D-dimensional lattice, described by the Heisenberg Hamiltonian $H = -J \sum_{\langle ij \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j$, the system is invariant under a global rotation of all spins. As the volume influenced by the symmetry operation occupies a D-dimensional region and in this case $d = D$.

In their simplest form, one which typically appears in compass models, d-dimensional symmetries are of the form

$$\prod_{j \in P} g_j \tag{6}$$

where g_j are group elements associated with a site j and P is a d-dimensional spatial region. In many cases, depending on the boundary conditions of the system, P correspond to entire open d-dimensional planes (as in 90° compass models; see, e.g., Fig. 2) or closed contours (when compass models are endowed with periodic boundary conditions). Defect creation operators (those that restore symmetries) and translations of defects are typically products of local group elements that do not span such an entire region P but rather a fragment of it (see, e.g., the open finite string in Fig. 2 with domain wall boundaries) generally leading to defects at the boundaries where the group element operations are applied [28].

2.1 Exact and emergent symmetries

A Hamiltonian H, and by extension the system it describes, can have two principal kinds of symmetries: exact and emergent ones. These are defined as follows.

(i) *Exact symmetries*. By this, one refers to the existence operators $\hat{O}$ that commute with the

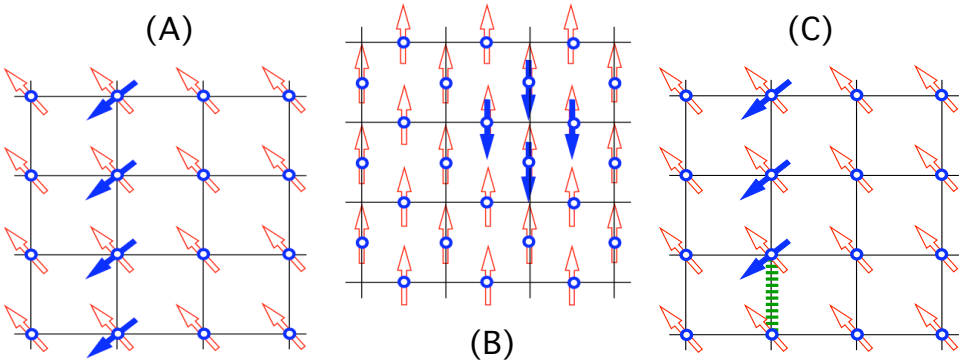


Fig. 2: *(A) The 90° square lattice compass model. The action of the $d = 1$ symmetry operation of Eq. (16) when the "plane" P is chosen to lie along the vertical axis. (B) A $d = 0$ (local) gauge symmetry. Defects within a gauge theory cost a finite amount of energy. Local symmetries such as the one depicted above for an Ising lattice gauge theory cannot be broken. (C) A defect in a semi-classical ground state of the two dimensional orbital compass model. Defects such as this do not allow for a finite on-site magnetization. The energy penalty for this defect is finite (there is only one bad bond – the dashed line) whereas, precisely as in $d = 1$ Ising systems, the entropy associated with such defects is monotonically increasing in system size [28].*

Hamiltonian

$$\left[H, \hat{O}\right] = 0. \tag{7}$$

Such operators, indicated by a hat, $\hat{}$, reflect symmetries of the Hamiltonian.
(ii) *Emergent symmetries*. In many compass (and numerous other) systems, there are operators $\tilde{O}$ that do not commute with the Hamiltonian,

$$\left[H, \tilde{O}\right] \neq 0 \tag{8}$$

i.e., do not satisfy Eq. (7) and are therefore indicated by a tilde, $\tilde{}$. Yet these operators do become symmetries when projected to a particular sector – a particular subset of states on which the Hamiltonian acts. That is,

$$[H, \mathcal{P}\tilde{O}\mathcal{P}] = 0, \tag{9}$$

where $\mathcal{P}$ is the relevant projection operator to that sector. In this case, if one defines $\mathcal{P}\tilde{O}\mathcal{P} = \hat{O}$ then $\hat{O}$ will be an exact symmetry satisfying Eq. (7).
The most prominent cases in condensed matter systems, including compass models in particular (yet also many others, e.g., [24, 29, 30]) relate to symmetries that appear in the *ground state sector* alone. In such instances, the symmetries are sometimes said to *emerge* in the low energy sector of the theory.
Although the formulation above is for quantum Hamiltonians, the same can, of course, be said for classical systems. There are numerous classical systems in which the application of a particular operation on an initial configuration will yield, in general, a new configuration with a differing energy. However, when such an operation is performed on a particular subset of configurations, such as the classical ground states, it will lead to other configurations that have precisely the same energy as the initial state. Similarly, certain quantum systems exhibit such particular symmetries

only in their large pseudo-spin (or classical) limit. In such cases, symmetries may be said to emerge in the large pseudo-spin (or classical) limit.

One should note that emergent low-energy symmetries are notably different from the far more standard situation of spontaneous symmetry breaking, wherein an invariance of the Hamiltonian (or action) is spontaneously broken in individual low energy states (which are related to one another by the symmetry operation at hand). In the condensed matter arena, the canonical example is rotationally symmetric ferromagnets in a spatial dimension larger than two, in which at sufficiently low temperature a finite magnetization points along a certain direction – thus breaking the rotational symmetry. Another canonical example is the discrete (*up* $\leftrightarrow$ *down* or) time reversal symmetry which is broken in Ising ferromagnets in dimensions large than one. Spontaneous symmetry breaking appears in systems that exhibit long-range order of some sort such as crystallization (breaking translational and rotational symmetries), superconductors (local gauge invariance and a Anderson-Higgs mechanism), or superfluid Helium. Other examples include the Higgs mechanism of particle physics, chiral symmetry breaking in quantum chromodynamics, nucleon pairing in nuclei, electro-weak symmetry breaking at low energies, and related mass generation.

In all of these textbook examples, the system is symmetric at high energies and exhibits low-energy states that do not have that symmetry. However, in low energy emergent symmetries, the situation is reversed: the system may become *more* symmetric in the low-energy sector. We will discuss explicit examples of exact and emergent symmetries in compass models in the following sections.

2.2 Consequences of intermediate symmetry

2.2.1 Degeneracy of spectrum

We now briefly discuss how the presence of a d-dimensional intermediate symmetry, either classical or quantum, implies an exponential degeneracy of the energy spectrum that corresponds to the Hamiltonian. The application of intermediate symmetries on disparate d-dimensional planes leads to inequivalent states that all share the same energy. If a symmetry transformation $\tilde{O}_P$ has its support on a d-dimensional plane P, then one can define the composite symmetry operators

$$\tilde{O}_{composite} = \tilde{O}_{P_1}\tilde{O}_{P_2}..\tilde{O}_{P_R}. \tag{10}$$

For a hypercubic lattice in D dimensions which is of size $L\times L\times L\cdots\times L$, the number of independent planes (R) in Eq. (10) scales as $R = \mathcal{O}(L^{d'})$ where

$$d' = D - d. \tag{11}$$

If each individual d-dimensional symmetry operation (exact or emergent) U_{P_i} leads to a degeneracy factor of m then the composite operation of Eq. (10) can lead to a degeneracy (of any state (for exact symmetries) or of the ground state (for emergent symmetries)) whose logarithm is of magnitude

$$\log_m \mathsf{degeneracy} = \mathcal{O}(L^{D-d}). \tag{12}$$

That this is indeed the case is clearer for classical systems with discrete symmetries than for quantum systems. Nevertheless, in the thermodynamic limit and/or on lattices whose boundaries are tilted the degeneracy factor of Eq. (12) associated with the intermediate d-dimensional symmetries becomes exact [31]. On hypercubic lattices, such as the square lattice of the planar 90° compass model discussed in subsection 2.3, whose boundaries are the same along the d' directions orthogonal to the planes P, the application of the operators of Eq. (10) does not lead to independent states for finite size systems. However, in the thermodynamic limit, the application of disparate operators of the form of Eq. (10) on a given initial state may lead to orthogonal states.

2.2.2 Dimensional reduction

The existence of intermediate symmetries has important consequences: it implies a dimensional reduction. The corresponding dimensional reduction is only with respect to expectation values of local quantities: the free energies of these systems and the transitions that they exhibit are generally those of systems in high dimensions [26, 27].

Theorem on Dimensional Reduction More precisely, the expectation value of any such quantity $\langle f \rangle$ in the original system (of dimension D) is bounded from above by the expectation value of the same quantity evaluated on a d dimensional region:

$$|\langle f \rangle| \leq |\langle f \rangle|_{H_d}. \tag{13}$$

The expectation value $\langle f \rangle$ refers to that done in the original system (or lattice) that resides in D spatial dimensions. The Hamiltonian H_d on the right-hand side is defined on a d dimensional subregion of the full lattice (system). The dimensionality $d \leq D$. The Hamiltonian H_d preserves the range of the interactions of the original systems. It is formed by pulling out of the full Hamiltonian on the complete (D dimensional) lattice, the parts of the Hamiltonian that appear within the d dimensional sub-region ($\mathcal{C}$) on which the symmetry operates. Fields (spins) external to $\mathcal{C}$ act as non-symmetry breaking external fields in H_d. The bound of Eq. (13) becomes most powerful for quantities that are not symmetry invariant as then the expectation values $\langle f \rangle_{H_d}$ need to vanish for low spatial dimensions d (as no spontaneous symmetry breaking can occur). This, together with Eq. (13), then implies that the expectation value of $\langle f \rangle$ on the full D dimensional spatial lattice must vanish. By "non invariant" we mean that $f(\phi_{\boldsymbol{i}})$ vanishes when summed over all arguments related to each other a d dimensional symmetry operation, $\sum_k f[\boldsymbol{g}_{\boldsymbol{i}k}(\phi_{\boldsymbol{i}})] = 0$. For continuous symmetries, non-invariance explicitly translates into an integral over the group elements $\int f[\boldsymbol{g}_{\boldsymbol{i}}(\phi_{\boldsymbol{i}})]\, d\boldsymbol{g} = 0$.
We will now summarize general corollaries of such symmetry based analysis for general systems.

Corollaries By choosing f to be the order parameter or a two-particle correlator, one arrives at the following general corollaries [26, 32, 27]:
Corollary I: Any local quantity that is not invariant under local symmetries ($d = 0$) or symmetries that act on one dimensional regions ($d = 1$) has a vanishing expectation value $\langle f \rangle_{H_d}$ at any finite

temperature. This follows as both zero- and one-dimensional systems cannot exhibit symmetry breaking: in one and two dimensional systems, the expectation value of any local quantities not invariant under global symmetries: $\langle f \rangle = 0$.

Physically, entropy overwhelms energetic penalties and forbids a symmetry breaking. Just as in zero- and one-dimensional systems, much more entropy is gained by introducing defects (e.g., domain walls in discrete systems), the same energy-entropy calculus is replicated when these symmetries are embedded in higher dimensions. An example with $d = 1$ domain walls in a two-dimensional systems is afforded by the planar 90° compass model (see Fig. 2); even though the planar compass model is two-dimensional, the energy cost of these domain walls is identical to that in a $d = 1$ system. The particular case of local ($d = 0$) symmetry is that of Elitzur's theorem [33] so well known in gauge theories. We may see it more generally as a consequence of dimensional reduction.

A discussion of how, by virtue of this consequence, such symmetries may protect and lead to topological quantum orders in systems at both finite and zero temperature appears in [34, 28].

Corollary II: One can push the consequences further by recalling that no symmetry breaking occurs for continuous symmetries in two spatial dimensions. Here again, free energy penalties are not sufficiently strong to induce order. When embedding continuous two dimensional symmetries in higher dimensions, the energy entropy balance is the same and the same result is attained $\langle f \rangle = 0$ at all finite temperatures for any quantity f that is not invariant under continuous $d \leq 2$ symmetries.

Further noting that order does not exist in continuous two dimensional systems also at zero temperature in the presence of a gap between ground and the next excited state, one similarly finds that for a $d \leq 2$ dimensional continuous symmetry the expectation value of any local quantity not invariant under this symmetry, strictly vanishes at zero temperature. Though local order cannot appear, multi-particle (including topological) order can exist. In standard gauge ($d = 0$) theories, the product of gauge degrees of freedom along a closed loop (the Wilson loop) can attain a non-zero value as it may be invariant under all $d = 0$ symmetries. In more general theories with higher d dimensional symmetries, similar considerations may lead to loop (or "brane") type correlators that involve multiple fields and are invariant under all low dimensional symmetries. Precisely such non-local correlation functions appear in Kitaev's honeycomb model and many other systems with topological orders [35, 28, 34, 36]. Symmetry breaking in the highly degenerate compass models often transpires by a fluctuation driven mechanism ("order by disorder") [37–39]. In this mechanism, entropic contributions to the free energy play a key role.

Corollary III: Not only can one make statements about the absence of symmetry breaking, we can also adduce fractionalization of non-symmetry invariant quantities in high dimensional system. That occurs if no (quasi-particle type) resonant terms appear in the lower dimensional spectral functions [32].

This corollary allows for fractionalization in quantum systems, where $d = 1, 2$. It enables symmetry invariant quasi-particles excitations to *coexist* with non-symmetry invariant fractionalized excitations. Fractionalized excitations may propagate in $D-d$ dimensional regions. Examples afforded by several frustrated spin models where spinons may drift along lines on the square

lattice [29] and in D dimensional regions on the pyrochlore lattice [30].
In what follows, we explicitly enumerate the symmetries that appear in various compass models. The *physical origin of dimensional reduction* in these systems can be seen examining intermediate symmetry restoring defects.

2.3 Symmetries of the 90° compass model

We now classify symmetries of the 90° compass model in various spatial dimensions, considering both quantum and classical versions. To highlight some aspects of the symmetries of this system, it is profitable to discuss the general anisotropic compass model, as given for $D = 2$ in Eq. (1) with general couplings J_x and J_y and in general spatial dimension D given by Eq. (4), without field

$$H_{D\square}^{90^\circ} = -\sum_{\boldsymbol{r},\gamma} J_\gamma \, \tau_{\boldsymbol{r}}^\gamma \tau_{\boldsymbol{r}+\boldsymbol{e}_\gamma}^\gamma \, . \tag{14}$$

The equivalent classical Hamiltonian on a D-dimensional hyper cubic lattice is

$$H_{D\square}^{90^\circ,\,\text{class}} = -\sum_{\boldsymbol{r},\gamma} J_\gamma \, T_{\boldsymbol{r}}^\gamma T_{\boldsymbol{r}+\boldsymbol{e}_\gamma}^\gamma \, . \tag{15}$$

In the quantum systems, T^γ are generators of the representations of SU(2) of size $(2T+1)$. For a pseudo-spin 1/2 system, $T^\gamma = \tau^\gamma/2$. In the classical arena, T^γ are the Cartesian components of normalized vector $\boldsymbol{T}$. These classical and quantum Hamiltonian systems exhibit both exact and emergent symmetries.

2.3.1 Exact discrete intermediate symmetries

Exact symmetries of both the square lattice and cubic lattice 90° compass model in any pseudo-spin representation are given by [26, 40–42, 6, 43]

$$\hat{O}^{(\gamma)} = \prod_{\boldsymbol{r}\in P_\gamma} e^{i\pi T_{\boldsymbol{r}}^\gamma} \tag{16}$$

where P_γ is any line (in the case of the two-dimensional model) or plane (in the case of the cubic lattice model) which is orthogonal to the external $\boldsymbol{e}_\gamma$ axis of the lattice. A schematic for the $D = 2$ dimensional case is provided in panel (a) of Fig. 2.
The exact nature of the symmetries of Eq. (16) is readily seen: the operators of Eq. (16) commute with the general Hamiltonian of Eq. (15): $[O^{(\gamma)},\, H] = 0$. Thus, rotations of individual planes about an orthogonal axis leave the system invariant. Written generally, for a 90° compass model in D dimensions, the planes P_γ are objects of spatial dimensionality $d = D-1$. In the $D = 3$ dimensional system, the symmetries of Eq. (16) are of dimension $d = 2$ as the planes P_γ are two-dimensional objects. On the square lattice, the symmetries are of dimension $d = 1$ as P_γ are lines. These symmetries hold for both the quantum system with arbitrary size pseudo-spin as well as the classical system in a high number of dimensions D. A consequence of these symmetries is an exponential in L^{D-1} degeneracy of each eigenstate of the Hamiltonian (including but

not limited to ground states) in systems with "tilted" boundary conditions that emulate the thermodynamic limit [31]. In pseudo-spin one-half realizations of this system, Eq. (14), on an $L \times L$ square lattice, a 2^L degeneracy was numerically adduced for anisotropic systems ($J_x \neq J_y$) in the thermodynamic limit [43]. Correlation functions involving the symmetry operators were examined in [44].

Now, here is an important point to which we wish to reiterate – that of the *physical origin of the dimensional reduction in this system.* In a $D = 2$ dimensional 90° compass model system, the energy cost for creating defects (domain walls) is identical to that in a $d = 1$ dimensional system (see Fig. 2). With the aid of the bound of Eq. (13), we then see the finite temperature expectation value $\langle \sigma_i^z \rangle = 0$ within the $D = 2$ orbital compass model. The physical engine behind the loss of on-site order of $\langle \sigma_i^z \rangle$ is the proliferation of solitons, see Fig. 2. Just as in $d = 1$ dimensional systems, domain walls (solitons) cost only a finite amount of energy while their entropy increases with system size. A schematic is provided in panel (c) of Fig. 2. The Hamiltonian $H_{d=1}$ defined on the vertical chain of Fig. 2 where these operations appear is none other than a one dimensional Ising Hamiltonian augmented by transverse fields generated by spins outside the vertical chain. Any fixed values of the spins outside the $d = 1$ dimensional chain lead to transverse fields that act on the chain. These along the Ising exchange interactions between neighboring spins along the chain lead in this case to the pertinent $H_{d=1}$ in Eq. (13): that of a transverse field Ising model Hamiltonian. By virtue of their location outside the region where the symmetry of Eq. (16) operates, the spins $\sigma^x_{i \notin P_x}$ do not break the discrete $d = 1$ symmetry associated with the plane P_x. These defects do not enable a finite temperature symmetry breaking.

2.3.2 Exact discrete global symmetries

When the couplings are not completely anisotropic (e.g., $J_x = J_y \neq J_z$ or $J_x = J_y = J_z$ on the cubic lattice or $J_x = J_y$ on the square lattice) there are additional discrete symmetries augmenting the $d = D-1$ Ising symmetries detailed above. For instance, when $J_x = J_y \neq J_z$ a global discrete rotation of all pseudo-spins on the lattice by an angle of 90° about the T^z direction leaves the Hamiltonian of Eq. (15) invariant. Such a discrete rotation essentially permutes the x and y oriented bonds which are all of equal weight in the isotropic case when these are summed over the entire square lattice. The same, of course, also applies for the square lattice model when $J_x = J_y$.

Yet another possible representation of essentially the same symmetry as it is pertinent to the exchange of couplings in the compass model is that of a uniform global rotation by 180° about the $(1, 1)/\sqrt{2}$ direction of the pseudo-spins. Similarly, when $J_x = J_y = J_z$, a uniform global rotation by 120° of all pseudo-spins about the internal $(1, 1, 1)/\sqrt{3}$ pseudo-spin direction is also a discrete symmetry; this latter symmetry is of the Z_3 type – if performed three times in a row, this will give back the identity operation.

These additional discrete symmetries endow the system with a higher degeneracy. For isotropic systems ($J_x = J_y$), numerically a 2^{L+1} fold degeneracy is seen in the pseudo-spin $T = 1/2$ system [43]; this additional doubling of the degeneracy is related to a global Ising operation of a

rotation by 180° about a chosen pseudo-spin direction that leaves the system invariant. These additional symmetries are global symmetries and thus of a dimension $d = D$ which is higher than that of the discrete lower dimensional that are present in both the anisotropic and isotropic systems ($d = D-1$). As a result, in, e.g., the isotropic $D = 2$ dimensional 90° compass model may exhibit a finite temperature breaking of such a discrete global symmetry associated with such a discrete rotation. By contrast, the $d = 1$ symmetries of the two-dimensional 90° compass model cannot be broken as discussed in section 2.2.2.

We note that in the classical anisotropic rendition of this system the degeneracy is exactly the same – i.e., 2^L, aside from continuous emergent symmetries that will be discussed in the next section. The classical isotropic case is somewhat richer. There, each uniform pseudo-spin state has an additional degeneracy factor of 2^{2L} associated with the $2L$ independent classical $d = 1$ Ising symmetries.

2.3.3 Emergent intermediate discrete symmetries: cubic 90° model

We now turn to intermediate symmetries that appear in the large pseudo-spin (or classical) limit of the 90° compass model in three dimensions. In its classical limit, the 90° compass model on the cubic lattice has $d = 1$ inversion (or reflection) symmetries along lines parallel to each of the three Cartesian axes x_a. Along these lines, we may set $\tau_i^a \to -\tau_i^a$ and not touch the other components. This corresponds to, e.g, a reflection in the internal xy pseudo-spin plane when we invert τ^z and not alter the x or y components.

We explicitly note that this transformation is not canonical and does not satisfy the commutation relation and is thus disallowed quantum mechanically; indeed, this appears only as an emergent symmetry in the classical limit of large pseudo-spin. Instead in the 90° compass model on the cubic lattice, quantum mechanically we have the $d = 2$ symmetries which we wrote earlier (which of course trivially also hold for the classical system). Thus, the quantum system is less symmetric than its classical counterpart.

By contrast to the cubic lattice case, for the square lattice 90° compass model, the intermediate $d = 1$ symmetries of Eq. (16) are are not emergent symmetries but rather exact quantum (as well as classical) symmetries.

2.3.4 Emergent continuous global symmetries

In addition to its exact symmetries, the 90° model also exhibits emergent symmetries in its isotropic version. As mentioned earlier, globally uniform pseudo-vector configurations are ground states of any classical isotropic ferromagnetic compass model. Thus any global rotation of all pseudo-spins is an emergent symmetry of the 90° models. In the $D = 2$ system, this corresponds to a global U(1) rotation of all angles of the planar pseudo-spins. In the $D = 3$ cubic lattice system, any SO(3) rotation of the three-dimensional pseudo-spins is an emergent symmetry. That a rotation does not change the energy of any uniform configuration is clear in the 90° model. Imagine that all pseudo-spins in the planar 90° model are oriented at an angle θ relative to the T^x axis. In such a case, the energy associated with the horizontal bonds,

$T^x_{\boldsymbol{r}} T^x_{\boldsymbol{r}+\boldsymbol{e}_x}$ will vary as $\cos^2\theta$ whereas that associated with the vertical bonds varies as $\sin^2\theta$. As $J_x = J_y = J$ in the isotropic system and as $\sin^2\theta + \cos^2\theta = 1$, any uniform pseudo-spin state will have the same energy and global rotations will not alter this energy.

3 Kitaev's honeycomb model

In 2006, Alexei Kitaev introduced a type of compass model that has interesting topological properties and excitations, which are relevant and much studied in the context of topological quantum computing [45]. The model is defined on a honeycomb lattice and is referred to either as *Kitaev's honeycomb model* or the *XYZ honeycomb compass model*. The lattice links on a honeycomb lattice may point along three different directions, see Fig. 3. One can label the bonds along these directions by $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$, where the angle between the three unit lattice vectors is 120°. With these preliminaries, the Kitaev's honeycomb model Hamiltonian $H^{Kitaev}_{\hexagon}$ reads

$$H^{Kitaev}_{\hexagon} = -J_x \sum_{\boldsymbol{e}_1-\text{bonds}} \tau^x_i \tau^x_j - J_y \sum_{\boldsymbol{e}_2-\text{bonds}} \tau^y_i \tau^y_j - J_z \sum_{\boldsymbol{e}_3-\text{bonds}} \tau^z_i \tau^z_j$$

One can re-express this model in the form of $H_{compass}$ introduced above, where

$$H^{Kitaev}_{\hexagon} = -\sum_{\boldsymbol{r},\gamma} J_\gamma \tau^\gamma_{\boldsymbol{r}} \tau^\gamma_{\boldsymbol{r}+\boldsymbol{e}_\gamma} \qquad \text{with} \qquad \begin{cases} \{\tau^\gamma\} = \{\tau^x, \tau^y, \tau^z\} \\ \{J_\gamma\} = \{J_x, J_y, J_z\} \\ \boldsymbol{e}_\gamma = \boldsymbol{e}_x \cos\theta_\gamma + \boldsymbol{e}_y \sin\theta_\gamma \\ \{\theta_\gamma\} = \{0, 2\pi/3, 4\pi/3\} \end{cases} \tag{17}$$

It was proven that for large J_z, the model Hamiltonian $H^{Kitaev}_{\hexagon}$ maps onto a square lattice model known as *Kitaev's toric code model* [16].

3.1 Features of Kitaev's honeycomb model

By its very nature, Kitaev's honeycomb model is very similar to the 90° compass models and other 120° models [4]. However, the Kitaev-model system has a number of very remarkable properties. These can be assessed in a crisp manner because the model is exactly solvable: it can be mapped exactly onto a system of non-interacting Majorana (as well as Dirac) fermions, as will be detailed in Sec. 3.2.2. This allows the derivation of all of the beautiful topological characteristics – its gapped bulk states, computable Chern numbers and Majorana excitations. Moreover, it will make evident that these Majorana excitations are coupled to a gauge field which embodies the topological charges, i.e., magnetic and electric like charges.
For future purposes it is useful to define an extension to this Hamiltonian $H^h_{\hexagon}$, which actually becomes relevant if the model is studied in an external field h. This term involves three pseudo-spins on sites i, j and k, and is the of form

$$H^h_{\hexagon} = -\kappa \sum_{ijk} \tau^x_i \tau^y_j \tau^z_k \tag{18}$$

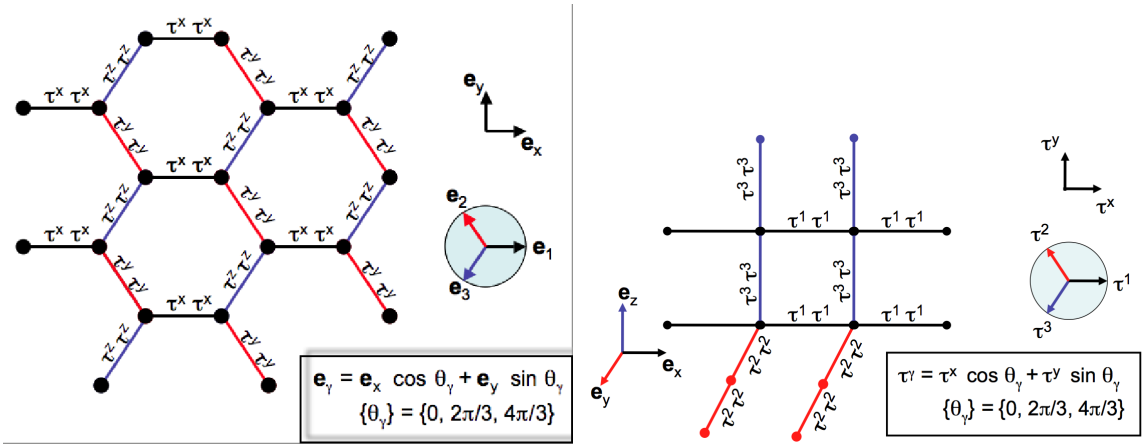


Fig. 3: *Left: Kitaev's compass model on a honeycomb lattice: the interaction of (pseudo-)spin degrees of freedom* $\boldsymbol{\tau} = (\tau^x, \tau^y, \tau^z)$ *along the three bonds that each site is connected to are* $\tau^x_{\boldsymbol{r}}\tau^x_{\boldsymbol{r}+\boldsymbol{e}_1}$, $\tau^y_{\boldsymbol{r}}\tau^y_{\boldsymbol{r}+\boldsymbol{e}_2}$ *and* $\tau^z_{\boldsymbol{r}}\tau^z_{\boldsymbol{r}+\boldsymbol{e}_3}$, *where the bond-vectors of the honeycomb lattice* $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ *are* $\{\boldsymbol{e}_x, (-\boldsymbol{e}_x+\sqrt{3}\boldsymbol{e}_y)/2, (-\boldsymbol{e}_x-\sqrt{3}\boldsymbol{e}_y)/2\}$, *respectively. Right: The 120° compass model on a cubic lattice: the interaction of (pseudo-)spin degrees of freedom* $\boldsymbol{\tau} = (\tau^x, \tau^y, \tau^z)$ *along the three bonds that each site is connected to are* $\hat{\pi}^1_{\boldsymbol{r}}\hat{\pi}^1_{\boldsymbol{r}+\boldsymbol{e}_x}$, $\hat{\pi}^2_{\boldsymbol{r}}\hat{\pi}^2_{\boldsymbol{r}+\boldsymbol{e}_y}$ *and* $\hat{\pi}^3_{\boldsymbol{r}}\hat{\pi}^3_{\boldsymbol{r}+\boldsymbol{e}_z}$, *where the different components* $\{\hat{\pi}^1, \hat{\pi}^2, \hat{\pi}^3\}$ *of the vector* $\hat{\boldsymbol{\pi}} = (\tau^x, (-\tau^x+\sqrt{3}\tau^y)/2, (-\tau^x-\sqrt{3}\tau^y)/2)$ *interact along the different bonds* $\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$.

where the sum over ijk is a sum over *all* sites connected by the two links $\langle ij\rangle$ and $\langle jk\rangle$. So here the link $\langle ij\rangle$ connects neighboring sites i and j, similarly for $\langle jk\rangle$, but sites i and k are *next* nearest neighbors. This form of the Hamiltonian might seem rather particular at this point, but when adding it, the model will stay exactly solvable. This term is essential in order to endow the non-Abelian excitations of Kitaev's honeycomb model with a gap. The Kitaev model reduces to the *toric code model* in the limit in which one coupling constant is far larger than all of the rest, e.g., $|J_z| \gg |J_{x,y}|$. The excitations in the toric code model, precisely have magnetic and electric charges.

3.1.1 Majorana excitations

The existence of edge-states in the Kitaev model constitutes an analogue to quantum Hall systems and other topological insulators. However, in integer quantum Hall systems, the edge-modes are bona fide fermions and not Majorana fermions. It is the Majorana character of the excitations that in principle enables the aforementioned fault tolerance relative to *all* local fluctuations – "errors" in the setting of quantum computing. The excitations of the Kitaev model flesh out the notions of anyonic statistics and afford very crisp realizations of non-trivial topology. The system also realizes one of the simplest examples of exotic ideas concerning fractionalization in strongly correlated electronic and spin systems. In its Abelian phase, the *magnetic* and *electric* excitations in the model may, respectively, be viewed [47] as counterparts of *vison* and *spinon* excitations in theories of doped quantum antiferromagnets [48] with relative "semionic" statistics which requires that when an excitation of one type is moved around another it picks up a phase factor of -1.

It should be stressed that while the existence of excitations of Majorana-type is a special feature of the Kitaev model, it is not necessarily a unique feature. In special situations three dimensional topological insulators may also exhibit Majorana fermion type of excitations, for instance on their surface when placed at an interface with a superconductor [49]. Majorana fermions may also manifest in some of the systems that we earlier referred to in the context of non-trivial statistics: the fractional quantum Hall systems such that of the state of filling fraction $\nu = 5/2$ [50], at cores of half-vortices in p-wave superconductors [51] and in semi-conductor [52, 53] and semi-conductor/(s-wave) superconductor systems [54].

3.2 Majorana representation – Abelian phases

As was emphasized earlier, the Kitaev model is exactly solvable in its ground state sector, for any set of coupling constants J_x, J_y and J_z. The original solution in [45] hinged on introducing several Majorana fermion degrees of freedom per site and making a projection on to a physical Hilbert space and symmetrization. Later approaches invoked a Jordan-Wigner (JW) transformation in two dimensions [55, 35, 56, 57], perturbative methods, e.g., [58] and slave fermion methods [59, 60]. Another approach, which will be followed here, is based on the direct use of a *bond algebra* [61]. It is rather straightforward and keeps directly track of the local symmetries that the Hamiltonian harbors, which are crucial to the solutions of $H^K_{\hexagon}$ (and the same model augmented by $H^h_{\hexagon}$). The explicit solution via the JW transformation [35] largely inspired the bond algebraic approach, but it is not as direct. The advantage of the bond algebraic method is that it enables the solution without enlarging the Hilbert space and making subsequent projections. Nor does it use at intermediate steps non-local string operators as in the Jordan-Wigner transformation.

3.2.1 Bond algebra, symmetries, and anyonic charge

In the Kitaev Hamiltonian $H^K_{\hexagon}$ three types of bonds $\{b_{jk}\}$ appear

$$\tau^x_j \tau^x_{j+\boldsymbol{e}_1},\ \tau^y_j \tau^y_{j+\boldsymbol{e}_2} \text{ and } \tau^z_j \tau^z_{j+\boldsymbol{e}_3}, \tag{19}$$

where $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ are unit vectors along the three directions of the hexagonal lattice. In terms of bond operators the Hamiltonian is

$$H^K_{\hexagon} = \sum_{\langle jk \rangle} J_{jk} b_{jk}, \tag{20}$$

with $J_{jk} = J_x, J_y$ or J_z depending on the orientation of bond $\langle jk \rangle$ along one of the three directions. One usually supplements this definition of the bond-Hamiltonian with an ordering convention of the bonds, the simplest one being that site j always lies below site k in the honeycomb lattice as for instance shown in Fig. 4. The pseudo-spin operators anticommute at any given site j, e.g., $\{\tau^x_j, \tau^z_j\} = 0$, and commute at different sites, e.g., $[\tau^x_j, \tau^z_p] = 0$ for any two sites $j \neq p$. The bonds therefore satisfy an extraordinarily simple algebra [61]:

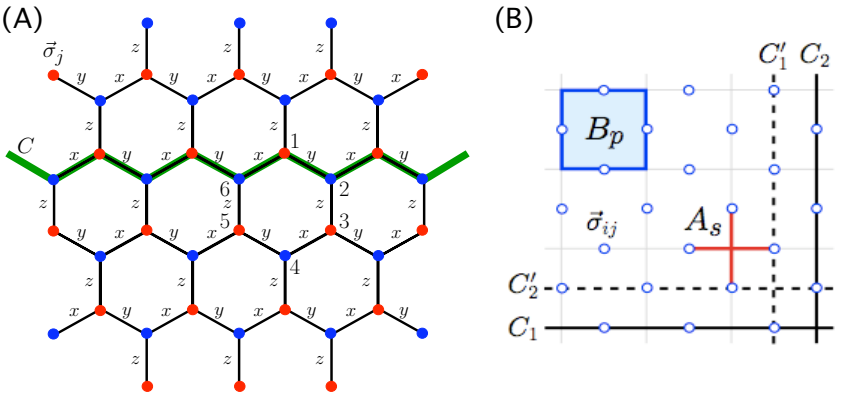


Fig. 4: *(A) Kitaev's model on a honeycomb lattice and three types of bonds. On each vertex there is an $S = 1/2$ degree of freedom indicated by a Pauli matrix $\vec{\sigma}_j$ (see text). (B) Elementary plaquette B_p and star A_s interaction terms in Kitaev's Toric code model. Hollow circles in the bonds (links) represent an $S = 1/2$ degree of freedom, while thick (dashed or solid) lines represent topological ($d = 1$) symmetry operators (see text). [46]*

(i) The square of each bond is one.

(ii) Two bonds that do not share any common site commute.

(iii) Two bonds that share one common site anti-commute.

There are no additional algebraic relations that the bonds that appear in the Hamiltonian $H^K_{\hexagon}$ need to satisfy. This set of all algebraic relations between the bonds in a general Hamiltonian is termed the *bond algebra* [12, 62, 7]. If we can write down another representation of the bonds in Eq. (19) for which all of the above algebraic relations are the same, then the Hamiltonian in the new representation and the original one will share the same spectrum and are thus related by a unitary transformation (and are thus dual to one another). Precisely such a change of representation underlies the exact solution of $H^K_{\hexagon}$ (as further elaborated on in subsections 3.2.2 and 3.2.3). Similar dualities (including those that lead to an effective dimensional reduction) can be established in numerous other compass models, e.g., [63, 64, 12, 62, 7, 65, 66, 34, 27].
We now pause to examine the symmetries of the Hamiltonian $H^K_{\hexagon}$. Exact local ($d = 0$) gauge symmetries are given by *products of pseudo-spins around each hexagon* [45]. For each hexagon i labeled by $\hexagon i$ as in Fig. 5, such a symmetry is given by

$$\hat{O}_{\hexagon i} = \tau_1^z \tau_2^x \tau_3^y \tau_4^z \tau_5^x \tau_6^y . \tag{21}$$

These local symmetries are *Ising gauge* symmetries (the square of these symmetry operators is identically equal to one and all of the symmetry operators for different hexagons commute with one another).
Each of the six sites of the hexagon contributes only one component τ^γ of its pseudo-spin operator to the product $\hat{O}_{\hexagon i}$, where γ is either x, y or z. Precisely which component of these three depends on the type of link that is *not* part of the hexagon – if on site j the bond operator

on the "non-hexagon link" is of type $\tau_j^\gamma \tau_{j+\boldsymbol{e}}^\gamma$ (thus with $j \in \hexagon i$ and $j+\boldsymbol{e} \notin \hexagon i$), the pseudo-spin component appearing in $\hat{O}_{\hexagon i}$ is τ_j^γ.
It can readily be verified that $\hat{O}_{\hexagon i}$ commutes with *any* bond-operator b_{jk} of Eq. (19) and consequently $[H_{\hexagon}^K, \hat{O}_{\hexagon i}] = 0$. These operators also mutually commute with one another: $[\hat{O}_{\hexagon i}, \hat{O}_{\hexagon j}] = 0$. Moreover the square of each such symmetry operator is one: $\hat{O}_{\hexagon i}^2 = 1$. When it attains a non-trivial eigenvalue, i.e., $\hat{O}_{\hexagon i} = -1$, the operator $\hat{O}_{\hexagon i}$ is said to depict an *anyonic charge* or *vorticity* on hexagon i, for reasons which will become clear later.
From the above follows that the system is composed of 2^{N_h} sectors with $N_h = N/2$ being the number of hexagons. Each sector is specified by the set of eigenvalues of the operators $\{\hat{O}_{\hexagon i}\}$, where $i = 1, ..., N_h$: $|O_{\hexagon 1} = \pm 1, O_{\hexagon 2} = \pm 1, ..., O_{\hexagon N_h} \pm 1\rangle$.
The model has more symmetries. When the system is placed on a torus, $H_{\hexagon}^K$ also has $d = 1$ symmetries, using the classification of symmetries of Section 2. For any loop C that spans the entire system the symmetry given by $\prod_{j \in C} \tau_j^\gamma$, where on each site j the component γ is determined by the character of one bond of site j that is not on C (i.e., the bond $\tau_j^\gamma \tau_{j+\boldsymbol{e}}^\gamma$ with $j \in C$ and $j+\boldsymbol{e} \notin C$). When C is for instance taken to be the zig-zag contour shown in Fig. 4 this symmetry is $\prod_{j \in C} \tau_j^z$, but actually any *closed* loop C represents a symmetry.

3.2.2 Majorana representation and fermionization

The relations (i)-(iii) of the previous section define the bond algebra of $H_{\hexagon}^K$ and it can readily be checked that they are also satisfied by the following substitution for the bonds in Eq. (19):

$$b_{jk} = 2i\eta_{jk} c_j c_k, \tag{22}$$

where the operators c_j represent *Majorana fermions*, obeying the Majorana algebra and η_{jk} are Ising-type gauge links: a number that is either $+1$ or -1 on any given link $\langle jk \rangle$. Since the Majorana fermions anticommute, we will choose the ordering convention such that the site j always lies below k. The set $\{\eta_{jk}\}$ encompassing all bonds constitutes a sector of gauge links. In any given sector $\{\eta_{jk}\}$, the Hamiltonian of Eq. (20) is quadratic in the Majorana fermions $\{c_i\}$ and *thus exactly solvable* [45, 61]. The local ($d = 0$) symmetries of Eq. (21) can be expressed in terms of the bonds as

$$\hat{O}_{\hexagon i} = \prod_{jk \in \hexagon i} \eta_{jk}. \tag{23}$$

That is, each sector of fixed $\{\eta_{jk}\}$ is an eigenstate of the symmetry operators of Eq. (21) with an eigenvalue that is determined only by $\{\eta_{jk}\}$. In $\hat{O}_{\hexagon i}$, as one multiplies η_{jk} for all links $\langle jk \rangle$ that are in the hexagon i, one keeps the bond indices j and k ordered with the previously chosen convention of j being below k.
The expression for $\hat{O}_{\hexagon i}$ above highlights the similarity between the local gauge symmetries in this system and such general symmetry (and fluxes) elsewhere. For instance, in a lattice version of electromagnetism, Eq. (23) relates to an Aharonov Bohm like phase. In the current context, Eq. (23) relates to the Ising version of such a phase ($O_{\hexagon i} = \pm 1$).
As each site belongs to three hexagons and each hexagon contains six sites, the number of hexagons is half the number of lattice sites ($N_h = N/2$). Thus to account for all eigenvalues of

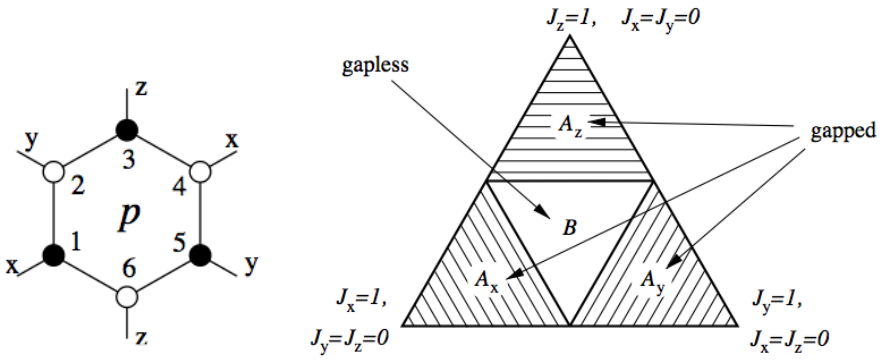


Fig. 5: *Left: Pictorial rendition of the local symmetry of Eq. (21) associated with every hexagon. Right: Phase diagram of the honeycomb Kitaev model. The triangle is the section of the positive octant* $(J_x, J_y, J_z) \geq 0$ *by the plane* $J_x+J_y+J_z = 1$. *The diagrams for the other octants are similar [45].*

the operators $\{O_{\hexagon_i}\}$, it suffices to allow the $N/2$ degrees of freedom η_{jk} on, for instance, all *vertical* bonds along $\boldsymbol{e}_3$ to attain a value of ± 1 and to pin η_{jk} on all other bonds (those along the $\boldsymbol{e}_1$ or $\boldsymbol{e}_2$ axis) to be 1. With this particular choice of the local gauge fields η, $\hat{O}_{\hexagon_i}$ in Eq. (23) reduces to the product of η_{jk} on the two vertical links that belong to each hexagon i.

The dimensionality of the original Hilbert space of N pseudo-spins is 2^N. Thus in each of the $2^{N/2}$ sectors of η_{jk}, there is a remaining Hilbert space of size $2^{N/2}$ on which the Majorana fermions are defined. One representation for the N Majorana fermions is in terms of $N/2$ spinless Dirac fermions. This may be explicitly done here by setting

$$c_j = d_{jk}+d_{jk}^\dagger \quad \text{and} \quad c_k = -i\big(d_{jk}-d_{jk}^\dagger\big), \tag{24}$$

with d_{jk} a spinless Dirac Fermi operator on the vertical link $\langle jk\rangle$ (that is, $k = j+\boldsymbol{e}_3$) [35]. The centers of the vertical links of the honeycomb lattice form a square lattice. It is therefore convenient to place the Fermi operators d_{jk} and $d_{jk}^\dagger$) at the centers of the vertical links $\langle jk\rangle$) and henceforth denote these by r, leaving us with the operators $d_r^\dagger$, d_r and the Ising degrees of freedom η_r. Denoting the unit vectors of the resulting square lattice by $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$, the Kitaev Hamiltonian reduces to

$$H_{\hexagon}^K = J_x\sum_r \big(d_r^\dagger+d_r\big)\big(d_{r+\boldsymbol{e}_x}^\dagger-d_{r+\hat{e}_x}\big)+J_y\sum_r \big(d_r^\dagger+d_r\big)\big(d_{r+\hat{e}_y}^\dagger-d_{r+\hat{e}_y}\big)+J_z\sum_r \eta_r\big(2d_r^\dagger d_r-1\big). \tag{25}$$

The last term constitutes an analogue of a "minimal coupling" term between gauge (i.e., the fields η_r on vertical links emanating from sites r) and matter (fermionic) degrees of freedom that is familiar from electromagnetism – in this specific case, an analogue of a coupling between the charge (or matter) density and an electrostatic-type potential.

An advantage of the fermionization procedure employed above is that it does not require the use of elaborate non-local JW transformations. That the representation in terms of spinless fermions is $2^{N/2}$ dimensional can be checked by realizing that there are $N/2$ vertical links $\langle jk\rangle$ and the dimensionality of each spinless Fermion operator is two: the bond $\langle jk\rangle$ can be either occupied or un-occupied by a fermion. Putting all of the pieces together, one sees that the problem of solving $H_{\hexagon}^K$ has now been reduced to a problem involving solely fermions and Ising gauge degrees of

freedom η_r, which at each site r can only attain the value ± 1. All excitations that appear in this system can be expressed in terms of the original spin variables $\boldsymbol{\tau}_j$ or, equivalently, in terms of fermions and Ising gauge fields. The *fusion rules* that will appear both in this system and its non-Abelian extension that we will review in Sec. 3.4.2 must relate to *fermionic* and *Ising gauge* type basic degrees of freedom.

3.2.3 Ground state of fermionized model

Within the ground state sector, for all hexagons all $\hat{O}_{\hexagon_i} = 1$, or equivalently on the square lattice $\eta_r = 1$ for all sites r. That the ground state must be vortex free is ensured by a corollary of a theorem due to Lieb [67] and has also been established numerically [45]. In momentum space, the fermionized Hamiltonian of Eq. (25) assumes the form

$$H^K_{\hexagon} = \sum_{\boldsymbol{q}} \varepsilon_q \, d^\dagger_{\boldsymbol{q}} d_{\boldsymbol{q}} + i \frac{\Delta_{\boldsymbol{q}}}{2} \big(d^\dagger_{\boldsymbol{q}} d^\dagger_{-\boldsymbol{q}} + d_{\boldsymbol{q}} d_{-\boldsymbol{q}} \big), \tag{26}$$

where $\boldsymbol{q} = (q_x, q_y)$ and

$$\varepsilon_{\boldsymbol{q}} = 2J_z - 2J_x \cos q_x - 2J_y \cos q_y, \quad \Delta_{\boldsymbol{q}} = 2J_x \sin q_x + 2J_y \sin q_y. \tag{27}$$

Interestingly, this Hamiltonian has the form of a *p-wave BCS type Hamiltonian on the square lattice* [35], which becomes explicit when the Hamiltonian is cast in the form of a Bogoliubov-de Gennes (BdG) Hamiltonian

$$H^K_{\hexagon} = \begin{pmatrix} d^\dagger_{\boldsymbol{q}} & d_{-\boldsymbol{q}} \end{pmatrix} H^K_{BdG}(\boldsymbol{q}) \begin{pmatrix} d_{\boldsymbol{q}} \\ d^\dagger_{-\boldsymbol{q}} \end{pmatrix}, \tag{28}$$

where $H^K_{BdG}(\boldsymbol{q})$ is a 2×2 matrix. It can be cast in the slightly more general form

$$H_{BdG}(\boldsymbol{q}) = h_{\boldsymbol{q}} \sigma_x + \Delta_{\boldsymbol{q}} \sigma_y + \varepsilon_{\boldsymbol{q}} \sigma_z = \boldsymbol{d}(\boldsymbol{q}) \cdot \boldsymbol{\sigma}, \tag{29}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ with Pauli matrices $\sigma_{x,y,z}$ and the last line defines the three-component vector $\boldsymbol{d}(\boldsymbol{q})$. Here an extra coupling $h_{\boldsymbol{q}}$ has been introduced for future reference. This coupling is not present within the pure honeycomb Kitaev model. Thus, $H^K_{BdG} = \lim_{h_{\boldsymbol{q}} \to 0} H_{BdG}(\boldsymbol{q})$.
In the Hamiltonian H_{BdG}, the vector $\boldsymbol{d}(\boldsymbol{q})$ acts as a "Zeeman field" applied to the "spin" $\boldsymbol{\sigma}$ of a two-level system. All of its eigenvalues come in pairs, at energies

$$E_{\boldsymbol{q}} = \pm d(\boldsymbol{q}) = \pm |\boldsymbol{d}(\boldsymbol{q})| = \pm \sqrt{\boldsymbol{d}(\boldsymbol{q}) \cdot \boldsymbol{d}(\boldsymbol{q})}. \tag{30}$$

Diagonalizing the Hamiltonian by a Bogoliubov transformation,

$$\gamma_{\boldsymbol{q}} = u_{\boldsymbol{q}} d_{\boldsymbol{q}} + v_{\boldsymbol{q}} d^\dagger_{-\boldsymbol{q}}, \tag{31}$$

with $|u_{\boldsymbol{q}}|^2 + |v_{\boldsymbol{q}}|^2 = 1$ and $|u_{\boldsymbol{q}}|^2 = \frac{1}{2}\sqrt{1 + \varepsilon_{\boldsymbol{q}}/E^2_{\boldsymbol{q}}}$, yields the energy spectrum

$$E_q = \pm \sqrt{\varepsilon_q^2 + |\tilde{\Delta}_q|^2}. \tag{32}$$

Note that no effective chemical potential appears here (i.e., $\mu = 0$). Thus, within the ground state, all fermionic states of negative energy ($E_q < 0$) are occupied while all states of positive energies are empty. The general ground state wavefunction corresponding to Eq. (32) is given by

$$|g\rangle = \prod_{\boldsymbol{q}} \left(u_{\boldsymbol{q}} + v_{\boldsymbol{q}} d^{\dagger}_{\boldsymbol{q}} d^{\dagger}_{-\boldsymbol{q}}\right)|0\rangle. \tag{33}$$

As mandated by particle-hole symmetry, non-zero eigenvalues of the BdG Hamiltonian of Eqs. (28) and (29) appear in *pairs*, at energies $\pm\varepsilon$ with $\gamma^{\dagger}_{-\varepsilon} = \gamma_{\varepsilon}$. Thus, generally, a fermion excitation eigenstate corresponds to two solutions of the BdG equations. At the energy $\varepsilon = 0$, the system is invariant under particle-hole symmetry, and a Majorana fermion appears (here $\gamma^{\dagger}_0 = \gamma_0$).

Whenever both ε_q and $\tilde{\Delta}_q$ vanish then a *Dirac node* may appear (i.e., the energy E_q will disperse linearly about its vanishing value at that point). Conversely, if for given J_x, J_y and J_z values there are no simultaneous solutions to the two conditions $\varepsilon_q = \tilde{\Delta}_q = 0$ then the spectrum will be gapped. In the vicinity of band extrema, the dispersions of Eq. (33) is parabolic when a gap appears between the two bands in the problem (i.e., $\min\{|E_q|\} > 0$) and, as just noted, is linear near the zeros of E_q when the system is gapless. Elementary calculations illustrate that when the moduli of the couplings $|J_x|, |J_y|$ and $|J_z|$ are such that they satisfy the triangle inequalities then the spectrum of Eq. (32) will be gapless. When the couplings violate the triangle inequalities, a gap emerges. The ground state of Eq. (33) corresponds to a BCS condensate. In [35], real space ground state wave-functions $|\Psi_0\rangle$ were explicitly constructed in the original spin representation in closed forms that do not require any implicit projections.

3.2.4 Gapless and gapped phases

To provide a better understanding of the spectrum, we focus on a particular set of couplings. At the symmetric system ($J_x = J_y = J_z$), the dispersion of Eq. (33) is, within the first Brillouin zone, zero at $\boldsymbol{q}^{(+)} \equiv \boldsymbol{K}_1/3 + 2\boldsymbol{K}_2/3, \boldsymbol{q}^{(-)} = 2\boldsymbol{K}_1/3 + \boldsymbol{K}_2/3$ where $\boldsymbol{K}_{1,2}$ denote the reciprocal lattice vectors along 1 and 2 directions [45] (and the equality holds modulo the addition of any reciprocal vectors). As the anisotropy of the coupling constants is increased (e.g., setting $|J_z|$ fixed and decreasing $|J_{x,y}|$), the two points $\boldsymbol{q}^{(\pm)}$ veer towards one another until they merge at the boundary between the gapped and gapless phases [45]. Beyond this point, as the $|J_{x,y}|$ are further decreased, and the system is in its gapped phase there are no real vectors $\boldsymbol{q}$ for which E_q is zero.

Similarly to the calculations above, the spectrum may be computed in other sectors of the Ising vortices. By looking at the algebra of the bonds appearing in the Hamiltonian, it is immediately clear that the spectrum is invariant under a change of sign of any of the exchange constants $J_\alpha \to -J_\alpha$. This, along with an overall global scale invariance of the gapless/gapped parameter regions under a uniform scaling of all coupling parameters $J_{x,y,z} \to cJ_{x,y,z}$ with c a constant enables us to delineate the boundaries of the gapless and gapped regions of the model. Such a phase diagram of the system is provided in Fig. 5 [45]. The existence of transitions between these phases are "topological" and as such cannot be discerned by any standard local measurements.

The inability of local measurements to discern between different phases underlies systems with topological order [16, 28, 34]. When expressed in terms of the basic spin degrees of freedom, anyons involve extended non-local lines. Amongst other probes, an interesting signature of the topological transitions between the Abelian phases in Kitaev's honeycomb model is afforded by quantum information theory measures, in particular the mutual information [68].
The condition for a gapless phase is tantamount to the triangle inequality via Eq. (27). This is so since (as alluded to in Section 3.2.3) the gapless phase implies $\varepsilon_q = \Delta_q = 0$ which in turn implies from Eq. (27) (via the law of cosines) that one can view q_x and q_y as angles in the triangle formed by the sides $\{J_x, J_y, J_z\}$ [35]. The relation between Kitaev's honeycomb model and the p-wave type pairing in Eqs. (25), (26) and (27) was further elucidated in several insightful works, [69, 70].

3.3 Braiding statistics

The Majorana fermion representation of Eq. (20) highlights another important property of this system – the braiding statistics formed by displacing one string of bonds around a closed loop. The product of bonds along any contour (open or closed) commutes with all other string products of the same form, including the symmetries O_h of Eqs. (21,23). That is, for any closed contour C drawn on the lattice, the operators

$$O_C = \prod_{ij \in C} b_{ij} \tag{34}$$

commute amongst themselves. For finite sized contours C, one has a sort of Stokes' theorem. That is, the symmetries of Eq. (34) can be written as [35]

$$O_C = \prod_{h \in C} O_h, \tag{35}$$

with the product taken over all hexagons h that are enclosed by the loop C. The right-hand side of Eq. (35) corresponds to the total anyonic charge enclosed by C. If an odd number of anyons (hexagons h for which $O_h = -1$) is circumscribed by C then $O_C = -1$. *This minus sign is the origin of the anyonic nature* of the braiding operations in Kitaev's model, e.g., [35, 71], in its gapped phase (known as the "A phase"). In the gapless phase of couplings of Kitaev's model ("B phase") the statistics of the vortices is ill defined. However, as will be elaborated on later, augmenting an additional external magnetic field term leads to the opening of a gap in the B phase. Within this gapped regime, the vortices exhibit well defined non-Abelian statistics.
An immediate corollary of the local ($d = 0$) symmetries O_h of Eqs. (21) and (23), is that, by Elitzur's theorem, only correlation functions that are invariant under all of these symmetries may attain a non-zero expectation value at finite temperatures [35]. Thus, all non-zero *correlation functions are composed of products of bonds (along closed or open contours) as in Eqs. (34) and (36).* Similar considerations also apply at zero-temperature. These considerations generally lead to string and "brane" type correlation functions. While the above considerations revolve around symmetries of the spins alone, an earlier work illustrated, by the use of Majorana fermions, that

all two-point correlation functions apart from those that form bonds vanish within the ground state and further related interesting consequences [72].
In the gapped phase, these string correlation functions are exponentially damped in spatial distance between the endpoints of the string (with a similar behavior concerning dynamic correlations). In the gapless phase, the string correlation functions decay algebraically in the distance. The striking difference between the slow algebraic vis-à-vis the rapid exponential decay of correlations is not limited only to the non-local string (or brane) correlation functions [35].

3.3.1 Fermion excitation and translation

The fermionization procedure discussed above enables the construction of anyons out of string operators [35]. A feature directly related to the symmetries of Eqs. (34) and (35) is that it is possible to create "fermionic" excitations alone sans anyons. One way to see this is by invoking symmetry and bond algebra arguments once again. Towards this end, one may consider the string product of bonds of the form of Eq. (34) yet now for *an open contour* Γ (as opposed to the closed contour C). That is, we may define the operator

$$O_\Gamma = \prod_{ij\in\Gamma} b_{ij} \tag{36}$$

along an open contour Γ. For the purposes of what follows, let us label the end points of Γ by U and V. Unlike O_c of Eq. (34), the operator of Eq. (36) is not a symmetry. That is, the operator O_Γ serves as a trivial symmetry for all bonds b_{kl} for which (1) $k, l \neq U$ or V and/or (2) lie along $\Gamma : k, l \in \Gamma$. For all such bonds, $[b_{kl}, O_\Gamma] = 0$. The above includes all bonds b_{kl} that have any number of sites along Γ (i.e., 0,1, or 2) such that the bonds do not touch Γ only at one point with that point being one of the endpoints U or V. However, if $k = U$ or V and $l \notin \Gamma$ or vice versa (i.e., $l = U$ or V and $k \notin \Gamma$) then b_{kl} will anti-commute with O_Γ:

$$\left\{O_\Gamma,\, b_{kl}\right\} = 0. \tag{37}$$

There are four such bonds b_{kl}. All other bonds commute with the operator of Eq. (36), $[O_\Gamma, b_{mn}] = 0$. As the exact solution that was outlined earlier [Eqs. (27), (33)] shows, the ground state sector of Kitaev's model is not highly degenerate. As O_Γ flips the energetic contributions of the four bonds b_{kU} and b_{Vl} that touch the endpoints U and V, all of this suggests that the application of general O_Γ (there is an exponentially large number of contours Γ) on a ground state cannot give back a ground state but rather must excite the system. The bonds at the end points of the contour Γ have been modified (by a change of sign) as a result of the anticommutation relation of Eq. (37) and together the four disrupted bonds at the two endpoints U and V that do not lie along Γ sum to yield a higher energy state. Thus, it is natural to associate defects created by the string operator of Eq. (36) at the endpoints U and V of the contour Γ. As seen from Eqs. (20) and (36), the string operator O_Γ involves only the Majorana fermions and not the anyons of Eqs. (21), (23) and their composites Eq. (35). Indeed, it is possible to verify that as each of the bonds of the lattice, b_{ij} of Eqs. (19) and (20) commutes with all anyonic charges

O_h of Eqs. (21) and (23), the operator of Eq. (36) does not create (nor, in general, remove or displace) any anyons: $[O_\Gamma,\, O_h] = 1$. For a closed loop however, the closed string operator of Eqs. (34) and (35) is a symmetry. The existence of general $d = 1$ symmetry operators that are products of defect creation operators along loops [28] has similar incarnations elsewhere (e.g., in quantum Hall systems with the creation of quasi-particle/quasi-particle pairs). By creating defects and moving these defects along entire closed cycles, the defects annihilate and the system returns to its low energy (ground state) sector. Putting all of the pieces together, one sees that it is possible to have fermionic excitations (generated by Eq. (36)) alone. It is possible to express all of these results in terms of the fermions directly similar to [35].
When a fermion is transported around a closed loop that encircles a single Ising vortex (for which $O_h = -1$), we see that Eqs. (34) and (35) reduce to an overall phase factor of -1. Thus, in such an instance the quantum state is multiplied by this overall phase factor [45].

3.3.2 Vortex pair creation and translation

It is common to think about excitations formed by the application of single spin operators (i.e., by a rotation of a single spin) or by a product of two on the ground state. As pointed out by [73], there are subtleties associated with a simple interpretation of the action of these operations within the low energy sector. In what follows, we will focus on such an excitation via general symmetry and bond algebraic considerations. Towards this end, we consider a single vertical link (ij). We define, similarly to [45, 74, 75], the three operators $X = \sigma_i^x \sigma_j^x$, $Y = \sigma_i^x \sigma_j^y$, and $Z = \sigma_j^z$. These operators are different from those of Eq. (36) (including the case of a single two site bond). Each of these three operators anti-commutes with two bond operators. For instance, Z anticommutes with the two bonds (other than b_{ij}) that have j as one of their endpoints. Similarly, the operators X and Y each anti-commute with exactly two bonds. When acting on the ground state, the flipping operations incurred by any of the operators X, Y or Z may increase the system energy. It is furthermore readily verified that Y and Z may each flip the anyonic charges O_h of two hexagonal plaquettes while X flips the anyonic charges of all four hexagonal plaquettes that contain either the site i or j (or both). The flipping of any of the bonds generated by each of these three operators can be accounted for by inverting the sign of the η field along the corresponding link following Eq. (20). The three operators satisfy the $S = 1/2$ spin algebra:

$$\{X, Y\}{=}\{X, Z\}{=}\{Y, Z\} = 0, \;\; X^2{=}Y^2{=}Z^2 = 1, \text{ and } \; XY{=}iZ, \; YZ{=}iX, \; ZX{=}iY. \quad (38)$$

It is natural to associate "particles" with states $X|\psi\rangle, Y|\psi\rangle, Z|\psi\rangle$ created by the application of the operators X, Y or Z on the ground state wavefunction. The identities of Eq. (38) generally suggests that a fusion of two particles into a third might be possible. This is indeed the case as has been worked out in some detail in various approaches and limits (especially that of $J_z \gg J_{x,y}$ lying within the A phases of the system, see Fig. 5 [45, 75]. In that limit, the energy of the excitation $X|\psi\rangle$ is nearly equal to the sum of energies corresponding to $Y|\psi\rangle$ and $Z|\psi\rangle$.
As the anyonic charges of Eq. (21) are symmetries, anyonic excitations are massive. That is, an anyonic excitation is stationary as it is an eigenstate of the Hamiltonian. As discussed in [73],

it is possible to create anyons without fermions by the combined use of one and three spin operations on the ground state. We now extend the discussion of the single bond operators above and present the general vortex translation (or anyon) operator. An approach related to ours, along with a detailed analysis of energies, is given in [71]. An insightful analysis is also provided in [73]. In order to analyze the Ising vortex translation operators, we introduce an operator that is identical to that of Eq. (36) apart from all important end point corrections that allow it to be expressed as O_Γ multiplied by two operators corresponding to the two endpoints. Specifically, we consider an open contour Γ. For each non endpoint vertex $i \in \Gamma$, there is only a single neighbor l that is not on Γ. For the two end points of Γ ($i_1 = U$ and $i_2 = V$), there are two neighbors l that do not lie on Γ. One may choose any of these neighbors for the two endpoints in what follows. (We will mark the chosen neighbors for the endpoints by l_1 and l_2 respectively.) We denote the direction of a ray parallel to the nearest neighbor link $\langle il \rangle$ by γ (that may be either x, y, or z). We then construct the open contour operator

$$\mathcal{T}_\Gamma = \prod_{i \in \Gamma} \sigma_i^\gamma. \tag{39}$$

Eq. (39) is of nearly identical form to (36) for all non-boundary points i. However, in Eq. (36), the component of the boundary spin operators that appear in the string operator are set equal to the two directions $\gamma_{1,2} = \langle i_{1,2} j_{1,2} \rangle$ with $j_{1,2}$ being the nearest neighbors of $i_{1,2}$ that *lie on* Γ (i.e., "going backwards" away from the endpoints $i_{1,2}$). By contrast, in Eq. (39), the components of the spins at the two endpoints that appear in the string product are set by the two directions $\gamma = \langle i_{1,2} l_{1,2} \rangle$ (with $l_{1,2}$ not on Γ).

For the two hexagonal plaquettes $h^* = h_{1,2}$ that have a single vertex at one of the endpoints of i_1 or i_2 of Γ and that furthermore include one of the vertices l_1 or l_2, we have that

$$\mathcal{T}_\Gamma O_{h*} \mathcal{T} = -O_{h^*}. \tag{40}$$

In Eq. (40), O_{h*} denotes the vortex charge of a plaquette $h*$ that lies at an endpoint of Γ. Similar to the operator of Eq. (36), for all other plaquettes $h \neq h^*$, we have that $\mathcal{T}_\Gamma O_h \mathcal{T}_\Gamma = O_h$ (with no change in the vortex charge).

It is readily verified that the operator $\mathcal{T}_\Gamma$, albeit flipping the sign of two bonds attached to the endpoints of Γ, does *not alter* the bond algebra of all bonds (all non-neighboring bonds commute, neighboring bonds anticommute, and the square of any bond is 1). The sole change triggered by the application of $\mathcal{T}_\Gamma$ is that two bond pre-factors η are multiplied by a factor of -1, and correspondingly two vortex charges are flipped. Thus, the effect of $\mathcal{T}_\Gamma$ is to flip the sign of the two vortices at its endpoints.

If the system has a single vortex $O_{h_1} = -1$ at plaquette h_1 that has only one (endpoint) on Γ and furthermore contains one of the two points $l_{1.2}$, then the application of $\mathcal{T}_\Gamma$ with the contour Γ having a single point in the plaquette h_1 (the latter plaquette also containing the point l_1) as one of its endpoints will move the vortex to another plaquette h_2 that lies at the other end of the contour Γ (and contains the point l_2). That is, $\mathcal{T}_\Gamma$ is a *vortex translation operator*. If Γ forms a complete closed contour C along a toric cycle (when h_1 and h_2 are identified as the same point

on the torus) then, similarly to O_Γ of Eq. (36), $\mathcal{T}_\Gamma$ veers towards the $d = 1$ dimensional symmetry of Eq. (34). In the above, we established that the sole effect of $\mathcal{T}_\Gamma$ is to displace a vortex without influencing the system energy from any of the bonds that do not touch the endpoints of the contour Γ.

3.4 Broken time reversal symmetry – the non-Abelian phase

Kitaev's model for a wide range of couplings corresponds, as discussed earlier to a gapless phase, the so-called "B" phase. It is only in the "corners" of the phase diagram of Fig. 5 (the so-called "A" phase where the $\{J_x, J_y, Jz\}$ differ substantially from one another and cannot form the sides of a triangle) that a gap opens up. In the A phase, gapped Abelian anyons are present. Our focus in this section will be on the B phase where gapless excitations of Eq. (33) were found. By a modification of Kitaev's honeycomb model, gapped non-Abelian excitations can arise. There are various ways in which such excitations can arise. For instance, these may be triggered by the geometry of the lattice (via, e.g., a decoration of the lattice wherein each vertex of the hexagonal lattice is replaced by a triangle [76]). In what follows, we consider the original investigation of [45] in which a gapped phase with non-Abelian excitations originates from the application of an external magnetic field to a point (J_x, J_y, J_z) in the space of coupling constants for which the system would have been gapless if no field were applied.

In the context of the broad link to topological insulator physics and, in particular, to the symmetry classification of topological insulators [77–79, 5], in the absence of any additional perturbations, the free fermion Kitaev honeycomb model lies in the "BDI" symmetry class. In the presence of an external field time reversal breaking field, however, the symmetry becomes that of the "D" class raising the specter of a non-trivial insulator as indeed occurs in the nontrivial B phase of the extended Kitaev model. As we now review, such a time reversal symmetry breaking field gives rise to an effective *next nearest neighbor coupling* between Majorana fermions. This additional hopping leads to a gapped spectrum with non-Abelian chiral modes. When a magnetic field $\boldsymbol{h}$ is applied along the $[111]$ direction, i..e., when Eq. (17) is augmented by a Zeeman coupling

$$H' = H - \sum_i \boldsymbol{h} \cdot \boldsymbol{\tau}_i, \tag{41}$$

a gap opens up in the core region (B phase) of the phase diagram of Fig. 5. The (time reversal broken) phase that arises from the application of this field is very interesting.

In particular, non-Abelian anyons appear in the former gapless phase (which includes the symmetric point $J_x = J_y = J_z$). The Hamiltonian of Eq. (41) is not exactly solvable. It can, however, be treated perturbatively and (ignoring unimportant corrections) reduced to an exactly solvable system [45]. That is, the magnetic field term in Eq. (41) gives rise (with $\kappa \sim h_x h_y h_z / J^2$ in the symmetric point $J_x = J_y = J_z = J$) to a (time reversal symmetry breaking) term of the form of Eq. (18) which we write here anew,

$$H_h = -\kappa \sum_{ijk} \sigma_i^x \sigma_j^y \sigma_k^z, \tag{42}$$

for all triplets of sites (i, j, k) formed by the union of the two bonds, (ij) and (jk), that impinge on site j. The *product of the three spin operators* of Eq. (42) can be expressed as a *product of two neighboring bonds* of Eqs. (19) and (22) by use of the relation $c_j^2 = 1/2$. For instance, for (oriented) links (ij) and (jk) along the x and z directions respectively (with $i_z < j_z$ and $j_z < k_z$), the product of the bonds of Eq. (22) reads $b_{ij}b_{jk} = -2\eta_{ij}\eta_{jk}c_ic_k$. Eq. (42) is seen to reduce to a *Majorana fermion bi-linear linking (all) next nearest neighbor sites*. The Majorana fermion bi-linear (c_ic_k) resulting from the product of two bonds has a real prefactor $(-2\eta_{ij}\eta_{jk})$ as opposed to the imaginary prefactors that are associated with single nearest neighbor bonds in Eq. (22). This relative phase factor of i reflects the time reversal symmetry breaking of the perturbation. Time reversal symmetry breaking also allows for the existence of chiral modes wherein fermionic modes may preferentially propagate in one (clockwise or anti-clockwise) direction. For any pair of next nearest neighbor sites (ik) on the honeycomb lattice, there is a unique three site path (and two bond product) that leads to the bi-linear form c_ic_k. The quadratic character of these three-spin perturbations of Eq. (42) in the Majorana fermions (and similarly also in the fermions following, e.g., Eq. (24)) ensures that even when the system is augmented by these perturbations, the final Hamiltonian

$$H_{K_h;h} \equiv H_{K_h} + H_h \tag{43}$$

formed by the sum of Eqs. (17) and (42) is *still exactly solvable*.

3.4.1 Solution of extended model with broken time reversal symmetry

The solution to the problem is of a similar character to the one that earlier led to Eqs. (27) and (33). As each spin product of the type $\sigma_i^x\sigma_j^y\sigma_k^z$ is given by a product of two bonds (each of which commutes with all of the symmetries of Eq. (21), it follows that the perturbation of Eq. (42) commutes with the operators O_h. As before, in any given sector one can employ the representation of Eqs. (19) with η related to the flux via the condition of Eq. (23). All of the earlier steps taken in Eqs. (27) and (26) can thus be exactly reproduced. However, unlike the nearest neighbor Hamiltonian that we studied earlier in the absence of an applied external field h (or an effective), the next-nearest neighbor Fermi interactions lead to new non-trivial results. In particular, the perturbation set non-zero $\boldsymbol{h}$ allows the earlier gapless phase in the absence of a field to become gapped and thus to support anyons which within this phase are non-Abelian [45]. The spectrum of $H_{K_h;h}$, in the vortex free sector ($O_h = 1$ for all h) is then seen to be given by Eq. (32) where the real p-wave type gap Δ_q [35] of Eq. (27) is now replaced by the complex

$$\tilde{\Delta}_q = \Delta_q + 4i\kappa\big(\sin q_1 - \sin q_2 + \sin(q_2 - q_1)\big). \tag{44}$$

As can be seen by some simple analysis, the former gapless points $\boldsymbol{q}^{(\pm)}$ of Eq. (33) now acquire a gap when $\kappa \neq 0$. The p-wave type gap function $\tilde{\Delta}$ now becomes complex [35]. This suggests that the physics will essentially be the same as that for $(p+ip)'$ superconductors [51] which is indeed the case. It is noteworthy that even when the Hamiltonian is time reversal invariant the ground states may spontaneously break time reversal. Indeed, by Kramers' theorem, this must

occur whenever the system is defined on a hexagonal lattice with an odd number of spins [35]. In the B phase of Kitaev's model wherein the gap was borne by the perturbation, the associated Chern number $\nu = \pm 1$ and the aforementioned non-trivial statistics [45] with non-Abelian topological anyons. We elaborate on these anyons and their features next.

3.4.2 Non-Abelian anyons in Kitaev's model and their properties

To conform with standard practice, we use σ to denote a vortex (defined, similarly to the Abelian phase, by having the plaquette product O_h of Eqs. (21) and (23) be -1, $O_h = -1$), ε to mark a fermionic mode, and I to denote the vacuum (having no anyons). The fusion rules are then of the form

$$\varepsilon \times \varepsilon = I, \quad \sigma \times \varepsilon = \sigma, \text{ and } \sigma \times \sigma = I + \varepsilon, \tag{45}$$

augmented by the trivial statement that the fusion of any particle with the identity operator leads back to that particle. As in the case of the Abelian anyons each particle is its own anti-particle. The non-trivial character of the non-Abelian anyons rears its head in the last line of Eq. (45). Two vortices (σ) may fuse in two different channels to either annihilate (I) each other or to form a fermion (ε). The vortex operators of Eq. (21) have, as always, Ising eigenvalues $O_h = \pm 1$. Anyons that satisfy the relations of Eq. (45) are called "Ising anyons". Unlike the case of Abelian anyons, *fusing two anyons may lead to non-unique outcomes*. In particular, if particles a and b are fused when $a = b = \sigma$ then this may yield an ε particle or the vacuum (I). In the limit of spatially infinitely distant vortices, the fermionic spectrum as adduced from the square lattice Hamiltonian with η_r on the vertical links of original the honeycomb lattice set by the vortices O_h of Eq. (23), exhibits a multitude of fermionic *zero modes* [80]. Thus the hybrid of two well separated vortices (σ) may lead to a state in which the vortices annihilate to form the vacuum (I) or a "zero energy" fermionic state (ε). This degeneracy is lifted once the vortices become close to one another wherein the fermionic modes ε attain a finite energy cost (or "mass"). Repeated applications of the last of Eqs. (45) rationalizes the $2^{n_\sigma/2-1}$-fold degeneracy that is present in a system of n_σ (with this number being an even integer) well separated vortices [81]. In formal terms, the quantum dimension of the vortices σ is $d_\sigma = \sqrt{2}$; the system degeneracy for n_σ vortices scales as $d_\sigma^{n_\sigma}$. Due to the unique outcome of all of the other fusion rules in Eq. (45), the quantum dimensions of ε and I are $d_\varepsilon = d_I = 1$.

References

[1] K. Kugel and D. Khomskii, Sov. Phys. Usp. **25**, 231 (1982)

[2] G. Jackeli and G. Khaliullin, Phys. Rev. Lett. **102**, 017205 (2009)

[3] J. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **105**, 027204 (2010)

[4] Z. Nussinov and J. van den Brink, Rev. Mod. Phys. **87**, 1 (2015)

[5] S. Trebst, Phys. Rep. **950**, 1 (2017)

[6] Z. Nussinov and E. Fradkin, Phys. Rev. B **71**, 195120 (2005)

[7] Z. Nussinov and G. Ortiz, Europhysics Lett. **84**, 36005 (2008)

[8] V.W. Scarola, K.B. Whaley, and M. Troyer, Phys. Rev. B **79**, 085113 (2009)

[9] C. Xu and J.E. Moore, Nucl. Phys. B **716**, 487 (2005)

[10] C. Xu and J.E. Moore, Phys. Rev. Lett. **93**, 047003 (2004)

[11] J.E. Moore and D.-H. Lee, Phys. Rev. B **69**, 104511 (2004)

[12] E. Cobanera, G. Ortiz, and Z. Nussinov, Phys. Rev. Lett. **104**, 020402 (2010)

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

[14] F. Wegner, J. Math. Phys. **12**, 2259 (1971)

[15] J.B. Kogut, Reviews of Modern Physics **51**, 659 (1979)

[16] A.Y. Kitaev, Ann. Phys. **303**, 2 (2003)

[17] E. Fradkin and S.H. Shenker, Phys. Rev. D **19**, 3682 (1979)

[18] Z. Nussinov, Phys. Rev. D **72**, 054509 (2005)

[19] V.J. Emery, E. Fradkin, S.A. Kivelson, and T.C. Lubensky, Phys. Rev. Lett. **85**, 2160 (2000)

[20] A. Vishwanath and D. Carpentier, Phys. Rev. Lett. **86**, 676 (2001)

[21] E. Fradkin and S.A. Kivelson, Phys. Rev. B **59**, 8065 (1999)

[22] A.H. MacDonald and M.P.A. Fisher, Phys. Rev. B **61**, 5724 (2000)

[23] A. Paramekanti, L. Balents, and M.P.A. Fisher, Phys. Rev. B **66**, 054626 (2002)

[24] J.W.F. Venderbos, M. Daghofer, J. van den Brink, and S. Kumar, Phys. Rev. Lett. **107**, 076405 (2011)

[25] D. Gaiotto, A. Kapustin, N. Seiberg and B. Willett, J. High Energy Phys. **2015**, 172 (2015)

[26] C.D. Batista and Z. Nussinov, Phys. Rev. B **72**, 045137 (2005)

[27] Z. Nussinov, G. Ortiz, and E. Cobanera, Ann. Phys. **327**, 2491 (2012)

[28] Z. Nussinov and G. Ortiz, Proc. Nat. Acad. Sci. U.S.A. **106**, 16944 (2009)

[29] C. Batista and S. Trugman, Phys. Rev. Lett. **93**, 217202 (2004)

[30] Z. Nussinov, C.D. Batista, B. Normand, and S.A. Trugman, Phys. Rev. B **75**, 094411 (2007)

[31] Z. Nussinov and G. Ortiz, Phys. Rev. B **107**, 045109 (2023)

[32] Z. Nussinov, C. Batista, and E. Fradkin, Int. J. Mod. Phys. B **20**, 5239 (2006)

[33] S. Elitzur, Phys. Rev. D **12**, 3978 (1975)

[34] Z. Nussinov and G. Ortiz, Ann. Phys. **324**, 977 (2009)

[35] H.-D. Chen and Z. Nussinov, J. Phys. A **41**, 075001 (2008)

[36] D. Perez-Garcia, M.M. Wolf, M. Sanz, F. Verstraete, and J.I. Cirac, Phys. Rev. Lett. **100**, 167202 (2008)

[37] J. Villain, Solid State Comm. **10**, 967 (1972)

[38] E. Shender, Sov. Phys. JETP **56**, 178 (1982)

[39] C. Henley, Phys. Rev. Lett. **62**, 2056 (1989)

[40] Z. Nussinov, M. Biskup, L. Chayes, and J. van den Brink, Europhys. Lett. **67**, 990 (2004)

[41] M. Biskup, L. Chayes, and Z. Nussinov, Commun. Math. Phys. **255**, 253 (2005)

[42] B. Doucot, M. Feigel'man, L. Ioffe, and A. Ioselevich, Phys. Rev. B **71**, 024505 (2005)

[43] J. Dorier, F. Becca, and F. Mila, Phys. Rev. B **72**, 024448 (2005)

[44] F. Lin and V.W. Scarola, Phys. Rev. Lett. **111**, 220401 (2013)

[45] A.Y. Kitaev, Ann. Phys. **321**, 2 (2006)

[46] Z. Nussinov and G. Ortiz, Phys. Rev. B **77**, 064302 (2008)

[47] S. Sachdev, arXiv:0901.410

[48] T. Senthil and M.P.A. Fisher, Phys. Rev. B **63**, 134521 (2001)

[49] J. Linder, Y. Tanaka, T. Yokoyama, A. Sudbo, and N. Nagaosa, Phys. Rev. Lett. **104**, 067001 (2010)

[50] N. Read and D. Green, Phys. Rev. B **61**, 10267 (2000)

[51] D.A. Ivanov, Phys. Rev. Lett. **86**, 268 (2001)

[52] J. Alicea, Phys. Rev. B **81**, 125318 (2010)

[53] J. Alicea, Y. Oreg, G. Refael, F. von Oppen, and M.P.A. Fisher, Nat. Phys. **7**, 412 (2011)

[54] J.D. Sau, R.M. Lutchyn, S. Tewari, and S. Das Sarma, Phys. Rev. Lett. **104**, 040502 (2010)

[55] H.-D. Chen, C. Fang, J. Hu, and H. Yao, Phys. Rev. B **75**, 144401 (2007)

[56] X.-Y. Feng, G.-M. Zhang, and T. Xiang, Phys. Rev. Lett. **98**, 087204 (2007)

[57] G. Kells, J.K. Slingerland, and J. Vala, Phys. Rev. B **80**, 125415 (2009)

[58] J. Vidal, K.P. Schmidt, and S. Dusuel, Phys. Rev. B **78**, 245121 (2008)

[59] F.J. Burnell and C. Nayak, Phys. Rev. B **84**, 125125 (2011)

[60] R. Schaffer, S. Bhattacharjee, and Y.-B. Kim, arXiv:1206.5814 (2012)

[61] Z. Nussinov and G. Ortiz, Phys. Rev. B **79**, 214440 (2009)

[62] E. Cobanera, G. Ortiz, and Z. Nussinov, Advances in Physics **60**, 679 (2011)

[63] W. Brzezicki and A.M. Oleś, European Physical Journal B **66**, 361 (2008)

[64] E. Eriksson and H. Johannesson, Phys. Rev. B **79**, 224424 (2009)

[65] J. Vidal, R. Thomale, K.P. Schmidt, and S. Dusuel, Phys. Rev. B **80**, 081104(R) (2009)

[66] V. Karimipour, Phys. Rev. B **79**, 214435 (2009)

[67] E.H. Lieb, Phys. Rev. Lett. **73**, 2158 (1994)

[68] J. Cui, C. Jun-Peng, and H. Fan, Physical Review A **82**, 022319 (2010)

[69] Y. Yu and Z. Wang, Europhysics Letters **84**, 57002 (2008)

[70] Y. Yu, Nuclear Physics B **799**, 345 (2008)

[71] G. Kells, A.T. Bolukbasi, V. Lahtinen, J.K. Slingerland, J.K. Pachos, and J. Valla, Phys. Rev. Lett. **101**, 240404 (2008)

[72] G. Baskaran, S. Mandal, and R. Shankar, Phys. Rev. Lett. **98**, 247201 (2007)

[73] S. Dusuel, K.P. Schmidt, and J. Vidal, Phys. Rev. Lett. **100**, 177204 (2008)

[74] J.K. Pachos, Int. J. Quantum Inf. **4**, 947 (2006)

[75] J.K. Pachos, Ann. Phys. **322**, 1254 (2007)

[76] H. Yao and S.A. Kivelson, Phys. Rev. Lett. **99**, 247203 (2007)

[77] P. Schnyder, A., S. Ryu, A. Furusaki, and A.W.W. Ludwig, Phys. Rev. B **78**, 195125 (2008)

[78] A. Kitaev, AIP Conference Proceedings **1134**, 22 (2009)

[79] S. Ryu, A.P. Schnyder, A. Furusaki, and A.W.W. Ludwig, New J. Phys. **12**, 065010 (2010)

[80] V. Lahtinen, G. Kells, A. Carollo, T. Stitt, J. Vala, and J.K. Pachos, Ann. Phys. **323**, 2286 (2008)

[81] C. Nayak and F. Wilczek, Nucl. Phys. B **479**, 529 (1996)

12 Kitaev Magnets

Simon Trebst
Institute für Theoretische Physik, Universität zu Köln
Zülpicher Straße 77, 50937 Köln

Contents

1 Introduction **2**

2 Spin liquids **2**
2.1 Frustrated magnets 3
2.2 Classical spin liquids 5
2.3 Quantum spin liquids 7

3 Kitaev honeycomb model **8**
3.1 Fractionalization and spin liquid ground states 9
3.2 Thermal signatures of fractionalization 10

4 Spin-orbit entangled Mott insulators **12**

5 Kitaev materials **15**
5.1 Honeycomb iridates Na_2IrO_3 and Li_2IrO_3 16
5.2 α-$RuCl_3$ 18
5.3 Other materials 19

6 Outlook **20**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

Kitaev materials – spin-orbit entangled Mott insulators with strong bond-directional Ising-like interactions – have attracted considerable interest over the past 15 years as candidate compounds to realize quantum spin liquid physics in experiment. This chapter will guide you on a journey through this field, which many expect to be fertile ground for many future discoveries, both experimentally and theoretically. We will start from building a conceptual perspective on the broader context by first providing a gentle introduction to spin liquids in frustrated magnets, both in the classical and quantum realm. A key player in this field is the Kitaev honeycomb model, to which we will devote a separate section. We will then move the conceptual underpinning to the materials side and introduce the broad class of spin-orbit entangled Mott insulators that are found in $4d$ and $5d$ materials and see how they distinguish themselves from more conventional Mott insulators, which have long been discussed in the context of cuprates and other $3d$ materials. Having set the stage as such, we will then turn to the family of Kitaev materials and discuss some prominent members such as $RuCl_3$ and the iridates Na_2IrO_3 and Li_2IrO_3. The chapter will close with an overview of more recent advances and an outlook what to expect in the near future.

This chapter is based on lecture notes [1], which I have prepared for a 2017 Jülich spring school under the heading "Topological Matter – Topological Insulators, Skyrmions and Majoranas" (48th IFF Spring School). Together with Ciarán Hickey these lecture notes were later turned into a substantially expanded 2022 review article [2] that gives a more in-depth introduction to this field. We should mention that a few other closely related reviews might be good pointers for the interested reader, such as two early reviews on spin-orbit entangled materials [3, 4], along with review-style articles directed towards Kitaev materials [5, 6]. We will mention additional, topical reviews in the subsequent sections that guide to pedagogical introductions or in-depth discussions of the broader context of Kitaev materials.

2 Spin liquids

Let us start our exploration of the conceptual background of (quantum) magnets by reminding ourselves of a paradigm that was first established in the context of the Ising model – spontaneous symmetry breaking. Cast in most general terms, the idea here is that the low-temperature ground state of a system has *less symmetry* than the high-temperature phase which still reflects all symmetries of the underlying Hamiltonian. Case in point of the Ising model is the magnetic ordering of the ground state which breaks the Z_2 symmetry of the original Ising Hamiltonian. This happens at a finite-temperature phase transition, at which it is precisely this Z_2 symmetry, still present in the high-temperature paramagnetic phase, which is spontaneously broken as one traverses the transition towards the low-temperature magnetically ordered phase. This is all very well understood – the finite-temperature transition itself arises from the competition of energy and entropy, while the formation of magnetic order accompanying the spontaneous symmetry breaking can be elegantly captured in terms of Landau-Ginzburg-Wilson theory.

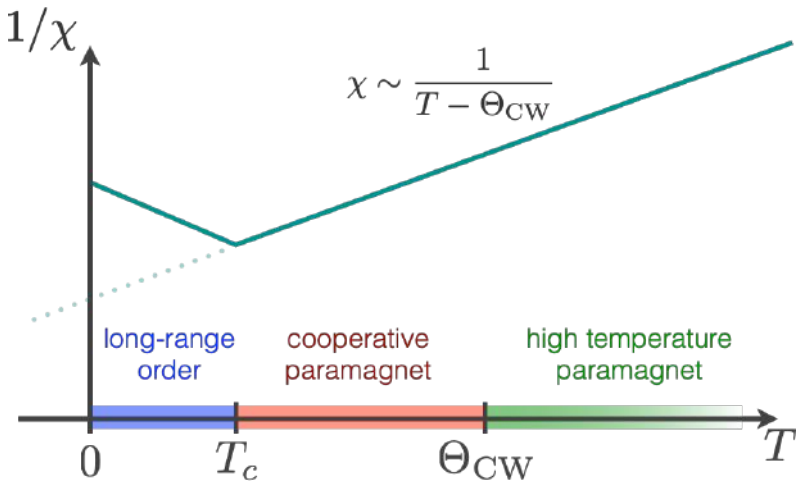


Fig. 1: Frustrated magnetism. *While conventional magnets are expected to show a magnetic ordering transition around the temperature scale associated with the Curie-Weiss scale $\Theta_{\rm CW}$, frustrated magnets instead exhibit an expanded temperature regime (below $\Theta_{\rm CW}$) in which the magnetic susceptibility χ continues to follow a Curie-Weiss law – as if it were still a paramagnet. This regime is often referred to as "cooperative paramagnet". Eventually, the system might order at some very low temperature scale T_c.*

But, quite intriguingly, the exact opposite can also happen – a magnetic system's ground state(s) can have *more symmetry* than the original Hamiltonian and associated high-temperature phase. This is, in fact, what one might define as one of the trademarks of frustrated magnets and the *emergence* of spin liquid physics.

2.1 Frustrated magnets

A frustrated magnet distinguishes itself from a conventional magnet by the absence or strong suppression of the finite-temperature phase transition to a magnetically ordered state. For any conventional magnet, we expect that this phase transition occurs roughly at the Curie-Weiss temperature $\Theta_{\rm CW}$ (set by the various magnetic couplings of a given system). For a frustrated magnet, in contrast, the magnetic susceptibility χ continues to follow a Curie-Weiss law, i.e.

$$\chi \propto \frac{1}{T - \Theta_{\rm CW}} ,$$

even way below the Curie-Weiss temperature. That is, the system keeps behaving as if it were in a paramagnetic phase. But since the system keeps loosing entropy as one goes to lower and lower temperature, there must be a distinction from the high-temperature paramagnet after all. In fact, the system might build up *local* correlations, which however do not reach correlation lengths of the order of the system size and the system therefore eludes the formation of long-range magnetic order. One often refers to this regime as "cooperative paramagnet" – a precursor of the spin liquid physics we might see at the very lowest temperatures (and which we will discuss in the next section) if that physics is not preempted by a magnetic ordering transition. The latter might occur also in a frustrated magnet, albeit at a much lower temperature than in a

conventional magnet as depicted in Fig. 1. In fact, the ratio of the Curie-Weiss temperature and the suppressed transition temperature T_c

$$f = \frac{\Theta_{\rm CW}}{T_c}$$

is a good quantifier of how frustrated a certain system really is and to distinguish a conventional magnet from a "highly-frustrated" magnet. For a conventional magnet $f \approx 1$, while one speaks of a highly-frustrated magnet if $f \gtrsim 10$. In the most extreme case of a system that exhibits no magnetic order transition whatsoever f goes to infinity.

So you might ask what is the microscopic origin of such frustration effects and the ultimate suppression of any magnetic order? The source here are *competing interactions* that cannot be simultaneously satisfied, e.g., by a single state (such as a magnetically ordered one). Instead it is a multitude of states that are all found to be equally well suited to satisfy most of the interactions, that is one finds a manifold of states that all exhibit the same (minimal) energy, though they might differ in their microscopic details. Such an emergence of a low-temperature *residual entropy* really is the defining signature of a frustrated magnet.

One of the most cited and earliest examples here is the triangular lattice Ising antiferromagnet, depicted on the left in Fig. 2. If all couplings are antiferromagnetic, each triangle will exhibit one bond that is left unsatisfied by the choice of the spin alignment around it. Going to a finite lattice with many such triangles this will lead to a ground-state degeneracy that will grow extensively with the number of spins (triangles). This emergence of a residual entropy was first discovered and quantitatively described by Wannier in a seminal 1950 work [7, 8]. Today, we refer to the underlying mechanism as *geometric frustration* to indicate that the source of residual entropy formation really arises from the underlying (non-bipartite) lattice geometry, which is simply non-commensurable with the formation of an antiferromagnetic Néel state.[1] Famous other examples of such non-bipartite lattice geometries are the kagome lattice in two spatial dimensions and the pyrochlore lattice in three spatial dimensions.

Another source of frustration, which will be more relevant in the context of the current chapter, is so-called *exchange frustration*. Consider the arrangement on the right-hand side of Figure 2 where a classical, three component Heisenberg spin is subject to three competing interactions that want to align this spin along one of the three principal spin axes via a pairwise interaction

$$\text{blue bond:} \quad S_i^x S_j^x \qquad \text{green bond:} \quad S_i^y S_j^y \qquad \text{brown bond:} \quad S_i^z S_j^z$$

to match a correspondingly $\{x, y, z\}$-aligned spin on the other side of the bond. Due to the orthogonality of the three principal spin axes, it is impossible to simultaneously satisfy all three exchange terms. Instead, if one picks one of the three principal spin axes (which energetically is more favorable than pointing, e.g., along the [111] direction) one has three equally good (or bad) choices, which again points to the formation of a residual entropy if one continues the

[1] One has to be rather careful, though, in designating geometric frustration to certain spin model. Note, for instance, that the antiferromagnetic *Heisenberg* model on the triangular lattice is *not* subject to geometric frustration. Here the spins will, at low temperatures, simply align in one of two 120 degree ordered states and as such there is no residual entropy.

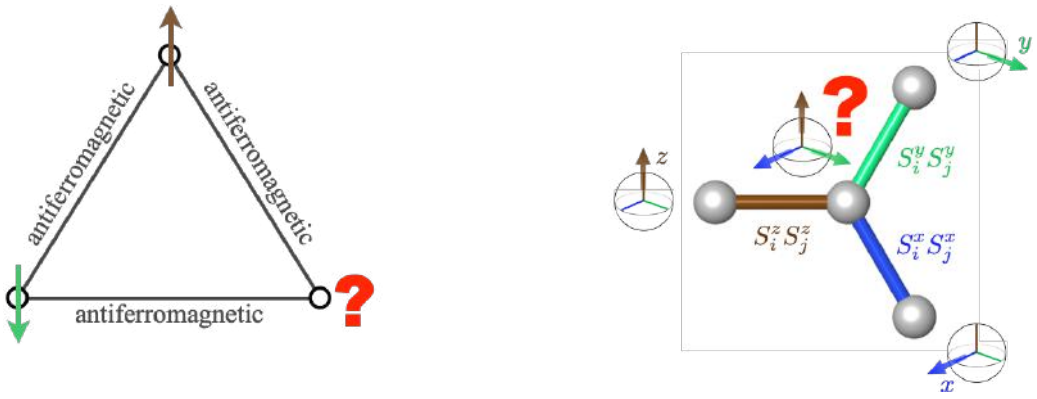


Fig. 2: Sources of Frustration. *(left) Geometric frustration arises from non-bipartite lattice geometries which are incompatible with the formation of a Néel state in antiferromagnetic systems, such as an Ising antiferromagnet on the triangular lattice. (right) Exchange frustration arises from competing interactions that cannot be simultaneously satisfied, even for a single site. An example of this are the three bond-directional interactions favoring alignment of the spins in orthogonal directions.*

tricolored bond assignment for all bonds of, say, a honeycomb lattice. Such exchange frustration is equally capable of suppressing magnetic order as the geometric frustration introduced above. Notably, exchange frustration can also occur for *ferro*magnetic interactions on a bipartite lattice geometry (as in the example above) – that is, in systems, which typically are considered to give rise to simple ferromagnets.

2.2 Classical spin liquids

We now want to consider a frustrated magnet, which indeed shows no magnetic order down to zero temperature. Examples in the realm of classical spins are the triangular lattice Ising antiferromagnet – the principal example of geometric frustration in Figure 2, and the Heisenberg model with bond-directional (ferromagnetic) Ising-like interactions on the honeycomb lattice – the principal example of exchange frustration in Figure 2. The latter example, which in fact is the Kitaev model of the subsequent section, will stay with us throughout the chapter, but for now we might think of it as simply a classical, spin-anisotropic Heisenberg model with no magnetic ordering. For both systems we know that as a result of frustration there will be no single, magnetically ordered ground state but instead there will be a significant residual entropy at zero temperature – in these cases extensive manifolds of states that all equally well satisfy the energetics of the underlying Hamiltonians. But what is the difference between these state manifolds and the high-temperature paramagnet? And can we identify, as alluded to in the introduction of this section, a higher symmetry in these ground states than what the Hamiltonians suggest?
The answers to these questions will introduce us to the concept of *classical spin liquids*. Let us approach such a classical spin liquid by considering the ground state manifold of the bond-directional Heisenberg model on the honeycomb lattice, i.e. the classical Kitaev model with ferromagnetic interactions. The aforementioned exchange frustration in this model results in

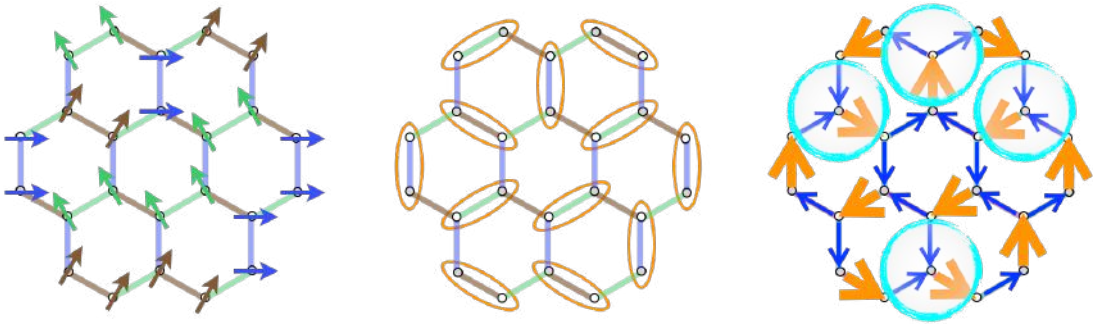


Fig. 3: Coulomb phase. *(left) Ground-state configuration of the classical Kitaev model where the bond-directional Ising-like interactions pair up spins in "dimers" (middle) Dimer covering of the honeycomb lattice, corresponding to the spin configuration on the left. (right) Mapping to a divergence-free magnetic field configuration where each site has an equal number of incoming and outgoing field lines if the orange arrows carry twice the field strengths of the blue arrows.*

a large number of ground states, which can be characterized as illustrated in the left panel of Figure 3: every spin pairs up with one of its three neighboring spins to form a "dimer" of spins pointing along one of the three principal spin axes (such that exactly one of three interaction terms per spin is fully satisfied while two remain completely unsatisfied). In more abstract terms, any such spin configuration can be conceptualized as a dimer covering of the honeycomb lattice[2] as illustrated in the middle panel of Fig. 3 for the exact same spin configuration. This simplifies the description, as a whole lot is known about dimer coverings such as, for instance, how many there are for a given lattice geometry and system size – that is, a direct measure of the residual entropy of our spin model at hand. Even more enticing is the fact that we can rewrite the local condition of "every spin is part of exactly one dimer" into a configuration of an artificial magnetic field that is divergence-free at every site, see the right panel in Fig. 3. This is a powerful correspondence, which tells us that the ground state of the classical Kitaev model (or the triangular lattice Ising antiferromagnet) is described by a *Coulomb phase* [9]. It readily lets us conclude that there are longer range, power-law decaying correlations in these ground states and that the elementary excitations are violations of the divergence-free conditions – magnetic monopoles. In other words, the ground state manifold of our classical magnets are described by *emergent magnetostatics*, a much more elegant description than what we might have anticipated when considering the quite ordinary nature of their underlying Hamiltonians.

It is precisely this theme of an *emergent* description of the low-energy states that sets apart the low-temperature phase of a highly frustrated magnet from both the high-temperature paramagnetic phase, which does not allow for an equally elegant description, or the symmetry-broken ground state of a conventional magnet. This naturally brings us to the question what additional effects zero-temperature quantum fluctuations might entail. On a pessimistic note one might argue that they will simply split the accidental degeneracy of the aforementioned classical ground

[2]Notably, Wannier showed that every ground state configuration of the triangular lattice Ising antiferromagnet can also be mapped to a dimer covering of the honeycomb lattice by marking the unsatisfied bonds of the triangular lattice to its dual honeycomb lattice. As such, all arguments applied to the ground state manifold of the classical Kitaev model also apply to the Ising case.

states in a mechanism referred to as order-by-disorder and thereby destroy all beauty. But this is, fortunately, not true after all and there will be even more to discover when going deep into the quantum realm of strongly fluctuating magnetic moments.

2.3 Quantum spin liquids

Like their classical counterparts, quantum spin liquids are not defined by the absence of magnetic order, but instead by the emergence of additional structures. As one might expect for a proper quantum system, this additional structure comes in the form of entanglement, or more precisely, *long-range entanglement* of the underlying quantum mechanical degrees of freedom. To discuss this, we need to recall some basic notions of quantum many-body entanglement. The latter is often quantified by an entanglement entropy defined via the reduced density matrix for a bipartition of the system into two subsystems (say, A and B). This entanglement entropy is quite distinct from a conventional thermal entropy (familiar from any statistical mechanics course) in that it is not extensive, but instead obeys a boundary law, i.e., it scales with the length of the boundary ∂A separating the two subsystems

$$S = a \cdot \partial A - \gamma + \dots ,$$

where a is some non-universal prefactor in the boundary law, γ refers to an important $O(1)$ correction, and the dots indicate further subleading terms. Our focus here should, in fact, be on the $O(1)$ correction γ that indicates an emergent topological quantum field theory (TQFT) description of the quantum state at hand – that is, an emergent structure that was not present in the Hamiltonian giving rise to the ground state in front of us, somewhat akin to what we had encountered in the classical context but on a whole new level. A TQFT is a complex theory whose constituents are, in general, different types of anyonic particles. An elementary example is the Ising TQFT with its ground state, denoted as 1, and two additional σ and ψ particles which have quantum dimensions[3] $1, \sqrt{2}, 1$, respectively. What is relevant here is that these quantum dimensions define the topological correction of the entanglement entropy [10, 11] in a universal manner as

$$\gamma = \mathrm{n}\left(\sqrt{\sum d_i^2}\right) ,$$

which in the case of the Ising TQFT reveals a correction of $\gamma = \mathrm{n}\, 2$. Importantly, the topological correction to the boundary law always results in a negative correction to the leading boundary-law. This is important as it indicates that we cannot deform the ground-state wavefunction into a simple product state, in which the boundary-law contribution would vanish and thereby turn the entanglement entropy negative – a scenario that is as forbidden as a negative thermal entropy. As such the emergence of such a topological correction instead signals the formation of long-range entanglement that can only be destroyed by driving the system through a quantum phase transition.

[3] The quantum dimension is a measure of how fast a Hilbert space spanned by N such particles grows with the number N. For conventional quantum spin-1/2 we are used to the idea that their quantum dimension is 2, while for the σ particle in the Ising TQFT it is apparently $\sqrt{2}$.

If the ground state of a quantum spin model exhibits such a form of long-range entanglement we have discovered a topological quantum spin liquid – a spin analogue of a fractional quantum Hall state as first envisioned by Kalmeyer and Laughlin [12] back in 1987. We will shortly see such a topological quantum spin liquid as the ground state of the Kitaev model in a magnetic field, and its relevance in the context of half-integer thermal quantum Hall states in the discussion of $RuCl_3$ in the last section.

The above description of quantum spin liquids is a pretty high-brow introduction using the abstract measures of entanglement, which theorists might love as a distinct measure but experimentalists will have a hard time to measure in the foreseeable future. Let us therefore introduce an alternate approach to describe the emergent phenomena of a quantum spin liquid, which might also be closer to experimental reality.

This alternative description of a spin liquid uses the concepts of *fractionalization* and *emergent gauge theories* – concepts that will come to life in the next section when we discuss the exact analytical solution of the quantum Kitaev model. Here we will simply introduce the main ideas and point the interested readers to the excellent review by Lucile Savary and Leon Balents on the subject [13]. In this framework, quantum spin liquids are emergent descriptions of the ground states of quantum magnets in which the elementary degrees of freedom, typically magnetic moments with spin-1/2 or spin-1, decompose into novel, fractionalized quantum particles – a parton (such as a Majorana or complex fermion) coupled to a gauge field. One then distinguishes the different types of possible quantum spin liquids by their different gauge structure into (i) Z_2 quantum spin liquids, (ii) $U(1)$ quantum spin liquids, and (iii) chiral spin liquids (with an emergent Chern-Simons theory). The aforementioned topological quantum spin liquid is an example of the latter, while the Kitaev model (without a magnetic field) is often considered as the quintessential model harboring a Z_2 quantum spin liquid ground state.

3 Kitaev honeycomb model

Let us now turn to the main motivation that has set off the search for Kitaev materials – the original Kitaev honeycomb model [14] and its rich physics [15]. The model itself is a variant of a quantum compass model [16], as discussed in much broader context in the accompanying chapter of Jeroen van den Brink. As such it has a deceivingly simple looking Hamiltonian of bond-directional, Ising-like interactions that couple elementary quantum spin-1/2 degrees of freedom on a honeycomb lattice

$$H_{\text{Kitaev}} = \sum_{i,\gamma} K_\gamma \, S_i^\gamma S_j^\gamma \, ,$$

where $\gamma = x, y, z$ denotes the three principal directions of the honeycomb lattice (depicted in blue/green/brown in Figure 4) and, at the same time, the three principal spin orientations.

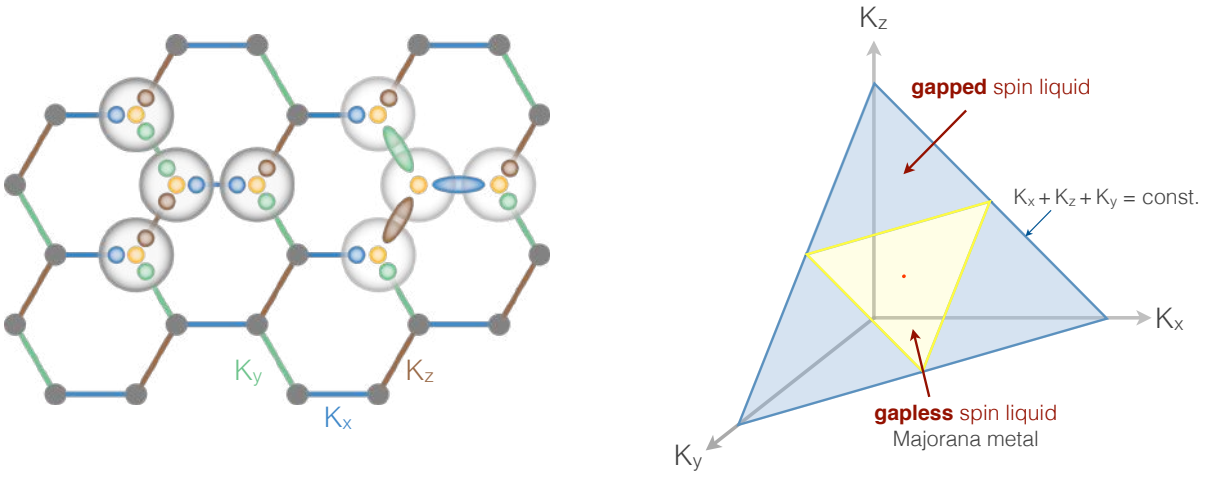


Fig. 4: Kitaev model. *(left) Bond-directional Ising-like interactions between spin-1/2 moments make the Kitaev model an example of a quantum compass model. Its analytical solution is sketched as the decomposition of the original spin degrees of freedom into four Majorana fermions (circles), which are then recombined in a pairwise fashion to result in a Z_2 gauge field on the bonds (ellipses) and a free, itinerant Majorana fermion (yellow circle) hopping on the lattice in the background of a static Z_2 gauge field. (right) The phase diagram of the Kitaev model in the plane $K_x+K_y+K_z = 1$, which exhibits three gapped spin liquid phases and an extended gapless spin liquid phase around the point of isotropic coupling.*

3.1 Fractionalization and spin liquid ground states

What sets this spin model apart from basically every other interacting quantum spin system is that Alexei Kitaev could solve this model exactly at zero temperature, i.e., he could analytically derive its entire ground-state phase diagram [14]. The latter is depicted in the right panel of Figure 4 and shows four highly non-trivial ground states as a function of the coupling parameters K_x, K_y, and K_z plotted in the plane defined by $K_x+K_y+K_z = 1$. If one of the three couplings dominates, corresponding to the light blue triangles, one finds a topological spin liquid ground state, i.e., a gapped spin liquid that is characterized by a non-trivial $\gamma = \mathrm{n}\,2$ correction to the boundary-law entanglement scaling and which corresponds to a toric code [17] phase. The phase around the point of isotropic coupling $K_x = K_y = K_z$ is, in contrast, a gapless spin liquid. The nature of this gapless phase becomes apparent when briefly describing how Kitaev solved the underlying spin model.

The analytical approach is quite ingenious in that it directly employs the fractionalization of the elementary spin degrees of freedom. Every spin-1/2 is rewritten in terms of four Majorana fermions (as depicted schematically by the four circles on the left in Figure 4), which are subsequently recombined by fusing two such Majorana fermions adjacent to a given bond into a single Z_2 variable on every bond (depicted by the ellipses in Figure 4). The latter is the Z_2 gauge field, while the "left-over" fourth Majorana fermion per spin is the complementary parton degree of freedom. Kitaev's solution thus explicitly introduces the emergent fractionalized degrees of freedom via an exact operator decomposition. Importantly, these two degrees

of freedom have very different dynamics. While the Majorana fermion is free to traverse the lattice (as a free fermion), the Z_2 gauge field turns out to be completely static, i.e., it does not fluctuate and any gauge excitations (so-called visons) are static as well and cannot move at all. This makes the emergent lattice gauge theory description of the Kitaev model particularly simple[4] and amenable to an exact solution. The problem factorizes, in that one can first identify the ground state of the gauge field – which due to another ingenious contribution of Elliott Lieb [18] we can readily identify with the flux-free configuration – and then solve for the free Majorana fermion problem with this fixed gauge configuration. But the latter is also trivial, since we know very well what the spectrum of free fermions on a honeycomb lattice is from the study of non-interacting electrons in graphene – a band structure with a Dirac cone dispersion, where due to the particle-hole symmetry of Majorana fermions the ground state of the spin model sits exactly at the tip of this Dirac cone. Returning to the phase diagram of the Kitaev model, this picture of Majorana fermions hopping in the background of a Z_2 lattice gauge structure, lets us readily understand the gapless quantum spin liquid in the center of the phase diagram as a Majorana metal (or, more precisely, a semi-metal with a point-singular Fermi surface due to the Dirac cone in the Majorana band structure).

Going away from the pure Kitaev model, its analytical solution still allows to understand the effect of certain perturbations. For instance, if one applies a magnetic field in the (111)-direction, i.e., a field that couples to all three spin components, one introduces a mass term in the Dirac equation. As a consequence the gapless spin liquid gaps out and turns into a chiral spin liquid (with non-Abelian topological order and gapless edge modes). When rephrased in terms of *complex* fermions this gapped state corresponds to a p-wave superconductor [19], which one can conceptualize to undergo a Higgs transition to a gapless metal that, recast into the language of the quantum spin model, would correspond to a gapless $U(1)$ spin liquid (with a spinon Fermi surface). It has been argued that this indeed what happens for the *anti*ferromagnetic Kitaev model for an intermediate-strength magnetic field [20–23]. Another important perturbation of the pure Kitaev model is the inclusion of an isotropic Heisenberg interaction [24, 25], which endows the vison excitations of the Z_2 gauge field with their own dynamics [26], i.e., they can start to disperse, become soft, condense and thereby drive the system into a magnetically ordered state. Of course, one could consider many other perturbations to the Kitaev model – a topos which we will return to when discussing the miscroscopics of the actual materials considered to realize some of this Kitaev physics.

3.2 Thermal signatures of fractionalization

Staying on the conceptual level, let us instead turn to the finite-temperature characteristics of the Kitaev model. While such features are not amenable to a direct analytical treatment, they can be captured by numerically exact quantum Monte Carlo (QMC) simulations. At first sight, this might sound counter-intuitive as the spin Hamiltonian seems to exhibit a strong sign prob-

[4] Typically one would expect to find a *fluctuating* gauge field, which then would have required further steps such as a mean-field decoupling, i.e., the application of some approximative approach.

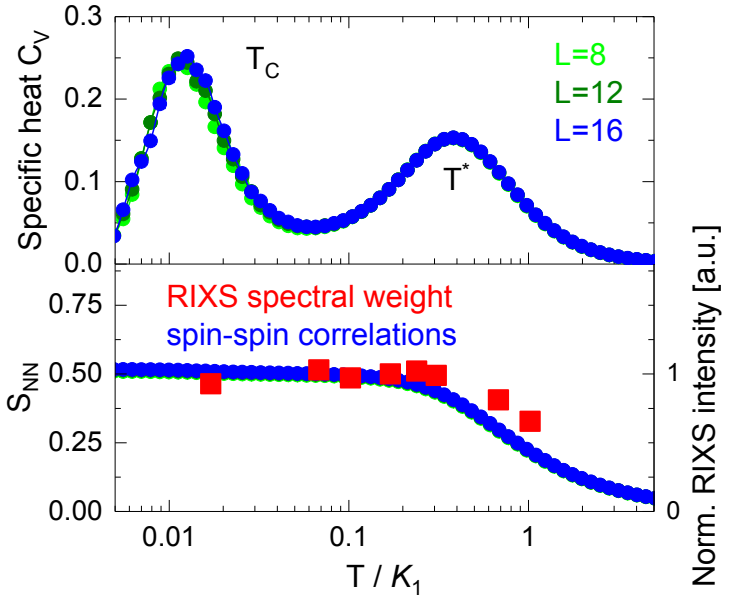


Fig. 5: Thermal signatures of fractionalization. *(top) The specific heat exhibits a characteristic two-peak structure. The feature at the higher temperature scale T^* is a thermal crossover associated with the fractionalization of the elementary spin degrees of freedom, while the lower temperature scale T_c is associated with the onset of order in the Z_2 gauge sector (for finite system sizes). (bottom) The nearest-neighbor spin-spin correlations are found to saturate already at T^*, the higher temperature scale. This is in very good agreement with resonant inelastic X-ray spectroscopy (RIXS) measurements discussed in Section 5. Figure reproduced from Ref. [27].*

lem [28] – at least when looking at it from the perspective of traditional QMC techniques such as the stochastic series expansion (SSE) [29], which is typically the first choice of QMC for quantum spin models. But at second sight and inspired by the exact solution, one should instead set up a QMC approach that samples in the fractionalized basis of Majorana fermions and the Z_2 gauge field. That is, an approach that samples the many different Z_2 gauge field configurations which become relevant at finite temperature as one allows for the thermal excitation of visons, while solving for the respective free Majorana fermion models for such modified Z_2 gauge configurations. This is precisely what Yuki Motome's group has spearheaded to arrive at a quasi-exact solution of the finite-temperature physics of the Kitaev model [30].

The key features of the thermodynamic behavior of the Kitaev model are summarized in Figure 5. The specific heat shows a distinct two-peak structure that is intimately linked to the physics of the Kitaev model. The higher-temperature feature appears at the scale of the exchange couplings and indicates the fractionalization of the elementary spin degrees of freedom and the formation of a Majorana fermion band structure. This fractionalization is a purely *local* phenomenon and as such the higher-peak is in fact a crossover phenomenon (as opposed to a phase transition, which due to its diverging length scales is rather sensitive to finite system sizes and therefore exhibits strong finite-size scaling effects whereas a crossover peak is completely insensitive to system sizes as found here). The second, lower-temperature feature is hugely

suppressed and occurs at a temperature scale of order $T_c \approx K/100$, i.e., two orders of magnitude lower than the high-temperature feature (note the logarithmic scale of the horizontal axis in Figure 5). This second feature is found to be associated with the ordering of the Z_2 gauge field, i.e., it is at this temperature scale that the system enters the flux-free ground state of the Z_2 lattice gauge theory. Above this temperature scale one finds thermally excited vison excitations and as such one might not be surprised to hear that the magnitude of this temperature scale is linked to the size of the vison gap [31]. Now, in two spatial dimensions (as it is the case for a honeycomb lattice geometry) such a vison excitation is a *point-like* excitation, e.g., a Z_2 flux threaded through a single plaquette, which can easily proliferate. This is an important observation which points us to the fact that in the thermodynamic limit of very large system sizes, this lower-temperature ordering transition scales to zero temperature, i.e., ceases to exist. This is a well known statement about two-dimensional Z_2 lattice gauge theories [32, 33], which in the context of the Kitaev model, tells us that the zero-temperature quantum spin liquids of the ground-state phase diagram are all unstable to finite-temperature fluctuations.[5] This puts the higher-temperature crossover feature back into the focus of a potential experimental signature of Kitaev materials. Indeed it should be noted that this crossover goes hand-in-hand with a build-up of nearest-neighbor spin-spin correlations as shown in the lower panel of Figure 5. Such strong correlations between neighboring spins (but not beyond) are indeed another hallmark of the Kitaev spin liquid states [35], which is also reflected in their dynamical structure factor [36, 37] and response in resonant inelastic X-ray spectroscopy (RIXS) [38, 27].

4 Spin-orbit entangled Mott insulators

Let us now enter the realm of materials physics and ask where we might look for microscopic situations that enable the emergence of Kitaev physics. At this stage, probably the most distinct feature of the Kitaev Hamiltonian is its lack of an $SU(2)$ spin symmetry and the bond-directionality of its interactions – both of which features that do not come naturally in conventional electronic Mott insulators, which typically exhibit spin-isotropic Heisenberg interactions and bond-by-bond variations are often limited to differences of the respective coupling strengths. So, we need to look further and this is where models of orbital moments might come to mind – for these it is quite natural to expect a strong bond-directionality and anisotropic interactions. Think, for instance, of a p-orbital model with p_x, p_y, and p_z orbital degrees of freedom where, simply due to the different spatial shapes and alignments of these three orbitals, the interactions in x, y, and z direction will take an Ising-like form, i.e., p_x orbitals couple strongly with one another along the x-direction but not at all along the y or z-directions and similarly for the other orbital components. So orbital physics seems to be the right ingredient to realize Kitaev-like interactions, but then orbital-only models are rare to find.

[5] As a side remark we note that this situation is very different for Kitaev models in three spatial dimensions [34], where the elementary vison excitations are extended flux loops and thereby allow for a different competition between energy and entropy leading to the existence of a true finite-temperature gauge ordering transitions, albeit ones that are still suppressed by two orders of magnitude with regard to the coupling strength [31].

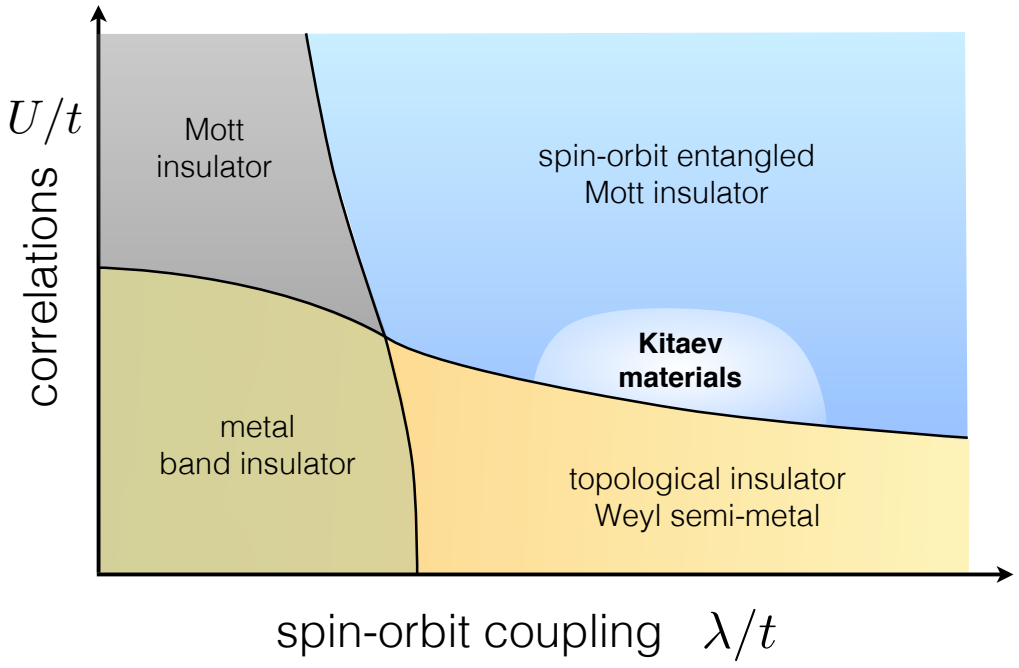


Fig. 6: Spin-orbit assisted Mott insulators. *Shown is a conceptual phase diagram in the presence of correlations (Hubbard U) and spin-orbit coupling λ. The four quadrants of this phase diagram exhibit distinct behavior. For small U and λ we are in the realm of non-interacting electronic band structures, which allow for the formation of metals and band insulators. Cranking up the spin-orbit coupling one can induce a band inversion to create a topological band insulator (or, similarly, a topological semi-metal). For strong correlations we expect to see Mott insulators, which in the presence of strong spin-orbit coupling can turn into a distinct class of Mott insulators with local, spin-orbit entangled moments. The Kitaev materials of interest in this chapter form a subclass of these spin-orbit entangled Mott insulators as discussed in the text. Figure adapted from Reference [3].*

This is where the new class of spin-orbit entangled Mott insulators enter the game. To conceptually understand what these new types of Mott insulators are, it is quite instructive to consider a general phase diagram in the presence of electronic correlations, i.e., a Hubbard U, and spin-orbit coupling λ as it is mapped out in Figure 6. Let us start in the lower left corner of this phase diagram, i.e., in the limit of small correlations and small spin-orbit coupling. This is the realm of non-interacting band theory which tells us that there are two principal states of matter – metals and band insulators. If one now adds substantial spin-orbit coupling, i.e., one moves to the right in the phase diagram, we have learned that, for a band insulator, this can lead to a band inversion and the formation of a topological band insulator [39,40] with protected gapless surface modes. Similarly, we can also create a topological metal via strong spin-orbit coupling, such as a Weyl semi-metal [41]. Now, let us add electronic correlations to the mix. In the regime of small spin-orbit coupling we know that electronic correlations can induce an insulating state of matter that is distinct from a band insulator – a so-called Mott insulator (which, in contrast to band insulators, can occur even at half filling). The local moments in such an electronic Mott insulator are typically $SU(2)$ spin-1/2 (or higher-spin) degrees of freedom – a good place to im-

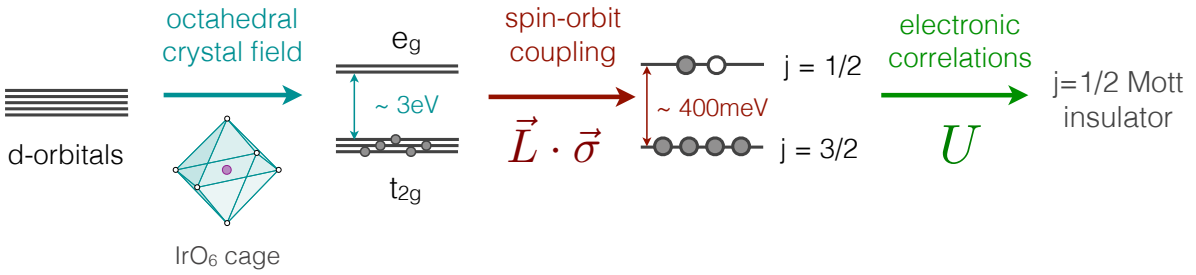


Fig. 7: Spin-orbit entanglement. *Schematic illustration of the formation of spin-orbit entangled $j{=}1/2$ moments in $5d^5$ or $4d^5$ materials with five electrons in the d orbital. Placing such d-orbitals within an octahedral crystal field such as an IrO_6 cage will split the five orbitals into three low-energy t_{2g} orbitals and two high-lying e_g orbitals. The five electrons with a total spin $s{=}1/2$ in the low-lying t_{2g} orbital with an effective orbital moment $l{=}1$ will experience further level splitting upon the introduction of strong spin-orbit coupling. This will result in a low-lying, completely filled $j{=}3/2$ and high-lying, half-filled $j{=}1/2$ state. This $j{=}1/2$ multiplet can now be turned into a Mott insulating states with an effective $j{=}1/2$ moment with relative small electronic correlations (Hubbard U).*

plement a Heisenberg model. If, however, we crank up the spin-orbit coupling in this strongly correlated regime we might end up in a different type of Mott insulator – so-called spin-orbit entangled Mott insulators, in which the local degrees of freedom have both spin and orbital components that are intimately linked to one another.

This scenario plays out in a class of $5d$ and $4d$ materials, which we will now zoom in on. Specifically, we will be interested in $5d^5$ and $4d^5$ materials, that is d-orbitals which are occupied with 5 electrons. This is, for instance, the case for the Iridates, in which the $5d$ iridium ions typically have an electronic Ir^{4+} configuration, and also for $RuCl_3$ with its $4d$ Ru^{3+} ions. As schematically illustrated in Figure 7, a conspiracy of crystal-field splitting, spin-orbit coupling and relatively weak electronic correlations can turn such materials into Mott insulators with local, spin-orbit entangled moments. Historically, this was somewhat unexpected to happen since one might expect $4d$ and particularly $5d$ materials to generically form metallic states, which – compared to the standard class of $3d$ Mott insulators such as the cuprates – exhibit much larger atoms resulting in relatively large electronic overlap in their crystalline structures, which would have to be compensated by strong electronic correlations. But the increased atomic size also gives rise to a much enhanced spin-orbit coupling, which in a crude estimation scales with the fourth power of the atomic number, $\lambda \propto Z^4$. It is through this spin-orbit coupling that the Mott lobes of the $j{=}1/2$ (and $j{=}3/2$) states exhibit a much smaller bandwidth and a relatively small amount of electronic correlations can split them to form Mott insulating states. As such these Mott insulators are also called "spin-orbit assisted" Mott insulators [3]. Their physical reality was first observed, some fifteen years ago, in 2008 in experiments on the perovskite iridate Sr_2IrO_4 [42, 43]. The latter is an isostructural analogue of La_2CuO_4, the parent compound of the cuprate superconductors, which, at the time, set off a flurry of activities searching for (topological) superconductivity in the presence of strong spin-orbit coupling [4].

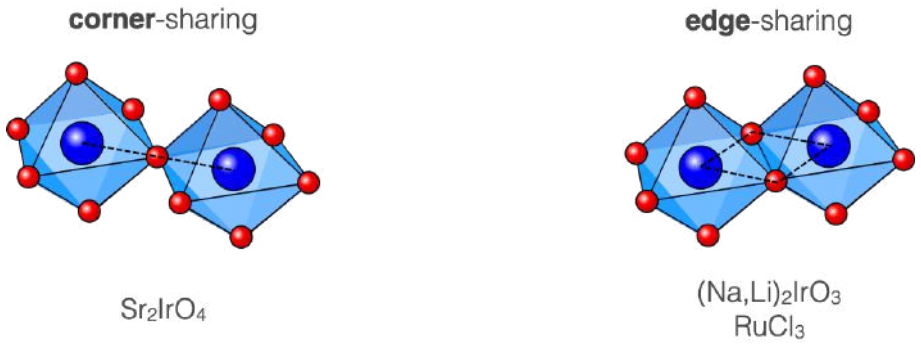


Fig. 8: Exchange path geometries. *The lattice geometries of perovskite and honeycomb iridates (and related materials) distinguish themselves by the way the elementary octahedral oxygen cages are connected – either in a corner-sharing fashion for the perovskite iridates such as $SrIr_2O_4$, while the honeycomb iridates such as Na_2IrO_3 exhibit edge-sharing geometries. The effect on the exchange of the $j{=}1/2$ moments (at the center of the octahedral cages) is profound: while the corner-sharing geometry gives rise to Heisenberg interactions, the edge-sharing geometry exhibits a dominant bond-directional Kitaev-type exchange.*

Prior to these experimental developments, Giniyat Khaliullin had already worked out a $j{=}1/2$ moments theory [44], which would consider the effect of different lattice geometries on the exchange paths of such spin-orbit entangled moments and their resulting effective interactions. In particular, two scenarios have turned out to make a crucial distinction as depicted schematically in Figure 8. Whereas the perovskite iridates exhibit a corner-sharing, square lattice geometry of the octahedral oxygen cages, other iridates with an underlying honeycomb lattice geometry would have edge-sharing oxygen cages. This difference in corner- versus edge-sharing geometries turns out to heavily influence the microscopic exchange of the spin-orbit entangled $j{=}1/2$ moments at the center of the octahedral cages. While in the corner-sharing scenario one finds an isotropic Heisenberg exchange, the edge-sharing scenario induces a suppression of this isotropic Heisenberg exchange (via destructive interference of two Ir-O-Ir exchange paths) turning the next-order bond-directional exchange into the dominant coupling. This turns out to be the sought-after Kitaev-type interaction which we have been looking for.

5 Kitaev materials

In 2009 Jackeli and Khaliullin turned this thinking about spin-orbit entangled $j{=}1/2$ Mott insulators with edge-sharing geometries into a concrete proposal – they went out and postulated that honeycomb iridates such as Na_2IrO_3 and Li_2IrO_3 should be an ideal place to look for Kitaev physics [45]. The boldness of this proposal should be appreciated – while we might have become accustomed to the idea that we can theoretically predict materials properties of weakly-coupled materials and then await experimental verification in a newly synthesized compound (this has been a recurring motif in the synthesis of topological insulators), such a conceptual prediction for strongly-correlated Mott materials has been without much precedent. Nevertheless, this proposal turned out to be extremely influential and opened the field of Kitaev spin liquid physics to experimental exploration of actual materials which rapidly happened.

5.1 Honeycomb iridates Na_2IrO_3 and Li_2IrO_3

The first samples of Na_2IrO_3 were synthesized basically within a year of the theoretical proposal by the groups of Takagi [46] and Gegenwart [47]. Many other groups would follow and today single crystals of Na_2IrO_3 are readily available in labs around the world. But there was an element of initial disillusion too – Na_2IrO_3 exhibits an ordering transition [48] around $T_N \approx 15$ K, i.e., it is clearly not showing a quantum disordered spin liquid ground state. Resonant X-ray magnetic scattering [49] and neutron scattering experiments [50, 51] would later reveal that the local moments form a zig-zag order. But on the positive side, the local moments themselves would turn out to be indeed the sought-after spin-orbit entangled $j{=}1/2$ moments, as seen from magnetic susceptibility measurements [47, 48]. Another uplift came when a direct experimental observation of bond-directional exchange was reported in diffuse magnetic X-ray scattering experiments [52], which later would be substantiated as Kitaev-type interactions in resonant inelastic X-ray spectroscopy (RIXS) [53]. It also posed new puzzles such as the observation of short-range spin-spin correlations in RIXS experiments [54, 27] on temperature scales far above the ordering temperature, see Figure 5. Such a high-temperature formation of spin-spin correlations would be a signature of a low-temperature Kitaev spin liquid as argued in the earlier section on thermal signatures of fractionalization. So maybe Na_2IrO_3 is, after all, not that far away from spin liquid physics?

Such puzzles coming out of the experimental exploration of the first Kitaev candidate material, Na_2IrO_3, have spurred more theoretical activity. Starting with the addition of a Heisenberg interaction to the pure Kitaev model [24, 25, 55–58], the effect of more and more perturbations of the Kitaev model have been explored leading to a refined microscopic model [59]

$$H = \sum_{\gamma-\text{bonds}} J\ \mathbf{S}_i\mathbf{S}_j + K\ S_i^\gamma S_j^\gamma + \Gamma\left(S_i^\alpha S_j^\beta + S_i^\beta S_j^\alpha\right),$$

which also includes a bond-directional, off-diagonal Γ-exchange. Complementing this more and more detailed microscopic understanding, the general concept of a "proximate spin liquid" [60, 61] was developed whose central idea is sketched in Figure 9. While additional "non-Kitaev" interactions (as in the Hamiltonian above) might induce magnetic ordering (as observed, for instance, in Na_2IrO_3), there might be a window of opportunity (indicated by the red arrow in the figure) to observe some of the thermal signatures of a nearby/proximate spin liquid ground state. Physically this finite-temperature proximate spin liquid regime opens up, as the free energy of a thermal spin liquid state (with all its fluctuations) is generically expected to be lower than the free energy of a thermally excited magnetically ordered state (which shows little fluctuations). It might be in this regime that we indeed see remnants of spin liquid physics above magnetically ordered states, such as signatures of fractionalization and the unexpected build-up of local spin-spin correlations [27].

The synthesis of the sister compound Li_2IrO_3 has led to yet another surprising discovery – this material exists in several polymorphs [48,62,63] which have been dubbed α-Li_2IrO_3, β-Li_2IrO_3, and γ-Li_2IrO_3 in the sequence of their discovery. The first one, α-Li_2IrO_3, is a honeycomb material akin to Na_2IrO_3, which shares most of its experimental signatures: local $j{=}1/2$ moments

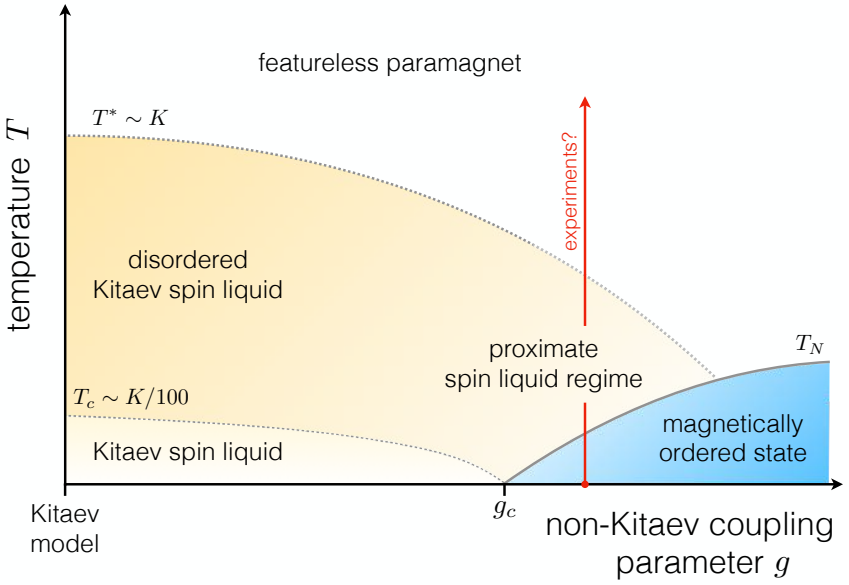


Fig. 9: Proximate spin liquid. *Conceptual phase diagram of the Kitaev model perturbed by some additional interactions parametrized by a coupling strength g, such as a Heisenberg and/or off-diagonal Γ-exchange, that will induce some type of magnetic ordering if sufficiently strong, $g > g_c$. At finite temperatures we would expect the spin liquid state, stabilized as a ground state for $g < g_c$, to dominate over the magnetically ordered state, since its fluctuations will lower its effective free energy more strongly than those of the ordered state. As such the magnetic ordering transition will bend to the right as indicated. This, however, opens a temperature window, indicated by the red arrow, in which one might see the thermal signatures of the proximate spin liquid such as the thermal fractionalization crossover and accompanying build-up of local spin-spin correlations shown in Figure 5.*

form, but undergo a magnetic ordering transition at $T_N \approx 15$ K to what is a somewhat unusual magnetic ordering pattern with counter-rotating spin spirals [64–66]. The other two polymorphs are materials in which the local $j{=}1/2$ moments are arranged in *three-dimensional* lattice geometries. These lattice structures, however, retain an important ingredient of the honeycomb lattice – they are still *tri-coordinated*, i.e., every site is connected to only three other sites (which is a highly unusual setting in a three-dimensional geometry), just as in the honeycomb case and the two polymorphs, β-Li_2IrO_3, and γ-Li_2IrO_3 have been dubbed hyperhoneycomb and stripy honeycomb to convey this point. Experimentally, these two 3D polymorphs turn out to be relatively close to their 2D counterpart in that they exhibit an ordering transition to a counter-rotating spin spiral state.

But again experimental discovery has spurred theoretical advances, here in the form of an investigation of three-dimensional generalizations of the Kitaev model [67,68]. The key observation is that the tri-coordination of the 3D lattice geometries still allows for the same analytical approach devised by Kitaev for the 2D honeycomb geometry (via a four Majorana fermion decomposition and subsequent bond-wise recombination, as illustrated in Figure 4). This has led to an extensive classification of three-dimensional Kitaev models, both in terms of their spin liquid ground states which can be recast as different types of Majorana metals [34] as well as their thermodynamic signatures [31].

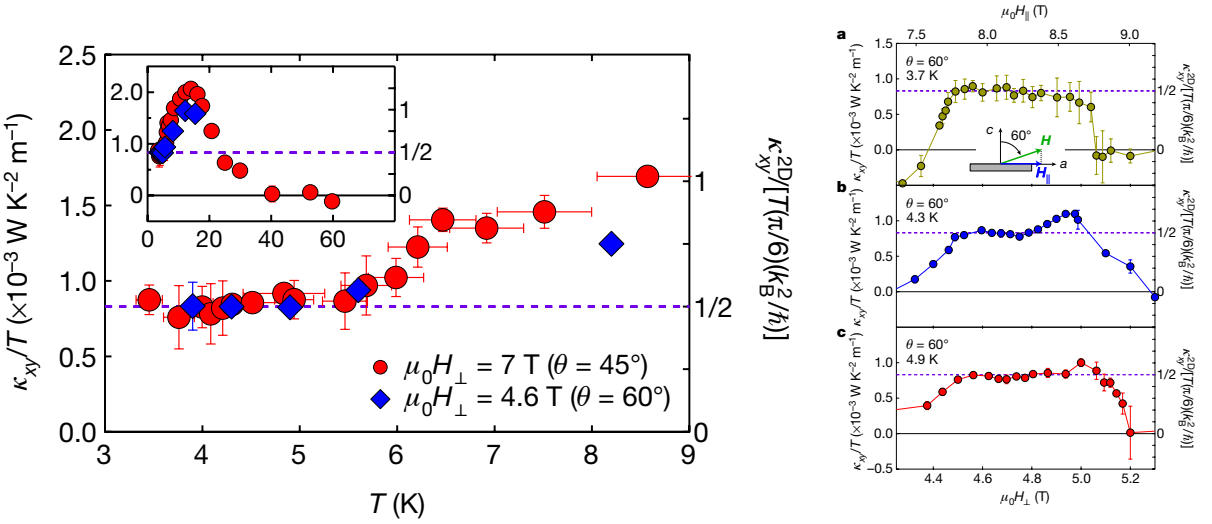


Fig. 10: Quantized Thermal Hall Effect. *Thermal Hall conductivity of α-RuCl$_3$ in a tilted magnetic field. (left) The thermal Hall conductivity κ_{xy}/T appears to saturate at a 1/2-quantized value within a temperature range of 4–6 K (for two tilted magnetic field configurations). The inset shows the higher-temperature behavior of κ_{xy}/T, overshooting the half-quantized value with increasing temperature and reaching a maximum at $\sim 14K$ before decreasing again for even higher temperatures. (right) The thermal Hall conductivity as a function of increasing magnetic field, at a fixed angle of 60° away from the c-axis for the three different temperatures. The plateau appears to be stable over an extended field and temperature range. Figure adapted from Ref. [75].*

5.2 α-RuCl$_3$

The biggest impact on the field of Kitaev materials has come in the form of the $4d$ compound α-RuCl$_3$, which had long existed but has been appreciated as another candidate material for honeycomb Kitaev physics only in 2014 [69]. By then a well-oiled sequence of initial experiments set in which quickly established the $j{=}1/2$ nature of the local moments [70, 71] and again the onset of zig-zag order [61], though at a slightly lower temperature of $T_N \approx 7$ K. But what sets α-RuCl$_3$ apart from the two honeycomb iridates discussed above has been a second round of spectacular experiments that have solidified its status as a front runner to indeed exhibit telltale signatures of Kitaev spin liquid physics. The first came in the interpretation of Raman scattering data as having *fermionic* excitations across a broad energy and temperature range [72–74], indicating the absence of a conventional magnetic state which only has bosonic excitations (magnons and phonons), but allowing for a more speculative state such as the Kitaev spin liquid which would indeed exhibit (Majorana) fermion excitations. The second in the form of the observation of a diffuse scattering continuum [60,61] in the inelastic neutron scattering of α-RuCl$_3$, reminiscent of what one would expect for fractional excitations such as spinons or, in the context of Kitaev models – spinless Majorana fermions. This observation at finite excitation energies above a magnetically ordered ground state is what triggered the conceptual idea of a proximate spin liquid, introduced in the previous section.

The most spectacular experimental result, reported by the Matsuda group in Kyoto [75], has come in the form of a field-induced state that appears to exhibit a half-quantized thermal Hall effect over a temperature range of 4–6 K, different orientations of the tilted magnetic field, and an extended plateau forming as a function of the magnetic field strength, as reproduced in Figure 10. It is precisely such a 1/2-quantized thermal Hall effect that one would theoretically expect for the field-induced, chiral spin liquid mentioned in the theory introduction. While gapped in the bulk, this chiral spin liquid would have gapless Majorana edge modes that would carry precisely a 1/2 (thermal) charge quantum. While similar to electronic fractional quantum Hall states (such as the one introduced [76] for filling fraction $\nu{=}5/2$, which gained experimental support through careful thermal Hall measurements [77]), an important difference is that the electronic Hall state arises due to the formation of Landau levels (and Coulomb interactions). In contrast, the chiral spin liquid state at hand is a non-trivial Chern insulator of Majorana fermions. One important distinction between the two scenarios is that the latter can also form in a *planar* field configuration, while the formation of Landau levels always requires an out-of-plane field component. As such, it has been quite reassuring to see that the Matsuda group could indeed reproduce their observation of a thermal Hall plateau for such an in-plane field configuration only [78]. There is grain of salt, however, in that these results have only been partially reproduced by the Takagi group [79], while others have openly questioned the quality of the quantization and have instead argued that the observation of quantum oscillations in fact points to the formation of a field-induced gapless spin liquid [80].

We are left to state that the observation of a field-induced Kitaev spin liquid remains a much-discussed topic in the community of Kitaev material aficionados that will come to a final conclusion only by further experimental evidence or all-encompassing theoretical models. Or to put it with a more positive spin – there is still much to be done here.

5.3 Other materials

Following the initial proposal [45] of the honeycomb iridates Na_2IrO_3 and Li_2IrO_3, a plethora of alternative compositions of honeycomb iridates have been put forward and synthesized, including $H_3LiIr_2O_6$ [81], $Ag_3LiIr_2O_6$ [82] and $Cu_3LiIr_2O_6$ [83], in which the interlayer alkali Li ions of Li_2IrO_3 have been replaced by H, Ag or Cu, respectively. In a similar vain, but starting from Na_2IrO_3, we have seen the synthesis of $Cu_3NaIr_2O_6$ [83] and Cu_2IrO_3 [84]. For readers interested in learning more about this "second generation" of honeycomb Kitaev materials we point to our recent review of Kitaev materials [2].

Going beyond the Jackeli-Khaliullin mechanism, much recent attention has gone into the question whether $3d^7$ materials such as cobaltates can also exhibit Kitaev physics [85,86], despite the reservation one might have that these systems will exhibit considerably smaller spin-orbit coupling than the $4d$ and $5d$ compounds discussed so far. Initial experimental efforts have focused on $Na_2Co_2TeO_6$ [87, 88] and $Na_3Co_2SbO_6$ [87] as potentially interesting materials, which like their d^5 counterparts both exhibit zigzag magnetic order at low temperatures [89–91]. Inelastic neutron scattering measurements have been argued to show evidence for dominant Kitaev

exchange interactions [92, 93], though this interpretation still remains under debate. Application of a magnetic field to $Na_2Co_2TeO_6$ points to a field-induced disordered state that could potentially harbor spin liquid physics [94, 95]. For another set of cobaltates, $BaCo_2(AsO_4)_2$ and $BaCo_2(PO_4)_2$, which have also been considered to be potential Kitaev materials, it by now seems increasingly likely that their original theoretical description, in terms of an XXZ J_1-J_2-J_3 model, better captures their essential physics [96–99].

In further broadening the search for Kitaev materials, it may be worthwhile to look beyond $4d$ and $5d$ transition metals and consider rare-earth magnets [100] whose $4f$ electrons are much more localized than the $5d$ or $4d$ electrons in iridates and ruthenates and at the same time experience a considerably stronger spin-orbit coupling – thus potentially providing another path to Kitaev materials in the future. New materials that realize antiferromagnetic Kitaev interactions, in contrast to the ferromagnetic interactions believed to occur in the current set of known Kitaev materials, will be particularly welcome both as a means to explore a new area of the extended phase diagrams and for their distinct in-field properties. At the level of materials synthesis, the search for Kitaev materials continues unabated.

If we adopt an even broader definition of what a Kitaev material (or Kitaev magnet) is by requiring that a given material is (i) a spin-orbit entangled Mott insulators with local $j{=}1/2$ moments, which (ii) interact via bond-directional Kitaev-type interactions, a much broader class of materials comes into view. In particular, these might realize lattice geometries well beyond the honeycomb structure of the original Kitaev model. This includes, for instance, the double perovskite Ba_2CeIrO_6, which turns out to be a pristine $j{=}1/2$ Mott insulator where the moments are subject to frustrated magnetism on the face-centered cubic lattice with Kitaev interactions [101].

6 Outlook

The field of Kitaev materials has come a long way since its inception by the introduction of the Kitaev honeycomb model [14] in 2006 and the bold proposal in 2009 to look for its physics in transition-metal oxides [45]. Along the way, we have seen one compound after the other being synthesized in unprecedented speed – broadening the search well beyond the initial trio of Na_2IrO_3, Li_2IrO_3, and α-$RuCl_3$ to other iridates, ruthenates, and even cobaltates. Experimental exploration has kept the pace and showered us with results, some of which reporting highly unusual observations that point to quantum spin liquid physics, such as Raman scattering data pointing to fermionic excitations, inelastic neutron scattering data showing a broad diffusive spectrum indicating fractionalized excitations, and a half-integer quantized thermal Hall effect. Further experimental efforts will need to go in the validation and verification of these results on different samples and by complementary approaches. On the theory side, the experimental discoveries have spurred lots of activities in refining the microscopic description of Kitaev materials beyond the pure Kitaev model, devising new concepts such as proximate spin liquid physics, and classifying Kitaev physics also in three-dimensional lattice geometries.

Looking into the future is impossible in a field with so much activity. But one might wish for some developments that seem promising today. On a materials synthesis side, this will include ideas to engineer novel Kitaev materials in hybrid devices such as heterostructures of α-$RuCl_3$ and graphene [102–104]. On the experimental side, novel probes such as two-dimensional coherent spectroscopy (2DCS) [105] to pick up the non-linear response of, for instance, fractional excitations might turn out to be rather insightful. Such novel spectroscopic methods should, of course, be complemented by theoretical activities predicting such non-linear response for the various Kitaev magnets of interest [106].

References

[1] S. Trebst, arXiv:1701.07056

[2] S. Trebst and C. Hickey, Phys. Rep. **950**, 1 (2022)

[3] W. Witczak-Krempa, G. Chen, Y.B. Kim, and L. Balents, Annu. Rev. Condens. Matter Phys. **5**, 57 (2014)

[4] J.G. Rau, E.K.-H. Lee, and H.-Y. Kee, Annu. Rev. Condens. Matter Phys. **7**, 195 (2016)

[5] S.M. Winter, A.A. Tsirlin, M. Daghofer, J. van den Brink, Y. Singh, P. Gegenwart, and R. Valentí, J. Phys.: Condens. Matter **29**, 493002 (2017)

[6] Y. Motome and J. Nasu, J. Phys. Soc. Jpn **89**, 012002 (2020)

[7] G.H. Wannier, Phys. Rev. **79**, 357 (1950)

[8] G.H. Wannier, Phys. Rev. B **7**, 5017 (1973)

[9] C.L. Henley, Annu. Rev. Condens. Matter Phys. **1**, 179 (2010)

[10] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006)

[11] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006)

[12] V. Kalmeyer and R.B. Laughlin, Phys. Rev. Lett. **59**, 2095 (1987)

[13] L. Savary and L. Balents, Rep. Prog. Phys. **80**, 016502 (2017)

[14] A. Kitaev, Ann. Phys. **321**, 2 (2006)

[15] M. Hermanns, I. Kimchi, and J. Knolle, Annu. Rev. Condens. Matter Phys. **9**, 17 (2018)

[16] Z. Nussinov and J. van den Brink, Rev. Mod. Phys. **87**, 1 (2015)

[17] A.Y. Kitaev, Ann. Phys. **303**, 2 (2003)

[18] E.H. Lieb, Phys. Rev. Lett. **73**, 2158 (1994)

[19] F.J. Burnell and C. Nayak, Phys. Rev. B **84**, 125125 (2011)

[20] C. Hickey and S. Trebst, Nat. Commun. **10**, 530 (2019)

[21] N.D. Patel and N. Trivedi, PNAS **116**, 12199 (2019)

[22] M. Gohlke, R. Moessner, and F. Pollmann, Phys. Rev. B **98**, 014418 (2018)

[23] Z. Zhu, I. Kimchi, D.N. Sheng, and L. Fu, Phys. Rev. B **97**, 241110 (2018)

[24] J. Chaloupka, G. Jackeli, and G. Khaliullin, Phys. Rev. Lett. **105**, 027204 (2010)

[25] H.-C. Jiang, Z.-C. Gu, X.-L. Qi, and S. Trebst, Phys. Rev. B **83**, 245104 (2011)

[26] X.-Y. Song, Y.-Z. You, and L. Balents, Phys. Rev. Lett. **117**, 037209 (2016)

[27] A. Revelli, M. Moretti Sala, G. Monaco, C. Hickey, P. Becker, F. Freund, A. Jesche, P. Gegenwart, T. Eschmann, F.L. Buessen, S. Trebst, P.H.M. van Loosdrecht, J. van den Brink, and M. Grüninger, Phys. Rev. Research **2**, 043094 (2020)

[28] E.Y. Loh, J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R.L. Sugar, Phys. Rev. B **41**, 9301 (1990)

[29] A.W. Sandvik, Phys. Rev. B **59**, R14157 (1999)

[30] J. Nasu, M. Udagawa, and Y. Motome, Phys. Rev. Lett. **113**, 197205 (2014)

[31] T. Eschmann, P.A. Mishchenko, K. O'Brien, T.A. Bojesen, Y. Kato, M. Hermanns, Y. Motome, and S. Trebst, Phys. Rev. B **102**, 075125 (2020)

[32] N. Read and S. Sachdev, Phys. Rev. Lett. **66**, 1773 (1991)

[33] T. Senthil and M.P.A. Fisher, Phys. Rev. B **62**, 7850 (2000)

[34] K. O'Brien, M. Hermanns, and S. Trebst, Phys. Rev. B **93**, 085101 (2016)

[35] G. Baskaran, S. Mandal, and R. Shankar, Phys. Rev. Lett. **98**, 247201 (2007)

[36] J. Knolle, D.L. Kovrizhin, J.T. Chalker, and R. Moessner, Phys. Rev. Lett. **112**, 207203 (2014)

[37] J. Knolle, D.L. Kovrizhin, J.T. Chalker, and R. Moessner, Phys. Rev. B **92**, 115127 (2015)

[38] G.B. Halász, N.B. Perkins, and J. van den Brink, Phys. Rev. Lett. **117**, 127203 (2016)

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

[40] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011)

[41] B. Yan and C. Felser, Annu. Rev. Condens. Matter Phys. **8**, 11.1 (2017)

[42] B.J. Kim, H. Jin, S.J. Moon, J.-Y. Kim, B.-G. Park, C.S. Leem, J. Yu, T.W. Noh, C. Kim, S.-J. Oh, J.-H. Park, V. Durairaj, G. Cao, and E. Rotenberg, Phys. Rev. Lett. **101**, 076402 (2008)

[43] B.J. Kim, H. Ohsumi, T. Komesu, S. Sakai, T. Morita, H. Takagi, and T. Arima, Science **323**, 1329 (2009)

[44] G. Khaliullin, Prog. Theor. Phys. Suppl. **160**, 155 (2005)

[45] G. Jackeli and G. Khaliullin, Phys. Rev. Lett. **102**, 017205 (2009)

[46] H. Takagi, *private communication* (2009)

[47] Y. Singh and P. Gegenwart, Phys. Rev. B **82**, 064412 (2010)

[48] Y. Singh, S. Manni, J. Reuther, T. Berlijn, R. Thomale, W. Ku, S. Trebst, and P. Gegenwart, Phys. Rev. Lett. **108**, 127203 (2012)

[49] X. Liu, T. Berlijn, W.-G. Yin, W. Ku, A. Tsvelik, Y.-J. Kim, H. Gretarsson, Y. Singh, P. Gegenwart, and J.P. Hill, Phys. Rev. B **83**, 220403 (2011)

[50] F. Ye, S. Chi, H. Cao, B.C. Chakoumakos, J.A. Fernandez-Baca, R. Custelcean, T.F. Qi, O.B. Korneta, and G. Cao, Phys. Rev. B **85**, 180403 (2012)

[51] S.K. Choi, R. Coldea, A.N. Kolmogorov, T. Lancaster, I.I. Mazin, S.J. Blundell, P.G. Radaelli, Y. Singh, P. Gegenwart, K.R. Choi, S.-W. Cheong, P.J. Baker, C. Stock, and J. Taylor, Phys. Rev. Lett. **108**, 127204 (2012)

[52] S. Hwan 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.M. Sala, M. Krisch, J. Chaloupka, G. Jackeli, G. Khaliullin, and B.J. Kim, Nat. Phys. **11**, 462 (2015)

[53] M. Magnaterra, K. Hopfer, C.J. Sahle, M.M. Sala, G. Monaco, J. Attig, C. Hickey, I.M. Pietsch, F. Breitner, P. Gegenwart, M.H. Upton, J. Kim, S. Trebst, P.H.M. van Loosdrecht, J. van den Brink, and M. Grüninger, arXiv:2301.08340

[54] J. Kim, J. c. v. Chaloupka, Y. Singh, J.W. Kim, B.J. Kim, D. Casa, A. Said, X. Huang, and T. Gog, Phys. Rev. X **10**, 021034 (2020)

[55] I. Kimchi and Y.-Z. You, Phys. Rev. B **84**, 180407 (2011)

[56] R. Schaffer, S. Bhattacharjee, and Y.B. Kim, Phys. Rev. B **86**, 224417 (2012)

[57] E. Sela, H.-C. Jiang, M.H. Gerlach, and S. Trebst, Phys. Rev. B **90**, 035113 (2014)

[58] I. Rousochatzakis, J. Reuther, R. Thomale, S. Rachel, and N.B. Perkins, Phys. Rev. X **5**, 041035 (2015)

[59] J.G. Rau, E.K.-H. Lee, and H.-Y. Kee, Phys. Rev. Lett. **112**, 077204 (2014)

[60] A. Banerjee, C.A. Bridges, J.Q. Yan, A.A. Aczel, L. Li, M.B. Stone, G.E. Granroth, M.D. Lumsden, Y. Yiu, J. Knolle, S. Bhattacharjee, D.L. Kovrizhin, R. Moessner, D.A. Tennant, D.G. Mandrus, and S.E. Nagler, Nat. Mater. **15**, 733 (2016)

[61] A. Banerjee, J. Yan, J. Knolle, C.A. Bridges, M.B. Stone, M.D. Lumsden, D.G. Mandrus, D.A. Tennant, R. Moessner, and S.E. Nagler, Science **356**, 1055 (2017)

[62] T. Takayama, A. Kato, R. Dinnebier, J. Nuss, H. Kono, L.S.I. Veiga, G. Fabbris, D. Haskel, and H. Takagi, Phys. Rev. Lett. **114**, 077202 (2015)

[63] K.A. Modic, T.E. Smidt, I. Kimchi, N.P. Breznay, A. Biffin, S. Choi, R.D. Johnson, R. Coldea, P. Watkins-Curry, G.T. McCandless, J.Y. Chan, F. Gandara, Z. Islam, A. Vishwanath, A. Shekhter, R.D. McDonald, and J.G. Analytis, Nat. Commun. **5**, 4203 (2014)

[64] I. Kimchi, R. Coldea, and A. Vishwanath, Phys. Rev. B **91**, 245134 (2015)

[65] E.K.-H. Lee and Y.B. Kim, Phys. Rev. B **91**, 064407 (2015)

[66] E.K.-H. Lee, J.G. Rau, and Y.B. Kim, Phys. Rev. B **93**, 184420 (2016)

[67] M. Hermanns and S. Trebst, Phys. Rev. B **89**, 235102 (2014)

[68] M. Hermanns, K. O'Brien, and S. Trebst, Phys. Rev. Lett. **114**, 157202 (2015)

[69] K.W. Plumb, J.P. Clancy, L.J. Sandilands, V.V. Shankar, Y.F. Hu, K.S. Burch, H.-Y. Kee, and Y.-J. Kim, Phys. Rev. B **90**, 041112 (2014)

[70] Y. Kobayashi, T. Okada, K. Asai, M. Katada, H. Sano, and F. Ambe, Inorg. Chem. **31**, 4570 (1992)

[71] J.M. Fletcher, W.E. Gardner, A.C. Fox, and G. Topping, J. Chem. Soc. A, 1038 (1967)

[72] L.J. Sandilands, Y. Tian, K.W. Plumb, Y.-J. Kim, and K.S. Burch, Phys. Rev. Lett. **114**, 147201 (2015)

[73] L.J. Sandilands, Y. Tian, A.A. Reijnders, H.-S. Kim, K.W. Plumb, Y.-J. Kim, H.-Y. Kee, and K.S. Burch, Phys. Rev. B **93**, 075144 (2016)

[74] Y. Wang, G.B. Osterhoudt, Y. Tian, P. Lampen-Kelley, A. Banerjee, T. Goldstein, J. Yan, J. Knolle, H. Ji, R.J. Cava, J. Nasu, Y. Motome, S.E. Nagler, D. Mandrus, and K.S. Burch, npj Quantum Materials **5**, 14 (2020)

[75] Y. Kasahara, T. Ohnishi, Y. Mizukami, O. Tanaka, Sixiao Ma, K. Sugii, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, T. Shibauchi, and Y. Matsuda, Nature **559**, 227 (2018)

[76] G. Moore and N. Read, Nucl. Phys. B **360**, 362 (1991)

[77] M. Banerjee, M. Heiblum, V. Umansky, D.E. Feldman, Y. Oreg, and A. Stern, Nature **559**, 205 (2018)

[78] T. Yokoi, S. Ma, Y. Kasahara, S. Kasahara, T. Shibauchi, N. Kurita, H. Tanaka, J. Nasu, Y. Motome, C. Hickey, S. Trebst, and Y. Matsuda, Science **373**, 568 (2021)

[79] J.A.N. Bruin, R.R. Claus, Y. Matsumoto, N. Kurita, H. Tanaka, and H. Takagi, Nat. Phys. **18**, 401 (2022)

[80] P. Czajka, T. Gao, M. Hirschberger, P. Lampen-Kelley, A. Banerjee, J. Yan, D.G. Mandrus, S.E. Nagler, and N.P. Ong, Nat. Phys. **17**, 915 (2021)

[81] K. Kitagawa, T. Takayama, Y. Matsumoto, A. Kato, R. Takano, Y. Kishimoto, S. Bette, R. Dinnebier, G. Jackeli, and H. Takagi, Nature **554**, 341 (2018)

[82] V. Todorova, A. Leineweber, L. Kienle, V. Duppel, and M. Jansen, J. Solid State Chem. **184**, 1112 (2011)

[83] J.H. Roudebush, K.A. Ross, and R.J. Cava, Dalton Trans. **45**, 8783 (2016)

[84] M. Abramchuk, C. Ozsoy-Keskinbora, J.W. Krizan, K.R. Metz, D.C. Bell, and F. Tafti, J. Am. Chem. Soc. **139**, 15371 (2017)

[85] H. Liu and G. Khaliullin, Phys. Rev. B **97**, 014407 (2018)

[86] R. Sano, Y. Kato, and Y. Motome, Phys. Rev. B **97**, 014408 (2018)

[87] L. Viciu, Q. Huang, E. Morosan, H. Zandbergen, N. Greenbaum, T. McQueen, and R. Cava, J. Solid State Chem. **180**, 1060 (2007)

[88] R. Berthelot, W. Schmidt, A. Sleight, and M. Subramanian, J. Solid State Chem. **196**, 225 (2012)

[89] E. Lefrançois, M. Songvilay, J. Robert, G. Nataf, E. Jordan, L. Chaix, C.V. Colin, P. Lejay, A. Hadj-Azzem, R. Ballou, and V. Simonet, Phys. Rev. B **94**, 214416 (2016)

[90] A.K. Bera, S.M. Yusuf, A. Kumar, and C. Ritter, Phys. Rev. B **95**, 094424 (2017)

[91] C. Wong, M. Avdeev, and C.D. Ling, J. Solid State Chem. **243**, 18 (2016)

[92] M. Songvilay, J. Robert, S. Petit, J.A. Rodriguez-Rivera, W.D. Ratcliff, F. Damay, V. Balédent, M. Jiménez-Ruiz, P. Lejay, E. Pachoud, A. Hadj-Azzem, V. Simonet, and C. Stock, Phys. Rev. B **102**, 224429 (2020)

[93] C. Kim, J. Jeong, G. Lin, P. Park, T. Masuda, S. Asai, S. Itoh, H.-S. Kim, H. Zhou, J. Ma, and J.-G. Park, arXiv:2012.06167

[94] G. Lin, J. Jeong, C. Kim, Y. Wang, Q. Huang, T. Masuda, S. Asai, S. Itoh, G. Günther, M. Russina, Z. Lu, J. Sheng, L. Wang, J. Wang, G. Wang, Q. Ren, C. Xi, W. Tong, L. Ling, Z. Liu, L. Wu, J. Mei, Z. Qu, H. Zhou, X. Wang, J.-G. Park, Y. Wan, and J. Ma, Nat. Commun. **12**, 5559 (2021)

[95] X. Hong, M. Gillig, R. Hentrich, W. Yao, V. Kocsis, A.R. Witte, T. Schreiner, D. Baumann, N. Perez, A.U.B. Wolter, Y. Li, B. Buchner, and C. Hess, Phys. Rev. B **104**, 144426 (2021)

[96] T. Halloran, F. Desrochers, E.Z. Zhang, T. Chen, L.E. Chern, Z. Xu, B. Winn, M.K. Graves-Brook, M.B. Stone, A.I. Kolesnikov, Y. Qui, R. Zhong, R. Cava, Y.B. Kim, and C. Broholm, arXiv:2205.15262 (2022)

[97] X. Liu and H.-Y. Kee, arXiv:2211.03737 (2022)

[98] S. Das, S. Voleti, T. Saha-Dasgupta, and A. Paramekanti, Phys. Rev. B **104**, 134425 (2021)

[99] P.A. Maksimov, A.V. Ushakov, Z.V. Pchelkina, Y. Li, S.M. Winter, and S.V. Streltsov, Phys. Rev. B **106**, 165131 (2022)

[100] F.-Y. Li, Y.-D. Li, Y. Yu, A. Paramekanti, and G. Chen, Phys. Rev. B **95**, 085132 (2017)

[101] A. Revelli, C.C. Loo, D. Kiese, P. Becker, T. Fröhlich, T. Lorenz, M. Moretti Sala, G. Monaco, F.L. Buessen, J. Attig, M. Hermanns, S.V. Streltsov, D.I. Khomskii, J. van den Brink, M. Braden, P.H.M. van Loosdrecht, S. Trebst, A. Paramekanti, and M. Grüninger, Phys. Rev. B **100**, 085139 (2019)

[102] B. Zhou, J. Balgley, P. Lampen-Kelley, J.-Q. Yan, D.G. Mandrus, and E.A. Henriksen, Phys. Rev. B **100**, 165426 (2019)

[103] S. Mashhadi, Y. Kim, J. Kim, D. Weber, T. Taniguchi, K. Watanabe, N. Park, B. Lotsch, J.H. Smet, M. Burghard, and K. Kern, Nano Lett. **19**, 4659 (2019)

[104] D.J. Rizzo, B.S. Jessen, Z. Sun, F.L. Ruta, J. Zhang, J.-Q. Yan, L. Xian, A.S. McLeod, M.E. Berkowitz, K. Watanabe, T. Taniguchi, S.E. Nagler, D.G. Mandrus, A. Rubio, M.M. Fogler, A.J. Millis, J.C. Hone, C.R. Dean, and D.N. Basov, Nano Lett. **20**, 8438 (2020)

[105] Y. Wan and N.P. Armitage, Phys. Rev. Lett. **122**, 257401 (2019)

[106] W. Choi, K.H. Lee, and Y.B. Kim, Phys. Rev. Lett. **124**, 117205 (2020)

13 Self Interaction Corrections to Density Functional Theory

Mark R. Pederson
Department of Physics
The University of Texas at El Paso
500 W University Avenue, El Paso, TX 79968, USA

Contents

1 Motivation **2**

2 Introduction to Fermi-Löwdin orbitals and preliminary applications **5**
2.1 Closed shell atoms 7
2.2 The closed-shell N_2 molecule 8
2.3 A closed shell electron gas: Wannier functions, FLOs and FOs 10

3 Fermi-Löwdin orbitals: an existence proof and their construction **12**
3.1 Optimizing Fermi orbital descriptors and Fermi-Löwdin orbitals using derivatives 14
3.2 Complex Fermi-orbital descriptors and complex Fermi-Löwdin orbitals 19
3.3 Atoms 21
3.4 Molecules 23
3.5 Returning to N_2: complex vs. real FLOs 24

4 Downward quantum learning: tricks for finding starting configurations **26**
4.1 Initializing FODs for principal quantum numbers with n=2, 3, and 4 29
4.2 Challenge: Simulating tetra-anionic Mn_{12}-Acetate in water 33

5 Summary and outlook **34**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Motivation

Since the earliest days of quantum mechanics researchers have philosophized about the physical meaning of electronic wavefunctions and have often sought to find rigorous means for interpreting both localized and canonical representations of the electronic degrees of freedom and for building more effective or efficient theories based upon localized orbitals [1–27]. Two of the earliest papers by Pauling and Koopmans highlight the dichotomy of viewpoints. Pauling [1] pointed out that localized sp^3 orbitals helped to explain energy differences between separated and condensed phases while Koopmans [2] showed that it was in fact canonical orbitals that best explained energy differences involving removal or addition of electrons. The discussion continues today with reasons, both conceptual and rigorous, for thinking in terms of both pictures. And the overlapping discussion spans all fields of quantum-mechanical inquiry especially in cases where computational constraints require consideration of quantum-mechanical formulations that are not unitarily invariant within their representation. Here these problems are discussed within the self-interaction correction to density functional theory and one aspect of this discussion partially unifies early work of Pauling [1], Wannier [3], Löwdin [4] and Edmiston, and Ruedenberg [5].
Before introducing the Fermi-Löwdin-Orbital formulation and discussing it within the framework of the self-interaction correction (SIC) to density-functional approximations (DFA), we motivate the use of the Fermi orbital with five questions, all of which have the same answer:

1. Can the concept of Wannier functions in condensed-matter physics [3], sp^n hybridized orbitals in atomic physics, localized molecular orbitals in molecular physics [4, 5] be formulated within a universal formulation?

2. Is there a means to generalize the concept of Wannier functions, which are currently only defined in insulating systems, to metallic systems?

3. Is there a way to create a unitary transformation, generally thought of as a discrete operator represented as an $N \times N$ matrix, that explicitly depends continuously on the one-particle density matrix?

4. Is there a way to start with a set of Kohn-Sham orbitals and define a quasi-classical "electronic geometry" that is, in some way, the reciprocal lattice of the Kohn-Sham orbitals?

5. How does one re-formulate the self-interaction correction [8] to the density-functional approximation in a manner that assures the resulting energy is both size consistent and unitarily invariant?

The answer to all of the above questions is that there is indeed a common construction, now commonly referred to as Fermi-Löwdin orbitals [19, 20] that are indeed the answer to all of the above questions. In the early 1980's Luken and coworkers [10], due to their interest in electron diffusion, considered the use of the so-called Fermi-exchange hole, $\rho(r, r')/\sqrt{\rho(r)}$, to define a set of orthonormal orbitals that sum to the total density of an electronic system. In

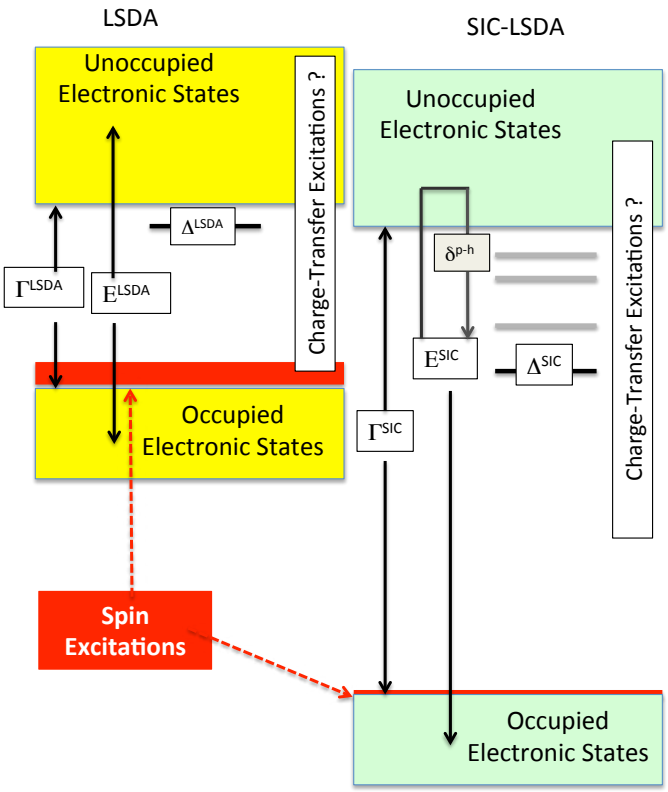


Fig. 1: *Qualitative comparison of electronic structures and excitations as approximated by the local-spin-density approximation and the self-interaction-corrected approximations to density functional theory: The HOMO/LUMO/Band gaps (Γ) increase. The location of the defect levels (δ) are moved away from the conduction band. The spin excitation spectrum narrows.*

the following, we discuss how such orbitals address one of the greatest challenges to the field of density-functional theory. In this chapter I try to compare and contrast the earliest versions of the self-interaction corrected density functional [8] with a new formalism referred to as the Fermi-Löwdin Self-Interaction Correction. The work discussed here was originally discussed as parts of Ref. [9,11,19–21,25,27] and involved collaborations with many authors who I thank and refer to here and in the acknowledgments. I have attempted to discuss the work in the most sensible way based on today's knowledge rather than on the chronological development.

SIC-induced improvements of spectroscopies

The qualitative differences between DFA and SIC-DFA are depicted in Fig. 1. The SIC pulls down the occupied states relative to the unoccupied states which generally leads to a gap (Γ) that is improved in comparison to experiment. Localized excitations in a vacuum or excitons in a wide-gap insulator (depicted as E) can be difficult to identify within LSDA calculations as they often appear above the conduction band. In DFA-SIC, with an approximation to the particle-hole interaction (δ), the description of such excitations can be improved. For defects in solids, where localized levels occupy the gap, LSDA and GGA calculations tend to place the defect levels (labeled by Δ) too close to, or overlapping, with the unoccupied conduction band. However, SIC-LSDA pulls the defect levels down and generally predicts shallow lev-

els in a qualitatively correct location relative to the conduction level. With inclusion of SIC and a particle-hole interaction, an unoccupied continuum of defect levels, expected from the Mott-Gurney theorem, begins to emerge. Consensus is emerging in regard to the differences between SIC and DFA charge-transfer excitations. Examples suggest that DFA can dramatically underestimate these energies and that an approximate self-interaction-corrected energy with a particle-hole interaction restores the correct asymptotic form for donor-acceptor and other charge-transfer excitations. For example, in a vacuum, these energies are found to be close to $I-A-1/R$, where I is the donor ionization energy, A is the acceptor electron affinity and R is the separation between the particle and hole. For spin excitations, determined from DFA-based derivations of Heisenberg Hamiltonians, a large number of calculations show that the spin-excitation energies are overestimated due to DFA's tendency to slightly delocalize the d-electrons, since the kinetic exchange interactions increase exponentially as the d-electrons delocalize. Therefore inclusion or partial inclusion of self-interaction corrections lowers the spin-excitation energies and improves agreement with experiment. While not depicted in the picture, vibrational spectra seem to be relatively well accounted for within DFA. However, since polarizabilities are dependent on SIC, the Raman intensities are also expected to show some dependence on the inclusion of self-interaction corrections. Original efforts and motivations within SIC sought to improve gaps and excitations through the use of eigenvalues. As alluded to at the very beginning of this chapter, there are conditions for which eigenvalues provide rigorous estimates for electronic processes but not for all possible changes in orbital rearrangements. While a generalized Koopmans' theorem is a good goal, changes in computational speed and the need for unambiguous accuracy will definitely favor total energy difference calculations rather than uses of Koopmans-like pictures for all possible excitations. One of the ways that the dialogue and interpretations about self-interaction-corrected and standard DFA eigenvalue differences needs to change is that when there are qualitative differences it often means that delta-SCF calculations are not possible within DFA but are within SIC. That is, a qualitatively correct electronic structure within SIC allows for total energy differences but a qualitatively incorrect DFT electronic structure does not.

In 1981 Perdew and Zunger formulated the self-interaction correction to ensure that any approximation to the density-functional would be correct in the one-electron limit. This was accomplished by modifying any approximation to the universal density-functional, $E_{xc}^{DFA} \equiv [n_\uparrow, n_\downarrow]$, according to

$$E_{xc}^{DFA} \rightarrow E_{xc}^{SIC\text{-}DFA} = E_{xc}^{DFA} - \sum_{i,\sigma} \left(U[\rho_{i,\sigma}] + E_{xc}^{approx}[\rho_{i,\sigma}, 0] \right). \tag{1}$$

In the above equation, the SIC (localized) orbitals $\{\varphi_{i\sigma}\}$ are used to define orbital densities according to: $\rho_{i\sigma}(\boldsymbol{r}) = |\varphi_{i\sigma}(\boldsymbol{r})|^2$. The terms $U[\rho_{i,\sigma}]$ and $E_{xc}^{approx}[\rho_{i,\sigma}, 0]$ are the exact self-Coulomb and approximate self-exchange-correlation energies, respectively. In the original formulation by Perdew and Zunger, the density-functional was approximated in terms of spin densities only but the formulation is not constrained by such an assumption. Modern-day approximations to the exchange-correlation energy generally include gradients of the spin densities which would then require corresponding gradients of orbital densities.

In reference to the work discussed in this section, the original PZ formulation led to a definition for the energy functional that did not transform like the density and posited that atomic-like orbitals might be the most appropriate set of orbitals for defining the SIC. This idea was formalized between 1981–1986 through the concept of localized and canonical orbitals in orbital-dependent density functional theory. It was demonstrated that, within the constraint that the orbitals used for constructing the SIC were orthonormal and constructed from a unitary transformation on the occupied orbital space, that the orbitals which minimize the self-interaction corrected functional satisfy the following equations [9]

$$\left(H_o + V_i^{SIC}\right)|\varphi_i\rangle = \sum_j \lambda_{ji}\,|\varphi_j\rangle \quad \text{and} \quad \left\langle\varphi_{i\sigma}\right| V_{i\sigma}^{SIC} - V_{j\sigma}^{SIC}\left|\varphi_{j\sigma}\right\rangle = 0, \tag{2}$$

with $V_{i\sigma}^{SIC}$ the partial functional derivative of Eq. (1) with respect to the orbital density $n_{i\sigma}$. A Jacobi-like approach for solution of the "localization equations" was shown to work well [9,11] and it was determined that in analogy to Koopmans' theorem in Hartree Fock, the eigenvalues of the Hermitian Lagrange multiplier matrix were expected to be in good agreement with experimental electron removal energies [11]. The localized orbitals obtained from these equations were found to be topologically similar to sp^3 hybrids in atoms, alternative energy-localized orbitals in molecules, and Wannier functions in insulating solids. While this formulation offered some advantages over density-functional approximations it was still not explicitly formulated as an energy that was dependent only on the density matrix. As a result this version of the self-interaction correction was neither unitarily invariant nor size consistent.

2 Introduction to Fermi-Löwdin orbitals and preliminary applications

The orbitals introduced here, for the purpose of implementation of the self-interaction correction, have been proposed by Luken *et al.* in the early 1980s and more recently (since 2014) considered for improving density-functional approximations. Within the group of scientists interested in the self-interaction error to density-functional theory they are now commonly referred to as Fermi-Löwdin orbitals because they are based on the concept of the exchange hole in Hartree-Fock, often referred to as the Fermi-exchange hole, and because their construction relies upon a technique known as Löwdin's method of symmetric orthormalization. To further motivate the derivation of the Fermi-Löwdin orbitals let us algebraically manipulate the expression for the exact exchange energy for spin σ as

$$\begin{aligned} E_\sigma^x &= -\frac{1}{2}\int d\vec{a}\int d\vec{r}\,\frac{\left|\sum_\alpha \psi_{\alpha\sigma}^*(\vec{r})\psi_{\alpha\sigma}(\vec{a})\right|^2}{|\vec{r}-\vec{a}|} = -\frac{1}{2}\int d\vec{a}\,\rho_\sigma(\vec{a})\int d\vec{r}\,\frac{\left|\sum_\alpha \psi_{\alpha\sigma}^*(\vec{r})\psi_{\alpha\sigma}(\vec{a})\right|^2}{\rho_\sigma(\vec{a})\,|\vec{r}-\vec{a}|} \\ &= -\frac{1}{2}\int d\vec{a}\,\rho_\sigma(\vec{a})\int d\vec{r}\left[\frac{\rho_\sigma(\vec{r},\vec{a})}{\sqrt{\rho_\sigma(\vec{a})}}\right]\left[\frac{\rho_\sigma(\vec{a},\vec{r})}{\sqrt{\rho_\sigma(\vec{a})}}\right]\frac{1}{|\vec{r}-\vec{a}|} \\ &= -\frac{1}{2}\int d\vec{a}\,\rho_\sigma(\vec{a})\int d\vec{r}\,\frac{\left|F_\sigma(\vec{r})\right|^2}{|\vec{r}-\vec{a}|} \end{aligned} \tag{3}$$

with the Fermi-orbital, defined as $F_{\vec{r}\sigma}(\vec{r}) \equiv \rho_\sigma(\vec{r},\vec{a})/\sqrt{\rho_\sigma(\vec{a})}$. In the above equation, the exact exchange energy has first been expressed in terms of the single-particle density matrix, $\rho_\sigma(\vec{r},\vec{a}) = \sum_\alpha \psi^*_{\alpha\sigma}(\vec{r})\,\psi_{\alpha\sigma}(\vec{a})$ and then the expression within the double integral has been multiplied and divided density $\rho_\sigma(\vec{a}) = \rho_\sigma(\vec{a},\vec{a})$. Examination of the expression shows that the Fermi-Exchange hole, evaluated at any point in space, $\vec{a}$, integrates to unity and that one can think about the exchange energy density as the interaction of the total density $\rho_\sigma(\vec{a})$ interacting with the Fermi-Exchange-hole density $F_{\vec{r}\sigma}(\vec{r})$ that is redefined at each point, $\vec{a}$.
Because the Fermi-Exchange-Hole transforms like the density under unitary transformations any quantity that depends on the Fermi-Exchange-Hole is also unitarily invariant. As such, the following reformulation of the Perdew-Zunger self-interaction was developed in terms of the spin-density-matrix and N electronic positions according to the following prescription:

1. For a trial set of Kohn-Sham (KS) orbitals $\{\psi_{\alpha\sigma}\}$ find a special set of N_σ positions in space $\{\boldsymbol{a}_{1\sigma}, \boldsymbol{a}_{2\sigma}, \ldots, \boldsymbol{a}_{N_\sigma\sigma}\}$ which provide a set of N_σ normalized linearly independent, but not orthogonal Fermi-orbitals $\{F_{1\sigma}, F_{2\sigma}, \ldots, F_{N_\sigma\sigma}\}$. By their construction from the density matrix, these Fermi-orbitals will always lie in the space spanned by the KS orbitals but are not guaranteed to span that space (For example if one defines each Fermi-orbital in terms of the same position the N Fermi-orbitals would be identical. These positions are now called Fermi-Orbital Descriptors (FODs).

2. When a set of N FODs, that provide a set of Fermi-orbitals that span the space of the Kohn-Sham orbitals is found, use Löwdin's method of symmetric orthonormalization to transform the set of FOs to a set of localized orthonormal orbitals $\{\varphi_{1\sigma}, \varphi_{2\sigma}, \ldots, \varphi_{N_\sigma\sigma}\}$ that are a unitary transformation on the KS orbitals. The resulting Fermi-Löwdin orbitals (FLOs) depend upon the set of FODs which means that the self-interaction energy also depends on the FODs.

3. Minimize the energy as a function of the KS orbitals and the set of FODs. The minimization with respect to the FODs can be performed using methods that are commonly used to optimize molecular geometries.

In this section some simple applications of this FLOSIC methodology are presented. The goal is to introduce both real and (briefly) complex SIC orbitals and compare them within the context of the original version of self-interaction corrections and the FLOSIC version of self-interaction corrections. By considering simple atoms, the N_2 molecule, and a cubic Brillouin zone with uniform density the reader should gain an understanding that complex Kohn-Sham orbitals do not require complex FODs or complex local orbitals, that there are times when symmetry considerations or energy considerations argue for the use of complex orbitals, and there are times when bond-breaking-considerations argue for FLOs that break spin symmetry. Appreciating these issues early on will help prepare practitioners for future improvements in the theory and implementation of the FLOSIC formulation. However the reader will also be prepared for the discussions about complex FLOs that appears later in this chapter.

2.1 Closed shell atoms

As a way of introducing Fermi-Löwdin orbitals and the Fermi-orbital-descriptors that define them and as a way of attempting to encourage universality across disciplines we can start by thinking about the $2sp^3$-hybrids. This also provides a simple analytical example that can be based upon the hydrogenic orbitals. For principal quantum numbers $n = 2$, the hydrogenic orbitals are given by

$$\psi_{2s}(\vec{r})=\frac{^{-r/2}\,(2-r)}{\sqrt{32\pi}} \quad \psi_{2p_x}(\vec{r})=\frac{x\quad^{-r/2}}{\sqrt{32\pi}} \quad \psi_{2p_y}(\vec{r})=\frac{y\quad^{-r/2}}{\sqrt{32\pi}} \quad \psi_{2p_z}(\vec{r})=\frac{z\quad^{-r/2}}{\sqrt{32\pi}}$$

Let us then guess that the set of FODs, that minimize the energy are determined by a tetrahedron with vertices chosen such that $\psi_{2s}(\vec{a}\,) = \psi_{2p_x}(\vec{a}\,) = \psi_{2p_y}(\vec{a}\,) = \psi_{2p_z}(\vec{a}\,)$. The condition that allows for this equality is to choose FODS which satisfy

$$|a_x| = |a_y| = |a_z| = 2 - |a| \quad \text{with} \quad |a| = \sqrt{3}|a_x|$$

Under these conditions, it then follows that a choice for the FLOs is given by

$$\begin{aligned}
|F_1\rangle &= \frac{1}{2}\big(|\psi_{2s}\rangle + |\psi_{2p_x}\rangle + |\psi_{2p_y}\rangle + |\psi_{2p_z}\rangle\big) = \big|\varphi_1^{FLO}\big\rangle \quad ! \\
|F_2\rangle &= \frac{1}{2}\big(|\psi_{2s}\rangle - |\psi_{2p_x}\rangle - |\psi_{2p_y}\rangle + |\psi_{2p_z}\rangle\big) = \big|\varphi_2^{FLO}\big\rangle \quad ! \\
|F_3\rangle &= \frac{1}{2}\big(|\psi_{2s}\rangle - |\psi_{2p_x}\rangle + |\psi_{2p_y}\rangle - |\psi_{2p_z}\rangle\big) = \big|\varphi_3^{FLO}\big\rangle \quad ! \\
|F_4\rangle &= \frac{1}{2}\big(|\psi_{2s}\rangle + |\psi_{2p_x}\rangle - |\psi_{2p_y}\rangle - |\psi_{2p_z}\rangle\big) = \big|\varphi_4^{FLO}\big\rangle \quad !
\end{aligned}$$

In the above, the exclamation mark has been added because this choice of FODs leads to the rare, but sought-after, condition that the Fermi-orbital and Fermi-Löwdin orbital are in fact identical. A second rarity is that for this special case, the FLOs also satisfy the localization equations within their subspace. A third rarity, and curiosity, is that if the set of four FODs is broken up into two pairs of FODs, the Fermi-Löwdin orbitals are invariant as the vertices are pinched toward one another. As a first illustration, calculations on some closed shell atoms will be discussed. The appearance of hybridization is common within the FLO formalism and not limited to sp-hybridization. In Fig. 2, the sp^3d^5 and $sp^3d^5f^7$ hybrids that are found for systems like Kr and Rn are illustrated.

In columns 1–4 of Table 1, the total energies for rare-gas atoms and a few lighter atoms with closed spin shells are presented. For completeness we also include the Hartree-Fock value of the energy that is calculated with the self-consistent FLOSIC orbitals. In a later section, the possibility of complex FLOs is introduced and the fifth column of this table will be discussed in that section. In preparation for that discussion it is a useful exercise to consider the possibility of multiplying any of the $2s/2p$ orbitals by $\sqrt{-1} = i$. The reader can convince themselves that the resulting FLOs would still be orthonormal and therefore a viable local orbital set within the original formulation of the self-interaction correction.

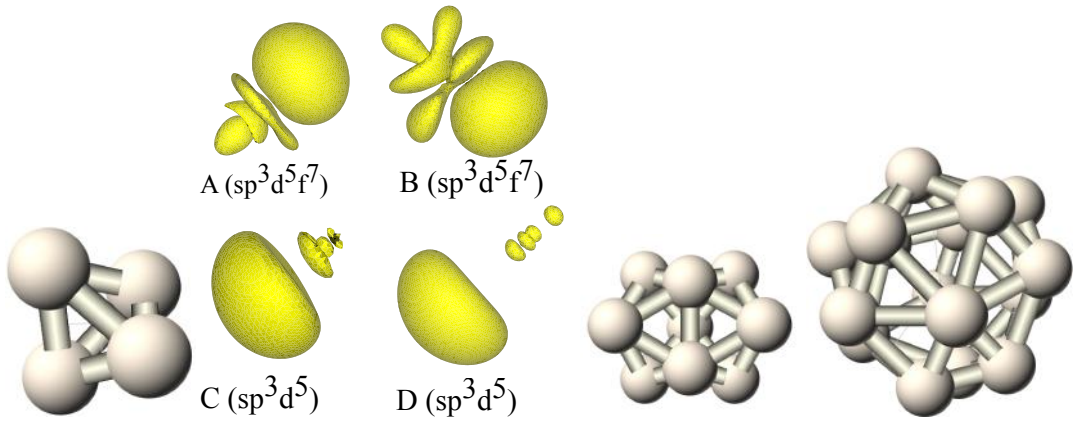


Fig. 2: *Relatively regularly polyhedra, referred to here as "flotonic solids", can be used for initializing FOD geometries that form a reciprocal lattice from closed shell atomic Kohn-Sham orbital sets. In this picture, isosurface plots of the orbital densities of inequivalent FLOs for the Radon $n = 4$ shells ($4s4p^3 4d^5 4f^7$, top) and $n = 3$ ($3s3p^3 3d^5$, bottom) shells. For the $n = 4$ shell, there are four FLOs that resemble FLO A and 12 FLOs that resemble FLO B. For the $n = 3$ shell, there are six FLOs that resemble FLO C and three FLOs that resemble FLO D. The flotonic solid for the $n = 2$ shell is a tetrahedron and it leads to standard sp^3 hybrids.*

Atom	HF (Ha)	Exp. (Ha)	LSDA (Ha)	rFLOSIC (Ha)	cFLOSIC (Ha)	ΔE (eV)
H	-0.5000	-0.500	-0.4786	-0.4999	-0.4999	0.00
He	-2.8615	-2.903	-2.8344	-2.9197	-2.9197	0.00
Li	-7.4320	-7.478	-7.3432	-7.5091	-7.5091	0.00
Be	-14.5715	-14.668	-14.4461	-14.7066	-14.7066	0.00
N	-54.3997	-54.612	-54.1342	-54.7407	-54.7578	-0.47
Ne	-128.5392	-129.053	-128.2297	-129.2805	-129.3339	-1.45
Ar	-526.7984	-528.223	-525.9395	-528.5365	-528.6767	-3.82
Kr	-2752.0206		-2750.1330	-2757.6071	-2758.0253	-11.39

Table 1: *Total energies (in Ha) of atoms from experiment, LSDA, rFLOSIC, cFLOSIC, and the total energy difference between the complex and real methods, ΔE = cFLOSICrFLOSIC (in eV). The cFLOSIC energies are evaluated using the self-consistent rFLOSIC electron density. Structures based upon the flotonic solids, described in the last section of this paper, appear to lead to the lowest energy solutions. It is only here that the Hartree-Fock energy has been systematically tracked as a function of the FLOSIC energy. So far it appears that the Hartree-Fock energy decreases as the FLOSIC energy decreases. (Calculations performed by Pederson and Withanage)*

2.2 The closed-shell N_2 molecule

The N_2 molecule, with a triple bond, is the second strongest diatomic molecule. Its isoelectronic cousin, carbon-monoxide, is slightly stronger due to similar covalent bonding and some degree of ionic enhancement. The molecule has cylindrical symmetry. Within the FLOSIC formulation, there are three inequivalent FODs given by $\vec{a}_{1s} = (0.628, 0.628, 0.628)$, $\vec{a}_{\textit{lone-pair}} = (1.311, 1.311, 1.311)$ and $\vec{a}_{\textit{banana}} = (0.702, -0.702, 0.000)$. To make contact with the discussion of sp^3 hybrids in the previous section, note that relative to the nitrogen atom, the lone-pair FOD

Mol.	LSD (PW92)	GGA (PBE)	FLOSIC-LSD (PW92)	Hartree-Fock	Expt
N_2	11.58	10.49	10.24	4.87	9.84

Table 2: *Atomization energies (eV) of N_2 as determined from LDA(PW92), GGA(PBE), FLOSIC-LSD(PW92), Hartree-Fock, and experiment. The same basis set is used for all calculations. For these calculations the nitrogen atoms were placed at the LDA equilibrium separation, at $R_\pm = \pm(0.598, 0.598, 0.598)$.*

and the three-equivalent banana-bond FODs form an almost perfect tetrahedron and thereby allow the nitrogen molecule to seamlessly dissociate into two atoms with anti-parallel spin polarization. The complete set of 14-FODs can be generated from the D_{3h} symmetry operations. Pictures of the resulting localized orbitals are shown in Fig. 3. The strongly covalent singlet N_2 molecule, $1\sigma_g^2 1\sigma_u^2 2\sigma_g^2 2\sigma_u^2 1\pi_u^4 3\sigma_g^2$, dissociates into an open-shell singlet with three unpaired $2p$ electrons per atom. For comparison, the LSD(PW92) energy functional gives an atomization energy of 11.54 eV at this bondlength and the GGA(PBE) energy functional gives an atomization energy of 10.54 eV. In Fig. 3, the valence FODs are shown pictorially. Complex FODs will be discussed in a later section. The real FODs $1s_{A\pm} = (1\sigma_g' \pm 1\sigma_u')/\sqrt{2}$, lone-pair states on the exterior of the molecule, $2sp_{A\pm} = (3\sigma_g' \pm 2\sigma_u')/\sqrt{2}$, and three bond-centered banana orbitals (e.g., $\varphi_n = \big(2\sigma_g' - \sqrt{2}[\ \ \mathrm{s}(\frac{2n\pi}{3})\pi_{ux} + \sin(\frac{2n\pi}{3})\pi_{uy}]\big)/\sqrt{3}$, with $n = -1, 0, +1$). The primes indicate that KS molecular orbitals of the same symmetry are mixed together by a unitary transformation within each irreducible representation to minimize Eq. (1). For example the $\{2\sigma_g', 3\sigma_g'\}$ are not perfect eigenstates. Instead they are determined by a nearly diagonal unitary mixture of the $\{2\sigma_g, 3\sigma_g\}$ KS eigenstates.

In the earliest implementation of self-interaction corrections to molecules a similar construction of localized orbitals for the σ states was envisioned. However for the π-states, a symmetry argument was made that the energy density should have the same symmetry as the total density. This mandates using localized SIC orbitals for the π-states as $|\varphi_\pm\rangle = \big(|\psi_{\pi_{ux}}\rangle + i\,|\psi_{\pi_{uy}}\rangle\big)/\sqrt{2}$.

Comments on Symmetry Breaking: Beginning with the earliest origins of the use of SIC there have been assertions that the use of SIC breaks symmetry. Here we should mention that there is spin-symmetry breaking, often needed in ordinary density-functional approximations, for the description of bond breaking, which is needed for most approximations to the universal functional. It is further quite common to hear assertions that there is unphysical symmetry breaking within the FLOSIC method. The experience of this author is that, while the inclusion of self-interactions leads to slightly less numerical stability than is available in standard DFT calculations, it is generally possible to find a lower energy solution that exhibits higher symmetry. A caution to readers is that when minor symmetry breaking is observed, it will always be the case that a nearby symmetrical solution will be a critical point and the goal should be to determine whether such solutions are lower in energy. Further it should be noted that when broken symmetry solutions exist, especially for spin-ordered systems, the solutions generally contain information about low-lying excited states.

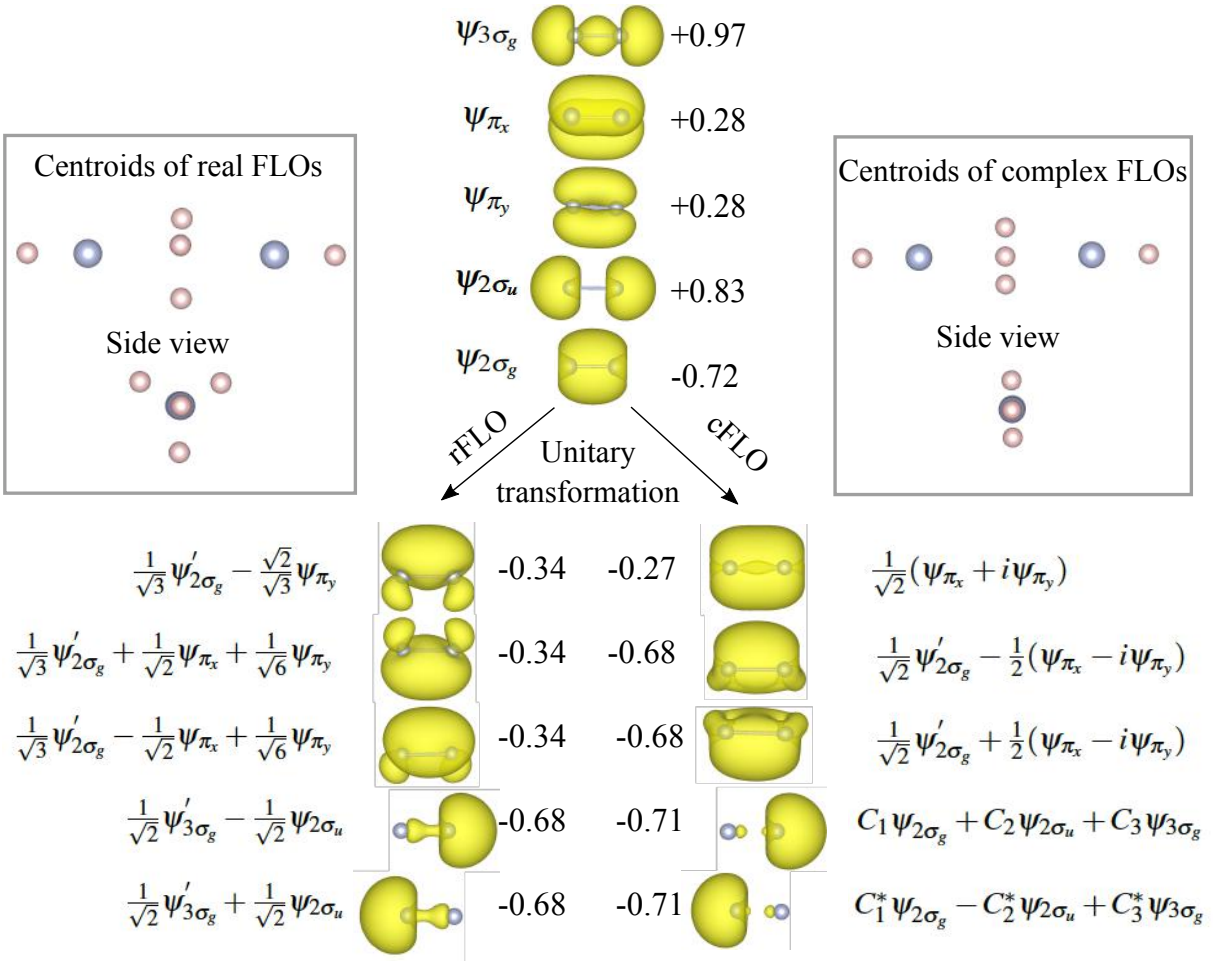


Fig. 3: *Valence orbital densities of N_2 for several different forms of density functional implementations. In the upper panel the standard canonical orbitals are shown. The shape of these orbitals have minimal variation regardless of functional choice (Exchange-Only, Perdew-Wang 92, PBE-GGA, Hartree-Fock, etc). When self-interaction corrections are included and orbitals are constrained to be real, the three bonding orbitals form banana bonds. These orbitals may be thought of as bonding combinations of sp^2 atomic orbitals) and this choice is, again, weakly dependent on the functional. However, if the constraint of reality is dropped and complex FODs are adopted, (cFLO) the orbitals soften their variability as shown on the far right. The SIC energy $E^{SIC}[n_{i\sigma}]$ (in eV) is indicated next to each orbital. The expansion of each rFLO and cFLO orbital in terms of the canonical orbitals is shown. The centroids of the rFLOs (left) and cFLOs (right) are indicated in the gray boxes. For the rFLOs, these correspond closely to the FOD positions. However, it is not generally true that the FOD position will correspond to the orbital centroid.*

2.3 A closed shell electron gas: Wannier functions, FLOs and FOs

Very early on, Wannier discovered a set of orthonormal functions, constructed from Bloch functions, that now bear his name. This work was accomplished prior to the computer age. Therefore, to make progress Wannier considered a uniform density gas. Rather that consider filling a Fermi-Sphere with plane waves, Wannier imagined approximating an insulating crystal by a cubic Brillouin zone. States on the inside of the "Wannier Cube" were occupied and states on

the outside were unoccupied. The title of his land-mark paper [3] is: *Structure of Electronic Excitation Levels in Insulating Crystals*. In it he wrote down Wannier functions of the form

$$\omega(r-R_\mu) = \frac{1}{\sqrt{\pi^3}} \prod_i \frac{\sin\left(q_F(x^i-X^i_\mu)\right)}{\sqrt{q_F}(x^i-X^i_\mu)}, \tag{4}$$

with $\boldsymbol{R}_\mu = (m_x, m_y, m_z)(\pi/q_F)$ and q_F is half the width of the cubic Brillouin zone. By inscribing the largest possible "Wannier cube" ($q_F = k_F/\sqrt{3}$) within the Fermi sphere, a specific set of Wannier functions may be derived. It has been shown early on by this author that this set of functions leads to a negative self-interaction correction for of the plane-wave states enclosed within the Wannier cube. So in the exchange-only limit, a limit that was perhaps the only tractable problem in 1937, it is highly probable that either exact exchange or the self-interaction correction would open up a HOMO-LUMO gap. By inspection, and related to the uncertainty principle, it is clear that the original Wannier functions get more localized as the magnitude of the q_F-vector gets larger. It is easy to verify that Wannier functions are a sub-class of the Fermi orbitals! The exclamation mark is included here because the original set of Wannier functions join the sp^3 hybrids as being an example where the Fermi orbitals are already orthogonal to one another. The possibility that, for any system, there is a set of FODs that leads to orthonormal FOs has been discussed and these orbitals were crowned "the most loved localized orbitals". To date there are very few exact cases, possibly only two, that satisfy this criterion. A few more comments illustrate their physical and chemical nature. At $\boldsymbol{r} = \vec{a}_{i\sigma}$, the value of the absolute square of the FO is identically equal to the total spin density at $\vec{r} = \vec{a}_{i\sigma}$. Further, the FO associated with any position, $\vec{a}_{i\sigma}$, in space is normalized to unity. Second, the absolute square of the FO is minus the exchange-hole density at $\boldsymbol{r}$ around an electron at $\vec{a}_{i\sigma}$. There have been several different attempts to find localized orbitals for the free-electron gas. Most of the work discussed in these references was in regard to the standard free-electron gas which leads to a set of occupied plane-wave states inside a Fermi sphere.

It is hypothesized that at low enough density, the limit of Wigner crystallization, despite the large increase in kinetic energy associated with the deformation of a spherical Fermi surface to a non-spherical Fermi surface, the derivative of the energy with respect to $n^{1/3}$ is more negative in the limit of $n \to 0$ (if the SIC-energy is indeed negative as is the case for the LSDA exchange-only functional). It was shown analytically that, in the low-uniform-density limit, a state that is based on a full band of plane waves/Wannier functions confined within a simple-cubic "Wannier cube" is lower in energy than the standard state composed of plane waves confined to the Fermi sphere. The author estimates that the crossover occurs at values of $\mathrm{r}_s > 35$ and that the uniform density "insulating state" (plane waves within a "Wannier cube") is lower than the metallic state. However with a new theorem that guarantees that one can find solutions of the $N-1$ electron problem when the N electron-problem is solved, it seems that some analytical progress toward defining FLOs for the metallic free-electron gas is now possible. Knowing this may very well determine if the sign of the SIC-energy of a localized orbital in the low-density limit must be negative.

3 Fermi-Löwdin orbitals: An existence proof and their construction

Given the infinite number of points, one expects that it would be rather easy to find a set of Fermi-orbital descriptors for construction of the Fermi-Löwdin orbitals. Further in the previous section, two simple examples requiring no computation have been shown to immediately provide FLOs – in fact the most loved ones.

In practice it is much more difficult to do than expected. Here we provide a proof that a solution always exists and that every time one initial solution is determined that an avalanche of solutions for smaller systems follows. To determine the Fermi-Löwdin orbitals one first constructs the overlap matrix

$$\underline{\mathrm{S}}_{ij} = \langle F_i | F_j \rangle = \frac{\sum_\alpha \psi_\alpha(\mathbf{a}_i)\psi_\alpha(\boldsymbol{a}_j)}{\sqrt{\rho(\boldsymbol{a}_i)\rho(\boldsymbol{a}_j)}} = \frac{\sum_\alpha W_{\alpha i} W_{\alpha j}}{\sqrt{\rho(\boldsymbol{a}_i)\rho(\boldsymbol{a}_j)}}\,. \tag{5}$$

Give the overlap matrix, an intermediate set of its eigenvectors is derived according to

$$\underline{\mathrm{S}}\,\vec{T}_\alpha = Q_\alpha \vec{T}_\alpha. \tag{6}$$

The eigenvectors Q_α must be greater than zero and $\sum_\alpha Q_\alpha = N$, the number of Kohn-Sham orbitals. The eigenvectors, $|T_\alpha\rangle$ are defined in terms of the Fermi orbitals according to $|T_\alpha\rangle = \sum_i |F_i\rangle$, with $\sum_\alpha T_{\alpha i} T_{\alpha j} = \delta_{ij}$ and $\sum_i T_{\alpha i} T_{\beta_i} = \delta_{\alpha\beta}$. To obtain Fermi-Löwdin orbitals from these intermediate orbitals we first normalize them, $|T'_\alpha\rangle \rightarrow |T_\alpha\rangle/\sqrt{Q_\alpha}$, and then back-transform them according to

$$|\Phi_i^{FLO}\rangle = \sum_\alpha \frac{1}{\sqrt{Q_\alpha}}\, T_{\alpha i}\, |T_\alpha\rangle. \tag{7}$$

Because the T-matrix is unitary, it is clear that in the limit of small overlap the eigenvalues (Q_α) tend to unity and that $|\Phi_i^{FLO}\rangle \rightarrow |F_i\rangle$. In the following section, several well-known cases in physics and chemistry are discussed. However, it is more common to find that some of the overlaps vanish and here the discussion focuses first on demonstrating that it is formally possible to find the Fermi-orbital overlaps. Subsequent to this discussion, ideas on how to find them for hard cases are included. To see that there is always a solution, let us start with Eq. (5) which relates the Fermi-orbital overlap matrix to the Kohn-Sham orbitals, and multiply both sides of the equation by $\sqrt{\rho(\vec{a_i})\rho(\vec{a_j})}$. We find

$$\sum_\alpha W_{\alpha i} W_{\alpha j} = \sqrt{\rho(\boldsymbol{a}_i)}\,\underline{\mathrm{S}}_{ij}\sqrt{\rho}(\boldsymbol{a}_j) = \sum_{pq} \sqrt{\rho(\boldsymbol{a}_p)}\;\delta_{ip}\;\underline{\mathrm{S}}_{pq}\;\delta_{qj}\;\sqrt{\rho(\boldsymbol{a}_j)}. \tag{8}$$

In the above equation $W_{\alpha i} = \psi_\alpha(\boldsymbol{a}_i)$, which reminds us that we can think of the left-hand side of the equation as a product of two matrices. The Kronecker delta function δ_{ip} is a convenient way to write the right-hand side as the product of three matrices. Since each element of the diagonal matrix, $\sqrt{\rho(\vec{a_p})}$ is now expressed as a manifestly positive definite matrix $P_{ip} \equiv \sqrt{\rho(\boldsymbol{a}_p)}\;\delta_{ip}$, Eq. (8) becomes, in matrix form,

$$\underline{\mathrm{W}} \times \underline{\mathrm{W}}^T = \underline{\mathrm{P}} \times \underline{\mathrm{S}} \times \underline{\mathrm{P}}^T, \tag{9}$$

and we can take the determinant of both sides which yields

$$(\underline{\mathrm{W}})^2 = \quad (\underline{\mathrm{P}})^2 \quad (\underline{\mathrm{S}}). \tag{10}$$

Inspection of the left-hand side of the equation shows that $(\underline{\mathrm{W}})$ is identically equal to the value of the many-electron Slater determinant, composed of the Kohn-Sham orbitals, evaluated at the geometry of the Fermi-orbital descriptors (multiplied by $\sqrt{N!}$). The $(\underline{\mathrm{P}})$ never vanishes except for the non-physical case that every Kohn-Sham orbital is zero at one or more FOD positions. Therefore the determinant of the FO-overlap matrix is not zero if (W) is. Since the (W) is the value of a Slater determinant constructed from the Kohn-Sham orbitals, a sufficient condition for obtaining a positive definite Fermi-orbital overlap matrix is that $\underline{\mathrm{W}}$ is invertible. Alternatively, a sufficient condition is that a many-electron wavefunction exists for the system. If a many-electron wavefunction exists, there must be at least one Slater determinant that is non-zero somewhere. Since the product of the eigenvalues of the FO-overlap matrix is identically equal to $(\underline{\mathrm{S}})$, it follows that a set of Fermi-Löwdin orbitals must exist if $(W)\neq 0$. Let us now rewrite the Slater-Determinant in terms of an alternative set of orbitals $\varphi_i = \sum U_{i\alpha}\psi_\alpha$. Since $(U\psi) = (U)(\psi)$, the many electron Slater determinant can, at most, change by a phase factor

$$\underline{\mathrm{W}} = \begin{vmatrix} \psi_1(\boldsymbol{a}_1) & \psi_2(\boldsymbol{a}_1) & \cdots & \psi_N(\boldsymbol{a}_1) \\ \psi_1(\boldsymbol{a}_2) & \psi_2(\boldsymbol{a}_2) & \cdots & \psi_N(\boldsymbol{a}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{a}_N) & \psi_2(\boldsymbol{a}_N) & \cdots & \psi_N(\boldsymbol{a}_N) \end{vmatrix} = e^{i\theta} \begin{vmatrix} \varphi_1(\boldsymbol{a}_1) & \varphi_2(\boldsymbol{a}_1) & \cdots & \varphi_N(\mathbf{a}_1) \\ \varphi_1(\boldsymbol{a}_2) & \varphi_2(\boldsymbol{a}_2) & \cdots & \varphi_N(\boldsymbol{a}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\boldsymbol{a}_N) & \varphi_2(\boldsymbol{a}_N) & \cdots & \varphi_N(\boldsymbol{a}_N) \end{vmatrix}. \tag{11}$$

The orbitals $(\varphi_1, \varphi_2, ...)$ can be any set of orbitals that are related to the Kohn-Sham orbitals (ψ_α) by a unitary transformation, including the FLOs, if the determinant of W is non-zero. In the above equation $e^{i\theta}$ would be the determinant of the unitary transformation. It follows that, if an N-electron wavefunction is not zero everywhere, there is at least one set of KS orbitals and one set of FODs for any non-zero N-electron wavefunction. Therefore it is guaranteed that it is always possible to find an initial set of KS orbitals and FODs. Once accomplished the gradient techniques discussed in Sec. 3.1 may be used reach a stationary point. It will be important for FLOSIC practitioners to learn when such stationary points are ground states and when they represent excitations.

An avalanche of solutions

Because it is in fact difficult to find a starting solution, it is then worthwhile to make the most out of every human-determined starting solution found. Here we consider the manifestations of finding one viable solution. Since an $N\times N$ determinant can be analytically represented in turns of a sum of N cofactors, it means that if an $N\times N$ determinant is non zero, that at least one of its co-factors ($N-1 \times N-1$ determinants) is non zero. This means that once *one* initial solution is found, a minimum of $N-1$ and a maximum of 2^N-2 new starting solutions are found for systems with fewer electrons. This provides a big advantage from the standpoint

of finding solutions. Further, once a non-positive-definite overlap matrix is found for a set of N-orthonormal orbitals, the resulting set of FLOs almost satisfy the following expression

$$\varphi_i^{FLO}(\boldsymbol{a}_j) \approx \sqrt{\rho(\boldsymbol{a}_i)}\,\delta_{ij}. \tag{12}$$

In terms of the ultimate goal of using quasi-classical electronic positions to better inform density-functional development, this proof suggests an alternative criterion for choosing FODs for defining the self-interaction correction (SIC). One could maximize the determinant of the FO-overlap matrix and then assert that the SIC energy is being defined by a region where the amplitude of the many-electron KS Slater determinant was large. However, for now, we stick to the necessity of subsequent optimization of the FODs based on energy minimization.

Existence proofs provide guarantees that solutions can be found but do not necessarily provide a set of directions on how to find them. We will return to strategies for finding initial starting points in a later section but first turn to the equations that are needed to find the optimal set of Fermi-orbital descriptors once an initial starting set has been found.

3.1 Optimizing Fermi orbital descriptors and Fermi-Löwdin orbitals using derivatives

Before embarking upon this section it is important to distinguish between Fermi-orbital descriptors (FODs) and Fermi-Löwdin-orbital centroids. FODs are the variational parameters. Fermi-Löwdin-orbital centroids correspond to the center of gravity $\langle\varphi_i^{FLO}|\vec{r}\,|\varphi_i^{FLO}\rangle$. The centroids are determined by the full set of descriptors. Sometimes, for example for sigma bonds, the FOD and FLOC are similar but generally this is not the case.

To efficiently use a Fermi-orbital based construction of localized orbitals in large systems it is necessary to have an analytic expression for the derivatives of the orbital-dependent energy terms with respect to these classical electronic positions. An additional goal for the next period of time should be to determine second derivatives of energy with respect to the Fermi-orbital descriptors (FODs). In the following the details that go into determining FOD gradients are discussed. The resulting equations depend on the evaluation of N sparse $N\times N$ matrices and one full $N\times N$ matrix. As such the evaluations of the FOD gradients and subsequent optimization of the SIC-DFT Hamiltonian is not necessarily more complicated than standard density-functional-based methods. For simplicity, spin indices on the orbitals and the spin-density matrices are suppressed. Since the goal of this section is to develop analytic expressions for transformations between various orbital sets, there are four different sets of orbitals that need to be considered. These four sets will were initially referred to as the Kohn-Sham orbitals (KS), the Fermi Orbitals (FO), the Intermediate Löwdin Orbitals (ILO), and the Localized Löwdin Orbitals (LLO), respectively [20, 21]. However, the nomenclature for the latter set is now Fermi-Löwdin Orbitals (FLOs). Also, in the original discussions of gradients [20, 21], the positions were imprecisely referred to as Fermi-orbital centroids rather than Fermi-orbital derivatives. The gradient of the FLO with respect to the FOD is needed to determine the optimal

Fermi-orbital descriptors for constructing the SIC energy. It is given by

$$\nabla_{\boldsymbol{a}_{i\sigma}} F_{i\sigma}(\boldsymbol{r}) = \frac{\sum_\alpha \psi^*_{\alpha\sigma}(\boldsymbol{r})\, \nabla_{\boldsymbol{a}_{i\sigma}} \psi_{\alpha\sigma}(\boldsymbol{a}_{i\sigma})}{\sqrt{\rho(\boldsymbol{a}_{i\sigma})}} - \frac{F_{i\sigma}(\boldsymbol{r}) \nabla_{\boldsymbol{a}_{i\sigma}} \rho(\boldsymbol{a}_{i\sigma})}{2\rho(\boldsymbol{a}_{i\sigma})} \tag{13}$$

$$= \sum_\alpha F^\alpha_{i\sigma} \left(\frac{\nabla_{\boldsymbol{a}_{i\sigma}} \psi_{\alpha\sigma}(\boldsymbol{a}_{i\sigma})}{\psi_{\alpha\sigma}(\boldsymbol{a}_{i\sigma})} - \frac{\nabla_{\boldsymbol{a}_{i\sigma}} \rho(\boldsymbol{a}_{i\sigma})}{2\rho(\boldsymbol{a}_{i\sigma})} \right) \quad \text{with} \quad F^\alpha_{i\sigma} \equiv \frac{\psi^*_{\alpha\sigma}(\boldsymbol{r}) \psi_{\alpha\sigma}(\boldsymbol{a}_{i\sigma})}{\sqrt{\rho(\boldsymbol{a}_{i\sigma})}}\,. \tag{14}$$

By construction, each of the FOs is normalized to unity and the set of Fermi orbitals spans the same space as the KS and FLO orbitals as long as the set of FODs, $a_{i\sigma}$, are far enough from one another. The ILO also span the space of the KS and FLO but they are neither orthogonal in orbital space nor normalized. The ILO are referred to as $|T_\alpha\rangle$ in the forthcoming equations. Their normalization, Q_α, is determined by the standard (i.e. non-general) diagonalization of the FO-overlap matrix according to

$$\sum_j S_{ij} T_{\alpha j} = Q_\alpha T_{\alpha i} \quad \text{where} \quad |T_\alpha\rangle = \sum_j T_{\alpha j} |F_j\rangle \quad \text{and} \quad S_{ij} = \langle F_i | F_j \rangle \tag{15}$$

The FLO, designated by φ_k, are then constructed from the ILO and their associated eigenvalues according to

$$|\varphi_k\rangle = \sum_{\alpha j} \frac{1}{\sqrt{Q_\alpha}} T_{\alpha k} T_{\alpha j} |F_j\rangle \equiv \sum_j \varphi^F_{kj} |F_j\rangle. \tag{16}$$

Because the FO overlap matrix is real and symmetric, the inverse transformation between Fermi orbitals and the FLO is also determined from

$$|F_l\rangle = \sum_{\beta n} T_{\beta l} \sqrt{Q_\beta} T_{\beta n} |\varphi_n\rangle. \tag{17}$$

Although no analytical use for the following equation has been identified, it is formally interesting and computationally useful to note that, since the FLO are unitarily equivalent to the KS orbitals, the Fermi orbitals may also be constructed from the FLO according to

$$F_{l\sigma}(\boldsymbol{r}) = \frac{\sum_n \varphi^*_{n\sigma}(\boldsymbol{r})\, \varphi_{n\sigma}(\boldsymbol{a}_{l\sigma})}{\sqrt{\sum_n |\varphi_{n\sigma}(\boldsymbol{a}_{l\sigma})|^2}} \equiv \sum_\alpha C^\sigma_{ln}\, \varphi_{n\sigma}(\boldsymbol{r}). \tag{18}$$

In other words, the inverse transformation coefficients determined by Eq. (17) are also determined by knowing the values of the FLO at the FODs. Possibly this equality could be used to derive something useful. Even if it is not found to be useful for analytic purposes, it is computationally useful since it shows that one can use a previous guess of FLOs to construct a new set of FO and that the resulting matrices will be sparse. Assuming an initial set of FODs, $\{\boldsymbol{a}_m\}$, have been determined the derivative of the total SIC energy with respect to a specific FOD is

given by

$$\frac{dE^{SIC}}{da_m} = \sum_k \Big(\langle \frac{d\varphi_k}{da_m} | V^k_{SIC} | \varphi_k \rangle + \langle \varphi_k | V^k_{SIC} | \frac{d\varphi_k}{da_m} \rangle \Big) \tag{19}$$

$$= \sum_{kl} \Big(\langle \frac{d\varphi_k}{da_m} | \varphi_l \rangle \langle \varphi_l | V^k_{SIC} | \varphi_k \rangle + \langle \varphi_k | V^k_{SIC} | \varphi_l \rangle \langle \varphi_l | \frac{d\varphi_k}{da_m} \rangle \Big) \tag{20}$$

$$= \sum_{kl} \lambda^k_{kl} \Big(\langle \frac{d\varphi_k}{da_m} | \varphi_l \rangle + \underbrace{\langle \varphi_l | \frac{d\varphi_k}{da_m} \rangle + \langle \frac{d\varphi_l}{da_m} | \varphi_k \rangle}_{= d\langle \varphi_l | \varphi_k \rangle / d} - \langle \frac{d\varphi_l}{da_m} | \varphi_k \rangle \Big) \tag{21}$$

$$= \sum_{kl} \lambda^k_{kl} \Big(\langle \frac{d\varphi_k}{da_m} | \varphi_l \rangle - \langle \frac{d\varphi_l}{da_m} | \varphi_k \rangle \Big) \equiv {\sum_{kl}}' \, \lambda^k_{kl} \, \Delta_{lk,m}, \tag{22}$$

with $\lambda^k_{kl} \equiv \langle \varphi_l | H_o + V^{SIC}_k | \varphi_k \rangle$. This is a general formula that does not depend on the Kohn-Sham orbitals being eigenstates of any Hamiltonian. Eq. (22) follows from Eq. (19) because the FO construction does not allow a FO to escape the space of the KS or FLO. In cases where one is determining analytic derivatives (such as Hellmann-Feynman forces), the orthonormality of the KS orbitals and the symmetry of the Lagrange-multiplier matrix leads to a simplification of the derivatives. In this more complicated case, the derivative of a FLO with respect to a FOD is determined by differentiating Eq. (16) and remembering that Q_α and $T_{\alpha k}$ depend on the entire set of FODs, $\{\vec{a}_m\}$, for any value of k but that the FO is only dependent on its own FOD. Since $|\varphi_k\rangle = \Sigma_{\alpha j} \frac{1}{\sqrt{Q_\alpha}} T_{\alpha k} T_{\alpha j} |F_j\rangle$ and remembering that only F_m depends on a_m, it follows that

$$\Big| \frac{d\varphi_k}{da_m} \Big\rangle = |D_{1,km}\rangle + |D_{2,km}\rangle + |D_{3,km}\rangle \equiv \sum_l \Delta^m_{kl} \, |\varphi_l\rangle \tag{23}$$

$$|D_{1,km}\rangle = \sum_{\alpha j} \frac{1}{\sqrt{Q_\alpha}} T_{\alpha k} T_{\alpha j} \Big| \frac{dF_j}{da_m} \Big\rangle = \sum_\alpha \frac{1}{\sqrt{Q_\alpha}} T_{\alpha k} T_{\alpha m} \Big| \frac{dF_m}{da_m} \Big\rangle \tag{24}$$

$$|D_{2,km}\rangle = -\frac{1}{2} \sum_{\alpha j} \frac{1}{Q^{3/2}_\alpha} \frac{dQ_\alpha}{da_m} T_{\alpha k} T_{\alpha j} |F_j\rangle \tag{25}$$

$$|D_{3,km}\rangle = \sum_{\alpha j} \frac{1}{\sqrt{Q_\alpha}} \Big(\frac{dT_{\alpha k}}{da_m} T_{\alpha j} + T_{\alpha k} \frac{dT_{\alpha j}}{da_m} \Big) |F_j\rangle \tag{26}$$

Each term $|D_{n,km}\rangle$ for $n = 1, 2, 3$ in the above equation is discussed separately:

Term 1

To determine the direct part of the FOD derivative, according to Eq. (22), the interest is in determining the difference of the matrix elements $\langle \varphi_l | D_{1,km} \rangle - \langle \varphi_k | D_{1,lm} \rangle$ which, using the expansion for the FLO in terms of the FO, Eq. (16), ($|\varphi_l\rangle = \sum_{\beta n} \frac{1}{\sqrt{Q_\beta}} T_{\beta l} T_{\beta n} |F_n\rangle$), can be rewritten as

$$\langle \varphi_l | D_{1,km} \rangle = \sum_\alpha \frac{T_{\alpha k} T_{\alpha m}}{\sqrt{Q_\alpha}} \langle \varphi_l | \frac{dF_m}{da_m} \rangle = \sum_{\alpha \beta n} \frac{T_{\alpha k} T_{\alpha m}}{\sqrt{Q_\alpha}} \frac{T_{\beta l} T_{\beta n}}{\sqrt{Q_\beta}} \langle F_n | \frac{dF_m}{da_m} \rangle \tag{27}$$

The difference of these two terms is then given by

$$\Delta^1_{lk,m} \equiv \langle\varphi_l|D_{1,km}\rangle - \langle\varphi_k|D_{1,lm}\rangle = \sum_{\alpha\beta n} \frac{T_{\alpha k}T_{\alpha m}T_{\beta l}T_{\beta n} - T_{\alpha l}T_{\alpha m}T_{\beta k}T_{\beta n}}{\sqrt{Q_\alpha Q_\beta}} \big\langle F_n\big|\frac{dF_m}{da_m}\big\rangle \tag{28}$$

$$= \sum_{\alpha\beta n} \frac{\big(T_{\alpha k}T_{\beta l} - T_{\alpha_l}T_{\beta k}\big)T_{\alpha m}T_{\beta n}}{\sqrt{Q_\alpha Q_\beta}} \big\langle F_n\big|\frac{dF_m}{da_m}\big\rangle. \tag{29}$$

Using $\varphi^F_{kj} = T_{\alpha k}T_{\alpha j}/\sqrt{Q_\alpha}$ defined in Eq. (16), it can be rewritten in the manifestly sparse form

$$\Delta^1_{lk,m} = \sum_n \big(\varphi^F_{km}\varphi^F_{ln} - \varphi^F_{lm}\varphi^F_{kn}\big)\big\langle F_n\big|\frac{dF_m}{da_m}\big\rangle = \varphi^F_{km}\big\langle\varphi_l\big|\frac{dF_m}{da_m}\big\rangle - \varphi^F_{lm}\big\langle\varphi_k\big|\frac{dF_m}{da_m}\big\rangle. \tag{30}$$

Term 2

Term 2 always vanishes: The term that is due to the gradient of the ILO eigenvalue with respect to a Fermi-orbital position becomes, inserting the expression for $|F_j\rangle$ given in Eq. (17)

$$|D_{2,km}\rangle = -\frac{1}{2}\sum_{\alpha j}\frac{1}{Q_\alpha^{3/2}}\frac{dQ_\alpha}{da_m}T_{\alpha k}T_{\alpha j}|F_j\rangle = -\frac{1}{2}\sum_{\alpha j\beta n}\frac{1}{Q_\alpha^{3/2}}\frac{dQ_\alpha}{da_m}T_{\alpha k}T_{\alpha j}\ T_{\beta j}\sqrt{Q_\beta}T_{\beta n}|\varphi_n\rangle. \tag{31}$$

It may be further simplified using the orthonormality of the ILOs, $\sum_j T_{\beta j}T_{\alpha j} = \delta_{\alpha\beta}$,

$$|D_{2,km}\rangle = -\frac{1}{2}\sum_{\alpha n}\frac{1}{Q_\alpha^{3/2}}\frac{dQ_\alpha}{da_m}T_{\alpha k}\sqrt{Q_\alpha}T_{\alpha n}|\varphi_n\rangle = -\frac{1}{2}\sum_{\alpha n}\frac{1}{Q_\alpha}\frac{dQ_\alpha}{da_m}T_{\alpha k}T_{\alpha n}|\varphi_n\rangle \tag{32}$$

so that, from the orthonormality, $\langle\varphi_l|\varphi_n\rangle = \delta_{ln}$, it follows that

$$\langle\varphi_l|D_{2,km}\rangle - \langle\varphi_k|D_{2,lm}\rangle - \frac{1}{2}\sum_\alpha\frac{1}{Q_\alpha}\frac{dQ_\alpha}{da_m}\big(T_{\alpha k}T_{\alpha l} - T_{\alpha l}T_{\alpha k}\big) = 0. \tag{33}$$

Term 3

In Ref. [20, 21] it was demonstrated that the analytical expression for the vector $\Delta_{kl,m}$ can be determined from sparse matrix manipulations and that the memory requirements scale better than N^2. In the above equations, the contributions from the direct $|D_{1,km}\rangle$ are analogous to the standard Hellmann-Feynman derivative since it arises from the explicit dependence of the FLO on the FOD. The terms involving $|D_{3,km}\rangle$ are more complicated. Qualitatively, these terms would be zero if there was a Hellmann-Feynman theorem or alternatively if the FLO actually satisfied the localization equations. However, because the FO-based formulation of the PZ-SIC already leads to unitary invariance there is neither required nor correct to force the localization equations [9, 11] need to be satisfied. The sum of the eigenvalues Q_α is always equal to the dimension of the KS space. The eigenvalues are bounded from below by zero and above by the dimension of the KS space. For reasonable guesses of the FODs, the eigenvalues are close to unity. For now, to make analytical progress, it is a assumed that FODs have been chosen that break all degeneracies of the FO-overlap derivatives, or that a small perturbation that breaks all

degeneracies has been added to Eq. (15). Then, a first-order perturbation analysis of Eq. (15) can be used

$$\frac{dT_{\alpha k}}{da_m} = \sum_{\beta\neq\alpha} T_{\beta k}\frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle}{Q_\alpha - Q_\beta} \tag{34}$$

to determine that

$$|D_{3,km}\rangle = \sum_j |F_j\rangle \sum_{\alpha,\beta\neq\alpha} \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle}{Q_\alpha^{1/2}(Q_\alpha - Q_\beta)}\big(T_{\beta k}T_{\alpha j} + T_{\alpha k}T_{\beta j}\big) \tag{35}$$

$$= \sum_j |F_j\rangle \sum_{\alpha,\beta>\alpha} \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle}{(Q_\alpha - Q_\beta)}\big(T_{\beta k}T_{\alpha j} + T_{\alpha k}T_{\beta j}\big)\left(\frac{1}{Q_\alpha^{1/2}} - \frac{1}{Q_\beta^{1/2}}\right) \tag{36}$$

$$= \sum_j |F_j\rangle \sum_{\alpha,\beta>\alpha} \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle\big(T_{\beta k}T_{\alpha j}+T_{\alpha k}T_{\beta j}\big)(Q_\beta^{1/2}-Q_\alpha^{1/2})}{(Q_\alpha - Q_\beta)(Q_\alpha Q_\beta)^{1/2}} \tag{37}$$

$$= -\sum_j |F_j\rangle \sum_{\alpha,\beta>\alpha} \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle\big(T_{\beta k}T_{\alpha j}+T_{\alpha k}T_{\beta j}\big)}{(Q_\alpha^{1/2}+Q_\beta^{1/2})(Q_\alpha Q_\beta)^{1/2}} \tag{38}$$

$$= -\sum_n |\varphi_n\rangle \sum_{\alpha,\beta>\alpha} \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle\big(T_{\beta k}T_{\alpha n}Q_\alpha^{1/2}+T_{\alpha k}T_{\beta n}Q_\beta^{1/2}\big)}{(Q_\alpha^{1/2}+Q_\beta^{1/2})(Q_\alpha Q_\beta)^{1/2}} \tag{39}$$

where in the last step we have used Eq. (17) and the orthonormality $\sum_j T_{\beta j}T_{\alpha j} = \delta_{\alpha\beta}$

$$\sum_j \big(T_{\beta k}T_{\alpha j}+T_{\alpha k}T_{\beta j}\big)|F_j\rangle = \sum_{\gamma n j} \big(T_{\beta k}T_{\alpha j}+T_{\alpha k}T_{\beta j}\big)T_{\gamma j}Q_\gamma^{1/2}T_{\gamma n}|\varphi_n\rangle \tag{40}$$

$$= \sum_n \big(T_{\beta k}T_{\alpha n}Q_\alpha^{1/2}+T_{\alpha k}T_{\beta n}Q_\beta^{1/2}\big)|\varphi_n\rangle. \tag{41}$$

It is noted that at the end of the analysis there is no division by zero in the first-order perturbative expressions even when the energy denominator vanishes! Now, to circle back to Eq. (22), it is possible to evaluate the difference

$$\Delta^3_{lk,m} \equiv \langle\varphi_l|D_{3,lm}\rangle - \langle\varphi_k|D_{3,lm}\rangle \tag{42}$$

$$= -\frac{1}{2}{\sum_{\alpha\beta}}' \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle\Big(\big(T_{\beta k}T_{\alpha n}Q_\alpha^{1/2}+T_{\alpha k}T_{\beta n}Q_\beta^{1/2}\big)-\big(T_{\beta k}T_{\alpha n}Q_\alpha^{1/2}+T_{\alpha k}T_{\beta n}Q_\beta^{1/2}\big)\Big)}{(Q_\alpha^{1/2}+Q_\beta^{1/2})(Q_\alpha Q_\beta)^{1/2}}$$

$$= -\frac{1}{2}{\sum_{\alpha\beta}}' \frac{\langle T_\beta|\frac{dS}{d}|T_\alpha\rangle}{(Q_\alpha^{1/2}+Q_\beta^{1/2})(Q_\alpha Q_\beta)^{1/2}}\big(T_{\alpha k}T_{\beta l} - T_{\alpha l}T_{\beta k}\big)\big(Q_\beta^{1/2}-Q_\alpha^{1/2}\big). \tag{43}$$

Expanding

$$\big\langle T_\beta\big|\frac{dS}{da_m}\big|T_\alpha\big\rangle = \sum_i \Big(T_{\beta i}\frac{dS_{im}}{da_m}T_{\alpha m} + T_{\beta m}\frac{dS_{mi}}{da_m}T_{\alpha i}\Big) = \sum_i \frac{dS_{im}}{da_m}\big(T_{\beta i}T_{\alpha m} + T_{\beta m}T_{\alpha i}\big) \tag{44}$$

and combining the above two equations together a computationally useful expression is found:

$$\Delta^3_{lk,m} = \frac{1}{2}\sum_{\alpha\beta i}\frac{dS_{im}}{da_m}\big(T_{\beta i}T_{\alpha m}+T_{\beta m}T_{\alpha i}\big)\big(T_{\alpha k}T_{\beta l}-T_{\alpha l}T_{\beta k}\big)\frac{Q_\alpha^{1/2}-Q_\beta^{1/2}}{(Q_\alpha^{1/2}+Q_\beta^{1/2})(Q_\alpha Q_\beta)^{1/2}}\,. \tag{45}$$

The above forms reduce the expression to the calculation of quantities that are symmetric under interchange of α and β and in terms of quantities that exhibit the sparsity. It shows that, if the calculation is performed in terms of Kohn-Sham eigenstates, the calculation of the derivatives may scale as poorly as N^4 in the small N limit. However, if one constructs the FOs in terms of the FLOs, the T matrix then becomes sparse and each bracketed quantity scales as N^2. The sparsity of the FLO SIC matrix elements ($\langle\varphi_k|H_o+V_k^{SIC}|\varphi_l\rangle$ already shows that one never needs to calculate all possible products of $T_{\alpha k}T_{\beta l}$. It is also noted that one does not need to calculate the contributions due to H_o since that contribution to the λ_{ij}^i matrices is always symmetric. To determine the final derivatives one simply evalutates $\Delta_{lk,m}$ in Eq. (22) according to

$$\Delta_{lk,m} = \Delta_{lk,m}^1 + \Delta_{lk,m}^3. \tag{46}$$

Once the gradients are calculated, the great expectation was that it would be as easy to step to the local minimum as it is for gradient-based optimizations of molecular, cluster, and solid-state geometries. While it is probably the case that the best possible optimization method has not been found, some lessons from optimization of molecular geometries still hold. First, as is the case for molecular geometries, the LBFGS method (implemented in FLOSIC by Jackson and Withanage) is generally more efficient than conjugate-gradient when one is reasonably close to a solution. Second, the use of approximations of second derivatives, determined by atomic calculations but then used for all systems, also provides for more efficient stepping to the minimum. Finally, the use of conjugate-gradient methods is more stable when one is far from solution as is the case for the optimization of molecular geometries. Additional efforts are being made to more efficiently use force information. A final comment is that since $\frac{d}{d_p}\langle\varphi_k|\varphi_l\rangle = 0$ for any values of m, k, l, and p, it is possible that higher order analytic derivatives can be determined in the frozen density regime.

3.2 Complex Fermi-orbital descriptors and complex Fermi-Löwdin orbitals

In one of the earliest applications of SIC to molecules, Pederson [9] suggested that the bonding π_u states should be complex to ensure that the symmetry was not broken. Early consideration of complex orbitals by Klüpfel *et al.* [18] were based on an energy minimization. The Fermi orbital is explicitly real if it is possible to represent the Kohn-Sham orbitals as real orbitals and if the FODs are constrained to be real. So in this sense the 2014 version of FLOSIC represented a self-consistent real theory. This section introduces the use of complex Fermi-orbital descriptors (FODs) in the Fermi-Löwdin self-interaction-corrected density functional theory (FLOSIC). With complex FODs, the Fermi-Löwdin orbitals (FLOs) used to evaluate the SIC correction to the total energy become complex. Complex FLO-SIC (cFLOSIC) calculations based on the local spin density approximation generally produce lower total energies than those found with FLOSIC restricted to real orbitals (rFLOSIC). The cFLOSIC results are qualitatively similar to earlier energy-localized Perdew-Zunger SIC (PZ-SIC) calculations using complex orbitals [18]. The energy lowering stems from the exchange-correlation part of the self-interaction correction.

The Hartree part of the correction is more negative in rFLOSIC. The energy difference between real and complex solutions is greater for more strongly hybridized FLOs in atoms and for FLOs corresponding to double- and triple-bonds in molecules. The case of N_2 is examined in detail to show the differences between the real and complex FLOs. We show that the complex triple-bond orbitals are simple and physically appealing combinations of π and σ_g orbitals that have been discussed only recently [25]. Consideration of complex FODs, and resulting unitary transformations, underscores the fact that FLO centroids are not necessarily good guesses for FOD positions in a FLOSIC calculation (It is for this reason that we pointed out earlier that the use of the term Fermi-orbital centroids was imprecise). Another reason to introduce complex FODs into FLOSIC is more practical. Gradient searches for optimal FOD positions fail when the FODs obtained at a given step in the search produce Fermi orbitals that are not linearly independent. Transition metal systems are particularly prone to this problem because the $n = 3$ orbitals have considerable spatial overlap. Complex FODs provide a larger parameter space and lead to smoother, numerically more tractable, orbitals. Using complex FODs as starting points helps avoid non-positive-definite FO-overlap matrices and makes the search process more efficient. The complex Fermi orbitals are defined by

$$F_i(\boldsymbol{r}) = \frac{\sum_\alpha \psi^*_\alpha(\boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i)\psi_\alpha(\boldsymbol{r})}{\sqrt{\sum_\alpha \psi^*_\alpha(\boldsymbol{a}_i+\mathrm{i}\mathbf{b}_i)\psi_\alpha(\boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i)}}\,. \tag{47}$$

In the above, the FODs are allowed to be complex, $\boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i$. Evaluating the ψ_α at complex positions leads to complex Fermi orbitals, F_i. The complex FOs have the same, and orthonormal complex FLOs (cFLOs) have similar characteristics. In calculations using Gaussian-type basis functions, evaluating $\psi_\alpha(\boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i)$ gives rise to positive exponential terms that can cause numerical difficulties. Because the terms appear in the numerator and denominator, they can be managed if handled carefully. Within a Gaussian-orbital construction, the wave function is ultimately decomposed in terms of polynomials (which become complex in cFLOSIC) and Gaussian envelope functions of the form $(-\beta(\boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i-\boldsymbol{A})^2)$, where $\boldsymbol{A}$ usually coincides with the position of an atom. The real part of the exponent becomes $-\beta(|\boldsymbol{a}_i-\boldsymbol{A}|^2-|\boldsymbol{b}_i|^2)$. In general, the values of β span many orders of magnitude ranging from 0.02 to approximately $50\,Z^{10/3}$ for atomic number Z. There is a combination of the quantities $\{\beta, \mathbf{A}, \boldsymbol{a}_i+\mathrm{i}\boldsymbol{b}_i\}$ that leads to the most positive exponent Γ_{max} that can be determined by sweeping through all combinations of Gaussian decay parameters, atomic positions, and FOD parameters. Γ_{max} can then be subtracted from all the exponents in the Gaussians prior to evaluating the exponential. This effectively multiplies the numerator and denominator of Eq. (47) by the same number and ensures that none of the terms diverge prior to division. This has no effect on the computer time. A similar approach will be needed for plane-wave basis functions (which as noted elsewhere will be interesting for finding FLOs for metallic model systems). In such cases, the envelope functions would have the form $(\boldsymbol{k}\boldsymbol{b}_i)$. Again, multiplying and dividing each plane wave by the appropriate largest exponential will ensure that there are no numerical problems associated with exponentiation. For starting guesses, it is possible to choose the imaginary part of the FOD such that:$|\mathbf{b_i}| < |\mathbf{a_i}-\mathbf{A}|$, where $\mathbf{A}$ is the nuclear position closest to $\mathbf{a_i}$, to ensure decaying functions. One way to do this

is to choose $\boldsymbol{b}_i = \quad \mathrm{s}(\alpha)|\boldsymbol{A}-\boldsymbol{a}_i|\boldsymbol{u}$ with $\boldsymbol{u} = \big(\sin(\theta) \quad \mathrm{s}(\phi), \sin(\theta)\sin(\phi), \quad \mathrm{s}(\theta)\big)$. For the cFLOSIC results shown below, the FOD positions, $\boldsymbol{a}_i+\boldsymbol{b}_i$, are updated using a gradient optimization scheme. To optimize the complex FOD positions, we began by adding small random imaginary parts to the optimized (real) FODs from the corresponding rFLOSIC calculation. For each atom or molecule studied, 100 random complex FOD sets were generated. For each, the cFLOs were created and the corresponding SIC energies were calculated. The set with the lowest SIC energy was chosen as the starting point for a gradient optimization using energy gradients corresponding to both the real and imaginary parts of the FODs. The gradients corresponding to the imaginary parts required computing a numerical derivative of the orbital with respect to the component. The basis sets and integration grids used in the rFLOSIC and cFLOSIC calculations reported here were identical.

3.3 Atoms

Table 1 presents the total energies of atoms from Hartree-Fock LSDA, rFLOSIC and cFLOSIC calculations in Hartree (Ha) units. The energy difference cFLOSIC – rFLOSIC is also shown in electron volts (eV) for each atom. For most of the atoms in Table 1 the LSDA total energy is higher than the experimental reference energy, while the rFLOSIC and cFLOSIC total energies are lower. Thus both rFLOSIC and cFLOSIC correct the atomic total energies in the right direction, but often over-correct. The rFLOSIC and cFLOSIC total energies are identical up to Be. After that, the cFLOSIC energies are always lower. For Ne, Ar, and Kr, for example, the cFLOSIC energy is lower by -1.45, -3.82, and -11.39 eV, respectively. The Jackson-Withanage analysis can be used to identify how close the FLOSIC energies are to the SIC calculations with full variational freedom. They previously compared the FLOSIC method against the traditional implementation of PZ-SIC with full variational freedom (where a localization condition is invoked, known as the SIC-LE method) and have shown that rFLOSIC orbitals satisfy the LEs (symmetric Lagrange multiplier matrix) up to carbon. In real FLOSIC, the number of constraints is $3N$ where N is the number of occupied orbitals while the number of constraints in real orbital SIC-LE is $N(N-1)/2$. We find that orbitals from cFLOSIC (where the number of constraints is $6N$) satisfy the LEs (Hermitian Lagrange multiplier matrix) for complex orbital SIC up to Ne with or without freezing the $1s$ FOD. We understand that this is due to the added variational freedom due to the complex FODs.

Since the rFLOSIC and cFLOSIC total energies in Table 1 are evaluated based on the same total electron density, the DFT parts of the energies are the same and any differences are entirely due to the SIC corrections. In Ref. [25] the values of the corrections for the valence electron local orbitals of selected first row atoms are shown. Values for $n = 3$ and $4s$ local orbitals for Zn are also shown. The corrections are shown for both the majority and minority spin channels (for the spin-polarized cases). For the first-row atoms, the local orbitals are hybrids of $2s$ and $2p$ canonical orbitals. The Zn $n = 3$ local orbitals are hybrids of $3s$, $3p$, and $3d$ canonical orbitals. The canonical orbitals contributing most to each FLO are listed in the table. The magnitude of the orbital corrections increases across the first row atoms as the orbitals become more compact

	$-\sum_i U_C[n_i]$		$-\sum_i U_{xc}[n_i]$	
Atom	rFLOSIC (eV)	cFLOSIC (eV)	rFLOSIC (eV)	cFLOSIC (eV)
Ne	-296.03	-288.49	267.44	258.44
Ar	-704.76	-683.55	634.10	609.07
Kr	-2046.10	-1962.77	1842.74	1748.02

Table 3: *The total self-Hartree and self-exchange-correlation contributions to the total SIC energy (in eV) for the closed-shell atoms Ne, Ar, and Kr. The cFLOSIC energies were obtained from calculations using the self-consistent rFLOSIC electron densities.*

with increasing atomic number. For example, for the Be $2s$ orbital, the correction is -0.17 eV in both theories. For the N $2s2p^3$ FLO, the correction is -0.60 and -0.71 eV for rFLOSIC and cFLOSIC, respectively. For the Ne $2s2p^3$ FLO the corrections are -1.02 and -1.20 eV. For the $n = 3$ Zn FLOs the corresponding corrections are -1.35 and -1.58 eV and for the $n = 4$ FLO, -0.10 and -0.11 eV. The difference between the cFLOSIC and rFLOSIC corrections depends on the nature of the FLO. For s-type FLOs, the corrections are equal in the two approaches. The difference increases with increasing p-character in the FLO. For example, for the $2s$ and $2sp$ FLOs of Be and B, the corrections are essentially equal for the cFLO and rFLO. For the C $2sp^2$ FLOs, the cFLO correction is -0.02 eV lower than for the rFLO. For the N $2s2p^3$ FLO, the cFLO correction is -0.11 eV lower than for the rFLO. Similarly, for the minority spin FLOs of F the $2sp^2$ cFLO correction is -0.04 eV lower than for the rFLO, while for the Ne $2s2p^3$, the cFLO is -0.18 eV lower than the rFLO. Finally, for the Zn $3s3p^33d^5$ FLOs, the cFLO corrections are -0.23 eV lower than the rFLO. Figure 2 shows iso-surface plots of the rFLO orbital densities, $n = \varphi^2$. (Only the density for first FLO listed for each atom is shown.) For the cFLOs, iso-surface plots of $\varphi_R^2 - \varphi_I^2$ may be found in Ref. [25]. These plots highlight an effective smoothing of the FLOs, referred to as lobedness by Perdew, that may be an important issue for higher-level functionals. The SIC corrections include self-Hartree and self-exchange-correlation components. The former are negative, while the latter are positive. Table 3 shows the total self-Hartree and self-exchange-correlation parts of the SIC corrections separately for the representative atoms Ne, Ar, and Kr. The self-Hartree contributions are more negative in rFLOSIC in every case. This implies that the rFLOs are more localized than the cFLOs. On the other hand, the self-exchange-correlation energies are less positive for cFLOSIC. The combined corrections are more negative in cFLOSIC than rFLOSIC, as seen in Table 1. This implies that the cFLOSIC – rFLOSIC difference between the self-exchange-correlation components must be more negative that the self-Hartree difference is positive. For Ne, for example, the rFLOSIC self-Hartree component is 7.54 eV more negative, while the cFLOSIC overall correction is 1.45 eV more negative (cf. Table 1). Thus, the magnitude of the self-exchange energy for Ne is 8.99 eV smaller in cFLOSIC than rFLOSIC. Allowing complex degrees of freedom does not result in more localized atomic orbitals, but instead decreases the magnitude of the self-exchange-correlation energy of the orbitals.

Molecule	HF	LSDA	rFLOSIC(g)	rFLOSIC	cFLOSIC	Δ E (eV)
Li_2	-14.8693	-14.7237	-15.0561	-15.0561	-15.0561	0.00
N_2	-108.9803	-108.6923	-109.8645	-109.8581	-109.9087	-1.38
C_2H_2	-76.8431	-76.6250	-77.6106	-77.6077	-77.6402	-0.88
CO	-112.7756	-112.4706	-113.6548	-113.6503	-113.6997	-1.34
HCN	-92.9016	-92.6541	-93.7304	-93.7244	-93.7662	-1.14
CH_4	-40.2103	-40.1187	-40.7021	-40.7003	-40.7014	-0.03
C_2H_6	-79.2537	-79.0720	-80.1912	-80.1878	-80.1900	-0.06
NH_3	-56.2173	-56.1067	-56.7742	-56.7729	-56.7779	-0.14
LiF	-106.9827	-106.7022	-107.7350	-107.7340	-107.7796	-1.24
HCl	-460.0940	-459.3330	-461.7452	-461.7451	-461.8655	-2.61

Table 4: *Total energies (in Ha) of molecules in Hartree-Fock (HF), LSDA, rFLOSIC(g), rFLOSIC, and cFLOSIC and the total energy difference ΔE = cFLOSIC – rFLOSIC (in eV). The cFLOSIC energies are evaluated using the self-consistent rFLOSIC electron density. rFLOSIC(g) corresponds to energies relaxed with FLOSIC while rFLOSIC corresponds to LSDA equilibrium geometries. rFLOSIC and cFLOSIC results, at LSDA geometries, are from Ref. [25].*

3.4 Molecules

For completeness HF, LSDA, rFLOSIC, and cFLOSIC total energies calculated for selected molecules, taken from Ref. [25], are shown in Table 4 in Ha units. The energy difference cFLOSIC – rFLOSIC is shown in the last column in eV. The rFLOSIC total energy is considerably lower than the LSDA energy for each molecule in the table and the cFLOSIC energy is lower than rFLOSIC in all cases except Li_2. The complex/real energy difference depends on the nature of the molecule. For molecules involving only C−H or N−H bonds the differences are less than 0.15 eV . The differences are much larger for molecules with multiple bonds (C=C, C≡N, C=O, N≡N, O=O). For example in N_2, which has a triple bond, the cFLOSIC energy is 1.4 eV lower than for rFLOSIC. For O=C=O, with two double bonds, the energy difference is 2.8 eV. The largest energy difference is obtained for HCl, where the cFLOSIC energy is 2.61 eV lower than the rFLOSIC energy. Cl is the heaviest atom appearing in our set of molecules and the cFLOSIC energy of the isolated Cl atom is 3.34 eV lower than for rFLOSIC.

Atomization energies were calculated as the difference in the total energies of the molecules (Table 4). The results (in eV) for LSDA, rFLOSIC, and cFLOSIC are shown in Table 5, along with reference experimental values from which zero-point energies have been removed in order to be directly comparable to the computed values. The cFLOSIC atomization energies are sometimes larger and sometimes smaller than in rFLOSIC, but mostly larger than the LSDA values. Exceptions where the cFLOSIC atomization energies are smaller than in LSDA are for F_2 and O_3. Compared to the reference atomization energies, the FLOSIC methods have significantly smaller mean errors (MEs) than LSDA. The ME for LSDA, rFLOSIC, and cFLOSIC are 1.92, 0.63, and 0.85. rFLOSIC and cFLOSIC also have smaller mean absolute errors (MAE) than LSDA. The MAE for cFLOSIC (1.21 eV) is somewhat worse than for rFLOSIC (1.08 eV).

Molecule	Ref.	LSDA (eV)	rFLOSIC (eV)	cFLOSIC (eV)	ΔE (eV)
Li_2	1.05	1.01	1.03	1.03	0.00
CH_3	13.27	14.69	14.62	14.65	0.03
CH_4	18.21	20.03	20.26	20.21	-0.05
C_2H_6	30.83	34.40	34.72	34.62	-0.10
NH_3	12.88	14.60	14.49	14.16	-0.39
LiF	6.03	6.80	5.70	6.18	0.48
F_2	1.65	3.42	1.70	0.92	-0.78
HCl	4.64	5.26	5.09	5.02	-0.07
C_2H_2	17.52	19.90	18.92	19.65	0.73
HCN	13.57	15.59	14.35	14.95	0.60
CO	11.32	12.95	11.14	11.80	0.66
CO_2	17.00	20.46	16.11	17.58	1.47
N_2	9.84	11.53	10.25	10.70	0.45
O_3	6.42	10.47	4.73	4.71	-0.02
ME		1.92	0.63	0.85	
MAE		1.92	1.08	1.21	

Table 5: *Atomization energy (in eV) of molecules in LSDA, rFLOSIC, and cFLOSIC, and difference ΔE = cFLOSIC – rFLOSIC. The reference atomization energies are zero-point energy corrected experimental values. Mean error (ME) and mean absolute error (MAE) for each method relative to the reference are also shown. The cFLOSIC values were obtained from calculations using the self-consistent rFLOSIC density. Calculations performed by K. Withanage.*

It is instructive to compare LiF and HCl which both have an outermost valence of 8 paired FODs that form a distorted tetrahedron. For LiF, small energy differences (0.0025 in the case of LiF) occur depending on whether the base or the vertex of the tetrahedron is found between the two atoms. For LiF there are three nearest FODs to the Li atom. For HCl the tetrahedron is inverted and there is only one FOD in close proximity to the hydrogen atom. This distinction is due to changes in the energy splitting between the occupied s and unoccupied p states on the column-1 element.

3.5 Returning to N_2: complex vs. real FLOs

Pictured in Fig. 3 are isosurface plots of the valence orbital densities for N_2. The figure includes the canonical orbitals (top), the rFLOs (left) and the cFLOs (right) from real and complex FLOSIC calculations, respectively. The figure displays the SIC energy $E^{SIC}[n_{i\sigma}]$ for each orbital, as well as the expansion of the rFLOs and cFLOs in terms of the canonical orbitals. (Note that the definition of the canonical orbitals varies slightly between the rFLOSIC and cFLOSIC calculations. Primes are used to indicate this difference.) The centroids of the rFLOs and cFLOs are also shown in the figure. For the rFLOs, the centroids are close to the FOD positions. The canonical $2\sigma_u$, $1\pi_{ux}, 1\pi_{uy}$ and $3\sigma_g$ orbitals shown in Fig. 3 have positive SIC energy. The SIC energies are all negative for the rFLOs and cFLOs. The centroids of the rFLOs associated with the triple bond are positioned at the vertices of an equilateral triangle in a plane perpendicular to the bond axis and passing through its midpoint. This is indicated in the gray inset on the left

side of Fig. 3. For the rFLOSIC case, the centroids are close to the optimized FOD positions. The topology of the triple-bond rFLOs is similar to that of the localized orbitals discussed by Ruedenberg and Edmiston and also by Klüpfel *et al*. The optimal cFLOs on the right of the figure appear very similar to the complex local orbitals shown by Klüpfel *et al*. The FLOs corresponding to the lone pairs in N_2 are similar in the rFLOSIC and cFLOSIC calculations, although the cFLOs are slightly more localized.

The SIC energies shown in the figure indicate that the cFLOSIC total energy is lower than the rFLOSIC total energy due to the SIC energies of the triple-bond orbitals. Interestingly, one cFLO has a less negative SIC energy than any of the three equivalent rFLOs, but this is compensated by a much larger difference for the other two orbitals that have a more negative correction than the rFLOs. The cFLO optimization thus accepts an energy penalty for one orbital in order to realize a larger energy reduction for the other two.

It is interesting to examine the unitary transformation connecting the rFLOs and cFLOs to the canonical orbitals. This is shown in Fig. 3. The entries indicate that the three-triple bond rFLOs are symmetry-related mixtures of the $2\sigma_g$ and π_x and π_y canonical orbitals, essentially equivalent to bonding combinations of sp^2 hybrids. For the cFLOs, one triple bond orbital is a complex combination of the π_x and π_y orbitals yielding a cylindrical density. The other two are σ-π hybrids resulting from a complex π orbital and a real $2\sigma_g$ orbital. To the best of our knowledge, these simple expressions detailing the cFLOs have not been published previously. The two lone pair orbitals have complex coefficients that essentially correspond to a real number times the same complex phase factor for all three. Thus, these orbitals can effectively be taken as real, showing that the cFLOSIC optimization may result in real orbitals, when these minimize the energy. The centroids of the cFLOs are also shown in the panel on the right of Fig. 3. For the triple bond orbitals, these fall on a line passing through the bond center perpendicular to the axis. It is worth noting that the centroid of the cylindrical orbital is at the bond center where the π orbitals have zero amplitude. The Fermi orbital corresponding to a real FOD placed at the center of the bond would therefore contain zero contribution from the π orbitals. This is a reminder that orbital centroids do not always coincide with FOD positions.

To summarize, a scheme to introduce complex local orbitals (cFLOs) into the Fermi-Löwdin orbital self-interaction correction (FLOSIC) method has been derived and tested.. The scheme rests on allowing the Fermi orbital descriptors (FODs) to be complex. FOD optimization is accomplished via gradient optimization as in the case of FLOSIC restricted to real orbitals (rFLOSIC) and requires only the additional calculation of energy gradients with respect to the imaginary part of the FODs. We demonstrated the complex FLOSIC (cFLOSIC) method through applications to an array of atoms and molecules. The results of these applications are similar to those obtained in complex PZSIC calculations. The cFLOSIC solutions are generally lower in energy than in rFLOSIC. We showed that the optimal cFLOs are less localized than the analogous rFLOs, as judged by having a less negative self-Hartree energy. The lower cFLOSIC total energies thus arise from reducing the magnitude of the self-exchange-correlation energy of the cFLOs relative to the rFLOs. Analyzed in terms of individual orbital corrections, we find that the cFLOs lower the energy more for strongly hybridized orbitals and, in molecules,

for double and triple bond orbitals. The cFLOSIC calculations reported here make use of the self-consistent occupied orbitals from corresponding rFLOSIC calculations. A next step will be to make the cFLOSIC calculations fully self-consistent. Self-consistent cFLOSIC calculations will result in somewhat lower total energies that those presented in this section, but we do not expect significant changes to any of our conclusions. It is also of interest to perform cFLOSIC calculations with more sophisticated functionals such as PBE and SCAN. Finally, the results presented in this section indicate that SIC calculations on molecules with multiple bonds or transition metal atoms may be particularly affected by the use of complex orbitals.

4 Downward quantum learning: Tricks for finding starting configurations

We now discuss a search method which we refer to as downward quantum learning. As discussed above, finding initial FODs that lead to a positive definite Fermi orbital overlap matrix, a necessity for obtaining FLOs, is difficult especially for f-electron systems and open-shell systems regardless of whether they are isolated or in molecules or solids. For systems where charge-transfer exists between a cation and an anion, the Kohn-Sham orbitals obtained from a starting calculation are generally inadequate for starting a calculation. The systems just mentioned are of course the systems for which self-interaction corrections are most needed. In an earlier section an existence proof was provided. Let us now think about the manifestations of this proof.

Existence of FLOs for the free-electron metallic state

It is generally not known how to find FODs and FLOs for metals. It is however known that Wannier functions for a metallic state do not exist and this is why Wannier functions are in fact a subset of FLOs. While there has not yet been a demonstration for an exact set of FLOs for the metallic Brillouin zone, or Fermi-sphere, of the free-electron gas, we argue here that one exists. Wannier's demonstration in 1937 showed Wannier functions exist for a cubic (insulating) Brillouin zone and this chapter as well as earlier works by the author showed here these functions coincide with FLOs. Therefore, there is now hope, in fact a guarantee, for a semi-analytic solution of this problem. Once can start by inscribing a Fermi Sphere inside a Wannier Cube and then successively remove one FOD and one FLO, from outside the sphere, and inside the cube until only the Fermi Sphere remains.

Existence of FLOs for all Atoms

With respect to atoms it is generally quite easy to find starting positions for any rare-earth atom and more generally for any atom/ion that has closed-shell spin states and a qualitatively correct shell filling. The theorem proved above stated that for each of these "easy" solutions, one can determine initial FODs for lighter atoms by removing one orbital and one FOD because at least one $N-1$ dimensional cofactor has to be non-zero if the N-particle determinant is non zero.

This means that by mining the quantum information contained within the resulting FLOs for the heaviest rare-earth atom (Z=118, Oganesson) it is possible to obtain starting configurations for every atom in the problem. In the following we demonstrate this capability. A recent demonstration of principle is repeated here.

The existence proof reiterated above shows that viable FOD positions and complementary Kohn-Sham orbitals do indeed exist if the many-electron wavefunction does not vanish everywhere. This solution is especially useful for open-shell atoms and ions from the d- and f-blocks of the periodic table.

Nomenclature for FLO issues

To improve the nomenclature used for discussing the FLOSIC results and formulation, we outline five technical problems that arise. Three of these inconveniences are unique to FLOSIC while the others are also present in standard DFT calculations. However, these issues have both positive and negative attributes. Their presence often hampers a calculation technically when the user is only interested in the ground state. However they also provide physical insights for cases where energy- or electron-transfer is of interest. We refer to these issues as Fermi-Orbital-Challenges (FOCs). First, the ability to start calculations by finding an initial set of FODs and a Kohn-Sham density matrix that leads to a positive definite FO-overlap matrix must be fully systematized (FOC1). FOC1 arises for all cases where density-functional algorithms give qualitatively incorrect shell fillings – a problem that is prevalent in atoms containing $3d$/$4s$ states and $4f$/$5d$ states. FOC2 is exemplified by systems that have multiple low-lying competing electronic configurations which, depending on whether one is thinking in terms of DFT or WF, are either Janak-like multi-reference systems respectively. For such systems there are generally non-integer occupation numbers associated with the Kohn-Sham orbitals that oscillate from iteration to iteration. For standard DFAs, this leads to poor convergence in systems as simple as the Nickel atom or Carbon dimer and results that are not easy to interpret. Within FLOSIC, rather than finding a single fractionally occupied solution, multiple low-energy stable solutions with integer occupancy are determined. This feature readily provides the correct pictures for problems such as charge transfer. When multiple solutions exist there are generally incompatibilities between the FODs for one solution, and Kohn-Sham orbitals for other solutions. These incompatibilities complicate the determination of starting solutions. While FOC2 is a significant frustration to users, it also identifies systems that may have interesting low-energy excitations which is generally relevant in energy applications. FOC3 presents the user with multiple stable states that are sometimes incorrectly associated with incorrect ground states but most likely contain information about collective excitations such as plasmon oscillation. Other *aufbau*-violating solutions associated with FOC3 may provide information about x-ray excitations. A spin-conserving example of FOC3 is charge transfer excitations between a halide and alkali in the stretched bond limit (NaCl $\rightarrow$ $Na^{+1}Cl^{-1}$) or the charge transfer excitations in light-harvesting systems. The latter are *aufbau* violating solutions at the DFT-level but may be *aufbau* consistent solutions within FLOSIC. FOC3 is always a clear indicator of mul-

tiple low-energy excitations or multi-reference character. It is a frustration when the primary interest is on the ground state but is advantageous when one is interested in studying molecular processes especially those related to luminescence and fluorescence in, for example, rare earths.

FOC4 and FOC5 are related to symmetry breaking in the density-functional formulation. It is generally accepted that the description of stretched bonds or dissociated atoms, within density-functional pictures, requires spin symmetry breaking with excess spin-up density on one atom and excess spin-down density congregating on an otherwise symmetrically equivalent atom. Such antiferromagnetic pictures lead to total densities that do not break the overall symmetry of the molecular system. Analogous symmetry breaking occurs within systems that are treated with FLOSIC (FOC4). Further when a partial open-shell structure occurs (ozone is a deceptively simple case), FLOSIC predicts spin separation but standard functionals do not. The conundrums presented by FOC4 and FOC5 are similar but one is driven by the self-interaction correction rather than the energy functional it self. As such we label FOC4 and FOC5 as separate challenges. Issues arising from FOC3–FOC5 raise the spectre for better descriptions of low-energy spin-conserving (FOC3) and spin-changing (FOC4–FOC5) excited states. The count-down paradigm discussed below is based on the rigorous existence proof presented above and provides a new tool for constructing self-consistent *aufbau*-violating states for systems where occupation number constraints might fail. The procedure does not circumvent FOC2 issues. Such issues will continue to require improvements on the iterative process and/or additional attention to occupation-number dependent FLOSIC formulations.

Generally speaking, closed shell atoms are the simplest cases for finding viable starting FODs. Yet even in those cases locating useful FOD starting points can be difficult. To give an indication of the challenge, we share our experiences for finding FODs in Radon (Z=86), We used a Monte Carlo approach to generate many sets of FOD positions for individual shells corresponding to the various principal quantum numbers n of the occupied orbitals. Each shell had a radius equal to the average radial expectation value of the corresponding orbitals. A total of n^2 positions were randomly placed on each shell (a spin unpolarized atom was assumed). This approach resulted in viable FODs in fewer than 8 percent of the trials. For other heavy atoms, especially those between La and Hg, the success rate was even lower. Using a solution for Rn, removing FODs, and starting from default atomic density-functional potentials, allowed successful calculations for a small number of atoms (Eu, Yb, Au–Rn). This success, while limited, highlighted the need for the solution discussed below.

Here we present a well-defined way to determine initial FOD positions to start a calculation. The method is based on a proof given below that a set of viable FOD positions has the property that a Slater determinant of the KS orbitals is non-zero when evaluated at the FOD positions. The ability to extract FLOs and FODs from clearly closed-shell systems and then systematically investigate on-atom Kossel-like solutions in the actinides and lanthanides is an additional need that the method described here addresses. For the calculations reported here, we used a legacy version of the NRLMOL software package that utilizes optimized cartesian gaussian basis sets and a highly accurate numerical integration scheme. The version employed was developed approximately 20 years ago to assess the possibility that scalar relativity could significantly

impact second-order anisotropy energies in $4d$ molecular magnets. The possibility of using f-electrons for post-processing of anisotropy Hamiltonians was incorporated at that time and basis sets for nuclear charges between Ba–Rn were generated using the techniques. Shortly after the MMQT group formed at UTEP in 2019, methods for performing self-consistent-field calculations with f-electrons were completed. These methods used integral transforms, albeit for gaussians. The Fermi-Löwdin methodology was simultaneously inserted into the legacy f-electron code. However in contrast to the earlier versions of the FLOSIC code, the formulation was implemented in a manner that preserved the group-theoretical methods that are part of the legacy NRLMOL codes. We viewed this as a necessary addition from the standpoint of efficiently finding FODs for heavy open-shell atoms. Group-theoretical techniques simplify the optimization of the basis sets for elements between Rn and Og (Z=86–118).

4.1 Initializing FODs for principal quantum numbers with *n*=2, 3, and 4

In the motivation we asked if there are lattices that are reciprocal to an arbitrary set of Kohn-Sham orbitals. For finite systems, we refer to these lattices, or polyhedra, as flotonic solids. We have used the T_d group operations to search for a set of equi-radial FODs with T_d symmetry that lead to a set of 16 points on the unit sphere with a large determinant. T_d symmetry allows for equivalent shells of 1, 4, 6, 12, and 24 points, respectively. To find a set of 16 points we have followed the prescription of placing 12 points on the unit sphere and found that, in addition to the dodecahedron there are many sets of 12-site shells that are compatible with T_d symmetry for which the 12×12 matrix constructed from l=2 and l=3 spherical harmonics leads to non-singular matrices. We listed approximately 20 of these solutions and then combined them with the 4-site T_d shells to create a family of 16-site FODs. We then re-ordered the family of 16-site solutions according to the determinant of the 16×16 $\underline{\mathrm{W}}$-matrix and found that there are two solutions with anomalously large determinants. The solution with second largest determinant turned out to be the best solution for the radon atom, strengthening the hypothesis that one should use spherical harmonics to find universal sets of FODs. It appears that a key figure of merit of a point group is that it must not contain the inversion operator since the presence of inversion prevents mixing of states with opposite parity. The octahedral point group, O_h, has two subgroups of order 16 which might provide additional solutions for the n=4 shells, but so far we have not found such solutions. As known from the periodic-table and represented in Table 6, noble-gas atoms correspond to atoms that contain subsets of electronic states that share a principal quantum number and have a similar atomic radius. For each principal quantum number n, there are n^2 hydrogenic-like orbitals with $n^2 = \sum_{l=0}^{n-1}(2l+1)$.

We then can consider the lowest n^2 spherical harmonics on a unit sphere. Depending on the value of n this yields a sequence of matrices of dimension $n^2 = \sum_{l=0}^{n-1}(2l+1)$, or 1, 4, 9, and 16 for n=1, 2, 3, and 4, respectively. For each of these cases a $\underline{\mathrm{W}}$-matrix can be constructed

according to

$$\underline{\mathrm{W}} = \begin{bmatrix} Y_{00}(\hat{a}_1) & Y_{1,-1}(\hat{a}_1) & Y_{1,0}(\hat{a}_1) & Y_{1,1}(\hat{a}_1) & \cdots & Y_{n-1,-n+1}(\hat{a}_1) & \cdots & Y_{n-1,n-1}(\hat{a}_1) \\ Y_{00}(\hat{a}_2) & Y_{1,-1}(\hat{a}_2) & Y_{1,0}(\hat{a}_2) & Y_{1,1}(\hat{a}_2) & \cdots & Y_{n-1,-n+1}(\hat{a}_2) & \cdots & Y_{n-1,n-1}(\hat{a}_2) \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ Y_{00}(\hat{a}_{n^2}) & Y_{1,-1}(\hat{a}_{n^2}) & Y_{1,0}(\hat{a}_{n^2}) & Y_{1,1}(\hat{a}_1) & \cdots & Y_{n-1,-n+1}(\hat{a}_{n^2}) & \cdots & Y_{n-1,n-1}(\hat{a}_{n^2}) \end{bmatrix}_{n^2 \times n^2} \quad (48)$$

The goal then is to pick sets of n^2 points on a unit sphere that, at the very minimum, guarantee that no spherical harmonic with angular momentum less than or equal to $n-1$ vanishes at all points but with the stricter and more difficult condition that the above matrix is non-singular. A systematic and fast way to find sets of points is to generate all sets of points on a cubic grid (n_x, n_y, n_z) with $|n_x| < 7$ (higher numbers do not change the outcome), and then normalize them onto a unit sphere. One can then search over all point groups to find high-symmetry arrangements of 4, 9, or 16 points. These can then be used to evaluate the determinant of $\underline{\mathrm{W}}$ in Eq. (48). The sets corresponding to the largest determinants then suggest the best sets of FODs for electrons sharing the same principal quantum number. The effect of these sets on the determinant of the FOD overlap matrix is shown in Table 7. The best 9-electron set located so far, consisting of staggered triangles, gives sp^3d^5 hybrids and probably limits the symmetry of the overall atomic spin density to C_{3v}. The optimal shells obtained for element 118 (Og) are presented and we show how these shells can be used to generate starting points for all other noble-gas atoms. The tetrahedral (T_d) shells, which describe sp^3 (n=2, n^2=4) systems have been well understood for quite some time. However, the sp^3d^5 (n=3, n^2=9) and $sp^3d^5f^7$ (n=4, n^2=16) shells are significantly more difficult to determine as random guesses and are deserving of additional discussion.

As a representative example, we describe the series of calculations beginning with the self-consistent closed-shell calculation for Ar (Z=18) that features filled $3s$ and $3p$ sub-shells for the outer electrons, in addition to completely filled n=1 and n=2 shells. For a shell with 4 electrons, FODs can be arranged at the vertices of a tetrahedron. The resulting FLOs are sp^3 hybrid orbitals. We adopt radii of 1.73 and 0.45 Bohr for the n=2 and n=3 tetrahedrons, respectively. We then computed the total energy self-consistently with the LSDA density functional, obtaining a PZSIC total energy of -529.9441 Ha, and an orbital energy for the highest occupied molecular orbital (HOMO) of -0.6141 Ha, corresponding to a predicted removal energy (the negative of the HOMO energy) of 16.7 eV. The corresponding value in an uncorrected LSDA calculation is 10.4 eV. The experimental ionization potential (IP) for Ar is 15.8 eV. These results reflect the well-known effect of self-interaction in atoms that the HOMO eigenvalues are smaller than ionization energies. Calculated LSDA and FLOSIC removal energies and corresponding experimental IPs are shown in Table 8.

From the self-consistent Ar calculation, one of the outermost FODs and one of the occupied Kohn-Sham orbitals are removed in order to create starting FLOs for a FLOSIC calculation for Cl. Similarly, two outer FODs and two KS orbitals are removed (of the same spin) to create a starting point for a FLOSIC calculation for S. This process was repeated for all the atoms of the second row. Predicted removal energies obtained from the calculations can be found in

Config.	$\vec{A}$	Sym.	N_{equiv}	R_{Og}	R_{Rn}	R_{Xe}	R_{Kr}	R_{Ar}	R_{Ne}	R_{He}
s	$(0,0,0)$	N/A	1	0.0000	0.000	0.00	0.00	0.00	0.00	0.00
sp	$\vec{A}_1$	T_d	4	0.0356	0.074	0.12	0.19	0.44		
spd	$\vec{A}_2^1$	T_d	6	0.1012	0.18	0.33				
spd	$\vec{A}_2^2$	C_{3V}	3	0.1012	0.18	0.33				
$spdf(A)$	$\vec{A}_3^1$	T_d	4	0.2808	0.43					
$spdf(B)$	$\vec{A}_3^2$	T_d	12	0.2808	0.43					
$spdf$	$-\vec{A}_3^1$	T_d	12	0.6138						
$spdf$	$-\vec{A}_3^2$	T_d	4	0.6138						
$spd(C)$	$-\vec{A}_2^1$	T_d	6	0.9045	1.00	0.85	0.60			
$spd(D)$	$-\vec{A}_2^2$	C_{3V}	3	0.9045	1.00	0.85	0.60			
sp	$-\vec{A}_1$	T_d	4	2.4656	2.47	2.28	1.92	1.68	1.07	

Table 6: *Reasonable starting FODs and radii for noble gas atoms. Symmetrized sets of unit-vectors that maximize the $n^2 \times n^2$ determinant of spherical harmonics on a unit sphere are given by: $\vec{A}_1 = (1,1,1)/\sqrt{3}$, $\vec{A}_2^2 = (1,0,0)$ and $\vec{A}_2^2 = (-1,-1,2)/\sqrt{6}$, $\vec{A}_3^1 = (1,1,1)/\sqrt{3}$ and $\vec{A}_3^2 = (-0.8789,-0,3373,-0.3373)$. Based on Linnett-like structures for inverting FOD-positions for opposite spins, the highest symmetry for the spin-densities in Ne and Ar is T_d while the symmetry of the density can be O_h. Similarly the highest symmetry for the spin-density for Kr–Og is C_{3v} while the symmetry for the total density can be D_{3d}. A variety of lower point group symmetries that would be compatible with incomplete shells of angular momenta are possible in other ions. The flotonic solids that result from these vectors and the resulting n^2 hybrids (for n=3 and 4) are illustrated in Fig. 2. While the shells for each principal quantum number close to resolved, the relative orientation of each shell is not resolved at the time of this writing. Standard optimization methods are not good at addressing such questions and additional automation is required for perfecting that part of the FOD optimization.*

Table 8, where it can be seen that the FLOSIC predictions are much closer to the experimental values. Similar downward-quantum-learning calculations have been performed starting from Og (Z=118), Rn (Z=86) and Ne (Z=10) [27].

Here, we have provided an existence proof that connects viable FOD positions to a non-zero value for the many-electron wave function constructed from the Kohn-Sham orbitals for the system and evaluated at the FOD positions. The proof relies only on the properties of determinants and it guarantees that if one finds viable FODs for a system of N electrons, it is always possible to find viable FODs for electronic systems with fewer electrons (in each spin channel), using a countdown algorithm that selectively removes FODs and orbitals from the N-electron solution. For any set of orbitals unitarily equivalent to the KS orbitals, the algorithm identifies at least N solutions of lighter systems constructed by successive removal of orbital-FOD pairs. It is very possible that successive removal of FLO-FOD pairs will generate a total of $\sum_{n=1}^{N-1} \binom{N}{n} = 2^N - 2$ initial solutions for atoms containing between 1 and $N-1$ electrons of a specific spin. We demonstrated that the algorithm can successfully generate FODs for lighter

Z	e $^{1/2}$	Z	e $^{1/2}$	Z	e $^{1/2}$	Z	e $^{1/2}$	Z	$D^{1/2}$	Z	$D^{1/2}$
59	0.0003	58	0.0003	57	0.0004	56	0.0004	55	0.0004	54	0.0008
53	0.0016	52	0.0033	51	0.0044	50	0.0059	49	0.0077	48	0.0101
47	0.0133	46	0.0171	45	0.0190	44	0.0210	43	0.0225	42	0.0241
41	0.0265	40	0.0284	39	0.0294	38	0.0333	37	0.0363	36	0.0377
35	0.0391	34	0.0403	33	0.0416	32	0.0426	31	0.0438	30	0.0453
29	0.0523	28	0.0600	27	0.0667	26	0.0740	25	0.0846	24	0.0967
23	0.1065	22	0.1298	21	0.1518	20	0.1606	19	0.1753	18	0.1943
17	0.2109	16	0.2289	15	0.2433	14	0.2638	13	0.3861	12	0.5644
11	0.8240	10	0.8683	9	0.9111	8	0.9456	7	0.9702	6	0.9870
5	0.9901	4	0.9943	3	0.9974	2	0.9993	1	1.0000		

Table 7: *Determinant ($^{1/2}$) of the Fermi-orbital overlap matrix for electronic configurations of Og^{+Q}. The fact that the determinant for every electronic configuration converges monotonically from the determinant of the neutral Og atom to 1 guarantees that it is possible to generate relatively physical starting points for any charge and spin state of any atom.*

atoms from a solution for the next-largest noble gas atom. This has been accomplished for all atoms below Ne and Ar, and for select atoms beginning from solutions for Rn and Og (Z=86) and Og (Z=118). The success of the algorithm across the periodic table demonstrates its utility and numerically confirms the theorem repeated above and first demonstrated in Ref. [27]. For atoms in d- and f-blocks of the periodic table, additional integration of existing and additional techniques are progressing. These techniques include: (1) single-shot Z-dependent scaling of the starting orbitals using a combination of the data in Table 1 and 6, and shell-by-shell virial-like scaling, (2) optimization of FODs at the frozen density, (3) facile but more sophisticated potential biasing of starting potentials that are discussed in Ref. [22] and [25], (4) new machine-learning strategies that aid predicting which of the $2N$ solutions identified from the countdown method are most likely to succeed, and (5) capitalizing on the ligand-induced changes in atom coordination that ultimately define the allowable $3d$ and $4f$ valence configurations in atoms. We expect that complete success in generating self-consistent solutions using the countdown algorithm will require scaling the starting orbitals obtained for the heavier atoms. Because of the larger Z, the wave functions of the closed-shell atoms are too compact for the lighter atoms. This can cause problems during the self-consistent iteration process that cause the calculation for the lighter atom to fail. An alternative approach will be to reduce the value of the nuclear charge of the lighter in several steps from that of the closed-shell atoms to its correct value, while generating self-consistent wave functions at each steps. The algorithm can be used to find a variety of solutions for lighter atoms, corresponding to various occupations of the orbitals. This means that the method could be used to generate a database of starting FODs for atoms throughout the periodic table, in various charge states and oxidation states, beginning from a solution for element Og, Z=118. With such a database automated starting points of viable FOD positions could also be created for molecules and other condensed systems.

The method proposed here succeeded by considering sets of 4, 9, or 16 points created by symmetry equivalent normalized vectors and was then used for the common special cases for par-

Atom	LDA	FLOSIC	EXPT.	Atom	LDA	FLOSIC	EXPT.
Ar	10.38	16.71	15.76	Ne	13.54	24.35	21.56
Cl	8.30	14.16	12.97	F	10.34	20.20	17.42
S	6.24	11.91	10.36	O	7.40	16.39	13.62
P	6.34	10.95	10.49	N	8.39	15.76	14.53
Si	4.65	8.89	8.15	C	6.11	12.46	11.26
Al	3.05	6.98	5.98	B	4.10	9.36	8.30
Mg	4.78	8.49	7.65	Be	5.59	10.84	9.32
Na	3.11	6.46	5.14	Li	3.17	6.92	5.39
Rn	7.98	11.90	10.75	Og	7.44	11.20	8.9
At	6.68	10.52	9.22	Ts	5.99	9.89	7.70
Po	5.35	9.19	8.42	Lv	4.95	8.67	8.64
Bi	5.40	8.70	7.29	Mc	5.11	8.42	5.68
Pb	4.18	7.35	7.42	Fl	3.96	7.25	8.53
Tl	2.96	6.10	6.11	Nh	2.83	6.07	7.31

Table 8: *Calculated (-1)HOMO energies from DFT-LDA and FLOSIC, and experimental ionization energies of atoms (in eV). Ionization energies for the superheavies may be found on the web. Some of the early experimental results are inconsistent with our FLOSIC results and trends expected from the Rn row.*

tially filled electronic shells. It points further to an improved method for rapid characterization of other solutions. If we loop over point groups, one can then create sets of equivalent points that are compatible with that point group. Given a set of Q points one can ask whether there are exactly Q combinations of the first 16 spherical harmonics that are linearly dependent on the unit sphere. When this condition holds FOC1 has been bypassed for a specific partially filled shell of angular momenta sharing the same principal quantum number. A further improvement for this case would be to hypothesize that the local coordinate system should be oriented such that some of the Q points coincide exactly with the zeros of the missing spherical harmonics. This point is expected to be particularly useful for molecular magnets for which local Jahn-Teller distortions, defined by the ligand structures, mandate shell fillings with holes in the frontier d- or f-shells.

4.2 Challenge: Simulating tetra-anionic Mn_{12}-Acetate in water

Learning how to control solar-induced splitting of water into oxygen and hydrogen would have immense value from the standpoint of the world energy and climate concerns. For computational materials scientists and physicists to help with this problem there are many multi-scale problems that need to be solved and quantum-mechanical methods for understanding these problems will require scientists to accurately simulate highly charged molecular systems in aqueous environments. Typical problems that occur when simulating a single ion near water is that fractionally occupied states occur at the Fermi level with the excess electron spread over the solvent (water) and the solute (anion). In Table 9 we show energy differences, calculated within PBE-GGA, for a chemical system containing four excess electrons. In going from S_0

State	Molecular Configuration	Energy (eV/H2O)
S_0	$Mn_{12}O_{12}(COOH)_{16}(H_2O)_4$ + $4H_2O$ + 4e-	0.00
S_1	$Mn_{12}O_{12}(COOH)_{16}(OH^-)_4$ + $4H_2O$ +$2H_2$	0.95
S_2	$Mn_{12}O_{12}(COOH)_{16}(H_2O)_4$ + $4(OH^-)$ +$2H_2$	1.34
S_3	$Mn_{12}O_{12}(COOH)_{16}(H_2O)_4^{4-}$ + 4(OH) +$2H_2$	2.62
S_4	$Mn_{12}O_{12}(COOH)_{16}(H_2O)_4$ + $4H_2O$ + 4e-	0.00

Table 9: *A cyclic catalytic water-splitting reaction sequence. The energy scale for various tetra-anion configurations are somewhat consistent with the experimental observation that four electrons are needed to split water. Full scale simulation on this type of problem requires corrections to the LUMO levels of solvated anions in water. Recent work suggests that FLOSIC will correct for such issues. [23]*

to S_1, terminating water molecules expel neutral hydrogens which then form molecular hydrogen leaving behind the isoelectronic hydroxyl anions in place of the waters. In going from S_1 to S_2, waters of solvation replace the hydroxyl anions which returns the molecule to its initial state. In the following two steps the electrons are transferred back to the molecule and the neutral hydroxyl radicals could convert into additional molecular oxygen and hydrogen. This is a straw-man hydrogen production cycle which may or many not hold water when put to rigorous computational testing. But the problem with computationally testing this hypothesis is, due to the self-interaction error, the HOMO level of the tetra-anion is predicted to be 6 eV above the LUMO of the surrounding molecule. FLOSIC calculations performed on the fragments suggest that inclusion of SIC for the entire system would place the tetra-anionic HOMO level very close to the LUMO level of the surrounding water molecules. Such conditions would be ideal for solar-induced hydrogen production. Problems such as this and other problems associated with highly charged anions in solution are one of many fertile areas for exploration with FLOSIC over the coming years. Additional discussion of this problem may be found in Ref. [23].

5 Summary and outlook

Including self-interaction corrections to density functional approximations (SIC-DFA) has been a long-standing challenge especially from the standpoint of maintaining the inherent efficiency of DFA methods in applications to molecular systems and devices. Early applications of SIC-DFA, based upon solutions of the localization equations [9, 11], succeeded in addressing the lack of unitary invariance in SIC-DFA and introduced a Koopmans' theorem [11]. We have reviewed a new implementation of the self-interaction-correction [19], now referred to as Fermi-Löwdin-Orbital Self-Interaction Correction (FLOSIC), that restores much of the formal structure expected from a DFA. Use of the density matrix, constructed from Kohn-Sham orbitals, and a physically appealing classical electron geometry, determines a density-dependent $N\times N$ unitary transformation that connects the occupied Kohn-Sham orbitals to the ideal localized orbitals for evaluation of the self-interaction correction. In the small N limit the FLOSIC method is at least N times slower than DFA. However, at the time of this writing, recent progress that

capitalizes on the intrinsic sparsity of the problem significantly reduces the cost. A downward-counting algorithm and existence proof have demonstrated applicability to all atoms in the periodic table [27]. This paper also provided an interesting connection between the FLOs and the amplitude of a Slater determinant composed of Kohn-Sham orbitals and raised the possibility of stronger connections between density-functional and wave-function pictures. Together with a sparse implementation of FLOSIC, there is now the possibility that the cost of a FLOSIC calculation, relative to LDA/GGA/SCAN, will have the exact same scaling as DFT and be within a factor of 10 of the cost. For example a soon to be reported application to a tri-anion-water system, $Cr^{III}(C_2O_4)_3$:$(H_2O)_{117}$, reduced the overhead from a factor of 1300 to, at most, 30.
The discussion here on the Fermi-Löwdin formulation of the self-interaction correction has attempted to provide the reader with the knowledge needed to embark upon their own original investigations. While the author opines that self-interaction corrections might decrease the need for spin-density-gradients in functionals, this has definitely not been proven here nor elsewhere. Complete analysis still requires a significant focus on implementation and efficiency but there are good reasons to expect that, as the FLOSIC community grows, new algorithms for implementation will be invented and the number of applications amenable to inquiry within the FLOSIC formalism will grow. The use of the FLO formalism is not limited to SIC and one can imagine applying the formulation to other quantum theories where unitary invariance is lacking. What is clear at this time is that basis set quality is seldom the accuracy limiting step in electronic structure calculations and it is probably still not time for the community to seriously invest their time in considering uncertainties due to basis sets. I have also provided my perspective on the status of the FLOSIC formulation and have tried to avoid encumbering the reader with too much cross comparison to other methods. The reader interested in works by others is encouraged to perform literature searches for the researchers mentioned in the acknowledgements. Finally, readers that are inspired to apply FLOSIC to problems that they are interested in, are encouraged to visit `https://www.flosic.org` to download the latest version of the publicly available FLOSIC code. Within that distribution, or via email to this author, it will also be possible to obtain a portable version of the legacy code which has some learning modules attached to it.

Acknowledgements

In finishing this book chapter and with confidence that the best days for the Fermi-Löwdin orbital formalism are still in front of us, I am thankful to competitors and collaborators for sharing ideas, questions, programming skills, and their individual contributions. I thank Dr. Eva Pavarini and Dr. Erik Koch for the opportunity to provide a perspective on this area of density-functional theory that has captured my curiosity for decades. Thanks to Dr. Alex Zunger and Dr. John Perdew for formulating SIC just as I was leaving college and entering graduate school and for many interesting discussions along the way. Thanks to Dr. Chun C. Lin for assigning an interesting dissertation topic and Dr. Richard A Heaton for collaboration during graduate school and for first suggesting, in August 1986, that Luken's early insights might provide a solution to the self-interaction problem. Thanks to Adrienn Ruszinzsky and John Perdew for collab-

oration on the introduction of FLOSIC. To Dr. Tunna Baruah, Dr. Jens Kortus, Dr. Der-You Kao, Dr. Javaria Batool, Dr. Torsten Hahn, and Dr. Koblar A. Jackson, I express my profoundest gratitude for their continuing interest in what are now called Fermi-Löwdin orbitals and for enthusiastic early efforts. There are numerous researchers from the Naval Research Laboratory, Max-Planck-Institut für Festkörperforschung, Daresbury National Laboratory, Linnaeus University, Freiberg Institute of Technology, University of Iceland and the University of Texas at El Paso that I have had stimulating discussions with on this problem. I note important contributions – both technical, computational, and scientific – from Dr. D.V. Porezag, Dr. J. Peralta, Dr. Yoh Yamamoto, Dr. Zahra Hooshmand, Dr. Kushantha Withanage, and Dr. Rajendre Zope. And thank you for reading this. I hope you will find time to read many interesting and important papers by these authors. The later part of this work (post 2020) was supported by the CCS FLOSIC project under the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under award number DE-SC001833.

References

[1] L. Pauling, J. Am. Chem. Soc. **53** 1367 (1931)

[2] T. Koopmans, Physica **1** 104 (1934)

[3] G.H. Wannier, Phys. Rev. **52**, 191 (1937)

[4] P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950)

[5] C. Edmiston and K.Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963)

[6] I. Lindgren, Int. J. Quantum Chem. Symp. **5**, 411 (1971)

[7] J.F. Janak, Phys. Rev. B **18**, 7165 (1978)

[8] J.P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981)

[9] M.R. Pederson, R.A. Heaton and C.C. Lin, J. Chem. Phys. **80**, 1972 (1984)

[10] W.L. Luken and J.C. Culberson, Theo. Chim. Acta **66** 279 (1984)

[11] M.R. Pederson, R.A. Heaton, and C.C. Lin, J. Chem. Phys **8**2, 2688 (1985)

[12] M.R. Pederson and C.C. Lin, Phys. Rev. B **35**, 2273 (1987)

[13] M.R. Pederson and C.C. Lin, J. Chem. Phys. **88** 1807-1817 (1988)

[14] A. Svane and O. Gunnarsson Phys. Rev. Lett. **65**, 1148 (1990)

[15] O.A. Vydrov and G.E. Scuseria, J. Chem. Phys. **121**, 8187 (2004)

[16] A. Ruzsinszky, J.P. Perdew, G.I. Csonka, O.A. Vydrov, and G.E. Scuseria, J. Chem. Phys. **125**, 194112 (2006)

[17] Z. Szotek, W.M. Temmerman, D. Ködderitzsch, A. Svane, L. Petit, and H. Winter, Phys. Rev. B **74**, 174431 (2006)

[18] S. S. Klüpfel, P. Klüpfel, and H. Jónsson, Phys. Rev. A **84**, 050501 (2011)

[19] M.R. Pederson, A. Ruzsinszky, and J.P. Perdew, J. Chem. Phys. **140**, 121103 (2014)

[20] M.R. Pederson, J. Chem. Phys. **142**, 064112 (2015)

[21] M.R. Pederson and T. Baruah: *Self-Interaction Corrections Within the Fermi-Orbital-Based Formalism*, Adv. At. Mol. Opt. Phys. **64**, 153-180, ed. by E. Arimondo, C.C. Lin, and S.F. Yelin (Academic Press, 2015)

[22] D.-Y. Kao, K. Withanage, T. Hahn, J. Batool, J. Kortus, and K.A. Jackson, J. Chem. Phys. **147**, 164107 (2017)

[23] J. Batool, T.Hahn, and M.R. Pederson, J. Comput. Chem. **40** 2301 (2019)

[24] S. Akter, Y. Yamamoto, C.M. Diaz, K.A. Jackson, R.R. Zope, and T. Baruah, J. Chem. Phys. **153**, 164304 (2020)

[25] K.P.K. Withanage, K.A. Jackson, and M.R. Pederson, J. Chem. Phys. **156** 231103 (2022)

[26] J.I. Melo, M.R. Pederson, and J.E. Peralta, J. Phys. Chem. A **127**, 527 (2023)

[27] M.R. Pederson, A.I. Johnson, K.P.K. Withanage, S. Dolma, G.B. Flores, Z. Hooshmand, K. Khandal, P.O. Lasode, T. Baruah, and K.A. Jackson, J. Chem. Phys. **158** (2023)

14 Coupled-Cluster Theory for Materials Science

Andreas Grüneis
Vienna University of Technology
Wiedner Hauptstraße 8-10/136, 1040 Wien

Contents

1 Introduction **2**

2 Fundamental concepts of coupled cluster theory **2**
2.1 Hartree-Fock theory and Slater determinants 3
2.2 The exponential Ansatz 3
2.3 Hausdorff expansion 4
2.4 Beyond the ground state 5
2.5 The coupled cluster doubles equations 6
2.6 Size consistency and extensivity 6
2.7 Caveats of coupled cluster theory 8

3 Coupled cluster theory and its relation to the RPA **10**

4 Coupled cluster theory applied to the uniform electron gas **11**

5 Conclusion and summary **13**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

Coupled cluster theories are widely-used to study many-body systems in nuclear physics, molecular quantum chemistry and solid state physics. This chapter introduces fundamental concepts of coupled cluster (CC) theory and discusses its application to the electronic structure theory problem. The present chapter serves as a primer to this topic. A more general overview of coupled cluster theory and its applications in quantum chemistry can be found in Ref. [1]. For an introduction to the theoretical formalism from the perspective of theoretical chemistry we recommend Refs. [2–5]. During the past decades a large body of well-written scientific articles and text books on coupled cluster theory has been published. Many of these can be found in the bibliographies of Refs. [1–4].

Coupled cluster theory was first proposed by Fritz Coester and Hermann Kümmel in the field of nuclear physics [6, 7]. Jiri Cizek and Josef Paldus introduced the method for electron correlation [8, 9]. Since then, coupled cluster theory has successfully been applied to study many-electron Hamiltonians for a wide range of systems, including atoms, molecules and even solids. At the same time, many of the most popular model Hamiltonians including lattice Hamiltonians and the uniform electron gas have also been explored and used to benchmark the accuracy of coupled cluster theories. It is therefore fair to say that CC theories are among the most successful approaches to treat many-body problems in quantum physics.

2 Fundamental concepts of coupled cluster theory

A fundamental approach to solve the time-independent Schrödinger equation for many-electron systems is based on finding accurate approximations to the true many-electron wavefunction $|\Psi\rangle$. An important challenge of these so-called wavefunction based methods revolves around finding a representation of the many-electron wavefunction that is at the same time compact and accurate in describing electronic correlation effects. Here, compactness not only means that the number of parameters used in the expansion of the wavefunction is within the limits of the available computational resources and scales with a favorable power law with respect to the system size. It also implies that the evaluation of the required matrix elements of quantum mechanical operators can be carried out in a computationally efficient manner. Successful wavefunction based methods typically optimize the balance between complexity in the ansatz and efficiency in evaluating matrix elements or expectation values. In this regard, for example, variational quantum Monte Carlo techniques achieve a good balance by combining stochastic integration techniques with a sophisticated many-body correlation function referred to as Jastrow factor. As will be explained in the present section, coupled cluster methods employ an Ansatz for the wavefunction that benefits significantly from an effective factorization of the many-electron wavefunction that can be systematically improved.

2.1 Hartree-Fock theory and Slater determinants

Hartree-Fock theory employs the simplest possible Ansatz to the many-electron wavefunction that is antisymmetric under exchange of two coordinates or orbitals, as required for a fermionic wave function. For an N-electron wavefunction in real space the Hartree-Fock wavefunction is given by a Slater determinant constructed from one-electron (Bloch) orbitals with an appropriate pre-factor to ensure normalization such that

$$\Phi^{\mathrm{HF}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{x}_1) & \cdots & \varphi_n(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x}_n) & \cdots & \varphi_n(\mathbf{x}_n) \end{vmatrix}. \tag{1}$$

The HF determinant is constructed from a set of orthonormal one-electron orbitals, $\varphi_i(\mathbf{x})$, that are obtained by minimizing the Hartree-Fock energy, whereby all the coupling terms of the Hamiltonian between the Hartree-Fock determinant and the corresponding single-excited Slater determinants vanish, which is also referred to as Brillouin's theorem. Here $\mathbf{x}$ is a compound index of spatial and spin coordinate. In periodic systems the index i is a compound index of the Bloch wave vector k_i used to sample the first Brillouin zone and the band index n_i. Hartree-Fock (HF) theory can be regarded as a low rank tensor approximation to the many-electron wavefunction, employing an antisymmetrized outer product of single electron orbitals to approximate the many-body wavefunction. Hence, by construction, HF theory neglects electronic correlation effects that cannot be captured using products of one-electron functions only. For brevity we will use to the following notation for the HF wavefunction $|0\rangle = \left|\Phi^{\mathrm{HF}}\right\rangle$.

2.2 The exponential Ansatz

The CC approximation is based on an exponential *Ansatz* for the electronic wavefunction [8, 1] acting on a single Slater determinant $|0\rangle$,

$$|\Psi_{\mathrm{CC}}\rangle = e^{\hat{T}} |0\rangle , \tag{2}$$

where the *cluster operator* consists of second-quantized neutral excitation operators

$$\hat{T} = \sum_{\mu} t_\mu \hat{\tau_\mu}, \qquad t_\mu \in \mathbb{C} \tag{3}$$

with μ labeling excitation configurations. For instance, when considering only singles and doubles excitations (Coupled Cluster Singles Doubles (CCSD)) the unrestricted CCSD cluster operator is given by

$$\hat{T} = \sum_{a,i} t_i^a \, \hat{a}_a^\dagger \hat{a}_i + \frac{1}{4} \sum_{a,b,i,j} t_{ij}^{ab} \, \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i \tag{4}$$

where the indices in $\{a, b, c, \ldots\}$ denote virtual or unoccupied spin orbitals and $\{i, j, k, \ldots\}$ denote occupied spin orbitals. Orbitals are occupied or unoccupied with respect to the reference

Slater determinant $|0\rangle$, which may come from a HF calculation. The excitation operators are defined such that they create excited determinants when acting on $|0\rangle$ such that

$$\hat{a}_a^\dagger \hat{a}_i \,|0\rangle = \left|\begin{matrix} a \\ i \end{matrix}\right\rangle \quad \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i \,|0\rangle = \left|\begin{matrix} ab \\ ij \end{matrix}\right\rangle . \tag{5}$$

Note that the following equation is satisfied.

$$\langle 0| \,\hat{a}_a^\dagger \hat{a}_i = 0. \tag{6}$$

We note in passing, that the nth-order CC ansatz including up to the nth-order excitation operator is exact for n-electron systems. One advantage of the different approximations to the cluster operator is that they constitute a hierarchy, which starting from the one-particle HF approximation, allows for a systematic treatment of the quantum many-body effects that are captured with an increasing level accuracy by employing CCSD, CCSDT and CCSDTQ theories. Calculated ground state properties typically exhibit decreasing errors using higher levels of theory.
Here, we will restrict the discussion to the case of CCSD. Applying the Coupled Cluster (CC) *ansatz* to the time-independent many-body electronic Schrödinger equation results in

$$\bar{H} \,|0\rangle = e^{-\hat{T}} \hat{H} e^{\hat{T}} \,|0\rangle = E_{\mathrm{CC}} \,|0\rangle \tag{7}$$

where E_{CC} is the coupled cluster energy, and we have implicitly defined the similarity transformed Hamiltonian $\bar{H}$. The state $|\Psi_{\mathrm{CC}}\rangle$ is parametrized by the coefficients t_μ, which can be obtained by projection. In the case of CCSD one projects the Schrödinger equation onto the singles and doubles sectors of the Hilbert space

$$E_{\mathrm{CC}} = \langle 0| \,\bar{H} \,|0\rangle \tag{8}$$

$$0 = \langle 0| \,\hat{a}_i^\dagger \hat{a}_a \bar{H} \,|0\rangle \tag{9}$$

$$0 = \langle 0| \,\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \bar{H} \,|0\rangle \,. \tag{10}$$

Equations (8–10) are a set of coupled non-linear equations in terms of the amplitudes t_i^a and t_{ij}^{ab} that are solved by iterative methods.

2.3 Hausdorff expansion

The similarity transformed Hamiltonian $\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$ occurring in the coupled cluster equations is an effective and non-Hermitian Hamiltonian, which can be expressed using the Hausdorff expansion

$$\bar{H} = \hat{H} + \big[\hat{H},\, \hat{T}\big] + \frac{1}{2!}\Big[\big[\hat{H},\, \hat{T}\big],\, \hat{T}\Big] + \frac{1}{3!}\Big[\Big[\big[\hat{H},\, \hat{T}\big],\, \hat{T}\Big],\, \hat{T}\Big] + \frac{1}{4!}\Big[\Big[\big[[\hat{H},\, \hat{T}],\, \hat{T}\big],\, \hat{T}\Big],\, \hat{T}\Big] + \cdots \tag{11}$$

Recalling that $\hat{H}$ in second quantization is given by

$$\hat{H} = \sum_{pq} h_{pq} \,\hat{a}_p^\dagger \hat{a}_q + \frac{1}{4} \sum_{pqrs} g_{pqrs} \,\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s \tag{12}$$

and

$$\left[\hat{a}_p^\dagger \hat{a}_q,\, \hat{a}_a^\dagger \hat{a}_i\right] = \hat{a}_p^\dagger \delta_{qa} \hat{a}_i - \hat{a}_a^\dagger \delta_{ip} \hat{a}_q, \tag{13}$$

it follows that Eq. (11) terminates exactly after the fourth nested commutator. We stress that the set of indices $\{p, q, r, s, \ldots\}$ denotes both occupied or unoccupied orbitals.
Substituting the expression for $\bar{H}$ in the equation for the coupled cluster energy yields

$$E_{\mathrm{CC}} = \langle 0|\, \bar{H}\, |0\rangle = \langle 0|\, \hat{H}\, |0\rangle + \langle 0|\, \hat{H}\hat{T}\, |0\rangle = E_{\mathrm{HF}} + \langle 0|\, \hat{H}\Big(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\Big)\, |0\rangle\,. \tag{14}$$

Note that this equation simplifies significantly using the Brillouin theorem ($\langle 0|\, \hat{H}\hat{a}_a^\dagger \hat{a}_i\, |0\rangle = 0$), the fact that $\hat{H}$ can de-excite at most two electrons and due to $\langle 0|\, \hat{a}_a^\dagger \hat{a}_i = 0$. Consequently, only singles and doubles amplitudes contribute to the CC energy. If higher-order excitations in the cluster operator are considered, their contribution to the energy is only indirect by the amplitude equations. We note that the correlation energy is implicitly defined as the difference between the exact ground state energy and the HF energy E_{HF}. Similarly to the energy, the singles and doubles amplitude equations defined in Equations (9–10) can be obtained.

2.4 Beyond the ground state

A common way to obtain excited states based on the CCSD theory is through diagonalizing the similarity transformed Hamiltonian $\bar{H}$ in a suitable subspace of the Hilbert space [10]. We present the neutral variant of this approach, also called electronically excited equation of motion, for which the number of electrons is conserved. In consequence, restricting from now on again the analysis to singles and doubles excitations, the ansatz for an excited state $\hat{R}\,|\Psi_{\mathrm{CC}}\rangle$ is

$$\hat{Q}\hat{H}\hat{R}\,|\Psi_{\mathrm{CC}}\rangle = \hat{Q}\hat{H}\hat{R}e^{\hat{T}}\,|0\rangle = E_R\hat{Q}\hat{R}\,|\Psi_{\mathrm{CC}}\rangle \tag{15}$$

where

$$\hat{R} = r_0 + \sum_{a,i} r_i^a\, \hat{a}_a^\dagger \hat{a}_i + \frac{1}{4}\sum_{a,b,i,j} r_{ij}^{ab}\, \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_j \hat{a}_i\,, \qquad r_\mu \in \mathbb{C} \tag{16}$$

is a linear excitation operator, E_R is its excitation energy and $\hat{Q}$ is the projector onto the singles and doubles excitations manifold of the Hilbert space, this is,

$$\hat{Q} = \sum_{a,i} \left|\begin{matrix} a \\ i \end{matrix}\right\rangle\!\left\langle\begin{matrix} a \\ i \end{matrix}\right| + \frac{1}{4}\sum_{a,b,i,j} \left|\begin{matrix} ab \\ ij \end{matrix}\right\rangle\!\left\langle\begin{matrix} ab \\ ij \end{matrix}\right|. \tag{17}$$

Equation (15) is equivalent to a commutator equation only involving $\bar{H}$ and the excitation energy difference ΔE_R between E_R and the correlated ground state E_{CC},

$$\left[\hat{Q}\bar{H},\, \hat{R}\right]|0\rangle = \Delta E_R\, \hat{Q}\hat{R}\,|0\rangle. \tag{18}$$

It is worthwhile noting that the commutator on the left-hand-side means that only connected diagrams need to be considered in the expansion. Equation (18) motivates the name *equation of motion* due to its resemblance to the time-dependent *Heisenberg picture* differential equation for the time evolution of an operator.

2.5 The coupled cluster doubles equations

For computer implementations it is necessary to rewrite the above equations only in terms of one- and two-electron integrals. Here we seek to give the final result of the corresponding expressions for CCD theory only. We note that in the case of, for instance, the uniform electron gas (UEG) Hamiltonian, due to the symmetry, single excitations are absent. Therefore we consider CCD a particularly instructive case to learn more about CC theory. The cluster amplitudes t_{ij}^{ab} are obtained by solving the quadratic amplitude equations $\left\langle {}_{ij}^{ab}\right| e^{-\hat{T}_2}\hat{H}e^{\hat{T}_2}||0\rangle = 0$ that in a spin-orbital basis read

$$\begin{aligned} t_{ij}^{ab} = \frac{1}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b}\Big(& \langle ij||ab\rangle + \langle cj||kb\rangle t_{ik}^{ac} + \langle ci||ka\rangle t_{jk}^{bc} + \langle cd||kl\rangle t_{lj}^{db} t_{ik}^{ac} \\ & + \frac{1}{2}\langle cd||ab\rangle t_{ij}^{cd} + \frac{1}{2}\langle ij||kl\rangle t_{kl}^{ab} + \frac{1}{4}\langle cd||kl\rangle t_{ij}^{cd} t_{kl}^{ab} \\ & - \langle cj||ka\rangle t_{ik}^{bc} - \langle ci||kb\rangle t_{jk}^{ac} - \langle cd||kl\rangle t_{lj}^{da} t_{ik}^{bc} \\ & + \frac{1}{2}\langle cd||kl\rangle \left[t_{lj}^{ab} t_{ik}^{cd} - t_{li}^{ab} t_{jk}^{cd} + t_{ij}^{db} t_{kl}^{ac} - t_{ij}^{da} t_{kl}^{bc}\right] \Big). \end{aligned} \tag{19}$$

In the above equation repeated indices are summed over. We recall that the indices i, j, k and l label occupied orbital indices, whereas a, b, c and d label virtual orbital indices. ε correspond to the HF one-electron energies and the anti-symmetrized two-electron integrals are defined by $\langle ij||ab\rangle = \langle ij|ab\rangle - \langle ij|ba\rangle$, where

$$\langle ij|ab\rangle = \int_\Omega\int_\Omega d\mathbf{x}_1 d\mathbf{x}_2 \frac{\varphi_i^*(\mathbf{x}_1)\varphi_j^*(\mathbf{x}_2)\varphi_a(\mathbf{x}_1)\varphi_b(\mathbf{x}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}. \tag{20}$$

In the above expression the spin-orbitals φ depend on the space-spin coordinate $\mathbf{x} = (\mathbf{r}, \sigma)$ and the spatial coordinates are integrated over all space. Equation (19) is solved for the amplitudes in an iterative manner by updating the amplitudes in every iteration using the right-hand side of Eq. (19). Convergence can be accelerated using standard techniques such as direct inversion of the iterative subspace (DIIS) [11]. Once the amplitudes are obtained, the CCD correlation energy can be calculated by

$$E_c^{\mathrm{CCD}} = \sum_{ijab}\frac{1}{4}\langle ij||ab\rangle t_{ij}^{ab}\,. \tag{21}$$

2.6 Size consistency and extensivity

We now discuss size extensivity and the convergence of computed ground state energies to the thermodynamic limit (TDL). These concepts are highly relevant for the application to solids. In contrast to molecular systems, properties of solids have to be calculated in the thermodynamic limit to enable a direct comparison to experiment. The TDL can be approached using; for example, (i) sampling of the Brillouin zone with increasingly dense k-point meshes and in periodic boundary conditions, (ii) studying increasingly large supercells in periodic boundary conditions, or employing (iii) increasingly large clusters with open boundary conditions and/or

embedding methods. Once the thermodynamic limit is approached with respect to the number of k-points or the number of atoms in the cluster, extensive properties such as the correlation energy per atom are converged to a constant value.

An important advantage of truncated coupled cluster theories compared to, for instance truncated configuration interaction methods is their size consistency. Size consistency is a concept of particular importance in quantum chemistry, which judges if the calculated quantities have the correct asymptotic size dependence or not. For extensive quantities, like the (correlation) energy, a given size-consistent method should yield the asymptotic K^1 dependence where K is the number of wave vector sampling points in the Brillouin zone [12]. Obviously, the methods with incorrect asymptotic K^α dependence of $\alpha < 1$, like the truncated configuration-interaction methods, lead, in the thermodynamic limit, to the total energy per unit cell equal to that of the HF mean-field approximation. The size consistency of coupled cluster theories can also be understood via either the diagrammatic criteria [13] or the supermolecule criterion [4]. It was argued that approximate post-HF correlation methods cannot capture the variational and size-consistent properties simultaneously [14].

The thermodynamic limit is approached as $N \to \infty$, where N is the number of particles in the simulation (super-)cell while the density is kept constant. Once the thermodynamic limit is approached, correlation energies per atom need to be converged to a constant for periodic systems, corresponding to $\alpha = 1$. Finite size errors are defined as the difference between the TDL and the finite simulation cell results. However, the convergence of calculated properties to the thermodynamic limit is very slow, often exceeding the computational resources of even modern supercomputers due to the steep scaling of the computational complexity of most post-HF methods with respect to system size. We stress that many properties such as the binding energy of molecules on surfaces converge slower than their counterparts calculated on the level of mean-field theories such as density-functional theory (DFT). This originates from the fact that correlated post-HF methods capture long-range electronic correlation effects such as van der Waals interactions explicitly. Even though the corresponding long-ranged contribution to the electronic correlation energy is small compared to short-ranged correlation energy contributions, the accumulation of weak van der Waals interactions can become a non-negligible contribution to the property of interest. Different strategies have been developed to correct for finite size errors that are defined as the difference between the thermodynamic limit and the finite simulation cell results. These strategies often involve extrapolation methods or range-separation techniques. Local theories that employ correlation energy expressions depending on localized electron pairs, can approximate correlation energy contributions of long-distant pairs using computationally more efficient yet less accurate theories. Alternatively local theories can account for electron pairs that are disregarded based on a distance criterion by using an R^{-6}-type extrapolation [15]. Canonical implementations of periodic post-HF methods employ scaling laws for extrapolations to the thermodynamic limit that are based on an analogue rationale [16–18]. Auxiliary field quantum Monte Carlo theory employs finite-size corrections that are based on parametrized density functionals obtained from finite uniform electron gas simulation cells [19].

2.7 Caveats of coupled cluster theory

2.7.1 Basis set convergence

The many-electron wavefunctions introduced above are expanded in a basis of Slater determinants constructed from (unoccupied) Hartree-Fock orbitals. The computational complexity of canonical coupled-cluster methods scales polynomially with respect to the number of unoccupied orbitals. Therefore the ability to span the relevant parts of the Hilbert space with as few orbitals as possible is crucial for the implementation of efficient periodic correlated methods. In practice all calculated quantities suffer from a basis-set incompleteness error that is caused by the truncation of the employed unoccupied orbital manifold. The optimal choice of the unoccupied orbital manifold minimizes the incompleteness error of the calculated quantity in a controllable manner.

Some of the most widely-used basis sets for the expansion of unoccupied orbitals include plane waves and Gaussian-type orbitals (GTOs). As an illustration of their respective characteristic properties we consider two limiting cases, the uniform electron gas and an atom in a box. From the perspective of the uniform electron gas, plane waves are the natural choice of basis to expand one- and many-electron wavefunction quantities. Plane waves are eigenfunctions of the kinetic energy operator, and exhibit the same periodicity as the simulation cell. In *ab initio* calculations, these plane waves also have a number of appealing features. A single cutoff parameter that limits the kinetic energy of the included plane waves is used to systematically expand the plane wave basis to completeness which is free from basis-set superposition errors (BSSE) and linear dependencies. However, there are obvious drawbacks to plane wave expansions. They lack reference to the nature of the atomic environment, having equal basis coverage throughout the cell. This can lead to a substantial waste of computational effort when studying an atom or molecule in a box [20].

For atoms or molecules, GTOs form a very compact orbital basis. Their widespread use in the field of quantum chemistry has lead to standardized tabulated basis sets of increasing size and flexibility [21, 22]. Orbitals beyond the core and valence shells are included to account for appropriate polarization of the atomic wavefunctions in bonding environments, and to provide a description of correlation effects. Basis sets are commonly arranged in hierarchies so that they can be systematically expanded to allow for consistent and extrapolatable convergence. Gaussian-type orbitals are used in a range of periodic electronic structure codes. However, the introduction of such local basis sets also leads to several shortcomings such as basis set superposition errors (BSSE) and linear dependencies of diffuse atom-centered basis functions in densely packed solids. These problems can partly be accounted for by counterpoise BSSE corrections and removing linearly dependent basis functions. The local nature of these functions is often used for reduced scaling techniques in order to approach linear scaling mean-field treatments, and can also be extended to local treatment of correlation.

For the calculation of energy differences such as the adsorption energy of a molecule on a surface it is beneficial to employ basis sets that can be truncated such that a large fraction of the incompleteness error cancels in a controllable manner. GTOs exhibit this advantageous prop-

erty, allowing for obtaining accurate estimates of interaction energies between weakly interacting fragments such as binding energies of physisorbed molecules on surfaces, despite suffering from large incompleteness errors in the respective absolute energies.

Another approach to obtain compact unoccupied orbital manifolds for the expansion of many-electron wavefunctions is provided by natural orbitals [23]. Natural orbitals are obtained by diagonalizing the unoccupied-unoccupied orbital block of the reduced density matrix and truncating the obtained natural orbital manifold according to their occupation number. This procedure yields for many applications an optimal unoccupied orbital manifold. To reduce the computational cost of this procedure it is possible to approximate the reduced density matrix at a lower-level of theory such as low-order perturbation theory only. Natural orbitals "down-fold" the unoccupied orbitals calculated using plane-wave basis sets for atoms and molecules in a box to manifolds that are similarly compact as GTOs [20].

Despite all the considerations outlined above, the convergence of the many-electron wavefunction and that of calculated expectation values such as the correlation energy is frustratingly slow with respect to the number of unoccupied orbitals [5, 24]. Therefore extrapolation techniques that remove the remaining basis set incompleteness error are needed on top of these fairly large basis set calculations. In the case of plane wave basis set calculations analytic and numerical results from perturbation theory suggest a $1/M$ decay of the basis-set incompleteness error where M is the number of plane waves used in the calculation, allowing for straightforward extrapolation to the complete basis set (CBS) limit [25]. Similar scaling laws are employed for the extrapolation of correlation energies to the complete basis set limit using GTOs [5].

The slow convergence of properties calculated using wavefunction based methods with respect to the number of orbitals originates from the difficulty to describe the many-electron wavefunction in the vicinity of the electron cusp. As the electrons coalesce, a derivative discontinuity or 'cusp' must arise, so that a divergence in the kinetic energy operator cancels an opposite one in the potential. The shape of the wavefunction at the cusp is exactly defined to first-order in the interelectronic distance by the Kato cusp conditions [26,27]. The $a\ priori$ inclusion of the cusp conditions in the wavefunction ansatz is a cornerstone of explicitly correlated or so-called F12 theories [28, 24, 29, 30]. Explicitly correlated methods augment the conventional wavefunction expansions discussed in the previous section with additional terms that account for the cusp conditions explicitly. Since electronic correlation is for the most part a short-ranged phenomenon, the proper description of the wavefunction shape at short interelectronic distances allows for capturing the largest fraction of the correlation energy in solids and molecules.

2.7.2 Computational cost

The scaling of the computational cost of canonical coupled cluster theory is dominated by the contractions present in the amplitude equations. Although some terms can be contracted efficiently by finding the optimal order of contraction over the nested summations over orbital indices, a limiting scaling remains. The scaling of CCSD and CCSDT is $\mathcal{O}(N^6)$ and $\mathcal{O}(N^8)$, where N is a measure of the system size (occupied or unoccupied orbitals) and arises from the

use of spatially delocalized canonical orbitals. Canonical orbitals are conceptually and computationally convenient. They are orthogonal and diagonalize the Fock matrix, greatly simplifying the post-HF correlation schemes. However, these orbitals are spatially delocalized and their use does not allow one to exploit the fact that electronic correlation is a short-ranged phenomenon. The use of spatially localized, instead of canonical, orbitals allows to construct coupled cluster algorithms that scale more favorably with system size, down to even $\mathcal{O}(N)$, at the price of a significant increase in complexity of the underlying equations with respect to their canonical counterparts. Some of the most notable amongst them are (based on) the local correlation method of Pulay and Saebø [31, 32], the so-called "Local Ansatz" of Stollhof and Fulde [33], the method of increments of Stoll [34–36], or the use of truncated pair natural orbitals [37, 38]. The reduced scaling algorithms mentioned above require that the occupied orbitals can be localized, i.e., a unitary transformation over the manifold of occupied orbitals can be found which optimizes the expectation value of an operator measuring the degree of localization. There exist systems which do not allow for a sufficient degree of orbital localization. In these systems, the character of electronic correlation is intrinsically more delocalized. A prominent example where this is the case corresponds to the uniform electron gas.

3 Coupled cluster theory and its relation to the RPA

The random phase approximation (RPA) to the correlation energy dates back to the 1950s. It was first introduced by Macke to predict convergent correlation energies [39] in the uniform electron gas and was also developed by Bohm and Pines [40] for the collective description of electron interactions. In the case of the uniform electron gas, the RPA captures the most important terms of the correlation energy expansion around the high-density limit ($r_s \to 0$). In the field of *ab initio* computational materials science the exact-exchange plus correlation in the random-phase approximation has attracted renewed and widespread interest in the last two decades. This is due to the fact that computationally increasingly efficient implementations have become available and that this method is capable of describing all interatomic bonding situations reasonably well: ionic, covalent, metallic, and even van der Waals bonding. The computational complexity can even be lowered to $\mathcal{O}(N^3)$ in real space formulations [41]. Thus, the complexity of an RPA calculation does not exceed that of a canonical hybrid density functional theory calculation, the prefactor is however considerably larger. The RPA correlation energy can be derived from many-electron Green function theory, or using the adiabatic-connection fluctuation-dissipation theorem (ACFDT) – or from coupled-cluster theory.
As shown in [42], it is possible to transform the RPA equations, that are usually expressed in a general eigenvalue problem, to a quadratic Riccati equation that reads

$$t_{ij}^{ab} = \frac{1}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b}\Big(\langle ij|ab\rangle + \langle cj|kb\rangle t_{ik}^{ac} + \langle ci|ka\rangle t_{jk}^{bc} + \langle cd|kl\rangle t_{lj}^{db} t_{ik}^{ac}\Big). \tag{22}$$

In the above equation we sum over repeated indices. We stress that in the above equation the ε correspond to the DFT one-electron energies. Once the amplitudes are obtained, the RPA

correlation energy can be calculated by

$$E_c^{\mathrm{RPA}} = \sum_{ijab} \frac{1}{2} \langle ij|ab\rangle \, t_{ij}^{ab}. \tag{23}$$

Although the above formulation does not allow for an efficient computer implementation of the RPA, it illustrates that the RPA and CCSD are closely related.
In the rings-only approximation, the second, third and fourth lines of Eq. (19) are disregarded. Furthermore the random-phase approximation includes the direct rings only. This implies that instead of using the (double bar) anti-symmetrized integrals, only $\langle ij|ab\rangle$ integrals are employed in the RPA amplitude and energy equations, making it necessary to employ a different prefactor in the correlation energy expression to stay consistent with many-body perturbation theory. Consequently, RPA can not be viewed as a wavefunction theory although it can be obtained from the coupled cluster amplitude equations as explained above. In a diagrammatic formulation the close relationship between coupled cluster theory and the RPA becomes more obvious for both ground and excited state properties as discussed in detail in Refs. [43, 42].

4 Coupled cluster theory applied to the uniform electron gas

One of the best studied systems in electronic structure theory is the uniform electron gas (UEG). Here, we seek to illustrate important concepts of CC theory for the example of the UEG. In the UEG, the one-electron orbitals are plane waves with wave vectors $\vec{k}_i$, $\vec{k}_j$ and $\vec{k}_a$, $\vec{k}_b$. This allows to write the two-electron repulsion integral as $\langle ij|ab\rangle = \upsilon_{ij}^{ab} = \upsilon(\mathbf{q})\delta_{\mathbf{k}_i-\mathbf{k}_a,\mathbf{k}_b-\mathbf{k}_j}$. The momentum transfer vector is given by $\mathbf{q} = \mathbf{k}_i - \mathbf{k}_a$. The Coulomb kernel is defined as $\upsilon(\mathbf{q}) = \frac{4\pi}{\Omega|\mathbf{q}|^2}$, with Ω being the volume of the simulation cell. We stress again that, due to the symmetry of the UEG Hamiltonian, single excitations are absent. With this, the CCD correlation energy is defined as

$$E_{\mathrm{c}}^{CCD} = \sum_{ijab} \upsilon_{ij}^{ab} \left(2t_{ij}^{ab} - t_{ji}^{ab}\right) = \sum_{\mathbf{q}} \upsilon(\mathbf{q}) S^{CCD}(\mathbf{q}). \tag{24}$$

The amplitudes t_{ij}^{ab} are obtained by solving the CCD amplitude equations. We note, however, that this section employs a notation where the orbital indices refer to spatial orbitals, i.e., the spin-coordinates have been integrated, which results in slightly different expressions. It is important to note that in the UEG t_{ij}^{ab} inherits the momentum conservation of the two-electron operator in the Hamiltonian such that $t_{ij}(\mathbf{q}) = t_{ij}^{ab}\, \delta_{\mathbf{q},\mathbf{k}_b-\mathbf{k}_j} \delta_{\mathbf{q},\mathbf{k}_i-\mathbf{k}_a}$. The quantity $t_{ij}(\mathbf{q})$ has two indices that correspond to occupied orbital indices, representing plane wave vectors inside the Fermi sphere, whereas $\mathbf{q}$ corresponds to a momentum transfer vector that excites electronic states into regions outside the Fermi sphere. Likewise the transition structure factor $S(\mathbf{q})$, introduced in Eq. (24), depends on $\mathbf{q}$, which gives access to the dependence of the correlation energy on the interelectronic interaction distance. The transition structure factor is defined as

$$S^{CCD}(\mathbf{q}) = \sum_{ijab} \frac{\delta \upsilon_{ij}^{ab}}{\delta \upsilon(\mathbf{q})} \left(2t_{ij}^{ab} - t_{ji}^{ab}\right). \tag{25}$$

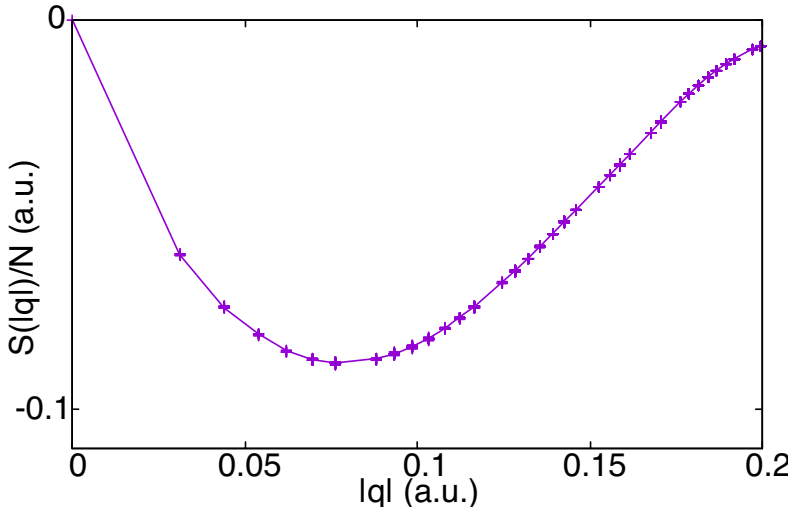


Fig. 1: *Twist-averaged structure factors for CCD and 246 electrons in the uniform electron gas simulation cell at density corresponding to* $r_s = 20$.

In the above equation, the functional derivative $\frac{\delta v_{ij}^{ab}}{\delta v(\mathbf{q})} = \delta_{\mathbf{q},\mathbf{k}_b-\mathbf{k}_j}\delta_{\mathbf{q},\mathbf{k}_i-\mathbf{k}_a}$ enables a concise notation.

We now study the behavior of the CCD correlation energy contribution for $\lim_{|\mathbf{q}|\to\infty}$ and short $\lim_{|\mathbf{q}|\to 0}$. Figure 1 depicts the computed CCD transition structure factor for 246 electrons at a density corresponding to $r_s = 20$ a.u., showing that $S(\mathbf{q})$ decays to zero in both limits. As already discussed by Bishop and Lührmann [44, 45] as well as Emrich and Zabolitzky [46], $\lim_{|\mathbf{q}|\to 0} S(\mathbf{q}) \propto |\mathbf{q}|$ and $\lim_{|\mathbf{q}|\to\infty} S(\mathbf{q}) \propto 1/|\mathbf{q}|^4$. It is important to note that the functional behavior in these limits originates from two important physical principles. The $\lim_{|\mathbf{q}|\to\infty}$ corresponds to the short-range limit in real space. It is known that as the electrons coalesce, a derivative discontinuity or 'cusp' must arise, so that a divergence in the kinetic energy operator cancels an opposite one in the interelectronic potential. Without proof, we stress that a linear behavior in the wavefunction as a function of the interelectronic distance results in a $1/|\mathbf{q}|^4$ behavior of $S(\mathbf{q})$. In fact the relatively slow convergence of $S(\mathbf{q})$ for large $|\mathbf{q}|$ is the cause for the slow basis set convergence of correlation energies. In other words, large numbers of unoccupied orbitals are needed in practice to capture significant contributions to the electronic correlation energy in the vicinity of the cusp. In practice one might perform several correlation energy calculations for different numbers of virtual orbitals and extrapolates the computed energies to the complete basis set (CBS) limit. The $\lim_{|\mathbf{q}|\to 0}$ corresponds to the long-range behavior of the electronic correlation energy. Unlike the short-range, the long-range behavior qualitatively differs between insulators and metals. Here, we consider the metallic uniform electron gas, which leads to a linear slope of $S(\mathbf{q})$ around $|\mathbf{q}| = 0$. This behavior can be understood by the close relationship between the RPA and CCD. Without proof, we stress that in the $\lim_{|\mathbf{q}|\to 0}$ RPA becomes identical to CCD, [46] which also explains the slow convergence of computed correlation energies to the thermodynamic limit.

5 Conclusion and summary

In this chapter we have discussed fundamental concepts of coupled cluster theory and its relation to other widely-used methods. CC theories offer a systematically improvable hierarchy of wavefunction based methods for the study of many-body problems. Due to their computational complexity, however, most coupled cluster calculations in electronic structure theory are performed at the truncation levels of CCSD and CCSD(T), where the effect of T is approximated in a perturbative manner. We have discussed for the example of the uniform electron gas, that converging the CC correlation energies to the complete basis set limit and the thermodynamic limit is difficult, which can partly be explained by the slow asymptotic convergence of the underlying electron structure factor. Another caveat of CC theories, which was not discussed in the present chapter, is that these approximate wavefunction theories are not well-suited to treat systems with strong static correlation effects, for example, bond dissociation problems. However, CCSD(T) theory is one of the most accurate *ab initio* methods currently available to compute, for example, reaction energies for a wide range of systems. Recent applications to solids have also shown that CCSD(T) can achieve a similar level of accuracy for semiconductors and insulators. More work remains to be done to expand the scope of CC theories to more complex systems and a larger number of properties which is beyond the scope of discussion of the present chapter.

Acknowledgement

I sincerely thank my coworkers Andreas Irmler, Tobias Schäfer, Felix Hummel, Alejandro Gallo, Nikolaos Masios and Faruk Salihbegovic for the fruitful discussions and cooperations on coupled cluster theory in the last years.

References

[1] R.J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007)

[2] T.D. Crawford and H.F. Schaefer: *An Introduction to Coupled Cluster Theory for Computational Chemists* (Wiley-Blackwell, 2007), pp. 33–136

[3] I. Shavitt and R. J. Bartlett: *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory* (Cambridge university press, 2009)

[4] A. Szabo and N.S. Ostlund: *Modern Quantum Chemistry* (McGraw-Hill, New York, 1996)

[5] T. Helgaker, P.Jørgensen, and J. Olsen: *Molecular Electronic-Structure Theory* (Wiley, 2000)

[6] F. Coester, Nucl. Phys. **1**, 421 (1958)

[7] F. Coester and H. Kümmel, Nucl. Phys. **17**, 477 (1960)

[8] J. Cizek, The Journal of Chemical Physics **45**, 4256 (1966)

[9] J. Cizek and J. Paldus, Int. J. Quantum Chem. **5**, 359 (1971)

[10] J.F. Stanton and R.J. Bartlett, J. Chem. Phys. **98**, 7029 (1993)

[11] P. Pulay, Chemical Physics Letters **73**, 393 (1980)

[12] S. Hirata, Theor Chem Acc **129**, 727 (2011)

[13] R. Bartlett, Ann. Rev. Phys. Chem. **32**, 359 (1981)

[14] S. Hirata and I. Grabowski, Theor Chem Acc **133**, 1 (2014)

[15] D. Usvyat, K. Sadeghian, L. Maschio, and M. Schütz, Phys. Rev. B **86**, 045412 (2012)

[16] G.H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013)

[17] M.D. Ben, J. Hutter, and J. Vandevondele, J. Chem. Theory Comput. **9**, 2654 (2013)

[18] J. McClain, Q. Sun, G.K.-L. Chan, and T.C. Berkelbach, J. Chem. Theory Comput. **13**, 1209 (2017)

[19] H. Kwee, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **100**, 126404 (2008)

[20] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theory Comput. **7**, 2780 (2011)

[21] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989)

[22] N. Balabanov and K. Peterson, J. Chem. Phys. **123**, 064107 (2005)

[23] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955)

[24] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2012)

[25] J.J. Shepherd, A. Grüneis, G.H. Booth, G. Kresse, and A. Alavi, Phys. Rev. B **86**, 035111 (2012)

[26] R. T. Pack and W. Byers Brown, J. Chem. Phys. **45**, 556 (1966)

[27] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957)

[28] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991)

[29] A. Grüneis, S. Hirata, Y.-y. Ohnishi, and S. Ten-no, J. Chem. Phys. **146**, 080901 (2017)

[30] G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009)

[31] S. Saebø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993)

[32] M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999)

[33] G. Stollhoff and P. Fulde, Z. Phys. B **29**, 231 (1978)

[34] H. Stoll, Phys. Rev. B **46**, 6700 (1992)

[35] B. Paulus, Phys. Rep. **428**, 1 (2006)

[36] P. Schwerdtfeger, B. Assadollahzadeh, and A. Hermann, Phys. Rev. B **82**, 205111 (2010)

[37] F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009)

[38] Q. Ma, M. Schwilk, C. Köppl, and H.-J. Werner, J. Chem. Theory Comput. **13**, 4871 (2017)

[39] W. Macke, Z. Naturforsch. **5a**, 192 (1950)

[40] D. Pines and D. Bohm, Phys. Rev. **85**, 338 (1952)

[41] M. Kaltak, J. Klimes, and G. Kresse, J. Chem. Theory Comput. **10**, 2498 (2014)

[42] G.E. Scuseria, T.M. Henderson, and D.C. Sorensen, J. Chem. Phys. **129**, 231101 (2008)

[43] T. C. Berkelbach, J. Chem. Phys. **149**, 041103 (2018)

[44] R. F. Bishop and K. H. Lührmann, Phys. Rev. B **17**, 3757 (1978)

[45] R. F. Bishop and K. H. Lührmann, Phys. Rev. B **26**, 5523 (1982)

[46] K. Emrich and J. G. Zabolitzky, Phys. Rev. B **30**, 2049 (1984)

15 Slave-Boson Theories of Multi-Orbital Correlated Systems

Nicola Lanatà
School of Physics and Astronomy
Rochester Institute of Technology
84 Lomb Memorial Drive, Rochester, NY 14623, USA

Contents

1 Introduction **2**

2 The multi-orbital Hubbard Hamiltonian **3**

3 Multi-orbital ghost Gutzwiller approximation (variational formulation) **4**

4 Reformulation using local reduced density-matrix **12**

5 Reformulation in terms of slave-boson amplitudes (connection with RISB) **13**

6 Reformulation in terms of embedding states (connection with DMET) **15**

7 Lagrange formulation of gGA (QE algorithmic structure) **19**

8 Generalizations, research directions and open problems **23**

9 Code availability **25**

A Useful mathematical definitions **26**

B Preliminaries on fermionic algebra and Fock states **27**

C One-body Hamiltonians and the Fermi-function matrix **31**

D Wick's theorem for one-body thermal states **34**

E Reduced density-matrix of a fermionic subsystem **36**

F Promoting of functions to independent variables **37**

G A useful matrix derivative **38**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Introduction

Strongly correlated electron systems, where electron-electron interactions are non-negligible, have captivated the condensed matter physics community due to the rich and often exotic physical phenomena they exhibit, including high-temperature superconductivity and magnetism. The theoretical study of such systems is indispensable in uncovering the underlying physics and holds tremendous promise for technological advancements. However, this is no small feat, as strongly correlated systems are notoriously difficult to describe accurately due to the complexity arising from these interactions. Advanced graduate students, Ph.D. students, and postdoctoral researchers venturing into the electronic structure of materials must be adept with an array of theoretical tools to effectively address these challenges.

Within the broad spectrum of techniques available for tackling strongly correlated systems, "slave-boson methods" have emerged as a versatile and powerful class of approaches [1–20]. These methods employ auxiliary particles, including both fermions and bosons, as subsidiary degrees of freedom to model strong electron-electron interactions, a concept shared across various theoretical frameworks such as tensor networks [21, 22] and neural-network quantum states [23]. In the context of slave-boson methods, the Gutzwiller Approximation (GA) theory and its extension, the ghost Gutzwiller Approximation (gGA), employ auxiliary fermionic degrees of freedom [10–12], while the Rotationally Invariant Slave Boson (RISB) theory and its extension, the ghost RISB (gRISB), utilize auxiliary bosonic degrees of freedom [19, 20].

Historically, slave-boson methods such as the GA were developed as computationally efficient alternatives to more demanding techniques, but this efficiency was achieved with a compromise on accuracy. However, recent advancements have shown that extensions like gGA [10–12], which incorporates auxiliary fermionic degrees of freedom to enrich the variational space, offer both computational efficiency and the potential for high accuracy. Notably, gGA has demonstrated an accuracy that is comparable to the more computationally demanding Dynamical Mean-Field Theory (DMFT) [24, 25], indicating that it might serve as an advantageous alternative, especially when aiming for a combination of accuracy and computational manageability.

This set of lecture notes is designed to provide a comprehensive overview of "slave-boson methods" with a particular focus on the gGA variational perspective. Through detailed technical expositions and unified, consistent notation, these notes aim to serve as a pedagogical resource for readers looking to delve into this field. While scientific literature often prioritizes conciseness over extensive derivations, this can sometimes leave out pedagogical explanations that are instrumental for learners and non-specialists. Our objective is to bridge this gap, furnishing the reader with a self-contained and in-depth comprehension of the subject matter. Furthermore, in the light of the active and burgeoning nature of this research area, we also elucidate the reformulations in terms of RISB/gRISB and explore connections with Quantum Embedding (QE) methods such as Density-Matrix Embedding Theory (DMET), while discussing their potential for catalyzing further theoretical and algorithmic advancements.

We suggest the reader begin with the Appendix, where we recapitulate some useful general notions of many-body theory. While the reader may already be familiar with these concepts, the

Appendix employs a consistent notation with the main text, making it a valuable starting point. Following this, in Sec. 3, we introduce the formalism from the variational perspective underlying the GA/gGA frameworks. In Sec. 5, we delve into the concept of slave-boson amplitudes; while keeping the focus on the GA/gGA formulation and the role of slave-boson amplitudes in formulating an efficient framework from this perspective, this is the juncture where the connection with RISB emerges, and we will guide the reader to the relevant literature exploring this connection. Sections 6 and 7 present the concept of embedding states; again, while maintaining a focus on the GA/gGA formulation, we highlight the role of embedding states in developing an efficient framework, and this is where the connection with Quantum Embedding theories such as DMET becomes apparent. Finally, Sec. 8 discusses further generalizations, new research directions, and open problems that aspiring researchers might find intriguing and rewarding to explore in their careers.

2 The multi-orbital Hubbard Hamiltonian

In this section, we describe the multi-orbital Hubbard Hamiltonian, which plays a fundamental role in the context of strongly correlated electron systems. We will elaborate on the terms and the notation involved. For readers not familiar with the mathematical structure of fermionic Fock spaces, a brief introduction is provided in Appedix B.
The multi-orbital Hubbard Hamiltonian encompasses local interactions as well as hopping terms. We contemplate a lattice system comprised of $\mathcal{N}$ fragments, each with multiple orbitals. The total Hamiltonian $\hat{H}$ can be formulated as:

$$\hat{H} = \sum_{i=1}^{\mathcal{N}} \hat{H}^i_{loc}[c^\dagger_{i\alpha}, c_{i\alpha}] + \sum_{i\neq j} \hat{T}_{ij} \,, \tag{1}$$

$$\hat{T}_{ij} = \sum_{\alpha=1}^{\nu_i} \sum_{\beta=1}^{\nu_j} [t_{ij}]_{\alpha\beta}\, c^\dagger_{i\alpha} c_{j\beta} \,, \tag{2}$$

where:

- i and j represent the indices of the fragments of the lattice.
- $\hat{H}^i_{loc}$ denotes an arbitrary local operator on fragment i, encompassing one-body and two-body terms.
- α and β index the fermionic modes (orbitals) within each fragment.
- $\hat{T}_{ij}$ symbolizes the hopping term between *different* fragments i and j.
- $[t_{ij}]_{\alpha\beta}$ are the matrix elements of the hopping term.

This Hamiltonian serves as the basis for our discussions on the ghost Gutzwiller approximation for multi-orbital systems in the subsequent sections.

3 Multi-orbital ghost Gutzwiller approximation (variational formulation)

In this section, we delve into the multi-orbital ghost Gutzwiller approximation (gGA) [10, 11], which is rooted in the variational principle and the limit of infinite dimensionality [1,2], building upon the multi-orbital Gutzwiller Approximation (GA) [1–9]. The gGA enriches the variational space by introducing auxiliary "ghost" fermionic degrees of freedom. This concept resonates with various theoretical frameworks such as extensions to DMET [26], matrix-product states and projected entangled pair states [27], ancilla qubit techniques [28], and extensions of neural network states [29], as well as the physical notions of "hidden Fermion" [30] and "hidden Fermi liquid" [31]. Given the close resemblance in derivation and algorithmic structure between gGA and multi-orbital GA, and the fact that the gGA framework includes the GA itself as a special case, we focus on the gGA framework for clarity and conciseness.

3.1 The gGA variational ansatz

Let us begin by introducing the structure of the variational ansatz used in gGA. We define a wavefunction $|\Psi_G\rangle$, which is obtained by applying an operator, indicated as $\hat{\mathcal{P}}_G$, to a reference single-particle wavefunction (Slater determinant) $|\Psi_0\rangle$

$$|\Psi_G\rangle = \hat{\mathcal{P}}_G|\Psi_0\rangle, \quad \text{with} \quad \hat{\mathcal{P}}_G = \prod_{i=1}^{\mathcal{N}} \hat{\mathcal{P}}_i\,, \tag{3}$$

Here, $|\Psi_0\rangle$ is the single-particle reference state, and $\hat{\mathcal{P}}_G$ is an operator composed of "local" operators $\hat{\mathcal{P}}_i$, whose precise mathematical structure will be described below.
Within our framework, the single-particle wavefunction $|\Psi_0\rangle$ is conceived in an auxiliary Hilbert space, while the operator $\hat{\mathcal{P}}_G$ maps states from the auxiliary space to the physical space. This is illustrated schematically in Fig. 1. Both $|\Psi_0\rangle$ and $\hat{\mathcal{P}}_i$ have to be optimized variationally, in order to minimize the variational energy

$$\mathcal{E}(\Psi_0, \hat{\mathcal{P}}_G) = \langle\Psi_G|\,\hat{H}\,|\Psi_G\rangle\,. \tag{4}$$

Specifically, we assume that the local operators $\hat{\mathcal{P}}_i$ (traditionally called "projectors" for historical reasons, but here not assumed to be projectors) have the following mathematical structure

$$\hat{\mathcal{P}}_i = \sum_{\Gamma=0}^{2^{\nu_i}-1} \sum_{n=0}^{2^{B\nu_i}-1} [\Lambda_i]_{\Gamma n}\,|\Gamma, i\rangle\langle n, i|, \tag{5}$$

$$|\Gamma, i\rangle = [c^\dagger_{i1}]^{q_1(\Gamma)} \ldots [c^\dagger_{iq_{\nu_i}}]^{q_{\nu_i}(\Gamma)}\,|0\rangle, \tag{6}$$

$$|n, i\rangle = [f^\dagger_{i1}]^{q_1(n)} \ldots [f^\dagger_{iq_{B\nu_i}}]^{q_{B\nu_i}(n)}\,|0\rangle\,. \tag{7}$$

In these equations we utilize the notation introduced in Sec. B.2, where $q_a(n)$ denotes the a-th occupation number of a Fock state $|n, i\rangle$, which is the a-th digit of the integer n in binary form. Additionally, the entries of the matrix Λ_i are variational parameters parametrizing $\hat{\mathcal{P}}_i$.

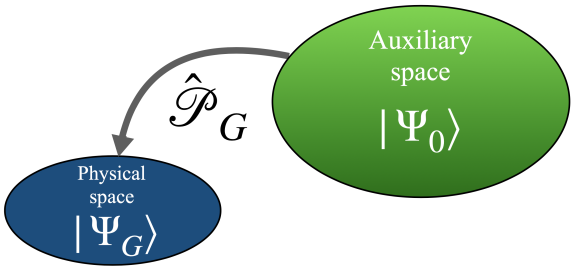


Fig. 1: *Diagrammatic representation of the gGA variational ansatz. The wavefunction $|\Psi_G\rangle$ is obtained by mapping a generic single-particle wavefunction $|\Psi_0\rangle$ through an operator $\hat{\mathcal{P}}_G$. Both $|\Psi_0\rangle$ and $\hat{\mathcal{P}}_G$ are optimized variationally.*

Remark. We point out that, in the traditional Gutzwiller Approximation (GA), the reference wavefunction $|\Psi_0\rangle$ is constructed directly within the physical Hilbert space, and the "Gutzwiller projector" $\hat{\mathcal{P}}_G$ in GA only operates within the physical space.
In contrast, in the gGA, the operator $\hat{\mathcal{P}}_G$ plays also the critical role of mapping the reference wavefunction $|\Psi_0\rangle$, which resides in the auxiliary space, onto a variational state $|\Psi_G\rangle$ residing in the physical space. Thus, the operator $\hat{\mathcal{P}}_G$ is inherently operating across both the auxiliary and physical spaces. It implements this mapping, while modulating the weights of the local electronic configurations, for optimizing the variational energy.

Remark. It is important to note that if the auxiliary space is configured as a mere replica of the physical space by setting $B = 1$ in Eqs. (5) and (7), gGA becomes mathematically equivalent to the traditional GA. In this specific case the resulting variational function $|\Psi_G\rangle$ would be as if $\hat{\mathcal{P}}_G$ and $|\Psi_0\rangle$ resided exclusively within the physical space. However, by setting B as an integer higher than 1 (generally an odd number for reasons that will be clarified later), the auxiliary space is enlarged, and gGA introduces additional degrees of freedom, thereby systematically enriching the variational ansatz. This gives the gGA the flexibility to explore the space of variational wavefunctions more comprehensively, establishing its primary distinction and advantage over the traditional GA.

While the remarks above are conceptually important, the gGA and the classic multi-orbital GA share essentially the same derivation and algorithmic structure. Consequently, for the sake of brevity and clarity, we will present the formalism directly in the context of the gGA framework.

3.2 Restriction to the "normal" variational states

We will focus on the normal phase, i.e., we will consider variational states $|\Psi_G\rangle$ that are eigenvectors of the physical number operator $\sum_{i\alpha} c^\dagger_{i\alpha} c_{i\alpha}$. It is important to note that, for $|\Psi_G\rangle$ to be an eigenstate of the number operator, it is not necessary that $\hat{\mathcal{P}}_i$ commutes with the number operator, as is the case in classic GA. Instead, it is sufficient that

- the auxiliary state $|\Psi_0\rangle$ is an eigenstate of the auxiliary-space number operator $\sum_{ia} f^\dagger_{ia} f_{ia}$,
- the coefficients $[\Lambda_i]_{\Gamma n}$ of $\hat{\mathcal{P}}_i$ satisfy the condition

$$\sum_{j=1}^{B\nu_i} q_j(n) - \sum_{j=1}^{\nu_i} q_j(\Gamma) = N(n) - N(\Gamma) = m_i \quad \forall\, \Gamma, n \,|\, [\Lambda_i]_{\Gamma n} \neq 0\,, \tag{8}$$

  where m_i is an integer.

This condition ensures that the $\hat{\mathcal{P}}_i$ operators, defined in Eqs. (5) and (7), map the auxiliary states $|n,i\rangle$ into physical states $|\Gamma,i\rangle$, with the number of physical fermions being reduced by an integer amount in each subsystem i. Consequently, the total number of physical fermions in $|\Psi_G\rangle$ is well defined and differs from the number of auxiliary fermions in $|\Psi_0\rangle$ by $\sum_{i=1}^{\mathcal{N}} m_i$.
In principle, m_i could be regarded as an additional variational parameters, to be optimized for minimizing the variational energy. However, previous work such as Ref. [11] showed that we can generally make the assumption that B is odd and

$$m_i = (B-1)\nu_i/2\,. \tag{9}$$

This particular choice for m_i, which reduces to $m_i = 0$ for $B = 1$ (consistent with the standard GA ansatz), has been empirically found to be the best variational choice. Therefore, in these notes we are going to make this assumption from now on.

3.3 Approximations for evaluating the variational energy

Our primary goal is to find the ground state of the Hamiltonian $\hat{H}$ by minimizing the variational energy. The variational energy $\mathcal{E}$ is a function of the wavefunction $|\Psi_0\rangle$ and the operator $\hat{\mathcal{P}}_G$, and is given by

$$\mathcal{E}(\Psi_0, \hat{\mathcal{P}}_G) = \langle\Psi_0|\hat{\mathcal{P}}_G^\dagger \hat{H}\hat{\mathcal{P}}_G|\Psi_0\rangle\,. \tag{10}$$

However, evaluating the variational energy in Eq. (10) is a highly non-trivial task due to the complexity associated with many-body interactions and the vastness of the Hilbert space. Without making approximations, numerical approaches such as Variational Monte Carlo could be used, but are computationally demanding.
To simplify the problem, let us first recognize that since $|\Psi_0\rangle$ is a single-particle wavefunction, Wick's theorem can be applied. According to Wick's theorem, the expectation value of a product of creation and annihilation operators can be decomposed into a sum of products of expectation values of pairs of operators (see appendix D for details). However, even with Wick's theorem, the number of possible contractions contributing to the expectation value of $\hat{H}$ in $\langle\Psi_0|\hat{\mathcal{P}}_G^\dagger \hat{H}\hat{\mathcal{P}}_G|\Psi_0\rangle$ is prohibitively large.

3.3.1 The Gutzwiller approximation and the Gutzwiller constraints

In light of the above, we must employ approximations that simplify the problem while still capturing the essential physics. Here, we introduce two main approximations:

1. **The Gutzwiller Constraints:** These constraints are limitations imposed on the variational wave function $|\Psi_G\rangle$, linking the variational parameters of $\hat{\mathcal{P}}_i$ and $|\Psi_0\rangle$

$$\langle\Psi_0|\,\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i\,|\Psi_0\rangle = \langle\Psi_0|\Psi_0\rangle = 1 \tag{11}$$

$$\langle\Psi_0|\,\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i\; f_{ia}^\dagger f_{ib}\,|\Psi_0\rangle = \langle\Psi_0|\,f_{ia}^\dagger f_{ib}\,|\Psi_0\rangle \qquad \forall\, a,b = 1,...,B\nu_i\,. \tag{12}$$

   These constraints slightly reduce the variational freedom, but, as we are going to see, they make the evaluation of the variational energy more manageable.

2. **The Gutzwiller Approximation:** The approximation consists in neglecting some of the Wick contractions arising when evaluating Eq. (10), based on the key observation that such terms would vanish in the particular limit of infinite coordination number. This approximation, that we are going to specify in detail below, constitutes a key connection with Dynamical Mean-Field Theory (DMFT), which is exact in this limit.

Let us now proceed to analyze how these approximations combined help in evaluating the variational energy efficiently.

3.3.2 Key consequence of the Gutzwiller constraints

For this purpose, it is essential to consider tthe left-hand side of Eq. (12)

$$\langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle \, , \tag{13}$$

using Theorem D.2 discussed in Sec. D.1 of the Appendix.

- In this context, we can treat the operator $\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i$, which resides entirely in the auxiliary space and is therefore some algebraic combination of the f modes, as the operator X of Theorem D.2. Following this theorem, we can write

$$\langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle = \langle \Psi_0 | \, \overbrace{\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i} \; \overbrace{f_{ia}^\dagger f_{ib}} \, | \Psi_0 \rangle + \langle \Psi_0 | \, \overbrace{\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib}} \, | \Psi_0 \rangle \, . \tag{14}$$

  This identity separates the expression into a disconnected term, where $f_{ia}^\dagger$ is contracted with f_{ib} and $\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i$ is contracted with itself, and a connected term, where $f_{ia}^\dagger$ and f_{ib} are both contracted with $\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i$.

- We can further simplify the first term on the right side of Eq. (14). Using the first Gutzwiller constraint [Eq. (11)], where it is specified that $\langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \, | \Psi_0 \rangle = 1$, the first term can be simplified as

$$\begin{aligned} \langle \Psi_0 | \, \overbrace{\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i} \; \overbrace{f_{ia}^\dagger f_{ib}} \, | \Psi_0 \rangle &= \langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \, | \Psi_0 \rangle \langle \Psi_0 | \, f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle \\ &= 1 \cdot \langle \Psi_0 | \, f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle = \langle \Psi_0 | \, f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle \, . \end{aligned} \tag{15}$$

  By substituting it in Eq. (14) we obtain

$$\langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle = \langle \Psi_0 | \, f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle + \langle \Psi_0 | \, \overbrace{\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib}} \, | \Psi_0 \rangle \tag{16}$$

- Comparing Eq. (16) with the second Gutzwiller constraint in Eq. (12)

$$\langle \Psi_0 | \, \hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle = \langle \Psi_0 | \, f_{ia}^\dagger f_{ib} \, | \Psi_0 \rangle \, , \tag{17}$$

  it follows that

$$\langle \Psi_0 | \, \overbrace{\hat{\mathcal{P}}_i^\dagger \hat{\mathcal{P}}_i \; f_{ia}^\dagger f_{ib}} \, | \Psi_0 \rangle = 0 \, . \tag{18}$$

- Applying the second part of Theorem D.2 to the left side of Eq. (18), we can express the sum of all connected terms in our expression as

$$\langle\Psi_0|\,\overbrace{\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i\;f_{ia}^\dagger f_{ib}}\,|\Psi_0\rangle = \sum_{a'b'}\xi_i^{a'b'}\langle\Psi_0|\,f_{ia'}f_{ia}^\dagger\,|\Psi_0\rangle\langle\Psi_0|\,f_{ib'}^\dagger f_{ib}\,|\Psi_0\rangle = 0\,, \tag{19}$$

  where the coefficients $\xi_i^{a'b'}$ depend only on $|\Psi_0\rangle$ and the operator $\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i$, but not on the indices a and b.

- Let us analyze the implications of Eq. (19). Defining the matrix Δ_i as

$$[\Delta_i]_{ab} = \langle\Psi_0|\,f_{ia}^\dagger f_{ib}\,|\Psi_0\rangle\,, \tag{20}$$

  we can rewrite Eq (19) as

$$\big(\mathbf{1}-\Delta_i\big)\xi_i\Delta_i = 0\,, \tag{21}$$

  from which it follows that, as long as neither Δ_i nor $\mathbf{1}-\Delta_i$ are degenerate, i.e., as long as none of the eigenvalues of Δ_i is equal to 0 or 1, we have

$$\xi_i = 0\,. \tag{22}$$

The significance of Eq. (22), which arises from the Gutzwiller constraints in Eqs. (11) and (12), is pivotal for streamlining the computation of the variational energy. This is due to its bearing on terms involving two contraction lines connecting $\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i$ with an arbitrary operator $\hat{X}$ built from algebraic combinations of the f and $f^\dagger$ modes.

Theorem 3.1. *Consider evaluating the expectation value $\langle\Psi_0|\,\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i\;\hat{X}\,|\Psi_0\rangle$. In this scenario, all terms comprising two contraction lines connecting $\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i$ with $\hat{X}$ do not contribute.*
More formally, if we depict the two contraction lines emerging from $\hat{\mathcal{P}}_i^\dagger$ and $\hat{\mathcal{P}}_i$, and linking them to $\hat{X}$, we can establish that

$$\langle\Psi_0|\overbrace{\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i\;\hat{X}}\,|\Psi_0\rangle = \sum_{a'b'}\xi_i^{a'b'}\langle\Psi_0|f_{ia'}f_{ia}^\dagger|\Psi_0\rangle\langle\Psi_0|f_{ib'}^\dagger f_{ib}|\Psi_0\rangle x^{ab} = 0\,. \tag{23}$$

Proof. The proof hinges on an observation we made in Sec. D.1. There, we noted that the coefficients $\xi_i^{a'b'}$ in Eq. (22) represent the sum of all terms stemming from self-contractions among the operators in $\hat{\mathcal{P}}_i^\dagger\hat{\mathcal{P}}_i$ after excluding a' and b', which are contracted with a and b respectively. Meanwhile, the coefficients x^{ab} signify the sum of all terms that originate from self-contractions among the operators in $\hat{X}$, excluding a and b which are contracted with a' and b' respectively.
Importantly, the coefficients $\xi_i^{a'b'}$ are only dependent on $|\Psi_0\rangle$, a' and b', and are unaffected by $\hat{X}$. As such, they are equivalent to the ones in Eq. (19), which we have proven to be zero due to the Gutzwiller constraints (see Eq. (22)). This completes the proof. □

3.3.3 The Gutzwiller approximation: Explicit definition

The Gutzwiller approximation is a key simplification used in conjunction with the Gutzwiller constraints, and it plays an essential role in streamlining the computation of the expectation value of the Hamiltonian with respect to the gGA variational states.
Specifically, the approximation involves neglecting terms in the expectation values with respect to $|\Psi_0\rangle$ that contain more than two non-local contractions. In this context, a non-local contraction refers to one that involves operators acting on different sites, such as $\langle\Psi_0| f^\dagger_{ia} f_{jb} |\Psi_0\rangle$, where $i \neq j$. The rationale behind this approximation is rooted in the behavior of these terms in the limit of infinite coordination number, where each site is infinitely connected. In this limit, all terms with more than two non-local contractions vanish. This aspect is particularly significant as dynamical mean-field theory (DMFT) becomes exact in this limit, establishing a meaningful link between the Gutzwiller approximation and DMFT.
In subsequent sections, we will delve into how the synergistic application of the Gutzwiller approximation and the Gutzwiller constraints considerably simplifies the evaluation of expectation values in our gGA variational framework.

3.4 Evaluation of the local expectation values

In this section, we will focus on one of the essential steps in the ghost Gutzwiller approximation (gGA) — the computation of the expectation value of local Hamiltonian terms with respect to the gGA wavefunction, $|\Psi_G\rangle$. To this end, let us consider a local Hamiltonian term, $\hat{H}^i_{\text{loc}}$, acting on fragment i as defined in Eq. (1). The expectation value of this term with respect to $|\Psi_G\rangle$ can be expressed as

$$\langle\Psi_G|\hat{H}^i_{\text{loc}}|\Psi_G\rangle = \langle\Psi_0|\left(\prod_{k=1}^{\mathcal{N}}\hat{\mathcal{P}}^\dagger_k\right)\hat{H}^i_{\text{loc}}\left(\prod_{k=1}^{\mathcal{N}}\hat{\mathcal{P}}_k\right)|\Psi_0\rangle = \langle\Psi_0|\prod_{k\neq i}\left(\hat{\mathcal{P}}^\dagger_k\hat{\mathcal{P}}_k\right)\left(\hat{\mathcal{P}}^\dagger_i\hat{H}^i_{\text{loc}}\hat{\mathcal{P}}_i\right)|\Psi_0\rangle\,. \tag{24}$$

At first glance, this expression appears challenging to evaluate without approximations due to the enormous number of possible Wick contractions.
Fortunately, the Gutzwiller constraints, as outlined in Eqs. (11) and (12), and the Gutzwiller approximation discussed previously, offer significant simplifications.
Consider a block $\hat{\mathcal{P}}^\dagger_k\hat{\mathcal{P}}_k$ within the product in Eq. (24). Wick's theorem allows us to categorize all possible terms contributing to this expectation value as follows:

- **Disconnected terms**: These are terms where there are no contraction lines between $\hat{\mathcal{P}}^\dagger_k\hat{\mathcal{P}}_k$ and the other operators. Utilizing the first Gutzwiller constraint, Eq. (11), the contribution of these terms is

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_k\hat{\mathcal{P}}_k|\Psi_0\rangle\langle\Psi_0|\prod_{k'\neq i,k}\left(\hat{\mathcal{P}}^\dagger_{k'}\hat{\mathcal{P}}_{k'}\right)\left(\hat{\mathcal{P}}^\dagger_i\hat{H}^i_{\text{loc}}\hat{\mathcal{P}}_i\right)|\Psi_0\rangle = \langle\Psi_0|\prod_{k'\neq i,k}\left(\hat{\mathcal{P}}^\dagger_{k'}\hat{\mathcal{P}}_{k'}\right)\left(\hat{\mathcal{P}}^\dagger_i\hat{H}^i_{\text{loc}}\hat{\mathcal{P}}_i\right)|\Psi_0\rangle\,. \tag{25}$$

- **Terms with two contraction lines**: These terms have two contraction lines between $\hat{\mathcal{P}}_k^\dagger \hat{\mathcal{P}}_k$ and the other operators. As demonstrated in previous sections, the contribution of these terms is zero due to the Gutzwiller constraints.

- **Terms with four contraction lines**: These terms have four contraction lines between $\hat{\mathcal{P}}_k^\dagger \hat{\mathcal{P}}_k$ and the other operators. Due to the Gutzwiller approximation, these non-local contractions are effectively zero.

We can now understand that the logic discussed in the itemized list above applies to each block $\hat{\mathcal{P}}_k^\dagger \hat{\mathcal{P}}_k$ independently. We can iteratively apply it to each block, effectively eliminating all of them one by one. This iterative process simplifies the expression until we are only left with the block corresponding to fragment i.
Combining these observations, we arrive at a significant simplification of Eq. (24)

$$\langle \Psi_G | \hat{H}^i_{\rm loc} | \Psi_G \rangle \approx \langle \Psi_0 | \hat{\mathcal{P}}_i^\dagger \hat{H}^i_{\rm loc} \hat{\mathcal{P}}_i | \Psi_0 \rangle \, . \tag{26}$$

This expression vastly reduces the complexity involved in computing the expectation values of local Hamiltonian terms, making it feasible for practical implementations.

3.5 Evaluation of the one-body non-local expectation values

In this section, we will extend the methodology discussed in the previous section for local operators to calculate the expectation values of one-body non-local operators within the ghost Gutzwiller approximation (gGA). Specifically, we will consider operators of the form $c^\dagger_{i\alpha} c_{j\beta}$, where i and j are fragment labels and α and β are additional quantum numbers, such as spin. We are interested in computing the expectation value of this operator with respect to the gGA wavefunction $|\Psi_G\rangle$.
The expectation value of this one-body non-local operator can be written as

$$\begin{aligned}
\langle \Psi_G | c^\dagger_{i\alpha} c_{j\beta} | \Psi_G \rangle &= \langle \Psi_0 | \left(\prod_{k=1}^{\mathcal{N}} \hat{\mathcal{P}}_k^\dagger \right) c^\dagger_{i\alpha} c_{j\beta} \left(\prod_{k=1}^{\mathcal{N}} \hat{\mathcal{P}}_k \right) | \Psi_0 \rangle \\
&= \langle \Psi_0 | \left(\prod_{k \neq i,j} \hat{\mathcal{P}}_k^\dagger \hat{\mathcal{P}}_k \right) \left(\hat{\mathcal{P}}_i^\dagger c^\dagger_{i\alpha} \hat{\mathcal{P}}_i \right) \left(\hat{\mathcal{P}}_j^\dagger c_{j\beta} \hat{\mathcal{P}}_j \right) | \Psi_0 \rangle \, .
\end{aligned} \tag{27}$$

Similar to the treatment in the previous section, we have grouped the terms with $k \neq i, j$ together. The additional complexity here arises from the fact that the non-local operator involves two different fragments, i and j.
We can employ the same considerations as in the previous section to simplify the expression in Eq. (27). Just as before, the Gutzwiller constraints and approximations allow us to iteratively eliminate all the $\hat{\mathcal{P}}_k^\dagger \hat{\mathcal{P}}_k$ blocks for $k \neq i, j$. In a manner analogous to the local terms, disconnected terms, terms with two contraction lines, and terms with four contraction lines can be handled exactly as was done for the local Hamiltonian terms, resulting in

$$\langle \Psi_G | c^\dagger_{i\alpha} c_{j\beta} | \Psi_G \rangle \approx \langle \Psi_0 | \left(\hat{\mathcal{P}}_i^\dagger c^\dagger_{i\alpha} \hat{\mathcal{P}}_i \right) \left(\hat{\mathcal{P}}_j^\dagger c_{j\beta} \hat{\mathcal{P}}_j \right) | \Psi_0 \rangle \, . \tag{28}$$

The expression in Eq. (28) can be further simplified by making several key observations:

- **Grouping terms by the number of Wick contractions**: We can organize the terms in Eq. (28) by grouping them according to the number of Wick contractions between the blocks i and j. Specifically, we classify terms by collecting together all terms with only one Wick contraction between the blocks i and j, then those with three, five, and so on.

- **Neglecting terms with more than one Wick contraction**: Due to the Gutzwiller approximation, terms with three or more Wick contraction between blocks i and j are neglected. Therefore, only the terms with a single Wick contraction need to be considered.

- **Auxiliary space representation**: It is important to recognize that $\hat{\mathcal{P}}_i^\dagger c_{i\alpha}^\dagger \hat{\mathcal{P}}_i$ operates entirely within the auxiliary space, so it can be represented as an algebraic combination of f_{ia} and $f_{ia}^\dagger$ modes. This applies analogously to the j block with f_{jb} and $f_{ja}^\dagger$ modes.

- **Matrix representation for self-contractions**: We can factor the contribution of all self-contractions within $\hat{\mathcal{P}}_i^\dagger c_{i\alpha}^\dagger \hat{\mathcal{P}}_i$ that are left after having contracted $f_{ia}^\dagger$ with an annihilation operator belonging to the j subsystem, and encode it into a $B\nu_i \times \nu_i$ matrix $[\mathcal{R}_i]_{a\alpha}$ as

$$\langle \Psi_0 | \left(\hat{\mathcal{P}}_i^\dagger c_{i\alpha}^\dagger \hat{\mathcal{P}}_i \right) \left(\hat{\mathcal{P}}_j^\dagger c_{j\beta} \hat{\mathcal{P}}_j \right) | \Psi_0 \rangle = \sum_{a=1}^{B\nu_i} \sum_{b=1}^{B\nu_j} \langle \Psi_0 | \left([\mathcal{R}_i]_{a\alpha} f_{ia}^\dagger \right) \left([\mathcal{R}_j]_{\beta b}^\dagger f_{jb} \right) | \Psi_0 \rangle \,, \tag{29}$$

  for all $i \neq j$.

- **Computing local expectation values using coefficient matrices**: The same coefficient matrices $\mathcal{R}_i$ would arise from self-contractions also in the following expression

$$\langle \Psi_0 | \hat{\mathcal{P}}_i^\dagger c_{i\alpha}^\dagger \hat{\mathcal{P}}_i f_{ia} | \Psi_0 \rangle = \sum_{b=1}^{B\nu_i} [\mathcal{R}_i]_{b\alpha} \langle \Psi_0 | f_{ib}^\dagger f_{ia} | \Psi_0 \rangle \,. \tag{30}$$

As we are going to see, the steps above can be used for calculating $\mathcal{R}_i$, therefore facilitating the computation of Eq. (29).

3.6 Recap: Evaluation of the variational energy

Let us take a moment to recapitulate the key developments in our evaluation of the variational energy within the ghost Gutzwiller approximation (gGA), employing the Gutzwiller constraints and the Gutzwiller approximation.

1. **Expectation value of local operators**: In Sec. 3.4 we derived the equation

$$\langle \Psi_G | \hat{H}_{\text{loc}}^i | \Psi_G \rangle \approx \langle \Psi_0 | \hat{\mathcal{P}}_i^\dagger \hat{H}_{\text{loc}}^i \hat{\mathcal{P}}_i | \Psi_0 \rangle \,. \tag{31}$$

2. **Expectation value of non-local one-body operators**: In Sec. 3.5 we derived

$$\langle\Psi_G|c^\dagger_{i\alpha}c_{j\beta}|\Psi_G\rangle \approx \sum_{a=1}^{B\nu_i}\sum_{b=1}^{B\nu_j}\langle\Psi_0|\left([\mathcal{R}_i]_{a\alpha}f^\dagger_{ia}\right)\left([\mathcal{R}_j]^\dagger_{\beta b}f_{jb}\right)|\Psi_0\rangle\,, \tag{32}$$

where we introduced the $\mathcal{R}_i$, characterized by

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i c^\dagger_{i\alpha}\hat{\mathcal{P}}_i f_{ia}|\Psi_0\rangle = \sum_{b=1}^{B\nu_i}[\mathcal{R}_i]_{b\alpha}\langle\Psi_0|f^\dagger_{ib}f_{ia}|\Psi_0\rangle\,. \tag{33}$$

Summing up the above contributions, the variational energy $\mathcal{E}$ can be expressed as

$$\mathcal{E} = \sum_{i,j=1}^{\mathcal{N}}\sum_{a,b=1}^{B\nu_i}\left[\mathcal{R}_i t_{ij}\mathcal{R}^\dagger_j\right]_{ab}\langle\Psi_0|\,f^\dagger_{ia}f_{jb}\,|\Psi_0\rangle + \sum_{i=1}^{\mathcal{N}}\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i\,\hat{H}^i_{loc}[c^\dagger_{i\alpha},c_{i\alpha}]\,\hat{\mathcal{P}}_i\,|\Psi_0\rangle\,,$$

where the matrices $\mathcal{R}_i$ are determined by Eq. (33).
This energy must be minimized subject to the fulfillment of the Gutzwiller constraints

$$\begin{aligned}\langle\Psi_0|\,\hat{\mathcal{P}}^\dagger_i\hat{\mathcal{P}}_i\,|\Psi_0\rangle &= \langle\Psi_0|\Psi_0\rangle = 1\,,\\ \langle\Psi_0|\,\hat{\mathcal{P}}^\dagger_i\hat{\mathcal{P}}_i\;f^\dagger_{ia}f_{ib}\,|\Psi_0\rangle &= \langle\Psi_0|\,f^\dagger_{ia}f_{ib}\,|\Psi_0\rangle \qquad \forall\, a,b=1,...,B\nu_i\,.\end{aligned}$$

In essence, we have reduced the problem to computing expectation values with respect to $|\Psi_0\rangle$ of local operators: those appearing in the Gutzwiller constraints, the terms of the local interactions, and the terms entering in the characterization of the system for computing the $\mathcal{R}_i$ matrices. We will learn how to calculate these terms systematically and efficiently in the following sections.

4 Reformulation using local reduced density-matrix

In order to compute all the expectation values of local observables with respect to $|\Psi_0\rangle$, which appear in our variational problem as summarized above in Sec. 3.6, we need to introduce the local reduced density-matrix of the i-th auxiliary-space subsystem i. According to Theorem E.1 from the appendix (following Refs. [6, 20, 11]), the local reduced density-matrix is given by

$$\hat{P}^0_i \propto \exp\left(-\sum_{a,b=1}^{B\nu_i}\left[\ln\left(\frac{\mathbf{1}-\Delta^T_i}{\Delta^T_i}\right)\right]_{ab}f^\dagger_{ia}f_{ib}\right), \tag{34}$$

where Δ_i is the $B\nu_i\times B\nu_i$ matrix with elements:

$$[\Delta_i]_{ab} = \langle\Psi_0|f^\dagger_{ia}f_{ib}|\Psi_0\rangle. \tag{35}$$

Utilizing these definitions and the matrix representation of the Gutzwiller projector in Eq. (5), it is straightforward to derive the relations

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i\hat{\mathcal{P}}_i\,|\Psi_0\rangle = \mathrm{Tr}\left[P^0_i\,\Lambda^\dagger_i\Lambda_i\right], \tag{36}$$

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i\hat{\mathcal{P}}_i\;f^\dagger_{ia}f_{ib}|\Psi_0\rangle = \mathrm{Tr}\left[P^0_i\,\Lambda^\dagger_i\Lambda_i\,\tilde{F}^\dagger_{ia}\tilde{F}_{ib}\right], \tag{37}$$

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i\,\hat{H}^i_{\rm loc}[c^\dagger_{i\alpha},c_{i\alpha}]\,\hat{\mathcal{P}}_i\,|\Psi_0\rangle = \mathrm{Tr}\left[P^0_i\,\Lambda^\dagger_i\,\hat{H}_{\rm loc}{}^i[F^\dagger_{i\alpha},F_{i\alpha}]\Lambda_i\right], \tag{38}$$

$$\langle\Psi_0|\hat{\mathcal{P}}^\dagger_i c^\dagger_{i\alpha}\hat{\mathcal{P}}_i\,f_{ia}|\Psi_0\rangle = \mathrm{Tr}\left[P^0_i\,\Lambda^\dagger_i\,F^\dagger_{i\alpha}\,\Lambda_i\,\tilde{F}_{ia}\right], \tag{39}$$

where Tr denotes the trace operator restricted within the $2^{B\nu_i}$-dimensional many-body Fock space of the i-th auxiliary-space subsystem. The matrices $F_{i\alpha}$, and $\tilde{F}_{ia}$ are representations of the local reduced density-matrix, the physical annihilation operators and the auxiliary annihilation operators in their own Fock basis, respectively, which are defined as

$$[F_{i\alpha}]_{\Gamma\Gamma'} = \langle \Gamma, i | c_{i\alpha} | \Gamma', i \rangle \qquad \big(\Gamma, \Gamma' \in \{0, \ldots, 2^{\nu_i} - 1\}\big), \tag{40}$$

$$[\tilde{F}_{ia}]_{nn'} = \langle n, i | f_{ia} | n', i \rangle \qquad \big(n, n' \in \{0, \ldots, 2^{B\nu_i} - 1\}\big), \tag{41}$$

while the matrix representation of $\hat{P}_i^0$ with entries $[P_i^0]_{nn'} = \langle n, i | \hat{P}_i^0 | n', i \rangle$ is given by

$$P_i^0 \propto \exp\left(- \sum_{a,b=1}^{B\nu_i} \left[\ln\left(\frac{\mathbf{1} - \Delta_i^T}{\Delta_i^T} \right) \right]_{ab} \tilde{F}_{ia}^\dagger \tilde{F}_{ib} \right). \tag{42}$$

Summing up the above contributions, with the local terms expressed in terms of the reduced density-matrix, the variational energy $\mathcal{E}$ can be expressed as

$$\mathcal{E} = \sum_{i,j=1}^{\mathcal{N}} \sum_{a,b=1}^{B\nu_i} \left[\mathcal{R}_i t_{ij} \mathcal{R}_j^\dagger \right]_{ab} \langle \Psi_0 | \, f_{ia}^\dagger f_{jb} \, | \Psi_0 \rangle + \sum_{i=1}^{\mathcal{N}} \mathrm{Tr} \left[P_i^0 \, \Lambda_i^\dagger \, \hat{H}_{\mathrm{loc}}^i [F_{i\alpha}^\dagger, F_{i\alpha}] \, \Lambda_i \right], \tag{43}$$

where the matrices $\mathcal{R}_i$ are determined by the relation

$$\mathrm{Tr} \left[P_i^0 \Lambda_i^\dagger F_{i\alpha}^\dagger \Lambda_i \tilde{F}_{ia} \right] = \sum_{b=1}^{B\nu_i} [\mathcal{R}_i]_{b\alpha} \langle \Psi_0 | f_{ib}^\dagger f_{ia} | \Psi_0 \rangle = \sum_{b=1}^{B\nu_i} [\mathcal{R}_i]_{b\alpha} [\Delta_i]_{ba} \, . \tag{44}$$

This energy must be minimized subject to the fulfillment of the Gutzwiller constraints, which in terms of the reduced density-matrix are

$$\mathrm{Tr} \left[P_i^0 \Lambda_i^\dagger \Lambda_i \right] = \langle \Psi_0 | \Psi_0 \rangle = 1 \, , \tag{45}$$

$$\mathrm{Tr} \left[P_i^0 \Lambda_i^\dagger \Lambda_i \tilde{F}_{ia}^\dagger \tilde{F}_{ib} \right] = \langle \Psi_0 | f_{ia}^\dagger f_{ib} | \Psi_0 \rangle = [\Delta_i]_{ab} \qquad \forall \, a, b = 1, ..., B\nu_i \, . \tag{46}$$

5 Reformulation in terms of slave-boson amplitudes (connection with RISB)

In this section, we delve into the concept of "slave-boson amplitudes" through the lens of the gGA. Within the Rotationally-Invariant Slave-Boson (RISB) approach [15, 19, 20], slave-boson amplitudes emerge from a distinct perspective, where auxiliary bosons are introduced to represent local modes within each system fragment. A remarkable aspect about this alternative perspective is that the gGA can be viewed as the mean-field approximation of the ghost RISB (gRISB) [32], and analogously, the Gutzwiller approximation (GA) bears a similar relationship to RISB [33, 6, 20]. As such, the RISB/gRISB formulation opens avenues for devising practical implementations that systematically incorporate quantum-fluctuation corrections toward obtaining the exact solution. Although we will not embark on a detailed derivation of RISB or gRISB within this section, we encourage readers to peruse the referenced literature for a more comprehensive understanding of this connection and the exciting possibilities it harbors for theoretical and algorithmic advancements in the many-body problem.

5.1 The slave-boson amplitudes

We can rewrite all the key local quantities in an alternative way in terms of the so-called matrices of slave-boson amplitudes. This is not only useful for technical purposes but also for establishing a formal connection with the rotationally-invariant slave-boson theory (RISB).
Let us introduce the matrix of slave-boson amplitudes, ϕ_i, as defined in Refs. [6, 33, 7, 11]

$$\phi_i = \Lambda_i \sqrt{P_i^0} = \Lambda_i [P_i^0]^{\frac{1}{2}} \,. \tag{47}$$

Substituting this into Eqs. (36)-(39) yields the following equations

$$\mathrm{Tr}\left[P_i^0 \Lambda_i^\dagger \Lambda_i\right] = \mathrm{Tr}\left[\phi_i^\dagger \phi_i\right] \,, \tag{48}$$

$$\mathrm{Tr}\left[P_i^0 \Lambda_i^\dagger \Lambda_i \tilde{F}_{ia}^\dagger \tilde{F}_{ib}\right] = \mathrm{Tr}\left[\phi_i^\dagger \phi_i [P_i^0]^{-\frac{1}{2}} \tilde{F}_{ia}^\dagger \tilde{F}_{ib} [P_i^0]^{\frac{1}{2}}\right] \,, \tag{49}$$

$$\mathrm{Tr}\left[P_i^0 \Lambda_i^\dagger \hat{H}^i_{\mathrm{loc}}[F_{i\alpha}^\dagger, F_{i\alpha}] \Lambda_i\right] = \mathrm{Tr}\left[\phi_i \phi_i^\dagger \hat{H}^i_{\mathrm{loc}}[F_{i\alpha}^\dagger, F_{i\alpha}]\right] \,, \tag{50}$$

$$\mathrm{Tr}\left[P_i^0 \Lambda_i^\dagger F_{i\alpha}^\dagger \Lambda_i \tilde{F}_{ia}\right] = \mathrm{Tr}\left[\phi_i^\dagger F_{i\alpha}^\dagger \phi_i [P_i^0]^{-\frac{1}{2}} \tilde{F}_{ia} [P_i^0]^{\frac{1}{2}}\right] \,. \tag{51}$$

We note that Eqs. (49) and (51) involve a similarity transformations of the matrix representations of the auxiliary-mode operators, such as those described in the Appendix, see Sec. B.4.2.
By applying theorems B.2 and B.3, we obtain

$$\begin{aligned}
[P_i^0]^{-\frac{1}{2}} \tilde{F}_{ia}^\dagger [P_i^0]^{\frac{1}{2}} &= e^{\frac{1}{2}\sum_{a,b=1}^{B\nu_i} \left[\ln\left(\frac{\mathbf{1}-\Delta_i^T}{\Delta_i^T}\right)\right]_{a'b'} \tilde{F}_{ia'}^\dagger \tilde{F}_{ib'}} \, \tilde{F}_{ia}^\dagger \, e^{-\frac{1}{2}\sum_{a'',b''=1}^{B\nu_i} \left[\ln\left(\frac{\mathbf{1}-\Delta_i^T}{\Delta_i^T}\right)\right]_{a''b''} \tilde{F}_{ia''}^\dagger \tilde{F}_{ib''}} \\
&= \sum_{a'=1}^{B\nu_i} \left[e^{\frac{1}{2}\ln\left(\frac{\mathbf{1}-\Delta_i^T}{\Delta_i^T}\right)}\right]_{a'a} \tilde{F}_{ia'}^\dagger = \sum_{a'=1}^{B\nu_i} \left[\frac{\mathbf{1}-\Delta_i^T}{\Delta_i^T}\right]_{a'a}^{\frac{1}{2}} \tilde{F}_{ia'}^\dagger = \sum_{a'=1}^{B\nu_i} \left[\frac{\mathbf{1}-\Delta_i}{\Delta_i}\right]_{aa'}^{\frac{1}{2}} \tilde{F}_{ia'}^\dagger \,,
\end{aligned} \tag{52}$$

$$[P_i^0]^{-\frac{1}{2}} \tilde{F}_{ia} [P_i^0]^{\frac{1}{2}} = \sum_{a'=1}^{B\nu_i} \left[e^{-\frac{1}{2}\ln\left(\frac{\mathbf{1}-\Delta_i^T}{\Delta_i^T}\right)}\right]_{aa'} \tilde{F}_{ia'} = \sum_{a'=1}^{B\nu_i} \left[\frac{\Delta_i^T}{\mathbf{1}-\Delta_i^T}\right]_{aa'}^{\frac{1}{2}} \tilde{F}_{ia'} = \sum_{a'=1}^{B\nu_i} \left[\frac{\Delta_i}{\mathbf{1}-\Delta_i}\right]_{a'a}^{\frac{1}{2}} \tilde{F}_{ia'} \,. \tag{53}$$

- By substituting Eq. (48) in Eq. (45) and Eq. (52) in Eq. (46) we obtain

$$\mathrm{Tr}\left[\phi_i^\dagger \phi_i\right] = \langle \Psi_0 | \Psi_0 \rangle = 1 \tag{54}$$

$$\mathrm{Tr}\left[\phi_i^\dagger \phi_i \tilde{F}_{ia}^\dagger \tilde{F}_{ib}\right] = \langle \Psi_0 | f_{ia}^\dagger f_{ib} | \Psi_0 \rangle = [\Delta_i]_{ab} \qquad \forall\, a, b = 1, \ldots, B\nu_i \,. \tag{55}$$

- By substituting Eq. (53) into Eq. (44) we obtain the following equation for $\mathcal{R}_i$

$$\mathrm{Tr}\left[\phi_i^\dagger F_{i\alpha}^\dagger \phi_i \tilde{F}_{ia}\right] = \sum_{c=1}^{B\nu_i} [\mathcal{R}_i]_{c\alpha} \, [\Delta_i(\mathbf{1}-\Delta_i)]_{ca}^{\frac{1}{2}} \,, \tag{56}$$

  which can always be inverted, as long as neither Δ_i nor $\mathbf{1}-\Delta_i$ are degenerate, which we already assumed after Eq. (21), to prove Eq. (22).

5.2 Recap: the variational problem in terms of slave-boson amplitudes

Summing up the above contributions, with the local terms expressed in terms of the slave-boson amplitudes, the variational energy $\mathcal{E}$ can be expressed as

$$\mathcal{E} = \sum_{i,j=1}^{\mathcal{N}} \sum_{a,b=1}^{B\nu_i} \left[\mathcal{R}_i t_{ij} \mathcal{R}_j^\dagger\right]_{ab} \langle \Psi_0 | \, f_{ia}^\dagger f_{jb} \, | \Psi_0 \rangle + \sum_{i=1}^{\mathcal{N}} \mathrm{Tr}\left[\phi_i \phi_i^\dagger \hat{H}_{\mathrm{loc}}^i [F_{i\alpha}^\dagger, F_{i\alpha}]\right] , \tag{57}$$

where the matrices $\mathcal{R}_i$ are determined by Eq. (56). It must be minimized subject to the Gutzwiller constraints, which in terms of the slave-boson amplitudes are given by Eqs. (54) and (55).

6 Reformulation in terms of embedding states (connection with DMET)

In this section, building on the foundation laid in the preceding section, we take a further step by expressing key local quantities in terms of "embedding states". This concept was first introduced in Ref. [7] within the context of the multi-orbital GA and later extended and further developed for the gGA in Refs. [10, 11]. The mapping presented in this section is computationally advantageous, as it makes it possible to reformulate the energy-optimization problem into a recursive computation of the ground state of an auxiliary "impurity model" with a finite bath. Furthermore, this perspective plays a critical role in bridging the gGA with quantum embedding theories such as DMET [34, 26], enabling a more unified understanding of these frameworks [35].

6.1 The embedding states

The embedding states are vectors belonging to an auxiliary Fock space and serve to map the slave-boson amplitudes ϕ_i to a fermionic impurity Hamiltonian.

Definition 6.1 (Embedding States)**.** The embedding states, denoted as $|\Phi_i\rangle$, are defined as

$$|\Phi_i\rangle = \sum_{\Gamma=0}^{2^{\nu_i}-1} \sum_{n=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}N(n)(N(n)-1)} [\phi_i]_{\Gamma n} |\Gamma; i\rangle \otimes U_{\mathrm{PH}} |n; i\rangle \, , \tag{58}$$

where:

- $|\Gamma; i\rangle$ and $|n; i\rangle$ are Fock states generated by auxiliary fermionic modes

$$|\Gamma; i\rangle = [c_{i1}^\dagger]^{q_1(\Gamma)} \dots [c_{iB\nu_i}^\dagger]^{q_{\nu_i}(\Gamma)} \, |0\rangle \, , \tag{59}$$

$$|n; i\rangle = [b_{i1}^\dagger]^{q_1(n)} \dots [b_{iB\nu_i}^\dagger]^{q_{B\nu_i}(n)} |0\rangle \, . \tag{60}$$

- Consistently with the notation introduced in Sec. B.2 (also used above in Sec. 3), $q_a(n)$ denotes the a-th occupation number of a Fock state $|n, i\rangle$, which is the i-th digit of the integer n in binary form, and

$$N(n) = \sum_{a=1}^{B\nu_i} q_a(n) \tag{61}$$

represents the total number of Fermions in each state $|n; i\rangle$.

- Additionally, U_{PH} represent a particle-hole transformation acting on the $|n;i\rangle$ states, defined by the following conditions

$$U_{\text{PH}}^\dagger b_{ia}^\dagger U_{\text{PH}} = b_{ia}\,, \tag{62}$$
$$U_{\text{PH}}^\dagger b_{ia} U_{\text{PH}} = b_{ia}^\dagger\,, \tag{63}$$
$$U_{\text{PH}}^\dagger c_{ia}^\dagger U_{\text{PH}} = c_{ia}^\dagger\,, \tag{64}$$
$$U_{\text{PH}}^\dagger c_{ia} U_{\text{PH}} = c_{ia}\,, \tag{65}$$
$$U_{\text{PH}}|0\rangle = \prod_{a=1}^{B\nu_i} b_{ia}^\dagger|0\rangle = |2^{B\nu_i}-1;i\rangle\,. \tag{66}$$

Remark. The basis vectors $|\Gamma;i\rangle \otimes U_{\text{PH}}|n;i\rangle$ in the expansion of Eq. (58) are orthogonal and thus linearly independent. This orthogonality implies that the slave-boson amplitudes $[\phi_i]_{\Gamma n}$ uniquely represent the expansion coefficients in this auxiliary Fock space. Consequently, we have established a one-to-one correspondence between the states $|\Phi_i\rangle$ in the Fock space and the variational parameters encoded in the slave-boson amplitudes.

Remark. The set of all embedding states forms a Fock space, which can be interpreted as a composite system consisting of a subsystem generated by the fermionic degrees of freedom $c_{i\alpha}^\dagger$ with $\alpha \in \{1,\dots,\nu_i\}$ and a subsystem (larger than the previous for $B>1$) generated by the fermionic degrees of freedom $b_{ia}^\dagger$ with $a \in \{1,\dots,B\nu_i\}$.

Theorem 6.1 (Half-filled Embedding States)**.** *Under the variational assumption made in Sec. 3.2*

$$N(n) - N(\Gamma) = m_i = (B-1)\nu_i/2 \tag{67}$$

(see Eq. (8)*), the embedding states $|\Phi_i\rangle$ as defined in Eq.* (58) *have a total of $(B+1)\nu_i/2$ Fermions, signifying that they are half-filled, i.e., they contain half of the maximum possible number of Fermions, which is the total number of modes.*

Proof. Consider the action of the total number operator $\hat{N}_{\text{tot}}$ on the embedding state $|\Phi_i\rangle$

$$\hat{N}_{\text{tot}}|\Phi_i\rangle = \left(\sum_{a=1}^{B\nu_i}(b_{ia}^\dagger b_{ia} + c_{ia}^\dagger c_{ia})\right)|\Phi_i\rangle \tag{68}$$
$$= \sum_{\Gamma=0}^{2^{\nu_i}-1}\sum_{n=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}N(n)(N(n)-1)}[\phi_i]_{\Gamma n}\big(N(\Gamma)+B\nu_i-N(n)\big)|\Gamma;i\rangle\otimes U_{\text{PH}}|n;i\rangle \tag{69}$$
$$= \underbrace{(-m_i+B\nu_i)}_{=(B+1)\nu_i/2}\underbrace{\sum_{\Gamma=0}^{2^{\nu_i}-1}\sum_{n=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}N(n)(N(n)-1)}[\phi_i]_{\Gamma n}|\Gamma;i\rangle\otimes U_{\text{PH}}|n;i\rangle}_{|\Phi_i\rangle} = \left(\frac{B+1}{2}\nu_i\right)|\Phi_i\rangle. \tag{70}$$

This demonstrates that the embedding state $|\Phi_i\rangle$ has a total of $(B+1)\nu_i/2$ electrons, and is therefore half-filled.

□

Remark. As a recap, it is worth highlighting that the significance of Theorem 6.1 is that we can reformulate the variational assumption made initially for ensuring that the gGA variational state has a well-defined number of fermions (see Eq. (8)) into the condition that the embedding states are "half-filled".

As we are going to show in the next sections, the mapping introduced above is not only computationally advantageous, but also plays a critical role in positioning the gGA within the context of quantum embedding frameworks such as Density-Matrix Embedding Theory and Dynamical Mean Field Theory, thereby fostering a unified perspective.

6.2 Expectation values of local operators in terms of embedding states

Here, we aim to study the expectation values of local operators, which were initially represented using slave-boson amplitudes, and establish equivalent representations using the embedding states.

- **Expression for $\mathrm{Tr}\left[\phi_i^\dagger \phi_i\right]$ in terms of embedding states:**

$$\begin{aligned}\langle \Phi_i|\Phi_i\rangle &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1}\sum_{n,n'=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}(N(n)(N(n)-1)-N(n')(N(n')-1))}[\phi_i]^*_{\Gamma n}[\phi_i]_{\Gamma' n'} \\ &\qquad \times \langle \Gamma; i|\Gamma'; i\rangle\langle n; i|U_{\mathrm{PH}}^\dagger U_{\mathrm{PH}}|n'; i\rangle \\ &= \sum_{\Gamma=0}^{2^{\nu_i}-1}\sum_{n=0}^{2^{B\nu_i}-1}[\phi_i]^*_{\Gamma n}[\phi_i]_{\Gamma n} = \mathrm{Tr}\left[\phi_i^\dagger\phi_i\right]. \end{aligned} \tag{71}$$

  We expanded the expression $\langle \Phi_i|\Phi_i\rangle$ according to the definition of embedding states, involving a summation over all Γ and n. Utilizing the orthonormality properties of the states $\langle \Gamma; i|\Gamma'; i\rangle$ and $\langle n; i|n'; i\rangle$, we retained terms for which $\Gamma = \Gamma'$ and $n = n'$. The summation over these terms involved the product of coefficients $[\phi_i]^*_{\Gamma n}[\phi_i]_{\Gamma n}$. The expression was then concisely written as the trace of $\phi_i^\dagger\phi_i$.

- **Expression for $\mathrm{Tr}\left[\phi_i\phi_i^\dagger \hat{H}^i_{\mathrm{loc}}[F^\dagger_{i\alpha}, F_{i\alpha}]\right]$ in terms of embedding states:**

$$\begin{aligned}\langle \Phi_i|\hat{H}^i_{\mathrm{loc}}[c^\dagger_{i\alpha}, c_{i\alpha}]|\Phi_i\rangle &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1}\sum_{n,n'=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}(N(n)(N(n)-1)-N(n')(N(n')-1))}[\phi_i]^*_{\Gamma n}[\phi_i]_{\Gamma' n'} \\ &\qquad \times \langle \Gamma; i|\hat{H}^i_{\mathrm{loc}}[c^\dagger_{i\alpha}, c_{i\alpha}]|\Gamma'; i\rangle\langle n; i|U_{\mathrm{PH}}^\dagger U_{\mathrm{PH}}|n'; i\rangle \\ &= \sum_{\Gamma=0}^{2^{\nu_i}-1}\sum_{n=0}^{2^{B\nu_i}-1}[\phi_i]^*_{\Gamma n}[\phi_i]_{\Gamma n}\langle \Gamma; i|\hat{H}^i_{\mathrm{loc}}[c^\dagger_{i\alpha}, c_{i\alpha}]|\Gamma; i\rangle \\ &= \mathrm{Tr}\left[\phi_i\phi_i^\dagger \hat{H}^i_{\mathrm{loc}}[F^\dagger_{i\alpha}, F_{i\alpha}]\right]. \end{aligned} \tag{72}$$

  The expectation value of the local Hamiltonian was expanded using embedding states, analogously to the initial equation. The orthonormality of the states helped to simplify the expression into a trace of the product of $\phi_i\phi_i^\dagger$ with the matrix representation of the local Hamiltonian in the $|\Gamma; i\rangle$ basis. (Eq. (72))

- **Expression for $\mathrm{Tr}\big[\boldsymbol{\phi}_i^\dagger \boldsymbol{\phi}_i \tilde{\boldsymbol{F}}_{ia}^\dagger \tilde{\boldsymbol{F}}_{ib}\big]$ in terms of embedding states:**

$$\begin{aligned}\langle \Phi_i | b_{ib} b_{ia}^\dagger | \Phi_i \rangle &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}(N(n)(N(n)-1)-N(n')(N(n')-1))} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} \\ &\qquad \times \langle \Gamma; i | \Gamma'; i \rangle \langle n; i | U_{\mathrm{PH}}^\dagger b_{ib} b_{ia}^\dagger U_{\mathrm{PH}} | n'; i \rangle \\ &= \sum_{\Gamma=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma n'} \langle n; i | b_{ib}^\dagger b_{ia} | n'; i \rangle \\ &= \sum_{\Gamma=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma n'} [\tilde{F}_{ib}^\dagger \tilde{F}_{ia}]_{nn'} \\ &= \sum_{\Gamma=0}^{2^{\nu_i}-1} \sum_{n=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma n'} [\tilde{F}_{ia}^\dagger \tilde{F}_{ib}]_{n'n} = \mathrm{Tr}\big[\phi_i^\dagger \phi_i \tilde{F}_{ia}^\dagger \tilde{F}_{ib}\big] . \end{aligned} \tag{73, 74}$$

Initially, the terms were expanded and the particle-hole transformation was incorporated. The orthonormality of $\langle \Gamma; i | \Gamma'; i \rangle$ allowed us to combine the summations over Γ and Γ'. The operators $b_{ib}^\dagger$ and b_{ia} were then expressed using their matrix representations. Furthermore, the real nature of the matrix elements of $\tilde{F}_{ib}$ and $\tilde{F}_{ia}^\dagger$ (as established in Sec. B.3) was employed to simplify the expression. Finally, a compact representation was obtained by writing it as the trace of a product of matrices.

- **Expression for $\mathrm{Tr}\big[\boldsymbol{\phi}_i^\dagger \boldsymbol{F}_{i\alpha}^\dagger \boldsymbol{\phi}_i \tilde{\boldsymbol{F}}_{ia}\big]$ in terms of embedding states:**

$$\begin{aligned}\langle \Phi_i | c_{i\alpha}^\dagger b_{ia} | \Phi_i \rangle &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}(N(n)(N(n)-1)-N(n')(N(n')-1))} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} \\ &\qquad \times \langle \Gamma; i | \langle n; i | U_{\mathrm{PH}}^\dagger c_{i\alpha}^\dagger b_{ia} U_{\mathrm{PH}} | \Gamma'; i \rangle | n'; i \rangle \\ &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} e^{\frac{i\pi}{2}(N(n)(N(n)-1)-N(n')(N(n')-1))} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} \\ &\qquad \times \langle \Gamma; i | \langle n; i | c_{i\alpha}^\dagger b_{ia}^\dagger | \Gamma'; i \rangle | n'; i \rangle \\ &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} (-1)^{n'} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} \langle \Gamma; i | \langle n; i | c_{i\alpha}^\dagger b_{ia}^\dagger | \Gamma'; i \rangle | n'; i \rangle \\ &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} \langle \Gamma; i | c_{i\alpha}^\dagger | \Gamma; i \rangle \langle n; i | b_{ia}^\dagger | n'; i \rangle \\ &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} [F_{i\alpha}^\dagger]_{\Gamma\Gamma'} [\tilde{F}_{ia}^\dagger]_{nn'} \\ &= \sum_{\Gamma,\Gamma'=0}^{2^{\nu_i}-1} \sum_{n,n'=0}^{2^{B\nu_i}-1} [\phi_i]^*_{\Gamma n} [\phi_i]_{\Gamma' n'} [F_{i\alpha}^\dagger]_{\Gamma\Gamma'} [\tilde{F}_{ia}]_{n'n} = \mathrm{Tr}\big[\phi_i^\dagger F_{i\alpha}^\dagger \phi_i \tilde{F}_{ia}\big] . \end{aligned} \tag{75}$$

The particle-hole transformation was applied first in this derivation. The relationship $N(n') = N(n)+1$ was then used to simplify the phase factor, yielding $(-1)^{n'}$. We then rewrote $\langle \Gamma; i|\langle n; i|c^\dagger_{i\alpha} b^\dagger_{ia}|\Gamma'; i\rangle|n'; i\rangle$ in terms of matrix representations of $c^\dagger_{i\alpha}$ and $b^\dagger_{ia}$. Interestingly, the phase factor that emerged from permutations nullified the $(-1)^{n'}$ from the earlier step. As in previous derivations, the realness of the entries of $\tilde{F}_{ia}$ allowed us to transpose it and take its Hermitian conjugate. Lastly, the summations were collected into a single trace expression.

6.3 Recap: the variational problem in terms of embedding states

Recalling the expressions obtained above, we can now rewrite the Gutzwiller constraints [Eqs. (54) and (55)] in terms of the embedding states $|\Phi_i\rangle$ as

$$\langle \Phi_i|\Phi_i\rangle = \langle \Psi_0|\Psi_0\rangle = 1\,, \tag{76}$$

$$\langle \Phi_i|b^\dagger_{ib} b_{ia}|\Phi_i\rangle = \langle \Psi_0|f^\dagger_{ia} f_{ib}|\Psi_0\rangle = [\Delta_i]_{ab}\,, \quad \forall a, b = 1, \ldots, B\nu_i\,. \tag{77}$$

We can also express the matrix $\mathcal{R}_i$, see Eq. (56), as the solution of the equation

$$\langle \Phi_i|c^\dagger_{i\alpha} b_{ia}|\Phi_i\rangle = \sum_{a=1}^{B\nu_i} [\mathcal{R}_i]_{a\alpha} \big[\Delta_i(\mathbf{1}-\Delta_i)\big]^{\frac{1}{2}}_{ab}\,. \tag{78}$$

With these expressions in terms of embedding states, the variational energy $\mathcal{E}$ takes the form

$$\mathcal{E} = \sum_{i,j=1}^{\mathcal{N}} \sum_{a,b=1}^{B\nu_i} \Big[\mathcal{R}^\dagger_i t_{ij} \mathcal{R}_j\Big]_{ab} f^\dagger_{ia} f_{jb} + \sum_{i=1}^{\mathcal{N}} \langle \Phi_i|\hat{H}^i_{\text{loc}}[c^\dagger_{i\alpha}, c_{i\alpha}]|\Phi_i\rangle\,, \tag{79}$$

where $\mathcal{R}_i$ is given by Eq. (78). This variational energy must be minimized with respect to the variational parameters, subject to the Gutzwiller constraints expressed in terms of the embedding states $|\Phi_i\rangle$, formulated with Eqs. (76) and (77).

7 Lagrange formulation of gGA (QE algorithmic structure)

In the previous section (Sec. 6.3), we discussed how the Gutzwiller approximation is formulated in terms of embedding states. However, this poses a complex optimization problem since the dimension of $|\Phi_i\rangle$, which is exponential in ν_i and B, is non-linear. To tackle this, we make use of a mathematical trick, as elaborated in Refs. [7, 20, 10], which reformulates the problem into a linear eigenvalue problem for $|\Phi_i\rangle$, with parameters to be computed recursively. This trick is facilitated by the theorem based on Lagrange multipliers, derived in Appendix F.

7.1 The gGA Lagrange function

We first define the Lagrange function, which encodes the gGA variational-energy function in Eq. (79) and the Gutzwiller constraints [Eqs. (76) and (77)] into a single function, and reduces to the variational energy when evaluated at the saddle point:

$$
\begin{aligned}
\mathcal{L}\left[\Phi, E^c; \mathcal{R}, \Lambda; \mathcal{D}, \Lambda^c; \Delta, \Psi_0, E\right] \; = \\
&\langle\Psi_0|\hat{H}_{\text{qp}}[\mathcal{R},\Lambda]|\Psi_0\rangle + E\left(1-\langle\Psi_0|\Psi_0\rangle\right) \\
&+\sum_{i=1}^{\mathcal{N}}\Big[\langle\Phi_i|\hat{H}^{\text{emb}}_i[\mathcal{D}_i,\Lambda^c_i]|\Phi_i\rangle + E^c_i\left(\mathbf{1}-\langle\Phi_i|\Phi_i\rangle\right)\Big] \\
&-\sum_{i=1}^{\mathcal{N}}\left[\sum_{a,b=1}^{B\nu_i}\left([\Lambda_i]_{ab}+[\Lambda^c_i]_{ab}\right)[\Delta_i]_{ab} + \sum_{c,a=1}^{B\nu_i}\sum_{\alpha=1}^{\nu_i}\Big([\mathcal{D}_i]_{a\alpha}\,[\mathcal{R}_i]_{c\alpha}\,[\Delta_i(\mathbf{1}-\Delta_i)]^{\frac{1}{2}}_{ca}+\text{c.c.}\Big)\right],
\end{aligned}
\tag{80}
$$

where $\mathcal{N}$ is the total number of unit cells. The Lagrange function introduces several Lagrange multipliers and variables. Specifically:

- E is a Lagrange multiplier that enforces the normalization condition $\langle\Psi_0|\Psi_0\rangle = 1$, which is the right-hand side of Eq. (76).
- E^c_i is a Lagrange multiplier that enforces the normalization condition $\langle\Phi_i|\Phi_i\rangle = 1$ for each embedding state $|\Phi_i\rangle$, corresponding to the left-hand side of Eq. (76).
- Δ_i has been promoted to a matrix of independent variables using the Lagrange multipliers Λ_i. Both Δ_i and Λ_i are $B\nu_i \times B\nu_i$ Hermitian matrices.
- Λ^c_i is a $B\nu_i \times B\nu_i$ Hermitian matrix, serving as a Lagrange multiplier to enforce the second Gutzwiller constraints presented in Eq. (77).
- $\mathcal{D}_i$ and $\mathcal{R}_i$ are rectangular matrices with dimensions $B\nu_i \times \nu_i$. $\mathcal{D}_i$ is introduced as a Lagrange multiplier for enforcing the definition of $\mathcal{R}_i$, which is given in Eq. (78).

Note that the Lagrange function has amalgamated all terms involving $|\Psi_0\rangle$ and $|\Phi_i\rangle$ into two auxiliary Hamiltonians, $\hat{H}_{\text{qp}}$ and $\hat{H}_{\text{emb}}$, respectively

$$
\hat{H}_{\text{qp}}[\mathcal{R},\Lambda] = \sum_{i,j=1}^{\mathcal{N}}\sum_{a,b=1}^{B\nu_i}\left[\mathcal{R}^\dagger_i t_{ij}\mathcal{R}^\dagger_j\right]_{ab} f^\dagger_{ia}f_{jb} + \sum_{i=1}^{\mathcal{N}}\sum_{a,b=1}^{B\nu_i}[\Lambda_i]_{ab}\,f^\dagger_{ia}f_{ib}\,, \tag{81}
$$

$$
\hat{H}^i_{\text{emb}}[\mathcal{D}_i,\Lambda^c_i] = \hat{H}^i_{\text{loc}}\left[c_{i\alpha},c^\dagger_{i\alpha}\right] + \sum_{a=1}^{B\nu_i}\sum_{\alpha=1}^{\nu_i}\Big([\mathcal{D}_i]_{a\alpha}\,c^\dagger_{i\alpha}b_{ia}+\text{H.c.}\Big) + \sum_{a,b=1}^{B\nu_i}[\Lambda^c_i]_{ab}\,b_{ib}b^\dagger_{ia}\,, \tag{82}
$$

where Eq. (81) is called "quasi-particle Hamiltonian" and Eq. (82), representing an impurity model consisting of the i-th fragment of the system coupled to a bath, is called embedding Hamiltonian (EH).
The introduction of the Lagrange function has converted the dependencies on $|\Psi_0\rangle$ and $|\Phi_i\rangle$ into linear ones. As we are going to see, this significantly simplifies the problem.

7.2 The gGA Lagrange equations

The saddle point conditions with respect to $|\Psi_0\rangle$ and E result in a Schrödinger equation for $\hat{H}_{\text{qp}}$. Similarly, the saddle point conditions with respect to $|\Phi_i\rangle$ and E_i^c result in a series of Schrödinger equations for $\hat{H}^i_{\text{emb}}$.

To write all Lagrange equations, including the remaining saddle point conditions with respect to the parameters Λ_i, Λ_i^c, Δ_i, $\mathcal{D}_i$, and $\mathcal{R}_i$, we rewrite Eq. (81) as

$$\hat{H}_*[\mathcal{R},\Lambda] = \sum_{i,j=1}^{\mathcal{N}} [\Pi_i h_* \Pi_j]_{ab}\, f^\dagger_{ia} f_{jb}\,, \tag{83}$$

where we introduce the matrix

$$h_* = \begin{pmatrix} \Lambda_1 & \mathcal{R}_1 t_{12} \mathcal{R}_2^\dagger & \dots & \mathcal{R}_1 t_{1\mathcal{N}} \mathcal{R}_{\mathcal{N}}^\dagger \\ \mathcal{R}_2 t_{21} \mathcal{R}_1^\dagger & \Lambda_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}_{\mathcal{N}} t_{\mathcal{N}1} \mathcal{R}_1^\dagger & \dots & \dots & \Lambda_{\mathcal{N}} \end{pmatrix} \tag{84}$$

and the projectors over the degrees of freedom corresponding to each fragment

$$\Pi_i = \begin{pmatrix} \delta_{i1}\,[\mathbf{1}]_{B\nu_1\times B\nu_1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \delta_{iM}\,[\mathbf{1}]_{B\nu_{\mathcal{N}}\times B\nu_{\mathcal{N}}} \end{pmatrix}, \tag{85}$$

where $[\mathbf{1}]_{n\times n}$ is the $n\times n$ identity matrix. We also represent the matrices Δ_i, Λ_i, and Λ_i^c as expansions in terms of an orthonormal basis of Hermitian matrices, denoted $[h_i]_s$ (with respect to the canonical scalar product $(A,B) = \mathrm{Tr}[A^\dagger B]$)

$$\Delta_i = \sum_{s=1}^{(B\nu_i)^2} \left[d_i^0\right]_s \left[h_i^T\right]_s \tag{86}$$

$$\Lambda_i = \sum_{s=1}^{(B\nu_i)^2} [l_i]_s\, [h_i]_s \tag{87}$$

$$\Lambda_i^c = \sum_{s=1}^{(B\nu_i)^2} [l_i^c]_s\, [h_i]_s\ , \tag{88}$$

where $[d_i^0]_s$, $[l_i]_s$, and $[l_i^c]_s$ are real-valued coefficients.

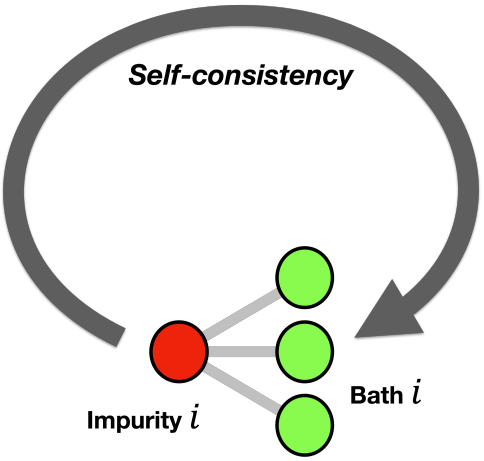


Fig. 2: *Representation of algorithmic structure for solving the gGA Lagrange equations.*

The saddle-point of the gGA Lagrange function $\mathcal{L}$ defined in Eq. (80) is given by the equations

$$\hat{H}_*[\mathcal{R},\Lambda]|\Psi_0\rangle = E_0|\Psi_0\rangle\,, \tag{89}$$

$$[\Delta_i]_{ab} = \langle\Psi_0|\, f^\dagger_{ia} f_{ib}\,|\Psi_0\rangle\,, \tag{90}$$

$$\sum_{c=1}^{B\nu_i} [\mathcal{D}_i]_{c\alpha}\,[\Delta_i\,(\mathbf{1}-\Delta_i)]^{\frac{1}{2}}_{ac} = \sum_j \left[t_{ij}\mathcal{R}^\dagger_j \Pi_j f\,(h_*)\,\Pi_i\right]_{\alpha a}\,, \tag{91}$$

$$[l^c_i]_s = -[l_i]_s - \sum_{c,b=1}^{B\nu_i}\sum_{\alpha=1}^{\nu_i} \frac{\partial}{\partial\,[d^0_i]_s}\left([\Delta_i\,(\mathbf{1}-\Delta_i)]^{\frac{1}{2}}_{cb}\,[\mathcal{D}_i]_{b\alpha}\,[\mathcal{R}_i]_{c\alpha} + \text{c.c.}\right)\,, \tag{92}$$

$$\hat{H}^i_{\text{emb}}|\Phi_i\rangle = E^c_i|\Phi_i\rangle\,, \tag{93}$$

$$\langle\Phi_i|b_{ib}b^\dagger_{ia}|\Phi_i\rangle = [\Delta_i]_{ab}\,, \tag{94}$$

$$\langle\Phi_i|c^\dagger_{i\alpha}b_{ia}|\Phi_i\rangle = \sum_{c=1}^{B\nu_i}[\Delta_i\,(\mathbf{1}-\Delta_i)]^{\frac{1}{2}}\,[\mathcal{R}_i]_{c\alpha}\,, \tag{95}$$

where f is the zero-temperature Fermi function.

Numerous numerical implementations have been proposed in the literature to solve these equations [7, 36, 37]. However, they all fundamentally consist of iteratively computing the ground state of $\hat{H}^i_{\text{emb}}$ (see Eq. (93)), which constitutes the computational bottleneck. This algorithmic structure is schematically represented in Fig. 2. We will list some examples of practical implementations in Sec. 9.

Remark. The matrix derivative in Eq. (92) is non-trivial, as $[h^T_i]_s$ and Δ_i do not commute. A method for computing it is outlined in Sec. G of the Appendix. It's important to note that the computational cost of evaluating this derivative is primarily determined by the diagonalization of Δ_i. However, this operation is computationally inexpensive, rendering the overall cost essentially negligible. Furthermore, a recent work exploring the connection between the gGA and DMET [35] provides an equivalent expression to Eq. (92), that avoids the need for matrix derivatives, potentially offering a more efficient approach for practical implementations.

7.3 Gauge invariance of the gGA equations

It can be readily shown that the gGA Lagrangian is invariant with respect to the following gauge transformation

$$|\Psi_0\rangle \to \mathcal{U}^\dagger(\theta)\,|\Psi_0\rangle \tag{96}$$

$$|\Phi_i\rangle \to U_i^\dagger(\theta_i)\,|\Phi_i\rangle \tag{97}$$

$$\mathcal{R}_i \to u_i^\dagger(\theta_i)\,\mathcal{R}_i \tag{98}$$

$$\mathcal{D}_i \to u^T(\theta_i)\,\mathcal{D}_i \tag{99}$$

$$\Delta_i \to u_i^T(\theta_i)\,\Delta_i\,u_i^*(\theta_i) \tag{100}$$

$$\lambda_i \to u_i^\dagger(\theta_i)\,\lambda_i\,u_i(\theta_i) \tag{101}$$

$$\lambda_i^c \to u_i^\dagger(\theta_i)\,\lambda_i^c\,u_i(\theta_i)\ , \tag{102}$$

with

$$u_i(\theta_i) = e^{i\theta_i} \tag{103}$$

$$U_i(\theta_i) = e^{i\sum_{a,b=1}^{B\nu_i}[\theta_i]_{ab}b_{ia}^\dagger b_{ib}} \tag{104}$$

$$\mathcal{U}(\theta) = e^{i\sum_i\sum_{a,b=1}^{B\nu_i}[\theta_i]_{ab}f_{Ria}^\dagger f_{Rib}}\ , \tag{105}$$

where $u_i(\theta_i) \in \mathbb{C}^{B\nu_i\times B\nu_i}$, $U_i(\theta_i) \in \mathbb{C}^{2^{B\nu_i}\times 2^{B\nu_i}}$ and $\mathcal{U}(\theta) \in \mathbb{C}^{2^{B\nu}\times 2^{B\nu}}$ (where $\nu = \sum_{i=1}^{\mathcal{N}}\nu_i$), and the θ_i are Hermitian matrices.

The name "gauge" here refers to the fact that modifications of the parameters generated by such a gauge transformation do not influence any physical observable. This property of the equations is relevant in relation to the connection with the RISB framework, which is based on an exact reformulation of the many-electron problem in terms of an actual gauge theory, that reduces to the Lagrange equations above at the mean-field level.

8 Generalizations, research directions and open problems

This section aims to explore further generalizations, new research directions, and open problems in the realm of the ghost Gutzwiller approximation (gGA). The evolution of gGA has led to several innovative adaptations and methodologies that harness its potential in various contexts. Here, we mention three of such avenues: the connection of gGA with dynamical mean-field theory (DMFT) through spectral functions, the extension of gGA for time-dependent dynamics, and the reformulation of gGA within quantum-embedding theories. These avenues represent exciting frontiers in the study of strongly correlated electron systems and may offer interesting opportunities for future research.

8.1 The spectral function (connection with DMFT)

Let us consider the gGA zero-temperature spectral function, defined as

$$\mathcal{A}_{i\alpha,j\beta}(\omega) = \langle\Psi_G|\, c_{i\alpha}\, \delta(\omega - \hat{H})\, c^\dagger_{j\beta}\, |\Psi_G\rangle + \langle\Psi_G|\, c^\dagger_{j\beta}\, \delta(\omega + \hat{H})\, c_{i\alpha}\, |\Psi_G\rangle\,. \tag{106}$$

Following Refs. [10, 11], in the limit for $\mathcal{N} \to \infty$, it is possible to write an approximate representation of excitations of $\hat{H}$ in terms of the gGA variational parameters, from which it is possible to obtain the following approximation to the physical Green's function

$$\mathcal{G}_{i\alpha,j\beta}(\omega) = \int_{-\infty}^{\infty} d\epsilon\, \frac{\mathcal{A}_{i\alpha,j\beta}(\omega)}{\omega - \epsilon} \simeq \Big[\mathcal{R}^\dagger_i\, \Pi_i \frac{1}{\omega - h_*} \Pi_j\, \mathcal{R}_j\Big]_{\alpha\beta}, \tag{107}$$

capturing both the low-energy quasi-particle excitations and the Hubbard bands.
From such a formula, it is possible to obtain a pole-expansion expression for the self-energy [10, 11, 36] that closely resembles an expansion proposed in previous DMFT literature [38] and was numerically shown to approach the DMFT solution in the limit of $B \to \infty$ on several examples of single-band and multi-orbital systems [36].
These analytical and numerical results suggest a profound connection with DMFT, which is subject of ongoing research.

8.2 Time-dependent dynamics

The time-dependent ghost Gutzwiller approximation (td-gGA), as introduced in Ref. [39], extends the gGA to the domain of non-equilibrium physics. Specifically, td-gGA builds upon the standard time-dependent Gutzwiller approximation [40,41], by systematically incorporating the auxiliary gGA degrees of freedom.
A key strength of td-gGA is its capability to capture the relaxation of local observables, which is something the standard time-dependent Gutzwiller method falls short of. Moreover, it offers comparable accuracy to the more computationally demanding time-dependent dynamical mean-field theory (td-DMFT), while requiring significantly fewer computational resources. Therefore, it can serve as a versatile tool for delving into the non-equilibrium properties of correlated electron systems, ranging from energy-related materials to quantum control, and other areas where the accurate treatment of strong correlations is required. As such, researchers and students venturing into the field of correlated electron systems may find this method to be a valuable addition to their toolkit.

8.3 Quantum-embedding reformulation (connection with DMET)

The formulation of gGA in terms of quantum embedding states, see Sec. 6.2, has led to the development of a conceptual connection between gGA and density-matrix embedding theory (DMET). In particular, within the ghost density-matrix embedding theory (gDMET) [35], the gGA equations are recast, based on quantum-embedding principles similar to those of DMET. Such alternative interpretation of the gGA equations may open up possibilities for new unexplored generalizations.

9 Code availability

In this section, we draw attention to the available codes for the Gutzwiller approximation (GA) and the ghost Gutzwiller approximation (gGA).

9.1 ComRISB for DFT+GA

The ComRISB package, developed by Yongxin Yao *et al.*, is an efficient tool for integrating density-functional theory (DFT) with the Gutzwiller approximation (GA) and the rotationally-invariant slave-boson (RISB) method. It specifically incorporates the gGA with a single bath site ($B = 1$), which is a particular case within the gGA framework. This package is valuable for studying correlated electron systems by effectively melding the electronic structure calculations of DFT with the Gutzwiller approximation's correlation treatment. ComRISB is available for download at ComRISB at BNL.
The URL is: `https://www.bnl.gov/comscope/software/downloads.php`
For further inquiries, you can contact Yongxin Yao at ykent@iastate.edu.

9.2 gGA code for the single band Hubbard Model

Marius Frank *et al.* have developed a simplified gGA code aimed at the single-band Hubbard model, which is a valuable resource especially for students seeking to comprehend the intricacies of the ghost Gutzwiller approximation (gGA) implementation. The code serves not only as an instructive material but also as a groundwork for researchers endeavoring to build more sophisticated multi-orbital implementations. This hands-on tool provides practical insights into the methodologies and techniques integral to gGA, consistent with the discussions in these lecture notes. The code is available for download at gGA Code at GitLab.
The URL is: `https://gitlab.com/collaborations3/g-ga-hubbard`
For further inquiries, you can contact Marius Frank at marius.frank@chem.au.dk.

Appendices

A Useful mathematical definitions

A.1 Function of an Hermitian matrix

In this section, we will introduce the notion of "function of an Hermitian matrix", that will be used extensively in these lecture notes.
Let H be a Hermitian matrix. It can be diagonalized by a unitary matrix U such that

$$H = UEU^\dagger \tag{108}$$

where E is a diagonal matrix with entries E_n, the eigenvalues of H. The matrix U can be written as $U = [U_1|U_2|\dots|U_N]$, where each U_n is an eigenvector of H corresponding to the eigenvalue E_n.
Given a real-valued function $f\colon \mathbb{R} \to \mathbb{R}$, we define the function of the Hermitian matrix H, denoted as $f(H)$, by

$$f(H) = U f(E)\, U^\dagger \tag{109}$$

where $f(E)$ is a diagonal matrix with entries $f(E_n)$.
This means that the action of $f(H)$ on the eigenvectors of H is the same as the action of H, but with eigenvalues $f(E_n)$ instead of E_n. Specifically, for any eigenvector U_n of H,

$$f(H)\, U_n = f(E_n)\, U_n\,. \tag{110}$$

A.2 Exponential of a Hermitian matrix through Taylor expansion

Another important approach to defining the function of a matrix, especially for the exponential function, is through the Taylor expansion. For a general function $f(x)$, its Taylor series (around $x = 0$) is given by:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n\,, \tag{111}$$

where $f^{(n)}(0)$ is the n-th derivative of f evaluated at 0.
When we focus on the exponential function, particularly e^{iH} or e^{H}, where H is an Hermitian matrix, the Taylor expansion becomes:

$$e^{iH} = \sum_{n=0}^{\infty} \frac{(iH)^n}{n!} \tag{112}$$

$$e^{H} = \sum_{n=0}^{\infty} \frac{(H)^n}{n!}\,. \tag{113}$$

This can be seen as an infinite sum of powers of H. It should be noted that the Taylor expansion gives an equivalent expression to the one provided in the previous section, where the exponential of an Hermitian matrix is defined through its eigenvalues and eigenvectors.

B Preliminaries on fermionic algebra and Fock states

B.1 Recap about the fermionic algebra

As a preamble to the discussion on slave-boson theories and the Gutzwiller approximation, we review some necessary algebraic tools. In particular, we will focus on the fermionic algebra.

Definition B.1 (Fermionic algebra)**.** Consider a set of operators $c_1, \ldots, c_\nu$, together with their adjoints $c_1^\dagger, \ldots, c_\nu^\dagger$. The fermionic algebra is the algebra generated by these operators and the complex numbers, subject to the following canonical anticommutation relations:

$$\{c_\alpha, c_\beta\} = 0\,, \tag{114}$$

$$\{c_\alpha^\dagger, c_\beta^\dagger\} = 0\,, \tag{115}$$

$$\{c_\alpha, c_\beta^\dagger\} = \delta_{\alpha\beta}\,, \tag{116}$$

where $\{A, B\} = AB{+}BA$ denotes the anticommutator, and $\delta_{\alpha\beta}$ is the Kronecker delta.

The anticommutation relations are central in describing fermionic systems. The first two relations reflect the exclusion principle, ensuring that states remain orthogonal under the action of these operators. The last relation essentially states that the operators act as creation and annihilation operators for fermions in the respective modes.

B.2 Recap about the fermionic Fock space

The next step is to construct a representation of the fermionic algebra on a linear space, which we will refer to as the Fock space. This construction is fundamental for analyzing many-body fermionic systems.

Definition B.2 (Fock Space)**.** Let us postulate the existence of a vacuum state $|0\rangle$, which is annihilated by all the annihilation operators c_α,

$$c_\alpha|0\rangle = 0, \quad \text{for all } \alpha\,.$$

Additionally, let there be an inner product with respect to which the conjugation operation is represented as the Hermitian conjugate.
From these properties, it follows that the linear space realizing this representation has dimension 2^ν and is spanned by a basis which we refer to as the Fock basis. In this text, we choose to represent elements of the Fock basis as

$$|\Gamma\rangle = [c_1^\dagger]^{q_1(\Gamma)} \cdots [c_\nu^\dagger]^{q_\nu(\Gamma)}|0\rangle\,,$$

where Γ is an integer ranging from 0 to $2^\nu{-}1$, and its binary representation is $\Gamma{=}q_1(\Gamma)\ldots q_\nu(\Gamma)$. Here, the digits $q_\alpha(\Gamma)$ represent the occupation numbers.

With the Fock basis defined, let us introduce the occupation number operators.

Definition B.3 (Occupation Number Operators)**.** For each mode α, the occupation number operator $\hat{n}_\alpha$ is defined as

$$\hat{n}_\alpha = c_\alpha^\dagger c_\alpha \, .$$

The Fock states are eigenvectors of these operators with eigenvalues $q_\alpha(\Gamma)$. That is,

$$\hat{n}_\alpha |\Gamma\rangle = q_\alpha(\Gamma)|\Gamma\rangle \, .$$

We can also define a full occupation number operator by summing over all modes.

Definition B.4 (Full Occupation Number Operator)**.** The full occupation number operator is given by

$$\hat{N} = \sum_{\alpha=1}^{\nu} c_\alpha^\dagger c_\alpha \, .$$

The Fock states are also eigenvectors of this operator. Specifically,

$$\hat{N}|\Gamma\rangle = \left(\sum_{\alpha=1}^{\nu} q_\alpha(\Gamma) \right) |\Gamma\rangle = N(\Gamma)|\Gamma\rangle \, ,$$

where $N(\Gamma)$ is defined as the sum of all occupation numbers, representing the total number of fermions.

B.3 Matrix representation of creation and annihilation operators

In this section, we will focus on the matrix representation of the creation and annihilation operators in the Fock basis. These matrix representations are essential for practical calculations in many-body fermionic systems.
By using the anticommutation rules, it can be readily verified that the elements of the matrix representation of the creation operator $c_\alpha^\dagger$ are given by the following equations:

$$[F_\alpha^\dagger]_{\Gamma,\Gamma'} = \langle \Gamma | c_\alpha^\dagger | \Gamma' \rangle = \delta_{q_\alpha(\Gamma), q_\alpha(\Gamma')+1} \prod_{s \neq \alpha} \delta_{q_s(\Gamma), q_s(\Gamma')} \, (-1)^{\sum_{s=1}^{\alpha-1} q_s(\Gamma)} \, . \tag{117}$$

The identity above means that the occupation numbers of Γ and Γ' are all equal except for $q_\alpha(\Gamma)$, which is raised by 1 unit with respect to $q_\alpha(\Gamma')$. Furthermore, an appropriate sign has to be taken into account, due to the anticommutation rules.

Example. Let us consider the case where $\nu = 2$, and so the Fock space has a dimension of 4. The matrices representing the creation and annihilation operators in this case are 4×4 matrices. Below are the matrix representations for both creation operators $c_1^\dagger$ and $c_2^\dagger$ and their Hermitian conjugates:

$$F_1^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \; F_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \; F_2^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \; F_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

B.4 Useful single-particle transformations

B.4.1 Unitary canonical transformations

Let us first consider a unitary transformation of the creation and annihilation operators, which is a familiar useful transformation in the context of many-body quantum systems. We will define a unitary operator $\hat{U}$ as

$$\hat{U} = \exp\left(i \sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta \right), \tag{118}$$

where h is a Hermitian matrix. We can now state the following theorem:

Theorem B.1. *The unitary operator $\hat{U}$ leaves the vacuum state unchanged, i.e.,*

$$\hat{U}|0\rangle = |0\rangle\,. \tag{119}$$

Furthermore, the transformed creation operator $\hat{U} c_a^\dagger \hat{U}^\dagger$ can be expressed in terms of the original creation operators $c_\alpha^\dagger$ as

$$\hat{U} c_\alpha^\dagger \hat{U}^\dagger = \sum_{\beta=1}^{\nu} \left[e^{ih}\right]_{\beta\alpha} c_\beta^\dagger\,, \tag{120}$$

where e^{ih} is the exponential of the Hermitian matrix h, understood as a function of an Hermitian matrix in the sense defined in Section A.1.

Proof. Let us begin with the first part of the proof. Since, by definition, the vacuum state $|0\rangle$ satisfies Eq. (B.2), the operator at the exponents $\hat{H} = \sum_{\alpha,\beta} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta$ annihilates the vacuum, i.e., $\hat{H}|0\rangle = 0$. In the Taylor expansion of the exponential, the zeroth-order term is the identity operator, and all other terms contain $\hat{H}$. Therefore, $\hat{U}|0\rangle = |0\rangle$.
Now, let us move on to the second part of the proof:
a) Let V be the unitary matrix such that $V^\dagger h V = e$, where e is diagonal.
b) Consider the transformed creation operators in the new basis:

$$f_a^\dagger = \sum_\alpha V_{\alpha a} c_\alpha^\dagger\,, \tag{121}$$

and write the reversed relation as

$$c_\alpha^\dagger = \sum_a V_{a\alpha}^\dagger f_a^\dagger\,. \tag{122}$$

c) Let us now express $\hat{U} c_\alpha^\dagger \hat{U}^\dagger$ explicitly using the definition of $\hat{U}$:

$$\hat{U} c_\alpha^\dagger \hat{U}^\dagger = \exp\left(i \sum_{\alpha',\beta'=1}^{\nu} h_{\alpha'\beta'} c_{\alpha'}^\dagger c_{\beta'} \right) c_\alpha^\dagger \exp\left(-i \sum_{\alpha'',\beta''=1}^{\nu} h_{\alpha''\beta''} c_{\alpha''}^\dagger c_{\beta''} \right) \tag{123}$$

$$= \exp\left(i \sum_{b'=1}^{\nu} e_{b'b'} f_{b'}^\dagger f_{b'} \right) \sum_a V_{a\alpha}^\dagger f_a^\dagger \exp\left(-i \sum_{b''=1}^{\nu} e_{b''b''} f_{b''}^\dagger f_{b''} \right) \tag{124}$$

$$= \sum_a V_{a\alpha}^\dagger e^{i e_{aa}} f_a^\dagger = \sum_\beta (V e^{ie} V^\dagger)_{\beta\alpha} c_\beta^\dagger = \sum_\beta [e^{ih}]_{\beta\alpha} c_\beta^\dagger\,. \tag{125}$$

Here, we replaced the number operators in the exponents with occupation numbers, and noticed that the left exponent has an extra mode, allowing the exponents to partially cancel out. We are then left with the phase factor $e^{ie_{aa}}$, and rewriting $f_a^\dagger$ back in terms of $c_\beta^\dagger$, we observe the coefficients are of the form $Ve^{ie}V^\dagger$, which is equal to e^{ih} based on the definition of a function of a matrix. □

B.4.2 Similarity transformations with anti-Hermitian generators

In the derivation of the multi-orbital GA and gGA framework, presented in the main text, we employ a similarity transformation related to the one described in the theorem above. However, in contrast to the unitary transformation, the transformation we use is not unitary. Specifically, the theorem can be generalized to transformations where there is no imaginary unit in the exponent.

Theorem B.2. *Let $\hat{U}$ be defined as*

$$\hat{U} = \exp\left(\sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta\right), \quad \textit{and} \quad \hat{U}^{-1} = \exp\left(-\sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta\right).$$

Then, the transformed creation operator $\hat{U}c_\alpha^\dagger\hat{U}^{-1}$ can be expressed in terms of the original creation operators $c_\alpha^\dagger$ as

$$\hat{U}c_\alpha^\dagger\hat{U}^{-1} = \sum_{\beta=1}^{\nu} \left[e^h\right]_{\beta\alpha} c_\beta^\dagger\,, \tag{126}$$

where e^h is the exponential of the Hermitian matrix h.

The proof proceeds analogously to the proof of the previous theorem. The only distinction lies in the absence of the imaginary unit in the exponents. The steps that involve the Taylor expansion of the exponential, replacement of number operators in the exponents with occupation numbers, and manipulation of coefficients as functions of matrices still apply. The result is obtained by simply removing the imaginary units from the exponents throughout the steps of the proof.

Theorem B.3. *From Eq.* (126) *and the fact that h, $\hat{U}$ and $\hat{U}^{-1}$ are both Hermitian, it also follows that:*

$$\hat{U}c_\alpha\hat{U}^{-1} = \sum_{\beta=1}^{\nu} \left[e^{-h}\right]_{\alpha\beta} c_\beta\,. \tag{127}$$

C One-body Hamiltonians and the Fermi-function matrix

In the study of quantum systems, particularly fermionic systems, one often encounters one-body Hamiltonians. These Hamiltonians describe the energy of the system in terms of single-particle states. A special case, which we will focus on, is when the one-body Hamiltonian is diagonal in the second quantization formalism. This simplifies the description and allows us to connect the properties of the system to the Fermi function.

C.1 Partition function and thermal distribution

Before diving into one-body Hamiltonians, let us first define the partition function Z and the thermal distribution for a generic Hamiltonian $\hat{H}$. The partition function is given by

$$Z(T) = \text{Tr}\left[e^{-\frac{1}{T}\hat{H}}\right], \tag{128}$$

where T is the temperature of the system, and the trace is taken over the entire Fock space. The thermal distribution at temperature T is defined as the normalized density-matrix

$$\rho_T = \frac{e^{-\frac{1}{T}\hat{H}}}{Z(T)}, \tag{129}$$

which allows us to compute expectation values of operators in the system at finite temperature. Specifically, for the expectation value of the number operator $f_a^\dagger f_b$, we define

$$n_{ab}(T) = \text{Tr}\left[\rho_T f_a^\dagger f_b\right]. \tag{130}$$

C.2 One-body diagonal Hamiltonian

Now, let us consider a specific one-body Hamiltonian that is diagonal,

$$\hat{H} = \sum_{a=1}^{\nu} e_{aa}\, f_a^\dagger f_a\,. \tag{131}$$

We can state the following theorem regarding the calculation of Z and n_{ab} for this Hamiltonian:

Theorem C.1. *For the one-body diagonal Hamiltonian $\hat{H} = \sum_a e_{aa} f_a^\dagger f_a$, the partition function $Z(T)$ and the thermal expectation values $n_{ab}(T)$ are given by*

$$Z(T) = \prod_{a=1}^{\nu}\left(1 + e^{-\frac{e_{aa}}{T}}\right), \tag{132}$$

$$n_{ab}(T) = \delta_{ab}\, f_T(e_{aa})\,, \tag{133}$$

where $f_T(x)$ is the Fermi function at temperature T defined as

$$f_T(x) = \frac{1}{e^{x/T} + 1}\,. \tag{134}$$

Proof. We begin by expressing the partition function $Z(T)$ as a trace over the Fock space,

$$\begin{aligned} Z(T) &= \mathrm{Tr}\left[e^{-\frac{1}{T}\sum_a e_{aa} f_a^\dagger f_a}\right] \\ &= \sum_{\Gamma=0}^{2^\nu-1} \langle \Gamma | e^{-\frac{1}{T}\sum_a e_{aa} f_a^\dagger f_a} | \Gamma \rangle = \sum_{\Gamma=0}^{2^\nu-1} \langle \Gamma | \exp\left(-\frac{1}{T}\sum_a e_{aa} q_a(\Gamma)\right) | \Gamma \rangle \,, \end{aligned} \tag{135}$$

where we have used the Fock basis representation $|\Gamma\rangle = [c_1^\dagger]^{q_1(\Gamma)} \cdots [c_\nu^\dagger]^{q_\nu(\Gamma)} |0\rangle$, and the relation $f_a^\dagger f_a |\Gamma\rangle = q_a(\Gamma)|\Gamma\rangle$.
We can further break down the sum over Γ into a sum over the occupation numbers q_a for each a,

$$\begin{aligned} Z(T) &= \sum_{q_1=0}^{1} \cdots \sum_{q_\nu=0}^{1} \exp\left(-\frac{1}{T}\sum_a e_{aa} q_a\right) \\ &= \prod_{a=1}^{\nu} \left(\sum_{q_a=0}^{1} \exp\left(-\frac{e_{aa} q_a}{T}\right)\right) = \prod_{a=1}^{\nu} \left(1 + \exp\left(-\frac{e_{aa}}{T}\right)\right) \,. \end{aligned} \tag{136}$$

Here, in the second step, we have used that the exponential of the sum is the product of the exponentials and separated the terms corresponding to each a. The expression obtained is the desired result for the partition function $Z(T)$ in terms of the eigenvalues of the one-body Hamiltonian. □

Proof. Let us first compute $Z(T)\, n_{ab}$, which can be written as a trace:

$$Z(T)\, n_{ab} = \mathrm{Tr}\left[e^{-\frac{1}{T}\sum_c e_{cc} f_c^\dagger f_c} f_a^\dagger f_b\right] = \sum_{\Gamma=0}^{2^\nu-1} \langle \Gamma | e^{-\frac{1}{T}\sum_c e_{cc} f_c^\dagger f_c} f_a^\dagger f_b | \Gamma \rangle \,. \tag{137}$$

If $a \neq b$, this is zero since expanding in terms of the Fock states will always have different occupation numbers on the left and right sides.
Now, let us focus on the case $a = b$,

$$\begin{aligned} Z(T)\, n_{aa} &= \sum_{\Gamma=0}^{2^\nu-1} \langle \Gamma | e^{-\frac{1}{T}\sum_c e_{cc} q_c(\Gamma)} q_a(\Gamma) | \Gamma \rangle \\ &= \sum_{q_1=0}^{1} \cdots \sum_{q_\nu=0}^{1} q_a \exp\left(-\frac{1}{T}\sum_c e_{cc} q_c\right) \\ &= \exp\left(-\frac{e_{aa}}{T}\right) \prod_{c\neq a} \left(1 + \exp\left(-\frac{e_{cc}}{T}\right)\right) = Z(T) \frac{\exp\left(-e_{aa}/T\right)}{1 + \exp\left(-e_{aa}/T\right)} \,. \end{aligned} \tag{138}$$

Dividing by $Z(T)$, we get

$$n_{aa} = \frac{\exp\left(-e_{aa}/T\right)}{1 + \exp\left(-e_{aa}/T\right)} = \frac{1}{e^{e_{aa}/T} + 1} = f_T(e_{aa}) \,, \tag{139}$$

which is the Fermi function at temperature T. □

C.3 Fermi-function matrix for non-diagonal one-body Hamiltonians

Now we consider a general one-body Hamiltonian, $\hat{H}$, given by

$$\hat{H} = \sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta \,. \tag{140}$$

We aim to generalize the results from the previous section for a diagonal Hamiltonian to this more general case.

Theorem C.2. *For an arbitrary one-body Hamiltonian $\hat{H}$, the thermal expectation value $\Delta_{\alpha\beta}(T) = \langle c_\alpha^\dagger c_\beta \rangle$ is given by*

$$\Delta_{\alpha\beta}(T) = [f_T(h)]_{\beta\alpha} \,, \tag{141}$$

where f_T is the Fermi function at temperature T.

Proof. We start by diagonalizing the matrix h. This can be done by writing $h = VeV^\dagger$, where V is a unitary matrix that diagonalizes h and e is the resulting diagonal matrix. The Hamiltonian can then be rewritten as $\hat{H} = \sum_a e_{aa} f_a^\dagger f_a$, where

$$f_a^\dagger = \sum_\alpha V_{\alpha a} c_\alpha^\dagger \,, \tag{142}$$

$$c_\alpha^\dagger = \sum_a V_{a\alpha}^\dagger f_a^\dagger \,. \tag{143}$$

Using this, we can write the thermal expectation value $\langle c_\alpha^\dagger c_\beta \rangle_T$ as

$$\Delta_{\alpha\beta}(T) = \langle c_\alpha^\dagger c_\beta \rangle_T \tag{144}$$

$$= \mathrm{Tr}\Big[\rho_T \Big(\sum_a V_{a\alpha}^\dagger f_a^\dagger\Big)\Big(\sum_b V_{\beta b} f_b\Big)\Big] \tag{145}$$

$$= \sum_a V_{a\alpha}^\dagger \sum_b V_{\beta b}\, n_{ab} = \sum_a V_{a\alpha}^\dagger \sum_b V_{\beta b}\, \delta_{ab}\, f_T(e_{aa}) = \big[V f_T(e) V^\dagger\big]_{\beta\alpha} \,, \tag{146}$$

where $f_T(e)$ is the Fermi function at temperature T defined as

$$f_T(e) = \frac{1}{e^{e/T}+1} \,. \tag{147}$$

Finally, using the definition of the function of a Hermitian matrix, as discussed in Section A.1, we can write

$$\Delta_{\alpha\beta}(T) = [f_T(h)]_{\beta\alpha} \,, \tag{148}$$

which concludes the proof. □

D Wick's theorem for one-body thermal states

Wick's theorem provides a powerful tool for evaluating the expectation values of products of creation and annihilation operators with respect to thermal states of one-body Hamiltonians, including the ground state, which is going to be essential within the formalism of the GA and the gGA.

Theorem D.1 (Wick's Theorem)**.** *Wick's Theorem provides a systematic method to decompose the expectation value of a product of creation and annihilation operators into a sum of products of expectation values of pairs of operators.*
Consider a system described by a generic one-body Hamiltonian of the form

$$\hat{H} = \sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta} c_\alpha^\dagger c_\beta \,, \tag{149}$$

and a corresponding generic thermal density-matrix ρ_T, as defined in Eq. (129)*.*
Let $O = o_1 o_2 \cdots o_n$ be a string of creation and annihilation operators, where each o_i is either $c_\alpha^\dagger$ or c_β.
Wick's theorem states that the thermal expectation value $\langle O \rangle_T$ can be expanded as a sum of products of contracted pairs:

$$\langle O \rangle_T = \sum_{\textit{all contractions}} (-1)^{\textit{crossings}} \prod_{\textit{contractions}} \langle o_i o_j \rangle_T \,. \tag{150}$$

Here, each term in the sum corresponds to a distinct way of pairing the creation and annihilation operators into contractions. A contraction between o_i and o_j is represented as $\langle o_i o_j \rangle_T$. The number of crossings is the number of times the contraction lines cross each other, and the sign is determined by the parity (even or odd) of the number of crossings.
The thermal expectation values for pairs of creation and annihilation operators are given by

$$\left\langle c_\alpha^\dagger c_\beta \right\rangle_T = [f_T(h)]_{\beta\alpha} \,, \tag{151}$$
$$\left\langle c_\beta c_\alpha^\dagger \right\rangle_T = \delta_{\alpha\beta} - [f_T(h)]_{\beta\alpha} \,. \tag{152}$$

This theorem, which is presented here without proof, is best understood through examples. The examples below illustrate how Wick's Theorem works in practice.

Example. Consider calculating the expectation value of $c_1^\dagger c_1 c_1^\dagger c_1$. This can be written as a sum of two terms corresponding to different contractions:

$$\begin{aligned}\langle c_1^\dagger c_1 c_1^\dagger c_1 \rangle_T &= \langle \overbrace{c_1^\dagger c_1}\, \overbrace{c_1^\dagger c_1} \rangle_T + \langle \overbrace{c_1^\dagger \overbrace{c_1 c_1^\dagger} c_1} \rangle_T \\ &= \left[f_T(h)\right]_{11}^2 + \left[f_T(h)\right]_{11}\left(1 - \left[f_T(h)\right]_{11}\right) = \left[f_T(h)\right]_{11} .\end{aligned}$$

The first contribution comes from the contraction lines between the pairs of creation and annihilation operators that are next to each other, while the second contribution is from the pairs that are more distant. The sum of these contributions gives $[f_T(h)]_{11}$. This is consistent with the fact that $c_1^\dagger c_1 c_1^\dagger c_1 = c_1^\dagger c_1$ at the operator level.

Example. Consider calculating the expectation value of $c_1^\dagger c_1^\dagger c_1 c_1$. This can be written as a sum of two terms corresponding to different contractions

$$\langle c_1^\dagger c_1^\dagger c_1 c_1\rangle_T = \langle c_1^\dagger c_1^\dagger c_1 c_1\rangle_T + \langle c_1^\dagger c_1^\dagger c_1 c_1\rangle_T = -\big[f_T(h)\big]_{11}^2 + \big[f_T(h)\big]_{11}^2 = 0\,.$$

The first contribution comes with a negative sign due to one crossing, while the second contribution comes with a positive sign. The sum of these contributions gives 0, consistent with the fact that $c_1^\dagger c_1^\dagger c_1 c_1 = 0$ at the operator level.

Example. Consider calculating the expectation value of $c_1^\dagger c_2^\dagger c_2 c_1$. This can be written as a sum of two terms corresponding to different contractions

$$\langle c_1^\dagger c_2^\dagger c_2 c_1\rangle_T = \langle c_1^\dagger c_2^\dagger c_2 c_1\rangle_T + \langle c_1^\dagger c_2^\dagger c_2 c_1\rangle_T = \big[f_T(h)\big]_{11}\big[f_T(h)\big]_{22} - \big[f_T(h)\big]_{12}\big[f_T(h)\big]_{21}\,.$$

The first contribution comes with no crossings, while the second contribution has one crossing and comes with a negative sign.

D.1 A useful observation involving Wick contractions

In this subsection, we present a useful observation based on Wick's theorem, which plays an important role in the formal derivation of the multi-orbital GA and gGA equations. We consider a thermal state of a one-body Hamiltonian (refer to Eq. (129)) and focus on the calculation of $\langle X c_\alpha^\dagger c_\beta\rangle_T$, where X represents a product of creation and annihilation operators.

Theorem D.2 (Classification of Contractions with Additional Operators)**.** *Let us consider a thermal state of a one-body Hamiltonian as defined in Wick's Theorem (see Theorem D.1), and let X be a generic fermionic operator, which can be represented as a linear combination of strings of creation and annihilation operators. Consider the expectation value $\langle X c_\alpha^\dagger c_\beta\rangle_T$. By applying Wick's theorem, we can classify the terms obtained into two types:*

1. ***Type 1:*** *Terms where $c_\alpha^\dagger c_\beta$ are contracted with each other. The sum of all such "disconnected" terms yields*

$$\langle X c_\alpha^\dagger c_\beta\rangle_T = \langle X\rangle_T \langle c_\alpha^\dagger c_\beta\rangle_T. \tag{153}$$

2. ***Type 2:*** *The sum of all remaining "connected" terms, which can be written in the form*

$$\langle X c_\alpha^\dagger c_\beta\rangle_T = \sum_{\alpha'\beta'} \xi_T^{\alpha'\beta'} \langle c_{\alpha'} c_\alpha^\dagger\rangle_T \langle c_{\beta'}^\dagger c_\beta\rangle_T\,. \tag{154}$$

   Here the coefficients $\xi_T^{\alpha'\beta'}$ depend only on T and X, but not on α and β. In fact, these coefficients correspond to the sum of all terms that arise from self-contractions among the operators remaining in X once we exclude α' and β' (which are contracted with α and β, respectively).

This observation is particularly helpful in simplifying calculations involving thermal expectation values with additional operators, and it is extensively employed in the derivation of equations within the multi-orbital GA and gGA formalisms.

E Reduced density-matrix of a fermionic subsystem

In this section, we discuss the reduced many-body density-matrix of a subsystem in a one-body fermionic system, as described in Wick's theorem (see Theorem D.1). This concept plays an important role in the formal derivation of the multi-orbital GA and gGA equations.
Let us recall that the Hamiltonian is given by

$$\hat{H} = \sum_{\alpha,\beta=1}^{\nu} h_{\alpha\beta}\, c_\alpha^\dagger c_\beta \,, \tag{155}$$

as stated in Theorem D.1, and the corresponding thermal states are represented by Eq. (129).
We denote the modes in the full system as $c_1, \ldots, c_\nu$ and the modes in a subsystem S as $c_1, \ldots, c_\eta$, where $\eta < \nu$.
Let us consider the so-called single-particle reduced density-matrix for the subsystem S, which is the following $\eta \times \eta$ matrix

$$\Delta_{\alpha\beta} = \langle c_\alpha^\dagger c_\beta \rangle_T = \big[f_T(h)\big]_{\beta\alpha} \quad \forall\, \alpha, \beta = 1, .., \eta \,, \tag{156}$$

where f_T is the Fermi function at temperature T, as shown in Sec. C.3.

Theorem E.1 (Reduced Many-Body Density-Matrix for a Subsystem)**.** *Let S be a subsystem with η modes, and let Δ be the single-particle reduced density-matrix for the subsystem. The reduced many-body density-matrix of S with respect to a thermal state of the Hamiltonian $\hat{H}$ is given by*

$$\rho_T^S = \frac{\exp\left(-\sum_{\alpha,\beta=1}^{\eta} F_{\alpha\beta}\, c_\alpha^\dagger c_\beta\right)}{\mathrm{Tr}_S\left[\exp\left(-\sum_{\alpha,\beta=1}^{\eta} F_{\alpha\beta}\, c_\alpha^\dagger c_\beta\right)\right]} \,, \tag{157}$$

where Tr_S denotes the trace over the many-body Fock space of the subsystem S, and F is an $\eta \times \eta$ matrix that can be expressed in terms of Δ as

$$F = \ln\left(\frac{1-\Delta^T}{\Delta^T}\right) \,, \tag{158}$$

where the superscript T denotes the transpose.
In other words, given any operator $\hat{O}$ acting on the subsystem S (i.e., any operator constructed as algebraic combinations of fermionic modes from 1 to η), its expectation value can be calculated as

$$\langle \hat{O} \rangle_T = \mathrm{Tr}_S\left[\rho_T^S \hat{O}\right] \,. \tag{159}$$

Proof. We will split the proof into two parts.
Part 1: First, we prove that Eq. (159) holds for operators of the form $\hat{O} = c_\alpha^\dagger c_\beta$ with $\alpha, \beta = 1, \ldots, \eta$. Notice that ρ_T^S has the form of a thermal density-matrix with $T = 1$ and a Hamiltonian parameterized by F. Therefore, we can apply the theorem proved before Sec. C.3 to express $\mathrm{Tr}_S[\rho_T^S c_\alpha^\dagger c_\beta]$ in terms of the corresponding Fermi function as

$$\mathrm{Tr}_S\big[\rho_T^S c_\alpha^\dagger c_\beta\big] = \big[f_{T=1}(F)\big]_{\beta\alpha} \,. \tag{160}$$

However, by the definition of F,

$$f_{T=1}(F) = \frac{1}{1+e^F} = \Delta^T. \tag{161}$$

This establishes the desired result for operators of the form $\hat{O} = c_\alpha^\dagger c_\beta$.

Part 2: Now, we extend the result to arbitrary operators $\hat{O}$ in the subsystem S. Since both sides of Eq. (159) resemble thermal expectation values, Wick's theorem (Theorem D.1) applies. Wick's theorem reduces the calculation of expectation values to algebraic combinations of Wick contractions. Since we have already proven that Eq. (159) holds for Wick contractions, it follows that Eq. (159) holds for arbitrary operators $\hat{O}$ acting on the subsystem S. □

Remark. It is important to note that we have implicitly assumed that the eigenvalues of Δ^T (equivalently, the eigenvalues of Δ) lie strictly between 0 and 1, that is, in the interval $(0, 1)$. This assumption is crucial for the well-definedness of F, as the logarithm in the expression for F would be ill-defined if the eigenvalues were 0 or 1.

F Promoting of functions to independent variables

In the context of formulating the ghost Gutzwiller approximation (gGA) equations, it proves beneficial to adopt a certain mathematical trick involving Lagrange multipliers. This trick is particularly useful for extremizing functions that have a specific structure, and it facilitates obtaining a QE algorithmic structure. Specifically, let us consider a real multivariable function $f(\mathbf{X})$ of the form $f(\mathbf{X}) = g\big(a_1(\mathbf{X}), \ldots, a_n(\mathbf{X})\big)$, where $\mathbf{X} = (X_1, \ldots, X_m)$ is a set of variables. We are interested in extremizing f with respect to $\mathbf{X}$.

Theorem F.1. *Let $L(\boldsymbol{\lambda}, \mathbf{a}, \mathbf{X})$ be a function constructed as*

$$L(\boldsymbol{\lambda}, \mathbf{a}, \mathbf{X}) = g(a_1, \ldots, a_n) - \sum_{k=1}^{n} \lambda_k \big(a_k - a_k(\mathbf{X})\big)\,, \tag{162}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\mathbf{a} = (a_1, \ldots, a_n)$. If $\bar{\mathbf{X}}$ is an extremum of $f(\mathbf{X})$, then there exist values $\bar{\boldsymbol{\lambda}}$ and $\bar{\mathbf{a}}$ such that $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{a}}, \bar{\mathbf{X}})$ is an extremum of $L(\boldsymbol{\lambda}, \mathbf{a}, \mathbf{X})$.

Proof. Let us examine the saddle point conditions for L. These conditions imply:

1. Differentiating with respect to λ_i yields

$$a_i = a_i(\bar{\mathbf{X}})\,. \tag{163}$$

2. Differentiating with respect to a_i yields

$$\frac{\partial g}{\partial a_i} = \lambda_i\,. \tag{164}$$

3. Differentiating with respect to X_l yields

$$\sum_{k=1}^{n} \lambda_k \frac{\partial a_k}{\partial X_l} = 0\,. \tag{165}$$

Substituting Eqs. (163) and (164) into Eq. (165), we obtain

$$\sum_{k=1}^{n} \frac{\partial g}{\partial a_k} \left(a_1(\bar{\mathbf{X}}), \ldots, a_n(\bar{\mathbf{X}})\right) \frac{\partial a_k}{\partial X_l} = \frac{\partial f}{\partial X_l}(\bar{\mathbf{X}}) = 0\,, \tag{166}$$

which is a necessary condition for an extremum of f. Hence, if $\bar{\mathbf{X}}$ is an extremum of f, it follows that $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{a}}, \bar{\mathbf{X}})$ is an extremum of L. □

This technique is particularly useful because it provides us with different options for how to implement the saddle-point search in practice. For example, we can first calculate the saddle-point conditions with respect to $\mathbf{X}$ for L, and this may be easier than deriving f with respect to $\mathbf{X}$, especially if the functions $a_i(\mathbf{X})$ are relatively simple compared to g. This is the kind of scenario that we will encounter in the main text.

G A useful matrix derivative

One of the gGA/GA equations requires to compute a matrix derivative of the form:

$$X = \frac{d}{d\lambda}[K + \lambda H]^{-\frac{1}{2}}\bigg|_{\lambda=0}\,, \tag{167}$$

where K and H are Hermitian matrices that do not commute. A simple way to compute this derivative is by reducing the calculation to a Sylvester equation [19], as follows.
Define

$$Y = \frac{d}{d\lambda}[K + \lambda H]^{-1}\bigg|_{\lambda=0} = -K^{-1}\,\frac{d}{d\lambda}[K + \lambda H]\bigg|_{\lambda=0} K^{-1}\,. \tag{168}$$

Deriving both sides of the equation $[K + \lambda H]^{-1} = ([K + \lambda H]^{-\frac{1}{2}})^2$, it follows that

$$XK^{-1} + K^{-1}X = Y\,, \tag{169}$$

which is a Sylvester equation that can be solved using standard methods.
Let us call U the unitary matrix that diagonalizes K, i.e., $U^\dagger K U = k$, where k is a diagonal matrix. By applying this unitary transformation to Eq. (169), we get:

$$U^\dagger X U\, k^{-1} + k^{-1}\, U^\dagger X U = U^\dagger Y U\,, \tag{170}$$

which can be easily inverted since k is diagonal:

$$[U^\dagger X U]_{ab} = \frac{[U^\dagger Y U]_{ab}}{k_{aa}^{-1} + k_{bb}^{-1}}\,. \tag{171}$$

The desired matrix X can be obtained by applying the inverse unitary transformation to both sides of Eq. (171).

References

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

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

[3] X.-Y. Deng, L. Wang, X. Dai, and Z. Fang, Phys. Rev. B **79**, 075114 (2009)

[4] K.M. Ho, J. Schmalian, and C.Z. Wang, Phys. Rev. B **77**, 073101 (2008)

[5] N. Lanatà, H.U.R. Strand, X. Dai, and B. Hellsing, Phys. Rev. B **85**, 035133 (2012)

[6] N. Lanatà, P. Barone, and M. Fabrizio, Phys. Rev. B **78**, 155127 (2008)

[7] N. Lanatà, Y.-X. Yao, C.-Z. Wang, K.-M. Ho, and G. Kotliar, Phys. Rev. X **5**, 011008 (2015)

[8] J. Bünemann, F. Gebhard, T. Ohm, S. Weiser, and W. Weber in A. Narlikar (Ed.) *Frontiers in Magnetic Materials* (Springer, Berlin, 2005), p. 117

[9] C. Attaccalite and M. Fabrizio, Phys. Rev. B **68**, 155117 (2003)

[10] N. Lanatà, T.-H. Lee, Y.-X. Yao, and V. Dobrosavljević, Phys. Rev. B **96**, 195126 (2017)

[11] M.S. Frank, T.-H. Lee, G. Bhattacharyya, P.K.H. Tsang, V.L. Quito, V. Dobrosavljević, O. Christiansen, and N. Lanatà, Phys. Rev. B **104**, L081103 (2021)

[12] D. Guerci, M. Capone, and M. Fabrizio, Phys. Rev. Materials **3**, 054605 (2019)

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

[14] R. Yu and Q. Si, Phys. Rev. B **86**, 085104 (2012)

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

[16] T. Li, P. Wölfle, and P.J. Hirschfeld, Phys. Rev. B **40**, 6817 (1989)

[17] R. Frésard and P. Wölfle, Int. J. Mod. Phys. B **06**, 685 (1992)

[18] R. Frésard and G. Kotliar, Phys. Rev. B **56**, 12909 (1997)

[19] F. Lechermann, A. Georges, G. Kotliar, and O. Parcollet, Phys. Rev. B **76**, 155102 (2007)

[20] N. Lanatà, Y.-X. Yao, X. Deng, V. Dobrosavljević, and G. Kotliar, Phys. Rev. Lett. **118**, 126401 (2017)

[21] S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997)

[22] R. Orús, Ann. Phys. **349**, 117 (2014)

[23] G. Carleo and M. Troyer, Science **355**, 602 (2017)

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

[25] V. Anisimov and Y. Izyumov: *Electronic Structure of Strongly Correlated Materials* (Springer, 2010)

[26] E. Fertitta and G.H. Booth, Phys. Rev. B **98**, 235132 (2018)

[27] F. Verstraete, V. Murg, and J. Cirac, Adv. Phys. **57**, 143 (2008)

[28] Y.-H. Zhang and S. Sachdev, Phys. Rev. Res. **2**, 023172 (2020)

[29] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, PNAS **119**, e2122059119 (2022)

[30] S. Sakai, M. Civelli, and M. Imada, Phys. Rev. B **94**, 115130 (2016)

[31] P.W. Anderson, Phys. Rev. B **78**, 174505 (2008)

[32] N. Lanatà, Phys. Rev. B **105**, 045111 (2021)

[33] J. Bünemann and F. Gebhard, Phys. Rev. B **76**, 193104 (2007)

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

[35] N. Lanatà, arXiv:2305.11895

[36] T.-H. Lee, C. Melnick, R. Adler, N. Lanatà, and G. Kotliar, arXiv:2305.1112823

[37] C. Mejuto-Zaera and M. Fabrizio, arXiv:2305.03329

[38] S.Y. Savrasov, K. Haule, and G. Kotliar, Phys. Rev. Lett. **96**, 036404 (2006)

[39] D. Guerci, M. Capone, and N. Lanatà, arXiv:2303.09584

[40] M. Schirò and M. Fabrizio, Phys. Rev. Lett. **105**, 076401 (2010)

[41] G. Seibold and J. Lorenzana, Phys. Rev. Lett. **86**, 2605 (2001)

16 DMFT for f-Electron Systems

Bernard Amadon
CEA, DAM, DIF,
Université Paris-Saclay
Laboratoire Matière en Conditions Extrêmes
91680 Bruyères-le-Châtel, France

Contents

1 Introduction to strong correlation: localized (f) orbitals **2**
1.1 Atomic orbitals . . . 2
1.2 Localization/delocalization of electrons . . . 3
1.3 The Hubbard model . . . 4

2 Introduction to DMFT and DFT+DMFT **7**
2.1 Dynamical Mean-Field Theory . . . 7
2.2 DFT+DMFT . . . 12

3 f-electron systems: pure metals **20**
3.1 Lanthanides . . . 20
3.2 Actinides . . . 24

4 f-electron systems: oxides **26**
4.1 Electronic structure of Ce_2O_3 . . . 26
4.2 Magnetic ordering in UO_2 . . . 28

5 Conclusion **28**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

In these notes, after a brief introduction to strong correlations and f-orbitals in solid state physics, I will introduce Dynamical Mean-Field Theory (DMFT) and the combination of Density Functional Theory (DFT) and DMFT (See also references [1], [2], and [3]). Then I will discuss some applications of DFT+DMFT on some metals and oxides.

1 Introduction to strong correlation: localized (f) orbitals

1.1 Atomic orbitals

We first focus on the spatial localization of orbitals as a function of their principal quantum numbers: $1s$, $2p$, $3d$ and $4f$ orbitals are orthogonal to lower orbitals only through the angular part because for their value of l, they are the orbitals with the lowest principal quantum number [3]. As a consequence, the radial part of the wavefunction is not constrained by orthogonality, and can lower its Coulomb interaction with the nucleus by being more localized. Thus, these orbitals are more localized than others. Fig. 1 illustrates the localization of $3d$ and $4f$ with respect to $4d$, $5d$ and $5f$ orbitals. Indeed, lanthanides are notorious examples of systems exhibiting very strong correlation effects.
Concerning the angular part, it has more and more nodal planes as l increases, thus in particular, d and f orbitals are fairly localized in space. As a consequence, $4d$, $5d$ and $5f$ orbitals exhibit also important correlation effects.

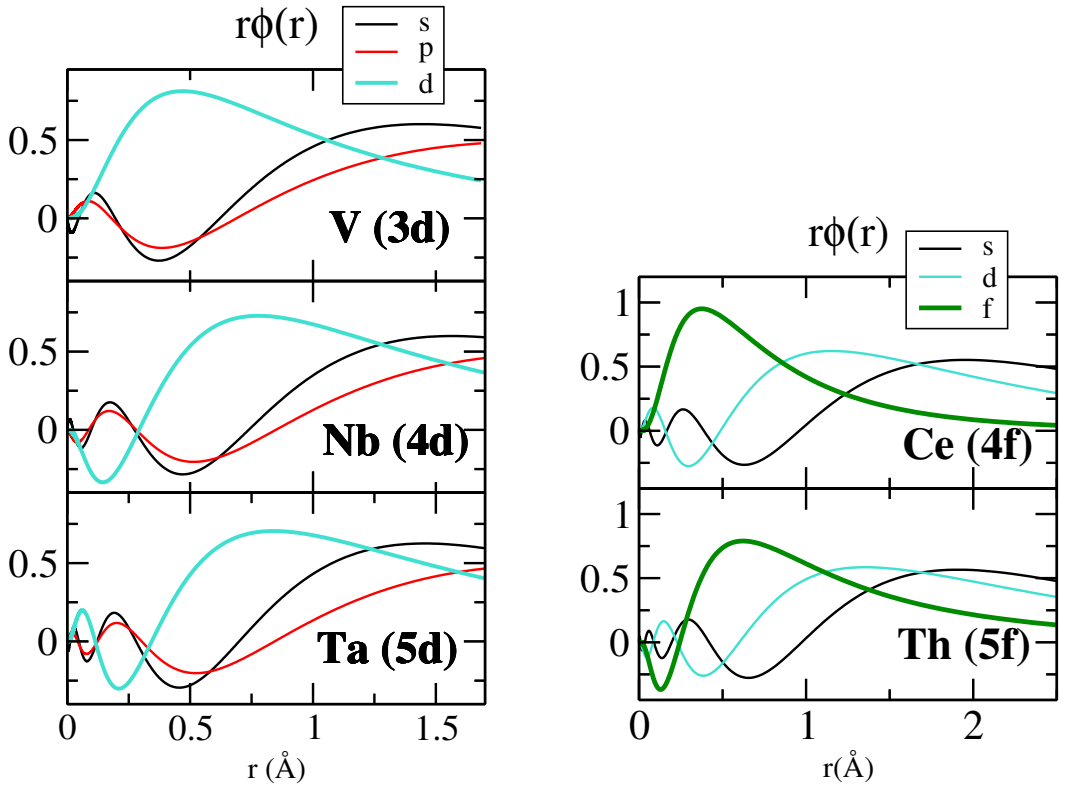


Fig. 1: *Radial part of valence wavefunctions of some atoms (computed in DFT/LDA).*

1.2 Localization/delocalization of electrons

Two main consequences arise from the localization of orbitals:

- When a solid is formed from atoms, the overlap of localized orbitals is weaker than for delocalized orbitals and thus the bandwidth W is smaller (see e.g [4]).
- The more localized the orbitals, the stronger the local interactions "U" between electrons, because electrons are closer to each other.

The simplest example to illustrate these two effects is the formation of the H_2 molecule. In this case, the competition between chemical bonding and interaction leads to a change of the ground state as the interatomic distance increases. At large distances, the electrons are localized each on its atom to avoid the Coulomb interaction with the other electrons (here $U > W$). At small distance, the two electrons delocalize in a bonding molecular orbital.
We see another consequence of this for solids on Fig. 2. It represents the evolution of the volume of pure elements as a function of the number of electrons, for $4d$, $4f$ and $5f$ elements. We now discuss each curve:

- $4d$ orbitals are fairly delocalized, and form bands. Thus, as Z increases, electrons fill the $4d$ band. First the bonding states are filled and then antibonding states. As a consequence, the volume decreases and then increases along the line. This experimental observation is thus coherent with the delocalization of $4d$ electrons.
- For lanthanides, the $4f$ electrons are localized, thus there is a negligible overlap between $4f$ orbitals. As a consequence, adding electrons inside the f-bands has no effect on the cohesion of the crystal. The decrease of the volume is called lanthanide contraction and is due to the poor shielding of nuclear charge by $4f$ orbitals. Eu and Yp have especially large volumes because one bonding electron is transferred to the non-bonding f-shell in order to reach a half-filled or a full f-shell.
- For actinides, the localization appears at the middle of the $5f$ line. One way to understand this is that as Z increases, the $5f$ orbitals are more and more localized because of the attraction with the nucleus. Starting from americium, the system localizes electrons to lower the energy.

This graph suggests that $4f$ electron systems will be closer to the atomic limit. We now discuss the equation of states of the lanthanides as a function of pressure. Refs. [5, 6] show that at ambient pressure and even for moderate pressure, the lanthanides exhibit compact structures. Under pressure, their structure changes becoming more complex and distorted [5, 6]. This suggests that under pressure, f-electrons participate in the bonding as in, e.g., uranium and neptunium (which also have low symmetry equilibrium structures).
We now discuss cerium, which is a model system for lanthanides (see phase diagram in Fig. 3). The α-γ transition in cerium is a first-order isostructural volume collapse transition. The transition line ends at a critical point (CP) around 1.5 GPa and 480 K [7] (see Fig. 3). The (larger

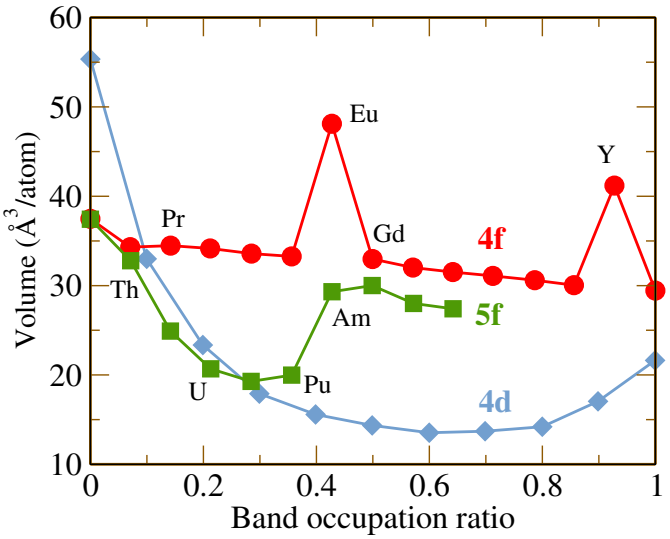


Fig. 2: *Evolution of volume (in Å^3 per atom) of pure elements are a function of the number of electrons for some rows of the periodic table (see also [5]).*

volume) γ-phase exhibits a Curie-Weiss behavior for the magnetic susceptibility while it is Pauli-like in the (smaller volume) α-phase (see Fig. 17). This is interpreted as $4f$ electrons, being localized in the γ-phase, giving rise to local moments and contributing weakly to the electronic bonding, whereas in the α-phase, the $4f$ electrons participate in both bonding [8] and formation of quasi-particles as shown in photoemission spectra [9, 10]. The $\alpha \rightleftharpoons \gamma$-phase transition of Ce is hence a model system for a volume collapse phase transitions due to the delocalization of localized electrons under an increase of pressure or decrease of temperature (see Fig. 3 for other examples in the lanthanides). This is a difference between cerium and other systems, such as praseodymium and plutonium [4], in which the volume collapse occurs directly from the compact structure to lower symmetry structures.

1.3 The Hubbard model

In this section, we will study the Hubbard model, which is appropriate to model the localization/delocalization competition. Using second quantization, one can write the Hamiltonian as

$$H = \sum_i \langle i|h|j\rangle \, c_i^\dagger c_j + \sum_{i,j,k,l} \langle ij|v|kl\rangle \, c_i^\dagger c_j^\dagger c_k c_l \tag{1}$$

In this equation, i, j, k, l belong to a complete one particle basis. If the sums in this equation are restricted to only one orbital per atoms, and if only on site interactions are kept, one recovers the Hubbard model

$$H = \sum_{i \neq j} t_{ij} c_i^\dagger c_j + \sum_i \varepsilon_0 (n_{i\uparrow} + n_{i\uparrow}) + \sum_i U n_{i\uparrow} n_{i\downarrow} \tag{2}$$

In this equation, as we have only one orbital per site, i, j refers to atomic sites. We will now try to understand two limits of this Hamiltonian.

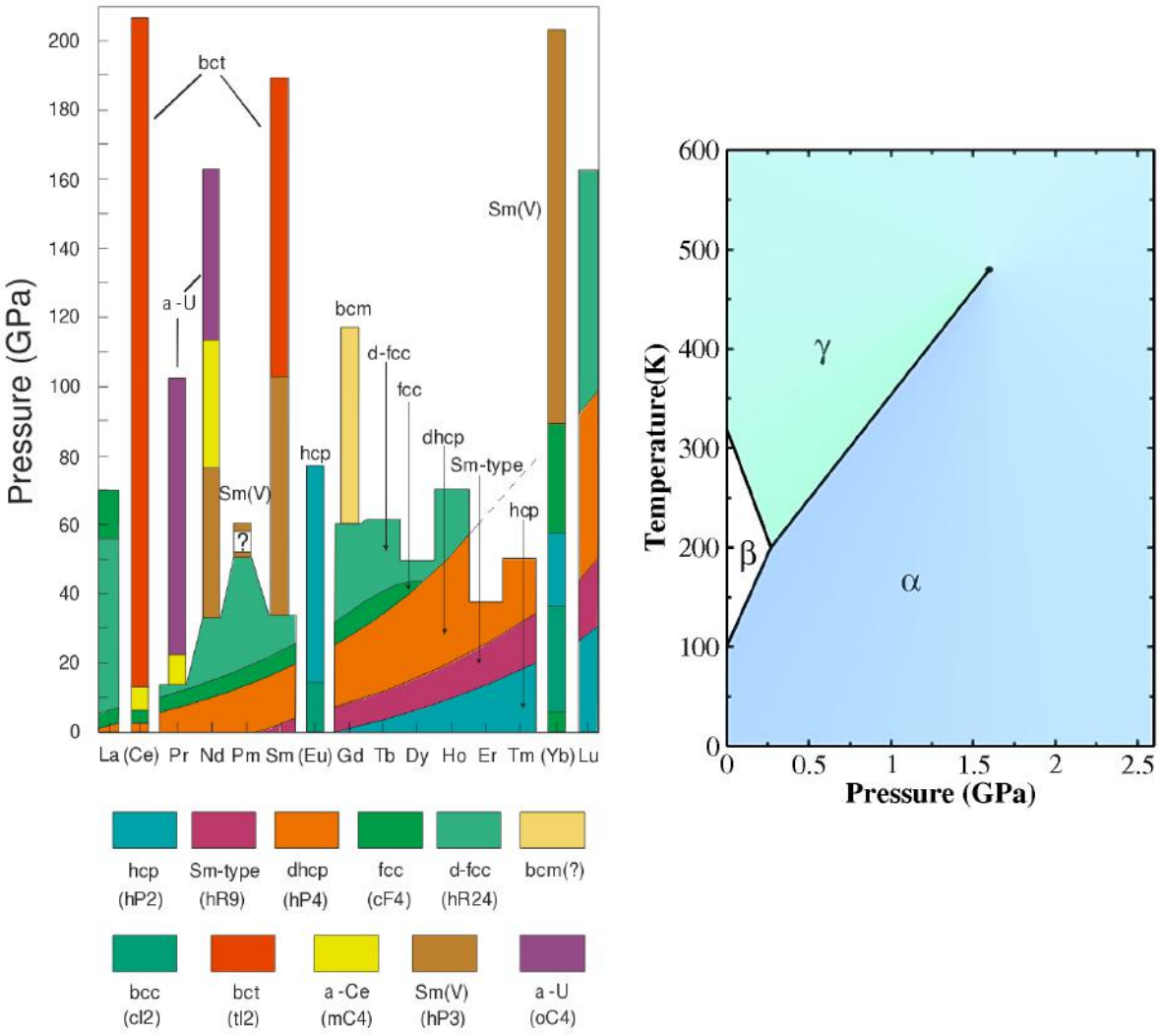


Fig. 3: *Left: Phases of the lanthanides as a function of pressure (From [6]). Right: Phase diagram of cerium.*

1.3.1 Delocalization limit: $U/t \ll 1$

In this limit, we neglect the interaction term.

$$H = \sum_{j \neq i} t_{ij} c_i^\dagger c_j + \varepsilon_0 (n_{i\uparrow} + n_{i\downarrow}) \tag{3}$$

One can easily solve this non-interacting Hamiltonian either by direct diagonalization or by using Bloch states: We can define Bloch states $|k\rangle$ as

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_i |T_i\rangle \ ^{ikT_i} \tag{4}$$

where $|T_i\rangle$ are atomic orbitals: $\langle r|T_i\rangle = \varphi(r - T_i)$ on site i. We also have

$$|T_i\rangle = \frac{1}{\sqrt{N}} \sum_k |k\rangle \ ^{-ikT_i} \tag{5}$$

and the change of basis for creation and annihilation operators

$$c_i^\dagger = \frac{1}{\sqrt{N}} \sum_k c_k^\dagger \ ^{-ikT_i} \qquad c_i = \frac{1}{\sqrt{N}} \sum_k c_k \ ^{ikT_i}. \tag{6}$$

As a consequence, one can show that

$$H = \sum_k \varepsilon_k c_k^\dagger c_k \quad \text{with} \quad \varepsilon_k = \frac{1}{N} \sum_{ij} t_{ij} \ ^{-ik(T_i - T_j)}. \tag{7}$$

In a simple case (one dimension and if t_{ij} is non-zero only for neighboring atoms), we have (note that t is negative) the following dispersion relation which gives the energy of levels as a function of the value of k.

$$\varepsilon_k = \varepsilon_0 + 2t \quad \mathrm{s}(ka) \tag{8}$$

According to the Bloch theorem, k must belong to the first Brillouin Zone $[-\frac{\pi}{}, \frac{\pi}{}]$, where a is the distance between the atoms. We now suppose that we have one electron per atom, whereas there are two states per atoms. As a consequence, the band is half filled. The system is thus metallic and the Fermi levels is at ε_0.

1.3.2 Localization limit: $U/t \gg 1$

$$H = \sum_i U n_{i\uparrow} n_{i\uparrow} + \sum_i \varepsilon_0 (n_{i\uparrow} + n_{i\downarrow}) \tag{9}$$

In this case, the Hamiltonian is a sum of Hamiltonians for independent atoms! We thus have to solve the atomic problem. The size of the Hilbert space for this system is four (2^2) and the states for each number of electrons per atom are

- 0 electron: $|0\rangle : E = 0$
- 1 electron: $|\uparrow\rangle$ and $|\downarrow\rangle$: $E = \varepsilon_0$
- 2 electrons: $|\uparrow\downarrow\rangle$: $E = 2\varepsilon_0 + U$

Let us now suppose that the system has one electron per atom. From the above energies, one can compute the cost associated to the hopping of one electron from one atom to another atom (see Fig. 4). The cost is U. The order of magnitude of U in a real system is a few eV. As a consequence, at room temperature, electrons will be localized and the system is called a Mott insulator. The insulating character comes from interactions and not from a band structure effect. One can also compute the photoemission spectra, i.e., the energy to add ($E(|\uparrow\downarrow\rangle) - E(|\uparrow\rangle) = \varepsilon_0 + U$) or remove ($E(|\uparrow\rangle) - E(|0\rangle) = \varepsilon_0$) an electron. We see that the photoemission spectra (or spectral function) will have two peaks located at ε_0 and $\varepsilon_0 + U$. The two peaks are separated by U, which is thus the photoemission gap. These peaks correspond to the Hubbard bands.

1.3.3 The Mott transition

From these two limits, one can deduce that as U/t will increase, the systems will change from a metal to a Mott insulator. This is the phenomenon of the Mott transition. But we need an approximation to find the solution for intermediate U/t. This is one of the purposes of DMFT.



Fig. 4: *For $U/t \gg 1$, the cost of a hopping is $E(|\uparrow\downarrow\rangle) + E(|0\rangle) - 2E(|\uparrow\rangle) = U$*

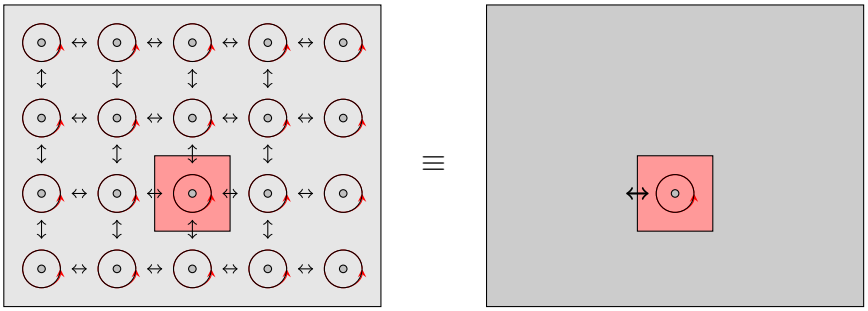


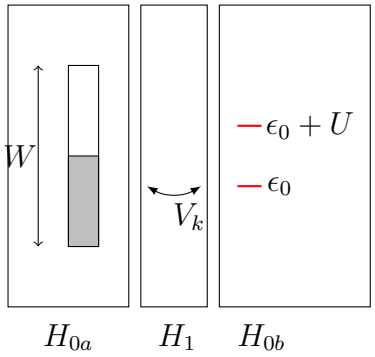
Fig. 5: *In DMFT, the Hubbard model is simplified by an Anderson model. On the left, the Hubbard model is represented schematically. On the right, one atom in red is interacting with the effective field of the others.*

2 Introduction to DMFT and DFT+DMFT

2.1 Dynamical Mean-Field Theory

2.1.1 The Anderson model

We would like to describe correctly the on-site Coulomb interactions which are essential to describe the atomic limit. The first idea of DMFT is to mimic the Hubbard model by a correlated atom with the exact Coulomb interaction, embedded in a bath of uncorrelated orbitals (see Fig. 5). The effect of all the other atoms are gathered in the bath. This is an Anderson model [11] whose Hamiltonian writes



$$H_{\text{Anderson}} = \underbrace{\sum \omega_k \, a^\dagger_{k,\sigma} a_{k\sigma}}_{H_a} + \underbrace{\sum_{k,\sigma} V_k \left(a^\dagger_{k,\sigma} c_\sigma + c^\dagger_\sigma a_{k,\sigma} \right)}_{H_1} + \underbrace{\varepsilon_0 (n_\uparrow + n_\downarrow) + U n_\uparrow n_\downarrow}_{H_b} \tag{10}$$

The use of the Anderson model is a mean-field idea that goes beyond static mean field. Indeed, we do not have only one electron in an effective field as in usual static mean-field theory, but a whole many-body atom whose interactions are taken into account exactly. Moreover the Anderson model will be defined self-consistently, but we will discuss this in the next section. In this section we just focus on the physics of the Anderson model. Once again, we are going to investigate several limiting cases for this Hamiltonian

The isolated atom limit: $V_k = 0$. This case is identical to the limit $U/t \gg 1$ of the Hubbard model: the correlated atom is insulating.

The $U = 0$ limit. The hamiltonian writes

$$H_{\text{Anderson}} = \sum \omega_k \, a^\dagger_{k,\sigma} a_{k\sigma} + \sum_{k,\sigma} V_k \left(a^\dagger_{k,\sigma} c_\sigma + c^\dagger_\sigma a_{k,\sigma} \right) + \varepsilon_0 (n_\uparrow + n_\downarrow).$$

This is again a non-interacting system which can be solved easily: The atomic level will hybridize with the levels at ω_k, and will acquire a width.
Let us compute the Green function of this system. We can use the equation of motion of the (non-interacting) Green function, we thus have

$$\big(\omega I - H\big) G = I. \tag{11}$$

The dimension of these matrices is equal to $N+1$. It contains the N uncorrelated orbitals ω_k and the correlated orbital ε_0. The diagonal elements of H are the energies ω_k and ε_0. The only off diagonal terms are the coupling elements of the level at ε_0 to the levels at ω_k.

$$H = \begin{pmatrix} \varepsilon_0 & V_1 & \cdots & V_k & \cdots & V_N \\ V_1 & \omega_1 & 0 & 0 & 0 & 0 \\ \vdots & 0 & \ddots & 0 & 0 & 0 \\ V_k & 0 & 0 & \omega_k & 0 & 0 \\ \vdots & 0 & 0 & 0 & \ddots & 0 \\ V_N & 0 & 0 & 0 & 0 & \omega_N \end{pmatrix}$$

We can easily invert this matrix and compute the Green function of the correlated orbital (Using $A^{-1} = \quad \mathrm{m}(A)^T / \quad A$ to invert $\omega I - H$). We obtain

$$G(\omega) = \frac{1}{\omega - \varepsilon_0 - \Delta(\omega)} \qquad \text{with} \quad \Delta(\omega) = \sum_k \frac{V_k^2}{\omega - \omega_k} \tag{12}$$

Where $\Delta(\omega)$ is called the hybridization function.
We can now compute the spectral function of this system by computing:

$$A(\omega) = -\frac{1}{\pi} \mathrm{Im} G^R(\omega + i\delta) \tag{13}$$

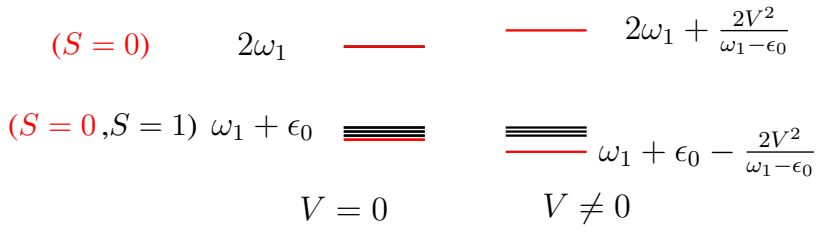


Fig. 6: *Two electrons levels with $V=0$ and V small.*

We need[1]

$$\Delta(\omega+i\delta) = \sum_k \frac{V_k^2}{\omega-\omega_k} - i\pi \sum_k |V_k|^2\, \delta(\omega-\omega_k) \tag{14}$$

If $\Delta = 0$, then the spectral function has a peak at ε_0. Using the last two equations, the peak at ε_0 in the spectral function will be shifted by the real part of Δ and will be broadened by the imaginary part of Δ. Interestingly, the imaginary part of Δ recovers the Fermi golden rule (width of the level coupled to the continuum is $\pi \sum_k |V_k|^2\, \delta(\omega-\omega_k)$).

The Anderson Molecule. We now simplify the bath by replacing it by a single level. We reproduce here the example detailed in Ref. [12, 13].
We suppose that U is infinite, so a double occupation of the localized level is impossible. Finally we limit our study to the states with 2 electrons. With these constraints, we have 5 possible states in the Hilbert space: the first is the double occupation of the bath level (this is a state with S=0) and there are four states with one electron in the bath level and one electron in the localized level. Among these four states, there are one triplet (S=1) and one singlet (S=0). If $V=0$, then the four levels are degenerate. In Fig. 6, the singlet states are in red.
If the hybridization V is non-zero, then the two states with S=0 couple (see Fig. 6). As a consequence we have a possible transition between the singlet and the triplet state at very low energy. This phenomenon has the same physical origin as the Kondo effect. It shows that a resonance exists even in a very strongly correlated system (U large) and this resonance will be at the Fermi level. We can define a temperature T^* corresponding to the difference of energy between the singlet and triplet state. Below this temperature, the system will exhibit the resonance, which will fade as temperature increases. Moreover, below this temperature, the magnetic moment of the localized level is completely cancelled by the formation of the singlet state.

2.1.2 The self consistency condition and the DMFT loop

Following the definition of the Green function in the local orbital basis (i and j)

$$G_{ij}(t) = -i\langle N|T(c_i^\dagger(t)c_j(0)|N\rangle \tag{15}$$

[1] We use $\delta \to \frac{1}{\pi}\frac{\delta}{2}\frac{\delta}{\delta^2} = \delta\ x$.

one can compute the Bloch Green function

$$G_{\mathbf{k}}(t) = \frac{1}{N} \sum_{ij} \; ^{i\mathbf{k}(T_i - T_j)} \, G_{ij}(t) \tag{16}$$

thus, in frequency

$$G_{\mathbf{k}}(\omega) = \frac{1}{N} \sum_{ij} \; ^{i\mathbf{k}(T_i - T_j)} \, G_{ij}(\omega) \tag{17}$$

Using the equation of Motion of the Green function $(\omega - H - \Sigma)G = 1$, the lattice Green function for the Hubbard model is written

$$G_{\mathbf{k}}(\omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma_{\mathbf{k}}(\omega)} \,, \tag{18}$$

where the self-energy is unknown. The local Green function of the lattice is

$$G_{ii}(\omega) = \frac{1}{N} \sum_{\mathbf{k}} \; ^{ik(T_i - T_i)} \, G_{\mathbf{k}}(\omega) = \frac{1}{N} \sum_{\mathbf{k}} G_{\mathbf{k}}(\omega). \tag{19}$$

Besides, the Green function for the Anderson impurity model is

$$G_{\text{Anderson}}(\omega) = \frac{1}{\omega - \varepsilon_0 - \Delta(\omega) - \Sigma(\omega)} \,. \tag{20}$$

The DMFT idea [1] is to identify the local Green function of the Hubbard model with the Green function of Anderson model and the self-energy of the Hubbard model to be equal to the self-energy of the Anderson model:[2] it is the *self-consistency relation* of DMFT. This implies in particular that the local one particle excitations of the Hubbard model will be the same as the one particle excitations of the Anderson model. This writes

$$\frac{1}{N} \sum_{k} \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma(\omega)} = \frac{1}{\omega - \varepsilon_0 - \Delta(\omega) - \Sigma(\omega)} \,. \tag{25}$$

[2] Thus we have expressed also the self-energy in the Bloch basis. We start from the expression of the self-energy as a sum of local self-energies on different sites

$$\Sigma = \sum_{T_i} | \quad \rangle \Sigma \; \omega \; \langle \quad | \tag{21}$$

We use then

$$| \quad \rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} |\mathbf{k}\rangle \mathrm{e} \quad ^{\mathbf{k}T_i} \tag{22}$$

Thus by replacing the last equation in Eq. (21), we have

$$\Sigma = \sum_{\mathbf{k}} |\mathbf{k}\rangle \Sigma \; \omega \; \langle \mathbf{k}' | \tag{23}$$

Thus

$$\langle \mathbf{k} | \Sigma | \mathbf{k} \rangle = \Sigma \; \omega \tag{24}$$

Thus for a local self-energy, the self-energy in the basis of Bloch states is equal to the local self-energy.

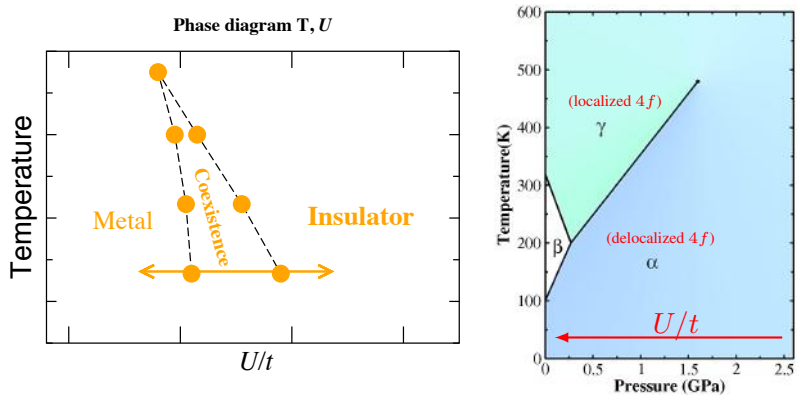


Fig. 7: *Phase diagram of the Hubbard model in DMFT compared to that of Cerium.*

This equation enables us to find $\Delta(\omega)$ as a function of the self-energy $\Delta = \Delta[\Sigma]$ and also $\varepsilon_0 = \frac{1}{N}\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}}$.[3] Besides, the solution of the Anderson model enables us to have the self energy from the value of ε_0 and Δ.[4] So this creates a system of two equations that can be solved self-consistently. These two equations constitute the DMFT self-consistent loop that can be solved by iteration.

2.1.3 The Hubbard model in DMFT

The main results for the Hubbard model phase diagram in DMFT (in infinite dimension, without magnetism and for one electron per atom) are summed up on Figs. 7 and 8. First, the metal insulator transition is a first-order phase transition. Moreover, the paramagnetic insulating phase with a degeneracy of two can be described in contrary to the static mean-field approximation which breaks the symmetry and creates magnetism.

We now describe the evolution of the spectral function. At large value of U, the Hubbard bands are present but still broadened by hybridization. Then, a resonance peak appears at the Fermi level and grows, as U/t decreases. Importantly thus, the metallic phase appears whereas the Hubbard band are still present in the photoemission spectra. Such feature is also not described by the static mean-field approximation. Let us finally outline that in DMFT, both the atomic limit ($U/t \gg 1$) and the delocalization limit ($U/t \ll 1$) are exact.

We compare on Figs. 7 and 8 the physics of the Hubbard model to the physics of an f electron system, namely pure cerium. As was discussed above, pure cerium exhibits two phases, one with delocalized f-electrons and one with localized electrons and the transition between these two phases can be controlled with pressure or temperature. It can be rationalized in comparison to Hubbard model: applying pressure is equivalent to reducing t. Besides, phases of cerium both exhibit Hubbard bands. The α-phase, exhibits less intense Hubbard band and larger quasi-particle peak, as expected for more delocalized electrons. In the next section, we will discuss DFT+DMFT in order to apply DMFT ideas to a real system such as cerium.

[3] Equivalently, one can say that the non-interacting Green function $\mathcal{G}_0 = \omega - \varepsilon_0 - \Delta(\omega)$ can be obtained from the self-energy.

[4] Equivalently, one can obtain the Green function and the self energy of the Anderson model as a function of the non-interacting Green function

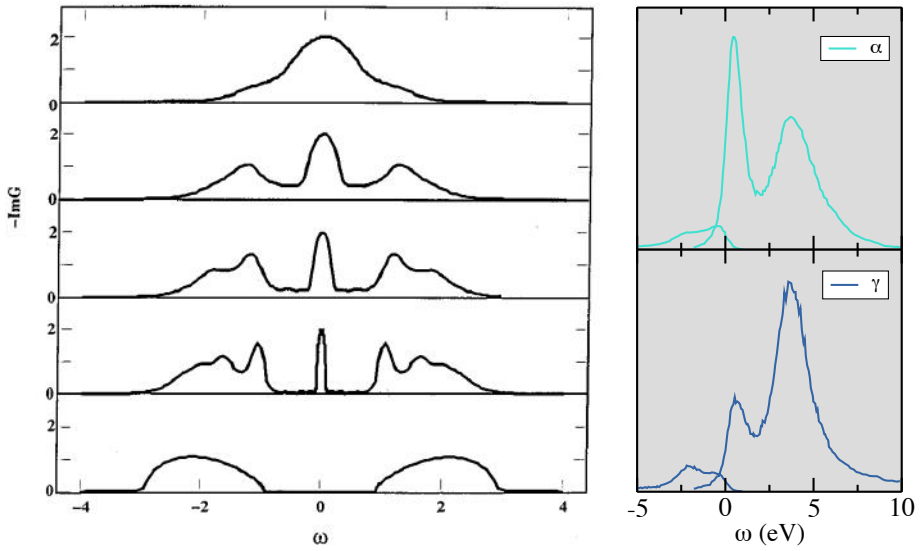


Fig. 8: *Evolution of the spectral function in the Hubbard model (from Ref [1]) compared to photoemission spectra in cerium (From Ref. [9, 10])*

2.2 DFT+DMFT

The rough idea of DFT+DMFT is to use the DFT/LDA Hamiltonian to define the one body term of the Hubbard model and then supplement it by an exact Coulomb interaction for the correlated orbital subset. However, to apply DMFT ideas to a real system, one need first to define correlated orbitals [14–16] and a corresponding value of the effective interaction U [17]. This is important and is discussed in the next subsection. I just outline here that there is no unique way to define correlated orbitals. So several choices have been made in the literature. They include atomic orbitals, or Wannier orbitals. When the choice of correlated orbitals is done, the effective interaction for these orbitals has to be computed. One formalism is the constrained Random Phase Approximation (cRPA) method [17] which considers that the effective interaction between correlated electrons is the bare interaction screened by all the non-correlated electrons. It thus requires the calculation of a non-interacting polarizability and the inverse dielectric function. It is described in the second subsection.

Then, using correlated orbitals and related values of U, a DFT+DMFT calculation can be carried out. The scheme will be presented in the third subsection. Lastly, I will briefly discuss solvers for the Anderson model in the fourth subsection.

2.2.1 Correlated orbitals: how to define them?

In this section, we briefly discuss Wannier functions, which are a flexible tool to define correlated orbitals.

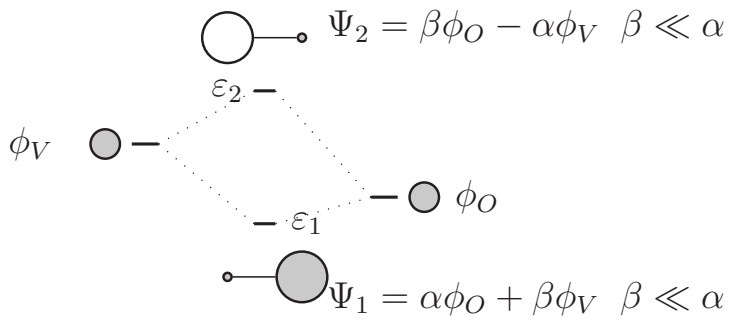


Fig. 9: *Simple molecular orbital diagram for an diatomic molecule containing an electropositive atom (V) and an electronegative atom (O). It can be viewed as a simple model for $SrVO_3$ or UO_2 or Ce_2O_3. In the text, we use this model with 3 electrons: the ε_2 level is thus half-filled.*

Projected Local Orbitals Wannier functions and Maximally Localized Wannier functions. Projected Wannier functions are defined in Refs. [14, 16]. We first introduce the auxiliary wavefunctions $|\tilde{\chi}^{\mathbf{R}}_{\mathbf{k}m}\rangle$ as

$$|\tilde{\chi}^{\mathbf{R}}_{\mathbf{k}m}\rangle \equiv \sum_{\nu\in\mathcal{W}} |\Psi^{\sigma}_{\mathbf{k}\nu}\rangle\langle\Psi^{\sigma}_{\mathbf{k}\nu}|\chi^{\mathbf{R}}_{\mathbf{k}m}\rangle. \tag{26}$$

For a given atomic site $\mathbf{R}$, we call $|\chi^{\mathbf{R}}_{\mathbf{k}m}\rangle$ the Bloch transform of isolated atomic orbitals with projected angular momentum m. $|\Psi^{\sigma}_{\mathbf{k}\nu}\rangle$ are the Kohn-Sham orbitals for k-point $\mathbf{k}$, band index ν and spin σ. $|\tilde{\chi}^{\mathbf{R}}_{\mathbf{k}m}\rangle$ is thus a weighted sum of Kohn-Sham orbitals. This sum extends over a given number of Kohn-Sham orbitals that can be defined by an index range or alternatively by an energy window $\mathcal{W}$. For a finite $\mathcal{W}$, the orthonormalization of $|\tilde{\chi}^{\mathbf{R}}_{\mathbf{k}m}\rangle$ leads to well defined Wannier functions $|w^{\mathbf{R}}_{\mathbf{k}m}\rangle$, unitarily related to $|\Psi^{\sigma}_{\mathbf{k}\nu}\rangle$ by[5]

$$|w^{\mathbf{R}}_{\mathbf{k}m}\rangle = \sum_{\nu\in\mathcal{W}} |\Psi^{\sigma}_{\mathbf{k}\nu}\rangle\langle\Psi^{\sigma}_{\mathbf{k}\nu}|w^{\mathbf{R}}_{\mathbf{k}m}\rangle. \tag{27}$$

It is important to notice that the localization of $|w^{\mathbf{R}}_{\mathbf{k}m}\rangle$ will decrease if $\mathcal{W}$ decreases. Maximally localized Wannier functions are Wannier functions whose extension is minimized [18]. They have the advantage over projected local orbital Wannier functions that they are uniquely defined.

Wannier functions: a pedagogical simplified molecular model In order to illustrate the localization of Wannier functions, we consider a simple diatomic molecule containing an electropositive atom (V) and an electronegative atom (O). In this model molecule, we consider only one orbital per atom, and for simplification, we assume it is of s symmetry. It can be viewed as a very simple model for $SrVO_3$ or UO_2. The bonding state at ε_1 is the analogue of the O-p like band, whereas the antibonding state at ε_2 is the analogue of the V-d like bands.
We now suppose that the system contains 3 electrons. The V-like band is thus half filled. We use Eq. (26) with $\chi = \phi_V$ to compute Wannier functions for two energy windows:

[5] In the limit of a infinite number of Kohn-Sham bands, the projection in Eq. (26) becomes complete and the Wannier functions $|w^{\mathbf{R}}_{\mathbf{k}}\rangle$ become equivalent to atomic orbitals $|\chi^{\mathbf{R}}_{\mathbf{k}}\rangle$.

Eigenstates included in the energy window $\mathcal{W}$	$\{\varepsilon_2\}$	$\{\varepsilon_1, \varepsilon_2\}$
Corresponding Wannier function w_V	$\|\Psi_2\rangle$	$\|\phi_V\rangle$
Corresponding number of electrons n_V	1	$2\beta^2+\alpha^2$

Table 1: *Wannier function on vanadium and corresponding numbers of electron in a simple diatomic molecule for two different choices of the energy window.*

- If $\mathcal{W} = \{\varepsilon_2\}$, then Eq. (26) contains only one term: $|\widetilde{\chi}\rangle = |\Psi_2\rangle\langle\Psi_2|\phi_V\rangle = \alpha|\Psi_2\rangle$. After renormalizing $\widetilde{\chi}$, we find that the Wannier function is a molecular orbital $|w_V\rangle = |\Psi_2\rangle$ that contains an oxygen contribution. The number of electrons in this orbital is 1.
- If $\mathcal{W} = \{\varepsilon_1, \varepsilon_2\}$, then Eq. (26) contains two terms. One shows easily that the Wannier function is a localized atomic orbital $|w_V\rangle = |\phi_V\rangle$ and is thus much more localized than in the previous case. The number of electrons in $2\beta^2+\alpha^2 \neq 1$.

So this simple example illustrates how the localization of the Wannier functions is modified by the choice of the energy windows of Kohn-Sham bands. A similar illustration in the case of $SrVO_3$ can be seen in [15]. We will now discuss the calculation of effective interactions for these orbitals.

2.2.2 How to compute the effective coulomb interaction?

There are several ways to compute effective interactions [19, 20, 17]. Here, we will briefly present the constrained random phase approximation method [17, 21] (cRPA) which is particularly adapted to DFT+DMFT.
The cRPA method depends on the definition of a many-body model [21]. The many-body model is defined by a set of local orbitals together with the interactions among them. A cRPA calculation of a model would require first the definition of a set of local orbitals, and second a consistently calculated screened interaction: it considers that the effective interaction between correlated electrons is the bare interaction screened by all the non-correlated electrons. Importantly, thus the value of effective interactions depends on the definition of correlated orbitals discussed above.

The cRPA method We call here χ_0 the non-interacting (Kohn-Sham) polarizability of the system. Let's now separate the correlated states (They could be f-states but the method is more general and correlated orbitals could gather several orbitals from, e.g., different atoms) from the rest (r). We thus have [17]

$$\chi_0 = \chi_0^{\text{correl}} + \chi_0^r. \tag{28}$$

Thus, we can rewrite the inverse dielectric matrix as

$$\epsilon^{-1} = \frac{1}{1 - v(\chi_0^{\text{correl}}+\chi_0^r)}\,. \tag{29}$$

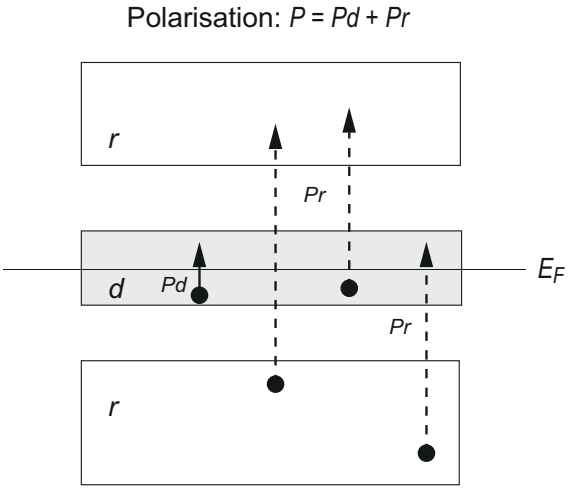


Fig. 10: *Schematic density of states with the transitions that are involved in the polarizability of correlated electrons $P_d \hat{=} \chi_0^{correl}$ and the other transitions that are involved in $P_r \hat{=} \chi_0^r = P - \chi_0^{correl}$ (From [22]).*

We now define the dielectric functions due to non-correlated electrons and the corresponding screened interaction

$$\epsilon_r^{-1} \hat{=} \frac{1}{1 - v\chi_0^r} \quad \text{and} \quad W_r = \epsilon_r^{-1} v = \frac{v}{1 - v\chi_0^r} \,. \tag{30}$$

The dielectric function due to correlated electrons acting on W_r is thus

$$\epsilon_{\text{correl}}^{-1} \hat{=} \frac{1}{1 - W_r \chi_0^{\text{correl}}} \,. \tag{31}$$

With these definitions, one shows that

$$\epsilon_{\text{correl}}^{-1} \epsilon_r^{-1} = ... = \frac{1}{1 - v\chi_0^r - v\chi_0^{\text{correl}}} = \epsilon^{-1} \quad \text{and thus} \quad W \hat{=} \epsilon^{-1} v = \epsilon_{\text{correl}}^{-1} \epsilon_r^{-1} v. \tag{32}$$

The fully screened RPA interaction is the combination of two screening processes. First, the bare interaction is screened by non-correlated electrons (r), giving rise to a screened interaction W_r. Secondly, the screening of this interaction by correlated electrons recovers the fully screened interaction [17].

The main idea of cRPA is to use this separation of screening processes. The screened interaction W_r can be computed explicitly, using the equation written above, which is valid for electrons interacting only with the Hartree potential. Removing the screening by correlated electrons is easy in cases where the correlated orbitals are unitarily related to an isolated set of bands (see Fig. 10). In other cases, several schemes have been proposed.

Then the interaction in the basis of correlated orbitals is computed with

$$U_{1234}(w) = \langle w_1 w_2 | W_r(w) | w_3 w_4 \rangle. \tag{33}$$

Finally, the screening by correlated electrons is explicitly taken into account by solving exactly the correlations with the interaction U_{1234} by a dedicated method (e.g. DMFT).

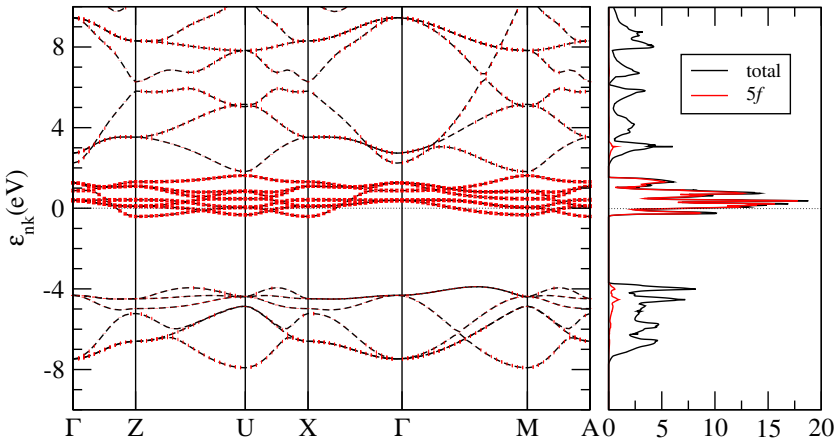


Fig. 11: *Band structure of UO_2 in DFT/LDA [23]. Between –8 eV and –3 eV, bands are mainly O-p like. Then, there are 14 bands of $5f$ character that are at the Fermi level. If Wannier correlated orbitals are constructed only with f-like bands, then, the cRPA interactions can be computed by removing transitions inside the 14 isolated f-bands.*

Many-body models We now briefly consider different kinds of models as discussed in Ref. [21, 25] in the case of UO_2. UO_2 is a paramagnetic insulator above the Néel temperature of 30 K with the fluorine structure.

- f-f model: The model is built from the U-f like bands only: Wannier functions are built from U-f like bands and transitions among the U-f like bands define χ^{correl}.
- fp-fp model: The model is built from the U-f like and O-p like bands. Wannier functions are build from U-f like bands and O-p like bands and transitions among these bands define χ^{correl}. In this case, effective interactions U_{ff}, U_{fp} and U_{pp} can be computed and used in a calculation using explicitly all these interactions.

We now first discuss the value of the interactions. We can first compare the bare interactions for $SrVO_3$ (19.1 eV), cerium (24 eV) and UO_2 (16.1 eV) [23]. It highlights the large localization of orbitals in lanthanides and the fact that interactions in actinides are lower.
We now discuss effective interactions U. The value of U depends largely on the model [23]. It is larger in the fp mode. This comes from the fact that the orbitals are more localized in the fp model and also that screening is lower. Ref [24] discusses the appropriate choice of U in uranium. However, it is expected to be between 4 and 6 eV in UO_2 to reproduce spectral properties [26]. In lanthanides, the values of U required to describe experiments are larger.

model	v	U
f-f	16.0	3.4
fp-fp	17.1	6.2

Table 2: *Bare (v), and cRPA (U) Coulomb interactions for UO_2 computed for different models. Results are from [23] (see also [24])*

2.2.3 The DFT+DMFT scheme

We now suppose that correlated orbitals and effective interactions are defined. We now discuss the DFT+DMFT scheme. In a first step, we are going to write the expression of the lattice Green function for the solid with a local self-energy. The self-energy will later be the DMFT self-energy.

The lattice Green function In comparison to the case of the one-band Hubbard model,

- the number of correlated orbitals is larger (14 for f-elements),
- the number of bands in the system is large, and larger than the number of correlated orbitals because some orbitals are not correlated. For example, in cerium, the $4f$ orbitals are correlated but the $6s$ are not. However, $6s$ and $4f$ states are strongly mixed in the DFT/LDA density of states.

We are going to express all quantities (G, Σ) in the Kohn-Sham wavefunctions basis because the DFT Hamiltonian has a simple expression in this basis

$$H_{\text{LDA}} = \sum_{n\mathbf{k}} |\Psi_{\mathbf{k}n}\rangle \, \varepsilon_{n\mathbf{k}} \, \langle \Psi_{\mathbf{k}n}| \,. \tag{34}$$

The self-energy in DMFT is local and is computed in the local basis (of correlated orbitals), thus the self energy operator is a sum of identical self-energies on each correlated site (m and m' are the quantum numbers of the projection of the angular momentum)

$$\Sigma = \sum_{mm'\mathbf{T}_i} |\chi_{\mathbf{T}_i m}\rangle \, \Sigma_{m,m'}(\omega) \, \langle \chi_{\mathbf{T}_i m'}| \,. \tag{35}$$

We can define the Bloch transform of the localized functions $\chi_{\mathbf{T}_i m}$ as before as

$$|\chi_{\mathbf{T}_i m}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} |\chi_{\mathbf{k}m}\rangle \;\; ^{-i\mathbf{k}\mathbf{T}_i}. \tag{36}$$

Thus by replacing the last equation in Eq. (35), we have

$$\Sigma = \sum_{mm'\mathbf{k}} |\chi_{\mathbf{k}m}\rangle \, \Sigma_{m,m'}(\omega) \, \langle \chi_{\mathbf{k}m'}| \,. \tag{37}$$

It can be expressed in the Kohn-Sham basis directly using the fact that $\langle \Psi_{\mathbf{k}n} | \chi_{\mathbf{k}'m} \rangle$ is zero if $\mathbf{k}$ and $\mathbf{k}'$ are different,

$$\Sigma_{nn'}(\mathbf{k}, \omega) = \sum_{m,m'} \langle \Psi_{\mathbf{k}n} | \chi_{\mathbf{k}m} \rangle \, \Sigma_{mm'}(\omega) \, \langle \chi_{\mathbf{k}m'} | \Psi_{\mathbf{k}n'} \rangle. \tag{38}$$

In order to avoid a double-counting of the correlation for correlated orbitals, the self-energy has to contain a so called "double-counting term" which should cancel the DFT/LDA Hartree and exchange-correlation potential for the correlated electrons

$$\Sigma = \Sigma_{\text{DMFT}} - \Sigma_{\text{DC}}. \tag{39}$$

The Green function will obey the usual equation of motion:

$$\big(\omega I - H - \Sigma\big) G = I. \tag{40}$$

Thus the Green function is

$$G_{nn'}(\mathbf{k},\omega) = \big(\omega I - H(\mathbf{k}) - \Sigma(\mathbf{k},\omega)\big)^{-1}\Big|_{nn'}, \tag{41}$$

where $H(\mathbf{k})$ is a diagonal matrix containing the Kohn-Sham eigenvalues and $\Sigma(\mathbf{k},\omega)$ is non-diagonal (see above). The local Green function is simply

$$G^{\text{loc}}_{mm'}(\omega) = \sum_{\mathbf{k}} \langle \chi_{\mathbf{k}m} | \Psi_{\mathbf{k}n} \rangle \, G_{nn'}(\mathbf{k},\omega) \, \langle \Psi_{\mathbf{k}n'} | \chi_{\mathbf{k}m'} \rangle. \tag{42}$$

This equation is the generalization of Eq. (19).

The Self-Consistency Condition The DMFT self-consistency relation equals the local Green function and the Green function of the Anderson model:

$$G^{\text{loc}}_{mm'}(\omega) = G^{\text{Anderson}}_{mm'}(\omega) \tag{43}$$

where

$$G^{\text{Anderson}}_{mm'}(\omega) = \big(\omega I - E_0 - \Delta(\omega) - \Sigma(\omega)\big)^{-1}\Big|_{mm'}. \tag{44}$$

E_0 is a diagonal matrix with the levels of correlated orbitals in the (multiorbital) Anderson model, Σ and Δ are the self-energy and hybridization matrices in the correlated orbital basis. E_0 and Δ are obtained from the self-consistency condition.

The Anderson impurity model E_0, Δ and U, J define the Anderson model. The solution of the Anderson model gives the local self-energy, which is used again to compute the lattice Green function, Eq. (41).

The DFT loop The density can be obtained from the full Green function using:

$$n(\mathbf{r}) = -i \sum_{n,\mathbf{k}} \Psi_{\mathbf{k}n}(\mathbf{r}) \, G_{nn'}(\mathbf{k}, t - t' = 0^-) \, \Psi_{\mathbf{k}n'}(\mathbf{r}) \tag{45}$$

and is used to reconstruct the Kohn-Sham Hamiltonian (a functional of the density) and thus the eigenvalues. The scheme is summarized in Fig. 12.

2.2.4 Solvers for the Anderson impurity model

The resolution of the Anderson impurity model, Eq. (10), can be done by several methods. Simplifications can be done in some cases. In some localized systems (e.g. lanthanide oxides), one can neglect the bath in the resolution of the Anderson model. In this case, the hybridization and crystal field play a role only through the renormalization of the atomic levels. This is the basis of the Hubbard I method.

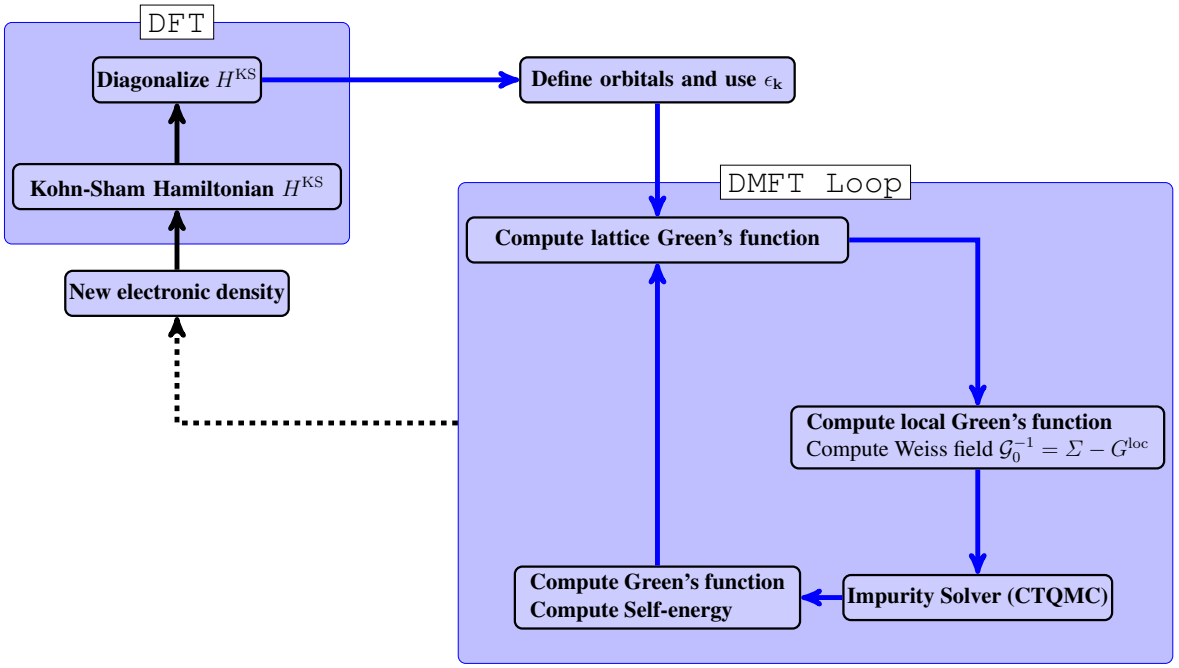


Fig. 12: *Scheme of the DFT+DMFT loops. $\mathcal{G}_0$ is the non-interacting Green function of the Anderson model and is linked to the hybridization by $\mathcal{G}_0^{-1} = \omega I - E_0 - \Delta(\omega)$. The Dyson Equation is $\mathcal{G}_0^{-1} - G^{-1} = \Sigma$.*

A more drastic solution is the static mean-field solution, which is equivalent to the DFT+U method [19]. To briefly present this approach, we suppose that density matrix of correlated orbitals is diagonal and $J = 0$. In this case, the self-energy for orbital i is simply $U(n_i - 1/2)$. It has the effect of opening a gap in Mott insulators by shifting occupied states downwards and empty states upward (see also [27]). However, even the atomic physics is not recovered and e.g paramagnetic insulators cannot be described with this approach. Some approximations can also be made if the system is on the metallic side (e.g. the Gutzwiller approximation).

In the general case, the solution can be provided by quantum Monte Carlo methods. The main idea is that the atomic problem can be solved exactly and the bath problem can be solved exactly. Most continuous-time quantum Monte Carlo codes, but not all, use an expansion as a function of H_1 in Eq. (10) and sample all relevant terms in the expansion.

2.2.5 Conclusion on DFT+DMFT

In this section, I have first reminded that the Hubbard model describes localization/delocalization transitions. Then I have presented the mapping to the Anderson model, and the DMFT equations. For real systems, I underlined that correlated orbitals and effective interactions have to be defined. Lastly, I detailed the DFT+DMFT scheme. The next section is devoted to applications to f-electron systems.

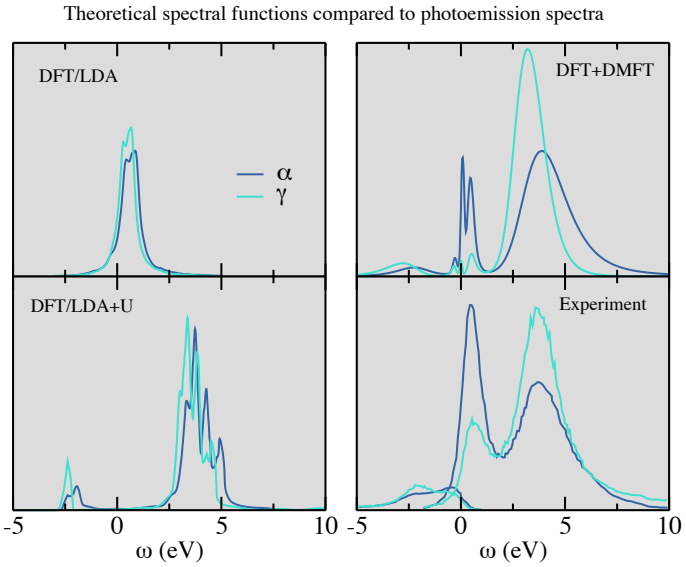


Fig. 13: *Theoretical spectral functions of cerium compared to photoemission spectra [9, 10]. Both the LDA and LDA+U spectra [28] are unable to differentiate the two phases of cerium. LDA+DMFT [29] describes qualitatively the appearance of the noticeable quasiparticle peak in the α-phase.*

3 f-electron systems: pure metals

3.1 Lanthanides

As discussed in the first section, f-electrons in lanthanides are localized and do not participate to the bonding at ambient pressure. So indeed, as DFT+DMFT correctly describes the atomic limit and the metal insulator transition, it should be adopted to describe these systems and their behavior under pressure. We will review in this section, photoemission, optical properties, magnetic and structural properties as described by DFT+DMFT on lanthanides.

3.1.1 Photoemission spectra

In this section, we discuss photoemission spectra of two lanthanides: cerium and terbium. For cerium, direct and inverse photoemission spectra were measured for both phases [9, 10] and are reproduced in Fig. 13. Importantly, both phases exhibit high-energy structures that can be interpreted as Hubbard bands. Only the α-phase has a sizable peak at the Fermi level, which suggests that the $4f$ electron is partly delocalized. We now compare to spectral functions obtained in DFT, DFT+U, and DFT+DMFT. We see that only DFT+DMFT is able to describe the important differences between the spectra of the α- and γ-phases, in contrast to LDA+U and LDA. LDA only describes the quasiparticle peak whereas LDA+U only opens a gap in the f-bands and thus describes only Hubbard bands.

We now discuss terbium spectra. Locht *et al.* [30] have computed photoemission spectra of all lanthanides using DFT+DMFT. At ambient pressure as discussed above, electrons are localized

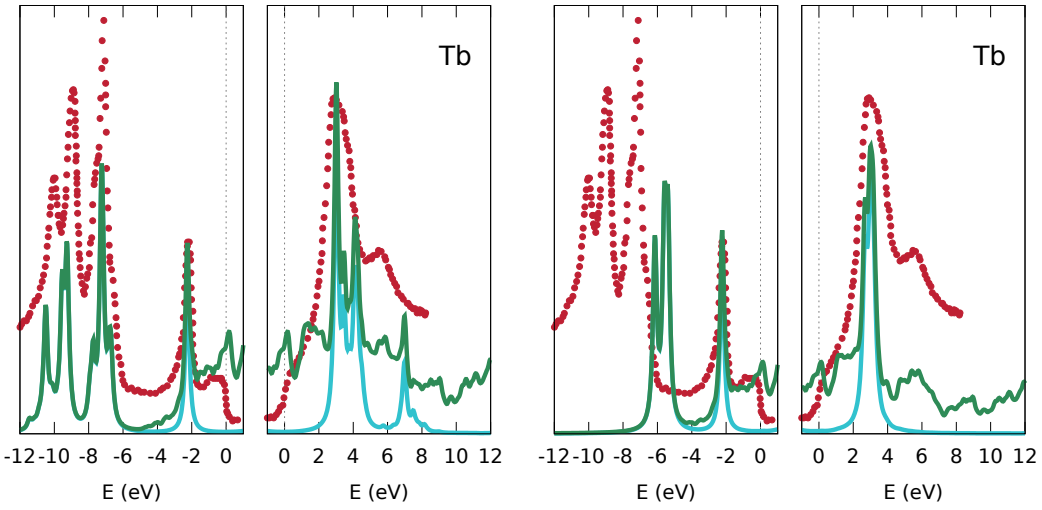


Fig. 14: *Comparison between experimental spectra (red dots) and the spectra calculated with DFT+DMFT (left) and with DFT+U (right). The green line is the calculated total spectral function. The blue line is the calculated $4f$ contribution to the spectral function [30].*

so the authors used the Hubbard I approximation. Fig. 14 shows theoretical spectra computed in DFT+DMFT and DFT+U compared to experimental spectra in the example of terbium. The direct photoemission spectrum contains peaks corresponding to the f^7 multiplets whereas inverse photoemission corresponds to f^9 states. These atomic features are correctly described by DMFT, provided that effective interactions and double-counting is correctly adjusted. In comparison, static mean-field LDA+U describes the system with only one Slater determinant and cannot reproduce the experimental data.

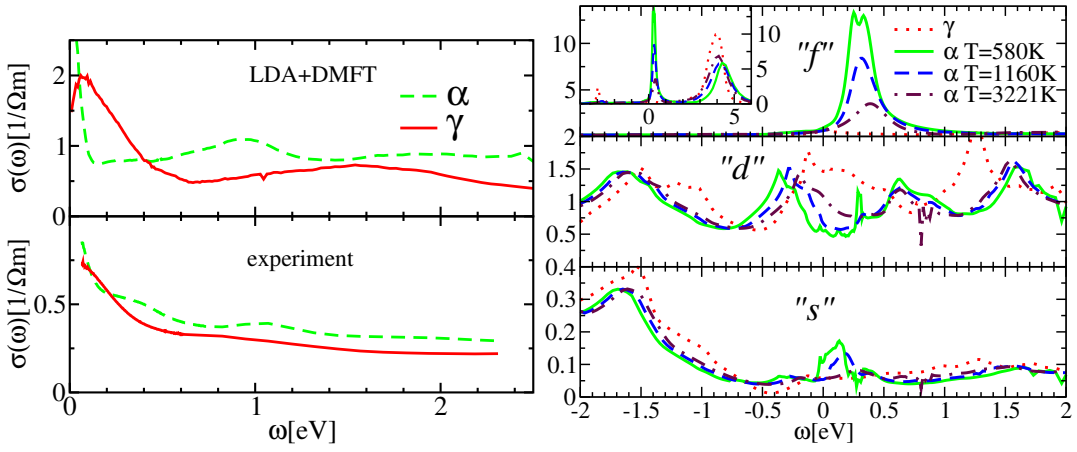


Fig. 15: *(left): calculated and experimental optical conductivity of phases of cerium (from [31]). (right): from top to bottom: f and d spectral functions, and hybridization function [31].*

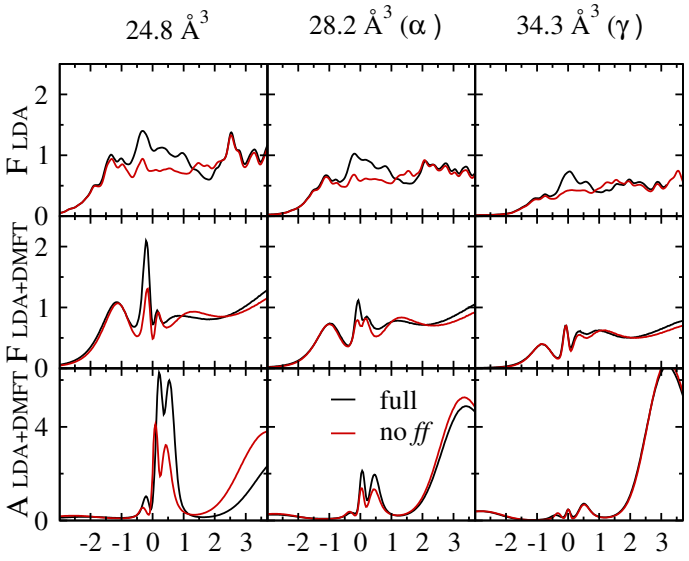


Fig. 16: *LDA and LDA+DMFT hybridizations and LDA+DMFT spectral function for three volumes of fcc cerium, including the α- and γ-phases [32].*

3.1.2 Optical conductivity

In Fig. 15, the experimental optical conductivity spectra of α- and γ-cerium are compared to DFT+DMFT calculations [31]. The DFT+DMFT calculations successfully reproduce the larger Drude peak in the γ-phase and the 1 eV peak in the α-phase spectra. The right panel of Fig. 15 provides insight into these spectra and the origin of these features. Indeed, at low temperature, when the quasiparticle f-peak at the Fermi level is large, the hybridization between f- and d-electrons creates a pseudo-gap in the d density of states, explaining the 1 eV peak in the α-phase. The large Drude peak at the Fermi level originates from the unhybridized d spectral function. This peak is visible mainly in the γ-phase where there is no pseudo-gap and thus an important d-weight at the Fermi level.
Two simplified models exist to describe the α-γ phase transition in cerium. They differ by the orbitals involved in the delocalization of f-electrons under pressure. The Mott transition assumes the f-f hoppings variation is most important, whereas in the Kondo Volume Collapse, it is the f-c hoppings that are dominant (c stands for s, p or d orbitals). The above analysis suggests that the main hybridization that drives the delocalization of f-electrons under pressure is the f-d hybridization. This is compatible with the qualitative view of strongly localized f-orbitals that have weak overlap.

3.1.3 Electronic model for the transition

To complement the above analysis, it is possible to simulate photoemission spectra with or without f-f hoppings to check the importance of this term. Fig. 16 shows the evolution of the $4f$ hybridization computed in LDA (top row), and LDA+DMFT (middle row), and the spectral functions computed in LDA+DMFT (bottom row) as a function of volume (in black). In

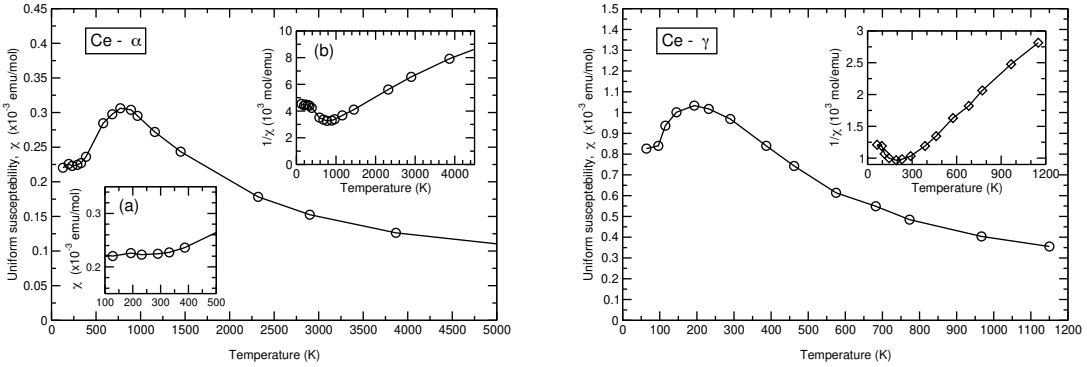


Fig. 17: *Magnetic susceptibility computed in DFT+DMFT [34] as a function of temperature for α- and γ-phases of cerium. We can analyze these data with the help of the phase diagram of Ce (Fig. 3). The magnetic susceptibility is constant as a function of temperature in the α-phase (below 500 K), whereas it has a Curie-Weiss behavior in the γ-phase (above 150 K). The γ-phase has thus a local moment. This is in qualitative agreement with experiment [33].*

red the same quantity is plotted without f-f hoppings [32]. The comparison of the hybridizations shows that f-c hybridization is clearly dominant at all volumes. Nevertheless, for lower volumes, the f-f hoppings play an increasing role in the amplitude of the hybridization. The amplitude of the quasiparticle peak is also impacted by the f-f hopping. This highlights that both hybridizations have to be taken into account and thus the need to go beyond models for a quantitative description of cerium.

3.1.4 Magnetism

Fig. 17 reproduces the magnetic susceptibility of cerium calculated within DFT+DMFT [34]. It exhibits a Curie-Weiss susceptibility at high temperature and low pressure in the γ-phase. This is in agreement with the spectral functions, that show mainly Hubbard bands and no quasiparticle peak, and suggest that this system is localized. In the α-phase, the quasiparticle peak is present, and a Pauli-like behavior in the α-phase (at low temperature) is observed. This is in qualitative agreement with experiments from Ref. [33].

3.1.5 Equation of states

Regarding cerium, numerous studies have extensively investigated the thermodynamics of its transition, highlighting the profound interplay between thermodynamic and electronic properties within the framework of density functional theory combined with dynamical mean-field theory (DFT+DMFT). Notably, these investigations have clearly demonstrated that the presence of a quasiparticle peak in the α-phases contributes to the stabilization of the phase at low temperatures. The inclusion of entropy and spin-orbit coupling has been found crucial for accurately describing the equation of states. In Fig. 18, the internal energy, entropy, and free energy at the transition are depicted, both with and without consideration of spin-orbit coupling. Let us start by discussing the internal energy. In both sets of calculations conducted within the volume range of the transition, we observe a softening of the energy's second derivative, indicating a

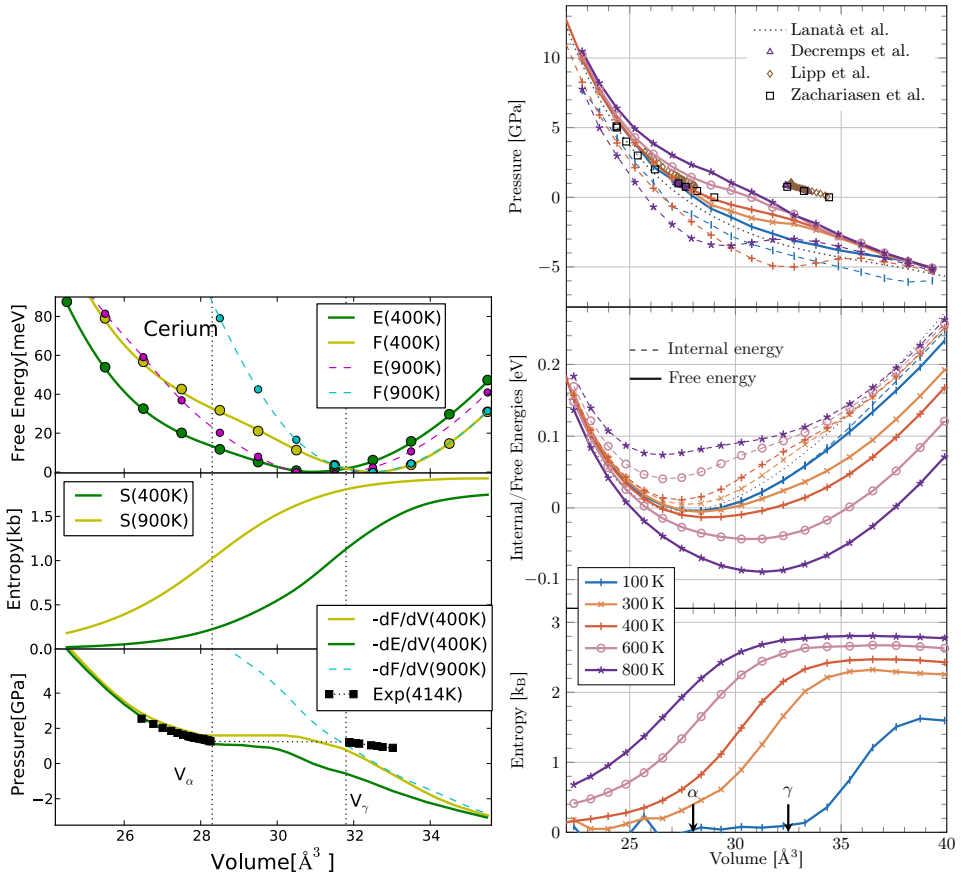


Fig. 18: *Internal, free energy, entropy, and pressure as a function of volume compared to experiment with spin-orbit coupling [35] and without spin-orbit coupling [29].*

decrease in the bulk modulus. However, it is crucial to consider the contribution of entropy, as its magnitude plays a significant role. It is worth noting that this magnitude is influenced by the degeneracy of the f-levels, which in turn depends on the inclusion of spin-orbit coupling. Finally, the resulting free energy exhibits a negative curvature, which is an indication of the presence of a phase transition. Notably, this distinctive feature is observed solely in the calculation that incorporates spin-orbit coupling. In this particular case, one can deduce the equation of state, which demonstrates a good agreement with the experimental equation of state.

3.2 Actinides

3.2.1 Structural properties

As discussed above, the volumes of actinides show a jump along the line. Uranium and neptunium have a low volume and a low symmetry structure whereas americium and curium have a large volume and compact phases. So plutonium is at the transition between two types of behavior. Depending on the temperature, it can exhibits low symmetry phase with low volume (α) or compact phase with high volume (δ). Concerning magnetism, all phases are non-magnetic, and there is no local moment except for curium.

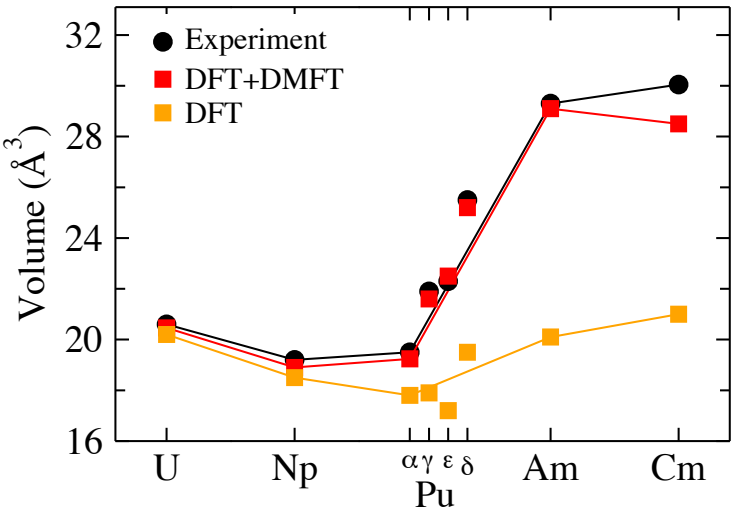


Fig. 19: *Volumes of uranium, neptunium, various phases of plutonium, americium and curium computed in non-magnetic GGA and DFT+DMFT [36].*

We now discuss theoretical results [36] reproduced on Fig. 19. GGA fails to recover the volume jump in actinides, unless an artificial magnetism is taken into account [36]. On contrary, DFT+DMFT is able to describe this transition. Interestingly spin-orbit coupling and Hund's exchange play a key role in this transition. In particular it is noticeable that the volume jump happens near the complete filling of the $J=5/2$ shell. This can be interpreted as an enhancement of correlation near the filling of this shell.

3.2.2 Branching ratio and importance of spin-orbit coupling

X-ray absorption spectroscopy (XAS) is especially useful to study f-electron systems because, as we will see below, it gives indication on the strength of the spin-orbit coupling (SOC). We focus here on the $4d$ to $5f$ transition. Spin-orbit coupling is very important for core states. So $4d$ orbitals are largely split in $4d_{5/2}$ and $4d_{3/2}$. $5f$ orbitals are less split by SOC. The absorption intensities $I_{5/2}$ and $I_{3/2}$, which correspond to $4d_{5/2} \rightarrow 5f_{5/2,7/2}$ and $4d_{3/2} \rightarrow 5f_{5/2}$ transitions, enable to define the branching ratio as $B = I_{5/2}/(I_{5/2}+I_{3/2})$. One can show that it is directly linked to the average of the spin-orbit operator. We can thus compare the experimental values of this operator to exact limits computed in the Russel-Saunders coupling scheme or the jj coupling scheme, and to ab-initio calculations. In the LS (Russel-Saunders) picture (see Fig. 20), we have in particular a cancellation of the spin-orbit coupling when $L=0$ for $n_f=7$. Whereas in jj coupling, the value increases until the $j=5/2$ shell is filled, after which it decreases.

Experimentally, the value is rather close to the jj case, except near $n_f=7$ where the stabilization of the half filled shell decreases the value of the spin-orbit operator. In Fig. 20, we compare the experimental value to LDA and LDA+DMFT calculations [37]. LDA overestimates the bandwidth, so the value of the spin-orbit coupling is underestimated. LDA+DMFT improves the agreement with experiment. This emphasizes the fact that LDA+DMFT correctly captures atomic physics.

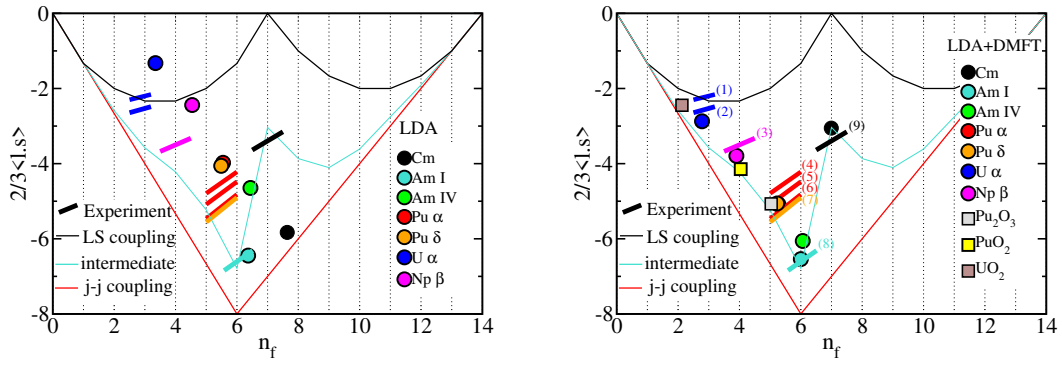


Fig. 20: *Value of spin-orbit operator deduced from the experimental branching ratio compared to exact limit (LS and jj coupling) and LDA (left) or LDA+DMFT (right) [37]*

4 f-electron systems: oxides

Most f-electron oxides are Mott insulators, and thus DFT+DMFT is the method of choice to describe them. DFT+U does describe also very well the structural and spectral properties, but not the magnetic properties. We will illustrate this with two examples of Mott insulators.

4.1 Electronic structure of Ce_2O_3

Ce_2O_3 is a Mott antiferromagnetic insulator with an optical band gap of 2.4 eV [61] and a Néel temperature of 9 K. We discuss in the following first the density of states and experiments, and secondly, the ground state density.

4.1.1 Density of states

Ce_2O_3 contains only one f-electron, resulting in the absence of multiplets in this system. Both DFT+U and DFT+DMFT methodologies successfully reproduce the energy gap of this compound. Fig. 21 shows the spectral function in DFT+DMFT. The DFT+U spectrum is very similar. These methods accurately capture the magnitude of the band gap which is between the two Hubbard bands. However, as we will discuss below, beneath these apparent similarities lie profound differences between the two approaches.

4.1.2 Electronic Density

To illustrate the differences between the LDA+DMFT and LDA+U electronic densities, we have plotted in Fig. 22 the isosurfaces of the differences between the LDA+DMFT and LDA electronic densities (on the left) and between LDA+U and LDA electronic densities (on the right). We see that the difference between LDA+DMFT and LDA is weaker than the difference between LDA+U and LDA. This suggests that the occupation matrices of the correlated orbitals are different in LDA+U and LDA+DMFT.

We now explain the origin of this difference. In DFT, local interaction is not taken into account, so the f-electron is delocalized over several f-orbitals. In DFT+U, because of the localization,

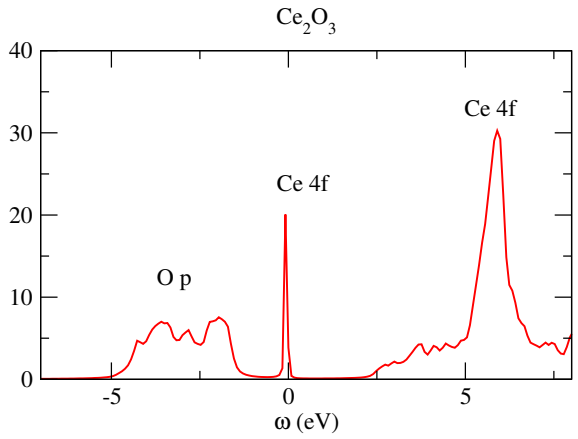


Fig. 21: *DFT+DMFT spectral function of Ce_2O_3 [38]*

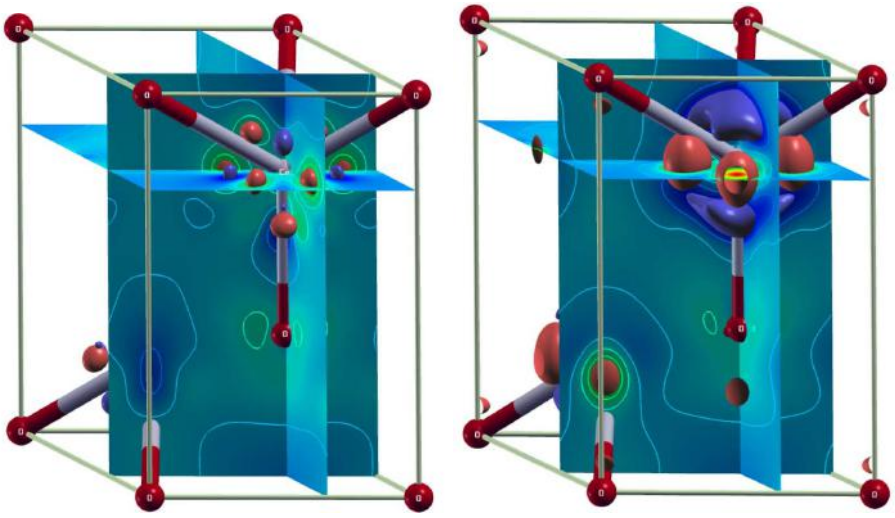


Fig. 22: *Difference between the electron densities computed using LDA+U (right) and LDA+DMFT (left) approximations and in the LDA approximation. The density of the filled orbital is clearly visible on the right. The blue (respectively green-red) area corresponds to a positive (respectively negative) value of the difference. [38]*

numerous low-energy one-electron nearly-degenerate states are present in these systems. Each of these states can be calculated using a single Slater determinant method like DFT+U [39] and corresponds to the localization of the f-electron in one f-orbital. As a consequence, the average occupation of one orbital is either close to 0 or 1 (to reduce the Coulomb interaction). This explains the large difference of the electron densities between DFT+U and DFT as seen in the right plot of Fig. 22.

However, to accurately describe a paramagnetic insulator without introducing unwanted orbital anisotropy and magnetism, a method capable of handling multiple determinants, such as DFT+DMFT, is required. In DFT+DMFT, the instantaneous occupation of orbitals is 0 or 1. But the average occupations of f-orbitals are lower and closer to the DFT/LDA values. The left plot of Fig. 22 highlights the similarity of the DFT/LDA and DFT+DMFT densities, even if it comes from completely different physical effects.

4.2 Magnetic ordering in UO_2

As f-orbitals are particularly localized, they are less sensitive to crystal-field effects. As a consequence, the orbital angular momentum is not quenched and plays an important role in magnetism. In particular, multipolar degrees of freedom can play a role: their interactions can lead to ordering of their arrangement.
This has been studied in UO_2 to resolve the long standing question of the 3**k** magnetic order in the antiferromagnetic phase, below 30 K. This peculiar ordering involves 4 different orientations of magnetic moments. Pourovkii *et al.* [40] proposed to compute nearest-neighbor interactions between the states belonging to the Γ_5 ground state using DFT+DMFT. DFT+DMFT describes well the band structure and paramagnetism in UO_2 [26]. Using the Hubbard I approximation, they were able to compute superexchange parameters for dipolar and quadrupolar interactions. Using these interactions, they were able to stabilize the experimental magnetic ordering and the phase transition to paramagnetism at a temperature close to experiment. Another recent work used a flavor of DFT+U to show the role of quadrupolar-quadrupolar interactions in the formation of 3**k** magnetic order.

5 Conclusion

f-electron systems exhibit significant orbital localization for lanthanides and comparatively lesser localization for actinides. As a consequence electrons are sensitive to electronic interaction and somewhat less to crystal fields in comparison to the d-elements. Spin-orbit coupling is also important for these heavy elements. These features require a good description of atomic physics and hybridization, which is provided by DFT+DMFT.

References

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

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

[3] A. Georges, AIP Conf. Proc. **715**, 3 (2004)

[4] S. Hecker, Metall. Mater. Trans. A **35**, 2207 (2004)

[5] A. McMahan, C. Huscroft, R. Scalettar, and E. Pollock, J. Comput. Aided Mater. Des. **5**, 131 (1998)

[6] A. Schiwek, F. Porsch, and W.B. Holzapfel, U. Krell, J.R. Schneider, M. von Zimmermann (eds.): Scientific Contributions. Part 1, HASYLAB Annual Report Hamburg, 2002

[7] F. Decremps, L. Belhadi, D.L. Farber, K.T. Moore, F. Occelli, M. Gauthier, A. Polian, D. Antonangeli, C.M. Aracne-Ruddle, and B. Amadon, Phys. Rev. Lett. **106**, 065701 (2011)

[8] B. Johansson, W. Luo, S. Li, and R. Ahuja, Sci. Rep. **4**, (2014)

[9] E. Weschke, C. Laubschat, T. Simmons, M. Domke, O. Strebel, and G. Kaindl, Phys. Rev. B **44**, 8304 (1991)

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

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

[12] P. Fulde, J. Phys. F: Metal Phys. **18**, 601 (1988)

[13] E. Pavarini and E. Koch (Eds.): *Simulating Correlations with Computers*, Modeling and Simulation, Vol. 10 (Forschungszentrum Jülich, 2011) `http://www.cond-mat.de/events/correl21`

[14] V.I. Anisimov, D.E. Kondakov, A.V. Kozhevnikov, I.A. Nekrasov, Z.V. Pchelkina, J.W. Allen, S.-K. Mo, H.-D. Kim, P.Metcalf, S. Suga, A. Sekiyama, G. Keller, I. Leonov, X. Ren, and D. Vollhardt, Phys. Rev. B **71**, 125119 (2005)

[15] F. Lechermann, A. Georges, A. Poteryaev, S. Biermann, M. Posternak, A. Yamasaki, and O.K. Andersen, Phys. Rev. B **74**, 125120 (2006)

[16] B. Amadon, F. Lechermann, A. Georges, F. Jollet, T.O. Wehling, and A.I. Lichtenstein, Phys. Rev. B **77**, 205112 (2008)

[17] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A.I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004)

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

[19] V.I. Anisimov and O. Gunnarsson, Phys. Rev. B **43**, 7570 (1991)

[20] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005)

[21] L. Vaugier, H. Jiang, and S. Biermann, Phys. Rev. B **86**, 165105 (2012)

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

[23] B. Amadon, T. Applencourt, and F. Bruneval, Phys. Rev. B **89**, 125110 (2014)

[24] P. Seth, P. Hansmann, A. van Roekeghem, L. Vaugier, and S. Biermann,
Phys. Rev. Lett. **119**, 056401 (2017)

[25] R. Sakuma and F. Aryasetiawan, Phys. Rev. B **87**, 165118 (2013)

[26] J. c. v. Kolorenč, A.B. Shick, and A.I. Lichtenstein, Phys. Rev. B **92**, 085125 (2015)

[27] E. Pavarini, E. Koch, F. Anders, and M. Jarrell (Eds.): *From Models to Materials*
Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
`http://www.cond-mat.de/events/correl12`

[28] B. Amadon, F. Jollet, and M. Torrent, Phys. Rev. B **77**, 155104 (2008)

[29] J. Bieder and B. Amadon, Phys. Rev. B **89**, 195132 (2014)

[30] I.L.M. Locht, Y.O. Kvashnin, D.C.M. Rodrigues, M. Pereiro, A. Bergman, L. Bergqvist, A.I. Lichtenstein, M.I. Katsnelson, A. Delin, A.B. Klautau, B. Johansson, I. Di Marco, and O. Eriksson, Phys. Rev. B **94**, 085137 (2016)

[31] K. Haule, V. Oudovenko, S.Y. Savrasov, and G. Kotliar,
Phys. Rev. Lett. **94**, 036401 (2005)

[32] B. Amadon and A. Gerossier, Phys. Rev. B **91**, 161103 (2015)

[33] T. Naka, T. Matsumoto, and N. Môri, Physica B: Condensed Matter **205**, 121 (1995)

[34] S.V. Streltsov, E. Gull, A.O. Shorikov, M. Troyer, V.I. Anisimov, and P. Werner,
Phys. Rev. B **85**, 195109 (2012)

[35] K. Haule and T. Birol, Phys. Rev. Lett. **115**, 256402 (2015)

[36] B. Amadon, Phys. Rev. B **97**, 039903 (2018)

[37] J H. Shim, K. Haule, and G. Kotliar, EPL (Europhys. Lett.) **85**, 17007 (2009)

[38] B. Amadon, Journal of Physics: Condensed Matter **24**, 075604 (2012)

[39] B. Dorado, B. Amadon, M. Freyss, and M. Bertolus, Phys. Rev. B **79**, 235125 (2009)

[40] L.V. Pourovskii and S. Khmelevskyi, Phys. Rev. B **99**, 094439 (2019)

17 Super-QMC: Strong Coupling Perturbation for Lattice Models

Alexander Lichtenstein
I. Institut für Theoretische Physik
Universität Hamburg, 20355 Hamburg

Contents

1 Idea of reference system **2**
1.1 Generic Hamiltonian 5

2 Numerically exact lattice QMC **6**
2.1 Hirsch-Fye DQMC 6
2.2 Continuous-time QMC 7

3 The DF-QMC method **8**
3.1 Real-space scheme 8
3.2 k-space scheme 12

4 Results for 8×8 lattices **15**

5 Discussion **19**

E. Pavarini and E. Koch (eds.)
Orbital Physics in Correlated Matter
Modeling and Simulation Vol. 13
Forschungszentrum Jülich, 2023, ISBN 978-3-95806-689-2
http://www.cond-mat.de/events/correl23

1 Idea of reference system

In this lecture we give an introduction the to theoretical description of interacting electron systems based on non-perturbative, strong-coupling expansions around optimal reference system. Density Functional Theory (DFT) and its Local Density Approximation (LDA) is based on the simplest reference system related with a homogeneous electron gas with constant external potential with the same Coulomb electron-electron interactions (see Fig. 1). Such a reference system can be solved via a numerically exact diffusion Monte Carlo scheme for theground state energy as a function of electron density [1]. On other hand, the Dynamical Mean-Field Theory (DMFT) [2] for strongly interacting fermionic systems is based on a strong coupling expansion around an effective impurity reference system (Fig. 1). This scheme become exact in the limit of infinite lattice dimension [3].

In the finite lattice dimension we can start from the DMFT reference system and use an analytical perturbation for non-local correlation effects. The frequency dependent effective impurity DMFT problem nowadays can be efficiently solved within continuous time quantum Monte Carlo (CT-QMC) schemes [4]. Therefore the perturbation theory needs to be formulated in the action path integral formalism. We discuss here a general way to include correlations beyond the reference system [5] which is based on the dual-fermion path-integral formalism [6].

For DMFT an effective impurity model, tailored to the problem of strong correlations, serves as the reference system, see Fig. 1. Since in the zeroth-order of this perturbative expansion, i.e., on the level of the DMFT problem, we already have an interacting problem and since the perturbation is momentum and frequency dependent, one is forced to replace the Hamiltonians by actions within the path-integral formalism. Note that the fermion path integral can also be used to formulate the DMFT itself [2]. The dual-fermion approach is not necessarily bound to a specific starting point.

We recently developed a strong-coupling perturbation scheme for generic Hubbard models around a half-filled particle-hole-symmetric reference system, which is free from the fermionic sign problem [7]. The approach is based on the lattice determinant quantum Monte Carlo (QMC) method in continuous and discrete time for large periodic clusters in a fermionic bath. Considering the first-order perturbation in the shift of the chemical potential and of the second-neighbor hopping gives an accurate electronic spectral function for a parameter range corresponding to the optimally doped cuprate system for temperatures of the order of T/t=0.1, the region hardly accessible for straightforward lattice QMC calculations. We discuss the formation of a pseudogap and the nodal-antinodal dichotomy for doped Hubbard systems in the strong-coupling regime with interaction parameter U equal to the bandwidth and the optimal value of the next-nearest-neighbor hopping parameter t' for high-temperature superconducting cuprates. Extensive investigation of the fermionic sign problem in the lattice DQMC for the t-t'-U Hubbard model for $U/t = 6$, $t'/t = -0.2$ and its relation with a quantum critical point [8] gives the generic "sign" phase diagram presented in Fig. 2. The red region in the temperature-doping (chemical potential) diagram presents an "unacceptable" sign problem, where one can not do any accurate simulations. It is interesting that the d-wave superconducting dome lies exactly in-

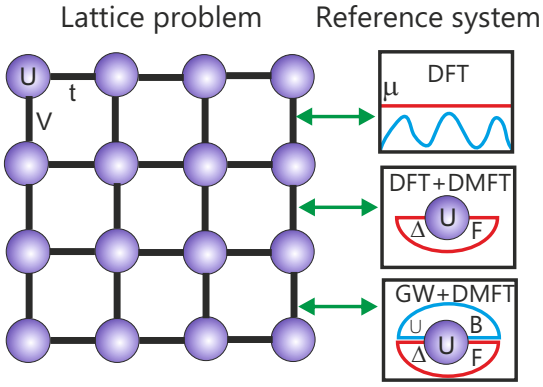


Fig. 1: *Schematic representation of reference systems in many-body approaches to lattice-fermion models: (i) Density-functional theory (DFT) with the interacting homogeneous electron gas as a reference system, defined by a constant external potential μ. (ii) Dynamical mean-field theory (DMFT) with an effective impurity problem as a reference system, defined by a fermionic bath, specified by the hybridization function Δ. (iii) GW+DMFT with a correlated atom in a fermionic (Δ) and a bosonic bath (Λ) due to effects of the frequency-dependent screening of long-range Coulomb (V) interactions.*

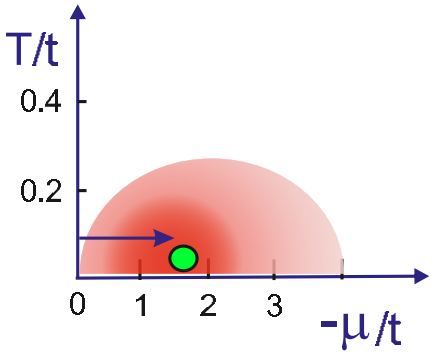


Fig. 2: *Schematic representation of the sign-problem in QMC lattice-fermion calculations. The green circle shows the approximate position of maximum d-wave superconductivity with 15% hole doping. The blue arrow give the "direction" of the reference system approach.*

side the "red-region". The arrow gives the directions which we would like to pursuit in order to overcome the sign problem. We will start from the sign-free half-field particle-hole symmetric case with μ=0 and move closer to the quantum critical point related with d-wave superconductivity in the cuprates. Comparing the density of states in Fig. 3 for the undoped case with a pseudogap, which is free from the sign-problem, and the hole doped case with μ/t=-2 with t'/t=-0.3 with a very sharp peak at the Fermi level, one can understand that a non-trivial reason for the sign-problem may be related with a strong anomaly in the many-body spectrum.

The search for numerically exact solutions of the t-t'-U Hubbard model in the thermodynamic limit at arbitrary interaction strength, long-range hopping and doping δ or, equivalently, chemical potential μ at low temperature T=$1/\beta$ is tremendously difficult. Modern computational approaches, based on lattice determinant quantum Monte Carlo (QMC) methods have made tremendous progress in the half-filled case without t' [9], but face an unacceptable fermionic sign problem for the general doped case related to the high-temperature superconductivity

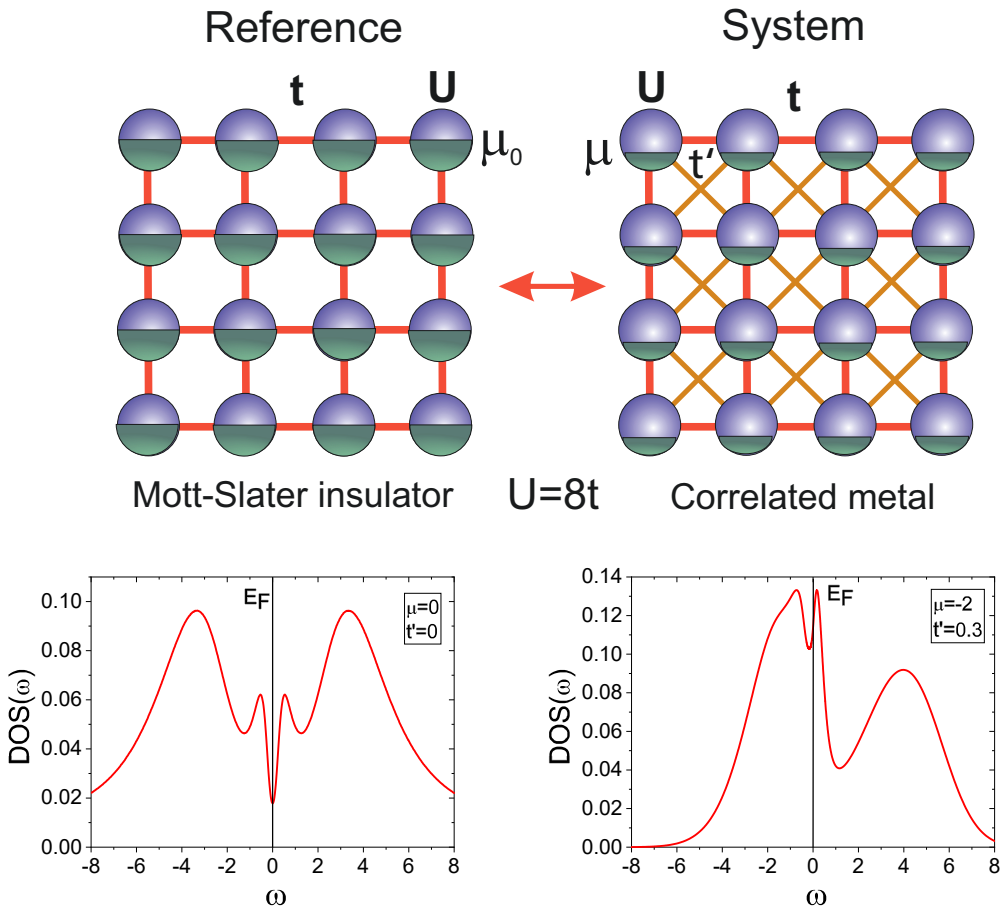


Fig. 3: *Schematic representation of a half-filled reference system for the doped square lattice. Bellow: calculated density of states (DOS) in presented scheme for $U/t = 8$, Left: undoped case $\mu = 0$ with $t' = 0$, Right: doped case $\mu = -2$ with $t'/t = -0.3$*

(HTSC) problem, which is the main factor restricting the accuracy of QMC calculations for interacting fermion systems [10–12, 8]. A very important and largely unresolved problem is related to the next-nearest-neighbor hopping t' in the Hubbard model and its role in superconductivity [13–19]. There are two recent successful attempts to resolve this long-standing problem using zero-temperature a variational QMC scheme for realistic HTSC systems [20] in combination with DMRG for the t-t'-U Hubbard model on a large ribbon geometry [21].

On the other hand, a new class of diagrammatic Monte Carlo scheme [22] is claimed to have a "sign blessing" property which helps to reduce the effects of high-order diagrams. The state-of-the-art diagrammatic Monte Carlo scheme in the connected determinant mode (C-DET) [23], based on efficient Continuous Time Quantum Monte Carlo(CT-QMC) scheme in the weak coupling technique (CT-INT) [24], gives unprecedented accuracy for the doped Hubbard model [25, 26]. It becomes possible to study the formation of the pseudogap already at the beginning of the strong coupling case with U/t=6 [25]. Nevertheless, the exponential convergence of the C-DET scheme for weak interactions [27, 28] turns into a divergence at large U values due to poles in the complex U-plane [26]. This means that calculations for interactions close to the bandwidth $U/t \approx 8$ and temperature $T/t \approx 0.1$ are still within a prohibited area in the phase diagram [26].

There is a recent interesting attempt to use a dynamical variational QMC scheme for the doped Hubbard model [29,30], which gives a very reasonable description of the spectral function. The existence of the pseudogap can be explained in a simple model of electron fractionalization and the appearance of "dark" fermions which is supported by 2×2 cluster Dynamical Mean Field Theory (C-DMFT) [31, 17]. Moreover, the experimental RIXS spectra [32] of doped cuprate materials can be interpreted in such a theoretical model of the pseudogap formation. Larger clusters in the C-DMFT scheme for the doped case have, however, an unacceptable fermionic sign problem within the QMC scheme.

Here we discuss a different route to tackle the "sign problem" in the determinant lattice QMC scheme and design a strong-coupling perturbative solution for a general Hubbard model. The starting point is related to the "reference system" idea [33] which is basically quite simple and straightforward. The conventional choice of the noninteracting Hamiltonian as the reference system for the perturbation [34] is motivated by Wick's theorem which allows to calculate exactly any many-particle Green functions: they are all expressed in terms of single-particle Green functions. The choice of a single-site approximation like dynamical mean-field theory [35] as the reference system leads to the dual fermion technique [6, 33]. Actually, the reference system can be arbitrary, assuming that we can calculate its Green functions of arbitrary order. It is worthwhile to mention here the very successful Peierls-Feynman-Bogoliubov variational principle [36–38]. In this case, a good variational estimate of the system's free energy F with the Hamiltonian H_1 is achieved on an optimal reference system with the Hamiltonian H_0, namely $F_1 \leq F_0 + \langle H_1 - H_0 \rangle_0$. One can hope therefore that even first-order corrections to the properly chosen reference system will already give a rich and adequate enough physical picture.

1.1 Generic Hamiltonian

The simplest model describing interacting fermions on a lattice is the single band Hubbard model, defined by the Hamiltonian

$$\hat{H}_\alpha = -\sum_{i,j,\sigma} t_{ij}^\alpha \, c_{i\sigma}^\dagger c_{j\sigma} + \sum_i U\Big(n_{i\uparrow} - \frac{1}{2}\Big)\Big(n_{i\downarrow} - \frac{1}{2}\Big) \tag{1}$$

where t_{ij} are hopping matrix elements including the chemical potential μ in the diagonal part

$$t_{ij}^\alpha = \begin{cases} t & \text{if } i \text{ and } j \text{ are nearest neighbors,} \\ \alpha t' & \text{if } i \text{ and } j \text{ are next-nearest neighbors,} \\ \alpha\mu & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$. We introduce a "scaling" parameter $\alpha{=}0, 1$, which distinguishes a reference system H_0 for $\alpha{=}0$ and corresponds to the half-filled Hubbard model ($\mu_0{=}0$) with only nearest neighbors hopping ($t_0'{=}0$) from the final system H_1 for $\alpha{=}1$ with given μ and t'. Note that long-range hoping parameters can be trivially included in the present formalism similar to t'.

The reference system now corresponds to the half-filled (μ=0) particle-hole symmetric (t'=0) case (Fig. 3), where lattice Monte Carlo has no sign problem and the numerically exact solution for any practical value of U is possible within a broad range of temperatures [39]. Then we apply the lattice dual fermion perturbation theory [6,5,33] to find the first-order perturbative corrections in μ and t'. To this end, it is sufficient to calculate the two-particle Green function, or, equivalently, the four-leg vertex, which can be done with sufficient accuracy within continuous time quantum Monte Carlo. Our reference system already has the main correlation effects in the lattice and shows the characteristic "four-peak" structure [40] with high-energy Hubbard bands around $\pm U/2$ and antiferromagnetic Slater bands close to the insulating gap (which can be seen in the density of states in figure 3, left panel). After the dual fermion perturbation scheme correlated metallic states with the DMFT-like "three peak" structure appear with a pseudogap-like feature at high temperature (the density of states in figure 3, right panel). Results for the strong-coupling case (U=W=$8t$) with practically interesting values of the chemical potential and next-nearest-neighbor hoppings corresponding to cuprate superconductors have shown the formation of a pseudogap and a nodal-antinodal dichotomy (that is, well-defined quasiparticles in the nodal part of the Fermi surface and strong quasiparticle damping for the antinodal part), which gives this approximation a perspective for practical applications.

2 Numerically exact lattice QMC

We briefly introduce here the main ideas of two different lattice QMC approaches for large periodic clusters in a bath. The first is based on the Hubbard-Stratonovich transformation of the local interaction in Eq. (2), the other is related with the continuous-time interaction (CT-INT) expansion scheme. Both QMC methods are used here for practical computations.

2.1 Hirsch-Fye DQMC

We use a path-integral formalism with Grassmann variables $[c_i^*, c_i]$. The space-time bare Green function $\mathcal{G}_{ij}$ describes the non-interacting part of the Hamiltonian in Eq. (2) for $N_x \times N_y$ 2D-space ($N = N_x \cdot N_y$) and $L \times L$ discretized times in an effective bath representing the external infinite lattice with space-time index here $i \equiv (\mathbf{r}, \tau)$. Imaginary time slicing corresponds to the mesh $\tau = l * \Delta\tau$ with $l = 0, \ldots, L-1$ and $\Delta\tau = \beta/L$ with inverse temperature β. The interaction part of the Hamiltonian Eq. (2) is decoupled by mapping to auxiliary Ising fields s_i via a discrete Hirsch-Hubbard-Stratonovich transformation [41]

$$\big(-U\Delta\tau(n_{i\uparrow}n_{i\downarrow} - (n_{i\uparrow}+n_{i\downarrow})/2)\big) = \frac{1}{2}\sum_{s_i=\pm 1} \qquad \big(\lambda s_i(n_{i\uparrow}-n_{i\downarrow})\big), \tag{3}$$

where $\lambda =$ s $(e^{U\Delta\tau/2})$ and for the best convergence of DQMC one uses the following "rule of thumb" $U\Delta\tau/2 \lesssim 1$. Then the effective lattice action become Gaussian

$$S[c^*, c] = -\sum_{i,j,\sigma} c^*_{i\sigma}\, G^{-1}_{ij\sigma}\, c_{j\sigma} \quad \text{with} \quad G^{-1}_{ij\sigma}(s) = \mathcal{G}^{-1}_{ij\sigma} - \delta_{i,j}\lambda s_i \sigma\,, \tag{4}$$

where $s \equiv \{s_i\}$ with $i = 1, \ldots, N{\cdot}L$. Note that in time space the delta function should be anti-periodic for fermions [42, 43, 2] and Eq. (4) has a schematic form. For such a Gaussian action we can integrate-out fermionic the degrees of freedom and get for the partition function the following formula used in the determinant QMC scheme

$$Z = \frac{1}{2^{NL}} \sum_{s} \prod_{\sigma} \quad [G_{\sigma}^{-1}(s)] \,, \tag{5}$$

where the sum over Ising auxiliary fields s_i performed with an importance-sampling Monte Carlo algorithm with probability $P(s) = \quad [G_{\uparrow}^{-1}(s)] \cdot \quad [G_{\downarrow}^{-1}(s)]$ which is always positive for the half-field particle-hole symmetric Hubbard model [42]. Within the DQMC scheme the exact single-particle Green function of the reference system can be calculated as

$$g_{ij}^{\sigma} = \frac{1}{Z} \sum_{s} P(s)\, G_{ij}^{\sigma}(s) \,. \tag{6}$$

In practice of DQMC one uses a so-called fast-update formalism to calculate the lattice Green function Eq. (4) with a single Ising spin-flip [42].

2.2 Continuous-time QMC

The interaction expansion (CT-INT) continuous-time quantum Monte Carlo algorithm for fermions is based on a formal series expansion for the partition function in the interaction term of the action [24]. In a schematic form with short notation $i_k \equiv (\mathbf{r}_k, \tau_k)$ we have

$$Z = \int \mathcal{D}[c^*, c]\, e^{-S\ [c^*, c]} \sum_{k=0}^{\infty} \frac{(-U)^k}{k!} \int_0^{\beta} d\tau_{1 \ldots k}\, c_{i_1\uparrow}^{*} c_{i_1\uparrow}\, c_{i_1\downarrow}^{*} c_{i_1\downarrow} \ldots c_{i_k\uparrow}^{*} c_{i_k\uparrow}\, c_{i_k\downarrow}^{*} c_{i_k\downarrow}, \tag{7}$$

where S_0 is the Gaussian part of the action related with $\mathcal{G}_{ij}^{\sigma}$. In this case we can integrate-out the fermionic path integral in Eq. (7) to get the determinant of the $k{\times}k$ bare Green function $\mathcal{G}^{\sigma}$

$$Z = Z_0 \sum_{k=0}^{\infty} (-U)^k \int_0^{\beta} d\tau_1 \ldots \int_{\tau_{k-1}}^{\beta} d\tau_k \prod_{\sigma} \quad \mathcal{G}_k^{\sigma} \,. \tag{8}$$

In order to overcome a trivial sign problem related with factor $(-U)^k$ one uses a particle-hole transformation related with a so-called α-shift [24]. The CT-INT scheme performes important sampling in the space of $k{\times}k$ fermionic determinants. The probability to change k to $k{+}1$-order in the Metropolis algorithm is related with ratio of the fermionic determinants [24]

$$P(k \to k{+}1) = \min \left(1, \frac{\beta U}{k{+}1} \prod_{\sigma} \frac{(\quad \mathcal{G}_{k+1}^{\sigma}}{\mathcal{G}_k^{\sigma}} \right) . \tag{9}$$

The optimal order of k-perturbation, which corresponds to the maximum of the distribution function of the fermionic determinants [24] for a cluster of N-sites is of the order $k_{opt} \sim \beta N U$. Finally, the exact reference Green function in CT-INT formalism is calculated as

$$g_{ij}^{\sigma} = \mathcal{G}_{ij}^{\sigma} - \sum_{k,k'} \mathcal{G}_{ik}^{\sigma} \cdot M_{k,k'}^{\sigma} \cdot \mathcal{G}_{k'j}^{\sigma} \,, \tag{10}$$

where the M-matrix is equal to the Monte Carlo average the of inverse fermionic matrix in Eq. (8).

3 The DF-QMC method

We start with the strong-coupling theory of the dual fermion scheme [6, 44] for the t-t'-U Hubbard model on a square lattice. There are many important works on pure strong-coupling expansions in the hopping t for Hubbard model [45–49]. The dual-fermion scheme [6] differs from the pure strong-coupling expansion in the hopping t in a very important way: it is an expansion from a reference system to the final system, or in the "difference" $\tilde{t}$ (Fig. 3) which converges much better. The general strategy of the dual fermion approach as a strong coupling theory is related to the formally exact expansion around an arbitrary reference system [33].

3.1 Real-space scheme

Let us consider a general lattice fermion model with local Hubbard-like interaction vertex U. Using the path-integral formalism the partition function of a general fermionic lattice system (Fig. 3) can be written as the functional integral over Grassmann variables $[c^*, c]$

$$Z_\alpha = \int \mathcal{D}[c^*, c]\, e^{-S_\alpha[c^*,c]}\,. \tag{11}$$

For the super-perturbation in the lattice Monte Carlo scheme we use a general dual-fermion expansion around an arbitrary reference system within the path-integral formalism [6, 33], similar to a strong coupling expansion [47, 48]. In this case our $N\times N$ lattice and corresponding reference systems represents an $N\times N$-piece cut from the infinite lattice and periodize the bare Green function $\mathcal{G}_\alpha$. The general lattice action for a discretized $N_x\times N_y\times L$ space-time lattice (for the CT-INT scheme imaginary time τ is continuous in the interval $[0, \beta)$) with general interaction term reads

$$S_\alpha[c^*, c] = -\sum_{1,2} c_1^*\, (\mathcal{G}_\alpha)^{-1}_{12}\, c_2 + \frac{1}{4}\sum_{1234} U_{1234}\, c_1^* c_2^*\, c_4 c_3\,. \tag{12}$$

In order to keep the notation simple, it is useful to introduce the combined index $|1\rangle \equiv |i, \tau, \sigma\rangle$ (i being the site index).

To calculate the bare propagators $(\mathcal{G}_\alpha)_{12}$ we start from the $N_x\times N_y$ cluster which is cut from the infinite lattice and then force translation symmetry and periodic boundary conditions on the finite $N_x\times N_y$ system. This procedure is easy to realize in the k-space, by doing first a double Fourier transform of the bare Green function for a non-periodic $N\times N$ cluster $\mathcal{G}^\alpha_{\mathbf{k},\mathbf{k}'}$, keeping only the periodic part, $\mathcal{G}^\alpha_{\mathbf{k}}\,\delta_{\mathbf{k},\mathbf{k}'}$.

The perturbation matrix related with the difference of the one-electron part of the action is

$$\tilde{t} = \mathcal{G}_0^{-1} - \mathcal{G}_1^{-1}. \tag{13}$$

In order to formulate an expansion around the reference action S_0, we express a connection to the final action $S \equiv S_1$ with the same local interaction in the following form

$$S[c^*, c] = S_0[c^*, c] + \sum_{12} c_1^*\, \tilde{t}_{12} c_2\,. \tag{14}$$

The main idea of the dual fermion transformation is the change of variables from strongly correlated fermions (c^*, c) to weakly correlated "dual" Grassmann fields (d^*, d) in the path integral representation for the partition function from Eq. (14), followed by a simple perturbation treatment. The new variables are introduced through a Hubbard-Stratonovich transformation with the matrix $\tilde{t}_{12}$ in real-space (assuming Einstein summation convention over repeated indices)

$$e^{-c_1^* \, \tilde{t}_{12} \, c_2} = Z_t \int \mathcal{D}[d^*, d] \, e^{d_1^* \, \tilde{t}_{12}^{-1} \, d_2 + d_1^* c_1 + c_1^* d_1} \tag{15}$$

with $Z_t = \quad \left[-\tilde{t}\right]$ and we always imply matrix inversion: $\tilde{t}_{12}^{-1} \equiv (\tilde{t}^{-1})_{12}$. Using this transformation, the lattice partition function becomes

$$Z = Z_0 Z_t \int \mathcal{D}[c^*, c, d^*, d] \, e^{d_1^* \, \tilde{t}_{12}^{-1} \, d_2} \, \left\langle e^{d_1^* c_1 + c_1^* d_1} \right\rangle_0 \tag{16}$$

with the standard definition of average over S_0

$$\langle \cdots \rangle_0 = \frac{1}{Z_0} \int \mathcal{D}[c^*, c] \, \cdots \, e^{-S \, [c^*, c]}. \tag{17}$$

Now we can integrate-out the c^*, c fermions and show that the average over S_0 can be rewritten in the cumulant expansion [48] of connected correlators $\langle \cdots \rangle_{0c}$

$$\left\langle e^{d_1^* c_1 + c_1^* d_1} \right\rangle_0 = \quad \left[\sum_{n=1}^{\infty} \frac{(-1)^n}{(n!)^2} \gamma^{(2n)}_{1 \cdots n, n' \cdots 1'} d_1^* \cdots d_n^* d_{n'} \cdots d_{1'} \right] \tag{18}$$

with cumulant of the reference system that can be calculated within QMC

$$\gamma^{(2n)}_{1 \cdots n, n' \cdots 1'} = (-1)^n \, \langle c_1 \cdots c_n c^*_{n'} \cdots c^*_{1'} \rangle_{0c} \, . \tag{19}$$

We can write the effective action for "dual fermions" $\tilde{S}[d^*, d]$ in the lowest-order approximation for the dual interaction [7]. The first term in the cumulant expansion, Eq. (18), with $n = 1$ ($\gamma^{(2)}_{11'}$), which is bilinear over the $[d_1^*, d_2]$ Grassmann variable, corresponds to the exact Green function of the reference system

$$g_{12} = -\langle c_1 c_2^* \rangle_0 = \frac{-1}{Z_0} \int \mathcal{D}[c^*, c] \, c_1 c_2^* \, e^{-S \, [c^*, c]} \, . \tag{20}$$

Note, that all correlators of the reference system will be written in lowercase. Together with the term $\tilde{t}_{12}^{-1}$ in Eq. (16) it gives the bare Green function for the dual fermions

$$\tilde{G}^0_{12} = \left[\, \tilde{t}^{-1} - \hat{g} \, \right]^{-1}_{12} \, . \tag{21}$$

The second term in the cumulant expansion, Eq. (18), with $n = 2$ ($\gamma^{(4)}_{122'1'}$), which is biquadratic over the $[d_1^*, d_2]$ Grassmann variable, gives the effective two-particle interaction among the dual-fermions. The corresponding connected four-point vertex has the form

$$\gamma_{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_0 - \langle c_1 c_4^* \rangle_0 \langle c_2 c_3^* \rangle_0 + \langle c_1 c_3^* \rangle_0 \langle c_2 c_4^* \rangle_0 \tag{22}$$

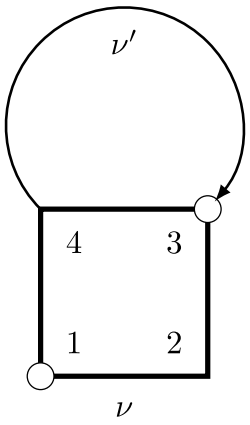


Fig. 4: *Feynman diagram for the first-order dual fermion perturbation for the self-energy $\tilde{\Sigma}_{12}(\nu)$: a line represents the non-local dual Green function $\tilde{G}_{43}(\nu')$ and a box is the two-particle vertex (cumulant) γ_{1234}, (σ, σ') are spin-indices.*

with four-point correlator, or two-particle Green function, for the reference system

$$\langle c_1 c_2 c_3^* c_4^* \rangle_0 = \frac{1}{Z_0} \int \mathcal{D}[c^*, c] \; c_1 c_2 c_3^* c_4^* \; e^{-S \; [c^*, c]} \, . \tag{23}$$

Finally, the dual-fermion action in the two-particle approximation has the form

$$\tilde{S}[d^*, d] = - \sum_{12 \, \nu\sigma} d^*_{1\nu\sigma} \; (\tilde{G}^0_\nu)^{-1}_{12} \; d_{2\nu\sigma} + \frac{1}{4} \sum_{1234} \gamma_{1234} d^*_1 d^*_2 d_4 d_3 \, . \tag{24}$$

Note, that we change sign for the interaction terms using anti-commutation rules for Grassmann variables in order to be consistent with the standard form for Coulomb interactions (Eq. (12)). The first-order correction to the dual self-energy is given by the diagram shown in Fig. 4 and can be calculated for a large system within the QMC-scheme as

$$\tilde{\Sigma}^{(1)}_{12} = \sum_{s-QMC} \sum_{3,4} \gamma^d_{1324}(s) \, \tilde{G}^0_{43} \tag{25}$$

where the density vertex reads

$$\gamma^d_{1234} = \gamma^{\uparrow\uparrow\uparrow\uparrow}_{1234} + \gamma^{\uparrow\uparrow\downarrow\downarrow}_{1234}. \tag{26}$$

The main trick for practical large system computations related to the fact that within the determinant DQMC scheme using the Ising-fields $\{s\}$ or within the CT-INT with stochastic sampling of interaction order expansion $\{s\}$, for two-particle correlators we can use Wick's theorem

$$\gamma_{1234}(s) \equiv \langle c_1 c_2 c_3^* c_4^* \rangle_s = \langle c_1 c_4^* \rangle_s \, \langle c_2 c_3^* \rangle_s - \langle c_1 c_3^* \rangle_s \, \langle c_2 c_4^* \rangle_s. \tag{27}$$

In order to find exact relationship between real and dual Green function, we perform a variation of $\mathrm{n}\, Z$ in Eq. (17) and Eqs. (11,14) with respect to $\tilde{t}$ [50]

$$G_{12} = \frac{\delta \; \mathrm{n}\, Z}{\delta \, \tilde{t}_{21}} = -\tilde{t}^{-1}_{12} + \tilde{t}^{-1}_{13} \tilde{G}_{34} \tilde{t}^{-1}_{42}. \tag{28}$$

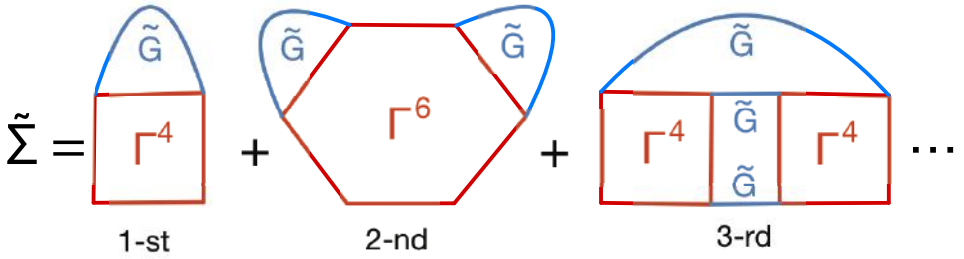


Fig. 5: *Diagrammatic series for the dual self-energy up to the 3-rd order in the $\tilde{G}$.*

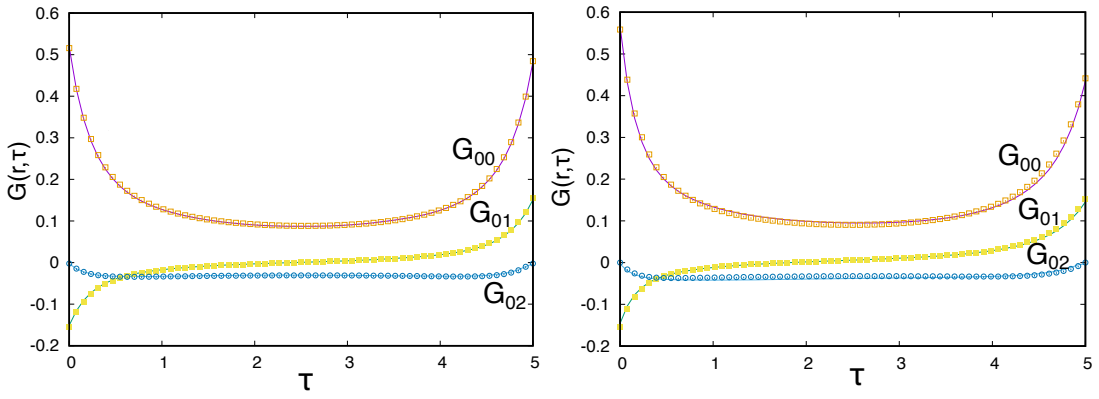


Fig. 6: *Three non-equivalent components of the Green functions for a 2×2 system as function of imaginary time for $U = 5.56$, $\beta = 5$ and t'/t=−0.1, μ=0 (left), and t'/t=−0.3, μ=−1.3 (right). Note, that here we use the QMC definition with positive local Green function.*

Using the definition of exact dual Green function $\tilde{G}^{-1} = \tilde{G}_0^{-1} - \tilde{\Sigma}$, we can get the expression for the real Green function

$$G_{12} = \left[\left(g + \tilde{\Sigma}\right)^{-1} - \tilde{t}\right]^{-1}_{12}. \tag{29}$$

The dual fermion transformation allows to use arbitrary reference systems and transforms the strongly correlated lattice fermion problem to an effective action of weakly-coupled dual quasi-particles. In this case even the lowest-order approximation can give reasonable results. The exact diagrammatic series for the dual self-energy presented is in Fig. 5. The second-order diagram in $\tilde{G}$ which includes $\gamma^{(6)}$ is local within the cluster and can be calculated with a similar QMC scheme.

For small systems of 2×2 clusters in the bath we can calculate the matrix of Green function of Eq. (29) directly in the real space formalism. In this case we do not need any additional periodization since the 2×2 cluster is "self-periodic". Since there is almost no sign problem in the DQMC method for the doped 2×2 cluster in the bath, we can compare the first-order dual-fermion perturbation with numerically exact DQMC results. The all three non-equivalent Green functions for 2×2 system are shown in Fig. 6, using the first-order DF-correction within the Hirsch-Fye QMC formalism. For a small perturbation, $\Delta\mu = -0.3$ and $\Delta t' = 0$, a comparison with exact DQMC results (points on Fig. 6) is perfect. For a large perturbation, $\Delta\mu = -1.5$ and $\Delta t' = 0.15$, one can already see small differences from the exact DQMC Green function. Nevertheless, the results of DF-QMC with only first-order corrections for the dual self-energy are very satisfactory.

3.2 k-space scheme

For large system ($N \geq 4$) it is much faster to calculate the dual self-energy in k-space with within the QMC Markov chain. The dual action in k-space reads

$$\tilde{S}[d^*, d] = -\sum_{\mathbf{k}\nu\sigma} d^*_{\mathbf{k}\nu\sigma}\, \tilde{G}^{-1}_{0\mathbf{k}\nu}\, d_{\mathbf{k}\nu\sigma} + \frac{1}{4}\sum_{1234} \gamma_{1234} d^*_1 d^*_2 d_4 d_3\,. \tag{30}$$

Using the short notation $k \equiv (\mathbf{k}, \nu_n)$ with the fermionic Matsubara frequencies $\nu_n = (2n+1)\pi/\beta$, $n \in \mathbf{Z}$, the dual Green function is equal to

$$\tilde{G}^0_k = \left(\tilde{t}^{-1}_k - \hat{g}_k\right)^{-1}. \tag{31}$$

Since the bare dual Green function is calculated in the independent QMC run for the reference system, it is fully translationally invariant $\tilde{G}^0_{34} \equiv \tilde{G}^0(3{-}4)$ and we use the Fourier transform to calculate the k-space dual Green function $\tilde{G}^0_k$. Within the QMC Markov chain the lattice auxiliary Green function is not translationally invariant, therefore for $g^{s\sigma}_{12} = -\langle c_{1\sigma} c^*_{2\sigma}\rangle_s$ we use the double Fourier transform to calculate $g^{s\sigma}_{kk'}$. Note that here we have explicitly written the fermionic spin σ and the auxiliary Ising spins s_i. To include the "disconnected part" of the vertex in equation Eq. (22) we just subtract the exact Green function from the previous QMC run of the paramagnetic reference system

$$\tilde{g}^{s\sigma}_{12} = g^{s\sigma}_{12} - g_{12}. \tag{32}$$

In the k-space this subtractions has the form

$$\tilde{g}^{s\sigma}_{kk'} = g^{s\sigma}_{kk'} - g_k\, \delta_{kk'}. \tag{33}$$

For the transformation of the vertex γ^d_{1234} in Eq. (26) within the QMC step in k-space we take into account that indices $(3, 4)$ are "diagonal" in k-space due to the multiplication by the translationally invariant dual Green function $\tilde{G}^0_{34}$, which transforms as $\tilde{G}^0_k \delta_{kk'}$, and indices $(1, 2)$ become translationally invariant after the QMC-summation, which finally leads us to the equation for the spin-up components of the first-order dual self-energy $\tilde{\Sigma}_k$ in the paramagnetic phase

$$\tilde{\Sigma}^{(1)}_k = \frac{-1}{(\beta N)^2 Z_{QMC}} \sum_{s-QMC} \sum_{k'} \left(\tilde{g}^{s\uparrow}_{kk} \tilde{g}^{s\uparrow}_{k'k'} - \tilde{g}^{s\uparrow}_{kk'} \tilde{g}^{s\uparrow}_{k'k} + \tilde{g}^{s\uparrow}_{kk} \tilde{g}^{s\downarrow}_{k'k'}\right) \tilde{G}^0_{k'}. \tag{34}$$

The additional normalization factor $\frac{1}{(\beta N)^2}$ comes from the Fourier transform in $\mathbf{k}$ and from the $\mathbf{k}'$-sum with N lattice sites and summation over Matsubara frequency: $\frac{1}{\beta}\sum_{\nu'}(...)$. For paramagnetic calculations we average over the two spin projections. The corresponding lattice Green function reads

$$G_k = \left[\left(g_k{+}\tilde{\Sigma}_k\right)^{-1} - \tilde{t}_k\right]^{-1}. \tag{35}$$

We note that if we neglect the dual self-energy, $\tilde{\Sigma}_k = 0$, this approximation is equivalent to cluster-perturbation theory (CPT) [51] and was recently implemented in the DQMC scheme [52].

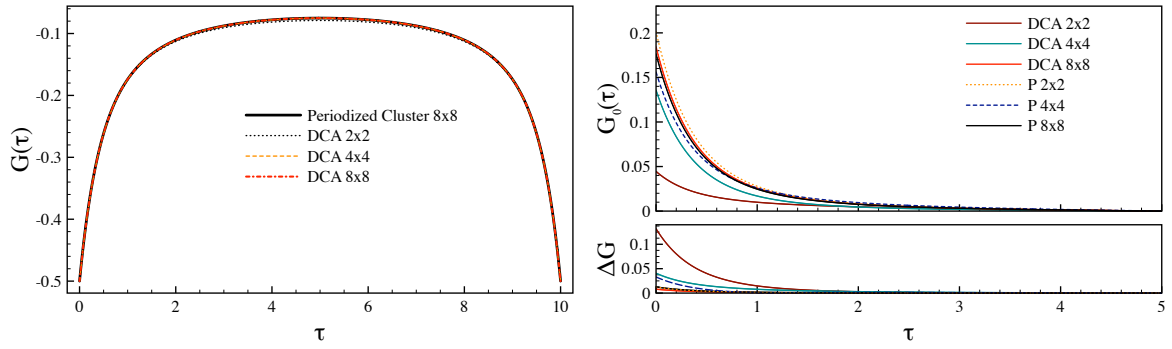


Fig. 7: *Periodized Green function in imaginary time τ for clusters of size $N\times N$ with $N = 2, 4, 8$ for our scheme compared to the DCA approach for the reference system with U=5.56 and β=10 for t'=0 and μ=0. The local Green function (left) and first nearest-neighbor (right).*

Tests for different system sizes show reasonable convergence of the first-order dual-fermion approximation for small perturbations.

For practical calculations of a bare Green function for $N_x\times N_y$ system we use a special scheme to reduce the dependence on the cluster size. We start from the non-interacting Green function with given t'/t and μ for a infinite lattice (in practice $50N_x\times 50N_y$ with periodic boundary conditions). We then cut the Green function to only our small $N\times N$ system, which results in a non-periodic Green function $\mathcal{G}_{ij}(\nu_n)$ with $(i, j = 0, N-1)$. In order to periodize the Green function for the small system, we average the corresponding distance, for example $\mathcal{G}_{0,n}$ and $\mathcal{G}_{0,N/2-n}$. In practice, we use the "double" Fourier transform on i and j from $\mathcal{G}_{ij}$ to $\mathcal{G}_{kk'}$ and take the diagonal (periodic) part $\mathcal{G}_k\delta_{kk'}$. In this way the local Green function does not depend on the size of our cluster and the non-local part (Fig. 7) has much faster converge in comparison to the standard periodic DCA cluster scheme [53]. The reason for this fast convergence of the non-local Green function and exact local Green function is related with the real space periodization, while DCA makes the average patches in the k-space. For the 8×8 system both periodization schemes converge for the nearest-neighbors Green function in comparison with 16×16 'test' cases (Fig.7).

We analyze the performance of the DF-QMC formalism as a function of $\Delta\mu$ and $\Delta t'$ for the 4×4 periodic cluster with and without fermionic bath. The DQMC sign-problem for 4×4 systems is also mild and we can compare our DF-QMC with numerically exact tests for the same μ and t'. We use a value of U=5.56 which corresponds to a degenerate ground state of the plaquette [19]. For all Hirsch-Fye DQMC calculations we use imaginary time discretization with L=64 slices. Fig. 8(a) shows DF-QMC results for a small perturbation $t'/t = -0.1$ and $\beta = 5$ in comparison with exact DQMC. The agreement is very good which shows the strength of dual fermion QMC theory for a small perturbation. Next, we compare for the $t'/t = -0.3$ case which corresponds to optimal next-nearest hopping in cuprate materials (Fig. 8(b)). In this case one can see the difference from exact DQMC results on the first Matsubara frequency, but still the overall agreement in all 6 non-equivalent k-points of the 4×4 system is quite satisfactory. The effects of a chemical potential shift μ=-0.5 is presented in Fig. 8(c). Qualitatively, agreement between perturbative DF-QMC and exact DQMC is similar to the case of t'/t=-0.3, but the

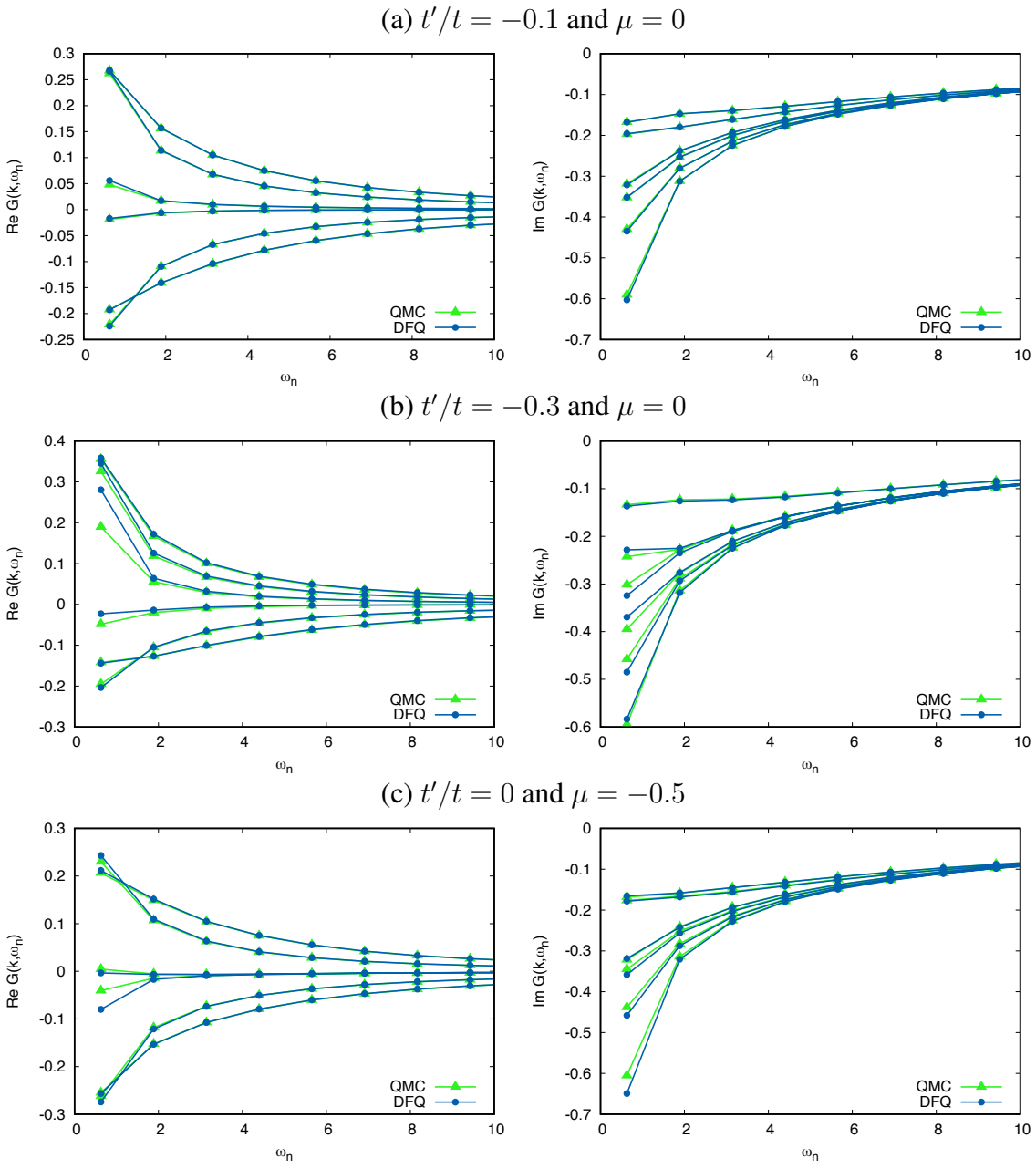


Fig. 8: *Green functions from DF-QMC (DFQ) in comparison with numerically exact DQMC results (QMC) for the* 4×4 *system in Matsubara space with* U=5.56, β=5 *and* t'/t, μ *as incated above the plots. Real part (left) imaginary part (right).*

structure of all 6 non-equivalent Green functions of the 4×4 system is very different. Still, the dual-fermion strong coupling perturbation in k-space works reasonably well.

Fig. 9 shows the combined effect of a strong chemical-potential shift μ=-1 and next-nearest hopping t'/t=-0.3. In this case we show results of the CT-INT approach with β=10. The agreement is very good, and only one Green function (the lowest imaginary part) which corresponds to the X-point ($\mathbf{k} = (\pi, \pi)$) and is located close to the Fermi level (the corresponding real part is close to zero) the dual perturbation shows a small discrepancy. In principle, one can reduce the error of the dual perturbation if one can choose the reference closer to the target system. The main condition is a weak sign problem for the reference system. Since this is

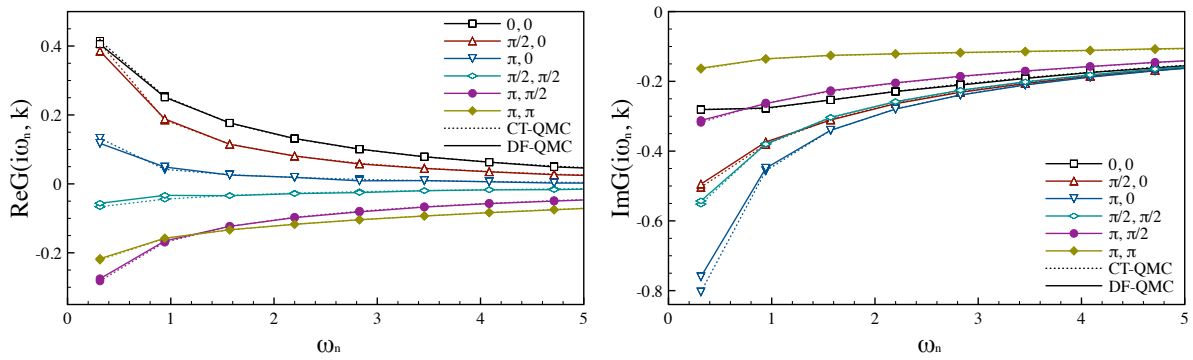


Fig. 9: *Green functions from DF-CT-QMC (DF-QMC) in comparison with numerically exact QMC results (CT-QMC) for the* 4×4 *system in Matsubara space with* U=5.56, t'/t=−0.3, μ=−1 *and* β=10. *Real part (left) and imaginary part (right).*

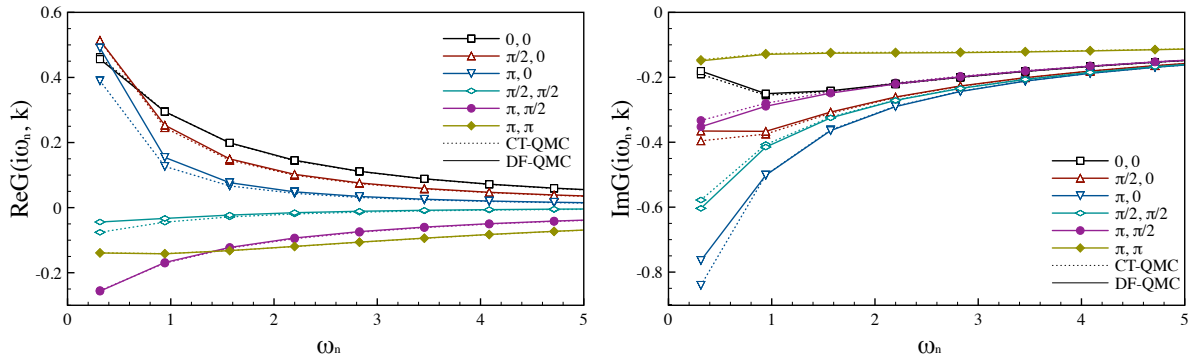


Fig. 10: *Green functions from DF-CT-QMC (DF-QMC) perturbation for* t'/t=−0.3, *starting from a reference system with* t'/t=−0.1 *in comparison with numerically exact QMC results (CT-QMC) for the* 4×4 *system in Matsubara space with* U=5.56, μ=0 *and* β=10. *Real part (left) and imaginary part (right).*

always the case for small 4×4 clusters, we can show in the Fig. 10 the CT-INT results starting from a reference system corresponding to t'/t=−0.1 for the target system with t'/t=−0.3. In this case, for lower temperature, T=$t/10$, the DF-QMC results are still in a good agreement with the exact solution. This example shows that we can also use an "over-doped" Hubbard model ($-\mu/t \backsimeq 4$ in Fig. 3) as a reference system in order to tackle the optimally doped case with super-QMC scheme.

4 Results for 8×8 lattices

For large 8×8 systems at $\beta = 10$, the average sign in DQMC is of the order of 10^{-3} even with a fermionic bath, and calculations of the test Green function are no longer possible. For much larger temperatures, corresponding to β=3 and not so large U=5.56 the sign problem is not severe and it is still possible to prepare a DQMC test. Fig. 11 shows a comparison of the Matsubara Green functions for all 15 non-equivalent k-points in dual-fermion perturbation with Hirsch-Fye QMC and DQMC-test. The agreement is quite good, but one should remember the very high temperature of this test (T=$t/3$) which results in metallic behavior of all Green functions.

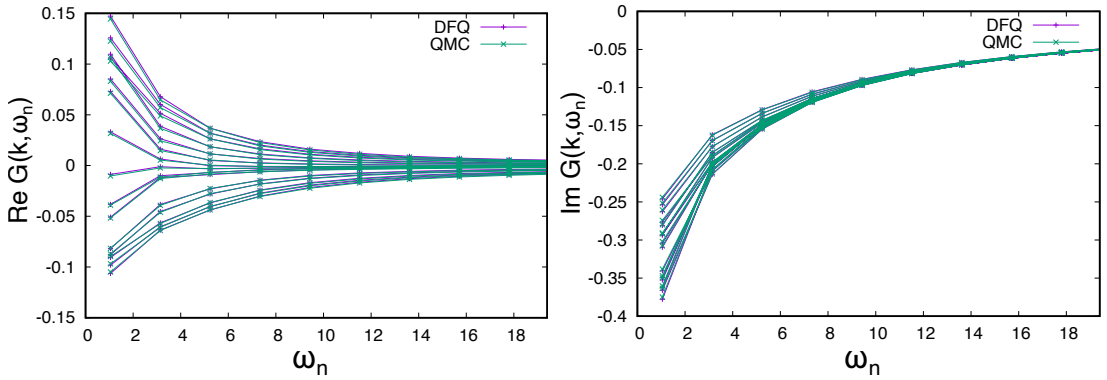


Fig. 11: *Green function for the* 8×8 *lattice with* $U = 5.56$, $t'/t = -0.3$, $\mu = -1.5$ $\beta = 3$ *for DF-QMC (DFQ) in comparison with DQMC test (QMC).*

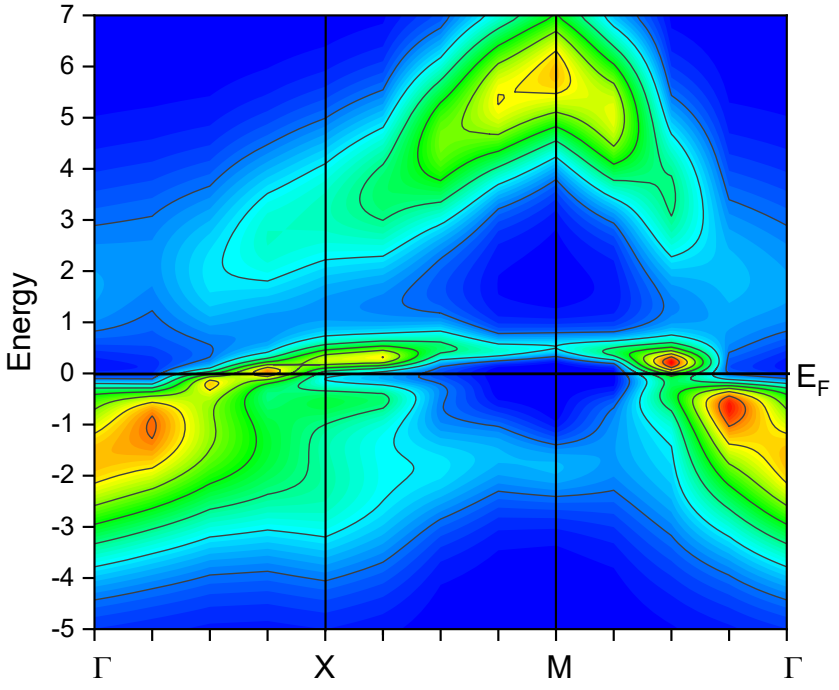


Fig. 12: *Spectral function* $-\Im G(\boldsymbol{k},\omega)/\pi$ *from dual-fermion QMC (CT-INT) for a* 8×8 *lattice with* $U/t = 8$, $t'/t = -0.3$, $\mu = -2.0$, *and* $\beta = 10$.

We have calculated the Green function for the doped two-dimensional Hubbard model for a periodic 8×8 system with U/t=8, t'/t=-0.3 and μ/t=-2 for β=$10/t$ using a CT-INT version of the CT-QMC scheme [24]. Note that for the non-interacting Green function we used the infinite-lattice limit with periodic boundary conditions for the calculated 8×8 system. This scheme reduces the cluster-size dependence for the bare Green function: in particular, the local one does not depend at all on the choice of the "simulation box". On the other hand, it may underestimate the effect of U-interactions, since they appears only in the calculated cluster. This may explain a small gap in the half-filled reference system compared to a standard lattice determinant QMC scheme [40]. The results for the first-order dual-fermion perturbation from the half-filled system indicate the formation of a correlated pseudogap electronic structure. Fig. 12 shows the color map of the spectral function along the irreducible path (Γ-X-M-Γ) in the square Brillouin zone. For analytical continuation we used the newly developed scheme [54].

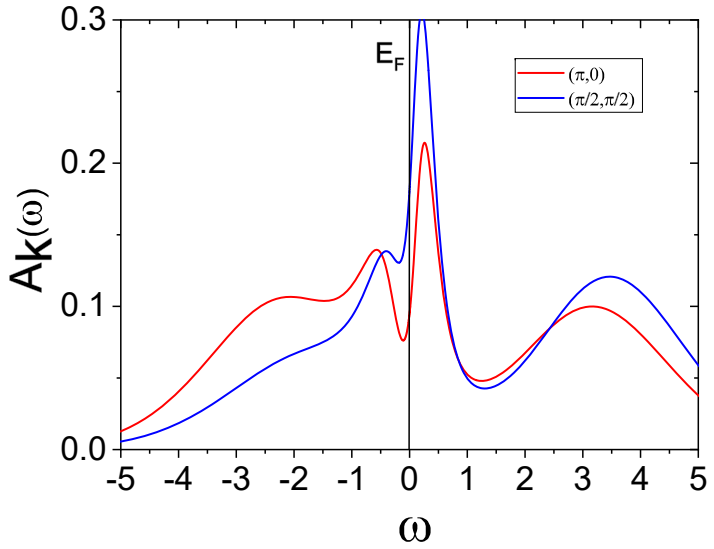


Fig. 13: *Spectral function* $-\Im G(\boldsymbol{k},\omega)/\pi$ *for two different* k*-points corresponding to the anti-nodal* $(\pi,0)$ *and nodal* $(\pi/2,\pi/2)$ *point in the dual fermion QMC (CT-INT) scheme for a* 8×8 *lattice with* $U/t=8$ $t'/t=-0.3$, $\mu=-2.0$ *and* $\beta=10$.

Several characteristic features of the correlated metallic phase in generic cuprate systems can be detected: the formation of an extended pseudogap region around the X-point towards the M-point, a shadow antiferromagnetic band at energy $-2t$ near the M-point, a strongly renormalized metallic band near the nodal point around $(\Gamma$-$M)/2$. Overall, the spectral function for $U=W$ clearly shows strong correlation features of the electronic structure far beyond a simple renormalized band paradigm.

In order to see the pseudogap and nodal-antinodal dichotomy more clearly we plot the energy dependence of the two spectral functions at the X- and the $(\Gamma$-$M)/2$-point in the Brillouin zone (Fig. 13). While at the X=$(\pi,0)$-point there is a reasonably deep pseudogap formation already at β=10, the nodal spectral function at $(\Gamma$-$M)/2$=$(\pi/2,\pi/2)$ shows correlated metallic behavior. A more unusual feature of the strong-coupling spectral function in Fig. 12 is related with a "shark mouth" pseudogap dip starting at X in the direction to M until half way. One can see from the energy dependence of the spectral function in the direction of X-M (Fig. 14 (middle)) that the pseudogap splitting of the sharp quasiparticle peak at zero for the $(X$-$M)/4$ point is even larger than at the X-point. The same feature was observed for a self-energy in a diagrammatic Monte Carlo (C-DET) investigation of the doped Hubbard model at $U/t=6$ [25]. We would like to point out that all these effects are not simply an artifact of the analytical continuation with the MaxEnt scheme and can be detected by inspection of the original complex Matsubara Green function from DF-QMC calculations (Fig. 15). If we compare the X=$(\pi,0)$ and $(X$-$M)/4$=$(\pi,\pi/4)$ points then both quasiparticle peaks are located almost at the Fermi energy (the real part of $G(\mathbf{k},\omega_n)$ is close to zero) but the pseudogap or upturn of the imaginary part of $G(\mathbf{k},\omega_n)$ for the first Matsubara frequencies are more pronounced at the $(\pi,\pi/4)$-point. We have also checked this characteristic feature with the Hirsch-Fye QMC scheme [43] and different MaxEnt implementations. The general structure of this spectral function is similar to recent results of dynamical variational Monte Carlo schemes [29, 30].

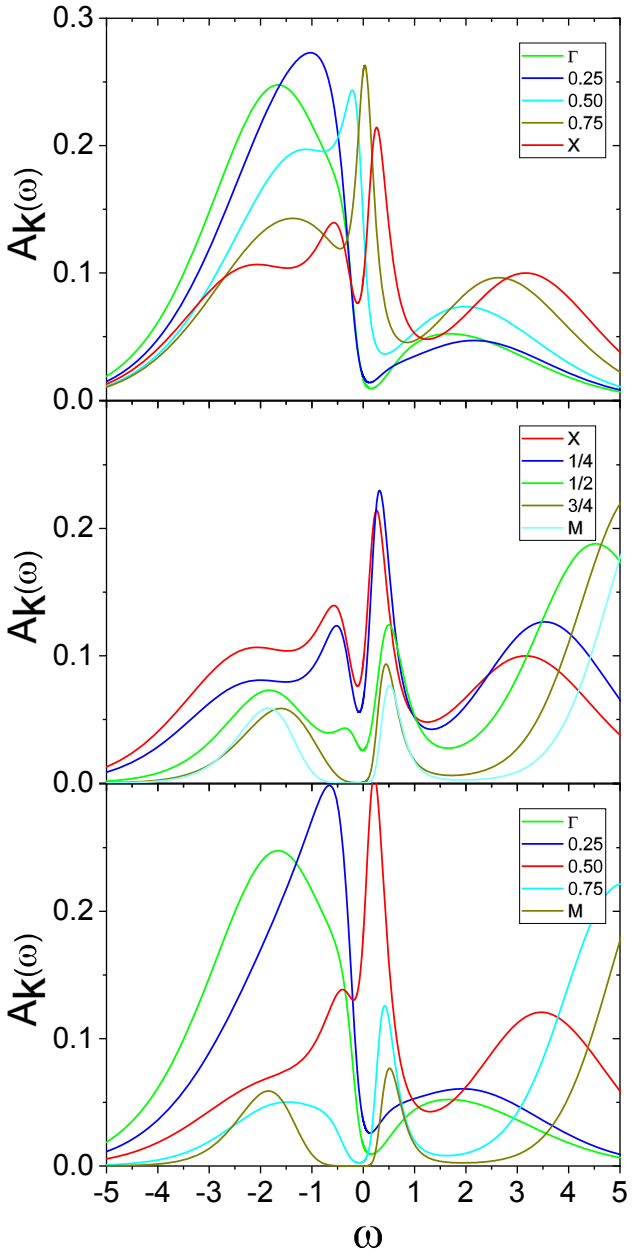


Fig. 14: *Spectral function* $-\Im G(\boldsymbol{k},\omega)/\pi$ *for three different* k*-directions in the Brillouin Zone, (top)* Γ*-*X*, (middle)* X*-*M *and (bottom) (*Γ*-*M*) dual fermion QMC (CT-INT) for a* 8×8 *lattice with* $U/t = 8$ $t'/t = -0.3$*,* $\mu = -2.0$ *and* $\beta = 10$*.*

We plot a broadened Fermi surface using the momentum-dependent spectral function for the first Matsubara frequency (Fig. 16). Comparison with the non-interacting tight-binding Fermi surface for the same doping shows a large region of the pseudogap around the X-point and formation of Fermi arcs near the nodal point. Moreover, one can understand that the pseudogap is more pronounced a bit away from the X-point towards the M-point, where the non-interacting Fermi surface crosses the Brillouin zone. We also compare the Fermi surface plot for smaller values of U/t=5.6, which was investigated with the diagrammatic Monte Carlo technique [55,56]; this value is related to a plaquette degenerate point [19]. While the Fermi surface for small U/t=5.6 agrees well with the results of the diagrammatic Monte Carlo approach [56]

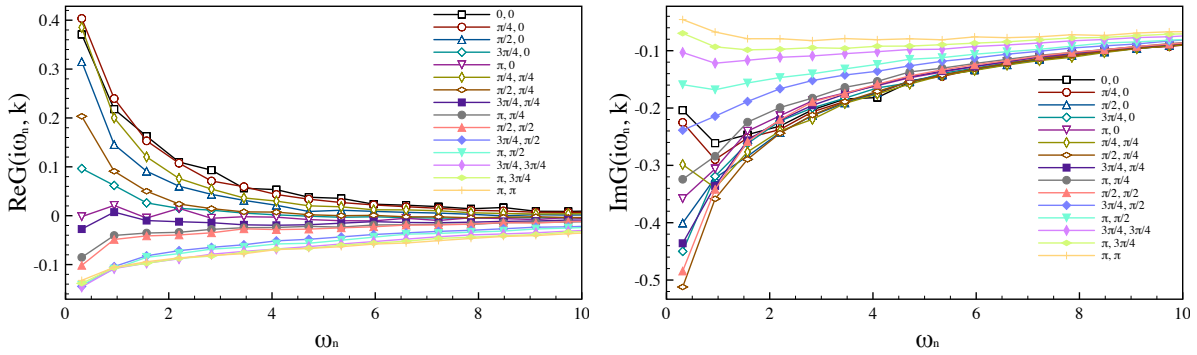


Fig. 15: *Green function $G(\boldsymbol{k}, \omega_n)$ on the Matsubara axes for all 15 non-equivalent k-points in the Brillouin Zone for 8×8 system, (left) Real part and (right) imaginary part for dual fermion QMC (CT-INT) with $U/t = 8$ $t'/t = -0.3$, $\mu = -2.0$ and $\beta = 10$.*

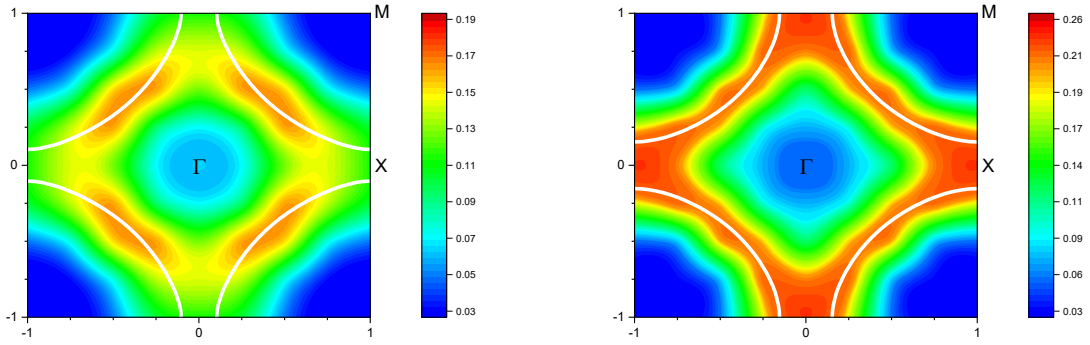


Fig. 16: *Fermi surface of the square-lattice Hubbard model or k-dependent spectral function at the first Matsubara frequency $-\Im G(\boldsymbol{k}, \omega_0)/\pi$ for dual fermion QMC (CT-INT) with $t'/t = -0.3$, $\beta = 10$ and $U/t = 8$, $\mu = -2.0$ (left) $U/t = 5.6$, $\mu = -0.9$ (right). The non-interacting Fermi surface with the same doping is shown for comparison as a white contour.*

and resembles the tight-binding one with only large broadening around the X-point, the U/t=8 results show already the formation of the pseudogap and Fermi arcs, that is, a nodal-antinodal dichotomy.

5 Discussion

We developed, for Hubbard-like correlated lattice models, the first-order strong-coupling dual fermion expansion in the shift of the chemical potential (doping) and in the second-neighbor hopping (t'). The starting reference point corresponds to the half-filled particle-hole symmetric system which can be calculated numerically exactly, without a fermionic sign problem. For the physically interesting parameter range of cuprate-like systems (around 10% doping and t'/t=-0.3 we can obtain a reasonable Green function for a periodic 8×8 lattice for the temperature T=$0.1t$. The formation of the pseudogap around the antinodal X-point and the nodal-antinodal dichotomy are clearly seen in the present approach.

We would like to point out a few main reasons why such a "super-perturbation" scheme works: first of all, the reference system already contains the main correlation effects which results in the four-peak structure of the density of states for the half-filled lattice Monte Carlo calculations [40]; second, the first-order strong-coupling perturbation relies on the lattice four-point vertex γ_{1234} (Eq. (23)) which is obtained numerically exactly and has all the information about the spin and charge susceptibilities of the lattice; and third, if the dual perturbation Green function $\tilde{G}^0_{12}$ (Eq. (21)) is relatively small, results will be reasonable. The complicated question of convergence for such a dual-fermion perturbation can be checked numerically by calculating the second-order contribution in $\tilde{\Sigma}_{12}$. For this term one needs to calculate within lattice QMC a six-point vertex $\gamma^{(6)}$ which will be a direction of future developments. In principle, one can also discuss an instability towards antiferromagnetism or d-wave superconductivity, introducing symmetry-breaking fields [15], which we also plan to investigate.
It is worthwhile to mention that for the starting reference system we can choose not only the half-filled case, but any doped case where the sign problem is mild, so we can use a QMC calculation to expand this numerically exact solution to "Terra Incognita" regions where the sign problem is unacceptable for direct QMC calculations.

Acknowledgements

The CT-INT part of the DF-QMC project was performed by Sergei Iskakov from the University of Michigan. We acknowledge valuable communications with Mikhail Katsnelson, Alexei Rubtsov, Evgeny Stepanov, Igor Krivenko, Richard Scalettar, Fedor Šimkovic IV, and Riccardo Rossi.

References

[1] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)

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

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

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

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

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

[7] S. Iskakov, M. Katsnelson, and A. Lichtenstein, arXiv (2023)

[8] R. Mondaini, S. Tarat, and R.T. Scalettar, Science **375**, 418 (2022)

[9] T. Schäfer, N. Wentzell, F. Šimkovic, Y.-Y. He, C. Hille, M. Klett, C.J. Eckhardt, B. Arzhang, V. Harkov, F.-M. Le Régent, A. Kirsch, Y. Wang, A.J. Kim, E. Kozik, E.A. Stepanov, A. Kauch, S. Andergassen, P. Hansmann, D. Rohe, Y.M. Vilk, J.P.F. LeBlanc, S. Zhang, A.-M.S. Tremblay, M. Ferrero, O. Parcollet, and A. Georges, Phys. Rev. X **11**, 011058 (2021)

[10] H. De Raedt and A. Lagendijk, Phys. Rep. **127**, 233 (1985)

[11] E.Y. Loh, J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R.L. Sugar, Phys. Rev. B **41**, 9301 (1990)

[12] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005)

[13] H.-C. Jiang and T.P. Devereaux, Science **365**, 1424 (2019)

[14] Y.-F. Jiang, J. Zaanen, T.P. Devereaux, and H.-C. Jiang, Phys. Rev. Res. **2**, 033073 (2020)

[15] C.-M. Chung, M. Qin, S. Zhang, U. Schollwöck, and S.R. White, Phys. Rev. B **102**, 041106 (2020)

[16] 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)

[17] M. Harland, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **94**, 125133 (2016)

[18] M. Harland, S. Brener, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **101**, 045119 (2020)

[19] M. Danilov, E.G. C.P. van Loon, S. Brener, S. Iskakov, M.I. Katsnelson, and A.I. Lichtenstein, npj Quant. Mater. **7**, 50 (2022)

[20] M.T. Schmid, J.-B. Morée, Y. Yamaji, and M. Imada, arXiv (2023)

[21] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S.R. White, and S. Zhang, arXiv (2023)

[22] N. Prokof'ev and B. Svistunov, Phys. Rev. Lett. **99**, 250201 (2007)

[23] R. Rossi, Phys. Rev. Lett. **119**, 045701 (2017)

[24] A.N. Rubtsov, V.V. Savkin, and A.I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005)

[25] F. Šimkovic, R. Rossi, A. Georges, and M. Ferrero, arXiv (2022)

[26] F. Šimkovic, R. Rossi, and M. Ferrero, Phys. Rev. Res. **4**, 043201 (2022)

[27] R. Rossi, N. Prokof'ev, B. Svistunov, K.V. Houcke, and F. Werner, EPL (Europhys. Lett.) **118**, 10004 (2017)

[28] A.J. Kim, N.V. Prokof'ev, B.V. Svistunov, and E. Kozik, Phys. Rev. Lett. **126**, 257001 (2021)

[29] M. Charlebois and M. Imada, Phys. Rev. X **10**, 041023 (2020)

[30] P. Rosenberg, D. Sénéchal, A.-M.S. Tremblay, and M. Charlebois, Phys. Rev. B **106**, 245132 (2022)

[31] S. Sakai, M. Civelli, and M. Imada, Phys. Rev. Lett. **116**, 057003 (2016)

[32] A. Singh, H.Y. Huang, J.D. Xie, J. Okamoto, C.T. Chen, T. Watanabe, A. Fujimori, M. Imada, and D.J. Huang, Nat. Commun. **13**, 7906 (2022)

[33] S. Brener, E.A. Stepanov, A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Ann. Phys. **422**, 168310 (2020)

[34] A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinski: *Methods of quantum field theory in statistical physics* (Dover, New York, 1975)

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

[36] R. Peierls, Phys. Rev. **54**, 918 (1938)

[37] R.P. Feynman: *Statistical mechanics: A set of lectures* (Benjamin/Cummings, 1972)

[38] N.N. Bogolyubov, Sov. Phys. Dokl. **3**, 292 (1958)

[39] R. Scalettar, R. Noack, and R. Singh, Phys. Rev. B **44**, 10502 (1991)

[40] D. Rost, E.V. Gorelik, F. Assaad, and N. Blümer, Phys. Rev. B **86**, 155109 (2012)

[41] J.E. Hirsch, Phys. Rev. B **28**, 4059 (1983)

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

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

[44] H. Hafermann, S. Brener, A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, JETP Lett. **86**, 677 (2008)

[45] S.K. Sarker, J. Phys. C: Solid State Phys. **21**, L667 (1988)

[46] W. Metzner, Phys. Rev. B **43**, 8549 (1991)

[47] S. Pairault, D. Sénéchal, and A.-M.S. Tremblay, Phys. Rev. Lett. **80**, 5389 (1998)

[48] S. Pairault, Sénéchal, D., and A.-M.S. Tremblay, Eur. Phys. J. B **16**, 85 (2000)

[49] N. Dupuis and S. Pairault, Int. J. Mod. Phys. B **14**, 2529 (2000)

[50] S. Brener, H. Hafermann, A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **77**, 195105 (2008)

[51] C. Gros and R. Valentí, Phys. Rev. B **48**, 418 (1993)

[52] E.W. Huang, S. Ding, J. Liu, and Y. Wang, Phys. Rev. Res. **4**, L042015 (2022)

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

[54] J. Fei, C.-N. Yeh, D. Zgid, and E. Gull, Phys. Rev. B **104**, 165111 (2021)

[55] W. Wu, M. Ferrero, A. Georges, and E. Kozik, Phys. Rev. B **96**, 041105 (2017)

[56] R. Rossi, F. Šimkovic, and M. Ferrero, EPL (Europhys. Lett.) **132**, 11001 (2020)

Index

Symbols
d- and f-orbitals, 7.8
d-shell orbitals, 9.5
e_g and t_{2g} orbitals, 4.17
e_g orbital Hubbard model, 6.7
f-orbitals, 16.3
$sp^3d^5f^7$ hybrids, 13.8
t_{2g} orbital Hubbard model, 6.24
10Dq, 4.17

A
Anderson condition, 3.13
Anderson model, 16.7
atomic limit, 5.23
atomic multiplets, 7.8
atomic orbitals, 1.40
atomic units, 1.40

B
Berry phase, 2.15
beyond-dipole terms, 7.3
Bohr magneton, 5.2
braiding statistics, 11.24
Buckminsterfullerene, 2.18

C
cerium, 16.3, 16.20
character table, 1.7
charge excitations in a Mott insulator, 6.10
charge fluctuations, 7.16
charge transfer, 4.19
  TM oxides, 9.7
charge-density wave, 3.2, 3.4
Clebsch-Gordan coupling coefficient, 2.5
co-operative Jahn-Teller distortion, 1.2
compass model, 11.3
complex Fermi Löwdin orbitals, 13.19
complex oxide heterostructures, 8.2
continuous-time quantum Monte Carlo, 17.7
Coulomb exchange, 5.4
Coulomb integral, 5.4
Coulomb phase, 12.6
covalency effects, 7.14
cRPA method, 16.14
crystal field, 4.14
  cubic, 1.6

D
degenerate Hubbard model, 6.10
  Kanamori parameters $\{U, J_H\}$, 6.9
  rotational invariance of interactions, 6.10
DFT+DMFT, 16.12
DFT+U, 16.19
diagonal sum-rule, 4.9
dimensional reduction, 11.10
Dirac monopole, 2.17
direct exchange, 5.8
DMFT, 16.9
double exchange, 5.18
downfolding, 1.16, 5.9, 5.29
downward quantum learning, 13.26
dual-fermion action, 17.10

E
effective interactions, 16.14
electron-lattice interaction, 3.2, 3.4
electronic structure of graphene-like $NdNiO_3$ (111), 9.22
embedding Hamiltonian, 15.20
entanglement, 5.9
exact diagonalization, 4.11
exchange
  Coulomb, 5.4
  direct, 5.8
  double, 5.18
  ferromagnetic superexchange, 5.16
  kinetic, 5.7
  orbital, 5.21
  superexchange, 5.13
exchange integral, 5.4
expectation values
  hopping, 15.10
  local, 15.9

F
Fermi Löwdin orbitals, 13.6
Fermi liquid, 10.21

fractionalized, 10.21
Fermi orbital descriptors, 13.14
Fermi sign, 4.12
Fermi's golden rule, 8.5
Fermi-function matrix, 15.33
Fermi-surface nesting, 3.2, 3.4
fermionic matrix representations, 15.28
fermionic similarity transformations, 15.30
FLO Centroids, 13.14
FLOSIC Software, 13.35
flotonic solids, 13.29
Fock space, 15.27
fractionalization, 12.9
frustrated magnet, 12.3
frustration, 10.7
exchange, 10.12
geometric, 10.7
function of matrix, 15.26

G
gauge invariance, 15.23
gauge-like symmetry, 11.8
Gaunt coefficient, 4.6
Germanium POSS, 2.8
Goodenough-Kanamori rules, 5.17, 5.19, 6.14, 6.23
Grassmann variables, 17.6
group theory, 1.6
Gutzwiller approximation, 15.9
Gutzwiller constraints, 15.7
gyromagnetic ratio, 5.2

H
hard X-ray photoelectron spectroscopy or HAXPES, 9.22
Hartree-Fock, 5.12
Heisenberg model, 5.25
Heisenberg-Kitaev model, 6.29
Heun equation, 2.13
honeycomb lattice, 3.19
Hubbard I, 16.18, 16.20
Hubbard model, 5.23, 16.4, 16.11, 17.4
Hubbard model of TM oxides, 9.6
Hubbard U, 4.7
Hubbard-Stratonovich transformation, 17.6, 17.9
Hund's exchange, 6.10
Hund's rules, 4.8, 5.6
hybridizing combinations, 4.21

I
interface
2D electron gas (2DEG), 9.13
charge transfer, 9.13
correlated oxides, 9.8
growth, 9.11
spin-polarized 2DEG, 9.15
tri-color superlattice, 9.15
interface reconstructions, 8.2, 8.3
ionic model of TM oxides, 9.4
irreducible representations, 1.8

J
Jahn-Teller effect, 1.19, 3.2
$E \times e$, 2.9
$T \times (e + t_2)$, 2.18
band, 3.4, 3.7
Mexican hat potential, 2.9, 2.11
pseudo, 2.7
Jahn-Teller theorem, 2.2

K
Kagome lattice, 3.21
$KCuF_3$, 1.14
Born-Mayer repulsion, 1.38
kinetic exchange, 5.7
Kitaev materials, 12.15
honeycomb iridates, 12.16
other, 12.19
$RuCl_3$, 12.18
Kitaev model, 11.16, 12.8
Kondo-breakdown transition, 10.24
Kugel-Khomskii mechanism, 3.2, 5.21
Kugel-Khomskii model, 1.28, 3.3, 6.16, 6.18
general form, 1.28
$LaVO_3$, a realization of, 1.37

L
Lagrange equations, 15.21
Lagrange function, 15.19
Landau-Ginzburg-Wilson theory, 10.2
lattice energy, 9.2
lattice engeneering
geometrical or GLE, 9.9
graphene-like $NdNiO_3$ (111), 9.19

level repulsion, 4.19
ligand field, 1.6
localization equations, 13.5
Luttinger theorem, 10.21

M
magnon dressing by orbitals, 6.21
Majorana representation, 11.18
metal, polar correlated, 9.13
metal-insulator transition, 5.24
Mott insulator, 3.2, 6.2, 9.6
Mott transition, 5.23, 9.7, 10.22, 16.6
    orbital-selective, 10.24
multi-orbital gGA ansatz, 15.4
multi-orbital Hubbard Hamiltonian, 15.3
multiband Hubbard models, 4.23
multiplets, 4.3, 5.6, 16.21

N
non-Fermi liquid, 10.21
non-resonant inelastic scattering, 7.2

O
octahedral coordination, 4.15
on-site Coulomb repulsion U, 6.2, 6.9
optical spectral weight, 6.13, 6.24, 6.25
orbital
    fluctuations, 6.2, 6.25, 6.28
    order, 1.2, 3.2, 5.21, 6.13, 6.20, 6.25
    polarization, 8.21, 8.23
orbital degrees of freedom, 1.2
orbital imaging, 7.6
orbital reflectometry, 8.19
order by disorder, 10.11

P
path-integral formalism, 17.2
Pauli matrices, 5.28
Pauling rules, 9.2
Peierls transition, 3.2, 3.4, 3.12
perovskite crystal structure, 8.2, 9.3
pseudospin, 1.2
pulsed laser deposition or PLD, 9.11
pyrochlore lattice, 3.5

Q
quantum critical point, 6.19
quantum criticality, 10.2
    conventional, 10.13
    deconfined, 10.18
    fractionalized, 10.15
quantum dynamics, 15.24
quantum Monte Carlo, 17.2
quantum phase transition, 10.2, 10.13, 10.15, 10.20
quantum-embedding reformulation, 15.24
quantum-embedding states, 15.15
quasiparticle Hamiltonian, 15.20

R
real harmonics, 1.41
reduced many-body density-matrix, 15.36
reference system, 17.6
representation theory
    irreducible representations, 1.8
resonant elastic X-ray scattering, 8.12
rutile structure, 2.2

S
scattering tensor, 8.6
second quantization, 5.10
sign problem, 17.5
Slater-Condon parameters, 4.7
Slater-Koster tables, 4.20
Slater-Koster two-center integrals, 1.42
slave-boson amplitudes, 15.13
slave-boson methods, 15.2
sodium trimer, 2.10
spectral function, 15.24
spin liquid, 12.2
    classical, 10.8, 12.5
    Kitaev, 10.12
    proximate, 12.16
    quantum, 10.9, 12.7
      concept, 9.24
      design, 9.25
      in $CoCr_2O_4$ (111) spinel, 9.26
    thermal signatures, 12.10
spin-density wave, 3.2, 3.4
spin-orbit coupling, 4.13, 6.3, 16.23, 16.25
spin-orbit entanglement, 12.12
spin-orbit interaction, 3.2
spin-orbital entanglement, 6.18, 6.20, 6.25, 6.30
    in one dimension, 6.3
    von Neumann entropy, 6.4
spin-orbital liquid, 10.12

spin-orbital model, 6.12
for K_2CuF_4 , 6.20
for $KCuF_3$, 6.16
for $LaMnO_3$, 6.22
for $LaVO_3$, 6.24
for $NaTiO_2$, 6.27
spinel crystal structure, 9.4
square lattice, 3.7
strong coupling, 17.2
expansion, 17.8
sum rules, 8.9, 8.11
superexchange, 5.13
superexchange interaction, 1.28, 6.2
superlattice, 8.4, 8.14, 8.19
symmetry
emergent, 11.8
exact, 11.8
gaugle-like, 11.8
symmetry breaking in FLOSIC, 13.9
symmetry group
I_h, 2.18
O_h, 2.2, 2.8, 2.18
$SO(5)$, 2.21

T
Tanabe-Sugano diagrams, 4.17
thermal Hall effect, 12.18
tight-binding method, 1.13
d bands, 1.13
triangular lattice, 3.10, 3.16

U
UO_2, 16.15, 16.28

W
Wannier functions, 13.10, 16.12
Wick's theorem, 15.34, 17.10

X
X-ray absorption cross section, 8.8
X-ray absorption spectroscopy, 8.6
X-ray linear dichroism, 8.9, 8.17
X-ray magnetic circular dichroism, 8.9, 8.11
X-ray resonant reflectivity, 8.14

Z
Zaanen-Sawatzky-Allen phase diagram, 9.8

Schriften des Forschungszentrums Jülich
Modeling and Simulation

1. **The LDA+DMFT approach to strongly correlated materials**
   Lecture Notes of the Autumn School 2011 Hands-on LDA+DMFT
   edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2011), 420 pages
   ISBN 978-3-89336-734-4

2. **Correlated Electrons: From Models to Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2012
   edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (2012), 450 pages
   ISBN 978-3-89336-796-2

3. **Emergent Phenomena in Correlated Matter**
   Lecture Notes of the Autumn School on Correlated Electrons 2013
   edited by E. Pavarini, E. Koch, and U. Schollwöck (2013), 520 pages
   ISBN 978-3-89336-884-6

4. **DMFT at 25: Infinite Dimensions**
   Lecture Notes of the Autumn School on Correlated Electrons 2014
   edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2014), 450 pages
   ISBN 978-3-89336-953-9

5. **Many-Body Physics: From Kondo to Hubbard**
   Lecture Notes of the Autumn School on Correlated Electrons 2015
   edited by E. Pavarini, E. Koch, and P. Coleman (2015), 500 pages
   ISBN 978-3-95806-074-6

6. **Quantum Materials: Experiments and Theory**
   Lecture Notes of the Autumn School on Correlated Electrons 2016
   edited by E. Pavarini, E. Koch, J. van den Brink, G. Sawatzky (2016), 420 pages
   ISBN 978-3-95806-159-0

7. **The Physics of Correlated Insulators, Metals, and Superconductors**
   Lecture Notes of the Autumn School on Correlated Electrons 2017
   edited by E. Pavarini, E. Koch, R. Scalettar, and R. Martin (2017), 450 pages
   ISBN 978-3-95806-224-5

8. **DMFT: From Infinite Dimensions to Real Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2018
   edited by E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (2018), 480 pages
   ISBN 978-3-95806-313-6

9. **Many-Body Methods for Real Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2019
   edited by E. Pavarini, E. Koch, and S. Zhang (2019), 520 pages
   ISBN 978-3-95806-400-3

Schriften des Forschungszentrums Jülich
Modeling and Simulation

10. **Topology, Entanglement, and Strong Correlations**
    Lecture Notes of the Autumn School on Correlated Electrons 2020
    edited by E. Pavarini and E. Koch (2020), 500 pages
    ISBN 978-3-95806-466-9

11. **Simulating Correlations with Computers**
    Lecture Notes of the Autumn School on Correlated Electrons 2021
    edited by E. Pavarini and E. Koch (2021), 420 pages
    ISBN 978-3-95806-529-1

12. **Dynamical Mean-Field Theory of Correlated Electrons**
    Lecture Notes of the Autumn School on Correlated Electrons 2022
    edited by E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt (2022), 450 pages
    ISBN 978-3-95806-619-9

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