1 SCIENTIFIC HIGHLIGHT OF THE MONTH

Dynamical Screening Effects in Correlated Electron Materials – A Progress Report on Combined Many-Body Perturbation and Dynamical Mean Field Theory: “GW+DMFT”

Silke Biermann

Centre de Physique Théorique, Ecole Polytechnique, 91128 Palaiseau, France

Abstract

We give a summary of recent progress in the field of electronic structure calculations for materials with strong electronic Coulomb correlations. The discussion focuses on developments beyond the by now well established combination of density functional and dynamical mean field theory dubbed “LDA+DMFT”. It is organized around the description of dynamical screening effects in the solid. Indeed, screening in the solid gives rise to dynamical local Coulomb interactions $U_{\omega}$ [1, 2], and this frequency-dependence leads to effects that cannot be neglected in a truly first principles description [3]. We review the recently introduced extension of LDA+DMFT to dynamical local Coulomb interactions “LDA+$U_{\omega}$+DMFT” [4, 5]. A reliable description of dynamical screening effects is also a central ingredient of the “GW+DMFT” scheme [6, 7, 8], a combination of many-body perturbation theory in Hedin’s GW approximation and dynamical mean field theory. Recently, the first GW+DMFT calculations including dynamical screening effects for real materials have been achieved, with applications to SrVO$_3$ [9] and adatom systems on surfaces [10]. We review these and comment on further perspectives in the field.

This review is an attempt to put elements of the original works [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] into the broad perspective of the development of truly first principles techniques for correlated electron materials.

1 Electronic Correlations from First Principles?

The electronic properties of solids are dominated by the electronic states in the immediate proximity to the Fermi level. This observation – together with the insights from renormalisation group techniques – are the motivation for the quest of low-energy effective models that describe the physical phenomena taking place in condensed matter. In the early days of correlated electron physics models were most often phenomenologically motivated, without the ambition of a microscopic derivation, let alone a quantitative description. Within the last decade, however, a new research field has developed at the interface of many-body theory and first principles electronic structure calculations. The aim is the construction of materials-specific parameter-free many-body theories that preserve the ab initio nature of density functional based electronic structure calculations, but incorporate at the same time a many-body description of Coulomb interactions beyond the independent-electron picture into the description of spectroscopic or finite-temperature properties.
Such “correlation” effects, that is, effects beyond the effective one-particle picture, are indeed most striking in spectroscopic probes, where they take the form of quasi-particle renormalisations or broadening due to finite lifetimes, and give rise to satellite features or atomic multiplets. An intrinsic temperature dependence of the electronic structure of a metal, with a coherence-incoherence crossover delimiting Fermi liquid properties, or a strongly temperature-dependent gap – beyond what can be explained by a Fermi factor – are further hallmarks of electronic correlations [11].

Historically, the first non-perturbative electronic structure techniques for correlated materials evolved from many-body treatments of the multi-orbital Hubbard Hamiltonian with realistic parameters. The general strategy of these so-called “LDA++” approaches consists in the extraction of the parameters of a many-body Hamiltonian from first principles calculations and then solving the problem by many-body techniques. The procedure becomes conceptually involved, however, through the need of incorporating effects of higher energy degrees of freedom on the low energy part, the so-called “downfolding”.

For the one-particle part of the Hamiltonian, downfolding techniques have been the subject of a vast literature [121, 122], and are by now well established. Downfolding of the interacting part of a many-body Hamiltonian is a less straightforward problem, which has attracted a lot of attention recently. The challenge is an accurate description of screening of low-energy interactions by high-energy degrees of freedom. Indeed, the net result of the rearrangement of the high-energy degrees of freedom as response to a perturbation of the system is an effective reduction of the perturbation strength in the low-energy space. It is for this reason that the effective Coulomb interaction in a low-energy effective model for a correlated system is in general an order of magnitude smaller than the matrix element of the bare Coulomb interaction. Nevertheless, the latter is recovered in the limit of high-frequencies of the perturbation, when screening becomes inefficient. The crossover – as a function of frequency – from the low-energy screened regime to the high-frequency bare matrix element of \( \frac{e^2}{|r-r'|} \) takes place at a characteristic screening (plasma) frequency where the dielectric function exhibits a pole structure.

This frequency-dependence of the effective local Coulomb interactions, the dynamical Hubbard \( U(\omega) \) and its consequences on the electronic structure of correlated materials are at the center of the present review. We first recall the formalism of the constrained random phase approximation, as the simplest means to obtain a quantitative estimate for the dynamical Hubbard interactions. We then review a recent scheme to incorporate the dynamical nature of the Hubbard interactions into dynamical mean field based electronic structure calculations. This “Bose factor ansatz” also gives a transparent physical interpretation of the observed new features, such as plasmon satellites and renormalisations of spectral weight at low energies. Electronic structure calculations with frequency-dependent interactions – using a scheme that should best be called “LDA+\( U(\omega) \)+DMFT” – for the iron pnictide compound BaFe\(_2\)As\(_2\) illustrate the importance of these effects, while at the same time revealing new unexpected many-body behavior in the form of an incoherent (non-Fermi-liquid) regime.

Dealing with frequency-dependent interactions at the DMFT level has been a major bottleneck in the implementation of the combined “GW+DMFT” scheme since its proposal in 2003 [6]. The recent advances concerning this issue, both concerning Monte Carlo techniques and through the
Bose factor ansatz, have now unblocked the situation: two calculations within GW+DMFT taking into account dynamical interactions have been achieved recently, for SrVO$_3$ [9] and for systems of adatoms on surfaces [10]. We review these calculations, together with systematic studies of an extended Hubbard model [7], which demonstrate how the GW+DMFT scheme enables an additional type of “downfolding”: effects of long-range interactions can in fact be “backfolded” into a purely local effective quantity, a generalised Hubbard $U(\omega)$, which acquires its frequency-dependence due to screening by non-local processes. The strength of these screening processes are shown to be strongly system-dependent when the true long-range nature of Coulomb interactions is taken into account, while simple rules of thumb work relatively well in the case of an extended Hubbard model with nearest-neighbor interactions only.

Finally, we define current open questions and comment on further perspectives in the field.

2 Calculating effective local Coulomb interactions from first principles: from “Hubbard U” to dynamical Hubbard $U(\omega)$

We now turn to a mathematical formulation of the above described frequency-dependent behavior of the effective Coulomb interactions (“Hubbard U”) to be used within low-energy effective descriptions. The simplest way to describe the frequency-dependence stemming from downfolding higher energy degrees of freedom is given by the “constrained random phase approximation” (cRPA) technique [1]. The cRPA provides an (approximate) answer to the following question: given the Coulomb Hamiltonian in a large Hilbert space, and a low-energy Hilbert space that is a subspace of the former, what is the effective bare interaction to be used in many-body calculations dealing only with the low-energy subspace, in order for physical predictions for the low-energy Hilbert space to be the same for the two descriptions? A general answer to this question not requiring much less than a full solution of the initial many-body problem, the cRPA builds on two approximations: it assumes (i) that the requirement of the same physical predictions be fulfilled as soon as in both cases the same estimate for the fully screened Coulomb interaction, Hedin’s $W$, is obtained and (ii) the validity of the random phase approximation to calculate this latter quantity.

The cRPA starts from a decomposition of the polarisation of the solid in high- and low-energy parts, where the latter is defined as given by all screening processes that are confined to the low-energy subspace. The former results from all remaining screening processes:

$$P^\text{high} = P - P^\text{low},$$

One then calculates a partially screened interaction

$$W^\text{partial}(1, 2) \equiv \int d^3 \varepsilon^\text{-1}_{\text{partial}}(1, 3) v(3, 2).$$

using the partial dielectric function

$$\varepsilon_{\text{partial}}(1, 2) = \delta(1 - 2) - \int d^3 P^\text{high}(1, 3) v(3, 2).$$

Here, the numbers represent space and time coordinates in a shorthand notation.
Screening $W^{\text{partial}}$ by processes that live within the low-energy space recovers the fully screened interaction $W$. This justifies the interpretation of the matrix elements of $W^{\text{partial}}$ in a localized Wannier basis as the interaction matrices to be used as bare Hubbard interactions within a low-energy effective Hubbard-like Hamiltonian written in that Wannier basis.

Hubbard interactions – obtained as the static ($\omega = 0$) limit of $\langle |W^{\text{partial}}| \rangle$ within cRPA – have by now been obtained for a variety of systems, ranging from transition metals [1] to oxides [12, 13, 14, 2], pnictides [15, 16, 17, 18], or f-electron compounds [19], and several implementations within different electronic structure codes and basis sets have been done, e.g. within linearized muffin tin orbitals [1], maximally localized Wannier functions [12, 20, 16], or localised orbitals constructed from projected atomic orbitals [2]. The implementation into the framework of the Wien2k package [2] made it possible that Hubbard $U$’s be calculated for the same orbitals as the ones used in subsequent LDA+DMFT calculations, see e.g. [21]. Systematic calculations investigating the basis set dependence for a series of correlated transition metal oxides revealed furthermore interesting trends, depending on the choice of the low-energy subspace. In contrast to common belief until then, Hubbard interactions increase for example with the principal quantum number when low-energy effective models encompassing only the $t_{2g}$ orbitals are employed. These trends can be rationalised by two counteracting mechanisms, the increasing extension of the orbitals with increasing principal quantum number and the less efficient screening by oxygen states [2].

![Figure 1](image.png)

Figure 1: Screened Coulomb interaction $W$ and partially screened Coulomb interaction $U$ for SrVO$_3$: Re$U$, Re$W$ (top panel), Im$U$, Im$W$ (bottom panel), as calculated within cRPA. Note the small low energy value of Re$U(0)$, compared to the matrix element of the bare Coulomb interaction Re$U(\infty)$, and the plasmon excitation at 15 eV. Adapted from [9].

In general, values obtained within cRPA have for a long time been believed to be slightly “too small”, since quite systematically not only constrained LDA techniques result in larger values but also many-body calculations that fix the interactions in order to obtain agreement with experiments usually employ slightly larger values than those obtained within cRPA. This puzzle has been recently solved [4, 3]: the key was found to lie in the frequency-dependence of the
interactions leading to additional renormalisations of the one-body Hamiltonian. Indeed, as can be seen from Eq. (2), \( W_{\text{partial}}(\omega) \) is a function of frequency, and so are matrix elements derived from it, in particular its local part, the Hubbard \( U(\omega) \). An example is given in Fig. 1. The consequences of this dynamical nature of the effective interactions are the subject of the following sections.

3 Dynamical screening effects: plasmons, spectral weight transfers and electronic polarons

The explicit treatment of many-body problems with dynamical Hubbard interactions has by now become possible even in the realistic multi-orbital case. This progress is due both to quite impressive advances in Monte Carlo techniques and to the development of extremely accurate efficient approximations, that reduce the problem to a static one, at least in the antiadiabatic limit when the characteristic screening frequencies are much larger than other relevant energy scales of the problem (bandwidth and static Hubbard interaction \( U(\omega = 0) \)). Several applications to materials have appeared, namely for SrVO\(_3\) in [4, 9, 22], and to BaFe\(_2\)As\(_2\) in [5]. Alternatively to a direct explicit treatment of the dynamical interactions, in the antiadiabatic limit a mapping onto an effective low-energy model with static interactions can also be performed, if only low-energy properties, living on energy scales considerably smaller than the plasma frequency, are of interest [3]. This procedure is reviewed in Appendix A.

In the remainder of this section, we focus on the treatment of dynamical Hubbard interactions within the “Bose factor ansatz” (BFA), which allows for a transparent physical interpretation of the observed effects. Moreover, comparison of calculations for SrVO\(_3\) within the BFA in [4] and within Monte Carlo in [22] demonstrate the impressive accuracy of this approach in the regime relevant for real materials applications.

Extending the philosophy of the LDA+DMFT scheme to dynamically screened interactions requires the use of a framework that allows for a description of an explicit frequency-dependence of the interactions \( U(\omega) \). One possibility is to switch from the Hamiltonian formulation of the “LDA+++” approach to an action description where the frequency-dependent nature of the interaction is readily incorporated as a retardation in the interaction term

\[
S = - \int_0^\beta \int_0^\beta d\tau d\tau' U(\tau - \tau') n(\tau) n(\tau')
\]

where we have assumed that the retarded interaction couples only to the density \( n(\tau) \). Alternatively, it is possible to stick to a Hamiltonian formulation. In order to describe the retardation effects in the interaction one then needs to introduce additional bosonic degrees of freedom that parametrise the frequency-dependence of the interaction. Indeed, from a physical point of view, screening can be understood as a coupling of the electrons to bosonic screening degrees of freedom such as particle-hole excitations, plasmons or more complicated composite excitations giving rise to shake-up satellites or similar features in spectroscopic probes. Mathematically, a local retarded interaction can be represented by a set of bosonic modes of frequencies \( \omega \) coupling to the electronic density with strength \( \lambda_\omega \). The total Hamiltonian

\[
H = H_{\text{LDA+++}} + H_{\text{screening}}
\]
is then composed by a part of “LDA++” form but with the local interactions given by the unscreened local matrix elements of the bare Coulomb interactions $V$ and the Hund’s exchange coupling $J$ (assumed not to be screened by the bosons and thus frequency-independent)

$$H_{\text{LDA++}} = H^{KS} + \frac{1}{2} \sum_{i m m' \sigma} V_{i m m' \sigma} n_{i m \sigma} n_{i m' \sigma} + \frac{1}{2} \sum_{i m \neq m' \sigma} (V_{i m} - J_{i m}) n_{i m \sigma} n_{i m' \sigma} \quad (6)$$

and a screening part consisting of the local bosonic modes and their coupling to the electronic density:

$$H^{\text{screening}} = \sum_i \int d\omega \left[ \lambda_i \omega (b_i^\dagger \omega + b_i \omega) \sum_{m \sigma} n_{i m \sigma} + \omega b_i^\dagger \omega b_i \omega \right].$$

Here, $H^{KS}$ represents a one-body Hamiltonian defined by the DFT Kohn-Sham band structure, suitably corrected for double counting terms. As in standard LDA+DMFT, many-body interactions are included for a selected set of local orbitals, assumed to be “correlated”. The sums thus run over atomic sites $i$ and correlated orbitals $m$ centered on these sites.

Integrating out the bosonic degrees of freedom would lead back to a purely fermionic action with retarded local interactions

$$U(\omega) = V + \int d\omega' \lambda_{\omega}^2 \left( \frac{1}{\omega - \omega'} - \frac{1}{\omega + \omega'} \right) \quad (7)$$

The above Hamiltonian thus yields a parametrisation of the problem with frequency-dependent interactions provided that the parameters are chosen as $\text{Im} U(\omega) = \pi \lambda_{\omega}^2$. The zero-frequency (screened) limit is then given by $U_0 = V - 2 \int d\omega \lambda_{\omega}^2$.

The above form of the Hamiltonian corresponds to a multi-orbital multi-mode version of the familiar Hubbard-Holstein Hamiltonian describing a system of fermions coupled to bosonic modes. The emergence of retarded interactions in the case of electron-phonon coupling has been investigated in detail for its role in the BCS theory of pairing arising in conventional superconductors. In the current situation where the bosons represent plasmons and other screening modes typical energy scales are radically different, and the regime of importance is most often the antiadiabatic one. Indeed, plasma frequencies of typical transition metal based materials – transition metals themselves, their oxides, pnictides etc. – are of the order of 15 to 25 eV, whereas both the typical bandwidth and static ($\omega = 0$) Hubbard interaction are rather of the order of a few eV. This hierarchy gives rise to a separation of energy scales that enables a simple and transparent physical interpretation of the solution of the Hubbard-Holstein Hamiltonian.

We illustrate this fact on the most simple example, a half-filled single-orbital Hubbard-Holstein model with a single local bosonic mode on a Bethe lattice with semi-circular density of states. The frequency $\omega_0$ of the bosonic mode is chosen to be the largest energy scale of the problem so that the chosen parameter set places the system deep in the antiadiabatic limit. The spectral function in this case is plotted in Fig. 2. It corresponds to a sequence of features located at energies that are positive or negative multiples of the plasma frequency. They correspond to electron removal or addition processes where the (inverse) photoemission process itself is accompanied by the creation or annihilation of a certain number of screening bosons. The low-energy part of the spectral function, close to the Fermi level – chosen to be the origin of energies – is given by electron removal or addition processes that do not change the number of screening
bosons. In the present simple half-filled case in a moderate correlation regime, it displays a “three-peak structure”, with a central quasi-particle peak and upper and lower Hubbard bands, typical of correlated metals. Interestingly, however, even this part is modified by the coupling to the bosons: indeed, since the full spectral function is normalised, spectral weight appearing in plasmon replicae of the main line reduces the weight contained in the latter. The coupling to the bosonic degrees of freedom thus leads to an additional mass renormalisation of the low-energy fermionic degrees of freedom. This effect corresponds to the mass enhancement due to the formation of “electronic polarons”, fermions dressed by their screening bosons just as usual polarons can be understood as electrons dressed by the polarisation of the surrounding lattice. In the case of core level spectroscopies, such effects have been extensively discussed, and the electron-boson couplings above can be viewed as a local version of Hedin’s “fluctuation potentials” (albeit, in the cRPA sense, that is parametrising not the fully screened interaction \( W \) but rather the Hubbard \( U \)) \cite{23}.

When solving a many-body problem with static interactions – to simplify the notation we discuss here the case of a single-orbital problem only – within DMFT, the central ingredient is the solution of an effective local problem (“impurity problem”)}

\[
S = - \int d\tau d\tau' \sum_{\sigma} \epsilon_{\sigma}^{\dagger}(\tau')(\delta(\tau - \tau')\delta_{\tau} - \Delta(\tau - \tau'))c_{\sigma}(\tau) + \int d\tau U_{n_{\uparrow}n_{\downarrow}}
\]  

with an effective local bath propagator \( \mathcal{G}^{-1}_{0}(i\omega_{n}) = i\omega_{n} + \mu - \Delta(i\omega_{n}) \), the “dynamical mean field”. The latter has to be determined through a self-consistency condition, expressing the translational invariance of the solid and thus the equivalence of different atomic sites. For further details we refer the interested reader to the many excellent reviews about DMFT. In the present context, we restrict ourselves to a discussion of how the above construction is modified when dynamical interactions are taken into account. What is the relevant impurity model if we want to solve a lattice model with purely local but dynamical effective Hubbard interactions? The answer is a
straightforward generalisation to frequency-dependent $U(\omega)$ of the above action:

$$S = -\int d\tau d\tau' \sum_{\sigma} c^\dagger(\tau')(\delta(\tau - \tau')\partial_\tau - \Delta(\tau - \tau'))c(\tau) + \int d\tau d\tau' U(\tau - \tau') n(\tau)n(\tau')$$  \hspace{1cm} (9)

where $n(\tau) = n_\uparrow(\tau) + n_\downarrow(\tau)$.

An extremely efficient scheme for the solution of this problem, suitable in the antiadiabatic regime, is the recently introduced [4] “Boson factor ansatz” (BFA). It approximates the local Green’s function of the dynamical impurity model as follows:

$$G(\tau) = -\langle T c(\tau)c^\dagger(0) \rangle = \left( \frac{G(\tau)}{G_{\text{stat}}(\tau)} \right) G_{\text{stat}}(\tau) \sim \left( \frac{G(\tau)}{G_{\text{stat}}(\tau)} \right) \bigg|_{\Delta=0} G_{\text{stat}}(\tau)$$  \hspace{1cm} (10)

where $G_{\text{stat}}$ is the Green’s function of a fully interacting impurity model with purely static interaction $U=U(\omega=0)$, and the first factor is approximated by its value for vanishing bath hybridization $\Delta$ [4]. In this case, it can be analytically evaluated in terms of the frequency-dependent interaction:

$$B(\tau) = \left( \frac{G(\tau)}{G_{\text{stat}}(\tau)} \right) \bigg|_{\Delta=0} = e^{-\int_0^\infty \frac{\text{Im} U(\omega)}{\beta \omega} (K_+(\omega)-K_0(\omega))}$$  \hspace{1cm} (11)

with $K_+(\omega) = \frac{\exp(-\tau\omega)+\exp(-\beta(\tau-\omega))}{1-\exp(-\beta\omega)}$. In the regime that we are interested in, namely when the plasma frequency that characterises the variation of $U$ from the partially screened to the bare value, is typically several times the bandwidth, this is an excellent approximation, as was checked by benchmarks against direct Monte Carlo calculations in Ref. [4]. The reason can be understood when considering the solution of the dynamical local model in the dynamical atomic limit $\Delta = 0$, that is, when there are no hopping processes possible between the impurity site and the bath. In this case the BFA trivially yields the exact the solution, and the factorisation can be understood as a factorisation into a Green’s function determined by the static Fourier component of $U$ only and the exponential factor $B$ which only depends on the non-zero frequency components of $U$. The former fully determines the low-energy spectral function of the problem, while the latter is responsible for generating high-energy replicae of the low-energy spectrum. For finite bath hybridisation, the approximation consists in assuming that the factorisation still holds and that the finite bath hybridisation modifies only the low-energy static-$U$ Green’s function, leaving the general structure of the plasmon replicae generation untouched. The approximation thus relies on the energy scale separation between low-energy processes and plasmon energy; it becomes trivially exact not only in the atomic limit but also in the static limit, given by small electron-boson couplings or large plasmon energy.

The BFA lends a precise mathematical meaning to the physical discussion of the generation of plasmon replicae. Indeed, the factorisation of the Green’s function corresponds in frequency space to a convolution of the spectral representations of the low-energy Green’s function $G_{\text{static}}$ and the bosonic factor $B$. In terms of the spectral function $A_{\text{stat}}(\omega)$ of the static Green’s function $G_{\text{stat}}(\omega)$ and the (bosonic) spectral function $B(\epsilon)$ of the bosonic factor $B(\tau)$ defined above the spectral function $A(\omega)$ of the full Green’s function $G(\tau)$ reads:

$$A(\omega) = \int_{-\infty}^\infty d\epsilon \frac{1 + e^{-\beta\omega}}{(1 + e^{-\beta(\omega-\epsilon)})(1 - e^{-\beta\epsilon})} A_{\text{stat}}(\omega - \epsilon).$$  \hspace{1cm} (12)
Figure 3: Spectral function of a low-energy ($t_{2g}$-only) Hamiltonian within “LDA+$U(\omega)$+DMFT” as compared to a standard LDA+DMFT calculation with static interactions, see text. From [4].

In the case of a single mode of frequency $\omega_0$, the bosonic spectral function consists of sharp peaks at energies given by that frequency, and the convolution generates replicae of the spectral function $A_{\text{stat}}(\omega)$ of the static part. Due to the overall normalisation of the spectral function, the appearance of replicae satellites is necessarily accompanied by a transfer of spectral weight to high-energies. This mechanism induces a corresponding loss of spectral weight in the low-energy part of the spectral function. Indeed, it can be shown [3] that the spectral weight corresponding to the low-energy part as defined by a projection on zero boson states is reduced by the factor

$$Z_B = \exp\left(-\frac{1}{\pi} \int_0^\infty d\nu \frac{\text{Im}U(\nu)}{\nu^2}\right).$$

(13)

Estimates of $Z_B$ for typical transition metal oxides vary between 0.5 and 0.9, depending on the energy scale of the plasma frequency and the efficiency of screening (as measured e.g. by the difference between bare Coulomb interaction $\langle |r-r'|| \rangle = U(\omega = \infty)$ and the static value $U(\omega = 0)$).

We reproduce in Fig.3 the low-energy spectral function of an “LDA+$U(\omega)$+DMFT” calculation for the $d^1$ ternary transition metal perovskite SrVO$_3$, demonstrating the reduction of spectral weight compared to a static-U calculation [3]. It should be noted however, that the calculation included the $t_{2g}$ states only. We will show below that the contribution of unoccupied $e_g$ states dominates at energies as low as $\sim 2.5$ eV. Also, non-local self-energy effects stemming from screened exchange interactions are non-negligible in this compound and alter quite considerably the unoccupied part of the $t_{2g}$ spectrum. We will come back to this point below, within the discussion of fully dynamical GW+DMFT calculations for SrVO$_3$ [9].

4 The Example of the Iron Pnictide BaFe$_2$As$_2$

As a non-trivial example for the generalised “LDA+$U(\omega)$+DMFT” approach, we review calculations on the iron pnictide compound BaFe$_2$As$_2$. This materials is the prototypical compound of the so-called “122-family” of iron pnictide superconductors. It exhibits superconductivity under
Figure 4: Frequency-dependent Hubbard interaction $U(\omega)$ for BaFe$_2$As$_2$: real and imaginary parts as well as the mode distribution function $Im[U(\omega)]/\omega^2$. From [5].

pressure [54, 55] or hole- as well as electron-doping [56, 57]. Many experimental probes including angle-resolved and angle-integrated photoemission spectroscopy [58, 59, 60, 61, 62, 63, 64], optics and transport, Raman and neutron scattering and nuclear magnetic resonance have been employed to characterise the electronic properties [65]. Experimental estimates of the (doping-dependent) mass enhancements vary substantially with doping; literature values range from about 1.4 [66] to 5, at least for the orbital pointing towards the As-sites [68]. The orbital character of the Fermi surface pockets are still subject to debate, but there seems to emerge a consensus about stronger correlation effects for holes than electrons.

The constrained RPA result for the average intra-orbital Coulomb repulsion $U$ of BaFe$_2$As$_2$ is shown in Fig. 4a. Here, $U(\omega)$ represents a partially screened Coulomb interaction for the Fe-$d$ states, which accounts for screening by all degrees of freedom except the Fe-$d$ states themselves. The real part ranges from the static value $U_0 \equiv ReU(\omega = 0) = 3.6$ eV to the bare interaction $V$ of about 20 eV at large $\omega$. In this case, screening does not arise from a single well-defined plasmon excitation. Instead, $Im[U(\omega)]$ is characterized by a broad structure beginning from a peak at $\sim 26$ eV and extending down to a few eV, implying that any plasmon excitations overlap strongly with the one-particle excitations. We note moreover that the term “plasmon” is used here a bit abusively for any bosonic excitation mode that screens the Coulomb interactions, regardless of the precise nature (plasmon, particle-hole excitations or any other many-body satellite feature).

In a standard DFT+DMFT calculation without dynamical effects, the relatively small value of
the interaction $U_0$ would result in a rather weakly correlated picture. This is demonstrated in Fig. 5, which presents the Fe-$d$ spectral function obtained by using the static $U_0$ in a standard DFT+DMFT calculation. Interaction effects lead to a moderate renormalization of the Fe-$d$ states, with a mass enhancement of 1.6. Comparison to the DFT density of states shows that the peaks at $-3$ eV and 1 eV are weakly renormalized band states. No Hubbard satellites or other correlation features appear, in agreement with previous studies [72, 26].

The function $\frac{\text{Im}U(\omega)}{\omega^2}$ is plotted as the red dash-dotted line in Fig. 4. Besides a first peak at 3.8 eV, which comes from the rapid decay of $\text{Im}U(\omega)$ at small frequencies, there are prominent peaks at 6.1 eV, 16 eV and 26 eV, as well as smaller features at 10 eV and 12 eV.

The sharp low-energy peak in the $d$-electron spectral function results in weak, but well-defined satellites, as discussed above. The inset of Fig. 5 shows the high energy tail of the occupied part of the spectrum, with arrows marking the most prominent satellites at $-6.1$ eV, $-12$ eV and $-16$ eV. The observation of satellites at $-6.5$ eV and $-12$ eV was emphasized in the photoemission study of Ding and collaborators [60]. While Ref. [27] confirms a hump in the $d$-electron spectral function around $-6.5$ eV, these authors suggest that the feature at $-12$ eV is an As-4s line. Our calculation suggests that a $d$-feature, originating from the structure in the frequency dependent interaction, is superimposed to the As-4s spectral contribution.

The $-16$ eV feature is probably not visible in experiments, because it overlaps with Ba-5p states, while a satellite which we predict at $-3.8$ eV is masked by structures arising from $p-d$
hybridization.

A detailed analysis of the many-body self-energy revealed a further interesting aspect of this compound, namely a pronounced “non-Fermi liquid” (incoherent) regime in the metallic phase near optimal doping, characterised by square-root behavior of the self-energy. We refer the interested reader to the original Ref. [5], where also implications for the doping and temperature dependence of the low energy electronic structure, in comparison with angle-resolved photoemission were discussed.

5 The combined “GW+DMFT” approach

The solution of the DMFT equations for a frequency-dependent (dynamical) Hubbard interaction, is also a key step in the combined “GW+DMFT” method, as proposed in [6]. Indeed, this scheme was proposed a few years ago, in order to avoid the ad hoc nature of the Hubbard parameter and the double counting inherent to conventional combinations of dynamical mean field theory with the LDA. Moreover, the theory provides momentum dependence to quantities (such as the self-energy) that are local within pure DMFT.

The starting point is Hedin’s GW approximation (GWA)[28, 23], in which the self-energy of a quantum many-body system is obtained from a convolution (or product) of the Green’s function $G$ with the screened Coulomb interaction $W = \epsilon^{-1}V$. The dielectric function $\epsilon$, which screens the bare Coulomb potential $V$, is – within a pure GW scheme – obtained from the random phase approximation. The GW+DMFT scheme, as proposed in [29], combines the first principles description of screening inherent in GW methods with the non-perturbative nature of DMFT, where local quantities such as the local Green’s function are calculated to all orders in the interaction from an effective reference system (“impurity model”)\(^1\). In DMFT, one imposes a self-consistency condition for the one-particle Green’s function, namely, that its on–site projection equals the impurity Green’s function. In GW+DMFT, the self-consistency requirement is generalized to encompass two-particle quantities as well, namely, the local projection of the screened interaction is required to equal the impurity screened interaction. This in principle promotes the Hubbard $U$ from an adjustable parameter in DMFT techniques to a self-consistent auxiliary function that incorporates long-range screening effects in an \textit{ab initio} fashion. Indeed, as already alluded to above and further elaborated upon in Sect. 7, not only higher energy degrees of freedom can be downfolded into an effective dynamical interaction, but one can also aim at incorporating non-local screening effects into an effective dynamical $U(\omega)$. The theory is then free of any Hubbard \textit{parameter}, and the interactions are directly determined from the full long-range Coulomb interactions in the continuum.

From a formal point of view, the GW+DMFT method, as introduced in [6], corresponds to a specific approximation to the correlation part of the free energy of a solid, expressed as a functional of the Green’s function $G$ and the screened Coulomb interaction $W$: the non-local part is taken to be the first order term in $W$, while the local part is calculated from a local impurity model as an \textit{ab initio} formalism. This leads to a set of self-consistent

\(^1\)The notion of locality refers to the use of a specific basis set of atom-centered orbitals, such as muffin-tin orbitals, or atom-centered Wannier functions.
equations for the Green’s function $G$, the screened Coulomb interaction $W$, the self-energy $\Sigma$ and the polarization $P$ [30, 31] (which are reviewed in Appendix B). Specifically, the self-energy is obtained as $\Sigma = \Sigma_{\text{local}} + \Sigma_{\text{non-local}}^{GW}$, where the local part $\Sigma_{\text{local}}$ is derived from the impurity model. In practice, however, the calculation of a self-energy for (rather delocalized) $s$- or $p$-orbitals has never been performed within DMFT, and it appears to be more physical to approximate this part also by a GW-like expression. For these reasons Ref. [9] proposed a practical scheme, in which only the local part of the self-energy of the “correlated” orbitals is calculated from the impurity model and all other local and non-local components are approximated by their first order expressions in $W$.

6 From “LDA+$U(\omega)$+DMFT” to “GW+DMFT”: the Example of SrVO$_3$

The very first dynamical (albeit not yet fully self-consistent) implementation of “GW+DMFT” was achieved in 2012, and applied to the ternary transition metal oxide SrVO$_3$ [9]. In this section, we review the results of these calculations, before entering more in detail into questions of the formalism and finally describing a – fully self-consistent – implementation in the single-orbital case and its application to surface systems (see Section 8).

Our target material, SrVO$_3$, has been thoroughly studied, both, experimentally and theoretically. It crystallizes in the cubic perovskite structure, splitting the V-d states into a threefold degenerate $t_{2g}$ manifold, filled with one electron per V, and an empty $e_g$ doublet. It has been characterized as a correlated metal with a quasiparticle renormalization of about 0.6 [32, 33, 34], and a photoemission (Hubbard-) satellite at around -1.6 eV binding energy [35]. Inverse photoemission has located the electron addition $d^1 \rightarrow d^2$ peak at an energy of about 2.7 eV [35].

Figure (6) summarizes the LDA electronic structure: the O-p states disperse between -2 and -7 eV, separated from the $t_{2g}$ states whose bandwidth extends from -1 eV to 1.5 eV. While the $t_{2g}$ and $e_g$ bands are well separated at every given k-point, the partial DOS slightly overlap, and the $e_g$ states display a pronounced peak at 2.3 eV. Finally, peaks stemming from the Sr-d states are located at 6.1 eV and 7.1 eV. We have superimposed to the LDA DOS the experimental PES and BIS curves taken from [35, 33]. The comparison reveals the main effects of electronic correlation in this material: as expected on quite general grounds, LDA locates the filled O-p states at too high and the empty Sr-d manifold at too low energies. The $t_{2g}$ manifold undergoes a strong quasi-particle renormalization and a concomitant shift of spectral weight to the lower Hubbard band, both of which are effects beyond the one-particle picture. The GW approximation (see the spectral function in Fig. (7)) increases the O-p to Sr-d distance, placing both manifolds at energies nearly in agreement with experiment $^2$. Most interestingly, however, a peak at 2.6 eV emerges from the d-manifold, which we find to be of $e_g$ character. Indeed, the GW approximation enhances the $t_{2g}$-$e_g$ splitting and places the maximum of the $e_g$ spectral weight at the location of the experimentally observed $d^1 \rightarrow d^2$ addition peak. The panel also displays the GW partial $t_{2g}$ contribution. These states show two interesting features: their width is narrowed from the LDA

$^2$The persisting slight underestimation is expected, since the polarization function that determines $W$ is calculated using the bare $t_{2g}$ bands, whose itinerant character is largely overestimated by the LDA.
Figure 6: Orbital-resolved LDA density of states, in comparison to experimental spectra from photoemission (PES)[35, 33] and inverse photoemission (BIS)[33]. From [9].

value by about 0.5 eV and satellite structures appear below and above the main quasiparticle peak at ±3 eV, as well as above 15eV. The origin of these features was analysed in [9], where it was argued that the peak at 15eV corresponds to the physical plasmon discussed in Section 2, while the lower energy satellites are spurious features due to the perturbative nature of the GW approximation.

In the implementation of Ref. [9] the GW+DMFT equations were solved self-consistently only at the DMFT level, that is for a fixed screened interaction (corresponding to the cRPA one) and fixed non-local GW self-energies. However, the fully dynamical interactions were retained, and the GW+DMFT equations solved within the BFA [4] reviewed above.

Figure 7: Spectral function from GW and GW+DMFT in comparison to experiments. t_{2g} orbital contribution is resolved. The Fermi energy is set to zero. From Ref. [9].

The results for the spectrum of SrVO₃ are plotted in Fig 7, which displays the local spectral function. In Fig 8 we compare the momentum resolved GW+DMFT spectral function to pure GW results as well as to the most recent angle resolved photoemission experiments (ARPES)[36]. The low-energy part of the spectral function is dominated by the t_{2g} contribution, which is profoundly modified within GW+DMFT compared to the pure GW spectrum. A renormalized
Figure 8: Momentum-resolved $t_{2g}$ spectral functions from GW and GW+DMFT, compared to experimental angle resolved photoemission spectra of Refs. [36] and [34]. From Ref. [9].

quasi-particle band disperses around the Fermi level, with e.g. at $\Gamma$ a sharp peak in the occupied part of the spectrum at about -0.5 eV binding energy, corresponding to a strong renormalisation of the LDA Kohn-Sham state that has, at this momentum, an energy of -1 eV. At the X-point, the three $t_{2g}$ bands are no longer degenerate, and surprisingly weakly renormalized xz/yz states are observed at 0.9 eV, while the yz band is located at nearly the same energy as at the $\Gamma$ point. This is in excellent agreement with experiments: In photoemission spectroscopy a renormalized quasi-particle band structure is dispersing between the $\Gamma$ and X points. At binding energies of -1.6 eV a weakly dispersive Hubbard band forms, whose intensity varies significantly as a function of momentum [34]. In the GW+DMFT spectral function, the Hubbard band is observed at about -1.6 eV, and its k-dependent intensity variation is indeed quite strong. As anticipated from the total spectral function discussed above, the GW spectrum also displays a satellite structure, which is however of plasmonic origin, arising from the structure in Im$W$.

This is a well-known failure of the GWA, which has been analyzed in detail in [37]: the simple form of the GW self-energy is not able to encode multiple satellites, and the single plasmon is then located at a too high energy. The lower Hubbard band is absent in GW, as expected.

As discussed above, the low-energy features in $W$ are absent from the dynamical Hubbard interaction entering the combined GW and dynamical calculation, consistently with the fact that the full GW+DMFT calculation does not show spurious features at 3 eV as does the GW.

The overall picture of the GW+DMFT spectra results in an occupied band structure that resembles closely the dynamical mean field picture (see e.g. [38]) though the lower Hubbard band is located slightly closer to the Fermi level, at about -1.6 eV, in agreement with experiments. This improvement is an effect of the relatively smaller zero-frequency $U$ value (3.6 eV) compared
to parameters commonly used in standard LDA+DMFT calculations (around 4 to 5 eV). In the latter, larger $U$ values are required to match the experimentally observed mass enhancements, thus necessarily placing the lower Hubbard band at energies slightly too far away from the Fermi level. The dynamical screening, included in our calculations, results in additional spectral weight transfers [4, 3], thus yielding at the same time a good description of mass enhancements and the Hubbard band.

In the unoccupied part of the spectrum non-local self-energy effects are larger. Interestingly, our total spectral function, right panel of Fig.7, does not display a clearly separated Hubbard band. The reason is visible from the k-resolved spectra: the upper Hubbard band is located at around 2 eV, as expected from the location of the lower Hubbard band and the fact that their separation is roughly given by the zero-frequency value of $U$ (=3.6 eV). The peak around 2.7 eV that appears in the inverse photoemission spectrum [35] – commonly interpreted as the upper Hubbard band of $t_{2g}$ character in the DMFT literature – arises from $e_g$ states located in this energy range. The non-local self-energy effects lead, in the unoccupied part of the spectrum, to overlapping features from different k-points and the impression of an overall smearing out of the total spectral function. From the k-resolved spectra, it is also clear that while non-local self-energy effects stemming from the GW part have little influence on the occupied part of the spectrum, they widen the bands in the unoccupied part substantially. A rough estimate of the various renormalisation effects on the overall bandwidth leads to a – at first sight – astonishing conclusion: effects of the dynamical tail of the Hubbard $U$ have been estimated to roughly lead to a band renormalisation of $Z_B \sim 0.7$ [3], and the renormalisation due to the static part $U(\omega = 0)$ still adds to this. Nevertheless, the final position of the empty quasi-particle bands after the GW+DMFT calculation nearly coincides with the initial LDA energies. This gives an order of magnitude for the substantial widening of the band induced by nonlocal effects. The picture is consistent with the observation of Refs. [39, 40], that a purely local GW calculation in fact leads to much stronger renormalisation effects than the full GW calculation, and the band structure within the latter results from subtle cancelations of band narrowing due to the local self-energy and widening due to its non-local parts. One could thus be tempted to conclude that renormalisation effects due to local dynamical interactions and widening due to non-local self-energies cancel, giving new justifications to combined LDA+++ schemes with static interactions. There are several reasons why this conclusion would be too quick: First, the widening effects rather selectively act on the unoccupied band structure, since the exchange-correlation potential of LDA is a much better approximation to the many-body self-energy for occupied states than for empty ones. Second, the renormalisation effect due to dynamical interactions goes hand in hand with a spectral weight loss at low-energies. These are barely observable in photoemission spectroscopy since spectra are generally not measured in absolute units, and even then would matrix element effects make a comparison of absolute normalisations intractable. Probes that can assess absolute units, such as optical spectroscopy, however, can be expected to be sensitive to such shifts of spectral weight.
In calculations of the effective local Coulomb interactions within the constrained random phase approximation the dynamical nature of the interaction is generally supposed to stem only from the downfolding of higher-energy degrees of freedom. It can then be directly assessed by a cRPA construction of a dynamical lattice model. The construction of the impurity model serves in this case merely as a tool to solve the dynamical lattice problem, with a fixed local dynamical interaction that is assumed to be the local part of the cRPA one. The GW+DMFT formalism, on the other hand, demonstrates that it is possible to adopt a more general point of view, that also incorporates the contribution of nonlocal interactions and nonlocal screening effects in a solid, giving rise to an additional frequency dependence of the effective local interaction. The Hubbard interaction $U$ in this case should no longer be interpreted as the local part of the physical Coulomb interaction of a downfolded model, but as an effective quantity that incorporates both, the effects of screening by downfolding and by representing a lattice model by a local model. This perspective goes beyond the cRPA view of the problem, but can be consistently formulated within extended DMFT or GW+DMFT. Indeed, when implemented in a fully self-consistent fashion, GW+DMFT provides a prescription on how to calculate both the one-particle part of the Hamiltonian and the effective Hubbard interactions of a correlated material from first principles. The idea is to calculate the nonlocal part not only of the self-energy but also of the polarization to lowest order in the screened Coulomb interaction. These non-local quantities are then combined with their local counterparts as calculated from a dynamical impurity model. One thus represents two physical quantities, namely, the local Greens function $G$ of the solid and the local part of the fully screened Coulomb interaction $W$ by a local model, defined by some effective Weiss field $G_0$ and the auxiliary Coulomb interaction $U(\omega)$. The latter is constructed such that the solution of the impurity model yields the local part of $W$. This is akin in spirit to other theories in solid state physics, where a physical quantity is represented by the self-consistent solution of an effective auxiliary model, famous examples being density functional theory or DMFT itself. In DFT, the physical density of a system is represented by an auxiliary system in an effective one-particle (Kohn-Sham) potential; in DMFT, the local lattice Greens function is constructed from an impurity model with an effective Weiss field or, equivalently, a local self-energy. The auxiliary quantities such as the Kohn-Sham potential of DFT or the impurity self-energy acquire the role of Lagrange multipliers fixing the density (in DFT) or the local Greens function (in DMFT) to their physical values. In extended DMFT, a nonlocal interaction in the original Hamiltonian gives rise to a dynamical impurity model representing the physical quantities of the model, and it is the polarisation that takes over the role of the corresponding Lagrange multiplier. This carries through to the combined GW+DMFT scheme where the non-local polarisation moreover adds to the frequency-dependence of the effective local interaction. The formalism of GW+DMFT, leading to a closed set of coupled equations for the one- and two-particle quantities $G$, $W$, the one- and two-body self-energies $\Sigma$ and $P$ and

\[ \text{See however the recent works aiming at the construction of an effective local Hubbard interaction to be used within the DMFT context from a cRPA scheme where } P_{\text{low}} \text{ is restricted to represent local screening only [119, 41].} \]
the auxiliary quantities $G_0$ and $U(\omega)$ is reviewd in Appendix B.

8 Exploring the Self-consistent “GW+DMFT” scheme: Systems of Adatoms on Semiconductor Surfaces

While the first dynamical implementation of GW+DMFT for a real material – the calculation for SrVO$_3$ reviewed above – was not yet fully self-consistent, the full self-consistent cycle has in the meanwhile been explored on the example of a different class of materials: Systems of adatoms on semiconducting surfaces, such as Si(111):X with X=Sn, C, Si, Pb, present the advantage that their electronic structure is determined by a low-energy Hilbert space spanned by a single narrow half-filled surface band, which lies – well-separated – in the gap of Si. Proposed early on [83] to be good candidates for observing low dimensional correlated physics, these systems have for a long time been considered to be realisations of the one-band Hubbard model on the triangular lattice, and have been the subject of a variety of experimental ([84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96]) and theoretical [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109] studies. Experimentally, the so-called $\alpha$-phases show a remarkable variety of interesting physics including commensurate charge density wave (CDW) states [87, 88, 90] and isostructural metal to insulator transitions (MIT)[93].

In Ref. [10], a low-energy effective Hamiltonians for describing the surface state was derived \textit{ab initio} from density functional theory and the constrained random phase approximation (cRPA) scheme [1] in the implementation of [2] (see also the extension to surface systems in [45]). The results for the interaction parameters are given in Table 1. The local interactions vary from 1eV to 1.4eV, depending on the adatom species. This is not only much larger than the bandwidths ($\sim 0.5$eV) confirming indeed the importance of correlation effects for these systems. It is also somewhat larger than what was previously assumed in model studies that treated the Hubbard interaction as adjustable parameter in order to fit experiments. Even more interestingly, however, long-range interactions are large, with the interaction between electrons in Wannier orbitals located on nearest neighbor atoms being about 50% of their onsite counterparts. The relevant low-energy Hamiltonian is then a single-orbital Hubbard Hamiltonian, extended to comprise the full tail of Coulomb interactions, where the value is reduced by screening but not the range. We stress that all parameters entering these Hamiltonians are calculated from first principles. Momentum-resolved spectral functions – to be compared to angle-resolved (inverse) photoemission spectra – and the two particle charge susceptibility were then calculated within the fully self-consistent GW+DMFT scheme applied to the lattice Hamiltonians. To this end, the GW+DMFT implementation of [7, 8] was generalised to the case of realistic Hamiltonians and long-range interactions using an Ewald technique.

Fig. 10 reproduces the momentum-resolved spectral functions for all compounds of the series: As expected from the large onsite interactions compared to the bandwidth we obtain insulating spectra for all four compounds. Interestingly, however, for the Pb compound two stable solutions – one metallic, one insulating – were found at the temperature of our study ($T = 116$K), stressing the proximity of this system to the metal-insulator transition (and charge order instabilities, as born out of an analysis of the charge-charge correlation function plotted in Fig. 11).
Table 1: Values of the static (partially) screened interactions \( U_0 = U(i\nu = 0) \) for on- and intersite nearest neighbor (nn) interaction parameters as determined from cRPA for use as bare interactions within the Hilbert space of the single surface band. Also reported are the values of the static self-consistent \( U_{\omega=0}^{GW+DMFT} \), incorporating non-local screening processes through the GW+DMFT “spatial downfolding” procedure. These values should be compared to the energy-scale of the Hubbard-gap in the spectral function, see text.

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>Si</th>
<th>Sn</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>bandwidth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_0 )</td>
<td>1.4</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>( U_1 )</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>( U_2 )</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>( : )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_n )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( U_{\omega=0}^{GW+DMFT} )</td>
<td>1.3</td>
<td>0.94</td>
<td>0.84</td>
<td>0.67(ins.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.54(met.)</td>
</tr>
</tbody>
</table>

In the context of the present review focusing on dynamical screening effects, the most interesting observation is however obtained from an analysis of the spectral functions: The insulating state is formed, as expected, from a splitting of the single half-filled surface band into upper and lower Hubbard bands. The energetic separation of the Hubbard bands is – in the single orbital case – a direct measure of the effective local Coulomb interaction. However, this energy is no longer set by the effective local Hubbard \( U_0 \) discussed above. Indeed, non-local screening processes – included within the GW+DMFT scheme through the self-consistency over two-particle quantities – lead to a reduction of the Hubbard interaction to a smaller value \( U_{\omega=0}^{GW+DMFT} \). The frequency-dependent \( U_{\omega}^{GW+DMFT} \) obtained within GW+DMFT for all four systems is plotted in Fig.11. The shape of this quantity is reminiscent of screened interactions, as calculated, e.g., within cRPA[1], where retardation effects result from downfolding of high-energy degrees of freedom. The GW+DMFT \( U_{\omega}^{GW+DMFT} \) can be viewed as an effective interaction, where the dynamical character results from downfolding non-local degrees of freedom into a local quantity. At large frequencies, screening is not efficient and, hence, \( U_{\omega}^{GW+DMFT} (\omega = \infty) = U_0 \). On the other hand, the static value \( U_{\omega}^{GW+DMFT} (\omega = 0) \) can be significantly reduced (up to nearly a factor of 2 for Si(111):Pb). The transition between the unscreened high frequency behavior and the static value takes place at an energy scale \( \omega_0 \) (plasmonic frequency) characteristic of the non-local charge fluctuations. The last line in Table 1 summarises the static values of the effective \( U_{\omega}^{GW+DMFT} (\omega = 0) \). These values are to be compared with the energetic separation of upper and lower Hubbard band, that is, the insulating gap. An estimate from the center of mass of the Hubbard bands yields values of 1.3eV for Si(111):C, 0.8eV Si(111):Si, 0.7eV Si(111):Sn, and 0.5eV for the insulating solution of Si(111):Pb. These values are – for the experimentally investigated systems, and in particular Si:Sn which is the most studied one – in excellent agreement with experimental results.
Interestingly, the behavior of the dynamical interactions is strikingly different between the different systems: $U(\omega)$ in Si(111):C [Si(111):Si] is [nearly] unaffected by non-local interactions and there is barely any screening. For Si(111):Sn and Si(111):Pb, however, the static values $U^{GW+DMFT}(\omega = 0)$ are substantially reduced compared to the onsite interaction: to 0.84eV for Si(111):Sn and to 0.67eV (0.54eV) for the insulating (metallic) solution for Si(111):Pb. Plasmonic resonances at energies between 0.6eV and 0.8eV stress the importance of non-local interactions/charge-fluctuations for these systems. It was proposed recently [115] that the effect of intersite interactions in extended Hubbard models for surface systems could simply be described by a reduction of the onsite interaction by the value of the nearest-neighbor one. While full GW+DMFT calculations for an extended Hubbard model with on- and intersite interactions only [7] indeed confirm the validity of this simple rule of thumb, the present more realistic series demonstrates that this procedure would lead to a large underestimation of the local interactions, when truly long-range interactions are present in the original Hamiltonian. In these cases, the efficiency of the non-local interaction in screening the local ones seems to depend quite significantly on the long-range interactions and/or the underlying one-particle band structure.

9 Perspectives

From a technical point of view, fully dynamical GW+DMFT calculations remain a challenge, even nowadays, and full self-consistency could at present be achieved only for the surface system.
Figure 10: Momentum-resolved spectral function at $T = 116$K of Si(111):X with X=Sn, Si, C, Pb obtained by analytical continuation of GW+DMFT imaginary time data. The Fermi energy is set to $\varepsilon_F = 0$ and indicated by the white dashed line. From Ref. [10] thanks to the simplifications that come along with a single orbital description. Self-consistent GW+DMFT calculations for realistic multi-orbital systems are thus an important goal for future work. In this context, the orbital-separated scheme proposed in the application to SrVO$_3$ reviewed above, where only a subset of low-energy states is treated within DMFT presents appealing features. For late transition metal oxides, and in particular charge-transfer systems, interactions between the correlated shell and the ligand electrons are likely to be important for an accurate estimation of the charge transfer energy and thus the whole electronic structure, and exploring the performance of the GW+DMFT scheme in this context is an interesting open problem. One could still expect a perturbative treatment of the intershell interactions to be sufficient, so that explicitly including an intershell self-energy $\Sigma_{pd}$ from GW in the orbital-separated scheme discussed above seems a promising way.

At the same time, it would be most interesting if it was possible to set up simpler schemes that could – at least approximately – reproduce the results of the full GW+DMFT calculations. Two routes could be pursued:
At the GW level, is it possible to avoid full GW calculations, and to base the combination with DMFT on a quasi-particle self-consistent GW (qsGW) scheme, for example the one by Kotani and Schilfgaarde [46]? 

At the DMFT level, is it possible to avoid the solution of a local problem with frequency-dependent interaction, replacing it by a problem with static $U$?

Recent work provided important insights into the first question [40, 124]. Ref. [40] argued that at least for certain classes of materials – the authors investigated iron pnictide compounds – nonlocal and dynamic contributions to the many-body self-energy are mostly separable, so that a purely local self-energy correction to a – in some sense – optimized one-particle band structure would be sufficient for an accurate description. The authors thus proposed a combination of quasi-particle self-consistent GW [46] with DMFT. Tests along these lines – although without attempting to avoid double counting of the local self-energies (see below) and based on a one-shot GW calculation – were performed in [124] for SrVO$_3$.

A subtle issue arises from the necessity of avoiding double counting of the local part of the GW self-energy, when “quasi-particilising” the GW calculations. Indeed, only the nonlocal part of the GW self-energy enters the combined GW+DMFT self-energy, which means that when a qsGW scheme is employed the local part $\Sigma_{GW\text{local}}(\omega) = \sum_{k'} GW$ has to be be explicitly substracted. This is not entirely trivial, since this term acquires both $k$- and state-dependence through the “quasi-particilisation” of its original frequency-dependence. Indeed, the correction acting on a quasi-particle state $\epsilon_{kn}$ reads $\Sigma_{GW\text{local}}(\omega = \epsilon_{kn})$. This term is responsible for the widening of the unoccupied bands discussed above (see Sect. 6) and can thus by no means be considered as negligible. The implementation of efficient yet double-counting-free qsGW+DMFT schemes remains therefore an important open challenge.

An optimistic answer can be given to the second question, concerning the need of using frequency-dependent interactions, at least when only the very low-energy electronic structure is of interest.
Indeed, the same trick as discussed above for reducing dynamical Hubbard models to static ones at the price of a renormalisation of the one-body part of the Hamiltonian can be applied also to a GW calculation corrected for its local term (“non-local GW”), “quasi-particledised” or not. On the other hand, the gain in computational cost obtained by this procedure is not tremendous in this case, since at least within the BFA the solution of a dynamical local model can anyhow be achieved at the cost of the solution of the corresponding static model.

Finally, the GW+DMFT construction also introduces some non-locality into the otherwise purely local self-energy (and polarisation) of (extended) DMFT. The first realistic GW+DMFT calculations for SrVO$_3$ have evidenced a crucial effect of this non-local self-energy on the unoccupied part of the Kohn-Sham band structure. A widening of the band due to the exchange interaction, which – for the empty states – is not well described by the Kohn-Sham potential leads to a substantial reinterpretation of the electronic structure. Nevertheless, this correction is essentially static and given by the Fock exchange, the truly frequency-dependent non-local part of the self-energy correction is small. A recent systematic study on the two-dimensional extended Hubbard model [7] has in fact shown that unless considering a system close to the charge-ordering transition, these non-local self-energy effects are tiny. To capture the strong k-dependence of the self-energies in the immediate vicinity of the Mott transition, experimentally observed e.g. in the form of highly k-dependent variations of the effective masses in doped Mott insulators, extending GW+DMFT to include not only fluctuations in the charge but also in the spin channel seems necessary. This is yet another promising topic for future work.

10 Conclusions

We have reviewed a series of recent papers dealing with dynamical screening of the Coulomb interactions in materials with correlated electrons. Quite generally, dynamically screened interactions arise as a “representative” within a restricted subspace of the full long-range Coulomb interactions in the original Hilbert space. In practice, it appears to be useful to disentangle two different mechanisms: (1) screening by high-energy degrees of freedom, when the original Coulomb Hamiltonian is downfolded to a low-energy effective Hamiltonian, (2) screening by non-local degrees of freedom, when a Hamiltonian with long-range interaction is backfolded into one with purely local interactions. The cRPA provides a simple and transparent description of the first mechanism, and the energetic separation of the different degrees of freedom can probably in many cases justify the neglect of vertex corrections. This is less obvious in the case of spatial downfolding, where both “backfolded” and retained degrees of freedom live on the same energy scales. The GW+DMFT scheme provides an elegant construction of effective local interactions, while including a local vertex in a non-perturbative manner.

This review has discussed the effects of dynamical interactions, specifically on the examples of the ternary transition metal compound SrVO$_3$ and the iron pnictide BaFe$_2$As$_2$. It was argued that the dynamical nature of the Hubbard interaction leads to many-body satellites in the spectral function (corresponding to plasmons, particle-hole excitations or more general bosonic excitations), and that the corresponding transfer of spectral weight leads to additional renormalisations in the low-energy electronic structure.
For SrVO$_3$, which was the first compound treated within GW+DMFT in a fully dynamical (albeit not fully selfconsistent) manner, the non-local self-energy effects introduced by this scheme moreover lead to profound modifications of the unoccupied spectra, as compared to simple LDA+DMFT calculations. In particular, the peak at 2.7 eV seen in inverse photoemission experiments was identified as an $e_g$ feature, while the energy scale of the upper Hubbard ($\sim 2$ eV) makes this feature hard to separate from the quite dispersive unoccupied $t_{2g}$ band states. These findings require a reinterpretation of early LDA+DMFT calculations using a $t_{2g}$-model for this compound, where the peak at 2.7 eV was commonly interpreted as an upper Hubbard band of $t_{2g}$ character. They do reconcile however DMFT-based electronic structure calculations with cluster model calculations for SrVO$_3$ [47]. The results are furthermore encouraging, since this first benchmark on SrVO$_3$ confirms the ability of the GW+DMFT approach to describe simultaneously Hubbard bands, higher energy satellite structures and corrected energy gaps, in an *ab initio* fashion. Quite generally, materials with a “double LDA failure”, an inappropriate description of correlated states, and deficiencies of LDA for the more itinerant states, such as an underestimated “pd-gap”, can be treated within this scheme.

Finally, we have reviewed an illustrative example of spatial downfolding within the self-consistent GW+DMFT scheme for the low-energy electronic structure of two-dimensional systems of atoms adsorbed on the Si(111) surface. Here, the frequency-dependence of the interactions stemming from downfolding higher energy degrees of freedom is weak (and thus neglected), but non-local screening leads to (system-dependent) reductions of the effective local interaction of up to nearly 50%.

11 Acknowledgments

This review summarises elements of the series of works [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] that I carried out together with different coauthors. I thank F. Aryasetiawan, T. Ayral, M. Casula, A. Georges, P. Hansmann, H. Jiang, A. Millis, T. Miyake, A. Rubtsov, J.M. Tomczak, L. Vaugier, P. Werner most warmly for the fruitful and enjoyable collaborations, all the way from the first proposal of the GW+DMFT scheme [6] until its first dynamical implementations for realistic materials [9, 10]. I furthermore acknowledge useful discussions with M. Ferrero, M. Imada, M.I. Katsnelson, A.I. Lichtenstein, O. Parcollet, L. Pourovskii, A. van Roekeghem, S. Sakai, and G. Sawatzky on various aspects related to the topics of this review.

This work was supported by the French ANR under projects SURMOTT and PNICTIDES, and IDRIS/GENCI under project 139313.

Appendix A: Effective low-energy Hamiltonian incorporating renormalisations due to dynamical screening – the electronic polaron

It has been discussed above how dynamical Hubbard interactions can be incorporated into the many-body description by LDA+DMFT, and that they are an integral part of the combined
GW+DMFT scheme. As seen on the above examples, even if the characteristic screening frequency (plasma frequency) is much larger than other relevant low-energy energy scales of the system (bandwidth and static Hubbard $U$), dynamical screening leads to substantial renormalisations of the low-energy electronic structure. One can thus ask the question if in this antiadiabatic limit it is possible to construct a low-energy effective Hamiltonian with static Hubbard interactions that reproduces the spectral properties of the dynamical problem in a low-energy window around the Fermi energy.

The answer is yes, and has been worked out explicitly in [3], where it was shown that in the antiadiabatic limit the dynamical effect of the interaction can be captured by a simple rescaling procedure of the one-particle part of the Hamiltonian. In the case of an Hamiltonian containing only low-energy degrees of freedom treated as “correlated”, the effect amounts to a simple scaling factor on the one-particle part of the Hamiltonian which is multiplied by (13). The derivation of this result is given in [3], and relies on a Lang-Firsov transformation to a polaronic Hamiltonian, where the coupling of electrons and plasmons (or bosonic particle-hole excitations) is eliminated at the price of passing to more complex (polaronic) degrees of freedom and then projecting onto the low-energy space containing no plasmon- or particle-hole contributions.

The final result is a downfolded many-body Hamiltonian that has the form of an extended Hubbard model, where the one-particle hoppings have been renormalised (for simplicity we only give here the single-orbital case, but the generalisation to multi-orbital systems, also in the presence of further uncorrelated orbitals was given in [3]):

$$\hat{H} = Z_B \sum_{ij\sigma} t_{ij} \left( \hat{c}^{\dagger}_{i\sigma} \hat{c}_{j\sigma} + \text{h.c.} \right) + H_{\text{int}}[U(\omega = 0)] \quad (14)$$

Here, $\hat{c}^{\dagger}_{i\sigma}$($\hat{c}_{i\sigma}$) creates (annihilates) an electron with spin $\sigma$ at lattice site $i$; $t_{ij}$ is the hopping amplitude between the Wannier orbitals on lattice sites $i$ and $j$, and $H_{\text{int}}[U(\omega = 0)]$ is the interaction term of (extended) Hubbard form (in the multi-orbital case possibly including Hund’s rule coupling etc.). The resulting prescription is thus rather simple:

- the one-particle part stems from a one-particle-downfolding procedure, supplemented by a rescaling by $Z_B$.

- the interaction is the zero-frequency limit of the dynamical Hubbard $U$ (as e.g. calculated within the cRPA). In the one-orbital case, this is simply the matrix element

$$U = \langle \phi_R \phi_R | W^{\text{test}}(0) | \phi_R \phi_R \rangle. \quad (15)$$

The problem with dynamical interactions can thus be mapped back again onto a static problem, but with renormalised fermions, corresponding to the mass enhancement due to the coupling to the bosonic degrees of freedom. In analogy to the electron-phonon coupling problem, the resulting fermionic charge carriers of enhanced mass – electrons dressed by screening bosons (plasmons or particle-hole excitations) – are called “electronic polarons” [28, 3].
Appendix B: Derivation of GW+DMFT from a Free Energy Functional

In this Appendix we review the derivation of the GW+DMFT scheme from a functional point of view. The discussion follows closely the original derivation in [6].

As noted in [48, 49], the free energy of a solid can be viewed as a functional $\Gamma[G, W]$ of the Green’s function $G$ and the screened Coulomb interaction $W$. The functional $\Gamma$ can trivially be split into a Hartree part $\Gamma_H$ and a many body correction $\Psi$, which contains all corrections beyond the Hartree approximation: $\Gamma = \Gamma_H + \Psi$. The Hartree part can be given in the form

$$\Gamma_H[G, W] = \text{Tr} \ln G - \text{Tr}[(G^{-1} - G^{-1}G)G] - \frac{1}{2} \text{Tr} \ln W + \frac{1}{2} \text{Tr}[(V_q^{-1} - W^{-1})W]$$

(16)

The $\Psi$-functional is the sum of all skeleton diagrams that are irreducible with respect to both, one-electron propagator and interaction lines. $\Psi[G, W]$ has the following properties:

$$\frac{\delta \Psi}{\delta G} = \Sigma^{xc}$$

$$\frac{\delta \Psi}{\delta W} = P.$$  

(17)

The $\Psi$ functional was first derived in [48]. A detailed discussion in the context of extended DMFT can be found in [49].

The GW approximation consists in retaining the first order term in the screened interaction $W$ only, thus approximating the $\Psi$-functional by

$$\Psi[G, W] = -\frac{1}{2} \text{Tr}(GWG).$$

(18)

We then find trivially

$$\Sigma = \frac{\delta \Psi}{\delta G} = -GW$$

(19)

$$P = \frac{\delta \Psi}{\delta W} = GG.$$  

(20)

Extended DMFT, on the other hand, would calculate all quantities derived from this function from a local impurity model, that is, one can formally write

$$\Psi = \Psi_{\text{imp}}[G^{RR}, W^{RR}].$$

(21)

In [6], an approximation to the $\Psi$ functional was constructed that corresponds to the combined GW+DMFT scheme. It approximates the $\Psi$ functional as a direct combination of local and non-local parts from GW and extended DMFT respectively:

$$\Psi = \Psi_{\text{GW non-loc}}^{\text{GW}}[G^{RR}, W^{RR}] + \Psi_{\text{imp}}[G^{RR}, W^{RR}]$$

(22)
More explicitly, the non-local part of the GW+DMFT $\Psi$-functional is given by

$$
\Psi_{GW}^{non-loc}[G^{RR'}, W^{RR'}] = \Psi_{GW}[G^{RR'}, W^{RR'}] - \Psi_{GW}^{loc}[G^{RR'}, W^{RR'}] \tag{23}
$$

while the local part is taken to be an impurity model $\Psi$ functional. Following (extended) DMFT, this onsite part of the functional is generated from a local quantum impurity problem (defined on a single atomic site). The expression for its free energy functional $\Gamma_{\text{imp}}[G_{\text{imp}}, W_{\text{imp}}]$ is analogous to (16) with $G$ replacing $G_H$ and $U$ replacing $V$:

$$
\Gamma_{\text{imp}}[G_{\text{imp}}, W_{\text{imp}}] = \operatorname{Tr} \ln G_{\text{imp}} - \frac{1}{2} \operatorname{Tr} \left( (G - 1 - G_{\text{imp}}^{-1}) G_{\text{imp}} \right) - \frac{1}{2} \operatorname{Tr} \left( (U - 1 - W_{\text{imp}}^{-1}) W_{\text{imp}} \right) + \Psi_{\text{imp}}[G_{\text{imp}}, W_{\text{imp}}] \tag{24}
$$

The impurity quantities $G_{\text{imp}}, W_{\text{imp}}$ can thus be calculated from the effective action:

$$
S = \int d\tau d\tau' \left[ - \sum L c^\dagger_{L} G_{LL'}^{-1}(\tau - \tau') c_{L'}(\tau') \right] + \frac{1}{2} \sum_1^L c^\dagger_{L_1} c_{L_2} U_{L_1 L_2 L_3 L_4}(\tau - \tau') : c^\dagger_{L_3} c_{L_4} : \right] \tag{25}
$$

where the sums run over all orbital indices $L$. In this expression, $c^\dagger_L$ is a creation operator associated with orbital $L$ on a given sphere, and the double dots denote normal ordering (taking care of Hartree terms).

The construction (22) of the $\Psi$-functional is the only ad hoc assumption in the GW+DMFT approach. The explicit form of the GW+DMFT equations follows then directly from the functional relations between the free energy, the Green’s function, the screened Coulomb interaction etc. Taking derivatives of (22) as in (17) it is seen that the complete self-energy and polarization operators read:

$$
\Sigma_{xc}(k, i\omega_n)_{LL'} = \Sigma_{GW}^{xc}(k, i\omega_n)_{LL'} - \sum_k \Sigma_{xc}^{GW}(k, i\omega_n)_{LL'} + [\Sigma_{xc}^{imp}(i\omega_n)]_{LL'} \tag{26}
$$

$$
P(q, i\nu_n)_{\alpha\beta} = P_{GW}(q, i\nu_n)_{\alpha\beta} - \sum_q P_{GW}(q, i\nu_n)_{\alpha\beta} + P_{\text{imp}}(i\nu_n)_{\alpha\beta} \tag{27}
$$

The meaning of (26) is transparent: the off-site part of the self-energy is taken from the GW approximation, whereas the onsite part is calculated to all orders from the dynamical impurity model. This treatment thus goes beyond usual E-DMFT, where the lattice self-energy and polarization are just taken to be their impurity counterparts. The second term in (26) substracts the onsite component of the GW self-energy thus avoiding double counting. At self-consistency this term can be rewritten as:

$$
\sum_k \Sigma_{GW}^{xc}(\tau)_{LL'} = - \sum_{L_1 L_1'} W_{LL_1 L_1'}^{imp}(\tau) G_{L_1 L_1'}^{\tau} \tag{28}
$$

so that it precisely substracts the contribution of the GW diagram to the impurity self-energy. Similar considerations apply to the polarization operator.

We now outline the iterative loop which determines $G$ and $U$ self-consistently (and, eventually, the full self-energy and polarization operator):
• The impurity problem (25) is solved, for a given choice of $G_{LL'}$ and $U_{\alpha\beta}$: the "impurity" Green’s function

$$G_{\text{imp}}^{LL'} \equiv -\langle T_\tau c_L(\tau)c_L^{\dagger}(\tau') \rangle_S$$

(29)

is calculated, together with the impurity self-energy

$$\Sigma_{\text{imp}}^{xc} \equiv \frac{\delta \Psi_{\text{imp}}}{\delta G_{\text{imp}}} = G - 1 - \frac{1}{G_{\text{imp}}}.$$  

(30)

The two-particle correlation function

$$\chi_{L_1L_2L_3L_4} = \langle \langle c_L^{\dagger}L_1(\tau)c_L^{\dagger}L_2(\tau) :: c_L^{\dagger}L_3(\tau')c_L^{\dagger}L_4(\tau') \rangle_S \rangle$$

(31)

must also be evaluated.

• The impurity effective interaction is constructed as follows:

$$W_{\text{imp}}^{\alpha\beta} = U_{\alpha\beta} - \sum_{L_1,..,L_4} \sum_{\gamma\delta} U_{\alpha\gamma} O_{\beta}^{\gamma} L_1 L_2 \chi_{L_1L_2L_3L_4} [O_{\delta}^{\gamma} L_3 L_4]^* U_{\delta\beta}$$

(32)

where $O_{\alpha}^{\gamma} L_1 L_2 \equiv \langle \phi_{L_1}^{\gamma} \phi_{L_2}^{\alpha} | B^\alpha \rangle$ is the overlap matrix between two-particle states and products of one-particle basis functions. The polarization operator of the impurity problem is then obtained as:

$$P_{\text{imp}} = -2 \frac{\delta \Psi_{\text{imp}}}{\delta W_{\text{imp}}} = U - 1 - W_{\text{imp}}^{-1}$$

(33)

where all matrix inversions are performed in the two-particle basis $B^\alpha$ (see the discussion in [31, 30]).

• From Eqs. (26) and (27) the full $k$-dependent Green’s function $G(k, i\omega_n)$ and effective interaction $W(q, i\nu_n)$ can be constructed. The self-consistency condition is obtained, as in the usual DMFT context, by requiring that the onsite components of these quantities coincide with $G_{\text{imp}}$ and $W_{\text{imp}}$. In practice, this is done by computing the onsite quantities

$$G_{\text{loc}}(i\omega_n) = \sum_k [G_H^{-1}(k, i\omega_n) - \Sigma^{xc}(k, i\omega_n)]^{-1}$$

(34)

$$W_{\text{loc}}(i\nu_n) = \sum_q [V_{q}^{-1} - P(q, i\nu_n)]^{-1}$$

(35)

and using them to update the Weiss dynamical mean field $G$ and the impurity model interaction $U$ according to:

$$G^{-1} = G_{\text{loc}}^{-1} + \Sigma_{\text{imp}}$$

(36)

$$U^{-1} = W_{\text{loc}}^{-1} + P_{\text{imp}}$$

(37)

The set of equations (28) to (36) is iterated until self-consistency.

References


28


Ding, H., et al., Observation of Fermi-surface-dependent nodeless superconducting gaps in Ba0.6K0.4Fe2As2. *Europhys. Lett.* **83**, 47001 (2008).


[120] L. Vaugier, H. Jiang and S. Biermann. (to be published)
[123] The structure with silicon as an adatom does not exist. It is calculated/presented as a hypothetical structure to complete the series - the pure silicon 111 surface is not stable in the $\sqrt{3} \times \sqrt{3}$ phase, but reorders to a very large $7 \times 7$ unit cell.