Topical Review

Dynamical screening in correlated electron systems—from lattice models to realistic materials

Philipp Werner\textsuperscript{1} and Michele Casula\textsuperscript{2}

\textsuperscript{1} Department of Physics, University of Fribourg, Chemin du Musée 3, 1700 Fribourg, Switzerland
\textsuperscript{2} Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Sorbonne Universités, Université Pierre et Marie Curie, CNRS, IRD, MNHN, 4 place Jussieu, 75252 Paris, France

E-mail: werner@phys.unifr.ch

Received 28 January 2016, revised 30 May 2016
Accepted for publication 3 June 2016
Published 20 July 2016

Abstract

Recent progress in treating the dynamical nature of the screened Coulomb interaction in strongly correlated lattice models and materials is reviewed with a focus on computational schemes based on the dynamical mean field approximation. We discuss approximate and exact methods for the solution of impurity models with retarded interactions, and explain how these models appear as auxiliary problems in various extensions of the dynamical mean field formalism. The current state of the field is illustrated with results from recent applications of these schemes to $U$-$V$ Hubbard models and correlated materials.

Keywords: dynamical mean field theory, dynamical screening, GW+DMFT, ab initio simulations, strongly correlated systems, impurity solvers

(Some figures may appear in colour only in the online journal)

1. Introduction

Screening describes the way the electrons rearrange themselves to reduce the inter-particle interaction and total energy of the system, while correlation describes how the electrons move collectively in the system. Both concepts are closely related and connected. In Hartree–Fock theory, the electrons are not correlated except for like-spin particles, which are constrained by the Pauli exclusion principle. A consistent framework to improve upon this theory was proposed by Hedin in his seminal 1965 paper [1]. He expanded the self-energy up to the first-order in the dynamically screened interaction $W$, and thereby laid the foundation of the GW theory [2–4]. The screened interaction $W$ is the bare interaction reduced by the dielectric function, determined by the charge–charge response function, which in turn is sensitive to electronic correlations. Screening is thus a dynamical process, as it depends on how the electrons respond to a given perturbation. It is obviously sensitive to the nature of the underlying electronic states. One of the fingerprints of screening are the plasmon satellites, collective excitations associated with long-range charge fluctuations, which are seen in electron spectroscopy (such as, for instance, photoelectron spectroscopy (PES) [5] and electron energy loss spectroscopy (EELS) [6]).

In correlated electron systems, where local atomic-like interactions are dominant, screening is also crucial to set the actual value of the Coulomb repulsion, and thus to determine the level of correlation in the system. Describing within the same framework both the long-range nature of screening and its role in tuning the local repulsion $U$ has been a long-standing theoretical challenge. The GW approximation, while neglecting vertex corrections, takes into account reasonably well dynamical long-range screening effects [7, 8], but it usually fails for large and local Coulomb repulsions, which are better described within a Hubbard model. This pushed Hedin to write in his review [3]: 'Clearly, the GW approximation, describing
long-range charge fluctuations, and a Hubbard model focusing on local on-site correlations, are two extremes’.

These two extremes can now be merged into a coherent and unified picture, thanks to recent progress in Green’s function embedding schemes, which have been developed as extended dynamical mean field theories. In dynamical mean field theory (DMFT) [9–11] the correspondence between the local lattice Green’s function $G_{\text{loc}}$ and the solution of an auxiliary Anderson impurity problem is realized through the dynamical Weiss field $G_{\text{w}}$, which mimics the effect of the lattice environment on the impurity site. In extended dynamical mean field theory (EDMFT) [12–14], the embedding is extended to the local screened interaction $W_{\text{loc}}$, which matches the impurity screened interaction and determines the dynamical bosonic field $U$, which represents the effect of nonlocal interactions on the auxiliary Anderson problem. Thus, the EDMFT embedding procedure maps nonlocal interaction effects onto local dynamical screening, and allows to treat these local interactions in a non-perturbative way. Indeed, since the advent of continuous time quantum Monte Carlo (CTQMC) algorithms [15], solving the Anderson impurity problem with retarded—i.e. frequency dependent—interactions has become feasible. This has been a major advance in the field. However, in the EDMFT framework the resulting self-energy is still local.

A further step forward is represented by the GW+DMFT theory, proposed in 2003 by Biermann and coworkers [16], where the embedding is performed at the Green’s function level between the local self-energy coming from the EDMFT solution of the impurity problem and the nonlocal self-energy, taken from GW. Therefore, nonlocality is included at the GW level, and the method can be applied to study first-principles Hamiltonians. Different schemes along these lines have become popular also in other contexts, such as in quantum chemistry [17–19], where the quantum system is divided into two parts, one of which is treated at the perturbative level, e.g. by the self-consistent second-order Green’s function method (GF2), while the other is solved at a higher level by the configuration interaction (CI) method [20, 21]. However, what distinguishes GW+DMFT from other approaches is the double embedding in both the Green’s function $G$ and the dynamically screened interaction $W$. This has many advantages, as we will see in this review; one of the most important is that it promotes the Hubbard $U$ parameter to the Weiss field $U$, which is determined self-consistently. Therefore, the underlying Hubbard Hamiltonian with retarded interactions is no longer a model, but is an auxiliary system, making the GW+DMFT scheme a truly $ab$ initio approach. At the same time, nonlocality is kept in both the self-energy $\Sigma$ and the electronic polarization $P$, which are related to $G$ and $W$, respectively. The GW+DMFT self-consistency in $G_{\text{loc}}$ and $W_{\text{loc}}$ closes the gap between the two extremes quoted by Hedin.

The general framework sketched above is conceptually appealing but still far from being a black-box machinery capable of solving generic correlated electron systems and first-principles Hamiltonians. However, in the last few years, significant progress has been made toward the final goal of a self-consistent description of dynamical screening effects in strongly correlated materials. In this review, we are going to introduce the theoretical and numerical tools which made this progress possible, such as, for instance, the CTQMC algorithm for frequency dependent interactions. Moreover, we are going to present the latest applications, both at the model level and from first principles. Model applications are useful to understand the limitations of the approximations underlying the EDMFT or GW+DMFT approaches, while the latest applications to real materials are extremely interesting, as they suggest promising routes to approximate GW+DMFT schemes, which are computationally more affordable, and can be applied to a wider class of $ab$ initio systems.

In our review we will not only explain the general theory, but also focus on practical aspects, with a twofold purpose: first, making tighter theoretical connections between various methods which deal with dynamical screening in $ab$ initio correlated electron systems. Second, presenting methodological details, which help the reader implement the methods in a computer program. Throughout the review, we will use $\text{eV}$ as energy units in the $ab$ initio applications, while the energy units will be set to the bandwidth, or some related hopping scale, in model applications.

The general organization of this review is as follows. In section 2 we explain how the screening is modeled from first principles in strongly correlated materials. We show that the retarded interactions emerge naturally from a multi-scale approach, where the coupling parameters are dynamically screened by higher-energy degrees of freedom, which are traced out in the derivation of the low-energy Hamiltonian.

In section 3, we introduce the main theoretical framework used throughout this review (section 3.1), namely the DMFT, employed to solve Hubbard-like low-energy Hamiltonians with retarded local interactions $U$. We describe several solvers capable of tackling the Anderson impurity problem with frequency dependent $U$, with a particular emphasis on the CTQMC method (section 3.2.3), which is formally exact and efficient in the case of density–density interactions. Different approximations are proposed for solving the impurity problem, such as the dynamic atomic–limit approximation (DALA) in section 3.2.1 and the Lang–Firsov approach in section 3.2.2. These methods are explained first for the solution of the Holstein–Hubbard model (section 3.2), where the dynamic nature of the local interaction $U$ derives from bosons (plasmons), locally coupled to the correlated electrons. This formalism can be generalized to deal with arbitrary $ab$ initio $U(\omega)$ (section 3.3) and with multi-band/multi-orbital systems typical of realistic materials (section 3.5), thanks to the mapping to a continuum of bosonic modes. The physics behind the generalized Holstein–Hubbard model can be made more transparent by the derivation of an effective static model, with renormalized hoppings and interactions due to screening effects, detailed in section 3.4.

With the tools of section 3, one would like to go beyond the local interaction picture, which may be too rough for realistic materials. Dealing with nonlocal interactions is a prerequisite for treating long-range screening effects, which lead to collective phenomena in solids. In section 4, we present different strategies for treating long-range interactions in strongly correlated systems, for both extended $U$-$V$ Hubbard models.
Computational methods play an important role in studying strongly correlated systems. Many interesting materials, such as the high-$T_c$ cuprates, have been studied in great detail using a broad range of experimental probes, which results in a detailed quantitative knowledge of the correlated electronic structure. A good theoretical description should be consistent with these experimental findings, and provide insights into the underlying physical mechanism. Ideally, the accuracy of the computational approach will enable quantitative predictions of material properties. However, even the simplest model used to investigate the physics of the cuprates, the two-dimensional (2D) single-band Hubbard model, does not admit an exact or at least numerically accurate solution in the most interesting parameter regimes, which are relevant for the experimental situations. For instance, the solution of this model near the Mott transition can be obtained in an approximate yet accurate way only if the most advanced numerical techniques are employed.

Moreover, the complexity and richness of strongly correlated materials puts limitations on the predictive power of simple models. The link between the model parameters and the experimental conditions, such as doping, pressure, chemical substitution, and temperature, may be difficult to establish. Yet, the validity of a theory is based on its ability to explain unambiguously and quantitatively a variety of situations. Therefore, a systematic procedure for deriving the model parameters is essential. Numerical approaches play a crucial role in the calculation of model parameters, which are seamlessly connected to the physics of actual compounds.

The idea is to use a multi-scale ab initio scheme, which is able to treat and predict material properties at a level of accuracy depending on the target energy. The common scheme is to first compute the full band structure by density functional theory (DFT) [35–37], usually used in the local density approximation (LDA), which takes into account electrons ranging from the deepest core levels to the virtual empty states, and covers an energy range of several tens or hundreds of eV. As it is well known, DFT is not the most accurate theory for strongly correlated materials, but it is able to reproduce general trends starting from the first-principle Hamiltonian, including several degrees of freedom. Its failure is usually related to the description of low-energy states, i.e., the ones near the Fermi level. In the case of strong correlations, they are poorly described by most of the functionals. At the same time, these states are primarily responsible for a material’s macroscopic properties. Thus, we would like to improve upon the DFT description, by applying more accurate, but computationally heavier, many-body methods to this low-energy sector. In order to do this, we need to derive a low-energy model by integrating out the high-energy degrees of freedom from the DFT solution. This intermediate, crucial step, is called downfolding [38–40]. The last step is the solution of the resulting model using advanced and hopefully accurate many-body techniques.

In order to downfold the DFT Hamiltonian, one needs to find a basis set which spans the low-energy space. Usually, one takes the maximally localized Wannier functions, $\phi_{mR}$ for the $m$th orbital centered in the unit cell $\mathbf{R}$, which give a reliable
representation of the correlated orbitals and their low-energy bands. Then, one projects the \textit{ab initio} Hamiltonian onto the basis set elements. The Coulomb matrix elements, which define the bare $U$, are therefore

$$
t_{mn}^{\text{bare}}(R) = \langle \phi_{m\alpha} | \phi_{n\beta} | \nu v_{\alpha\beta} \rangle = e^2 \int \mathcal{D}r dr' \frac{|\phi_{m\alpha}(r)|^2 |\phi_{n\beta}(r')|^2}{|r-r'|},
$$

where $v$ is the Coulomb potential. Without loss of generality, we restrict our discussion here to interactions of density–density form.

The above expression neglects the fundamental effect of screening from high-energy electrons on the low-energy manifold. In the random phase approximation (RPA) framework, the fully screened interaction is given by $W = v(1 - \nu P)$, where $P$ is the polarization function, defined as $P(r, r', t) = -2i \mathcal{G}_0(r, r', t) G_0(r, r, -t)$. In the latter definition, the factor of 2 comes from the spin summation in the spin-degenerate case, and $G_0$ is the zero temperature DFT Green’s function, which reads:

$$
\mathcal{G}_0(r, r', t) = \left\{ \begin{array}{ll}
\sum_l \psi_l(r) \psi_{l'}^*(r') e^{-i\nu t}, & \text{if } t < 0 \\
-i \sum_l \psi_l(r) \psi_{l'}^*(r') e^{+i\nu t}, & \text{if } t > 0
\end{array} \right.
$$

where the $\{\psi_l, \epsilon_l\}$ are the one-particle Bloch eigenfunctions and eigenvalues corresponding to the system’s band structure. In the frequency domain, and for temperature $T = 0$, the above expression for the polarization becomes

$$
P(r, r', \omega) = 2 \sum_{ij} \langle \psi_i(r) | \psi_j^*(r') \rangle \langle \psi_i^*(r') | \psi_j(r) \rangle \frac{1}{\omega - \epsilon_j + i0^+ - \frac{1}{\omega + \epsilon_j - \epsilon_i - i0^+}}.
$$

One can then separate $P$ into the two contributions $P = P^H + P^L$, with $P^L$ defined by polarization channels fully contained in the low-energy ($L$) window around the Fermi level, and $P^H$ its complement. Note that $P^H$ includes not only particle–hole excitations within the high-energy ($H$) sector, but also those connecting the high- and the low-energy sector. A schematic illustration of the division between $P^H$ and $P^L$ is presented in figure 1.

In the constrained RPA (cRPA) theory \cite{43, 44}, the partially screened interaction is given by $W^c = v(1 - \nu P^c)$. The physical interpretation is transparent: $W^c$ is the bare interaction screened by scattering processes leaking from the low-energy sector. If $W^c$ is further screened by the polarization $P^c$ of the low-energy sector, the fully screened interaction is recovered: $W = W^c/(1 - \nu P^c)$.

$W^c$ yields the matrix elements of the partially screened $U$:

$$
U_{mn}(r, r', \omega) = \left\langle \phi_{m\alpha} | \phi_{n\beta} \right| \frac{v}{1 - \nu P^c(\omega)} \left| \phi_{m\alpha} \phi_{n\beta} \right\rangle.
$$

The above $U$ defines the electron–electron interaction felt by the low-energy electrons in the \textit{downfolded} model, i.e. after the high-energy degrees of freedom have been integrated out. The important thing to note here is that the partially screened $U$ becomes \textit{frequency dependent}, a direct consequence of the frequency dependence of $P^c$ \cite{45}. In the time-domain, this implies that the effective interaction becomes retarded. This \textit{dynamical screening} is the main focus of our review. We will show that it can be explicitly taken into account in the solution of the low-energy models, and that it will affect the results in nontrivial ways.

It is interesting to remark here that the frequency dependence is particularly strong for monopole–monopole interactions (direct terms of the RPA expansion), while the screening is less effective for multipolar charge distributions (‘exchange’ terms in RPA), like the ones related to the estimate of the couplings $J$.
It turns out that $U_{mn}$ is typically reduced by an order of magnitude with respect to $U_{bare}$, while $J_{nm}$ is almost unaffected by screening (change of typically less than 20% [46]). As an illustration, we plot the cRPA results for the 3-band model of SrVO$_3$ in figure 2.

We would like to stress that the cRPA values of $U$ and $J$ are not adjusted by fitting schemes or empirical arguments. They are evaluated from first-principles [48–55], according to a rigorous procedure, which has been recently extended also to the more involved case of entangled bands [56, 57], where the separation between the low- and high-energy sectors is not sharp, due to the hybridization between correlated orbitals and more extended ones.

Although the cRPA scheme seems plausible, it has some limitations, which originate from two sources: first, it is built on the RPA method which is not exact; second, the underlying DFT band structure is not exact either. While the first limitation can lead to underscreening or overscreening depending on the particular case, the second one is most severe in materials which are close to a Mott transition (which leads to strong modifications in the starting DFT band structure, and rearrangements of the screening bands)\(^4\), or in the case where the ligands are not included in the model (as usually the $p$–$d$ hybridization is not well reproduced by DFT). A way to overcome the latter limitation is to use a better \textit{ab initio} scheme to solve the full Hamiltonian, e.g. by replacing DFT by GW, while the former one can be reduced by improving upon the RPA theory. A possible approach is the recently developed constrained functional renormalization group (cRFG) method [58]. Testing the reliability of the cRPA method is a subject of current research [59].

Besides the cRPA method, other methods have been proposed to evaluate the local electron–electron couplings from \textit{ab initio} calculations, such as the linear response approach [60, 61] and the constrained DFT [62, 63]. They however neglect the explicit frequency dependence of the Hubbard parameters. A slight variation of the cRPA method, which retains its full frequency dependence but is based on orbitals rather than bands, can be found in [64]. Along the same lines, i.e. working in the maximally localized orbitals space, Nomura \textit{et al} proposed to evaluate the partially screened $U$ by undressing the fully screened coupling via the local polarization (i.e. by also including nonlocal processes in the low-energy manifold as active screening channels) [65]. This should give a $U$ particularly suited for embedding theories such as DMFT, where only local correlations are taken into account. However, in our review, we will consider an alternative strategy, which is to include nonlocal processes in an explicit way, using extended-DMFT frameworks.

While the two-body part of the \textit{downfolded} Hamiltonian is provided by the matrix elements of the partially screened interaction, equations (4) and (5), the one-body part is given by

$$l_{nm}(\mathbf{R}) = \langle \phi_{n0}|H_{\text{DFT}}^{}|\phi_{m\mathbf{R}}\rangle,$$

where $H_{\text{DFT}}^{}$ is the DFT Hamiltonian determined in the first step. The $l_{nm}(\mathbf{R})$ already include some correlation effects, as described by $H_{\text{DFT}}^{}$. Therefore, one needs to correct the one-body part, i.e. subtract the low-energy correlation effects from the density functional, to avoid their \textit{double counting} (DC). The so-called ‘double-counting correction’ is common to all methods which augment the DFT band structure by explicit electron–electron interaction terms, such as DFT+$U$ [66, 67] or DFT+DMFT [68]. This is the weakest point of DFT-based downfolding procedure, as an exact expression for the DC term is hard to derive. This is because the exchange correlation potential $V_{xc}$ in $H_{\text{DFT}}^{}$ is a nonlinear functional of the total density, which makes it impossible to separate it into low- and high-energy contributions.

Several forms of approximate DC corrections have been proposed in the literature. One of the most successful for strongly correlated systems has proven to be the fully localized limit (FLL) form [69], which reads

$$\Sigma_{\text{DC}}^{\text{FLL}}(n_d) = \delta_{nm} \left( U \delta_{n_d - \frac{1}{2}} - J \delta_{n_d - \frac{1}{2}} \right).$$

\(^4\) Because the low-energy polarization $P^L$ is removed in cRPA, the changes in $P^H$ resulting from a Mott transition are irrelevant. However, the rest polarization $P^H$ has contributions from transitions between low-energy and high-energy bands (processes (2) and (3) in figure 1), which are not correctly estimated by cRPA in the case of a splitting or renormalization of the low-energy bands.

![Figure 2. cRPA $U(\omega) \equiv U_{mn}(\omega)$ and $J(\omega) \equiv J_{nm}(\omega)$ for the 3-band model of SrVO$_3$. Note the different y-axis scales. (Reproduced with permission from [47].)](image-url)
where \( n_d^0 \) is the spin-resolved occupancy of the correlated orbitals, and \( n_d = n_d^+ + n_d^- \). In the above equation, \( U = \frac{1}{N} \sum_{n,m} U_{nm}(0,0) \) is the average local static Coulomb interaction (\( N \) is the number of correlated bands), while the Hund’s coupling \( J = \frac{1}{N(N-1)} \sum_{n,m \neq 0} U_{nm}(0,0) - J_{\text{ins}}(0,0) \). The form in equation (7) has been used in the ab initio applications presented in this review, unless otherwise stated.

Very recently, Haule proposed a new DC scheme, in which a \( V_{xc}^{\text{imp}} \) potential for the impurity system is constructed in terms of the local Green’s functions and a local energy functional built on screened interactions [70]. He found that this procedure leads to double counting shifts in good agreement with an approximate DC correction obtained from equation (7), by replacing \( n \) with its closest integer value \( n^0 = [n_d] \) (nominal DFT occupancy [71]), i.e.

\[
\delta_{\text{nominal DC}} = \delta_{\text{ren}} \left( U \left( n^0 - \frac{1}{2} \right) - \frac{J}{2} (n^0 - 1) \right).
\]

Another interesting solution of the double counting problem has been proposed recently in [34]. It is based on the replacement of the exchange-correlation functional \( V_{xc} \) by the perturbative expansion of the potential to first order in the fully screened interaction \( W \), as it is done in the GW approach [1]. In contrast to \( V_{xc} \) in the GW framework one can easily separate the low-energy contributions (\( \Sigma^L = G^L W \)) from the high-energy ones (\( \Sigma^H = G^H W \)), where \( G^H \) (\( G^L \)) is the unperturbed Green’s function living in the high- (low-) energy subspace. The off-diagonal matrix elements connecting the two subspaces are disregarded in the quasiparticle approximation. This leaves us with the operator \( H^{\text{DFT}}_{\text{imp}} - V_{xc} + \Sigma^H \), which is well defined in the high-energy space, although being frequency dependent. To get rid of the frequency dependence, the dynamic part included in the \( \Sigma^H \) self-energy is treated as a first-order variation around the Fermi energy, and condensed into a bandwidth renormalization factor \( Z^H = (1 - \frac{\partial \Sigma^H}{\partial \omega})^{-1} \). The corresponding DC-free hoppings read

\[
I_{\text{ren}}^{\text{DC}}(\mathbf{R}) = \langle \phi_{\mathbf{R}}^{\mathbf{F}} | \phi_{\mathbf{R}}^{\mathbf{D}} \rangle \langle \phi_{\mathbf{R}}^{\mathbf{D}} | \phi_{\mathbf{R}}^{\mathbf{F}} \rangle,
\]

with \( H^{\text{eff}} = Z^H (H^{\text{DFT}}_{\text{imp}} - V_{xc} + \text{Re} \Sigma^H(0)) \). It turns out that, due to the analytical properties of the self-energy, the effective band structure in equation (9) has a larger bandwidth than the one in equation (6). Therefore, a proper treatment of the double counting correction leads to a much more complex rearrangement of the low-energy bands than the simple rigid (orbital independent) shift of the correlated manifold implied by equation (7). In most of the applications presented in this review, we are however going to use the latter correction, even though it is less accurate. It is simpler to compute, because it does not require a single-shot GW calculation of the full problem.

Usually, in the downfolded model one considers only local contributions in the two-body part, i.e. the contributions for \( \mathbf{R} = \mathbf{0} \) in equations (4) and (5), as the screening reduces the range of the interaction, making \( U_{nm}(\mathbf{R}, \omega) \) a localized function in space. Of course the localization of these matrix elements depends on the type of system and on the correlation strength. In fact, the elimination of the low-energy metallic screening (by unscreening \( W \) with \( \text{PL} \)) makes the partially screened interactions longer-ranged than the fully screened ones. Therefore, Hubbard-type approximations with only on-site interactions could be problematic. In fact, the restriction to Hubbard-like terms in most existing calculations is mainly due to the difficulty of dealing with long-ranged interactions in the solution of the low-energy model. Downfolded extended \( U-V \) models have however been taken into account in recent DMFT studies [72], and one of the purposes of this review is to explain the techniques which enable these simulations.

There has also been a vast amount of work focusing on the explicit solution of extended lattice models. These calculations keep the full spatial dependence of the Hamiltonian, both in the one- and two-body parts, but neglect dynamical effects. Successful methods of this kind are for instance lattice and diagrammatic quantum Monte Carlo (although affected by the sign problem and/or by finite-size effects) [27, 73–77], exact diagonalization (which is limited to very small sizes), and methods based on matrix product states and their variants (most appropriate for low-dimensional systems) [78–81]. For models with purely local interactions, the exact diagonalization approach is also useful within cluster embedding theories [22], such as the variational cluster approximation (VCA) [82], cellular dynamical mean field theory (CDMFT) [83], or the dynamical cluster approximation (DCA) [84]). Here, we focus our attention on Green’s function methods, and Green’s function embedding approaches, such as the DMFT, where significant progress has been made in the consistent treatment of dynamically screened interaction parameters and the corresponding retardation effects.

From the point of view of including dynamically screened nonlocal interactions, a very appealing framework is the GW+DMFT theory, which treats the \( k \)-dependent electron correlations at the GW level, while the local diagrams are summed up to infinite order by the DMFT solver. This is justified by the interaction-range reduction from screening, which makes the \( U \) matrix more localized in \( \mathbf{R} \). Therefore, a perturbative expansion of the self-energy may be a good approximation for the nonlocal \( U \) terms. However, we recall that a \( k \)-dependent \( \Sigma \) can also arise from purely local interactions, particularly in low dimensional systems (1D and 2D), where the strength of the resulting \( k \)-modulation can seriously challenge the validity of the perturbative nature of GW. As it will be explained in section 4.1.2, in the GW+DMFT formalism the dynamically screened interaction \( U(\omega) \) is replaced by an auxiliary local interaction \( U(\omega) \), which becomes frequency dependent due to screening provided not only by high-energy states (related to downfolding) but also by nonlocal contributions which enter the local polarization function (related to two-particle embedding). This double nature of screening is captured by the GW+DMFT self-consistency construction, where the effective hybridization \( \gamma_0 \) and effective local interaction \( U \) are fixed by a double constraint imposed on the local Green’s function and the local polarization, respectively.
3. Solving low-energy models with dynamically screened \( U(\omega) \)

Once a low energy lattice model with a few correlated orbitals and a frequency-dependent (retarded) monopole interaction \( U \) has been derived, we need a suitable method to solve it. Here, we focus on dynamical mean field (DMF) methods, which are adequate for the description of systems with a large coordination number, or systems in the strongly correlated (Mott insulating) regime, where the local physics is dominant. We will start by introducing the dynamical mean field approximation (section 3.1), and then discuss various strategies for dealing with the retarded interaction originating from the coupling to a single bosonic mode (section 3.2). These techniques will then be generalized to arbitrary \( U(\omega) \) in section 3.3. We also discuss the proper definition of effective low-energy models with static interactions (section 3.4).

3.1. Dynamical mean field theory

For simplicity, we discuss the dynamical mean field approximation [11] for a simple, but fundamentally important model for correlated electron materials, the single-orbital Hubbard model with a retarded on-site interaction \( U(\tau) \). Written in terms of fermionic creation and annihilation operators \( d^{\dagger} \) and \( d \), the action of the lattice model is given by

\[
S_{\text{Hubbard}} = -t \sum_{\langle ij \rangle} \int_0^\beta d\tau d^{\dagger}_{j}(\tau) d_{i}(\tau) + d^{\dagger}_{i}(\tau) d_{j}(\tau) - \mu \sum_{i} \int_0^\beta d\tau d^{\dagger}_{i}(\tau) n(\tau)n(\tau'),
\]

with \( n_{\tau} = d^{\dagger}_{\tau} d_{\tau} \), \( n = n_{\uparrow} + n_{\downarrow} \), \( t \) the hopping amplitude and \( \beta = 1/T \) the inverse temperature. The partition function of the lattice model is \( Z = \text{Tr}[\mathcal{T}e^{-S_{\text{Hubbard}}}] \), with \( \mathcal{T} \) the imaginary-time ordering operator. Note that the retarded interaction also couples electrons of the same spin. In the spirit of mean-field approximations, we now focus on one particular site of the lattice (black dot in the left panel of figure 3) and replace the remaining degrees of freedom of the model by a hybridization term, which describes the hopping of electrons in and out of the impurity. The effective single-site problem then becomes an Anderson impurity model [10] with retarded interaction,

\[
S_{\text{imp}} = \sum_{\sigma} \int_0^\beta d\tau d^{\dagger}_{\sigma}(\tau) \Delta_{\sigma}(\tau - \tau') d_{\sigma}(\tau') - \mu \sum_{\sigma} \int_0^\beta d\tau n_{\sigma}(\tau) + \frac{1}{2} \int_0^\beta d\tau d\tau' U(\tau - \tau') n(\tau)n(\tau').
\]

The hybridization function \( \Delta(\tau) \) plays the role of the dynamical mean field, which is computed self-consistently in such a way that the Anderson impurity model mimics the lattice environment as closely as possible. More precisely, the self-consistent solution is constructed such that the impurity Green’s function \( G_{\text{imp}}(\omega_n) \) reproduces the local lattice Green’s function \( G_{\text{loc}}(\omega_n) \equiv G_{\text{imp}}(\omega_n) \). If \( G(\mathbf{k}, \omega_n) \) is the momentum-dependent lattice Green’s function of the Hubbard model, we thus seek a hybridization function such that

\[
\int (d\mathbf{k}) G(\mathbf{k}, \omega_n) \equiv G_{\text{imp}}(\omega_n),
\]

where \( \int (d\mathbf{k}) \) denotes a normalized integral over the Brillouin zone.

It is also useful to introduce the Green’s function of the non-interacting impurity (‘Weiss Green’s function’) \( G_0 \), which is related to the hybridization function by

\[
[G_0]^{-1}(\omega_n) = \omega_n + \mu - \Delta(\omega_n).
\]

Depending on the method used, it is more natural to work with the Weiss Green’s function \( G_0 \) rather than the hybridization function \( \Delta \).

3.1.1. DMFT approximation. We obtain the solution of equation (12) iteratively. However, it is not immediately clear how we can use this self-consistency condition to update the dynamical mean field. To define a practical procedure, we have to relate the left-hand-side of equation (12) to impurity model quantities. This step involves, as the essential approximation of the DMFT method, a simplification of the momentum-dependence of the lattice self-energy [11, 85].

The self-energy describes the effect of interactions on the propagation of electrons. In the non-interacting model, the lattice Green’s function is \( G_0(\mathbf{k}, \omega_n) = [\omega_n + \mu - \epsilon_{\mathbf{k}}]^{-1} \), with \( \epsilon_{\mathbf{k}} \) the Fourier transform of the hopping matrix. The Green’s function of the interacting model is \( G(\mathbf{k}, \omega_n) \) \( = [\omega_n + \mu - \epsilon_{\mathbf{k}} - \Sigma(\mathbf{k}, \omega_n)]^{-1} \) with \( \Sigma(\mathbf{k}, \omega_n) \) the lattice self-energy. Therefore

\[
\Sigma(\mathbf{k}, \omega_n) = G_0^{-1}(\mathbf{k}, \omega_n) - G^{-1}(\mathbf{k}, \omega_n),
\]
Similarly, we obtain the impurity self-energy as
\[ \Sigma_{\text{imp}}(i\omega_n) = G_0^{-1}(i\omega_n) - G_{\text{imp}}^{-1}(i\omega_n), \]
(15)
with \( G_0^{-1} \) defined in equation (13). The DMFT approximation is the identification of the lattice self-energy with the momentum-independent impurity self-energy,
\[ \Sigma(k, i\omega_n) \approx \Sigma_{\text{imp}}(i\omega_n). \]
(16)
This approximation enables us to rewrite the self-consistency equation (12) as
\[ \int (dk)[i\omega_n + \mu - \epsilon_k - \Sigma_{\text{imp}}(i\omega_n)]^{-1} \equiv G_{\text{imp}}(i\omega_n). \]  
(17)
Since both \( G_{\text{imp}}(i\omega_n) \) and \( \Sigma_{\text{imp}}(i\omega_n) \) are determined by the hybridization function \( \Delta(\tau) \) (or \( G_0(\tau) \)), equation (17) defines a self-consistency condition for these functions.

3.1.2. DMFT self-consistency loop. We now formulate the self-consistency loop for the Weiss Green’s function \( G_{\text{imp}}(i\omega_n) \). Starting from an arbitrary initial \( G_0(i\omega_n) \) (for example, the local Green’s function of the non-interacting lattice model), we iterate the following steps until convergence:

(a) Solve the impurity problem, that is, compute the impurity Green’s function \( G_{\text{imp}}(i\omega_n) \) for the given \( G_0(i\omega_n) \).
(b) Extract the self-energy of the impurity model: \( \Sigma_{\text{imp}}(i\omega_n) = G_0^{-1}(i\omega_n) - G_{\text{imp}}^{-1}(i\omega_n) \).
(c) Identify the lattice self-energy with the impurity self-energy, \( \Sigma(k, i\omega_n) = \Sigma_{\text{imp}}(i\omega_n) \) (DMFT approximation), and compute the local lattice Green’s function \( G_{\text{loc}}(i\omega_n) = \int (dk)[i\omega_n + \mu - \epsilon_k - \Sigma_{\text{imp}}(i\omega_n)]^{-1} \).
(d) Apply the DMFT self-consistency condition, \( G_{\text{loc}}(i\omega_n) = G_{\text{imp}}(i\omega_n) \), and use it to define a new Weiss Green’s function \( G_0^{-1}(i\omega_n) = G_{\text{loc}}^{-1}(i\omega_n) + \Sigma_{\text{imp}}(i\omega_n) \). The computationally expensive step is the solution of the impurity problem (step (a)). When the loop converges, the Weiss Green’s function contains information about the topology of the lattice (through the density of states), and about the phase (metal, Mott insulator, antiferromagnetic insulator, ...). The impurity thus behaves, at least to some extent, as if it were a site of the lattice.

Obviously, a single-site impurity model does not capture all the physics. In particular, the DMFT approximation neglects all spatial fluctuations. These fluctuations are important, for example, in low-dimensional systems or near 2nd order phase transitions. The DMFT formalism is believed to provide a qualitatively correct description of three-dimensional unfrustrated lattice models (away from 2nd order phase transitions). It becomes exact in the limit of infinite dimension [9, 86] or infinite coordination number (where spatial fluctuations are negligible), in the non-interacting limit (\( U = 0 \) implies \( \Sigma = 0 \)), and in the atomic limit (\( t = 0 \) implies \( \Delta = 0 \)).

3.2. Holstein–Hubbard model
In a Hamiltonian formulation, a lattice model with dynamically screened \( U(\omega) \) can be represented by coupling the electron density on a given site to a continuum of bosonic modes with frequency \( \omega \), with appropriately chosen coupling strengths \( g_\omega \). If there is only a single bosonic mode with frequency \( \omega_0 \) and coupling \( g \), this corresponds to the Holstein–Hubbard model
\[ H_{\text{HH}} = -t \sum_{\langle ij \rangle} (d_{i \sigma}^\dagger d_{j \sigma} + d_{j \sigma}^\dagger d_{i \sigma}) + U \sum_i n_i n_i - \mu \sum_i n_i + g \sum_i (b_i^\dagger + b_i)(n_i + n_i - 1) + \omega_0 \sum_i b_i^\dagger b_i. \]
(18)
Here, \( b_i^\dagger \) denotes the boson creation operator. In the DMFT approximation, this lattice model is mapped onto a single-site impurity model with Hamiltonian \( H_{\text{imp}} = H_{\text{loc}} + H_{\text{hyb}} + H_{\text{bath}} \), where
\[ H_{\text{loc}} = U n_i n_i - \mu (n_i + n_i) + g(n_i + n_i - 1)(b_i^\dagger + b_i) + \omega_0 b_i^\dagger b_i, \]
\[ H_{\text{hyb}} = \sum_{\rho \rho} [V_{\rho \rho} d_{\rho}^\dagger d_{\rho} + V_{\rho \rho}^\dagger d_{\rho}^\dagger d_{\rho}], \]
\[ H_{\text{bath}} = \sum_{\rho \rho} \varepsilon_\rho b_{\rho}^\dagger b_{\rho}. \]
In terms of the harmonic oscillator position and momentum operators
\[ X = (b_i^\dagger + b_i)/\sqrt{2}, \quad P = i(b_i^\dagger - b_i)/\sqrt{2}, \]
(22)
the local Hamiltonian can (up to a constant) be written as
\[ H_{\text{loc}} = U n_i n_i + \sqrt{2} g (n_i + n_i - 1) X + \frac{\omega_0}{2} (X^2 + P^2), \]
(23)
so the physics of this model is that the charge couples to bosons describing either local lattice distortions (phonons) or local density fluctuations (plasmons).

The hybridization term \( H_{\text{hyb}} \) and the fermionic bath \( H_{\text{bath}} \) are defined such that the parameters \( V_{\rho \rho} \) and \( \varepsilon_\rho \) encode the hybridization function
\[ \Delta_\rho(i\omega_n) = \sum_{\rho} \frac{|V_{\rho \rho}|^2}{i\omega_n - \varepsilon_\rho}. \]
(24)
On the other hand, integrating out the bosons yields the frequency-dependent interaction
\[ U(i\omega_n) = U + \frac{2 g^2 \omega_0}{(i\omega_n)^2 - \omega_0^2}, \]
(25)
or, after analytical continuation to the real-frequency axis,
\[ \text{Re} U(i\omega_n) = U + \frac{2 g^2 \omega_0}{\omega^2 - \omega_0^2}, \]
\[ \text{Im} U(i\omega_n) = -g^2 \pi (\delta(\omega - \omega_0) - \delta(\omega + \omega_0)). \]
(26)
The real part of the frequency-dependent interaction, illustrated in the left panel of figure 4, therefore has poles at the
boson energy $\pm \omega_0$ and ranges from the unscreened interaction $U_{\text{bare}} = U$ at high frequencies to a static (screened) interaction

$$U_{\text{scr}} = U - 2g^2/\omega_0$$  

(28)

at $\omega = 0$.

While the DMFT approximation simplifies the problem considerably, by mapping the Holstein–Hubbard lattice model onto an auxiliary single-site impurity model, this effective model is still a complicated interacting many-body system. The electron-boson coupling introduces additional energy scales, besides the bandwidth and Kondo scale of the Anderson impurity model, namely the boson frequency $\omega_0$ and the effective coupling strength $\lambda = 2g^2/\omega_0$. (In the high-frequency limit, the Holstein–Hubbard model simplifies to the Hubbard model with interaction $U_{\text{scr}} = U - \lambda$.) Even in the DMFT approximation, and in the absence of long-range order, the Holstein–Hubbard model features a rich phase diagram with metallic, Mott insulating and bipolaronic insulating phases (section 3.2.4) [87–91]. Antiferromagnetic, charge-ordered, superconducting and supersolid phases can also be found [92–94] if symmetry breaking is allowed. In the following, we will discuss efficient, yet accurate numerical approaches for solving the Holstein–Hubbard impurity problem, and also show how these techniques can be generalized to models with a coupling to a continuum of bosonic modes (or arbitrary retarded interactions). In fact, in the context of DMFT based ab initio simulations of correlated materials, the numerical challenge of treating dynamically screened interactions has been a major bottleneck which has hampered the implementation of advanced LDA+DMFT or GW+DMFT schemes for many years. The techniques introduced in the following sections eliminate this bottleneck.

3.2.1. DALA approach. An approximate, but often very useful scheme is the so-called dynamic atomic-limit approximation (DALA) [95]. The main idea behind the DALA is to assume the following ansatz for the Green’s function in imaginary time:

$$G(\tau) = G^{\text{static}}(\tau, \mu, U_{\text{scr}}) \exp \left( \frac{1}{\beta} \sum_{n=0} U(\nu_n) - U_{\text{scr}}(\nu_n^2) / \nu_n^2 (e^{i\omega_0 \tau} - 1) \right).$$

(29)

where $G^{\text{static}}(\tau, \mu, U_{\text{scr}})$ is the ‘standard’ Green’s function of the model with static interaction $U_{\text{scr}}$. The factor which multiplies this Green’s function takes into account the dynamic nature of the interaction. For the Holstein–Hubbard model, $U(\nu_n)$ is defined in equation (25) and $U_{\text{scr}}$ is the static screened interaction, defined in equation (28) as the zero-frequency limit of $U(\omega)$. This ansatz can be extended to arbitrary retarded interactions, fillings and number of orbitals. The latter dependence enters only through $G^{\text{static}}$. The above ansatz is exact in the atomic limit [96], that is, when the hybridization function $\Delta$ is set to zero and the impurity is isolated from the bath. In that limit the Green’s function exactly factorizes into the Bose factor

$$F_{\text{DALA}}(\tau) = \exp \left( \frac{1}{\beta} \sum_{n=0} U(\nu_n) - U_{\text{scr}}(\nu_n^2) / \nu_n^2 (e^{i\omega_0 \tau} - 1) \right),$$

(30)

depending only on dynamic $U$ quantities, and a fermionic part $G^{\text{static}}$, which is the Green’s function of the atomic limit with static screened $U_{\text{scr}}$. Other trivial exact limits are the static and non-interacting regimes, where $F_{\text{DALA}} = 1$.

Away from the atomic limit, one can still define the Bose factor $F(\tau)$ by

$$F(\tau) = \frac{G(\tau)}{G^{\text{static}}(\tau)},$$

(31)

where $G(\tau)$ is the exact Green’s function of the dynamic $U(\omega)$ problem and $G^{\text{static}}(\tau)$ is the exact Green’s function of the Hubbard model with the same hybridization $\Delta$ as $G$ but with the static Hubbard repulsion $U_{\text{scr}}$. One may then approximate the exact $F$ by its atomic limit form $F_{\text{DALA}}$ in equation (30). On the other hand, $G^{\text{static}}(\tau)$ can be obtained quite accurately by the solution of the Anderson model with static $U_{\text{scr}}$ via ‘standard’ methods. Therefore, in the DALA framework, the Green’s function $G$ of the full problem is approximated by the product $G^{\text{static}}F_{\text{DALA}}$.

Several remarks are in order here: (i) the DALA approach is non-perturbative in both the electron–electron screened interaction (through the $G^{\text{static}}$ solution) and the electron-boson interaction, as one may argue from the exponential form of $F_{\text{DALA}}$. In fact, $F_{\text{DALA}}$ can also be seen as a cumulant expansion of the Green’s function in the retarded interaction around $\Delta = 0$. A similar formalism has been used to compute the
multiple plasmon satellites in silicon within the GW framework [8]. (ii) It has been shown that the DALA works quite well even away from the atomic limit, as long as $U_{\text{scr}}$ and $\omega_0$ are large compared to $\Delta$. The approximation is problematic when the hybridization becomes essential for the low-energy low-temperature properties of the system. In particular, the DALA does not fulfill the Friedel sum rule, which sets the zero frequency limit of the exact Green’s function, $\text{Im}G(i\omega^0) = -4i/D$, in a half-filled single-band system with a semicircular density of states of bandwidth $D$. On the other hand, the DALA gives the right high-energy properties of the Green’s function. (iii) Once the DALA solution of the Anderson impurity model with retarded interaction (equation (11)) is plugged into the DMFT cycle, the high-energy tails of the impurity Green’s function, correctly modulated by the Bose factor $F_{\text{DALA}}$, can have a strong impact on the self-consistent solution. In general, in the strongly correlated regime, the Mott transition line modified by retardation effects is quite accurately described by the DALA-DMFT solution of the lattice problem, as shown by a direct comparison with more advanced and accurate methods such as CTQMC (see section 3.2.3 and figure 6). (iv) The very simple structure of the factorization ansatz used in the DALA framework is suggestive of the physics of the system. Indeed, the analytic form of $F_{\text{DALA}}$ is the basis for a very effective analytic continuation method, described in section 5, which can produce the correct strength and position of the plasmon/phonon satellites in the spectral function of $G$. These high energy features are fingerprints of the dynamically screened electron–electron interaction and the DALA inspired analytic continuation can successfully resolve them.

### 3.2.2. Lang–Firsov approach

To introduce the Lang–Firsov (LF) approach, let us start with $H_{\text{imp}}$, the single-band single-phonon impurity Hamiltonian, introduced in section 3.2. The Lang–Firsov transformation is a unitary transformation $O$ of the fermion and boson local (impurity) operators, which eliminates the explicit electron-boson interaction and introduces dressed fermion quasiparticles [97]. It is defined as follows:

$$O = \exp \left( -\frac{g}{\omega_0} (b^\dagger + b)(n_{\uparrow} + n_{\downarrow} - 1) \right).$$

(32)

$$\tilde{b} = O^{-1}bO = b - \frac{g}{\omega_0}(n_{\uparrow} + n_{\downarrow} - 1),$$

(33)

$$\tilde{d} = O^{-1}dO = \exp \left( -\frac{g}{\omega_0}(b^\dagger - b) \right)d,$$

(34)

where $\tilde{b}$ and $\tilde{d}$ are the transformed boson and fermion local operators, respectively. Therefore, the transformed impurity Hamiltonian $\tilde{H}_{\text{imp}} = O^{-1}H_{\text{imp}}O$ reads

$$\tilde{H}_{\text{imp}} = \sum_\sigma (-\mu + g^2/\omega_0)d_\sigma^\dagger d_\sigma + (U - 2g^2/\omega_0)n_{\uparrow}n_{\downarrow} + \omega_0b^\dagger b$$

$$+ \sum_{\sigma\rho} \left[ V_{\sigma\rho} \exp \left( \frac{g^2}{2\omega_0} (b^\dagger - b) \right) d_\sigma^\dagger a_{\rho\sigma} + V^*_{\rho\sigma} \exp \left( \frac{g^2}{2\omega_0} (b^\dagger - b) \right) a_{\rho\sigma}^\dagger d_\sigma \right] + \sum_{\rho\sigma} \epsilon_\rho a_{\rho\sigma}^\dagger a_{\rho\sigma}. $$

(35)

A nontrivial effect of the LF transformation is that it changes the bare local Hubbard repulsion to the screened value $U_{\text{scr}}$ (equation (28)). The interaction between dressed quasiparticles $\tilde{d}(0)$ is reduced by the presence of bosons. Analogously, we can write the impurity Green’s function expressed in the transformed coordinates as

$$G(\tau) = Z^{-1}(\tau)\tilde{d}(0)d(\tau)\exp(-\beta\tilde{H}).$$

(36)

While the LF transformation reveals the role of bosons as mediators of the effective screened electron–electron interaction, which is the core of the screening theory presented in this review, the transformed $H_{\text{imp}}$ in equation (35) is not immediately useful, as it shifts all the complexity of the electron-boson interaction into ‘electronic polaron’, the dressed quasiparticles, i.e. electrons coupled to their surrounding polarization cloud. We need to make some approximation, which exploits the form of equation (35). A widely used one is to project the full Fock space onto the zero boson mode, the so-called LF approximation. The dynamic nature of the bosons is thereby reduced to their lowest harmonic level. It is clear that this approximation is good when $\omega_0 \gg E^*$, with $E^*$ the relevant energy scale (bandwidth, hybridization, or Hubbard $U$ repulsion) of the purely fermionic part. In this regime, the fermionic degrees of freedom are well separated from the bosonic high-energy ones. The LF approximation is exact in the antiadiabatic $\omega_0 \rightarrow \infty$ limit. It turns out that many realistic cases fall into the antiadiabatic regime, and the characteristic screening frequency is usually the plasma frequency $\omega_0$ of the homogeneous electron gas evaluated at the same average density as the real system. As shown in [95], $\omega_0 \gtrsim 15$ eV in many correlated materials, which is normally larger than the energy scale of the correlated bands.

Once $\tilde{H}_{\text{imp}}$ in equation (35) is projected onto the boson vacuum, the resulting LF approximated Anderson model becomes a standard (non-retarded) one with renormalized hybridization, shifted chemical potential, and screened on-site interaction:

$$H_0 = \sum_\sigma (-\mu + g^2/\omega_0)d_\sigma^\dagger d_\sigma + (U - 2g^2/\omega_0)n_{\uparrow}n_{\downarrow} + \omega_0b^\dagger b$$

$$+ \sum_{\sigma\rho} \left[ V_{\sigma\rho} \exp \left( \frac{g^2}{2\omega_0} (b^\dagger - b) \right) d_\sigma^\dagger a_{\rho\sigma} + V^*_{\rho\sigma} \exp \left( \frac{g^2}{2\omega_0} (b^\dagger - b) \right) a_{\rho\sigma}^\dagger d_\sigma \right] + \sum_{\rho\sigma} \epsilon_\rho a_{\rho\sigma}^\dagger a_{\rho\sigma}. $$

(37)

Analogously, the LF approximated form of the Green’s function in equation (36) is

$$G_{\text{LF}}(\tau) = \exp\left( -\frac{g^2}{2\omega_0} \right)\sum \left(\tau d(0)d(\tau)\exp(-\beta H_0)\right)$$

$$\equiv \exp\left( -\frac{g^2}{2\omega_0} \right)G_{\text{HF}}(\tau).$$

(38)

After integration over the fermionic bath, one obtains the following Lang–Firsov Green’s function for the Anderson–Holstein model in equations (19)–(21):
\[ G_{L2}(\omega_n) = \frac{\exp(-g^2\omega_n^2)}{\omega_n + \mu - g^2\omega_n - \exp(-g^2\omega_n^2)\Delta(i\omega_n) - \Sigma(U_{\text{int}})(i\omega_n)}, \]

where \( \Delta(i\omega_n) = \sum_{p\sigma}\left| V_{p\sigma} \right|^2/(i\omega_n - \epsilon_p) \) is the hybridization function and \( \Sigma(U_{\text{int}})(i\omega_n) \) is the self-energy, which depends on the screened value of \( U \).

We note that the same \( \exp(-g^2\omega_n^2) \) factor renormalizes both the hybridization function and the Green's function. Therefore, the physical effect of retarding bosons, at least in the antiadiabatic regime, is threefold: (i) They reduce the Hubbard repulsion \( U \) to its screened value, (ii) they reduce the hopping elements between the impurity and the bath, which in the DMFT framework means reducing the bandwidth of the related lattice model, and (iii) they reduce the weight of the low-energy Green's function, which is a manifestation of the spectral weight transfer from the correlated manifold to higher-energy satellites, due to boson shake-off processes. All these physical insights provided by the LF approximation will be used in section 3.4 to derive an effective static model which properly describes the low-energy effects of screening in realistic strongly correlated systems.

In contrast to the DALA approximation, the LF Green's function of equation (39) fulfills the Friedel sum rule. Therefore the LF approximation can be used to correct the low-energy behavior of the DALA, as suggested in [95]. However, the combination of the two theories is quite cumbersome. If a numerically exact result is desirable, it is better to use the continuous-time Monte Carlo approach extended to dynamic interactions, which we describe in the next section.

3.2.3. Continuous-time Monte Carlo approach. Continuous-time Monte Carlo simulations rely on an expansion of the partition function into a series of diagrams and the stochastic sampling of (collections) of these diagrams. We represent the partition function as a sum (or, more precisely, integral) of configurations \( c \) with weight \( w_c \),

\[ Z = \sum_c w_c, \tag{40} \]

and implement a random walk \( c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow \ldots \) in configuration space in such a way that each configuration can be reached from any other in a finite number of steps (ergodicity) and that detailed balance is satisfied,

\[ |w_{c_1}| p(c_1 \rightarrow c_2) = |w_{c_2}| p(c_2 \rightarrow c_1). \tag{41} \]

This assures that each configuration \( c \) is visited with a probability proportional to \( |w_c| \). One can thus obtain an estimate for the Green's function from a finite number of measurements:

\[ G = \frac{\sum_{c_1} w_{c_1} G_{c_1}}{\sum_{c} w_{c}} = \sum_{i=1}^{N_{\text{MC}}} w_{G_i} \frac{\sum_{c} \text{sign}_{c} G_{c}}{\sum_{c} |\text{sign}_{c}|} = \frac{(\text{sign} \cdot G)_{\text{MC}}}{(\text{sign})_{\text{MC}}}, \tag{42} \]

The error on this estimate decreases like \( 1/\sqrt{N} \).

The first step in the derivation of the continuous-time impurity solver [15] is to rewrite the partition function \( Z = \text{Tr}[e^{-\beta H_{\text{imp}}}] \) as a time ordered exponential using some interaction representation. We split the impurity Hamiltonian into two parts, \( H_{\text{imp}} = H_1 + H_2 \) and define the imaginary-time dependent operators in the interaction picture as \( O(\tau) = e^{i\beta H_0}Oe^{-i\beta H_2} \). We furthermore introduce the operator \( A(\beta) = e^{i\beta H_0}e^{-i\beta H_2} \) and write the partition function as \( Z = \text{Tr}[e^{-\beta H_0}A(\beta)] \). The operator \( A(\beta) \) satisfies \( dA(\beta)/d\beta = e^{i\beta H_0}(H_1 - H)e^{-i\beta H_2} = -H_2A(\beta) \) and can be expressed as \( A(\beta) = \text{Tr}[e^{i\beta H_0}O(\beta)] \).

In a second step, the time-ordered exponential is expanded into a power series,

\[ Z = \text{Tr}[e^{-\beta H_0}T_0] = \sum_{n=0}^{\infty} \int_0^{\beta} d\tau_1 \int_0^{\beta} d\tau_2 \cdots \int_0^{\beta} d\tau_n \text{Tr}[e^{-i\tau_1 \eta_1 H_1(\tau_1)} \cdots e^{-i\tau_n \eta_n H_1(\tau_n)}] e^{-\tau_n \eta_n H_2}], \tag{43} \]

which is a representation of the partition function of the form (40), namely the sum of all configurations \( c = \{\eta_1, \ldots, \eta_n\} \) with weight \( w_c = \text{Tr}[e^{i\beta H_0}O(\tau_1 \eta_1 H_1(\tau_1) \cdots \tau_n \eta_n H_1(\tau_n)) e^{-\tau_n \eta_n H_2}]d\tau_n \). (44)

In the following we will discuss the so-called ‘hybridization-expansion’ approach [98, 99], which is based on an expansion of \( Z \) in powers of the impurity-bath hybridization term \( H_{\text{hyb}} \), and an interaction representation in which the time evolution is determined by the local part \( H_{\text{loc}} + H_{\text{bath}} \).

In the case of the Holstein–Hubbard model [100] the trace in equation (43) is over the Fock states of the impurity, the fermionic bath and the bosonic bath (\( \text{Tr} = \text{Tr}_f\text{Tr}_{t}\text{Tr}_b \)), or an equivalent basis. After the expansion in the hybridization operators, the time-evolution from one hybridization event to the next does no longer couple the impurity and the fermionic bath. Since the fermionic bath is noninteracting, the trace over the \( a \)-states can be evaluated analytically, resulting in two determinants (one for each spin) of matrices \( M_{\sigma} \), whose elements are the hybridization functions \( \Delta_{\sigma} \) evaluated at the time-intervals determined by the hybridization operator positions. The weight of a Monte Carlo configuration corresponding to a perturbation order \( \sum_{\alpha} n_\alpha \) is \( (n_\alpha)_{\text{MC}} \) evaluated at the time-intervals determined by the hybridization operator positions. The time-ordered exponential is thus expressed as

\[ w(O(\tau) = Z_{0,\alpha}\text{Tr}_f\text{Tr}_b \left[ T e^{i\beta H_{\text{loc}}} \prod_{\alpha} d_{\alpha}(\tau_{\alpha,\eta_\alpha}) d_{\alpha}^\dagger(\tau_{\alpha,\eta_\alpha}) \cdots d_{t}(\tau_{t,\eta_t}) d_{t}^\dagger(\tau_{t,\eta_t}) \right] (\text{det} M_\alpha)^{-1}(d\tau)^{2n}, \tag{44} \]

where the matrix elements are \( M_{\sigma,i,j} = \Delta_{\sigma}(\tau_{\sigma,i,j}) \). We denote the time-ordered sequence of impurity creation and annihilation operators by \( \{O(\tau_i)\} (1 \leq i \leq 2n, n = \sum_{\alpha} n_\alpha) \). At this stage, the time-evolution from one operator to the next still includes the coupling to the bosons. In order to evaluate the trace over the boson states, we perform the Lang–Firsov [97] transformation introduced in the previous section (equation (32)). Here, we express it in terms of the boson position and momentum operators \( P \) and \( X \) defined in equation (22).
The unitary transformation specified by $e^{ip_{XY}}$ shifts $X$ by $X_0$, and if we choose $X_0 = (\sqrt{2} g / \omega_0)(n_1 + n_1 - 1)$, the transformed local Hamiltonian (23),
\[
\hat{H}_{\text{loc}} = e^{-ip_{XY}} \hat{H}_{\text{loc}} e^{ip_{XY}} = \hat{U} \hat{h} \hat{r} - \hat{\mu} (\hat{n}_1 + \hat{n}_2) + \frac{\omega_0}{2} (X^2 + P^2),
\]
has no explicit electron–phonon coupling anymore. $\hat{H}_{\text{loc}}$ is of the Hubbard form but with modified chemical potential and interaction strength,
\[
\hat{\mu} = \mu - g^2 / \omega_0, \quad (46)
\]
\[
\hat{U} = U - 2g^2 / \omega_0, \quad (47)
\]
Note that $\hat{U} = U_{\text{act}}$ (equation (28)). The transformation also affects the impurity creation and annihilation operators, which acquire a boson factor (equation (34)):
\[
\hat{a}_\sigma^\dagger = e^{ip_{XY}} \hat{d}_\sigma^\dagger e^{-ip_{XY}} = e^{g / \omega_0 (b^\dagger - b)} \hat{d}_\sigma^\dagger, \quad (48)
\]
\[
\hat{a}_\sigma = e^{ip_{XY}} \hat{d}_\sigma e^{-ip_{XY}} = e^{-g / \omega_0 (b^\dagger - b)} \hat{d}_\sigma, \quad (49)
\]
After this transformation, the electron and boson sectors are decoupled and the trace in equation (44) becomes the product of a term involving electron operators which is identical to the weight appearing in a Hubbard model simulation (with shifted $\hat{U}$ and $\hat{\mu}$), and a phonon term which is the expectation value of a product of exponentials of boson operators, to be evaluated with the noninteracting boson Hamiltonian $\omega_0 b^\dagger b$. The phonon contribution $w_{\text{b}}$ to the weight (44) can be written as
\[
w_{\text{b}}(\{O(\tau)\}) = Z_{0, b} (e^{ip_{XY}} \{O(\tau)\} e^{ip_{XY}} - 1), \quad (50)
\]
with $Z_{0, b}$ the noninteracting boson partition function, $\{\ldots\}_b = \text{Tr}_b \{\ldots\}$, $0 \leq \tau_1 < \cdots < \tau_2 b < \beta$, $s_i = 1$ ($-1$) if the $i$th operator is a creation (annihilation) operator and $A(\tau)$ given by
\[
A(\tau) = \frac{g}{\omega_0} (e^{\omega_0 b^\dagger} - e^{-\omega_0 b}). \quad (51)
\]
The expectation value is to be taken in the thermal state of free bosons. Using $e^{iA(\tau)} = e^{-g^2 / \omega_0 (b^\dagger - b)} e^{-g^2 / \omega_0 (b^\dagger - b)}$, one finds the disentangled expression
\[
w_{\text{b}}(\{O(\tau)\}) = Z_{0, b} e^{-g^2 / \omega_0 (b^\dagger - b)} \left\{ \sum_{s_i = 1} e^{iA(\tau)} \prod_{\tau < \tau_j < \beta} \sum_{s_j = 1} e^{iA(\tau_j)} \prod_{\tau_j < \tau_k < \beta} e^{iA(\tau_k)} \prod_{\beta < \tau_k} e^{-iA(\tau_k)} \right\} b, \quad (52)
\]
For thermal expectation values, we have the formula [101]
\[
\langle e^{iA} b^\dagger b \rangle_b = e^{\omega_0 / (\omega_0 - 1)}, \quad (53)
\]
which gives the final expression for the bosonic weight
\[
w_{\text{b}}(\{O(\tau)\}) = Z_{0, b} \exp \left\{ \sum_{2 \beta > \beta_1 > \beta_2} s_\beta K(\tau - \tau_2) \right\}, \quad (53)
\]
\[
K(\tau) = - \frac{g^2}{\omega_0^2} \frac{\cosh(\omega_0(\beta - \tau)) - \cosh(3\omega_0 / 2)}{\sinh(3\omega_0 / 2)}, \quad (54)
\]
Figure 5. Illustration of a segment configuration with one segment for spin up and spin down. The interaction and chemical potential contribution to the configuration weight can be obtained from the lengths and overlaps of the segments. The boson contribution is indicated by the blue dashed lines, which connect all pairs of operators.

The total weight (44) can be expressed as
\[
w(\{O(\tau)\}) = w_{\text{b}}(\{O(\tau)\}) w_{\text{Hubbard}}(\{O(\tau)\}), \quad (55)
\]
where $w_{\text{Hubbard}}$ denotes the weight of a corresponding configuration in the pure Hubbard impurity model (with parameters modified according to equations (46) and (47)). This weight can be efficiently computed using the segment picture [98] (see figure 5). For each spin, we mark the imaginary-time intervals corresponding to an occupied impurity state by a segment. Then, $w_{\text{Hubbard}}(\{O(\tau)\})$ can be obtained from the total length $l_\sigma$ of these segments and the total length $l_{\text{overlap}}$ of the overlaps between up-spin and down-spin segments:
\[
w_{\text{Hubbard}}(\{O(\tau)\}) = s Z_{0, \sigma} e^{(i/\omega_0) \hat{U}_{\text{act}}} \prod_{\sigma} \left( \det M_{\sigma}^{-1} \right) (d\tau)^{2d}, \quad (56)
\]
with $s$ a permutation sign related to the time-ordering of the impurity creation and annihilation operators.

In a segment insertion move, a pair of $\sigma$-operators is added to the sequence $\{O(\tau)\}$. For the Metropolis test, one has to compute the ratio $w_{\text{new}} / w_{\text{old}}$ and a determinant ratio $\det M_{\sigma}^{-1} / \det M_{\sigma, \text{old}}$ both at a cost $O(n)$. The ratio of the traces over the impurity states (evaluated from the segment lengths and overlaps) is also at most $O(n)$. If the move is accepted, the matrix $M_{\sigma}$ needs to be updated. This is the computationally most expensive step requiring an effort $O(n^3)$, but this is identical to the effort in the Hubbard model without electron–phonon coupling [15]. Hence, the treatment of the bosons does not affect the scaling of the algorithm.

Observables such as the Green’s function or double occupation can be measured as in the case with static interaction [98], i.e. from the elements of the inverse hybridization matrix and the segment overlaps. Specifically, the formula for the Green’s function reads
\[
G_\sigma(\tau) = \left\{ - \frac{1}{\beta} \sum_{\theta} \delta(\tau, \tau - \theta) (M_{\sigma, \theta}^2) \right\}, \quad (57)
\]
with $\delta(\tau', \tau) = \delta(\tau' - \tau)$. For $\tau' > 0$, $\delta(\tau, \tau') = -\delta(\tau' - \tau - \beta)$ for $\tau' < 0$, and $\delta(\tau, \tau')$ denoting the time of the $i$th annihilation ($i$th creation) operator. Efficient measurement schemes for the self-energy and vertex function [102] on the other hand need to be specifically adapted to the case of retarded interactions. A detailed discussion of these techniques can be found in [103].
As a side remark, we note that the DALA Bose factor reported in equation (30) of section 3.2.1 can be written in terms of the $K$ function (54) as $F_{\text{DALA}}(\tau) = \exp(-K(\tau))$. In the atomic limit, i.e. zero hybridization, only the 0th order term in the hybridization expansion survives, leading to a total Green’s function $G_\text{a}(\tau)$ consisting of the product of the global factor $F_{\text{DALA}}(\tau)$ and the static Green’s function, as explained in section 3.2.1. Therefore, the exact CTQMC hybridization expansion algorithm with retarded interactions naturally reduces to the DALA result in the atomic limit.

3.2.4. Phase diagram. We illustrate the DMFT phase diagram of the half-filled Holstein–Hubbard model [104] in the left panel of figure 6. The results are for a semi-circular density of states with bandwidth 4 and inverse temperature $\beta = 50$. A paramagnetic solution is enforced. At weak boson coupling $g$, there is a metallic phase at small $U_{\text{bare}}$ and a Mott insulating phase for sufficiently large $U_{\text{bare}}$. As $g$ is increased, the effective on-site interaction is reduced and eventually, the system makes a transition to a bipolaronic insulating phase. The dashed line indicates the value of $g$ at which the screened interaction $U_{\text{scr}}$ changes sign. Right panel: phase diagram in the space of screened and bare interaction. The approximate DALA results (section 3.2.1), corrected by the low-energy Lang–Firsov behavior, are also reported for $\omega_0 = 20$ and different coupling strengths. (Adapted with permission from [104].)

![Figure 6](image-url)

**Figure 6.** Left panel: half-filled metal-insulator phase diagram for $\beta = 50$ in the space of bare interaction $U_{\text{bare}}$ and boson coupling strength $g$ for different values of the screening frequency $\omega_0$ (DMFT solution for a semi-circular density of states with bandwidth 4). The dashed lines indicate the boson coupling at which the screened interaction $U_{\text{scr}}$ changes sign. Right panel: phase diagram in the space of screened and bare interaction. The approximate DALA results (section 3.2.1), corrected by the low-energy Lang–Firsov behavior, are also reported for $\omega_0 = 20$ and different coupling strengths. (Adapted with permission from [104].)

3.3. General $U(\omega)$

It follows from equation (27) that the Holstein–Hubbard model corresponds to a frequency-dependent interaction $U(\omega)$, whose imaginary part is a $\delta$-function at $\omega = \pm \omega_0$, with a weight given by $\mp g^2 \pi$. An arbitrary $U(\omega)$ can thus be thought of as arising from a Holstein-type coupling to a continuum of bosonic modes with energy $\omega$ and coupling strength $g_\omega$ given by $g_\omega^2 = -\Im U(\omega)/\pi$. According to the derivation in section 3.2.3 each boson contributes an effective ‘interaction’ $K(\tau) = -g_\omega^2 \sinh(\omega/2) \cosh((\omega - \tau)/2)$ between impurity creation or annihilation operators at imaginary times $\tau$ and $\tau$ (equations (53) and (54)). Hence, the hybridization expansion Monte Carlo simulation for a model with general $U(\omega)$ proceeds exactly as in the case of the Holstein–Hubbard model, but with the $K$-function replaced by [104]

$$K(\tau) = \int_0^\infty d\omega' \frac{\Im U(\omega') \cosh(\omega(\beta/2 - \tau)) - \cosh(\beta\omega/2)}{\sinh(\beta\omega/2)}$$

and the shifted interaction and chemical potential (equations (46) and (47)) given by

$$\tilde{\mu} = \mu + \int_0^\infty d\omega \frac{\Im U(\omega)}{\pi \omega},$$

$$\tilde{U} = U + 2 \int_0^\infty d\omega \frac{\Im U(\omega)}{\pi \omega} = U_{\text{scr}}.$$

The last identity follows from the Kramers–Kronig relation and the antisymmetry of $\Im U(\omega)$.

These formulas can also be derived from the impurity action with retarded $U(\tau)$, without the detour through the Hamiltonian representation [105]. In fact, it is easy to evaluate the interaction contribution $\frac{1}{\tau_0} \int_0^{\beta} d\tau d\tau' U(\tau - \tau') n(\tau)n(\tau')$ in the action (11) for a given segment configuration. In a first step, we split the interaction into its instantaneous contribution and a retarded part:

$$U(\tau) = U_{\text{bare}} \delta(\tau) + U_{\text{ret}}(\tau - \tau'),$$

$$U(\omega) = U_{\text{bare}} + U_{\text{ret}}(\omega).$$

(61)

(62)
\[ U_{\text{ret}}(\omega) \] has the same imaginary part as \( U(\omega) \), but its real part approaches 0 in the high frequency limit. Therefore, we have the spectral representation (\( 0 \leq \tau \leq \beta \))

\[
U_{\text{ret}}(\tau) = -\frac{1}{\pi} \int_{0}^{\infty} \text{d} \omega \text{Im} U(\omega) \frac{e^{-\omega \tau} - e^{-\omega \beta}}{e^{\omega \beta} - 1}.
\]

Plugging equation (61) into the expression for the interaction contribution and using \( n = n_1 + n_i \), as well as \( n_s^2 = n_s \), we find

\[
\begin{align*}
\frac{1}{2} \int_{0}^{\beta} \text{d} \tau' \text{d} \tau U(\tau - \tau') n(\tau) n(\tau') = U_{\text{bare}} \int_{0}^{\beta} \text{d} \tau n_i(\tau) \\
+ \frac{U_{\text{bare}}}{2} \int_{0}^{\beta} \text{d} \tau n_i(\tau) + n_i(\tau) \\
+ \int_{0}^{\beta} \text{d} \tau' n_i(\tau) U_{\text{ret}}(\tau - \tau') n_i(\tau') \\
+ \sum_{\sigma} \frac{1}{2} \int_{0}^{\beta} \text{d} \tau' n_{\sigma,i}(\tau) U_{\text{ret}}(\tau - \tau') n_{\sigma,i}(\tau').
\end{align*}
\]  

The first term yields \( U_{\text{bare}} \delta \text{overlap} \) and the second term a shift of the chemical potential. The contribution from the retarded interaction, which acts both between same and opposite spin electrons, is given by the last two terms. Since \( n_s = 1 \) only on a segment of spin \( \sigma \), and zero otherwise, we can express the retarded contribution as

\[
\begin{align*}
\sum_{k, k_{-1}} \int_{k_{-1}} \text{d} \eta \int_{k_{-1}} \text{d} \eta U_{\text{ret}}(\eta - \tau_2) + \frac{1}{2} \sum_{k} \int_{k} \text{d} \eta \int_{k} \text{d} \tau U_{\text{ret}}(\eta - \tau_2),
\end{align*}
\]

where \( k_i \) denotes the segment \( i \). The first term is the inter-segment contribution and the second term the intra-segment contribution of the retarded interaction. Now, let us assume that we know a \( \beta \)-periodic function \( h(\tau) \), which is even, and satisfies \( \frac{d^2}{d \tau^2} = U_{\text{ret}}(\tau) \) in the interval \( 0 < \tau < \beta \). With the help of this function, we can express the integral over a pair of segments as \( \int_{k_i} \text{d} \eta \int_{k_{-1}} \text{d} \tau U_{\text{ret}}(\eta - \tau_2) = -h(\tau_1 - \tau_2) + h(\tau_1 - \tau_2) + h(\tau_1 - \tau_2) - h(\tau_1 - \tau_2) \), where \( \tau_1(\tau_2) \) denotes the start (end) point of segment \( k \). Similarly, the double integral over segment \( k \) evaluates to \( h(\tau - \tau_2) - h(0) \). Hence, the retarded interaction energy for a segment configuration with \( n \) creation and \( n \) annihilation operators can be written as

\[
- \sum_{i > j} s_i s_j (h(\tau_i - \tau_j) - h(0)),
\]

where we assumed that the operators are time-ordered (\( \tau_i > \tau_j \) for \( i > j \)) and the sign \( s_i \) is +1 if the \( i \)th operator is a creation operator, and −1 if it is an annihilation operator.

From the expression (63) we find by double-integration that \( h(\tau) - h(0) = K(\tau) \), the function defined in equation (58). Hence, the retarded interaction contribution to the weight becomes \( \exp \sum_{i,j} s_i s_j K(\tau_i - \tau_j) \), as derived in section 3.2.3.

There is one subtlety which we need to consider: due to the slope-discontinuity in the \( \beta \)-periodic function \( h(\tau) \) at \( \tau = 0 \), the second derivative yields a delta-function contribution with weight

\[
- \frac{2}{\pi} \int_{0}^{\infty} \text{d} \omega \frac{\text{Im} U(\omega)}{\omega} = -\text{Re} U_{\text{ret}}(\omega = 0) = -(U_{\text{ret}} - U_{\text{bare}}).
\] 

So, if we use this \( h(\tau) \) in the formula for the retarded interaction contribution, we have to subtract \( -(U_{\text{ret}} - U_{\text{bare}})h(\tau) \) from the instantaneous interaction. The total instantaneous interaction thus becomes \( (U_{\text{bare}} + (U_{\text{ret}} - U_{\text{bare}}))h(\tau) = U_{\text{slow}}h(\tau) \), which means that the overlap contribution to the weight of a segment configuration has to be evaluated with the screened interaction. A similar shift is introduced to the chemical potential, so that the end result is identical to equations (59) and (60).

### 3.4. Effective static model

Even though there exist efficient numerical methods for treating dynamically screened monopole interactions, it is conceptually interesting to ask how to define a low-energy model with static interactions, which properly captures the low-energy physics. To derive this model, we again first consider the simple case of the Holstein–Hubbard model (18). Applying the Lang–Firsov transformation (97) \( \mathcal{H}_{\text{HF}} \to \mathcal{H}_{\text{LF}} = e^{\mathcal{F}_{\text{scr}}} \mathcal{H}_{\text{HF}} \) with \( S = \frac{e}{\tau_{\omega}} \sum (n_i + n_i - 1)(b_i^+ + b_i) \) yields a decoupled Hamiltonian in terms of the polaron operators (48) and (49):

\[
\mathcal{H}_{\text{LF}} = -i \sum_{\langle ij \rangle \sigma} (d_{ij,\sigma}^a d_{ji,\sigma}^a + d_{ij,\sigma}^a d_{ji,\sigma}) + U_{\text{scrb}} n_i n_{i+1} - \mu \sum_{i} n_{i \sigma} + \sum_{i} \omega_{b_i} b_i,
\]

with \( U_{\text{scrb}} = U - 2g^2 / \omega_0 \). As shown in [106], the low-energy effective model can now be defined by projecting this Hamiltonian onto the zero-boson subspace:

\[
\mathcal{H}_{\text{eff}} = \langle 0 | \mathcal{H}_{\text{LF}} | 0 \rangle = -Z_{\text{LL}} \sum_{\langle ij \rangle \sigma} (d_{ij,\sigma}^a d_{ji,\sigma}^a + \bar{d}_{ij,\sigma} \bar{d}_{ji,\sigma}) + U_{\text{scrb}} \sum n_i n_{i+1} - \mu \sum_{i} n_{i \sigma}.
\]

\( \mathcal{H}_{\text{eff}} \) is of the Hubbard form and involves the original fermionic operators \( d^i \) and \( d \), the screened interaction \( U_{\text{scrb}} \), and a hopping term reduced by a factor

\[
Z_{\text{LL}} = \exp[-g^2 / \omega_0^2].
\]

Hence, a crucial effect of the dynamical screening is an effective bandwidth reduction.

The projection onto the zero boson subspace is a good approximation if the screening frequency \( \omega_0 \) is large. As an illustration of this, we list in Table 1 the critical interaction strengths for the Mott transition in a DMFT simulation of the Holstein–Hubbard model with semi-circular density of states of bandwidth 4, at inverse temperature \( \beta = 100 \). The result from the exact DMFT treatment based on the method explained in section 3.2.3 is compared to the result of a DMFT treatment of the effective static model (69), with the bandwidth renormalization (70).
In a model with a general $U(\omega)$ we can again view the frequency dependence as arising from a coupling to a continuum of bosonic modes with frequency $\omega$ and coupling strength $g^2 = -\text{Im}U(\omega)/\pi$. The screened interaction is then given by equation (60) (or simply $\text{Re}U(\omega = 0)$), and the formula for the Lang–Firsov renormalization factor becomes

$$Z_b = \exp\left[\frac{1}{\pi} \int_0^\infty d\omega \frac{\text{Im}U(\omega)}{\omega^2}\right].$$

(71)

In a model with strongly correlated ‘$d$’ states and itinerant ‘$p$’ states, and associated hopping parameters $T_{pp}, T_{pd}$ and $T_{dd}$, the Lang–Firsov transformation and subsequent projection onto the zero-boson sector leads to a renormalization of each $d$ operator by a factor $\sqrt{Z_b} = (0 | \exp(\sum_{\omega}(b_l - b_l^\dagger)) | 0)$. Hence, the hopping part of the one-particle Hamiltonian is renormalized as [106]

$$\left(p_d^d Standard\right)\left(\begin{array}{c}
T_{pp} \\
\sqrt{Z_b} T_{pd} \\
Z_b T_{dd}
\end{array}\right)\left(p_d\right).$$

(72)

This expression shows that the bandwidth reduction implied by the effective model cannot simply be translated into an effective increase of the on-site interaction in the multi-band situation typically considered in first-principles calculations.

### 3.5. Multiorbital systems

Up to now, we have discussed dynamically screened on-site interactions in a single-orbital model. The extension of these techniques to multi-orbital systems is straightforward if the total charge on a given site is screened, that is, if the electron-boson coupling in a Hamiltonian formulation of the type

$$H_{\text{el}b} = \sum_{i,\lambda} g_{i,\lambda} n_i^{\text{tot}} n_{i,\lambda}^{\text{ann}},$$

(73)

with $i$ the site index, $b_{i,\lambda}$ the annihilation operator for the oscillator of frequency $\omega_{i,\lambda}$ at site $i$ and $g_{i,\lambda}$ the corresponding coupling strength. The total charge on site $i$ is $n_i^{\text{tot}} = \sum_{\alpha=\uparrow,\downarrow} n_i^{\alpha} = \sum_{\alpha=\uparrow,\downarrow} n_{i,\alpha}$. Within DMFT, one then has to solve a multi-orbital impurity system with an analogous electron-boson coupling. In the hybridization expansion formalism introduced in section 3.2.3, each hybridization operator changes the total charge on the multi-orbital impurity by $\pm 1$ and hence excites the bosons. After integrating out the bosons, one again finds an ‘interaction’ of the form (58) between each operator pair, with a sign that depends on the operator types (creation or annihilation operators). Now, $U(\omega)$ denotes the dynamically screened monopole interaction. In the trace calculation, the intra- and inter-orbital interaction parameters are replaced by the static values, in analogy to equation (60). Note that this procedure works both in the case of density–density interactions (segment formalism [98]), and for models with rotationally invariant interactions (matrix formalism [99]), because spin-flips and pair-hoppings do not change the total charge on the impurity and hence do not couple to the bosons.

A more complicated situation arises if the Hund coupling has a significant frequency dependence, such as in alkali-doped fullerides [53, 107], where Jahn–Teller screening may even produce an overscreened (negative) static $J$. In this case, the changes in the orbital occupation associated with spin-flips and pair-hoppings will couple to bosonic modes and the rotationally invariant system becomes much more difficult to simulate. After integrating out the phonons, the interaction part of the action reads [108]

$$S_{\text{int}} = \frac{1}{2} \sum_{\alpha,\beta} \int_0^\beta d\tau \int_0^\beta d\tau' \text{Im}U_{\alpha,\beta}(\tau - \tau') n_{\alpha,\beta}(\tau)$$

$$+ \frac{1}{2} \sum_{\alpha<\beta} \int_0^\beta d\tau \int_0^\beta d\tau' \text{Im}U_{\alpha,\beta}(\tau - \tau')$$

$$[X_{\alpha}^{\text{dd}}(\tau) X_{\beta}^{\text{dd}}(\tau') + X_{\beta}^{\text{dd}}(\tau) X_{\alpha}^{\text{dd}}(\tau') + X_{\beta}^{\text{dd}}(\tau) X_{\beta}^{\text{dd}}(\tau')],$$

(74)

with $X_{\alpha}^{\text{dd}} = \sum_{\lambda} g_{i,\lambda} n_{i,\lambda}^{\uparrow\downarrow}$ and retarded density–density interaction and Hund coupling parameters.

An algorithm based on a double-expansion in the hybridization and Hund coupling has recently been presented in [108]. The full problem with retarded spin-flips and pair-hoppings however suffers from a sign problem, so that in practice it seems necessary to treat these operators as instantaneous and keep the retardation only in the density–density component. Note that even the simulation of such a simplified model requires a double-expansion approach. To illustrate this, let us consider an on-site electron-boson coupling term of the form

$$H_{\text{el}b} = \sum_{i,\lambda} g_{i,\lambda} n_i^{\text{ann}} (b_{i,\lambda}^\dagger + b_{i,\lambda}),$$

(49)

$$\tilde{d}_{\alpha} = e^{\Sigma_{\alpha}^{\text{dd}}(i b_{i,\lambda}^\dagger - b_{i,\lambda})} \tilde{d}_{\alpha},$$

(48)

which transformation leaves density operators unchanged, $\bar{n}_i = n_i$, while a pair hopping operator $d_{i,\alpha}^\dagger d_{i,\beta}^\dagger d_{i,\beta} d_{i,\alpha}$ acquires a factor $e^{\Sigma_{\alpha,\beta}^{\text{dd}}(i b_{i,\lambda} - b_{i,\lambda})}$, which for example in the case of a Jahn–Teller coupling ($g_{\alpha} = -g_{\beta}$) is nonzero. Hence, we cannot keep such non-density–density terms in the time-evolution operators, but need to expand in them.
4. Towards a self-consistent description of screening

In the previous sections, we have discussed the cRPA technique for computing low-energy effective models for correlated materials, and different approximate and exact schemes for treating the resulting dynamically screened interactions within DMFT. We also mentioned the fact that the elimination of low-energy screening processes in cRPA generically leads to rather long-ranged interactions. In this section, we would like to improve the DMFT description by (i) including the screening effect from long-ranged Coulomb interactions, and (ii) considering nonlocal exchange and correlation effects. First, we will focus on the single-orbital extended Hubbard model (section 4.1), which allows us to introduce these advanced schemes in the simplest possible set-up, and then discuss how some of these techniques can be implemented in an ab initio context (section 4.2). The goal is to provide a self-consistent description of screening and correlations within the low-energy window defined by the cRPA downfolding, and in the future perhaps even within a larger window containing some of the screening bands.

4.1. U-V Hubbard model

4.1.1. Extended DMFT. The screening from nonlocal Coulomb interactions can be described by the so-called extended DMFT (EDMFT) formalism [12–14, 105]. This method is still based on an effective single-site impurity model, but involves two self-consistently computed dynamical mean fields: the Weiss Green’s function $G_{\beta}(\omega_n)$ (or hybridization function $\Delta(\omega_n)$), which controls the hopping of electrons in and out of the impurity site, and the dynamical on-site interaction $U(\nu_k)$, which incorporates the effect of screening. While $G_{\beta}$ is $\beta$-anti-periodic, $U$ is $\beta$-periodic. We denote the fermionic Matsubara frequencies by $\omega_n$ and the bosonic Matsubara frequencies by $\nu_k$.

For simplicity, we derive the EDMFT formalism for the U-V Hubbard model

$$H_{\text{UV-Hubbard}} = \sum_{\langle i,j \rangle} (d_{i\sigma}^\dagger d_{j\sigma} + d_{j\sigma}^\dagger d_{i\sigma}) + U \sum_{i} n_{i\sigma} - \mu \sum_{i} n_{i\uparrow} + V \sum_{\langle i,j \rangle} n_{i\sigma} n_{j\sigma}$$

with $n_i = n_{i\uparrow} + n_{i\downarrow}$. Here, we assume that the hoppings and off-site interactions are between nearest-neighbor sites, although the generalization to arbitrary hoppings and longer range interactions is straightforward. We furthermore take the bare interactions $U$ and $V$ as static. If the U-V Hubbard model is a low-energy effective theory derived from some downfolding procedure, then these bare parameters can themselves have a frequency-dependence. Again, the generalization of the following formalism to the case of frequency-dependent bare interactions is straightforward, and merely involves the replacement of $U$ by $U(\tau)$ or $U(\nu_k)$ and similarly for $V$.

We start by writing the action of the lattice model in terms of Grassmann fields $d^\dagger, d$ as

$$S = \int_0^\beta d\tau \left[ \sum_{\langle i,j \rangle} d_{i\sigma}^\dagger (\partial_\tau - \mu) d_{j\sigma} + U \sum_{i} n_{i\sigma}(\tau) + \frac{1}{2} \sum_{i,j} \nu_k n_{i\sigma}(\tau) n_{j\sigma}(\tau) \right]$$

$$= \int_0^\beta d\tau \left[ \sum_{\langle i,j \rangle} d_{i\sigma}^\dagger (\partial_\tau - \mu) d_{j\sigma} + U \sum_{i} n_{i\sigma}(\tau) + \frac{1}{2} \sum_{i,j} \nu_k n_{i\sigma}(\tau) n_{j\sigma}(\tau) \right]$$

(76)

In the second expression, we have written the interaction contributions in terms of $\tilde{\nu}_k = U\nu_k + \nu_k$ and shifted the chemical potential as $\tilde{\mu} = \mu + \frac{U}{2}$. Note that for the U-V Hubbard model, $t_j = -t\delta(\nu_j)$ and $\nu_k = V\delta(\nu_k)$ but it is more convenient to use the general notation.

Since we want to map the lattice model onto a single-site impurity model, we next decouple the interaction term by a Hubbard–Stratonovich transformation, thereby replacing the (on-site and off-site) interaction by a local coupling to a real, $\beta$-periodic field $\phi$ [109]:

$$\exp\left[ -\frac{1}{2} \int_0^\beta d\tau \sum_{\langle j \rangle} n_{j\sigma}(\tau) \tilde{\nu}_k n_{j\sigma}(\tau) \right] = ((2\pi)^{\nu_k} \det \tilde{\nu})^{-1/2} \int D[\phi_\sigma, \ldots, \phi_{3\sigma}]$$

$$\times \exp\left[ -\frac{1}{2} \sum_{\langle j \rangle} \phi_{j\sigma}(\tau) \tilde{\nu}_k \phi_{j\sigma}(\tau) + i \sum_{j} \phi_{j\sigma}(\tau) n_{j\sigma}(\tau) \right]$$

(77)

After this decoupling, the action of the lattice model can be written as

$$S = \int_0^\beta d\tau \left[ -\sum_{\langle i,j \rangle} d_{i\sigma}^\dagger(\tau)(G_{\tau}^{-1})_{ij} d_{j\sigma}(\tau) + \frac{1}{2} \sum_{\langle j \rangle} \phi_{j\sigma}(\tau) \tilde{\nu}_k \phi_{j\sigma}(\tau) + i \sum_{j} \phi_{j\sigma}(\tau) n_{j\sigma}(\tau) \right]$$

(78)

where we have introduced the inverse of the noninteracting lattice Green’s function, $(G_{\tau})_{ij} = ((-\partial_\tau + \tilde{\mu}) \delta_{ij} - t_j)$ to simplify the first term. In EDMFT, this lattice model is self-consistently mapped onto an impurity model with action

$$S_{\text{EDMFT}}^{\psi} = -\int_0^\beta d\tau d\tau’ \sum_{\sigma} d_{\sigma}(\tau) \tilde{G}^{-1}(\tau - \tau’) d_{\sigma}(\tau’) + \frac{1}{2} \int_0^\beta d\tau d\tau’ \phi(\tau) \tilde{U}^{-1}(\tau - \tau’) \phi(\tau’) + i \int_0^\beta d\tau \phi(\tau) n(\tau).$$

(79)

A detailed derivation can be found in [14] and [105]. For the present purpose it suffices to note that in addition to the fermionic Weiss field $\tilde{G}$, there also appears a $\beta$-periodic Weiss field $\tilde{U}$, which has to be adjusted in such a way that the local dynamics of the Hubbard–Stratonovich field is well reproduced. In the self-consistency loop, one computes the impurity Green’s functions.
and identifies them with the corresponding local lattice Green’s functions \( G^\text{lat}_{i,i} = \langle T_d(\tau) d_i(0) \rangle_{\text{lat}} \) and \( W^\text{lat}_{i,i} = \langle T_d(\tau) \phi(0) \rangle_{\text{lat}} \), where the latter are computed using a local approximation for the lattice self-energies \( \Sigma \) and \( P \):

\[
\Sigma(k, \omega_n) \approx \Sigma^\text{imp}(i\omega_n), \quad \Sigma^\text{imp}(i\omega_n) = G^{-1}(i\omega_n) - G^{-1}_{\text{imp}}(i\omega_n),
\]

\[
P(k, i\nu_n) \approx P^\text{imp}(i\nu_n), \quad P^\text{imp}(i\nu_n) = U^{-1}(i\nu_n) - W^{-1}_{\text{imp}}(i\nu_n).
\]

Physically, \( P(k, i\nu_n) \) (\( P^\text{imp}(i\nu_n) \)) represents the lattice (impurity) polarization. For the actual calculations, it is convenient to integrate out the \( \phi \)-field from the impurity action (79), to obtain

\[
S^\text{EDMFT} = -\int d\tau d\tau' \sum_\sigma d^\dagger_\sigma(\tau) G^{-1}(\tau - \tau') d_\sigma(\tau')
\]

\[
+ \frac{1}{2} \int d\tau d\tau' n(\tau) U(\tau - \tau') n(\tau') - \frac{1}{2} \text{Tr} \ln U.
\]

This impurity problem with retarded density–density interaction can be solved for example using the hybridization expansion Monte Carlo method discussed in section 3.2.3. While the fermionic Green’s function (80) can be measured directly, the evaluation of the bosonic Green’s function (81) requires an intermediate step. From equation (79) it follows that \( \frac{\partial S^\text{lat}}{\partial \chi_{\text{loc}}} = -2U \chi_{\text{loc}} U \), while equation (84) implies \( \frac{\partial S}{\partial U} = \frac{1}{2} \langle T_d(\tau) (n(0)) \rangle_{\text{lat}} - \frac{1}{2} U \). Combining the two expressions yields the measurement formula

\[
W^\text{imp} = U - U \chi_{\text{loc}} U, \quad \chi_{\text{loc}} = \langle T_d(\tau) n(0) \rangle_{\text{lat}}.
\]

The density–density correlation function \( \chi_{\text{loc}} \) can be easily evaluated in the hybridization expansion Monte Carlo method discussed in section 3.2.3.

The fermionic self-consistency loop in an EDMFT calculation is identical to usual DMFT:

(a) Compute \( G^\text{imp}(i\omega_n) \) for the given \( S^\text{EDMFT} \).

(b) Extract fermionic self-energy: \( \Sigma^\text{imp}(i\omega_n) = G_0^{-1}(i\omega_n) - G^{-1}_{\text{imp}}(i\omega_n) \).

(c) Use DMFT approximation \( \Sigma(k, i\omega_n) = \Sigma^\text{imp}(i\omega_n) \) to compute the local lattice Green’s function \( G^\text{loc}(i\omega_n) = \int d^k d(k) G_0^{-1}(i\omega_n, k) - \Sigma^\text{imp}(i\omega_n) \)^{-1}.

(d) Use DMFT self-consistency condition \( G^\text{imp}(i\omega_n) = G^\text{loc}(i\omega_n) \) to define a new Weiss Green’s function \( G_0^{-1}(i\omega_n) = G^{-1}_{\text{loc}}(i\omega_n) + \Sigma^\text{imp}(i\omega_n) \), while the bosonic self-consistency loop is analogous:

(a) Compute \( W^\text{imp}(i\nu_n) \) for the given \( S^\text{EDMFT} \) (equation (85)).

(b) Extract bosonic self-energy: \( P^\text{imp}(i\nu_n) = U^{-1}(i\nu_n) - W^{-1}_{\text{imp}}(i\nu_n) \).

(c) Use DMFT approximation \( P(k, i\nu_n) = P^\text{imp}(i\nu_n) \) to compute the local lattice Green’s function \( W^\text{loc}(i\nu_n) = \int d^k d(k) P_0^{-1}(i\nu_n, k) - P^\text{imp}(i\nu_n) \)^{-1}.

(d) Use DMFT self-consistency condition \( W^\text{imp}(i\nu_n) = W^\text{loc}(i\nu_n) \) to define a new retarded interaction \( U^{-1}(i\nu_n) = W^{-1}_{\text{loc}}(i\nu_n) + P^\text{imp}(i\nu_n) \).

In an EDMFT calculation, these two loops are typically solved in parallel, i.e. both the Weiss Green’s function \( G \) and the retarded interaction \( U \) is updated before the next impurity calculation is started. We finally remark that in the case of frequency dependent \( U \) and \( V \), the bosonic Dyson equation, which allows to update \( W^\text{loc} \) from \( P^\text{imp} \), will have frequency dependent \( (\Gamma)^{-1} \). Therefore, retarded \( U \) and \( V \) are readily included in this framework, as already mentioned before.

### 4.1.2. GW+DMFT

We next discuss the implementation of the GW+DMFT method for the \( U-V \) Hubbard model (75), for which fully self-consistent calculations have recently been implemented \([105, 110, 111]\). GW+DMFT is based on the EDMFT framework, but involves momentum-dependent fermionic and bosonic self-energies, which are obtained by combining the (local) EDMFT self-energies with the nonlocal components of the GW self-energies:

\[
\Sigma^{\text{GW+DMFT}}_{jk}(i\omega_n) = \Sigma^{\text{EDMFT}}_{jk}(i\omega_n) \delta_{jk} + \Sigma^{\text{GW}}_{jk}(i\omega_n)(1 - \delta_{jk}),
\]

\[
P^{\text{GW+DMFT}}_{jk}(i\nu_n) = P^{\text{EDMFT}}_{jk}(i\nu_n) \delta_{jk} + P^{\text{GW}}_{jk}(i\nu_n)(1 - \delta_{jk}).
\]

We note that this is not the only combination which avoids a double counting of self-energy diagrams. In fact, the subtraction of all the local GW diagrams also removes contributions (e.g. with nonlocal polarization bubbles) which are not accounted for in the EDMFT self-energy. An alternative strategy would be to remove the subset of GW diagrams which contains only local propagators: \( \Sigma^{\text{GW+DMFT}}_{jk}(i\omega_n) = \Sigma^{\text{EDMFT}}_{jk}(i\omega_n) \delta_{jk} + \Sigma^{\text{GW}}_{jk}(i\omega_n) - \Sigma^{\text{GW}}_{jk}(i\omega_n) G_{jk}(i\omega_n) \delta_{jk} \), and similarly for \( P \). In the case of the two-dimensional Hubbard model in the weak-coupling regime, both double counting corrections were found to produce similar results \([28]\), but in more general situations, the effect of different double counting schemes has not yet been studied systematically.

The computational steps in the self-consistent GW+DMFT calculation are the following:

(a) Start, e. g., from the converged EDMFT solution \( \Sigma^\text{imp}(i\omega_n), P^\text{imp}(i\nu_n) \), and define approximate lattice self-energies:

\[
\Sigma(k, i\omega_n) = \Sigma^\text{imp}(i\omega_n), \quad P(k, i\nu_n) = P^\text{imp}(i\nu_n).
\]

(b) Update the lattice Green’s functions:

\[
G(k, i\omega_n) = [G_0^{-1}(k, i\omega_n) - \Sigma(k, i\omega_n)]^{-1},
\]
\begin{equation}
W(k, \nu_n) = \frac{1}{2} \Sigma^{ij}(k) - P(k, \nu_n)^{-1},
\end{equation}
(c) Compute the local lattice Green’s functions and the new Weiss fields $G_0$ and $U$:
\begin{align}
G_{\text{loc}}(i\omega_n) &= \int (dk) G(k, i\omega_n) \\
&\rightarrow G_{\text{loc}}^{-1}(i\omega_n) = G_{\text{loc}}(i\omega_n)^{-1} + \Sigma_{\text{imp}}(i\omega_n),
\end{align}
(91)
\begin{align}
W_{\text{loc}}(i\omega_n) &= \int (dk) W(k, i\nu_n) \\
&\rightarrow U^{-1}(i\nu_n) = W_{\text{loc}}(i\nu_n)^{-1} + P_{\text{imp}}(i\nu_n),
\end{align}
(92)
(d) Solve the impurity problem, i.e. compute $G_{\text{imp}}$ and $\chi_{\text{loc}} \rightarrow W_{\text{imp}} = U - U\chi_{\text{loc}} U$. Compute the fermionic and bosonic self-energies
\begin{align}
\Sigma_{\text{imp}}(i\omega_n) &= G_0^{-1}(i\omega_n) - G_{\text{imp}}^{-1}(i\omega_n),
\end{align}
(93)
\begin{align}
P_{\text{imp}}(i\nu_n) &= U^{-1}(i\nu_n) - W_{\text{imp}}^{-1}(i\nu_n),
\end{align}
(94)
(e) Calculate the GW+DMFT self-energies:
- Calculate the GW self-energies,
\begin{align}
\Sigma^{GW}(k, i\omega_n) &= - \frac{T}{N_k} \sum_q \sum_{\nu_n} G(q, i\omega_n - i\nu_n) W(k - q, i\nu_n),
\end{align}
(95)
\begin{align}
P^{GW}(k, i\nu_n) &= 2 \frac{T}{N_k} \sum_q \sum_{\omega_n} G(q, \omega_n) G(k - q, \omega_n - i\nu_n),
\end{align}
(96)
- Extract the nonlocal parts,
\begin{align}
\Sigma_{\text{nonlocal}}^{GW}(k, i\omega_n) &= \Sigma^{GW}(k, i\omega_n) - \int (dk) \Sigma^{GW}(k, i\omega_n),
\end{align}
(97)
\begin{align}
P_{\text{nonlocal}}^{GW}(k, i\nu_n) &= P^{GW}(k, i\nu_n) - \int (dk) P^{GW}(k, i\nu_n),
\end{align}
(98)
- Combine GW and EDMFT self-energies
\begin{align}
\Sigma(k, i\omega_n) &= \Sigma_{\text{imp}}(i\omega_n) + \Sigma_{\text{nonlocal}}^{GW}(k, i\omega_n),
\end{align}
(99)
\begin{align}
P(k, i\nu_n) &= P_{\text{imp}}(i\nu_n) + P_{\text{nonlocal}}^{GW}(k, i\nu_n),
\end{align}
(100)
(f) Go back to (ii) until convergence is reached.

Note that the GW self-energy defined in equation (95) contains a Hartree contribution involving the bare local interaction. This contribution is however removed in the definition of $\Sigma_{\text{nonlocal}}^{GW}$ (equation 97).

An interesting question is how accurately the GW+DMFT method captures the momentum and energy dependence of the nonlocal self-energies. While a complete picture is lacking, some systematic tests against numerically exact diagrammatic Monte Carlo results [112] have recently been performed for the two-dimensional square-lattice Hubbard model in the weak-coupling regime [28]. It was shown that the DMFT approximation provides a very good description of $\Sigma_{\text{loc}}(i\omega_n)$, and that the GW+DMFT result is of comparable accuracy (figure 7).

While the nonlocal components are of the correct order of magnitude in the weak-coupling regime, their relative errors are large. Apparently, the GW approximation does not capture the correct momentum dependence at weak $U$ and for filling $n \lesssim 0.8$, and the result can only be expected to get worse in the intermediate coupling regime and closer to half-filling. In particular, GW+DMFT does not reproduce the strong differen-
tiation between node and antinode which is found in cluster DMFT simulations [24, 25] in the intermediate coupling regime. Figure 7 also shows the comparison to alternative many-body perturbation theory+DMFT schemes, namely the bare second order perturbation theory $(\Sigma^{(2)})$+DMFT [14] and the fluctuation exchange approximation (FLEX)+DMFT methods. They are of similar accuracy, but also fail to correctly capture the nonlocal components, $\Sigma^{(2)}$+DMFT at least ensures the correct high-frequency behavior of the local self-energy.

While GW+DMFT produces rather poor results for the momentum dependence of the two-dimensional Hubbard model, it should be kept in mind that (i) DMFT based methods are by construction most appropriate for high-dimensional systems, and (ii) that the main advantage of the GW+DMFT lies in the self-consistent description of the screening, and thus in the possibility to self-consistently compute the appropriate ‘Hubbard-U’ parameters in an ab initio simulation.

Finally, let us note that the factor of two in the GW polarization (96) comes from the sum over spin orientations. For the Hubbard model, with its spin-dependent instantaneous on-site interaction, the RPA polarization diagrams should in fact only include odd numbers of bubbles with alternating spin. To avoid unphysical diagrams, one should implement a spin-dependent GW formalism, which involves 2 $\times$ 2 matrices in spin space. While the polarization $P$ is diagonal, the Hubbard interaction becomes an off-diagonal matrix $U_\sigma$. Therefore, in a spin-dependent GW calculation for the Hubbard model, the self-energy $\Sigma$ is constructed with the following diagonal element of the screened interaction:
\begin{equation}
W_{\sigma\sigma}(k, \nu_n) = \frac{U^2 P(k, \nu_n)}{1 - [UP(k, \nu_n)]^2}.
\end{equation}
(101)

Systematic tests of the spin-dependent and spin-independent GW schemes for the two-dimensional Hubbard model [28] have shown that the spin-dependent formulation indeed cures the most obvious deficiencies of the spin-independent scheme. However, in models with nonlocal interactions and realistic material simulations within GW or GW+DMFT, this issue becomes less relevant.

4.1.3. Dual bosons. The dual boson formalism [113, 114] is a systematic extension of EDMFT, which incorporates momentum-dependent correlations and enables a consistent description of collective excitations. This method is still under active development, so we will content ourselves here with a sketch of the main ideas, and references to the original papers, where the mathematical details can be found. As in the previous sections, we will consider the $U$-$V$ Hubbard model and start the discussion by rewriting the Grassmann path-integral for the lattice action (76) in the Matsubara formalism:
where in practice one only retains low-order or ladder-type graphs. In the present context are called ‘dual bosons’. At the same time, the fermionic hopping and hybridization terms in the dual bosons become the zeroth order of this dual perturbation theory, in which case the dual boson formalism becomes a systematic approximation to the numerically exact diagrammatic Monte Carlo results for the square lattice Hubbard model with bandwidth 8. The left panels show the imaginary part of the local self-energy for indicated values of the interaction $U$ and filling $n$. The error bars are estimated from different cutoff orders in the diagrammatic sampling. The right panels show the imaginary part of the nonlocal self-energy for nearest neighbor sites, with the gray and blue shaded bands corresponding to the diagrammatic Monte Carlo result. Note the much smaller $y$-axis scale compared to the left panels. (Reproduced with permission from [28].)

The explicit form of the dual action is

$$S = -T \sum_{\omega, \sigma} d^\dagger_\sigma(i\omega_n)\Delta_{\sigma}(i\omega_n)d_\sigma(i\omega_n) + \frac{U_T}{2} \sum_{\nu, \sigma} n \Delta(\nu) + \frac{V_T}{2} \sum_{\nu, \sigma} n \Delta(\nu) - \sum_{\nu, \sigma} n \Delta(\nu) + \sum_{\nu, \sigma} \Delta(\nu),$$

with at this stage unspecified hybridization functions $\Delta_{\sigma}$, which in the present context are called ‘dual bosons’. At the same time, the fermionic hopping and hybridization terms in $S_{\text{rest}}$ are decoupled by an appropriate Hubbard–Stratonovich transformation. The decoupling of the interactions is analogous to EDMFT (equation (77)) and introduces the bosonic fields $\phi_{\nu}$, which thus becomes a sum of EDMFT-type impurity actions and a rest

$$S = \sum_{\nu} S_{\text{imp}, \nu} + S_{\text{rest}},$$

where the bare dual propagators, expressed in terms of the impurity Green’s function $G_{\text{imp}}(i\omega_n) = \langle d^\dagger(i\omega_n)d(i\omega_n) \rangle_{S_{\text{rest}}}$ and impurity charge susceptibility $\chi_{\text{imp}}(i\omega_n) = -\langle n(i\omega_n)n(-i\omega_n) \rangle_{S_{\text{imp}}}$ are [114]

\begin{align}
\tilde{G}_{\text{imp}}(i\omega_n) &= \left[ G_{\text{imp}}(i\omega_n) + \Delta(i\omega_n) - \epsilon_k \right]^{-1} - G_{\text{imp}}(i\omega_n), \\
\tilde{X}_{\text{imp}}(i\omega_n) &= \left[ X_{\text{imp}}(i\omega_n) + D(i\omega_n) - V_k \right]^{-1} - \chi_{\text{imp}}(i\omega_n),
\end{align}

with $\epsilon_k$ and $V_k$ the Fourier transforms of the hopping and nearest neighbor interaction.

The dual action (105) is treated in perturbation theory, where in practice one only retains low-order or ladder-type diagrams in $V$. The main idea is to exploit the freedom of choosing $\Delta(i\omega_n)$ and $D(i\omega_n)$, to define these quantities in such a way that the strong correlation effects are captured at the level of the impurity model $S_{\text{imp}}$, which can be solved exactly using the technique described in section 3.2.3, so that only weaker correlations must be treated by the dual perturbation theory. For example, the EDMFT solution can be incorporated as the zeroth order of this dual perturbation theory, in which case the dual boson formalism becomes a systematic approximation to the numerically exact diagrammatic Monte Carlo results for the square lattice Hubbard model with bandwidth 8. The left panels show the imaginary part of the local self-energy for indicated values of the interaction $U$ and filling $n$. The error bars are estimated from different cutoff orders in the diagrammatic sampling. The right panels show the imaginary part of the nonlocal self-energy for nearest neighbor sites, with the gray and blue shaded bands corresponding to the diagrammatic Monte Carlo result. Note the much smaller $y$-axis scale compared to the left panels. (Reproduced with permission from [28].)
expansion around EDMFT. Better choices for $\Delta$ and $D$, which take into account a feedback from the lattice solution onto the impurity problem may exist, as has been recently discussed in [114].

From the dual perturbation theory, one obtains a dual self-energy $\Sigma^i_k(i\omega_n)$ and a dual polarization $P^i_k(i\omega_n)$. These can then be used to obtain the lattice Green’s function $G^i_k(i\omega_n)$ and lattice susceptibility $\chi^i_k(i\omega_n)$ of the original $d$-fermions:

$$G^i_k(i\omega_n) = [G^\text{imp}(i\omega_n) + G^\text{imp}(i\omega_n)\Sigma^i_k(i\omega_n)G^\text{imp}(i\omega_n)]^{-1} + \Delta(i\omega_n) - \epsilon_k.$$  

$$\chi^i_k(i\omega_n) = [\chi^\text{imp}(i\omega_n) + \chi^\text{imp}(i\omega_n)P^i_k(i\omega_n)\chi^\text{imp}(i\omega_n)]^{-1} + D(i\omega_n) - V_k.$$  

While an appropriately formulated dual boson theory is self-consistent both on the single-particle and two-particle level [114], in contrast to GW+DMFT, this appealing feature comes at the cost of having to calculate and manipulate vertex functions. This makes it challenging to apply this formalism to realistic multiband systems.

### 4.2. Realistic materials

#### 4.2.1. GW+DMFT. The full implementation of the GW+DMFT scheme as proposed in [16] is at present doable only for simple Hamiltonians, such as the $U$-$V$ extended Hubbard model discussed above, or within a low-energy subspace of small size. This represents already a significant step forward with respect to the situation just a few years ago, when only static and non-self-consistent GW+DMFT applications had been performed. The methods which allow to treat the dynamically screened nature of $U$, inherent in the GW+DMFT formalism, and to determine it self-consistently in some particular cases, have been discussed in sections 3.2 and 3.3.

As already highlighted in the Hubbard model context of section 4.1.2, one of the main virtues of GW+DMFT is the appropriate description of dynamical screening. Clearly, this is also true for realistic materials. Additionally, GW+DMFT improves upon DFT or LDA+DMFT by adding important nonlocal exchange contributions to the self-energy. Functionals which depend strictly on the density include exchange contributions only in an approximate way, usually through the local density mapping to the homogeneous electron gas. This is the case, for instance, of the LDA functional. Recently, functionals including a fraction of the exact nonlocal exchange energy, such as B3LYP [115, 116], PBE0 [117], and HSE [118], have become very popular, although—strictly speaking—they are not pure density functionals as they depend explicitly on the Kohn–Sham orbitals. Moreover, their attempt to describe nonlocal exchange effects is empirical, as the weight of the exact exchange part is usually provided by ad hoc recipes. On the other hand, the GW framework, as well as GW+DMFT, include those effects in a truly first principle fashion.

However, in a full ab initio calculation, the local dynamical impurity problem is too large to be solved in a reasonable computer time. Indeed, the local basis set of an ab initio Hamiltonian can be very large, and in the original GW+DMFT formulation all local orbitals should be taken into account in the embedded site, as they all contribute to the screening of the effective local interaction $U$.

A more practical GW+DMFT implementation recently introduced is the so-called ‘orbital-separated’ scheme, where only the correlated orbitals are kept in the impurity problem, whose size becomes then manageable by state-of-the-art CTQMC solvers, provided the low energy model is a single-site multiorbital system, and not a cluster. On one hand, it is reasonable to include only the most correlated orbitals in the impurity problem. On the other hand, separating the self-energy contribution of the local, correlated manifold from the nonlocal or noncorrelated one has no unique solution, as there is no rigorous free energy functional which generates this separation unambiguously, as discussed in [119]. Therefore, in defining the local self energy, one needs to make an ad hoc choice. The one which seems the most general is...
\[ \Sigma(k, \omega_n)_{LL} = \Sigma_{GW}(k, \omega_n)_{LL} - \sum_k \left[ \Sigma^d_{GW}(q, \omega_n)_{LL} + \Sigma^d_{imp}(\omega_n)_{LL} \right], \]

where \( L \) is the full-orbital index, and \( d \) denotes the projection onto the low-energy correlated space. The corresponding equation for the total polarization is

\[ P(k, \omega_n)_{\alpha \beta} = P_{GW}(k, \omega_n)_{\alpha \beta} - \sum_q \left[ P^d_{GW}(q, \omega_n)_{\alpha \beta} + P^d_{imp}(\omega_n)_{\alpha \beta} \right], \]

where \( \alpha \) and \( \beta \) are indices of a two-particle basis, constructed from the full one-body basis set \( L \). Equations (110) and (111) are analogous to equations (86) and (87), written for the GW+DMFT calculation of the U-V Hubbard model. \( \Sigma^d_{imp} \) and \( P^d_{imp} \) are the (fermionic) self-energy and polarization (bosonic self-energy) computed as solutions of the impurity problem in the correlated local basis (see the extended DMFT description in section 4.1.1). From \( \Sigma(k, \omega_n)_{LL} \) and \( P(k, \omega_n)_{\alpha \beta} \), one obtains the dressed Green’s function \( G(k, \omega_n)_{LL} \) and the fully screened interaction \( W(k, \omega_n)_{\alpha \beta} \) by standard procedures. As in step (ii) of the GW+DMFT loop in section 4.1.2, one then computes the ‘Weiss’ fields \( G_0 \) and \( U \), after projection of the local \( G^{\text{loc}} \) and \( W^{\text{loc}} \) onto \( d \). Thus, the orbital-separated framework follows the usual GW+DMFT self-consistency loop, where the non-perturbative many-body solution is provided only in a correlated subspace, and the convergence is reached when the local lattice Green’s function and polarization projected to the correlated subspace become identical to the impurity Green’s function and impurity polarization, respectively.

Further simplifications may be needed to reduce the cost of the GW calculations within the self-consistency loop. With this aim, Boehnke et al [120] introduced an intermediate subspace \( I \), containing the local correlated subspace \( C \), in which the GW self-energy and polarization are self-consistently computed, while the components outside of \( I \) are treated at the \( G^W^{\Omega} \) or cRPA level. This ‘multi-tier’ approach can be illustrated by the formula

\[ G^{-1}_K = \left( G_{\text{LDA}}^{\Omega} \right)^{-1} \left[ \begin{array}{c} \omega_n + \mu - \varepsilon_{K}^{\text{LDA}} + V_{\text{xc}}^{\text{LDA}} - \Sigma^G_{K} - \Sigma^W_{K} \right] \]

where \( A_{I,I} \) represents the restriction of \( A \) to the \( I \) subspace. Here, \( G_{\text{LDA}}^{\Omega} \) is the LDA propagator, and \( \Sigma^G_{K} \) and \( \Sigma^W_{K} \) the effect of the rest subspace (everything outside \( I \)) on the ‘bare’ propagator \( G_{0,k}^{I} \) of the intermediate subspace. The analogous procedure has also been implemented for the polarization,

\[ W^{-1}_{q,q} = v_q^{-1} - p^G_{q} - p^{GG}_{q} - p^{EDMFT}_{q}, \]

where \( P^{GG}_{q} \) the GW estimate of the polarization, and \( P^{EDMFT}_{q} \) the more accurate EDMFT result for the local polarization. Here, \( v_q \) denotes the bare Coulomb interaction, and \( P^{\text{loc}}_{q} \) the screening effect of the rest subspace, which determines the effective ‘bare’ interaction \( U_q \) of the intermediate subspace. The self-consistent calculation in the scheme of [120] is restricted to the subspace \( I \) (tier I + tier II), which may be chosen much smaller than the full space with typically hundreds of bands.

Significant effort has also been devoted to the development of numerically more efficient approaches, which are based on additional approximations. On the one hand, the double self-consistency in the Green’s function and polarization has been replaced by only one based on the Green’s function. In these calculations, the polarization is frozen to the cRPA value, but the corresponding retarded interaction is kept in the impurity problem. This approximate scheme is detailed in section 4.2.2. On the other hand, a series of approximations has been proposed, which simplify the frequency and spatial dependence of the GW self-energy, which is one of the heaviest ingredients to compute. We will give a short survey of these simplified methods, which in order of increasing complexity are (i) SEX+DMFT (section 4.2.3), (ii) DMFT@nonlocal-GW (section 4.2.4), and (iii) quasiparticle self-consistent GW (QSGW)+DMFT (section 4.2.5).

4.2.2. Frozen polarization: \( U \) replaced by the cRPA \( U^{\Omega}(\omega) \). Instead of computing explicitly \( P^{\text{imp}} \) to update the bosonic Weiss field \( U \) at each GW+DMFT iteration, one can approximate it by its RPA value, i.e., \( P^{\text{imp}} = 2G^{\text{loc}}G^{\text{loc},d} \), with \( G^{\text{loc},d} \) the local starting Green’s function, taken from LDA and projected onto the correlated manifold [119, 121]. Therefore, the total polarization in equation (111) can be written as \( P = 2G^{\text{LDA}}G^{\text{LDA}} \). Its value is frozen during the self-consistency cycle, which is performed only on \( G \). This also implies that the interaction \( U \) of the impurity model is frozen. Moreover, instead of evaluating this interaction as \( U^{-1}(\omega) = W_{\text{loc}}(\omega)^{-1} + P^{\text{imp}}(\omega) \), step (ii) of the GW+DMFT loop, which involves local quantities only, \( U \) is calculated as

\[ U = \sum_{q} \left[ W^{-1} + P^{d}[-1] \right], \]

where \( W \) is the fully screened interaction. Hence, \( W \) is undressed by \( P^{d} \), which is the RPA polarization function containing electron–hole processes in the \( d \) manifold only. Note that the ‘locality’ operation \( (\sum_{q} \omega^{-1} - \Sigma_{k}^{\text{LDA}}) \) is performed after undressing \( W \), and the matrix is projected onto the \( d \)-manifold only at the end. Equation (114) is the partially screened cRPA value of \( U_{\text{imp}}(R = 0, \omega) \) in equation (4), where \( n, m \) are indices of the \( d \) subspace. Therefore, in this approach, \( U \) is kept fixed at the cRPA value. At a first glance, this might seem a very rough approximation, with respect to the double loop on \( G \) and \( W \). However, if compared to available electron energy loss spectroscopy (EELS) measurements, the RPA polarization function computed from the LDA band structure looks usually reasonable. This suggests that the cRPA estimate of
$U$ is quantitatively correct, particularly at not-so-low frequencies, where the interplay with the low-energy correlated manifold is supposed to be weak. In the impurity calculation, the cRPA frequency dependence of $U$ is taken into account and the self-consistent solution for $G$ is hence affected by retardation effects contained in the impurity model.

As far as $\Sigma_{GW}$ is concerned, one needs to carry out a one-shot GW calculation on top of the LDA band structure to compute the initial nonlocal self-energy part. In the simplified implementation based on fixed $P$, the nonlocal part does not change, while the local part is changed according to the solution of the dynamic impurity model. Thus, the resulting lattice Green’s function is

$$G(k, i\omega) = (i\omega + \mu - H_0 - \Sigma(k, i\omega))^{-1},$$

where $H_0 = H_{LDA} - V_{xc}^{LDA}$ is the LDA Hamiltonian without the exchange-correlation potential, and $\Sigma(k, i\omega)$ is the one defined in equation (110).

We note that if instead of replacing $V_{xc}$ by $\Sigma_{GW}$ one keeps $V_{xc}$ and adds a local $\Sigma$ only, the above scheme reduces to the DFT+$U$ approach. From this perspective, the DFT+$DMFT + U(\omega)$ method can be regarded as an embryo of the GW+DMFT method, which lacks nonlocality and a proper treatment of double counting, but where the dynamical nature of both local and nonlocal screening effects is taken into account in the effective interaction of the impurity problem via the cRPA estimate of $U(\omega)$.

In the DFT+$DMFT + U(\omega)$ scheme, one could ask what is the correct double counting term in the presence of a retarded $U$. By assuming that the spectroscopic high-energy features described by the coupling with plasmons cancel out in the zero-temperature mean-field solution to recover the potential exact DFT ground state energy, it turns out [122, 123] that the appropriate double counting is the same as the one introduced in section 2 for static $U$, which in this case takes the value of the screened static limit of $U(\omega)$ ($U = U(\omega \to 0)$).

4.2.3. SEX+DMFT. In the GW approach, the COHSEX approximation [1] is a way to simplify greatly the calculation of the self-energy, by separating it into two static contributions $\Sigma = \Sigma_{SEX} + \Sigma_{COH}$, where:

$$\Sigma_{SEX}(\mathbf{r}, \mathbf{r}') = -\sum_{\mathbf{k}\sigma} \psi^\dagger(\mathbf{r}) \psi(\mathbf{r}') W(\mathbf{r}, \mathbf{r}', 0),$$

(115)

$$\Sigma_{COH}(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \delta(\mathbf{r} - \mathbf{r}') (W(\mathbf{r}, \mathbf{r}', 0) - v(\mathbf{r} - \mathbf{r}')), $$

(116)

called screened exchange and Coulomb hole, respectively [2–4]. $W$ is the static fully screened interaction. While the first term significantly improves upon the exchange contribution in the Hartree–Fock theory, the second one describes the contribution to the self-energy due to interactions between the quasiparticle and its surrounding hole. As it is apparent in equations (115)–(116), the first term is nonlocal, while the second one is local.

The idea behind the SEX+DMFT theory [124] is to replace the static local Coulomb hole self-energy by a dynamic one provided by the DMFT solution of a downfolded Hubbard model with retarded $U$, i.e. $\Sigma = \Sigma_{SEX} + \Sigma_{DMFT}$. The clear advantage with respect to the COH self-energy is that the DMFT one is dynamical and non-perturbative. Moreover, we would like to point out that the locality of the DMFT self-energy is defined with respect to the local (or Wannier) basis set sitting on the same unit cell, while the locality of the COH term is much tighter, being related to the more specific condition $r = r'$. The validity of the SEX+DMFT theory is supported by the observation, verified in the iron pnictides and transition metal oxides such as SrVO$_3$, that at low-energy scales the nonlocal contributions to the self-energy are essentially static, while the local ones are dynamic (see, for instance, [119, 125], and section 7.3). The self-energy separation between static nonlocal terms on the one side and dynamic local terms on the other side is implemented in the SEX+DMFT by merging SEX and DMFT. This is done in the same spirit as in LDA+DMFT, except that the $H^{LDA} + V_{xc}$ Hamiltonian is replaced by $H_0 + \Sigma_{SEX}$. We note that SEX+DMFT is not double-counting error free, because the DMFT Hamiltonian contains a local Hartree term already included in $H_0$. This can be easily estimated by a mean-field approximation of the Hubbard terms. Therefore, a double counting correction is needed as in regular LDA+DMFT, or LDA+DMFT with dynamic $U$.

Despite this fact, SEX+DMFT improves upon the LDA+DMFT method with dynamic $U$. The reason is that the SEX part yields wider bands than LDA, which partially compensates the band narrowing produced by the frequency dependence of $U$. As found in the case of the compound BaCo$_2$As$_2$ (section 7.2), which is isostructural to the more famous BaFe$_2$As$_2$ and only moderately correlated, these opposite effects almost cancel each other in the final result, and rather accurate quasiparticle energies are obtained.

In the practical implementation of [124], the screened exchange contribution is calculated as a Fock exchange with the screened potential in the limit of long wavelengths, i.e.

$$W \approx W_{TF} = \frac{\kappa_{TF}^2}{\kappa_{TF}^2 + k^2},$$

where $\kappa_{TF}$ is the Thomas–Fermi wavevector or inverse screening length. In the Thomas–Fermi theory, its value depends on the density of states (DOS) at the Fermi level. Thus, a self-consistent determination of $W_{TF}$ can be devised, as a given $W_{TF}$ yields a new DOS, that implies a new $\kappa_{TF}$, that closes the loop by finally fixing a new $W_{TF}$, and so on. In the actual calculation of [124], this self-consistency has been replaced by a simple manual inspection to check that the guessed $k_{TF}$ is consistent with the final DOS.

The band widening produced by SEX has the same origin as the larger bandwidth found in the nonlocal self-energy framework by a number of authors [34, 126, 127]. The simplified static nonlocal version bears some important physics naturally present in the more involved full GW convolution. SEX+DMFT is the simplest theory capable of including nonlocal screening effects of the exchange part, besides the non-perturbative local correlations provided by DMFT. Therefore, SEX+DMFT goes in the direction of extending DMFT in a fully ab initio fashion, to include longer-range interactions beyond the Hubbard type.

4.2.4. DMFT@nonlocal-GW. The DMFT@nonlocal-GW approach [119], as the previous SEX+DMFT method, is
based on the observation (see section 7.3.1) that the local and nonlocal self-energy contributions are dynamically separable, with the former one frequency dependent and the latter one static. The DMFT self-consistency condition for the one-body quantities requires the local Green’s function to satisfy

\[ G^{\text{loc}}(i\omega) = \sum_{\mathbf{k}} \left[ i\omega + \mu - H_0(\mathbf{k}) - \Sigma_{\text{nonloc}}^{\text{GW}}(\mathbf{k}, i\omega) - \Sigma_{\text{imp}}(i\omega) \right]^{-1}, \]

(117)

with \( H_0 = H_{\text{LDA}} - V_{\text{Sc}}^{\text{LDA}}, \) and \( \Sigma_{\text{nonloc}}^{\text{GW}} \) the nonlocal part of the full GW \( \Sigma_{\text{GW}} \) self-energy. As the nonlocal correlation self-energies are purely static in the low-energy window, i.e. \( \Sigma_{\text{nonloc}}^{\text{GW}}(\mathbf{k}, \omega) = \Sigma_{\text{nonloc}}^{\text{GW}}(\mathbf{k}) \), one can construct an effective quasiparticle Hamiltonian that also comprises these exchange-correlation effects:

\[ H^{\text{QP}}(\mathbf{k}) = H_0(\mathbf{k}) + \text{Re}\Sigma_{\text{GW}}^{\text{nonloc}}(\mathbf{k}). \]

(118)

\( H^{\text{QP}} \) is a simplified one-shot analogue of the QSGW Hamiltonian \( H^{\text{QSGW}} \) that was proposed in the context of the QSGW+DMFT formalism [125, 128]. Then the DMFT self-consistency is much simpler since quantities are either frequency or momentum dependent, but not both, which drastically reduces memory requirements. Of course, the simplified DMFT@nonlocal-GW scheme is not expected to give quantitatively accurate results outside the quasiparticle energy range. In particular the dispersion of collective excitations will not be captured. However, their position in the local spectrum which is determined by the structure of the dynamic interaction \( U(\omega) \) is still meaningful. The results of the DMFT@nonlocal-GW for SrVO\(_3\) have been presented by Tomczak et al in [119]. A previous attempt on the same compound was presented by Taranto et al [129] by using a one-shot version of QSGW. However, in the latter work the authors obtained quite different results from Tomczak’s, with a much stronger quasiparticle renormalization (\( Z_\text{f} = 0.36 \)) when the Zc factor was used, probably due to a different way of dealing with the local self-energy subtraction at the QSGW level.

4.2.5. QSGW+DMFT. In the quasiparticle self-consistent GW (QSGW)+DMFT [125, 128], one defines an effective static Hamiltonian, which includes nonlocal and dynamic correlation effects, through the fully self-consistent QSGW construction [130, 131]. In QSGW+DMFT, an additional self-consistency on the GW-level is performed which circumvents the full GW+DMFT self-consistency that is computationally very demanding and has so far only been achieved on the model level [105, 110, 111], the simpler case of a two-dimensional system of adatoms on surfaces [72], and a \( t_2g \) model of SrVO\(_3\) [120]. In the QSGW+DMFT framework, once a self-consistent QSGW calculation is performed and the quasiparticles energies \( E_{\text{kn}} \) are found, the local part of the self-energy is subtracted to avoid double counting, and the nonlocal self-energy is evaluated at the corresponding \( E_{\text{kn}} \) energies. In this way, a static correction to the initial Hamiltonian is obtained, and incorporated into a modified Hamiltonian \( H^\text{al} \). Then, a DMFT calculation with dynamical \( U \) can be performed on top of \( H^\text{al} \). In contrast to DMFT@nonlocal-GW, dynamical nonlocal corrections can be incorporated into the QSGW+DMFT through the iterative QSGW construction. Suggested in [125] and later in [119], a simplified variant of this approach has recently been applied to the Mott insulator La\(_2\)CuO\(_4\) [132].

4.2.6. Concluding remarks on simplified GW+DMFT approaches. All methods in sections 4.2.3–4.2.5 make the calculation of the nonlocal GW part faster and more efficient. However, the self-consistency is performed at the \( G \) level only. Therefore, these schemes are not fully self-consistent GW+DMFT approaches. However, in the SEX+DMFT approach, it is possible to perform an additional self-consistency on the screening Fermi wave-vector, through the evaluation of the density of states at the Fermi level. This replaces the self-consistency at the polarization level in a purely static screening approach.

We remark that the Thomas–Fermi model used for the SEX part in [124] has several well-known limitations. Its exponential decay form is valid only for metals; in insulators there is a longer-range decay instead [133]. Moreover, the Thomas–Fermi model (as well as RPA) overestimates screening in metals. In fact, due to the singularity at \( 2k_F \), there are Friedel oscillations with a \( 2k_F \) period, decaying as \( 1/r^3 \), while in the Thomas–Fermi model the decay is always exponential. On the other hand, SEX+DMFT solves one of the major problems of the fully static COHSEX approximation, namely the lack of quasiparticle renormalization coming from the \( Z \) factor. Indeed, in SEX+DMFT, the DMFT part provides a non-perturbative frequency-dependent local self-energy, which usually yields a good estimate of \( Z \). Therefore, as future perspective, it is worth trying to implement better approximations for the SEX part, to go beyond the Thomas–Fermi model, within the SEX+DMFT framework.

5. Analytical continuation

At present, the Monte Carlo technique discussed in section 3.2.3 is the method of choice for the solution of impurity problems with dynamically screened interactions. For this reason, the extended DMFT or GW+DMFT calculations are most conveniently implemented on the Matsubara axis. For the interpretation of the results, it is however often important to have access to spectral functions. The analytical continuation from the Matsubara-frequency to the real-frequency axis is a delicate problem and particularly challenging in the case where high-energy features exist and need to be resolved. Both the Padé [134] and maximum entropy methods [135], which usually work well for low-frequency features in systems at low enough temperature, are not well-suited for capturing the high-frequency part of the spectrum. A pre- and post-processing of the data is needed in order to use these methods as part of a somewhat elaborate analytical continuation scheme. In section 5.1, we explain the scheme for calculating the fermionic spectral function corresponding to some \( \text{LDA} + \text{DMFT} + U(\omega) \) calculation. We will also briefly address the calculation of the bosonic spectral functions from EDMFT or GW+DMFT calculations, focusing in this case on models with static bare interactions (section 5.2).
5.1. Fermionic spectral function

The idea proposed in [95] is to split the Green’s function \( G(\tau) \) into a product of a bosonic function \( B(\tau) \) and an auxiliary fermionic Green’s function \( G_{\text{aux}}(\tau) \):

\[
G(\tau) = B(\tau)G_{\text{aux}}(\tau). \tag{119}
\]

From the corresponding spectral functions \( \rho_{B} \) and \( \rho_{\text{aux}} \) of the two factors one can then obtain the spectral function \( \rho(\omega) \) of the original Green function using the convolution

\[
\rho(\omega) = \int_{-\infty}^{\infty} d\tau \rho_{B}(\omega - \tau) \rho_{\text{aux}}(\tau), \tag{120}
\]

The bosonic function \( B(\tau) \) can be chosen arbitrarily, as long as the factorization (119) does not lead to unphysical properties of \( G_{\text{aux}} \). A natural choice, which often works in practice, is

\[
B(\tau) = e^{-K(\tau)}, \tag{121}
\]

where \( K(\tau) \) is the twice-integrated screening function defined in equation (58). The rationale for this choice is that such a factorization holds in the atomic limit (see section 3.2).

Substituting \( \tau = i\tau \) in equation (58), and expressing the factor \( \frac{\cos(\omega(\beta/2 - i\tau))}{\sinh(\omega/2)} \) as \( e^{-i\omega\tau} + \frac{2\cosh(\omega/2)}{e^{\omega/2} - 1} \), one finds

\[
K(t) = \int_{0}^{\infty} d\omega \frac{\text{Im} U(\omega)}{\pi\omega^{2}} \left( e^{-i\omega t} + \frac{2\cosh(\omega/2)}{e^{\omega/2} - 1} \right).
\]

The last step we used the fact that at low temperatures, \( e^{-\omega\tau} \gg 1 \), except near \( \omega = 0 \), where \( \text{Im} U(\omega) \) vanishes. At low enough temperature, we can therefore express the bosonic factor on the real-time axis as

\[
B(t) = e^{-K(t)} = \exp\left[ \int_{0}^{\infty} d\omega \frac{\text{Im} U(\omega)}{\pi\omega^{2}} \right] \exp\left[ -\int_{0}^{\infty} d\omega \frac{\text{Im} U(\omega)}{\pi\omega^{2}} e^{-i\omega t} \right]. \tag{123}
\]

The first term corresponds to the Bose factor defined in equation (71), while the second term oscillates around 1 as \( t \to \infty \). We thus split off \( Z_B \) and write

\[
B(t) = Z_B + B_{\text{reg}}(t), \tag{124}
\]

with the regular term \( B_{\text{reg}}(t) = Z_B \exp\left[ -\int_{0}^{\infty} d\omega \frac{\text{Im} U(\omega)}{\pi\omega^{2}} e^{-i\omega t} \right] - 1 \). The corresponding spectral density \( \rho_{B_{\text{reg}}}(\omega) \) can be obtained numerically from an appropriate Fourier transform. We have the spectral representation

\[
B_{\text{reg}}(t) = \int_{-\infty}^{\infty} d\omega' \rho_{B_{\text{reg}}}(\omega') \frac{e^{-i\omega't}}{1 - e^{-\beta\omega'}}, \tag{125}
\]

and therefore

\[
\int_{-\infty}^{\infty} d\omega e^{i\omega t} B_{\text{reg}}(t) = \int_{-\infty}^{\infty} d\omega' \int_{-\infty}^{\infty} d\omega'' \frac{e^{i\omega'' \tau} \rho_{B_{\text{reg}}}(\omega'')}{1 - e^{-\beta\omega'}} = 2\pi \rho_{B_{\text{reg}}}(\omega) \frac{1 - e^{-\beta\omega}}{1 - e^{-\beta\omega}} , \tag{126}
\]

Using \( B_{\text{reg}}(-t) = [B_{\text{reg}}(t)]^{*} \), we can write the left hand side as \( 2\text{Re} \int_{0}^{\infty} d\omega e^{i\omega t} B_{\text{reg}}(t) \), which finally yields the expression

\[
\rho_{B_{\text{reg}}}(\omega) = \frac{1 - e^{-\beta\omega}}{\pi} \text{Re} \int_{0}^{\infty} d\omega e^{i\omega t} B_{\text{reg}}(t). \tag{127}
\]

It immediately follows from this derivation that the constant \( Z_B \) in equation (124) gives a non-regular contribution

\[
\rho_{\text{non-regular}}(\omega) = Z_B(1 - e^{-\beta\omega})\delta(\omega)
\]

to the spectral density. Substitution of (128) into equation (120) gives

\[
\rho_{\text{non-regular}}(\omega) = Z_B \rho_{\text{aux}}(\omega), \tag{129}
\]

which is the expected renormalization of the quasiparticle peak by the Bose factor (see section 3.4 and equation (71)).

In summary, the spectral function \( \rho \) can be computed from the maximum entropy result for \( \rho_{\text{aux}}(\omega) \) and either the cRPA result or some numerical estimate for \( \text{Im} U(\omega) \) using the following formulas:

\[
\rho(\omega) = Z_B \rho_{\text{aux}}(\omega) + \int_{-\infty}^{\infty} d\omega' \rho_{B_{\text{reg}}}(\omega') \frac{1 + e^{-\beta\omega'}}{1 - e^{-\beta\omega}} \rho_{\text{aux}}(\omega - \omega'), \tag{130}
\]

\[
\rho_{B_{\text{reg}}}(\omega) = \frac{1 - e^{-\beta\omega}}{\pi} \text{Re} \int_{0}^{\infty} d\omega e^{i\omega t} B_{\text{reg}}(t), \tag{131}
\]

\[
B_{\text{reg}}(t) = Z_B \left[ \exp\left[ -\int_{0}^{\infty} d\omega \frac{\text{Im} U(\omega)}{\pi\omega^{2}} e^{-i\omega t} \right] - 1 \right]. \tag{132}
\]

5.2. Bosonic spectral function

Here we discuss a useful strategy for calculating the spectral function \( \text{Im} U(\omega) \) corresponding to some retarded interaction \( U_{\text{ret}}(\tau) \), as it is obtained for example in the self-consistency loop of an EDMFT or GW+DMFT simulation [111]. In a maximum entropy approach, it is important to work with a properly normalized spectral function. In the bosonic case, we can use the relation

\[
U_{\text{ret}} - U = 2 \int_{0}^{\infty} d\omega \frac{1}{\pi} \frac{\text{Im} U(\omega)}{\omega}, \tag{133}
\]

to define such a normalized distribution function:

\[
B(\omega) = \frac{2}{\pi(U_{\text{ret}} - U)} \frac{\text{Im} U(\omega)}{\omega}, \quad \int_{0}^{\infty} d\omega B(\omega) = 1. \tag{134}
\]

Equation (63), which connects the retarded interaction to the spectral density can then be written in the form

\[
U_{\text{ret}}(\tau) = \int_{0}^{\infty} d\omega K(\omega, \tau) B(\omega), \tag{135}
\]

with the bosonic kernel

\[
K(\omega, \tau) = \frac{\omega(U_{\text{ret}} - U) \cosh(\omega(\beta/2 - \tau/2))}{2 \sinh(\omega(\beta/2))}. \tag{136}
\]
Figure 9. Paramagnetic phase diagram of the half-filled $U$-$V$ Hubbard model on the 2D square lattice (a) and 3D cubic lattice (b). The unit of energy is $4t$ and the temperature is $T = 0.01$. Results for nonlocal interactions up to nearest neighbors (NN), next-nearest neighbors (NN+NNN) and third nearest neighbors (NN+NNN+3NN) are shown. FL denotes the metallic phase, CO the charge ordered insulating phase, and MI the Mott insulating phase. The insets show the phase diagrams of the models with NN interactions with axes rescaled by the bandwidth. (Reproduced with permission from [111].)

(If the $\omega$ integration is taken from $-\infty$ to $\infty$, the kernel becomes $K(\omega, \tau) = \frac{e^{-\beta \omega}}{1 - e^{-\beta}}$. The factor $(U_{\text{scr}} - U)$ is known from the solution on the Matsubara axis.

We can now use the maximum entropy method [135] to solve equation (135) for $B(\omega)$, and finally equation (134) to find $\text{Im}U(\omega)$. The real part can as usual be obtained from the antisymmetry of $\text{Im}U(\omega)$ and the Kramers–Kronig relation

$$
\text{Re}U(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\text{Im}U(\omega')}{\omega' - \omega} d\omega' = \frac{1}{\pi} \int_{-\infty}^{\infty} \text{Im}U(\omega') \left( \frac{\text{Im}U(\omega') - \text{Im}U(\omega)}{\omega' - \omega} \right). \quad (137)
$$

In the last step, we have re-expressed the integral in a form which is suitable for numerical treatment.

6. Applications to model systems

6.1. Extended DMFT

6.1.1. $U$-$V$ Hubbard model. The EDMFT solution for the half-filled $U$-$V$ Hubbard model (75) on the cubic lattice predicts a paramagnetic phase diagram with three phases: a metallic phase for small $U$ and small $V$, a Mott insulating phase for large $U$ and small $V$, and a charge ordered insulating phase for sufficiently large $V$ [14]. Recently, the low-temperature phase diagrams on the 2D square and 3D cubic lattices have been mapped out using the efficient and unbiased impurity solver described in section 3.2.3 [105, 111]. In figure 9 we show the results for generalizations of model (75) with nonlocal interactions up to the third-nearest neighbors. Here, the parameter $V$ encodes the strength of the nonlocal interactions, i.e. the nonlocal interactions are scaled as $V_{ij} = a \tau_{ij}$, with $a$ the lattice spacing and $r_{ij}$ the positions of the nearest-neighbor, next-nearest neighbor or third nearest neighbor sites. The unit of energy is the hopping.

In the square-lattice case with only nearest-neighbor interactions, the $V_c(U)$ phase boundary jumps near the intersection with the Mott transition line. In the models with longer-ranged interactions, the metallic phase extends to larger values of $U$, forming a ‘nose’ which separates the Mott insulator and charge ordered phases at low temperature. The jump in $V_c(U)$ disappears, so that the phase diagram looks qualitatively similar to that of the Hubbard–Holstein model with large phonon frequency (figure 6). However, in the Hubbard–Holstein case, the boundary to the charge ordered phase does not exhibit a slope change near the critical $U$ for the Mott transition, which indicates that the slope change in the EDMFT phase diagram originates from changes in the dominant screening modes near $U_c$ [111], and from the absence of a nonlocal Fock term in standard EDMFT calculations [136].

Overall, the phase diagrams for the 2D and 3D lattice are similar, with the main difference being the larger extent of the metallic nose in the 3D case, and a different dependence of the metal to charge order phase boundary on the interaction range. The latter can be explained by considering the lattice geometry and our definition of next-nearest and third-nearest neighbors [111]: in the 2D case, the third-nearest neighbor interactions act between sites on the same sublattice, and hence frustrate the charge order, while in the 3D case, they act between sites on different sublattices.

In order to identify the dominant screening modes, and understand their origin, it is instructive to compute the bosonic and fermionic spectral functions, as illustrated in figure 10 for a metallic and Mott insulating system. Let us focus first on the half-filled case (blue lines and symbols). As seen in panels (c) and (f), $\text{Im}W(\omega)$, which is proportional to the square of the coupling strength of the screening modes with frequency $\omega$, exhibits two peaks near $U$ and $U/2$ in the metallic system, and a single peak near $U$ in the insulating case. The comparison to the fermionic spectral functions plotted in panels (a) and (d) suggests that the peak at $U$ is related to transitions between the Hubbard bands, while the peak at $U/2$ in the metallic system originates from transitions between the quasiparticle band and one of the Hubbard bands.

If the chemical potential is shifted away from the particle–hole symmetric value (green curves), the spectral functions change. In the metallic system, the peak near $U/2$ grows and shifts to lower energies, consistent with the increased weight of the quasiparticle peak and the reduced separation between...
Figure 10. Fermionic and bosonic spectral functions for the paramagnetic Hubbard model with NN interactions on the square lattice. The unit of energy is 4\(t\) and the temperature is \(T = 0.01\). The blue curves show the results for a half-filled system and the green curves for a particle–hole asymmetric system with a chemical potential shift of \(\delta\mu = 0.4\) and 0.6 relative to the half-filled value \(\mu = U/2\). Panels (a)–(c) show results for a metallic system with \(U = 2.4\) and \(V = 0.2\), while panels (d)–(f) show results for a Mott insulating system with \(U = 3.6\) and \(V = 1.0\). The left panels plot the fermionic spectral function \(A(\omega)\), the middle panels the real part of the screened interaction \(\text{Re} W(\omega)\) on the Matsubara axis, and the right panels the bosonic spectral functions \(\text{Im} W(\omega)\). (Reproduced with permission from [111].)

this peak and the upper Hubbard band in the fermionic spectral function. The high energy peak also grows, which suggests that the asymmetric shape and population of the Hubbard bands enhances the screening effect. Despite the uncertainties in the analytical continuation, it appears that one can even identify a third low-energy mode associated with transitions within the quasiparticle peak. In the insulating case, the reduction in the gap size and the broadening of the lower Hubbard band are reflected in a shift of the peak in \(\text{Im} W(\omega)\) to lower energies, and a broader distribution of screening modes.

In the middle panels, we plot the real part of the screened interactions on the Matsubara axis. In the high-frequency limit \(\text{Re} W(\omega)\) approaches the bare on-site interaction \(U\), while below an energy scale determined by the dominant screening modes, the screened interaction is reduced. Here, we should recall the fact that this reduction is dominated by the low-energy modes (see equation (28)). Hence, \(\text{Re} W(0)\) is not much smaller than \(U\) in the insulator, while the low-energy screening modes in the metallic case lead to a substantially reduced static interaction.

6.1.2. Dynamical screening in photo-doped Mott insulators.

The real-time dynamics of screening has been recently investigated using a nonequilibrium generalization of EDMFT [137]. As discussed in [138], the DMFT formalism can be applied to nonequilibrium problems by solving the DMFT equations on a 3-branch Kadanoff–Baym contour which runs from 0 to some maximum time \(t_{\text{max}}\) along the real time axis, back to time 0 along the real time axis, and then to \(-i\beta\) along the imaginary time axis. The solution on the imaginary-time branch corresponds to the usual DMFT solution for the initial equilibrium state (with \(\beta\) the inverse temperature), while the solution on the real-time branches allows to describe the evolution of the system after some perturbation or in the presence of external fields. In a similar manner, the bosonic self-consistency loop of EDMFT can be solved on the Kadanoff–Baym contour and the resulting nonequilibrium EDMFT formalism then captures the changes in the screening properties resulting from the nonthermal state of the system. In [137], the \(U-V\) Hubbard model (75) on the square lattice was driven out of an initially Mott insulating state by a single-cycle electric-field pulse with frequency \(\Omega_{\text{pulse}} \approx U\) directed along the lattice diagonal. As shown in the right hand panel of figure 11, \(\text{Im} W(\omega)\), which represents the distribution of screening modes, initially exhibits a single broad peak centered at energy \(U\), similar to figure 10(f). This is because in a Mott insulator, the screening processes involve particle excitations across the gap.

However, as soon as the field pulse creates doublon-hole pairs, a second screening mode appears at lower energy, and the imaginary part of \(W(\omega, t)\) qualitatively resembles the equilibrium result for a metallic system (figure 10(c)). An essential difference to the equilibrium metallic system however is the absence of a coherent quasiparticle band [139], which implies that the low-energy screening modes in the photo-doped system are not associated with transitions between quasiparticle and Hubbard bands, but rather with screening transitions within the photo-doped Hubbard bands. This also explains the broader energy distribution. As shown in figure 11, the low-energy mode grows while the pulse (which lasts up to \(t \approx 1.5\)) produces additional carriers, and then essentially saturates. The subsequent slower evolution of the bosonic spectral function (not shown) is governed by
changes in the energy distribution of the photo-doped carriers, and on much longer timescales by doublon-hole recombination processes.

Looking at the real part of the screened interaction, which is plotted in the left hand panel, we see that the high-energy mode produces only a small screening effect, as expected in a Mott insulator, while the low-energy screening modes linked to screening transitions within the photo-doped Hubbard bands lead to a significant reduction of $\text{Re}W(\omega = 0, t)$. The evolution of the bosonic Weiss field $U(\omega, t)$ looks qualitatively similar to that of $W(\omega, t)$, but with smaller screening effects. The reduction of $U$ by the enhanced screening in the photo-doped Mott insulator results in a shrinking of the gap size and potentially even in a screening-induced transition to a transient metallic state.

6.2. GW+DMFT

6.2.1. U-V Hubbard model. The first self-consistent GW+DMFT calculations have been presented for the U-V Hubbard model (75) in [110] and [105]. The nonlocal component of the self-energy $\Sigma$ was found to be much smaller than the local contribution, even close to the Mott transition. This result is inconsistent with cluster DMFT data, which (for the half-filled Hubbard model) predict a strong momentum differentiation between the nodal and antinodal region, and even a momentum-selective metal-insulator transition [24]. However, as mentioned in section 4.1.2, the main purpose of the GW+DMFT scheme is not to provide an accurate description of the momentum-dependent self-energy, but to enable a self-consistent description of screening. Hence, it is more interesting to look at the polarization function $P(k, \omega)$. It was shown in [105] that the nonlocal components of the polarization can be of the same order of magnitude as the local polarization if the system is close to the charge ordering instability.

A relevant question is to what extent the self-consistent feedback of the momentum-dependent polarization and self-energy affects the converged result. To illustrate this, we compare in figure 12 different approximations for the polarization [105]: the local polarization obtained from EDMFT (panel (a)), the bubble $GG$ computed with EDMFT lattice Green’s functions (panel (c)), the sum of the local EDMFT polarization and the nonlocal part of the bubble (panel (b)), and the self-consistent GW+DMFT polarization (panel (d)). Let us first consider panels (a) and (c). While the bubble diagram $GG$ yields a $\mathbf{k}$-dependent polarization, it lacks vertex corrections beyond those built into the EDMFT propagators. The local EDMFT polarization, which is calculated from the density–density correlation function (85), contains the local vertex. An advantage of the GW+DMFT method is that it incorporates both this local vertex and the momentum-dependence of the bubble in a self-consistent manner. The effect of the self-consistent treatment becomes apparent by comparing panels (b) and (d). While we can combine the local EDMFT polarization and the nonlocal component of the EDMFT polarization bubble in the spirit of GW+DMFT, such a calculation lacks a self-consistent feedback, and as a result, the momentum-dependence of the polarization looks quite similar to the bubble result, away from the $(0, 0)$ point. In the GW+DMFT polarization, some of the momentum-dependent structures differ significantly from the ‘one-shot’ result in panel (b), especially near the $(\pi, 0)$ point.

The self-consistent feedback of the $\mathbf{k}$-dependent self-energy and polarization also has an effect on local observables, such as the local spectral function. In [110, 105] and [111] it was found that in the half-filled U-V Hubbard model, the self-consistent GW+DMFT calculation yields stronger correlation effects than EDMFT, as exemplified by a larger mass enhancement or more pronounced Hubbard bands. This is however only the case if the nonlocal Fock term is neglected in GW+DMFT (as it is in EDMFT). With the nonlocal Fock term included in the GW+DMFT calculations, the result is opposite [136]. Indeed, it was found in a recent study of the Hubbard model away from half-filling that the inclusion of the nonlocal GW

Figure 11. Time-evolution of the screened interaction $W$ in a Mott insulator which is excited by a mono-cycle electric field pulse with $\Omega_{\text{pulse}} = U$ (2D square lattice, bandwidth 8). The electric field pulse has the form $E(t) = E_0 \sin(\Omega_{\text{pulse}} (t - t_0)) \exp(-4.6(t - t_0)^2/\tau_0^2)$ with $t_0 = 2\pi/\Omega_{\text{pulse}}$ and $E_0 = 9. U = 10, V = 2$, initial equilibrium temperature $T = 0.2$. (Reproduced with permission from [137].)
Topical Review

28

diagrams reduces the correlation effects [28]. The latter study, which was restricted to the weak-coupling regime, concluded that GW$^+$DMFT provides slightly more accurate results for the local self-energy than DMFT, while the nonlocal components are not improved with respect to pure GW.

In the strong-coupling regime, where the failure of self-consistent GW to produce Hubbard bands is well known, the combination with DMFT enables physically meaningful self-consistent calculations [110]. Therefore, the main advantage of the GW$^+$DMFT scheme is that it enables a self-consistent treatment of screening effects at arbitrary interaction strength.

6.2.2. Adatom systems on semiconductor surfaces. An interesting playground to explore correlation effects in two-dimensional lattice systems are periodic systems of adatoms on a semiconductor surface [140]. A recent ab initio study of Si(111):X (X = Sn, C, Si, or Pb) based on cRPA downfolding and a self-consistent GW+DMFT solution of the low-energy effective theory has revealed the importance of nonlocal Coulomb interactions and dynamical screening effects in these systems and consistently explained material trends for this series of adatoms [72]. In this work, the one-particle part of the Hamiltonian was calculated in the LDA approximation, yielding a half-filled single band of predominantly $p_z$ character near the Fermi level, with a bandwidth of approximately 0.5 eV for all systems considered. The interaction parameters (partially screened Coulomb matrix elements) were calculated using cRPA and a low-energy window containing the surface band. The resulting static interaction parameters for the on-site ($U$) and nearest-neighbor ($V$) interactions are listed in table 2. Also shown for comparison are the bare interaction values, which ignore the screening effects from higher-energy bands.

While the on-site interaction is large, about 2–3 times the bandwidth, the nonlocal interactions are also substantial, and the resulting nonlocal screening effects are essential for understanding the properties of the different adatom systems. It is furthermore evident that the nearest-neighbor interaction is almost independent of the adatom type. The reason is the relatively large distance of 6 Å between the adatoms.

Table 2. Bare (effective) on-site interactions $U^{bare}$ ($U$) and nearest-neighbor interactions $V^{bare}$ ($V$). Also shown is the static value of the effective on-site interaction $\mathcal{U}(\omega = 0)$ obtained from GW+DMFT. All values are in eV. (From [72].)

|            | C | Si | Sn | Pb |
|------------|---|----|----|----|
| $U$        | 1.4 | 1.1 | 1.0 | 0.9 |
| $V$        | 0.5 | 0.5 | 0.5 | 0.5 |
| $U^{bare}$ | 6.0 | 4.7 | 4.4 | 4.3 |
| $V^{bare}$ | 2.8 | 2.8 | 2.7 | 2.8 |
| $\mathcal{U}(\omega = 0)$ | 1.3 | 0.94 | 0.84 | 0.67 (insulator) | 0.54 (metal) |

Figure 12. Comparison of various approximations for the momentum-dependent polarization function of the 2D $U$-$V$ Hubbard model. The unit of energy is $4t$ and the temperature is $T = 0.01$, $U = 1.5$, and $V = 0.4$. (a) Im $P_{amp}(\omega)$ (EDMFT), (b) Im $(P_{amp}(\omega) + G_{nonlocal}(k, \omega))$ (EDMFT), (c) Im $G_{nonlocal}(k, \omega)$ (EDMFT) and (d) Im $(P_{amp}(\omega) + G_{nonlocal}(k, \omega))$ (GW+EDMFT). (Reproduced with permission from [105].)
which implies that the intersite Coulomb energy is essentially that of two point charges. It was furthermore found that \( V \) is very close to the value of \( V_{\text{bare}} \) divided by the static dielectric constant of the silicon surface, which suggests that one can compute the long-range interactions by rescaling \( V \) with \( a/r \), where \( a \) is the nearest neighbor distance. In this sense, \( V \) parametrizes the strength of all the nonlocal interactions. The \( 1/r \) tail can be treated by an Ewald summation.

The low-energy model with these static on-site and off-site interactions (a particular realization of the single-orbital \( U-V \) Hubbard model) was solved using self-consistent GW+DMFT \[72\]. The dynamical interaction \( U(\omega) \) obtained within this scheme is plotted in the left panel of figure 13 and reflects the nonlocal screening effect on the local interaction. While at high frequencies, screening is not effective and \( U(\omega \rightarrow \infty) = U \), the static value of \( U \) can be substantially reduced, especially in a metallic system (Si(111):Pb). As a result, also the gap values are smaller than they would be in the absence of nonlocal screening.

The simulation results, which reproduce the experimentally observed materials trends, are summarized in the schematic phase diagram of figure 13 (right panel). While all the considered systems have the same strength of the nonlocal interactions, their on-site interactions differ. As a result of this, the screening effect in SrVO\(_3\):C and SrVO\(_3\):Si is small (see left panel), which places these systems in the Mott insulating region of the phase diagram, while SrVO\(_3\):Sn and Si(111):Pb are close to a metallic solution (the latter one is in fact in a coexistence regime). It was also found that Si(111):Pb is close to a charge ordering instability.

### 7. Applications to realistic materials

#### 7.1. LDA+DMFT+U(\(\omega\))

The frequency-dependent interaction parameters derived from cRPA have been employed in several recent \textit{ab initio} simulations based on the LDA+DMFT framework \[55, 122, 141, 142\]. These simulations have produced high-energy satellites and enhanced correlations in often good agreement with experiments. In this section, we illustrate the important effects of the dynamical \( U(\omega) \) by focusing on three materials: SrVO\(_3\) (section 7.1.1), hole-doped BaFe\(_2\)As\(_2\) (section 7.1.2) and undoped La\(_2\)CuO\(_4\) (section 7.1.3).

#### 7.1.1. Mass enhancement and satellites in SrVO\(_3\)

The correlated metal SrVO\(_3\), with an undistorted perovskite structure, has been studied extensively within LDA+DMFT, LDA+DMFT+U(\(\omega\)) and variants of GW+DMFT. It is a suitable test material for new computational schemes, because of its relatively simple band structure, with a well-defined low-energy window containing the three \( t_{2g} \) bands. The LDA bandwidth is 2.6 eV. Experiments indicate a substantial narrowing of these bands, by about a factor of two, and the appearance of satellites below and above the renormalized quasiparticle band \[143\]. Conventionally, these satellites have been interpreted as Hubbard bands, but recent theoretical results force us to reconsider this interpretation.

LDA+DMFT calculations with a static \( U = 5 \) eV and a Hund coupling parameter \( J = 0.68 \) eV were shown in \[144\] to produce the correct band renormalization compared to photoemission experiments \[145\]. The resulting \( k \)-integrated spectrum features a pronounced lower and upper Hubbard bands at \(-1.8 \) and \( 3 \) eV, respectively. A more recent study \[146\] employed \( U = 5.5 \) eV, which results in an even larger splitting between the Hubbard bands. LDA+DMFT calculations with a dynamically screened \( U(\omega) \) have been performed within the DALA approximation (section 3.2.1) in \[95\] and using the full cRPA interaction and the CTQMC scheme of section 3.2.3 in \[141, 147\]. The frequency dependent \( U \) used in \[147\] is plotted in the left panel of figure 2. It has a static value of 3.4 eV and a dominant pole structure near \( \omega = 14 \) eV. The width of the renormalized quasiparticle band predicted by the LDA+DMFT+U(\(\omega\)) calculation is 0.9 eV, which is too narrow compared to experiment. As we will argue in section 7.3.1, the missing ingredient in this calculation is the widening of the band due to the momentum dependence of the self-energy. In this section, we would like to comment on the proper static-\( U \) description (section 3.4), and the implications for the position of the Hubbard bands.

If the band-widening effect of the \( k \)-dependent self-energy is neglected, the static description should involve the cRPA \( U(\omega = 0) \) and a bandwidth reduced by the Bose factor \( 71 \), which for SrVO\(_3\) is \( Z_B = 0.689 \). Since the material has three
equivalent $t_{2g}$ bands, shrinking the bandwidth by $Z_B$ is equivalent to increasing the static interaction to $3.4/Z_B = 4.93$ eV and rescaling the frequency axis by $Z_B$.

The left panel of figure 14 compares the local spectral function from the dynamic-$U$ simulation (red line) to a static-$U$ simulation with enhanced on-site interaction before (dashed black line) and after (blue line) the rescaling of the frequency axis. The dashed line is essentially the result of the previous LDA+DMFT simulations [144]. After the rescaling of the frequency axis by the factor $Z_B$, one recovers the mass renormalization of the quasiparticle band and the Hubbard band positions of the $U(\omega)$ calculation. The upper Hubbard band is placed at 2 eV, instead of 3 eV, and the lower Hubbard band is shifted from about −2 eV to −1.6 eV. The latter position is consistent with photo-emission experiments [143]. The spectral weight which is lost by the rescaling of the frequency axis is shifted to high-energy satellites near ±15 eV, and also at ±5 eV (see right panel), which cannot be extracted from a static-$U$ simulation. Looking at the cRPA result for $U(\omega)$ (figure 2), we see that the energies of these satellites are determined by the dominant screening modes (peaks in $\text{Im} U(\omega)$). In fact, the analytical continuation procedure discussed in section 5 by construction leads to satellite features at the corresponding energy offsets.

This example illustrates the general fact that LDA+DMFT+$U(\omega)$ simulations can produce the same mass renormalizations as static-$U$ LDA+DMFT simulations, but with a lower $U(\omega = 0)$ (here, 3.4 eV instead of 5 eV). As a consequence, the splitting between the Hubbard bands is reduced, or the Hubbard bands may not even be well defined anymore, as in the example discussed in the following section.

### 7.12. Spin-freezing crossover in hole-doped BaFe$_2$As$_2$

BaFe$_2$As$_2$ is a prototypical compound of the so-called 122 family of iron based superconductors. It becomes superconducting under pressure, or by hole- and electron-doping. The hole-doped compound exhibits nontrivial correlation effects, even in the normal phase (above the maximum $T_c$ of 38 K [148]), as exemplified by the widely varying experimental estimates of the mass enhancement [149, 150]. The remarkable sensitivity of the electronic structure to changes in temperature, pressure or doping was shown [122] to be related to the proximity of the optimally hole-doped compound to a spin-freezing crossover. Spin-freezing [151] has recently been recognized as a generic and important phenomenon affecting the properties of multi-orbital systems with Hund coupling in a certain regime of filling, interaction strength and temperature [152]. Inside the spin-frozen regime, long-lived magnetic moments appear, which leads to strong scattering and bad metallic behavior. The boundary of the spin freezing regime is characterized by fluctuating local moments [153] and non-Fermi liquid properties, in particular a self-energy which varies as a square-root of frequency (rather than linearly with frequency) on the Matsubara axis [151]. Such a square-root self-energy leads to strong band renormalizations at low energy, even in system which do not exhibit Hubbard bands. Due to the dramatic changes in the electronic structure in the spin-freezing crossover regime, an \textit{ab initio} simulation of hole-doped BaFe$_2$As$_2$ requires an accurate estimation of the interaction parameters.

The left panel of figure 15 shows the real and imaginary parts of the partially screened interaction $U(\omega)$ for the Fe-$d$ states estimated from cRPA [122]. The real part varies from the static value $U(\omega = 0) = 3.6$ eV to a bare value $U_{\text{bare}} \approx 20$ eV. The plasmon excitation near $\omega \approx 26$ eV overlaps with single-particle excitations, which results in a broad peak in $\text{Im} U(\omega)$. In order to properly judge the importance of the different features we also plot $\text{Im} U(\omega)/\omega^2$. As discussed in section 3 it is the twice-integrated retarded interaction, or $\text{Im} U(\omega)/\omega^2$, which enters into the calculation of the diagram weights in a hybridization expansion solver (equation (58)) and the Bose renormalization factor (equation (71)). Apart from a dominant peak at 3.8 eV, which results from the lack of ‘high-energy’ screening processes below this frequency, there are additional peaks at 6.1, 12 and 16 eV. As in the case of SrVO$_3$ discussed in the previous section, such sharp structures can be expected to lead to side-bands in the $d$-electron spectral function at the corresponding energies. Indeed, as shown in the middle panel (inset) of figure 15, we can identify these peaks in the spectral function computed with the procedure described in section 5. It is interesting to note that a satellite at approximately $-6.5$ eV has been seen in photoemission experiments [154, 155].
The middle panel also shows a comparison of the LDA + DMFT + $\omega$ spectral function for optimally doped BaFe$_2$As$_2$ to the result obtained with LDA + DMFT using the static interaction parameter $U_{\text{static}} = U(\omega = 0)$ and to the LDA density of states. (A static Hund coupling parameter $J = 0.675$ was used in the DMFT calculations.) The much stronger renormalization of the quasiparticle peak in the dynamic-$U$ calculation is due to the non-Fermi liquid self-energy in the spin-freezing crossover regime: near optimal doping and for the simulation temperature of 145 K, the effective increase of the Coulomb interaction due to the barely screened fast charge fluctuations pushes the system closer to the spin-freezing region, resulting in large mass enhancements. In the underdoped regime, the behavior is more Fermi liquid like and the effect of the dynamic $U$ on the electronic structure is less pronounced, while in the overdoped region, the spin-freezing leads to very short quasiparticle life-times. The location of the spin-freezing region in the space of doping and temperature is sketched in the right-hand panel of figure 15 (yellow region).

The calculated renormalized band structure of optimally doped BaFe$_2$As$_2$ is shown in figure 16 with the left panel plotting the result from the static-$U$ approximation and the middle panel the much more strongly renormalized and smeared out bands obtained in the dynamic-$U$ simulation. As discussed in section 3.4 the proper static model involves an effectively reduced bandwidth. In the case of BaFe$_2$As$_2$ the renormalization factor $Z_B = 0.59$ is rather low [106]. The right hand panel of figure 16 shows the renormalized low-energy bands from a static-$U$ calculation with such a renormalized bandwidth. Due to the enhanced correlation effects, we now get the proper mass enhancement and also the broadening of the bands due to the scattering with local moments. While the low-energy physics is correctly reproduced by the model with reduced bandwidth, this static description will of course not produce any high-energy satellites.

7.1.3. Mott gap and $-13\text{eV}$ satellite in La$_2$CuO$_4$. Low energy models of cuprates usually involve the Cu $d_{x^2-y^2}$ and O $p_x$ and $p_y$ orbitals. The one-band description considers the antibonding combination of these orbitals, while the three-band model also takes into account the bonding combination of Cu $d_{x^2-y^2}$ and O $p_x$, as well as the non-bonding $p$ orbital. LDA+DMFT+$U(\omega)$ calculations based on the cRPA estimate of $U$ have recently been analyzed by both models in [142]. The conclusion of this study was that in both models, a static approximation $U = U_{\text{static}}$ fails to open a Mott gap, while in the three-band model the dynamic-$U$ calculations yields a gap.
of approximately 1.9 eV, in good agreement with experiment [156], provided that p-d interactions are accounted for within the DMFT self-consistency loop, at least at the Hartree level. The important role of interatomic Hartree potentials in the correct positioning of the p-bands and the opening of the gap is consistent with previous results [157] for a related three-band model.

The partially screened $U(\omega)$ for the three-band model of La$_2$CuO$_4$ is plotted in the left panel of figure 17. The broad peak in $\text{Im}U(\omega)$ centered at $\omega = 30$ eV is a plasmon excitation coupled to single-particle excitations. At lower energies, $\text{Im}U_{pd}$ exhibits a sharp peak at $\omega = 9$, which is absent in $\text{Im}U_{pp}$ and $\text{Im}U_{pd}$. This indicates that the collective excitation associated with the 9 eV peak is localized on the Cu site. The prominent 9 eV peak is primarily responsible for the low band filling of approximately 1.9 eV, in good agreement with experiment [156], provided that p-d interactions are accounted for within the DMFT self-consistency loop, at least at the Hartree level. The important role of interatomic Hartree potentials in the correct positioning of the p-bands and the opening of the gap is consistent with previous results [157] for a related three-band model.

The partially screened $U(\omega)$ for the three-band model of La$_2$CuO$_4$ is plotted in the left panel of figure 17. The broad peak in $\text{Im}U(\omega)$ centered at $\omega = 30$ eV is a plasmon excitation coupled to single-particle excitations. At lower energies, $\text{Im}U_{pd}$ exhibits a sharp peak at $\omega = 9$, which is absent in $\text{Im}U_{pp}$ and $\text{Im}U_{pd}$. This indicates that the collective excitation associated with the 9 eV peak is localized on the Cu site. The prominent 9 eV peak is primarily responsible for the low band filling of approximately 1.9 eV, in good agreement with experiment [156], provided that p-d interactions are accounted for within the DMFT self-consistency loop, at least at the Hartree level. The important role of interatomic Hartree potentials in the correct positioning of the p-bands and the opening of the gap is consistent with previous results [157] for a related three-band model.

The k-integrated spectral function and its p-electron and d-electron contributions are plotted in the right hand panel of figure 17. At the simulated temperature of 1200 K, antiferromagnetic correlations do not play an important role, so that the gap in the spectral function is a Mott gap. The upper Hubbard band has a width of about 2 eV , while the lower Hubbard band obtained in this simulation covers the same energy range as the p-bands and is hence not easily identified. The states near the lower gap edge, which have mixed p-d character, may be interpreted as Zhang–Rice singlet states [158]. The most interesting feature in this spectral function, as far as dynamical screening is concerned, are the two satellites at $-13$ eV and $+10$ eV. They originate from the sharp 9 eV peak in $\text{Im}U(\omega)$, and should not be confused with the Hubbard bands. A satellite feature at $-13$ eV has indeed been observed in photoemission measurements [159].

72. SEX+DMFT results for BaCo$_2$As$_2$

In the quest for new iron-based high-T$_c$ superconductors, a significant effort has been devoted to alloying iron with other ferromagnetic 3d elements, such as Co and Ni. Ba(Fe,Co)$_2$As$_2$ and Ba(Fe,Ni)$_2$As$_2$ are superconducting, but with a lower $T_c$ than the ‘parent’ BaFe$_2$As$_2$ compound. While pristine BaNi$_2$As$_2$ is still a superconductor, BaCo$_2$As$_2$ is a paramagnetic metal close to a ferromagnetic instability, as suggested by a high Wilson ratio [160]. The latter material is particularly interesting from the theoretical viewpoint, because it challenges ab initio methods to predict the correct magnetism and ARPES data. Density functional theory in the standard local spin density approximation (LSDA) gives a ferromagnetic ground state, which is not consistent with experiment. Moreover, the LDA bandwidth in the paramagnetic solution is a factor 1.5 too large compared to ARPES. The LDA density of states is peaked near the Fermi level, with a quite large value of 2.12 states/Co/spin/eV , which fulfills the Stoner criterion for ferromagnetism [160].

In [124], the sex+DMFT scheme, presented in section 4.2.3, has been applied to BaCo$_2$As$_2$, in order to see whether a more refined treatment of exchange and correlation effects results in a better agreement with the experimental situation. On theoretical grounds, we expect BaCo$_2$As$_2$ to be less correlated than the superconducting BaFe$_2$As$_2$, despite the very similar $U(\omega)$, because the nominal occupation of the d-orbital lattice site goes from the ideal spin-freezing value of 6 electrons in BaFe$_2$As$_2$ [122] to 7 in BaCo$_2$As$_2$. By moving farther away from half-filling, the onset of strong electron correlations is pushed up to larger values of $U$. Therefore, the electronic structure of BaCo$_2$As$_2$ is supposed to be characterized by coherent quasiparticles, in contrast to BaFe$_2$As$_2$, where coherence is lost very quickly as the temperature increases, particularly in the hole-doped compounds (see section 7.1.2).

ARPES data confirm this scenario [161], by reporting quite sharp low-energy spectra for BaCo$_2$As$_2$, typical of a weakly-to-moderately correlated material, as shown in figure 18. A peculiar feature of its band structure is represented by the $d_{x^2-y^2}$ band, which is very flat along the $\Gamma$–M direction just below the empty part of the spectrum. This band contributes to a quite tall
peak in the LDA density of states, exactly located at the Fermi level, and to the ferromagnetic nature of the LSDA solution. The situation encountered here is common in iron-based superconductors, where the tendency to magnetism is largely overestimated by density functional theory. In this sense, it is useful to perform a detailed study of BaCo$_2$As$_2$ as a benchmark system. The authors of [124] tested LDA$+$DMFT calculations of different flavors, namely LDA$+$DMFT with the cRPA static $U$, LDA$+$DMFT$+U(\omega)$, and SEX$+$DMFT$+U(\omega)$. The resulting spectral functions are shown in figure 18. It was found that the regular LDA$+$DMFT performs quite well, whereas the dynamic $U$ worsens the LDA$+$DMFT results particularly around the M$k$-point and below $-0.5$ eV, due to a too large band renormalization. On the other hand, these features are corrected by the SEX$+$DMFT, as the ZB band narrowing is compensated by nonlocal exchange effects included in the SEX part, which yields instead a band widening, as it is apparent in figure 19.

Therefore, both ingredients, the dynamically screened $U$ and the nonlocal screened exchange, are essential for a consistent description of this moderately correlated material. Electron correlations included in the non-perturbative solution of the impurity problem with retarded $U$ are key to the broadening of the spectral function peak at the Fermi level (figure 19), which results in a density of states below the Stoner threshold.

One thus recovers the correct paramagnetic phase in agreement with experimental conditions. SEX$+$DMFT is a promising scheme that should be applied to other strongly correlated compounds for further benchmarks and predictions.

7.3. GW$+$DMFT results for SrVO$_3$

7.3.1. Band widening from $k$-dependent $\Sigma$. The simplest way of combining GW and DMFT is to add the local self-energy of a converged LDA$+$DMFT$+U(\omega)$ calculation to the nonlocal self-energy from a separate GW calculation [147]. To avoid a double-counting of interaction effects, the local component of the GW self-energy must be subtracted. Figure 20 compares the resulting quasiparticle band structure of SrVO$_3$, obtained as the solution of the quasiparticle equation $E_{i,k} - \epsilon_{i,k} \sim \text{Re} \Sigma_{i,i}(k, E_{i,k}) = 0$, with $i$ the band index, to LDA, GW and LDA$+$DMFT$+U(\omega)$ data. The bandwidth of 1.2 eV predicted by the one-shot GW$+$DMFT scheme is in good agreement with photoemission data [143]. In particular, the one-shot GW$+$DMFT corrects the overestimation of the correlation effect in LDA$+$DMFT$+U(\omega)$ (section 7.1.1), while producing a substantial renormalization of the GW band structure, especially in the unoccupied part of the spectrum.
This is illustrated in the lower panel, where a rescaled GW quasiparticle band structure is compared to the one-shot GW+DMFT result. Note that the stronger renormalization of the bands in the unoccupied part is the result of the local self-energy, while the overall GW+DMFT band structure is wider than the DMFT one, because of the nonlocal self-energy components coming from the GW part.

It has been argued in [126] that the partial cancellation between the band-narrowing effect of a dynamically screened interaction (section 3.4) and the band-widening effect of the nonlocal self-energy contributions is a generic phenomenon, which highlights the advantage of formalisms such as GW+DMFT, which incorporate both effects. In fact, it is useful and instructive to analyze the band widening effect due to the nonlocal exchange at the GW level. A one-shot GW calculation of SrVO$_3$ starting from the LDA band structure and wavefunctions produces the $t_{2g}$ spectral function shown in the left panel of figure 21. The GW bandwidth is renormalized with respect to LDA, such that the effective mass is increased by a factor $m_{GW}/m_{LDA} = 1.3$. However, the quasiparticle weight $Z_k = 1/(1 - \partial_{\omega} \Re \Sigma(k, \omega) )$ is ~ 0.53, where the self-energy is defined with respect to the LDA exchange-correlation potential: $\Sigma = \Sigma_{GW} - V_{LDA}^c$. This would give a mass enhancement of ~ 2 in the absence of nonlocal self-energy effects. However, the expression for the group velocity within GW, from where the total mass enhancement is extracted, reads:

$$\frac{dE_{k\omega}}{dk} = \frac{\bra{\psi_k} \partial_{\omega} \{ H_{LDA}(k) + \Re \Sigma(k, \omega) \} \ket{\psi_k}}{(1 - \bra{\psi_k} \partial_{\omega} \Re \Sigma(k, \omega) \ket{\psi_k})}$$

which shows that the additional renormalization via the non-locality of the self-energy, $\partial_{\omega} \Re \Sigma(k, \omega)$, must be taken into account.

In order to illustrate and quantify the effect of the nonlocal components, one can remove the local part from $\Sigma_{GW}$ by defining $\Sigma_{GW}^{\text{nonloc}}(k, \omega) = \Sigma_{GW}(k, \omega) - \Sigma_{GW}^{\text{loc}}(k, \omega)$. The spectral function of the corresponding Green’s function $\tilde{G}(k, \omega) = \frac{1}{\omega + H_{LDA}(k) + V_{LDA}^c - \Sigma_{GW}^{\text{nonloc}}(k, \omega)}$ is plotted in the right panel of figure 21. It is apparent that nonlocal contributions, which mainly come from the exchange part, lead to a significant band widening, yielding a bandwidth ~ 1.5 times larger than LDA. We also note a small asymmetry in the band-widening, that is, the effect is more prominent in the empty part of the spectrum, as pointed out in [119] and [124].

Another noteworthy feature of the GW self-energy of SrVO$_3$ is the separation of the local and nonlocal self-energy effects into dynamic and static ones, respectively. A way to quantify the nonlocality of dynamical renormalizations is to compute the generalized $k$-dependent quasiparticle weight

$$Z_k(\omega) = \left[ 1 - \frac{\partial_{\omega} \Re \Sigma(k, \omega)}{\partial_{\omega}} \right]^{-1},$$

and its $k$-variations, defined by

$$\Delta k Z = \frac{\sqrt{\sum_k \text{Tr}[Z_k(\omega) - Z_k^{\text{loc}}(\omega)]^2}}{\sum_k \text{Tr}[Z_k(\omega)]},$$

where the local quantity is, as usual, $Z_k^{\text{loc}}(\omega) = \sum_k Z_k(\omega)$. $\Delta k Z$ and $Z_k^{\text{loc}}(0)$ are plotted in the bottom panel of figure 22. It turns out that the $k$-dispersion of $Z_k$ around $Z_k^{\text{loc}}$ is very weak in the frequency window from ~2 to 2 eV, signaling that dynamical effects are local at low-energy. Conversely, nonlocal effects are static in the same energy window. This is confirmed by the inspection of the real and imaginary parts of the GW self-energy, plotted for some selected high symmetry $k$-points in the top and middle panels of figure 22, respectively. One easily sees that at low energy the frequency dependence is $k$-insensitive in both the real and imaginary parts, while the $k$-dependence is $\omega$-independent in the real part, leading to a $k$-independent rigid shift of the self-energy curves. This effect has not only been observed in SrVO$_3$ [119, 162], but also in the iron pnictides and chalcogenides [125], and in the 3D Hubbard model [163].

We come back to the asymmetry between the occupied and empty parts of the spectrum, which results from the nonlocal part of the self-energy. This asymmetry is not only present in the band widening, as seen in figure 21, but most prominently in the quasiparticle lifetimes. Indeed, the imaginary part of the GW self-energy is largest on the unoccupied orbitals for $\omega > 0$, as shown in the middle panel of figure 22, which implies a stronger effect of electron–electron scattering in the empty part of the spectrum. The imaginary part of the self-energy keeps growing in a non-symmetric way even outside the energy window where the dynamic and static effects are separable in frequency. At larger frequencies it reaches a peak, which will produce $k$-dependent satellite features, stronger in the empty part of the spectrum.
7.3.2. Self-consistent calculation of $\Sigma$. The partially self-consistent implementation of [119, 121] goes a step beyond the one-shot calculation of figure 20 by performing a self-consistency loop for $\Sigma$ and $G$. However, compared to the full GW+DMFT scheme discussed in section 4.1.2, a number of approximations have been made. The bosonic Weiss field $U$ was kept frozen to the $(\omega = U_{R0})_{\text{cRPA}}$ value, as described in section 4.2.2. In other words, the self-consistency loop for $P$ and $W$ was not performed. However, the RPA polarization function computed from the LDA band structure of SrVO$_3$, and its corresponding dielectric function, compare favorably to electron energy loss spectroscopy measurements of SrTiO$_3$, an isostructural compound with a $d^0$ occupation, where data of this kind are available up to 40 eV.

The results for the GW and the GW+DMFT spectra are shown in figure 23. The dynamically screened impurity problem of GW+DMFT has been solved by DALA. Compared to LDA, the GW spectral function, plotted in the leftmost panel of figure 23, gives a better position of the O-2$p$ and Sr-4$d$ states, which are closer to the experimental photoemission spectroscopy (PES) and Bremsstrahl-Isochromat spectroscopy (BIS) curves taken from [145] and [164]. However, GW yields a too strong quasiparticle peak at the Fermi level, of full $t_{2g}$ character, and a too weak mass enhancement. GW+DMFT, reported in the middle panel of figure 23, corrects for these deficiencies, by strongly renormalizing the height of this peak, and producing a mass enhancement of $\sim 2$. The renormalization is accompanied by a spectral weight transfer from the quasiparticle peak to the lower Hubbard band, correctly located at $-1.6$ eV (see rightmost panel in figure 23), and to plasmon satellites, which can be identified in the figure at $-4$ eV, 5 eV, and 15 eV. Therefore, dynamic screening and correlation effects play a major role in determining the renormalized low-energy properties of the material.

Surprisingly, the upper Hubbard band is not visible in the local GW+DMFT spectral function computed in [119, 121]. It may be hidden by nonlocal self-energy effects, particularly large in the empty part of the spectrum. However, other calculations, such as the one-shot GW+DMFT scheme of [147], or the simplified SEX+DMFT [162], both using a CTQMC solver, show a quite sharp upper Hubbard band well separated from the quasiparticle peak. The missing Hubbard band could thus also be a consequence of an underestimation of the bandwidth reduction by DALA with respect to QMC (reported in [141]), or by nonlocal dynamical self-energy effects just above the Fermi liquid regime, which are neglected in [162] and treated non-self-consistently in [147].

In any case, within the partially self-consistent GW+DMFT scheme of [119, 121] the upper Hubbard band should be located around 2 eV, as its separation from the lower one must be the static value of $U$ (here 3.3 eV). While it is washed out due to the strong dispersion of the quasiparticle, it can indeed be identified in this energy region in the $k$-resolved...
spectra. This implies that the BIS signal detected at 2.7 eV is not the upper Hubbard band of the $t_{2g}$ manifold, as previously claimed. Overlapping the GW+DMFT spectrum with the BIS data shows that the peak just above the Fermi level has an $e_g$ origin, instead, as seen in figure 23.

7.3.3. Self-consistent calculation of $\Sigma$ and $P$. The first fully self-consistent GW+DMFT results for SrVO$_3$ have been presented very recently in [120]. These simulations are based on the multi-tier approach introduced in section 4.2.1. Starting from a LDA and $G_0W^0$ calculation for a large energy window containing 200 bands, the self-energy and polarization functions projected onto the low-energy $t_{2g}$ subspace (the ‘intermediate subspace’ $I$ in equations (112) and (113)) were calculated and used to define an effective bare propagator $G_{t_2g,k}^0$ and bare interaction $U_{t_2g,q}$ for this subspace. With this $ab\ initio$ input, a self-consistent GW+EDMFT calculation (analogous to section 4.1.2) was performed in the $t_{2g}$ subspace. In the calculation of $\Sigma$, the local components of $\Sigma_{t_{2g,k}}^{G_0W}$ were replaced by the more accurate estimates from the EDMFT impurity problem, $\Sigma_{t_{2g,k},loc}^{EDMFT}$ (see equation (112)), while in the calculation of $P$, the local components of $P_{t_{2g,k}}^{GG}$ were replaced by $P_{t_{2g,k},loc}^{EDMFT}$ (see equation (112)).

The remarkable outcome of this hitherto most sophisticated $ab\ initio$ simulation of SrVO$_3$ is that the full self-consistency within the $t_{2g}$ subspace leads to a substantial reduction of the effective local interaction. As a result of nonlocal screening processes within this subspace, the fully screened interaction is reduced below an energy of about 2 eV in a way which resembles the $G_0W^0$ result. The imaginary part of $W$ (distribution of screening modes) for $G_0W^0$, self-consistent GW and self-consistent GW+DMFT is shown in the top panel of figure 24. In the EDMFT impurity problem, these screening processes lead to a substantial reduction of the static impurity interaction $\mathbf{U}(0)$ from the cRPA value of 3.4 eV to about 2.2 eV.

The implications are profound: in contrast to all the previous schemes discussed in this review, the self-consistent GW+DMFT results are incompatible with a Hubbard band interpretation of the satellite features in the $t_{2g}$ spectral function of SrVO$_3$. Instead, these satellites are naturally explained as plasmon sidebands of the renormalized quasiparticle band [120], as illustrated in the lower panel of figure 24, which shows the local $t_{2g}$ spectral function obtained from $G_0W^0$, self-consistent GW and self-consistent GW+DMFT. The addition of the nonperturbative EDMFT impurity self-energy enables meaningful self-consistent calculations, which in the end produce a result that is rather close to $G_0W^0$ (albeit with more pronounced satellites whose positions are in better agreement with available photoemission and inverse photoemission experiments [143, 145, 164]). This is an interesting observation in connection with the empirical fact that $G_0W^0$ often

![Figure 23](image1.png)

**Figure 23.** GW (left panel) and GW+DMFT (middle panel) spectral functions of SrVO$_3$ in comparison to photoemission and inverse photoemission spectra. Right panel: zoom of the middle panel around the lower Hubbard band region. (Reproduced with permission from [119].)

![Figure 24](image2.png)

**Figure 24.** Top panel: imaginary part of the screened interaction $W$ for SrVO$_3$ in the $G_0W^0$, self-consistent GW and self-consistent GW+DMFT (denoted GW+EDMFT) approximation ($t_{2g}$ subspace, inverse temperature 1/15 eV). Lower panel: local $t_{2g}$ spectral functions from the same approximations. The static value $\mathbf{U}(0)$ of the EDMFT impurity interaction is comparable to the width of the quasiparticle band. The satellites are split off by the plasmon energy $\omega_p$ associated with collective excitations of the $t_{2g}$ electrons. (Reproduced with permission from [120].)
yields a much more accurate electronic structure than self-consistent GW [2].

At the very least, this recent work emphasizes the importance of a self-consistent treatment of screening and interactions. It will be interesting to test the self-consistent GW+EDMFT scheme on broad classes of materials, to judge its accuracy, and to gain a deeper understanding of the effect of screening processes within the low-energy model subspace. If the large effect on local correlations demonstrated in [120] turns out to be generic for metallic systems, then a potentially large number of previous LDA+DMFT+\(U(\omega)\) results will have to be reconsidered.

8. Conclusions and outlook

We have discussed the effect of dynamical screening in strongly correlated lattice systems and materials, and detailed some recently developed or implemented techniques based on extensions of the DMFT framework. At this point, the most advanced scheme which allows a self-consistent treatment of screening and correlation effects in materials is the combination of the GW ab initio method and DMFT [16]. A fully self-consistent GW+DMFT calculation yields retarded interaction parameters for the DMFT impurity problem, which incorporate the effect of local and nonlocal screening processes in the correlated system. This method also captures some extent the momentum dependence of the self-energy, and hence the competition between the band-narrowing effect of the frequency-dependent interaction and the band-widening effect of the nonlocal self-energy components. Self-consistent GW+DMFT calculations of multi-band systems and realistic materials have only been implemented very recently [120]. While these results demonstrate large nonlocal screening effects, more work on a broad range of materials is needed to judge the reliability and predictive power of the scheme. Also, it should be emphasized that in current implementations, the self-consistent calculation is performed only within a low-energy subspace containing just a few bands, after a cRPA downfolding or one-shot GW calculation. In the near future, this scheme should be extended to a multi-tier approach with a larger intermediate subspace for the self-consistent GW calculations. In such a scheme, the GW+DMFT estimate of the self-energy of the low-energy space has to be merged with the GW self-energy for the remaining orbitals in the intermediate subspace, and the calculations within the intermediate and low-energy space should be iterated until a self-consistent solution for \(G, W\), and the auxiliary impurity problem is obtained.

Apart from the implementation of these hopefully accurate, but numerically demanding schemes, the further exploration of simplified versions of the GW+DMFT framework will produce useful insights into the effects of screening and nonlocal correlations. We have discussed the partially self-consistent formulation of Tomczak et al., which performs a self-consistency loop on \(\Sigma\) and \(G\), while approximating the effective interaction by the cRPA estimate [119]. If the self-energy is furthermore approximated as local, one ends up with the LDA+DMFT+\(U(\omega)\) method, which has been used in recent years by several groups to study plasmon satellites in strongly correlated materials [122, 141, 142, 147]. Another recently proposed variant, which has been applied to correlated materials with promising results, is SEX+DMFT+\(U(\omega)\), which takes into account the band-widening effect of the nonlocal screened exchange [124]. Further applications and comparisons of these different simplified schemes will provide valuable insights into the roles played by the different ingredients of the GW+DMFT formalism, and the importance of self-consistency.

It is also essential to clarify the accuracy and limitations of the cRPA method, or related RPA schemes, which underpin the material simulations discussed in this review. The RPA polarization lacks vertex corrections, and in principle, a downfolding onto a low-energy subspace should generate higher-order interaction terms which are not contained in an effective Hubbard model description. Some of these issues have recently been addressed in simple model set-ups containing a small number of screening bands [58, 59], but more work is needed to properly judge the realistic situation with a large number of high-energy screening bands.

It should be kept in mind that GW+DMFT and related methods, which combine the local self-energy from an effective impurity model with the nonlocal self-energy of some weak-coupling perturbation theory, cannot be expected to capture the strong nonlocal correlation effects in low dimensional systems. This has been explicitly demonstrated for the two-dimensional Hubbard model in the weak-coupling regime [28], while at intermediate coupling, there are obvious inconsistencies between the weakly momentum-dependent GW+DMFT results, and the strong momentum-variation predicted by cluster DMFT simulations [24, 165]. Hence, we should view GW+DMFT primarily as a method which is suitable for capturing dynamical screening effects in three dimensional compounds. For two-dimensional systems, the combination of (extended) cluster DMFT and many-body perturbation theory may be a promising strategy. However, efficient impurity solvers capable of handling dynamically screened interactions in cluster impurity problems have yet to be developed.

In fact, progress in this field is intricately linked to further improvements of the available impurity solvers and analytical continuation techniques. The methods reviewed in this paper enable an efficient simulation of single-site, multi-orbital impurity problems with dynamically screened monopole interactions. Retarded spin-flip terms cannot be handled efficiently with methods based on a Lang–Firsov decoupling of the electron-boson interaction and must be dealt with in a double-expansion approach [108, 166], which may suffer from a sign problem. The analytical continuation procedure explained in section 5 is also based on a Lang–Firsov picture, and works reliably only in systems with a clear energy separation between low-energy and satellite features. More flexible and powerful methods will be essential in particular for the eventual application of more advanced schemes, such as TRILEX [167], dual bosons [113], or extended cluster methods to realistic materials.
Another new and interesting research direction is the extension of the methods described here to nonequilibrium systems [137, 138]. A proper description of dynamical screening effects should be very important in materials perturbed by intense laser fields, especially if the laser pulse produces inter-band transitions. GW+DMFT is a promising starting point for the development of a formalism, which enables \textit{ab initio} predictions of nonequilibrium phenomena in strongly correlated materials.

Acknowledgments

We would like to thank F Aryasetiawan, T Ayral, S Biermann, L Boehnke, V Brouet, M Eckstein, D Golez, J Gukulberger, L Huang, A JMillis, F Nilsson, T Miyake, Y Nomura, R Sakuma, A van Roekeghem, H Shinaoka, J Tomczak, and L Vaugier for helpful discussions and close collaborations on the topics covered in this review. PW acknowledges support from ERC starting grant 278023, DFG FOR 1346 (SNSF Grant 200021-149122), SNSF Grant 200021-140648 and NCCR Marvel. Some of the results presented in this review were obtained using HPC resources from GENCI 096493 allocation.

References

[1] Hedlin L 1965 Phys. Rev. 139 A796
[2] Aryasetiawan F and Gunnarsson O 1998 Rep. Prog. Phys. 61 237
[3] Hedlin L 1965 J. Phys.: Condens. Matter 11 R489
[4] Onida G, Reining L and Rubio A 2002 Rev. Mod. Phys. 74 601
[5] Jing S P, Lad R J and Henrich V E 1991 Phys. Rev. B 43 11971
[6] Martlon L, Simpson J A, Fowler H A and Swanson N 1962 Phys. Rev. 126 182
[7] Aryasetiawan F, Hedlin L and Karlsson K 1996 Phys. Rev. Lett. 77 2268
[8] Guzzo M, Lani G, Sottile F, Romaniello P, Gatti M, Kas J J, Rehr J J, Silly M G, Sirotti F and Reining L 2011 Phys. Rev. Lett. 107 166401
[9] Metzner W and Vollhardt D 1989 Phys. Rev. Lett. 62 324
[10] Georges A and Kotliar G 1992 Phys. Rev. B 45 6479
[11] Georges A, Kotliar G, Krauth W and Rozenberg M J 1996 Rev. Mod. Phys. 68 13
[12] Sengupta A M and Georges A 1995 Phys. Rev. B 52 10295
[13] Si Q and Smith J L 1996 Phys. Rev. Lett. 77 3391
[14] Sun P and Kotliar G 2002 Phys. Rev. B 66 085120
[15] Gull E, Millis A J, Lichtenstein A I, Rubtsov A N, Troyer M and Werner P 2011 Rev. Mod. Phys. 83 349
[16] Biermann S, Aryasetiawan F and Georges A 2003 Phys. Rev. Lett. 90 086402
[17] Zgid D and Chan G K L 2011 J. Chem. Phys. 134 094115
[18] Kananenka A A, Gull E and Zgid D 2015 Phys. Rev. B 91 121111
[19] Lar N T, Kananenka A A and Zgid D 2015 J. Chem. Phys. 143 214102
[20] Zgid D, Gull E and Chan G K L 2012 Phys. Rev. B 86 165128
[21] Go A and Millis A J 2015 Phys. Rev. Lett. 114 016402
[22] Maier T, Jarrell M, Pruschke T and Hettler M H 2005 Rev. Mod. Phys. 77 1027
[23] Tocchio L F, Becca F, Parola A and Sorella S 2008 Phys. Rev. B 78 041101
[24] Werner P, Gull E, Parcollet O and Millis A J 2009 Phys. Rev. B 80 045120
[25] Gull E, Parcollet O, Werner P and Millis A J 2009 Phys. Rev. B 80 245102
[26] Chang C C and Zhang S 2010 Phys. Rev. Lett. 104 116402
[27] Shi H and Zhang S 2013 Phys. Rev. B 88 125132
[28] Guikelberger J, Huang L and Werner P 2015 Phys. Rev. B 91 235114
[29] Zheng B X and Chan G K L 2015 2016 Phys. Rev. B 93 035126
[30] LeBlanc J P F et al 2015 Phys. Rev. X 5 031041
[31] Rubtsov A N, Katsnelson M I and Lichtenstein A I 2008 Phys. Rev. B 77 033101
[32] Toschi A, Katanin A A and Held K 2007 Phys. Rev. B 75 045118
[33] Imada M and Miyake T 2010 J. Phys. Soc. Japan 79 112001
[34] Hirayama M, Miyake T and Imada M 2013 Phys. Rev. B 87 195144
[35] Hohenberg P and Kohn W 1964 Phys. Rev. 136 B864
[36] Kohn W and Sham L J 1965 Phys. Rev. 140 A1133
[37] Jones R O 2015 Rev. Mod. Phys. 87 897
[38] Löwdin P O 1951 J. Chem. Phys. 19 1396
[39] Löwdin P O 1955 Phys. Rev. 97 1474
[40] Lambrecht W R L and Andersen O K 1986 Phys. Rev. B 34 2439
[41] Marzari N and Vanderbilt D 1997 Phys. Rev. B 56 12847
[42] Miyake T and Aryasetiawan F 2008 Phys. Rev. B 77 085122
[43] Aryasetiawan F, Imada M, Georges A, Kotliar G, Biermann S and Lichtenstein A I 2004 Phys. Rev. B 70 195104
[44] Aryasetiawan F, Karlsson K, Jepsen O and Schönenberger U 2006 Phys. Rev. B 74 125106
[45] Springer M and Aryasetiawan F 1998 Phys. Rev. B 57 4364
[46] Çağaloglu E, Friedrich C and Blügel S 2011 Phys. Rev. B 83 121101
[47] Sakuma R and Aryasetiawan F private communication
[48] Miyake T, Pouroufski L, Vildosola V, Biermann S and Georges A 2008 J. Phys. Soc. Japan 77 99
[49] Nakamura K, Arita R and Imada M 2008 J. Phys. Soc. Japan 77 093711
[50] Tomczak J M, Miyake T, Sakuma R and Aryasetiawan F 2009 Phys. Rev. B 79 235133
[51] Miyake T, Nakamura K, Arita R and Imada M 2010 J. Phys. Soc. Japan 79 044705
[52] Tomczak J M, Miyake T and Aryasetiawan F 2010 Phys. Rev. B 81 115116
[53] Nomura Y, Nakamura K and Arita R 2012 Phys. Rev. B 85 155452
[54] Vaugier L, Jiang H and Biermann S 2012 Phys. Rev. B 86 165105
[55] Sakuma R and Aryasetiawan F 2013 Phys. Rev. B 87 165118
[56] Souza I, Marzari N and Vanderbilt D 2001 Phys. Rev. B 65 035109
[57] Miyake T, Aryasetiawan F and Imada M 2009 Phys. Rev. B 80 155134
[58] Kinza M and Honerkamp C 2015 Phys. Rev. B 92 045113
[59] Shinaoka H, Troyer M and Werner P 2015 Phys. Rev. B 91 245156
[60] Cococcioni M and de Gironcoli S 2005 Phys. Rev. B 71 035105
[61] Hsu H, Umemoto K, Cococcioni M and Wentzcovitch R 2009 Phys. Rev. B 79 125124
[62] Anisimov I V and Gunnarsson O 1991 Phys. Rev. B 43 7570
[63] Anisimov I V, Korotin D M, Korotin M A, Kozhevnikov A V, Kunes J, Shorikov A O, Skornyakov S L and Streltsov S V J. Phys.: Condens. Matter 21 075602
[64] Kutepov A, Haule K, Savrasov S Y and Kotliar G 2010 Phys. Rev. B 82 045105
