State and superstate sampling in hybridization-expansion continuous-time quantum Monte Carlo

Alexander Kowalski,[1] Andreas Hausoel,[1] Markus Wallerberger,[2] Patrik Gunacker,[3] and Giorgio Sangiovanni[1]

[1]*Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany*
[2]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[3]*Institute of Solid State Physics, TU Wien, 1040 Vienna, Austria*

Due to the intrinsic complexity of the quantum many-body problem, quantum Monte Carlo algorithms and their corresponding Monte Carlo configurations can be defined in various ways. Configurations corresponding to few Feynman diagrams often lead to severe sign problems. On the other hand, computing the configuration weight becomes numerically expensive in the opposite limit in which many diagrams are grouped together. Here we show that for continuous-time quantum Monte Carlo in the hybridization expansion the efficiency can be substantially improved by dividing the local impurity trace into fragments, which are then sampled individually. For this technique, which also turns out to preserve the fermionic sign, a modified update strategy is introduced in order to ensure ergodicity. Our (super)state sampling is particularly beneficial to calculations with many d orbitals and general local interactions, such as full Coulomb interaction. For illustration, we reconsider the simple albeit well-known case of a degenerate three-orbital model at low temperatures. This allows us to quantify the coherence properties of the "spin-freezing" crossover, even close to the Mott transition.

I. INTRODUCTION

Continuous-time quantum Monte Carlo algorithms are state-of-the-art, numerically exact methods for the solution of the Anderson impurity model (AIM) [1–4]. These are widely used for the description of the physics of magnetic impurities, Kondo systems, transport through quantum junctions and are also employed as auxiliary models in dynamical mean field theory (DMFT) calculations for lattice models of correlated electron systems. Several high-level open source implementations of DMFT and of its merger with density functional theory have been recently made available [5–9].

One of the most successful flavors of continuous-time quantum Monte Carlo algorithms is the strong-coupling hybridization expansion (CT-HYB) [10]. CT-HYB is the method of choice for multiorbital impurity models with general interactions because one observes only a moderate sign problem provided that the bath problem has sufficient symmetry. This is because CT-HYB splits each Monte Carlo configuration into a noninteracting bath part and a fully interacting impurity part and solves the impurity part using an exact diagonalization/FullCI-type method. However, the dimension of the impurity Hamiltonian grows exponentially with the number of orbitals and so does the computational effort with it. In practice, correlated d or f shells as well as small correlated molecules can be treated with CT-HYB.

Yet, reaching low temperatures is still challenging due to the quadratic scaling of the impurity problem with inverse temperature. This follows from the fact that the mean order of diagrammatic expansion grows linearly with inverse temperature, and both the computational cost of evaluating a single configuration as well as the observed autocorrelation time between configurations scale linearly with expansion order.

While the exponential scaling with the number of orbitals and the quadratic scaling with the inverse temperature are intrinsic to the local problem, potentially model-dependent improvements to the prefactor of this overall scaling can be achieved.

Common approaches to such optimization are block diagonalization of the local Hamiltonian using conserved quantities [3,11] and binning, tree [4], or equivalent [12] algorithms in so-called "matrix-matrix" implementations of CT-HYB. Additionally, with a similar motivation as for our method, outer truncation of the local trace to the few dominant contributions and calculation of those with more efficient sparse-matrix methods has been applied particularly to large systems at low temperatures [13]. Other more advanced strategies are local updates in imaginary time [14], a fast-rejection/acceptance algorithm by calculating upper/lower boundaries of the weight [12], or a partial summation of diagrams to extract more information out of one Monte Carlo configuration [15].

Here we consider a matrix-vector version of the CT-HYB algorithm as implemented in the *w2dynamics* package [9] and investigate the possibility of sampling the sum over the eigenstates of the local impurity in the Monte Carlo simulation. A hard outer truncation of high energy states (so far typically used in calculations with *w2dynamics*) constitutes an approximation and it is unclear whether it retains ergodicity. The approach proposed here, instead, is numerically exact and furthermore exceeds the performance benefits of hard truncation substantially.

We formulate two versions: the "superstate" -sampling algorithm, where states grouped by the blocks of the Hamiltonian are sampled together, and the "state" -sampling algorithm, where each many-body state of the impurity is sampled

individually. Conceptually, these methods can be interpreted as an equivalent of the segment algorithm for general interactions. Our improvements touch the core of the exponential scaling of CT-HYB and manage to significantly reduce the computation time of a Monte Carlo weight. Furthermore, they are in principle compatible with all of the other above-mentioned algorithmic improvements. Using a five-orbital AIM with the most general form of the electron-electron Coulomb interaction as an example, we achieve speed-up factors verging on three orders of magnitude.

First, we review the basic formulas of CT-HYB in Sec. II. In Secs. III and IV, the superstate and state sampling methods are introduced. In Sec. V, we comment on the performance and the average sign, while in Sec. VI we demonstrate the capabilities of the (super)state sampling algorithm with a simple physical example.

II. HYBRIDIZATION EXPANSION

Let us start with a brief review of the hybridization-expansion continuous-time quantum Monte Carlo algorithm, focusing on the formulas needed to explain our changes. For a complete introduction to this method, see Ref. [4]. We are concerned with the solution of a multiorbital Anderson impurity model, whose Hamiltonian can be written as

$$H = H_{\text{loc}}[d, d^{\dagger}] + \sum_{p\lambda}(V_{p\lambda}f_{p}^{\dagger}d_{\lambda} + \text{H.c.}) + \sum_{p}\tilde{E}_{p}f_{p}^{\dagger}f_{p}, \quad (1)$$

where d_{λ} annihilates a fermion on the impurity, which consists of spin orbitals $\lambda \in \{1, \ldots, 2N_{\text{orb}}\}$, and f_{p} annihilates a fermion on the bath, where the quantum number p can be continuous. $V_{p\lambda}$ and $\tilde{E}_{p}$ parametrize the hybridization and bath levels, respectively, while H_{loc} is a generic interacting local (impurity) Hamiltonian.

The expansion of the partition function $Z = \text{Tr}\,e^{-\beta H}$ in terms of the bath hybridization can be written as [4]:

$$Z = Z_{0}\sum_{k=0}^{\infty}\sum_{\lambda_{1},\lambda_{1}'}\int_{0}^{\beta}d\tau_{1}\int_{0}^{\beta}d\tau_{1}'\cdots\sum_{\lambda_{k},\lambda_{k}'}\int_{\tau_{k-1}}^{\beta}d\tau_{k}\int_{\tau_{k-1}'}^{\beta}d\tau_{k}'$$

$$\times \text{Tr}\left[T_{\tau}e^{-\beta H_{\text{loc}}}\prod_{i=1}^{k}d_{\lambda_{i}}^{\dagger}(\tau_{i})\,d_{\lambda_{i}'}(\tau_{i}')\right]\det[\Delta_{\lambda_{i}\lambda_{j}'}(\tau_{i} - \tau_{j}')]_{ij}, \tag{2}$$

where Z_{0} is the partition function for the bath part, $\beta = 1/T$ is the inverse temperature, T_{τ} denotes path ordering in imaginary time, and $\Delta(\tau) = V^{\dagger}(\partial_{\tau} - \tilde{E})^{-1}V$ is the hybridization function, which encodes the total retardation effect of the bath on the local fermions.

Schematically, Eq. (2) can be written as follows:

$$Z = \sum_{\mathcal{C}}w_{\text{loc}}(\mathcal{C})w_{\text{bath}}(\mathcal{C}), \tag{3}$$

where $\mathcal{C} := (\tau_{1}, \tau_{1}', \lambda_{1}, \lambda_{1}', \ldots, \tau_{k}, \tau_{k}', \lambda_{k}, \lambda_{k}')$ denotes an infinitesimal term in the expansion, i.e.,

$$\sum_{\mathcal{C}} := \sum_{k=0}^{\infty}\sum_{\lambda_{1},\lambda_{1}'}\int_{0}^{\beta}d\tau_{1}\int_{0}^{\beta}d\tau_{1}'\cdots\sum_{\lambda_{k},\lambda_{k}'}\int_{\tau_{k-1}}^{\beta}d\tau_{k}\int_{\tau_{k-1}'}^{\beta}d\tau_{k}'. \tag{4}$$

In the conventional continuous-time hybridization expansion quantum Monte Carlo (CT-HYB) algorithm, each $\mathcal{C}$ is taken as a Monte Carlo configuration, and the sum (4) is performed using Markov chain Monte Carlo:

$$Z = \underbrace{\sum_{\mathcal{C}}}_{\text{QMC}}w_{\text{loc}}(\mathcal{C})w_{\text{bath}}(\mathcal{C}). \tag{5}$$

With $w_{\text{bath}} = \det[\ldots]$ we denote the bath part, corresponding to a determinant of noninteracting hybridization functions, which can be computed in $\mathcal{O}(k^{3})$ and updated in $\mathcal{O}(k^{2})$ time. The quantity

$$w_{\text{loc}}(\mathcal{C}) = \sum_{s}\langle s|\hat{\mathcal{C}}|s\rangle \tag{6}$$

$$:= \sum_{s}\langle s|T_{\tau}e^{-\beta H_{\text{loc}}}\prod_{i=1}^{k}d_{\lambda_{i}}^{\dagger}(\tau_{i})d_{\lambda_{i}'}(\tau_{i}')|s\rangle \tag{7}$$

is the local weight or local trace of a configuration, where s indexes the $4^{N_{\text{orb}}}$ many-body eigenstates of the local impurity Hamiltonian H_{loc}, and we have introduced the shorthand $\hat{\mathcal{C}}$ for the sequence of local operators of the current configuration. A naïve implementation of Eq. (7) involves k multiplications of $4^{N_{\text{orb}}} \times 4^{N_{\text{orb}}}$ matrices, which scales as $\mathcal{O}(k\exp(\alpha N_{\text{orb}}))$ with a constant α. Reducing the computational impact of the calculation of $w_{\text{loc}}(\mathcal{C})$ is thus usually the main objective of optimizing CT-HYB codes.

In general, the local Hamiltonian H_{loc} conserves a set of quantum numbers. Consequently, the many-body Hilbert space can be partitioned into a set of linear subspaces $\{\mathcal{S}\}$, so-called "superstates" [3], and the Hamiltonian can be brought into a block-diagonal form with respect to these superstates. We can write Eq. (6) as:

$$w_{\text{loc}}(\mathcal{C}) = \sum_{\mathcal{S}}w_{\text{loc},\mathcal{S}}(\mathcal{C}) := \sum_{\mathcal{S}}\sum_{s\in\mathcal{S}}\langle s|\hat{\mathcal{C}}|s\rangle. \tag{8}$$

In defining the quantum numbers, we impose the additional requirement that the impurity operators do not take a state from one superstate to more than one other superstate, thereby possibly merging multiple blocks of the Hamiltonian into one superstate. This implies that $w_{\text{loc},\mathcal{S}}$ can be calculated independently for each $\mathcal{S}$. The scaling is now controlled by the size of the largest superstate (in the worst case), i.e., by something much smaller than $4^{N_{\text{orb}}}$. Since the application of an impurity operator corresponds to a one-to-one mapping between different superstates (and giving zero if it violates the Pauli principle), a further optimization is possible: For each superstate $\mathcal{S}$, one can follow the sequence of superstates by using this mapping starting with $\mathcal{S}$ at $\tau = 0$ until reaching $\tau = \beta$. If one reaches zero at any point or ends up in a different superstate at the other end, $w_{\text{loc},\mathcal{S}}$ is exactly zero and does not need to be calculated using possibly much more costly linear algebra. We will refer to this procedure as *quantum number checking* in the following sections. This concludes our overview of what we refer to as conventional CT-HYB method.

III. SUPERSTATE SAMPLING

A. General description

The main idea of this work is to transform the deterministic summation over the eigenstates of the impurity in Eq. (6) into a stochastic summation. The original Monte Carlo configuration is split into many "smaller" weights. While having been proposed [16], it has never been implemented to the best of our knowledge.

We focus on two strategies in particular, one that partitions the sum into subsets by quantum numbers and one that breaks it up entirely. We call them

(1) "Superstate sampling": the summation over all superstates $\mathcal{S}$ is now done by Monte Carlo sampling:

$$Z = \underbrace{\sum_{\mathcal{C}} \sum_{\mathcal{S}}}_{QMC} \sum_{s \in \mathcal{S}} \langle s|\hat{\mathcal{C}}|s\rangle w_{\text{bath}}(\mathcal{C}), \qquad (9)$$

where each Monte Carlo configuration now contains the trace over all states s within a superstate.

(2) "State sampling": the summation over all states s in Eq. (8) is now done by Monte Carlo sampling:

$$Z = \underbrace{\sum_{\mathcal{C}} \sum_{\mathcal{S}} \sum_{s \in \mathcal{S}}}_{QMC} \langle s|\hat{\mathcal{C}}|s\rangle w_{\text{bath}}(\mathcal{C}), \qquad (10)$$

where each Monte Carlo configuration now contains the local configuration evaluated for a single outer state s.

In this section, we will focus on superstate sampling, while state sampling will be discussed in Sec. IV.

The fragmentation of the sum reduces the amount of calculations needed for one local weight and allows us to move faster through phase space. It is thus particularly beneficial in systems with low symmetry, which can have many superstates with small but nonzero contributions to the local weight. On the other hand, if many quantum numbers can be used in a calculation with the conventional sampling, an increase in β results in an effective reduction of the number of possible outer superstates. This "help" is a side effect of the large number of operators in the trace present in the low-T limit. It results in a very high chance of quantum number violation and it hence substantially restricts the room for maneuver for the outer superstates. The advantage of superstate sampling is therefore twofold: at any temperature an easy and natural selection of the most important outer superstates and much less need for quantum number checking, particularly beneficial at low T [17].

By sampling superstates, we are sampling a sum of terms with potentially different signs. This may induce a sign problem, which would in general be expected to worsen exponentially with decreasing temperature. (This is why it is important to combine all possible bath configurations into a bath determinant in CT-HYB [2].) Yet, we do not observe any worsening of the average sign in superstate sampling compared to the original algorithm (cf. Sec. V). A heuristic argument for this can be summarized thusly: Since the mean expansion order grows linearly with the inverse temperature β [2], the average number of superstates that violate the Pauli principle increases, until at a certain β, we are often
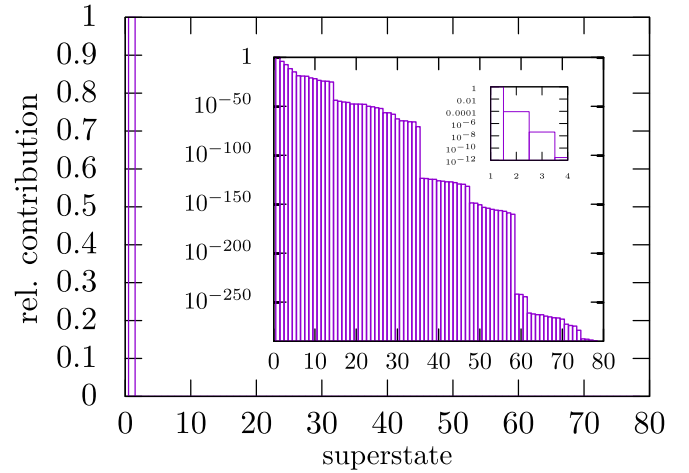

FIG. 1. Average relative contributions of outer superstates to the local weight per configuration for a typical simulation with five orbitals and Kanamori interaction [18]. For comparability, the superstates are ordered by their contribution (i.e., absolute value of the part of the local weight sum from all states contained in the superstate) for each individual configuration, i.e., superstate "1" does not denote one specific constant superstate but always refers to the biggest contributor.

left with only one outer superstate. For example, we have observed that for a typical metallic systems, a temperature of the order of 10^{-2} of the electronic bandwidth is about the point where many configurations have only a single superstate contributing to the trace. At such low temperatures, the local weight in conventional sampling is the sum over only one outer superstate, so switching to superstate sampling should not affect the sign.

When multiple superstates contribute, such as in simulations at high temperatures, the switch to superstate sampling could in principle cause a difference depending on the superstates' relative weight and relative sign. To illustrate the typical superstate weight distribution in such cases, we consider relatively high-temperature simulations. For a five-orbital model with Kanamori interaction [18], we show the average distribution of the local weight of a configuration onto the outer superstates in Fig. 1. It is clear how the local weight of each configuration is strongly dominated by the contribution of one superstate. Similar results can be obtained for a simpler two-orbital model. We can therefore expect the method to be useful for high temperatures as well, as it allows us to sample configurations with their "ideal" outer superstates.

B. Sampling and ergodicity

Since we extend the configurations in the superstate sampling method by an outer superstate, the simulation must be able to reach every configuration with nonzero weight independent of its outer superstate to preserve ergodicity.

a. Inner pair moves. While the possibility to change the outer superstate needs to be available for ergodicity, we observe that it only needs to be done comparatively rarely in the simulation. Therefore, we do not change the outer superstate when inserting or removing a pair of operators, which are
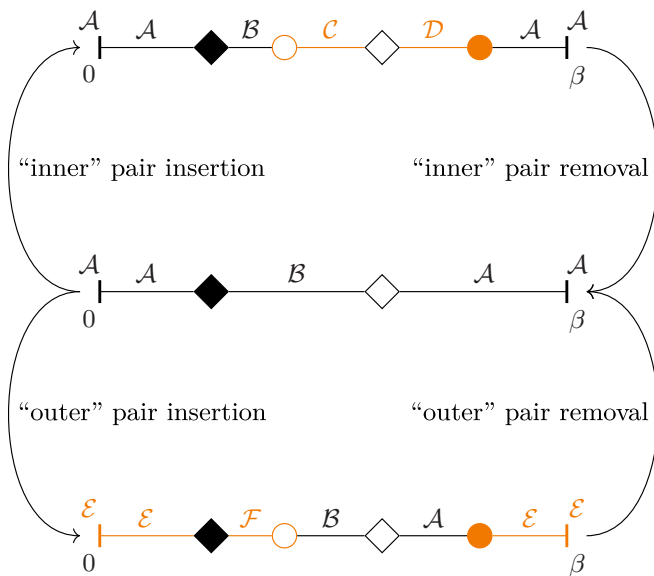
FIG. 2. Three configurations represented as line segments with calligraphic letters denoting the superstates. An "inner" ("outer") pair move is shown in the upper (lower) part of the figure. The impurity operators are represented by symbols, whose filledness tells whether they are creators or annihilators. The imaginary times of the operators is given by their position along the segment, which represents the imaginary time axis. The letter above the $\tau = 0$ mark is the outer superstate of the configuration.

the most common moves proposed. We call this variant the "inner" pair moves: Because we fix the outer superstate, only states that lie between the inserted or removed operators in imaginary time can change and need to be recalculated.

Consider the example of the "inner" pair insertion move shown in the upper half in Fig. 2: The old configuration $\mathcal{C}_{\text{old}}$ is the one shown in the middle panel. We want to perform an "inner" pair move that inserts the two orange operators with random times and flavors in the top diagram, which represents the resulting configuration $\mathcal{C}_{\text{inner}}$. In this most commonly proposed type of move, new configurations $\mathcal{C}_{\text{inner}}$ are only proposed with the same outer superstate as the old configuration $\mathcal{C}_{\text{old}}$. The superstate sequence between the two new orange operators is new whereas the part outside them remains as in the old configuration. These moves are in a sense most closely connected to the pair moves of conventional sampling as the outer superstates with the biggest contributions are not likely to change in local moves.

b. Outer pair moves. In the other local pair moves we consider, the "outer" moves, the prescription for the superstate sequence is the opposite. The superstates between the inserted operators are to be left unchanged, and the sequence must be continued from there to $\tau = 0$ to determine the superstate that should be used as new outer superstate. An example for an outer insertion is the move from the configuration $\mathcal{C}_{\text{old}}$ in the middle of Fig. 2 to $\mathcal{C}_{\text{outer}}$ on the bottom, where the inserted operators are the same as in the inner move example for easy comparison. Since we fix the inner part of the superstate sequence, the entire part "outside" of the orange operators changes.

The acceptance of outer pair moves is in practice however significantly smaller than that of the "inner" pair moves. This can be understood thinking about the limit of local-in-τ moves: These, in order to be considered local in the "outer" case, are subject to the additional constraint of having the two operators at opposite ends of the trace. In the next subsection, we discuss a more efficient way to ensure ergodicity with respect to the outer superstate, the so-called global τ-shift move. We will also show in Appendix A that the global τ-shift moves induce an equivalence between inner and outer moves, which however does not imply equal acceptance rates.

A final noteworthy detail of this sampling procedure is the choice of the outer superstate for the initial configuration at the beginning of the simulation. While it should not influence the simulation after thermalization, for many highly excited outer superstates the local weight is close to zero. We thus select the initial outer superstates randomly with probabilities proportional to their local weights.

Let us note that one could think of simpler techniques than the presented moves to ensure ergodicity, e.g., the addition of a move that changes *only* the outer superstate or the possibility to change the outer superstate randomly during each move. Both of these turn out to be inefficient ways compared to those we present here. Adding such a move is, however, necessary for ergodicity in a simulation of a system in the atomic limit, i.e., without hybridization, because outer (and inner) insertions would always be rejected, operators for outer removals are not present, and the move presented in the following section does not change a configuration without operators at all. Therefore, we do occasionally propose a change of just the outer superstate in configurations without any operators.

C. Global τ-shift moves

A global "τ-shift move" shifts the positions of all operators in imaginary time by a random $\Delta \tau \in [0, \beta]$, which can equivalently be thought of as a shift of the imaginary time axis. At the same time, the new outer superstate is by construction chosen to be consistent with this shift of the origin of the imaginary time axis (see Fig. 3).

Using just inner pair moves and the global τ-shift move, a superstate sampling simulation is ergodic if and only if it is ergodic using conventional sampling with pair moves. This is because if any configuration can be built up using pair moves in conventional sampling, any configuration can be built up using inner pair moves with the outer superstate being one of the contributing ones in superstate sampling. A proof of ergodicity can be found in Appendix A.

Let us now consider the properties of the τ-shift move. Over the course of an entire simulation, the proposal probability for a specific outer superstate in this kind of move is proportional to the average relative amount of imaginary time it covers. Since the superstate sequence is cyclic and effectively also just shifted along the τ axis, there is no need to perform quantum number checking. A global move similar to our τ shift was introduced by Shinaoka *et al.* [5] for a different technical reason.
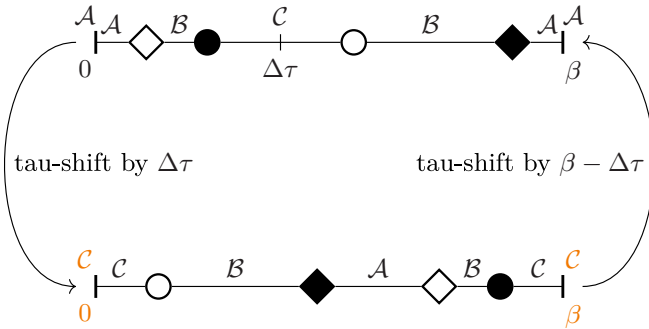
FIG. 3. Diagrams representing two configurations that can be obtained from one another using a τ-shift move. The symbols on the imaginary time segments represent the impurity operators, with shapes standing in for flavors and filledness for whether they are creators or annihilators. Calligraphic letters denote superstates with the ones above the $\tau = 0$ and $\tau = \beta$ mark giving the outer superstate of the configuration. When a τ-shift move is performed, all operators are shifted according to $\tau \to \tau - \Delta\tau$ (wrapping around to $\tau \to \tau + \beta - \Delta\tau$ if necessary).
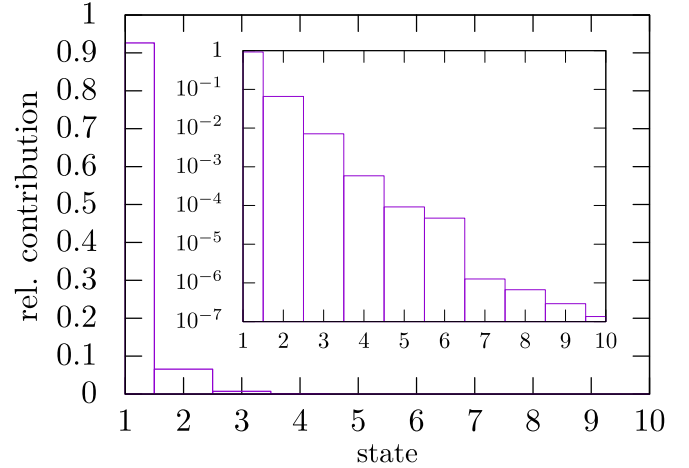

FIG. 4. Average relative contribution of outer states within one superstate [19], ordered by the size of contribution per configuration, normalized to the total weight of that superstate. This graph was obtained from a typical simulation of a five-orbital system with Kanamori interaction [18].

The proposal probabilities of a τ-shift move and its reverse are equal. The acceptance probability of this move is 1. In Appendix B we prove that the bath determinant remains unchanged, as the action of the τ shift on the hybridization matrix move effectively corresponds to a number of permutations and multiplications of rows and columns. Additionally in Appendix C there is a proof that the local trace remains unchanged under a combined cyclic permutation of the operators and corresponding change of the outer superstate.

Let us discuss why τ shift moves allow us to preserve ergodicity. Inner moves alone cannot change the outer superstate, but only superstates in parts not including $\tau = 0 = \beta$. τ shifts indeed move "the section with the outer superstate" away from $\tau = 0 = \beta$ and can hence shift the operator/superstate sequence in such a way that the outer superstate changes while the configuration remains otherwise equivalent. In combination with the ordinary inner pair moves it can hence change the superstate of any section without need for outer pair moves.

Due to its favorable characteristics compared to the outer moves, we usually add just τ-shift moves to the always necessary inner moves to ensure ergodicity of simulations with respect to the outer superstate. We choose to propose τ-shift moves as 0.5% of all moves by default, which was also the ratio used in all calculations shown later. While a smaller ratio might improve performance, the potential speedup in usual cases would be small as the τ-shift moves usually do not take up the majority of the time.

Additionally, we also allow random changes of the outer superstate in our implementation during other global moves that can be used in CT-HYB but which we do not further discuss in this paper. For such global moves as the flavor permutations used in *w2dynamics*, procedures mapping the old superstate to a new proposal based on the specific move could actually be thought of, but since these moves only serve to go between badly connected areas of phase space, we prefer not to restrict them more than necessary.

IV. STATE SAMPLING

A. General description

The superstate sampling method of the last section already significantly reduces the cost of computing a local weight. Similarly to the predominance of a single superstate in the local weight (Fig. 1), we often find that within one superstate, the individual eigenstates show a similar trend (Fig. 4, using the same model [18]): The contribution of one superstate $\mathcal{S}$ is dominated by the contribution of one or a small group of its eigenstates $s \in \mathcal{S}$. This suggests trying to apply the principle of superstate sampling one level deeper in the form of an "(eigen)state sampling."

We split the local weight further into even smaller parts, where the summation over all states s within a superstate $\mathcal{S}$ is also done as a Monte Carlo sum:

$$Z = \underbrace{\sum_{\mathcal{C}} \sum_{\mathcal{S}} \sum_{s \in \mathcal{S}}}_{\text{QMC}} w_{\text{loc},s}(\mathcal{C}) w_{\text{bath}}(\mathcal{C}), \tag{11}$$

with

$$w_{\text{loc},s}(\mathcal{C}) = \langle s | \hat{\mathcal{C}} | s \rangle. \tag{12}$$

To avoid confusion, let us stress at this stage that our method does *not* make any assumption on the contribution of the superstates to the local weight. It is an exact sampling with no approximation involved.

Unlike for superstate sampling, a heuristic argument from conventional sampling for a sign close to 1 cannot be given (unless a system with superstates containing a single state each is considered). This stems from the fact that at least quantitatively, the weight in state sampling is always different from the one in conventional sampling, as the latter sums up contributions from at least one entire superstate. However, as stated earlier, we empirically find the contribution of a superstate to the local weight is often dominated by one of its eigenstates. This suggests that the sign should often not be

much worse than in superstate sampling, as those dominating states should be sampled considerably more often than other ones. Because the local weight in state sampling when such a dominating state s is the outer state has the same sign as the local weight of the corresponding configuration with outer superstate $\mathcal{S} \ni s$ in superstate sampling, the maximum deterioration of the sign as compared to superstate sampling should be related to the extent to which individual states dominate the superstate weight contributions. As the time evolution further suppresses states of higher energy at lower temperature, causing the lower energy states to dominate the superstates' contributions more and more, the "sign gap" between the methods should also decrease with decreasing temperature.

Finally, let us note that this sampling technique only improves performance over superstate sampling if the summation over the outer states is actually performed as the outermost summation in the numerical implementation as well, as opposed to multiplying the operator matrices first. While performing the summation over outer *superstates* first is common in implementations of conventional CT-HYB sampling as this allows the use of quantum number checking for performance improvement, the summation over the outer *states* is instead often performed only after the multiplication of the operator matrices (per superstate), as this allows the use of optimizations employing tree structures. As opposed to these matrix-matrix implementations, there are so-called "matrix-vector" ones that perform the summation over the outer states as the outermost one. For more detailed information, see Appendix D.

B. Choice of the outer state within a superstate

The crucial point of moves in state sampling compared to those in superstate or conventional sampling is how to choose the outer state. This is also the only point in which our state sampling moves differ from the ones we propose for superstate sampling. In superstate sampling the choice of the new outer superstate is clear from the way in which the superstate sequence is changed by the move. This is the case for both outer moves and the τ-shift move (cf. Secs. III B and III C). In state sampling, the situation is different because a qualitative equivalent of the superstate sequence for states does not exist. At $\tau = 0$ there is indeed an eigenstate of the Hamiltonian, but it will change into a linear combination of multiple eigenstates after the first application of an impurity operator.

In the inner pair moves (cf. Sec. III B, Fig. 2), we simply keep the outer state fixed, just as we kept the outer superstate fixed in the corresponding moves in superstate sampling. For the kinds of moves which we want to use to change the outer state, we randomly choose the outer state to be proposed from a suitable set (cf. Fig. 5). This involves first following essentially the same procedure for a move as in superstate sampling, which allows us to obtain an outer superstate proposal which we could call the target superstate. From this target superstate, we randomly choose a state to propose as outer state, in our specific implementation according to the distribution in Eq. (13). In this optimized probability distribution for the outer state proposal, we take only the eigenenergies of the

different possible outer states into account, as we have found this to be a useful way to increase the sampling efficiency.

Let us take a look at some detailed examples for these latter kinds of moves: When the outer pair insertion (cf. Sec. III B) depicted by Fig. 2 is performed in state sampling, the initial configuration depicted in the middle is of course extended by the specification of one outer state $s \in \mathcal{A}$, and one of the states contained in $\mathcal{E}$ is randomly selected as the outer state of the proposed new configuration depicted below it. Similarly, in the τ-shift move (cf. Sec. III C) depicted by Fig. 3, the initial configuration depicted on top is extended by the specification of an outer state $s \in \mathcal{A}$ and one of the states contained in $\mathcal{C}$ is randomly chosen as the outer state of the proposed configuration. In any further global moves with no particular connection to the superstate sampling technique, we chose to randomly propose one of the contributing superstates. Therefore when they are performed in state sampling, we randomly choose one of the states contained by the contributing superstates— the only case in our implementation where outer states from more than one different superstate could be proposed in one otherwise identical move.

Proposing one of the possible states with uniform probability, however, causes a lower acceptance rate compared to the analogous moves in superstate sampling, where, e.g., the τ-shift move even has acceptance rate 1. Therefore, we employ a more efficient strategy: Since close to $\tau = 0$ there is always just the outer state s propagating with its eigenenergy, we can make the procedure more efficient by "transferring" the time evolution of this state from the acceptance to the proposal probability; i.e., we include it in the proposal probability so that in the standard Metropolis acceptance probability formula, it cancels with the equivalent factor in the weight of the configuration. This proposal probability weighting allows us to incorporate our prior (or easier to calculate) knowledge to avoid wasting time on proposals that would likely be rejected: For a well chosen proposal probability, i.e., one close to the actual weight distribution, we raise the acceptance probability for all outer states since those outer states that would be rejected more often with uniform proposal probability are simply proposed less often. To include the aforementioned time evolution close to $\tau = 0$, we use the proposal probability

$$p_{\text{prop}}(s) = \frac{\exp(-(E_s - E_0) \cdot (\tau_f + \beta - \tau_l))}{\sum_{k \in \mathcal{T}} \exp(-(E_k - E_0) \cdot (\tau_f + \beta - \tau_l))}, \quad (13)$$

where $\mathcal{T}$ contains all states from all outer superstates that may be proposed and τ_f and τ_l are, respectively, the imaginary times of the first and last operator after the move. In this way, the acceptance rate for outer state changes is significantly increased and, e.g., reaches about 50% for the global τ-shift move in typical calculations. The best choice of proposal probability for such an optimization depends on the weight we expect: In this case, we essentially assume that the potentially excited state at $\tau = 0$ will be brought closer to the ground state by the impurity operators, since if we expected the energy to stay at the level of the outer state, we could choose a better proposal probability assuming propagation over the entire imaginary time with the energy of the outer state (which corresponds to replacing $\tau_f + \beta - \tau_l$ by β in the proposal probability).
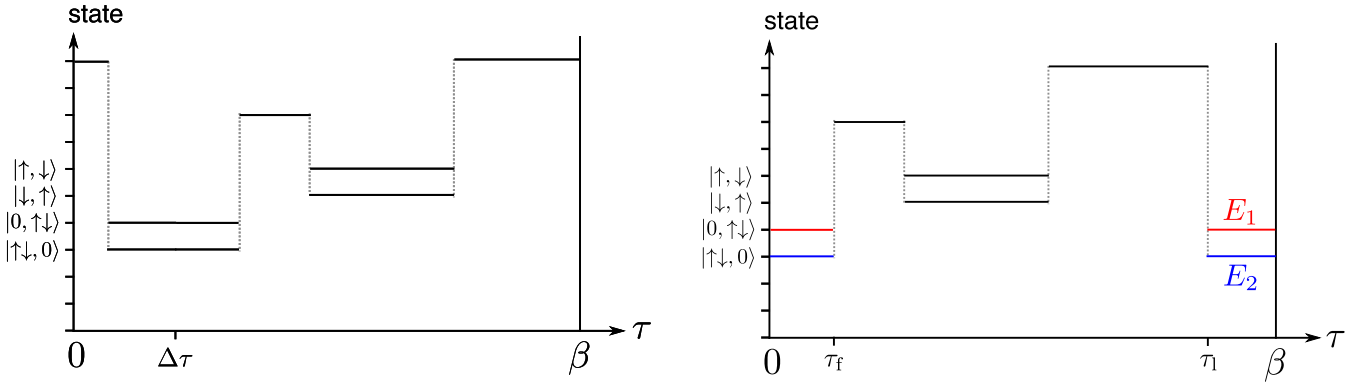
FIG. 5. Left: a superstate sampling configuration for a two orbital model with Kanamori interaction. Bold horizontal lines denote a time evolution of an eigenstate, the dotted vertical lines the operators that cause transitions between superstates. Here all superstates have size 1, except the spin-flip and pair-hopping one. Right: the configuration resulting from application of a global τ shift by $\Delta\tau$ to all operators of the one shown on the left. τ_f and τ_l denote the imaginary times of the first and last operator after the shift. In the superstate sampling algorithm the local weight is the sum of the red and blue state as outer state of the trace, whereas in the state sampling only one of the two is selected.

V. SPEEDUP QUANTIFICATION

To check the correctness of the results, we use a two-orbital model with a Kanamori interaction, i.e., density-density, pair-hopping, and spin-flip terms, and a finite number of bath sites for which we have a reference solution obtained using exact diagonalization. Both the reference self-energy as well as one calculated using our CT-HYB solver are shown in Fig. 13 (Appendix E).

In order to analyze the performance, we use a five-orbital Hamiltonian modeling a realistic transition-metal impurity on the surface of a metal [18] with both the full (spherically symmetric) Coulomb tensor as well as one derived from the same interaction matrix restricted to Kanamori-like terms only. We perform all calculations using *w2dynamics* [9], with either the implementation of conventional sampling [20] found in older versions, or superstate and state sampling proposed here; other than the sampling method, there are no differences with significant performance impact between the calculations. To quantify the performance improvement we compare the sampling rates, i.e., the raw amount of generated (correlated) samples per time. For the autocorrelation time we found a minor increase of about 10% for superstate sampling but only about 3% for state sampling as compared to the old sampling method. This can be considered negligible in comparison to the speedup factors. The mean sign is about 1.0 for the model with Kanamori interaction using both conventional and superstate sampling and about 0.98 using state sampling. For the model with full Coulomb interaction the sign is significantly less than 1 in all cases, and it is slightly smaller with state sampling than with the other methods [21]. For the calculations used to measure the speedup factors, the absolute values of the mean sign for all three methods can be found in Fig. 7.

Figure 6 shows the achieved speedup factors (top panel) and the absolute sampling rates measured in simulations (bottom panel) of the impurity model with Coulomb interaction. We obtain a speedup of the Monte Carlo sampling up to a factor of about 700 in the considered temperature range, depending on the used sampling method, temperature, and interaction. Remarkably, the speedup of our five-orbital

example was never smaller than 100 for the most arduous case, i.e., the full interaction. The reason why the speedup factors for the Coulomb interaction are larger under otherwise equal conditions is that a larger number of superstates
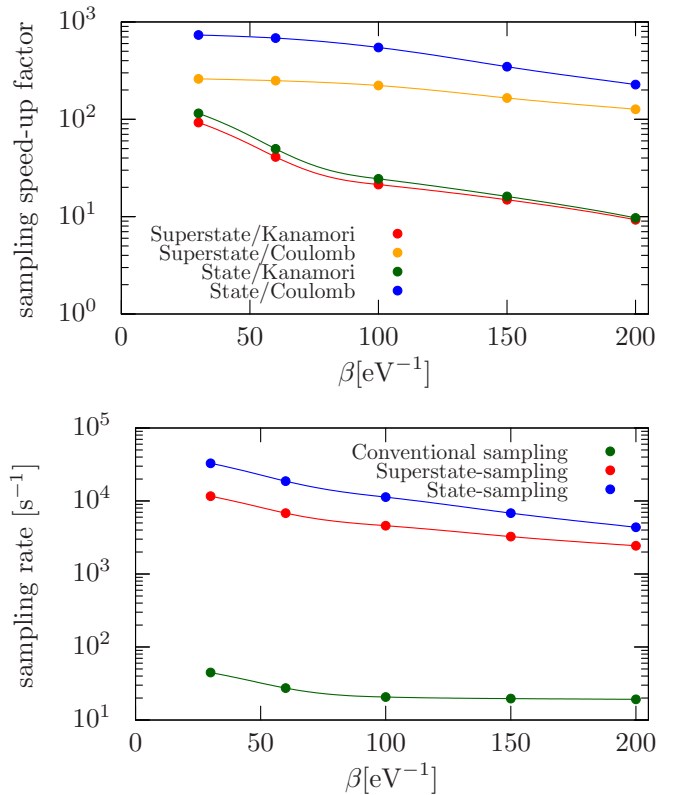

FIG. 6. Comparison of Monte Carlo sampling speed, i.e., number of individual, mostly local, updates per CPU time, not including measurements, for a five-orbital system with Kanamori or Coulomb interaction and cubic interpolation curves plotted logarithmically. Top panel: speedup factors of the sampling methods compared to conventional sampling for Kanamori and Coulomb interaction. Bottom panel: speed of all the sampling methods for Coulomb interaction.
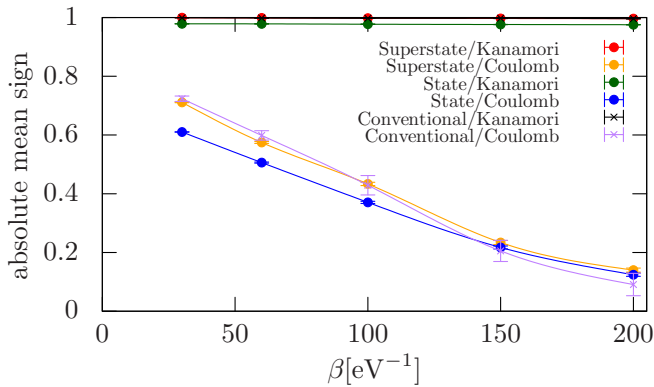
FIG. 7. Absolute value of the mean sign using our sampling methods for a five-orbital system with Kanamori or Coulomb interaction with cubic interpolation. The sign obtained in simulations using conventional sampling is usually not better than with superstate sampling, which we confirmed for the case with Kanamori interaction. Note that the error of the sign for the model with Coulomb interaction is considerably larger using conventional sampling than using other algorithms. The graph also shows that the sign obtained in state sampling is only slightly worse that that obtained using the other methods.

contribute on average in this case. A general observation is that the speedup decreases with decreasing temperature because the number of superstates contributing to the trace decreases with decreasing temperature as more operators tend to cause more quantum number violations. Therefore only the quantum number checking can be avoided at lower temperatures, whereas at higher temperatures other trace contributions are present for which matrix-vector products need not be calculated any more.

Yet, there is still a noticeable speedup even at lower temperatures where quantum-number checking takes a large amount of the total time [22]. As the speed-up affects only the trace calculation, it will also continue to decrease with β because the computational complexity of the bath determinant (which scales with β^3) is asymptotically larger than that of the trace calculation.

Another noticeable feature is the bigger advantage of state sampling over superstate sampling for Coulomb interaction, which is due to the larger average size of the outer superstates in that case. The amount of calculated matrix-vector products is reduced by approximately that factor in state sampling compared to superstate sampling, as only one of the outer states is chosen in the former case. This optimization is only advantageous for a matrix-vector solver like ours, as additional outer states can be included at negligible further cost if the entire product of the operator matrices for a specific outer superstate has been calculated, cf. Sec. IV and Appendix D. A similar optimization is also possible without splitting configurations in the superstate sampling method by using the cyclical invariance of the trace and starting the trace calculation at a τ where the superstate of the configuration is smaller than the outer superstate of the configuration, but this can interfere with time savings from caching of intermediate state vectors and even the size of the smallest superstate may

be greater than 1. More details on the methods can be found in Ref. [23].

In conclusion, we find that superstate sampling improves performance without significant drawbacks to such an extent that it should always be preferable to the conventional sampling method. Especially in simulations with few good quantum numbers, state sampling can provide an additional speedup, though it can also impact the mean sign. In our examples, the speedup in the case with full Coulomb interaction is big enough to clearly outweigh the marginally reduced sign, but this may depend on characteristics of the system and the implementation.

VI. APPLICATION: THE SPIN-FREEZING CROSSOVER

In Ref. [24], Werner *et al.* applied DMFT to a model with three degenerate orbitals and rotationally-invariant Coulomb interaction. Upon changing the filling n, they identified a sharp change in the qualitative behavior of the local spin susceptibility. For small fillings, the latter becomes rapidly small at large imaginary times τ, as in a standard metal. Approaching the Mott transition at $n = 3$, it starts instead to closely resemble that of an atomic insulator, i.e., it seems to become essentially constant in τ. This is surprising, as it happens for fillings still on the metallic side, before reaching the metal-insulator Mott-Hubbard transition. The sudden loss of coherence was interpreted as an abrupt crossover—or even a true quantum phase transition ("spin-freezing")—to a bad metal characterized by violations of the Fermi-liquid properties, along a line in the zero-temperature n-U phase diagram.

An independent analysis showed, however, convincing evidence that the same model remains in a metallic phase away from integer fillings [25]. What changes upon getting close to the Mott transition is the coherence of the quasi-particle excitations. The "spin-freezing" is therefore a finite-temperature—though surprisingly rapid—crossover to a "bad-metal" rather than a $T = 0$ phase transition. As long as the DMFT self-consistence does not lead to a gap in the spectral function of the bath of the corresponding Anderson impurity model, the solution at zero temperature must indeed be a Fermi liquid. This conclusion was already demonstrated more rigorously in Refs. [26,27].

Yin, Haule, and Kotliar [28] and, immediately after them, Georges, de' Medici, and Mravlje [29] performed CT-HYB calculations of hitherto unprecedented efficiency, reaching temperatures 1000–1500 times smaller than the half band-width (see Fig. 6 and Fig. 7 in the two papers, respectively). The focus was on the functional dependence of $\text{Im}\Sigma(i\omega_n)$ which, even after the spin-freezing crossover, was shown to follow a Fermi-liquid scaling at extremely low frequencies, visible at these temperatures. A recent study nailed this down using an advanced multiorbital numerical renormalization group solver [30].

Similar types of crossovers have been discussed in the presence of spin-orbit interaction [31] always showing the change of behavior in the local spin susceptibility at a fixed temperature. Yet, one would like to unambiguously demonstrate that this is actually the physics of a crossover from a good to a bad metal with a coherence temperature that

becomes fairly small upon approaching the Mott transition at half filling.

Here, we consider the same model of Ref. [24] as a function of doping focusing in particular on the temperature dependence. Using the CT-HYB implementation of *w2dynamics* [9], featuring both superstate sampling and sliding window sampling [14], we are able to obtain a clear picture of the temperature evolution of the local spin response, identifying a coherence scale, even deep into the "spin-freezing region." The quantities of interest are the electronic self-energy $\Sigma(i\omega_n)$ and the static local spin susceptibility

$$\chi_{\text{loc}}^{\omega=0}(T) = \int_0^\beta d\tau \, \chi_{\text{loc}}(\tau), \qquad (14)$$

i.e., the $\omega = 0$-Fourier component of the spin-spin response function $\chi_{\text{loc}}(\tau) = g^2 \sum_{ij} \langle S_z^i(\tau) S_z^j(0) \rangle$ (with i and j running over the three orbitals). The half-bandwidth of the semicircular noninteracting density of the states of each orbital is D (corresponding to $2t$ in Ref. [24]). The coupling constants of
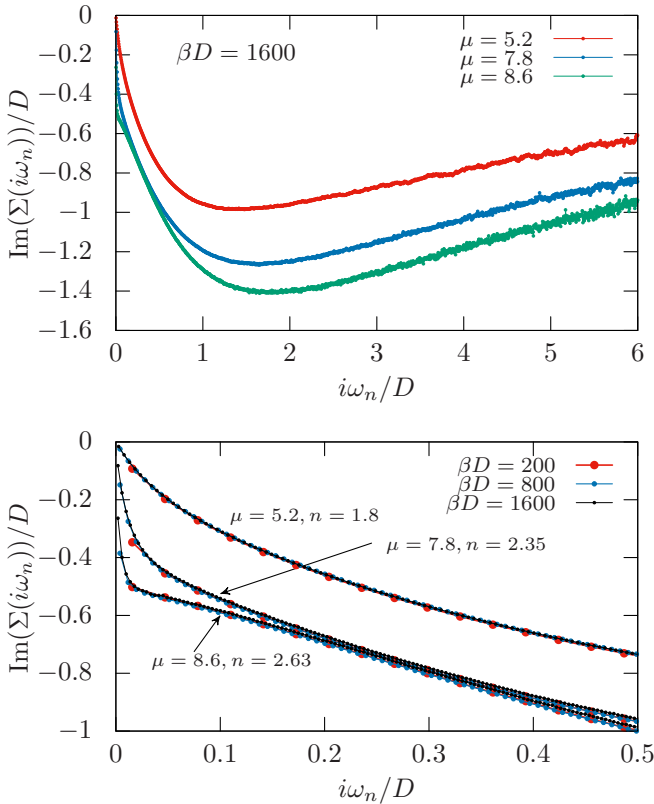
the three-orbital Kanamori interaction are the usual Hund-J and Hubbard-U (with $V = U - 2J$) at fixed $J/U = 1/6$ ratio.

We first focus on the imaginary part of the Matsubara self-energy $\text{Im}\Sigma(i\omega_n)$ for three different fillings at the lowest temperature $T = 1/\beta = D/1600$. As shown in Fig. 8, for filling $n = 1.8$ the extrapolation of $\text{Im}\Sigma(i\omega_n)$ for $i\omega_n \to 0$ is not dramatically influenced by the temperature. This indicates that the system has reached a coherent Fermi-liquid state and further lowering T does not change the shape of $\text{Im}\Sigma(i\omega_n)$ but only makes the Matsubara frequencies denser, remaining on the same "straight" line. This is a manifestation of the so-called "first-Matsubara" rule [32–34], according to which a T^2 scattering rate characteristic of a Fermi liquid gives rise to a linear-in-T value for $\text{Im}\Sigma(i\omega_{n=0})$.

The situation is drastically different for larger values of n. Note that at this U, $n = 2.35$ and 2.63 had been already assigned to the "spin-freezing region" in the original paper by Werner *et al.* (see inset to Fig. 11). At these fillings the low-frequency part of $\text{Im}\Sigma(i\omega_n)$ is highly nonlinear and it is clear that to recover linearity one has to consider the lowest temperatures (and probably even lower than $T = D/1600$ at $n = 2.63$). This unambiguously hints at a sudden drop of the Fermi-liquid coherence temperature upon increasing the filling n.

An inspection of the local spin susceptibility confirms that the physics at $n = 2.35$ and 2.63 is not qualitatively different from the good-metal fillings but it is just the result of a strong renormalization of the coherence properties. The results are shown in Fig. 9. For an atom, $\chi(\tau)$ is perfectly flat independently of the temperature, so that its integral from 0 to β is proportional to β (Curie law). For a Fermi liquid, its shape instead has to change with temperature in such a way that its integral gives a constant Pauli susceptibility. Even though the speed of the decay for $n = 2.35$ and 2.63 is greatly reduced compared to $n = 1.8$ (in agreement with Werner *et al.*), a pronounced temperature dependence of $\chi(\tau)$ is present also for the larger fillings, revealing the Fermi-liquid properties.

To conveniently represent the evolution with the filling we look at the value of the susceptibility at $\tau = \beta/2$: In the

FIG. 8. Imaginary part of the self-energy on the Matsubara axis with $U/D = 4$, $J/D = 2/3$. Upper panel: three different chemical potentials at low temperature $\beta D = 1600$. The corresponding fillings are reported in the lower panel and represent values across the spin-freezing crossover (see inset to Fig. 11). A closeup of the low-frequency region is shown in the lower panel together with the evolution with temperature. The Fermi liquid behavior for $n = 2.63$ ($\mu = 8.6$), i.e., deep in the "spin-freezing region," is recovered only when considering the lowest temperatures. It indeed occurs at such a low-energy scale that its existence is not foreseeable by extrapolating the data at, e.g., $\beta D = 200$.
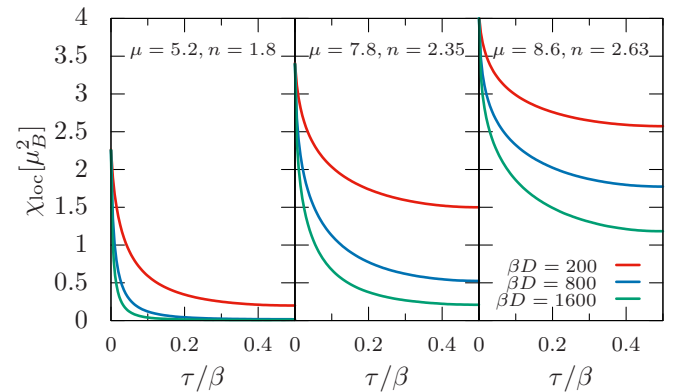

FIG. 9. Local spin susceptibility for three different fillings across the spin-freezing crossover. While the instantaneous moment at $\tau = 0$ gets larger upon increasing the filling, the value at $\tau = \beta/2$ can be made arbitrarily small upon reaching low enough temperatures. This supports the picture of a Fermi liquid with progressively lower and lower coherence scale.
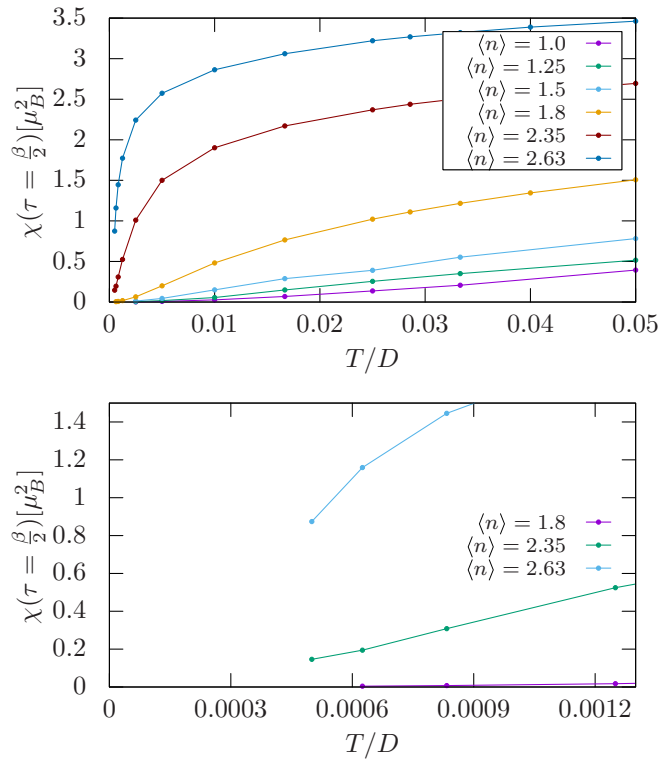
FIG. 10. Behavior of $\chi(\tau = \beta/2)$ for different temperatures and fillings. The lower panel contains a closeup of the three fillings shown in Fig. 9 across the spin-freezing crossover.
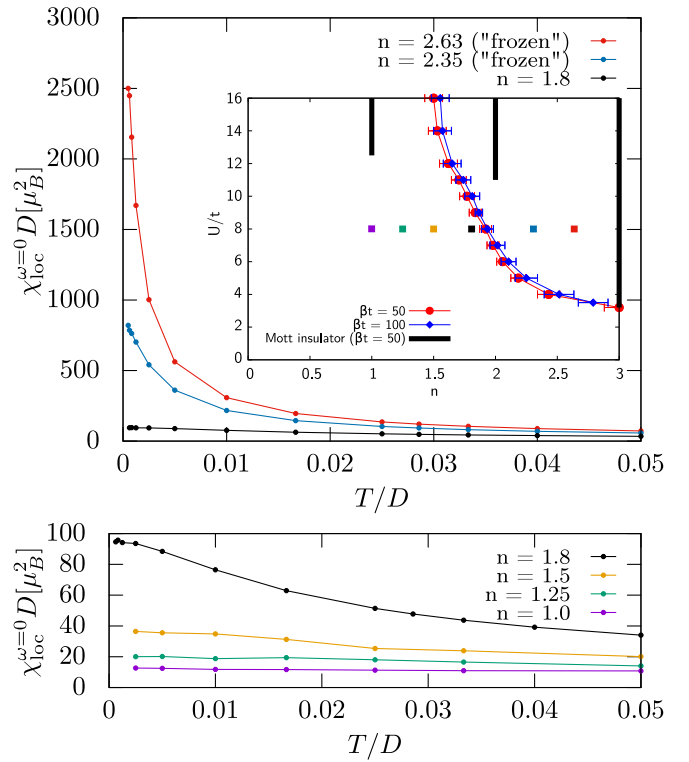

FIG. 11. Temperature dependence of $\chi_{\text{loc}}^{\omega=0}(T)$ for different fillings (indicated by square markers of the same colors in the inset taken from Ref. [24]). The low-T Pauli-like behavior is visible in the "nonfrozen" regime (lower panel). At higher temperature there is a crossover to Curie-Weiss physics. The latter gets more and more dominant in the "frozen" regime (upper panel; upper right region of parameter space in the inset).

Fermi-liquid case this has to go (quadratically) to 0 upon reducing T. The coherence temperature can be estimated for instance from the inflection point of $\chi(\tau = \beta/2)$ (see also Fig. 12). By looking at Fig. 10 one immediately understands how the filling efficiently reduces the temperature scale at which $\chi(\tau = \beta/2)$ approaches 0 and hence how it makes the coherence temperature drop. At the same time our results reveal how the physics of this model is qualitatively the same at all metallic fillings. The difference between the curves for different n is indeed only the velocity of the renormalization and the temperature scale at which the Fermi-liquid scaling is recovered.

The crossover from Curie to Pauli upon reducing the temperature can also be visualized in Fig. 11. Upon approaching the spin-freezing crossover the Pauli behavior gets progressively pushed to lower and lower temperatures. In the curves at $n = 2.63$ for instance it is clear that even lower temperatures would be needed to fully resolve it. For this reason it is hard to unambiguously prove that deep into the "spin-freezing region" the coherence scale is actually exponentially small. Even in the good-metal region a precise estimate of the crossover temperature T_{coh} is not a trivial task. First of all there is a dependence on the observable one is focusing on. Second, even by looking at the same physical quantity, different criteria give somewhat different answers.

In Fig. 12 we quantify T_{coh} via two approaches based on $\chi(\tau = \beta/2)$. These two ways of quantitatively estimating T_{coh} give compatible results though affected by sizable errors. Furthermore, even though the crossover from the

high-temperature $\sim 1/T$ Curie to the Pauli region becomes relatively sharp (i.e., in principle more easily identifiable) after crossing the "spin-freezing" crossover line, we have difficulties in precisely determining T_{coh}, as the latter is very small there and we do not have many susceptibility data points at such low temperatures. Nevertheless, the existence
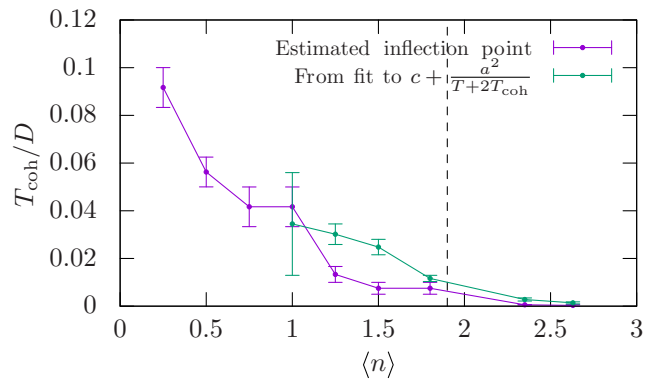

FIG. 12. Coherence temperature T_{coh} determined with two different methods: with the inflection point of the curves shown in Fig. 10 and by fitting the high temperature region with Wilson's formula [34,35]. We have not been able to obtain good results using the latter method at the lowest densities.
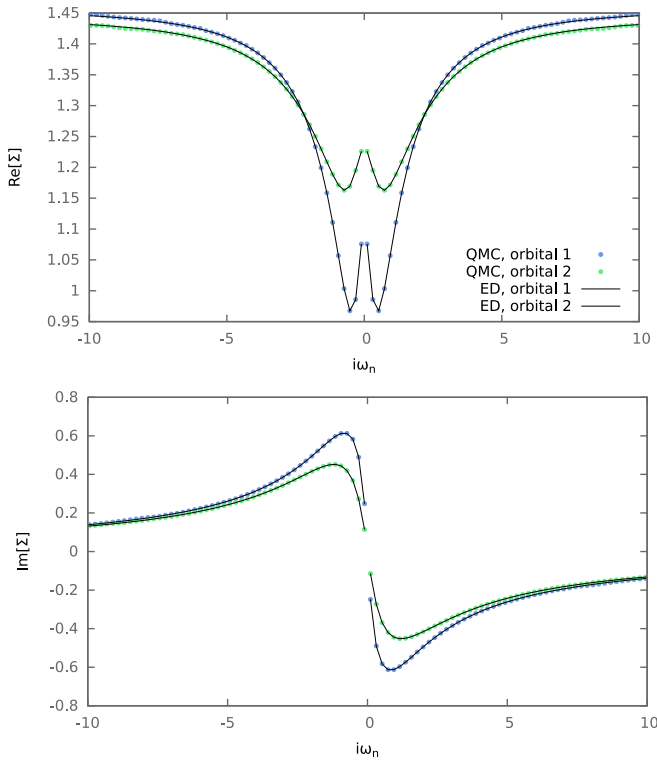
FIG. 13. Comparison of the CT-HYB self-energy calculated using the superstate sampling method (points) to exact diagonalization (lines) for a two-orbital system with Kanamori interaction.

of a *sudden* drop of $T_{\rm coh}$ approaching half filling, as in fact pioneered by Werner *et al.* in Ref. [24], is undoubted. Similar conclusions are corroborated by high-precision numerical renormalization group studies, published in Refs. [36,37]. The reason why this crossover is so sharp, as well as its shape in the doping-U diagram, are not fully clear yet [38].

In real materials, the position of the coherence scale can be strongly influenced by several factors, such as the nonlocal hybridization between orbitals, absent in the model Hamiltonian studied here. One of these factors has been also identified in the presence of sharp peaks in the noninteracting density of the states [34], something often coexisting with the many-body physics in strongly correlated materials.

VII. CONCLUSIONS

We have shown that the sum over all impurity eigenstates of the local problem in CT-HYB can be divided into smaller pieces and sampled individually. This fragmentation leads to a remarkable gain in the algorithm's efficiency, to some extent against the general intuition. This is due to the exponential character of the imaginary time evolution $e^{-H_{\rm loc}\tau}$, which very sensitively damps the amplitudes of high energy excitations. Acting on the core of the exponential scaling in CT-HYB we manage to achieve speedup factors of the order of 10^3, with essentially no worsening of the average sign. Additional work has to be carried out in order to show whether the impact of the exponential scaling of the local problem can be reduced further by employing methods based on our ideas.

The speedup figures have been obtained for a five-orbital model with full-Coulomb interaction, representing physically relevant situations such as realistic transition-metal impurities deposited on metallic substrates. We also discussed the well-known spin-freezing crossover obtained in three-orbital Hubbard-model calculations. Reaching very low temperature allows us to quantify the coherence temperature and the recovery of the Fermi-liquid properties of the self-energy, even when this physics is pushed to very low scales by the proximity to the Mott transition.

APPENDIX A: PROOF OF ERGODICITY OF τ SHIFT MOVES

To prove ergodicity, rigorously at least for the case of density-density interaction, by connecting two configurations differing only in their outer superstates, we first take an arbitrary configuration $\mathcal{C}$ and connect it to the "empty" configuration, which has no operators: For an arbitrary outer superstate, consider the occupation number basis states contained by the superstates. To get from one superstate to another, impurity operators need to be applied such that a state from the first superstate would be transformed into a state from the second one. To do this in a simulation, perform inner pair moves to build up this sequence of operators right after τ; the second operator from each pair can be placed anywhere resulting in a configuration of nonzero weight, with one simple possibility being $\beta - \tau$ if the first is placed at τ. Now, perform one tau-shift move with $\Delta\tau$ being the point right after the last operator of the inserted sequence, and the outer superstate is changed from the old, "first" one to an arbitrary "second" one. By removing all operator pairs in the opposite order, an "empty" configuration with an arbitrary outer superstate is reached. We can now run the procedure in reverse to build up a new configuration $\mathcal{C}'$. This implies ergodicity, as any two configurations $\mathcal{C}$ and $\mathcal{C}'$ are connected by a finite number of moves.

It is simple to demonstrate that ergodicity is not lost if outer pair moves are replaced by global τ-shift moves or vice versa: An outer move can be replaced by a sequence of a τ-shift move, an inner move, and the τ-shift move inverse to the first one. In the first move, $\Delta\tau$ just needs to be sufficiently big to change the order of the operator pair that would be affected by the outer move in τ, then it can be performed as an inner move instead because the section of the trace that is changed has entirely been moved inside. A τ-shift move can be replaced by a sequence of outer and inner moves: Remove all operators

in pairs using outer and inner moves such that an operator-free configuration with the desired outer superstate is obtained, then put all operator pairs shifted by $\Delta\tau$ back in using outer and inner moves.

As a final remark, it should be stressed that even though a τ-shift move has acceptance 1 in superstate sampling, the acceptance of the outer pair moves remains smaller than that of the inner pair moves. This is due to the fact that outer moves are connected to inner moves by *specific* τ shifts that have a proposal probability smaller than 1.

APPENDIX B: INVARIANCE OF THE BATH WEIGHT UNDER GLOBAL τ-SHIFT MOVE

For the proof of the invariance of the bath determinant (adapted from Ref. [23]), let us consider the form of the hybridization matrix elements with time ordering along both dimensions,

$$\Delta_{ij} = \Delta(\tau_i - \tau_j^\dagger), \tag{B1}$$

where $\Delta(\tau)$ is the hybridization function, an antiperiodic function with period β, τ_i the imaginary time of the ith annihilator (ordered by imaginary time), and $\tau_j^\dagger$ the imaginary time of the jth creator. The number of annihilators and creators shifted across $\tau = 0$ by the move will be denoted as N_A and N_C in the following.

Due to the τ-shift move, the imaginary time of operators with $\tau < \Delta\tau$ will be transformed as $\tau \to \tau + \beta - \Delta\tau$ and that of other operators as $\tau \to \tau - \Delta\tau$. The arguments of the hybridization functions are only time differences in which the shift parameter $\Delta\tau$ always cancels, but in cases where exactly one of the two operators had a $\tau < \Delta\tau$, the corresponding matrix element changes its sign.

Additionally, since the ordering of the operators is cyclically permuted, the rows and columns are cyclically permuted such that the N_A first rows become the N_A last rows and the N_C first columns become the N_C last rows, where N_A is the number of annihilators with $\tau < \Delta\tau$ and N_C the number of creators with $\tau < \Delta\tau$. A cyclic permutation that moves every column or row exactly one position toward the front ("wrapping around" from the beginning to the end) is equivalent to swapping adjacent columns or rows $k - 1$ times, where k is the size of the matrix in that dimension (as the hybridization matrix is a $k \times k$ matrix for hybridization expansion order k).

Each swap causes the determinant to change its sign, and the sign change of matrix elements where only one operator wrapped around the end is equivalent to multiplying all wrapped rows and columns by -1, where every multiplication of a column or row causes the determinant to change its sign as well. In total, expressed using the number of wrapped operators $N_{\mathrm{wrap}} = N_A + N_C$, the determinant thus accumulates an additional factor of $((-1)^{k-1})^{N_{\mathrm{wrap}}} \cdot (-1)^{N_{\mathrm{wrap}}} = (-1)^{kN_{\mathrm{wrap}}}$, where the matrix size k is the expansion order.

This extra factor is compensated by the sign that is incurred due to time ordering. That the change of this extra sign is equal to the factor acquired by the determinant may be proven by considering the amounts of permutations necessary to restore the ordering after performing a τ-shift move that wraps exactly one operator around the origin. From this, the general case follows.

APPENDIX C: INVARIANCE OF THE LOCAL WEIGHT UNDER GLOBAL τ-SHIFT MOVE

To prove the invariance of the local weight in superstate sampling under a τ-shift move, we use the cyclic invariance of the trace (adapted from Ref. [23]). Due to the way superstates are chosen by definition we know that if the trace is restricted in such a way that nonzero components are left for only one superstate at any τ, the result will be the same as if done so everywhere. Our local weight is effectively the conventional local weight with such a restriction applied at $\tau = 0$ and $\tau = \beta$, and if projection operators $P^{(x)}$ are inserted onto the outer superstate x at the beginning and end of the product of time-evolution, creation, and annihilation operators corresponding to current configuration, it can be written as a proper trace:

$$w_{\mathrm{loc}} = \sum_{s \in x} \langle s | \hat{\mathcal{C}} | s \rangle \tag{C1}$$

$$= \mathrm{tr}(P^{(x)} \hat{\mathcal{C}} P^{(x)}). \tag{C2}$$

As the time evolution does not mix states from different superstates, the superstate projection operator commutes with all time-evolution operators,

$$[\exp(-\tau \hat{H}), P^{(x)}] = 0, \tag{C3}$$

for any superstate x and any τ. Because the creators and annihilators map each source superstate to one unique target superstate and vice versa, a projector onto a superstate x on one side of an annihilator or creator can be replaced by a projector onto the superstate y which the operator maps x to on the other side of the operator:

$$d^{(y|x)} P^{(x)} = P^{(y)} d^{(y|x)}, \tag{C4}$$

where the mapping of superstates relevant for the specific case is given in parentheses in superscript with the meaning that applying $d^{(y|x)}$ to a state vector (to the right) in the subspace of superstate x will produce a state vector in the subspace of superstate y.

Using these two relations, we can commute the projectors all the way through the product to any other point and also to any imaginary time by splitting time evolution at that time into two consecutive time evolutions if necessary. The projectors will not necessarily be projectors onto the old outer superstate any more, but onto the superstate that can be found at that point in the sequence for the current configuration. After commuting both projectors to the position in the product corresponding to the imaginary time $\Delta\tau$, the product can be cyclically permuted such that one of the projectors ends up at each end of the trace. It is then equivalent to the local weight of the superstate sampling configuration after a τ shift by $\Delta\tau$. This shows that the local weight does not change after a τ-shift move.

APPENDIX D: COMPATIBILITY OF STATE SAMPLING WITH IMPLEMENTATIONS

Whether the performance of state sampling actually exceeds that of superstate sampling depends on the type of CT-HYB implementation it is used with. When reviewing

standard CT-HYB in Sec. II, we simply expressed the weight of a configuration as the trace of the product of the corresponding impurity operators. Typically, there are two ways to calculate it: In a matrix-matrix implementation, the matrices representing the impurity operators are multiplied with one another and the trace is obtained from the diagonal elements of the total matrix product. In the matrix-vector flavor, we instead explicitly perform a sum over the outer states: For each outer ket-vector, we repeatedly calculate the matrix-vector product with the operators, starting from the first all the way through to the outer bra vector, with which we finally compute the scalar product.

By decomposing the problem into superstates (i.e., block diagonalizing the Hamiltonian and choosing the blocks such that the operators connect them in a one-to-one way), we simplify the calculation of the trace in that the outer states (per superstate) follow independent, noncrossing paths through the superstates. As a result, in the matrix-vector implementation one can just reduce the size of the initial operator matrices. In a matrix-matrix implementation the trace of the product can be decomposed into the sum of traces per outer superstate,

which introduces the explicit summation whose terms can be calculated using matrices of reduced size as well.

If we consider such an implementation as the starting point, we just have to restrict the summation to one outer superstate in either case to implement superstate sampling. While we can obviously reduce the number of needed calculations by restricting the outer sum to one state only in the matrix-vector implementation, i.e., implementing state sampling, there is no way to beneficially implement this in a matrix-matrix algorithm. We could use only one of the diagonal elements of the resulting products, but this does not make the product calculation simpler and therefore does not improve performance but would only waste the other contributions that could be included at negligible further cost.

APPENDIX E: EXACT DIAGONALIZATION CROSS CHECK

Here we show that the results of the CT-HYB results with superstate sampling agree with an exact-diagonalization benchmark.

---

[1] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005).

[2] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, Phys. Rev. Lett. **97**, 076405 (2006).

[3] K. Haule, Phys. Rev. B **75**, 155113 (2007).

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

[5] H. Shinaoka, E. Gull, and P. Werner, Comput. Phys. Commun. **215**, 128 (2017).

[6] P. Seth, I. Krivenko, M. Ferrero, and O. Parcollet, Comput. Phys. Commun. **200**, 274 (2016).

[7] K. Haule, C.-H. Yee, and K. Kim, Phys. Rev. B **81**, 195107 (2010).

[8] L. Huang, Y. Wang, Z. Y. Meng, L. Du, P. Werner, and X. Dai, Comput. Phys. Commun. **195**, 140 (2015).

[9] M. Wallerberger, A. Hausoel, P. Gunacker, A. Kowalski, N. Parragh, F. Goth, K. Held, and G. Sangiovanni, Comput. Phys. Commun. **235**, 388 (2019).

[10] P. Werner and A. J. Millis, Phys. Rev. B **74**, 155107 (2006).

[11] N. Parragh, A. Toschi, K. Held, and G. Sangiovanni, Phys. Rev. B **86**, 155158 (2012).

[12] P. Sémon, C.-H. Yee, K. Haule, and A.-M. S. Tremblay, Phys. Rev. B **90**, 075149 (2014).

[13] A. M. Läuchli and P. Werner, Phys. Rev. B **80**, 235117 (2009).

[14] H. Shinaoka, M. Dolfi, M. Troyer, and P. Werner, J. Stat. Mech. Theory Exp. (2014) P06012.

[15] P. Augustinský and J. Kuneš, Comput. Phys. Commun. **184**, 2119 (2013).

[16] M. Wallerberger, w2dynamics: continuous time quantum Monte Carlo calculations of one- and two-particle propagators, Ph.D. thesis, Technische Universität Wien, Wien (2016).

[17] One may somewhat reduce the residual advantage of superstate sampling visible at large β in Fig. 6 upon optimizing the quantum number checking or using more sophisticated schemes, such as those proposed in Ref. [12].

[18] We used a five-orbital Anderson impurity model with a local Slater-Kanamori interaction, parameterized by $U \approx 5.03$, $J \approx 0.64$, and $U' = U - 2J$. The temperature was $\beta = 30$ where not stated otherwise and the chemical potential was chosen to reach a mean filling of $\langle N \rangle = 8$ in the interacting model. The hybridization function was obtained from DFT for a Cobalt impurity on Cu(001) as described in Ref. [39].

[19] Note that the number of states per superstate varies and the axis range of 10 was sufficient to capture the maximum number of contributing states over the entire course of the simulation due to many good quantum numbers.

[20] Specifically, a matrix-vector implementation of CT-HYB with time evolution in eigenbasis. Quantum number checking (cf. Sec. II) is used to avoid the calculation of zero contributions to the local weight using a partitioning of the Hilbert space into superstates using conserved quantities for Kanamori-like interaction [11] and additionally automatic partitioning [3–6] for full Coulomb interaction. We do not use outer truncation of the local trace, sliding-window-style local updates, tree algorithms, or other optimizations even if mentioned in the introduction unless explicitly stated.

[21] We used about the same amount of CPU time for all sampling methods for this example, so due to the worse performance of the conventional implementation its error is considerably larger. While the error in this example is thus sufficiently large to allow other conclusions about the relative sign of the methods in some temperature ranges, data from many other calculations we did with all methods not specifically for the purpose of this article strongly indicate equal signs for conventional and superstate sampling and a sign closer to zero (with model-dependent extent) for state sampling.

[22] In this work we consider the computational cost of quantum-number checking in *w2dynamics* [9] with decreasing temperature. The magnitude of this speedup might of course be lower in other implementations.

[23] A. Kowalski, *Improved Sampling in Continuous-Time Quantum Monte Carlo Algorithms for Fermions*, Master's thesis, Julius-Maximilians-Universität Würzburg, Würzburg (2017).

[24] P. Werner, E. Gull, M. Troyer, and A. J. Millis, Phys. Rev. Lett. **101**, 166405 (2008).

[25] L. de' Medici, J. Mravlje, and A. Georges, Phys. Rev. Lett. **107**, 256401 (2011).

[26] L. De Leo, Non-Fermi liquid behavior in multi-orbital Anderson impurity models and possible relevance for strongly correlated lattice models, Ph.D. thesis, SISSA (2004).

[27] L. De Leo and M. Fabrizio, Phys. Rev. B **69**, 245114 (2004).

[28] Z. P. Yin, K. Haule, and G. Kotliar, Phys. Rev. B **86**, 195141 (2012).

[29] A. Georges, L. d. Medici, and J. Mravlje, Annu. Rev. Condens. Matter Phys. **4**, 137 (2013).

[30] K. M. Stadler, G. Kotliar, A. Weichselbaum, and J. von Delft, Ann. Phys. (2018), doi:10.1016/j.aop.2018.10.017.

[31] A. J. Kim, H. O. Jeschke, P. Werner, and R. Valentí, Phys. Rev. Lett. **118**, 086401 (2017).

[32] A. V. Chubukov and D. L. Maslov, Phys. Rev. B **86**, 155136 (2012).

[33] L. V. Pourovskii, J. Mravlje, A. Georges, S. I. Simak, and I. A. Abrikosov, New J. Phys. **19**, 073022 (2017).

[34] A. Hausoel, M. Karolak, E. Şaşıoğlu, A. Lichtenstein, K. Held, A. Katanin, A. Toschi, and G. Sangiovanni, Nat. Commun. **8**, 16062 (2017).

[35] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).

[36] K. M. Stadler, Z. P. Yin, J. von Delft, G. Kotliar, and A. Weichselbaum, Phys. Rev. Lett. **115**, 136401 (2015).

[37] A. Horvat, R. Žitko, and J. Mravlje, Phys. Rev. B **94**, 165140 (2016).

[38] L. de' Medici, "*Hund's Metals Explained*", The Physics of Correlated Insulators, Metals, and Superconductors Modeling and Simulation, Vol. 7, edited by E. Pavarini, E. Koch, R. Scalettar, and R. Martin (Forschungszentrum Juelich, Jülich, 2017).

[39] M. P. Bahlke, M. Karolak, and C. Herrmann, Phys. Rev. B **97**, 035119 (2018).