

Unsupervised learning of effective quantum impurity models

Jonas B. Rigo^{1,2,3,*} and Andrew K. Mitchell^{2,3,†}¹*Forschungszentrum Jülich GmbH, Peter Grünberg Institute, Quantum Control, 52425 Jülich, Germany*²*School of Physics, University College Dublin, Belfield, Dublin 4, Ireland*³*Centre for Quantum Engineering, Science, and Technology, University College Dublin, Dublin 4, Ireland*

(Received 15 June 2024; accepted 18 September 2024; published 17 October 2024)

Generalized quantum impurity models—which feature a few localized and strongly correlated degrees of freedom coupled to itinerant conduction electrons—describe diverse physical systems, from magnetic moments in metals to nanoelectronics quantum devices such as quantum dots or single-molecule transistors. Correlated materials can also be understood as self-consistent impurity models through dynamical mean-field theory. Accurate simulation of such models is challenging, especially at low temperatures, due to many-body effects from electronic interactions, resulting in strong renormalization. In particular, the interplay between local impurity complexity and Kondo physics is highly nontrivial. A common approach, which we further develop in this work, is to consider instead a simpler effective impurity model that still captures the low-energy physics of interest. The mapping from a bare to an effective model is typically done perturbatively, but even this can be difficult for complex systems, and the resulting effective model parameters can nevertheless be quite inaccurate. Here we develop a nonperturbative, unsupervised machine learning approach to systematically obtain low-energy effective impurity-type models, based on the renormalization-group framework. The method is shown to be general and flexible, as well as accurate and systematically improvable. We benchmark the method against exact results for the Anderson impurity model, and we provide an outlook for more complex models beyond the reach of existing methods.

DOI: [10.1103/PhysRevResearch.6.043044](https://doi.org/10.1103/PhysRevResearch.6.043044)

In the field of condensed-matter physics, the microscopic Hamiltonian describing a quantum many-body system (the *bare model*) is in many cases well known. However, for interacting systems, the complexity of these models grows quickly with the number of quantum degrees of freedom (for example, orbitals or spins), such that a brute force solution becomes analytically and/or numerically intractable for many realistic scenarios of interest. The challenge in many-body theory is therefore not in writing down the bare model, but rather in solving it. However, in many situations it is not necessary to consider the entire configuration space of the system, because only a (relatively) small *active* subspace controls the phenomena of interest [1]. Thus, *effective models* can be devised that faithfully capture the phenomena of interest while only keeping the active part of the configuration space. Such effective models have a reduced complexity and increased expressiveness. The challenge in many-body theory can therefore be restated as one of finding the best *solvable* model that describes approximately but accurately the physics of interest.

As an example, it is often more convenient to analyze an effective lattice model in second quantized form than

the *ab initio* treatment involving an all-orbital description of the constituent atoms [2–4]. Methods such as the constrained random-phase approximation [5,6], coupled cluster downfolding [7], density matrix downfolding [8], or the Pariser-Parr-Pople model for molecules [9] were devised to systematically eliminate inactive degrees of freedom and account for them with renormalized parameters of a reduced model.

These methods have in common that they require some prior knowledge about the bare model to derive the effective model. This situation shares some similarity with that of the *inverse problem* [10] encountered in statistical inference or machine learning [11], which seeks to infer a probabilistic model from observed data. Recently machine learning techniques have been explored to construct effective Hamiltonians from simulated data [12–15] or from experimentally measured data [16–18]. Another field where the inference of Hamiltonians has gained considerable relevance is in the analog simulation of quantum Hamiltonians on quantum hardware [19–21]. Indeed, machine learning inspired process- and Hamiltonian-tomography methods have been developed to infer the precise Hamiltonian that NISQ devices actually simulate, given hardware imperfections [22–27].

A good effective Hamiltonian must not just reproduce desired properties of the bare model, it must also be meaningfully simpler than the bare model. The canonical method to systematically eliminate degrees of freedom and obtain effective models is the renormalization group (RG) [28–30]. On the level of the thermal density matrix, a single RG transformation acts as the partial trace over the high-energy degrees

*Contact author: j.rigo@fz-juelich.de†Contact author: andrew.mitchell@ucd.ie

of freedom [31]. In the context of quantum physics, a central concept from information theory is that thermal states encode the corresponding Hamiltonian [32], such that comparing the ensemble of thermal states from two systems is equivalent to comparing the two Hamiltonians. This provides a way to measure the “difference” between a Hamiltonian before and after renormalization. By minimizing the distance in Hamiltonian space, one can therefore in principle optimize a simplified effective model to best approximate some more complicated bare Hamiltonian after renormalization. We rigorously derive such an approach in this paper for a specific class of systems known as “quantum impurity” models [33]. These models involve localized interacting quantum degrees of freedom (the “impurity” $\hat{H}^{\text{imp}}$) coupled (through $\hat{H}^{\text{hyb}}$) to one or more continuum baths of noninteracting conduction electrons ($\hat{H}^{\text{bath}}$), as described generically by

$$\hat{H} = \hat{H}^{\text{imp}} + \hat{H}^{\text{bath}} + \hat{H}^{\text{hyb}}. \quad (1)$$

Generalized quantum impurity models describe semiconductor quantum dot devices [34] and complex single-molecule junctions [35]. They also underpin our understanding of correlated materials through dynamical mean-field theory [36]. Developing a strategy for systematically and accurately deriving simple effective models to better understand and simulate complex systems is the ultimate application of this work.

A prerequisite for the design and use of complex quantum nanoelectronics devices with advanced functionality beyond the classical paradigm (for example in quantum metrology [37,38]) is a fundamental understanding of their low-temperature correlated electron physics and quantum transport properties. However, this is a notoriously difficult theoretical challenge because of the subtle interplay between the orbital and spin complexity of the nanostructure, determined by its structure and chemistry; strong electron interactions due to quantum confinement; and the coupling to $\sim 10^{23}$ conduction electrons in the external circuit. This results in nontrivial quantum phenomena such as Coulomb blockade [39], various forms of the Kondo effect [35,39,40], and quantum interference [41,42]—all of which strongly affect low-temperature electronic conductance through the device, and hence its functionality. As with coupled quantum dot devices [43–45], entangled spin and charge degrees of freedom can give rise to new physics in single molecule junctions. It is therefore a formidable task to derive simplified effective models that can still describe this range of physics in these kinds of quantum device.

A basic illustrative example is given in Fig. 1. Here we show the temperature dependence of a physical observable, in this case the entropy $S(T)$, for a bare quantum impurity model (blue), compared with that of a simpler effective model (orange). One can immediately identify that the effective model has lower complexity than the bare one, because in the high-temperature limit it has a lower entropy—signifying that the effective model has a smaller number of active degrees of freedom than the bare model. At high temperatures, the physical properties of the two models are distinctly different. But below some low-energy scale, the behavior of the two models becomes essentially indistinguishable. The low-energy physics is perfectly captured by the simpler effective

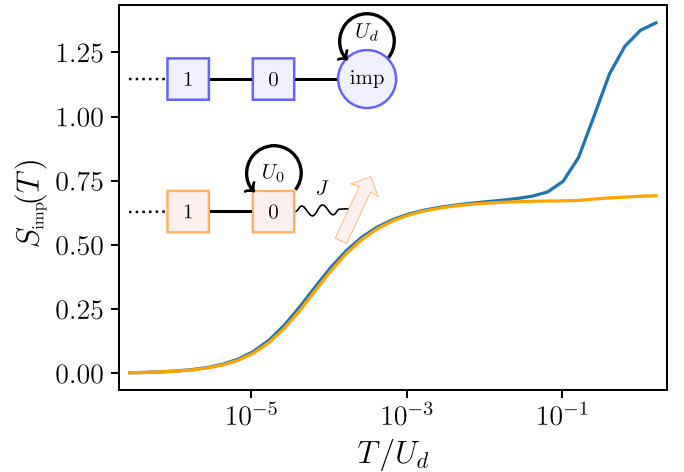


FIG. 1. Schematic of the simplest bare (blue) and effective (orange) impurity models, together with an illustrative comparison of their impurity entropies after model optimization.

model. We will return to the specific details of this example later.

In the context of molecular electronics, the complex microscopic models describing single-molecule junctions can be mapped to much simpler effective two-channel Kondo models using a perturbative approach, as described in Refs. [35,42]. This method captures simultaneously the effect of quantum interference and Kondo physics—but is inevitably approximate due to the perturbative treatment. Since thermodynamic observables flow under RG and the bare and effective model are defined at different energy scales, it is also not *a priori* clear whether such minimal Kondo models are sufficiently general to reproduce local observables of interest in the bare model [46]. In this paper, we show how an RG analysis of effective interaction terms can be used to determine thermodynamic observables that are comparable across different energy scales. This allows us to introduce a novel machine learning (ML) methodology to derive accurate effective models for complex quantum impurity problems that works by optimizing generalized (minimally constrained) models, and ensuring that local observables are correctly reproduced. We show by way of explicit examples that the low-energy Kondo physics is simultaneously captured. The parameters of the effective model are optimized by minimizing the Kullback-Leibler divergence (KLD) [47] that compares its ensemble of thermal states to that of the bare model. Information on the target is extracted from a numerical simulation of the system of interest. However, the full solution of the bare model is *not* required: minimization of the KLD requires only an estimation of thermal expectation values corresponding to specific local operators at relatively high temperatures. This can be achieved with any suitable quantum impurity solver [48–51]. Furthermore, the KLD can be shown to be convex under reasonable assumptions (see Appendix A) and its gradient is known analytically in closed form [52], which makes it an ideal optimization problem. We refer to this method as unsupervised model learning (UML). We demonstrate the efficacy of the UML method by application to the Anderson impurity

model (AIM) [33], and we obtain some new nonperturbative results for this old problem. Finally, we give an outlook to the application of this framework to more complex problems, where the development of tractable effective models is essential for the study of nontrivial correlated electron physics at low temperatures.

I. OVERVIEW OF THE METHOD

Unsupervised learning is a type of ML with the goal to recreate the probability distribution of some target data. Examples such as the Boltzmann machine [10] achieve this by minimizing the distance between the probabilistic ansatz, at the core of the machine, and the heuristic estimation of the target distribution given by some sample data [11]. The distinguishability between probability distributions can be computed using the aforementioned KLD [47]. The KLD can be generalized for quantum density matrices $\hat{\rho}$ in the form of the von Neumann relative entropy [53,54],

$$D_{\text{KL}}[\hat{\rho}_1 : \hat{\rho}_2] = \text{tr}[\hat{\rho}_1 \log(\hat{\rho}_1)] - \text{tr}[\hat{\rho}_1 \log(\hat{\rho}_2)]. \quad (2)$$

The generalized KLD quantifies how distinguishable $\hat{\rho}_1$ is from $\hat{\rho}_2$. The thermal density matrix $\hat{\rho} = \frac{1}{\mathcal{Z}} e^{-\beta \hat{H}}$ is fully defined by the system Hamiltonian $\hat{H}$ (and inverse temperature β). Here $\mathcal{Z} = \text{tr}[e^{-\beta \hat{H}}]$ is the partition function of $\hat{H}$. Thus, the thermal density matrix can be seen as a proxy for its defining Hamiltonian, and the KLD as a measure of distinguishability between two Hamiltonians. To emphasize this, we denote the KLD for two thermal density matrices $\hat{\rho}_1 = \frac{1}{\mathcal{Z}_1} e^{-\beta \hat{H}_1}$ and $\hat{\rho}_2 = \frac{1}{\mathcal{Z}_2} e^{-\beta \hat{H}_2}$ as $D_{\text{KL}}[\hat{H}_1 : \hat{H}_2]$.

Given the target Hamiltonian $\hat{H}_{\text{bare}}$, we seek to optimize the simpler effective model,

$$\hat{H}_{\text{eff}}(\theta) = \sum_i \theta_i \hat{h}_i, \quad (3)$$

by minimizing $D_{\text{KL}}[\hat{H}_{\text{bare}} : \hat{H}_{\text{eff}}(\theta)]$ with respect to the set of parameters $\{\theta_i\}$ corresponding to the operators $\{\hat{h}_i\}$. The minimization yields the vector of optimal couplings θ^* to represent $\hat{H}_{\text{bare}}$ with $\hat{H}_{\text{eff}}(\theta^*)$. In general, this is an extremely challenging problem.

Specializing now to quantum impurity models of the type Eq. (1), the KLD can be evaluated by representing the reduced thermal density matrix of the impurity after tracing out the bath, $\hat{\rho}_{\text{imp}} = \text{tr}_{\text{bath}}[e^{-\beta \hat{H}} / \mathcal{Z}]$, as a *classical* probability distribution. The fact that $\hat{\rho}_{\text{imp}}$ can be viewed as a classical probability distribution can be seen as follows. Using the hybridization perturbation expansion [55], one can write the partition function $\mathcal{Z}_{\text{bare}}$ of $\hat{H}_{\text{bare}}$ as a sum over the weights w_x of all impurity occupation diagrams, denoted x . We further subdivide each such diagram x into a sum over all corresponding diagrams $w(\{\alpha\}_x)$ labeled in terms of impurity eigenstates $\{\alpha\}_x$,

$$\begin{aligned} \mathcal{Z} &= \mathcal{Z}^{\text{bath}} \sum_x \sum_{\{\alpha\}_x} w(\{\alpha\}_x), \\ w(\{\alpha\}_x) &= e^{-\langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x}} \Lambda_{\{\alpha\}_x} \det(\Delta_x), \end{aligned} \quad (4)$$

where $w(\{\alpha\}_x)$ is the weight of a distinct Feynman diagram labeled by the eigenstate diagram $\{\alpha\}_x$, Δ_x is the antiperiodic hybridization matrix [55], $\Lambda_{\{\alpha\}_x}$ the sequence of impurity operators comprising the diagram $\{\alpha\}_x$ while being projected onto the eigenbasis of $\hat{H}^{\text{imp}}$, and $\langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x}$ is the average value of the impurity Hamiltonian over the diagram $\{\alpha\}_x$. We have also defined $\mathcal{Z}^{\text{bath}}$ as the partition function of the free (decoupled) bath $\hat{H}^{\text{bath}}$. The details of the derivation and precise definitions of the terms appearing in Eq. (4) are given in Appendix A.

From Eq. (4) we extract the distribution

$$P(\{\alpha\}_x) = (\mathcal{Z}^{\text{bath}} / \mathcal{Z}) w(\{\alpha\}_x),$$

which can be interpreted as a classical probability distribution provided that $w(\{\alpha\}_x) > 0$. This distribution acts as a proxy for the impurity density matrix and hence also for the impurity Hamiltonian. As with the classical Boltzmann machine, the probability distribution P is in the form of an energy-based model, with the weights $w(\{\alpha\}_x)$ here distributed according to the impurity Hamiltonian, $\hat{H}^{\text{imp}}$. We can therefore evaluate the KLD

$$D_{\text{KL}}[\hat{H}_{\text{bare}} : \hat{H}_{\text{eff}}(\theta)] = \sum_x \sum_{\{\alpha\}_x: \text{ad}} P_{\text{bare}}(\{\alpha\}_x) \log \left[\frac{P_{\text{bare}}(\{\alpha\}_x)}{P_{\text{eff}}(\{\alpha\}_x)} \right], \quad (5)$$

where we have used the term *admissible* (ad) to denote diagrams that involve eigenstates of $\hat{H}_{\text{bare}}^{\text{imp}}$ for which $\hat{H}_{\text{eff}}^{\text{imp}}$ has analogous eigenstates. This is important because the dimensionality of the impurity subspace is lower for the effective model than in the bare model, by construction. To minimize the distinguishability between $\hat{H}_{\text{bare}}$ and $\hat{H}_{\text{eff}}(\theta)$, we use gradient descent (GD) methods [11]. For this, we need the analytic gradient of the KLD,

$$\begin{aligned} \nabla_{\theta} D_{\text{KL}}[\hat{H}_{\text{bare}} : \hat{H}_{\text{eff}}(\theta)] &= \beta \langle \hat{\Omega}_{\text{ad}}^{\dagger} \nabla_{\theta} \hat{H}_{\text{eff}} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} \\ &\quad - \beta \langle \nabla_{\theta} \hat{H}_{\text{eff}} \rangle_{\text{eff}}, \end{aligned} \quad (6)$$

where the *admissibility operator* $\hat{\Omega}_{\text{ad}}$ connects the effective and bare Fock space by mapping the effective eigenstates to analogous bare eigenstates (see Sec. II A for a detailed derivation and discussion). Thus, $\hat{\Omega}_{\text{ad}}$ eliminates all nonadmissible states from the bare thermal average such that $\langle \hat{\Omega}_{\text{ad}}^{\dagger} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} < 1$, whereas on the effective Hilbert space it holds that

$$\langle \hat{\Omega}_{\text{ad}} \hat{\Omega}_{\text{ad}}^{\dagger} \rangle_{\text{eff}} = \langle \hat{\mathbb{1}} \rangle_{\text{eff}} = 1. \quad (7)$$

From Eq. (6) it follows immediately that the minimum is found when the impurity observables match,

$$\langle \hat{\Omega}_{\text{ad}}^{\dagger} \hat{h}_i \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} = \langle \hat{h}_i \rangle_{\text{eff}} \quad (8)$$

for all effective impurity operators $\hat{h}_i$ of the effective model (from now on we simplify the notation $\langle \hat{\Omega}_{\text{ad}}^{\dagger} \hat{h}_i \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} \rightarrow \langle \hat{h}_i \rangle_{\text{bare}}$). Thus, in the precise sense of Eq. (5), the optimal low-energy effective model $\hat{H}_{\text{eff}}$ for a given bare model $\hat{H}_{\text{bare}}$ is the one that matches the thermodynamic expectation values for all effective impurity operators $\hat{h}_i$ in the bare model. We emphasize that the KLD Eq. (5) does not itself ever have to be evaluated. For the GD optimization, it is sufficient to evaluate the gradient of the KLD, which can be expressed in terms of the physical and readily computable observables in Eq. (6).

We refer to the process of optimizing an effective model to best represent the low-energy physics of our microscopic bare Hamiltonian by matching these key observables as unsupervised model learning (UML).

A particularly attractive feature of UML for quantum impurity systems is the efficiency of the optimization problem. This follows from the fact that the second-order derivative of the KLD corresponds to the second-order derivative of the effective free energy $\mathcal{F}_{\text{eff}}(\theta) = -1/\beta \log \text{tr}[\exp(-\beta \hat{H}_{\text{eff}}(\theta))]$. Therefore, provided all operators $\{\hat{h}_i\}$ mutually commute, the KLD is convex—see Appendix A. Thus we have a single, global minimum, which a gradient descent algorithm is guaranteed to find, making this a trivial optimization problem. We note that the gradient obtained from the KLD, Eq. (6), is equivalent to the gradient obtained in the *maximum entropy principle* that was recently applied to the problem of Hamiltonian tomography in Ref. [22].

The optimization problem is efficient, given the convexity of Eq. (5). The computational cost of the optimization is therefore mainly controlled by the method used to compute the observables comprising the gradient in Eq. (6). The bottleneck is then the calculation of the observables $\langle \hat{h}_i \rangle_{\text{bare}}$ in the bare model. However, this needs to be done just once [e.g., with quantum Monte Carlo (QMC) methods [48,56]]. On the other hand, although multiple calculations of $\langle \hat{h}_i \rangle_{\text{eff}}$ are required during the optimization, these are performed on the simplified effective model and are therefore by construction inexpensive [they can be done with, e.g., the numerical renormalization group (NRG) method [49–51]]. For the results presented in the following, we have used continuous-time QMC (CT-QMC) and NRG to calculate the observables and the Adagrad or Adam GD methods [57].

II. DESIGNING AN EFFECTIVE MODEL

The UML methodology described above makes no assumptions about the form of the bare impurity Hamiltonian (for example, it could be an *ab initio* type model or in tight-binding form). However, for illustrative purposes, and without loss of generality, we assume here that the bare impurity Hamiltonian is akin to a *Pariser-Parr-Pople* (extended Hubbard) model [58],

$$\hat{H}_{\text{bare}}^{\text{imp}} = \sum_{i,j,\sigma} \lambda_{ij} \hat{d}_{i\sigma}^\dagger \hat{d}_{j\sigma} + \sum_i U_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} + \sum_{i>j} U'_{ij} \hat{n}_i \hat{n}_j, \quad (9)$$

where $\hat{d}_{i\sigma}^\dagger$ ($\hat{d}_{i\sigma}$) creates (annihilates) a spin- σ electron in impurity orbital i , whereas $\hat{n}_{i\sigma} = \hat{d}_{i\sigma}^\dagger \hat{d}_{i\sigma}$ and $\hat{n}_i = \sum_\sigma \hat{n}_{i\sigma}$ are number operators. This form of $\hat{H}_{\text{bare}}^{\text{imp}}$ is particularly suited to describe molecules in the context of single-molecule transistors. In the following, we use the term “molecule” to describe the impurity degrees of freedom, as this emphasizes the important role of orbital complexity in the bare model. The total thermodynamic average charge on the molecule, $\langle \hat{Q} \rangle \equiv Q$, can be controlled externally by means of a gate voltage $\hat{H}^{\text{gate}} = V_g \hat{Q}$, where $\hat{Q} = \sum_i \hat{n}_i$.

In practice, the molecule degrees of freedom can be coupled to several noninteracting fermionic baths (metallic leads). For simplicity and concreteness, we now consider one single bath with particle-hole symmetry, $\hat{H}^{\text{bath}} = \sum_{k,\sigma} \epsilon_k \hat{c}_{k\sigma}^\dagger \hat{c}_{k\sigma}$. We

omit the “bare” label here because the bath will be common to both bare and effective models. Without any loss of generality, we cast this in the form of a 1D tight-binding chain,

$$\hat{H}^{\text{bath}} = \sum_{n=0}^{\infty} \sum_{\sigma} t_n (\hat{c}_{n\sigma}^\dagger \hat{c}_{n+1\sigma} + \hat{c}_{n+1\sigma}^\dagger \hat{c}_{n\sigma}), \quad (10)$$

where the chain parameters $\{t_n\}$ encode the bath density of states. This tridiagonal form can be derived explicitly using the Lanczos method [59]. On-site potentials are zero here due to particle-hole symmetry, but they may be reinstated if needed without affecting the following discussions. For the RG analysis below, we note that the *Wilson chain* [49] takes exactly the same form as Eq. (10), but with chain parameters that behave as $t_n \sim D\Lambda^{-n/2}$ for large n , where D is the bare conduction electron half-bandwidth, and $\Lambda > 1$ is the logarithmic discretization parameter. In that case, we replace $\hat{H}^{\text{bath}} \rightarrow \hat{H}_{\text{disc}}^{\text{bath}}$.

The impurity degrees of freedom couple to the bath at the end of the chain via the hybridization term,

$$\hat{H}_{\text{bare}}^{\text{hyb}} = \sum_{i,\sigma} V_i (\hat{d}_{i\sigma}^\dagger \hat{c}_{0\sigma} + \hat{c}_{0\sigma}^\dagger \hat{d}_{i\sigma}). \quad (11)$$

We denote the collection of all bare Hamiltonian couplings using the vector

$$\mathbf{C} \equiv \{\lambda_{11}, \lambda_{12}, \dots, U_1, U_2, \dots, U'_{11}, U'_{12}, \dots, V_1, V_2, \dots, V_g\}.$$

By varying V_g , the ground state of the “molecule” can be tuned to a specific charge sector, which may be spin-degenerate. Of particular interest is the case in which the molecule hosts an odd number of electrons and the ground state spin is $S = \frac{1}{2}$, although the arguments are general. With E_{GS} the energy of the (possibly degenerate) ground state and $E_{\text{ex},1}$ the energy of the first excited state, the active part of the molecule Fock space $\mathcal{H}_{\text{bare}}^{\text{imp}}$ at low temperatures $T \ll E_{\text{ex},1} - E_{\text{GS}}$ is dominated by the Fock space of the ground-state manifold. We can then map the full impurity Fock space of our bare model to an effective one spanning only the ground-state manifold,

$$\mathcal{H}_{\text{bare}}^{\text{imp}} \rightarrow \mathcal{H}_{\text{eff}}^{\text{imp}}. \quad (12)$$

In terms of effective Hamiltonians, the above argument implies that we can map $\hat{H}_{\text{bare}}^{\text{imp}} + \hat{H}_{\text{bare}}^{\text{hyb}} \rightarrow \hat{H}_{\text{eff}}^{\text{imp}} + \hat{H}_{\text{eff}}^{\text{hyb}}$, with the bath Hamiltonian left unchanged. In particular, for a spin-degenerate ground state, the entire molecule can be represented by a single spin $\hat{S}$ degree of freedom, and the molecule-bath hybridization term becomes a spin-exchange interaction. This mapping constitutes an immense reduction in complexity: for a molecule with N orbitals, the dimension of the impurity Fock space is reduced from 4^N to $2S + 1$. In the following, we discuss the details of how this mapping is achieved in practice, and we provide a simple example demonstration.

For some parametrized effective model $\hat{H}_{\text{eff}}(\theta)$ we seek to optimize the coupling constants θ so that properties of the bare model are faithfully reproduced at low energies (as in Fig. 1). A crucial aspect of the GD optimization of Eq. (5) is to compare bare and effective physical observables. This must be done with care, since bare and effective models live in different Fock spaces and are defined at different energy scales. In the case of dynamical quantities such as Green’s functions, a

meaningful comparison of renormalized correlation functions can be achieved by rescaling procedures [60]. But for static quantities, like the thermodynamic expectation values we are comparing, subtleties arise. In particular, a map between bare and effective eigenstates given by the admissibility operator $\hat{\Omega}_{\text{ad}}$ is required.

A. Admissibility

Effective model operators are only defined on the effective model Fock space $\mathcal{H}_{\text{eff}}^{\text{imp}}$, and thus to apply them to a state in the bare Fock space $\mathcal{H}_{\text{bare}}^{\text{imp}}$, we must project the states using the admissibility operator. One approach to construct the admissibility operator is to label bare and effective eigenstates with quantum numbers (QNs) corresponding to the symmetries that are common to both models,

$$\begin{aligned}\hat{H}^{\text{imp}}|n\rangle &= E_n|n\rangle \mapsto \hat{H}^{\text{imp}}|\mathbf{Q}, \mathbf{q}; m_{\mathbf{Q}, \mathbf{q}}\rangle \\ &= E_{\mathbf{Q}, \mathbf{q}; m_{\mathbf{Q}, \mathbf{q}}}|\mathbf{Q}, \mathbf{q}; m_{\mathbf{Q}, \mathbf{q}}\rangle,\end{aligned}$$

where $\mathbf{Q} = (Q_1, Q_2, \dots)$ denotes a vector of non-Abelian QNs, and $\mathbf{q} = (q_1, q_2, \dots)$ is a vector of Abelian QNs. The label $m_{\mathbf{Q}, \mathbf{q}} \in \{1, 2, \dots, M_{\mathbf{Q}, \mathbf{q}}\}$ denotes an index that distinguishes multiplets with the same set of QNs. $M_{\mathbf{Q}, \mathbf{q}}$ is the number of such multiplets in a given sector, which is in general smaller in the effective model than in the bare model $M_{\text{eff}} \leq M_{\text{bare}}$. This is required in order to have a meaningful reduction in complexity.

By labeling our states with these QNs, we can identify which bare states are *admissible* (“ad” for short) on the level of symmetries. Importantly, we can remove *inadmissible* states in the bare model that have no analog in the effective model. For example, the full molecule Fock space of the bare model may contain high-lying spin $S = 1$ states, whereas the effective model retains only the ground states with spin $S = \frac{1}{2}$. Therefore, in this case only $S = \frac{1}{2}$ states of the bare model would be admissible. We regard any state that is labeled with a QN combination that exists in the effective model Fock space as an admissible state of the bare model.

For illustration, we now assume that $M_{\text{eff}} = 1$. Then the admissibility operator can be written as a projector,

$$\hat{\Omega}_{\text{ad}} = \sum_{\mathbf{Q}, \mathbf{q}: \text{ad}} \left[|\mathbf{Q}, \mathbf{q}\rangle_{\text{eff}} \times \sum_{m_{\mathbf{Q}, \mathbf{q}}=1}^{M_{\text{bare}}} \langle \mathbf{Q}, \mathbf{q}; m_{\mathbf{Q}, \mathbf{q}} |_{\text{bare}} \right], \quad (13)$$

where the notation “ $\mathbf{Q}, \mathbf{q} : \text{ad}$ ” indicates that the sum runs only over the quantum numbers that label the eigenstates of $\hat{H}_{\text{eff}}^{\text{imp}}$. Thus, operators $\hat{h}_i$ of the effective model can be meaningfully computed in the bare Fock space, via

$$\chi_i = \langle \hat{\Omega}_{\text{ad}}^\dagger \hat{h}_i \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}}. \quad (14)$$

We provide specific examples of this in action shortly.

There are exceptions where the reference-state quantum numbers differ in the bare and effective models and a little more care has to be taken to make them comparable. In particular, this applies to the definition of the charge quantum number, since the charge depends on the number of (occupied, fermionic) impurity degrees of freedom. But the number of degrees of freedom is smaller in the effective model than the bare one. For ex-

ample, a molecule with N orbitals at half-filling has a ground-state charge $Q = N$. But the effective model may replace these degrees of freedom with a single local moment spin $S = 1/2$, which can be regarded as a fermionic site subject to the constraint that the charge is $Q = 1$. Indeed, the bare model may have $S = 1/2$ representations in multiple charge sectors. One practical solution is to take only the charge sector $\tilde{Q}$ of the bare model which contains the lowest-energy $S = 1/2$ states. Alternatively, one can sum over all charge sectors $\tilde{Q}$ that contain $S = 1/2$ states. Different spin- S multiplets with the same $\tilde{Q}$ in the bare model can also be summed over.

B. RG analysis

The UML approach yields the best couplings θ^* for the chosen Hamiltonian operators $\{\hat{h}_i\}$ in the effective model, Eq. (3), given the target Hamiltonian $\hat{H}_{\text{bare}}$. It may be advantageous to choose $\{\hat{h}_i\}$ in the most general and unbiased way possible, for maximum expressiveness and to facilitate automation of the process. However, we may also utilize physical insights in making this choice to target the desired physics.

On the other hand, the RG procedure provides a systematic way to identify the structure of certain effective models. We briefly summarize the basics of the RG method below, as it will be useful in framing the following discussion. For a full description, see Ref. [30].

The Wilsonian RG map [28] R_l for a Hamiltonian of the form $\hat{H}(\mathbf{C}) = \sum_i \theta_i \hat{h}_i$ [Eq. (3)], with couplings $\mathbf{C} = (\theta_1, \theta_2, \dots)$ and associated interaction terms $\{\hat{h}_i\}$, allows one to eliminate correlations between states above and below a running energy scale cutoff $\exp(-l)D < D$, where D is the bare model cutoff and $l > 1$. In the context of impurity models, D is the conduction electron half-bandwidth. As the map R_l is successively applied, we therefore focus on progressively lower and lower energy scales. The RG map produces a new Hamiltonian that is characterized by *renormalized* couplings $\mathbf{C}'$,

$$R_l[\hat{H}(\mathbf{C})] = \hat{H}(\mathbf{C}'), \quad (15)$$

while leaving the free energy invariant,

$$\mathcal{F}(\mathbf{C}) = \mathcal{F}(\mathbf{C}'). \quad (16)$$

Within this framework, all terms $\hat{h}_i$ consistent with the symmetries and Fock space of the system are allowed to appear in $\hat{H}(\mathbf{C})$. For simplicity, we assume that the set of interactions $\{\hat{h}_i\}$ is finite and does not change during the RG procedure. The continuous flow of the couplings $\mathbf{C}$ is described by the infinitesimal RG-transformation

$$\frac{d\mathbf{C}}{dl} = \lim_{\delta l \rightarrow 0} \frac{1}{\delta l} [R_{l+\delta l}(\mathbf{C}) - \mathbf{C}] \equiv \beta_l(\mathbf{C}), \quad (17)$$

with β_l being the Gell-Mann–Low β -function [30].

An important property of RG is the existence of fixed points (FPs), which are characterized by a set of couplings $\mathbf{C}^*$ that are invariant under the RG transformation,

$$R_l[\hat{H}(\mathbf{C}^*)] = \hat{H}(\mathbf{C}^*) \equiv \hat{H}^*, \quad (18)$$

and we call $\hat{H}^*$ the FP Hamiltonian. The RG transformation can be linearized near the FP Hamiltonian,

$$R_l[\hat{H}^* + \delta\hat{H}] = \hat{H}^* + \mathcal{R}_l^* \delta\hat{H} + O(\delta\hat{H}^2), \quad (19)$$

where $\delta\hat{H}$ is a perturbation to the fixed point. The FP can be further analyzed by finding the eigensystem of the linearized RG transformation, $\mathcal{R}_l^* \cdot \hat{O}_i^* = \lambda_i^* \hat{O}_i^*$. The eigenvalue λ_i^* tells us whether a given operator $\hat{O}_i^*$ drives the system away from the FP (for $|\lambda_i^*| > 1$) or brings us closer to the FP (for $|\lambda_i^*| < 1$). Physical perturbations to the FP can be decomposed in terms of these operators, $\delta\hat{H} = \sum_i \alpha_i \hat{O}_i^*$. If there are finite contributions to $\delta\hat{H}$ from operators with $|\lambda_i^*| > 1$, the perturbation is said to be *relevant*. Conversely, if there are only contributions from operators with $|\lambda_i^*| < 1$, the perturbation is *irrelevant*.

Within this framework, we can identify a class of *minimal* effective models. As in Ref. [46], we define a minimal effective model here as one that retains only the most RG relevant corrections to the low-energy stable FP. Elaborated effective models can also be formulated by including leading RG irrelevant operators. By contrast, *minimally constrained* effective models contain all possible operators consistent with the symmetries of the bare model that live in the reduced Fock space. We consider these various scenarios below.

We emphasize that although the RG procedure described above provides useful information on the *structure* of FP Hamiltonians and their leading corrections, the RG equations for interacting quantum impurity models cannot be solved analytically exactly. Perturbative RG methods give useful insights [61] but cannot provide quantitatively accurate predictions for effective model parameters in versatile settings. Nonperturbative solutions such as the numerical renormalization group [49] allow exact results to be extracted [62] but are only viable for sufficiently simple bare models, where $\hat{H}_{\text{bare}}^{\text{imp}}$ can be diagonalized exactly. In this paper, we are interested precisely in the class of bare models that are beyond the reach of a direct solution using NRG, or at low temperatures inaccessible to QMC.

C. Schrieffer-Wolff revisited

A paradigmatic example for the derivation of an effective model is the mapping of the bare symmetric single Anderson impurity model (AIM) to the effective Kondo model. The former features a spinful fermionic interacting orbital $\hat{d}_\sigma$ on the impurity, whereas the latter involves only a spin-half impurity operator $\hat{S}_d$. The mapping eliminates the impurity charge states with zero or two electrons, and thus reduces the four-dimensional impurity Fock space of the AIM to the two-dimensional Fock space of the Kondo impurity spin.

We take the AIM as our bare Hamiltonian,

$$\begin{aligned} \hat{H}_{\text{bare}}^{\text{AIM}} = & \hat{H}_{\text{disc}}^{\text{bath}} + \frac{1}{2} U_d (\hat{n}^d - 1)^2 \\ & + V \sum_{\sigma} (\hat{d}_{\sigma}^{\dagger} \hat{c}_{0\sigma} + \hat{c}_{0\sigma}^{\dagger} \hat{d}_{\sigma}), \end{aligned} \quad (20)$$

where $\hat{n}^d = \sum_{\sigma} \hat{d}_{\sigma}^{\dagger} \hat{d}_{\sigma}$. As defined, the model possesses particle-hole symmetry and is hence at half-filling (the ground state has impurity charge $Q = 1$). For $U_d \gg \Gamma$, with $\Gamma = \pi V^2 / 2D$ the impurity-bath hybridization, local moment

formation on the impurity is expected. Under these circumstances, second-order perturbation theory in the impurity-bath coupling V yields the Kondo model,

$$\hat{H}_{\text{eff}}^{\text{K}} = \hat{H}_{\text{disc}}^{\text{bath}} + J \hat{S}_d \cdot \hat{S}_0, \quad (21)$$

where $\hat{S}_0 = \frac{1}{2} \sum_{\sigma, \sigma'} \sigma_{\sigma\sigma'} c_{0\sigma}^{\dagger} c_{0\sigma'}$ is the conduction electron spin density at the impurity position. This is the famous Schrieffer-Wolff transformation (SWT) [33,63]. The Kondo model is as such the *minimal* effective model for the AIM, in the sense defined in the previous section.

The second-order perturbative calculation yields an estimate for the spin-exchange coupling, $J_{\text{SW}} = 8V^2/U_d$, which is naturally a rather crude approximation at finite U_d/Γ . Indeed, as discussed in Ref. [46], physical properties such as the thermodynamic entropy and Kondo temperature of the effective model Eq. (21) computed with J_{SW} do not match well those of the bare AIM Eq. (20) unless $U_d \gg D$. In fact, even when the SWT is carried out to infinite order in the impurity-bath hybridization [64], the predicted value of J still does not give accurate results for low-energy properties [46]. The reason for this is well known: the SWT does not account for the bandwidth renormalization $D \rightarrow D_{\text{eff}}$. In the small U_d/D limit, Haldane found a perturbative estimate for this effect [65], with the intuitive result that $D_{\text{eff}} \sim U_d$. This was put on a more concrete footing by Krishnamurthy *et al.* in Ref. [49] using RG arguments. Only recently in Ref. [46] was the precise form determined numerically by nonperturbative methods for all U_d and V . The conclusion is that even with this highly simplified minimal effective model, the free energy and hence low-temperature thermodynamics can be faithfully reproduced—provided the correct effective coupling J is found.

D. $\mathcal{F}$ -learning for the Kondo model

Since the AIM and the Kondo model are connected by the RG flow of the AIM, the low-temperature thermodynamics of the AIM, such as the impurity entropy flow, magnetic susceptibility, and the Kondo temperature T_K , can be exactly reproduced in an effective Kondo model with a properly chosen coupling constant J , provided $T_K \ll U_d$. As shown in Ref. [46], this “true” J can be found numerically exactly by solving the AIM and Kondo models nonperturbatively using, e.g., NRG, and matching their free energies

$$\mathcal{F}^{\text{eff}}(J) = \mathcal{F}^{\text{bare}}(U_d, V), \quad (22)$$

which according to Eq. (16) also matches their RG flows. This should be done at sufficiently low temperatures $T \ll U_d$ where the RG flows of the two models are expected to coincide. We refer to this model learning technique as $\mathcal{F}$ -learning. This is a powerful method if one is interested in reproducing the low-temperature thermodynamics or low-frequency dynamical correlation functions of some bare model using a simpler effective model—provided that both bare and effective models can be solved accurately down to a temperature $T \ll U_d$. However, for complex impurity problems, such a direct solution of the bare model might not be feasible. Furthermore, a coinciding free energy in bare and effective models does *not* guarantee that all local thermodynamic observables will agree, as discussed in the next subsection. If

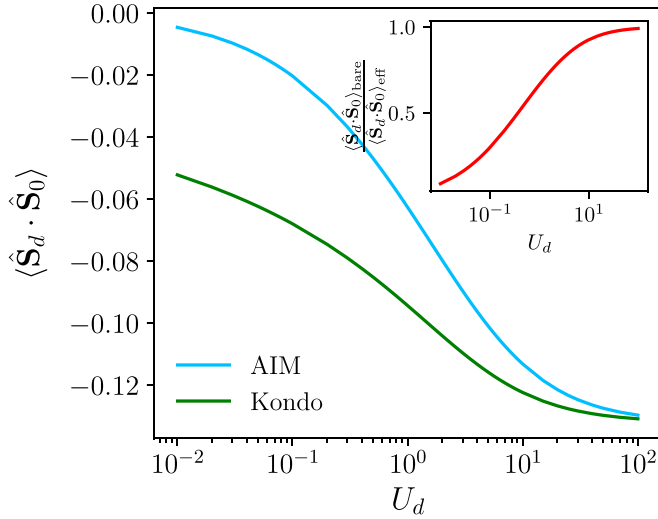


FIG. 2. Impurity-bath spin-spin expectation value $\langle \hat{S}_d \cdot \hat{S}_0 \rangle$ computed with NRG at $T = 0$ for the AIM Eq. (20) (blue line) compared with the corresponding value in the Kondo model Eq. (21), optimized by $\mathcal{F}$ -learning (green). The inset shows their ratio, which captures the different renormalizations along the two paths to the same low-energy fixed point. Plotted as a function of U_d for $8V^2/U_d = 0.3$ and $D = 1$.

such observables are the desired outcome from an effective model, $\mathcal{F}$ -learning alone is not the best choice.

E. Renormalization of observables

Thermodynamic observables may themselves *flow* under RG, and the full path taken, from high to low energies, affects the result. Since the effective model is defined on an energy scale below some cutoff, it cannot capture renormalization effects in the bare model at energies above this cutoff. The important implication of this is that we cannot expect even an optimal minimal effective model to capture simultaneously low-energy scales such as the Kondo temperature T_K , as well as the behavior of thermal expectation values. The former is controlled by the low-temperature free energy, whereas the latter depends on the entire RG flow.

We demonstrate this point explicitly for the simplest example of the mapping from the AIM to the Kondo model. Using $\mathcal{F}$ -learning [46], we extracted the effective J of the Kondo model that reproduces exactly the low-temperature thermodynamics and Kondo temperature T_K for reference AIM systems with different bare U_d (keeping V^2/U_d constant and setting $D \equiv 1$). For each bare AIM and corresponding $\mathcal{F}$ -optimized Kondo model, we computed with NRG the zero-temperature expectation value of the spin-spin correlator $\langle \hat{S}_d \cdot \hat{S}_0 \rangle$. The results are shown in the main panel of Fig. 2, demonstrating markedly different renormalization of this observable in bare and effective models, especially at small U_d . The ratio of the observables calculated in bare and effective models, shown in the inset, characterizes the different degree of renormalization in the two models along the RG flow. Only for $U_d \gtrsim 100D$ do the observables agree; this is consistent with the fact that the perturbative SWT becomes asymptotically exact in the large U_d/Γ limit.

One can understand this result using the framework of Wilsonian RG described above. It is instructive to interpret the RG procedure as a reparametrization of the free energy, $\mathcal{F}$. In this picture, the renormalization is achieved by replacing the original coordinate system $\mathbf{C}$ with a new coordinate system $\mathbf{C}'$ that leaves $\mathcal{F}$ invariant, as required by Eq. (16). Thus we have $\mathbf{C} \rightarrow \mathbf{C}'(\mathbf{C})$ and thermodynamic observables can be interpreted as covariant tensors that can be transformed as such between coordinate systems [60]. For the effective model Eq. (3), we may therefore write

$$\begin{aligned} \langle \hat{h}_i \rangle_{\text{eff}} &= \frac{\partial}{\partial \theta_i} \mathcal{F}_{\text{eff}}(\boldsymbol{\theta}) \\ &= \sum_j \frac{\partial \theta_j^*}{\partial \theta_i} \frac{\partial}{\partial \theta_j^*} \mathcal{F}_{\text{eff}}(\boldsymbol{\theta}^*) = \sum_j \chi_{\theta_j}^* \frac{\partial \theta_j^*}{\partial \theta_i}, \end{aligned} \quad (23)$$

with $\boldsymbol{\theta}^*$ being the renormalized effective model parameters at the low-energy FP, and where $\chi_{\theta_j}^*$ is the expectation value of the operator $\hat{h}_j$ evaluated in the FP Hamiltonian. The matrix of derivatives $\frac{\partial \theta_j^*}{\partial \theta_i}$ contains information about the RG flow to the FP, and can in principle be obtained from the β -function, Eq. (17). With this information we can make a connection between the renormalized value of the observable and its FP value. The desired observable is found to be a linear combination of all possible FP expectation values.

Taking again the example of the AIM to Kondo model mapping, the low-energy stable FP (of both models) is of course the strong-coupling (SC) FP [28,49]. In the effective Kondo model, the SC FP Hamiltonian is characterized by $J^* \rightarrow \infty$. Therefore, $\chi_J^* \equiv \partial_{J^*} \mathcal{F}_{\text{eff}}(J^*) = \langle \hat{S}_d \cdot \hat{S}_0 \rangle^*$ evaluated at the SC FP takes the value

$$\chi_J^* = -3/4, \quad (24)$$

which as such embodies the spin-singlet formed between the impurity and bath. However, the actual value of the correlator as measured in the Kondo model, $\chi_J \equiv \partial_J \mathcal{F}_{\text{eff}}(J)$, carries information about the RG flow,

$$\langle \hat{S}_d \cdot \hat{S}_0 \rangle = -\frac{3}{4} \frac{\partial J^*}{\partial J}. \quad (25)$$

It is tempting to use perturbative scaling techniques [66] to evaluate the nontrivial factor $\frac{\partial J^*}{\partial J}$, but we find that these yield rather poor approximations to the true value of the correlator obtained by NRG.

To understand the flow of the same observables $\langle \hat{h}_i \rangle$ in the bare model, we have to add $\hat{h}_i$ to the bare model as a source term. This is because the effective interactions $\hat{h}_i$ do not typically appear in the original bare model. Thus we may write

$$\begin{aligned} \langle \hat{h}_i \rangle_{\text{bare}} &= \left. \frac{\partial}{\partial \theta_i} \mathcal{F}_{\text{bare}}(\mathbf{C}; \theta_i) \right|_{\theta_i=0} \\ &= \left[\frac{\partial \theta_i^*}{\partial \theta_i} \frac{\partial}{\partial \theta_i^*} + \sum_j \frac{\partial C_j^*}{\partial \theta_i} \frac{\partial}{\partial C_j^*} \right] \mathcal{F}_{\text{bare}}(\mathbf{C}^*; \theta_i^*) \Big|_{\theta_i=0} \\ &= \chi_{\theta_i}^* \frac{\partial \theta_i^*}{\partial \theta_i} \Big|_{\theta_i=0} + \sum_j \chi_{C_j}^* \frac{\partial C_j^*}{\partial \theta_i} \Big|_{\theta_i=0}, \end{aligned} \quad (26)$$

where $\mathbf{C}^*$ are the bare model FP parameters, and $\chi_{\theta_i}^*$ or $\chi_{C_j}^*$ are expectation values evaluated in the FP Hamiltonian of the bare model.

In the low-energy limit, both bare and effective models share the same FP, by construction. Therefore, in the bare model we may write the FP parameters $\theta_j^* \equiv \theta_j^*(\mathbf{C})$ as functions of the bare model parameters. From this it follows that

$$\langle \hat{h}_i \rangle_{\text{bare}} = \sum_j \chi_{\theta_j}^* \left. \frac{\partial \theta_j^*}{\partial \theta_i} \right|_{\theta_i=0}, \quad (27)$$

where $\chi_{\theta_j}^*$ are the FP expectation values of the effective model operators $\hat{h}_j$. Since the FP is the same, these expectation values are the same in bare and effective models. For the AIM to Kondo mapping, we find

$$\langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle_{\text{eff}} \leq \langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle_{\text{bare}}, \quad (28)$$

as demonstrated numerically in Fig. 2. This implies that

$$\left[\frac{\partial J^*}{\partial J} \right]_{\text{bare}} \bigg/ \left[\frac{\partial J^*}{\partial J} \right]_{J=0} \bigg|_{\text{eff}} \leq 1, \quad (29)$$

where equality is only approached in the large- U_d limit of the AIM where the RG flow is equivalent to that of the Kondo model.

The above has consequences for automated model learning. When optimizing an effective model with UML using GD, Eq. (6), we should match the full set of observables $\langle \hat{h}_i \rangle_{\text{bare}} = \langle \hat{h}_i \rangle_{\text{eff}}$ for all effective model operators $\hat{h}_i$. This in turn implies that we enforce

$$\left[\frac{\partial \theta_j^*}{\partial \theta_i} \right]_{\text{eff}} = \left[\frac{\partial \theta_j^*}{\partial \theta_i} \right]_{\theta_i=0} \bigg|_{\text{bare}}. \quad (30)$$

However, the RG flow to reach the low-energy FP is not the same in bare and effective models since the latter typically has fewer FPs. Therefore, enforcing Eq. (30) will give a solution for which the free energy of the effective model is in general not the same as that of the bare model. The low-temperature thermodynamic entropy, Kondo scales, etc., are therefore only in good agreement if the matched observables do not flow *strongly*. For the Kondo model, we see that $\langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ is in general renormalized rather differently compared with the AIM.

Our conclusion is that the KLD Eq. (5) is only a good measure of Hamiltonian distinguishability if the bare and effective models have similar RG flows. A *minimal* effective model may therefore be poorly suited to this kind of optimization.

The question addressed in the next section is, can more general effective model structures in principle capture the Kondo scale and thermal expectation values simultaneously, at least approximately? If so, can UML be used to automate the optimization while remaining computationally tractable?

III. MINIMALLY CONSTRAINED EFFECTIVE MODELS

We now go beyond minimal models, like the Kondo model, and introduce a family of minimally constrained (MC) effective models. With a given choice for the reduced Fock space of the effective model, MC models are characterized by a completely general operator structure, constrained only to respect

the *symmetries* of the bare model. This allows for the inclusion and parametrization of all possible RG relevant, marginal, and irrelevant operators. We show in the following sections that such an effective model can much more accurately capture the physics of the bare model. Because its RG flow is more comparable with the bare model, an MC model is better able to capture both the behavior of local observables as well as the Kondo temperature and associated universal low-energy Kondo physics. This also means that the UML strategy can be used to find good effective models, when the starting point is an MC model.

To construct an MC model, we first must decide on the reduced Fock space in which the effective model lives. In this regard, we note that by replacing the impurity part of the bare Hamiltonian by operators spanning only the ground-state manifold, we also implicitly change the nature of the impurity-bath hybridization term. Consider, for example, the classic mapping of the AIM to the Kondo model. The impurity degrees of freedom $\hat{d}_\sigma$ are mapped to a spin- $\frac{1}{2}$ operator $\hat{\mathbf{S}}_d$, but the impurity-bath tunneling $V \sum_\sigma (\hat{d}_\sigma^\dagger \hat{c}_{0\sigma} + \text{H.c.})$ in Eq. (20) is also replaced by a spin-spin exchange coupling $J \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$ in Eq. (21).

An alternative way to formulate the mapping is to take an extended view on what constitutes the “impurity” to include also the local bath orbitals to which the bare impurity degrees of freedom couple. In so doing, we also redefine our bath to exclude these local orbitals. We thus repartition the bare model as $\hat{H}_{\text{bare}} = \hat{H}_{\text{bare}}^{\text{imp}} + \hat{H}_1^{\text{bath}} + \hat{H}_1^{\text{hyb}}$, where our extended impurity is now given by

$$\hat{H}_{\text{bare}}^{\text{imp}} = \hat{H}_{\text{bare}}^{\text{imp}} + \hat{H}_{\text{bare}}^{\text{hyb}}. \quad (31)$$

The extended impurity Fock space is thus

$$\tilde{\mathcal{H}}_{\text{bare}}^{\text{imp}} = \mathcal{H}_{\text{bare}}^{\text{imp}} \otimes \mathcal{H}_0^{\text{bath}}. \quad (32)$$

The bath and hybridization terms are correspondingly redefined,

$$\hat{H}_N^{\text{bath}} = \sum_{n=N}^{\infty} \sum_{\sigma} t_n (\hat{c}_{n\sigma}^\dagger \hat{c}_{n+1\sigma} + \hat{c}_{n+1\sigma}^\dagger \hat{c}_{n\sigma}), \quad (33)$$

$$\hat{H}_N^{\text{hyb}} = \sum_{\sigma} t_{N-1} (\hat{c}_{N-1\sigma}^\dagger \hat{c}_{N\sigma} + \hat{c}_{N\sigma}^\dagger \hat{c}_{N-1\sigma}), \quad (34)$$

where $N = 1$ indicates the new starting site index of the redefined bath chain. This has the advantage that the new bath and hybridization terms are always the same in bare and effective models. We have emphasized this by now removing the “bare” label from these terms.

The MC effective model is similarly structured,

$$\hat{H}_{\text{eff}}^{\text{MC}}(\boldsymbol{\theta}) = \hat{H}_{\text{eff}}^{\text{imp}}(\boldsymbol{\theta}) + \hat{H}_1^{\text{bath}} + \hat{H}_1^{\text{hyb}}, \quad (35)$$

where $\hat{H}_{\text{eff}}^{\text{imp}}(\boldsymbol{\theta}) = \sum_i \theta_i \hat{h}_i$ and the effective impurity operators $\hat{h}_i$ live in the extended MC effective Fock space,

$$\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}} = \mathcal{H}_{\text{eff}}^{\text{imp}} \otimes \mathcal{H}_0^{\text{bath}}. \quad (36)$$

As before, $\mathcal{H}_{\text{eff}}^{\text{imp}}$ spans the ground-state manifold of $\hat{H}_{\text{bare}}^{\text{imp}}$, and $\mathcal{H}_0^{\text{bath}}$ is the Fock space of the local bath orbital(s).

A. Corrections to the Kondo model

As the simplest example, we again consider the particle-hole symmetric AIM as our bare model, Eq. (20). The MC effective model can be found by following the above arguments. With a spin- $\frac{1}{2}$ operator $\hat{S}_d$ representing the impurity and $\hat{c}_{0\sigma}$ for the local bath orbital, the effective impurity Fock space $\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}$ is eight-dimensional. Since we have conserved spin S , spin projection S_z , and charge Q , we may label the basis states of the MC effective impurity Hamiltonian by these quantum numbers, $|Q, S; S_z\rangle$. Then $\tilde{H}_{\text{eff}}^{\text{imp}}(\theta)$ assumes a block-diagonal structure, with each quantum number block in this case being simply a 1×1 scalar. The allowed configurations are

Q	S	S_z
± 1	$\frac{1}{2}$	$\pm \frac{1}{2}$
0	0	0
0	1	0, ± 1

where we specify the charge Q with respect to the two-electron half-filled state. Due to SU(2) spin symmetry, states with the same Q and S are degenerate; and due to particle-hole symmetry, states with $Q = \pm 1$ are degenerate. The most general MC effective Hamiltonian defined in this Fock space, consistent with these symmetries, can therefore be parametrized by just three distinct Hamiltonian matrix elements. Because in this case each quantum number block contains a unique basis state, the Hamiltonian operator can be entirely specified in terms of projectors onto these symmetry spaces,

$$\hat{P}_{QS} = \sum_{S_z} |Q, S; S_z\rangle \langle Q, S; S_z|, \quad (37)$$

according to

$$\tilde{H}_{\text{eff}}^{\text{imp}}(\theta) = \theta_{\pm 1, 1/2} (\hat{P}_{-1, 1/2} + \hat{P}_{+1, 1/2}) + \theta_{0, 0} \hat{P}_{0, 0} + \theta_{0, 1} \hat{P}_{0, 1}. \quad (38)$$

These projectors can be written out explicitly as

$$\begin{aligned} \hat{P}_{0, 1} &= |\uparrow, \uparrow\rangle \langle \uparrow, \uparrow| + |\downarrow, \downarrow\rangle \langle \downarrow, \downarrow| + \frac{1}{2} (|\uparrow, \downarrow\rangle \langle \uparrow, \downarrow| \\ &\quad + |\downarrow, \uparrow\rangle \langle \downarrow, \uparrow| + |\downarrow, \uparrow\rangle \langle \uparrow, \downarrow| + |\uparrow, \downarrow\rangle \langle \downarrow, \uparrow|), \\ \hat{P}_{\pm 1, 1/2} &= (|0\rangle \langle 0| + |\uparrow\downarrow\rangle \langle \uparrow\downarrow|) \otimes (|\uparrow\rangle \langle \uparrow| + |\downarrow\rangle \langle \downarrow|), \\ \hat{P}_{0, 0} &= \frac{1}{2} (|\uparrow, \downarrow\rangle \langle \uparrow, \downarrow| + |\downarrow, \uparrow\rangle \langle \downarrow, \uparrow| - |\uparrow, \uparrow\rangle \langle \uparrow, \uparrow| \\ &\quad - |\downarrow, \downarrow\rangle \langle \downarrow, \downarrow|), \end{aligned} \quad (39)$$

where $|\phi, \sigma\rangle \equiv |\phi\rangle_0 \otimes |\sigma\rangle_d$ with $\phi = \{0, \uparrow, \downarrow, \uparrow\downarrow\}$ for the local bath orbital and $\sigma = \{\uparrow, \downarrow\}$ for the impurity spin.

The identity operator can also be resolved in this basis,

$$\hat{\mathbb{1}}_{\text{eff}} = \sum_{Q, S} \hat{P}_{QS}, \quad (40)$$

such that $\langle \hat{\mathbb{1}}_{\text{eff}} \rangle_{\text{eff}} = 1$. We can therefore use the identity operator to eliminate one of the projectors in Eq. (38). As an

example, we eliminate the $\hat{P}_{0, 0}$ projector and write

$$\tilde{H}_{\text{eff}}^{\text{imp}}(\theta) = \theta_{\pm 1, 1/2} (\hat{P}_{-1, 1/2} + \hat{P}_{+1, 1/2}) + \theta_{0, 1} \hat{P}_{0, 1} + \theta_{\perp} \hat{\mathbb{1}}_{\text{eff}}, \quad (41)$$

where $\theta_{\perp} = \theta_{0, 0}$ and the (as yet still unknown) parameters $\theta_{\pm 1, 1/2}$ and $\theta_{0, 1}$ have been suitably rescaled. This is useful because the identity operator strictly does not flow under RG. Thus, the term $\theta_{\perp} \hat{\mathbb{1}}_{\text{eff}}$ cannot change the RG flow of the effective model or affect thermal expectation values of the projectors $\langle \hat{P}_{QS} \rangle_{\text{eff}}$ evaluated in the effective model. Further, setting $\theta_{\perp} = 0$ corresponds to a uniform shift of all energy levels and does not affect the dynamics. For the symmetric AIM, the ground state lives in the $Q = 0, S = 0$ sector, so this choice is equivalent to measuring energies relative to the ground-state energy. Importantly, the UML optimization, which is based on matching thermal expectation values, is sensitive to this elimination of a projector, as the unassuming operator $\hat{\mathbb{1}}_{\text{eff}}$ becomes nontrivial in the bare model $\hat{\Omega}_{\text{ad}}^{\dagger} \cdot \hat{\mathbb{1}}_{\text{eff}} \cdot \hat{\Omega}_{\text{ad}} \neq \hat{\mathbb{1}}_{\text{bare}}$. The intricacies that arise from this sensitivity are discussed in detail in Sec. III B.

In any case, we are left with an MC effective model featuring just two tunable parameters. The physical interpretation of Eq. (41) is manifested by expressing the projectors in second-quantized form,

$$\begin{aligned} \hat{P}_{0, 1} &= \frac{3}{4} (\hat{n}_{0\uparrow} + \hat{n}_{0\downarrow} - 2\hat{n}_{0\uparrow}\hat{n}_{0\downarrow}) + \hat{S}_d \cdot \hat{S}_0, \\ \hat{P}_{0, 0} &= \frac{1}{4} (\hat{n}_{0\uparrow} + \hat{n}_{0\downarrow} - 2\hat{n}_{0\uparrow}\hat{n}_{0\downarrow}) - \hat{S}_d \cdot \hat{S}_0, \\ \hat{P}_{\pm 1, 1/2} &= \hat{\mathbb{1}}_{\text{eff}} - \hat{n}_{0\uparrow} - \hat{n}_{0\downarrow} + 2\hat{n}_{0\uparrow}\hat{n}_{0\downarrow}, \end{aligned} \quad (42)$$

where $\hat{n}_{0\sigma} = \hat{c}_{0\sigma}^{\dagger} \hat{c}_{0\sigma}$. Recombining these projectors yields a different parametrization of Eq. (41) (but with the same expressibility). We write it as

$$\tilde{H}_{\text{eff}}^{\text{imp}}(\theta) \equiv \hat{H}_{\text{eff}}^{JU_0} = J \hat{S}_d \cdot \hat{S}_0 + \frac{1}{2} U_0 (\hat{n}_0 - \hat{\mathbb{1}}_{\text{eff}})^2 + L \hat{\mathbb{1}}_{\text{eff}}, \quad (43)$$

where $\hat{n}_0 = \sum_{\sigma} \hat{n}_{0\sigma}$ and $J = \theta_{0, 1} - \theta_{0, 0}$, $U_0 = 2\theta_{\pm 1, 1/2} - \frac{3}{2}\theta_{0, 1} - \frac{1}{2}\theta_{0, 0}$, $L = \frac{3}{4}\theta_{0, 1} + \frac{1}{4}\theta_{0, 0} + \theta_{\perp}$. Again, the term involving $\hat{\mathbb{1}}_{\text{eff}}$ can be neglected and we set $L = 0$. The effective model written in this way is equivalent to a Kondo model with an additional electron-electron interaction term on the local bath orbital.

Equation (41), and equivalently Eq. (43), are the most general forms of the effective model that live in the extended Fock space $\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}$, consistent with SU(2) spin symmetry, U(1) charge conservation, and particle-hole symmetry.

Although the minimal effective model for the AIM is the Kondo model Eq. (21), the MC effective model $\hat{H}_{\text{eff}}^{\text{MC}}(J, U_0)$, with $\tilde{H}_{\text{eff}}^{\text{imp}}$ parametrized according to Eq. (43), is precisely the Kondo model *plus the leading RG-irrelevant perturbation* to the local moment fixed point identified by Krishnamurthy, Wilkins, and Wilson in Ref. [49]. The extended-impurity approach thus allows us to define more general models that include such RG marginal and irrelevant terms, beyond those obtained by, e.g., a Schrieffer-Wolff transformation.

In terms of the interpretability of effective models learned by UML, we note that the symmetry analysis already offers insights into the kinds of coupling terms that can arise in a given extended impurity space. The fact that the physical impurity hosts a local moment is exploited directly to reduce

the dimensionality of the active impurity space; this has implications for the interactions appearing in the effective model. By formulating the resulting symmetry projectors in second-quantized form, as per Eq. (42), we obtain a more physical interpretation of these terms. Indeed, recombining these in the form of Eq. (43) makes a connection to the RG theory of the low-energy properties of the bare model. However, the numerical values of the optimized coupling constants remain somewhat inscrutable. Only in perturbative limits do these effective parameters take a simple form in terms of the bare model parameters.

When the AIM is the bare model, the extended impurity Fock space dimension is reduced modestly from 16 to 8, but the parametric complexity $(U_d, V) \rightarrow (J, U_0)$ is the same and we do not appear to have gained much. However, we emphasize that the AIM is *not* the intended application of this method. Rather, we are interested in being able to systematically derive expressive effective models for complex impurity systems. We provide examples of this in the following. The AIM is, however, still a useful benchmark test case that we continue to explore.

B. UML optimization of the MC effective model

With the form of the MC effective model now at hand, we turn to the problem of optimizing its parameters using the UML method. As explained in Sec. I, this amounts to matching the observables $\langle \hat{\Omega}_{\text{ad}}^\dagger \hat{h}_i \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} = \langle \hat{h}_i \rangle_{\text{eff}}$ in bare and effective models, where $\hat{h}_i$ are the operators comprising the MC effective model, and $\hat{\Omega}_{\text{ad}}$ is the admissibility operator that projects onto the reduced Fock space of the effective model. The explicit form of these will be discussed below, and an explicit example of the UML optimization is given at the level of the AIM.

First, we note that for the MC effective model [in either form, Eq. (41) or Eq. (43)], we have two observables to match and *two* parameters to learn when comparing with the bare model. In particular, we emphasize that the identity $\hat{\mathbb{1}}_{\text{eff}}$ is *not* a matchable observable, because the Fock spaces of bare and effective models are different by construction. Therefore, although $\langle \hat{\mathbb{1}}_{\text{eff}} \rangle_{\text{eff}} = 1$ in the effective model, we have $\langle \hat{\Omega}_{\text{ad}}^\dagger \cdot \hat{\mathbb{1}}_{\text{eff}} \cdot \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} < 1$ in the bare model. From this it follows that it is in fact *impossible* to demand that

$$\langle \hat{\Omega}_{\text{ad}}^\dagger \cdot \hat{P}_{QS} \cdot \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}} = \langle \hat{P}_{QS} \rangle_{\text{eff}} \quad (44)$$

be satisfied simultaneously for the full set of projectors $\hat{P}_{QS}$. Importantly, we conclude that the optimization problem Eq. (5) is *overdetermined* when using these projectors. Therefore, at least one projector must be omitted from the UML observable matching process. In Eq. (41) we eliminated $\hat{P}_{0,0}$, but other choices can be made.

Another key point is that matching different combinations of observables may yield different results for the optimized effective model parameters. The operator content and scaling dimensions for different observables means they are renormalized differently. Since bare and effective models have different RG flows in the UV, different effective model parameters are in general needed to ensure agreement between for a given set of observables. We return to the systematic selection of

optimal observables below, but here we simply note that observables that flow the *least* are the best choice.

Although the impurity Hamiltonian terms in $\hat{H}_{\text{eff}}^{JU_0}$ in Eq. (43) are arguably more physically meaningful, the decomposition in terms of symmetry projectors in Eq. (41) turns out to be most advantageous for UML. This is because (i) the projectors have a direct meaning in the bare model, since bare and effective models have the same *symmetries*; (ii) the projectors mutually commute, and so the optimization problem is *convex*; and (iii) the action of the admissibility operator on the projectors is straightforward.

1. UML for the AIM

We first look at the AIM as our bare model. As explained above, we have two observables to match within the UML optimization procedure. Working with the projectors in Eq. (41) we tune the effective model parameters $\theta_{\pm 1, 1/2}$ and $\theta_{0,1}$ via GD according to Eq. (6). Importantly, the KLD loss function itself need not be computed at any stage; the optimization can be done with a knowledge of its gradient, which involves only the observables $\langle \hat{P}_{\pm 1, 1/2} \rangle_{\text{eff}}$ and $\langle \hat{P}_{0,1} \rangle_{\text{eff}}$ in the effective model, and $\langle \hat{\Omega}_{\text{ad}}^\dagger \hat{P}_{\pm 1, 1/2} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}}$ and $\langle \hat{\Omega}_{\text{ad}}^\dagger \hat{P}_{0,1} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}}$ in the bare model. The optimal solution is found when both

$$\begin{aligned} \langle \hat{P}_{\pm 1, 1/2} \rangle_{\text{eff}} &= \langle \hat{\Omega}_{\text{ad}}^\dagger \hat{P}_{\pm 1, 1/2} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}}, \\ \langle \hat{P}_{0,1} \rangle_{\text{eff}} &= \langle \hat{\Omega}_{\text{ad}}^\dagger \hat{P}_{0,1} \hat{\Omega}_{\text{ad}} \rangle_{\text{bare}}. \end{aligned} \quad (45)$$

The Q and S quantum numbers have physical meaning in the enlarged Fock space $\mathcal{H}_{\text{bare}}^{\text{imp}}$ of the bare model, and so $\hat{P}_{QS}$ can be defined in the bare model. The admissibility operator projects onto the $S = \frac{1}{2}$ states of the impurity. For the AIM, this means the two local moment impurity states with $n_d = 1$. Because of the simplicity of the AIM and the symmetry constraints, we can construct $\hat{\Omega}_{\text{ad}} = \hat{n}_d - 2\hat{n}_{d\uparrow}\hat{n}_{d\downarrow}$ and use the definitions of the projectors given already in Eq. (42).

A key advantage of UML is that the optimization is done via physical quantities that can be obtained using any numerical method able to compute static thermodynamic observables, including *ab initio* methods. Indeed, different methods can be used for bare and effective models, if desired. For the AIM considered in this section, we use NRG to compute the observables in both bare and effective models at temperature $T = 0$ [50,67]. We consider the role of finite temperature in the next subsection. Note that the (relatively expensive) bare model “reference” observables need to be computed only once. The (computationally cheap) effective model observables are iteratively calculated during the GD until Eq. (45) is satisfied. Once the optimized effective model parameters have been found, the effective model can be fully solved to obtain the physical properties of interest.

In Fig. 3 we show one such example calculation for the AIM to MC model mapping. The temperature dependence of the impurity entropy is shown for the bare AIM (points), compared with that of the optimized MC effective model (lines) for three different values of U_d , keeping $8V^2/U_d = 0.3$ fixed. The inset shows a comparison with UML optimization of only J in the minimal Kondo model Eq. (21) by matching just $\langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ (dashed line), as well as UML optimization of both J and U_0 in Eq. (43) by matching $\langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ and $\langle (\hat{n}_0 - \hat{\mathbb{1}}_{\text{eff}})^2 \rangle$

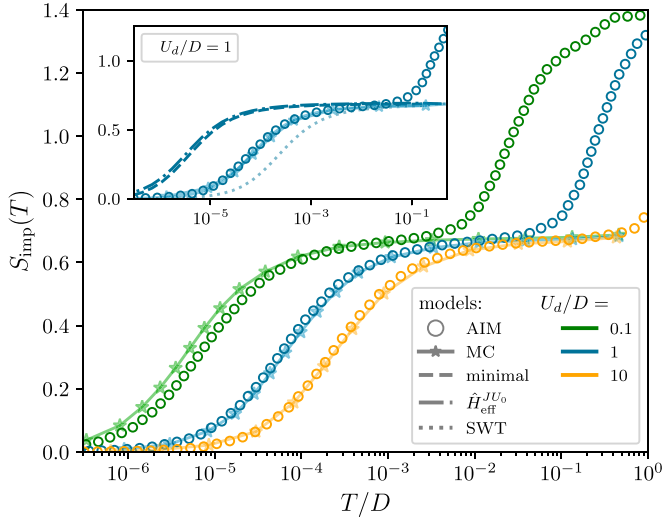


FIG. 3. Demonstration of the UML optimization for the MC effective model. The impurity contribution to the entropy of the bare AIM (points) is computed as a function of T/D ($D = 1$) using NRG, and compared with that of the optimized MC effective model (lines) for fixed $8V^2/U_d = 0.3$ and $U_d/D = 0.1, 1, 10$. Inset shows a comparison for $U_d = 1$ with the minimal Kondo model using the SWT value of J (dotted) and UML-optimized J (dashed), as well as UML-optimized model $\hat{H}_{\text{eff}}^{JU_0}$ (dot-dashed). Training done at $T = 0$ with NRG.

(dot-dashed line). Also given for reference is the minimal Kondo model with J obtained perturbatively from SWT (dotted line).

The numerical results show that UML using the MC effective model Eq. (41) represents a significant improvement over the other methods, already at relatively high U_d (of the order of the bandwidth $U_d \sim D$) where the RG flows of the bare and effective models only mildly differ. However, we also see that the low-energy physics, characterized by the low-temperature crossover scale T_K , is still not perfectly reproduced by UML for $U_d/D \ll 1$. We show in Sec. IV how the results can be systematically improved.

2. Optimizing the set of matched observables

As noted above, there is some ambiguity in Eq. (38) as to which projector to eliminate in the UML observable matching process. In Eq. (42) we somewhat arbitrarily eliminated the projector $\hat{P}_{0,0}$ by using the identity operator, but in principle $\hat{P}_{0,1}$ or $\hat{P}_{\pm 1,1/2}$ could have been eliminated. Here we present a method for identifying which projector should be eliminated, and we validate this numerically by comparing effective models based on different choices. For the case of the MC effective model, it turns out that indeed Eq. (42) is the optimal model. However, for the more complicated effective models considered in Sec. IV, which feature a larger number of observables, one requires a systematic approach

The first step is to define the Frobenius scalar product (FSP) on the operator space $\langle \hat{A}, \hat{B} \rangle = \text{tr}[\hat{A}^\dagger \cdot \hat{B}]$ and the corresponding norm $\|\hat{A}\| = \sqrt{\langle \hat{A}^\dagger, \hat{A} \rangle}$. Here we use the FSP to decompose an operator $\hat{h}_i$ from the effective model into contributions from the low-energy fixed point Hamiltonian of the

bare model $\hat{H}^*$, and contributions from RG irrelevant corrections to the fixed point, $\hat{\Lambda}_i$. We define $\alpha_i = \langle \hat{h}_i, \hat{H}^* \rangle / \|\hat{H}^*\|^2$ and $\beta_i = \langle \hat{h}_i, \hat{\Lambda}_i \rangle / \|\hat{\Lambda}_i\|^2$, noting that $\langle \hat{\Lambda}_i, \hat{H}^* \rangle = 0$. The operator $\hat{\Lambda}_i$ can be found by the Gram-Schmidt procedure $\hat{\Lambda}_i = \hat{h}_i - \alpha_i \hat{H}^*$. We now decompose the projector as $\theta_i \hat{h}_i = \theta_i (\alpha_i \hat{H}^* + \beta_i \hat{\Lambda}_i) \equiv \gamma_i \hat{H}^* + \lambda_i \hat{\Lambda}_i$, which can be used for the substitution,

$$\frac{\partial}{\partial \theta_i} \mathcal{F}_{\text{bare}}(\mathbf{C}; \theta_i) \Big|_{\theta_i=0} = \left[\underbrace{\beta_i \frac{\partial}{\partial \lambda_i}}_{\text{RG irrelevant}} + \underbrace{\alpha_i \frac{\partial}{\partial \gamma_i}}_{\text{RG relevant}} \right] \mathcal{F}_{\text{bare}}(\mathbf{C}; \lambda_i, \gamma_i) \Big|_{\gamma_i=\lambda_i=0}. \quad (46)$$

The coefficients α_i and β_i in Eq. (46) tell us the relative weight of RG relevant and RG irrelevant terms in the decomposition of a given operator $\hat{h}_i$. In particular, the larger α_i is, the stronger $\langle \hat{h}_i \rangle$ flows under RG. For the example of the *minimal* Kondo model with $\hat{h}_i = \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$, we have trivially that $\beta_i = 0$ and the operator is entirely RG relevant, $\alpha_i = 1$. The projectors $\hat{P}_{QS}$ in Eq. (42), on the other hand, have a nontrivial decomposition. To render the overlap parameter α_i comparable for different projectors, we introduce the normalized quantity, $\tilde{\alpha}_i = |\langle \hat{h}_i, \hat{H}^* \rangle| / \|\hat{H}^*\| \|\hat{h}_i\|$. This provides a means to compare the strength of the renormalization of the effective interactions. Since we wish to compare only the most weakly renormalized observables in bare and effective models, we can use the FSP to eliminate the projector with the largest $\tilde{\alpha}_i$. For the effective interactions of $\hat{H}_{\text{eff}}^{JU_0}$ and $\hat{H}_{\text{eff}}^{MC}$ and the fixed point Hamiltonian $\hat{H}^* = \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$, the FSP yields

	$\hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$	$(\hat{n}_0 - 1)^2$	$\hat{P}_{0,1}$	$\hat{P}_{0,0}$	$\hat{P}_{\pm 1,1/2}$
$\tilde{\alpha} \approx$	1	0	0.5	0.87	0

Based on these calculated values of $\tilde{\alpha}_i$, the best effective Hamiltonian is $\hat{H}_{\text{eff}}^{MC}$ with $\theta_{0,0} = 0$, since this excludes the term that gets renormalized the strongest. This is precisely Eq. (42).

We put this hypothesis to the test by performing the UML optimization for all four effective Hamiltonians and comparing the results: Eqs. (43) and (38) with each of the three projectors eliminated in turn. Our numerical results are presented in Fig. 4.

In the top-left of panel (a), we perform UML optimization on $\hat{H}_{\text{eff}}^{JU_0}$, matching the two effective model operators $\hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$ and $(\hat{n}_0 - 1)^2$. The $T = 0$ expectation values of these operators evaluated in bare and effective models using NRG are plotted as a function of U_d . Both observables can be matched in the effective model for all bare model parameters. In the top-right panel of Fig. 4(a) we show instead the UML optimization of $\hat{H}_{\text{eff}}^{MC}$ [Eq. (38)] with $\theta_{0,0} = 0$. Here we match $\langle \hat{P}_{0,1} \rangle$ and $\langle \hat{P}_{\pm 1,1/2} \rangle$, which is also achievable for all U_d . Although $\langle \hat{P}_{0,0} \rangle$ is not matched during the optimization, we can still compare its calculated expectation value in the optimized effective model with that of the bare model. We find, as expected, that they only agree in the

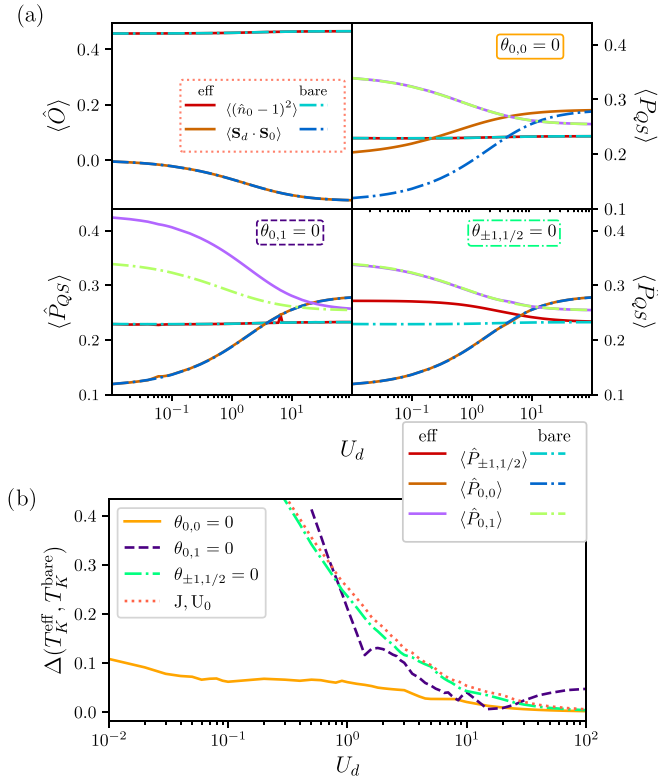


FIG. 4. Analysis of UML optimization of MC effective models for the AIM bare model. (a) Top-left panel: $\hat{H}_{\text{eff}}^{JU_0}$ [Eq. (43)], with observables compared in bare and effective models, as a function of U_d . Other panels: similarly for $\hat{H}_{\text{eff}}^{\text{MC}}$ [Eq. (38)] with $\theta_{0,0} = 0$ (top-right); $\theta_{\pm 1, 1/2} = 0$ (bottom-right); $\theta_{0,1} = 0$ (bottom-left). (b) Relative (logarithmic) error for the Kondo temperature T_K [Eq. (47)] in the optimized model as a function of U_d for the four cases considered in panel (a). The orange line corresponds to Eq. (41). Plotted for fixed $8V^2/U_d = 0.3$ as a function of U_d , with results obtained at $T = 0$ by NRG.

perturbative limit $U_d \gg 1$. It is not possible to match all projectors simultaneously in general. The bottom-left and bottom-right panels of Fig. 4(a) correspond to optimization of Eq. (38) with $\theta_{0,1} = 0$ and $\theta_{\pm 1, 1/2} = 0$, respectively—a similar result pertains.

In Fig. 4(b), we compare the accuracy of these differently optimized effective models by calculating their Kondo temperature T_K , and comparing with that in the bare model. Our figure of merit is the (logarithmic) relative error, defined as

$$\Delta(T_K^{\text{eff}}, T_K^{\text{bare}}) = \left| 1 - \frac{\ln[\max(T_K^{\text{eff}}, T_K^{\text{bare}})]}{\ln[\min(T_K^{\text{eff}}, T_K^{\text{bare}})]} \right|. \quad (47)$$

Given the exponential dependence of the Kondo temperature on the coupling J in the minimal Kondo model [33], Eq. (47) captures the error in the effective J . Our numerical results in Fig. 4(b) show that including the U_0 term as a correction to the Kondo model as per Eq. (43) (dotted line) does not appreciably improve on the results of the minimal Kondo model. Furthermore, UML optimization of the MC effective model Eq. (42), with $\theta_{0,1} = 0$ (dashed) or $\theta_{\pm 1, 1/2} = 0$ (dot-dashed), does not produce markedly better results for $\Delta(T_K^{\text{eff}}, T_K^{\text{bare}})$. We attribute this to the injudicious attempt to match

observables that flow strongly under RG. However, when the most RG relevant operator is eliminated, $\hat{P}_{0,0} = 0$ (orange line), the Kondo temperature is well approximated at all U_d . This shows an enormous improvement over the observable-matching minimal model. We conclude that the FSP is a good way to establish RG relevance of Hamiltonian terms and to identify ill-fitted interactions for the optimization. Observables with weaker flow are better suited for the KLD distinguishability measure. This analysis provides a guide to the proper set of observables to match within UML.

3. Temperature dependence of the optimization

Minimization of the KLD loss function during UML involves computing static thermal expectation values, and comparing these in bare and effective models. In principle, this can be done at any temperature, T . However, one obtains meaningful results only when the temperature is sufficiently low that the *active* Fock space of the bare model is comparable to that of the effective model, $T \ll E_{\text{ex},1} - E_{\text{GS}}$. For the AIM, this means in practice $T \ll U_d$, since then impurity charge fluctuations are frozen, and the description of the impurity as a local moment spin- $\frac{1}{2}$ degree of freedom is applicable.

So far we performed UML by calculating observables at zero temperature using NRG. We note that other methods operating at $T = 0$ that target the ground state can similarly be used, for example quantum-chemical methods [4], the density matrix renormalization group (DMRG) [68–70], or *ab initio* methods based on the variational principle [71,72]. However, here we show that the results of the effective model optimization are rather insensitive to temperature provided $T \ll U_d$. The important consequence of this is that the bare model need not be completely solved down to $T = 0$, or the exact ground state of the many-body system determined. Methods based on imaginary-time evolution that favor relatively high temperatures $T \gtrsim 10^{-2}D$ [48] can therefore be used for complex bare model calculations. Indeed, imaginary-time *ab initio* methods can also be used for a first-principles treatment of the bare model [73,74].

In Fig. 5 we plot the temperature dependence of the optimized MC effective model parameters θ_{QS} [panel (a)] and the corresponding thermal observables $\langle \hat{P}_{QS} \rangle$ [panel (b)]. For this demonstration, we use the AIM as the bare model and use NRG as our impurity solver. Once the AIM reaches the local moment regime for $T/U_d \lesssim 0.1$, the optimized parameters θ_{QS} become essentially temperature-independent. The observables can also be precisely matched in this regime, down to $T = 0$. Demonstrating that the calculated observables can be matched and do not change appreciably with temperature provides an independent check on the convergence of the UML optimization. By contrast, it is clear that a good effective model cannot be formulated for $T/U_d \gtrsim 1$.

4. UML for complex impurities

Finally, we turn to UML optimization of the MC effective model for more complex impurity systems. For a bare model with a unique doublet ground state, Eq. (41) still constitutes a good effective model and UML involves matching the

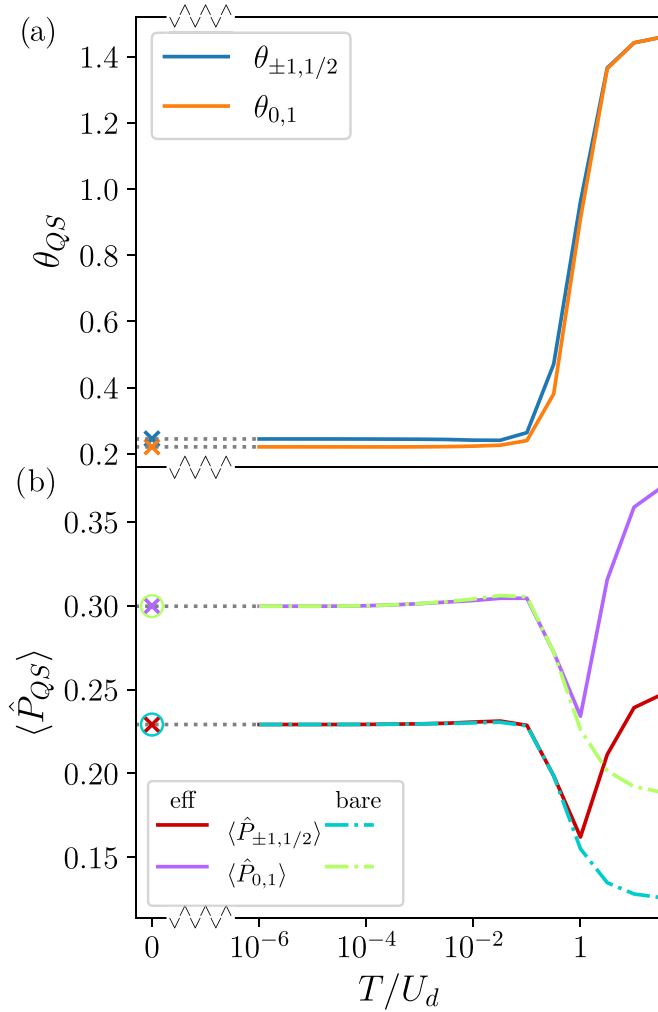


FIG. 5. (a) Evolution of optimized parameters of the MC effective model Eq. (41) with temperature for the AIM with $U_d = 0.4D$ and $8V^2/U_d = 0.3$. (b) Corresponding observables evaluated in the bare and effective models. The logarithmic temperature axis includes a cut to $T = 0$ to show that optimization can be done using ground-state properties if desired.

observables as per Eq. (45). With conserved charge Q and spin S in the bare model, we can again define projectors $\hat{P}_{QS}$. However, care must now be taken because (i) a multiorbital bare model might have a reference charge $\tilde{Q}$ that is not the same as in the MC effective model; (ii) in general the bare model will have several multiplets for a given symmetry subspace defined by the quantum numbers Q and S ; and (iii) the admissibility operator projects onto the physical impurity $S = \frac{1}{2}$ space of $\mathcal{H}_{\text{bare}}^{\text{imp}}$ but the operators are defined in the extended impurity Fock space $\tilde{\mathcal{H}}_{\text{bare}}^{\text{imp}}$.

In the following, we address these issues by interpreting the operators $\hat{P}_{QS} = \hat{\Omega}_{\text{ad}}^\dagger \hat{P}_{QS} \hat{\Omega}_{\text{ad}}$ in the bare model as meaning the projector onto the spin- S subspace of the *enlarged* impurity Fock space $\tilde{\mathcal{H}}_{\text{bare}}^{\text{imp}}$ with spin $S = \frac{1}{2}$ on the physical impurity Fock space $\mathcal{H}_{\text{bare}}^{\text{imp}}$, where the local bath orbital $\hat{c}_{0\sigma}$ has a charge Q relative to half-filling. We assume a trace over different charge states and multiplets of the physical impurity space that are compatible with these conditions. Thus, we may

define

$$\begin{aligned} \langle \hat{P}_{\pm 1, 1/2} \rangle_{\text{bare}} &= \langle (\hat{n}_0 - 1)^2 \rangle_{\text{bare}}^{S=1/2}, \\ \langle \hat{P}_{0, 1} \rangle_{\text{bare}} &= \langle (\hat{n}_0 - 2\hat{n}_{0\uparrow}\hat{n}_{0\downarrow}) \left(\frac{5}{4} - \frac{1}{3}\hat{S}_{\text{imp}}^2 \right) \rangle_{\text{bare}}^{S=1}, \end{aligned} \quad (48)$$

where $\hat{S}_{\text{imp}} = \sum_i \hat{S}_i$ is an operator for the total spin of the physical impurity space and $\langle \cdots \rangle_{\text{bare}}^S$ denotes the thermal expectation value in the full bare model of an operator defined in the spin S block of the enlarged MC impurity Fock space.

We emphasize that any method that can compute such observables can be used for the UML optimization. This can be done at any temperature where bare and effective models should agree (that is, $T \ll E_{\text{ex}, 1} - E_{\text{GS}}$ and including $T = 0$ ground-state calculations). For example, QMC can be used to compute reference observables in the bare model at relatively high temperatures, and NRG can be used to iteratively refine the MC effective model via GD. With the optimized effective model at hand, one can then solve the effective model down to $T = 0$ cheaply with NRG, and calculate real-frequency dynamical correlation functions as might be needed for a DMFT calculation, electrical conductance via the Kubo formula, or the temperature dependence of thermodynamic quantities such as entropy or magnetic susceptibility.

Ultimately, we envision that UML will be integrated with *ab initio* methods to tackle problems that are currently out of reach for brute-force calculations. This is beyond the scope of the present work, but we provide a discussion of how few-body approximants to bare-model observables might play a role in UML in Appendix D.

5. Application: Triple quantum dot device

As a nontrivial demonstration of the UML methodology, we now consider its application to a bare model describing a triangular triple quantum dot (TQD) coupled to a metallic lead in a magnetic field—see Fig. 6(a). This system supports a rich range of physics [75–77]. The physical impurity Hilbert space has dimension $d_{\mathcal{H}} = 4^3 = 64$, and $\text{SU}(2)$ -spin symmetry as well as particle-hole symmetry is broken. We consider a regime of filling in which the TQD hosts a net spin- $\frac{1}{2}$ moment, delocalized across the three interacting quantum dot sites. The bare model Hamiltonian is given by

$$\begin{aligned} \hat{H}_{\text{bare}}^{\text{TQD}} &= \hat{H}_{\text{disc}}^{\text{bath}} + \frac{1}{2}U_d \sum_{i=1}^3 (\hat{n}_i^d - 1)^2 + \zeta \hat{N}^d + B \hat{S}_z^d \\ &+ \sum_{\sigma} [t \hat{d}_{1\sigma}^\dagger (\hat{d}_{2\sigma} + \hat{d}_{3\sigma}) + t' \hat{d}_{2\sigma}^\dagger \hat{d}_{3\sigma} + \text{H.c.}] \\ &+ V \sum_{\sigma} (\hat{d}_{1\sigma}^\dagger \hat{c}_{0\sigma} + \hat{c}_{0\sigma}^\dagger \hat{d}_{1\sigma}), \end{aligned} \quad (49)$$

where $\hat{n}_i^d = \sum_{\sigma} \hat{d}_{i\sigma}^\dagger \hat{d}_{i\sigma}$ is the number operator for each quantum dot, $\hat{N}^d = \sum_i \hat{n}_i^d$ is the total TQD number operator, and $\hat{S}_z^d = \frac{1}{2} \sum_i (\hat{d}_{i\uparrow}^\dagger \hat{d}_{i\uparrow} - \hat{d}_{i\downarrow}^\dagger \hat{d}_{i\downarrow})$ is the total TQD spin projection.

The UML analysis for the TQD proceeds similarly to that of the AIM. However, with the application of a magnetic field B , the spin multiplets previously labeled by the quantum number S are now no longer $(2S+1)$ -fold degenerate (although total S_z is still conserved). Due to the broken particle-hole symmetry, $Q = +1$ and -1

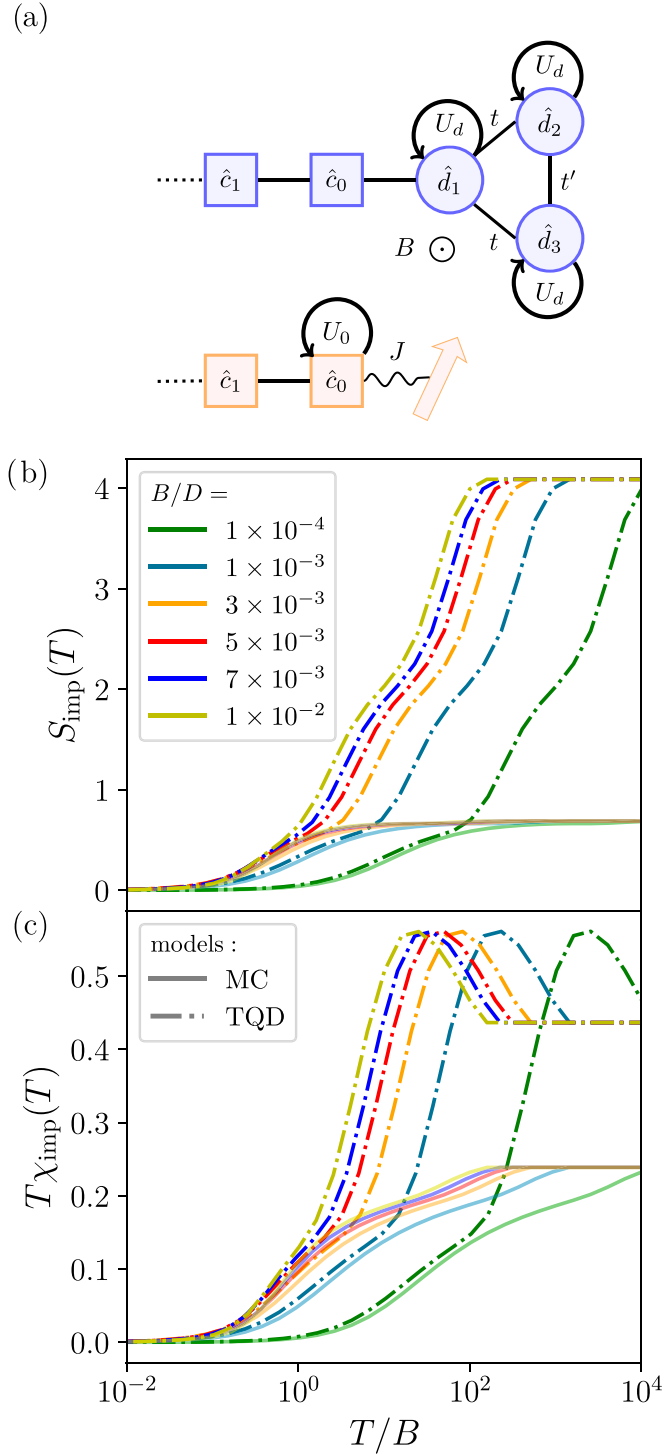


FIG. 6. (a) Schematic of the triple quantum dot bare model (blue) and the associated MC effective model (orange). (b) Entropy $S_{\text{imp}}(T)$ vs T/B for the TQD bare model (dot-dashed lines) and the UML-optimized effective model (solid lines) for different magnetic field strengths B . (c) Corresponding behavior of the magnetic susceptibility $T \chi_{\text{imp}}(T)$. Plotted for TQD model parameters $U_d = 1.1$, $V = 0.25$, $t = 0.10$, $t' = 0.15$, $\zeta = 0.05$, $D = 1$. Both bare and effective models solved using NRG. Quantitative agreement is found in all cases in the regime $T \ll U_d$ where the mapping is expected to hold.

sectors are no longer degenerate. The lower symmetry of the bare TQD model therefore implies that we also have spin and charge anisotropies in the MC effective model. As a consequence, UML for the TQD entails matching eight independent observables, rather than three for the symmetric AIM. Thermal expectation values of the symmetry projectors $\hat{P}_{Q,S_z}$ are the observables that we compare in bare and effective models, with $(Q, S_z) = (0, 0)$, $(0, 1)$, $(0, -1)$, $(1, 1/2)$, $(1, -1/2)$, $(-1, 1/2)$, $(-1, -1/2)$. The $(0,0)$ block is a diagonal 2×2 matrix. These observables can be defined and computed (here using NRG) in both bare and effective models.

The effective model is then optimized by minimizing the KLD loss function by gradient descent. Since the KLD is still convex and its gradient, Eq. (6), is expressed in terms of the computed observables $\langle \hat{P}_{Q,S_z} \rangle$, the optimization remains highly efficient, despite the increase from 3 to 8 in the number of parameters to be tuned. The gradient descent method is easily able to treat optimization problems of this complexity [11]. Indeed, as shown later, systems with a far larger number of parameters in the nonconvex case are also readily optimized.

Each step of the optimization process involves solving the MC effective model using the updated coupling constants, and computing the updated set of observables. The compute time simply grows linearly in the number of observables. These are compared with those of the bare model from a one-shot calculation. Even though the effective model has to be solved many times, these calculations are comparatively cheap, since by construction the Hilbert space dimension of the impurity in the effective model is less than that in the bare model. Although the effective model may have a larger number of coupling constants than the original bare model, it is defined in a reduced Hilbert space and therefore is of lower computational complexity to solve.

The iterative solution and refinement of the effective model with UML has essentially fixed computational cost, and is highly efficient. The potential bottleneck is in the one-shot solution of more complex bare models, as noted above. For the TQD, however, NRG can still be used to obtain exact benchmark results.

We demonstrate the viability of UML for the TQD, Eq. (49), in Fig. 6. Thermodynamic properties of the bare TQD (dot-dashed lines) and the optimized MC effective model (solid lines) are compared, for different temperatures T and field strengths B . Panel (b) shows the impurity contribution to the total entropy, $S_{\text{imp}}(T)$, and (c) shows the magnetic susceptibility $T \chi_{\text{imp}}(T)$, both versus T/B . The simplified effective model successfully captures the physics of the TQD for $T \ll U_d$. At higher temperatures, TQD degrees of freedom not included in the effective model contribute to thermodynamic averages.

As B is increased in the bare TQD model, we see a crossover from Kondo screening of the TQD local moment to polarization of the local moment by the field. Both the entropy and the magnetic susceptibility are fully quenched as $T \rightarrow 0$ for any B . However, the scale on which this happens depends on the underlying mechanism. For small fields $B \ll T_K$ (green line) screening occurs due to the Kondo effect, and the physics is insensitive to B . However, at larger B the Kondo effect, which depends on dynamical spin-flip scattering, is destroyed.

The TQD local moment is then screened on the scale of B . For $B \gg T_K$ we therefore see scaling collapse in terms of T/B , with polarization of the local moment setting in for $T/B \sim 1$ (see the yellow, blue, and red lines). The low-energy behavior of the optimized effective model captures this crossover and the onset of scaling versus T/B accurately.

We conclude that more complex models, including those with more interacting degrees of freedom and lower symmetries, can be treated efficiently and accurately with UML. Physical properties of the resulting optimized effective models agree quantitatively at low temperatures with bare model benchmark reference results.

IV. SYSTEMATIC IMPROVEMENT OF EXTENDED MODELS

We now put the above discussions on a more general footing, and consider systematic improvements to the MC effective models explored so far.

The expressibility of our effective model to capture different aspects of the physics of the bare model depends on the types of operators that can be included. Going beyond the minimal Kondo model, which includes only the most RG-relevant operator, our MC effective model lives in an extended impurity Fock space Eq. (36), which allows for inclusion of more general coupling terms, Eq. (41). Our starting point in this section is to note that the quality of the effective model can be systematically improved by further enlarging the effective impurity Fock space:

$$\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}} = \mathcal{H}_{\text{eff}}^{\text{imp}} \otimes \mathcal{H}_0^{\text{bath}} \otimes \mathcal{H}_1^{\text{bath}}, \quad (50)$$

which now includes both $\hat{c}_{0\sigma}$ and $\hat{c}_{1\sigma}$ bath orbitals. The extended effective model then takes the form

$$\hat{H}_{\text{eff}}^{\text{ext}}(\theta) = \tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}(\theta) + \hat{H}_2^{\text{bath}} + \hat{H}_2^{\text{hyb}}, \quad (51)$$

where the extended effective impurity Hamiltonian $\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}(\theta)$ lives in the extended impurity Fock space $\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}$ and correspondingly the bath and hybridization terms have been redefined according to Eqs. (33) and (34). The bath chain now starts on site index $N = 2$. The impurity part can of course be further extended if required. The MC effective model discussed in Sec. III can be viewed as the first member of a family of such extended models.

When the original bath Eq. (10) takes the form of a Wilson chain [49], we note that the $\hat{c}_{n\sigma}$ operators become progressively less RG relevant as the chain index n increases [specifically, the most relevant operator of the linearized RG transformation Eq. (19) involved in the decomposition of $\hat{c}_{n\sigma}$ near the low-energy fixed point has an eigenvalue $|\lambda^*|$ that decreases with n]. Therefore, expanding our effective impurity Hamiltonian to include operators from Wilson orbitals with larger n results in additional terms that are progressively more RG irrelevant. On the one hand, strongly RG irrelevant corrections to our effective model become less important for determining the low-energy physics, but on the other hand, expectation values of these operators flow very weakly under RG and therefore can be accurately matched even at relatively

high energy scales with corresponding observables in the bare model during UML optimization.

To proceed with the construction and UML optimization of our generalized effective models, we first note that the symmetry representations of U(1) charge Q and SU(2) spin S on $\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}$ still commute, and so we can construct a basis on the effective Fock space labeled by quantum numbers S for the non-Abelian spin, S_z for the Abelian magnetic quantum number of the Cartan subalgebra of S , and Q for the Abelian charge,

$$\langle Q, S, S_z; m | Q', S', S'_z; m' \rangle = \delta_{QQ'} \delta_{SS'} \delta_{S_z S'_z} \delta_{mm'}, \quad (52)$$

where m is the *multiplet* index that distinguishes different basis states with the same set of quantum numbers Q, S, S_z . Unlike in the basic MC example, in general the symmetry blocks of the extended impurity Hamiltonian are not simply 1×1 scalars. Furthermore, such multiplets are not unique since linear combinations of different multiplets with the same set of quantum numbers are also valid multiplets. This introduces a gauge degree of freedom in the choice of the basis that means we cannot unambiguously compare degenerate multiplets between bare and effective models based on symmetry alone. We show below that fixing the gauge by lifting the multiplet degeneracy allows us to reestablish a one-to-one correspondence between bare and effective models.

Accounting for the multiplet structure of the symmetry spaces, we construct projectors by generalizing Eq. (37),

$$\hat{P}_{QS} = \sum_{S_z, m} |Q, S, S_z; m\rangle \langle Q, S, S_z; m|. \quad (53)$$

One can use these projectors as a mutually commuting operator basis to construct the effective impurity Hamiltonian,

$$\tilde{\mathcal{H}}_{\text{eff}}^{\text{imp}}(\theta) = \sum_{Q, S} \theta_{QS} \hat{P}_{QS}. \quad (54)$$

The allowed quantum number configurations compatible with Eq. (50) are

Q	S	S_z	M_{eff}
0	$\frac{1}{2}$	$\pm \frac{1}{2}$	4
0	$\frac{3}{2}$	$\pm \frac{1}{2}, \pm \frac{3}{2}$	1
± 1	0	0	2
± 1	1	0, ± 1	2
± 2	$\frac{1}{2}$	$\pm \frac{1}{2}$	1

where the charge Q is again specified relative to half-filling, M_{eff} is the number of multiplets with a given Q and S , and we take the impurity to be a spin- $\frac{1}{2}$ object. We therefore have 15 QS multiplets that span the extended impurity space.

The advantage of formulating the effective model in terms of such symmetry projectors is that equivalent projectors can be identified in the bare model, since the two models share the same symmetries. The projectors so defined also mutually commute, which means the UML optimization problem is convex (Appendix A).

A disadvantage is that several QS blocks contain more than one multiplet, which means that the projectors cannot distinguish between multiplets with the same QS . The effective model Eq. (54) is therefore *not* the most general extended impurity model one can write down. Interactions in the bare model will in general lift the energy degeneracy of these multiplets, but this effect cannot be captured by a model of the form of Eq. (54).

A. Fixed basis

A partial solution is to fix the basis according to physical intuition of the problem, and add terms to Eq. (54) that lift the energy degeneracy of multiplets with the same Q and S quantum numbers. This should be done in such a way that (i) the new set of operators still all mutually commute, so as to guarantee convexity of the UML optimization problem; and (ii) the added terms have a physical correspondence in the bare model. This approach is still not the most general form of the extended MC effective model, because we make a basis choice. We return to the fully general case later.

Since the $\hat{c}_{0\sigma}$ and $\hat{c}_{1\sigma}$ bath orbitals are common to both bare and effective models, and there is a physical tunneling matrix element between these orbitals in both cases, we introduce the operator

$$\hat{T} = \sum_{\sigma} (\hat{c}_{0\sigma}^{\dagger} \hat{c}_{1\sigma} + \hat{c}_{1\sigma}^{\dagger} \hat{c}_{0\sigma}). \quad (55)$$

The tunneling operator $\hat{T}$ has support only on the physical bath Fock space and therefore does not depend on the structure of the impurity in the bare model. However, one might anticipate that $\langle \hat{T} \rangle_{\text{bare}}$ will depend on details of the bare impurity model. Action of the tunneling operator conserves charge and spin and so $\hat{T}$ commutes with $\hat{Q}$ and $\hat{S}^2$ defined on the extended impurity space. We can use the eigenvalues T of the tunneling operator to label our basis states, as shown explicitly in Appendix B. With basis states $|Q, S, T, S_z; m\rangle$ we almost fully remove multiplet degeneracies,

Q	S	T	S_z	M_{eff}
0	$\frac{1}{2}$	0	$\pm \frac{1}{2}$	2
0	$\frac{1}{2}$	± 2	$\pm \frac{1}{2}$	1
0	$\frac{3}{2}$	0	$\pm \frac{1}{2}, \pm \frac{3}{2}$	1
± 1	0	± 1	0	1
± 1	1	± 1	0, ± 1	1
± 2	$\frac{1}{2}$	0	$\pm \frac{1}{2}$	1

The $(Q, S, T) = (0, \frac{1}{2}, 0)$ subspace still has $M_{\text{eff}} = 2$ multiplets, but this degeneracy can be lifted by introducing an additional operator (which we take to act only in this subspace), $\hat{W} = \hat{n}_{0\uparrow}\hat{n}_{0\downarrow} + \hat{n}_{1\uparrow}\hat{n}_{1\downarrow}$, which measures double-occupancy on the bath sites included in the extended impurity space. $\hat{W}$ has eigenvalues 0 or 1 only in the subspace where the operator acts, so we employ a concise notation in which the (T, W) eigenvalues are augmented into one label $(T, W) \mapsto T$.

As such, all multiplet degeneracies are lifted when we express

$$\hat{H}_{\text{eff}}^{\text{imp}}(\theta) = \sum_{Q,S,T} \theta_{QST} \hat{P}_{QST}, \quad (56)$$

where the spin multiplets $|Q, S, T, S_z\rangle$ are now unique (we drop the multiplet index m) and $\hat{P}_{QST} = \sum_{S_z} |Q, S, T, S_z\rangle \langle Q, S, T, S_z|$.

Fixing the multiplet basis in such a way allows us to address each multiplet individually and make a comparison of their projector expectation values in bare and effective models. This is the best one can do while retaining a fully commuting set of extended impurity Hamiltonian operators, as required for convexity. However, the most general extended effective model would not constrain the basis in this way.

If desired, the projectors $\hat{P}_{QST}$ can be formulated in second-quantized form (see Appendix D). New types of coupling term arise in the extended MC effective model, relative to the basic MC model. In particular, we see *assisted hopping* terms in which an impurity spin-flip is accompanied by electronic tunneling between bath sites. This might be expected from the physical interpretation of the Kondo effect in terms of spin-flip scattering.

B. Generalized basis

The most general extended effective impurity model must include off-diagonal terms coupling different multiplets. This can be captured by operators of the type

$$\hat{X}_{QS}^{T,T'} = \sum_{S_z} |Q, S, T, S_z\rangle \langle Q, S, T', S_z| + \text{H.c.} \quad (57)$$

such that our effective model reads

$$\hat{H}_{\text{eff}}^{\text{imp}}(\theta) = \sum_{Q,S} \sum_{T,T'} \theta_{QS}^{T,T'} \hat{X}_{QS}^{T,T'}. \quad (58)$$

The model thus features 25 distinct interaction terms. Due to the cross terms $T \neq T'$, the operators $\hat{X}$ do not in general commute, and so the UML optimization problem is no longer convex. This is unavoidable for more complex effective models, but the appearance of local minima in the search landscape can still be dealt with in practice using more sophisticated GD-based routines.

The impact of including the off-diagonal terms $T \neq T'$ can be quantified using the FSP. An optimized representation of any operator $\hat{Y}$ using the set of operators contained in the model Eq. (58) can be found by minimizing the normalized FSP [78],

$$\mathcal{L}(\theta, \hat{Y}) = 1 - \frac{|\langle \hat{H}_{\text{eff}}^{\text{imp}}(\theta), \hat{Y} \rangle|}{\|\hat{H}_{\text{eff}}^{\text{imp}}(\theta)\| \cdot \|\hat{Y}\|}. \quad (59)$$

As a demonstration of the expressibility of the model Eq. (58), we consider its ability to represent the Kondo Hamiltonian operator $\hat{Y} = \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0$ from Eq. (21). In the limit $U_d \gg 1$ where the perturbative SWT holds, the effective model should reduce to Eq. (21), so this is a nontrivial and stringent check that a good effective model should satisfy. By minimizing $\mathcal{L}(\theta, \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0)$ with respect to θ using standard gradient descent, we find $\mathcal{L}(\theta_{\text{opt}}, \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0) = 0$, meaning that Eq. (58)

can perfectly represent this operator. However, if we use a fixed basis (no off-diagonal terms) as with Eq. (56), then $\mathcal{L}(\theta_{\text{opt}}, \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0) \approx 0.3$. The fixed basis, therefore, substantially reduces the expressibility of the extended model. Therefore, even though the fully general extended model lacks convexity, we argue that this feature should be sacrificed and that Eq. (58) should be used.

C. UML optimization of extended models

Extended MC effective models with a general multiplet basis such as Eq. (58) can represent any Hamiltonian term in the extended Fock space of Eq. (50), consistent with SU(2) spin and U(1) charge symmetries.

Within UML, the GD optimization of the effective model is achieved by matching observables in the bare and effective models, Eq. (6). As noted in Sec. III, however, one cannot hope to match the *full* set of symmetry projectors because the optimization problem is overdetermined. The solution discussed in Sec. III B 2 is to eliminate the observable that flows strongest under RG, as quantified by the FSP Eq. (59).

However, even by omitting the observable with the largest $\tilde{\alpha}$, we are still left with 24 observables to compute in the effective model Eq. (58). It is not obvious that for generalized effective models one can always simultaneously match all such observables in bare and effective models. Indeed, we find that even in the optimized effective model, a small number of observables are not matched, and the KLD loss function can therefore only be approximately minimized. To quantify this, we introduce a measure of the relative error in a given observable,

$$\Delta(\langle \hat{h}_i \rangle_{\text{eff}}, \langle \hat{h}_i \rangle_{\text{bare}}) = \left| 1 - \frac{\max(\langle \hat{h}_i \rangle_{\text{eff}}, \langle \hat{h}_i \rangle_{\text{bare}})}{\min(\langle \hat{h}_i \rangle_{\text{eff}}, \langle \hat{h}_i \rangle_{\text{bare}})} \right|. \quad (60)$$

When learning a representation for complex bare models (for example *ab initio* models of a realistic molecular junction), the Kondo temperature T_K will not typically be a good figure of merit to assess the quality of the effective model. This is because, for any practical application of UML, the bare model will not itself be solvable down to very low but finite temperatures—hence the need in the first place for a simplified effective model that can be fully solved. Therefore, one will not in general have access to the reference T_K for the bare model, from which to compute the relative error Eq. (47). The error in computable observables is thus more practical; the KLD loss function gradient depends on these observables, which are matched during the optimization process. For simple bare model test cases such as the AIM, we can, however, still compare with T_K and other quantities.

Taking the AIM again as our bare model for demonstration, we perform UML optimization on the extended MC effective model Eq. (58). For representative AIM parameters, in Fig. 7(a) we plot the relative error of selected (diagonal) observables labeled by the projector quantum numbers (Q, S, T) as a function of the optimization step (epoch). We find that most effective model observables converge exponentially quickly towards their reference bare model counterparts, whereas a few are essentially stationary and cannot be matched. Indeed, it is in general not possible to

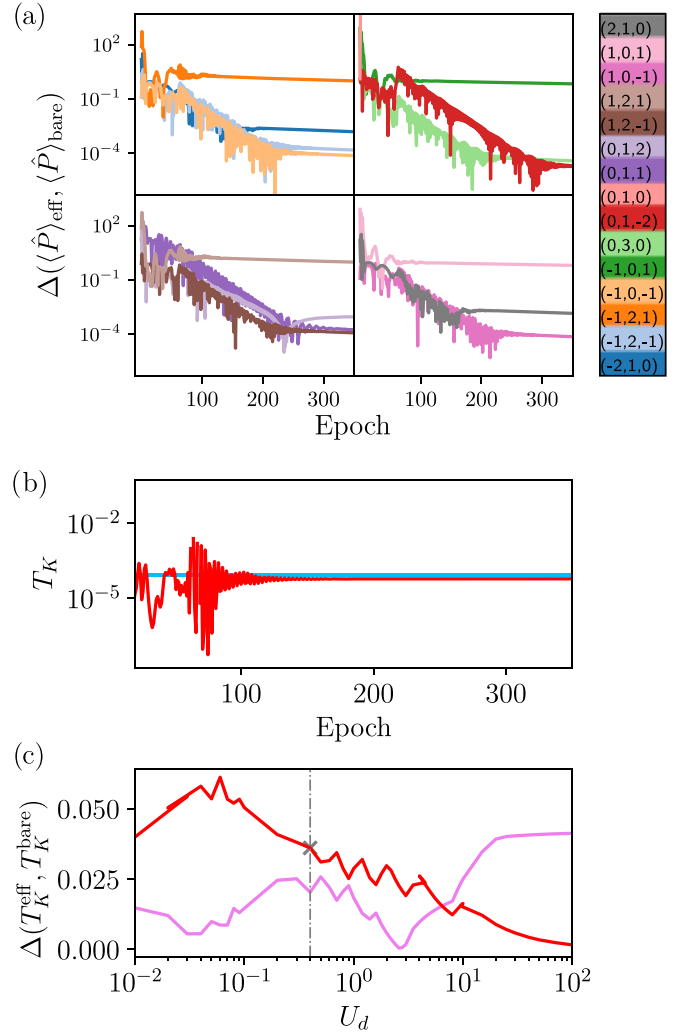


FIG. 7. Analysis of UML optimization of the extended MC effective models [Eqs. (56) and (58)], taking the AIM as our bare model. (a) Comparison of relative errors in computed observables [Eq. (60)] as a function of the GD optimization parameter update step (epoch). Lines are color-coded according to the key, labeled by the (diagonal) projector quantum numbers (Q, S, T). (b) Convergence of the Kondo temperature T_K with epoch (red line) compared with the reference result in the bare model (blue line). (c) Comparison of relative (logarithmic) error in the Kondo temperature for the fixed-basis effective model Eq. (56) (pink line) and generalized-basis model Eq. (58) (red line). Plotted for $8V^2/U_d = 0.3$ and $U_d = 0.4$, with the optimization performed at $T = 0$ using NRG as the impurity solver for both bare and effective models.

perfectly match multiple observables simultaneously, and the minimum relative error is small but remains finite after convergence. This is because the effective model has good but not unlimited expressibility: minimization of the KLD loss function typically involves a compromise in the optimized solution. However, as Fig. 7(b) shows, this does not make much difference to the quality of the optimized effective model, as quantified by the Kondo temperature T_K . After ~ 100 or so training steps, T_K for the effective model (red line) converges quite accurately to that of the bare model (blue line).

In Fig. 7(c) we show the relative (logarithmic) error in T_K in the fully effective model, as a function of AIM interaction strength U_d , with the red line for the fully general extended MC effective model Eq. (58), compared with the fixed-basis model Eq. (56) as the purple line. We see that T_K is reproduced rather well by the effective model in both cases for all values of U_d . As expected from the above arguments, the generalized model does better at large U_d . However, somewhat unexpectedly, the fixed-basis model has a smaller error at small U_d . We attribute this to the fact that optimization of the fixed-basis model requires us to match fewer observables, which is easier to achieve in practice. Furthermore, the additional (off-diagonal) observables in Eq. (58) appear to flow more strongly under RG at small U_d than the projectors in Eq. (56), meaning that matching them may ironically *reduce* the quality of the optimized effective model. Despite these subtleties of the training process, the extended MC effective models do represent a systematic improvement over simpler effective models.

We compare the different approaches in the following.

V. COMPARISON OF MODELS

In this section, we consider the relative performance of the various effective models introduced in this paper. Each is optimized by UML to capture the physics of the AIM, which serves as our benchmark bare model. As we have shown, it is in general not possible to simultaneously capture within a simplified effective model the universal low-temperature thermodynamics (as characterized by the Kondo temperature T_K) as well as the behavior of static thermal expectation values of local operators. Any effective model must therefore strike a compromise in reproducing these features. Different methods may target different aspects of the physics and be better suited to different applications.

The $\mathcal{F}$ -learning approach [46] requires calculation of the full free energy of bare and effective models within the optimization process. It guarantees to reproduce the low-energy thermodynamics and T_K within a minimal Kondo model, but fails to capture local observables except in the perturbative limit $U_d \gg D$ where the SWT also holds as a good description. However, it might not be possible to accurately calculate the free energy of a complex bare model in practice, so other methods might be preferred.

UML, by contrast, is based on matching certain local observables, and so by construction these observables agree in bare and optimized effective models. However, applied to crude effective models, such as minimal Kondo-type models that only feature the most RG-relevant terms, UML yields solutions that do not capture well the low-energy thermodynamics and Kondo scale T_K . By including minimally constrained models in an expanded effective impurity Fock space, we systematically improve upon results obtained by UML. Such models include RG-marginal and RG-irrelevant corrections, and UML optimization is done using the set of observables that do not flow strongly. Depending on the level of theory employed, one can optimize with UML generalized effective models that do capture local observables while also approximating T_K with good accuracy.

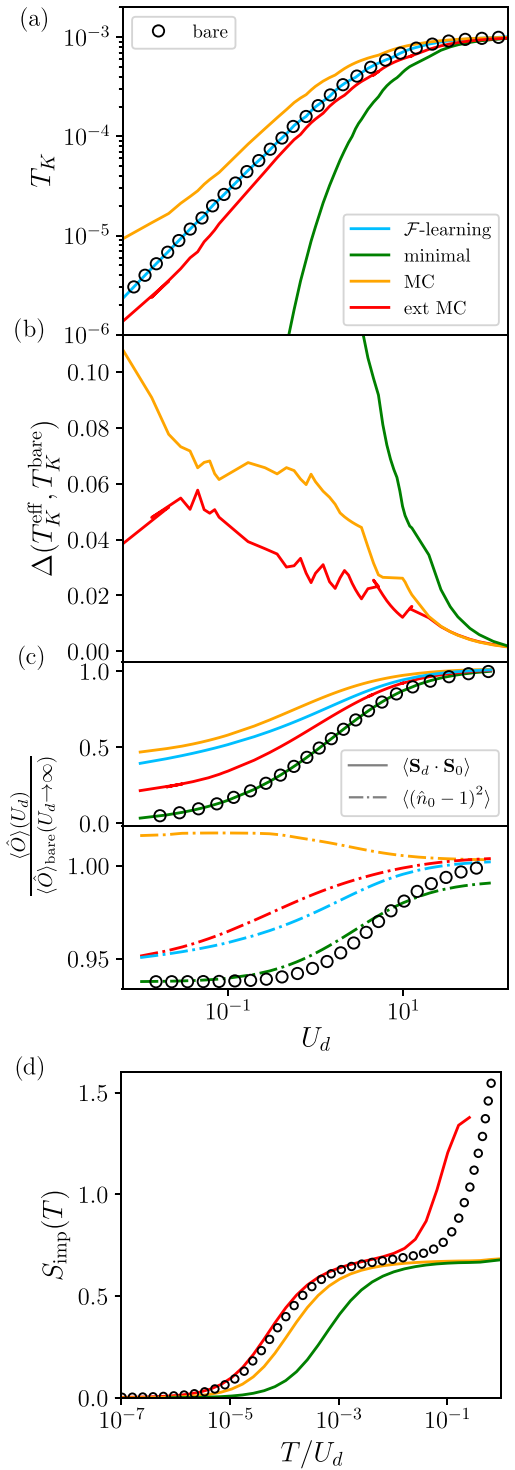


FIG. 8. Comparison of performance metrics for the optimized effective models considered in this work, taking the AIM as the bare model. (a) Kondo temperature T_K vs U_d , comparing the AIM (points) and $\mathcal{F}$ -learned Kondo result (blue line) with the UML-optimized effective models: minimal (green), MC (orange), and extended MC (red). Corresponding T_K relative error shown in (b). Observables $\langle \hat{O} \rangle$ normalized by their $U_d \rightarrow \infty$ values shown in panel (c), with $\hat{O} = \mathbf{S}_d \cdot \mathbf{S}_0$ (solid lines, top) and $(\hat{n}_0 - 1)^2$ (dashed lines, bottom). (d) Entropy $S_{\text{imp}}(T)$ vs T/U_d for the bare AIM ($U_d = 0.4$, points) and the optimized effective models (lines). We set $8V^2/U_d = 0.3$ throughout.

These features are illustrated in Fig. 8, where we compare the performance of the different effective models.

The UML method is based on minimizing the KLD loss function Eq. (5) that quantifies the distinguishability between bare and effective model. In general ML applications, the value of the loss function is a useful metric to judge the training progress and ultimately the performance of the model. In the case of Eq. (5), however, the KLD is essentially impossible to compute in practice. The training progress cannot therefore be established in the standard way. On the other hand, we have shown that to minimize the KLD via GD we only need to know its gradient, Eq. (6). This can be done without actually ever evaluating the KLD itself. For a commuting set of local impurity operators (for example, the symmetry projectors in MC models), the KLD is convex and so the gradient vanishes only at saddle points or global minima of the KLD. Thus, the gradient is itself a good replacement for the loss function to determine whether the training of a model has converged. For the cases considered in Fig. 8, the KLD gradient can always be made to vanish.

We therefore turn to our other metrics to quantify the relative performance of the models. With our NRG solution of the bare and optimized effective models, we can compute the Kondo temperature directly. This is shown in Fig. 8(a) as a function of U_d . The exact results for the bare AIM (points) are compared with the effective models (lines): the $\mathcal{F}$ -learning and UML result for the minimal Kondo model shown are as blue and green lines; the MC and extended MC models are shown as orange and red lines. Panel (b) shows the corresponding relative error $\Delta(T_K^{\text{eff}}, T_K^{\text{bare}})$ evaluated using Eq. (47). In panel (c) we plot the $T = 0$ value of static observables $\langle \hat{O} \rangle = \langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ and $\langle (\hat{n}_0 - 1)^2 \rangle$, comparing results in bare AIM (points) with those of the optimized effective models (lines). We have normalized the results by their value in the $U_d/D \rightarrow \infty$ limit.

First, we note that by construction, $\mathcal{F}$ -learning perfectly reproduces the T_K value for all U_d ; see the blue line in Fig. 8(a). However, the behavior of $\langle \hat{O} \rangle = \langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$, which is not directly optimized by $\mathcal{F}$ -learning, is poorly captured; see panel (c). On the other hand, UML for the minimal Kondo model (green line) directly matches the observable $\langle \hat{O} \rangle = \langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ in panel (c) and therefore by construction agrees perfectly with the bare AIM results. The T_K in this case is not directly matched, and panel (a) shows that UML for the minimal model performs very poorly using that metric, except in the SWT (perturbative) limit $U_d/D \rightarrow \infty$.

The behavior of the Kondo temperature and the observables is a stringent nontrivial performance check for the MC and extended MC models, since neither T_K nor $\langle \hat{\mathbf{S}}_d \cdot \hat{\mathbf{S}}_0 \rangle$ is directly matched within UML. We find that UML represents a good compromise on the performance of these models for the two metrics. Good performance is seen for each over all U_d , with a systematic improvement for extended MC over plain MC.

With the optimized effective models at hand, we compute using NRG the full temperature-dependence of the thermodynamic entropy for a representative example of the bare AIM with $U_d = 0.4D$, comparing the results in Fig. 8(d). Within the regime of validity of the effective models ($T/U_d \ll 1$), we see

that the MC and extended MC models do accurately capture the physics of interest for the AIM.

VI. CONCLUSION

In this paper, we focused on developing new ML-inspired methodologies for the automatic derivation of expressive and accurate effective models for quantum impurity systems. To establish the feasibility and performance of our UML method, we studied in detail the AIM as the bare model. This is because the AIM is a very well-known paradigm [33], which we can solve exactly down to $T = 0$ using NRG [50] for benchmarking purposes. The same approach can in principle be used for more complicated bare model systems as described in Sec. III B 5. However, we recognize that useful applications of UML may require integration with *ab initio* methods to treat realistic systems [4,70–74]. This is beyond the scope of the current work, but we discuss possible ways to do this as an outlook in Appendix D.

The UML framework that we have introduced is based on unsupervised ML techniques, with the goal to optimize parameters of a variational effective Hamiltonian, to best describe the low-temperature physics of a more complex bare model quantum impurity system. In a well-defined sense, the optimized effective Hamiltonian is the “best fit” description at a given complexity level. We propose a loss function to quantify the distinguishability between bare and effective models, based on a classical distribution of quantum Feynman diagrams within a hybridization expansion of the partition function. The loss function can be minimized using GD by computing certain physical observables. In the standard case we discuss, the resulting optimization problem is convex, resulting in a highly efficient learning process. The optimized effective model is the one in which local static observables match in bare and effective models.

Care must be taken, however, since thermodynamic observables are not invariant under RG and so bare and effective models with different RG flows may have different static expectation values. Thus, the naive matching of observables does not imply that low-energy scales, such as the Kondo temperature, are correctly recovered. We confirm this directly, already at the simplest possible example of the Anderson-Kondo model mapping. To mitigate this, we identify and construct the best observables to match, based on an analysis of their RG relevance. We show that this can be done while maintaining convexity of the optimization, and that low-energy scales can be recovered. The accuracy of our effective models can be systematically improved by increasing the complexity of the effective impurity. Although this clearly offers no advantage at the level of the AIM, UML produces dramatically simplified effective models when applied to more complex systems, consisting of three or more interacting active orbitals. The accuracy and efficiency of UML may be further improved by formulating bare and effective models in the “natural orbital” (NO) basis [79]. Since entanglement between NOs and the rest of the fermionic bath is small by construction, local observables to be matched in UML should flow weakly, which we have showed is a desirable property when formulating effective models.

An important aspect of the UML framework is that it can be performed at relatively high temperatures without compromising precision. This opens the door to applying the algorithm to vastly more complex systems than could be treated by brute-force NRG calculations alone. For example, a one-shot CT-QMC calculation of observables for the bare model could be used as a reference to optimize a simpler effective model treated cheaply within NRG, which can then be solved down to $T = 0$. Integration with *ab initio* methods may permit the realistic study of complex strongly correlated systems that are currently out of reach.

ACKNOWLEDGMENTS

The authors would like to thank Sudeshna Sen, Alexander Munoz, Lucas K. Wagner, and Denjoe O'Connor for the insightful discussion. We acknowledge funding from the Irish Research Council Laureate Awards 2017/2018 through Grant No. IRCLA/2017/169 (J.B.R. and A.K.M.), from Science Foundation Ireland through Grant No. 21/RP-2TF/10019 (A.K.M.) and from the Helmholtz Initiative and Networking Fund, Grant No. VH-NG-1711 (J.B.R.).

APPENDIX A: HYBRIDIZATION EXPANSION AND CONVEXITY OF THE LOSS FUNCTION

For a general Hamiltonian, which can be bipartitioned $\hat{H} = \hat{H}_0 + \hat{H}_1$, one can compute the partition function as an expansion in powers of $\hat{H}_1$ by [65]

$$\mathcal{Z} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} T_{\tau} \int_0^{\beta} d\tau_1 \cdots \int_0^{\beta} d\tau_n \times \text{tr}[e^{-\beta \hat{H}_0} \hat{H}_1(\tau_n) \cdots \hat{H}_1(\tau_1)], \quad (\text{A1})$$

where τ is the imaginary time. This expansion is discussed extensively in the context of continuous-time quantum Monte Carlo (CT-QMC) [80–82]. In the case of hybridization expansion CT-QMC [82], $\hat{H}_1 = \hat{H}^{\text{hyb}}$ describes the hybridization between a noninteracting bath $\hat{H}^{\text{bath}}$ and an interacting quantum impurity $\hat{H}^{\text{imp}}$. For the following discussion, we therefore consider Hamiltonians of the form

$$\hat{H}_0 = \hat{H}^{\text{bath}} + \hat{H}^{\text{imp}}, \quad (\text{A2a})$$

$$\hat{H}_1 = \hat{H}^{\text{hyb}} = \sum_k \sum_{\sigma} V_k^{\sigma} \hat{d}_{\sigma}^{\dagger} \hat{c}_{k\sigma} + \text{H.c.}, \quad (\text{A2b})$$

where we assume that the hybridization tensor is diagonal in the spin quantum number σ [82]. For this type of Hamiltonian, Eq. (A1) becomes

$$\begin{aligned} \mathcal{Z} = & \sum_{n=0}^{\infty} \int_0^{\beta} d\tau_1 \cdots \int_{\tau_{n-1}}^{\beta} d\tau_n \int_0^{\beta} d\tau'_1 \cdots \int_{\tau'_{n-1}}^{\beta} d\tau'_n \\ & \times \sum_{\substack{a_1, \dots, a_n, k_1, \dots, k_n \\ a'_1, \dots, a'_n, k'_1, \dots, k'_n}} V_{k_1}^{a_1} V_{k'_1}^{a'_1*} \cdots V_{k_n}^{a_n} V_{k'_n}^{a'_n*} \\ & \times \text{tr}[T_{\tau} e^{-\beta \hat{H}^{\text{bath}}} c_{k'_n a'_n}^{\dagger}(\tau'_n) \hat{c}_{k_n a_n}(\tau_n) \cdots c_{k'_1 a'_1}^{\dagger}(\tau'_1) \hat{c}_{k_1 a_1}(\tau_1)] \\ & \times \text{tr}[T_{\tau} e^{-\beta \hat{H}^{\text{imp}}} \hat{d}_{a'_n}^{\dagger}(\tau'_n) \hat{d}_{a_n}^{\dagger}(\tau_n) \cdots \hat{d}_{a'_1}^{\dagger}(\tau'_1) \hat{d}_{a_1}^{\dagger}(\tau_1)]. \quad (\text{A3}) \end{aligned}$$

Equation (A3) can be interpreted as a sum over all possible diagrams obtained by allowing electrons to hop between the impurity and the bath. Since the bath is noninteracting it can be integrated out, and using Wick's theorem the *antiperiodic hybridization function* can be obtained,

$$\begin{aligned} & \det_{ij} [V_{k_i}^{\sigma_i} V_{k'_j}^{\sigma'_j*} \text{tr}(T_{\tau} e^{-\beta \hat{H}^{\text{bath}}} \hat{c}_{k_i \sigma_i}^{\dagger}(\tau_i) \hat{c}_{k'_j \sigma'_j}(\tau'_j))] \\ & = \mathcal{Z}^{\text{bath}} \det_{ij} [V_{k_i}^{\sigma_i} V_{k'_j}^{\sigma'_j*} \langle T_{\tau} \hat{c}_{k_i \sigma_i}^{\dagger}(\tau_i) \hat{c}_{k'_j \sigma'_j}(\tau'_j) \rangle^{\text{bath}}] \\ & = \mathcal{Z}^{\text{bath}} \det(\Delta^{(x)}), \quad (\text{A4}) \end{aligned}$$

where $x = (n, \{k_i, k'_i, a_i, a'_i, \sigma_i, \sigma'_i, \tau_i, \tau'_i\}_{i=1}^n)$ denotes an impurity diagram in terms of the sequence of impurity operators [65] such that

$$\int_0^{\beta} d\tau_1 \cdots \int_{\tau_{n-1}}^{\beta} d\tau_n \sum_{a_1, \dots, a_n} \sum_{k_1, \dots, k_n} \mapsto \sum_x. \quad (\text{A5})$$

Following the approach of Ref. [55], we bring the impurity operators into the eigenbasis $\{(|E_{\alpha}, |\alpha\rangle)\}$ of $\hat{H}^{\text{imp}}$ and thus $e^{-\tau \hat{H}^{\text{imp}}}$ can be trivially evaluated, with

$$d_a^{\dagger}(\tau) = e^{-\tau \hat{H}^{\text{imp}}} d_a^{\dagger} e^{\tau \hat{H}^{\text{imp}}} = \sum_{\alpha, \alpha'} e^{\tau(E_{\alpha'} - E_{\alpha})} |\alpha\rangle \langle \alpha'| |\alpha| d_a^{\dagger}(\tau) |\alpha'\rangle \quad (\text{A6})$$

and instead of occupation diagrams we obtain diagrams involving impurity eigenstates $|\alpha\rangle_x \equiv |\alpha_1 \cdots \alpha_k, \alpha'_1 \cdots \alpha'_k\rangle_x$. Thus, Eq. (A1) can be rewritten in terms of the weights of the sum of the eigenstate diagrams,

$$\mathcal{Z} / \mathcal{Z}^{\text{bath}} = \sum_x \sum_{\{\alpha\}} \Lambda_{\{\alpha\}_x} \det(\Delta_x) e^{-(\hat{H}^{\text{imp}})_{\{\alpha\}_x}}, \quad (\text{A7})$$

with $\Lambda_{\{\alpha\}_x}$ denoting the contribution from impurity operators in the impurity eigenbasis

$$\Lambda_{\{\alpha\}_x} = \prod_{i=1}^n \langle \alpha_i | d_{a'_i}^{\dagger} | \alpha'_i \rangle \langle \alpha'_i | d_{a_i}^{\dagger} | \alpha_{i+1} \rangle \quad (\text{A8})$$

and $\langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x}$ is the average impurity energy over a diagram

$$\begin{aligned} \langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x} = & \sum_{i=1}^n [E_{\alpha_i}(\tau'_i - \tau_{i-1}) + E_{\alpha'_i}(\tau_i - \tau'_i)] \\ & + E_{\alpha_n}(\beta - \tau_n). \quad (\text{A9}) \end{aligned}$$

Having reformulated the expansion of the impurity partition Eq. (A3) as Eq. (A7), we now look to prove the convexity of KLD loss function Eq. (5).

To show the convexity of Eq. (5), it suffices to show that the free energy $\log(\mathcal{Z})$ is convex in the variational parameters θ . The core assumption of our calculation is that the effective impurity Hamiltonian is constructed as follows:

$$H^{\text{imp}} = \sum_i \theta_i \hat{h}_i, \quad [\hat{h}_i, \hat{h}_j] = 0, \quad (\text{A10})$$

which is the case for the effective models in the main text. For the following proof, it is convenient to define a *weight* w for

eigenstate diagrams to reformulate Eq. (A7),

$$\mathcal{Z} = \mathcal{Z}^{\text{bath}} \sum_x \sum_{\{\alpha\}_x} w(\{\alpha\}_x), \quad (\text{A11})$$

$$w(\{\alpha\}_x) = e^{-\langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x}} \Lambda_{\{\alpha\}_x} \det(\Delta_x), \quad (\text{A12})$$

where we also assume that $w(\{\alpha\}_x) > 0$, such that w acts as a probability distribution upon normalization. To show that $\log(\mathcal{Z})$ is convex, we need to show its Hessian is positive-semidefinite,

$$\partial_{\theta_i} \partial_{\theta_j} \log(\mathcal{Z}) \succcurlyeq 0. \quad (\text{A13})$$

In the following, we use the shorthand notation ∂_i for ∂_{θ_i} for concision. It is straightforward to compute the first-order derivative using the fact that the trace is invariant under circular shifts,

$$\begin{aligned} \partial_i \ln(\mathcal{Z}) &= -\frac{1}{\mathcal{Z}} \int_0^\beta d\tau \text{tr} [T_\tau e^{-\int_0^\tau d\tau' \hat{H}} \hat{h}_i(\tau) e^{-\int_\tau^\beta d\tau' \hat{H}}] \\ &= -\frac{1}{\mathcal{Z}} \int_0^\beta d\tau \text{tr} [e^{-\int_0^\tau d\tau' \hat{H}} - \int_\tau^\beta d\tau' \hat{H} \hat{h}_i(\tau)] \\ &= -\frac{1}{\mathcal{Z}} \int_0^\beta d\tau \sum_\alpha e^{-\beta E_\alpha} \langle \alpha | e^{-\tau E_\alpha} \hat{h}_i e^{\tau E_\alpha} | \alpha \rangle \\ &= -\beta \text{tr}[\hat{\rho} \hat{h}_i] = -\beta \langle \hat{h}_i \rangle. \end{aligned}$$

However, we can also carry out the derivative directly on the diagrammatic expansion of the partition function,

$$\begin{aligned} \partial_i \log(\mathcal{Z}) &= \frac{1}{\mathcal{Z}} \partial_i \left[\mathcal{Z}^{\text{bath}} \sum_x \sum_{\{\alpha\}_x} e^{-\langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x}} \Lambda_{\{\alpha\}_x} \det(\Delta_x) \right] \\ &= \frac{1}{\mathcal{Z}} \left[\mathcal{Z}^{\text{bath}} \sum_x \sum_{\{\alpha\}_x} \partial_i w(\{\alpha\}_x) \right]. \end{aligned} \quad (\text{A14})$$

It holds that a given eigenbasis $\{(E_\alpha, |\alpha\rangle)\}$ of $\hat{H}^{\text{imp}}$ depends in the following way on the parameters θ :

$$\partial_i E_\alpha = \langle \alpha | \hat{h}_i | \alpha \rangle, \quad (\text{A15})$$

$$\partial_i |\alpha\rangle = \sum_{\alpha \neq \beta} \frac{\langle \beta | \hat{h}_i | \alpha \rangle}{E_\alpha - E_\beta} |\beta\rangle. \quad (\text{A16})$$

With the assumptions Eq. (A10) in place, all operators in the Hamiltonian are mutually commuting, implying that they

share a common set of eigenvectors. This property allows us to trivially evaluate these derivatives,

$$\partial_i E_\alpha = \epsilon_i^\alpha, \quad (\text{A17})$$

$$\partial_i |\alpha\rangle = 0, \quad (\text{A18})$$

with $\hat{h}_i |\alpha\rangle = \epsilon_i^\alpha |\alpha\rangle$. With these relationships it is straightforward to show that

$$\partial_j \Lambda_{\{\alpha\}_x} = 0, \quad (\text{A19})$$

$$\partial_j \langle \hat{H}^{\text{imp}} \rangle_{\{\alpha\}_x} = \langle \hat{h}_j \rangle_{\{\alpha\}_x}. \quad (\text{A20})$$

This allows the calculation of the derivative of w , which amounts to

$$-\partial_i w(\{\alpha\}_x) = \langle \hat{h}_i \rangle_{\{\alpha\}_x} w(\{\alpha\}_x). \quad (\text{A21})$$

This shows that w is a generating functional of moments of $\langle \hat{h}_i \rangle_{\{\alpha\}_x}$. We therefore write

$$\begin{aligned} -\partial_i \log(\mathcal{Z}) &= -\frac{1}{\mathcal{Z}} \left[\mathcal{Z}^{\text{bath}} \sum_x \sum_{\{\alpha\}_x} \partial_i w(\{\alpha\}_x) \right] \\ &= \mathbb{E}_w[\langle \hat{h}_i \rangle_{\{\alpha\}_x}]. \end{aligned} \quad (\text{A22})$$

Accordingly, the second-order derivative becomes

$$\begin{aligned} \partial_j \partial_i \log(\mathcal{Z}) &= \mathbb{E}_w[\langle \hat{h}_i \rangle_{\{\alpha\}_x} \langle \hat{h}_j \rangle_{\{\alpha\}_x}] \\ &\quad - \mathbb{E}_w[\langle \hat{h}_i \rangle_{\{\alpha\}_x}] \mathbb{E}_w[\langle \hat{h}_j \rangle_{\{\alpha\}_x}], \end{aligned} \quad (\text{A23})$$

which can be identified as the covariance. Since the covariance is a positive-semidefinite matrix, we have

$$\partial_j \partial_i \log(\mathcal{Z}) = \text{cov}_w[\langle \hat{h}_i \rangle_{\{\alpha\}_x}, \langle \hat{h}_j \rangle_{\{\alpha\}_x}] \succcurlyeq 0, \quad (\text{A24})$$

making the Hessian of $\log(\mathcal{Z})$ positive-semidefinite, and the same applies to the Hessian of the KLD loss function Eq. (5) under the aforementioned assumptions Eq. (A10).

APPENDIX B: EXTENDED IMPURITY BASIS

To construct the projectors of the extended MC Hamiltonian, we require the basis $|Q, S, T, S_z; m\rangle$ on the extended Fock space Eq. (50). The first step in the construction is to determine the charge $\hat{Q}$, total spin $\hat{S}^2$, and $\hat{S}_z$ eigenstates $|Q, S, S_z; m\rangle$. For the occupation basis, we use the following labeling convention:

$$|\text{orbital 2 (L), orbital 1 (R), local moment (M)}\rangle. \quad (\text{B1})$$

Thus, the Hilbert space is spanned by the states

$ -2, 1/2, S_z; m\rangle$	$ -1, 1, S_z; m\rangle$	$ -1, 0, S_z; m\rangle$
$ 0, 0, \uparrow\rangle$	$ \uparrow, 0, \uparrow\rangle$	$1/\sqrt{2}(\uparrow, 0, \downarrow\rangle - \downarrow, 0, \uparrow\rangle)$
$ 0, 0, \downarrow\rangle$	$ \downarrow, 0, \downarrow\rangle$	$1/\sqrt{2}(0, \uparrow, \downarrow\rangle - 0, \downarrow, \uparrow\rangle)$
	$1/\sqrt{2}(\uparrow, 0, \downarrow\rangle + \downarrow, 0, \uparrow\rangle)$	
	$ 0, \uparrow, \uparrow\rangle$	
	$ 0, \downarrow, \downarrow\rangle$	
	$1/\sqrt{2}(0, \uparrow, \downarrow\rangle + 0, \downarrow, \uparrow\rangle)$	

$ 0, 3/2, S_z; m\rangle$	$ 0, 1/2, S_z; m\rangle$	$ 0, 1/2, S_z; m\rangle$
$ \uparrow, \uparrow, \uparrow\rangle$	$\sqrt{2/3} \uparrow, \uparrow, \downarrow\rangle - \sqrt{1/6}(\downarrow, \uparrow, \uparrow\rangle + \uparrow, \downarrow, \uparrow\rangle)$	$ \uparrow\downarrow, 0, \uparrow\rangle$
$1/\sqrt{3}(\downarrow, \uparrow, \downarrow\rangle + \downarrow, \downarrow, \uparrow\rangle + \uparrow, \downarrow, \downarrow\rangle)$	$-\sqrt{2/3} \downarrow, \downarrow, \uparrow\rangle + \sqrt{1/6}(\downarrow, \uparrow, \downarrow\rangle + \uparrow, \downarrow, \downarrow\rangle)$	$ \uparrow\downarrow, 0, \downarrow\rangle$
$1/\sqrt{3}(\downarrow, \uparrow, \uparrow\rangle + \uparrow, \uparrow, \downarrow\rangle + \uparrow, \downarrow, \uparrow\rangle)$		$ 0, \uparrow\downarrow, \uparrow\rangle$
$ \downarrow, \downarrow, \downarrow\rangle$		$ 0, \uparrow\downarrow, \downarrow\rangle$
		$\sqrt{1/2}(\downarrow, \uparrow, \uparrow\rangle - \uparrow, \downarrow, \uparrow\rangle)$
		$\sqrt{1/2}(\downarrow, \uparrow, \downarrow\rangle - \uparrow, \downarrow, \downarrow\rangle)$

The states $|2, 1/2, S_z; m\rangle$, $|1, 1, S_z; m\rangle$, and $|1, 0, S_z; m\rangle$ can be obtained straightforwardly by replacing all 0 entries with $\uparrow\downarrow$. It is important to note that when using only Q and S quantum numbers, the $Q = -1, S = 1$ subspace is twofold-degenerate and the $Q = 0, S = 1/2$ subspace is threefold-degenerate. Therefore, using only the projectors onto the Q and S quantum number subspaces limits the expressibility of our effective model. Full tunability of the model requires that we can address each basis state individually. The next step is thus to lift these degeneracies. We may do this by diagonalizing the hopping operator in the degenerate subspaces,

$$\hat{T} = \sum_{\sigma} \hat{c}_{L\sigma}^{\dagger} \hat{c}_{R\sigma} + \hat{c}_{R\sigma}^{\dagger} \hat{c}_{L\sigma}. \quad (\text{B2})$$

In the $|QS\rangle = |\pm 1, 1\rangle$ and $|\pm 1, 0\rangle$ subspaces, we can find a common eigenbasis $|Q, S, T, S_z; m\rangle$ of $\hat{T}$ and $\hat{S}^2$ that allows us to completely lift the multiplet degeneracy, e.g.,

$ -1, 1, +1, S_z; m\rangle$	$ -1, 1, -1, S_z; m\rangle$
$1\sqrt{2}(\uparrow, 0, \uparrow\rangle + 0, \uparrow, \uparrow\rangle)$	$1\sqrt{2}(\uparrow, 0, \uparrow\rangle - 0, \uparrow, \uparrow\rangle)$
$1/\sqrt{4}(0, \uparrow, \downarrow\rangle + 0, \downarrow, \uparrow\rangle + \uparrow, 0, \downarrow\rangle + \downarrow, 0, \uparrow\rangle)$	$1/\sqrt{4}(0, \uparrow, \downarrow\rangle + 0, \downarrow, \uparrow\rangle - \uparrow, 0, \downarrow\rangle - \downarrow, 0, \uparrow\rangle)$
$1\sqrt{2}(\downarrow, 0, \downarrow\rangle + 0, \downarrow, \downarrow\rangle)$	$1\sqrt{2}(\downarrow, 0, \downarrow\rangle - 0, \downarrow, \downarrow\rangle)$

In the $Q = 0$ subspace, one can show that the hopping has no effect on the quadruplet and the doublet. However, the $Q = 0, S = 1/2$ subspace behaves nontrivially under the hopping:

$$\begin{aligned} \hat{T}|\uparrow\downarrow, 0, \uparrow\rangle &= |\downarrow, \uparrow, \uparrow\rangle - |\uparrow, \downarrow, \uparrow\rangle, \\ \hat{T}|0, \uparrow\downarrow, \uparrow\rangle &= |\downarrow, \uparrow, \uparrow\rangle - |\uparrow, \downarrow, \uparrow\rangle, \\ \hat{T}(|\downarrow, \uparrow, \uparrow\rangle - |\uparrow, \downarrow, \uparrow\rangle)/\sqrt{2} &= \sqrt{2}(|\uparrow\downarrow, 0, \uparrow\rangle + |0, \uparrow\downarrow, \uparrow\rangle). \end{aligned}$$

We can combine these states such that they become eigenstates of $\hat{T}$:

$ 0, 1/2, 0, S_z; m\rangle$	$ 0, 1/2, +2, S_z; m\rangle$	$ 0, 1/2, -2, S_z; m\rangle$
$\sqrt{\frac{2}{3}} \uparrow, \uparrow, \downarrow\rangle - \sqrt{\frac{1}{6}}(\downarrow, \uparrow, \uparrow\rangle + \uparrow, \downarrow, \uparrow\rangle)$	$\frac{1}{\sqrt{4}}(\downarrow, \uparrow, \uparrow\rangle - \uparrow, \downarrow, \uparrow\rangle + \uparrow\downarrow, 0, \uparrow\rangle + 0, \uparrow\downarrow, \uparrow\rangle)$	$\frac{1}{\sqrt{4}}(\downarrow, \uparrow, \uparrow\rangle - \uparrow, \downarrow, \uparrow\rangle - \uparrow\downarrow, 0, \uparrow\rangle - 0, \uparrow\downarrow, \uparrow\rangle)$
$-\sqrt{\frac{2}{3}} \downarrow, \downarrow, \uparrow\rangle + \sqrt{\frac{1}{6}}(\downarrow, \uparrow, \downarrow\rangle + \uparrow, \downarrow, \downarrow\rangle)$	$\frac{1}{\sqrt{4}}(\downarrow, \uparrow, \downarrow\rangle - \uparrow, \downarrow, \downarrow\rangle + \uparrow\downarrow, 0, \downarrow\rangle + 0, \uparrow\downarrow, \downarrow\rangle)$	$\frac{1}{\sqrt{4}}(\downarrow, \uparrow, \downarrow\rangle - \uparrow, \downarrow, \downarrow\rangle - \uparrow\downarrow, 0, \downarrow\rangle - 0, \uparrow\downarrow, \downarrow\rangle)$
$\frac{1}{\sqrt{2}}(\uparrow\downarrow, 0, \uparrow\rangle - 0, \uparrow\downarrow, \uparrow\rangle)$		
$\frac{1}{\sqrt{2}}(\uparrow\downarrow, 0, \downarrow\rangle - 0, \uparrow\downarrow, \downarrow\rangle)$		

The *fourfold* degenerate multiplet is now almost completely lifted. Only the $|0, 1/2, 0, S_z; m\rangle$ subspace is still twofold-degenerate. This last degeneracy can be lifted by introducing a further operator,

$$\hat{W} = \hat{n}_{\uparrow}^L \hat{n}_{\downarrow}^L + \hat{n}_{\uparrow}^R \hat{n}_{\downarrow}^R,$$

which has the eigenvalues $W \in \{0, 1\}$ in the $|0, 1/2, 0, S_z; m\rangle$ subspace. Instead of introducing a new label for only this subspace, the eigenvalues for $\hat{T}$ and $\hat{W}$ are added together in the T label. This gives the states

$ 0, 1/2, 1, S_z; m\rangle$	$ 0, 1/2, 0, S_z; m\rangle$
$\sqrt{\frac{2}{3}} \uparrow, \uparrow, \downarrow\rangle - \sqrt{\frac{1}{6}}(\downarrow, \uparrow, \uparrow\rangle + \uparrow, \downarrow, \uparrow\rangle)$	$\frac{1}{\sqrt{2}}(\uparrow\downarrow, 0, \uparrow\rangle - 0, \uparrow\downarrow, \uparrow\rangle)$
$-\sqrt{\frac{2}{3}} \downarrow, \downarrow, \uparrow\rangle + \sqrt{\frac{1}{6}}(\downarrow, \uparrow, \downarrow\rangle + \uparrow, \downarrow, \downarrow\rangle)$	$\frac{1}{\sqrt{2}}(\uparrow\downarrow, 0, \downarrow\rangle - 0, \uparrow\downarrow, \downarrow\rangle)$

The dimension of each subspace spanned by these labels has thereby been reduced to $M = 1$, as desired.

APPENDIX C: DETAILS OF NRG CALCULATIONS

All calculations in this work were carried out using Wilson's NRG method [50,83] using the full density matrix approach [67,84], which allows the precise calculation of the static observables discussed here. The impurity entropy, which is used to estimate T_K via $S_{\text{imp}}(T_K) = \frac{1}{2}$, was computed using standard thermodynamic NRG [50]. For the simulation of bare and effective models, the number of kept states is $M_K = 2000$ and we use a Wilson chain discretization parameter of $\Lambda = 2.5$ at a Wilson chain length of $N = 40$. Total charge and spin projection quantum numbers are exploited.

APPENDIX D: INTEGRATION WITH *AB INITIO* METHODS

At the heart of the UML method is the computation of local observables in bare and effective models. The effective models are simple enough that this can be done accurately at any temperature using, e.g., NRG. However, evaluating objects such as $\langle \hat{P}_{QS} \rangle$ in the bare model within an *ab initio* description is in general challenging.

For methods that operate in second quantization, we present here a straightforward way to construct the bare projectors $\hat{P}_{QS}$ that project onto the quantum number states $|Q, S\rangle$. First we introduce the auxiliary operators,

$$\hat{X}_Q = \hat{Q} - Q, \quad (\text{D1})$$

$$\hat{Y}_S = \hat{S}^2 - S(S+1), \quad (\text{D2})$$

where $\hat{S}^2$ is the total spin operator of the extended impurity, and $\hat{Q}$ is the total charge operator of the extended impurity. The operator $\hat{X}_Q$ eliminates all contributions with charge Q to a state $|\psi\rangle$ and similarly $\hat{Y}_S$ eliminates all contributions with total spin S . Using these operators, we can now write the projector onto the multiplet subspace of Q and S as

$$\hat{P}_{QS} = \frac{1}{N_{QS}} \prod_{Q' \neq Q} \hat{X}_{Q'} \times \prod_{S' \neq S} \hat{Y}_{S'}, \quad (\text{D3})$$

where

$$N_{QS} = \prod_{Q' \neq Q} (Q - Q') \prod_{S' \neq S} [S(S+1) - S'(S'+1)] \quad (\text{D4})$$

is a normalization constant. One can similarly construct the admissibility operator $\hat{\Omega}_{\text{ad}}$. However, these operators are manifestly of high order: Eq. (D3) is an N -body operator, where

N is the number of particles in the extended bare impurity Hilbert space.

Within an *ab initio* framework, expectation values of such N -body operators are typically intractable, because they are usually expressed in terms of reduced density matrices (RDMs) [85]. For a given basis of molecular orbitals $\{\phi_i\}_{i=1}^M$ one can express the spinless single particle RDM (1-RDM) as

$$\rho_1(r'; r) = \sum_{i,j}^M {}^1D_j^i \phi_i(r') \phi_j(r), \quad (\text{D5})$$

where ${}^1D_j^i = d_i \times d_j = \langle \psi | \hat{c}_i^\dagger \hat{c}_j | \psi \rangle$ is the 1-RDM, which encodes a single-particle wave function as

$$\psi(r) = \sum_i^M d_i \phi_i(r). \quad (\text{D6})$$

The 1-RDM is not a many-body object, although it is related to the 2-RDM 2D as [85]

$${}^2D_{st}^{pq} = 2 {}^1D_p^s \wedge {}^1D_t^q + {}^2\Delta_{st}^{pq}, \quad (\text{D7})$$

where ${}^2\Delta$ is a cumulant that encodes the relations between multiple determinants. The tensor $D_{st}^{pq} = \langle \psi | \hat{c}_p^\dagger \hat{c}_q^\dagger \hat{c}_s \hat{c}_t | \psi \rangle$ encodes two-body correlations. The construction of the N -RDM is therefore possible for an arbitrary number for particles N , but it becomes prohibitively expensive rather quickly.

Instead of attempting to compute Eq. (D3) exactly within this framework, we instead sketch a method for reducing the projectors $\hat{P}_{QS}$ to a simplified but approximate form that might be compatible with *ab initio* methods using one- or two-body RDMs.

We wish to find the best variational approximate representation to Eq. (D3) using at most two-body operators. We can do this using the FSP, Eq. (59). We therefore construct a full basis of operators for the bare Hilbert space,

$$\hat{A} \in \Xi_n \in \{ \{ \hat{c}_i^\dagger \hat{c}_j \}, \{ \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k \hat{c}_l \}, \{ \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k^\dagger \hat{c}_l^\dagger \hat{c}_m \hat{c}_n \hat{c}_o \hat{c}_p \}, \dots \}. \quad (\text{D8})$$

Now retaining only the one- and two-body terms, we have a basis $\{\hat{h}_i\}$ for the approximate representation of $\hat{P}_{QS}$. For a complete basis Ξ_n , the decomposition $\hat{P}_{QS} = \sum_i \theta_i \hat{h}_i$ is unique and can be achieved like a regular spectral decomposition of a vector using the FSP, viz.,

$$\theta_i = \frac{\langle \hat{P}_{QS}, \hat{h}_i \rangle}{\|\hat{h}_i\|}, \quad \hat{h}_i \in \Xi_n. \quad (\text{D9})$$

However, we note that the basis Eq. (D8) is generally overcomplete and therefore any representation of $\hat{P}_{QS}$ according to Eq. (59) is not necessarily unique. Constructing a complete

operator basis on the bare Hilbert space can be challenging. On the other hand, the variational approach can find optimal representations of any operator even for an overcomplete basis, and the value of the optimized loss function

can serve as an error estimation for the decomposition. We argue that this is a controlled approach to obtain two-body approximations of the UML operators $\hat{P}_{QS}$. We leave development and implementation of these methods for future work.

-
- [1] D. J. Thouless, *The Quantum Mechanics of Many-body Systems* (Academic Press, New York, NY, 1964).
 - [2] J. Hubbard, Electron correlations in narrow energy bands, *Proc. R. Soc. London A* **276**, 238 (1963).
 - [3] J. Kanamori, Electron correlation and ferromagnetism of transition metals, *Prog. Theor. Phys.* **30**, 275 (1963).
 - [4] A. Calvo-Fernández, M. Kumar, D. Soler-Polo, A. Eiguren, M. Blanco-Rey, and P. Jelínek, Theoretical model for multi-orbital Kondo screening in strongly correlated molecules with several unpaired electrons, [arXiv:2405.15958](https://arxiv.org/abs/2405.15958).
 - [5] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein, Frequency-dependent local interactions and low-energy effective models from electronic structure calculations, *Phys. Rev. B* **70**, 195104 (2004).
 - [6] E. Şaşıoğlu, C. Friedrich, and S. Blügel, Effective coulomb interaction in transition metals from constrained random-phase approximation, *Phys. Rev. B* **83**, 121101(R) (2011).
 - [7] N. P. Bauman and K. Kowalski, Coupled cluster downfolding theory: towards universal many-body algorithms for dimensionality reduction of composite quantum systems in chemistry and materials science, *Mater. Theor.* **6**, 17 (2022).
 - [8] H. Zheng, H. J. Changlani, K. T. Williams, B. Busemeyer, and L. K. Wagner, From real materials to model Hamiltonians with density matrix downfolding, *Front. Phys.* **6**, 43 (2018).
 - [9] R. Pariser and R. G. Parr, A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I., *J. Chem. Phys.* **21**, 466 (1953); A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II, **21**, 767 (1953).
 - [10] H. C. Nguyen, R. Zecchina, and J. Berg, Inverse statistical problems: from the inverse ising problem to data science, *Adv. Phys.* **66**, 197 (2017).
 - [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2017).
 - [12] H. S. Bhat, K. Ranka, and C. M. Isborn, Machine learning a molecular Hamiltonian for predicting electron dynamics, *Int. J. Dyn. Control* **8**, 1089 (2020).
 - [13] Z. Wang, S. Ye, H. Wang, J. He, Q. Huang, and S. Chang, Machine learning method for tight-binding Hamiltonian parameterization from *ab-initio* band structure, *npj Comput. Mater.* **7**, 11 (2021).
 - [14] R. C. Sawaya and S. R. White, Constructing Hubbard models for the hydrogen chain using sliced-basis density matrix renormalization group, *Phys. Rev. B* **105**, 045145 (2022).
 - [15] M. Nakhaee, S. Ketabi, and F. Peeters, Machine learning approach to constructing tight binding models for solids with application to BiTeCl, *J. Appl. Phys.* **128**, 215107 (2020).
 - [16] M. Prüfer, T. V. Zache, P. Kunkel, S. Lannig, A. Bonnin, H. Strobel, J. Berges, and M. K. Oberthaler, Experimental extraction of the quantum effective action for a non-equilibrium many-body system, *Nat. Phys.* **16**, 1012 (2020).
 - [17] S. Yu, Y. Gao, B.-B. Chen, and W. Li, Learning the effective spin Hamiltonian of a quantum magnet, *Chin. Phys. Lett.* **38**, 097502 (2021).
 - [18] T. Schuster, M. Niu, J. Cotler, T. O'Brien, J. R. McClean, and M. Mohseni, Learning quantum systems via out-of-time-order correlators, [arXiv:2208.02254](https://arxiv.org/abs/2208.02254).
 - [19] I. Buluta and F. Nori, Quantum simulators, *Science* **326**, 108 (2009); I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, *Rev. Mod. Phys.* **86**, 153 (2014).
 - [20] P. Barthélemy and L. M. Vandersypen, Quantum dot systems: a versatile platform for quantum simulations, *Ann. Phys.* **525**, 808 (2013); T. Hensgens, T. Fujita, L. Janssen, X. Li, C. Van Diepen, C. Reichl, W. Wegscheider, S. Das Sarma, and L. M. Vandersypen, Quantum simulation of a fermi-hubbard model using a semiconductor quantum dot array, *Nature (London)* **548**, 70 (2017).
 - [21] W. Pouse, L. Peeters, C. L. Hsueh, U. Gennser, A. Cavanna, M. A. Kastner, A. K. Mitchell, and D. Goldhaber-Gordon, Quantum simulation of an exotic quantum critical point in a two-site charge Kondo circuit, *Nat. Phys.* **19**, 492 (2023).
 - [22] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, Sample-efficient learning of interacting quantum systems, *Nat. Phys.* **17**, 931 (2021).
 - [23] N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory, Robust method for estimating the lindblad operators of a dissipative quantum process from measurements of the density operator at multiple time points, *Phys. Rev. A* **67**, 042322 (2003).
 - [24] J. Zhang and M. Sarovar, Quantum Hamiltonian identification from measurement time traces, *Phys. Rev. Lett.* **113**, 080401 (2014).
 - [25] M. Kieferová and N. Wiebe, Tomography and generative training with quantum Boltzmann machines, *Phys. Rev. A* **96**, 062327 (2017).
 - [26] T. Olsacher, L. Pastori, C. Kokail, L. M. Sieberer, and P. Zoller, Digital quantum simulation, learning of the floquet Hamiltonian, and quantum chaos of the kicked top, *J. Phys. A* **55**, 334003 (2022).
 - [27] S. Nandy, M. Schmitt, M. Bukov, and Z. Lenarčič, Reconstructing effective Hamiltonians from nonequilibrium (pre-) thermal steady states, *Phys. Rev. Res.* **6**, 023160 (2024).
 - [28] K. G. Wilson, Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture, *Phys. Rev. B* **4**, 3174 (1971).
 - [29] S. D. Glazek and K. G. Wilson, Renormalization of Hamiltonians, *Phys. Rev. D* **48**, 5863 (1993).
 - [30] P. Kopietz, L. Bartosch, and F. Schütz, *Introduction to the Functional Renormalization Group* (Springer Science & Business Media, 2010), Vol. 798.
 - [31] C. Béný and T. J. Osborne, Information-geometric approach to the renormalization group, *Phys. Rev. A* **92**, 022330 (2015).
 - [32] J. R. Garrison and T. Grover, Does a single eigenstate encode the full Hamiltonian? *Phys. Rev. X* **8**, 021026 (2018).

- [33] A. C. Hewson, *The Kondo Problem to Heavy Fermions*, Cambridge Studies in Magnetism (Cambridge University Press, 1993).
- [34] D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, and M. Kastner, Kondo effect in a single-electron transistor, *Nature (London)* **391**, 156 (1998).
- [35] A. K. Mitchell, K. G. L. Pedersen, P. Hedegård, and J. Paaske, Kondo blockade due to quantum interference in single-molecule junctions, *Nat. Commun.* **8**, 15210 (2017).
- [36] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [37] G. Mihailescu, A. Kiely, and A. K. Mitchell, Quantum sensing with nanoelectronics: Fisher information for an adiabatic perturbation, [arXiv:2406.18662](https://arxiv.org/abs/2406.18662).
- [38] G. Mihailescu, A. Bayat, S. Campbell, and A. K. Mitchell, Multiparameter critical quantum metrology with impurity probes, *Quantum Sci. Technol.* **9**, 035033 (2024); G. Mihailescu, S. Campbell, and A. K. Mitchell, Thermometry of strongly correlated fermionic quantum systems using impurity probes, *Phys. Rev. A* **107**, 042614 (2023).
- [39] J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruña, P. L. McEuen, and D. C. Ralph, Coulomb blockade and the Kondo effect in single-atom transistors, *Nature (London)* **417**, 722 (2002).
- [40] G. D. Scott and D. Natelson, Kondo resonances in molecular devices, *ACS Nano* **4**, 3560 (2010).
- [41] C. M. Guédon, H. Valkenier, T. Markussen, K. S. Thygesen, J. C. Hummelen, and S. J. Van Der Molen, Observation of quantum interference in molecular charge transport, *Nat. Nanotechnol.* **7**, 305 (2012).
- [42] S. Sen and A. K. Mitchell, Many-body quantum interference route to the two-channel kondo effect: Inverse design for molecular junctions and quantum dot devices, *Phys. Rev. Lett.* **133**, 076501 (2024).
- [43] A. Keller, S. Amasha, I. Weymann, C. Moca, I. Rau, J. Katine, H. Shtrikman, G. Zaránd, and D. Goldhaber-Gordon, Emergent $s(4)$ kondo physics in a spin-charge-entangled double quantum dot, *Nat. Phys.* **10**, 145 (2014).
- [44] Z. Iftikhar, A. Anthore, A. Mitchell, F. Parmentier, U. Gennser, A. Ouerghi, A. Cavanna, C. Mora, P. Simon, and F. Pierre, Tunable quantum criticality and super-ballistic transport in a "charge" kondo circuit, *Science* **360**, 1315 (2018); A. K. Mitchell, L. A. Landau, L. Fritz, and E. Sela, Universality and scaling in a charge two-channel kondo device, *Phys. Rev. Lett.* **116**, 157202 (2016).
- [45] A. K. Mitchell, A. Liberman, E. Sela, and I. Affleck, $SO(5)$ non-Fermi liquid in a Coulomb box device, *Phys. Rev. Lett.* **126**, 147702 (2021).
- [46] J. B. Rigo and A. K. Mitchell, Machine learning effective models for quantum systems, *Phys. Rev. B* **101**, 241105(R) (2020).
- [47] S.-i. Amari, *Manifold, divergence and dually flat structure*, in *Information Geometry and Its Applications* (Springer Japan, Tokyo, 2016), pp. 3–30.
- [48] P. Seth, I. Krivenko, M. Ferrero, and O. Parcollet, TRIQS/CTHYB: A continuous-time quantum Monte Carlo hybridisation expansion solver for quantum impurity problems, *Comput. Phys. Commun.* **200**, 274 (2016).
- [49] H. R. Krishna-Murthy, J. W. Wilkins, and K. G. Wilson, Renormalization-group approach to the anderson model of dilute magnetic alloys. I. Static properties for the symmetric case, *Phys. Rev. B* **21**, 1003 (1980).
- [50] R. Bulla, T. A. Costi, and T. Pruschke, Numerical renormalization group method for quantum impurity systems, *Rev. Mod. Phys.* **80**, 395 (2008).
- [51] A. K. Mitchell, M. R. Galpin, S. Wilson-Fletcher, D. E. Logan, and R. Bulla, Generalized Wilson chain for solving multichannel quantum impurity problems, *Phys. Rev. B* **89**, 121105(R) (2014); K. M. Stadler, A. K. Mitchell, J. von Delft, and A. Weichselbaum, Interleaved numerical renormalization group as an efficient multiband impurity solver, **93**, 235101 (2016).
- [52] Alternative optimization methods in computational physics have employed exact gradients from automatically differentiable algorithms; see, e.g., Refs. [86–88].
- [53] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Quantifying entanglement, *Phys. Rev. Lett.* **78**, 2275 (1997).
- [54] H. J. Kappen, Learning quantum models from quantum or classical data, *J. Phys. A* **53**, 214001 (2020).
- [55] K. Haule, Quantum monte carlo impurity solver for cluster dynamical mean-field theory and electronic structure calculations with adjustable cluster base, *Phys. Rev. B* **75**, 155113 (2007).
- [56] J. Li, M. Wallerberger, and E. Gull, Diagrammatic Monte Carlo method for impurity models with general interactions and hybridizations, *Phys. Rev. Res.* **2**, 033211 (2020).
- [57] J. Duchi, E. Hazan, and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Machine Learn. Res.* **12**, 2121 (2011).
- [58] J. Linderberg and Y. öhrn, Derivation and analysis of the Pariser–Parr–Pople model, *J. Chem. Phys.* **49**, 716 (1968).
- [59] A. Allerdt and A. E. Feiguin, A numerically exact approach to quantum impurity problems in realistic lattice geometries, *Front. Phys.* **7**, 67 (2019).
- [60] B. P. Dolan, A geometrical interpretation of renormalization group flow, *Int. J. Mod. Phys. A* **09**, 1261 (1994).
- [61] F. B. Kugler and J. von Delft, Multiloop functional renormalization group for general models, *Phys. Rev. B* **97**, 035162 (2018).
- [62] A. Hewson, A. Oguri, and D. Meyer, Renormalized parameters for impurity models, *Eur. Phys. J. B* **40**, 177 (2004).
- [63] J. R. Schrieffer and P. A. Wolff, Relation between the Anderson and Kondo Hamiltonians, *Phys. Rev.* **149**, 491 (1966).
- [64] R. Chan and M. Gulácsi, The exact Schrieffer–Wolff transformation, *Philos. Mag.* **84**, 1265 (2004).
- [65] F. D. M. Haldane, Theory of the Atomic limit of the Anderson model. I. Perturbation expansions re-examined, *J. Phys. C* **11**, 5015 (1978).
- [66] P. W. Anderson, A poor man's derivation of scaling laws for the Kondo problem, *J. Phys. C* **3**, 2436 (1970).
- [67] A. Weichselbaum and J. von Delft, Sum-rule conserving spectral functions from the numerical renormalization group, *Phys. Rev. Lett.* **99**, 076402 (2007).
- [68] L. G. G. V. Dias da Silva, F. Heidrich-Meisner, A. E. Feiguin, C. A. Büsser, G. B. Martins, E. V. Anda, and E. Dagotto, Transport properties and kondo correlations in nanostructures: Time-dependent dmrg method applied to quantum dots coupled to Wilson chains, *Phys. Rev. B* **78**, 195317 (2008).

- [69] A. Holzner, I. P. McCulloch, U. Schollwöck, J. von Delft, and F. Heidrich-Meisner, Kondo screening cloud in the single-impurity anderson model: A density matrix renormalization group study, *Phys. Rev. B* **80**, 205114 (2009).
- [70] T. Zhu, L. Peng, H. Zhai, Z.-H. Cui, and G. K. Chan, Towards an exact electronic quantum many-body treatment of Kondo correlation in magnetic impurities, [arXiv:2405.18709](https://arxiv.org/abs/2405.18709).
- [71] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Mod. Phys.* **73**, 33 (2001).
- [72] R. J. Bartlett and M. Musiał, Coupled cluster theory in quantum chemistry, *Rev. Mod. Phys.* **79**, 291 (2007).
- [73] G. Sugiyama and S. Koonin, Auxiliary field Monte-Carlo for quantum many-body ground states, *Ann. Phys.* **168**, 1 (1986).
- [74] M. Motta and S. Zhang, *Ab initio* computations of molecular systems by the auxiliary-field quantum Monte Carlo method, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1364 (2018).
- [75] A. K. Mitchell, T. F. Jarrold, and D. E. Logan, Quantum phase transition in quantum dot trimers, *Phys. Rev. B* **79**, 085124 (2009); A. K. Mitchell, T. F. Jarrold, M. R. Galpin, and D. E. Logan, Local moment formation and Kondo screening in impurity trimers, *J. Phys. Chem. B* **117**, 12777 (2013).
- [76] T. Numata, Y. Nisikawa, A. Oguri, and A. C. Hewson, Kondo effects in a triangular triple quantum dot: Numerical renormalization group study in the whole region of the electron filling, *Phys. Rev. B* **80**, 155330 (2009).
- [77] A. K. Mitchell and D. E. Logan, Two-channel Kondo phases and frustration-induced transitions in triple quantum dots, *Phys. Rev. B* **81**, 075126 (2010).
- [78] L. Che, C. Wei, Y. Huang, D. Zhao, S. Xue, X. Nie, J. Li, D. Lu, and T. Xin, Learning quantum Hamiltonians from single-qubit measurements, *Phys. Rev. Res.* **3**, 023246 (2021).
- [79] C. Yang and A. E. Feiguin, Unveiling the internal entanglement structure of the Kondo singlet, *Phys. Rev. B* **95**, 115106 (2017).
- [80] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, Continuous-time quantum Monte Carlo method for fermions, *Phys. Rev. B* **72**, 035122 (2005).
- [81] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, *Rev. Mod. Phys.* **83**, 349 (2011).
- [82] P. Werner and A. J. Millis, Hybridization expansion impurity solver: General formulation and application to Kondo lattice and two-orbital models, *Phys. Rev. B* **74**, 155107 (2006).
- [83] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, *Rev. Mod. Phys.* **47**, 773 (1975).
- [84] F. B. Anders and A. Schiller, Real-time dynamics in quantum-impurity systems: A time-dependent numerical renormalization-group approach, *Phys. Rev. Lett.* **95**, 196801 (2005).
- [85] D. A. Mazziotti, Two-electron reduced density matrix as the basic variable in many-electron quantum chemistry and physics, *Chem. Rev.* **112**, 244 (2012).
- [86] H. Xie, J.-G. Liu, and L. Wang, Automatic differentiation of dominant eigensolver and its applications in quantum physics, *Phys. Rev. B* **101**, 245139 (2020).
- [87] L. Coopmans, D. Luo, G. Kells, B. K. Clark, and J. Carrasquilla, Protocol discovery for the quantum control of majoranas by differential programming and natural evolution strategies, *PRX Quantum* **2**, 020332 (2021).
- [88] I. Luchnikov, M. Krechetov, and S. Filippov, Riemannian optimization and automatic differentiation for complex quantum architectures, *New J. Phys.* **23**, 073006 (2021).