Downfolding from ab initio to interacting model Hamiltonians: comprehensive analysis and benchmarking of the DFT+cRPA approach

Check for updates

Yueqing Chang[1], Erik G. C. P. van Loon [2], Brandon Eskridge [3,4], Brian Busemeyer[4], Miguel A. Morales[4], Cyrus E. Dreyer[4,5], Andrew J. Millis [4,6], Shiwei Zhang [4], Tim O. Wehling[7,8], Lucas K. Wagner [9] ✉ & Malte Rösner [10] ✉

Model Hamiltonians are regularly derived from first principles to describe correlated matter. However, the standard methods for this contain a number of largely unexplored approximations. For a strongly correlated impurity model system, here we carefully compare a standard downfolding technique with the best possible ground-truth estimates for charge-neutral excited-state energies and wave functions using state-of-the-art first-principles many-body wave function approaches. To this end, we use the vanadocene molecule and analyze all downfolding aspects, including the Hamiltonian form, target basis, double-counting correction, and Coulomb interaction screening models. We find that the choice of target-space basis functions emerges as a key factor for the quality of the downfolded results, while orbital-dependent double-counting corrections diminish the quality. Background screening of the Coulomb interaction matrix elements primarily affects crystal-field excitations. Our benchmark uncovers the relative importance of each downfolding step and offers insights into the potential accuracy of minimal downfolded model Hamiltonians.

The computational cost of solving electronic Hamiltonians increases rapidly with the size of the electronic Hilbert space, i.e., with the number of orbitals and electrons. This presents a substantial constraint on electronic structure calculations for molecules and an essential problem for the study of correlated solids. To overcome this dilemma, it is desirable to construct models with fewer electronic degrees of freedom in a systematic and controllable way. For this purpose, one needs to select a smaller *target* Hilbert space and determine the structure of the target Hamiltonian and its matrix elements. In practice, typically, no more than a handful of orbitals per unit cell are kept in the target space. Altogether, this procedure is known as downfolding[1] and serves as a bridge connecting ab initio methods such as density functional or GW theory with higher-level many-body treatments including (extended) dynamical mean field[2] or dual theories[3,4], ab initio DΓA[5], self-energy

embedding[6] or Eliashberg theory[7], which cannot be performed on the full Hilbert space. There have been various suggestions for these downfolding schemes, ranging from constrained density-functional perturbation theory for phononic degrees of freedom[8–10] to the constrained random-phase approximation (cRPA)[11], constrained density-functional theory[12], constrained GW[13] and constrained functional renormalization group[14] approaches for purely electronic Hamiltonians, next to embedding schemes, e.g., starting from coupled-cluster references[15,16] or via projection schemes[17]. Our focus here is on interacting target space Hamiltonians of the form $H = \hat{t}_{\alpha\beta} c_\alpha^\dagger c_\beta + \hat{U}_{\alpha\beta\gamma\delta} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta$, where $\hat{U}$ is a static Coulomb interaction in the target space, $\hat{t}$ defines the single-particle energies and the Greek indices run over the electronic target space. Generalized Hubbard model Hamiltonians of this form are the standard in "DFT++"[18] (or density-functional

[1]Center for Materials Theory, Department of Physics & Astronomy, Rutgers University, Piscataway, NJ 08854, USA. [2]NanoLund and Division of Mathematical Physics, Department of Physics, Lund University, Lund, Sweden. [3]Department of Physics, College of William and Mary, Williamsburg, VA 23187, USA. [4]Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA. [5]Department of Physics & Astronomy, Stony Brook University, Stony Brook, NY 11794, USA. [6]Department of Physics, Columbia University, New York, NY 10027, USA. [7]I. Institute for Theoretical Physics, Universität Hamburg, Notkestraße 9-11, 22607 Hamburg, Germany. [8]The Hamburg Centre for Ultrafast Imaging, 22761 Hamburg, Germany. [9]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [10]Institute for Molecules and Materials, Radboud University, 6525 AJ Nijmegen, The Netherlands.
✉e-mail: lkwagner@illinois.edu; m.roesner@science.ru.nl

theory based embedding) approaches to describe correlation effects in solids for which various solvers have been developed and implemented.

Although the idea of downfolding is conceptually simple, and reductions of the Hilbert space are ubiquitous in quantum physics, the question has remained of *how* one should do this in a systematic and practical way. By definition, the original Hilbert space is too large to do exact calculations, so one needs to use approximate methods whose errors are usually not controllable. At the same time, one needs to avoid double counting of correlation effects. This arises because of the interplay between the two-body term $\hat{U}$ and the single-particle term $\hat{t}$ in the Hamiltonian. The latter is frequently extracted from a density-functional theory (DFT) calculation that already contains some interaction and correlation effects. Several single-particle double-counting corrections to $\hat{t}$ exist in the literature[19–22], with some physical arguments supporting their use, but there is no general consensus on how to tackle this problem. In fact, the uncertainty surrounding these single-particle double-counting corrections can be a limiting factor for the accuracy of DFT++ approaches, see, e.g., refs. 23,24 for further discussions.

In terms of the interaction in the target space, downfolding a priori generates many-fermion interactions of arbitrary order[14] and not just one- and two-particle terms. Furthermore, these interactions might be non-instantaneous (frequency-dependent)[11], requiring an action rather than a Hamiltonian formalism. Even if one accepts the restriction to a static two-particle Coulomb interaction tensor, its strength should be different from the corresponding Coulomb interaction in the full space to account for the screening by electrons outside of the target space, while screening by the other electrons inside the target space should not be included. The latter is to avoid double counting of screening processes to the Coulomb interaction $\hat{U}$ within the target space. A variety of constrained methods[8–13,25,26] has been developed to avoid this two-particle double-counting issue, e.g., via partial (constrained) screening approaches.

Evaluating the performance of downfolding techniques has proven challenging in practice, as the matrix elements of the downfolded model are not measurable quantities themselves and as the models are often solved using approximate solvers. This multi-step procedure makes it hard to establish how accurate the downfolding by itself is since deviations from full first-principles calculations or experimental data might also originate in approximations made while solving the low-energy model.

Given the widespread use of cRPA calculations based on DFT input to derive interacting model Hamiltonians and their promise of accurately describing correlation effects at low numerical costs, it is critical to quantify their accuracy. To this end, we analyze here the sensitivity of this method at the hand of the vanadocene molecule. The latter is a good proxy for correlated bulk materials or embedded correlated defects with well-defined interacting sub-spaces within gapped semiconducting backgrounds (referred to as rest spaces). Because vanadocene is a relatively small system, we are able to use state-of-the-art first-principles methods to derive reliable reference data, and the downfolded models are solved exactly, which allows us to focus on errors in the downfolding procedure. To avoid obtaining correct spectra with incorrect states, we assess both energies and the character of the states. We systematically assess how the model Hamiltonian form, the target space basis, the impact of double-counting corrections, and the (constrained) screening models affect the ground state and excited state energies and wave functions.

In the following, we first explain our benchmarking strategy and introduce the vanadocene test system. Afterward, we discuss the first-principles reference data before we present our step-wise benchmark, showing how each downfolding step affects the comparison to the reference system. The impact of our findings is discussed at the end.

Results
Benchmarking strategy
To obtain accurate reference data, we use a combination of several state-of-the-art first-principles methods which each provides systematically improvable many-body wave functions to treat the electronic correlations at an attainable cost. Specifically, we apply equation of motion coupled cluster

(EOM-CC)[27,28], auxiliary-field quantum Monte Carlo (AFQMC)[29], and fixed-node diffusion Monte Carlo (DMC)[30], which have been shown to accurately describe the excited states for transition metal molecules[31–34].

As the test system, we have chosen the unperturbed eclipsed vanadocene molecule VCp_2[35], which consists of a V atom situated between two planar and parallel C_5H_5 rings as illustrated in Fig. 1. The unperturbed molecule has D_{5h} symmetry with a five-fold rotation axis, five two-fold rotational axes orthogonal to the five-fold rotational axis, and a mirror plane in the xy-plane. The vanadocene molecule may also exist in a staggered geometry with D_{5d} symmetry. Previous DFT calculations predict a difference of only 13 meV in the ground state energy between the D_{5h} and D_{5d} structures with the D_{5h} structure as the global ground state, consistent with experiment[36,37]. On the level of DFT, there are 5 V $3d$ dominated Kohn-Sham states forming the highest occupied and the lowest unoccupied molecular orbitals, which host 3 electrons in total. The D_{5h} symmetry of the crystal field generated by the carbon rings leads to two doubly degenerate (e_1, e_2) and one non-degenerate (a_1) single-particle energies for these d-dominated molecular orbitals as depicted in Fig. 1. This way, VCp_2 forms an optimal test bed for cRPA-based ab initio downfolding approaches: it hosts a well-defined partially filled correlated target space spanned by the V d orbitals, which is well disentangled from the carbon ring "background" electronic structure, which has a significant single-particle gap of more than 6 eV. The latter is important to guarantee that the RPA treatment of the background screening is adequate[38].

Our goal is to benchmark the effects of electronic correlations in this molecule. For this purpose, we use the same frozen positions of the nuclei in all methods and do not consider any electron-phonon coupling. Furthermore, all methods start from the same ccECP pseudopotentials[39] and ignore spin-orbit coupling. In this sense, our benchmark should be understood as a comparison of several computational methods for the interacting electron Hamiltonian defined by this pseudopotential and these atomic positions rather than as a benchmark to experimental data.

For the downfolding, we start from conventional DFT calculations and project onto a set of localized orbitals. These are subsequently used to perform constrained random-phase approximation[1,11,40] (cRPA) calculations, which are widely used for deriving Coulomb interaction matrix elements for downfolded models[41–43]. The resulting downfolded models are evaluated using exact diagonalization to avoid any artifacts from approximate model solutions.

Using the reference data from accurate quantum chemistry calculations, we assess the effects of changing the downfolding procedure as a sort of sensitivity analysis, in which an independent variable (the choice in downfolding procedure) is varied to determine the sensitivity of the output (the eigenstates of the embedded problem). For each step in the DFT+cRPA procedure, we consider several reasonable and common choices and analyze the sensitivity of the results.
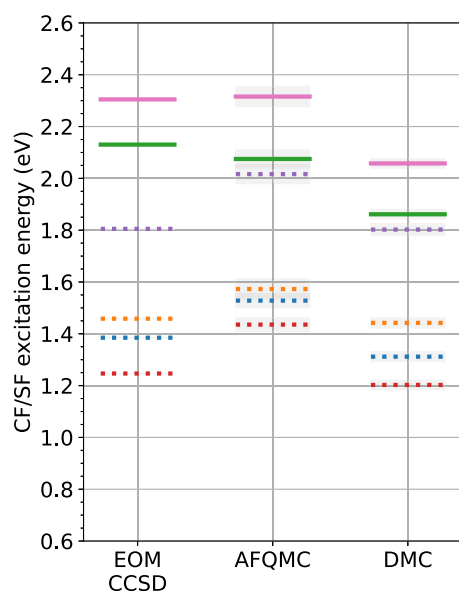
First-principles reference
Before we delve into the benchmark of the downfolded results, we first establish the first-principles references for the vanadocene molecule. In this study, we focus on the lowest four spin-flip and the lowest two crystal-field excitations, using three first-principles methods: EOM-CCSD, AFQMC, and DMC. They are among the most accurate quantum chemical methods for the investigation of excited states in correlated systems. Calculations have been systematically converged, and details of the methods can be found in the section "Methods". Calculations have been performed without imposing specific point group symmetry properties on many-body wave functions.

We found with all reference methods that the ground state is $^4A_2$ with $S = 3/2$, consistent with previous experimental and theoretical studies[36,37,44–47]. Specifically, the DMC ansatzes were chosen to be of $S_z = 1/2$ on the $S = 3/2$ quartet. The configuration of the DMC ground state is determined to be $(e_2)^2(a_1)^1(e_1)^0$, characterized by the one-body reduced density matrix of a natural orbital basis (see section "Methods" for details). The symmetry of the many-body state computed using AFQMC is inherited
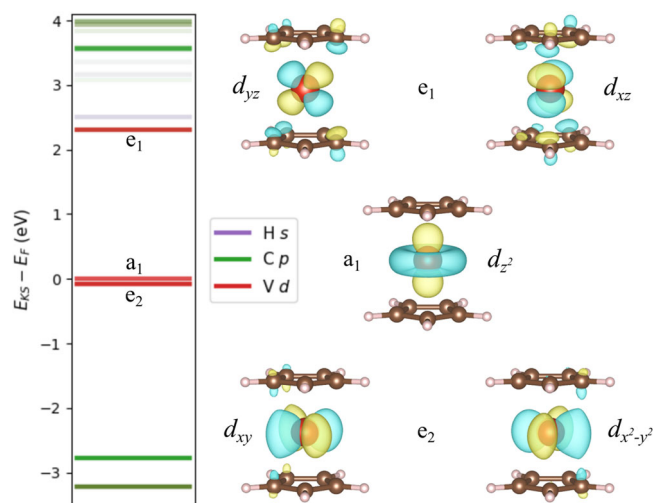
Fig. 1 | Single-particle properties. Left: DFT Kohn-Sham energies. Purple, green, and red colors indicate the projected orbital weight of each Kohn-Sham state, being mostly H s, C p, or V d, respectively. Red lines mark single-particle energies resulting from diagonalization of the downfolded single-particle Hamiltonian, which perfectly overlaps with the Kohn-Sham ones. Right: Maximally localized Wannier functions used within the downfolding procedure.

Fig. 2 | First-principles excitation energies obtained using EOM-CCSD, AFQMC, and DMC. The ground state energies obtained from all the methods have been shifted to align at zero. Dashed and solid lines represent spin-flip (SF) and crystal-field (CF) excitations, respectively, and the shaded regions around the energy levels show statistical uncertainties.

from the trial wave function used. For the ground state, AFQMC calculations were performed using a CASSCF trial wave function with $(e_2)^2(a_1)^1(e_1)^0$ 3d-orbital occupancy.

Figure 2 shows a summary of the charge-neutral excitation spectra obtained using these first-principles methods. The ground state energies are aligned at zero, and the dashed (solid) lines represent spin-flip (crystal-field) excitations, identified by states' zero (finite) occupation on the high-energy e_1 manifold. All three first-principles methods are in qualitative agreement regarding the order and degeneracy of the spin-flip and crystal-field excitation energies. Quantitatively, all methods agree to within about 200 meV.

The energy difference between excitation energies of the same type (i.e., within either the set of spin-flip excitations or the set of crystal-field

excitations) is remarkably similar across all methods. According to Fig. 2, the lowest three spin-flip excitation energies are found to be within windows of < 250 meV, while the fourth spin-flip excitation energy is higher than the lowest three spin-flip states by 350 meV (DMC). The crystal-field excitations are identified to be higher than the lowest four spin-flip excitations. The two crystal-field excitations are found to be separated by about 200 meV for each first-principles method. Overall, the first-principles methods obtain the same ordering and character of the many-body eigenstates of interest but differ in energy by approximately 60–200 meV.

In the following sections, we proceed with the discussion of the six lowest charge-neutral many-body excitations with $S_z = 1/2$, which take place within the V d shell and are thus describable within the minimal subspace of the downfolding calculations.

Downfolding results

Now that we are familiar with the test system and the first-principles reference data, we turn to the results obtained using our downfolding procedure. We aim to systematically test and scrutinize all relevant steps in the DFT+cRPA procedure, especially those for which multiple strategies exist in the literature such that choices need to be made. Specifically, we will discuss the chosen form of the model Hamiltonian, the target space basis, the single-particle double-counting procedure, and finally, the Coulomb screening model. The downfolding procedure is qualitatively assessed based on the predicted ground state properties, the ordering of the excitation energies, the many-body characteristics of the excited states, and the shape of their charge densities. To quantify the agreement of the (charge-neutral) excitation energies of the downfolded model with the first-principles reference data, we further calculate

$$\chi^2(\{E_n\}) = \frac{1}{2\sigma^2}\sum_{i,n}\left(E_n^{\mathrm{ref},i} - E_n\right)^2, \qquad (1)$$

where E_n are the charge-neutral excitation energies in the downfolded model and $E_n^{\mathrm{ref},i}$ are the corresponding excitation energies according to the ith first-principles method with $i \in \{$EOM-CCSD,AFQMC,DMC$\}$. χ^2 is the log-likelihood from a Bayesian inference of the embedding model probability, based on the reference values. It naturally accounts for the fact that some of the states are in close agreement among the reference sets, and some are not. We determine σ as the average over the variances of the n excitations within the reference set (EOM-CCSD, AFQMC, DMC), which results in $\sigma \approx 0.1$ eV. χ^2 accounts for the fact that the reference data itself has errors, and so it represents our uncertainty about the exact result.

The purpose of this quantifier is to summarize how consistent a given downfolding procedure is with the excitation spectrum of the reference data in a single number, given that there is uncertainty within the reference data set itself. The smaller χ^2, the better the match.

Note that this procedure requires us to identify which excitations belong together (the label n, the colors of the energy levels in the figures). We achieve this based on the quantum numbers, the spatial structure of the spin-resolved charge densities ρ_n^σ, and the many-body characteristics of the corresponding excited wave functions for each state n as explained in the section "Wannier basis sets".

We consider several variations of the DFT+cRPA procedure. The choice of localized orbitals for the embedded space was either constructed using maximally localized Wannier functions[48] (MLWF) or first-guess Wannier functions (FGWF). The single-particle energies were constructed either using double-counting corrections, no double-counting corrections, and by explicitly modifying crystal-field energies, which emulates using different DFT functionals to create the one-particle energies. We then move to considering the construction of the interacting part of the downfolded Hamiltonian by considering mono- and miltipole screening channels and the frequency dependence.

Reference DFT+cRPA procedure. To limit the number of calculations, we start with a set of common choices that yield reasonable agreement

Fig. 3 | DMC and downfolded ground states charge densities. $\rho^{\uparrow/\downarrow}(r)$ from each method is normalized by the corresponding $\rho^{\downarrow}(r)$ from the same method.
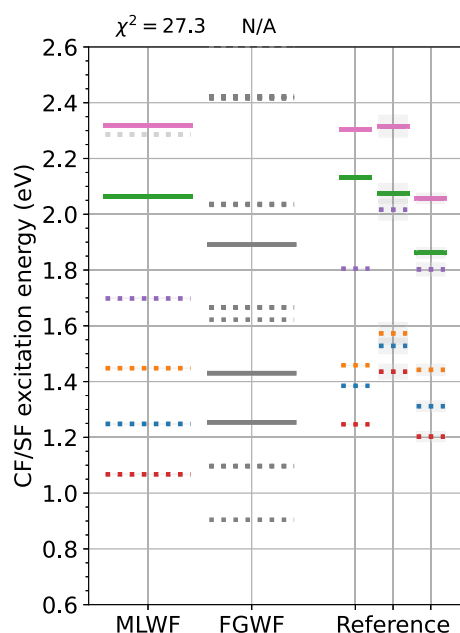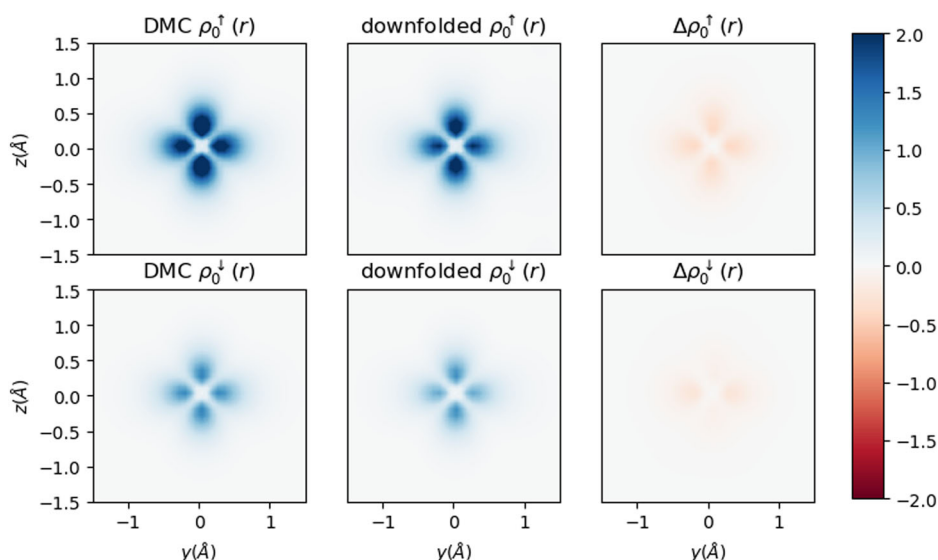
Fig. 4 | Basis set influence on downfolded excitation energies. Comparison between maximally localized and first-guess Wannier function basis sets using static cRPA Coulomb interactions. Dashed and solid lines represent spin-flip (SF) and crystal-field (CF) excitations, respectively. (Light) Gray excitation energies could not be (have not been) characterized.

with the reference data. We use MLWF localized orbitals, single-particle energies with no double-counting corrections, and the full four index cRPA-screened Coulomb matrix elements in the $\omega \to 0$ limit. As shown for the ground state in Fig. 3 and for all excited states in section "Wannier basis sets" this model is in qualitative agreement with the reference data, with a loss function χ^2 of 27.3. We will consider modifications to this strategy, and examine the changes in χ^2; if χ^2 varies upon a change in the choice, then the DFT+cRPA results are sensitive to the choice, and if it does not, then the DFT+cRPA procedure is not sensitive to the choice. In this way, we can identify what approximations deserve further analysis.

Target space basis set. We proceed with the discussion of the chosen basis sets. For comparison to the conventional MLWF, which is

constrained here to reproduce the DFT single-particle energies, we also use a projector-like[49–52] first-guess Wannier function (FGWF) basis set to calculate all model Hamiltonian matrix elements. The FGWF Wannier function basis set differs in its construction by not applying any inner (or "frozen") energy window, such that the DFT single-particle energies are not reproduced, and by using a much larger outer energy window, such that the resulting wave functions are more localized. All details can be found in the section "Methods". In this case, both basis sets are capable of reproducing the correct ground state, so we proceed with the analysis of the charge-neutral excitation spectrum. Figure 4 shows all spin-flip and crystal-field excitations between 0.6 and 2.6 eV. For the FGWF basis, various double and four-fold degenerate spin-flip excitations appear in the energy window of interest, with significantly different characters, such that a 1:1 mapping to the spin-flip excitations from the other methods cannot be established and χ^2 cannot be defined. Correspondingly, the FGWF excitation energies are colored gray in Fig. 4.

This comparison strikingly shows that projector-like first-guess Wannier functions, which are not constrained to reproduce the DFT single-particle energies (see section "Methods" for details), can result in quantitatively and even qualitatively wrong many-body excitation properties.

As shown in Table 1, the single-particle energies in the FGWF basis are not as different from MLWF basis functions as other modifications; however, the interactions as shown in Tables 2 and 3 differ by up to 27%. Thus, at least in this case, the basis functions are important mainly because of their influence on the interactions rather than the single-particle energies.

Single-particle energies. Although it is convenient to extract $\hat{t}$ from the DFT energies, there are several uncontrolled approximations in doing so. First of all, the Kohn-Sham (KS) DFT energies are known to be auxiliary quantities that are not guaranteed to accurately represent the excitations of the system[53,54], even if the exact density-functional were available. Secondly, the DFT calculation already incorporates some electronic interaction effects, while the (ED) solution of the downfolded Hamiltonian will add further interaction effects. Correcting for this double counting could be either achieved by starting from constrained GW calculations[13] or by using a so-called double-counting correction term in the model Hamiltonian, which aims to remove any Coulomb interaction effects from the single-particle starting point. We consider two modifications to the single-particle energies: a Hartree double-counting correction[20,55,56], which decreases the energy difference between a_1 and e_1 by more than 1 eV such that the crystal-field excitations shift significantly up in energy (see Table 1), and rigid shifts of the energy difference

Table 1 | Single-particle energies (in eV) from DFT output (KS), after wannierization (MLWF and FGWF), and including double-counting corrections (MLWF+H)

type	e_2	a_1	e_1
KS cons.	−0.082	0.0	2.310
KS sme.	−0.061	0.0	2.311
MLWF	−0.082	0.0	2.310
FGWF	0.036	0.0	1.986
MLWF + H_{DC}^{Har}	−0.451	0.0	1.140
MLWF + $H_{CFC}^{(a_1)}$	0.061	0.0	2.453
MLWF + $H_{CFC}^{(e_1)}$	−0.082	0.0	2.160

DFT energies are given for constrained and smeared occupations.

Table 2 | Bare and cRPA-screened density-density (v_{iijj}) and Hund's exchange (v_{ijij}) Coulomb matrix elements in the Wannier basis of the MLWF in eV

	v_{iijj}^{bare}			v_{ijij}^{bare}		
	e_2	e_1	a_1	e_2	e_1	a_1
e_2	13.454	12.438	12.985		0.437	0.637
e_1	12.438	13.357	13.135	0.437		0.317
a_1	12.985	13.135	14.466	0.637	0.317	
	U_{iijj}^{cRPA}			U_{ijij}^{cRPA}		
	e_2	e_1	a_1	e_2	e_1	a_1
e_2	5.757	4.998	5.153		0.396	0.570
e_1	4.998	5.900	5.379	0.396		0.302
a_1	5.153	5.379	6.247	0.570	0.302	
	U^{cRPA} monopole screening only					
	e_2	e_1	a_1	e_2	e_1	a_1
e_2	5.838	4.956	5.096		0.436	0.647
e_1	4.956	5.979	5.390	0.436		0.319
a_1	5.096	5.390	6.335	0.647	0.319	

Table 3 | Bare density-density (v_{iijj}) and Hund's exchange (v_{ijij}) Coulomb matrix elements in the Wannier basis of the FGWF in eV

|	v_{iijj}^{bare}			v_{ijij}^{bare}		
	e_2	e_1	a_1	e_2	e_1	a_1
e_2	18.450	15.058	16.733		0.578	0.838
e_1	15.058	14.531	15.314	0.578		0.407
a_1	16.733	15.314	18.127	0.838	0.407	
	U_{iijj}^{cRPA}			U_{ijij}^{cRPA}		
	e_2	e_1	a_1	e_2	e_1	a_1
e_2	6.599	5.349	5.491		0.508	0.307
e_1	5.349	6.146	5.809	0.508		0.379
a_1	5.491	5.809	7.062	0.307	0.379	

Fig. 5 | Single-particle double-counting correction influence on downfolded excitation energies using the MLWF basis and static cRPA Coulomb interactions. Dashed and solid lines represent spin-flip (SF) and crystal-field (CF) excitations, respectively. Light gray excitation energies have not been characterized.

between a_1 and e_1 levels by approx. ± 150 meV. We use the rigid shifts rather than varying the functional because varying the DFT functional also changes orbitals, which could result in conflated effects. For discussions on how to possibly vary the single-particle correction scheme for different DFT functionals, we refer the interested reader to refs. 20,55.

Figure 5 shows how modifications to the single-particle terms affect the excitation energies in the downfolded model. The spin-flip excitations are barely affected by single-particle energy changes, as they are mainly defined by exchange Coulomb matrix elements, which are unaffected by single-particle double-counting corrections. The crystal-field excitations can, however, change significantly: Increasing the energy difference between a_1 and e_1 with $H_{CFC}^{a_1}$ by about 150 meV lowers the crystal-field excitation energies while decreasing the difference between a_1 and e_1 by about 150 meV using $H_{CFC}^{e_1}$ shifts the crystal-field excitations up in energy. This trend also holds using the Hartree double-counting correction, which decreases the energy difference between a_1 and e_1 by more than 1 eV such that the crystal-field excitations shift significantly up in energy (see Table 1 for further details).

In all of these cases, we can exactly map the excited many-body states to the reference ones, allowing us to calculate χ^2. We see that $H_{CFC}^{a_1}$ and $H_{CFC}^{e_1}$ only mildly affect χ^2. The Hartree single-particle double-counting correction, however, significantly increases χ^2, yielding a poor agreement with the reference data. From this, we can conclude that the Hartree double counting is not a favorable correction scheme for charge-neutral local excitations in our vanadocene test system. A similar effect from Hartree double-counting corrections has already been observed for Fe impurities in AlN[20]. In fact, we find the best agreement with the first-principles reference without applying a double-counting correction at all. This is in line with conventional double-counting corrections, as regularly applied in DFT++ schemes[57–60], which do not have an orbital dependence within the correlated space.

The effects of the single-particle energies on the accuracy of the model are relatively straightforward compared to the other terms. As long as the energies are not too much in error, the ground state and spin excitations are robust to their precise value. The crystal-field excitations are shifted roughly linearly with the single-particle energies.

Static screening model. We now investigate how different screening models affect the charge-neutral excitation spectrum. To this end, we show in Fig. 6 the original data as obtained from using the conventional screened cRPA U_{ijkl} tensor (referred to as "full"), together with results obtained from a cRPA tensor which is predominantly screened in the monopole channel (defined here as the leading eigenvalue of the screening, as explained in section "Methods"), and results from using the bare (unscreened) Coulomb interaction tensor. The "monopole" cRPA
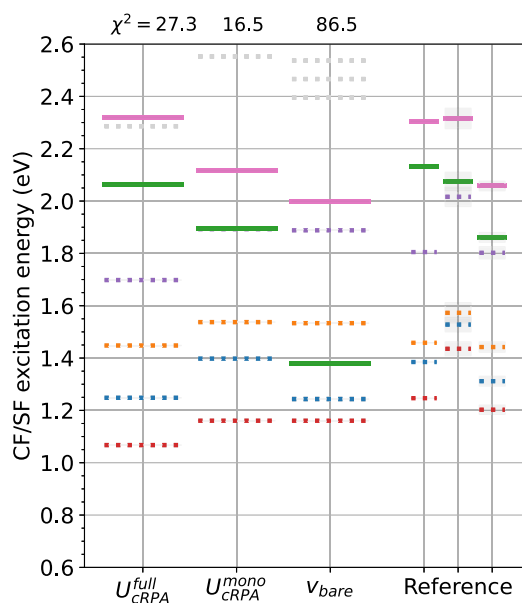
Fig. 6 | Static screening model influence on downfolded excitation energies using the MLWF basis and no double-counting correction. Dashed and solid lines represent spin-flip (SF) and crystal-field (CF) excitations, respectively. Light gray excitation energies have not been characterized.
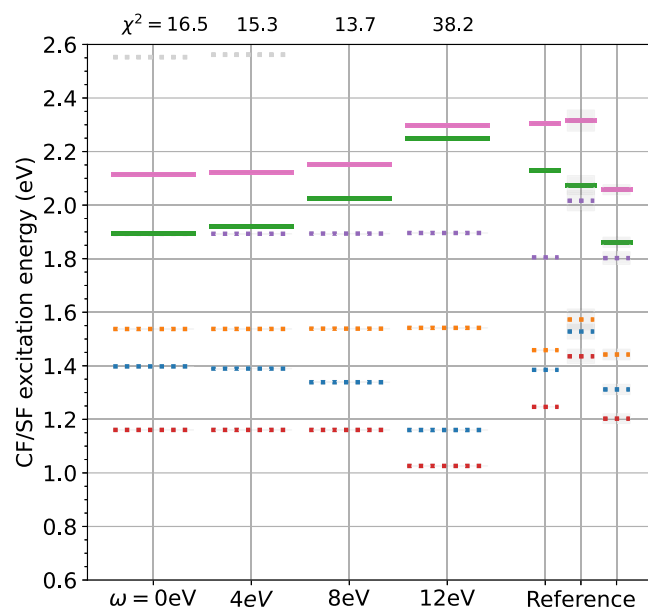
Fig. 7 | cRPA screening frequency influence on downfolded excitation energies within the monopole channel using the MLWF basis and no double-counting correction. Dashed and solid lines represent spin-flip (SF) and crystal-field (CF) excitations, respectively. Light gray excitation energies have not been characterized.

model is inspired by Scott and Booth[61], and further details can be found in the section "Methods".

The full cRPA results that have already been discussed yield a $\chi^2 = 27.3$. Restricting the cRPA screening predominantly to the monopole channel, such that predominately density-density Coulomb matrix elements are screened, yields a better agreement with $\chi^2 = 16.5$. This additional constraint on the cRPA screening shifts the crystal-field excitation energies down by about 200 meV, and clusters the spin-flip excitations into 3:1 pairs, with the highest spin-flip excitations being closer to the lowest crystal-field excitations (in Fig. 6 they actually now overlap). This results in a better agreement with the reference data as quantified by the reduced χ^2 value. When screening is completely neglected, i.e., when we use the bare v_{ijkl} Coulomb tensor, the agreement with the reference data as measured by χ^2 becomes, however, unsatisfactory compared to the (monopole) cRPA cases. In the bare case, the crystal-field excitation energies are strongly underestimated, which mixes the overall ordering of the spin-flip and crystal-field excitations, resulting in a large error, $\chi^2 = 86.5$. The spin-flip excitations, however, also show the 3:1 clustering similar to the monopole screening case.

From this study, we learn that the cRPA screening model is indeed better than bare Coulomb interactions. However, the screening model is not perfectly accurate; a modification of only screening the monopole channel, motivated by the underlying RPA screening by long-range charge fluctuations[61], results in better agreement of the model eigenstates with the quantum chemistry reference results, at least for vanadocene.

Finite frequency screening. The Coulomb tensor in the cRPA approximation is frequency-dependent[11], i.e., $\hat{U}_{cRPA}(\omega)$. To obtain a static Hamiltonian which is easier to solve, it is conventional to evaluate $\hat{U}_{cRPA}(\omega = 0)$, based on the idea that the low-energy target space physics is slow compared to the screening processes. To assess this approximation, we study how the results change when the cRPA Coulomb matrix elements are evaluated at finite frequency. Here, we perform this analysis based on a simple model of the frequency dependence, the plasmon pole model (see Methods for details).

Figure 7 shows the excitation energies using a Coulomb tensor $\hat{U}(\omega)$ evaluated at various ω, which are all below the estimated background screening plasmon pole frequency of approximately 16.5 eV. Note that these are all calculated using the monopole screening model of Fig. 6, which

corresponds to $\omega = 0$ here. For $\omega = 4$ eV and $\omega = 8$ eV, we find an increase in the crystal-field excitation energies and a concomitant small improvement in the match with the reference data as quantified by χ^2. The spin excitation energies barely change for these ω values. $\omega = 12$ eV is close to the plasmon pole frequency of 16.5 eV, so the Coulomb matrix elements change more rapidly, and the excitation energies change more as a result. At this point, the agreement with the reference data deteriorates substantially. We should note that the theoretical justification of cRPA relies on the screening frequency ω being smaller than the gap in the rest space[38], which is roughly 6 eV in vanadocene (Fig. 1). I.e., for $\omega > 6$ eV, the dynamics of particle-hole excitations start to become relevant, which we cannot capture in the static Hamiltonian formalism.

With regard to the frequency choice in the cRPA background screening, we thus find slightly better agreement with the reference data upon using $\omega > 0$ in comparison to $\omega = 0$. Among all benchmarked quantities, however, the background screening frequency has the smallest effect.

Discussion

By comparing the ground state and charge-neutral excited states obtained directly from first-principles descriptions and from minimal downfolded Hamiltonians, we have provided a systematic and quantitative sensitivity analysis of the DFT+cRPA scheme for the ab initio derivation of interacting model Hamiltonians for the description of correlation effects in the vanadocene molecule. We expect that the results in this study will be most applicable to similar systems, such as defects in gapped semiconductors or correlated bulk materials with well-separated correlated sub-spaces. These systems are ideal for cRPA, as noted by some of us[38], and even so, we were able to quantify errors in every step of the DFT+cRPA process.

To summarize, we find the following about the DFT+cRPA process: Orbital shapes are crucial. The single-particle energies for FGWF are not that different from MLWF but the model is extremely inaccurate because the two-body terms highly depend on the orbital shapes. Naïve application of the orbital-dependent double-counting corrections results in very poor estimations of crystal-field excitations, because it fails to describe the one-particle levels. Small corrections do not change much within our uncertainties. The cRPA-screened interactions are much better than the bare ones. Monopole-only screening is clearly better than traditional. Frequency dependence, in this case, is not that big, unless one goes to very high energies.
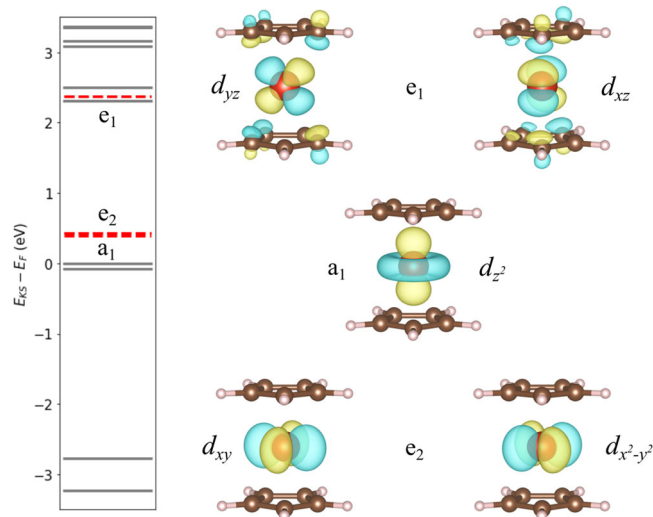
Fig. 8 | Single-particle properties. Left: DFT Kohn-Sham energies. Red lines mark single-particle energies resulting from diagonalization of the downfolded Hamiltonian using first-guess Wannier functions. Right: First-guess Wannier functions used within the downfolding procedure.

From this data, we can make a few recommendations. One should be very careful about the basis orbitals, especially in cases where disentanglement of bands is necessary since orbital shapes can completely change the results. Similarly, the function used to choose localized orbitals might have significant effects on the interactions computed. Double-counting corrections do not necessarily improve the accuracy of the Hamiltonian. These "corrections" can also be conflated with the tendency of DFT to incorrectly compute the one-particle energies. Even in the case of vanadocene, which is a best-case scenario for cRPA, a naïve application of cRPA screening obtains interactions that are better than bare, but appear to be limiting accuracy. In this case, they can be improved by monopole-only screening in particular, although it remains to be seen if such a method is generally more accurate.

Looking forward, we have demonstrated that it is possible to generate reference quantum mechanical data of sufficient quality and quantity to systematically analyze heuristic (but inexpensive) methods like DFT +cRPA, at least for model systems such as vanadocene. The application of multiple first-principles methods allows us to evaluate the uncertainty in the reference data, partially mitigating the problem of overfitting. The data from this work can serve as a reference for other methods, such as *GW*-based approaches, and could be applied to a variety of materials and systems and has been made publicly available[62].

Methods
DFT calculations
We use Quantum Espresso (QE)[63–65] and embed the vanadocene molecule in a $15 \times 15 \times 15$ Å^3 supercell to minimize spurious wave function overlap and undesired screening between repeated cells. The plane-wave cut-off was set to 440 Ryd, and we applied the spin-restricted (spin-unpolarized) generalized gradient approximation within a PBE functional[66]. In these spin-restricted calculations, we find an ordering of $e_2 < a_1 < e_1$, which is different from spin-unrestricted results for both C_{2h} and D_{5h} VCp$_2$ where the $e_2 < a_1$ ordering is reversed between the spin channels[45,67]. The occupations of the three lowest d orbitals are constrained to exactly 1 to avoid partial state occupations due to the conventional smearing methods in plane-wave DFT codes. This constraint does, however, not significantly affect the Kohn-Sham energies compared to calculations with a large smearing, c.f. Table 1.

Wannier basis sets
We use the three highest (partially) occupied and the two lowest unoccupied molecular orbitals with predominant V d-orbital character to construct five

maximally localized Wannier functions (MLWF) $|\phi_i\rangle$ utilizing RESPACK[68]. For the construction of these MLWFs, we use an inner (frozen) window to constrain the energies of the five Kohn-Sham states of interest. The resulting Wannier orbitals resemble the expected V d orbital symmetries, however, with small lobes at the carbon rings as visualized in Fig. 1, which is in good agreement with ref. 67. These indicate a finite hybridization between the V d states and the carbon rings.

To investigate this specific choice of Wannier functions and to qualitatively compare to projector-based methods, we construct a second Wannier basis without using the frozen (inner) window, utilizing a larger Wannierization (outer) window, and without performing maximal localizations. The resulting first-guess Wannier function (FGWF) single-particle energies do not perfectly reproduce the KS ones, as indicated in Fig. 8 and listed in Table 1, and cannot be generated via some uniform rotation of the MLWFs. These FGWF orbitals are nevertheless similar to the MLWFs, but do have a slightly different shape, which is reflected in the overlap matrix elements between Wannier orbitals of the same symmetry:

$$\langle \phi_{e_2}^{\text{MLWF}} | \phi_{e_2}^{\text{FGWF}} \rangle \approx 0.936,$$

$$\langle \phi_{a_1}^{\text{MLWF}} | \phi_{a_1}^{\text{FGWF}} \rangle \approx 0.961,$$

$$\langle \phi_{e_1}^{\text{MLWF}} | \phi_{e_1}^{\text{FGWF}} \rangle \approx 0.987.$$

Model Hamiltonian matrix elements and cRPA screening
With these Wannier basis sets, we evaluate the hopping matrix elements $t_{ij} = \langle \phi_i | H_{\text{DFT}} | \phi_j \rangle$ and the bare $v_{ijkl} = \langle \phi_i \phi_j | v | \phi_k \phi_l \rangle$ as well as the statically cRPA-screened $U_{ijkl} = \langle \phi_i \phi_j | U(\omega = 0) | \phi_k \phi_l \rangle$ Coulomb matrix elements within RESPACK[68]. The resulting single-particle energies are given in Table 1, and the density-density and Hund's exchange matrix elements are listed in Tables 2 and 3. We see that density-density elements are screened by about 55 to 60% (up to 7.8 eV), while Hund's elements are reduced by only 5 to 10% (up to 60 meV). Analysing the leading and sub-leading screening contributions yields a screening of $\varepsilon_1 = \varepsilon_{\text{mono}} \approx 2.44$ in the predominate monopole channel and $\varepsilon_{i>1} = \varepsilon_{\text{multi}} \approx 1.04$ to 1.21 in the sub-leading multipole channels. These screening channels are defined by $\varepsilon_i = v_i/U_i$ with v_i being the eigenvalues of the full v_{ijkl} tensor evaluated in a product basis. U_i is the approximated eigenvalues of U_{ijkl} obtained using the eigenbasis of the bare v_{ijkl}[69]. The full cRPA screening is thus most dominant in the monopole channel, which mostly affects density-density interactions, while the screening becomes small in the multipole channels, which mostly affect Hund's exchange elements.

Inspired by Scott and Booth[61] we also construct a cRPA Coulomb tensor, which is screened mostly in the leading monopole channel, such that only $\varepsilon_1 > 1$, while all other $\varepsilon_{i>1} = 1$. The resulting matrix elements are listed in Table 2. As expected, the approx. "monopole" cRPA Coulomb matrix elements agree with the full cRPA ones in the density-density and approx. with the bare ones in the Hund's exchange channels. However, note that the monopole-screened Hund's exchange elements are up to 77 meV larger than the fully screened ones, which is relevant for the spin-flip excitation energies.

For the analysis of finite background screening frequencies, we calculate $U_{ijkl}(i\omega_n) = \langle \phi_i \phi_j | U(i\omega_n) | \phi_k \phi_l \rangle$ for a few Matsubara frequencies $i\omega_n$. Afterwards we fit the leading eigenvalue to a single plasmon pole model of the form $U_1(i\omega_n) = v_1 - g^2 \frac{2\omega_p}{\omega_n^2 + \omega_p^2}$ yielding $g \approx 17.6$ eV and $\omega_p \approx 16.5$ eV. This allows us to adjust the real frequency of the leading eigenvalue as $U_1(\omega) \approx v_1 + g^2 \frac{2\omega_p}{\omega^2 - \omega_p^2 + i\delta}$ for which we use $\delta = 2.5$ eV. This way, we can continuously adjust the leading eigenvalue in the monopole screening model as done for Fig. 7. The rather large $\delta = 2.5$ eV is used here to mimic the effects of screening channels beyond the single plasmon pole model. The dependence of the density-density monopole-screened Coulomb interaction matrix elements is depicted in Fig. 9.

Although the FGWFs show deviations of just up to 6.4% in comparison to the MLWF ones (as measured by the overlaps given above), the resulting bare Coulomb matrix elements are different by up to 27% as a result of the ϕ_i^4 dependence of v and as visible from Table 3. This is a result of the enhanced atomic character of the FGWFs and their reduced hybridization with the

carbon rings, yielding a smaller spread and, thus, enhanced bare Coulomb matrix elements.

Double counting

We choose three orbital-dependent correction schemes, which all act on the single-particle t_{ij} hopping matrix elements according to the following terms, which we add to the model Hamiltonian:

$$H_{\text{DC}}^{\text{Har}} = \sum_{ij,\sigma} c_{i\sigma}^\dagger c_{j\sigma} \sum_{kl} \rho_{kl} U_{iljk}, \quad (2)$$

with ρ_{kl} the single-particle density-matrix in the Wannier basis obtained from the DFT calculation. This is the conventional Hartree double-counting correction as discussed in detail in refs. 20,55,56. We also apply ad hoc chosen crystal-field corrections of the form

$$H_{\text{CFC}}^{(a_1/e_1)} = \sum_{i,\sigma} \Delta_i^{(a_1/e_1)} c_{i\sigma}^\dagger c_{i\sigma} \quad (3)$$

with only $\Delta_{i=a_1}^{(a_1)} = -0.143$ eV or $\Delta_{i=e_1}^{(e_1)} = -0.150$ eV non-zero elements for the respective H_{CFC}. These corrections thus solely shift the a_1 or e_1 states, while they preserve all symmetries. With these, we aim to investigate the effects of possibly wrong single-particle energies.

Exact Solutions and Benchmark Quantities

We use the exact diagonalization (ED) routines (atom_diag) from TRIQS[70] to solve the downfolded many-body Hamiltonian $H = \hat{t}_{\alpha\beta} c_\alpha^\dagger c_\beta + \hat{U}_{\alpha\beta\gamma\delta} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta$, giving us access to both the energies and the many-body wave functions in the target space. Here, we restrict ourselves to charge-neutral
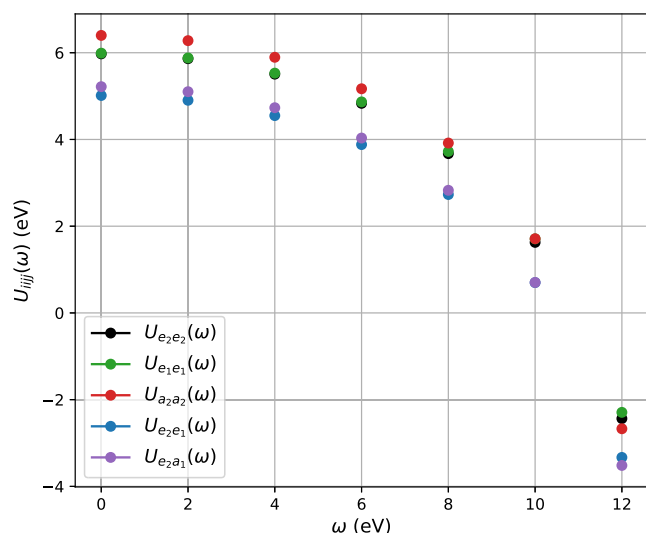
Fig. 9 | Approximated frequency dependence of the monopole-screened cRPA Coulomb matrix elements in the density-density channel.
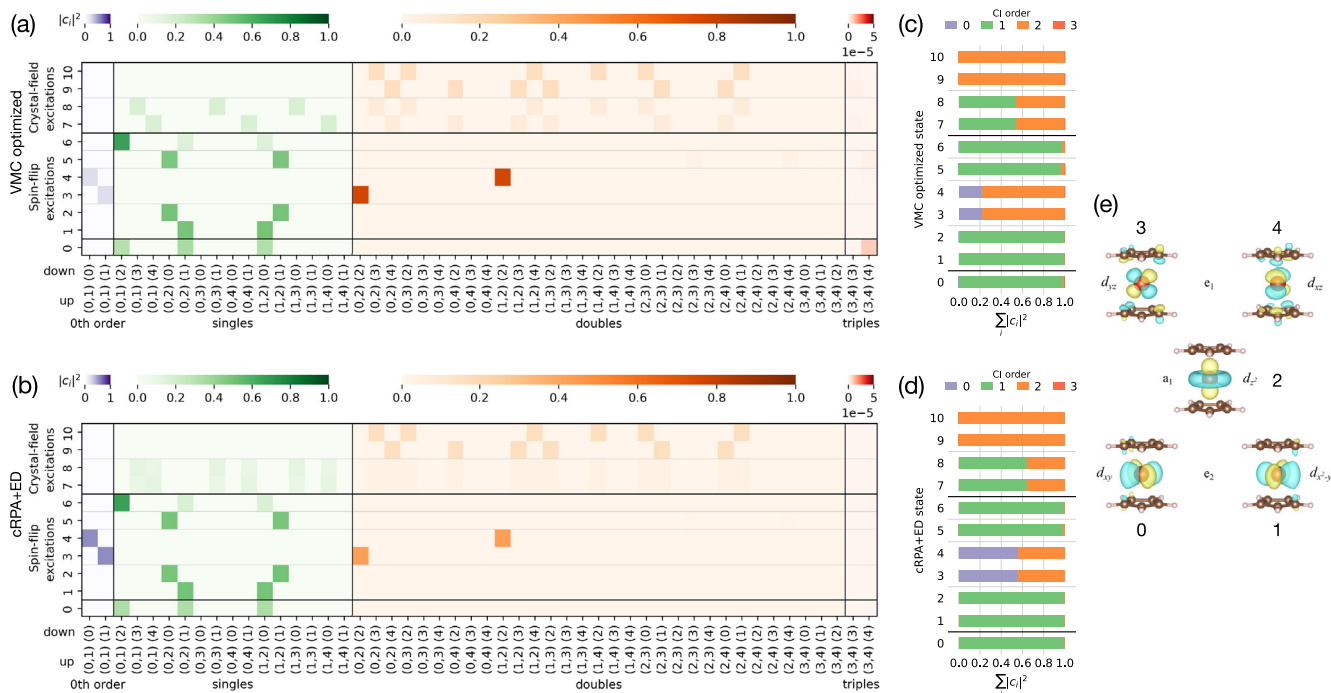
Fig. 10 | Comparisons of many-body wave functions from quantum chemistry and downfolded model. a, b The squares of configuration interaction (exact diagonalization) coefficients ($|c_i|^2$) of the optimized multi-Slater-Jastrow wave functions obtained using variational Monte Carlo and eigenstates obtained from ED using the model given by the MLWFs. The four panels correspond to the 0th-order configuration interaction (purple), single excitations (green), double excitations (orange), and triple excitations (red) within the CAS(5, (2,1)) space. The tick labels on *the x*-axis, (i, j) (k), represent a Slater determinant $c_{k\downarrow}^\dagger c_{j\uparrow}^\dagger c_{i\uparrow}^\dagger |\Phi_0\rangle$ (see panel **e** for the MLWFs), where $|\Phi_0\rangle$ is the Slater determinant of all the valence electrons doubly

occupying the orbitals up to the V d shell. Panels **c, d** show the cumulated contributions in each order of CIs for the VMC wave functions and the model eigenstates. The black horizontal lines group the many-body states into ground state, spin-flip excitations, and crystal-field excitation, whereas the gray horizontal lines group the degenerate states together. We can see that states 3 and 4 (degenerate in energy and correspond to the 2nd excited state, labeled by blue dashed lines in Fig. 2) and all the crystal-field excitations show significant contributions from the double excitations.

excitations, i.e., those with $N = 3$ electrons in total, such that the relevant many-body wave functions are of the form $|\Psi\rangle = \sum_{\alpha\beta\gamma} A_{\alpha\beta\gamma} c_\alpha^\dagger c_\beta^\dagger c_\gamma^\dagger |0\rangle$, where the sum runs over spin and orbital indices. For three electrons, the possible values of S_z are $\pm 3/2$ and $\pm 1/2$. Note that our model has $SU(2)$ spin symmetry, so the spin S is a good quantum number, and states with the same S and different S_z have the same energy. We can thus restrict ourselves to states with $S_z = 1/2$.

Using the Wannier functions $\phi_\alpha(\mathbf{r})$, we calculate the electron density $\rho_i(\mathbf{r})$ in real-space for a given many-body wave function i as $\rho_i(\mathbf{r}) = \sum_{i,\alpha} \langle \Psi_i | c_\alpha^\dagger c_\alpha | \Psi_i \rangle |\phi_\alpha(\mathbf{r})|^2$ using the many-body states Ψ_i as obtained from the exact diagonalization of the Hamiltonian. For the spin-resolved density $\rho_i^\sigma(\mathbf{r})$ we use a similar expression yielding $\int d\mathbf{r} \rho_i^\uparrow(\mathbf{r}) = 2$ and $\int d\mathbf{r} \rho_i^\downarrow(\mathbf{r}) = 1$ in the $S_z = 1/2$ channel. In addition to this, we also calculate the 1-RDMs $(\rho_i)_{\alpha\beta} = \langle \Psi_i | c_\alpha^\dagger c_\beta | \Psi_i \rangle$ for each wave function Ψ_i.

Based on these quantities, that is, the many-body state energies, wave functions, 1-RDMs, and charge densities, we can unambiguously identify all states among all applied methods. A corresponding comparison between the many-body wave functions represented as linear combinations of Slater determinants from DFT+cRPA and the multi-Slater part of the trial wave functions for DMC is depicted in Fig. 10. The corresponding comparisons for the 1-RDMs and the many-body charge densities are given in the Supplemental Notes 1 and 2.

Real-space fixed-node Diffusion Monte Carlo (DMC)
In order to obtain approximate many-body eigenstates using fixed-node diffusion Monte Carlo (DMC), we constructed the trial wave functions in the following multi-Slater-Jastrow form,

$$\Psi_T(\mathbf{R}) = \Psi_T(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$$
$$= e^{J(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)} \sum_k c_k D_k^\uparrow \left(\chi_{i\uparrow}(\mathbf{r}_j) \right) D_k^\downarrow \left(\chi_{i\downarrow}(\mathbf{r}_j) \right), \quad (4)$$

where $\mathbf{r}_i$ represents the real-space coordinates of the ith electron and $\mathbf{R}$ encompasses the configurations of all the electrons. The multi-Slater-determinant expansion, given by $\sum_k c_k D_k^\uparrow(\chi_{i\uparrow}(\mathbf{r}_j)) D_k^\downarrow(\chi_{i\downarrow}(\mathbf{r}_j))$, was generated from restricted Hartree-Fock (RHF) followed by state-averaged complete active space self-consistent field (CASSCF)[71] using the PySCF package[72] with correlation-consistent effective core potentials, and a corresponding triple-zeta basis set[39]. The system consists of 63 valence electrons, and in CASSCF, we chose an active space that includes two spin-up electrons, one spin-down electron, and five molecular orbitals that are V-3d-like determined by the atomic valence active space (AVAS) procedure[73]. For all the eigenstates, the parameters in the two-body Jastrow factor J, all the determinant coefficients c_k, and the molecular orbitals in the active space were fully optimized using variational Monte Carlo[30] using the method described in refs. [74,75]. The excited states were optimized with the constraint that they are orthogonal to the optimized lower-energy eigenstates[32]. After the trial wave functions Ψ_T (of either the ground state or the excited states) were fully optimized, the lowest-energy wave functions that have the same nodal structure were projected out using DMC[30]. We then compute the energies and reduced density matrices of the fixed-node wave functions of all the eigenstates of interest. All the variational and diffusion Monte Carlo calculations, including the optimization of trial wave functions and DMC, were performed using the PyQMC package[76] interfaced with PySCF.

Using reduced density matrices in natural orbitals to characterize many-body eigenstates.
For a wave function $|\Psi\rangle$, its one-body and two-body reduced density matrices (1(2)-RDMs) in a given single-particle basis are defined as

$$\rho_{ij}^\sigma[\Psi] = \left\langle \Psi | c_{i\sigma}^\dagger c_{j\sigma} | \Psi \right\rangle,$$
$$\rho_{ijkl}^{\sigma\sigma'}[\Psi] = \left\langle \Psi | c_{i\sigma}^\dagger c_{k\sigma'}^\dagger c_{l\sigma'} c_{j\sigma} | \Psi \right\rangle, \quad (5)$$

where $c_{i\sigma}^\dagger$ ($c_{i\sigma}$) creates (annihilates) an electron in single-particle orbital $\chi_{i\sigma}$. Almost all the one- and two-body observables can be computed using the 1- and 2-RDMs. For example, the spin-resolved electronic density is evaluated as

$$\rho^\sigma(\mathbf{r}) = \sum_i \chi_{i\sigma}^*(\mathbf{r}) \rho_{ii}^\sigma[\Psi] \chi_{i\sigma}(\mathbf{r}). \quad (6)$$

To directly compare the observables that characterize the fixed-node eigenstates with those of the downfolded eigenstates, it is crucial to ensure that the RDMs are calculated using an identical set of single-particle orbitals. Given that the downfolding and DMC calculations were performed using separated packages and employed a different basis to expand the active space (MLWFs for the downfolded results and CASSCF molecular orbitals for DMC), we rotate the RDMs obtained from each method to the same basis set to align the many-body eigenstates.

The natural orbitals are defined as the principal components, that correspond to the largest five singular values, of all the 1-RDMs up to the mth downfolded excited state[77],

$$\varrho = U\Sigma V,$$
$$\varrho = \left(\rho^\uparrow[\Psi_0], \cdots, \rho^\uparrow[\Psi_m], \rho^\downarrow[\Psi_0], \cdots, \rho^\downarrow[\Psi_m] \right), \quad (7)$$

where $\rho^\sigma[\Psi_a]$ is eigenstate Ψ_a's 1-RDM written in the Wannier basis $\{\phi_i(\mathbf{r})\}$. We then transform all the 1- and 2-RDMs computed from both downfolding and DMC to these natural orbitals. This ensures that we characterize downfolded and DMC eigenstates within the same Hilbert space. We found $\mathrm{Tr}(\rho_{\mathrm{downfolded}}[\Psi_i]) - \mathrm{Tr}(\rho_{\mathrm{DMC}}[\Psi_i])$ to be 0.1 ~ 0.2, for all the Ψ_i's of interest (up to the 2nd charge excitation). With three electrons in the active space, this suggests that when using the MLWFs derived from Wannierization of Kohn-Sham orbitals to characterize the low-energy excitations in VCp$_2$, we get about 5% error in terms of their 1-RDMs.

Auxiliary-field quantum Monte Carlo (AFQMC)
We computed the low-energy many-body spectrum, including spin-flip, and crystal-field excitations, of the vanadocene molecule using phaseless AFQMC[29,78,79]. All AFQMC calculations were performed using the Flatiron Institute, Center for Computational Quantum Physics's production quality AFQMC code. A detailed description of the AFQMC method was recently reviewed in great detail[78]. AFQMC was recently used to compute the ionization potentials of several 3d-transition metal complexes, including vanadocene, producing results that compare favorably with experiment[33]. Our AFQMC procedure, especially regarding the choice of trial wave function, is guided by that benchmark. AFQMC calculations were performed using multi-Slater determinant trial wave functions which were derived from CASSCF calculations. The active space for CASSCF was chosen using the atomic valence active space (AVAS) procedure[73] and the ANO-RCC basis to define V-3d reference atomic orbitals, with an AVAS threshold of 0.1. This yields an active space consisting of 13 active electrons and 15 active orbitals. CASSCF orbitals were optimized for the ground state, and excited-state CAS-wave functions were computed in one-shot state-specific CASCI calculations using the ground-state CASSCF orbitals to define the active space. AFQMC trial wave functions are truncated CAS-wave functions where Slater determinants with weight less than 0.0014 are discarded. For states with $S = 3/2$, including the ground state and both crystal-field excitations, the trial wave functions consisted of only about 100 Slater determinants. For all spin-flip excitations, the trial wave functions consisted of about 150 Slater determinants.

Equation-of-motion coupled cluster (EOMCC)
Equation-of-motion coupled-cluster calculations were performed at the single-, and double-excitation level (EOM-CCSD), as implemented in PySCF[72]. The correlation-consistent effective core potential (ccECP)/pseudopotential and the corresponding ccECP-cc-pVQZ basis were used for all atoms. The ROHF ground state determinant, with $S_z = 3/2$ and

$S^2 = 3.75$, was used as a reference for EOM-CCSD. Due to the very large dimension of the Hilbert space in the ccECP-cc-pVQZ basis, we performed EOM-CCSD in a CAS-like active space with 502 active orbitals and including all electrons in the active space. We checked that the excitation energies are converged by comparing them with similar calculations performed in an active space consisting of 452 orbitals. The average absolute deviation of the excitation energies, computed over the states in Fig. 2, between EOM-CCSD(502o,63e) and EOM-CCSD(452o,63e) is 28(15) meV with a maximum absolute deviation of 53 meV from the first crystal-field excitation.

Data availability
The data that support the findings of this work are available from the corresponding author upon reasonable request. All ab initio and modeling results (AFQMC, DMC, EOM-CC, DFT, cRPA, and ED) are furthermore publicly available[62].

Code availability
All non-commercial numerical codes to reproduce the findings of this study are available from the corresponding author upon reasonable request.

References
1. Aryasetiawan, F. & Nilsson, F. *Downfolding Methods in Many-Electron Theory* (AIP Publishing LLC, 2022). https://aip.scitation.org/doi/10.1063/9780735422490.
2. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
3. Rubtsov, A. N., Katsnelson, M. I. & Lichtenstein, A. I. Dual fermion approach to nonlocal correlations in the Hubbard model. *Phys. Rev. B* **77**, 033101 (2008).
4. Rubtsov, A., Katsnelson, M. & Lichtenstein, A. Dual boson approach to collective excitations in correlated fermionic systems. *Ann. Phys. (N.Y.)* **327**, 1320–1335 (2012).
5. Galler, A. et al. The AbinitioDΓA Project v1.0: Non-local correlations beyond and susceptibilities within dynamical mean-field theory. *Comput. Phys. Commun.* **245**, 106847 (2019).
6. Kananenka, A. A., Gull, E. & Zgid, D. Systematically improvable multiscale solver for correlated electron systems. *Phys. Rev. B* **91**, 121111 (2015).
7. Lee, H. et al. Electron-phonon physics from first principles using the EPW code. *npj Comput. Mater* **9**, 156 (2023).
8. Falter, C. & Selmke, M. Renormalization of the dielectric response with applications to effective ion interactions and phonons. *Phys. Rev. B* **24**, 586–594 (1981).
9. Falter, C. A unifying approach to lattice dynamical and electronic properties of solids. *Phys. Rep.* **164**, 1–117 (1988).
10. Nomura, Y. & Arita, R. Ab initio downfolding for electron-phonon-coupled systems: Constrained density-functional perturbation theory. *Phys. Rev. B* **92**, 245108 (2015).
11. Aryasetiawan, F. et al. Frequency-dependent local interactions and low-energy effective models from electronic structure calculations. *Phys. Rev. B* **70**, 195104 (2004).
12. McMahan, A. K., Martin, R. M. & Satpathy, S. Calculated effective Hamiltonian for La_2CuO_4 and solution in the impurity Anderson approximation. *Phys. Rev. B* **38**, 6650–6666 (1988).
13. Hirayama, M., Miyake, T. & Imada, M. Derivation of static low-energy effective models by an ab initio downfolding method without double counting of Coulomb correlations: Application to $SrVO_3$, FeSe, and FeTe. *Phys. Rev. B* **87**, 195144 (2013).
14. Honerkamp, C. Effective interactions in multiband systems from constrained summations. *Phys. Rev. B* **85**, 195129 (2012).
15. Bartlett, R. J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **79**, 291–352 (2007).
16. Lyakh, D. I., Musiał, M., Lotrich, V. F. & Bartlett, R. J. Multireference nature of chemistry: the coupled-cluster view. *Chem. Rev.* **112**, 182–243 (2012).
17. Kowalski, K. & Bauman, N. P. Quantum flow algorithms for simulating many-body systems on quantum computers. *Phys. Rev. Lett.* **131**, 200601 (2023).
18. Lichtenstein, A. I. & Katsnelson, M. I. Ab initio calculations of quasiparticle band structure in correlated systems: LDA++ approach. *Phys. Rev. B* **57**, 6884–6895 (1998).
19. Anisimov, V. I., Aryasetiawan, F. & Lichtenstein, A. I. First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+U method. *J. Phys. Condens. Matter* **9**, 767 (1997).
20. Muechler, L. et al. Quantum embedding methods for correlated excited states of point defects: Case studies and challenges. *Phys. Rev. B* **105**, 235104 (2022).
21. Sheng, N., Vorwerk, C., Govoni, M. & Galli, G. Green's function formulation of quantum defect embedding theory. *J. Chem. Theory Comput.* **18**, 3512–3522 (2022).
22. Haule, K. Exact double counting in combining the dynamical mean field theory and the density functional theory. *Phys. Rev. Lett.* **115**, 196403 (2015).
23. Kristanovski, O., Shick, A. B., Lechermann, F. & Lichtenstein, A. I. Role of nonspherical double counting in DFT+DMFT: Total energy and structural optimization of pnictide superconductors. *Phys. Rev. B* **97**, 201116 (2018).
24. Kotliar, G. et al. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **78**, 865–951 (2006).
25. Dederichs, P. H., Blügel, S., Zeller, R. & Akai, H. Ground states of constrained systems: application to cerium impurities. *Phys. Rev. Lett.* **53**, 2512–2515 (1984).
26. Zhang, Z. & Satpathy, S. Electron states, magnetism, and the Verwey transition in magnetite. *Phys. Rev. B* **44**, 13319–13331 (1991).
27. Koch, H. & Jo/rgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **93**, 3333–3344 (1990).
28. Stanton, J. F. & Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **98**, 7029–7039 (1993).
29. Zhang, S. & Krakauer, H. Quantum Monte Carlo method using phase-free random walks with slater determinants. *Phys. Rev. Lett.* **90**, 136401 (2003).
30. Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
31. Balabanov, N. B. & Boggs, J. E. Ab initio study of molecular and electronic structures of early transition metal trihydrides MH_3 (M = Sc, Ti, V, Fe). *J. Phys. Chem. A* **104**, 1597–1601 (2000).
32. Pathak, S., Busemeyer, B., Rodrigues, J. N. B. & Wagner, L. K. Excited states in variational Monte Carlo using a penalty method. *J. Chem. Phys.* **154**, 034101 (2021).
33. Rudshteyn, B. et al. Calculation of metallocene ionization potentials via auxiliary field quantum Monte Carlo: toward benchmark quantum chemistry for transition metals. *J. Chem. Theory Comput.* **18**, 2845–2862 (2022).
34. Shee, J., Weber, J. L., Reichman, D. R., Friesner, R. A. & Zhang, S. On the potentially transformative role of auxiliary-field quantum Monte Carlo in quantum chemistry: a highly accurate method for transition metals and beyond. *J. Chem. Phys.* **158**, 140901 (2023).
35. Prins, R. & Voorst, J. D. W. V. Bonding in sandwich compounds. *J. Chem. Phys.* **49**, 4665–4673 (1968).
36. Xu, Z.-F., Xie, Y., Feng, W.-L. & Schaefer, H. F. Systematic investigation of electronic and molecular structures for the first

transition metal series metallocenes M(C_5H_5)$_2$ (M = V, Cr, Mn, Fe, Co, and Ni). *J. Phys. Chem. A* **107**, 2716–2729 (2003).

37. Gard, E., Haaland, A., Novak, D. P. & Seip, R. The molecular structures of dicyclopentadienylvanadium, (C_5H_5)$_2$V, and dicyclopentadienylchromium, (C_5H_5)$_2$Cr, determined by gas phase electron diffraction. *J. Organomet. Chem* **88**, 181–189 (1975).

38. van Loon, E. G. C. P., Rösner, M., Katsnelson, M. I. & Wehling, T. O. Random phase approximation for gapped systems: Role of vertex corrections and applicability of the constrained random phase approximation. *Phys. Rev. B* **104**, 045134 (2021).

39. Annaberdiyev, A. et al. A new generation of effective core potentials from correlated calculations: 3d transition metal series. *J. Chem. Phys.* **149**, 134108 (2018).

40. Aryasetiawan, F., Karlsson, K., Jepsen, O. & Schönberger, U. Calculations of Hubbard *U* from first-principles. *Phys. Rev. B* **74**, 125106 (2006).

41. Nomura, Y. et al. Effective on-site interaction for dynamical mean-field theory. *Phys. Rev. B* **86**, 085117 (2012).

42. Martins, C., Aichhorn, M., Vaugier, L. & Biermann, S. Reduced effective spin-orbital degeneracy and spin-orbital ordering in paramagnetic transition-metal oxides: Sr_2IrO_4 versus Sr_2RhO_4. *Phys. Rev. Lett.* **107**, 266404 (2011).

43. Werner, P., Sakuma, R., Nilsson, F. & Aryasetiawan, F. Dynamical screening in La_2CuO_4. *Phys. Rev. B* **91**, 125142 (2015).

44. Prins, R. & Voorst, J. D. Wv Bonding in sandwich compounds. *J. Chem. Phys.* **49**, 4665–4673 (1968).

45. Jackson, T. A. et al. Vanadocene de novo: spectroscopic and computational analysis of Bis(η^5-cyclopentadienyl)vanadium(II). *Organometallics* **31**, 8265–8274 (2012).

46. Phung, Q. M., Vancoillie, S. & Pierloot, K. A multiconfigurational perturbation theory and density functional theory study on the heterolytic dissociation enthalpy of first-row metallocenes. *J. Chem. Theory Comput.* **8**, 883–892 (2012).

47. Nain, S., Khurana, R. & Ali, M. E. Harnessing colossal magnetic anisotropy in sandwiched 3d^2-metallocenes. *J. Phys. Chem. A* **126**, 2811–2817 (2022).

48. Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847–12865 (1997).

49. Amadon, B. et al. Plane-wave based electronic structure calculations for correlated materials using dynamical mean-field theory and projected local orbitals. *Phys. Rev. B* **77**, 205112 (2008).

50. Haule, K., Yee, C.-H. & Kim, K. Dynamical mean-field theory within the full-potential methods: Electronic structure of CeIrIn$_5$, CeCoIn$_5$, and CeRhIn$_5$. *Phys. Rev. B* **81**, 195107 (2010).

51. Karolak, M., Wehling, T. O., Lechermann, F. & Lichtenstein, A. I. General DFT++ method implemented with projector augmented waves: electronic structure of $SrVO_3$ and the mott transition in $Ca_{2-x}Sr_xRuO_4$. *J. Phys. Condens. Matter* **23**, 085601 (2011).

52. Schüler, M. et al. Charge self-consistent many-body corrections using optimized projected localized orbitals. *J. Phys. Condens. Matter* **30**, 475901 (2018).

53. Perdew, J. P. & Levy, M. Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities. *Phys. Rev. Lett.* **51**, 1884–1887 (1983).

54. Sham, L. J. & Schlüter, M. Density-functional theory of the band gap. *Phys. Rev. B* **32**, 3883–3889 (1985).

55. Bockstedte, M., Schütz, F., Garratt, T., Ivády, V. & Gali, A. Ab initio description of highly correlated states in defects for realizing quantum bits. *npj Quant. Mater.* **3**, 1–6 (2018).

56. Ma, H., Govoni, M. & Galli, G. Quantum simulations of materials on near-term quantum computers. *npj Comput. Mater.* **6**, 1–8 (2020).

57. Solovyev, I. V., Dederichs, P. H. & Anisimov, V. I. Corrected atomic limit in the local-density approximation and the electronic structure of *d* impurities in Rb. *Phys. Rev. B* **50**, 16861–16871 (1994).

58. Czyżyk, M. T. & Sawatzky, G. A. Local-density functional and on-site correlations: The electronic structure of La_2CuO_4 and $LaCuO_3$. *Phys. Rev. B* **49**, 14211–14228 (1994).

59. Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard *U* instead of Stoner *I. Phys. Rev. B* **44**, 943–954 (1991).

60. Liechtenstein, A. I., Anisimov, V. I. & Zaanen, J. Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators. *Phys. Rev. B* **52**, R5467–R5470 (1995).

61. Scott, C. J. C. & Booth, G. H. Rigorous screened interactions for realistic correlated electron systems *Phys. Rev. Lett.* **132**, 076401(2023).

62. Chang, Y., van Loon, E. G. C. P., Eskridge, B. & Rösner, M. DFT+cRPA benchmarking in vanadocene molecule, GitHub repository. https://github.com/YueqingChang/Downloading_benchmark_vanadocene/tree/main (2024).

63. Giannozzi, P. et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).

64. Giannozzi, P. et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* **29**, 465901 (2017).

65. Giannozzi, P. et al. Quantum ESPRESSO toward the exascale. *J. Chem. Phys.* **152**, 154105 (2020).

66. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

67. Sergentu, D.-C., Gendron, F. & Autschbach, J. Similar ligand–metal bonding for transition metals and actinides? 5f^1 U(C_7H_7)$_2$ − versus 3d^n metallocenes. *Chem. Sci.* **9**, 6292–6306 (2018).

68. Nakamura, K. et al. RESPACK: an ab initio tool for derivation of effective low-energy model of material. *Comput. Phys. Commun.* **261**, 107781 (2021).

69. Rösner, M., Şaşıoğlu, E., Friedrich, C., Blügel, S. & Wehling, T. O. Wannier function approach to realistic Coulomb interactions in layered materials and heterostructures. *Phys. Rev. B* **92**, 085102 (2015).

70. Parcollet, O. et al. TRIQS: a toolbox for research on interacting quantum systems. *Comput. Phys. Commun.* **196**, 398–415 (2015).

71. Ruedenberg, K., Schmidt, M. W., Gilbert, M. M. & Elbert, S. Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model. *Chem. Phys.* **71**, 41–49 (1982).

72. Sun, Q. et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **153**, 024109 (2020).

73. Sayfutyarova, E. R., Sun, Q., Chan, G. K.-L. & Knizia, G. Automated construction of molecular active spaces from atomic valence orbitals. *J. Chem. Theory Comput.* **13**, 4063–4078 (2017).

74. Pathak, S. & Wagner, L. K. Non-orthogonal determinants in multi-Slater-Jastrow trial wave functions for fixed-node diffusion Monte Carlo. *J. Chem. Phys.* **149**, 234104 (2018).

75. Pathak, S. & Wagner, L. K. A light weight regularization for wave function parameter gradients in quantum Monte Carlo. *AIP Adv.* **10**, 085213 (2020).

76. Wheeler, W. A. et al. PyQMC: an all-Python real-space quantum Monte Carlo module in PySCF. *J. Chem. Phys.* **158**, 114801 (2023).

77. Chang, Y., Joshi, S. & Wagner, L. K. Learning emergent models from *ab initio* many-body calculations. Preprint at https://arxiv.org/abs/2302.02899 (2023).

78. Motta, M. & Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *WIREs Comput. Mol. Sci.* **8**, e1364 (2018).

79. Shi, H. & Zhang, S. Some recent developments in auxiliary-field quantum Monte Carlo for real materials. *J. Chem. Phys.* **154**, 024107 (2021).

Author contributions

Y.C., E.v.L., T.O.W., L.K.W., and M.R. designed the study. Y.C., E.v.L., B.E., B.B., and M.R. performed all calculations. All authors contributed to the discussion and interpretation of the results. The manuscript was written by Y.C., E.v.L., T.O.W., L.K.W., and M.R. with further contributions from M.A.M., C.D., A.J.M., and S.Z.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01314-6.

Correspondence and requests for materials should be addressed to Lucas K. Wagner or Malte Rösner.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.