

# Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods

Mario Motta,<sup>1</sup> David M. Ceperley,<sup>2</sup> Garnet Kin-Lic Chan,<sup>3</sup> John A. Gomez,<sup>4</sup> Emanuel Gull,<sup>5</sup> Sheng Guo,<sup>3</sup> Carlos A. Jiménez-Hoyos,<sup>3</sup> Tran Nguyen Lan,<sup>6,5,\*</sup> Jia Li,<sup>5</sup> Fengjie Ma,<sup>7</sup> Andrew J. Millis,<sup>8</sup> Nikolay V. Prokof'ev,<sup>9,10</sup> Ushnish Ray,<sup>3</sup> Gustavo E. Scuseria,<sup>4,†</sup> Sandro Sorella,<sup>12,13</sup> Edwin M. Stoudenmire,<sup>14</sup> Qiming Sun,<sup>3</sup> Igor S. Tupitsyn,<sup>9,10</sup> Steven R. White,<sup>14</sup> Dominika Zgid,<sup>6</sup> and Shiwei Zhang<sup>1,‡</sup>

(Simons Collaboration on the Many-Electron Problem)

<sup>1</sup>Department of Physics, College of William and Mary, Williamsburg, Virginia 23187, USA

<sup>2</sup>Department of Physics, University of Illinois at Urbana-Champaign, Champaign, Illinois 61801, USA

<sup>3</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

<sup>4</sup>Department of Chemistry, Rice University, Houston, Texas 77005, USA

<sup>5</sup>Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA

<sup>6</sup>Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

<sup>7</sup>Department of Physics, Beijing Normal University, Beijing, Beijing 100875, China

<sup>8</sup>Department of Physics, Columbia University, New York, New York 10027, USA

<sup>9</sup>Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003, USA

<sup>10</sup>National Research Center "Kurchatov Institute," 123182 Moscow, Russia

<sup>11</sup>Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA

<sup>12</sup>SISSA-International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy

<sup>13</sup>Democritos Simulation Center CNR-IOM Istituto Officina dei Materiali,

Via Bonomea 265, 34136 Trieste, Italy

<sup>14</sup>Department of Physics and Astronomy, University of California, Irvine, California 92697-4575, USA

(Received 1 May 2017; revised manuscript received 3 August 2017; published 28 September 2017)

We present numerical results for the equation of state of an infinite chain of hydrogen atoms. A variety of modern many-body methods are employed, with exhaustive cross-checks and validation. Approaches for reaching the continuous space limit and the thermodynamic limit are investigated, proposed, and tested. The detailed comparisons provide a benchmark for assessing the current state of the art in many-body computation, and for the development of new methods. The ground-state energy per atom in the linear chain is accurately determined versus bond length, with a confidence bound given on all uncertainties.

DOI: [10.1103/PhysRevX.7.031059](https://doi.org/10.1103/PhysRevX.7.031059)

Subject Areas: Chemical Physics,  
Computational Physics,  
Condensed Matter Physics

## I. INTRODUCTION

One of the grand challenges in modern science is the accurate treatment of interacting many-electron systems. In condensed phase materials, the challenge is increased by the need to account for the interplay between the electrons and the chemical and structural environment. Progress in addressing this challenge will be fundamental to the realization of “materials genome” or materials by design initiatives.

Often the physical properties of materials and molecules are determined by a delicate quantitative balance between competing tendencies, so that accurate computations are required to predict the outcome. The theoretical framework for these calculations, the many-particle Schrödinger equation, is known [1]. However, the solution of the Schrödinger equation in a many-electron system presents fundamental difficulties originating from combinatorial growth of the dimension of the Hilbert space involved, along with the high degree of entanglement produced by the combination of Fermi statistics and electron-electron interactions. Computational methods need to reach beyond the incredible success afforded by density functional theory (DFT) and capture electron correlation effects with sufficient accuracy across different physical parameter regimes.

No general, numerically exact method presently exists that can treat many-electron systems with low computational cost. Except for special cases, known methods either have systematic errors which cannot be easily quantified or

\*On leave from: Ho Chi Minh City Institute of Physics, VAST, Ho Chi Minh City, Vietnam.

†shiwei@wm.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

the computational burden scales exponentially or as a very high power of the system size.

Recent years have witnessed remarkable progress in the development of new theories, concepts, methodologies, software, and algorithms that have pushed the conceptual horizons and technical boundaries of computational many-body methods and considerably improved our understanding of interacting electrons in solids and molecules. A vast suite of methods exist which have different strengths and weaknesses and different domains of applicability, and ever more are being developed.

It is imperative, under these circumstances, to develop systematic knowledge by detailed benchmark studies. Comparison of different methods allows characterization of relative accuracy and capabilities, which provides a survey of the state of the art to guide applications. Applying complementary methods synergistically to the same problem enables cross-check and validation, leading to a powerful new paradigm of attack on difficult problems. Cases where results from different methods agree provide valuable benchmarks against which new methods can be tested, thereby facilitating further development and accelerating progress.

Detailed benchmark studies of extended systems have been rare. A major reason is the nature of the problems involved: while it is reasonably straightforward to compare results obtained for finite systems (such as molecules), or at the thermodynamic limit in an independent-electron picture, it is challenging to make reliable calculations in the thermodynamic limit with many-body methods. A recent success is the benchmark study of the Hubbard model [2], and a subsequent multimethod study of the ground-state order in the celebrated 1/8-doped case [3]. With real materials, two more challenges arise. First, the general long-range Coulomb interaction must be treated accurately. Second, many-body calculations often require the use of incomplete one-electron basis sets, whose systematic errors must be removed in order to reach the continuous space, or complete basis set, limit for physical observables.

In this work, we undertake a comprehensive benchmark study of state-of-the-art many-body methods in a more realistic context. We choose the linear hydrogen chain—introduced in Ref. [4] and studied at finite lengths and finite basis sets by several groups [5–10]—and investigate its ground state versus bond length under the Born-Oppenheimer approximation of fixed nuclear positions, at the thermodynamic and complete basis-set limits.

Hydrogen is the first element in the periodic table and the most abundant in nature. Studies of the H atom,  $H_2^+$  cation, and  $H_2$  molecule have served as landmarks in quantum physics and chemistry. Despite their deceptive simplicity, bulk H systems are rich and complex. The ground-state properties of the hydrogen chain can differ significantly from those of simpler systems such as the Hubbard model, and are, in fact, not completely understood. The linear H chain captures key features that are essential for generalizing

model-system methods to real materials, in particular: the presence of strong electron correlations of diverse nature as the H-H distance is varied, and the need to treat the full physical Coulomb interaction and to work in the continuous space and thermodynamic limits.

Compared to the one-dimensional Hubbard model, the hydrogen chain has multiple (in principle, infinite) orbitals per site, as well as long-range interactions. The use of basis sets defines models of the hydrogen chain of increasing complexity. In a minimal basis, there is only one band, and the problem resembles a one-dimensional Hubbard model with long-range interactions. Larger, more realistic, basis sets bring back characteristics of real materials. Thus, one can neatly and systematically connect from a fundamental model of strong electron correlation to a real material system. On the other hand, the H chain eliminates complexities of other materials systems such as the need to separately treat core electrons or incorporate relativistic effects, and is thus accessible to many theoretical methods at their current state of development. As such, the linear hydrogen chain is an ideal first benchmark system for testing the ability of many-body theoretical methods to handle the challenges posed by real materials.

We study finite chains of increasing length and cross-check the results against calculations performed using periodic boundary conditions. We also present results from calculations formulated in the thermodynamic limit. Most of the methods employed use a one-particle basis set, and we investigate convergence by obtaining results for a systematic quantum chemistry sequence of basis sets of increasing size. These results are compared to methods formulated directly in real space. With extensive direct comparisons and cross-validations between different methods, we characterize the uncertainties in each approach in detail.

This study presents results obtained from more than a dozen many-body computational methods currently used or under development in physics and chemistry. A vast amount of data are produced, providing detailed information in finite-length chains and with finite basis sets. In the largest systems treated, the size of the Hilbert space exceeds  $10^{100}$ . We anticipate that our data, which are made available in the appendixes and in online repositories [11], will be useful for benchmarking other existing and future electronic structure methods. In addition to the results and comparisons, we introduce a variety of methodological developments which are spurred by the benchmark, including new approaches. Combining the strengths of complementary methods, we are able to determine the energy per atom in the thermodynamic limit to sub-milli-Hartree accuracy. We hope that the results presented here will serve as a preview of what can be achieved in the predictive computation of the properties of real materials and provide a firm basis for theoretical progress in condensed matter physics, quantum chemistry, materials science, and related fields.

The rest of the paper is organized as follows. In Sec. II, we introduce the linear H chain systems that are studied and

define notation. In Sec. III, we give a brief overview of the many-body methods employed in the present work. More details on the methods and some of the computational details of each method are given in Appendix A. In Sec. IV, we present results for a finite  $H_{10}$  chain: In Sec. IVA, we summarize the results within a minimal basis set, in Sec. IV B, we present the results in the complete basis set (CBS) limit, and in Sec. IV C, we describe the extrapolation to the CBS limit from finite basis-set results. In Sec. V, we present results in the thermodynamic limit: Sections VA and VB contain results for the minimal basis and the CBS limit, respectively, while Sec. VC presents finite-size results and discusses the procedure for reaching the thermodynamic limit. A brief discussion and summary of our work, along with future prospects, is then given in Sec. VI. In the Appendixes, we include further descriptions of the methods, present tables that summarize our data, and provide further details of our benchmark results and procedures. A database of results is also made available electronically [11].

## II. HYDROGEN CHAIN

We consider a system composed of  $N$  protons, at fixed equispaced positions along a line, with  $N$  electrons. This system is described by the Hamiltonian

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i<j=1}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,a=1}^N \frac{1}{|\mathbf{r}_i - \mathbf{R}_a|} + \sum_{a<b=1}^N \frac{1}{|\mathbf{R}_a - \mathbf{R}_b|}, \quad (1)$$

where  $(\mathbf{r}_1 \dots \mathbf{r}_N)$  and  $(\mathbf{R}_1 \dots \mathbf{R}_N)$  are the coordinates of electrons and nuclei, respectively. We use atomic units throughout; i.e., lengths are measured in bohr [ $a_B = \hbar^2 / (m_e e^2)$ ] and energies in Hartree ( $E_{\text{Ha}} = e^2 / a_B$ ). In the thermodynamic limit of infinite system size at zero temperature, which is our primary focus, such a system is characterized by only one parameter, the bond length  $R$  separating two adjacent atoms.

The electron coordinates are continuous in three-dimensional space, while the nuclear coordinates are fixed on a line; e.g.,  $\mathbf{R}_a = aR\mathbf{e}_z$ , with  $R$  the interproton separation and  $a = 1, \dots, N$ . Most of our calculations are on such finite-size systems, referred to as open boundary conditions (OBC). We also perform calculations using periodic boundary conditions (PBC), in which case a periodic supercell containing  $N$  atoms is treated.

Among the methods employed here, diffusion Monte Carlo (DMC) and variational Monte Carlo (VMC) methods operate in first quantization and treat the Hamiltonian in Eq. (1) directly. The other methods use a finite one-electron basis set, with a total of  $M$  orbitals  $\{\varphi_p\}_{p=1}^M$ , i.e.,  $m \equiv M/N$  basis functions *per atom* (including occupied and virtual orbitals). The Hamiltonian is written in second-quantized form,

$$\hat{H} = \sum_{pq=1}^M h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs=1}^M v_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \quad (2)$$

where the creation and annihilation operators  $\hat{a}^\dagger$  and  $\hat{a}$  obey fermionic anticommutation relations and the indices  $p, q, r, s$  run over all  $M$  single-electron basis functions.

Most calculations are performed using standard Gaussian basis sets, but specialized density-matrix renormalization group (DMRG) calculations using a grid along the nuclear axis and Gaussians along the other two directions (sliced basis) are also performed, as discussed in Sec. 1 b. Within the Gaussian basis, the matrix elements  $\{h, v\}$  in Eq. (2) are readily obtained from standard quantum chemistry packages. The basis functions are centered on the protons. We use the correlation-consistent cc-pV $x$ Z basis set, with  $x = \text{D, T, Q, and 5}$ , which correspond to  $m = 5, 14, 30,$  and  $55$  orbitals per atom, respectively. For small  $N$ , explicit correlation using the F12 technique [12,13] was also considered to help ascertain the approach to the CBS limit. Our procedure for extrapolating to the CBS limit is described in Secs. IV C and V C.

H chains with nearest-neighbor proton separation (bond length)  $R$  of 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.4, 2.8, 3.2, and 3.6 bohr are studied. We focus, in this work, on the ground-state energy  $\mathcal{E}(N, R)$  for different chain sizes and lengths, and obtain the *energy per atom*,  $E(N, R) = \mathcal{E}(N, R)/N$ , at the thermodynamic limit (TDL),

$$E_{\text{TDL}}(R) = \lim_{N \rightarrow \infty} E(N, R). \quad (3)$$

We only consider spin-unpolarized systems, with  $N_\uparrow = N_\downarrow = N/2$ . Below, when presenting results on finite chains, we follow the chemistry convention and use the term potential energy curve (PEC), although we always refer to energy per atom,  $E(N, R)$ . When presenting results for the TDL, we use the term equation of state (EOS) to refer to  $E_{\text{TDL}}(R)$  versus  $R$  at zero temperature. Most of the methods consider chains with  $N = 10\text{--}102$  atoms. The procedure for extrapolating to  $N \rightarrow \infty$  is discussed in Sec. V C.

## III. OVERVIEW OF COMPUTATIONAL METHODS

The methods employed in this work include the following: AFQMC, auxiliary-field quantum Monte Carlo [14–16]. BDMC, bold diagrammatic Monte Carlo [17–19]. DMET, density-matrix embedding theory [20–22]. DMRG, density-matrix renormalization group with a quantum chemistry basis [23–26]. FCI, full configuration interaction, i.e., exact diagonalization. GF2, self-consistent second-order Green’s function [27–32]. LR-DMC, lattice-regularized [33] diffusion Monte Carlo [34]. MRCI and MRCI +  $Q$ , multireference configuration interaction without (MRCI) and with (MRCI +  $Q$ ) the Davidson correction [35,36].

NEVPT2, partially (PC) and strongly contracted (SC) variants of the  $N$ -electron valence state second-order perturbation theory [37].

RCCSD and RCCSD(T), coupled-cluster (CC) theory with full treatment of singles and doubles (RCCSD) and perturbative treatment of triple excitations [RCCSD(T)], using RHF as a reference state [38–40].

RHF, restricted Hartree-Fock [41].

SBDMRG, specialized DMRG with sliced basis sets [42].

SC-*GW*, fully self-consistent *GW* [27,43–45].

SEET, self-energy embedding theory [46–50].

UCCSD and UCCSD(T), CC theory with full treatment of singles and doubles (UCCSD) and perturbative treatment of triple excitations [UCCSD(T)], using UHF as a reference state [38–40].

UGF2, spin-unrestricted GF2 [29].

UHF, unrestricted Hartree-Fock [51].

VMC: variational Monte Carlo [52,53].

Detailed descriptions of the methods and specific calculational details are presented in Appendix A. We focus on many-body methods in this work. The independent-electron methods RHF and UHF are listed above, because they are used by many of our methods as reference, initial, or trial states. Further, we use the RHF to define the correlation energy for the purpose of extrapolation to the

CBS limit [regardless of the nature of the mean-field reference (if any) used in the method].

In Table I, we summarize the methods using several characteristics or criteria. At a high level, the methods can be distinguished by general categories such as wave function, embedding, and diagrammatic. Wave-function methods (AFQMC, CC, DMRG, FCI, LR-DMC, MRCI, NEVPT2, and VMC) formulate an *Ansatz* for the ground state and compute expectation values of observables and correlation functions with respect to the wave function. The *Ansatz* for the wave function can be explicit (as in VMC and most quantum chemistry methods) or reached via an iterative procedure (as in AFQMC, LR-DMC). The accuracy of a wave-function method is determined by the quality of the underlying *Ansatz* (e.g., form of trial wave function in VMC, size of truncated space, order of perturbation) and by approximations (if any) in the realization of the *Ansatz* (e.g., constraints in QMC) and in the evaluation of observables (e.g., nonvariational estimators in CC methods, mixed estimate or backpropagation in QMC). Extrapolations in  $N$  (and, in many cases, basis-set size  $M$ ) are needed for wave-function methods.

Embedding methods (DMET, SEET) evaluate the properties of a large system by partitioning it within a basis, e.g., the spatial or energy basis, into a collection of fragments, embedded in a self-consistently determined environment

TABLE I. Summary of the methods employed in the present work. Methods are classified by type: wave function, embedding, or diagrammatic; the use of one-electron Gaussian basis sets (b), sliced-basis sets (sb), or continuous electron positions (cs); whether a self-consistency procedure is involved; whether the method is deterministic or stochastic in nature; and whether the method is variational. The scaling of the computational cost of key pieces of the algorithm is shown, versus the number of electrons ( $N$ ), basis set size ( $M$ ), etc. In DMRG,  $D$  is the bond dimension kept in the calculation. In SBDMRG,  $N_o$  is the number of basis function per slice and  $D(N_o)$  the compressed MPO size. In Green's function methods,  $n_\tau$  is the size of the imaginary time grid. In DMET,  $N_f$  is the number of atoms in the fragment and  $D$  denotes the bond dimension kept by the DMRG solver. The first scaling corresponds to a DMET calculation with a single fragment using translational invariance, while the second corresponds to treating multiple fragments tiling the chain. In SEET,  $N_{\text{imp}}$  is the number of impurities, while  $n_e$  denotes the number of electrons in the impurity and  $M_s = M^A + M_b$  the number of orbitals treated, where  $M_b$  is the number of bath orbitals. The first scaling corresponds to SEET(CI/GF2), while the second corresponds to SEET(CI/HF). In BDMC,  $n$  denotes the diagrammatic order.

Method	Deterministic	Basis set	Self-consistent	Variational	Scaling	
Wave function	CCSD	Yes	b	Yes	No	$N^2M^4 + N^3M^3$
	CCSD(T)	Yes	b	Yes	No	$N^3M^4$
	DMRG	Yes	b	Yes	Yes	$D^3M^3 + D^2M^4$
	SBDMRG	Yes	sb	Yes	Yes	$NRD^3[N_o^3 + D(N_o)]$
	HF	Yes	b	Yes	Yes	$M^4$
	FCI	Yes	b	No	Yes	$\binom{M}{N}$
	MRCI	Yes	b	No	Yes	$> \binom{N}{N/2}N^4 + N^2M^4$
	NEVPT2	Yes	b	No	No	$\binom{N}{N/2}N^8$
	AFQMC	No	b	No	No	$N^2M^2 + M^2N$
	VMC	No	cs	No	Yes	$N^2M + N^3$
LR-DMC	No	cs	No	Yes	$N^2M + N^3$	
Embedding	DMET	Yes	b	Yes	No	$N_f^3D^3 + N_f^2D^4[(N_f^3D^3 + N_f^2D^4)M]$
	SEET	Yes	b	Yes	No	$N_{\text{imp}} \binom{M_s}{n_e} + M^5 n_\tau [N_{\text{imp}} \binom{M_s}{n_e} + M^4]$
Diagrammatic	SC- <i>GW</i>	Yes/No	b	Yes	No	$M^4 n_\tau$
	GF2	Yes	b	Yes	No	$M^5 n_\tau$
	BDMC $_n$	No	b	Yes	No	$e^{n \ln n}$

treated at a more approximate level. The accuracy of an embedding method is determined by a combination of several factors including the size of the fragments, the level of accuracy in the treatment of the embedded fragments and environment, and the level of convergence of the self-consistency loop. Extrapolations in the fragment size (and in the basis-set size  $M$  if a basis set is used) are needed for embedding methods.

Diagrammatic methods (BDMC, GF2, SC-GW) evaluate, either deterministically (GF2, SC-GW) or stochastically (BDMC), terms in a diagrammatic expansion of a system property. Diagrammatic expansions can be formulated either in a basis or directly in the continuum, and can be applied to finite systems or directly in the thermodynamic limit. GF2 provides an exact self-consistent evaluation of all second-order terms in a perturbative expansion in the interaction; SC-GW evaluates a small subset of diagrams to all orders in perturbation theory via the solution of self-consistent equations; BDMC accounts for higher-order vertex corrections within the skeleton expansion by performing a stochastic sampling of diagram topologies and internal integrations, and is limited to situations where the series converges. [We note an important ambiguity in the formulation of diagrammatic approximations: Hamiltonian terms that are identically zero because of the Pauli exclusion principle give rise to diagrams that may not sum to zero (self-interaction error) in approximations that do not consider all terms at a given order. The effect of these zero terms on SC-GW results is illustrated in Appendix A 4.]

Many other types of classification are possible. For example, one could characterize a method as deterministic or stochastic, whether a basis set is used and if so, what type, whether a self-consistent procedure is involved, whether the computed ground-state energy is variational, etc. Table I lists some of these classifiers, in addition to the computational scaling of key pieces in each algorithm. It is important to note that the classification is meant only as a general guide, and is in many cases fluid. For instance, embedding methods could also be classified by their solver for the embedding fragments, as wave function (DMET) or diagrammatic (SEET). Depending on the particular form of the solver, they could be deterministic (e.g., DMET with a DMRG solver as in the present study) or stochastic (if a QMC solver is used [54]). Various methods are shown as needing a Gaussian basis set, but can also be implemented using other bases (e.g., plane wave with pseudopotentials for AFQMC, and SC-GW). The choice of basis set can affect the computational scaling. Note also that the meaning of self-consistency can vary and depend on the type of methods. In wave-function methods, we use it to indicate whether a self-consistent procedure is involved, although this can still have ambiguity since there are sometimes multiple ways to obtain the solution. Finally, the scaling we report refers to the canonical implementation of these methods, without any specialized optimizations.

## IV. RESULTS FOR THE H<sub>10</sub> CHAIN

In this section, we present results on a finite chain of ten atoms. This relatively simple system provides a good intermediate step in the benchmark, as it removes one of the major challenges, namely, the approach to the thermodynamic limit. Detailed comparisons can be made as in quantum chemistry, providing insights for the more challenging case of the TDL. We emphasize the approach to the continuous space limit, with extensive studies on the removal of any residual finite basis errors. In Sec. IV A, results in the minimal basis are given, for which exact results from FCI are available for detailed comparison. Final results are presented for the CBS limit in Sec. IV B. Then, in Sec. IV C, we include results using finite basis sets and discuss the approach to the CBS limit.

### A. Benchmark in the minimal basis set

Figure 1 shows a detailed comparison of the potential energy curve,  $E(N = 10, R)$  versus  $R$ , obtained by a variety of methods, in the minimal STO-6G basis. For all methods, the PEC features a familiar short-range repulsion, due to the combined effect of Coulomb repulsion and Pauli exclusion, followed by a decrease to a minimum value  $E_0$ , attained at the equilibrium bond length  $R_e$ . Beyond  $R_e$ , the PEC monotonically increases to an asymptotic value  $E_\infty$ , the ground-state energy of a single H atom. The well depth gives a dissociation energy  $D_e = (E_\infty - E_0)$ . Owing to the small size of this chain and the STO-6G basis, the PEC can be calculated using the FCI method, giving the exact values  $R_e = 1.786a_B$ ,  $E_0 = -0.542457E_{\text{Ha}}$ , and  $E_\infty = -0.471039E_{\text{Ha}}$ .

The overall agreement between all the many-body methods is quite good. Deviations from FCI are shown in the lower panel of Fig. 1 in a magnified view. We see that the different methods agree with each other and with the exact result at the level of  $\sim 2mE_{\text{Ha}}$ , i.e., about 5% of the dissociation or cohesive energy. The agreement is better for  $R \leq 2.0a_B$  and tends to worsen in many methods as  $R$  increases, because electronic correlations become more pronounced, increasing the multireference character of the system.

MRCI and especially MRCI +  $Q$  are seen to be uniformly accurate for this system, with discrepancies of  $\mu E_{\text{Ha}}$  or less from FCI. This is also confirmed in the larger cc-pVDZ basis set (see Sec. IV C), where MRCI +  $Q$  and DMRG results are virtually indistinguishable. MRCI +  $Q$  can be conveniently carried out for even larger basis sets (but not for large  $N$ ), and we use it as our reference in the following sections for the ten-atom chain.

Among the other approaches, AFQMC gives results accurate to within  $0.2 mE_{\text{Ha}}$ . Bias from the CP approximation is visible in the intermediate region, where the energy is slightly overestimated, and also at large bond lengths, where there is an underestimation. Coupled-cluster methods, especially RCCSD(T), are very accurate near equilibrium. Although it is in principle possible to

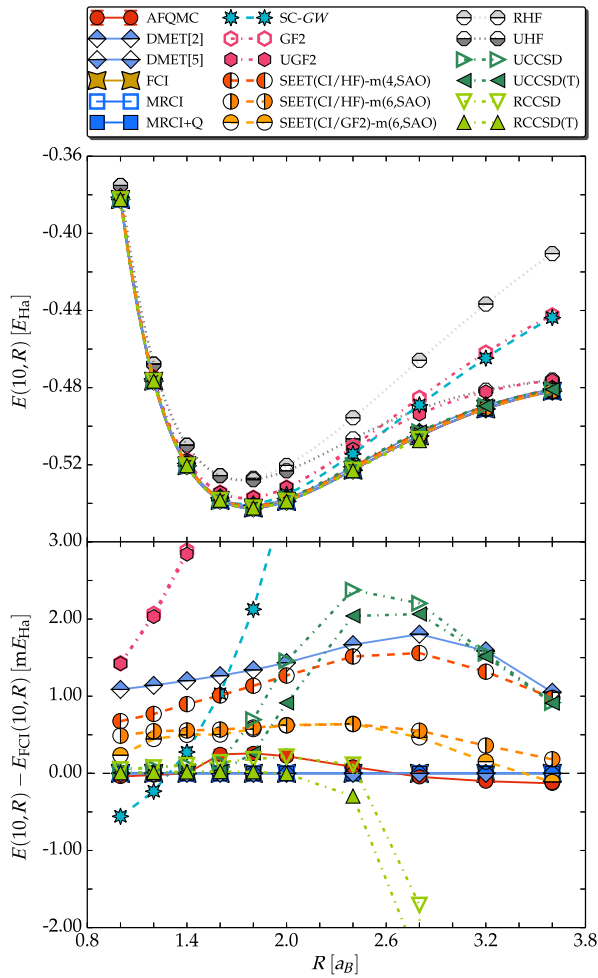


FIG. 1. Potential energy curve of  $H_{10}$  (top) and deviations from FCI (bottom), in the minimal STO-6G basis.

dissociate the H chain to the correct energy within RCCSD (as a product of dissociated H dimers), there can be multiple CC solutions, and in practice a correctly dissociating solution is hard to find [38]. UCCSD and UCCSD(T) provide accurate results at equilibrium and approach the correct result at longer bond lengths, but have large errors at the intermediate bond lengths due to spin recoupling.

In embedding methods, extending the number  $X$  of embedded atoms (in DMET[ $X$ ]) or the number  $Y$  of impurity orbitals (in SEET- $m[Y]$ ) leads to noticeable improvement. In the minimal basis, DMET[ $X$ ] uses  $X$  impurity orbitals because there is only one orbital per atom. The maximum error in DMET[2] is about  $2mE_{\text{Ha}}$ , while DMET[5] is exact by construction. While the SEET- $m[4]$  curve is at a similar level of accuracy to DMET[2], a substantial improvement is obtained within the mixing scheme (SEET- $m[6]$  curves), especially at large bond lengths.

As weak coupling methods, the diagrammatic GF2 and SC-GW methods have difficulties in the strong coupling regime at large bond lengths. Allowing methods to break spin symmetry may lead to an improvement of the energetics. As

illustrated with GF2, using an unrestricted reference state provides a better estimate of the ground-state energy in that regime but generates a spurious magnetization. Deviations at small distances (corresponding to the weak coupling regime) show that terms beyond the bare second-order or screened first-order approximation are needed to reach the accuracy of other methods. We also note that the cancellation of self-interaction error in SC-GW is subtle and depends on the treatment of exclusion principle violating terms in the Hamiltonian (see Appendix 4 d).

## B. Potential energy curve in the complete basis-set limit

In Fig. 2, we show the final computed potential energy curves of  $H_{10}$  in the continuous space (complete basis set) limit, including results obtained from VMC and LR-DMC, which work in continuous space.

For all our methods that require a basis set, we employ the correlation-consistent polarized valence  $x$ -zeta (cc-pV $x$ Z) sequence [55], which is designed to include successively larger shells of polarization functions ( $x = 2, 3, 4, 5$

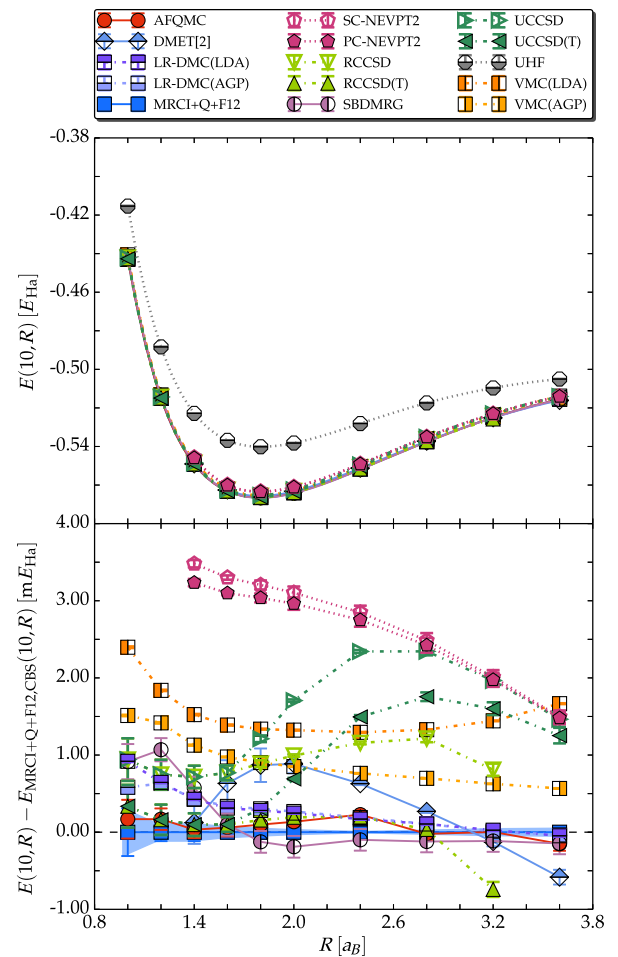


FIG. 2. Top: PEC of  $H_{10}$  in the continuous space, or complete basis set (CBS), limit. Bottom: Detailed comparison of the final  $H_{10}$  PECs, using MRCI +  $Q$  + F12 results at CBS as reference.

corresponding to  $D$ ,  $T$ ,  $Q$ , 5 respectively). The results are extrapolated to the CBS limit, following procedures described in Sec. IV C with further details given in Appendix C 1.

Our final results in this system give an equilibrium geometry  $R_e = 1.801(1)$  and energy  $E_0 = -0.5665(1)$ . The computed PECs are tabulated in the Appendix B. Deviations from the reference curve are shown in the bottom panel, where the combined uncertainties in the reference curve (primarily from the extrapolation to the CBS limit) are indicated. Our reference curve for this system is obtained from MRCI +  $Q$ , extrapolated to the CBS with basis sets up to  $x = 5$ . This is confirmed by a separate extrapolation including an F12 correction, which gives results in agreement within the statistical uncertainties.

Trends similar to the minimal basis results are observed. LR-DMC, which works in continuous electron coordinate space, has only a weak dependence on basis sets originating from the representation of the Slater determinant in the trial wave function (and hence the position of the nodes). LR-DMC provides an upper bound for the ground-state energy. Its quality is determined by the nodal surface of the trial wave function. At large bond length, the nodal structure is simpler, consistent with the more quasi-one-dimensional nature of the system. The LR-DMC results are very accurate in this regime, indicating that the DFT-LDA

determinant gives a good description of the nodal structure. The Antisymmetrized Geminal Power (AGP) trial wave function allows a more sophisticated, multideterminant description of the many-body nodes. Improvement with the AGP trial wave function is seen only at the smallest bond length. The excellent consistency between the LR-DMC results and the basis-set methods provides another assurance on the robustness of the approach to the CBS limit in the latter.

### C. Reaching the complete basis-set limit

For each method that utilizes a basis set, the computational cost grows as the basis size  $M$  is increased, in some cases very rapidly. In Fig. 3, we show the PEC yielded by several methods, in the cc-pV $x$ Z bases, taking MRCI +  $Q$  as reference. The accuracy of MRCI +  $Q$  is further validated by its excellent agreement with DMRG. The general trends seen at the minimal basis level are mostly confirmed with the larger basis sets. Most methods show errors that remain consistent throughout this family of basis sets which, though not surprising, is reassuring. As mentioned, improvements of the results are possible within certain methods, via larger embedding clusters, using better trial wave functions, or going to higher orders. Examples are

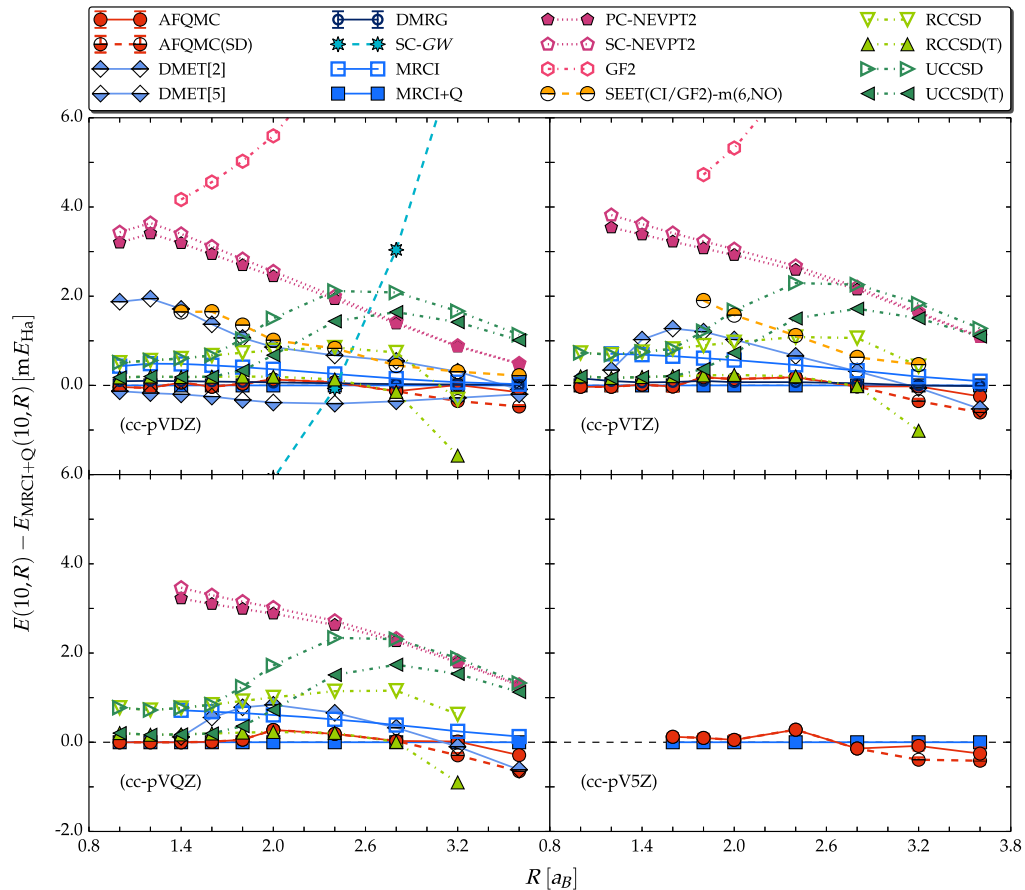


FIG. 3. Detailed comparison of the  $H_{10}$  PEC in each basis set (cc-pV $x$ Z,  $x = 2, 3, 4, 5$ ), using MRCI +  $Q$  as reference.

shown for DMET and SEET; for instance, at the cc-pVDZ level, increasing the number of embedded atoms from DMET[2] to DMET[5] reduces the maximum error from  $\sim 2 mE_{\text{Ha}}$  to  $0.5 mE_{\text{Ha}}$ . (In the cc-pVTZ and cc-pVQZ bases, cancellation of errors means that the maximum DMET[2] error is  $\sim 1 mE_{\text{Ha}}$ .) Since SEET(CI/GF2) works in the energy basis, increasing the number of impurity orbitals results in significant improvements. The mixing scheme, which illustrates the strong correlation present in the active space, recovers results of NEVPT2 quality while solving impurity problems with only six orbitals.

In AFQMC, a multideterminant trial wave function is used in the dissociation region (last two points,  $R > 3$ ) as discussed in Sec. A, which improves its accuracy by  $\sim 0.36 mE_{\text{Ha}}$  over that with the UHF trial wave function (shown as half filled red circles). For CC methods, the improvement from the inclusion of the perturbative triples is systematic and evident. MRCI +  $Q$  energies for the shortest bond lengths and large basis sets rely on a correction as discussed in Appendix B.

Increasing the basis-set size has a dramatic effect on the total energy, as seen in Fig. 4. The basis-set dependence is stronger at short bond lengths, with an energy difference of  $\sim 21 mE_{\text{Ha}}$  between cc-pVDZ and CBS, compared to  $\sim 3 mE_{\text{Ha}}$  in the bond-breaking regime.

The effect of the sliced basis used in the SBDMRG method is also illustrated in the main panel in Fig. 4. At the STO-6G level there are several competing effects that account for the deviations between the sliced basis and standard basis results. In the large  $R$  limit, the single basis function of STO-6G poorly describes an isolated hydrogen atom, and the increased flexibility in the chain direction of

the sliced basis can partially compensate for this. At shorter distances, the overlapping basis functions between adjacent atoms of standard STO-6G give additional degrees of freedom in the transverse directions, which can improve the energy at both the Hartree-Fock level and in terms of transverse correlations. In contrast, the sliced basis set has nearly ideal resolution in the longitudinal direction. At very short distances, the STO-6G basis becomes nearly linearly dependent while the sliced basis does not and consequently performs significantly better. These complicated competing transverse and longitudinal effects make it unsurprising that the differences between the two energies change sign as a function of  $R$ . For the cc-pVDZ bases, the sliced version is uniformly slightly better, probably because the dominant effect is its improved longitudinal correlation.

We extrapolate the finite basis-set results to the CBS limit by standard procedures [56,57], taking care to reach large basis sets. We first fit the RHF energies  $E_{\text{RHF},x}(N, R)$  computed at the cc-pV $x$ Z basis-set level to an exponential function

$$E_{\text{RHF},x}(N, R) = A(N, R) + B(N, R)e^{-xC(N,R)}. \quad (4)$$

The correlation energy

$$E_{\text{corr},x}(N, R) \equiv E(N, R) - E_{\text{RHF},x}(N, R) \quad (5)$$

is then fitted to a power law:

$$E_{\text{corr},x}(N, R) = \alpha(N, R) + \frac{\beta(N, R)}{x^3}. \quad (6)$$

The CBS result is taken as  $\alpha(N, R) + A(N, R)$ , with a combined uncertainty estimated from the fitting procedures. We find that using UHF as a reference gives numerically indistinguishable results, except for very short bond lengths and large sizes, where convergence of UHF shows more sensitivity.

To deal with basis-set linear dependence, which becomes relevant at the shortest bond lengths and largest basis sets, we apply a threshold to eigenvalues of the overlap matrix. Threshold values are given under method descriptions in Appendix A.

The final CBS results are verified with a separate set of MRCI +  $Q$  calculations augmented by F12 explicit correlation, as illustrated in the inset in Fig. 4. These data are extrapolated using the same procedure as above. That these results are consistent with those from the continuous-space LR-DMC provides a further validation of the procedure.

## V. RESULTS IN THE THERMODYNAMIC LIMIT

In this section, we present our results in the TD. The section is structured similarly to the previous one. Results for the minimal basis, which makes the H chain resemble an extended Hubbard model, can be valuable for model

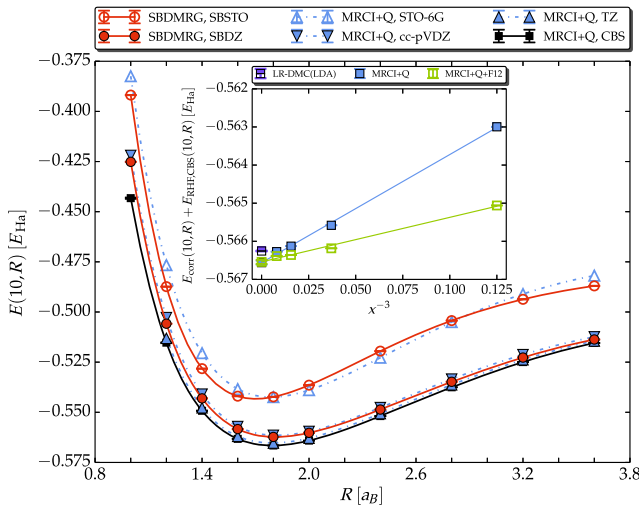


FIG. 4. Basis set dependence of the PEC in  $\text{H}_{10}$  and extrapolation to the CBS limit. MRCI +  $Q$  and SBDMRG results are shown in the main figure for selected basis sets (to avoid cluttering). The inset shows the extrapolation of the correlation energy (from MRCI +  $Q$ ) for two sequences of basis sets, cc-pV $x$ Z and with F12, together with LR-DMC, for  $R = 1.8$ .



studies and are described in some detail in Sec. VA. The final results for the hydrogen chain at the joint continuous space and thermodynamic limits are given in Sec. VB. Section VC discusses our procedures and cross-validations for approaching the thermodynamic limit.

### A. Benchmark in the minimal basis set

The minimal basis-set hydrogen chain is similar to an extended Hubbard model. As such, the results in this basis provide a quantitative connection to model studies. The computed EOS is shown in Fig. 5 for the STO-6G basis set. DMRG calculations can be carried out for large system sizes in this basis, and serve as a near-FCI quality benchmark. DMRG results for finite chains, after extrapolation to  $N \rightarrow \infty$ , yield an equilibrium geometry  $R_e = 1.831(3)a_B$  and ground-state energy per atom of  $E_0 = -0.5407(2)E_{\text{Ha}}$  at  $R_e$ .

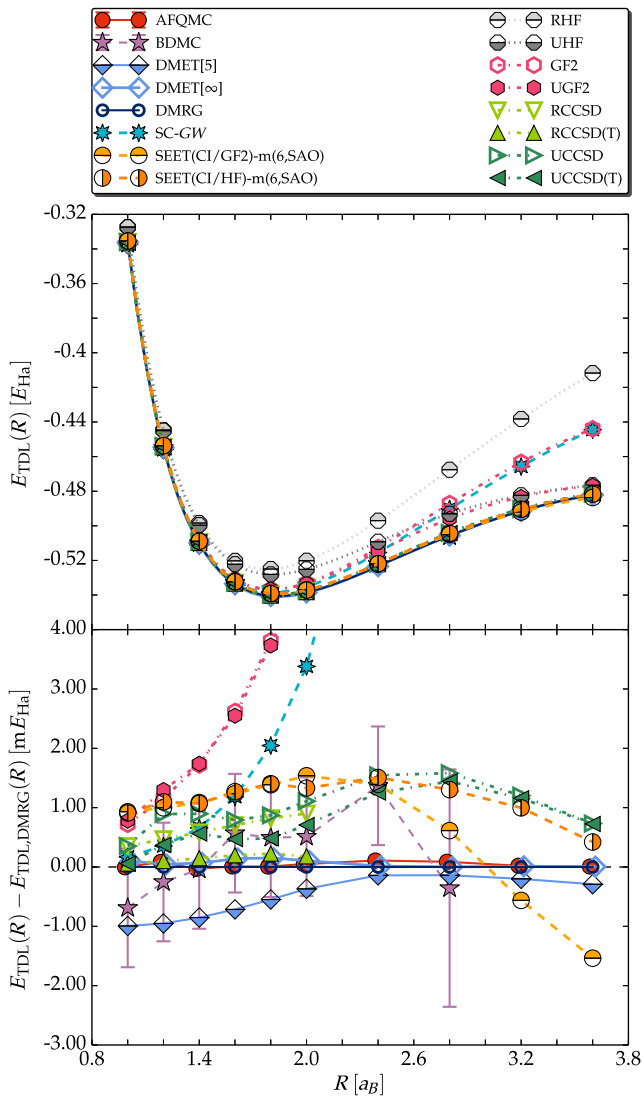


FIG. 5. Top: Computed equation of state in the minimal basis at the thermodynamic limit. Bottom: Detailed comparison using DMRG results as reference.

In the lower panel of Fig. 5, a detailed comparison is shown using DMRG as a reference. Most methods show similar behaviors as in finite chains. Coupled-cluster methods display the same general trends, with RCCSD(T) in particular giving extremely accurate results before the breakdown at larger bond lengths ( $R > 2a_B$ ). AFQMC yields energies accurate to within  $0.15 mE_{\text{Ha}}$  per particle across the bond lengths.

SEET is extrapolated to the TDL with respect to the chain length, with the number of orbitals treated by an accurate method fixed to 6. With this constraint, SEET(CI/HF)-m(6,SAO) shows accuracy at the thermodynamic limit comparable to the ten-atom chain when FCI is used to treat the impurity and HF to treat the environment, with a maximum error of  $\sim 1 mE_{\text{Ha}}$ . For stretched distances, SEET-m results improve if HF is used instead of GF2 since the latter [SEET(CI/GF2)-m(6,SAO)] results in overcorrelation.

As we discuss in Sec. 3 a, two types of DMET calculations are performed for the minimal basis. DMET[5] is from the first type, treating finite chains with fragment size  $N_f = 5$ , followed by extrapolation of the chain size  $N$  similar to the procedure used by most other methods whose results are shown here; this gives a maximum error of  $\sim 1 mE_{\text{Ha}}$ . DMET[ $\infty$ ] shows results from the second type, which works directly in the large- $N$  limit, and extrapolates the fragment cluster size  $N_f$ . Details of the extrapolation procedure are given in Sec. 3 a. DMET[ $\infty$ ] results should approach the exact limit similar to DMRG; the DMET[ $\infty$ ] and DMRG energies agree to better than  $0.15 mE_{\text{Ha}}$  per particle across all bond lengths.

BDMC<sub>3</sub> yields converged results up to  $R = 2.4$ . For  $R = 2.8$  convergence is reached only at the level of BDMC<sub>5</sub>; reaching convergence for larger values of  $R$  requires even higher orders. The calculated EOS is in good agreement with the exact results, and its final error bar of  $1 mE_{\text{Ha}}$  for  $R \leq 2.4$  is dominated by the resolution of the grid of 512 Matsubara frequencies used. For  $R = 2.8$  the error bar of  $2 mE_{\text{Ha}}$  is dominated by statistical noise in high diagrammatic orders. (The performance of lower-order BDMC<sub>*n*</sub> calculations is discussed in Appendix A 4, as well as their relations with SC-GW and other diagrammatic techniques.)

### B. Equation of state in the complete basis set and thermodynamic limits

Our final results for the equation of state of the hydrogen chain are presented in Fig. 6. Detailed numerical data are tabulated and included in Appendix. For these results, VMC and LR-DMC are extrapolated to the TDL, while basis-set methods are extrapolated to the joint TDL + CBS limit. We carry out extensive self-consistency and cross-checks in order to validate the extrapolations, as discussed in Secs. IV C and V C.

In the bottom panel of Fig. 6, AFQMC results are used as a reference, based on its accuracy from the systems that have been benchmarked. Large system sizes and basis sets

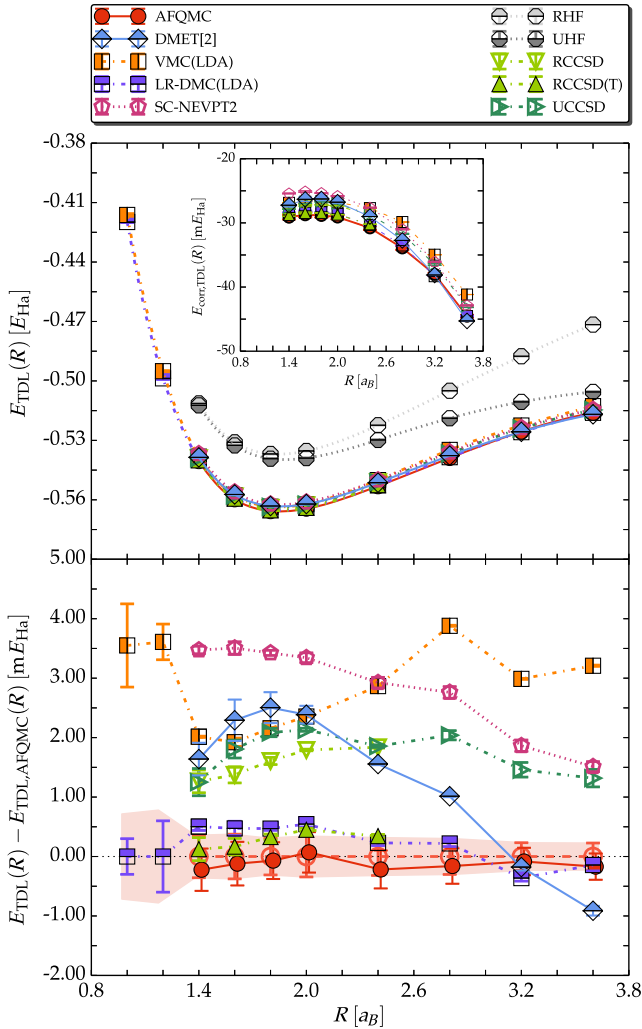


FIG. 6. Top: Computed equation of state of the hydrogen chain in the thermodynamic limit. The inset shows the corresponding correlation energy per particle. Bottom: Detailed comparison using AFQMC results as reference (LR-DMC for the shortest two bond lengths). The empty circles indicate AFQMC results after a correction is applied from the difference with DMRG at the cc-pVDZ level. The pink error bands indicate all statistical uncertainties.

can be reached to minimize the uncertainty in the extrapolation to the TDL and CBS limits. We can further quantify the residual systematic errors of the constraint in AFQMC from cross-checks with DMRG, by estimating their difference,  $E_{\text{TDL,DMRG}}(R) - E_{\text{TDL,AFQMC}}(R)$ , at the cc-pVDZ basis level. This “correction” can be applied to the AFQMC equation of state at the CBS limit,  $E_{\text{TDL,AFQMC}}(R)$ . The result is shown by the empty circles and dashed lines in Fig. 6, while original AFQMC data are shown with solid circles and lines.

Agreement is seen between these results and that from RCCSD(T), which provides another consistency check. We find an equilibrium bond length of  $R_{\text{eq}} = 1.859(3) a_B$  with an energy of  $E_0 = -0.5659(3) E_{\text{Ha}}$  at

$R_{\text{eq}}$  in the thermodynamic limit. The computed correlation energy, defined with respect to the RHF energy, is shown in the inset of Fig. 6.

At the smallest bond lengths ( $R \leq 1.2$ ), there is significant linear dependence in the basis sets. This causes an effective reduction in the size of the basis, which can render the usual *Ansatz* for basis-set extrapolations unreliable. We thus avoid performing CBS extrapolations. (All finite-basis data are listed in the Appendix B and repository [11]. It will be valuable to develop specifically designed basis-set sequences or correction methods in this regime.) LR-DMC results are shown, which provide an upper bound for the energy. Based on the results in Sec. IV B, the fixed-node error is estimated to be  $< 1 mE_{\text{Ha}}$  per atom, which is indicated by the pink error bands on these two points.

LR-DMC results are obtained directly in real space and provide an independent validation. At large bond lengths, the fixed-node error in LR-DMC is minimal, as we have seen in the finite-chain benchmarks. Furthermore, we perform PBC calculations using LR-DMC to provide a separate extrapolation to the TDL. The excellent agreement between LR-DMC and AFQMC at large  $R$  is thus a further indication of the robustness of our procedures for reaching the infinite basis set and thermodynamic limits. Note that the VMC results exhibit a different trend from the corresponding LR-DMC, suggesting that the variational many-body wave function is best at intermediate bond lengths. This is likely a reflection of the balance between the two parts that form the VMC *Ansatz*, namely, the LDA Slater determinant and the optimized Jastrow factor. The former becomes more accurate in describing the nodal surface as  $R$  increases, where the latter is evidently more effective at weaker correlation. Only the determinant part, via the nodes that it defines, impacts the DMC results.

The DMET[2] results provide an example of an embedding calculation at the thermodynamic and complete basis-set limits, with a modest impurity size. The limitation on the impurity size is from the use of a DMRG impurity solver, which becomes expensive in the large basis-set limit.

We comment that various correction schemes can be applied to our finite-basis and/or finite-size data to provide additional estimates from methods not included in Fig. 6. For example, a residual basis-set correction could be obtained either from a different method or using a lower-order theory (if available) of the same method, and applied to DZ or TZ basis results to estimate the CBS limit. These can be readily retrieved for assessment from the detailed data provided in the Appendix B and repository [11].

### C. Reaching the complete basis-set and thermodynamic limits

A key challenge in the *ab initio* computation of bulk materials is to remove finite-size and finite-basis effects so

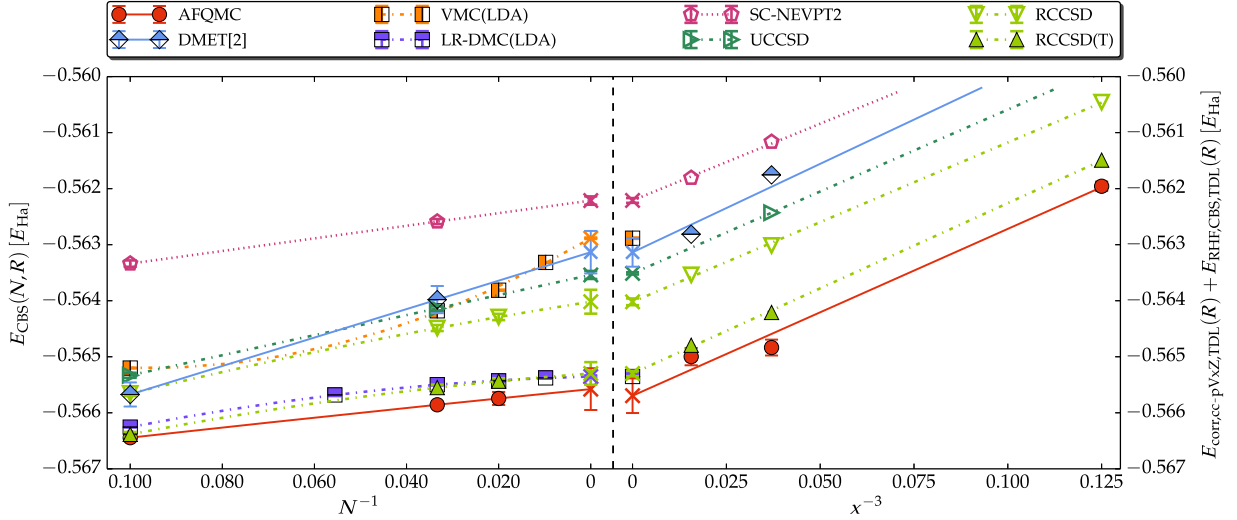


FIG. 7. Illustration of the extrapolations to the CBS and TDL limits. Results are shown for  $R = 1.8 a_B$ . The left-hand panel shows extrapolation of  $E_{\text{CBS}}(N)$  versus  $1/N$ , while the right-hand panel shows extrapolation of  $E_{\text{cc-pVxZ}}(N \rightarrow \infty)$  versus  $1/x^3$ . (The correlation energy is shown on the right, shifted by the CBS RHF energy.) Final results are consistent within statistical errors and independent of the order with which the limits are taken.

as to obtain results for the continuous and thermodynamic limits. This is important in order to make reliable predictions about materials properties and allow direct comparisons with experiments. Various choices exist in the calculation. These can be at the level of the type of many-body methods, for example, the use of particular embedding approaches (versus those that treat a cluster only, whether finite or periodic), or the use of coordinate space methods like DMC versus basis-set methods. They can also be common to classes of methods and decoupled from and independent of the details of the underlying many-body methods, for example, the use of periodic supercells versus finite clusters, or the choice of basis sets, etc. By employing many state-of-the-art methods, we are able to investigate these factors extensively and with great care in the present work.

Many of our calculations are performed using OBC, i.e., treating a finite chain. We find that, somewhat surprisingly, OBC calculations show faster convergence to the TDL than PBC in the hydrogen chain for all but the smallest few bond lengths (see Appendix C 2). To extrapolate the finite- $N$  results to the TDL, we assume that the PEC has the following size dependence:

$$E(N, R) = \sum_{i=0}^k \frac{A_i(R)}{N^i}, \quad (7)$$

where  $k$  is a small integer. For  $k = 1$ , this gives the subtraction trick based on a division of surface and bulk terms, namely,  $A_0 = [N_1 E(N_1) - N_2 E(N_2)] / [N_1 - N_2]$  (omitting  $R$ ), which has been used, for example, in DMRG calculations before [58]. In this work, we typically use  $k = 2$ , employing  $N = 10, 30, 50$  and, when necessary,  $N = 18, 22, 70$ , and 102.

Under this choice, there are still multiple strategies for finite basis-set methods to approach the combined limits. One could extrapolate to the CBS limit for each finite chain of fixed  $N$  following the procedure described in Sec. IV C, and then extrapolate the results in  $N$  to the TDL. Alternatively, one could extrapolate each basis set to the TDL, and then extrapolate to the CBS limit, or use a joint *Ansatz* and extrapolate both simultaneously. As illustrated in Fig. 7, exchanging the order of the extrapolation leads to consistent and robust results.

With the exception of the minimal basis, the TDL extrapolation for the DMET data is performed using the same OBC size dependence described above. In the minimal basis, we also carry out DMET calculations directly in the TDL, as mentioned earlier. Additional details on this extrapolation scheme and a comparison of the two is discussed briefly in the Appendix C 2. Calculations in PBC (including those with BDMC, which uses a ring geometry, and LR-DMC) are extrapolated to the TDL using the form  $E(N, R) = A_0(R) + A_2(R)N^{-2}$  [59], and statistical error bars are propagated following standard procedures in the extrapolations.

## VI. CONCLUSIONS

We present a comprehensive investigation of the hydrogen chain, deploying a vast suite of cutting-edge many-body numerical methodologies and obtaining a detailed and quantitative understanding of current computational capabilities for treating correlated quantum materials. We show how finite-size effects and finite spatial or other basis-set resolutions can be systematically removed to reach the physically relevant infinite system size and complete basis-set limits.

Through the synergistic use of complementary methods, we accurately determine the ground-state energy as a function of interatomic distance. This serves as a proof of concept for a new mode of attack on correlated materials by *ab initio* calculations. The benchmark results will provide a reference on the state of the art in many-body computation of real materials.

Our study captures many of the salient features of predictive computations in real materials. The ability of each many-body method to correctly capture important physical properties will depend on the material system under study. For example, perturbative or diagrammatic methods can have better or worse accuracy in systems with different amounts of electron correlation, the qualities of the constraints on the sign problem in QMC methods can vary with the physical nature of the problem, the rate of convergence with the fragments or the requirement on the impurity solver in embedding methods can differ from material to material, the scaling and computational feasibility of DMRG can change, etc. More benchmark studies of this kind will be highly desirable to broaden the understanding and identify further limitations as well as opportunities of development.

The computational cost of each method depends on various factors, including the degree to which the algorithm and codes have been optimized, the level to which one wishes to take the calculation (the order in perturbative or diagrammatic methods, or the statistical accuracy in Monte Carlo methods), etc. The results in this benchmark are obtained with moderate computing (order of days on platforms ranging from local clusters to medium-sized supercomputers). The computational scaling, which is summarized in Sec. III, together with the corresponding accuracy achieved by each method in the benchmark, will provide a rule of thumb on their computational cost.

The benchmark results indicate that many of the methods tested here are capable of reaching an accuracy of 5% of the cohesive energy or better, across wide parameter regimes of strong electron correlation. A subset of these methods predicts the equation of state systematically to sub-milli-Hartree accuracy. Further development may turn these into post-DFT methods of choice for ground-state studies, and a concerted effort to build open-source codes will be invaluable. Other techniques can more naturally address dynamical and thermodynamical properties, many of which are the outcome of recent research. Continued development along these lines will further improve their accuracy and time to solution. Further benchmark studies of dynamical and thermal effects, building on the work done here on the equation of state, would also be very desirable.

It is important to continue to expand the benchmark studies to more complex materials. Even in this relatively simple system of the hydrogen chain, important questions remain on the physics which are of strong interest and relevance to some of the key issues in strongly correlated

systems in general. For example, how does the nature of the charge and magnetic orders vary with the bond length? We are presently investigating these and related questions.

## ACKNOWLEDGMENTS

We gratefully acknowledge the Simons Foundation for funding. We thank E. Kozik, H. Krakauer, M. van Schilfhaarde, H. Shi, B. Svistunov, and N. Tubman for valuable interactions. Support from National Science Foundation (NSF) (Grant No. DMR-1409510) is acknowledged for method development work at William & Mary. F. M. was also supported by Department of Energy (DOE) (Grant No. DE-SC0001303). The work at the California Institute of Technology was supported by the Department of Energy, through DOE-SC0008624. G. K.-L. C. is a Simons Investigator. The work at Rice University was supported by Grant No. NSF-CHE-1462434. J. A. G. acknowledges support from the National Science Foundation Graduate Research Fellowship Program (DGE-1450681). G. E. S. is a Welch Foundation Chair (C-0036). I. S. T. and N. V. P. acknowledge NSF under Grant No. PHY-1314735. S. S. acknowledges computational resources provided through the High-Performance Computing Infrastructure (HPCI), Advanced Institute for Computational Science (AICS) projects No. hp120174, No. hp140092, No. hp160126, and No. hp170079. S. R. W. and E. M. S. acknowledge support from the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Grant No. DE-SC008696. E. G. was also supported by DOE Grant No. ER 46932, J. L. by NSF DMR 1606348, and computer resources were provided by TG-DMR130036. D. Z. and T. N. L. were also supported from DOE Grant No. ER16391.

## APPENDIX A: DESCRIPTION OF COMPUTATIONAL METHODS

In this Appendix, we provide further descriptions of the individual methods used. Following the main text, we group methods into the following categories: deterministic wave function (CC, DMRG, MRCI, NEVPT2), stochastic wave function (AFQMC, LRDMC, VMC), embedding (DMET, SEET), and diagrammatic (SC-GW, GF2, BDMC). The categories are by no means rigid, as a method can fit into multiple groups; they are meant to provide a general guide and help with organization of the discussions.

### 1. Deterministic wave function methods

Deterministic wave-function methods (HF, CC, DMRG, FCI, MRCI, NEVPT2) range in quality between the mean-field HF and the exact FCI. Correspondingly, their computational costs vary a great deal. These methods rely on different types of *Ansätze*, the nature of which is ultimately responsible for their accuracy and computational cost.

### a. Coupled cluster (CC)

Coupled-cluster theory [38–40] is widely applied, often providing accurate and systematically improvable ground-state energies when systems are neither too large nor too strongly correlated. The CC wave function is written as

$$|\text{CC}\rangle = e^{\hat{T}}|0\rangle, \quad (\text{A1})$$

where  $|0\rangle$  is a single-determinant reference state, and the cluster operator is given by

$$\hat{T} = \sum_{\mu} t_{\mu} \hat{t}_{\mu}, \quad (\text{A2})$$

where  $\hat{t}_{\mu}$  creates an excited determinant  $|\mu\rangle$  containing  $\mu$  particle-hole pairs relative to  $|0\rangle$ , with amplitude  $t_{\mu}$ . Standard CC theory constructs a similarity-transformed Hamiltonian

$$\hat{H}' = e^{-\hat{T}} \hat{H} e^{\hat{T}}, \quad (\text{A3})$$

and the energy and amplitudes  $t_{\mu}$  are obtained by solving the Schrödinger equation projectively:

$$E = \langle 0 | \hat{H}' | 0 \rangle, \quad (\text{A4a})$$

$$0 = \langle \mu | \hat{H}' | 0 \rangle, \quad \forall \mu. \quad (\text{A4b})$$

In other words, CC theory diagonalizes the similarity-transformed Hamiltonian in the space spanned by the mean-field reference and excitation manifold. Because  $\hat{T}$  is an excitation operator, the commutator expansion used to evaluate the amplitudes in Eq. (A4) terminates after four nested commutators for all values of  $\mu$ , because the Hamiltonian contains only one- and two-particle terms. In this work, we limit  $\hat{T}$  to  $\mu \leq 2$ , i.e., CC with single and double excitations (CCSD) on restricted Hartree-Fock (RCCSD) and unrestricted Hartree-Fock (UCCSD) references [60,61]. Since the quality of the reference determines the quality of the CC wave function, a reference obtained from a symmetry-broken mean field can be critical to getting good CC energies when systems become strongly correlated. When multiple Hartree-Fock solutions exist, the question of which one to use as a reference for the CC calculations can be subtle. In general, our CC calculations are performed on the lowest-energy Hartree-Fock determinants we could find. In some calculations, we also perturbatively include triple-excitation effects, denoted CCSD(T) [62].

All CC calculations we show here are widely available in standard quantum chemistry packages [63]. To carry out the larger calculations, we use the high-performance implementation in the PYSCF package. This uses an AO-driven implementation to reduce the input-output (IO) costs

associated with accessing integrals on a disk [64]. We carry out RCCSD and RCCSD(T) calculations for  $\text{H}_{30}$  with cc-pVQZ and cc-pV5Z basis and  $\text{H}_{50}$  with cc-pVTZ and cc-pVQZ basis with this implementation. Even with the AO-driven technique, the RCCSD calculation for  $\text{H}_{50}$  in the cc-pV5Z basis would require at least 4 TB of disk space. Although technically feasible, we do not perform it here. For the largest basis sets cc-pVTZ, cc-pVQZ, and cc-pV5Z at geometries  $R < 1.6 a_B$  and cc-pVDZ at  $R < 1.4 a_B$ , the Gaussian basis is nearly linearly dependent. We remove linearly dependent vectors with an overlap threshold of  $10^{-7}$ . Although a smaller threshold could be used in the Hartree-Fock calculation, changing the energy by  $\sim 10^{-6} E_{\text{Ha}}$  per atom, we find the resulting CCSD calculations with smaller thresholds to be numerically unstable.

### b. Density-matrix renormalization group (DMRG, SBDMRG)

DMRG is a low-entanglement wave-function approximation [23]. The wave function can be written as

$$|\Psi\rangle = \sum_{\{n\}} \mathbf{A}^{n_1} \mathbf{A}^{n_2} \dots \mathbf{A}^{n_M} |n_1 n_2 \dots n_M\rangle, \quad (\text{A5})$$

where  $M$  denotes the number of orbitals in the system. Each  $\mathbf{A}^n$  is a  $D \times D$  matrix of real numbers associated with a single-particle basis function, except for the boundary terms  $\mathbf{A}^{n_1}$  and  $\mathbf{A}^{n_M}$ , which are length- $D$  vectors.  $D$  denotes the bond dimension and controls the accuracy of the simulation; as  $D$  increases, the wave function converges to the exact correlated state. In linear systems such as the hydrogen chains we consider here, provided that the gap of the system does not close, the bond dimension required for a given accuracy per atom stays close to a constant, independent of the number of atoms. The energy of the wave function may be stably computed and variationally optimized through the DMRG sweep algorithm.

In this work, we consider two different kinds of single-particle basis functions in the DMRG calculations. In the standard quantum chemistry formulation of DMRG [24–26], the single-particle basis is simply an orthogonal basis in the space of Gaussian orbitals of the system. This is what we refer to as DMRG in the calculations in this work, and details can be found in standard references to DMRG in the quantum chemistry literature [65]. We carry out Gaussian-based DMRG calculations using the implementation in the BLOCK code, with the standard settings described in Ref. [66]. DMRG energies are computed for  $\text{H}_{10}$  (STO-6G, cc-pVDZ, cc-pVTZ bases),  $\text{H}_{30}$  (STO-6G, cc-pVDZ), and  $\text{H}_{50}$  (STO-6G), up to a maximum bond dimension of  $D = 2000$ . For the STO-6G basis, the DMRG single-particle basis is the set of symmetrically orthogonalized AO orbitals. For all the other bases, we use split-localized molecular orbitals. The split-localized orbitals are ordered by the exchange Fiedler vector [66,67]. The estimated maximum

uncertainty in the Gaussian-based DMRG energies is less than  $0.05mE_{\text{Ha}}$  per atom.

In addition, we also introduce the sliced basis DMRG method (SBDMRG). Here, instead of a 3D Gaussian basis, one uses a grid in one direction ( $z$ ), while using a Gaussian-derived basis in the two transverse directions ( $x, y$ ). Formally, one can write the basis as

$$\phi_{jn}(x, y, z) = \varphi_{jn}(x, y)\delta^{1/2}(z - z_n), \quad (\text{A6})$$

where  $n$  labels grid points along the  $z$  direction (with grid spacing  $a$ ) and  $j$  labels the transverse basis function at that grid point (or “slice”). The  $1/2$  power on the Dirac delta function indicates that the basis functions are square-normalized. The kinetic energy terms in the Hamiltonian are approximated with a fourth-order finite-difference second-derivative approximation. For the data presented in Sec. V B, we use a grid spacing of  $a = 0.1$ , for which we estimate an error of about  $0.1mE_{\text{Ha}}$  per atom. The transverse basis functions  $\varphi_{jn}(x, y)$  are derived from a standard Gaussian, atom-centered basis set. Functions from the standard basis are projected onto each slice, then these functions are orthogonalized, keeping the most significant ones up to as many as the number of contracted orbitals on each atom. Compared to the original basis, a sliced basis has approximately the same transverse resolution, but its resolution in the  $z$  direction is essentially perfect. Thus, energies in the sliced basis are generally lower than in the original basis, due to the improved  $z$  correlation. The key advantage of SBDMRG over the standard quantum chemistry formulation of DMRG is that the local support of the basis functions along the grid direction makes the number of Hamiltonian terms proportional to  $M^2$ . Using matrix product operator compression techniques, which are quite simple to apply in the sliced basis formulation, the cost of SBDMRG is further reduced to  $M$ , which is the same scaling as applying DMRG to the 1D Hubbard model. For more details, see the recent paper [42].

### c. Multireference configuration interaction (MRCI)

MRCI is a method that incorporates 1- and 2-external excitations on top of an active space wave function. It is a commonly used method for high-accuracy simulations of multireference electronic structure in small molecules. Here, we use a variant of internally contracted MRCI described by Werner and Knowles [35,36], implemented in the MOLPRO package [68]. We start with a CASSCF (complete active space self-consistent-field) wave function  $|\Psi_0\rangle$ . The variational *Ansatz* is then

$$|\Psi\rangle = c_0|\Psi_0\rangle + \sum_I c_I|\Psi_I\rangle + \frac{1}{4}\sum_{ijab} c_{ij}^{ab}\hat{E}_i^a\hat{E}_j^b|\Psi_0\rangle, \quad (\text{A7})$$

where  $|\Psi_0\rangle$  is the CASSCF wave function,  $|\Psi_I\rangle$  is a configuration state function (CSF) with a single external orbital, and  $\hat{E}_i^a$  is the spin-summed excitation operator,

$\sum_{\sigma}\hat{a}_{i\sigma}^{\dagger}\hat{a}_{i\sigma}$ . The parameters  $c_0$ ,  $c_I$ , and  $c_{ij}^{ab}$  are determined variationally.

We also consider the explicitly correlated (F12) MRCI approximation [69–71]. Explicit correlation accelerates convergence to the complete basis-set limit by introducing 2-external amplitudes with explicit  $r_{12}$ -dependent functions. The associated integrals are computed through an auxiliary Gaussian basis. In this work, we use the default F12 settings and auxiliary bases in MOLPRO, including the singles corrections in the complementary auxiliary basis-set space.

The MRCI wave function does not give an extensive energy. Defining the MRCI correlation energy as  $\Delta E = \langle\Psi|\hat{H}|\Psi\rangle - \langle\Psi_0|\hat{H}|\Psi_0\rangle$ , we define the approximate size-extensive correlation energy ( $Q$ ) through the scaling  $\Delta E \rightarrow \Delta E(1 - c_0^2)/c_0^2$ .

With the above techniques, we compute MRCI +  $Q$  and MRCI-F12 +  $Q$  wave functions and energies for  $\text{H}_{10}$  in the STO-6G and cc-pVxZ ( $x = 2-5$ ) bases, using a (10, 10) CASSCF initial state.

### d. $N$ -electron valence state perturbation theory

$N$ -electron valence state perturbation theory (NEVPT) [37] is a multireference second-order perturbation theory that is size extensive and free of intruder state problems. It uses a zeroth-order CASSCF wave function and Dyall’s Hamiltonian [72] as the zeroth-order Hamiltonian. From this starting point, the first-order wave function and second-order energy are defined using the usual Rayleigh-Schrödinger perturbation theory. Typically, the first-order equation is not solved exactly, but rather in a restricted variational space. In partially contracted NEVPT2 (PC-NEVPT2), the first-order wave function is expanded in the space of 1- and 2-external excitation operators acting on the ground-state wave function. In strongly contracted NEVPT2 (SC-NEVPT2), the expansion space of the first-order wave function is restricted further such that the amplitudes can be determined without solving any linear equations, and simply from expectation values of the zeroth-order wave function.

For the  $\text{H}_{10}$  chain, we compute SC-NEVPT2 and PC-NEVPT2 using the MOLPRO package, starting from a (10,10) CASSCF zeroth-order state. For  $\text{H}_{30}$ , we carry out DMRG-SC-NEVPT2 calculations [73] using the PYSCF package, starting from a (30,30) DMRG-CASSCF zeroth-order state computed with split-localized orbitals, using the BLOCK package. The basis linear dependency threshold is set to  $10^{-8}$ . The DMRG-CASSCF calculation is carried out with bond dimension  $D = 1000$ , leading to an estimated energy error of less than  $0.1 mE_{\text{Ha}}$ . The zeroth-order wave function is constructed by compressing the DMRG-CASSCF wave function down to bond dimension  $D = 500$ , with a compression error in the total energy of less than  $0.3 mE_{\text{Ha}}$  except at the shortest geometry,

$R = 1.0 a_B$ , where it is  $10 mE_{\text{Ha}}$ . The semi-internal components of the DMRG-SC-NEVPT2 wave function and energy are approximated using the matrix product state (MPS) compression scheme described in Refs. [74,75], with a first-order wave function bond dimension of  $D = 1500$ ; these contributions are determined with an estimated accuracy of  $0.1 mE_{\text{Ha}}$ .

## 2. Stochastic wave-function methods

Stochastic wave-function methods (AFQMC, VMC, LRDMC) rely on Monte Carlo sampling to construct an *Ansatz* for the ground state of the system and compute expectation values of observables. AFQMC and LRDMC are both based on mapping the imaginary-time evolution onto a random walk. AFQMC is formulated in nonorthogonal Slater determinant space. LRDMC (and VMC) conducts the random walk in coordinate space. The fermion sign problem has different manifestations in the different manifolds, and the constraints to control them lead to different approximations.

### a. Auxiliary-field quantum Monte Carlo (AFQMC)

The AFQMC is a wave-function method, which estimates the ground-state properties of a many-fermion system by statistically sampling the wave function  $e^{-\beta\hat{H}}|\Psi_0\rangle \propto |\Psi_\beta\rangle \rightarrow |\Psi_G\rangle$ , where  $\Psi_0$  is an initial wave function, nonorthogonal to the ground state  $\Psi_G$  of  $\hat{H}$  [15,16]. For sufficiently large  $\beta$ , expectation values computed over  $\Psi_\beta$  give ground-state averages. AFQMC projects  $\Psi_0$  towards  $\Psi_G$  iteratively, writing  $e^{-\beta\hat{H}} = (e^{-\delta\beta\hat{H}})^n$ , where  $\delta\beta = (\beta/n)$  is a small imaginary-time step. The propagator is represented as  $e^{-\delta\beta\hat{H}} = \int d\mathbf{x} p(\mathbf{x}) \hat{B}(\mathbf{x})$ , where  $\hat{B}(\mathbf{x})$  is a mean-field propagator of the form of an exponential of one-body operators that are dependent on the vector  $\mathbf{x}$ , and  $p(\mathbf{x})$  is a probability distribution [76,77]. This representation maps the original interacting system onto an ensemble of noninteracting systems subject to a fluctuating potential. The imaginary-time projection can be realized as an open-ended random walk over the auxiliary-field (i.e., mean-field potential) configurations [15,16]. Sampling the trajectories of the random walk leads to a stochastic representation of  $\Psi_\beta$  as an ensemble of Slater determinants.

For general two-body interactions AFQMC has a sign or phase problem, which is controlled by a phaseless gauge constraint (CP) on the Slater determinants using a trial wave function  $\Psi_T$  [16,78]. (For Hamiltonians that satisfy certain symmetry properties, e.g., the Hubbard model at half filling, AFQMC is free of the sign problem.) The trial wave function is typically taken from Hartree-Fock or DFT. A self-consistent constraint is possible [79] but is not used in this work. The accuracy of the CP AFQMC was extensively benchmarked in both real materials [5,80,81] and lattice models [2,79]. The CP AFQMC provides an

alternative and complementary way to address the sign problem with respect to fixed-node DMC. The random walks take place in the overcomplete manifold of Slater determinants, in which fermion antisymmetry is by construction maintained in each walker. Applications have indicated that often this reduces the severity of the sign problem and, as a result, the phaseless approximation has weaker reliance on the trial wave function [82].

For *ab initio* materials computations, AFQMC can be carried out using either a plane-wave basis and pseudopotentials [16,83], localized basis sets such as standard Gaussian-type orbitals [84], or general basis sets using DFT orbitals [85]. In this work, we apply AFQMC implemented for Gaussian basis sets to finite chains. In all our calculations we use a linear-dependence threshold of  $10^{-7}$  in the one-electron basis. The two-body matrix elements  $v_{pqrs}$  are decoupled into bilinear form with the modified Cholesky approach using a tolerance of  $\delta \leq 10^{-5}$  [86]. Results are extrapolated to the TDL and CBS limits. The total projection time is typically  $\beta = 80 E_{\text{Ha}}$ , although calculations with  $\beta = 220 E_{\text{Ha}}$  are performed in some cases. The convergence error from the use of a finite  $\beta$  is negligible. Most calculations use  $\delta\beta = 0.005 E_{\text{Ha}}^{-1}$ . Extrapolations are performed where the associated Trotter error is greater than the other uncertainties. The reported QMC error bars are estimated as 1 standard deviation statistical errors. The CP bias leads to a non-variational estimator of the ground-state energy. In our calculations, the UHF ground state is taken to be  $\Psi_T$ . For the two largest bond lengths ( $R = 3.2$  and  $3.6$ ), motivated by the analogy between the H chain and the Heisenberg model, we employ in this work a new form of trial wave function, a linear combination of spinon excitations on top of the UHF state,  $|\Psi_T\rangle = \sum_{k=0}^s \sum_{i_1 < \dots < i_k} C_{i_1 \dots i_k} |\Psi_{i_1 \dots i_k}\rangle$ . Spinon excitations  $\Psi_{i_1 \dots i_k}$  are constructed using atomic positions with antiferromagnetic spin ordering, and  $k$  pairs  $(i_1, i_1 + 1) \dots (i_k, i_k + 1)$  of adjacent spins flipped, as initial conditions for the UHF self-consistence procedure, and coefficients  $C_{i_1 \dots i_k}$  are finally optimized variationally. We use  $k = 1, 2$  for  $R = 3.2, 3.6$ , respectively.

### b. Variational and lattice-regularized diffusion Monte Carlo (LR-DMC, VMC) method

The LR-DMC is a projection method [33] that uses the lattice regularization for applying the imaginary time propagator  $\exp(-\beta\hat{H})$  to a trial function  $\Psi_T$  defined in the continuous space. In this work, improvements including new use of localized basis sets are introduced, which drastically improves accuracy over previous results [10]. The main approximation is to write the Laplacian by means of its discretized expression on a lattice with a grid with lattice space  $a$ , e.g., for a single electron wave function depending only on one variable  $x$ :  $\nabla^2\Psi = \partial_x^2\Psi \rightarrow \nabla_a^2\Psi = (1/a^2)[\Psi(x+a) + \Psi(x-a) - 2\Psi(x)]$ . The extension of

$\nabla_a$  to higher dimensions is straightforward, and the continuous space can be sampled ergodically even for finite  $a > 0$  with a much simpler algorithm than the original proposal [33], namely, by randomizing the directions along which the Laplacian is discretized [87]. For  $a \rightarrow 0$  and  $\beta \rightarrow \infty$ , the exact ground-state wave function  $\Psi_G$  can be obtained if  $\langle \Psi_T | \Psi_G \rangle \neq 0$ . However, due to the fermion “sign problem” an approximation is employed to achieve small and controlled statistical errors: it is required that, during the projection, the sign of the propagated wave function is constrained to the one of the chosen trial function:

$$\Psi_T(x) \langle x | \exp(-\beta \hat{H}) | \Psi_T \rangle \geq 0, \quad (\text{A8})$$

for any electronic configuration  $x$  where the spins and the electron positions are defined. For  $a \rightarrow 0$ , the results coincide with the standard fixed-node approximation introduced a long time ago [34]. This scheme is usually employed within the diffusion short time ( $\Delta$ ) approximation of the propagator, in a way that in the  $\Delta \rightarrow 0$  limit, by applying it  $(\beta/\Delta)$  times, the exact fixed-node projection scheme Eq. (A8) is recovered, with the well-established variational property on the estimated energy. In this work, we use the lattice regularization, just because it is more conveniently implemented in the TurboRVB package [88], and also because the extrapolations for  $a \rightarrow 0$  are very well behaved and easily controlled in an automatic way.

This method is very weakly dependent on the dimension of the basis set chosen to represent the nodes of the Slater determinant, and we verify that a negligible error in the DMC energy is obtained by using the standard cc-pVTZ basis, where the largest  $Z_{1s}$  ( $Z_{1s} = 33.87$ ) is removed. Indeed, too large exponents are also not necessary in this approach because the cusp conditions are fulfilled by a one-body Jastrow factor of the type

$$u_{1\text{-body}}(r) = -\frac{1 - \exp(-\sqrt{2}r)}{\sqrt{2}}. \quad (\text{A9})$$

This one-body Jastrow is included also in the GTO basis set for the DFT (LDA) calculation. The DFT is also defined (within the TurboRVB package [88]) on a mesh of lattice space  $a = 0.1$  or smaller, until convergence is reached within  $0.001 E_{\text{Ha}}$  in the total energy. The use of the one-body Jastrow factor drastically improves the convergence for  $a \rightarrow 0$ , and the quality of the basis set as the DFT energy is much lower than the standard one in the original cc-pVTZ basis. For  $R < 1.4$ , we find that the cc-pVDZ basis can be significantly improved by using  $p$  diffusive Gaussian orbitals with small exponents ( $Z_{1p} = 0.2$ ,  $Z_{2p} = 0.05$ ), allowing us to obtain the best variational LR-DMC estimates for  $R = 1.0$  and  $R = 1.2$ , even better than the ones corresponding to the larger cc-pVTZ basis. Within periodic boundary conditions in the direction of the

chain, assumed to be along the  $z$  direction, we use a supercell of dimension  $L_x \times L_x \times L_z$ , with  $L_x = L_y = 40 a_B$ , that is large enough for safely neglecting the interaction between the periodic images in the  $x, y$  directions (error less than  $0.0001 E_{\text{Ha}}$  per atom). Moreover, the basis set (standard for open systems) has been periodized according to the standard procedure described in Ref. [89], when PBC are applied.

Before the application of the LR-DMC algorithm, a more accurate Jastrow factor is used to define the trial function  $\Psi_T$ . This contains the so-called two-, three-, and four-body contributions that are expanded in a localized basis different from the determinant one. All these terms are efficiently optimized, using the scheme described in Ref. [90]. Since the Jastrow is not affecting the results for  $a \rightarrow 0$  we present here, we do not describe in detail its form and the standard optimization methods used [91].

### 3. Embedding theories

Quantum embedding theories (DMET, SEET) are based on the idea of combining two different types of quantum calculations: high-level calculations on one or more active regions of interest, called fragments, and low-level calculations on the environment surrounding fragments. In various methods, these fragments can be chosen either in the energy or in the local basis.

A quantum embedding theory determines the coupling between fragments and environment self-consistently, using a variable of interest to provide for feedback. DMET and SEET, respectively, use the one-particle density matrix and the self-energy as variables of interest.

#### a. Density-matrix embedding theory (DMET)

DMET [20,21] provides a framework to approximate expectation values of a large system from embedded calculations. A mean-field wave function  $\Phi$  over the full system is used to define the embedding of a fragment defined in terms of a set of  $L$  local orbitals. The embedding of  $\Phi$  splits the occupied and virtual orbitals into ones with and without weight in the local fragment. The occupied and virtual set *with* weight in the fragment fully span the  $L$  local orbitals used to define the embedding, as well as an additional set of (at most  $L$ ) *bath* orbitals. In DMET, a high-level calculation is carried out in the fragment + bath orbital space; as the size of the embedded fragment  $L$  approaches that of the full system, the resulting DMET energy converges to that of a high-level calculation on the full system. While DMET can be used to study finite systems, it provides a natural framework to study systems directly in the TDL. In this work, we use different DMET strategies for calculations in finite chains and in the TDL; we refer the reader to Appendix C 2 for further details about the latter.

As outlined above, the splitting of the full system into fragments requires the introduction of a set of local orbitals. In this work, we use intrinsic atomic orbitals (IAOs) [92] to define the local basis in the valence space.



In our finite-chain calculations, we split the system into fragments of  $x$  atoms by considering the corresponding local valence orbitals. These local orbitals, through the embedding construction, generate a set of bath orbitals. To this embedding (fragment + bath) space, we further add a set of local virtual orbitals, built as projected atomic orbitals (PAOs), on the constituent atoms. We use the acronym DMET[X] for calculations using fragments of size  $X = 2, 5$ .

Expectation values in DMET (such as the energy and particle number) are computed by partial traces [93] (using the local orbitals in a given fragment) of the contraction of integrals with density matrices. A self-consistency loop can be used to uniquely define  $\Phi$  [20,93]. In this work, however, we use the RHF wave function without further optimization. A global chemical potential is used to control the total number of electrons in the system.

Our DMET results are reported using FCI as a solver for STO-6G basis calculations and larger basis calculations with fragments of size 2. Other calculations use DMRG as a solver using a bond dimension of  $D = 1000$ . The error due to the DMRG solver is expected to be significantly smaller than the error due to the fragment sizes considered.

We introduce new schemes for basis set and TDL extrapolations in this work, further details for which are provided in Appendix. C 2.

### b. Self-energy embedding theory (SEET)

The SEET [46–50] relies on the assumption that orbitals in the system can be separated into  $S$  different intersecting or nonintersecting subsets  $A_i$ , each containing  $M_i^A$  orbitals, while  $M^R$  orbitals are contained in the remainder  $R$  such that  $M_i^A \ll M$  for each  $i$ . In SEET, orbitals within subsets are strongly correlated and treated nonperturbatively; on the other hand, intersubset correlations can be treated either perturbatively or nonperturbatively. In a case of a perturbative treatment, the intersubset correlations between two different orbital sets  $A_i$  and  $A_j$ , where  $i \neq j$ , or correlations in the remainder  $R$  are assumed to be weaker. Various ways of choosing the orbital subsets are possible. New versions implemented in this work have led to much improved accuracy. We employ a selection based on the occupancies of natural orbitals (NOs) as well one based on the spatial locality of symmetrically orthogonalized atomic orbitals (SAOs) and localized molecular orbitals (LMOs). For details concerning each procedure, see Refs. [48,50].

In SEET, the solution of the whole physical system is approximated by an affordable but frequently not so accurate  $\Phi$ -derivable method suitable for treating weakly correlated systems. Subsequently, this approximation is corrected within chosen strongly correlated orbital subspaces by a nonperturbative method. We demonstrate that the general SEET functional can be written as

$$\begin{aligned} \Phi_{\text{MIX}}^{\text{SEET}} &= \Phi_{\text{weak}}^{\text{tot}} + \sum_i^{\binom{N}{k}} (\Phi_{\text{strong}}^{A_i} - \Phi_{\text{weak}}^{A_i}) \\ &\pm \sum_{k=K-1}^{k=1} \sum_i^{\binom{N}{k}} (\Phi_{\text{strong}}^{B_i^k} - \Phi_{\text{weak}}^{B_i^k}), \end{aligned} \quad (\text{A10})$$

where the contributions with  $\pm$  signs are used to account correctly for the possible double counting; for details, see Ref. [50]. In this paper,  $\Phi_{\text{weak}}^{\text{tot}}$  is determined from GF2 or HF. In general, other choices such as the *GW* method [94] are also possible.  $\Phi_i^{A_i}$  stands for all those terms in  $\Phi$  with all four indices  $i, j, k, l$  of two-body interactions  $v_{ijkl}$  contained inside orbital subspace  $A_i$ . Here,  $\Phi_{\text{weak}}^{A_i}$  is the solution for subset  $A_i$  within the weak-coupling method, here GF2.  $\Phi_{\text{strong}}^{A_i}$  is the solution in the  $M_i^A$  subspace evaluated using a higher-order method suitable for treating “strong correlation.” We denote this way of performing SEET calculations as SEET(method strong/method weak)- $m([M^A \text{o}]/\text{basis})$  since here self-energies from intersecting orbital subspaces with  $M^A$  strongly correlated orbitals are “mixed” between each other. While in a general case the self-energy has to be evaluated for  $\binom{M}{M^A}$  orbital subgroups, where  $\binom{M}{M^A}$  can be a fairly large number, in practice one can quite trivially reduce it by identifying the most important subgroups containing  $M^A$  orbitals that lead to the significant lowering of the ground-state energy; see Ref. [50].

Since calculations can be performed in either the energy or spatial basis employing NOs or SAOs and LMOs, respectively, as basis functions, we denote these choices using the “basis” keyword, where  $\text{tbasis} = \text{NO}, \text{SAO}, \text{or LMO}$ .

In SEET, the self-energy is constructed as a functional derivative of the  $\Phi^{\text{SEET}}$  functional and the total SEET self-energy contains diagrams from both the strong- and weak-coupling methods.

Consequently, each strongly correlated subspace self-energy is embedded into a weakly correlated self-energy generated by all orbitals outside the strongly correlated subspace and accounting for all the nonlocal interactions on the strongly correlated orbital groups [95]. For details explaining how to evaluate  $\Sigma_{\text{strong}}^{A_i}$  and  $\Sigma_{\text{weak}}^{A_i}$ , see Refs. [46,49].

We converge the electronic energy to  $10^{-4} E_{\text{Ha}}$ . The inverse temperature  $\beta$  is set at 100 or 200  $E_{\text{Ha}}^{-1}$  depending on the geometry. The Matsubara frequency grid is generated using the spline’s interpolation [96] with the maximum number of points varying between 20 000 and 50 000. In this paper, for the weakly correlated method in SEET, we use GF2 as well as HF. The strongly correlated part of the SEET self-energy is evaluated from the Anderson impurity model using FCI or versions of restricted active space CI (RASCI). Note that in all the calculations presented here, SEET is based on the RHF reference.

## 4. Diagrammatic methods

The diagrammatic methods discussed in this work (SC-*GW*, GF2, BDMC) evaluate, either deterministically

(GF2, SC-GW) or stochastically (BDMC), a subset of the terms in a diagrammatic interaction expansion. The methods are distinguished by the subsets of diagrams or series terms that are included in the calculation. The diagrammatic methods we discuss in this work are based on the Feynman diagrammatic technique formulated in terms of self-consistent propagators and bare or renormalized interactions [97]. They are formulated at finite temperature but evaluated at low enough temperature that the  $T \rightarrow 0$  limit can be taken. Many of the methods are implemented specifically for the present study. The BDMC<sub>n</sub> calculations are the first attempt to use high-order skeleton diagrams in materials to systematically improve *GW*.

### a. Self-consistent second-order Green's function theory (GF2)

The fully self-consistent second-order Green's function theory (GF2) [27–32] includes all second-order skeleton diagrams dressed with the renormalized second-order propagators and bare interactions. GF2 is formulated as a low-level approximation to the exact Luttinger-Ward (LW) functional [43,98] and therefore is  $\Phi$  derivable, thermodynamically consistent, and conserving [43,99]. We solve all the nonlinear equations self-consistently at nonzero temperature using on an imaginary-time mesh [96,100]. At each iteration, the self-energy, Green's function, and Fock matrix are updated until convergence is reached, so that the converged solution is reference independent.

In the weakly correlated regime, GF2 preserves the advantages of the second-order Møller-Plesset perturbation theory (MP2), while at the same time avoiding the divergences appearing in non-self-consistent zero-temperature formulations of finite-order perturbation theories.

We converge the electronic energy to the threshold of  $10^{-6} E_{\text{Ha}}$ . The inverse temperature  $\beta$  is set at 100 or  $200 E_{\text{Ha}}^{-1}$  depending on the geometry. The Matsubara frequency grid is generated using splines interpolation [96] with between 20 000 and 50 000 points on the Matsubara axis. The calculations we present here are based on a RHF or UHF reference. We use GF2 to denote the version that is based on RHF and does not allow for spin symmetry breaking. The acronym UGF2 is used to denote a spin-unrestricted version based on UHF.

### b. Diagrammatic methods with renormalized interactions

The BDMC and SC-GW methods are diagrammatic approximations formulated in terms of renormalized propagators  $G$  and renormalized (“screened”) interactions  $W$ . They can be written as approximations to the LW functional  $\Phi$  [43,98], which implies that they are thermodynamically consistent and conserving [43,99]. The methods require the self-consistent determination of propagators  $G$ , self-energies  $\Sigma$ , screened interactions  $W$ , and polarizations  $P$ . While the expressions for  $\Sigma$  and  $P$  are different in the individual methods, the Dyson equations,

$$G = G_0 - G_0 \Sigma G, \quad W = V + VPW, \quad (\text{A11})$$

determine  $G$  and  $W$ , where  $G_0$  and  $V$  are the bare electronic propagator and interaction.

### c. Self-consistent GW (SC-GW)

The SC-GW approximation truncates the skeleton sequence at the lowest-order graph, so that only the first-order contribution in the renormalized interaction is considered and the second-order exchange diagram is neglected. We implement a deterministic procedure of this approximation closely following Refs. [27,43–45]: The Green's function is initialized using the HF approximation result. We then construct the polarization  $P = GG$  and obtain  $W$  from Eq. (A11). After computing the GW self-energy  $\Sigma = -GW$ , we obtain the updated  $G$  by solving Dyson's equation, thus closing the self-consistency loop. The method is formulated in a grand canonical ensemble, but the chemical potential  $\mu$  used in each step is chosen to preserve the desired electron number. After convergence, the total energy is computed from  $\Sigma$  and  $G$ .

While SC-GW benefits from the conservation of the average particle number, energy, momentum, and angular momentum, the size and complexity of  $W$  call for appropriate controlled simplifications. Instead of introducing physically motivated approximations that may not respect the conserving properties, we perform systematic linear algebra decompositions and truncations on  $V$ ,  $W$ , and  $G$ , which vastly reduce the numerical effort [101]. We converge our calculations to a relative precision of  $10^{-7}$ . Convergence is reached at inverse temperature  $\beta = 100 E_{\text{Ha}}^{-1}$ , and for about 8000 Matsubara frequencies. A combination of power and uniform meshes, adaptive grids, and spline grids is used for imaginary time and Matsubara data [96]. The code is based on the ALPS libraries [102] and uses integrals generated by LIBINT [103].

### d. Treatment of zero terms in the Hamiltonian

We note an important ambiguity in formulating certain diagrammatic approximations: Hamiltonian terms that are identically zero because of the Pauli principle, and thus have no observable effect on the physical properties of the system, can be arbitrarily added to a Hamiltonian. While these terms will evaluate to zero in the exact solution, they may evaluate to nonzero values in approximations that do not consider all terms at a given order. The *GW* approximation is such an approximation, whereas the GF2 approximation does not suffer from this problem.

To illustrate the point, consider an ideal spin-polarized lattice Fermi gas and add a contact interaction term:

$$\hat{H}_0 = -\sum_{ij} t_{ij} \hat{\psi}_{j\uparrow}^\dagger \hat{\psi}_{i\uparrow} + U \sum_i \hat{\psi}_{i\uparrow}^\dagger \hat{\psi}_{i\uparrow} \hat{\psi}_{i\uparrow} \hat{\psi}_{i\uparrow}. \quad (\text{A12})$$

The system remains noninteracting, and, correspondingly, in the diagrammatic expansion based on the bare

interaction vertex  $U$ , all diagrams of the same order cancel each other exactly. However, low-order self-consistent theories like  $GW$  may break this cancellation by including some, but not all, higher-order contributions in  $U$ . As a result, a low-order self-consistent method would produce different answers for Eq. (A12) with  $U = 0$  and  $U \neq 0$ .

When the interaction Hamiltonian is projected on the orbital basis, similar considerations apply to all terms that create or delete two electrons in the same state, meaning that one has a choice of keeping or dropping Hamiltonian terms based on matrix elements  $v_{abad}$  and  $v_{bada}$ ; in what follows, we call them “zero terms.” In calculations with realistic Hamiltonians, these terms are usually kept, and we refer to this choice as Hamiltonian  $\hat{H}$ . In contrast, the lattice model Hamiltonian community usually omits zero terms explicitly (or nullifies the corresponding matrix elements); we refer to this choice as Hamiltonian  $\hat{H}'$ . In an exact solution of the problem, all physical properties of  $\hat{H}$  and  $\hat{H}'$  are identical.

One question we answer in this work is how the SC-GW results depend on the Hamiltonian representation with respect to zero terms. The two curves labeled as SC-GW( $\hat{H}$ ) and SC-GW( $\hat{H}'$ ) in Fig. 8 correspond to the outcomes of the SC-GW method when it is applied to  $\hat{H}$  and  $\hat{H}'$ , respectively. The effect of zero terms is profound. While the SC-GW( $\hat{H}$ ) curve at higher energy is more accurate for separations up to the equilibrium distance, it becomes less accurate than the SC-GW( $\hat{H}'$ ) curve at  $R > 2.4a_B$ , and SC-GW( $\hat{H}'$ ) appears to produce more consistent energy differences at the large separation range.

Given that the two SC-GW answers surround the variational estimate, the difference between them can be used as a crude estimate of the accuracy of the SC-GW approximation. This point of view is confirmed by our study of vertex corrections. When the second-order vertex corrections are accounted for, the result for Hamiltonian  $\hat{H}'$  shifts upwards by an amount comparable to the difference between the SC-GW( $\hat{H}$ ) and SC-GW( $\hat{H}'$ ) curves. The BDMC result starts converging to the best variational estimate when higher-order corrections are included, as shown in Fig. 8.

### e. Bold diagrammatic Monte Carlo (BDMC)

We also develop a stochastic implementation of the  $G^2W$  formalism that is able to go beyond the lowest-order diagrams. Within the bold diagrammatic Monte Carlo framework (see, e.g., Refs. [17–19]), the configuration space of skeleton diagrams for  $\Phi$  is sampled stochastically starting from vertex corrections to SC-GW. The method can be applied to any system at nonzero temperature with arbitrary dispersion relation (doped and undoped) and with arbitrary shape of the interaction potential [17–19,104–106]. Both  $\Sigma$  and  $P$  are computed as sums of skeleton graphs, up to order  $n$ ; we denote these sums as  $\Sigma_n$  and  $P_n$  and abbreviate the corresponding level of approximation as BDMC $_n$ . The lowest-order contributions to  $\Sigma_1$  and  $P_1$  are based on products of the  $G$  and  $W$  functions

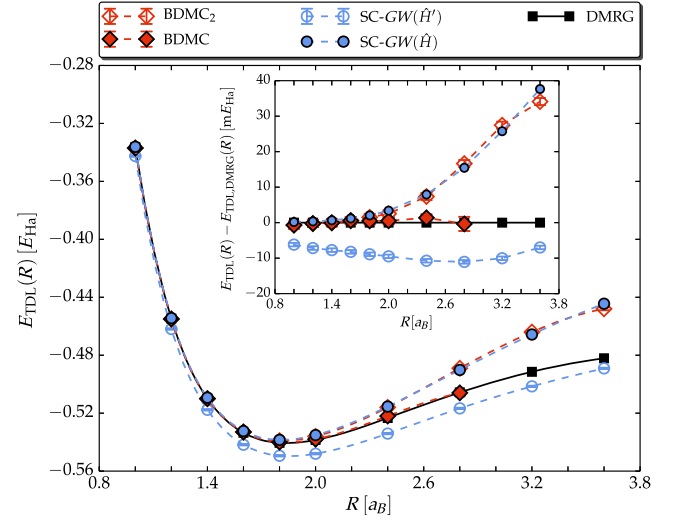


FIG. 8. Diagrammatic equation of state in the TDL, at STO-6G level. Two SC-GW curves underline the ambiguity of formulating the Hamiltonian terms that are identically zero because of the Pauli principle (see text). The SC-GW( $\hat{H}$ ) curve corresponds to the protocol usually used for realistic *ab initio* Hamiltonians, while the SC-GW( $\hat{H}'$ ) curve corresponds to the one used in the lattice model Hamiltonian community. By accounting for higher-order vertex corrections, the BDMC $_n$  results are observed to converge towards the estimates from other methods.

mentioned above, and BDMC $_1$  is identical to SC-GW. To obtain final answers we either perform an extrapolation to the  $n \rightarrow \infty$  limit or observe good convergence with increasing the diagram order. (This was demonstrated for several Coulomb systems in Refs. [19,107]).

In the orbital representation, each interaction line depends on four site or atom indices  $\{(i, j); (k, l)\}$ , four orbital indices  $\{(\alpha, \beta); (\gamma, \delta)\}$ , and two spin indices  $\{\sigma; \sigma'\}$  (the Coulomb interaction vertex does not change spin); to simplify notations we also use a composite index  $a = (i, \alpha, \sigma)$ . It is worth mentioning that terms with  $u = v = 0$ , where  $u = i - j$  and  $v = k - l$ , are relative “distances” between the orbitals, represent the “density-density” part of the interaction potential, and their contribution is dominating in the final answer. Accounting for nonzero values of  $(u, v)$  in the Dyson equation for  $W$  changes the answer at the subpercent level, and the corresponding contribution quickly saturates when separation in space between orbital indices, limited by cutoffs  $u^*$  and  $v^*$ , is increased. For hydrogen atoms in the single-orbital case, we find that energies per atom obtained with unrestricted summation over  $(u, v)$  and with  $u^* = v^* = 2$  coincide at the level of  $\sim 10^{-5}$  in relative units even at the smallest values of lattice constant  $R$  considered in this work (the agreement is better at larger values of  $R$ ).

## APPENDIX B: TABLES OF RESULTS AND ADDITIONAL BENCHMARK DATA

In Tables II–VI, we include the numerical values of the results presented in some of the figures in the main

TABLE II. Potential energy curve of H<sub>10</sub> with the minimal (STO-6G) basis. DMET[5], MRCI, and MRCI + *Q* energies coincide with FCI to within 10<sup>-6</sup>.

<i>R</i>	AFQMC	DMET[2]	FCI	GF2	SC-GW	SEET (CI/HF)-m (4,SAO)	SEET (CI/HF)-m (6,SAO)	SEET (CI/GF2)-m (6,SAO)	RCCSD	RCCSD(T)	UCCSD	UCCSD(T)	RHF	UHF
1.0	-0.38248(3)	-0.381351	-0.382439	-0.381012	-0.382996	-0.3817	-0.3819	-0.3822	-0.382387	-0.382432	-0.382387	-0.382432	-0.375174	-0.375174
1.2	-0.47667(3)	-0.475497	-0.476638	-0.474570	-0.476870	-0.4758	-0.4760	-0.4761	-0.476561	-0.476626	-0.476561	-0.476626	-0.467805	-0.467806
1.4	-0.52051(5)	-0.519307	-0.520509	-0.517622	-0.520234	-0.5196	-0.5199	-0.5200	-0.520401	-0.520492	-0.520401	-0.520492	-0.509862	-0.509862
1.6	-0.53819(6)	-0.537170	-0.538436	-0.534488	-0.537403	-0.5374	-0.5378	-0.5379	-0.538292	-0.538414	-0.538292	-0.538414	-0.525628	-0.525628
1.8	-0.54218(6)	-0.541097	-0.542439	-0.537111	-0.540312	-0.5413	-0.5418	-0.5418	-0.542254	-0.542418	-0.541742	-0.542175	-0.527014	-0.527745
2.0	-0.53874(5)	-0.537527	-0.538963	-0.531852	-0.535304	-0.5376	-0.5383	-0.5383	-0.538745	-0.538966	-0.537504	-0.538049	-0.520347	-0.523137
2.4	-0.52271(8)	-0.521125	-0.522794	-0.510567	-0.514336	-0.5212	-0.5221	-0.5221	-0.522681	-0.523088	-0.520417	-0.520753	-0.495606	-0.506571
2.8	-0.50507(5)	-0.503220	-0.505024	-0.485323	-0.488955	-0.5034	-0.5044	-0.5045	-0.506729	-0.507486	-0.502820	-0.502959	-0.465843	-0.491183
3.2	-0.49114(2)	-0.489450	-0.491038	-0.461886	-0.464687	-0.4897	-0.4906	-0.4908	-0.590076	-0.592683	-0.489482	-0.489521	-0.436679	-0.481323
3.6	-0.48200(2)	-0.480818	-0.481870	-0.442549	-0.443705	-0.4808	-0.4816	-0.4819	-0.664459	-0.668234	-0.480951	-0.480959	-0.410493	-0.476035

TABLE III. Potential energy curve of H<sub>10</sub> extrapolated to the CBS limit. The MRCI + *Q* value at *R* = 1.0 uses the AFQMC energy with a correction estimated from the difference between MRCI + *Q* and AFQMC energies at *R* = 1.2. RCC breaks down at *R* = 3.6 and, for large basis sets and the shortest bond lengths, DMET[2], PC-NEVPT2, and SC-NEVPT2 are nonconvergent due to linear dependency issues.

<i>R</i>	AFQMC	DMET[2]	LR-DMC(AGP)	LR-DMC(LDA)	MRCI + <i>Q</i>	MRCI + <i>Q</i> + F12	PC-NEVPT2	SC-NEVPT2
1.0	-0.44284(24)	...	-0.442430(7)	-0.442093(9)	-0.44324(32)	-0.44301(31)	...	...
1.2	-0.51489(14)	...	-0.514432(6)	-0.514418(7)	-0.51529(15)	-0.51506(12)	...	...
1.4	-0.54914(11)	-0.54905(28)	-0.548742(6)	-0.548748(7)	-0.54926(15)	-0.54917(12)	-0.54593(3)	-0.54569(3)
1.6	-0.56315(12)	-0.56258(30)	-0.562878(6)	-0.562900(7)	-0.56324(8)	-0.56321(7)	-0.56011(1)	-0.55991(1)
1.8	-0.56644(4)	-0.56567(22)	-0.566234(6)	-0.566255(7)	-0.56655(5)	-0.56654(5)	-0.56350(5)	-0.56333(6)
2.0	-0.56396(6)	-0.56321(12)	-0.563831(5)	-0.563852(7)	-0.56411(3)	-0.56410(3)	-0.56113(7)	-0.56099(8)
2.4	-0.55164(4)	-0.55124(2)	-0.551679(6)	-0.551699(8)	-0.55189(1)	-0.55187(1)	-0.54911(8)	-0.54902(9)
2.8	-0.53755(8)	-0.53726(2)	-0.537423(6)	-0.537424(9)	-0.53754(3)	-0.53753(3)	-0.53510(9)	-0.53504(9)
3.2	-0.52499(7)	-0.52512(6)	-0.524987(7)	-0.524974(9)	-0.52499(6)	-0.52500(5)	-0.52302(9)	-0.52299(9)
3.6	-0.51568(9)	-0.51612(9)	-0.515571(7)	-0.515582(11)	-0.51549(8)	-0.51553(5)	-0.51405(8)	-0.51403(8)

<i>R</i>	SBDMRG	RCCSD	RCCSD(T)	UCCSD	UCCSD(T)	UHF	VMC(AGP)	VMC(LDA)
1.0	-0.44209(21)	-0.44207(26)	-0.44268(21)	-0.44207(26)	-0.44268(21)	-4.15363(16)	-0.441499(9)	-0.44061(1)
1.2	-0.51399(15)	-0.51432(19)	-0.51490(19)	-0.51432(19)	-0.51490(19)	-4.88368(5)	-0.513646(9)	-0.513225(8)
1.4	-0.54859(15)	-0.54846(14)	-0.54907(14)	-0.54846(14)	-0.54907(14)	-5.22845(4)	-0.548046(7)	-0.547650(8)
1.6	-0.56311(14)	-0.56240(10)	-0.56311(9)	-0.56240(10)	-0.56311(9)	-5.36820(3)	-0.562237(5)	-0.561823(7)
1.8	-0.56666(14)	-0.56565(4)	-0.56638(4)	-0.56533(4)	-0.56623(4)	-5.40150(2)	-0.565634(5)	-0.565201(6)
2.0	-0.56428(14)	-0.56311(1)	-0.56391(1)	-0.56239(1)	-0.56341(1)	-5.38220(2)	-0.563249(4)	-0.562776(6)
2.4	-0.55196(14)	-0.55071(2)	-0.55167(3)	-0.54953(2)	-0.55037(2)	-5.281261(6)	-0.551106(4)	-0.550575(5)
2.8	-0.53765(14)	-0.53632(5)	-0.53749(6)	-0.53519(5)	-0.53577(6)	-5.174240(5)	-0.536832(3)	-0.536201(5)
3.2	-0.52511(14)	-0.52419(8)	-0.52574(10)	-0.52303(8)	-0.52340(8)	-5.096702(6)	-0.524373(4)	-0.523557(5)
3.6	-0.51568(14)	...	...	-0.51407(10)	-0.51428(10)	-5.050387(7)	-0.514968(3)	-0.513869(6)

TABLE IV. Potential energy curve of H<sub>10</sub> at cc-pVxZ, from the MRCI + *Q* method. For large basis sets and the shortest bond lengths, calculations are nonconvergent due to linear dependency issues. Additional calculations are performed with F12 corrections.

<i>R</i>	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z
1.0	-0.421954	...	...	...
1.2	-0.502812	-0.513118	...	...
1.4	-0.540966	-0.547756	-0.548548	...
1.6	-0.557110	-0.561985	-0.562710	-0.562932
1.8	-0.561486	-0.565330	-0.566068	-0.566277
2.0	-0.559552	-0.562892	-0.563645	-0.563852
2.4	-0.547661	-0.550663	-0.551426	-0.551651
2.8	-0.533470	-0.536307	-0.537099	-0.537349
3.2	-0.521270	-0.523817	-0.524586	-0.524858
3.6	-0.512397	-0.514463	-0.515150	-0.515425

TABLE V. Equation of state of the hydrogen chain at the TDL, computed with the minimal basis (STO-6G).

$R$	AFQMC	BDMC	DMET[5]	DMET[ $\infty$ ]	DMRG	SC-GW	SEET (CI/GF2)-m (6,SAO)	SEET (CI/HF)-m (6,SAO)	RCCSD	RCCSD(T)	UCCSD	UCCSD(T)	UHF
1.0	-0.33631(5)	-0.337(1)	-0.337 30	-0.336 21	-0.336 31	-0.336 14	-0.3353	-0.3354	-0.335 96	-0.336 25	-0.335 96	-0.336 25	-0.327 42
1.2	-0.45466(5)	-0.455(1)	-0.455 69	-0.454 70	-0.454 74	-0.454 39	-0.4537	-0.4536	-0.454 28	-0.454 64	-0.454 37	-0.454 37	-0.445 04
1.4	-0.50999(4)	-0.510(1)	-0.510 81	-0.509 89	-0.509 96	-0.509 30	-0.5088	-0.5088	-0.509 37	-0.509 81	-0.509 06	-0.509 39	-0.499 58
1.6	-0.53356(5)	-0.533(1)	-0.534 28	-0.533 42	-0.533 56	-0.532 38	-0.5323	-0.5323	-0.532 86	-0.533 37	-0.532 78	-0.533 10	-0.522 18
1.8	-0.54048(4)	-0.540(1)	-0.541 04	-0.540 34	-0.540 49	-0.538 44	-0.5391	-0.5390	-0.539 67	-0.540 26	-0.539 62	-0.540 01	-0.528 06
2.0	-0.53846(7)	-0.538(1)	-0.538 88	-0.538 40	-0.538 50	-0.535 12	-0.5369	-0.5371	-0.537 60	-0.538 33	-0.537 40	-0.537 80	-0.525 17
2.4	-0.52326(4)	-0.522(1)	-0.523 50	-0.523 35	-0.523 36	-0.515 44	-0.5219	-0.5218	...	...	-0.521 82	-0.522 11	-0.509 20
2.8	-0.50556(5)	-0.506(2)	-0.505 78	-0.505 64	-0.505 64	-0.490 18	-0.5050	-0.5043	...	...	-0.504 06	-0.504 20	-0.493 05
3.2	-0.49145(4)	...	-0.491 66	-0.491 45	-0.491 46	-0.465 70	-0.4920	-0.4904	...	...	-0.490 26	-0.490 30	-0.482 38
3.6	-0.48210(5)	...	-0.482 40	-0.482 10	-0.482 11	-0.444 46	-0.4836	-0.4816	...	...	-0.481 37	-0.481 38	-0.476 57

TABLE VI. Final equation of state for the hydrogen chain at the thermodynamic limit. LR-DMC results for  $R = 1.0$  and  $1.2$  are  $-0.4198(3)$  and  $-0.4987(6)$ , respectively.

$R$	AFQMC	AFQMC+ $\Delta_{\text{DMRG}}$	DMET[2]	LR-DMC (LDA)	SC-NEVPT2	RCCSD	RCCSD(T)	UCCSD	RHF	UHF	VMC(LDA)
1.4	-0.54044(35)	-0.54022(35)	-0.53858(27)	-0.53971(6)	-0.53674(7)	-0.53895(19)	-0.54009(19)	-0.53897(22)	-0.51133(3)	-0.51234(3)	-0.53820(1)
1.6	-0.55971(36)	-0.55959(37)	-0.55730(34)	-0.55912(12)	-0.55608(11)	-0.55821(13)	-0.55942(14)	-0.55778(15)	-0.53097(2)	-0.53256(2)	-0.55767(7)
1.8	-0.56569(30)	-0.56562(31)	-0.56312(26)	-0.56515(3)	-0.56220(4)	-0.56401(6)	-0.56529(5)	-0.56354(7)	-0.53685(2)	-0.53925(1)	-0.56346(1)
2.0	-0.56444(34)	-0.56451(34)	-0.56213(14)	-0.56397(4)	-0.56117(8)	-0.56272(1)	-0.56406(3)	-0.56238(2)	-0.53535(1)	-0.53894(1)	-0.56216(1)
2.4	-0.55313(31)	-0.55291(31)	-0.55135(1)	-0.55268(4)	-0.54999(9)	-0.55107(1)	-0.55257(3)	-0.55105(2)	-0.52235(1)	-0.52978(1)	-0.55004(1)
2.8	-0.53886(29)	-0.53870(30)	-0.53768(1)	-0.53848(5)	-0.53594(10)	...	...	-0.53666(7)	-0.50495(1)	-0.51883(1)	-0.53482(1)
3.2	-0.52557(23)	-0.52549(23)	-0.52566(5)	-0.52585(5)	-0.52362(10)	...	...	-0.52403(12)	-0.48754(1)	-0.51058(1)	-0.52250(1)
3.6	-0.51611(22)	-0.51594(22)	-0.51685(8)	-0.51608(9)	-0.51443(9)	...	...	-0.51462(14)	-0.47157(4)	-0.50556(1)	-0.51273(1)

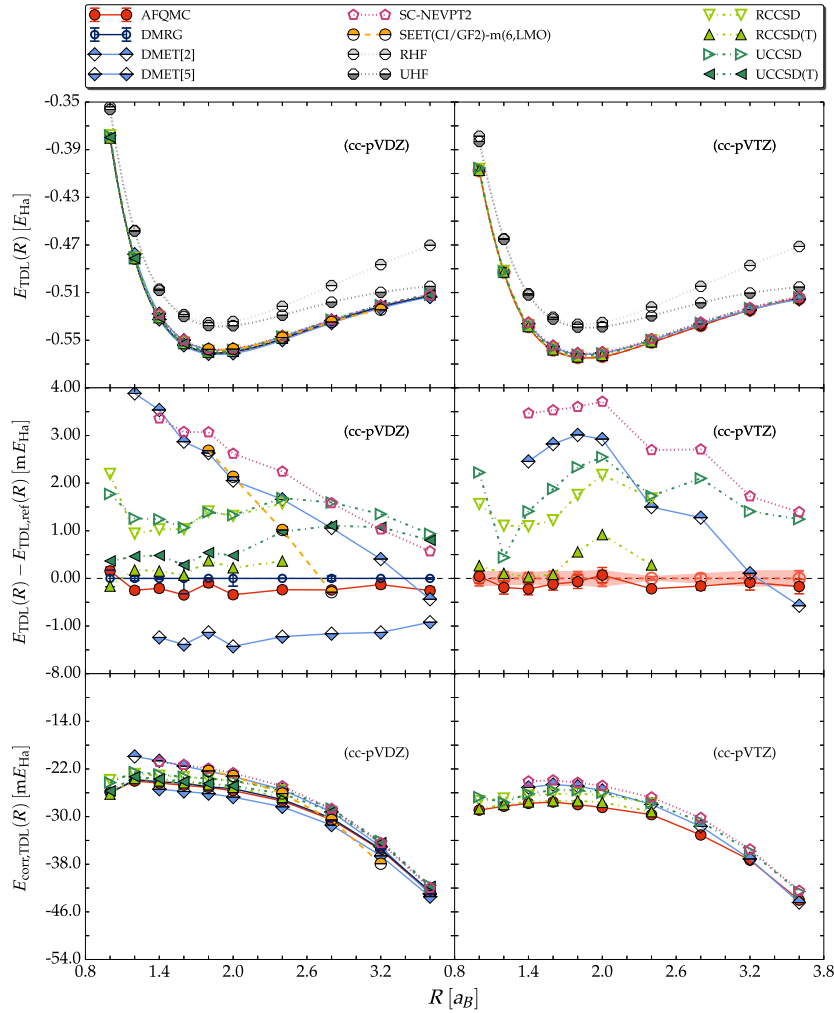


FIG. 9. Top: EOS in the thermodynamic limit computed with finite basis sets (cc-pVDZ and cc-pVTZ). Middle: Detailed comparison using DMRG and AFQMC as reference. Error bars on the DMRG data indicate estimates of the TDL extrapolation uncertainties based on the results at the STO-6G level in Fig. 15. AFQMC +  $\Delta_{\text{DMRG}_{\text{DZ}}}$  is shown as reference for TZ (empty red circle), where the correction is obtained from the energy difference between DMRG and AFQMC at DZ. Bottom: Correlation energy per particle in the TDL, at cc-pVDZ (left) and cc-pVTZ (right) level.

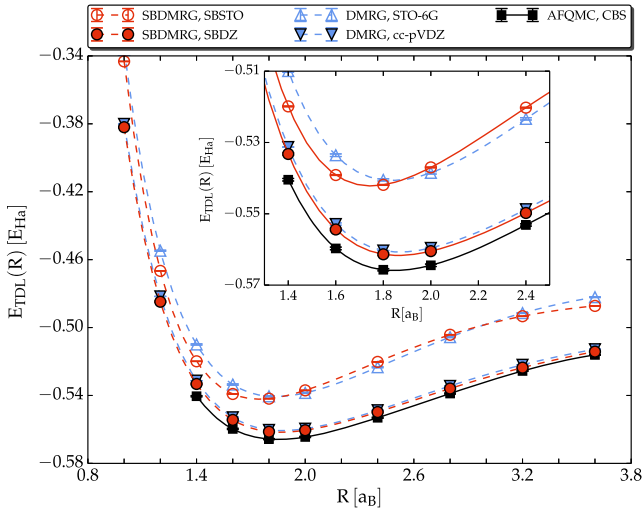
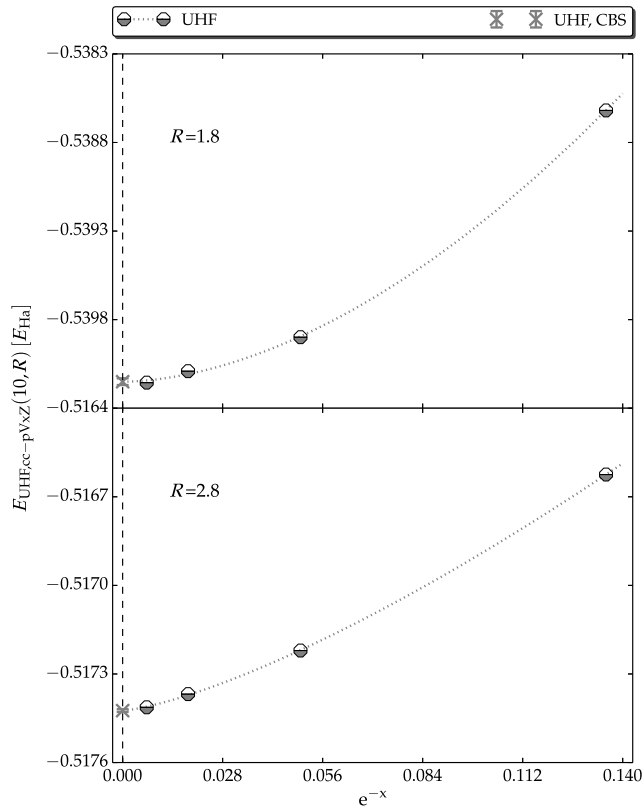


FIG. 10. Equation of state in the thermodynamic limit with select finite basis sets, by DMRG and SBDMRG. AFQMC results at the CBS limit are also shown for reference.

text, as well as additional finite-basis and finite-size data. Lengths are measured in Bohr and energies in Hartree. Data not included in the appendixes are available online [11].



A variety of finite-size and/or finite-basis-set data are available. Figure 9 shows the equation of state extrapolated to the TDL at cc-pVDZ and cc-pVTZ level. At cc-pVDZ level, DMRG provides a highly accurate EOS, with equilibrium bond length and energy  $R_{\text{eq}} = 1.880(2) a_B$ ,  $E_0 = -0.5608(2) E_{\text{Ha}}$ . The corresponding correlation energies, using RHF energies as reference, are shown in the middle panel of Fig. 9. Figure 10 shows the EOS extrapolated to the TDL using SBDMRG and DMRG for STO-6G and cc-pVDZ level basis sets.

## APPENDIX C: ADDITIONAL DETAILS ON REACHING THE COMPLETE BASIS SET AND THERMODYNAMIC LIMITS

### 1. Extrapolation to the CBS limit

Extrapolations of the UHF and UHF-based correlation energy to the CBS limit are illustrated in Fig. 11, for the representative bond lengths  $R = 1.8, 2.8 a_B$ . Using RHF as references gives indistinguishable results.

Figure 12 shows the effect of the F12 correction on MRCI+ $Q$  energies. As illustrated in the main text, extrapolations to the CBS limit obtained with and without F12 correction agree with each other to within the fitting

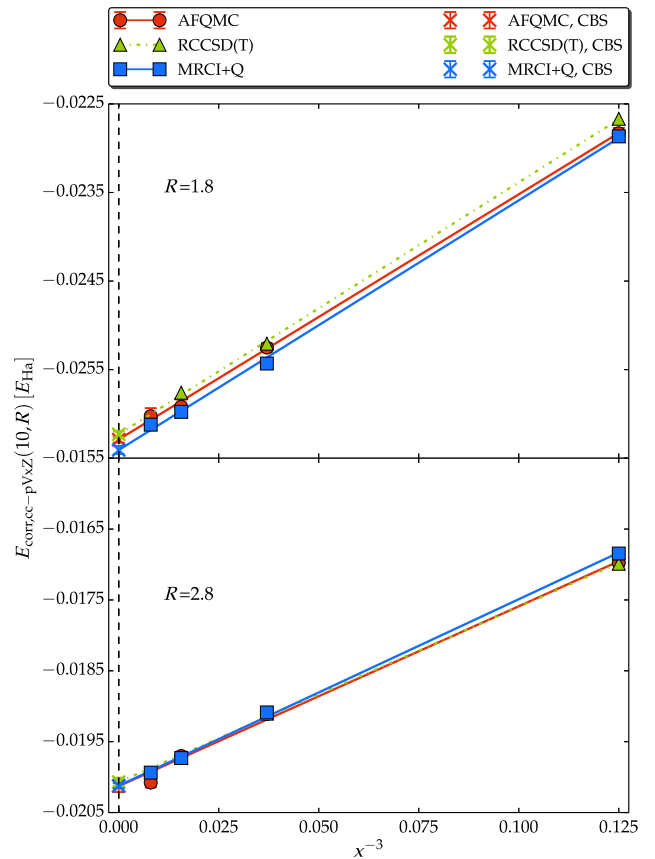


FIG. 11. Illustration of the extrapolation to the CBS limit. Results are shown for  $H_{10}$  at bond lengths,  $R = 1.8$  and  $2.8 a_B$ . The unrestricted Hartree-Fock energy is fitted to an exponential function of the index  $x = 2, 3, 4, 5$  of the cc-pV $x$ Z basis, the correlation energy to the power law  $\alpha + \beta x^{-3}$ .

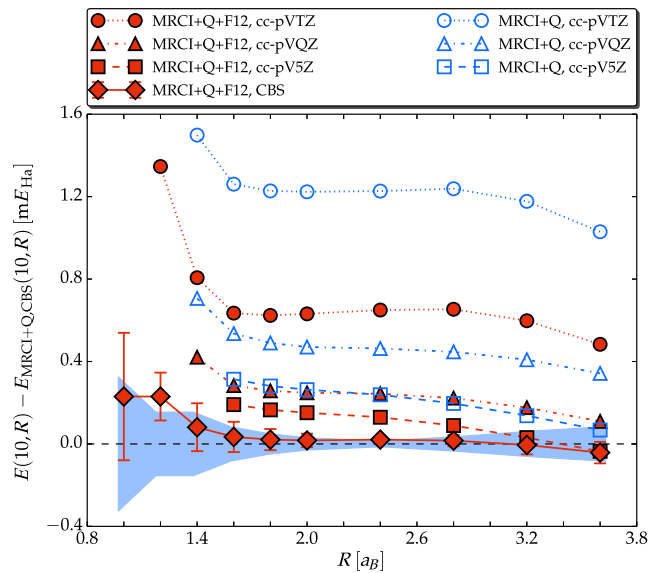


FIG. 12. Effect of the F12 correction on MRCI+ $Q$  data. A cc-pV $x$ Z + F12 calculation yields an energy of cc-pV $(x+1)$ Z quality. Extrapolations to the CBS agree with each other within uncertainties from the fitting procedure.

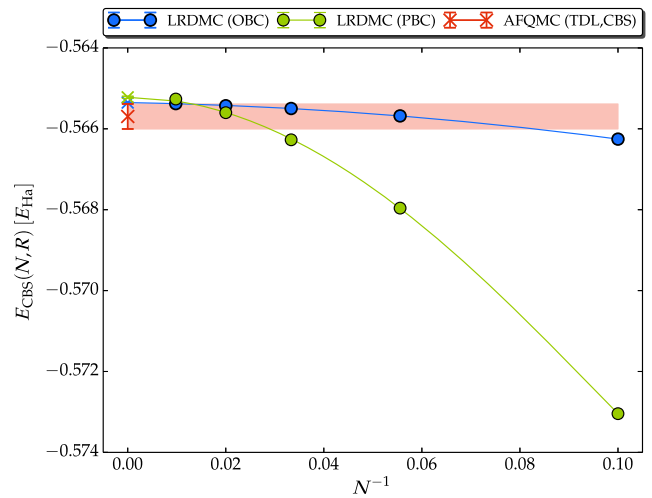


FIG. 13. Extrapolation to the thermodynamic limit using periodic (PBC) versus open (OBC) boundary conditions. Results are shown from LR-DMC for  $R = 1.8 a_B$ . Both sets of calculations converge to the same result for  $N \rightarrow \infty$ , but convergence is more rapid with OBC. The final result for  $R = 1.8 a_B$  from AFQMC (from Fig. 6) is also indicated for reference.

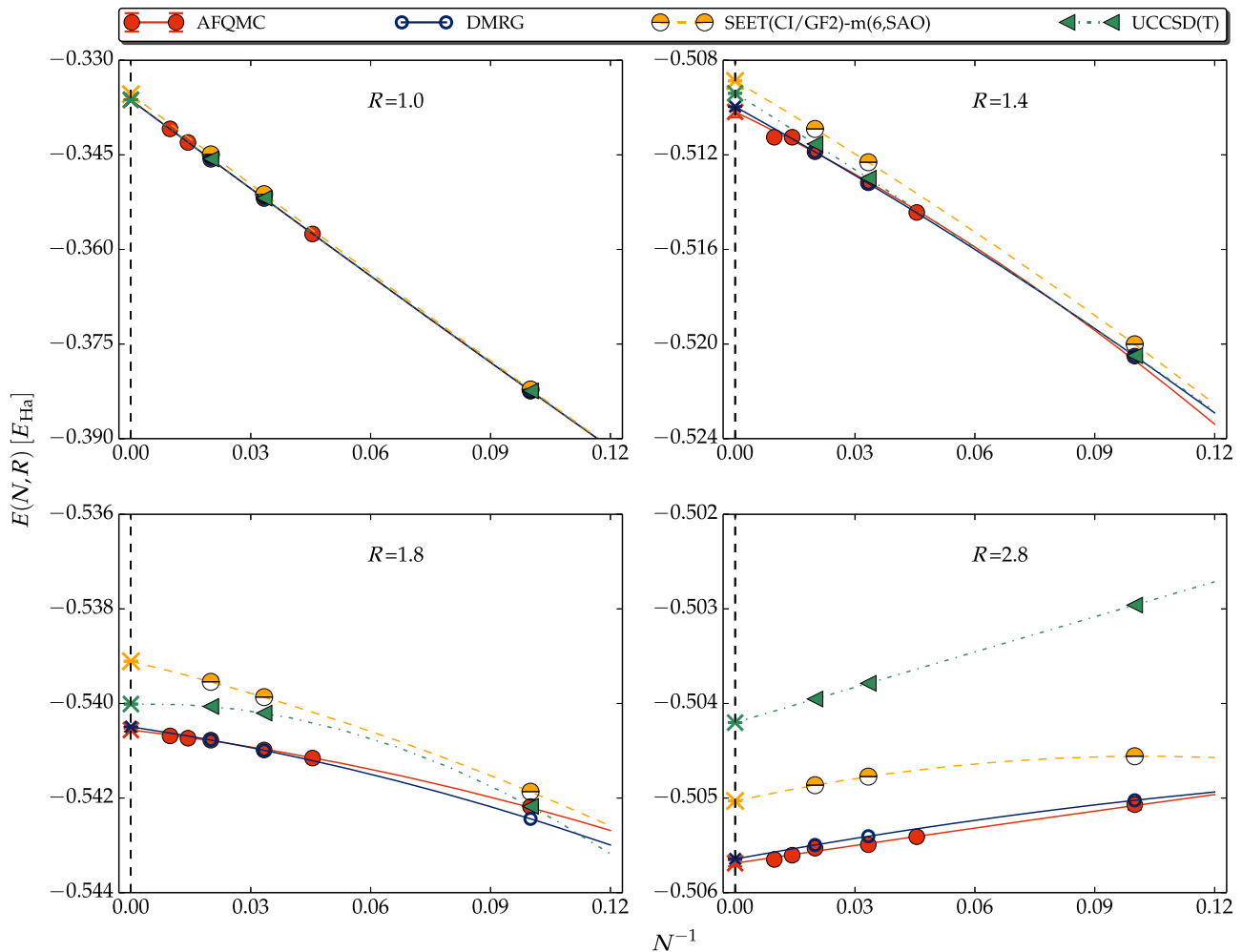


FIG. 14. Extrapolation of the EOS to the TDL limit, in the minimal STO-6G basis. The energy per particle is fitted to a second-order polynomial  $A_0 + A_1 N^{-1} + A_2 N^{-2}$  in  $N^{-1}$ , reflecting the presence of a bulk and a boundary contribution.

uncertainties, confirming the robustness of the extrapolation procedure.

## 2. Extrapolation to the TDL

As mentioned, we find that the use of finite clusters (chains) tends to give better convergence to the TDL than with periodic boundary conditions, except for very short bond lengths. This is illustrated in Fig. 13. The faster convergence with OBC than PBC is somewhat surprising and counter to commonly held belief. The quasi-one-dimensional nature of the hydrogen chain is likely an important factor.

Extrapolations to the TDL are illustrated in the minimal basis in Fig. 14 for the bond lengths  $R = 1.0, 1.4, 1.8, 2.8$ . The importance of the  $A_2(R)N^{-2}$  correction is evident in capturing the size effects. Note that the finite-size effects are larger at shorter bond lengths (ranging from roughly  $50 mE_{\text{Ha}}$  for  $R = 1.0 a_B$  to less than  $1 mE_{\text{Ha}}$  at large separation as the chain turns into a collection of uncoupled

H atoms). This is expected from the nature of the long-range Coulomb interaction.

In Fig. 15, we examine the robustness of the extrapolation. Results from the simple subtraction trick separating surface and bulk [i.e.,  $k = 1$  in Eq. (7)] are compared with the reference extrapolation using  $E_{N_1, N_2, N_3}(R)$ . Extrapolations agree with each other to well within  $1 mE_{\text{Ha}}$ , and approach  $E_{N_1, N_2, N_3}(R)$  as  $N_1, N_2$  are increased. These results suggest that extrapolations  $E_{N_1, N_2, N_3}(R)$  to the TDL have a resolution of the order of  $0.1 mE_{\text{Ha}}$  per particle. For most methodologies, therefore, the uncertainty on the TDL extrapolation is 1 order of magnitude smaller than the bias due to the underlying approximations. It is also well within the uncertainty bound in our final best estimate of the EOS.

Our DMET strategy for calculations that directly access the TDL is somewhat different. We start from a RHF calculation on a large system ( $\text{H}_{150}, \text{H}_{300}, \dots$ ). A single fragment around the central H atom is constructed, using

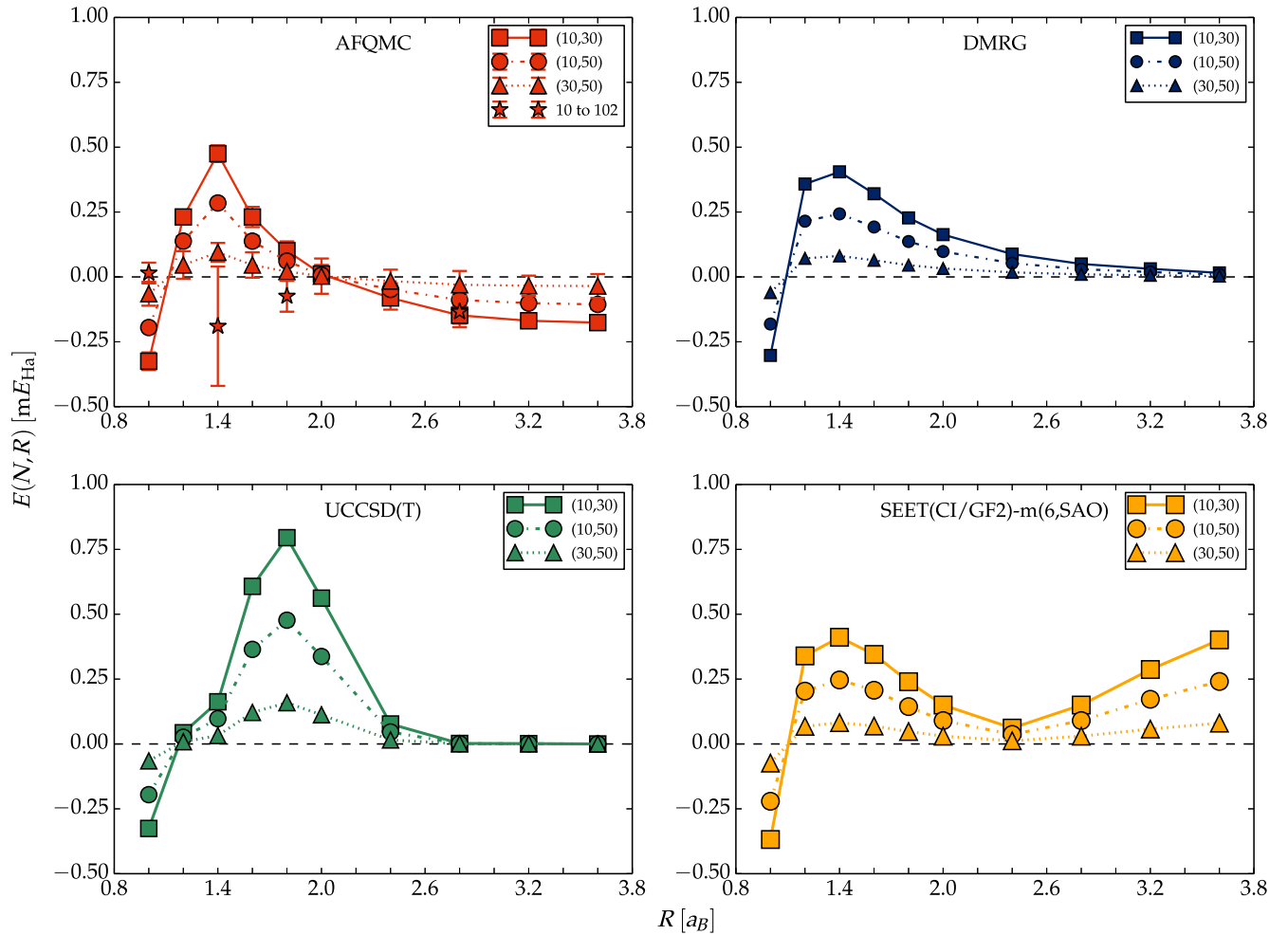


FIG. 15. Illustration of the extrapolation to TDL. Residual errors of fits from different pairs  $(N_1, N_2)$  of finite sizes are shown versus the reference extrapolation from  $N = 10, 30, 50$ . Data are for the minimal STO-6G basis. Large system sizes are investigated in some cases to verify convergence.



$x - 1$  neighbors around it. The energy contribution due to the central H atom is taken as the energy per atom, while the chemical potential is adjusted in the central H atom such that its particle number contribution is 1.

In order to more efficiently access large impurity sizes, we truncate the embedding space such that the bath orbitals with a small norm ( $< 0.01$ ) are excluded from the high-level calculation. Our DMET results are reported using DMRG as a solver and a bond dimension of  $D = 1000$ . The error from the DMRG solver is less than  $1 \mu E_{\text{Ha}}$  in the energy per atom. To converge to the TDL it is necessary to converge both the full system as well as the fragment size. We find that it is necessary for the full system to be very large to converge to the TDL at short bond lengths. We carry out calculations on systems of increasing size (up to

$H_{1950}$  at  $R = 1.0 a_B$ ) using a fixed fragment size [22], until the change in the energy per atom is smaller than  $0.01 mE_{\text{Ha}}$ . Using this suitably defined full system size [108], we perform calculations on larger and larger fragments [see Fig. 16(b)] and perform a quadratic extrapolation with the inverse of fragment size. The fragment size extrapolation using the largest 4 or 5 fragment sizes yields the same limit to better than  $0.02 mE_{\text{Ha}}$ .

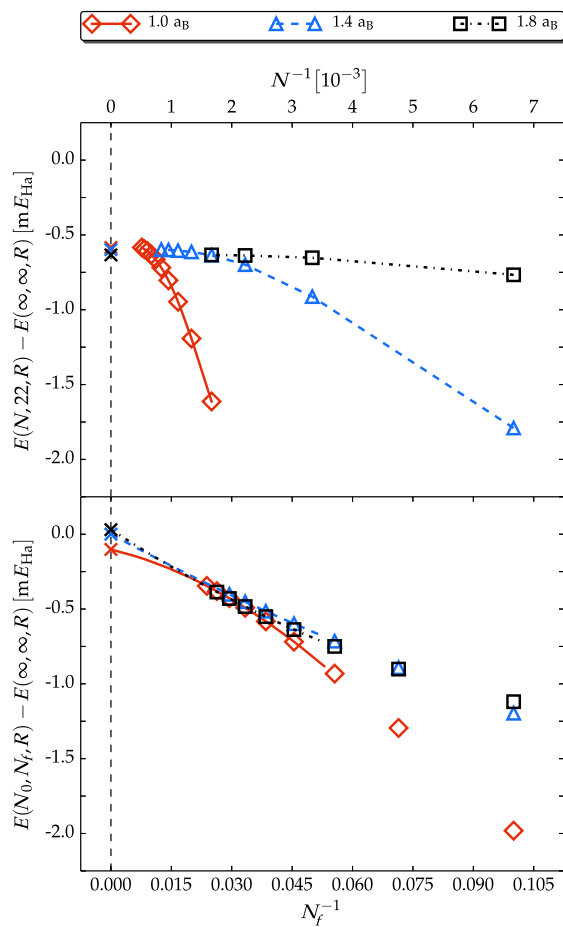


FIG. 16. Top: Difference with respect to extrapolated energies  $E_{\text{TDL}}(R) \equiv E(\infty, \infty, R)$  in STO-6G DMET calculations of increasing full system size  $N$  (using a fixed fragment of size [22]) for three different geometries. (b) Difference with respect to extrapolated energies  $E_{\text{TDL}}(R) \equiv E(\infty, \infty, R)$  in STO-6G DMET calculations of increasing fragment size  $N_f$  (using a fixed full system size  $N_0$ :  $H_{1200}$  for  $1.0$  and  $1.4 a_B$  and  $H_{450}$  for  $1.8 a_B$ ). The difference between the asymptotic value  $E(N_0, \infty, R)$  and the final extrapolation  $E(\infty, \infty, R)$  is estimated as  $E(\infty, \infty, R) - E(N_0, \infty, R) \approx E(\infty, 22, R) - E(N_0, 22, R)$

- [1] P. A. M. Dirac, *Quantum Mechanics of Many-Electron Systems*, Proc. R. Soc. A **123**, 714 (1929).
- [2] J. P. F. LeBlanc *et al.*, *Solutions of the Two-Dimensional Hubbard Model: Benchmarks and Results from a Wide Range of Numerical Algorithms*, Phys. Rev. X **5**, 041041 (2015).
- [3] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehler, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, *Stripe Order in the Underdoped Region of the Two-Dimensional Hubbard Model*, arXiv:1701.00054.
- [4] J. Hachmann, W. Cardoen, and G. K.-L. Chan, *Multi-reference Correlation in Long Molecules with the Quadratic Scaling Density Matrix Renormalization Group*, J. Chem. Phys. **125**, 144101 (2006).
- [5] W. A. Al-Saidi, S. Zhang, and H. Krakauer, *Bond Breaking with Auxiliary-Field Quantum Monte Carlo*, J. Chem. Phys. **127**, 144101 (2007).
- [6] T. Tsuchimochi and G. E. Scuseria, *Strong Correlations via Constrained-Pairing Mean-Field Theory*, J. Chem. Phys. **131**, 121102 (2009).
- [7] A. V. Sinititskiy, L. Greenman, and D. A. Mazziotti, *Strong Correlation in Hydrogen Chains and Lattices Using the Variational Two-Electron Reduced Density Matrix Method*, J. Chem. Phys. **133**, 014104 (2010).
- [8] D. A. Mazziotti, *Large-Scale Semidefinite Programming for Many-Electron Quantum Mechanics*, Phys. Rev. Lett. **106**, 083001 (2011).
- [9] N. Lin, C. A. Marianetti, A. J. Millis, and D. R. Reichman, *Dynamical Mean-Field Theory for Quantum Chemistry*, Phys. Rev. Lett. **106**, 096402 (2011).
- [10] L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, *Strong Electronic Correlation in the Hydrogen Chain: A Variational Monte Carlo Study*, Phys. Rev. B **84**, 245117 (2011).
- [11] M. Motta *et al.*, Github repository, <https://github.com/simonsfoundation/hydrogen-benchmark-PRX/>.
- [12] W. Kutzelnigg,  *$R12$ -Dependent Terms in the Wave Function as Closed Sums of Partial Wave Amplitudes for Large  $l$* , Theor. Chim. Acta **68**, 445 (1985).
- [13] L. Kong and E. F. Valeev, *Perturbative Correction for the Basis Set Incompleteness Error of Complete-Active-Space Self-Consistent Field*, J. Chem. Phys. **133**, 174126 (2010).
- [14] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, *Monte Carlo Calculations of Coupled Boson-Fermion Systems. I*, Phys. Rev. D **24**, 2278 (1981).
- [15] S. Zhang, J. Carlson, and J. E. Gubernatis, *Constrained Path Monte Carlo Method for Fermion Ground States*, Phys. Rev. B **55**, 7464 (1997).

- [16] S. Zhang and H. Krakauer, *Quantum Monte Carlo Method Using Phase-Free Random Walks with Slater Determinants*, *Phys. Rev. Lett.* **90**, 136401 (2003).
- [17] K. Van Houcke, E. Kozik, N. Prokof'ev, and B. Svistunov, *Diagrammatic Monte Carlo*, *Phys. Procedia* **6**, 95 (2010).
- [18] S. A. Kulagin, N. Prokof'ev, O. A. Starykh, B. Svistunov, and C. N. Varney, *Bold Diagrammatic Monte Carlo Method Applied to Fermionized Frustrated Spins*, *Phys. Rev. Lett.* **110**, 070601 (2013).
- [19] I. S. Tupitsyn, A. S. Mishchenko, N. Nagaosa, and N. Prokof'ev, *Coulomb and Electron-Phonon Interactions in Metals*, *Phys. Rev. B* **94**, 155145 (2016).
- [20] G. Knizia and Garnet Kin-Lic Chan, *Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory*, *Phys. Rev. Lett.* **109**, 186404 (2012).
- [21] G. Knizia and G. K. -L. Chan, *Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory*, *J. Chem. Theory Comput.* **9**, 1428 (2013).
- [22] S. Wouters, C. A. Jimenez-Hoyos, Q. Sun, and G. K. -L. Chan, *A Practical Guide to Density Matrix Embedding Theory in Quantum Chemistry*, *J. Chem. Theory Comput.* **12**, 2706 (2016).
- [23] S. R. White, *Density Matrix Formulation for Quantum Renormalization Groups*, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [24] S. R. White and R. L. Martin, *Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group*, *J. Chem. Phys.* **110**, 4127 (1999).
- [25] G. K. -L. Chan and M. Head-Gordon, *Highly Correlated Calculations with a Polynomial Cost Algorithm: A Study of the Density Matrix Renormalization Group*, *J. Chem. Phys.* **116**, 4462 (2002).
- [26] G. K. -L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, *Matrix Product Operators, Matrix Product States, and Ab Initio Density Matrix Renormalization Group Algorithms*, *J. Chem. Phys.* **145**, 014102 (2016).
- [27] J. J. Phillips and D. Zgid, *Communication: The Description of Strong Correlation within Self-Consistent Green's Function Second-Order Perturbation Theory*, *J. Chem. Phys.* **140**, 241101 (2014).
- [28] A. A. Rusakov and D. Zgid, *Self-Consistent Second-Order Green's Function Perturbation Theory for Periodic Systems*, *J. Chem. Phys.* **144**, 054106 (2016).
- [29] J. J. Phillips, A. A. Kananenka, and D. Zgid, *Fractional Charge and Spin Errors in Self-Consistent Green's Function Theory*, *J. Chem. Phys.* **142**, 194108 (2015).
- [30] A. R. Welden, A. A. Rusakov, and D. Zgid, *Exploring Connections between Statistical Mechanics and Green's Functions for Realistic Systems. Temperature Dependent Electronic Entropy and Internal Energy from a Self-Consistent Second-Order Green's Function*, *J. Chem. Phys.* **145**, 204106 (2016).
- [31] N. E. Dahlen and U. von Barth, *Variational Energy Functionals Tested on Atoms*, *Phys. Rev. B* **69**, 195102 (2004).
- [32] N. E. Dahlen and U. von Barth, *Variational Second-Order Møller-Plesset Theory Based on the Luttinger-Ward Functional*, *J. Chem. Phys.* **120**, 6826 (2004).
- [33] M. Casula, C. Filippi, and S. Sorella, *Diffusion Monte Carlo Method with Lattice Regularization*, *Phys. Rev. Lett.* **95**, 100201 (2005).
- [34] P. Reynolds, D. Ceperley, B. Alder, and J. Lester, *Fixed-Node Quantum Monte Carlo for Molecules*, *J. Chem. Phys.* **77**, 5593 (1982).
- [35] H.-J. Werner and P. J. Knowles, *An Efficient Internally Contracted Multiconfiguration-Reference Configuration Interaction Method*, *J. Chem. Phys.* **89**, 5803 (1988).
- [36] P. J. Knowles and H.-J. Werner, *An Efficient Method for the Evaluation of Coupling Coefficients in Configuration Interaction Calculations*, *Chem. Phys. Lett.* **145**, 514 (1988).
- [37] C. Angeli, M. Pastore, and R. Cimiraglia, *New Perspectives in Multireference Perturbation Theory: The N-Electron Valence State Approach*, *Theor. Chem. Acc.* **117**, 743 (2007).
- [38] J. Paldus and X. Li, *A Critical Assessment of Coupled Cluster Method in Quantum Chemistry*, *Adv. Chem. Phys.* **110**, 1 (1999).
- [39] R. J. Bartlett and M. Musial, *Coupled-Cluster Theory in Quantum Chemistry*, *Rev. Mod. Phys.* **79**, 291 (2007).
- [40] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, New York, 2009).
- [41] C. C. J. Roothaan, *Self-Consistent Field Theory for Open Shells of Electronic Systems*, *Rev. Mod. Phys.* **32**, 179 (1960).
- [42] E. Stoudenmire and S. R. White, *Sliced Basis Density Matrix Renormalization Group for Electronic Structure*, *Phys. Rev. Lett.* **119**, 046401 (2017).
- [43] C.-O. Almbladh, U. V. Barth, and R. V. Leeuwen, *Variational Total Energies from  $\Phi$ - and  $\Psi$ -Derivable Theories*, *Int. J. Mod. Phys. B* **13**, 535 (1999).
- [44] A. Stan, N. E. Dahlen, and R. van Leeuwen, *Fully Self-Consistent GW Calculations for Atoms and Molecules*, *Europhys. Lett.* **76**, 298 (2006).
- [45] A. Stan, N. E. Dahlen, and R. van Leeuwen, *Levels of Self-Consistency in the GW Approximation*, *J. Chem. Phys.* **130**, 114105 (2009).
- [46] A. A. Kananenka, E. Gull, and D. Zgid, *Systematically Improvable Multiscale Solver for Correlated Electron Systems*, *Phys. Rev. B* **91**, 121111 (2015).
- [47] T. N. Lan, A. A. Kananenka, and D. Zgid, *Communication: Towards Ab Initio Self-Energy Embedding Theory in Quantum Chemistry*, *J. Chem. Phys.* **143**, 241102 (2015).
- [48] T. Nguyen Lan, A. A. Kananenka, and D. Zgid, *Rigorous Ab Initio Quantum Embedding for Quantum Chemistry Using Green's Function Theory: Screened Interaction, Nonlocal Self-Energy Relaxation, Orbital Basis, and Chemical Accuracy*, *J. Chem. Theory Comput.* **12**, 4856 (2016).
- [49] D. Zgid and E. Gull, *Finite Temperature Quantum Embedding Theories for Correlated Systems*, *New J. Phys.* **19**, 023047 (2017).
- [50] T. N. Lan and D. Zgid, *Generalized Self-Energy Embedding Theory*, arXiv:1703.06981.
- [51] J. A. Pople and R. K. Nesbet, *Self-Consistent Orbitals for Radicals*, *J. Chem. Phys.* **22**, 571 (1954).

- [52] W. L. McMillan, *Ground State of Liquid He<sup>4</sup>*, *Phys. Rev.* **138**, A442 (1965).
- [53] D. Ceperley, G. V. Chester, and M. H. Kalos, *Monte Carlo Simulation of a Many-Fermion Study*, *Phys. Rev. B* **16**, 3081 (1977).
- [54] B.-X. Zheng, J. S. Kretchmer, H. Shi, S. Zhang, and Garnet Kin-Lic Chan, *Cluster Size Convergence of the Density Matrix Embedding Theory and Its Dynamical Cluster Formulation: A Study with an Auxiliary-Field Quantum Monte Carlo Solver*, *Phys. Rev. B* **95**, 045103 (2017).
- [55] T. H. Dunning, Jr., *Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen*, *J. Chem. Phys.* **90**, 1007 (1989).
- [56] D. Feller, *Application of Systematic Sequences of Wave Functions to the Water Dimer*, *J. Chem. Phys.* **96**, 6104 (1992).
- [57] T. Helgaker, W. Klopper, H. Koch, and J. Noga, *Basis-Set Convergence of Correlated Calculations on Water*, *J. Chem. Phys.* **106**, 9639 (1997).
- [58] E. Stoudenmire and S. R. White, *Studying Two-Dimensional Systems with the Density Matrix Renormalization Group*, *Annu. Rev. Condens. Matter Phys.* **3**, 111 (2012).
- [59] S. Chiesa, D. M. Ceperley, R. M. Martin, and M. Holzmann, *Finite-Size Error in Many-Body Simulations with Long-Range Interactions*, *Phys. Rev. Lett.* **97**, 076404 (2006).
- [60] G. D. Purvis III and R. J. Bartlett, *A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples*, *J. Chem. Phys.* **76**, 1910 (1982).
- [61] G. E. Scuseria, A. C. Scheiner, T. J. Lee, J. E. Rice, and H. F. Schaefer III, *The Closed-Shell Coupled Cluster Single and Double Excitation (CCSD) Model for the Description of Electron Correlation. A Comparison with Configuration Interaction (CISD) Results*, *J. Chem. Phys.* **86**, 2881 (1987).
- [62] J. A. Pople, M. Head-Gordon, and K. Raghavachari, *Quadratic Configuration Interaction. A General Technique for Determining Electron Correlation Energies*, *J. Chem. Phys.* **87**, 5968 (1987).
- [63] M. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria *et al.*, Gaussian 09, Revision D.01, Gaussian, Inc., 2013.
- [64] H. Koch, O. Christiansen, R. Kobayashi, P. Jorgensen, and T. Helgaker, *A Direct Atomic Orbital Driven Implementation of the Coupled Cluster Singles and Doubles (CCSD) Model*, *Chem. Phys. Lett.* **228**, 233 (1994).
- [65] S. Wouters and D. Van Neck, *The Density Matrix Renormalization Group for Ab Initio Quantum Chemistry*, *Eur. Phys. J. D* **68**, 272 (2014).
- [66] R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, *The Ab-Initio Density Matrix Renormalization Group in Practice*, *J. Chem. Phys.* **142**, 034102 (2015).
- [67] G. Barcza, Ö. Legeza, K. H. Marti, and M. Reiher, *Quantum-Information Analysis of Electronic States of Different Molecular Structures*, *Phys. Rev. A* **83**, 012508 (2011).
- [68] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, *MOLPRO: A General-Purpose Quantum Chemistry Program Package*, *Comput. Mol. Sci.* **2**, 242 (2012).
- [69] T. Shiozaki, G. Knizia, and H.-J. Werner, *Explicitly Correlated Multireference Configuration Interaction: MRCI-F12*, *J. Chem. Phys.* **134**, 034113 (2011).
- [70] T. Shiozaki and H.-J. Werner, *Multireference Explicitly Correlated F12 Theories*, *Mol. Phys.* **111**, 607 (2013).
- [71] T. Shiozaki and H.-J. Werner, *Explicitly Correlated Multireference Configuration Interaction with Multiple Reference Functions: Avoided Crossings and Conical Intersections*, *J. Chem. Phys.* **134**, 184104 (2011).
- [72] K. G. Dyall, *The Choice of a Zeroth-Order Hamiltonian for Second-Order Perturbation Theory with a Complete Active Space Self-Consistent-Field Reference Function*, *J. Chem. Phys.* **102**, 4909 (1995).
- [73] S. Guo, M. A. Watson, W. Hu, Q. Sun, and G. K.-L. Chan, *N-Electron Valence State Perturbation Theory Based on a Density Matrix Renormalization Group Reference Function, with Applications to the Chromium Dimer and a Trimer Model of Poly (p-Phenylenevinylene)*, *J. Chem. Theory Comput.* **12**, no. 1583 (2016).
- [74] S. Sharma and G. Chan, *Communication: A Flexible Multi-Reference Perturbation Theory by Minimizing the Hylleraas Functional with Matrix Product States*, *J. Chem. Phys.* **141**, 111101 (2014).
- [75] S. Sharma, G. Knizia, S. Guo, and A. Alavi, *Combining Internally Contracted States and Matrix Product States to Perform Multireference Perturbation Theory*, *J. Chem. Theory Comput.* **13**, 488 (2017).
- [76] R. L. Stratonovich, *On a Method of Calculating Quantum Distribution Functions*, *Sov. Phys. Dokl.* **2**, 416 (1957).
- [77] J. Hubbard, *Calculation of Partition Functions*, *Phys. Rev. Lett.* **3**, 77 (1959).
- [78] S. Zhang, *Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems*, in *Emergent Phenomena in Correlated Matter*, edited by E. Pavarini, E. Koch, and U. Schollwöck (Forschungszentrum Jülich, Jülich, Germany, 2013), Vol. 3.
- [79] M. Qin, H. Shi, and S. Zhang, *Coupling Quantum Monte Carlo and Independent-Particle Calculations: Self-Consistent Constraint for the Sign Problem Based on the Density or the Density Matrix*, *Phys. Rev. B* **94**, 235119 (2016).
- [80] W. Purwanto, H. Krakauer, and S. Zhang, *Pressure-Induced Diamond to  $\beta$ -Tin Transition in Bulk Silicon: A Quantum Monte Carlo Study*, *Phys. Rev. B* **80**, 214116 (2009).
- [81] Y. Virgus, W. Purwanto, H. Krakauer, and S. Zhang, *Stability, Energetics, and Magnetic States of Cobalt Adatoms on Graphene*, *Phys. Rev. Lett.* **113**, 175502 (2014).
- [82] W. Purwanto, S. Zhang, and H. Krakauer, *Excited State Calculations Using Phaseless Auxiliary-Field Quantum Monte Carlo: Potential Energy Curves of Low-Lying C<sub>2</sub> Singlet States*, *J. Chem. Phys.* **130**, 094107 (2009).
- [83] F. Ma, S. Zhang, and H. Krakauer, *Auxiliary-Field Quantum Monte Carlo Calculations with Multiple-Projector Pseudopotentials*, *Phys. Rev. B* **95**, 165103 (2017).

- [84] W. A. Al-Saidi, S. Zhang, and H. Krakauer, *Auxiliary-Field Quantum Monte Carlo Calculations of Molecular Systems with a Gaussian Basis*, *J. Chem. Phys.* **124**, 224101 (2006).
- [85] F. Ma, W. Purwanto, S. Zhang, and H. Krakauer, *Quantum Monte Carlo Calculations in Solids with Downfolded Hamiltonians*, *Phys. Rev. Lett.* **114**, 226401 (2015).
- [86] W. Purwanto, H. Krakauer, Y. Virgus, and S. Zhang, *Assessing Weak Hydrogen Binding on  $\text{Ca}^+$  Centers: An Accurate Many-Body Study with Large Basis Sets*, *J. Chem. Phys.* **135**, 164105 (2011).
- [87] M. Casula, S. Moroni, S. Sorella, and C. Filippi, *Size-Consistent Variational Approaches to Nonlocal Pseudopotentials: Standard and Lattice Regularized Diffusion Monte Carlo Methods Revisited*, *J. Chem. Phys.* **132**, 154113 (2010).
- [88] S. Sorella, TurboRVB Quantum Monte Carlo Package, <http://people.sissa.it/~sorella/web/>.
- [89] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. D. L. Pierre, P. D'Arco, Y. Noel, M. Causa, M. Rerat, and B. Kirtman, *CRYSTAL14: A Program for the Ab Initio Investigation of Crystalline Solids*, *Int. J. Quantum Chem.* **114**, 1287 (2014).
- [90] A. Zen, Y. Luo, S. Sorella, and L. Guidoni, *Molecular Properties by Quantum Monte Carlo: An Investigation on the Role of the Wave Function Ansatz and the Basis Set in the Water Molecule*, *J. Chem. Theory Comput.* **9**, 4332 (2013).
- [91] M. Marchi, S. Azadi, M. Casula, and S. Sorella, *Resonating Valence Bond Wave Function with Molecular Orbitals: Application to First-Row Molecules*, *J. Chem. Phys.* **131**, 154116 (2009).
- [92] G. Knizia, *Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts*, *J. Chem. Theory Comput.* **9**, 4834 (2013).
- [93] S. Wouters, C. A. Jiménez-Hoyos, Q. Sun, and G. K.-L. Chan, *A Practical Guide to Density Matrix Embedding Theory in Quantum Chemistry*, *J. Chem. Theory Comput.* **12**, 2706 (2016).
- [94] L. Hedin, *New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem*, *Phys. Rev.* **139**, A796 (1965).
- [95] A. A. Rusakov, J. J. Phillips, and D. Zgid, *Local Hamiltonians for Quantitative Green's Function Embedding Methods*, *J. Chem. Phys.* **141**, 194105 (2014).
- [96] A. A. Kananenka, A. R. Welden, T. N. Lan, E. Gull, and D. Zgid, *Efficient Temperature-Dependent Green's Function Methods for Realistic Systems: Using Cubic Spline Interpolation to Approximate Matsubara Green's Functions*, *J. Chem. Theory Comput.* **12**, 2250 (2016).
- [97] L. Hedin, *New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem*, *Phys. Rev.* **139**, A796 (1965).
- [98] J. M. Luttinger and J. C. Ward, *Ground-State Energy of a Many-Fermion System. II*, *Phys. Rev.* **118**, 1417 (1960).
- [99] G. Baym, *Self-Consistent Approximations in Many-Body Systems*, *Phys. Rev.* **127**, 1391 (1962).
- [100] A. A. Kananenka, J. J. Phillips, and D. Zgid, *Efficient Temperature-Dependent Green's Functions Methods for Realistic Systems: Compact Grids for Orthogonal Polynomial Transforms*, *J. Chem. Theory Comput.* **12**, 564 (2016).
- [101] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, *Self-Consistent GW: All-Electron Implementation with Localized Basis Functions*, *Phys. Rev. B* **88**, 075105 (2013).
- [102] A. Gaenko, A. E. Antipov, G. Carcassi, T. Chen, X. Chen, Q. Dong, L. Gamper, J. Gukelberger, R. Igarashi, S. Isakov, M. Könz, J. P. F. LeBlanc, R. Levy, P. N. Ma, J. E. Paki, H. Shinaoka, S. Todo, M. Troyer, and E. Gull, *Updated Core Libraries of the ALPS Project*, [arXiv: 1609.03930](https://arxiv.org/abs/1609.03930).
- [103] E. F. Valeev, LIBINT: A Library for the Evaluation of Molecular Integrals of Many-Body Operators over Gaussian Functions (version 2.1.0), <http://libint.valeev.net/>.
- [104] N. Prokof'ev and B. Svistunov, *Bold Diagrammatic Monte Carlo Technique: When the Sign Problem is Welcome*, *Phys. Rev. Lett.* **99**, 250201 (2007).
- [105] R. Rossi, *Determinant Diagrammatic Monte Carlo in the Thermodynamic Limit*, *Phys. Rev. Lett.* **119**, 045701 (2017).
- [106] R. Rossi, N. V. Prokof'ev, B. Svistunov, K. Van Houcke, and F. Werner, *Polynomial Complexity Despite the Fermionic Sign*, *Europhys. Lett.* **118**, 10004 (2017).
- [107] I. S. Tupitsyn and N. V. Prokof'ev, *Stability of Dirac Liquids with Strong Coulomb Interaction*, *Phys. Rev. Lett.* **118**, 026403 (2017).
- [108] For  $R = 1.0$  and  $1.2a_B$  we perform fragment-size extrapolations on  $\text{H}_{1200}$ , which is not converged with respect to the full system size. Our extrapolated result includes a full-system correction determined from calculations on larger systems, with a fragment of size [22].