MANY-BODY PHYSICS

Variational benchmarks for quantum many-body problems

Dian Wu[1,2], Riccardo Rossi[1,3], Filippo Vicentini[2,4,5], Nikita Astrakhantsev[6], Federico Becca[7], Xiaodong Cao[8], Juan Carrasquilla[9,10], Francesco Ferrari[11], Antoine Georges[4,5,8,12], Mohamed Hibat-Allah[9,13,14,15], Masatoshi Imada[16,17,18,19], Andreas M. Läuchli[1,20], Guglielmo Mazzola[21], Antonio Mezzacapo[22], Andrew Millis[8,23], Javier Robledo Moreno[8,24], Titus Neupert[6], Yusuke Nomura[25,26], Jannes Nys[1,2], Olivier Parcollet[8,27], Rico Pohle[17,19], Imelda Romero[1,2], Michael Schmid[17], J. Maxwell Silvester[28], Sandro Sorella[29]†, Luca F. Tocchio[30], Lei Wang[31,32], Steven R. White[28], Alexander Wietek[33], Qi Yang[31,34], Yiqi Yang[35], Shiwei Zhang[8], Giuseppe Carleo[1,2]*

The continued development of computational approaches to many-body ground-state problems in physics and chemistry calls for a consistent way to assess its overall progress. In this work, we introduce a metric of variational accuracy, the V-score, obtained from the variational energy and its variance. We provide an extensive curated dataset of variational calculations of many-body quantum systems, identifying cases where state-of-the-art numerical approaches show limited accuracy and future algorithms or computational platforms, such as quantum computing, could provide improved accuracy. The V-score can be used as a metric to assess the progress of quantum variational methods toward a quantum advantage for ground-state problems, especially in regimes where classical verifiability is impossible.

A key aspect of the quantum many-body problem, for systems ranging from the subatomic to molecules and materials, is determining the ground-state properties and energy. Knowing the ground state, one can predict which systems are stable and whether these systems exhibit useful and exotic phases, such as superconductivity or spin liquid states. However, because of the exponential complexity of the quantum wave function, finding the ground state of a many-body system can be very challenging, which limits exact numerical studies to a small number of particles. Efficiently solving the general ground-state problem is largely believed to be intractable. However, this does not necessarily apply to any particular system or class of systems, which may admit powerful approximations for ground states. Decades of research have focused on devising computational methods to find approximate solutions for specific cases of interest.

These computational methods have widely varying degrees of accuracy, and typically each method is much more successful on some systems than on others. Some of the most widely used methods include quantum Monte Carlo (QMC) (*1*–*3*), tensor networks (TNs) (*4*, *5*), and dynamical mean field theory (DMFT) and its extensions (*6*, *7*). It is known that the applicability of QMC methods is negatively affected by the frustration of the quantum system and particle statistics (*8*); similarly, high entanglement and large correlation lengths limit the applicability of TNs (*9*) and DMFT (*7*), respectively. Variational approaches based on physically motivated ansatzes (*10*, *11*) or neural networks (*12*) are not explicitly affected by the aforementioned issues. However, it is more difficult to assess their applicability and accuracy for a given quantum many-body system.

Quantum computers provide an alternative platform to attack quantum many-body problems (*13*). Notably, the dynamics of quantum many-body systems can be efficiently simulated by a digital quantum computer when the initial states are easy to prepare (*14*). Besides dynamics, substantial attention has been devoted to preparing ground states that are difficult to study with classical algorithms. Quantum algorithms for this task include phase estimation (*15*), variational approaches (*16*–*19*), adiabatic passage (*20*), imaginary time evolution (*21*), and subspace and Lanczos methods (*22*, *23*).

A fundamental challenge in assessing newly established computational methods based on classical or quantum computing is defining a consistent accuracy metric. Especially for ground-state problems, such a metric is necessary to clearly identify target Hamiltonians of broad interest, which cannot be solved with sufficient accuracy by existing methods. Also, this metric is crucial to quantify the improvements of computational approaches with time. In the context of assessing quantum computing–based methods, this issue pertains to the broader problem of determining in what cases quantum computers have an advantage over classical ones (*24*, *25*).

Determining a consistent metric for physically and chemically relevant ground-state problems is one of the goals of this work. To this end, we provide a large, curated collection of variational and numerically exact results on strongly correlated lattice models obtained by both state-of-the-art and baseline methods. The data that we provide include multiple approaches, such as exact diagonalization (ED), QMC (*1*) in the auxiliary field algorithm (*26*–*30*), matrix product states (MPSs) (*4*), variational wave function formulated on a lattice (*31*), and neural network–based methods (*12*). In addition to providing the data, we introduce an indicator of the variational accuracy of these results, named the V-score, that is suitable for directly comparing classical and quantum computing–based variational approaches. The V-score, obtained as a combination of the mean energy and its variance of a given variational state, allows us to identify what Hamiltonians and regimes are hard to approximate with classical variational methods without prior knowledge of the exact solution. Furthermore, we argue that the V-score can be used as a controlled benchmark to quantify the continued progress of quantum algorithms and quantum hardware to simulate those challenging target Hamiltonians.

[1]Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. [2]Center for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. [3]Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75005 Paris, France. [4]CPHT, CNRS, Ecole Polytechnique, IP Paris, F-91128 Palaiseau, France. [5]Collège de France, 75005 Paris, France. [6]Department of Physics, University of Zurich, 8057 Zurich, Switzerland. [7]Dipartimento di Fisica, Università di Trieste, Strada Costiera 11, I-34151 Trieste, Italy. [8]Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA. [9]Vector Institute, MaRS Centre, Toronto, ON M5G 1M1, Canada. [10]Institute for Theoretical Physics, ETH Zürich, 8093 Zurich, Switzerland. [11]Institut für Theoretische Physik, Goethe-Universität, 60438 Frankfurt am Main, Germany. [12]Department of Quantum Matter Physics, Université de Genève, CH-1211 Geneva, Switzerland. [13]Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada. [14]Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada. [15]Department of Applied Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada. [16]Toyota Physical and Chemical Research Institute, 41-1 Yokomichi, Nagakute, Aichi, 480-1192, Japan. [17]Waseda Research Institute for Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan. [18]Physics Division, Sophia University, Chiyoda-ku, Tokyo, 102-8554, Japan. [19]Department of Applied Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. [20]Laboratory for Theoretical and Computational Physics, Paul Scherrer Institute, 5232 Villigen, Switzerland. [21]Institute for Computational Science, University of Zurich, 8057 Zurich, Switzerland. [22]IBM Quantum, T. J. Watson Research Center, Yorktown Heights, NY 10598, USA. [23]Department of Physics, Columbia University, New York, NY 10027, USA. [24]Center for Quantum Phenomena, Department of Physics, New York University, New York, NY 10003, USA. [25]Department of Applied Physics and Physico-Informatics, Keio University, 3-1471 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan. [26]Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan. [27]Université Paris-Saclay, CNRS, CEA, Institut de physique théorique, 91191 Gif-sur-Yvette, France. [28]Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA. [29]SISSA, International School for Advanced Studies, 34136 Trieste, Italy. [30]Institute for Condensed Matter Physics and Complex Systems, Department of Applied Science and Technology (DISAT), Politecnico di Torino, I-10129 Torino, Italy. [31]Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [32]Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China. [33]Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany. [34]University of Chinese Academy of Sciences, Beijing 100049, China. [35]Department of Physics, College of William and Mary, Williamsburg, VA 23187, USA.

***Corresponding author. Email: giuseppe.carleo@epfl.ch**

†Deceased.

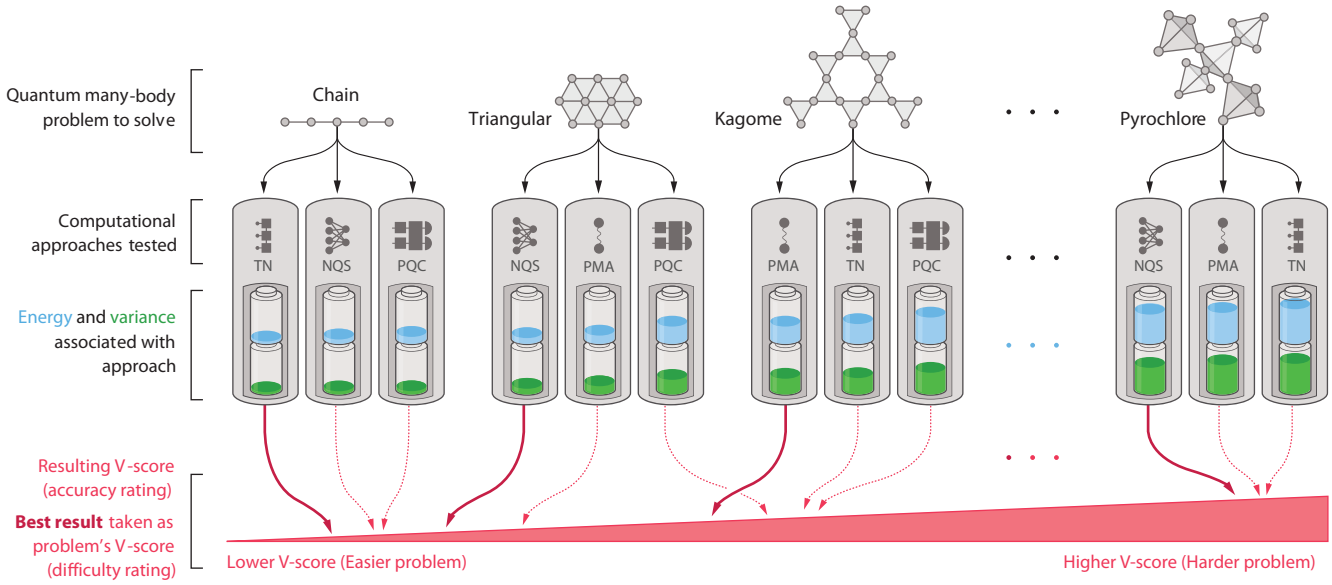

Fig. 1. Sketch of the V-score as a metric of simulation hardness. In this work, we present an extensive dataset of computational results for quantum many-body ground-state problems (for this sketch, we have selected a spin-½ system on a chain, a kagome, and a pyrochlore lattice). For each Hamiltonian in the dataset, we compute the mean energy and its variance with different variational techniques, including physically motivated ansatzes (PMAs), neural quantum states (NQSs), TNs, and PQCs. The energy and the variance are combined into the V-score, a metric of variational accuracy that we introduce in the main text. A low V-score is associated with high accuracy. The best result per Hamiltonian is then taken as the V-score for that Hamiltonian and used to rank the Hamiltonians in terms of simulation accuracy, highlighting which quantum many-body models are hard to simulate with current methods.

Benchmarking variational algorithms

We focus our study on benchmarking classical and quantum variational algorithms in approximating ground states of quantum many-body systems. On the classical side, these algorithms involve explicitly maintained variational representations of wave functions, such as TN or variational Monte Carlo (VMC)–based approaches. On the quantum side, the variational methods of major interest involve parameterized quantum circuits (PQCs) or other state preparation techniques based on local unitary transformations. In all cases, we assume that the methods to be benchmarked allow unbiased estimates of expectation values for Hamiltonians with few-body interactions (k-local operators, in the language of quantum information). Such expectation values can be obtained with a controllable statistical error, as in the case of classical Monte Carlo–based techniques, or as a result of statistical noise caused by measurements on quantum hardware.

Choice of problems

There is large freedom in the choice of many-body quantum problems that can be used to benchmark computational techniques. In this work, we have decided to focus on lattice Hamiltonians. These are minimal models of strong correlations and typically capture the essence of many physical systems. Lattice models first rose to prominence within classical statistical mechanics with the definition of the Ising model (*32*). Within solid-state physics, they stem from tight binding approaches to describe the electronic band structure (*33*). More recently, within the second quantization formalism, they are routinely used in different areas of physics to understand the low-energy behavior of unconventional quantum phases and transitions among them (*34*, *35*). In this

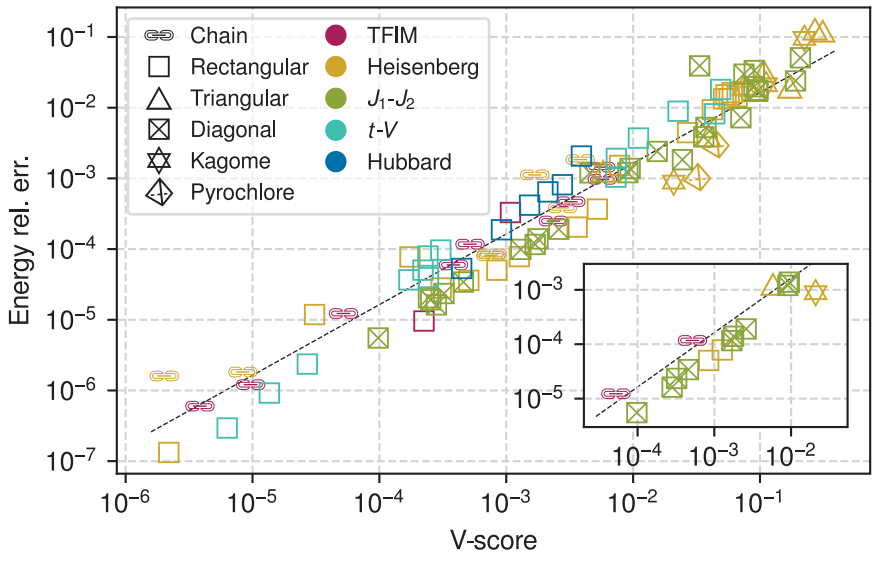

Fig. 2. Validation of V-score against exact results. We compare V-scores versus energy relative errors on various strongly correlated models for which exact results (ED or QMC) are available. The black dashed line is a least-squares fit of log(energy rel. err.) = log(V-score) + C, where C = −1.80 ± 0.08. The inset focuses on PQC results run on classical hardware (no shot noise included). The symbol shapes and colors correspond to the lattice geometry and model type, respectively (see legend).

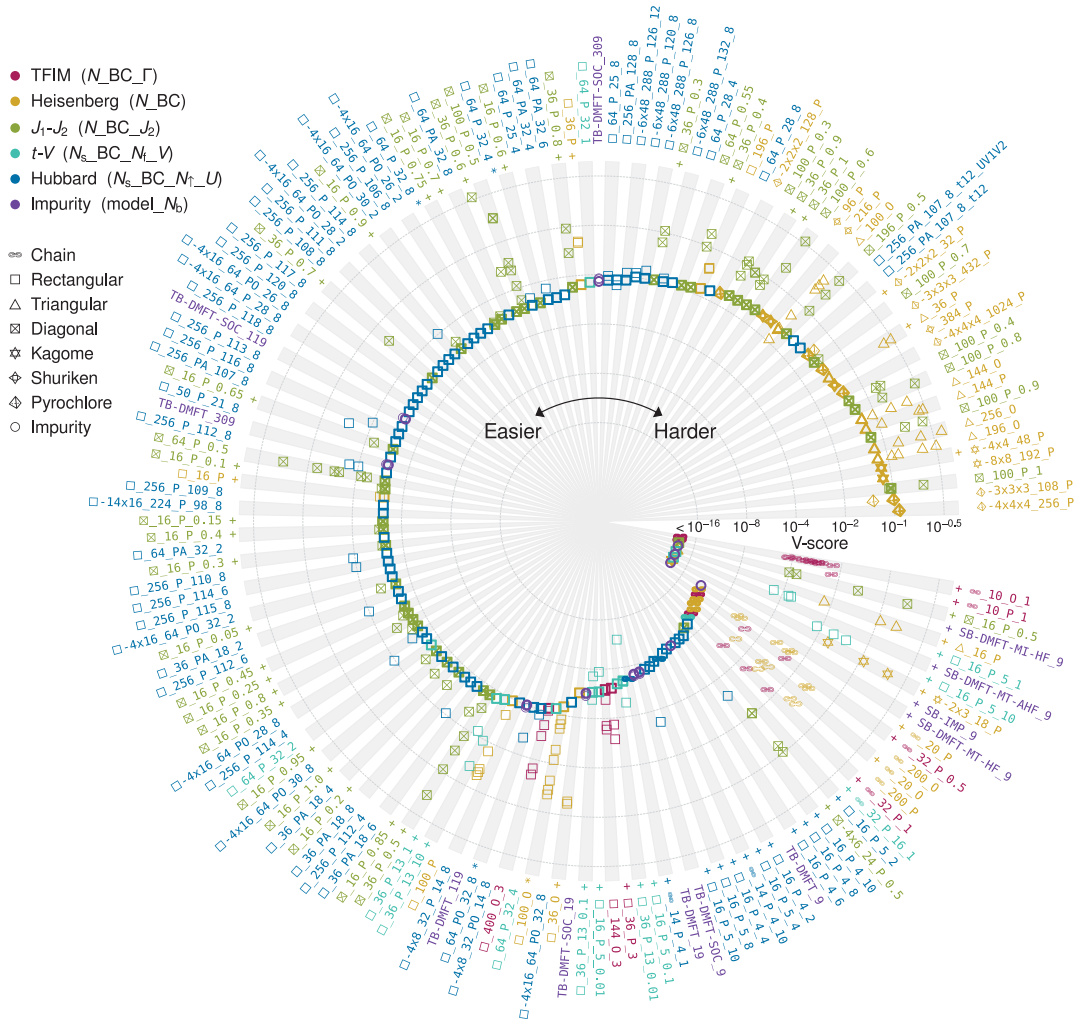

Fig. 3. V-score rankings of the complete dataset of variational benchmark results. Each radial slice shows the results for a given Hamiltonian. The bold marker in the slice shows the method with the lowest variational energy (rather than the lowest V-score) for that Hamiltonian, and its V-score determines the clockwise ranking of that Hamiltonian. The radial axis uses the doubly log scale to show the results across a wide range of V-score magnitude. The results with V-scores less than 10^{-16} are indistinguishable from the exact solutions because of limited numerical precision. The label outside each slice is the name of the Hamiltonian in the dataset, which describes the lattice geometry, the number of spins or particles N_s, the boundary conditions (BCs), and the Hamiltonian parameters, such as the interaction strength (Γ, J_2, V, U) and the number of particles (N_f, $N_\uparrow$, N_b). If a rectangular or diagonal lattice is nonsquare, its two edge lengths are shown, such as "□-4×8." Otherwise, it is assumed to be square, and it can be tilted, allowing for numbers of spins, such as 50. In a kagome or pyrochlore lattice, the number of copies of the unit cell in each spatial dimension is shown, such as "☼-4×4." The boundary conditions are O (open), P (periodic), or A (antiperiodic). See legend for the meaning of the various parameters associated with each Hamiltonian type. Additionally, we include results on the generalized Hubbard model with second nearest neighbor hopping t_2, nearest neighbor repulse V_1, and second nearest neighbor repulse V_2, as indicated by the suffixes "_t12" and "_t12_UV1V2." Formal definitions of the Hamiltonian parameters and specific types of impurity models are provided in the supplementary materials (*62*). The "+" marker near the label indicates that an ED solution is available for that Hamiltonian, and the "*" marker indicates a numerically exact QMC solution.

regard, the transverse-field Ising model (TFIM) provides the simplest example of a zero-temperature phase transition purely driven by quantum fluctuations between a paramagnet and a ferromagnet as seen, for example, in the Ising ferromagnet $LiHoF_4$ (*34*, *36*). Other prominent examples are the various quantum impurity models, in which a localized interacting degree of freedom is embedded into a non-interacting bulk, such as the Anderson impurity model (*37*). Quantum impurity models are central to quantum embedding methods, such as DMFT (*7*), and have applications to nanoelectronic devices (*38*). Their lattice generalizations, such as the Kondo lattice model, describe heavy fermion systems with 4f or 5f atoms, such as Ce or U (*39*).

Similarly, the Hubbard model (*40*–*42*) has been widely used to capture the essence of strong correlation in solids and has been shown to be relevant to the study of high-temperature superconductivity in cuprate compounds—e.g., $La_{2-x}Sr_xCuO_4$ (*43*)—and the Mott metal-insulator transitions in a variety of compounds (*44*). A descendant of the Hubbard model, the Heisenberg model describes a wide range of magnetic phases, for example with ferromagnetic or antiferromagnetic orders (*45*). In addition, when defined on geometrically frustrated lattices, possibly with anisotropic super-exchange couplings, the Heisenberg model gives rise to a wealth of phenomena, including spin liquid phases with topological order and exotic critical points (*46*, *47*). In this respect, the rare earth compound $YbMgGaO_4$ (*48*) and the mineral herbertsmithite $ZnCu_3(OH)_6Cl_2$ (*49*) have offered examples for unconventional quantum phases on triangular and kagome lattices.

V-score

To quantify the accuracy of two or more variational methods applied on the same ground-state approximation task, a key indicator is the expectation value of the energy $E = \langle \hat{H} \rangle$, an unbiased metric to assess the relative accuracy of variational methods; here, $\hat{H}$ is the relevant Hamiltonian. Given, for example, two independent methods preparing approximate ground states with variational energies E_a and E_b, the one providing the lower energy can be considered more accurate. From a practical point of view, however, it is preferable to have an absolute metric capable of predicting the accuracy of a method without comparing it with other methods. This would, for instance, allow comparing the performance of a given method on different tasks. Nonetheless, it is unlikely we could find such a metric that is provably

applicable in all cases because its existence would also allow the solution of NP-hard problems (*50*). We are therefore forced to settle for an empirically applicable metric. Moreover, the metric should be easy to estimate with variational methods.

Apart from the mean energy, for most variational methods, we also have a controllable estimate of the energy variance $\mathrm{Var}E = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2$. It has the important property that it exactly vanishes if computed on the exact ground state. Therefore, Var$E$ can be used to infer some information about the distance of the variational energy E from the exact, and a priori unknown, ground-state energy E_0. After early empirical observations (*51*), it has been shown that Var$E$ scales linearly with the deviation $E - E_0$ (*52*–*54*), so it can be used as a measure of the accuracy of the variational state.

We can use E and Var$E$ to create a dimensionless, intensive combination

$$\mathrm{V-score} = \frac{N\mathrm{Var}E}{(E - E_\infty)^2} \quad (1)$$

where N is the number of degrees of freedom, which is the number of spins for spin models and the number of particles for fermionic models. The constant E_∞ serves as a zero point of the energy, compensating for any global energy shift in the Hamiltonian definition. The V-score is dimensionless in energy units and system size for the variational states that we consider and is also invariant under energy shifts by construction. The procedure to benchmark the V-score is sketched in Fig. 1, and a detailed discussion of the definition of the V-score is presented in the supplementary materials (*55*).

To further justify the definition of the V-score, in Fig. 2, we present a comparison of this quantity against the energy relative error $(E - E_0)/(E_\infty - E_0)$ for a wide range of Hamiltonians and variational methods, where the ground-state energy E_0 is obtained by ED or numerically exact QMC. Despite the great diversity of Hamiltonians and variational methods considered, the V-score is a notably consistent and reliable estimator for the order of magnitude of the energy relative error, as shown by the linear fit in Fig. 2. In the inset of Fig. 2, we show that the same linear fit also well describes classically simulated PQCs, optimized with the variational quantum eigensolver (VQE) algorithm. These results validate the V-score as an absolute performance metric for both classical and quantum variational algorithms, at least for the Hamiltonians and the techniques we consider in the paper.

Identifying hard problems

We can now discuss which Hamiltonians are hard for the state-of-the-art variational methods presented in our collection. Given the intrinsic exponential complexity of the problem, it is no longer possible to obtain ED results on larger system sizes. Thus, using the V-score as a guide in this task is crucial. In Fig. 3, we show the V-score of all Hamiltonians and methods in our dataset. We first select the best method for each Hamiltonian by choosing the one with the lowest variational energy. We then use this method's V-score as an absolute hardness metric in the ground-state approximation task, which we refer to as the V-score of this Hamiltonian. The V-score of the best-performing method is marked in bold in Fig. 3. For additional clarity, in Fig. 4, we classify all those results by Hamiltonian types and lattice geometries.

It is well known that one-dimensional (1D) (chain) geometries are easy to solve with the density matrix renormalization group (DMRG). The small values of their V-scores in Fig. 3 and Fig. 4 clearly label them as accessible, particularly for spin models. Unfrustrated spin models typically also have small V-scores, ranging from 10^{-6} to 10^{-4} for the TFIM and the Heisenberg model on square lattices with open boundary conditions (OBCs). Moreover, these models can be efficiently simulated with unbiased stochastic techniques, such as QMC, and thus are easy to study on classical computers. On the other hand, the V-scores show that frustrated geometries, such as triangular, kagome, pyrochlore, and the J_1-J_2 square lattice, as well as fermionic models, such as the Hubbard model, are the most demanding for variational algorithms. With the help of Fig. 2, we can infer the order of magnitude of the energy relative error from the V-score. When this correspondence is applied to those hard problems, it predicts that we cannot expect an accuracy on the energy better than one or two digits.

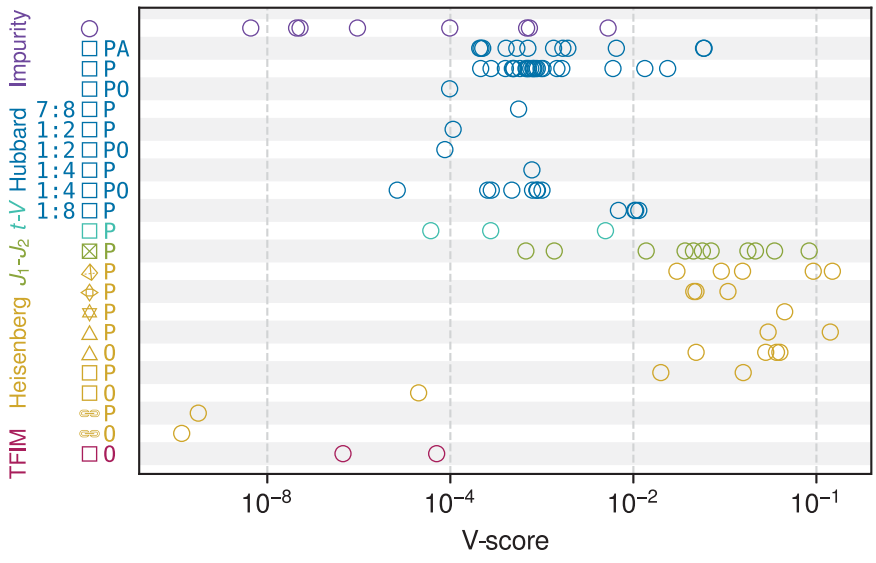

Fig. 4. V-scores of Hamiltonians, classified by Hamiltonian types and lattice geometries. A Hamiltonian's V-score is defined as the V-score of the available variational method with the lowest energy on this Hamiltonian. Only the Hamiltonians without ED results are shown. The ratios to the left of the lattice icons are aspect ratios of rectangular lattices. The letters to the right are boundary conditions (O, open; P, periodic; and A, antiperiodic).

A perspective on quantum advantage

Many recent theoretical and experimental efforts have been dedicated to showing the computational advantage of quantum computers over classical computers. Informed by theoretical computer science arguments, random circuit sampling has been proposed as a specific task to show such advantage (*56*, *57*). Still, it is unclear whether current noisy experimental quantum computing platforms can provide a solid and scalable advantage over classical ones (*58*–*60*). In addition to tasks of purely theoretical interest, there is also growing interest in finding practical quantum advantage (*61*), where quantum devices show a speedup for problems of scientific or technological relevance. The benchmarks introduced in this work belong to the family of approaches that can be useful to assess a quantum advantage that is practically relevant to physics.

In the context of variational ground-state algorithms, the V-score can readily be used as a good quality metric to assess quantum advantage. Furthermore, the V-score can be used as an absolute indicator of hardness for Hamiltonians. In this respect, Hamiltonians with large classical V-scores are identified as hard problems that are

not yet satisfactorily solved by classical computers and can be targeted by quantum computations. Finally, in the absence of classical verifiability of the quantum solutions, the V-score can benchmark the progress of variational quantum computing–based approaches in solving ground-state problems relevant to physics.

In Fig. 3, we show the V-scores of the classical variational methods that we have analyzed. From these results, we can infer that there is little room for quantum advantage in 1D geometries, where DMRG already achieves V-scores less than 10^{-8}. In higher dimensions, unfrustrated spin models, such as the TFIM or the Heisenberg model on the square lattice, are similarly well approximated by existing classical methods, with V-scores less than 10^{-4}. On the contrary, specific regimes of higher-dimensional frustrated spin models constitute a clear challenge for existing classical methods. For example, the pyrochlore or kagome Heisenberg models typically present V-scores greater than 10^{-2}—substantially higher than their unfrustrated counterparts. A similar scenario emerges for the Hubbard model in two dimensions with V-scores on the order of 10^{-2}. In the specific regimes of interaction strengths, geometries, and frustration identified by large V-scores, these models represent natural targets for variational quantum algorithms. Impurity models with multiple bands also represent an ideal terrain for practical quantum advantage because the ability of classical algorithms to simulate them rapidly degrades upon increasing the number of bands, as Fig. 3 shows for three-band models, and because of their importance for materials science.

We can also provide an early assessment of the V-scores obtained by the type of variational states that can be efficiently prepared on quantum computers. In this respect, it is encouraging to remark that PQCs perform well compared with classical variational methods, as shown in the inset of Fig. 2, at least for the small system sizes that we consider, where PQCs can be classically simulated and ideally optimized. Applications on quantum hardware are much more challenging because of stochastic fluctuations and noise. However, the baseline of ideally optimized PQCs is promising for applications.

Discussion and outlook

In this work, we have introduced the V-score, an empirical metric to quantify the absolute accuracy of variational solutions to strongly interacting quantum models. Supplemented with state-of-the-art results obtained by a large variety of numerical methods, this metric allows us to clearly identify models, geometries, and regimes in which existing approaches are currently less accurate. With the introduction of future computational techniques and improved computing architectures, the outcomes of this analysis will naturally evolve in time, revealing in a certifiable manner the continuous improvements happening in the field. In this respect, the dataset presented in this work can be a standardized way of taking snapshots of the evolution of quantum many-body techniques with time.

Besides the importance of these benchmarks for future developments in computational techniques based on classical computers, it will be especially interesting to use the V-score to directly measure the impact of quantum computing–based approaches. The hardest classical problem instances identified by large V-scores can be good candidates for studies based on quantum algorithms. In that context, the V-score can be used as a metric to assess progress in quantum variational state preparation, in the absence of classical verifiability.

REFERENCES AND NOTES

1. D. Ceperley, B. Alder, *Science* **231**, 555–560 (1986).
2. F. Becca, S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge Univ. Press, 2017).
3. E. Pavarini, E. Koch, S. Zhang, Eds., *Many-Body Methods for Real Materials*, vol. 9, Schriften des Forschungszentrums Jülich Modeling and Simulation (Verlag des Forschungszentrum Jülich, 2019).
4. S. R. White, *Phys. Rev. Lett.* **69**, 2863–2866 (1992).
5. R. Orús, *Nat. Rev. Phys.* **1**, 538–550 (2019).
6. A. Georges, G. Kotliar, *Phys. Rev. B* **45**, 6479–6483 (1992).
7. A. Georges, G. Kotliar, W. Krauth, M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13–125 (1996).
8. M. Troyer, U.-J. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005).
9. J. I. Cirac, D. Pérez-García, N. Schuch, F. Verstraete, *Rev. Mod. Phys.* **93**, 045003 (2021).
10. W. L. McMillan, *Phys. Rev.* **138**, A442–A451 (1965).
11. D. Ceperley, G. V. Chester, M. H. Kalos, *Phys. Rev. B* **16**, 3081–3099 (1977).
12. G. Carleo, M. Troyer, *Science* **355**, 602–606 (2017).
13. R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467–488 (1982).
14. S. Lloyd, *Science* **273**, 1073–1078 (1996).
15. A. Y. Kitaev, arXiv:quant-ph/9511026 (1995).
16. A. Peruzzo *et al*., *Nat. Commun.* **5**, 4213 (2014).
17. A. Kandala *et al*., *Nature* **549**, 242–246 (2017).
18. M. Cerezo *et al*., *Nat. Rev. Phys.* **3**, 625–644 (2021).
19. K. Bharti *et al*., *Rev. Mod. Phys.* **94**, 015004 (2022).
20. E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, arXiv:quant-ph/0001106 (2000).
21. M. Motta *et al*., *Nat. Phys.* **16**, 205–210 (2020).
22. R. M. Parrish, P. L. McMahon, arXiv:1909.08925 [quant-ph] (2019).
23. W. Kirby, M. Motta, A. Mezzacapo, *Quantum* **7**, 1018 (2023).
24. S. Bravyi, D. Gosset, R. König, *Science* **362**, 308–311 (2018).
25. A. Bouland, B. Fefferman, C. Nirkhe, U. Vazirani, *Nat. Phys.* **15**, 159–163 (2019).
26. R. Blankenbecler, D. J. Scalapino, R. L. Sugar, *Phys. Rev. D* **24**, 2278–2286 (1981).
27. S. Sorella, S. Baroni, R. Car, M. Parrinello, *Europhys. Lett.* **8**, 663–668 (1989).
28. M. Imada, Y. Hatsugai, *J. Phys. Soc. Jpn.* **58**, 3752–3780 (1989).
29. S. Zhang, J. Carlson, J. E. Gubernatis, *Phys. Rev. B* **55**, 7464–7477 (1997).
30. H. Shi, S. Chiesa, S. Zhang, *Phys. Rev. A* **92**, 033603 (2015).
31. P. Horsch, T. A. Kaplan, *J. Phys. C Solid State Phys.* **16**, L1203 (1983).
32. M. Niss, *Arch. Hist. Exact Sci.* **63**, 243–287 (2008).
33. J. C. Slater, G. F. Koster, *Phys. Rev.* **94**, 1498–1524 (1954).
34. S. Sachdev, *Quantum Phase Transitions* (Cambridge Univ. Press, 2001).
35. X.-G. Wen, *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons* (Oxford Univ. Press, 2004).
36. D. Bitko, T. F. Rosenbaum, G. Aeppli, *Phys. Rev. Lett.* **77**, 940–943 (1996).
37. P. W. Anderson, *Phys. Rev.* **124**, 41–53 (1961).
38. L. Kouwenhoven, L. Glazman, *Phys. World* **14**, 33–38 (2001).
39. A. C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, 1993).
40. J. Hubbard, *Proc. R. Soc. Lond. A* **276**, 238–257 (1963).
41. J. Kanamori, *Prog. Theor. Phys.* **30**, 275–289 (1963).
42. M. C. Gutzwiller, *Phys. Rev. Lett.* **10**, 159–162 (1963).
43. P. A. Lee, N. Nagaosa, X.-G. Wen, *Rev. Mod. Phys.* **78**, 17–85 (2006).
44. M. Imada, A. Fujimori, Y. Tokura, *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
45. E. Manousakis, *Rev. Mod. Phys.* **63**, 1–62 (1991).
46. L. Balents, *Nature* **464**, 199–208 (2010).
47. L. Savary, L. Balents, *Rep. Prog. Phys.* **80**, 016502 (2017).
48. Y. Li *et al*., *Phys. Rev. Lett.* **115**, 167203 (2015).
49. M. R. Norman, *Rev. Mod. Phys.* **88**, 041002 (2016).
50. F. Barahona, *J. Phys. A Math. Gen.* **15**, 3241–3253 (1982).
51. Y. Kwon, D. M. Ceperley, R. M. Martin, *Phys. Rev. B* **48**, 12037–12046 (1993).
52. M. Imada, T. Kashima, *J. Phys. Soc. Jpn.* **69**, 2723–2726 (2000).
53. T. Kashima, M. Imada, *J. Phys. Soc. Jpn.* **70**, 2287–2299 (2001).
54. S. Sorella, *Phys. Rev. B* **64**, 024512 (2001).
55. See the supplementary materials, section S2, for a detailed discussion on the definition of the V-score.
56. S. Boixo *et al*., *Nat. Phys.* **14**, 595–600 (2018).
57. F. Arute *et al*., *Nature* **574**, 505–510 (2019).
58. F. Pan, K. Chen, P. Zhang, *Phys. Rev. Lett.* **129**, 090502 (2022).
59. X. Gao *et al*., *PRX Quantum* **5**, 010334 (2024).
60. D. Aharonov, X. Gao, Z. Landau, Y. Liu, U. Vazirani, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, 2023), pp. 945–957.
61. A. J. Daley *et al*., *Nature* **607**, 667–676 (2022).
62. See the supplementary materials, section S1, for the definitions of the Hamiltonians used in this manuscript.
63. J. M. Silvester *et al*., varbench/varbench: v1.1.0, Zenodo (2024); https://doi.org/10.5281/zenodo.13377360.
64. L. Tocchio *et al*., varbench/methods: Methods v1.0.0, Zenodo (2024); https://doi.org/10.5281/zenodo.13377380.
65. M. Fishman, S. R. White, E. M. Stoudenmire, arXiv:2007.14822 [cs.MS] (2020).
66. T. Misawa *et al*., *Comput. Phys. Commun.* **235**, 447–462 (2019).
67. F. Vicentini *et al*., *SciPost Phys. Codebases*, 7 (2022).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Sandro Sorella, a dear friend, esteemed colleague, and coauthor of this paper, who passed away before its submission. We acknowledge discussions with A. Sandvik and M. Stoudenmire. We thank L. Reading-Ikkanda (Simons Foundation) for preparing Fig. 1. The Flatiron Institute is a division of the Simons Foundation. **Funding:** G.C. acknowledges support by the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant no. 205602). D.W. and G.C. acknowledge support from the Swiss National Science Foundation under grant no. 200336. R.R. and G.C. acknowledge support from SEFRI through grant no. MB22.00051 (NEQS–Neural Quantum Simulation). N.A. and T.N. acknowledge support from the European Union's Horizon 2020 research and innovation program (ERC-StG-Neupert-757867-PARATOP) and from the Swiss National Science Foundation (grant PP00P2_176877). J.C. acknowledges support from the Natural Sciences and Engineering Research Council (NSERC), the Shared Hierarchical Academic Research Computing Network (SHARCNET), Compute Canada, and the Canadian Institute for Advanced Research (CIFAR) AI chair program. M.I., Y.N., R.P., and M.S. acknowledge support by MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (Basic Science for Emergence and Functionality in Quantum Matter – Innovative Strongly Correlated Electron Science by Integration of Fugaku and Frontier Experiments, JPMXP1020200104) together with computational resources of supercomputer Fugaku provided by the RIKEN Center for Computational Science (project ID: hp210163 and hp220166). M.I. acknowledges the support of MEXT KAKENHI, Grant-in-Aid for Transformative Research Areas, grant nos. JP22H05111 and JP22H05114 and JSPS KAKENHI grant no. JP19H00658. M.I. and Y.N. acknowledge support from MEXT as "Program for Promoting Researches on the Supercomputer Fugaku" (grant no. JPMXP1020230411). G.M. acknowledges support from the Swiss National Science Foundation (grant PCEFP2_203455). Y.N. acknowledges support from Grant-in-Aids for Scientific Research (JSPS KAKENHI) (grant nos. JP23H04869, JP23H04519, JP23K03307, and JP21H01041) and JST (grant no. JPMJPF2221).

J.M.S. and S.R.W. acknowledge the support of the US National Science Foundation under grant DMR-2110041. L.W. and Q.Y. acknowledge the support of the National Natural Science Foundation of China under grant nos. 92270107, 12188101, T2225018, and T2121001. A.W. acknowledges support by the DFG through the Emmy Noether program (509755282). Y.Y. thanks the Center for Computational Quantum Physics (CCQ) of the Flatiron Institute, Simons Foundation for support and hospitality. Research at the Perimeter Institute is supported in part by the government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities. **Author contributions:** D.W. implemented the research infrastructure and curated the data. R.R. designed the final form of the V-score. G.C. designed and coordinated the project with the help of M.I. and R.R. The main text was written by D.W., R.R., F.V., and G.C. with input from all authors. All authors contributed to the supplementary materials. A detailed record of the authors' contributions to the dataset can be found on the GitHub repository. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The VarBench dataset and the code to analyze the benchmark results and reproduce the figures in this manuscript are available at Zenodo (*63*). Every benchmark result either has a reference to a paper reporting it or can be reproduced using the code at Zenodo (*64*). The software packages used to reproduce the data include ITensor (*65*), mVMC (*66*), and NetKet (*67*). Future updates to the dataset can be found at https://github.com/varbench/varbench.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg9774
Supplementary Text
Figs. S1 and S2
References (*68*–*155*)

Submitted 9 February 2023; accepted 12 September 2024
10.1126/science.adg9774