Realizing large-size quantum spin chains on cloud quantum computers

Hongye Yu,1 Yusheng Zhao,1 Tzu-Chieh Wei1,2*

1Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, NY 11794-3800, USA
2C. N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, NY 11794-3840, USA

*To whom correspondence should be addressed; E-mail: tzu-chieh.wei@stonybrook.edu.

Abstract

Quantum computers have the potential to efficiently simulate other quantum systems, especially when their size is large and classical simulation is not feasible. Despite the total qubit number reaching over one hundred in several existing quantum devices, their applicability is still plagued by the prevalent noise and errors. Here we report experiments on various IBM quantum computers to simulate ground states of quantum spin chains with qubit numbers ranging from 4 to 102. The estimated ground-state energies from these experiments across different backends and with varying sizes of system reach the expected values within a few error percentages. The enabling factors for accurate large experiments include physics-motivated variational ansätze and efficient and scalable energy measurement and error mitigation, particularly using a reference state in the zero-noise extrapolation.

Introduction

The notion of quantum computers traces back to the early 1980s in the works of Bennioff [1], Mannin [2] and, Feynman [3]. In particular, Feynman suggested using quantum computers to simulate other quantum systems more efficiently than classical ones, giving rise to the notion of a universal quantum simulator [4]. A critical breakthrough was made by Shor about a decade later who devised a factoring quantum algorithm almost exponentially faster than the current best classical algorithm [5]. Experimental progress has made a long way, and recent technological development has led to the burgeoning of quantum devices in several different physical implementations, with the total qubit number exceeding one hundred and the gate quality continuing to be improved. However, current machines are still regarded as noisy intermediate-scale quantum (NISQ) devices [6], not yet suitable for full-scale quantum error correction and fault-tolerant quantum computation. But there are concurrent efforts to develop error mitigation techniques [7,8,9,10,11,12,13,14,15] and algorithms [16] suitable
Figure 1: Variational ansatz, layout of a 102-qubit quantum computer, and simulation results of the ansatz. (a) The variational ansatz structure, (b) the layout of the 127-qubit ibm\_washington backend, (c) the fidelity of the optimal ansatz state with the ground state (GS) of the open-chain Heisenberg model vs. the total number of qubits $N$ for one to six layers in the ansatz; (d) the corresponding relative error in the GS energy. In (b), an chain of 102 qubits is illustrated by the thick, shaded line, starting from qubit 3 and ending at qubit 114.
for NISQ devices to realize the potential quantum advantage \cite{17}. Recent notable experimental achievements include random quantum circuits with 53 qubits \cite{18} and with 56 qubits \cite{19}, boson sampling with 76 photons in 100 modes \cite{20} and with 219 photons in 216 modes \cite{21}, the Hartree-Fock method for quantum chemistry with 12 qubits \cite{22}, topological-state realization with 31 qubits \cite{23}, and quantum walks on a 62-qubit superconducting processor \cite{24}.

Some of these tasks with large system sizes cannot be efficiently simulated by classical computers, such as the random circuits and the boson sampling, so their validation and verification rely on comparing small-size simulations with corresponding small-size experiments. Others, such as the Hartree-Fock method, topological-state realizations, and quantum walk do not necessarily focus on the advantage perspective and can be compared with properties accessible by classical computers. These latter quantum simulations and emulations provide useful benchmarks for gauging the performance of current and future quantum computers and can thus lend credibility to their performance of large-scale tasks not necessarily verifiable classically.

In this work, we present experimental realizations of approximate ground states of spin chains, up to 102 qubits, and report the estimated ground-state energy accurate to within a few percent of relative error. The ability to perform such large-scale experiments and extract accurate ground-state energy values relies on using a physics-motivated variational ansatz, efficient and scalable approaches for measuring energy and mitigating the experimentally acquired results. Those quantum circuits used in the gate-error mitigation include up to 77 layers of CNOT gates and, in the 102-qubit experiments, up to a total of 3186 CNOT gates, and the introduction of a reference state in the zero-noise extrapolation substantially improves the accuracy of the results.

**Heisenberg and XXZ models and the ansatz for ground states**  Quantum spin systems, such as the Heisenberg \cite{25} and XXZ models \cite{26, 27}, have sparked analytical development and understanding of quantum phases, as well as served as a testbed for the development of numerical techniques. Here, we explore the creation of approximate ground states of XXZ spin chains that can be potentially realized on quantum computers and used to characterize the physical properties, such as the ground-state energy and entanglement. Let us begin with the Hamiltonian of the paradigmatic spin-1/2 XXZ spin chain with the open boundary condition

$$\hat{H}_{XXZ}(\Delta) = \sum_{j=1}^{N-1} \hat{h}_{XXZ}[j,j+1](\Delta) = \sum_{j=1}^{N-1} \left( \sigma_x[j] \sigma_x[j+1] + \sigma_y[j] \sigma_y[j+1] + \Delta \sigma_z[j] \sigma_z[j+1] \right), \quad (1)$$

where $N$ is the total number of sites, the superscripts $[j]$’s are the site indices, $\sigma_x$, $\sigma_y$, and $\sigma_z$ are Pauli matrices associated with spin-1/2 angular momentum operators, $\Delta$ represents the anisotropy parameter.
Table 1: Results related to the open-chain Heisenberg model: optimized parameters \( \{ \theta^* \} \), energy values from the optimal ansatz \( E^*_{\text{ansatz}} \) and from the MPS obtained ground-state value \( E_{\text{gs}} \), and the relative error \( \epsilon \) and ground-state fidelity \( f \), as well as the experimental results \( E_{\text{exp}} \) and their errors. The numerical calculation was done with the MPS method using a bond dimension \( \chi = 64 \) with the termination condition to be a relative error of the energy in the iteration being \( 10^{-6} \) or less. We note that \( f = |\langle \psi_{\text{gs}}|\psi_{\text{ansatz}}(\{\theta^*\})\rangle| \) and \( \epsilon = |E_{\text{ansatz}}(\{\theta^*\}) - E_{\text{gs}}|/|E_{\text{gs}}| \). For the one-layer ansatz, there are only two variational parameters used, \( \{ \theta_{\text{even}}, \theta_{\text{odd}} \} \). The ‘error’ in the last column represents the relative error between the experimentally estimated value \( E_{\text{exp}} \) and the exact ground-state energy \( E_{\text{gs}} \). ∗Note that these values were obtained by averaging results over different backends and/or different groups of physical qubits; see Tables S.3 and S.2 for the complete list of results.

| \( N \) | \( \theta^*_{\text{even}} \) | \( \theta^*_{\text{odd}} \) | \( E^*_{\text{ansatz}} \) | \( E_{\text{gs}} \) | \( \epsilon \) | \( f \) | \( E_{\text{exp}} \) | \( \text{error} \) |
|-----|----------------|----------------|----------------|----------------|-----|-----|----------------|-----|
| 4   | 0.151748       | 0.215765       | -6.464102      | -6.464102      | 0   | 1.0000         | -6.5(1.6) | 0.56% |
| 6   | 0.141671       | 0.216088       | -9.880996      | -9.974099      | 0.94% | 0.9923       | -9.9(1.9)* | 0.19% |
| 8   | 0.138569       | 0.216093       | -13.299823     | -13.499730     | 1.48% | 0.9796       | -13.2(2.2) | 2.22% |
| 10  | 0.13710        | 0.216102       | -16.719307     | -17.032141     | 1.84% | 0.9639       | -16.7(1.3)* | 1.95% |
| 12  | 0.136248       | 0.216110       | -20.139037     | -20.568363     | 2.09% | 0.9462       | -20.3(2.1) | 1.30% |
| 14  | 0.135688       | 0.216115       | -23.58885      | -24.106899     | 2.27% | 0.9271       | -23.6(1.8) | 2.10% |
| 16  | 0.135293       | 0.216120       | -26.978800     | -27.649499     | 2.42% | 0.9072       | -25.8(1.6)* | 6.68% |
| 18  | 0.134999       | 0.216123       | -30.398756     | -31.188044     | 2.53% | 0.8867       | -30.7(0.7)* | 1.56% |
| 20  | 0.134773       | 0.216126       | -33.818738     | -34.729893     | 2.62% | 0.8659       | -33.0(0.5)* | 4.98% |
| 30  | 0.134132       | 0.216134       | -50.918850     | -52.445423     | 2.91% | 0.7614       | -50.2(2.0)* | 4.28% |
| 40  | 0.133832       | 0.216139       | -68.019098     | -70.165893     | 3.06% | 0.6629       | -68.5(2.0)* | 2.34% |
| 50  | 0.133658       | 0.216141       | -85.119397     | -87.884441     | 3.15% | 0.5737       | -85.0(2.8)* | 3.29% |
| 60  | 0.133544       | 0.216143       | -102.219721    | -105.612060    | 3.21% | 0.4946       | -99(4)      | 6.26% |
| 70  | 0.133464       | 0.216144       | -119.320058    | -123.36305     | 3.26% | 0.4253       | -125(7)     | 1.35% |
| 80  | 0.133405       | 0.216145       | -136.420403    | -141.06947     | 3.29% | 0.3649       | -138.5(2.5) | 1.82% |
| 90  | 0.133359       | 0.216146       | -153.520754    | -158.785857    | 3.32% | 0.3126       | -153(5)     | 3.64% |
| 98  | 0.133329       | 0.216146       | -167.201038    | -172.965924    | 3.33% | 0.2760       | -168.1(2.6) | 2.81% |
| 100 | 0.133323       | 0.216146       | -170.621109    | -176.510957    | 3.34% | 0.2675       | -173(9)     | 1.99% |
| 102 | 0.133316       | 0.216146       | -174.041180    | -180.055995    | 3.34% | 0.2592       | -177.5(2.7) | 1.42% |
in the spin-spin coupling, and we have set the overall coupling strength to unity. For $\Delta = 1$, the model reduces to the isotropic antiferromagnetic Heisenberg spin chain $\hat{H}_{\text{Heis}} = \hat{H}_{\text{XXZ}}(\Delta = 1)$. The XXZ model is known to possess three distinct quantum phases: (i) for $\Delta < -1$, it is a ferromagnetic phase, i.e., any classical ferromagnetic states, such as $|\uparrow\uparrow\ldots\rangle$ and $|\downarrow\downarrow\ldots\rangle$ and their uniform spin rotations are ground states. (ii) For $-1 < \Delta < 1$, it is a gapless, critical phase, including the special point $\Delta = 0$, i.e., the XY model. (iii) For $\Delta > 1$, it is an antiferromagnetic phase and becomes the antiferromagnetic Ising model when $\Delta \to \infty$. We will mainly focus on the range of $\Delta > -1$, as the ferromagnetic phase $\Delta < -1$ has simple ground states.

We explain in the Materials and Methods section in Appendix how we arrive at a physics-motivated variational ansatz, schematically shown in Fig. 1a, and it is captured in the following equation,

$$\psi_{\text{ansatz}}(\theta) = U_{\text{init}} |0\ldots0\rangle,$$

where $U_{\text{init}}$ denotes the initialization step that takes the fiducial state $|0\ldots0\rangle \equiv |\uparrow\ldots\uparrow\rangle$ to a product of singlets or Bell pairs labelled as $|\psi_{\text{singlets}}\rangle$, $U_{\text{even/odd}}^{(l)}(\theta_e/o)$’s are the $l$-th layer variational gates in a total of $N_L$ layers, and the parameters $\{\theta_e/o\}$’s are to be determined by minimizing the energy expectation value $E_{\text{ansatz}}(\theta) \equiv \langle \psi_{\text{ansatz}}(\theta) | H_{\text{XXZ}} | \psi_{\text{ansatz}}(\theta) \rangle$.

Variational ansätze and trial wavefunctions are commonly used in physics. Despite the drawback that they may not exactly represent the ground state, they capture the essential properties of the underlying physical model. Well-known examples include the BCS wave function for superconductivity [28] and the Laughlin wave function for the fractional quantum Hall effect [29].

But how accurately can this family of ansatz efficaciously simulate the ground-state wave functions and approximate their energy? To do this, we numerically minimize $E_{\text{ansatz}}(\theta)$ to obtain the optimal parameters $\{\theta^*\}$ and then compare the ground-state fidelity $f \equiv |\langle \psi_{gs} | \psi_{\text{ansatz}}(\{\theta^*\}) \rangle|$ and the relative error in the ground-state energy $\epsilon \equiv |E_{\text{ansatz}}(\{\theta^*\}) - E_{gs}|/|E_{gs}|$, where $|\psi_{gs}\rangle$ denotes the ground-state wavefunction and $E_{gs}$ its exact energy. We can use exact matrix multiplication and diagonalization for a small number of qubits. For larger numbers of qubits, we resort to numerical approximation by matrix product states (MPS) [31, 32, 33] to implement the DMRG algorithm [34], which is efficient for one-dimensional gapped systems. We find that for $N = 4$, the one-layer ansatz can reach the exact ground state for $\Delta \geq 0$ (with both open and periodic boundary conditions); with two layers, we can obtain the exact ground state for all $\Delta > -1$; see Fig. S.7a & c. For $N = 6$ (see Fig. S.7b & d), the 2-layer ansatz with the periodic boundary condition can reach the exact ground state for $\Delta > -1$ and the 2-layer ansatz with the open boundary condition achieves this exactness with a slightly narrower region. For $N = 6$, it takes three layers in the open-boundary ansatz to achieve the exactness for all $\Delta > -1$. 
Figure 2: Zero-noise gate-error extrapolation for experimental realisations of ansatz states of Heisenberg spin chains. (a) & (b) a 102-qubit Heisenberg chain on \textit{ibm.washington}; (c) & (d), a 50-qubit Heisenberg chain on \textit{ibmq.brooklyn}, (e) & (f) two 40-qubit random ansatz states on \textit{ibmq.brooklyn}. The dots represent the experimental data, whose error bars are small and thus not particularly visible in the plot. The curves represent the fitting with an exponential function, \( f_E(m) = a \exp(-bm) + c \), to fit the data, where \( m = 2n + 1 \) is the total number of \( U \) and \( U^{-1} \) in the circuit to prepare the ansatz state. The stars represent the extrapolated values at \( m = 0 \). In (e), the parameters \([\theta_{\text{even}}, \theta_{\text{odd}}] = [3.5, 0.7]\) were used for the Heisenberg and the exact ansatz energy is \(-16.0669\). In (f), parameters \([0.3, 1.7]\) were used and the exact ansatz energy is \(-48.0625\). Note that separate experiments (results not shown in plots) with 40-qubit Bell pairs gives a naive extrapolation of the Bell pairs energy to be \(-67.0(4.0)\), whose ideal value should be \(-60\). The mitigated values with the reference state for (e) and (f) are \(-15.4 \pm 0.7\) and \(-46.1 \pm 2.4\), respectively.
We note that even for \( N = 3 \) electrons, Laughlin’s droplet state is not an exact wavefunction \([29, 35]\), but Laughlin’s wavefunctions are the key to understand the fractional quantum effect. That variational \( \text{ansätze} \) contain the exact ground states is a desired feature, as one can use it to ascertain optimality of variational parameters.

In Table 1, we show the optimized parameters, energy, and overlap with MPS diagonalized ground-state wave function using one layer of our \( \text{ansatz} \) for the Heisenberg model. We check that the results agree with the exact computation for the qubit number \( N \leq 12 \). As expected, the fidelity decreases rapidly with the number of qubits, similar to the case of the Laughlin wave function with many electrons. In contrast, the approximate ground-state energy seems to reach about 3% of error even for large chains using just one layer in the \( \text{ansatz} \) (e.g. 3.33% even for \( N = 100 \)). Using also multiple (up to six) layers of the \( \text{ansatz} \), we compare its ground-state fidelity and the relative error in the ground-state energy in Fig. 1cd with the qubit number ranging from 4 to 50. Although we cannot assert that the solutions we found are the absolute optimum, it is obvious that by increasing the number of layers in the \( \text{ansatz} \), the fidelity with the ground state (e.g. above 0.9 for \( N = 50 \) and six layers) and the accuracy in the ground-state energy (e.g. above 99.75%) improve substantially. We also perform a similar multi-layer analysis for the case of the periodic boundary condition, whose results are shown in Fig. S.9 in the Appendix.

**Experiments and results**

We mainly use the Bell-state measurement approach to measure the system’s energy, as it requires the least resource compared to two other approaches; see their comparison in the Materials and Methods section in Appendix. This approach takes advantage of the specific form of the XXZ Hamiltonian, and we have performed experiments on nine different backends, which contain 27, 65 and 127 qubits on three different layouts, as illustrated in Fig. S.6 and whose properties are listed in Table S.1. The results on the Heisenberg spin chains (some averaged over different backends) are also included in Table 1, ranging from 4 to 102 qubits, and are compared to the numerical simulations; their relative errors with the ground-state energy values are within a few percentages. These experimental data can then be used to extract the energy per site in the thermodynamic limit, as shown in Fig. S.4, which agrees with the exact Bethe ansatz calculation within 3.4% of deviation. For the purpose of demonstration, we mostly use the numerically optimized parameters to run the state creation circuits. But still, we also test the feasibility of the hybrid classical-quantum approach by performing experiments with random parameters.

In Fig. 2, we present exemplary experimental data with readout mitigation \([37, 36, 38, 39, 40, 41]\) applied for both the optimal one-layer \( \text{ansatz} \) and the Bell pairs, where we have also performed the forward \( U \) and backward \( U^{-1} \) repetitions of the circuits before a final forward \( U \) circuit (up to 4 or
5 repetitions) in order to carry out gate-error mitigation \cite{10, 11, 12, 13}; for details of readout- and gate-error mitigations, see the **Materials and Methods** section in Appendix. This allows us to apply our proposed reference-state zero-noise extrapolation (rZNE), illustrated with the 102-qubit Heisenberg chain on ibm\textunderscore washington, whose layout is also illustrated in Fig. 1b, and with the 50-qubit case on ibmq\textunderscore brooklyn. From the 102-qubit experimental data in Fig. 2ab, we can extrapolate the ansatz and Bell pairs’ energy to \( m = 0 \) and obtain the respective energy values, -199.2 and 169.8, where \( m \) is the total number of \( U \) and \( U^{-1} \) in the circuit. Combining the extrapolated Bell pairs’ energy value \(-169.8\) and the ideal value \(-153\) allows us to naively correct the extrapolated ansatz energy from \(-199.2\) to a value \(-199.2/(169.8/153) \approx -179\), which is close to the numerical value \(-174.04\) by the MPS algorithm. We note that in this set of experiments, the maximal CNOT depth is 63, and the total number of CNOTs used is 3186. Similar experimental results on a set of 50 qubits in ibmq\textunderscore brooklyn are presented in Fig. 2cd. In the **Materials and Methods** section of Appendix, we also provide a slightly more sophisticated estimation, and the results may differ but usually not by much for the Heisenberg model, and an empirical rule that the two approaches agree very well is when the two curves from the ansatz energy and the Bell pairs energy can be rescaled and well collapsed into a single one. We refer to Tables S.3 and S.2 in the Appendix for a comprehensive list of mitigated results. These backends possess different qubit numbers, quantum volumes, and noise and error rates (see in Table S.1), but the success across all these backends demonstrates the utility of such a simple and scalable rZNE approach.

To illustrate a proof-of-principle demonstration of the potential hybrid quantum-classical approach, we have performed two additional sets of experiments on 40 qubits with random parameters in the ansatz, and using our rZNE method, the mitigated energy values agree very well with the numerically calculated values; see Fig. 2ef. This demonstrates that it is feasible to use quantum devices to measure the expectation values accurately (mitigated via the rZNE) and, based on these, estimate the next iteration of the variational parameters by classical computers. The mitigated VQE can potentially become more practical and accurate as the processors’ speed, their qubit number, and error rates continue to improve. We note that the rZNE can also be developed at the pulse level and other sophisticated error mitigation and protection/correction schemes could be also incorporated to improve accuracy, especially at a stage close to convergence.

In Fig. 3 we present the two sets of experimental results for the ansatz energy for a wide range of \( \Delta \in [-0.8, 1.4] \) in the XXZ model with \( N = 40 \) and \( N = 80 \) spins, respectively. The experiments were performed on two separate backends, ibmq\textunderscore brooklyn and ibm\textunderscore washington. The mitigated values agree well with the anticipated ansatz values and fall within several percentages of the numerical ground-state energy values obtained from the MPS approach.
Summary

In this work, we have demonstrated that variational quantum algorithms with short-depth circuits could be applied to large systems of qubits, with up to 102 qubits performed on real devices. Despite the presence of substantial noise and errors in current devices, we have been able to improve and implement efficient error mitigation schemes to deduce accurate ground-state energy from experiments on large systems, including the use of reference states for zero-noise extrapolation, i.e., the rZNE technique. Our work thus opens up the potential practical use of error mitigated VQE on large quantum computing backends for improved accuracy. One first applies our rZNE (combining readout mitigation and possibly further mitigation) to obtain the extracted observable value(s) and/or its gradients from quantum devices, then uses classical computers to search for variational parameters to be used in the subsequent iteration of experiments with mitigation. The procedure is iterated until the mitigated observable value(s) converge to within certain accuracy. Such an error-mitivated, rZNE VQE approach, though not yet practical for large systems in the current cloud-based setting, due to limited allocated time and long job queues, seems plausible in dedicated experiments. Our experiments using randomly chosen parameters already demonstrated agreement with the expected ansatz energy. To enter a regime where quantum advantage may be realized, we will likely need to go beyond one dimension, e.g., two dimensions, where classical simulations of quantum many-body systems become intractable as the system size increases. Toward this goal, we have also analyzed a two-leg ladder and showed the applicability of our ansatz (see Figure 3: The energy results for XXZ spin chains. (a), A 40-qubit XXZ chain on ibmq_brooklyn. The energy is obtained using the Bell measurement approach (labelled ‘Bell’) on physical qubits [38,41,42,43,52,56,57,58,59,60,53,47,46,45,39,31,30,29,24,15,16,17,11,4,5,6,7,8,12,21,20,19,25,33,34,35,40,49,50,51]. (b), An 80-qubit XXZ chain on ibm_washington, with physical qubits being [97,96,95,94,90,75,76,77,71,58,57,56,52,37,38,39,33,20,21,22,23,24,25,26,27,28,29,30,31,32,36,51,50,49,48,47,46,45,44,43,42,41,53,60,61,62,63,64,65,66,67,68,69,70,74,89,88,87,93,106,107,108,112,126,125,124,123,122,111,104,103,102,101,100,110,118,117,116,115,114]; see Fig. 1b for the layout of ibm_washington and Fig. S.6 for ibmq_brooklyn.
A recent analogue quantum simulation experiment on ultra-cold atoms was carried out to realize such a two-leg ladder with different intra- and inter-leg Heisenberg couplings, and the Haldane phase was realized \cite{42}. Future work will be extended to these two- and multiple-leg ladders as well as two-dimensional systems on digital quantum simulation.

References

[1] P. Benioff, The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. Journal of Statistical Physics, 22, 563–591 (1980).

[2] Y. I. Manin, Vychislimoe i nevychislimoe [Computable and Noncomputable] (in Russian). Sovetskoe Radio (1980).

[3] R. P. Feynman, Simulating physics with computers. International Journal of Theoretical Physics, 21, 467–488 (1982).

[4] S. Lloyd, Universal Quantum Simulators. Science 273, 1073-1078 (1996).

[5] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring. In Proceedings 35th Annual Symposium on Foundations of Computer Science (pp. 124–134). IEEE (1994).

[6] J. Preskill, Quantum Computing in the NISQ era and beyond. Quantum 2, 79 (2018).

[7] S. Endo, S. C. Benjamin, Y. Li, Practical Quantum Error Mitigation for Near-Future Applications. Phys. Rev. X 8, 031027 (2018).

[8] K. Temme, S. Bravyi, J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits. Phys. Rev. Lett. 119, 180509 (2017).

[9] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor. Nature, 567, 491–495 (2019).

[10] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, P. Lougovski, Cloud Quantum Computing of an Atomic Nucleus. Phys. Rev. Lett. 120, 210501 (2018).

[11] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, M. J. Savage, Quantum-classical computation of Schwinger model dynamics using quantum computers. Phys. Rev. A 98, 032331 (2018).
[12] M. Urbanek, B. Nachman, V. R. Pascuzzi, A. He, C. W. Bauer, W. A. de Jong, Mitigating Depolarizing Noise on Quantum Computers with Noise-Estimation Circuits. Phys. Rev. Lett. 127, 270502 (2021).

[13] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation. Preprint at http://arXiv.org/abs/2005.10921 (2020).

[14] J. J. Wallman, J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling. Phys. Rev. A 94, 052325 (2016).

[15] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, Scalable error mitigation for noisy quantum circuits produces competitive expectation values. Preprint at https://arxiv.org/abs/2108.09197 (2021).

[16] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms. Rev. Mod. Phys. 94, 015004 (2022).

[17] J. Preskill, Quantum computing and the entanglement frontier, Preprint at https://arxiv.org/abs/1203.5813 (2012).

[18] F. Arute et al. Quantum supremacy using a programmable superconducting processor. Nature 574, 505-510 (2019).

[19] Y. Wu et al. Strong Quantum Computational Advantage Using a Superconducting Quantum Processor. Phys. Rev. Lett. 127, 180501 (2021).

[20] H.-S. Zhong et al. Quantum computational advantage using photons. Science 370, 1460-1463 (2020).

[21] L. S. Madsen et al. Quantum computational advantage with a programmable photonic processor. Nature 606, 75–81 (2022).

[22] F. Arute et al. Hartree-Fock on a superconducting qubit quantum computer. Science 369, 1084-1089 (2020).

[23] K. J. Satzinger et al. Realizing topologically ordered states on a quantum processor. Science 374, 1237–1241 (2021).
[24] M. Gong et al. Quantum walks on a programmable two-dimensional 62-qubit superconducting processor. Science 372, 948-952 (2021)

[25] W. Heisenberg, Zür Theorie des Ferromagnetismus (On the theory of ferromagnetism). Zeitschrift für Physik 49, 619–636 (1928).

[26] P. W. Kasteleijn, The lowest energy state of a linear antiferromagnetic chain. Physica 18, 104-113 (1952).

[27] V. E. Korepin, N. M. Bogoliubov, A. G. Izergin, Quantum Inverse Scattering Method and Correlation Functions, Cambridge University Press (Cambridge, UK, 1993).

[28] J. Bardeen, L. N. Cooper, J. R. Schrieffer, Theory of Superconductivity. Phys. Rev. 108 (5), 1175–1204 (1957).

[29] R. B. Laughlin, Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations. Phys. Rev. Let. 50 (18), 1395–1398 (1983).

[30] M. Nielsen, I. Chuang, Quantum Computation and Quantum Information (Cambridge Univ. Press, 2000).

[31] F. Verstraete, J. I. Cirac, V. Murg, Matrix Product States. Projected Entangled Pair States, and variational renormalization group methods for quantum spin systems. Adv. Phys. 57, 143 (2008).

[32] S. Östlund, S. Rommer, Thermodynamic Limit of Density Matrix Renormalization. Phys. Rev. Lett. 75, 3537 (1995).

[33] M. Fannes, B. Nachtergaele, R. F. Werner, Finitely correlated states on quantum spin chains. Commun. Math. Phys. 144, 443 (1992).

[34] S. R. White, Density matrix formulation for quantum renormalization groups. Phys. Rev. Lett. 69, 2863 (1992).

[35] F. D. M. Haldane, Fractional Quantization of the Hall Effect: A Hierarchy of Incompressible Quantum Fluid States, Phys. Rev. Lett. 51, 605 (1983).

[36] Y. Chen, M. Farahzad, S. Yoo, T.-C. Wei, Detector Tomography on IBM 5-qubit Quantum Computers and Mitigation of Imperfect Measurement, Phys. Rev. A 100, 052315 (2019).
[37] A. C. Keith, C. H. Baldwin, S. Glancy, E. Knill, Joint quantum-state and measurement tomography with incomplete measurements. Phys. Rev. A 98, 042318 (2018).

[38] M. R. Geller, M. Sun, Efficient correction of multiqubit measurement errors. Quantum Sci. Technol. 6, 025009 (2021).

[39] F. B. Maciejewski, Z. Zimborás, M. Oszmaniec, Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. Quantum 4, 257 (2020).

[40] S. Bravyi, S. S. Sheldon, A. Kandala, D. C. Mckay, J. M. Gambetta, Mitigating measurement errors in multi-qubit experiments. Phys. Rev. A 103, 042605 (2021).

[41] P. D. Nation, H. Kang, N. Sundaresan, J. M. Gambetta, Scalable mitigation of measurement errors on quantum computers. Preprint at https://arXiv.org/abs/2108.12518 (2021).

[42] P. Sompet, S. Hirthe, D. Bourgund, T. Chalopin, J. Bibo, J. Koepsell, P. Bojović, R. Verresen, F. Pollmann, G. Salomon, C. Gross, T. A. Hilker, I. Bloch, Realising the Symmetry-Protected Haldane Phase in Fermi-Hubbard Ladders. Preprint at https://arXiv.org/abs/2103.10421 (2021).

Additional References Used in the Appendix

[43] E. Farhi, J. Goldstone, S. Gutmann, A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem. Preprint at https://arXiv.org/abs/1412.6062 (2014).

[44] F. Vatan, C. Williams, Optimal Quantum Circuits for General Two-Qubit Gates. Phys. Rev. A 69, 032315 (2004).

[45] D. Wierichs, J. Izaac, C. Wang, C. Y.-Y. Lin, General parameter-shift rules for quantum gradients. Preprint at https://arXiv.org/abs/2107.12390.

[46] J. Li, X. Yang, X. Peng, C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control. Phys. Rev. Lett. 118, 150503 (2017).

[47] K. Mitarai, M. Negoro, M. Kitagawa, K. Fujii, Quantum circuit learning. Phys. Rev. A 98, 032309 (2018).

[48] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, N. Killoran, Evaluating analytic gradients on quantum hardware. Phys. Rev. A 99, 032331 (2019).
[49] O. Kyriienko, V. E. Elfving, Generalized quantum circuit differentiation rules. Phys. Rev. A 104, 052417 (2021).

[50] W. K. Wootters, Entanglement of Formation of an Arbitrary State of Two Qubits. Phys. Rev. Lett. 80, 2245 (1998).

[51] L. K. Shalm et al. Strong loophole-free test of local realism. Phys. Rev. Lett. 115, 250402 (2015).

[52] S. L. Todaro, V. B. Verma, K. C. McCormick, D. T. C. Allcock, R. P. Mirin, D. J. Wineland, S. W. Nam, A. C. Wilson, D. Leibfried, D. H. Slichter, State Readout of a Trapped Ion Qubit Using a Trap-Integrated Superconducting Photon Detector. Phys. Rev. Lett. 126, 010501 (2021).

[53] R. Gehr, J. Volz, G. Dubois, T. Steinmetz, Y. Colombe, B. L. Lev, Long, R., Estève, J. & Reichel, J. Cavity-Based Single Atom Preparation and High-Fidelity Hyperfine State Readout. Phys. Rev. Lett. 104, 203602 (2010).

[54] J. Yoneda, K. Takeda, A. Noiri, T. Nakajima, S. Li, J. Kamioka, T. Kodera, S. Tarucha, S. Quantum non-demolition readout of an electron spin in silicon, Nature Communications 11, 1144 (2020).

[55] Q. Zhang, Y. Guo, W. Ji, M. Wang, J. Yin, F. Kong, Y. Lin, C. Yin, F. Shi, Y. Wang, J. Du, High-fidelity single-shot readout of single electron spin in diamond with spin-to-charge conversion. Nature Communications 12, 1529 (2021).

[56] F. Lecocq, F. Quinlan, K. Cicak, J. Aumentado, S. A. Diddams, J. D. Teufel, Control and readout of a superconducting qubit using a photonic link. Nature 591, 575–579 (2021).

[57] B. Abdo, O. Jinka, N. T. Bronn, S. Olivadese, M. Brink, High-Fidelity Qubit Readout Using Interferometric Directional Josephson Devices, PRX Quantum 2, 040360 (2021).

[58] A. Luis, L. L. Sanchez-Soto, Complete characterization of arbitrary quantum measurement processes. Phys. Rev. Lett. 83, 3573–3576 (1999).

[59] J. Fiurasek, Maximum-likelihood estimation of quantum measurement. Phys. Rev. A 64, 024102 (2001).

[60] E. Magesan, J. M. Gambetta, J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes. Phys. Rev. Lett. 106, 180504 (2011).

[61] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, J. M. Gambetta, Validating quantum computers using randomized model circuits. Phys. Rev. A 100, 032328 (2019).
[62] J. Zhang, T.-C. Wei, R. Laflamme, Experimental Quantum Simulation of Entanglement in Many-Body Systems. Phys. Rev. Lett. 107, 010501 (2011).

[63] S. S. Roy, H. S. Dhar, D. Rakshit, A. Sen(De), U. Sen, Detecting phase boundaries of quantum spin-1/2 XXZ ladder via bipartite and multipartite entanglement transitions. Journal of Magnetism and Magnetic Materials 444, 227-235 (2017).

Acknowledgements

The research in this work on physical and entanglement properties of the XXZ model and the development of the associated the toolkit for measuring and extrapolating them was supported by the National Science Foundation under Grant No. PHY 1915165 (T.-C.W.). The research on the design and analysis of the variational ansatz in this work was supported by the U. S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under contract number DE-SC0012704 (H.Y. and T.-C.W.). This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, and the Brookhaven National Laboratory operated IBM-Q Hub. The results presented in this work do not reflect the view of IBM and its employees.
Appendix

Materials and Methods

Ansatz structure from an adiabatic Hamiltonian

For the XXZ interaction $\hat{H}_{XXZ} (\Delta)$ on a bond involving two nearest-neighbor qubits, the singlet pair $|\Psi^-\rangle = (|01\rangle - |10\rangle) / \sqrt{2}$ has an energy value $-2 - \Delta$, the triplet $|\Psi^+\rangle = (|01\rangle + |10\rangle) / \sqrt{2}$ has energy $2 - \Delta$, and both $|00\rangle$ and $|11\rangle$ (or equivalently the two entangled triplets $|\Phi^\pm\rangle = (|00\rangle \pm |11\rangle) / \sqrt{2}$) have energy $\Delta$. Note that we have used the notation $|0/1\rangle$ to replace $|\uparrow/\downarrow\rangle$, the eigenstates of the Pauli Z operator $\sigma_z$. Thus, the singlet is the ground state of the simple two-qubit XXZ interaction for $\Delta > -1$. This means that the following Hamiltonian with interaction only on odd numbers of bonds only even $N$,

$$\hat{H}_{\text{odd}} = \sum_{j=1}^{N/2-1} \left( \sigma_x^{[2j-1]} \sigma_x^{[2j]} + \sigma_y^{[2j-1]} \sigma_y^{[2j]} + \Delta \sigma_z^{[2j-1]} \sigma_z^{[2j]} \right),$$

(S.1)

has its unique ground state being the product of singlets over these odd bonds, i.e., a linear chain of valence-bond state,

$$|\psi_{\text{singlets}}\rangle = \frac{1}{\sqrt{2}^{N/2}} \prod_{j=1}^{N/2} (|01\rangle - |10\rangle)_{2j-1,2j}.$$  

(S.2)

We then expect that $|\psi_{\text{singlet}}\rangle$ is adiabatically connected to the ground state of the XXZ model, by connecting $\hat{H}_{\text{odd}}$ to the full XXZ Hamiltonian $\hat{H}_{XXZ}$ via the following linear interpolation,

$$\hat{H}(s) = (1 - s) \hat{H}_{\text{odd}} + s \hat{H}_{XXZ} = \hat{H}_o(s) + \hat{H}_e(s).$$  

(S.3)

We regroup it into interaction terms on even and odd bonds, denoted collectively by $\hat{H}_o(s)$ and $\hat{H}_e(s)$, respectively, where $\hat{H}_o(s) = \hat{H}_{\text{odd}}$, but

$$\hat{H}_e(s) = s \sum_{j=1}^{N/2-1} \left( \sigma_x^{[2j]} \sigma_x^{[2j+1]} + \sigma_y^{[2j]} \sigma_y^{[2j+1]} + \sigma_z^{[2j]} \sigma_z^{[2j+1]} \right),$$

(S.4)

is a rescaled version of the XXZ model on even bonds.

We check the spectral properties of this Hamiltonian for small $N$ and find that $\hat{H}(s)$ is gapped in the range $s \in [0, 1]$ for $\Delta > -1$; see Fig. S.1. This means that the product of singlets $|\psi_{\text{singlet}}\rangle$ is adiabatically connected to the ground state of the XXZ model via the evolution $|\psi(1)\rangle = U_{\text{evo}} |\psi_{\text{singlets}}\rangle = e^{-i \int_0^1 ds \hat{H}(s)} |\psi_{\text{singlets}}\rangle$. Discretizing the evolution operator $U_{\text{evo}}$, we have the following Trotterized approximation

$$U_{\text{evo}} \approx \prod_{l=1}^{N_L} e^{-i \hat{H}(s_l) \delta s} \approx \prod_{l=1}^{N_L} \left( e^{-i \hat{H}_o(s_l) \delta s} e^{-i \hat{H}_e(s_l) \delta s} \right),$$

(S.5)
where $N_L$ is the number of discretized time steps or layers and $\delta_s = 1/N_L$ is the dimensionless step size. To allow for flexibility, we turn the discretized evolution into a variational form and arrive at the structure of the ansatz shown in Fig. 1a with gates in the $l$-th layer being

$$U^{(l)}_{\text{even/odd}}(\{\theta\}) = \bigotimes_{j \in \text{even/odd}} \left[ e^{-i\theta \sigma_z^{(l)}\sigma_z^{(j+1)}} - i\theta \sigma_y^{(l)}\sigma_y^{(j+1)} e^{-i\theta \sigma_z^{(l)}\sigma_z^{(j+1)}}\right],$$

(S.6)

where $\{\theta\}'s$ are a set of variational parameters, the subscripts $e/o$ denote the association with even and odd bonds, respectively. Therefore, we arrive at the following $N_L$-layer variational ansatz state

$$|\psi_{\text{ansatz}}(\{\theta\})\rangle = N_L \bigotimes_{l=1}^{N_L} \left[ U_{\text{even}}^{(l)}(\{\theta_e\}) U_{\text{odd}}^{(l)}(\{\theta_o\}) \right] |\psi_{\text{singlets}}\rangle.$$  

(S.7)

We remark that our construction is similar to how the Quantum Approximate Optimization Algorithm (QAOA) ansatz originates from discretizing the Ising interaction and the transverse field [43]. But they differ in the goal: the QAOA aims to minimize the energy of a classical Ising Hamiltonian using the transverse-field part as a driver, whereas our goal is to optimize the energy of an intrinsically quantum Hamiltonian as a whole.

**Creation of singlets**

Each singlet pair in $|\psi_{\text{singlets}}\rangle$ can be created from $|00\rangle$ by simple single-qubit gates (the Hadamard $H$ and the Pauli $X$ gates) followed by a CNOT gate [30],

$$|\Psi^{-}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = U_{\text{singlet}}|00\rangle = \begin{array}{c} \text{X} \\ \text{H} \end{array}$$

and, thus, the product of such singlet pairs can be created in parallel with these circuits,

$$|\psi_{\text{singlets}}\rangle = U_{\text{init}}|0\ldots0\rangle = \bigotimes_{i=1}^{N/2} U_{\text{singlet}}^{[i]}|0\ldots0\rangle,$$

(S.9)

with the superscript $i$ denoting the pair of qubits for the singlet creation. We note that the reverse of the latter part corresponds to Bell measurement,

$$\begin{array}{c} \text{X} \\ \text{H} \end{array}$$

(S.10)

which can be used to measure the energy contribution of a pair of qubits.
Gate decomposition

To explain the gate decomposition, we consider the following two-qubit gates (where we have used the standard notation $X \equiv \sigma_x$, $Y \equiv \sigma_y$ and $Z \equiv \sigma_z$ in quantum information),

\[
R_{zz}(\theta_z) = e^{-i\frac{\theta_z}{2}Z \otimes Z},
\]
\[
R_{xx}(\theta_x) = e^{-i\frac{\theta_x}{2}X \otimes X} = H \otimes HR_{zz}(\theta_x)H \otimes H,
\]
\[
R_{yy}(\theta_y) = e^{-i\frac{\theta_y}{2}Y \otimes Y} = R_x\left(-\frac{\pi}{2}\right) \otimes R_x\left(-\frac{\pi}{2}\right)R_{zz}(\theta_y)R_x\left(\frac{\pi}{2}\right) \otimes R_x\left(\frac{\pi}{2}\right)
\]
\[
= (SH) \otimes (SH)R_{zz}(\theta_y)(HS^\dagger) \otimes (HS^\dagger),
\]

where $R_{\alpha}(\theta) = e^{-i\theta\sigma_{\alpha}/2}$ is the single-qubit rotation around $\alpha$-axis ($\alpha = x, y, z$) by an angle $\theta$, and $S$ is the one-qubit phase gate $S = e^{i\pi/4}R_z(\pi/2)$ \[30\]. The above two-qubit gates can each be implemented by the following circuits with the combination of one-qubit rotation gates and two-qubit CNOTs \[30\].

![Circuit Diagram](image)

Now let us define the essential two-qubit $R_{xyz}$ gate that we need,

\[
R_{xyz}(\theta_x, \theta_y, \theta_z) \equiv e^{-i(\theta_x/2)\sigma_x \otimes \sigma_x - i(\theta_y/2)\sigma_y \otimes \sigma_y - i(\theta_z/2)\sigma_z \otimes \sigma_z},
\]

where a factor of 1/2 is inserted in the definition of the $R_{xyz}$ gate to match the convention of single-qubit rotation and we have used the tensor product notation ‘$\otimes$’ to emphasize the two-qubit structure in the gate. If we naively cascade the above three $R_{xx}$, $R_{yy}$ and $R_{zz}$ gates, it will requires six CNOTs. However, we find that $R_{xyz}(\theta_x, \theta_y, \theta_z) = R_{xx}(\theta_x)R_{yy}(\theta_y)R_{zz}(\theta_z)$ can be simplified from the naive 6-CNOT circuit by cancelling out two pairs of CNOTs after simple circuit identities to one with only 3 CNOTs,

![Circuit Diagram](image)
Thus, the gate $U_{\text{even}/\text{odd}}(\theta) = R_{xyz}(2\theta, 2\theta, 2\theta)$ will be used for the Heisenberg model, and for the XXZ model, due to the ZZ anisotropy, we will allow $\theta_z = 2\theta_2$ parameter to be independent from $\theta_x = \theta_y = 2\theta_1$, and thus $R_{xyz}(2\theta_1, 2\theta_1, 2\theta_2)$ is needed.

Note that as the circuit action is symmetric with respect to swapping the two qubits, one can flip the circuit in the last line to fit the desired or natural direction of the CNOT gate. One can also replace the Hadamard gate $H$ by a combination of the square root of $X$ gate (or equivalently $R_x(\pi/2)$, which is among the native gates in IBM Quantum Computers), and the phase gate $S$ via the identity $H = SR_x(\pi/2)S$. Note that $S$ is equivalent to $R_z(\pi/2)$ up to an irrelevant global phase factor, and, therefore, the circuit can be expressed entirely in terms of IBM Q’s native gate set: $\{R_z, S_X, \text{CNOT}, X\}$. Note that the variational parameters used via our Rxyz gate reside in the single-qubit rotation, so we can employ the parameter shift rule and its generalization to evaluate the gradient of the circuit, especially on real devices; see e.g. Refs. [45, 46, 47, 48, 49].

**Measuring energy: other approaches**

Here we describe three approaches that we have used to measure the energy expectation value. Ideally the three approach should give the same results and we have indeed tested all three experimentally and verified that they give the same results within a few percentages of errors for small systems; see Fig. S.2.

**1. Tomography-based approach.** For the models we consider, the Hamiltonian terms are of the form $\sigma_{i,\alpha}^{[j]}\sigma_{i,\alpha}^{[j+1]}$, where $j$ is the site number and $\alpha$ is the spin direction (x, y or z). Naively, if we can obtain the reduced density $\rho_{j,j+1}$ matrix for the pair $(j, j+1)$ then we can calculate the energy contribution from $\text{Tr}(\rho_{j,j+1}\sigma_{i,\alpha}^{[j]}\sigma_{i,\alpha}^{[j+1]})$. But this requires state tomography and seems to need to run $9N_{\text{bond}}$ different circuits for the total energy, where $N_{\text{bond}}$ is the number of nearest pairs or bonds, e.g. $N - 1$ for an open chain and $N$ for a periodic one. However, we could improve the efficiency by performing the state tomography in parallel. Doing so, we just need two sets of state tomography circuits (for even and odd bonds respectively) to obtain the reduced density matrices of neighboring pairs of qubits. This is made possible by the assumption that measurement is independent on whether other qubits are measured or not. However, in real devices, there may be crosstalk [36]. We will later discuss the measurement mitigation on pairs of qubits associated with bonds in order to extract reliable energy contribution. Doing the tomography in parallel reduces the number of circuits to measure to $9 \times 2$, which is independent of the...
model size. The benefit of this is that any one- and two-qubit observables are readily available, such as the local spin observables and the concurrence [50] which quantifies nearest-neighbor entanglement, and it applies to all nearest-neighbor interacting spins. (Extension to finite-ranged interaction is straightforward but requires more sets of measurements and multi-qubit state tomography.) For our experimental results on measuring the concurrence in a 8-qubit XXZ chain, see Fig. S.3.

(2) **XYZ measurement.** The second, slightly reduced measurement method is to measure separately the two neighboring qubits on each bond in basis $\alpha$ and then average over the classical results $\sigma_\alpha[k] \sigma_\alpha[k+1]$, treating $\sigma_\alpha = \pm 1$ from the measurement outcome assignment. This naively requires $3N_{\text{bond}}$ different circuits for the total energy. But a much simplified implementation is to measure all qubits in $\alpha$ basis and calculate the average $\sigma_\alpha[k] \sigma_\alpha[k+1]$ for all bonds. This only requires 3 different measurement settings to obtain the total energy. Such simplification is applicable for the same reason mentioned in the previous approach (1).

(3) **Bell measurement.** There is another method that uses Bell-state measurement on all bonds. As explained above, for the XXZ interaction on a bond, $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ has energy $-2 - \Delta$, the triplet $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ has energy $2 - \Delta$, and both $|00\rangle$ and $|11\rangle$ (or equivalently $|\Phi^\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$) have energy $\Delta$. We can identify the Bell state on a particular bond with Bell measurement. The energy contribution of that bond is the energy corresponding to the Bell state obtained from the measurement outcome. The total energy can be calculated by adding up every bond’s energy contribution. In practice, the Bell-state measurement requires a short two-qubit circuit including a CNOT gate, shown above, which may have a larger error rate than single-qubit gates. However, the effect could be merged with and considered as part of the measurement error. We describe a mitigation method that corrects for the CNOT gate below.

Naively, this approach of measuring energy requires $N_{\text{bond}}$ different measurement patterns appended at the end of state creation as readout for the total energy if each bond is measured separately. However, we can divide the bonds into even and odd groups, as above, and can perform the Bell measurement in parallel within each group. Then we only need to perform two different sets of measurements. This turns out to be the approach we used to perform large-system (up to $N = 102$ qubits) experiments on real devices to obtain the total energy.

**Error mitigation**

**Measurement/Readout Error Mitigation.** For certain qubit systems, such as photon polarizations [51], trapped ions [52] and neutral atoms [53], readout fidelity can be as good as 99% or higher. For others, such as solid-state-based qubits, the readout fidelity is lower but keeps improving [54, 55, 56, 57]. For
superconducting qubits, the readout error can be as large as 10% or more \[36\]. It is therefore crucial to mitigate the measurement error in order to calculate the correct energy of the created state on the real device. Due to the expanding deployment of cloud quantum computers, the interest in the issue of state preparation and readout error has recently been rekindled \[37, 36, 38, 39, 40, 41\]. The key idea is to first characterize the measurement pattern dependent on the state input, such as from the detector tomography \[58, 59\] or simply measuring the probability matrix \(M\) that relates the input states to the measured outcomes, i.e. \(\vec{P}_{\text{measured}} = M\vec{P}_{\text{ideal}}\) \[36, 39\], where \(\vec{P}_{\text{measured}}\) and \(\vec{P}_{\text{ideal}}\) represent respectively the measured and ideal probability distribution. By properly inverting the relation with the constraint that the outcome distribution \(P_{\text{ideal}}\) be non-negative, one can obtain the mitigated distribution to evaluate the observables.

For \(N\) qubits, the complete matrix \(M\) is of size \(2^N \times 2^N\) and requires preparation of \(2^N\) computational states, thus is not efficient and is only doable for a small number of qubits. As the models we consider here contain only nearest-neighbor interactions, we are mainly concerned with measurement mitigation for pairs of qubits in a bond, i.e., involved in the interacting Hamiltonian, and such simplification allows us to deal with large systems in a practical way. We can perform readout mitigation pairwise for the nearest-neighbor two qubits on all bonds. Similar to the energy measurement, this can be reduced to two sets of mitigation, i.e., on pairs of even and odd bonds, respectively. Each mitigation requires 4 different inputs from all two-qubit computational basis and measurement in the same basis gives rise to a \(4 \times 4\) matrix \(M\), which we can then use to infer the ideal two-qubit measurement distribution so as to obtain the mitigated energy contribution.

**Bell-measurement Mitigation.** In our experiments with large numbers of qubits, the local energy for a pair of qubits is obtained by measuring in the Bell-state basis, which uses an inverse circuit for Bell-state preparation and involves CNOT gates. To mitigate potential errors caused by imperfect CNOT gates, we adopt the above readout mitigation for the Bell measurement. Specifically, for each pair in the bonds, we prepare the four Bell states and then immediately measure qubits pairwise in the Bell-state basis, such as the circuit shown previously, to obtain a \(4 \times 4\) Bell-state assignment matrix \(M_{\text{Bell}}\) for each pair. With this we can mitigate the outcome distribution and hence the energy value obtained from the Bell-state measurement.

**Gate Error Mitigation.** By doing readout mitigation we are probing the properties of the state actually created on the quantum devices. However, the observable expectation is affected by gate errors as well that prevent us from obtaining the idealized value. In order to estimate the latter, prior works have considered pulse and gate error mitigation by extrapolating to the zero-error limit \[7, 8, 9\], and this is an extrapolation of the physical observables, not the actual observable values associated with the quantum
states created. Nevertheless, it is still important to see how well quantum computers can estimate these values despite the noise and errors, especially in the regime where direct classical calculations might not be feasible.

However, in order to perform accurate gate mitigation, one needs to have substantial access to the hardware performing pulse-level optimization and operations [9], which is still not practical for dealing with a large number of qubits. (Note that recent experiments have been carried on 26 qubits using pulse-level zero-noise extrapolation [15].) Instead, we will use the gate-level zero-noise extrapolation (ZNE) approach discussed in Refs. [10, 11, 12, 13]. In particular, our approach builds on the idea in Ref. [13] and we prepare the circuits to create $|\psi_n\rangle = U(U^{-1}U)^n|0...0\rangle$, where $n$ is a non-negative integer and $U = U_{\text{var}}(\{\theta\})U_{\text{init}}$, as in Eq. (2), denotes the circuit to prepare the ansatz state from the fiducial state $|0...0\rangle$, i.e., $|\psi_{\text{ansatz}}\rangle = U(\{\theta\})|0...0\rangle$, and then use several forward-backward repetitions in $U$ to evaluate the observable $O_n = \langle \psi_n | \hat{O} | \psi_n \rangle$, as a function of $n$. Ideally, different $n$ should give the same state and hence the same value for the observable $\hat{O}$. However, noise and errors spoil this and the state with larger $n$ should be noisier. The extrapolation to the gate-level zero-noise limit is done by a fitting to $O_n$ with $m = 2n + 1 \to 0$ limit.

Reference-state Gate Error Mitigation. Building on this, we propose to use a reference state (or possibly multiple ones), which is contained in the ansatz family, for example, the product of Bell pairs that we use below (via setting all $\theta$’s to zero, i.e., $|\psi_{\text{singlets}}\rangle = U(\{\theta\} = 0)|0...0\rangle$), with a known exact energy value, to improve the extrapolation of the energy value or other observables. Running the energy experiment for this reference state with the above gate-level mitigation, we obtain the naively-extrapolated experimental value and hence the possible mismatch with the exact value. Using such knowledge for the reference, we can estimate the expected value of the ansatz state from the naive experimental value. Combining both gate and readout error mitigation, we are able to reach the accuracy of the extrapolated energy with a few percentages of the exact value for all ranges of the qubit number that we have tested on real devices. We expect that this reference-state ZNE (rZNE) may be applied to the general VQE platform. It does not require additional circuits from randomized compiling, as done, e.g., in Ref. [12, 14], but averaging the results from these randomized circuits can be used to further improve the accuracy.

**Extrapolating the ansatz energy with a reference state.** In Fig. 2 we present experimental data with readout mitigation for both the optimal one-layer ansatz and the Bell pairs, where we have also performed the forward and backward repetitions of the circuits before a final forward circuit (up to 4, 5 or sometimes 6 repetitions). We illustrate the naive zero-noise extrapolation (ZNE) in Fig. 2 with the 102-qubit Heisenberg chain on ibm\_washington and the 50-qubit case on ibmq\_brooklyn. Note that
the statistical errors for each data point of the energy obtained from the experiments are very small, with the largest being 0.21 (each data point is obtained from 50 repetitions of the same circuit with 40K shots of measurement). In fitting the energy data, we use an exponential function $f_E(m) = a \exp(-bm) + c$, where $m = 2n + 1$ is the total number of $U$ or $U^{-1}$ in the circuit to prepare the ansatz state. For most experiments we used $n$ from 0 to 4; for some we had $n$ from 0 to 5; for few others we had $n$ from 0 to 3 (due to failure of jobs with large circuits at $n = 4$, for reasons that we could not identify). Let us consider the 102-qubit experiments with the data shown in Fig. 2a & b. First, the energy for the optimal ansatz state from the naive extrapolation to the zero-noise limit gives $-199.2 \pm 1.8$. The energy for Bell pairs, whose ideal energy is $-153$, is obtained from the the naive extrapolation to be $-169.8 \pm 3.2$. The error bars from the extrapolated values are estimated from the extreme fit parameters values within the 68.27% confidence interval, corresponding to one standard deviation.

We note that the ideal energy for the Bell pairs is $-3N/2$, and for $N = 102$, the energy value is $-153$. Naively we can adjust the extrapolated ansatz energy to be $-199.2/(169.8/153) \approx -179$, with an error bar of 4 approximately. This should be compared to the ansatz value -174.04, obtained numerically by the MPS algorithm with a bond dimension $\chi = 64$. The naive reference-state ZNE (rZNE) thus achieves the expected ansatz value within 2.85% of error and the MPS ground-state energy (with $\chi = 64$) within 0.59% of error.

A slightly more sophisticated approach to do this final extrapolation is to take into account of the residual energy values for large $m$. Assuming that the noise and errors for large circuit depths will make at least the local two-qubit density matrices completely mixed, we should expect that the residual value $c$ in the fitting should be zero for the XXZ model. However, in practice it may not be, as we have seen in our data. So in determining how to rescale and obtain a more sensible value of the energy, we should only rescale the energy drop from $m = 0$ to $m \to \infty$, i.e., the $a$ parameter. To be more precise, the rescale factor $r$ is obtained via $a_B \cdot r + c_B = E_{\text{bell}}$, where $E_{\text{bell}}$ is the exact Bell pairs energy and the subscript $B$ in $a$ and $c$ denotes the parameters obtained from fitting the experiments for Bell pairs. Assume the experiments for the optimal ansatz experience similar noise and errors, as their circuit structure and depth are identical (except the rotation parameters), we obtain the extrapolated experimental ansatz energy to be $E_{\text{exp}} = a_E \cdot r + c_E$, where the subscript $E$ denotes the parameters obtained from fitting the experiments for the ansatz. From our experience, the results obtained this way do not differ much from the naive rescaling in most of our experiments. Using the refined estimation, we obtain almost the same result (up to rounding): $E_{\text{exp}} = -179.1 \pm 3.1$. In another run of the same 102-qubit experiments (with 25 repetitions), the naive extrapolation gives $E_{\text{exp}} = -175 \pm 5$ and the modified extrapolation gives $E_{\text{exp}} = -174 \pm 5$. The average of the two sets of experiments gives $E_{\text{exp}} = -177.5 \pm 2.7$. 

S.8
In Fig. S.4, we show the energy per site for all the experiments we have carried out on the Heisenberg model and compare the results to the energy density in the thermodynamic limit by the Bethe ansatz and they agree within a relative error of 3.38%.

Quantum observable depth

In fitting the energy data, we use an exponential function \( f_E(m) = a \exp(-bm) + c \), where \( m = 2n+1 \) is the total number of \( U \) or \( U^{-1} \) in the circuit to construct the state. We note that each \( U \) contains 7 layers of CNOT gates. When such an exponential-decay fitting works, the quantity \( 7/b \), roughly speaking, represents the decay depth in the quantum circuit for the total energy, which we will refer to as the quantum observable depth (QOD), with the observable being the total energy here. It basically provides a practical way to measure how the experimental observable value degrades with the number of CNOT layers (as CNOT gates have the largest error rates in basis gate set). From the ansatz energy data of the 102-qubit experiment on \texttt{ibm\_washington}, we obtain its \( b \) parameter to be \( b_E = 0.567 \pm 0.03 \) and hence about 12.3 ± 0.7 value of the QOD. For the Bell pairs data, we extract that its \( b \) parameter to be \( b_B = 0.53 \pm 0.05 \) and hence a value of 13.1 ± 1.2 for the QOD. These two values seem to agree and we average them to yield a QOD of 12.7 ± 0.7. (The other set of 102-qubit experiments gives a QOD of 12.59 ± 0.34.) The QOD depends on the qubits used in the experiment and possibly on the number of qubits as well. The 50-qubit experiments on \texttt{ibmq\_brooklyn} give a QOD of 18.7(1.5). Among all the experiments carried out on the backend \texttt{ibm\_washington}, we find the experiment using the 10 qubits \{30,31,32,36,51,50,49,48,47,35\} gives the best QOD value of 44 ± 7. For the QOD from other experiments and other backends, see Tables S.3 and S.2. The QOD serves as a quality measure of those qubits in the quantum processor involved in the benchmark, analogous to but different from the metrics, such as the randomized benchmarking \[60\] and the quantum volume \[61\]. We note the QOD will depend on the choice of the observable and the model used, in particular, its value varies across different values of \( \Delta \) in the XXZ model; see e.g. the decay coefficient \( b \) extracted for the 80-qubit XXZ model in Fig. S.5. Moreover, the form of the fitting function may be different; e.g. for some prior experiments with small number of qubits, both linear and quadratic fits were used in the CNOT-gate mitigation \[10,11,12\]. In these cases, we may need to use other quantities (such as the slope) to define the notion similar to the QOD.

Properties of quantum backends and the choice of qubits

The properties of the nine quantum backends of IBM are listed in Table S.1 and there are three different layouts, as illustrated in Fig. S.6. Seven of the backends have 27 qubits, the backend \texttt{ibmq\_brooklyn}
has 65 qubits, and \texttt{ibm\_washington} has 127 qubits, with the last also shown in Fig. 1b. Before experiments were performed, we examined the detailed error rates reported on the service website and chose a path with a desired total number of sites along those connected qubits so as to avoid CNOT links with high error rates. For large system sizes, it is inevitable that we encounter a few CNOT links that may have somewhat high error rates. We note that the detailed noise and error rates may drift over time as the devices are regularly calibrated and this impact large paths more than small ones. For example, in order to perform the 80-qubit XXZ model experiments in Fig. 3b, we had to use a different path from the one used previously for the Heisenberg model (reported in Table S.2) to avoid certain CNOT links with large error rates.

\section*{Additional Text}

\textbf{Exact 4-qubit ground state for the open chain}

For four-qubit XXZ model with the open-boundary condition, we could assume the ground state to be of the form

$$|\psi_0\rangle = a \frac{1}{\sqrt{2}}(|0101\rangle + |1010\rangle) + b \frac{1}{\sqrt{2}}(|0011\rangle + |1100\rangle) + c \frac{1}{\sqrt{2}}(|1001\rangle + |0110\rangle),$$  \hspace{1cm} (S.19)

and evaluate the energy, giving

$$\langle \psi_0 | H_{XXZ} | \psi_0 \rangle = 4ab + 2c + (b^2 - c^2 - 3a^2)\Delta.$$  \hspace{1cm} (S.20)

In order to minimize the energy with the constraint that \(a^2 + b^2 + c^2 = 1\), we can introduce a Lagrange multiplier to enforce the constraint to the optimization of a quadratic function of \(a, b\) and \(c\). It can be formulated as solving an eigenvalue problem for a \(3 \times 3\) matrix, and we find that the exact ground-state energy \(E_0(\Delta)\) can be obtained from the lowest real root of a third-order polynomial,

\[(12\Delta - 3\Delta^3) - (20 + \Delta^2)x + 3\Delta x^2 + x^3 = 0.\]  \hspace{1cm} (S.21)

The coefficients \((a, b, c)\) in the ground-state wavefunction can be obtained as follow,

\begin{align*}
\tilde{a}(\Delta) &= (4\Delta - 3\Delta^3 - 12E_0(\Delta) - \Delta^2E_0(\Delta) + 3\Delta E_0(\Delta)^2 + E_0(\Delta)^3)/32,  \hspace{1cm} (S.22) \\
\tilde{b}(\Delta) &= 1,  \hspace{1cm} (S.23) \\
\tilde{c}(\Delta) &= (-4 - 3\Delta^3 - 12E_0(\Delta) - \Delta^2E_0(\Delta) + 3\Delta E_0(\Delta)^2 + E_0(\Delta)^3)/7,  \hspace{1cm} (S.24)
\end{align*}

with \((a, b, c) = (\tilde{a}, \tilde{b}, \tilde{c})/\sqrt{\tilde{a}^2 + \tilde{b}^2 + \tilde{c}^2}.

S.10
One can also write down the equation for the parameters in the one-layer ansatz to arrive at the above four-qubit state (S.19), up to a global phase, and in principle solve for the optimal parameters for the ground state. However, we could not find a clean expression for the parameters.

For the periodic boundary condition, one can use the anastz,

$$|\psi_0\rangle = a \frac{1}{\sqrt{2}} (|0101\rangle + |1010\rangle) + b \frac{1}{2} (|0011\rangle + |0110\rangle + |1100\rangle + |1001\rangle),$$  \hspace{1cm} (S.25)

and the problem reduces to solving a $2 \times 2$ matrix, which was previously presented in the supplemental materials of Ref. [62], and the analytic expression for the four-qubit ground state is available there.

**Gap structure of the interpolated Hamiltonian**

We have introduced the interpolated Hamiltonian $\hat{H}(s) = (1-s)\hat{H}_{\text{odd}} + s \hat{H}_{\text{XXZ}}$ in the main text. One can check the spectral properties of this Hamiltonian for small $N$ and finds that it is gapped (for finite $N$) in the range of our interest $s \in [0, 1]$. This is illustrated in Fig. S.1 for two different $\Delta$ values using 8 qubits. This means that the product of singlets is adiabatically connected to the ground state of the Heisenberg model via the above interpolating Hamiltonian.

**Analysis for the ansatz**

For the XXZ model, we find that for $N = 4$, the one-layer ansatz can give the exact ground state for $\Delta \geq 0$ (with both open and periodic boundary conditions); with two layers, we can obtain the exact ground state for all $\Delta > 0$; see Fig. S.7a & c. For $N = 6$ (see Fig. S.7b & d), the 2-layer ansatz with the periodic boundary condition is able to give the exact ground state for $\Delta > -1$ and the 2-layer ansatz with the open boundary condition achieves this with a slightly narrower region. With the periodic boundary condition, the 3-layer ansatz to achieve exactness for all $\Delta > -1$ for $N = 6$. For $N \geq 8$, we could not find an ansatz with a small number of layers that contains the exact ground state, but only approaches the ground state as the number of layers increases; see Fig. S.8. For example, for the periodic Heisenberg chain with $N = 8$, the 5-layer and 6-layer ansätze both achieve the ground-state fidelity of unity within an error less than $10^{-6}$.

**Analysis for the ansatz–Heisneberg chain with the periodic boundary condition**

In Fig. S.9, we present the results our analysis on the periodic Heisenberg chain up to 50 sites using ansätze with one to six layers. Given our ansatz breaks the translation invariance (down to two sites), it will take a few layers to approximately restore the invariance. Thus, we expect the fidelity and the accuracy in the GS energy will be worse than the results for the case of the open-boundary condition.
Nevertheless, we do see general improvement in both quantities as the number of layers in the ansatz increases.

**Additional experimental results**

**Results on Heisenberg chains.** The nine IBM Q backends we use have three different layouts, as illustrated in Fig. S.6. The complete list of the results from the experiments for the Heisenberg model on various backends and with various number of qubits is shown in Tables S.3 and S.2. These were carried out using the Bell-measurement approach. In Fig. S.4 we show the energy per site for all the experiments we have carried out on the Heisenberg model and compare the results to the energy density in the thermodynamic limit by the Bethe ansatz and they agree within a relative error of 3.38%.

**Results on XXZ chains.** We have also performed experiments for 8-qubit XXZ model on ibmq_montreal, with $\Delta$ ranging from -0.8 to 1.4, and use two different measurement methods to calculate the energy, as shown in the Fig. S.2. The two methods of the XYZ measurement and of the Bell measurement agree with each other. In addition, we have also used quantum state tomography to measure the total energy at two different values of $\Delta$ (0 and 1); the energy values obtained from tomography also agree with the other two approaches. In particular, the energy results from the state tomography give $-13.46 \pm 0.31$ at $\Delta = 1$ and $-9.3 \pm 0.8$ at $\Delta = 0$.

**Concurrence results.** With the tomography approach, we have obtained additionally the concurrence for all the bonds, and the experimental results are compared to those of the ansätze and the exact solution in Fig. S.3. Due to the open boundary condition, the concurrence alternates from large to small between odd and even bonds. The entanglement on all even bonds is identically zero for the one-layer ansatz. This is due to the initial state being product of singlet pairs on odd bonds and the one-layer entangling operation on even bonds is not strong enough to make the pairs on even bonds entangled. For odd bonds, the concurrence values inferred from the experiments are \{0.890372, 0.767076, 0.683096, 0.768255\} at $\Delta = 1$ and \{0.850059, 0.663988, 0.648982, 0.812279\} at $\Delta = 0$. As the quantum phase transition at the Heisenberg point $\Delta = 1$ is infinite-order, the concurrence does not exhibit singularity across the transition, so we did not perform experiments for the concurrence over a wide of $\Delta$, but only for $\Delta = 0&1$ as an illustration. These concurrence values were obtained by use our rZNE approach with the naive extrapolation using Bell pairs as the reference. In doing ZNE, we had to repeat $(UU^{-1})$ several times, but the resulting reduced density matrices become unentangled for $n \geq 2$ and this makes a fitting not possible. The error bar is thus not directly accessible, but can be estimated from the energy curves.

**Exponential decay rate in the observable.** In Fig. S.5 we show the dependence of the decay coefficient $b$ extracted for the 80-qubit XXZ model, which is related to the QOD via $7/b$. These were performed on
the same set of physical qubits \{97, 96, 95, 94, 90, 75, 76, 77, 71, 58, 57, 56, 52, 37, 38, 39, 33, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 36, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 53, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 74, 89, 88, 87, 93, 106, 107, 108, 112, 126, 125, 124, 123, 122, 111, 104, 103, 102, 101, 100, 110, 118, 117, 116, 115, 114\] on ibm_washington.

**Beyond one dimension**

We expect that our approach used earlier (including energy measurement and mitigation) can be generalized to the Heisenberg and XXZ models in two-dimensional structures, but numerical tools to calculate the optimal variational parameters need to be developed in order to test and implement the optimal ansatz by simulations and on real devices with large systems. Each layer of the ansatz will consist of $z$ sub-layers, where $z$ is the largest coordination number and we assume bipartite graphs. (For example, how well our approach works on non-bipartite graphs themselves is worth studying as there are more exotic states, such as spin liquids on the kagome lattice.)

As a concrete example, we consider a two-leg ladder of the XXZ model as shown in Fig. S.10a and present the analysis of the ansatz, illustrated in Fig S.10. For this system, we prepare all pairs of top and bottom qubits (i.e. belonging to vertical bonds) first in singlets. Then we apply the XXZ ansatz gates first to the odd bonds (both to top and bottom chains), and then to the even bonds, followed by the XXZ ansatz gates applied to the vertical bonds. The above constitutes a one-layer circuit, which contains 6 variational parameters (3 for the Heisenberg interaction) and a CNOT depth of 10. One can extend this to multiple layers.

Since we consider the two-leg system, we have the total spin $N$ being even and there are $N/2$ spins in each rung. We find that for $N = 6$ (total number of spins in the ladder), the two-layer ansatz can achieve the exact ground state for $\Delta > -1$ up to machine precision. We illustrate in Fig. S.11 the ground-state fidelity for a two-leg ladder with $N = 6$ and $N = 8$ sites using a few layers in the ansatz. With this two-leg system, it opens up possibility to study the coupled XXZ model with different coupling strengths in the horizontal bonds from the vertical ones. There is already an interesting phase diagram from the model \[63\] on the two-leg lattice, including a Haldane phase.

In order to assess how our ansatz performs for this two-leg system, we use the MPS technique to compute the ansatz state and its energy and then minimize the energy with respect to the variational parameters. We also use the MPS algorithm to obtain the ground-state wavefunction and its energy. With the one-layer ansatz, the optimal ansatz gives the energy within about 5% error to the ground-state energy for $N = 50$ spins, as shown in Fig. S.12. Its improves to less than 0.7% and the ground-state fidelity exceeds 0.95 even for $N = 50$. Unfortunately, existing devices on IBM Q do not contain such a two-leg
Figure S.1: The energy gap of 8-qubit XXZ model with open-boundary condition that interpolates from one with interaction on odd bonds only to one with interaction on all bonds for (a) $\Delta = 1$, i.e. the Heisenberg model, and (b) $\Delta = -0.8$.

structure, which, however, did exist before on the retired backend IBM Q 16 Melbourne. Therefore, we could not implement the two-leg model. Nevertheless, the above analysis shows the potential of our approach for multiple legs and possibly two-dimensional systems, whose analysis will be left for future work.

| Backend       | Qubit no. | Processor type | Average CNOT error | Average readout error | Average $T_1$ time | Average $T_2$ time |
|---------------|-----------|----------------|---------------------|-----------------------|--------------------|--------------------|
| ibmq_auckland (64) | 27        | Falcon r5.11   | $1.042 \times 10^{-2}$ | $1.439 \times 10^{-2}$ | 178.38 $\mu$s     | 152.09 $\mu$s     |
| ibmq_brooklyn (32) | 65        | Hummingbird r2 | $2.842 \times 10^{-2}$ | $2.928 \times 10^{-2}$ | 74.35 $\mu$s      | 77.66 $\mu$s      |
| ibm_cairo (64)   | 27        | Falcon r5.11   | $7.909 \times 10^{-2}$ | $1.352 \times 10^{-2}$ | 101.71 $\mu$s     | 132.53 $\mu$s     |
| ibm_hanoi (64)   | 27        | Falcon r5.11   | $4.444 \times 10^{-2}$ | $1.357 \times 10^{-2}$ | 151.26 $\mu$s     | 116.79 $\mu$s     |
| ibmq_kolkata (128) | 27       | Falcon r5.11   | $4.801 \times 10^{-2}$ | $1.556 \times 10^{-2}$ | 118.67 $\mu$s     | 96.82 $\mu$s      |
| ibmq_montreal (128) | 27      | Falcon r4      | $1.943 \times 10^{-2}$ | $3.426 \times 10^{-2}$ | 119.39 $\mu$s     | 102.78 $\mu$s     |
| ibmq_mumbai (128) | 27       | Falcon r5.1    | $7.984 \times 10^{-2}$ | $2.665 \times 10^{-2}$ | 135.78 $\mu$s     | 117.56 $\mu$s     |
| ibmq_toronto (32) | 27        | Falcon r4      | $8.680 \times 10^{-2}$ | $6.050 \times 10^{-2}$ | 115.84 $\mu$s     | 104.92 $\mu$s     |
| ibm_washington (64) | 127     | Eagle r1       | $4.734 \times 10^{-2}$ | $2.789 \times 10^{-2}$ | 94.38 $\mu$s      | 90.82 $\mu$s      |

Table S.1: Properties of various IBM Q backends used in this work. Q. volume is the quantum volume. The basis gate set of these backends include CX, ID, RZ, SX, and X, where CX denotes the CNOT gate, ID is the identity gate, RZ is the z-rotation gate, and SX is the square root of the Pauli X gate.
Figure S.2: The energy results for an 8-qubit XXZ chain on ibmq_montreal. (a) The energy is obtained using the Bell measurement approach (labelled ‘Bell’) on physical qubits [15,12,13,14,16,19,22,25]; (b) The energy is obtained using the XYZ measurement approach (labelled ‘xyz’) on physical qubits [11,14,16,19,22,25,24,23]. We have also performed energy measurement using quantum state tomography for $\Delta = 0$ and 1. The three methods for measuring energy agree very well in experiments.

Figure S.3: Ground-state entanglement property—concurrence—for a chain of 8-spin XXZ model with the open-boundary condition: (a) $\Delta = 1$ and (b) $\Delta = 0$. The concurrence is calculated for two neighboring qubits $(j, j + 1)$ on $j$-th bond ($j \in [1, 7]$) using quantum states obtained from optimizing one-layer, two-layer, and three-layer ansatzes, as well as from exact diagonalization of the XXZ Hamiltonian and from the experiment done on ibmq_montreal with the one-layer ansatz. Note that with just one layer, the concurrence on the even bonds is zero. For a periodic chain, the concurrence for the exact ground state is independent of which bond is chosen. But the ansatz breaks the translation invariance. From these we observe that the entanglement is decreasing from the 1-layer optimal ansatz to 2- and to the 3-layer one, towards the exact solution. The reason is that the initial state of the ansatz is a product of singlet Bell states on odd bonds, which possesses a very high global entanglement. The gates on even bonds act to decrease the entanglement of Bell states (on odd bonds) to increase the entanglement on even bonds.
Figure S.4: Approximated ground-state energy per site vs. the total number $N$ in the spin chain obtained from experiments performed on various IBM Q backends using the one-layer variational ansatz; see Table I for the numerical values of the energy for all data points. The dashed line is a fit from the data: $-1.713 + 0.393/N$ and the approximated ground-state energy at the $N \to \infty$ limit is $-1.713 \pm 0.046$, which is compared to the exact result from the Bethe ansatz solution $4(\ln 2 - 1) \approx -1.773$.

Figure S.5: The decay coefficient $b$ from extracting the ansatz energy ($b_\text{E}$) and from the Bell pairs energy ($b_\text{B}$) vs. the anisotropy parameter $\Delta$ in the 80-qubit XXZ model, performed in ibm_washington. The QOD is related to the decay rate $b$ as $\text{QOD} = 7/b$. 
Figure S.6: Illustration of the layout of some backends used in this work: (a) the layout of 27-qubit machines, such as ibm_auckland, ibm_cairo, ibm_hanoi, ibm_kolkata, ibm_montreal, ibm_mumbai, and ibm_toronto; (b) the 65-qubit layout of ibm_brooklyn; and (c) the 127-qubit layout of ibm_washington. An edge between two qubits indicates that a direct CNOT gate can be executed between them. See Table S.1 for certain properties of these backends.
Figure S.7: The fidelity of the optimal ansatz states with the exact ground state of the XXZ model as a function of $\Delta$ with (a) & (b): the open boundary condition and (c) & (d): the periodic boundary condition. (a) The chain size $N = 4$ with one-layer and two-layer ansatz states. We see that the optimal ansatz states are the exact ground state for $\Delta \geq 0$ for one layer and for all $\Delta > -1$ with two layers. (b) The chain size $N = 6$ with one-layer, two-layer, and three-layer ansatz states. We see that the optimal ansatz states are the exact ground states for all $\Delta > -1$ with three layers. (c) For $N = 4$, we see that the optimal ansatz states are the exact ground state for $\Delta \geq 0$ for one layer and for all $\Delta > -1$ with two layers. (d) For $N = 6$ with two layers, the optimal ansatz states are the exact ground state.
Figure S.8: The open-boundary XXZ chain with $N = 8$ spins, using 1-layer, 2-layer, and 3-layer ansätze. (a) The respective ground-state fidelity for different layers of ansätze. (b) Their relative error in the ground-state energy.

Figure S.9: The fidelity (left panel) and relative energy error (right panel) of the optimal ansatz state and the exact ground state of the (periodic-boundary-condition) Heisenberg model with the total number of qubits $N$ for one to three layers in the ansatz. With these data, we can estimate and extrapolate the GS fidelity and the relative error in GS energy to a larger number $N$ of qubits. The bond dimension used in the MPS calculations is $\chi = 80$. We note that each layer contains two variational parameters and so there are 12 such parameters to optimize in the six layers. Although the results may not necessarily be the global minimum, adding more layers in general improves the ground-state energy and fidelity.
Figure S.10: The anastz for the two-leg ladder. (a) We initialize the state in a product of singlets which are formed between the upper spins and the lower spins. One layer ansatz includes (b), (c) and (d), where the gates are indicated by shaded rectangles.

Figure S.11: The fidelity of the optimal ansatz state (with one to three layers) with the exact ground state for $N = 6$ (a) and $N = 8$ (b) two-leg ladder XXZ model parameterized by $\Delta$. 
| N | $E_{\text{exp}}$ | $\epsilon_{\text{aim}}$ | $\epsilon_{\text{go}}$ | shots | rep. |
|---|---|---|---|---|---|
| 10 | -16.9(3.4) | 1.08% | 0.78% | 44(7) | 40K |
| 10 | -16.5(1.4) | 1.31% | 1.08% | 31(4) | 40K |
| 20 | -33.8(1.4) | 0.055% | 2.67% | 20.3(1.0) | 40K |
| 30 | -51.1(2.3) | 3.16% | 2.57% | 19.6(1.1) | 40K |
| 40 | -69.2(1.6) | 1.74% | 1.31% | 3.12% | 40K |
| 50 | -86.9(1.8) | 2.09% | 1.31% | 3.12% | 40K |
| 60 | -99(4) | 3.16% | 1.31% | 3.12% | 40K |
| 70 | -125(7) | 4.76% | 1.35% | 13.1(0.8) | 40K |
| 80 | -138.5(2.5) | 1.52% | 1.82% | 12.79(0.23) | 40K |
| 90 | -153(5) | 1.34% | 1.82% | 3.64% | 12.5(0.4) | 40K |
| 98 | -168.1(2.6) | 0.54% | 2.81% | 12.32(0.27) | 40K |
| 100 | -173(9) | 1.39% | 1.99% | 11.9(0.14) | 40K |
| 102 | -177.5(2.7) | 1.99% | 1.42% | 12.23(0.17) | 40K |

Table S.2: Various Heisenberg spin-chain experiments performed on the 127-qubit ibm_washington backend/device of IBM Q.
Figure S.12: The fidelity (left panel) and relative energy error (right panel) of the optimal ansatz state and the exact ground state of the (two-leg) Heisenberg model with the total number of qubits $N$ for one to three layers in the ansatz. With these data, we can estimate and extrapolate the GS fidelity and the relative error in GS energy to a larger number $N$ of qubits. The bond dimension we used in the MPS calculation is $\chi = 64$.

Table S.3: Heisenberg spin-chain experiments performed on all available 27-qubit backends/devices and the 65-qubit ibmq_brooklyn of IBM Q.