PNAS

www.pnas.org

Supplementary Information for

"Quantum Approximate Optimization of the Long-Range Ising Model with a Trapped-Ion Quantum Simulator"

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Guido Pagano
E-mail: pagano@umd.edu

This PDF file includes:

Figs. S1 to S6
SI References

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

1 of 11
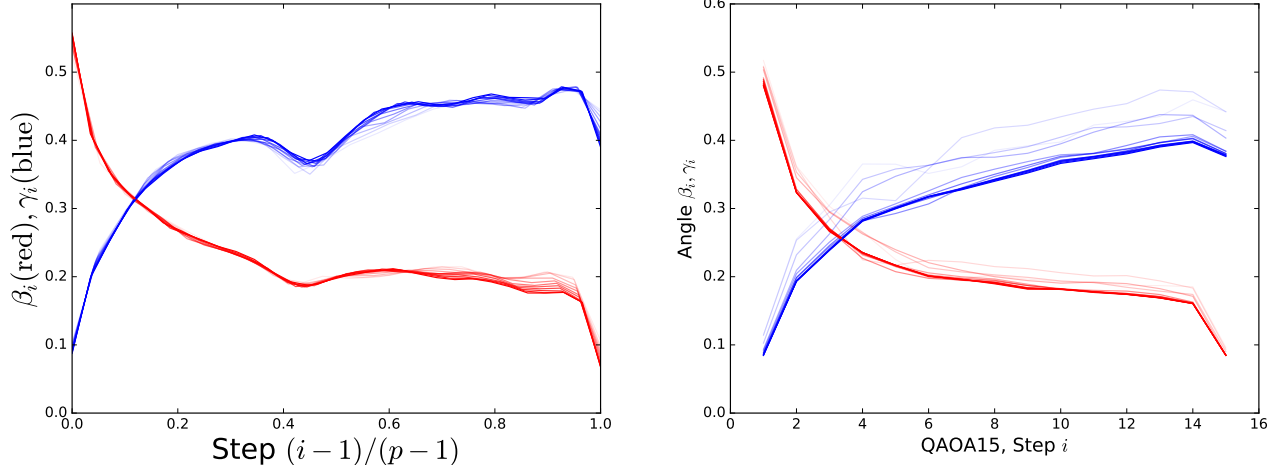
Figure S1. Convergence in p and N. Convergence of optimal angle curves with increasing QAOA layers p (left), and number of spins N (right). The p-convergence plot was generated for an $N = 8$ spin system, for p ranging from 20 up to 30, with higher p shaded darker. The N-convergence figure was generated for a 15 layer QAOA, for N in the range of 4 to 14, with higher N curves shaded darker.

1. Quantum Approximate Optimization Algorithm (QAOA)

The QAOA is an approximate optimization algorithm first introduced in 2014 by Farhi *et al.* (1), and has since enjoyed growing interest. The QAOA uses alternating evolutions under two non-commuting operators, typically a problem (or cost) Hamiltonian H_A that encodes the cost function on the diagonal in (say) the σ^x basis, and a transverse term $H_B = -\sum_{i=0}^{N} \sigma_i^y$ that generates transitions between bit strings, such that the initial state $|+\rangle_y^{\otimes N}$ evolves into an approximate ground state of H_A.

Practically, the most valuable feature of the QAOA seems to be its "learnability" via a classical outer loop optimizer, where the discovery of the evolution angles in the optimal QAOA schedule is achieved via the discovery of structure in the angle sequences (2–4). These patterns are seen quite generally across local Hamiltonian problems, and while steps towards a theory describing optimal QAOA sequences have been taken (3), several questions surrounding it remain open. Regardless, the structure in optimal QAOA schedules may be harnessed to implement approximate state preparation in a scalable manner and with a low overhead on quantum resources. We present a new heuristic method that helps achieves this goal.

First, we discuss how to discover optimal QAOA1 schedules, i.e., QAOA schedules for $p = 1$.

A. QAOA, $p = 1$. Despite its apparent simplicity, the $p = 1$ QAOA (or QAOA1) can be a powerful state preparation ansatz. For example, hardness-of-sampling results are known for QAOA1 circuits (5), closely mirroring the hardness of sampling from instantaneous quantum polynomial (IQP) circuits (see next section for details). Furthermore, it is known that the performance of the QAOA1 for certain combinatorial optimization problems can be competitive with the best classical algorithms for the same problems (6). Another desirable feature of the QAOA1 for local spin Hamiltonians is the tractability of computing energy expectation values, as observed in (1). A very similar result has also been known in the setting of quantum dynamics (7, 8). For a two-local transverse field spin Hamiltonian as in Eq. (1) in the main text, this leads to a formula for the energy expectation under a state produced by the QAOA1, starting from the product state $|+\rangle^{\otimes N}$. These formulas are applicable to many cases of interest in quantum state preparation and optimization. Importantly, the time complexity to compute the formula is $O(N^3)$ in the worst case, making it tractable to optimize the QAOA1 protocols for large spin chains.

B. QAOA, $p > 1$. The general analytical formula for $p = 1$ does not extend to the case where we apply the QAOA for more than one layer. Here, we must turn to classical numerical methods to find the optimal QAOA angles β_i, γ_i for each layer i. For p layers, this is an optimization on a $2p$-dimensional space that grows exponentially with the depth of the circuit. However, numerics done here and in (2, 4) have identified the existence of minima that exhibit patterns in the optimal QAOA angles, namely that the angles, when plotted as a function of their index i, form smooth curves for any p. While this observation points to a deeper theoretical mechanism at play, it does not directly simplify the optimization problem, since we must still search over all approximately smooth sequences of the angles. Zhou *et al.* (2) have exploited the smoothness of the functions by carrying out searches in the Fourier domain. Here, we follow a different route that arises from some novel observations of these family of minima.

For each p, denote the special optimal angles by $\left\{ \left(\boldsymbol{\beta}^{*(p)}, \boldsymbol{\gamma}^{*(p)} \right) \right\}_p$, which we can also think of as a pair of angle curves (as a function of step index i). As p is varied, we may think of these minima as a family. We numerically find that this family exhibits the following desirable features (for p sufficiently large):

1. The angles are non-negative, small and bounded.

 G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

2. For p sufficiently large, the two angle sequences $\beta^{*(p)}$ and $\gamma^{*(p)}$ are approximately smooth.

3. The angle sequence $\beta^{*(p)}$ (and correspondingly, $\gamma^{*(p)}$) when viewed as a function on the normalized time parameter $s_i = \frac{i-1}{p-1}$, is convergent in the parameter p. In other words, as p is increased, the angle sequences $\beta^{*(p)}$ and $\gamma^{*(p)}$ approach a smooth, asymptotic curve (See Fig. S1.)

4. The energy expectation $E(\beta^{*(p)}, \gamma^{*(p)})$ approaches the global minimum as $p \to \infty$, and hence this family is asymptotically optimal.

The significance of the first point is that in experimental settings, large evolution times are infeasible to implement due to decoherence, so these minima correspond to practicable QAOA protocols. The third and fourth points suggest an inductive algorithm where a locally optimal schedule for a given p may be discovered using the optimal schedule for $p-1$ as a prior.

Point 3 in the above list is a novel observation that allows us to construct a heuristic that is efficiently scalable for large p. The main idea behind this construction is that the minimal angle curves for a larger p may be guessed from the optimal curve of a smaller $p' < p$ by interpolation. It should be noted that the angles in these curves remain roughly the same size as p increases. Furthermore this size is large enough that interpretation of QAOA as a Trotterized product series is not feasible with the corresponding error terms being non-vanishing. Therefore, while tempting, it is not theoretically accurate to interpret these curves as a Trotterized annealing path. The theoretical underpinnings of these curves are still under investigation.

Using the above points, we use a bootstrapping algorithm to find the optimal angle sequences, $\beta^{*(p)}$ and $\gamma^{*(p)}$, for a given p, as described below. Let $q = 1, \dots, p$ denote an intermediate angle index. Then:

1. For $q = 1$, use an analytic formula to find $\beta^{*(1)}$ and $\gamma^{*(1)}$.

2. For $q = 2$, choose an initial guess of $\beta^{(2)} = \left(\beta^{*(1)}, \beta^{*(1)} - 0.2\right)$ and $\gamma^{(2)} = \left(\gamma^{*(1)}, \gamma^{*(1)} + 0.2\right)$.

3. Perform a local optimization of $\beta^{(2)}$ and $\gamma^{(2)}$ in order to find $\beta^{*(2)}$ and $\gamma^{*(2)}$.

4. Repeat the next steps (5-7) for $q = 3, \dots, p$.

5. Create interpolating functions through the angle sequences, $\beta^{*(q-1)}$ and $\gamma^{*(q-1)}$, using the normalized time $s_i = \frac{i-1}{q-2}$ as the independent parameter (we use a linear interpolation for $q = 3$ and cubic for $q > 3$).

6. Choose the initial guesses for $\beta^{(q)}$ and $\gamma^{(q)}$ by sampling the interpolating function from (5) at evenly spaced points separated by a normalized time distance of $\Delta s = 1/(q-1)$.

7. Perform a local optimization of $\beta^{(q)}$ and $\gamma^{(q)}$ in order to find $\beta^{*(q)}$ and $\gamma^{*(q)}$.
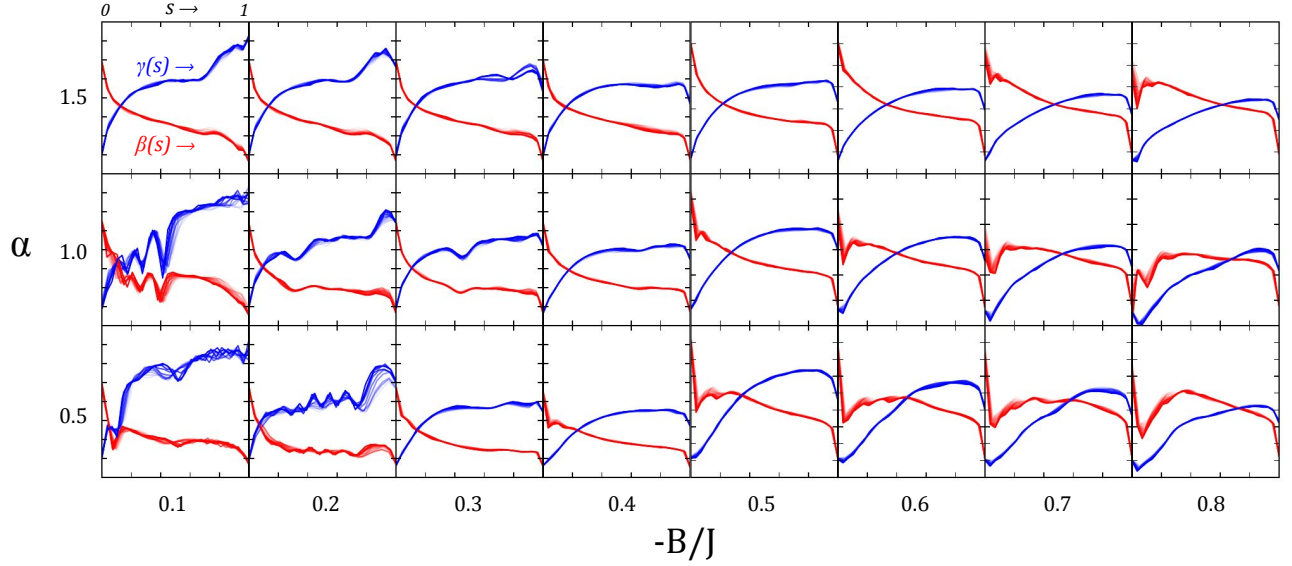
The resulting angles $\beta^{*(p)}$ and $\gamma^{*(p)}$ should be at least a good local minimum of the energy expectation value and approaches the global minimum as $p \to \infty$.

The $q = 2$ interpolation in step 2 is based on our observation that the β angles tend to curve down at the end and the γ angles tend to curve up.

An important feature of our algorithm is that its asymptotic runtime is expected to be efficient in p. This feature is predicated on the previous result that the angle curves are generally convergent as p tends to infinity. The argument proceeds as follows: if we assume a maximal deviation of the initial guess for layer q to be $\epsilon_q \geq 0$, then the total l_2-norm distance between the initial guess and the optimized curve is no greater than $\epsilon_q \sqrt{q}$, by the Cauchy-Schwarz inequality. Therefore, the local search algorithm is confined to a ball of radius at most $\epsilon_q \sqrt{q}$, and for a fixed error tolerance, the convergence time for a standard local optimizer is $O(\epsilon_q^2 q)$. Summing over convergence times for all from $q = 1, \dots, p$, we have

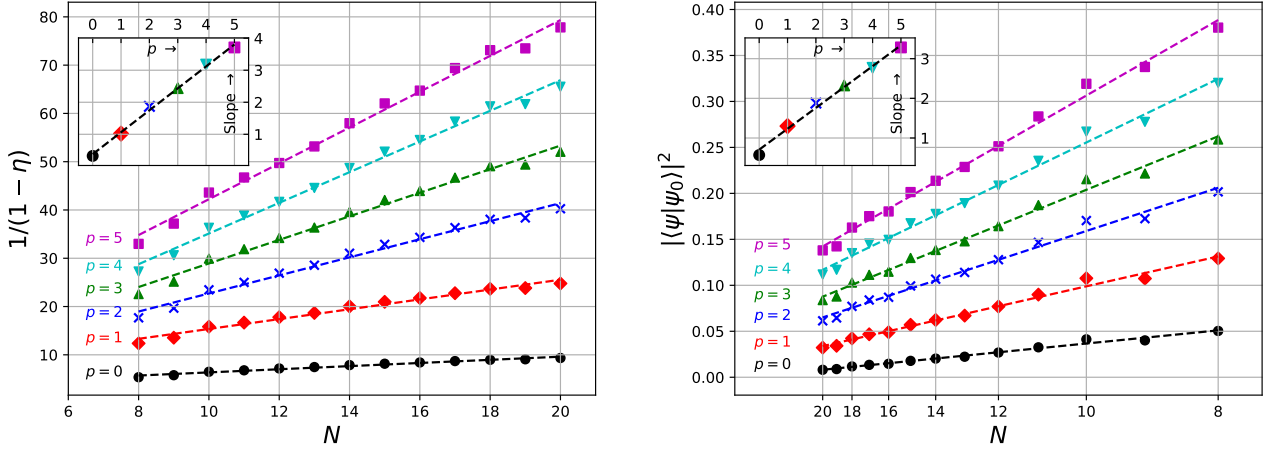$$T = O\left(\sum_{q=1}^{p} q\epsilon_q^2\right) \leq O(p^2) \qquad [1]$$

The last inequality above comes about as follows: while the summand depends on the convergence rate of the sequence $\{\epsilon_q\}_{q=1}^{p}$, it is upper bounded by $O(q)$ for a converging set of paths and an initial error ϵ_1 of order 1. The latter is true since our angle search domain is bounded and independent of N. Therefore, the sum is no greater than $O(p^2)$. In practice, even faster runtimes are possible. Therefore, the bootstrap algorithm exploits the structure of the special minima and provides a scalable route to multi-step QAOA for the long-range TFIM. In fact, as discussed in the supplement and in (3), there is mounting numerical evidence that the path approach applies across a very general variety of models on discrete as well as continuous systems.

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Figure S2. Angle sequence curves. A collage of angle sequence curves, arranged by the Hamiltonian parameters for which they were computed. In each subplot, curves for different p ranging from 20 to 30 are overlaid, with higher p curves shaded darker. The horizontal axis represents fractional step $s = (i-1)/(p-1)$ ranging from 0 to 1, while the vertical axis gives the value of the angles β (red), and γ (blue) in the range $[0, 0.6]$. The subplots are arranged horizontally by $-B/J_0$, increasing from 0.1 to 0.8 in steps of 0.1 (from left to right), and vertically by the long-range power $\alpha = 0.5, 1.0, 1.5$ (bottom to top). This collage shows the persistence of structure in the optimal angle sequences for a range of Hamiltonians within the same family.

C. Convergence in N. In the previous sections, we introduced a bootstrap algorithm that is asymptotically efficient in the number of layers p. However, in order to be fully scalable the algorithm must also be scalable in the system size N. This may not be possible in general (say for random spin models), as the optimized angles for a particular small system may have no bearing on the angles for a larger system. However, for the long-range TFIM, and indeed any translationally-invariant model with a well-defined notion of metric and dimension arising from the functional form of the coupling coefficients J_{ij}, it is reasonable to expect that the optimized angles depend on system size in a predictable way. This is indeed the case for the long-range TFIM. There, it can be seen that the angle curves for varying N appear similar in shape. Usefully, the curves also appear to be *convergent* to an idealized curve for a hypothetical continuous, long-range spin chain. Once again, this feature suggests that the optimized QAOA angle curves for small systems may be used as initial guesses for larger systems *within the same Hamiltonian family.*

While it is not clear (due to numerical limitations) how fast the curves converge, we argue that the rate should be weakly dependent (or independent) of the system size N. For a given coupling function (such as inverse power-law) that decays as a function of distance, we define a characteristic length scale, which may be called the *skin depth δ*, that is the number of sites from the boundary that the coupling is a factor of e smaller than the nearest-neighbour value. In other words, we define δ such that $J_{i,i+\delta} \sim J_{i,i+1}/e$. Clearly, δ is independent of the system size N and depends only on the parameters of the coupling function. For instance, for the long-range TFIM, $\delta \sim e^{1/\alpha}$. As N tends to infinity, the *fractional* skin depth δ/N then "falls away" and becomes vanishing with respect to the bulk region of the chain. Now, we make the assumption that any deviations in the optimal QAOA schedules from N to $N+1$ arise from change in the fractional skin depth, which is reasonable for a translationally invariant model. The incremental change in the fractional skin depth from N to $N+1$ is $\delta/N - \delta/(N+1) \sim O(1/N^2)$. Therefore, if the change in the optimal QAOA curves ϵ_N (in, say, l_1-norm distance) is a smooth function of the the fractional skin depth, then we expect it to vary as $\epsilon_N \sim 1/\mathrm{poly}(N)$. Therefore, the total running time of a bootstrap from small system sizes to a given size N should be $O\left(\sum_{k=1}^{N} 1/\mathrm{poly}(N)\right)$ which is sub-linear in N. Combining this observation with the convergence in p, we see that for a given Hamiltonian family, optimized QAOA angle curves for small p may be used as a rubric for the optimization for longer circuit depths. Furthermore, if the Hamiltonian is translationally-invariant with decaying interactions, the optimized QAOA schedules are expected to scale with N as well. Therefore, the state preparation procedure under the QAOA for such a Hamiltonian family is scalable in circuit "volume", for a wide range of Hamiltonian parameters (Fig. S2). This is our main theoretical contribution in this work.

     G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Figure S3. Performance scaling in p, N. Behaviour of performance parameters η (left) and squared ground state overlap (right) with increasing number of spins N (x axis) and p (colors), for ideal power-law coupling with $\alpha = 1.1$. We find that for each p, $1/(1 - \eta)$ grows linearly in N with a slope that depends on p. (Inset) The slope is linear in p, suggesting that the performance converges to 1 as $\eta \sim 1 - 1/(pN)$. On the right, we empirically observe that $|\langle \psi | \psi_0 \rangle|^2 \sim p/N$, indicating that constant overlap with the ground state can be achieved with linear depth QAOA. The x axis has been scaled as $1/N$ so that the linear relationship with the squared overlap is apparent. The inset shows the linear trend with p.

D. Scaling of η in p, N. Our performance parameter η, defined as

$$\eta \equiv \frac{E(\vec{\beta}, \vec{\gamma}) - E_{max}}{E_{gs} - E_{max}}, \tag{2}$$

measures how close (in energy) the prepared state is to the ground state of the system. As described in previous sections, the optimal angle curves for QAOA appear to converge to a smooth, hypothetical curve, as a function of p as well as N. We show that under the assumption that such a curve exists, there is a fast heuristic for finding optimal angles for any finite p that is time-efficient in p and the number of spins N (when used in conjunction with the quantum device). In this section, we show that not only is the search efficient, but the quality of the optimum is numerically seen to improve with p, N as well.

In Fig. S3, we show the result of the numerical study. We chose as the target Hamiltonian an idealized transverse field Ising model with inverse power-law couplings, with the power $\alpha = 1.1$ chosen to closely mimic the experimental Hamiltonian. The number of spins was varied from $N = 8$ to 20. Via DMRG, the critical value of the transverse field for a finite chain can be located by maximizing the von Neumann entropy at half-cut. This was done independently for each value of N. Then, using our heuristic, we located the optimal angle curve, and computed η for the final state prepared using this angle sequence, for each N. The plot shows the trend of $1/(1 - \eta)$ with N, for a range of $p = 0, 1, 2, 3, 4, 5$, with 0 corresponding to a trivial protocol where the initial state is returned. While the number of spins could not be extended beyond 20 due to computational limitations, the trend is clear. We see that $1/(1 - \eta)$ grows linearly with N and p (inset). While the linear trend in N is encouraging, we similarly expect the inverse spectral gap (and indeed, the density of low-lying states) to increase with N. Empirically for the target Hamiltonian, we observe a gap scaling of $\sim 1/N^2$. Assuming the density of low-lying states scales similarly, this suggests that the squared overlap with the ground state should fall off with N. Numerics confirm this expectation and indicate a scaling of the squared overlap of $|\langle \psi | \psi_0 \rangle|^2 \sim p/N$.

The linear scaling with p for both the energy and fidelity metric, combined with a polynomial-time search heuristic, suggests that for any desired energy (or probability) threshold ϵ, our approach allows us to approximate the state to within $1 - \epsilon$ (in energy or fidelity) in time and number of layers that scale as $\text{poly}(N, 1/\epsilon)$.

E. Characteristic scale for η. The figure of merit η characterizes how close the final state is to the ground state of the system. At $\eta = 0$, the system is in the highest excited configuration, while $\eta = 1$ corresponds to a perfectly prepared ground state. QAOA, starting from the initial state $|+\rangle^{\otimes n}$, gives a state with figure of merit $\eta \in [0, 1]$, from the initial value of η_0. The difference between the final η and η_0 indicate the success of our QAOA protocol.

While η is normalized to the range $[0, 1]$, differences in η are still somewhat arbitrary. In long-range Ising models with a transverse field η_0 is not 0 but typically greater than 0.5, making the difference in η an unsatisfactory metric of success. Therefore, in addition to the initial and final η, we must provide a characteristic scale for η that quantifies the typical deviation from η_0. A natural choice is the standard deviation of η for QAOA with random angles.

For QAOA1 with evolution angles β, γ, it is possible to estimate the standard deviation analytically as a function of the underlying model parameters B and J_0 and on the number of qubits N. This derives from the analytical formula for the energy expectation $E(\beta, \gamma)$ which can be stated as follows:

$$E(\beta, \gamma) = E_I + E_{II} + E_{III} \tag{3}$$

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

where

$$E_I = B \sum_{i=1}^{N} \prod_{k \neq i} \cos\left(2\gamma J_{ik}\right) \tag{4}$$

$$E_{II} = -\frac{\sin\left(4\beta\right)}{2} \sum_{i,j} J_{ij} \sin\left(2\gamma J_{ij}\right) \prod_{k \neq i,j} \cos\left(2\gamma J_{ik}\right) \tag{5}$$

$$E_{III} = -\frac{\sin^2\left(2\beta\right)}{4} \sum_{s=\pm 1, i, j} J_{ij} \prod_{k \neq i,j} \cos\left(2\gamma \left(J_{ik} + (-1)^s J_{jk}\right)\right) \tag{6}$$

where the Hamiltonian has long-range power law couplings $J_{ij} \sim \frac{1}{|i-j|^\alpha}$ (with $J_{ii} = 0$), and a transverse field of strength B. Then, our goal is to compute the standard deviation (normalized by the spectral bandwidth $\Delta := E_{max} - E_{gs}$),

$$\frac{\sigma_E}{\Delta} = \frac{\sqrt{\langle E^2 \rangle_{\beta,\gamma} - \langle E \rangle_{\beta,\gamma}^2}}{\Delta} \tag{7}$$

which gives us the characteristic scale for η. We define the average $\langle \cdot \rangle_{\beta,\gamma}$ as

$$\langle f \rangle_{\beta,\gamma} := \lim_{T_\beta, T_\gamma \to \infty} \frac{1}{4T_\beta T_\gamma} \int_{-T_\beta}^{T_\beta} \int_{-T_\gamma}^{T_\gamma} f(\beta, \gamma) d\beta d\gamma \tag{8}$$

In the limit, the average is precisely the constant term of the Fourier transform of f. Since the function is a sum of trigonometric monomials, its moments over the angle variables β, γ can be computed analytically term by term. We will need the following properties of the coupling function:

1. (Symmetry) Since the inverse power law only depends on distance between nodes, we have $J_{ij} = J_{(2j-i)j}$ In other words, the inverse power-law is symmetric under a lateral flip (or "mirroring"). We assume a finite, open chain, and therefore couplings J_{ij} with $|j - i| > N - j$ do not have an image under mirroring.

2. (Incommensurateness) The coupling strengths J_{ij} are, in general, mutually indivisible irrational numbers whose sums and differences are also irrational and mutually distinct, e.g. for $i \neq j, k \neq l$, $J_{ik} \pm J_{jk} \neq J_{il} \pm J_{jl}$ (with a very small set of exceptions due to, say, symmetry).

The mean $\langle E \rangle_{\beta,\gamma}$ consists of three parts corresponding to the terms E_I, E_{II}, E_{III}. Performing the β integral first, we see that $\langle E_{II} \rangle_{\beta,\gamma} = 0$. Next, we may argue that in products of the form $\prod_k \cos(2\gamma J_{ik})$, the cosine factors are of degree one if they have no mirror images, and degree two otherwise. The only way to have a non-zero expectation is if all terms are systematically paired up by mirroring, so that the overall product is quadratic in a product of cosines. For the summand in E_I, this can only happen if N is odd and i is exactly at the center of the chain, in which case the average is $B/2^{(N-1)/2}$. When N is even, the mean is 0. Finally, for general i, j the last term is zero by property 2, since the cosines are generically incommensurate and therefore barring very few exceptions, most phases do not cancel out. However, in the special case that i, j are mirror images, i.e. $i = N - j$, we have perfectly paired terms when N is even (and one unpaired term at $k = \lfloor N/2 \rfloor$ when N is odd). Counting all occurrences of this case, the mean is approximately $\frac{1}{2^{N/2+1}} \sum_{i=1}^{N} J_{i(N-i)} \lesssim NJ_0/2^{N/2}$ where J_0 is the nearest-neighbor coupling in the chain. Note that asymptotically in N, $\langle E \rangle_{\beta,\gamma} \sim O(N/2^{N/2})$ which approaches 0 in the infinite N limit.

Next, we estimate the term $\langle E \rangle_{\beta,\gamma}^2$. By the orthogonality of trigonometric polynomials in β, we first have that $\langle E \rangle_{\beta,\gamma}^2 = \langle E_I \rangle_{\beta,\gamma}^2 + \langle E_{II} \rangle_{\beta,\gamma}^2 + \langle E_{III} \rangle_{\beta,\gamma}^2$. Therefore, we estimate each term separately. As before, we require that the cosines pair up so that their phases can cancel. First, we have

$$\langle E_I \rangle_{\beta,\gamma}^2 = B^2 \sum_{i,j} \prod_{k=1}^{N} \cos(2\gamma J_{ik}) \cos(2\gamma J_{jk}) \tag{9}$$

Each summand is a product of $2N$ cosines, and only survives averaging if every cosine is paired. This happens exactly when either $i = j$ or $i = N - j$ (There is also the "disconnected" contribution that cancels with the mean). In each case, the squared cosines give a factor of $1/2$ from averaging. Moreover, using mirror symmetry we can have fourth powers of some of the cosines, which give a factor $3/8$ from averaging. In all, the mean (minus the disconnected part) is no greater than

$$\langle E_I \rangle_{\beta,\gamma}^2 \lesssim 4NB^2 \left(\frac{3}{8}\right)^{(N-1)/2} \tag{10}$$

6 of 11   **G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe**

A similar reasoning for E_{II}, E_{III} give us the following estimates:

$$\langle E_{II}\rangle^2_{\beta,\gamma} \lesssim \frac{1}{4} N J_0^2 \left(\frac{3}{8}\right)^{(N-1)/2} \qquad [11]$$

$$\langle E_{III}\rangle^2_{\beta,\gamma} \lesssim \frac{3}{16} N J_0^2 \left(\frac{3}{8}\right)^{(N-1)/2} \qquad [12]$$

Finally, this gives

$$\langle E\rangle^2_{\beta,\gamma} \lesssim N \left(\frac{3}{8}\right)^{N/2} \left[8B^2 + J_0^2\right] \sim O(N \cdot (3/8)^{N/2}) \qquad [13]$$

Therefore, we see that the standard deviation $\sigma_\eta = \sigma_E/\Delta \sim \frac{\sqrt{8B^2 + J_0^2}}{\Delta} \cdot N^{1/4} (3/8)^{N/4}$, which is exponentially suppressed for large N. For $N = 20$ ions, we have $N^{1/4} \cdot (3/8)^{N/4} \sim 0.02$. While this is already small, the normalization $\frac{\sqrt{8B^2 + J_0^2}}{\Delta}$ will have an additional linear N factor in the denominator, making the scale for η about 0.002. Therefore, a typical final QAOA performance of $\eta \gtrsim 0.95$ is several standard deviations above a typical $\eta_0 \sim 0.85$.

2. Evidence for hardness of sampling from general QAOA circuits

In this section we expand upon previous work (5) that gives evidence for exact sampling hardness of QAOA circuits, using the techniques of Refs. (10, 11) to give evidence for hardness of approximate sampling. First we relabel the bases $Y \to X \to Z$ so that the $p = 1$ experiment is equivalent to preparing a state $|\psi_0\rangle = |\uparrow\rangle^{\otimes N}$, evolving under a Hamiltonian H_z diagonal in the computational basis, followed by a uniform rotation $\tilde{H} = e^{-i\beta \sum_i \sigma_i^x}$ and measurement in the computational basis. Following Ref. (1), it suffices to consider QAOA circuits with $\beta = \pi/4$. The output state is $\tilde{H}^{\otimes N} e^{-i\gamma H_z} H^{\otimes N} |0^N\rangle$ for some cost function C diagonal in the computational basis.
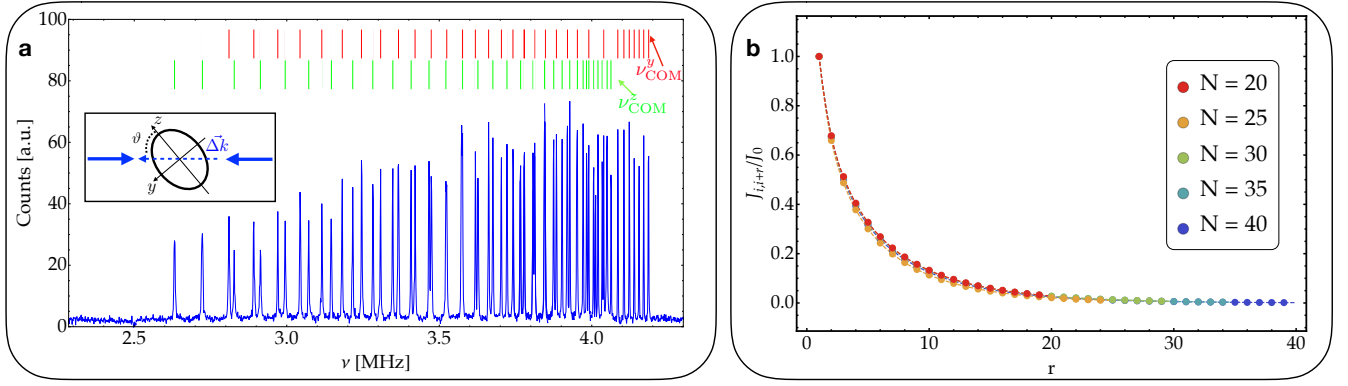
A. Generalized gap of a function. The main idea behind proving exact sampling hardness is to examine a particular output amplitude, say the amplitude of the $|0^N\rangle$ basis state. In Ref. (10), the output state after a so-called IQP circuit (which only differs from the one here in that the final rotation is a global Hadamard $H^{\otimes N}$ instead of $\tilde{H}^{\otimes N}$) has an amplitude proportional to a quantity known as the *gap* of a Boolean function, $\text{gap}(f) = \sum_{x:f(x)=0} 1 - \sum_{x:f(x)=1} 1$, the difference in the number of inputs that map to 1 and the number of inputs that map to 0 under f. Finding the gap of a general function is a GapP-complete problem. This is a very hard problem since the class GapP includes #P, which in turn includes the whole of NP. The authors of Ref. (10) prove that the gap of a degree-3 polynomial over $\mathbb{Z}_2$, f, may be expressed as an output amplitude of an IQP circuit. They also show that the finding the gap of such functions f is still GapP-complete. Following Ref. (10), we examine the $|0^N\rangle$ output amplitude of a QAOA state:

$$\langle 0^N | \tilde{H}^{\otimes N} e^{-i\gamma H_z} H^{\otimes N} |0^N\rangle = \frac{1}{2^N} \sum_{x,y} \langle y | i^{\sum_i y_i + \bar{f}(x)} |x\rangle, \qquad [14]$$

where now we define the function $\tilde{f}$ to have the range $\mathbb{Z}_4$ and the Hamiltonian H_z satisfies $e^{-i\gamma H_z}|x\rangle = i^{\tilde{f}(x)}|x\rangle$ for a computational basis state $|x\rangle$. The output amplitude is thus proportional to a 'generalized gap' $\text{ggap}(f) := \sum_{x:f(x)=0} 1 + i \sum_{x:f(x)=1} 1 + i^2 \sum_{x:f(x)=2} 1 + i^3 \sum_{x:f(x)=3} 1$ of a function $f(x) = \tilde{f}(x) + \text{wt}(x)$, where $\text{wt}(x)$ is the Hamming weight of x. This modified function $f(x)$ is also a degree-3 polynomial over $\mathbb{Z}_4$. Note that this restriction to degree-3 comes from the fact that the gates Z, CZ and CCZ are universal for classical computation (indeed, the Toffoli alone is universal for classical computation) and there is a natural degree-3 polynomial coming from this construction. The quantity we have defined, $\text{ggap}(f)$, can be easily shown to be GapP-hard to compute, by reducing gap to ggap. This suffices for exact sampling hardness assuming the polynomial hierarchy (PH) does not collapse.

B. Approximate sampling hardness. For approximate sampling hardness, we need two other properties, namely anti-concentration and a worst-to-average case reduction. Anti-concentration of a circuit roughly says that the output probability is sufficiently spread out among all possible outcomes so that not many output probabilities are too small. We choose a random family of QAOA circuits by choosing H_z such that the function $f(x)$ is a degree-3 polynomial $\sum_{i,j,k} a_{i,j,k} x_i x_j x_k + \sum_{i,j} b_{i,j} x_i x_j + \sum_i c_i x_i$ with uniformly random weights $b_{i,j}$ and $c_i \in \mathbb{Z}_4$. Anti-concentration then follows from the Paley-Zygmund inequality and Lemma 4 of the Supplemental Material of Ref. (10) (with $r = s = 4$).

Finally, we need to show that the problem of approximating the generalized gap is average-case hard. Currently, no scheme for quantum computational supremacy has achieved this, and the best known result in this direction is in Ref. (11), where the authors show a worst-to-average case reduction for the problem of *exactly* computing an output probability of a random quantum circuit. The authors remark that their techniques may be extended to any distribution parametrized by a continuous variable. In principle, we have such a parameter γ available here, which continuously changes the parameters $b_{i,j}$ and c_i. However, we have only shown anti-concentration when the weights $b_{i,j}$ and c_i are chosen from a finite set. It remains to be seen whether one can have the property of anti-concentration and average-case hardness holding at the same time for some specific QAOA output distribution.

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Figure S4. System 2 characterization. (a) Sideband resolved spectroscopy of a 32 ion chain with frequencies $\nu_{\mathrm{COM}}^{y} = 4.18$ MHz and $\nu_{\mathrm{COM}}^{z} = 4.06$ MHz, with both transverse families identified. Inset: geometrical configuration of the global Raman beams (blue arrows) with respect to the transverse principal axes of the trap (black arrows). The ellipsoid shows qualitatively an equipotential surface of the trap. (b) Average spin-spin interaction matrix element $J_{i,i+r}$ as a function of ion separation $r = |i - j|$ for the data taken in Fig. 2c in the main text, calculated with the system parameters directly measured with sideband spectroscopy, using Eq. (16). The results are normalized to the average nearest-neighbour coupling J_0 for each system size.

3. Trapped-ion experimental systems

In this work two quantum simulators have been used, referred to as system 1 and 2. System 1 (12) is a room-temperature ion-trap apparatus, consisting of a 3-layer linear Paul trap with transverse center-of-mass (COM) motional frequency $\nu_{\mathrm{COM}} = 4.7$ MHz and axial center-of-mass frequencies ν_x ranging from 0.39 to 0.6 MHz depending on the number of trapped ions. In this system Langevin collisions with the residual background gas in the ultra high vacuum (UHV) apparatus are the main limitation to ion chain lifetime (13). These events can melt the crystal and eject the ions from the trap because of rf-heating or other mechanisms.
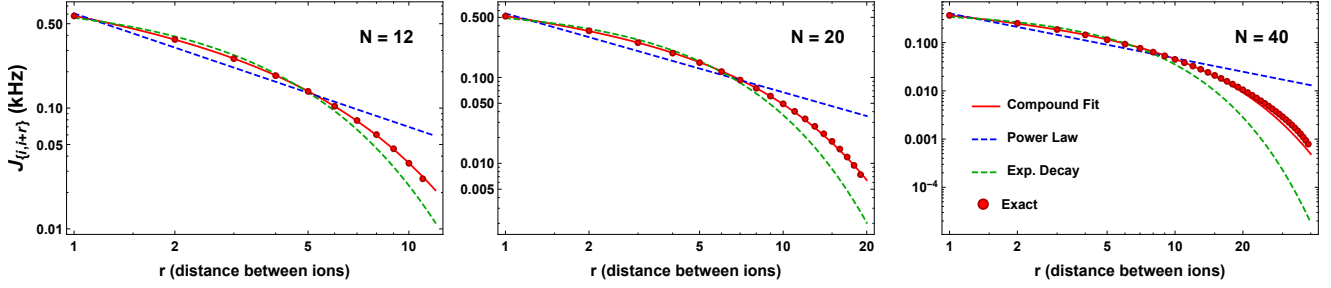
System 2 (14) is a cryogenic ion-trap apparatus based on a linear blade trap with four segmented gold coated electrodes. The trap is held at 6.5 K in a closed cycle cryostat, where differential cryo-pumping reduces the background pressure at low 10^{-12} Torr level, which allows for long storage times of large ion chains. For this reason system 2 has been used to perform the QAOA with a large number of qubits (Fig. 2b) or when a large number of measurements was required (Fig. 4). The two transverse trap frequencies are $\nu_{\mathrm{COM}}^{y} = 4.4$ MHz and $\nu_{\mathrm{COM}}^{z} = 4.26$ MHz, and the axial frequency ranges from 0.27 to 0.46 MHz.

A. State preparation. The qubit is initialized by applying resonant 369.5 nm light for about 20 μs to optically pump into the $|\downarrow\rangle_z$ state. To perform global rotations in the Bloch sphere, we apply two far-detuned, non-copropagating Raman beams whose beatnote is tuned to the hyperfine splitting $\nu_0 = 12.642821$ GHz of the clock states $^2S_{1/2}|F = 0, m_F = 0\rangle$ and $^2S_{1/2}|F = 1, m_F = 0\rangle$ encoding the qubit (15). State preparation in our implementation of the QAOA requires qubit initialization in the $|\downarrow\rangle_z$ state by optically pumping the ions and then a global rotation into the $|\uparrow\rangle_y$ state using stimulated Raman transitions. We detect the state of each ion at the end of each experimental sequence using state-dependent fluorescence, with single site resolution. In order to improve the accuracy of global qubit rotations, we employ a composite pulse sequence based on the dynamical decoupling BB1 scheme (16). This allows us to compensate for inhomogeneity due to the Raman beam's Gaussian profile and achieve nearly 99% state preparation fidelity. The BB1 four pulse sequence is:

$$U_1(\pi/2) = e^{-i\frac{\pi}{2}\sigma_i^{\theta}} e^{-i\pi\sigma_i^{3\theta}} e^{-i\frac{\pi}{2}\sigma_i^{\theta}} e^{-i\frac{\pi}{4}\sigma_i^{x}},$$

where after the first $\pi/2$ rotation $e^{-i\frac{\pi}{4}\sigma_i^{x}}$, three additional rotations are applied: a π-pulse along an angle $\theta = \cos^{-1}(-1/16) = 93.6°$, a 2π-pulse along 3θ, and another π-pulse along θ. The axes of these additional rotations are in the x-y plane of the Bloch sphere with the specified angle referenced to the x-axis.

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Figure S5. Log-log plot of spin-spin interactions: red points represent the average Ising couplings between spins separated by distance $r = |i - j|$, calculated from experimental parameters using Eq. 15. These plots show the exact average couplings and fits corresponding to the $N = 12$ and $N = 20$ gradient descent experiments (Fig. 3 in the main text) and the $N = 40$ exhaustive search experiment (Fig. 2c in the main text). The power law fit (blue dashed curve) fails to match the couplings for larger spin separations, as does an exponential fit (green dashed curve). The compound formula (Eq. 18) fits well the actual couplings for all spin separations, even for a chain of 40 ions. The fitted parameters $\{J_0, \alpha', \beta'\}$ for N = 12, 20, and 40 are $\{0.580, 0.322, 0.229\}$, $\{0.517, 0.318, 0.181\}$, and $\{0.369, 0.383, 0.134\}$ respectively.

B. Generating the Ising Hamiltonian. We generate spin-spin interactions by employing a spin dependent force with a pair of non-copropagating 355 nm Raman beams, with a wavevector difference Δk aligned along the transverse motional modes of the ion chain. The two off-resonant Raman beams are controlled using acousto-optic modulators which generate two interference beatnotes at frequencies $\nu_0 \pm \mu$ in the Mølmer-Sørensen configuration (17). In the Lamb-Dicke regime, the laser-ion interaction gives rise to the effective spin-spin Hamiltonian in Eq. (1) in the main text, where the coupling between the i-th and j-th ion is:

$$J_{ij} = \Omega^2 \nu_R \sum_m \frac{b_{im} b_{jm}}{\mu^2 - \nu_m^2}. \qquad [15]$$

Here Ω is the Rabi frequency, $\nu_R = h \Delta k^2 / (8\pi^2 M)$ is the recoil frequency, ν_m is the frequency of the m-th normal mode, b_{im} is the eigenvector matrix element for the i-th ion's participation to the m-th normal mode ($\sum_i |b_{im}|^2 = \sum_m |b_{im}|^2 = 1$) (18), and M is the mass of a single ion.

Differently from system 1, where the wavevector difference Δk of the Raman beams is aligned along one of the principal axes of the trap, in system 2 the spin-spin interaction stems from the off-resonant coupling to both families of transverse normal modes. Eq. (15) is then generalized to:

$$
\begin{aligned}
J_{ij} &= J_{ij}^y + J_{ij}^z, \\
J_{ij}^\ell &= \Omega_\ell^2 \nu_R^\ell \sum_m \frac{b_{im} b_{jm}}{\mu^2 - \left(\nu_m^\ell\right)^2}, \ \ell = y, z,
\end{aligned}
\qquad [16]
$$

where ν_R^ℓ is the recoil frequency given by the projection of the Raman wavevector Δk along the two transverse principal axes of the trap $\ell = y, z$. We infer an angle $\vartheta \sim 40^o$ between Δk and the z principal axis (see inset in Fig. S4a) from the ratio between the resonant spin-phonon couplings to the two transverse COM modes. Before every experiment, we perform Raman sideband cooling on both the COM and the two nearby tilt modes for both transverse mode families.

As we scale up the number of qubits (see Fig. 2c in the main text), we vary the axial confinement in order to maintain a self-similar functional form of the spin-spin interaction (see Fig. S4b). For the data in Fig. 2c in the main text, we set the detuning to $\delta = \mu - \omega_{\text{COM}}^y = 2\pi \times 45$ kHz and the axial frequency to $\nu_x = 0.46, 0.37, 0.36, 0.31, 0.27$ MHz, for $N = 20, 25, 30, 35, 40$ respectively. For the data in Fig. 4 in the main text, the detuning is $\delta/2\pi = 45$ kHz and the $\nu_x = 0.54$ MHz.

C. Fitting Ising Couplings to Analytic Form. By directly measuring trap parameters and spin-phonon couplings, we can calculate the spin-spin interaction matrix J_{ij} with Eqs. (15) and (16). However, in order to efficiently compute the ground state energy of the Hamiltonian in Eq. (1) (see main text) for $N \gtrsim 25$ using DMRG, we approximate the Ising couplings using a translational invariant analytic function of the ion separation $r = |i - j|$. For $N < 20$ the spin-spin coupling J_{ij} between the two qubits at distance r is well approximated by a power law decay:
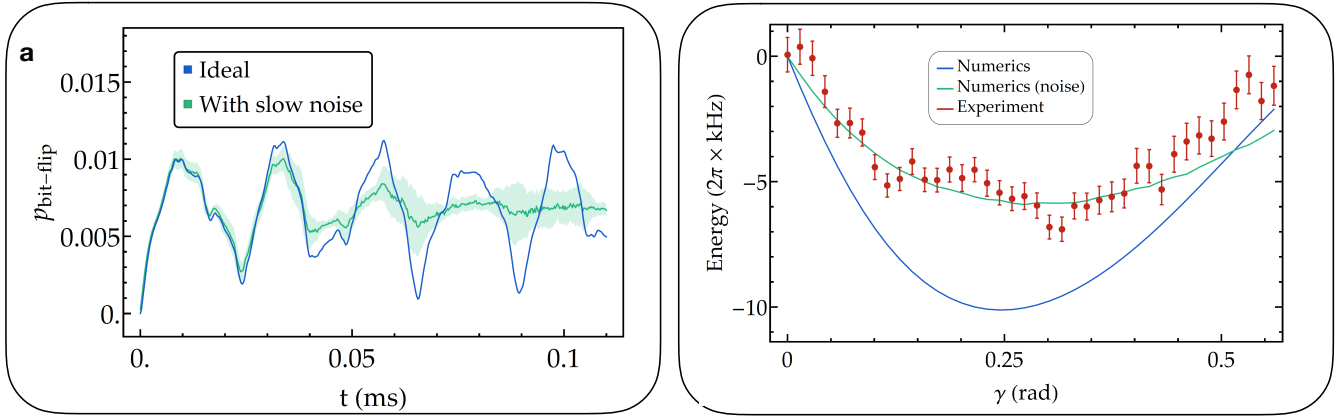
$$J_{ij} \approx \frac{J_0}{r^\alpha}, \qquad [17]$$

where, as stated in the main text, J_0 is the average nearest-neighbor coupling and α is the power law exponent (19). However for larger system sizes, this approximation fails to capture the actual decay of the interaction matrix.

In order to use the DMRG algorithm to accurately compute the ground state energies, we developed a compound function to better fit our couplings. This function is a product of a power law decay and an exponential decay parametrized by J_0, α' and β':

$$J_{ij} \approx \frac{J_0}{r^{\alpha'}} e^{-\beta'(r-1)} \qquad [18]$$

As seen in Fig. S5, this functional form fits well the exact Ising couplings even for a chain of 40 ions, while both a power law and a pure exponential fit diverge significantly.

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

Figure S6. Errors in trapped-ion quantum simulator: (a) Phonon-assisted bit-flips per ion predicted by evolving the coherent off-resonant spin-phonon drive for 12 ions. The simulation includes slow drifts of the trap frequency and of the laser power over 500 shots, each including a Hamiltonian evolution of 0.11 ms, with $\delta/2\pi = 45$ kHz and $\Omega/2\pi = 440$ kHz. The shaded region is defined as the average p_i plus and minus one standard deviation (see main text for details). (b) Energy as a function of the γ parameter scan for Fig. 4 in the main text. Taking into account our total bit-flip error budget together with uncompensated light shift, we explain most of the discrepancy between our experimental performance and the ideal QAOA energy output.

D. State Detection. We detect the ion spin state by globally rotating all the spins into the measurement basis with a composite BB1 $\pi/2$ pulse as described above, to rotate the x or y basis into the z basis), followed by the scattering of resonant laser radiation on the $^2\mathrm{S}_{1/2}|F = 1\rangle \leftrightarrow {}^2\mathrm{P}_{1/2}|F = 0\rangle$ cycling transition (wavelength near 369.5 nm and radiative linewidth $\gamma/2\pi \approx 20$ MHz). If the atom is projected in the $|\uparrow\rangle_z$ "bright" state, it fluoresces strongly, while if projected in the $|\downarrow\rangle_z$ "dark" state it fluoresces almost no photons because the laser is far from resonance (15).

In both systems the fluorescence of the ion chain is imaged onto an Electron Multiplying Charge Coupled Device (EMCCD) camera (Model Andor iXon Ultra 897) using an imaging objective with 0.4 numerical aperture and a magnification of 90x for both systems. The fluorescence of each ion covers roughly a 7x7 array of pixels on the EMCCD. After collecting the fluorescence for an integration time of 0.65 (1) ms for system 1 (2), we use a binary threshold to determine the state of each ion, discriminating the quantum state of each ion with near 98% (97%) accuracy in system 1 (2). The residual 2 (3)% errors include off-resonant optical pumping of the ion between spin states during detection as well as detector cross-talk between adjacent ions, readout noise, and background counts.

In system 2 the individual ion range-of-interests (ROIs) on the camera are updated with periodic diagnostic images, acquired by applying a nearly resonant cooling laser for 50 ms so that each ion fluoresces strongly regardless of its state. The signal to background noise ratio in the diagnostic shots is larger than 100, yielding precise knowledge of the ions' center locations and taking into account the slow $\sim 2\,\mu$m pk-pk drift due to thermal expansion/contraction of the cryostat. Ion separations range from 1.5 μm to 3.5 μm depending on the trap settings and the distance from the chain center, and are always much larger than the resolution limit of the imaging system. We utilize the pre-determined ion centers to process the individual detection shots and optimize the integration area on the EMCCD camera to collect each ion's fluorescence while minimizing cross-talk. We estimate cross-talk to be dominated by fluorescence from nearest-neighbor, which can cause a dark ion to be erroneously read as bright.

E. Error sources. The fidelity of the quantum simulation is limited by experimental noise that causes the system to depart from the ideal evolution and that can have several sources that are reviewed below. One important error source is off-resonant excitation of motional modes of the ion chain, which causes residual spin motion-entanglement. When the motion is traced out at the end of the measurement this results in a finite probability of an unwanted bit-flip. The probability of this error to occur on the ith ion (12) is proportional to $p_i \sim \sum_{m=1}^{N} (\eta_{im}\Omega/\delta_m)^2$, where $\eta_{im} = b_{im}\sqrt{\nu_R/\nu_{\mathrm{COM}}}$ (see Eq. (15)) and $\delta_m = \mu - \omega_m$ is the beatnote detuning from the m-th normal mode. We trade off a lower error for a weaker spin-spin coupling by choosing a δ_{COM} such that $(\eta_{\mathrm{COM}}\Omega/\delta_{\mathrm{COM}})^2 \lesssim 1/10$. By considering the off-resonant contributions of all the modes (see Fig. S6), we estimate the phonon error to cause about 1% bit-flip per ion. Additionally, bit-flip errors are affected by fluctuations in the trap frequency and laser light intensity at the ions' location. To take this into account, we included slow drifts and fluctuations of the trap frequency and of the laser power on the timescale of 500 experimental repetitions assuming noise spectral density falling as $1/f$. Given our typical trap frequency and laser power fluctuations, we assume a relative standard deviation $\Delta\Omega/\Omega \sim 2\%$ and $\Delta\delta_{\mathrm{COM}}/\delta_{\mathrm{COM}} \sim 9\%$ over the timescale required to average over quantum projection noise and we end up estimating an average bit-flip probability $p_i \sim 1\%$ (see Fig. S6a). Moreover, laser intensity, beam steering and trap frequency slow drifts over the time scale of a few hours required for data-taking cause averaging over different Ising parameters J_0. In particular, beam steering fluctuations create an imbalance between the red and blue $\nu_0 \pm \mu$ beatnotes at the ions, producing an effective B_z noisy field, that has been estimated to be as high as $\sim 0.65J_0$. To take into account these drifts, we calculated several evolutions sampling from a Gaussian distribution of values of B_z and J_0, using as a variance the standard deviations ($\sigma_{J_0} = 0.18J_0$ and $\sigma_{B_z} = 0.4J_0$) observed in the experiment. Another source of bit-flip errors is imperfect detection. Off-resonant pumping limits

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe

our average detection fidelity to 98%(97%) for system 1 (2). A detection error is equivalent to a random bit-flip event so the two errors will sum up. A specific source of noise in system 2 is mechanical vibrations at 41 Hz and 39 Hz due to residual mechanical coupling to the cryostat (14). This is equivalent to phase-noise on the Raman beams, which leads to dephasing of the qubits. Other less important noise sources are related to off-resonant Raman scattering errors during the Ising evolution (estimated in $7 \cdot 10^{-5}$ per ion) and RF heating of the transverse COM motional mode of the ion chain in system 1.

In Fig. S6b, we plot the experimentally measured energy as a function of γ, and the corresponding theoretical curves with and without incorporating errors. Using the time dependent average bit-flip probability evolution that we estimated from our error model considering phonons and detection errors and averaging over slow drifts in experimental parameters J_0 and B_z, we get a good agreement with the experimental data (see also Fig. 2c in the main text, where the same parameters have been used), showing that we have a good understanding of the noise sources in our system.

References

1. Farhi E, Goldstone J, Gutmann S (2014) A Quantum Approximate Optimization Algorithm. *arXiv:1411.4028*.
2. Zhou L, Wang ST, Choi S, Pichler H, Lukin MD (2018) Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices. *arXiv e-prints* p. arXiv:1812.01041.
3. Brady L, Bapat A, Gorshkov A (2019) in preparation.
4. Crooks GE (2018) Performance of the quantum approximate optimization algorithm on the maximum cut problem. *arXiv:1811.08419*.
5. Farhi E, Harrow AW (2016) Quantum supremacy through the quantum approximate optimization algorithm. *arXiv:1602.07674*.
6. Farhi E, Goldstone J, Gutmann S (2014) A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem. *arXiv:1412.6062*.
7. Dylewsky D, Freericks J, Wall M, Rey A, Foss-Feig M (2016) Nonperturbative calculation of phonon effects on spin squeezing. *Physical Review A* 93(1):013415.
8. Hadfield S (2018) Quantum algorithms for scientific computing and approximate optimization. *arXiv:1805.03265*.
9. Farhi E, Harrow AW (2016) Quantum supremacy through the quantum approximate optimization algorithm. *arXiv:1602.07674v1*.
10. Bremner MJ, Montanaro A, Shepherd DJ (2016) Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.* 117(8):080501.
11. Bouland A, Fefferman B, Nirkhe C, Vazirani U (2019) On the complexity and verification of quantum random circuit sampling. *Nat. Phys.* 15(2):159–163.
12. Kim K, et al. (2009) Entanglement and tunable spin-spin couplings between trapped ions using multiple transverse modes. *Physical Review Letters* 103(12).
13. Wineland D, et al. (1998) Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl. Inst. Stand. Technol.* 103:259–328.
14. Pagano G, et al. (2019) Cryogenic trapped-ion system for large scale quantum simulation. *Quantum Science and Technology* 4(1):014004.
15. Olmschenk S, et al. (2007) Manipulation and detection of a trapped yb^+ hyperfine qubit. *Phys. Rev. A* 76(5):052314.
16. Brown KR, Harrow AW, Chuang IL (2004) Arbitrarily accurate composite pulse sequences. *Phys. Rev. A* 70(5):052318.
17. Sørensen A, Mølmer K (1999) Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* 82(9):1971–1974.
18. James DFV (1998) Quantum dynamics of cold trapped ions with application to quantum computation. *Applied Physics B* 66:181.
19. Porras D, Cirac JI (2004) Effective Quantum Spin Systems with Trapped Ions. *Phys. Rev. Lett.* 92(20):207901.

G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe