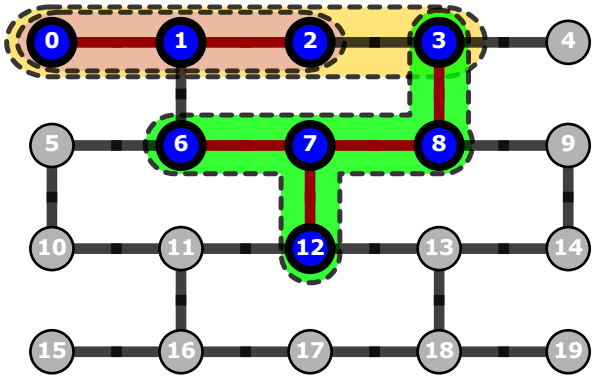
Supplementary material for "Analyzing the Performance of Variational Quantum Factoring on a Superconducting Quantum Processor"

Amir H. Karamlou[1,2,*] William A. Simon[1,*] Amara Katabarwa[1],
Travis L. Scholten[3], Borja Peropadre[1], and Yudong Cao[1]
[1] *Zapata Computing, Boston, MA 02110 USA*
[2] *Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and*
[3] *IBM Quantum, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598*

SUPPLEMENTARY METHODS

Device Information



Supplementary Figure 1. **IBMQ Boeblingen connectivity map**. To factor 1099551473989 we use qubits 0, 1, and 2 (shaded in maroon). To factor 3127 we use qubits 0, 1, 2, and 3 (shaded in yellow). To factor 6557 we use qubits 3, 6, 7, 8, and 12 (shaded in green).

For our experiments we use the 20 qubit Boeblingen system. This quantum processor consists of 20 fixed-frequency transmon qubits, with microwave-driven single and two-qubit gates, and with a connectivity map as depicted in Supplementary Figure 1. For the experiments discussed in this paper use the CNOT as our native two-qubit gate which is driven via the cross-resonance interaction between coupled qubits. Experiments were conducted on the highlighted qubits, for their respective single and two qubit gate fidelities have a higher fidelity, as well as a relatively low readout error. Qubit properties and two-qubit gate characterization can be found in Supplementary Table I and II respectively. For factoring 1099551473989 we use Q0, Q1, and Q2, for factoring 3127 we use Q0, Q1, Q2, and Q3, and for factoring 6557 we use Q3, Q6, Q7, Q8, and Q12.

Residual ZZ-coupling

In addition to the well-known incoherent processes of relaxation and dephasing, there are coherent noise mechanisms that may affect the overall algorithm performance. Taking into account these coherent errors has become increasingly important in near term algorithms, for their effects cannot be captured by standard benchmark techniques (such as randomized benchmarking), which may impact the overall performance of a quantum algorithm. For superconducting qubits in general, and especially for fixed-frequency transmons, these coherent sources of error are qubit crosstalk

---

* These authors contributed equally.

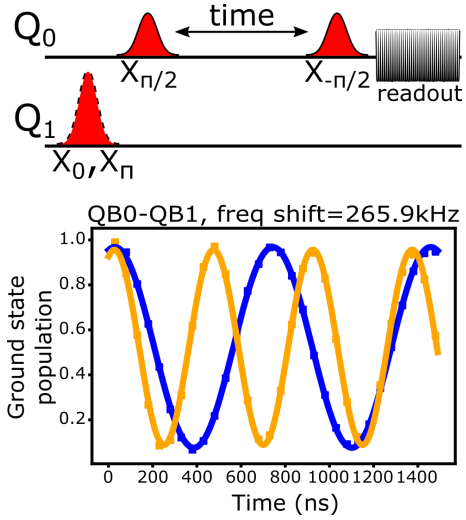
Property		Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q12
1QB Gate Error	mean	4.28e-04	3.08e-04	2.83e-04	4.03e-04	7.07e-04	5.19e-04	7.37e-04	3.16e-04	4.20e-04	4.08e-04	3.47e-04
	std	1.56e-04	5.09e-05	7.98e-05	6.61e-05	4.13e-04	1.05e-04	5.230e-04	4.80e-05	8.72e-05	6.20e-05	3.54e-05
Frequency (GHz)	mean	5.05	4.85	4.70	4.77	4.37	4.91	4.73	4.55	4.66	4.77	4.74
	std	2.42e-06	4.35e-06	3.18e-06	3.73e-06	1.01e-05	9.48e-05	1.49e-05	4.26e-06	4.84e-06	6.37e-05	3.26e-06
Readout Error (%)	mean	2.48	2.73	3.68	2.28	5.13	1.95	4.69	3.80	5.76	6.34	4.59
	std	.523	.348	.964	.373	4.27	.711	1.54	.805	.352	2.43	.465
T1 (μs)	mean	62.8	59.1	105	80.2	87.3	80.0	64.2	81.0	44.8	57.5	80.5
	std	21.3	11.5	21.7	8.98	17.4	19.5	15.0	14.0	12.4	5.61	17.7
T2 (μs)	mean	97.4	86.5	104	42.7	76.9	54.2	75.0	88.9	69.6	84.4	107
	std	38.6	21.9	31.2	4.38	32.3	16.3	20.7	17.7	18.1	40.0	26.1

Supplementary Table I. **Single-qubit characterization**. Data acquired over the span of 14 days.

Coupling	2QB Gate Error	Time (ns)
Q0-Q1	7.38e-03 ± 9.7e-04	220
Q1-Q2	6.87e-03 ± 9.9e-04	334
Q2-Q3	9.75e-03 ± 20.8e-04	277
Q6-Q7	13.55e-03 ± 55.8e-04	256
Q7-Q8	10.76e-03 ± 10.4e-04	412
Q7-Q12	14.96e-03 ± 86.1e-04	306
Q8-Q3	10.06e-03 ± 12.3e-04	363

Supplementary Table II. **Two-qubit properties**. Data acquired over the span of 14 days.

due to microwave leaking out from one qubit to another while driving single qubit gates, and residual ZZ-coupling between connected qubits, due to the relatively small anharmonicity of the transmons [1]. Qubit crosstalk is relatively harmless in variational algorithms, for its effect, namely single qubit overrotations, can be learned and cancelled out by the classical optimizer on each optimization step. However, the residual ZZ noise which accumulates during the gates cannot be learned by the optimizer, as it doesn't commute with the native ZX term from the CR gate, and needs to be taken into account in the quantum circuit.



Supplementary Figure 2. **Measuring the strength of the residual ZZ-coupling between pairs of qubits.** For each of the pairs of qubits, we measure the Ramsey oscillation frequency of each qubit when the coupled qubit in question is in the $|0\rangle$ state and in the $|1\rangle$ state. The strength of the residual ZZ-coupling is then the difference between these frequencies.

In order to estimate the effect of the residual ZZ-coupling in actual experiments, it is crucial to derive an accurate

Coupling	$\xi/2\pi$ (kHz)	Coupling	$\xi/2\pi$ (kHz)
Q0-Q1	275.1 ± 6.7	Q8-Q9	135.6 ± 0.6
Q1-Q2	226.7 ± 3.4	Q10-Q11	-249.8 ± 0.4
Q1-Q6	66.1 ± 5.8	Q11-Q12	242.6 ± 3.7
Q2-Q3	86.3 ± 0.8	Q11-Q16	79.9 ± 0.4
Q3-Q4	-116.0 ± 3.5	Q12-Q13	79.3 ± 3.0
Q3-Q8	54.7 ± 1.5	Q13-Q14	67.7 ± 3.3
Q5-Q6	186.4 ± 3.7	Q13-Q18	53.9 ± 2.9
Q5-Q10	81.1 ± 1.6	Q15-Q16	119.5 ± 3.2
Q6-Q7	117.8 ± 5.2	Q16-Q17	108.0 ± 1.6
Q7-Q8	69.8 ± 0.2	Q17-Q18	98.8 ± 13.4
Q7-Q12	74.9 ± 13.2	Q18-Q19	383.6 ± 13.3

Supplementary Table III. **Reporting the measured strength of the residual ZZ-coupling between all pairs of qubits.** Data is averaged over two trials per connected pair of qubits.

Hamiltonian model that captures the effect of higher energy levels in the energy spectrum. To this end, we sketch the derivation of an effective Hamiltonian for two transmon qubits interacting through a common transmission line resonator (which corresponds to qubits connected through solid lines in Supplementary Figure 1). We model the transmon qubit j as a Duffing oscillator

$$H_{\mathrm{qb}_i} = \omega_0(b_i^\dagger b_i + 1/2) + \frac{\delta_i}{2} b_i^\dagger b_i (b_i^\dagger b_i - 1), \tag{1}$$

that interacts with other transmons via the transmission line resonator through the exchange of virtual excitations

$$H = \omega_{\mathrm{res}} a^\dagger a + H_{\mathrm{qb}_1} + H_{\mathrm{qb}_2} + \sum_{j=1}^{2} g_j (a^\dagger b_j + H.c), \tag{2}$$

where g_j represents the coupling constant between transmon j and the resonator. We can diagonalize this Hamiltonian through a Schrieffer-Wolff transformation to adiabatically remove the resonator effect, getting an effective transmon-transmon interaction that reads

$$H = \sum_{j=1}^{2} \omega_j b_j^\dagger b_j + J(b_1^\dagger b_2 + b_2^\dagger b_1), \tag{3}$$

where $J = -g_1 g_2 \times (\Delta_1 + \Delta_2)/2(\Delta_1 \Delta_2)$, is the effective coupling between transmons, and $\Delta_j = \omega_{\mathrm{res}} - \omega_j$ the detuning between the transmon j and resonator. We now bring the above Hamiltonian to its diagonal form, as it represents the physical basis where the qubits are actually measured, and project it to the two-qubit subspace, yielding

$$H = \sum_{j=1}^{2} \tilde{\omega}_j Z_j + \xi Z_1 Z_2, \tag{4}$$

where $\tilde{\omega}_j$ are (dressed) qubit frequencies that are actually measured in an experiment, and $\xi = 2J^2(\delta_1 + \delta_2/(\delta_1 - \Delta)(\delta_2 + \Delta))$. The residual ZZ-coupling between two qubits, Q_i and Q_j, can be experimentally measured by finding the difference in the Ramsey oscillation frequency of Q_i while Q_j is in the $|0\rangle$ and $|1\rangle$ state (Supplementary Figure 2). The residual ZZ-coupling between the qubits used in our experiments can be found in Supplementary Table III.

In order to model the effects of the residual ZZ-coupling in simulation, we implement the $\mathrm{ECR}_{2-pulse}$ protocol as follows:

$$\mathrm{ECR}_{2-\mathrm{pulse}}(\mathrm{t}) = [XI]CR(-t/2)[XI]CR(t/2) \tag{5}$$

As mentioned, in the presence of the residual ZZ-coupling between qubits, the axis of rotation for the cross-resonance gate is fundamentally altered. When accounting only for the residual ZZ-coupling coupling between the qubits involved in the cross-resonance gate, we model the effect as follows:

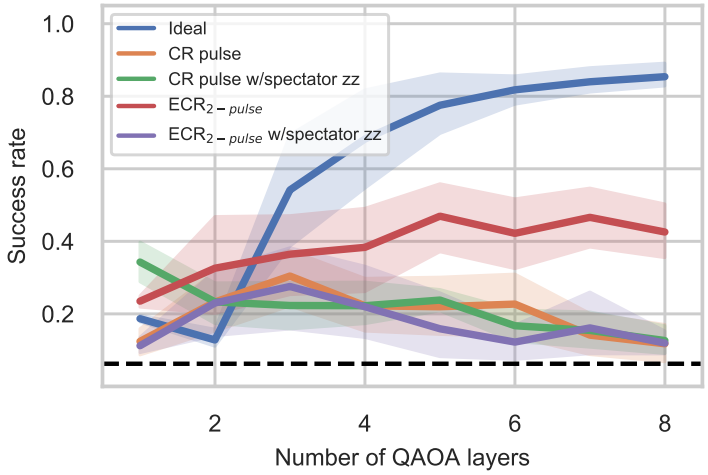
$$\Xi = \xi_{c,t} Z_c Z_t \tag{6}$$

where $\xi_{c,t}$ is the strength of the residual ZZ-coupling between the control (c) qubit and the target (t) qubit.

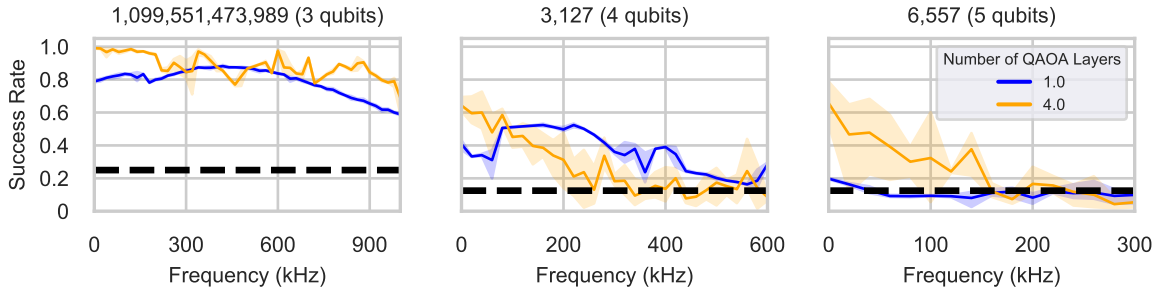
However, the presence of residual ZZ-couplings between each of these two qubits and its respective spectator qubits also alters the effective cross-resonance gate. When accounting for the involved spectator qubits, the effect can be modeled as:

$$\Xi = \xi_{c,t} Z_c Z_t + \sum_{i=0}^{U} \xi_{c,i} Z_c Z_i + \sum_{j=0}^{V} \xi_{t,j} Z_t Z_j \tag{7}$$

where U and V are the set of spectator qubits of the control qubit and target qubit respectively. All values of ξ are as reported in Supplementary Table III. In Supplementary Figure 3 we show how different implementations of the effect of the residual ZZ-couplings impact the success rate of VQF on 6557. We note that when spectator qubits are not considered, the $\mathrm{ECR}_{2-pulse}$ protocol [2] shows significant improvement over the single cross-resonance gate. However, when the residual ZZ-coupling of spectator qubits is considered, this improvement is lost and the success rates trend rapidly toward the random success rate.



Supplementary Figure 3. **Effects of residual ZZ-coupling on spectator qubits with different cross-resonance gate decompositions**. We compare the effects of incorporating the residual ZZ-coupling noise on spectator qubits during the cross resonance gate under both the single CR pulse and $\mathrm{ECR}_{2-pulse}$ protocols. The shaded regions indicate one standard deviation of uncertainty.



Supplementary Figure 4. **Effect of residual ZZ-coupling frequency on success rate**. For each VQF instance, 1,099,551,473,989 (left), 3,127 (middle), and 6,557 (right), we show how different frequencies of the residual ZZ-coupling affect the overall success rate of the algorithm when using 1 and 4 layer(s) of the ansatz. The shaded regions indicate one standard deviation of uncertainty.

We show how the average frequency of the residual ZZ-coupling between qubits impacts the various problem instances to differing degrees in Supplementary Figure 4. The results for the three qubit instance show that the success rate is not significantly diminished until the average frequency surpasses approximately 600 kHz. We note that, for the device used in this work, the largest frequency measured for a specific coupling was less than 400 kHz

(Supplementary Table III). However, for the four and five qubit instances, especially when using 4 ansatz layers, we see that any non-zero frequency of the residual ZZ-coupling causes the success rate to be reduced. For both of these instances, the success rate approaches the random success rate for frequencies that were measured on the device used in this work.

Impact of coherent error on the QAOA landscape

We start from analyzing how the crosstalk term affects the implementation of a CNOT gate, and in turn affects the implementation of a pairwise evolution term $\exp(-i\gamma Z_1 Z_2)$. Based on the discussion in the previous supplemental section, a cross resonance gate with qubit cross talk takes the form of

$$\widetilde{\mathrm{CR}}(t) = \exp\left[-i(\Gamma Z_1 X_2 + \xi Z_1 Z_2)t\right], \tag{8}$$

in contrast with the ideal cross resonance gate $\mathrm{CR}(t) = \exp(-iZ_1X_2t)$. For simplicity in discussion, we introduce normalization $\bar{\Gamma} = \frac{\Gamma}{\sqrt{\Gamma^2+\xi^2}}$ and $\bar{\xi} = \frac{\xi}{\sqrt{\Gamma^2+\xi^2}}$ and rescaling $\tau = \sqrt{\Gamma^2+\xi^2}t$. To gauge the impact of the coherent noise caused by qubit crosstalk, we compare the ideal CNOT gate sequence U_{CNOT} with the actual sequence $\widetilde{U}_{\mathrm{CNOT}}$ implemented:

$$U_{\mathrm{CNOT}} = e^{-i\frac{\pi}{4}} e^{i\frac{\pi}{4}Z_1} e^{-i\frac{\pi}{4}Z_1X_2} e^{i\frac{\pi}{4}X_2} \tag{9}$$

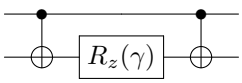
$$\widetilde{U}_{\mathrm{CNOT}} = e^{-i\frac{\pi}{4}} e^{i\frac{\pi}{4}Z_1} e^{-i\frac{\pi}{4}Z_1(\bar{\Gamma}X_2+\bar{\xi}Z_2)} e^{i\frac{\pi}{4}X_2}. \tag{10}$$

This comparison leads to $\widetilde{U}_{\mathrm{CNOT}} = U_{\mathrm{CNOT}} + Q$, where the difference Q reads

$$Q = \frac{I - iZ_1}{2}\left[(\bar{\Gamma}-1)X_2 + \bar{\xi}Z_2\right]\frac{I+iX_2}{\sqrt{2}}. \tag{11}$$

Note that $\|Q\| \le \sqrt{2}\cdot\sqrt{(\bar{\Gamma}-1)^2+\bar{\xi}^2}$. Here we will use the quantity $\delta = \sqrt{(\bar{\Gamma}-1)^2+\bar{\xi}^2}$ to measure the amount of deviation that the crosstalk is causing from the ideal operations.

Suppose we would like to realize the pairwise evolution term $U_{\mathrm{ZZ}} = \exp(-i\gamma Z_1 Z_2)$ using the following circuit (with $R_z(\gamma) = \exp(-i\gamma Z)$):

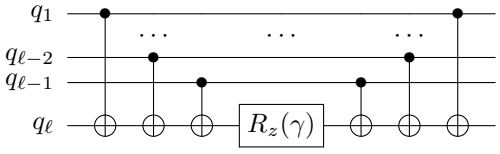


(12)

Then the deviation of the coherent error can be described as

$$\widetilde{U}_{\mathrm{ZZ}} = U_{\mathrm{ZZ}} + \underbrace{Qe^{-i\gamma Z_2}U_{\mathrm{CNOT}} + U_{\mathrm{CNOT}}e^{-i\gamma Z_2}Q}_{O(\delta)} + \underbrace{Qe^{-i\gamma Z_2}Q}_{O(\delta^2)}. \tag{13}$$

Note that here the deviation contains contributions that are both first and second order in δ, because there are two CNOT gates in the circuit (14). We can generalize this set up to the case where we need to implement an ℓ-local evolution $\exp(-i\gamma Z_1 Z_2 \cdots Z_\ell)$. In this scenario the circuit implementation will be:



(14)

The first order contribution of the coherent error is then $O(\ell\delta)$. For a Hamiltonian $H = \sum_{k=1}^{K}\alpha_k H_k$ where each H_k is a product of local Z operators, the first contribution of the coherent error to the QAOA ansatz step $e^{-i\gamma H}$ is $O(K\ell\delta)$. To demonstrate more concretely such contribution, we use the following 2-qubit example: The Hamiltonian is $H = Z_1Z_2$ and the ansatz $U(\gamma,\beta) = e^{-i\beta(X_1+X_2)}e^{-i\gamma Z_1Z_2}$. In the ideal case, the optimization landscape is

$$f(\gamma,\beta) = \langle\gamma,\beta|H|\gamma,\beta\rangle = \langle++|U^\dagger(\gamma,\beta)HU(\gamma,\beta)|++\rangle = 2\sin(4\beta)\sin(2\gamma). \tag{15}$$

In the case where there is coherent error due to qubit crosstalk, using the circuit in (14) for implementing the time evolution $e^{-i\gamma Z_1 Z_2}$, we find that up to first order in δ the impact of coherent error on the optimization landscape can be characterized by the new optimization landscape:

$$\begin{aligned}\widetilde{f}(\gamma,\beta) &= 2\left[1-\frac{1}{2}\mathrm{Re}\left(\langle++|Q|++\rangle-\langle--|Q|+-\rangle\right)\right]\sin(4\beta)\sin(2\gamma)\\ &+ \mathrm{Re}\left(\langle+-|Q|++\rangle+\frac{1}{2}\langle--|Q|++\rangle-\frac{1}{2}\langle++|Q|+-\rangle\right)\cos(4\beta)\cos(2\gamma)\\ &+ \mathrm{Re}\left(\langle+-|Q|++\rangle+\frac{1}{2}\langle--|Q|++\rangle+\frac{1}{2}\langle++|Q|+-\rangle\right)\cos(4\beta)+O(\delta^2)\end{aligned} \tag{16}$$

Here the matrix elements of Q can be calculated from (11). By comparing (15) and (16) one can observe that the coherent error does not affect the frequency spectrum of the landscape fluctuation as a function of either β or γ, but instead alters the weights or coefficients of different frequency components. This can be explained by noting that the coherent error only affects the CNOT gates, which are independent of any of the parameters β or γ. In general, for a QAOA ansatz, the optimization landscape $f(\gamma,\beta)$ for the last layer of the ansatz is essentially a trigono-multilinear function in β and γ:

$$f(\gamma,\beta)=\sum_{\omega} a_\omega \sin(2n\beta)\sin(\omega\gamma)+b_\omega\cos(2n\beta)\sin(\omega\gamma) \tag{17}$$

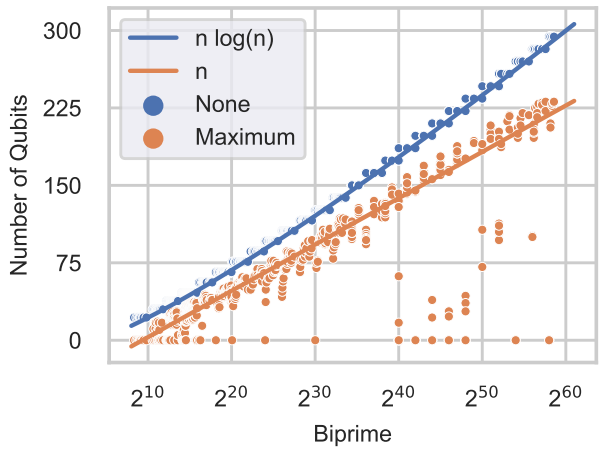
where we ignore the constant shift terms and ω sums over a set of possible values that depend on the coeffients $\{\alpha_k\}$ in the Hamiltonian $H=\sum_k \alpha_k H_k$. The general impact of coherent error on the landscape is that it alters the coefficients a_ω and b_ω of the various frequency components ω:

$$\widetilde{f}(\gamma,\beta)=\sum_{\omega} \widetilde{a}_\omega \sin(2n\beta)\sin(\omega\gamma)+\widetilde{b}_\omega\cos(2n\beta)\sin(\omega\gamma). \tag{18}$$

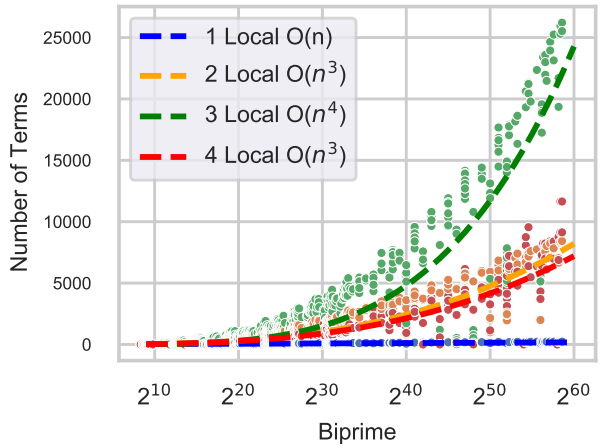
VQF Preprocessing

By simplifying the clauses using binary rules we reduce the quantum resources required for executing the VQF algorithm. The classical preprocessing procedure iterates through all clauses $\{C_i\}$ a constant number of times, using a set of binary rules to solve or make deductions where possible. Assuming $x,y,z\in F_2, a\in Z^+$ we apply the following classical preprocessing rules:

1. $xy=1\Rightarrow x=y=1$
2. $x+y=1\Rightarrow xy=0$
3. $x+y=2z\Rightarrow x=y=z$
4. $\sum_{i=1}^{a} x_i=a\Rightarrow x_i=1$
5. if $x_i=1$ violates $\max(lhs)=\max(rhs)\Rightarrow x_i=0$
6. if $\min(lhs)<\min(rhs)$ and x is the only variable on lhs $\Rightarrow x=1$
7. the parity of the lhs must be the same as the rhs



Supplementary Figure 5. **Empirical results on number of qubits required for factoring a biprime N after classical preprocessing.** n represents the number of bits needed to represent the given biprime.



Supplementary Figure 6. **Empirical results on number of n-local terms in the hamiltonain for factoring a biprime N after classical preprocessing.**

The runtime of the classical preprocessor is $O(n^2)$ [3]. Without preprocessing the VQF qubit requirements scale as $O(n\log n)$, whereas this resource bound empirically gets reduced to $O(n)$ using the classical preprocessor (Supplementary Figure 5). We additionally note that for certain biprimes, the preprocessor is able to significantly reduce the number of required qubits, in some cases completely solving the clauses. Additionally, in Supplementary Figure 6, we show the scaling of 1-local, 2-local, 3-local, and 4-local terms in the resulting hamiltonians, after classical preprocessing, for factoring a given biprime N.

Metrics

$$\hat{H}_{1099551473989} = 2.0 + 0.5(Z_0Z_1) + 1.0(Z_0Z_1Z_2) + 1.5(Z_2) \tag{19}$$

$$\hat{H}_{3127} = 4.5 - 0.5(Z_0) - 0.5(Z_0Z_1Z_2) - 1.0(Z_0Z_3) + 2.0(Z_1Z_2) + 1.5(Z_1Z_2Z_3) + 3.0(Z_3) \tag{20}$$

$$\hat{H}_{6557} = 12.0 + 2.0(Z_0Z_1) + 1.5(Z_0Z_1Z_2) + 1.0(Z_0Z_1Z_3) - 0.5(Z_0Z_1Z_4) + 4.0(Z_2) - 4.0(Z_3) - 5.0(Z_3Z_4) + 2.0(Z_4) \tag{21}$$

	CNOT gates	Single-qubit gates	Circuit depth
1st layer	4	22	13
2nd layer	8	41	25
3rd layer	12	60	37
4th layer	16	79	49
5th layer	20	98	61
6th layer	24	117	73
7th layer	28	136	85
8th layer	32	155	97

Supplementary Table IV. **Properties of QAOA circuits 1099551473989 (3 qubits)**. The qubit mapping used for this instance is: [0, 1, 2].

	CNOT gates	Single-qubit gates	Circuit depth
1th layer	18	12	21
2th layer	37	20	39
3th layer	59	28	58
4th layer	81	36	77
5th layer	103	44	96
6th layer	125	52	115
7th layer	147	60	134
8th layer	169	68	153

Supplementary Table V. **Properties of QAOA circuits 3127 (4 qubits)**. The qubit mapping used for this instance is: [0, 1, 2, 3].

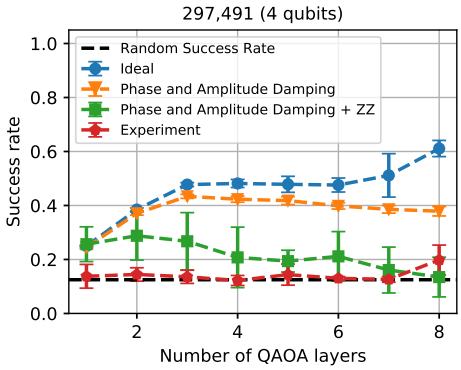
	CNOT gates	Single-qubit gates	Circuit depth
1st layer	13	44	25
2nd layer	29	86	52
3rd layer	45	128	76
4th layer	94	203	116
5th layer	95	230	132
6th layer	147	308	176
7th layer	147	334	197
8th layer	162	375	226

Supplementary Table VI. **Properties of QAOA circuits on 6557 (5 qubits)**. The qubit mapping used for this instance is: [6, 7, 12, 8, 3].

In equations 19, 20, and 21 we show the hamiltonians for each of the problem instances studied and in Tables IV, V, and VI, we show the relevant properties for the circuits used in both simulation and experiment. The terms in the hamiltonians are mapped to the physical qubit subsets shown in Supplementary Figure 1. We use the Qiskit transpiler, with the optimization level set to 1, to create the circuits with the correct gate set for IBMQ devices. Due to the non-deterministic nature of this transpiler, we independently run the transpilation process 10 times, using the circuit with the least number of cx gates.

297491 (4 qubits)

In Supplementary Figure 7, we show the results of running the VQF algorithm on the integer 297491 using 4 qubits (6, 7, 8, and 12). We present the success rate of the VQF algorithm as a function of the number of layers in our QAOA ansatz and compare results from experiment with those from simulation with different noise models. In contrast to the



Supplementary Figure 7. **Success rates for running VQF on the integer 297491 using 4 qubits**. The qubits used for this instance are 6, 7, 8, and 12. We compare the results obtained from experiments (red line) to ideal simulation (blue line), simulation containing phase and amplitude damping noise (orange line), and the residual ZZ-coupling (green line). The amplitude damping rate (T1≈64μs), dephasing rate (T2≈82μs), and residual ZZ-coupling (Supplementary Table III) reflect the values measured on the quantum processor. The error bars indicate one standard deviation of uncertainty.

results for factoring 1099551473989, 3127, and 6557, we observe no improvement in the success rate during experiment versus random sampling when factoring 297491 with 4 qubits. Additionally, the residual ZZ-coupling between qubits does not alone explain the degradation in the success of the algorithm when run on the quantum processor. This suggests the presence of an additional source of noise that is unaccounted for, even though the qubits used in this instance are a subset of the ones used to factor 6557.

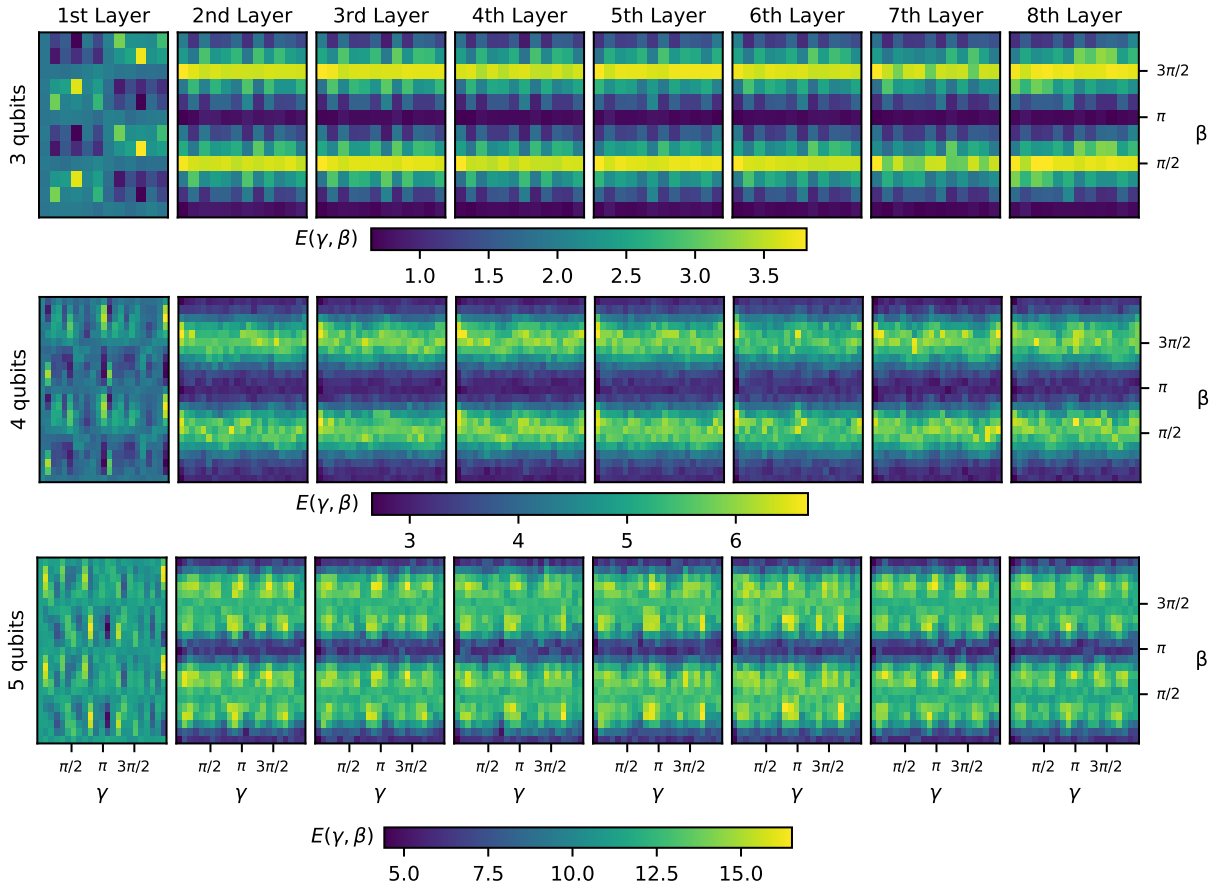
Higher depth energy landscapes

In Supplementary Figure 8 we display the energy landscapes for the instances representing 1099551473989, 3127, and 6557 using 3, 4, and 5 qubits respectively for $p = (1, 2, \cdots, 8)$. We observe that the landscapes for $p = (2, \cdots, 8)$ follow a similar pattern, not only between layers, but across instances as well, where the landscape depicts a negligible dependence on the variational ansatz parameter, γ, and a convex relationship with β. These relationships can be reasonably explained by the physical interpretation of the variational ansatz parameters and the implemented parameter initialization strategy.

The QAOA algorithm is designed to mimic the behavior of quantum annealing, where the wavefunction, $|\psi\rangle$, is initialized in an easy to prepare ground state and then alternately evolved under a mixer hamiltonian and then under a cost hamiltonian. The time under which the wavefunction is evolved by these two hamiltonians is represented by the variational ansatz parameters, β and γ, respectively. Consequently, if a state closely approximates the ground state of the cost hamiltonian, further time evolution under the cost hamiltonian should not yield a significant change in the prepared state. In turn, this implies that the energy of $|\psi\rangle$ with respect to the cost hamiltonian will not depend on γ. Similarly, yet opposite in effect, further evolution under the mixer hamiltonian would only bring the prepared state further from the ground state. Hence, creating a periodic relationship between β and the energy, where minima occur at points where the prepared state does not evolve further under the mixer hamiltonian.

Since this arises when $|\psi\rangle$ approximates the ground state, the layer-by-layer grid search strategy employed where we fix $\boldsymbol{\gamma} = (\gamma_0, .., \gamma_{p-2})$ and $\boldsymbol{\beta} = (\beta_0, .., \beta_{p-2})$ to the optimal parameters found in previous $p-1$ iterations, produces the condition where $|\psi\rangle$ already approximates the ground state of the cost hamiltonian after the first $p-1$ layers of the ansatz. Hence, the repeated pattern in the energy landscapes, shown in Supplementary Figure 8, emerges at $p > 1$ and becomes more pronounced as p increases, for each instance. However, it is worth noting that although these additional layers do not always produce better ground state approximations immediately after the parameter initialization, the second step of the employed optimization process - wherein all ansatz parameters, $(\gamma_0, .., \gamma_{p-1})$ and $(\beta_0, .., \beta_{p-1})$, are refined - typically leads to better solution states.

Additionally, we note that we continue to observe a discernible structure in the landscapes for high-depth circuits. As reported in Supplementary Table VI, the landscape produced for 6557 with 8 layers requires approximately 162 CNOT gates, 375 single-qubit gates, and a circuit depth of 226.



Supplementary Figure 8. **QAOA optimization landscapes**. For each VQF instance (rows), we show the resulting optimization landscape as a function of the number of QAOA layers in the ansatz (columns). For the 3 qubit instance (upper), the landscapes are obtained with a 12x12 grid, resulting in 144 uniformly spaced points. Whereas the landscapes obtained for the 4 and 5 qubit instances (middle and bottom, respectively) are created using a 23x23 grid with 529 uniformly spaced points.

[1] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Applied Physics Reviews **6**, 21318 (2019), arXiv:1904.06560.

[2] N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta, PRX Quantum **1**, 020318 (2020).

[3] E. R. Anschuetz, J. P. Olson, A. Aspuru-Guzik, and Y. Cao, (2018), arXiv:1808.08927v1.