Title: Optimizing Quantum Optimization

Speakers: Stefan Leichenauer

Collection: Machine Learning for Quantum Design

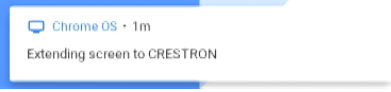
Date: July 10, 2019 - 10:45 AM

URL: http://pirsa.org/19070008

Abstract: Variational algorithms for a gate-based quantum computer, like the QAOA, prescribe a fixed circuit ansatz --- up to a set of continuous parameters --- that is designed to find a low-energy state of a given target Hamiltonian. After reviewing the relevant aspects of the QAOA, I will describe attempts to make the algorithm more efficient. The strategies I will explore are 1) tuning the variational objective function away from the energy expectation value, 2) analytical estimates that allow elimination of some of the gates in the QAOA circuit, and 3) using methods of machine learning to search the design space of nearby circuits for improvements to the original ansatz. While there is evidence of room for improvement in the circuit ansatz, finding an ML algorithm to effect that improvement remains an outstanding challenge.

Optimizing Quantum Optimization

Stefan Leichenauer
X, the moonshot factory

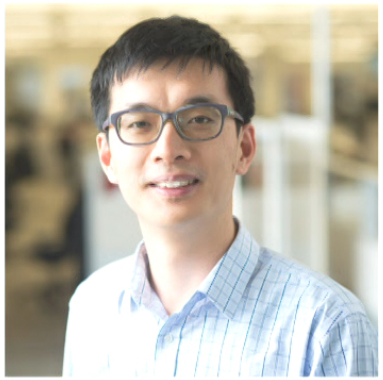


Collaborators

Friends from the Google Accelerated Science (GAS) team

Li Li

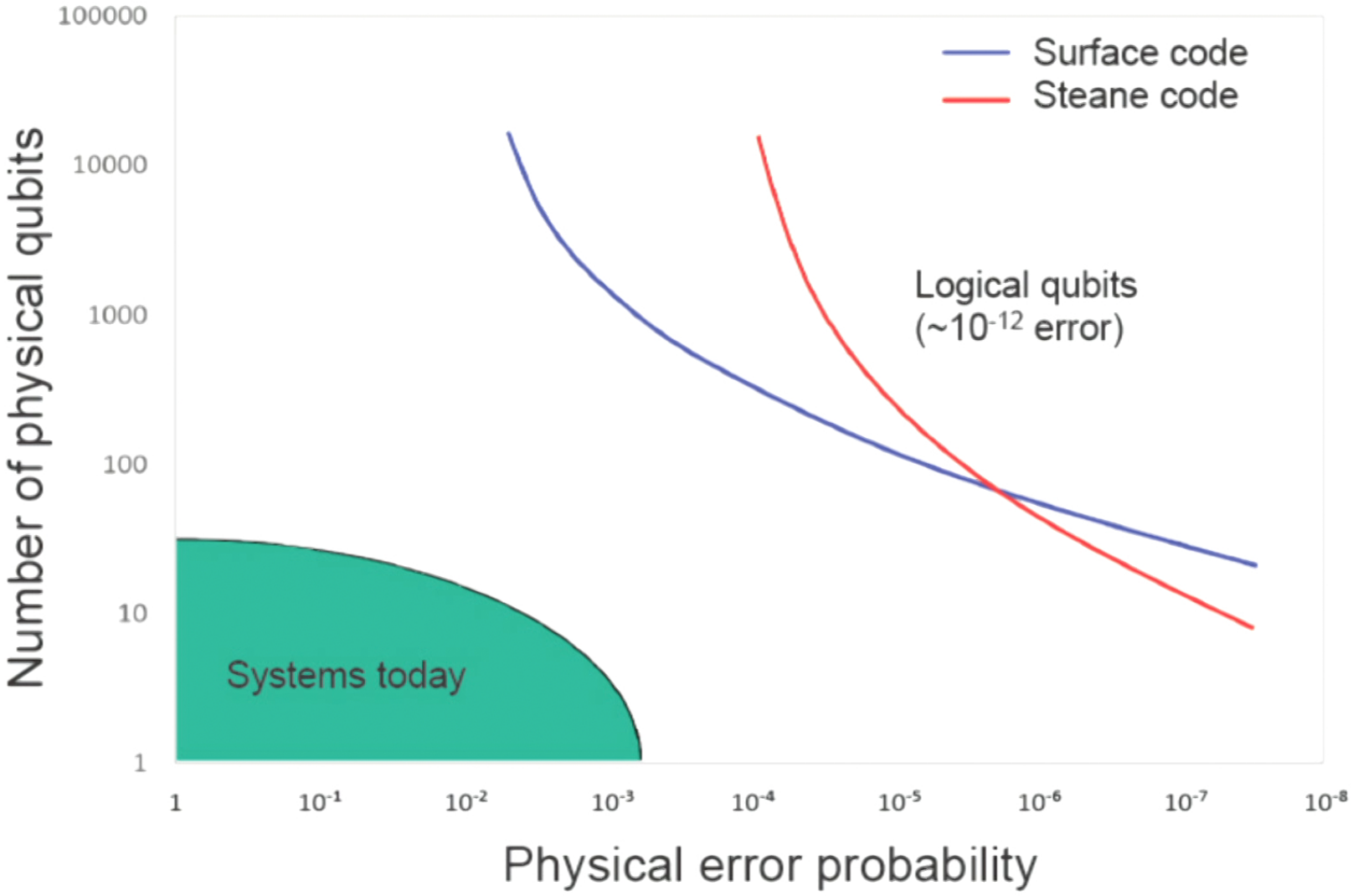
Minjie Fan



Patrick Riley

Marc Coram

Source: NAS Report Ch. 7



Quantum Computing in the NISQ Era

Using a Quantum Computer

Output is always a bitstring drawn from the Born distribution.

Try to arrange the distribution so that the bitstrings are useful.

1000s of samples in a reasonable amount of time? 10000? 100000?

Part of the goal is to limit the number of calls to the quantum computer.

Optimization with the QAOA

Quantum Approximate Optimization Algorithm

Farhi, Goldstone, Gutmann (1411.4028)

Objective: Find a low-energy state of a given computational-basis Hamiltonian.

More complicated Hamiltonians will be more difficult, but the recipe is general.

Variational: there are parameters to tune.

Based on adiabatic theorem/algorithm, but different in practice.

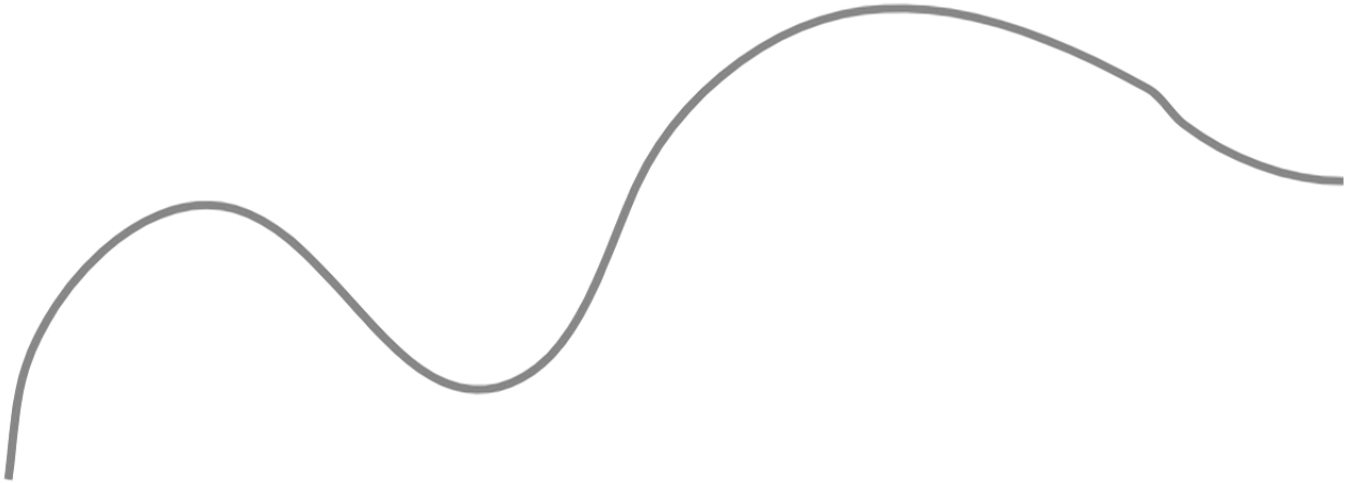
Quantum Approximate Optimization Algorithm

$$|\psi\rangle = \exp(-i\beta_p X)\exp(-i\gamma_p E)\cdots\exp(-i\beta_1 X)\exp(-i\gamma_1 E)H^{\otimes n}|0\rangle$$

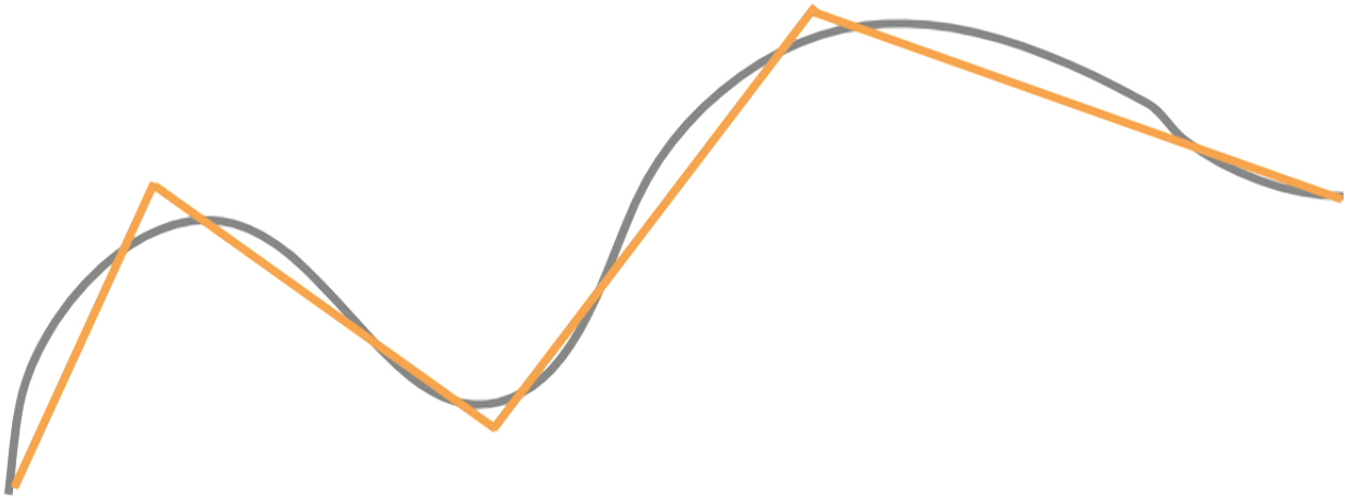
2p parameters to tune, try to get |ψ> to be a low energy state.

Really: try to get |ψ> to have a large amplitude in low-energy states. You don't actually care if |ψ> itself is low-energy.

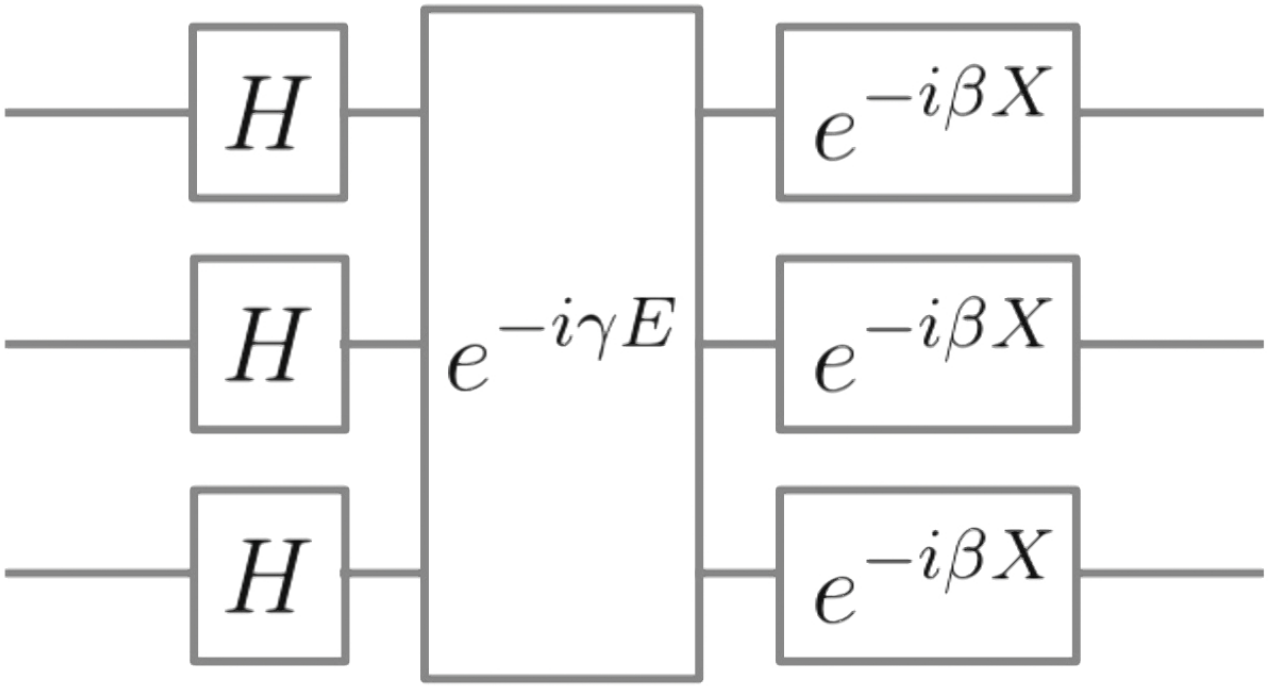
Quantum Approximate Optimization Algorithm



Quantum Approximate Optimization Algorithm



Quantum Approximate Optimization Algorithm



The QAOA According to Schrodinger

Compute the final state after applying the circuit.

Superposition over bitstrings, hopefully with more weight on the low-energy ones.

Can see how certain choices of the parameters lead to trivial results.

$$|\psi\rangle = \exp(-i\beta X)\exp(-i\gamma E)H^{\otimes n}|0\rangle$$

The QAOA According to Schrodinger

Initial state:

$$H^{\otimes n}|0\rangle = |+\rangle = \frac{1}{2^{n/2}} \sum_{z \in \{-1,1\}^n} |z\rangle$$

Uniform superposition over all bitstrings (equivalently, strings of +/-1).

Eigenstate of Pauli-X on each qubit.

The QAOA According to Schrodinger

Next apply Hamiltonian time evolution:

$$\exp(-i\gamma E)|+\rangle = \frac{1}{2^{n/2}} \sum_{z\in\{-1,1\}^n} e^{-i\gamma E(z)}|z\rangle$$

Hamiltonian is diagonal in computational basis.

Born probabilities still uniform on bitstrings.

The QAOA According to Schrodinger

Finally apply mixing Hamiltonian:

$$\exp(-i\beta X)\exp(-i\gamma E)|+\rangle = \frac{1}{2^{n/2}} \sum_{z\in\{-1,1\}^n} e^{-i\gamma E(z)} e^{-i\beta X}|z\rangle$$

Now interference can happen which depends on the energy levels.

By tuning the parameters we can arrange the state to be advantageous to us.

The QAOA According to Feynman

Compute the propagator:

$$\begin{aligned}\langle z'|e^{-i\beta X}e^{-i\gamma E}|z\rangle &= \langle z'|\prod(\cos\beta - iX\sin\beta)|z\rangle e^{-i\gamma E(z)} \\ &= \langle z'|\prod(1 - iX\tan\beta)|z\rangle \cos^n\beta e^{-i\gamma E(z)} \\ &= (-i\tan\beta)^{d(z',z)}\cos^n\beta e^{-i\gamma E(z)}\end{aligned}$$

d(z,z') is the Hamming distance

The QAOA According to Feynman

Suppose z_0 is the ground state, and each flip of a bit costs energy ε. Sum the propagator over uniformly over initial bitstrings:

$$\langle z_0 | e^{-i\beta X} e^{-i\gamma E} | + \rangle = \frac{\cos^n \beta}{2^{n/2}} \sum_z (-i \tan \beta e^{-i\gamma\epsilon})^{d(z_0,z)}$$

$$= \frac{\cos^n \beta}{2^{n/2}} \sum_d \binom{n}{d} (-i \tan \beta e^{-i\gamma\epsilon})^d$$

$$= \frac{1}{2^{n/2}} (\cos \beta - i \sin \beta e^{-i\gamma\epsilon})^n$$

The QAOA According to Feynman

The probability of measuring the ground state is then

$$P(z_0) = \frac{1}{2^n}|\cos\beta - i\sin\beta e^{-i\gamma\epsilon}|^{2n}$$
$$= \frac{1}{2^n}(1 - \sin 2\beta \sin\gamma\epsilon)^n$$

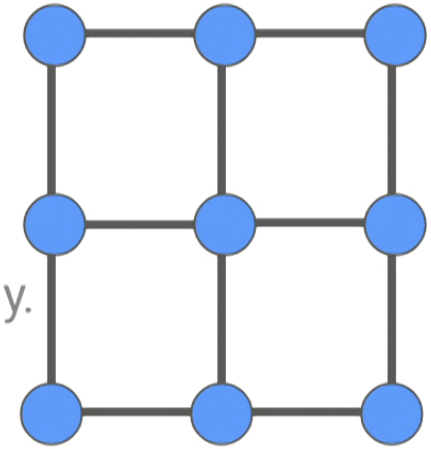
We can tune the parameters to make this probability equal to one.

Still need to use the quantum computer to extract z_0!

The QAOA According to Heisenberg

Focus: Ising model on a graph (e.g., square lattice) with random couplings.

$$E = \sum_{\langle ij \rangle} J_{ij} Z_i Z_j$$



Why this model?

Easy to put on a quantum computer, hard to solve exactly.

The QAOA According to Heisenberg

Hamiltonian evolution is a product over “ZZ” gates, simple two-qubit gates:

$$\exp\left(-i\gamma\sum_{\langle ij\rangle} J_{ij}Z_iZ_j\right) = \prod_{\langle ij\rangle}\exp(-i\gamma J_{ij}Z_iZ_j)$$

Easy to realize with single-qubit rotations and controlled phases:

$$aZ_i + bZ_j + c(1-Z_i)(1-Z_j) = (a-c)Z_i + (b-c)Z_j + c + cZ_iZ_j$$

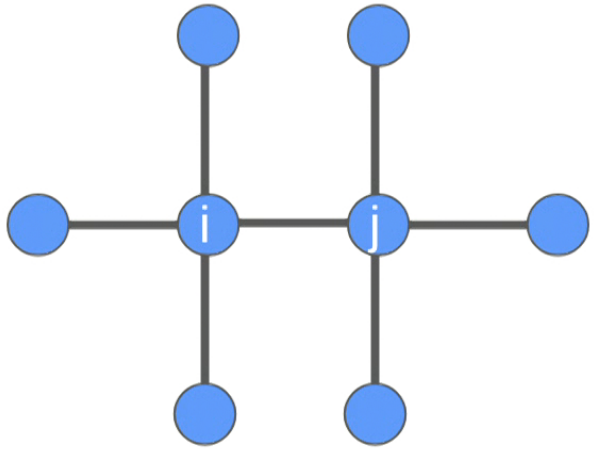
Only nearest-neighbor two-qubit gates.

The QAOA According to Heisenberg

Prescription is to minimize

$$\langle \psi | E | \psi \rangle$$

Isolate the ij edge. Use Heisenberg evolution for Z_i Z_j. Only neighbors of the ij edge will matter.



The QAOA According to Heisenberg

The mixing Hamiltonian just rotates the Pauli operators:

$$\exp(i\beta X)Z_i b \exp(-i\beta X) = Z_i \cos 2\beta + Y_i \sin 2\beta$$

Now the edge operators:

$$e^{i\gamma J_{ik} Z_i Z_k}(Z_i \cos 2\beta + Y_i \sin 2\beta)e^{-i\gamma J_{ik} Z_i Z_k} = Z_i \cos(2\beta) + Y_i e^{-i2\gamma J_{ik} Z_i Z_k} \sin 2\beta$$

We can sum over k in the exponent.

The QAOA According to Heisenberg

$$Z_i \cos(2\beta) + Y_i e^{-i2\gamma \sum_k J_{ik} Z_i Z_k} \sin 2\beta$$

There is a similar expression for site j as well. To have a nonzero expectation value in the |+> state, you need an even number of Z operators on each site. Then we can toss out several terms involving the neighbors of site i that are not site j:

$$Z_i \cos(2\beta) + Y_i e^{-i2\gamma J_{ij} Z_i Z_j} \prod_{k \neq j} \cos(2\gamma J_{ik}) \sin 2\beta$$

The QAOA According to Heisenberg

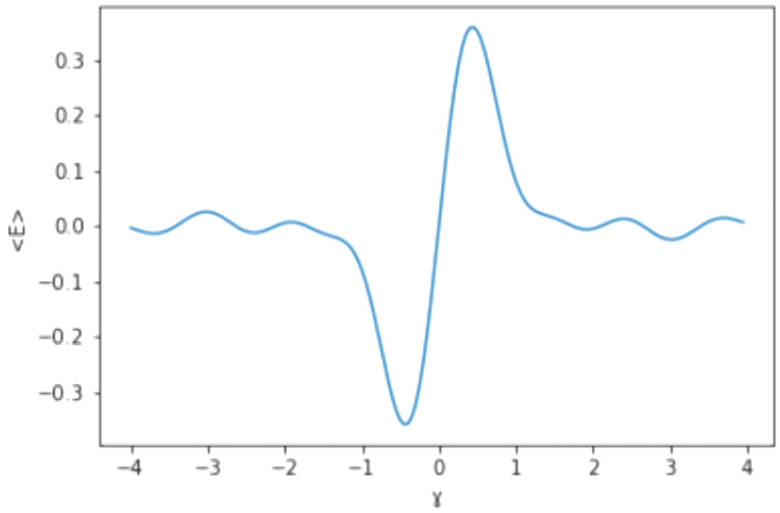
I'll spare you the rest of the details, but just a few lines from here leads to

$$\langle E \rangle = \frac{1}{2} \sin 4\beta \sum_{\langle ij \rangle} J_{ij} \tan(2\gamma J_{ij}) \left[\prod_k \cos(2\gamma J_{ik}) + \prod_k \cos(2\gamma J_{jk}) \right]$$

Beta is completely determined!

Ganna remains a variational parameter.

Solve for gamma, call quantum computer.



Can we do better?

Better objective function?

Why is <E> the thing we are optimizing?

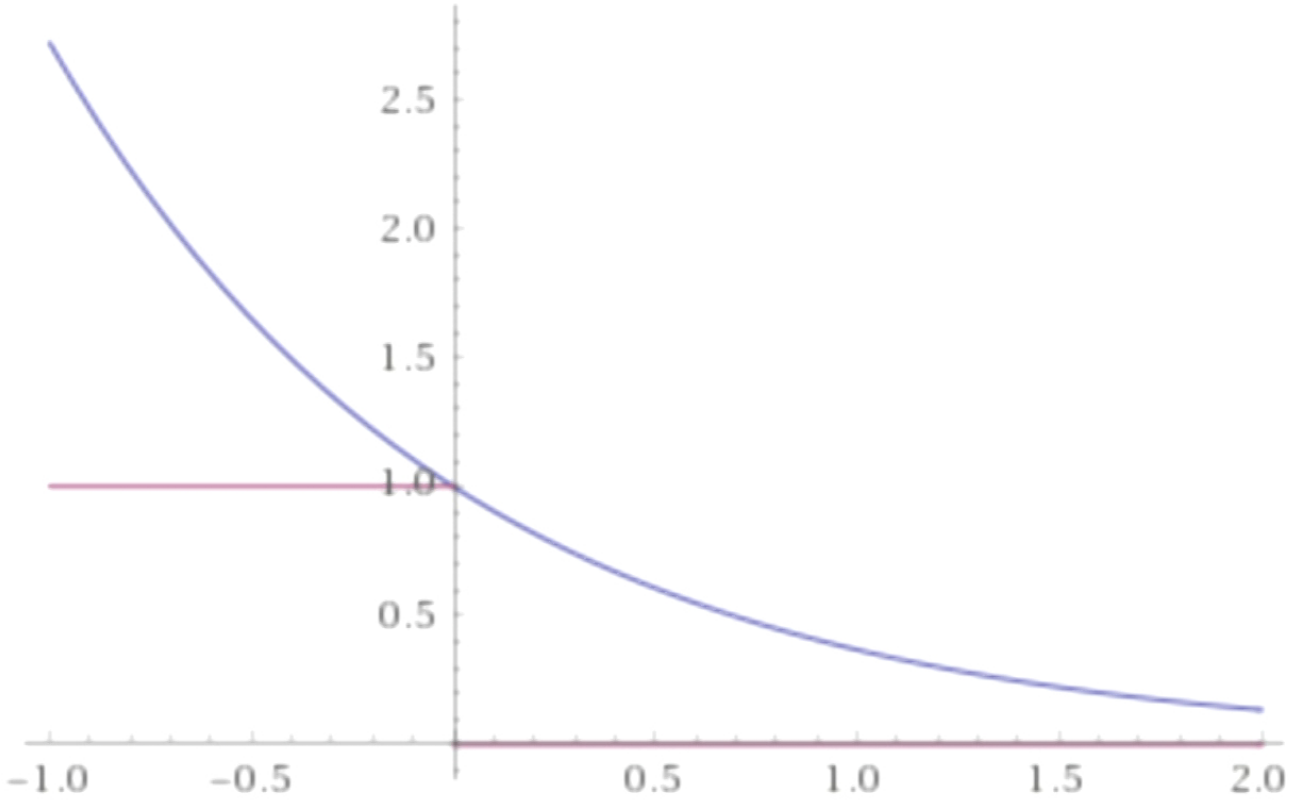
Want to maximize the probability of finding a low-energy state.

We'd rather be optimizing something like $\langle \Theta(E_0 - E) \rangle$.

This may be worth a try! But which E_0 do you pick? Also seems annoyingly discrete.

Better objective function?

Exponential decay is similar to a step function...



Gibbs Objective Function

Introduce a single new hyperparameter.

$$f(\beta, \gamma, \eta) = -\log\langle e^{-\eta E}\rangle$$

Small η: equivalent to <E>

Large η: isolates probability of getting the ground state

η is analogous to an inverse temperature.

Can estimate a good value of η if you know the energy scales of the problem.

A different ansatz?

Can we find a better circuit ansatz?

Enlarge the search space beyond those unitaries connected by the continuous parameters.

One idea: Include discrete variations obtained by adding/removing gates.

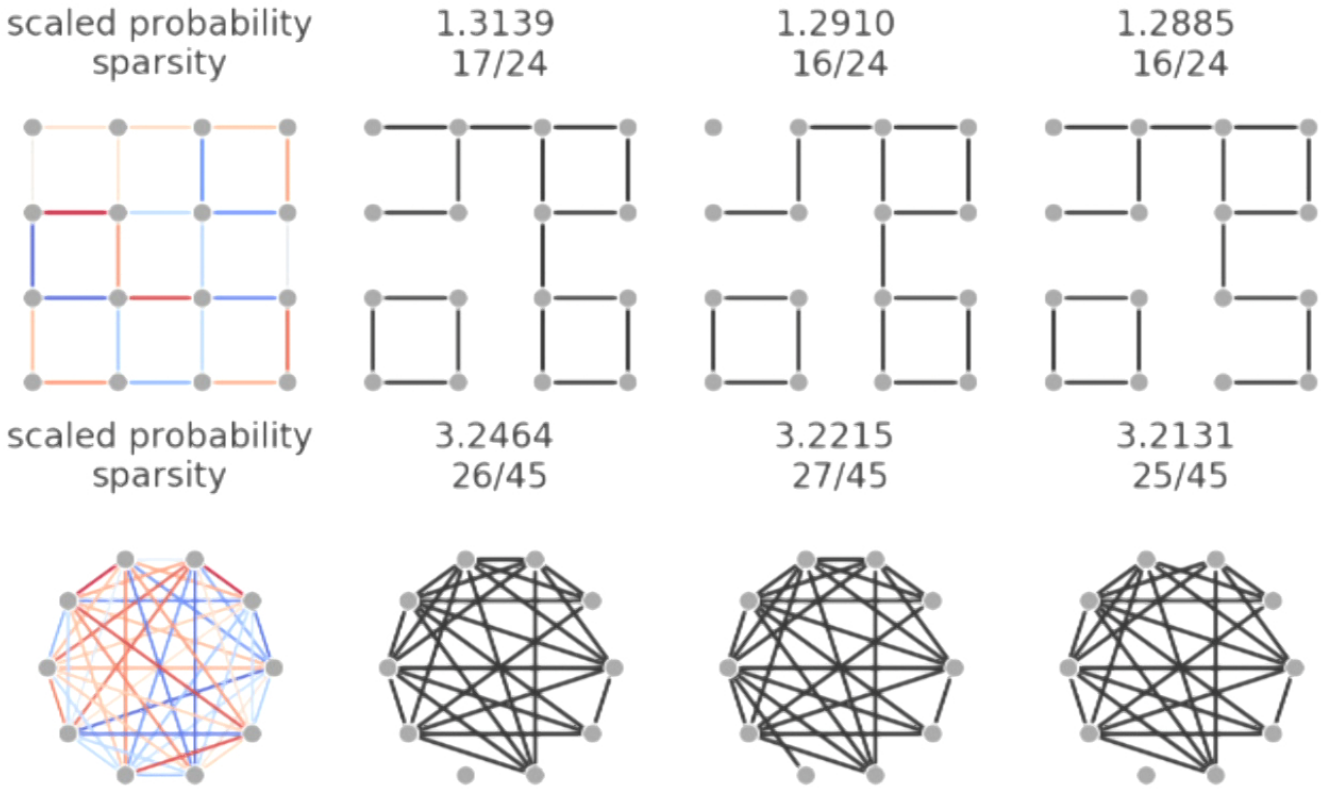
Focus on removing two-qubit gates.

Fewer gates is better for NISQ, but can we don't want to sacrifice on ansatz quality.

A different ansatz?

Removing 30-40% of the two qubit gates leads to improvement most of the time.

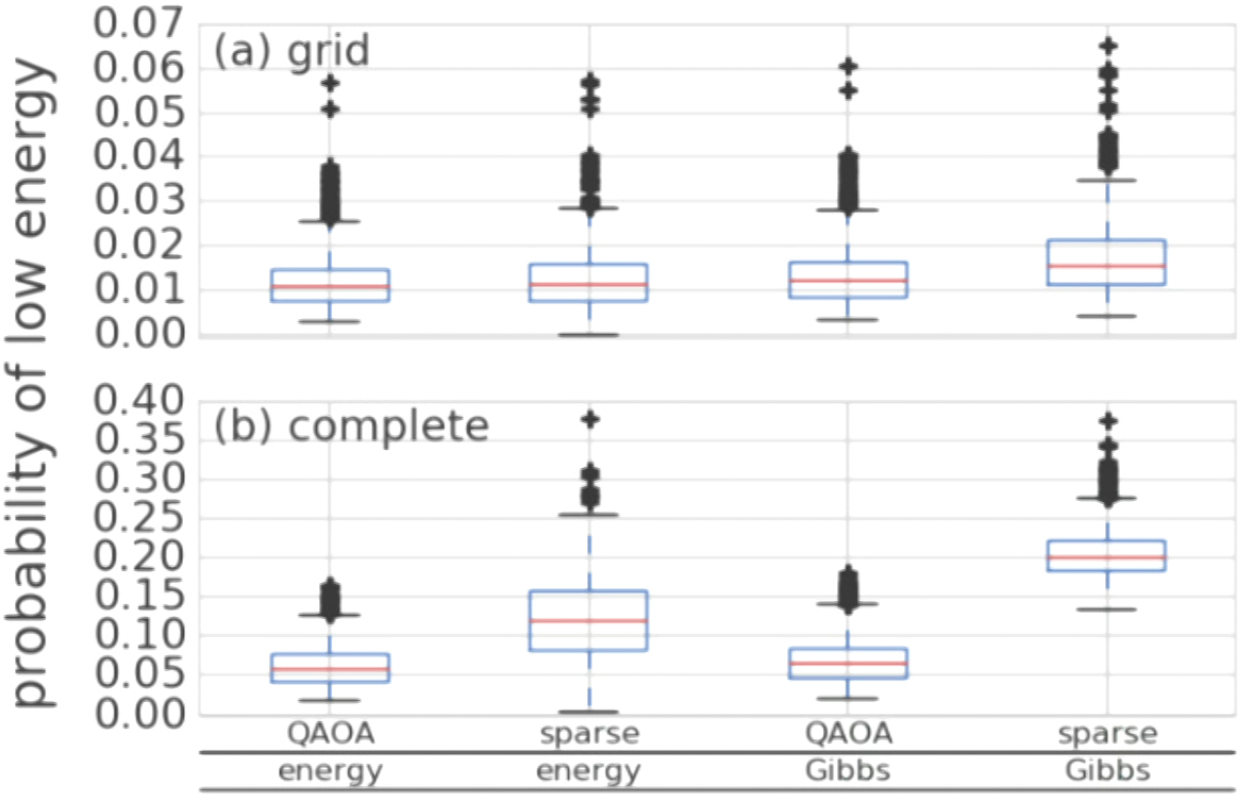
The trick: which 30% do you remove?



A different ansatz?

Gibbs objective function + gate removal are a potent combination.

Works especially well for the fully-connected graph. (Why?)

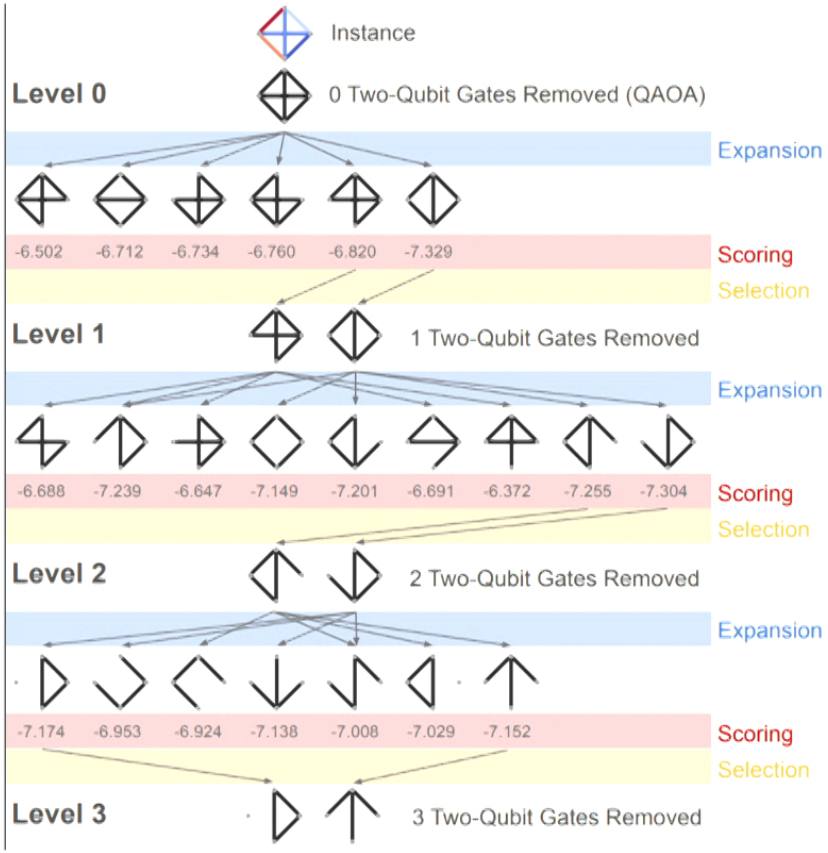


Architecture Search

Beam search to find optimal circuit architectures.

How do we do the scoring at each level?

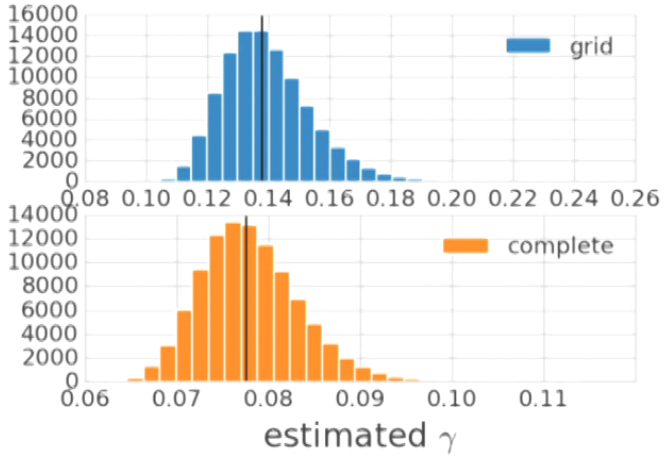
Full simulation (expensive), some sort of heuristic, ML model...

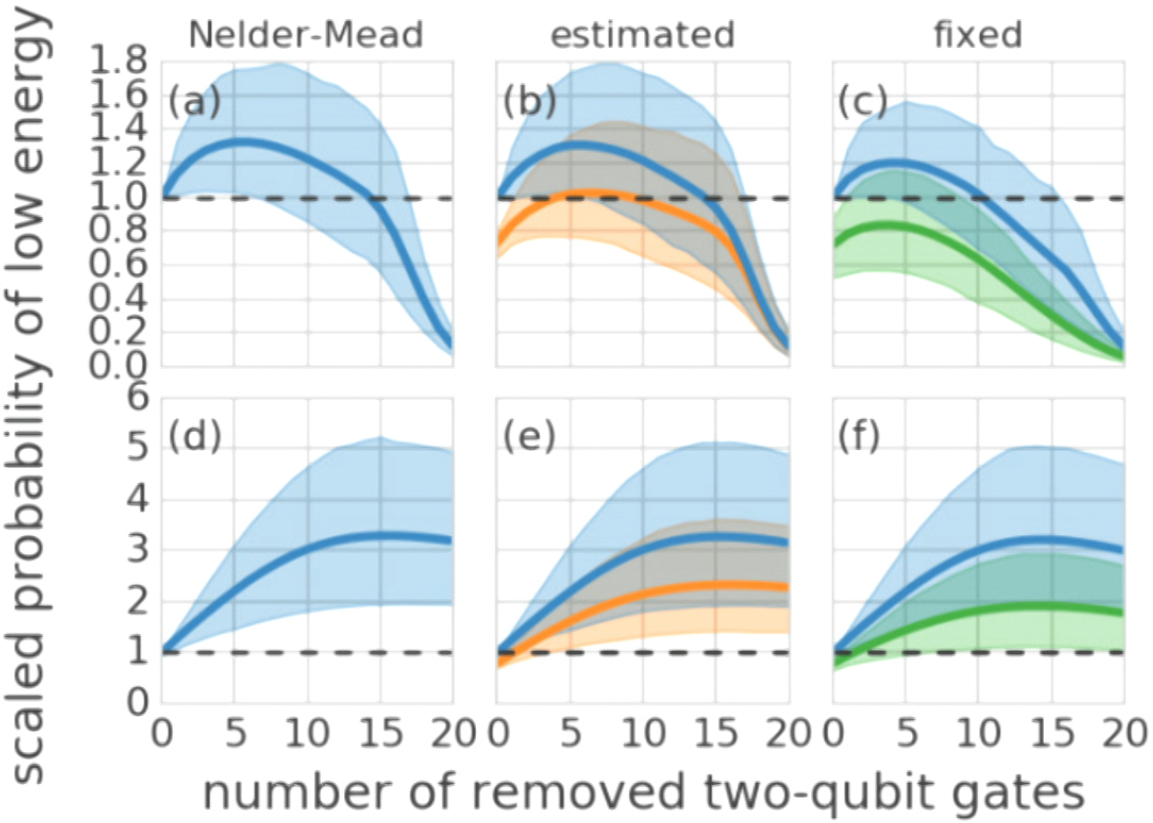


Parameter Estimation

Estimated parameters to save optimization time:

$$\gamma^* = -\sqrt{\frac{\sum_{ij} J_{ij}^2}{6\left(\sum_{ijk,j\neq k} J_{ki}^2 J_{ij}^2 + \frac{1}{3}\sum_{ij} J_{ij}^4\right)}}$$





The Energy Heuristic

What happens to <E> when we remove one of the two-qubit gates in the model? From the derivation, this is equivalent to setting J_{ij} =0 for that edge.

Use $O(\gamma^3)$ analytic approximation for <E> find change in energy after removing an edge:

$$\Delta\langle E\rangle_{ij} \approx -2\gamma J_{ij}^2 \left[1 - 2\gamma^2 \left(\sum_{k\neq i} J_{kj}^2 + \sum_{k\neq j} J_{ki}^2 + \frac{1}{3} J_{ij}^2\right)\right]$$

It's better to remove the gate if its neighbors have large couplings.

ML Guided Search?

It turns out that this is quite hard.

We don’t need to actually simulate the quantum circuit, we only need to be able to rank the circuits in terms of their performance.

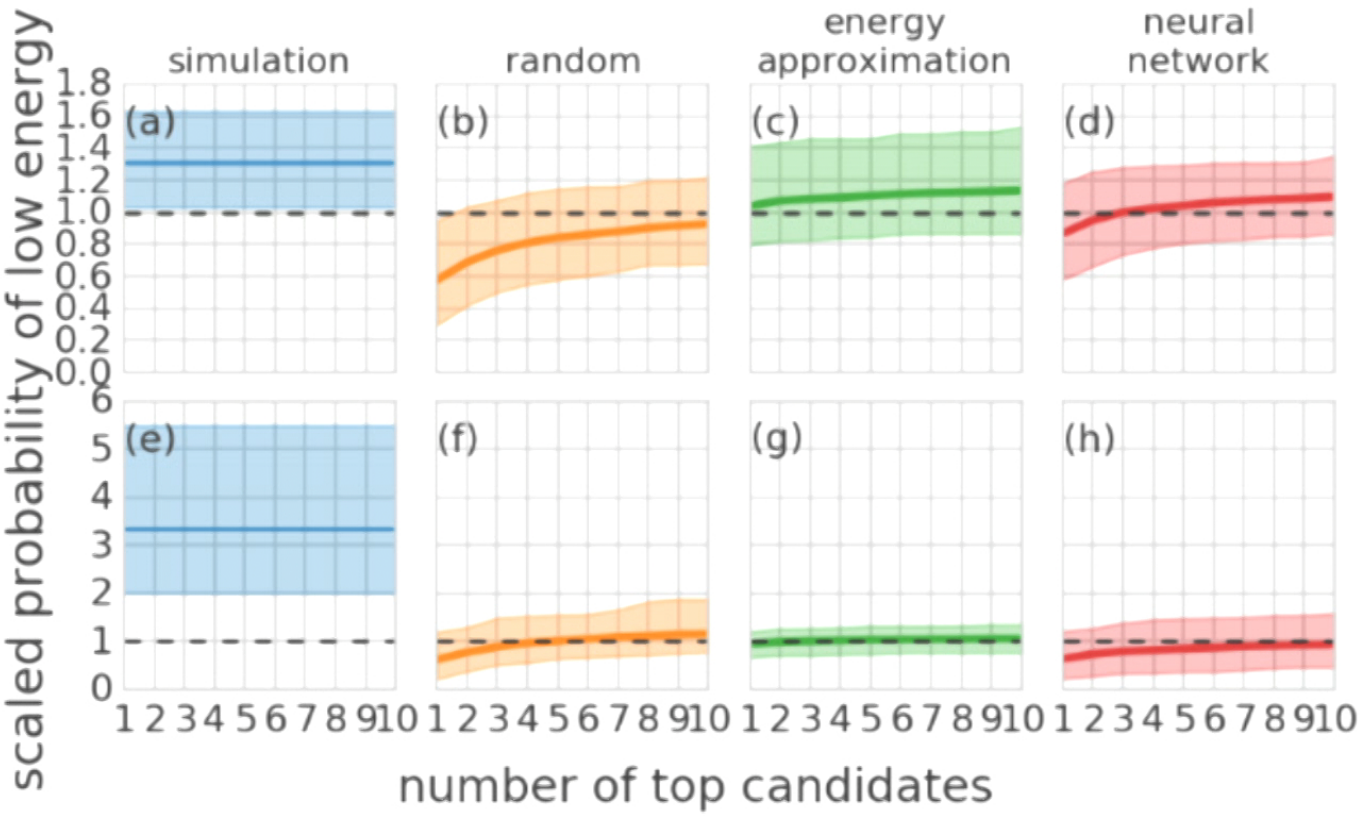
We can beat random chance, but fall far short of the potential gains.

Is it worthwhile?

Results

Greedy simulation vs large beam width (100) for heuristics.

Simulate the top n candidates at the end of the search for comparison.



What next?

There is something interesting happening here, and it could be valuable for NISQ devices.

Can we improve the ML or find a better heuristic?