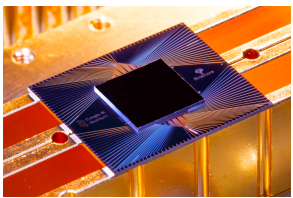
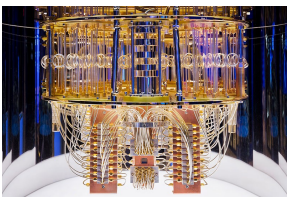
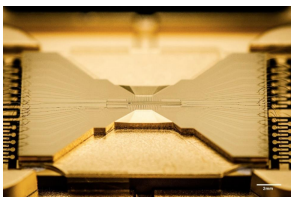
QAOA, VQE, NISQ

Ashley Montanaro

Phasecraft Ltd
School of Mathematics, University of Bristol







University of BRISTOL

EPSRC
Engineering and Physical Sciences Research Council

erc
European Research Council
Established by the European Commission

PHASECRAFT

Advert

Phasecraft is hiring! We're looking for postdoctoral-level quantum software researchers, sponsored PhD students, visiting fellows, and interns.

Visit phasecraft.io for more.

References about variational algorithms:

- Noisy intermediate-scale quantum algorithms, Bharti et al, Rev. Mod. Phys. 94, 015004 (2022)
- Variational quantum algorithms, Cerezo et al, Nature Reviews Physics volume 3, pages 625–644 (2021)

Noisy Intermediate-Scale Quantum computing

Today's quantum computers are:

- Noisy: they experience fairly significant errors, uncorrected by quantum error-correction
- Intermediate-Scale: they are at or beyond the capacity of classical computing to simulate, but not large enough to implement full fault-tolerance (say 50-1000 qubits)

Preskill (2018) coined the acronym NISQ for this era of quantum computing.

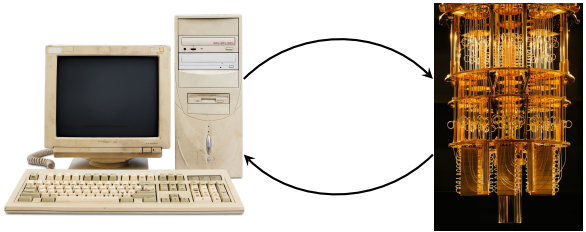
What can we do in the NISQ era?

- With an error rate of 0.005 per 2-qubit gate (as achieved e.g. by Google's quantum hardware), 1000 gates $\implies$ Pr[no error] $= 0.007$
- Breaking RSA-2048: 2.7×10^9 Toffoli gates [Gidney and Ekera '19]
- To bridge this gap we need different applications and error mitigation.

Variational quantum algorithms

One of the most prominent families of quantum algorithms in the NISQ era is variational algorithms.

These algorithms optimise over a family of quantum circuits to solve a problem.



The variational quantum eigensolver [Peruzzo et al '14]

- VQE is an approach to find the ground state of a quantum Hamiltonian H.
- Based on the variational principle of quantum mechanics: for all states $|\psi\rangle$,

$$\langle\psi|H|\psi\rangle \geqslant E_0$$

  where E_0 is the ground energy of H.
- So if we optimise over quantum circuits from some family (variational ansatz) to produce states $|\psi\rangle$, the lowest energy found is an upper bound on E_0.
- We also hope that $|\psi\rangle$ is a good approximation to the ground state itself.
- We use the quantum computer to produce energies $\langle\psi|H|\psi\rangle$ and a classical optimisation loop to optimise over parameters of the quantum circuit.

The variational quantum eigensolver [Peruzzo et al, 2014]

This overall framework leaves a lot of questions open:

- What is a good family of circuits to optimise over?
- How can we efficiently measure the energy of $|\psi\rangle$ with respect to H?
- What is a good method to optimise over the variational ansatz?
- How best to represent H on the quantum computer in the first place?

The art of a variational quantum algorithm designer is answering these questions.

Encoding the Hamiltonian onto a quantum computer

Various families of quantum Hamiltonians are particularly interesting:

- k-local Hamiltonians: $H = \sum_i H_i$, and each H_i acts nontrivially on $\leqslant k$ qubits
  - e.g. Ising model $H = \sum_{\langle i,j \rangle} Z_i Z_j$, Heisenberg model $H = \sum_{\langle i,j \rangle} X_i X_j + Y_i Y_j + Z_i Z_j$
- Fermionic Hamiltonians (e.g. molecules):

$$H = \sum_{i,j} h_{ij} a_i^\dagger a_j + \sum_{i,j,k,l} h_{ijkl} a_i^\dagger a_j a_k^\dagger a_l$$

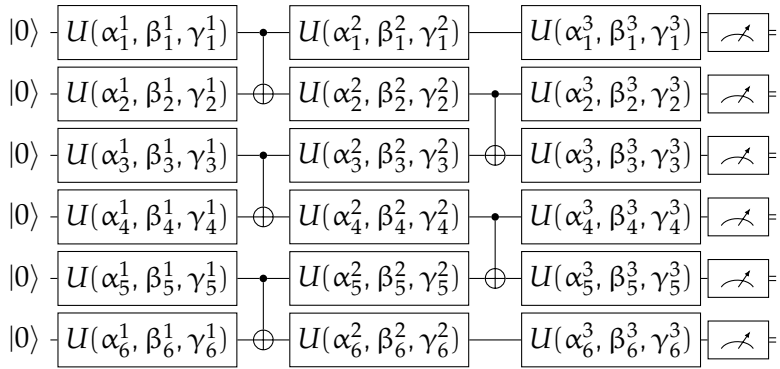
Here $a_i^\dagger$, a_i are fermionic creation and annihilation operators, which can be represented on a quantum computer by (e.g.) the Jordan-Wigner transform:

$$a_i^\dagger \mapsto Z_1 \otimes Z_2 \otimes \cdots \otimes Z_{i-1} \otimes |1\rangle\langle 0|, \quad a_i \mapsto Z_1 \otimes Z_2 \otimes \cdots \otimes Z_{i-1} \otimes |0\rangle\langle 1|$$

Some variational ansätze

A variational ansatz (quantum circuit family) usually consists of a quantum circuit with a certain structure, where some of the gates are parametrised by real numbers.

For example, a hardware-efficient ansatz (e.g. [Kandala et al 2017]):



Some variational ansätze

Some ansätze are targeted at quantum simulation. For example the Hamiltonian Variational ansatz [Wecker et al '15]:

Assume that:

- we want to find the ground state of $H = \sum_j H_j$
- we can write $H = H_A + H_B$
- we have an efficient quantum algorithm for preparing the ground state of H_A.

Then:

1. Prepare the ground state of H_A
2. For each of L layers l, implement $\prod_k e^{it_{lk}H_k}$ for some times $t_{lk} \in \mathbb{R}$

Intuition for this comes from the quantum adiabatic theorem: as $L \to \infty$, this ansatz provably can represent the ground state of H.

Energy measurements

To apply the VQE framework, we need to measure energies $E_\psi = \langle\psi|H|\psi\rangle$.

- If we can write $H = \sum_j H_j$ for some "simple" terms H_j (for example, Pauli matrices) we can measure $E_j := \langle\psi|H_j|\psi\rangle$ for each j and compute the overall energy as a sum
- Computing $E_j \pm \epsilon$ can be achieved using $O(1/\epsilon^2)$ measurements
- Terms that commute can be measured simultaneously, which can reduce the complexity substantially (e.g. diagonal terms can be measured in the computational basis)

Many works have looked at optimising measurement strategies, e.g. [Crawford et al '21, Gokhale et al '19, Bonet-Monroig et al '20]

Classical optimiser

A key ingredient in the VQE loop is the classical numerical optimiser used to minimise the energy. This has to satisfy some desiderata:

- Ability to cope with noisy energy measurements
- (Ideally) no need for gradient information

Many optimisers have been used, e.g.:

- SPSA (Simultaneous Perturbation Stochastic Approximation) [Spall '92]: evaluate the energy in a random line at the current position, and go downhill
- Model Gradient Descent [Sung et al '20]: fit a quadratic model around the current parameters and minimise it (BayesMGD [Stanisic et al '21]: use Bayes' rule to maintain uncertainty in model; analytic descent [Koczor and Benjamin '22]: use the particular form of energies)
- Coordinate Descent (sequential minimal optimisation [Nakanishi et al '20], Rotosolve [Ostaszewski et al '21], Jacobi diagonalisation [Parrish et al '19])

Energies as trigonometric polynomials

- Assume variational gates are of the form $e^{i\theta A}$, $\lambda_j(A) \in \{-M, \ldots, M\}$.
- Then the energy $E(\theta)$ is of the form

$$\langle\psi|e^{-i\theta A}V^\dagger HVe^{i\theta A}|\psi\rangle = \sum_{j,k\in\{-M,\ldots,M\}} e^{i\theta(j-k)}\langle\psi|P_k V^\dagger HVP_j|\psi\rangle = \sum_{l\in\{-2M,\ldots,2M\}} e^{il\theta}w_l$$

  which is a trigonometric polynomial in θ of degree $2M$.
- So it can be determined completely by evaluating it at $4M+1$ points and then (e.g.) minimised directly! (Coordinate descent)
- For the special case of Pauli matrices A:

$$E(\theta) = e^{-2i\theta}w_{-2} + w_0 + e^{2i\theta}w_2, \quad \frac{dE}{d\theta} = -2ie^{-2i\theta}w_{-2} + 2ie^{2i\theta}w_2$$

  so we can, for example, find $\frac{dE}{d\theta}$ by computing $E(\pm\pi/4)$ to learn $w_{\pm 2}$ (the parameter shift rule)

The quantum approximate optimisation algorithm [Farhi et al '14]

We can apply the VQE framework to solve classical optimisation problems by setting

$$H = \sum_{x \in \{0,1\}^n} C(x)|x\rangle\langle x|$$

where $C(x)$ is a cost function. The ground state of H is then the lowest-cost x.

[Farhi et al '14] proposed the following variational method to find the ground state:

1. Start with $|+^n\rangle$
2. Apply $e^{i\gamma H}$
3. Apply $e^{i\beta \sum_j X_j}$

Repeat steps 2 and 3 (with different parameters) p times. Then optimise over the parameters $\beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_p$.

Notes on QAOA

Why should we imagine that QAOA works?

- Intuition from the adiabatic theorem:
  1. We start in the ground state of $H_A := -\sum_j X_j$ and want to get to the ground state of $H_B := H$
  2. If we smoothly change $H_A \to H_B$ slowly enough, we remain in the ground state
  3. Discretising this process gives terms of the form $e^{i\theta H_A}$, $e^{i\theta H_B}$
  4. So if we take large enough p, QAOA can find the ground state of H.

- QAOA essentially encompasses Grover's search algorithm [Jiang et al '17]:
  1. Let H correspond to unstructured search (everything has cost 1 except the marked item, which has cost 0)
  2. The mixer term $e^{i\beta \sum_j X_j}$ is analogous to the Grover diffusion operator

An essentially identical algorithm (without the classical optimisation step) was developed by [Hogg '00]

Applying QAOA to MAX-CUT

A particularly simple example of a combinatorial optimisation problem where QAOA can be applied is MAX-CUT.

- We are given a graph $G = (V, E)$ with n vertices and asked to find a cut (partition of the vertices into sets A and B) that maximises the number of edges across the cut

To map this to a Hamiltonian:

- Let $x \in \{0,1\}^n$ correspond to a cut ($x_i = 1 \Leftrightarrow i \in B$)
- Define

$$C_{ij}(x) = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{if } x_i \neq x_j \end{cases}$$

  and set $C(x) = \sum_{(i,j)\in E} C_{ij}(x)$
- We have $\langle x|Z_iZ_j|x\rangle = 2C_{ij}(x) - 1$. So set

$$H = \sum_{(i,j)\in E} Z_iZ_j$$

Performance of QAOA for MAX-CUT

- Easy to implement – all operations are 2-qubit gates, all terms in H commute
- Provable bounds on the approximation ratio (e.g. $p = 1$, 3-regular graphs: cut ~ 0.69x size of optimal cut)
- Numerical evaluation on random graphs for large enough p outperforms the classical Goemans-Williamson algorithm [Crooks '18]
- ...but for some graphs, choosing $p = O(1)$ cannot outperform Goemans-Williamson [Bravyi et al '19]
- Provably hard to simulate on a classical computer (subject to computational complexity assumptions) [Farhi and Harrow '16]

Some variants of QAOA

- Recursive QAOA (RQAOA) [Bravyi et al '19]:
  1. Run QAOA
  2. Measure correlations $\langle Z_i Z_j \rangle$ in the output state
  3. Find the maximally correlated pair of bits and fix one of them
  4. Recursively solve the resulting problem on $n-1$ bits.

- Warm start QAOA [Egger et al '21]:
  1. Run another algorithm to produce an initial (product) distribution on $x \in \{0,1\}^n$
  2. Use the corresponding distribution as initial state for QAOA instead of $|+\rangle^{\otimes n}$.

Challenges for variational quantum algorithms

- Lack of theoretical justification for performance
- Difficulty of training
  - NP-hard to find optimal variational parameters for a quantum circuit, even for problems which are efficiently solvable classically [Bittel and Kliesch '21]
  - (Actually worse than this: QCMA-hard to find optimal parameters for VQE)
  - VQE iterations on real hardware can be slow (round trip to cloud $\Rightarrow$ each iteration may take seconds)
  - Barren plateaus
- Difficulty of finding a good variational ansatz
- Complexity of actually implementing variational ansatz on hardware

Barren plateaus [McClean et al '18]

For most input parameters, and a sufficiently expressive variational ansatz, function values are exponentially close to their mean:

- If we have a truly random state $|\psi\rangle \in \mathbb{C}^{2^n}$, and $\operatorname{tr} H = 0$, $\operatorname{tr} H^2 = O(m2^n)$ then

$$\mu := \int \langle\psi|H|\psi\rangle d\psi = \operatorname{tr} H \int |\psi\rangle\langle\psi| d\psi = \frac{\operatorname{tr} H}{2^n} = 0$$

and

$$\int (\langle\psi|H|\psi\rangle - \mu)^2 \, d\psi = \int \langle\psi|H|\psi\rangle^2 d\psi = \operatorname{tr} H^{\otimes 2} \int |\psi\rangle\langle\psi|^{\otimes 2} d\psi = \operatorname{tr} H^{\otimes 2} \left(\frac{I + F}{2^n(2^n + 1)} \right)$$

and the latter quantity is $O(\operatorname{tr} H^2/2^{2n}) = O(m2^{-n})$.

Barren plateaus 2 [McClean et al '18]

- Imagine we have a family of states $|\psi\rangle$ which forms an (approximate) 2-design:

$$\mathbb{E}_{|\psi\rangle\sim D}\left[|\psi\rangle\langle\psi|^{\otimes 2}\right] \approx \int |\psi\rangle\langle\psi|^{\otimes 2} d\psi$$

- Then

$$\mathsf{Var}(E) = \mathbb{E}_{|\psi\rangle\sim D}\left(\langle\psi|H|\psi\rangle - \mu\right)^2 = O(m2^{-n}).$$

- So for "most" portions of the parameter space, the energy is exponentially close to 0.
- Many "hardware efficient" ansätze form approximate 2-designs.
- Similar statements can be shown for gradients and other relevant quantities.

Concrete estimates for VQE complexities

How complicated a circuit do we need to solve post-classical problems?

- 5×5 Fermi-Hubbard model: may be able to represent the ground state using a variational ansatz with circuit depth several hundred 2-qubit gates [Cade et al '20]
- Antiferromagnetic Heisenberg model on the Kagome lattice with 50 qubits: 2-qubit gate circuit depth < 200 [Bosse and AM '21, Kattemölle and van Wezel '21]
- One layer of QAOA for the Sherrington-Kirkpatrick model on n vertices (MAX-CUT on the complete graph): 2-qubit gate depth $\sim n$.

Characteristics of the hardware are important, e.g. gate topology, clock speed.

Some implementations on quantum hardware

Problem	Ansatz	Platform	Size	Reference
He-H^+	UCC	Optics	2 qubits	[Peruzzo et al '14]
BeH_2	HEA	IBM	6 qubits	[Kandala et al '18]
MAX-CUT	QAOA	Google	23 qubits	[Harrigan et al '20]
Hydrogen chain	Hartree-Fock	Google	12 qubits	[Arute et al '20]
Fermi-Hubbard	HVA	Google	16 qubits	[Stanisic et al '21]

These different implementations use varying optimisers and achieve varying levels of accuracy. The most complex use 100–200 2-qubit gates.

Many smaller-scale implementations of QAOA have been run.

Error mitigation

In the NISQ world, we don't have access to full fault-tolerance. However, we can improve the performance of our quantum hardware by error mitigation.

Assume that we want to estimate $a := \operatorname{tr} A\rho$, where ρ is the desired state produced by the quantum circuit.

Some techniques that have been developed to improve accuracy of $\widetilde{a} \approx a$:

- Zero-noise extrapolation [Temme et al '17]: compute $\widetilde{a}_1, \ldots, \widetilde{a}_n$ with different noise rates, and extrapolate the results to noise rate 0
- Probabilistic error cancellation [Temme et al '17]: expand an ideal quantum circuit as a linear combination of noisy quantum circuits
- Virtual distillation [Huggins et al '21, Koczor '21]: produce M copies of $\widetilde{\rho}$, and use these to compute $\operatorname{tr}(A\widetilde{\rho}^M)/\operatorname{tr}(\widetilde{\rho}^M)$

Error mitigation (2)

- Error mitigation by training [Czarnik et al '20, AM and Stanisic '21]: use a family of quantum circuits which can be simulated efficiently classically to infer a map $\operatorname{tr} A\widetilde{\rho} \mapsto \operatorname{tr} A\rho$
- Readout noise inversion (e.g. [Kandala et al '17, Maciejewski et al '20]): handle measurement errors by learning the noise map and inverting it
- Dynamical decoupling (e.g. [Pokharel et al '18]): add a sequence of single-qubit gates designed to cancel out errors from interaction between the system and a bath
- Postselection on a symmetry (e.g. [Cade et al '20]): discard runs where a property which should hold is violated
- . . . and more! VQE may also experience a level of inherent noise tolerance.

Summary and future directions

- Variational quantum algorithms are an intriguing way to push the limits of near-term quantum hardware.
- A lot of interesting proposals have been developed, but much more remains to be done

Some interesting future directions:

- More convincing evidence that QAOA etc could outperform classical algs
- Find better theoretical justification for the performance and trainability of variational algorithms
- Develop better classical optimisers
- Evade the known barriers to variational algorithms working
- Design new error mitigation schemes
- Find NISQ algorithms directly and render variational algorithms obsolete!