The Variational Quantum Eigensolver

An optimizer's guide to finding quantum ground states

Adam Godel

September 4, 2026

BU Quantum Seminar

Outline

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

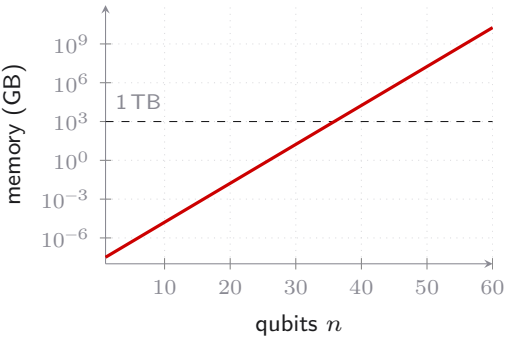
VQE is a derivative-free classical optimizer whose objective function is evaluated by a quantum computer.

What we are computing

$$\hat{H}\,|\psi_k\rangle = E_k\,|\psi_k\rangle\,, \qquad E_0 = \lambda_{\min}(\hat{H}), \qquad \dim \hat{H} = 2^n$$

- $\hat{H}$ is Hermitian: real spectrum, orthonormal eigenbasis
- The physics is present in the smallest eigenvalue, E_0
- **Nobody is diagonalizing this** — at $n = 50$ you cannot store one vector

Why it is hard



Qubits	Amplitudes	Memory
10	10^3	16 KB
30	10^9	16 GB
50	10^{15}	16 PB
100	10^{30}	—

Where ground-state energies show up

Chemistry

Binding energies, reaction barriers, catalysis

Materials

Phase boundaries in lattice models

Combinatorics

3-SAT, MaxCut, portfolio selection

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

Minimize a black box

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^p} f(\boldsymbol{\theta})$$

- No analytic gradient
- Many local minima
- Expensive evaluations
- **Noisy evaluations**

How much do you know about f?

What you can ask for	Methods	Cost / step	Under noise
Gradient + curvature	Newton, BFGS	grad + Hessian	poor
Gradient	SGD, Adam	1 gradient	good
Function values only	**Nelder–Mead, COBYLA, SPSA**	**1–2 evaluations**	**workable**

COBYLA

Constrained Optimization BY Linear Approximations — Powell (1994)

1. Initialize $p + 1$ points, and the value of f at each
2. $p + 1$ points pin down exactly one flat surface — fit it
3. Trust that surface only within a radius ρ of the best point
4. Go to its lowest point inside that radius, evaluate f there
5. Better: accept, drop the worst point. Worse: shrink ρ

COBYLA: two objects

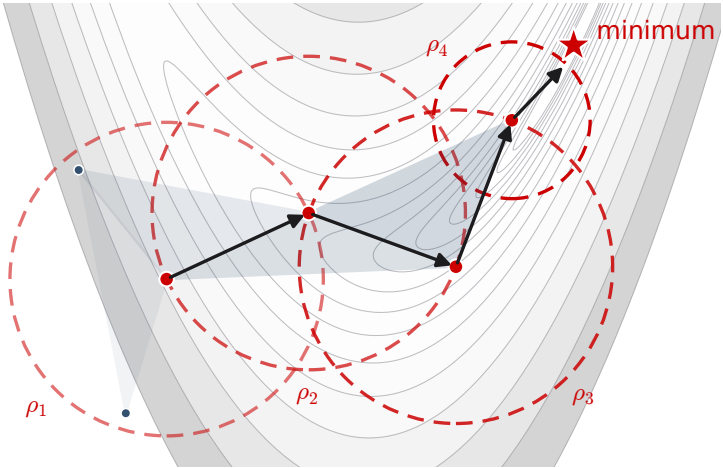
The flat surface through the points

$$L(\boldsymbol{\theta}) = f(\boldsymbol{\theta}_0) + \mathbf{g}^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \quad \text{s.t.} \quad L(\boldsymbol{\theta}_i) = f(\boldsymbol{\theta}_i),\ i = 0, \dots, p$$

Step, but no further than ρ

$$\boldsymbol{\theta}^+ = \arg \min_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \rho} L(\boldsymbol{\theta})$$

The search region walks downhill



■ the $p+1$ points ● the best point, at the center ○ how far we trust the fit, ρ

COBYLA in eight lines

initialize $p+1$ points $\{\theta_0 \dots \theta_p\}$, $\rho \leftarrow \rho_{\text{begin}}$
for each point θ: $y \leftarrow f(\theta)$
repeat
    L $\leftarrow$ flat surface through the $(\theta_i,\ y_i)$
    $\theta^+ \leftarrow$ lowest point of L within ρ of θ_{best}
    $y^+ \leftarrow f(\theta^+)$
    if $y^+ < y_{\text{best}}$: accept, drop the worst point
    else: $\rho \leftarrow \rho / 2$
until $\rho \leq \rho_{\text{end}}$

Where COBYLA fails

Scales poorly in p

A flat surface cannot bend

Degrades under noise

Noisy samples give noisy slopes

No global guarantee

Initialization matters

For contrast: Adaptive moment estimation (Adam)

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla f(\boldsymbol{\theta}_t)$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \nabla f(\boldsymbol{\theta}_t)^{\odot 2}$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \, \hat{\mathbf{m}}_t \oslash \left(\sqrt{\hat{\mathbf{v}}_t} + \delta \right)$$

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

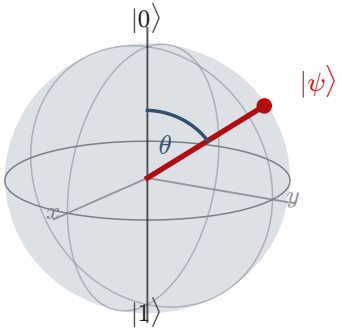
The dictionary

Classical optimization	Quantum counterpart
Parameter vector $\boldsymbol{\theta}$	?
Trial point	?
One function evaluation	?
Objective value $f(\boldsymbol{\theta})$	?
Evaluation noise	?

A bit becomes a direction

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \qquad |\alpha|^2 + |\beta|^2 = 1$$

- A bit is an element of $\{0, 1\}$
- A qubit is a unit vector in $\mathbb{C}^2$
- Can be represented as a unit sphere (Bloch sphere)
- **A continuous surface, not a finite set**



Trial point — the exponential arrives

$$|\Psi\rangle \in (\mathbb{C}^2)^{\otimes n} \cong \mathbb{C}^{2^n}, \qquad |\Psi\rangle = \sum_{x \in \{0,1\}^n} c_x \, |x\rangle$$

- Tensor products, not Cartesian products
- n qubits carry 2^n complex amplitudes
- **This is the space the eigenvalue problem already exists in**

Entanglement is the part that does not factor

$$\underbrace{\frac{1}{\sqrt{2}}\left(\,|00\rangle + |11\rangle\,\right)}_{\text{entangled}} \qquad \text{vs.} \qquad \underbrace{\frac{1}{2}\left(\,|0\rangle + |1\rangle\,\right) \otimes \left(\,|0\rangle + |1\rangle\,\right)}_{\text{product}}$$

- A product state of n qubits needs $2n$ numbers, not 2^n
- Ground states of interacting Hamiltonians are generically entangled
- **Prediction: hardware error will track entanglement**

One function evaluation — programs become products of unitaries

$$|\Psi_{\text{out}}\rangle = U_L \cdots U_2 U_1 |\Psi_{\text{in}}\rangle, \qquad U^\dagger U = \mathbb{1}$$

Reversible

$U^{-1} = U^\dagger$; no branching inside a circuit

Norm-preserving

Probabilities still sum to one, for free

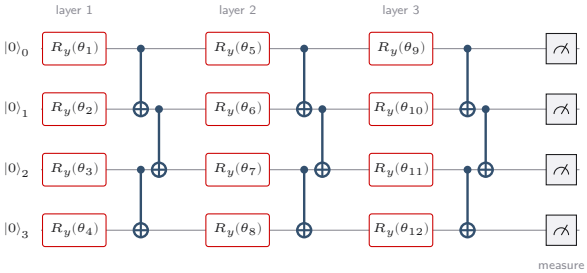
Local and sparse

2×2 or 4×4, tensored with $\mathbb{1}$

Parameter vector θ — where θ actually goes

$$R_y(\theta) = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$

- Smooth, periodic in θ
- p gates $\Rightarrow \boldsymbol{\theta} \in (-\pi, \pi]^p$
- **Rotations are parameterized**



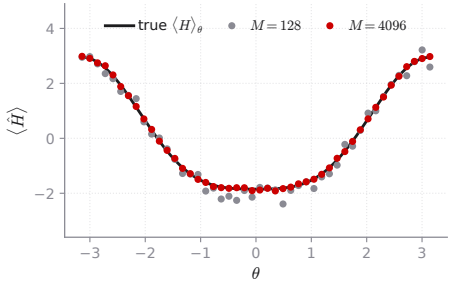
One function evaluation — you cannot read the state

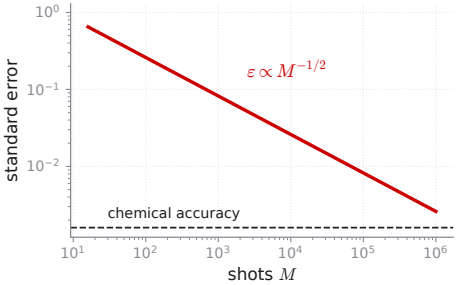
$$\Pr[\,\text{outcome } x\,] = \left|\,\langle x|\Psi\rangle\,\right|^2$$

- One run returns *one* bitstring
- The state is destroyed — a second sample requires a second run
- Only the squared magnitudes of amplitudes are observed

Objective value $f(\theta)$ — the result each iteration

$$\langle \hat{H} \rangle_{\boldsymbol{\theta}} \approx \frac{1}{M} \sum_{s=1}^{M} h_s, \qquad \text{s.e.} = \frac{\sigma}{\sqrt{M}}$$





Which averages you are allowed to take

$$\hat{H} = \sum_{k=1}^{K} c_k P_k, \qquad P_k \in \{\mathbb{1}, X, Y, Z\}^{\otimes n}, \quad c_k \in \mathbb{R}$$

- Pauli strings are an orthogonal basis for Hermitian operators
- Commuting P_k share one measurement setting — G groups, $G \ll K$
- **Hubbard: 3 groups. 3-SAT: 1 group (diagonal).**

The filled dictionary

Classical optimization	**Quantum counterpart**
Parameter vector $\boldsymbol{\theta}$	rotation angles inside a circuit
Trial point	$\lvert\Psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,\lvert\mathbf{0}\rangle$
One function evaluation	prepare, measure, average over M shots
Objective value $f(\boldsymbol{\theta})$	$\langle\widehat{H}\rangle_{\boldsymbol{\theta}}$
Evaluation noise	shot noise + hardware noise

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

The variational principle

$$E_0 \leq \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad \text{for } \textit{every } |\Psi\rangle$$

An upper bound

You can never overshoot

Lower is better

Optimize the *states* to get to E_0

Noise breaks it

The bound holds for $\langle \hat{H} \rangle$, not your estimate of it

One-line proof

$$\langle\Psi|\,\hat{H}\,|\Psi\rangle \;=\; \sum_k E_k \left|\,\langle\psi_k|\Psi\rangle\,\right|^2 \;\geq\; E_0 \sum_k \left|\,\langle\psi_k|\Psi\rangle\,\right|^2 = E_0\,\langle\Psi|\Psi\rangle$$

Restrict, then minimize

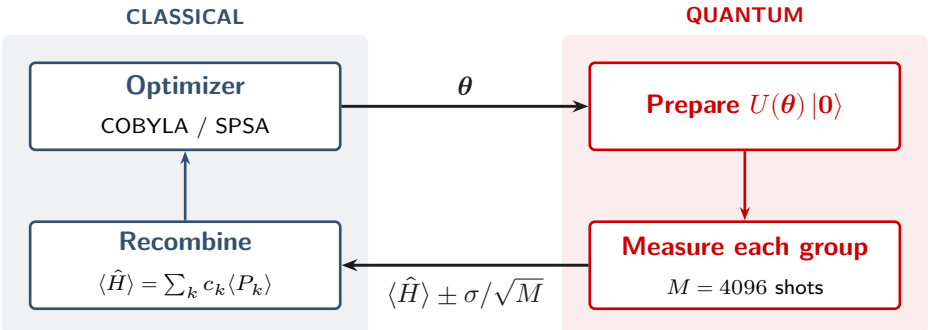
Trial state

$$|\Psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\mathbf{0}\rangle\,, \qquad |\mathbf{0}\rangle = |0\rangle^{\otimes N}\,, \qquad \boldsymbol{\theta} \in (-\pi, \pi]^p$$

The objective

$$E_{\mathrm{VQE}} = \min_{\boldsymbol{\theta}} \; \langle \mathbf{0}|\, U^\dagger(\boldsymbol{\theta})\, \hat{H}\, U(\boldsymbol{\theta})\, |\mathbf{0}\rangle \;\geq\; E_0$$

The loop


CLASSICAL
QUANTUM
Optimizer
COBYLA / SPSA
θ
Prepare $U(\theta)\,|0\rangle$
Recombine
$\langle \hat{H} \rangle = \sum_k c_k \langle P_k \rangle$
$\langle \hat{H} \rangle \pm \sigma/\sqrt{M}$
Measure each group
$M = 4096$ shots

Four things you choose

1 Qubit mapping
Sets qubit count and Pauli weight

2 Ansatz $U(\boldsymbol{\theta})$
Sets what states are reachable

3 Measurement
Sets the variance per evaluation

4 Optimizer
Sets the number of evaluations

Choice 1 — the mapper

Fermi–Hubbard: Jordan–Wigner

$$a_j = \Big(\prod_{k<j} Z_k\Big)\frac{X_j + iY_j}{2}, \qquad \hat{n}_j = a_j^\dagger a_j = \frac{\mathbb{1} - Z_j}{2}$$

$$a_i^\dagger a_j + \text{h.c.} \;\longmapsto\; \tfrac{1}{2}(X_i Z_{i+1}\cdots Z_{j-1} X_j + Y_i Z_{i+1}\cdots Z_{j-1} Y_j)$$

$$\hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \;\longmapsto\; \tfrac{1}{4}(\mathbb{1} - Z_{i\uparrow})(\mathbb{1} - Z_{i\downarrow})$$

Random 3-SAT: one variable, one qubit

$$x_j = \frac{\mathbb{1} - Z_j}{2}, \qquad \hat{H} = \sum_{c=1}^{m} \prod_{j\in c} \frac{\mathbb{1} + s_{cj} Z_j}{2}, \qquad s_{cj} = \pm 1$$

Choice 2 — the ansatz

$$U_{\mathrm{EP}}(\theta,\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} & 0 \\ 0 & -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix}$$

Hardware-efficient

Expressive, but structurally ignorant of your problem

Symmetry-preserving

Fixes particle number, so the search stays in the N_{occ} sector

Choice 3 — what one evaluation costs

	Fermi–Hubbard	**Random 3-SAT**
Qubits	8	4
Commuting Pauli groups G	3	1 (diagonal)
Optimizer	SPSA $\times$ 5	COBYLA $\times$ 50
Circuits	135 / 272	200 / 402
Shots	5.5×10^5 / 1.15×10^6	8.2×10^5 / 1.68×10^6
QPU time	**0.62 s / 2.12 s**	**0.62 s / 1.98 s**

Choice 4 — the classical optimizer returns

Optimizer	Circuits / step	Under shot noise
COBYLA	G	moderate
SPSA	$2G$	good
Adam	$2pG$	good, if affordable
BFGS / Newton	gradient + curvature	unusable

VQE in a nutshell

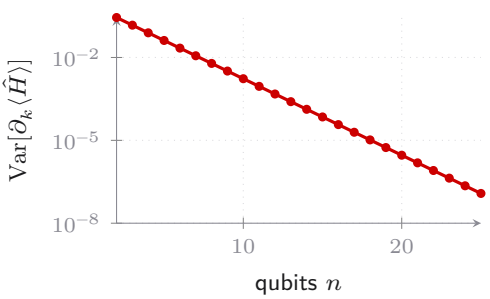
1. encode H, split into Pauli groups
2. choose ansatz U(θ), initial θ_0
3. repeat:
   - QPU: prepare U(θ)$|0\rangle$, measure $\rightarrow \langle \mathrm{H} \rangle$
   - CPU: optimizer proposes θ^+
4. until converged
5. report $\mathrm{E}_{\mathrm{VQE}} \geq \mathrm{E}_0$

1. The problem
2. A classical optimizer
3. Into quantum
4. The algorithm
5. Reality

Problem 1 — barren plateaus

$$\mathrm{Var}\left[\partial_k \langle \hat{H} \rangle\right] \in \mathcal{O}(2^{-n})$$

- Flat almost everywhere, minimum in a crevasse
- Gradient-free methods fare no better
- **Exponentially many shots to get out of a local minimum**



Problem 2 — the measurement wall

$$M = \mathcal{O}\left(\varepsilon^{-2}\right)$$

$100\times$
shots to buy $10\times$ more precision

$\times\, G$
G commuting groups, one circuit each

$\times\, T$
T steps before the optimizer stops

Problem 3 — the device is not your circuit

Decoherence	Gate error	Readout error
T_1, T_2 set a wall-clock budget	Compounds with circuit depth	The measurement itself misfires

Problem 4 — ansatz bias

$$\min_{\text{states the ansatz can reach}} \neq \min_{\text{all possible states}}$$

0.011
DMRG mean absolute error

0.90
VQE, 5-layer hardware-efficient

Problem 5 — the competition is really good

- Example: DMRG minimizes $\langle \hat{H} \rangle$ over matrix product states
- Near-exact in 1D and quasi-1D, where entanglement obeys an area law
- Strains in higher-dimensional regimes, but other tensor network methods fare better
- **It runs on a laptop**

Error mitigation — not error correction

M3 readout

Invert the measurement confusion matrix

DD decoupling

Pulse trains on otherwise idle qubits

ZNE extrapolation

Run at noise $\times\{1, 3\}$, extrapolate to $\times 0$

The experiments — IQM Garnet

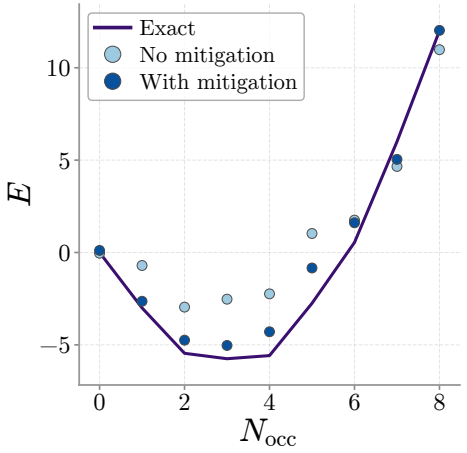
20
qubits, square lattice

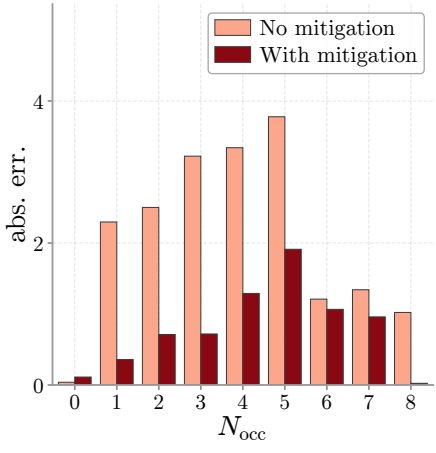
4,096
shots per circuit

322 ns
readout duration

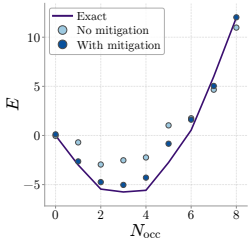
	Fermi–Hubbard	**Random 3-SAT**
Width	8 qubits, 2×1 honeycomb	4 qubits, $n = 4$ variables
Swept	$N_{\text{occ}} = 0 \ldots 8$; $t = 1$, $U = 3$	$\alpha \in \{2, 3, 4, 5\}$
Ansatz	excitation-preserving	real-amplitudes
Reference	exact diagonalization	exact diagonalization

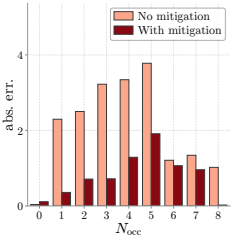
Result 1 — Fermi–Hubbard on hardware





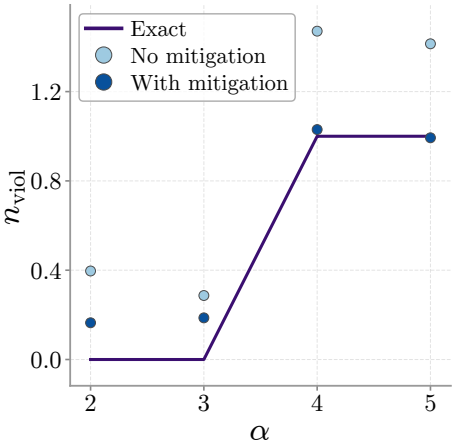
Reading result 1

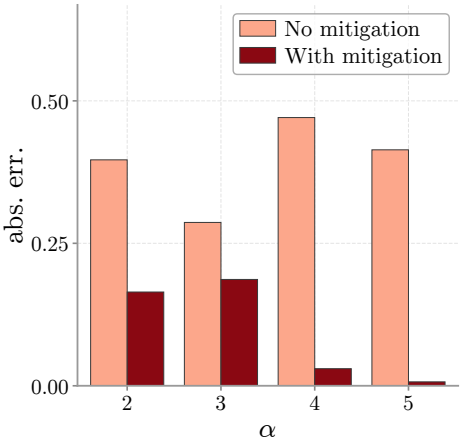




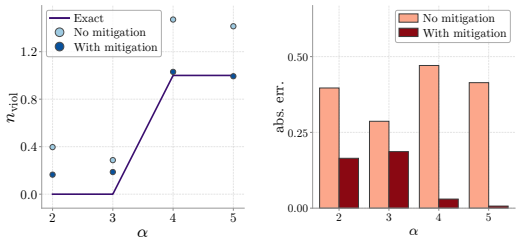
- Shape recovered — minimum near half filling
- $2.08 \to 0.79$ mean absolute error, $2.6\times$
- Exact at the band edges: a product state the ansatz gets for free
- Worst over $N_{\mathrm{occ}} = 1$–5, where the state is entangled

Result 2 — random 3-SAT on hardware





Reading result 2



- The exact curve is a step: satisfiable at $\alpha = 2, 3$
- $0.39 \to 0.10$ clauses, $4.0\times$
- Shallower: 4 qubits, one commuting group, no basis change
- Rounding recovers the exact integer at *every* α

Takeaways

1. VQE can be thought of as an optimizer with an exotic oracle
2. The variational principle is what makes that oracle usable
3. Everything hard is in the oracle, not the optimizer
4. **Shallow, structured, discrete problems can run on today's hardware**

References

- Peruzzo *et al.*, *Nat. Commun.* **5**, 4213 (2014) — VQE
- Tilly *et al.*, *Phys. Rep.* **986** (2022) — review
- Powell, *Adv. Optim. Numer. Anal.* (1994) — COBYLA
- Spall, *IEEE TAC* **37** (1992) — SPSA
- Kingma & Ba, *ICLR* (2015) — Adam
- McClean *et al.*, *Nat. Commun.* **9**, 4812 (2018) — barren plateaus
- Jordan & Wigner, *Z. Physik* **47** (1928)
- White, *PRL* **69**, 2863 (1992) — DMRG
- Sarovar *et al.*, HamLib (2024)
- Godel *et al.*, QuantumBenchPhase (2026)
- IQM Quantum Computers

Questions?