Google Cloud

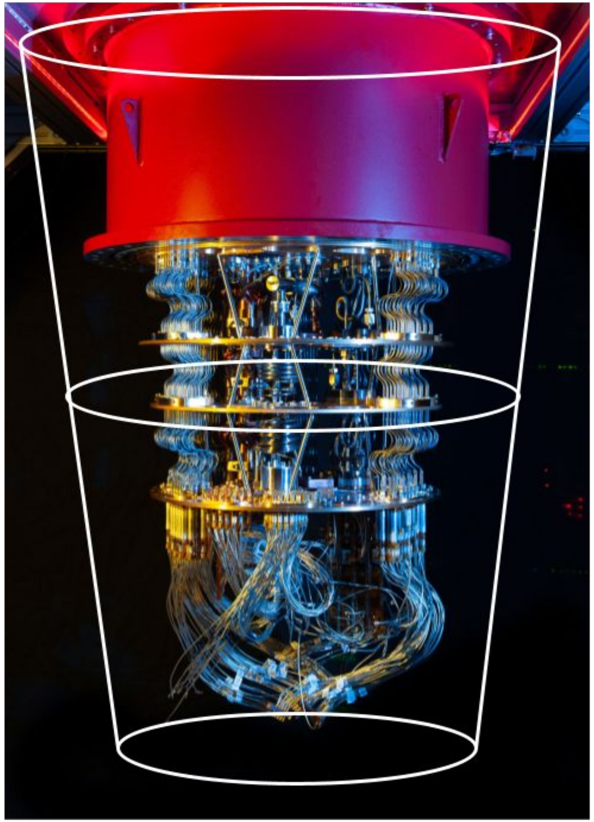
Architectural Implications of Hybrid Quantum/Classical Computing

Kevin D. Kissell, Google

Salishan Conference on High Speed Computing 2022

The Quantum Glass is Half Full

- The theories of quantum computing have consistently been proven correct
- Real quantum processors are being built, and real quantum parallelism is being demonstrated
- The body of quantum algorithmic work is rapidly growing and maturing
- Engineering problems have consistently proved more challenging than expected
- Quantum advantage from quantum speedup requires large problem scales to get traction
- Major technical hurdles remain to scale systems

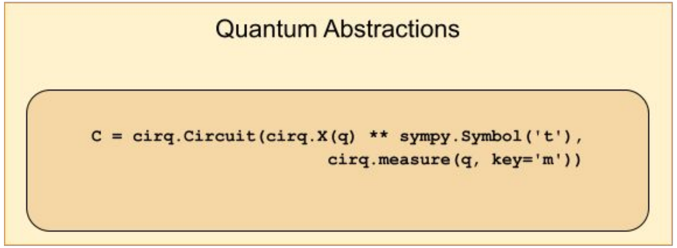


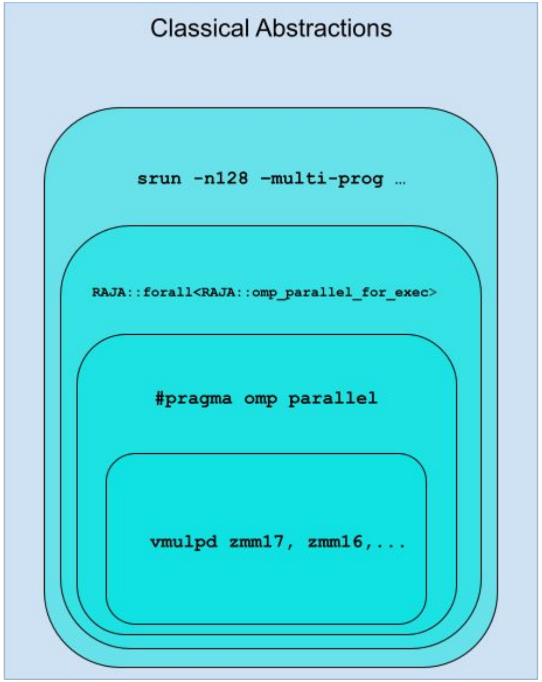
Immaturity of Abstraction, Immaturity of Architecture

Classical parallelism is abstracted at different levels of hardware and software that allow evolution at different levels at different rates

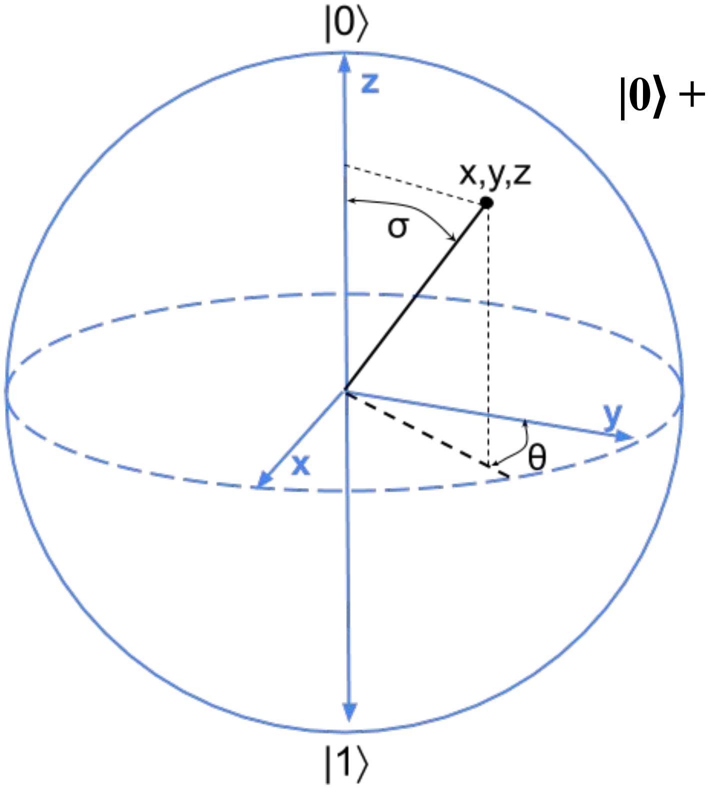
Quantum is so new that application writers are still manipulating individual bits

Laboratory implementation is architecture

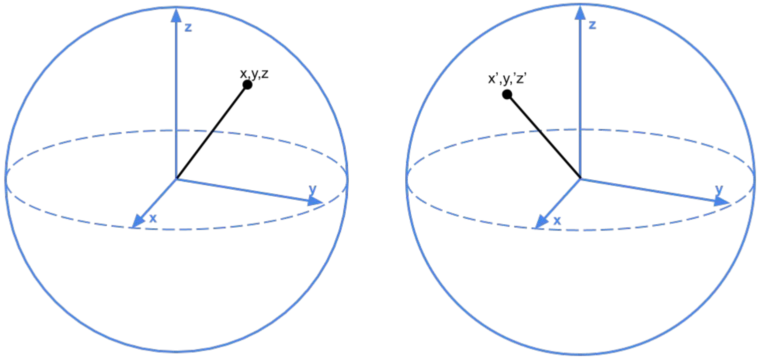




Quantum Bits, Superposed States



$$|0\rangle + |1\rangle$$

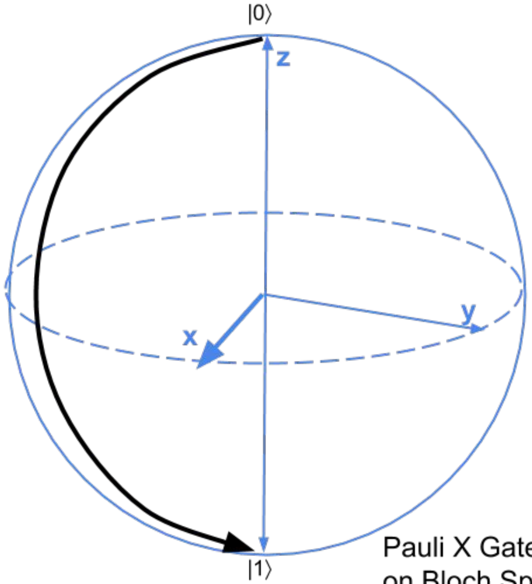


$$(\,|0\rangle+|1\rangle\,)^2$$
$$=\ |00\rangle+|01\rangle+|10\rangle+|11\rangle$$

For N entangled qubits, 2^N possible superposed states

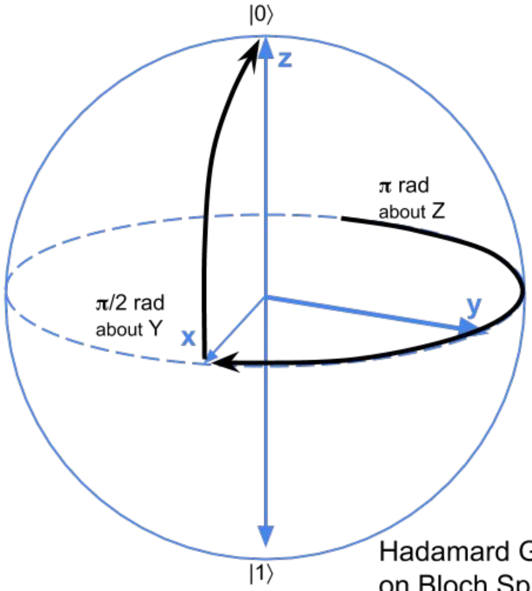
Proprietary + Confidential

Visualizing Quantum "Gates"



Pauli X Gate Operation on Bloch Sphere

$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} =$$



Hadamard Gate Operation on Bloch Sphere

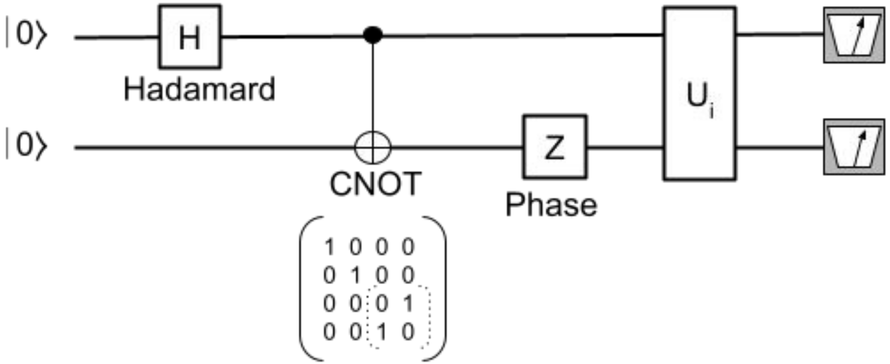
$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} =$$

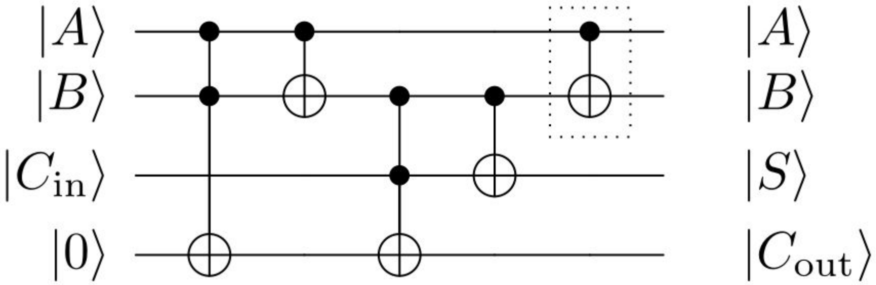
Proprietary + Confidential

Predicated Operations and Entanglement

Gate model programs have no "branches", only conditional operations that look a bit like predicated execution.

The Toffoli "conditional-conditional-not" gate is a universal gate, like the boolean NAND





Toffoli Gate

(Feynman, 1986)

"A man's reach should exceed his grasp. Or what's a heaven for?"

Photosynthesis fixes atmospheric nitrogen at 25 C and normal atmospheric pressure.

1-2% of all human energy consumption into doing it with the Haber process, which needs 400 C and 200 atm.

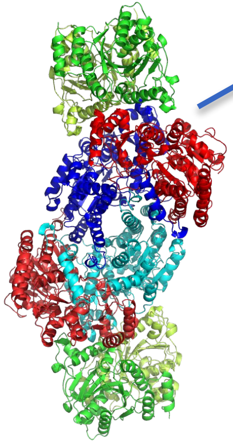
Nature's secret is the enzyme catalyst nitrogenase.

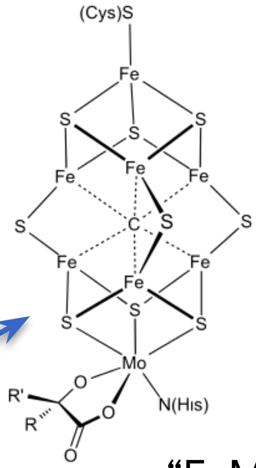
Electronic structure and substrate attachment almost totally unknown and beyond all current classical computational methods.

Should be solvable with ~200 **stable** logical qubits.

But O(100) stable qubits are still years away...

$$N_2 + 3\,H_2 \rightarrow 2\,NH_3$$





"FeMoco"

Early Hybrid: “Hardware Efficient” Generic Ansatz for Quantum Machine Learning

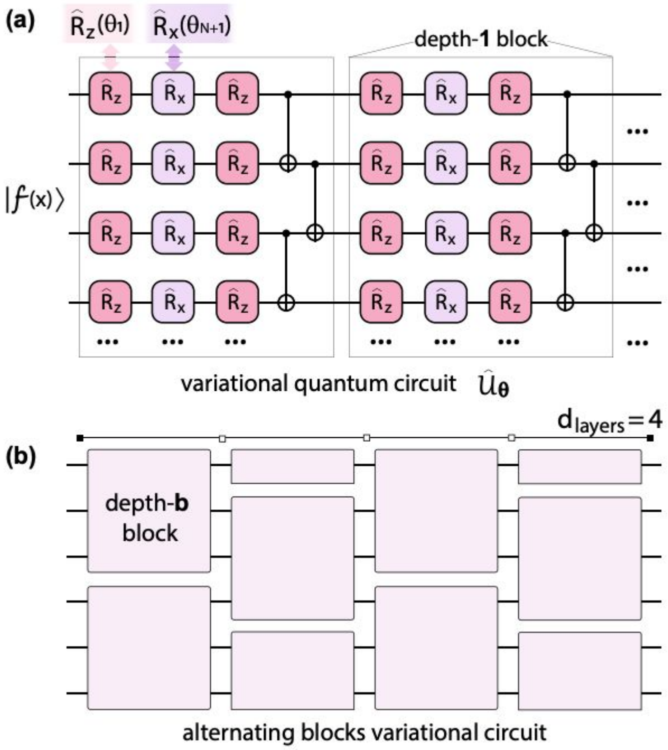
Instead of building exact quantum models of a thing, “train” a quantum circuit to approximate it.

Very much analogous to machine learning, where instead of adjusting weights, one adjusts the operational parameters of quantum gates.

In theory, a desired quantum distribution can be arrived at in many different ways.

“Ansatz” fairly arbitrary, simplistic

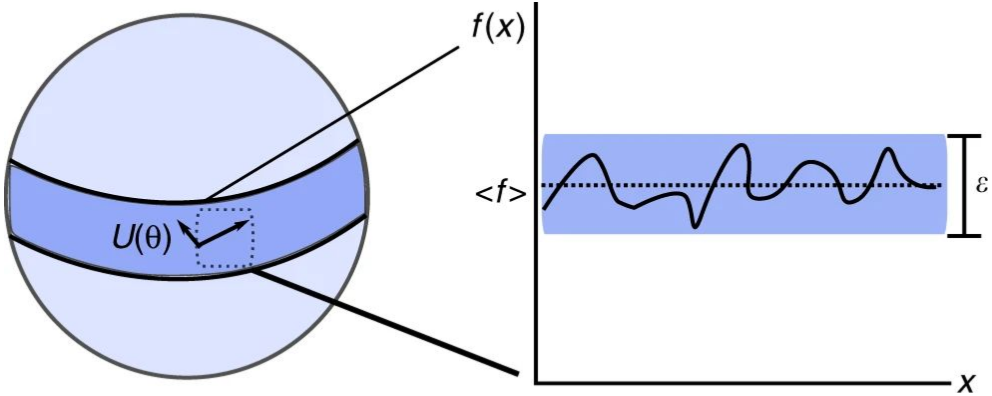
Good early results with small problems, but none of the legitimately interesting problems are small.



https://arxiv.org/pdf/2011.10395.pdf

The Pitfall of the Barren Plateau

The sphere depicts the phenomenon of concentration of measure in quantum space: the fraction of states that fall outside a fixed angular distance from zero along any coordinate decreases exponentially in the number of qubits.. This implies a flat plateau where observables concentrate on their average over Hilbert space and the gradient is exponentially small. The fact that only an exponentially small fraction of states fall outside of this band means that searches resembling random walks will have an exponentially small probability of exiting this "barren plateau"



https://arxiv.org/pdf/1803.11173.pdf

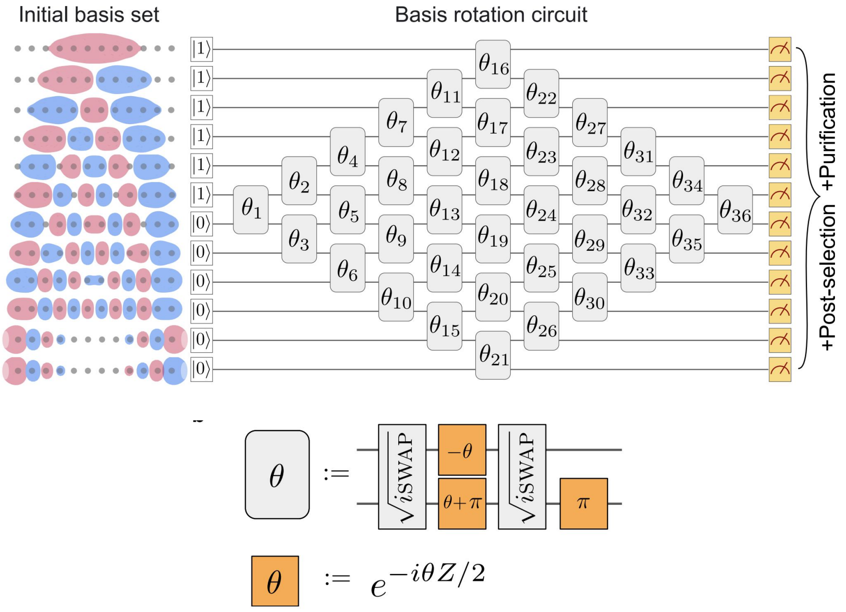
“Fit for Purpose” Ansatz for Molecular Ground States

Initial ansatz designed for molecular energy minimization.

Qubit state represents electron orbital occupation.

The problem can be simplified to terms of quantum basis rotation, so the building block is a 2-input unitary that performs a parametric basis rotation.

Successfully tested on chained hydrogen molecules of length 6, 8, 10, and 12.



https://arxiv.org/pdf/2004.04174.pdf

Getting Close Enough to See Cool Stuff

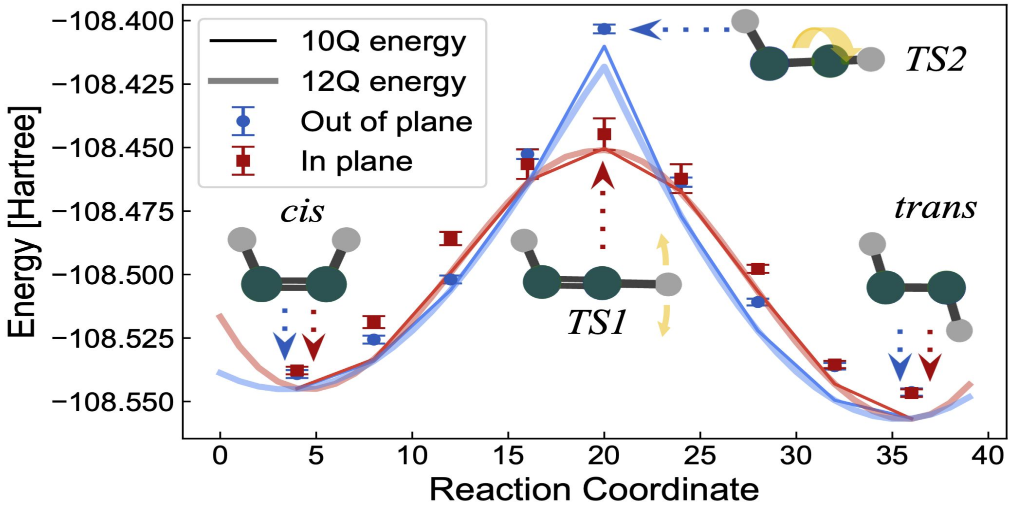
Diazene model run on 10 and 12 qubits

Simplistic Hartree-Fock model used

Roughly 40 gates deep (close to current limits)

Each circuit variation run 250,000 times to get clear sampling profile from noisy machine

Clear identification of in-plane and out-of-plane reaction paths made to chemical accuracy



https://arxiv.org/pdf/2004.04174.pdf

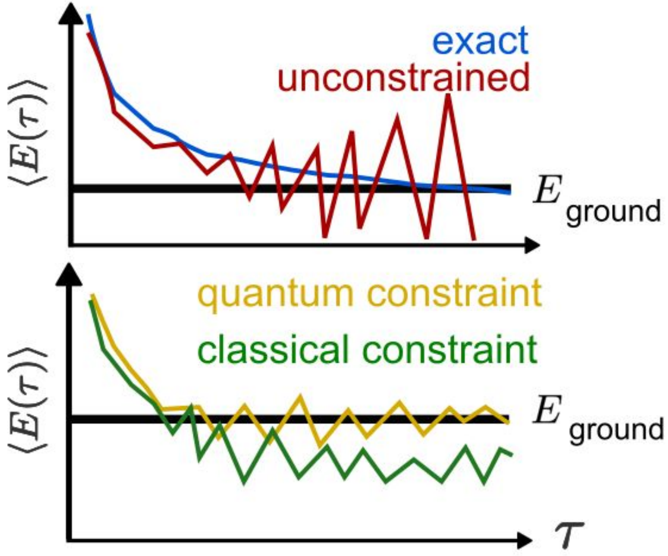
Turning the Tables - Quantum Driven Classical Monte Carlo

Projector Monte Carlo is an efficient technique for classical modeling of quantum behavior, but use with fermions (e.g. electrons) has a sign problem.

One common solution uses a “trial function” that approximates the ground state well enough to serve as a constraint that catches and discards bad results.

Classical trial functions are limited in the models they can approximate, and are often very inexact

Quantum functions can potentially approximate a larger class, and do a better job.

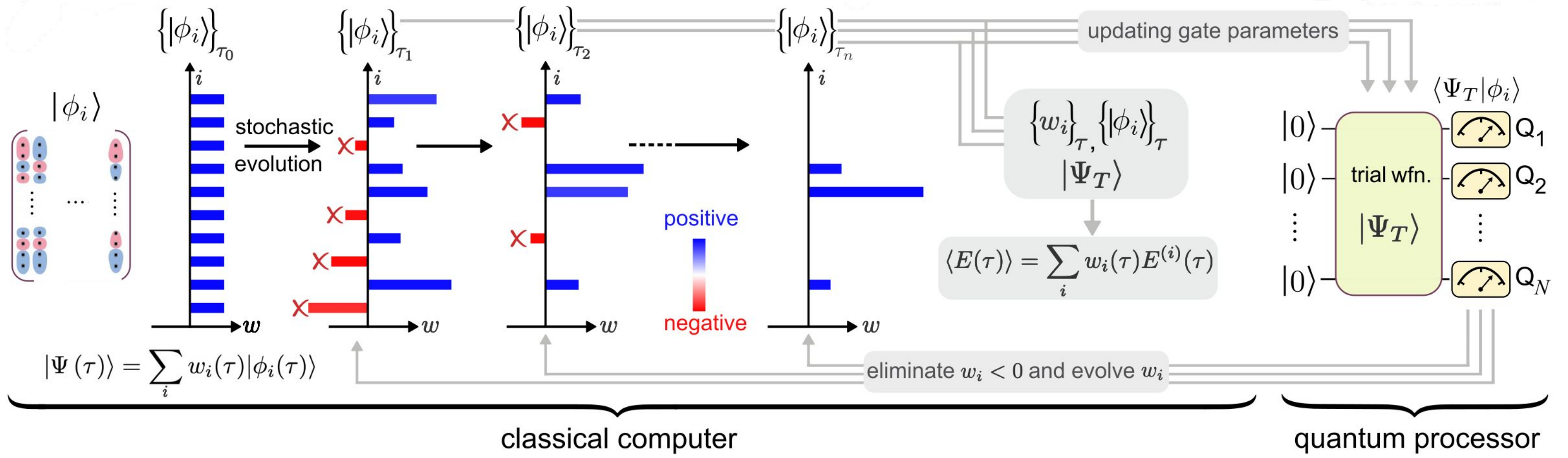


https://arxiv.org/pdf/2106.16235.pdf

The Classical/Quantum Hybrid Loop

Classical QMC with Quantum Computed Trial Function

The quantum computer estimates energies and checks for acceptable prediction “overlap”



https://arxiv.org/pdf/2106.16235.pdf

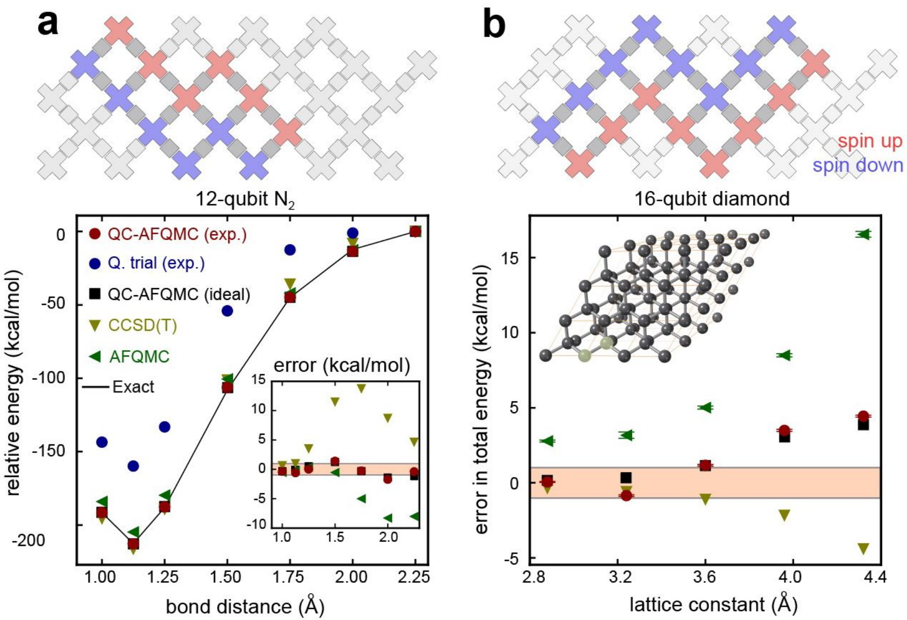
Diamond in the Rough

The basic, untuned Hartree-Fock quantum trial function is far from exact.

The quantum-corrected AFQMC code, however, is more accurate than the classical AFQMC and CCSD(T) models, particularly at high energies.

The diamond model at 16 qubits is the largest successful quantum computation of molecular dynamics so far.

Even with a reduced number of qubits, restricting the number of orbitals modeled, the trial function proved almost as effective.



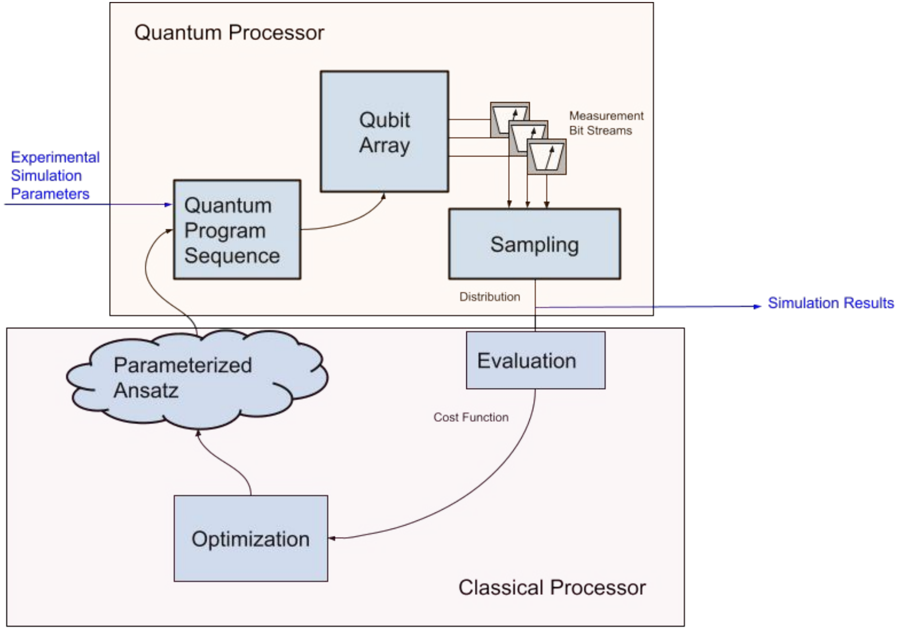
https://arxiv.org/pdf/2106.16235.pdf

Mapping the Classical/Quantum Frontier

Current experiments use laboratory prototype quantum processors, loosely coupled to classical compute servers.

The actual bandwidth required between the domains is modest. The output of 10s to 100s of qubit measurements is 10s to 100s of bits. The parameterization of 100s of gates is 1000s of bits.

For variational quantum algorithms, the ratio of quantum to classical compute time is highly variable with the classical optimization algorithm used.

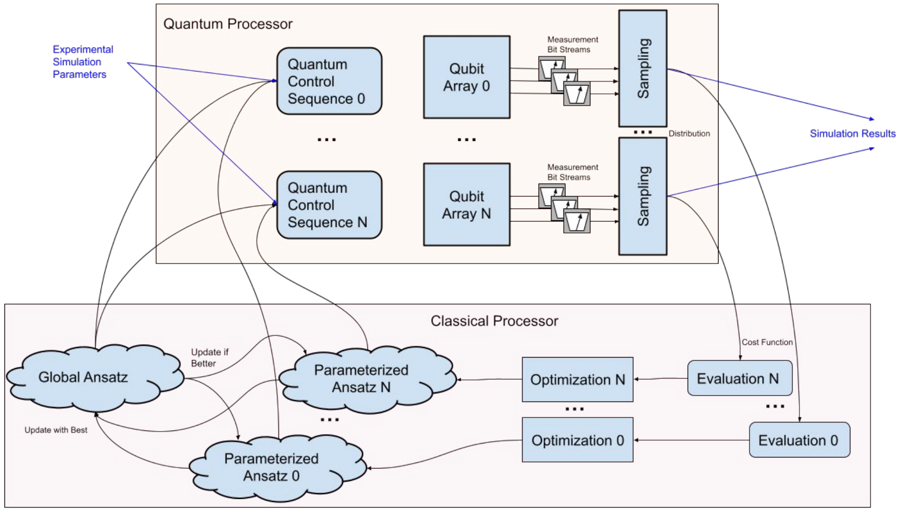


Architecting for Performance

Coarse-grained parallelism is abundant in variational quantum approaches.

A given trial parameterization can be run simultaneously on multiple quantum processors, with linear speedup.

Multiple trial parameterizations can be explored in parallel, with the potential for voting/pruning based on competition between them.

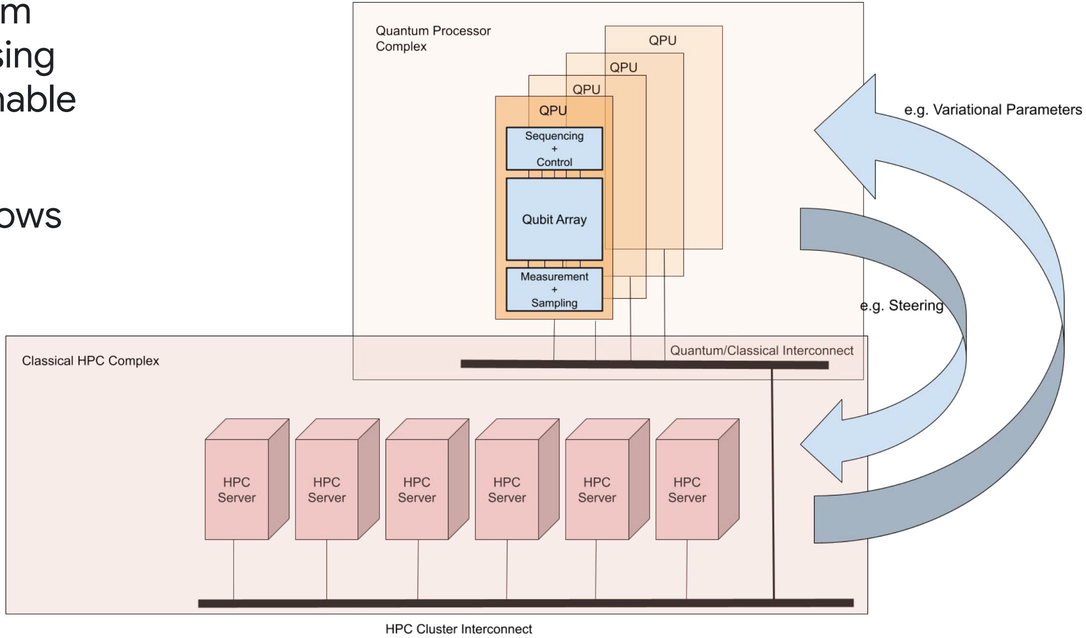


Future Hybrid Systems Architectures

For more closed-cycle hybrids like quantum corrected QMC, the greatest value is in using large volumes of quantum resources to enable large scale hybrid computations.

The potential for quantum acceleration grows exponentially with scale.

Expect to see best-in-class qubits deployed close to HPC clusters already adapted to large quantum problems.



Thank You for Your Interest!

Resources:

Cirq

https://github.com/quantumlib/Cirq

TensorFlow Quantum

https://www.tensorflow.org/quantum