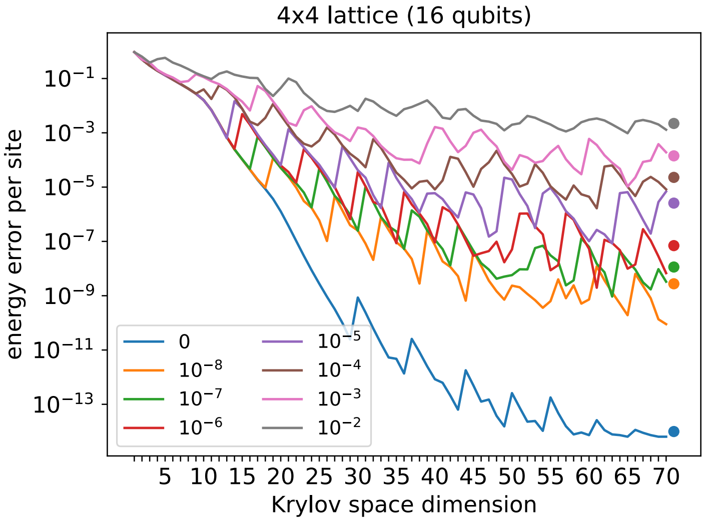
Quantum Krylov algorithms for ground state energy approximation

Will Kirby

IBM Quantum

IBM Research Cambridge, 314 Main St, Cambridge, MA



Motivation

Estimate ground state energy of quantum Hamiltonian.

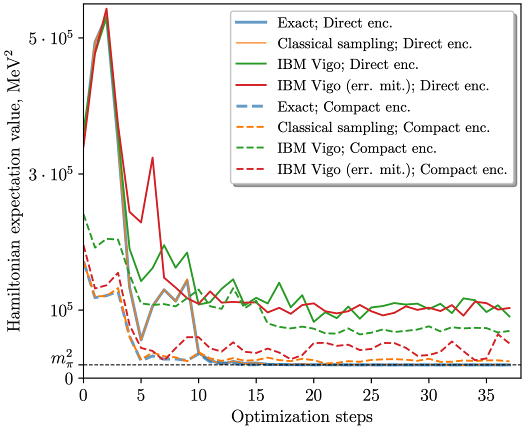
Abstract perspective: approximate lowest eigenvalue of Hermitian matrix.

Goal is classically challenging due to exponential Hilbert space dimension*.

*Assuming general, hard case: sufficiently entangled, supported on exponentially-many basis states, etc.

Example applications:

- Quantum chemistry.
- Condensed matter physics.
- Nuclear physics.
- High-energy physics.



M. Kreshchuk *et al.*, Entropy 23, 59 (2021).

Lanczos method

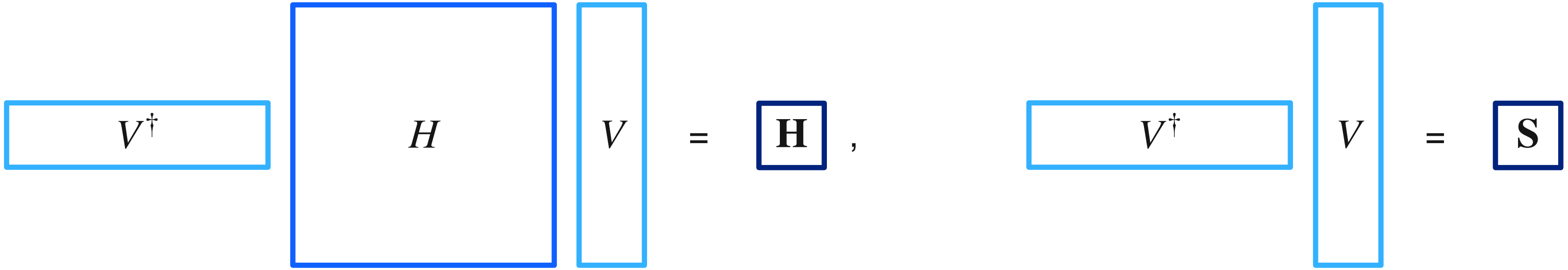
= classical method for approximating lowest eigenvalues.

High-level idea:

1. Initial guess $|\psi_0\rangle \quad \Rightarrow \quad H|\psi_0\rangle \quad \Rightarrow \cdots \Rightarrow \quad H^{D-1}|\psi_0\rangle$

Krylov space

2. $(\mathbf{H}, \mathbf{S})$ = project H onto $\text{span}\underbrace{[\,|\psi_0\rangle,\, H|\psi_0\rangle,\, H^2|\psi_0\rangle,\, \ldots,\, H^{D-1}|\psi_0\rangle]}_{V}$



3. Lowest eigenvalue of $(\mathbf{H}, \mathbf{S})$ i.e., of $\mathbf{Hv} = \lambda \mathbf{Sv}$, approximates lowest eigenvalue of H

Lanczos method

Caveat: typically in classical Lanczos(-like) methods, would orthogonalize along the way… challenging in quantum implementations.

Advantage: converges exponentially with D (in ∞ precision arithmetic).

Disadvantage: classically, requires storing entire statevectors $H^i|\psi_0\rangle$ ⇒ exponential overhead.

Can we construct a quantum version that mitigates statevector overhead while keeping fast convergence?[1]

[1] Klymko *et al.*, PRX Quantum 3, 020323 (2022); Epperly *et al.*, SIAM J. Mat. An. Appl. 43, 1263-1290 (2022); and many more!

Quantum “Lanczos method” = “Quantum Krylov”

Options for generating Krylov space: multiply $|\psi_0\rangle$ by…

- Powers of H — same as original Lanczos ⇒ nontrivial on quantum but possible in principle.[1]
- $e^{-Hk\,dt}$ — this version claimed “Qlanczos.”
- $e^{iHk\,dt}$ — many good options e.g. Trotterization, qubitization, etc.
- $T_k(H)$ — arises naturally from block encoding.

Will focus on last two in this talk.

[1]Seki and Yunoki, PRX Quantum 2, 010333 (2021)

Quantum Krylov with real time-evolutions

Majority of works have focused on Krylov states generated by real time-evolution:

$$V = [\,|\psi_0\rangle, \quad U|\psi_0\rangle, \quad U^2|\psi_0\rangle, \ldots, U^{D-1}|\psi_0\rangle] \text{ for } U = e^{iH\,dt}$$

Need to estimate

$$\mathbf{H_{jk}} = \langle\psi_0|(U^j)^\dagger H U^k|\psi_0\rangle,$$

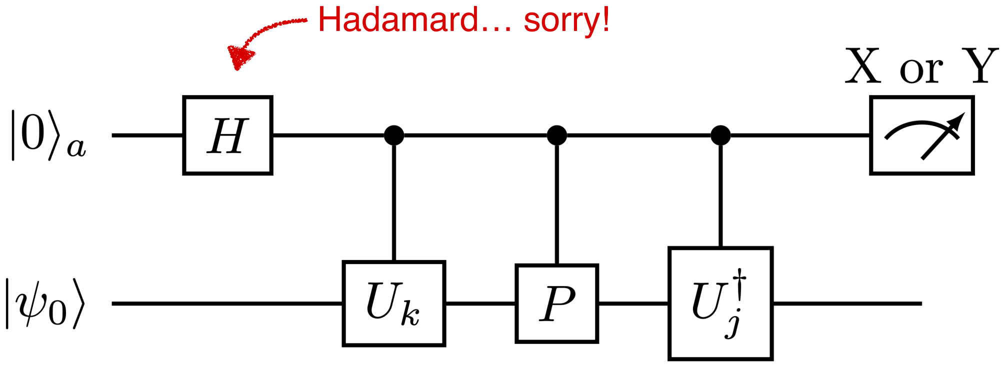
$$\mathbf{S_{jk}} = \langle\psi_0|(U^j)^\dagger U^k|\psi_0\rangle$$

for each $j, k = 0,1,...,D-1$.

Estimating matrix elements (simple version)

Targets: $\mathbf{H_{jk}} = \langle \psi_0 | (U^j)^\dagger H U^k | \psi_0 \rangle, \quad \mathbf{S_{jk}} = \langle \psi_0 | (U^j)^\dagger U^k | \psi_0 \rangle$.

Can approach using Hadamard test:*



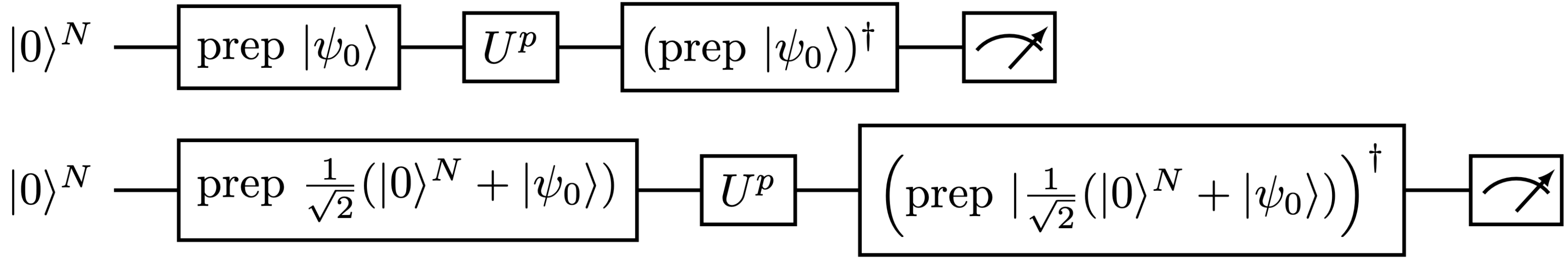
Yields $\langle X \rangle_a = Re\left[\langle \psi_0 | (U^j)^\dagger P U^k | \psi_0 \rangle\right], \quad \langle Y \rangle_a = Im\left[\langle \psi_0 | (U^j)^\dagger P U^k | \psi_0 \rangle\right]$

*Cortes and Gray, 2021.

Estimating matrix elements (better version)

Change target slightly: $\mathbf{U_{jk}} = \langle \psi_0 | (U^j)^\dagger U U^k | \psi_0 \rangle, \quad \mathbf{S_{jk}} = \langle \psi_0 | (U^j)^\dagger U^k | \psi_0 \rangle$.

Suppose Hamiltonian preserves particle number (Hamming weight)…



Circuit #2: $\frac{1}{4} \left| (\langle 0 | + \langle \psi_0 |) | U^p | (|0\rangle + |\psi_0\rangle) \right|^2$ must have nonzero Hamming weight

$$= \frac{1}{4} \left(|\langle 0 | U^p | 0 \rangle|^2 + |\langle \psi_0 | U^p | \psi_0 \rangle|^2 + 2\text{Re}\left[\langle 0 | U^p | 0 \rangle \langle \psi_0 | U^p | \psi_0 \rangle\right] \right)$$

Circuit #1

*Cortes and Gray, 2022.

Quantum Krylov with real time-evolutions

Summary:

- Estimate $\mathbf{H_{jk}}$, $\mathbf{S_{jk}}$ via Hadamard(-ish) tests and repeated sampling.
- Depending on Hamiltonian, can sometimes avoid controlled time-evolutions using symmetries.[1]
- Advantage: can use crude approximations for time-evolution to get low circuit depth.
- Disadvantage: time-evolution always approximated — more accuracy requires more depth.

[1]Cortes and Gray, Phys. Rev. A 105, 022417 (2022).

Quantum Krylov from block encoding

The most accurate simulations of time-evolution require Hamiltonian input as block encoding.[1]

Block encoding: for H on n qubits (s.t. $||H|| \leq 1$), find U on $m+n$ qubits s.t.

$$U = \begin{pmatrix} H & \cdot \\ \cdot & \cdot \end{pmatrix}, \qquad R = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & & -1 & & \\ & 0 & & & \ddots & \\ & & & & & -1 \end{pmatrix}$$

[1] Low and Chuang, Quantum 3 (2019).

Quantum Krylov from block encoding

$$U^2 = 1 \quad \Rightarrow \quad (RU)^j = \begin{pmatrix} T_j(H) & \cdot \\ \cdot & \cdot \end{pmatrix}$$

Brief proof:

1. Let $H|\lambda\rangle = \lambda|\lambda\rangle$ and $U = |G\rangle\langle G| \otimes H + \ldots$

2. $\Rightarrow \quad U|G\rangle|\lambda\rangle = \lambda|G\rangle|\lambda\rangle + \sqrt{1-\lambda^2}|\perp\rangle \quad \Rightarrow \quad U \sim \begin{pmatrix} \lambda & \cdot \\ \sqrt{1-\lambda^2} & \cdot \end{pmatrix}$

3. $U^2 = 1 \quad \Rightarrow \quad U \sim \underbrace{\begin{pmatrix} \lambda & \sqrt{1-\lambda^2} \\ \sqrt{1-\lambda^2} & -\lambda \end{pmatrix}}_{\text{reflection}} \quad \Rightarrow \quad RU \sim \underbrace{\begin{pmatrix} \lambda & \sqrt{1-\lambda^2} \\ -\sqrt{1-\lambda^2} & \lambda \end{pmatrix}}_{\text{rotation}}$

Quantum Krylov from block encoding

$$(RU)^j = \begin{pmatrix} T_j(H) & \cdot \\ \cdot & \cdot \end{pmatrix}$$

$\Rightarrow$ Can use block encoding to exactly construct $T_j(H)\,|\psi_0\rangle$

Recall: Lanczos method ~ project H onto

$$\begin{aligned} &\text{span}\{\,|\psi_0\rangle,\, H|\psi_0\rangle,\, H^2|\psi_0\rangle,\, \ldots,\, H^{D-1}|\psi_0\rangle\} \\ = \;&\text{span}\{\,|\psi_0\rangle,\, T_1(H)|\psi_0\rangle,\, T_2(H)|\psi_0\rangle,\, \ldots,\, T_{D-1}(H)|\psi_0\rangle\} \end{aligned}$$

Quantum Krylov from block encoding

$$\mathbf{H_{jk}} = \langle \psi_0 | T_j(H) H T_k(H) | \psi_0 \rangle$$

$$= \frac{1}{4}\left(\left\langle T_{i+j+1}(H) \right\rangle_0 + \left\langle T_{|i+j-1|}(H) \right\rangle_0 + \left\langle T_{|i-j+1|}(H) \right\rangle_0 + \left\langle T_{|i-j-1|}(H) \right\rangle_0 \right)$$

$$\mathbf{S_{jk}} = \langle \psi_0 | T_j(H) T_k(H) | \psi_0 \rangle = \frac{1}{2}\left(\left\langle T_{i+j}(H) \right\rangle_0 + \left\langle T_{|i-j|}(H) \right\rangle_0 \right)$$

for $j, k = 0,1,2,\ldots,D-1$. In other words, need to estimate

$$\langle T_k(H) \rangle_0 = \langle \psi_0 | T_k(H) | \psi_0 \rangle$$

for $k = 0,1,2,\ldots,2D-1$.

Quantum Krylov from block encoding

Since $(RU)^j = \begin{pmatrix} T_j(H) & \cdot \\ \cdot & \cdot \end{pmatrix}$

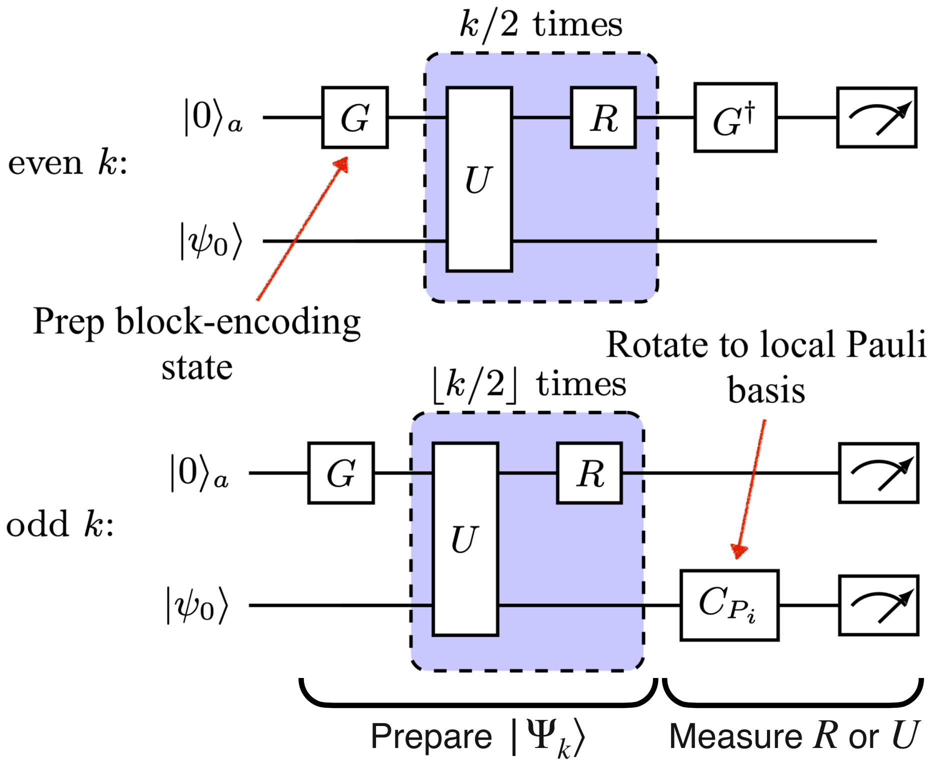
$$\Rightarrow \quad \langle T_k(H)\rangle_0 = \langle \psi_0 | T_k(H) | \psi_0 \rangle = (\langle G | \otimes \langle \psi_0 |)(RU)^k(|G\rangle \otimes |\psi_0\rangle)$$

$$= \begin{cases} (\langle G | \otimes \langle \psi_0 |)(UR)^{\lfloor k/2 \rfloor} R\, (RU)^{\lfloor k/2 \rfloor}(|G\rangle \otimes |\psi_0\rangle) & \text{if } k \text{ is even,} \\ \underbrace{(\langle G | \otimes \langle \psi_0 |)(UR)^{\lfloor k/2 \rfloor}}_{\langle \Psi_k |} U \underbrace{(RU)^{\lfloor k/2 \rfloor}(|G\rangle \otimes |\psi_0\rangle)}_{|\Psi_k\rangle} & \text{if } k \text{ is odd.} \end{cases}$$

Quantum Krylov from block encoding

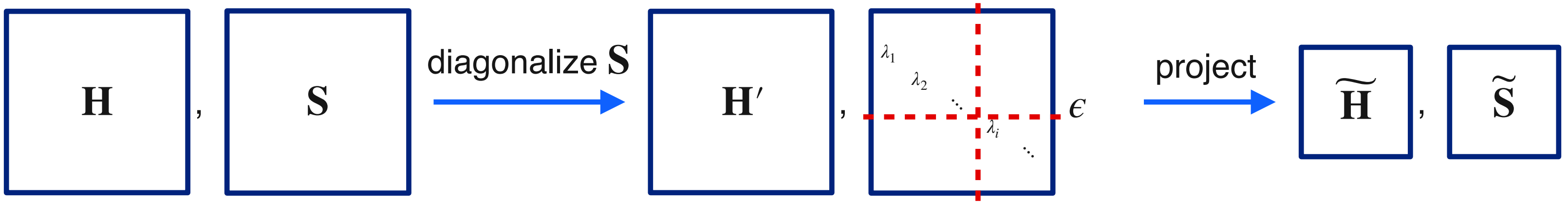
That yields the necessary circuits: for each $k = 0,1,2,\ldots,D-1\ldots$

$$U = \sum_i |i\rangle\langle i| \otimes P_i$$



Regularizing the Krylov space

- Either real-time or block-encoding Krylov methods yield noisy estimates of $(\mathbf{H}, \mathbf{S})$.
- Want to solve $\mathbf{Hv} = \lambda \mathbf{Sv}$.
- Ill-conditioned if $\mathbf{S}$ is ill-conditioned ⇒ need to regularize.
- “Canonical orthogonalization” or “thresholding”: project $(\mathbf{H}, \mathbf{S})$ onto eigenspaces of $\mathbf{S}$ above threshold ϵ.



- $\mathbf{S}$ is metric in Krylov space ⇒ choose threshold ~ noise rate ⇒ discards vectors compatible with 0.

Error analysis

- Real-time: analyzed in Epperly *et al.*, SIAM J. Mat. An. Appl. 43, 1263-1290 (2022).
- Block-encoding: in our paper, modification of Epperly's analysis.
- Three main error terms:

$$\underbrace{\text{error from Krylov space} + \text{error from thresholding}}_{\text{Requires modification}} + \underbrace{\text{error from noise}}_{\text{Same for real-time and block-encoding}}$$

- First two terms **high-level idea of proof:**
  - Krylov space = span[$|\psi_0\rangle, T_1(H)|\psi_0\rangle, T_2(H)|\psi_0\rangle, \ldots, T_{D-1}(H)|\psi_0\rangle$] = poly$_{D-1}(H)|\psi_0\rangle$
  - $\Rightarrow$ Best poly approx to delta function at E_0 = approx ground space projector in Krylov space.
  - Thresholding $\Rightarrow$ perturbation of Chebyshev expansion coefficients of projector.

Error analysis

"In practice" results:[1] to reach energy error $\mathscr{E}$, require…

$$\text{Krylov space dimension } D = \Theta\left[\left(\log\frac{1}{|\gamma_0|} + \log\frac{1}{\mathscr{E}}\right)\min\left(\frac{1}{\mathscr{E}}, \frac{1}{\Delta}\right)\right],$$

$$\text{Measurements per dimension } M = \Theta\left(\frac{1}{\mathscr{E}^2} + \frac{1}{\mathscr{E}|\gamma_0|^4}\right)$$

where $\gamma_0 =$ initial state overlap with low-energy subspace, $\Delta =$ spectral gap.

[1] "In practice" because theoretical bound only guarantees first term in M is $\mathscr{E}^{-p}$ for $p \in [2,3]$; $p = 2$ is based on numerics.

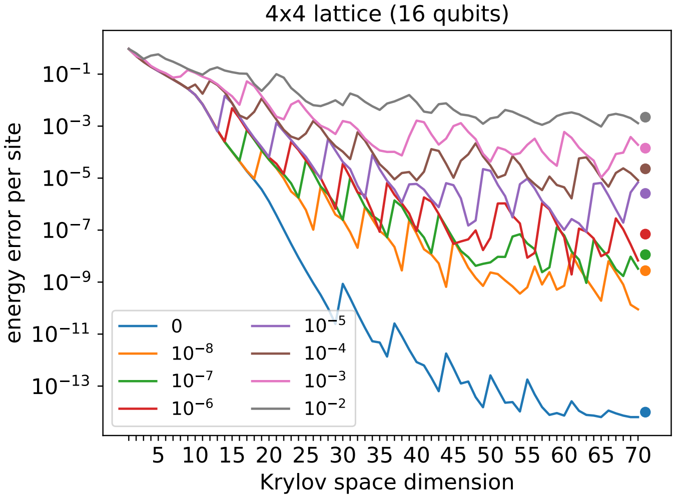
[2] Above is for block-encoding; real-time Krylov space analysis is similar.

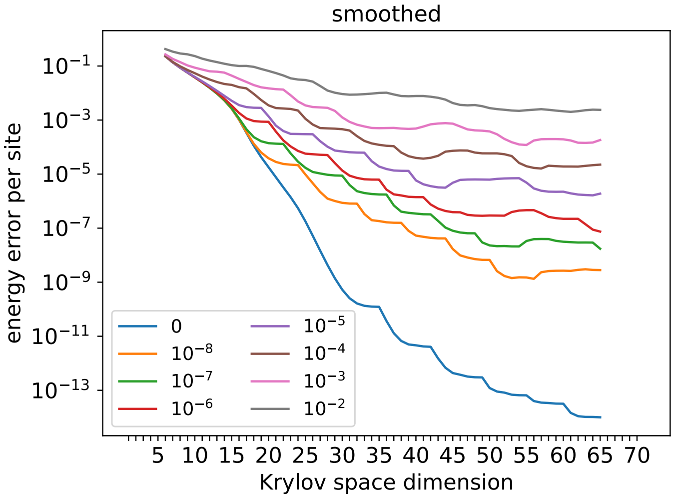
Comparing theory to numerics

$$\mathcal{E} \leq O\left(\frac{1}{\sqrt{M}} + \frac{\sqrt{\delta}}{|\gamma_0|^2\sqrt{M}} + \delta + \frac{1}{|\gamma_0|^2}\left(1+\frac{\delta}{2}\right)^{-D}\right)$$

Error from noise — Error from Krylov space

δ free, best choice is $\delta = \Theta\big(\max(\text{target error}, \Delta)\big)$





Thank you!

Questions?

Will Kirby, Mario Motta, and Antonio Mezzacapo, Quantum 7, 1018 (2023),

https://quantum-journal.org/papers/q-2023-05-23-1018/.