Exqaliber Stage 1 Report

James Cruise[1], Joseph Tedds[1], Camille de Valk[2], and Walden Killick[1]

[1]Cambridge Consultants, Cambridge, UK
[2]Capgemini QLab

TBD

1 Introduction

Quantum amplitude estimation (QAE) is a fundamental subroutine in many application areas including quantum chemistry, machine learning, finance, and more. It is the problem of, given access to an $(n+1)$-qubit oracle $\mathcal{A}$ such that $\mathcal{A}|0\rangle = \sqrt{a}|\phi_1\rangle|1\rangle + \sqrt{1-a}|\phi_0\rangle|0\rangle$, estimating a. This is a problem of great current interest due to its promise of a quantum quadratic speedup, including potentially on noisy near-term devices.

The first such algorithm, proposed by Brassard et al. [1], works by combining the quantum phase estimation (QPE) algorithm [2] with the Grover iteration operator. Although this achieves a provable quantum speedup, this approach is unsuitable for use on near-term devices primarily due to its large circuit depth stemming from (1) the quantum Fourier transform (QFT) [3] and (2) the controlled unitaries required to prepare the input state of the QFT.

Several recent approaches have studied algorithms achieving the same or similar asymptotic speedups without the use of the QFT, in particular by repeated sampling of the state after only repeated applications of the Grover operator [4, 5, 6]. The underlying motif of this paradigm is the learning of the amplitude by statistical sampling, which is in contrast to Brassard et al.'s method which allows for direct access to the amplitude with high probability. Each of these approaches has its own problems with regards to practical implementation on near-term devices; most notably, rapidly growing gate depths, large constant-factor overheads, and expensive controlled-Grover operations. All of these are problematic in the presence of noise as repeated applications of the Grover operator causes the state to decohohere; after which, without a noise model, information supposedly gained about the amplitude may be incorrect or skewed.

In this work, we present hybrid quantum-classical algorithms for QAE with a focus on minimal circuit depth and measurement counts, and taking into account the noisiness of near-term devices. (tbc. after we know what's in the rest of the report)

James Cruise: james.cruise@cambridgeconsultants.com
Joseph Tedds: joseph.tedds@cambridgeconsultants.com
Camille de Valk: camille.de.valk@capgemini.com
Walden Killick: walden.killick@cambridgeconsultants.com

1.1 Amplitude estimation

Problem 1 (Quantum Amplitude Estimation)**.** *Given some* $(n+1)$*-qubit oracle* $\mathcal{A}$ *which generates the state*

$$\mathcal{A}\left|0\right\rangle = \sqrt{a}\left|\phi_1\right\rangle\left|1\right\rangle + \sqrt{1-a}\left|\phi_0\right\rangle\left|0\right\rangle,$$

where $\left|\phi_1\right\rangle$ *and* $\left|\phi_0\right\rangle$ *are arbitrary normalised* n*-qubit states and* $0 \leq a \leq 1$*, quantum amplitude estimation (QAE) is the problem of estimating the unknown amplitude* $\sqrt{a}$*.*

The 'naive', classical way to solve this problem is to simply prepare and measure this state many (say N) times, counting the number of times $\left|\phi_1\right\rangle\left|1\right\rangle$ was observed (say K), then define the estimate for $\sqrt{a}$ as $\sqrt{K/N}$. Chebyshev's inequality tells us that taking $N \in O(\varepsilon^{-2})$ samples is sufficient to approximate $\sqrt{a}$ within additive error ε (with, say, 99% confidence). Furthermore, this is asymptotically optimal [7]. By contrast, in a manner which will be made more precise in following sections, quantum computers allow us to modify the state before measurement, which allows us to achieve the same degree of accuracy using only $O(\varepsilon^{-1})$ queries (known as the Heisenberg limit), achieving a quadratic quantum speedup [1].

An important and general application of QAE is Monte Carlo estimation [8, 9, 5] - a method for estimating the mean value of a function via random sampling. More precisely, given some function $f : \{0,1\}^n \to [0,1]$, the task is to estimate the mean value

$$\mathbb{E}[f(X)] = \frac{1}{2^n}\sum_{x=0}^{2^n-1} f(x).$$

Montanaro [9] presents a method for encoding this value in the amplitude of a quantum state, thereby reducing this problem to a case of amplitude estimation, which in turn offers a quadratic speedup for Monte Carlo estimation over classical methods.

1.1.1 Statistical amplitude estimation

Although the original QAE algorithm based on QPE is still the superior algorithm to run on a fault-tolerant quantum computer, the noisiness of near-term devices makes all algorithms utilising the QFT, including the aforementioned algorithm, very difficult to implement (primarily due to the large circuit depth). Thus, the current paradigm in the design of QAE algorithms is to sample the state after amplitude amplification of the state $\mathcal{A}\left|0\right\rangle$, refining the estimate for $\sqrt{a}$ and increasing confidence in the estimate with consecutive samples.

By setting $\theta = \sin^{-1}\sqrt{a}$, we can rewrite the state as

$$\mathcal{A}\left|0\right\rangle = \sin\theta\left|\phi_1\right\rangle\left|1\right\rangle + \cos\theta\left|\phi_0\right\rangle\left|0\right\rangle.$$

Now consider the Grover iterate given by

$$\mathcal{U} = \mathcal{A}\left(2\left|0^{n+1}\right\rangle\!\!\left\langle 0^{n+1}\right| - I_{2^{n+1}}\right)\mathcal{A}^{-1}\left(I_{2^n}\otimes Z\right)$$

as in Figure 1. In the subspace spanned by $\left|\phi_1\right\rangle\left|1\right\rangle$ and $\left|\phi_0\right\rangle\left|0\right\rangle$, by a sequence of reflections, each application of $\mathcal{U}$ performs a rotation of angle 2θ towards $\left|\phi_1\right\rangle\left|1\right\rangle$. That is,

$$\mathcal{U}^m\mathcal{A}\left|0\right\rangle = \sin(2m+1)\theta\left|\phi_1\right\rangle\left|1\right\rangle + \cos(2m+1)\theta\left|\phi_0\right\rangle\left|0\right\rangle$$

and so $\mathcal{U}^m$ takes the probability of measuring $\left|\phi_1\right\rangle\left|1\right\rangle$ from $\sin^2\theta$ to $\sin^2(2m+1)\theta$.

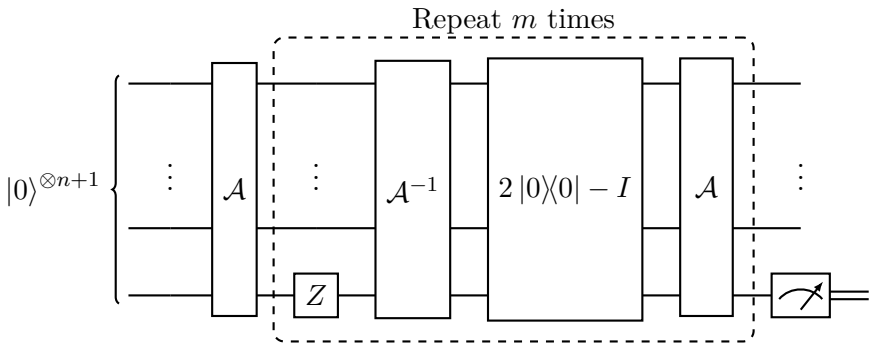

Figure 1: Circuit for amplitude amplification. The probabilities for the measurement outcomes are $\Pr(|1\rangle) = \sin^2(2m+1)\theta$ and $\Pr(|0\rangle) = \cos^2(2m+1)\theta$.

1.1.2 Decoherent noise model

Performing Bayesian inference without a noise model on a device which in reality is noisy may cause incorrect information to be attained about the angle θ. For example, if the true angle $\theta = 0$, then, if the system were truly noiseless, no number of Grover iterates should affect the state and make it possible to measure $|\phi_1\rangle|1\rangle$. However, in the realistic noisy case, depolarisation stemming from the application of gates may cause the probability of measuring $|\phi_1\rangle|1\rangle$ to rise above 0 nonetheless. If we measure this state, without a noise model, our understanding of the angle effectively eliminates the case $\theta = 0$. The presence of a noise model mitigates this issue by effectively decreasing confidence in the measurement outcome as the number of Grover iterations increases, thereby retaining the possibility that the non-zero probability was the result of noise.

In the presence of depolarising noise, it is reasonable to assume that consecutive applications of $\mathcal{A}$ and its inverse cause, on average, the probability of measuring $|\phi_1\rangle|1\rangle$ to tend to $\frac{1}{2}$, at which point no information about the original state is recoverable.

In the ideal (fault-tolerant) model, the probability of measuring $|\phi_1\rangle|1\rangle$ after m iterations of U to $\mathcal{A}|0\rangle$ is $\sin^2(2m+1)\theta$, which can be rewritten as

$$\sin^2(2m+1)\theta = \frac{1}{2}\left(1 - \cos(4m+2)\theta\right).$$

If we assume that each iteration of $\mathcal{A}$ or its inverse dampens the oscillating term (which contains information about θ) by some factor $e^{-\lambda}$, where $\lambda \geq 0$ is a constant corresponding to the noisiness of the system, then we can model the new probability of measuring $|\phi_1\rangle|1\rangle$ after m iterations of $\mathcal{U}$ as

$$\Pr(|\phi_1\rangle|1\rangle) = \frac{1}{2}\left(1 - e^{-\lambda(2m+1)}\cos(4m+2)\theta\right).$$

Here, the exponent contains the term $2m+1$ since we need one application of $\mathcal{A}$ to prepare the state, then each iteration of $\mathcal{U}$ requires one call to $\mathcal{A}$ and another to its inverse.

(maybe write some more about what we do with this noise model - tbc. after the rest of the report is written)

1.2 Related work

Brassard et al. [1] were the authors of the first QAE algorithm, now often referred to as canonical QAE. The idea is that the Grover operator $\mathcal{U}$ as defined above contains the

amplitude of the corresponding state in its eigenvalues, and thus the QPE algorithm can be used with $\mathcal{U}$ to extract this value. This method allows for direct access to the amplitude to within additive error ε with high probability using $O(\varepsilon^{-1})$ queries to $\mathcal{A}$, which makes it still the superior algorithm for QAE on a fault-tolerance quantum computer. However, its reliance on controlled Grover iterates and the QFT make it very difficult to implement on near-term devices, which has inspired much subsequent work on QAE without the use of the QFT.

Suzuki et al. [5] propose a QFT-free QAE algorithm based on maximum likelihood estimation. Although lacking a rigorous proof, numerical simulation seems to show that their algorithm achieves the asymptotically optimal scaling of measurements. However, the scheme which achieves this scaling incurs an exponentially increasing number of Grover iterates at each step, which is problematic especially in the presence of noise. Furthermore, the choice of number of Grover iterates is not dependent on previous measurement outcomes, which suggests that there is further optimisation to be achieved in minimising the circuit depth.

Wie [6] sketches another QFT-free QAE algorithm based on Hadamard tests, similar to that which is used in iterative phase estimation as presented by Kitaev [2]. Apart from also lacking a proof of optimality, Wie's algorithm uses the more expensive controlled-Grover operations.

Aaronson and Rall [4] also present an algorithm for QFT-free QAE, and were the first to rigorously prove a quadratic speedup for such an algorithm. Although this algorithm achieves the optimal asymptotic complexity, large constant-factor overheads make it impractical for use even in the fault-tolerant case [10].

Grinko et al. [10] present an algorithm called iterative QAE which combines ideas from previous works and greedily chooses the number of Grover iterates at each step to maximise the quantum Fisher information. They prove that their algorithm is optimal up to a double-logarithmic factor and has much smaller constant factors than comparable algorithms.

Giurgica-Tiron et al. [11] design algorithms which interpolate between classical and quantum amplitude estimation algorithms with the aim of utilising parallelism to minimise overall circuit depth. Their algorithms achieve a query complexity of $\tilde{O}\left(\varepsilon^{-(1+\beta)}\right)$ where $\beta \in (0, 1]$ is some parameter corresponding to the balance between classical and quantum queries.

Smith et al. [12] present an algorithm for QFT-free QPE (which can equivalently be used for QAE by replacing the unitary operator with the Grover iterate) which is also based on maximising Fisher information. They prove that it achieves the Heisenberg limit in the noiseless case, and prove that it achieves the best possible query complexity in the presence of depolarising noise.

1.2.1 Quantum phase estimation

tbc. how much should we write about this? already discussed it a bit in previous sections

The first algorithm for QAE, presented by Brassard et al. [1], achieves its objective by applying the QPE algorithm [2] to the Grover iterate defined above.

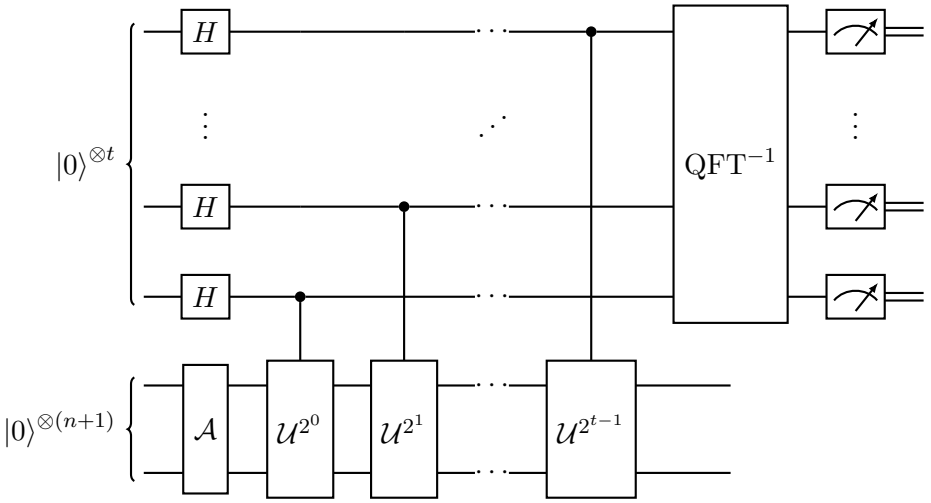

Figure 2: Amplitude estimation by phase estimation

1.3 Contribution

(to be written once we know what's in the rest of the paper)

2 Algorithm

2.1 Algorithm description

Give a statement of our algorithm, phrase as an algorithm rather than talking about Bayes updates etc. What is implemented

State is current estimate and variance Rules for updating state given result Rules for selecting next update

Include detail of stopping criteria

2.1.1 Adjustments for noise

Changes needed for decoherence noise

2.2 Bayesian framework

Introduction to Bayesian stats, prior, posterior, and adaptive termination

Introduce idea of using normal distribution

2.2.1 Posterior properties

Posterior mean, variance and expected variance under noiseless and decoherence noise model

2.2.2 Normal approximation of posterior

Details on the use of a normal approximation for posterior distribution and next prior. Include comments about both the circular issue and tails due to indeterminacy.

2.3 Selecting next sample depth

Details of greedy choice (possibly give indication of dynamic programming approach) Noise and noiseless

3 Experimentation

We conduct a series of experiments for the noiseless and noisy simulations of our algorithm in this section. We empirically compare against the IAE and MLAE algorithms. As the update steps for these algorithms are equivalent to sampling from a Bernoulli distribution for a given choice of θ_0 and noise level we sample directly from the Bernoulli distribution instead of simulating the dynamics of a quantum system for each algorithm. This algorithm has also been implemented and tested using Qiskit [13], with a view for future experiments with quantum simulators and quantum hardware.

3.1 Simulation

3.1.1 Noiseless

First, we explore the convergence time for a range $\theta \in [0, \pi/2]$. Note that as shown in [14], so-called exceptional points with poor convergence occur near to rational multiples of π, therefore we sample values of θ from continuous distributions instead of equally spaced angles. Figure 3 shows that the convergence for θ close to 0 is several orders of magnitude larger than convergence for other values. For this reason, we consider partitioning the space $\Theta = [0, \pi/2]$ into the central region $\Theta_0 = [\frac{\pi}{12}, \frac{5\pi}{12}]$ and the edge region $\Theta_1 = \Theta \backslash \Theta_0 = [0, \frac{\pi}{12})$.

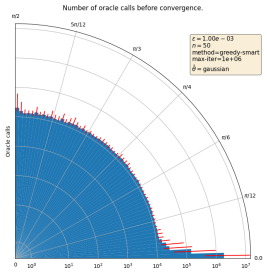

Figure 3: Median number of oracle calls for SAE to converge to a precision of $\varepsilon = 10^{-3}$, with 25% and 75% quartiles shown. Each bar is a region of width $\pi/120$ with the median number of oracle calls calculated from 50 values of true value θ_0 selected uniformly at random. The prior for each iteration is taken to be $N(\tilde{\theta}_0, 1)$, where $\tilde{\theta}_0$ is sampled from a $N(\theta_0, 0.1)$ distribution and success probability $1 - \alpha$ with $\alpha = 0.01$.

(The following is all conjecture as we don't have a comparison graph.)
As seen in Figure 4, the set Θ_1 performs poorly for all methods that take a statistical approach. This region contains points where the variance reduction factor approaches 0, with the extreme points where $\theta = 0, \pi$.

- We should consider how the QoPrime and Power-law AE algorithms deal with Θ_1
- May need a gradient graph here?
- Or graph of variance update ratios - there are some times when it's negative for both, negative for neither or negative for only one.

Figure 4: Median number of oracle calls for $\theta \in \Theta_1$ and fixed error rate $\varepsilon = 10^{-3}$ for different statistical amplitude estimation routines. Each section of the histogram is a region of width $\pi/120$ with the median number of oracle calls calculated from 30 values of true value θ_0 selected uniformly at random. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$.

From now on, we restrict to considering $\theta \in \Theta_0$.

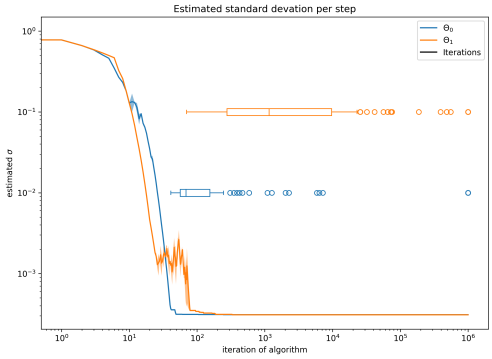

Figure 5: Median standard deviation for fixed error rate $\varepsilon = 10^{-3}$ against time step for $\theta_0 \in \Theta_0$ and $\theta_0 \in \Theta_1$. We sample θ_0 from a uniform distribution of x and 50 samples for Θ_0 and Θ_1 respectively. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$. The box plots show the Q1, Q2 and Q3 algorithm termination times for the recorded runs.

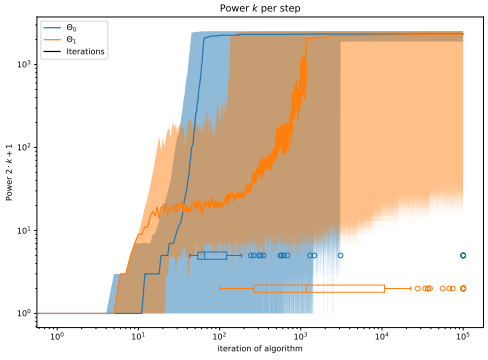

Figure 6: Median depth for fixed error rate $\varepsilon = 10^{-3}$ against time step for $\theta_0 \in \Theta_0$ and $\theta_0 \in \Theta_1$. We sample θ_0 from a uniform distribution of x and 50 samples for Θ_0 and Θ_1 respectively. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$. The box plots show the Q1, Q2 and Q3 algorithm termination times for the recorded runs.

A source of potential error in this algorithm is the normal approximation made at each step, which removes the theoretical guarantees of the Bernstein-von-Mises theorem. This can be seen in Figure 7, where should expect $(1-\alpha)\%$ of the points to lie above the diagonal. As a mitigation technique, we post-process the measurement data using the MLE estimator, and recover some of the desired accuracy.

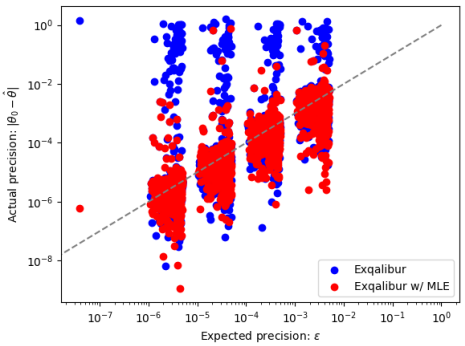

Figure 7: Actual precision of the final estimate for SAE with and without MLE post-processing. We sample $\theta_0 \in \Theta_0$ from a uniform distribution of x and 50 for target precisions of $\varepsilon = 10^{-3}, 10^{-4}, \ldots, 10^{-7}$. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1-\alpha$ with $\alpha = 0.01$.

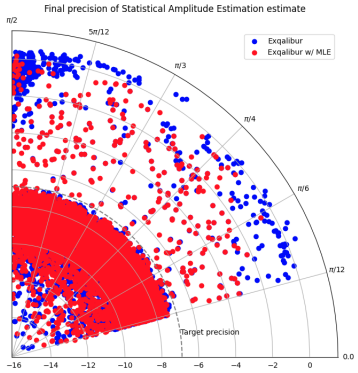

Figure 8: Actual precision of the final estimate for SAE with and without MLE post-processing. We sample $\theta_0 \in \Theta_0$ from a uniform distribution of x and 500 for a target precision of $\varepsilon = 10^{-3}$. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1-\alpha$ with $\alpha = 0.01$.

We compare the performance of SAE to IAE and MLAE, and observe that in the noiseless case we achieve similar query to IAE. However, there still remain some points for which SAE fails to satisfactorily converge.

3.1.2 Decohering Noise

We explore the performance of our algorithm for a range of noise rates.

As shown in Figure 11, the algorithm will not go beyond a depth of $\sim \frac{1}{\lambda}$.

- Concerned that error comparisons make no sense here. These algorithms have no capacity to cope with noise so obviously they're bad - but others e.g. Hitachi, QoPrime, Power-law AE do

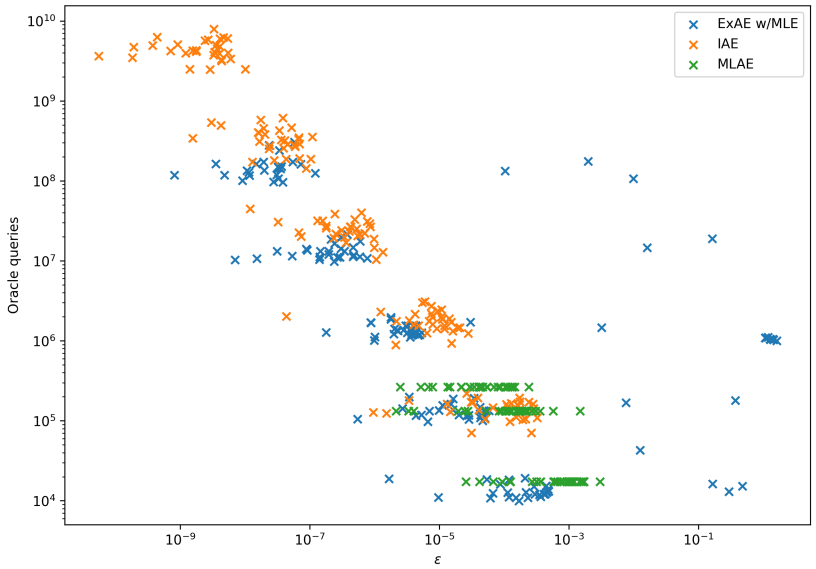

Figure 9: Actual precision versus total number of oracle queries used for IAE, MLAE and SAE. We target precisions of $\varepsilon = 10^{-3}, 10^{-4}, \ldots, 10^{-7}$, with each point a single value of $\theta_0 \in \Theta_0$ and take the final precision to be half the width of the confidence interval. We sample uniformly from Θ_0 for 30 values of θ_0. Each value of θ_0 is evaluated for all algorithms and each target ε. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$.

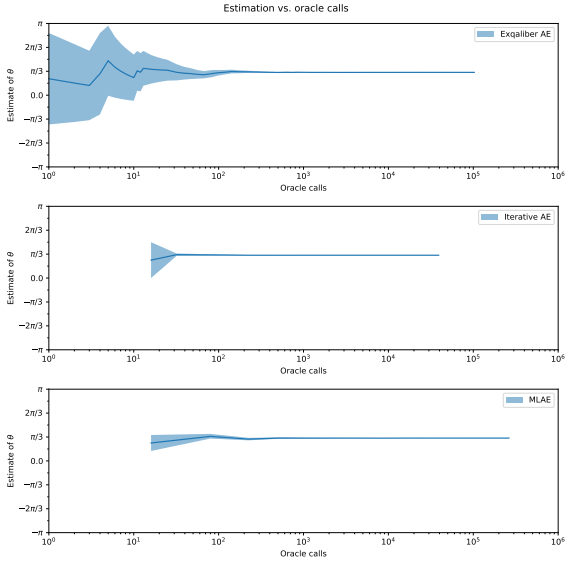

Figure 10: Confidence intervals for a single run of SAE, IAE and MLAE with $\theta_0 = 1$. The prior for SAE is taken to be $N(\pi/2, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$. As IAE and MLAE use a fixed number of shots per step, the number of oracle calls does not start at 0.

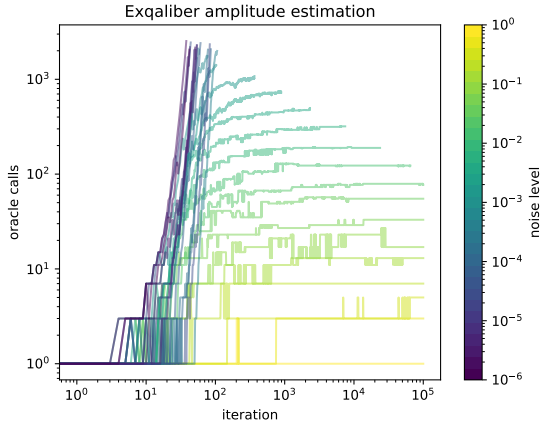

Figure 11: Depth of SAE for $\theta_0 = 1$ at varying levels of decohering noise characterised by $\lambda = 10^0, 10^{-1}, \ldots 10^{-6}$. The prior for each iteration is taken to be $N(\pi/2, 1)$, we target $\varepsilon = 10^{-3}$ and success probability $1 - \alpha$ with $\alpha = 0.01$.

Figure 12: Total number of oracle queries for SAE with decohering noise characterised by $\lambda = 10^{-1}, 10^{-2}, \ldots, 10^{-7}$. Each point is a single value of $\theta_0 \in \Theta_0$. We sample uniformly from Θ_0 for 30 values of θ_0. Each value of θ_0 is evaluated for all decohering noise values and each target ε. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$.

Figure 13: Final precision versus total number of oracle queries used for IAE, MLAE and SAE with decohering noise characterised by $\lambda = 10^{-3}$. We target precisions of $\varepsilon = 10^{-3}, 10^{-4}, \ldots, 10^{-7}$, with each point a single value of $\theta_0 \in \Theta_0$. We sample uniformly from Θ_0 for 30 values of θ_0. Each value of θ_0 is evaluated for all algorithms and each target ε. The prior for each iteration is taken to be $N(\theta_0, 1)$ and success probability $1 - \alpha$ with $\alpha = 0.01$.

4 Discussion

This is for internal discussions not external release Topics to include here include:

- Directions for future including reinforcement learning
- Superconducting and trapped ion thought experiment
- Discussion about general dynamic programming approach for considering error mitigation and correction
- Connection to QPE and applicability of thinking to that setting (maybe include small section on the more complicated statistical challenge of general state)
- Circular distributions and what we know about them (probably subsection of this section or algorithm)

References

[1] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. "Quantum amplitude amplification and estimation". Contemporary Mathematics **305**, 53–74 (2002).

[2] A Yu Kitaev. "Quantum measurements and the abelian stabilizer problem" (1995). arXiv:quant-ph/9511026.

[3] Don Coppersmith. "An approximate fourier transform useful in quantum factoring" (2002). arXiv:quant-ph/0201067.

[4] Scott Aaronson and Patrick Rall. "Quantum approximate counting, simplified". Pages 24–32. Society for Industrial and Applied Mathematics. (2020). arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611976014.5.

[5] Yohichi Suzuki, Shumpei Uno, Rudy Raymond, Tomoki Tanaka, Tamiya Onodera, and Naoki Yamamoto. "Amplitude estimation without phase estimation". Quantum Information Processing **19**, 1–17 (2020).

[6] Chu-Ryang Wie. "Simpler quantum counting" (2019). arXiv:1907.08119.

[7] Paul Dagum, Richard Karp, Michael Luby, and Sheldon Ross. "An optimal algorithm for monte carlo estimation". SIAM Journal on computing **29**, 1484–1496 (2000). arXiv:https://doi.org/10.1137/S0097539797315306.

[8] Stefan Heinrich. "Quantum summation with an application to integration". Journal of Complexity **18**, 1–50 (2002).

[9] Ashley Montanaro. "Quantum speedup of monte carlo methods". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471**, 20150301 (2015). arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2015.0301.

[10] Dmitry Grinko, Julien Gacon, Christa Zoufal, and Stefan Woerner. "Iterative quantum amplitude estimation". npj Quantum Information **7**, 52 (2021). arXiv:https://doi.org/10.1038/s41534-021-00379-1.

[11] Tudor Giurgica-Tiron, Iordanis Kerenidis, Farrokh Labib, Anupam Prakash, and William Zeng. "Low depth algorithms for quantum amplitude estimation". Quantum **6**, 745 (2022). arXiv:https://doi.org/10.22331/q-2022-06-27-745.

[12] Joseph G. Smith, Crispin H. W. Barnes, and David R. M. Arvidsson-Shukur. "An adaptive bayesian quantum algorithm for phase estimation" (2023). arXiv:2303.01517.

[13] Qiskit contributors. "Qiskit: An open-source framework for quantum computing" (2023).

[14] Adam Callison and Dan E. Browne. "Improved maximum-likelihood quantum amplitude estimation" (2022). arXiv:2209.03321.

A Technical details

Details of technical work if needed