AFRL-AFOSR-VA-TR-2022-0116

(Quantum Accelerator) Coreset Quantum Computing: Addressing Large Data Sets with Small Quantum Computers

FREDERIC CHONG
UNIVERSITY OF CHICAGO THE
5801 S ELLIS AVE
CHICAGO, IL,
US

02/24/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
20220224	Final	**START DATE** 20201115	**END DATE** 20211114

4. TITLE AND SUBTITLE
(Quantum Accelerator) Coreset Quantum Computing: Addressing Large Data Sets with Small Quantum Computers

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
	FA9550-21-1-0033	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Frederic Chong

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
UNIVERSITY OF CHICAGO THE 5801 S ELLIS AVE CHICAGO, IL US	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR RTB1	AFRL-AFOSR-VA-TR-2022-0116

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Quantum computers offer a powerful new approach to tackling computational challenges that are otherwise intractable. However, the current state of hardware has severe constraints on the quality and quantity of quantum bits (qubits). In this near-term regime, where quantum computers have just tens of noisy qubits, how can we aim to address optimization problems involving thousands or millions of variables? This fundamental research question aligns with the Innovate Quantum U Tech Challenge's call for "strategies for performing large computations on limited qubit machines, including gate and problem decompositions".
Our research responds to this call by exploring the connection between quantum computers and coresets. A coreset is a compact summary of a large classical data set, such that solving a problem on the coreset is equivalent to solving a problem on the full data set, up to an Ïµ error guarantee. Recently, Harrow proposed in [Har20] to use coresets so that a quantum computer with a small number of qubits, N, can address a classical data set of size M >> N. However, the algorithms studied in [Har20] are untenable on noisy quantum computers.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	9
U	U	U		

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
GRACE METCALFE	(703) 696-9740

AFOSR Final Report

Principal Investigator:
Fred Chong
Professor
Department of Computer Science
University of Chicago
Chicago, IL 60637
chong@cs.uchicago.edu

Project Title:
[Quantum Accelerator] Coreset Quantum Computing: Addressing Large Data Sets with Small Quantum Computers

Grant Number:
FA9550-21-1-0033

Period of Performance:
November 2020 to November 2021

Report Date:
February 2022

Report Documentation Page (Form)

1. Introduction

Quantum computers offer a powerful new approach to tackling computational challenges that are otherwise intractable. However, the current state of hardware has severe constraints on the quality and quantity of quantum bits (qubits). In this near-term regime, where quantum computers have just tens of noisy qubits, how can we aim to address optimization problems involving thousands or millions of variables? This fundamental research question aligns with the Innovate Quantum U Tech Challenge's call for "strategies for performing large computations on limited qubit machines, including gate and problem decompositions".

Our research responds to this call by exploring the connection between quantum computers and *coresets*. A coreset is a compact summary of a large classical data set, such that solving a problem on the coreset is equivalent to solving a problem on the full data set, up to an ϵ error guarantee. Recently, Harrow proposed in [Har20] to use coresets so that a quantum computer with a small number of qubits, N, can address a classical data set of size M >> N. However, the algorithms studied in [Har20] are untenable on noisy quantum computers. For example, Grover search does not have a quantum speedup in the presence of noise [RS08].

We instead focus on techniques that could enable quantum computers to address coresets in the near-term, when operations on qubits are noisy. Successful execution of these ideas would extend the reach of near-term quantum computers to enable them to address clustering problems on data sets with sizes greatly exceeding the number of qubits. This goal is compatible with the DOD and Air Force's interest in "Artificial Intelligence for Command and Control of Air Power" [V19]. For example, prior examples of clustering for aircraft identification [BSB+17] consider data sets with 5,000 points—far out of range of < 100 qubit computers unless we exploit coresets.

2. Summary of Accomplishments

2.1 Refinement of Prior Quantum + Coreset Work

Our initial effort pertained to refining, extending, and polishing our own recent prior work in [TGA+20], previously an exploratory arXiv preprint. Our prior work recast the techniques in [Har20] to be compatible with near-term quantum computers. Specifically, we studied the k-means clustering problem, which aims to partition a large classical data set into k clusters. We demonstrated how clustering on a weighted coreset can be cast as a Hamiltonian minimization problem, which in turn can be addressed by the Quantum Approximate Optimization Algorithm (QAOA) [FGG14]. It is worth emphasizing a few aspects of our prior work, which make it particularly realistic and suitable for near-term hardware. First, we optimized the hardware realization of our QAOA instantiation by mapping with a swap network [5, 6]. This enabled us to execute our algorithm experimentally on a 5-qubit computer from IBM with linear connectivity, with no loss in overhead due to SWAPs. Our experimental results were competitive with state-of-the-art classical techniques Second, our work demonstrates how Taylor expansions in the Hamiltonian construction can be used to break assumptions about cluster sizes.

Finally, unlike prior QAOA proposals which can be classically simulated due to entanglement bounds, our QAOA instance solves MAXCUT on a complete graph and therefore spreads pairwise entanglement across qubits within a single QAOA layer. Thus, the potential for quantum speedup is intact at low depth.

Our AFOSR funding enabled us to continue work on [TGA+20], which has culminated in its publication in the *Electronics* peer-reviewed journal. Among key funded contributions, we refined the underlying mathematical model for converting coreset k-means problems to PUBO (Polynomial Unconstrained Binary Optimization) Hamiltonian formulations. In addition, we improved our data analysis and visualizations to better understand the settings in which coreset-equipped quantum computers can outperform classical computers. In particular, we find that the coreset + quantum combination is most likely to outperform classical-only approaches for anomaly-detection oriented datasets with unbalanced cluster sizes. Finally, we studied protocols for generating coresets interactively, with the quantum computer playing an active role in the generation of coresets. While these interactive protocols did not outperform static classically-generated coresets, it provided a rich framework for ongoing work. In particular, the ability of quantum algorithms such as QAOA to sample from low-temperature probability distributions is valuable.

2.2 Ongoing Exploration of Adaptive Coresets.

Our recent work has entailed further exploration of adaptive coresets, where construction of the coreset is done in an iterative manner, querying the problem to be solved at each step. We looked at interfacing iterative Bayesian coreset algorithms with our quantum formulation of the 2-means problem. In this paradigm, the coreset algorithm queries our quantum computer for a model that solves the current 2-means problem. In effect this discretizes the continuous model space and enables the coreset algorithm to calculate cheap log-likelihoods of the original data points with the sampled models. Treating each data point's log-likelihoods as a vector, the Bayesian coreset algrithm can make a geometrically informed decision on which data point to add to the coreset and how to update weights. For example, Greedy Iterative Geodesic Ascent (GIGA), considers these vectors under geodesic alignment criterion and then samples a new point greedily [CB18]. In these algorithms, the quantum computer is enabling feedback each iteration for the coreset in the context of the specific problem being solved. We find that a quantum computer solving 2-means instances using QAOA as in our prior work, coupled with an adaptive coreset algorithm like GIGA, enables generation of near-optimal coresets of size $N \ll M$, greatly reducing the number of qubits needed to solve 2-means compared to the original dataset of size M.

We also extend our prior coreset work by exploring coresets in machine learning (ML) algorithms beyond k-means clustering. Specifically, we're looking at Quantum Boltzmann Machines (QBMs). Believed to be an ideal candidate for quantum advantage, QBMs can be used ubiquitously in both generative and discriminative machine learning settings [AAR+18].

QBMs require creating a model that learns on a large dataset and either generates or classifies future data points, depending on the use-case. This learning process is done iteratively, where the model's parameters are tuned in a direction that minimizes some loss function describing the current error. Efficiently calculating this loss function and updating the model is the crux for lots of popular ML algorithms, both in the classical domain and the quantum domain. For QBMs, which require exponential space on a classical computer, this step has been prohibitive for their usage. Even on a quantum computer, the loss function in the case of QBMs requires sampling from thermal states, which although explored on current noisy quantum computers, is still lacking in efficiency and accuracy. In fact, preparing these thermal states ends up being the main bottleneck for efficiently training QBMs, so any technique that can minimize the number of thermal states that needs to be prepared could enable QBMs to be trained on large enough datasets for a potential quantum advantage. This is the precise use case we're exploring for coresets. In most QBM approaches, the number of thermal states that needs to be prepared scales with the size of the dataset. By succinctly describing the classical dataset of size M with a weighted coreset of size N << M, we can expect the number of thermal states needed to decrease by the fraction N/M. We find that this application of coresets is compatible with both quantum and classical strategies for training QBMs. Figure 1 shows a high-level overview of the interaction between the coreset algorithm, the QBM training process, and thermal state sampling on a quantum device.

In the classical case, exact training of a QBM is exponentially difficult, and so approximate methods to sample from thermal states are employed instead. A common classical approximation that we use requires mapping our d-dimensional quantum system to a (d+1)-dimensional classical system, known as the Trotter-Suzuki mapping [S76]. Using this mapping, we can approximately describe the state of our quantum QBM system with a classical system with typically ~10 times as many units. In this classical system, we can calculate the energy of a state in polynomial time, circumventing the exponential requirement associated with the quantum system at the cost of accuracy. We then emulate prior work in classically training QBMs and use this system in population annealing with path integral Monte Carlo updates. We maintain a population of replicas of the system and then iteratively resample the population according to the relative classical energies, called the Boltzmann weights, of replicas. Each iteration also entails a path integral Monte Carlo update of the replicas, where replicas are probabilistically transformed into new states, again weighted by the relative classical energies. This algorithm serves the purpose of approximately sampling from our classical thermal state. Performing this iterative approximation, although not exponentially difficult, still takes notable time in practice and is the main contributor to the training time of QBMs in the classical case as it needs to be done for each thermal state. As such, using a well-constructed coreset allows us to reduce the number of times the algorithm is run, decreasing the overall training time by the same fraction, N/M.

In the quantum case, recent research has largely explored variational approaches to preparing and sampling from thermal states both inside and outside the context of QBMs. Using

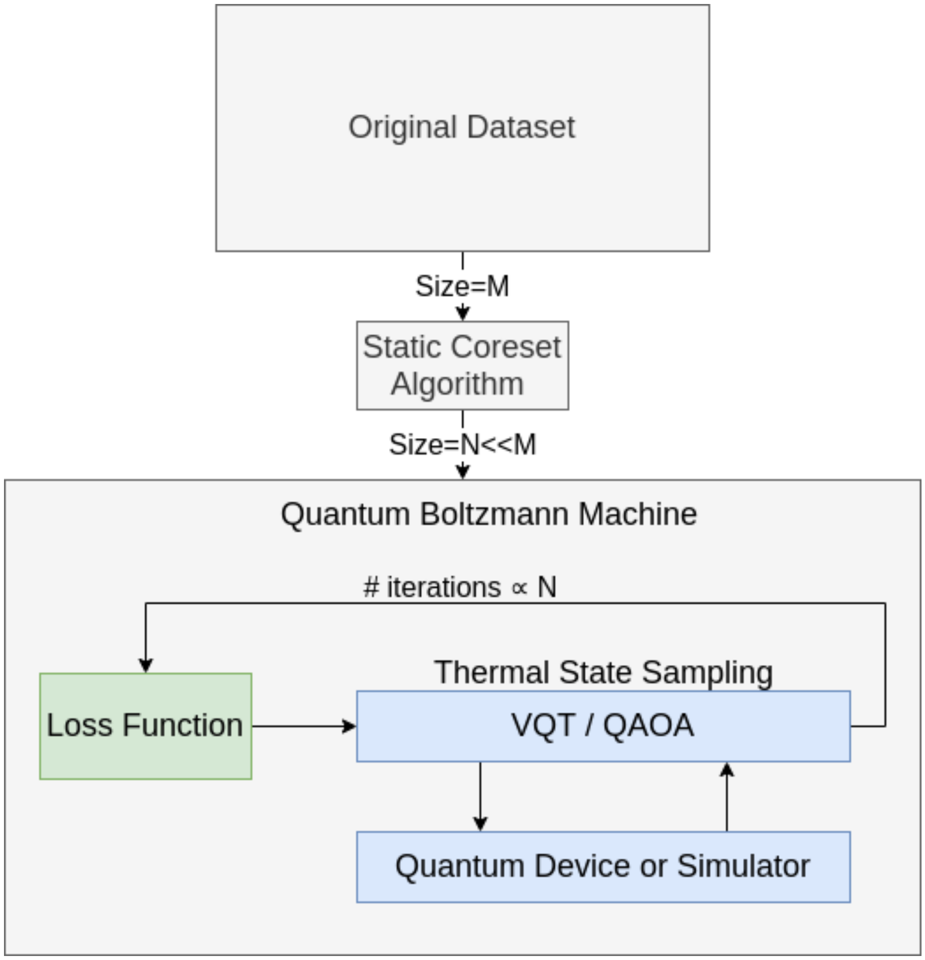

Figure 1. Overview of QBM training aided by coresets and using VQT/QAOA for sampling from thermal states.

QAOA, it's been theorized that at depth p=1, which is tractable on current machines, thermal states with Gaussian perturbations, named pseudo-Boltzmann states, can be generated and sampled from [DPG22]. Other variational approaches include the Variational Quantum Thermalizer (VQT) algorithm [VMN+19]. A visualization of a thermal state prepared exactly, with VQT, and through the Trotter-Suzuki mapped classical system is shown in Figure 2. Our ongoing work is looking at how these approaches can be applied to training QBMs with coresets on actual quantum devices.

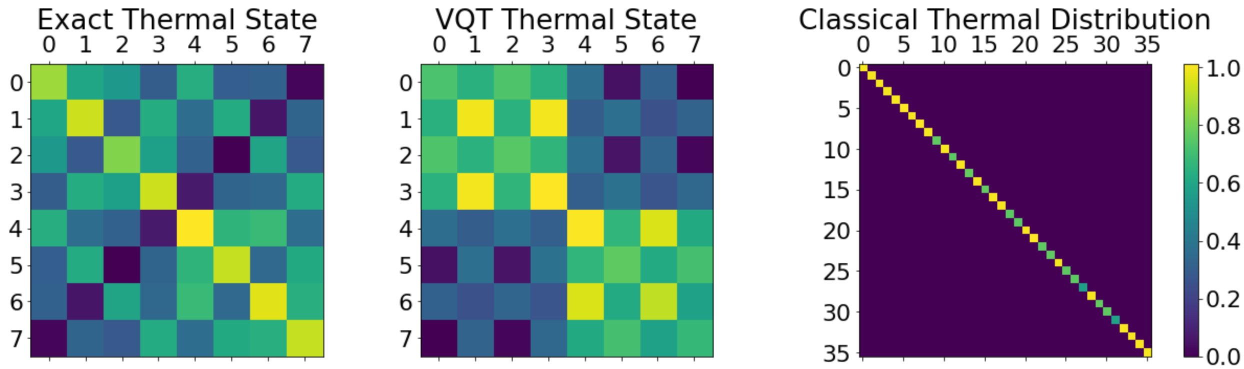

Figure 2. Thermal state/distribution prepared exactly, with VQT, and from the Trotter-Suzuki mapped classical system. This is for a small Restricted Quantum Boltzmann Machine with randomized initial weights and all transverse field parameters, Γ = 2.

3. Summary and Discussion

The application of quantum computers to real-world problems is of great interest, but faces barriers due to mismatches between the size of data sets and the size (in number of qubits) of near-term quantum computers. Our work demonstrates that the use of *coresets* can help bridge this gap.

The key results of our AFOSR funded research were (a) the refinement of our prior quantum + coreset work and (b) our ongoing exploration of adaptive coresets. The former led to a publication in the *Electronics* peer-reviewed journal and the latter is currently under preparation as an anticipated conference paper.

4. List of Publications

Published in peer-reviewed journal:

- Tomesh, Teague, Pranav Gokhale, Eric R. Anschuetz, and Frederic T. Chong. "Coreset clustering on small quantum computers." *Electronics* 10, no. 14 (2021): 1690.

Currently under preparation:
- Viszlai, Joshua, Teague Tomesh, Pranav Gokhale, Eric R. Anschuetz, and Frederic T. Chong. "Variational training of Quantum Boltzmann Machines aided by coresets (tentative title)." (2022).

5. Personnel Supported

Joshua Viszlai was supported for a portion of his doctoral work on this grant. Professor Fred Chong carried out work on this project, but was supported from other sources.
At the subcontractor (Super.tech Labs Inc.), Dr. Pranav Gokhale was supported by the subcontract for this grant.

References

[AAR+18] Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B., & Melko, R. (2018). Quantum boltzmann machine. Physical Review X, 8(2), 021050.

[BSB+17] Blanks, Z., Sedgwick, A., Bone, B., & Mayerchak, A. (2017, April). Identification of flight maneuvers and aircraft types utilizing unsupervised learning with big data. In 2017 Systems and Information Engineering Design Symposium (SIEDS) (pp. 180-185). IEEE.

[CB18] Campbell, T., & Broderick, T. (2018, July). Bayesian coreset construction via greedy iterative geodesic ascent. In International Conference on Machine Learning (pp. 698-706). PMLR.

[CLW20] Chowdhury, A. N., Low, G. H., & Wiebe, N. (2020). A variational quantum algorithm for preparing quantum Gibbs states. arXiv preprint arXiv:2002.00055.

[DPG22] Díez-Valle, P., Porras, D., & García-Ripoll, J. J. (2022). QAOA pseudo-Boltzmann states. arXiv preprint arXiv:2201.03358.

[FGG14] Farhi, E., Goldstone, J., & Gutmann, S. (2014). A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028.

[Har20] Harrow, A. W. (2020). Small quantum computers and large classical data sets. arXiv preprint arXiv:2004.00026.

[KMW+18] Kivlichan, I. D., McClean, J., Wiebe, N., Gidney, C., Aspuru-Guzik, A., Chan, G. K. L., & Babbush, R. (2018). Quantum simulation of electronic structure with linear depth and connectivity. Physical review letters, 120(11), 110501.

[OHR+19] O'Gorman, B., Huggins, W. J., Rieffel, E. G., & Whaley, K. B. (2019). Generalized swap networks for near-term quantum computing. arXiv preprint arXiv:1905.05118.

[RS08] Regev, O., & Schiff, L. (2008, July). Impossibility of a quantum speed-up with a faulty oracle. In International Colloquium on Automata, Languages, and Programming (pp. 773-781). Springer, Berlin, Heidelberg.

[S76] Suzuki, M. (1976). Relationship between d-dimensional quantal spin systems and (d+ 1)-dimensional ising systems: Equivalence, critical exponents and systematic approximants of the partition function and spin correlations. Progress of theoretical physics, 56(5), 1454-1469.

[TGA+20] Tomesh, T., Gokhale, P., Anschuetz, E. R., & Chong, F. T. (2020). Coreset Clustering on Small Quantum Computers. *arXiv preprint arXiv:2004.14970*.

[VMN+19] Verdon, G., Marks, J., Nanda, S., Leichenauer, S., & Hidary, J. (2019). Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. arXiv preprint arXiv:1910.02071.

[V19] Voke, M. R. (2019). Artificial Intelligence for Command and Control of Air Power. Air Command and Staff College.