Particle Track Classification Using Quantum Associative Memory

Final Technical Report DE-SC0019497

Lauren Ice, Gregory Quiroz, Andrea Delgado, Travis Humble

December 21, 2020

Overview

This project explored the use of quantum-assisted algorithms for pattern matching in sub-atomic physics experiments. Pattern matching algorithms are commonly employed to prune data of random noise and to help discriminate between signals generated by particle tracks of interest and signals generated by background events. The quantum-assisted algorithms explored in this project were based on an Ising formulation of quantum associative model (QAMM) recall and quantum content-addressable memory (QCAM) recall. The recall is performed by comparing a probe pattern with those stored in a library of patterns encoded in the QAMM/QCAM model.

The classification accuracy of QAMM and QCAM recall was determined as a function of detector resolution, noise, and efficiency and pattern density, where pattern density is defined as the ratio of the number of reference signal patterns encoded in the library to each pattern's length. We found that QAMM achieved high classification accuracy when applied to datasets with low pattern density. QCAM achieved high classification accuracy for datasets with high pattern density and was found to be more robust to detector noise. The project methodology and results are described in detail in our arXiv preprint (arXiv:2011.11848)[1].

This project was conducted by scientists at the Johns Hopkins University Applied Physics Laboratory and Oak Ridge National Laboratory from August 2018 to August 2020 and was supported by DOE grant DE-SC0019497.

Background

Track recognition and classificaiton is a challenging and necessary part of most subatomic physics experiments. Track reconstruction is the process of mapping detector signals produced by a particle traversing a detector back into the particle's trajectory. This process is complicated by random noise, detectors with high spatial resolution or low efficiency, and experiments with many simultaneous particles traversing the detector region. Pattern matching algorithms are commonly employed to prune data of random noise, which simplifies the track reconstruction task. Pattern matching algorithms quickly identify potential track candidates by comparing the pattern of detector signals to a library of patterns known to be from events of interest. However, the performance of these algorithms is sensitive to the amount of random detector noise, background processes, and the detector's spatial resolution.

Pattern matching for track recognition can be accomplished through several methods, two of which are associative memory model (AMM) recall and content-addressable memory (CAM) recall. AMM is a

[1] Gregory Quiroz, Lauren Ice, Andrea Delgado, and Travis S. Humble. Particle Track Classification Using Quantum Associative Memory. *arXiv preprint arXiv:2011.11848*, 2020

supervised learning model for pattern matching when given incomplete or incorrect information.[2] CAM is a form of associative memory that stores the association between a key and a value and recalls this association by retrieving the key when given the exact or approximate value.[3] It has been shown that AMM[4] and CAM[5] can be solved using quantum annealing (QA), a probabilistic algorithm that exploits quantum phenomena to find the global minimum to an objective function.[6] This quantum AMM (QAMM) and quantum CAM (QCAM) approach has the potential to provide improvements to the learning capacities of the models, and ultimately lead to increased performance in pattern classification algorithms for track reconstruction. While currently available QA processors are limited in size, devices such as the D-Wave 2000Q-6 processor provides a useful testbed of evaluating toy models and experimentally investigating recall methodologies for track classification and pattern matching.

Project goals

This aim of this project was to evaluate QAMM and QCAM recall performance when used to perform pattern matching for particle track classification in subatomic physics experiments.

This goal was split into three main tasks. The first task was to generate a dataset of patterns based on a detector simulation. Using a simulated dataset that could be easily modified allowed for the classification accuracy of QAMM and QCAM recall to be determined as a function of parameters such as detector granularity, efficiency, and the level of detector noise. The second task was to develop, optimize, and characterize the QCAMM and QCAM recall algorithms' classification accuracy. The QAMM and QCAM performances were optimized for each simulated dataset by selecting the learning rule used to train the models, determining the optimal classifier or combination of classifiers for QAMM and QCAM, and optimizing the annealing schedule of the D-Wave QPU. The third task was to compare QAMM and QCAM to the classical approach using experimental data.

The ultimate goal was to compare the classification accuracy using QAMM and QCAM to the classical tree search method to determine potential improvements in the rejection of noise signals, computational speed, and track reconstruction. This benchmark was to be achieved by applying QAMM and QCAM recalls

---

[2] John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences.* 79(8):2554-2558, 1982

[3] Hopfield, 1982

[4] Hadayat Seddiqi and Travis S Humble. Adiabatic quantum optimization for associative memory recall. *Frontiers in Physics*, 2:79, 2014; Siddhartha Santra, Omar Shehab, and Radhakrishnan Balu. Ising formulation of associative memory models and quantum annealing recall. *Physical Review A*, 96(6):062330, 2017.

[5] Jonathan Schrock, Alex J McCaskey, Kathleen E Hamilton, Travis S Humble, and Neena Imam. Recall performance for content-addressable memory using adiabatic quantum optimization. *Entropy*, 19(9):500, 2017.

[6] Giuseppe E. Santoro, Roman Martonak, Erio Tosatti, and Roberto Car. Theory of quantum annealing of an ising spin glass. *Science*, 295(5564):2427–2430, 2002; M. W. Johnson and et al. Quantum annealing with manufactured spins. *Nature*, 473:194–198, 2011; Catherine C. McGeoch. Adiabatic quantum computation and quantum annealing: Theory and practice. *Synthesis Lectures on Quantum Computing*, 5(2):1–93, 2014; Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Rev. Mod. Phys.*, 90:015002, Jan 2018.

models to an experimental dataset from the OLYMPUS experiment[7] and comparing the performance to that of a classical tree search method.

Accomplishments

Task 1: Generate datasets of patterns based on a detector simulation

The detector simulation was done using the Geant4 simulation and modeling toolkit.[8] It consisted of three parallel detector planes that could be segmented into individual detection segments to control the detector granularity, and correspondingly, the number of bits in each pattern. A diagram of the simulated detector is shown in Figure 1.

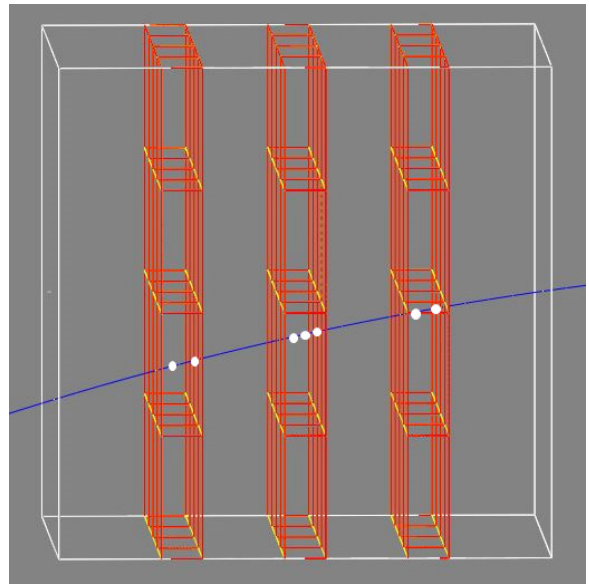

Figure 1: A diagram of the simulated detector with three parallel planes. The blue line represents a track of an electron traversing the detector.

The detector segments were mapped to an array of Boolean elements of size V. The detector shown has 16 segments per plane (V = 48 segments in total), which creates a Boolean pattern of 48 bits.

The created patterns that corresponded to particle trajectories, electrons and positrons were initialized left of the detector and traversed the detector region interacting with each detector plane. If a given cell was activated, the corresponding bit in the array of Boolean elements was set to 1, and 0 if it was not. Then, various random detector noise rates, η, and detector inefficiency, γ, were included by flipping some of the Boolean elements. Background patterns were also created by setting the bits in the array of Boolean elements randomly to either 0 or 1.

For this project, we explored detector granularity of 24-segments, 30-segments, 42-segments, 48-segments, and 54-segments. The random detector noise levels considered were $\gamma = [0.\ 0.02, 0.04, 0.06, 0.08]$ and the detector efficiencies considered were $\eta = [1, 0.98, 0.96, 0.94, 0.92]$.

Task 2: Characterization QCAMM and QCAM Recall Algorithms

The classification accuracy for QAMM and QCAM recall was determined as a function of γ, η and pattern density α_s, defined as the ratio of the number of signal patterns encoded into the library to the number of bits in the pattern, V. First, we explored different approaches to defining a classifier for QAMM and QCAM. Second, the classification accuracy was determined for noiseless patterns with perfect detector efficiency for each classification method and both QAMM and QCAM recall. Next, the classification accuracy was determined for patterns that contained noise signals and inefficiencies. Last, the QA control

---

[7] Henderson, B. S., L. D. Ice, D. Khaneft, C. O'Connor, R. Russell, A. Schmidt, J. C. Bernauer et al. "Hard two-photon contribution to elastic lepton-proton scattering determined by the OLYMPUS experiment." *Physical Review Letters* 118, no. 9 (2017): 092501.

[8] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G 2 Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* 506(3):250–303, 2003.

schedules were manipulated through a method called reverse annealing to increase the classification accuracy for noisy patterns.

Determination of classifiers

More than one classifier can be used for QCAM and QAMM recall to distinguish between signal and background patterns. Two approaches were explored here. The first classifier that can be used is the key of the recalled pattern. This requires the datasets to include both signal and background patterns, labeled by keys. The second classifier is the energy of the solution states resulting from QA. Energy-based classification requires only signal patterns to be encoded. For each probe pattern recalled, 100 annealing runs are carried out. From this, the average energy $\langle E \rangle$ and key $\langle k \rangle$ can be calculated for each probe pattern.

We define β as the tolerance of $\langle E \rangle$ and key $\langle k \rangle$ of a particular probe pattern to be different from that of the rest of the set. We allowed the tolerance β for both energy $\langle E \rangle$ and key $\langle k \rangle$ to range widely from $\beta = [0, 10]$ and display the classification accuracies using receiver operator characteristic (ROC) curves. This allows for the true positive rate (TPR) and the false positive rate (FPR) to be shown on a single chart, as a function of β.

Noiseless pattern recall

We first examine QCAM and QAMM recall performance for both classifiers under the condition of noiseless pattern recall. Noiseless signal probe patterns are defined as signal patterns that perfectly match patterns encoded within the model. Noiseless background probe patterns are defined similarly with the caveat that background probes will only perfectly match patterns within the training set when performing key-based classification. We summarize the noiseless recall results in Figure 2 for varying detector size and encoded signal pattern density. ROC curves for energy-based classification for QCAM and QAMM are shown in panels (a)-(d) and (e)-(h), respectively. Panels (i)-(l) denote key-based QCAM performance, while panels (m)-(p) convey QAMM recall using key-based classification. Each column contains a fixed encoded pattern density, where each panel displays ROC curves for a variety of detector sizes. The data is generated from 5 unique training sets and 50 unique probe patterns (25 signal and 25 background). ROC curves are produced by varying β over the range discussed above. Each data set includes a logistic fit to help guide the eye. Note that ROC curves favoring each subplot's top-left corner denote optimal discriminators, while those tending towards the diagonal, black, dotted line designate models with the least discriminatory capabilities. In certain cases, for example, panel (o), the data drops below the diagonal, denoting reciprocated class discrimination. This behavior is likely due to the presence of spurious memories, which is elaborated upon in arXiv:2011.11848.

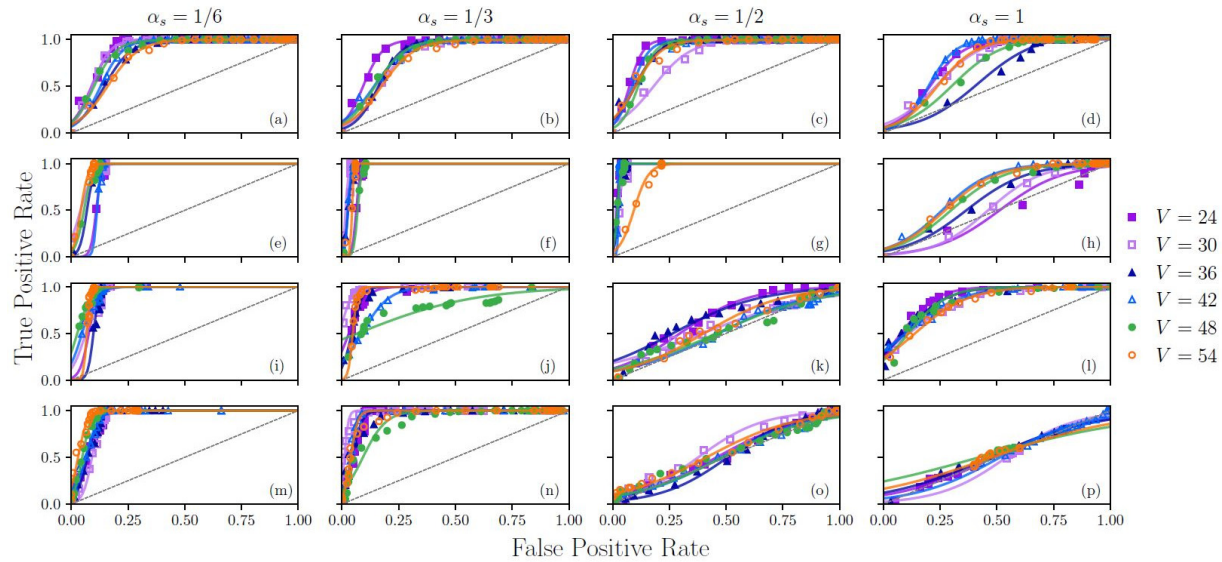

Figure 2: Comparison between recall models as a function encoded signal pattern density α_s for varying detector size V. Models considered include: energy based QCAM [panels (a)-(d)], energy based QAMM [panels (e)-(h)], key based QCAM [panels (i)-(l)], and key-based QAMM [panels (m)-(p)]. Data is averaged over 5 unique training sets and 50 unique probes (25 signal and 25 background patterns).

As these results indicate, the energy-based QAMM (panels (e)-(h)] show the best performance for $\alpha_s < 1$ and the classification accuracy appears to be relatively robust as V increases. For high pattern density ($\alpha_s = 1$) key-based QCAM provides the highest classification accuracy.

Faulty pattern recall

Next, the classification accuracy as a function of inefficiencies and noise was explored. Here, faulty probe patterns with either missing signal bits (due to detector inefficiency) or additional signal bits (due to detector noise) are recalled. Because the energy-based QAMM and key-based QCAM performed the best with the noiseless data, these recall methods were used to explore how inefficiencies and noise impact the classification accuracy.

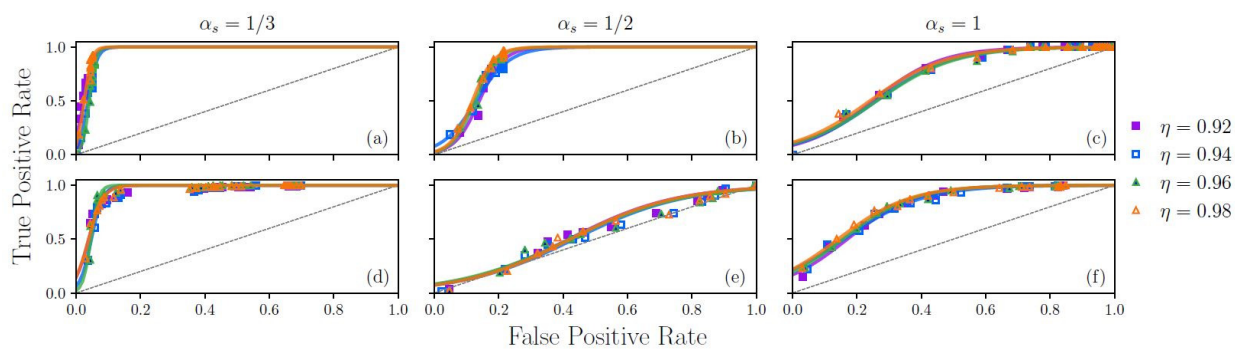

Figure 3: Comparison between recall models as a function encoded signal pattern density α_s for varying detector efficiency η and V = 54. Models considered include: energy based QAAM [panels (a)-(c)] and key based QCAM [panels (d)-(f)].

Figure 3 shows the ROC curves for energy-based QAMM [panels (a)-(c)] and key-based QCAM [panels (d)-(f)] for recalling patterns with missing signals due to detector inefficiencies. This plot only shows the data

for $V = 54$. As this reveals, the classification accuracy does not have a strong dependence on detector inefficiencies.

While the classification accuracy for energy-based QAMM and key-based QCAM appeared to be robust to detector inefficiencies, the classification accuracy was strongly impacted when recalling noisy patterns.

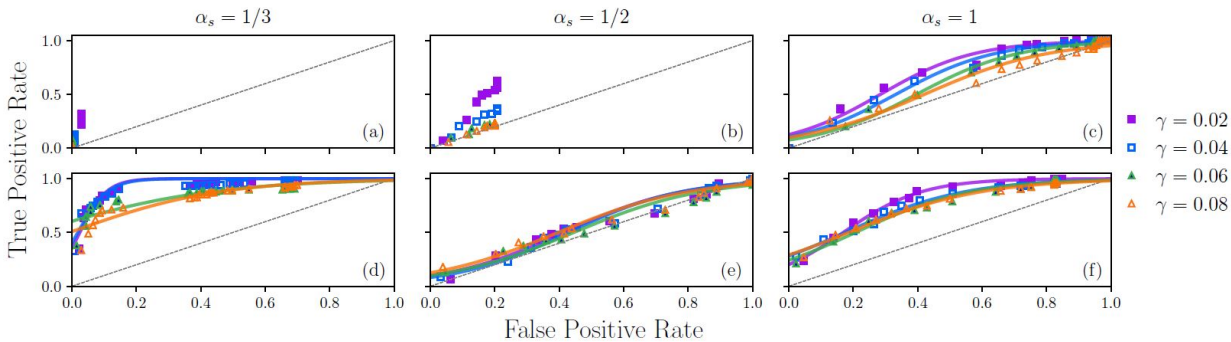

Figure 4: Comparison between recall models as a function encoded signal pattern density α_s for varying detector noise γ and V = 54. Models considered include: energy based QAAM [panels (a)-(c)] and key based QCAM [panels (d)-(f)].

Figure 4 shows the results for noisy data. Again, figures are for $V = 54$. As shown, the amount of noise in the dataset significantly decreases the classification performance, especially for energy-based QAMM. However, as shown in panel (d), Key-based QCAM appears to be slightly robust to noisy datasets for small α_s. This implies that key-based QCAM might be a better classifier for noisy experimental data.

Reverse annealing

In efforts to improve the noisy pattern recall performance, some of the D-Wave's QPU advanced control features were exploited. This effort focused improvements to the key-based QCAM approach as this showed the best performance on noisy data. By manipulating the QA control schedule through a process known as reverse annealing, an increase the TPR was observed with only minor increase to FPR for low pattern density. Figure 5 shows this for $V = 54$, where the control path was chosen such that the quantum system was initialized in a candidate classical state provided by a forward annealing optimization run. The system is then driven in "reverse" by ramping the control field such that the contribution to the dynamics that drive transitions in the system becomes more dominant. This process allows quantum fluctuations to increase and contribute to the dynamics over a fixed time duration. After a selected period of time has expired, the system is driven according to the standard forward anneal and subsequently measured. The choice of the reverse annealing profile was chosen by preliminary studies of recall performance as a function of ramp duration and pause time; see arXiv:2011.11848 for further details.

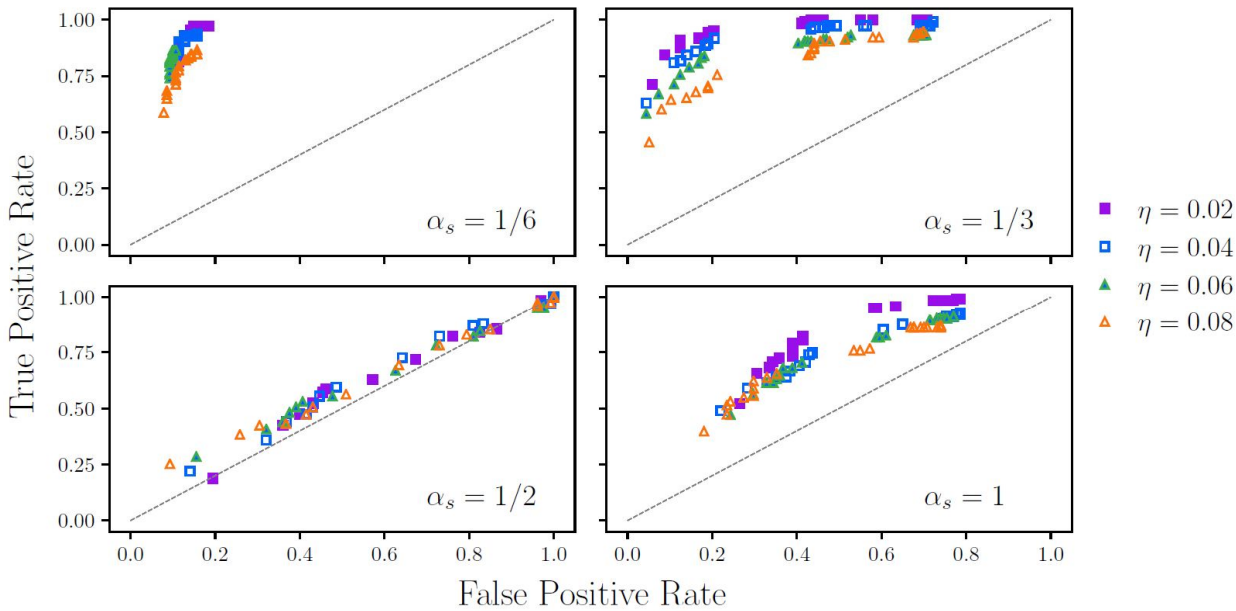

Figure 5: Key-based QCAM recall performance as a function of detector noise with reverse annealing control.

This report only summarizes the achievements and results. Detailed discussion of these results can be found in arXiv:2011.11848. This paper includes an extensive discussion on the datasets analyzed and the methods utilized for this project including quantum annealing, the quantum associative memory model, the quantum content addressable memory model and projection rule learning. Additionally, the results are discussed at length.

Task 3: Comparison to classical methods

Current hardware significantly limits both the size of the patterns that can be recalled and the number of patterns stored in the library. Because of the current hardware limitations, we determined that QAMM and QCAM recall cannot be directly applied to pattern recognition for most experimental datasets. Before we could attempt to apply QAMM and QCAM recall to the OLYMPUS dataset, we needed to focus on fully characterizing QAMM and QCAM recall as a function of pattern size, number of patterns in the library, and the rate of faulty patterns (those with noise or inefficiencies). This helped identify which datasets QAMM and QCAM recall could provide high classification accuracy, especially as the hardware improves.

Although we did not accomplish the last goal of applying the recall methods to experimental data, we explored a methodology that could transform experimental datasets to datasets suitable for QAMM and QCAM recall given the current hardware constraints. As discussed in arXiv:2011.11848, a Hough-Transform (HT) can be used to subdivide a large experimental dataset into smaller datasets, according to pattern features, such as track length and initial angle. The robustness of the HT method was explored using the Geant4-simulated detector with much higher detector resolution (360 segments) and the same noise and detector efficiencies as explored for QAMM and QCAM recall. Using the HT method, large experimental pattern libraries can be split into multiple pattern banks of shorter patterns providing suitable pattern libraries for QAMM and QCAM recall.

Results dissemination

A detailed description of the background, methodology, and results of this project are publicly available on ArXiv (arXiv:2011.11848) and are currently being reviewed for publication.

Conference Presentations

1. "Particle Track Pattern Recognition via Content Addressable Memory and Adiabatic Quantum Optimizaiton", Department of Energy Quantum Information Science PI-Kickoff Meeting (2019)
2. "Particle Track Pattern Recognition via Content Addressable Memory and Adiabatic Quantum Optimization", Qubits North America (2019)
3. "Performance Improvements of a Quantum Annealer Using Optimized Quantum Control", APS March Meeting (2019)
4. "Particle Track Pattern Recognition via Quantum Associative Memory", Qubits 2020.
5. "Quantum Algorithms for Event Reconstruction and Simulation in High-Energy Physics", IEEE Quantum Week (2020)