Novel machine learning algorithms for quantum annealing with applications in high energy physics

Alexander Zlokapa
California Institute of Technology
Overview

Higgs boson classification (QAML-Z):

• Phrase error minimization in an Ising model
• Use multiple anneals to zoom into the energy surface
Overview

Higgs boson classification (QAML-Z):

- Phrase error minimization in an Ising model
- Use multiple anneals to zoom into the energy surface

Charged particle tracking:

- Adapt large-scale computations to NISQ hardware
- Match state-of-the-art classical tracking algorithms
QAML-Z: Higgs boson classification
QAML algorithm


“Quantum annealing for machine learning” (QAML)

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers
QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

$$\arg\min_{s_i} \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i c_i(x_\tau) \right|^2$$
QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers
QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

\[
\arg\min_{s_i} \sum_{\tau=1}^{S} (y_{\tau} - \sum_{i=1}^{N} s_i c_i(x_{\tau}))^2
\]

Training set \hspace{1cm} Training label \hspace{1cm} Weak classifier = ±1/N

Training input
QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

\[
\text{argmin}_{s_i} \sum_{\tau=1}^S y_{\tau} - \sum_{i=1}^N s_i c_i(x_{\tau}) = \pm \frac{1}{N}
\]
QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

\[ H_{\text{Ising}} = \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{\tau=1}^{S} s_i c_i(x_{\tau}) s_j c_j(x_{\tau}) - \sum_{i=1}^{N} \sum_{\tau=1}^{S} s_i c_i(x_{\tau}) y_{\tau} \]
Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?
Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?
Higgs problem construction

Higgs boson
Higgs problem construction

Higgs boson

Other Standard Model (SM) processes
Higgs problem construction

Eight kinematic observables assembled from decay photons:

\[ \frac{p_T^1}{m_{\gamma\gamma}}, \frac{p_T^2}{m_{\gamma\gamma}}, \frac{(p_T^1 + p_T^2)^2}{m_{\gamma\gamma}}, \frac{(p_T^1 - p_T^2)^2}{m_{\gamma\gamma}}, \frac{p_T^{\gamma\gamma}}{m_{\gamma\gamma}}, \Delta \eta, \Delta R, |\eta^{\gamma\gamma}| \]

Diphoton angle
Higgs problem construction

Thirty-six weak classifiers constructed from division and multiplication of eight observables
Higgs problem construction

Thirty-six weak classifiers constructed from division and multiplication of eight observables
Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events
Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events
Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events
QAML-Z algorithm


Two improvements:

- Zoom into the energy surface — continuous optimization
- Augment the set of classifiers — stronger ensemble
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals

Energy

QAML: take discrete values ±1
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals

QAML: take discrete values ±1
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals.

Energy

-1

μ₀ = 1

s₀ = 1

1

Classifier weight

Ising model spin

QAML: take discrete values ±1
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals

QAML-Z: search for weights in [-1, 1]
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals

\[ \mu_0(0) = 0 \]

\[ \mu_0(1) = 0.5 \]

\[ s_0 = 1 \]
QAML-Z algorithm: Zooming

Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals

\[
\mu_0(0) = 0 \\
\mu_0(1) = 0.5 \\
\mu_0(2) = 0.25 \\
s_0 = -1
\]
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut.
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut
QAML-Z algorithm: Augmentation

Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut
Higgs classification results

QAML-Z vs. QAML

Improves advantage over DNN by ~40% for small training sets

Shrinks disadvantage to DNN by ~50% for large training sets
Higgs classification results

**QAML-Z vs. QAML**

Improves advantage over DNN by ~40% for small training sets

Shrinks disadvantage to DNN by ~50% for large training sets
Higgs classification results

QAML-Z vs. QAML

Implements advantage over DNN by ~40% for small training sets

Shrinks disadvantage to DNN by ~50% for large training sets
Higgs classification results

Both zooming and augmentation improve performance
Higgs classification results

Both zooming and augmentation improve performance
Charged particle tracking
Track reconstruction

Cluster “hits” in a detector by particle instance
Track reconstruction

Cluster “hits” in a detector by particle instance
Track reconstruction

Cluster “hits” in a detector by particle instance
Track reconstruction

Cluster “hits” in a detector by particle instance

Track reconstruction

Cluster “hits” in a detector by particle instance

Classical methods

Upgrade of LHC to high luminosity increases the number of hits per event by a factor of 5

Current tracking (Kalman filter) is thought to scale exponentially with the number of hits

Possibility of quantum speedup?

Ising model formulation

Make each edge a binary variable: turn edge “on” or “off”
Ising model formulation


\[ H_1 = - \sum_i \sum_{j > i} J_{ij} s_i s_j - \sum_i h_i s_i \]

Affinity between edges \(i\) and \(j\)

1 if edge is on; 0 if edge is off

Prior expectation on edge \(i\)
Expect helical tracks due to a charged particle moving in a uniform magnetic field
Ising model formulation

Expect helical tracks due to a charged particle moving in a uniform magnetic field

\[-\left(\frac{\cos^2(\theta_{abc})}{r_{ab} + r_{bc}}\right) s_{ab} s_{bc}\]
Ising model formulation

\[ E = - \sum_{a,b,c} \left( \frac{\cos^4(\theta_{abc}) + \rho \cos^4(\phi_{abc})}{r_{ab} + r_{bc}} \right) s_{ab} s_{bc} + \eta \sum_{a,b,c} \left( \frac{z_c - z_c - z_a}{r_c - r_a} r_c \right) s_{ab} s_{bc} + \alpha \left( \sum_{b \neq c} s_{ab} s_{ac} + \sum_{a \neq c} s_{ab} s_{cb} \right) + \sum_{a,b} \left( \gamma - \beta P(s_{ab}) \right) s_{ab} \]

Helical tracks
High-momentum bias
Track bifurcation penalty
Global edge penalty
Beam spot geometry
Edge orientation probability
(Gaussian kernel density estimation)
Dimension challenge

Higgs event at LHC: $10^3$ to $10^4$ detector hits $\implies \sim 10^7$ edges

- Divide into 16 sectors: $\sim 10^5$ edges
- Remove edges with Gaussian KDE: $\sim 10^3$ edges
Dimension challenge

Higgs event at LHC: $10^3$ to $10^4$ detector hits $\implies \sim 10^7$ edges

- Divide into 16 sectors: $\sim 10^5$ edges
- Remove edges with Gaussian KDE: $\sim 10^3$ edges

D-Wave 2X: 33 fully-connected qubits

- Sparse Ising model weights: $\sim 10^2$ qubits
- Split into disjoint sub-graphs: $\sim 10$ problems per sector
Dimension challenge

Disjoint sub-graphs: prune and divide

Initial graph

True graph
Dimension challenge

Disjoint sub-graphs: prune and divide

Pruned graph

True graph
Dimension challenge

Disjoint sub-graphs: prune and divide

Disjoint sub-graphs

True graph
Dimension challenge

Higgs event at LHC: $10^3$ to $10^4$ detector hits $\implies \sim 10^7$ edges

- Divide into 16 sectors: $\sim 10^5$ edges
- Remove edges with Gaussian KDE: $\sim 10^3$ edges

D-Wave 2X: 33 fully-connected qubits

- Sparse Ising model weights: $\sim 10^2$ qubits
- Split into disjoint sub-graphs: $\sim 10$ problems per sector
Higgs event at LHC: $10^3$ to $10^4$ detector hits $\implies \sim 10^7$ edges

- Divide into 16 sectors: $\sim 10^5$ edges
- Remove edges with Gaussian KDE: $\sim 10^3$ edges

D-Wave 2X: 33 fully-connected qubits

- Sparse Ising model weights: $\sim 10^2$ qubits
- Split into disjoint sub-graphs: $\sim 10$ problems per sector

Result: $\sim 100$ Ising model variables on $\sim 100$ qubits
Results

Performance metrics: efficiency (recall) and purity (precision) measured on the TrackML dataset

\[
\text{efficiency} = \frac{\# \text{ true tracks reconstructed}}{\# \text{ true tracks}}
\]

\[
\text{purity} = \frac{\# \text{ true tracks reconstructed}}{\# \text{ tracks reconstructed}}
\]
Results

![Graph showing Efficiency vs. Particles/event for QA, Random, and SA.](image)
Results

Maximum efficiency (after pre-processing)

![Graph showing efficiency vs. Particles/event]

- QA
- Random
- SA
Results

Maximum efficiency (after pre-processing)

D-Wave 2X
Simulated annealing

Efficiency

Particles/event

QA Random SA
Results

Maximum efficiency (after pre-processing)

D-Wave 2X

Simulated annealing

Higgs in 2011-2012

Efficiency

Particles/event

QA  Random  SA
Results
Results

Efficiency

Purity
Results

**Efficiency**

- Track Efficiency: 97%

**Purity**

- Track Purity: 99%
Conclusion
Beyond HEP: What’s new in QML?

Substantial improvement demonstrated by QAML-Z

- Widespread applicability of successive anneals on iteratively refined problems
Beyond HEP: What’s new in QML?

Substantial improvement demonstrated by QAML-Z

- Widespread applicability of successive anneals on iteratively refined problems

Successful encoding of big data in the era of NISQ

- General methodology of pruning Ising models with a successful outcome
Beyond HEP: What’s new in QML?

Substantial improvement demonstrated by QAML-Z

- Widespread applicability of successive anneals on iteratively refined problems

Successful encoding of big data in the era of NISQ

- General methodology of pruning Ising models with a successful outcome

Competitive results with state-of-the-art classical algorithms
Thank you
Supplementary slides
Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?
Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?
Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?
QAML-Z algorithm: Hamiltonian

Zooming: replace each binary weight $s_i$ with continuous weight $\mu_i(t)$ governed with search breadth $\sigma(t) = 1/2^t$.

Augmentation: generate multiple shifted classifiers $c_{il}(x_\tau)$ for each original classifier $c_i(x_\tau)$.

Anneal for iterations $t = 0, 1, 2, \ldots, T - 1$. 
QAML-Z algorithm: Hamiltonian

Each iteration $t$, anneal:

$$H(t) = \sum_{l=-A}^{A} \left[ \sum_{i=1}^{N} \left( -C_{il} + \sum_{j>i}^{N} \mu_{jl}(t)C_{ijl} \right) \sigma(t)s_{il} + \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ijl}\sigma^2(t)s_{il}s_{jl} \right]$$

where we have defined:

$$C_{il} = \sum_{\tau=1}^{S} c_{il}(x_{\tau})y_{\tau} \quad C_{ijl} = \sum_{\tau=1}^{S} c_{il}(x_{\tau})c_{jl}(x_{\tau})$$
QAML-Z algorithm: Hamiltonian

Each iteration $t$, anneal:

$$H(t) = \sum_{l=-A}^{A} \left[ \sum_{i=1}^{N} \left( -C_{il} + \sum_{j>i}^{N} \mu_{jl}(t) C_{ijl} \right) \sigma(t) s_{il} + \sum_{i=1}^{N} \sum_{j>i}^{N} C_{ijl} \sigma^2(t) s_{il} s_{jl} \right]$$

and update continuous weights from spins:

$$\mu_{il}(t + 1) = \mu_{il}(t) + s_{il} \sigma(t + 1)$$
ROC curve

Metric of performance: area under receiver operating characteristic (ROC) curve
ROCC curve

Metric of performance: area under receiver operating characteristic (ROC) curve
ROC curve

Metric of performance: area under receiver operating characteristic (ROC) curve

- True positive rate (signal efficiency)
- False positive rate (background rejection)

Perfect classifier
Random classifier
ROC curve

Metric of performance: area under receiver operating characteristic (ROC) curve
Higgs classification results

Energy vs. Iteration number for Training set size 1000:
- Dashed line: Augmented classifiers, no zoom
- Solid line: QAML
- Blue line: QAML-Z

Dashes indicate test set, solid line indicates training set

AUROC vs. Iteration number for Training set size 1000:
- Red line: QAML
- Blue line: QAML-Z
Dataset

TrackML Challenge: top quark events with 15% noise

Bias towards high momentum tracks that are more important

\[- \left( \frac{\cos^2(\phi_{abc})}{r_{ab} + r_{bc}} \right) s_{ab}s_{bc} \]
Ising model formulation

Bias towards high momentum tracks that are more important

\[-\left( \frac{\cos^2(\phi_{abc})}{r_{ab} + r_{bc}} \right) s_{ab}s_{bc} \]
Ising model formulation

Tracks should point towards the beam spot at the origin

\[
\left( z_c - \frac{z_c - z_a}{r_c - r_a} \right) s_{ab}s_{bc}
\]
Ising model formulation

In general, tracks shouldn’t split at or merge into a single hit

\[ s_{ab}s_{ac} + s_{ab}s_{cb} \]
Use Gaussian kernel density estimation to provide a prior on an edge being “on” or “off” based on orientation and position.

\[(\gamma - P(s_{ab}))s_{ab}\]
Results

Ground state energy

True solution energy
Quantum speedup?

Inconclusive results for quantum speedup
Quantum speedup?

Inconclusive results for quantum speedup

- Classical: $O(\exp(\# \text{ of hits}))$
- Quantum: preprocess at $O\left((\# \text{ of hits})^2\right)$, inconclusive
  QUBO-solving time
Inconclusive results for quantum speedup

- Classical: $O(\exp(\# \text{ of hits}))$

- Quantum: preprocess at $O((\# \text{ of hits})^2)$, inconclusive QUBO-solving time

Could use specialized classical hardware for particle tracking at the trigger level: 1 PB/s reduced to 1 GB/s
Quantum speedup?

Pre-processing to construct the Ising model scales like $O(h^2)$ where an event has $h$ hits.
Quantum speedup?

Pre-processing reduces the simulated annealing solving time from $O(\exp(ch^2))$ where an event has $h$ hits.
Quantum speedup?

The problem remains NP-hard after pre-processing, so SA is exponential in the number of Ising model variables.

For an event divided into $K$ sub-graphs with $m_i$ edges each, we expect solving time

$$O\left( \sum_{i=1}^{K} \exp\left( cm_i \right) \right)$$
Quantum speedup?

Quantum annealing is expected to reduce the size of \( c \) but leave the problem exponential

\[
O\left( \sum_{i=1}^{K} \exp\left(c m_i\right) \right)
\]

Can’t measure a single time to solution, but can change annealing time and measure change in performance
Quantum speedup?