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

Alexander Zlokapa
California Institute of Technology

A. Anand, Harvard University
A. Mott, DeepMind
J. Job, Lockheed Martin
J.-R. Vlimant, Caltech
J. M. Duarte, Fermilab/UC San Diego
D. Lidar, University of Southern California
M. Spiropulu, Caltech

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

Caltech

QAML algorithm

A. Mott, J. Job, J.-R. Vlimant, D. Lidar, M. Spiropulu. "Solving a Higgs optimization problem with quantum annealing for machine learning." *Nature* 550.7676 (2017): 375.

"Quantum annealing for machine learning" (QAML)

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

Caltech

QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

$$\underset{s_i}{\operatorname{argmin}} \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i \, c_i(\mathbf{x}_\tau) \right|^2$$

Caltech

QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

Training label

Training set

$$\underset{s_i}{\operatorname{argmin}} \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i \, c_i(\mathbf{x}_\tau) \right|^2$$

Training input

Caltech

QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

Training label

Weak classifier = ±1/*N*

Training set

$$\underset{s_i}{\operatorname{argmin}} \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i\, c_i(\mathbf{x}_\tau) \right|^2$$

Training input

Caltech

QAML algorithm

Rationale: minimize squared error

Method: create strong classifier from sum of weak classifiers

Training label *Weak classifier = ±1/N*

Training set

$$\underset{s_i}{\operatorname{argmin}} \sum_{\tau=1}^{S} \left| y_\tau - \sum_{i=1}^{N} s_i \, c_i(\mathbf{x}_\tau) \right|^2$$

Classifier weight *Training input*

QAML algorithm

Rationale: minimize squared error

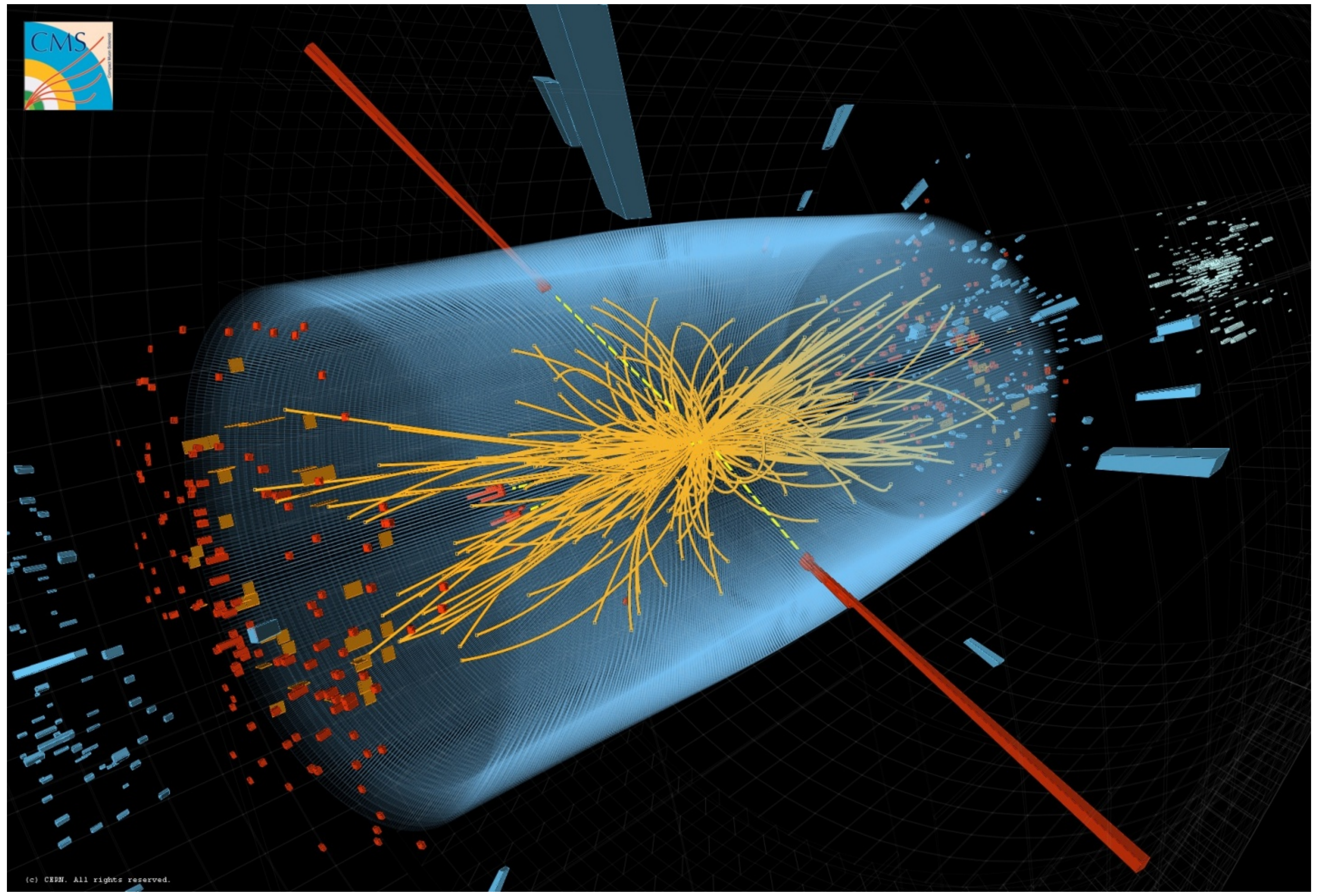
Method: create strong classifier from sum of weak classifiers

$$H_{\text{Ising}} = \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{\tau=1}^{S} s_i\, c_i(\mathbf{x}_\tau)\, s_j\, c_j(\mathbf{x}_\tau) - \sum_{i=1}^{N} \sum_{\tau=1}^{S} s_i\, c_i(\mathbf{x}_\tau)\, y_\tau$$

Caltech

Higgs problem construction

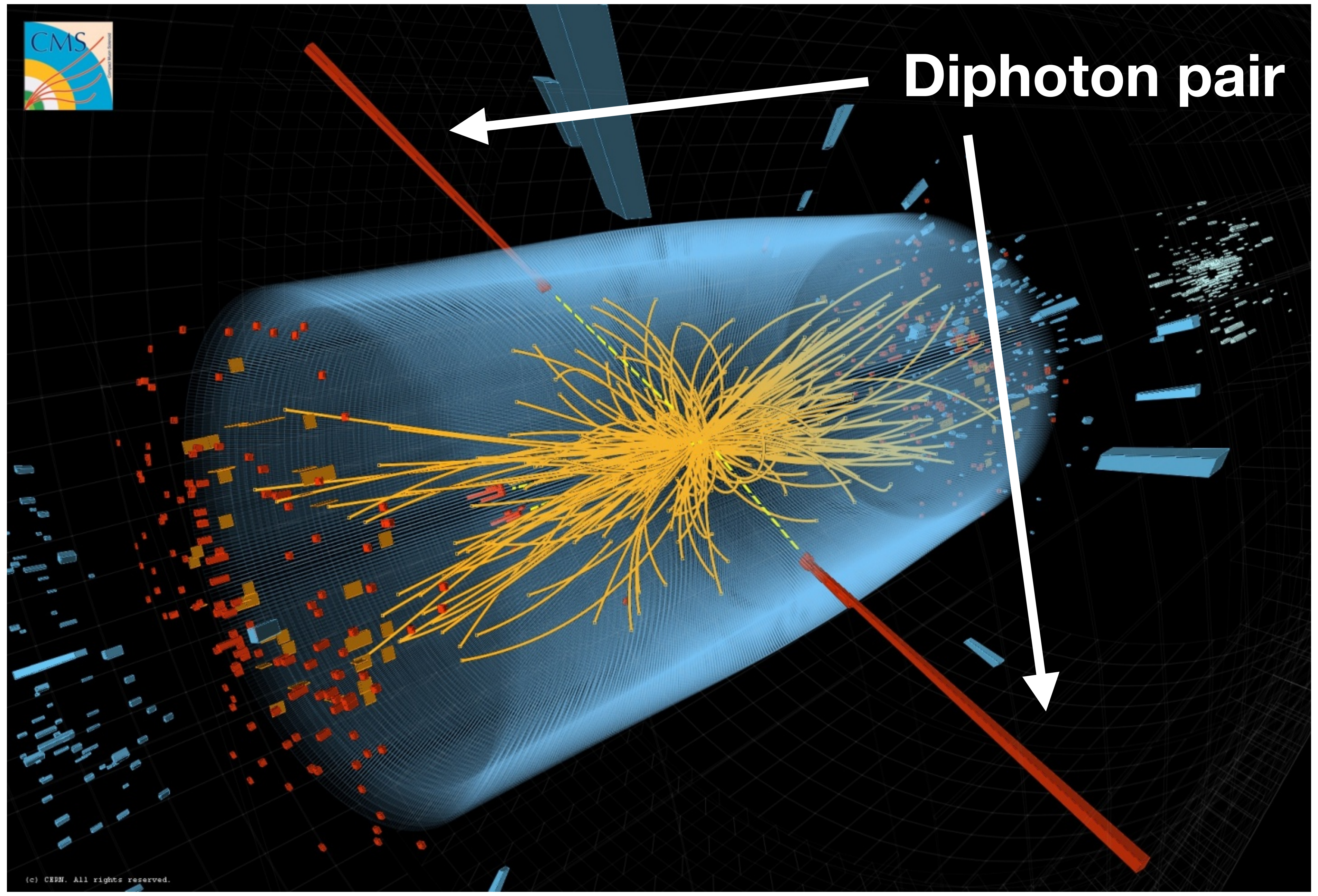
Can we “rediscover” the Higgs boson with QAML?



Caltech

Higgs problem construction

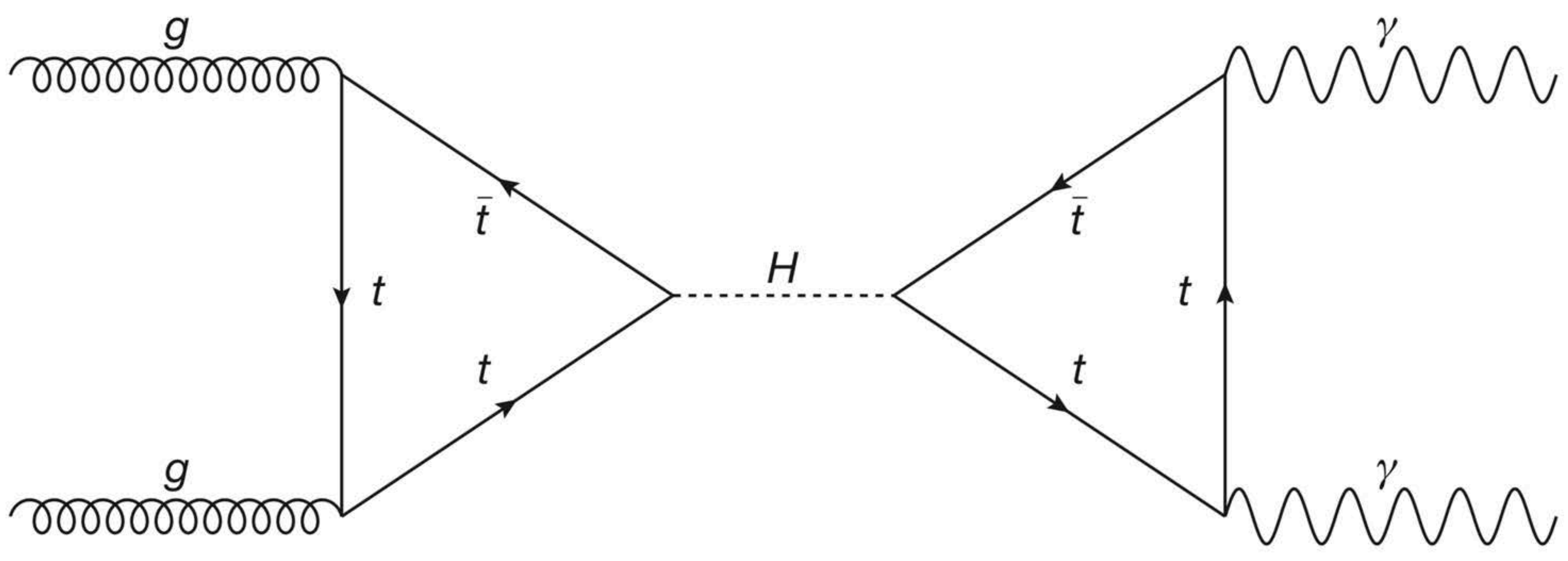
Can we “rediscover” the Higgs boson with QAML?



Caltech

Higgs problem construction

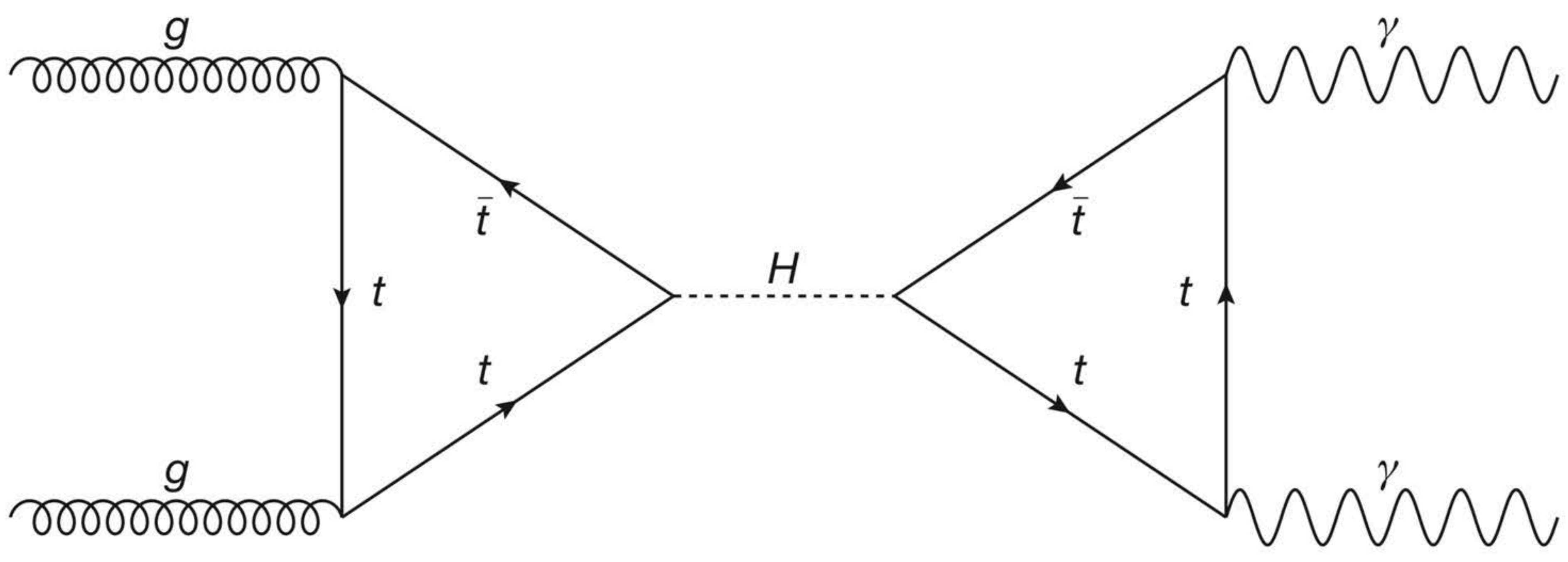
Higgs boson



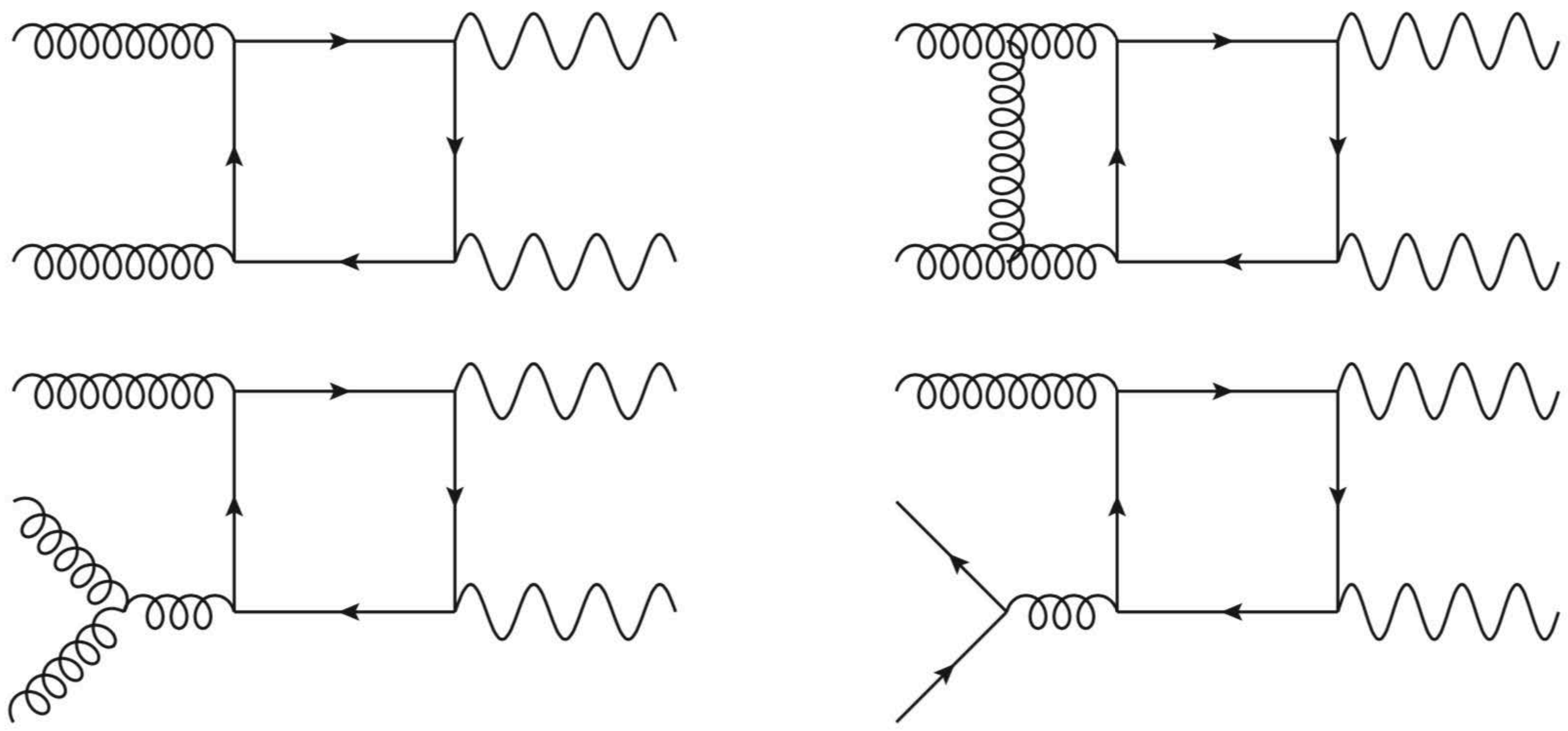
Caltech

Higgs problem construction

Higgs boson



Other Standard Model (SM) processes



Caltech

Higgs problem construction

Eight kinematic observables assembled from decay photons:

Transverse momentum + diphoton mass

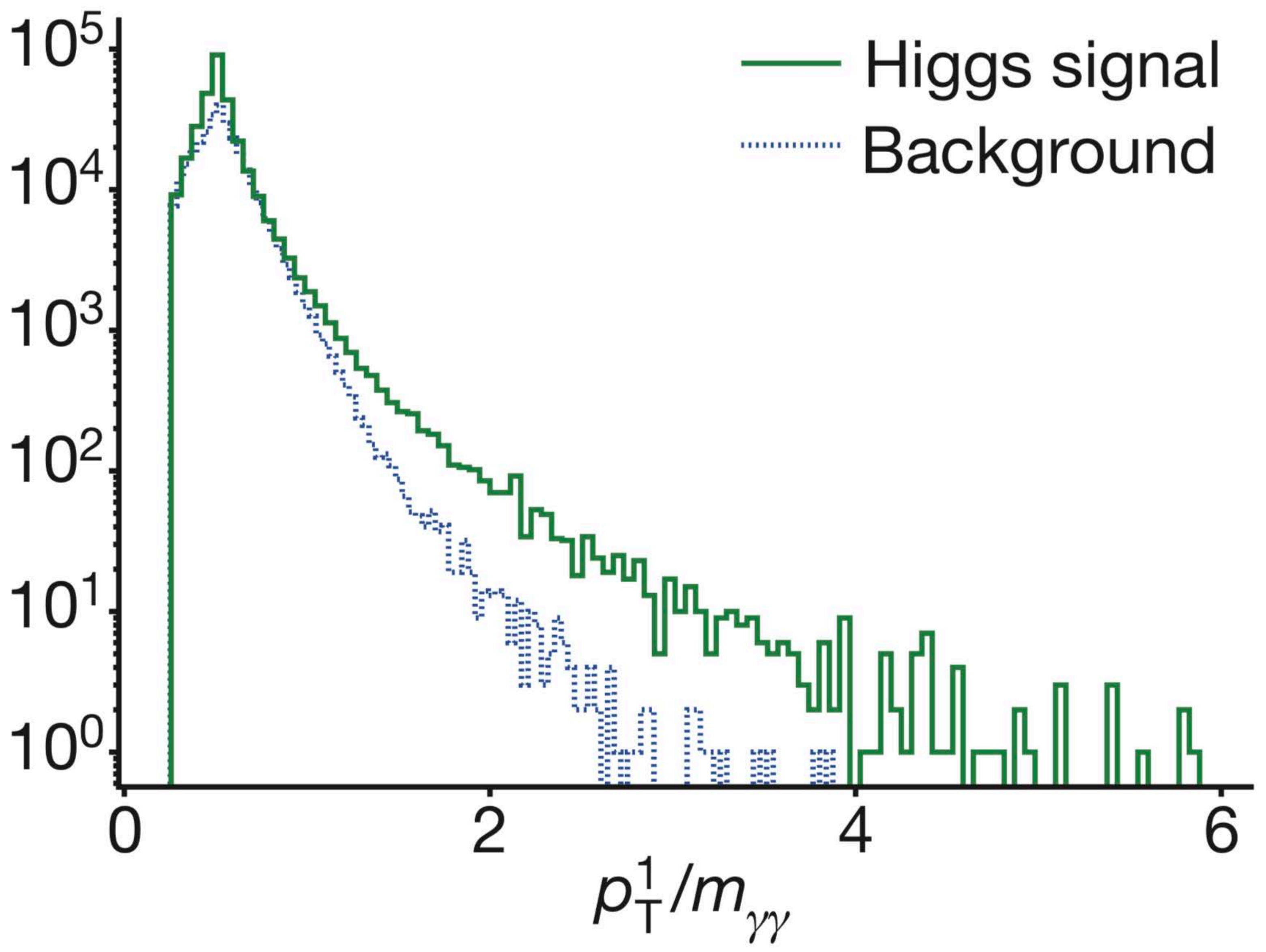
$$p_{\mathrm{T}}^{1}/m_{\gamma\gamma}\,,\;\; p_{\mathrm{T}}^{2}/m_{\gamma\gamma}\,,\;\; (p_{\mathrm{T}}^{1}+p_{\mathrm{T}}^{2})^{2}/m_{\gamma\gamma}\,,\;\; (p_{\mathrm{T}}^{1}-p_{\mathrm{T}}^{2})^{2}/m_{\gamma\gamma}\,,\;\; p_{\mathrm{T}}^{\gamma\gamma}/m_{\gamma\gamma}\,,\;\; \Delta\eta\,,\;\; \Delta R\,,\;\; |\eta^{\gamma\gamma}|$$

Diphoton angle

Caltech

Higgs problem construction

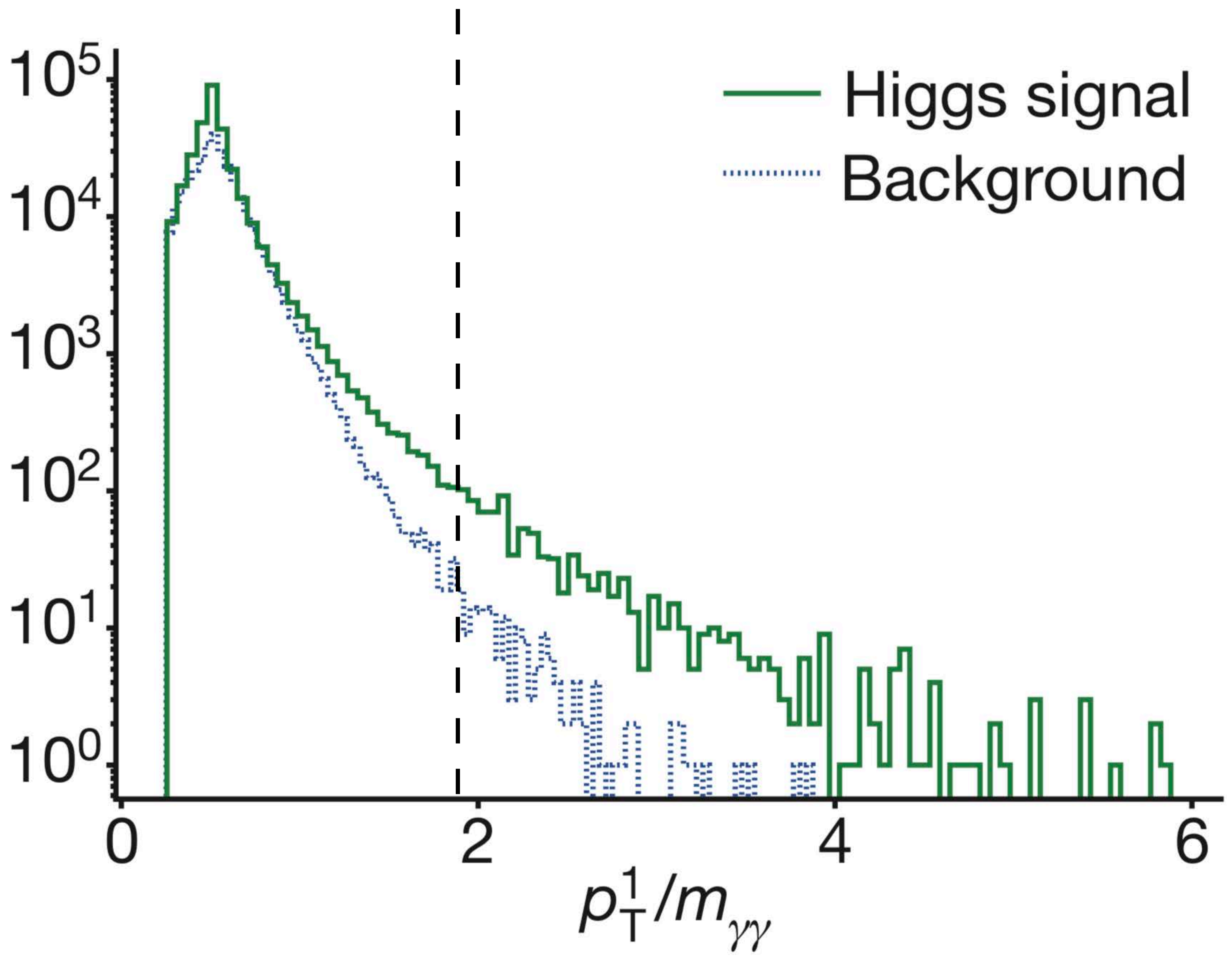
Thirty-six weak classifiers constructed from division and multiplication of eight observables



Caltech

Higgs problem construction

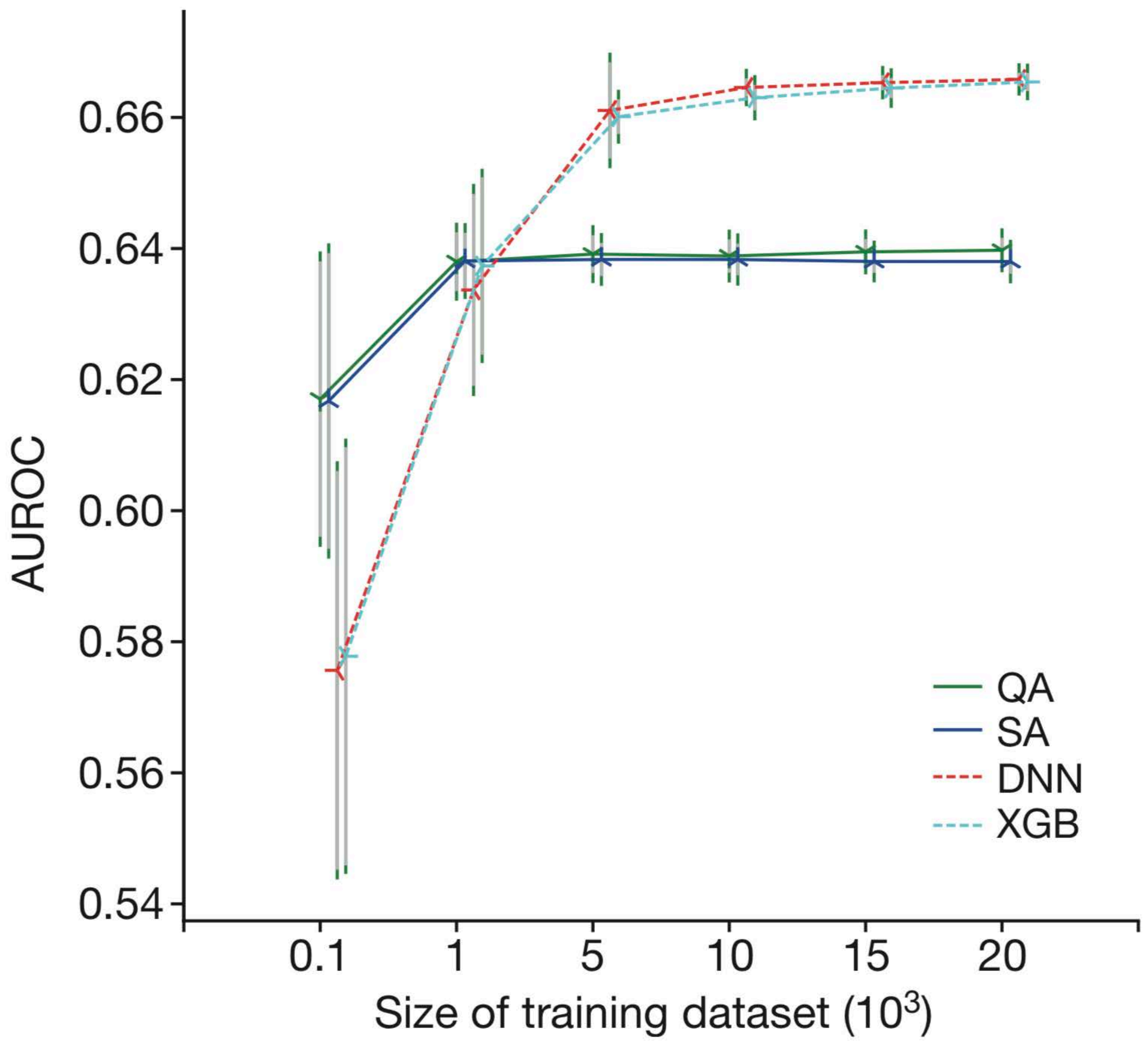
Thirty-six weak classifiers constructed from division and multiplication of eight observables



Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events

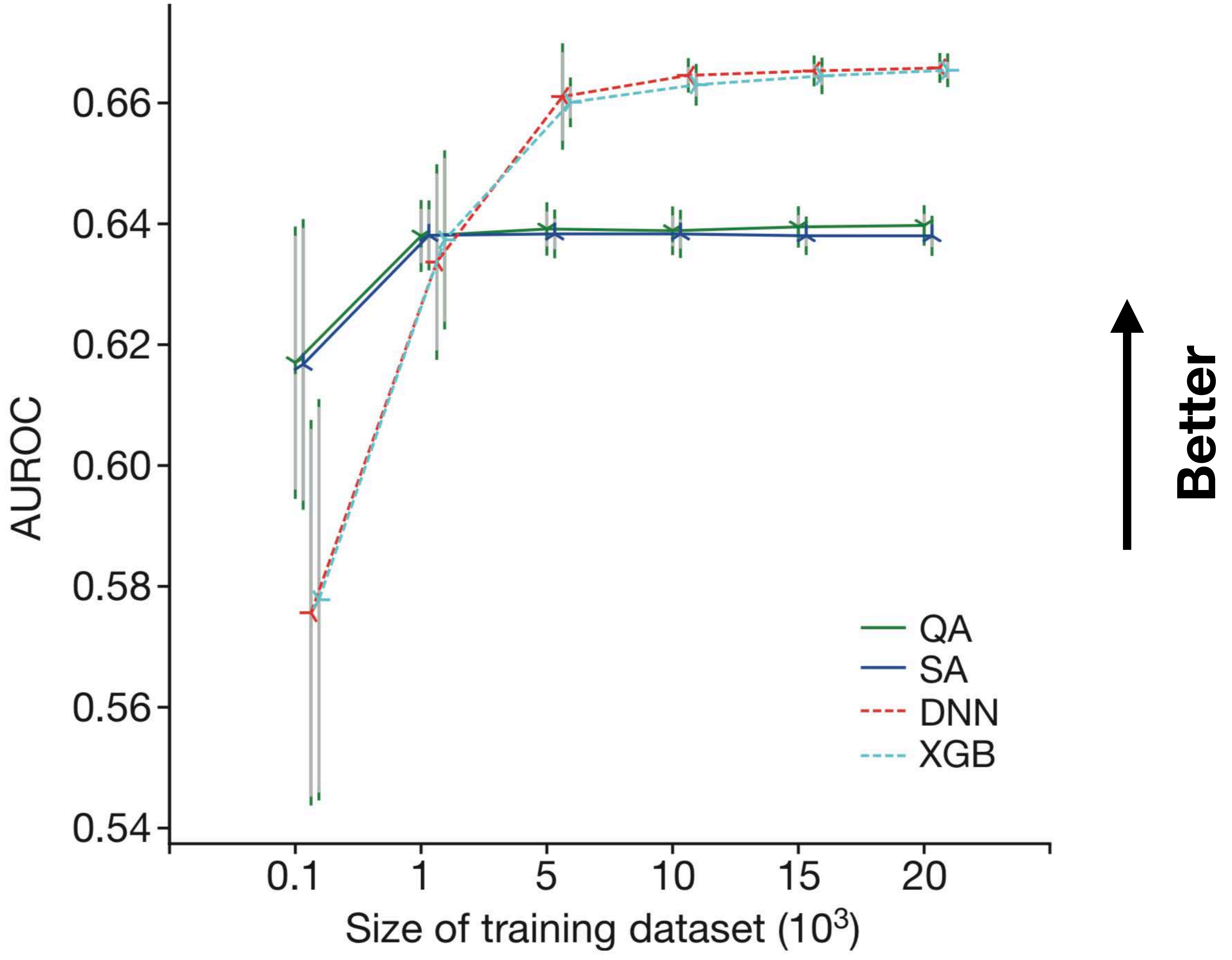


Caltech

Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events

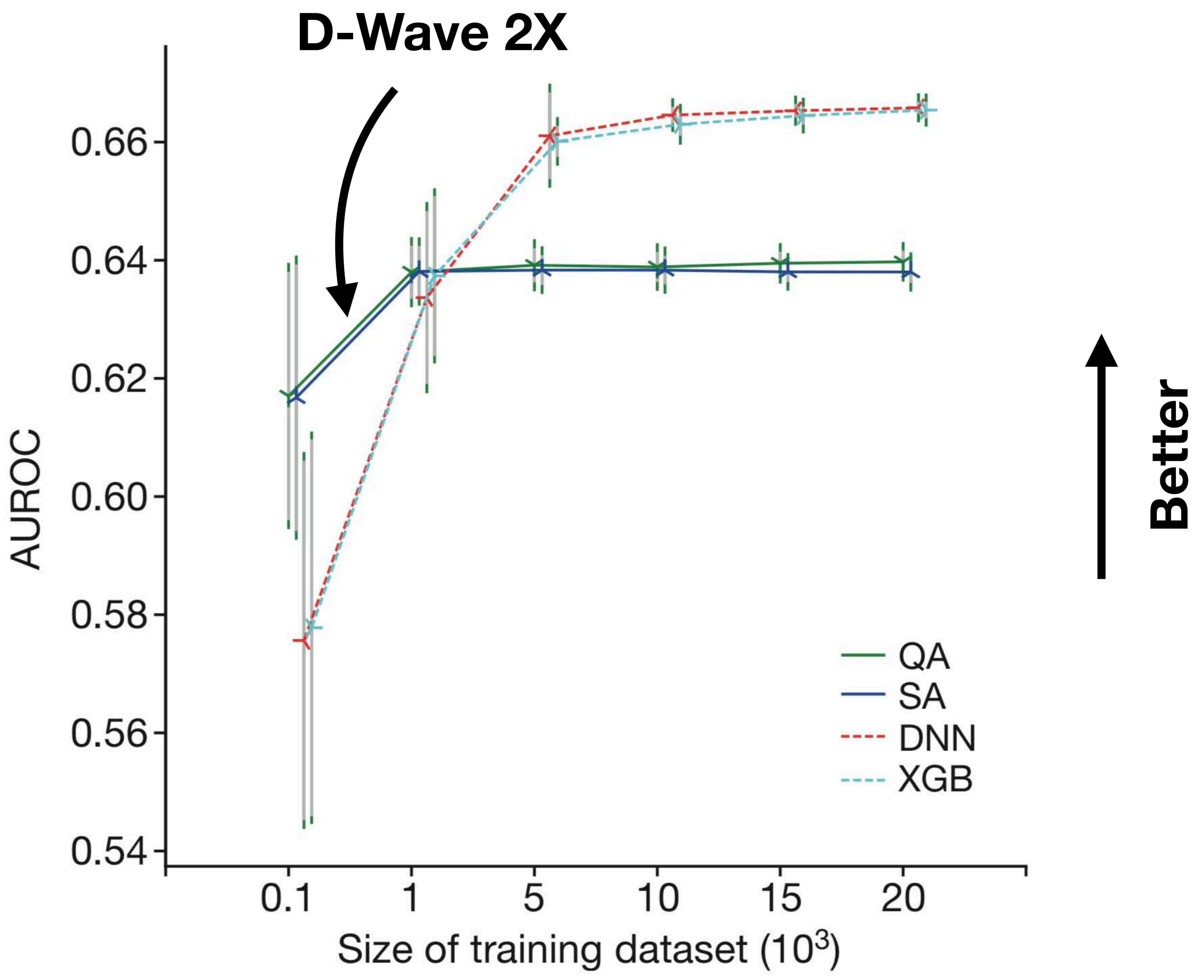


Caltech

Higgs classification results

Optimize simulated annealing, deep neural network, and XGBoost hyperparameters

Measure area under ROC curve on 200,000 simulated events



Caltech

QAML-Z algorithm

A. Zlokapa, A. Mott, J. Job, J.-R. Vlimant, D. Lidar, M. Spiropulu.
"Quantum adiabatic machine learning with zooming." arXiv:
1908.04480 [quant-ph] (2019).

Two improvements:

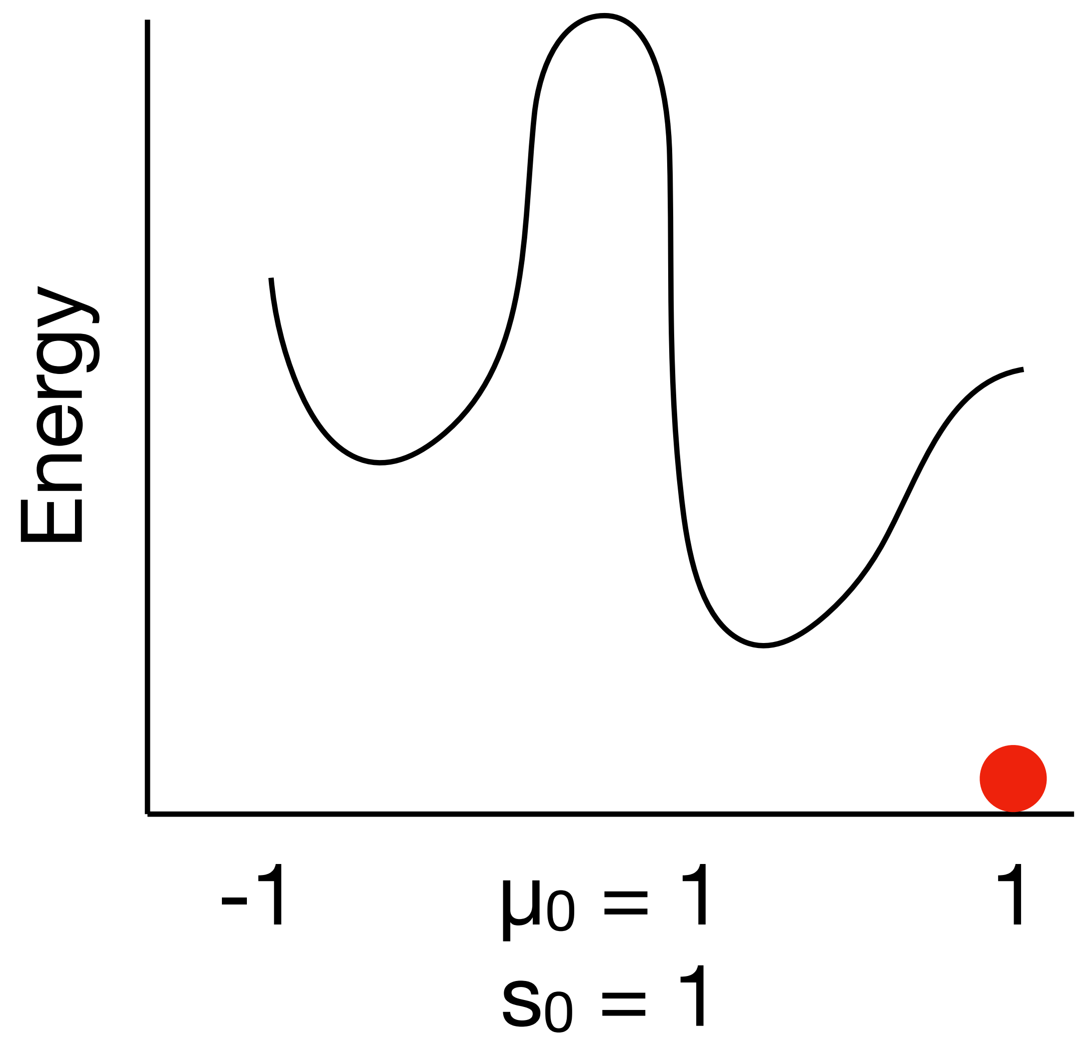
- Zoom into the energy surface — continuous optimization
- Augment the set of classifiers — stronger ensemble

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

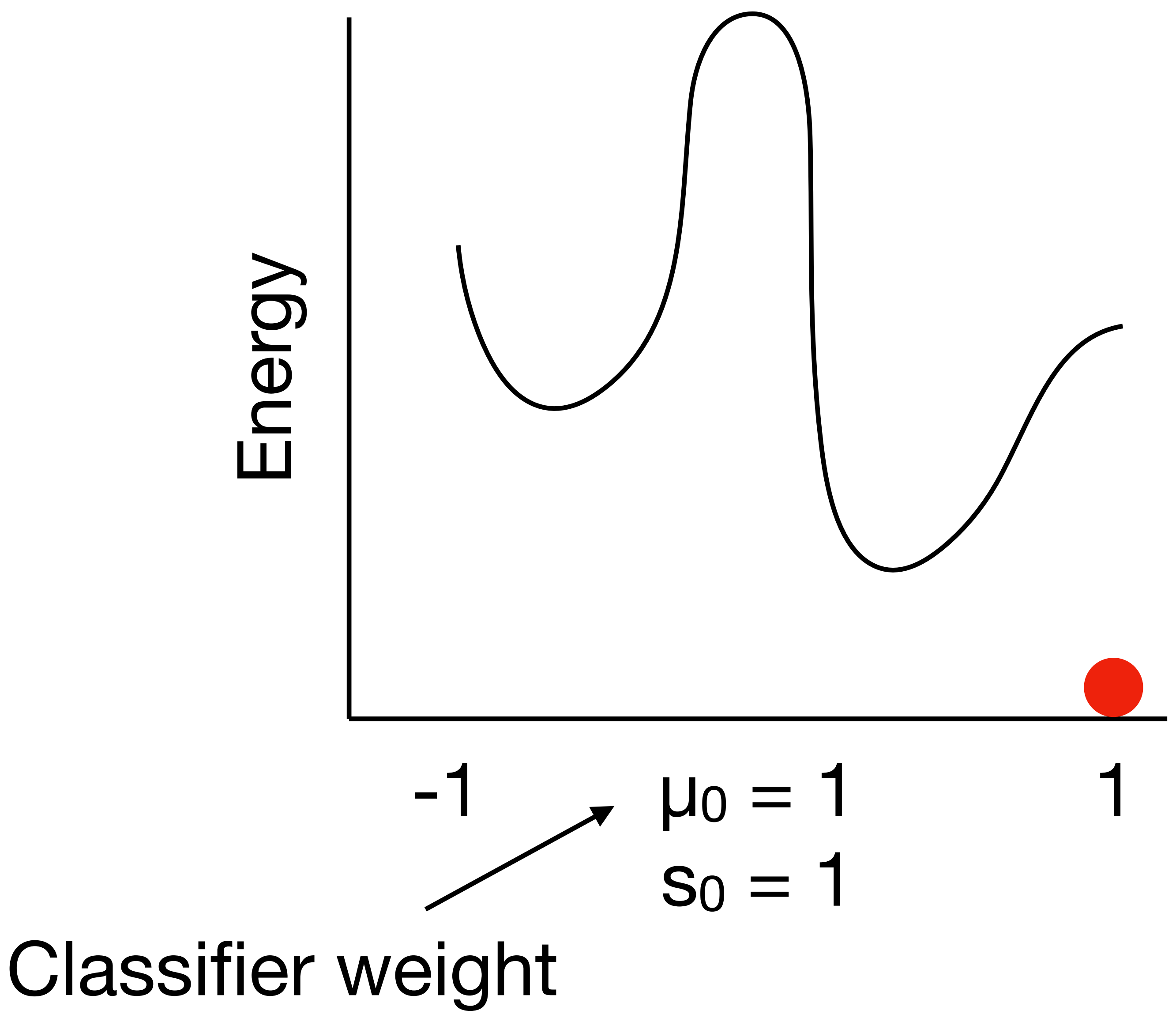


QAML: take discrete values ±1

Caltech

QAML-Z algorithm: Zooming

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

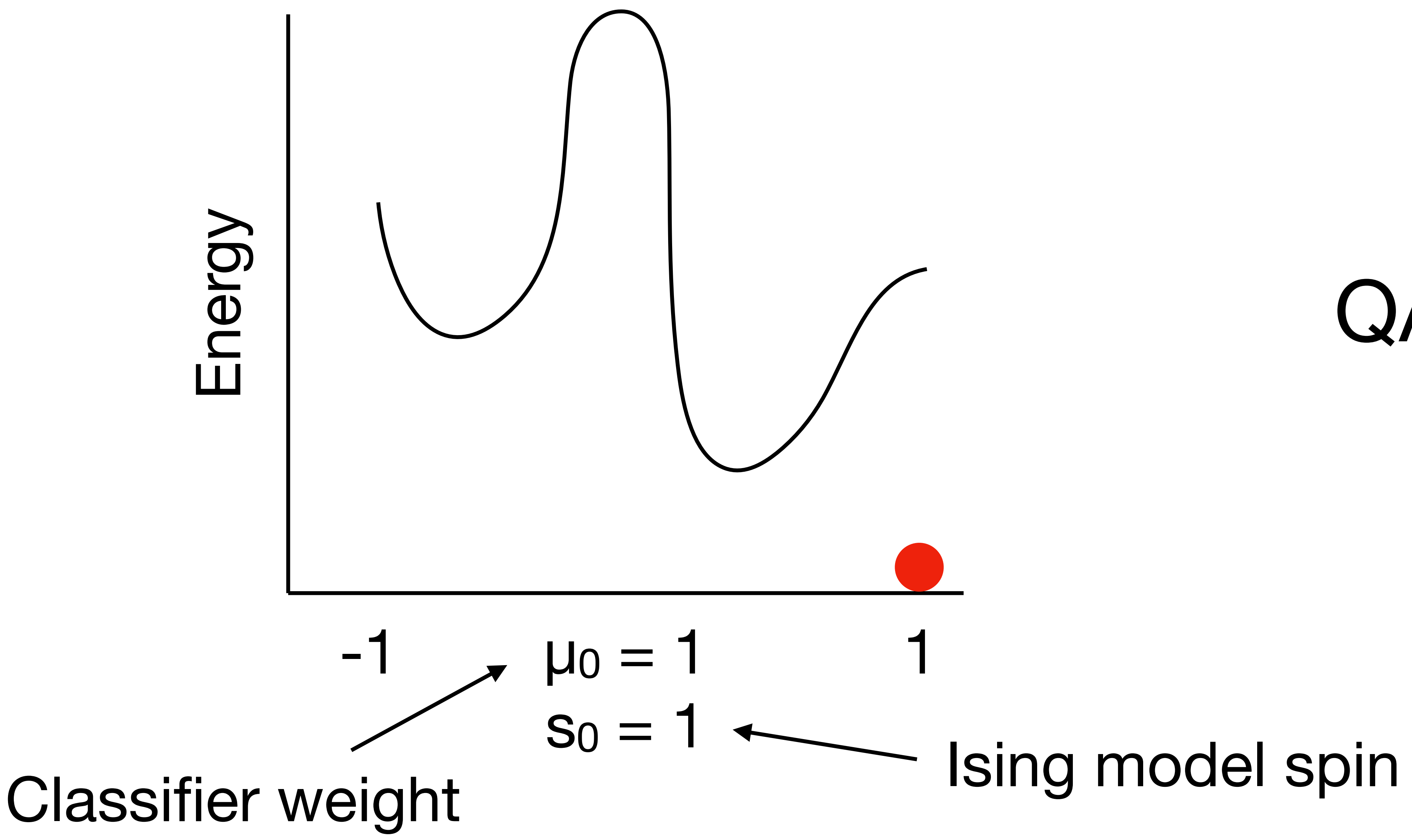


QAML: take discrete values ±1

Caltech

QAML-Z algorithm: Zooming

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

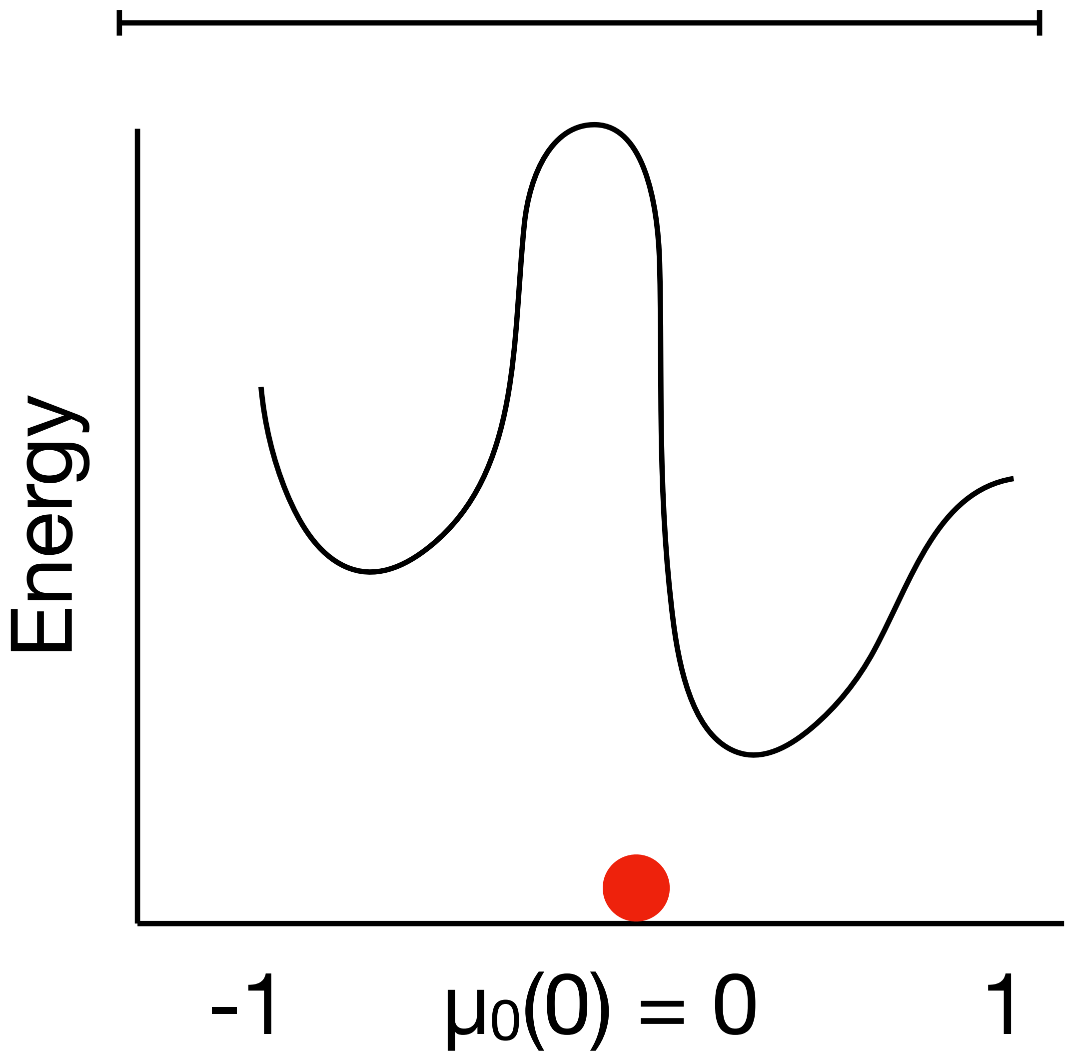


QAML: take discrete values ±1

Caltech

QAML-Z algorithm: Zooming

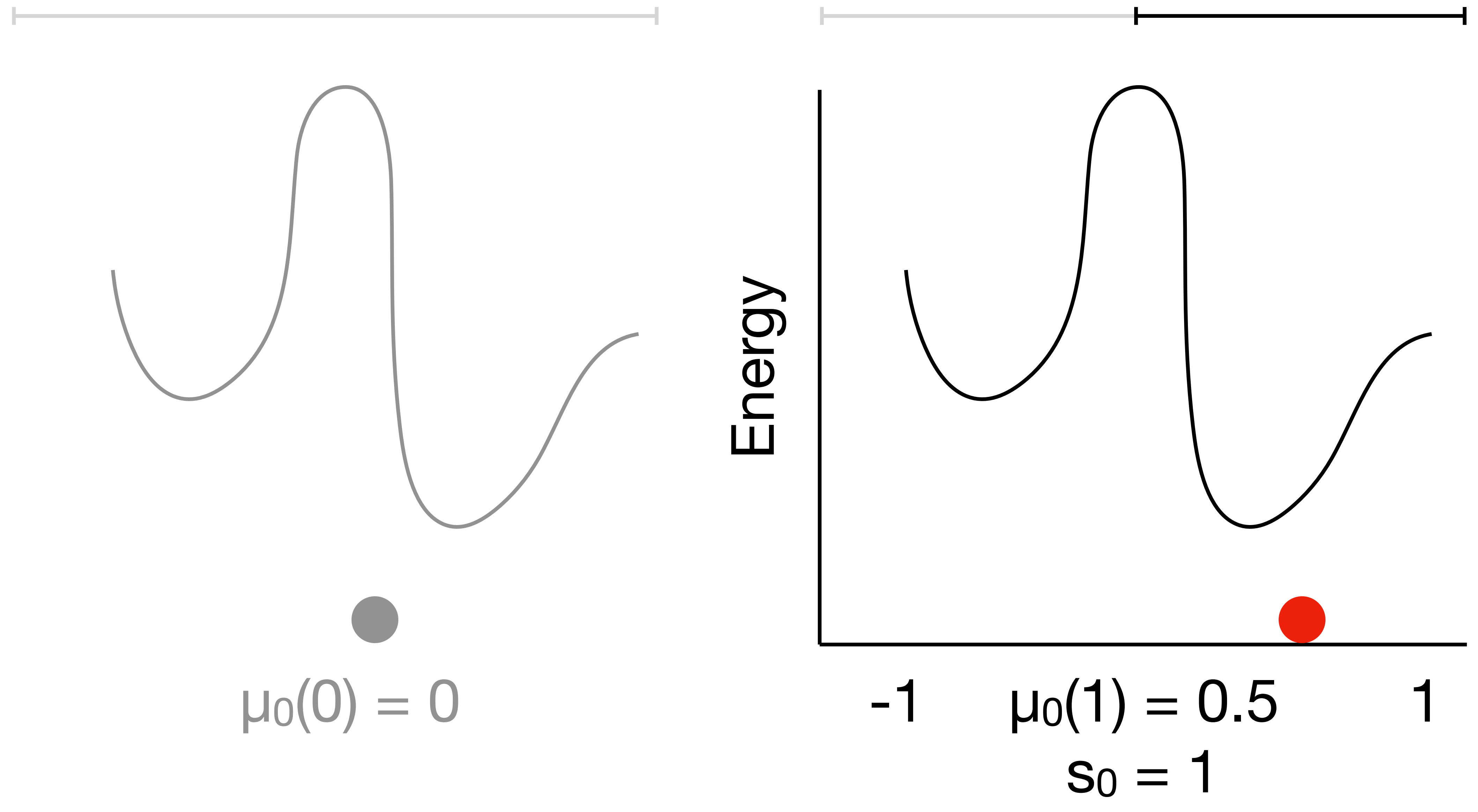
Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals



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

QAML-Z algorithm: Zooming

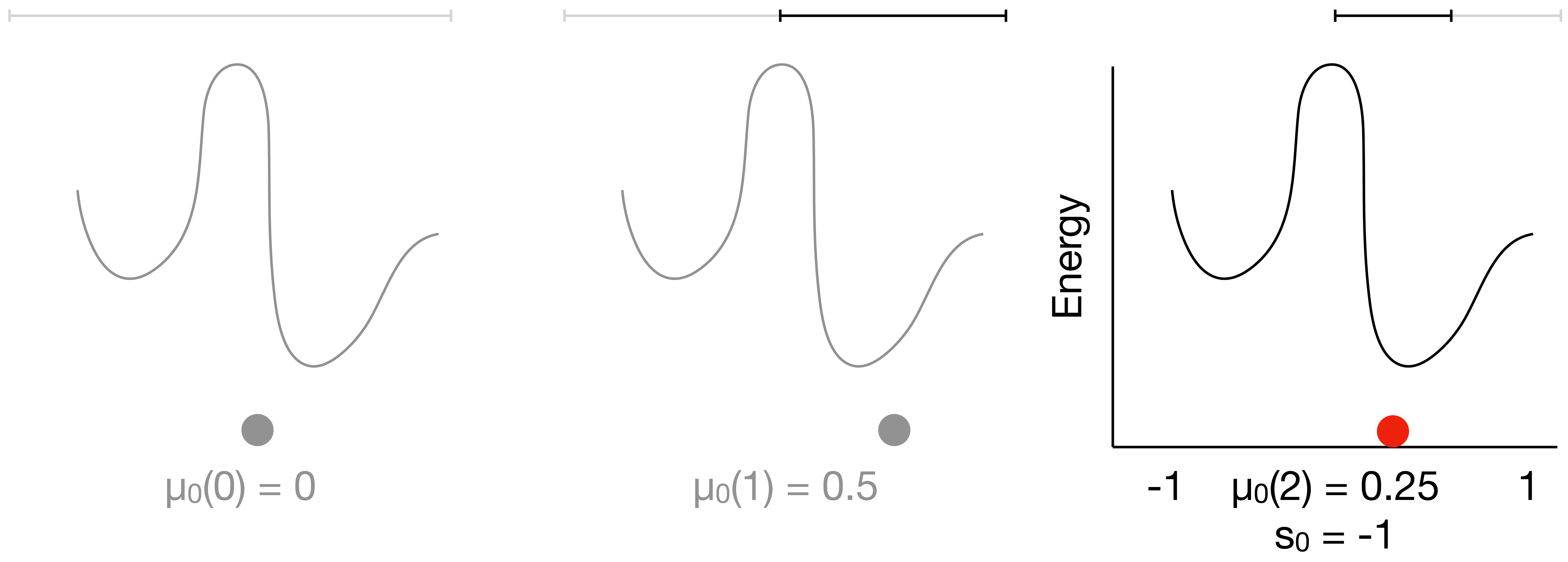
Zooming: perform a binary search on continuous classifier weights by running multiple quantum anneals



Caltech

QAML-Z algorithm: Zooming

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



Caltech

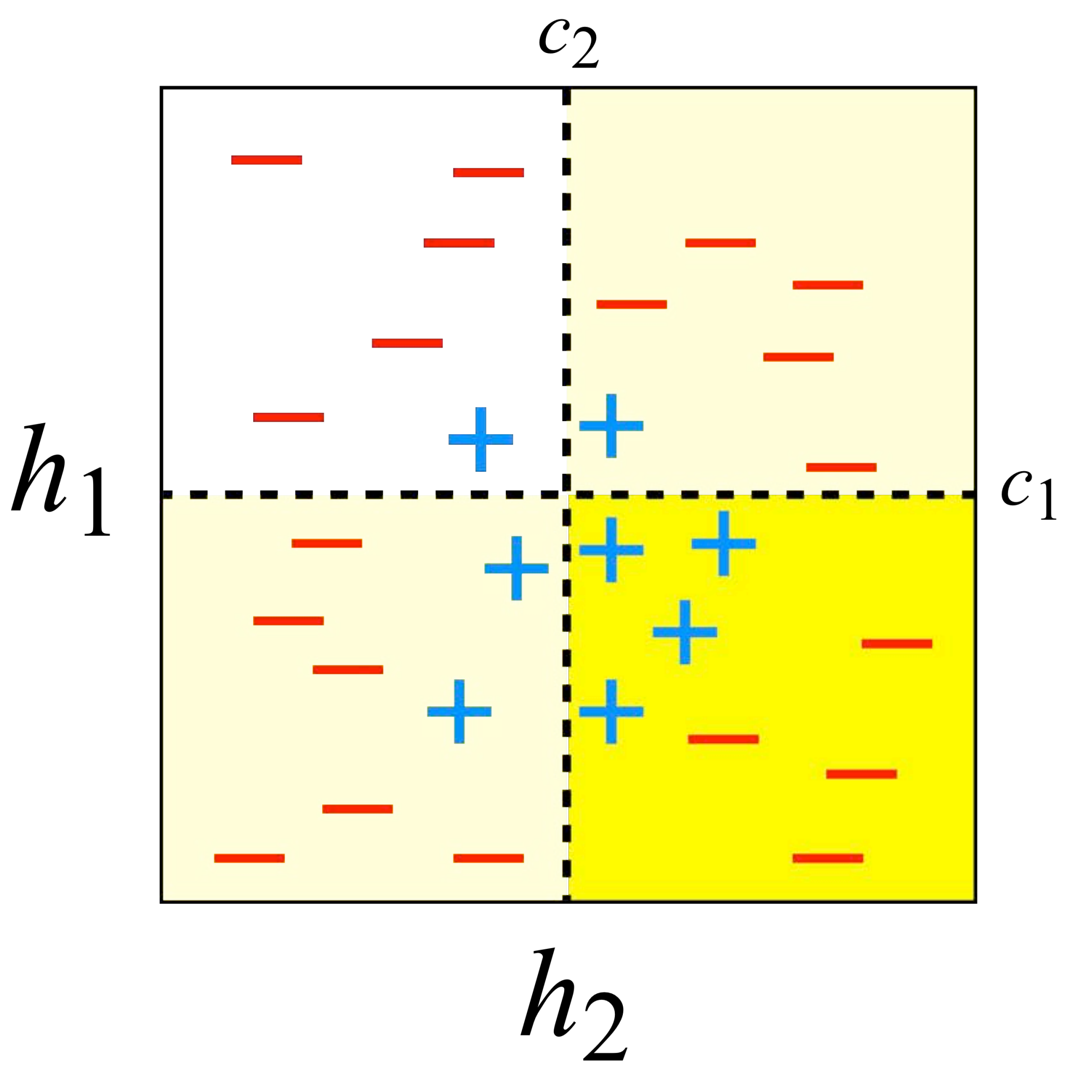
QAML-Z algorithm: Augmentation

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

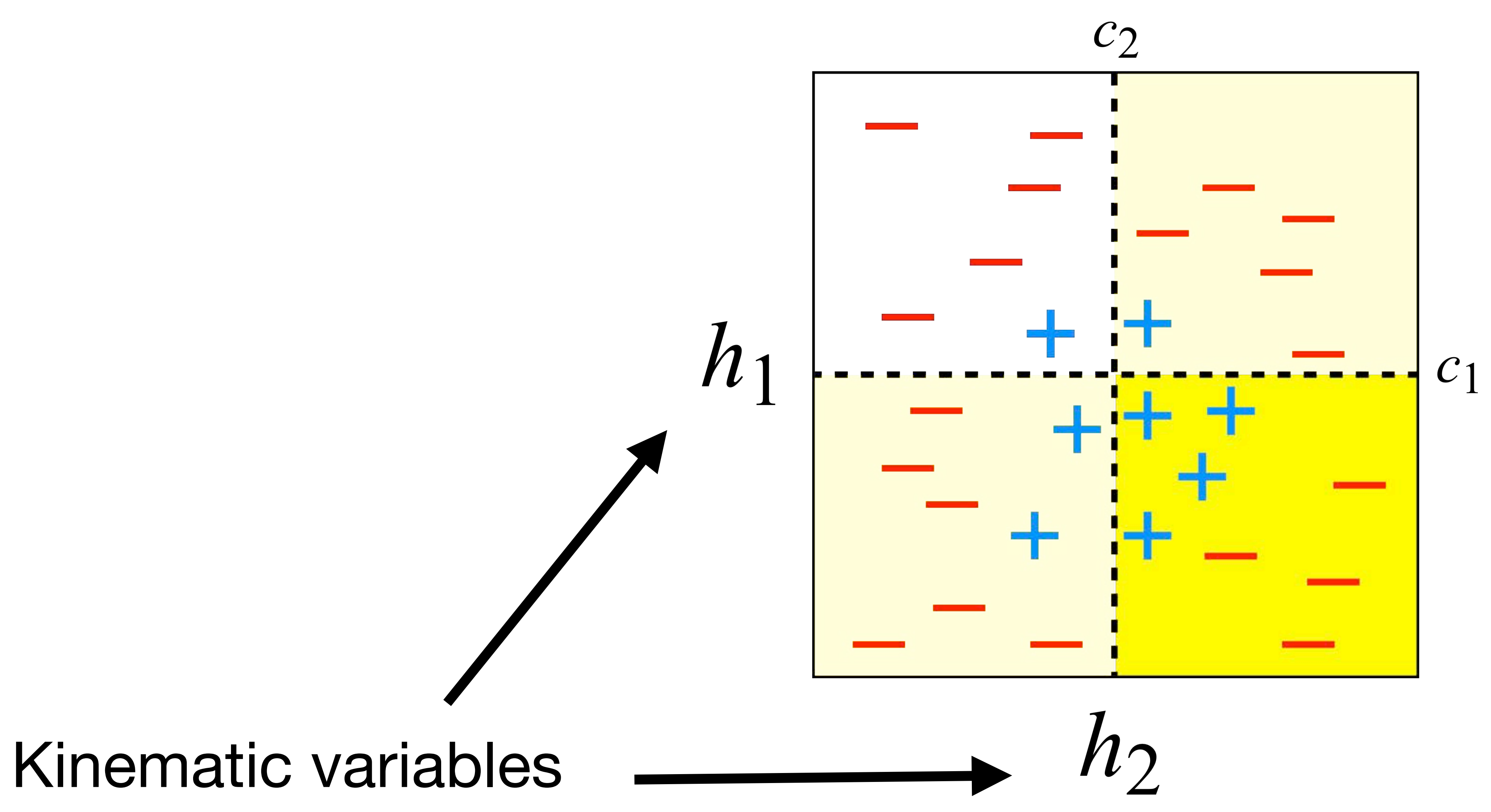


Caltech

QAML-Z algorithm: Augmentation

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

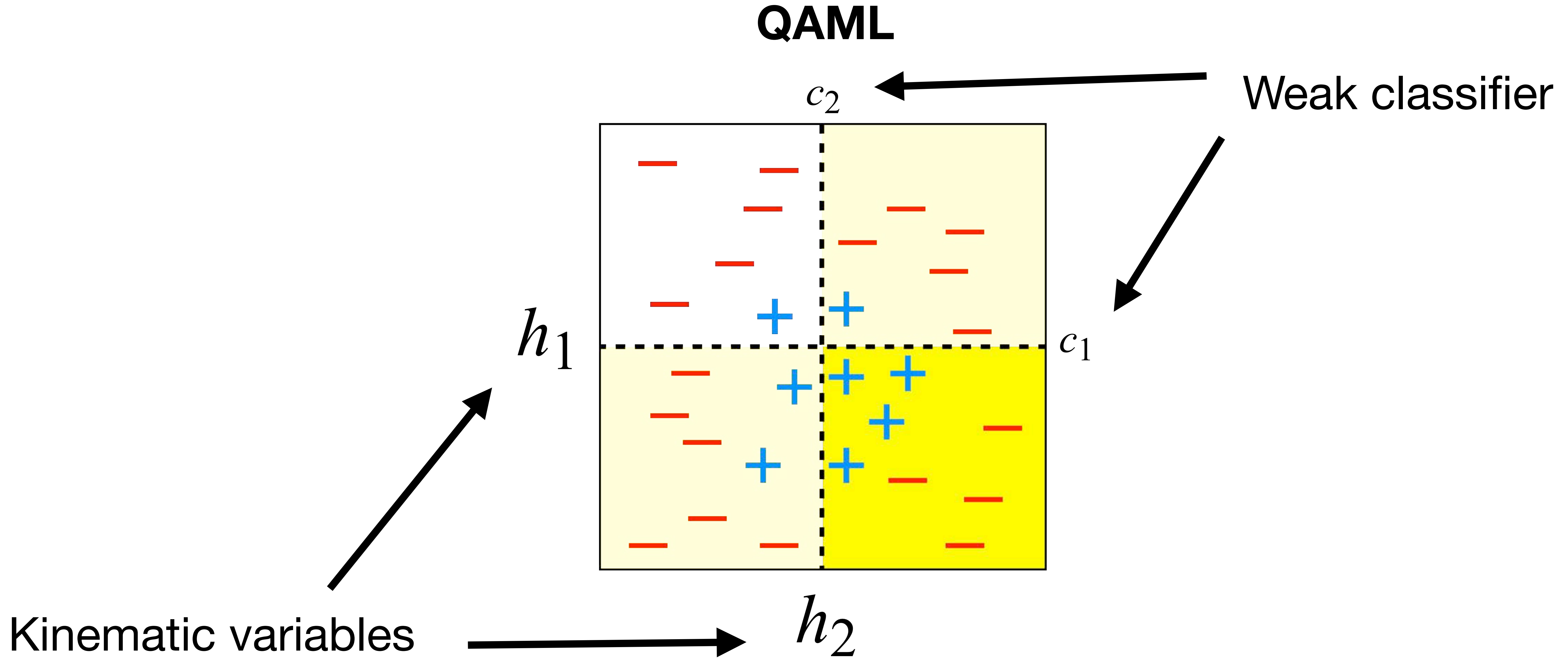
QAML



Caltech

QAML-Z algorithm: Augmentation

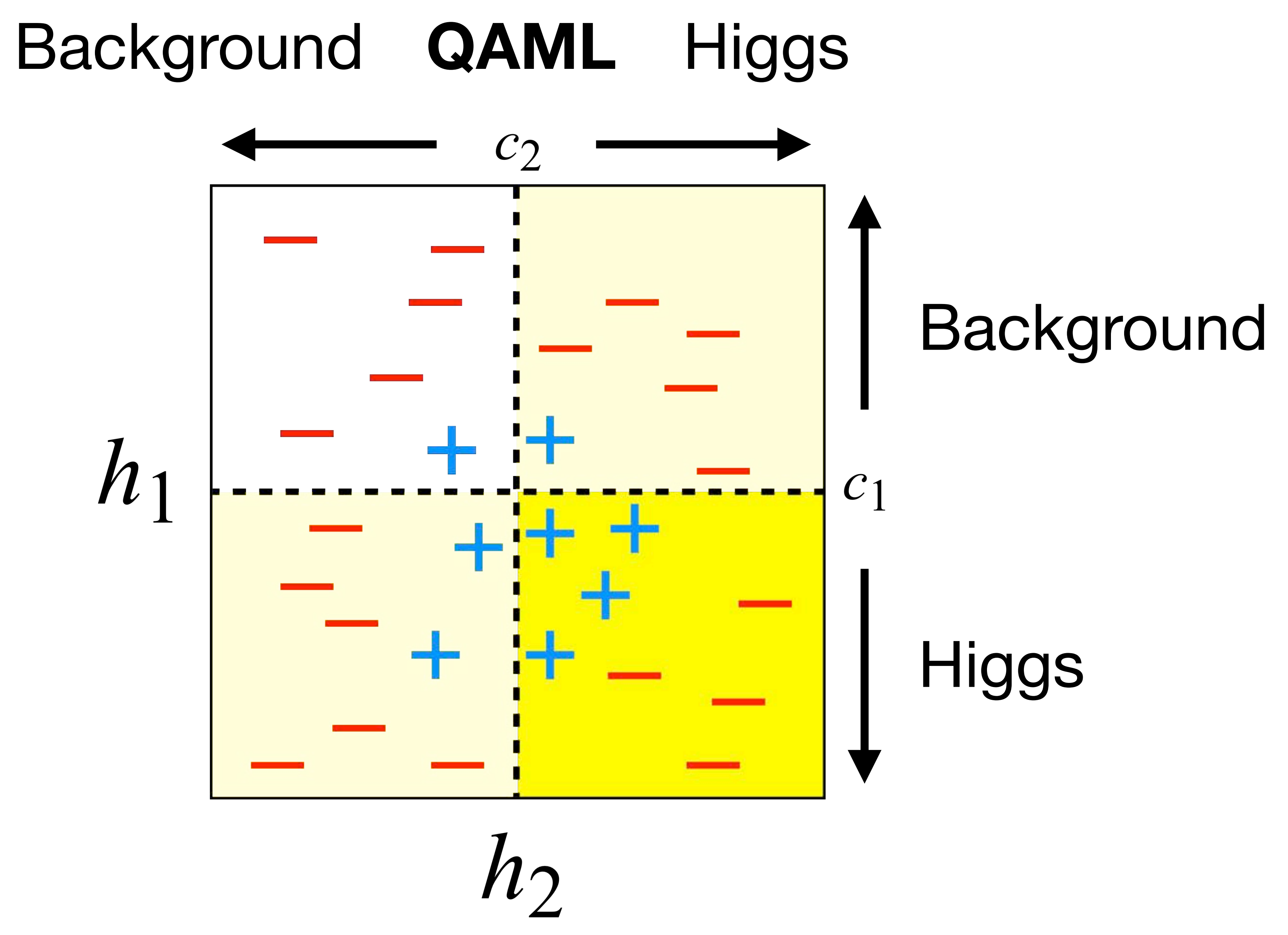
Augmentation: create multiple classifiers from the same combination of physical variables by offsetting distribution cut



Caltech

QAML-Z algorithm: Augmentation

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

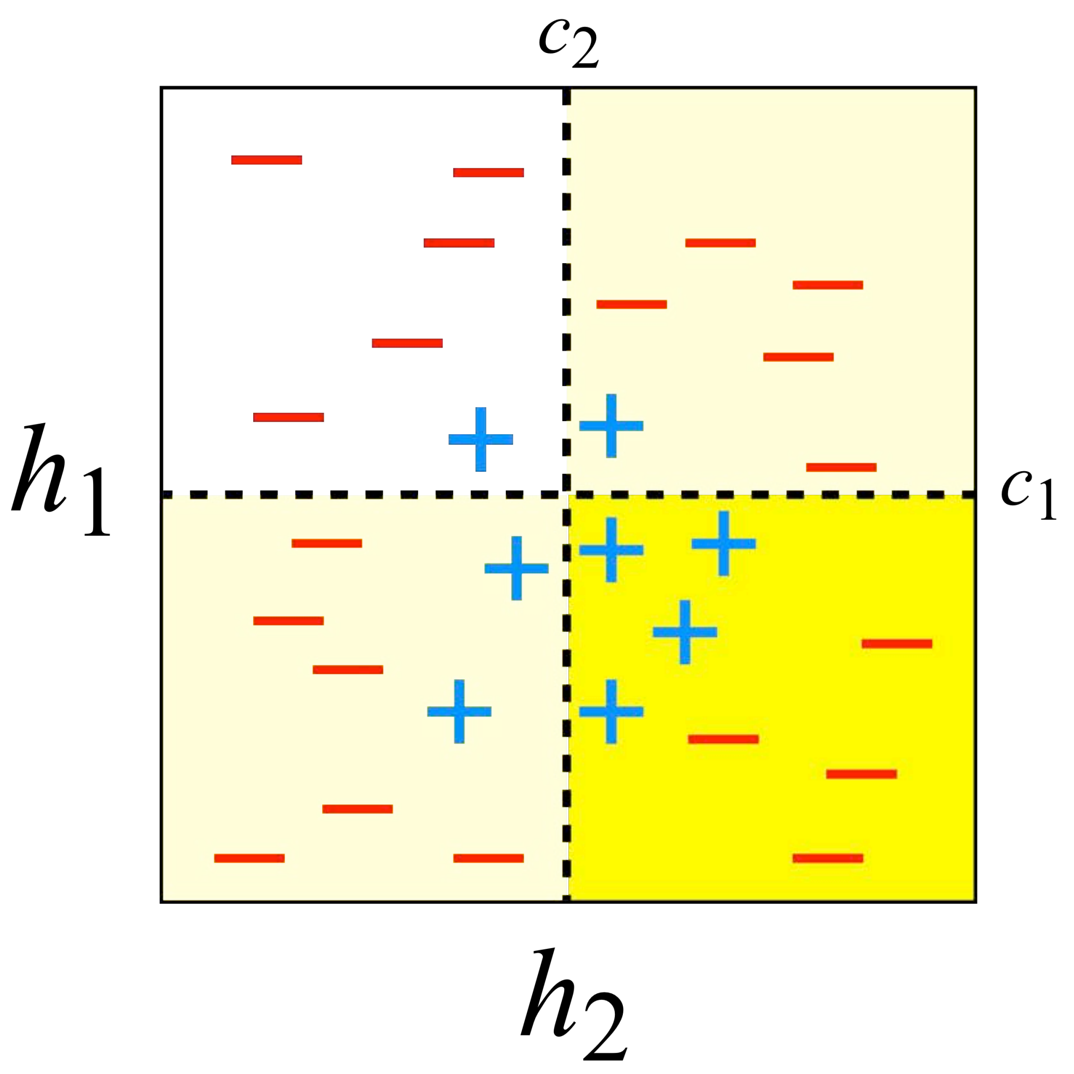


Caltech

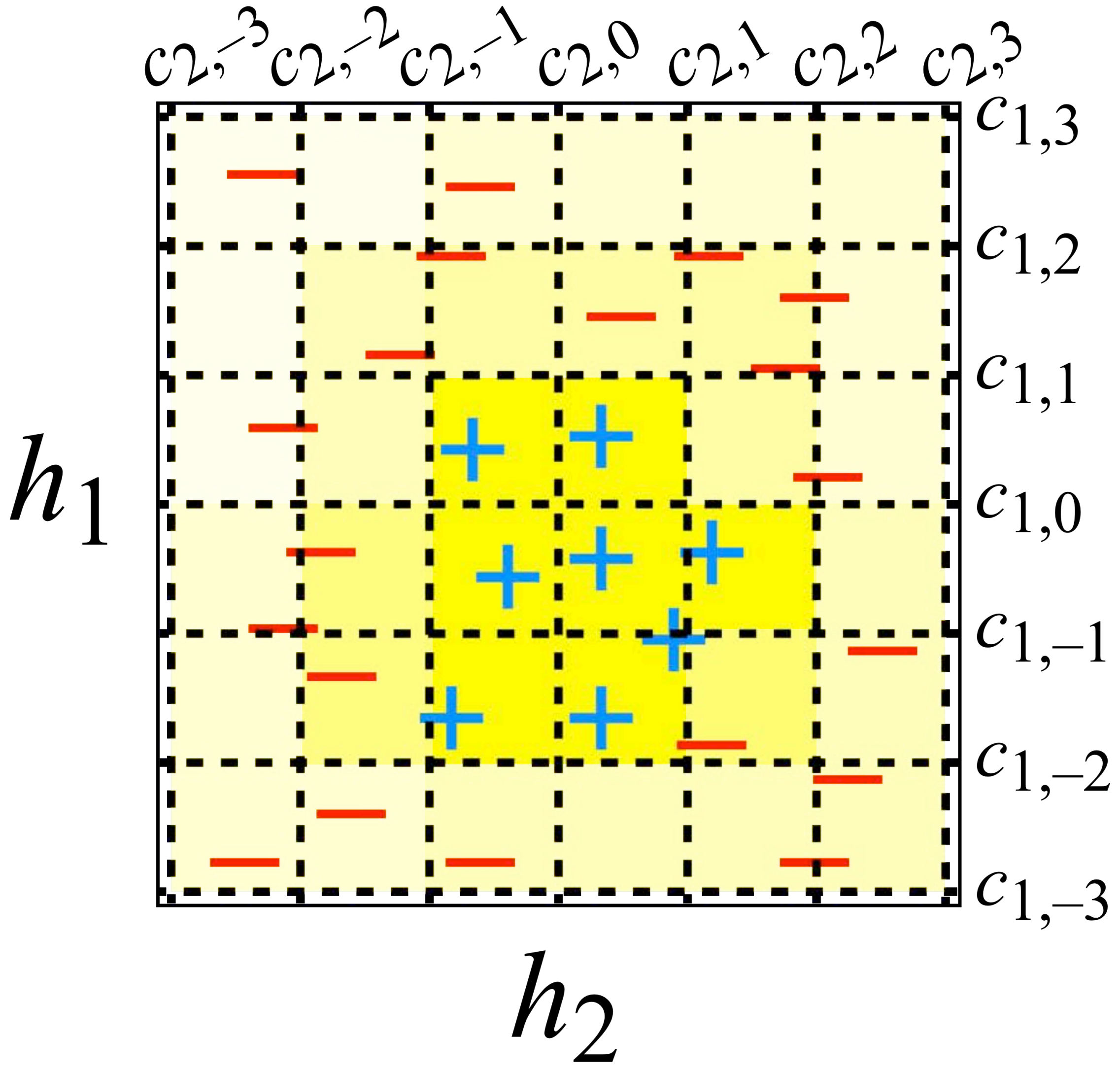
QAML-Z algorithm: Augmentation

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

QAML



QAML-Z



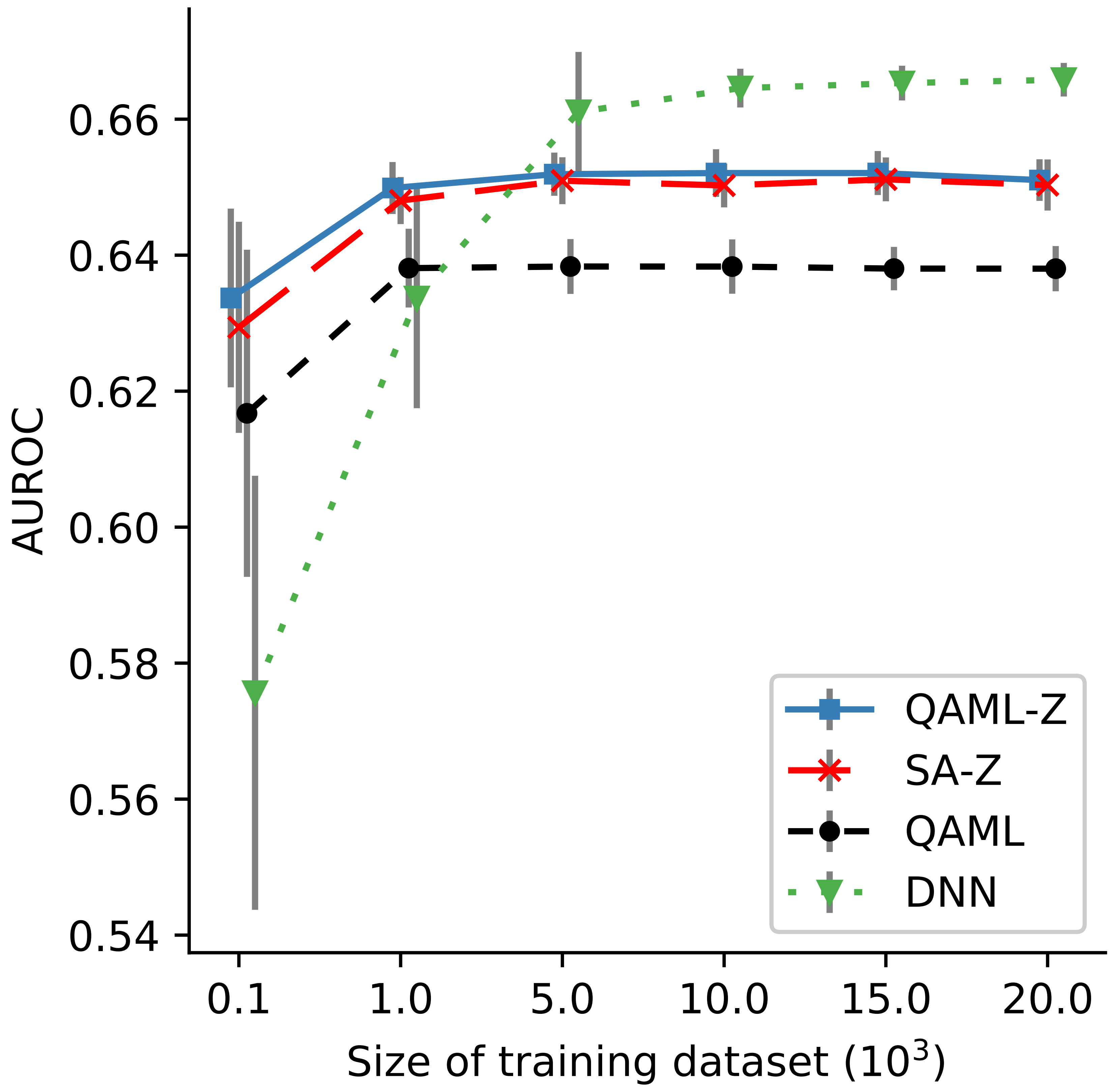
Caltech

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



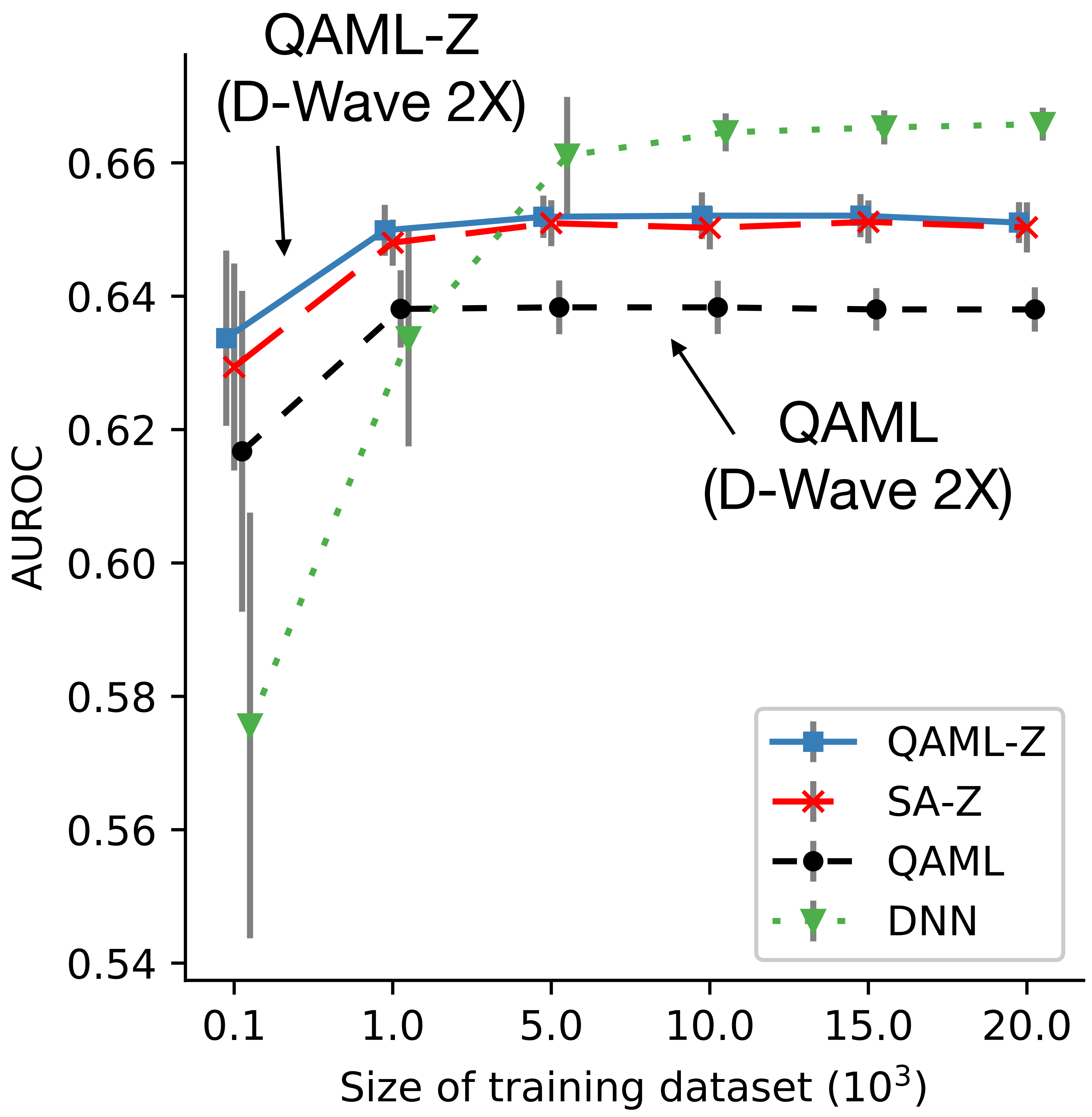
Caltech

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



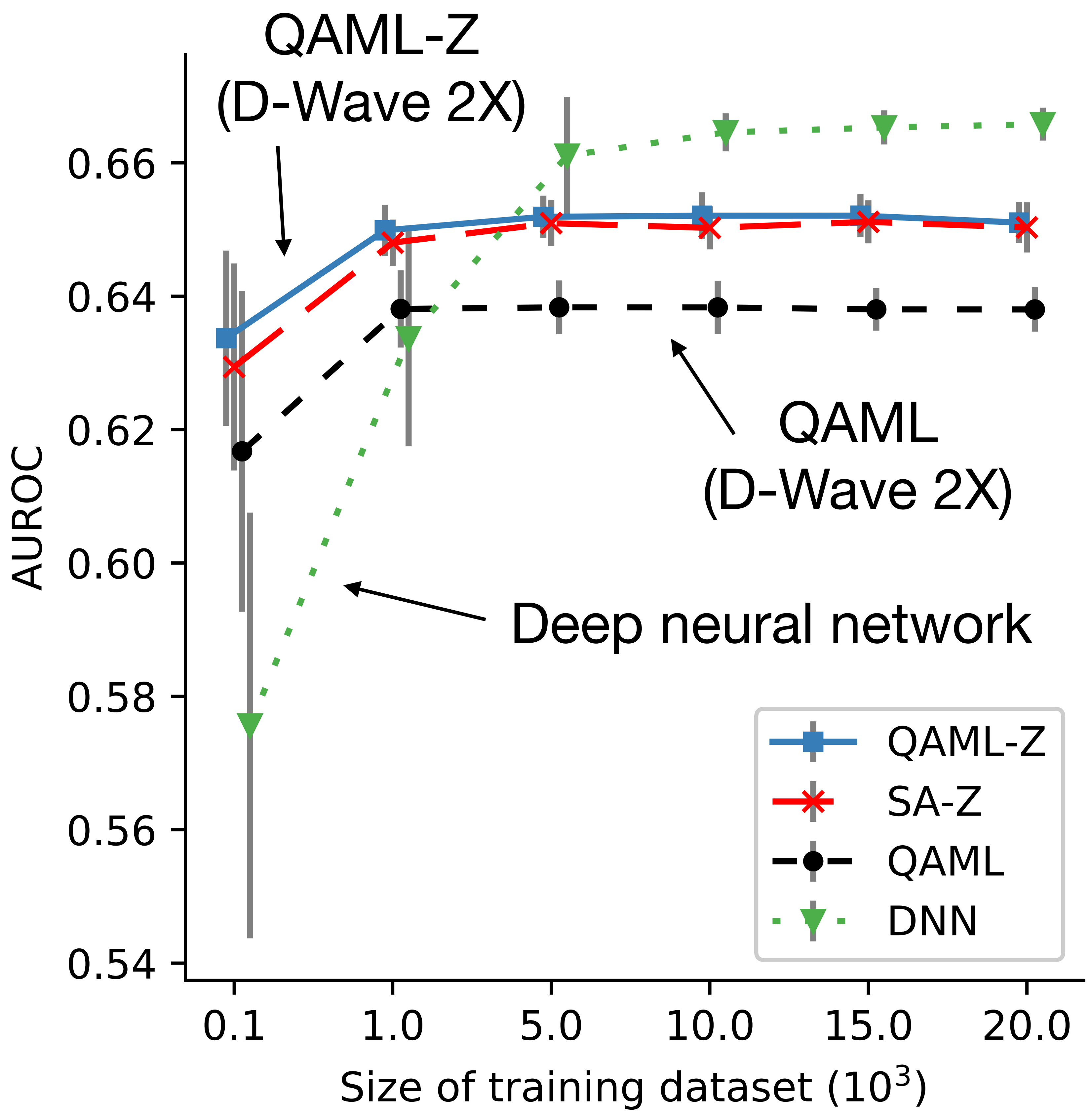
Caltech

Higgs classification results

QAML-Z vs. QAML

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

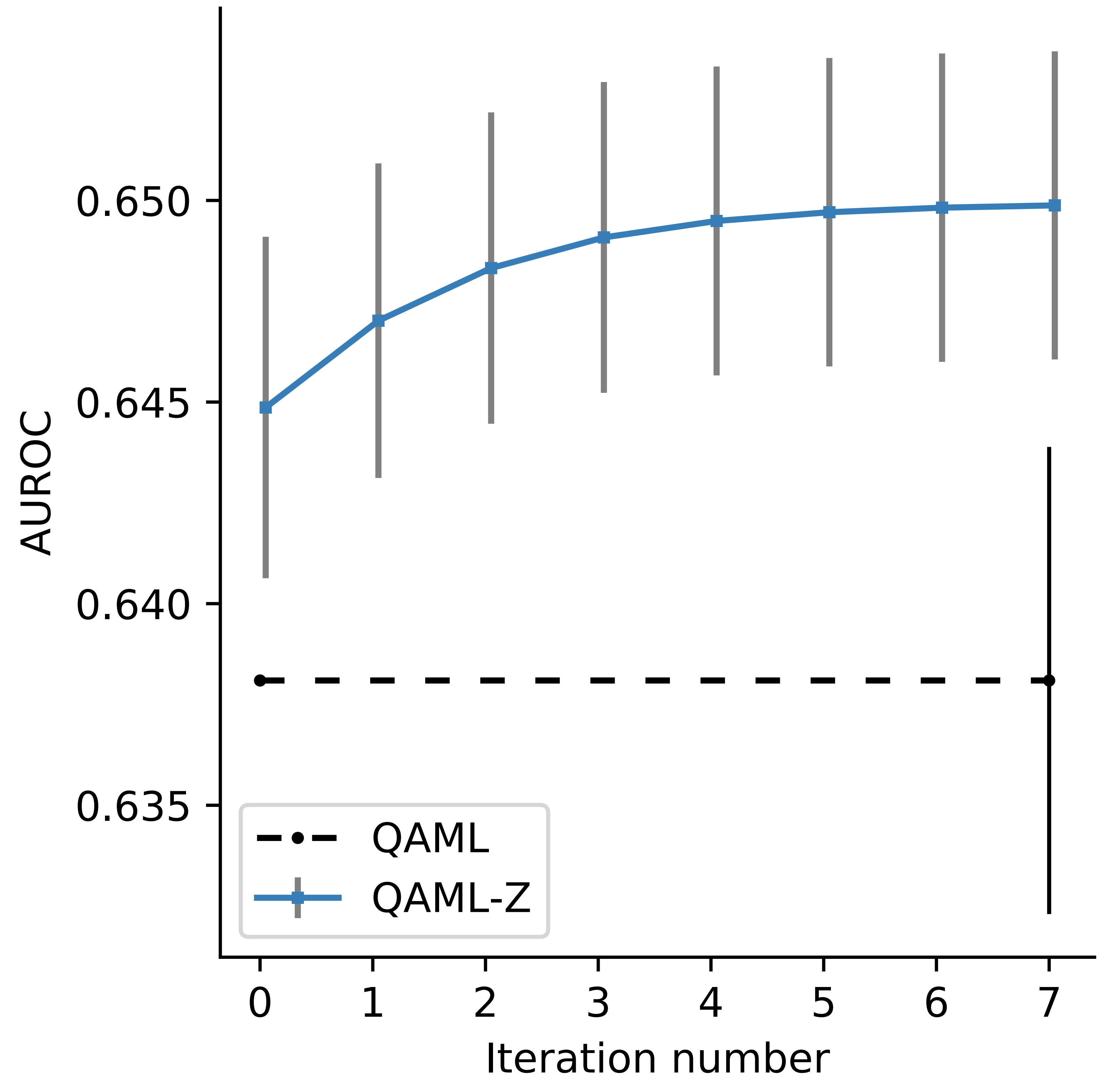
Shrinks disadvantage to DNN by ~50% for large training sets



Caltech

Higgs classification results

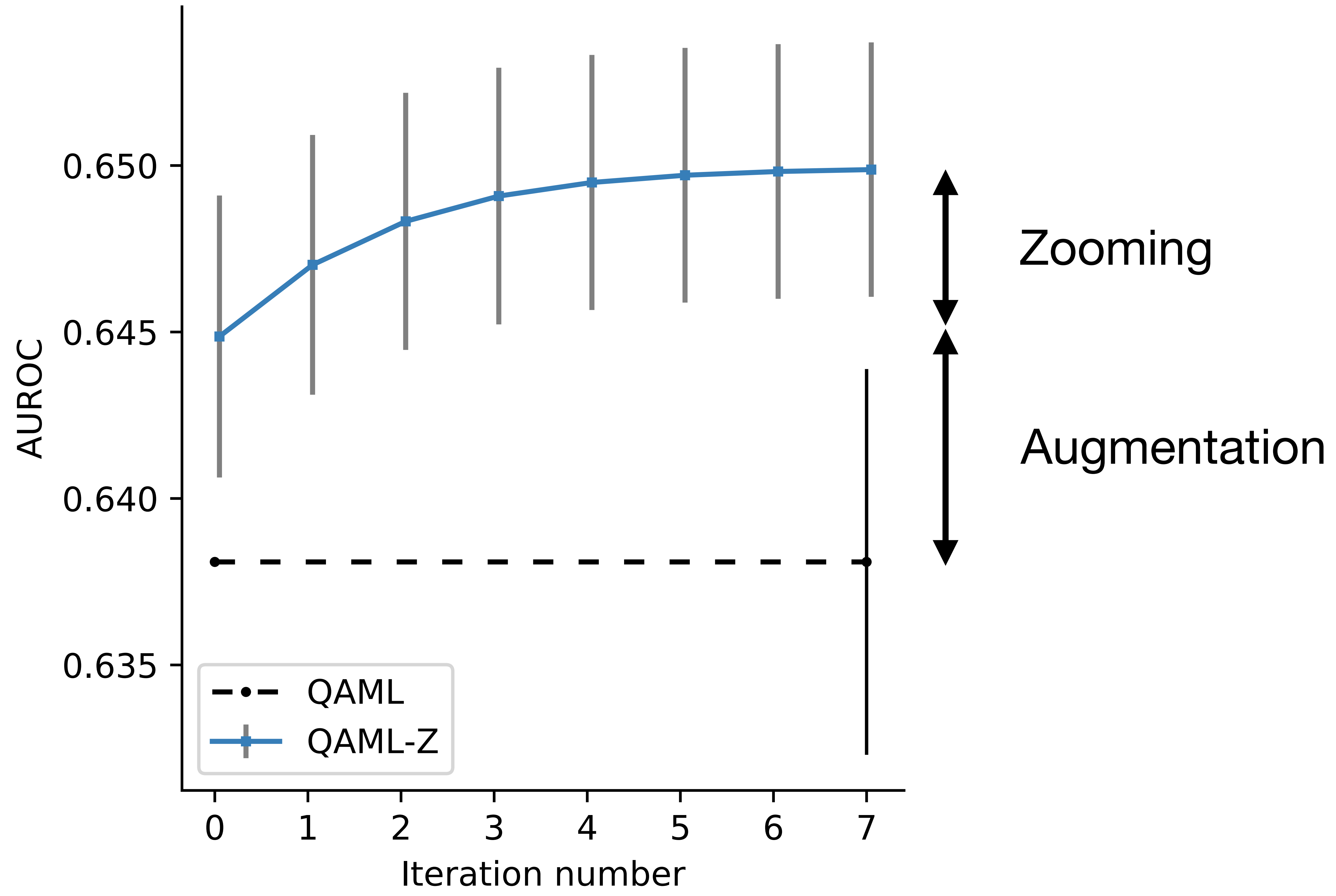
Both zooming and augmentation improve performance



Caltech

Higgs classification results

Both zooming and augmentation improve performance

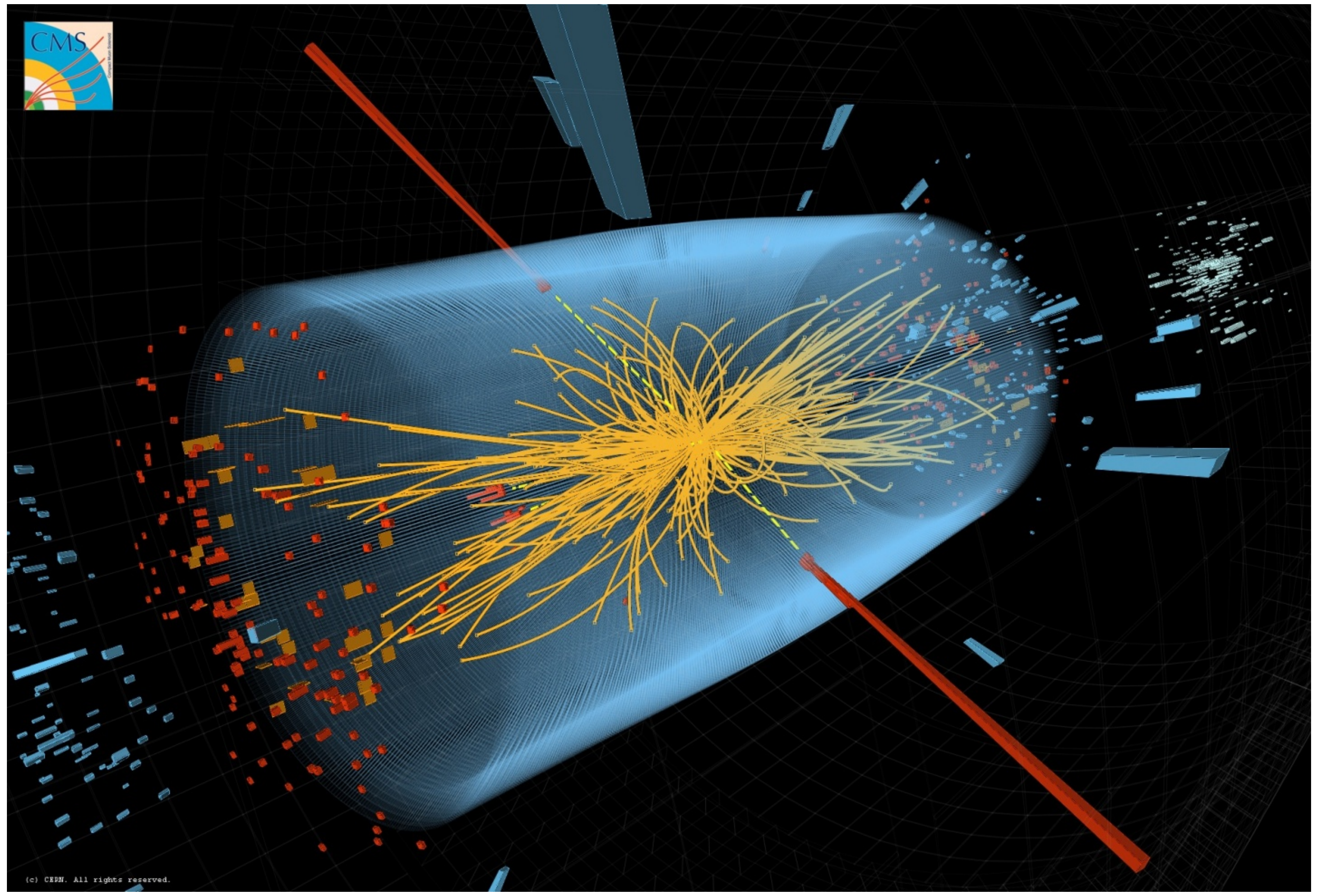


Caltech

Charged particle tracking

Track reconstruction

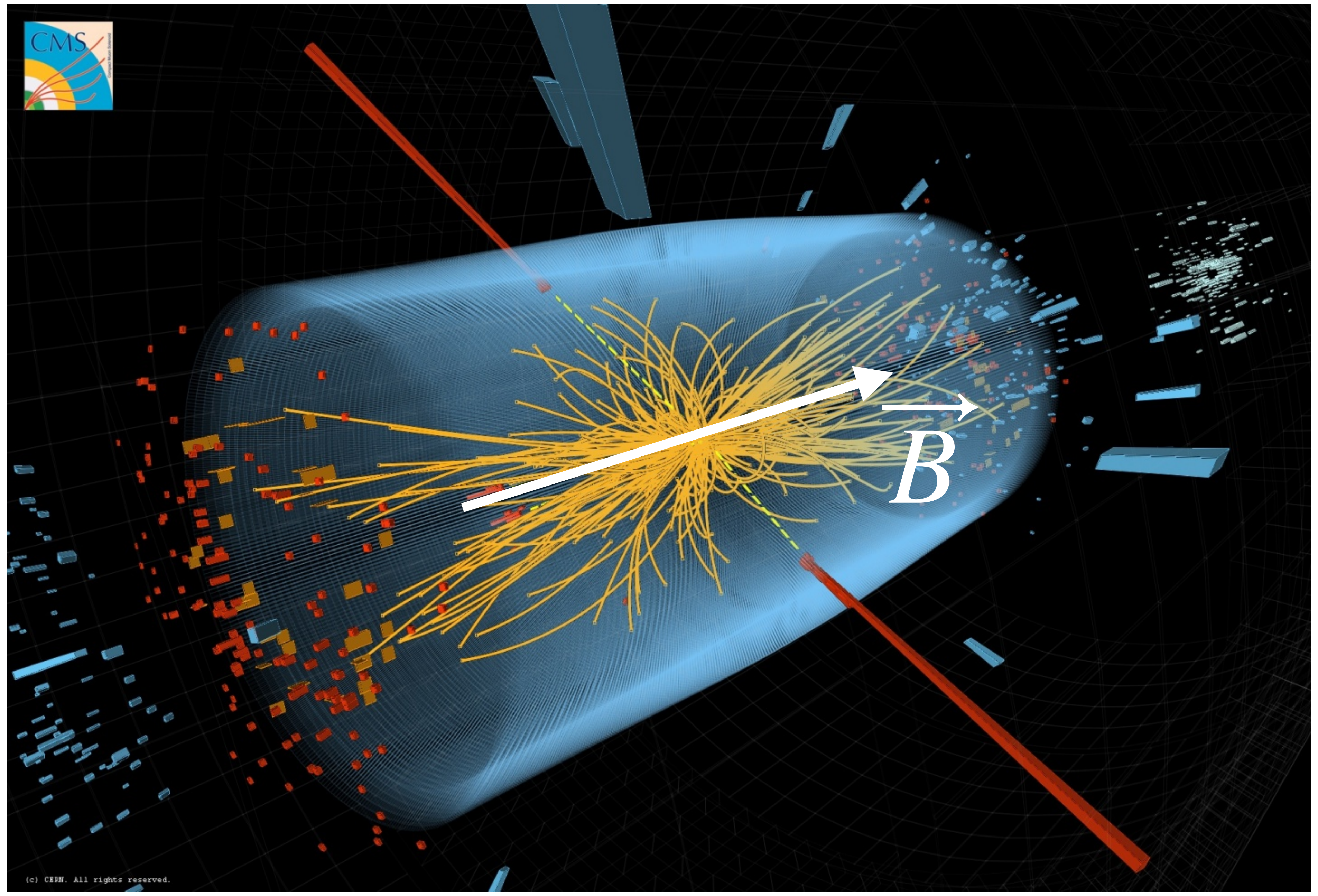
Cluster “hits” in a detector by particle instance



Caltech

Track reconstruction

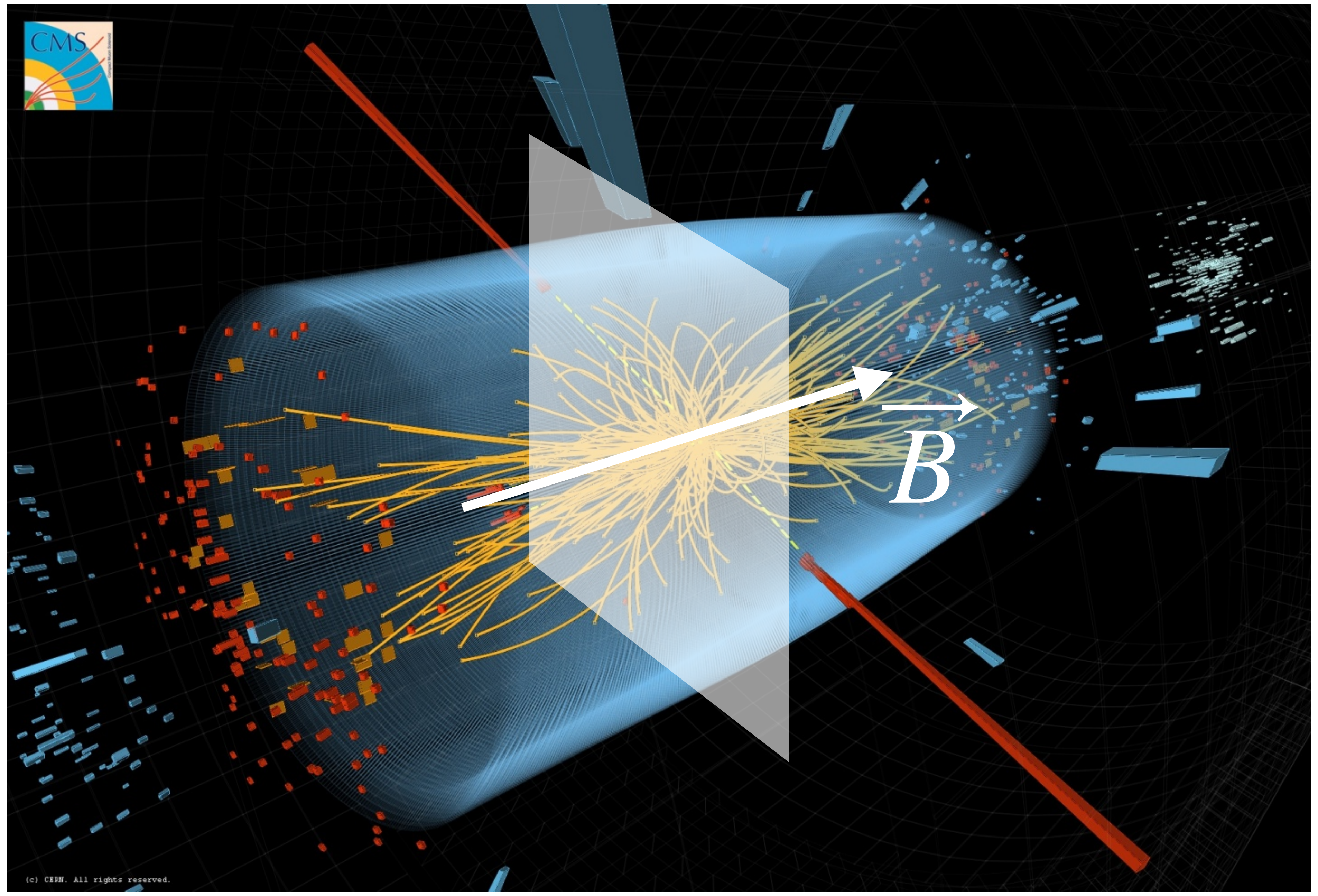
Cluster “hits” in a detector by particle instance



Caltech

Track reconstruction

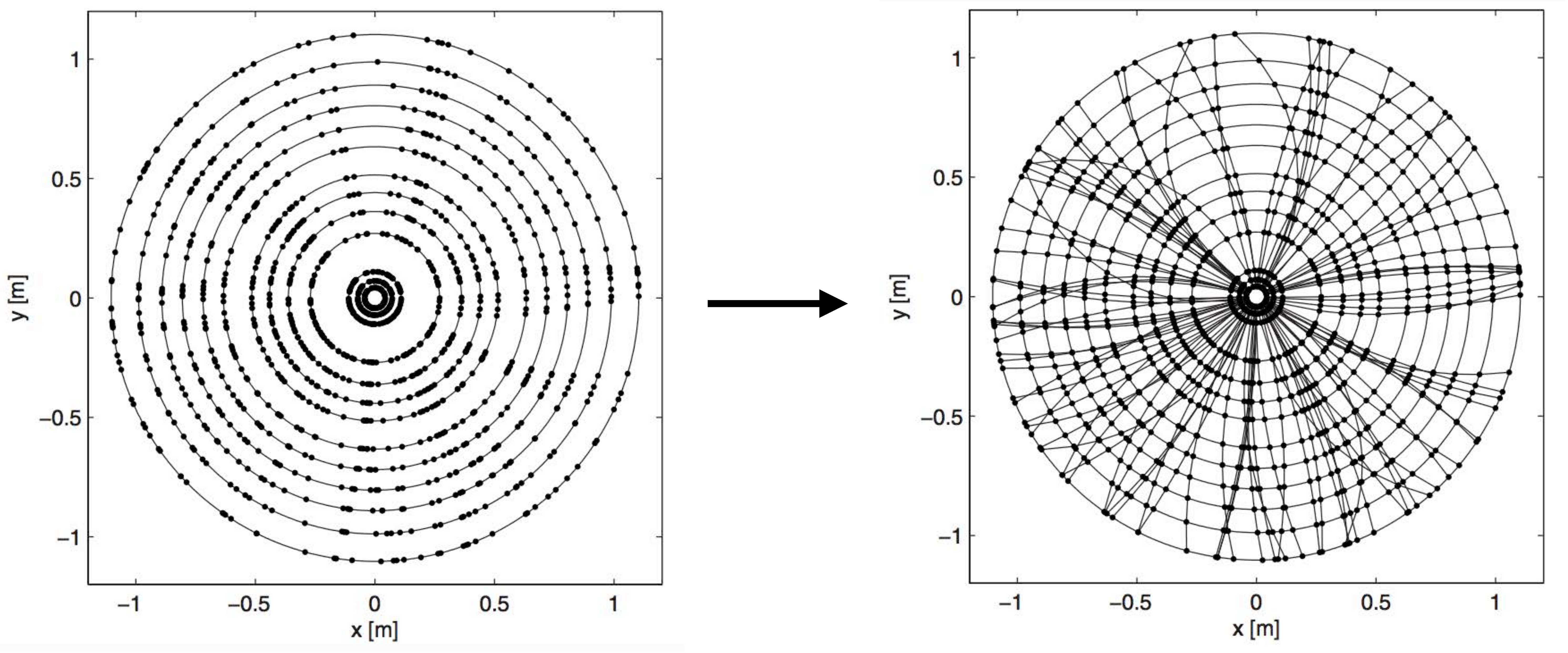
Cluster "hits" in a detector by particle instance



Caltech

Track reconstruction

Cluster “hits” in a detector by particle instance

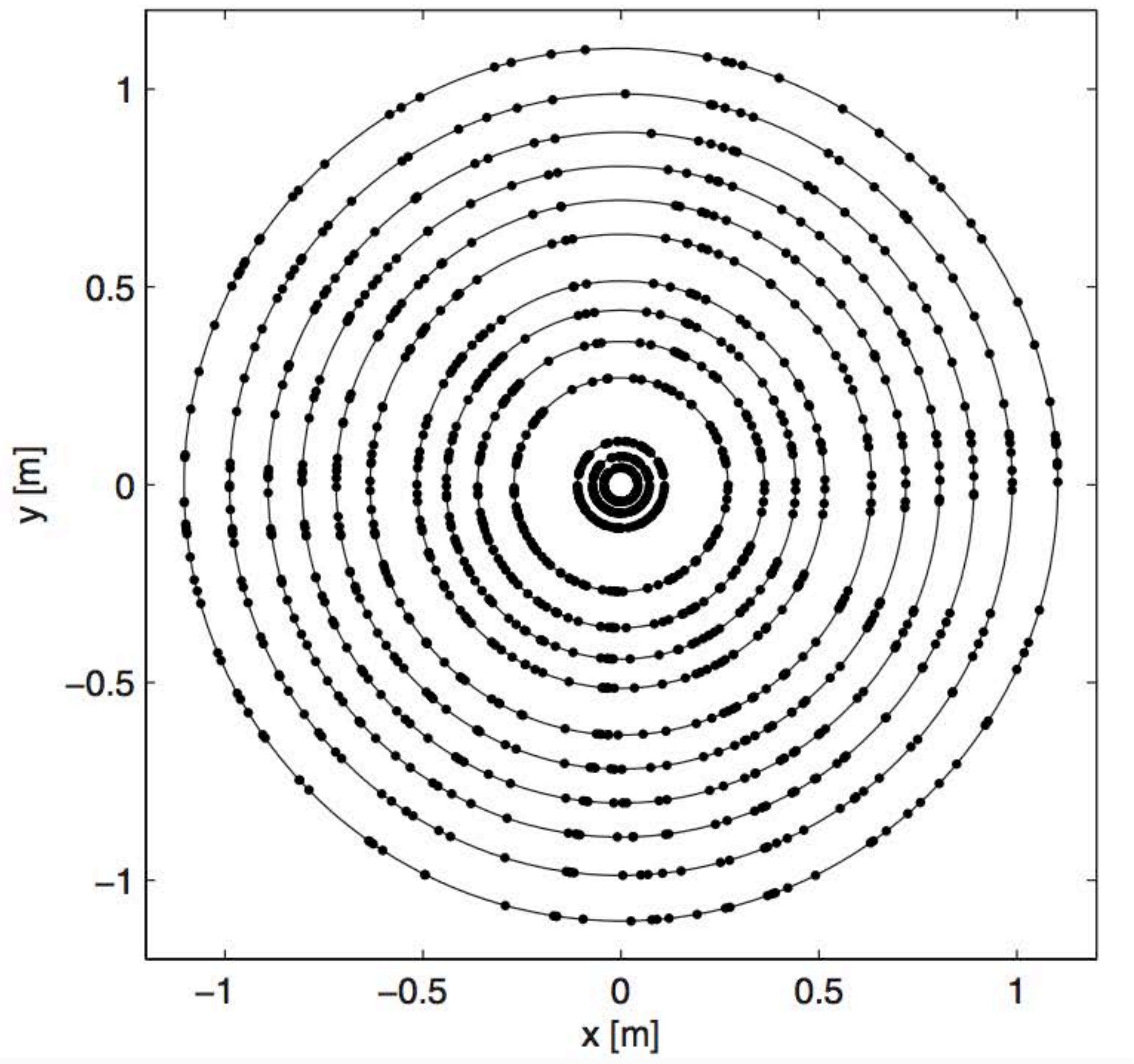


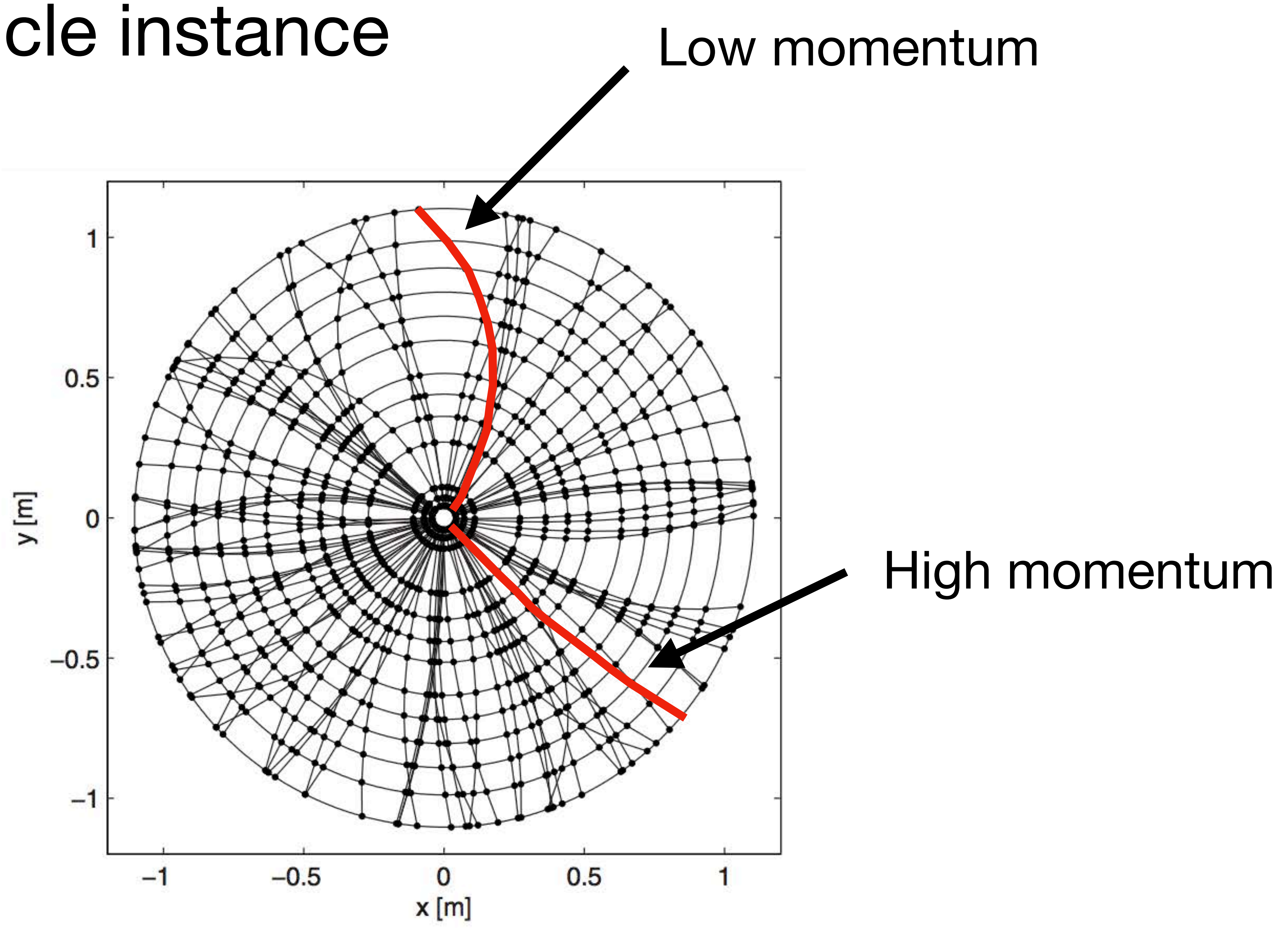
Strandlie, Are, and Rudolf Frühwirth. "Track and vertex reconstruction: From classical to adaptive methods." Reviews of Modern Physics 82.2 (2010): 1419.

Caltech

Track reconstruction

Cluster “hits” in a detector by particle instance





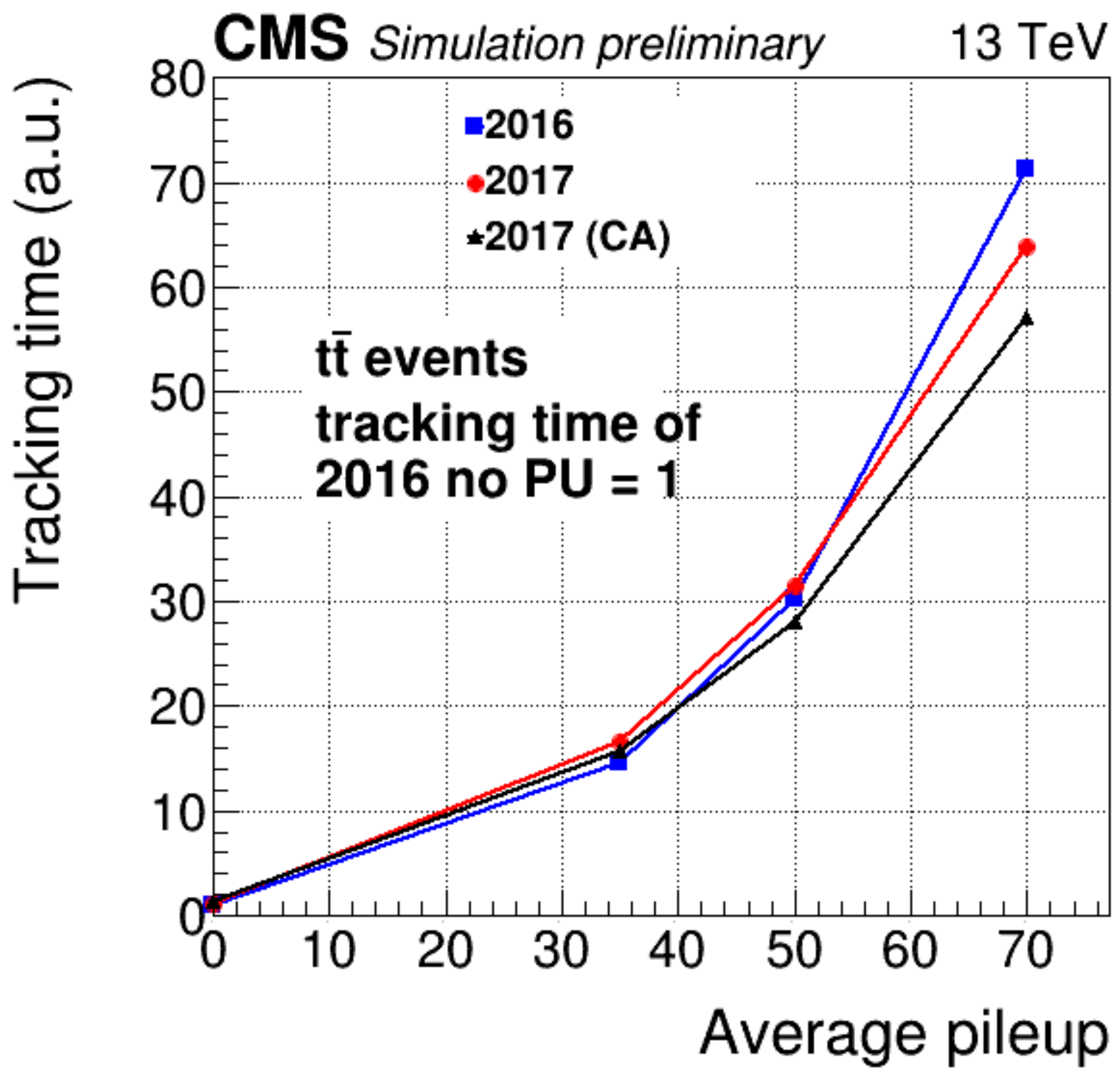
Strandlie, Are, and Rudolf Frühwirth. "Track and vertex reconstruction: From classical to adaptive methods." Reviews of Modern Physics 82.2 (2010): 1419.

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?

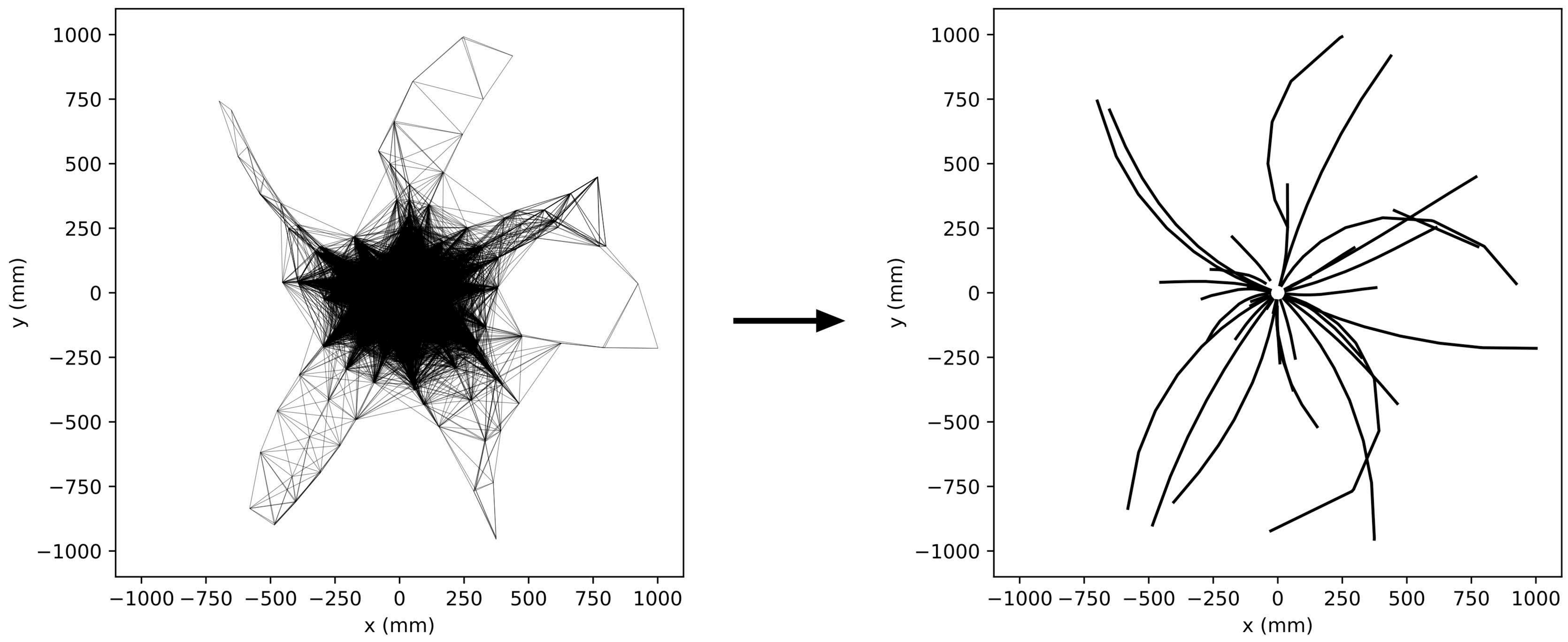


CMS Collaboration. "CMS Tracking POG Performance Plots For 2017 with PhaseI pixel detector." (2017).

Caltech

Ising model formulation

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



Caltech

Ising model formulation

A. Zlokapa, A. Anand, J.-R. Vlimant, J. M. Duarte, J. Job, D. Lidar, M. Spiropulu. "Charged particle tracking with quantum annealing-inspired optimization." arXiv:1908.04475 [quant-ph] (2019).

Affinity between edges *i* and *j*

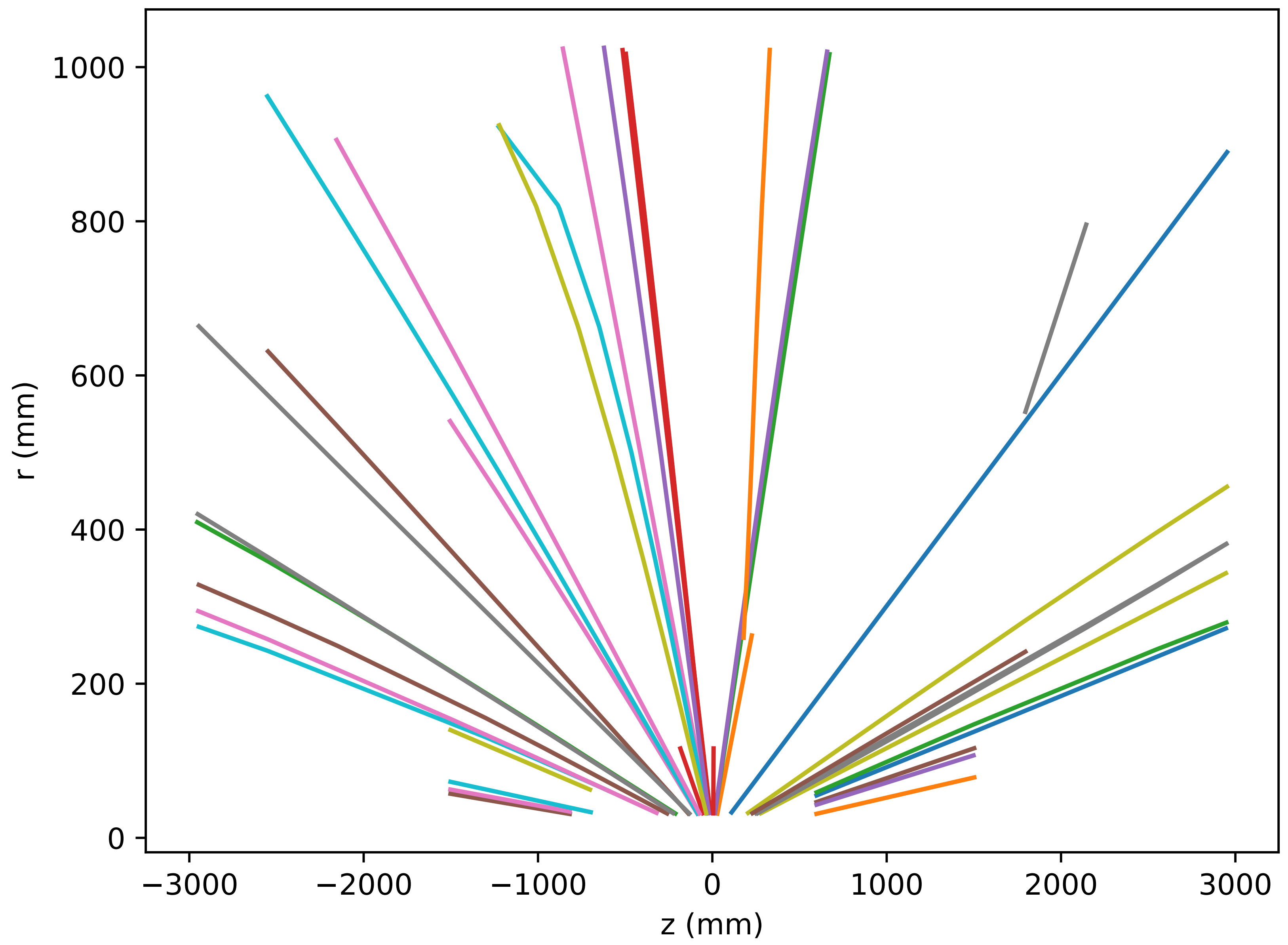
1 if edge is on; 0 if edge is off

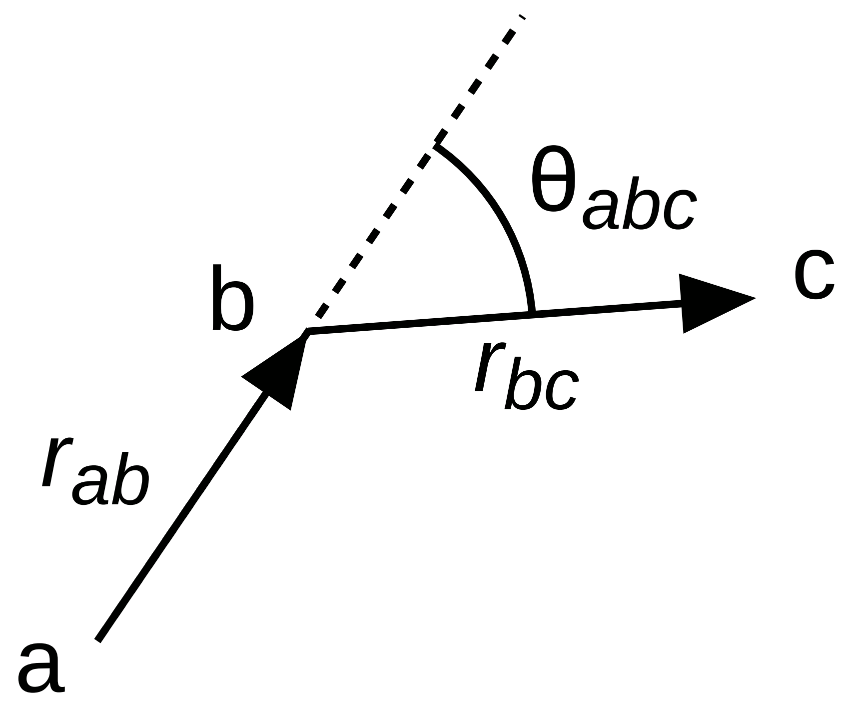
$$H_1 = -\sum_i \sum_{j>i} J_{ij} s_i s_j - \sum_i h_i s_i$$

Prior expectation on edge *i*

Ising model formulation

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



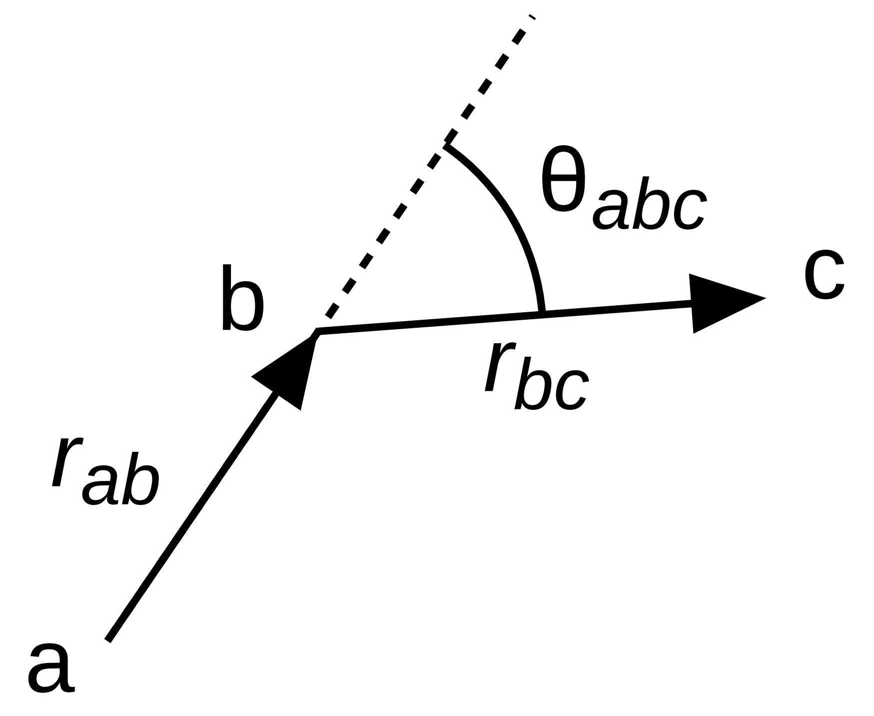


Caltech

Ising model formulation

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

$$-\left(\frac{\cos^{\lambda}(\theta_{abc})}{r_{ab}+r_{bc}}\right)s_{ab}s_{bc}$$



Caltech

Ising model formulation

Helical tracks

High-momentum bias

Track bifurcation penalty

Global edge penalty

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

Beam spot geometry

Edge orientation probability
(Gaussian kernel density estimation)

Dimension challenge

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

- Divide into 16 sectors: ~10^5 edges
- Remove edges with Gaussian KDE: ~10^3 edges

Dimension challenge

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

- Divide into 16 sectors: ~10^5 edges
- Remove edges with Gaussian KDE: ~10^3 edges

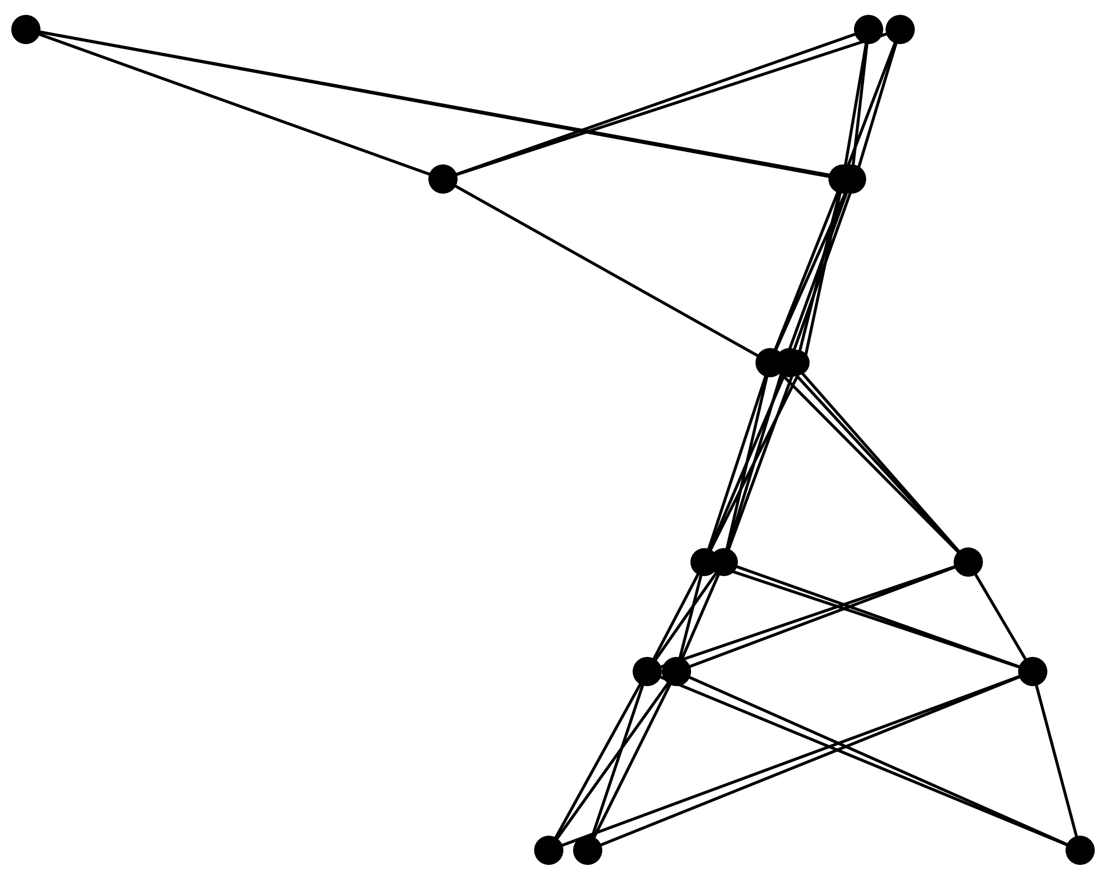
D-Wave 2X: 33 fully-connected qubits

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

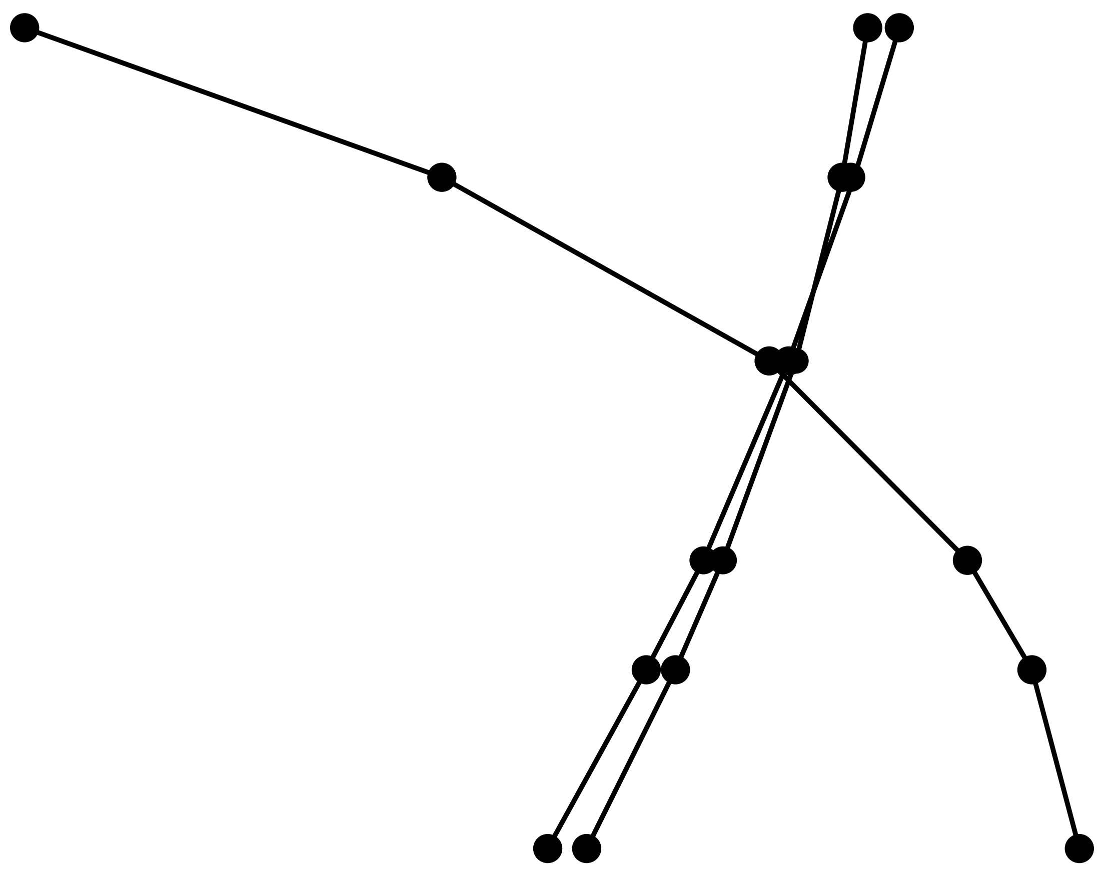
Caltech

Dimension challenge

Disjoint sub-graphs: prune and divide



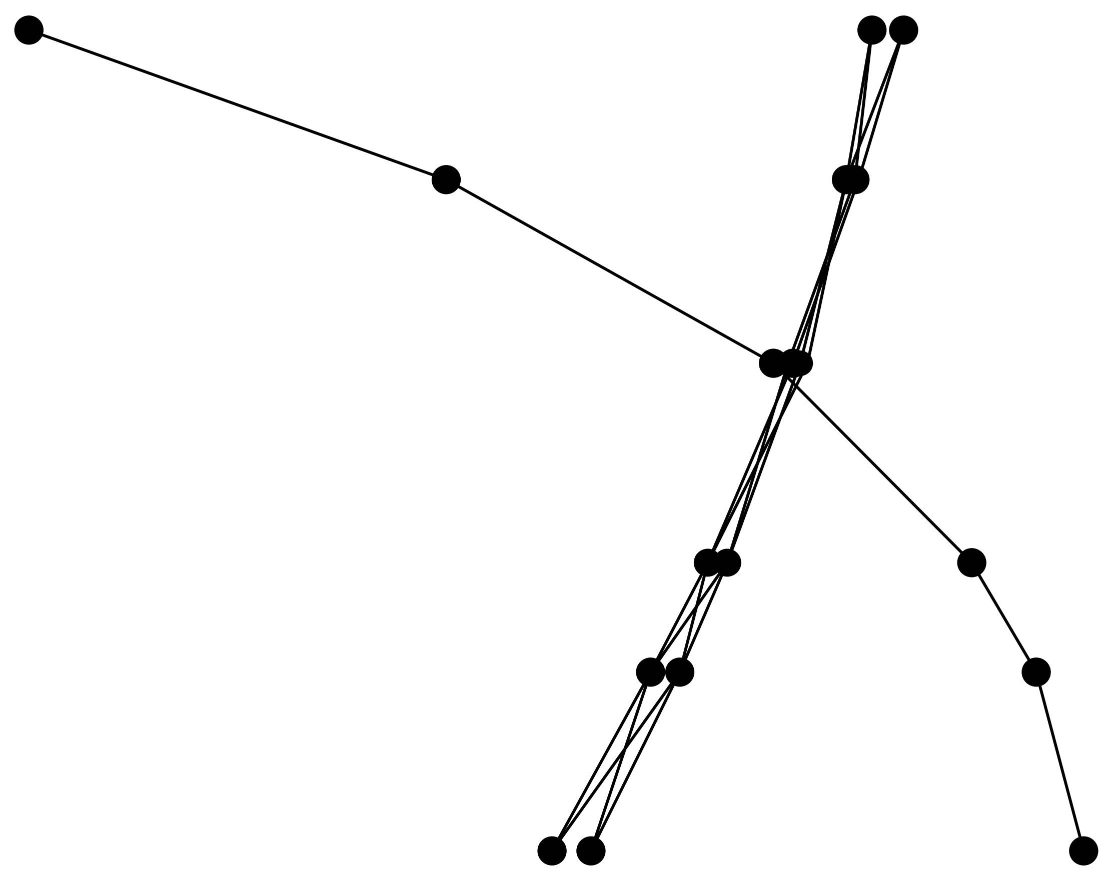
Initial graph



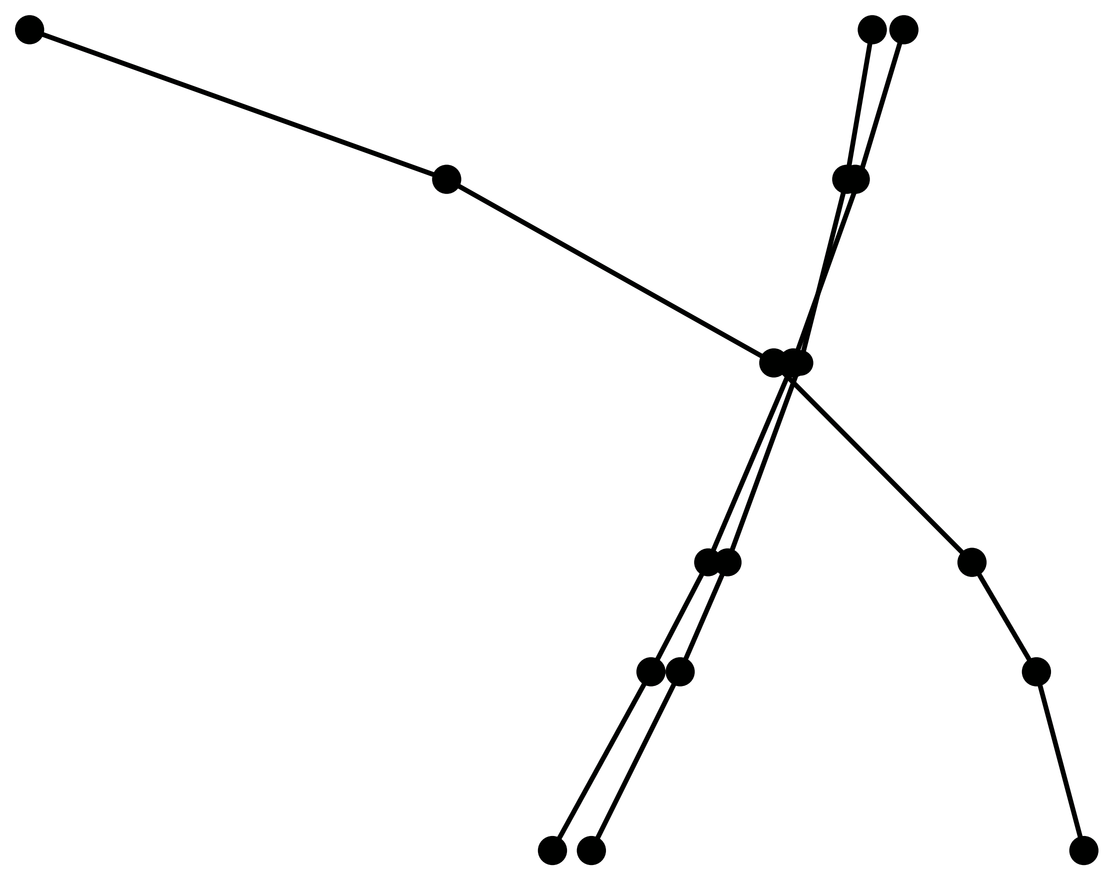
True graph

Dimension challenge

Disjoint sub-graphs: prune and divide



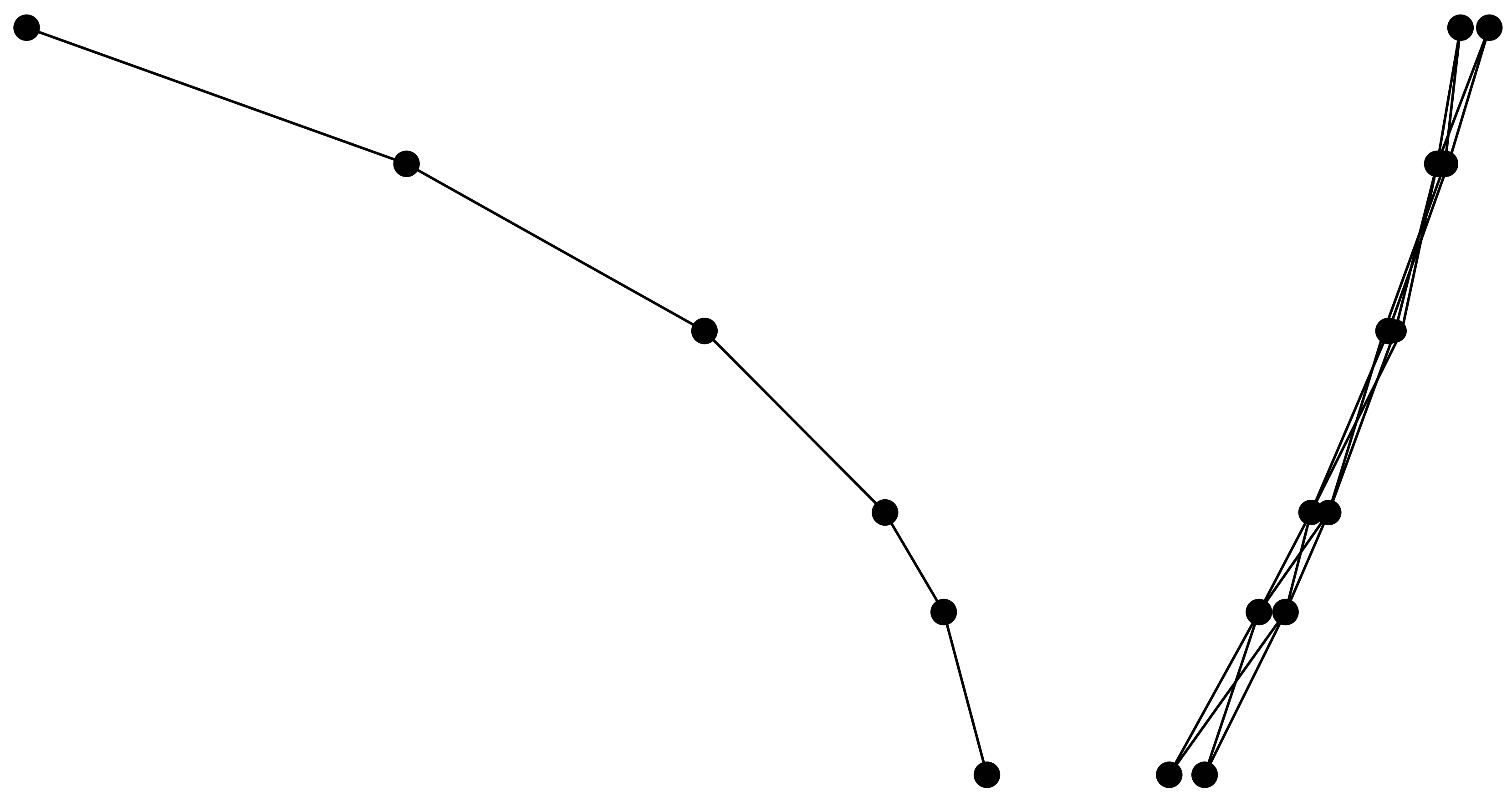
Pruned graph



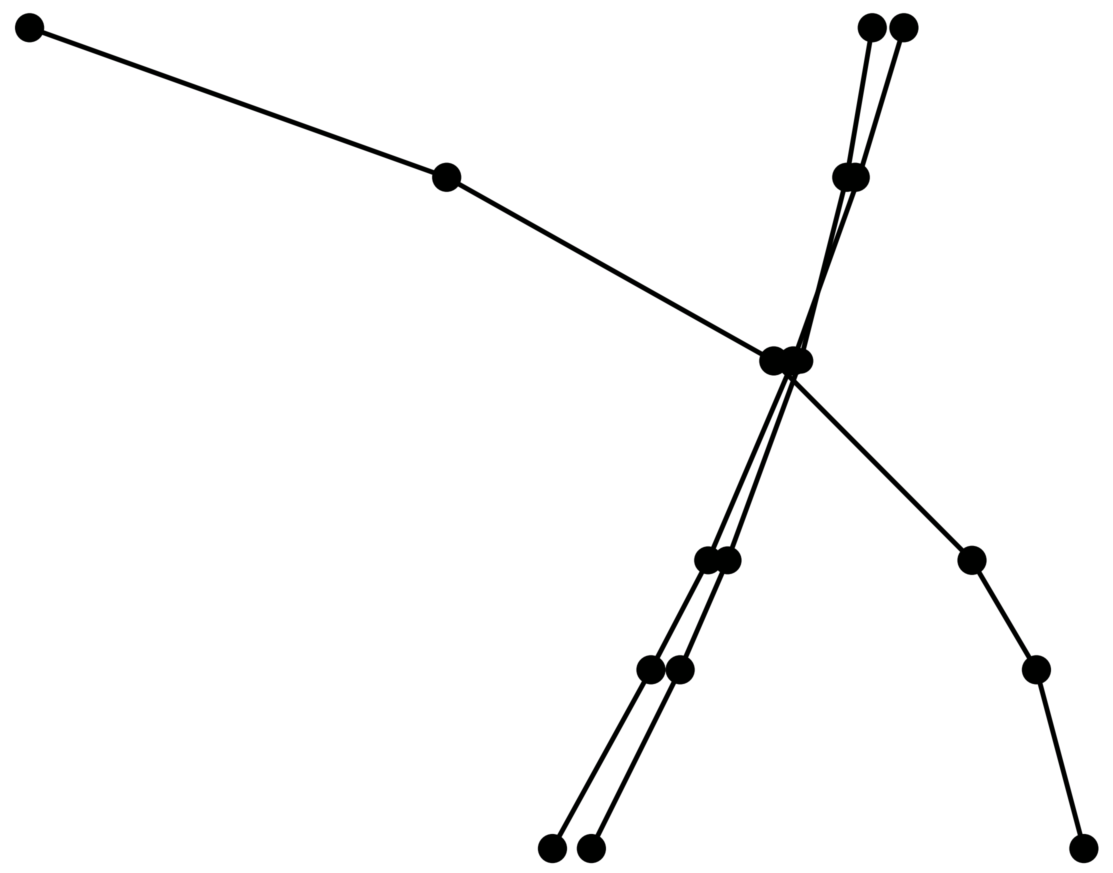
True graph

Dimension challenge

Disjoint sub-graphs: prune and divide



Disjoint sub-graphs



True graph

Caltech

Dimension challenge

Higgs event at LHC: 10^3 to 10^4 detector hits $\Longrightarrow$ $\sim 10^7$ 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: ~ 10 problems per sector

Dimension challenge

Higgs event at LHC: 10^3 to 10^4 detector hits $\Longrightarrow$ $\sim 10^7$ 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: ~10 problems per sector

Result: ~100 Ising model variables on ~100 qubits

Caltech

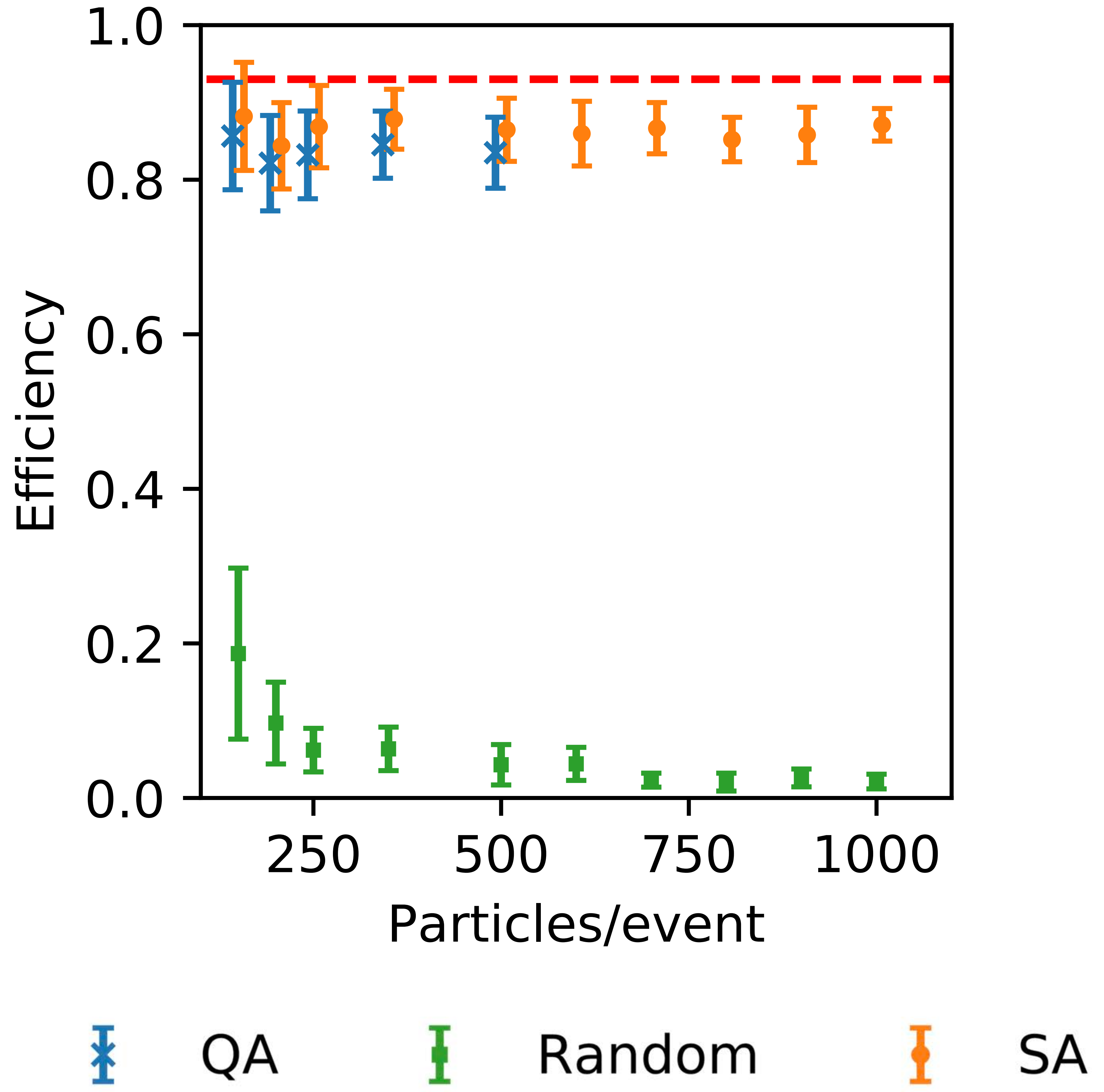
Results

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

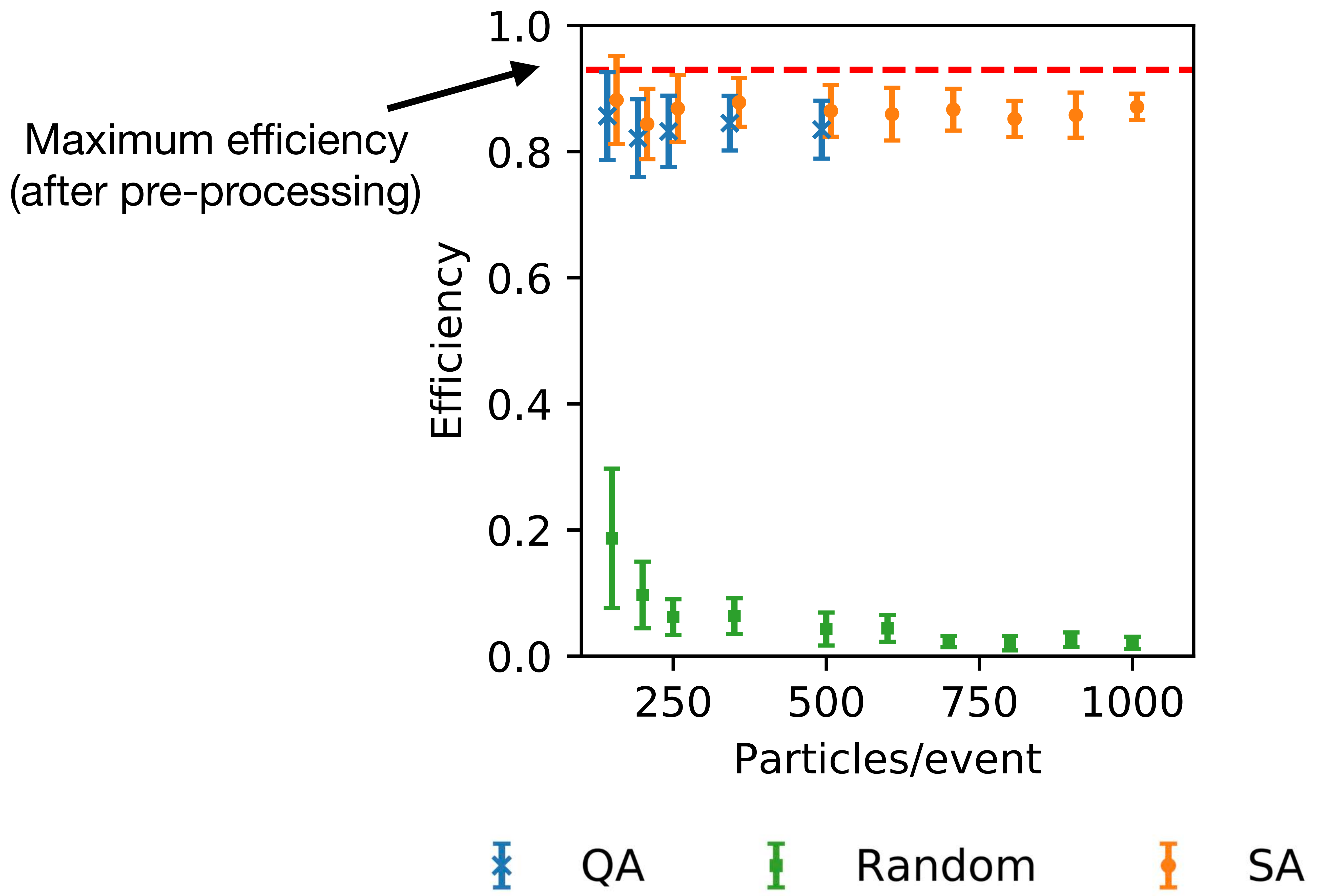
$$\text{efficiency} = \frac{\text{\# true tracks reconstructed}}{\text{\# true tracks}}$$

$$\text{purity} = \frac{\text{\# true tracks reconstructed}}{\text{\# tracks reconstructed}}$$

Results

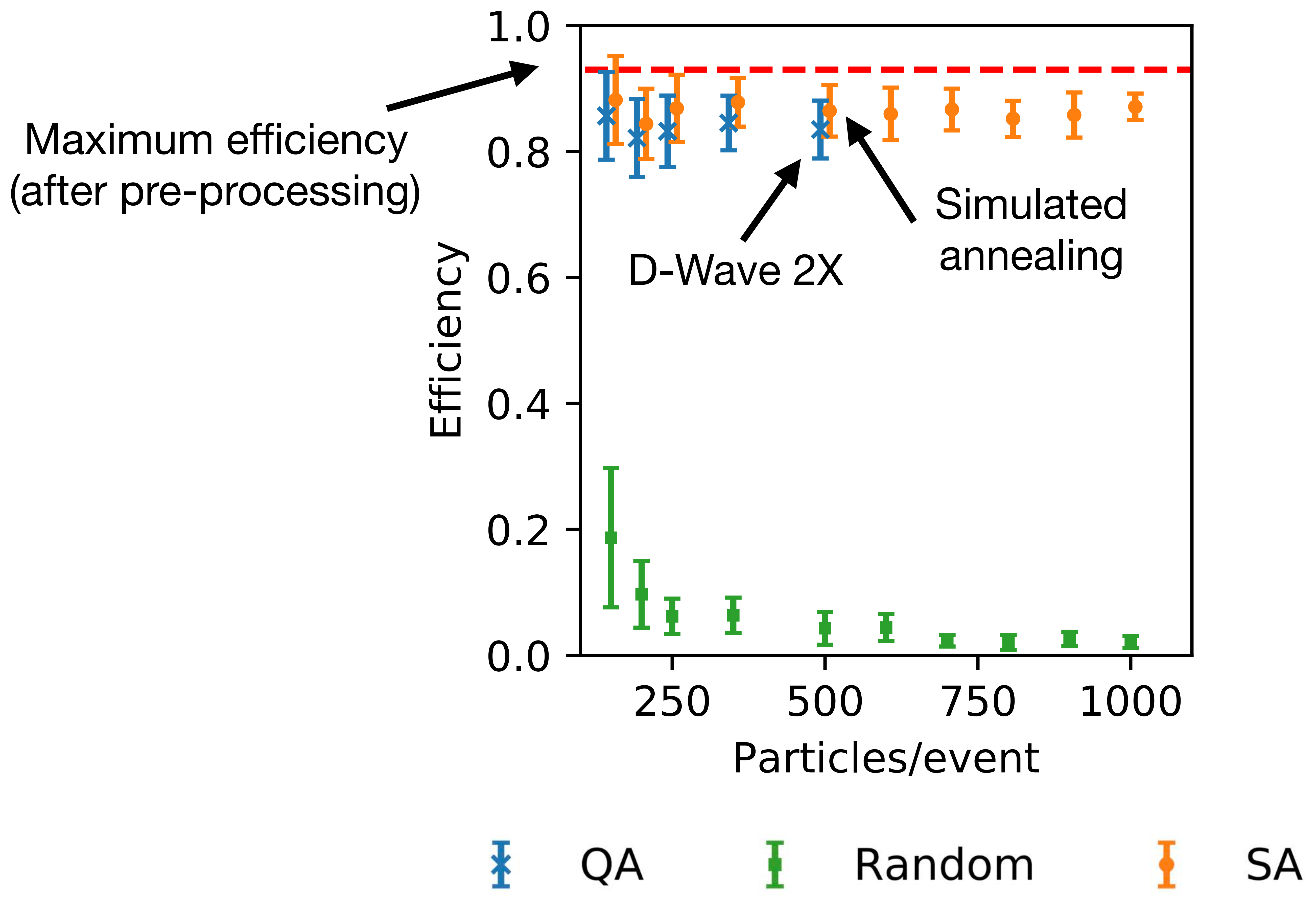

1.0
0.8
0.6
0.4
0.2
0.0
Efficiency
250
500
750
1000
Particles/event
QA
Random
SA

Results



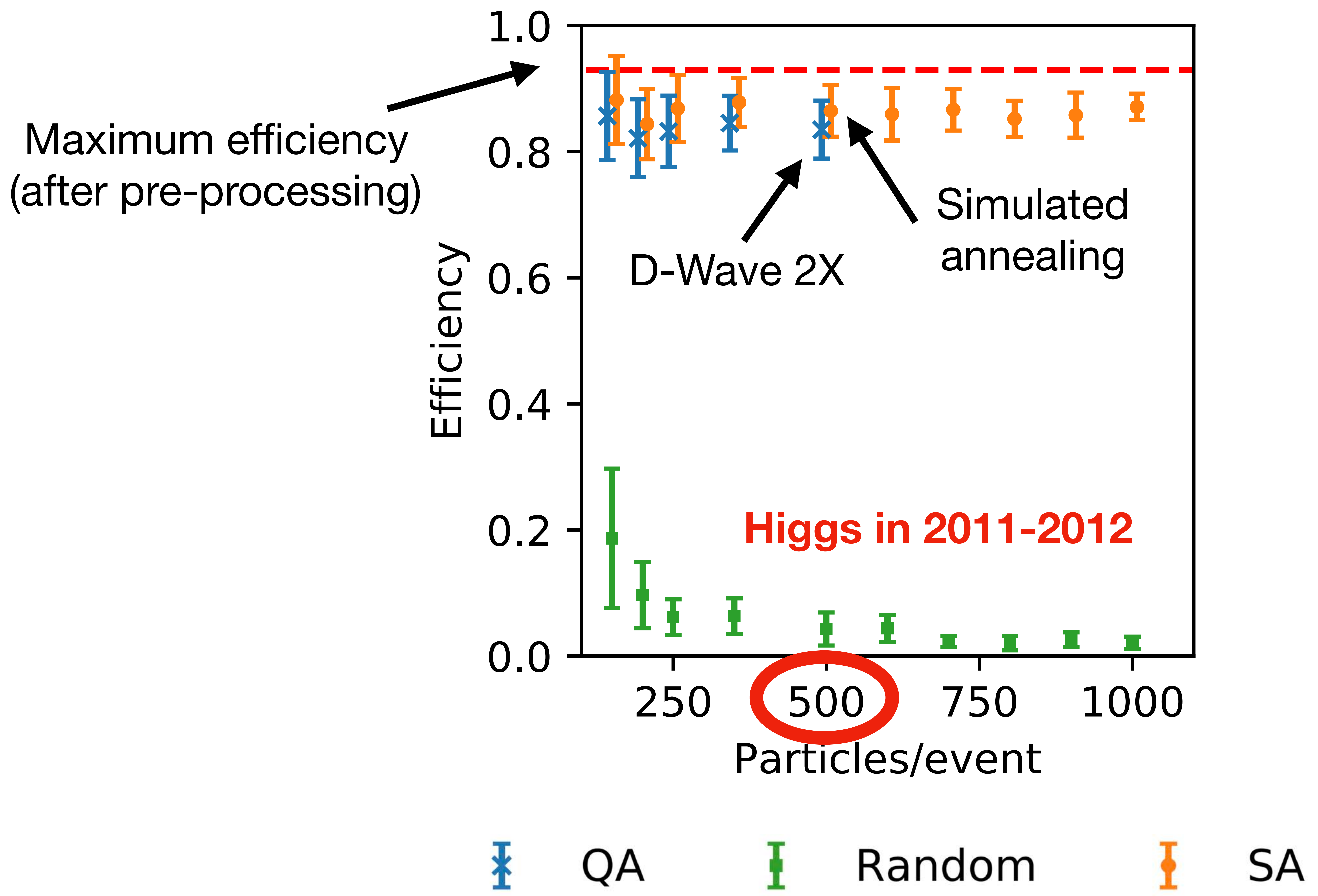
Caltech

Results



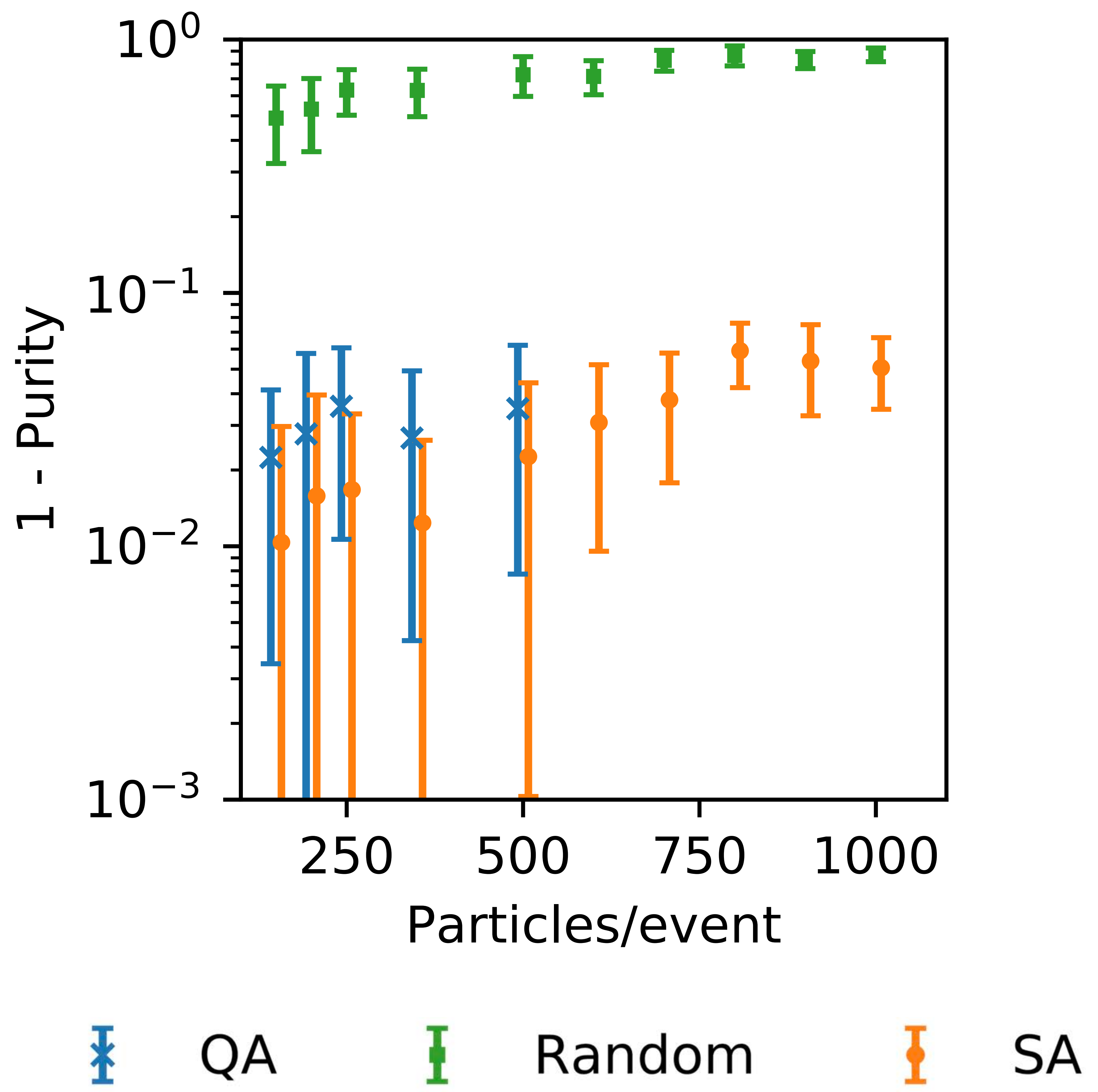
Caltech

Results



Caltech

Results

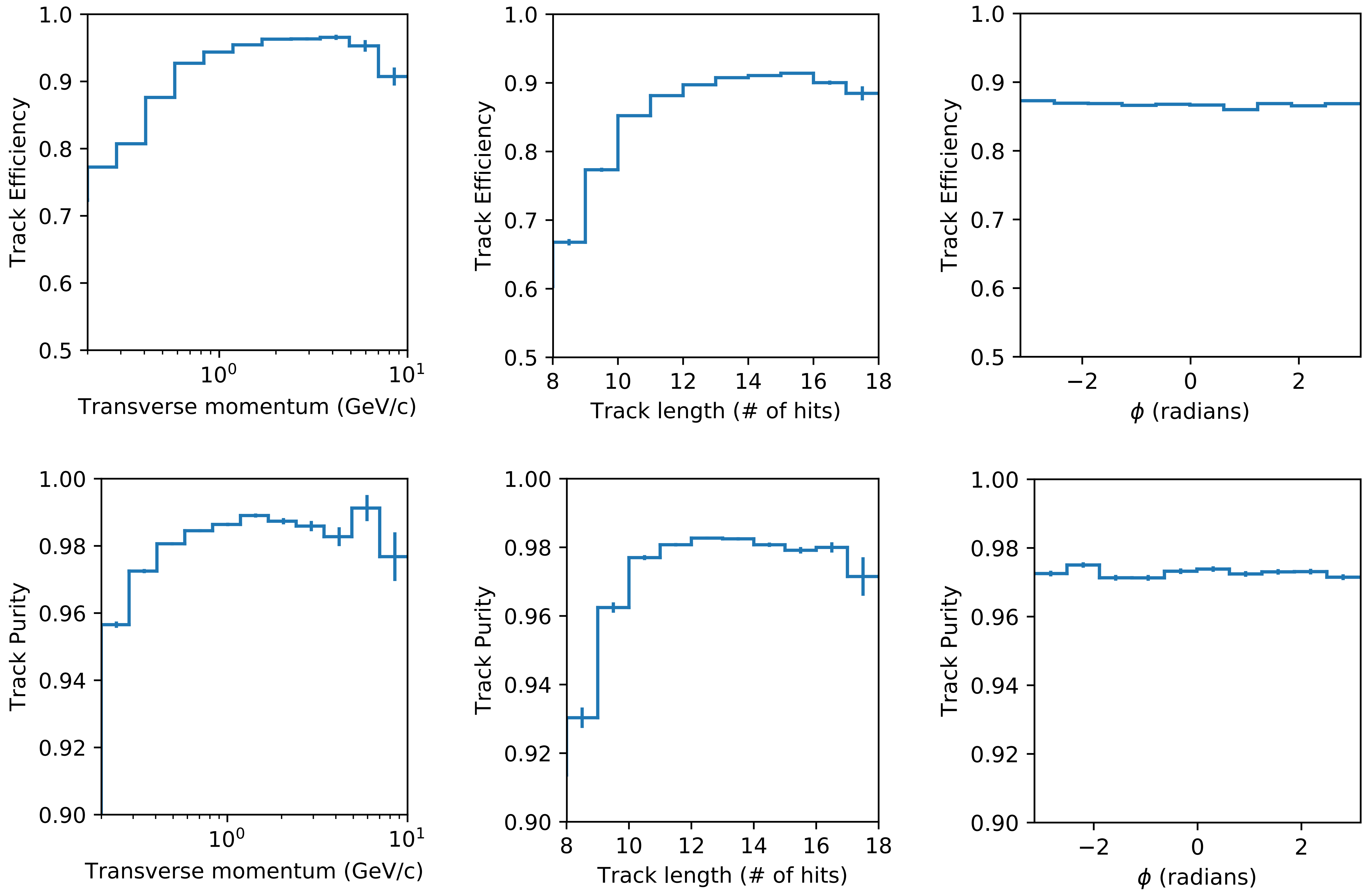


Caltech

Results

Efficiency

Purity

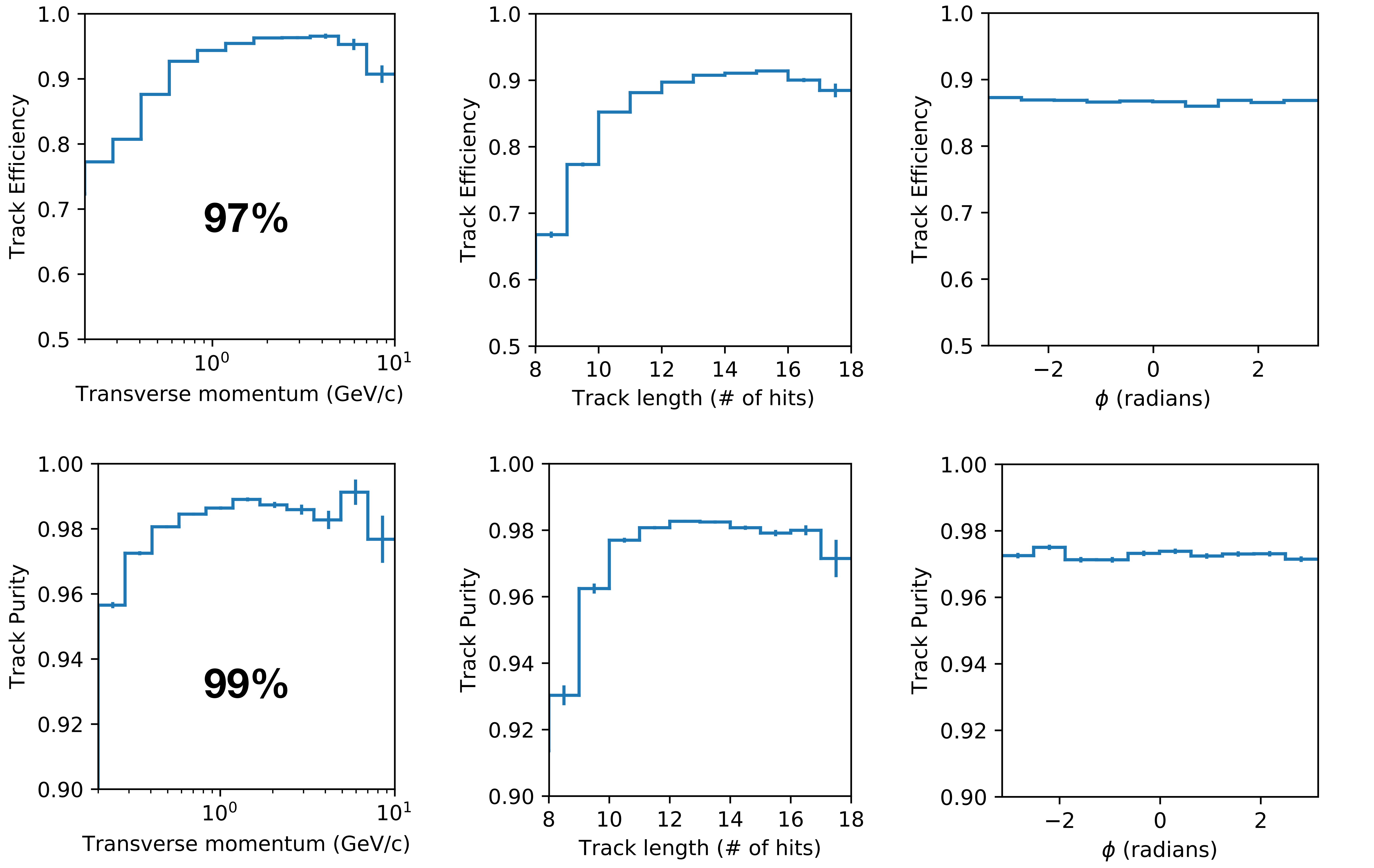


Caltech

Results

Efficiency

Purity



Caltech

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

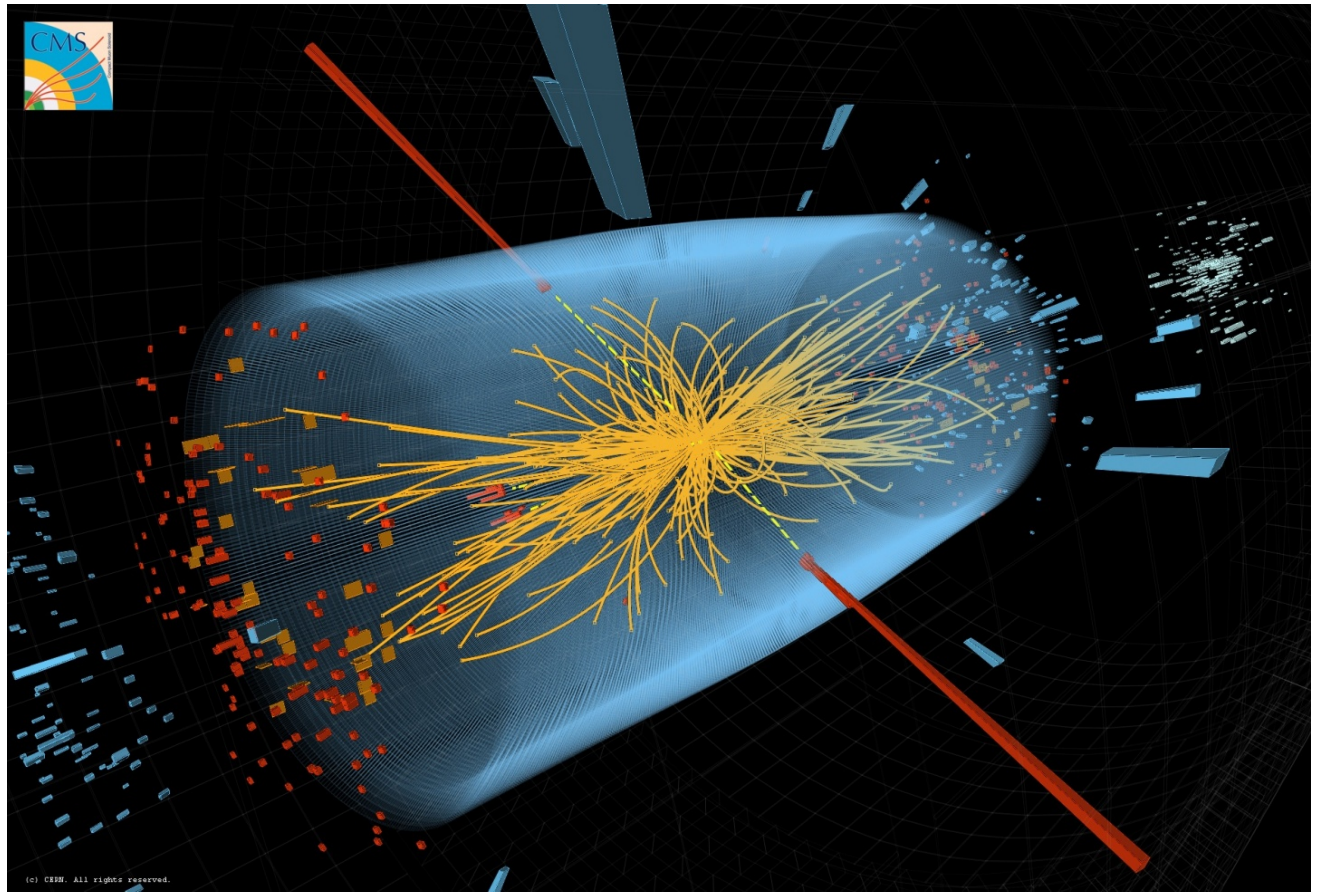
Competitive results with state-of-the-art classical algorithms

Thank you

Supplementary slides

Higgs problem construction

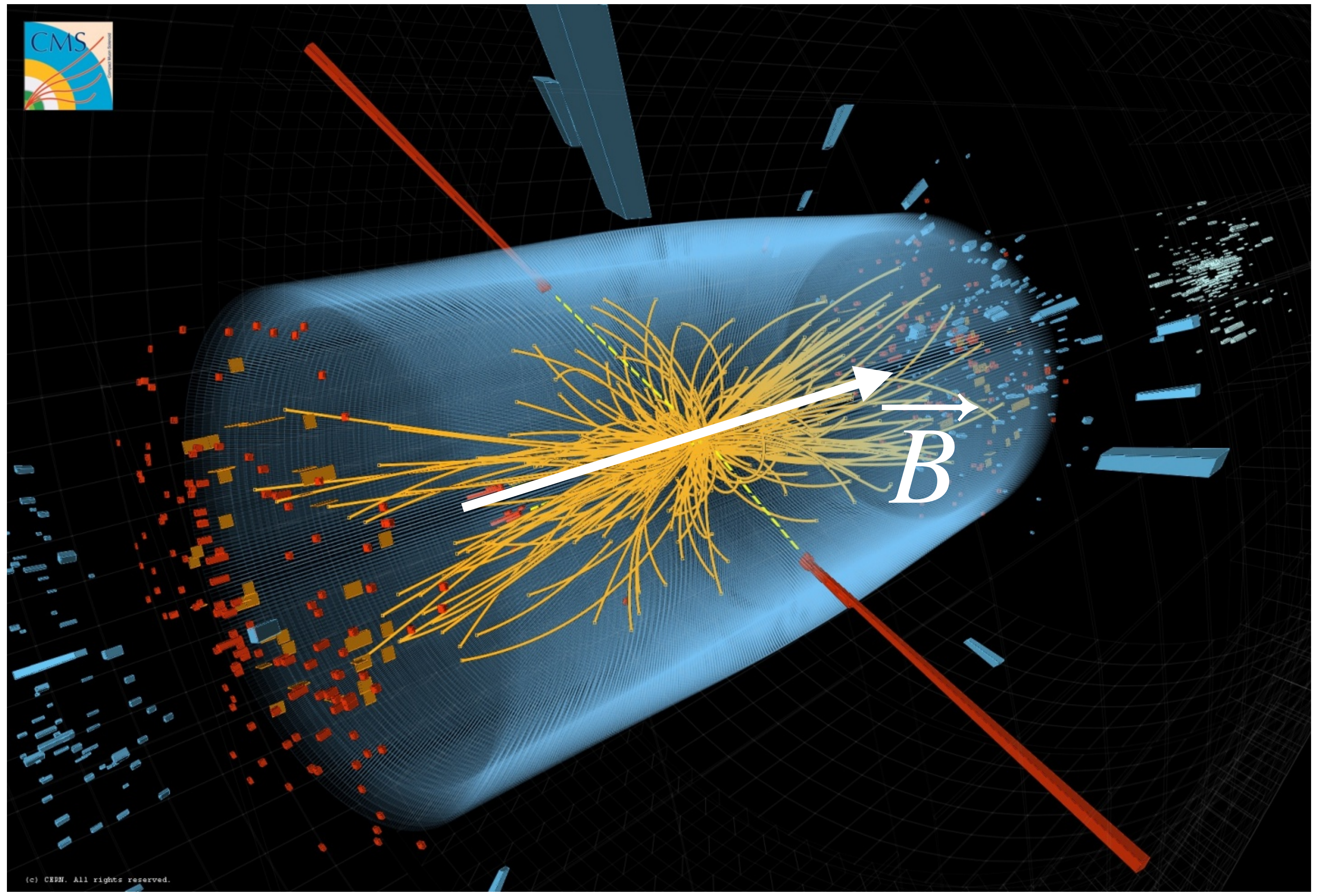
Can we “rediscover” the Higgs boson with QAML?



Caltech

Higgs problem construction

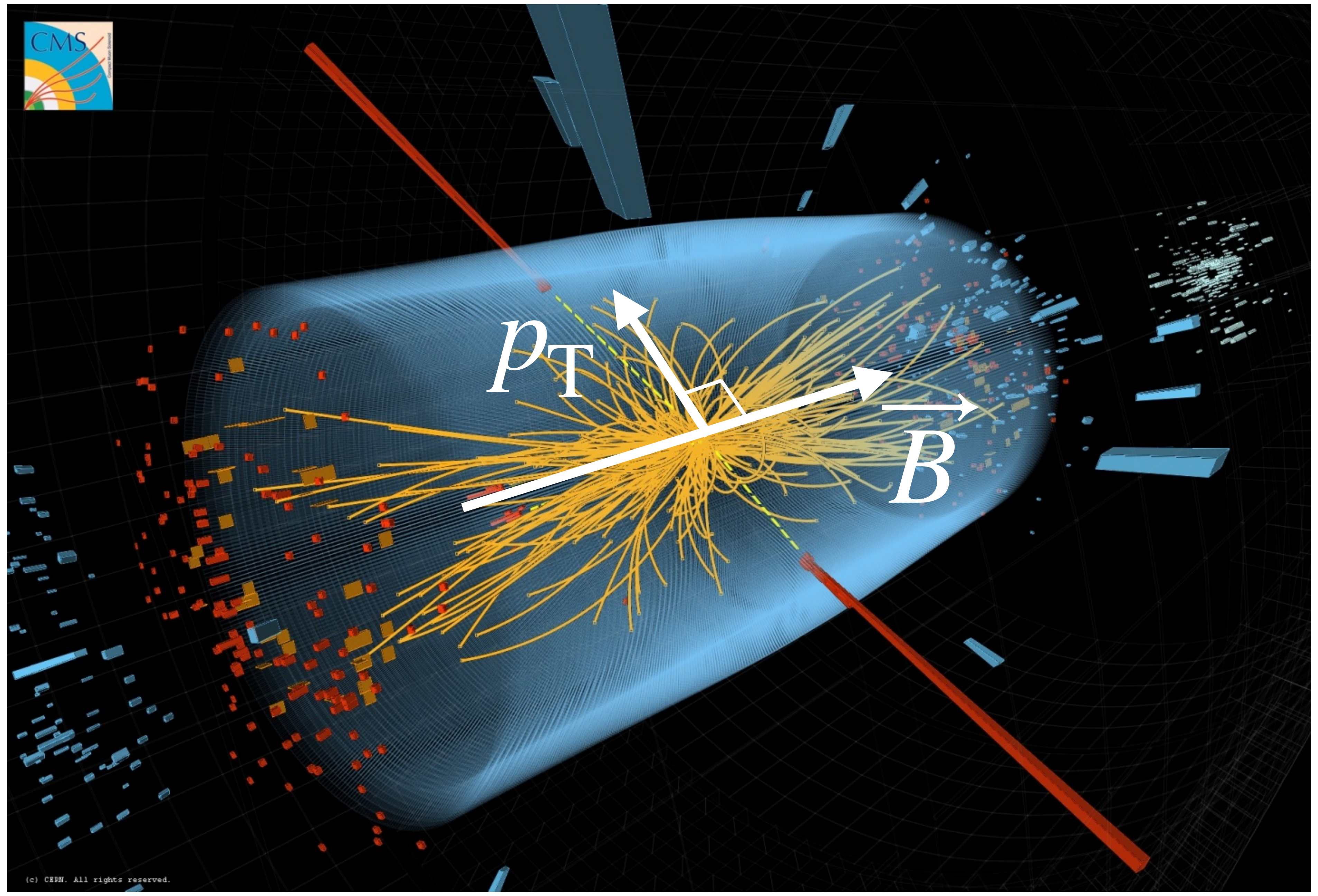
Can we “rediscover” the Higgs boson with QAML?



Caltech

Higgs problem construction

Can we “rediscover” the Higgs boson with QAML?



Caltech

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 $c_{il}(\mathbf{x}_\tau)$ for each original classifier $c_i(\mathbf{x}_\tau)$.

Anneal for iterations $t = 0, 1, 2, \ldots, T-1$.

Caltech

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:

$$C_{il} = \sum_{\tau=1}^{S} c_{il}(\mathbf{x}_\tau) y_\tau \qquad C_{ijl} = \sum_{\tau=1}^{S} c_{il}(\mathbf{x}_\tau) c_{jl}(\mathbf{x}_\tau)$$

Caltech

QAML-Z algorithm: Hamiltonian

Each iteration t, anneal:

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

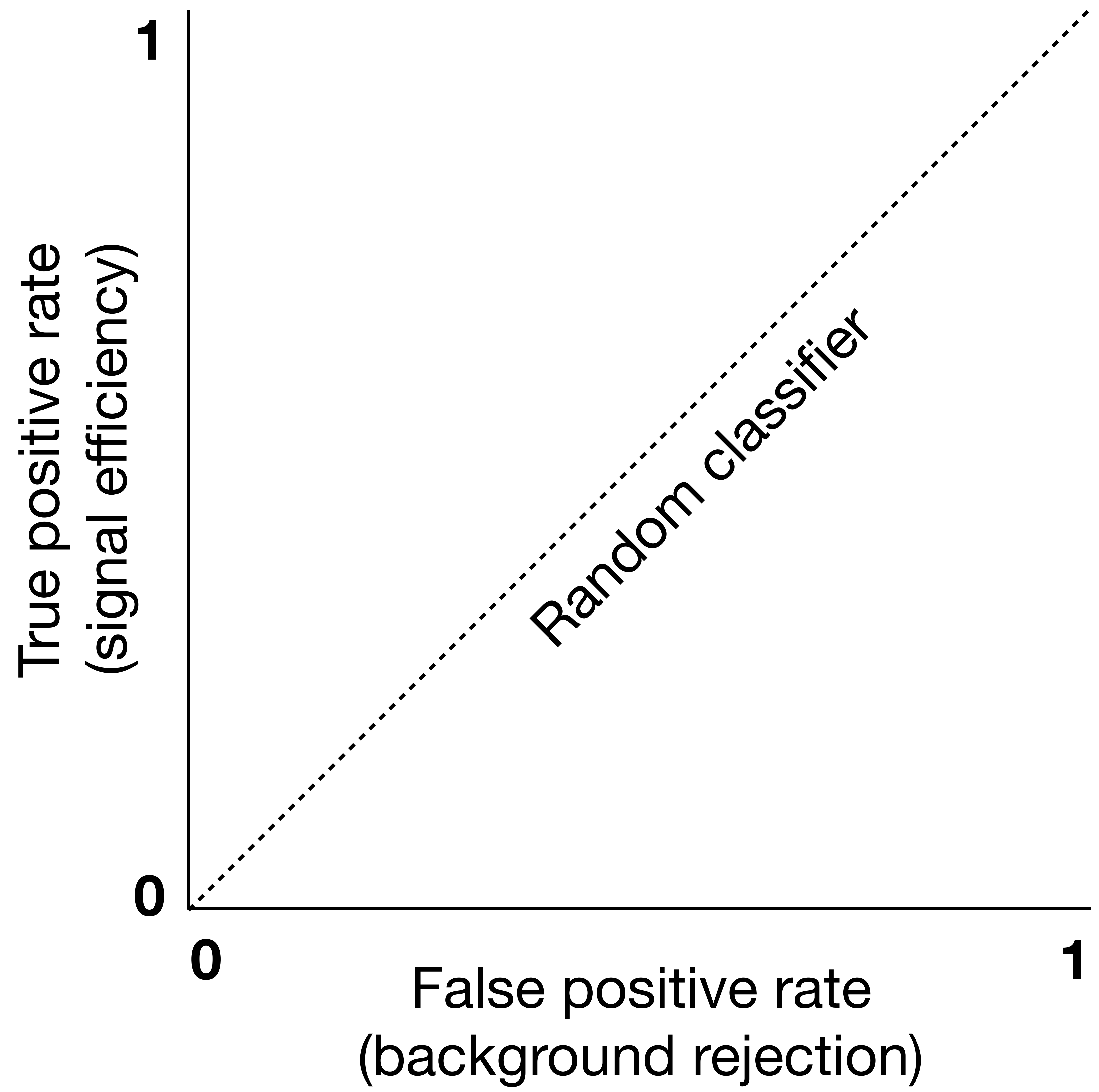
and update continuous weights from spins:

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

Caltech

ROC curve

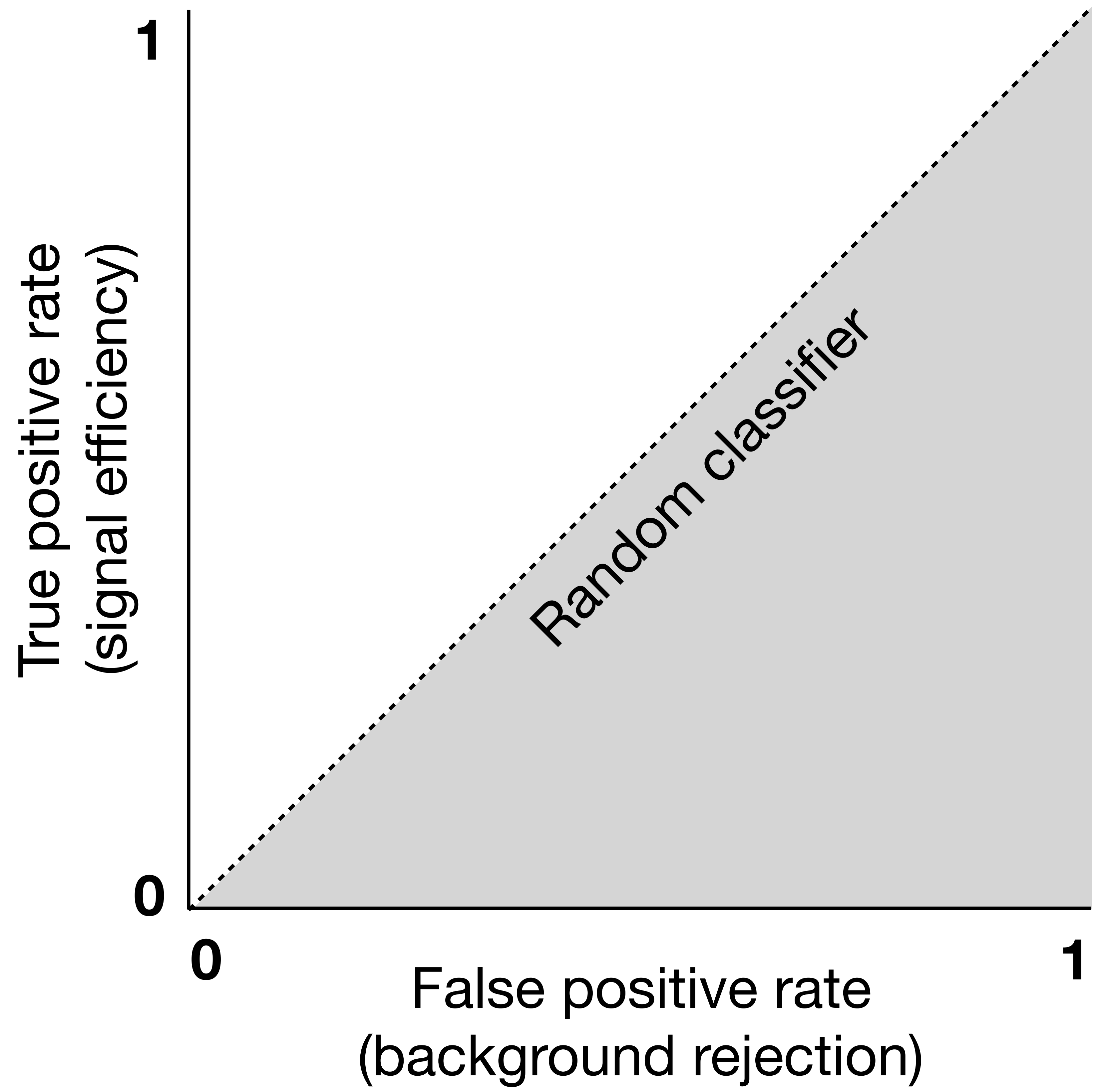
Metric of performance: area under receiver operating characteristic (ROC) curve



Caltech

ROC curve

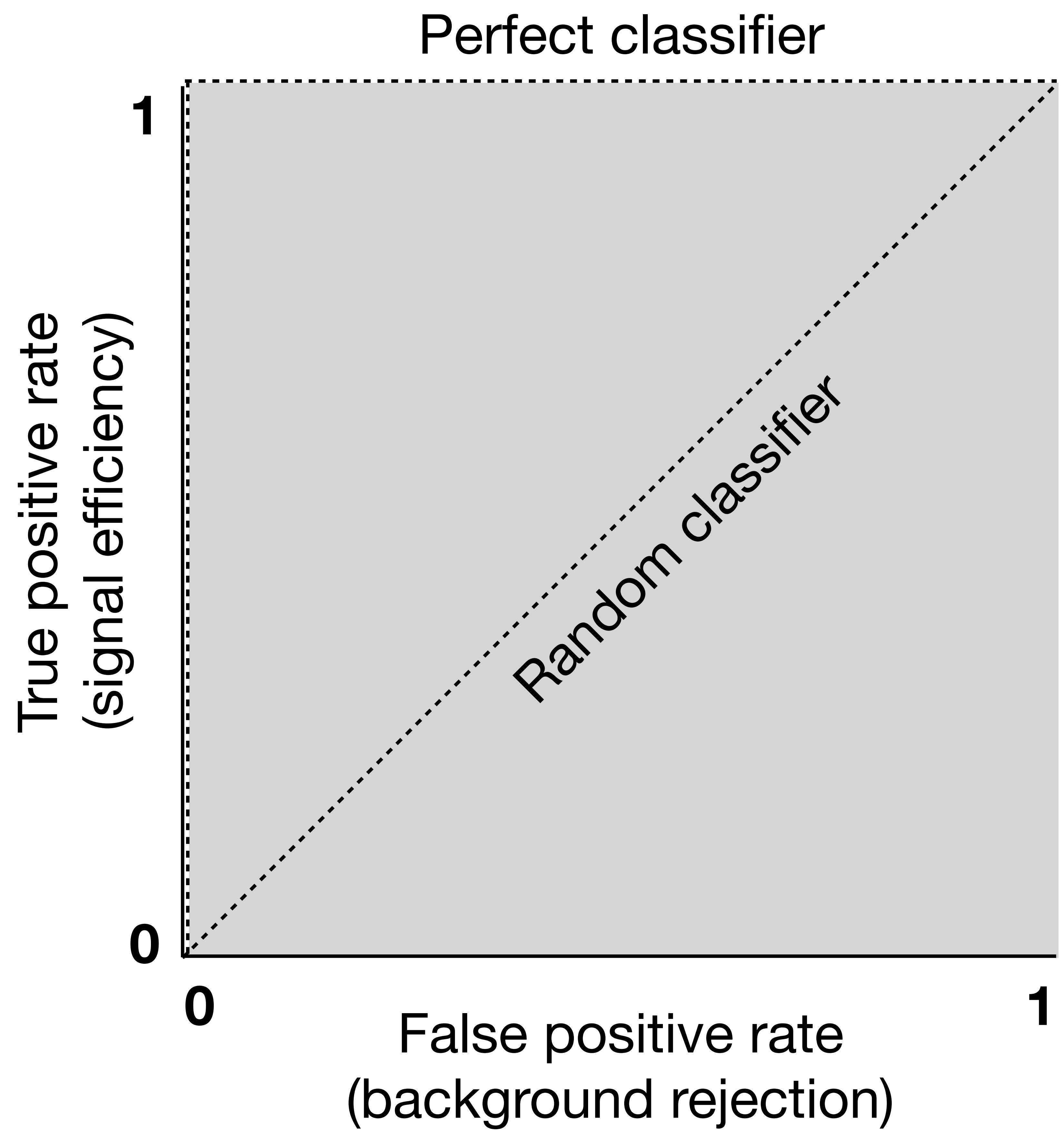
Metric of performance: area under receiver operating characteristic (ROC) curve



Caltech

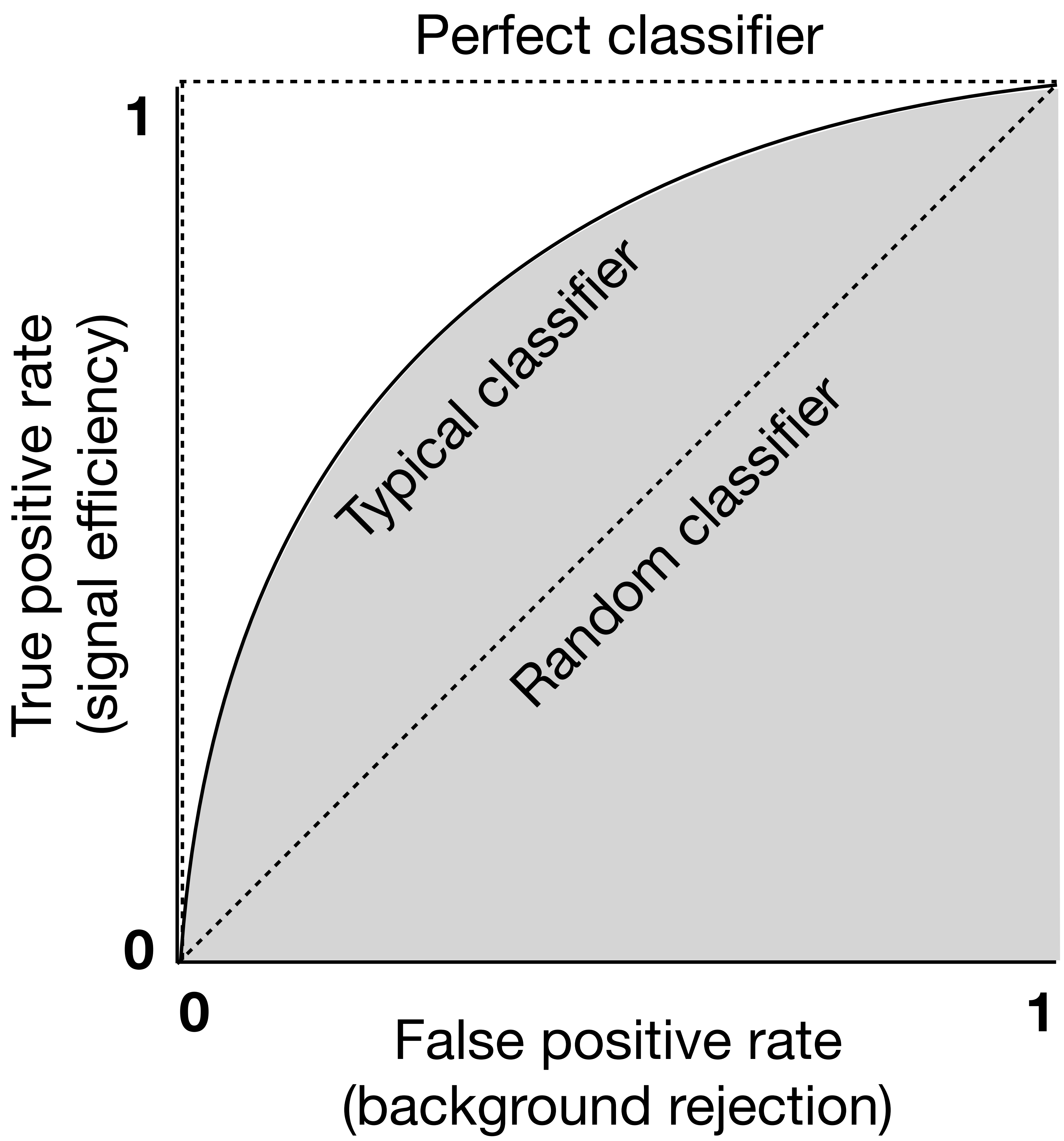
ROC curve

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

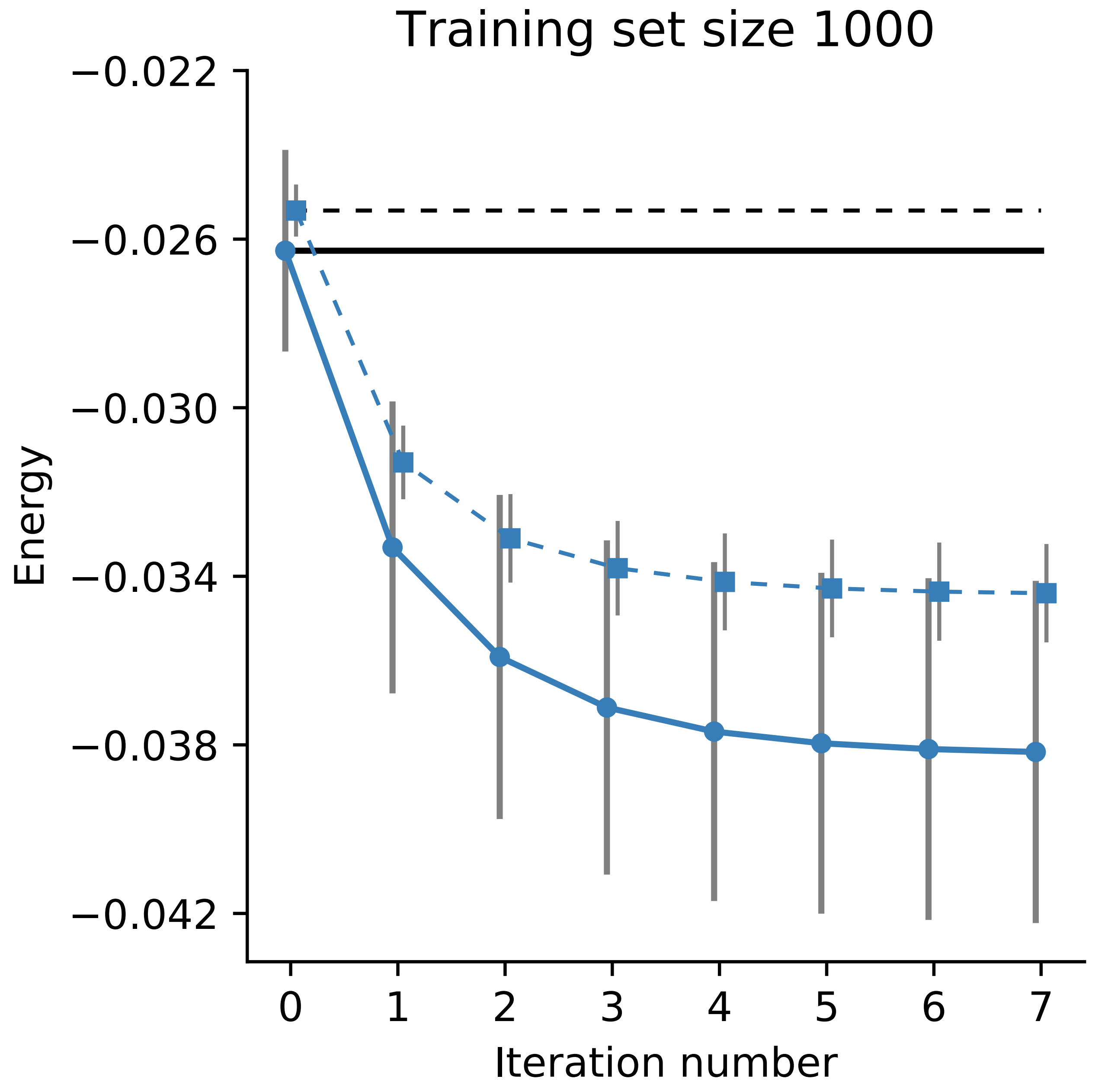


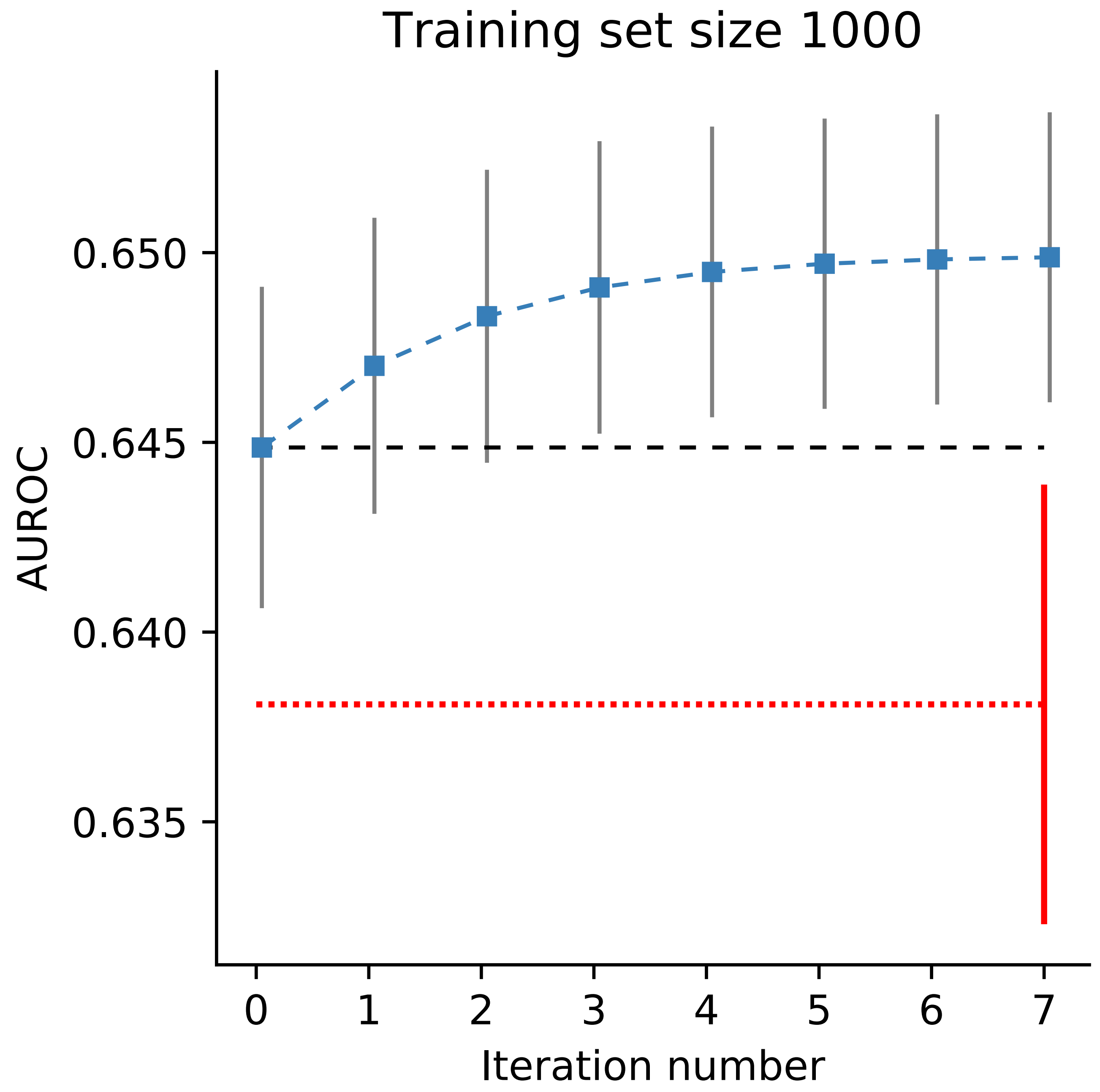
ROC curve

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



Higgs classification results





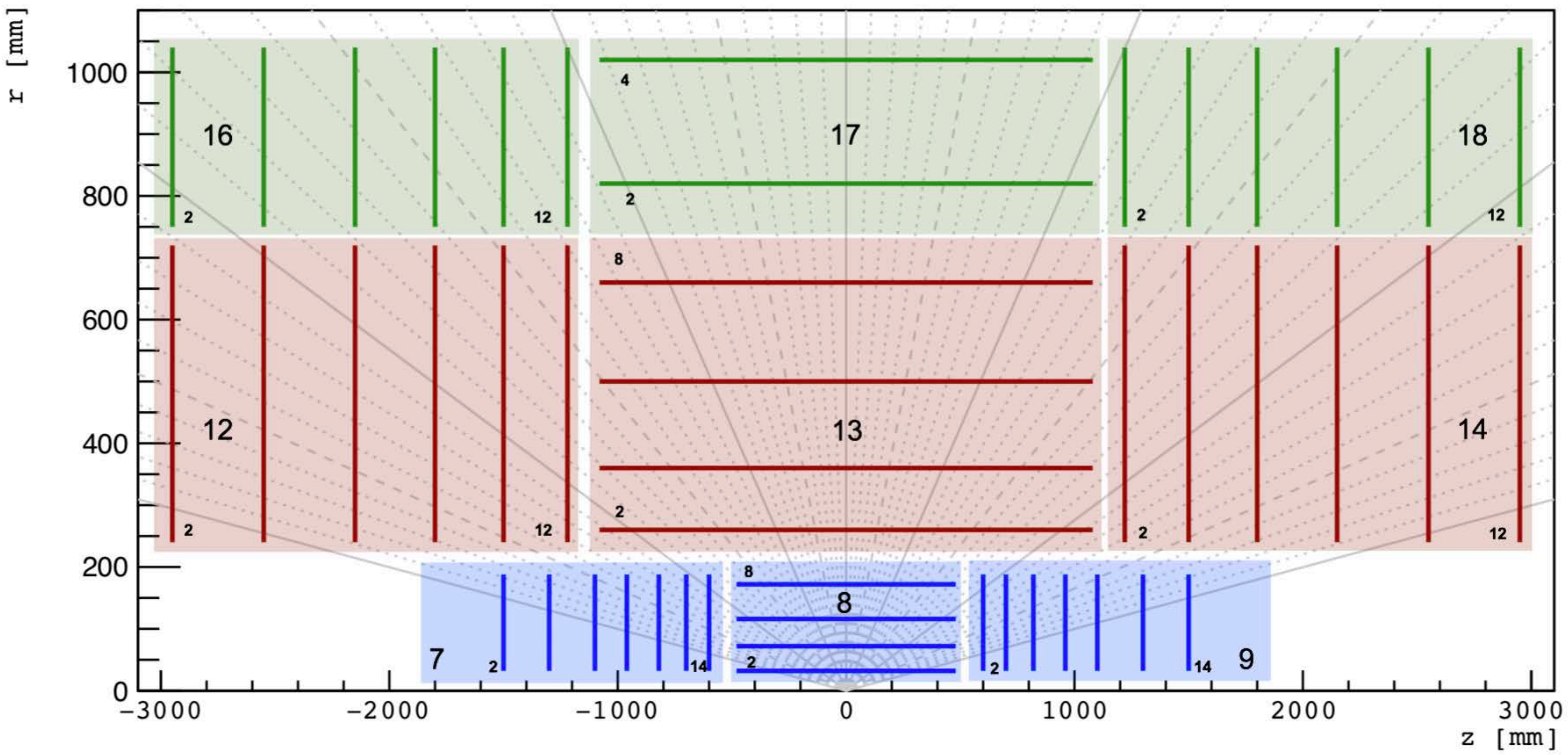
Augmented classifiers, no zoom
QAML
QAML-Z

Dashes indicate test set, solid line indicates training set

Caltech

Dataset

TrackML Challenge: top quark events with 15% noise

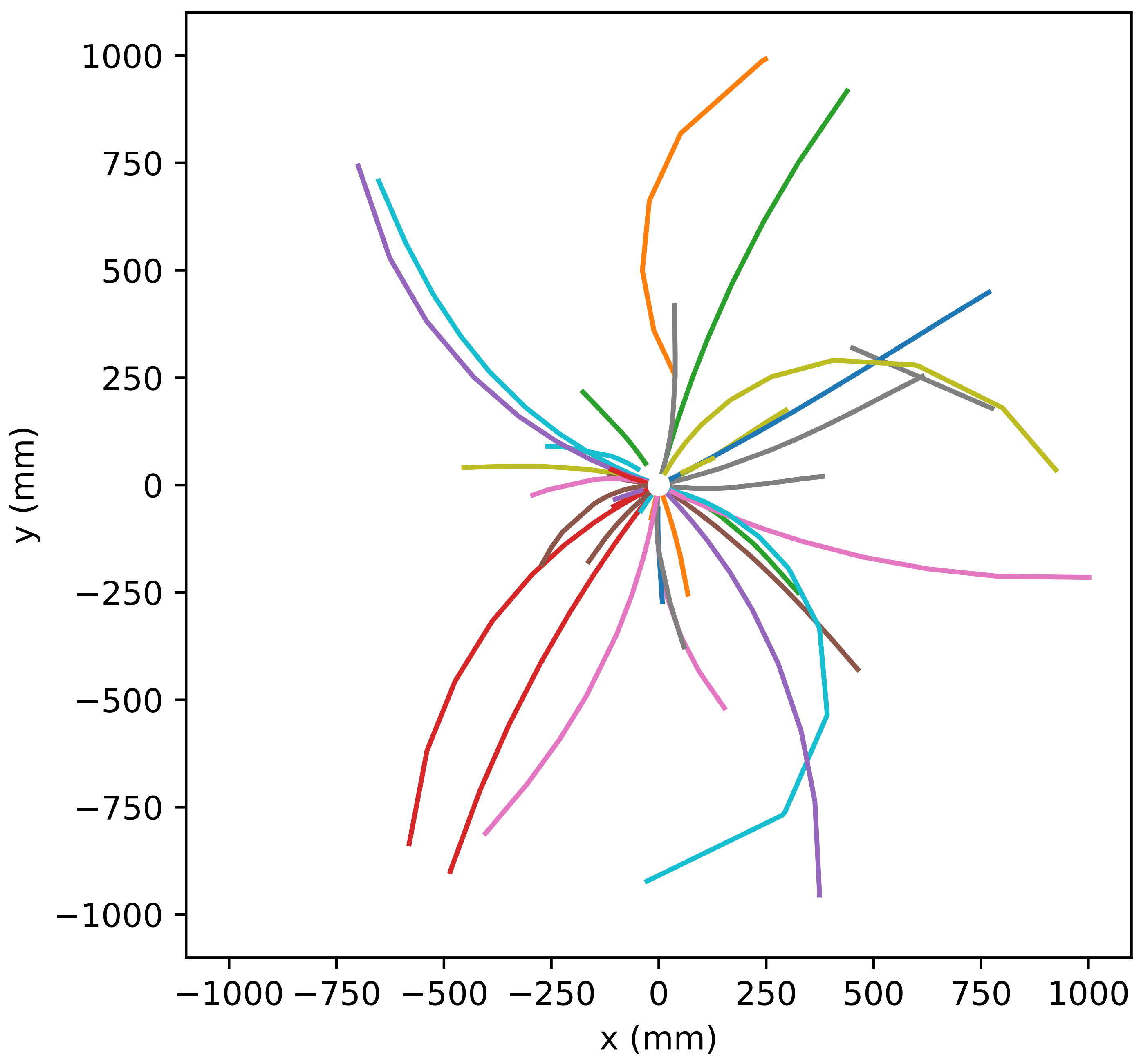


Amrouche, Sabrina, et al. "The Tracking Machine Learning challenge: Accuracy phase." arXiv preprint arXiv:1904.06778 (2019).

Ising model formulation

Bias towards high momentum tracks that are more important

$$-\left(\frac{\cos^{\lambda}(\phi_{abc})}{r_{ab} + r_{bc}}\right) s_{ab} s_{bc}$$

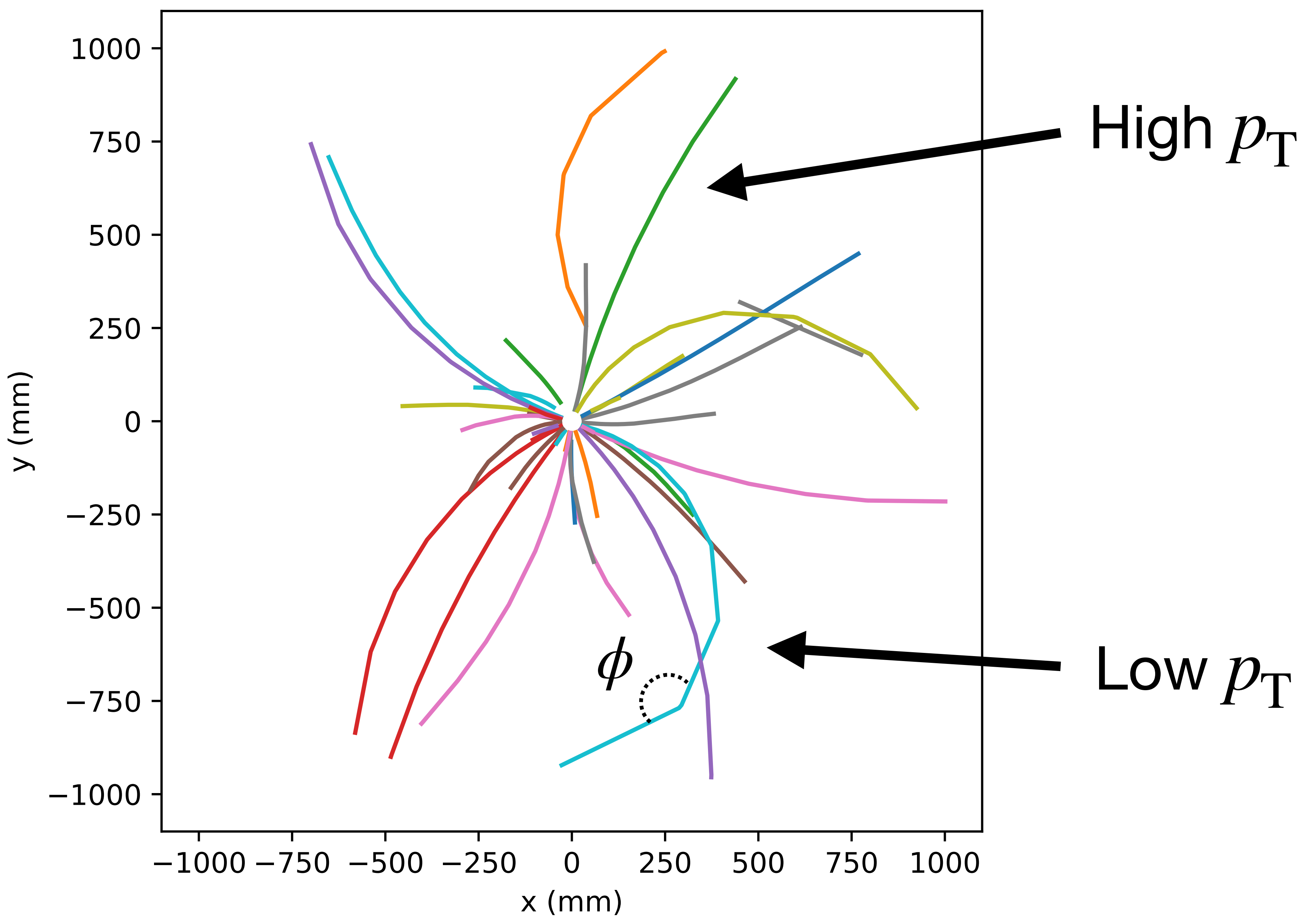


Caltech

Ising model formulation

Bias towards high momentum tracks that are more important

$$-\left(\frac{\cos^{\lambda}(\phi_{abc})}{r_{ab}+r_{bc}}\right)s_{ab}s_{bc}$$

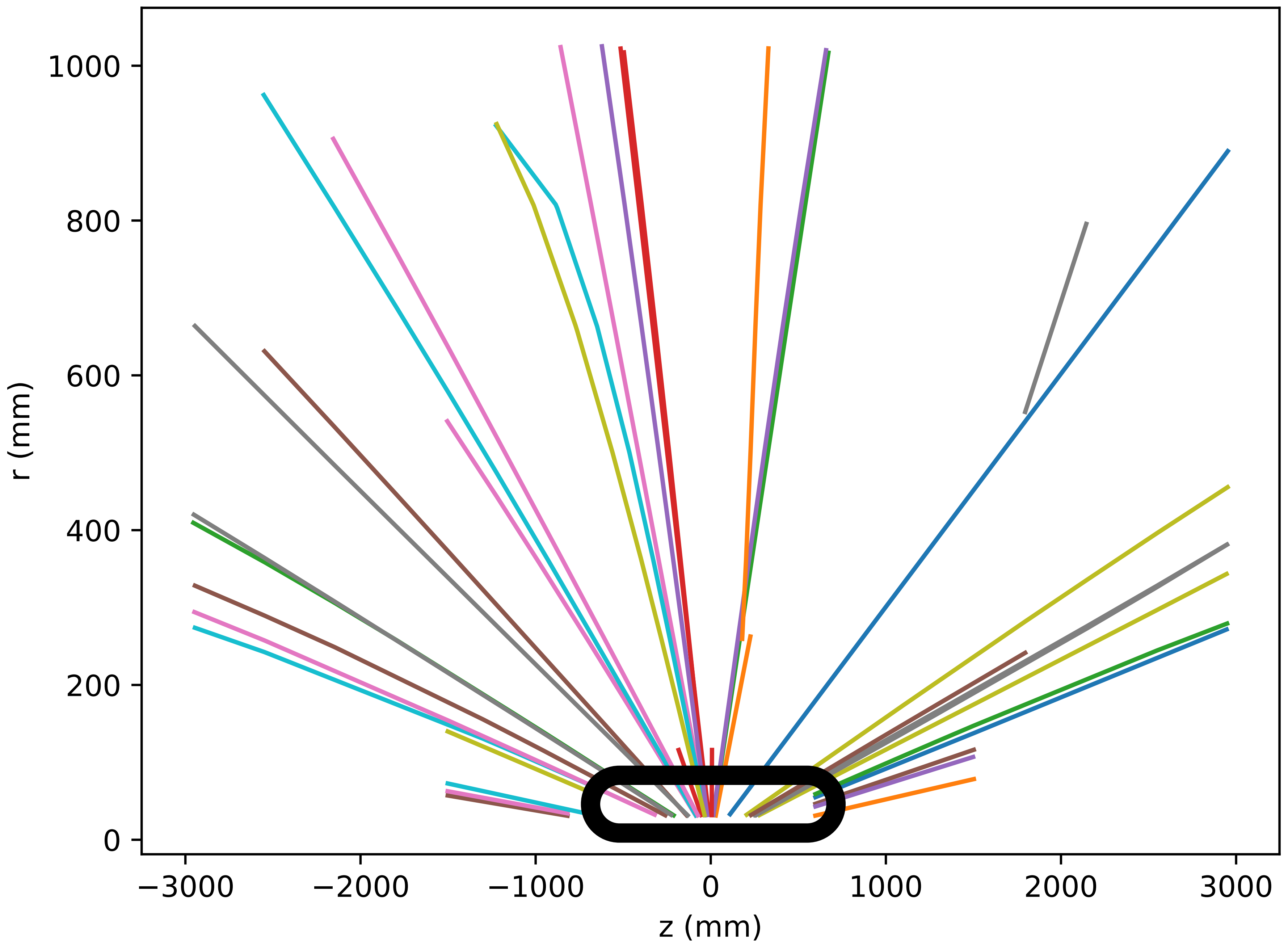


Caltech

Ising model formulation

Tracks should point towards the beam spot at the origin

$$\left(z_c - \frac{z_c - z_a}{r_c - r_a}\right) s_{ab} s_{bc}$$

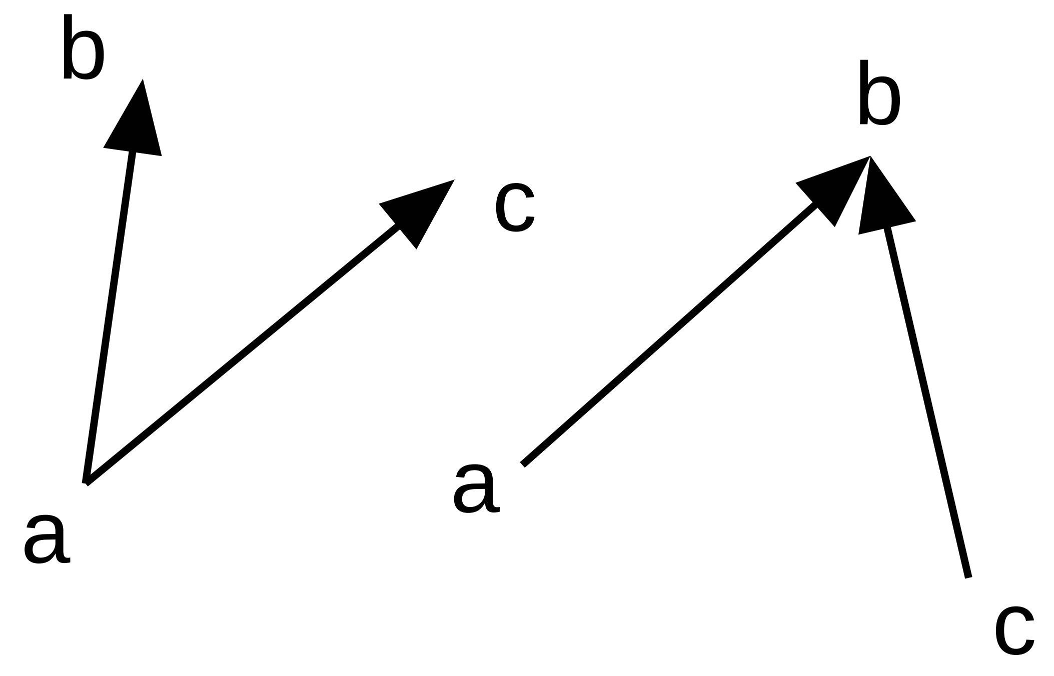


Caltech

Ising model formulation

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

$$s_{ab}s_{ac} + s_{ab}s_{cb}$$

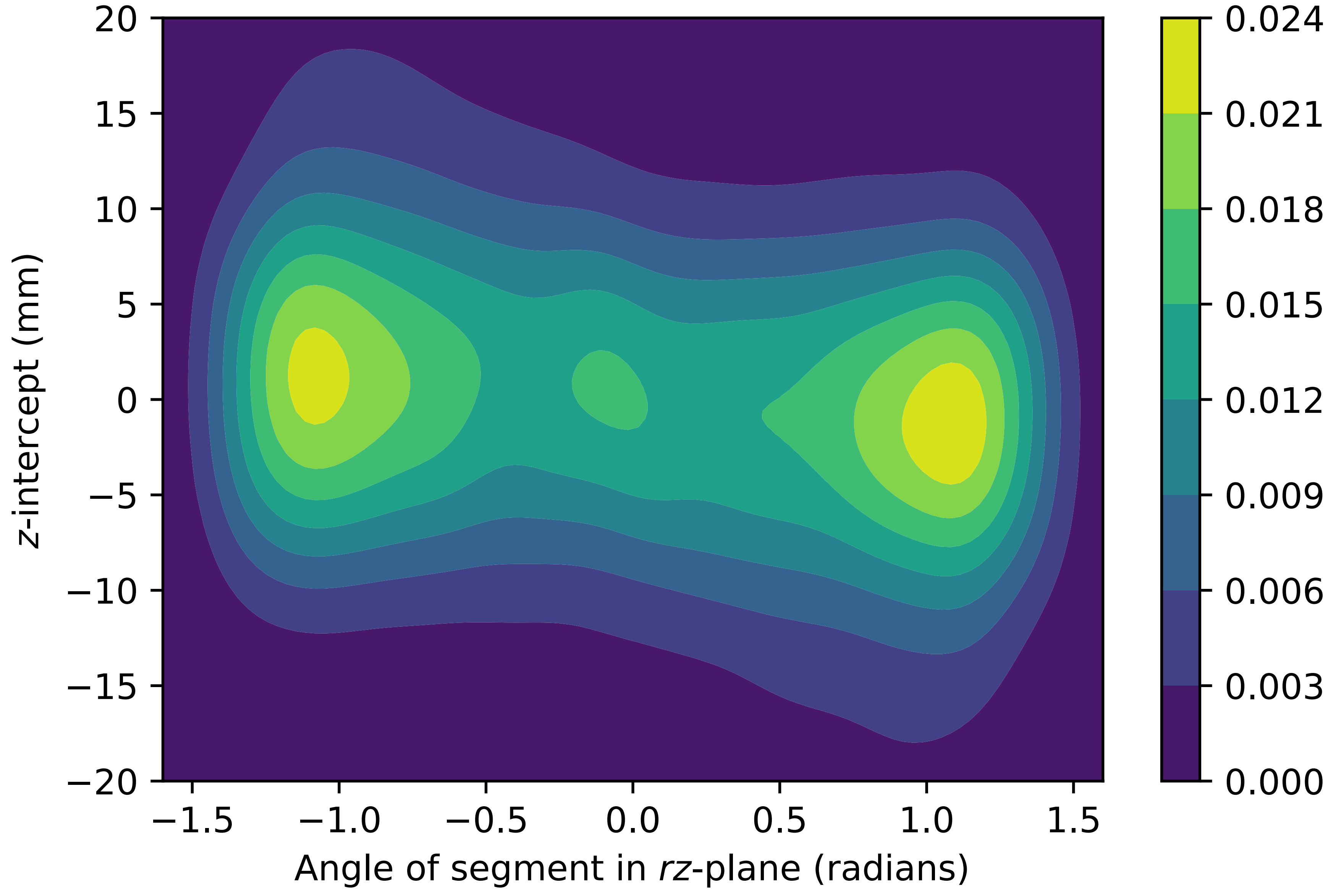


Caltech

Dimension challenge

Use Gaussian kernel density estimation to provide a prior on an edge being "on" or "off" based on orientation and position

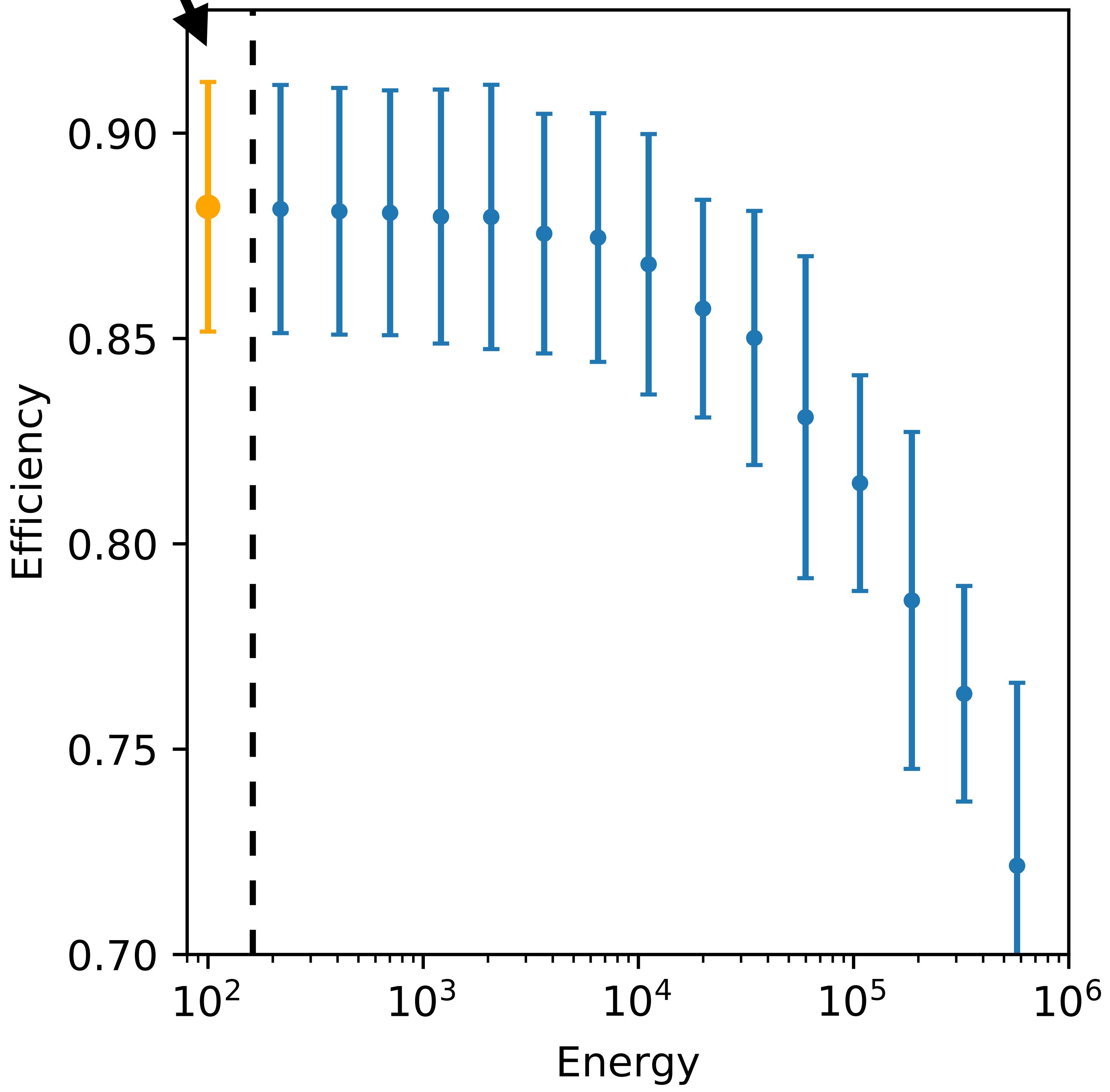
$$(\gamma - P(s_{ab}))s_{ab}$$



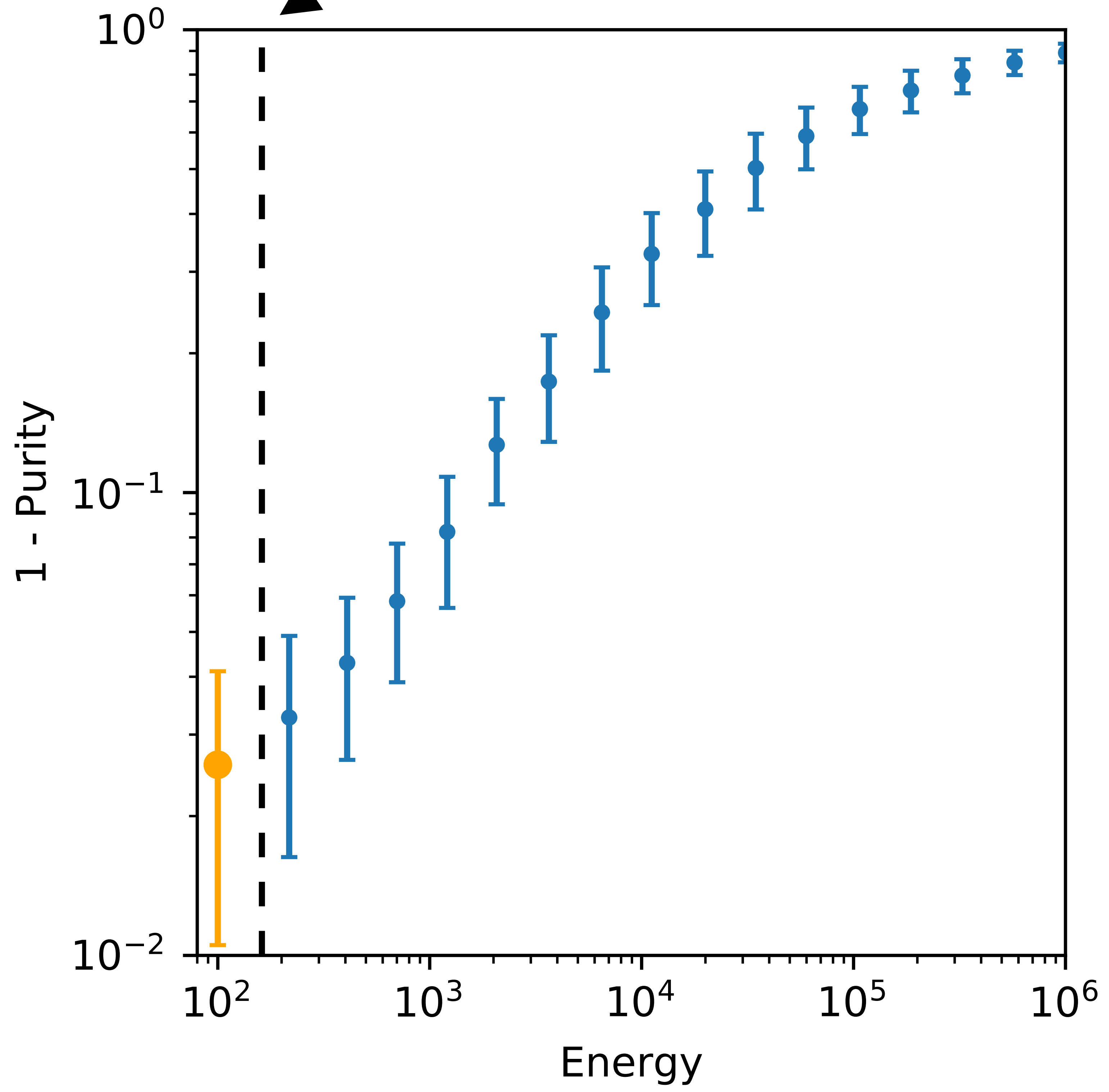
Caltech

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

Caltech

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

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

Caltech

Quantum speedup?

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

Singlet selection
with Gaussian KDE
$O(h^2)$

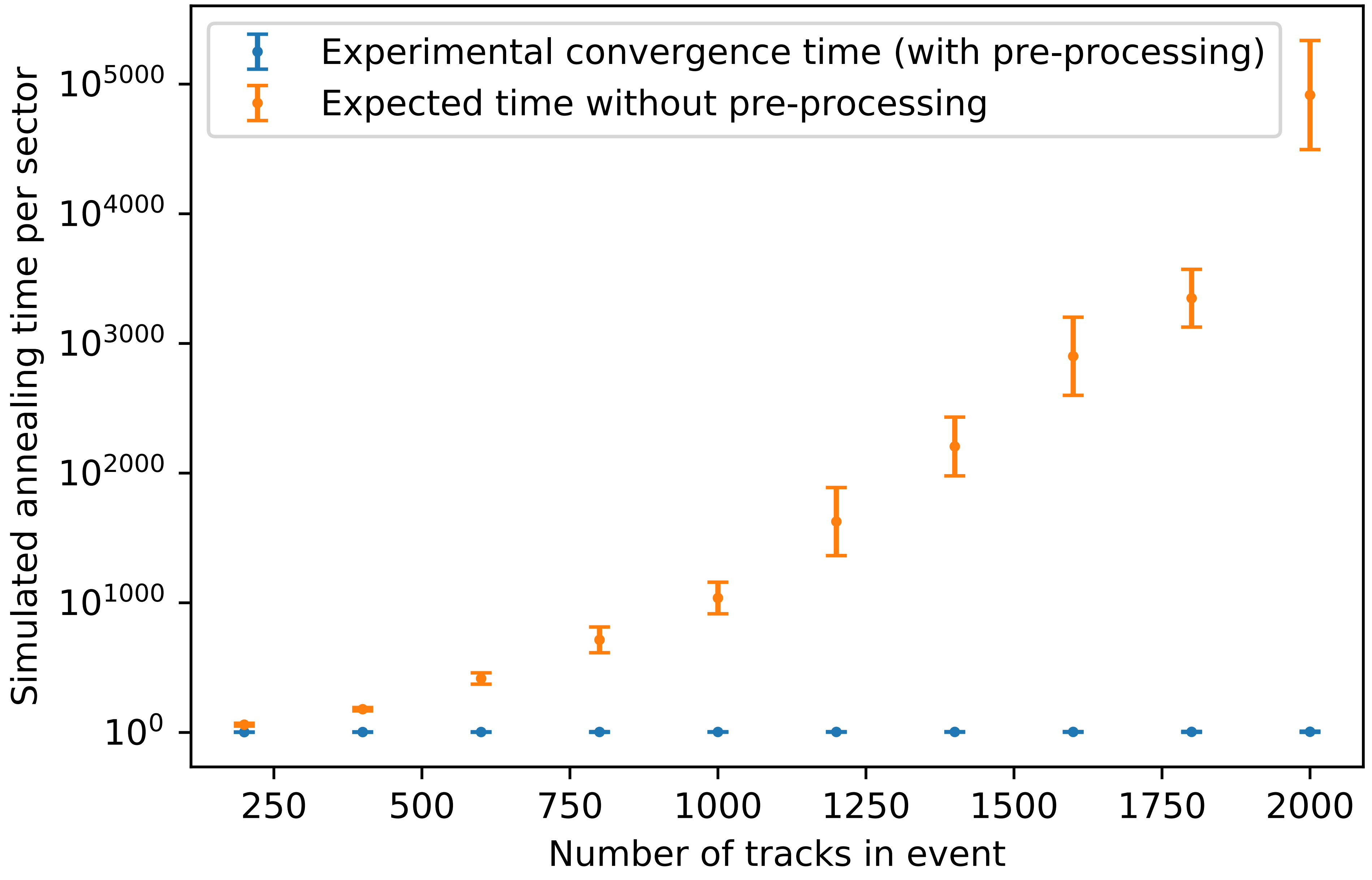
Ising model construction
$O(h^2)$

Disjoint sub-graph
flood-fill search
$O(h^2)$

Annealing
Classical: $O(\exp(ch^2))$
Quantum: $O(?)$

Quantum speedup?

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



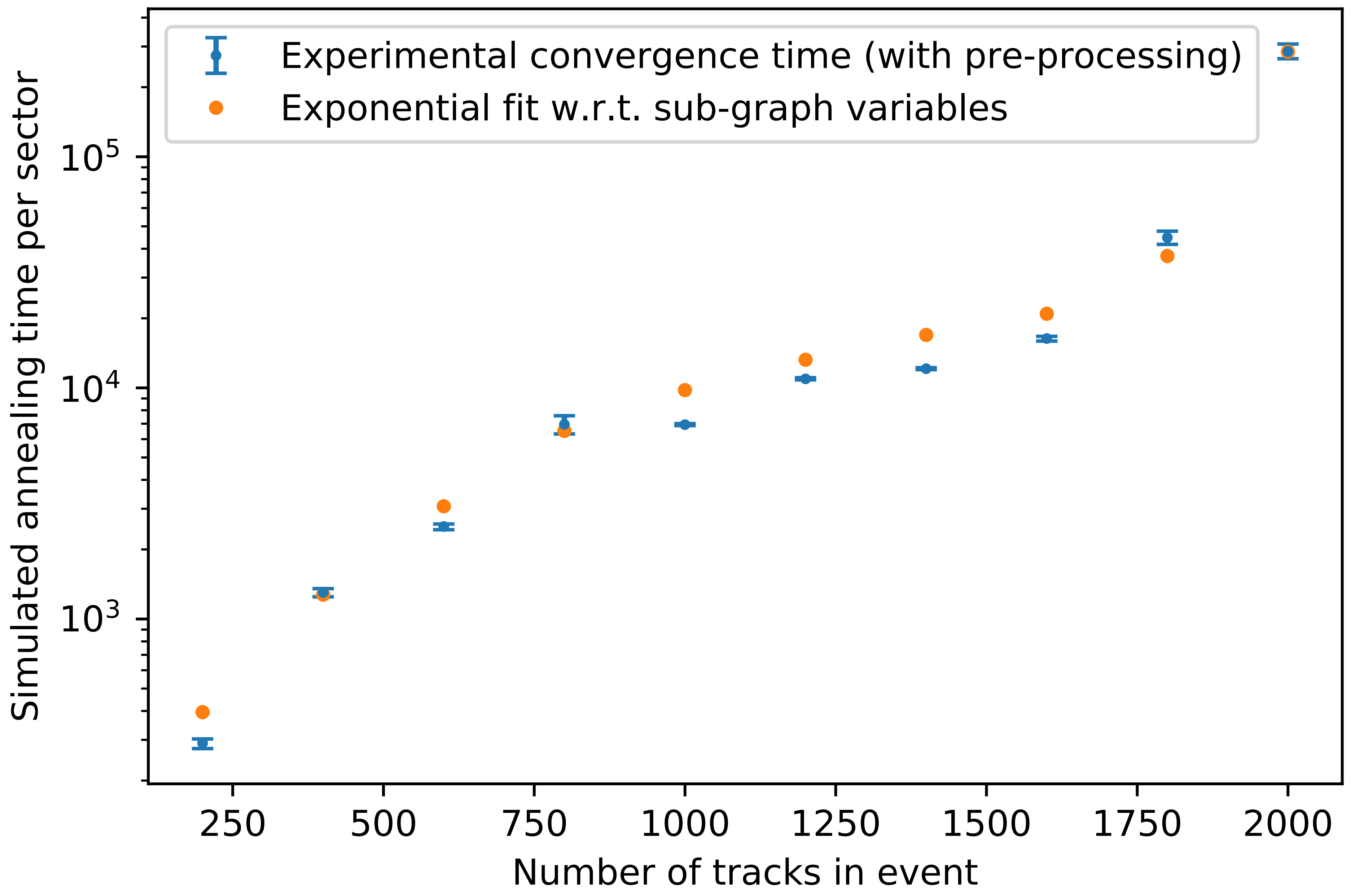
Caltech

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

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



Caltech

Quantum speedup?

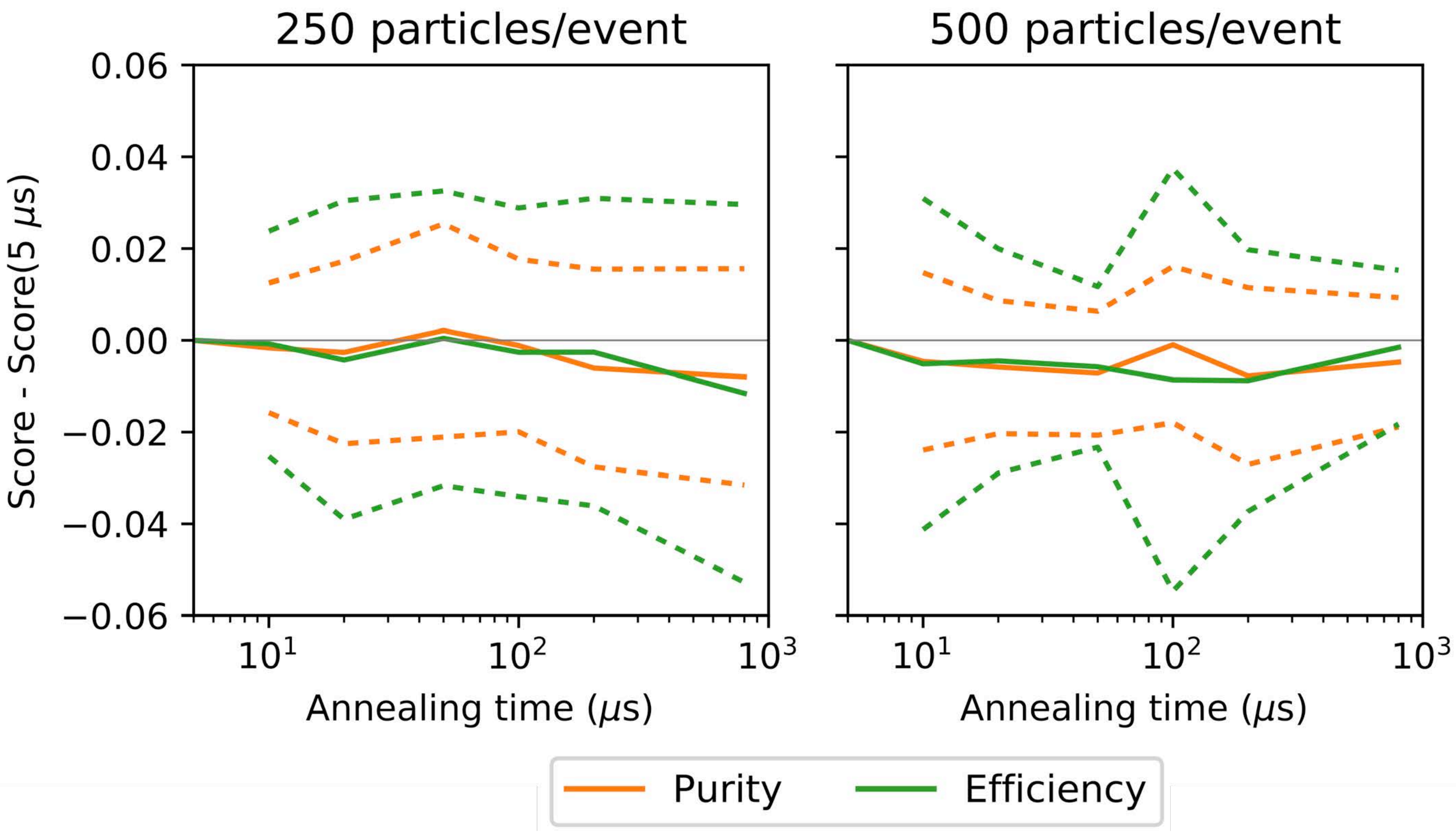
Quantum annealing is expected to reduce the size of c but leave the problem exponential

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

Can't measure a single time to solution, but can change annealing time and measure change in performance

Caltech

Quantum speedup?



Caltech