D:Wave
The Quantum Computing Company™

Classification with Quantum Annealing on the D-Wave System

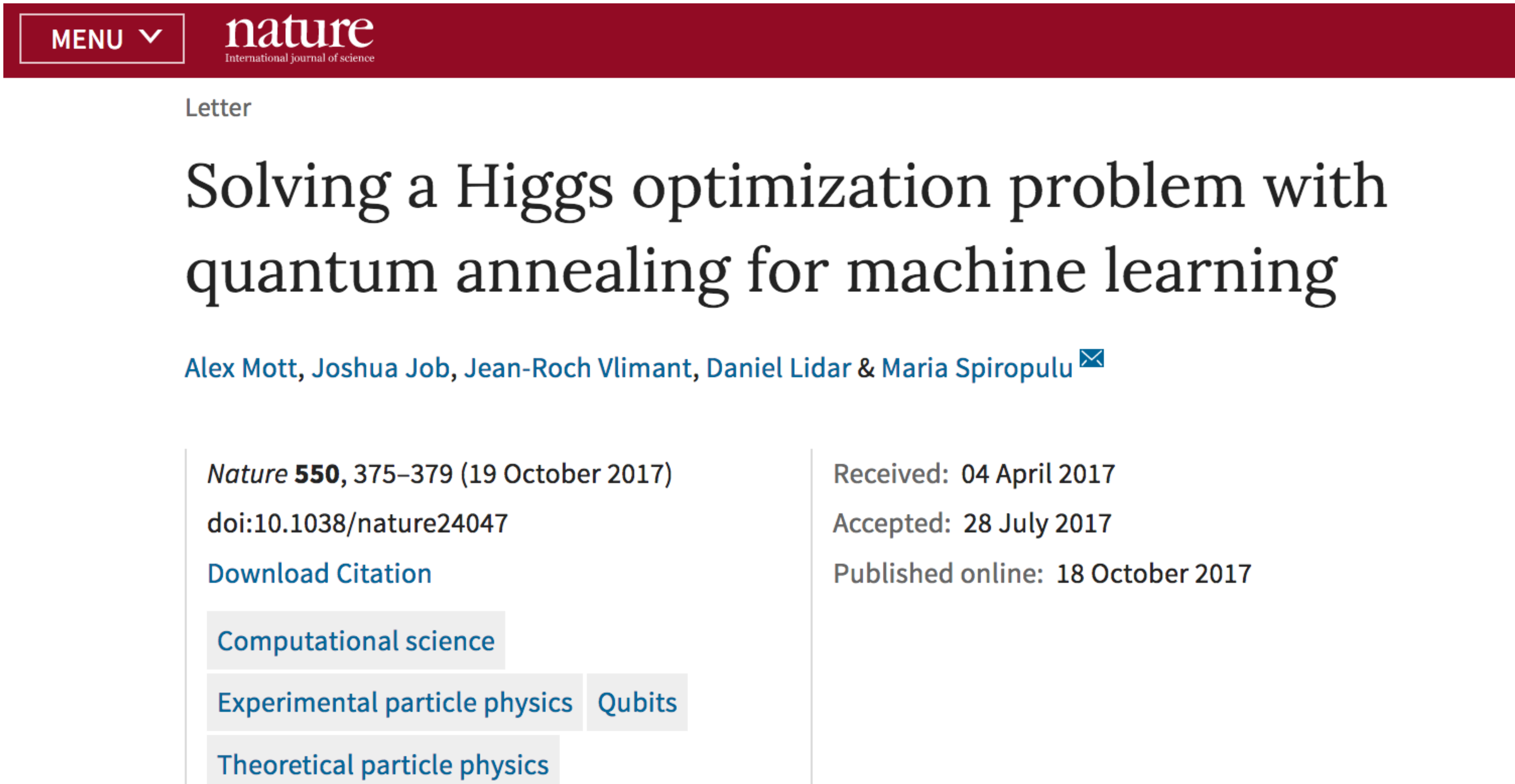
Seminar at the Laboratoire de l'Accelerateur Lineaire - LAL
Jan 26, 2018

Jean-Roch Vlimant
jvlimant@caltech.edu
California Institute of Technology

Outline

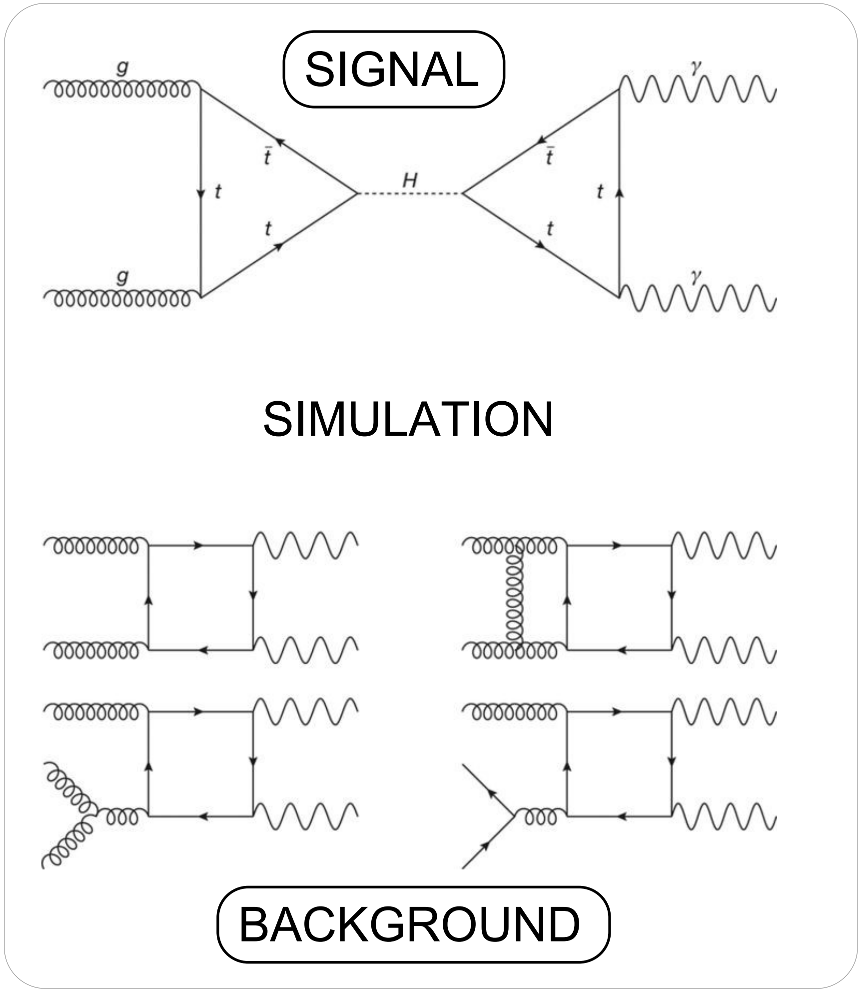
- Overview
  - Quantum Annealing
  - QA Machine Learning
  - A Higgs dataset
- Experiments
- Outlooks

Overview

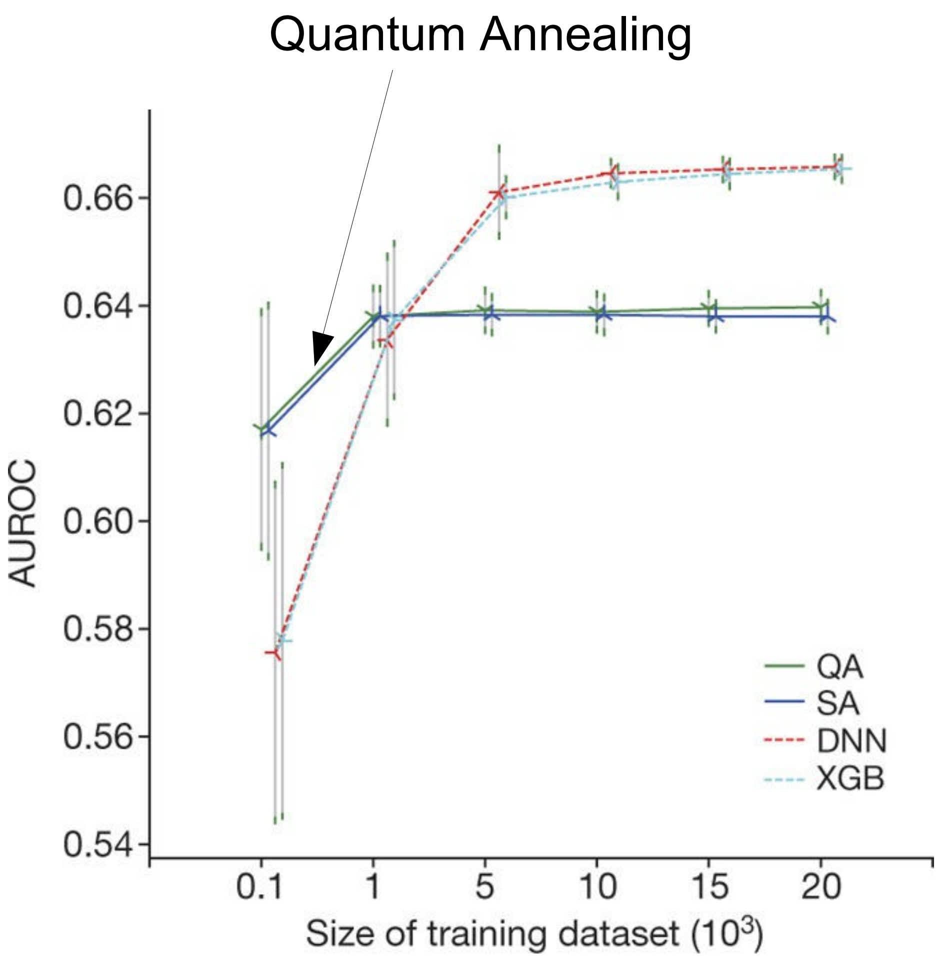


https://www.nature.com/articles/nature24047

Experiment

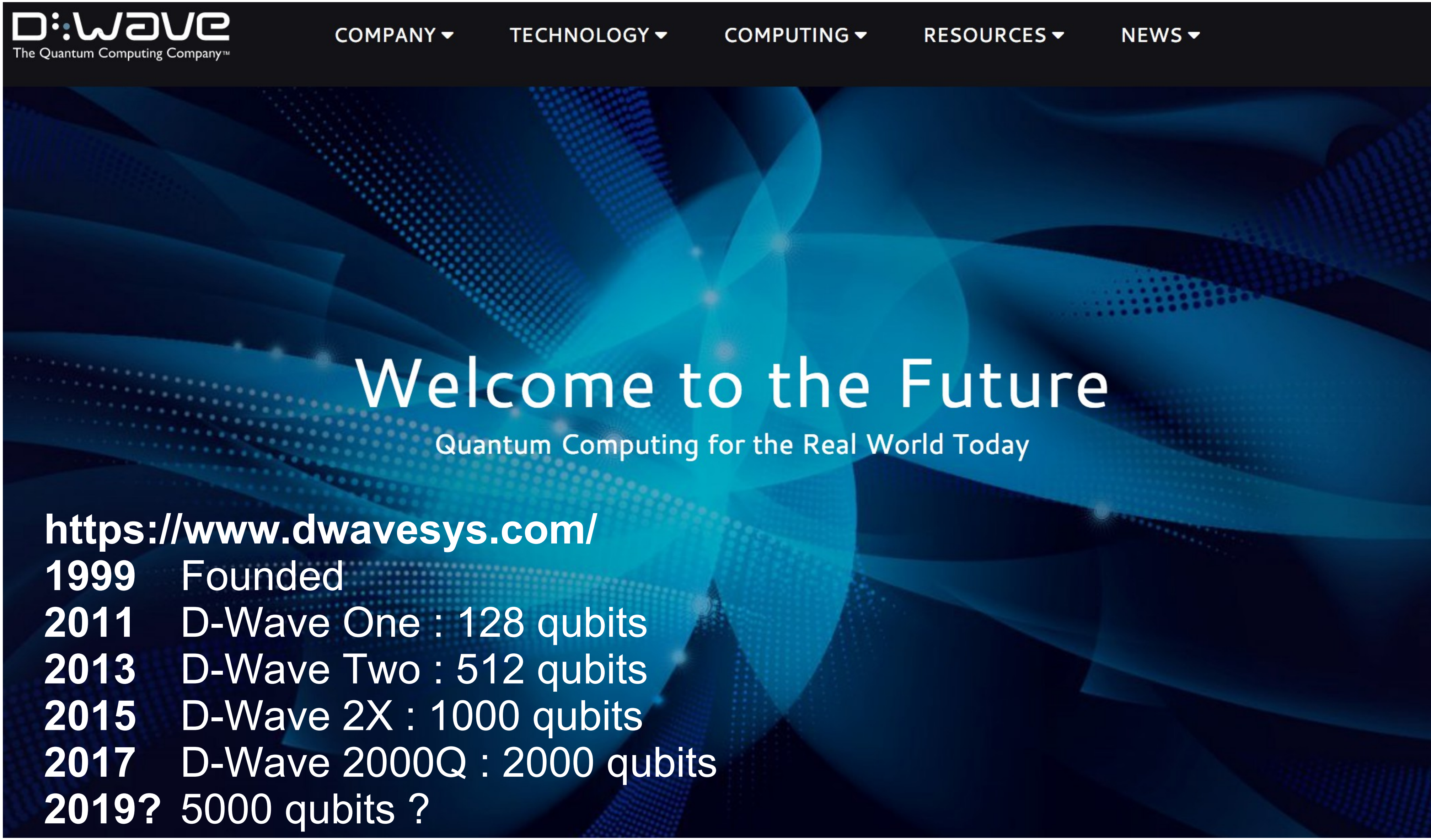


Classification



The D-Wave Computing System

The D-Wave Company

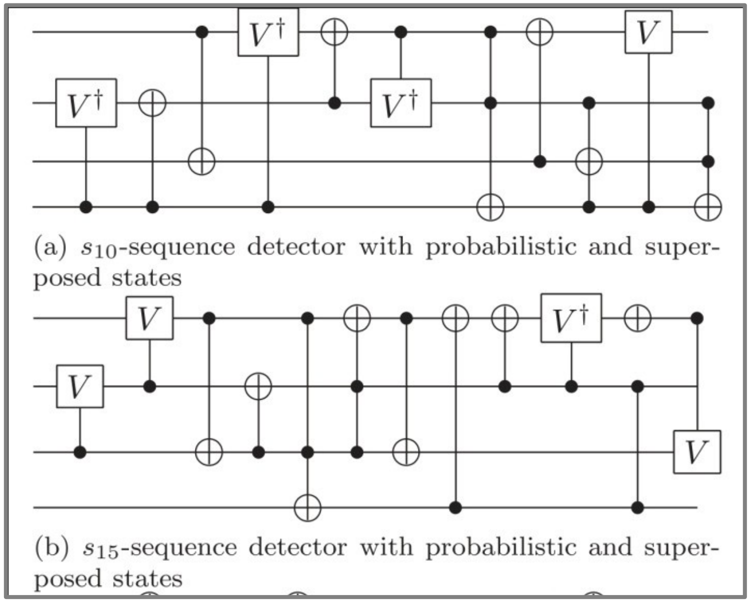


https://www.dwavesys.com/

1999 Founded
2011 D-Wave One : 128 qubits
2013 D-Wave Two : 512 qubits
2015 D-Wave 2X : 1000 qubits
2017 D-Wave 2000Q : 2000 qubits
2019? 5000 qubits ?

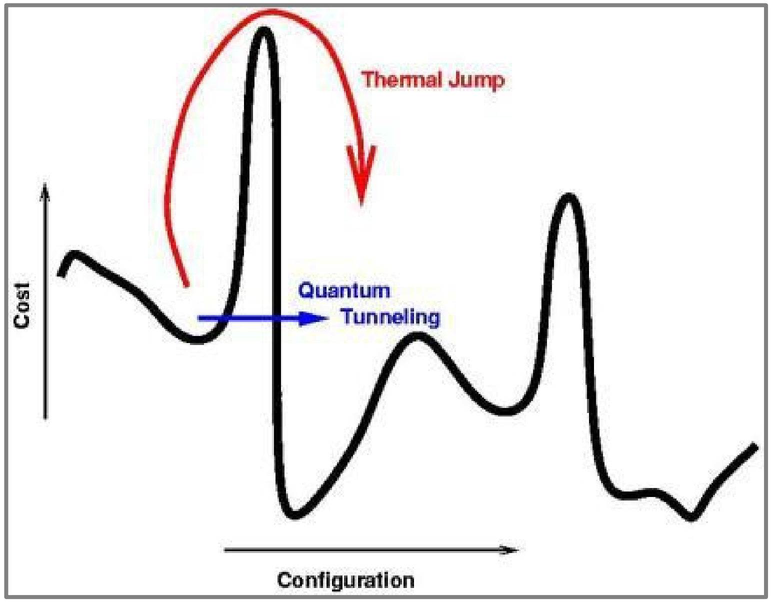
D-Wave 2X™

qubit and qubit



<u>Quantum Circuits</u>

Series of quantum gates operating on a set of quantum states.



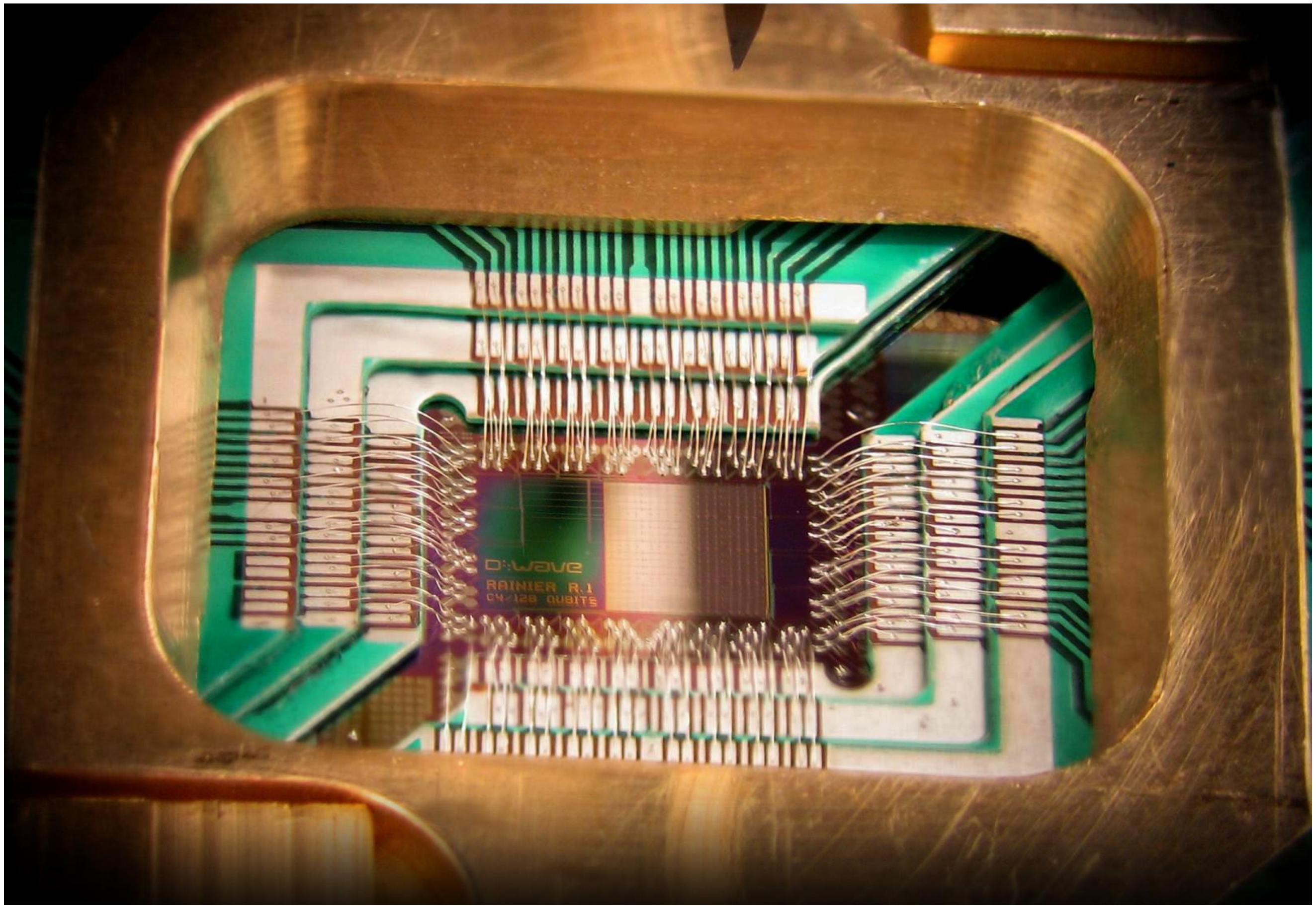
<u>Quantum Annealing</u>

Evolution of a quantum system to a low T Gibbs state

That's D-Wave !

D-Wave Chip

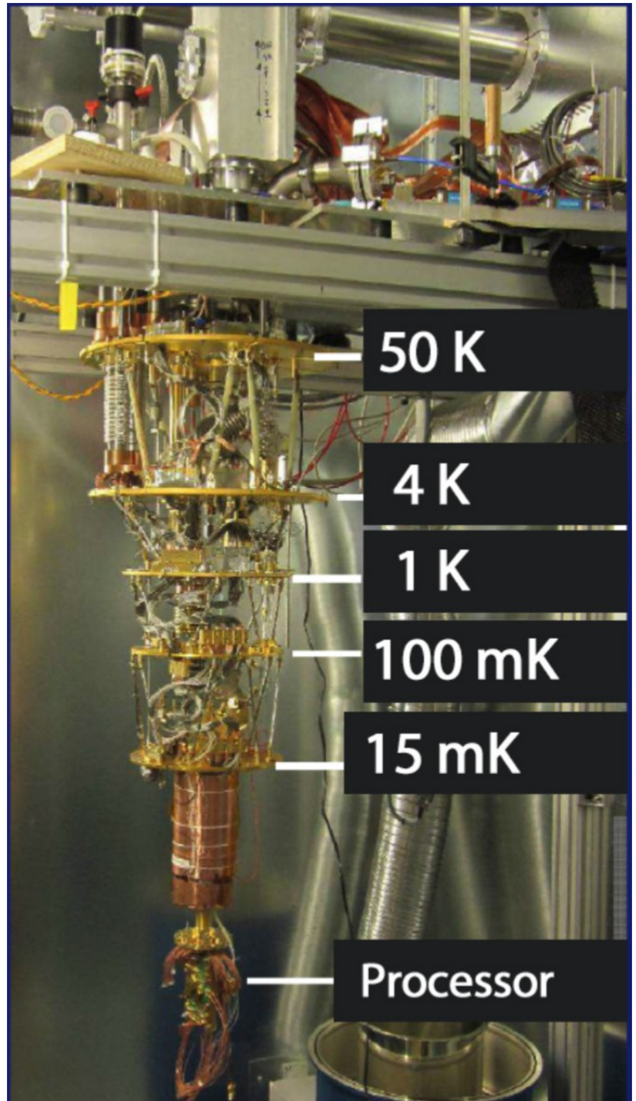
D-Wave's quBit



- Each qubit is a pair of Josephson junction (JJ)
- Able to apply local magnetic fields with programable digital-to-analog flux converters (DAC)
- Operates at 15 mK to remove noise

https://doi.org/10.1109/TASC.2014.2318294

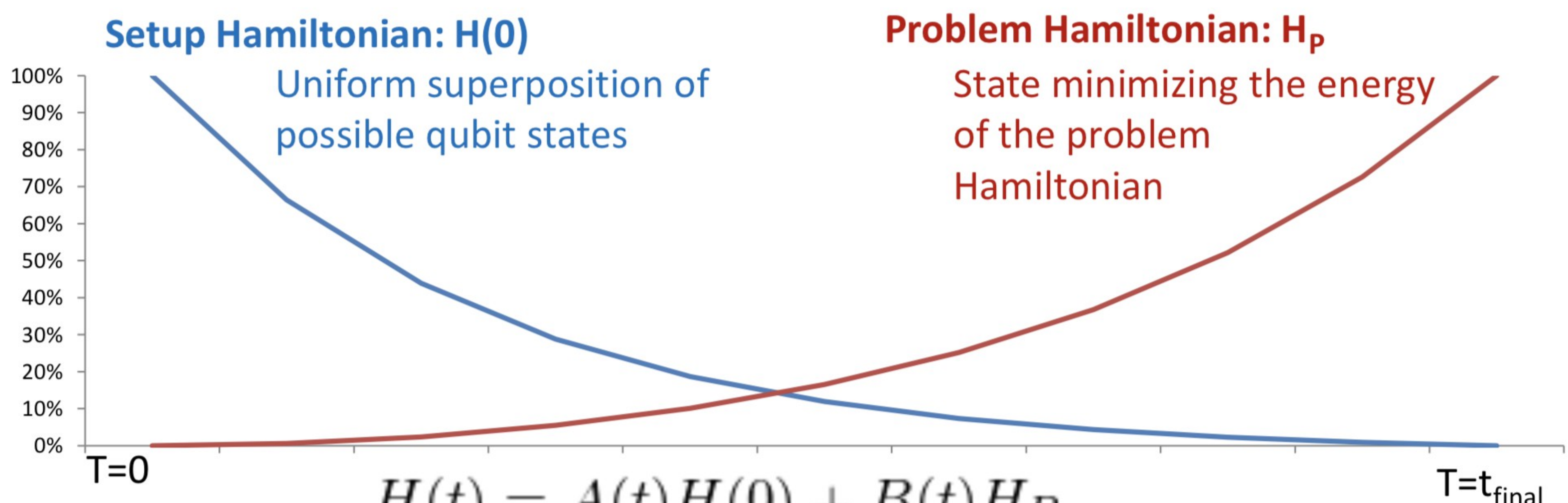
Thermal Noise Isolation



Quantum Annealing

Adiabatic Quantum Annealing

- System setup with trivial hamiltonian H(0) and ground state
- Evolve adiabatically the hamiltonian towards the desired Hamiltonian H_p
- **Adiabatic theorem** : with a slow evolution of the system, the state stays in the ground state.



$$H(t) = A(t)H(0) + B(t)H_P$$

https://arxiv.org/abs/quant-ph/0001106
https://arxiv.org/abs/quant-ph/0104129

D-Wave Hamiltonian
And
Chimera Graph

D-Wave Hamiltonian

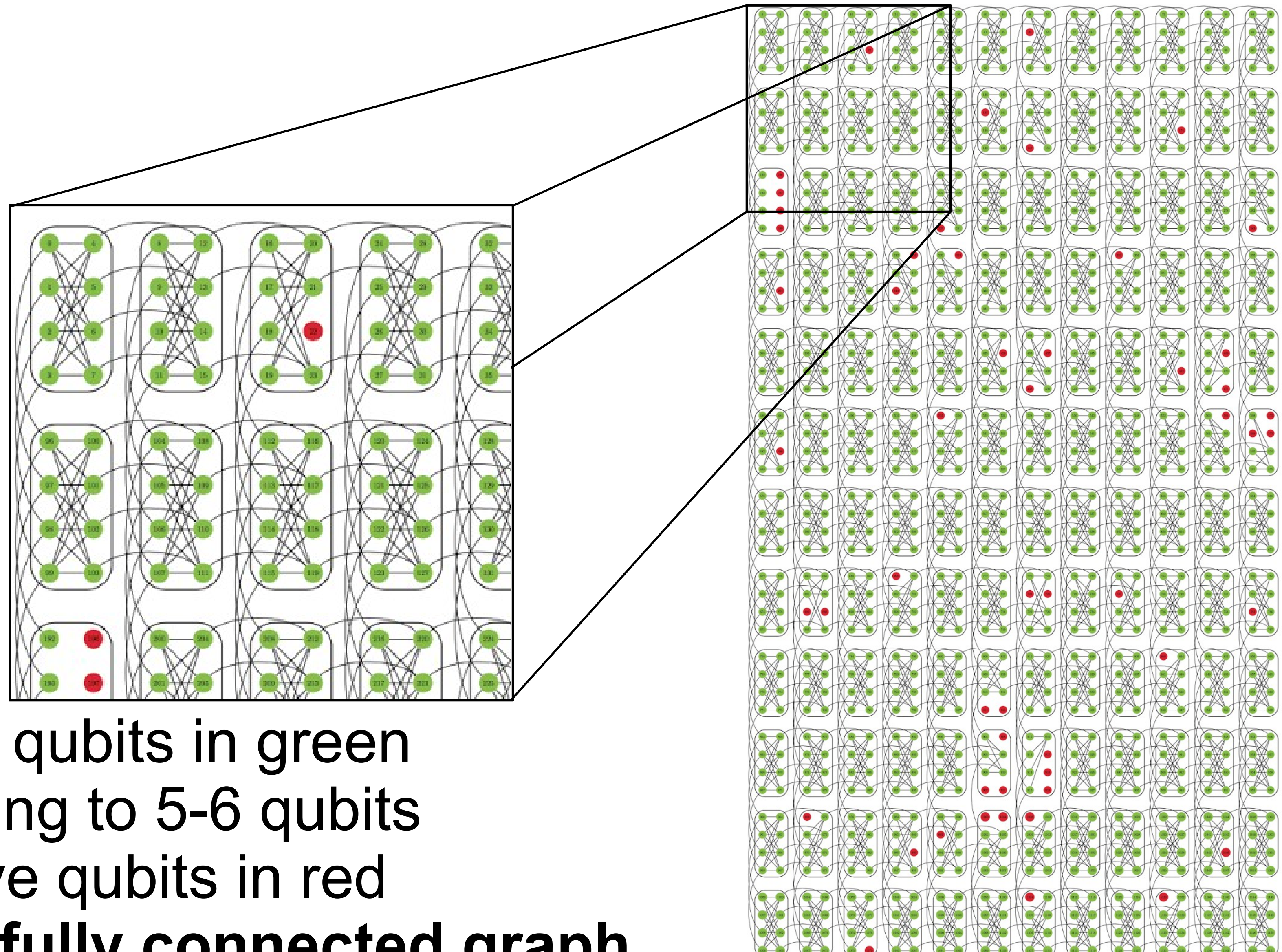
Runs over
adjacent quBits

$$H_{\text{Ising}} = \sum_i h_i \sigma_i^z + \sum_{ij} J_{ij} \sigma_i^z \sigma_j^z$$

External
magnetic field

Interactions

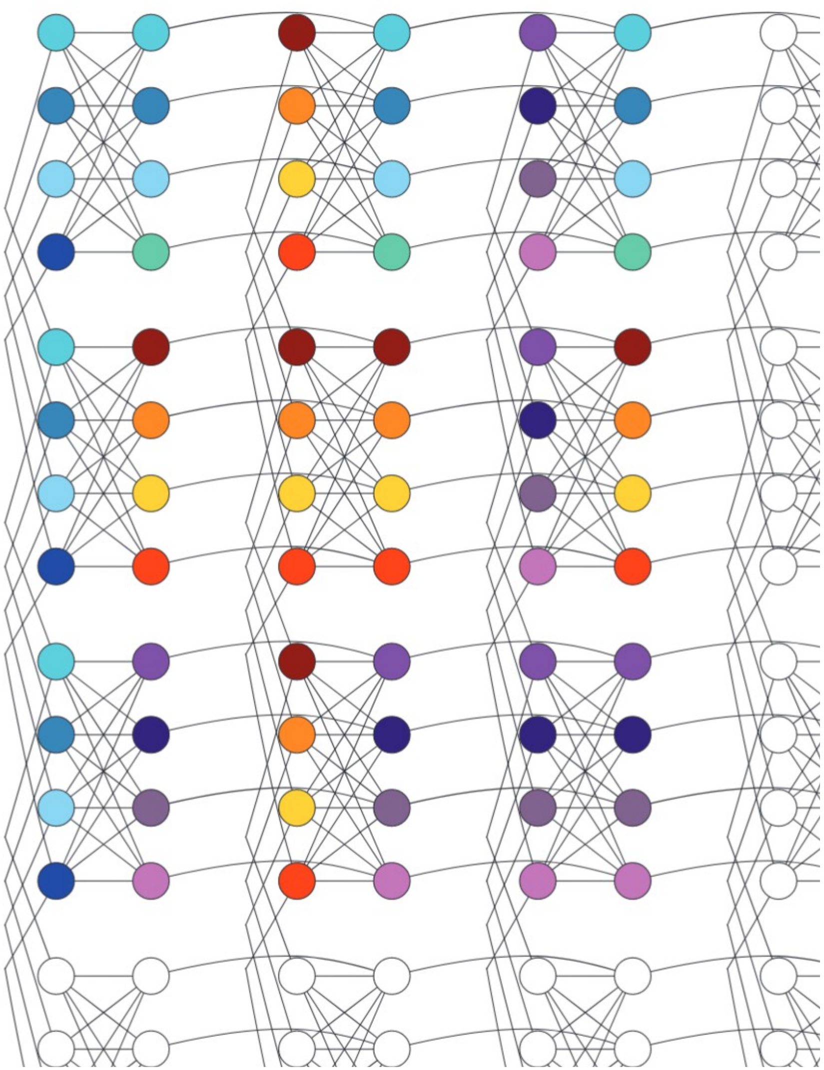
D-Wave qubit Adjacency



Active qubits in green
Coupling to 5-6 qubits
Inactive qubits in red
Not a fully connected graph

Model Embedding

Full Ising Model



- Create chains of spins through the chimera graph
- Split local fields across all qubits in the chain
- Tightly couple (J_F=6)

- Non-unique embedding. Heuristic approach.

- Suppressing spin flip within chain as error correction.
- Use majority vote

→ Approximately full Ising Model with ~<40 spins

https://arxiv.org/abs/1210.8395

Ising Hamiltonian

Runs over **all** quBit pairs

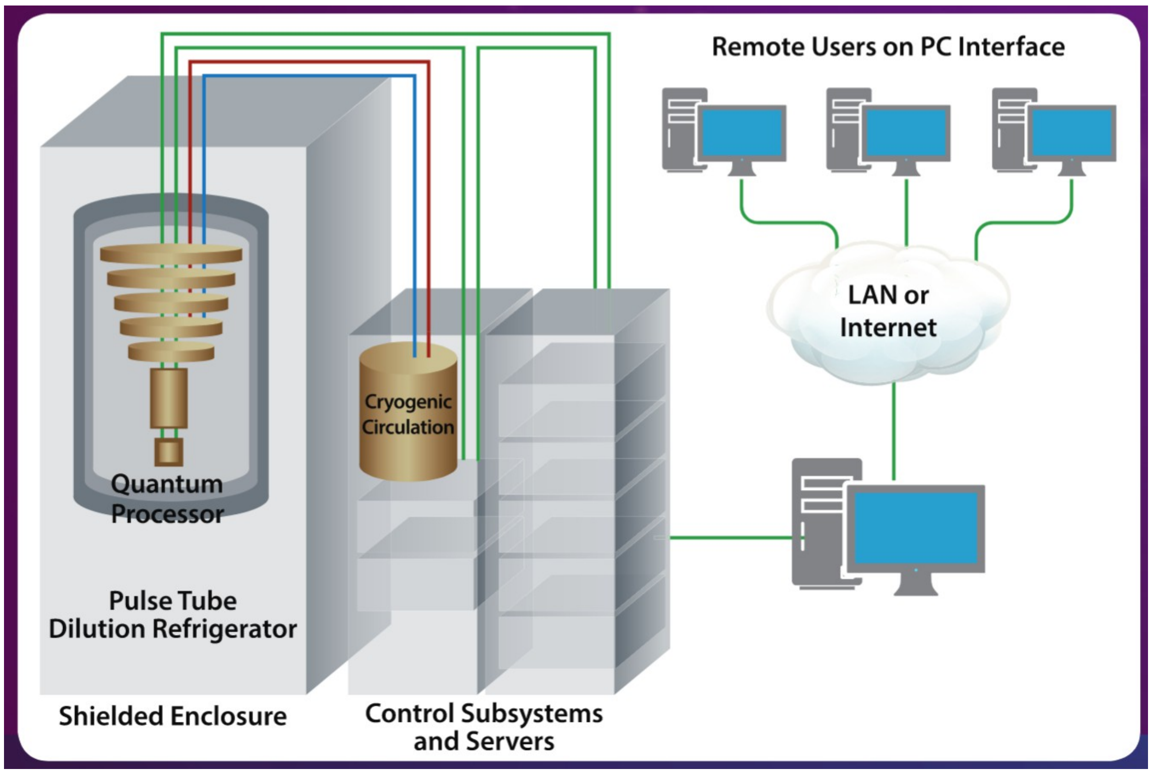
$$H_{\text{Ising}} = \sum_i h_i \sigma_i^z + \sum_{ij} J_{ij} \sigma_i^z \sigma_j^z$$

External magnetic field

Interactions

D-Wave
Computing Interface

Working on a D-Wave



- Web Interface to post the problem settings (Hp).
- Asynchronous processing.
- Solution is made available for download.

- Distributed library for performing embedding
  - Retain full intellectual property.

- Equivalent restapi to submit and retrieve solutions
  - D-Wave processor as a service

D-Wave Sampling Solutions

User provides

$(\mathbf{h}_i , \mathbf{J}_{ij})$

Multiple 5 μs
annealing cycles

D-Wave provides sampling
from lowest energy levels
(appox. Gibbs)

$(\{\sigma_i\}_k, N(\{\sigma_i\}_k))$

Classification with Quantum Annealing

QA Machine Learning

Formalism to transform a binary classification into an Ising model hamiltonian optimization

K.L. Pudenz, D.A. Lidar
https://arxiv.org/abs/1109.0325

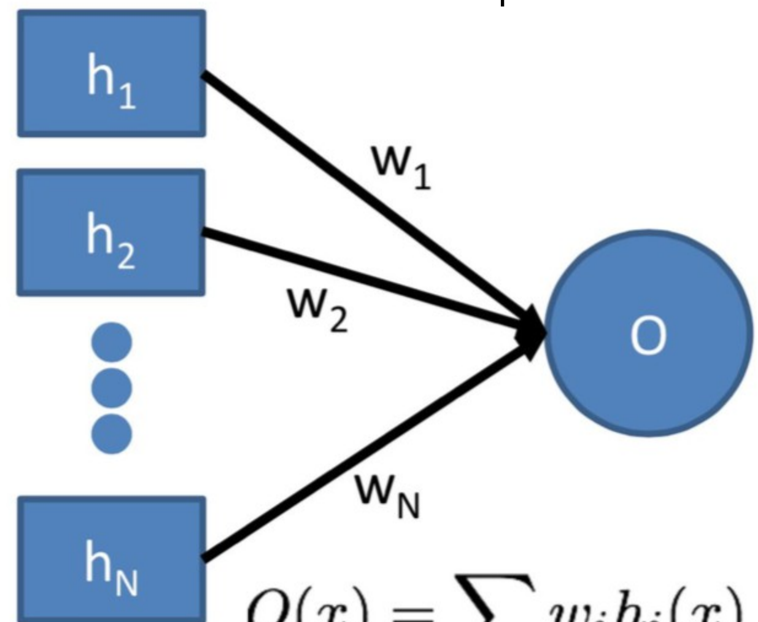
QAML Weak/Strong Classifier

Define functions $\mathbf{h_i}$ of the input variables into [-1,1] such that

- P(signal|h>0) > P(bkg|h>0)
- P(bkg|h<0) > P(signal|h<0)

i.e. Most signal on h>0, most bkg on h<0

Define $\mathbf{w_i}$ as binary linear combination of h_i



$$O(x) = \sum_i w_i h_i(x)$$

https://arxiv.org/abs/1109.0325

QAML Target/Objective

Define as a "target" function

$$y(x) = \begin{cases} +1, & \text{if } \in S, \\ -1, & \text{if } \in B \end{cases}$$

Per event error

$$E(x) = y(x) - \sum_{i=1}^{N} w_i h_i(x)$$

Full error

$$\delta(\vec{w}) \propto \sum_{i,j} C_{ij} w_i w_j + \sum_{i} (\lambda - 2C_{iy}) w_i$$

- C_{ij} and C_{iy} are summations over the values of h_i over the training set
- λ is a parameter penalizing the number of non-zero w_i

https://arxiv.org/abs/1109.0325

QUBO

Quadratic Unconstrained Binary Optimization

$$\delta(\vec{w}) \propto \sum_{i,j} C_{ij} w_i w_j + \sum_i (\lambda - 2C_{iy}) w_i$$

Simple conversion of binary weights to ±1

$$H_{\text{Ising}} = \sum_i h_i \sigma_i^z + \sum_{ij} J_{ij} \sigma_i^z \sigma_j^z$$

QAML End-to-End

QAML Discriminant

Signal and Background samples
Less than 40 features

Run on D-Wave

Mask (w_i) of features contributing to

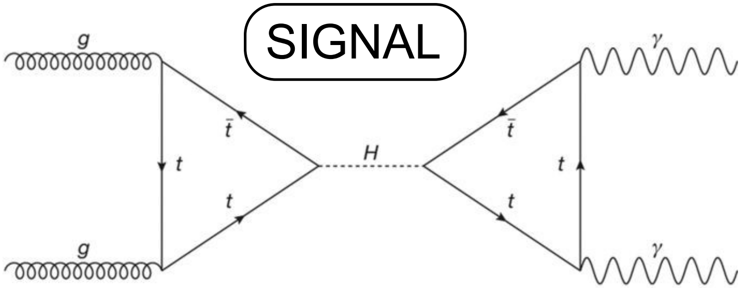
$$\sum_{i=1}^{N} w_i h_i(x)$$

continuous discriminant function.
One mask per energy level.

A Higgs- background dataset

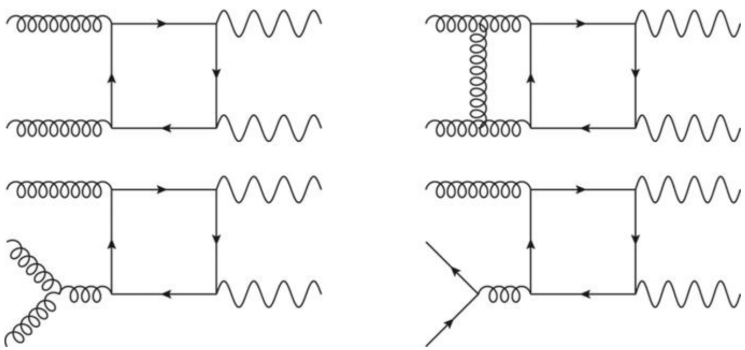
Generated Samples

SIGNAL



Generated with PYTHIA 6.4 at 8TeV proton-proton c.o.m energy

BACKGROUND



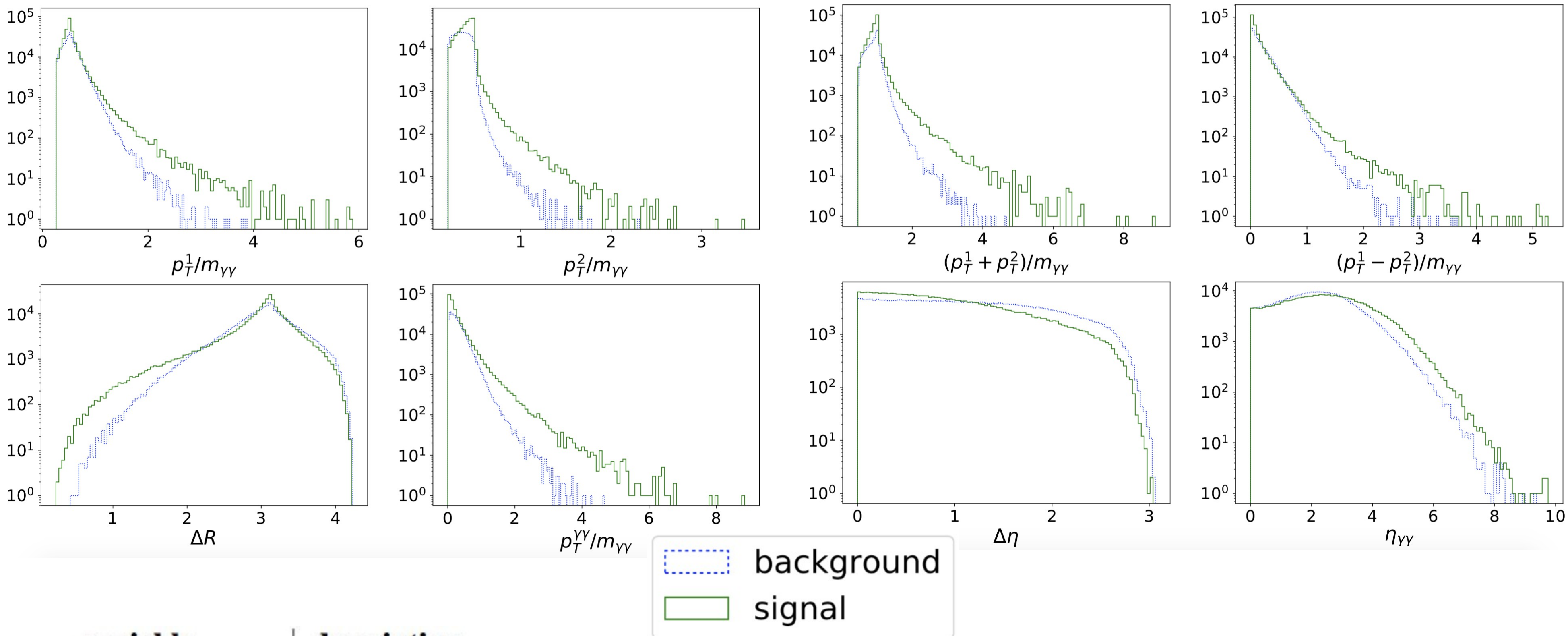
Generated with SHERPA at 8TeV proton-proton c.o.m energy

- Photon pT of 32 GeV and 25 GeV for realistic trigger selection
- Di-photon mass [122.5, 127.5] GeV
- Higgs candidate |η|<2.5

Sample Size and Folding

- 300k signal + 300k background total sample
  - **Training set**
    - 20 stratified, independent splits of sizes 100, 1000, 5000, 10k, 15k, 20k events
    - → Spread of classifier performance over the folds reported as the **uncertainty due to the choice of training sample, and initialization.**
  - **Testing set**
    - Remaining 100k+100k independent sample
    - → **Statistical error** on the classifier performance estimated using bootstrapping

Characterizing Variables



variable	description
$p_T^1/m_{\gamma\gamma}$	transverse momentum of the highest p_T photon divided by the invariant mass of the diphoton pair
$p_T^2/m_{\gamma\gamma}$	transverse momentum of the second-highest p_T photon divided by the invariant mass of the diphoton pair
$(p_T^1+p_T^2)/m_{\gamma\gamma}$	sum of the transverse momentum of the two photons divided by their invariant mass
$(p_T^1-p_T^2)/m_{\gamma\gamma}$	difference of the transverse momentum of the two photons divided by their invariant mass
$p_T^{\gamma\gamma}/m_{\gamma\gamma}$	transverse momentum of the diphoton system divided by its invariant mass
$\Delta\eta$	difference in $\eta = -\log\tan\left(\frac{\theta}{2}\right)$, where θ is the angle with the beam axis
ΔR	sum in quadrature of the separation of and ϕ, the azimuthal angle of the two photons $(\sqrt{\Delta\eta^2+\Delta\phi^2})$
$\lvert\eta^{\gamma\gamma}\rvert$	the η value of the diphoton system

Weak Classifier Function

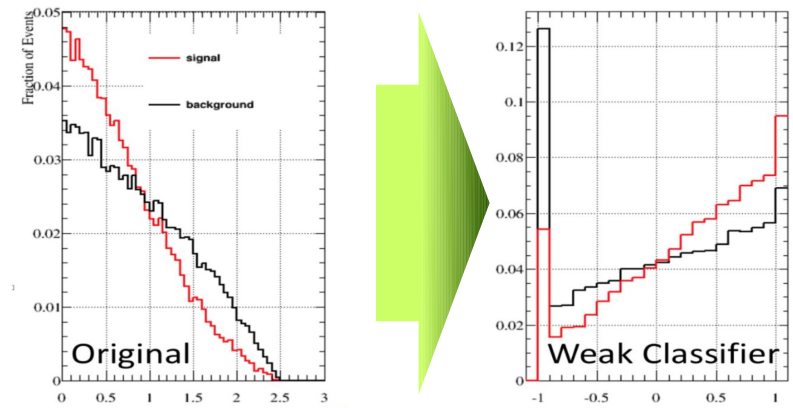
Define v_{shift}

- Based on 70^{th} and 30^{th} percentile of the signal distribution (s_{70}, s_{30})
- If the percentile of background at s_{70} is less than 70%, then translate to s_{70} and invert the variable
- Else, check the percentile of background at s_{30}, and if more than 30%, then translate to s_{30}.
- Else, the two distributions are “too overlapping” and we discard the variable.

Define h

- v_{+1} and v_{-1} are the 10^{th} and 90^{th} percentile of v_{shift}

$$h(v) = \begin{cases} +1 & \text{if } v_{+1} < v^{shift}(v) \\ \frac{v^{shift}(v)}{v_{+1}} & \text{if } 0 < v^{shift}(v) \leq v_{+1} \\ \frac{v^{shift}(v)}{|v_{-1}|} & \text{if } v_{-1} < v^{shift}(v) \leq 0 \\ -1 & \text{if } v^{shift}(v) < v_{-1} \end{cases}$$



Applied to all variables and their product (inverse if flipped)

Weak Classifiers Naming

1	2	3	4	5	6	7	8	9
p_T^1	p_T^2	ΔR	$p_T^{\gamma\gamma}$	$p_T^1 + p_T^2$	$p_T^1 - p_T^2$	$\Delta\eta$	$\eta_{\gamma\gamma}$	$(p_T^1 + p_T^2)\eta_{\gamma\gamma}$
10	11	12	13	14	15	16	17	18
$\frac{p_T^2}{p_T^1 - p_T^2}$	$\frac{p_T^2}{\Delta\eta}$	$p_T^2\eta_{\gamma\gamma}$	$\frac{1}{\Delta R p_T^{\gamma\gamma}}$	$\frac{p_T^1 + p_T^2}{\Delta R}$	$\frac{1}{\Delta R(p_T^1 - p_T^2)}$	$\frac{1}{\Delta R \Delta\eta}$	$\frac{\eta_{\gamma\gamma}}{\Delta R}$	$\frac{1}{(p_T^1 - p_T^2)\Delta\eta}$
19	20	21	22	23	24	25	26	27
$p_T^1 p_T^2$	$\frac{p_T^1}{\Delta R}$	$\frac{p_T^1}{p_T^{\gamma\gamma}}$	$p_T^1(p_T^1 + p_T^2)$	$\frac{p_T^1}{p_T^1 - p_T^2}$	$\frac{p_T^1}{\Delta\eta}$	$\frac{p_T^1}{\eta_{\gamma\gamma}}$	$\frac{p_T^2}{\Delta R}$	$\frac{\eta_{\gamma\gamma}}{p_T^1 - p_T^2}$
28	29	30	31	32	33	34	35	36
$\frac{p_T^2}{p_T^{\gamma\gamma}}$	$p_T^2(p_T^1 + p_T^2)$	$\frac{p_T^1 + p_T^2}{p_T^{\gamma\gamma}}$	$\frac{\eta_{\gamma\gamma}}{p_T^{\gamma\gamma}}$	$\frac{1}{p_T^{\gamma\gamma}\Delta\eta}$	$\frac{1}{p_T^{\gamma\gamma}(p_T^1 - p_T^2)}$	$\frac{p_T^1 + p_T^2}{p_T^1 - p_T^2}$	$\frac{p_T^1 + p_T^2}{\Delta\eta}$	$\frac{\eta_{\gamma\gamma}}{\Delta\eta}$

Experiments

Outline

- Train classical classifiers as a baseline measurement of performance.
- Evaluate the exact solution of the problem using simulating annealing of the Ising model.
- Scan for λ, penalty on number of weak classifiers.
- Classification performance depending on the size of the training set.
- Scan on the fraction of exited states included in the classifier.

Baseline Classifiers

Classical Baseline

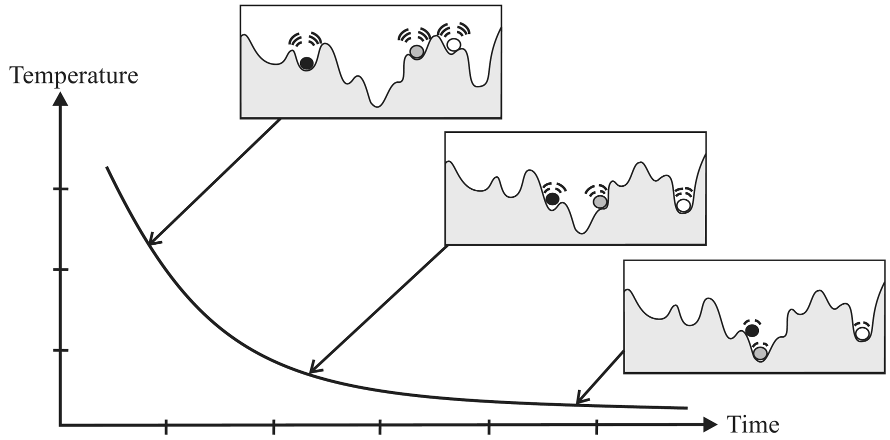
➔ XGBoost (XGB)
- Extremely efficient library for training decision trees
- http://xgboost.readthedocs.io
- Discovered during the higgs-ml challenge
  https://www.kaggle.com/c/higgs-boson
- Moderately optimize the hyper-parameters

➔ Deep Neural Network (DNN)
- Simple fully connected model 2 layers 1000 nodes
- https://keras.io/
  http://deeplearning.net/software/theano/
- Moderately optimize the hyper-parameters

Simulated Annealing

Ising Model Heuristic Solution

- Monte-Carlo based method to find ground state of energy functions
- Random walk across phase space
  - ➔ accepting descent
  - ➔ accepting ascent with probability $e^{-\Delta E/kT}$
- Decrease T with time



Applied to the QUBO problem, and finds the **ground state** reasonably well. SA in the legends.

Variable Importance

Weak Classifier Penalty

$$\delta(\vec{w}) \propto \sum_{i,j} C_{ij} w_i w_j + \sum_i (\lambda - 2C_{iy}) w_i$$

Penalize for using many weak classifiers

Surviving Weak Classifiers

λ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	0	20	20	20	19	20	20	0	5	20	20	20	20	19	20	17	20	20
0.01	0	20	20	20	19	20	20	0	4	20	20	20	20	19	20	17	20	20
0.02	0	20	20	20	19	20	20	0	4	20	20	20	20	19	20	16	20	20
0.05	0	20	20	20	19	20	20	0	1	20	20	20	20	19	20	10	20	17
0.1	0	20	20	20	19	20	20	0	0	20	20	20	20	19	20	6	14	2
0.2	0	20	20	20	19	20	20	0	0	20	14	20	20	12	20	4	1	0
0.4	0	20	0	2	19	20	20	0	0	20	17	20	20	0	20	1	0	0
0.8	0	20	0	0	0	0	9	0	0	18	0	0	20	0	2	0	0	0

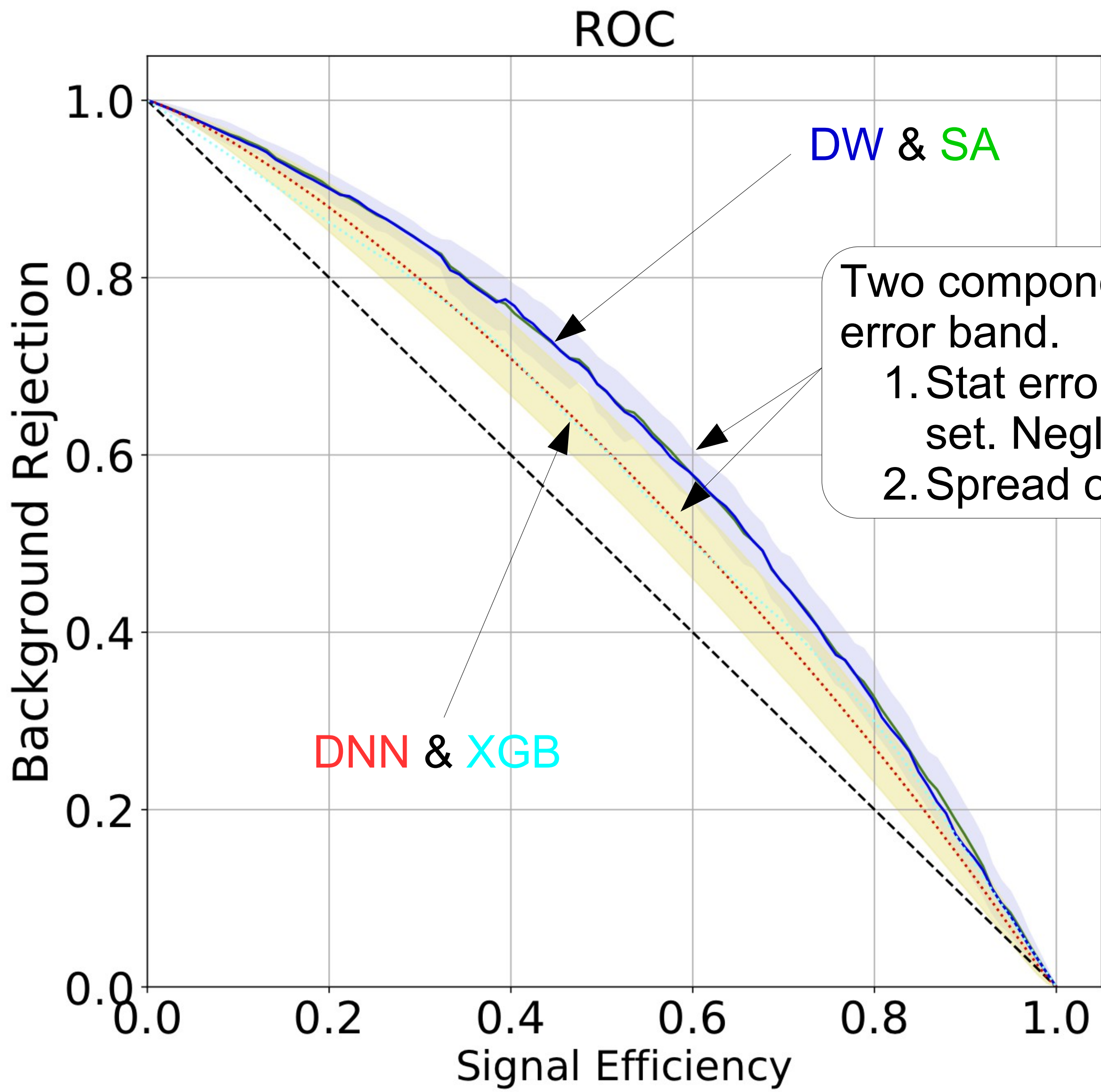
λ	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
0	20	0	0	19	0	20	0	3	0	20	19	7	0	15	0	19	20	20
0.01	20	0	0	19	0	20	0	2	0	20	19	6	0	15	0	19	20	20
0.02	20	0	0	19	0	20	0	1	0	20	19	4	0	15	0	19	20	20
0.05	20	0	0	19	0	20	0	0	0	20	16	1	0	11	0	19	20	20
0.1	20	0	0	1	0	20	0	0	0	20	1	0	0	5	0	16	20	20
0.2	18	0	0	0	0	20	0	0	0	20	0	0	0	0	0	0	20	20
0.4	0	0	0	0	0	7	0	0	0	20	0	0	0	0	0	0	20	3
0.8	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	19	0

Table : Number of times the weak classifier of a given variable is used in the ground state solution, as a function of the penalty

Three major variables (2, 13, 28) : p_T^2, $(\Delta R p_T^{\gamma\gamma})^{-1}$, and $\frac{p_T^2}{p_T^{\gamma\gamma}}$
Relates to the creation of a heavy particle (Higgs) with less transverse energy than typical QCD in the same mass range.

Classification Performance

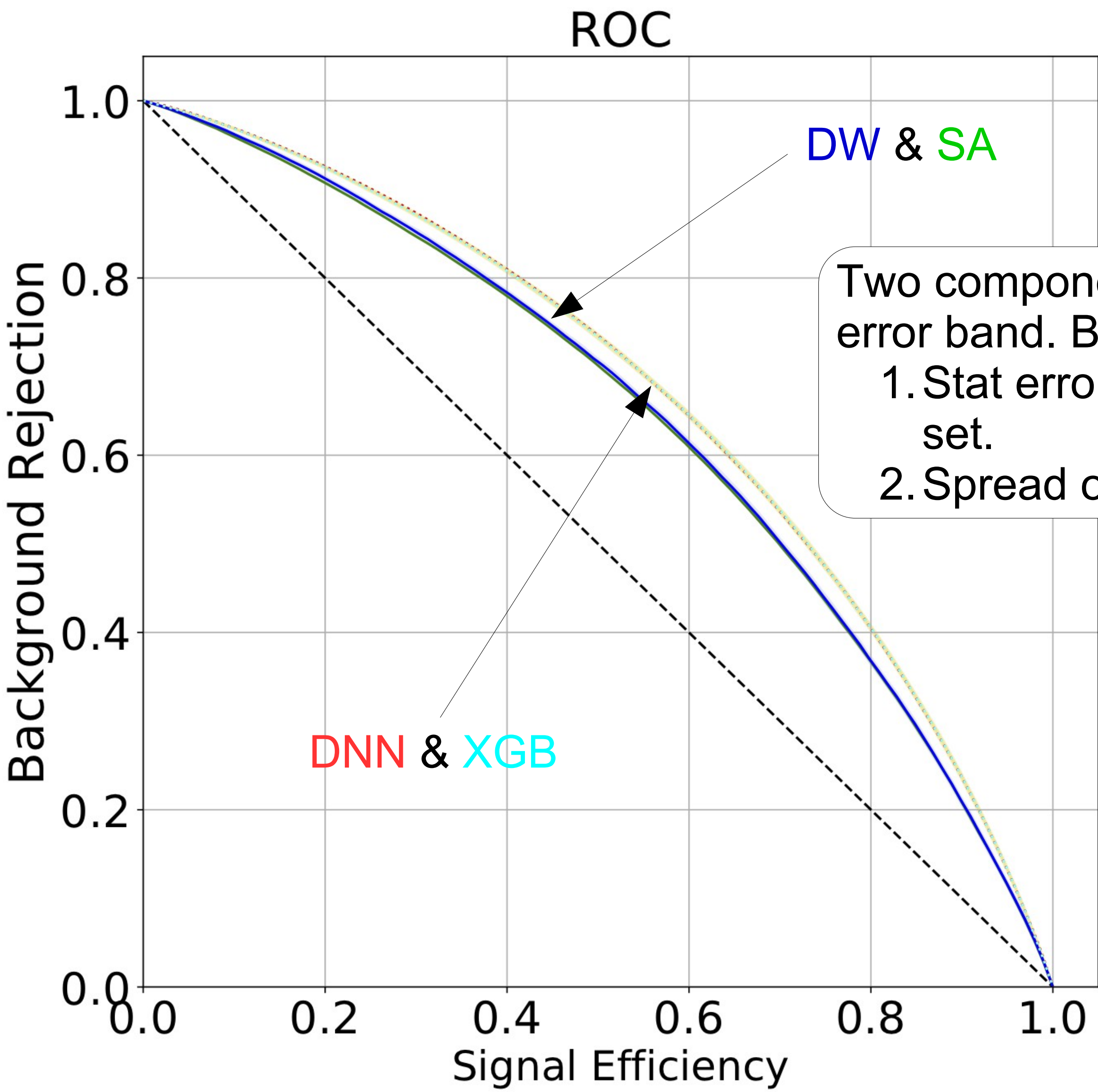
Sample Size of 100 Events



Two components in the error band.

1. Stat error of the test set. Negligible.
2. Spread over the folds.

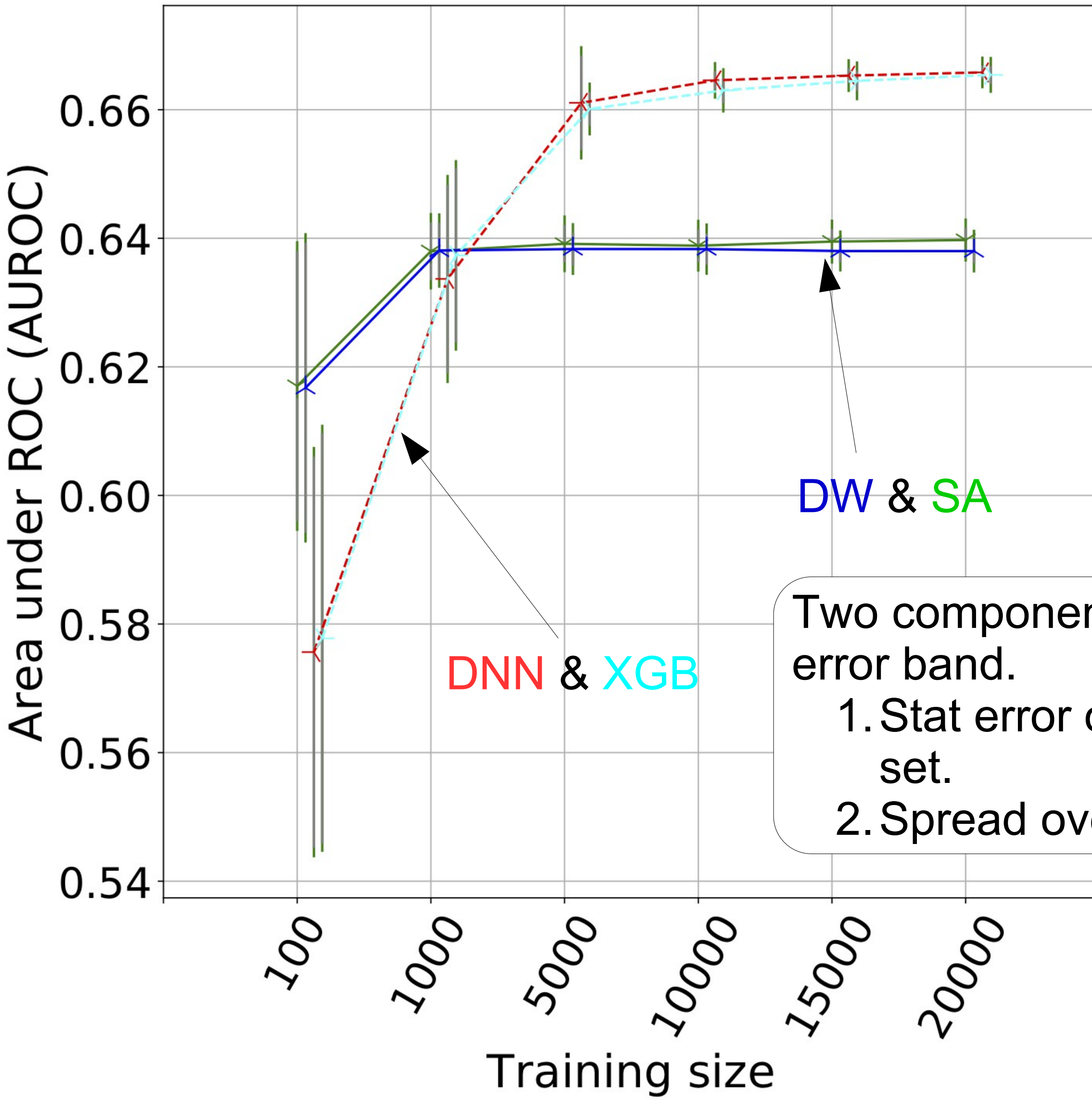
Sample Size of 20k Events



Two components in the error band. Both negligible

1. Stat error of the test set.
2. Spread over the folds.

Evolution With Training Size



Two components in the error band.

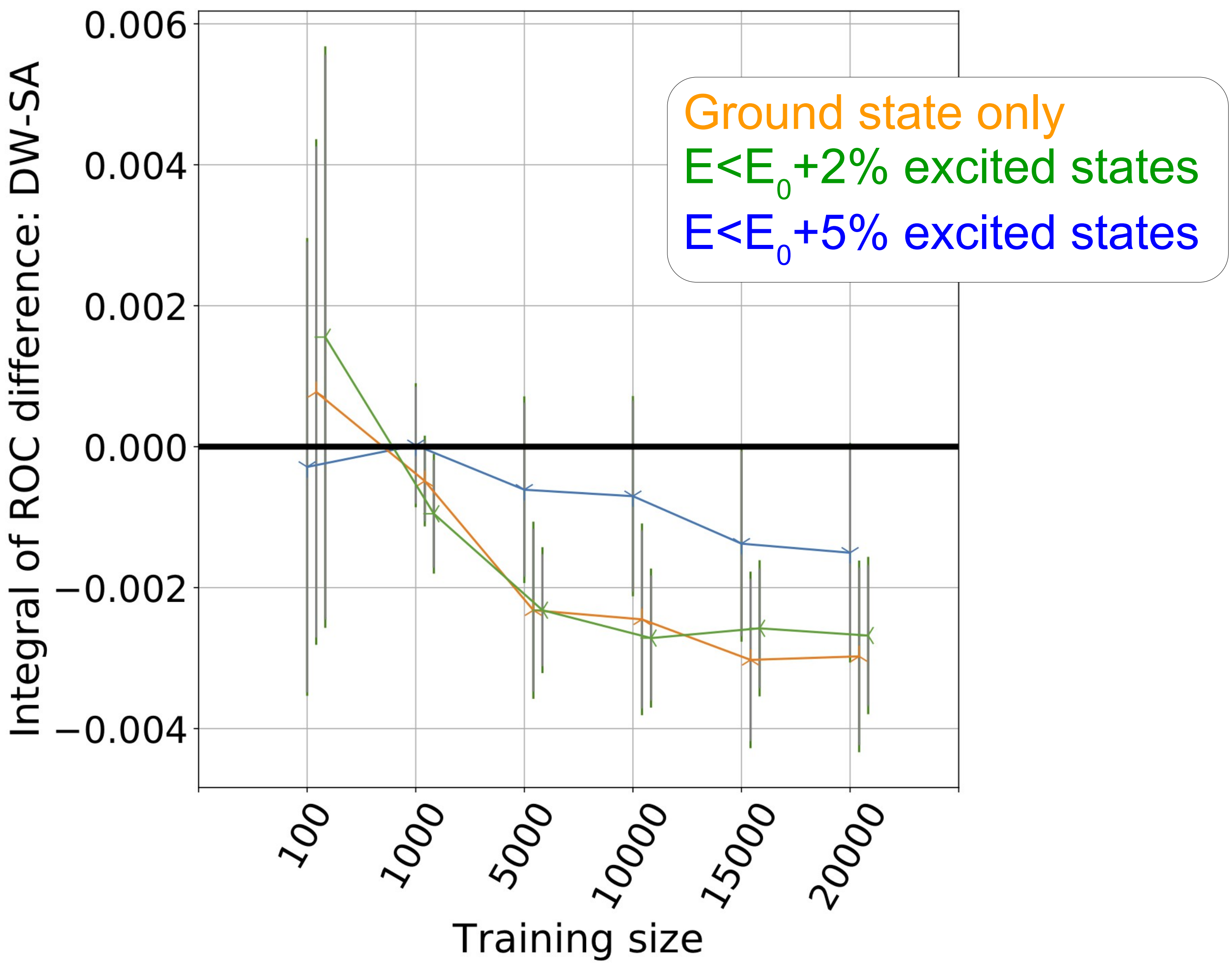
1. Stat error of the test set.
2. Spread over the folds.

Hybrid Classifier

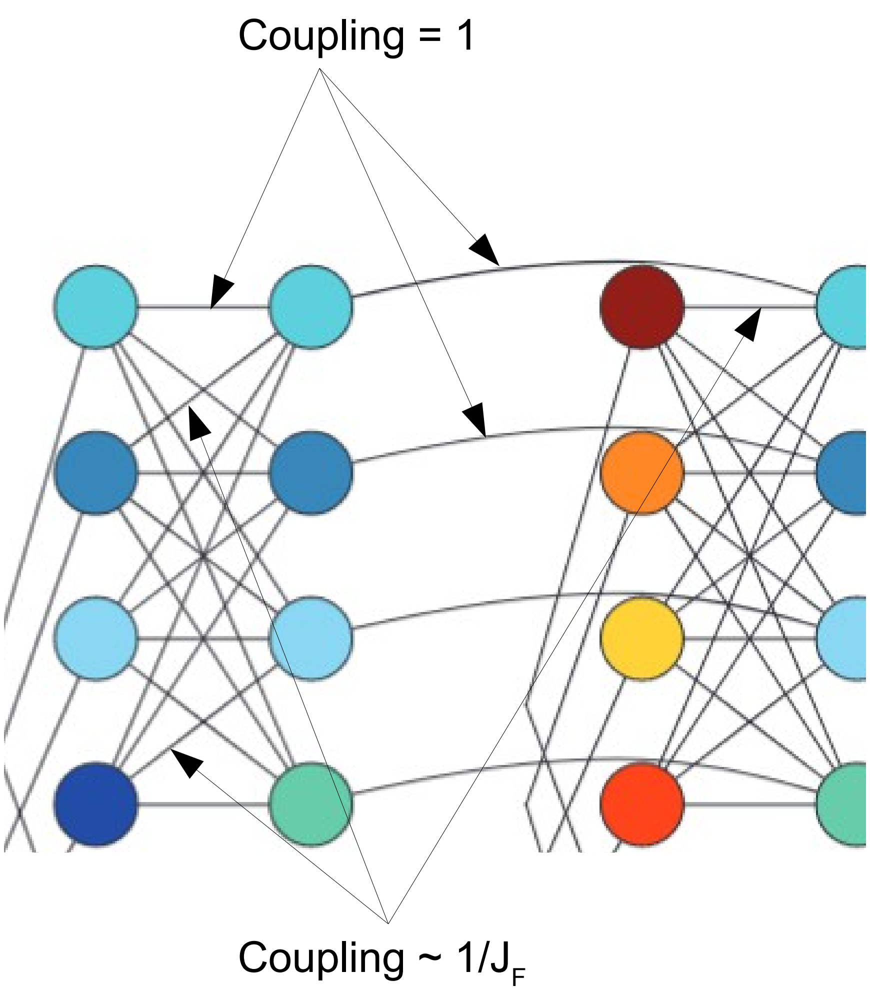
In presence of Excited States

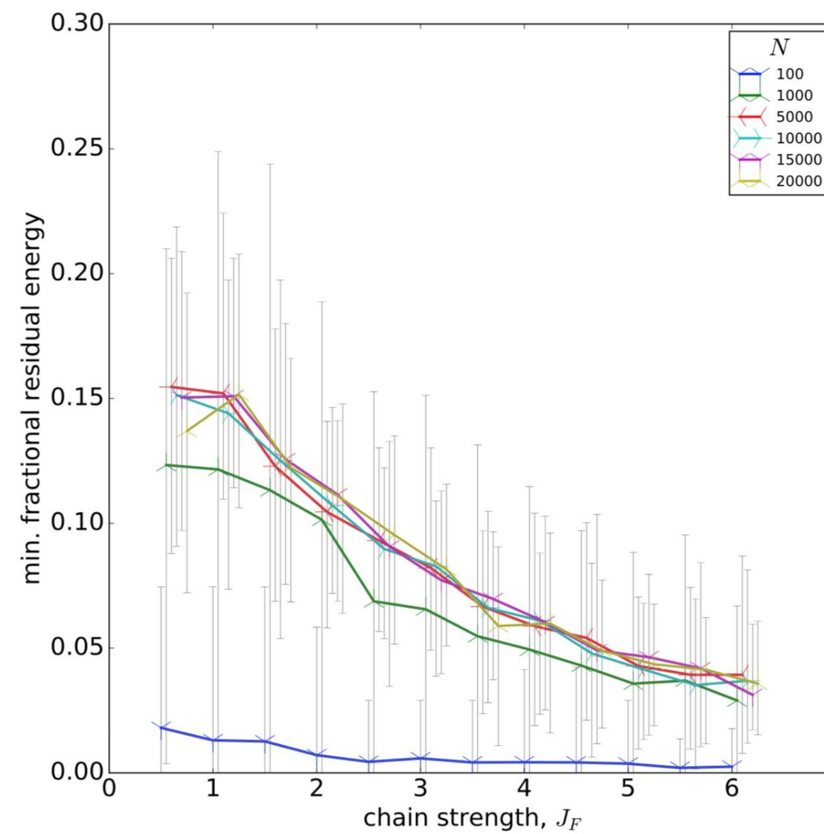
- Excited states bring new mask of weak classifier to build discriminators
- Building a discriminator including excited states
  - Take x% of the levels above the ground state
  - For each evaluate the average ROC on the training folds
  - At each signal efficiency, pick the energy state that has the most rejecting discriminator

Comparison DW/SA



Chimera Spin Chains

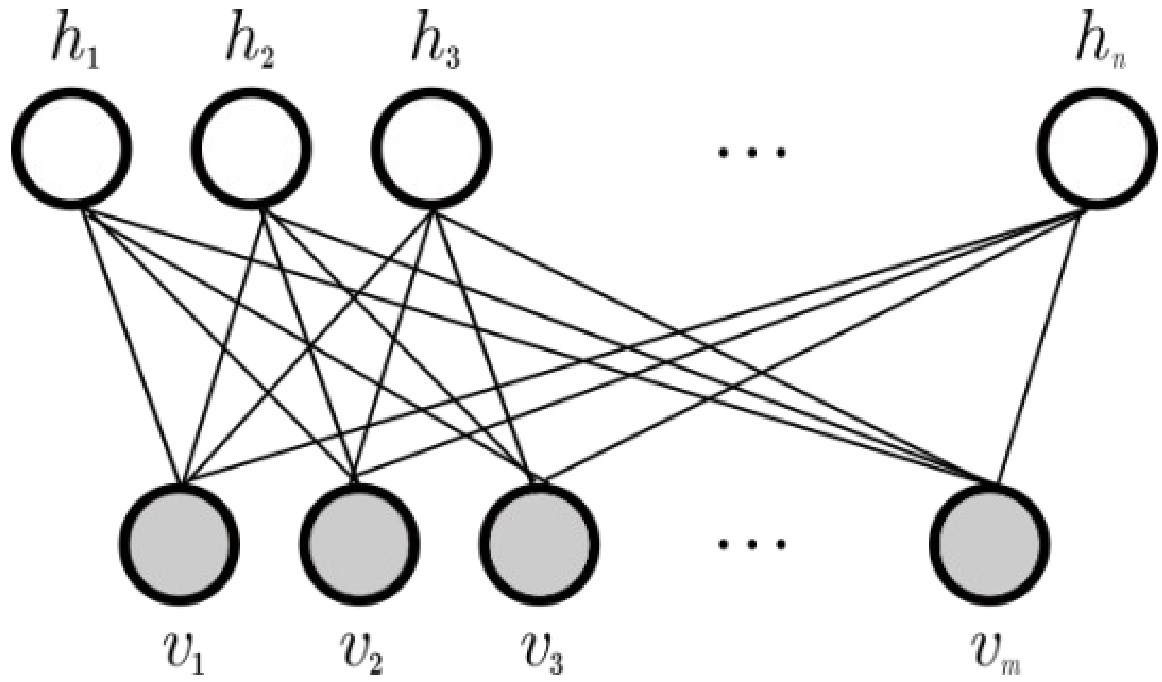




- Quantum Annealing gets better at finding the **ground truth** ground state with strong chain strength.
- SA does not include spin chains.
- Feedback to machine developers

Application Outlook

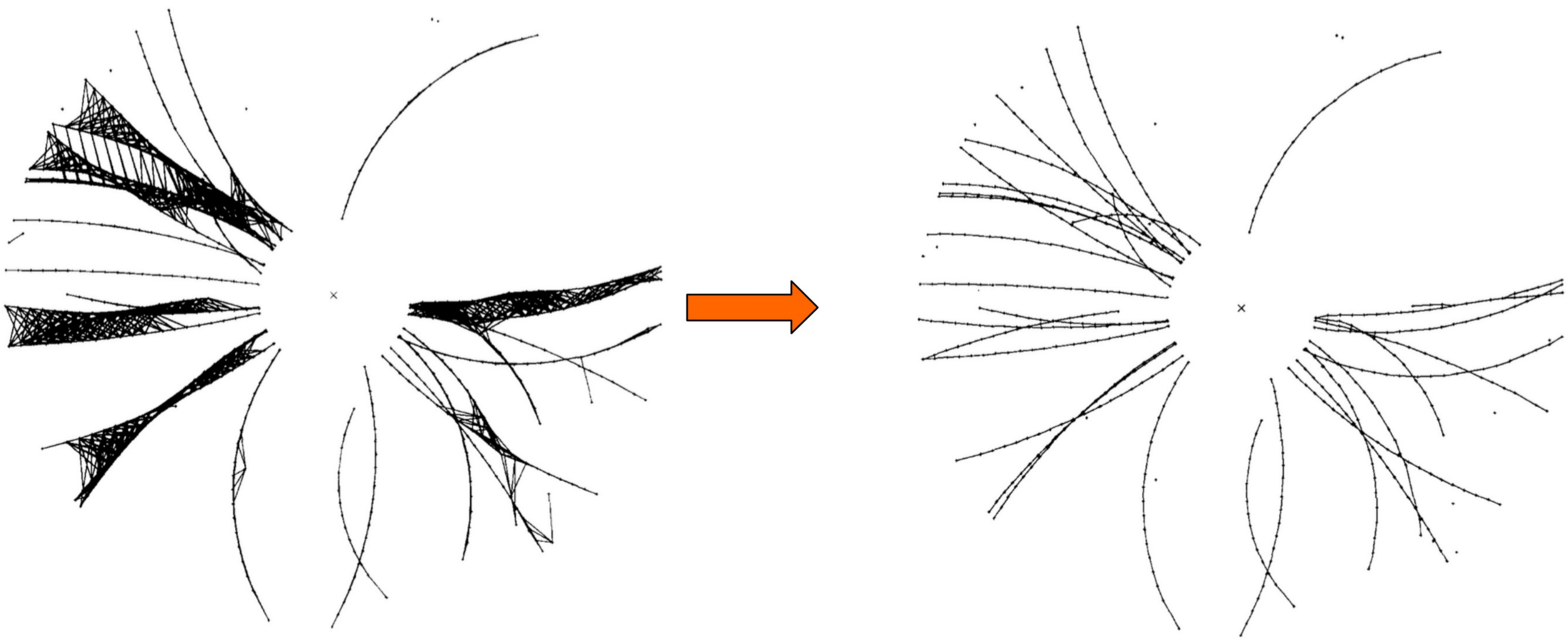
Boltzmann Machine



$$E = -\left(\sum_{i<j} w_{ij}\, s_i\, s_j + \sum_i \theta_i\, s_i \right)$$

https://arxiv.org/abs/1601.02036

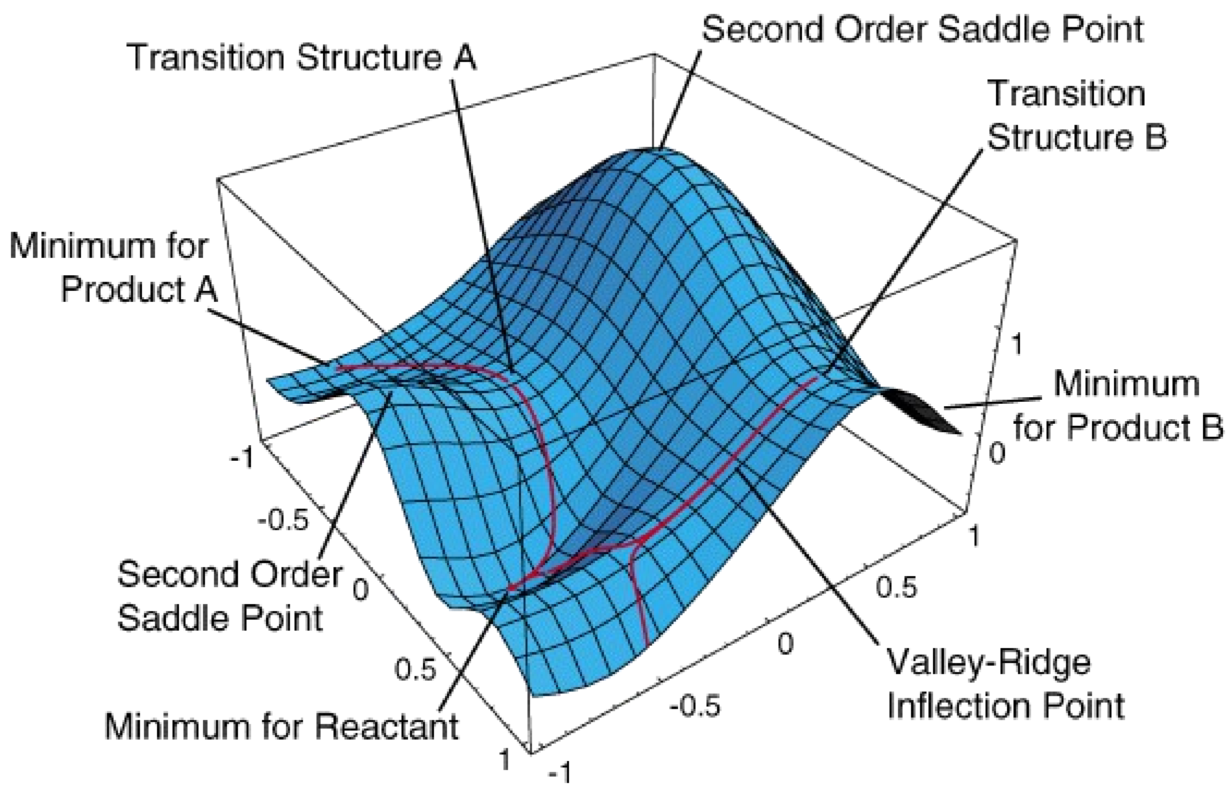
Hopfield Network



$$E = -\frac{1}{2}\sum_{i,j} w_{ij} s_i s_j + \sum_i \theta_i s_i$$

https://tinyurl.com/yayglkf4

Second Order Optimization



$$y = f(\mathbf{x} + \Delta\mathbf{x}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}} \Delta\mathbf{x} + \frac{1}{2!} \Delta\mathbf{x}^{\mathrm{T}} \mathbf{H}(\mathbf{x}) \Delta\mathbf{x}$$

Summary

- First application of D-Wave quantum annealing capability to High Energy Physics use case. Raises interesting questions.
- Scope of the Quantum Annealing on the D-Wave computing device is the solving the Ising model. Limited but powerful.
- Potential applicability of Quantum Annealing to other problems.

Thanks

This project is supported in part by the United States Department of Energy, Office of High Energy Physics Research Technology Computational HEP and Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359. The project is also supported in part under ARO grant number W911NF-12-1-0523 and NSF grant number INSPIRE-1551064. The work is supported in part by the AT&T Foundry Innovation Centers through INQNET, a program for accelerating quantum technologies. We wish to thank the Advanced Scientific Computing Research program of the DOE for the opportunity to first present and discuss this work at the ASCR workshop on Quantum Computing for Science (2015). We acknowledge the funding agencies and all the scientists and staff at CERN and internationally whose hard work resulted in the momentous H(125) discovery in 2012.

With special thanks to Joshua Job.