Reconstructing proton-proton collision positions at the Large Hadron Collider with a D-Wave quantum computer

Andrew Wildridge, Souvik Das, Sachin Vaidya, Andreas Jung

DPF2021: 2021 Meeting of the Division of Particles and Fields of the American Physical Society, July 12th - 14th 2021,
Florida State University, Video only (Virtual World)

Quantum Computers

Annealers

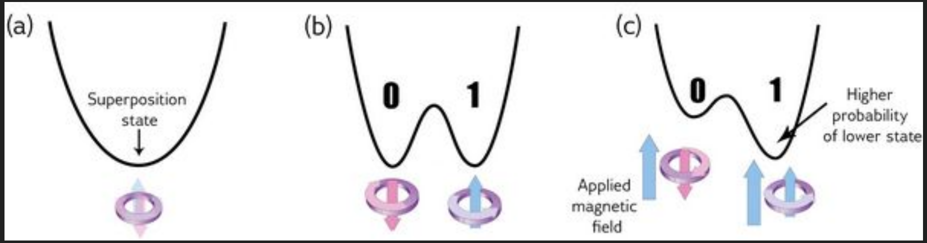


Image courtesy of: https://docs.dwavesys.com/docs/latest/c_gs_2.html

Circuits/Gates

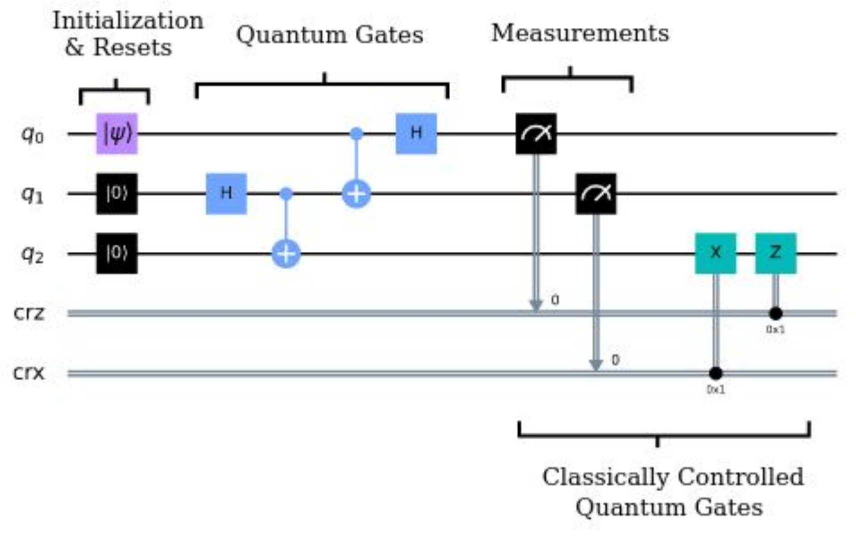
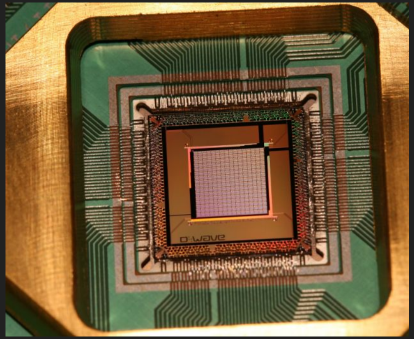


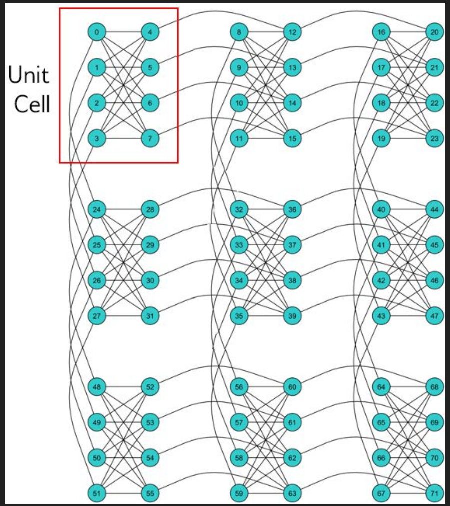
Image courtesy of: https://qiskit.org/textbook/ch-algorithms/defining-quantum-circuits.html

Quantum Annealers — D-Wave

- Quantum Processing Unit (QPU) made of rf-SQUIDs (radio frequency-superconducting quantum interference device) acting as qubits
  - Programmable external biases and couplings between qubits are made available
  - Not a fully connected graph of qubits
- System can be modeled as an **Ising model**

$$H_p = \sum_i h_i \sigma_z^i + \sum_i \sum_{j>i} J_{ij} \sigma_z^i \sigma_z^j$$



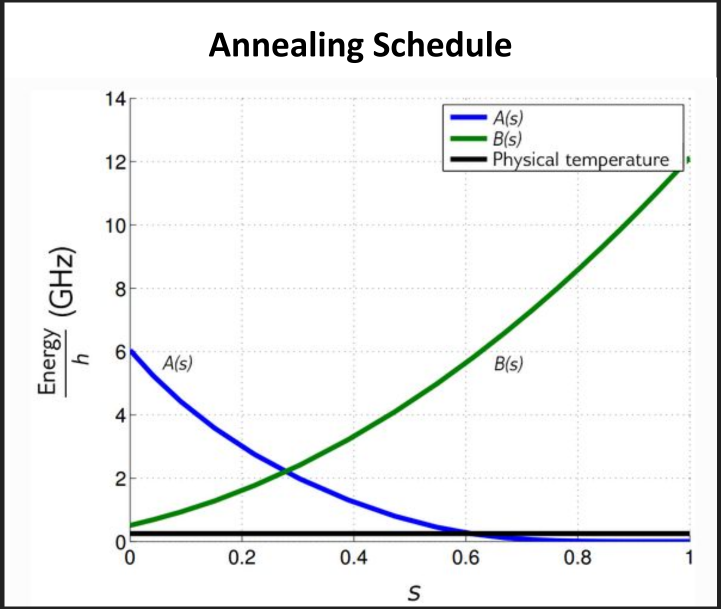


The chimera graph showcasing the limited connectivity of the qubits. [Link to D-Wave]

Quantum Annealers — Annealing Schedule

- Practical approximation to an adiabatic quantum computer
- *Adiabatic Theorem - A physical system remains in its instantaneous eigenstate if a given perturbation is acting on it slowly enough and if there is a gap between the eigenvalue and the rest of the Hamiltonian's spectrum* [1]
- Final state is the ground state and the optimal solution to the problem Hamiltonian

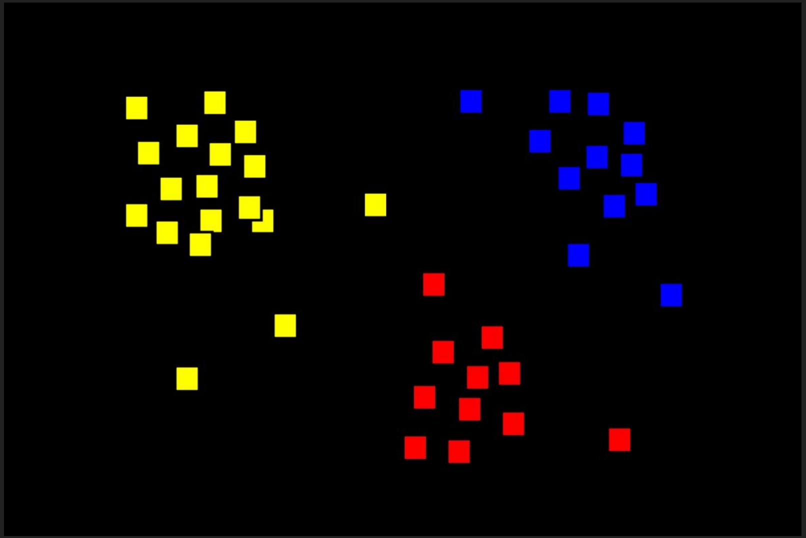
$$\mathcal{H} = -\frac{A(s)}{2}\left(\sum_i \sigma_x^i\right) + \frac{B(s)}{2}\left(H_p\right)$$



The annealing schedule and functions A(s) and B(s). [Link to D-Wave]

What is a Good Problem to Solve with a Quantum Annealer?

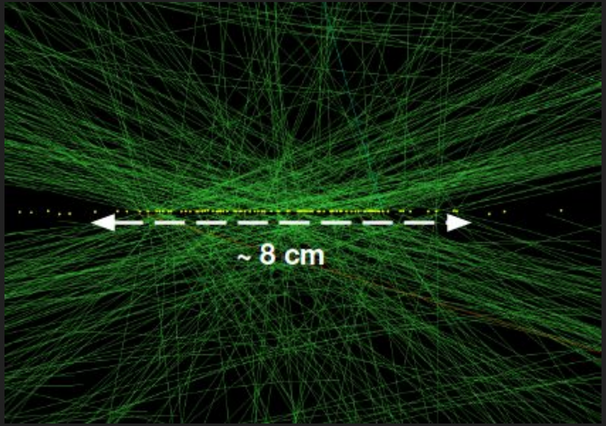
1. Combinatorial optimization problems
2. Fast to check, hard to solve (classically)



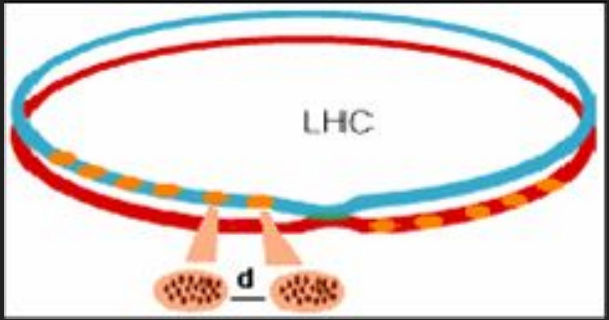
Solution...Clustering!

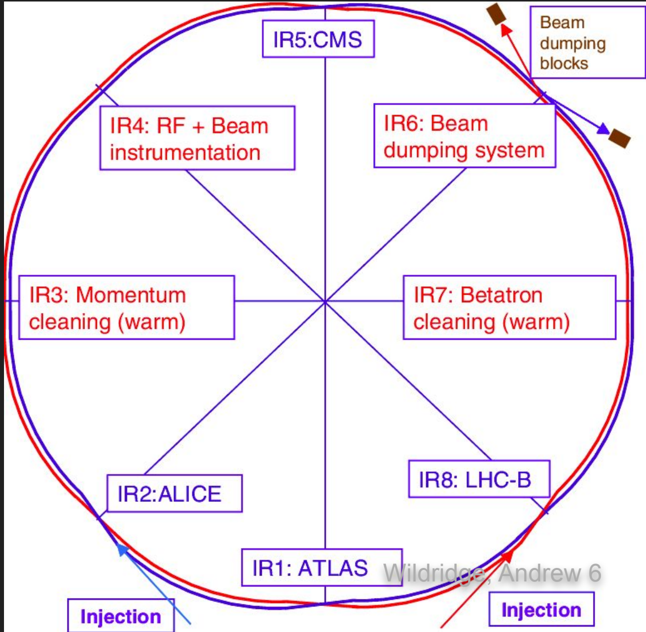
Introduction

- Counter-rotating beams of bunches of protons cross, producing multiple collisions of protons
- Clustering resulting tracks determines the p-p collision points
- Centroid-based clustering is NP-hard

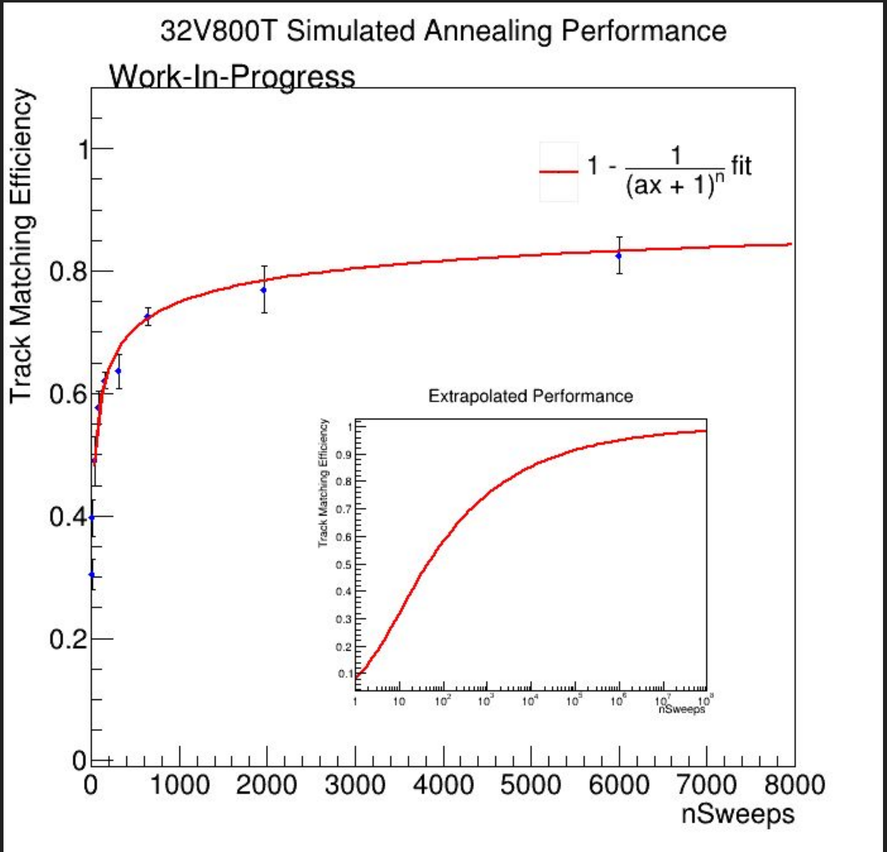


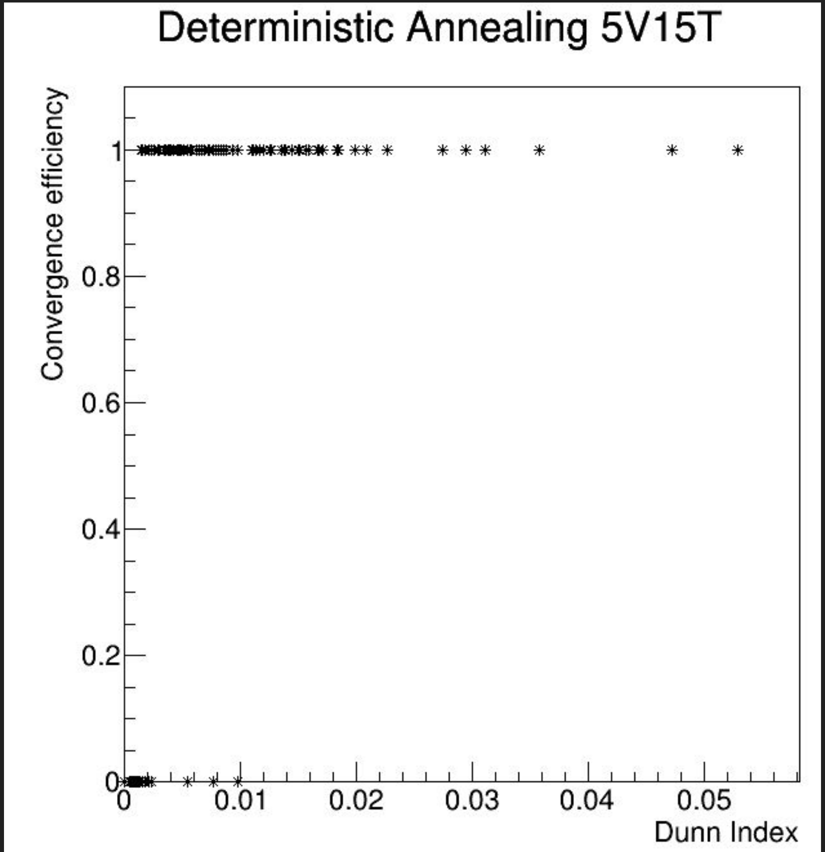
An event in CMS with 78 p-p collisions. Green lines are charged particle tracks, yellow dots are p-p collisions.



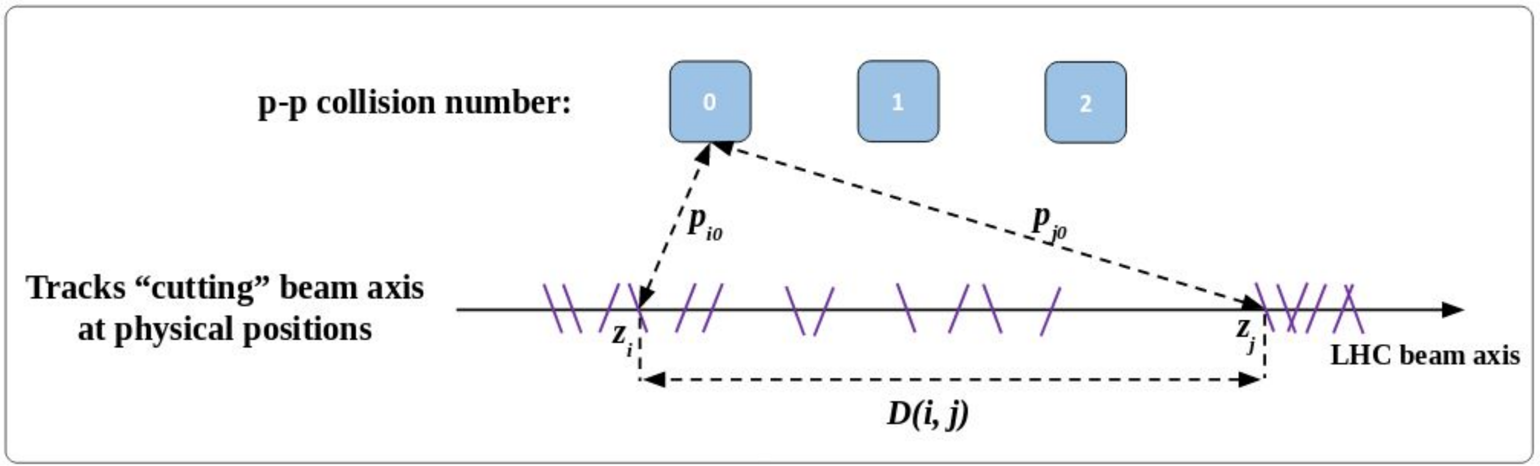


Classical Formulations





The Formulation



A graphical representation of the algorithm

The Formulation

$$H_p = \sum_{k}^{n_V} \sum_{i}^{n_T} \sum_{j>i}^{n_T} p_{ik}\, p_{jk}\, g\left(D(i,j); m\right) + \lambda \sum_{i}^{n_T} \left(1 - \sum_{k}^{n_V} p_{ik}\right)^2$$

The Formulation

$$H_p = \sum_{k}^{n_V} \sum_{i}^{n_T} \sum_{j>i}^{n_T} \underbrace{p_{ik}\, p_{jk}}\, g\left(D(i,j); m\right) + \lambda \sum_{i}^{n_T} \left(1 - \sum_{k}^{n_V} p_{ik}\right)^2$$

p_{ik} is the probability that the i^{th} track belongs to the k^{th} cluster. $p_{ik} \in \{0, 1\}$

The Formulation

$$D(i,j) = \frac{|z_i - z_j|}{\sqrt{\delta z_i^2 + \delta z_j^2}}$$

z_i is the location of closest approach to the beam axis for the particle track

$$H_p = \sum_k^{n_V} \sum_i^{n_T} \sum_{j>i}^{n_T} \underbrace{p_{ik}\, p_{jk}}\, g\left(\overbrace{D(i,j)}; m\right) + \lambda \sum_i^{n_T} \left(1 - \sum_k^{n_V} p_{ik}\right)^2$$

p_{ik} is the probability that the i^{th} track belongs to the k^{th} cluster. $p_{ik} \in \{0, 1\}$

The Formulation

$$D(i,j) = \frac{|z_i - z_j|}{\sqrt{\delta z_i^2 + \delta z_j^2}}$$

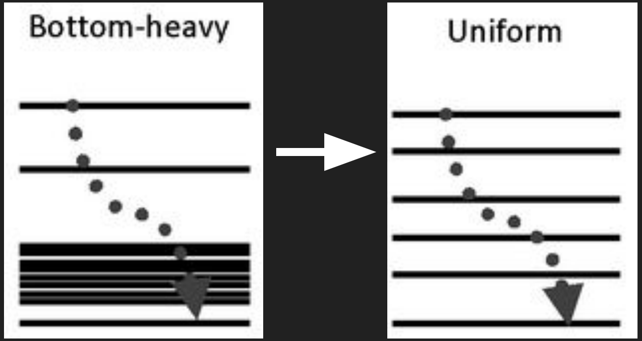
z_i is the location of closest approach to the beam axis for the particle track

$$H_p = \sum_{k}^{n_V} \sum_{i}^{n_T} \sum_{j>i}^{n_T} \underbrace{p_{ik}\, p_{jk}}\, g\underbrace{\Big(\overbrace{D(i,j)}; m\Big)} + \lambda \sum_{i}^{n_T} \left(1 - \sum_{k}^{n_V} p_{ik}\right)^2$$

p_{ik} is the probability that the i^{th} track belongs to the k^{th} cluster. $p_{ik} \in \{0, 1\}$

Distortion function used to distribute energy levels more uniformly

$$g(x; m) = 1 - e^{-mx}$$



The Formulation

$$D(i,j) = \frac{|z_i - z_j|}{\sqrt{\delta z_i^2 + \delta z_j^2}}$$

Constraint to ensure a track belongs to a single cluster

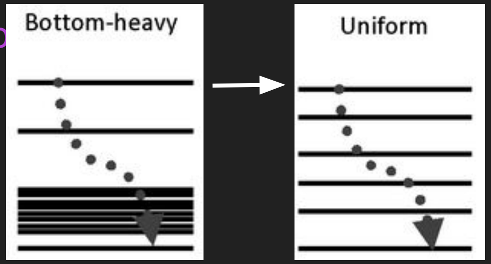
z_i is the location of closest approach to the beam axis for the particle track

$$H_p = \sum_k^{n_V} \sum_i^{n_T} \sum_{j>i}^{n_T} p_{ik}\, p_{jk}\, g\left(D(i,j); m\right) + \lambda \sum_i^{n_T} \left(1 - \sum_k^{n_V} p_{ik}\right)^2$$

p_{ik} is the probability that the i^{th} track belongs to the k^{th} cluster. $p_{ik} \in \{0, 1\}$

Distortion function used to distrib
energy levels more uniformly

$$g(x;m) = 1 - e^{-mx}$$



The Formulation

$$D(i,j) = \frac{|z_i - z_j|}{\sqrt{\delta z_i^2 + \delta z_j^2}}$$

z_i is the location of closest approach to the beam axis for the particle track

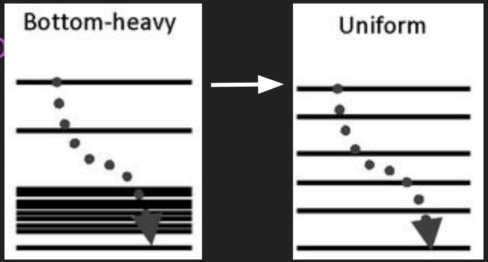
Constraint to ensure a track belongs to a single cluster

$$H_p = \sum_{k}^{n_V} \sum_{i}^{n_T} \sum_{j>i}^{n_T} \underbrace{p_{ik}\, p_{jk}}\, g\underbrace{\left(\overbrace{D(i,j)}; m\right)} + \underbrace{\lambda} \sum_{i}^{n_T} \left(\overbrace{1 - \sum_{k}^{n_V} p_{ik}}\right)^2$$

p_{ik} is the probability that the i^{th} track belongs to the k^{th} cluster. $p_{ik} \in \{0, 1\}$

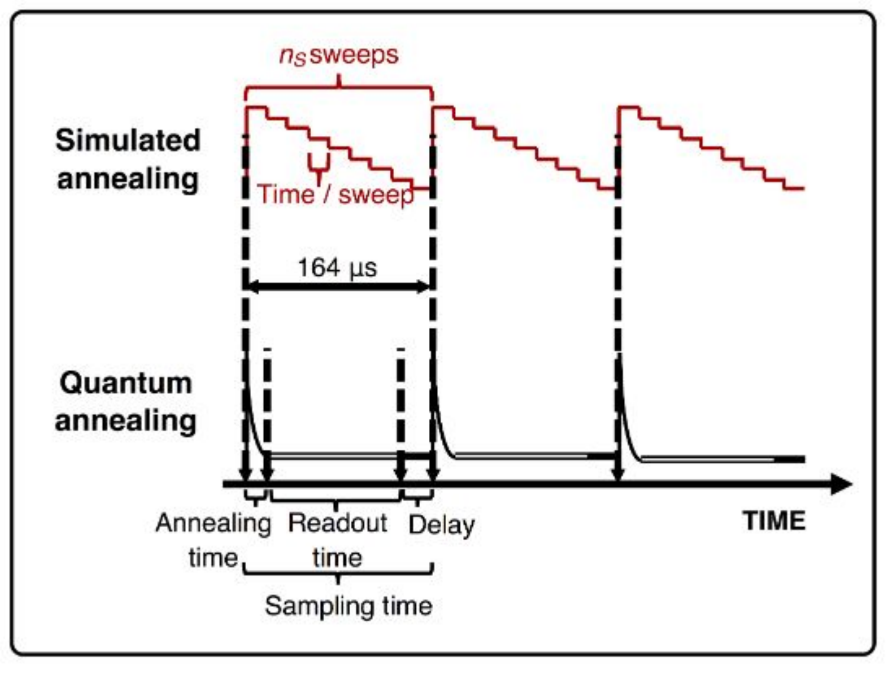
Distortion function used to distrib energy levels more uniformly

$$g(x; m) = 1 - e^{-mx}$$

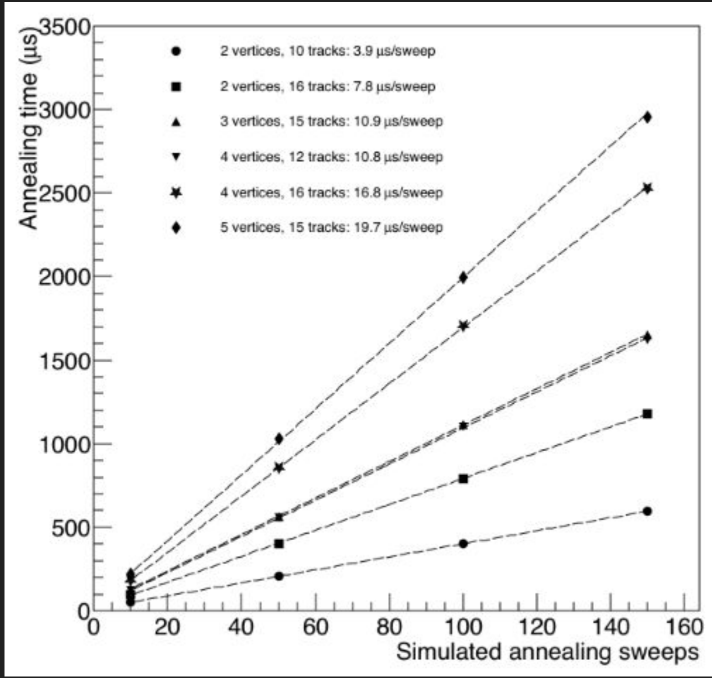


Penalty strength parameter

Benchmarking

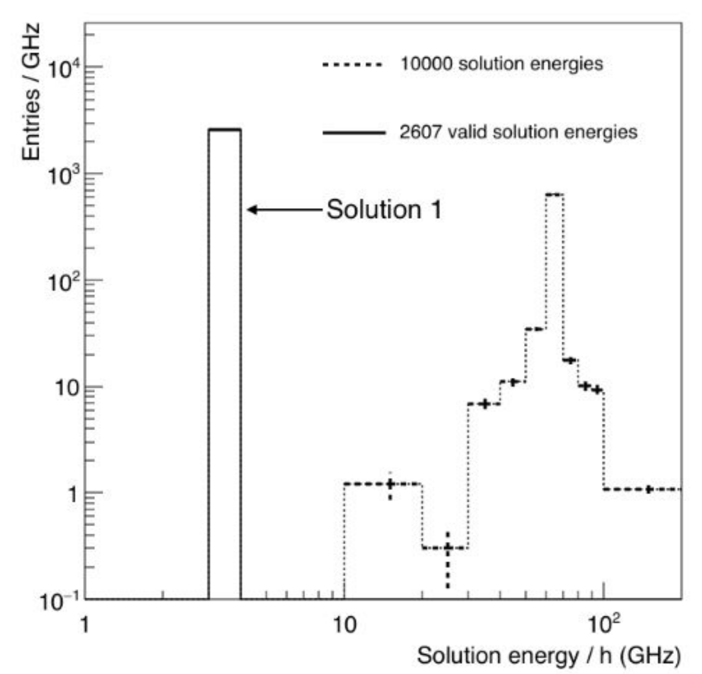


A diagram showing how we are restricting the amount of time the CPU is allowed to perform computations for

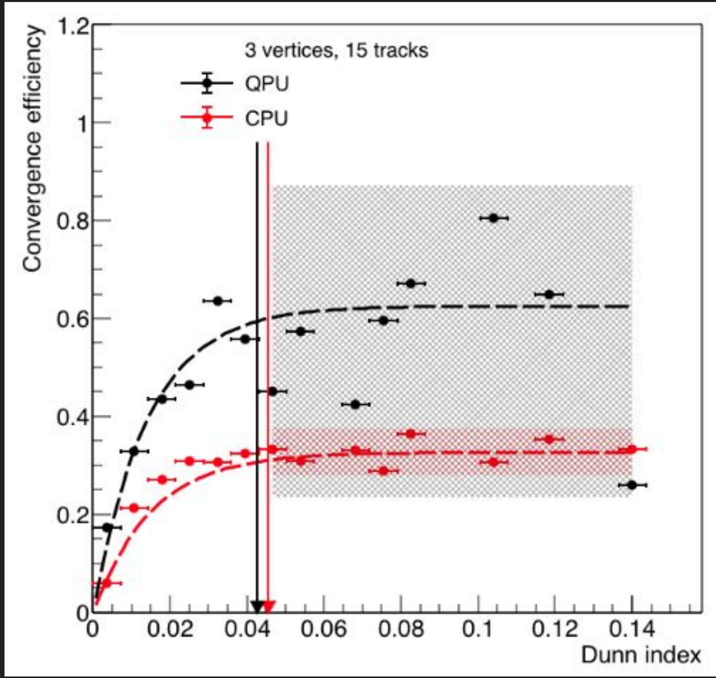


Linear regressions used to determine the time per sweep in the SA algorithm

Results



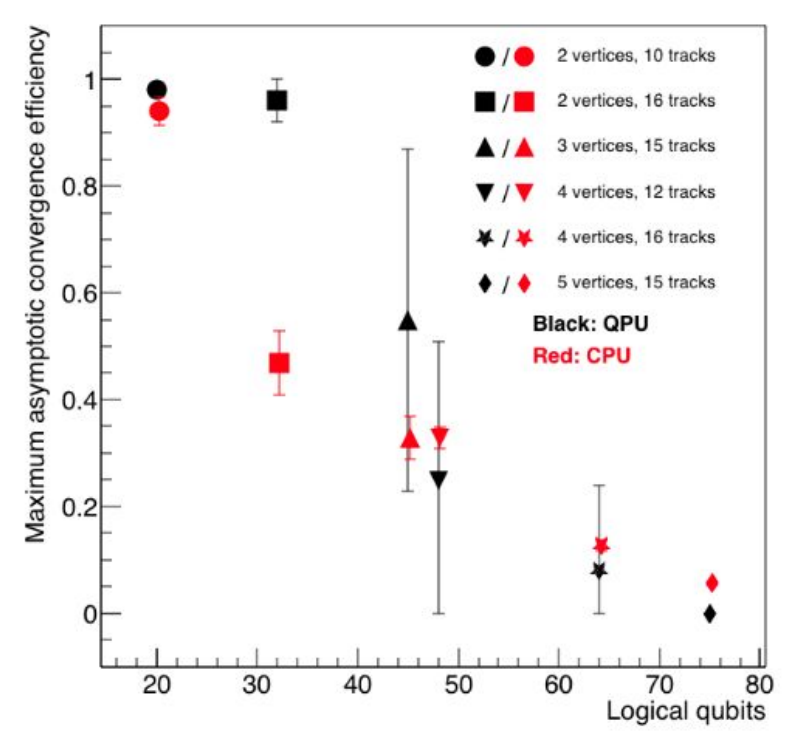
The energy spectrum of solutions for one event with 3 p-p collisions and 15 tracks explored by the QPU with 10,000 samples



A histogram of QPU convergence efficiency for 3 p-p collisions and 15 tracks using 100 events.

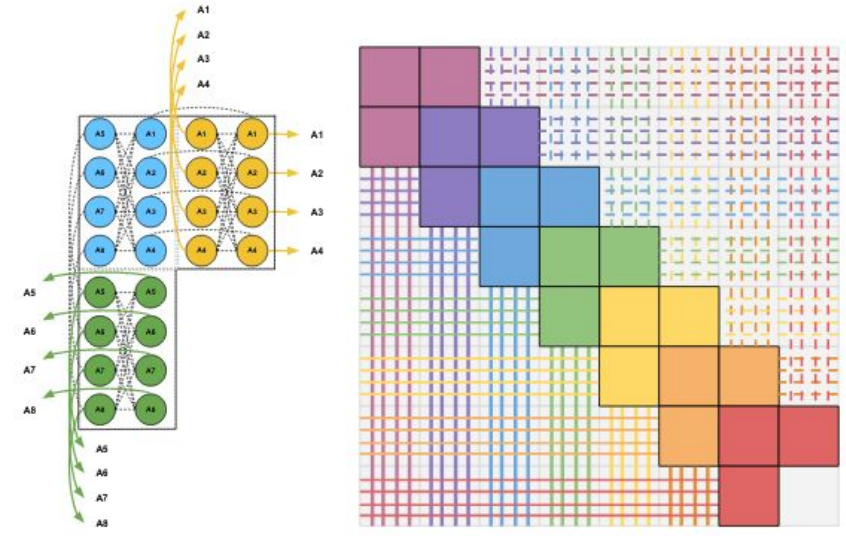
Intermediate Results

- Efficiency decreases with problem complexity
- Could have been used for Tevatron
- Interesting “quantum advantage” for 2 vertices 16 tracks
  - “Sweet spot” for QPU?



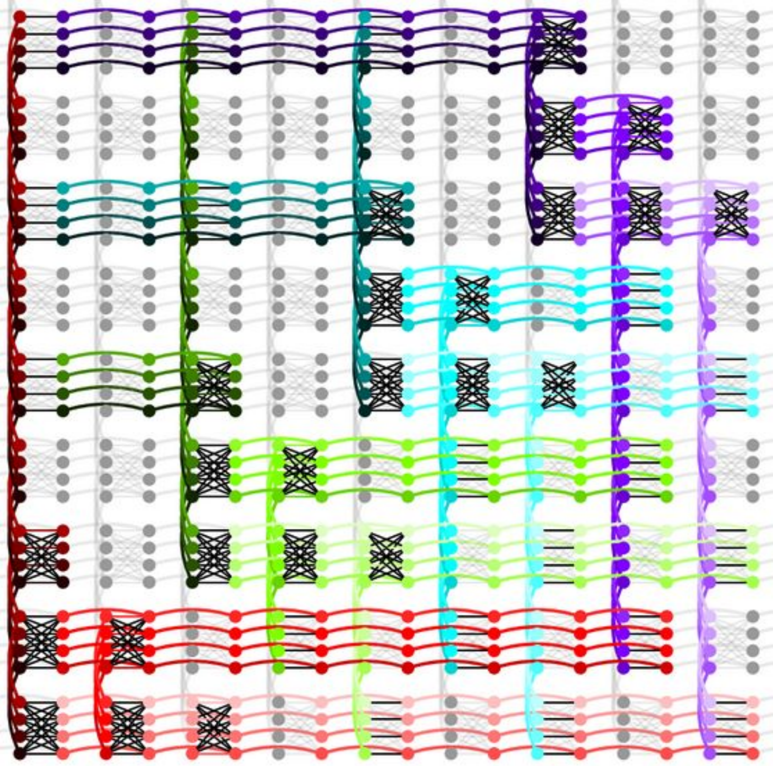
Plot of convergence efficiencies for various event topologies

Optimizations - Deterministic Embedding



(a) A possible nexus choice for a K_8 graph

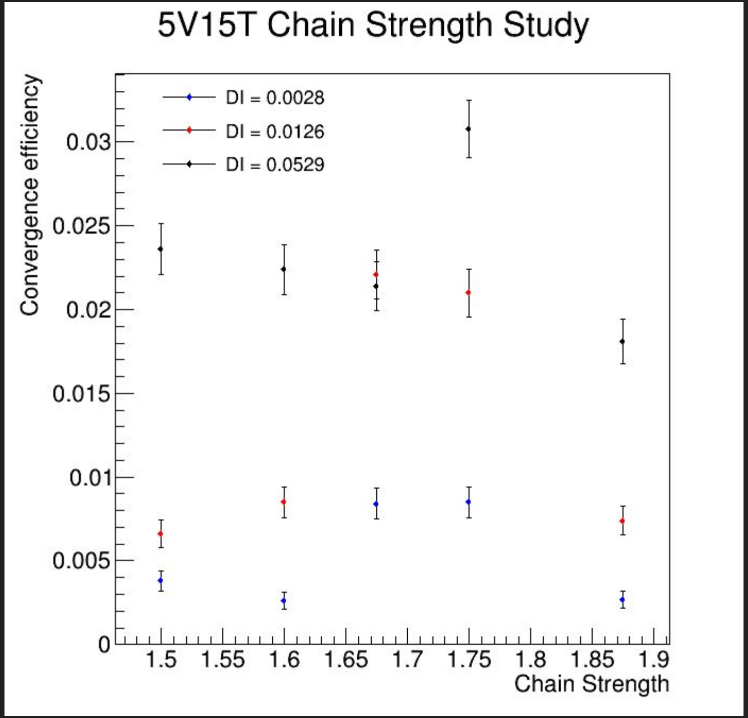
(b) An embedding for $K_8 \square K_7$



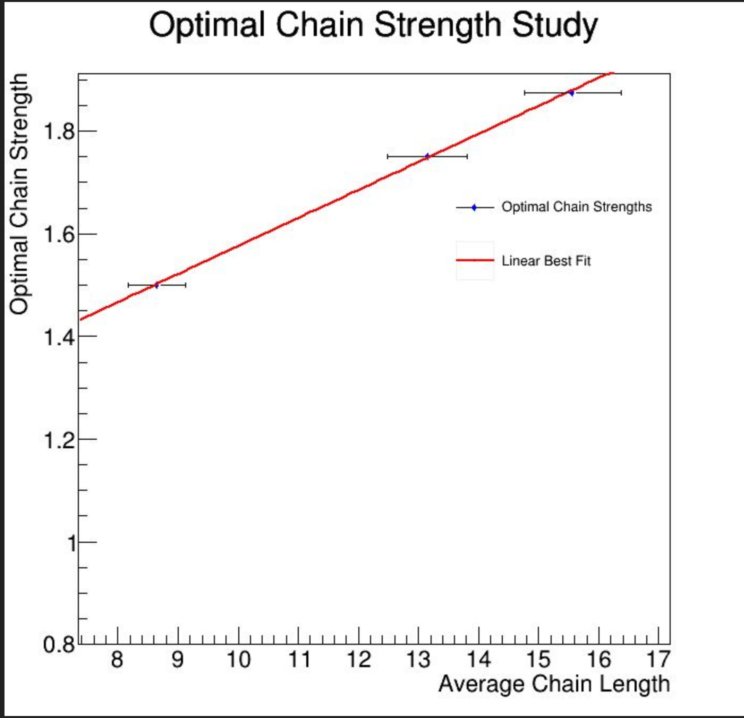
Our implementation for 4 vertices, 12 tracks

Deterministic embedding adapted from:
https://arxiv.org/pdf/1602.04274.pdf

Optimizations - Chain Strength Optimizations



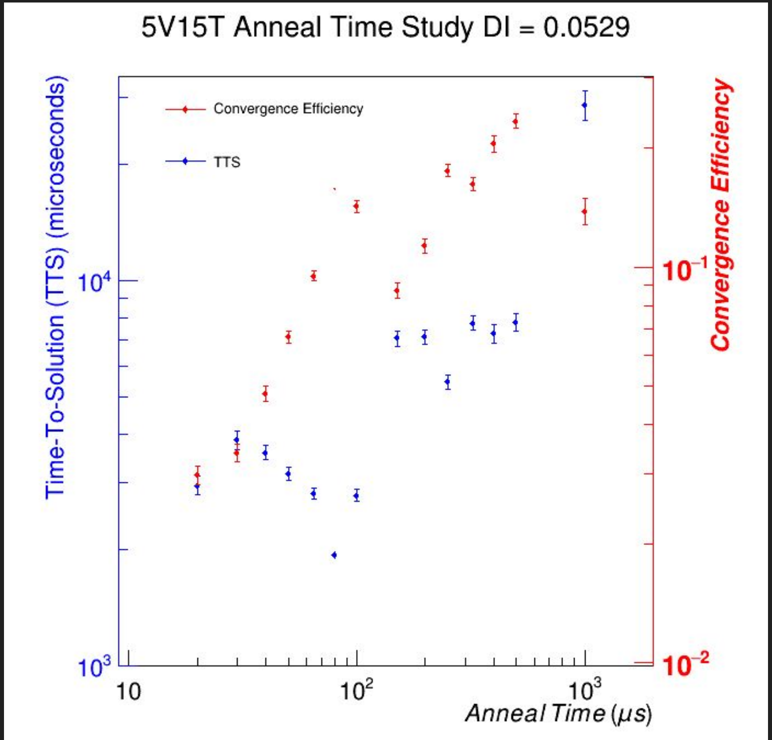
Efficiency solving 5 vertices, 15 tracks, for a variety of dunn indices while varying chain strength



Linear regression showing the relationship between the average chain length in the embedding and the optimal chain strength

Optimizations - Anneal Time

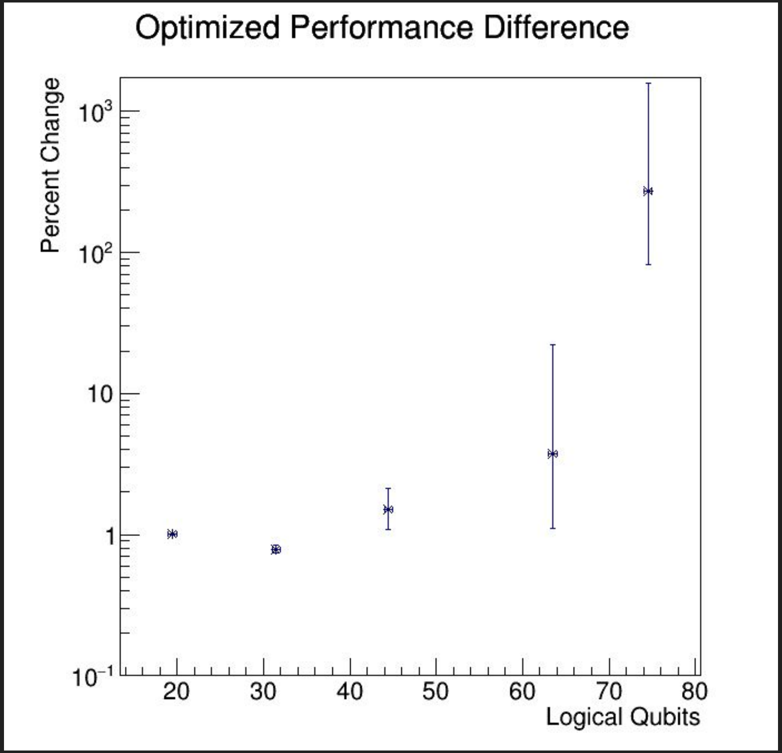
- Looked at optimizing anneal time for 3V15T, 5V15T, and 4V20T
- Logical qubits < 50 → anneal time = 20 microseconds
- Logical qubits > 50 → anneal time = 80 microseconds



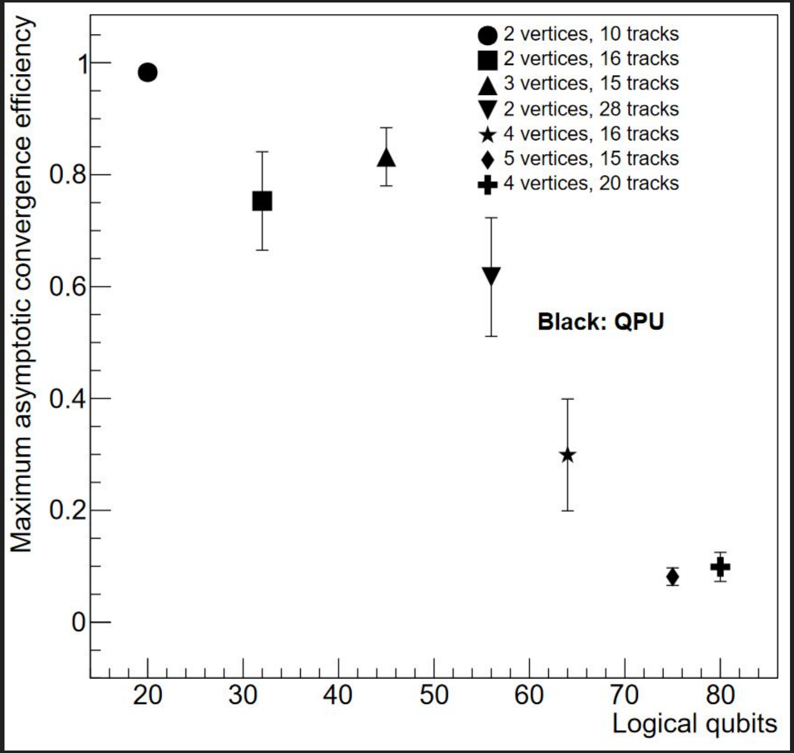
$$\mathrm{TTS}(t_f) = t_f R(t_f) \frac{N}{N_{\max}}, \quad R(t_f) = \frac{\ln(1-p_{\mathrm{d}})}{\ln[1-p_{\mathrm{S}}(t_f)]}$$

Final Results

Huge improvement at high logical qubits!



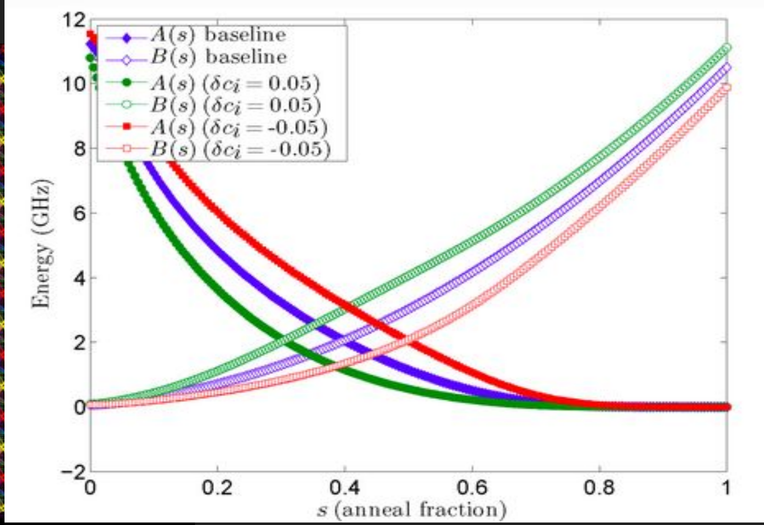
Ratio plot for comparing old results versus optimized results



Optimized results for a variety of event complexities

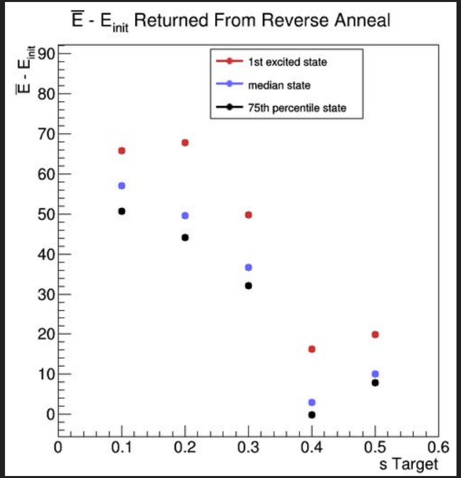
Summary

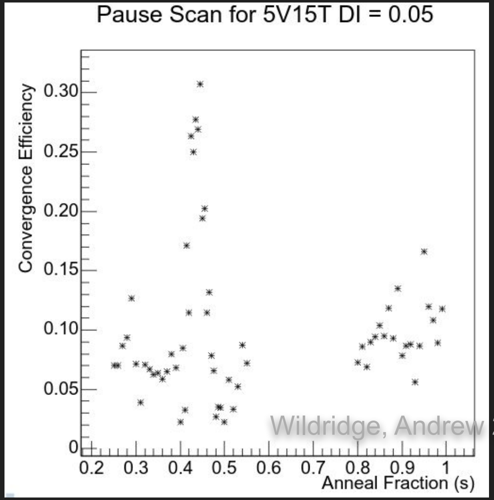
- Determining p-p collision points with track clustering is possible with QA
- Could have been used at Tevatron, does not currently scale to LHC
- Tons more optimizations out there
  - Reverse Annealing, Pegasus architecture, Pause during anneal, Anneal offsets, ….



D-WAVE

Effect of anneal offsets on A(s) and B(s). [Link to D-Wave]





Questions?

E-mail: awildrid@purdue.edu

Backup

Table of Contents

1. QA vs Gate
2. CPCG Embedding
3. Optimization Breakdown
4. Dataset Information
5. Simulated Annealing
6. Is D-Wave Quantum?
7. P, NP, NP-complete
8. Deterministic Annealing
9. Reverse Annealing
10. Anneal Pauses
11. Optimizing Squeeze Strength

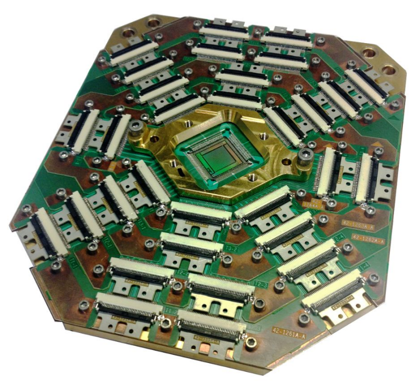
Quantum Annealing vs. Gate Model

Table of Contents

Quantum Computers

Annealers

- 5436 qubits (Advantage_system1.1)
- Non-universal**
- 37,440 couplings between qubits



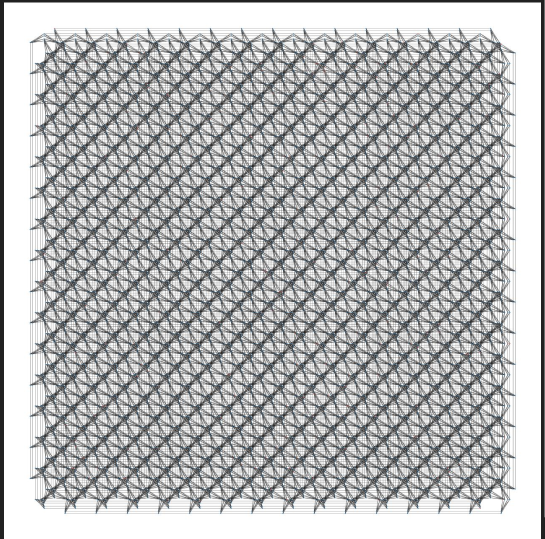
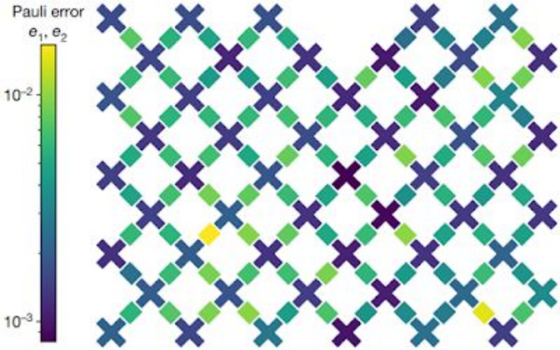
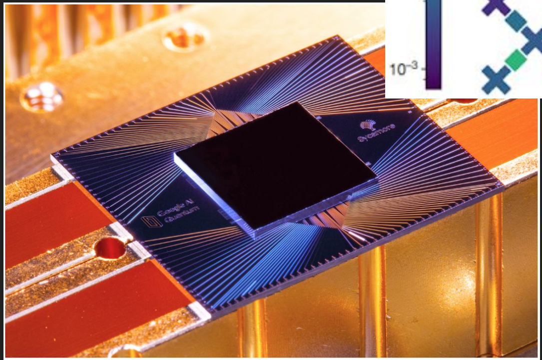


Image courtesy of:
https://www.dwavesys.com/sites/default/files/14-1047A-A_Practical_Quantum_Computing_An_Update_0.pdf

Circuits/Gates

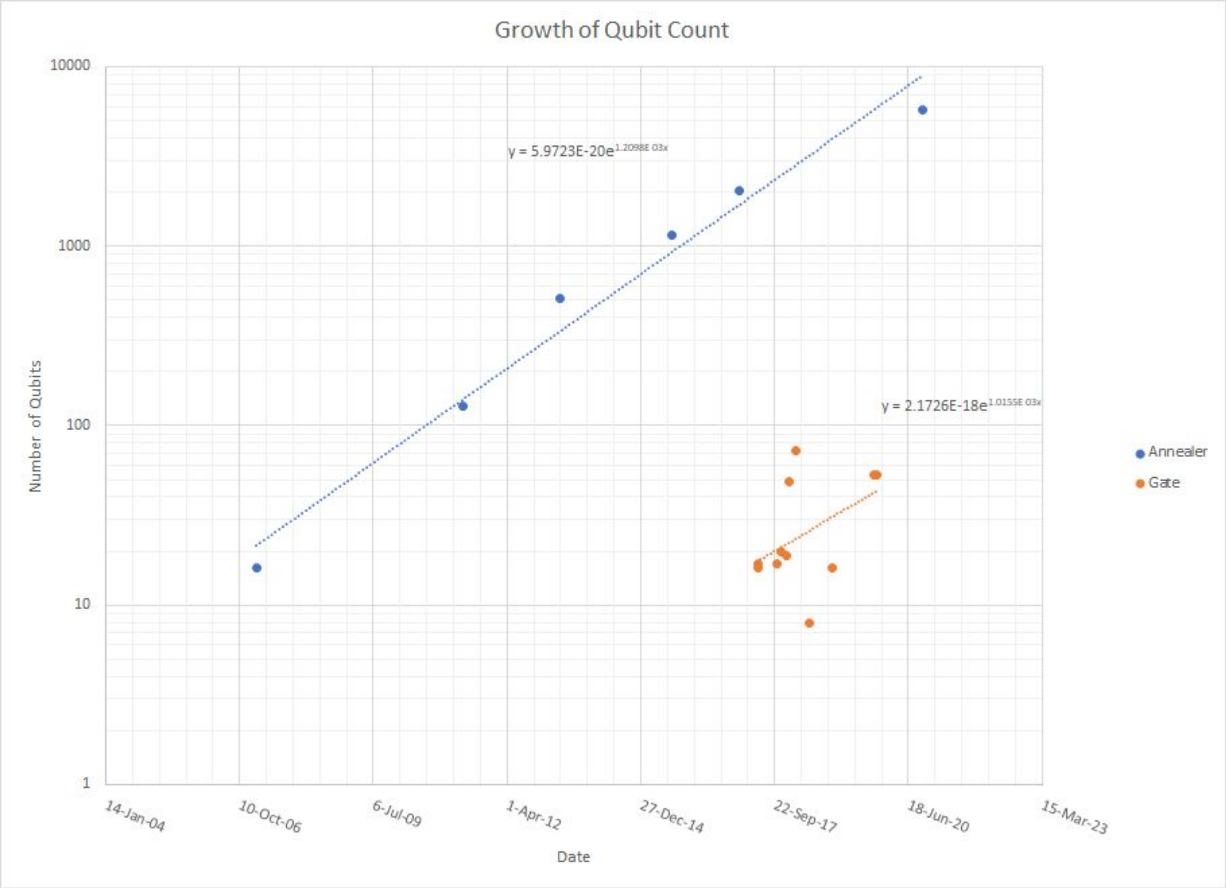
- 53 qubits*
- Universal
- 86 couplings between qubits





Images courtesy of:
https://ai.googleblog.com/2019/10/quantum-supremacy-using-programmable.html

Quantum Computers

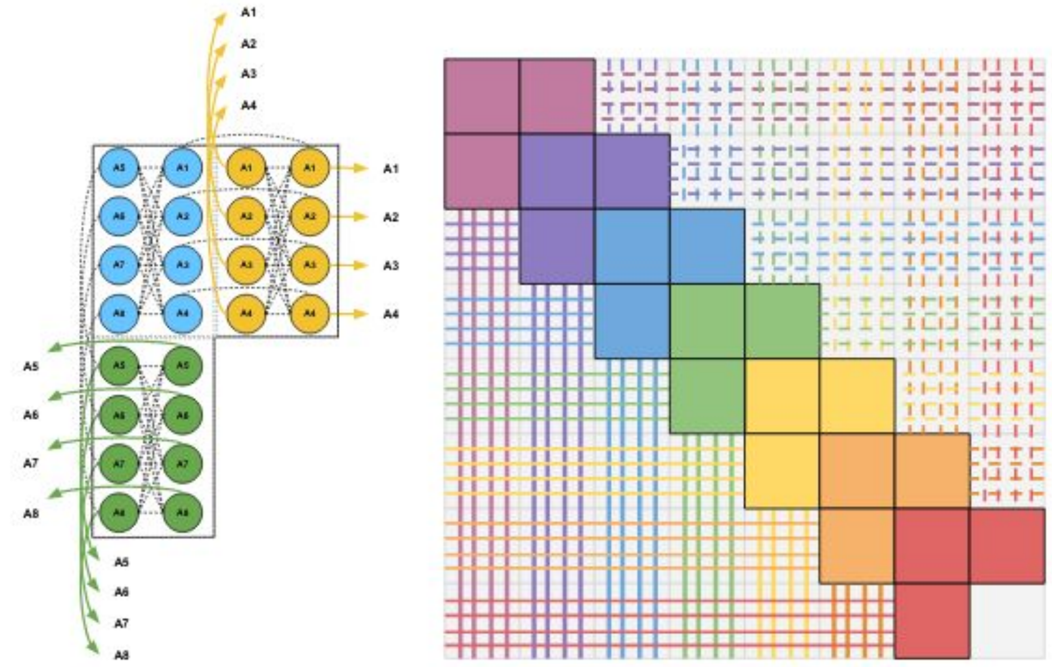


Cartesian Product of Completely Connected Graphs Deterministic Embedding

Table of Contents

Minor Embedding a Cartesian Product of Fully Connected Graphs

- https://arxiv.org/pdf/1602.04274.pdf

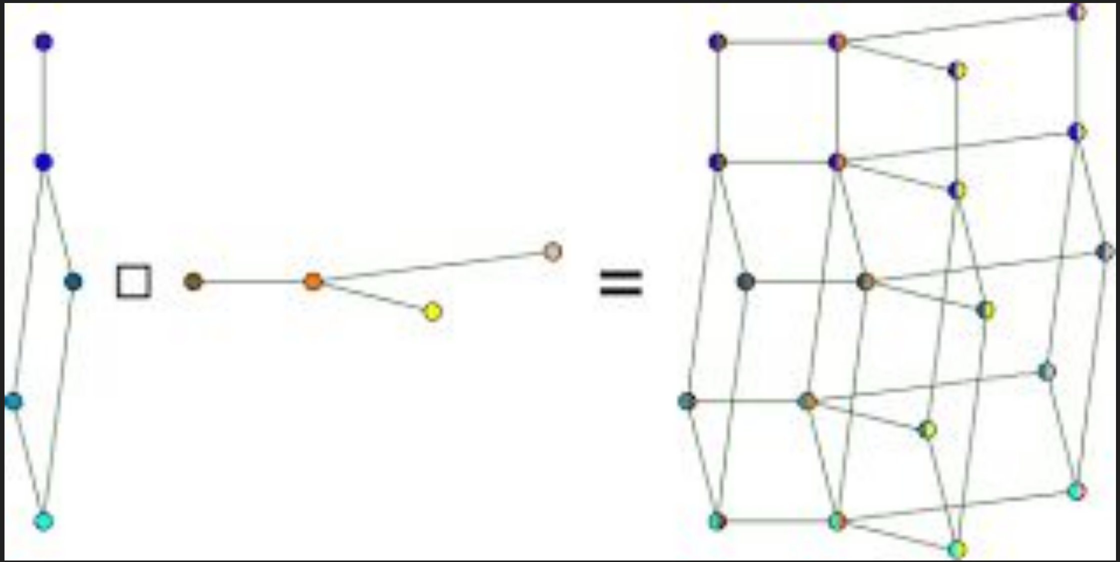


(a) A possible nexus choice for a K_8 graph

(b) An embedding for $K_8 \square K_7$

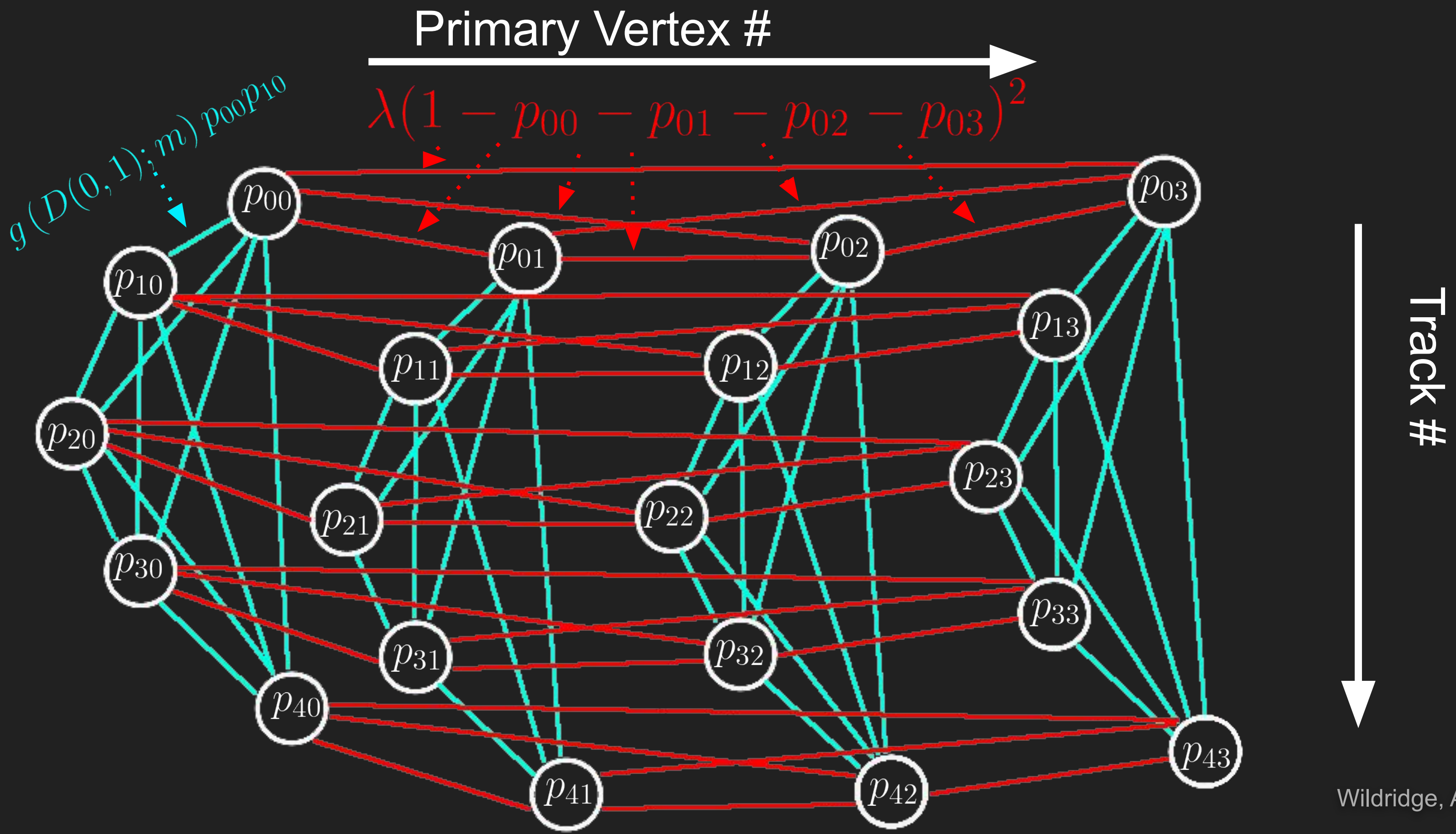
Cartesian Product

- □ Denotes Cartesian product of graphs



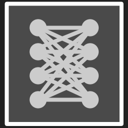
Example of a Cartesian product between two graphs

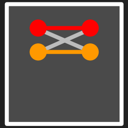
Example: 4 Vertices 5 Tracks



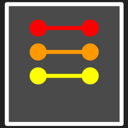
“Nexus” Embedding

Unit Cell

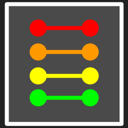




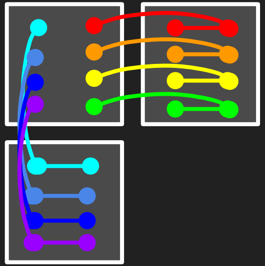
K_2



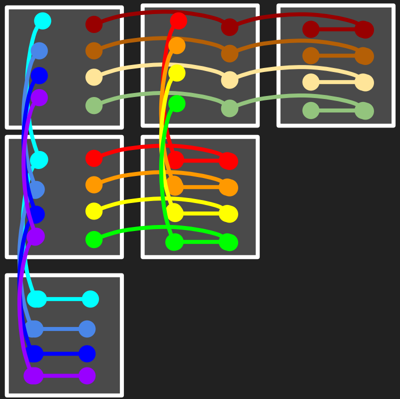
K_3



K_4



K_8

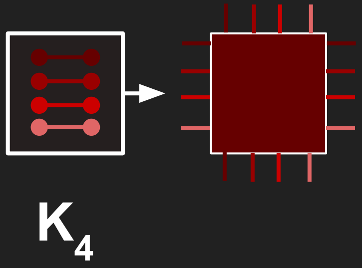


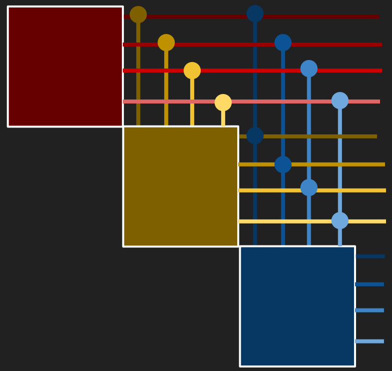
K_{12}

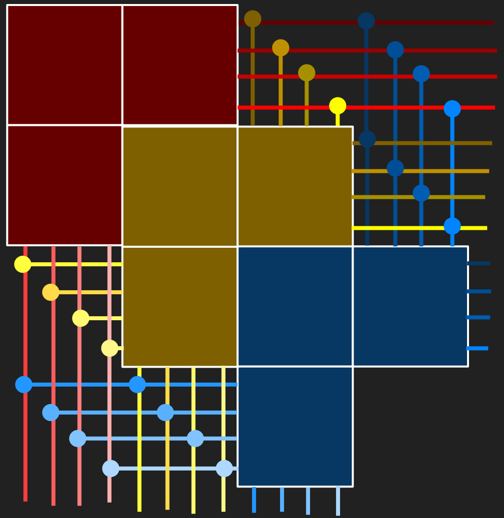
“Nexus” Embedding

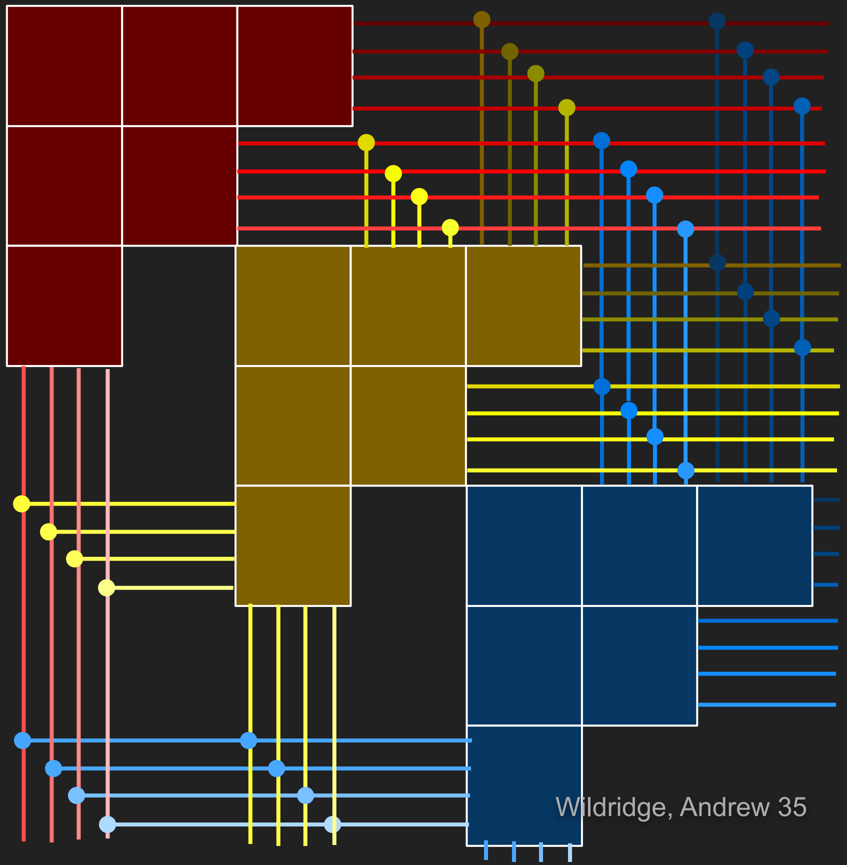
- Free to choose either K_{n_T} or K_{n_V} as nexus size
    - If nexus is K_{n_T}, then you repeat nexus n_V times
    - Vice versa for n_T
- I do both and pick one that requires the smallest Chimera

"Bus" Connections - Connections between Nexuses

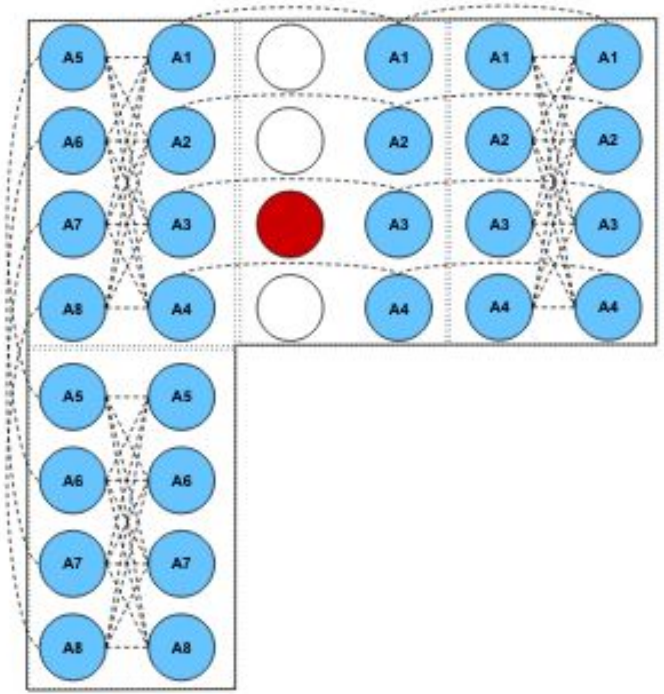




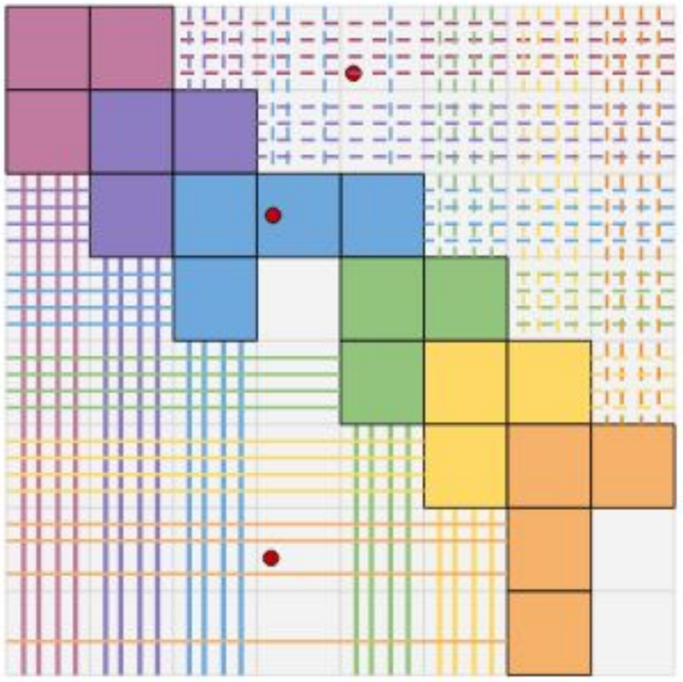




Minor Embedding a Cartesian Product of Fully Connected Graphs

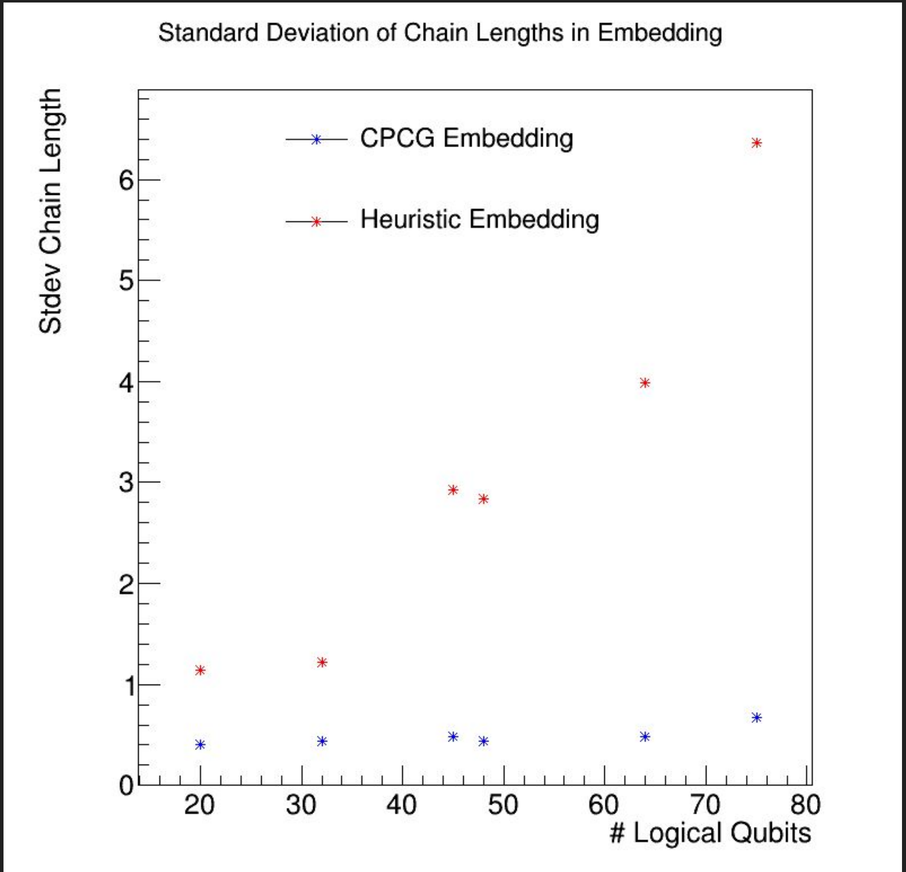


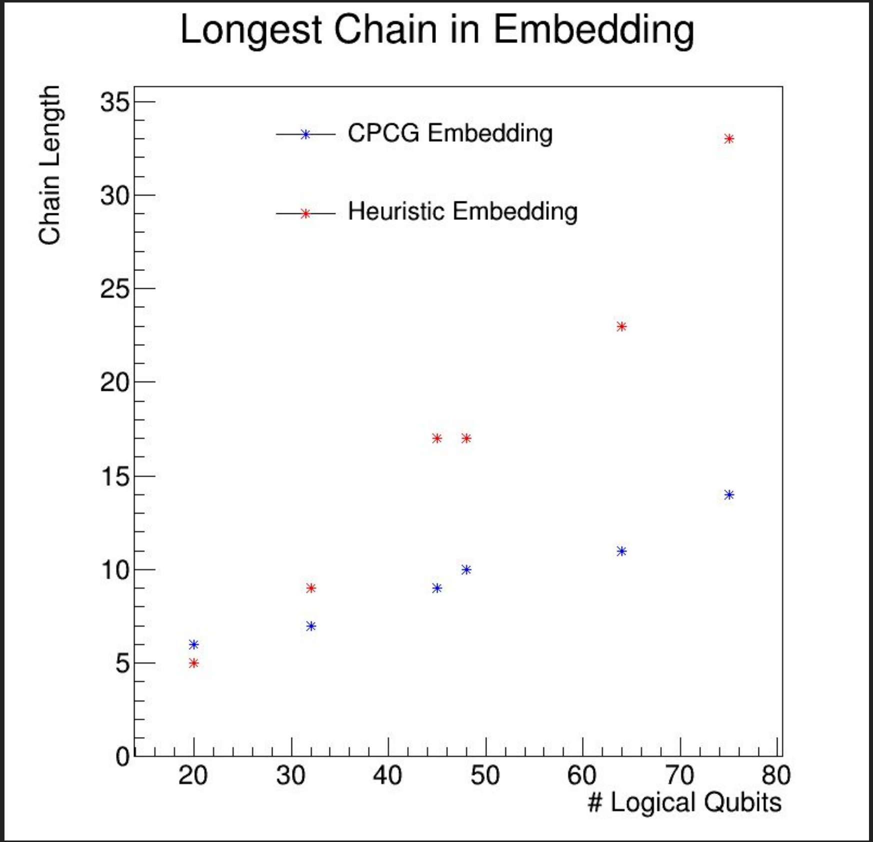
(a) A possible K_8 nexus extension with an inoperable qubit



(b) A K_8 nexus on a chip with inoperable qubits

CPCG Embedding vs Heuristic Embedding





Optimization Breakdown

Breakdown of Each Optimization

3V 15T Improvements		
QPU + Improvement	Convergence Efficiency (%)	
	Dunn Index = 0.02418	Dunn Index = 0.00290
DW_2000Q_2_1	0.6615 ± 0.0047	0.0800 ± 0.0027
DW_2000Q_6	0.6774 ± 0.0047	0.3010 ± 0.0046
DW_2000Q_6 + CPCG Embedding	0.7664 ± 0.0042	0.4618 ± 0.0050
DW_2000Q_6 + CPCG Embedding + Chain Strength	0.9063 ± 0.0029	0.6515 ± 0.0048

Information on Dataset

Table of Contents

Data

- Artificial events generated from known CMS event distributions
- Multiple event topologies are explored
- https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPerformance2017MC#Expected_resolutions_on_track_pa

Simulated Annealing

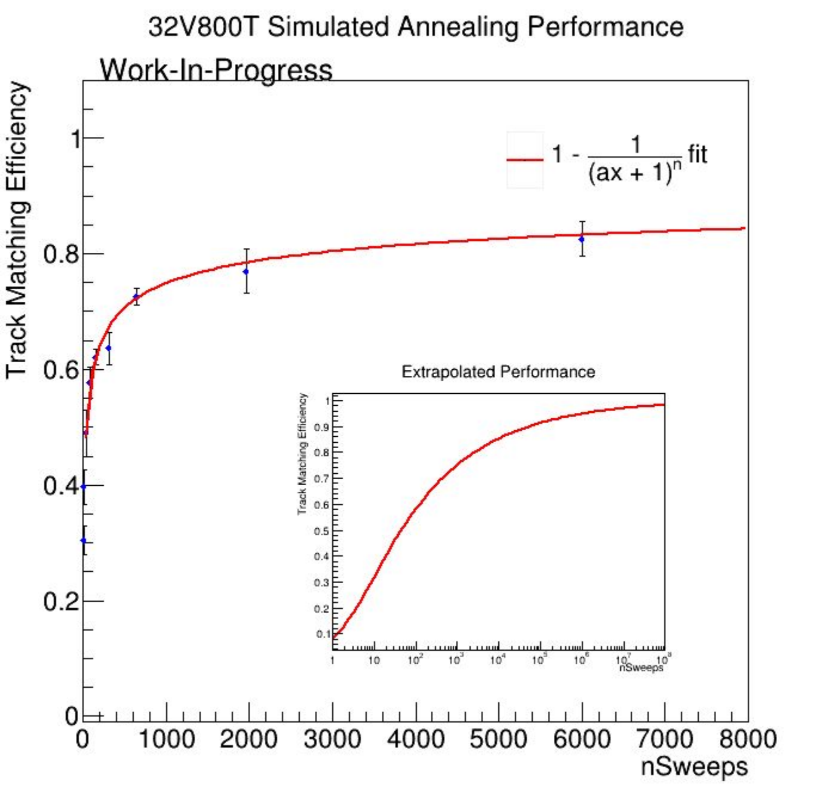
Table of Contents

Overview

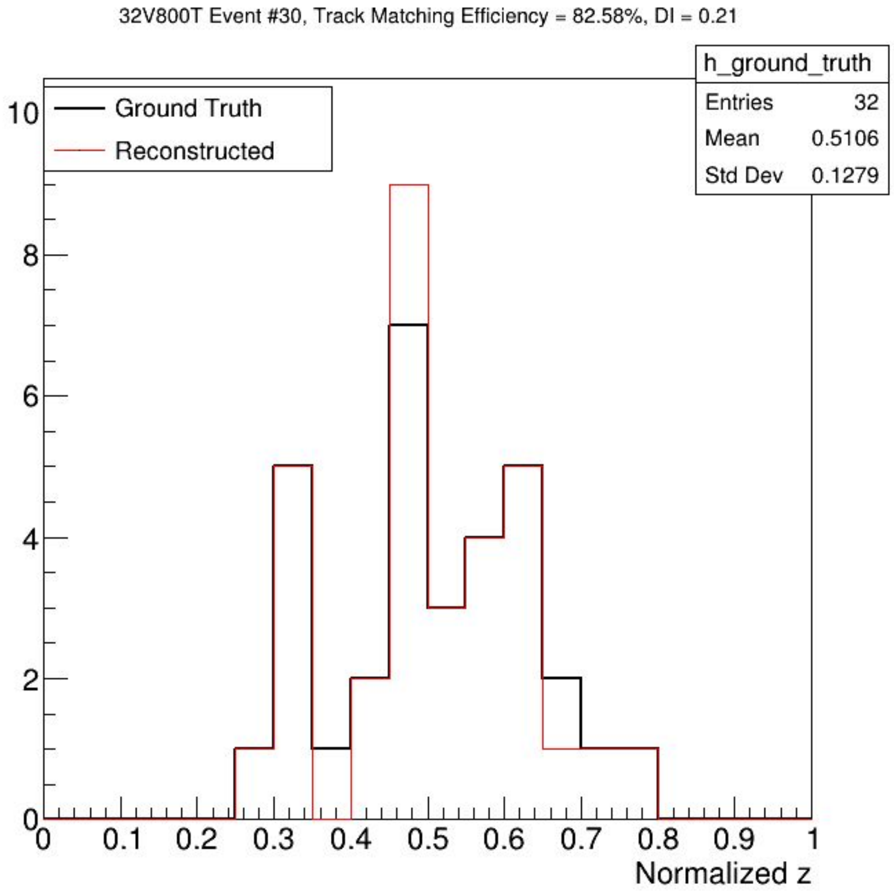
- Provided QUBO used by Quantum Annealing algorithm
- Uses algorithm based on Metropolis algorithm to flip bits
- Sweeps over all bits and flips with random probability based on energy difference of flip
- Repeats nSweeps times

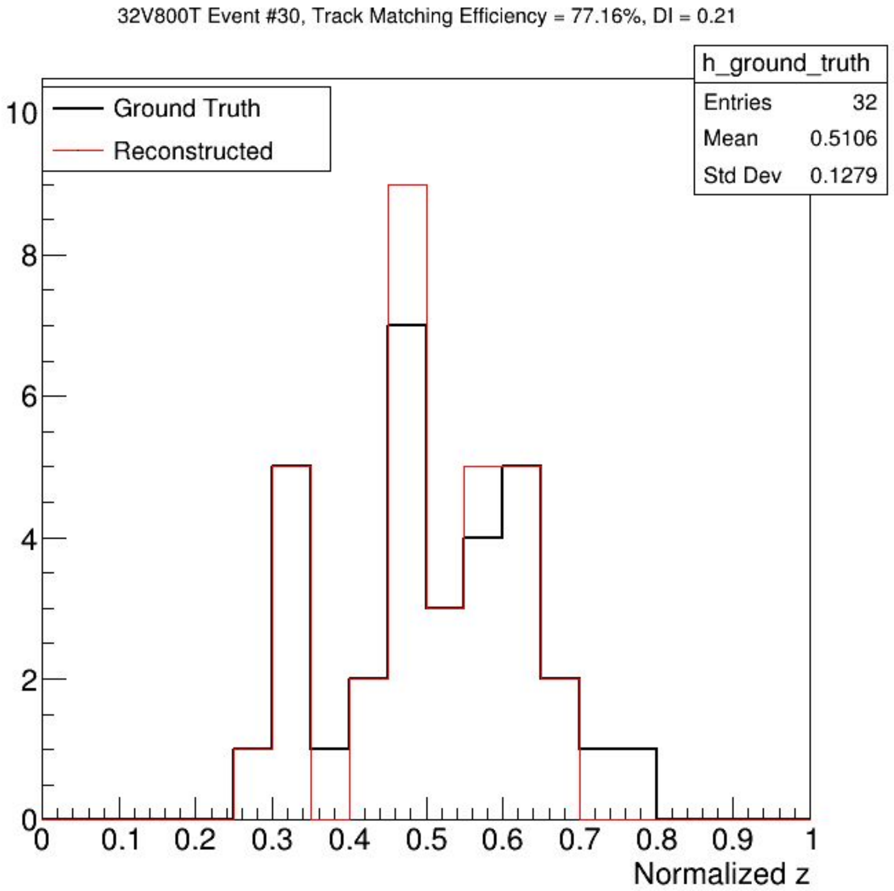
Simulated Annealing at LHC Scale Event Topologies

- Trying to figure out why performance plateaus at ~80% and plateaus very fast
- This is highest dunn index (DI) for 32V800T (realistic LHC conditions), DI = 0.21
  - Reminder: higher DI → easier to cluster

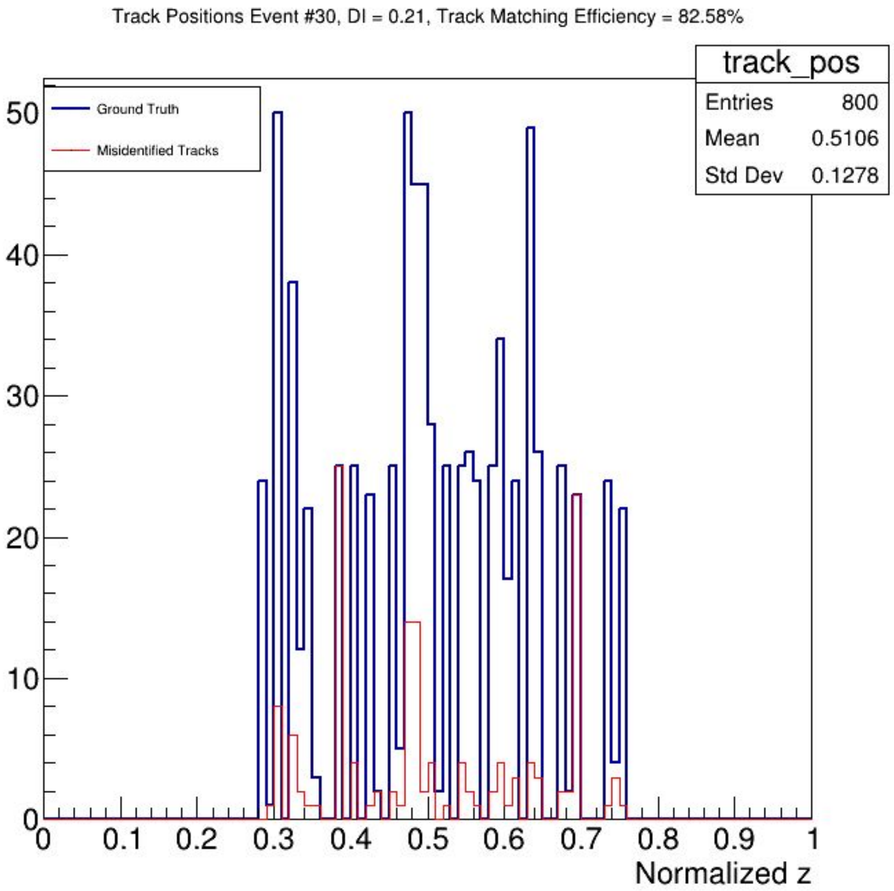


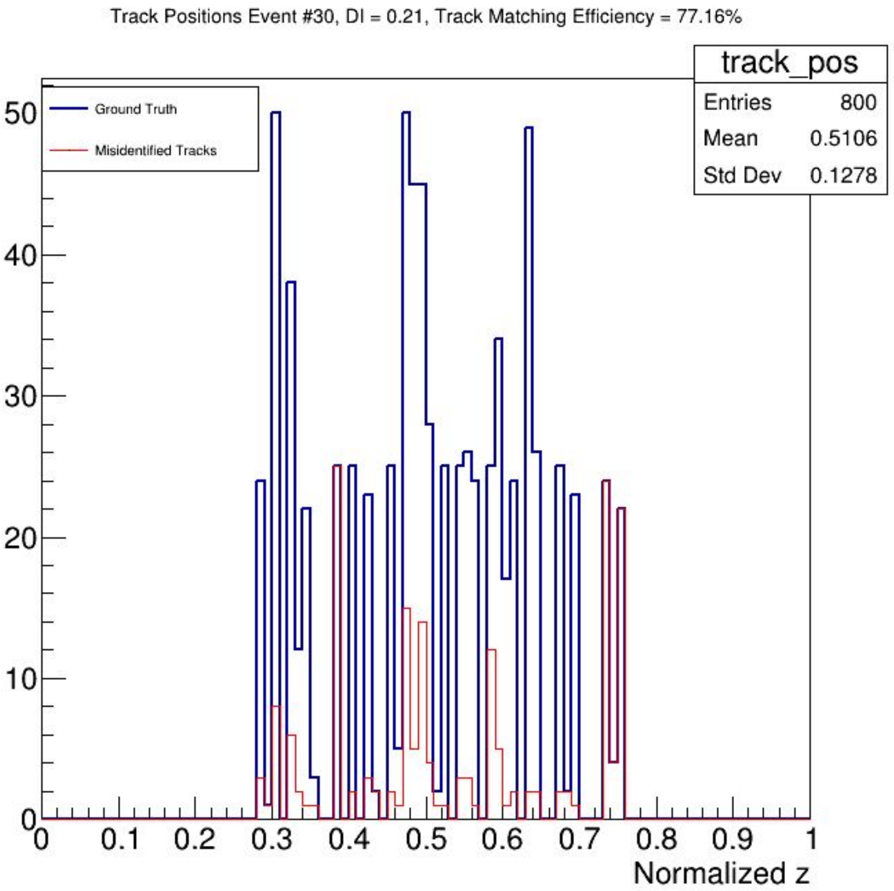
Reco vs Ground Truth MC





Misidentified Track Positions





Is D-Wave Quantum?

Table of Contents

Is D-Wave Quantum?

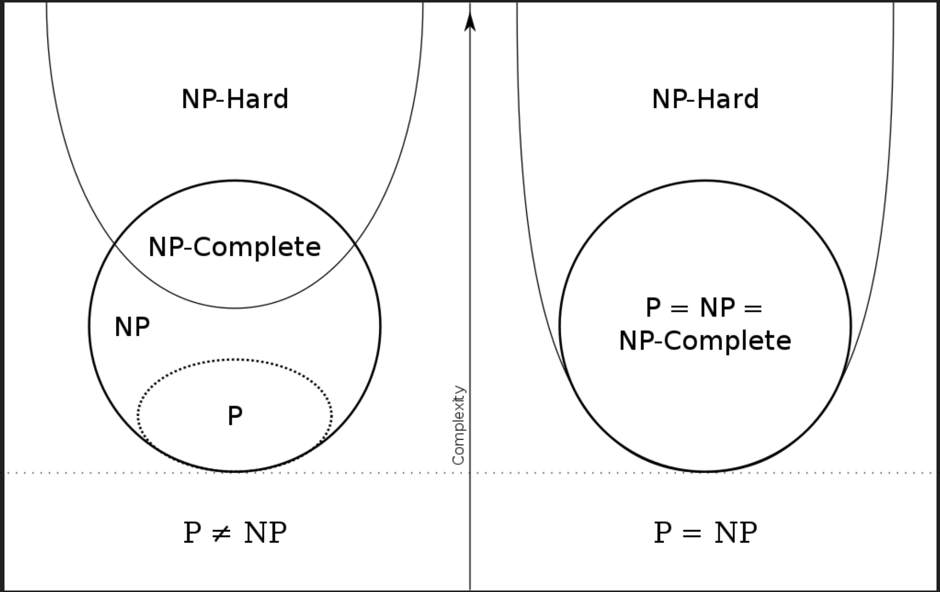
- **Entanglement in a Quantum Annealing Processor, T. Lanting *et al.*** DOI: 10.1103/PhysRevX.4.021041
  - Showed quantum entanglement and coherence existed for 2 qubit and 8 qubit systems
- **Quantum annealing with manufactured spins, M. W. Johnson et al.** *Nature* **volume 473**, pages 194–198 (12 May 2011)
  - Showed quantum annealing performs better than thermal annealing
  - Has a temperature dependence that is quantum

P vs. NP vs. NP-complete

Table of Contents

P vs NP vs NP-hard vs NP-complete

- **P** - can be solved and verified in polynomial time
- **NP** - can be verified in polynomial time
- **NP-Hard** - is “harder” than any other NP problem. “Hard” to solve, “hard” to check (for now)
- **NP-Complete** - is “harder” than any other NP problems and is in NP



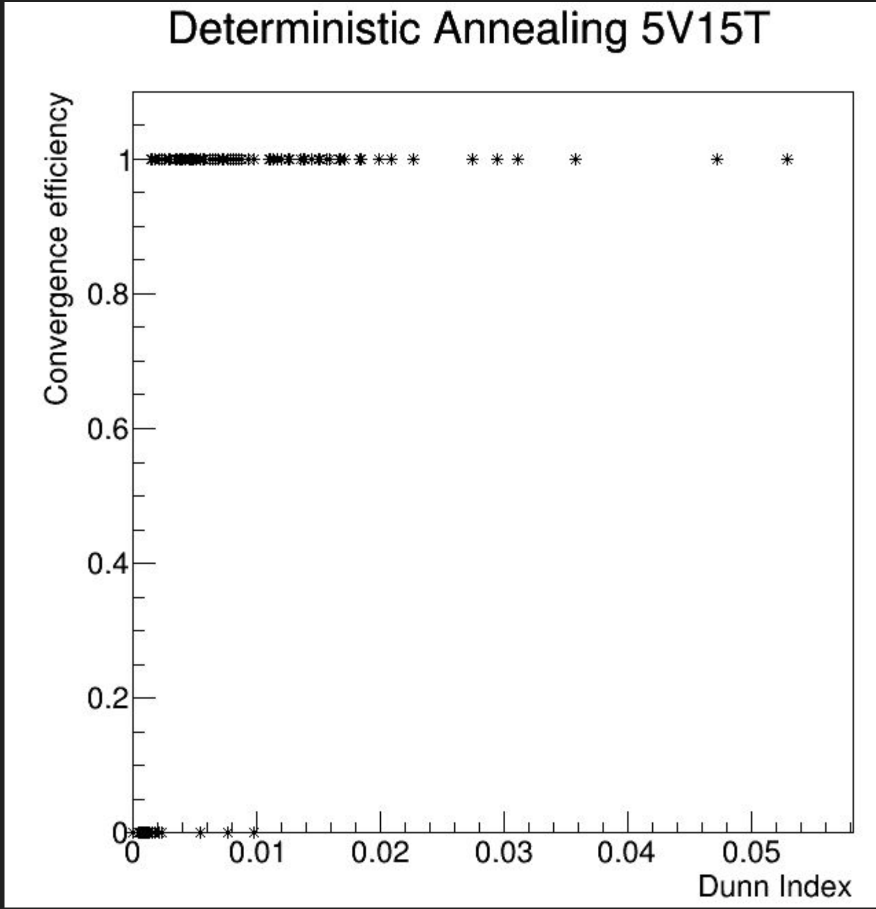
Euler diagram for P, NP, NP-Complete, NP-Hard [wikipedia]

Deterministic Annealing

Table of Contents

Deterministic Annealing - 5V15T

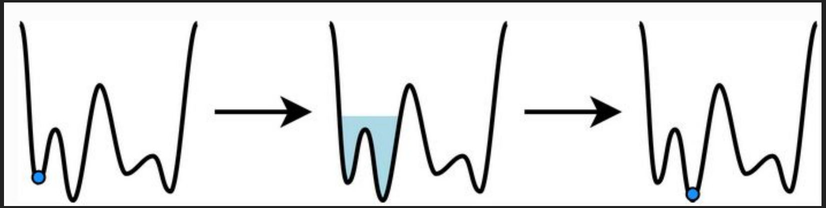
- Step function because it is deterministic
  - No need for sampling
  - You either get it or you don't
- Ran with default values given by CMSSW



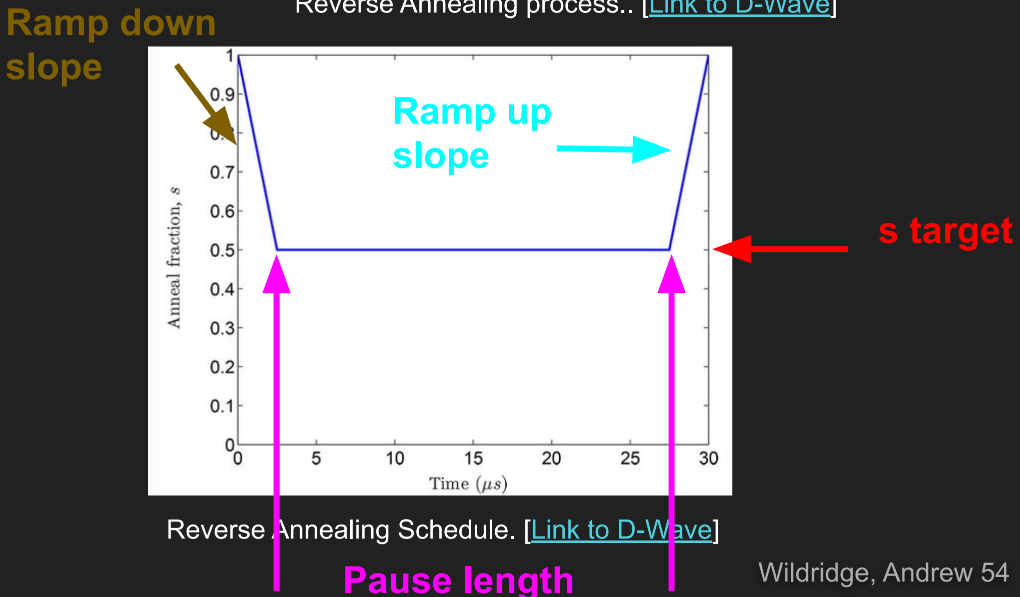
Reverse Annealing

Reverse Annealing - Overview

- Performs a local search to try and find a better solution
- Must provide an initial state that is classical
- You weaken the strength of the problem Hamiltonian a little and increase the transverse Hamiltonian to try and “scramble” or “shake” the solution a bit to go into a new minima
- You hope this new minima is your global minimum
- Can be performed iteratively/”back-to-back”, end state from 1st reverse anneal is beginning state for 2nd reverse anneal etc.



Reverse Annealing process.. [Link to D-Wave]



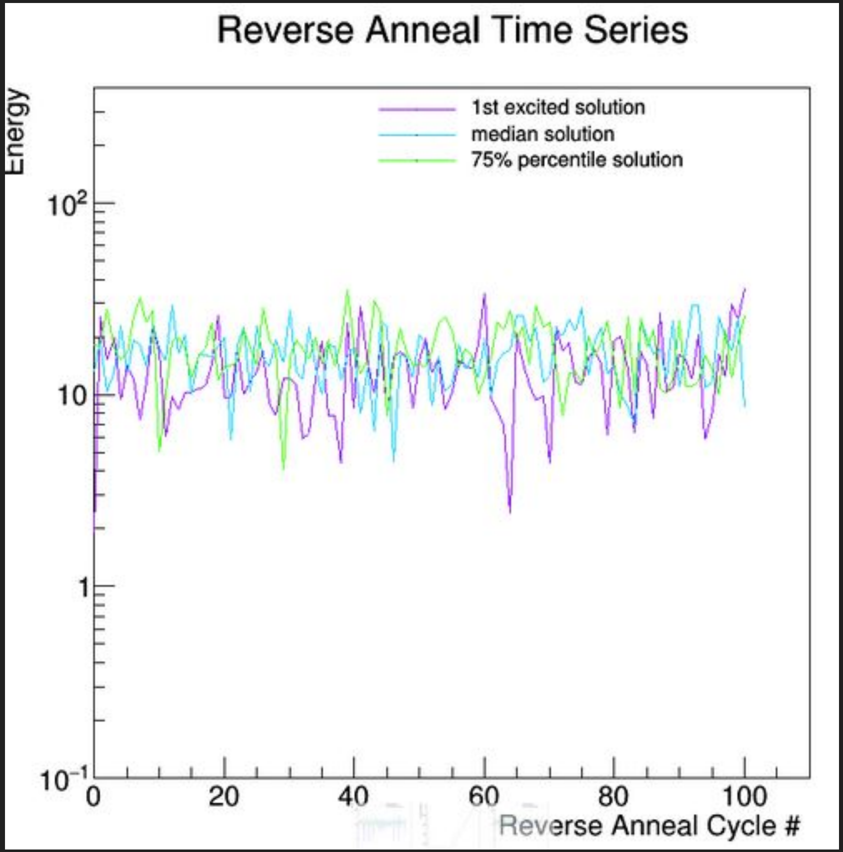
Reverse Annealing Schedule. [Link to D-Wave]

Reverse Annealing - Game Plan

- Look at moderately hard event for 5V15T
  - Event #4, Dunn Index = 0.0126
  - November Convergence Efficiency = 3.4%; Asymptotic Efficiency = 7.55%
- Reverse anneal 3 classical solutions returned from November, back-to-back 100 times; ground state = 0.495 energy
  - 1st excited = 1.885 energy
  - median = 13.164 energy
  - 75th percentile = 18.243 energy
- Perform a grid search on parameters for reverse anneal schedule
  - s target: 0.1, 0.2, 0.3, 0.4, 0.5
  - Pause length: 10, 40, 80

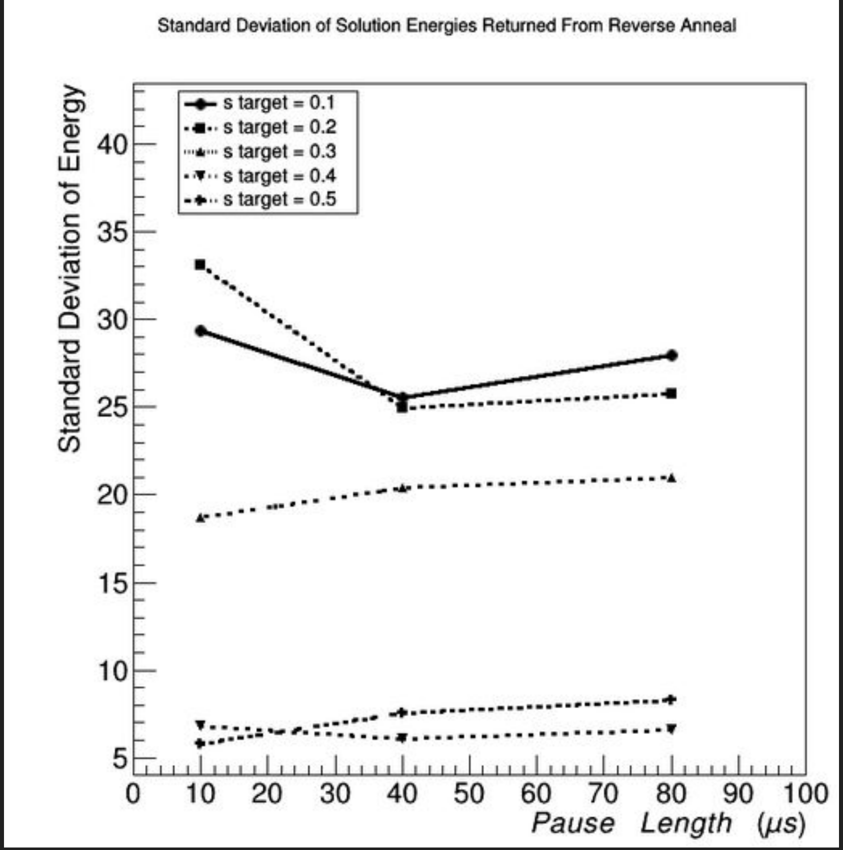
Reverse Annealing - Example

- This is what we see
- Going to try and attempt to characterize these time series and identify any trends
  - Standard deviation
  - Initial energy - mean energy
  - Slope as a function of anneal cycles



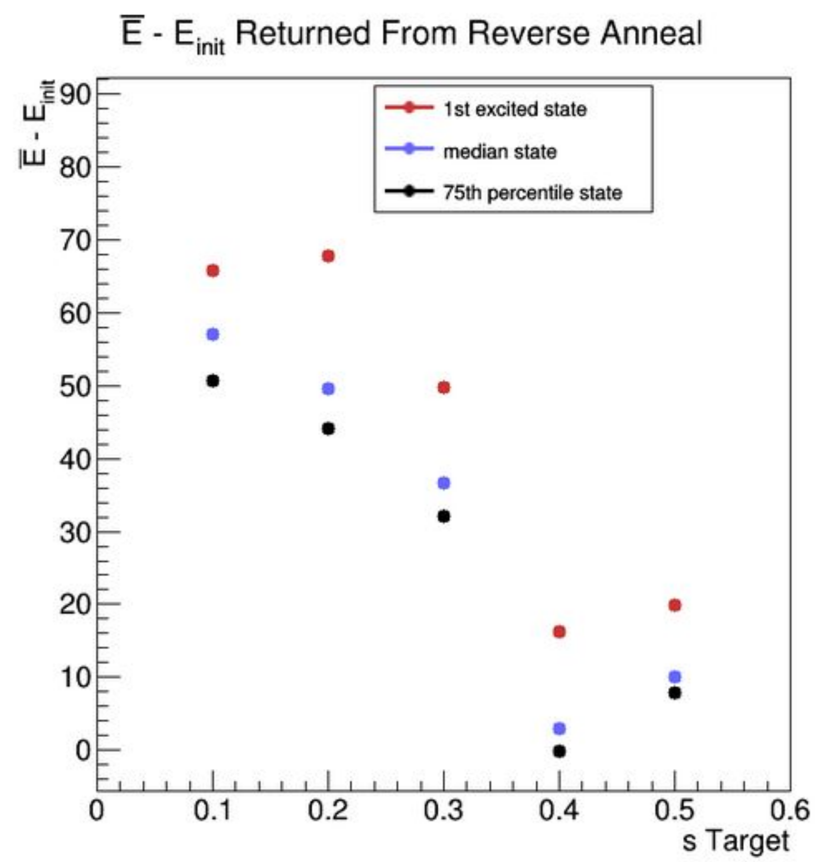
Reverse Annealing - Standard Deviation

- Averaged over initial states
- The standard deviation of the time series is dominated by s target
- Smaller s target results in a much broader search
- Large jump in broadness of search between s target = 0.3 to s target = 0.4
- Pause length seems to have a higher order effect on standard deviation



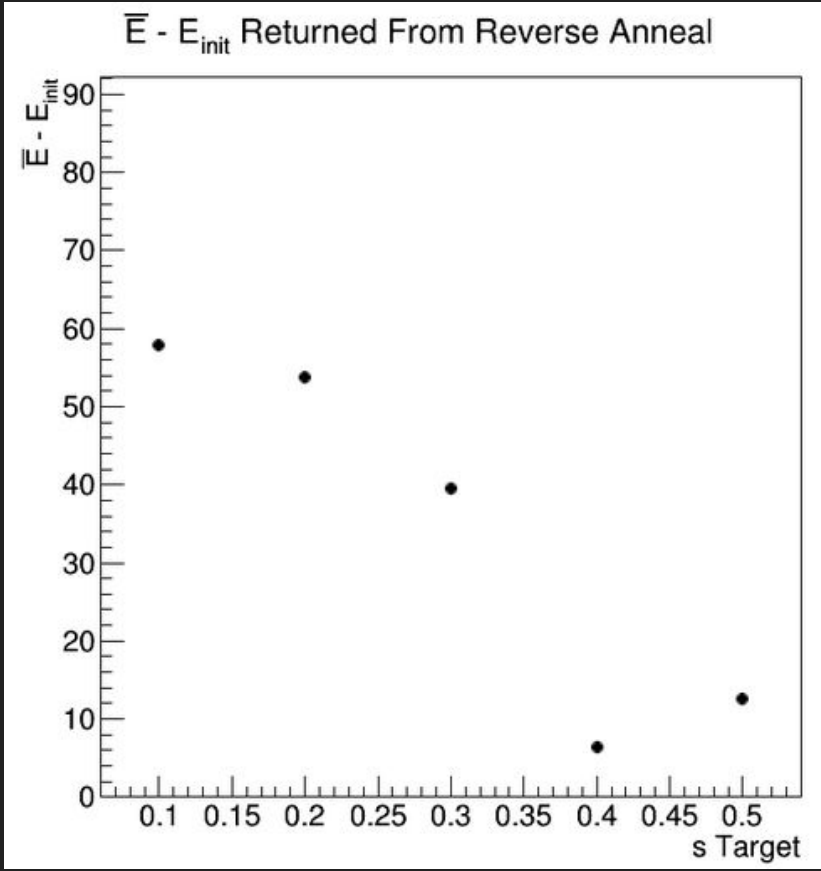
Reverse Annealing: <E> - E_init

- Averaged over pause length



Reverse Annealing: <E> - E_init

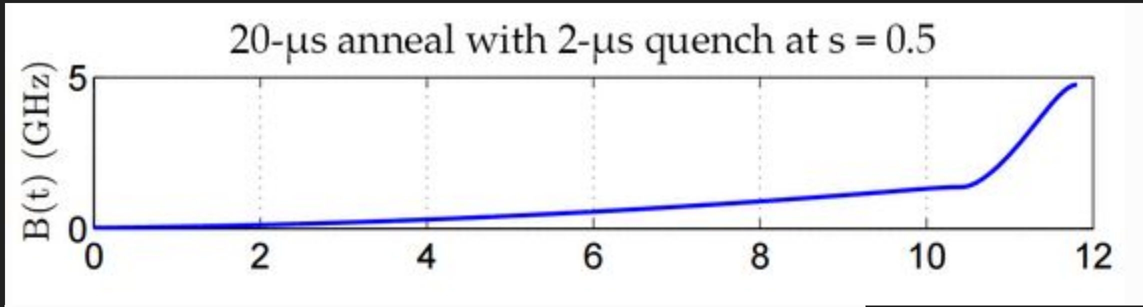
- Averaged over different states and pause length

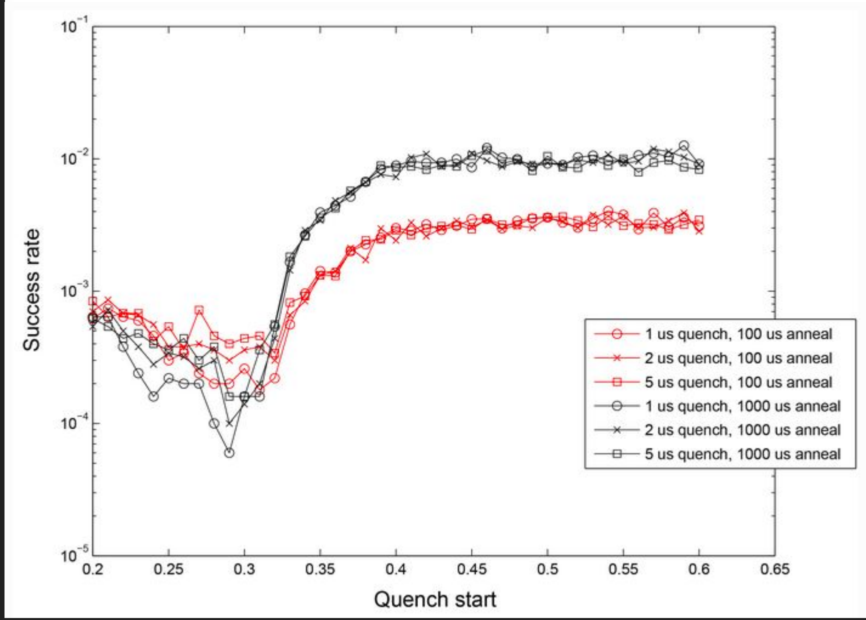


Anneal Pauses

Quench Scan - Overview

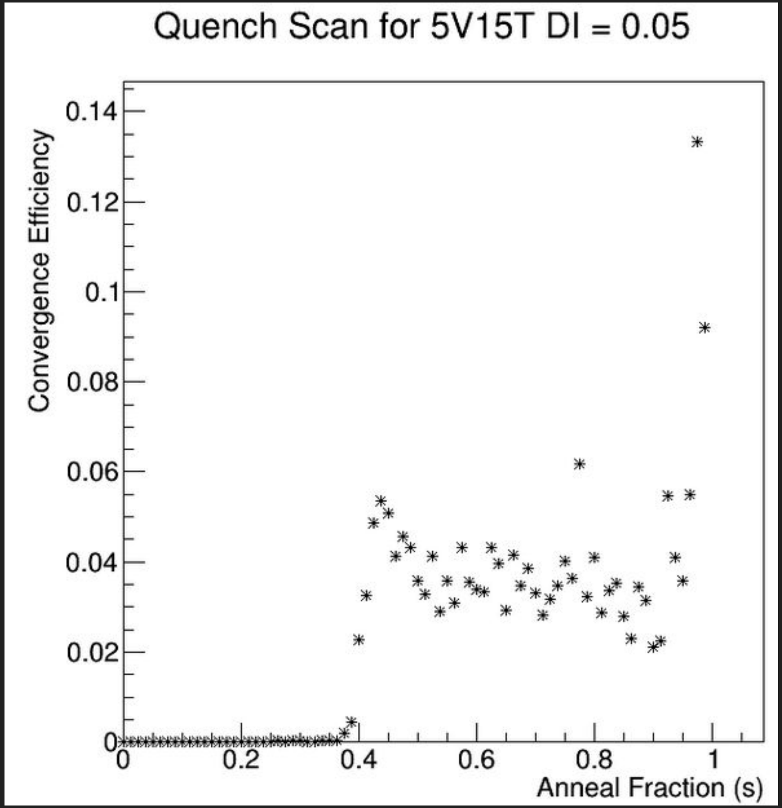
- Quenches act as a method to peer into the dynamics of your system during the anneal
- By abruptly turning the anneal fraction to 1.0, we “freeze” the system where it was and can look at it classically
- There is a point during the anneal in which the ground state solution is returned much more frequently





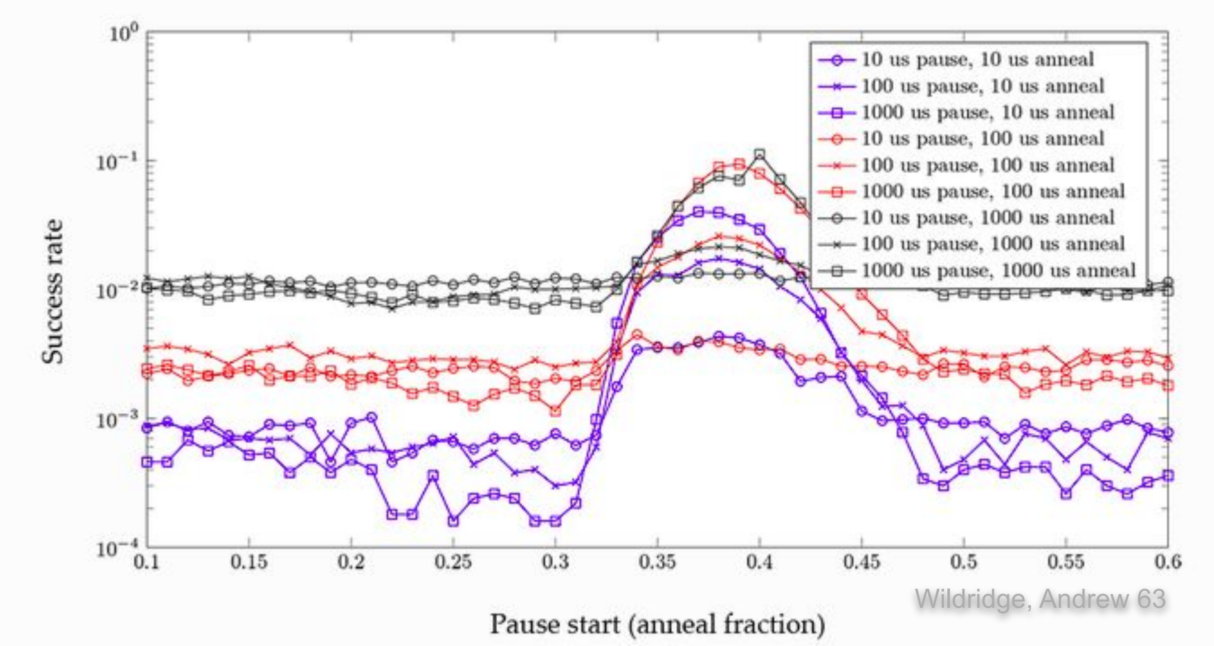
Quench Scan

- Event #23, high dunn index
- Important point in dynamics seems to be in the range s = [0.35, 0.55]
- Much larger scatter than D-Wave's example
- Second change in dynamics around range [0.9, 1.0]?



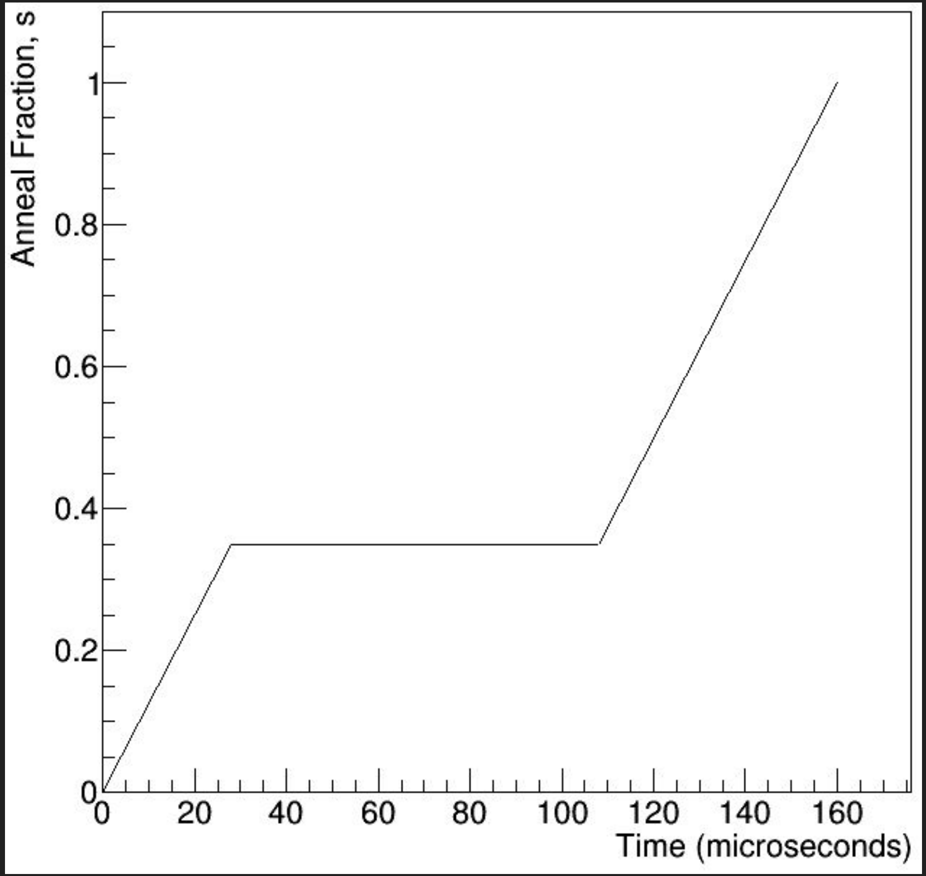
Anneal Pauses - Overview

- Much more efficient for shorter anneal schedules
  - 100 us pause + 10 us anneal is about same level of performance as 100 us pause + 100 us anneal and 100 us pause + 1000 us anneal
- Pause needs to be of comparable length to anneal, 10 us pause + 1000 us anneal saw no improvement
- Anneal time has diminishing returns, 1000 us pause + 100 us anneal is same performance as 1000 us pause + 1000 us anneal
- 100 us pause + 100 us anneal has better/same performance as 100 us pause + 1000 us anneal



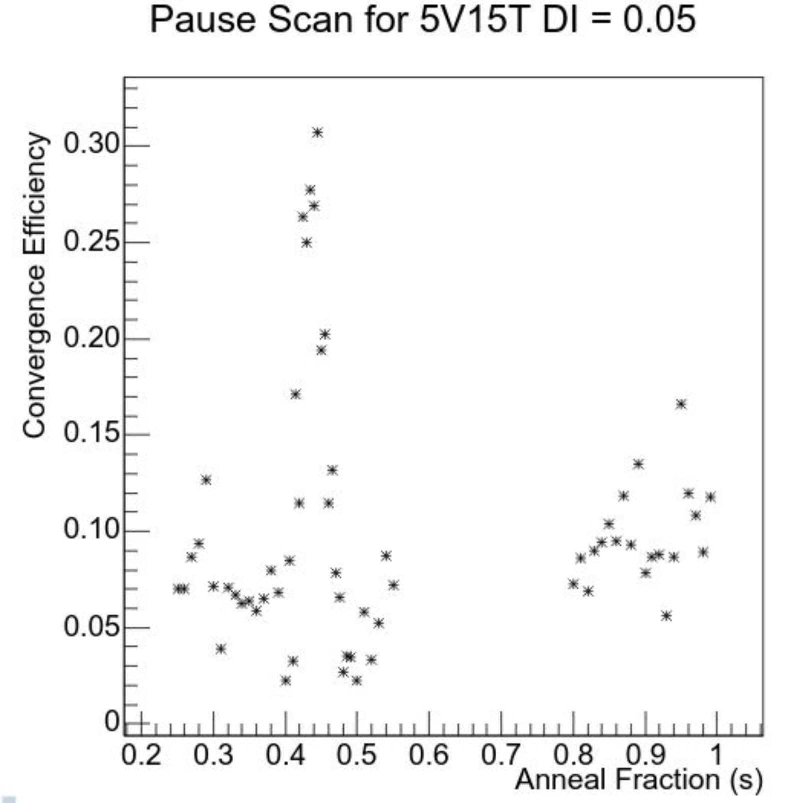
Anneal Pauses - Game Plan

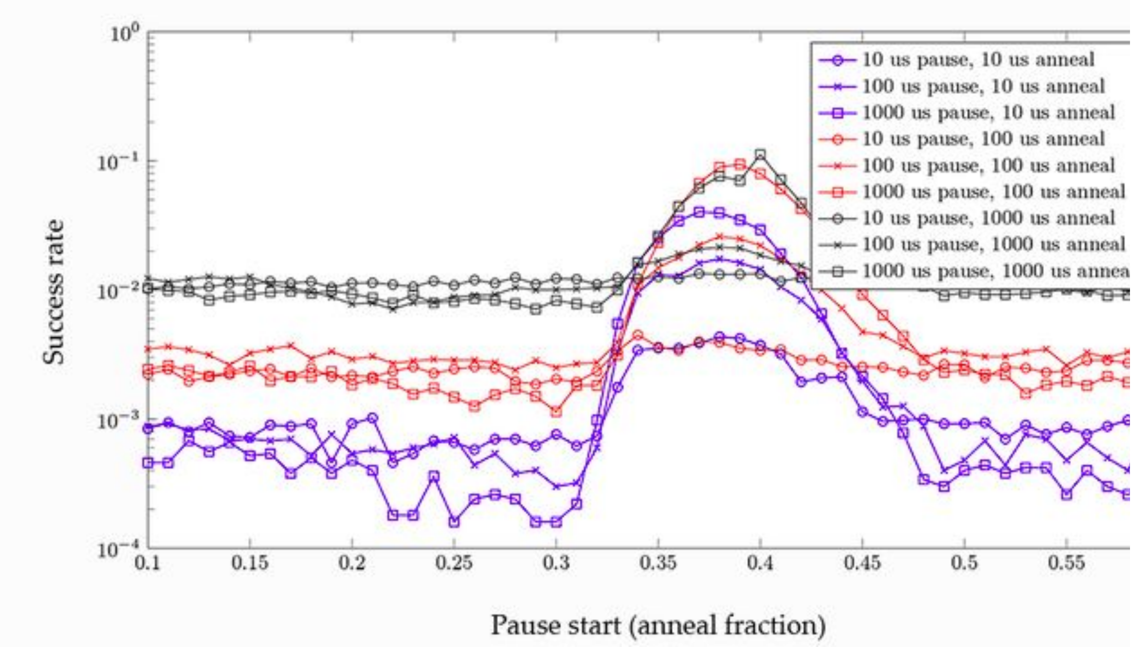
- 10k num_reads per pause point
- Delta pause point = 0.01 anneal fraction
- Try 20 different pause points around interesting region
- 80 microsecond forward anneal with a 80 microsecond pause for a total 160 microsecond anneal
- On right is pausing at anneal fraction = 0.35, 0.35*80=28

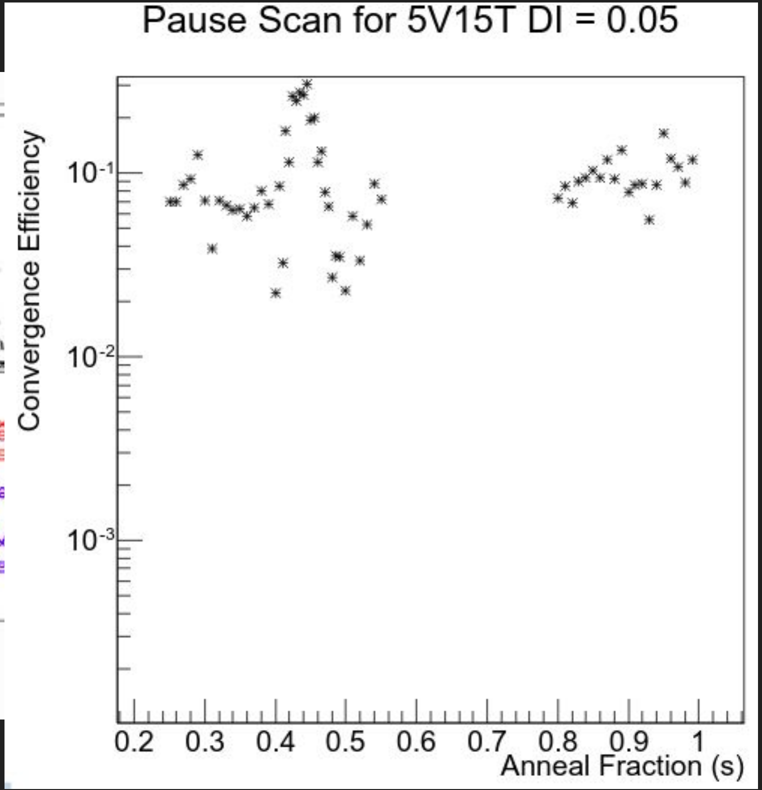


Anneal Pause

- Optimal pause point = 0.445
- Factor of 5x improvement
- No secondary pause point seen in range [0.8, 1.0]




Success rate
Pause start (anneal fraction)
10 us pause, 10 us anneal
100 us pause, 10 us anneal
1000 us pause, 10 us anneal
10 us pause, 100 us anneal
100 us pause, 100 us anneal
1000 us pause, 100 us anneal
10 us pause, 1000 us anneal
100 us pause, 1000 us anneal
1000 us pause, 1000 us annea


Pause Scan for 5V15T DI = 0.05
Convergence Efficiency
Anneal Fraction (s)

Optimizing Squeeze Strength

Table of Contents

Optimizing Squeeze Strength - Overview

- Distortion function is meant to space small distances out more so that these are less susceptible to coupler noise
  - These are the important distances
- Large distances we don't care about → They are all set to about 1.0
- These new distances can be used to calculate a new "effective" dunn index after distortion
- **Idea: Maximize effective dunn index to find optimal squeeze strength, m.**

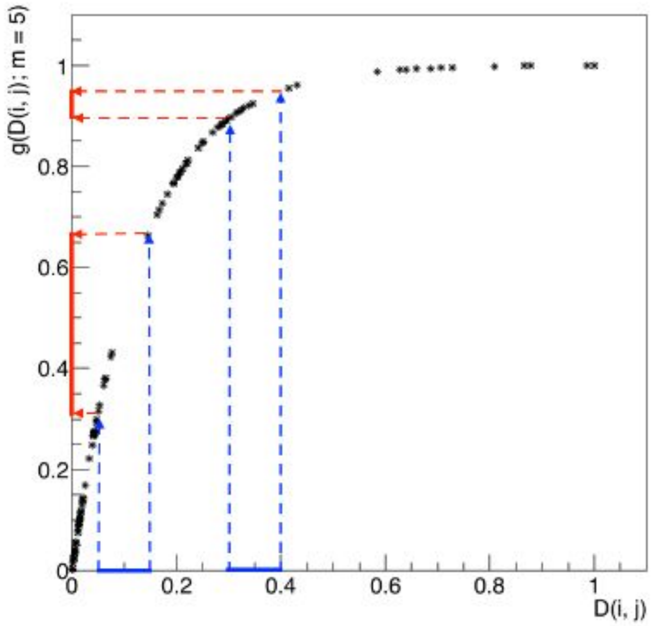
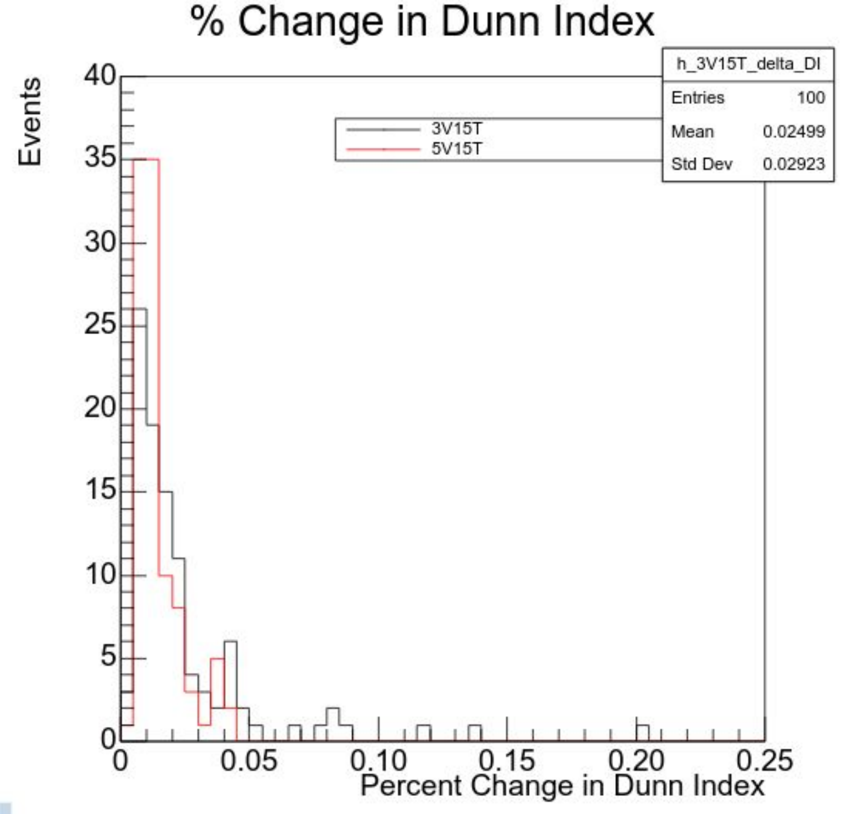


Figure 2: The distortion function $g(x; m = 5)$, defined in Eq. 6, applied to $D(i, j)$ defined in Eq. 3. This raises and increases spacing between small values of $D(i, j)$ to mitigate their sensitivity to coupler noise.

$$g(x; m) = 1 - e^{-mx}$$

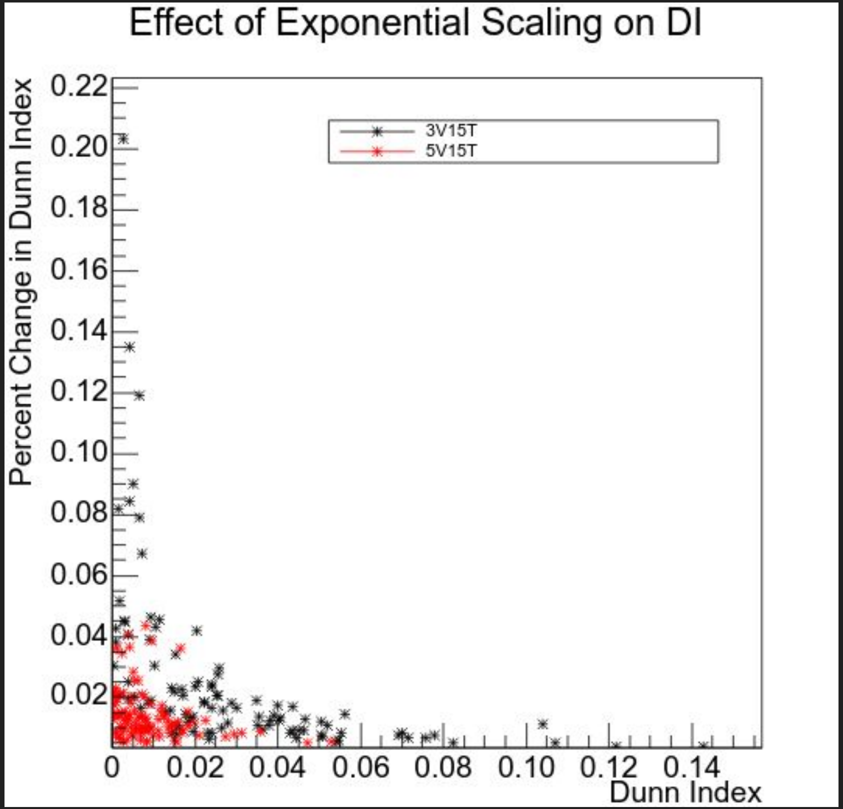
Optimizing Squeeze Strength - 3V15T vs. 5V15T

- m = 5 was found by studying 3V15T
- Comparing change in dunn index due to squeezing, 3V15T has a much larger relative change in dunn index
- More mass is shifted to higher dunn index as well



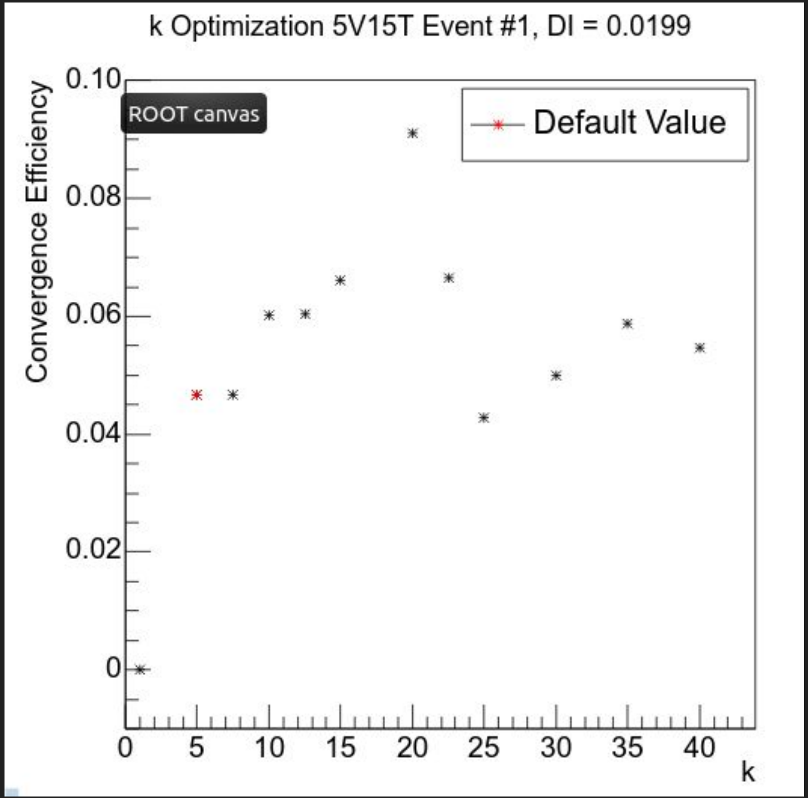
Optimizing Squeeze Strength - 3V15T vs. 5V15T

- Distortion function causes larger relative changes in dunn index for low dunn index events



Optimizing Squeeze Strength - 5V15T

- Did a grid search to optimize moderately hard event, event #1, dunn index = 0.0199
- Factor of 2x improvement finding a new optimal squeeze strength



Optimizing Squeeze Strength - 5V15T

- How to make this more robust?
  - Change distortion function to Lp norm and optimize p?
  - Create a new metric instead of effective dunn index