Applications of Quantum Annealing in Statistics

Joint Statistical Meetings 2019

arXiv:1904.06819

Los Alamos NATIONAL LABORATORY EST. 1943

Robert C. Foster, CCS-6

2 Aug. 2019

NNSA National Nuclear Security Administration

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

LA-UR-19-27143

LANL D-Wave 2X Quantum Computer

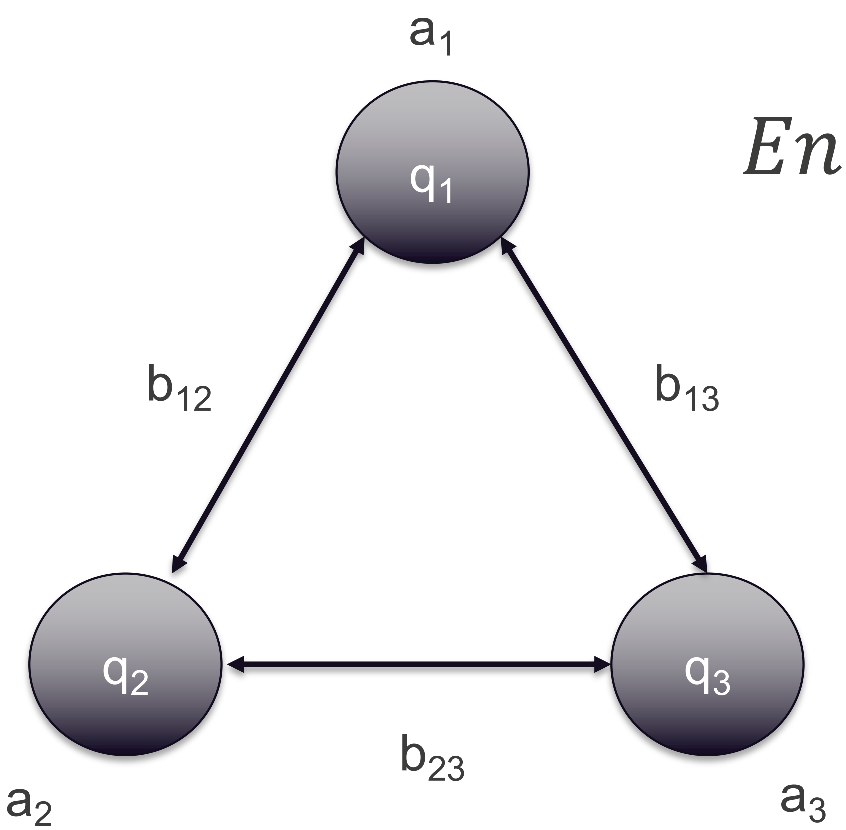
- The D-Wave is ***NOT*** a universal quantum computer

The D-Wave only performs energy minimization for Ising or QUBO models.

- Rather than operating on qubits directly, it sets up a problem and lets physics manipulate the qubits towards the solution.

What does the D-Wave do?

The D-Wave solves exactly one (1) type of problem…..



$$Energy = \sum_{i} a_i q_i + \sum_{j>i} b_{ij} q_i q_j$$

The q_i can be either…

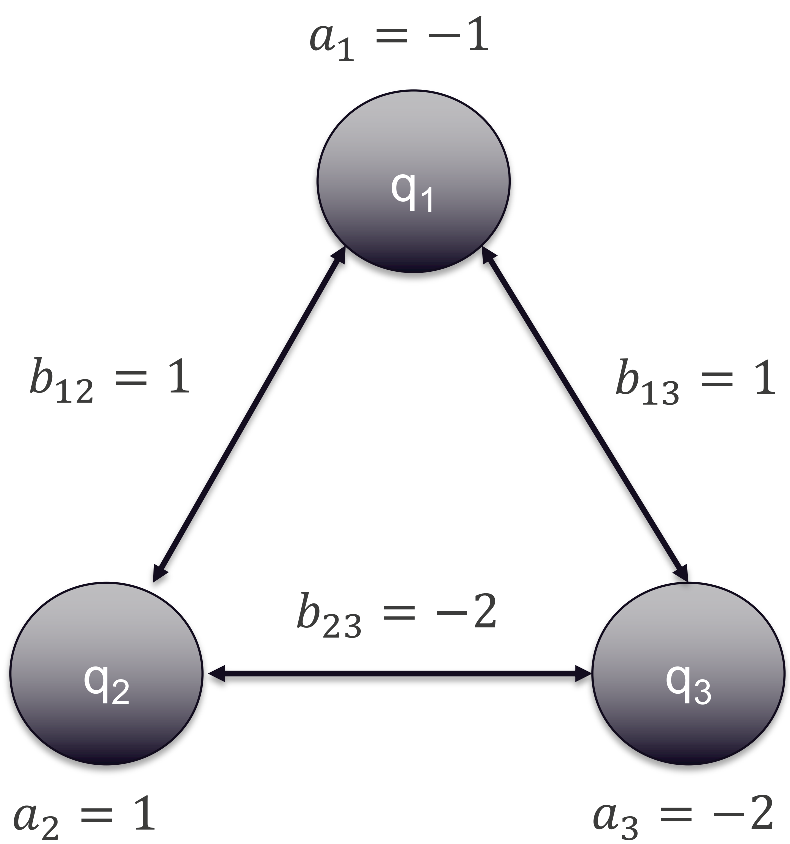
Native

1. Ising model: $q_i \in \{-1,1\}$
2. QUBO model: $q_i \in \{0,1\}$

Alternative

The D-Wave finds the set of q_i that minimize the energy of the system.

Example: QUBO Notation

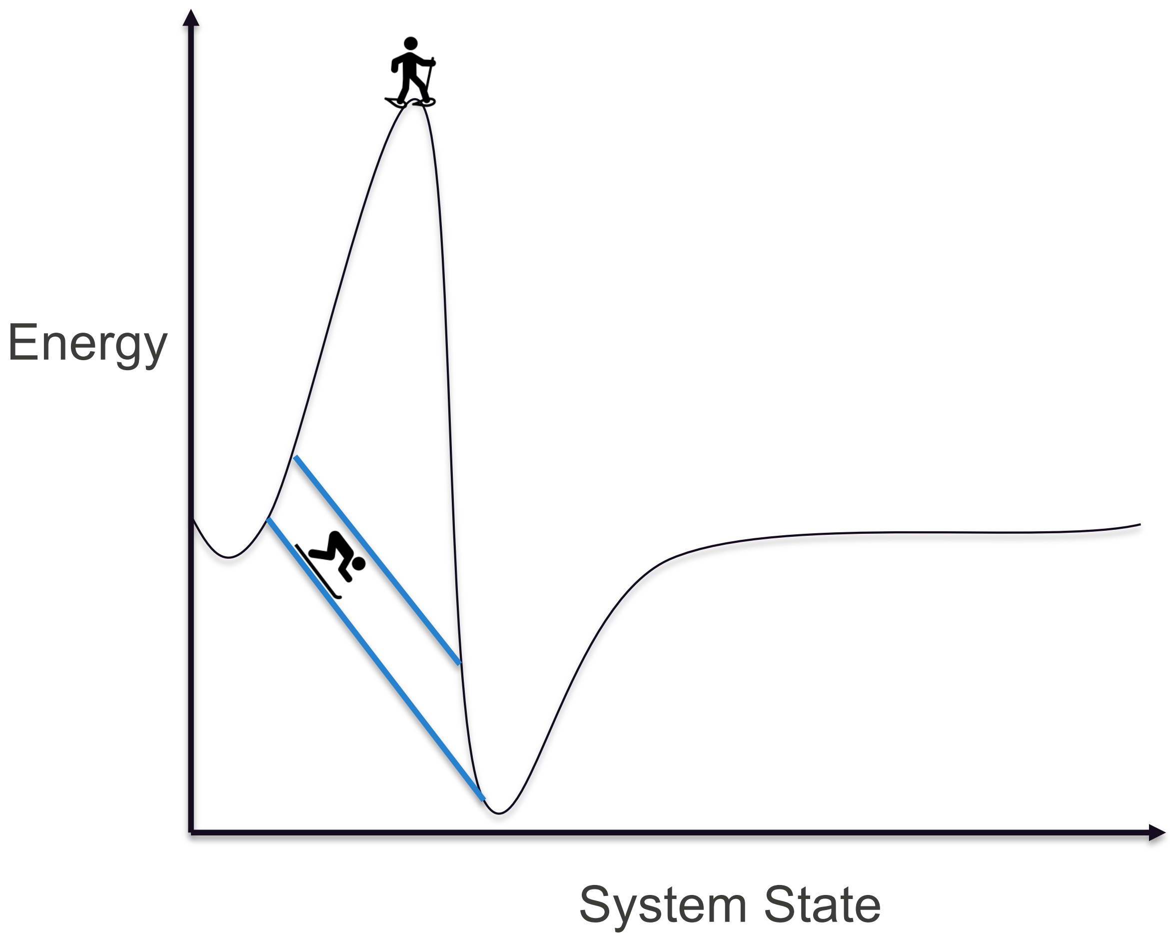


$$Energy = \sum_i a_i q_i + \sum_{j>i} b_{ij} q_i q_j$$

q_1	q_2	q_3	Energy
0	0	0	0
0	0	1	-2
0	1	0	1
0	1	1	-3
1	0	0	-1
1	0	1	-2
1	1	0	1
1	1	1	-2

Visual Metaphor for how the D-Wave Works

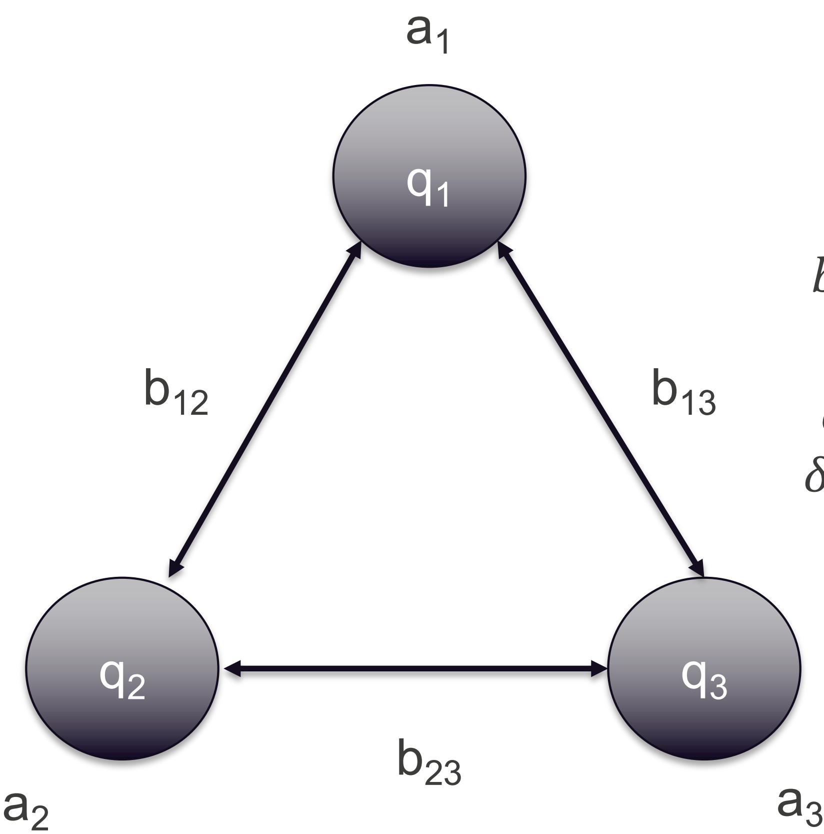
Simulated and Quantum Annealing



- Sammie A. Neal climbs the entire mountain just to get to the other side
- Quentin A. Neal tunnels directly through the mountain
- Where peaks are tall and thin, Quentin will outperform Sammy
- (It's actually incredibly hard to find real problems where Quentin beats Sammie)

Problem: Noise in the System

What you Input

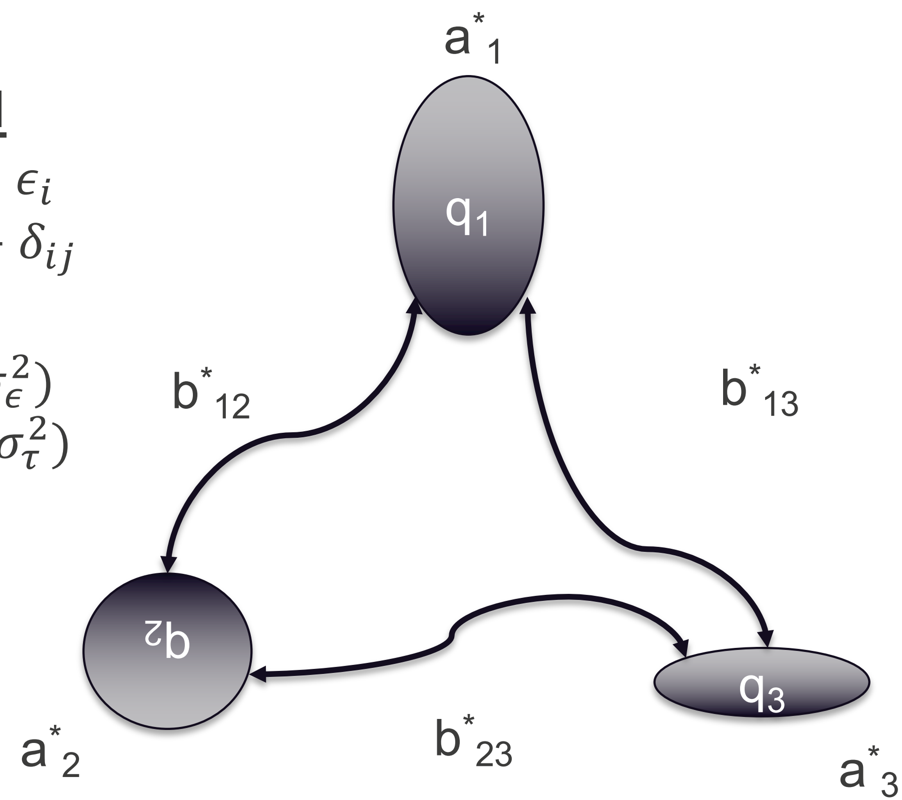


Model

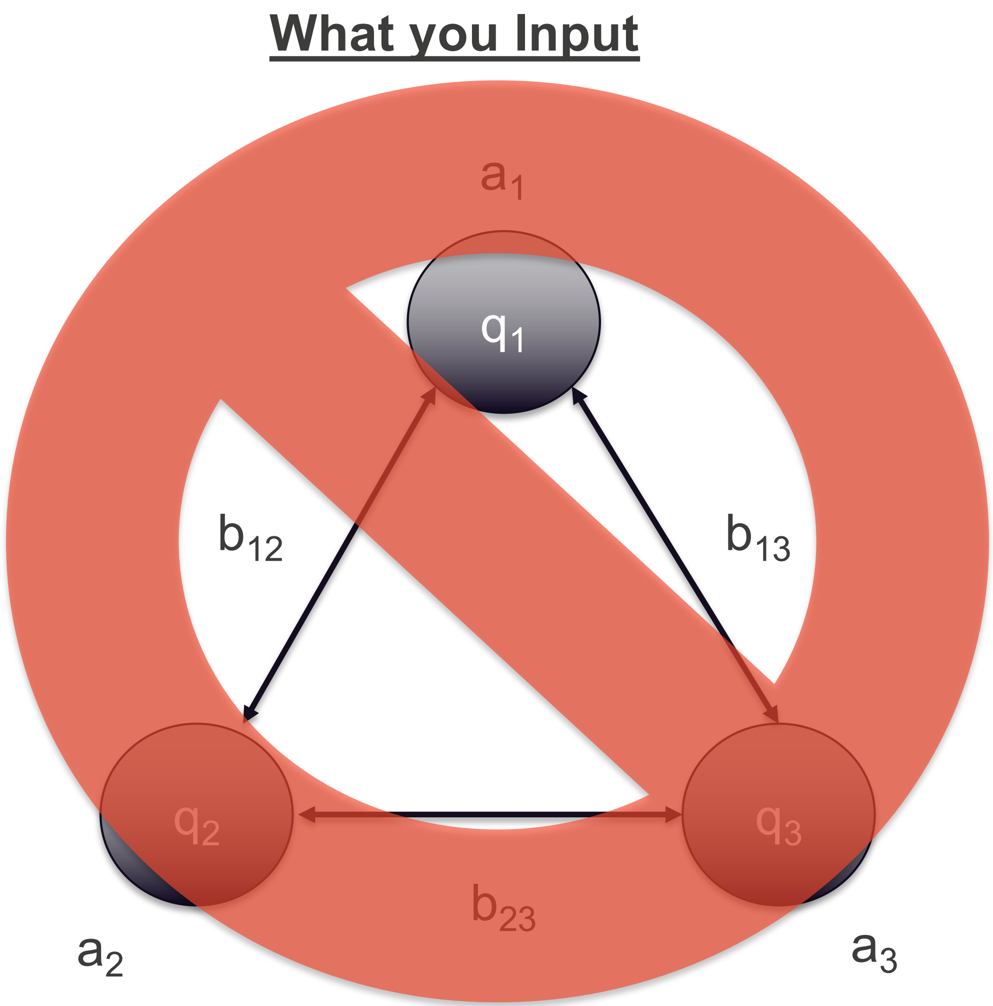
$$a_i^* = a_i + \epsilon_i$$
$$b_{ij}^* = b_{ij} + \delta_{ij}$$

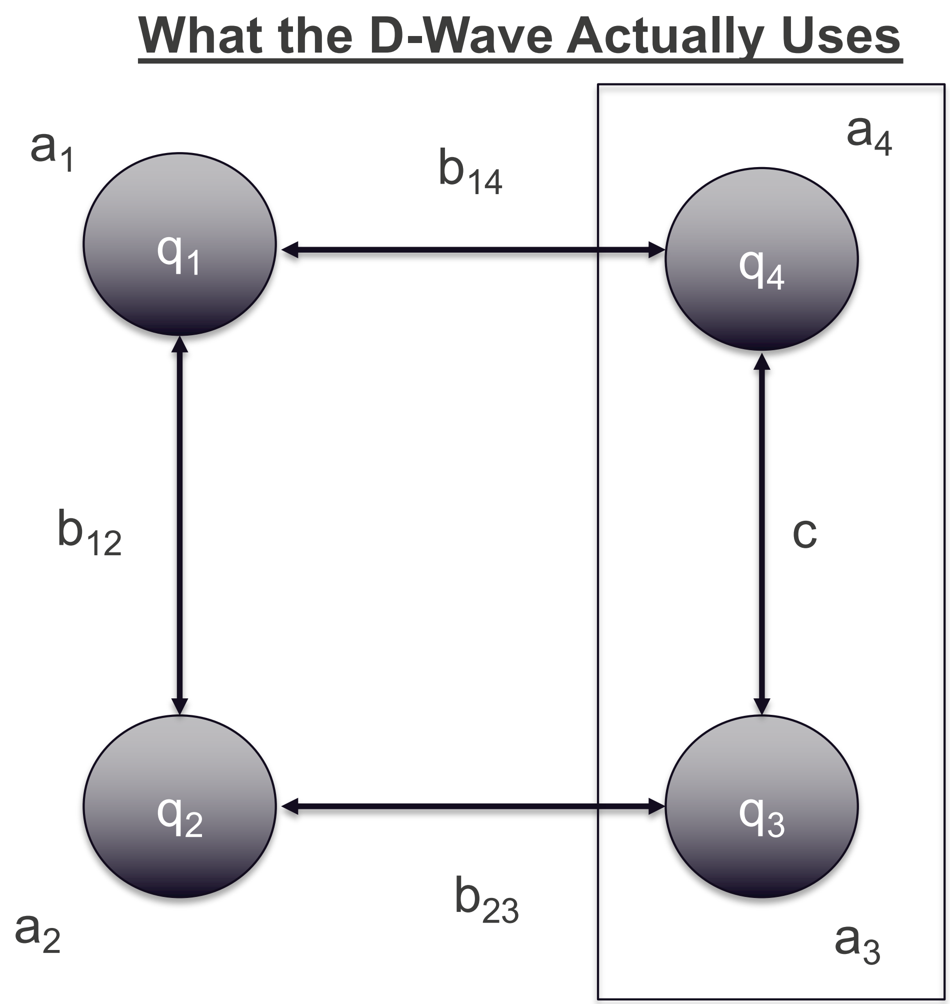
$$\epsilon \sim N(0, \sigma_\epsilon^2)$$
$$\delta_{ij} \sim N(0, \sigma_\tau^2)$$

What the D-Wave Actually Uses

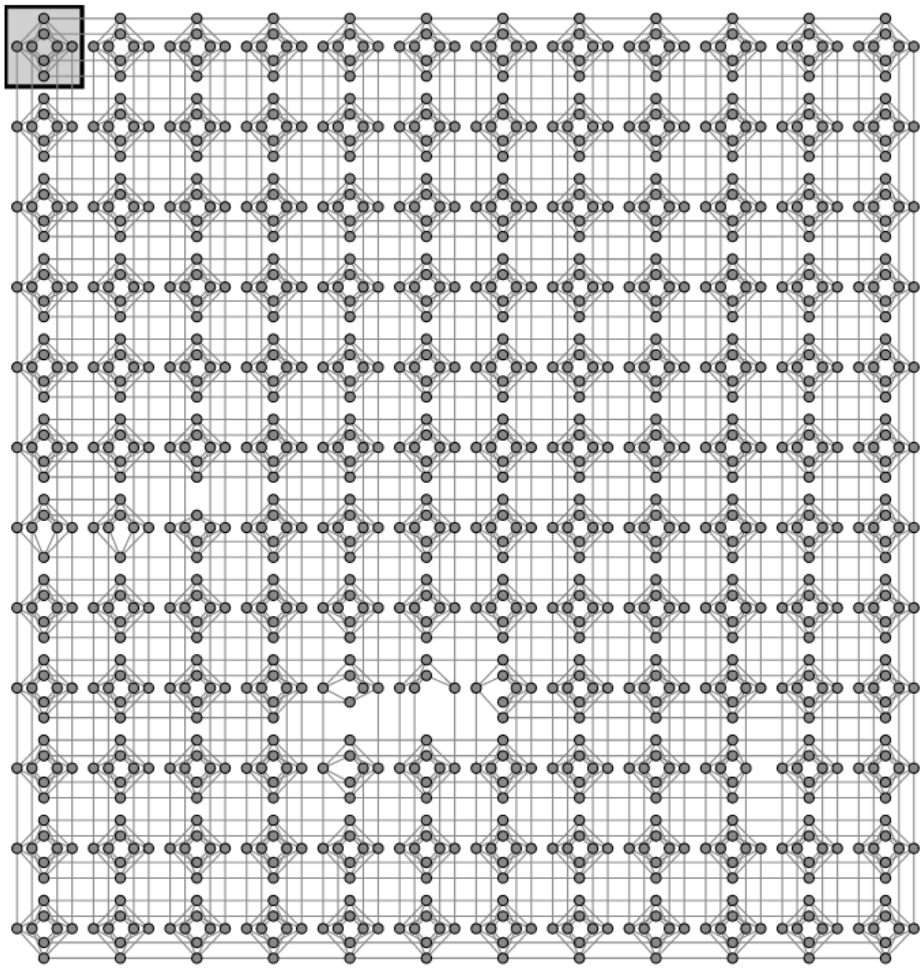


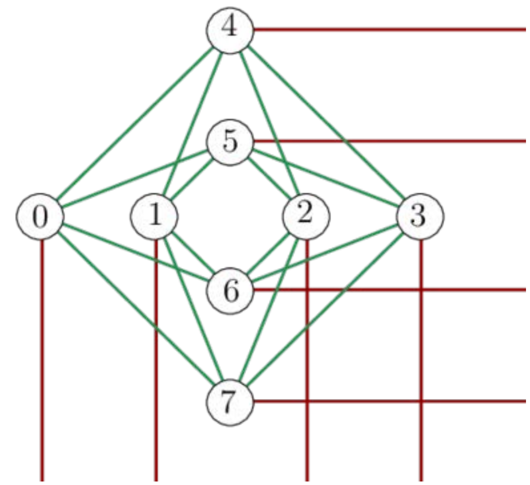
Problem: Hardware Constraints





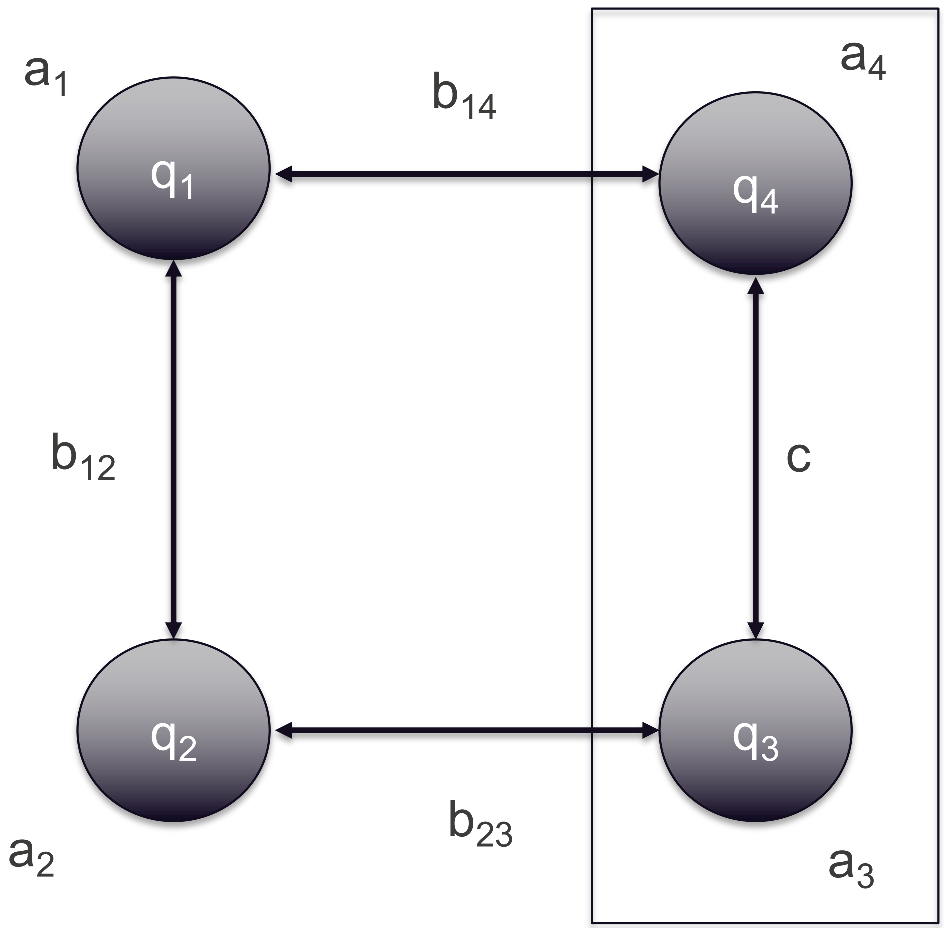
Chimera Graph Network





- Graphs must map onto physical D-Wave hardware network.
- D-Wave uses special graph structure called Chimera, chosen to satisfy physical constraints

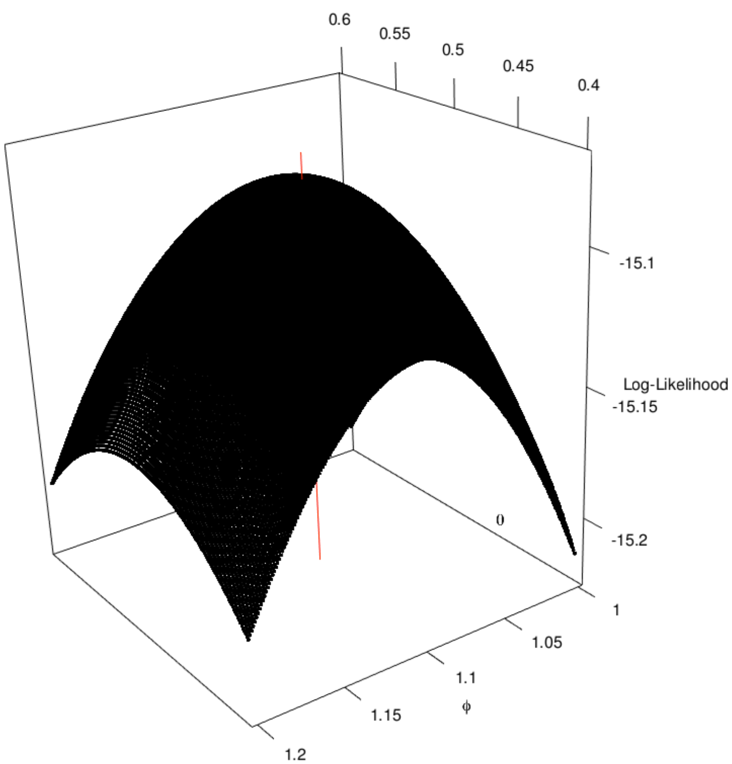
Solution: Embedding



- Qubits q_3 and q_4 treated as a single qubit with chain strength c chosen so that all lowest energy solutions have $q_3 = q_4$
- Long chains can be used to create complex graph structures
- D-Wave includes automated tools to do this for you!

Maximum Likelihood Estimation on the D-Wave

We have independent data x from a model with parameters θ and ϕ



$$x_d \sim f(x \mid \theta, \phi)$$

$$\ell(\theta, \phi \mid x) = \sum_{d=1}^{n} \log[f(x_d \mid \theta, \phi)]$$

$$(\hat{\theta}, \hat{\phi}) = \operatorname{argmax}_{\phi,\theta} \ \ell(\theta, \phi | x)$$

Assumptions:

1. The x_d are iid
2. Both $\theta > 0$ and $\phi > 0$ for computational simplicity

Formulating the Problem as a

Read the full paper at

arXiv:1904.06819

The likelihood needs to be written as

$$\ell(\theta, \phi \mid x_d) = \sum_i a_i q_i + \sum_{j>i} b_{ij} q_i q_j$$

1. Write $\theta = \sum_i 2^{p_{i,\theta}} q_{i,\theta}$ and $\phi = \sum_i 2^{p_{i,\phi}} q_{i,\phi}$

$$\theta = 2^{p_{1,\theta}} q_{1,\theta} + 2^{p_{2,\theta}} q_{2,\theta} + 2^{p_{3,\theta}} q_{3,\theta} + \ \dots$$

$$\phi = 2^{p_{1,\phi}} q_{1,\phi} + 2^{p_{2,\phi}} q_{2,\phi} + 2^{p_{3,\phi}} q_{3,\phi} + \ \dots$$

2. Perform a multivariate Taylor expansion on $\ell(\theta, \phi | x_d)$

$$\begin{aligned}\ell(\theta, \phi \mid x_d) \approx\ & \ell(\theta_0, \phi_0 | \, x_d) + \ell_\theta(\theta_0, \phi_0 | \, x_d)(\theta - \theta_0) + \ell_\phi(\theta_0, \phi_0 | \, x_d)(\phi - \phi_0) \\ & + \frac{1}{2}[\ell_{\theta\theta}(\theta_0, \phi_0 | \, x_d)(\theta - \theta_0)^2 + 2\ell_{\theta\phi}(\theta_0, \phi_0 \mid x_d)(\theta - \theta_0)(\phi - \phi_0) \\ & + \ell_{\phi\phi}(\theta_0, \phi_0 | x_d)(\phi - \phi_0)^2]\end{aligned}$$

Iterating the Procedure

Problem: The two-term Taylor series is only an approximation to the likelihood function, and will give wrong estimates $\hat{\theta}$ an

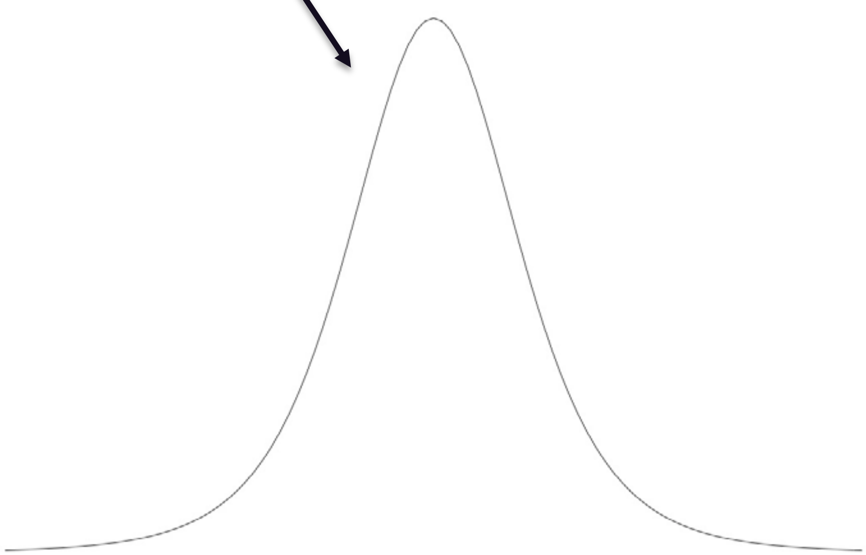
Solution: Iterate the procedure unti

1. Choose initial values (θ_0, ϕ_0)
2. Find maximum likelihood estimates $(\hat{\theta}, \hat{\phi})$ as maxima of Taylor series expanded around (θ_0, ϕ_0)
3. Take $(\hat{\theta}, \hat{\phi})$ as new expansion points (θ_0, ϕ_0)
4. Repeat 2-3 until stopping criterion met

At true maximum likelihood estimates $(\hat{\theta}, \hat{\phi})$, should get back $(\hat{\theta}, \hat{\phi}) = (\theta_0, \phi_0)$

Example: $N(\theta, \phi^2)$

Normal Density



Simplest case: can a quantum computer estimate the mean and variance of a normal distribution?

Data: $x = \{-2.296, -0.216, -0.082, 0.231, 1.127, 1.164, 1.189, 1.236, 1.272, 1.373\}$

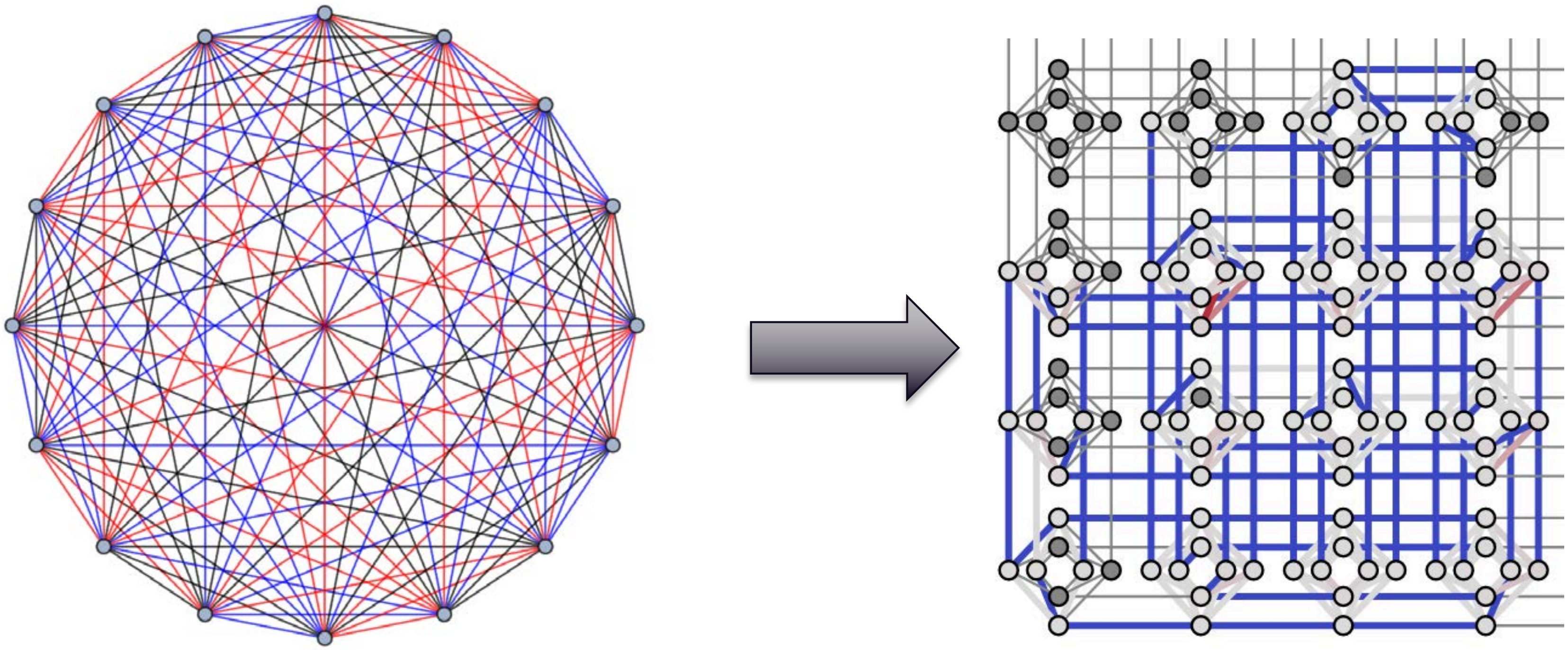
Maximum likelihood estimates:

$$\hat{\theta} = 0.4998 \qquad \hat{\phi} = 1.093$$

Required for algorithm:

1. All first and second order derivatives of log-likelihood function
2. Starting values θ_0 and ϕ_0

Details of Embedding

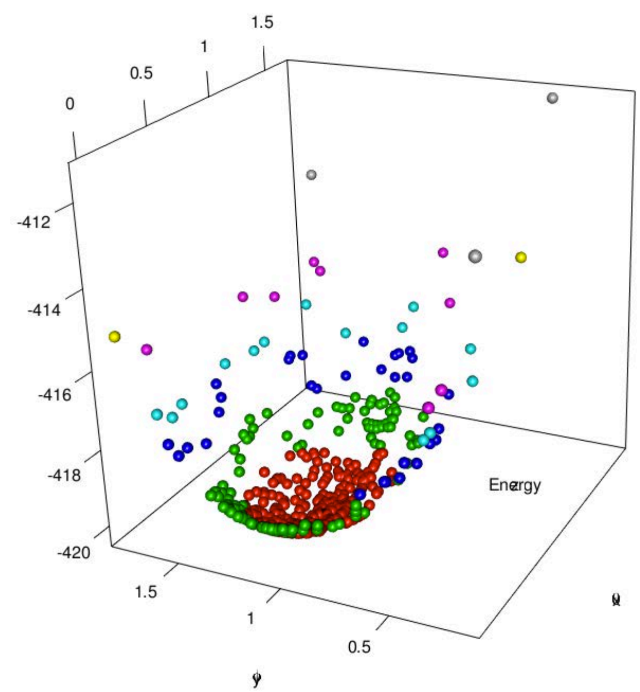


- 8 qubits used for each of θ and ϕ (16 total), ranging from powers 2^0 to 2^{-7}
- Complete K_{16} graph embedded onto Chimera hardware configuration

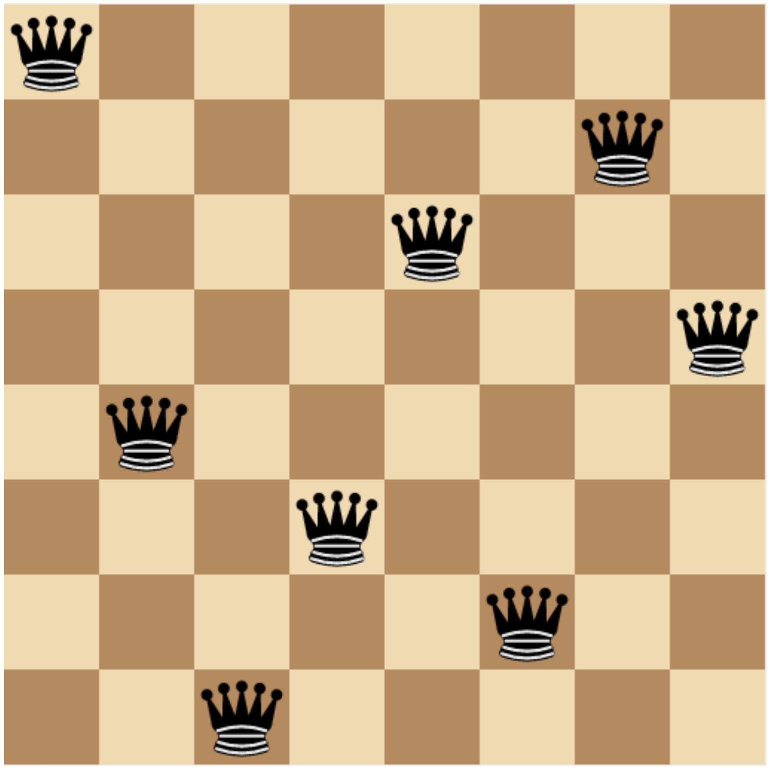
Results

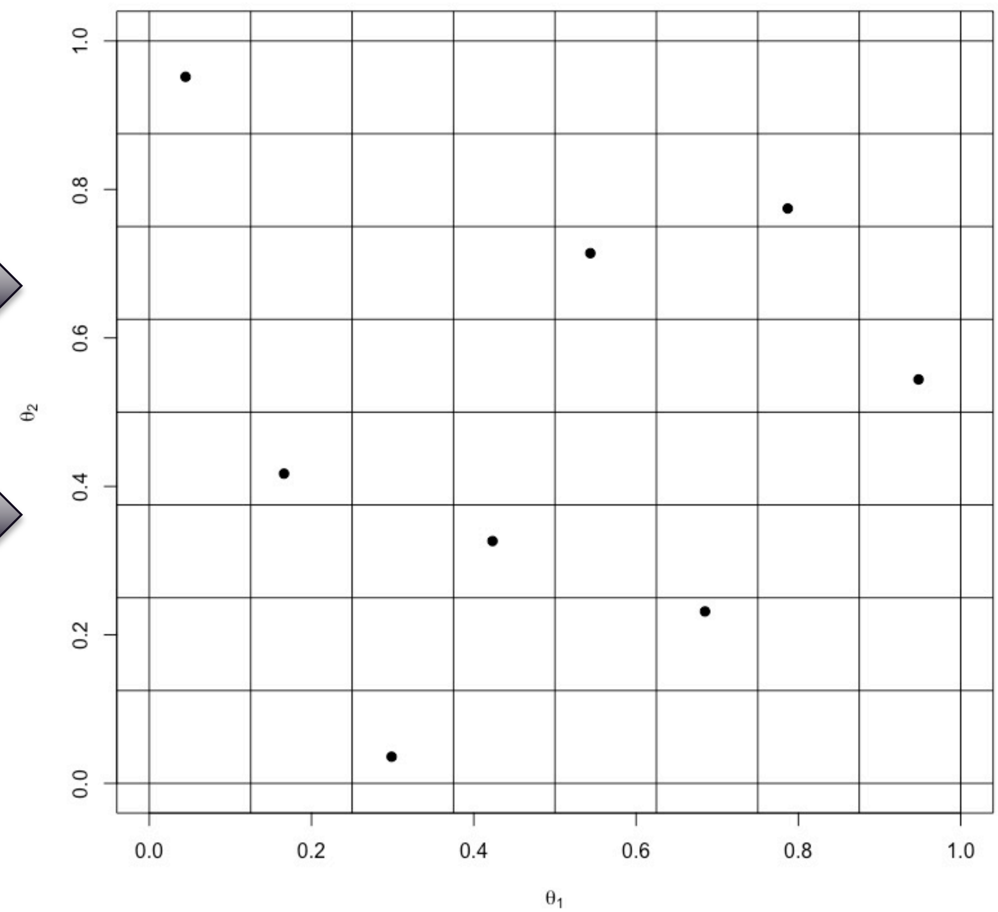
Iteration	$\hat{\theta}$	$\hat{\phi}$	Energy
1	0.5078125	0.9765625	-423.439
2	0.5	1.0625	-422.93
3	0.515625	1.0859375	-420.865
4	0.5	1.09375	-420.434
5	0.5	1.0859375	-420.283
6	0.4765625	1.09375	-420.42
7	0.53125	1.09375	-420.263
8	0.484375	1.09375	-420.308
9	0.5	1.09375	-420.272
10	0.5	1.0859375	-420.283
Truth	0.4998	1.093	N/A

Starting values of $\theta_0 = 0$ and ϕ_0 = 1 used.



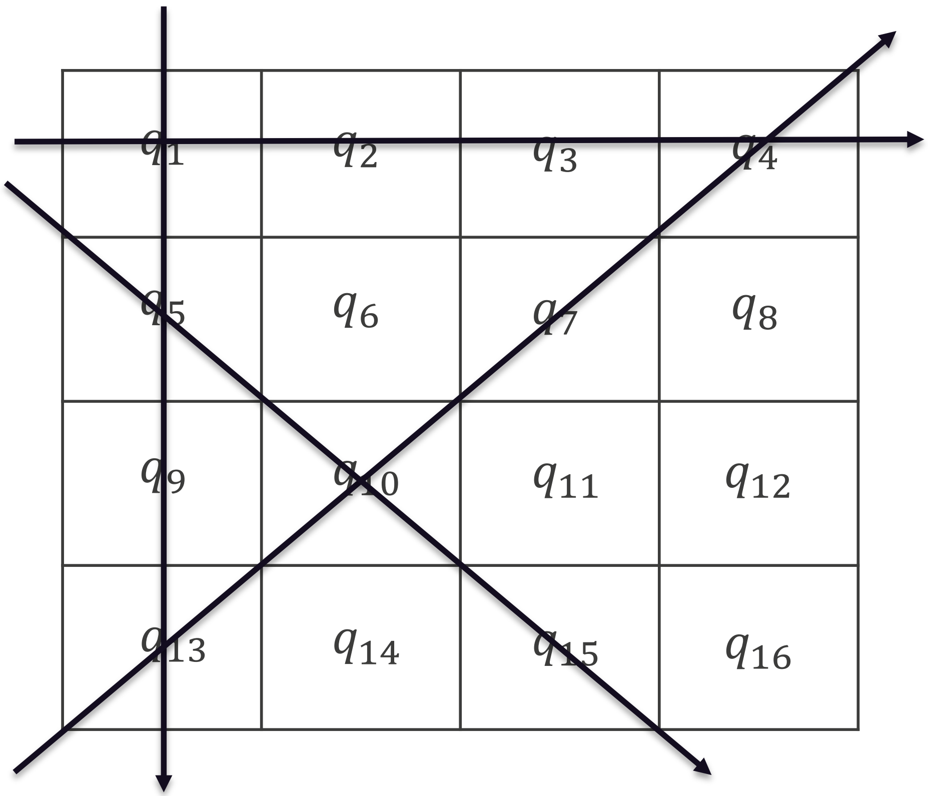
N-Queens Problem





Place *N* Queens on a chessboard so that none attack each other

Formulation as QUBO Matrix



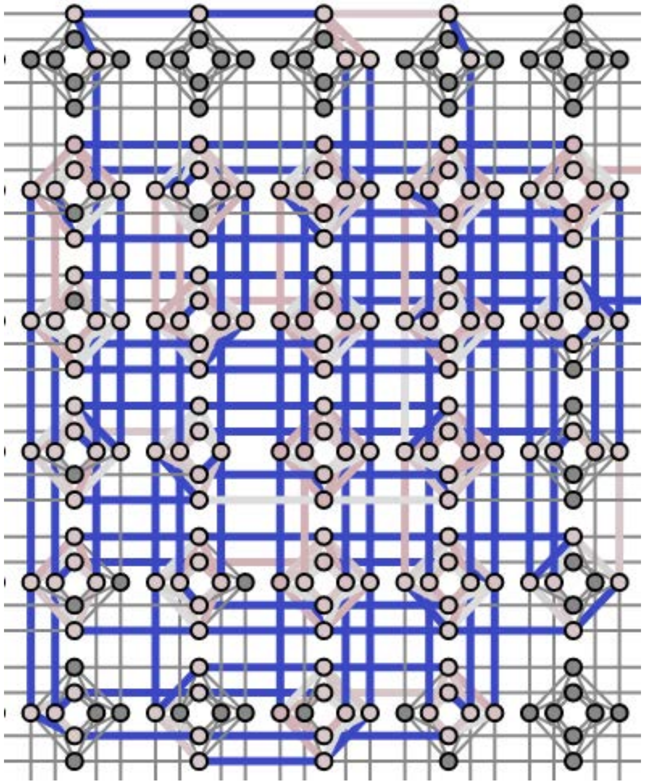
$$Energy = \sum_{i} a_i q_i + \sum_{j>i} b_{ij} q_i q_j$$

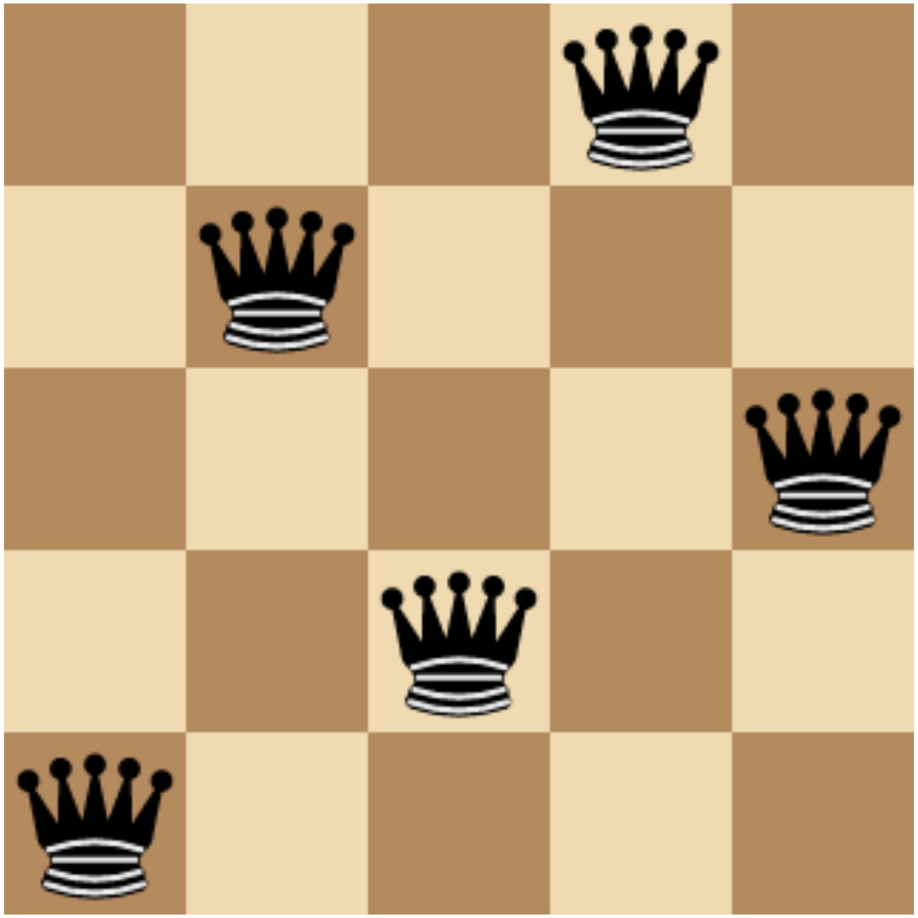
$a_i = -2$ for all i

$b_{ij} = 2$ for i, j in
the same row or same column

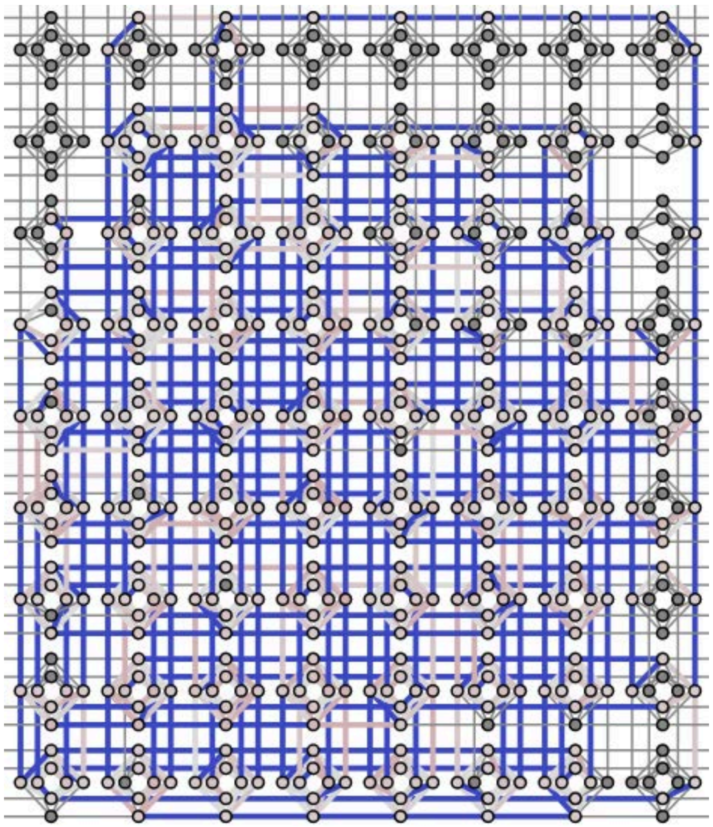
$b_{ij} = 1$ for i, j in
the same diagonal

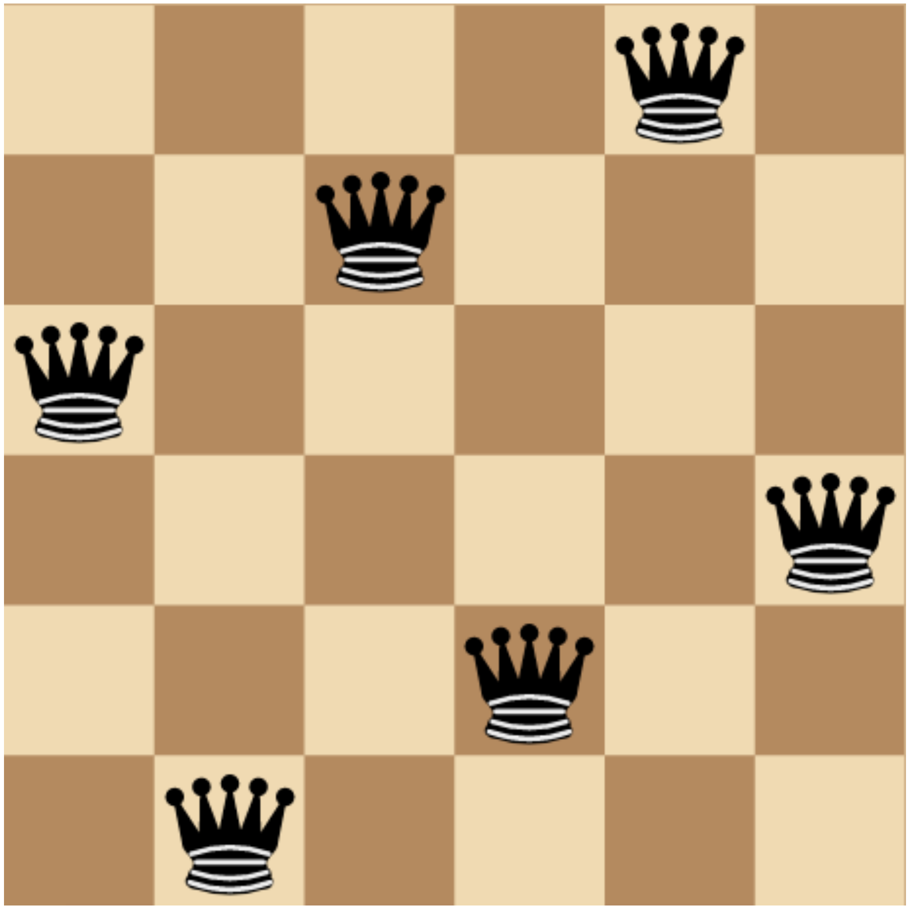
5 x 5 Embedding and Result



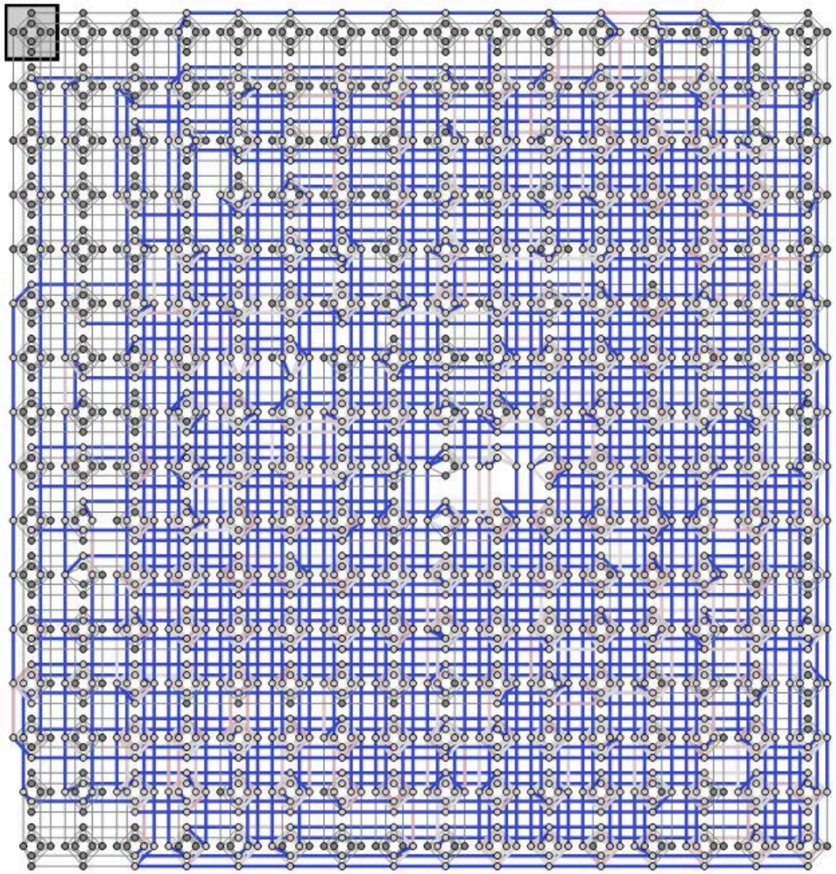


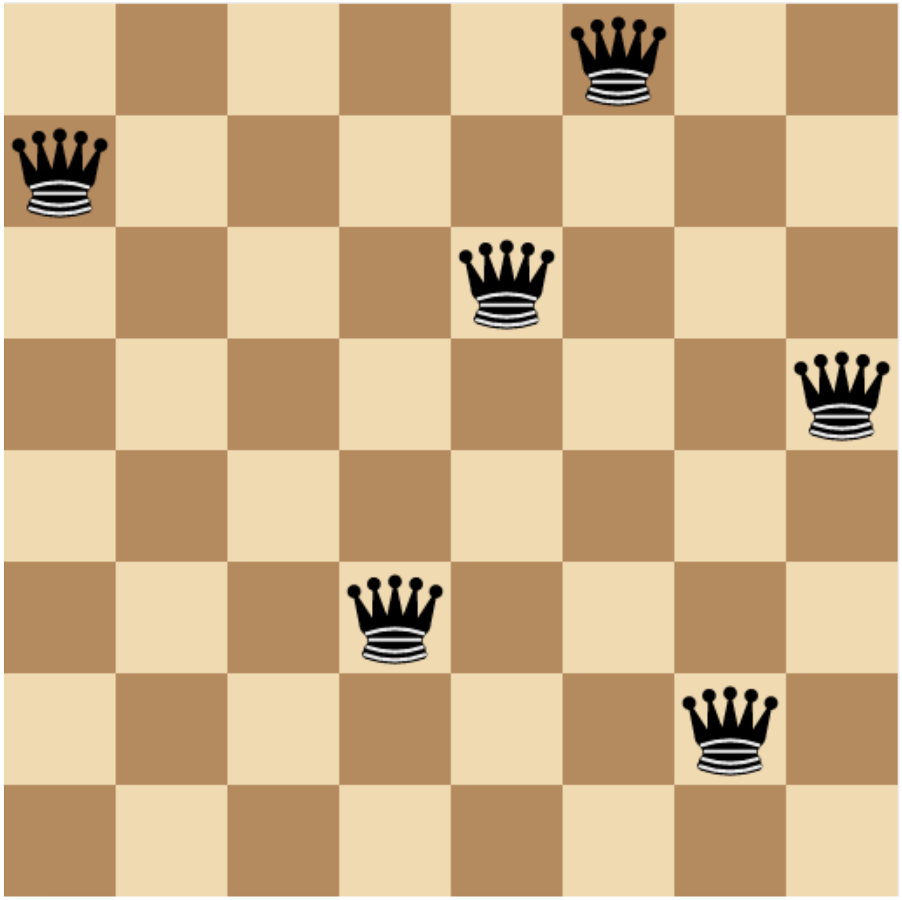
6 x 6 Embedding and Result





8 x 8 Embedding and Result





Example: Matrix Inversion

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix} \times \begin{bmatrix} V_{11} & V_{12} & V_{13} & V_{14} \\ V_{21} & V_{22} & V_{23} & V_{24} \\ V_{31} & V_{32} & V_{33} & V_{34} \\ V_{41} & V_{42} & V_{43} & V_{44} \end{bmatrix} = I_{4\,x\,4}$$

For a given column k of $V = A^{-1}$, define the column energy as

$$E_k = (\sum_j A_{1j}V_{jk})^2 + \ ... + (\sum_j A_{kj}V_{jk} - 1)^2 + \ ... + (\sum_j A_{nj}V_{jk})^2$$

The inverse is the set of V_{ij} which minimize the sum E of column energies.

Each energy E_k can be minimized independently of the others to obtain one column of the inverse matrix. This could even be done in parallel!

Results

Problem Setup

1. Goal: invert 3×3 input matrix, 3 anneal steps
2. 6 qubits per V_{ij} entry, 18 qubits total

$$A = \begin{bmatrix} 1.344 & 0.418 & -0.935 \\ -1.018 & 1.095 & -0.250 \\ 0.277 & -0.384 & 0.755 \end{bmatrix}$$

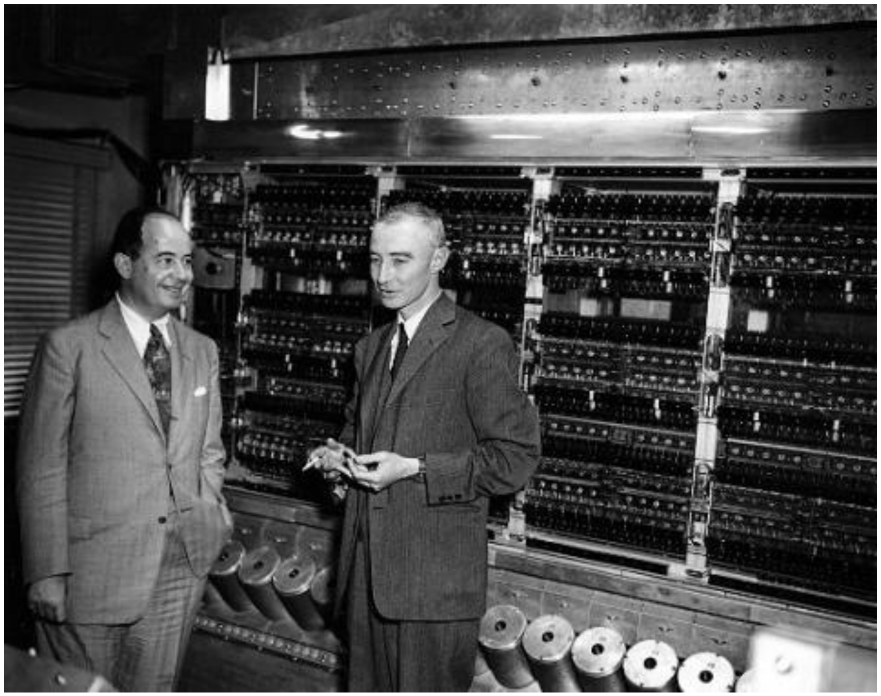
$$A \times \begin{bmatrix} 0.613 & 0.037 & 0.772 \\ 0.586 & 1.068 & 1.080 \\ 0.074 & 0.531 & 1.592 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Truth

$$A \times \begin{bmatrix} 0.625 & 0.0 & 0.75 \\ 0.5 & 1.0625 & 1.1875 \\ 0.0 & 0.5625 & 1.6875 \end{bmatrix} = \begin{bmatrix} 1.049 & -0.082 & -0.073 \\ -0.088 & 1.023 & 0.116 \\ -0.019 & 0.016 & 1.025 \end{bmatrix}$$

Quantum Estimate

Conclusions



1. Quantum computers are here, and you can use them to impress your friends and colleagues
2. But they're fairly limited both in problem scope and scale
3. Even then, the results aren't great
4. Look for problems with lots of local optima that you might use simulated annealing on

arXiv:1904.06819