LA-UR-19-21739
Approved for public release; distribution is unlimited.

Title: Challenges with Chains: Testing the Limits of a D-Wave Quantum Annealer for Discrete Optimization

Author(s): Coffrin, Carleton James

Intended for: external release

Issued: 2019-02-27

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Challenges with Chains

Testing the Limits of a D-Wave Quantum Annealer for Discrete Optimization

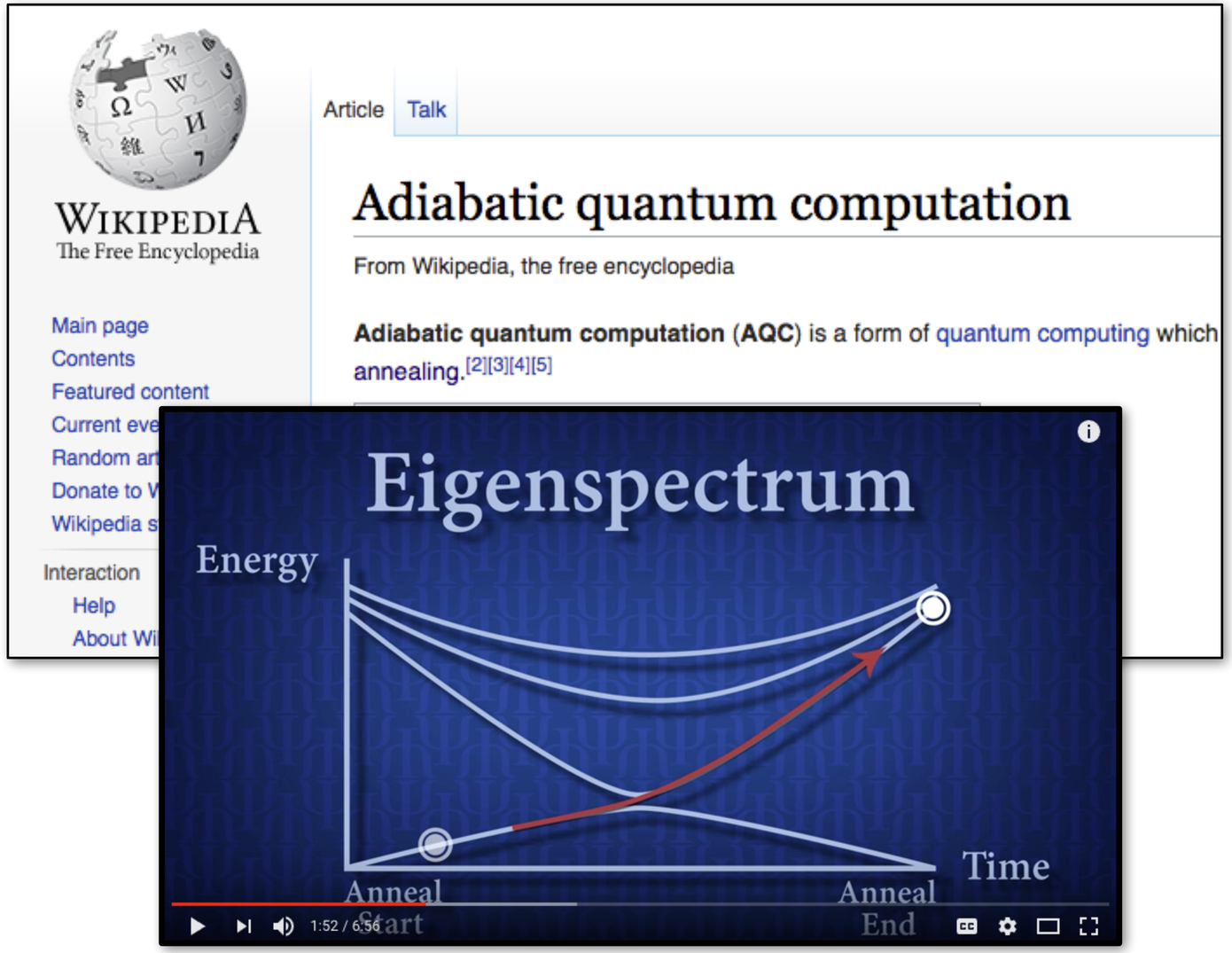
Carleton Coffrin

Los Alamos National Laboratory

Advanced Network Science Initiative

Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA

So about that D-Wave…



The D-Wave User Manual (first order approximation)

Ising Model

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{c}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{c}_i \sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$

Quadratic Unconstrained Binary Optimization (QUBO)

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{c}_{ij} x_i x_j + \sum_{i \in \mathcal{N}} \boldsymbol{c}_i x_i$$

s.t.

$$x_i \in \{0, 1\} \; \forall i \in \mathcal{N}$$

So what?

WOW!

Max-Cut

Max-Independent Set

Max-Clique

Research Agenda

1. **Pick your favorite combinatorial optimization benchmark library**

2. **Encode as an Ising model and solve on the D-Wave**

3. **Revolutionize the field of discrete optimization**

4. **Graciously accept Turing Award**

5. **Retire**

A First Attempt: DIMACS Max-Clique

http://iridia.ulb.ac.be/~fmascia/maximum_clique/DIMACS-benchmark

dimacs benchmark set

This is a selected set of instances form the Second DIMACS Implementation Challenge (1992-1993).
In the following table, the clique number ω(G) corresponds to the global optimum or to the lower bound as indicated by the
For some instances the lower bound has been confirmed to coincide with the global optimum. For such instances, the cliqu
The exact algorithms that confirmed the bound are reported later on this page.

instance	ω(g)	best known	nodes	edges	graph degrees		best degrees	
					median	iqr	median	iqr
C125.9	34*	34	125	6 963	112.0	(5.00)	114.5	(4.75)
C250.9	44*	44	250	27 984	224.0	(6.00)	227.0	(5.00)
C500.9	≥ 57	57	500	112 332	449.0	(9.00)	455.0	(9.00)
C1000.9	≥ 68	68	1 000	450 079	900.0	(13.00)	907.0	(11.25)
C2000.9	≥ 80	80	2 000	1 799 532	1 800.0	(18.00)	1 803.0	(15.25)
DSJC1000_5	15	15	1 000	499 652	500.0	(20.00)	503.0	(23.00)
DSJC500_5	13	13	500	125 248	250.0	(16.00)	259.0	(14.00)
C2000.5	16*	16	2 000	999 836	999.0	(30.00)	1 006.0	(11.50)

A First Attempt: DIMACS Max-Clique

ISTI RR 2016

(Straw Man)

GUROBI OPTIMIZATION

D:wave

			gurobi			dwave				
Case	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	Best Sol.	Opt. Gap	Time	Best Sol.	Best	Inf.	Samples	Time
C015_9	15	12	-11	0%	<1	-11	9073	1	10000	0+3
C020_9	20	17	-14	0%	<1	-14	8370	85	10000	0+3
C030_9	30	44	-16	0%	<1	-16	5651	123	10000	0+3
C040_9	40	77	-18	0%	<1	-18	3865	316	10000	0+4
C050_9	50	108	-24	0%	<1	-24	16	1254	10000	0+4
C060_9	60	158	-25	0%	<1	-25	22	5465	10000	0+5
C070_9	70	215	-27	0%	<1	-26	1	9855	10000	4+5
C080_9	80	306	-29	0%	<1	F.E.	-	-	-	T.L.
C090_9	90	407	-29	0%	1.0	F.E.	-	-	-	T.L.
C100_9	100	508	-30	0%	2.0	F.E.	-	-	-	T.L.
C110_9	110	615	-32	0%	5.1	F.E.	-	-	-	T.L.
C120_9	120	729	-32	0%	45	F.E.	-	-	-	T.L.
C125_9	125	787	-34	0%	55	F.E.	-	-	-	T.L.
C250_9	250	3141	-43	40%	T.L.	F.E.*	-	-	-	T.L.

F.E.
Failed
Embed

T.L.
Time
Limit
(1 hour)

Hard Library Problems :: nodes 4,000 - edges 4,000,000

An Inconvenient Reality

Every attempt to show D-Wave supremacy via combinatorial optimization has failed…

What's Going On?

(Straw Man)

GUROBI OPTIMIZATION

D-Wave

			gurobi			dwave				
Case	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	Best Sol.	Opt. Gap	Time	Best Sol.	Best	Inf.	Samples	Time
C015_9	15	12	-11	0%	<1	-11	9073	1	10000	0+3
C020_9	20	17	-14	0%	<1	-14	8370	85	10000	0+3
C030_9	30	44	-16	0%	<1	-16	5651	123	10000	0+3
C040_9	40	77	-18	0%	<1	-18	3865	316	10000	0+4
C050_9	50	108	-24	0%	<1	-24	16	1254	10000	0+4
C060_9	60	158	-25	0%	<1	-25	22	5465	10000	0+5
C070_9	70	215	-27	0%	<1	-26	1	9855	10000	4+5
C080_9	80	306	-29	0%	<1	F.E.	-	-	-	T.L.
C090_9	90	407	-29	0%	1.0	F.E.	-	-	-	T.L.
C100_9	100	508	-30	0%	2.0	F.E.	-	-	-	T.L.
C110_9	110	615	-32	0%	5.1	F.E.	-	-	-	T.L.
C120_9	120	729	-32	0%	45	F.E.	-	-	-	T.L.
C125_9	125	787	-34	0%	55	F.E.	-	-	-	T.L.
C250_9	250	3141	-43	40%	T.L.	F.E.*	-	-	-	T.L.

F.E.
Failed Embed

T.L.
Time Limit (1 hour)

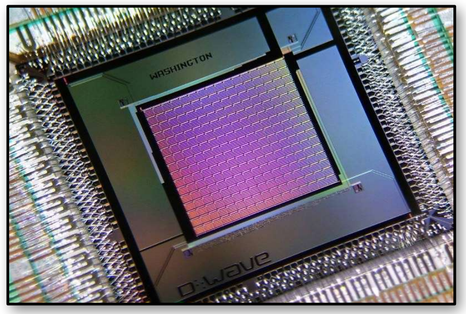
Overview

- What does the D-Wave Compute?
- The basic approach to Discrete Optimization via the D-Wave
- Mathematics of Boltzmann Sampling
- Mathematics of Chains
- The Indecisive Ising Model
- The Consensus QUBO Model
- Mitigations and Conclusions

What does a D-Wave Compute?

The D-Wave User Manual (first order approximation)

$$\mathcal{G} = (\mathcal{N}, \mathcal{E})$$
$$\mathcal{G} \subseteq C_{16}$$



$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

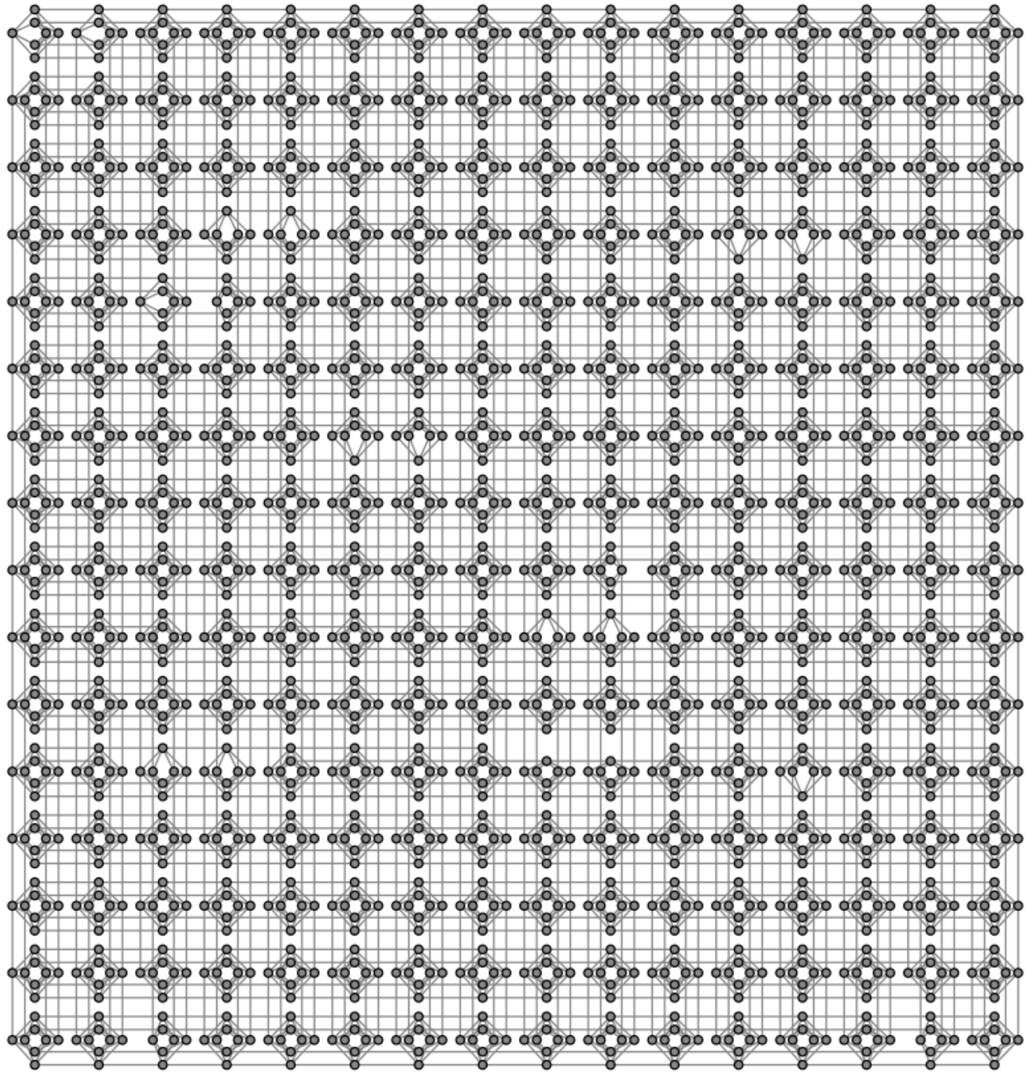
Solves

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i \sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$

DW_2000Q_3



The User's Perspective

Ising Model Specification

$$f(\sigma) = \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$

Runtime Parameters (e.g. # of replicates)

D-Wave

Variable Assignments

count	σ_1	σ_2	$\ldots$	σ_n
342	-1	1	$\ldots$	1
173	1	-1	$\ldots$	-1
12	1	1	$\ldots$	-1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

Hold the phone…

Beyond the D-Wave User Manual (second order approx.)

The Ising model: teaching an old problem new tricks

Zhengbing Bian, Fabian Chudak, William G. Macready*, Geordie Rose

D-Wave Systems

August 30, 2010

Abstract

In this paper we investigate the use of hardware which physically realizes quantum annealing for machine learning applications. We show how to take advantage of the hardware in both zero- and finite-temperature modes of operation. At zero temperature the hardware is used as a heuristic minimizer of Ising energy functions, and at finite temperature the hardware allows for sampling from the corresponding Boltzmann distribution. We rely on quantum mechanical processes to perform both these tasks more efficiently than is possible through software simulation on classical computers. We show how Ising energy functions can be sculpted to solve a range of supervised learning problems. Finally, we validate the use of the hardware by constructing learning algorithms trained using quantum annealing on several synthetic and real data sets. We demonstrate that this novel approach to learning using quantum mechanical hardware can provide significant performance gains for a number of structured supervised learning problems.

https://www.dwavesys.com/resources/publications?type=internal

Discrete Optimization

$$\min : \sum_{i,j\in\mathcal{E}} \boldsymbol{c}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{c}_i\sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\}\ \forall i \in \mathcal{N}$$

Boltzmann Sampler

$$P(\sigma) \propto \boldsymbol{e}^{\frac{\sum_{i,j\in\mathcal{E}} \boldsymbol{c}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{c}_i\sigma_i}{\tau}}$$

$$\sigma_i \in \{-1, 1\}$$

$\boldsymbol{\tau}$?

Learning a Model of the D-Wave

$$\mu(\underline{\sigma}) \propto \exp\left(\underbrace{\sum_i h_i \sigma_i}_{\text{1st Order}} + \underbrace{\sum_{ij} J_{ij} \sigma_i \sigma_j}_{\text{2nd Order}} + \underbrace{\sum_{ijk} J_{ijk} \sigma_i \sigma_j \sigma_k}_{\text{3rd Order}} + \underbrace{\sum_{ijkl} J_{ijkl} \sigma_i \sigma_j \sigma_k \sigma_l}_{\text{4th Order}} + \cdots \right)$$

Sidhant Misra

Marc Vuffray

Andrey Lokhov

Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models

M. Vuffray, S. Misra, A. Lokhov, M. Chertkov

(2016)

Optimal Structure and Parameter Learning of Ising Models

A. Lokhov, M. Vuffray, S. Misra, M. Chertkov

(2018)

GitHub https://github.com/lanl-ansi/inverse_ising

Learned D-Wave Model

Ising Model Specification

$$\min : \sum_{i,j\in\mathcal{E}} \boldsymbol{c}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{c}_i\sigma_i$$

s.t.

$$\sigma_i \in \{-1,1\}\ \forall i \in \mathcal{N}$$

Runtime Parameters (e.g. # of replicates)

D-Wave

Boltzmann Sampler

$$P(\sigma) \propto e^{\frac{\left(\sum_{i,j\in\mathcal{E}} \boldsymbol{c}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{c}_i\sigma_i\right)}{\tau}}$$

$$\sigma_i \in \{-1,1\}$$

$$\boldsymbol{\tau} \approx 0.1$$

Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models

M. Vuffray, S. Misra, A. Lokhov, M. Chertkov

(2016)

SciAdv AAAS

Optimal Structure and Parameter Learning of Ising Models

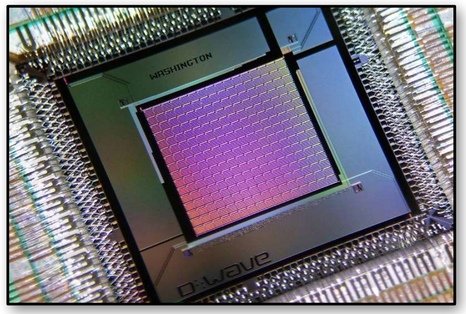
A. Lokhov, M. Vuffray, S. Misra, M. Chertkov

(2018)

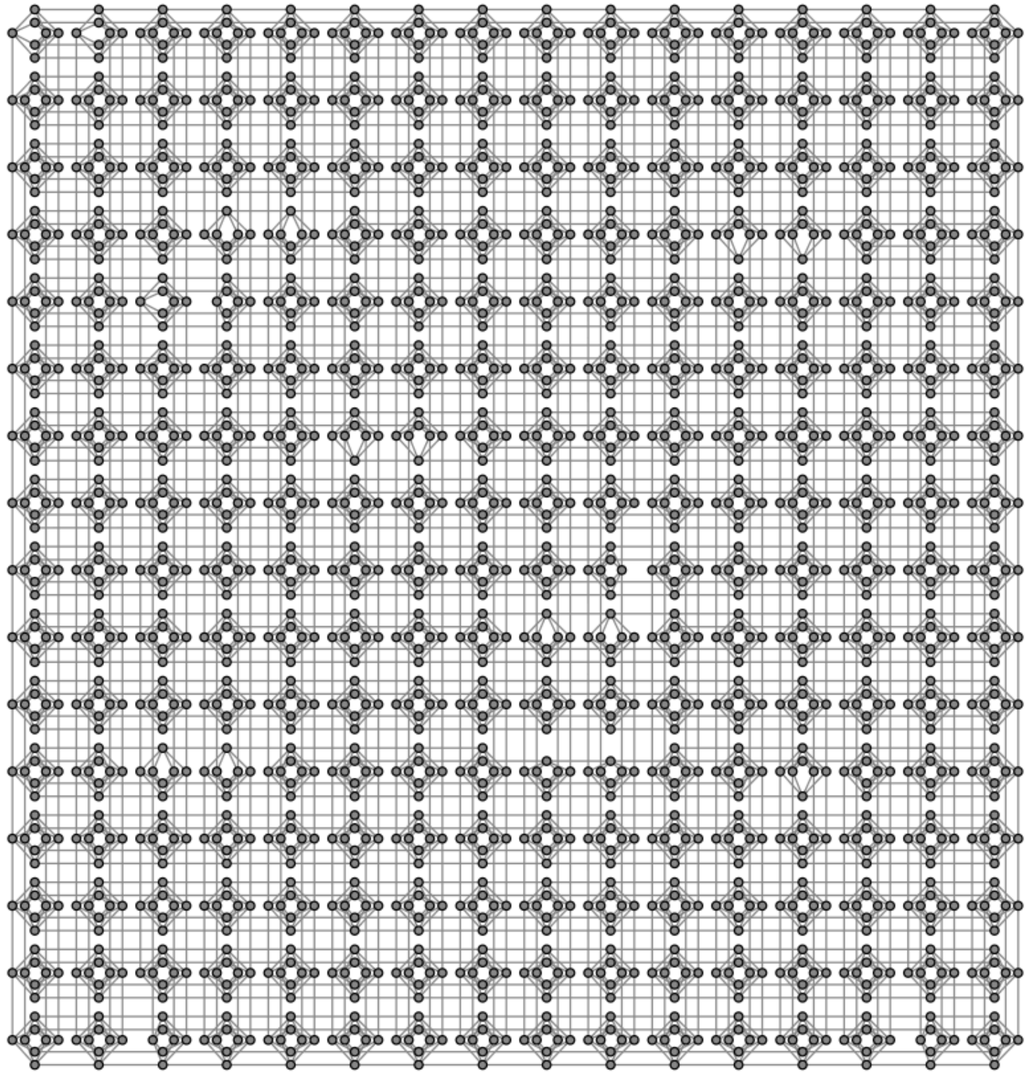
GitHub https://github.com/lanl-ansi/inverse_ising

Revised Model (second order approximation)

$$\mathcal{G} = (\mathcal{N}, \mathcal{E})$$
$$\mathcal{G} \subseteq C_{16}$$



DW_2000Q_3



$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

$$P(\sigma) \propto \boldsymbol{e}^{\frac{\left(\sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i\right)}{0.1}}$$

$$\sigma_i \in \{-1, 1\}$$

Optimization on D-Wave

The Challenge

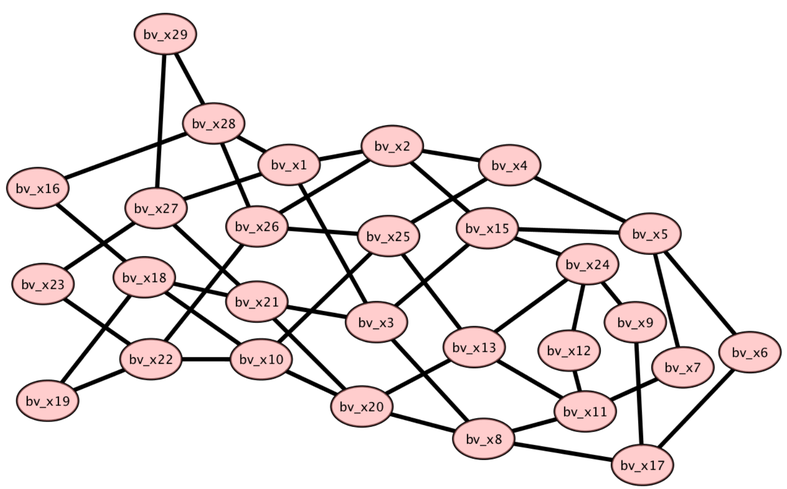
Source Problem

QUBO

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{c}_{ij} x_i x_j + \sum_{i \in \mathcal{N}} \boldsymbol{c}_i x_i$$

s.t.

$$x_i \in \{0, 1\} \ \forall i \in \mathcal{N}$$



qblib_3867

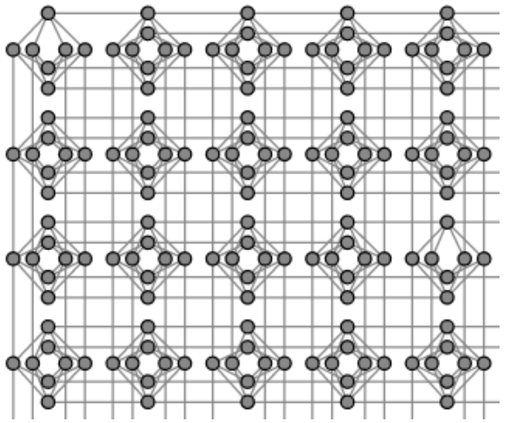
Target Problem

Ising

$$f(\sigma) = \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i \sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \ \forall i \in \mathcal{N}$$



$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \ \ \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \ \ \forall i \in \mathcal{N}$$

Core Challenges

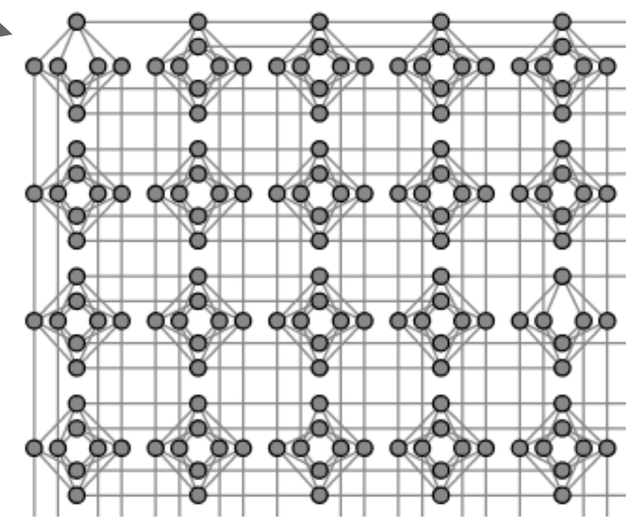
- **Ising Formulation**

$$f(\sigma) = \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \ \forall i \in \mathcal{N}$$

- **Graph Sparsity**



- **Parameter Range**

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \ \ \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \ \ \forall i \in \mathcal{N}$$

Ising Formulation

QUBO

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{c}_{ij} x_i x_j + \sum_{i \in \mathcal{N}} \boldsymbol{c}_i x_i$$

s.t.

$$x_i \in \{0, 1\} \ \forall i \in \mathcal{N}$$

$$x_i = \frac{\sigma_i + 1}{2}$$

Ising

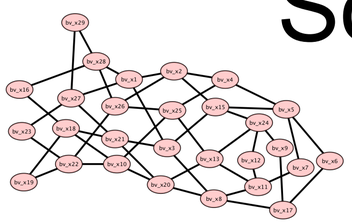
$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i \sigma_i + \boldsymbol{o}$$

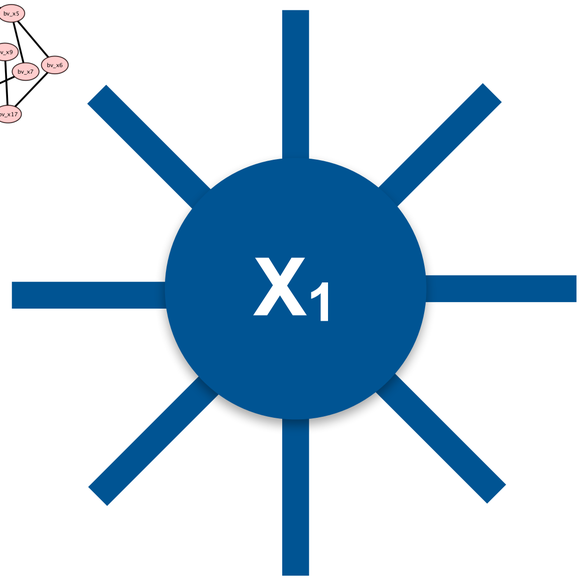
s.t.

$$\sigma_i \in \{-1, 1\} \ \forall i \in \mathcal{N}$$

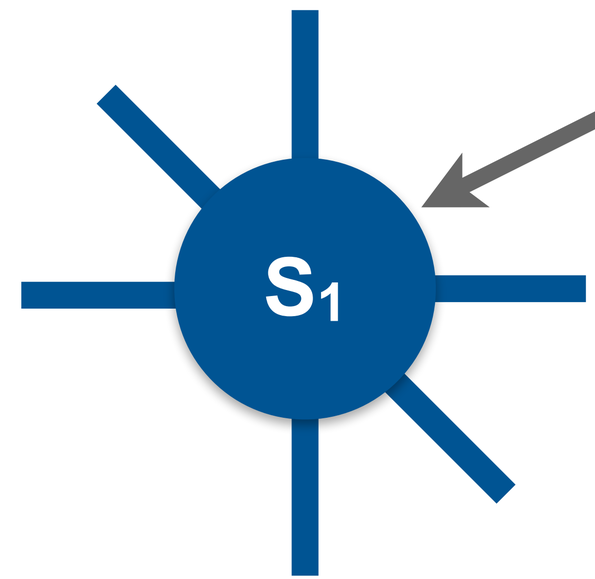
Graph Sparsity

Source Graph

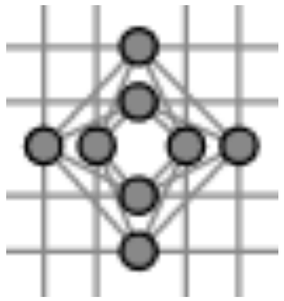




Target Graph

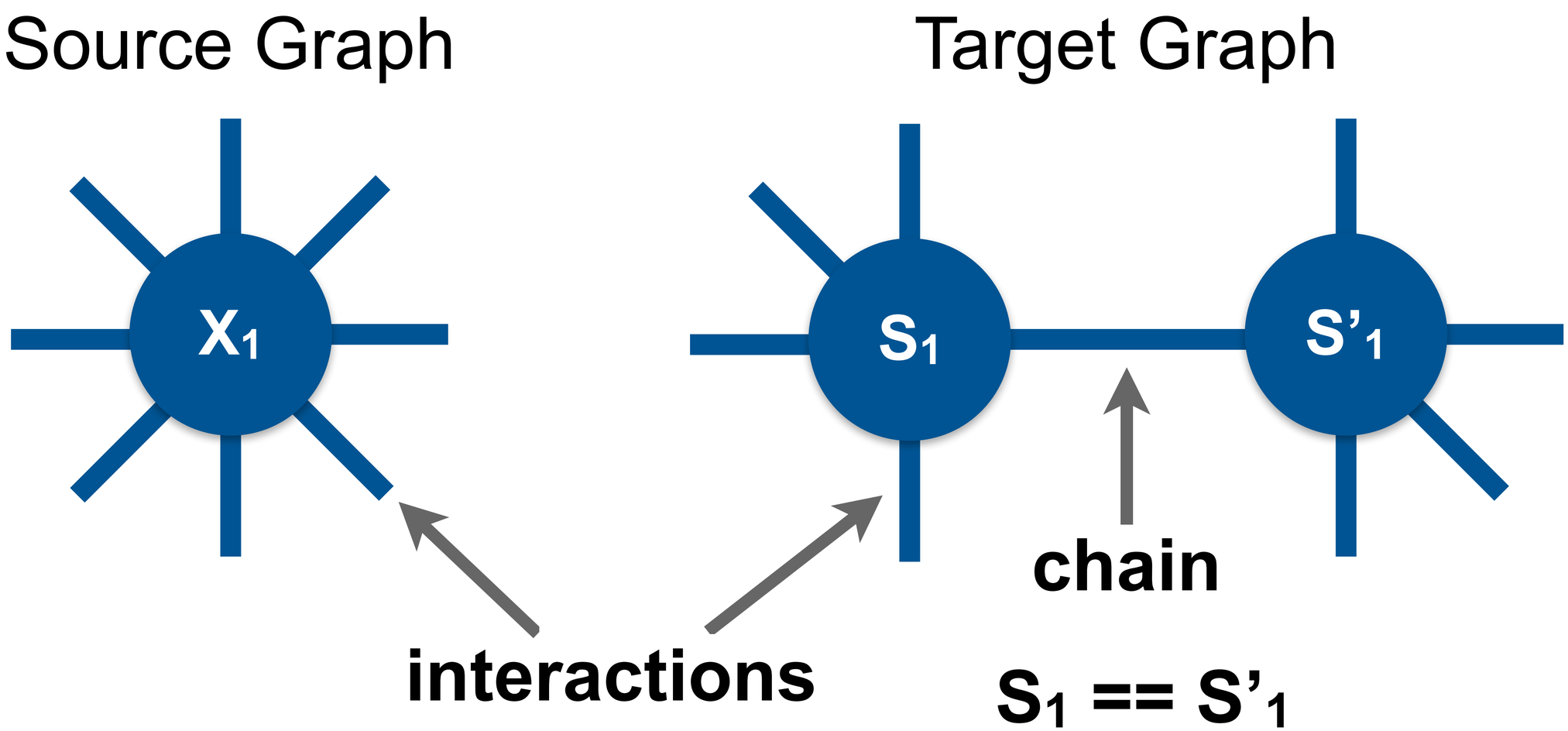


only 6 edges!



Idea: Lift to a higher dimensional space

Graph Sparsity



Warning: broken chains = infeasible solution

Parameter Range

Scaling Factor: $\boldsymbol{s}$

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

Ising

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i \sigma_i + \boldsymbol{o}$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$
$$\mathcal{E} \subseteq \mathcal{C}_{12}$$

$$\boldsymbol{J}'_{ij} = \boldsymbol{s}\boldsymbol{J}_{ij} \; \forall i,j \in \mathcal{E}$$
$$\boldsymbol{h}'_i = \boldsymbol{s}\boldsymbol{h}_i \; \forall i \in \mathcal{N}$$
$$\boldsymbol{o}' = \boldsymbol{s}\boldsymbol{o}$$

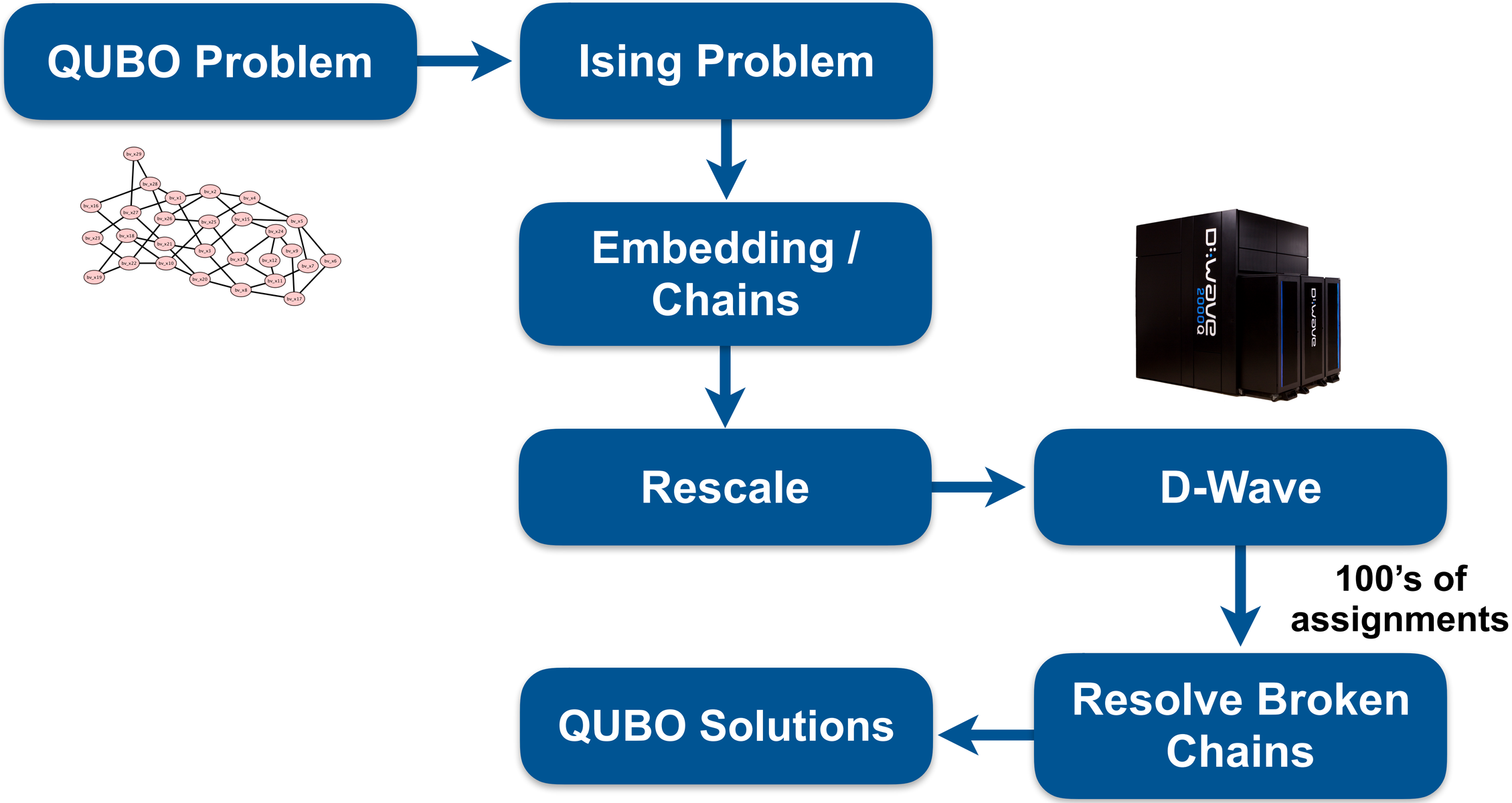
Rescaled Ising

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}'_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}'_i \sigma_i + \boldsymbol{o}'$$

s.t.

$$-1 \leq \boldsymbol{J}'_{ij} \leq 1 \; \forall i,j \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}'_i \leq 2 \; \forall i \in \mathcal{N}$$
$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$
$$\mathcal{E} \subseteq \mathcal{C}_{12}$$

Typical D-Wave Algorithm

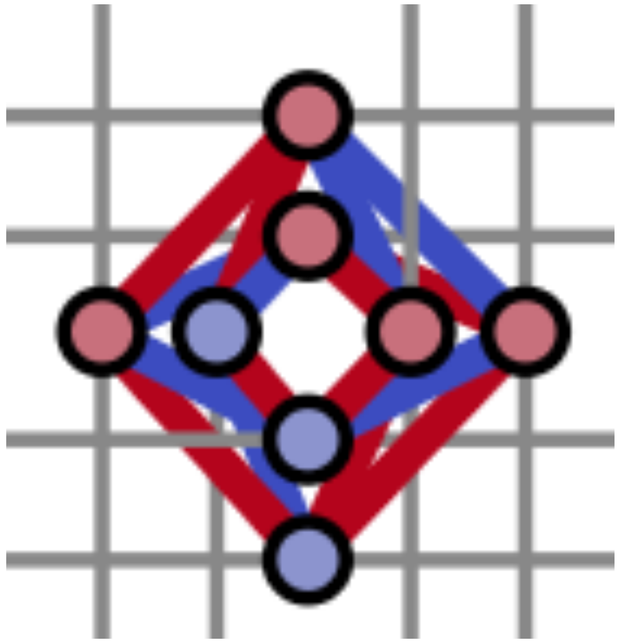


Mathematics of Boltzmann Sampling

A Random Example

$$f(\sigma) = \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i \qquad \mathcal{N} = \{1,2,3,4,5,6,7,8\}$$

$\boldsymbol{h}_1 = 0.1$	$\boldsymbol{J}_{15} = -0.1$	$\boldsymbol{J}_{35} = -0.1$
$\boldsymbol{h}_2 = -0.1$	$\boldsymbol{J}_{16} = 0.1$	$\boldsymbol{J}_{36} = 0.1$
$\boldsymbol{h}_3 = 0.1$	$\boldsymbol{J}_{17} = -0.1$	$\boldsymbol{J}_{37} = 0.1$
$\boldsymbol{h}_4 = 0.1$	$\boldsymbol{J}_{18} = -0.1$	$\boldsymbol{J}_{38} = -0.1$
$\boldsymbol{h}_5 = -0.1$	$\boldsymbol{J}_{25} = -0.1$	$\boldsymbol{J}_{45} = 0.1$
$\boldsymbol{h}_6 = -0.1$	$\boldsymbol{J}_{26} = 0.1$	$\boldsymbol{J}_{46} = -0.1$
$\boldsymbol{h}_7 = -0.1$	$\boldsymbol{J}_{27} = 0.1$	$\boldsymbol{J}_{47} = -0.1$
$\boldsymbol{h}_8 = 0.1$	$\boldsymbol{J}_{28} = 0.1$	$\boldsymbol{J}_{48} = -0.1$

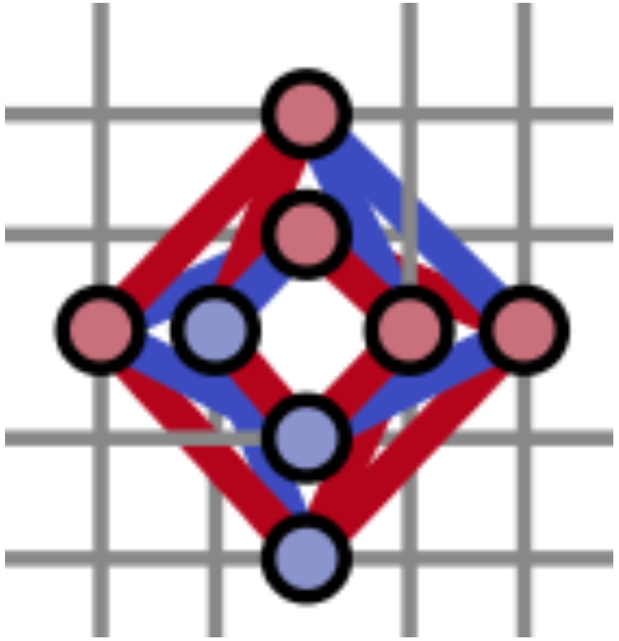


Low Energy Solutions

obj.	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8
-1.2	-1	1	1	-1	1	-1	-1	-1
-1.2	-1	1	-1	-1	1	1	-1	-1
-1.2	-1	-1	-1	1	-1	1	1	-1
-1.0	1	1	1	-1	1	-1	-1	-1
-1.0	-1	1	-1	-1	-1	1	-1	-1
-1.0	-1	1	-1	1	-1	1	1	-1
-1.0	-1	-1	-1	1	-1	1	1	1
-1.0	-1	1	1	-1	1	1	-1	-1
-1.0	1	1	-1	-1	1	-1	1	-1
-1.0	-1	1	-1	-1	1	-1	-1	-1
-1.0	-1	1	-1	-1	1	1	1	-1
-1.0	-1	-1	-1	-1	-1	1	1	-1

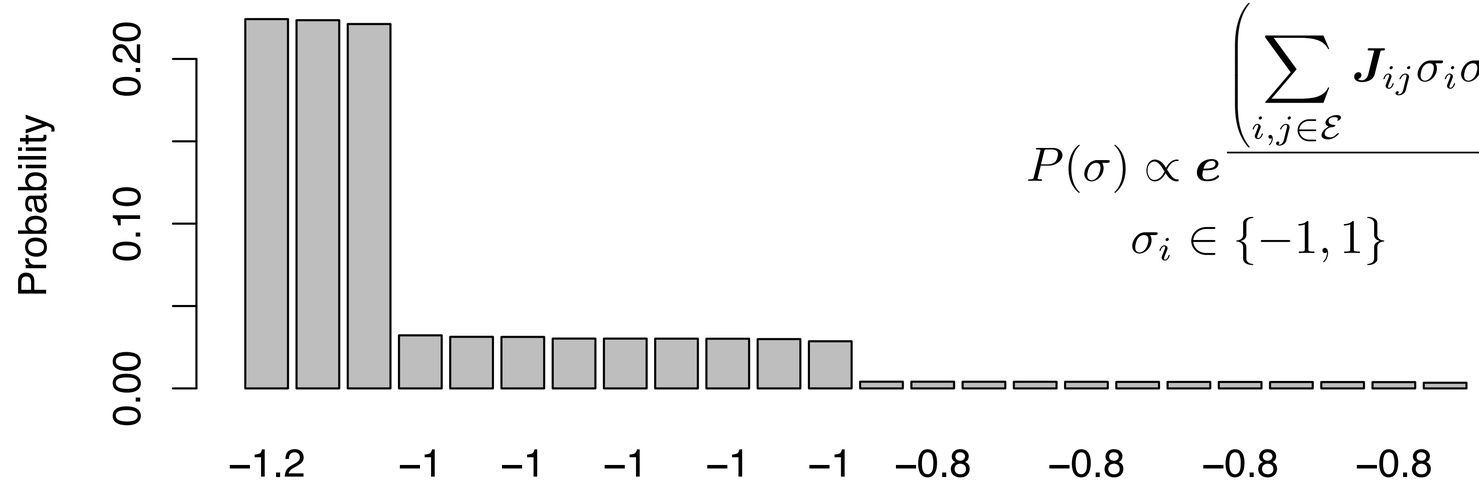
...

Optimal



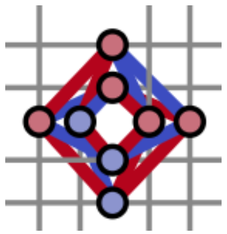
Boltzmann Sampling Comparison

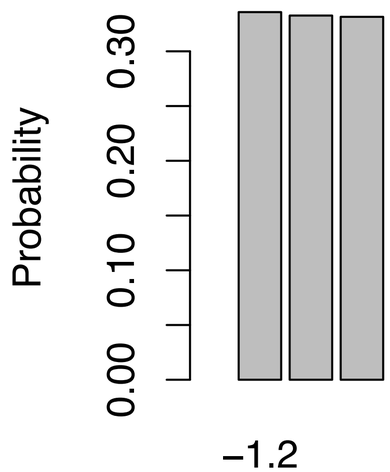
Low Temp.
Boltzmann Sampling
$\tau \approx 0.1$



$$P(\sigma) \propto e^{\frac{\left(\sum_{i,j\in\mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{h}_i\sigma_i\right)}{\tau}}$$

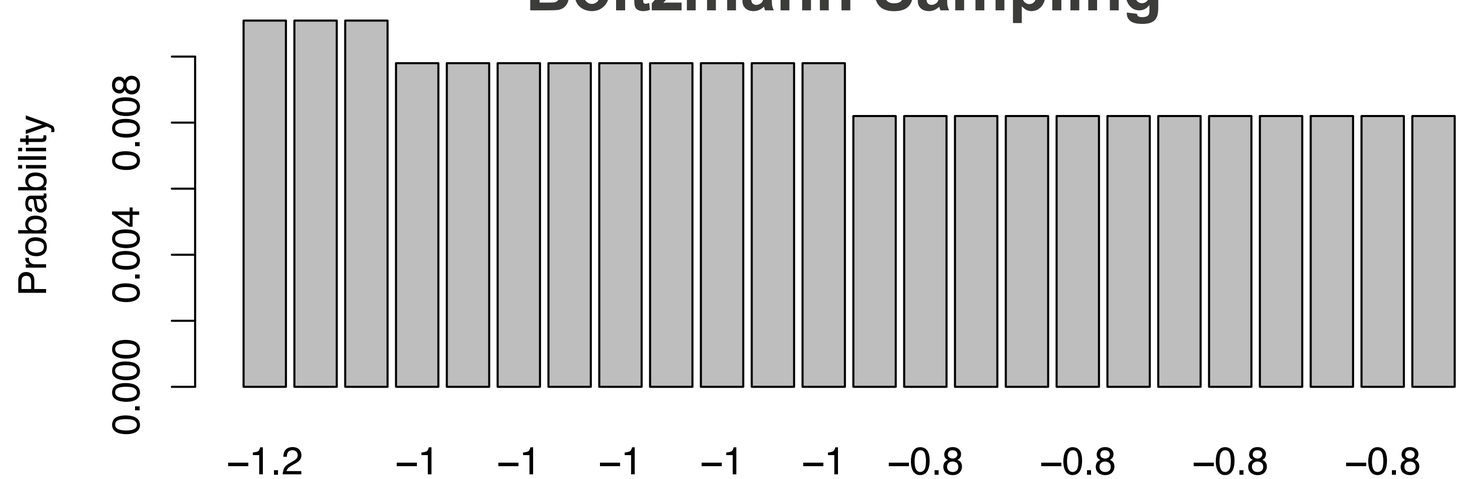
$$\sigma_i \in \{-1, 1\}$$





Zero Temp.
Boltzmann Sampling
$\tau \approx 0$

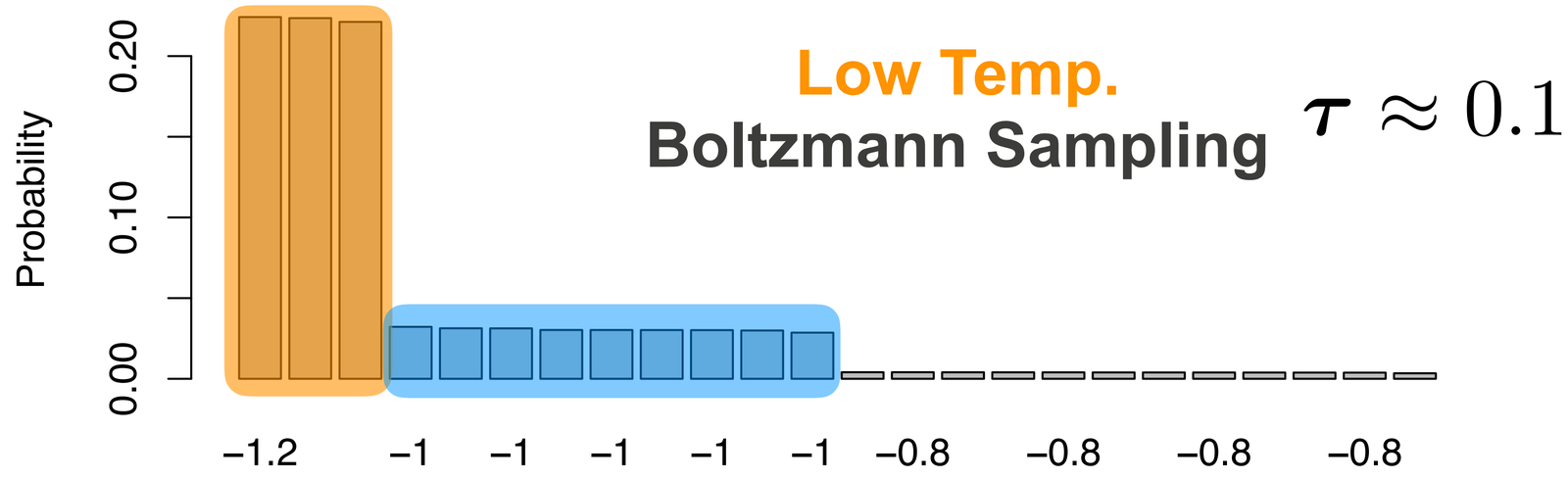
High Temp.
Boltzmann Sampling
$\tau \approx 1.0$



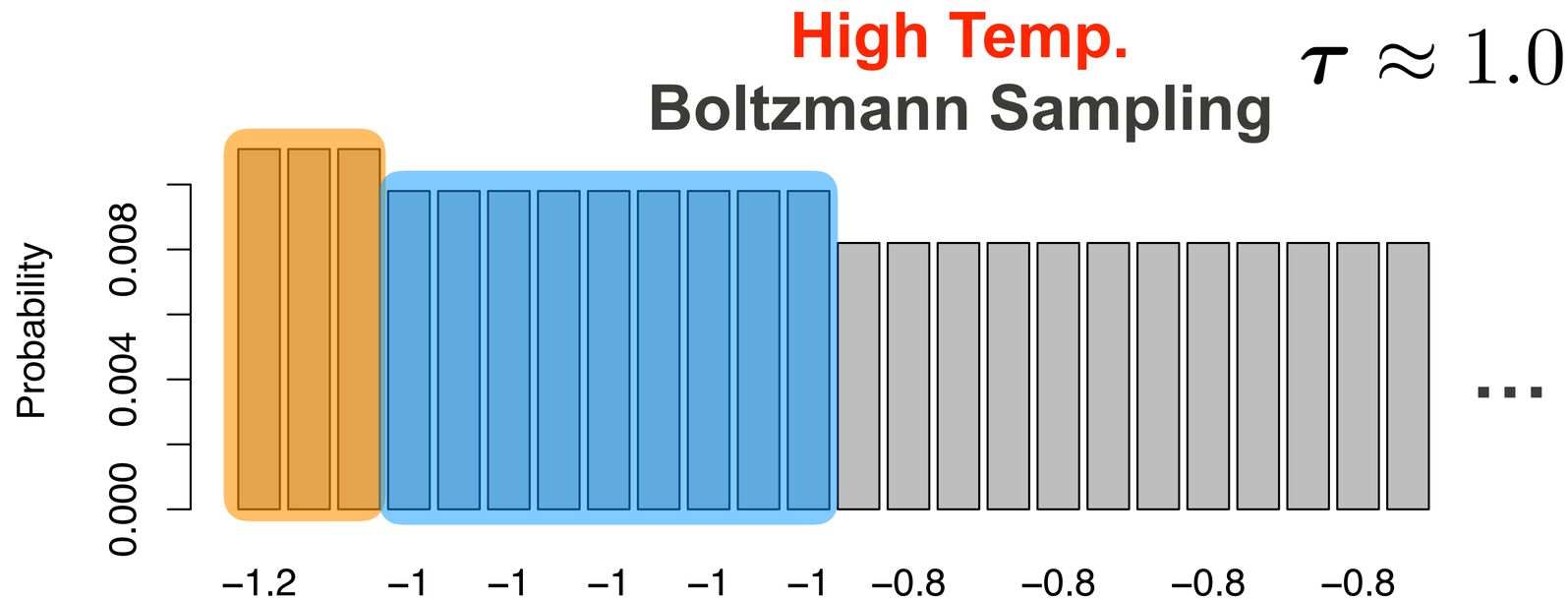
...

The Problem for Optimization

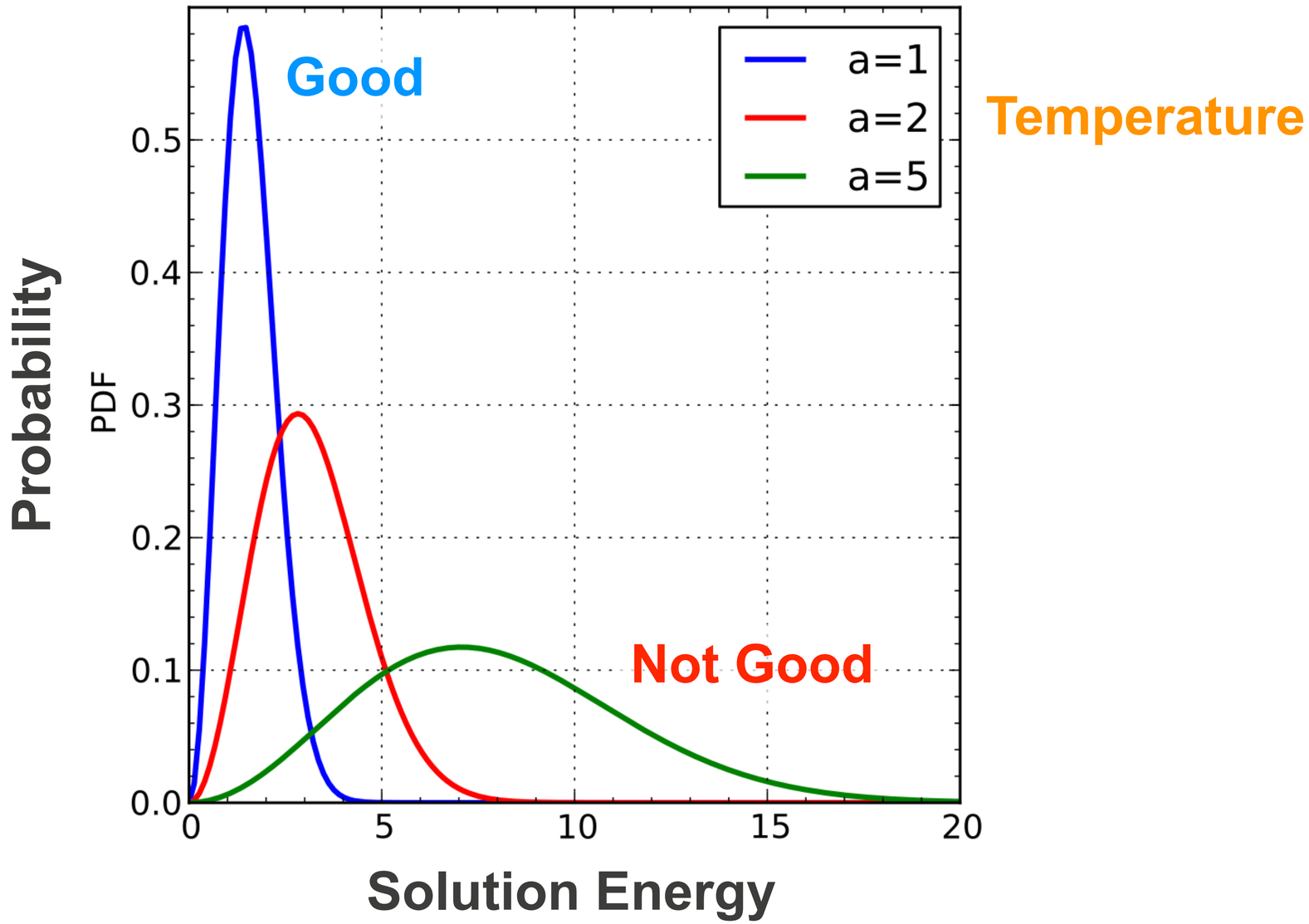
τ	$P(\sigma^*)$	99% Confidence
0.0	1.00	1
0.1	0.67	5
1.0	0.03	155



- **Benchmarking Challenges**
- **All states have non-zero probability**
- **Will always see OPT, with sufficient samples…**



The Problem for Optimization



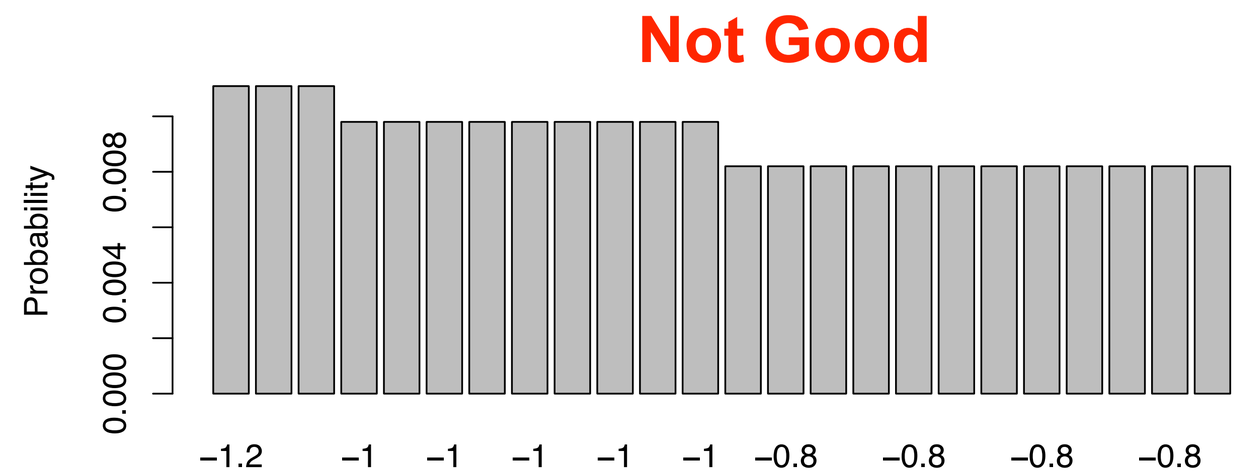
https://upload.wikimedia.org/wikipedia/commons/thumb/1/19/Maxwell-Boltzmann_distribution_pdf.svg/1049px-Maxwell-Boltzmann_distribution_pdf

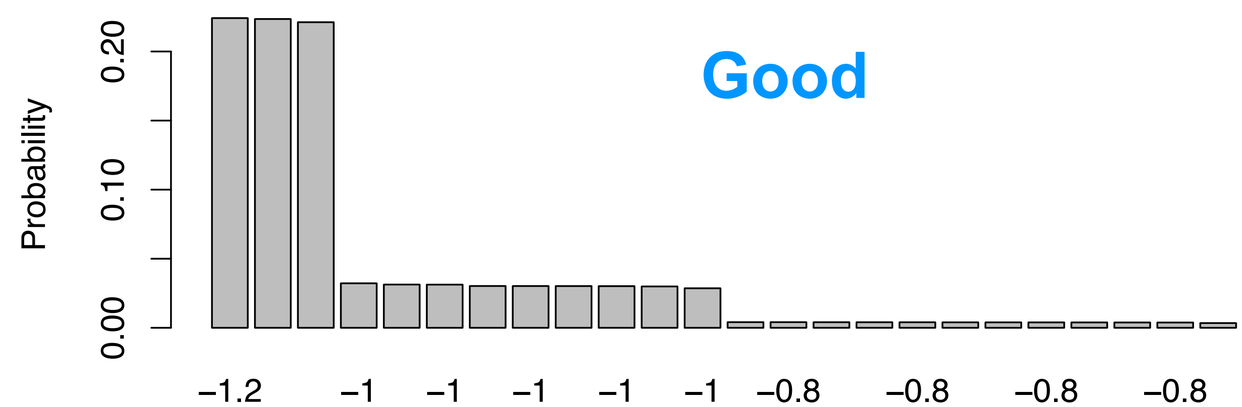
Goals for Optimization

- **Worst case scenario**
  - Boltzmann Sampler devolves into an i.i.d. sampler from all states
- **Mitigations?**
  - lower temperature (i.e. tau)
  - increase coefficients (i.e. J,h)

$$P(\sigma) \propto e^{\frac{\left(\sum_{i,j\in\mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{h}_i\sigma_i\right)}{\tau}}$$

$$\sigma_i \in \{-1, 1\}$$





D-Wave Constraints

- **Tau is fixed to 0.1**

- **Make J/h as large as possible**
  - But limited!

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

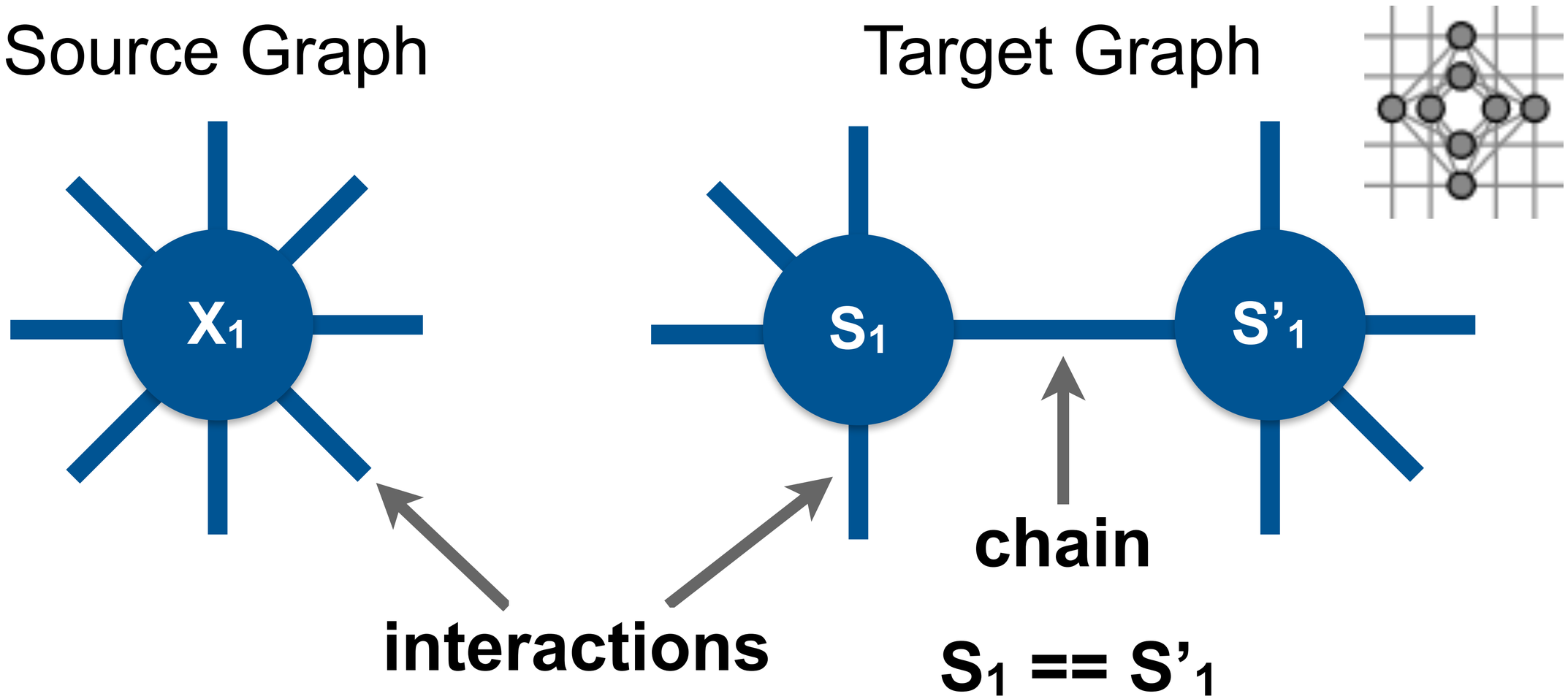
- **By default D-Wave "auto scales"**
  - Scales J/h to largest possible values

- **Max samples per job 10,000**
  - My goal, see OPT w.h.p. in 10,000 samples

$$P(\sigma) \propto \boldsymbol{e}^{\frac{\left(\sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i\right)}{0.1}}$$
$$\sigma_i \in \{-1, 1\}$$

Mathematics of Chains

What are Chains?



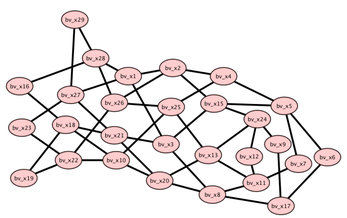
A Mathematical Perspective

Logical Problem

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i\sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$



Lifted Problem

$$\min : \sum_{i,j \in \hat{\mathcal{E}}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i \in \hat{\mathcal{N}}} \boldsymbol{h}_i\sigma_i$$

s.t.

$$\sigma_i = \sigma_j \; \forall i, j \in \mathcal{L}$$

$$\sigma_i \in \{-1, 1\} \; \forall i \in \hat{\mathcal{N}}$$

$$\hat{\mathcal{E}} \subseteq \mathcal{C}_{12}$$

?

Lagrangian Relaxation

$$\min : \sum_{i,j \in \mathcal{L}} (\sigma_i - \sigma_j)^2$$

$$\sigma_i = \sigma_j \; \forall i, j \in \mathcal{L}$$

$$\sigma_i - \sigma_j = 0 \; \forall i, j \in \mathcal{L}$$

$$(\sigma_i - \sigma_j)^2 = 0 \; \forall i, j \in \mathcal{L}$$

$$\sigma_i = \sigma_j \; \forall i, j \in \mathcal{L}$$

$$\Downarrow$$

$$\min : \sum_{i,j \in \mathcal{L}} -\boldsymbol{\lambda}_{ij}\sigma_i\sigma_j$$

Mathematical Properties

- **Lambda >= 0**

- **If lambda is sufficiently large, optimal solutions will match**

$$\boldsymbol{\lambda}_{ij} \geq \sum_{k,l \in \mathcal{E}} |\boldsymbol{J}_{kl}| + \sum_{k \in \mathcal{N}} |\boldsymbol{h}_k| \; \forall i, j \in \mathcal{L}$$

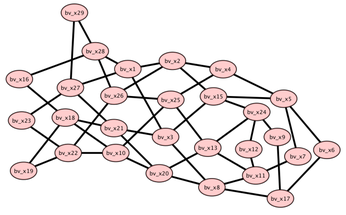
- **Finding the smallest possible setting of lambda can be NP-Hard**
  - Weak bounds, heuristics and iterative algorithms usually work well in practice

Logical Problem

$$\min : \sum_{i,j \in \mathcal{E}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \mathcal{N}} \boldsymbol{h}_i \sigma_i$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \mathcal{N}$$



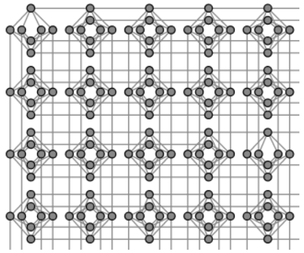
Unconstrained Lifted Problem

$$\min : \sum_{i,j \in \hat{\mathcal{E}}} \boldsymbol{J}_{ij} \sigma_i \sigma_j + \sum_{i \in \hat{\mathcal{N}}} \boldsymbol{h}_i \sigma_i - \sum_{i,j \in \mathcal{L}} \boldsymbol{\lambda}_{ij} \sigma_i \sigma_j$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \hat{\mathcal{N}}$$

$$\hat{\mathcal{E}} \cup \mathcal{L} \subseteq \mathcal{C}_{12}$$



An Illustrative Example

Logical Problem

$$\min : 10\sigma_1 - 20\sigma_2$$

s.t.

$$\sigma_1 = \sigma_2$$

$$\sigma_i \in \{-1, 1\} \; \forall i \in \{1, 2\}$$

Objective	σ_1	σ_2
10	-1	-1
inf.	1	-1
inf.	-1	1
-10	1	1

Unconstrained Problem

$$\min : 10\sigma_1 - 20\sigma_2 - \boldsymbol{\lambda}_{12}\sigma_1\sigma_2$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \{1, 2\}$$

$$\boldsymbol{\lambda}_{12}?$$

An Illustrative Example

$$\min : 10\sigma_1 - 20\sigma_2 - \boldsymbol{\lambda}_{12}\sigma_1\sigma_2$$

s.t.

$$\sigma_i \in \{-1, 1\} \; \forall i \in \{1, 2\}$$

$\boldsymbol{\lambda}_{12}$	Objective	σ_1^*	σ_2^*
0	-30	-1	1
1	-29	-1	1
5	-25	-1	1
11	-21	1	1
30	-40	1	1

$$\boldsymbol{\lambda}_{ij} \geq \sum_{k,l \in \mathcal{E}} |\boldsymbol{J}_{kl}| + \sum_{k \in \mathcal{N}} |\boldsymbol{h}_k| \; \forall i, j \in \mathcal{L}$$

$$\boldsymbol{\lambda}_{12} \geq 30$$

Smallest Lambda?

Mathematical Correctness Requires

$$\boldsymbol{\lambda}_{12} \geq 10 + \boldsymbol{\epsilon}$$

D-Wave Constraints

- **J/h range is limited**

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

- **Want smallest possible Lambda**
  - To avoid rescaling

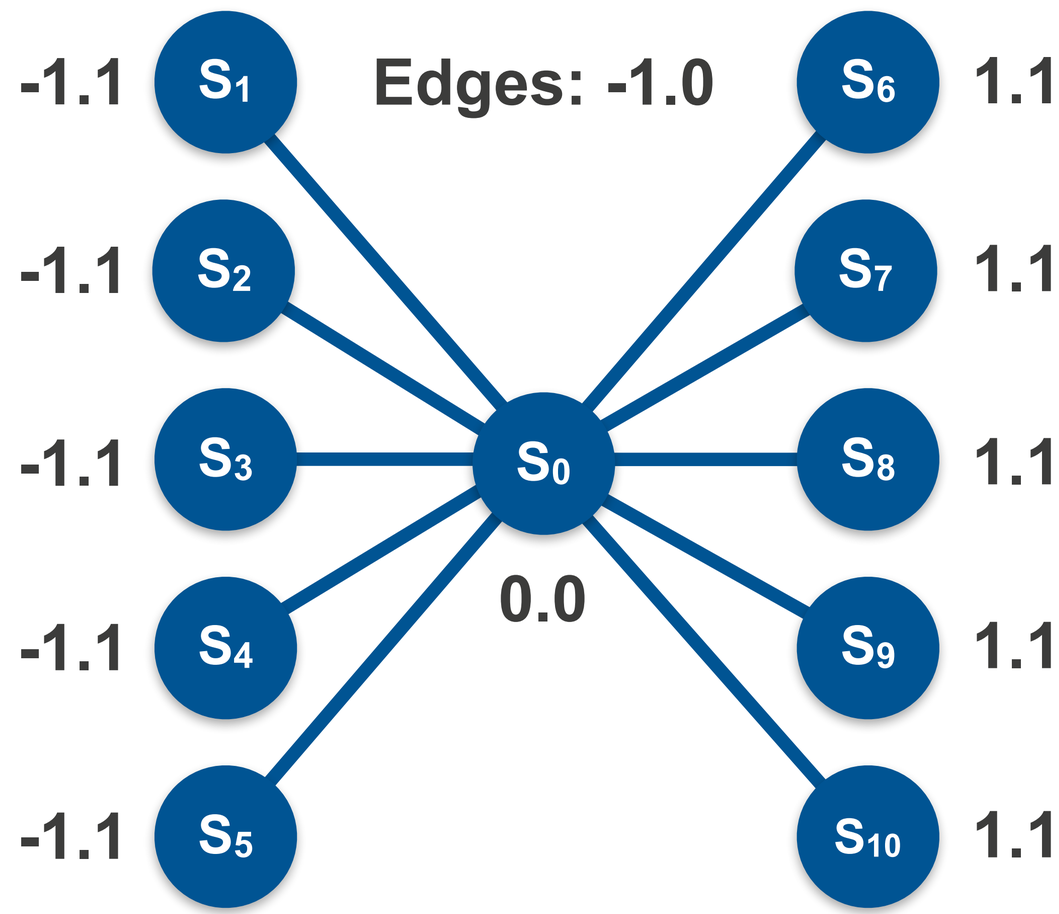
Two Problematic Examples

The Indecisive Ising Model

(Chaining)

Indecisive Ising Model

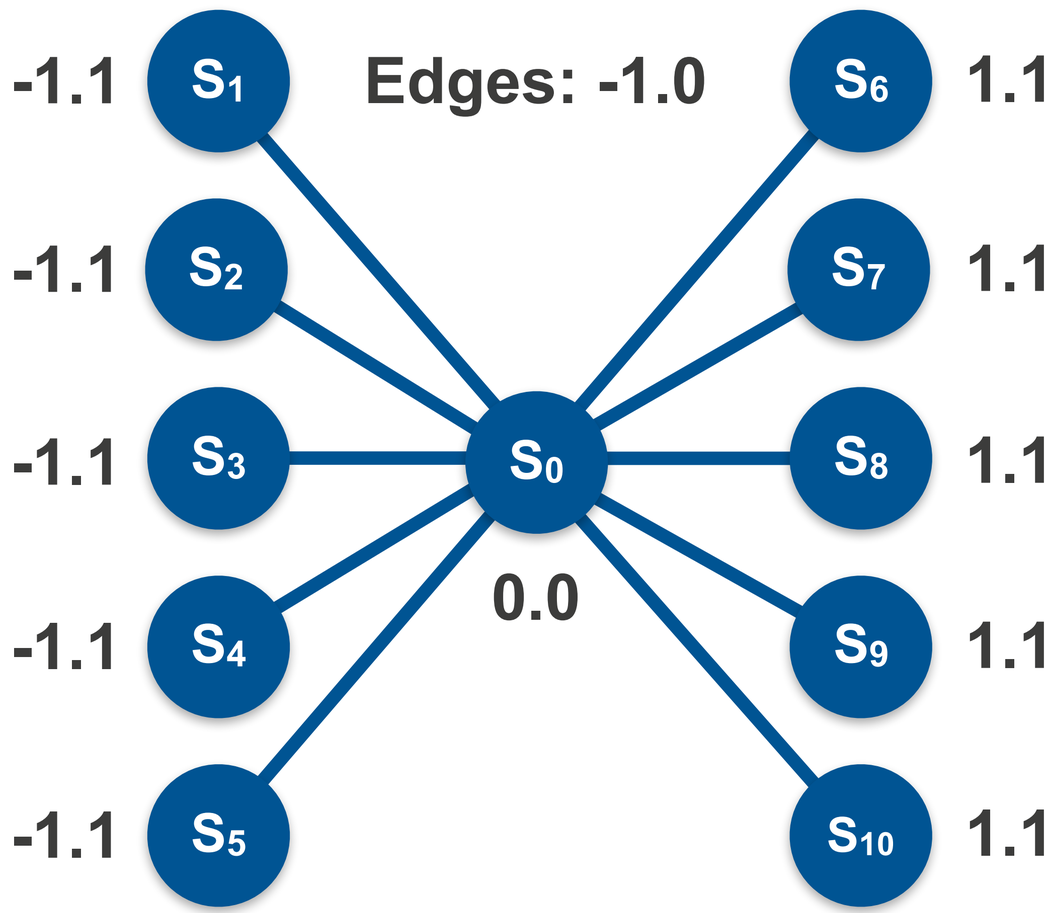
- -1/1 variable space
- Left leaves want to take 1 value
- Right leaves want to take -1 value
- The hub wants to match the side values
- Does not matter what value hub takes
  - two symmetric optimal solutions



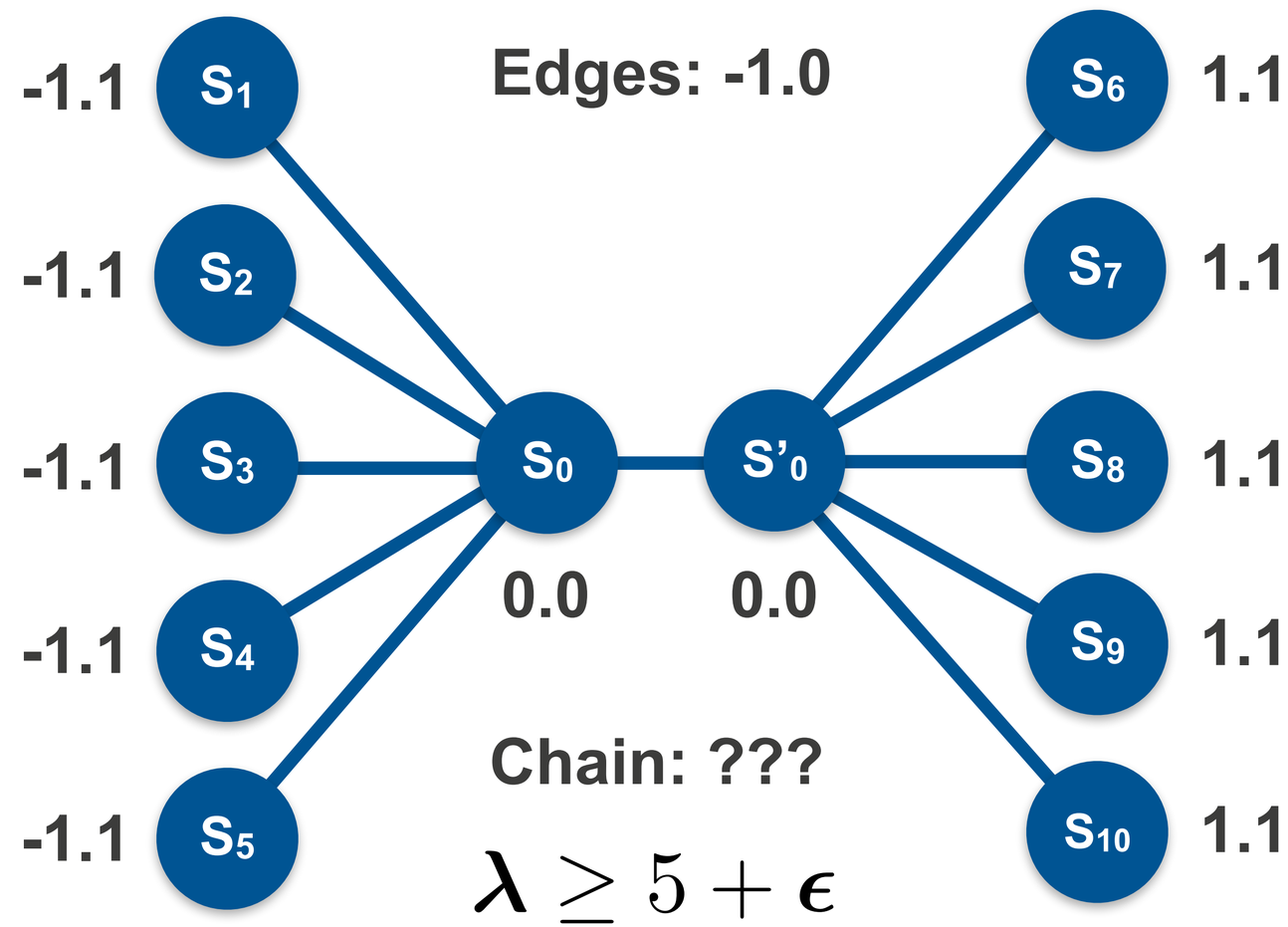
Indecisive Ising Model

- **D-Wave implementation requires a chain!**

Logical Model



Embeddable Model

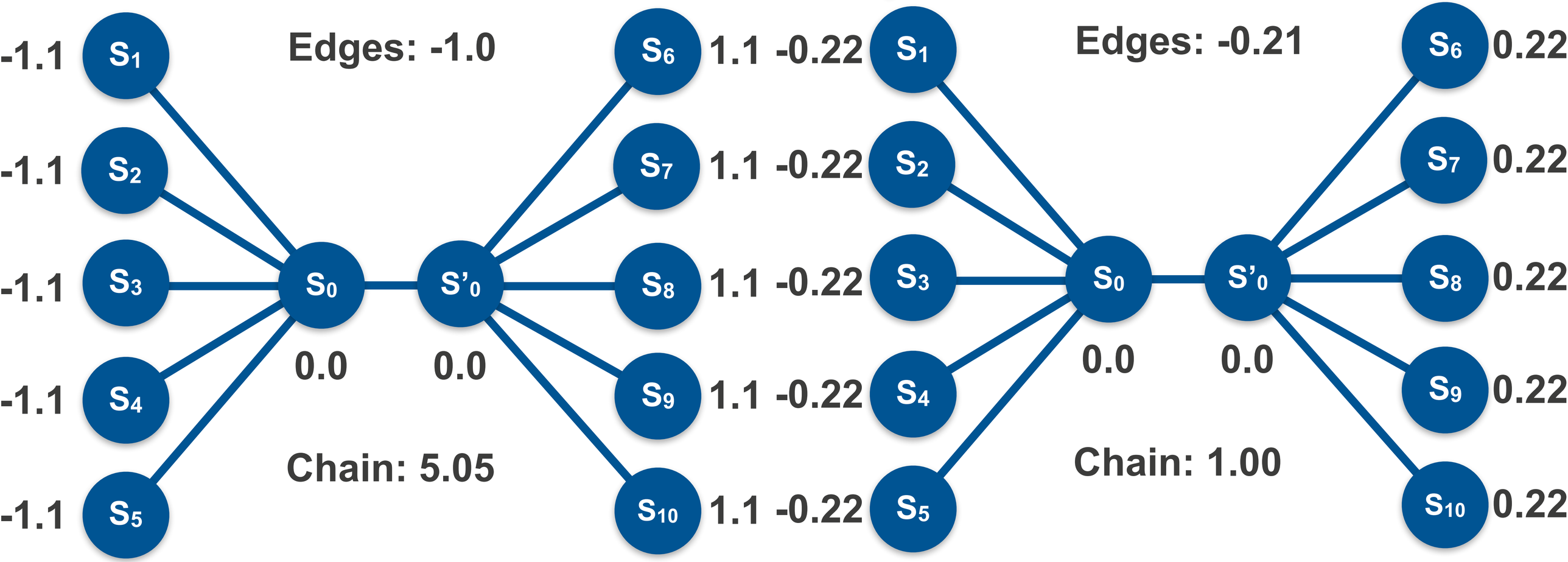


$$\lambda \geq 5 + \epsilon$$

Indecisive Ising Model

- **D-Wave implementation requires rescaling!**

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

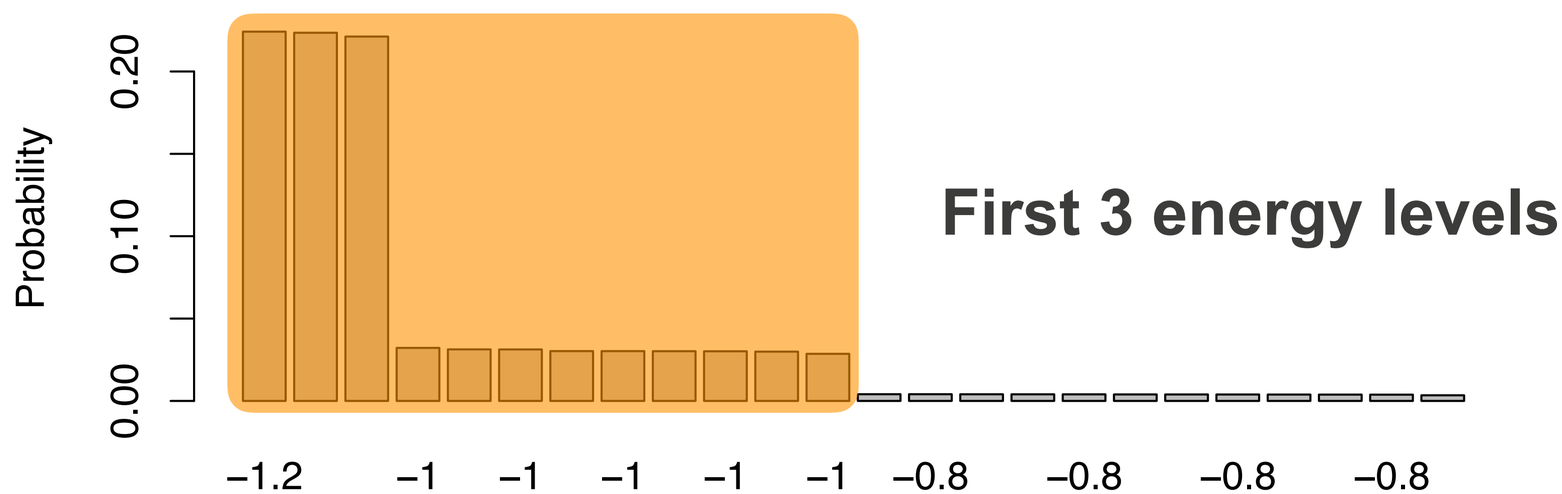


Indecisive Ising Model

Idealized Boltzmann Sampler

$$\tau \approx 0.1$$

Ising **Ising Rescaled** **D-Wave**



Indecisive Ising Model

Idealized Boltzmann Sampler

$$\boldsymbol{\tau} \approx 0.1$$

Ising

Probability	$f(\sigma)$	feas.
0.2415	-16.05	✓
0.2415	-16.05	✓
0.0888	-15.95	x
0.0327	-15.85	✓

Ising Rescaled

Probability	$f(\sigma)$	feas.
0.0369	-3.178	✓
0.0369	-3.178	✓
0.0303	-3.158	x
0.0248	-3.139	✓

D-Wave

Probability	$f(\sigma)$	feas.
0.0288	-3.178	✓
0.0270	-3.178	✓
0.0157	-3.158	x
0.0225	-3.139	✓
...	...	...
0.0196	-3.139	✓

So What?

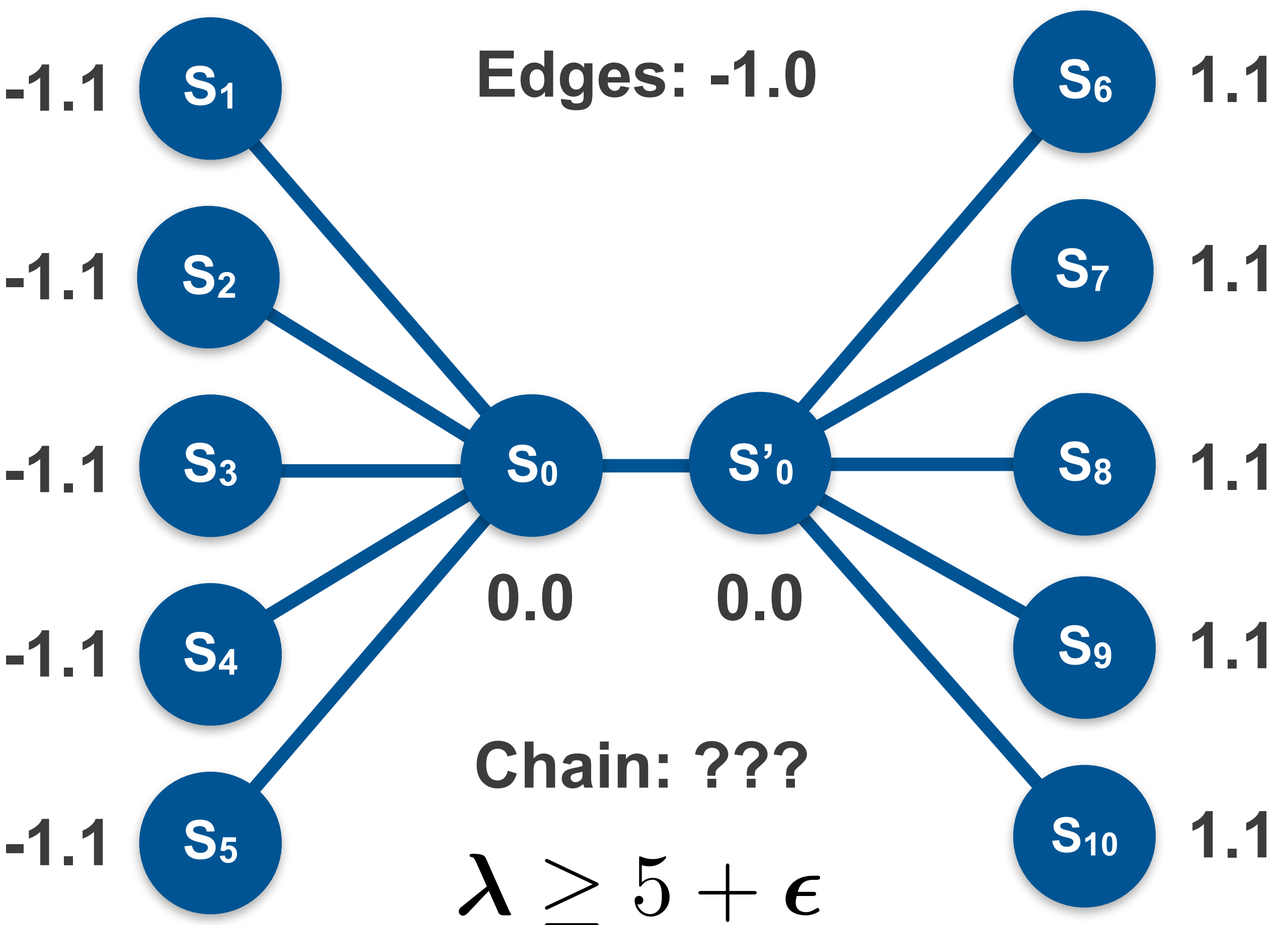
- **A 0.05 to 0.07 probability of seeing OPT seems pretty good…**
  - only need about 100 samples to see OPT w.h.p. (below 10,000 goal)

- **The problem is very small (i.e. n=12)**
  - targeting problems with 100-1000's

- **As connectivity grows, so will this problem**
  - Lets consider doubling the connectivity (i.e. 20 edges)

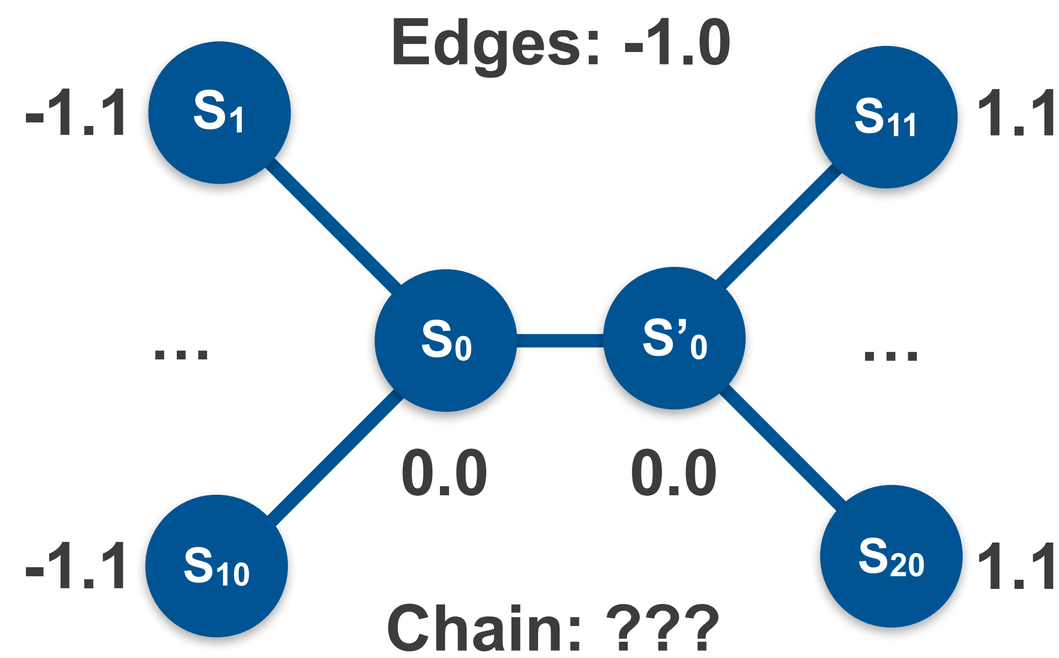
Indecisive Ising Model

2000Q



$$\lambda \geq 5 + \epsilon$$

Possible Future



$$\lambda \geq 10 + \epsilon$$

Indecisive Ising Model

Idealized Boltzmann Sampler

$$\boldsymbol{\tau} \approx 0.1$$

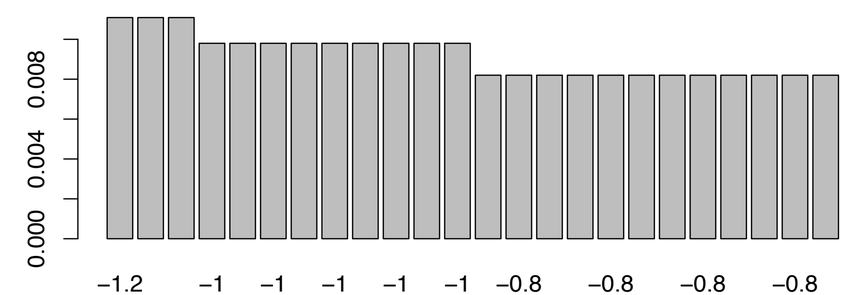
Ising

Probability	$f(\sigma)$	feas.
0.1336	-32.05	✓
0.1336	-32.05	✓
0.0492	-31.95	x
0.0181	-31.85	✓

Ising Rescaled

Probability	$f(\sigma)$	feas.
0.0011	-3.189	✓
0.0011	-3.189	✓
0.0010	-3.179	x
0.0009	-3.169	✓

Forced High Temp.
Boltzmann Sampling

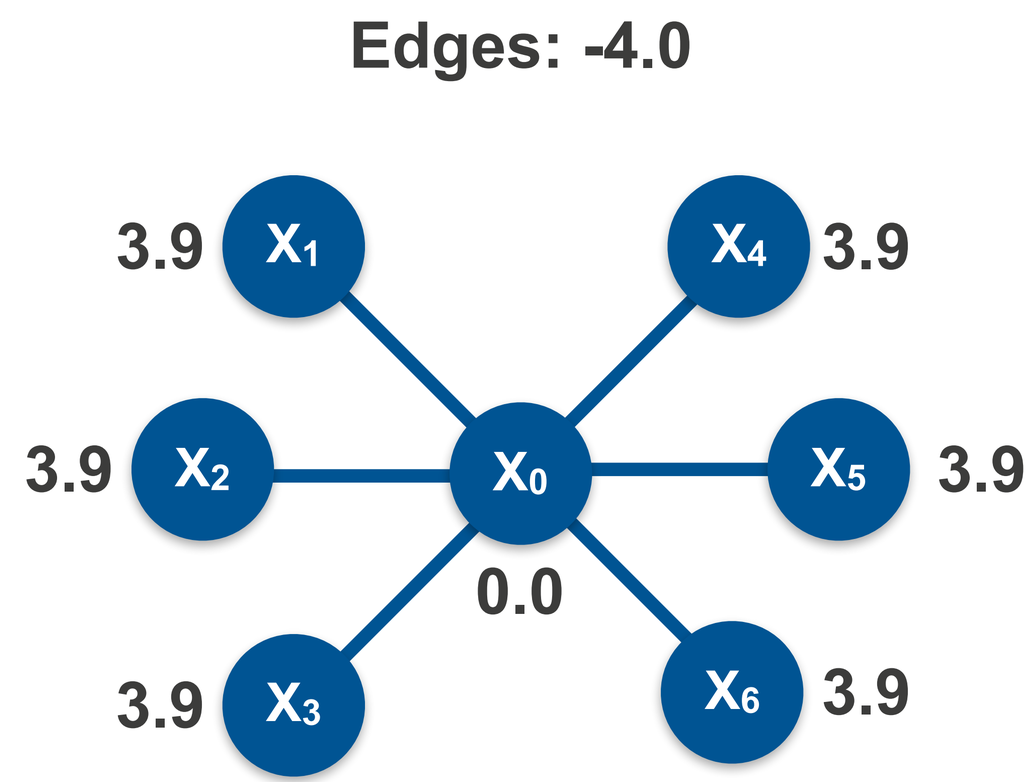


The Consensus QUBO Model

(Rescaling)

Consensus QUBO Model

- 0/1 variable space
- Leaves want to take 0 value
- Hub can take any value
- Edges want all nodes to be 1
- Edges win in OPT, just barely



Consensus QUBO in Ising Space

- **D-Wave implementation requires an Ising model!**

$$x_i = \frac{\sigma_i + 1}{2}$$

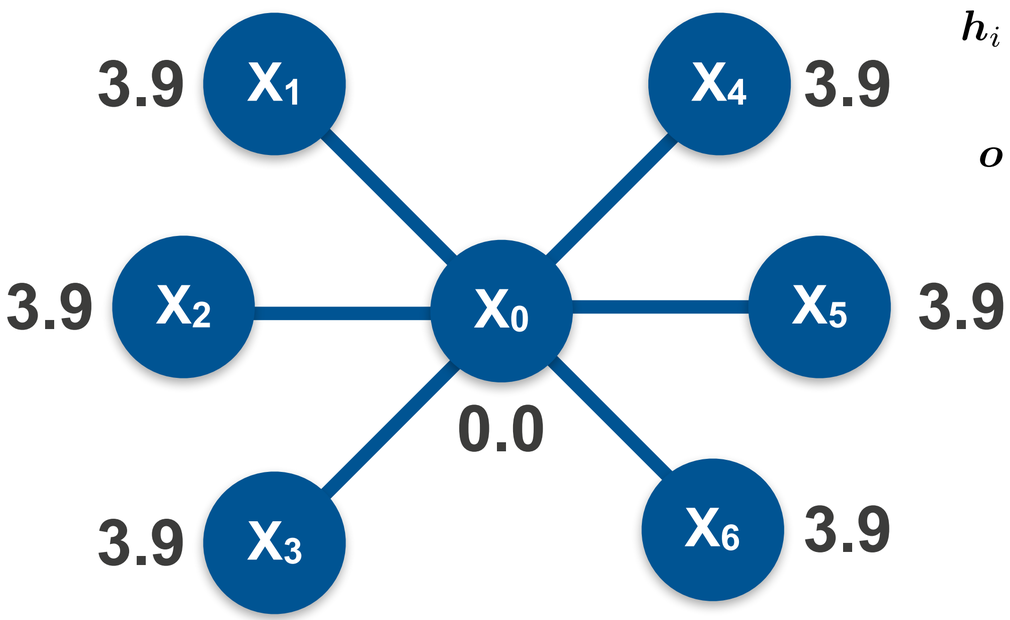
$$\boldsymbol{J}_{ij} = \frac{\boldsymbol{c}_{ij}}{4} \; \forall i, j \in \mathcal{E}$$

$$\boldsymbol{h}_i = \frac{\boldsymbol{c}_i}{2} + \sum_{i,j \in \mathcal{E}_i} \frac{\boldsymbol{c}_{ij}}{4} \; \forall i \in \mathcal{N}$$

$$\boldsymbol{o} = \sum_{i \in \mathcal{N}} \frac{\boldsymbol{c}_i}{2} + \sum_{i,j \in \mathcal{E}} \frac{\boldsymbol{c}_{ij}}{4}$$

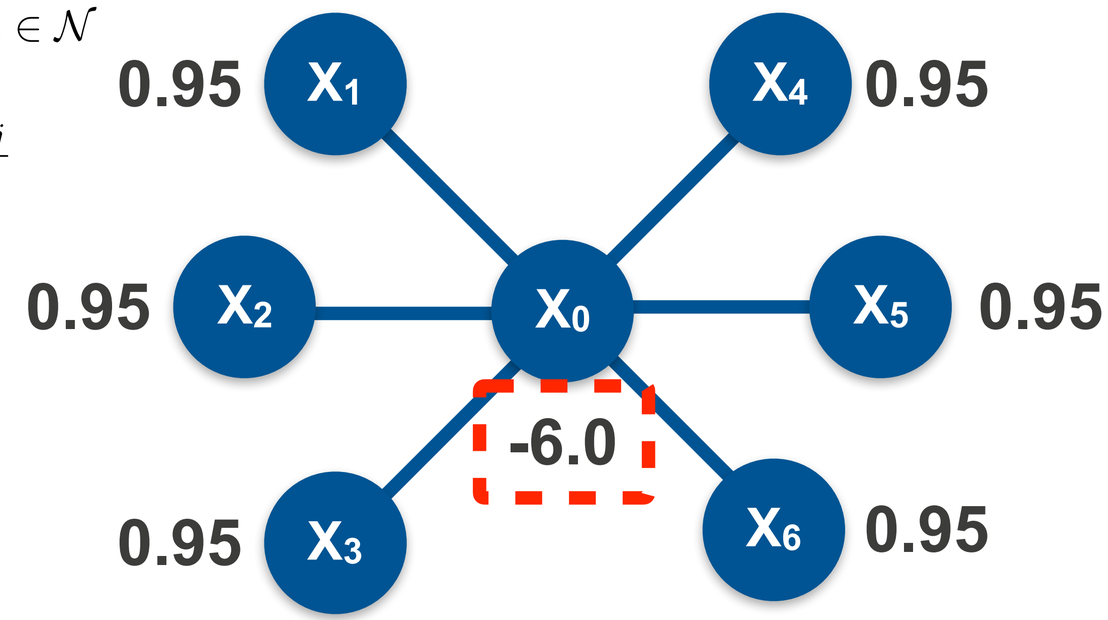
QUBO

Edges: -4.0



Ising Space

Edges: -1.0



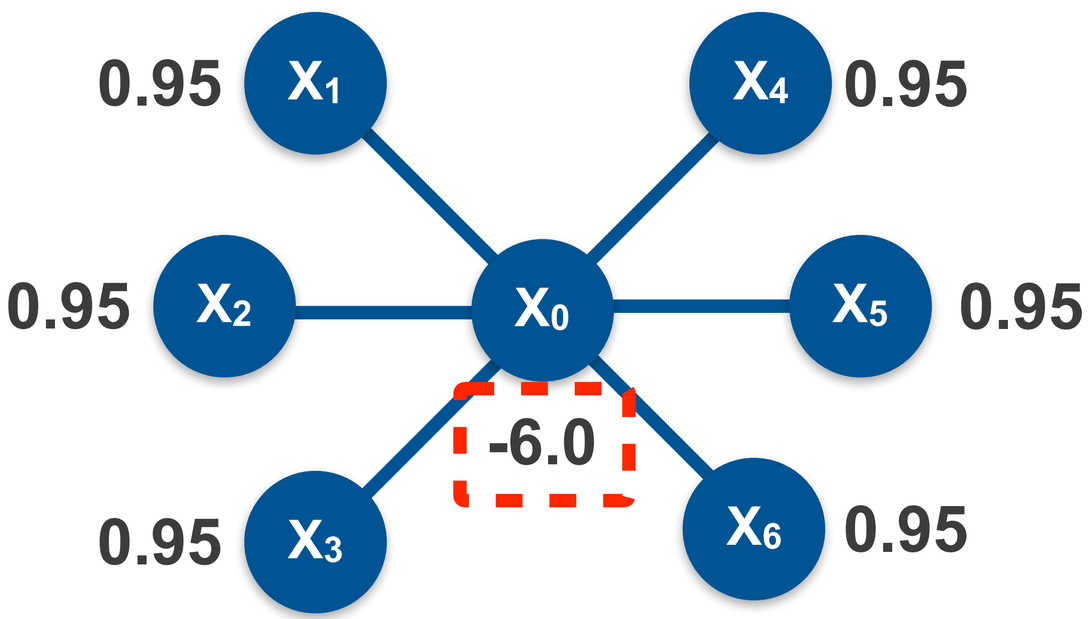
Consensus QUBO in Ising Space

- **D-Wave implementation requires rescaling!**

$$-1 \leq \boldsymbol{J}_{ij} \leq 1 \;\; \forall (i,j) \in \mathcal{E}$$
$$-2 \leq \boldsymbol{h}_i \leq 2 \;\; \forall i \in \mathcal{N}$$

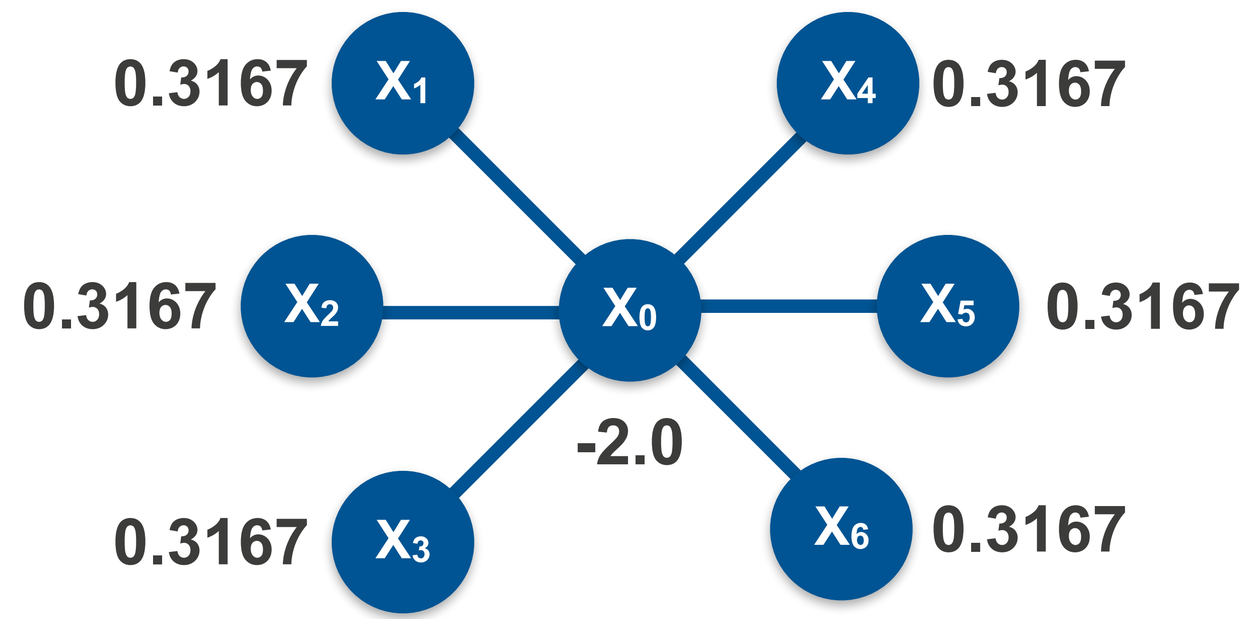
Ising

Edges: -1.0



Ising Rescaled

Edges: -0.3333

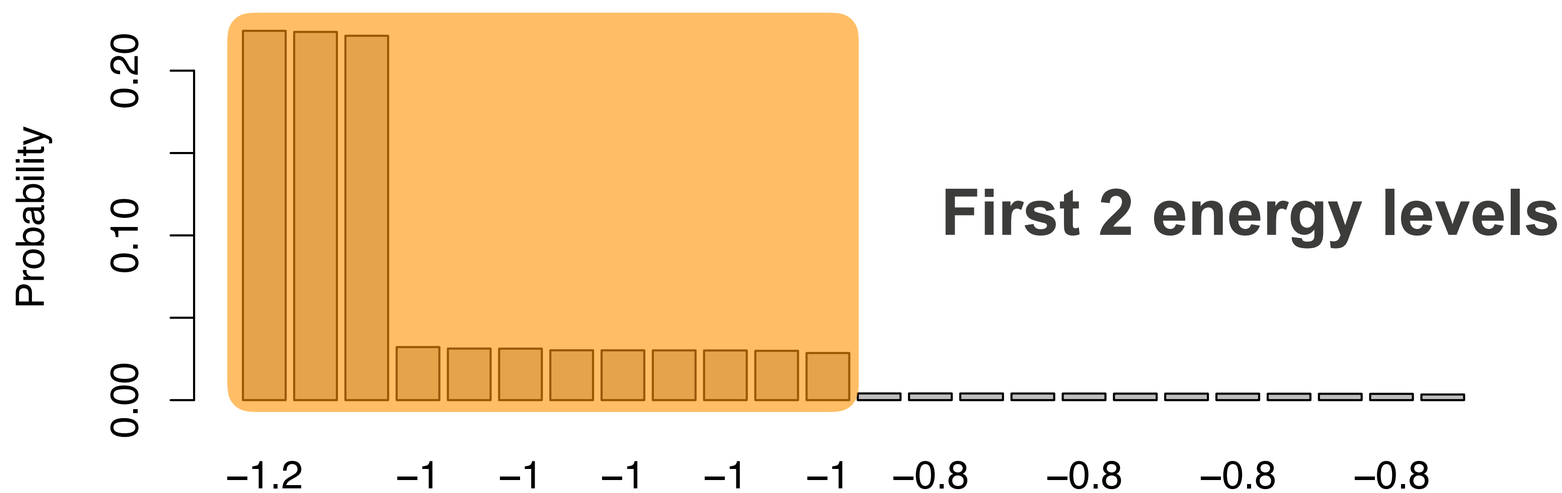


Consensus QUBO in Ising Space

Idealized Boltzmann Sampler

$$\boldsymbol{\tau} \approx 0.1$$

Ising **Ising Rescaled** **D-Wave**



Consensus QUBO in Ising Space

Idealized Boltzmann Sampler

$$\boldsymbol{\tau} \approx 0.1$$

Ising

Probability	$f(\sigma)$
0.153	-6.3
0.056	-6.2
0.056	-6.2
0.056	-6.2
0.056	-6.2
0.056	-6.2
0.056	-6.2

Ising Rescaled

Probability	$f(\sigma)$
0.039	-2.100
0.028	-2.067
0.028	-2.067
0.028	-2.067
0.028	-2.067
0.028	-2.067
0.028	-2.067

D-Wave

Probability	$f(\sigma)$
0.083	-2.100
0.042	-2.067
0.041	-2.067
0.041	-2.067
0.040	-2.067
0.039	-2.067
0.038	-2.067

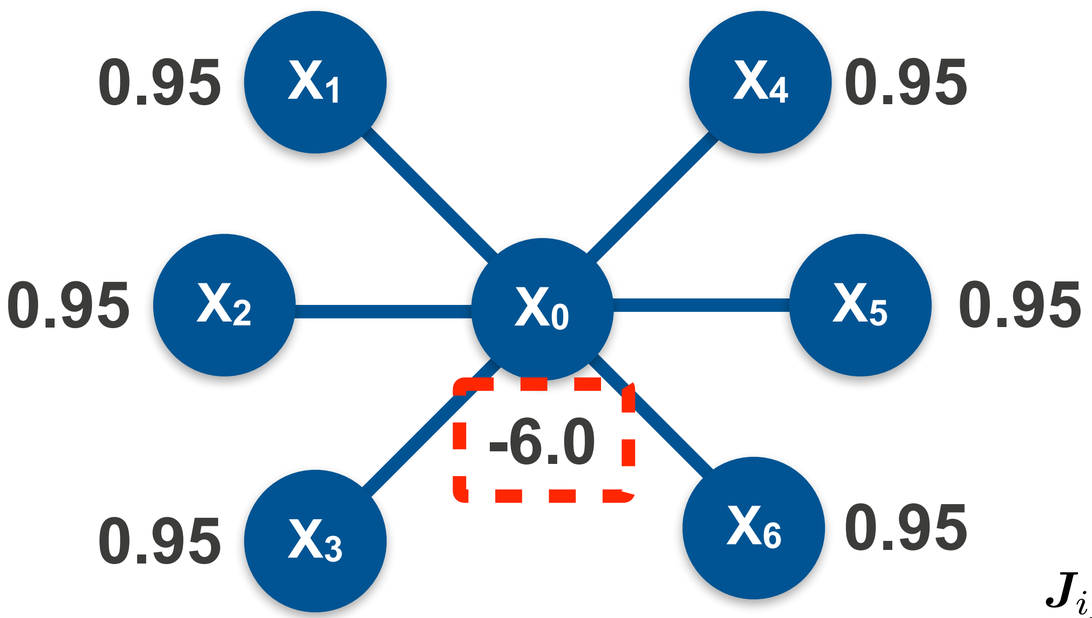
So What?

- **A 0.04 to 0.08 probability of seeing OPT seems pretty good…**
  - only need about 100 samples to see OPT w.h.p. (below 10,000 goal)
- **The problem is very small (i.e. 7), targeting problems with 100-1000's**
- **We want high connectivity graphs**
  - To avoid chains
- **As connectivity grows, so will this problem**
  - Lets consider doubling the connectivity (i.e. 12 edges)

Consensus QUBO in Ising Space

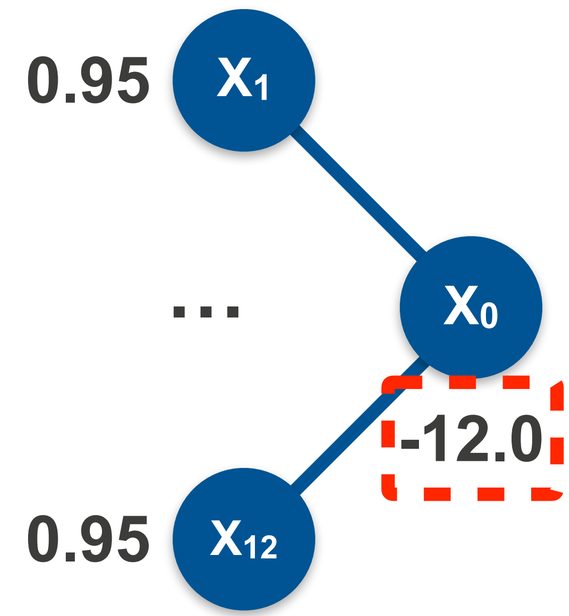
2000Q

Edges: -1.0



Possible Future

Edges: -1.0



$$\boldsymbol{J}_{ij} = \frac{\boldsymbol{c}_{ij}}{4} \; \forall i, j \in \mathcal{E}$$

$$\boldsymbol{h}_i = \frac{\boldsymbol{c}_i}{2} + \boxed{\sum_{i,j \in \mathcal{E}_i} \frac{\boldsymbol{c}_{ij}}{4}} \; \forall i \in \mathcal{N}$$

$$\boldsymbol{o} = \sum_{i \in \mathcal{N}} \frac{\boldsymbol{c}_i}{2} + \sum_{i,j \in \mathcal{E}} \frac{\boldsymbol{c}_{ij}}{4}$$

12 Edge Hub and Spoke QUBO

Idealized Boltzmann Sampler

$$\tau \approx 0.1$$

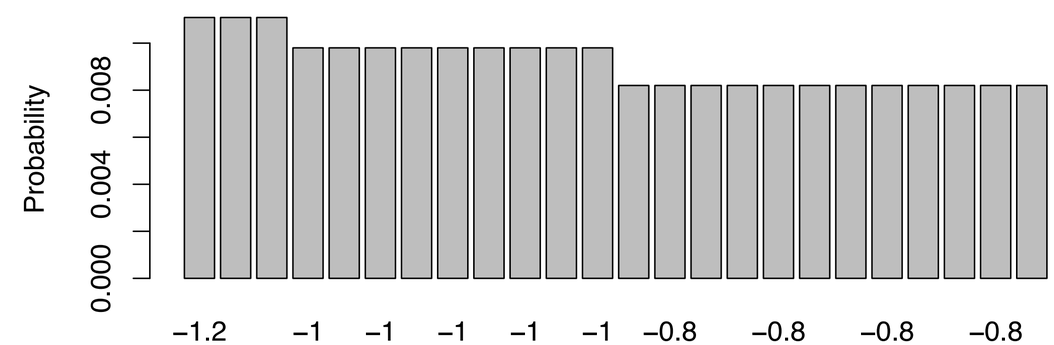
Ising

Probability	$f(\sigma)$
0.0233	-12.600
0.0086	-12.500
...	...

Ising Rescaled

Probability	$f(\sigma)$
0.0006	-2.100
0.0005	-2.083
...	...

Forced High Temp. Boltzmann Sampling

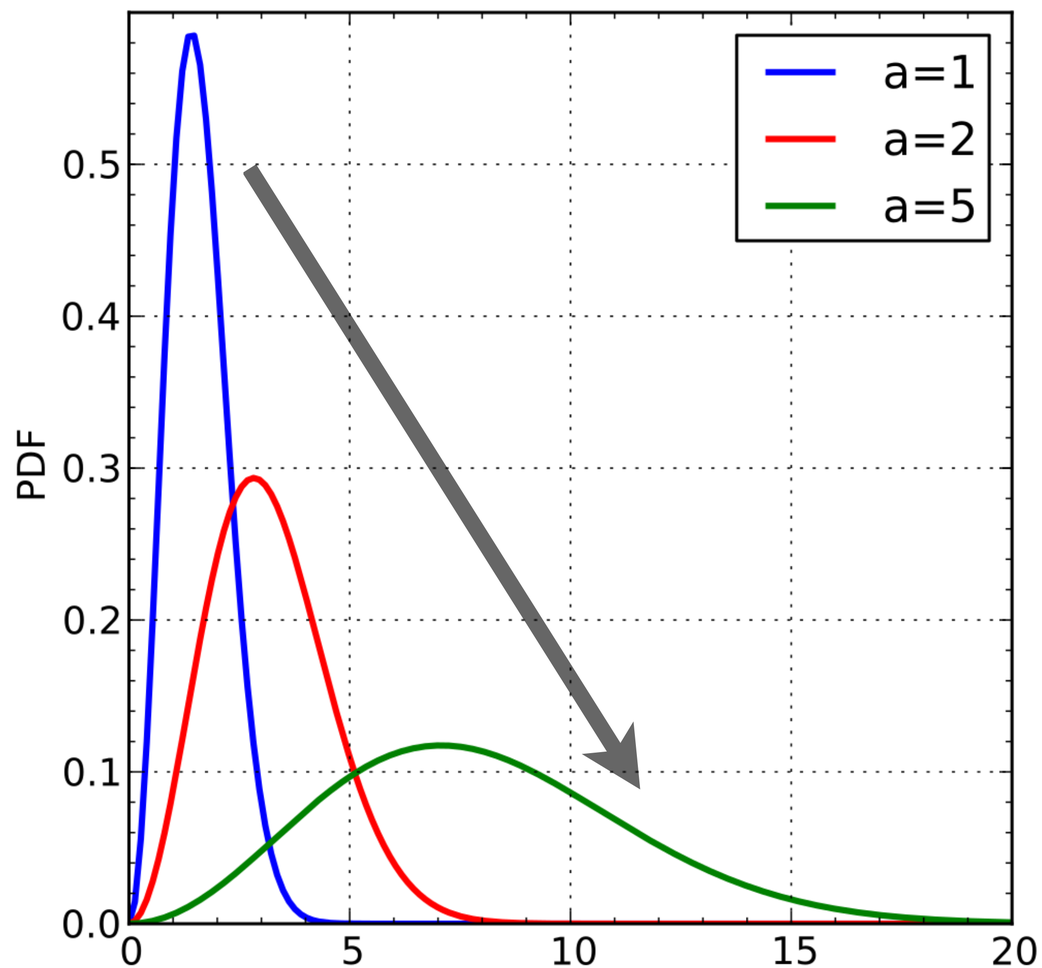


Back to the Original Question

What's Going On?

$$P(\sigma) \propto e^{\frac{\left(\sum_{i,j\in\mathcal{E}} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i\in\mathcal{N}} \boldsymbol{h}_i\sigma_i\right)}{0.1}}$$

$$\sigma_i \in \{-1, 1\}$$



D:wave

dwave				
Best Sol.	Best	Inf.	Samples	Time
-11	9073	1	10000	0+3
-14	8370	85	10000	0+3
-16	5651	123	10000	0+3
-18	3865	316	10000	0+4
-24	16	1254	10000	0+4
-25	22	5465	10000	0+5
-26	1	9855	10000	4+5
F.E.	-	-	-	T.L.
F.E.	-	-	-	T.L.
F.E.	-	-	-	T.L.
F.E.	-	-	-	T.L.
F.E.	-	-	-	T.L.
F.E.	-	-	-	T.L.
F.E.*	-	-	-	T.L.

F.E.

Failed Embed

T.L.

Time Limit (1 hour)

What's Going On?

- **For a consistent energy gap, the effective temperature should decrease as the problem size increases**

Temperature scaling law for quantum annealing optimizers

Tameem Albash,[1,2] Victor Martin-Mayor,[3,4] and Itay Hen[1,2]

[1] *Information Sciences Institute, University of Southern California, Marina del Rey, California 90292, USA*
[2] *Department of Physics and Astronomy and Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*
[3] *Departamento de Física Teórica I, Universidad Complutense, 28040 Madrid, Spain*
[4] *Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Zaragoza, Spain*

Physical implementations of quantum annealing unavoidably operate at finite temperatures. We point to a fundamental limitation of fixed finite temperature quantum annealers that prevents them from functioning as competitive scalable optimizers and show that to serve as optimizers annealer temperatures must be appropriately scaled down with problem size. We derive a temperature scaling law dictating that temperature must drop at the very least in a logarithmic manner but also possibly as a power law with problem size. We corroborate our results by experiment and simulations and discuss the implications of these to practical annealers.

Introduction.— Quantum computing devices are becoming sufficiently large to undertake computational tasks that are infeasible using classical computing [1–7]. The theoretical underpinning for whether such tasks exist with physically realizable quantum annealers remains lacking, despite the excitement brought on by recent technological breakthroughs that have made programmable quantum annealing (QA) [8–12] optimizers

tion, or final Hamiltonian H, they are to solve. The adiabatic theorem of quantum mechanics ensures that the overlap of the final state of the system with the ground state manifold of H, approaches unity as the duration of the process increases [31, 32]. For physical quantum annealers that operate at positive temperatures ($T > 0$), there is no equivalent guarantee of reaching the ground state with high probability. For long runtimes, an ideal

Is it time to give up?

NO!

Don’t Get Me Wrong

- **Big D-Wave Advocate**
  - Every experiment I have run on *well suited* inputs has had very good results
  - See https://arxiv.org/abs/1707.00355

- **My hunch, it is possible to show notable performance gains NOW!**
  - Just need to be carful on problem selection (most likely very contrived)
  - And have a clever D-Wave encoding

- **It will be easy to show supremacy on next generation**
  - But still have to mitigate the issues presented here

- **Guidelines and Suggestions**

Some Mitigations

- **Avoid problems requiring chains**
  - i.e. subgraph of chimera graph

- **Avoid working in QUBO space**
  - Or check that the coefficient ranges in re-scaled Ising space are reasonable

- **Avoid needing multiple coefficient levels**
  - It would be ideal if the problem only requires -1,0,1 / -2,-1,0,1,2

- **Max-Cut is an interesting problem**
  - Meets above criteria
  - APX-Hard (i.e. no PTAS)

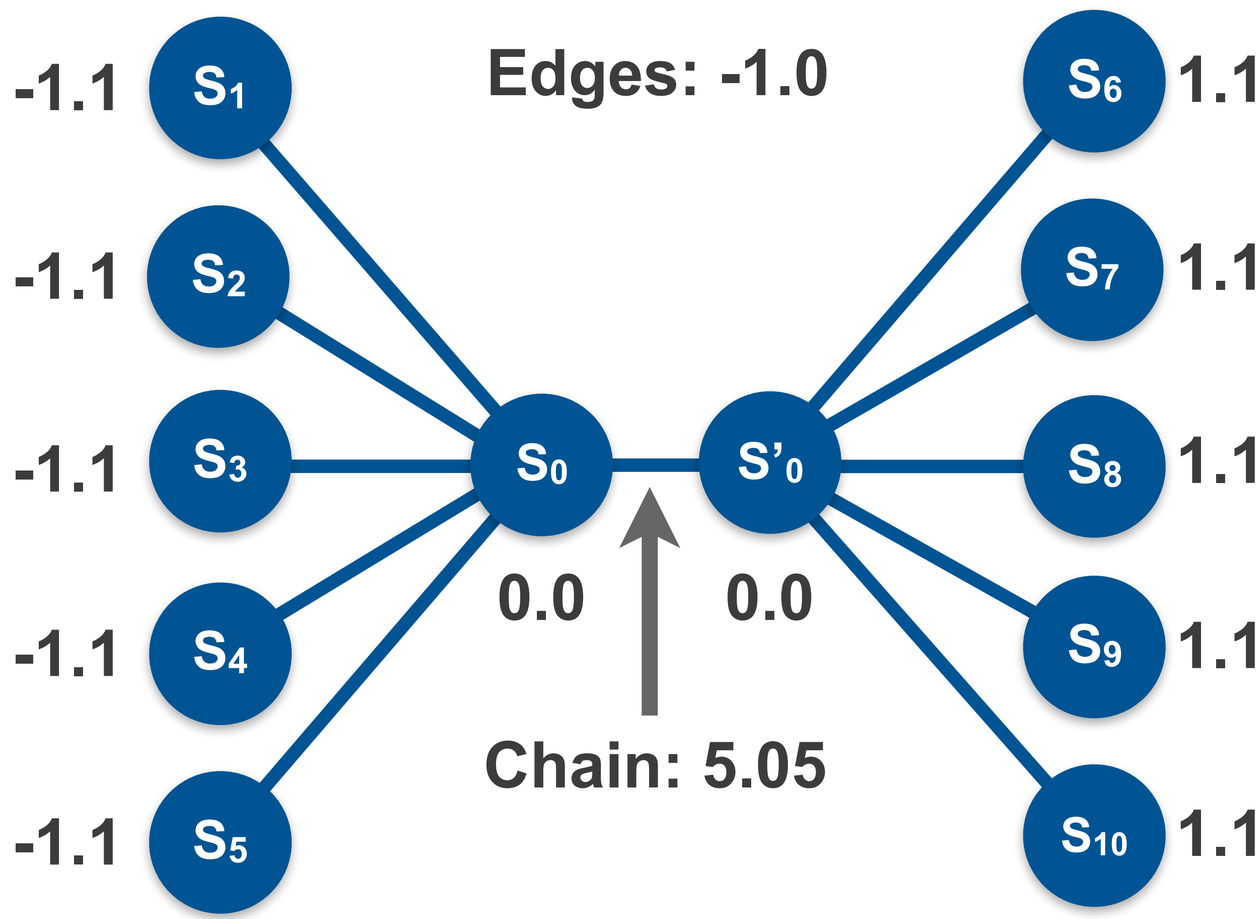
Some Mitigations

- **What if “your favorite problem” does not have that structure?**
  - Classical post-processing (VFYC)
  - Design a hybrid algorithm that extracts sub-tasks with the proposed structure

- **Extended J range / Annealing Offsets**
  - DW_2000Q_LANL Extended J range -4.0 to 1.0
    - Possibly mitigates rescaling

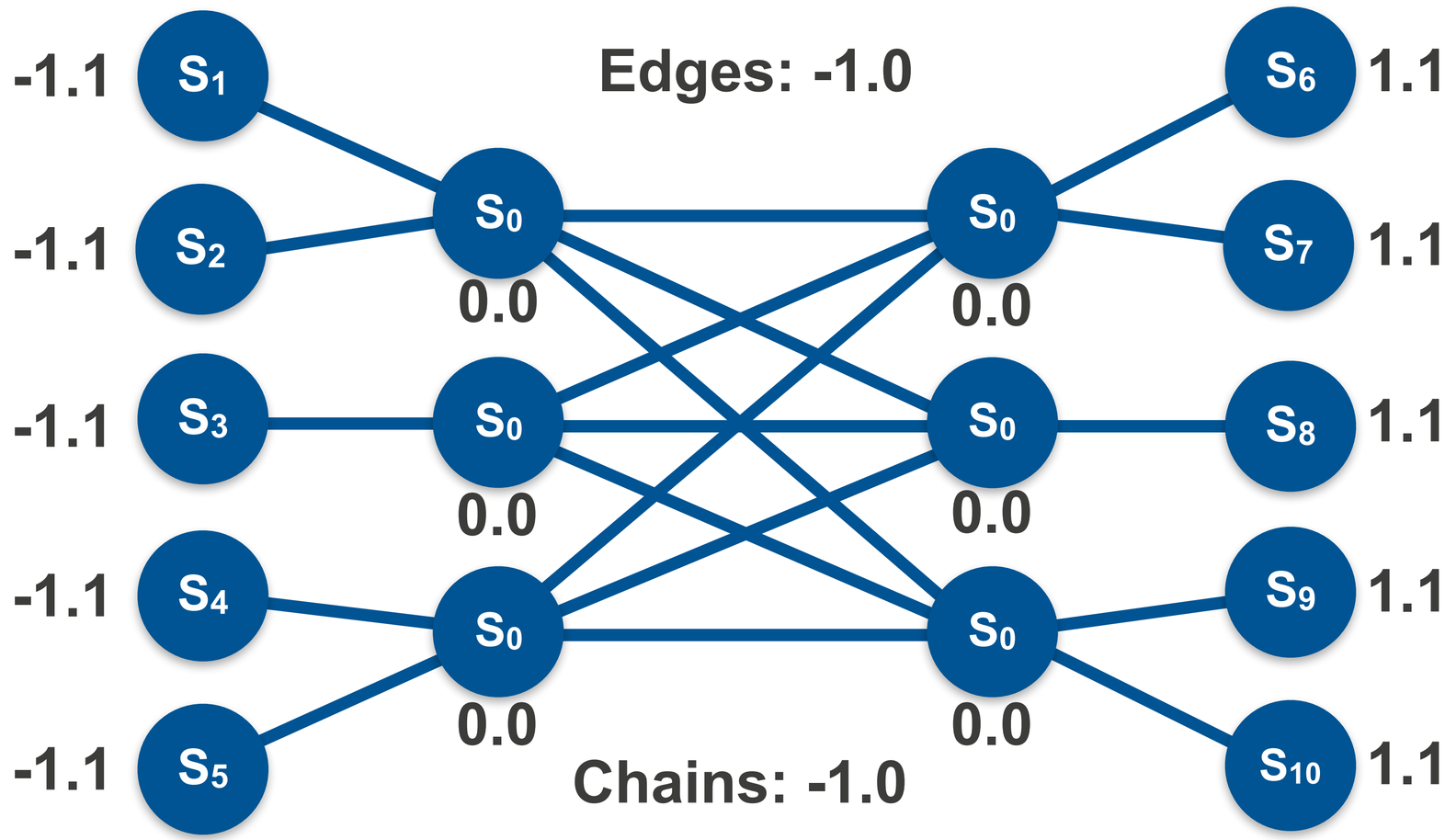
- **Much more clever problem encodings**
  - Lift to much higher dimensions

Indecisive Ising Model



Indecisive Ising Model Revisited

- An alternative mathematically correct encoding



No Rescaling Required!!!

Indecisive Ising Model Revisited

$\tau \approx 0.1$

Ising

Probability	$f(\sigma)$	feas.
0.2415	-16.05	✓
0.2415	-16.05	✓
0.0888	-15.95	x
0.0327	-15.85	✓

Probability	$f(\sigma)$	feas.
0.2651	-20.00	✓
0.2651	-20.00	✓
0.0359	-19.80	✓
0.0359	-19.80	✓

Ising Rescaled

Original Implementation

Probability	$f(\sigma)$	feas.
0.0369	-3.178	✓
0.0369	-3.178	✓
0.0303	-3.158	x
0.0248	-3.139	✓

Spin Cluster Implementation

Probability	$f(\sigma)$	feas.
0.2651	-20.00	✓
0.2651	-20.00	✓
0.0359	-19.80	✓
0.0359	-19.80	✓

D-Wave

Probability	$f(\sigma)$	feas.
0.0288	-3.178	✓
0.0270	-3.178	✓
0.0157	-3.158	x
0.0225	-3.139	✓
...	...	...
0.0196	-3.139	✓

???

Probability	$f(\sigma)$	feas.
0.1291	-20.00	✓
0.1105	-20.00	✓
0.0554	-19.80	✓
...	...	...
0.0298	-19.80	✓

Conclusions

- **The D-Wave's hardware constraints present significant challenges for discrete optimization**
  - Fixed coefficient range and effective temperature (0.1)
  - Sparse hardware graph

- **Finding problems that are well suited to the hardware is challenging**

- **Very clever encodings are likely required to find optimal solutions w.h.p.**

- **Another possibility, consider problems beyond discrete optimization**
  - sampling applications
  - stochastic optimization

Thanks!

Questions?