Sample-Optimal Learning of Graphical Models (and opening the D-Wave annealer's quantum box)

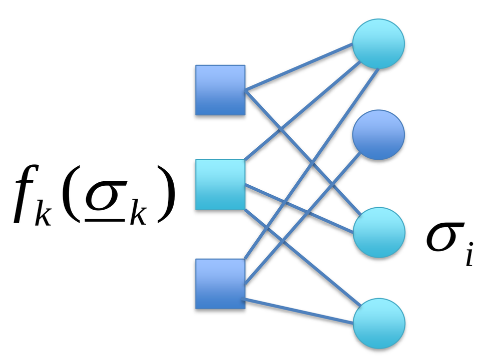
Los Alamos NATIONAL LABORATORY EST. 1943

Andrey Lokhov

with M. Chertkov, C. Coffrin, Y. Kharkov, S. Misra, M. Vuffray

Theoretical Division, Los Alamos National Laboratory

Learning General Markov Random Fields



Arbitrary alphabet $\sigma_i \in A_i,\ |A_i| \le q$

Arbitrary interaction orders $|\sigma_k| \le L$

Arbitrary basis functions $f_k(\sigma_k)$

$$P(\underline{\sigma}) = \frac{1}{Z}\exp\left(\sum_{k\in K}\theta_k^* f_k(\vec{\sigma}_k)\right) \qquad \alpha = \min\left|\theta_k^*\right|,\ \gamma = \max_i \sum_{k\in K_i}\left|\theta_k^*\right|$$

Example: binary variables and multi-body interactions

$$P(\underline{\sigma}) \propto \exp\left(\sum_i h_i^*\sigma_i + \sum_{ij} J_{ij}^*\sigma_i\sigma_j + \sum_{ijk} J_{ijk}^*\sigma_i\sigma_j\sigma_k + \cdots\right)$$

Learning problem: given M i.i.d. samples $\{\underline{\sigma}^{(1)},\dots,\underline{\sigma}^{(M)}\}$, with probability $1-\delta$ learn the graph of conditional dependencies and estimate couplings $\underline{\theta}^*$

Our Method: Interaction Screening

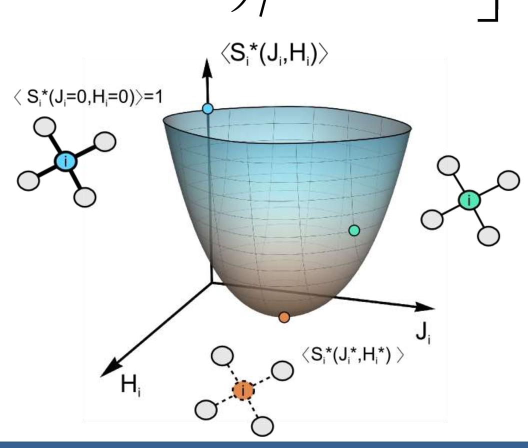
Regularized Interaction Screening Estimator (RISE): convex optimization

$$(\hat{\underline{J}}_i, \hat{h}_i) = \arg\min_{\underline{J}_i, h_i}\left[\left\langle \exp\left(-h_i\sigma_i - \sum_j J_{ij}\sigma_i\sigma_j - \sum_{jk} J_{ijk}\sigma_i\sigma_j\sigma_k - \cdots\right)\right\rangle + \lambda\|\underline{J}_i\|_1\right]$$

✓**Exact**, tractable & parallelizable algorithm

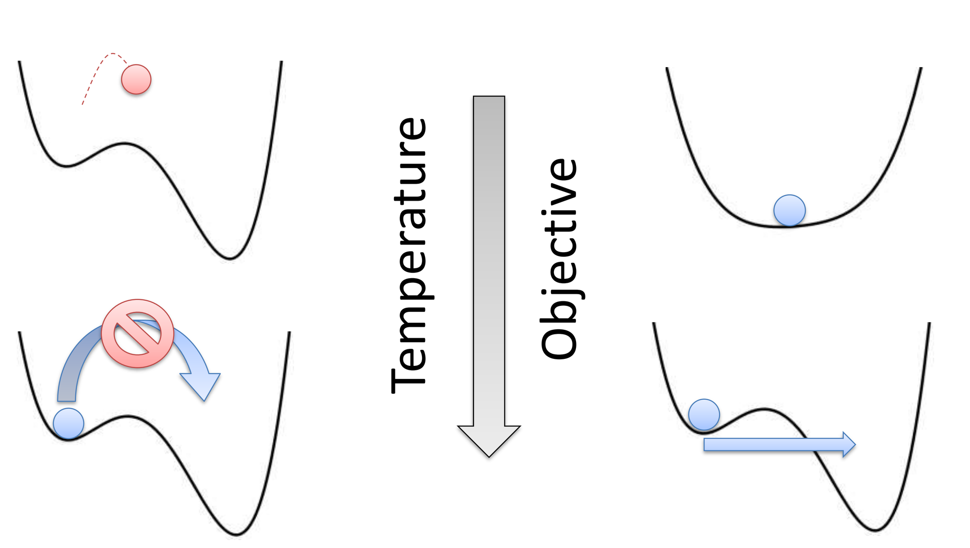
✓Requires near **IT optimal** number of samples

✓Works for **any models**: low temperature, spin glass, …

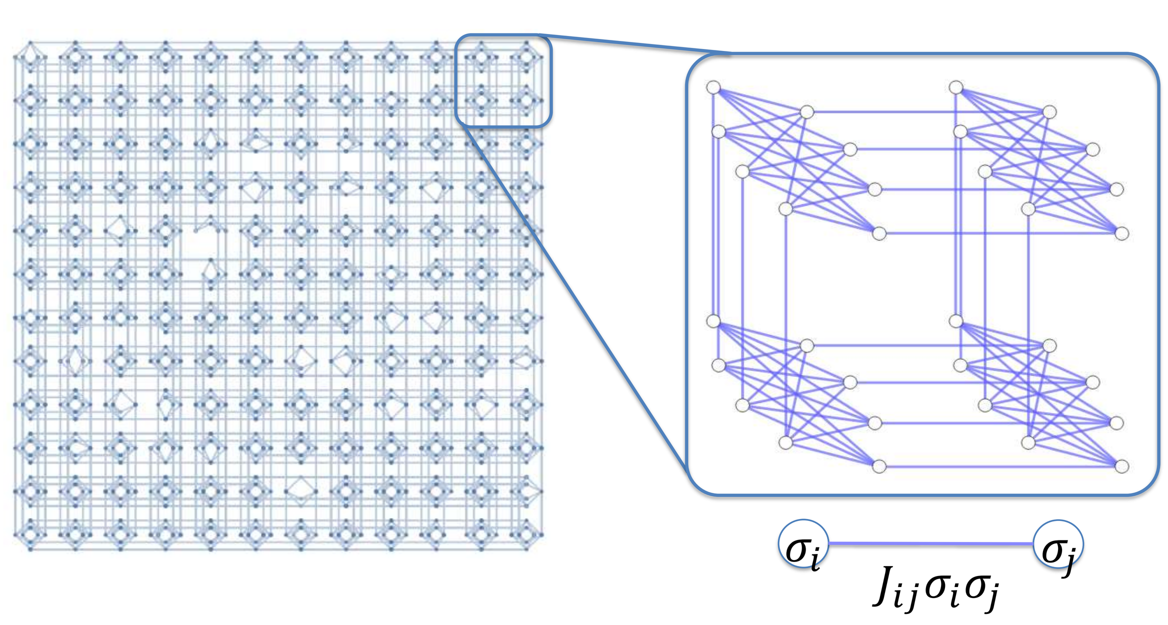


Quantum Annealing and Inverse Ising Problem

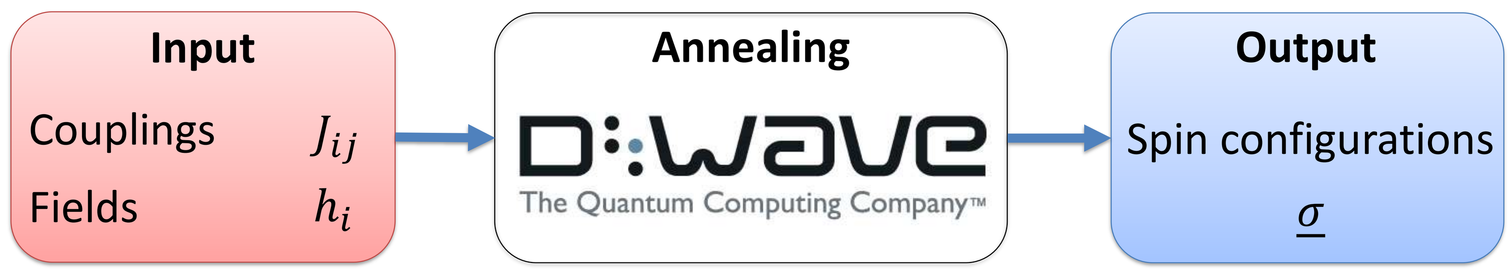
Classical annealing Quantum annealing $H(s) = (1-s)H^x + sH^z$



D-Wave computer is an implementation of quantum annealing for Ising models



Motivation: assessing performance and quantify uncertainties



Question 1: Characterize the probability distribution of output data

Solving Inverse Ising problem

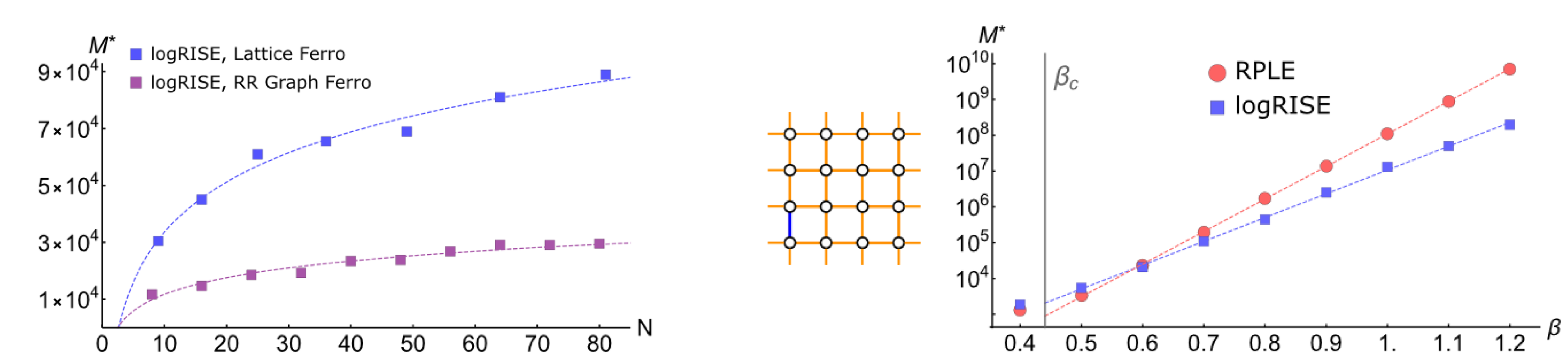
Question 2: What can we learn about the machine from this distribution?

- Type of interactions?
- Structure of interactions?
- Input-output response?

Performance and Rigorous Guarantees

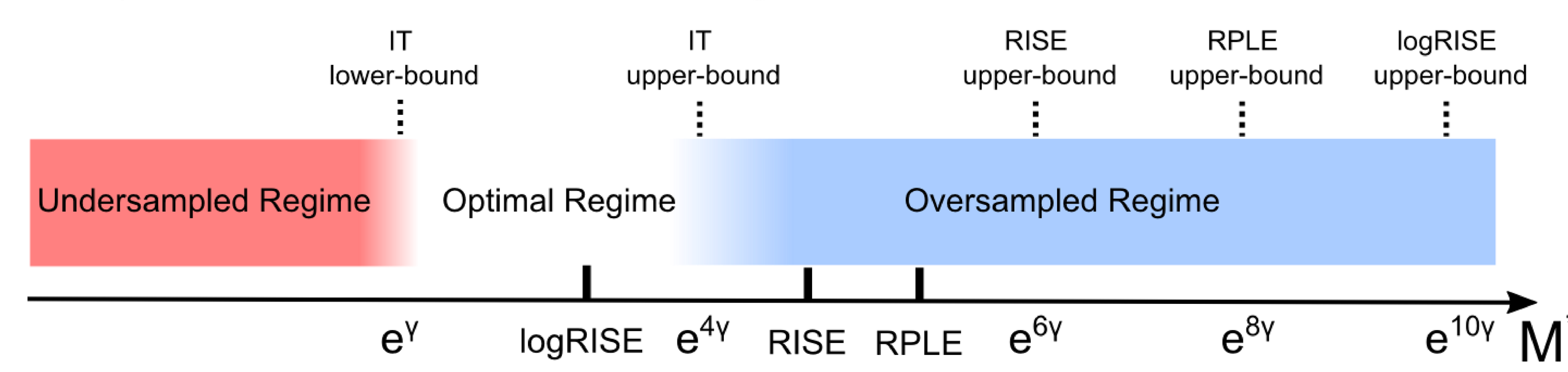
Theorem : If $M > C\frac{q^{2L}\exp(6\gamma)}{\alpha^2}\ln\frac{N}{\delta}$, then $\left\|\hat{\underline{\theta}}_i - \underline{\theta}_i^*\right\|_2 \le \frac{\alpha}{2}$ with prob. $1-\delta$

+ assumptions-free proof for logistic regression loss (pseudo-likelihood)



Generalizations: dense models, reconstruction in the dynamic setting, …

Comparison with **IT bounds for** Ising models:



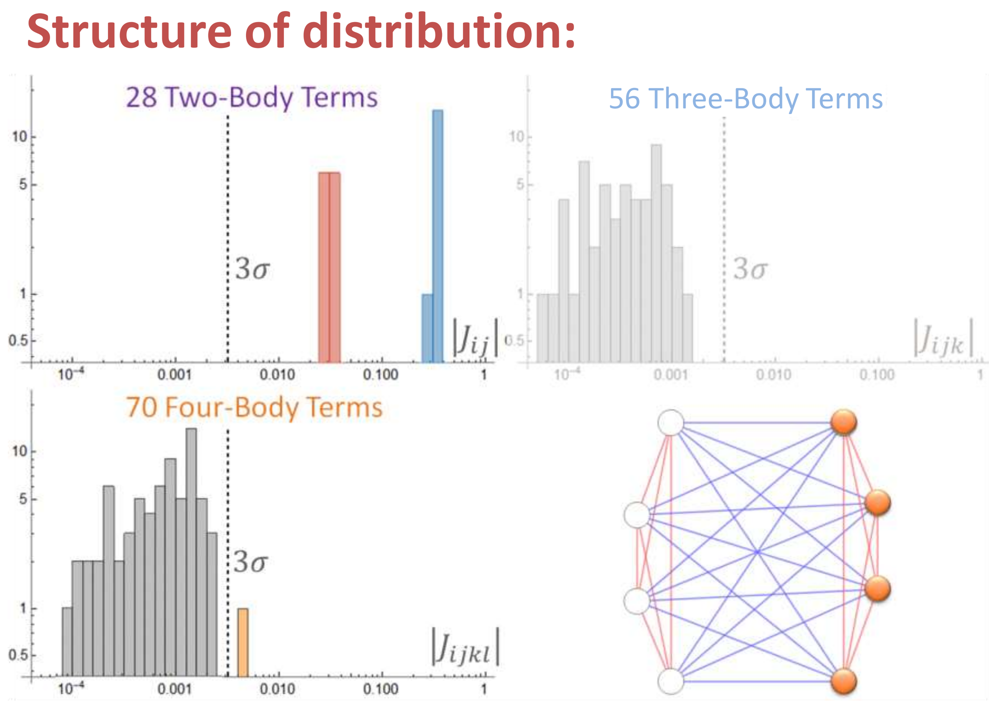
Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models
M. Vuffray, S. Misra, A. Lokhov, M. Chertkov (2016)

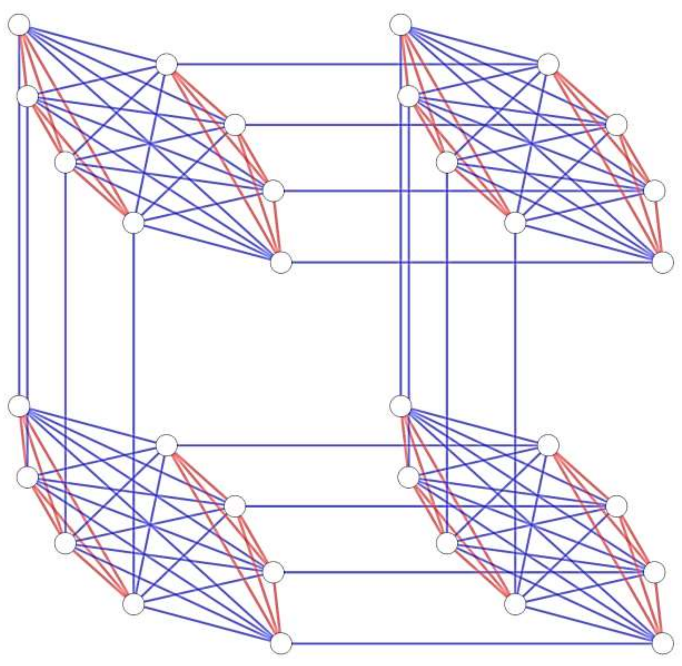
Optimal Structure and Parameter Learning of Ising Models
A. Lokhov, M. Vuffray, S. Misra, M. Chertkov (2018)

Learning Discrete Graphical Models with Generalized Interaction Screening
S. Misra, A. Lokhov, M. Vuffray (submitted)

Understanding the Behavior of D-Wave Annealer

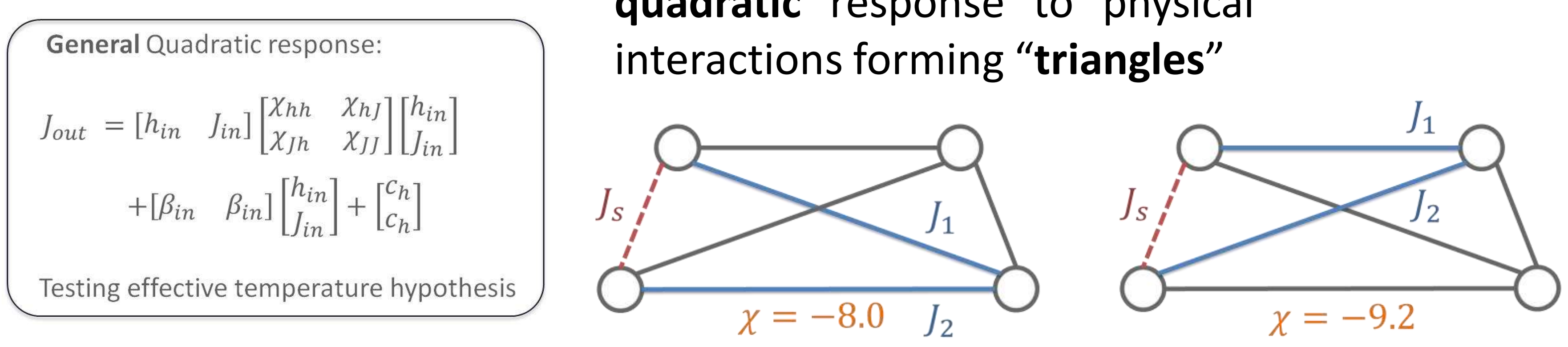
Structure of distribution:



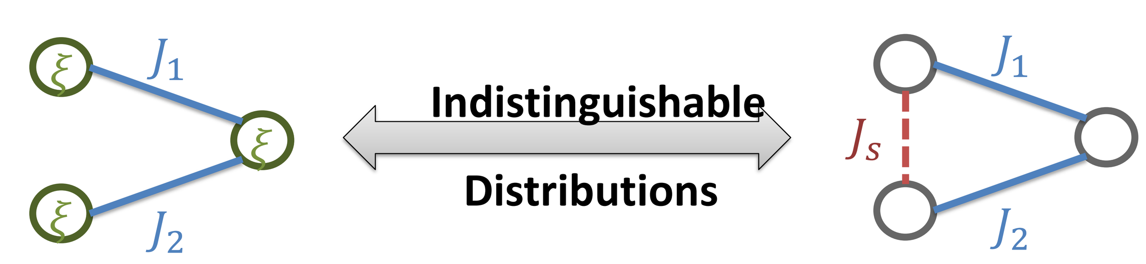


Input-output response:

Spurious interactions are **quadratic** response to physical interactions forming "**triangles**" $J_s = \chi J_1 \times J_2$

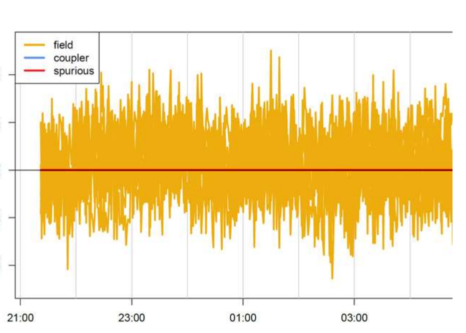


Explanation for spurious links:



Mixture of models with **fluctuating** magnetic fields $\xi \sim \mathcal{N}(0,\sigma_\xi)$

Single model with a **spurious** coupling $J_s \approx -\sigma_\xi^4 J_1 J_2$



Statistical reconstruction of **instantaneous** field **fluctuations** ⟶ **spurious** links with **quadratic** response $J_s \approx -3.2\,J_1J_2$

Opening the Quantum Box: Understanding the Behavior of Analogue Annealers with Statistical Learning A. Lokhov, Y. Kharkov, C. Coffrin, M. Vuffray (submitted)