Density Propagation and Improved Bounds on the Partition Function

Stefano Ermon* Carla Gomes* Ashish Sabharwal+ Bart Selman*

*Computer Science Department, Cornell University, USA

+IBM Watson Research Center, NY, USA

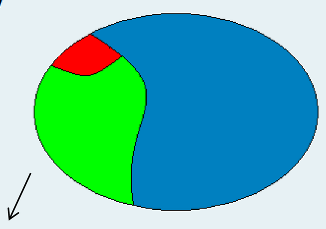
Partition Function, Density of States, and Density Propagation

Partition Function = normalization constant for factored probabilistic models

EXAMPLE: factor graph representation

$$p(x,\Theta)=\frac{1}{Z(\Theta)}\exp(\Theta\cdot\phi(x)) \qquad Z(\Theta^*)=\sum_{x\in\mathcal{X}}\exp(\Theta^*\cdot\phi(x))$$

sum over exponentially many states ⟶ hard to compute



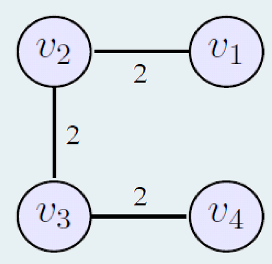
partition of the set of all possible configurations (according to energy)

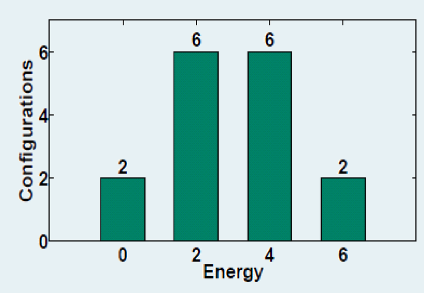
Density of States [Wang et al., Ermon et al.]:

Distribution that for any likelihood value, gives the number of configurations with that probability

$$n(E,\Theta)=\sum_{x\in\mathcal{X}}\delta\left(E-\Theta\cdot\phi(x)\right)$$

one Dirac delta for each possible variable assignment x, centered at its energy





The density of states **gives the partition function** $Z(\Theta^*)=||n(E,\Theta^*)\exp(E)||_1$

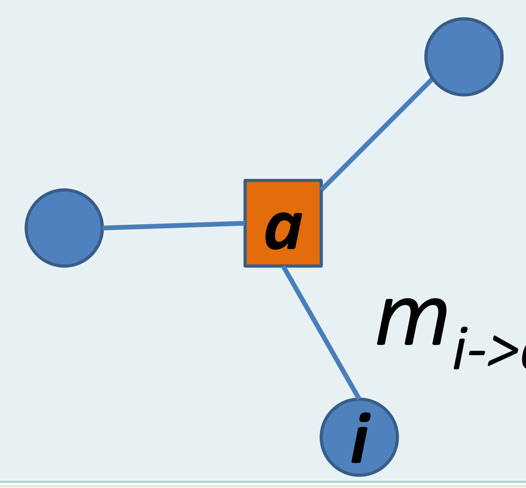
(1) Density Propagation (DP): a new message passing algorithm to compute the density of states of tree-structured models

Message Updates:

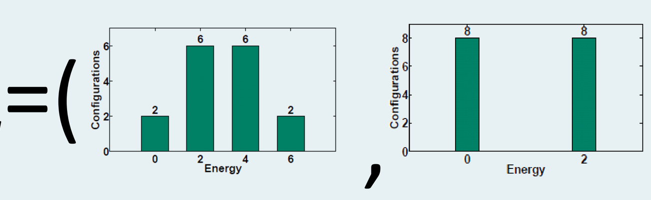
$$m_{i\to a}(x_i)=\bigotimes_{b\in\mathcal{N}(i)\setminus a} m_{b\to i}(x_i)$$

Convolution (sum of conditionally independent RV)

$$m_{a\to i}(x_i)=\sum_{\{x\}_{\alpha\setminus i}}\left(\bigotimes_{j\in\mathcal{N}(a)\setminus i} m_{j\to a}(x_j)\right)\bigotimes\delta_{E_\alpha(\{x\}_\alpha)}$$



DP messages are distributions

$m_{i\text{->}a}$=()

DP generalizes Belief-Propagation and Max-Product

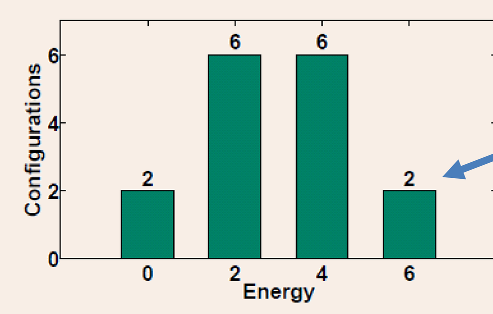
Max Product and Belief Propagation message updates can be derived from DP messages.

Belief Propagation (BP) only considers "total weight"

$$\mu_{i\to j}(x_j)=||m_{i\to j}(x_j)(E)\exp(E)||_1=\int_{\mathbb{R}} m_{i\to j}(x_j)(E)\exp(E)\mathrm{d}E$$

e.g., 2+6exp(2)+6exp(4)+2exp(6)

DP messages are distributions



Max Product (MP) only considers the highest probability entry e.g. exp(6)

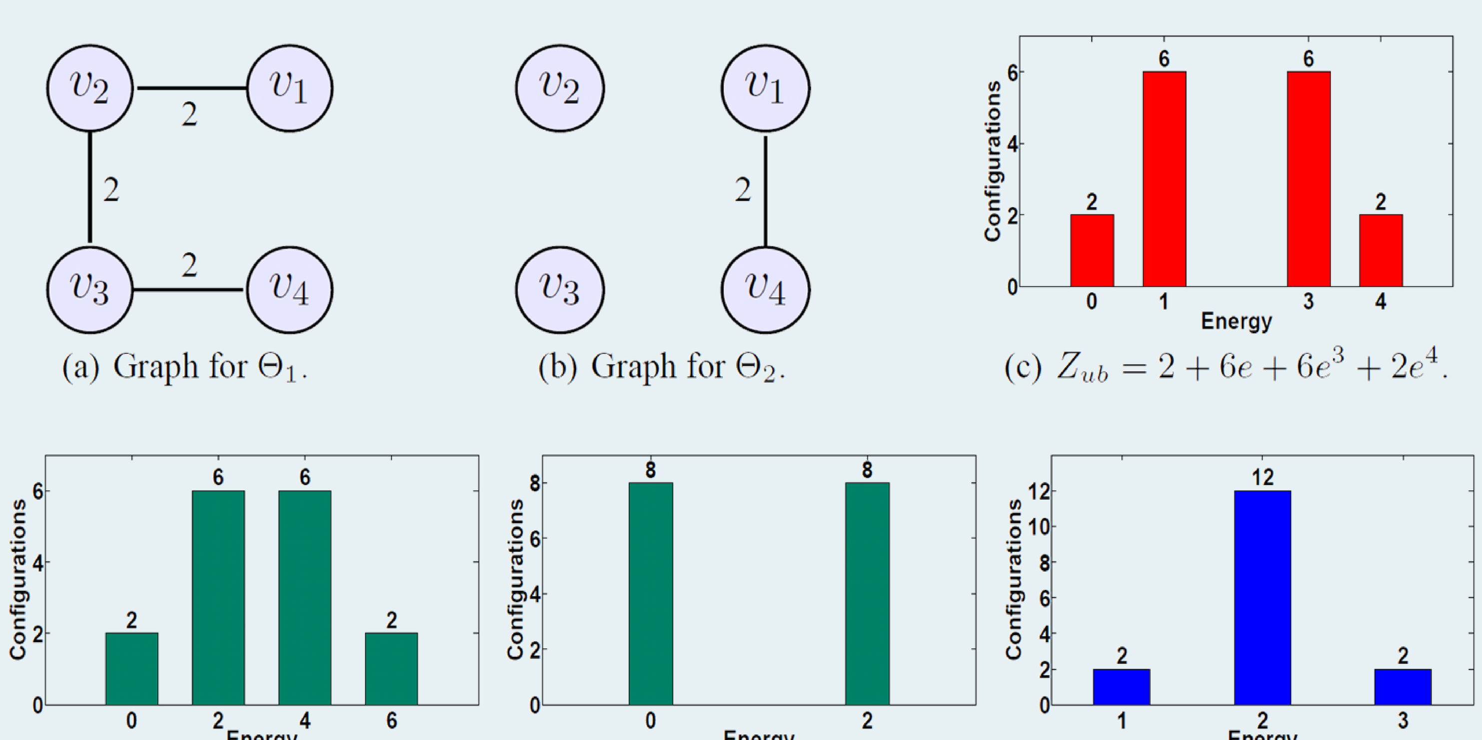
(2) DP messages carry strictly more information than BP and MP

Improved, Matching-Based Bounds on the Partition Function

Decomposition of loopy models into tractable families

[Wainwright et al., Liu et al.]:

Example: 2x2 Ising model decomposed as $\Theta^*=\sum_{i=1}^n\gamma_i\Theta_i$



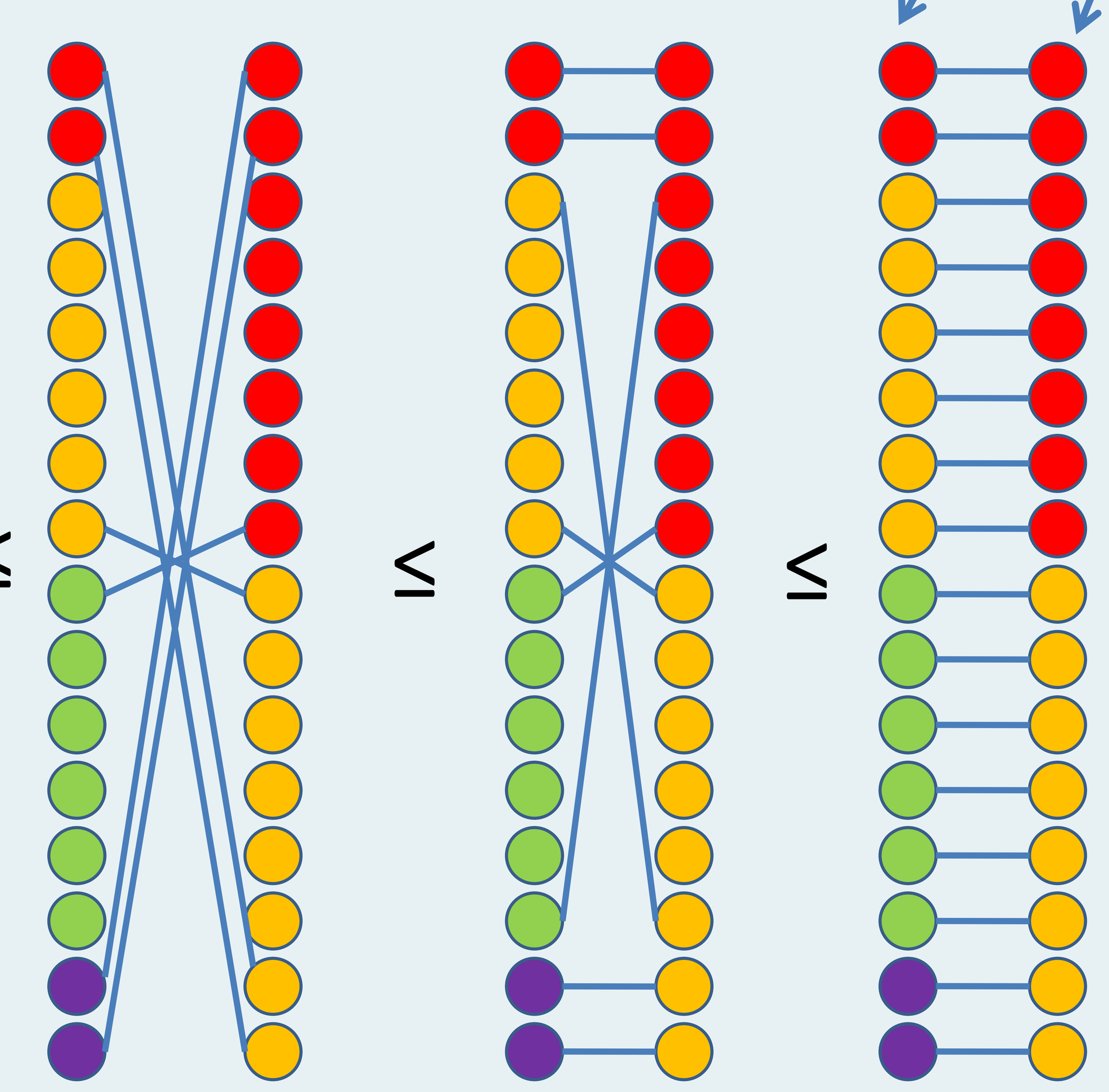
(a) Graph for Θ_1. (b) Graph for Θ_2. (c) $Z_{ub}=2+6e+6e^3+2e^4$.

(d) Histogram $n(E,\Theta_1)$ (e) Histogram $n(E,\Theta_2)$ (f) $Z_{lb}=2e+12e^2+2e^3$

● = energy 0 ● = energy 4
● = energy 2 ● = energy 6

edge weight = $e^{(6+2)/2}$

density of states of tractable subproblems Θ_1 and Θ_2

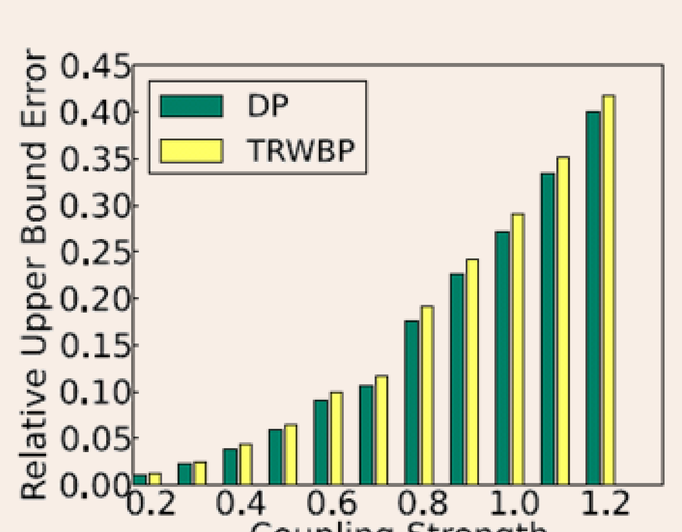
Negative TRWBP [Liu et al.] ≤  ≤ TRWBP [Wainwright et al.]

Min Matching (f) ≤ **(unknown) matching** Sum of edge weights = partition function ≤ **Max Matching (c)**

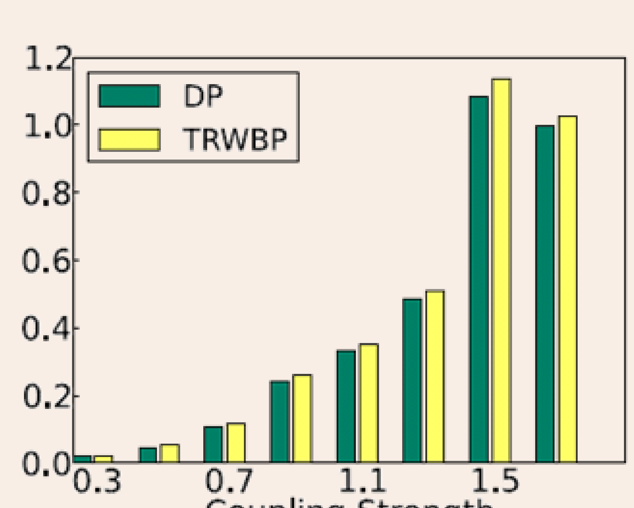
(3) For any decomposition, new matching-based upper and lower bounds provably stronger than convexity-based ones

(when Holder inequality is strict)

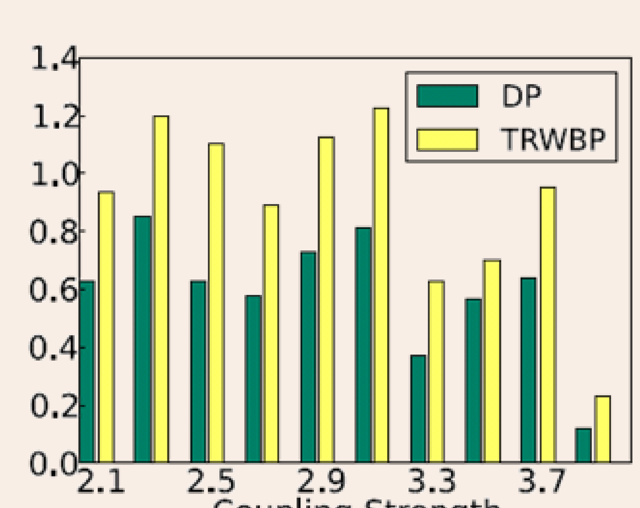
Experimental results



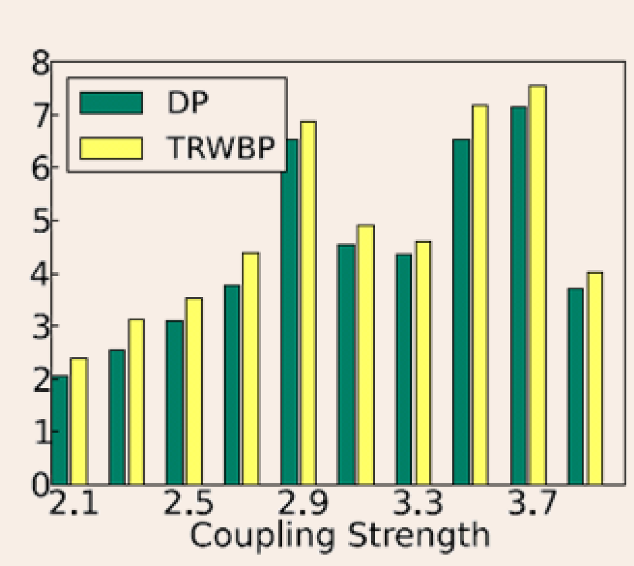
(a) 15 × 15 grid, attractive.



(b) 10 × 10 grid, mixed.



(c) 15-Clique, attractive.



(d) 15-Clique, mixed.

Max-Matching based upper bound always improves over the convexity based one (TRWBP)