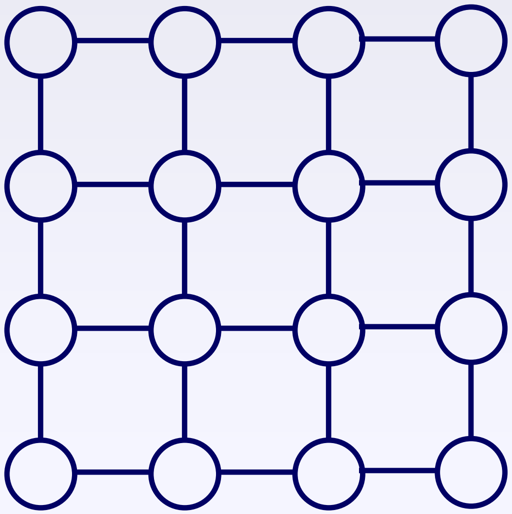
# Loop Series and Bethe Variational Bounds in Attractive Graphical Models

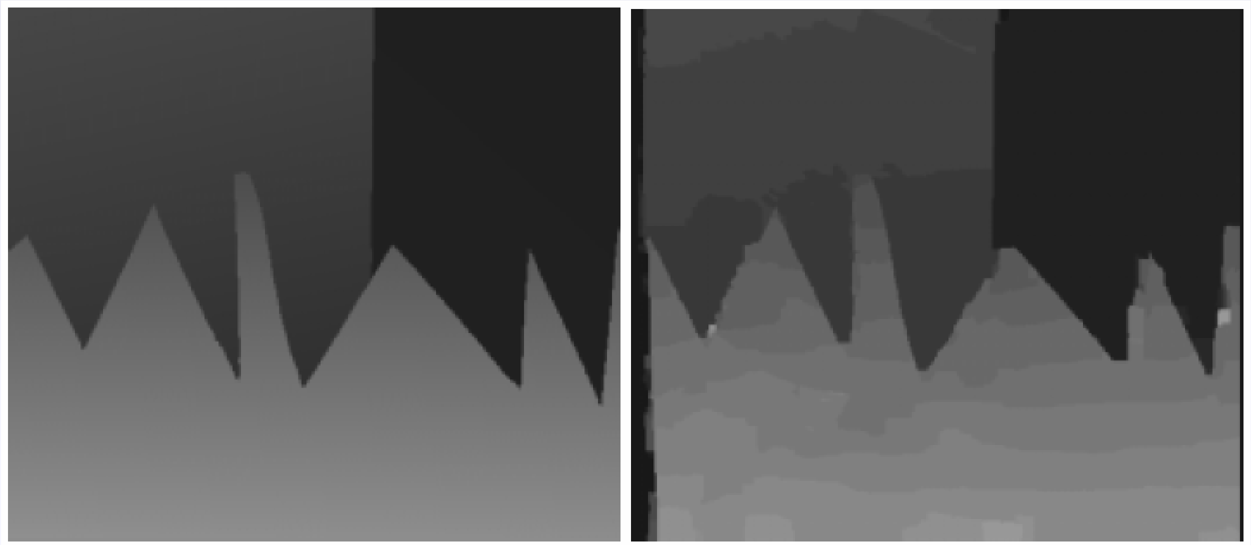

Erik Sudderth

Electrical Engineering & Computer Science
University of California, Berkeley

*Joint work with* Martin Wainwright
Alan Willsky

# Loopy BP and Spatial Priors

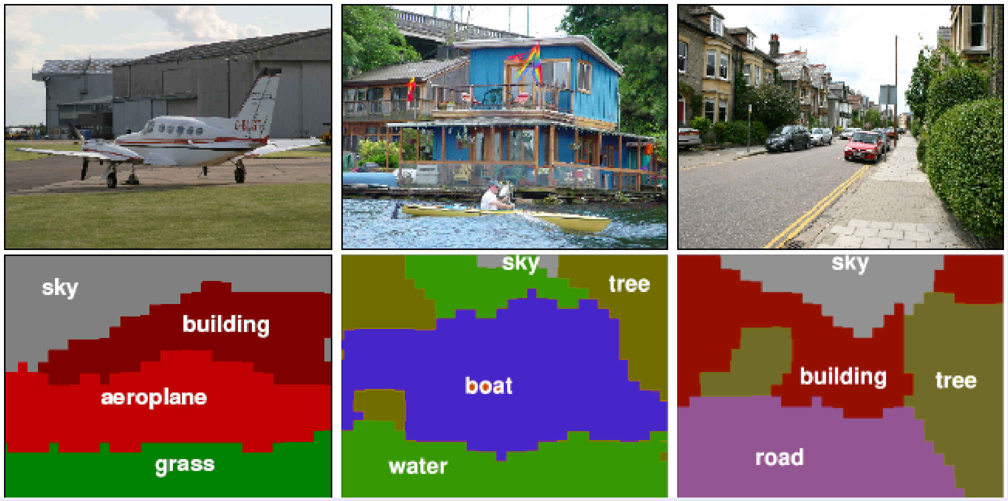




***Dense Stereo Reconstruction*** *(Sun et. al. 2003)*

***Image Denoising***

*(Felzenszwalb & Huttenlocher 2004)*

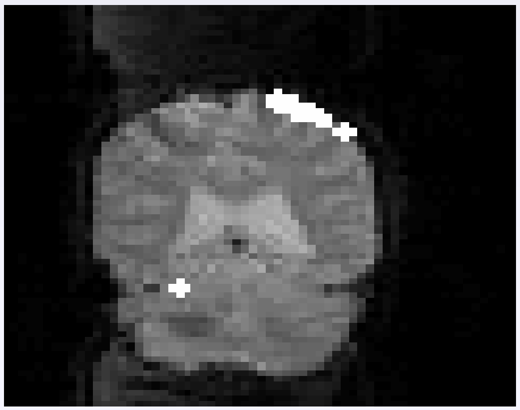



***Segmentation & Object Recognition***

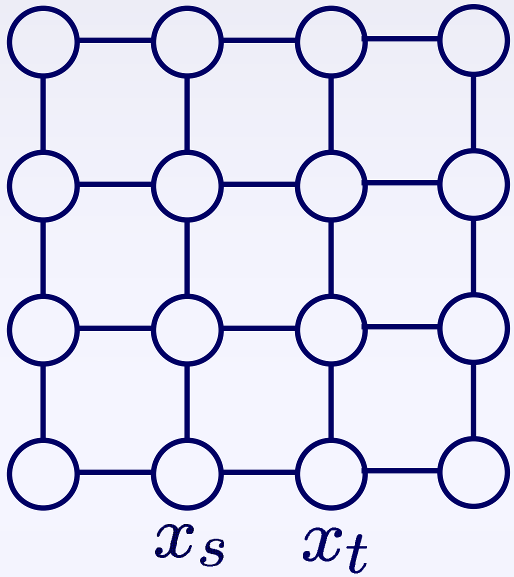
*(Verbeek & Triggs 2007)*

# What do these models share?

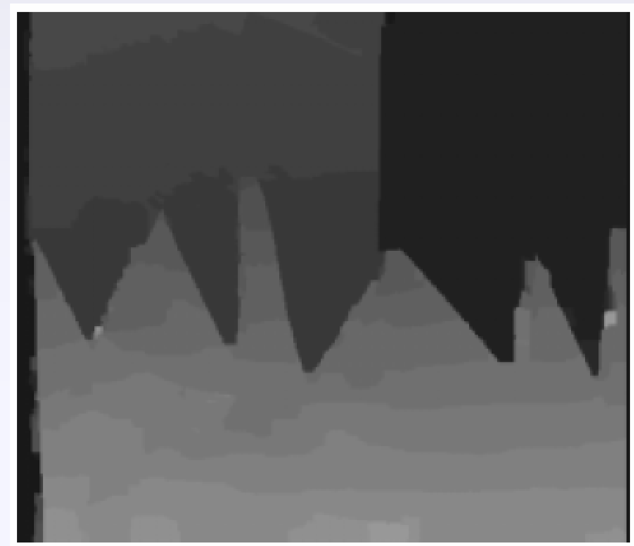
***fMRI Analysis***



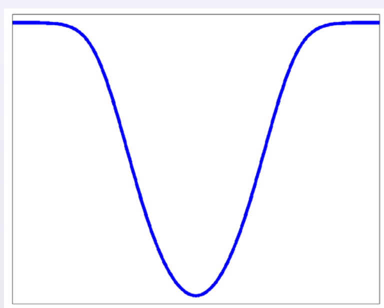
*Kim et. al. 2000*




***Dense Stereo***



$$\phi_{st}(x_s, x_t) = \begin{cases} 0 & x_s = x_t \\ \theta_{st} > 0 & \text{otherwise} \end{cases}$$



$$\phi_{st}(x_s, x_t) = D\left(\frac{x_s - x_t}{\sigma_{st}}\right)$$

*pairwise energies are attractive to encourage spatial smoothness*

# Outline

## Graphical Models & Belief Propagation

- Pairwise Markov random fields
- Variational methods & loopy BP

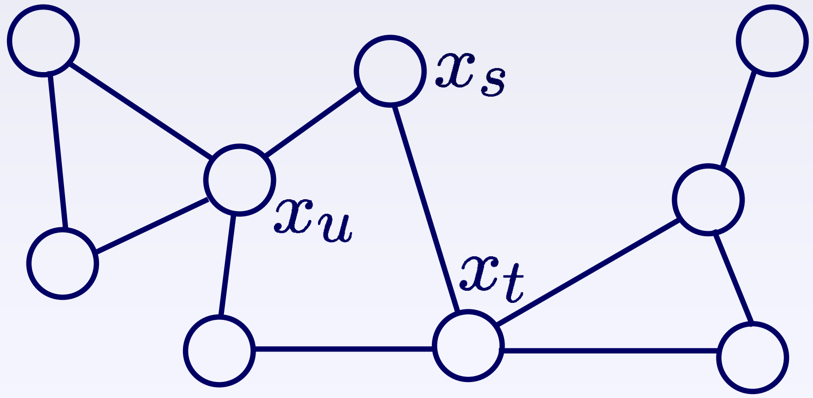
## Binary Markov Random Fields

- Attractive pairwise interactions
- Loop series expansion of the partition function

## Bounds & the Bethe Approximation

- Conditions under which BP provides bounds
- Empirical comparison to mean field bounds

# Pairwise Markov Random Fields




$$p(x) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

$V \longrightarrow$ set of $N$ nodes representing *random variables* $x_s$

$E \longrightarrow$ set of edges $(s, t)$ connecting pairs of nodes, inducing dependence via positive *compatibility functions*

$Z \longrightarrow$ normalization constant or *partition function*

# Why the Partition Function?

$$Z = \sum_x \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

**Statistical Physics**

- Sensitivity of physical systems to external stimuli

**Hierarchical Bayesian Models**

- Marginal likelihood of observed data
- Fundamental in hypothesis testing & model selection

**Cumulant Generating Function**

- For exponential families, derivatives with respect to parameters provide marginal statistics

***PROBLEM:*** Computing $Z$ in general graphs is intractable

# Gibbs Variational Principle

$$\psi(x) := \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

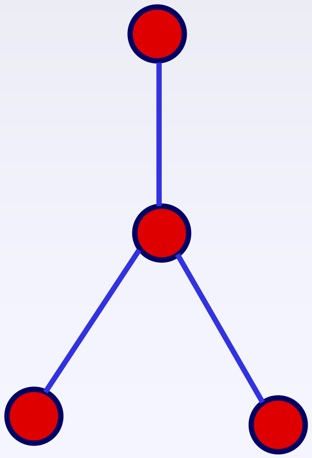
$$\log Z = \max_{q \in \mathcal{Q}} \overbrace{\underbrace{-\sum_x q(x) \log q(x)}_{\textit{Entropy}} - \underbrace{\left[ -\sum_x q(x) \log \psi(x) \right]}_{\textit{Average Energy}}}$$

*All Joint Distributions*

*Negative Gibbs Free Energy* $= -D(q(x) \,||\, p(x)) + \log Z$

- Mean field methods optimize bound over a restricted family of *tractable* densities $\tilde{\mathcal{Q}}$
- Provide *lower bounds* on $Z$

# Belief Propagation in Trees



$$p(x) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

$$= \prod_{s \in V} p_s(x_s) \prod_{(s,t) \in E} \frac{p_{st}(x_s, x_t)}{p_s(x_s) p_t(x_t)}$$

*Exact Marginals*

- *Belief propagation (BP)* is a message passing algorithm that infers this *reparameterization*

*Tree structure leads to a simplified representation of the exact variational problem*

$$\log Z = \max_{q = \{q_s, q_{st}\}} H_\beta(q) + \sum_x q(x) \log \psi(x)$$

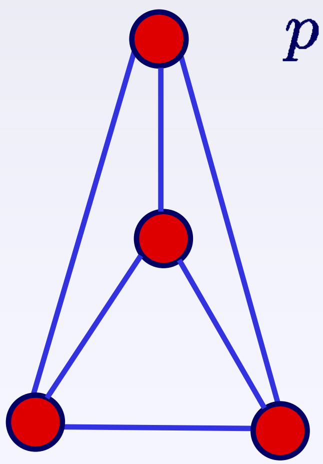
*subject to* $\sum_{x_s} q_{st}(x_s, x_t) = q_t(x_t) \qquad \sum_{x_s} q_s(x_s) = 1$

$$H_\beta(q) = \sum_{s \in V} H_s(q_s) - \sum_{(s,t) \in E} I_{st}(q_{st})$$

*Marginal Entropies*

*Mutual Information*

# Bethe Approximations & Loopy BP



$$p(x) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

$$= \frac{1}{Z(q)} \prod_{s \in V} q_s(x_s) \prod_{(s,t) \in E} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

*Pseudo-Marginals*

- Fixed points of *loopy BP* also correspond to reparameterizations of $p(x)$ *(Wainwright et. al. 2001)*

*Bethe variational approximation parameterized by pseudo-marginals which may be globally inconsistent*

$$\log Z_\beta = \max_{q=\{q_s, q_{st}\}} H_\beta(q) + \sum_x q(x) \log \psi(x)$$

*subject to* $\sum_{x_s} q_{st}(x_s, x_t) = q_t(x_t)$ $\qquad \sum_{x_s} q_s(x_s) = 1$

$$H_\beta(q) = \sum_{s \in V} H_s(q_s) - \sum_{(s,t) \in E} I_{st}(q_{st})$$

*Yedidia, Freeman, & Weiss 2000*

# When is Loopy BP Effective?

### Graphs with Long Cycles

*(Gallager 1963; Richardson & Urbanke 2001)*

- Turbo codes & low density parity check (LDPC) codes
- For long block lengths, graph becomes *locally tree-like*, and BP accurate with high probability

### Graphs with Weak Potentials

*(Tatikonda & Jordan 2002; Heskes 2004; Ihler et. al. 2005; Mooij & Kappen 2005)*

- If potentials are sufficiently weak, BP has a *unique fixed point*
- Analyzing compatibility strength in context of graph structure can sometimes guarantee message passing *convergence*

### Graphs with Attractive Potentials?

- Existing theory does not explain empirical effectiveness
- We will show that the Bethe approximation *lower bounds* the true partition function for a family of attractive models

# Outline

## Graphical Models & Belief Propagation

- Pairwise Markov random fields
- Variational methods & loopy BP

## Binary Markov Random Fields

- Attractive pairwise interactions
- Loop series expansion of the partition function

## Bounds & the Bethe Approximation

- Conditions under which BP provides bounds
- Empirical comparison to mean field bounds

# Binary Markov Random Fields

*Boltzmann Machines, Ising Models, …*

$$p(x) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

- Nodes associated with binary variables: $x_s \in \{0, 1\}$
- Parameterize *pseudo-marginal* distributions via moments:

$$\tau_s := q_s(X_s = 1)$$

$$\tau_t := q_t(X_t = 1)$$

$$\tau_{st} := q_{st}(X_s = 1, X_t = 1)$$

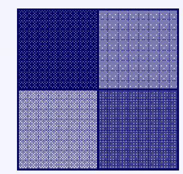
$$q_{st}(x_s, x_t) = \begin{array}{cc} \begin{array}{cc} \quad\quad 0 \quad\quad\quad & \quad\quad\quad 1 \end{array} & x_t / x_s \\ \begin{bmatrix} 1 - \tau_s - \tau_t + \tau_{st} & \tau_t - \tau_{st} \\ \tau_s - \tau_{st} & \tau_{st} \end{bmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

# Attractive Binary Models

$$p(x) = \frac{1}{Z} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

- A pairwise MRF has *attractive* compatibilities if all edges $(s,t)$ satisfy the following bound:

$$\psi_{st}(0,0)\ \psi_{st}(1,1) \geq \psi_{st}(0,1)\ \psi_{st}(1,0)$$



- Equivalent condition on reparameterized *pseudo-marginals:*

$$\mathrm{Cov}_{q_{st}}(X_s, X_t) = \tau_{st} - \tau_s \tau_t \geq 0$$

- In statistical physics, such models are *ferromagnetic*
- Extensive literature on *correlation inequalities* bounding moments of attractive fields: *GHS, FKG, GKS, …*

# Bounding Partition Functions

$$p(x) = \frac{1}{Z(\psi)} \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t)$$

*Original MRF*

$$= \frac{1}{Z(q)} \prod_{s \in V} q_s(x_s) \prod_{(s,t) \in E} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

*Reparam. MRF*

- Compatibilities differ by a positive, constant multiple:

| | *True Partition Function* | | *Bethe Approximation* |
|---|---|---|---|
| *Original MRF* | $Z(\psi)$ | $\geq$ | $Z_\beta(\psi)$ |
| *Reparam. MRF* | $Z(q)$ | $\geq$ | $Z_\beta(q) = 1$ |

$$\frac{Z(\psi)}{Z_\beta(\psi)} = \frac{Z(q)}{1}$$

- Focus analysis on partition function of *reparameterized* MRF

# Loop Series Expansions

$$Z(q) = \sum_{x \in \{0,1\}^n} \prod_{s \in V} q_s(x_s) \prod_{(s,t) \in E} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

- True log partition function can be expressed as a series expansion, whose first term is the Bethe approximation:

$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\left[ (X_s - \tau_s)^{d_s(F)} \right]$$

$F \longrightarrow$ nonempty subset of the graph's edges

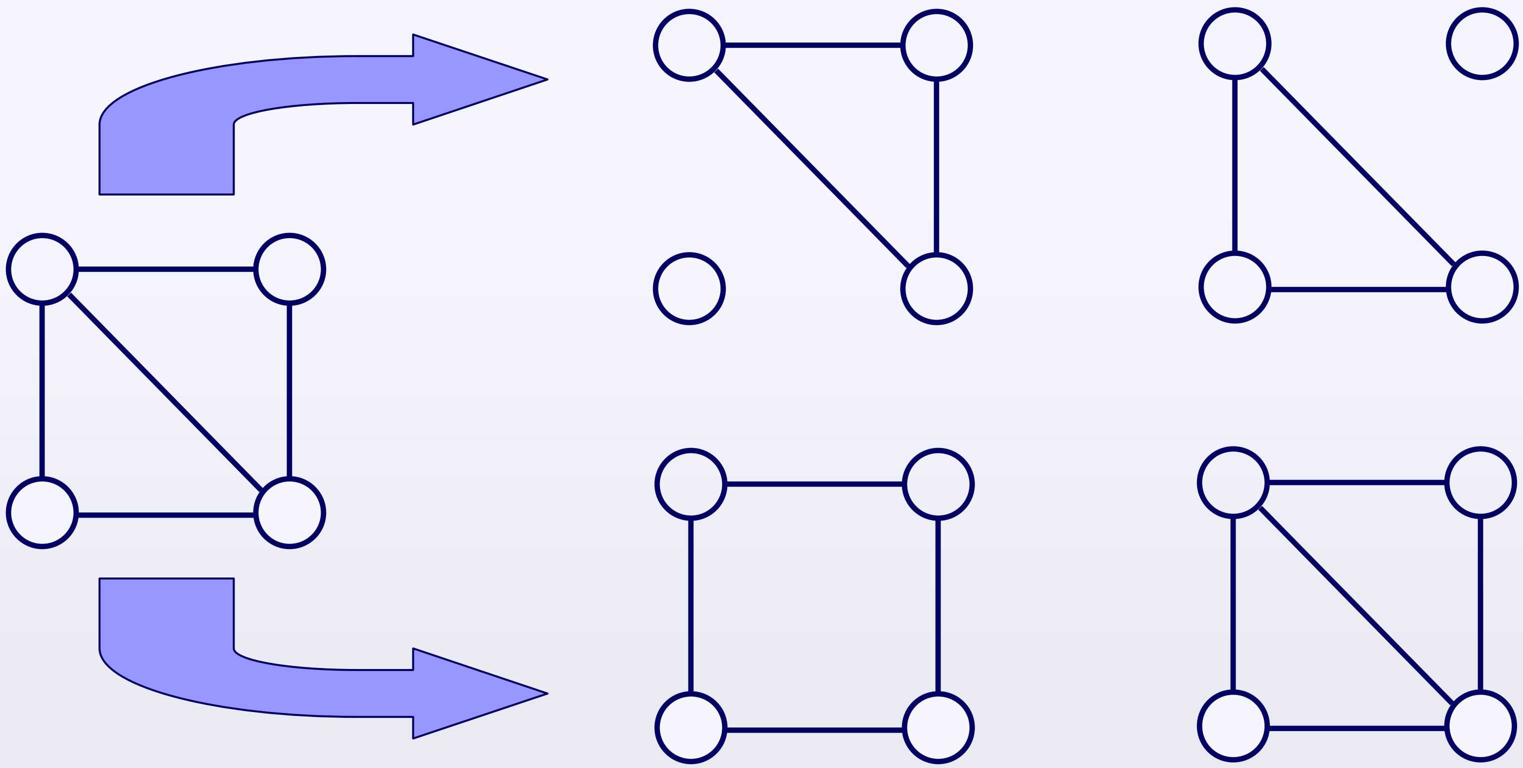
$\beta_F \longrightarrow$ scalar function of $\{q_{st}(x_s, x_t) \mid (s,t) \in F\}$

$d_s(F) \longrightarrow$ degree of node $s$ in *subgraph* induced by $F$

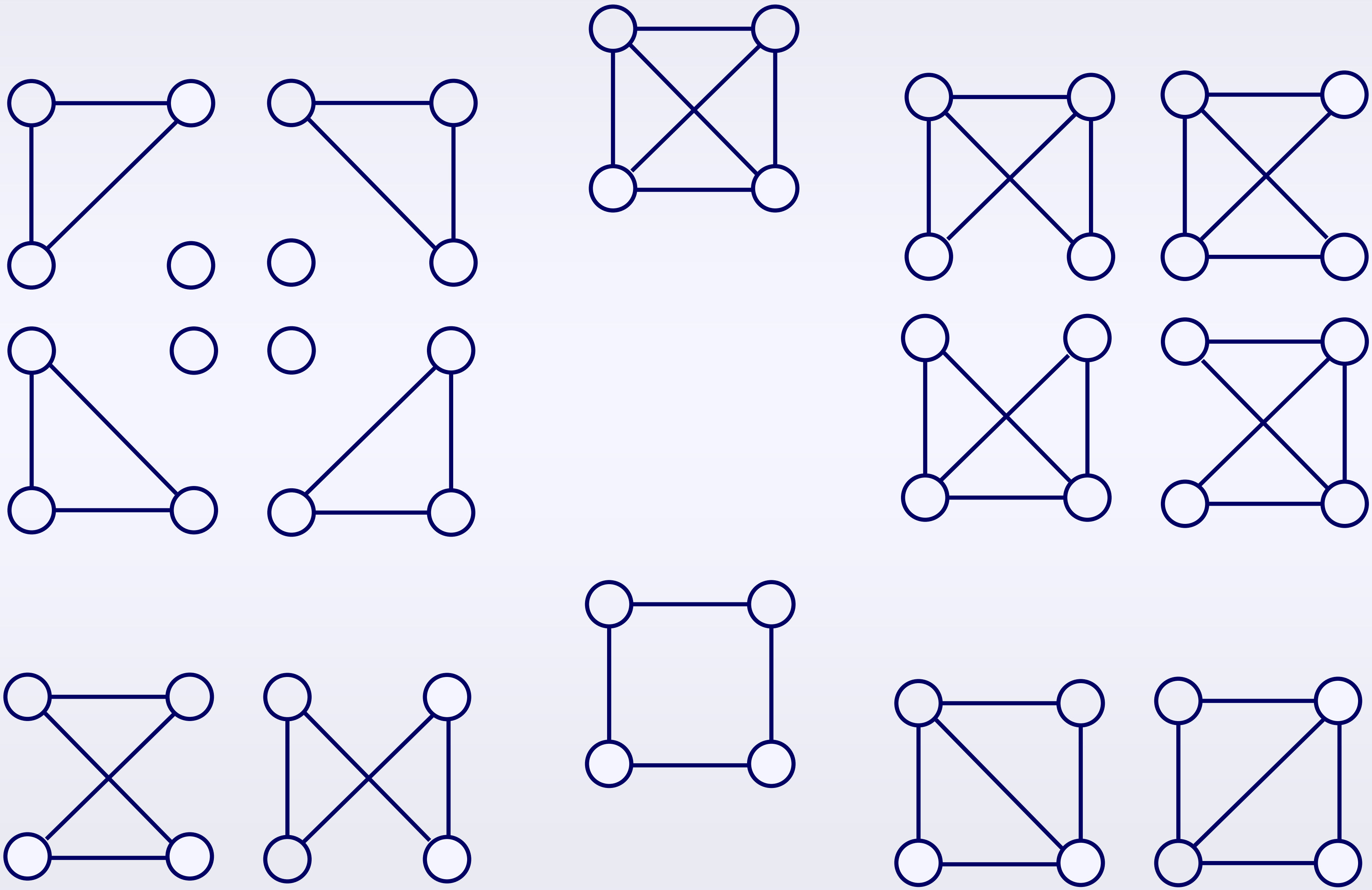
- These *loop corrections* are only non-zero when $F$ defines a *generalized loop* *(Chertkov & Chernyak, 2006)*

# Generalized Loops

- Subgraphs in which *all nodes* have degree $d_s(F) \neq 1$
- All *connected nodes* must have degree $d_s(F) \geq 2$



# Lots of Generalized Loops



# Deriving the Loop Series

**Two Existing Approaches** *(Chertkov & Chernyak 2006)*

- Saddle point approximation of BP fixed point based upon contour integration in a complex auxiliary field
- Employ Fourier representation of binary functions, and manipulate terms via hyperbolic trigonometric identities

**Our Contribution: A Probabilistic Derivation**

- Simple, direct derivation from reparameterization characterization of loopy BP fixed points
- Exposes probabilistic interpretations for loop series terms, and makes connections to other known invariants

# Loop Series: A Key Identity

$$Z(q) = \sum_{x \in \{0,1\}^n} \prod_{s \in V} q_s(x_s) \prod_{(s,t) \in E} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

- For binary variables, reparameterized pairwise compatibilities can be expressed as follows:

$$\frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)} = 1 + \beta_{st}(x_s - \tau_s)(x_t - \tau_t)$$

$$\beta_{st} := \frac{\tau_{st} - \tau_s \tau_t}{\tau_s(1 - \tau_s)\tau_t(1 - \tau_t)} = \frac{\mathrm{Cov}_{q_{st}}(X_s, X_t)}{\mathrm{Var}_{q_s}(X_s)\,\mathrm{Var}_{q_t}(X_t)}$$

- Straightforward (but tedious) to verify for $(x_s, x_t) \in \{0, 1\}^2$
- For *attractive* compatibilities, note that $\beta_{st} \geq 0$

# Loop Series Derivation

$$Z(q) = \sum_{x \in \{0,1\}^n} \prod_{s \in V} q_s(x_s) \prod_{(s,t) \in E} \frac{q_{st}(x_s, x_t)}{q_s(x_s) q_t(x_t)}$$

$$= \mathbb{E}_{\tilde{q}}\left[ \prod_{(s,t) \in E} \frac{q_{st}(X_s, X_t)}{q_s(X_s) q_t(X_t)} \right]$$

$$= \mathbb{E}_{\tilde{q}}\left[ \prod_{(s,t) \in E} 1 + \beta_{st}(X_s - \tau_s)(X_t - \tau_t) \right]$$

$$= 1 + \sum_{\emptyset \neq F \subseteq E} \mathbb{E}_{\tilde{q}}\left[ \prod_{(s,t) \in F} \beta_{st}(X_s - \tau_s)(X_t - \tau_t) \right]$$

Expectation over *factorized* distribution: $\tilde{q}(x) := \prod_{s \in V} q_s(x_s)$

Expand polynomial using *linearity* of expectations:

$$(1+a)(1+b)(1+c) = 1+a+b+c+ab+bc+ac+abc$$

# Pairwise Loop Series Expansion

$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \mathbb{E}_{\tilde{q}}\Big[ \prod_{(s,t) \in F} \beta_{st}(X_s - \tau_s)(X_t - \tau_t) \Big]$$

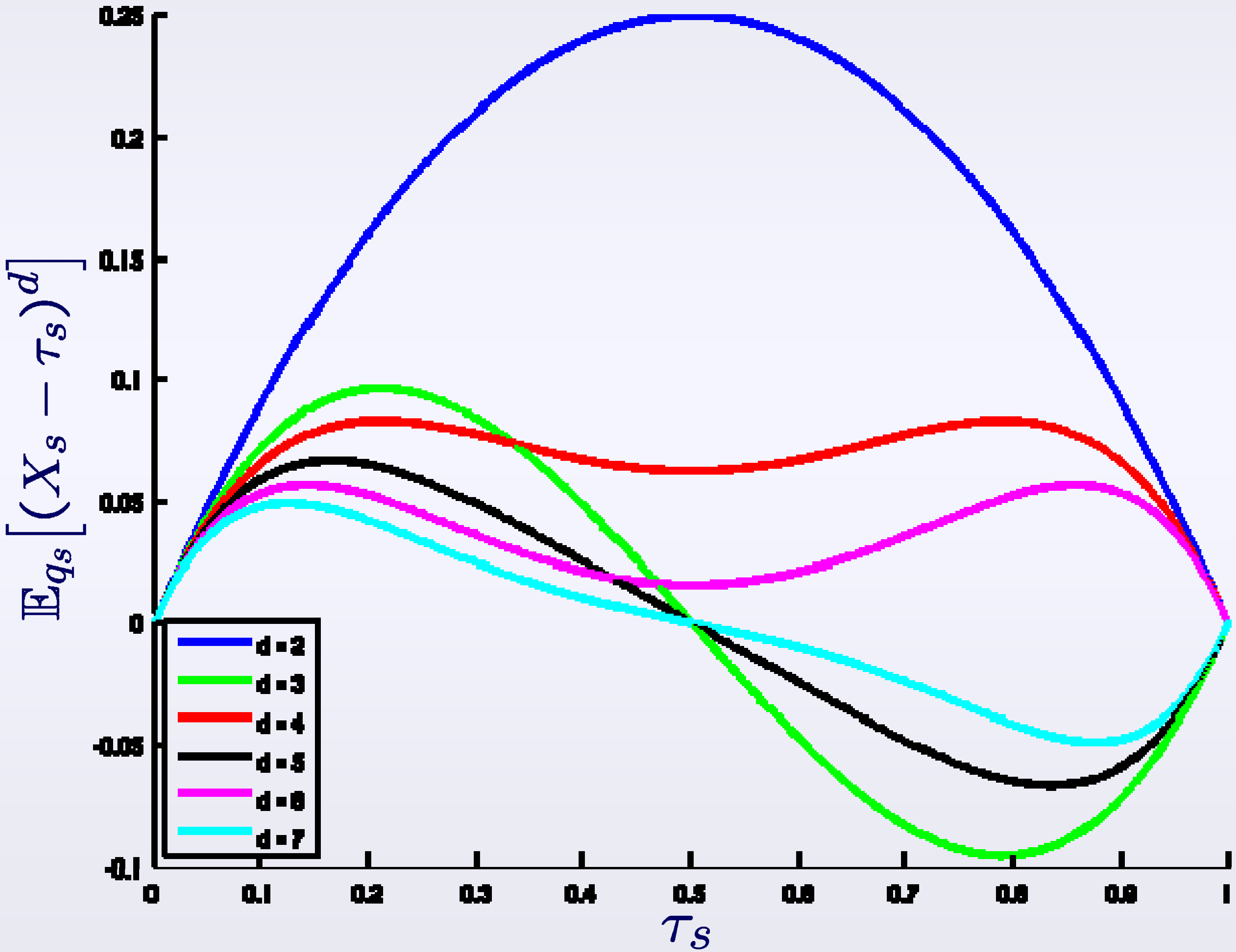
$$= 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\Big[ (X_s - \tau_s)^{d_s(F)} \Big]$$

$$\beta_F := \prod_{(s,t) \in F} \beta_{st}$$

$d_s(F) \longrightarrow$ degree of node $s$ in *subgraph* induced by $F$

- Depends on *central pseudo-moments* corresponding to loopy BP fixed point: $\mathbb{E}_{q_s}\Big[ (X_s - \tau_s)^{d_s(F)} \Big]$
- Only *generalized loops* are non-zero: $\mathbb{E}_{q_s}[X_s - \tau_s] = 0$

# Bernoulli Central Moments



# Outline

**Graphical Models & Belief Propagation**

- Pairwise Markov random fields
- Variational methods & loopy BP

**Binary Markov Random Fields**

- Attractive pairwise interactions
- Loop series expansion of the partition function

**Bounds & the Bethe Approximation**

- Conditions under which BP provides bounds
- Empirical comparison to mean field bounds

# Bethe Bounds in Attractive Models

$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\left[(X_s - \tau_s)^{d_s(F)}\right]$$

**Theorem:** For a *"large family"* of binary MRFs with attractive compatibilities, any BP fixed point provides a lower bound:

| | *True Partition Function* | | *Bethe Approximation* |
|---|---|---|---|
| *Original MRF* | $Z(\psi)$ | $\geq$ | $Z_\beta(\psi)$ |
| *Reparam. MRF* | $Z(q)$ | $\geq$ | 1 |

*Sufficient condition:* Show that all terms in the loop series are *non-negative*

**Conjecture:** For *all* binary MRFs with attractive compatibilities, the Bethe approximation always provides a lower bound

# Loop Series in Attractive Models

$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\left[ (X_s - \tau_s)^{d_s(F)} \right]$$

$$\beta_F = \prod_{(s,t) \in F} \beta_{st} \qquad\qquad \beta_{st} = \frac{\mathrm{Cov}_{q_{st}}(X_s, X_t)}{\mathrm{Var}_{q_s}(X_s)\,\mathrm{Var}_{q_t}(X_t)} \geq 0$$

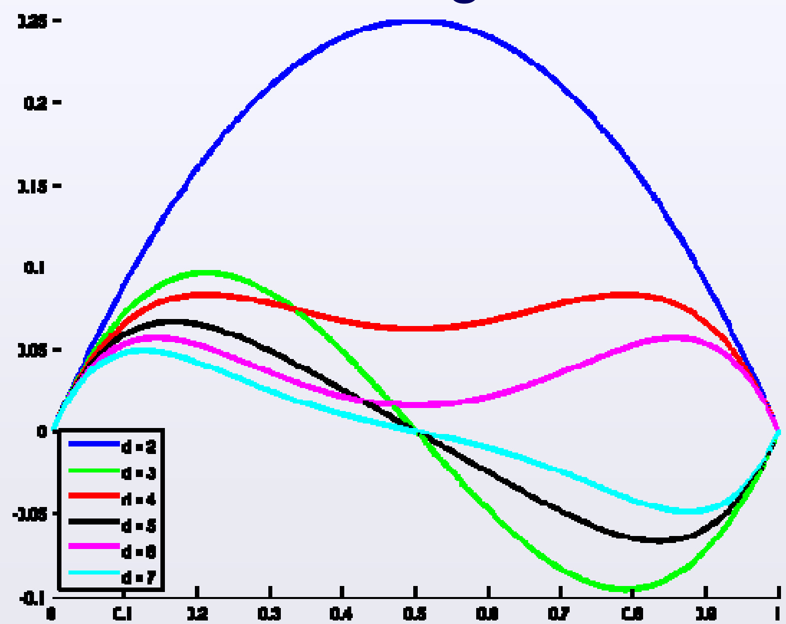
- When are binary pseudo-central moments non-negative?

- Bound holds when

$$\tau_s \leq \frac{1}{2} \quad \textit{for all nodes} \quad s \in V$$

***OR***

$$\tau_s \geq \frac{1}{2} \quad \textit{for all nodes} \quad s \in V$$



# Loop Series in Attractive Models

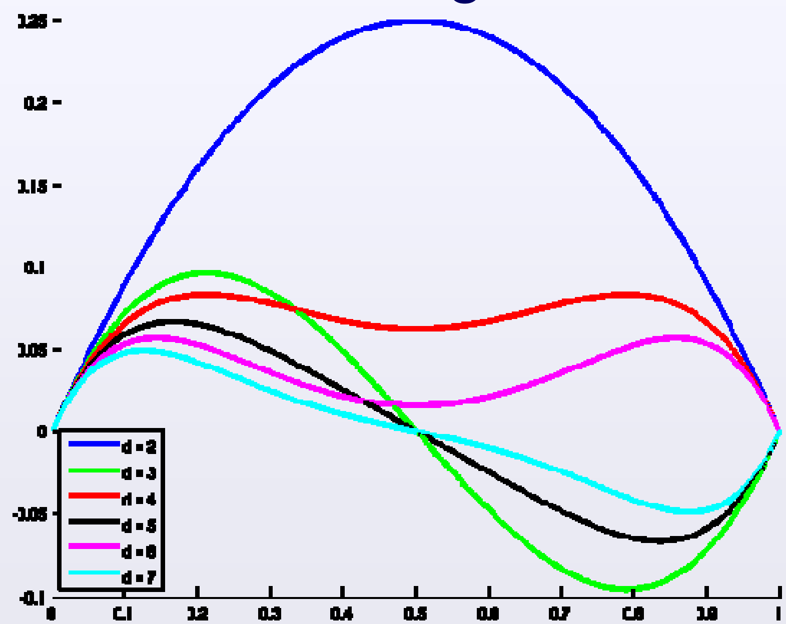
$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\left[(X_s - \tau_s)^{d_s(F)}\right]$$

$$\beta_F = \prod_{(s,t) \in F} \beta_{st} \qquad \beta_{st} = \frac{\mathrm{Cov}_{q_{st}}(X_s, X_t)}{\mathrm{Var}_{q_s}(X_s)\,\mathrm{Var}_{q_t}(X_t)} \geq 0$$

- When are binary pseudo-central moments non-negative?
- Only nodes with degrees

$$d_s(F) \geq 3$$

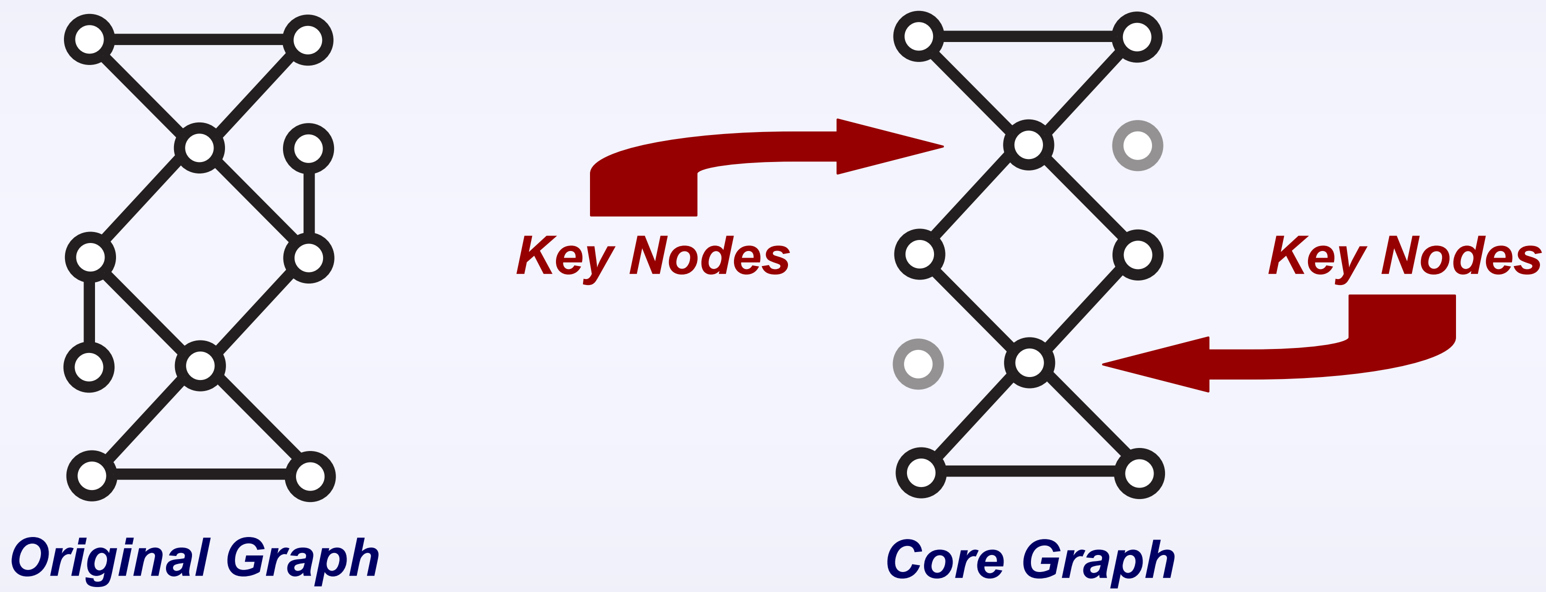
must agree in sign

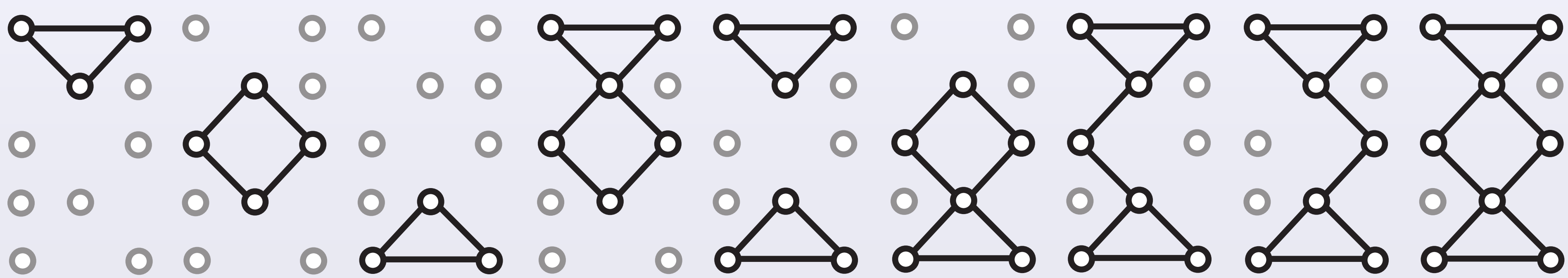



*Bound always holds for graphs with a single cycle*

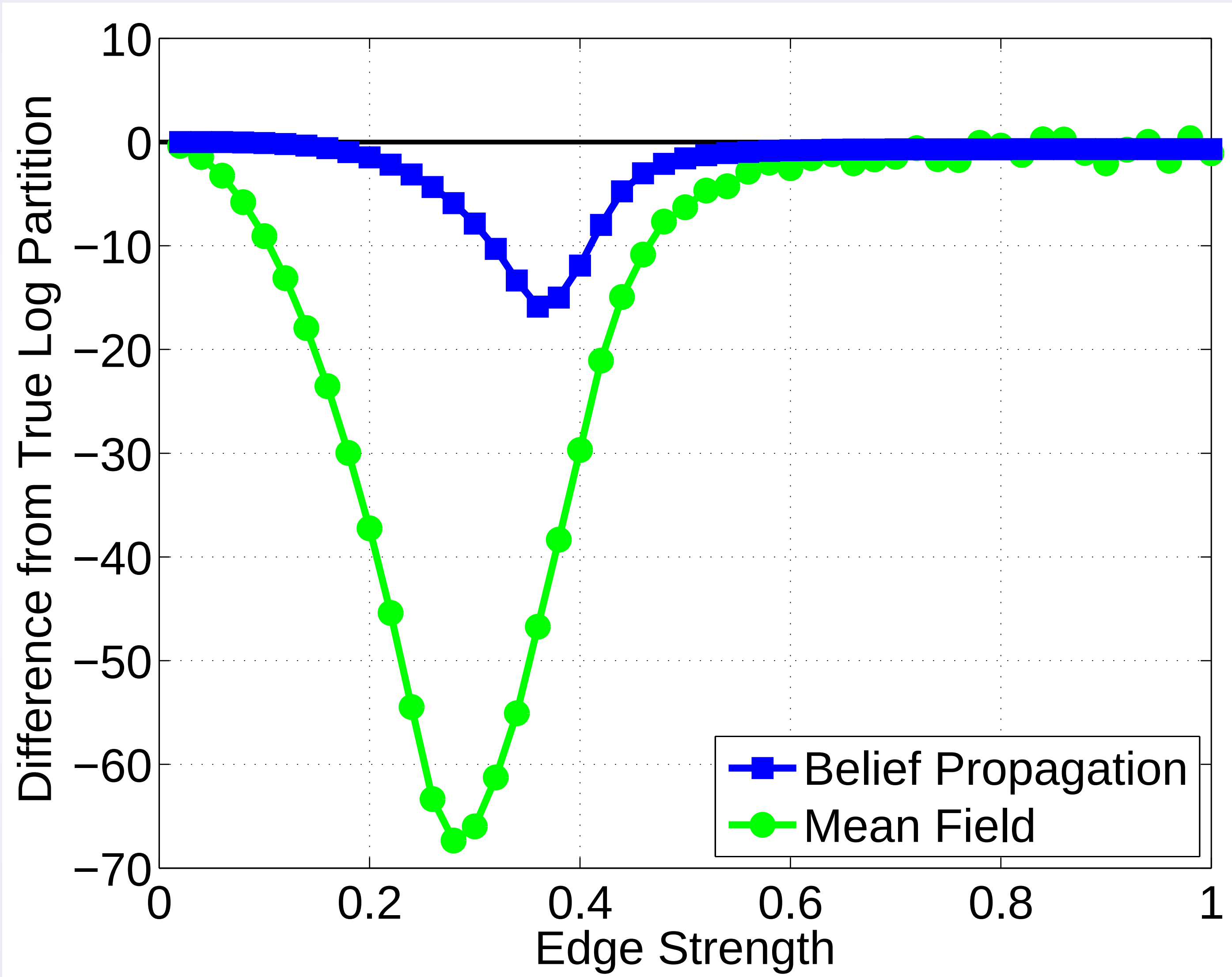
# Weaker Bound Conditions

$$Z(q) = 1 + \sum_{\emptyset \neq F \subseteq E} \beta_F \prod_{s \in V} \mathbb{E}_{q_s}\left[(X_s - \tau_s)^{d_s(F)}\right]$$



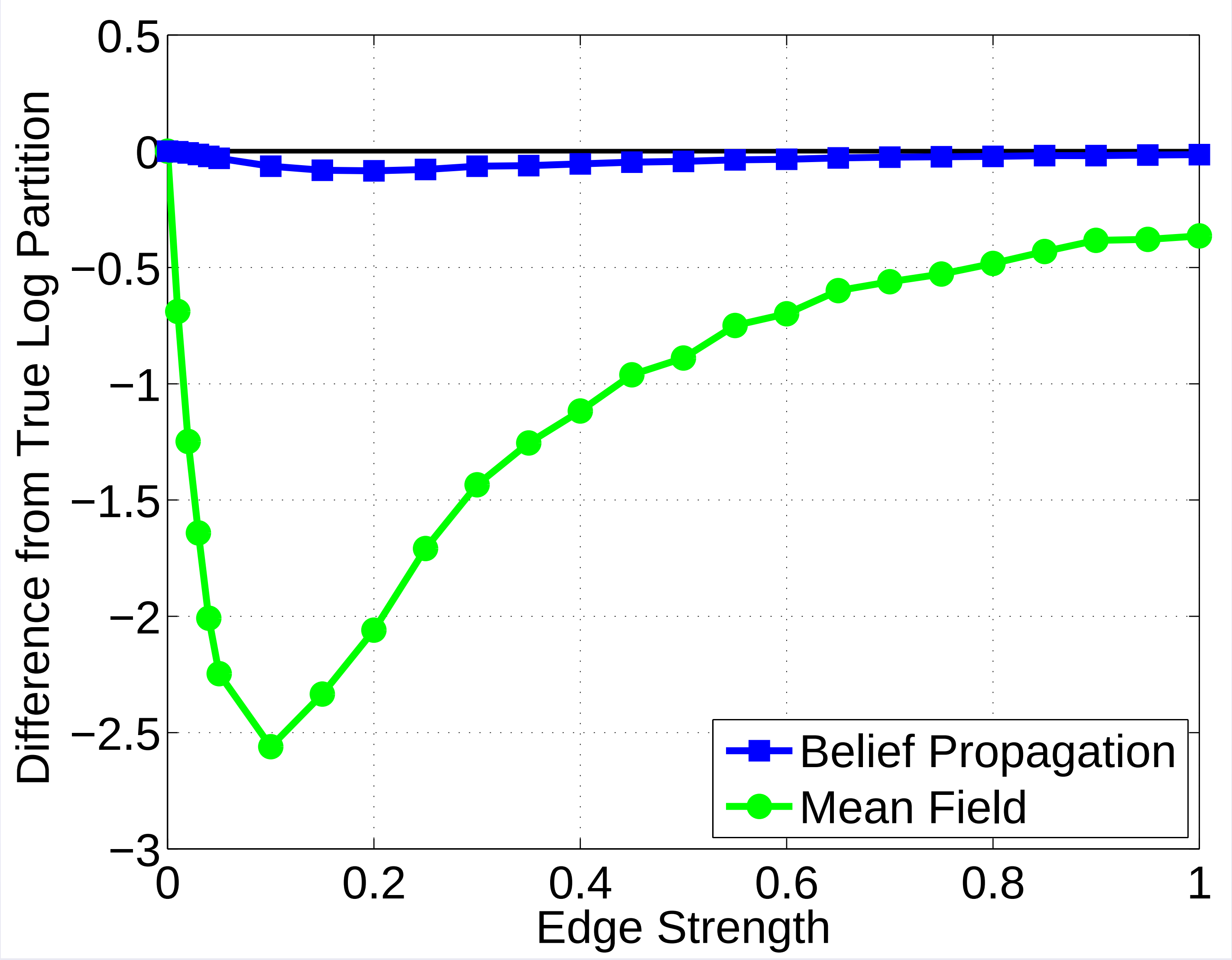



# Empirical Bounds: 30x30 Torus



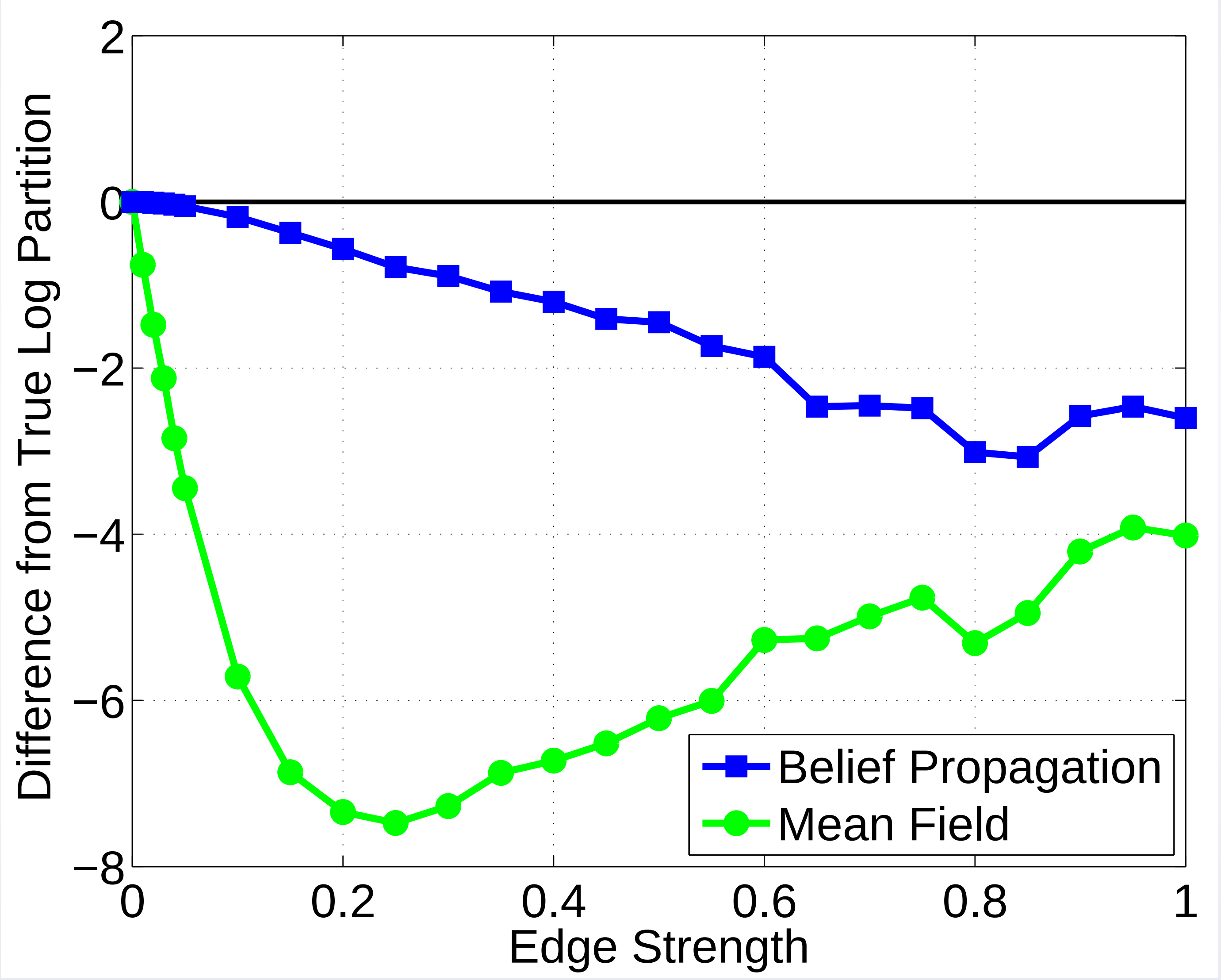

*Exact partition function via eigenvector method of Onsager (1944)*

# Empirical Bounds: 10x10 Grid



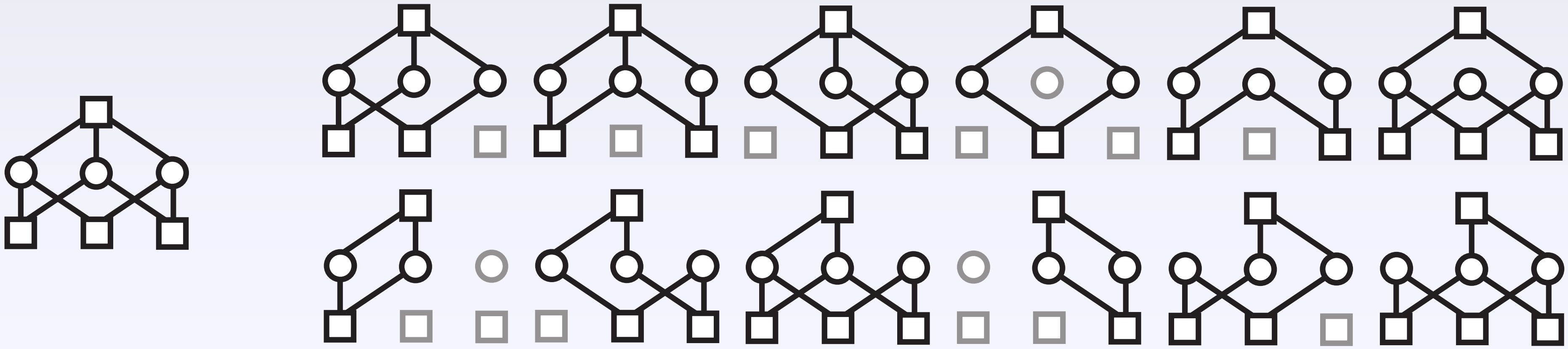

*All marginals have same bias, satisfying conditions of theorem*

# Empirical Bounds: 10x10 Grid




*Random marginals with mixed biases, so some negative loop corrections*

# Generalization: Factor Graphs



- Generalized loops: all connected *variable nodes* and *factor nodes* must have degree at least two
- *Probabilistic derivation* via reparameterization generalizes
- *Bethe lower bound* continues to hold for a higher-order family of attractive binary compatibilities

# Conclusions

## Belief Propagation & Partition Functions

- Simple, probabilistic derivation of the *loop series* expansion associated with fixed points of loopy BP
- Proof that the Bethe approximation lower bounds the true partition function in many *attractive* binary models

## Ongoing Research

- Generalize expansion & bounds to other model families: *higher-order discrete MRFs, Gaussian MRFs*
- Implications of results for BP *dynamics* in attractive models, and stability of *learning* algorithms based on loopy BP