KIT
Karlsruhe Institute of Technology

Open6GHub

SPONSORED BY THE Federal Ministry of Education and Research

erc European Research Council Established by the European Commission

CEL

Local Message Passing on Frustrated Systems

Luca Schmid, Joshua Brenk, Laurent Schmalen

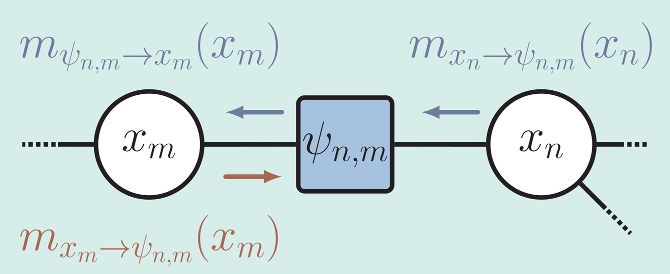
1. Motivation: Marginal Inference

- **Task**: Compute marginal distribution $p(x_n|\boldsymbol{y}) = \sum_{\sim\{x_n\}} p(x_1,\ldots,x_N|\boldsymbol{y})$
  - Marginalization $\sum_{\sim\{x_n\}}$: Sum over all variables except x_n
  - $\boldsymbol{y}$: Noisy observation of the system $x_1,\ldots,x_N$
  - Many practical problems require marginalization
- Efficient approach: **Message passing on graphical models**
  - Widely used: **Sum-product algorithm (SPA)** / belief propagation (BP)
- **Problem**: SPA is derived for graphical models with tree structure
  → **Suboptimal** for graphical models with many short cycles, which often occur in practical applications

2. Message Passing on Graphs with Cycles

- **Idea**: Instead of fixing the mismatched SPA, **find a new message update rule** that is especially taylored to graphs with (many) cycles!
- Message update rule:

  A mapping from one or multiple incident messages to one outgoing message, which is applied **locally** at a (factor) node.



- Sum-product algorithm (SPA):

$$m_{\psi\to x}(x) = \sum_{\sim\{x\}}\left(\psi(\mathcal{X})\prod_{x'\in\mathcal{X}\setminus\{x\}} m_{x'\to\psi}(x')\right),\quad \mathcal{X}\text{: Neighbors of } \psi$$

- General update rule for pairwise factors $\psi_{n,m}(x_m,x_n)$:

  Extrinsic: $\mathrm{FN}_{\mathrm{e}}(\psi_{n,m}) : m_{x_n\to\psi_{n,m}}(x_n) \mapsto m_{\psi_{n,m}\to x_m}(x_m)$

  novel: **Non-extrinsic:** $\mathrm{FN}(\psi_{n,m}) : \begin{pmatrix} m_{x_n\to\psi_{n,m}}(x_n) \\ m_{x_m\to\psi_{n,m}}(x_m)\end{pmatrix} \mapsto m_{\psi_{n,m}\to x_m}(x_m)$

- **cycBP/cycBP$_\mathrm{e}$**: **Represent** (non-)extrinsic **mapping** by compact multilayer perceptron (1 hidden layer à 7 neurons)
  - Optimize local mapping towards good **end-to-end inference performance**

3. Objective Functions for Marginal Inference

- Supervised Training: **Kullback-Leibler (KL) divergence**

$$\mathcal{L}_{\mathrm{KL}} := D_{\mathrm{KL}}\left(b_n(x_n)\|p(x_n)\right)$$

  - $b_n(x_n), b_{n,m}(x_n,x_m)$: single, pairwise beliefs of message passing
  - $p(x_n) = \sum_{\sim\{x_n\}} p(x_1,\ldots,x_N)$: Exact marginal distributions
- Unsupervised Training: **Regularized Bethe free energy**

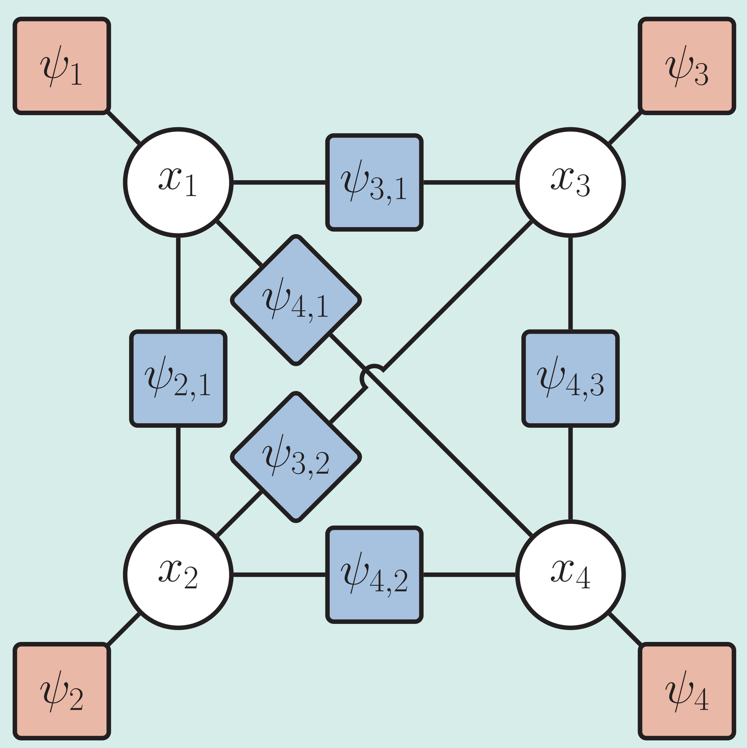
$$\mathcal{L}_{\mathrm{Bethe}} := F_{\mathrm{Bethe}} + \alpha\mathcal{L}_{\mathbb{L}},\quad \alpha\in\mathbb{R}^+$$

  - F_{Bethe}: Bethe free energy
  - **Bethe consistency distance** (*proposed*):

$$\mathcal{L}_{\mathbb{L}} := D_{\mathrm{KL}}\left(\sum_{x_m} b_{n,m}(x_n,x_m)\Big\| b_n(x_n)\right) + D_{\mathrm{KL}}\left(\sum_{x_n} b_{n,m}(x_n,x_m)\Big\| b_m(x_m)\right)$$

4. Example: 2×2 Ising Model

- System: $p(x_1,\ldots,x_N|\boldsymbol{y}) = \frac{1}{Z}\prod_{n=1}^{N}\psi_n(x_n)\prod_{n>m}\psi_{nm}(x_m,x_n)$
  - $\psi_n(x_n) = \exp(\theta_n x_n)$ with **local fields** θ_n, $n=1,\ldots,N$
  - $\psi_{n,m}(x_n,x_m) = \exp(E_{n,m}x_n x_m)$ with **local couplings** $E_{n,m}$, $n>m$
- $N=4$ binary variables $x_n\in\{+1,-1\}$
  - Use log-likelihood ratios $L_{\psi_{nm}\to x_n} := \log\left(\frac{m_{\psi_{nm}\to x_n}(x_n=+1)}{m_{\psi_{nm}\to x_n}(x_n=-1)}\right)$

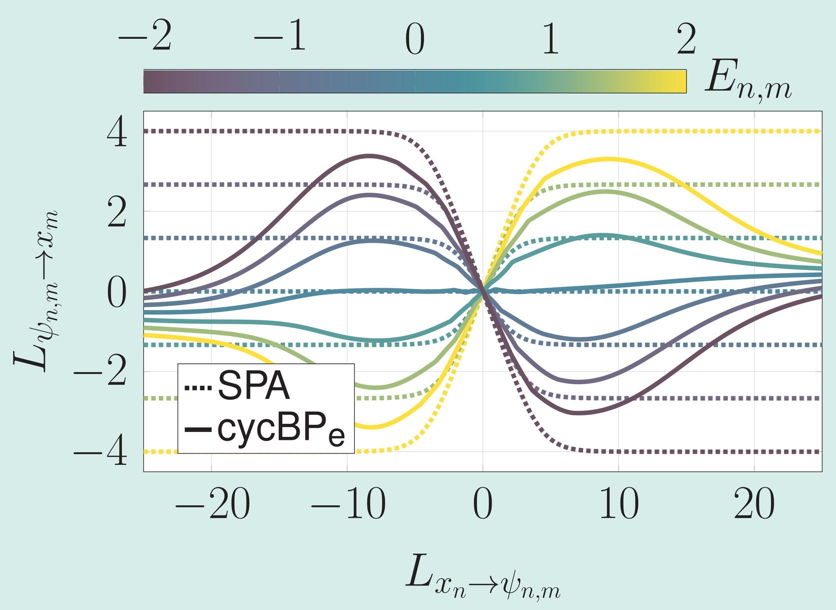


- **Spin glass**: θ_n and $E_{n,m}$ independently sampled from $\mathcal{U}[-2,+2]$
  - **Frustrated system**, e.g., for $E_{n,m}\ll 0,\ \forall n,m$

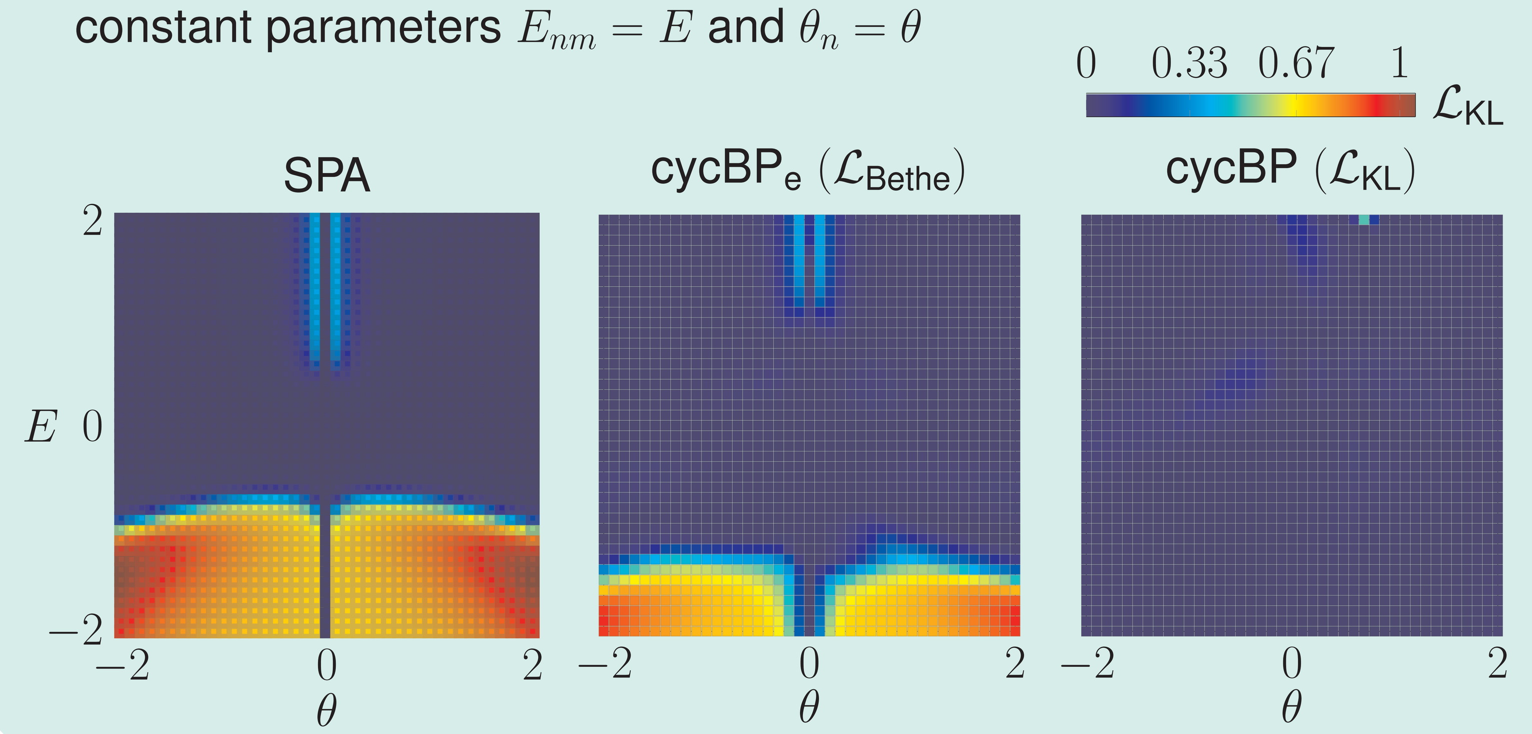
5. Experiment 1: 2×2 Ising Spin Glasses

- **Training** and **evaluation** on samples of the Ising spin glass model

Algo.	Loss	$\mathcal{L}_{\mathrm{KL}}$	$\mathcal{L}_{\mathbb{L}}$
SPA	-	0.087	0.30
cycBP$_\mathrm{e}$	$\mathcal{L}_{\mathrm{KL}}$	0.040	0.17
cycBP	$\mathcal{L}_{\mathrm{KL}}$	0.014	0.48
cycBP$_\mathrm{e}$	$\mathcal{L}_{\mathrm{Bethe}}$	0.030	0.11
cycBP	$\mathcal{L}_{\mathrm{Bethe}}$	0.027	0.027

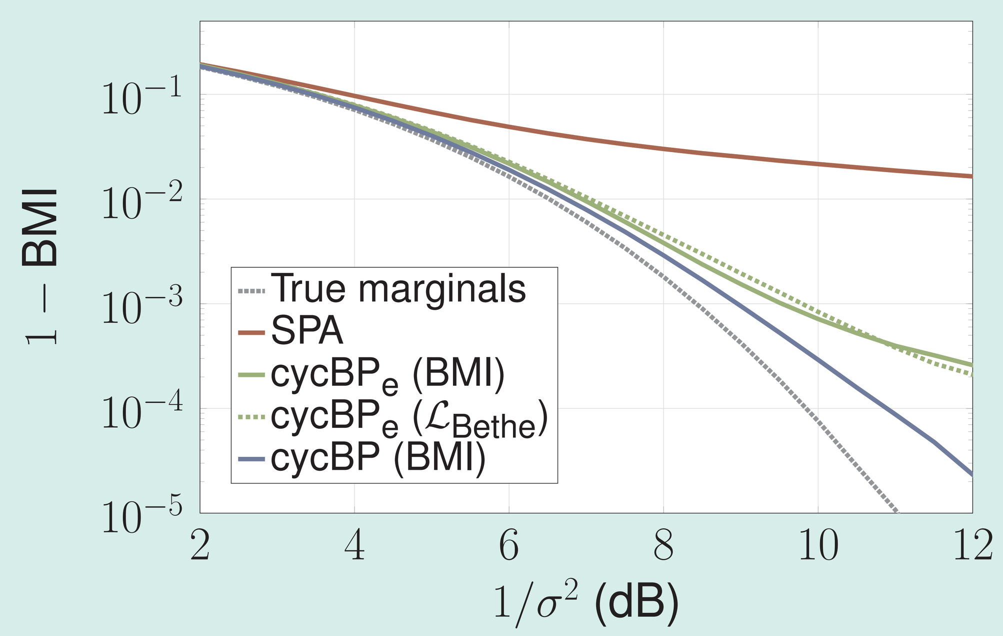


- **Generalizability**: Evaluate pretrained mappings on graphs with constant parameters $E_{nm}=E$ and $\theta_n=\theta$



7. Example 2: Symbol Detection

- **Digital transmission** of binary symbols $x_n\in\{\pm1\}$ over additive white Gaussian noise (AWGN) channel with linear inter-symbol interference
- Efficient symbol detection by message passing on factor graphs



- **BMI**: bitwise mutual information
- σ^2: Variance of the AWGN