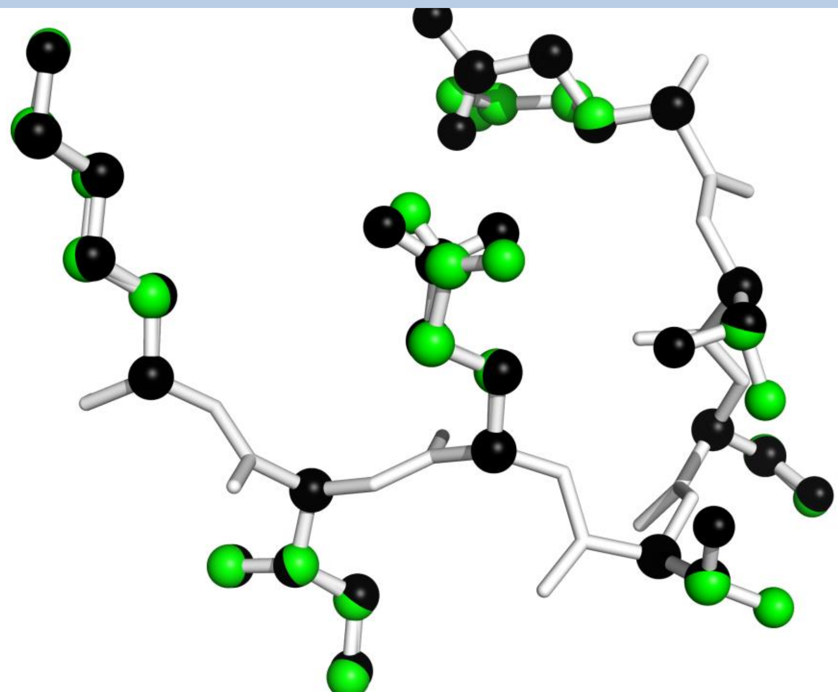
# Proteins, Particles, and Pseudo-Max-Marginals: A Submodular Approach

**Jason Pacheco**
**Erik Sudderth**

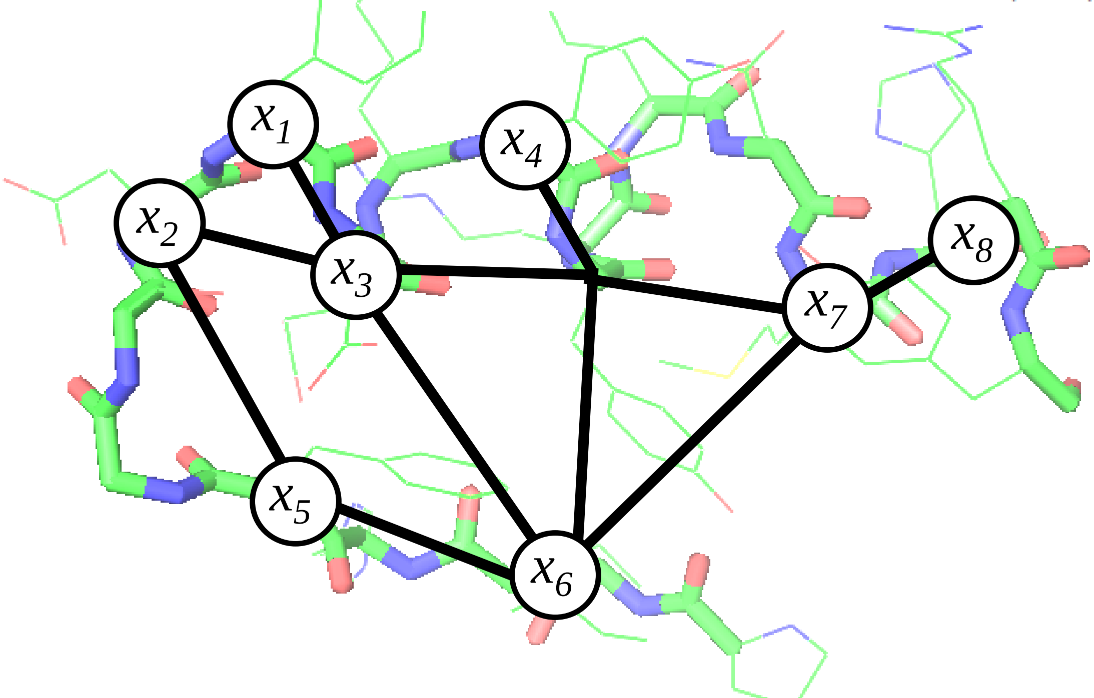
*Department of Computer Science*
*Brown University, Providence RI*

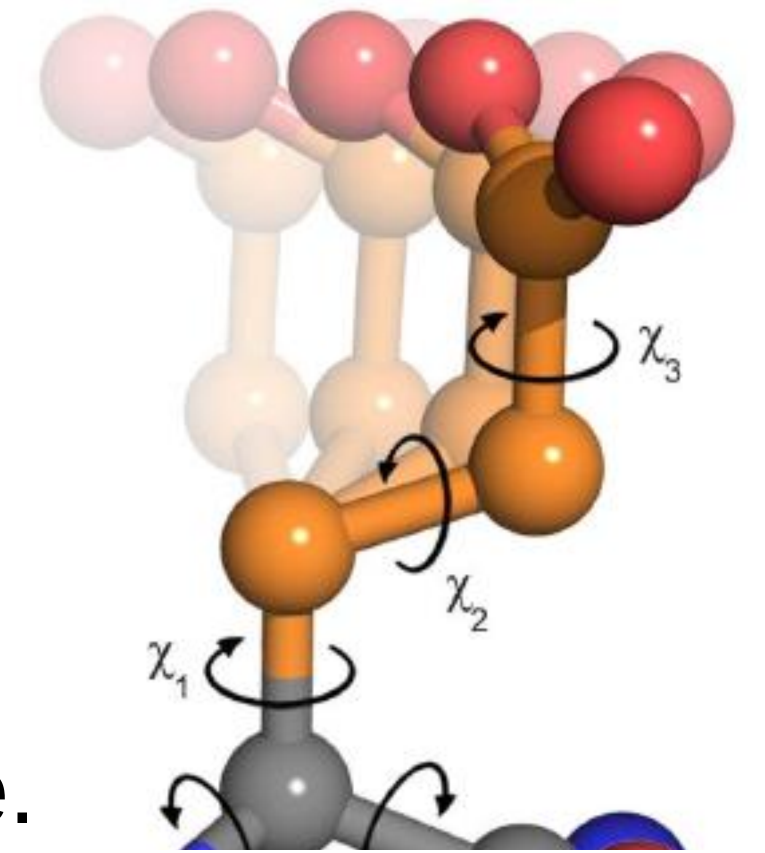


# Protein Side Chain Prediction

Estimate side chains from backbone.

$$p(x) \propto \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t)$$



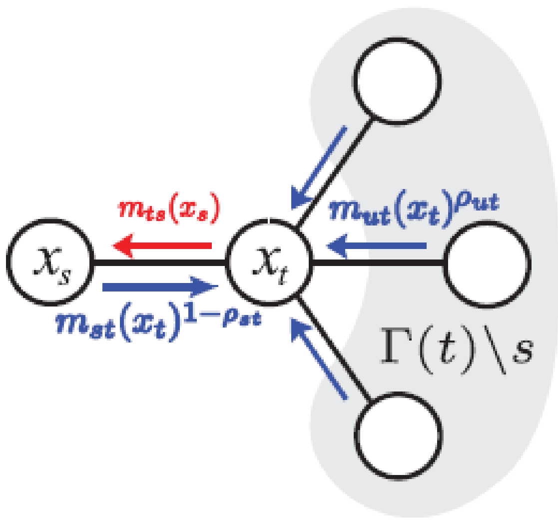



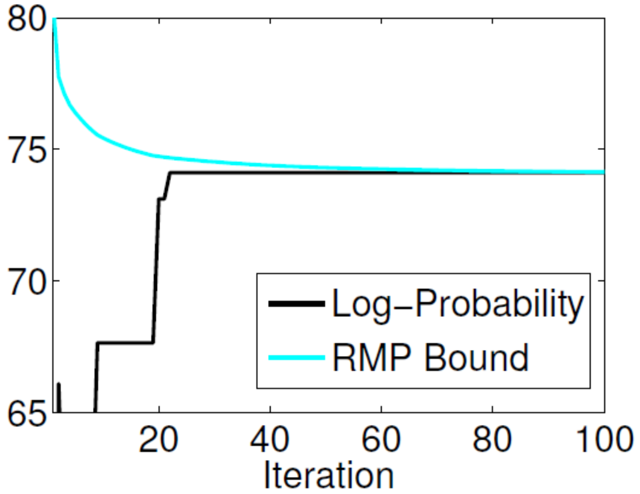

1D – 4D Continuous state.

[ Image: Harder et al., BMC Informatics 2010 ]

# Reweighted Max-Product (RMP)

## Message passing on discrete side chains




Max-marginal: $\mu_s(x_s) \propto \max_{\{x' : x'_s = x_s\}} p(x')$

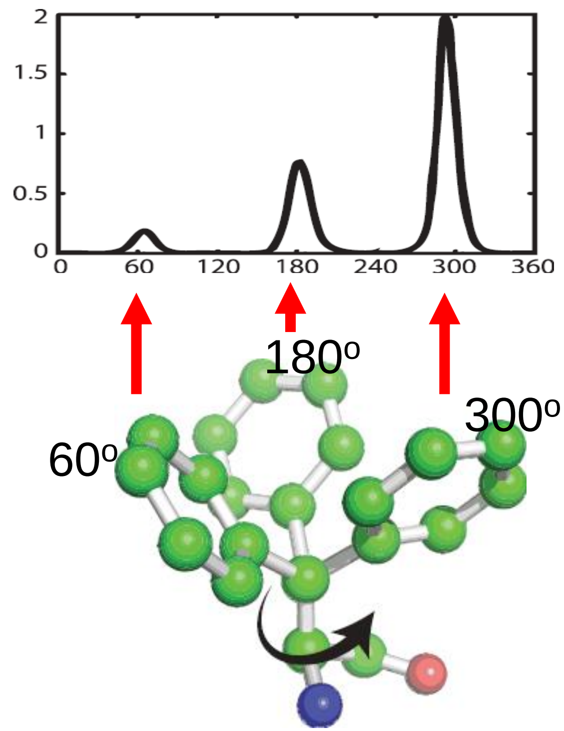
Pseudo-max-marginal: $\nu_s(x_s) \propto \psi_s(x_s) \prod_{u \in \Gamma(s)} m_{us}(x_s)^{\rho_{us}}$

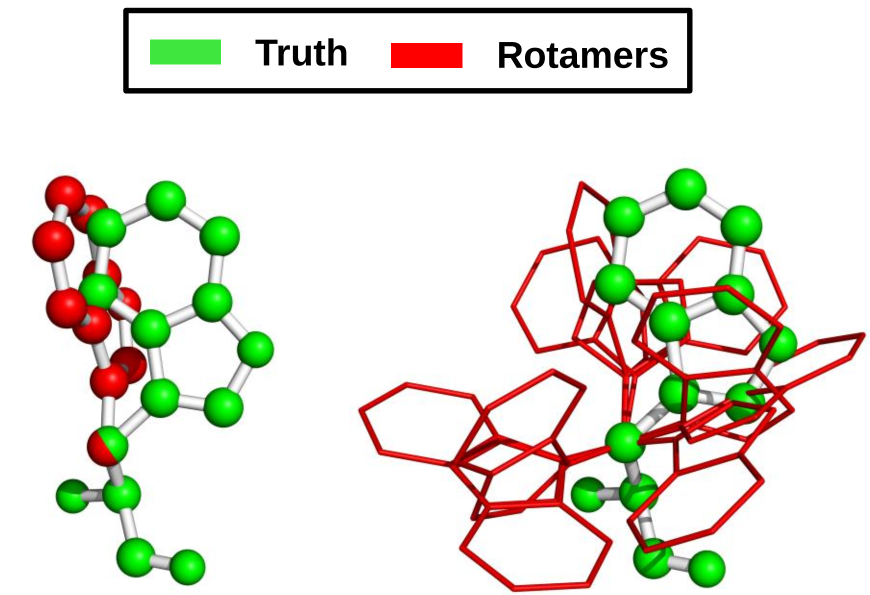
**Edge Appearance Probability**

# Rotamer discretization

## Fit to side chain marginal statistics

**Rotamers**



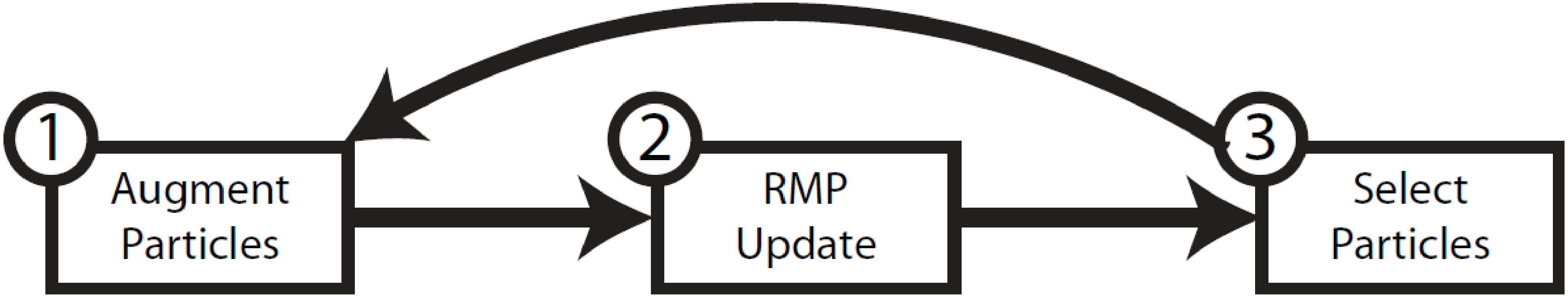




## Fails to capture side chain placement…

Penicillin Acylase Complex, Trp154 [Shapovalov & Dunbrack 2007]

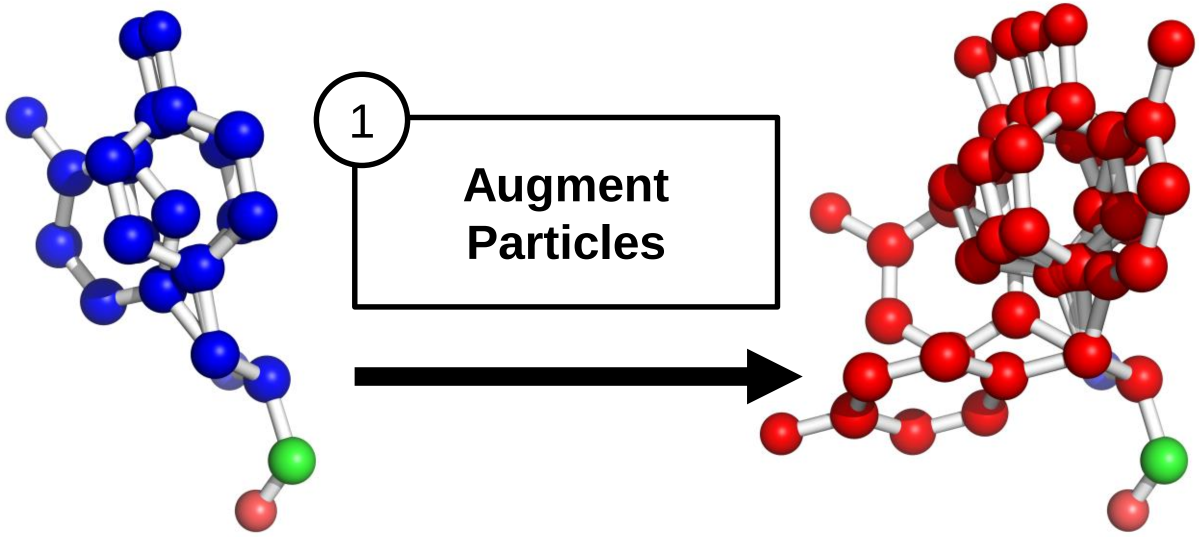
# Particle Max-Product (PMP)

Latent space is continuous…




…particle approximation of continuous RMP messages.

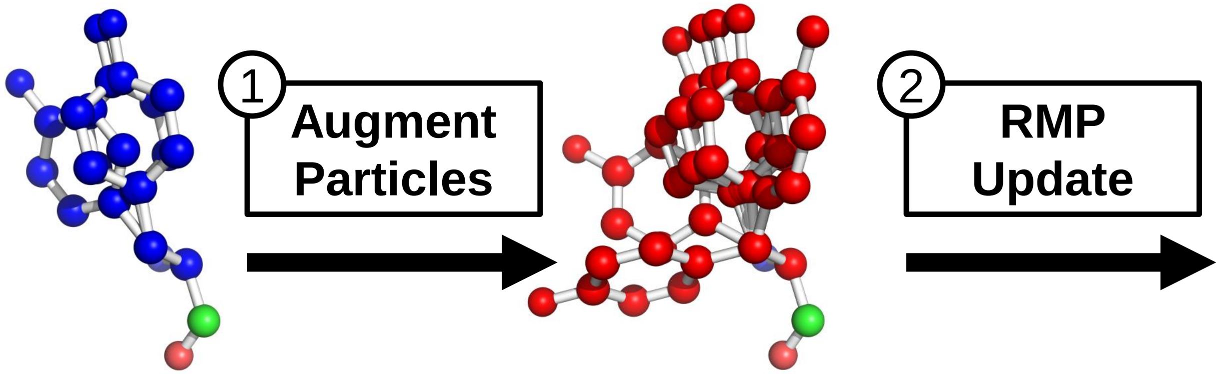
# Particle Max-Product (PMP)

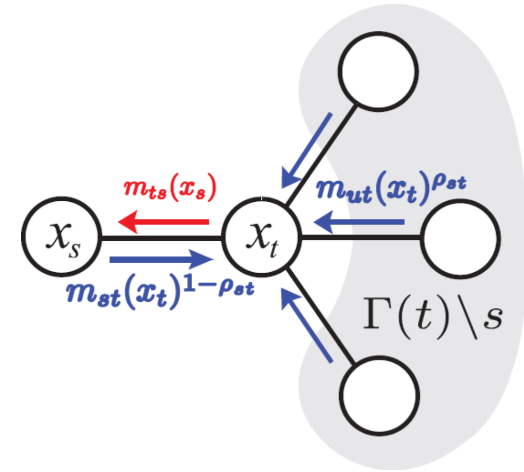



Sample new particles from proposals:

*( Random Walk, Likelihood, Neighbor, … )*

# Particle Max-Product (PMP)

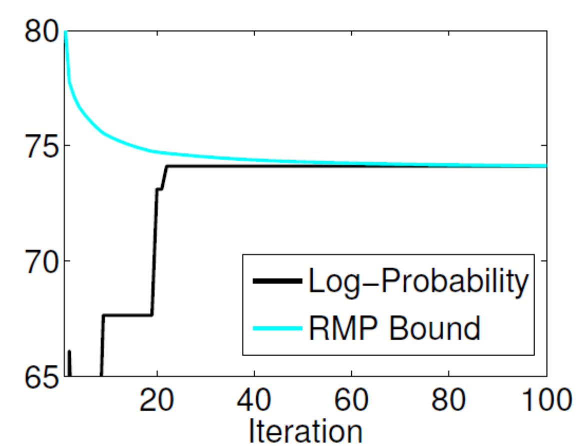


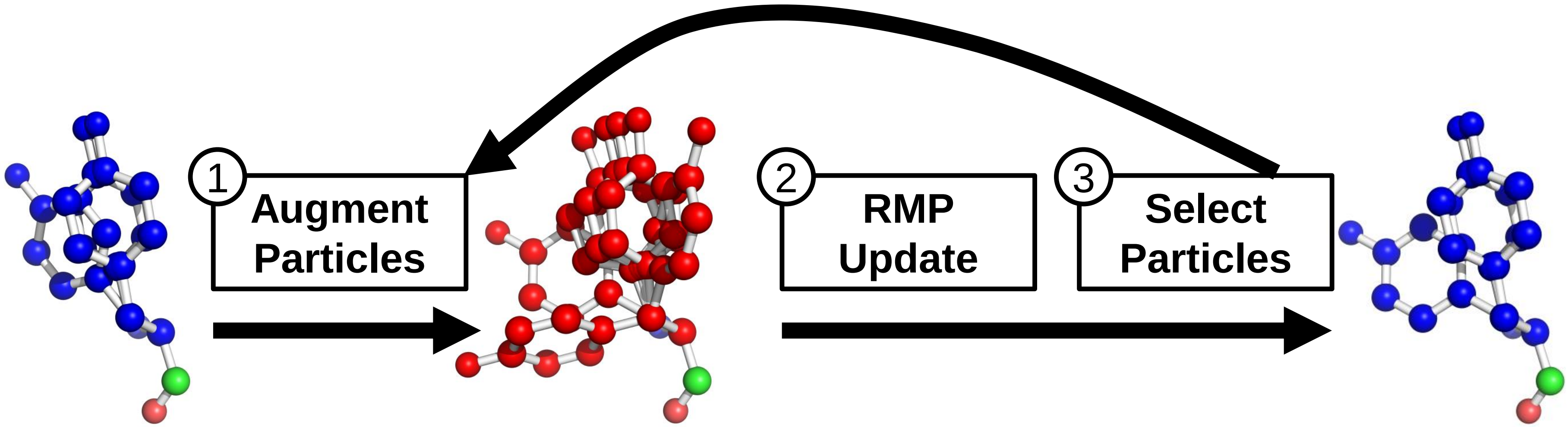

$$m_{ts}(x_s) = \max_{x_t} \psi_t(x_t)\psi_{st}(x_s, x_t)^{\frac{1}{\rho_{st}}} \frac{\prod_{u\in\Gamma(t)\setminus s} m_{ut}(x_t)^{\rho_{ut}}}{m_{st}(x_t)^{1-\rho_{st}}}$$

**Edge Appearance Probability**

Update RMP messages on augmented particles.

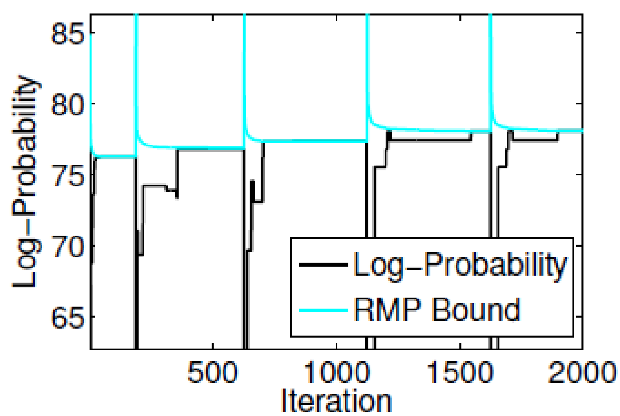


# Particle Max-Product (PMP)



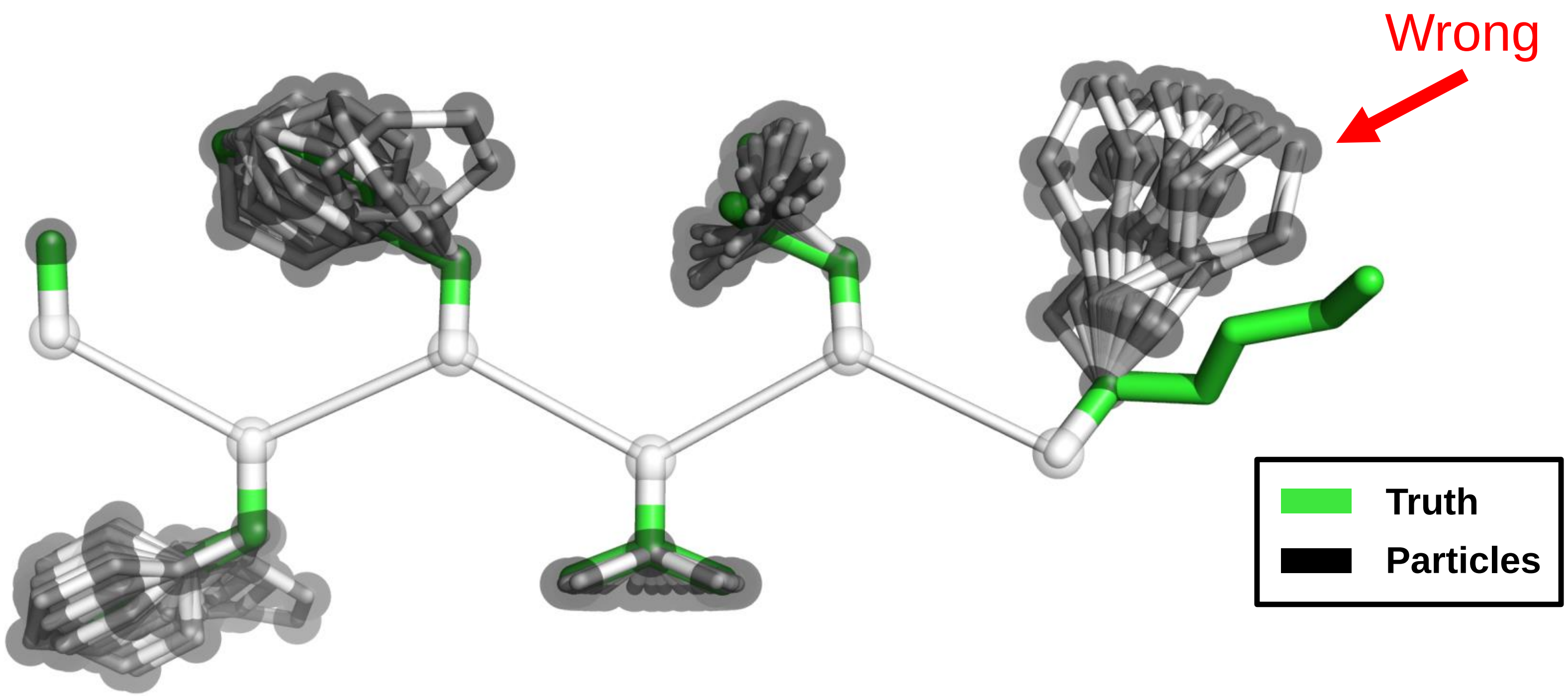

Select subset of good particles…

…Need particle selection method.



# Greedy PMP (G-PMP)

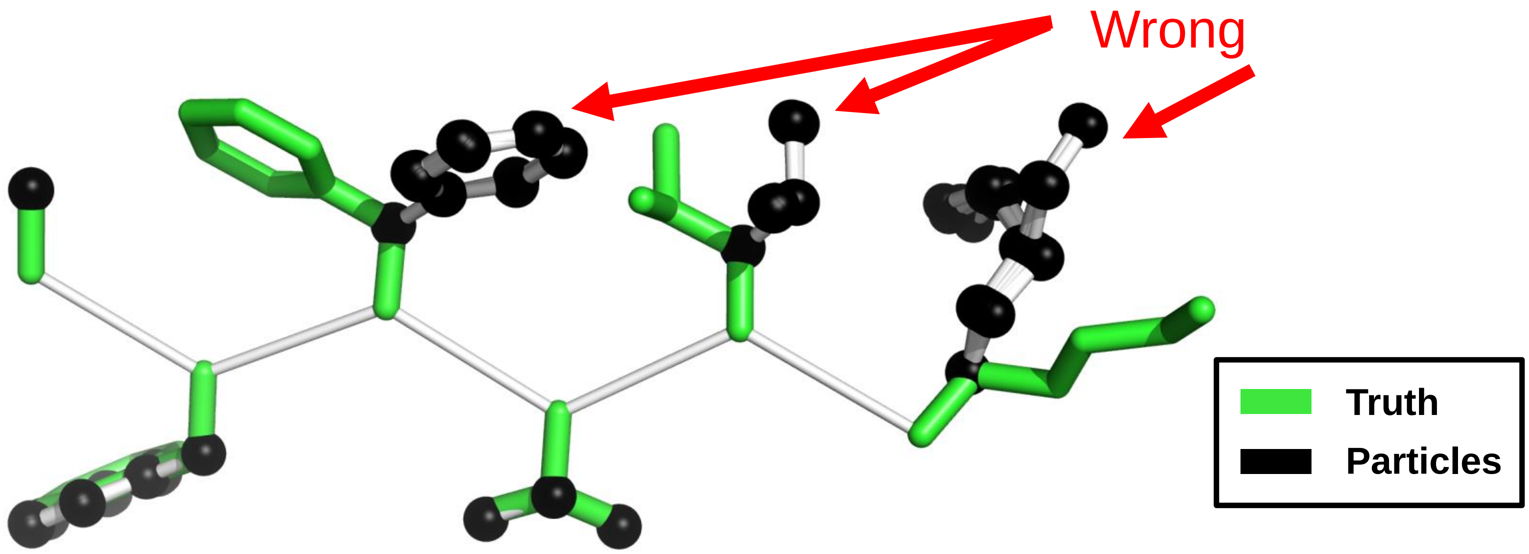
Select best particle, sample from random walk Gaussian. [ Trinh '09, Peng '11 ]




Naïve proposals do not exploit model.

# Top-N PMP (T-PMP)

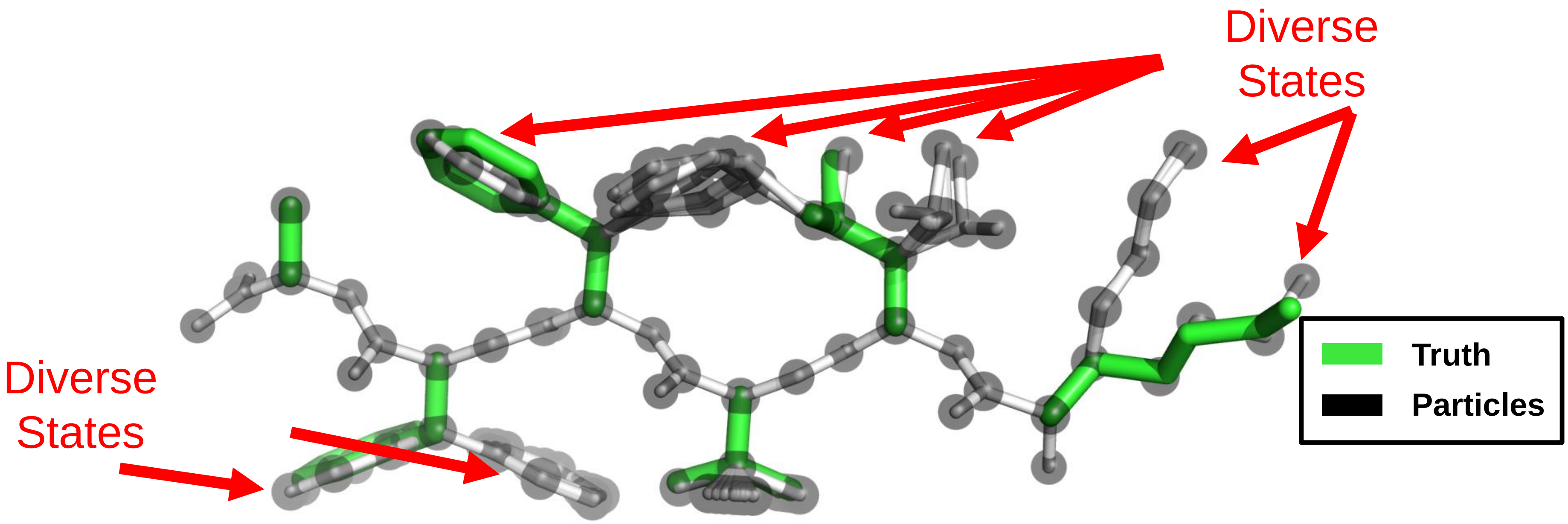
Select N-best particles ranked by pseudo-max-marginal values. [ Besse '12, Pacheco '14 ]




Particles collapse to single solution.

# Diverse PMP (D-PMP)

Select particles to preserve messages.



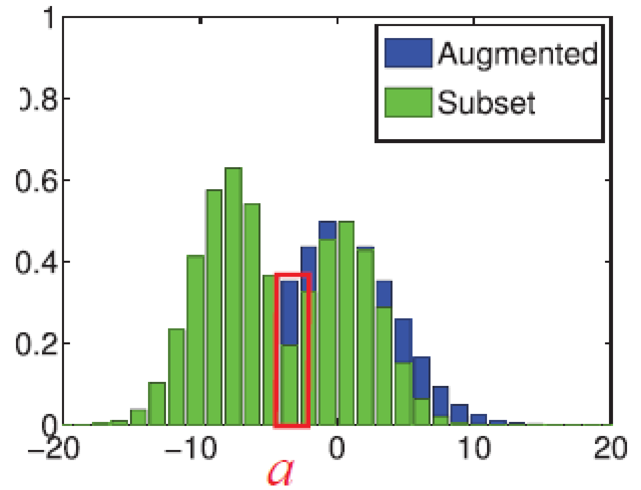

- ➢Encourages particle diversity
- ➢Robust to initialization

# Diverse Particle Selection

At node $t \in \mathcal{V}$ select particles to minimize maximum outgoing message error:

$$\min_{z} \max_{s \in \Gamma(t), a} m_{ts}(a) - \hat{m}_{ts}(a; z)$$

**Binary Selection Vector**



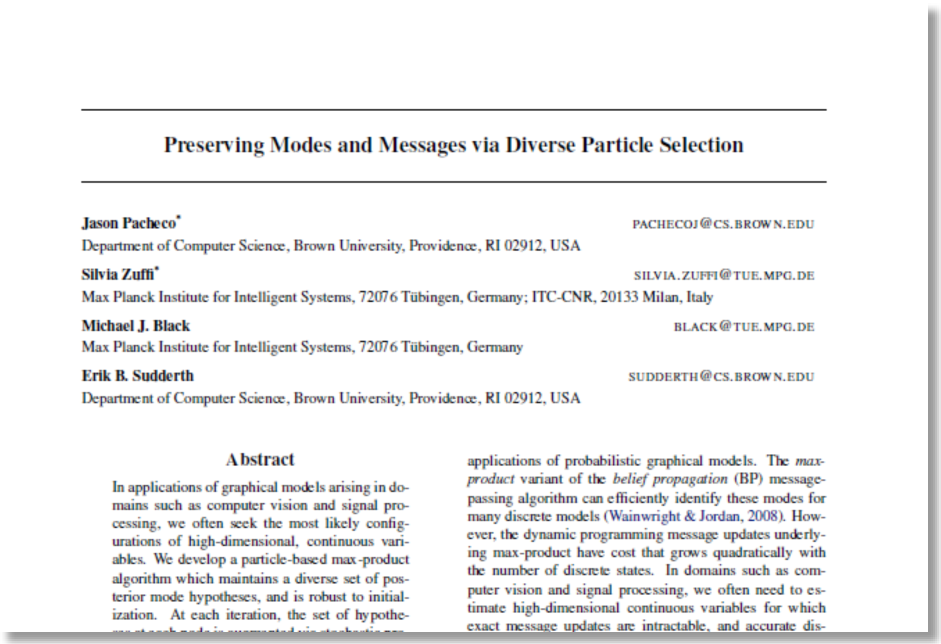

RMP message over subset:

$$\hat{m}_{ts}(a; z) = \max_{b} z(b) \psi_t(b) \psi_{st}(a, b)^{\frac{1}{\rho_{st}}} \frac{\prod_{u \in \Gamma(t) \setminus s} m_{ut}(b)^{\rho_{ut}}}{m_{st}(b)^{1-\rho_{st}}}$$

Approximate IP with greedy algorithm.

# Diverse Particle Selection

## Pacheco et al. ICML 2014



**Preserving Modes and Messages via Diverse Particle Selection**


**Jason Pacheco*** PACHECOJ@CS.BROWN.EDU
Department of Computer Science, Brown University, Providence, RI 02912, USA
**Silvia Zuffi*** SILVIA.ZUFFI@TUE.MPG.DE
Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany; ITC-CNR, 20133 Milan, Italy
**Michael J. Black** BLACK@TUE.MPG.DE
Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany
**Erik B. Sudderth** SUDDERTH@CS.BROWN.EDU
Department of Computer Science, Brown University, Providence, RI 02912, USA

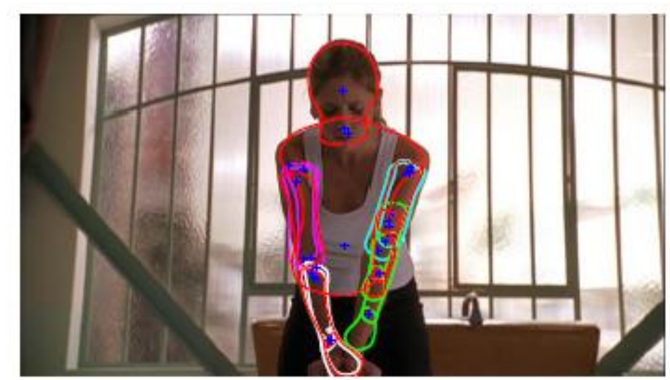


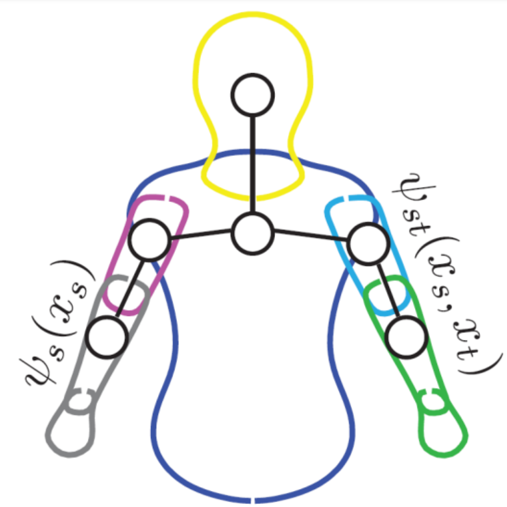
**Abstract**

In applications of graphical models arising in domains such as computer vision and signal processing, we often seek the most likely configurations of high-dimensional, continuous variables. We develop a particle-based max-product algorithm which maintains a diverse set of posterior mode hypotheses, and is robust to initialization. At each iteration, the set of hypothe-


applications of probabilistic graphical models. The *max-product* variant of the *belief propagation* (BP) message-passing algorithm can efficiently identify these modes for many discrete models (Wainwright & Jordan, 2008). However, the dynamic programming message updates underlying max-product have cost that grows quadratically with the number of discrete states. In domains such as computer vision and signal processing, we often need to estimate high-dimensional continuous variables for which exact message updates are intractable, and accurate dis-

Pose Estimation






- ➢Good empirical results
- ➢Difficult to analyze
- ➢Limited to tree-structured MRFs

# Diverse Particle Selection

Equivalent to minimizing $L_\infty$ norm.

Consider other norms, e.g. $L_1$ :

$$\begin{aligned}&\underset{z}{\text{minimize}} \sum_{s\in\Gamma(t)} \|m_{ts} - \hat{m}_{ts}(z)\|_1 \\ &\text{subject to} \quad \|z\|_1 \leq N, \;\; z \in \{0,1\}^{\alpha N}\end{aligned}$$

Easier to analyze…

**<u>Property 1:</u> Message error upper bounds pseudo-max-marginal error:**

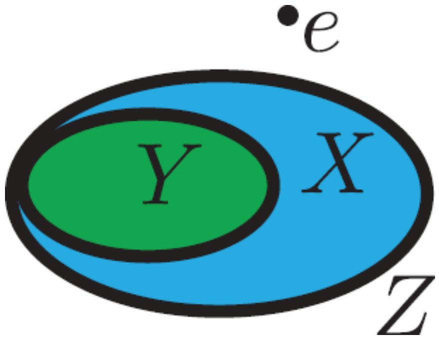
$$\|\nu_s - \hat{\nu}_s\|_1 \leq \sum_{t\in\Gamma(s)} \|m_{ts} - \hat{m}_{ts}\|_1^{\rho_{ts}}$$

# Submodular Particle Selection

**<u>Property 2:</u> Selection IP equivalent to submodular maximization.**

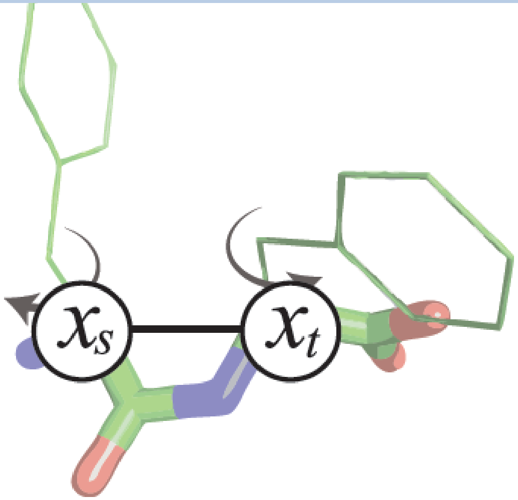
Set function $f : 2^Z \rightarrow \mathbb{R}$ is submodular iff diminishing marginal gains.

$$\underbrace{f(Y \cup \{e\}) - f(Y)}_{\text{Margin}} \geq f(X \cup \{e\}) - f(X)$$




**<u>P. 3:</u> Efficient LazyGreedy selection within $(1 - \frac{1}{e})$ factor of optimal value.**

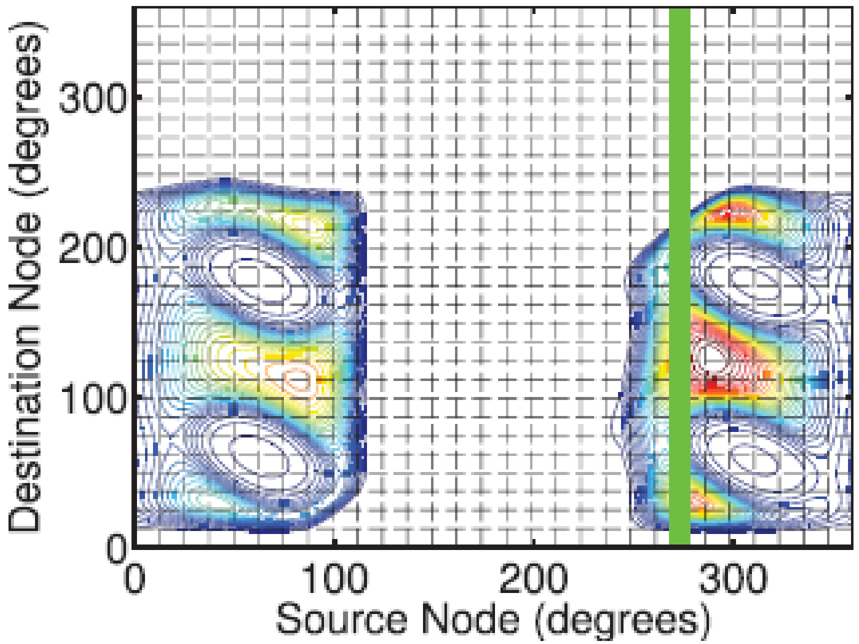
# LazyGreedy Selection



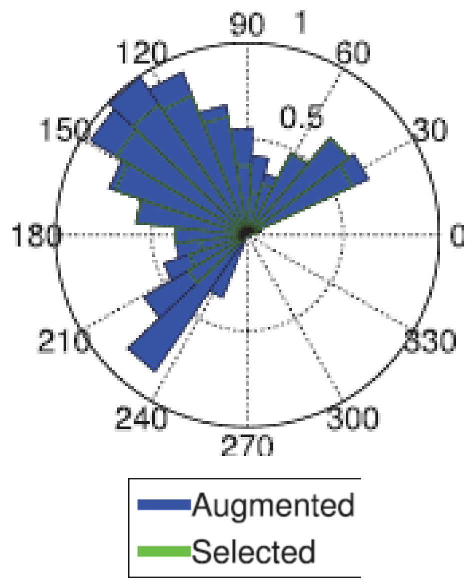

## Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

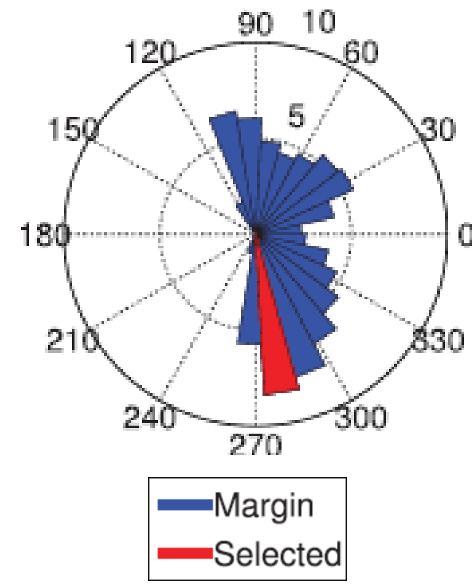
**Joint Probability**



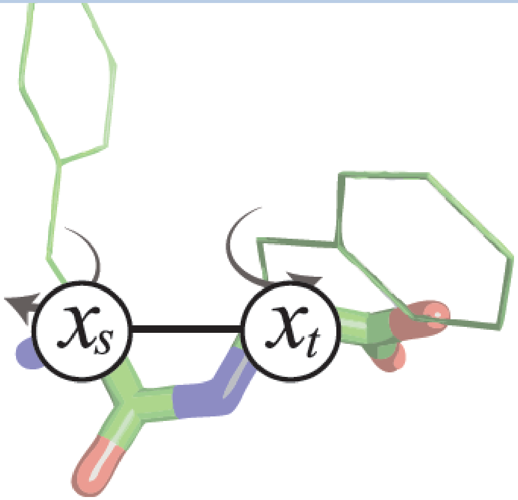

**Message**




**Margin**



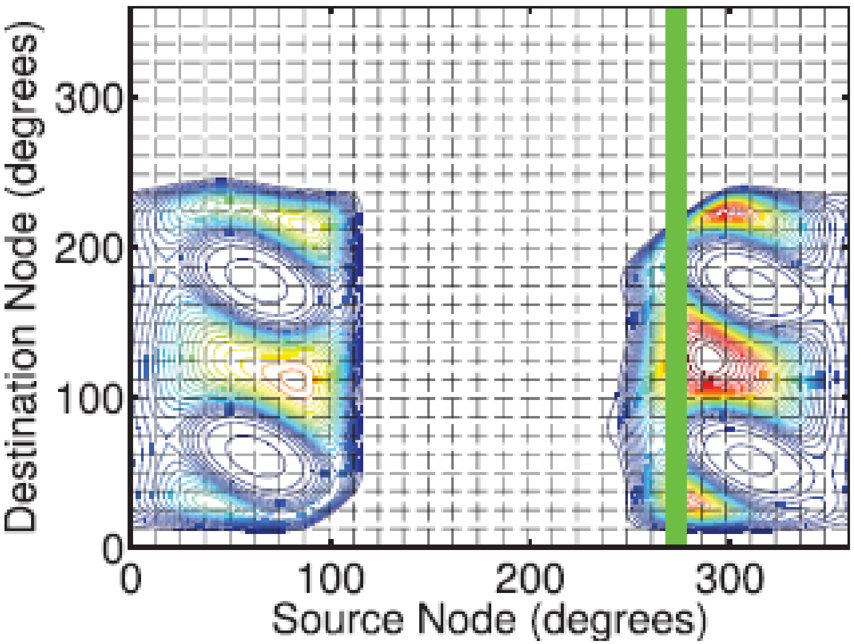
# LazyGreedy Selection



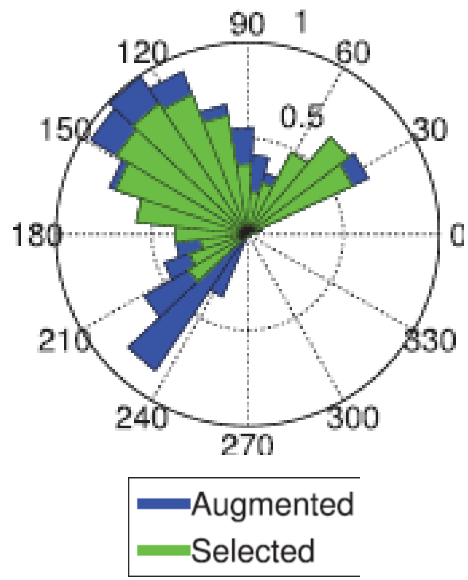

## Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

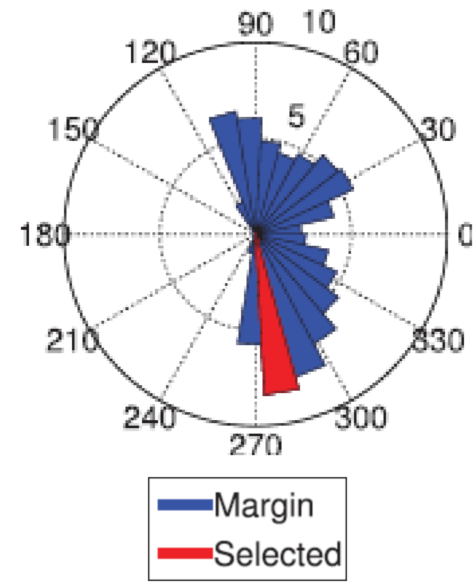
**Joint Probability**



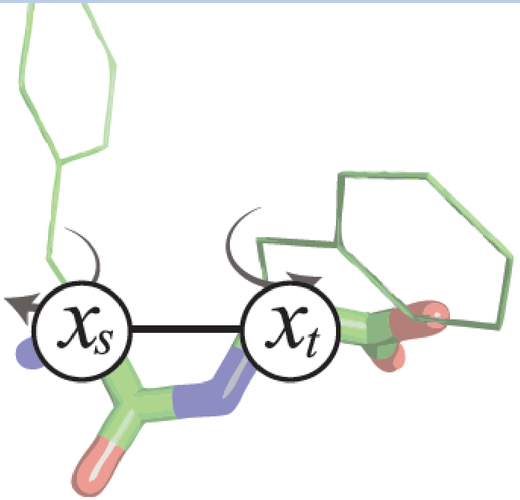

**Message**




**Margin**

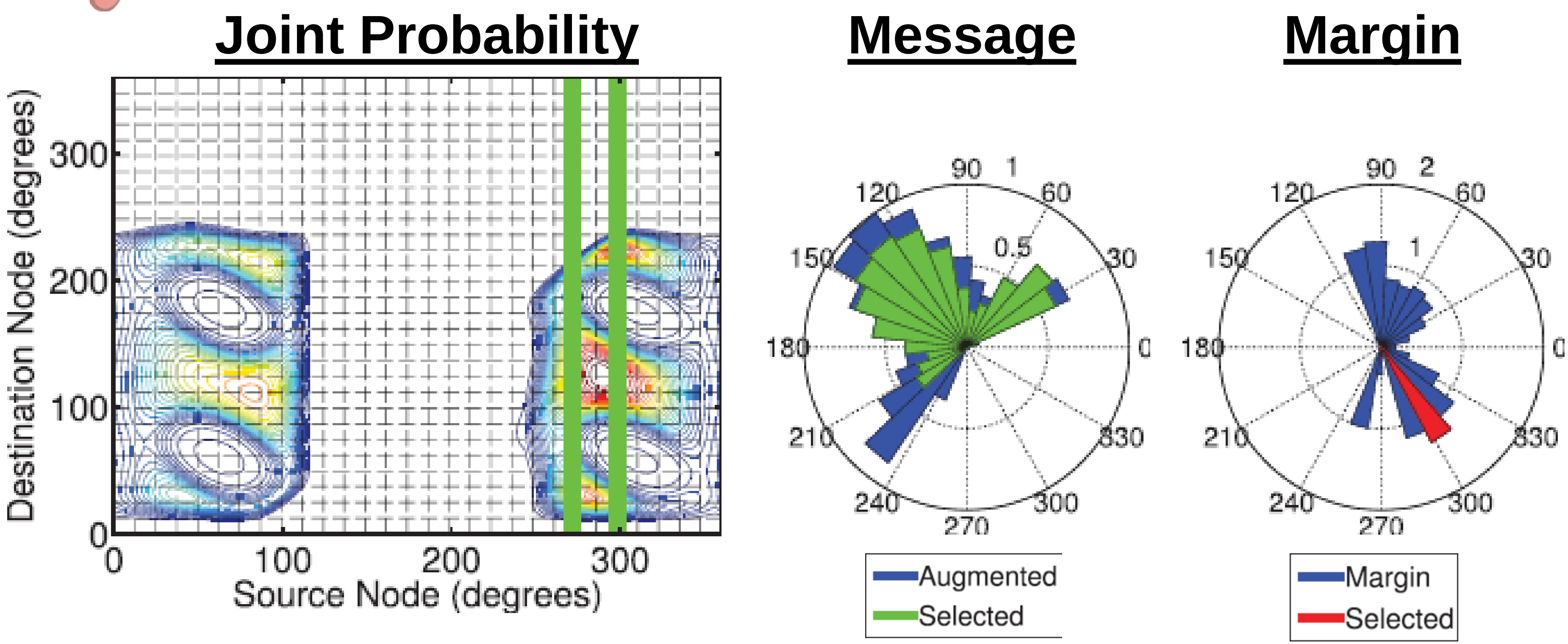


# LazyGreedy Selection

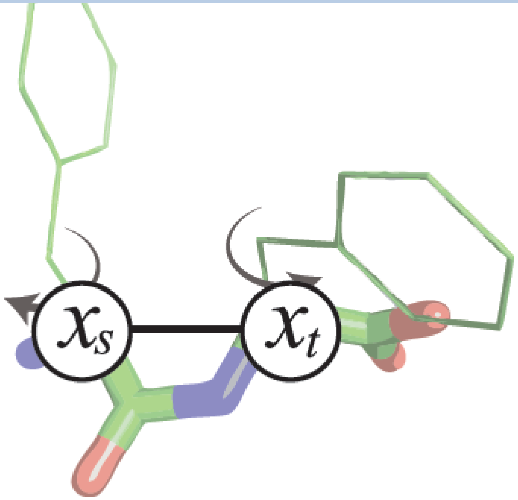



Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$



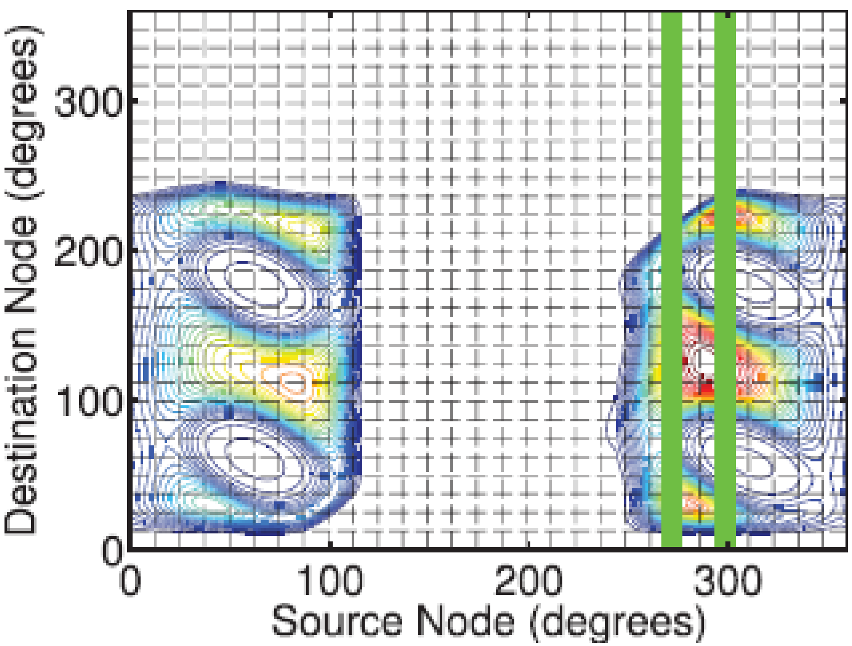
# LAZYGREEDY Selection



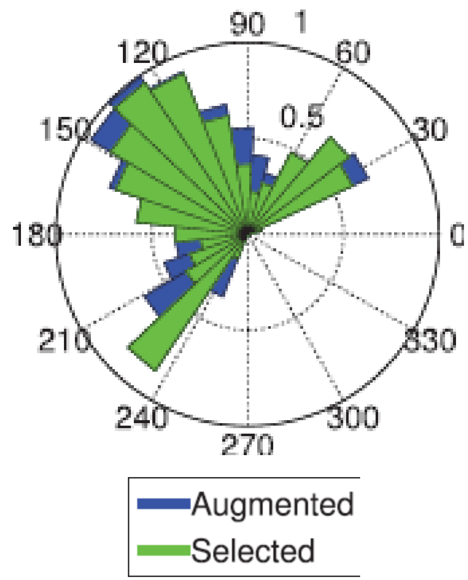

## Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

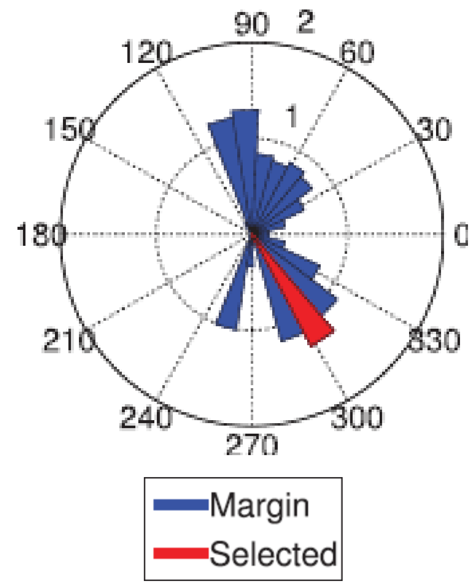
**Joint Probability**



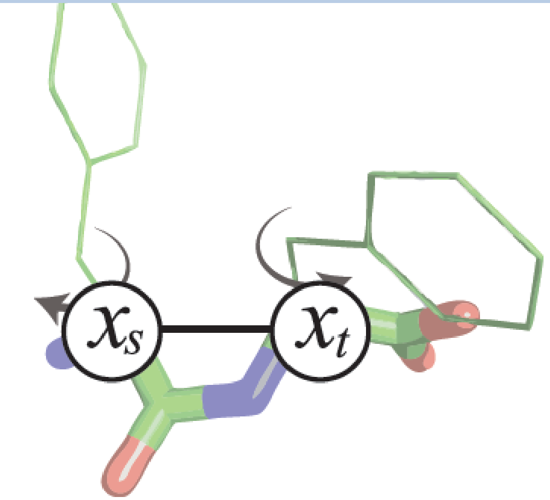

**Message**




**Margin**



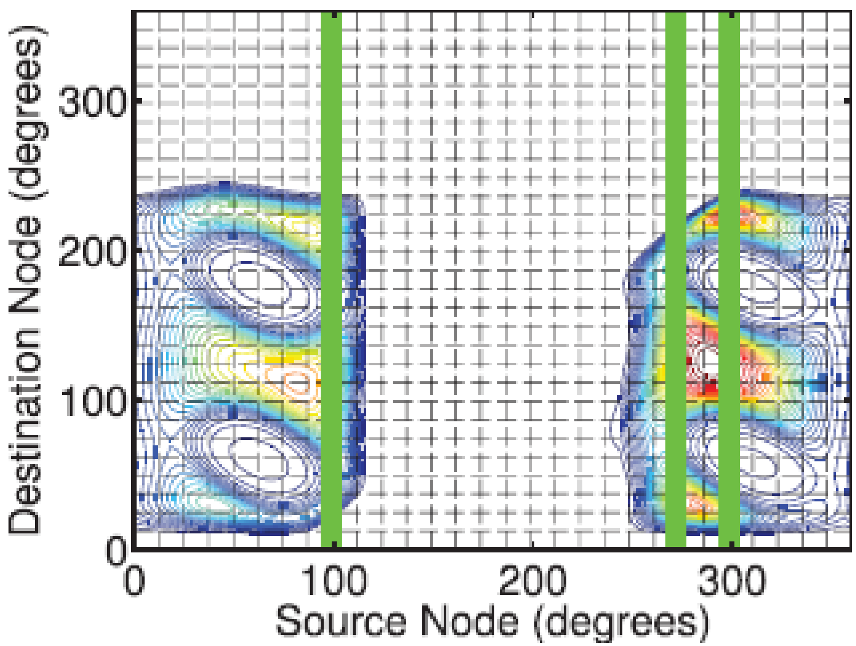
# LAZYGREEDY Selection



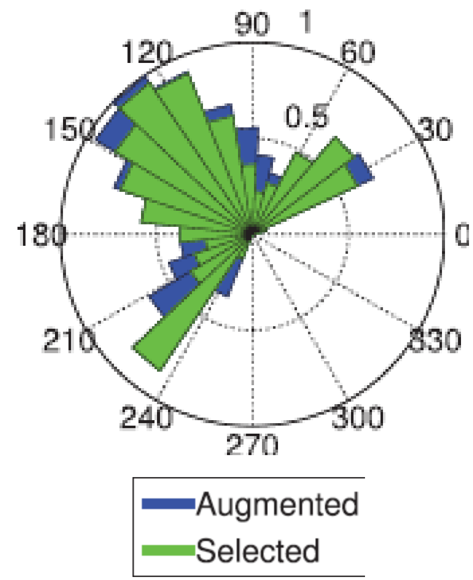

## Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

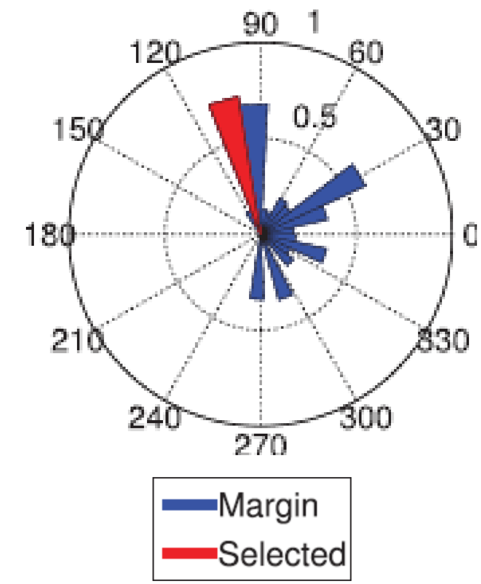
**Joint Probability**



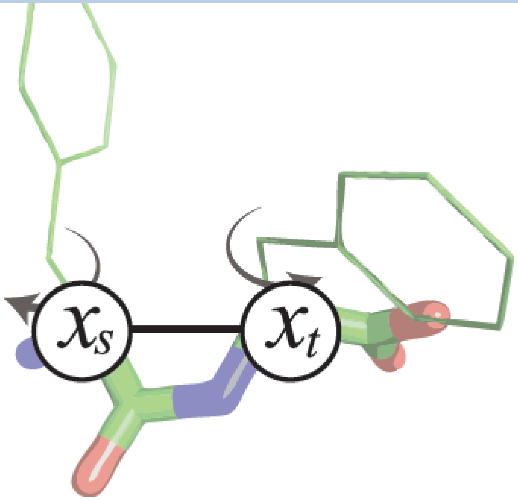

**Message**




**Margin**



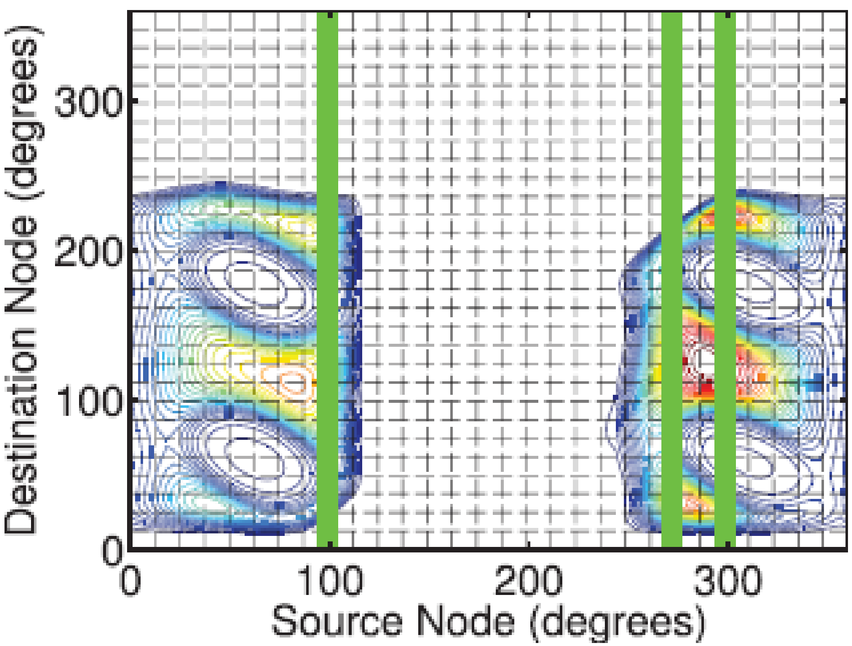
# LazyGreedy Selection



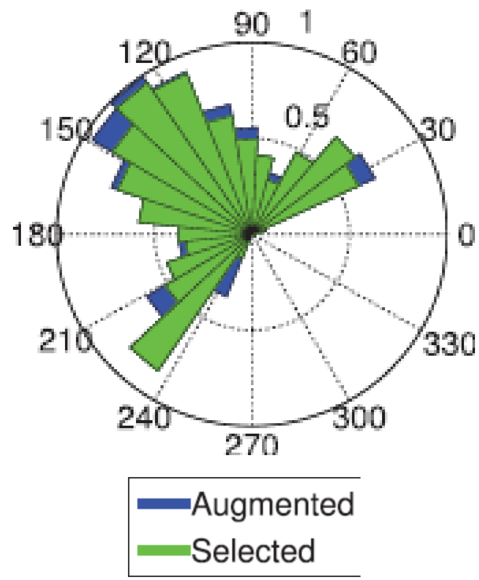

## Selection Objective:

$$\underset{z}{\text{minmize}} \sum_{s\in\Gamma(t)} \|m_{ts} - \hat{m}_{ts}(z)\|_1$$

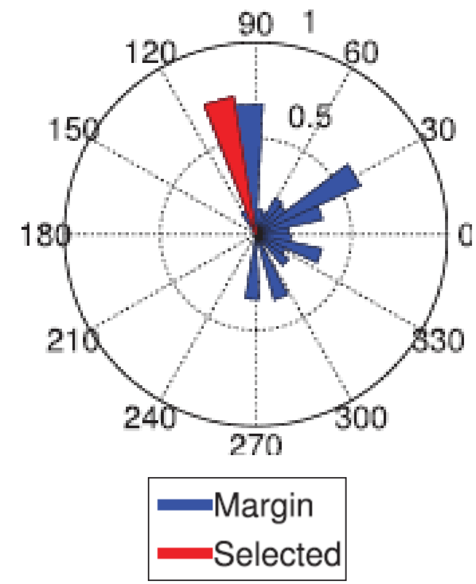
**Joint Probability**




**Message**



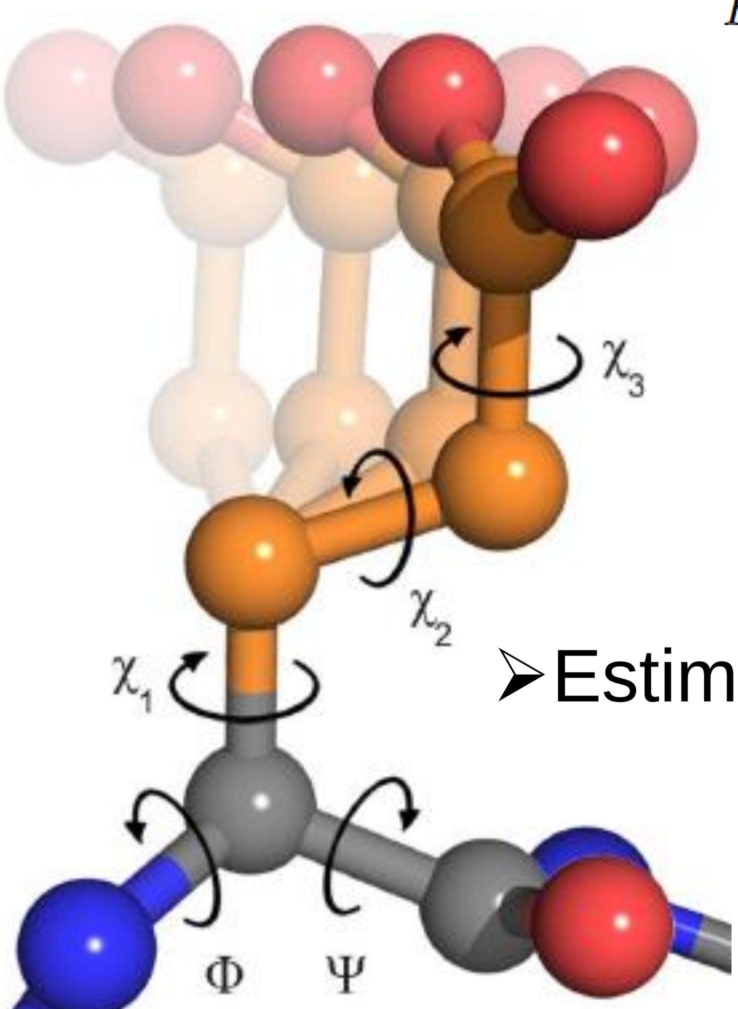

**Margin**



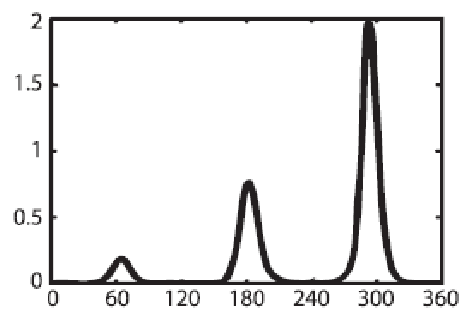
# Protein Side Chain Prediction

Pairwise Markov random field (MRF):

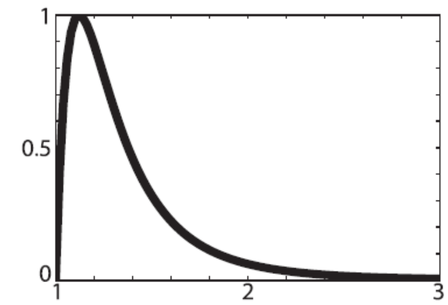
$$p(x) \propto \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t)$$




**Gaussian Mixture**



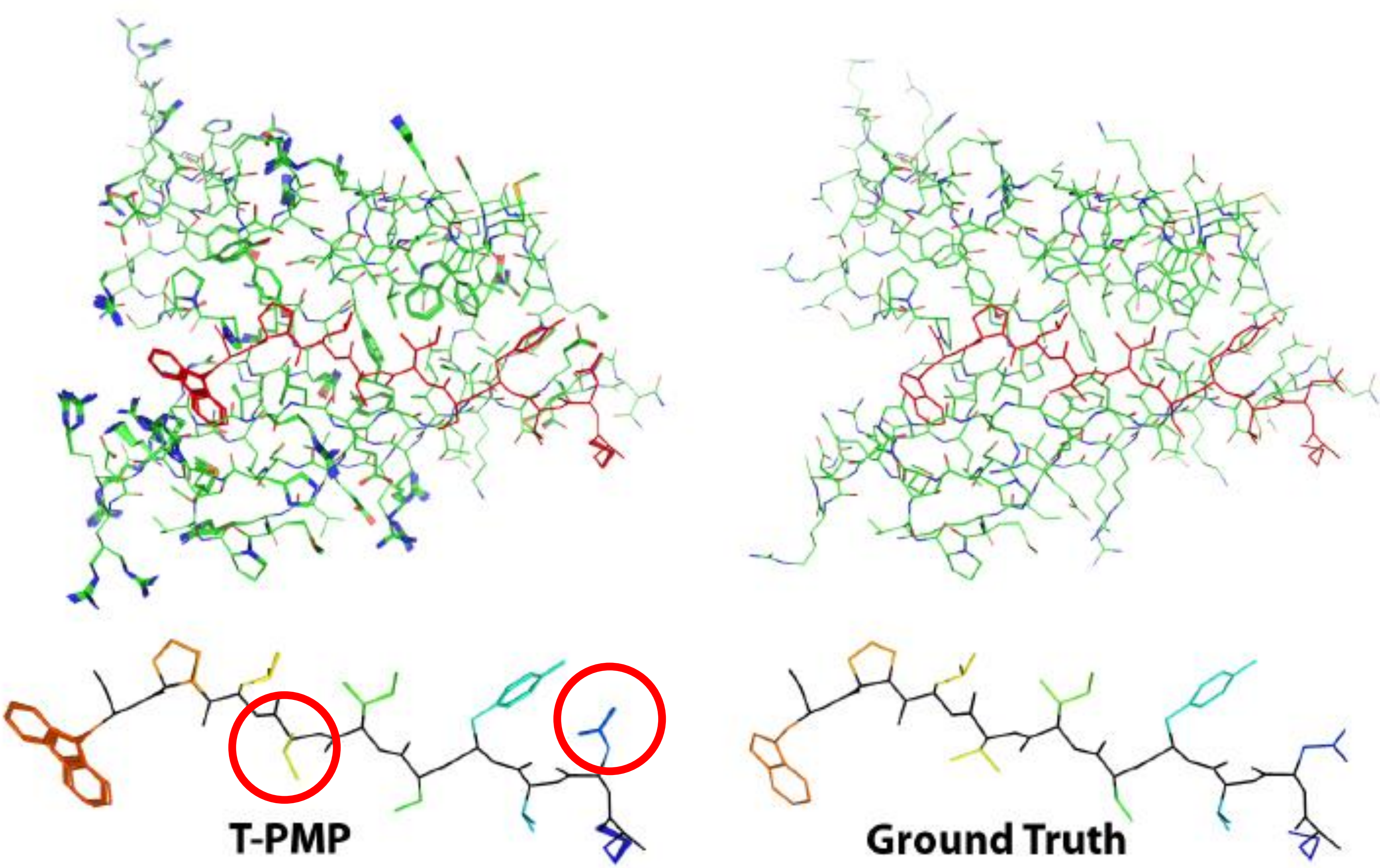

**Lennard-Jones**



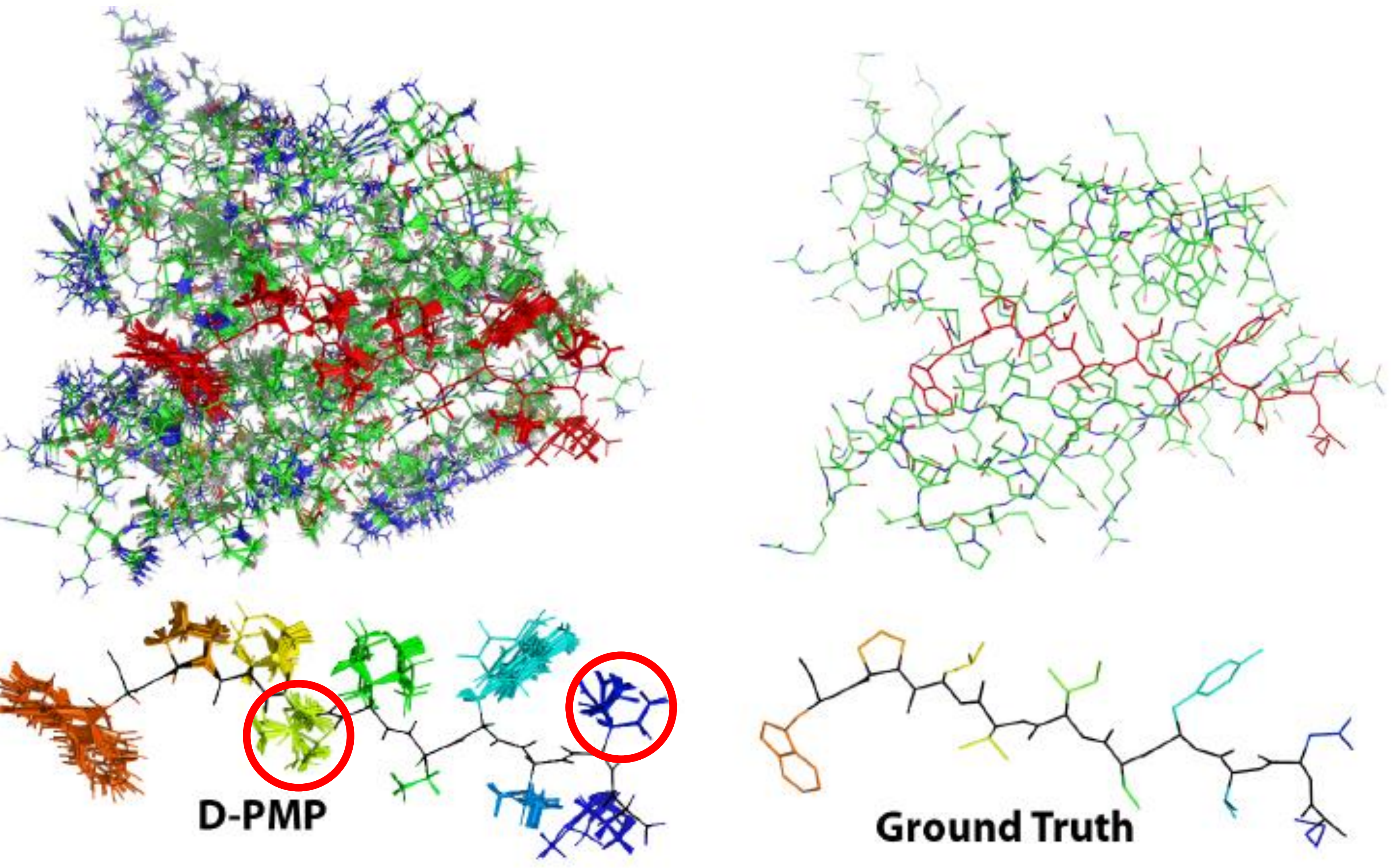

- ➢Estimate side chain for fixed backbone
- ➢1D to 4D continuous states.

[ Image: Harder et al., BMC Informatics 2010 ]

# Protein Side Chain Prediction



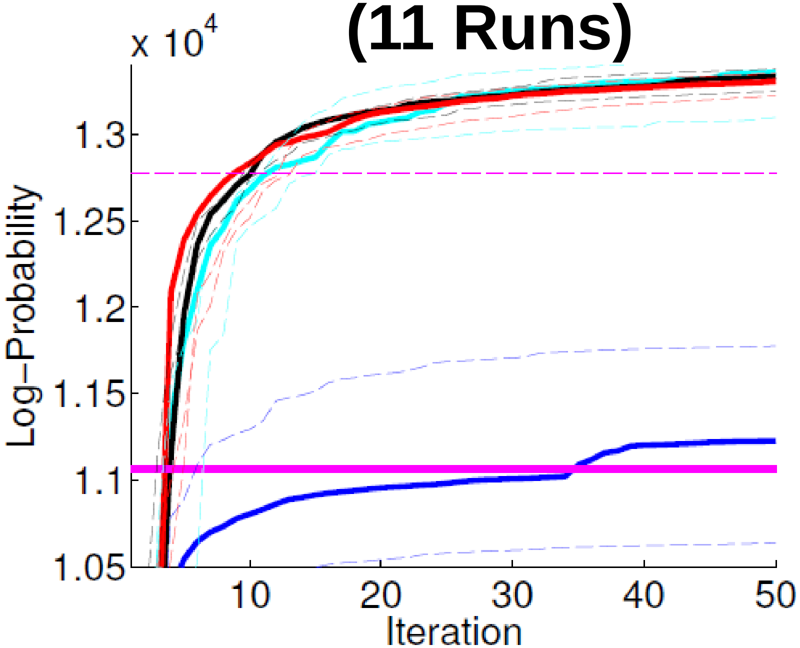
# Protein Side Chain Prediction



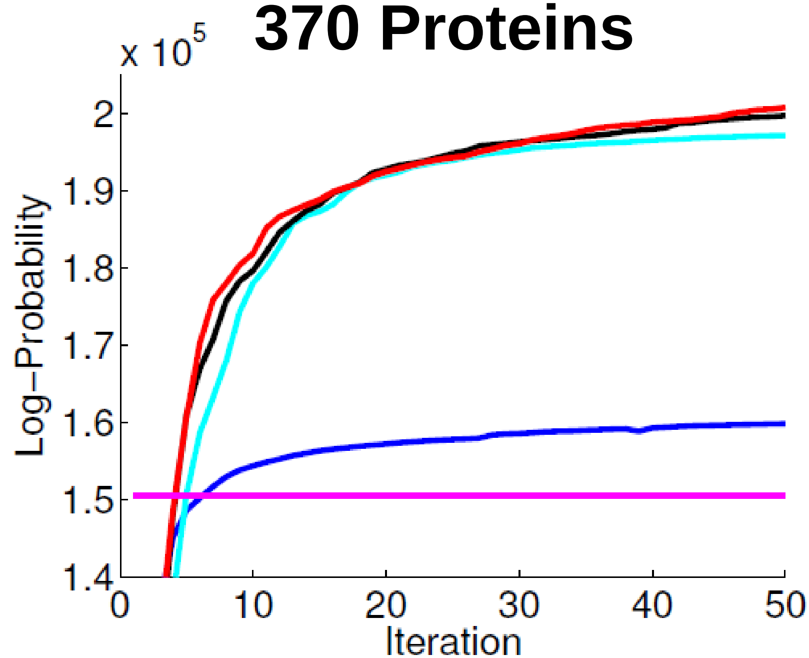
# Protein Side Chain Prediction

## Log-probability of MAP estimate for…

**20 Proteins (11 Runs)**




**370 Proteins**

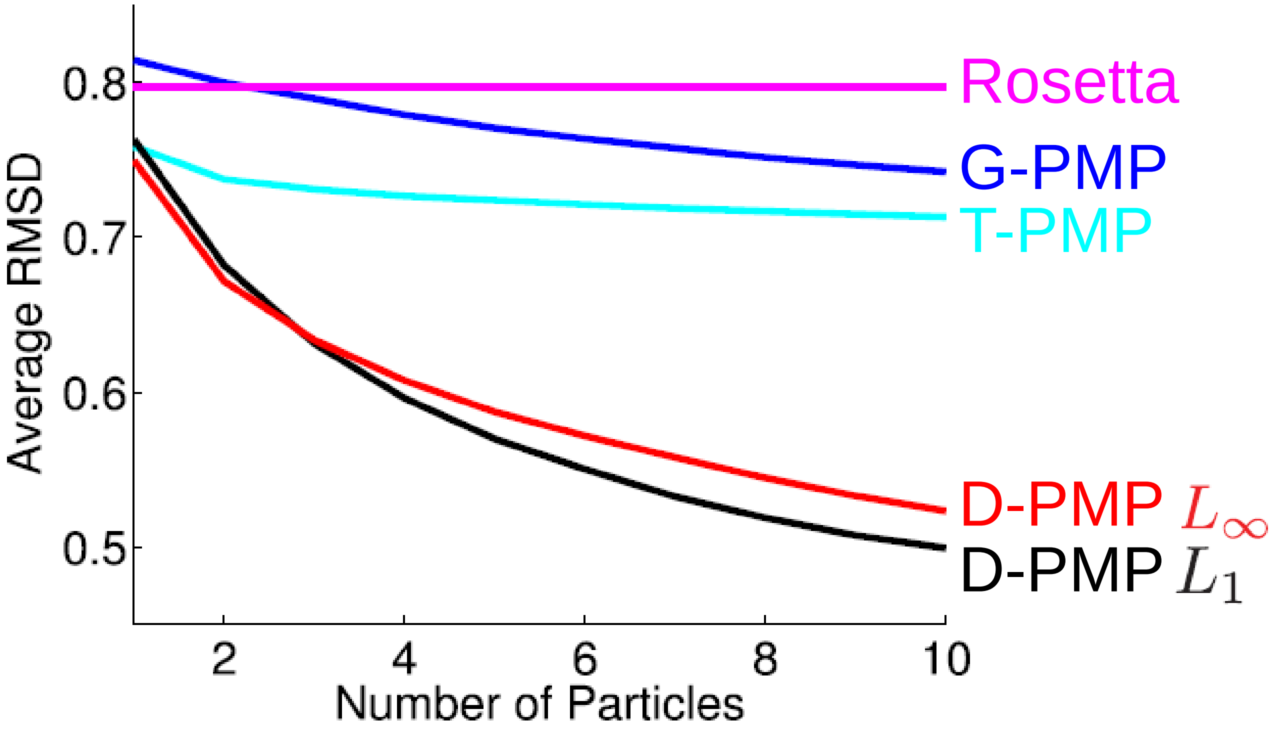



G-PMP, T-PMP, D-PMP $L_1$, D-PMP $L_\infty$

Rosetta simulated annealing [Rohl et al., 2004]

# Protein Side Chain Prediction

Root mean square deviation (RMSD) from x-ray structure.




Oracle selects best configuration in current particle set.

# Optical Flow

Estimate 2D motion for every superpixel.

Middlebury optical flow benchmark

[Baker et al. 2011]

# Optical Flow

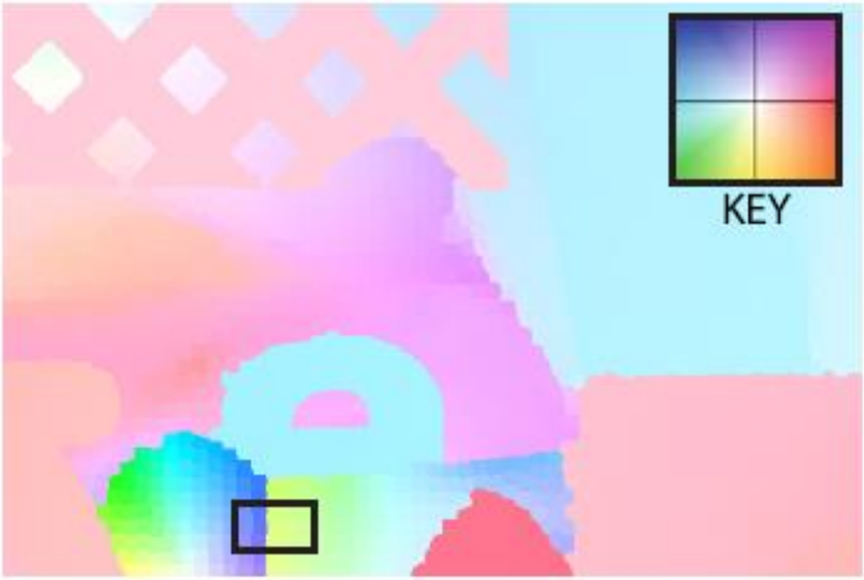
Estimate 2D motion for every superpixel.

Middlebury optical flow benchmark

[Baker et al. 2011]

# Optical Flow

Estimate 2D motion for every superpixel.



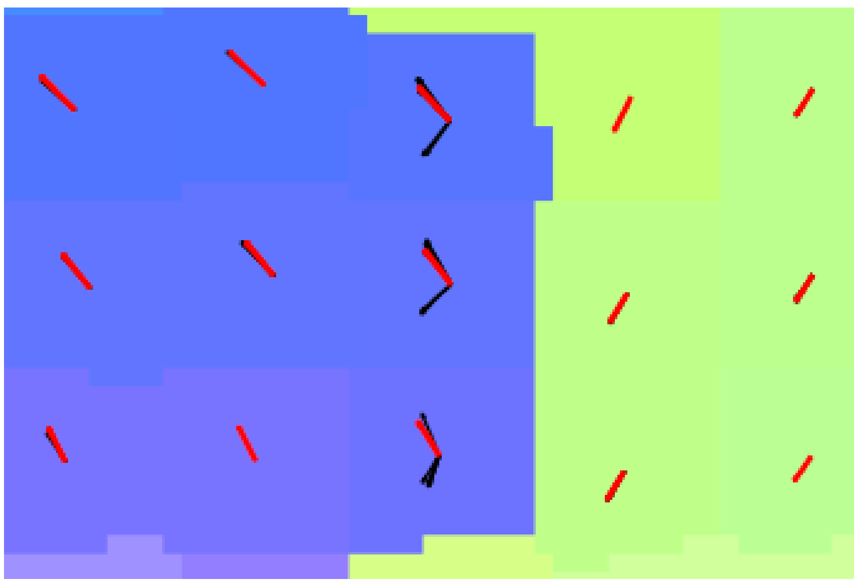

Middlebury optical flow benchmark

[Baker et al. 2011]

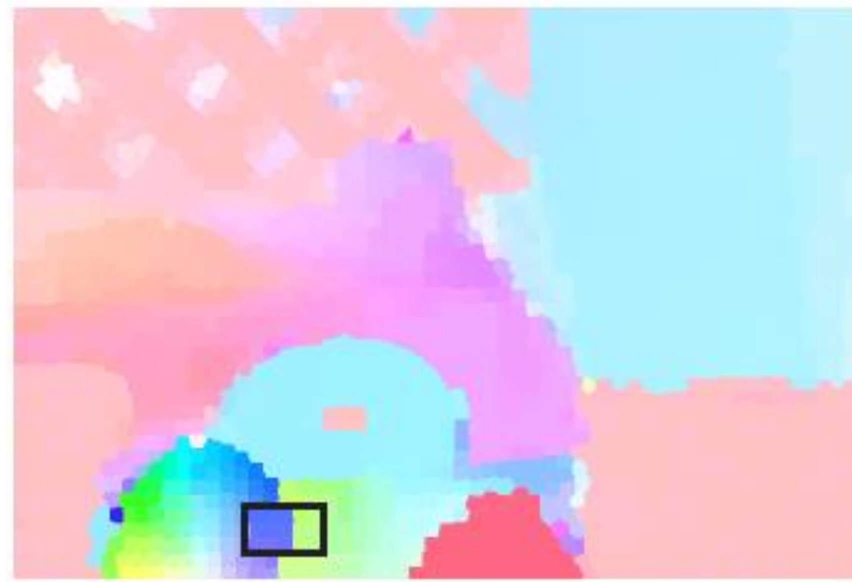
# Optical Flow

Flow ambiguity near object boundaries…

**D-PMP Particles**

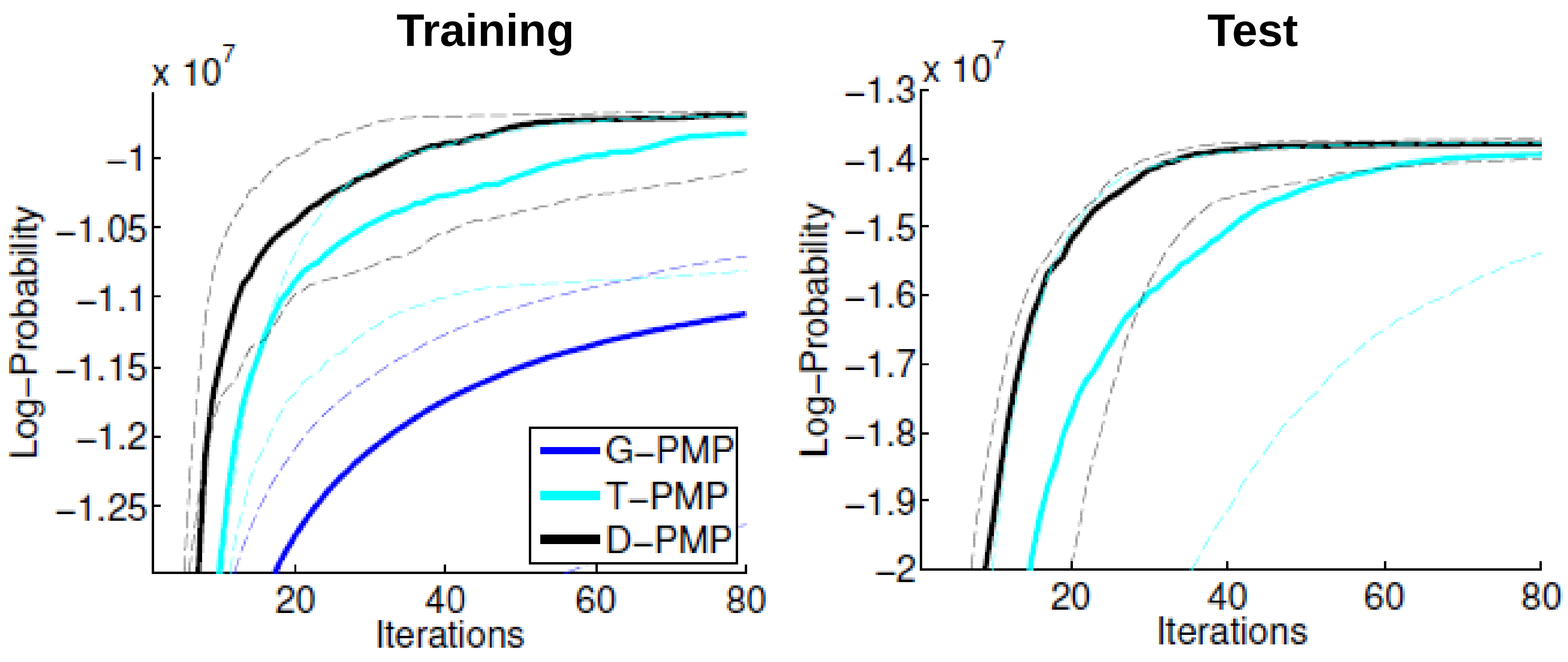


**D-PMP Estimate**



D-PMP particles reflect this.

# Optical Flow

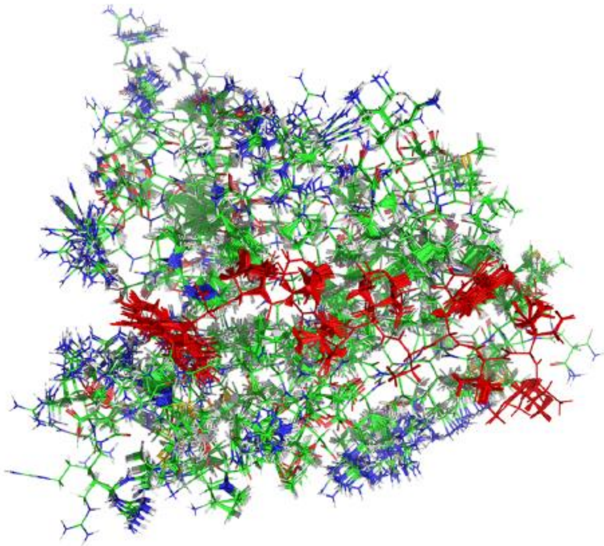


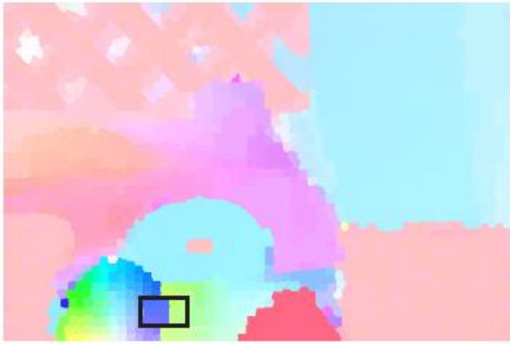

D-PMP accuracy equivalent to Classic-C

[Sun et al. 2014]

# Summary

**General purpose particle-based max-product for continuous graphical models with cycles.**





**Code Available: cs.brown.edu/~pachecoj**

# Diverse Particle Selection

Minimize sum of errors ( $L_1$ ):

Augmented Messages

Subset Messages

Selection Vector

$$\underset{z}{\text{minimize}} \sum_{s \in \Gamma(t)} \|m_{ts} - \hat{m}_{ts}(z)\|_1$$

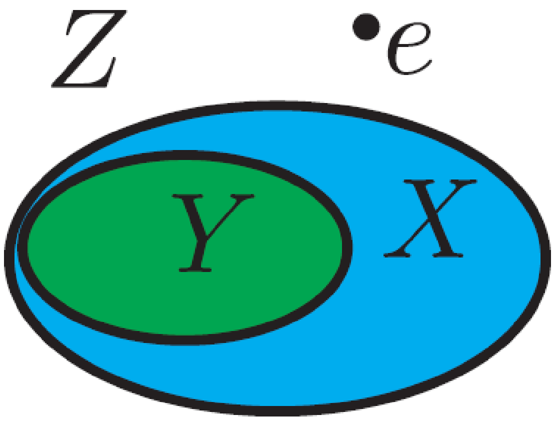
$$\text{subject to} \quad \|z\|_1 \leq N, \quad z \in \{0,1\}^{\alpha N}$$

Easier to analyze than $L_\infty$ selection…

**P1: Message error upper bounds pseudo-max-marginal error:**

$$\|\nu_s - \hat{\nu}_s\|_1 \leq \sum_{t \in \Gamma(s)} \|m_{ts} - \hat{m}_{ts}\|_1^{\rho_{ts}}$$

# Submodularity




A function $f : 2^Z \rightarrow \mathbb{R}$ is submodular iff diminishing marginal gains:

$$f(Y \cup \{e\}) - f(Y) \geq f(X \cup \{e\}) - f(X)$$

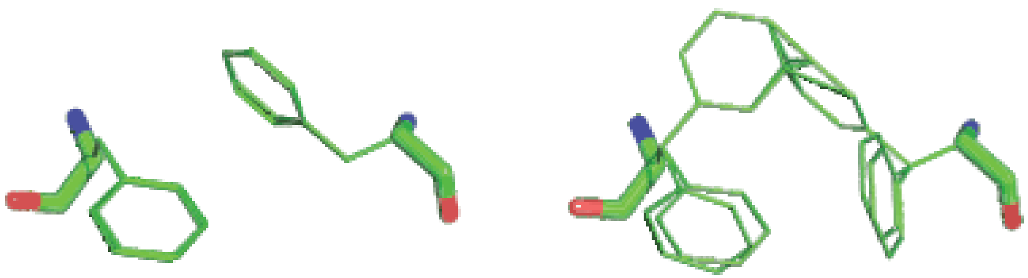
➢Diverse particle selection is *submodular maximization with cardinality constraint*

➢Efficient greedy approximation algorithm

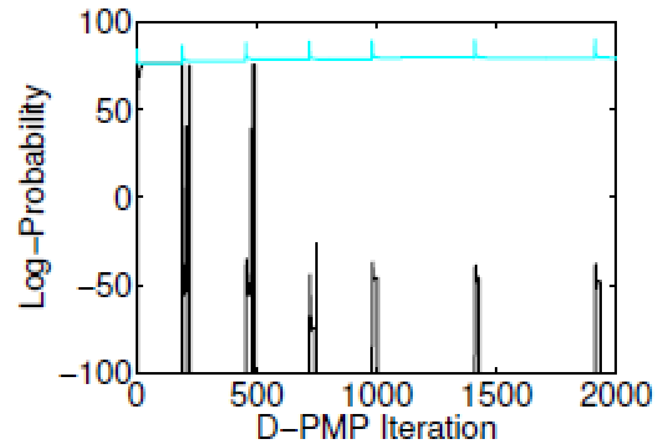
# Resolving Ties

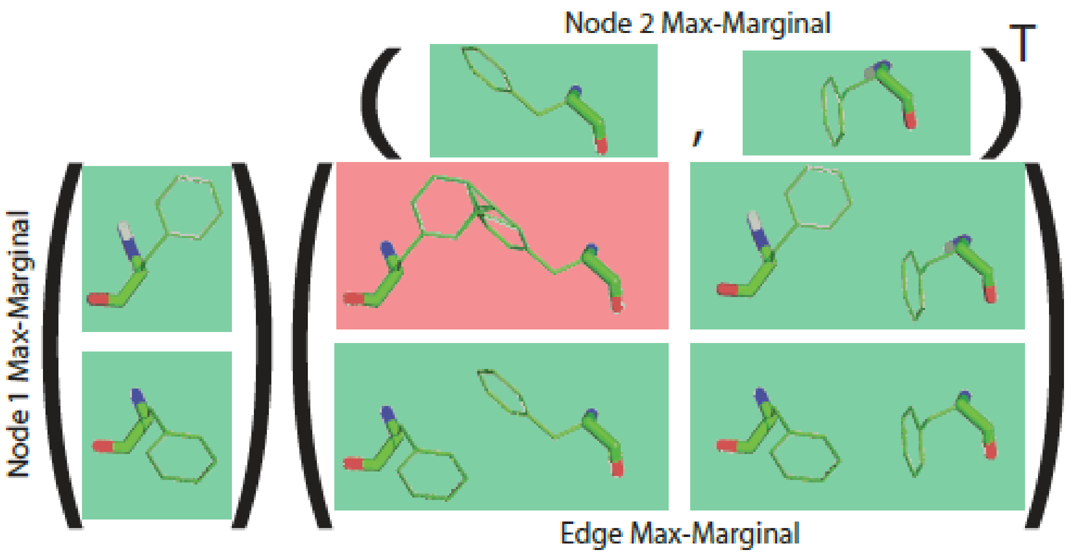
Particle diversity leads to more conflicts:

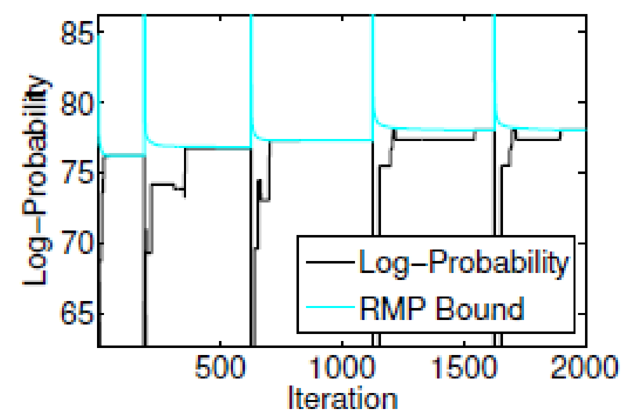
**Side Chain Particles**



**T-PMP** **D-PMP**



# Submodular Particle Selection

**Augmented Messages** | **Subset Messages** | **Selection Vector**

$$\underset{z}{\text{minimize}} \sum_{s \in \Gamma(t)} \|m_{ts} - \hat{m}_{ts}(z)\|_1$$

$$\text{subject to} \quad \|z\|_1 \le N, \quad z \in \{0,1\}^{\alpha N}$$

<u>Property 1:</u> Message reconstruction error bounds pseudo-max-marginal error:

$$\|\nu_s - \hat{\nu}_s\|_1 \le \sum_{t \in \Gamma(s)} \|m_{ts} - \hat{m}_{ts}\|_1^{\rho_{ts}}$$

<u>Property 2:</u> IP is equivalent to submodular maximization subject to cardinality constraints

# Submodular Particle Selection

Select particles to minimize sum of errors:

**Augmented Messages** **Subset Messages** **Selection Vector**

$$\underset{z}{\text{minimize}} \sum_{s \in \Gamma(t)} \|m_{ts} - \hat{m}_{ts}(z)\|_1$$

$$\text{subject to} \quad \|z\|_1 \leq N, \quad z \in \{0,1\}^{\alpha N}$$

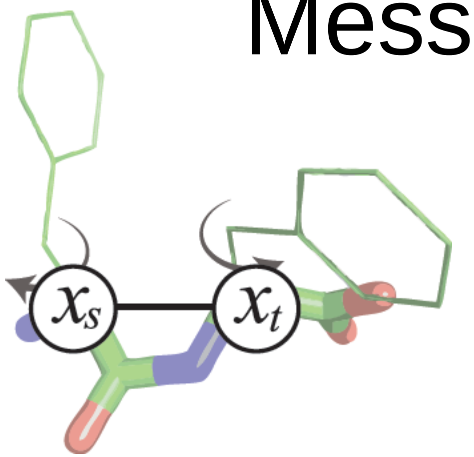
Good empirical results *and* we can analyze!

Property 1: Message error bounds pseudo-max-marginal:

$$\|\nu_s - \hat{\nu}_s\|_1 \leq \sum_{t \in \Gamma(s)} \|m_{ts} - \hat{m}_{ts}\|_1^{\rho_{ts}}$$

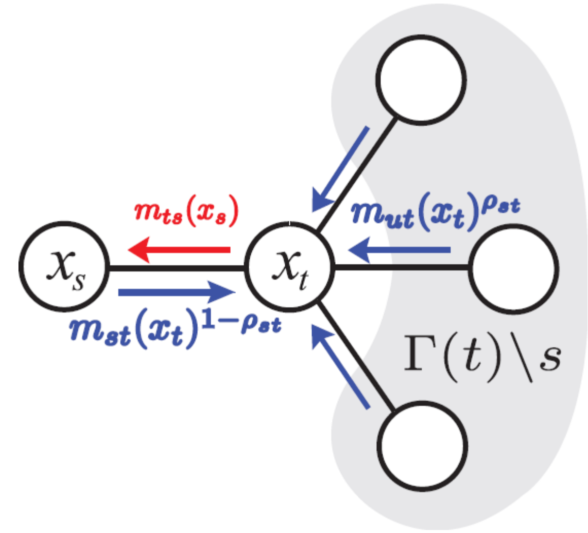
Property 2: Equivalent to submodular maximization subject to cardinality constraints

# Reweighted Max-Product (RMP)




Message passing on discrete side chain states.

But latent space is continuous…




RMP Messages:

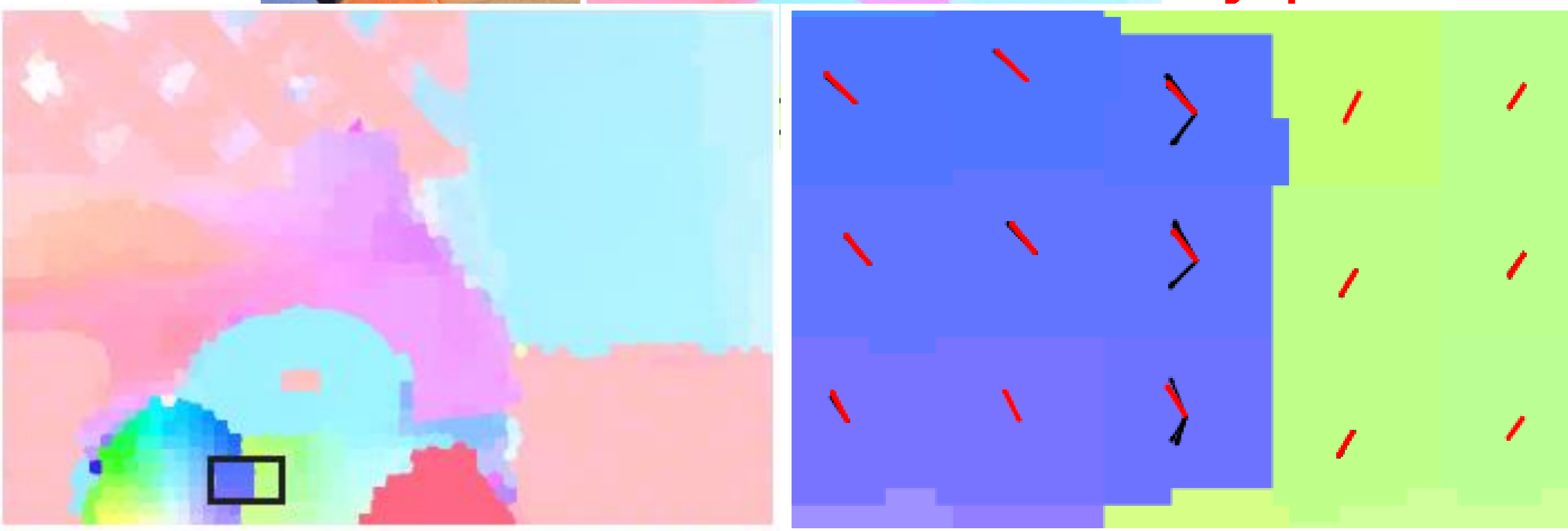
$$m_{ts}(x_s) = \max_{x_t} \psi_t(x_x)\psi_{st}(x_s,x_t)^{\frac{1}{\rho_{st}}} \frac{\prod_{u\in\Gamma(t)\backslash s} m_{ut}(x_t)^{\rho_{st}}}{m_{st}(x_t)^{1-\rho_{st}}}$$

**Edge Appearance Probability**

Pseudo-max-marginal: $\nu_s(x_s) \propto \psi_s(x_s) \prod_{u\in\Gamma(s)} m_{us}(x_s)^{\rho_{us}}$

# Optical Flow


Estimate mo ery pixel.



Diverse Particle Selection (D-PMP)

# Diverse Particle Selection

Minimize maximum message error ($L_\infty$):

Pose Estimation

Augmented Messages

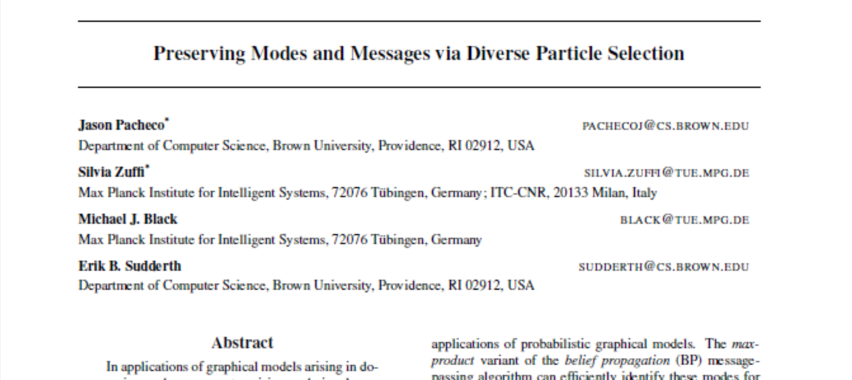
Subset Messages

$$\underset{z}{\text{minimize}} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_\infty$$

Selection Vector

$$\text{subject to} \quad \|z\|_1 \leq N, \quad z \in \{0,1\}^{\alpha N}$$

**[ Pacheco et al., ICML 2014 ]**



Preserving Modes and Messages via Diverse Particle Selection


**Jason Pacheco*** PACHECOJ@CS.BROWN.EDU
Department of Computer Science, Brown University, Providence, RI 02912, USA
**Silvia Zuffi*** SILVIA.ZUFFI@TUE.MPG.DE
Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany; ITC-CNR, 20133 Milan, Italy
**Michael J. Black** BLACK@TUE.MPG.DE
Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany
**Erik B. Sudderth** SUDDERTH@CS.BROWN.EDU
Department of Computer Science, Brown University, Providence, RI 02912, USA



Abstract

In applications of graphical models arising in do-

applications of probabilistic graphical models. The *max-product* variant of the *belief propagation* (BP) message-passing algorithm can efficiently identify these modes for


- Good empirical results
- No analysis/guarantees
- Limited to tree MRFs