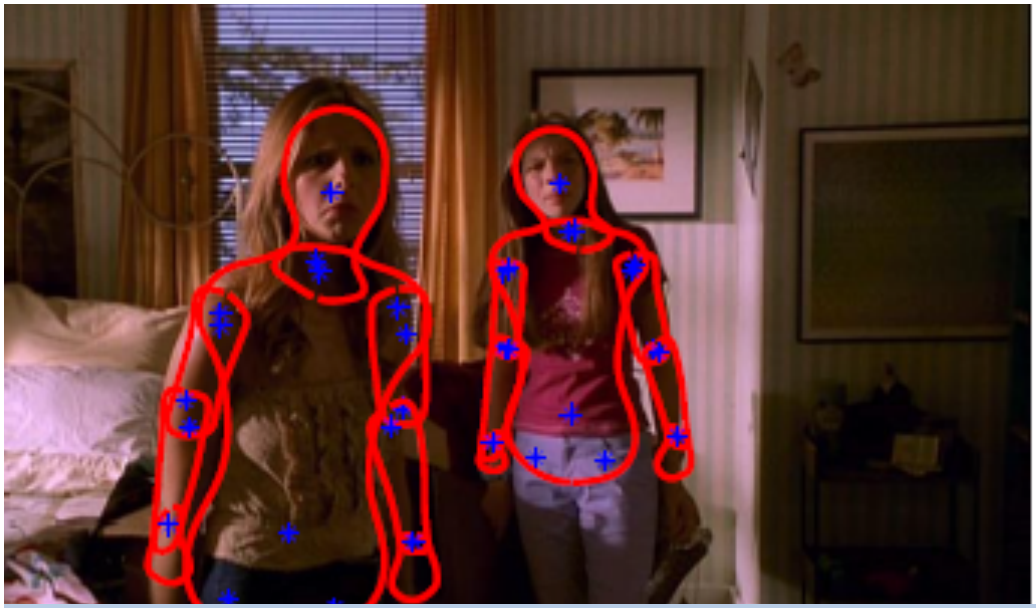
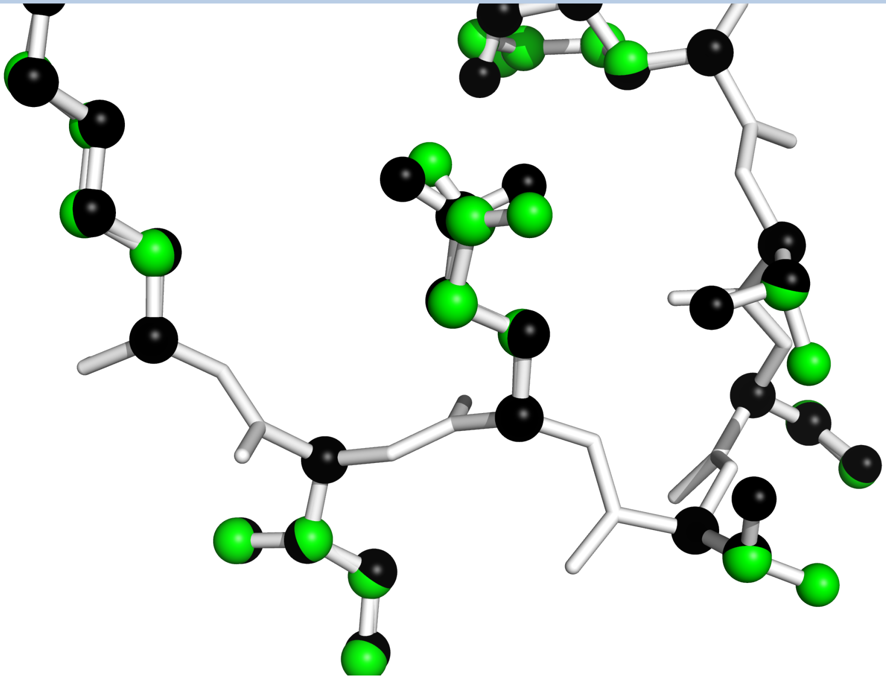


CS242: Probabilistic Graphical Models

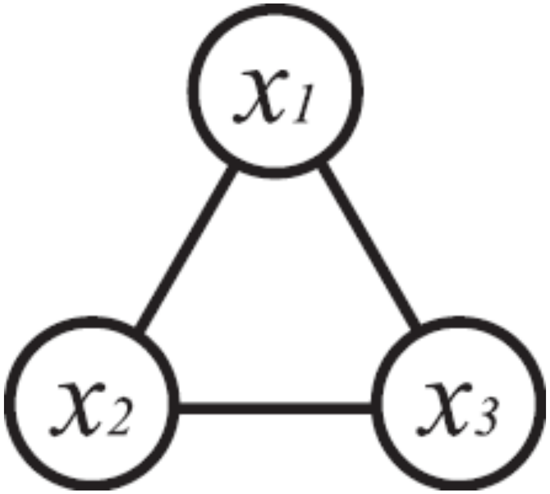
Lecture 7A: Particle Belief Propagation



Guest Lecturer:
Jason Pacheco

Some figures and materials courtesy of:
Silvia Zuffi, Michael Black, MPI Tubingen

Pairwise Graphical Models



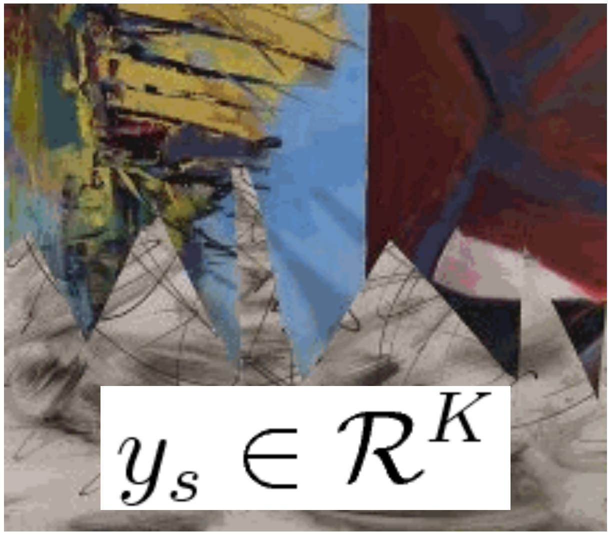
$$p(x) \propto \prod_{s \in \mathcal{V}} \underbrace{\psi_s(x_s)}_{\text{Likelihood}} \prod_{(s,t) \in \mathcal{E}} \underbrace{\psi_{st}(x_s, x_t)}_{\text{Compatibility}}$$

$\mathcal{G}(\mathcal{V}, \mathcal{E})$: Vertices $s \in \mathcal{V}$, edges $(s,t) \in \mathcal{E}$

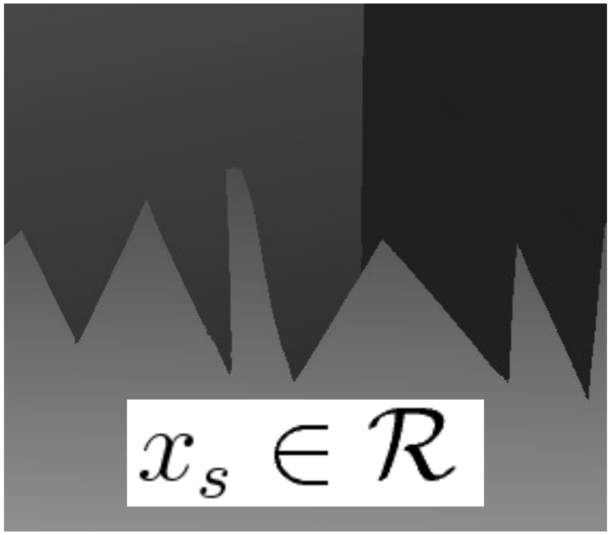
- Nodes are *continuous* random variables $x_s \in \mathbb{R}^d$
- *Potentials* $\psi(x)$ encode statistical relationships
- Edges indicate direct pairwise energetic interaction
- Facilitates efficient statistical inference

Example: Stereo Vision

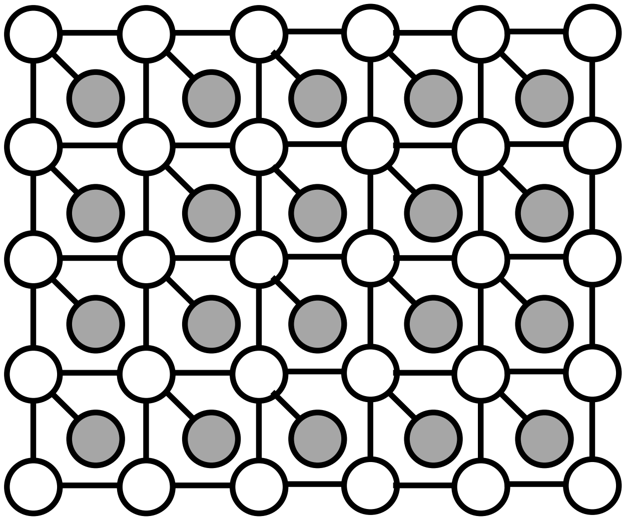
Input Images



Depth Map



Grid Graph



Non-Gaussian continuous model:

$\psi_s(x_s, y_s)$: Likelihood scores disparity / obeys occlusion

$\psi_{st}(x_s, x_t)$: Compatibility encourages similar disparities, allows discontinuity at image edges

Can't use particle filter, not a time series…

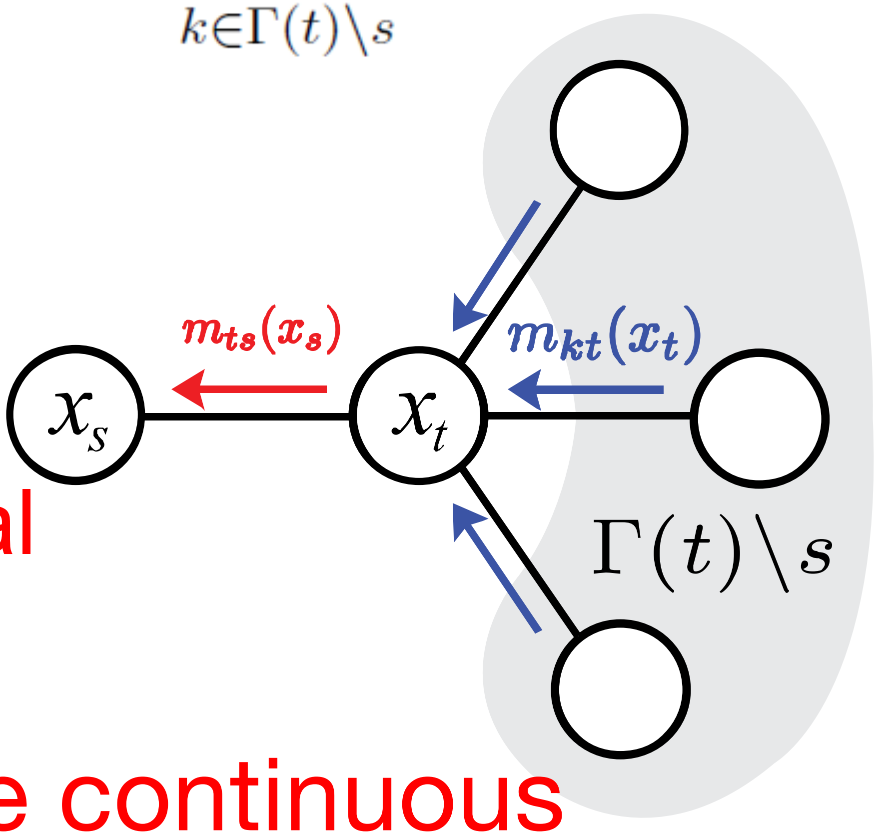
Sum-Product Belief Propagation

Passing *messages* in graphical model:

$$m_{ts}(x_s) = \int \psi_{st}(x_s, x_t)\psi_t(x_t) \prod_{k \in \Gamma(t) \setminus s} m_{kt}(x_t)$$

Discrete: BP updates well-defined, matrix-vector products.

Continuous: no general closed-form updates.



Solution: Approximate continuous messages with discrete *particles.*

Importance Sampling

- Draw samples from proposal distribution,

$$x^{(i)} \sim q(x), \quad \{x^{(1)}, \ldots, x^{(N)}\}$$

- Approximate expectation,

$$\mathbb{E}[g(x)] = \int_{\mathcal{X}} g(x)p(x)\; dx \approx \sum_{i=1}^{N} g(x^{(i)})w(x^{(i)}),$$

- Importance weights account for proposal,

$$w(x) \propto \frac{p(x)}{q(x)}, \quad \sum_{i=1}^{N} w(x^{(i)}) = 1$$

Main Idea: Approximate BP messages with importance sampling.

Particle Belief Propagation

- Rewrite BP message as expectation,

$$m_{ts}(x_s) = \mathbb{E}_{q_t}\left[\psi_{st}(x_s, x_t)\frac{\psi_t(x_t)}{q_t(x_t)}\prod_{k\in\Gamma(t)\backslash s} m_{kt}(x_t)\right]$$

- Importance weighted expectation,

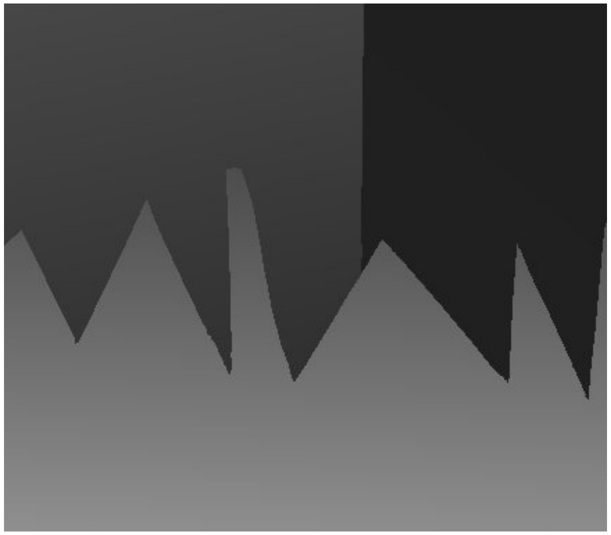
$$m_{ts}(x_s) \approx \sum_{i=1}^{N} \psi_{st}(x_s, x_t^{(i)})w_t(x_t^{(i)}) \triangleq \hat{m}_{ts}(x_s)$$

- Importance weights,

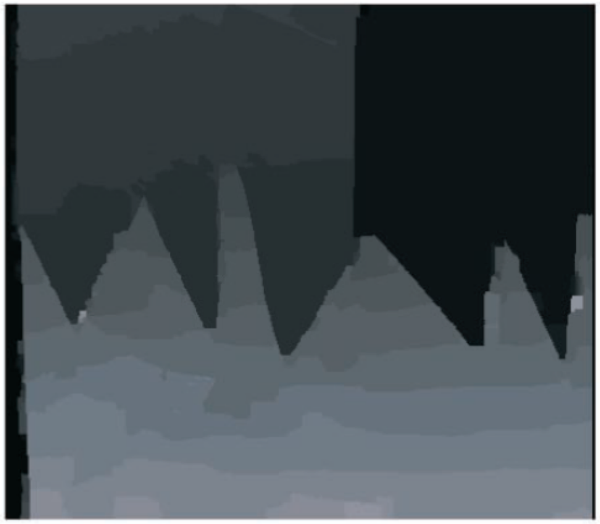
$$w_t(x_t^{(i)}) = \frac{\psi_t(x_t^{(i)})\prod_{k\in\Gamma(t)\backslash s} m_{kt}(x_t^{(i)})}{q_t(x_t^{(i)})}$$

Sample particles from BP marginal.

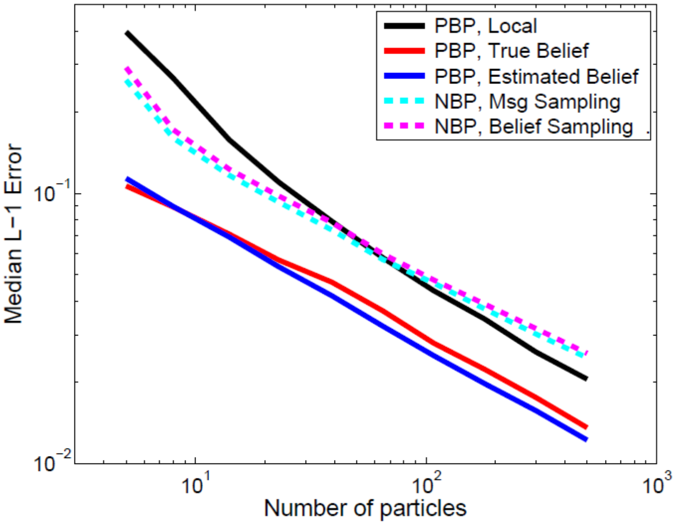
Stereo Vision Results



True Disparity



Estimate

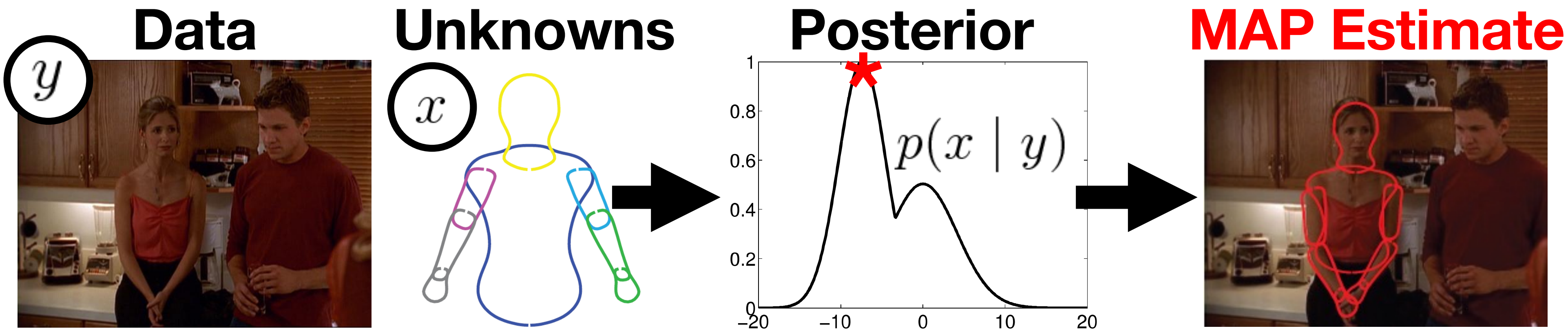


Prediction Error

- ➢Comparison to related Nonparametric BP
- ➢L_1 error w.r.t. true beliefs via discretization
- ➢Errors decrease at rate $1/\sqrt{N}$

A. Ihler and D. McAllester, AISTATS 2009

Maximum a Posteriori (MAP)



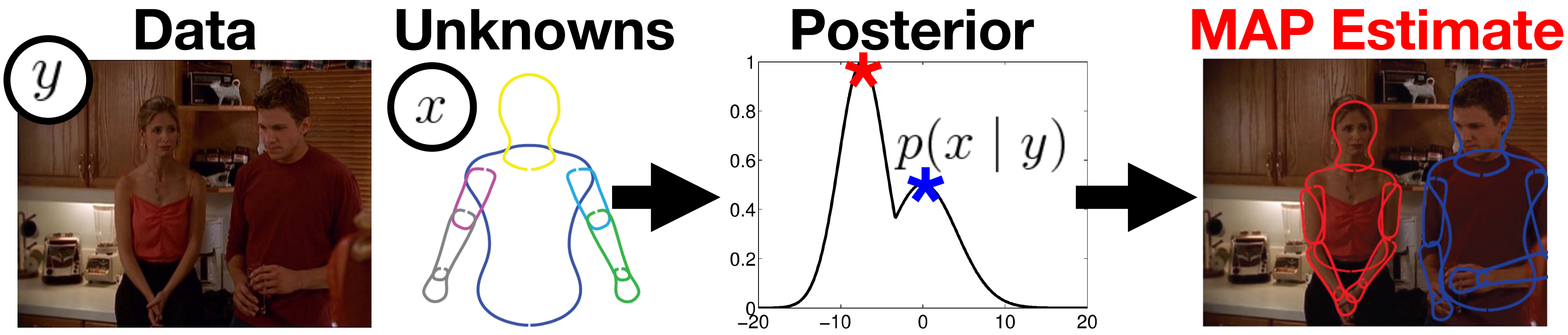
Maximizer of the posterior probability:

$$\color{red}{x^*}\color{black} = \underset{x}{\operatorname{argmax}}\, p(x \mid y)$$

Issues with continuous models:

➢ Analytically intractable posterior density

Maximum a Posteriori (MAP)



Maximizer of the posterior probability:

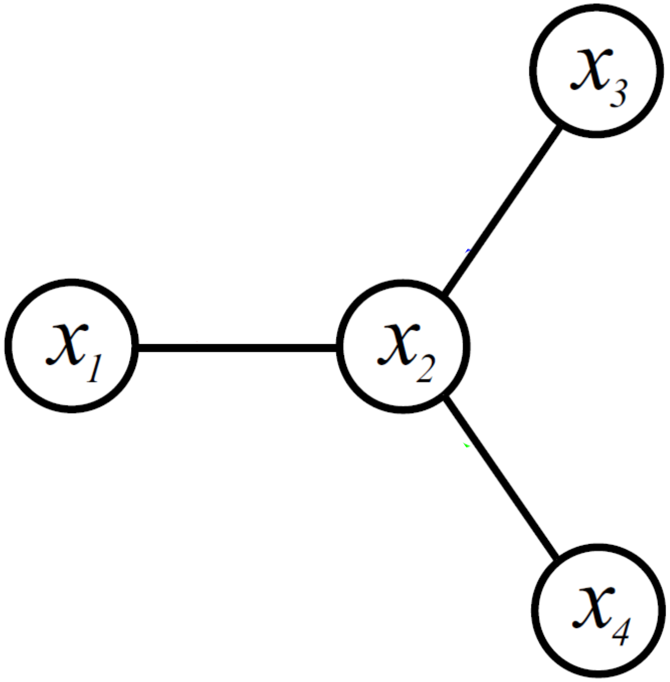
$$\color{red}{x^*} \color{black}{= \operatorname*{argmax}_x p(x \mid y)}$$

Issues with continuous models:

- ➢ Analytically intractable posterior density
- ➢ Multiple local optima (these can be useful too…)

Message Passing

Global MAP inference decomposes into local computations via graph structure…

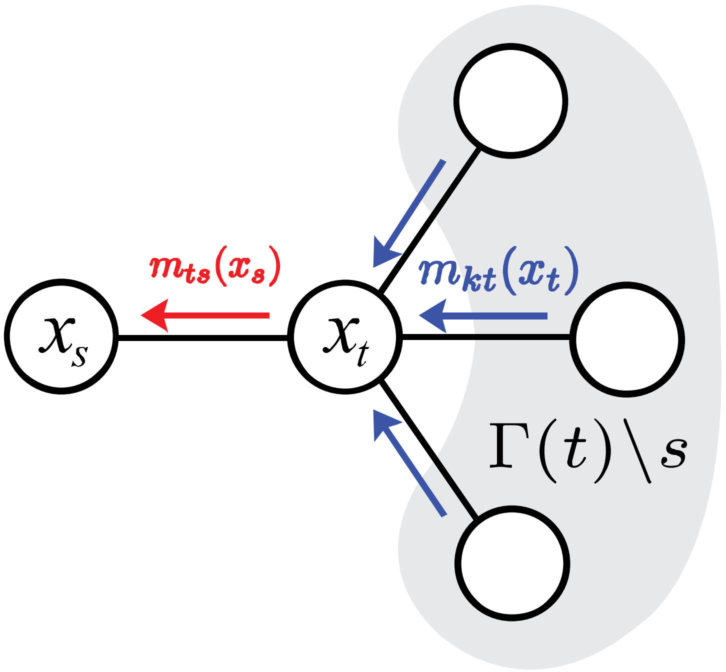


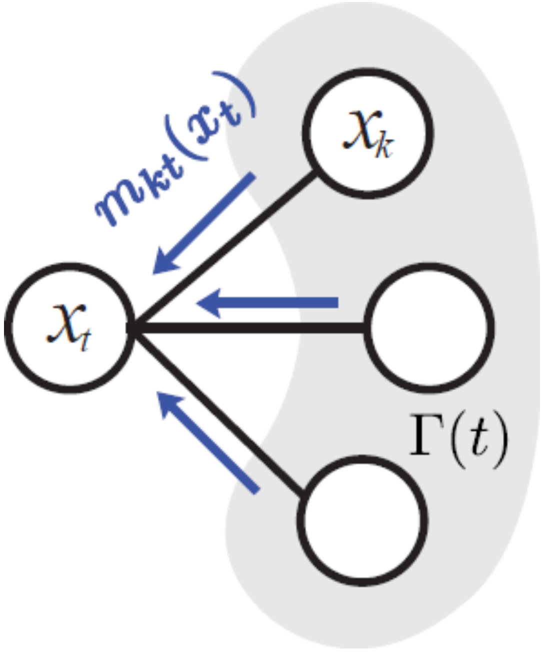
$$\max_x p(x) \propto \max_x \psi_{12}(x_1, x_2)\psi_{23}(x_2, x_3)\psi_{24}(x_2, x_4)$$

$$= \max_{x_1} \underbrace{\max_{x_2} \psi_{12}(x_1, x_2) \underbrace{\left[\max_{x_3} \psi_{23}(x_2, x_3)\right]}_{m_{32}(x_2)} \underbrace{\left[\max_{x_4} \psi_{24}(x_2, x_4)\right]}_{m_{42}(x_2)}}_{m_{21}(x_1)}$$

Max-Product (MP) Belief Propagation

Passing *messages* in a graphical model…



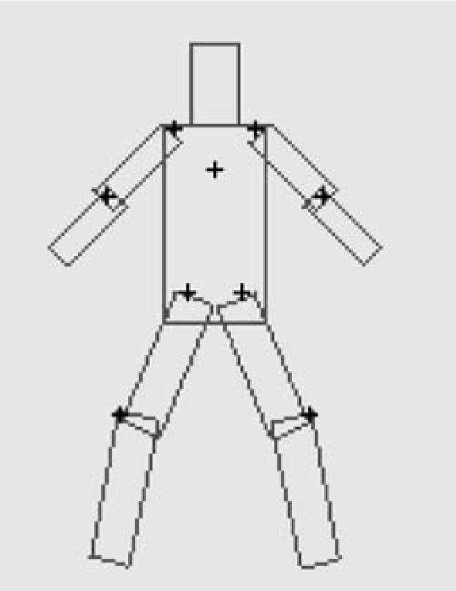


Message $m_{ts}(x_s) \propto \max_{x_t} \psi_{st}(x_s, x_t)\psi_t(x_t) \prod_{k\in\Gamma(t)\setminus s} m_{kt}(x_t)$

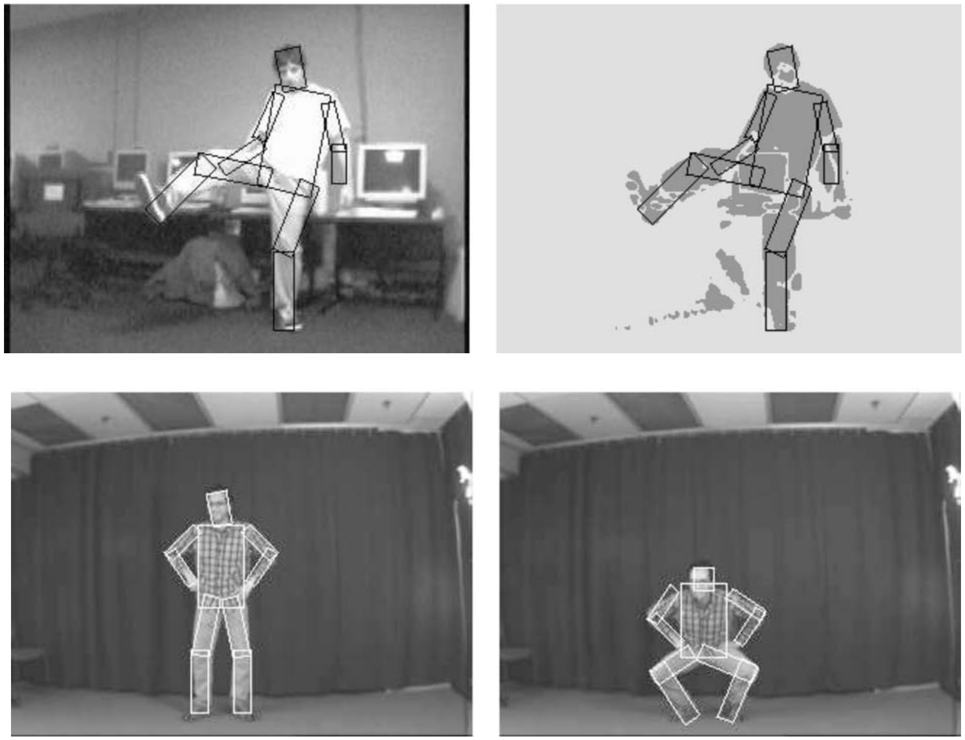
Max-Marginal $q_t(x_t) \propto \psi_t(x_t) \prod_{k\in\Gamma(t)} m_{kt}(x_t)$

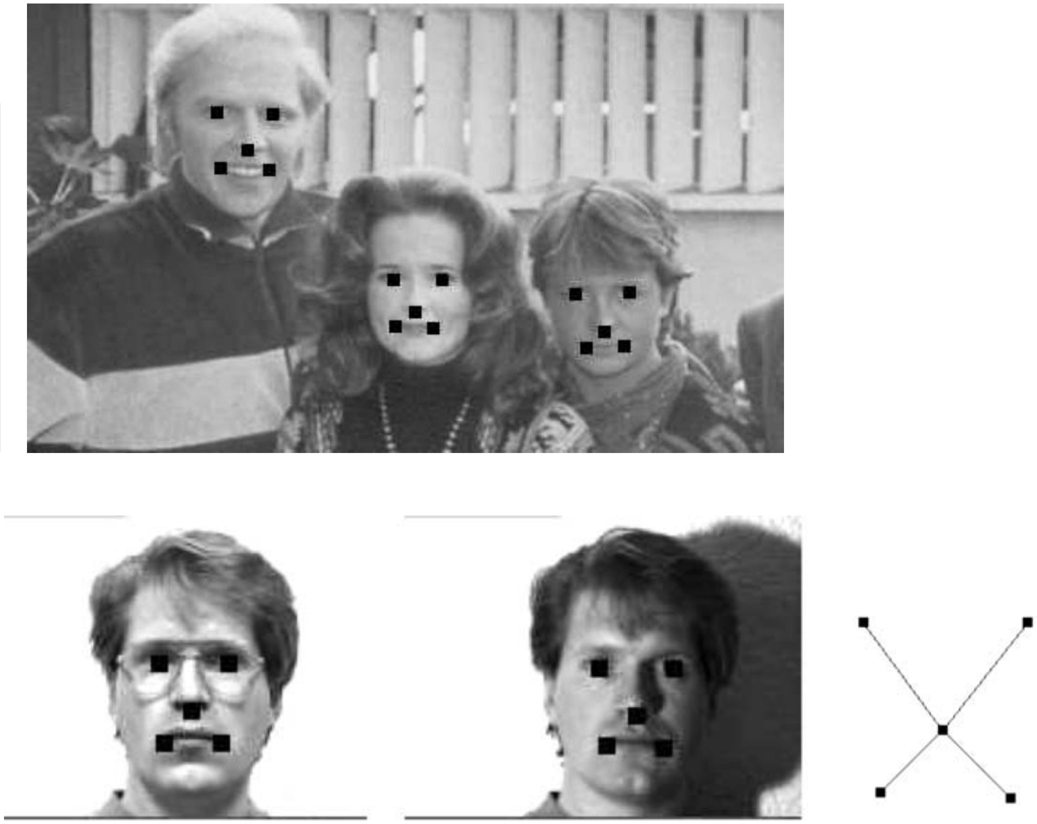
Poses & Discrete Probabilities

Felzenszwalb & Huttenlocher, 2005



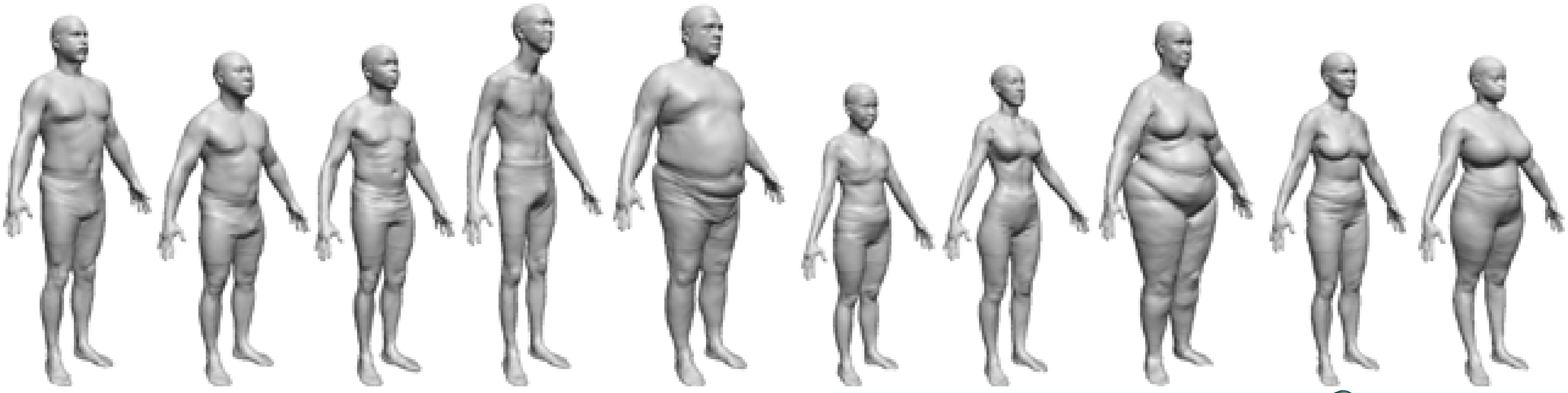
- Pairwise MRF with rigid geometry
- MAP estimate of pose via discrete max-product BP
- Discrete state space limits allowable deformations

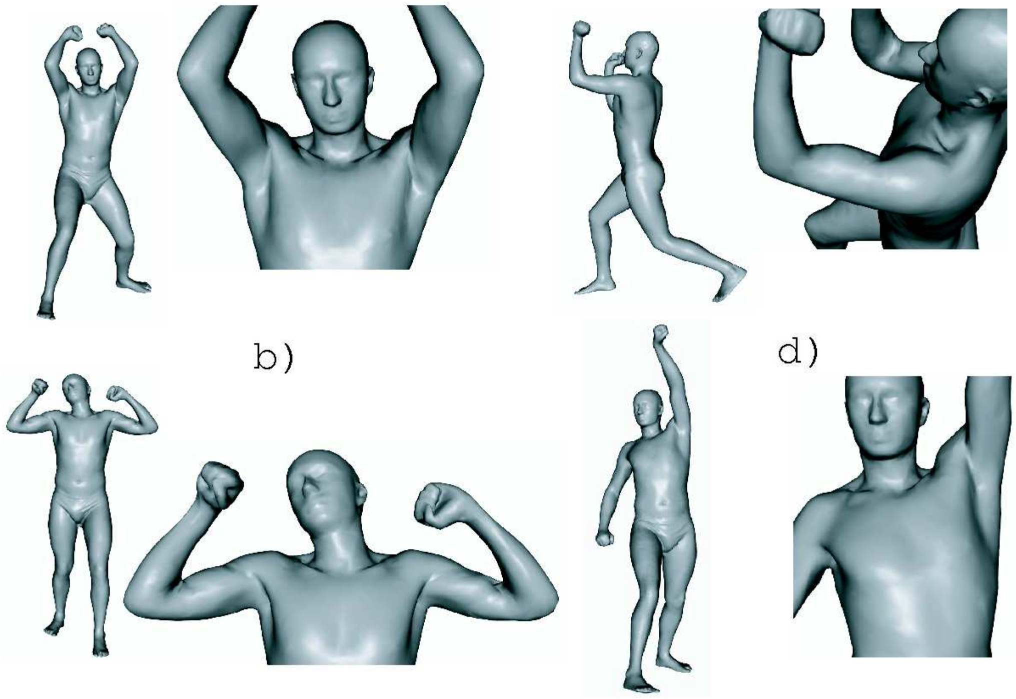


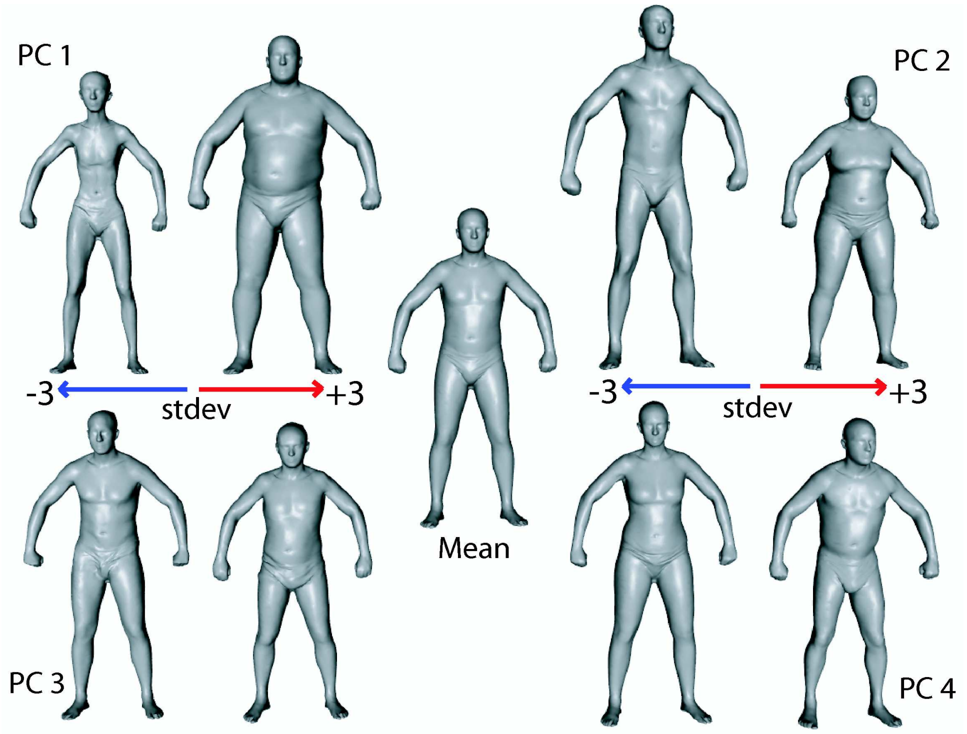


SCAPE

Shape Completion and Animation of People, Anguelov et al. 2004

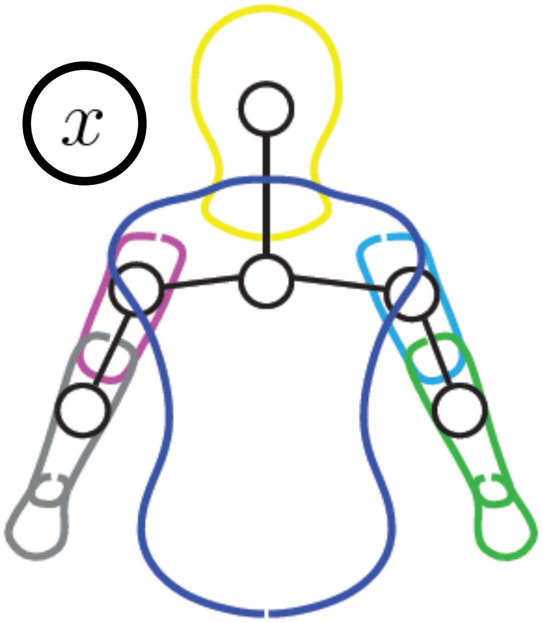






Deformable Structures

[Zuffi et al., CVPR 2012]

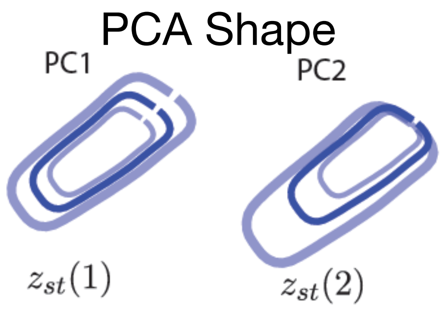


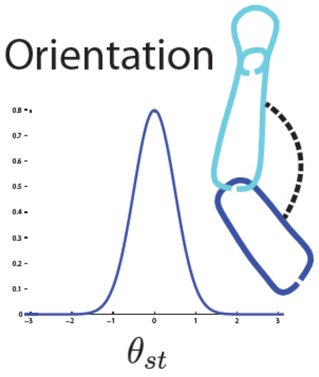
$$p(x,y) \propto \prod_{s\in\mathcal{V}} \psi_s(x_s,y) \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s,x_t)$$

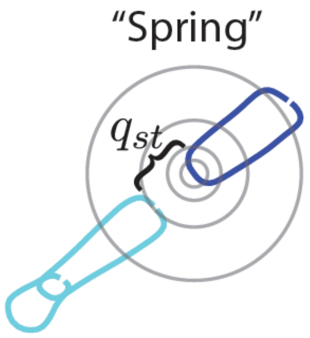
Complicated Likelihood

Non-Gaussian Compatibility

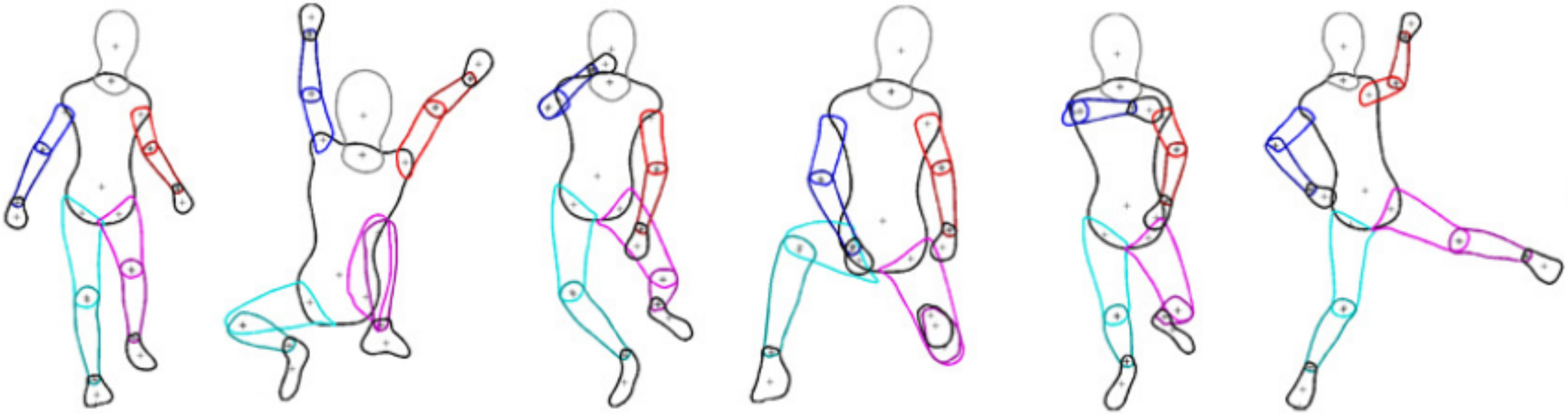
Continuous state $x_s \in \mathcal{X}_s$ for part *shape*, *location*, *orientation* and *scale*.



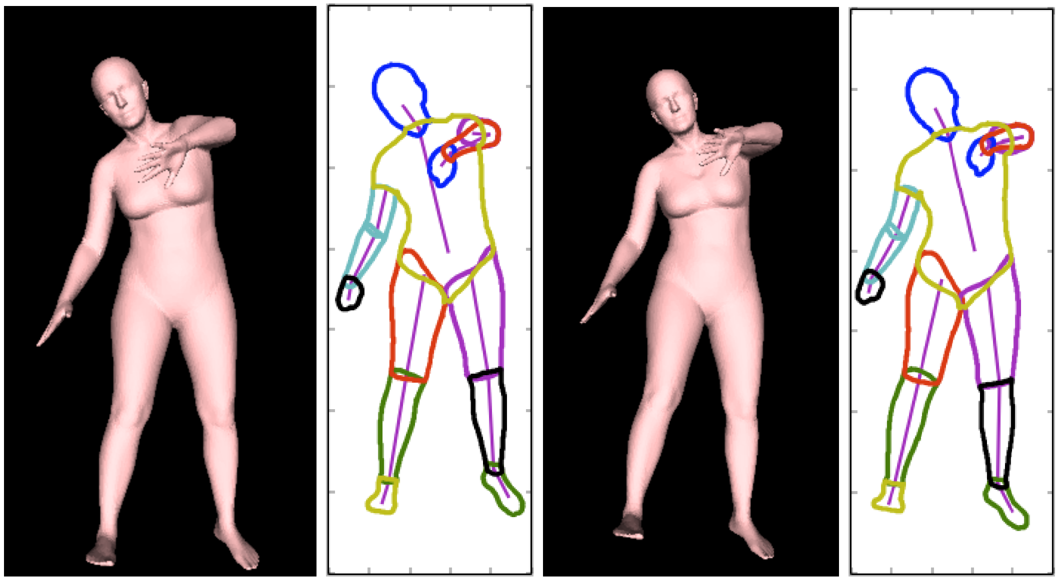




Samples From DS Prior



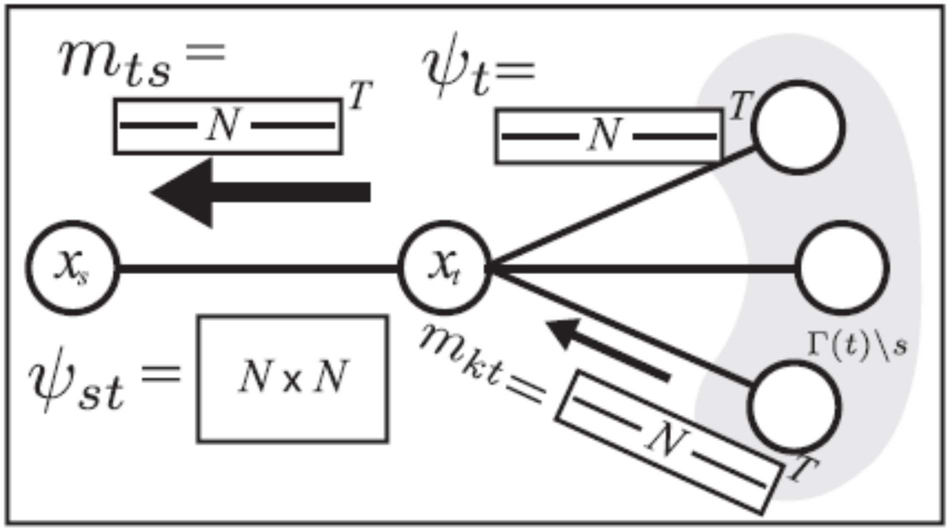
DS defines a joint probability from which we can sample human poses.



Max-Product Belief Propagation

Discrete

$$x \in \{1, \dots, N\}^D$$



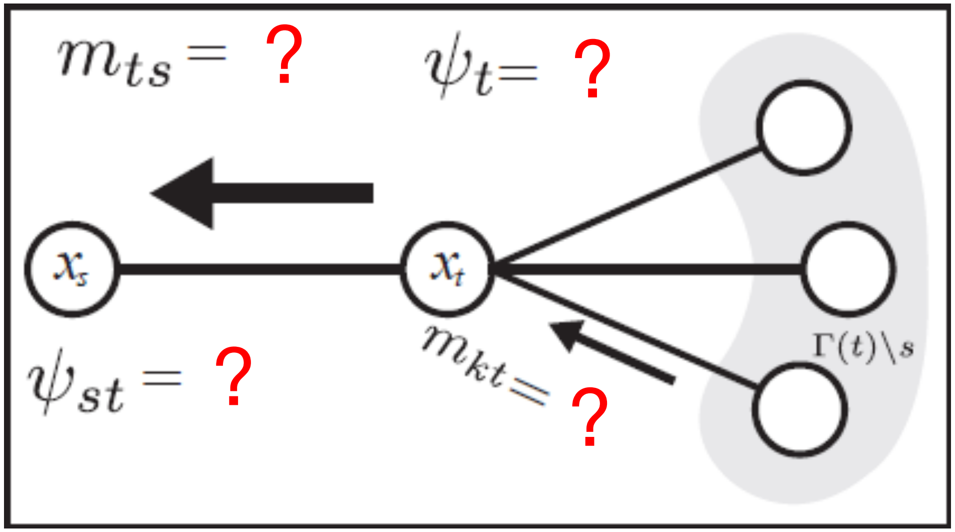
Message Update:

$$m_{ts} = \max_{x_t} \boxed{\psi_{st}} \left[\psi_t\right] \prod \left[m_{kt}\right]$$

Matrix-vector multiplication & discrete maximization

Continuous

$$x \in \mathcal{R}^D$$



Message Update:

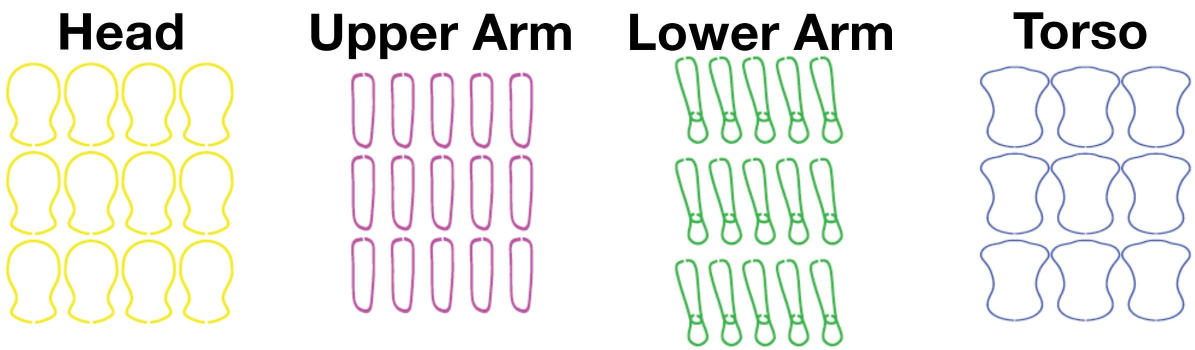
$$m_{ts}(x_s) = \dots$$

$$\max_{x_t} \psi_{st}(x_s, x_t)\psi_t(x_t) \prod m_{kt}(x_t)$$

Nonlinear optimization

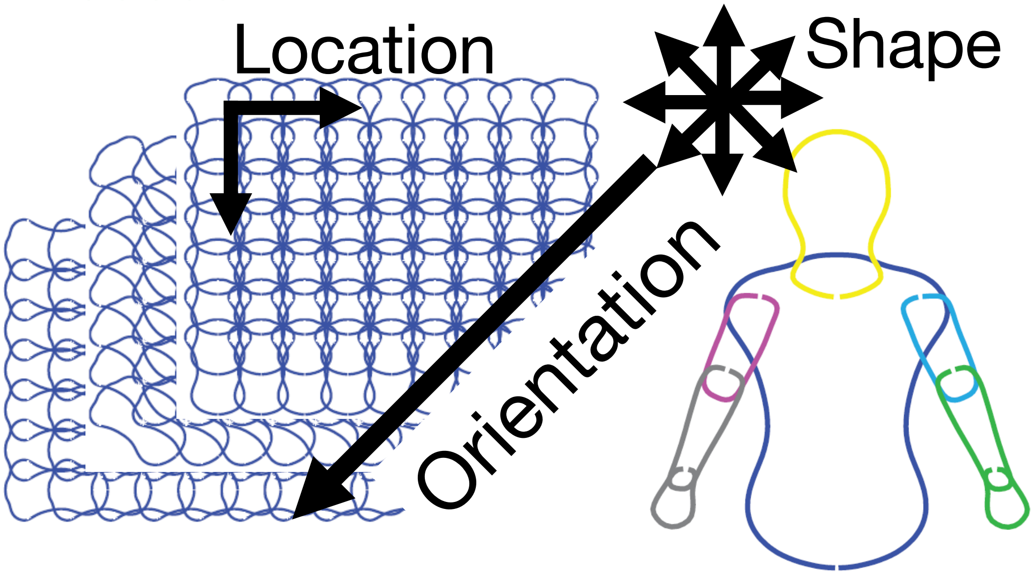
Regular Discretization

Approximate continuous max-product messages over regular grid of *points*



Example: Torso

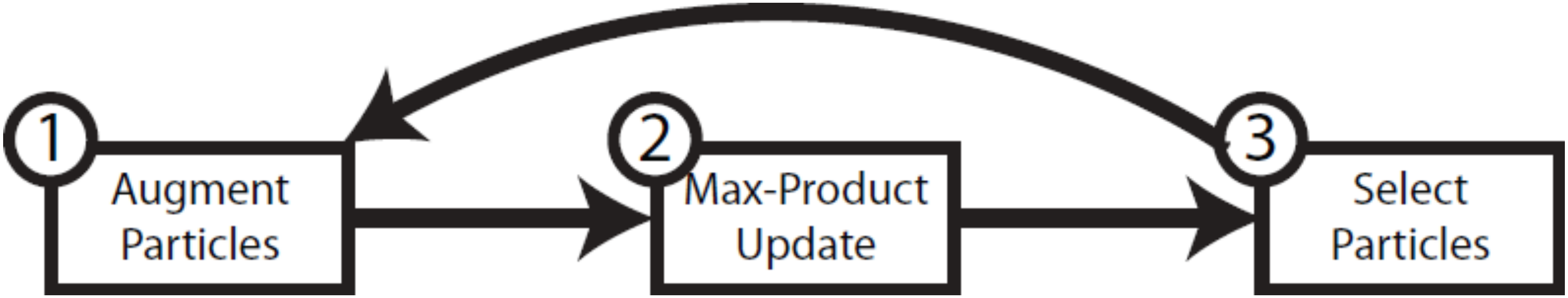
- ~10 dimensions.
- 10 grid points per dimension
- 10 Million points!



Infeasible for high dimensional models.

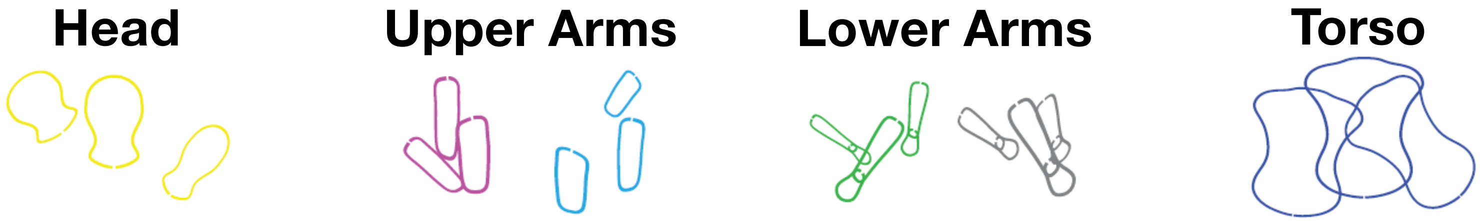
Particle Max-Product (PMP)

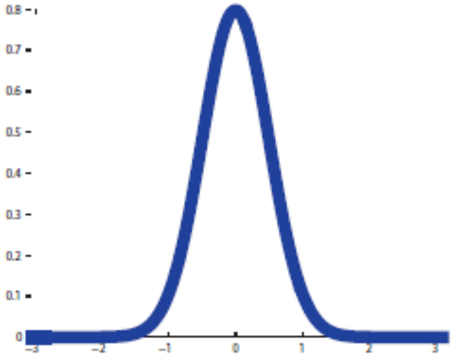
Combine particle filter ideas with max-product more effectively.



Particle approximation of continuous max-product (MP) messages.

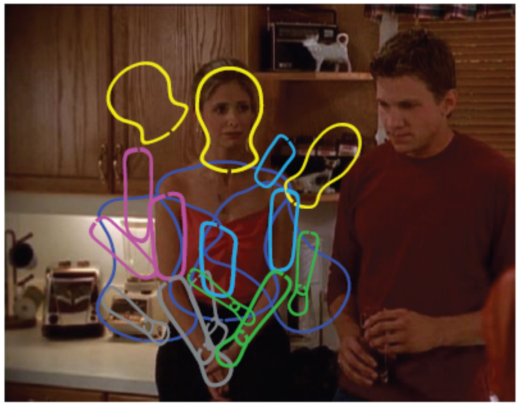
Particle Max-Product (PMP)





Proposal

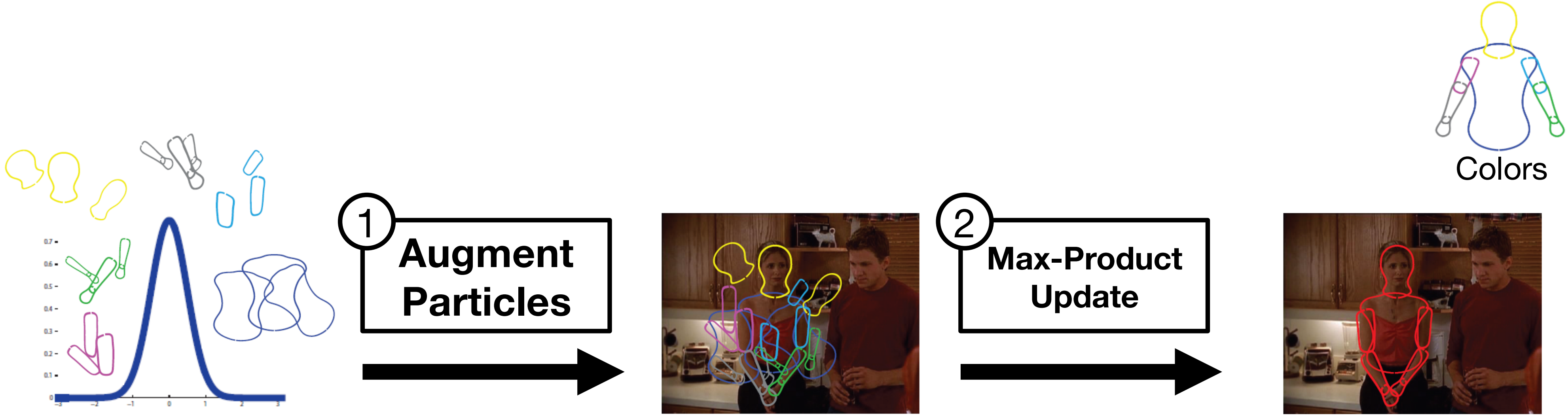
① **Augment Particles**



Augmented Set

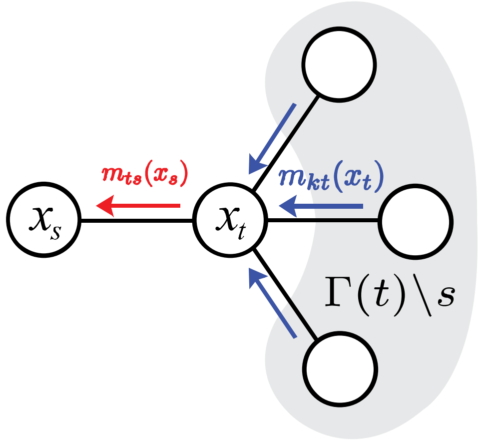
Sample new hypotheses at every node to grow particle set.

Particle Max-Product (PMP)

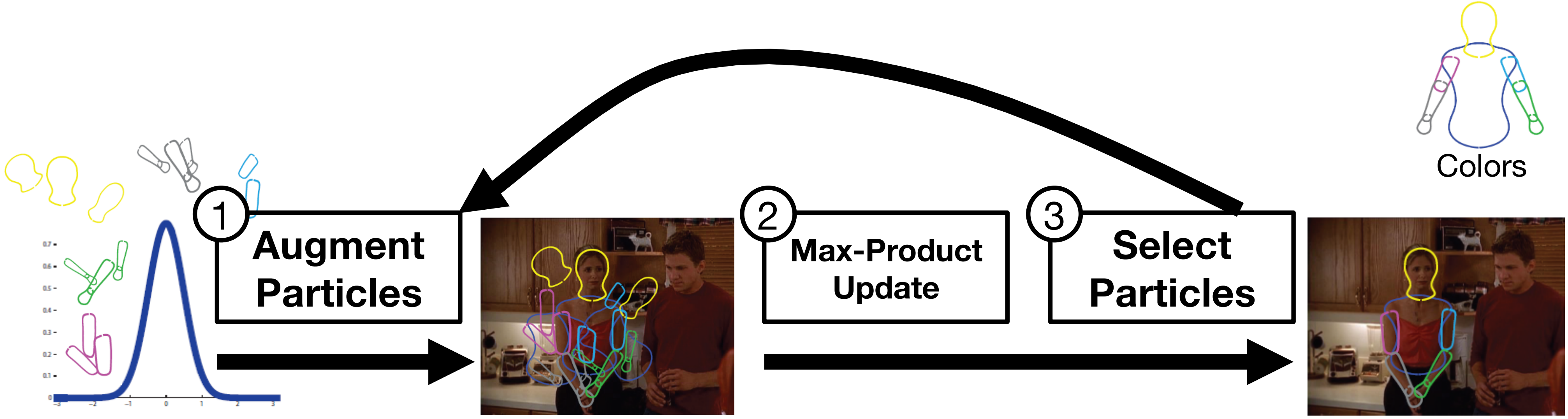


$$m_{ts}(x_s) \propto \max_{x_t} \psi_{st}(x_s, x_t)\psi_t(x_t) \prod_{k\in\Gamma(t)\setminus s} m_{kt}(x_t)$$

Update MP messages on augmented particles.



Particle Max-Product (PMP)



① **Augment Particles**

② **Max-Product Update**

③ **Select Particles**

Colors

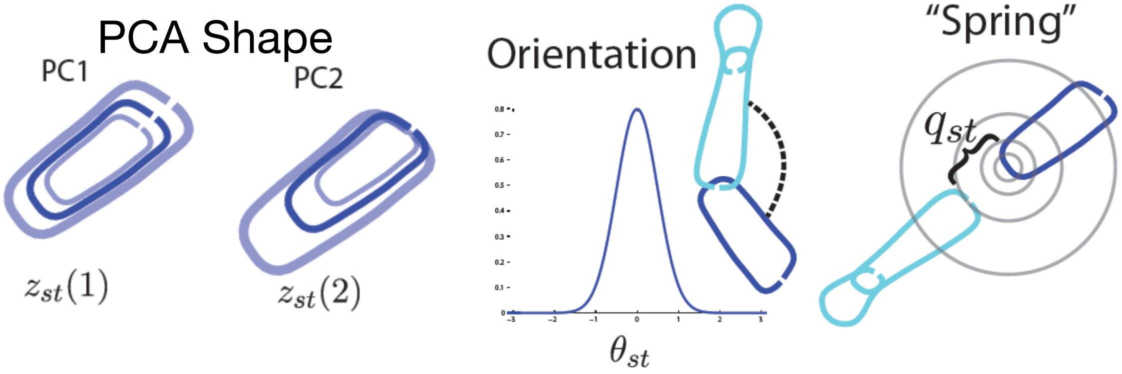
Select subset of *good* particles & repeat

Need a particle selection method…

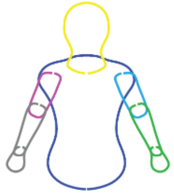
Synthetic Pose Estimation

Binary image of 4 silhouettes.

Model Truncated Gaussian pairwise potentials $\psi_{st}(x_s, x_t)$:



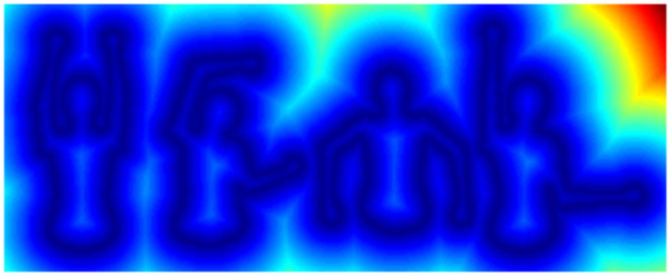
Likelihood $\psi_s(x_s)$ distance-map from silhouette contours.



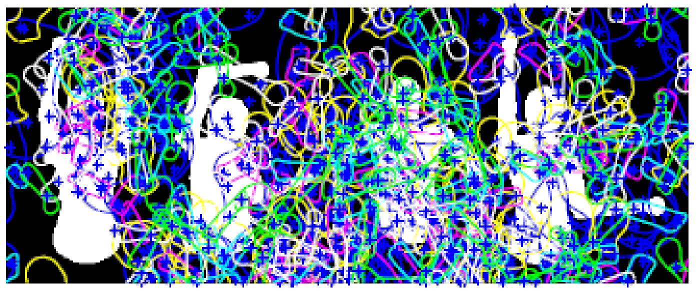
Colors

Synthetic Image

Distance Likelihood



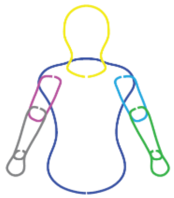
Random Initialization

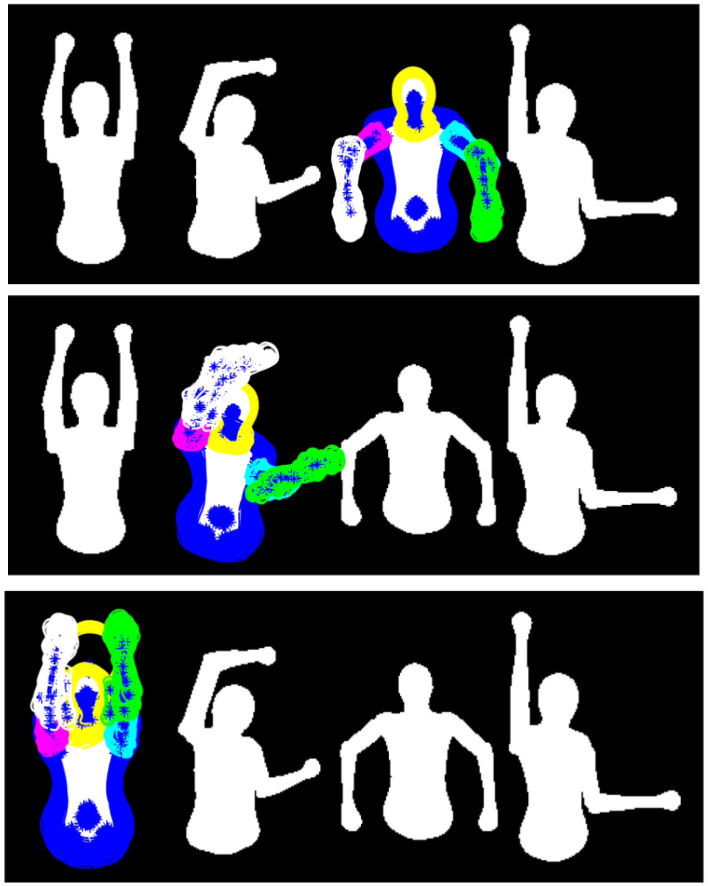


Top-N Particle Max Product (T-PMP)

- ➢ Keep N-best particles
- ➢ Sensitive to initialization
- ➢ Still too greedy; Selection reduces effective number of particles

Example Runs


Colors

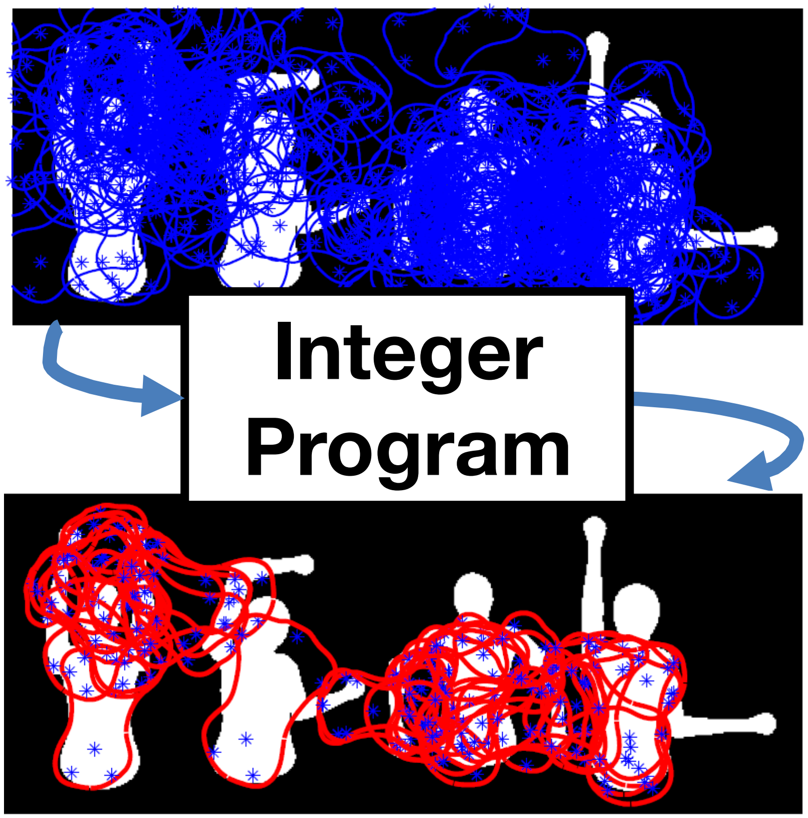


Maintain *diversity* in particles.

[Pacheco et al., ICML 2014]

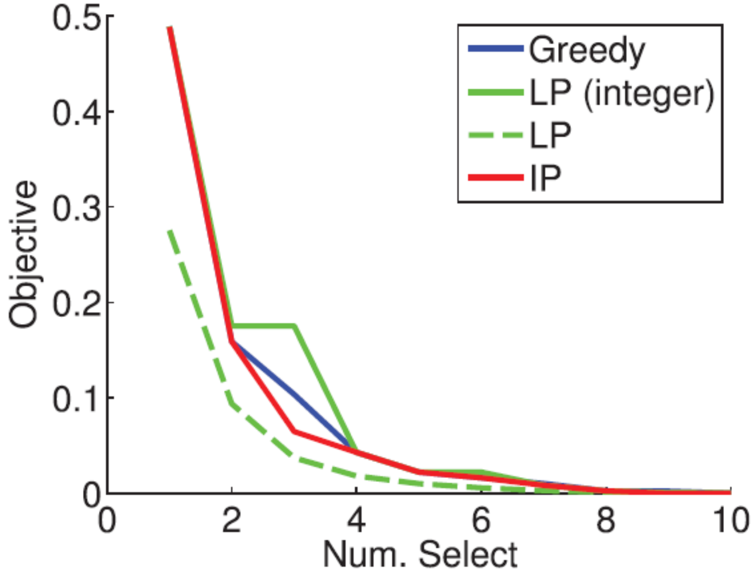
Diverse Particle Selection

Initial Particles



Diverse Selection

Integer Program (IP) solved with efficient greedy approximation:



LP : Linear Program relaxation

IP: Optimal solution by brute force

Greedy: Efficient approximation

Particles selected to minimize max-product message distortion.

Diverse Particle Selection

Minimize total message distortion:

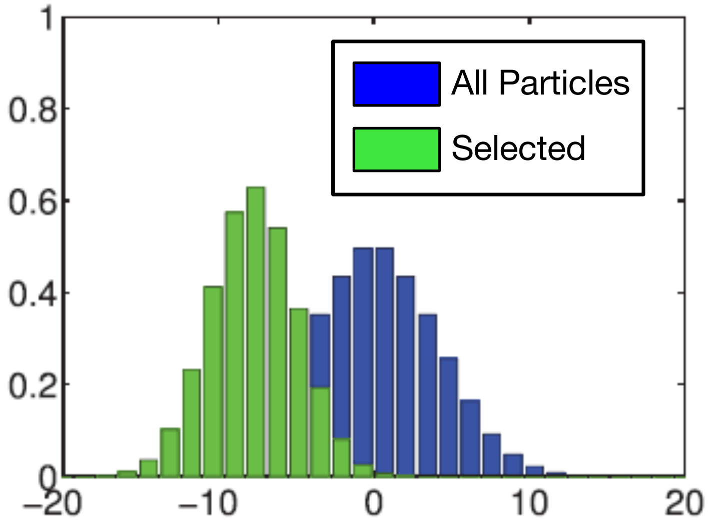
$$\underset{z}{\text{minimize}} \sum_{s\in\Gamma(t)} \sum_{a=1}^{\alpha N} (m_{ts}(a) - \hat{m}_{ts}(a,z))$$

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

✗ **NP-hard**

✓ **Submodular**

Good approximation qualities.

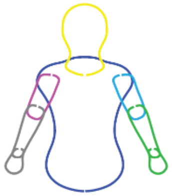


Diverse Particle Max-Product (D-PMP)

Augment:
- Data-Driven
- Neighbor
- Random Walk
- ...

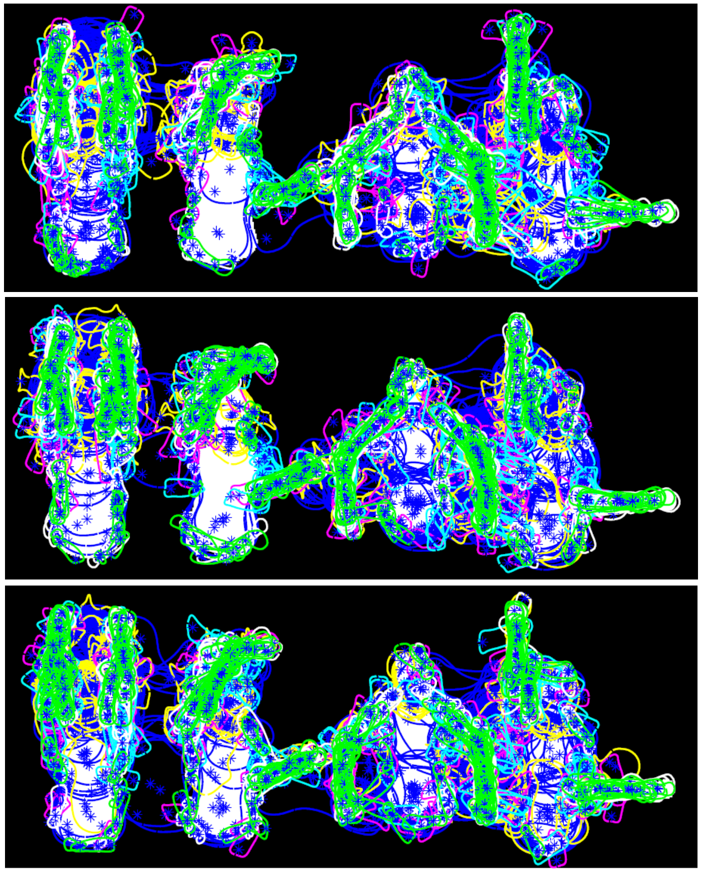
MP Update

Select Diverse



Colors

Example Runs

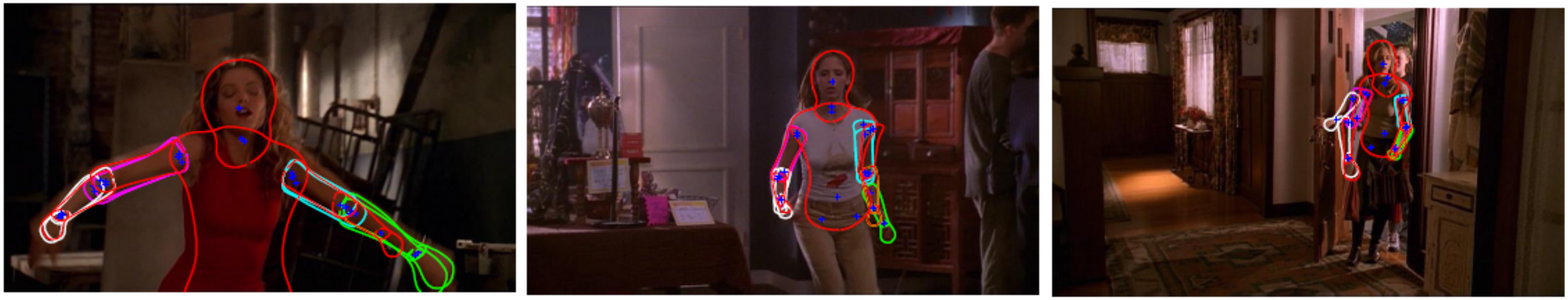


- ➢ No explicit diversity constraint
- ➢ Objective encourages diversity
- ➢ Efficient *Lazy* greedy algorithm
- ➢ Bounds on optimality

Avoids particle degeneracies by maintaining *ensemble* of *diverse solutions* near local modes.

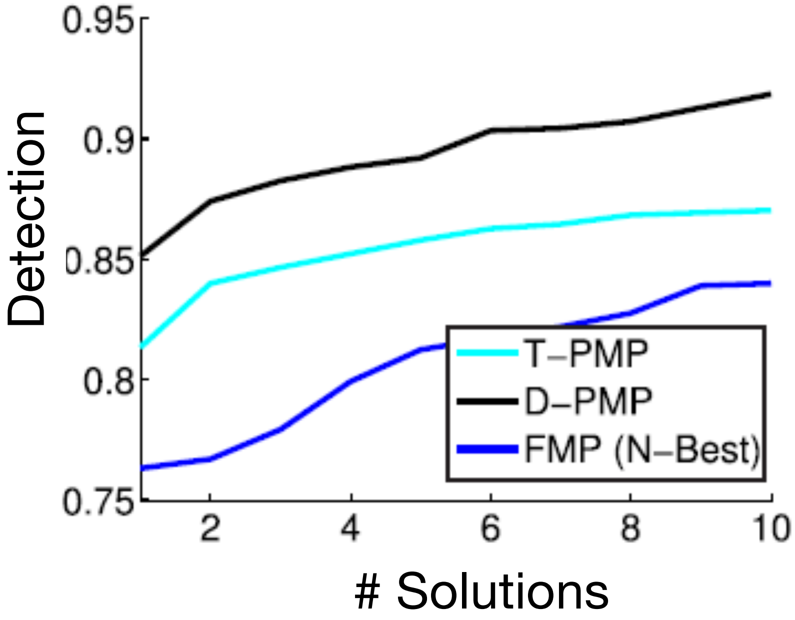
[Pacheco et al., ICML 2014]

Real Images (Single Person)



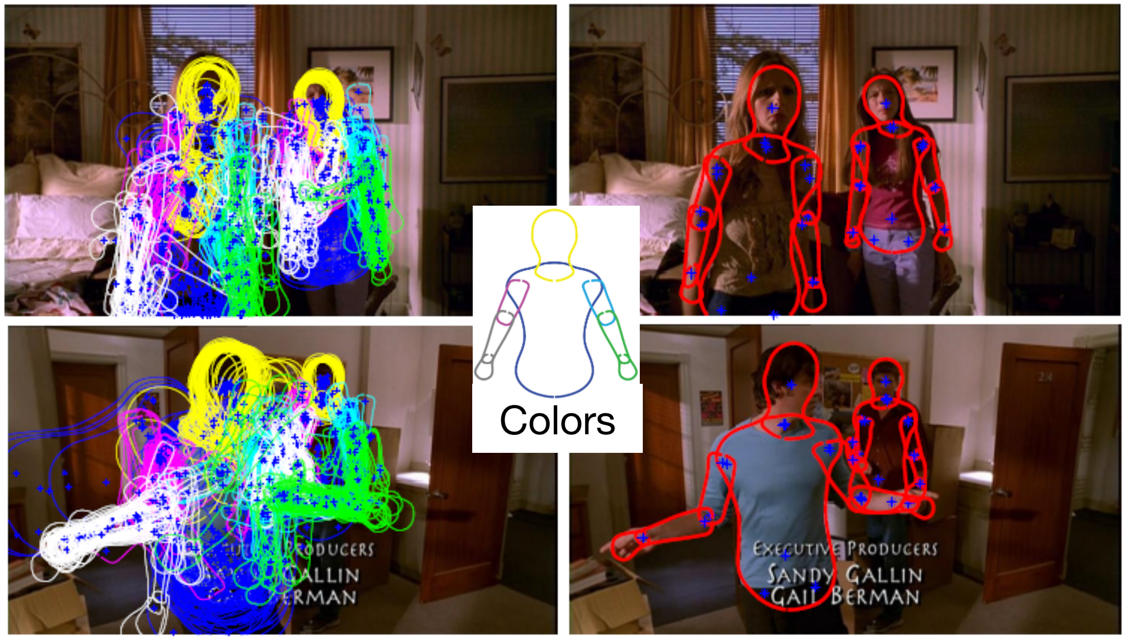
Top 3 arm hypotheses MAP estimate, 2nd and 3rd modes for upper arm (magenta, cyan), lower arm (green, white).

- ➢ "Buffy" dataset [Ferrari et al. 2008].
- ➢ Detections versus number of ranked hypotheses.
- ➢ Baseline: Flexible Mixture of Parts (FMP) [Yang & Ramanan 2013; Park & Ramanan 2011]

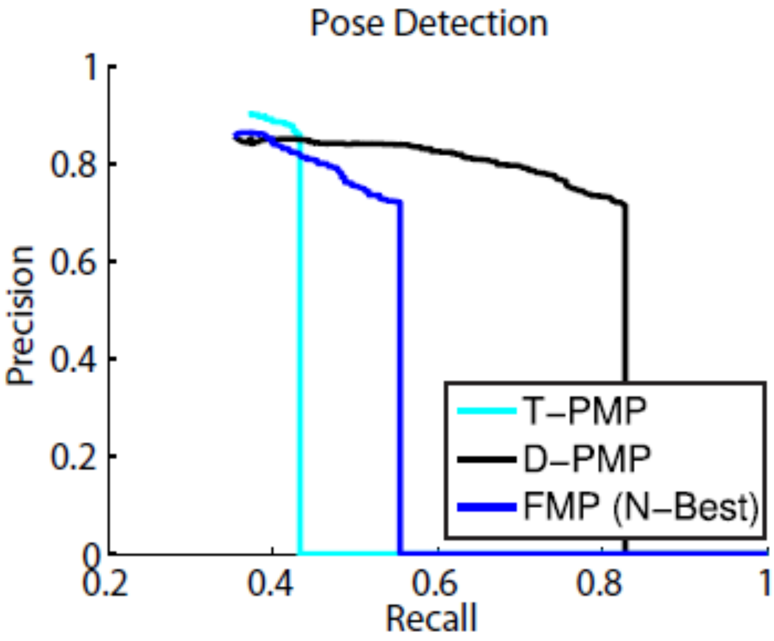


[Pacheco, Zuffi, Black & Sudderth, ICML 2014]

Real Images (Multiple People)



D-PMP Particles **Mode Estimates**



Precision-Recall for multi-person frames:

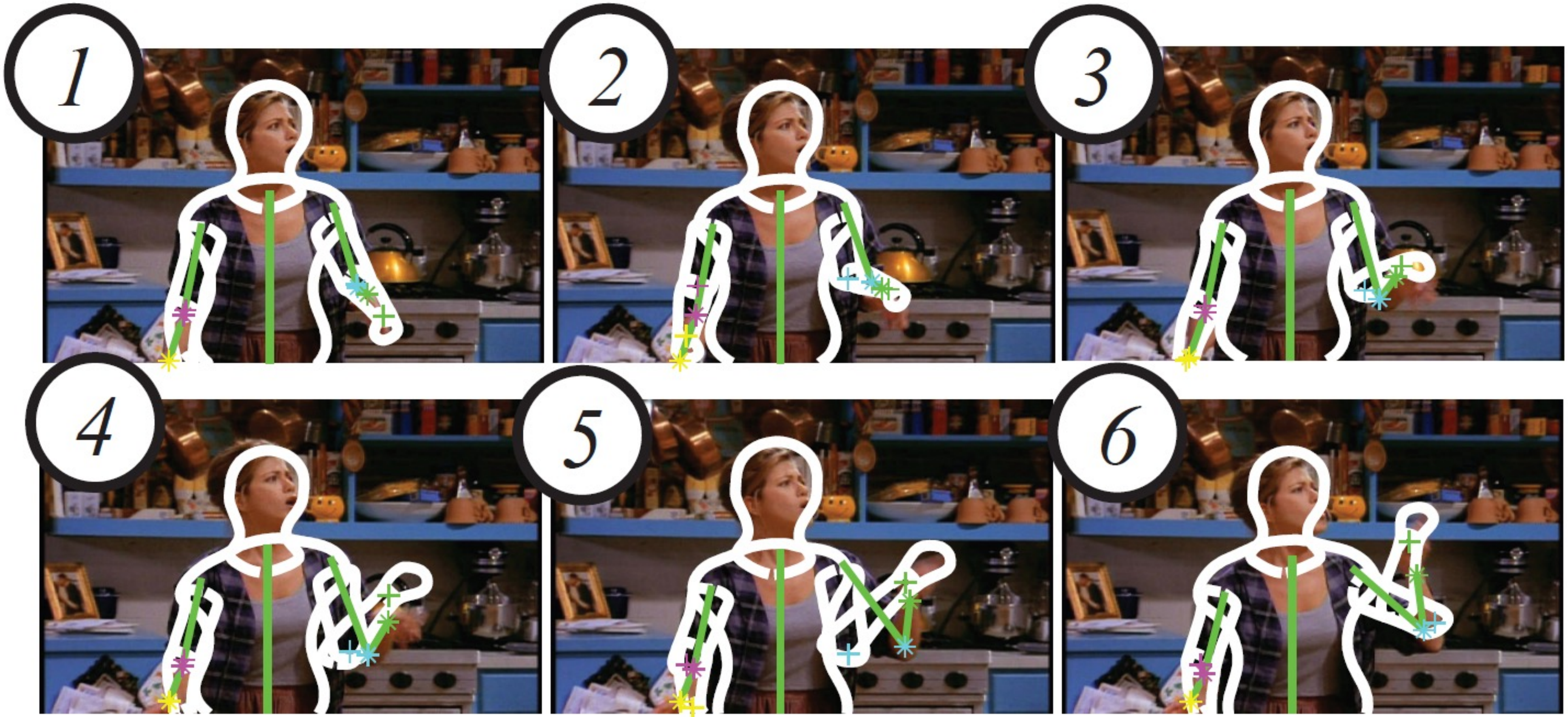
T-PMP : High precision, low recall, particles on one figure
D-PMP : Outperforms **FMP** and other particle methods

Note: G-PMP not reported due to poor performance.

[Pacheco, Zuffi, Black & Sudderth, ICML 2014]

Articulated Pose Tracking

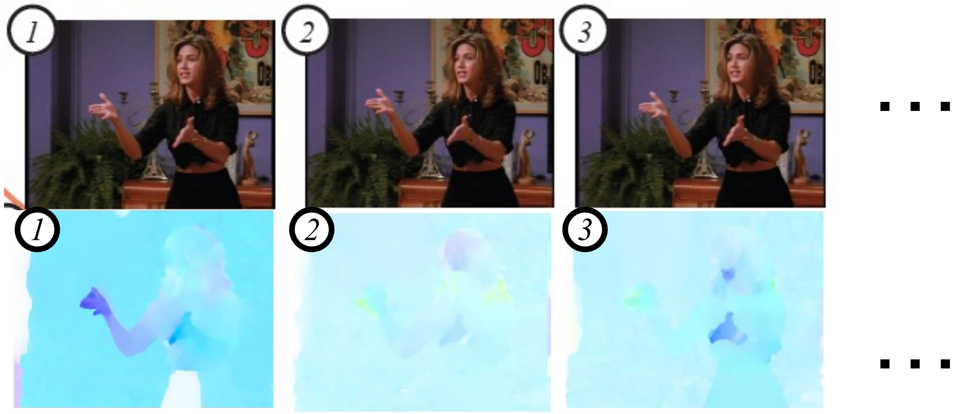
Prior work fails to show improvement by incorporating motion model.



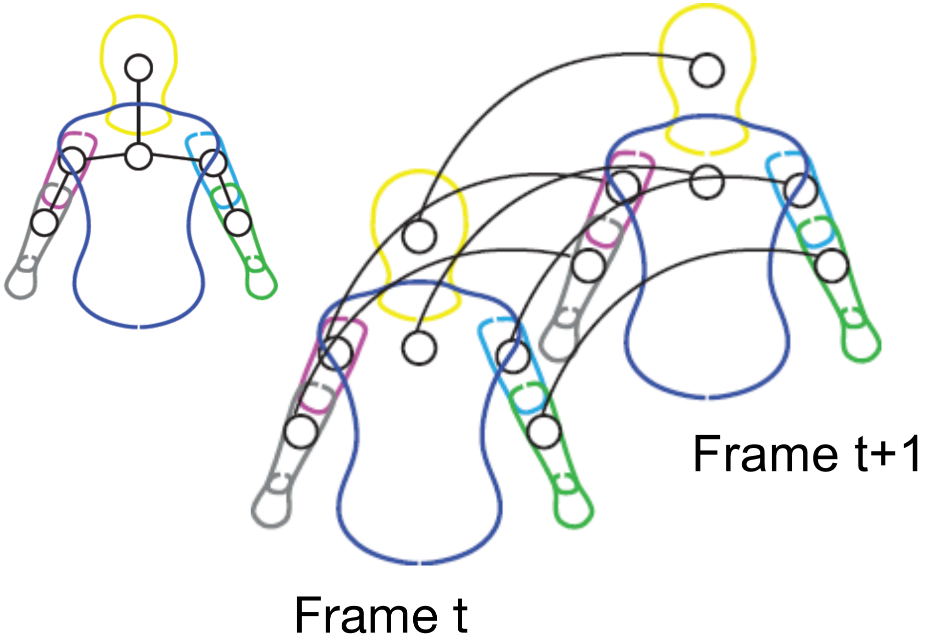
This is a failure of inference…

Articulated Pose Tracking

Data and Optical Flow

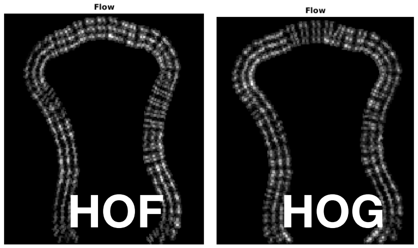


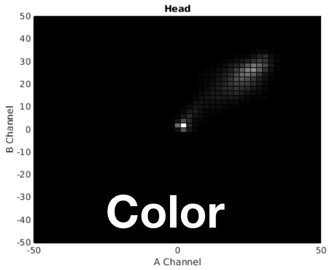
Extension of the *Flowing Puppets* model [Zuffi et al., 2013]



Part Likelihood

Gradients: Encode object and motion boundaries via HOG / HOF.

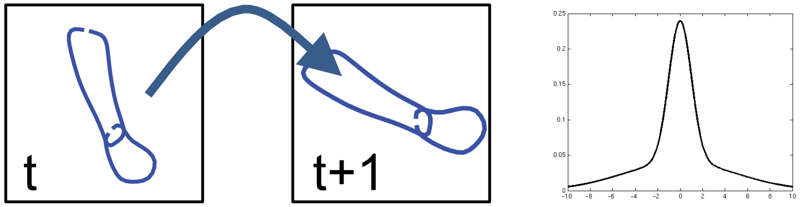




Appearance: 2D histogram of A/B color channels in L*a*b* space. Luminance ignored.

Prior

Part Motion: Scale mixture captures heavy tailed statistics of motion between frames.

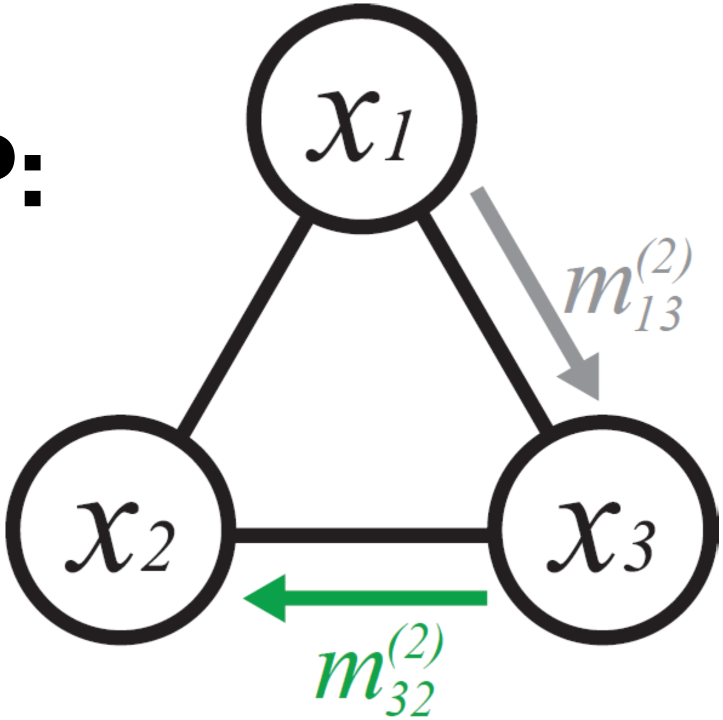


Structural prior identical to DS.

Loopy Max-Product BP

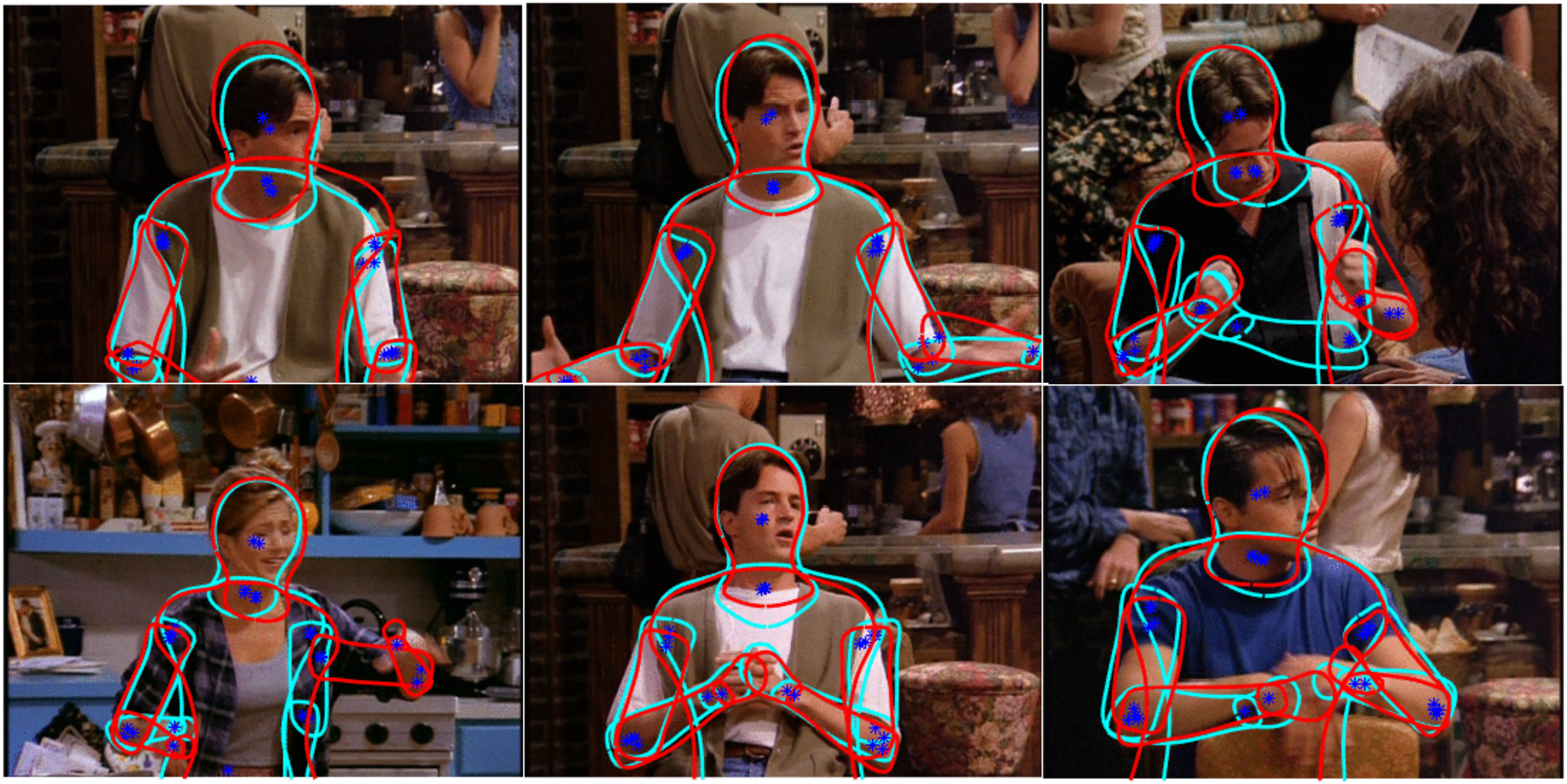
Many interesting models exhibit *cyclic* dependency structure...

Loopy Max-Product BP:
Iteratively update until converged.



Stay tuned later in the course for *reweighted* message passing…

VideoPose2 Experiments



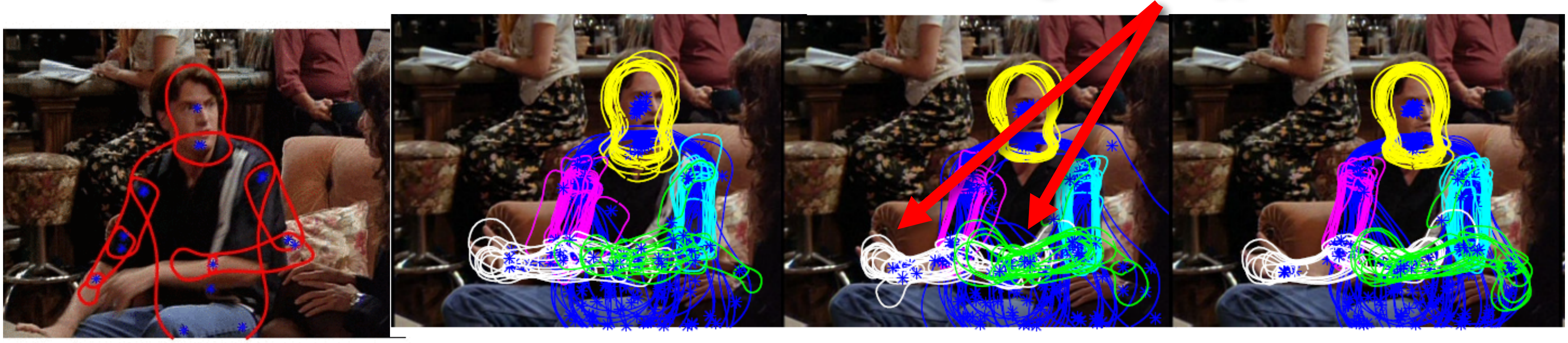
Comparison on VideoPose2 dataset of ~2,000 video frames from TV shows [Sapp et al., 2011]

🟦 T-PMP
🟥 D-PMP

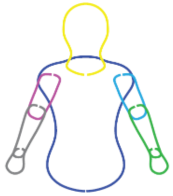
Pose Tracking Particles

T-PMP

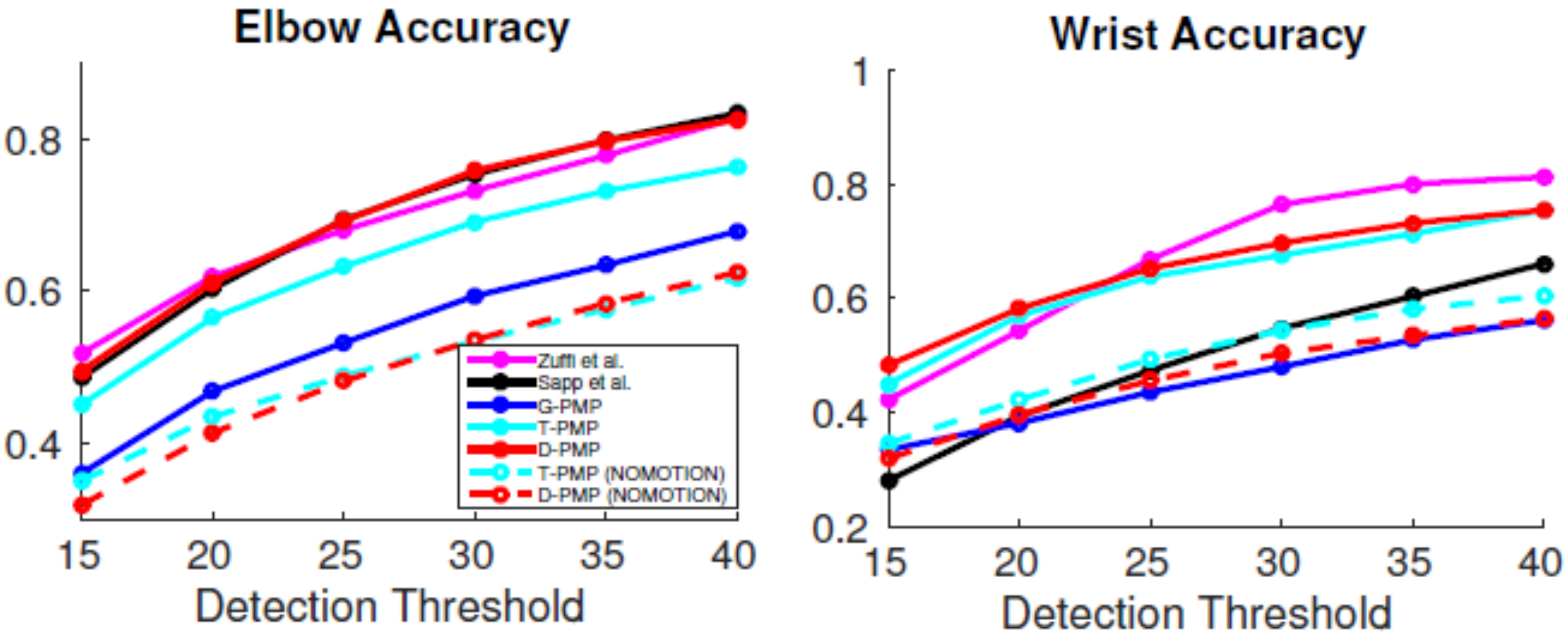
D-PMP **Both right arm hypotheses**



Greater diversity in particles allows D-PMP to reason more globally



VideoPose2 Experiments [Sapp et al. 2011]



- ➢ Superior to single-frame estimates (--,--)
- ➢ Clear improvement over Sapp et al. baseline
- ➢ D-PMP superior to Flowing Puppets in close detection ranges. Looking at failure cases.

D-PMP for 3D Mesh Alignment

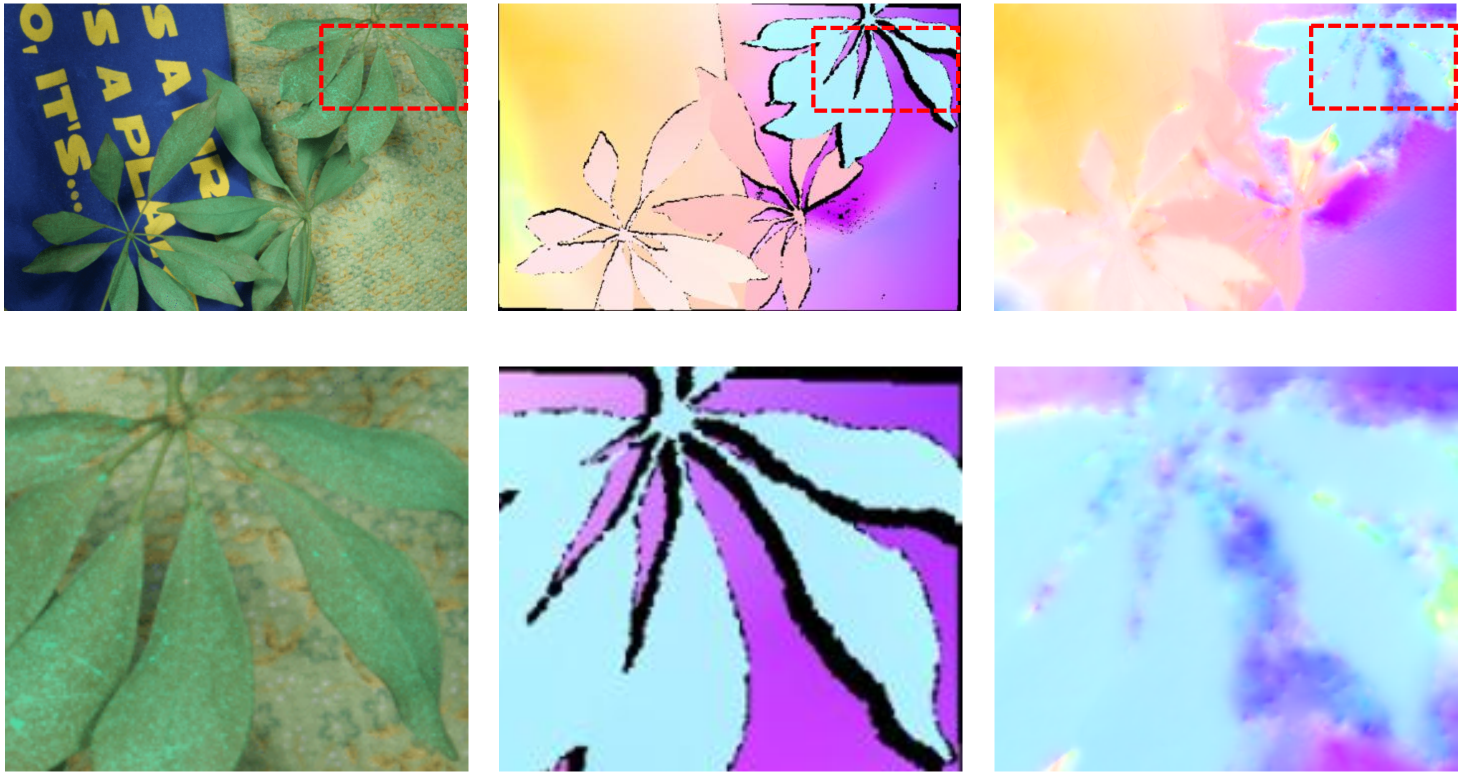
Independent work by Zuffi & Black, appeared at CVPR 2015.

Motion Estimation: Optical Flow

Occlusion boundaries are crucial for accurate motion estimation

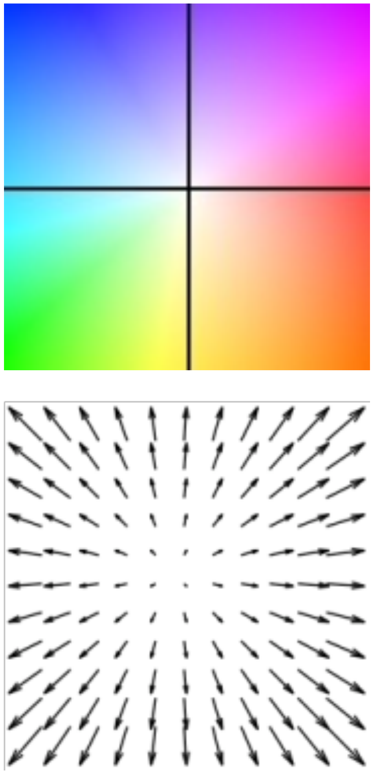
Middlebury Benchmark: Ground Truth

Horn & Schunck (1981)



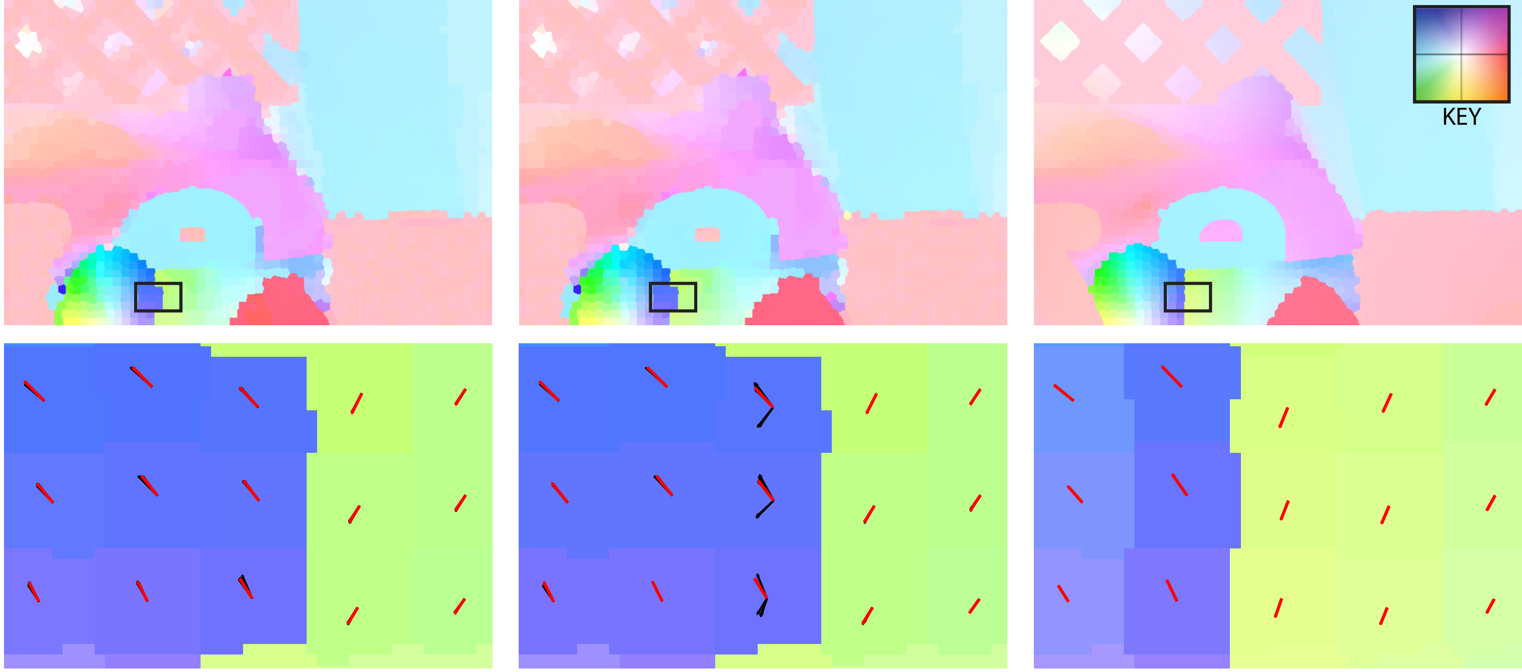
Occluded regions in black

Gaussian MRF

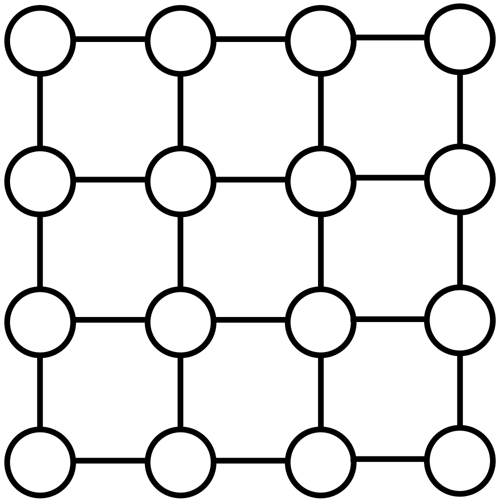


Need non-Gaussian models to capture natural motion statistics.

Optical Flow Estimation

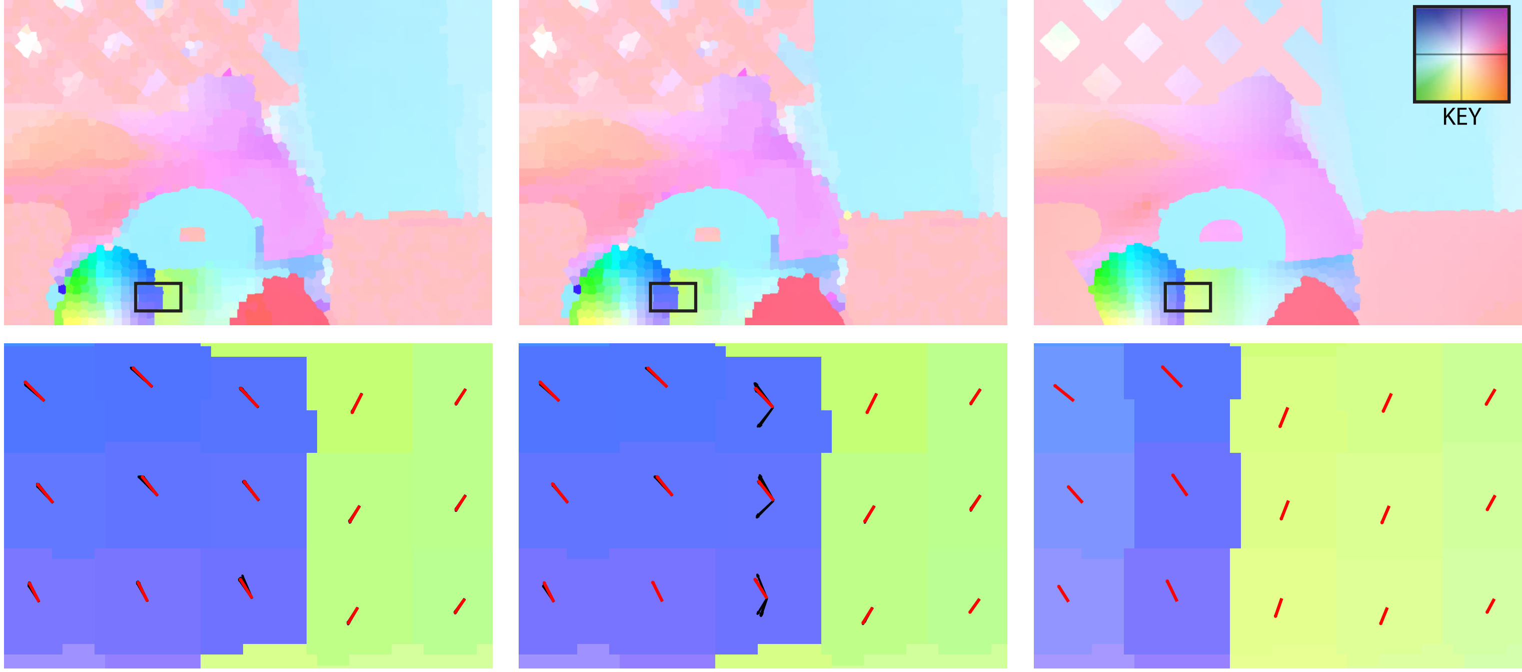


T-PMP **D-PMP** **Ground Truth**

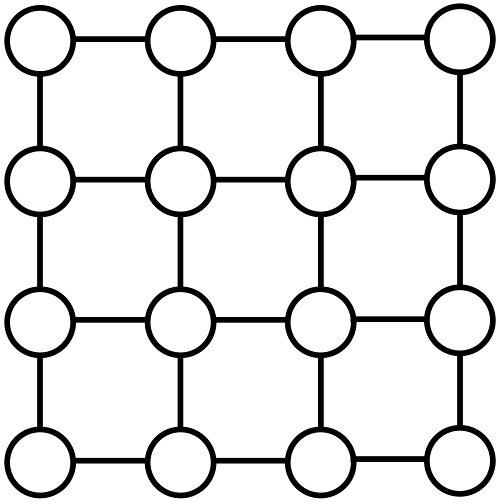


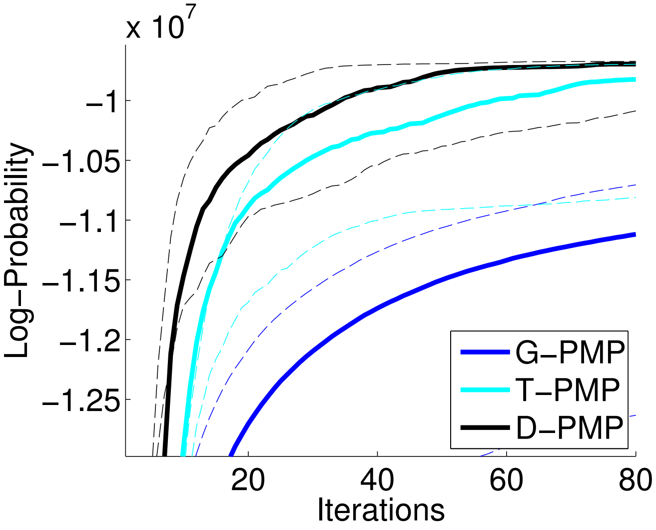
- Robust MRF (Sun, Roth, Black, IJCV 2014), discretization needs *~100,000 flow vectors per pixel* for good accuracy
- Low-level MRF often makes errors at occlusion boundaries, but *D-PMP preserves true flow as secondary hypothesis*
- Theory: Often have *global MAP on particle set*

Optical Flow Estimation

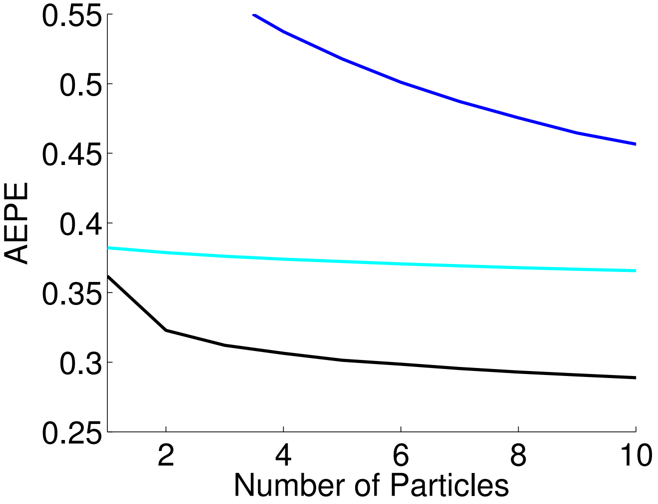


T-PMP **D-PMP** **Ground Truth**





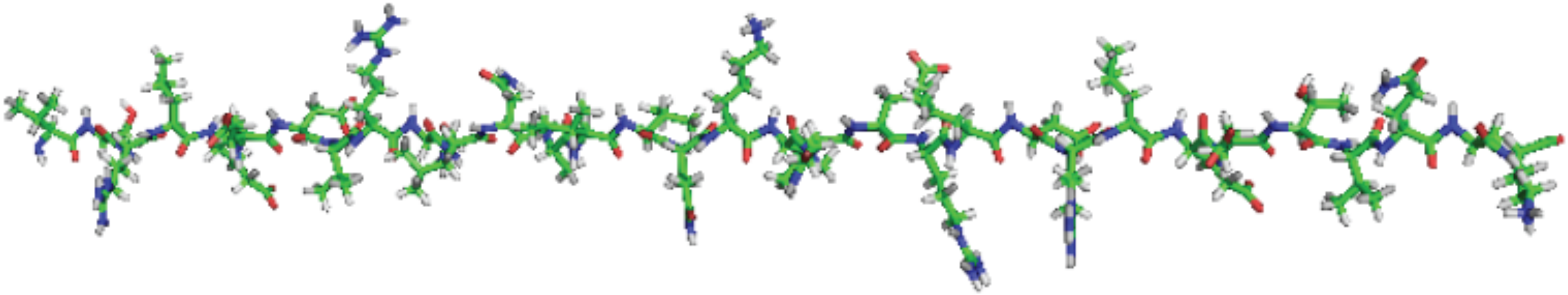
Log-Probability



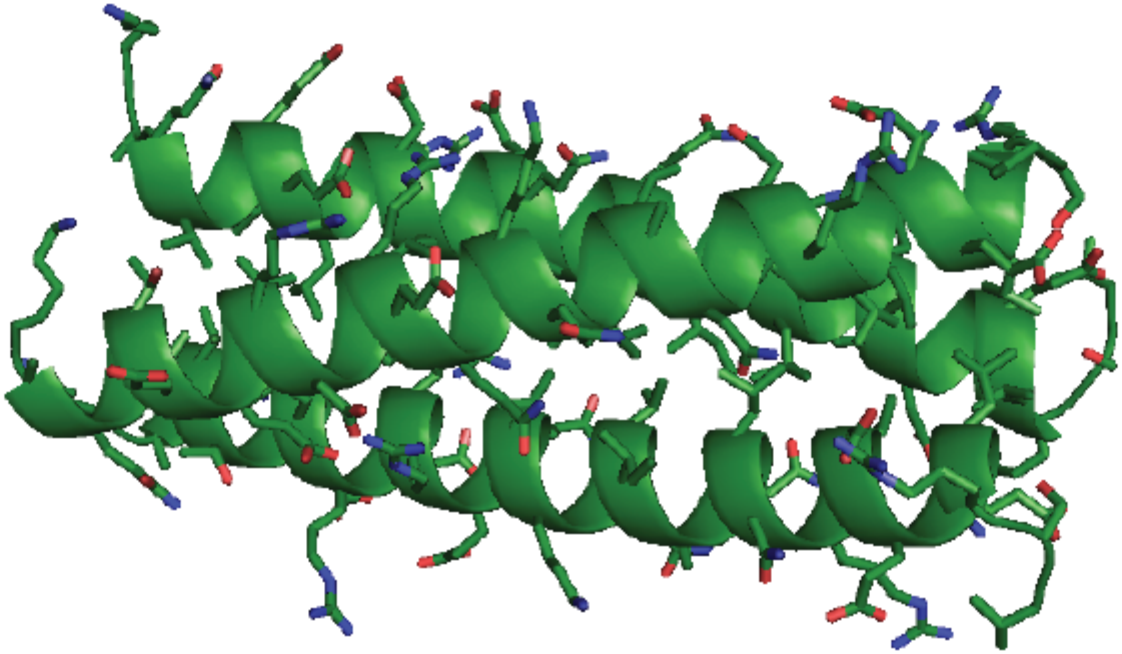
Oracle Avg. EPE

Reweighted PMP comparison on a "superpixel" graph with ~10,000 nodes

Protein Structure Prediction

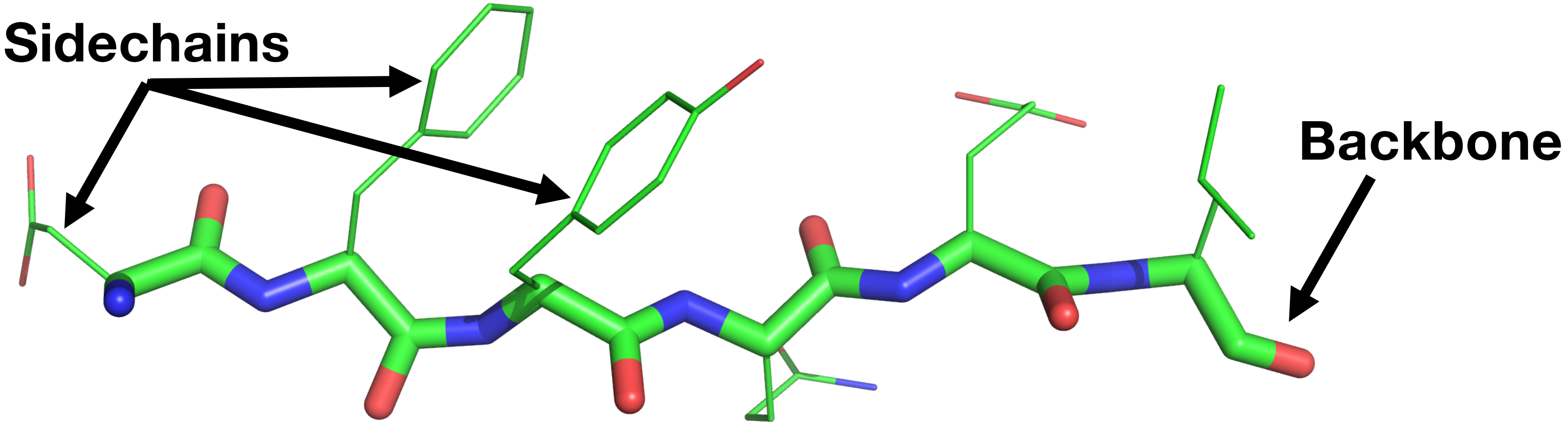


V-S-R-L-E-E-D-V-R-N-L-N-A-I-V-Q-K-L-Q-E-R-L-D-R-L-E-E-T-V-Q-A-K

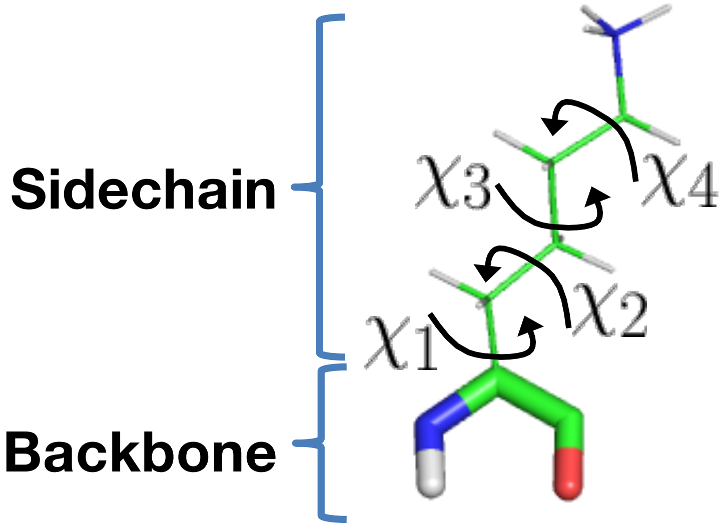


All information for predicting 3D structure encoded in amino acid sequence and physics

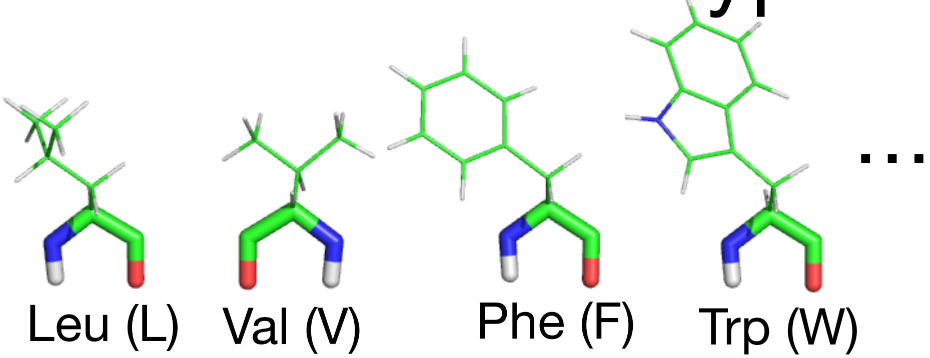
Protein Side Chains



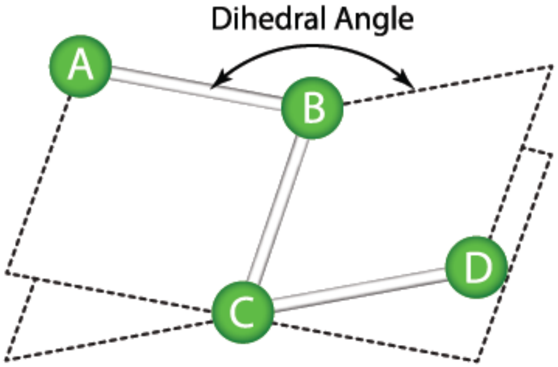
Side chain prediction: Estimate side chains given fixed backbone.



20 Amino Acid Types



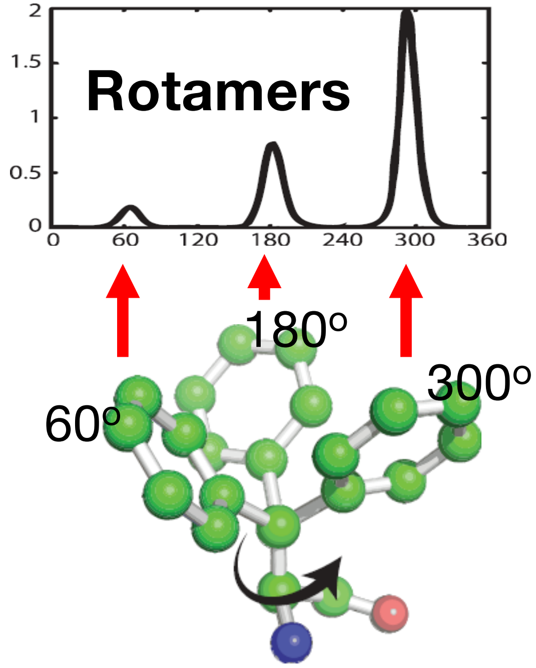
Dihedrals and Rotamers

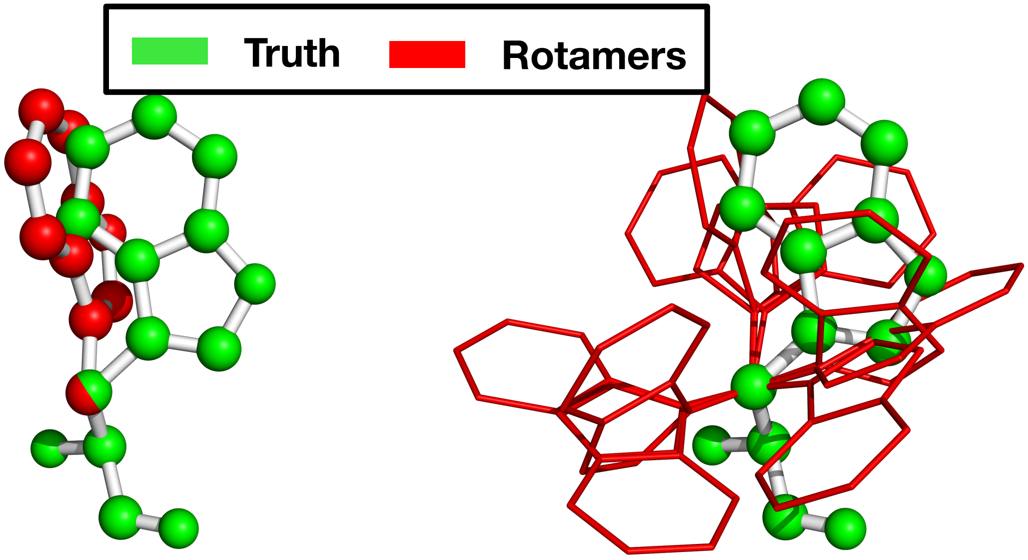


Dihedral Angles:

- ➢ Compact angular encoding
- ➢ 1D-4D continuous state

Rotamer discretization based on marginal statistics fails to capture fine details…

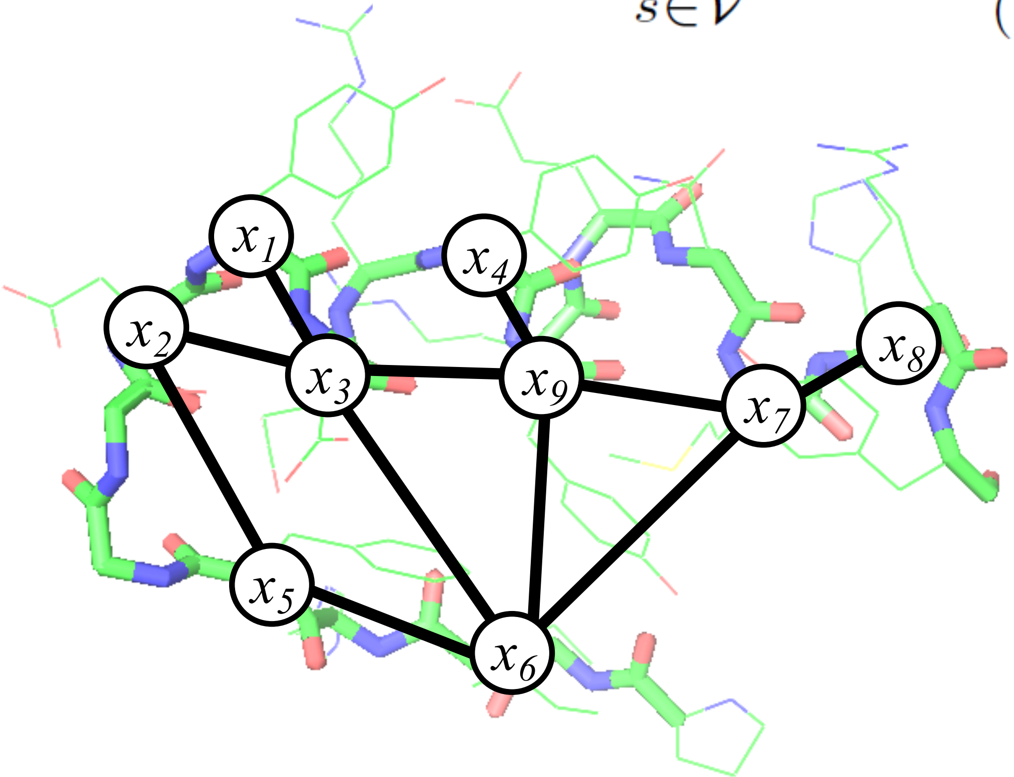




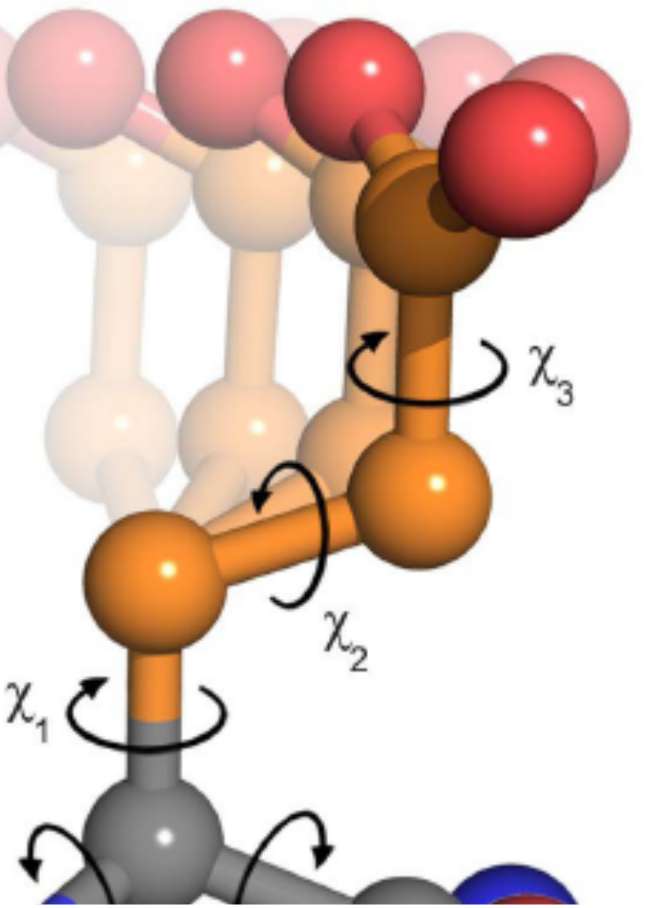
[Shapovalov & Dunbrack 2007]

Side Chain Prediction

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



Edges between amino acids within distance threshold.

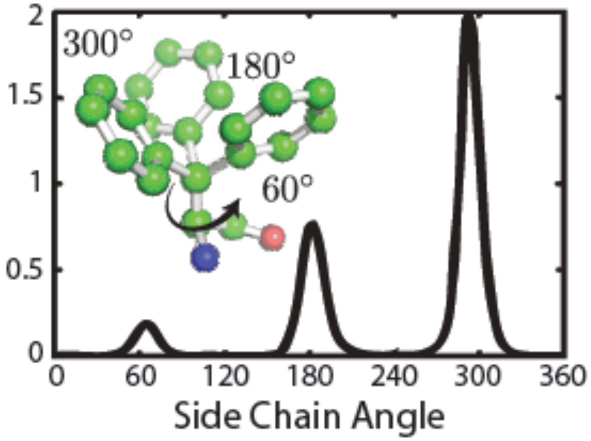


[Image: Harder et al., BMC Informatics 2010]

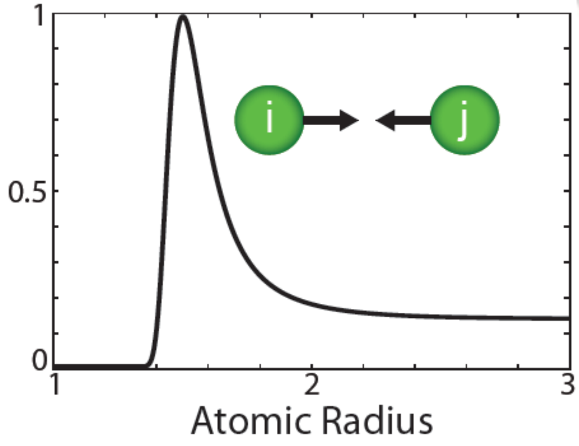
Side Chain Prediction

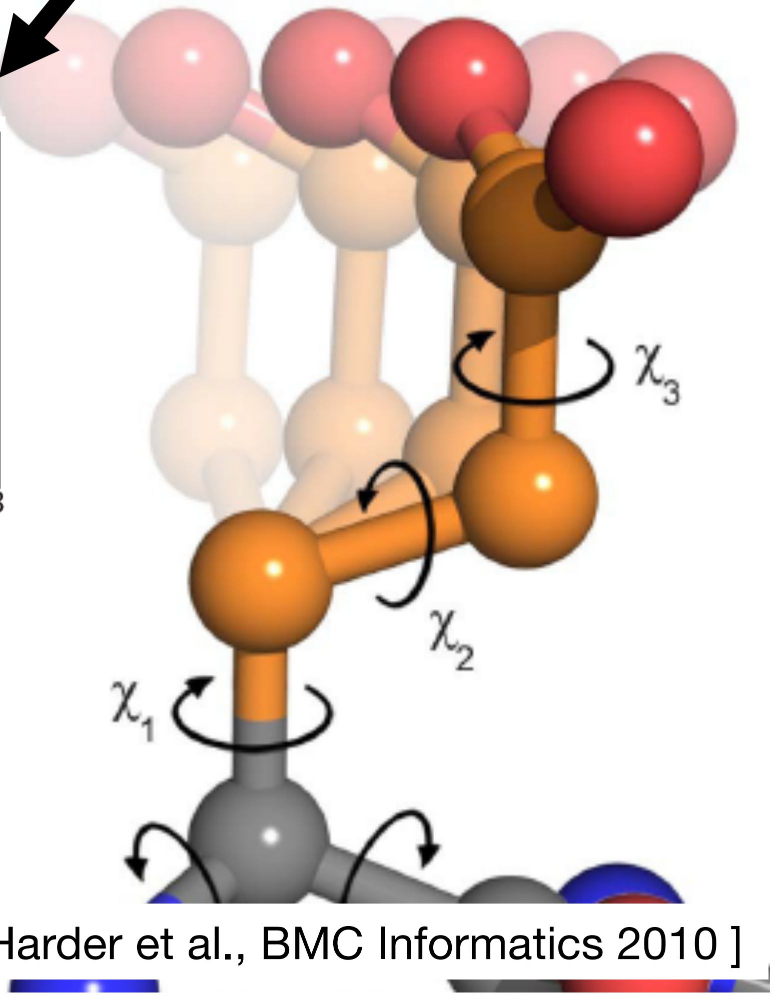
$$p(x) \propto \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t)$$

Rotamer Likelihood



Atomic Interaction

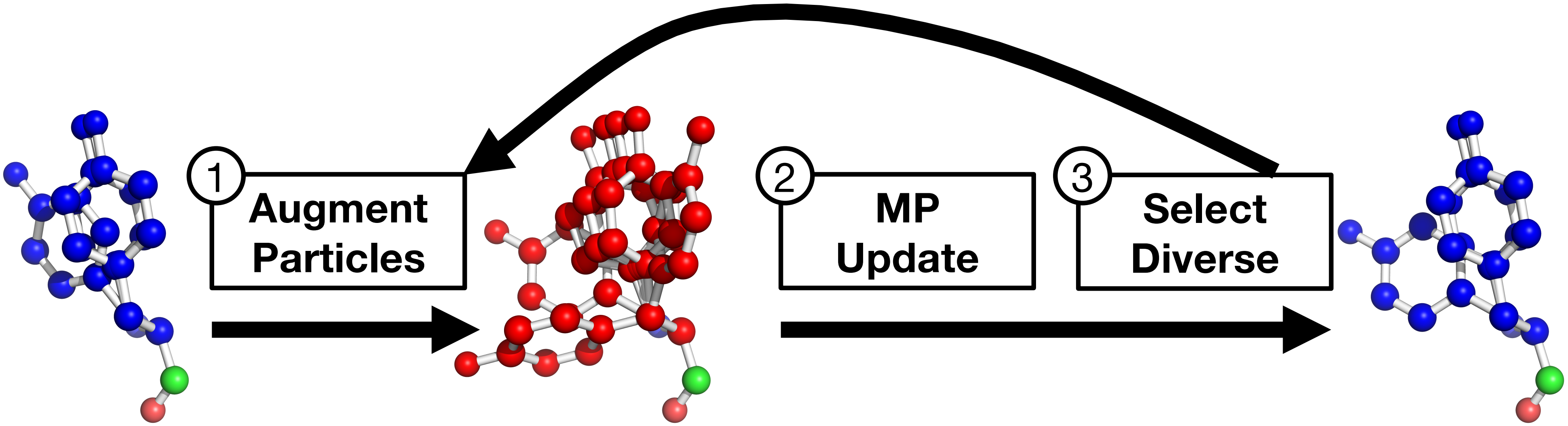




Statistical and physical potential functions.

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

D-PMP for Side Chains



① **Augment Particles**

② **MP Update**

③ **Select Diverse**

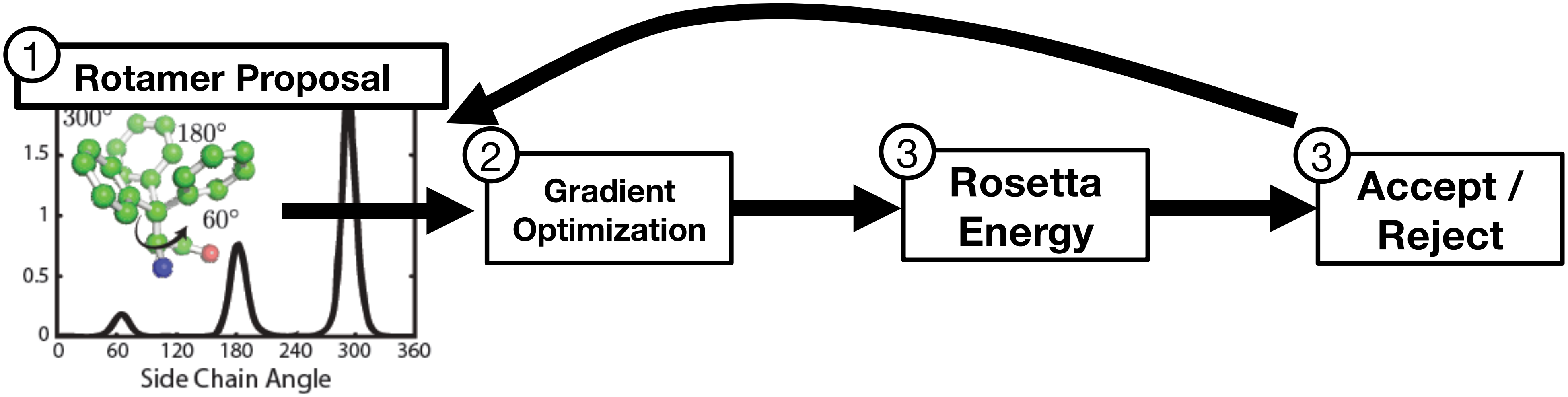
Continuous optimization of side chains:

➢Captures non-rotameric side chains

➢Conformational diversity

➢Likelihood-based proposals

Rosetta



① **Rotamer Proposal**

② **Gradient Optimization**

③ **Rosetta Energy**

③ **Accept / Reject**

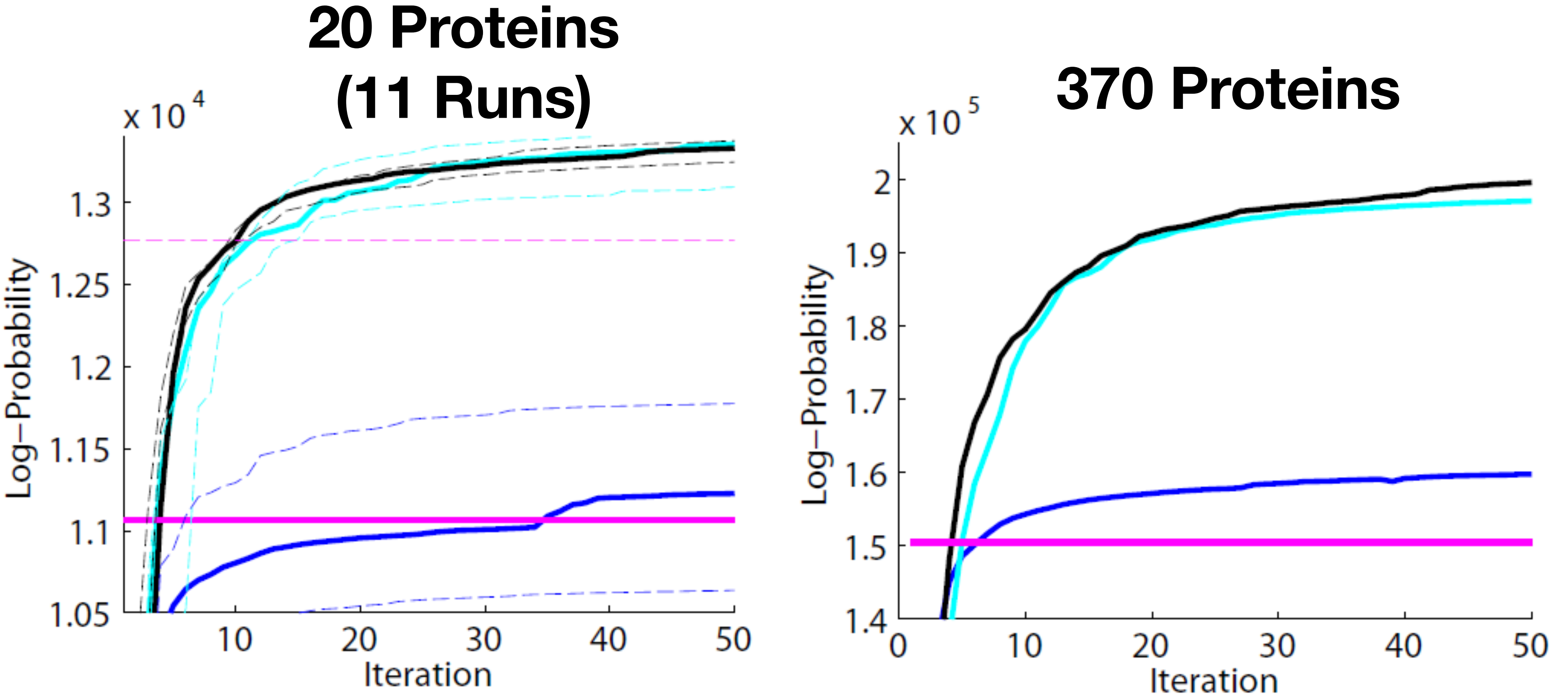
- ➢ Energy model used in *FoldIt* game
- ➢ Simulated annealing (SA) Monte Carlo
- ➢ Independent chains for multiple optima

Rosetta Commons

Replace SA with D-PMP. Use Rosetta as black-box energy method.

Protein Side Chain Prediction

Log-probability of MAP estimate for…



20 Proteins (11 Runs)

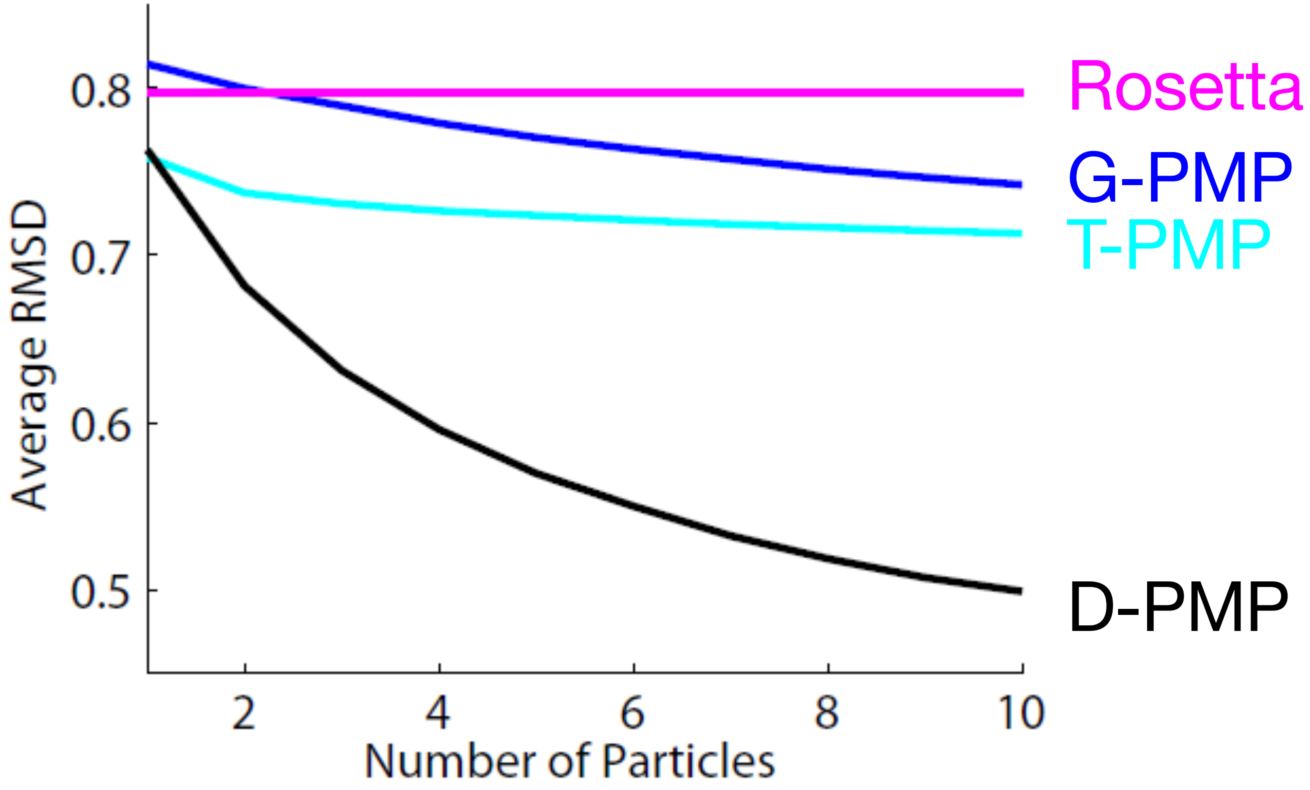
370 Proteins

G-PMP, T-PMP, D-PMP, Rosetta simulated annealing [Rohl et al., 2004]

[Pacheco et al., ICML 2015]

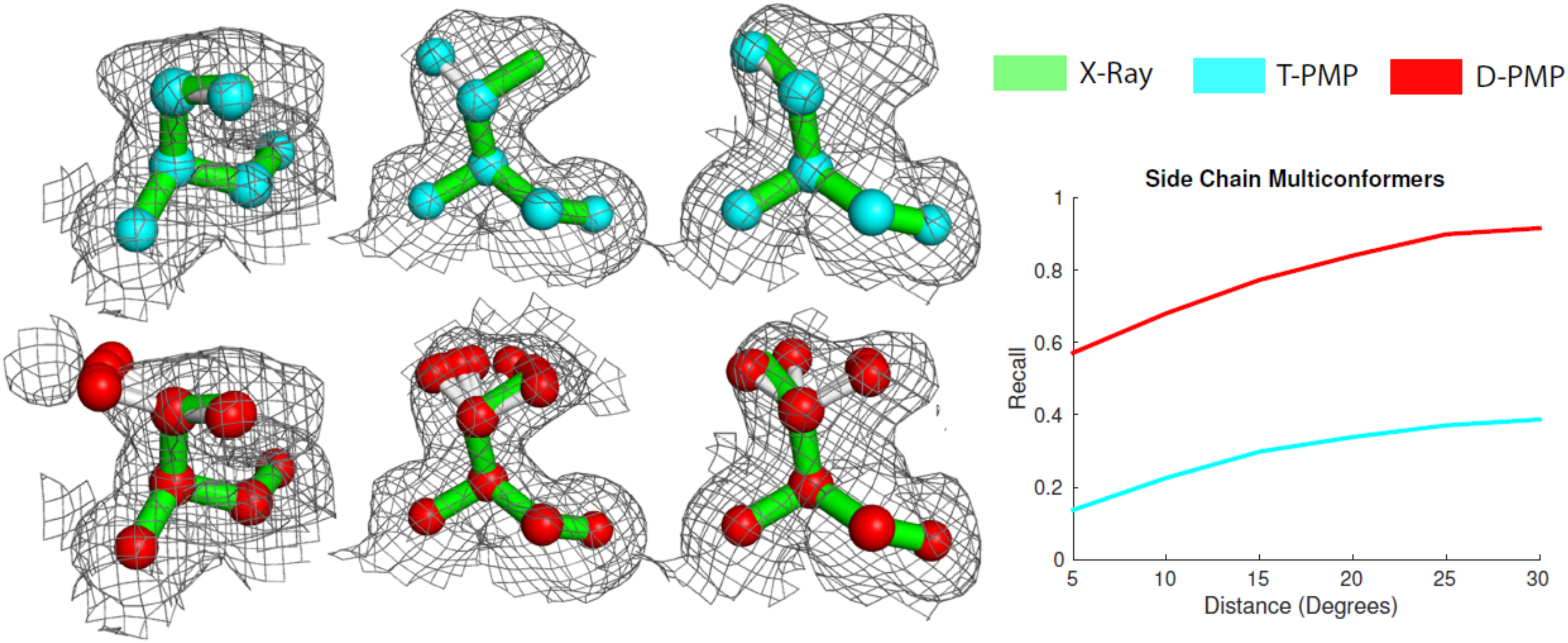
Protein Side Chain Prediction

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



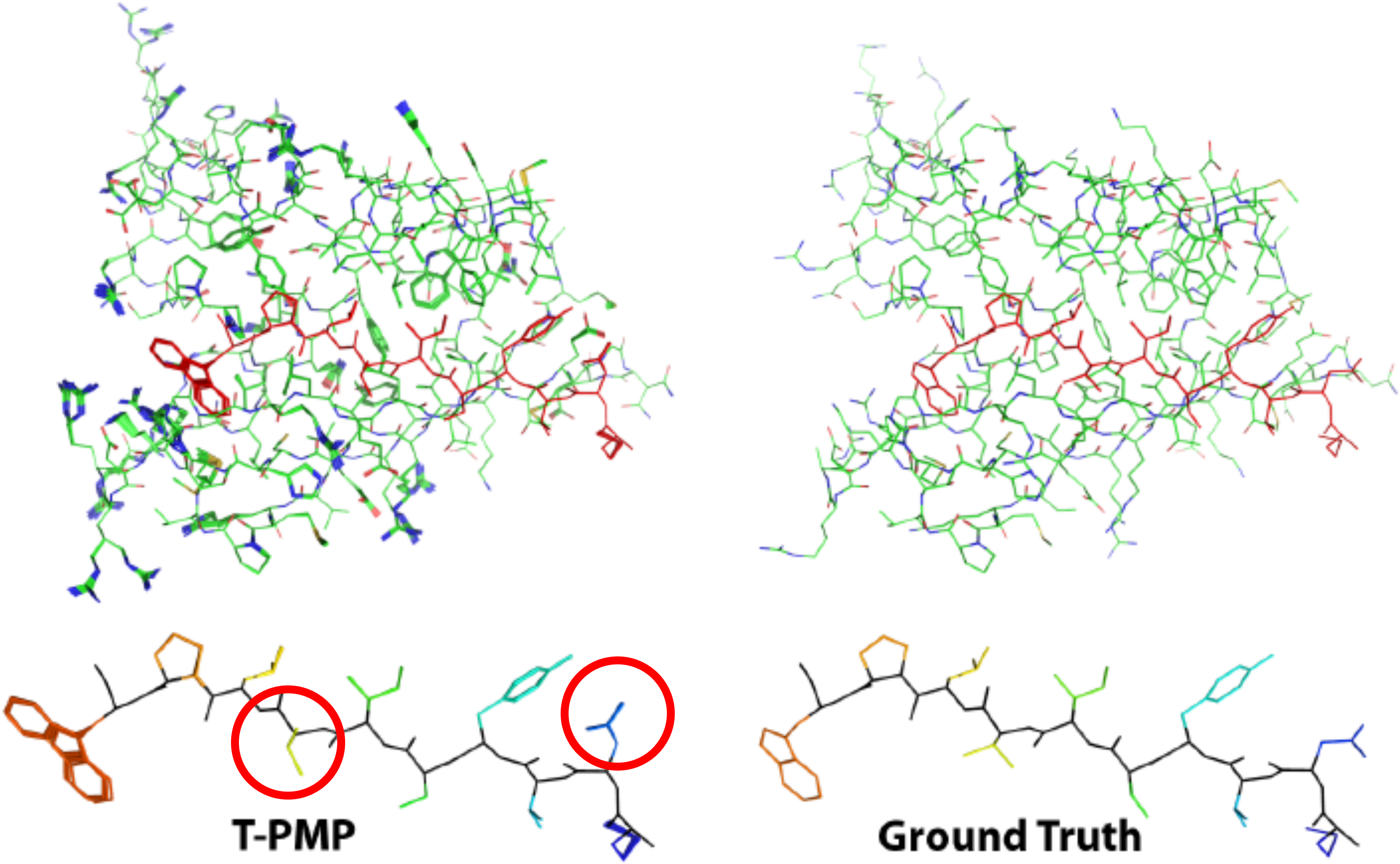
Oracle selects best configuration in current particle set.

Multiple Side Chain Conformations

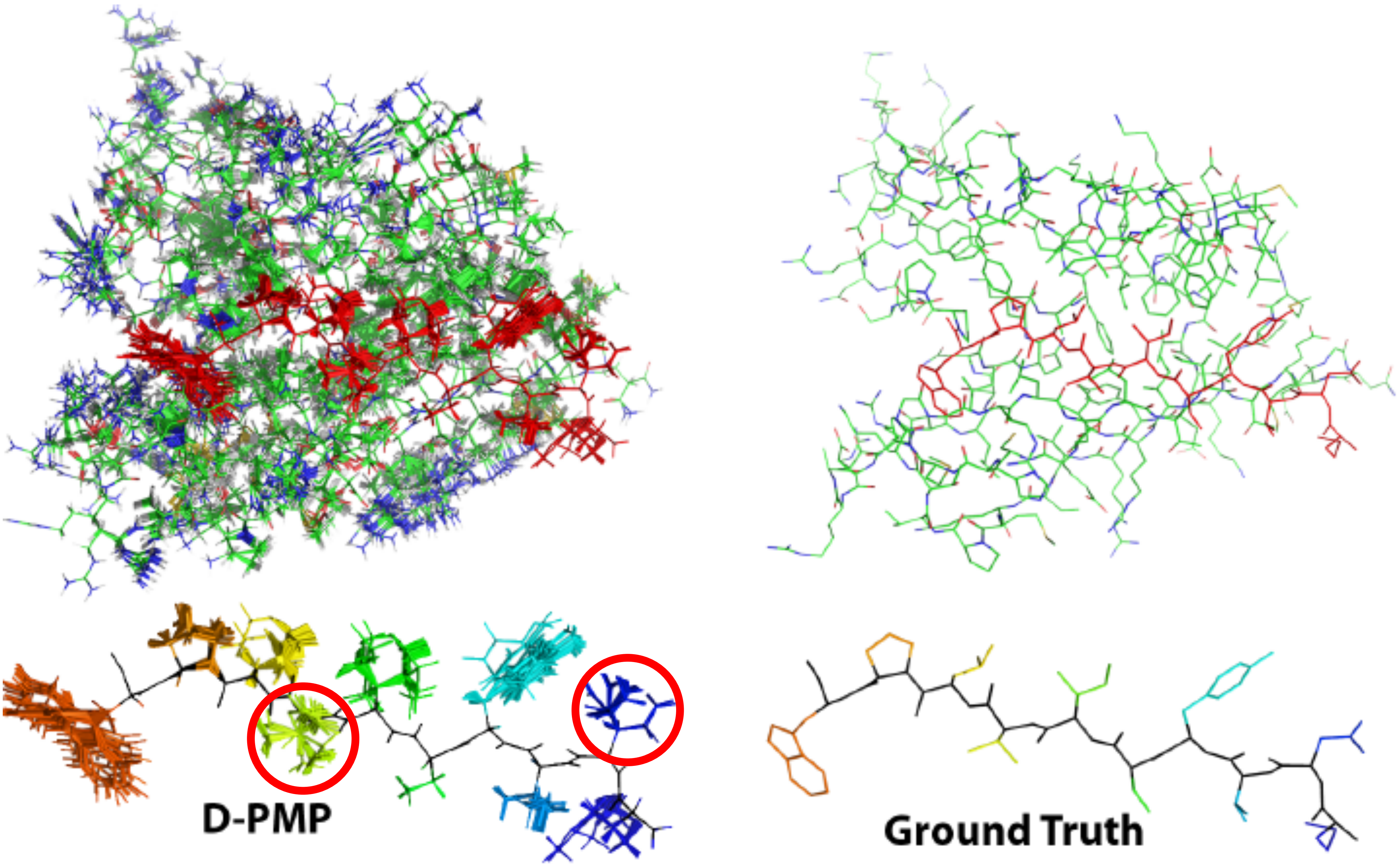


- Side chains don’t exist in a single conformation
- Diversity in D-PMP particles captures multiple alternate states.
- T-PMP particles get stuck in local optima

Protein Side Chain Prediction



Protein Side Chain Prediction



Contributions

Reliable particle-based MAP inference
for graphical models with continuous variables:
object shape, articulation, position, …

Validation: *Inference of multiple poses, motions, protein conformations, …*