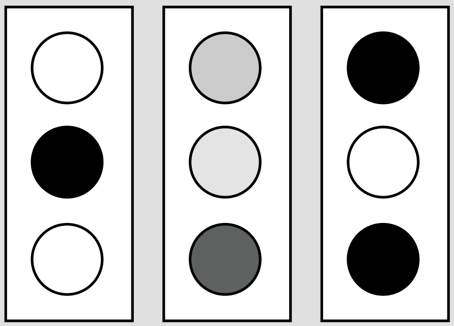
Composable, distributed-state models for high-dimensional time series

Graham Taylor

Examination Committee:

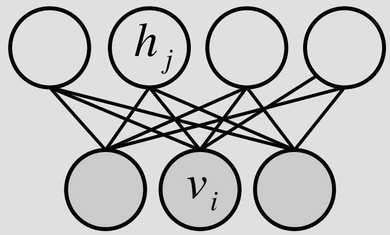
Professors G. Hinton (Co-supervisor), S. Roweis (Co-Supervisor), R. Zemel, C. Bregler (External Examiner) and D. J. Rowe (Chair)

Distributed



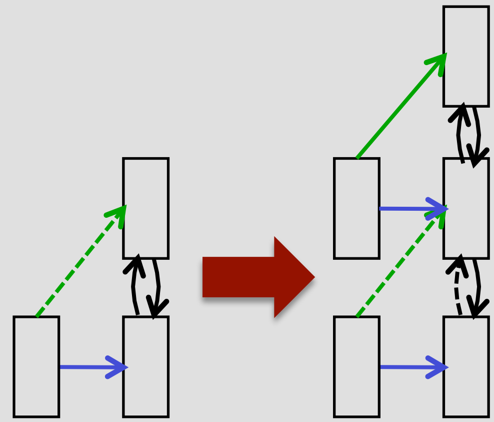
- Capable of representing data that is a product of multiple underlying influences

Undirected



- Using an RBM makes exact inference easy

Composable



- Train greedily, layer-by-layer

Modeling human motion

- Capture the movement of a subject as a time series of 3D cartesian coordinates
- High-dimensional (60-100), nonlinear, long-range deps
- Large repositories available
- Extension to tracking (Ch. 6)

Related work

Concatenation

Transforming existing motion

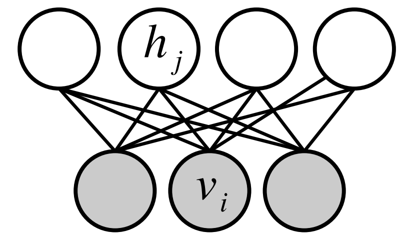
Interpolation/ Blending

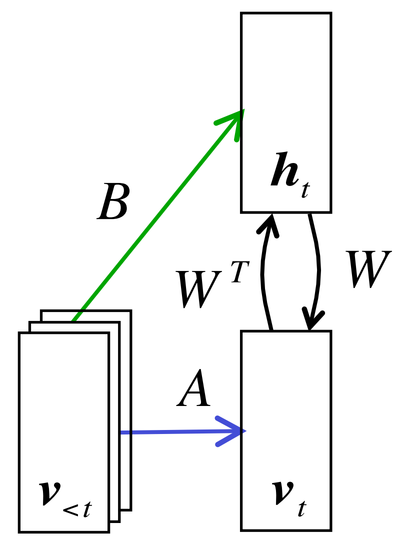
Physics-based methods

Generative models-
Our method is based on a "pure" learning approach. It is able to generalize well while avoiding the complexity of explicitly imposing physics-based constraints.

Conditional Restricted Boltzmann Machines

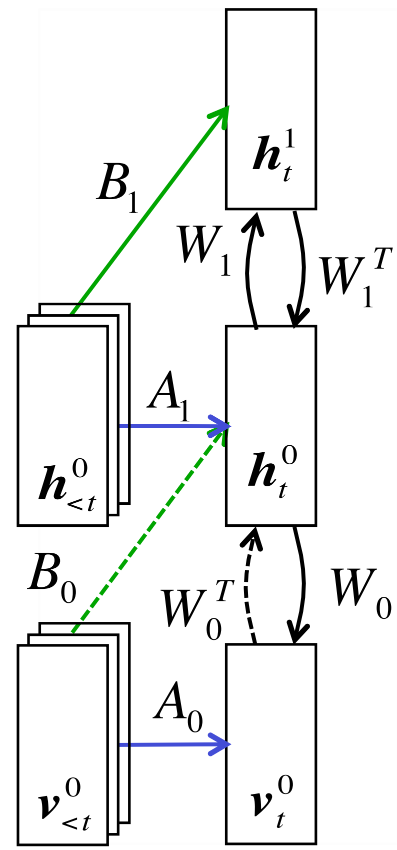
- Start with an RBM (binary-binary or real-binary)
- Add two types of directed connections
- Does not change inference and learning
- Autoregressive weights model short-term, linear structure



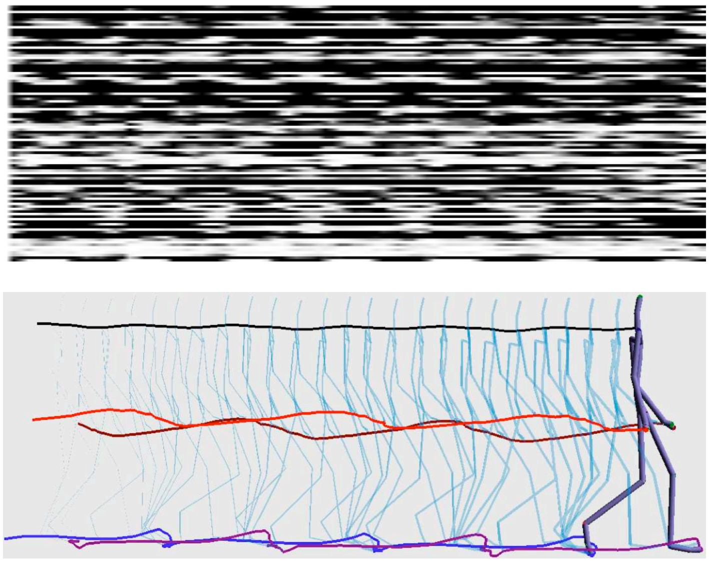


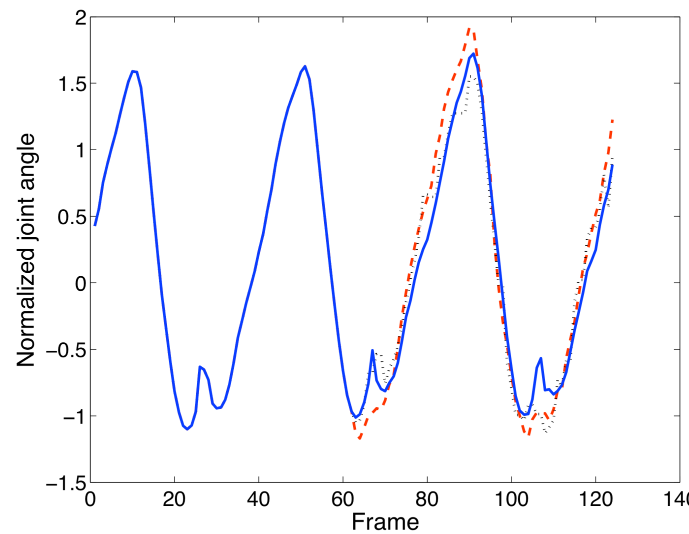
Conditional deep belief networks

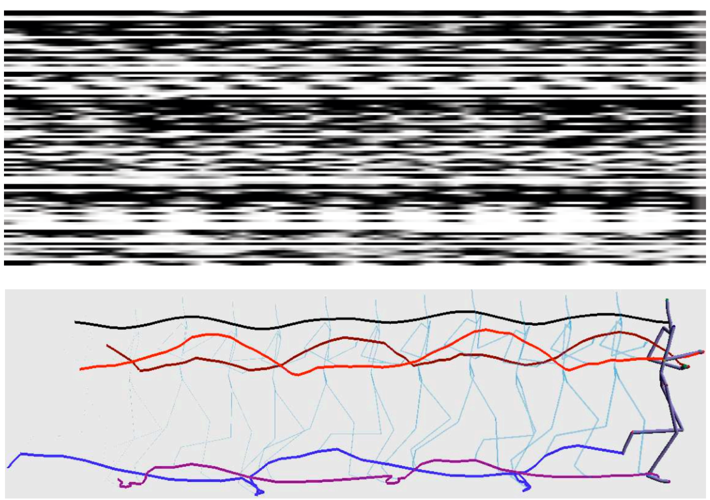
- CRBM defines *p(v^0,h^0)* – implicitly *p(h^0),p(v^0|h^0)*
- Consider "trading in" *p(h^0)* for a better model
- Subject to conditions (which we violate) – guaranteed to never decrease a variational lower bound on log prob

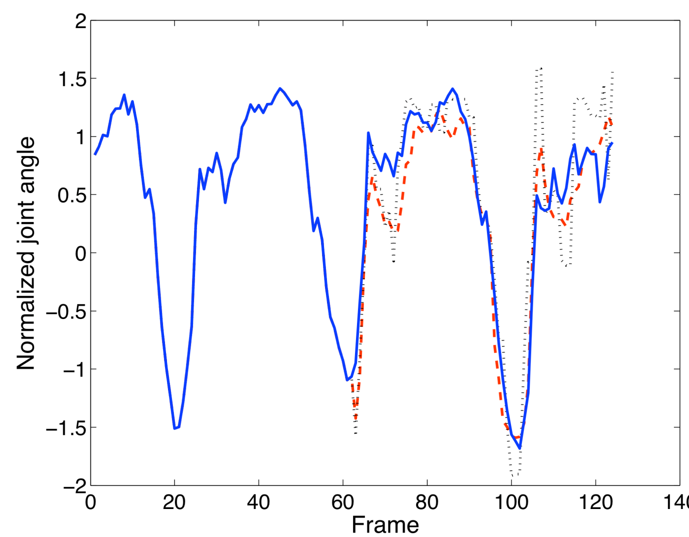


Results: synthesis and filling-in

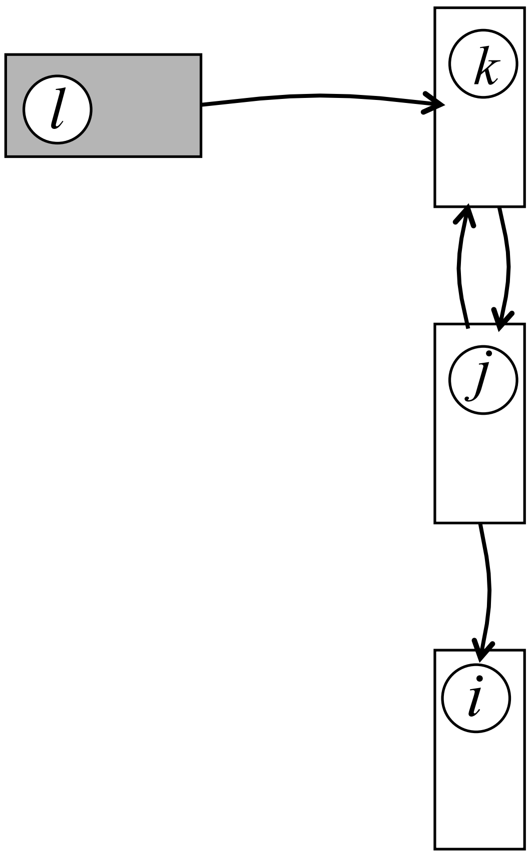


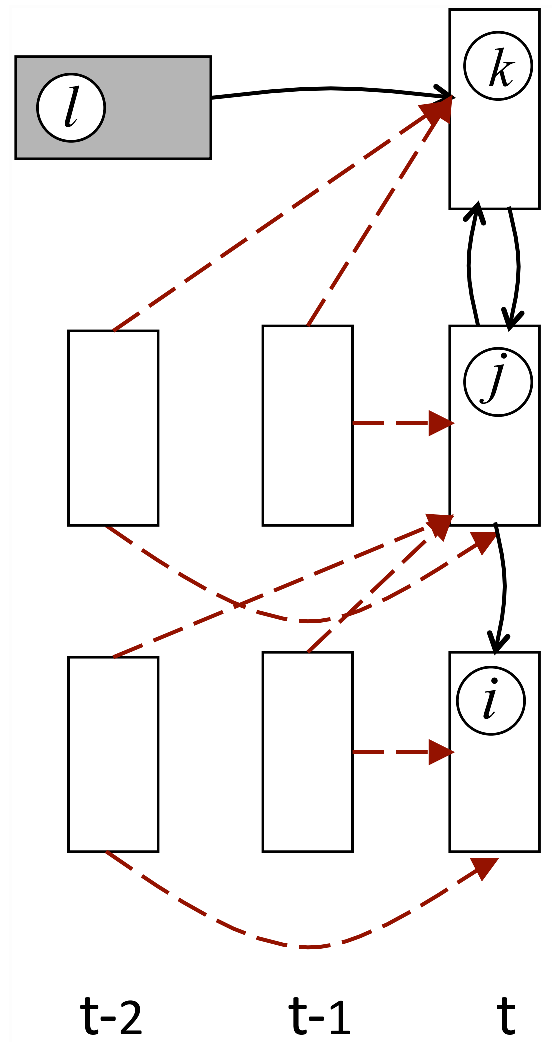






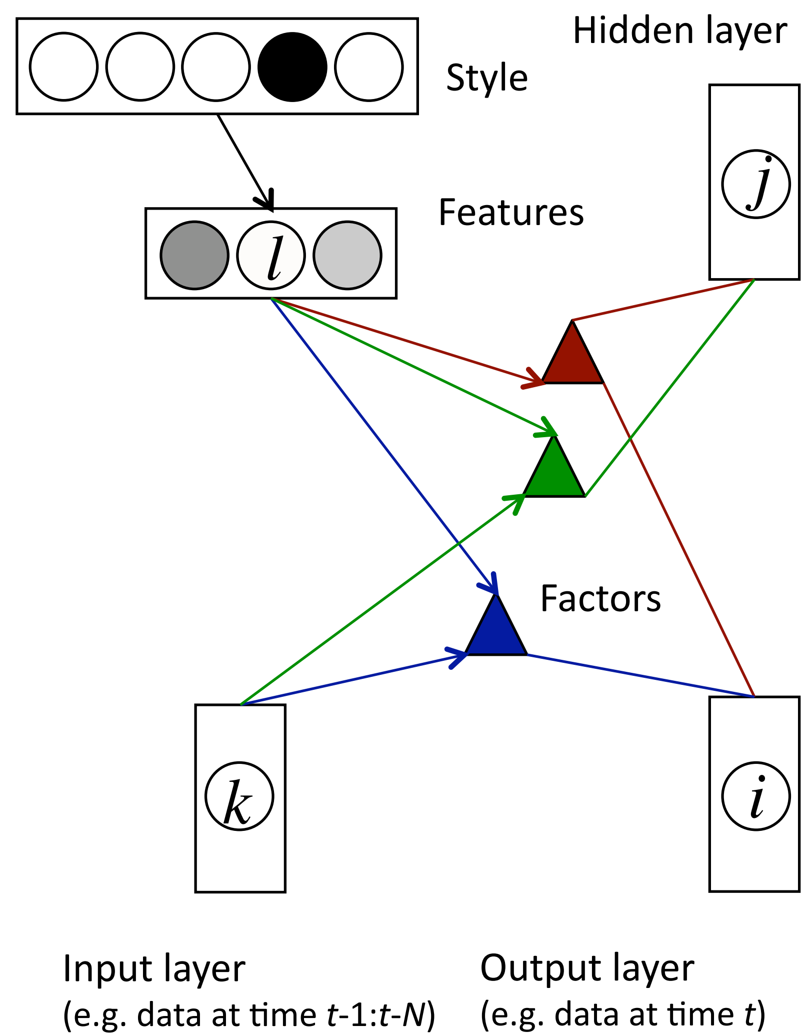
Learning style



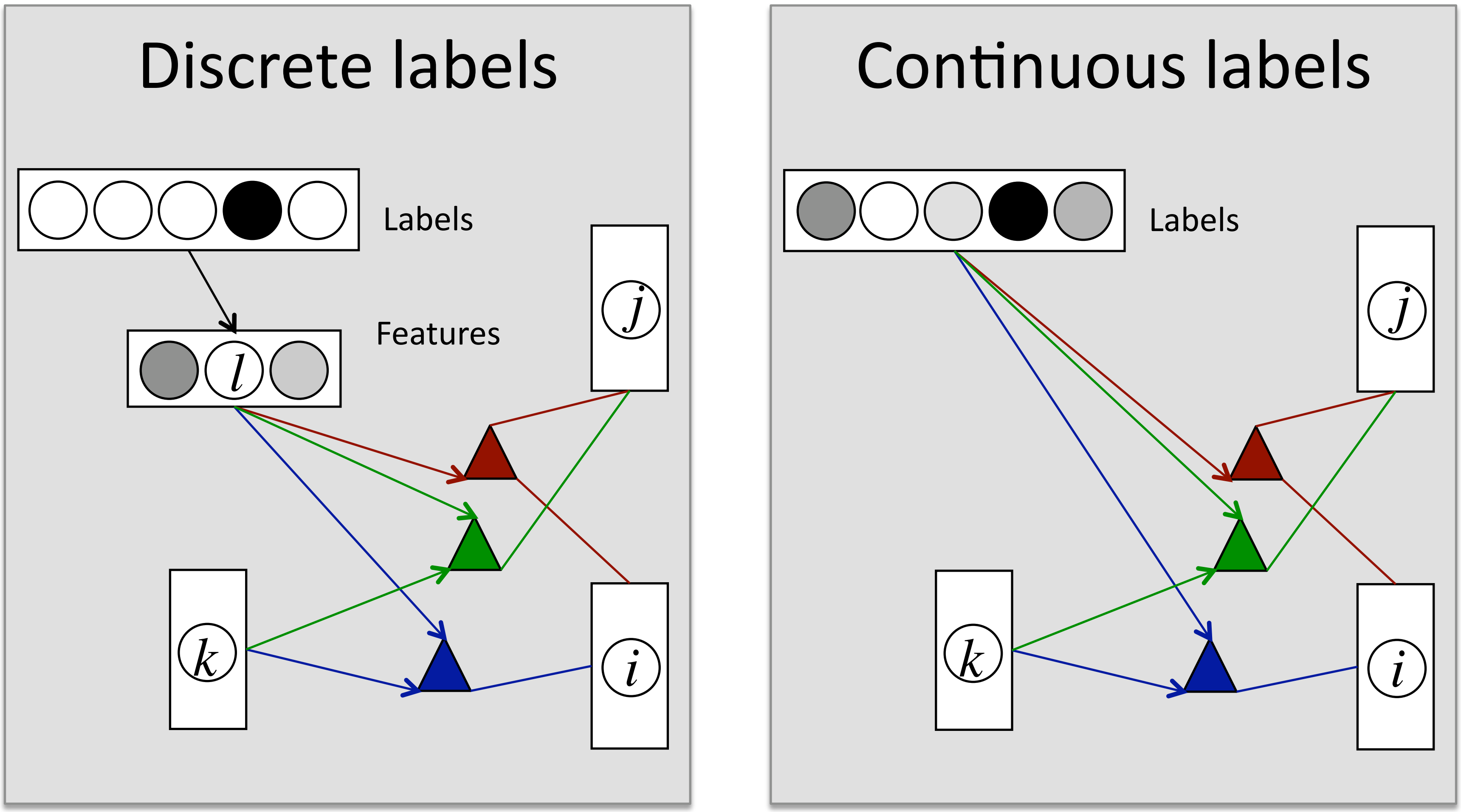


Factored Conditional RBMs

- Effective interaction weight between two units is modulated by the dynamic state of a third unit
- Factor implied three-way weight tensor, reducing parameters from $O(N^2)$ to $O(N^3)$
- Can blend and transition among motion styles

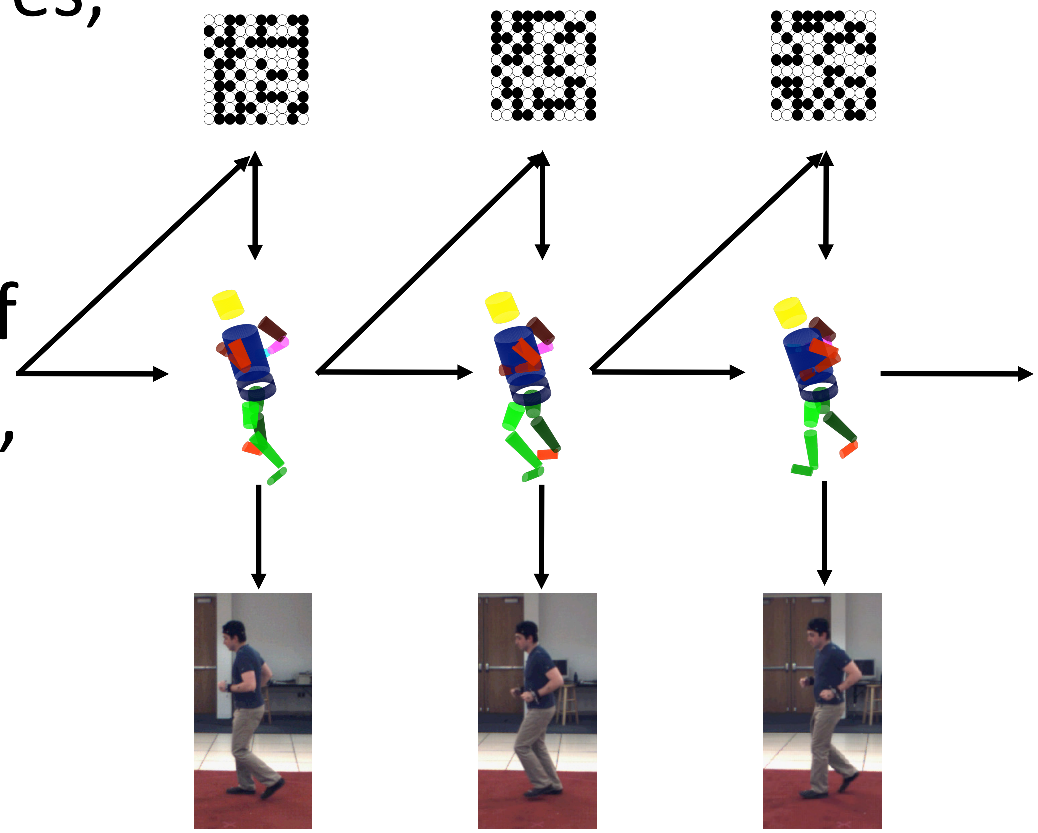


Results: synthesis



Tracking people with rich dynamical priors

- Given image features, infer 3-D pose at every frame
- Difficult because of noisy observations, occlusions, ambiguities
- Need good priors



Bayesian filtering

$$p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_t \mid \boldsymbol{x}_t)\, p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t-1})$$

Filtering distribution

Likelihood

Predictive distribution

$\boldsymbol{x}_t$: pose

$\boldsymbol{y}_t$: image features

$$p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t-1}) = \int_{\boldsymbol{x}_{t-1}} p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_{t-1}$$

Dynamical prior

Cannot be computed in closed form

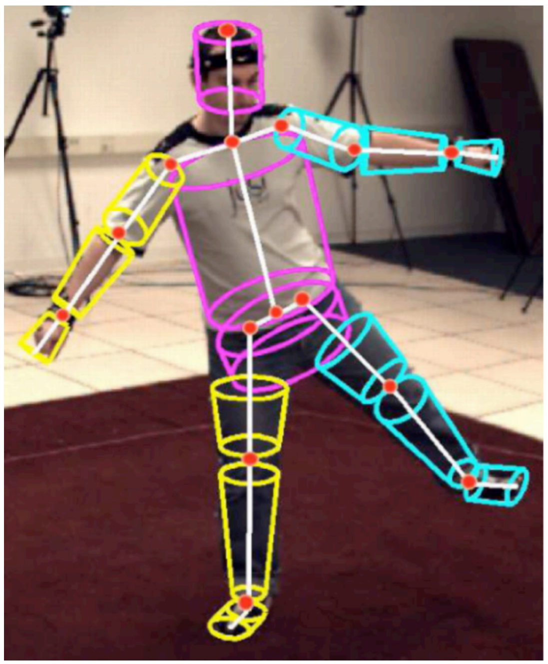
$$p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}) \approx \sum_{p=1}^{P} w_t^{(p)} \delta(\boldsymbol{x}_t - \boldsymbol{x}_t^{(p)})$$

$$\boldsymbol{x}_t^{(p)} \sim p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t-1})$$

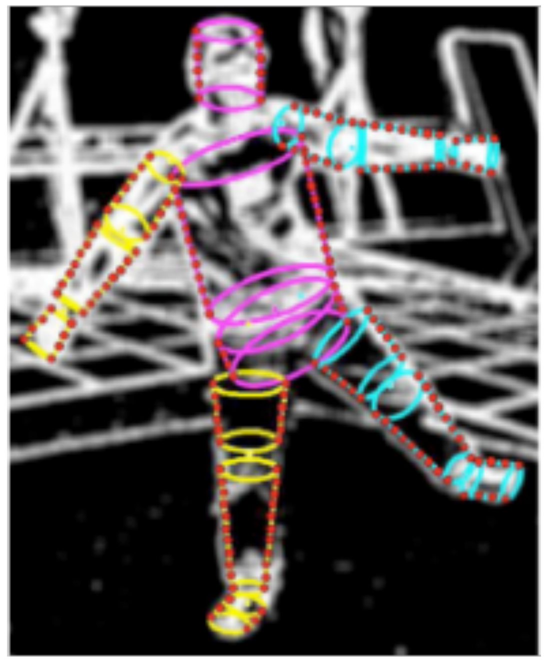
$$w_t^{(p)} \propto p(\boldsymbol{y}_t \mid \boldsymbol{x}_t^{(p)})$$

(Deutscher & Reid, 2005)

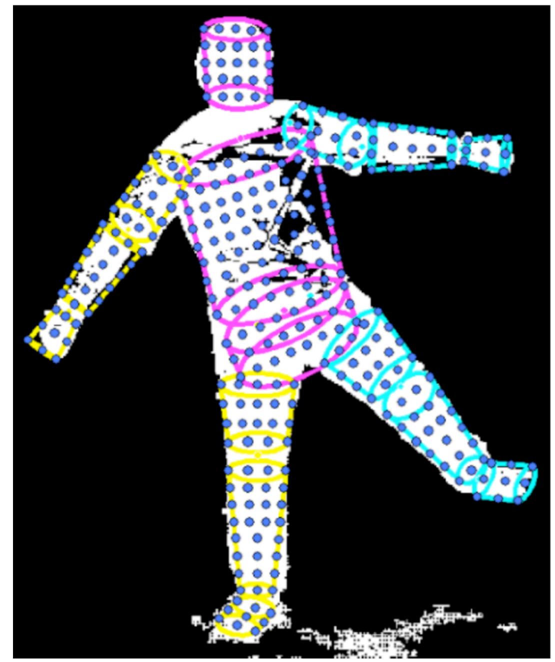
Cylindrical body model, likelihoods



Body model



Edge-based



Silhouette-based

(Images reproduced from Sigal, Balan, and Black, 2009)

Results: generalization over subject

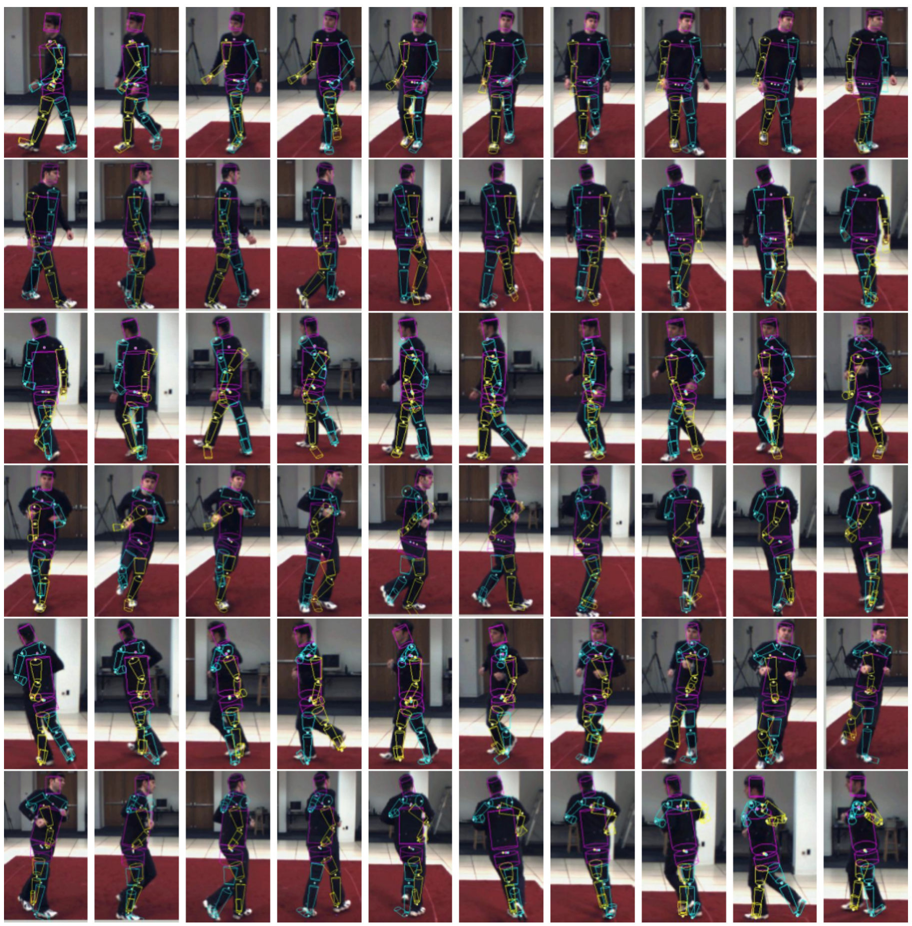
Training set	Baseline	Xu and Li (2007)	CRBM
S5	91.37±6.29	48.98	41.97±3.57
S1		51.66	38.45±0.80
S1S2S3		55.30	48.03±0.29

Training:Test	Baseline	Xu and Li (2007)	CRBM
S1S2S3:S1	129.18±19.47	140.35	55.43±0.79
S1:S1		-	48.75±3.72
S1S2S3:S2	162.75±15.36	149.37	99.13±22.98
S2:S2		-	47.43±2.86
S1S2S3:S3	180.11±24.02	156.30	70.89±2.10
S3:S3		-	49.81±2.19

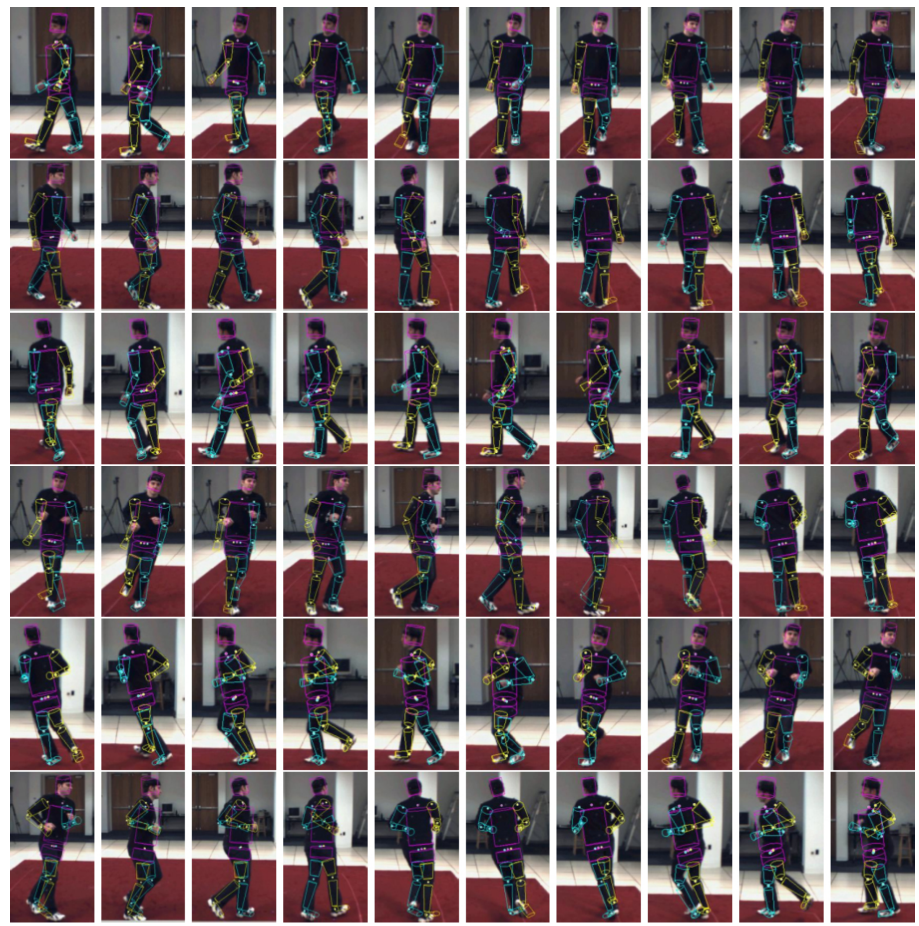
(Mean pose error in mm)

Results: multi-view tracking

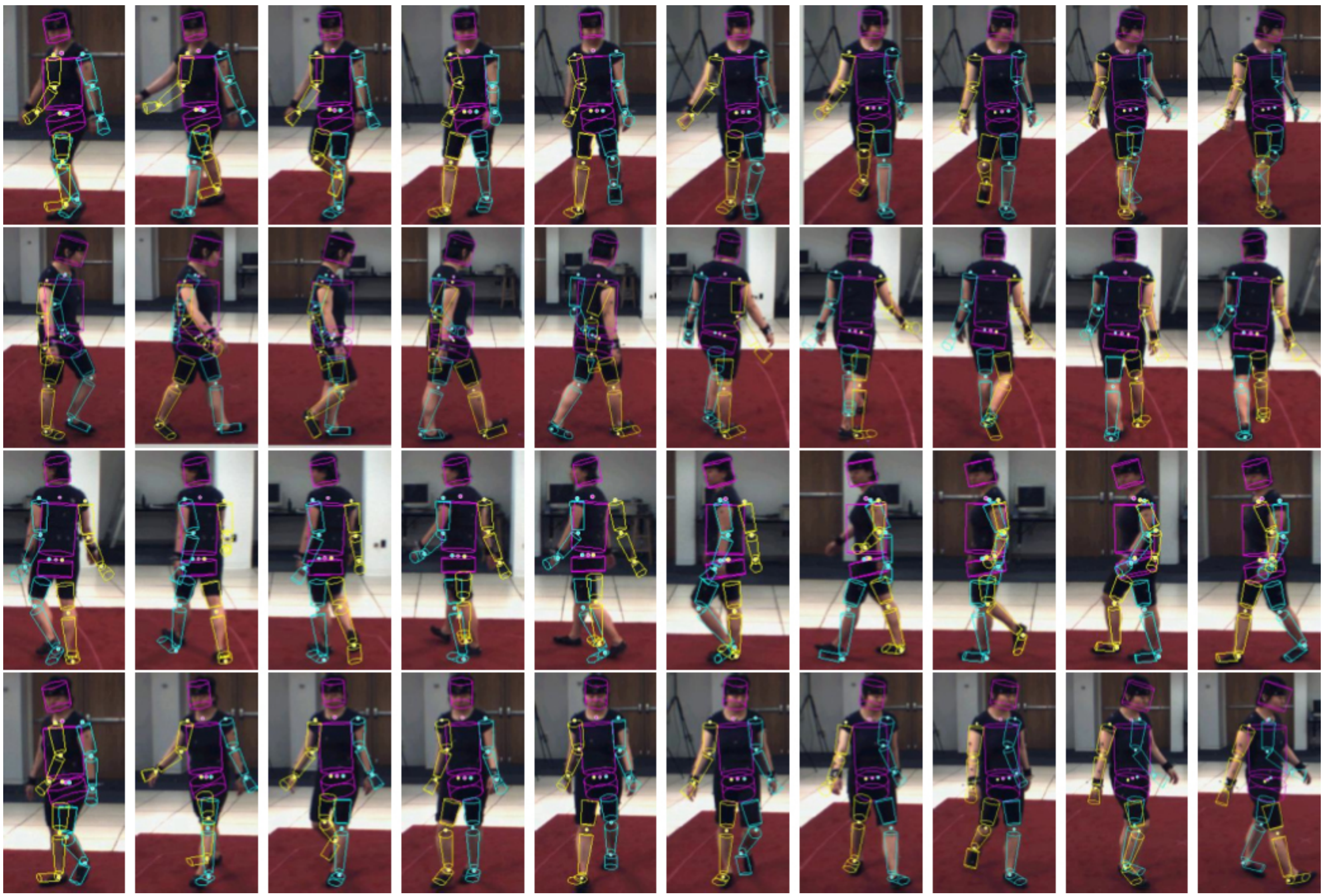
Baseline



CRBM prior



Results: monocular tracking



Conclusions

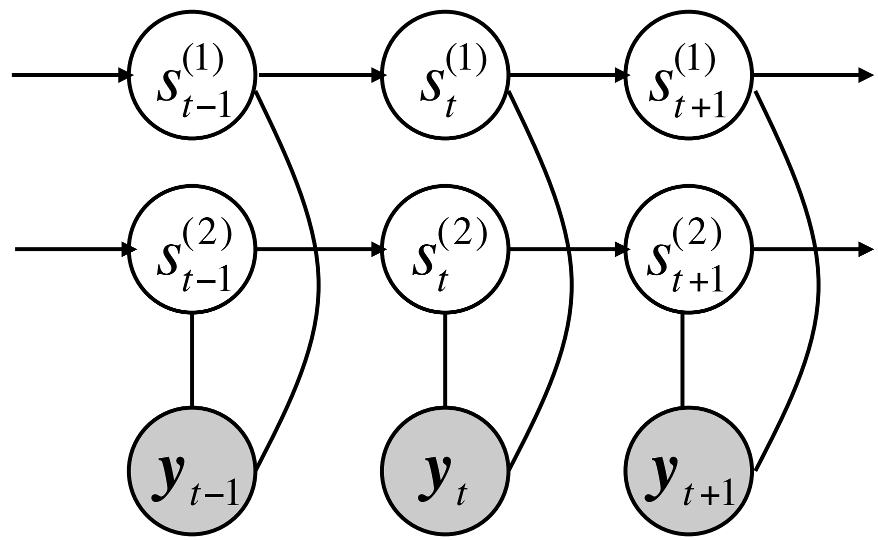
- CRBMs: distributed representations, exact inference and efficient approximate learning
  - Measuring likelihood, improving CD
- FCRBMs: multiplicative interactions with quadratic number of parameters
  - Unsupervised style discovery, deep models
- Tracking: prior aids in challenging settings
  - Augmenting models with identity or action
- PoHMMs: It takes N>1 to Tango

(Brown & Hinton 2001)

Products of Hidden Markov Models

$$p(\boldsymbol{y}_{1:T};\Theta) = \frac{\prod_{m=1}^{M} p^{(m)}(\boldsymbol{y}_{1:T};\theta^{(m)})}{\sum_{\boldsymbol{y}'_{1:T}} \prod_{m=1}^{M} p^{(m)}(\boldsymbol{y}'_{1:T};\theta^{(m)})}$$

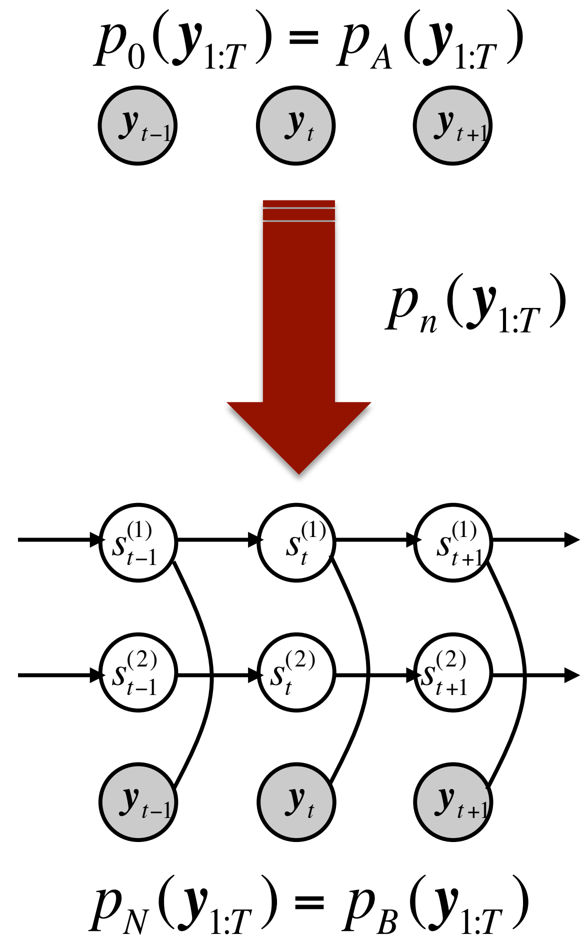
$$= \frac{\prod_{m=1}^{M} p^{(m)}(\boldsymbol{y}_{1:T};\theta^{(m)})}{Z(\Theta,T)}$$



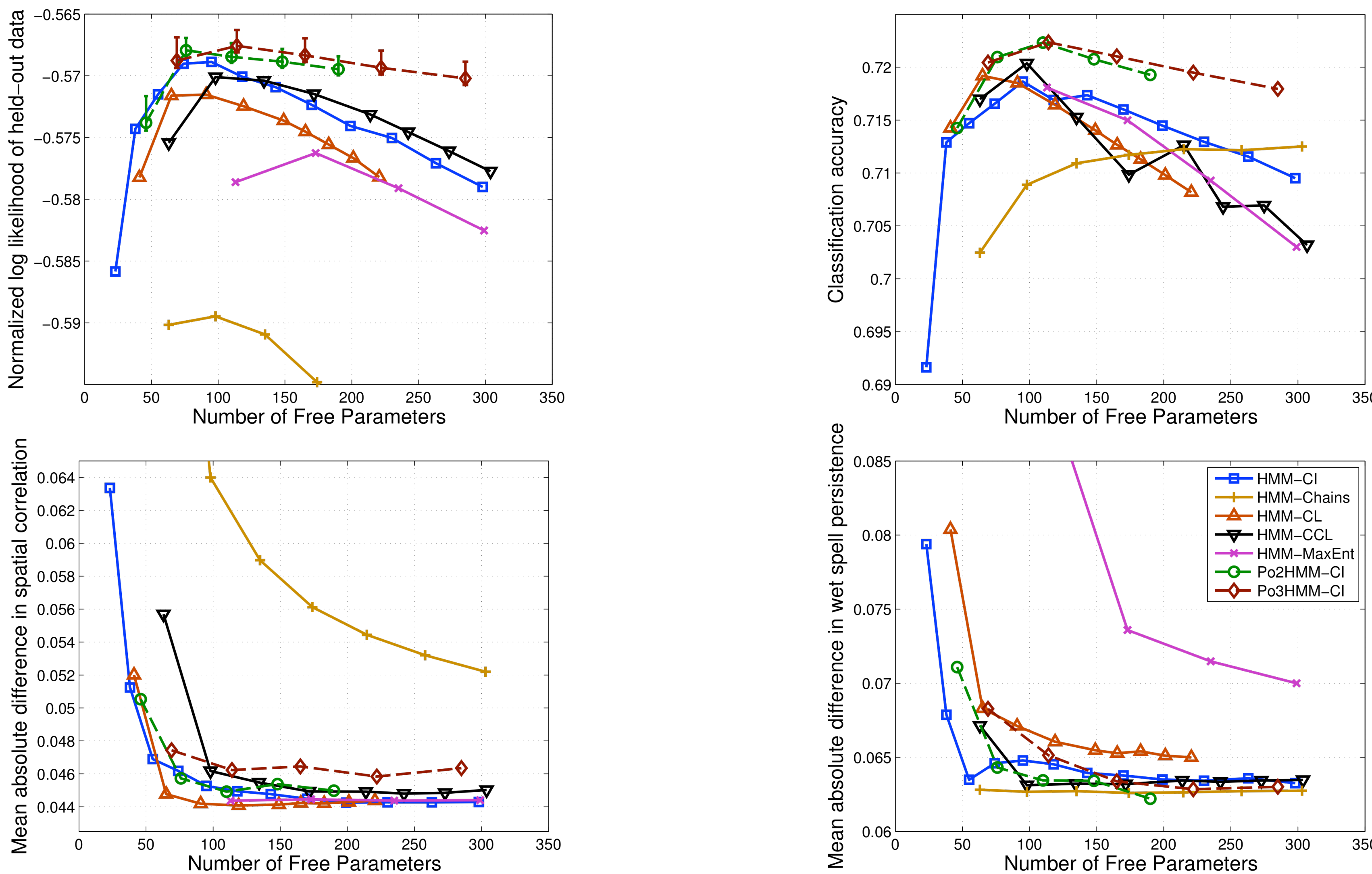
Partition function

Annealed Importance Sampling

- AIS works well for RBMs (Salakhutdinov & Murray 2008)
- PoHMMs and RBMs are both Products of Experts!
- MCMC approach to obtaining an unbiased estimate of the ratio of partition functions Z_B/Z_A
- Define intermediate distribution and transition operator that leaves this distribution invariant

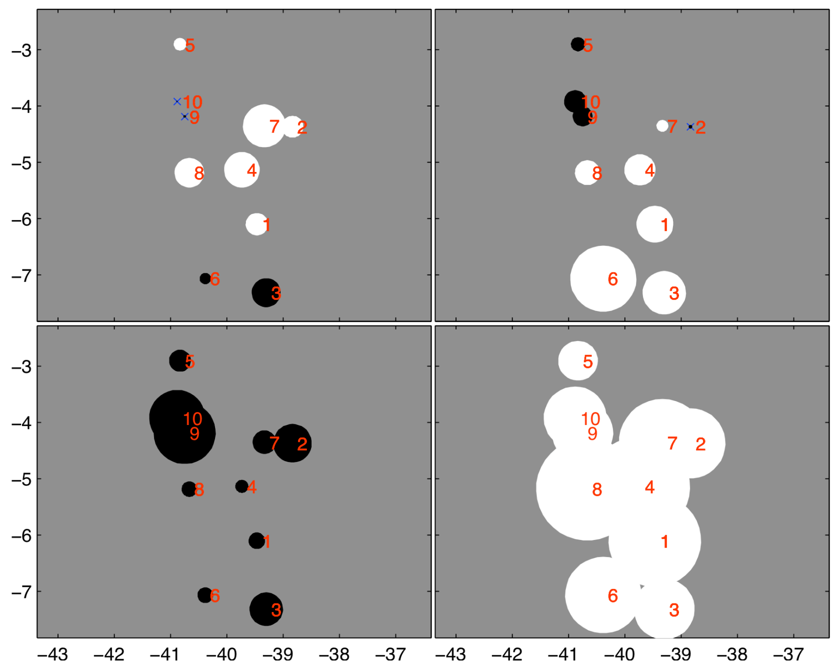


Results: rainfall occurrence data

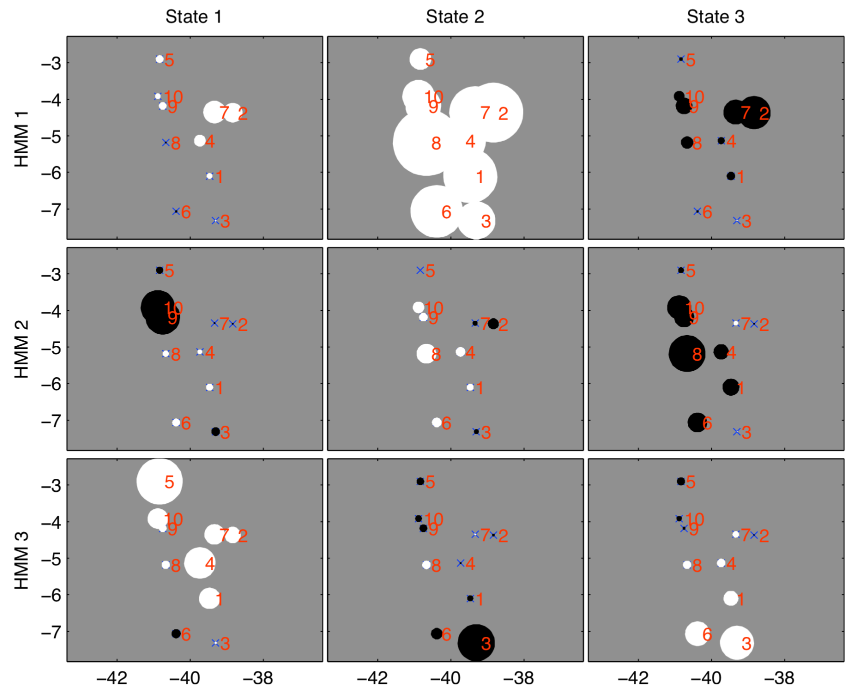


Results: rainfall occurrence data

4-state HMM

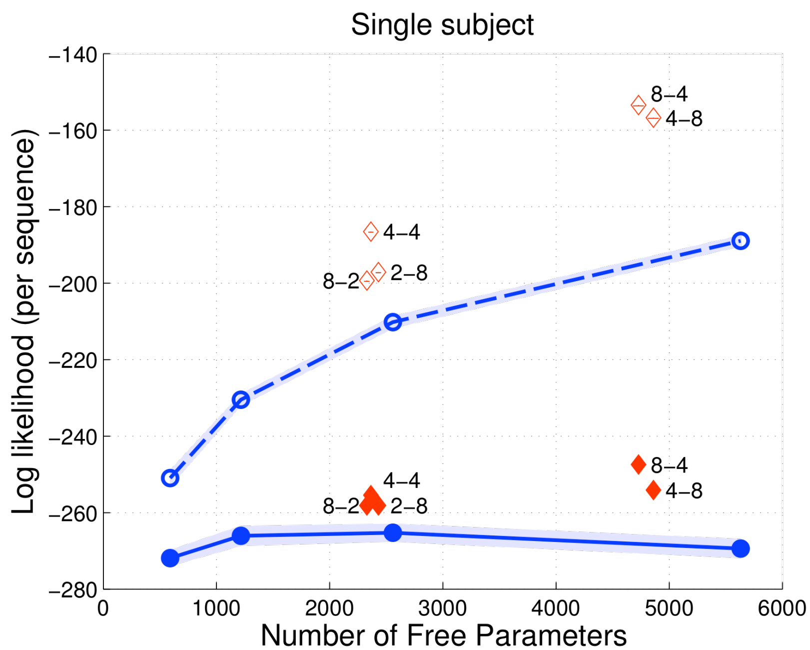


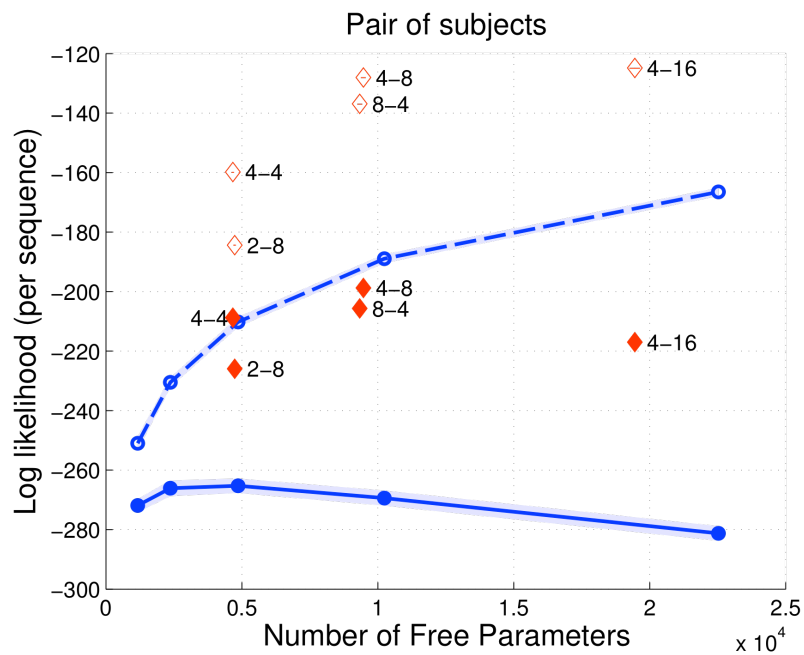
Product of 3 3-state HMMs



HMMs specialize locally

Results: weakly and strongly componential motion





Results: forecasting