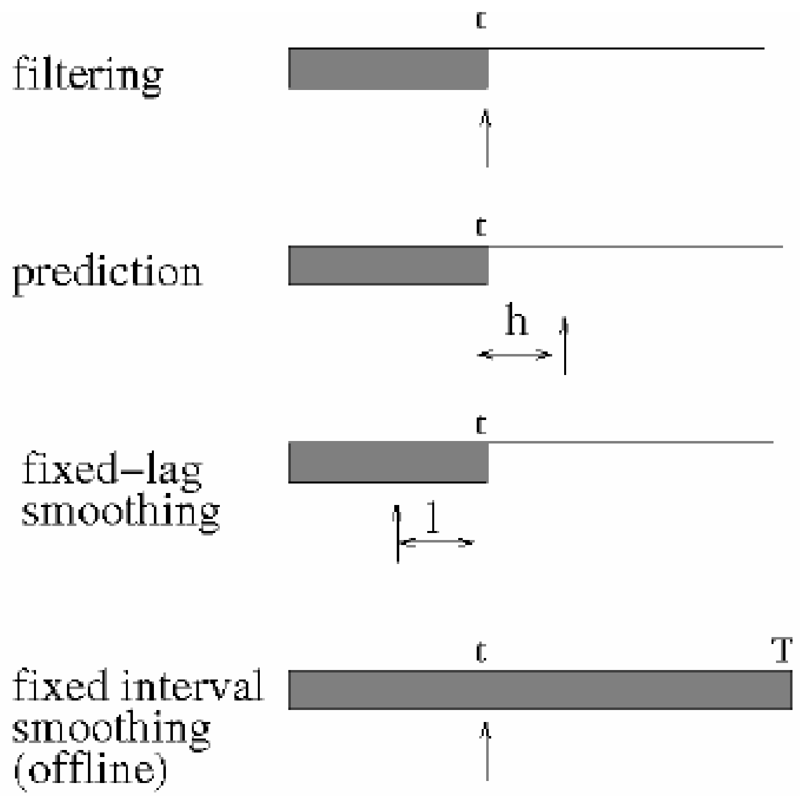
Stat 521A
Lecture 19

Outline

- Inference goals (15.1)
- Exact inference in DBNs (15.2)
- Factored belief sates (15.3.2)
- Particle filtering (15.3.3)
- Switching LDS (15.4.2)

Inference goals

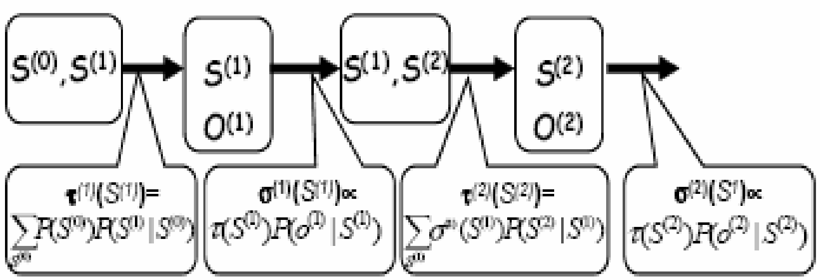


Exact filtering in HMMs

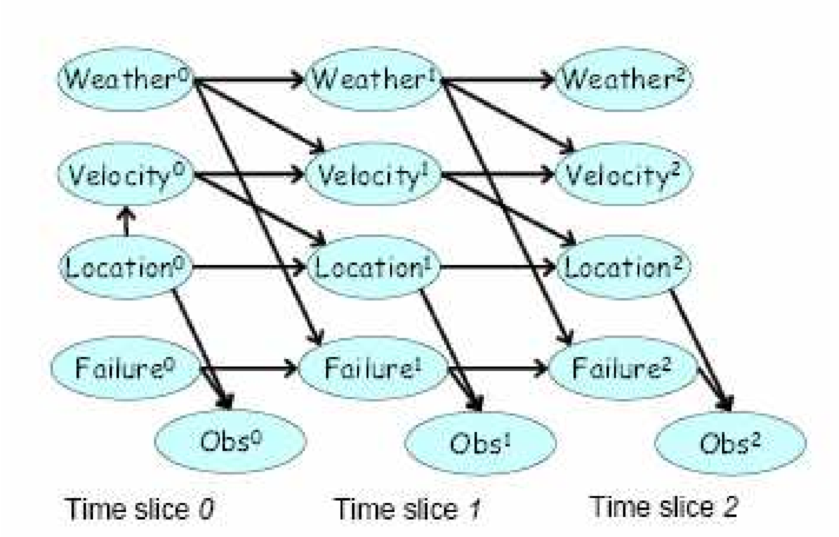
- We can apply the predict-update equations to any dynamical model

$$
\begin{aligned}
\sigma^{(\cdot t+1)}(X^{(t+1)}) &\triangleq P(X^{(t+1)} \mid o^{(1:t)}) \\
&= \sum_{X^{(t)}} P(X^{(t+1)} \mid X^{(t)}, o^{(1:t)}) P(X^{(t)} \mid o^{(1:t)}) \\
&= \sum_{X^{(t)}} P(X^{(t+1)} \mid X^{(t)}) \sigma^{(t)}(X^{(t)}).
\end{aligned}
$$

$$
\begin{aligned}
\sigma^{(t+1)}(X^{(t+1)}) &= P(X^{(t+1)} \mid o^{(1:t)}, o^{(t+1)}) \\
&= \frac{P(o^{(t+1)} \mid X^{(t+1)} \mid o^{(1:t)}) P(X^{(t+1)} \mid o^{(1:t)})}{P(o^{(t+1)} \mid o^{(1:t)})} \\
&= \frac{P(o^{(t+1)} \mid X^{(t+1)}) \sigma^{(\cdot t+1)}(X^{(t+1)})}{P(o^{(t+1)} \mid o^{(1:t)})}.
\end{aligned}
$$



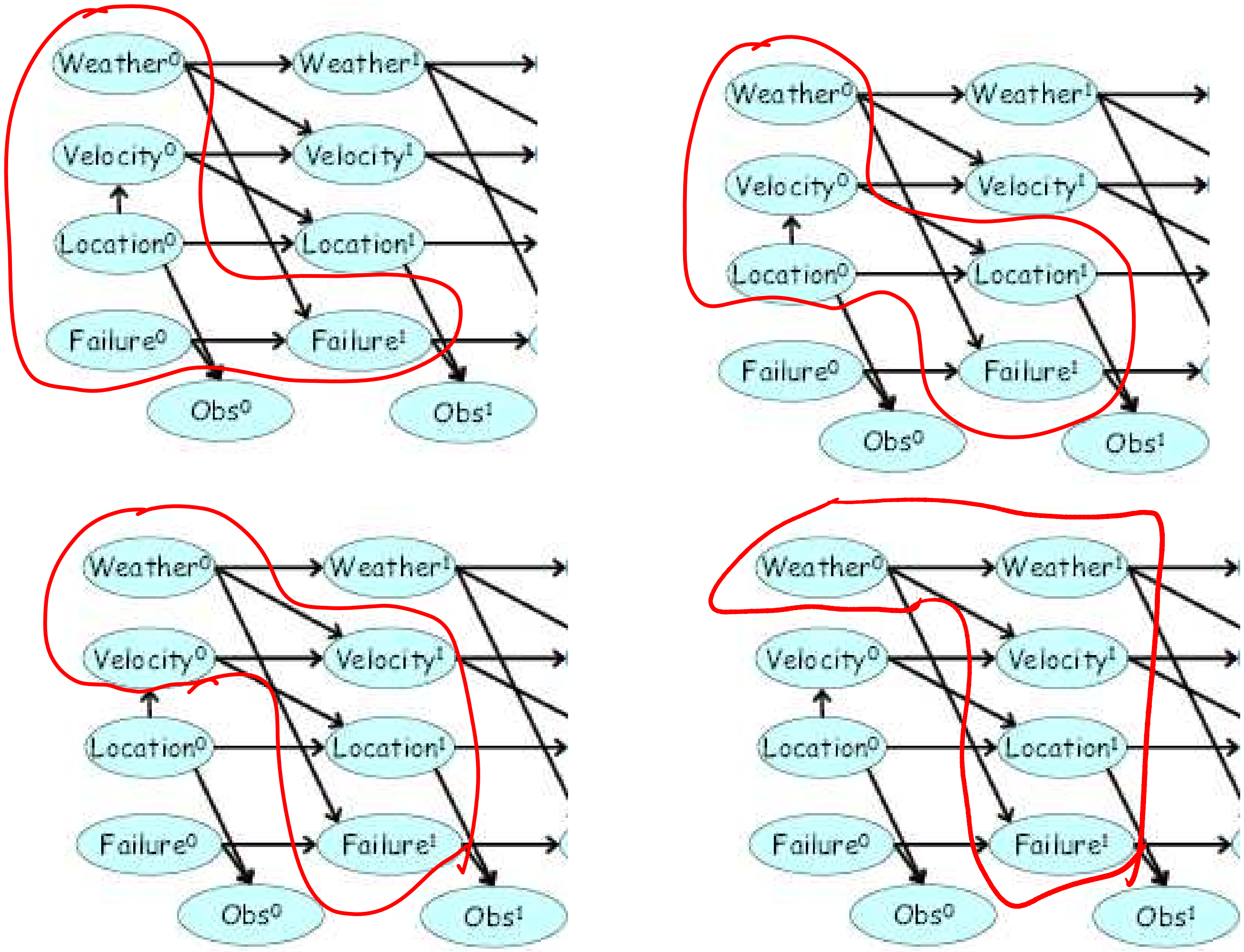
Entanglement



In the unrolled network, all the persistent nodes become correlated. Hence the belief state does not admit any factorization.

Frontier algorithm

- We need cliques that can store the interface variables



Factored frontier algorithm

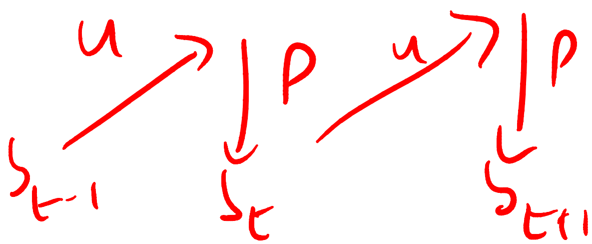
- Represent incoming belief state as a product of marginals

$$\hat{\sigma}^{(t)}(\mathcal{X}^{(t)}) = \prod_r (\beta_r^{(t)}[\mathbf{X}_r^{(t)}])^{\mu_r}.$$

- Perform calibration in the 2-slice jtree
- Compute posterior marginals (M projection onto factored distribution)
- Can also use conditionally factored belief states

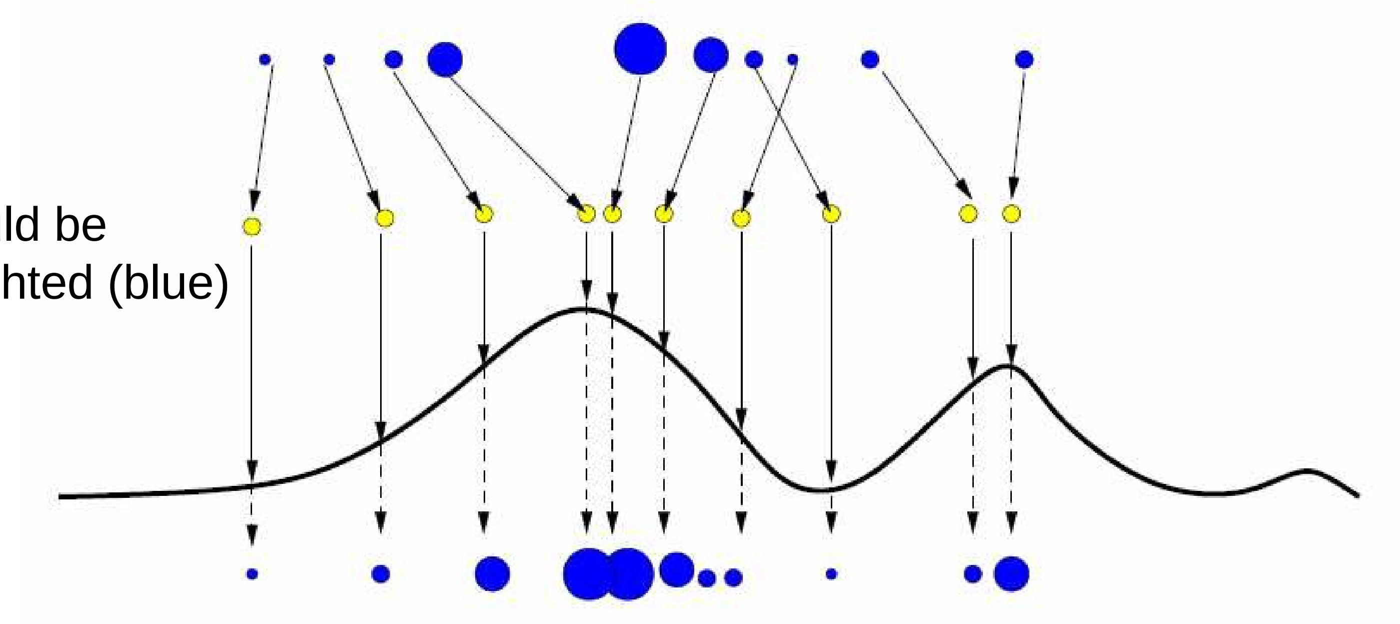
$$\left(\beta_g^{(t)}[\mathbf{Z}^{(t)}]\right)^{-(k-1)} \prod_{i=1}^{k} \beta_i^{(t)}[\mathbf{Z}^{(t)}, \mathbf{Y}_i^{(t)}],$$

- This is like EP without the backwards pass, aka ADF



Importance sampling

Should be
Weighted (blue)



Propose

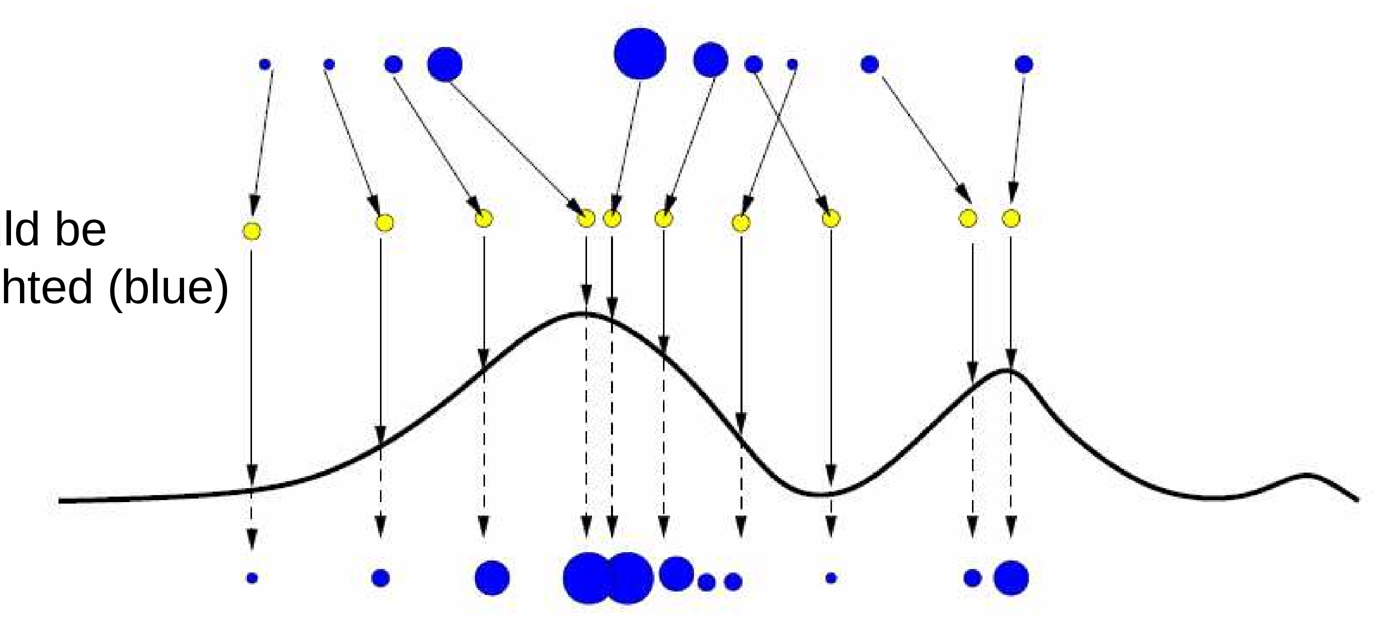
Weight

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \approx \frac{1}{N}\sum_{i=1}^{N}\tilde{w}_t^{(i)} I(\mathbf{x}_{0:t} = \mathbf{x}_{0:t}^{(i)}) \quad (1)$$

$$\tilde{w}_t^{(i)} \stackrel{\text{def}}{=} \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}} \quad (2)$$

$$w_t^{(i)} \stackrel{\text{def}}{=} \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_{0:t}^{(i)})} \quad (3)$$

Sequential Importance Sampling



Markov proposal

$$\pi(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \pi(\mathbf{x}_0)\prod_{k=1}^{t}\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1},\mathbf{y}_{1:t})$$

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)}\frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{1:t})}$$

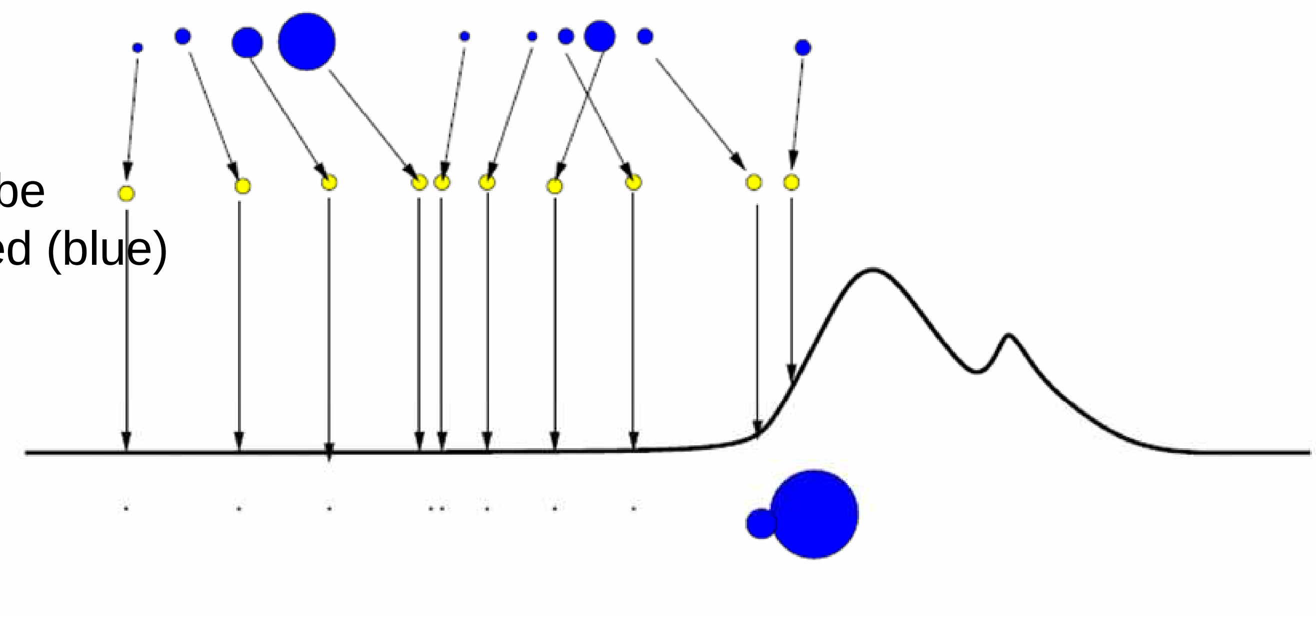
Propose from dynamical prior

$$\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)},\mathbf{y}_{1:t}) = p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})$$

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)}p(\mathbf{y}_t|\mathbf{x}_t^{(i)})$$

Problem with SIS

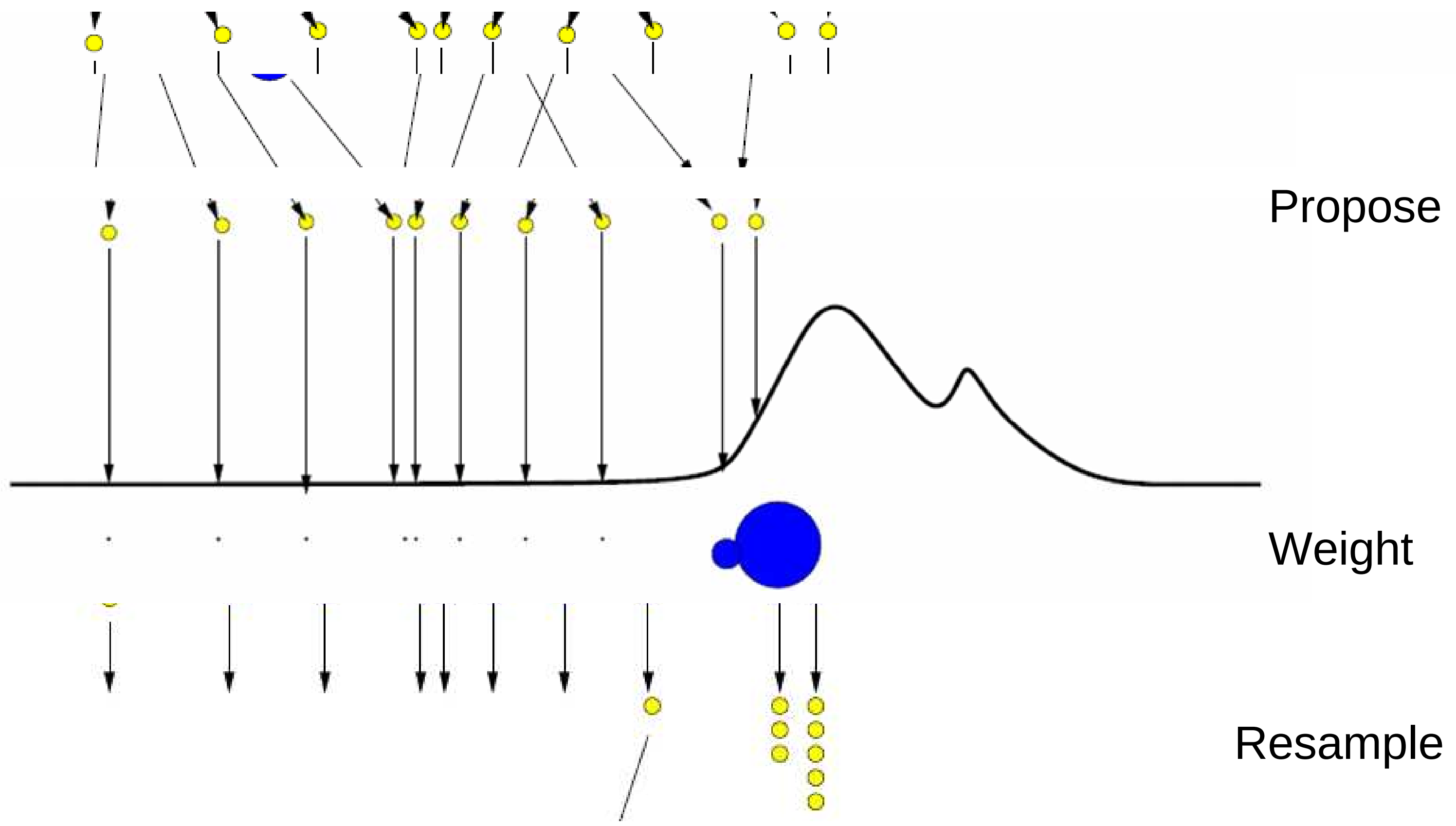
Should be
Weighted (blue)



Unlikely evidence “kills off” most particles
(Particle impoverishment) resulting in high variance estimate

SIR/ PF/ SOF/ SMC

Should be in diff locns



PF

1. Sequential importance sampling step

- For $i = 1, \ldots, N$, sample

$$\left(\widehat{\mathbf{x}}_t^{(i)}\right) \sim q(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t})$$

and set

$$\left(\widehat{\mathbf{x}}_{1:t}^{(i)}\right) \triangleq (\widehat{\mathbf{x}}_t^{(i)}, \mathbf{x}_{1:t-1}^{(i)})$$

- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalising constant:

$$w_t^{(i)} = \frac{p\left(y_t | \mathbf{x}_t^{(i)}\right) p\left(\widehat{\mathbf{x}}_t^{(i)} \middle| \widehat{\mathbf{x}}_{t-1}^{(i)}\right)}{q\left(\widehat{\mathbf{x}}_t | \widehat{\mathbf{x}}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right)}$$

- For $i = 1, \ldots, N$, normalise the importance weights:

$$\widetilde{w}_t^{(i)} = w_t^{(i)} \left[\sum_{j=1}^{N} w_t^{(j)}\right]^{-1}$$

2. Selection step

- Resample the discrete weighted measure $\{(\widehat{\mathbf{x}}_{1:t}^{(i)}, \tilde{w}_t^{(i)})\}_{i=1}^{N}$ to get an unweighted measure $\{(\mathbf{x}_{1:t}^{(i)}, \frac{1}{N})\}_{i=1}^{N}$

Example from Nando de Freitas

$$x_t = \frac{1}{2}x_{t-1} + 25\frac{x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t$$

$$y_t = \frac{x_t^2}{20} + w_t$$

where $x_0 \sim \mathcal{N}(0, \sigma_1^2)$, v_t and w_t are mutually independent white Gaussian noises, $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $w_t \sim \mathcal{N}(0, \sigma_w^2)$

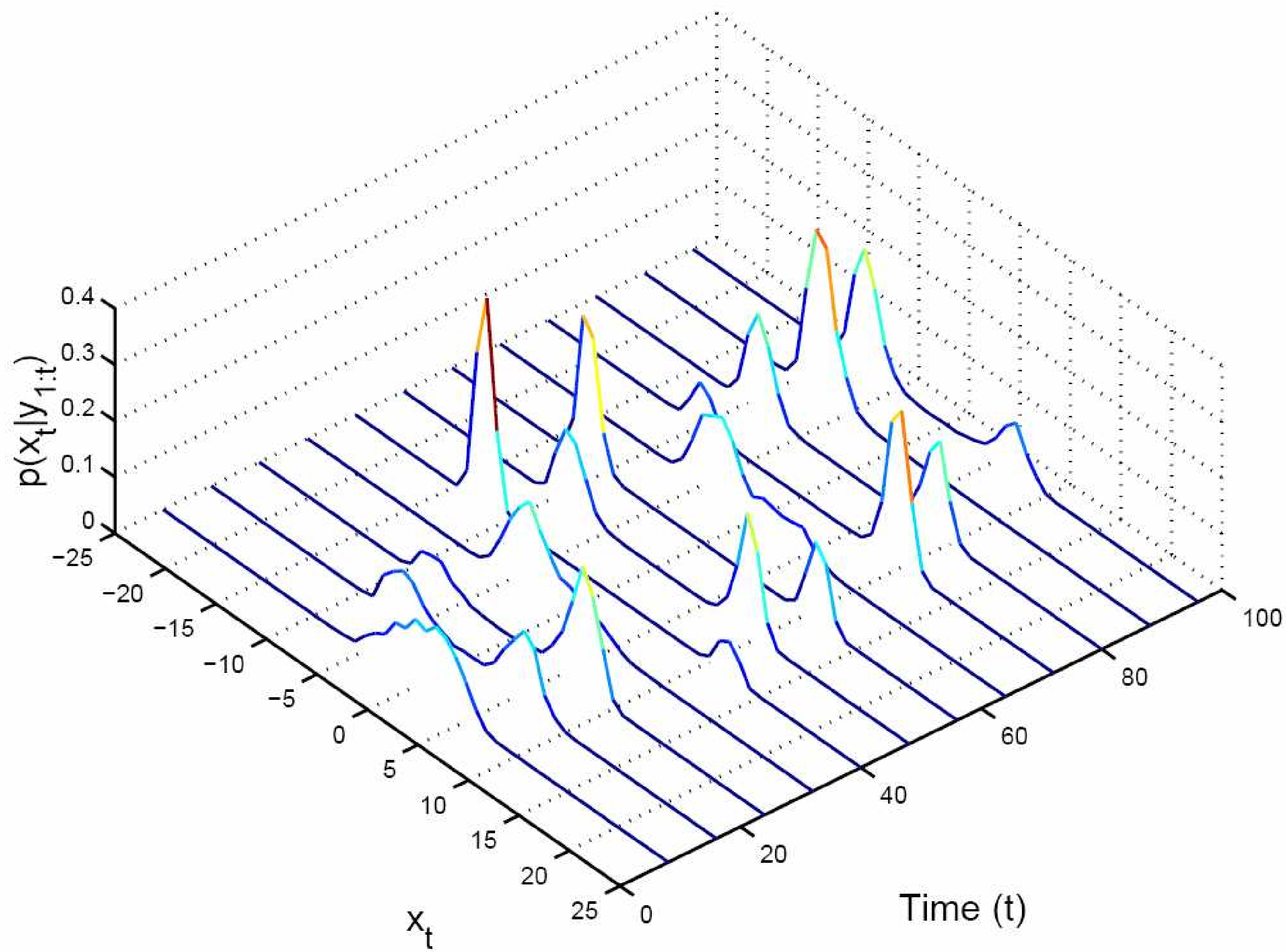
- For $i = 1, ..., N$, sample $\mathbf{x}_0^{(i)} \sim \mathcal{N}\left(0, \sigma_1^2\right)$
- For $i = 1, ..., N$, sample

$$x_t^{(i)} = \frac{1}{2} x_{t-1}^{(i)} + 25 \frac{x_{t-1}^{(i)}}{1 + x_{t-1}^{2(i)}} + 8 \cos(1.2t) + \mathcal{N}\left(0, \sigma_v^2\right)$$

- For $i = 1, ..., N$, evaluate the importance weights

$$\widetilde{w}_t^{(i)} = \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{1}{2\sigma_w^2}\left(y - \frac{x_t^{2(i)}}{20}\right)^2}$$

- Normalise the importance weights.
- Resample fittest samples (black-box).



0.4
0.3
0.2
0.1
0
p(x_t|y_1:t)
−25
−20
−15
−10
−5
0
5
10
15
20
25
x_t
0
20
40
60
80
100
Time (t)

PF for DBNs

PF

for $m = 1, \ldots, M$
  Sample $\bar{x}^{(0)}[m]$ from $\mathcal{B}_0$
  $w^{(0)}[m] \leftarrow 1/M$
for $t = 1, 2, \ldots$
  for $m = 1, \ldots, M$
    Sample $\bar{x}^{(0:t-1)}$ from the distribution $\hat{P}_{\mathcal{D}^{(t-1)}}$.
      // Select sample for propagation
    $(\bar{x}^{(0:t)}[m], w^{(t)}[m]) \leftarrow$ LW-2TBN$(\mathcal{B}_{\rightarrow}, \bar{x}^{(0:t-1)}, \boldsymbol{o}^{(t)})$
      // Generate time t sample and weight from selected sample $\bar{x}^{(t-1)}$
    $\mathcal{D}^{(t)} \leftarrow \{(\bar{x}^{(0:t)}[m], w^{(t)}[m]) \ : m = 1, \ldots, M\}$
    $\hat{\sigma}^{(t)}(x) \leftarrow \hat{P}_{\mathcal{D}^{(t)}}$

LW-2TBN

Let $X'_1, \ldots, X'_n$ be a topological ordering of $\mathcal{X}'$ in $\mathcal{B}_{\rightarrow}$
$w \leftarrow 1$
for $i = 1, \ldots, n$
  $\boldsymbol{u}_i \leftarrow (\xi, \boldsymbol{x}')\langle \mathrm{Pa}_{X'_i} \rangle$
    // Assignment to $\mathrm{Pa}_{X'_i}$ in $x_1, \ldots, x_n, x'_1, \ldots, x'_{i-1}$
  if $X'_i \notin O^{(t)}$ **then**
    Sample x'_i from $P(X'_i \mid \boldsymbol{u}_i)$
  else
    $x'_i \leftarrow \boldsymbol{o}^{(t)}\langle X'_i \rangle$ // Assignment to X'_i in $\boldsymbol{o}^{(t)}$
    $w \leftarrow w \cdot P(x'_i \mid \boldsymbol{u}_i)$ // Multiply weight by probability of desired value
return $(x'_1, \ldots, x'_n), w$

Condensation algorithm

Isard & Blake (ICCV98)

Monte Carlo Localization

Fox, Burgard, Dellaert, Thrun, AAAI'99

Optimal proposal distribution

- Optimal proposal is the posterior

$$\pi(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}) \quad = \quad p(\mathbf{x}_t|\mathbf{y}_t,\mathbf{x}_{t-1})$$

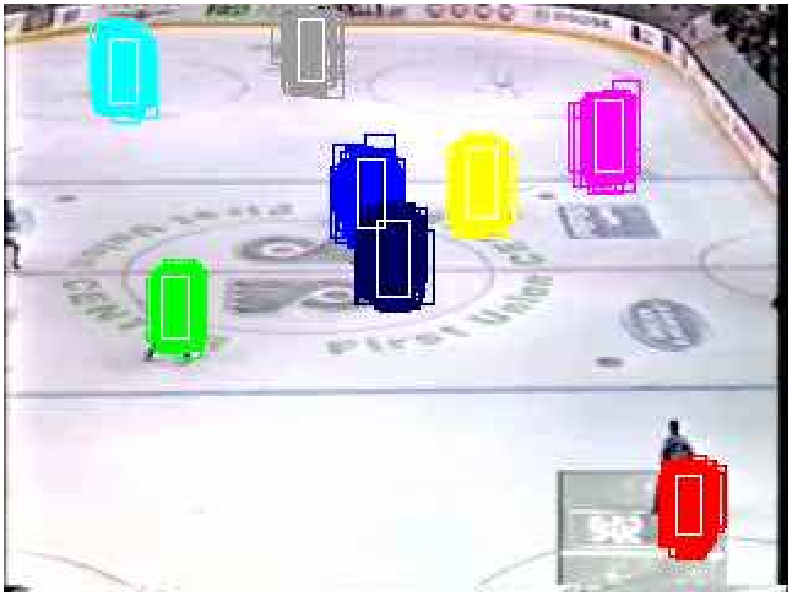
- Incremental weights are one-step-ahead predictive density

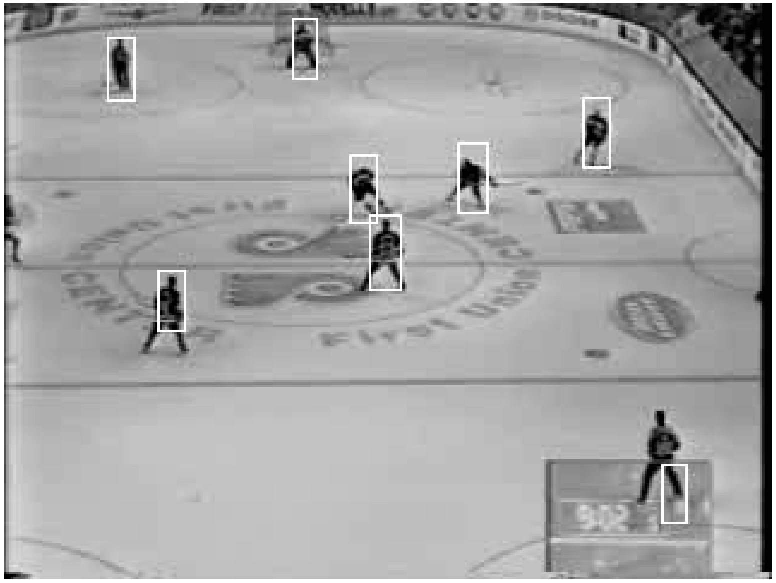
$$\begin{aligned}
\tilde{w}_t^{(i)} &\propto \tilde{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{p(\mathbf{x}_t^{(i)}|\mathbf{y}_t,\mathbf{x}_{t-1}^{(i)})} \\
&= \tilde{w}_{t-1}^{(i)} p(\mathbf{y}_t|\mathbf{x}_{t-1}^{(i)}) \\
p(\mathbf{y}_t|\mathbf{x}_{t-1}^{(i)}) &= \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})d\mathbf{x}_t
\end{aligned}$$

- Can approximate this using EKF, UKF, etc.

Boosted particle filter

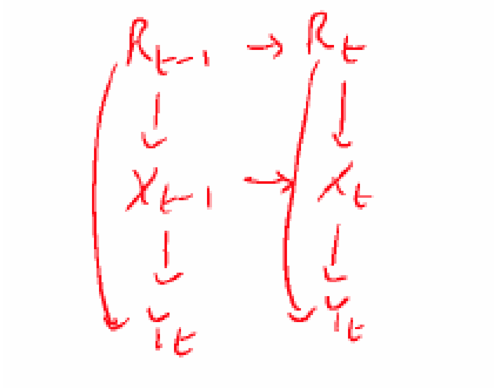
- Run a classifier, trained using boosting, to detect people, and use this as a proposal
- Okuma, Taleghani, de Freitas, Little, Lowe, ECCV04





RBPF

- Rao-Blackwellisation: integrate out X, sample R



- Distributional particles

$$\alpha^{i}_{t-1|t-1}(\mathrm{x}_{t-1}) \stackrel{\text{def}}{=} p(\mathrm{x}_{t-1}|\mathrm{y}_{1:t-1}, r^{(i)}_{1:t-1})$$

RBPF high level

Generic RBPF

1. Sequential importance sampling step

   - For $i = 1, \ldots, N$, sample

   $$\left(\widehat{r}_t^{(i)}\right) \sim q(r_t; r_{1:t-1}^{(i)}, y_{1:t})$$

   and set

   $$\left(\widehat{r}_{1:t}^{(i)}\right) \triangleq \left(\widehat{r}_t^{(i)}, r_{1:t-1}^{(i)}\right)$$

   - For $i = 1, \ldots, N$, evaluate the importance weights up to a normalising constant:

   $$w_t^{(i)} = \frac{p\left(y_t \middle| \mathbf{y}_{1:t-1}, \widehat{r}_{1:t}^{(i)}\right) p\left(\widehat{r}_t^{(i)} \middle| \widehat{r}_{1:t-1}^{(i)}, \mathbf{y}_{1:t-1}\right)}{q\left(\widehat{r}_t; \widehat{r}_{1:t-1}^{(i)}, y_{1:t}\right)}$$

   - For $i = 1, \ldots, N$, normalise the importance weights:

   $$\widetilde{w}_t^{(i)} = w_t^{(i)} \left[\sum_{j=1}^{N} w_t^{(j)}\right]^{-1}$$

2. Selection step

   - Resample the discrete weighted measure $\{(\widehat{r}_{1:t}^{(i)}, \widetilde{w}_t^{(i)})\}_{i=1}^N$ to get an unweighted measure $\{(r_{1:t}^{(i)}, \frac{1}{N})\}_{i=1}^N$

3. Exact step

   - Update $p(\mathbf{X}_t | \mathbf{y}_{1:t}, r_{1:t}^{(i)})$ given $p(\mathbf{X}_{t-1} | \mathbf{y}_{1:t-1}, r_{1:t-1}^{(i)})$, $r_t^{(i)}$, and y_t.

RBPF updates

Then, for each possible value of r_t, we perform a predict-update cycle:

$$\alpha^i_{t|t-1}(\mathbf{x}_t, r_t) \stackrel{\text{def}}{=} p(\mathbf{x}_t, r_t | \mathbf{y}_{1:t-1}, r^{(i)}_{1:t-1}, r_t) = \int p(\mathbf{x}_t | r_t, \mathbf{x}_{t-1}) p(r_t | r^{(i)}_{1:t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}, r_t) d\mathbf{x}_{t-1} \quad (21.28)$$

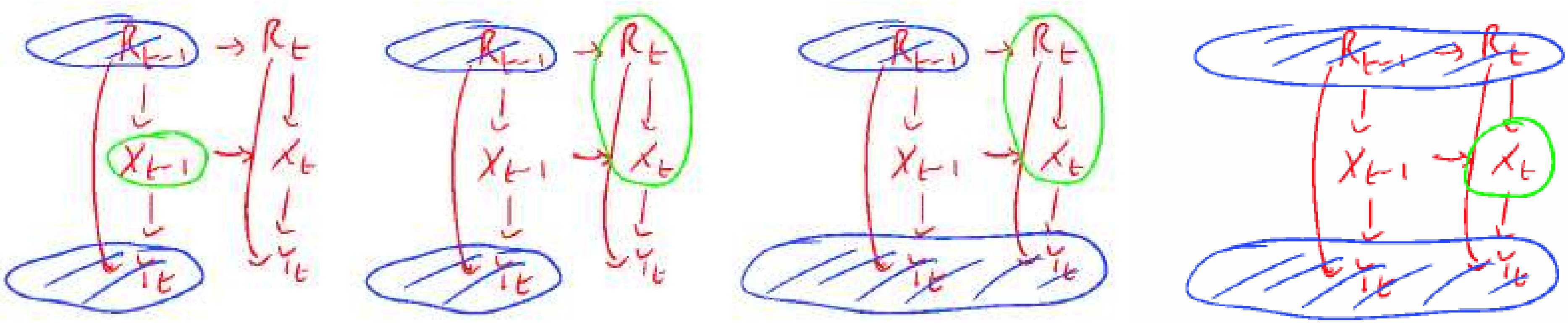
$$\alpha^i_{t|t}(\mathbf{x}_t, r_t) \stackrel{\text{def}}{=} p(\mathbf{x}_t, r_t | \mathbf{y}_{1:t-1}, \mathbf{y}_t, r^{(i)}_{1:t-1}, r_t) = \frac{p(\mathbf{y}_t | \mathbf{x}_t, r_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1}, r^{(i)}_{1:t-1}, r_t)}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, r^{(i)}_{1:t-1}, r_t)} \quad (21.29)$$

where

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, r^{(i)}_{1:t-1}, r_t) = \int p(\mathbf{y}_t | \mathbf{x}_t, r_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1}, r^{(i)}_{1:t-1}, r_t) d\mathbf{x}_t \quad (21.30)$$

Finally, once we have chosen $r^{(i)}_t$, we pick the corresponding updated distribution:

$$\alpha^i_{t|t}(\mathbf{x}_t) = \alpha^i_{t|t}(\mathbf{x}_t, r^{(i)}_t) \quad (21.31)$$



(a) (b) (c)

RBPF for Switching LDS

Algorithm 7: One step of the Mixture Kalman filter algorithm

for $i = 1 : N$ **do**
- $r_{t,i} \sim p(r | r_{t-1,i})$
- $(\boldsymbol{\mu}_{t,i}, \boldsymbol{\Sigma}_{t,i}, w_{t,i}) = \text{KFupdate}(\boldsymbol{\mu}_{t-1,i}, \boldsymbol{\Sigma}_{t-1,i}, \mathbf{y}_t, r_{t,i})$

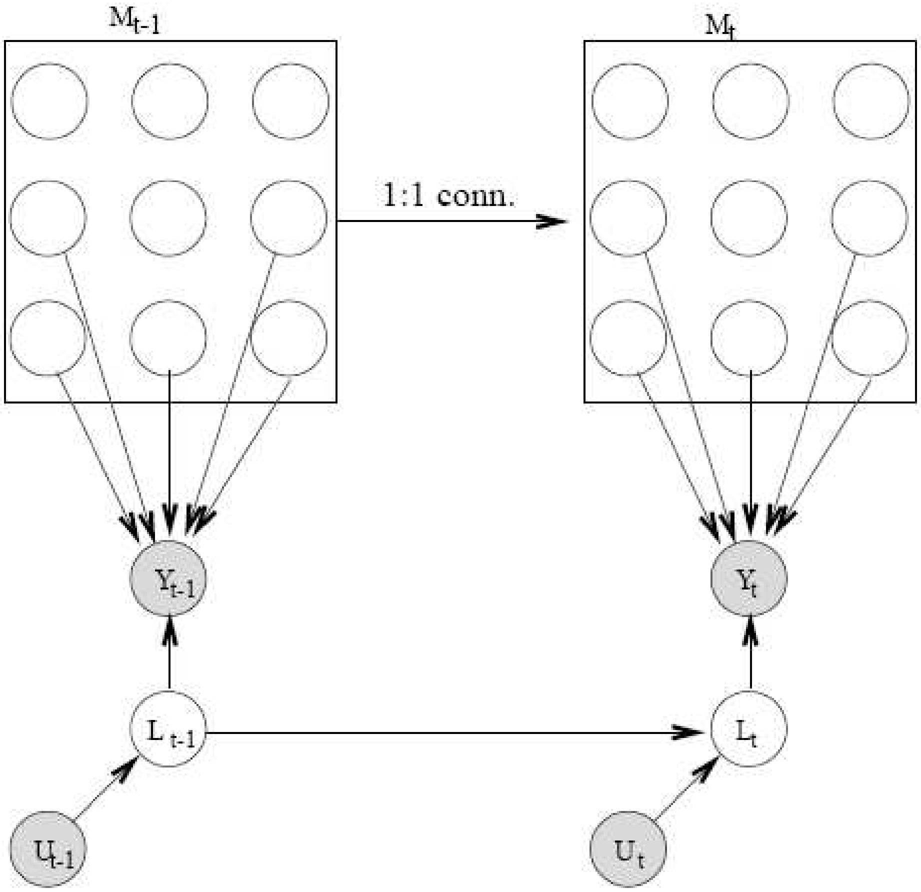
for $i = 1 : N$ **do**
- $\tilde{w}_{t,i} = w_{t,i} \left[\sum_j w_{t,j} \right]^{-1}$

π = resample$(\tilde{w}_{t,1:N})$

return $(r_{t,\pi}, \boldsymbol{\mu}_{t,\pi}, \boldsymbol{\Sigma}_{t,\pi})$

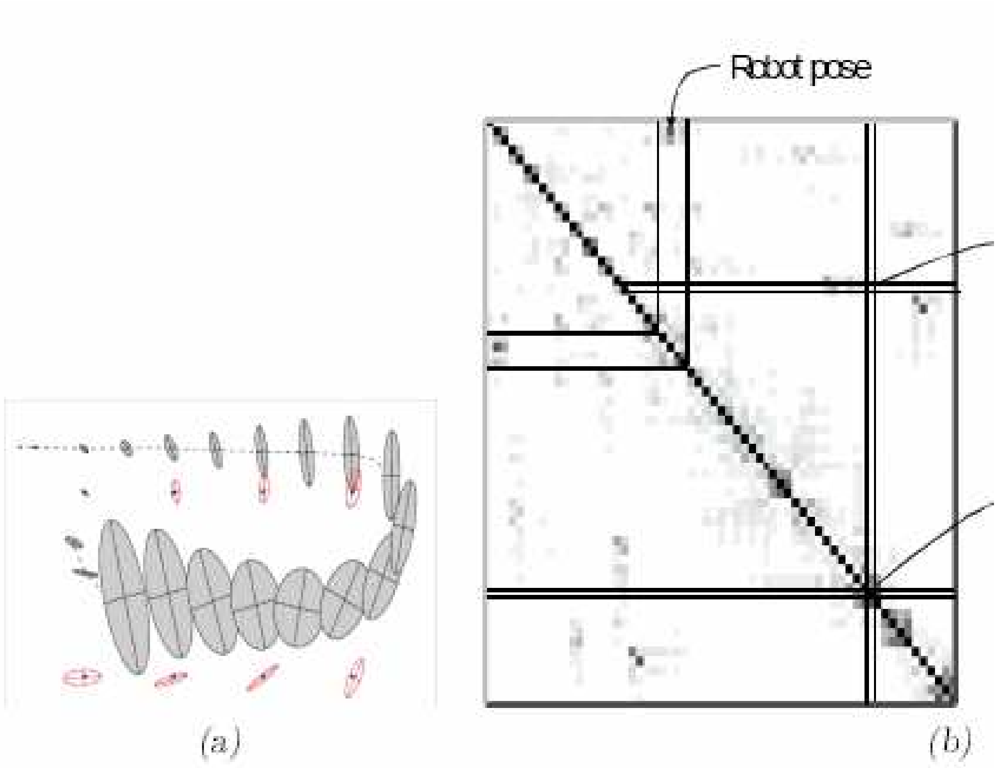
RBPF for SLAM

- Simultaneous Localization and Mapping
- Occupancy grid version (Murphy, NIPS'00)



FastSLAM

- Kalman filter version: replace covmat of size $(2K+2)^2$ with $P*K*2^2$ covmats, P=#particles, K=#num landmarks



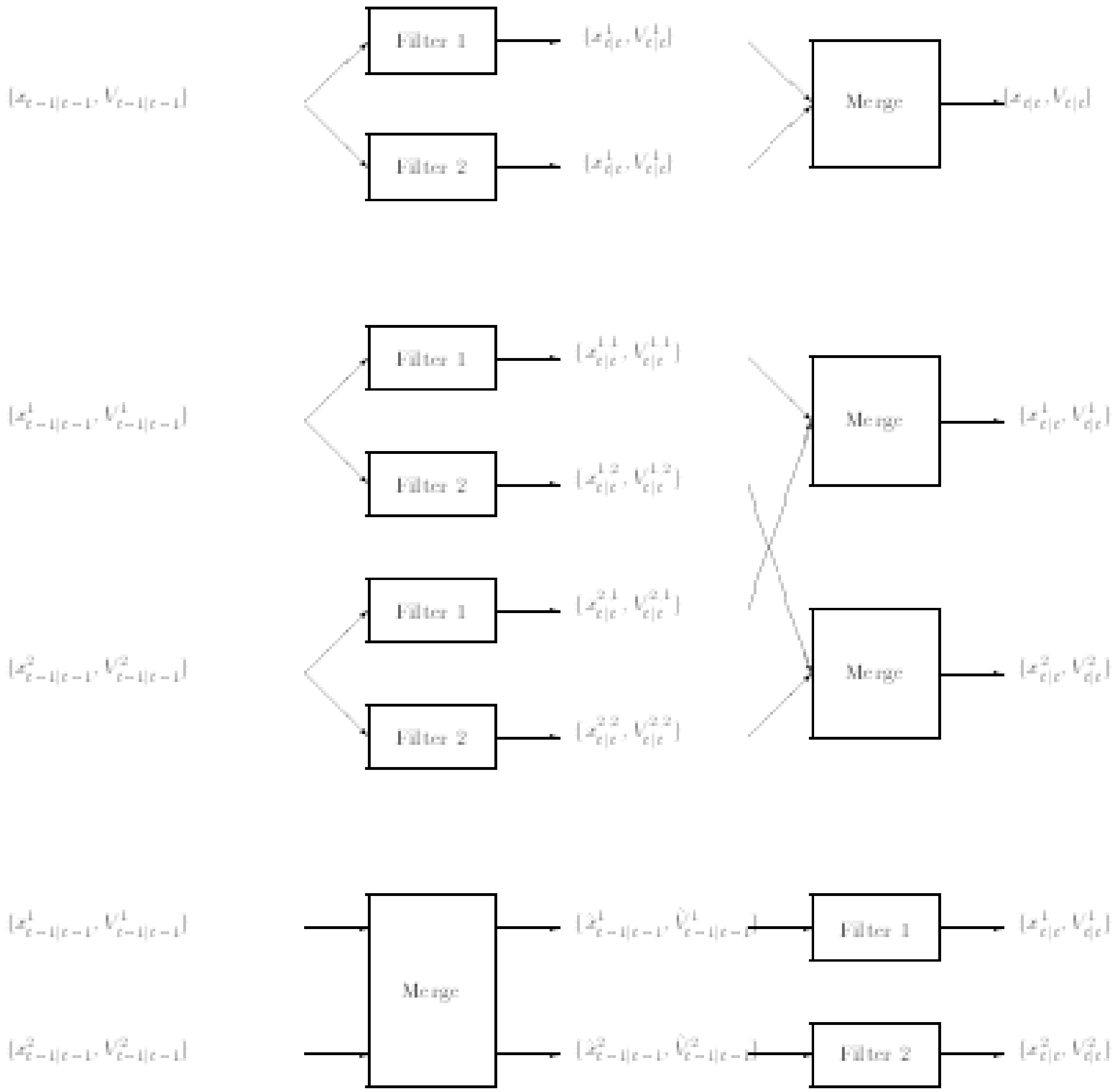
Montemerlo, Thrun, Koller, Wegbreit, AAAI'02

Switching LDS

GBP1

GBP2

IMM



EP approximations