A Few Applications of Particle Filtering to Digital Communications

The good, the bad and the ugly

Elena Punskaya - Arnaud Doucet

William J. Fitzgerald

Engineering Department, Cambridge University, UK

http://www-sigproc.eng.cam.ac.uk/~op205/

{op205,ad2,wjf}@eng.cam.ac.uk

Institut Henri Poincaré - Paris - 3 Décembre 2002.

Outline of the Talk

1. Introduction.
2. Demodulation in Fading Channels.
3. Joint Demodulation and Synchronization in SS Systems.
4. Discussion.

1.1– A Bayesian Approach to Digital Communications

- *Aim*: Recover one/several sequences of finite-valued symbols r_n which have been through (unknown) channels, noise, mixed etc.

- *Parameters of the model*
  - $\{r_n\}_{n\geq 0}$ sequence of unknown symbols.
  - $\{\theta_n\}_{n\geq 0}$ nuisance parameters (channels, variance of noise, delays etc.)
  - $\{y_n\}_{n\geq 0}$ noisy observations.

- *Generic Bayesian model*

  - Prior distribution on $\{r_n\}_{n\geq 0}$ AND nuisance parameters $\{\theta_n\}_{n\geq 0}$.
  - Likelihood function $g\left(y_n \middle| y_{1:n-1}, r_{1:n}, \theta_{1:n} \right)$.

- Given $y_{1:n}$, all inference relies on

$$p\left(r_{1:n},\theta_{1:n}\right|y_{1:n}\left.\right)\propto\underbrace{f\left(r_{1:n}\right)f\left(\theta_{1:n}\right)}_{\text{Prior}}\underbrace{\prod_{k=1}^{n}g\left(y_{k}\right|y_{1:k-1},r_{1:k},\theta_{1:k}\left.\right)}_{\text{Likelihood}}$$

We are particularly interested in

$$p\left(r_{1:n}\right|y_{1:n}\left.\right)=\int p\left(r_{1:n},\theta_{1:n}\right|y_{1:n}\left.\right)d\theta_{1:n},$$

$$p\left(r_{n}\right|y_{1:n}\left.\right)=\sum_{r_{1:n-1}}\int p\left(r_{1:n},\theta_{1:n}\right|y_{1:n}\left.\right)d\theta_{1:n}.$$

- No analytical expression in most applications.
- Even if $p\left(r_{1:n}\right|y_{1:n}\left.\right)$ known, too computationally intensive!

1.3– Bayesian Inference - Batch and Recursive

• *Batch Inference.* MCMC methods popular but remember that SMC apply too! See the MIMO example.

• *Sequential Inference.*

$$p\left(r_{1:n},\theta_{1:n}\right|y_{1:n}) \propto g\left(y_n\right|y_{1:n-1},r_{1:n},\theta_{1:n}) f\left(r_n\right|r_{1:n-1}) f\left(\theta_n\right|\theta_{1:n-1})$$
$$\times p\left(r_{1:n-1},\theta_{1:n-1}\right|y_{1:n-1}).$$

$\Rightarrow$ Sequential Monte Carlo Methods

• *Numerous previous work in Comms.* Optimal Filtering (e.g. Snyder 1969) & Bayes Estimation.

Mapping: k-bit digital sequence $\leftrightarrow$ transmitted signal waveform

$$s_{\text{trans}}(\tau) = \text{Re}[s_n(r_{1:n})h(\tau)\exp(j2\pi f_c\tau)], \quad r_{1:n} = (r_1, \ldots, r_n)^{\mathtt{T}},$$

where

- f_c - carrier frequency,
- $h(\tau)$ - real-valued signal pulse,
- $s_n(\cdot)$ - mapping function,
- r_n - indicator variable associated with one of $M = 2^k$ possible k-bit sequences.

Assumption: r_n is a first order, time-homogeneous, M-state, Markov process with known transition probabilities $p_{ij}, i, j \in \mathcal{R}$.

2.2– Noisy Rayleigh Fading Channel Model

- **Multiplicative disturbance** g_n: complex zero-mean low-pass filtered Gaussian process, i.e. ARMA(q, q) process (say Butterworth filter of order q).

$$g_n = \mathbf{a}^{\mathsf{T}}\mathbf{g}_{n-1:n-q} + \mathbf{b}^{\mathsf{T}}\mathbf{v}_{n:n-q}, \quad v_n \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0, 1\right).$$

- Filter coefficients **a**, **b** are chosen so that
  - transfer function of the filter ⇔ **power spectral density** of the channel (f_dT is the cut-off frequency of the filter/where f_dT - is the normalized Doppler frequency).
  - $var\{abs(g_n)\} = 1$.

- Define a **state** x_n so that

$$g_n = \mathbf{b}^{\mathsf{T}} x_{n:n-q+1}, \qquad x_n \in \mathbb{C}.$$

$$x_{n:n-q+1} = \mathbf{A} x_{n-1:n-q} + \mathbf{B} v_n, \quad v_n \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1).$$

- Complex **output** of the filter matched to $h(\tau)$

$$y_n = s_n(r_{1:n}) g_n + e_n, \qquad e_n \overset{i.i.d.}{\sim} \mathcal{N}_c\left(0, \sigma_j^2\right),$$

$$= \mathbf{C}(r_{1:n}) x_{n:n-q+1} + \mathbf{D} w_n, \;\; w_n \overset{i.i.d.}{\sim} \mathcal{N}_c(0,1).$$

where

- $x_{n:n-q+1}$, v_n, w_n mutually independent for all $n \geq 1$,
- $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}(r_{1:n})$, $\mathbf{D}$ known for each $r_n \in \mathcal{R}$, $z_n \in \mathcal{R}_z$,
- r_n, x_n, z_n *unknown* for $n \geq 1$.

- **Filtering**:
  - compute the filtering distribution $p\left(r_n \mid y_{1:n}\right)$,
  - MAP estimates of the symbols

$$\hat{r}_{n|n} = \arg\max p\left(r_n \mid y_{1:n}\right).$$

- **Fixed-lag smoothing:**
  - compute the fixed-lag smoothing distribution $p\left(r_n \mid y_{1:n+L}\right)$, $L \in \mathbb{N}^*$,
  - MAP estimates of the symbols

$$\hat{r}_{n|n+L} = \arg\max p\left(r_n \mid y_{1:n+L}\right).$$

- *Jump Markov Linear Systems*

$$\begin{aligned} x_n &= A(m_n)\,x_{n-1} + B(m_n)\,v_n,\ V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I) \\ y_n &= C(m_n)\,x_n + D(m_n)\,w_n,\ w_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I) \end{aligned}$$

where $\{m_n\}_{n\geq 0}$ is a discrete (Markov) sequence taking values in $\mathbb{M}$.

- *Multiple Applications* in communications, tracking, econometrics, etc.
- *Key idea*

$$p(x_{1:n}, m_{1:n} | y_{1:n}) = \underbrace{p(x_{1:n} | y_{1:n}, m_{1:n})}_{\text{Gaussian distribution}} \quad \underbrace{p(m_{1:n} | y_{1:n})}_{\text{known up to a normalizing constant}}$$

$$p(m_{1:n} | y_{1:n}) \propto \underbrace{g(y_{1:n} | m_{1:n})}_{\text{likelihood - Kalman filter}} \quad f(m_{1:n}).$$

- *Rao-Blackwellisation.* Use PF only to estimate $p(m_{1:n}|\, y_{1:n})$.

- *"Optimal" Importance Distribution*

$$q(m_n|\, y_{1:n}, m_{1:n-1}) \quad \propto \quad \underbrace{g(y_n|\, y_{1:n-1}, m_{1:n})}_{\text{Kalman Filter}} f(m_n|\, m_{n-1}),$$

$$w_n \quad \propto \quad \sum_{m_n \in \mathbb{M}} g(y_n|\, y_{1:n-1}, m_{1:n}) f(m_n|\, m_{n-1}).$$

- Might be too computationally expensive: suboptimal distribution based on standard algorithms.

- No need to propagate $m_{1:n}$ but just m_n, $E[m_n|\, y_{1:n}]$, $cov[m_n|\, y_{1:n}]$!

Assuming at time $n-1$, one has N unweighted particles $\left\{M_{1:n-1}^{(i)}\right\}$ distributed according to $p\left(m_{1:n-1}\middle|\,y_{1:n-1}\right)$.

- Set $\widetilde{M}_{1:n-1}^{(i)} = M_{1:n-1}^{(i)}$. Multiply/Discard particles $\left\{\widetilde{M}_{1:n}^{(i)}\right\}$ with respect to high/low weights $\left\{W_n^{(i)}\right\}$ to obtain N particles $\left\{M_{1:n}^{(i)}\right\}$ where

$$W_n^{(i)} \propto \sum_{m_n \in \mathbb{M}} g\left(y_n\middle|\,y_{1:n-1}, \widetilde{M}_{1:n}^{(i)}\right) f\left(\widetilde{M}_n^{(i)}\middle|\,\widetilde{M}_{n-1}^{(i)}\right), \quad \sum_{i=1}^{N} W_n^{(i)} = 1;$$

- For $i = 1, \ldots, N$ Sample $M_n^{(i)} \sim q\left(\cdot\middle|\,y_{1:n}, \widetilde{M}_{1:n-1}^{(i)}\right)$.

2.8– Demodulation of 8-DPSK Signals

- Mapping function

$$s_n = A_c \exp(j\theta_n), \quad \theta_n = \sum_{k=1}^{n} \sum_{m=1}^{M} \frac{2\pi m}{M} \mathbb{I}_{r_m}(r_k) .$$

- Channel characteristics: LPF white Gaussian noise with cutoff frequency corresponding to a normalized Doppler frequency $f_d T$.

- Parameters of the algorithm: $N = 50$, $L = 5$ for fixed-lag smoothing.

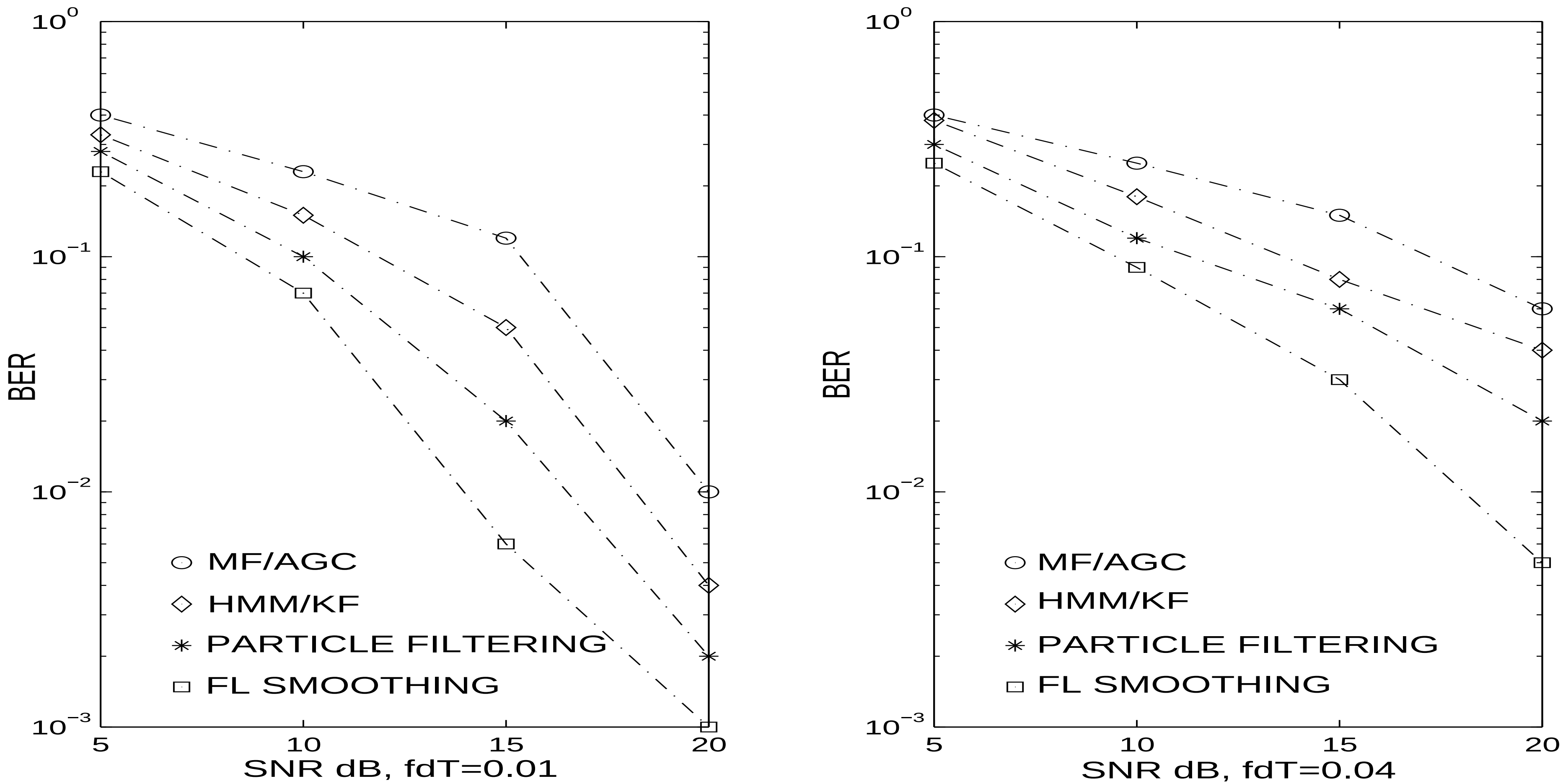

Figure 1: Bit error rate (additive Gaussian noise).

- *Frequency selective channel.*

- *Non Gaussian noise.* e_n distributed as a **mixture of** complex zero-mean **Gaussians**

$$e_n \sim \sum_{j=1}^{K} \lambda_j \mathcal{N}_c\left(0, \sigma_j^2\right),$$

with the latent **allocation variable** $z_n \in \{1, 2, \ldots, K\}$, $n = 1, 2, \ldots$

$$e_n|\, z_n \sim \mathcal{N}_c\left(0, \sigma_{z_n}^2\right), \quad , \quad \Pr(z_n = j) = \lambda_j, \text{ for } j = 1, \ldots, K, \quad \sum_{j=1}^{K} \lambda_j = 1.$$

- *Diversity*

- *CDMA*

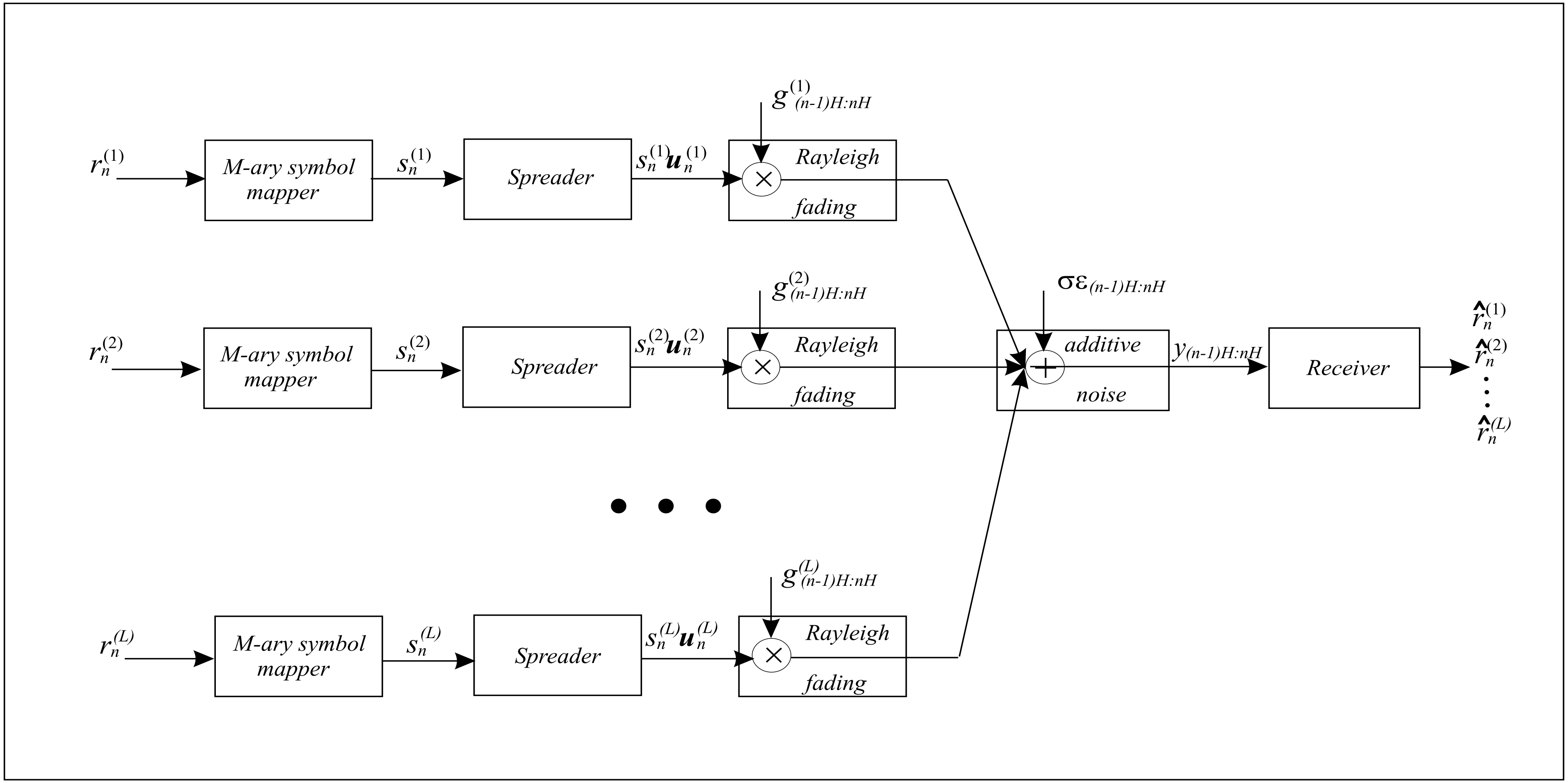

Figure 2: CDMA System

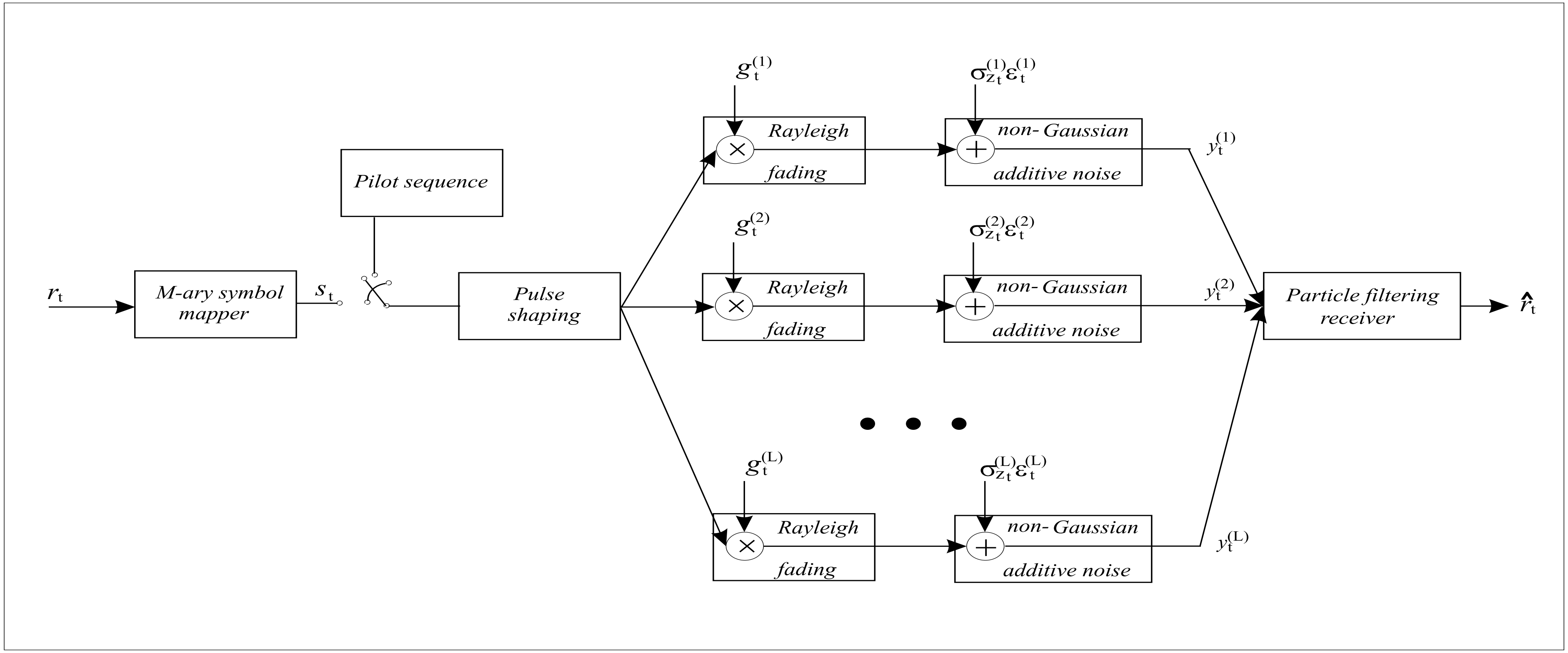

Figure 3: Diversity.

2.11– Particle Filtering for JMLS - The bad

- "What you're doing might not be very clever..." - Simon Maskell
- Two Points
  - Calculations for $N \times |\mathbb{M}|$ particles done! Why sampling then and not selecting the "best" N?
  - Keep the weights of particles!
- Resampling method by Fearnhead (1998)... necessary?

2.12– A Simple Deterministic Algorithm

Assuming at time $n-1$, one has N weighted particles $\left\{M_{1:n-1}^{(i)}\right\}$ by $\left\{W_{n-1}^{(i)}\right\}$ such that $p\left(m_{1:n-1}\middle|\, y_{1:n-1}\right) \approx \sum_{i=1}^{N} W_{n-1}^{(i)} \delta\left(m_{1:n-1} - M_{1:n-1}^{(i)}\right)$.

- For $i = 1, \ldots, N \quad j = 1, \ldots, |\mathbb{M}|$ compute

$$W_n^{(i,j)} \propto g\left(y_n \middle|\, y_{1:n-1}, M_{1:n-1}^{(i)}\right) f\left(M_n = j \middle|\, M_{n-1}^{(i)}\right) W_{n-1}^{(i)}.$$

- Keep the N particles with the highest weights among the $N\,|\mathbb{M}|$ candidates.

- **Problem:** Particles never forget the past... (Tugnait, 1979, Punskaya et al. IEEE SSP August 2001)

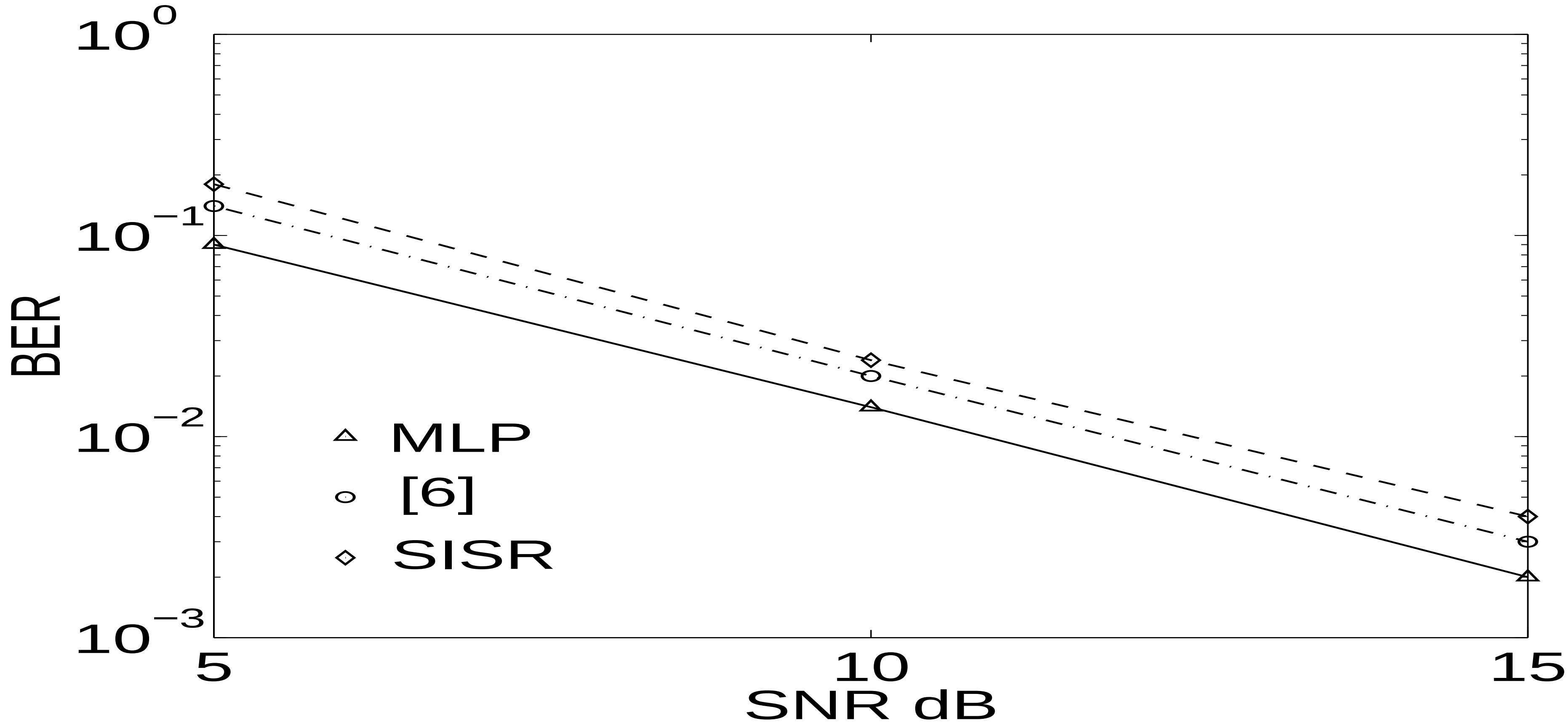

Figure 4: Frequency Selective Channels.

2.14– Deterministic vs Stochastic Methods = The Ugly

- N survivors Viterbi-like Algorithm

- In all applications tested (tracking, communications, mixtures etc)

$\Longrightarrow$ Deterministic works better!!!!

- Further improvement using discount factor idea (Del Moral 1994; see Ph.D. Punskaya 2002).

- Transmitted **spread-spectrum** signal waveform:

$$s_{\text{trans}}(\tau) = \text{Re}[s_n(d_n)PN(\tau)\exp(j2\pi f_c\tau)], \quad \text{for } (n-1)T_d < \tau \leq nT_d,$$

where

- d_n - nth information **symbol** transmitted in the symbol interval T_d,
- $s_n(.)$ - mapping function: digital sequence $\Leftrightarrow$ waveforms (corresponds to the *modulation* technique employed),
- f_c - carrier frequency,
- $PN(\tau)$ - wide-band **pseudo-noise** (PN) waveform $PN(\tau) = \sum_{h=1}^{H} a_h \eta(\tau - hT_c)$,
- $a_{1:H}$ - *spreading code* sequence (H **chips** per symbol with values $\{\pm 1\}$),
- $\eta(\tau - hT_c)$ - rectangular pulse of unit height and duration of chip interval $T_c = T_d/H$.

• **Model**: time-varying *tapped-delayed* line taps spaced at Nyquist sampling rate $T_s = T_c/2$.

• **Impulse response**: $h_{c,t} = \sum_{n_f=0}^{N_f-1} f_t^{(n_f)} \delta_{t,n_f}$, with N_f being the number of paths of the channel.

• **Channel coefficients $\mathbf{f}_t$** : first order *AR model*

$$\mathbf{f}_t = \mathbf{A}_f \mathbf{f}_{t-1} + \mathbf{B}_f \mathbf{v}_t, \mathbf{v}_t \overset{i.i.d.}{\sim} \mathcal{N}_c \left(\mathbf{0}, \mathbf{I}_{N_f} \right),$$

$\mathbf{A}_f \triangleq diag(\alpha_0, \ldots, \alpha_{N_f-1})$, $\mathbf{B}_f \triangleq diag(\sigma_{f,0}, \ldots, \sigma_{f,N_f-1})$; α_{n_f} accounting for the Doppler spread, σ^2_{f,n_f} being the noise variance

• **Code delay θ_t** : first order AR

$$\theta_t = \gamma \theta_{t-1} + \sigma_\theta \epsilon_t, \ \epsilon_t \overset{i.i.d.}{\sim} \mathcal{N} \left(0, 1 \right),$$

- Complex *output* of the channel

$$y_t = \mathbf{C}(d_{1:n}, \theta_{1:t}) + \sigma\varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_c(0, 1),$$

where

- $\mathbf{C}(d_{1:n}, \theta_{1:t}) = \sum_{n_f=0}^{N_f-1} f_t^{(n_f)} s\left((t - n_f) T_s - \theta_t\right),$
- σ^2 is the noise variance
- $d_n \leftrightarrow y_{2H(n-1)+1:2Hn}$ (i.e. $t = 2H(n-1)+1, \ldots, 2Hn$ samples correspond to the nth symbol transmitted)

• **Unknown parameters:** symbols d_n, channel characteristics $\mathbf{f}_t$, code delay θ_t: *continuous-valued* parameter is involved!

• **Objectives:** obtain *sequentially* an estimate of $p\left(\left. d_{1:n}, \mathbf{f}_{0:2Hn}, \theta_{0:2Hn} \right| y_{1:2Hn}\right)$, obtain estimates of unknown parameters $\mathbb{E}\left(\left. d_{1:n} \right| y_{1:2Hn}\right)$, $\mathbb{E}\left(\left. \mathbf{f}_{0:2Hn} \right| y_{1:2Hn}\right)$ and $\mathbb{E}\left(\left. \theta_{0:2Hn} \right| y_{1:2Hn}\right)$.

• **Problem:** No analytical solution $\Rightarrow$ approximate methods must be employed.

$$\hat{p}_N\left(\left. d_{1:n}, d\theta_{1:2Hn} \right| y_{1:2Hn}\right) = \sum_{i=1}^N \tilde{w}_n^{(i)} \delta_{\left(D_{1:n}^{(i)}, \Theta_{1:2nH}^{(i)}\right)}\left(d_{1:n}, d\theta_{1:2Hn}\right),$$

$$\widehat{p}_N\left(\left. \mathbf{f}_{1:2Hn} \right| y_{1:2Hn}\right) = \sum_{i=1}^N p(\left. \mathbf{f}_{1:2Hn} \right| y_{1:2Hn}, D_{1:n}^{(i)}, \Theta_{1:2nH}^{(i)}).$$

3.5– Joint Demodulation/Code Delay Estimation

Sampling Step

- For $i = 1, \ldots, N$, set $(\widetilde{D}_{0:n-1}^{(i)}, \widetilde{\Theta}_{0:2H(n-1)}^{(i)}) = \left(D_{0:n-1}^{(i)}, \Theta_{0:2H(n-1)}^{(i)} \right)$ and sample

$(\widetilde{D}_{n}^{(i)}, \widetilde{\Theta}_{2H(n-1)+1:2Hn}^{(i)})$

$\sim \pi(\cdot | \, \widetilde{D}_{1:n-1}^{(i)}, \widetilde{\Theta}_{1:2H(n-1)}^{(i)}, y_{1:2Hn})$ and evaluate

$$W_n^{(i)} \propto \frac{p(\widetilde{D}_n^{(i)}) \prod_{t=2H(n-1)+1}^{2Hn} p\left(y_t | \, \widetilde{D}_{1:n}^{(i)}, \widetilde{\Theta}_{1:t}^{(i)}, y_{1:t-1} \right) p(\widetilde{\Theta}_t^{(i)} \Big| \, \widetilde{\Theta}_{t-1}^{(i)})}{\pi(\widetilde{D}_n^{(i)}, \widetilde{\Theta}_{2H(n-1)+1:2Hn}^{(i)} \Big| \, \widetilde{D}_{1:n-1}^{(i)}, \widetilde{\Theta}_{1:2H(n-1)}^{(i)}, y_{1:2Hn})}.$$

Selection Step

- Multiply/discard particles with respect to high/low $\left\{ W_n^{(i)} \right\}$ to obtain N

particles $\left\{ D_{1:n}^{(i)}, \Theta_{1:2Hn}^{(i)} \right\}$

- **Remark 1.** For the **prior** $p(d_n)\prod_{t=2H(n-1)+1}^{2Hn} p(\,\theta_t|\,\theta_{t-1})$ taken as an importance distribution:

$$W_n^{(i)} \propto \prod_{t=2H(n-1)+1}^{2Hn} p\left(y_t|\,\widetilde{D}_{1:n}^{(i)}, \widetilde{\Theta}_{1:t}^{(i)}, y_{1:t-1}\right),$$

$\Rightarrow 2H$ one-step Kalman filter updates are required.

- **Remark 2.** For $H \gg 1$ - the state space to explore is large
  - Sample $\left(D_n^{(i)}, \Theta_{2H(n-1)+1}^{(i)}\right)$ and update the distribution with $y_{2H(n-1)+1}$
  - for $k = 2H(n-1)+2, \ldots, 2Hn :\ \Theta_k^{(i)} \sim p(\,\cdot|\,\Theta_{k-1}^{(i)})$ and update the distribution with y_k.

- **BPSK Modulation**.

- **DS spread-system characteristics:** number of chips $H = 15$.

- **Multipath fading:** channel B [Iitis 1990], number of paths $N_f = 4$, coefficients $\alpha_{n_f} = 0.95$, $\sigma^2_{f,n_f} = 0.01$, constant delay $\gamma = 0.95$, $\sigma^2_\theta = 0.01$

- **Parameters of the algorithm:** $N = 50$.

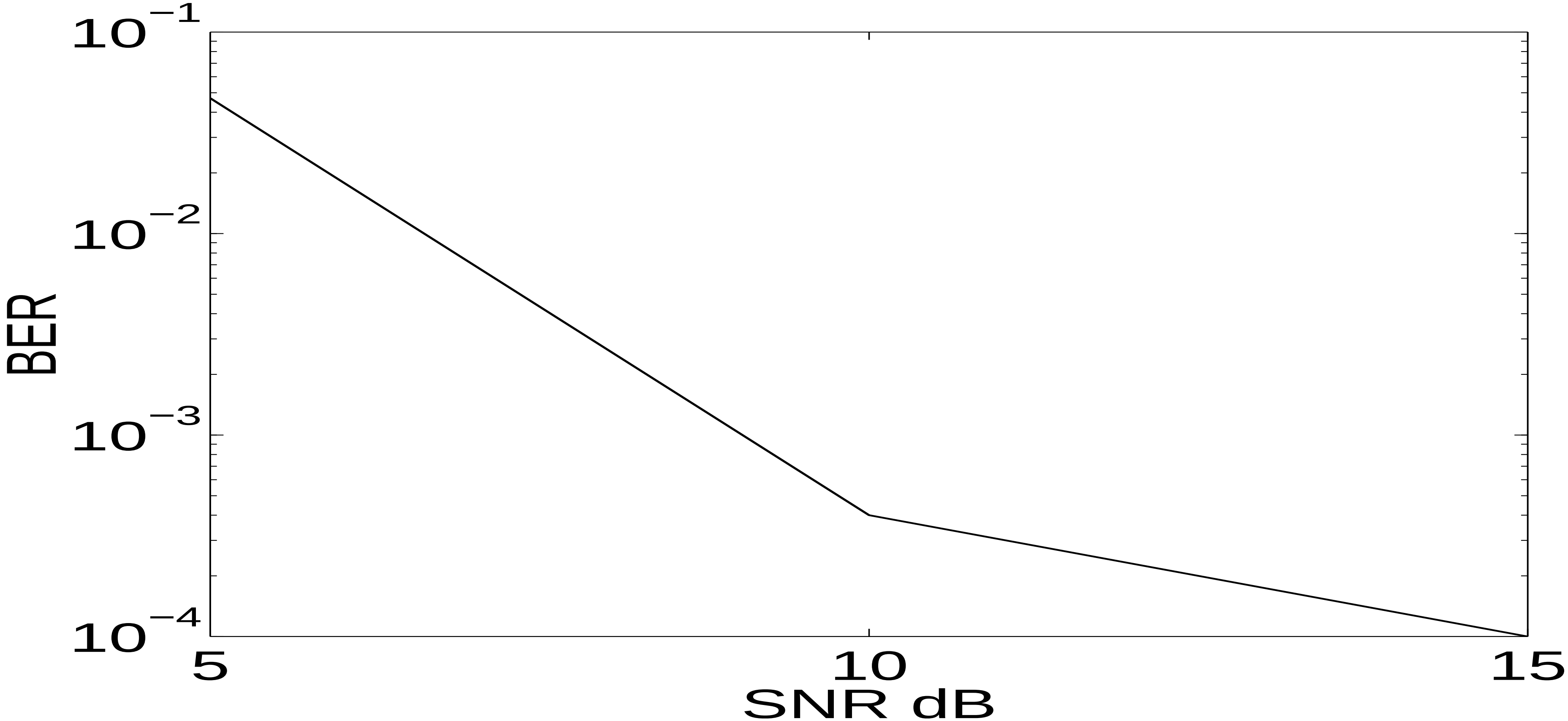

Figure 5: BER for Synchronization.

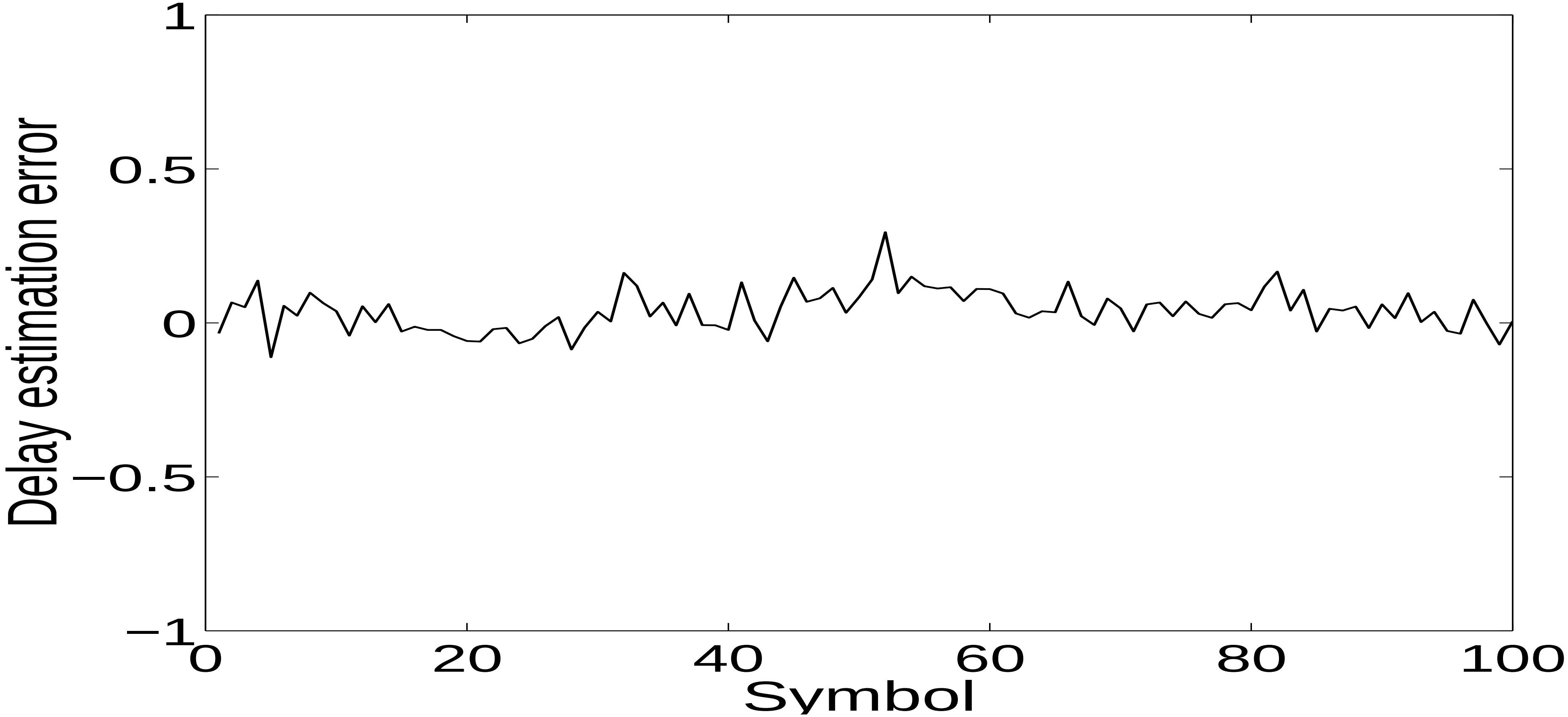

Figure 6: Delay Error for SNR 15dB.

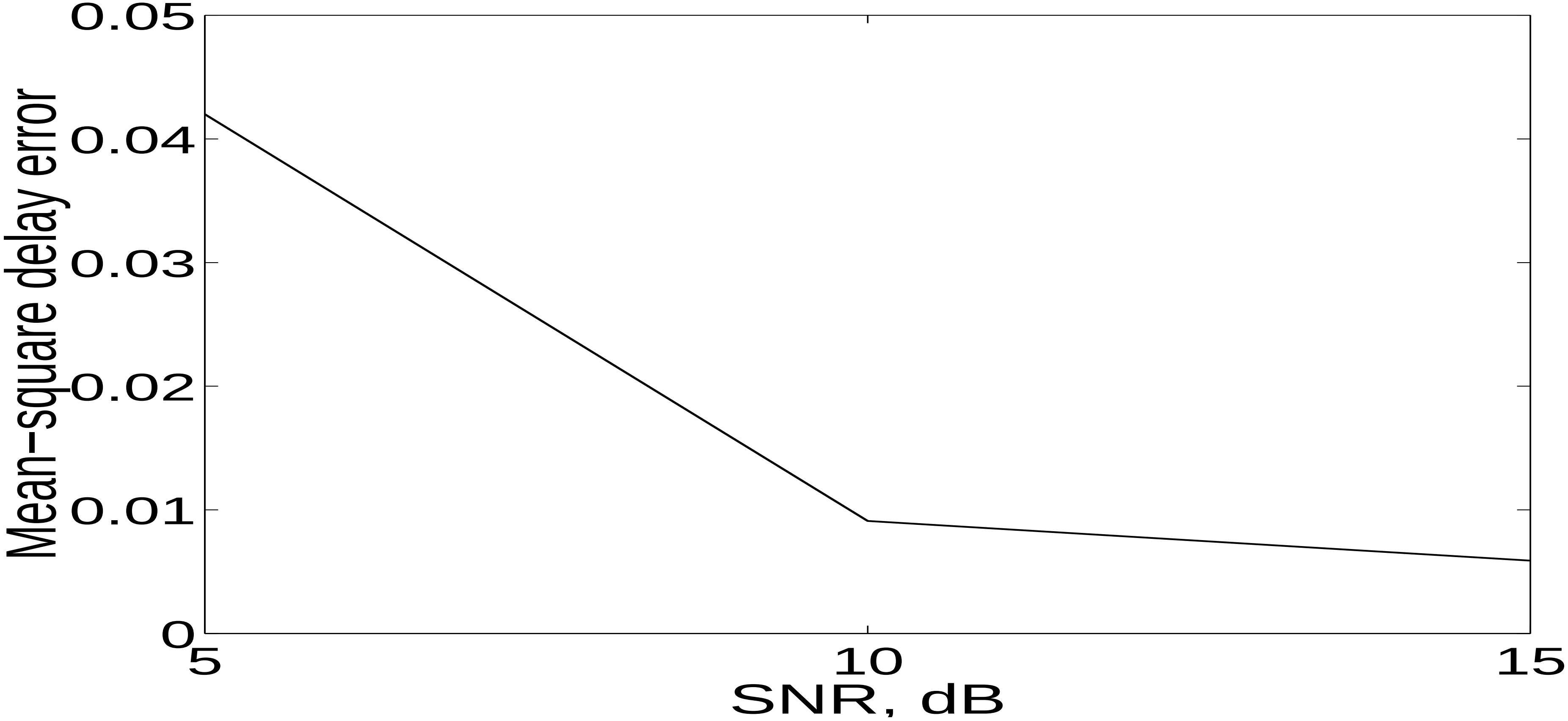

Figure 7: MSE Delay for various SNR.

3.9– Discussion

- Standard SMC methods do *not* necessarily outperform basic deterministic approaches when only discrete parameters to estimate.

- SMC methods appear to be *useful* when

    - continuous-valued unknown parameters are involved (synchronization, semi-parametric channels).
    - number of discrete states is very large (CDMA).

- *Advantages*: You can base algorithms on intuitions / current methods.

- *Drawbacks*: Given computational complexity, does it make sense in most comms applications???

4.1– References

- SMC Website: `http://www-sigproc.eng.cam.ac.uk/smc/index.html`

- *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, 2001.

- Special Issue *Applied Signal Processing*, Deadline April 30, 2003.

- Many potential algorithmical and theoretical developments / Numerous applications.