Efficient Markov Chain Monte Carlo Algorithms for Iterative Detection of Multiple Access Channels

Zhenning Shi
WSP group, National ICT Australia
Level 2, 216 Northbourne Avenue
Braddon, ACT 2612, Australia
zhenning.shi@nicta.com.au

NATIONAL ICT AUSTRALIA LIMITED

Introduction

- MAI and Multiuser Detection Techniques
- Turbo Receiver Structure
- MCMC Techniques and Their Applications to Multiple-Access Channels
- Some Properties of MCMC in CDMA
- An Iterative CDMA Receiver using MCMC Approach: MCRB
- Two MCMC detectors (MCRB-U, modified MCRB-U) that are robust wrt Heavily Loaded Channels and High SNRs

MAI and Multiuser Detection

In a Synchronous CDMA channel, the receive signal is

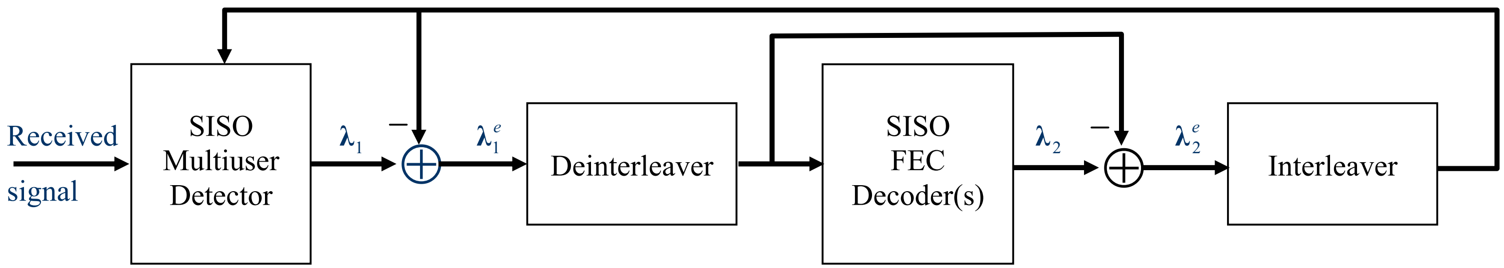
$$\mathbf{y} \quad = \quad \mathbf{A}\mathbf{d} + \mathbf{n} = \mathbf{a}_k d_k + \overbrace{\mathbf{A}_{-k}\mathbf{d}_{-k}}^{\text{MAI } I_k} + \mathbf{n} \tag{1}$$

- MAP detection ([Verdu 84])
- Minimum Mean-Square Error Filter ([Madhow&Honig 94], [Miller 95])
- Decorrelator; Parallel Interference Canceller (PIC)
- Successive Interference Canceller (SIC)

The system model is on synchronous CDMA using BPSK modulations, but the methodology can be easily extended to the treatment of MIMO channels

Turbo Receiver Structure

(CDMA) Turbo Receiver Diagram



- Two SISO modules: CDMA multiuser detector (MUD) and FEC decoders
- Extrinsic LLRs on the transmitted symbols are exchanged between two processing cores, and get improved over iterations; hard decision is made at the last iteration
- Iterative receiver designs: APP CDMA detector ([Moher]), MMSE filter ([Wang & Poor]), Multi-stage filters, Interference Canceller ([Schlegel et. al.])...

Problem in Multiuser Detection

- The MUD module (ideally) outputs the APP statistics that requires multi-dimensional summation:

$$P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e) = \sum_{\mathbf{d}_{-k}} P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e, \mathbf{d}_{-k}) P(\mathbf{d}_{-k}|\mathbf{y}, \boldsymbol{\lambda}_2^e)$$

- Monte Carlo intergration:

$$E_f\left[h(x)\right] = \int h(x)f(x)dx \approx \frac{1}{N_s}\sum_{n=1}^{N_s} h(x_n)$$

where $f(x)$ is a valid distribution, and $\{x(n)\}_{n=1}^{N_s}$ are samples taken from $f(x)$

MCMC Techniques

- MCMC obtains the statistical inference of the signals by simulating the underlying process through Markov Chains
- MCMC is suitable for addressing problems involving high-dimensional summations or integrals
- Instead of evaluating all summation terms (exponentail complexity), *average* over the samples from the complex distribution
- Techniques: importance sampling, metroplised algorithms (Gibbs Sampler), sequential MCMC methods ...
- Applications: molecular simulation, computational biology, target tracking, blind equalization, blind detections over fading channels, joint channel parameter estimation and detection ...

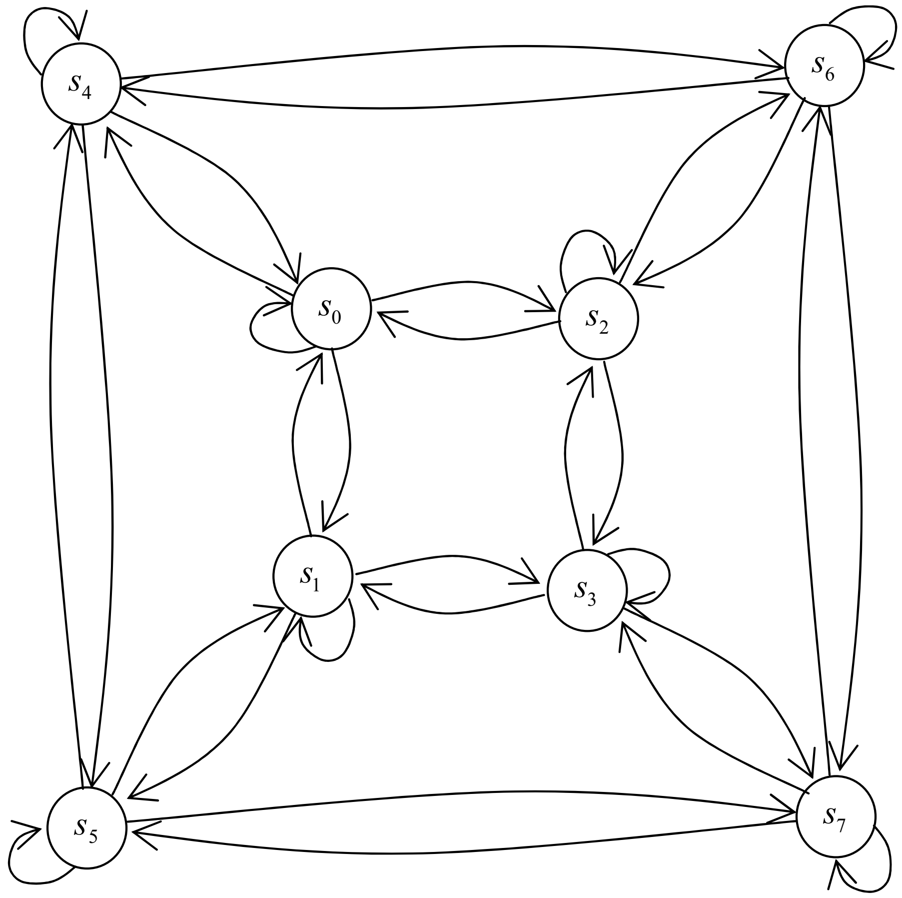
MCMC in Digital Communications

- Detection of Multiple Access Channels:
  - Gibbs Sampler: a special MCMC method that takes Markov samples marginally based on the conditional target distribution ([Schmidl et. al.], [Buchoux et. al.])
  - Gibbs Coupler: maintain the entire search space at the beginning, narrow down to one sample using the conditional probability bounds ([Huang & Djuric])
    - sample from the exact distribution, can determine the *burn-in* period
    - slow convergence at highly correlated scenarios
  - Particle Filtering: a sequential Markov chain method that recursively computes the relevant distribution, and use the associated weights to resample
    - address the particle degeneration problem
    - require Cholesky decomposition at front-end in detecting multiple-access channel (complexity, and performance in overloaded cases??)

Properties of MCMC of CDMA

State Diagram of the Markov chain for a three-user CDMA

$$
\begin{aligned}
s_0 &= \{-1,-1,-1\} \\
s_1 &= \{-1,-1,+1\} \\
s_2 &= \{-1,+1,-1\} \\
s_3 &= \{-1,+1,+1\} \\
s_4 &= \{+1,-1,-1\} \\
s_5 &= \{+1,-1,+1\} \\
s_6 &= \{+1,+1,-1\} \\
s_7 &= \{+1,+1,+1\}
\end{aligned}
$$



Properties of MCMC of CDMA

- The CDMA Markov chain is homogeneous (a static transition probability matrix);
- irreducible: any state in the Markov chain is reachable from another state within finite steps ($P(s_i \rightarrow s_j) \neq 0, \quad , i \neq j$)
- aperiodic: no transition period that is larger than one exists (self-returning edges)

Observation: Since the CDMA Markov chain is homogeneous, iirreducibleand aperiodic, it converges to a unique stationary distribution starting an arbitrary state with enough steps (if the stationary distribution is APP, Gibbs sampler asymptotically achieve optimal performance of MUD)

Properties of MCMC of CDMA

In a Markov chain with the transition matrix Π and the state space $\mathcal{S}$, a distribution $\mathbf{p}$ is said to be reversible, if for neighbouring states s_i and s_j

$$p_i \pi_{ij} = p_j \pi_{ji}$$

Theorem: For a homogeneous Markov chain, the reversible distribution $\mathbf{p}$ is also a stationary distribution of the chain

In the CDMA Markov chain, for neighbouring state pair (s_i, s_j), define

$$p_i = P(\mathbf{d}^i|\mathbf{y}, \boldsymbol{\lambda}_2^e), \qquad \pi_{i,j} = \frac{P\left(d_k^j|\mathbf{y}, \mathbf{d}_{-k}^j, \boldsymbol{\lambda}_2^e\right)}{P\left(d_k^i|\mathbf{y}, \mathbf{d}_{-k}^i, \boldsymbol{\lambda}_2^e\right) + P\left(d_k^j|\mathbf{y}, \mathbf{d}_{-k}^j, \boldsymbol{\lambda}_2^e\right)} \propto P(\mathbf{d}^j|\mathbf{y}, \boldsymbol{\lambda}_2^e)$$

$$p_j = P(\mathbf{d}^j|\mathbf{y}, \boldsymbol{\lambda}_2^e), \qquad \pi_{j,i} = \frac{P\left(d_k^i|\mathbf{y}, \mathbf{d}_{-k}^i, \boldsymbol{\lambda}_2^e\right)}{P\left(d_k^i|\mathbf{y}, \mathbf{d}_{-k}^i, \boldsymbol{\lambda}_2^e\right) + P\left(d_k^j|\mathbf{y}, \mathbf{d}_{-k}^j, \boldsymbol{\lambda}_2^e\right)} \propto P(\mathbf{d}^i|\mathbf{y}, \boldsymbol{\lambda}_2^e)$$

it is obvious that $p_i \pi_{ij} = p_j \pi_{ji}$, hence the CDMA Markov chain is reversible, and $p_i \left(= P(\mathbf{d}^i|\mathbf{y}, \boldsymbol{\lambda}_2^e)\right)$ is the unique stationary distribution for state s_i.

An Iterative MCMC Receiver (MCRB)

- Gibbs sampler used for obtaining Bayesian inference (Inputs: $\mathbf{y}, \boldsymbol{\lambda}_2^e$; Output: $\boldsymbol{\lambda}_1^e$)

A1 `Initialize` $\mathbf{d}^{(N_b)}$ `(random seed)`

A2 `FOR` $n = -N_b + 1$ `to` N_s

`draw` $d_1^{(n)} \sim P\left(d_1 | d_2^{(n-1)}, \ldots, d_K^{(n-1)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)$

`draw` $d_2^{(n)} \sim P\left(d_2 | d_1^{(n)}, d_3^{(n-1)}, \ldots, d_K^{(n-1)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)$

$\vdots$

`draw` $d_K^{(n)} \sim P\left(d_K | d_1^{(n)}, \ldots, d_{K-1}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)$

`END FOR`

A3 `Calculate the` *`a posteriori`* `probability`

$$P(d_k = 1 | \mathbf{y}, \boldsymbol{\lambda}_2^e) \approx \mathtt{Est}\left(\{\mathbf{d}^{(n)}\}_{n=1}^{N_s}\right), \quad k = 1, \ldots, K$$

A4 `Calculate the extrinsic LLR`

$$\lambda_1^e(d_k) = \ln\left(\frac{p(d_k=1|\mathbf{y},\boldsymbol{\lambda}_2^e)}{1-p(d_k=1|\mathbf{y},\boldsymbol{\lambda}_2^e)}\right) - \lambda_2^e(d_k), \quad k = 1, \ldots, K$$

An Iterative MCMC Receiver (MCRB)

- Transition probability $P\left(d_k|\mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)$ calculation

$$\begin{aligned}
\lambda_1^{(n)}(d_k) &= \ln \frac{P\left(d_k = +1|\mathbf{y}, \mathbf{d}_{-k}^{(n)}, \boldsymbol{\lambda}_2^e\right)}{P\left(d_k = -1|\mathbf{y}, \mathbf{d}_{-k}^{(n)}, \boldsymbol{\lambda}_2^e\right)} \\
&= \ln \frac{P\left(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = +1\right) P\left(\mathbf{d}_{-k}^{(n)}, d_k = +1|\boldsymbol{\lambda}_2^e\right)}{P\left(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = -1\right) P\left(\mathbf{d}_{-k}^{(n)}, d_k = -1|\boldsymbol{\lambda}_2^e\right)} \\
&= \ln \frac{P\left(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = +1\right)}{P\left(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = -1\right)} + \lambda_2^e(d_k) \\
&= \frac{1}{N_0}\left(|\mathbf{y} - \mathbf{A}_{-k}\mathbf{d}_{-k} + \mathbf{a}_k|^2 - |\mathbf{y} - \mathbf{A}_{-k}\mathbf{d}_{-k}^{(n)} - \mathbf{a}_k|^2\right) + \lambda_2^e(d_k) \\
&= \frac{4}{N_0}\mathbf{a}_k^T\left(\mathbf{y} - \mathbf{A}_{-k}\mathbf{d}_{-k}^{(n)}\right) + \lambda_2^e(d_k) \\
P(d_k = 1|\mathbf{y}, \mathbf{d}_{-k}^{(n)}, \boldsymbol{\lambda}_2^e) &= \frac{1}{1 + \exp(\lambda_1^{(n)}(d_k))}
\end{aligned}$$

An Iterative MCMC Receiver (MCRB)

- Bayesian Estimation Method Est $\left(\left\{\mathbf{d}^{(n)}\right\}_{n=1}^{N_s}\right)$ for calculation of $P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e)$
  - Method in the previous literature: Bit Counting

$$P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e) \approx \frac{1}{N_s} \sum_{n=1}^{N_s} \delta(d_k^{(n)} = 1)$$

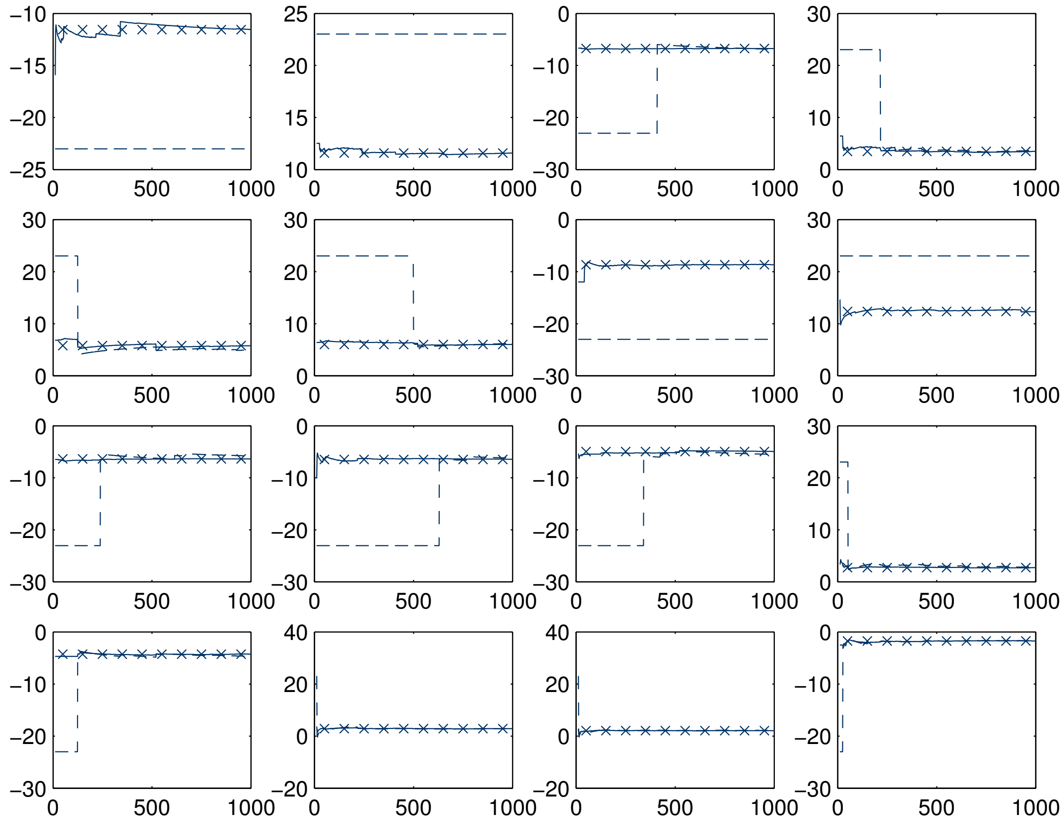
  - Proposed Method: Monte Carlo Rao-Blackwellization

$$\begin{aligned} P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e) &= \sum_{\mathbf{d}_{-k}} P\left(d_k = 1, \mathbf{d}_{-k}|\mathbf{y}, \boldsymbol{\lambda}_2^e\right) = \sum_{\mathbf{d}_{-k}} P\left(d_k = 1|\mathbf{y}, \mathbf{d}_{-k}\right) P\left(\mathbf{d}_{-k}|\mathbf{y}, \boldsymbol{\lambda}_2^e\right) \\ &\approx \frac{1}{N_s} \sum_{n=1}^{N_s} P\left(d_k = 1|\mathbf{y}, \mathbf{d}_{-k}^{(n)}, \boldsymbol{\lambda}_2^e\right) = \frac{1}{N_s} \sum_{n=1}^{N_s} \frac{1}{1 + \exp\left(\lambda_1^{(n)}(d_k)\right)} \end{aligned}$$

  - MCRB is the Rao-Blackwellization of Bit Counting, i.e., $P\left(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e, \mathbf{d}_{-k}\right) = E\left[\delta(d_k = 1)|\mathbf{y}, \boldsymbol{\lambda}_2^e, \mathbf{d}_{-k}\right]$, hence the former is much more efficient in evaluating the *a posteriori* probabilities

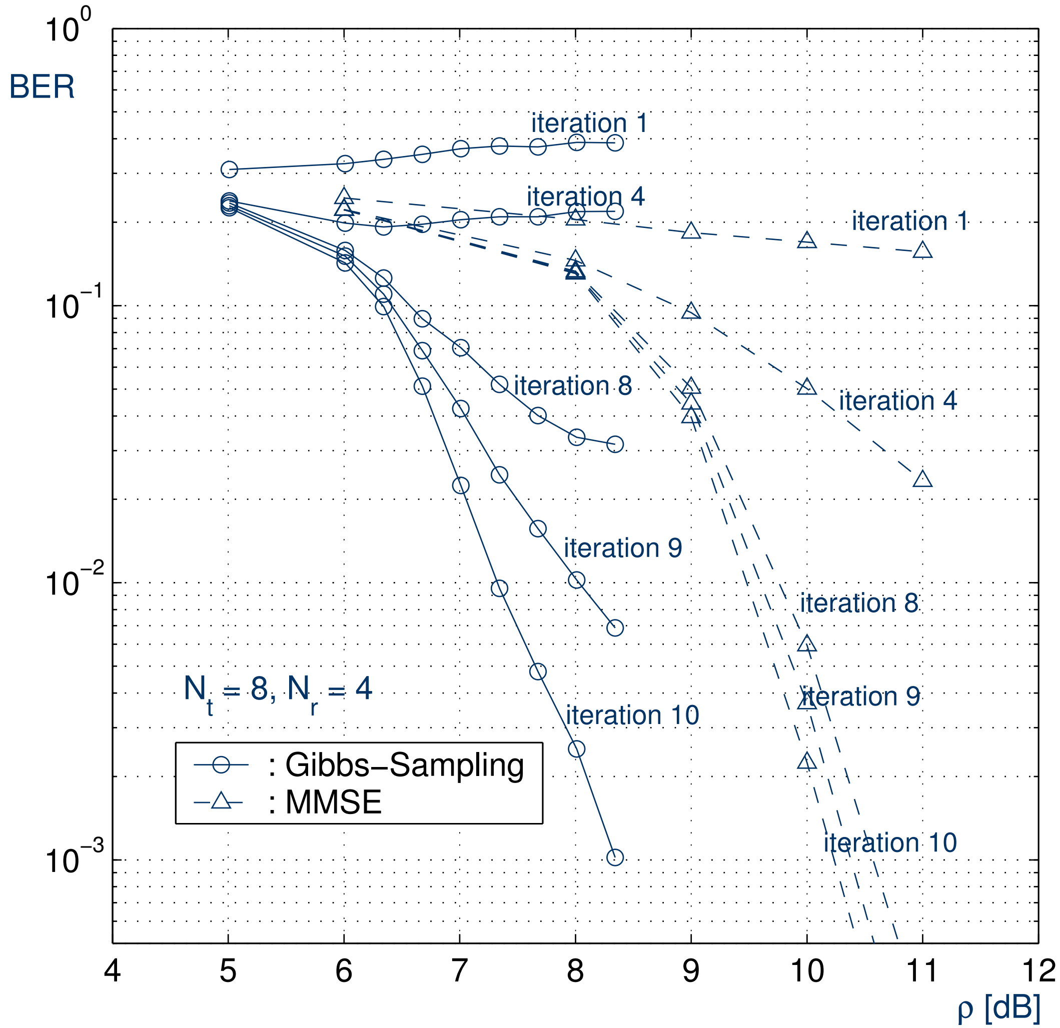
Comparison between Two Bayesian Estimation Methods

- A 16-user CDMA system, with processing gain of 8, SNR=3 dB
- The reliability of FEC feedbacks: $I\left(d_k, \lambda_2^e(d_k)\right) = 0.5$
- Symbol counting requires more samples to converge
- Symbol counting results diverge from the true values when L-values are large



Performance in a MIMO channel

- $N_t = 8, N_r = 4$, Rayleigh fading
- $(5, 7)$ CC, rate-$1/2$, four-state, QPSK
- burn in $N_b = 10$, average over $N_s = 20$ samples
- Gibbs sampler outperforms MMSE with less complexity
- Slower convergence as N_t increases (higher correlation); performance degeneration as SNR increases

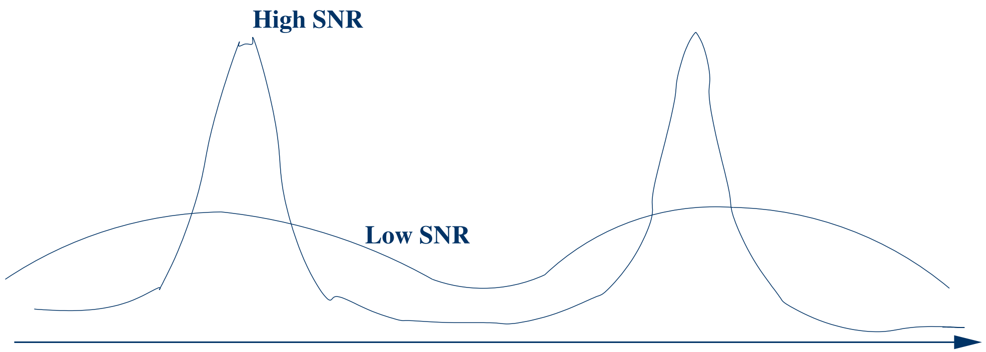


Problems associated with System Load and SNR

- Increase system load
  - Slow the Markov chain process, requires much longer burn-in period
  - Samples obtained in a finite number of stages don't converge to stationary distribution, cause iinsufficientstatistics
  - Grouping Technique may address this problem, at a cost of iincreasein complexity

- High Signal-to-Noise Ratio

- Leads to *peaked* stationary distribution
- Slow convergence, extreme case: when $\sigma^2 = 0$, Markov chain is no longer irreducible; trapped in one mode
- Markov chains of a fixed number of stages render *skewed* samples



MCRB-U Method: Motivation

- Sampled from stationary distribution (MCRB):

$$P\left(d_k = 1 | \mathbf{y}, \boldsymbol{\lambda}_2^e\right) \approx \frac{1}{N_s} \sum_{n=1}^{N_s} P\left(d_k = 1 | \mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)$$

Works if $\mathbf{d}_{-k}^{(n)} \sim P(\mathbf{d}_{-k} | \mathbf{y}, \boldsymbol{\lambda}_2^e)$

- Sampled from a uniform distribution (MCRB-U):

$$P\left(d_k = 1 | \mathbf{y}, \boldsymbol{\lambda}_2^e\right) \approx \frac{\sum_{n=1}^{N_s} P\left(d_k = 1 | \mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right) P\left(\mathbf{d}_{-k}^{(n)} | \mathbf{y}, \boldsymbol{\lambda}_2^e\right)}{\sum_{n=1}^{N_s} P\left(\mathbf{d}_{-k}^{(n)} | \mathbf{y}, \boldsymbol{\lambda}_2^e\right)}$$

- Due to the degeneration of Gibbs sampler, the samples obtained in a finite-step Markov chain do not follow the stationary distribution; summation over marginal probability weighted by appropriate sample weights (MCRB-U) provides more reliable statistics

MCRB-U Method: Algorithm

- MCRB-U Algorithm:

A1 Run M independent N_s-stage Gibbs sampler, construct a sample set $\mathcal{D} = \{\mathbf{d}^{(n)}\}, n = 1, 2, \ldots, MN_s$.

A2 Delete the repetitions of sample $\mathbf{d}^{(n)}$ in $\mathcal{D}$, build a new sample set $\mathcal{D}' \subseteq \mathcal{D}$, containing the distinct samples after pruning

A3 Calculate the APP probability as

$$P\left(d_k|\mathbf{y}, \boldsymbol{\lambda}_2^e\right) \approx \sum_{\mathcal{D}'} P\left(d_k|\mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right)\left(\frac{P(\mathbf{d}_{-k}^{(n)}|\mathbf{y}, \boldsymbol{\lambda}_2^e)}{\sum_{\mathcal{D}'} P(\mathbf{d}_{-k}^{(n)}|\mathbf{y}, \boldsymbol{\lambda}_2^e)}\right)$$

A4 Calculate the extrinsic LLR

$$\lambda_1^e(d_k) \quad = \quad \ln\left(\frac{p(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e)}{1 - p(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e)}\right) - \lambda_2^e(d_k), \qquad k = 1, \ldots, K$$

- Comments:
  - MCRB-U does not need burn-in steps, since every sample will improve the Bayesian inference

– The extrinsic LLR can be calculated without involving the *a priori* LLR of the target user

$$\lambda_1^e(d_k) = \ln \frac{P(d_k = 1|\mathbf{y}, \boldsymbol{\lambda}_2^e)}{P(d_k = -1|\mathbf{y}, \boldsymbol{\lambda}_2^e)} - \lambda_2^e(d_k)$$

$$= \ln \frac{\sum_{\mathcal{D}'} P\left(d_k = 1|\mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right) P\left(\mathbf{d}_{-k}^{(n)|\boldsymbol{\lambda}_2^e}\right)}{\sum_{\mathcal{D}'} P\left(d_k = -1|\mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right) P\left(\mathbf{d}_{-k}^{(n)}|\boldsymbol{\lambda}_2^e\right)} - \lambda_2^e(d_k)$$

Note that

$$P\left(d_k|\mathbf{d}_{-k}^{(n)}, \mathbf{y}, \boldsymbol{\lambda}_2^e\right) \propto P(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k)P(d_k)$$

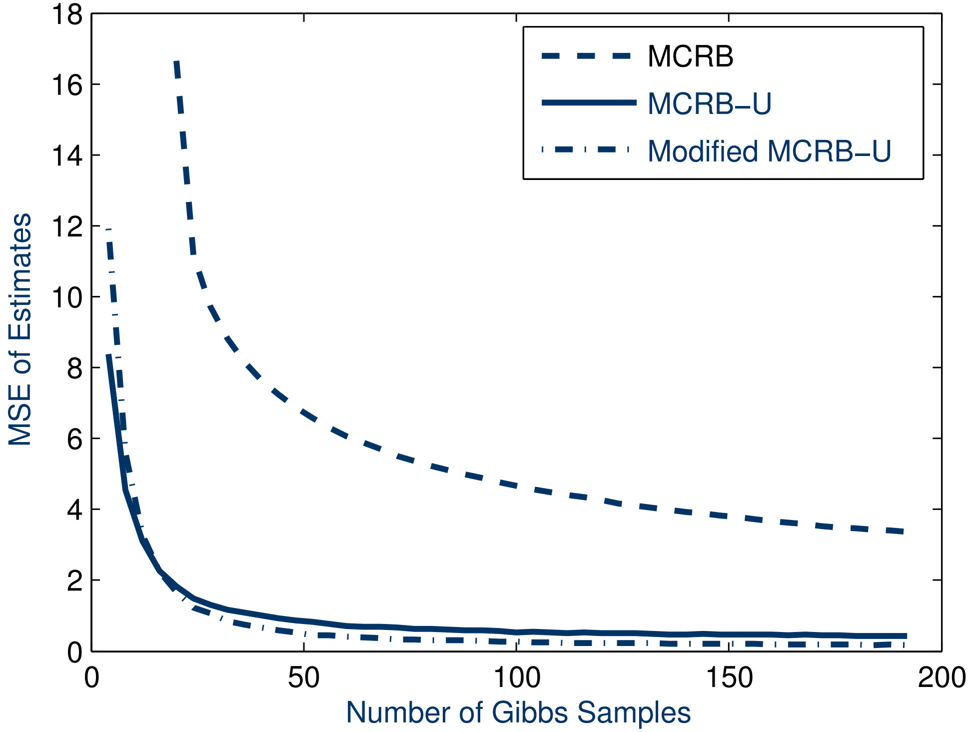
the extrinsic LLR can be computed as

$$\lambda_1^e(d_k) = \ln \frac{\sum_{\mathcal{D}'} P(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = 1)P(d_k^{(n)}|\boldsymbol{\lambda}_2^e)}{\sum_{\mathcal{D}'} P(\mathbf{y}|\mathbf{d}_{-k}^{(n)}, d_k = 1)P(d_k^{(n)}|\boldsymbol{\lambda}_2^e)}$$

– MCRB-U partially improves the efficacy of MCRB (Gibbs sampler) by the uniform procedure, but the samples ($\mathcal{D}$) so obtained remain fundamentally the same as those in MCRB
– To get set $\mathcal{D}^{'}$ from $\mathcal{D}$ requires storing all distinctive states through the Markov chain, and comparing them with the new state that the chain reaches. MaxLog algorithm can save the storage and (most) comparison operations by allowing a reasonable loss in power efficiency.

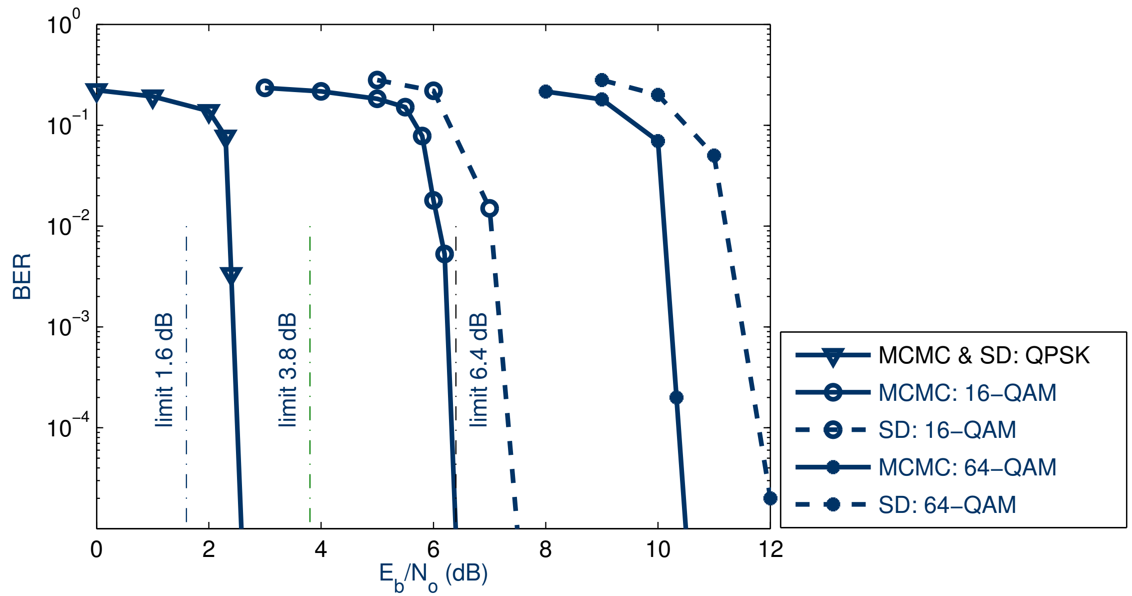
MSE Result of MCRB-U

- MIMO channel, $N_t = 4, N_r = 2$, QPSK
- MCRB: $N_s = \mathrm{NumIte} - N_b$, $N_b = 20$
- MCRB-U: $M = \mathrm{NumIte}/N_s$, $N_s = 4$
- SNR $= 8$ dB, $I(d_k, \lambda_2^e(d_k)) = 0.2$



BER Result of MCRB-U

- MIMO $N_t = 8, N_r = 8$
- rate-$\frac{1}{2}$ CC: $g(D) = \left[1 \ \frac{1+D^2}{1+D+D^2}\right]$
- QPSK: $M = 8, N_s = 5$
- 16-QAM: $M = 20, N_s = 20$
- 64-QAM: $M = 30, N_s = 30$



Modified MCRB-U

- Motivation: At high SNR, MCRB-U is observed to perform abnormally because the Markov chain is easily trapped in a state whose probability is a (*peaked*) local maximum

- Solutions: Modify the sampling distribution to increase the *mobility* of the Markov chain
  - Metropolized algorithms: slightly better than Gibbs sampler at high SNRs
  - Decrease the sampling SNR, weigh in favor of the FEC feedbacks

$$\lambda_1^{(n)}(d_k) = \frac{4}{N_0'}\mathbf{a}_k^T\left(\mathbf{y} - \mathbf{A}_{-k}\mathbf{d}_{-k}^{(n)}\right) + \lambda_2^e(d_k), \quad N_0' > N_0$$

- Modified MCRB-U lead the Markov chain to more important states within a finite number of steps, for the first few iterations of the turbo receiver

Modified MCRB-U

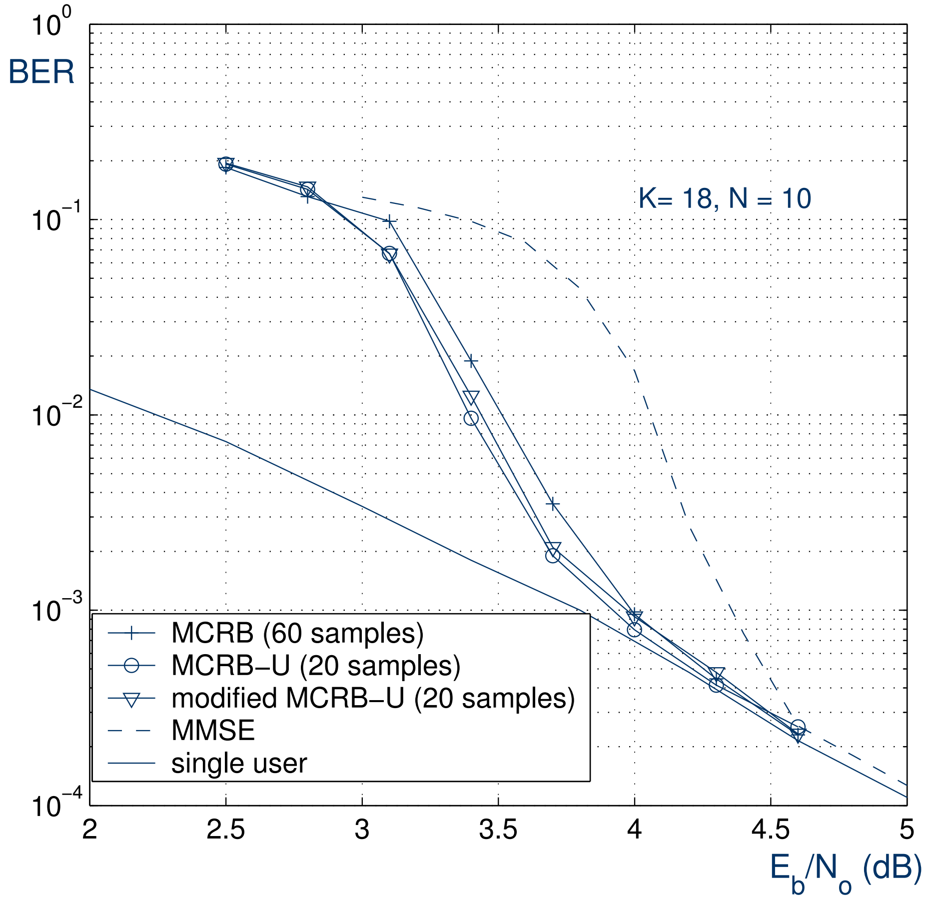
- Problem for reduced sampling SNR: the associated Markov chain converges to a different stationary distribution

$$P_{\mathrm{is}}\left(\mathbf{d}|\mathbf{y},\boldsymbol{\lambda}_2^e\right) = \exp\left(-\frac{1}{N_0'}\|\mathbf{y}-\mathbf{A}\mathbf{d}\|^2\right)P(\mathbf{d}|\boldsymbol{\lambda}_2^e)$$

- Solution: Increase the sampling SNR over turbo iterations, and set $N_0^{'} = N_0$ at the last iteration
  - Let $N_0^{'} \rightarrow N_0$ allows the Markov chain converge to the target distribution
  - The extrinsic feedback $\boldsymbol{\lambda}_2^e$ from FEC decoders become reliable as the turbo iterations proceed, and can direct the Markov chain to important states; here the priority is the *convergence*, not *mobility*

BER Simulations of Modified MCRB-U

- CDMA: $K = 18, N = 10$, BPSK
- Rate-$\frac{1}{2}$ CC: $g(D) = \left[1 \ \frac{1+D^2}{1+D+D^2}\right]$
- MCRB: $N_b = 30$, $N_s = 30$
- MCRB-U/modified MCRB-U: $M = 4, N_s = 5$
- First iteration: $N_0^{'} = 1.58N_0$; last iteration: $N_0^{'} = N_0$



BER Simulations of Modified MCRB-U

- CDMA: $K = 24, N = 10$, BPSK
- Rate-$\frac{1}{2}$ CC: $g(D) = \left[1 \ \frac{1+D^2}{1+D+D^2}\right]$
- MCRB: $N_b = 100$, $N_s = 100$, $M = 3$
- MCRB-U/modified MCRB-U:
  - 180 samples: $M = 6, N_s = 30$
  - 240 samples, $M = 8, N_s = 30$
- 1st iteration: $N_0^{'} = 2N_0$; 10th iteration: $N_0^{'} = N_0$