Efficient Markov Chain Monte Carlo Algorithms For MIMO and ISI channels

Rong-Hui Peng

2011/7/16

Summary of Past work

- Efficient MCMC algorithms for communication
  - Application to MIMO channel [Chen-Peng-Ashikhmin-Farhang]
  - Application to noncoherent channel [Chen-Peng]
  - Application to ISI channel and extend to underwater channel [Peng-Chen-Farhang]
  - Achieve near optimal performance with reduced complexity
  - Solve the slow convergence problem
- MIMO-HARQ schemes and combining algorithm [Peng-Chen]
  - Propose new retransmission scheme
  - Propose new combining algorithm
  - Increase the throughput significantly
  - Applicable to Wimax and LTE

Summary of Past work

- Nonbinary LDPC codes [Peng-Chen]
  - Nonbinary LDPC coded MIMO system, achieve good performance with reduced complexity
  - Code design based on EXIT chart
  - Hardware-friendly construction
    - Low encoding complexity
    - Parallel architecture
    - Low error floor
- Intern at MERL
  - Low complexity MIMO detection algorithm for MIMO-OFDMA
  - Wimax link-level simulation
  - 3G LTE antenna selection
  - 1 paper, 1 patent

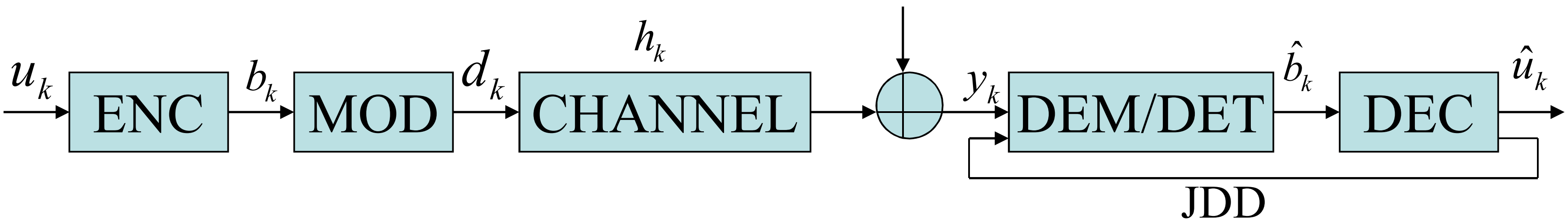
Summary of Past work

- Consultant at WiderNetworks
  - Synchronization and signal detection for 3G LTE driver scanner
- Work at SpiralGen
  - Vectorizing physical layer algorithm and map it to multiple-core processor
  - Implement physical layer software components in software communication architecture (SCA)

Outline

- Background and motivation
- Introduction to MCMC technology
- MCMC MIMO detection
- MCMC ISI equalization
- Conclusion

Background and motivation



$$\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{n}$$

MAP detection:

$$\hat{b}_k = \max P(b_k \mid \mathbf{y}, \mathbf{H})$$

$$P(b_k \mid \mathbf{y}, \mathbf{H}) = \sum_{\mathbf{b}_{-k}} p(\mathbf{Y} \mid \mathbf{H}, \mathbf{d}) P(\mathbf{d})$$

MAP detector : Optimal performance but exponential complexity

↓

Reduction in computational cost is essential for application to future system.

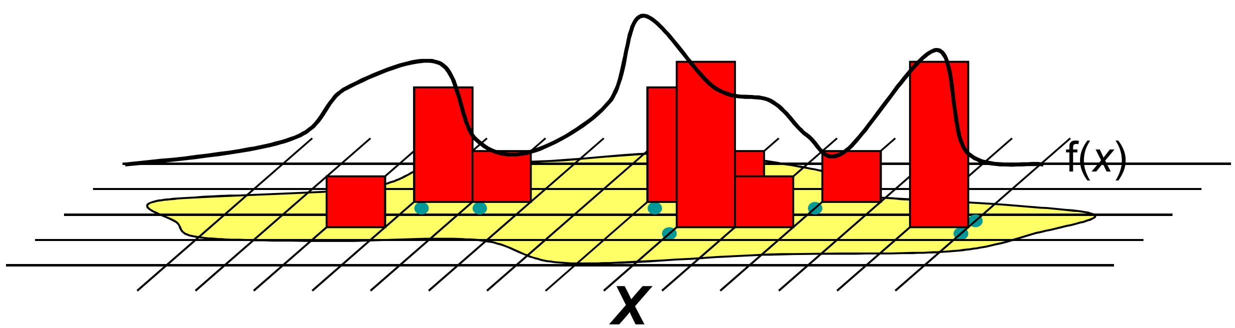
Our contribution

- Previous work on MCMC methods for signal processing and communication [Doucet-Wang'05]
  - Bit-counting: use MCMC to determine the frequency over which a bit occurs.
  - Many samples are needed.
  - Requires a burning period to allow Markov chain to converge.
  - Do not consider convergence problem

- Our proposed MCMC methods
  - Do not use bit-counting
  - No burning period is needed
  - Fewer samples even for large systems
  - Solve the slow-convergence problem

Markov Chain Monte Carlo (MCMC)

- MCMC obtains the statistical inference by sampling from a posterior distribution through Markov chain
- MCMC is suitable for addressing problems involving high-dimensional summations or integrals
- Instead of evaluating all summation terms (exponential complexity), *average* over the samples from the complex distribution



Gibbs sampler

- A MCMC method sampling from a multivariate distribution

Idea: Sample from the conditional of each variable given the settings of the other variables

Repeatedly:

1) pick i (either at random or in turn)

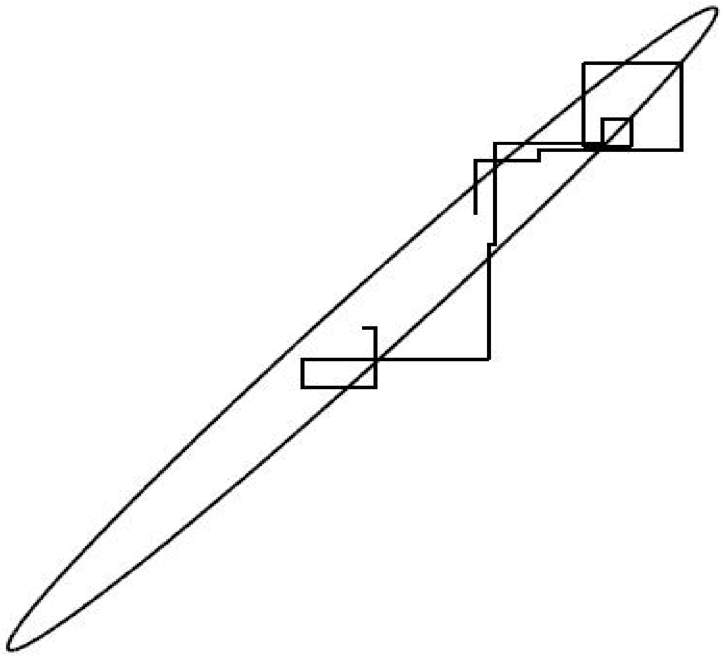
2) replace x_i by a sample from the conditic distribution

$$p(x_i \mid x_1, \Lambda, x_{i-1}, x_{i+1}, \Lambda, x_n)$$

Gibbs sampling is feasible if it is easy to s from the conditional probabilities.

This creates a Markov Chain

$$\mathbf{x}^{(1)} \rightarrow \mathbf{x}^{(2)} \rightarrow \mathbf{x}^{(3)} \Lambda$$



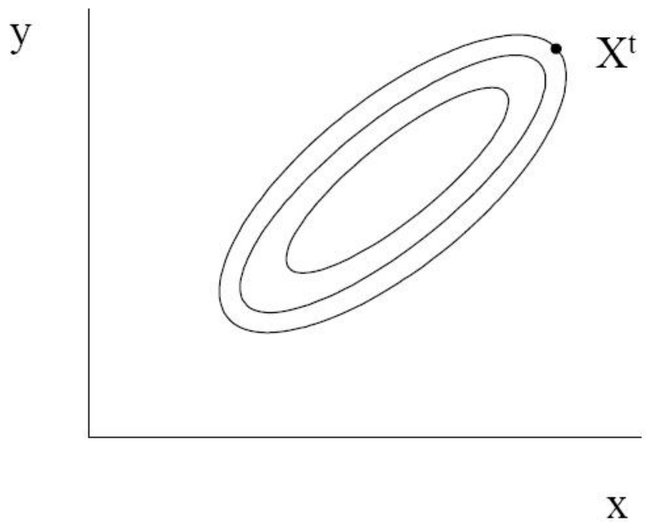
Example: 20 iterations of Gibbs sampling on a bivariate Guassian

Why Gibbs sampling works

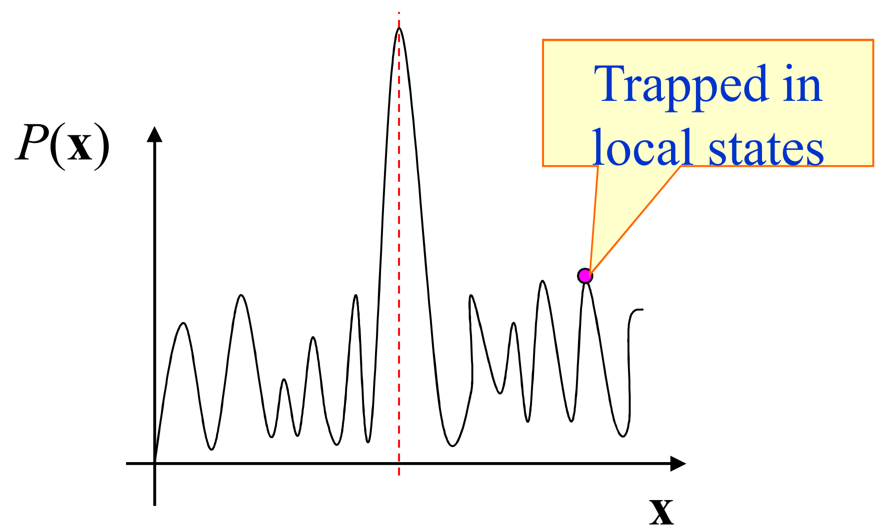
- Retains elements of the greedy approach
  - weighing by conditional PDF makes likely to move towards locally better solutions
- Allows for locally bad moves with a small probability, to escape local maxima

Limitations

- May have slow convergence problem
  - High posterior correlation, hard to move in other directions
  - Multi-modal distribution, proposing small changes causes the move between modes to become rare

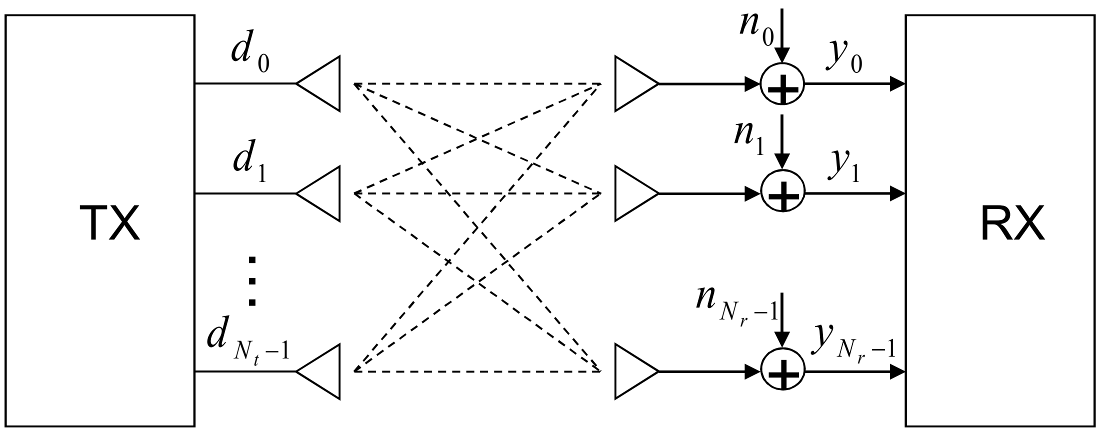


High posterior correlation



Multi-model distribution

MIMO channel



$$\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{n}$$

where

$\mathbf{d} \in \mathbf{C}^{N_t}$ is a vector of transmit symbols

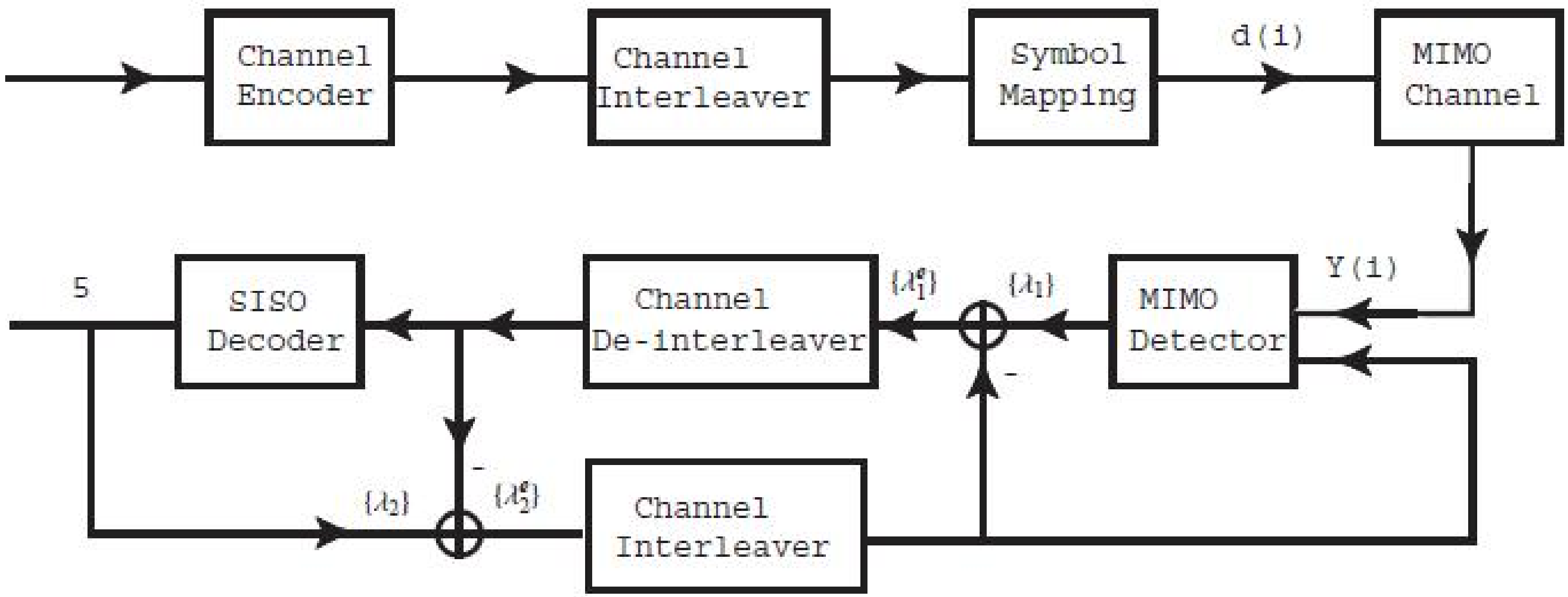
$\mathbf{y} \in \mathbf{C}^{N_r}$ is a vector of received signal

$\mathbf{H} \in \mathbf{C}^{N_r \times N_t}$ is the channel gain matrix

$\mathbf{n} \in \mathbf{C}^{N_r}$ is an additive noise vector

N_t, N_r is the number of tx and rx antennas

System Diagram of MIMO system



SDD: Separate detection and decoding

IDD : Iterative detection and decoding

MIMO detection

- Maximal likelihood (ML) detection or Maximum *a posteriori* (MAP), optimal but exponential complexity
- Low complexity sub-optimal detectors (ZF, MMSE, VBLAST), performance gap can be 20 dB
- Tree search based (sphere decoding (SD), QRD-M)
  - Excellent performance at high SNR
  - SD has variable complexity depending channel condition
  - Exponential complexity at low SNR
  - Complexity increase quickly with large system
- Markov chain Monte Carlo (MCMC)
  - Constant complexity, not increased much for large system
  - Excellent performance at low SNR
  - High SNR problem

Low complexity detection

- Approximate optimal detectors

$$\lambda_k = \ln \frac{P(b_k = +1 \mid \mathbf{y})}{P(b_k = -1 \mid \mathbf{y})} = \ln \frac{\sum_{\mathbf{b}_{-k}} P(b_k = +1, \mathbf{b}_{-k} \mid \mathbf{y})}{\sum_{\mathbf{b}_{-k}} P(b_k = -1, \mathbf{b}_{-k} \mid \mathbf{y})}$$

$$\approx \ln \frac{\max_{\mathbf{b}_{-k}} P(b_k = +1, \mathbf{b}_{-k} \mid \mathbf{y})}{\max_{\mathbf{b}_{-k}} P(b_k = -1, \mathbf{b}_{-k} \mid \mathbf{y})} \approx \ln \frac{\max_{\mathtt{I}_{-k}} P(b_k = +1, \mathbf{b}_{-k} \mid \mathbf{y})}{\max_{\mathtt{I}_{-k}} P(b_k = -1, \mathbf{b}_{-k} \mid \mathbf{y})}$$

where $\mathtt{I}_{-k} \subset \mathbf{b}_{-k}$

The searching is performed over a small subset (*Important set*) instead of a large full set.

MCMC MIMO detector

- MCMC detector finds important sets using Markov chain Monte Carlo
  - Create Markov chain with state space: **d** and stationary distribution P(**d**|**Y**)
  - Run Markov chain using Gibbs sampler
  - After Markov chain converge, the samples are generated according to P(**d**|**Y**). Those samples with large P(**d**|**Y**) generated with high probabilities
- The samples generated by Gibbs sampler are used to compute the soft message for soft decoder

MCMC MIMO detector (bit-wise)

S1. Generate initial $\mathbf{b}^{(0)}$

S2. FOR $n = 1$ to L

$$\text{draw } b_0^{(n)} \sim p(b_0 = b \mid b_1^{(n-1)}, \cdots, b_{N_t M_c - 1}^{(n-1)}, \mathbf{y})$$

$$\text{draw } b_1^{(n)} \sim p(b_1 = b \mid b_0^{(n)}, b_2^{(n-1)}, \cdots, b_{N_t M_c - 1}^{(n-1)}, \mathbf{y})$$

$$\vdots$$

$$\text{draw } b_{N_t M_c - 1}^{(n)} \sim p(b_1 = b \mid b_0^{(n)}, \cdots, b_{N_t M_c - 2}^{(n)}, \mathbf{y})$$

Add $\mathbf{b}^{(n)}$ to important set $\mathcal{I}$

END FOR

S3. Calculate LLR

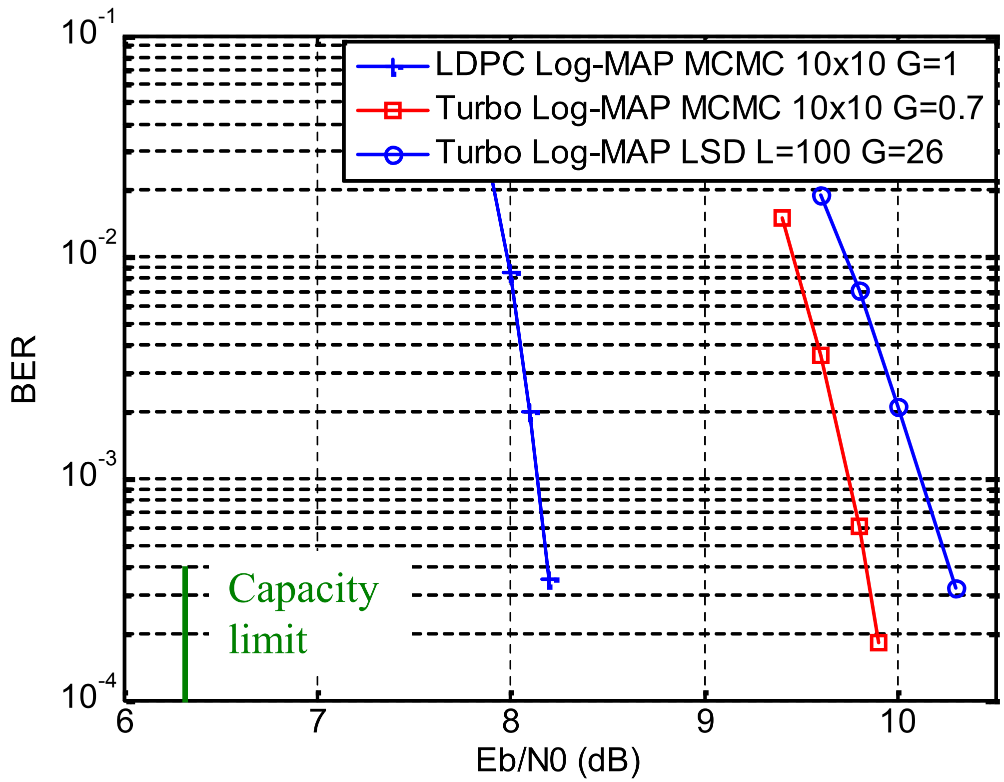
$$\lambda_k = \ln \frac{\max\limits_{\mathcal{I}_{-k}} P(b_k = +1, \mathbf{b}_{-k} \mid \mathbf{y})}{\max\limits_{\mathcal{I}_{-k}} P(b_k = -1, \mathbf{b}_{-k} \mid \mathbf{y})}$$

MCMC detector

- Calculate transition probability

$$
\begin{aligned}
& p(b_i = b \mid b_0^{(n)}, \cdots b_{i-1}^{(n)}, b_{i+1}^{(n-1)}, \cdots, b_{N_t M_c - 1}^{(n-1)}, \mathbf{y}) \\
\propto\ & p(\mathbf{y} \mid b_0^{(n)}, \cdots b_{i-1}^{(n)}, b_i = b, b_{i+1}^{(n-1)}, \cdots, b_{N_t M_c - 1}^{(n-1)}) p(b_i = b) \\
=\ & p(\mathbf{y} \mid \mathbf{d}) p(b_i = b)
\end{aligned}
$$

Simulation results (Low SNR)

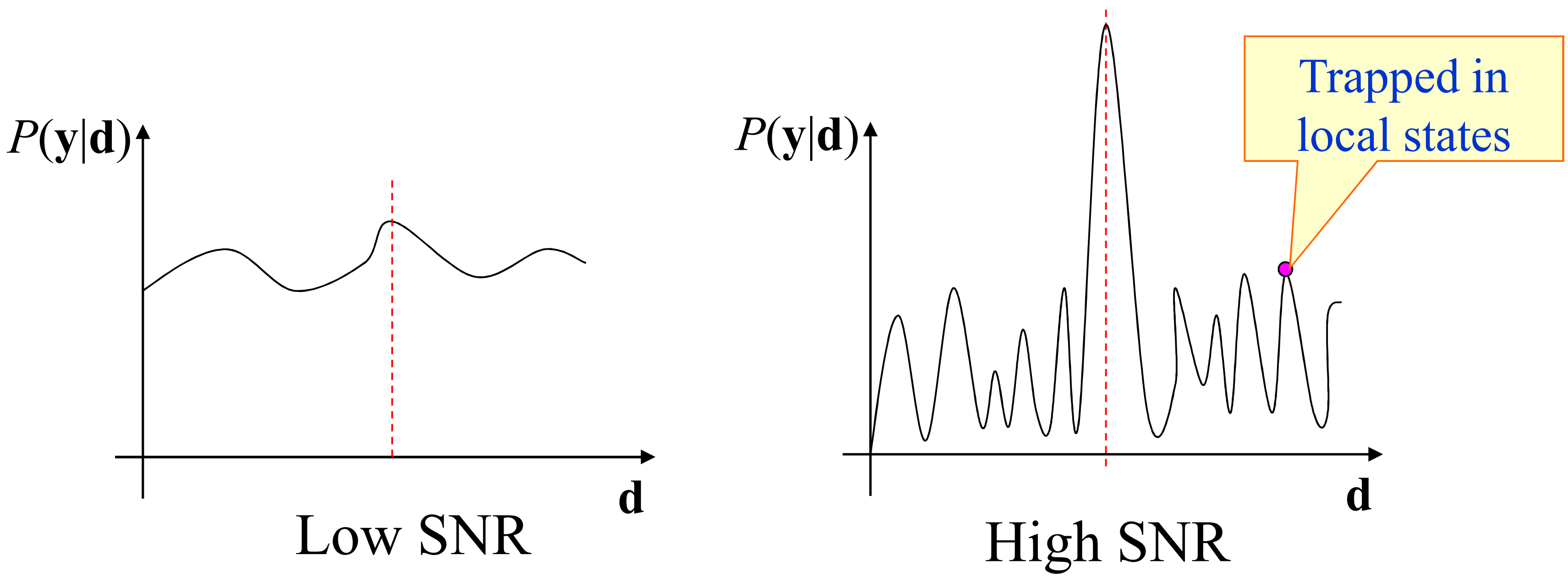


Performance of turbo and LDPC coded TX8 64QAM systems, code length 18432

- MCMC with LDPC performs best in performance and complexity
- Within 1.8 dB of capacity at 24 bits/channel use
- Turbo LSD is 25 times higher in running time

High SNR problem

- At high SNR, MCMC takes long time to converge, leads performance degradation, this is because of the multimodal property of channel PDF at high SNR



Solutions

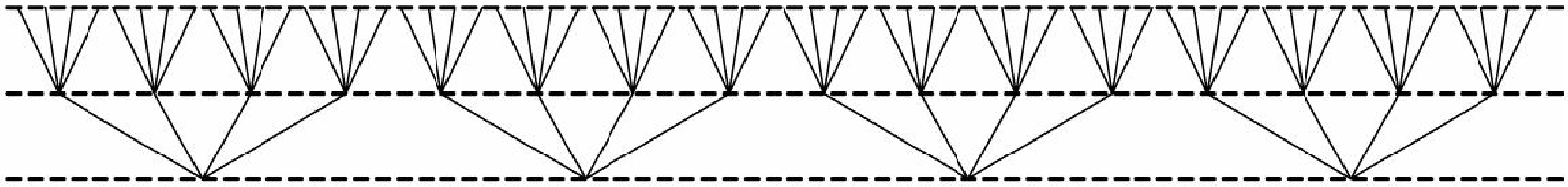
- Run *L* independent parallel Gibbs samplers and each Gibbs sampler run *I* iterations
- Generate good initial candidates using other low complexity detector (warm start)
  - MMSE/ZF
- Hybrid QRD-MCMC

QRD-M

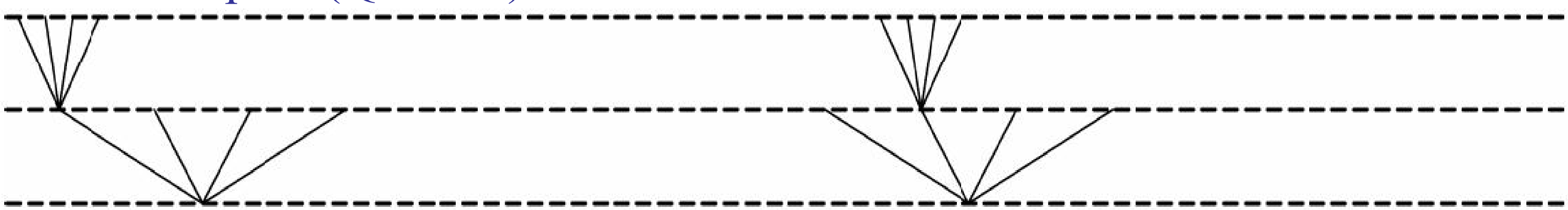
- QR decomposition

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \mathbf{Q}^H \mathbf{y} = \mathbf{R}\mathbf{d} + \mathbf{n}' = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & r_{1,4} \\ 0 & r_{2,2} & r_{2,3} & r_{2,4} \\ 0 & 0 & r_{3,3} & r_{3,4} \\ 0 & 0 & 0 & r_{4,4} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} + \begin{bmatrix} n_1' \\ n_2' \\ n_3' \\ n_4' \end{bmatrix}$$

Full path (MLD)



Select M path (QRD-M)

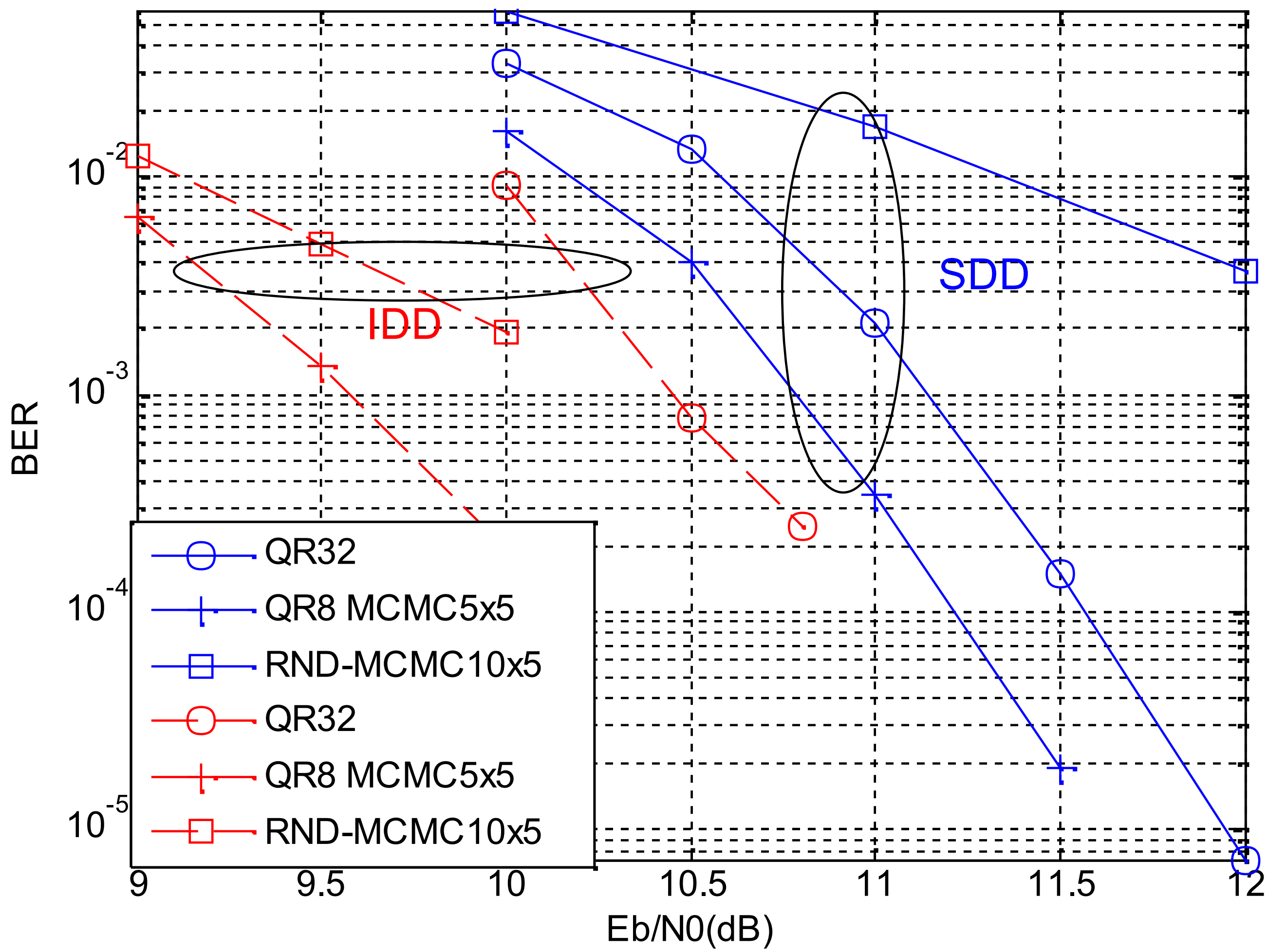


QPSK, M=1

Hybrid QRD-MCMC

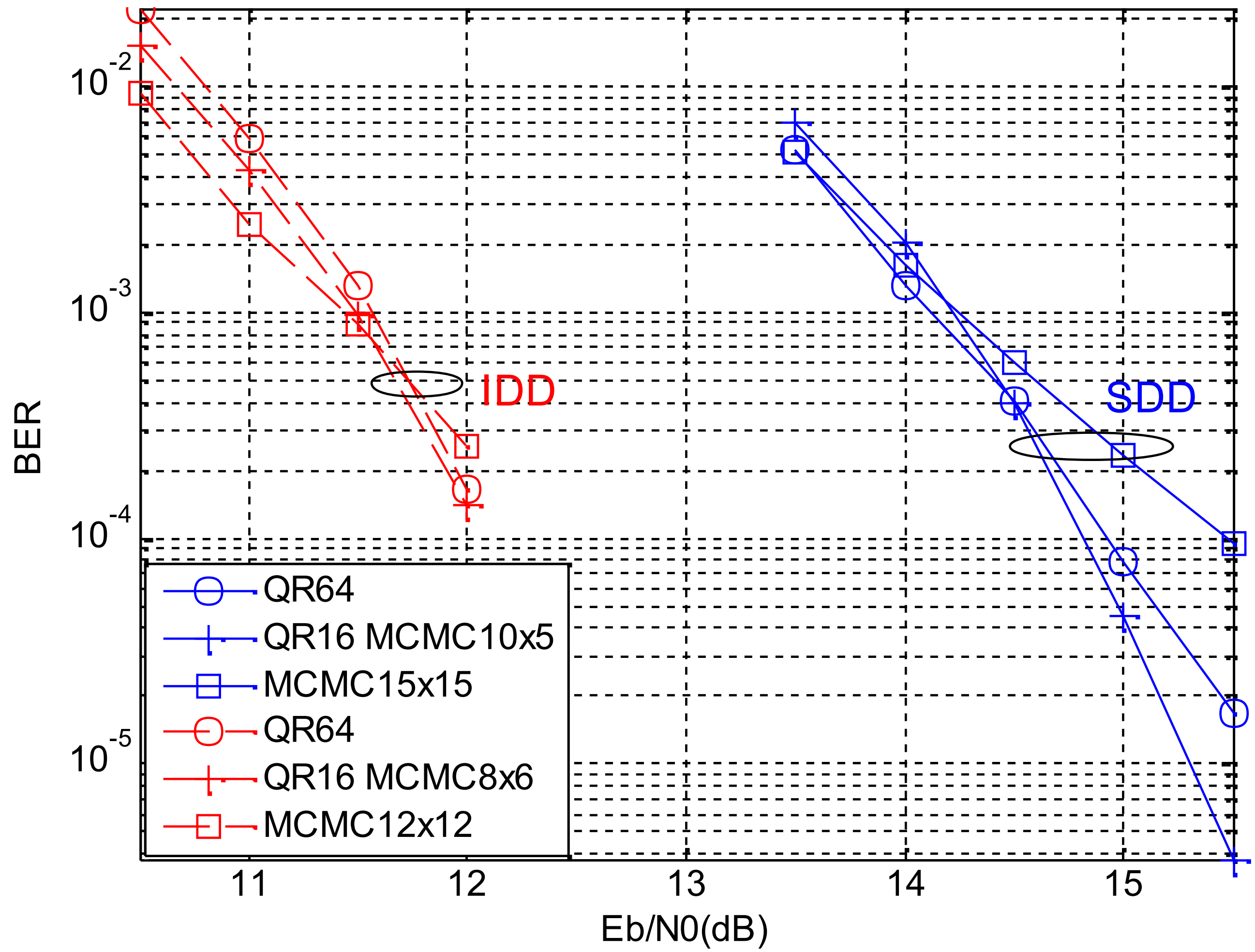
- Combine QRD-M and MCMC
  - A QRD-M with a small M is running first to generate initial important sets
  - The bit sequence with minimal path metric will be used to initialize one of L parallel MCMC
  - The important set produced by the QRD-M detector is incorporated by the MCMC detector
  - MCMC is running to generate refined important set
  - The soft message is computed using refined important set

Simulation Results



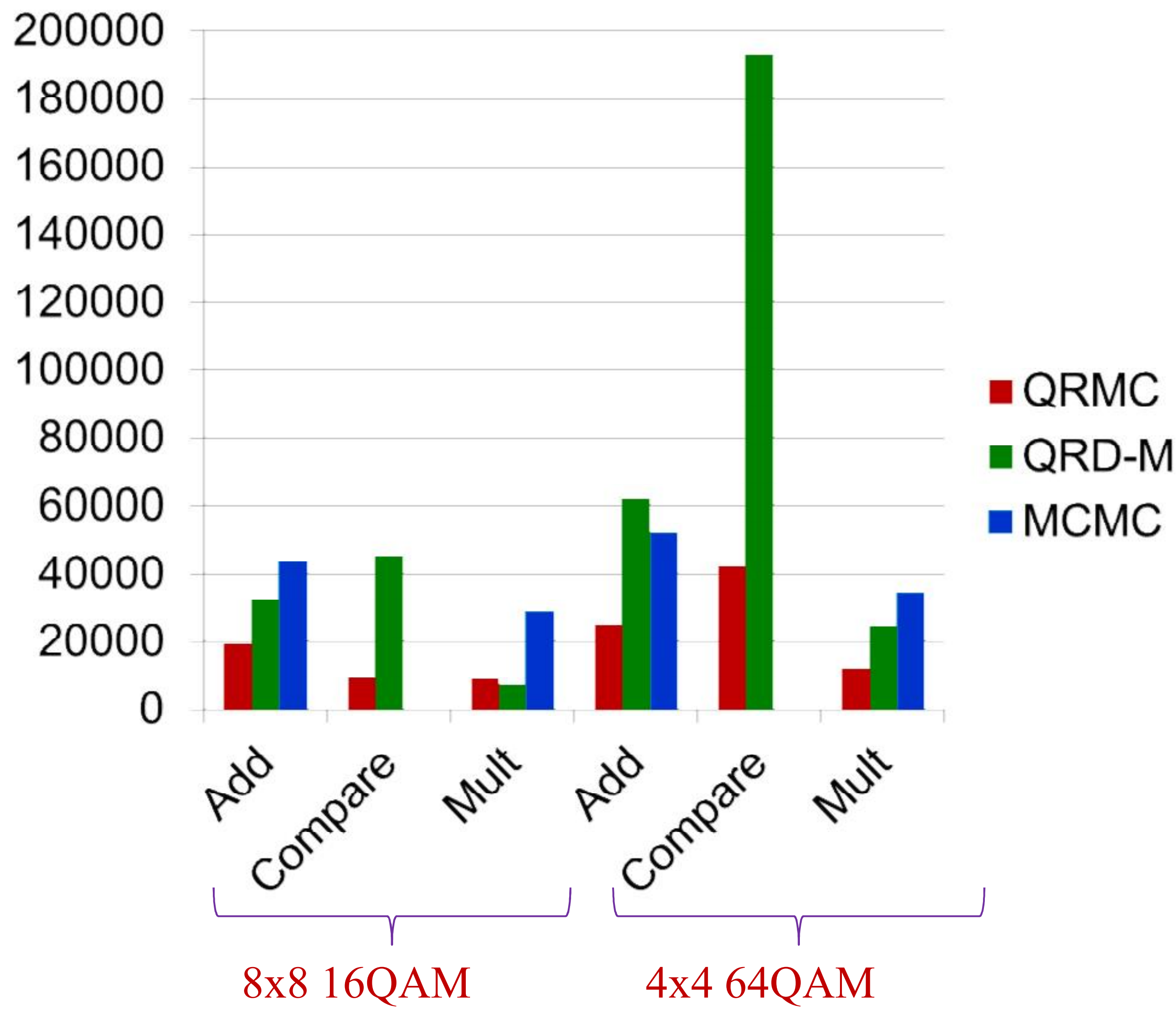
8x8 16QAM LDPC coded SDD and IDD systems, code length 2304

Simulation Results



4x4 64QAM LDPC coded SDD and IDD systems, code length 2304

Complexity


200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
Add
Compare
Mult
Add
Compare
Mult
QRMC
QRD-M
MCMC
8x8 16QAM
4x4 64QAM

Summary of results

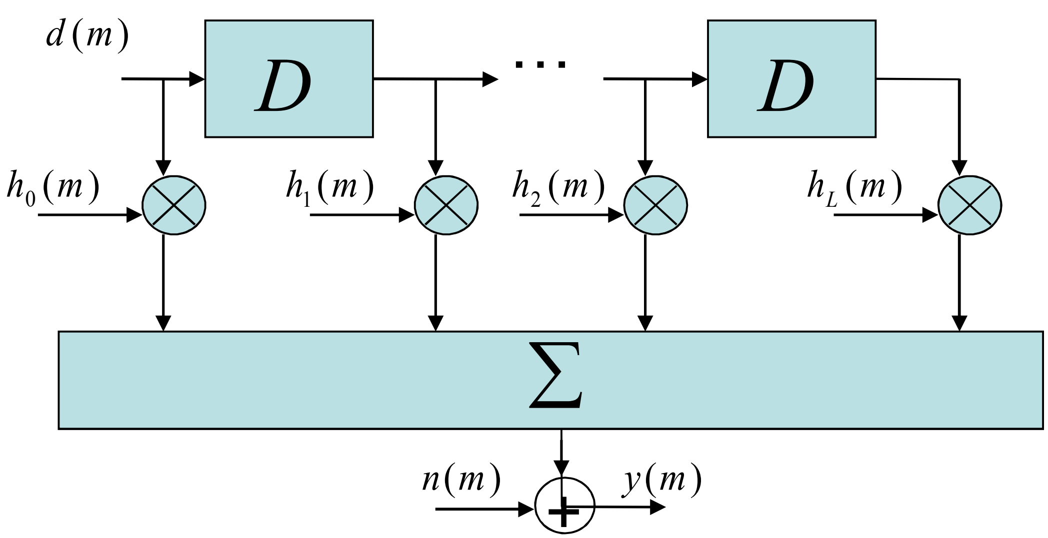
- MCMC detector for MIMO channel is studied
- Low complexity for large system
- Excellent performance at low SNR region
- High SNR problem can be solved by combining QRD-M with MCMC without increasing complexity

ISI channel

- Multipath fading channel

$$y(m) = \sum_{l=0}^{L} h_l(m) d(m-l) + n(m),$$

$$\text{for } m = 0, 1, \ldots, N-1+L$$



ISI channel

- In matrix form

$$\mathbf{y} = \mathbf{H}\mathbf{d} + \mathbf{n}$$

where

$\mathbf{d} \in \mathbf{C}^{N}$: transmitting vector

$\mathbf{y} \in \mathbf{C}^{N+L-1}$: received vector

$\mathbf{H} \in \mathbf{C}^{N+L-1 \times N}$: channel matrix

$\mathbf{n} \in \mathbf{C}^{N+L-1}$: dditive noise vector

N : Equalizer block length

L : Channel memory

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & \cdots & 0 & \cdots & & 0 \\ h_1 & h_0 & 0 & \cdots & & & 0 \\ \vdots & \ddots & & \ddots & & & \vdots \\ 0 & h_L & \cdots & h_0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & & \vdots \\ 0 & \cdots & 0 & h_L & h_{L-1} & \cdots & h_0 \\ 0 & \cdots & & 0 & \ddots & & \vdots \\ 0 & \cdots & & & 0 & h_L & h_{L-1} \\ 0 & \cdots & & \cdots & & 0 & h_L \end{bmatrix}$$

Detection for ISI channel

- MAP detection is still optimal
  - Efficient implementation of MAP detection is Forward backward algorithm (BCJR)
  - Still exponential complexity with channel memory
- Low complexity algorithm
  - MMSE
  - Decision-feedback
  - …

Bit-wise MCMC detector

- Transition probability

$$P(b_k = a \mid \overline{\mathbf{b}}_k, \mathbf{y}) \propto p(\mathbf{y} \mid \mathbf{b}^a) P(b_k = a)$$

$$= p(\mathbf{y} \mid \mathbf{d}^a) P(b_k = a)$$

$$= \prod_{j=0}^{N+L-1} p(y_j \mid \mathbf{d}^a_{j-L:j}) P(b_k = a)$$

$$= \boxed{\left\{ \prod_{j=0}^{i-1} p(y_j \mid \mathbf{d}^a_{j-L:j}) \prod_{j=i+L+1}^{N+L-1} p(y_j \mid \mathbf{d}^a_{j-L:j}) \right\}} \prod_{j=i}^{i+L} p(y_j \mid \mathbf{d}^a_{j-L:j}) P(b_k = a)$$

$$= C \cdot \prod_{j=i}^{i+L} p(y_j \mid \mathbf{d}^a_{j-L:j}) P(b_k = a)$$

$$\mathbf{b}^a = \{ b_0^{(n)}, \cdots, b_{k-1}^{(n)}, a, b_{k+1}^{(n-1)}, \cdots, b_{M_b N-1}^{(n-1)} \},$$

$\mathbf{d}^a$ denote the symbol vector corresponding to $\mathbf{b}^a$

b_k is mapped to d_i

Bit-wise MCMC detector

- Computing the a posteriori LLR
  - Accurate posteriori LLR involve computations over the whole block (may be very cumbersome since *N* can be very large)
  - Truncated window to approximate

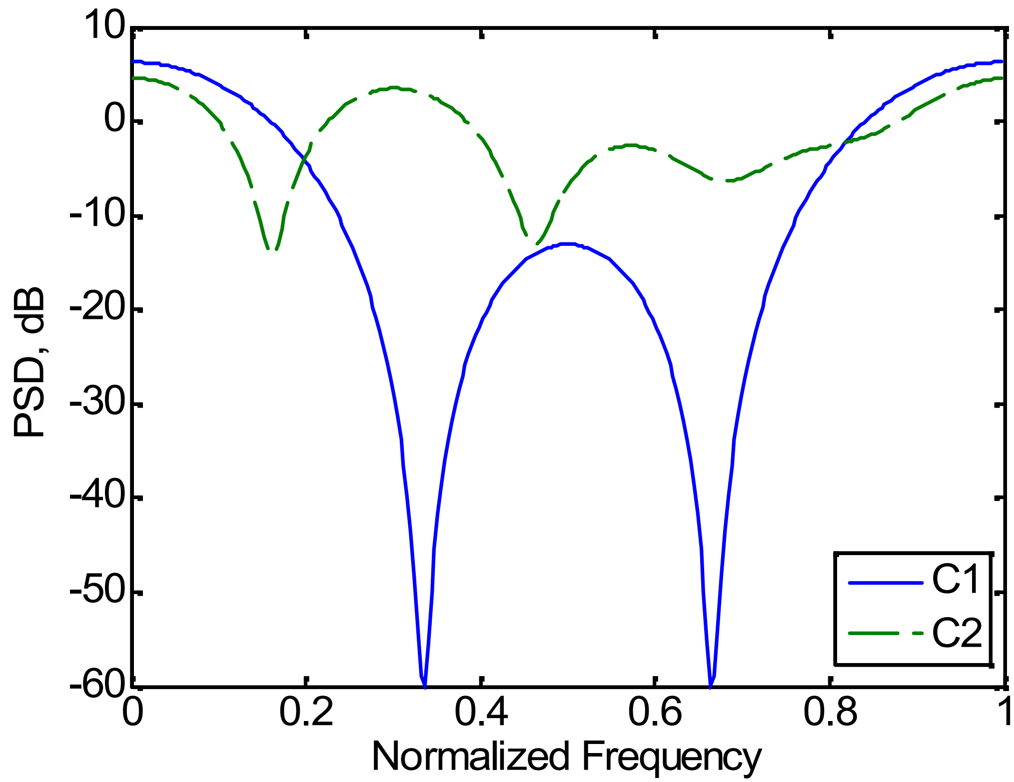
$$\lambda_{1,k}^{e} = \ln \frac{\sum_{\mathbf{b}_{i_1:i_2} \in \mathrm{I}_{i_1:i_2}^{k,0}} p(\mathbf{y}_{i:i+L} \mid \mathbf{b}_{i_1:i_2}) \prod_{l=i_1}^{i_2} P(b_l)}{\sum_{\mathbf{b}_{i_1:i_2} \in \mathrm{I}_{i_1:i_2}^{k,1}} p(\mathbf{y}_{i:i+L} \mid \mathbf{b}_{i_1:i_2}) \prod_{l=i_1}^{i_2} P(b_l)} - \lambda_{2,k}^{e}$$

$\mathrm{I}_{i_1:i_2}$ the set which truncate sequences in I

with bits $\{b_k, i_1 \le k \le i_2\}$

Severe ISI channel

- For channels with severe ISI, bit-wise MCMC suffers from slow convergence problem
- We found slow convergence problem for ISI channel is mainly caused by high posterior correlation



$$h_1[n] = 0.227\delta[n] + 0.46\delta[n-1] + 0.688\delta[n-2] + 0.46\delta[n-3] + 0.227\delta[n-4]$$

$$h_2[n] = (2+04i)\delta[n] + (15+18i)\delta[n-1] + \delta[n-2] + (12-13i)\delta[n-3] + (08+16)\delta[n-4]$$

Severe ISI channel

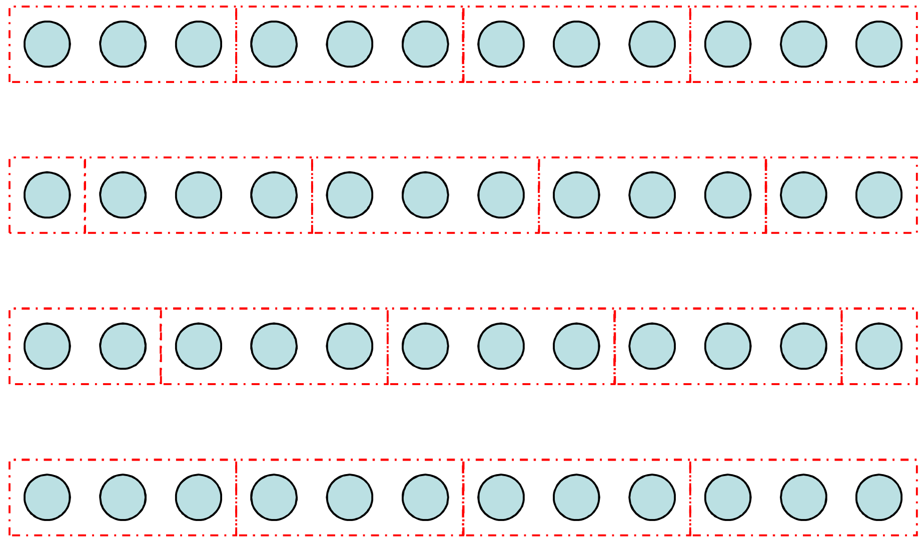
- Our solution
  - Grouping (blocking) the highly correlation variables in Gibbs sampler

1st iteration:

2nd iteration:

3rd iteration:

4th iteration:



Shift grouping different adjacent symbols over iterations to speed up the mixing rate of Gibbs sampler

Transition probability

- Grouping multiple variables allow a large sample space containing $r = 2^{M_b G}$ values

$$
\begin{bmatrix} y_0 \\ \vdots \\ y_{i-1} \\ y_i \\ \vdots \\ y_{i+P} \\ y_{i+P+1} \\ \vdots \\ y_{N+L-1} \end{bmatrix}
=
\begin{bmatrix}
h_0 & 0 & \cdots & 0 & \cdots & & 0 \\
h_1 & h_0 & 0 & \cdots & & & 0 \\
\vdots & \ddots & & \ddots & & & \vdots \\
0 & h_L & \cdots & h_0 & 0 & \cdots & 0 \\
\vdots & & \ddots & & \ddots & & \vdots \\
0 & \cdots & 0 & h_L & h_{L-1} & \cdots & h_0 \\
0 & \cdots & & 0 & \ddots & & \vdots \\
0 & \cdots & & & 0 & h_L & h_{L-1} \\
0 & \cdots & & \cdots & & 0 & h_L
\end{bmatrix}
\begin{bmatrix} d_0 \\ \vdots \\ d_{i-1} \\ d_i \\ \vdots \\ d_{i+G} \\ d_{i+G+1} \\ \vdots \\ d_{N-1} \end{bmatrix}
+
\begin{bmatrix} n_0 \\ \vdots \\ n_{i-1} \\ n_i \\ \vdots \\ n_{i+P} \\ n_{i+P+1} \\ \vdots \\ n_{N+L-1} \end{bmatrix}
$$

$$P = G + L + 1$$

Transition probability

$$\tilde{\mathbf{y}}_{i:i+P} = \mathbf{y}_{i:i+P} - \mathbf{H}_{i:i+P}^{i-L:i-1}\mathbf{d}_{i-L:i-1} - \mathbf{H}_{i:i+P}^{i+G:i+P}\mathbf{d}_{i+G:i+P} + \mathbf{n}_{i:i+P}$$

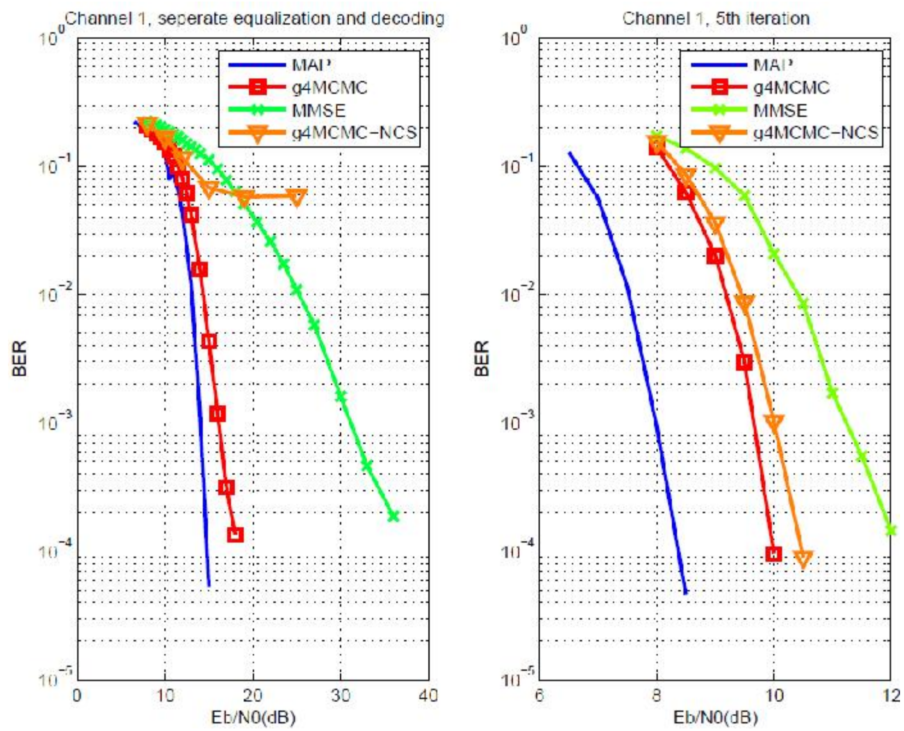
$$\boxed{= \mathbf{H}_{i:i+P}^{i+G:i+P}\mathbf{d}_{i:i+G-1} + \mathbf{n}_{i:i+P}}$$

- Can be considered as received signals of a MIMO channel with *G* transmit antenna *and P*+1 receive antenna
- QRD-M is applied to find $r_0 < r$ important states with large APPs
- Gibbs sampler randomly jump to one state from those important states

Simulation results

- A channel with strong ISI

$$h[n] = 0.227\delta[n] + 0.46\delta[n-1] + 0.688\delta[n-2] + 0.46\delta[n-3] + 0.227\delta[n-4]$$



Performance comparison of MAP, MMSE, g-MCMC equalizer for a strong ISI channel

- g-MCMC significantly outperforms MMSE equalizer (Tuchler-Singer-Koetter'02)
- b-MCMC does not work for such channel with strong ISI

Summary of results

- MCMC equalizer for ISI channel is studied
- Near optimal performance can be obtained
- Slow convergence is mainly caused by high posterior correlation
- QRD-M algorithm is applied to reduce the complexity of group-wise MCMC equalizer
- Parallel implementation is proposed

Conclusions

- Efficient MCMC algorithms for MIMO and ISI channel are studied
  - For MIMO, MCMC works well at low SNR region
  - For ISI, MCMC works well for moderate ISI
- Solutions for slow convergence are proposed
  - For MIMO, use QRD-M as good start point
  - For ISI,
    - Use group-wise Gibbs sampler to group high correlated variables
    - Use QRD-M to reduce the complexity of group-wise Gibbs sampler
- Future work
  - Study the sensitivity of proposed MIMO detector with more practical channel model
  - Study the time-varing ISI channel and adaptive grouping scheme for group-wise MCMC
  - Hardware implementation of MCMC

Thank you !