Multiuser Detection of M-QAM Symbols via Bit-Level Equalization and Soft Detection

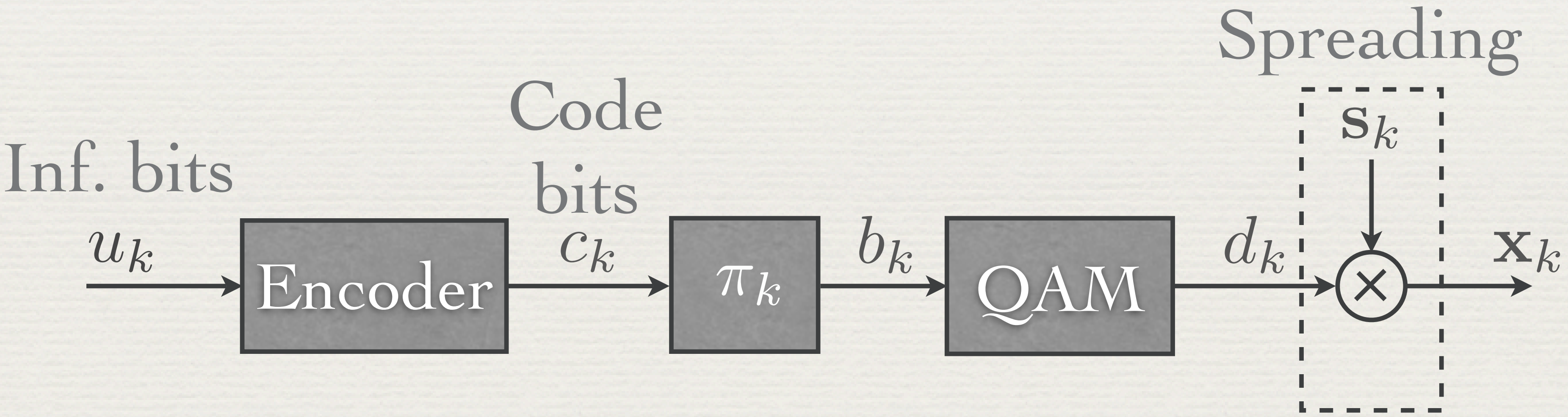
Darryl Lin and Teng Joon Lim
Department of Electrical & Computer Engineering
University of Toronto

Outline

- Multi-User Detection With Prior Knowledge
  - Needed for turbo multi-user detection
  - A number of heuristic techniques
- Our proposal:
  - Use variational inference as unifying design tool
  - Several turbo MUD methods can be derived like this
  - Can also deal with Gray-coded M-QAM in a natural way

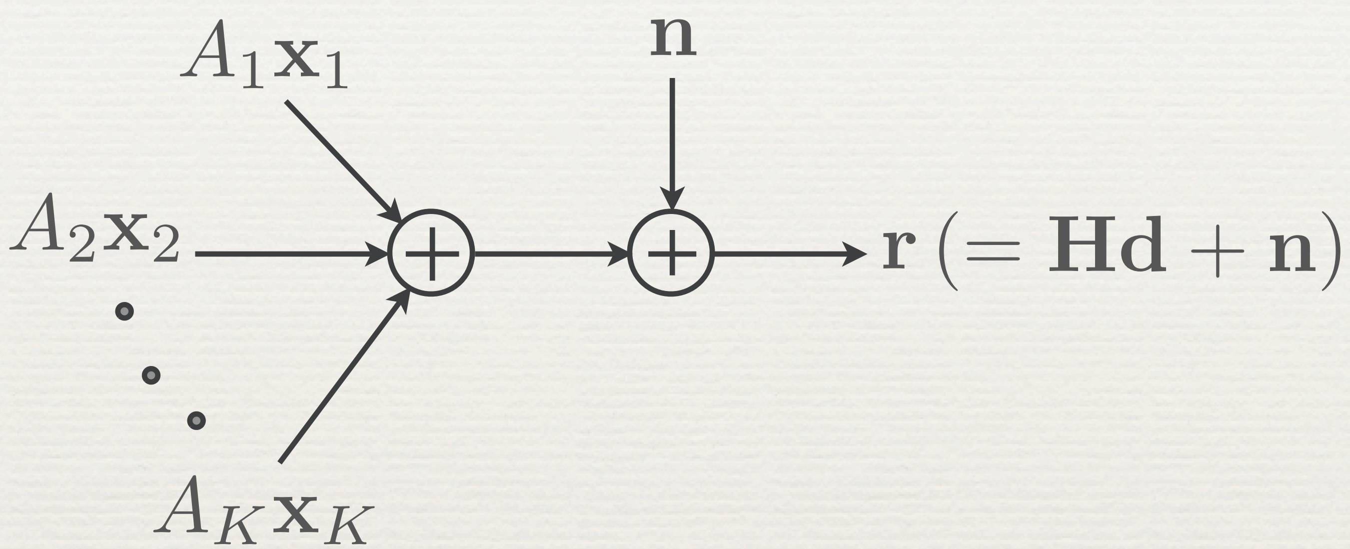
Decoding in an Interference Channel

- At tx'er: Coding - Interleaving - QAM bit-to-symbol mapping - Channelization (CDMA)



User k's transmitter

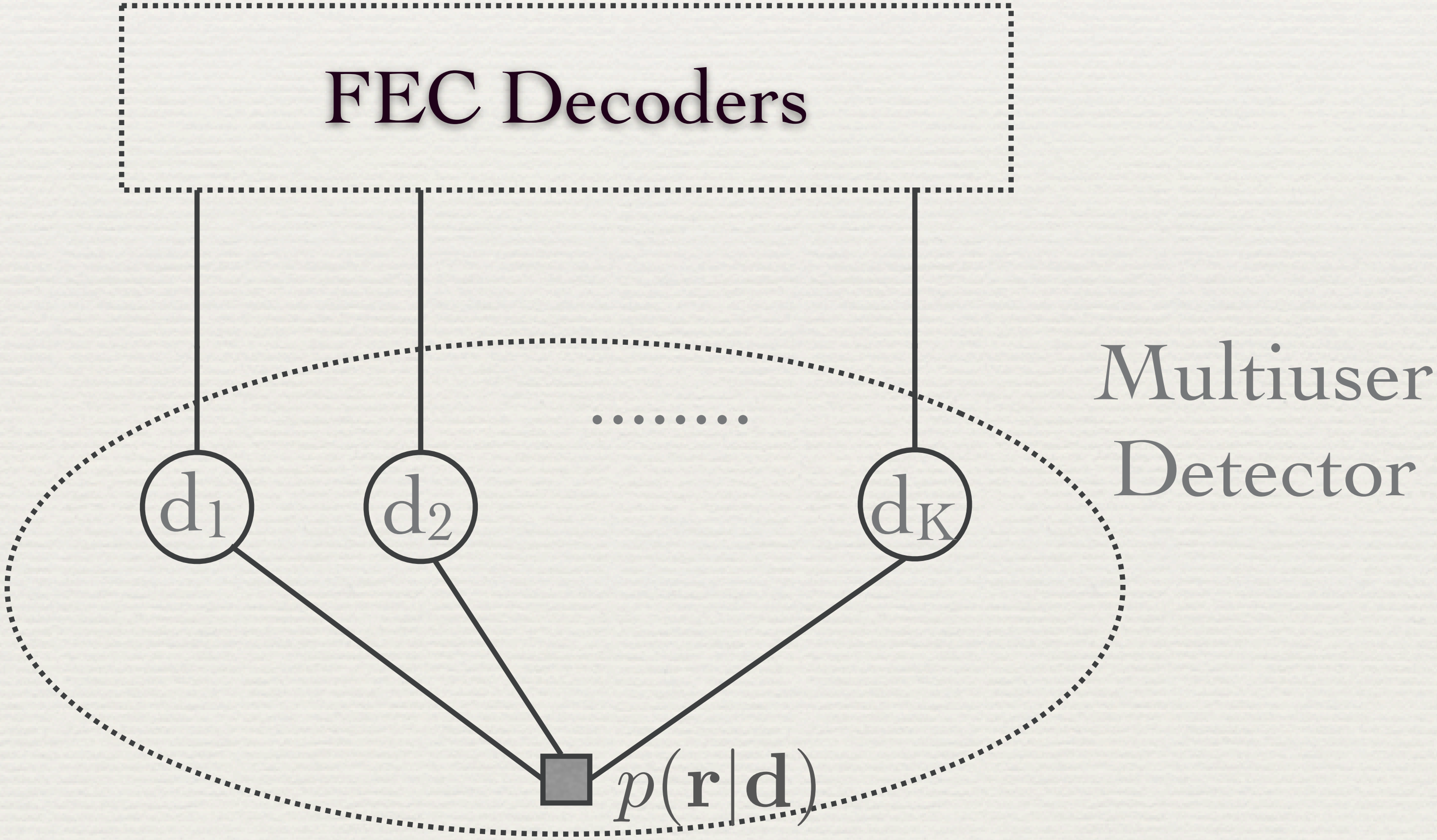
Decoding in an Interference Channel



- Channel: Multi-access interference, unfaded.
- At rx'er: Iterative decoding and multi-user detection.
  - Optimal decoding and detection too complex

Iterative Decoding Basics

"Straightforward" iterative decoding by sum-product algorithm requires full-blown APP updates in multi-user channel.



d_k = k-th user's channel symbol

Iterative Decoding Basics

- Need sub-optimal MUD to generate approximate posterior symbol probabilities.
- For example, use Wang/Poor's LMMSE-based detector:
  - Compute LMMSE filter output per iteration per user, $\hat{d}_k$.
  - Assume $\hat{d}_k = \mu_k d_k + \eta_k$, where η_k is Gaussian. With known channels and AWGN variance, both μ_k and variance of η_k can be found.

Iterative Decoding Basics

- Then APP of symbol d_k can be found, assuming
$$P(d_k|\mathbf{r}) = P(d_k|\hat{d}_k)$$
- From symbol APP, bit APPs can be found by summing over 2^{L-1} terms, in 2^L-ary modulation:
$$P(b_k^1 = 0|\mathbf{r}) = \sum_{d_k : b_k^1 = 0} P(d_k|\mathbf{r})$$
- Other turbo MUDs for M-ary modulation can be defined, using similar heuristic assumptions.
  - E.g. interference cancellation.

Variational Inference As A Unified Approach

- Variational Inference approximates the posterior distribution $p(\mathbf{d}|\mathbf{r})$ with a simpler one $Q(\mathbf{d})$
- The parameters (mean, variance, etc.) of the Q function are chosen to minimize the KL divergence b/w $Q(\mathbf{d})$ and $p(\mathbf{d}|\mathbf{r})$:

$$F(\boldsymbol{\lambda}) = \int Q(\mathbf{d}) \log \frac{Q(\mathbf{d})}{p(\mathbf{r}|\mathbf{d})p(\mathbf{d})} d\mathbf{d}$$

where $\boldsymbol{\lambda}$ denotes the set of parameters for the Q function.

Variational Inference

- Good choices of Q result in major simplifications of the original inference problem (finding $P(d_k|\mathbf{r})$)
  - Mean field approximation...

$$Q(\mathbf{d}) = \prod_{k=1}^{K} Q_k(d_k)$$

  - ...or Gaussianity

$$Q(\mathbf{d}) \propto \exp[(\mathbf{d} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{d} - \boldsymbol{\mu})]$$

Variational Inference

- We can also replace prior distribution $p(\mathbf{d})$ with postulated form, i.e. $p(\mathbf{d})$ needn't be the true expression.
- **Key**: KL divergence must be in closed form; optimal parameters must be obtainable.
- Previous Results: Obtained LMMSE-based turbo MUD, and IC-based turbo MUD with suitable choices of Q function and priors.

Discrete SISO MUD

Make mean field assumption, and let $b_k \in \{0,1\}$

$$\begin{aligned} P(\mathbf{b}) &= \prod_{k=1}^{K} \xi_k^{b_k} (1-\xi_k)^{1-b_k} \\ p(\mathbf{r}|\mathbf{b}) &= \mathcal{N}(\mathbf{Hb}, \sigma^2 \mathbf{I}) \\ Q(\mathbf{b}) &= \prod_{k=1}^{K} \gamma_k^{b_k} (1-\gamma_k)^{1-b_k} \end{aligned}$$

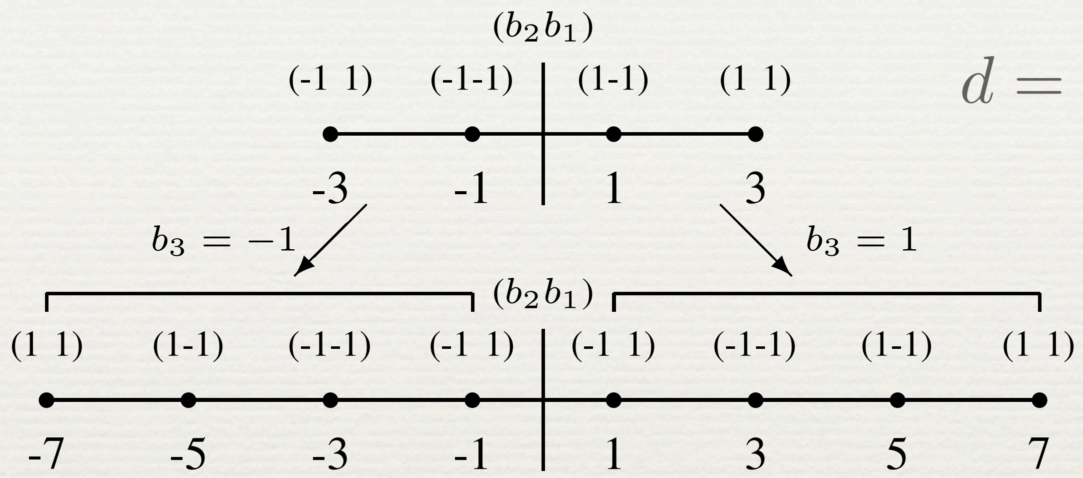
where ξ_k is the prior probability of $b_k = 1$, and

γ_k is the posterior prob. of $b_k = 1$.

Discrete SISO MUD

- The parameters of the Q function that appear in the KL divergence are $\{\gamma_1, \cdots, \gamma_K\}$.
- KL divergence has closed form which can be minimized using coordinate descent.
- As the MUD part of a turbo MUD, one form of this receiver is the IC-based turbo MUD of Alexander, et al.
- But M-QAM not easy to handle -- more than one parameter per user!

Gray Mapping for PAM



$$d = b^{(3)}b^{(2)}b^{(1)} + 2b^{(3)}b^{(2)} + 4b^{(3)}$$

$$b^{(q)} \in \{-1, +1\}$$

(q-th bit in symbol)

- In general, for 2^L-ary PAM, we have

$$d = \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} b^{(q)}$$

- This multi-linear transformation enables the extension of the variational approach to M-ary modulation.

M-QAM Turbo MUD

- For K users' symbols in a vector:

$$\mathbf{d} = \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} \mathbf{b}^{(q)}$$

where $\prod_{q} \mathbf{b}^{(q)}$ denotes element-wise product.

- Received signal is

$$\begin{aligned} \mathbf{r} &= \mathbf{Hd} + \mathbf{n} \\ &= \mathbf{H} \sum_{l=1}^{L} 2^{l-1} \prod_{q=l}^{L} \mathbf{b}^{(q)} + \mathbf{n} \end{aligned}$$

M-QAM Turbo MUD

- So we have
$$p(\mathbf{r}|\mathbf{d}) = p(\mathbf{r}|\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(L)})$$
which is known (except for noise variance).

- By the mean-field approximation, we have
$$p(\mathbf{d}) = \prod_{q=1}^{L} p(\mathbf{b}^{(q)}) = \prod_{q=1}^{L} \prod_{k=1}^{K} p(b_k^{(q)})$$
which is known from the FEC decoder in a turbo MUD.

M-QAM Turbo MUD

- In variational inference, we want to minimize KL divergence b/w $Q(\mathbf{d})$ and $p(\mathbf{r}|\mathbf{d})p(\mathbf{d})$.
- For tractability, let $Q(\mathbf{d})$ be factorizable i.e. assume all bits are conditionally independent:

$$Q(\mathbf{d}) = \prod_{q=1}^{L} Q(\mathbf{b}^{(q)}) = \prod_{q=1}^{L} \prod_{k=1}^{K} Q(b_k^{(q)})$$

- Then approx. marginal distribution can be found w/o summation or integration.
- Gaussian form also suitable.

M-QAM Turbo MUD

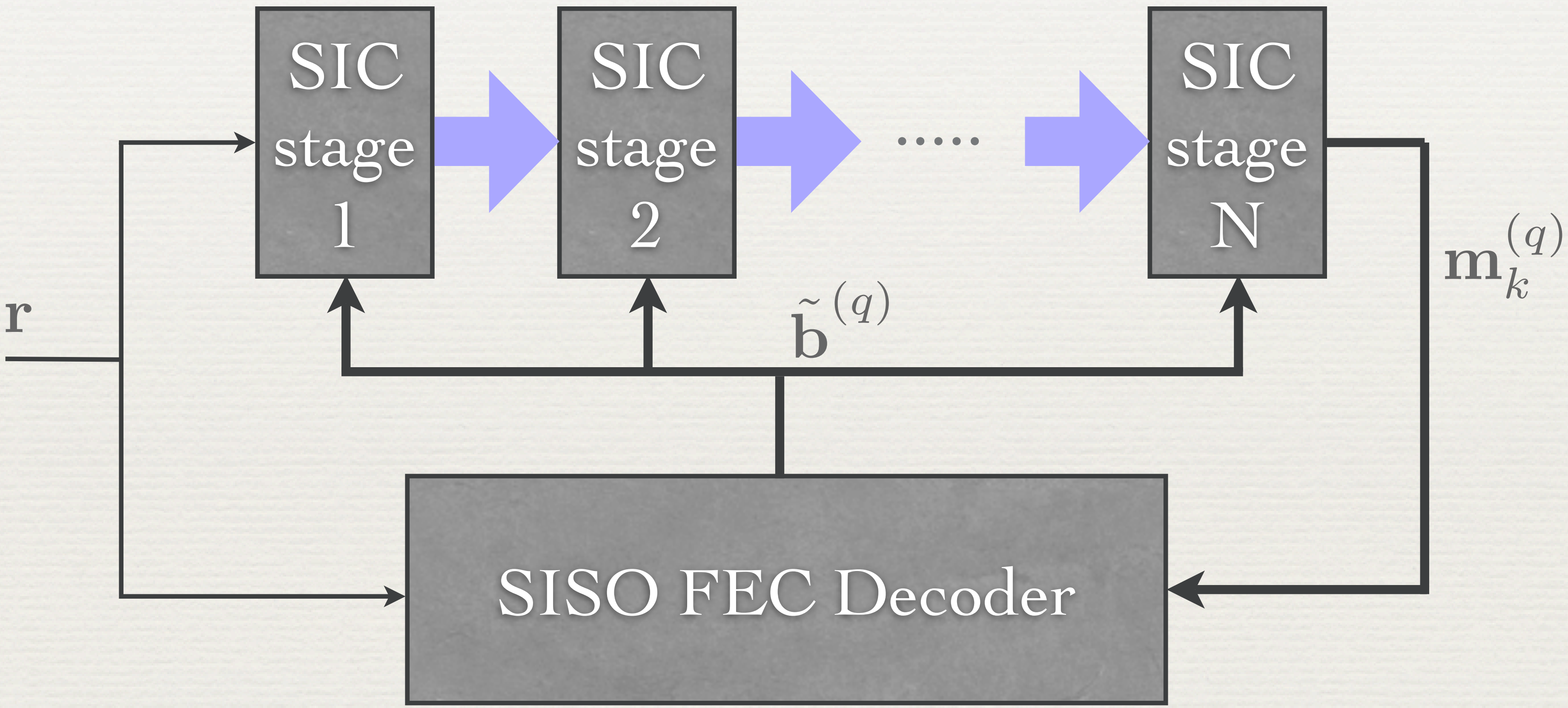
- Assuming binary distributions for $p(b_k^{(q)})$ and $Q(b_k^{(q)})$ we can find the KL divergence as a function of

$$m_k^{(q)} = E_Q(b_k^{(q)}) \quad \text{and} \quad \tilde{b}_k^{(q)} = E_p(b_k^{(q)})$$

- Setting the derivative of divergence w.r.t. $m_k^{(q)}$ to zero gives coordinate-descent updates (eq. 20 in the paper):

$$\log \frac{1 + m_k^{(q)}}{1 - m_k^{(q)}} = \log \frac{1 + \tilde{b}_k^{(q)}}{1 - \tilde{b}_k^{(q)}} + \text{IC-like update}$$

M-QAM Turbo MUD



$\mathbf{m}^{(q)}$ = Bit APP's from MUD

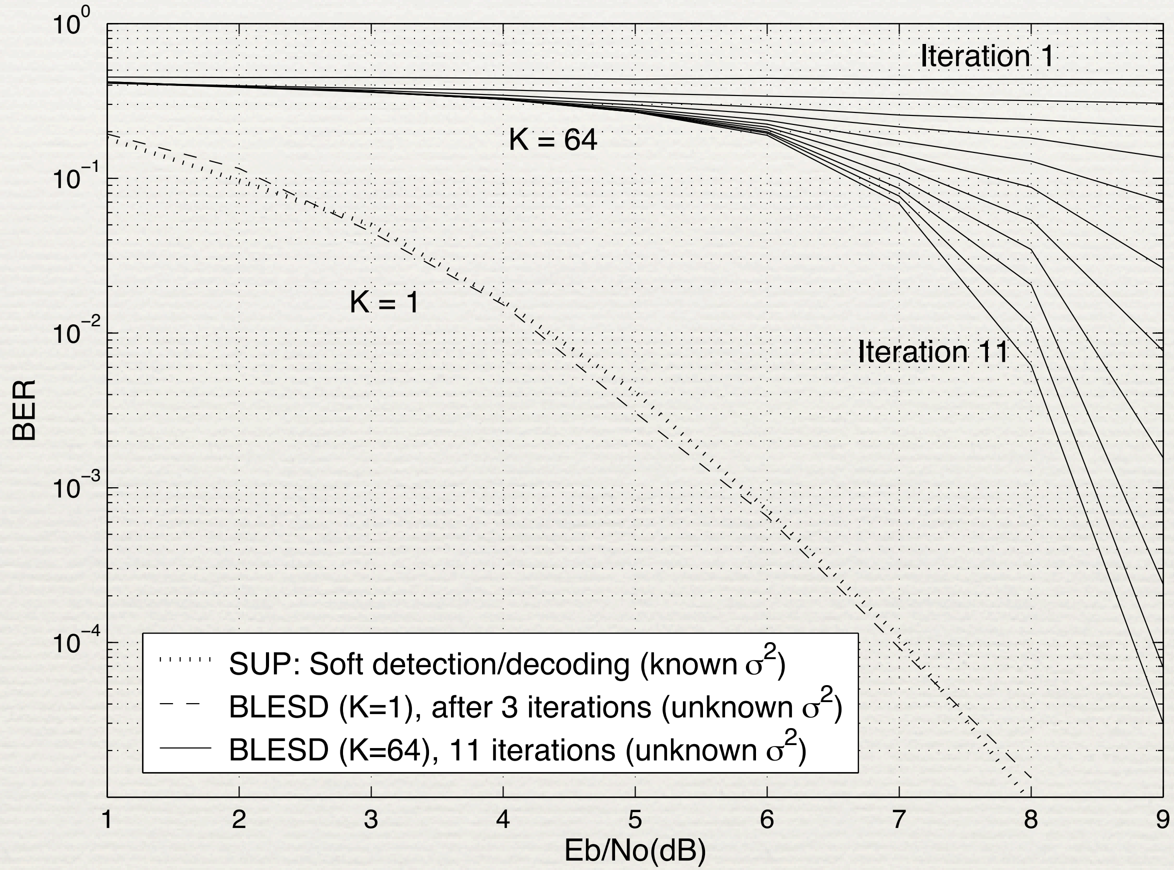
$\tilde{\mathbf{b}}^{(q)}$ = Bit priors to MUD

Unknown Noise

- By including σ^2 as an unknown variable to be estimated with variational inference, and using a "point distribution", we get a variational EM algorithm.
- Other unknowns e.g. channel can also be incorporated
    - But more unknowns usually means worse performance.

Simulation

- Random spreading, N = 64
- 4-PAM
- Rate 1/2 conv. code with generators 10011 and 11101.



Conclusions

- Most Important:
    - Traditional view of optimal MUD as too complex led to signal processing approaches e.g. MMSE, int. cancellation, adaptive filters, etc.
    - But these don't allow obvious link to FEC decoder in turbo MUD.
    - Variational approach keeps probabilistic inference viewpoint of optimal MUD, but uses distributions that are non-exact.

Conclusions

- Variational inference can be used as a unifying concept in detection
  - Different choices of $p(\mathbf{d})$ and $Q(\mathbf{d})$ can lead to various familiar detectors.
  - New viewpoint can lead to improved detectors e.g. variational EM.
- M-QAM (Gray coding) can be handled systematically, without first finding symbol APP's.