Non-Binary Belief Propagation in Large-Dimension Communication Receivers

A. Chockalingam

Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore

CEFIPRA Workshop 2014, IISc, Bangalore

14 January 2014
Outline

- Large MIMO systems
  - Motivation and challenges
- Signal detection
- BP based near-optimal detection
  - Binary BP
  - Non-binary BP
- Concluding remarks
Exploitation of large spatial dimensions

Potential to practically realize the theoretically predicted benefits of MIMO

- very high spectral efficiencies / sum rates
  - tens to hundreds of bps/Hz
- increased reliability
  - Transmit / receive diversity
- power efficiency
  - Green communications

Use of large number of antennas

- getting recognized as a good approach to fulfill high throughput requirements in future wireless systems
### MIMO capacity

- **$n_t$: # of transmit antennas, $n_r$: # receive antennas**

<table>
<thead>
<tr>
<th># Antennas</th>
<th>Error Probability ($P_e$)</th>
<th>Capacity ($C$), bps/Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SISO</strong></td>
<td>$P_e \propto SNR^{-1}$</td>
<td>$C = \log(SNR)$</td>
</tr>
<tr>
<td>$n_t = n_r = 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SIMO</strong></td>
<td>$P_e \propto SNR^{-n_r}$</td>
<td>$C = \log(SNR)$</td>
</tr>
<tr>
<td>$n_t = 1, n_r &gt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MIMO</strong></td>
<td>$P_e \propto SNR^{-n_t n_r}$</td>
<td>$C = \min(n_t, n_r) \log(SNR)$</td>
</tr>
<tr>
<td>$n_t &gt; 1, n_r &gt; 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Increased spectral efficiency (bps/Hz) with increased $n_t, n_r$
Large MIMO systems

(a) Point-to-point MIMO

(b) Multiuser MIMO

- Multiuser MIMO with hundreds of antennas at the BS: ‘Massive MIMO’
- System loading factor \( \alpha = K/N \leq 1 \)
Large MIMO systems

Release by January/February 2014
“This cutting-edge portrayal of large-scale MIMO systems provides a shrewd long-term outlook on this salient wireless subject.”

Lajos Hanzo
University of Southampton

“This is a very timely and useful book written by authors who are pioneers in the area of large MIMO systems.”

Vijay K. Bhargava
The University of British Columbia

“Large MIMO will power our wireless networks before this decade is out and the race is just starting. Chockalingam and Sundar Rajan and have compiled an excellent companion for this journey.”

Arogyaswami Paulraj
Stanford University
Challenges

- **Placement of large # of antennas in communication terminals**
  - Feasible in moderately sized communication terminals
  - Use high carrier frequencies for small carrier wavelengths (e.g., 5 GHz, 60 GHz)

- **RF technologies**
  - Multiple IF/RF transmit and receive chains

- **Signal detection**
  - Need low-complexity detectors

- **Channel estimation**
  - Estimation and feedback of large # of channel coefficients
Large multiuser MIMO systems

- Tens to hundreds of antennas at the base station
- Same or less number of users
  - Users can have one or more antennas

Uplink
- synchronization
- channel estimation, detection, decoding
- multi-cell operation

Downlink
- precoding
- CSI for precoding
- pilot contamination in multi-cell operation
System model

- $\mathbf{x}_c = [x_1, x_2, \cdots, x_K]^T$: transmitted vector
  - $x_k \in \mathcal{B}$: transmitted symbol from user $k$
  - $\mathcal{B}$: modulation alphabet (e.g., $M$-QAM)

- $\mathbf{H}_c \in \mathbb{C}^{N \times K}$: channel gain matrix
  - $h_{jk} \sim \mathcal{CN}(0, \sigma_k^2)$: gain from $k$th user to $j$th rx ant at BS

- $\mathbf{n}_c$: noise vector with i.i.d. $\mathcal{CN}(0, \sigma^2)$ entries

Complex system model: $\mathbf{y}_c = \mathbf{H}_c \mathbf{x}_c + \mathbf{n}_c$

Work with real-valued system model: $\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{n}$

- For $M$-QAM alphabet $\mathcal{B}$, elements of $\mathbf{x}$ come from underlying $\sqrt{M}$-PAM alphabet $\mathcal{A}$
Optimum detection

- Maximum likelihood (ML) decision rule:
  $$x_{ML} = \arg \min_{x \in \mathbb{A}^{2K}} \| y - Hx \|^2$$

- Maximum a posteriori (MAP) decision rule:
  $$x_{MAP} = \arg \max_{x \in \mathbb{A}^{2K}} \Pr(x \mid y, H)$$

- Complexity: exponential in $K$
Detection algorithms

- Low-complexity, near-optimal detection
  - a challenge in large dimensions

- Traditional sub-optimum solutions
  - Matched filter (MF) solution: $x_{MF} = H^T y$
  - Zero-Forcing (ZF) solution: $x_{ZF} = (H^T H)^{-1} H^T y$
  - MMSE solution: $x_{MMSE} = (H^T H + \sigma^2 I)^{-1} H^T y$
  - Problem: Poor performance in large dimensions

- Sphere decoder and variants
  - achieves ML / near-ML performance
  - Problem: doesn’t scale well beyond 32 dimensions

- Encouraging progress in large-MIMO detection
  - using algorithms rooted in AI and machine learning
Detection algorithms

Algorithms of demonstrated near-optimal performance and low complexity for large MIMO detection

1. Local neighborhood search based
   - Likelihood Ascent Search (LAS) and variants
   - Reactive Tabu Search (RTS) and variants

2. Belief propagation (BP) based
   - Message passing on graphical models
   - Scalar Gaussian approximation of interference

3. Probabilistic Data Association (PDA) based
   - Vector Gaussian approximation of interference

4. Markov Chain Monte Carlo (MCMC) based
   - Sampling from mixture distribution
Consider 4-QAM, $x_k \in \{\pm 1\}$

Each entry of $y$ is treated as a function (observation) node

Each symbol, $x_k \in \{\pm 1\}$, is treated as a variable node

Scalar Gaussian approximation of interference (GAI)

$$y_i = h_{ik}x_k + \sum_{j=1, j \neq k}^{2K} h_{ij}x_j + n_i, \quad i = 1, \ldots, 2N$$

$z_{ik}$ modeled as $\mathbb{C}\mathcal{N}(\mu_{z_{ik}}, \sigma^2_{z_{ik}})$ with

$$
\begin{align*}
\mu_{z_{ik}} &= \sum_{j=1, j \neq k}^{2K} h_{ij} \mathbb{E}(x_j), \\
\sigma^2_{z_{ik}} &= \sum_{j=1, j \neq k}^{2K} h_{ij}^2 \text{Var}(x_j) + \sigma^2
\end{align*}
$$

$h_{ij}$ is the $(i, j)$th element in $H$
BP Detection - Binary modulation

- LLR of \( x_k \) at observation node \( i \) is
  \[
  \Lambda^k_i = \log \frac{p(y_i | H, x_k = 1)}{p(y_i | H, x_k = -1)} = \frac{2}{\sigma^2_{z_{ik}}} h_{ik}(y_i - \mu_{z_{ik}})
  \]

- LLRs computed at observation nodes are passed to variable nodes
- Using these LLRs, variable nodes compute probabilities
  \[
  p^k_i(\Delta) = p_i(x_k = +1 | \mathbf{y}) = \frac{\exp(\sum_{l=1, l \neq i}^{2N} \Lambda^k_l)}{1 + \exp(\sum_{l=1, l \neq i}^{2N} \Lambda^k_l)}
  \]
  and pass them back to observation nodes
- This message passing done for a certain no. of iterations
- At the end, \( x_k \) is detected as
  \[
  \hat{x}_k = \text{sgn}\left(\sum_{i=1}^{2N} \Lambda^k_i\right)
  \]
BP Detection - Binary modulation

\[ \Lambda_i^1 = f(\{p_i^j\}, j \neq 1) \]

\[ \Lambda_i^2 \]

\[ \Lambda_i^{nt} \]

\[ x_1 \]

\[ p_i^{1+} \]

\[ y_i \]

\[ x_2 \]

\[ p_i^{2+} \]

\[ \ldots \]

\[ p_i^{nt+} \]

\[ x_{nt} \]

\[ \Lambda_i^1 \]

\[ \Lambda_i^2 \]

\[ \Lambda_i^{nt} \]

\[ x_1 \]

\[ x_2 \]

\[ x_{nt} \]

\[ y_1 \]

\[ \Lambda_1^k \]

\[ p_1^{k+} = g(\{\Lambda_i^k\}, l \neq 1) \]

\[ y_1 \]

\[ y_2 \]

\[ \Lambda_2^k \]

\[ x_k \]

\[ \Lambda_k^k \]

\[ p_2^{k+} \]

\[ y_2 \]

\[ y_{nr} \]

\[ \Lambda_{nr}^k \]

\[ p_{nr}^{k+} \]

\[ y_{nr} \]

**Figure**: Message passing between variable and observation nodes

- **Complexity**: \( O(KN) \)
- **MMSE Complexity**: \( O(KN^2) \)
BP Detection - Binary modulation

- Bit Error Rate (BER) improves with increasing $N=K$.
- Message damping factor = 0.4
- 4-QAM
- $\#$ BP iterations = 20
- SISO AWGN

Graph shows BER vs. Average received SNR (dB) for different values of $N=K$: 8, 16, 24, 32, 64.
Consider square $M$-QAM
- Each element of $\mathbf{x}$ belongs to underlying $\sqrt{M}$-PAM

**One approach**
- Write each $\sqrt{M}$-QAM symbol in the form of linear combination of $q = \log_2 \sqrt{M}$ bits

$$x_i = \sum_{j=0}^{q-1} 2^j b^{(j)}_i, \quad i = 0, 1, \ldots, 2K-1$$

- $\mathbf{c} \triangleq [2^0 2^1 \ldots 2^{q-1}]$, $\mathbf{b} \triangleq [b^{(0)}_0 \ldots b^{(q-1)}_0 \ldots b^{(0)}_{2K-1} \ldots b^{(q-1)}_{2K-1}]^T$

- $\mathbf{x} = (\mathbf{I}_{2K} \otimes \mathbf{c}) \mathbf{b} \triangleq \mathbf{H}'$

- System model in bit vector form: $\mathbf{y} = \mathbf{H}(\mathbf{I}_{2K} \otimes \mathbf{c}) \mathbf{b} + \mathbf{n}$
- Run binary message passing on this system model
Another approach
- Vector messages

\[ y_1 \quad x_1 \quad x_2 \quad \ldots \quad x_{2K} \quad y_{2N} \]

\[ \mathbf{v}_{ji} \quad \mathbf{a}_{ij} \]
y_i = h_{ij} x_j + z_{ij}, \quad i = 1, \cdots, 2N, \ j = 1, \cdots, 2K

z_{ij} \triangleq \sum_{l=1, l \neq j}^{2K} h_{il} x_l + w_i

approximate the scalar term z_{ij} as Gaussian with mean and variance

\mu_{ij} = \sum_{l=1, l \neq j}^{2K} h_{il} \mathbb{E}(x_l), \quad \sigma^2_{ij} = \sum_{l=1, l \neq j}^{2K} h_{il}^2 \text{Var}(x_l) + \sigma^2

a_{ij}: \sqrt{M}\text{-length vector message passed from } i\text{th observation node to } j\text{th variable node (likelihood)}

v_{ji}: \sqrt{M}\text{-length vector message passed from } j\text{th variable node to } i\text{th observation node (posterior probabilities)}
BP Detection - Higher-order modulation

- Likelihood and posterior probabilities are approximated as

\[
\Pr(y_i|H, x_j = s) \approx \frac{1}{\sigma_{ij} \sqrt{2\pi}} \exp \left( \frac{-(y_i - \mu_{ij} - h_{ij}s)^2}{2\sigma_{ij}^2} \right), \quad s \in A
\]

\[
\Pr(x_j = s|y, H) \propto \prod_{i=1}^{2N} \Pr(y_i|H, x_j = s) \approx \prod_{i=1}^{2N} \frac{\exp \left( \frac{-(y_i - \mu_{ij} - h_{ij}s)^2}{2\sigma_{ij}^2} \right)}{\sigma_{ij}}
\]

- Messages

\[
a_{ij}(s) = \frac{1}{\sigma_{ij} \sqrt{2\pi}} \exp \left( \frac{-(y_i - \mu_{ij} - h_{ij}s)^2}{2\sigma_{ij}^2} \right)
\]

\[
\nu_{ji}(s) = \prod_{l=1, l \neq i}^{2N} a_{lj}(s)
\]
BP Detection - Higher-order modulation

- Mean & variance at \(i\)th observation node are computed as

\[
\mu_{ij} = \sum_{l=1,l\neq j}^{2K} h_{il} s^T v_{li}
\]

\[
\sigma^2_{ij} = \sum_{l=1,l\neq j}^{2K} h^2_{il} \left( (s \odot s)^T v_{li} - (s^T v_{li})^2 \right) + \sigma^2
\]

\(s\): vector of all elements in \(\mathbb{A}\) (for \(M = 16\), \(s = [-3 \ -1 \ +1 \ +3]^T\))

- Symbol probabilities at the end (after iterations)

\[
P_{x_j}(s) \triangleq \Pr(x_j = s) \propto \prod_{l=1}^{2N} a_{lj}(s)
\]

- Bit decisions are made on probability values computed as

\[
\Pr(b^p_j = 1) = \sum_{\forall s \in \mathbb{A}: \text{\(p\)th bit in } s \text{ is } 1} P_{x_j}(s)
\]

\(b^p_j\): \(p\)th bit in the \(j\)th user’s symbol
BP Detection - Higher-order modulation

16-QAM

![Graph showing BER vs. Average SNR for different modulation techniques and block sizes. The graph includes lines for MF, MMSE, B-BP in [11], and Prop. NB-BP.]
BP Detection - Higher-order modulation

- Complexity: $O(KN\sqrt{M})$, MMSE complexity: $O(KN^2)$

(a) Performance

(b) Complexity
Channel estimation

- Transmission frame format

- Obtain MMSE channel estimate in the pilot phase
16-QAM, MMSE channel estimate
16-QAM, MMSE channel estimate

- 16-QAM
- MMSE channel estimate

Graph:
- Uncoded BER vs. Average SNR in dB
- Lines for different channel estimation methods:
  - MMSE with perfect CSI
  - Prop. NB-BP with perfect CSI
  - MMSE with estimated CSI
  - Prop. NB-BP with estimated CSI

Parameters:
- N=128, K=64
Design optimized $q$-ary LDPC codes
by matching EXIT charts of the proposed detector and the LDPC decoder
Concluding remarks

- **Large MIMO systems with tens to hundreds of antennas**
  - fast becoming a reality
  - enable multi-Gigabit wireless transmissions
  - potential candidate technology for 5G

- **Tools/algorithms from machine learning**
  - key enablers for large-dimension communication/signal processing
  - much more needs to be explored and remains to be exploited
Thank you