Graph-Based Decoding in the Presence of ISI

Linear Programming and Message Passing

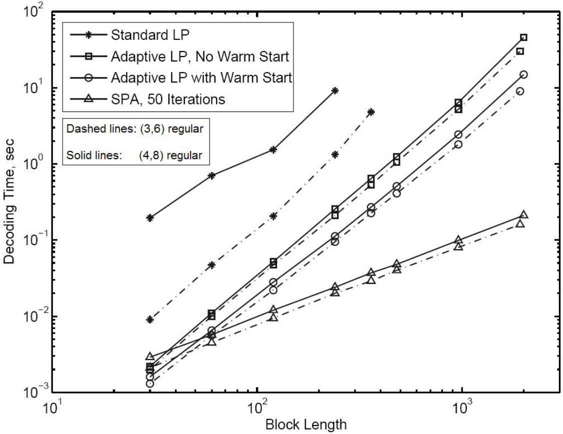
Mohammad H. Taghavi   Paul H. Siegel
(mtaghavi, psiegel)@ucsd.edu

Center for Magnetic Recording
University of California, San Diego

May, 2007

But First...

- Adaptive LP: Start with a small problem and add the constraints adaptively.



M. H. Taghavi and P. H. Siegel, "Adaptive methods for linear programming decoding," *preprint available at ArXiv*

Outline

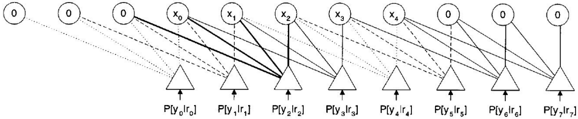
1. Graph-Based Detection
2. Uncoded Detection
   - Performance Analysis
   - Simulation Results
3. Combined Equalization and LDPC Decoding
   - Simulation Results
4. Conclusion

UCSD

Outline

Combined Channel Equalization and Decoding

- Gain obtained by combining equalization and decoding
- Need to exchange soft information between them.
  - SOVA / BCJR for equalization + message-passing
  - Exponential complexity in memory length
- Incorporate the ISI channel into the decoding graph
  - Can combine with the Tanner graph of the LDPC code
  - Use linear programming (LP) or iterative message passing (IMP) for decoding
- Kurkoski *et al.*: Bit-based detection
  - 4-cycles in the graph

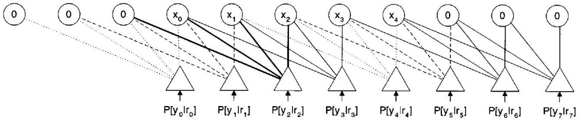


- Goal: Find a graph representation where LP can be applied.

UCSD

Combined Channel Equalization and Decoding

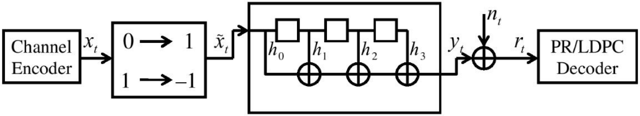
- Gain obtained by combining equalization and decoding
- Need to exchange soft information between them.
  - SOVA / BCJR for equalization + message-passing
  - Exponential complexity in memory length
- Incorporate the ISI channel into the decoding graph
  - Can combine with the Tanner graph of the LDPC code
  - Use linear programming (LP) or iterative message passing (IMP) for decoding
- Kurkoski *et al.*: Bit-based detection
  - 4-cycles in the graph



- Goal: Find a graph representation where LP can be applied.

ML Detection in a PR Channel

$y_t = \sum_{i=0}^{\mu} h_i \tilde{x}_{t-i}$



- Look for the codeword that minimizes

$$\sum_t (r_t - y_t)^2 = \sum_t \left[r_t^2 - 2r_t \sum_i h_i \tilde{x}_{t-i} + \left(\sum_i h_i \tilde{x}_{t-i} \right)^2 \right]$$

$$= \sum_t \left[\underbrace{r_t^2 + \sum_i h_i^2 \tilde{x}_{t-i}^2}_{const} \underbrace{-2r_t \sum_i h_i \tilde{x}_{t-i}}_{linear} + \underbrace{\sum_{i \neq j} \sum h_i h_j \tilde{x}_{t-i} \tilde{x}_{t-j}}_{nonlinear} \right]$$

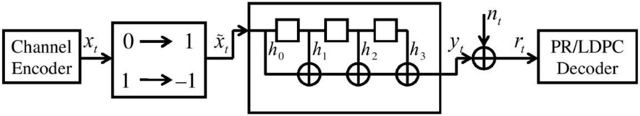
- Optimization problem in general matrix form

$$\begin{array}{ll} \text{Minimize} & -\underline{q}^T \underline{\tilde{x}} + \frac{1}{2} \underline{\tilde{x}}^T P \underline{\tilde{x}} \\ \text{Subject to} & \underline{x} \in \mathcal{C} \end{array}$$

- The general form of an integer quadratic programming problem (IQP)
- If no coding, $\mathcal{C} = \{0, 1\}^n$

ML Detection in a PR Channel

$y_t = \sum_{i=0}^{\mu} h_i \tilde{x}_{t-i}$



- Look for the codeword that minimizes

$$\sum_t (r_t - y_t)^2 = \sum_t \left[r_t^2 - 2r_t \sum_i h_i \tilde{x}_{t-i} + \left(\sum_i h_i \tilde{x}_{t-i} \right)^2 \right]$$

$$= \sum_t \left[\underbrace{r_t^2 + \sum_i h_i^2 \tilde{x}_{t-i}^2}_{const} \underbrace{-2r_t \sum_i h_i \tilde{x}_{t-i}}_{linear} + \underbrace{\sum_{i \neq j}\sum h_i h_j \tilde{x}_{t-i} \tilde{x}_{t-j}}_{nonlinear} \right]$$

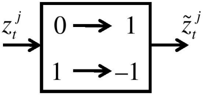
- Optimization problem in general matrix form

$$\begin{array}{ll} \text{Minimize} & -\underline{q}^T \underline{\tilde{x}} + \frac{1}{2} \underline{\tilde{x}}^T P \underline{\tilde{x}} \\ \text{Subject to} & \underline{x} \in \mathcal{C} \end{array}$$

- The general form of an integer quadratic programming problem (IQP)
- If no coding, $\mathcal{C} = \{0, 1\}^n$

Linearization of the Objective Function

- Define state variables:



$$\tilde{z}_{t,j} = \tilde{x}_t \cdot \tilde{x}_{t-j} \text{ or equivalently } z_{t,j} = x_t \oplus x_{t-j} \mod 2$$

- The IQP can be rewritten as a decoding a binary linear code:

$$\begin{aligned} \text{Minimize} \quad & \sum_t q_t x_t + \sum_t \sum_j \lambda_{t,j} z_{t,j}, \\ \text{Subject to} \quad & \underline{x} \in \mathcal{C}, \\ & z_{t,j} \oplus x_t \oplus x_{t-j} = 0 \mod 2, \; j = 1, \ldots, \mu, \\ & \qquad\qquad\qquad\qquad\qquad t = j+1, \ldots, n \end{aligned}$$

  - For the equalization problem

$$q_t = \sum_i h_i r_{t+i} \leftarrow \text{ Output of matched filter}$$

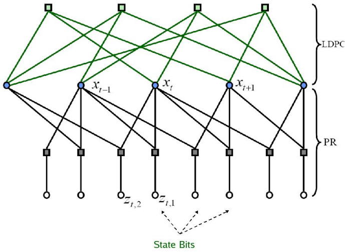
$$\lambda_{t,j} = \lambda_j = -\sum_{i=0}^{\mu-j} h_i h_{i+j} \leftarrow -1 \times \text{Correlation function of the channel}$$

Tanner Graph Representation

- PR layer:
  - $n\mu$ degree-1 state bit nodes and degree-3 check nodes
  - cycles of length 6 or more
- LP decoding
  - Parity check c with neighborhood N_c is relaxed to

$$\sum_{i \in V} x_i - \sum_{i \in N_c \setminus V} x_i \leq |V| - 1, \quad \forall\ V \subset N_c \text{ s.t. } |V| \text{ is odd}$$

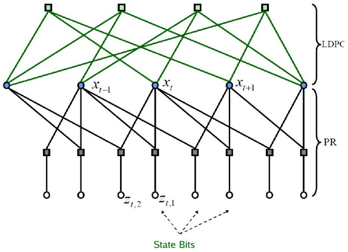
  - and $x_i \in \{0, 1\}$ is relaxed to $0 \leq x_i \leq 1$.
  - ML certificate property
- IMP Decoding
  - Use the objective coefficients as estimates of the log-likelihood ratios (LLR)
  - Complexity per iteration is linear in block length and channel memory size



UCSD

Tanner Graph Representation

- PR layer:
  - $n\mu$ degree-1 state bit nodes and degree-3 check nodes
  - cycles of length 6 or more
- LP decoding
  - Parity check c with neighborhood N_c is relaxed to



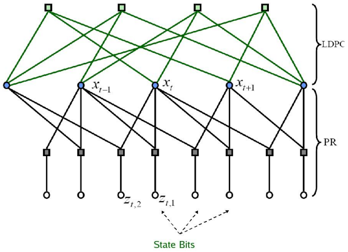
$$\sum_{i \in V} x_i - \sum_{i \in N_c \setminus V} x_i \leq |V| - 1, \quad \forall\ V \subset N_c \text{ s.t. } |V| \text{ is odd}$$

  - and $x_i \in \{0, 1\}$ is relaxed to $0 \leq x_i \leq 1$.
  - ML certificate property
- IMP Decoding
  - Use the objective coefficients as estimates of the log-likelihood ratios (LLR)
  - Complexity per iteration is linear in block length and channel memory size

UCSD

Tanner Graph Representation

- PR layer:
  - $n\mu$ degree-1 state bit nodes and degree-3 check nodes
  - cycles of length 6 or more
- LP decoding
  - Parity check c with neighborhood N_c is relaxed to



$$\sum_{i \in V} x_i - \sum_{i \in N_c \setminus V} x_i \leq |V| - 1, \quad \forall\; V \subset N_c \text{ s.t. } |V| \text{ is odd}$$

  - and $x_i \in \{0, 1\}$ is relaxed to $0 \leq x_i \leq 1$.
  - ML certificate property
- IMP Decoding
  - Use the objective coefficients as estimates of the log-likelihood ratios (LLR)
  - Complexity per iteration is linear in block length and channel memory size

Outline

Project the Problem Back to n-D

- The relaxation of the binary constraint $z_{t,j} = x_t \oplus x_{t-j}$ can be simplified as

$$|x_t - x_{t-j}| \leq z_{t,j} \leq 1 - |x_t + x_{t-j} - 1|.$$

- Depending on the sign of its coefficient, $\lambda_{t,j}$, $z_{t,j}$ will be equal to one of the two bounds.
- Solve $z_{t,j}$ in terms of x_t, and project the problem back to the n-D space:

$$\begin{aligned} \text{Minimize} \quad & f(\underline{x}) = \sum_t q_t x_t + \sum_{t,j:\lambda_{t,j}>0} \sum |\lambda_{t,j}||x_t - x_{t-j}| \\ & \quad + \sum_{t,j:\lambda_{t,j}<0} \sum |\lambda_{t,j}||x_t + x_{t-j} - 1|, \\ \text{Subject to} \quad & 0 \leq x_t \leq 1, \ \forall t = 1, \ldots, n \end{aligned}$$

  - Convex, piece-wise linear objective function.

UCSD

Project the Problem Back to n-D

- The relaxation of the binary constraint $z_{t,j} = x_t \oplus x_{t-j}$ can be simplified as
$$|x_t - x_{t-j}| \le z_{t,j} \le 1 - |x_t + x_{t-j} - 1|.$$
- Depending on the sign of its coefficient, $\lambda_{t,j}$, $z_{t,j}$ will be equal to one of the two bounds.
- Solve $z_{t,j}$ in terms of x_t, and project the problem back to the n-D space:

$$\begin{aligned} \text{Minimize} \qquad & f(\underline{x}) = \sum_t q_t x_t + \sum\sum_{t,j:\lambda_{t,j}>0} |\lambda_{t,j}||x_t - x_{t-j}| \\ & \qquad + \sum\sum_{t,j:\lambda_{t,j}<0} |\lambda_{t,j}||x_t + x_{t-j} - 1|, \\ \text{Subject to} \qquad & 0 \le x_t \le 1, \ \forall t = 1, \dots, n \end{aligned}$$

  - Convex, piece-wise linear objective function.

UCSD

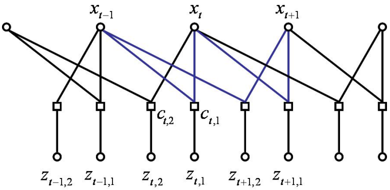
LP-Proper Channels: Guaranteed ML Performance

Theorem

LP detection is guaranteed to find the ML solution if and only if the channel satisfies:

> ***Weak Nonnegativity Condition (WNC)**: Every check node $c_{t,j}$ that is on a cycle in the Tanner graph corresponds to a nonnegative coefficient: $\lambda_{t,j} \geq 0$.*

- We call them *LP-proper channels*.
- Can interpret the problem as generalized min-cut



Corollary

The solution $\hat{\underline{x}}$ of LP detection on any channel is in $\{0, \frac{1}{2}, 1\}^n$.

UCSD

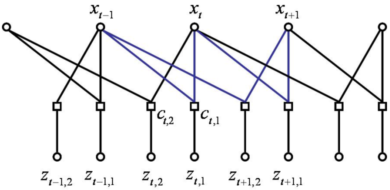
LP-Proper Channels: Guaranteed ML Performance

Theorem

LP detection is guaranteed to find the ML solution if and only if the channel satisfies:

> ***Weak Nonnegativity Condition (WNC)****: Every check node $c_{t,j}$ that is on a cycle in the Tanner graph corresponds to a nonnegative coefficient: $\lambda_{t,j} \geq 0$.*

- We call them *LP-proper channels*.
- Can interpret the problem as generalized min-cut

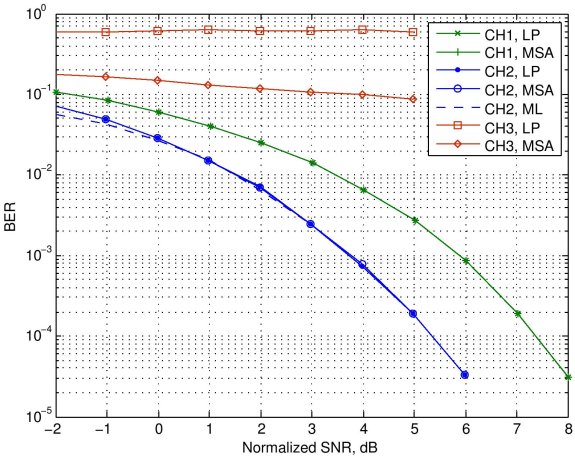


Corollary

The solution $\hat{\underline{x}}$ of LP detection on any channel is in $\{0, \frac{1}{2}, 1\}^n$.

UCSD

Simulation: LP and MSA



1. **CH1:** $h(D) = 1 - D - 0.5D^2 - 0.5D^3$ (satisfies WNC) $\leftarrow$ *LP-proper*
2. **CH2:** $h(D) = 1 + D - D^2 + D^3$ $\leftarrow$ *Asymptotically LP-proper*
3. **CH3:** $h(D) = 1 + D - D^2 - D^3$ $\leftarrow$ *LP-improper*

UCSD

General Channels: High SNR Analysis

Question

When is the performance of LP asymptotically close to ML?

- LP detection has two dominant types of failure
  - Type 1 (E_1): ML gives the correct solution $\underline{x}$, but LP gives a fractional solution, $\hat{\underline{x}}$.
  - Type 2 (E_2): Both LP and ML fail to find the correct solution.
- Two extreme cases:
  - $\Pr[E_1] \ll \Pr[E_2]$ at high SNR: LP asymptotically achieves ML performance $\longleftarrow$ *Asymptotically LP-Proper Channel*
  - $\Pr[E_1] \geq \beta > 0$, $\forall$ SNR: LP performs poorly $\longleftarrow$ *LP-Improper Channel*
- Sufficient condition for type-1 failure:

$$\exists \hat{\underline{x}} \in \left\{0, \frac{1}{2}, 1\right\}^n : \; f(\hat{\underline{x}}) - f(\underline{x}) \leq 0$$

  - Separate the signal and noise terms: $f(\hat{\underline{x}}) - f(\underline{x}) = \delta + \eta$
- If $\delta \leq 0$ for some $(\underline{x}, \hat{\underline{x}})$, the channel is LP-improper.
- To find the dominant error event, we should optimize over $\underline{x}$ and $\hat{\underline{x}}$.

UCSD

General Channels: High SNR Analysis

Question

When is the performance of LP asymptotically close to ML?

- LP detection has two dominant types of failure
  - Type 1 (E_1): ML gives the correct solution $\underline{x}$, but LP gives a fractional solution, $\hat{\underline{x}}$.
  - Type 2 (E_2): Both LP and ML fail to find the correct solution.
- Two extreme cases:
  - $\Pr[E_1] \ll \Pr[E_2]$ at high SNR: LP asymptotically achieves ML performance $\leftarrow$ *Asymptotically LP-Proper Channel*
  - $\Pr[E_1] \geq \beta > 0, \ \forall$ SNR: LP performs poorly $\leftarrow$ *LP-Improper Channel*
- Sufficient condition for type-1 failure:

$$\exists \hat{\underline{x}} \in \left\{0, \frac{1}{2}, 1\right\}^n : \ f(\hat{\underline{x}}) - f(\underline{x}) \leq 0$$

  - Separate the signal and noise terms: $f(\hat{\underline{x}}) - f(\underline{x}) = \delta + \eta$
- If $\delta \leq 0$ for some $(\underline{x}, \hat{\underline{x}})$, the channel is LP-improper.
- To find the dominant error event, we should optimize over $\underline{x}$ and $\hat{\underline{x}}$.

UCSD

General Channels: High SNR Analysis

Question

When is the performance of LP asymptotically close to ML?

- LP detection has two dominant types of failure
  - Type 1 (E_1): ML gives the correct solution $\underline{x}$, but LP gives a fractional solution, $\hat{\underline{x}}$.
  - Type 2 (E_2): Both LP and ML fail to find the correct solution.
- Two extreme cases:
  - $\Pr[E_1] \ll \Pr[E_2]$ at high SNR: LP asymptotically achieves ML performance $\longleftarrow$ *Asymptotically LP-Proper Channel*
  - $\Pr[E_1] \geq \beta > 0, \ \forall$ SNR: LP performs poorly $\longleftarrow$ *LP-Improper Channel*
- Sufficient condition for type-1 failure:

$$\exists \hat{\underline{x}} \in \left\{0, \frac{1}{2}, 1\right\}^n : \ f(\hat{\underline{x}}) - f(\underline{x}) \leq 0$$

  - Separate the signal and noise terms: $f(\hat{\underline{x}}) - f(\underline{x}) = \delta + \eta$
- If $\delta \leq 0$ for some $(\underline{x}, \hat{\underline{x}})$, the channel is LP-improper.
- To find the dominant error event, we should optimize over $\underline{x}$ and $\hat{\underline{x}}$.

General Channels: High SNR Analysis

Question

When is the performance of LP asymptotically close to ML?

- LP detection has two dominant types of failure
  - Type 1 (E_1): ML gives the correct solution $\underline{x}$, but LP gives a fractional solution, $\hat{\underline{x}}$.
  - Type 2 (E_2): Both LP and ML fail to find the correct solution.
- Two extreme cases:
  - $\Pr[E_1] \ll \Pr[E_2]$ at high SNR: LP asymptotically achieves ML performance $\leftarrow$ *Asymptotically LP-Proper Channel*
  - $\Pr[E_1] \geq \beta > 0, \ \forall$ SNR: LP performs poorly $\leftarrow$ *LP-Improper Channel*
- Sufficient condition for type-1 failure:

$$\exists \hat{\underline{x}} \in \left\{0, \frac{1}{2}, 1\right\}^n : \ f(\hat{\underline{x}}) - f(\underline{x}) \leq 0$$

  - Separate the signal and noise terms: $f(\hat{\underline{x}}) - f(\underline{x}) = \delta + \eta$
- If $\delta \leq 0$ for some $(\underline{x}, \hat{\underline{x}})$, the channel is LP-improper.
- To find the dominant error event, we should optimize over $\underline{x}$ and $\hat{\underline{x}}$.

UCSD

All-$\frac{1}{2}$ Event

- The most interesting case is when $\underline{\hat{x}} = [\frac{1}{2}, \cdots, \frac{1}{2}]$:

Lemma

If the transmitted sequence is i.i.d. Bernouli(1/2), as $n \to \infty$

$$\delta \to n\Big[|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\Big] \qquad \textit{and} \qquad \varsigma^2 \to \sigma^2 n\big[|\lambda_0|\big]$$

- Natural to define $\delta_\infty \triangleq \frac{1}{|\lambda_0|}\left(|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\right)$

Theorem

The WER of uncoded LP detection with an i.i.d. Bernouli$(1/2)$ sequence of transmitted symbols goes to 1 as the block length n goes to infinity for any SNR, i.e., the channel is LP-improper, if $\delta_\infty < 0$.

Lemma

LP-proper channels satisfy $\delta_\infty > \frac{1}{2}$.

All-$\frac{1}{2}$ Event

- The most interesting case is when $\underline{\hat{x}} = [\frac{1}{2}, \cdots, \frac{1}{2}]$:

Lemma

If the transmitted sequence is i.i.d. Bernouli(1/2), as $n \to \infty$

$$\delta \to n\Big[|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\Big] \qquad \textit{and} \qquad \varsigma^2 \to \sigma^2 n\big[|\lambda_0|\big]$$

- Natural to define $\delta_\infty \triangleq \frac{1}{|\lambda_0|}\left(|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\right)$

Theorem

The WER of uncoded LP detection with an i.i.d. Bernouli$(1/2)$ sequence of transmitted symbols goes to 1 as the block length n goes to infinity for any SNR, i.e., the channel is LP-improper, if $\delta_\infty < 0$.

Lemma

LP-proper channels satisfy $\delta_\infty > \frac{1}{2}$.

All-$\frac{1}{2}$ Event

- The most interesting case is when $\underline{\hat{x}} = [\frac{1}{2}, \cdots, \frac{1}{2}]$:

Lemma

If the transmitted sequence is i.i.d. Bernouli(1/2), as $n \to \infty$

$$\delta \to n\Big[|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\Big] \qquad \textit{and} \qquad \varsigma^2 \to \sigma^2 n\big[|\lambda_0|\big]$$

- Natural to define $\delta_\infty \triangleq \frac{1}{|\lambda_0|}\left(|\lambda_0| - \sum_{j=1}^{\mu} |\lambda_j|\right)$

Theorem

The WER of uncoded LP detection with an i.i.d. Bernouli$(1/2)$ sequence of transmitted symbols goes to 1 as the block length n goes to infinity for any SNR, i.e., the channel is LP-improper, if $\delta_\infty < 0$.

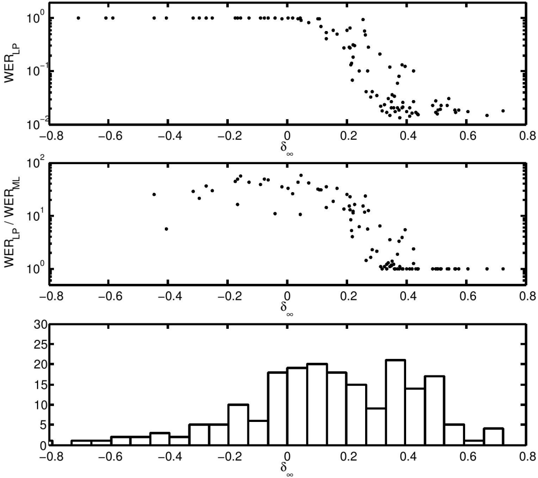
Lemma

LP-proper channels satisfy $\delta_\infty > \frac{1}{2}$.

- 200 randomly-generated channels of memory size 4.
- The channel taps are i.i.d. $\sim \mathcal{N}(0,1)$.
- Normalized to have unit power gain:

$$|\lambda_0| = \sum_j |h_i|^2 = 1$$

- SNR=11dB
- Strong correlation between the performance and δ_∞.



Outline

Coded LP Detection

- Add the relaxed parity-check constraints to the set of constraints.
- These constraints cut some of the existing pseudo-codewords, and add some new ones.

Corollary

Consider a linear code with no "trivial" (i.e., degree-1) parity check, used on a channel with $\delta_\infty < 0$. Then, coded LP detection on this system has a WER bounded below by a constant at all SNR for large block lengths.

Proof.

Follows from the analysis of uncoded detection and the fact that the all-$\frac{1}{2}$ vector satisfies all the non-trivial constraints of any linear code. □

UCSD

Coded LP Detection

- Add the relaxed parity-check constraints to the set of constraints.
- These constraints cut some of the existing pseudo-codewords, and add some new ones.

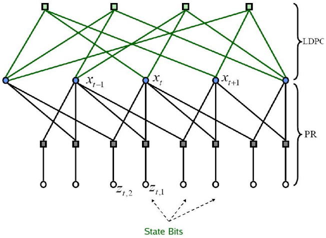
Corollary

Consider a linear code with no "trivial" (i.e., degree-1) parity check, used on a channel with $\delta_\infty < 0$. Then, coded LP detection on this system has a WER bounded below by a constant at all SNR for large block lengths.

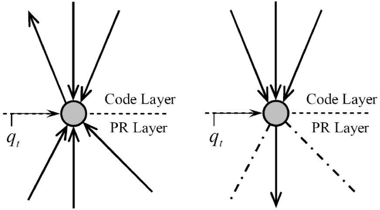
Proof.

Follows from the analysis of uncoded detection and the fact that the all-$\frac{1}{2}$ vector satisfies all the non-trivial constraints of any linear code. $\square$

Coded IMP Detection

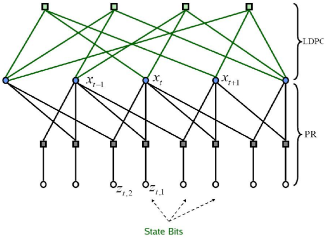


- Min-Sum Algorithm (MSA)
  - Use the LP coefficients $\{q_t\}$ and $\{\lambda_{t,j}\}$ as the costs.
- Sum-Product Algorithm (SPA)
  - Estimate "log-likelihood ratios" by multiplying $\{q_t\}$ and $\{\lambda_{t,j}\}$ by $2/\sigma^2$.
  - In the absence of ISI reduce to the true LLRs.
- Use a selective rule for combining messages in order to mitigate the effect of cycles in the PR layer.
  - To calculate the messages going to the PR layer only use the messages coming from the LDPC layer:

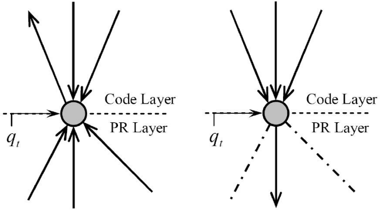


Coded IMP Detection

- Min-Sum Algorithm (MSA)
  - Use the LP coefficients $\{q_t\}$ and $\{\lambda_{t,j}\}$ as the costs.
- Sum-Product Algorithm (SPA)
  - Estimate "log-likelihood ratios" by multiplying $\{q_t\}$ and $\{\lambda_{t,j}\}$ by $2/\sigma^2$.
  - In the absence of ISI reduce to the true LLRs.

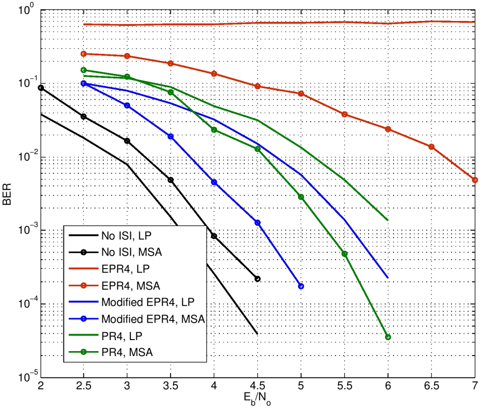


- Use a selective rule for combining messages in order to mitigate the effect of cycles in the PR layer.
  - To calculate the messages going to the PR layer only use the messages coming from the LDPC layer:



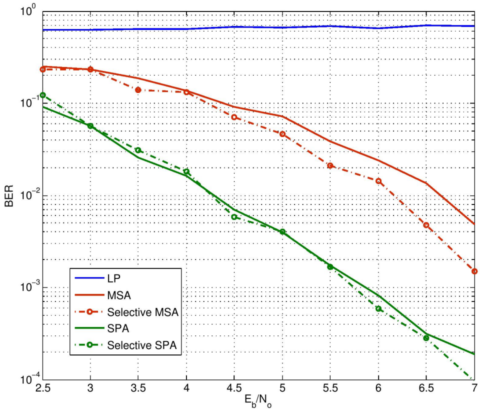
Simulation Results

- A randomly-generated regular LDPC code of length 200, rate $1/4$, and variable degree 3.
- The following PR channels:
  1. **No-ISI Channel:** $h(D) = 1$,
  2. **EPR4 Channel:** $h(D) = 1 + D - D^2 - D^3$ ($\delta_\infty = 0$, LP-improper),
  3. **Modified EPR4:** $h(D) = 1 + D - D^2 + D^3$ ($\delta_\infty = \frac{1}{2}$, Asymptotically LP-proper),
  4. **PR4 Channel:** $h(D) = 1 - D^2$ ($\delta_\infty = \frac{1}{2}$, LP-proper).



More on the EPR4 Channel

- With coding, there is a large gap between LP, MSA, and SPA.
  - Unlike LP, IMP works on LP-improper channels.
- Some gain for MSA by selective combining



Outline

Conclusion

1. Summary
   - Proposed a linear relaxation of the equalization problem
   - Easily applicable to combined equalization and decoding with LP or message passing
   - Derived necessary and sufficient conditions for optimal performance
   - Characterized the error events
   - IMP is superior to LP in combined channel equalization/decoding
2. Outlook
   - Modifying the constraints/combining rules to improve on LP-improper channels
   - Applications in the context of PRML detection
   - Applications to 2-D ISI channels
   - Exact performance analysis, especially with coding