LP Decoding

Allerton, 2003

Jon Feldman
jonfeld@ieor.columbia.edu
Columbia University

David Karger
karger@theory.lcs.mit.edu
MIT

Martin Wainwright
wainwrig@eecs.berkeley.edu
UC Berkeley

LP Decoding

- Linear Programming (LP):
  - Finding a solution to a set of linear inequalities that optimizes a linear objective function.
- Integer Linear Programming (ILP):
  - LP where variables constrained to be integers.
- LP Relaxation:
  - Using an LP to find a good (approximate) solution to an ILP.
- LP Decoding:
  - **LP relaxation for the Maximum-Likelihood (ML) decoding problem.**

LP Decoding

- Previous work on specific code families/constructions:
  - Turbo codes [FK, FOCS '02] [EH, A '03] [F '03].
  - LDPC codes [FKW, CISS '03] [F '03].
  - New iterative algs. [FKW, Allerton '02] [F '03].
- This paper: general treatment of LP decoding, for any binary code, memoryless channel (BSC, AWGN).
  - *Proper* polytope (ML certificate).
  - *LP pseudocodeword*.
  - *Fractional Distance*.
  - *Symmetric* polytope (linear codes).

Maximum-Likelihood (ML) Decoding

- Log-likelihood ratio (LLR) γ_i of y_i as a cost function:

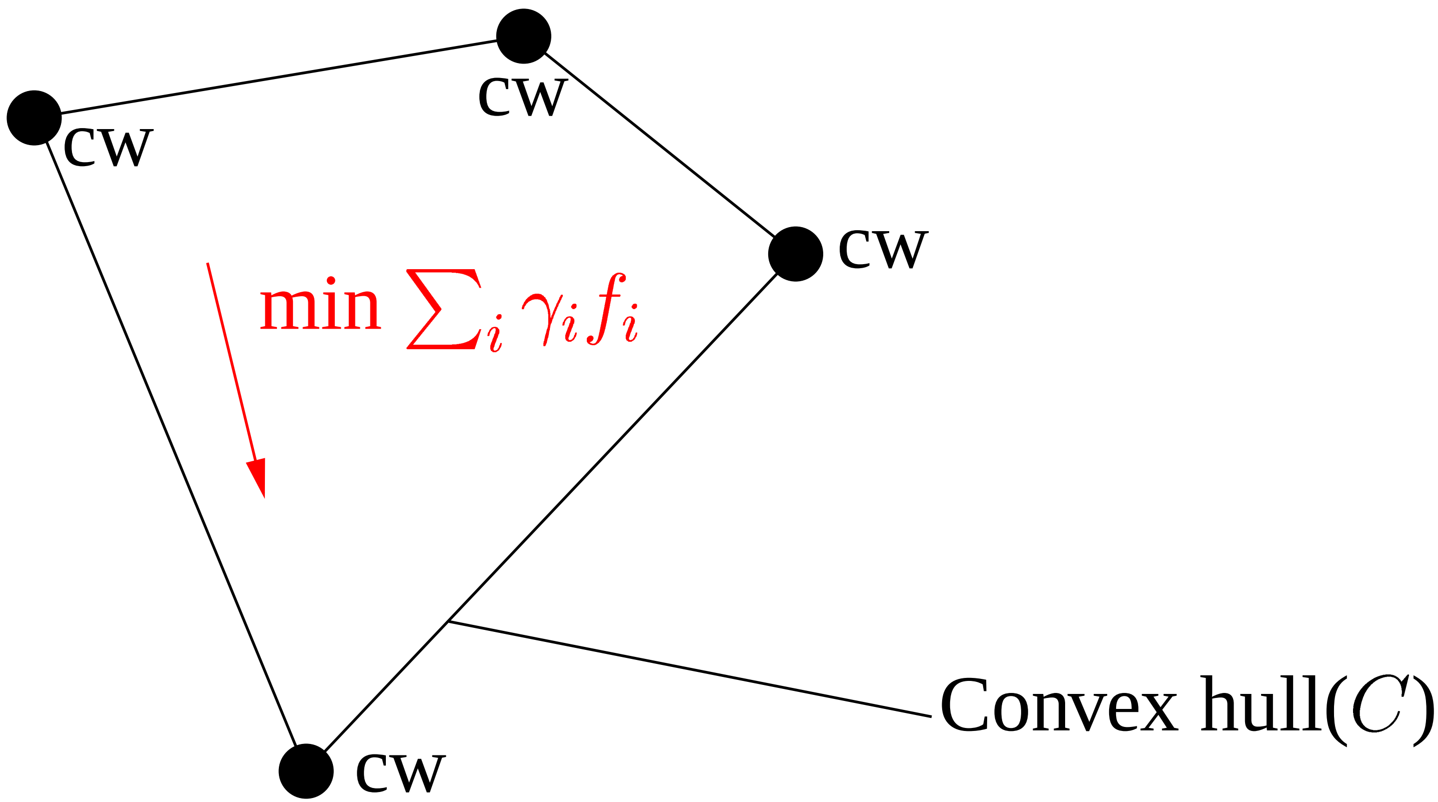
$$\gamma_i = \ln \left(\frac{\Pr[\hat{y}_i \mid y_i = 0]}{\Pr[\hat{y}_i \mid y_i = 1]} \right)$$

  - $\gamma_i > 0 \implies y_i$ more likely 0
  - $\gamma_i < 0 \implies y_i$ more likely 1

- For any binary-input memoryless channel:

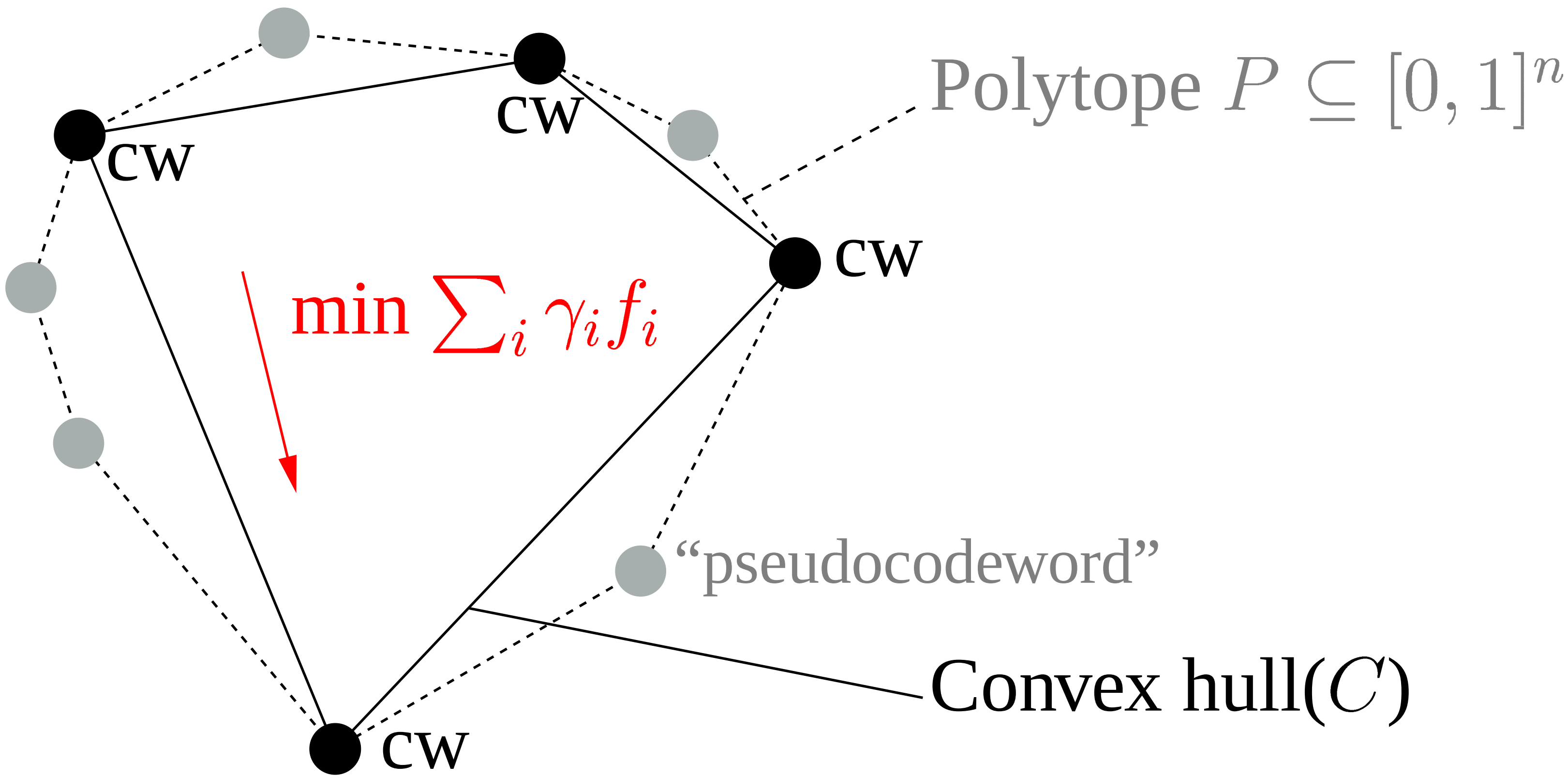
> ML DECODING: Given LLRs $\{\gamma_i, \ldots, \gamma_n\}$, find $y \in C$ such that $\sum_i \gamma_i y_i$ is minimized.

Maximum-Likelihood (ML) Decoding



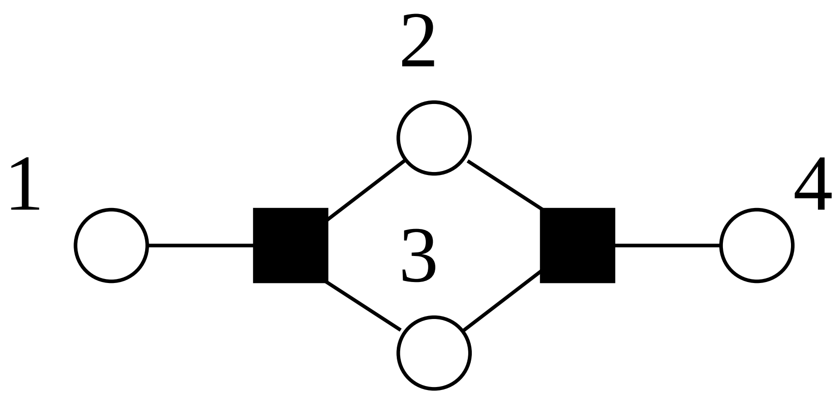
- CH(C) = convex hull of codewords; CH(C) $\subseteq [0,1]^n$.
- ML Decoding: Minimize $\sum_i \gamma_i f_i$ s.t. $f \in$ CH(C).
- Problem: CH(C) is too complex (not poly-size).

LP Decoding



- "Proper" relaxation polytope P: $P \cap \{0,1\}^n = C$.
- Alg: Solve LP. If f^* integral, output f^*, else "error."
- *ML certificate* property

LP Decoder Example



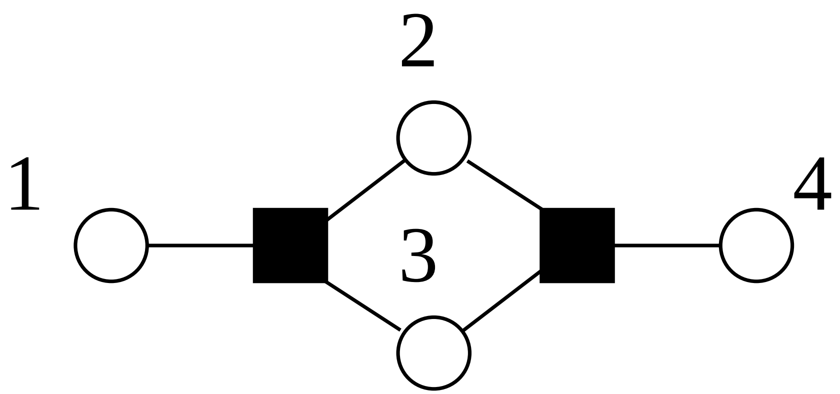
$C = \{\texttt{0000}, \texttt{1101}, \texttt{1011}, \texttt{0110}\}$

- Define polytope P on variables $\{f_1, f_2, f_3, f_4\}$:

$$
\begin{array}{ccc}
f_1 \le f_2 + f_3 & f_2 \le f_3 + f_4 & 0 \le f_1 \le 1 \\
f_2 \le f_1 + f_3 & f_3 \le f_2 + f_4 & 0 \le f_2 \le 1 \\
f_3 \le f_1 + f_2 & f_4 \le f_2 + f_3 & 0 \le f_3 \le 1 \\
f_1 + f_2 + f_3 \le 2 & f_2 + f_3 + f_4 \le 2 & 0 \le f_4 \le 1
\end{array}
$$

- Is P proper (does $P \cap \{0,1\}^n = C$) ?

LP Decoder Example



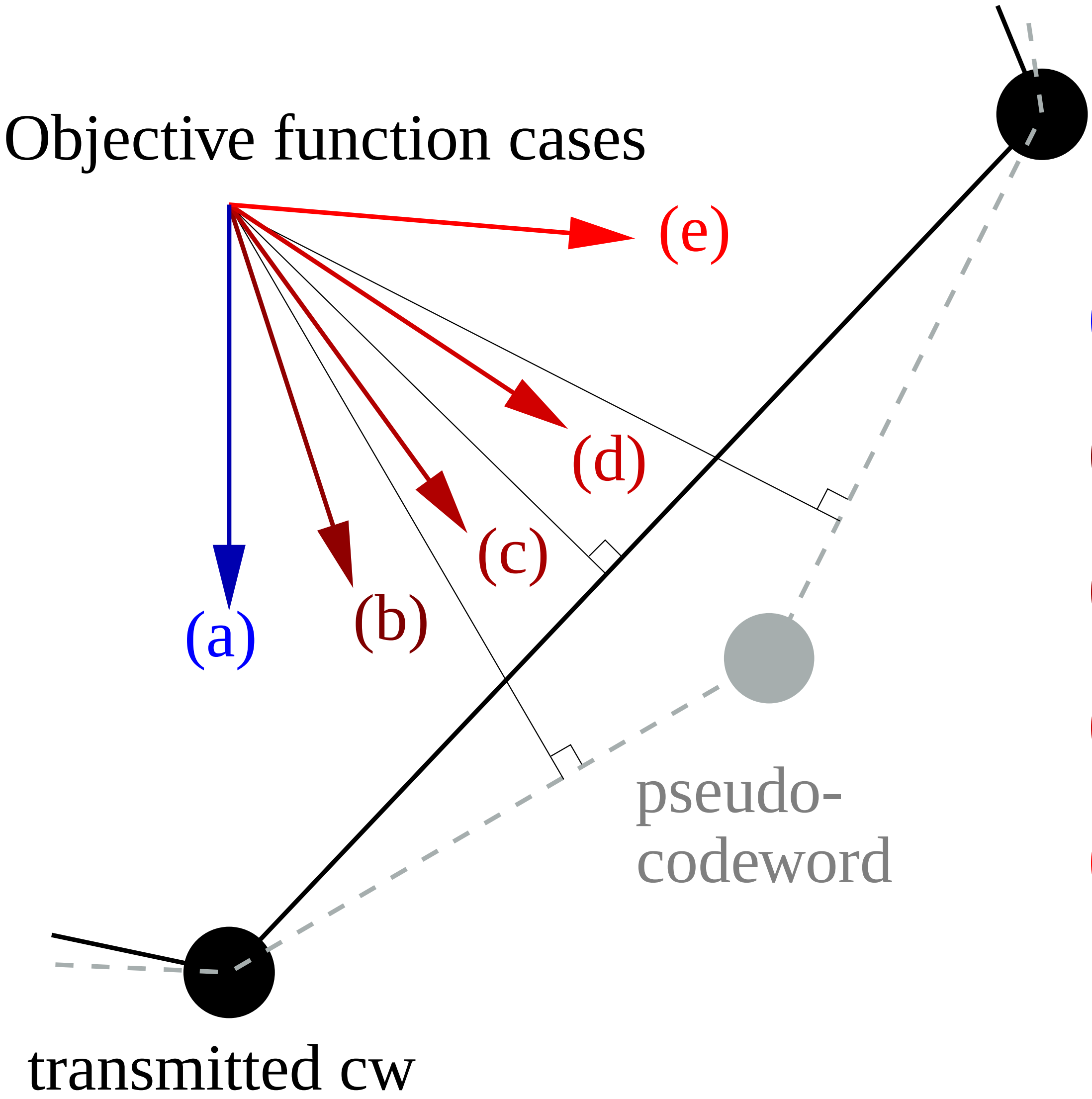
$C = \{0000, 1101, 1011, 0110\}$

- Polytope:

$$
\begin{array}{ccc}
f_1 \leq f_2 + f_3 & f_2 \leq f_3 + f_4 & 0 \leq f_1 \leq 1 \\
f_2 \leq f_1 + f_3 & f_3 \leq f_2 + f_4 & 0 \leq f_2 \leq 1 \\
f_3 \leq f_1 + f_2 & f_4 \leq f_2 + f_3 & 0 \leq f_3 \leq 1 \\
f_1 + f_2 + f_3 \leq 2 & f_2 + f_3 + f_4 \leq 2 & 0 \leq f_4 \leq 1
\end{array}
$$

- Vertices: $\{0000, 1101, 1011, 0110, 1\,\frac{1}{2}\,\frac{1}{2}\,0, 0\,\frac{1}{2}\,\frac{1}{2}\,1\}$

LP Decoding Success Conditions



(a) No noise

(b) Both succeed

(c) ML succeed, LP fail

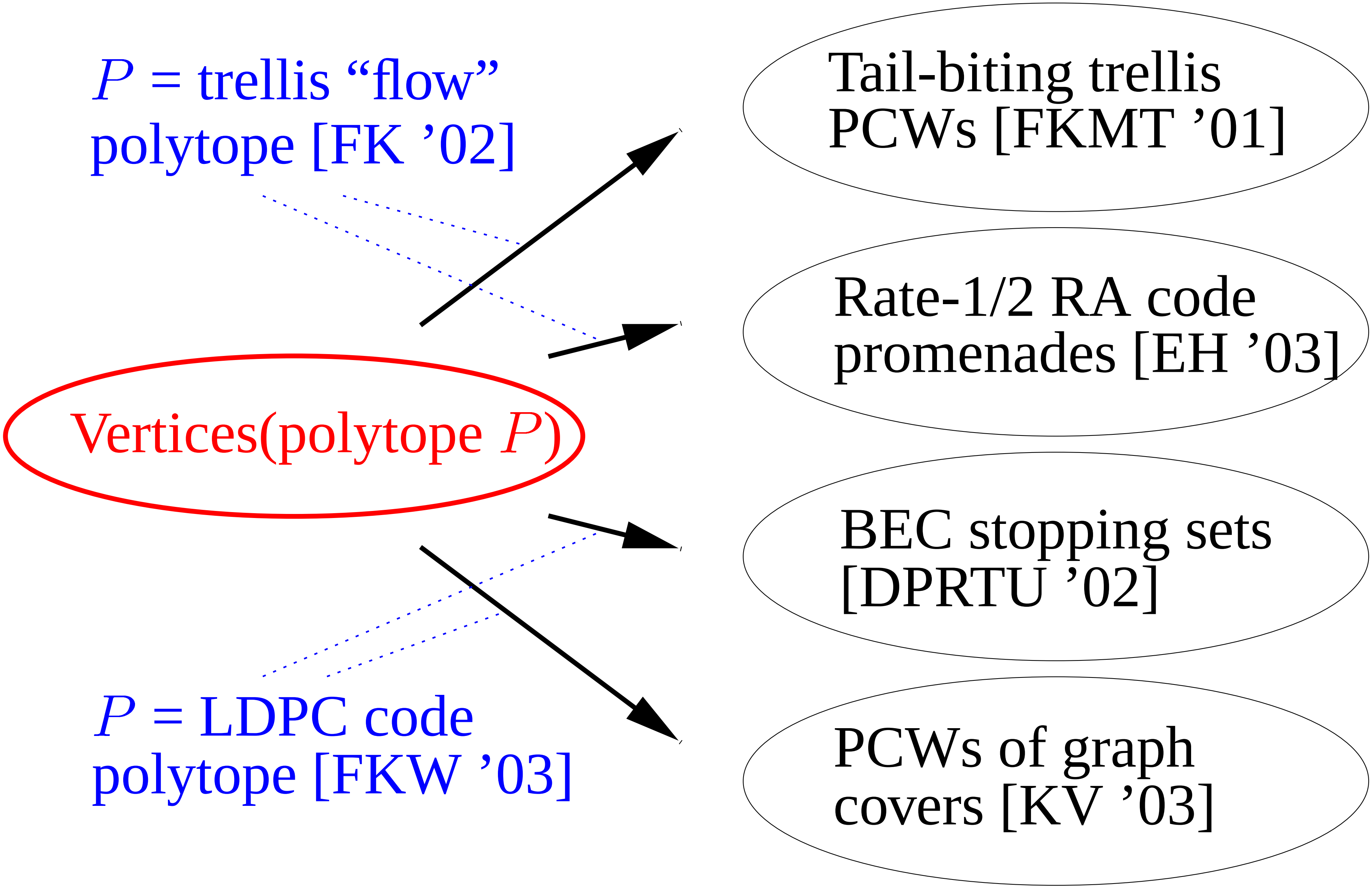
(d) Both fail, detected

(e) Both fail, undetected

LP Pseudocodewords

- In general, pseudocodewords are the set of possible results of a sub-optimal decoder :
  - PCWs $\supset$ codewords;
  - Algorithm finds min-cost PCW;
  - WER = Pr[transmitted cw = min-cost PCW].
- Example: It. decoding in the BEC [Di et. al, '02].
  - PCWs = "stopping sets" $\supset$ codewords;
  - Iterative decoding finds min-cost stopping set.
- LP Decoding:
  - PCWs = polytope vertices $\supset$ codewords
  - LP Decoder find min-cost polytope vertex.

Unifying Other Known PCWs



Using PCWs for Performance Bounds

- Turbo code polytope [FK '02, F '03]:

Theorem: In {BSC, AWGN}, for any $\alpha > 0$, if $\{p, \sigma^2\} < f(\alpha)$, then WER $\leq n^{-\alpha}$.

  - Bounds improved by [EH, Allerton '03].

- LDPC code polytope [FKW, CISS '03]: For any graph G with girth g, left-degree $\geq d_\ell$:

Theorem: LP decoding corrects $(d_\ell - 1)^{\lceil g/4 \rceil - 1}$ errors (adversarial).

  - With log-girth, can correect $\Omega(n^{1-\epsilon})$ errors.

Fractional Distance

- Another way to define (classical) distance d:
  - d = min l_1 dist. between two integral vertices of P.
- Fractional distance:
  - d_{frac} = min l_1 distance between an integral vertex and any other vertex of P.
  - Lower bound on classical distance: $d_{frac} \leq d$.

Theorem: In the binary symmetric channel, LP decoders can correct up to $\lceil d_{frac}/2 \rceil - 1$ errors.

- Linear codes: Given facets of P, fractional distance can be computed efficiently.

Symmetric Polytopes for Linear Codes

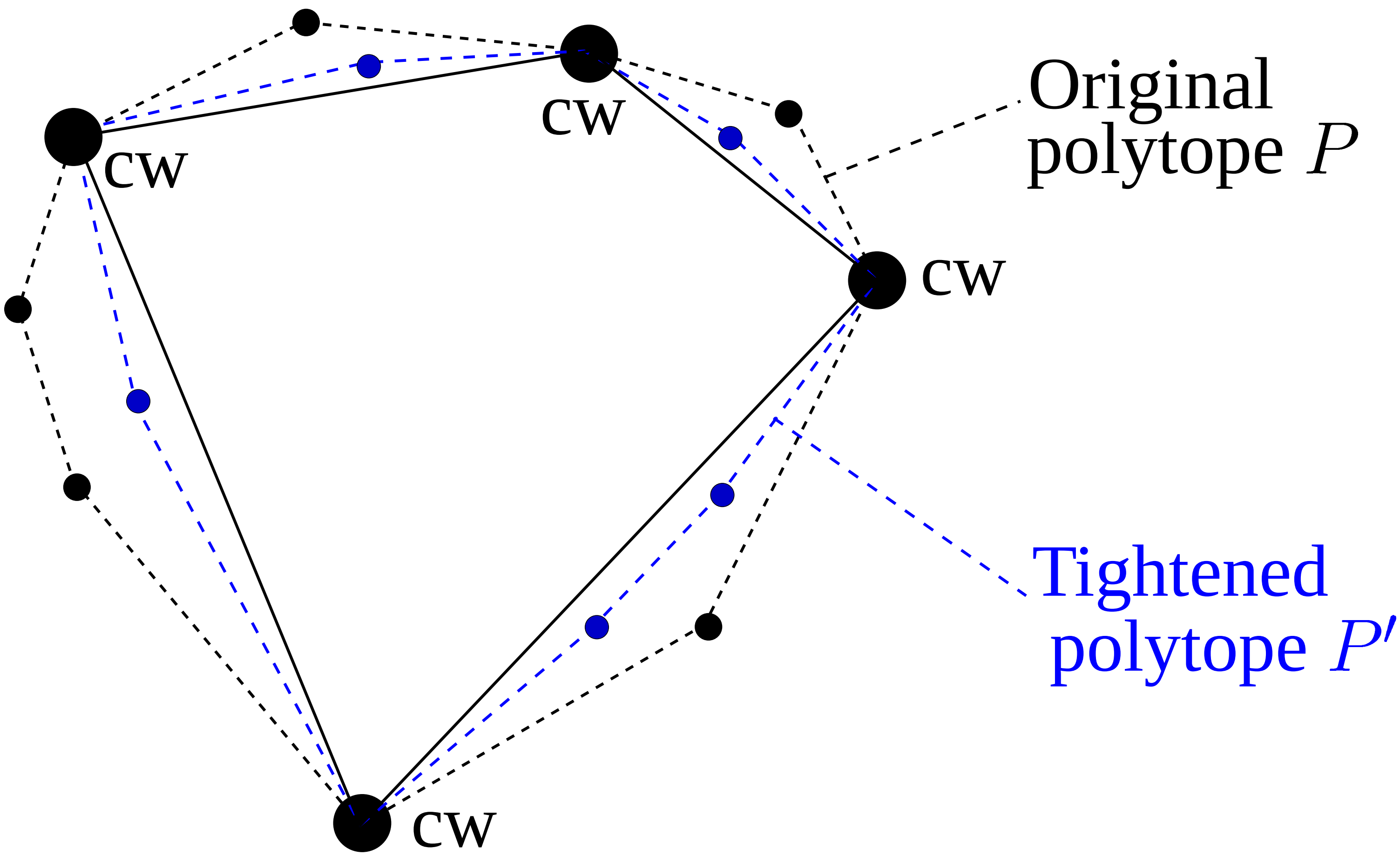
- ML decoding:
  - If C is linear, may assume $\mathtt{0}^n$ is transmitted.
  - Simplifies analysis, notation.
  - Min-distance = min-weight.
- Same assumption can be made for iterative algorithms, since pseudocodewords obey "symmetry."
- LP Decoding:

Definition: Polytope P is C-symmetric if, for all $f \in P$ and $y \in C$, we have $f^{[y]} \in P$ (where $f_i^{[y]} = |y_i - f_i|$).

Theorem: If polytope P is proper and C-symmetric, then WER of LP decoder using P is independent of the transmitted codeword.

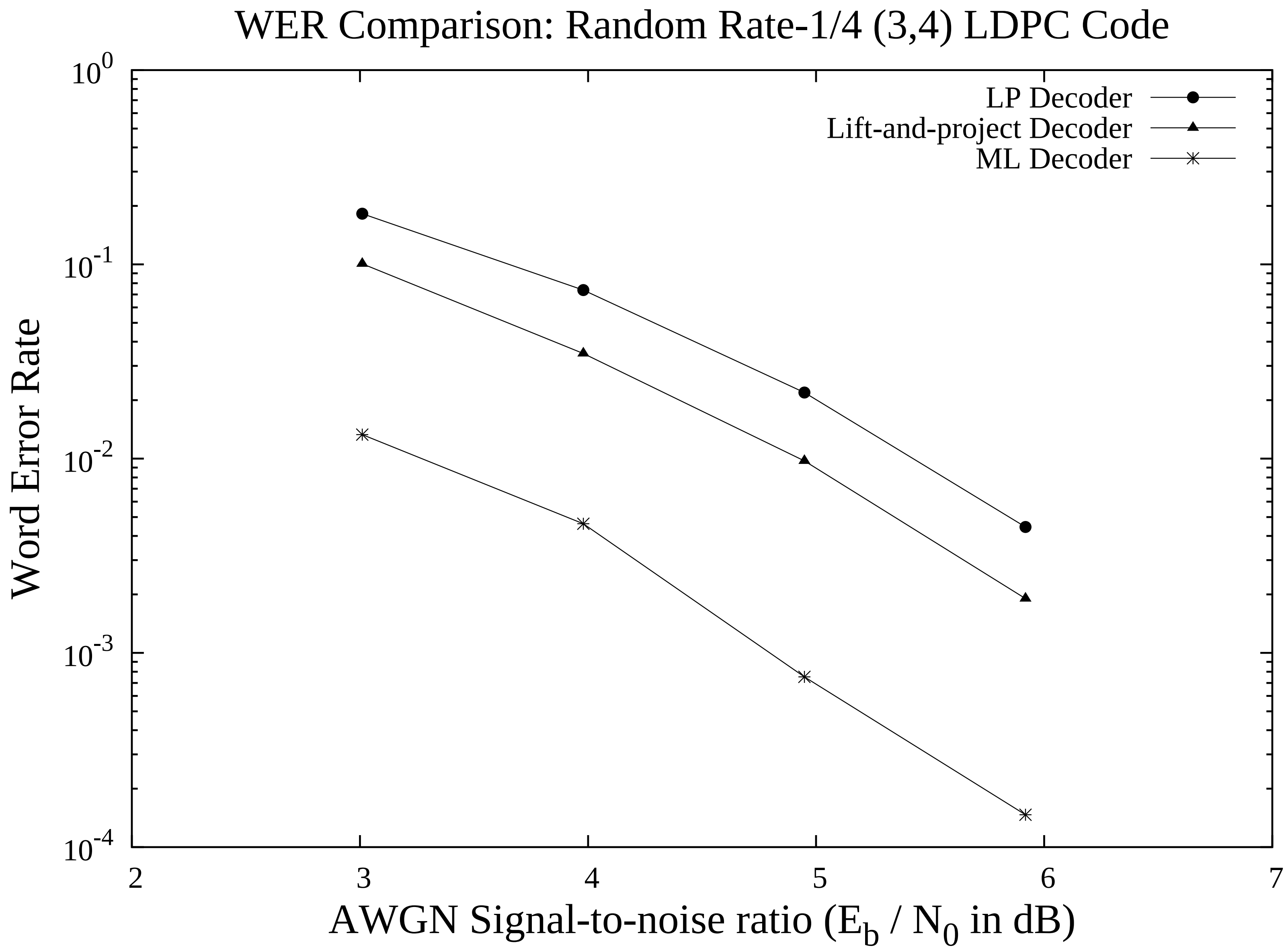
Tightening the Relaxation

- If constraints are added to the polytope, the decoder can only improve.



- Generic tightening techniques [LS ’91] [SA ’90].

Using Lift-And-Project

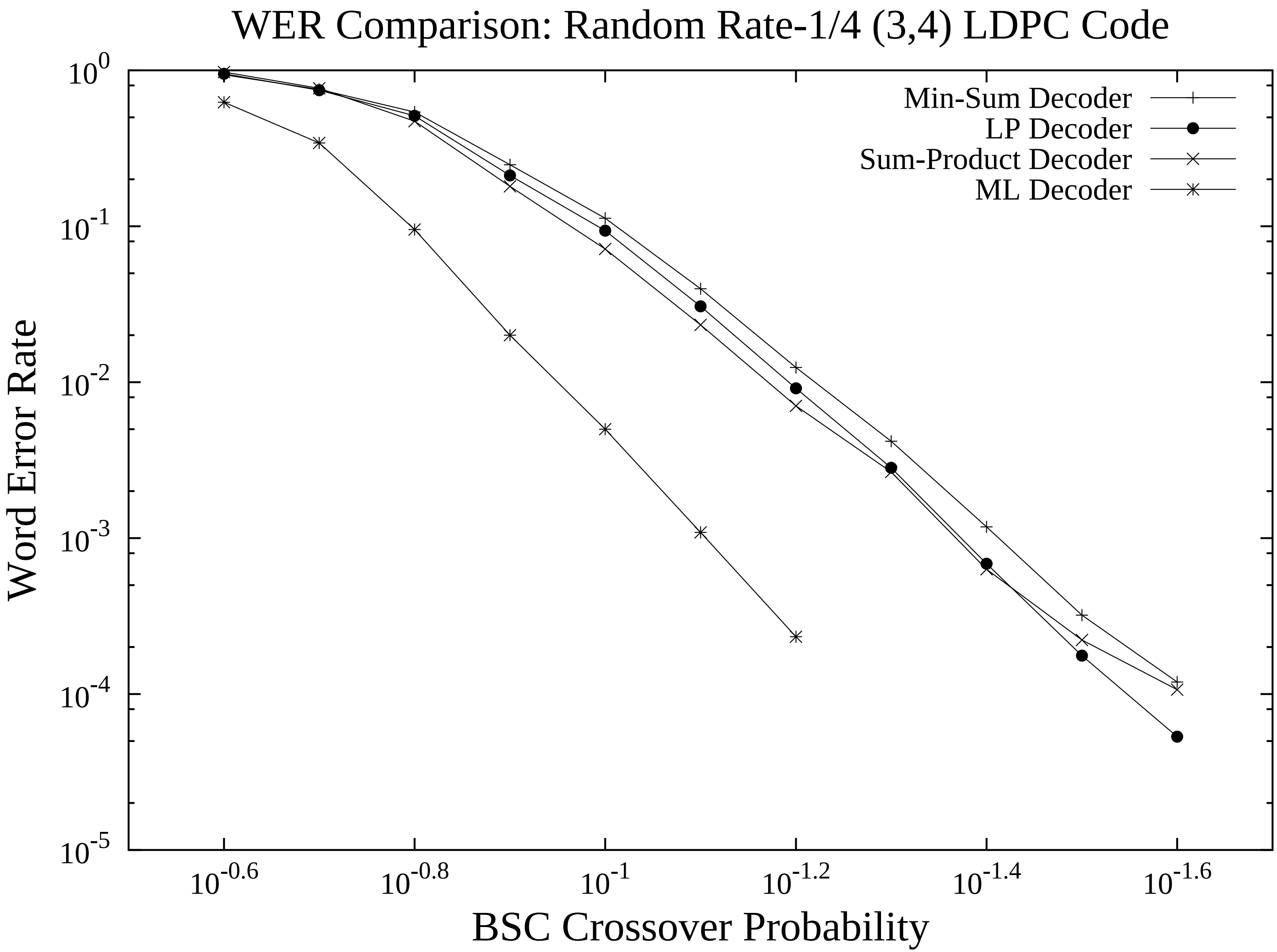


- Length 36, left degree 3, right degree 4.

Future Work

- New PCW-based performance bounds for turbo/LDPC polytopes?
  - Better turbo codes (rate-1/3 RA);
  - Other LDPC codes.
- New (better?) polytopes for turbo/LDPC codes?
- Using "lifting" procedures (generic, specialized) to tighten relaxation?
- Deeper connections to "sum-product" (belief-prop)?
- Improved running time over simplex/ellipsoid algorithm?
- LP decoding of new code families, channel models?

Performance Comparison



- Length 60, left degree 3, right degree 4.