LP Decoding Achieves Capacity

Jon Feldman Cliff Stein

Columbia University

Thank you: Alexander Barg, David Karger, Ralf Koetter,

Tal Malkin, Rocco Servedio, Pascal Vontobel,

Martin Wainwright, Gilles Zémor.

Codes, Noisy Channels, Achieving Capacity

$$x \in \{0,1\}^k$$
$$\downarrow$$
Encoder
$$\downarrow$$
$$y \in C \subseteq \{0,1\}^n$$
$$\downarrow$$
Noise
$$\downarrow$$
$$\hat{y} \in \mathbb{R}^n$$
$$\downarrow$$
Decoder
$$\downarrow$$
$y' \in C$ ($y' = y$?)

- Noise: probabilistic (Gaussian).
- Word Error Rate (WER) =

$$\Pr_{\text{noise}}\left[y' \neq y\right]$$

- Channel "capacity" $\mathcal{C}$: highest rate ($r = k/n$) s.t. $\exists$ code family, decoder, with WER $\leq 2^{-\Omega(n)}$.
- Shannon: characterized capacity for many channels.
- "Achieving capacity:" code family, decoder, with WER $\leq 2^{-\Omega(n)}$ for all $r < \mathcal{C}$.

Achieving Capacity with Poly-time Decoders

- Forney ('66):
  - ML/BD decoder
  - Concatenated codes (OPT $\diamond$ Reed-Solomon)
- Barg/Zémor ('02):
  - ML/message-passing decoder
  - Expander codes [SS '96][BZ '01-'04][GI '01-'04]
- This paper:
  - Linear Programming (LP) decoder
  - Same expander codes as Barg/Zémor
  - New feature: "Maximum-Likelihood (ML) Certificate" property: if codeword output, it maximizes Pr[correct].

Application to Practical Codes

- Turbo [BGT ’93], low-density parity-check (LDPC) [Gal ’63] codes:
  - Practical construction/encoding, moderate length
  - Perform well (experimentally) under message-passing decoder
- Most successful theory: density evolution [RU, LMSS, RSU, BRU, CFDRU, ..., ’99...present].
  - Non-constructive, assumes “local tree” structure.
- LP decoding bounds:
  - No tree assumption.
  - Bounds relevant for finite lengths.
  - Performance of message-passing is closely related [FKW 02, Fel02, KV02, KV04a, KV04b].

Maximum-Likelihood (ML) Decoding

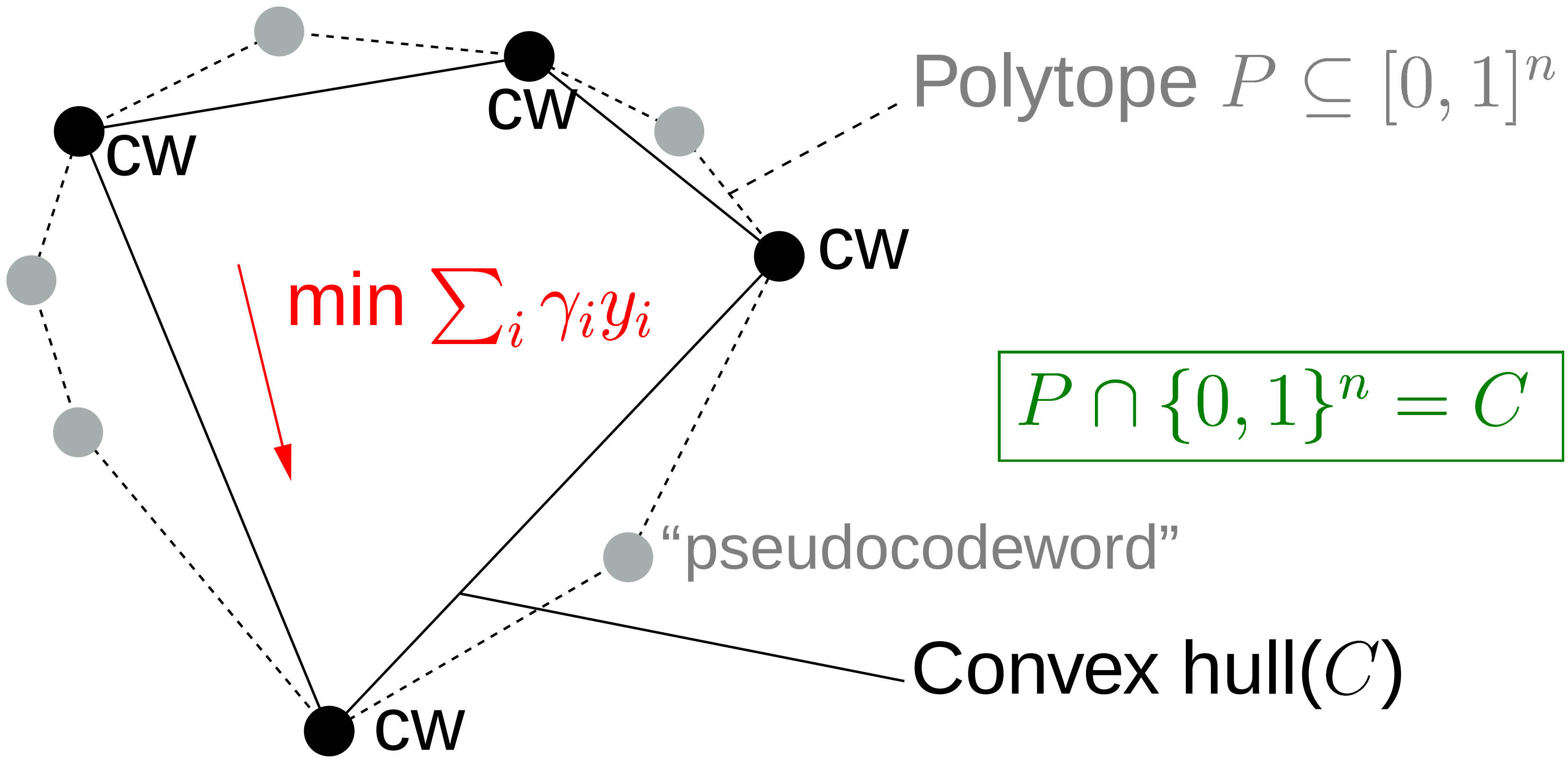
- "Memoryless" channel: each y_i is affected by noise indpendently.
  - Example: Gaussian, where $\hat{y}_i = \mathcal{N}(2y_i - 1, \sigma^2)$.
- Cost function γ_i: *log-likelihood ratio* of $\hat{y}_i$.

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

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

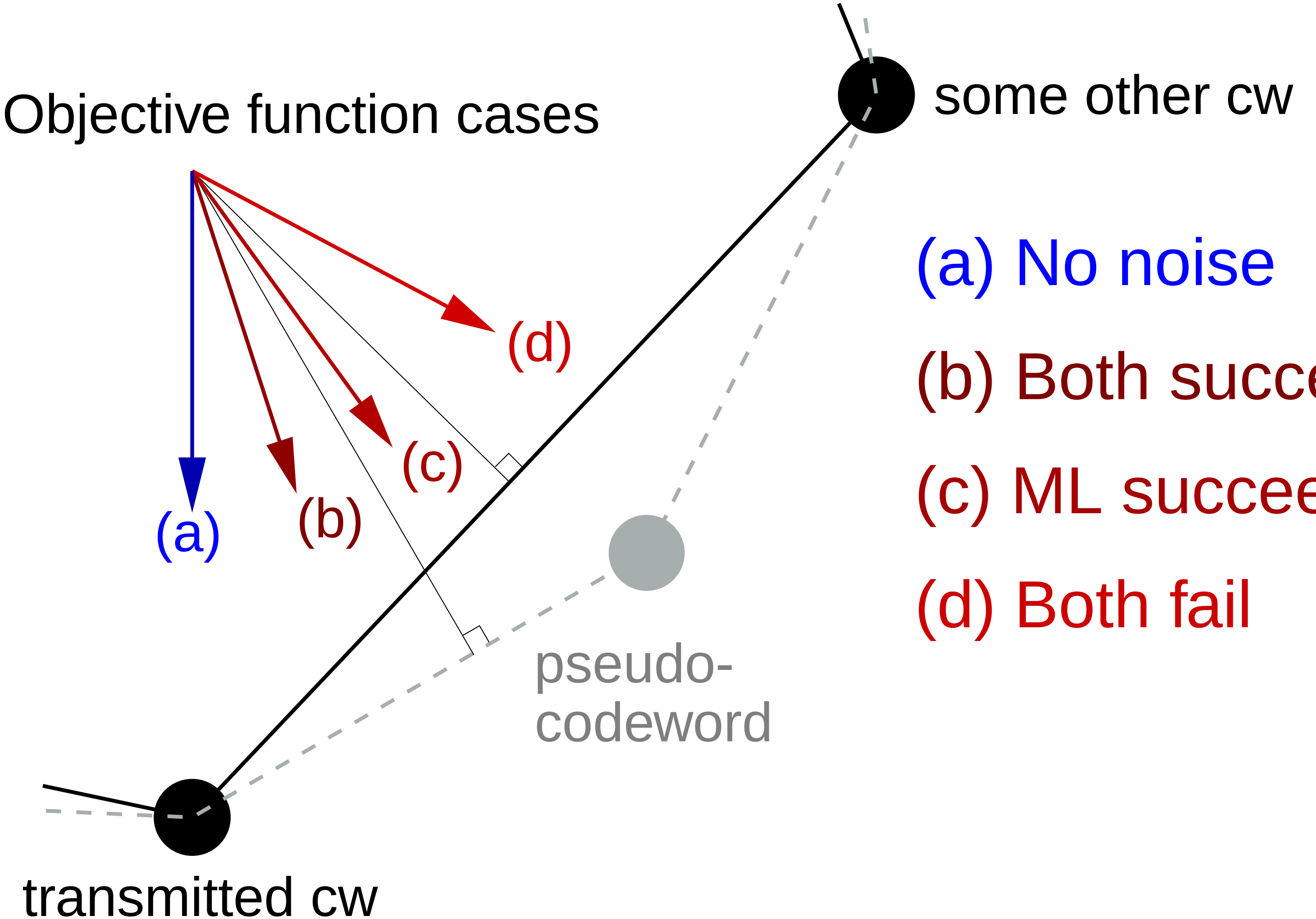
ML DECODING: Given corrupt codeword $\hat{y}$,
find $y \in C$ such that $\sum_i \gamma_i y_i$ is minimized.

LP Decoding [FK '02, Fel '03]



- LP variables y_i for each code bit, relaxed $0 \leq y_i \leq 1$.
- Alg: Solve LP. If y^* integral, output y^*, else "error."
- *ML certificate* property

LP Decoding Success Conditions



(a) No noise

(b) Both succeed

(c) ML succeed, LP fail

(d) Both fail

Using a Dual Witness to Prove Success [FMSSW '04]

- Assume 0^n is transmitted (polytope symmetry); assume unique LP optimum (no problem).

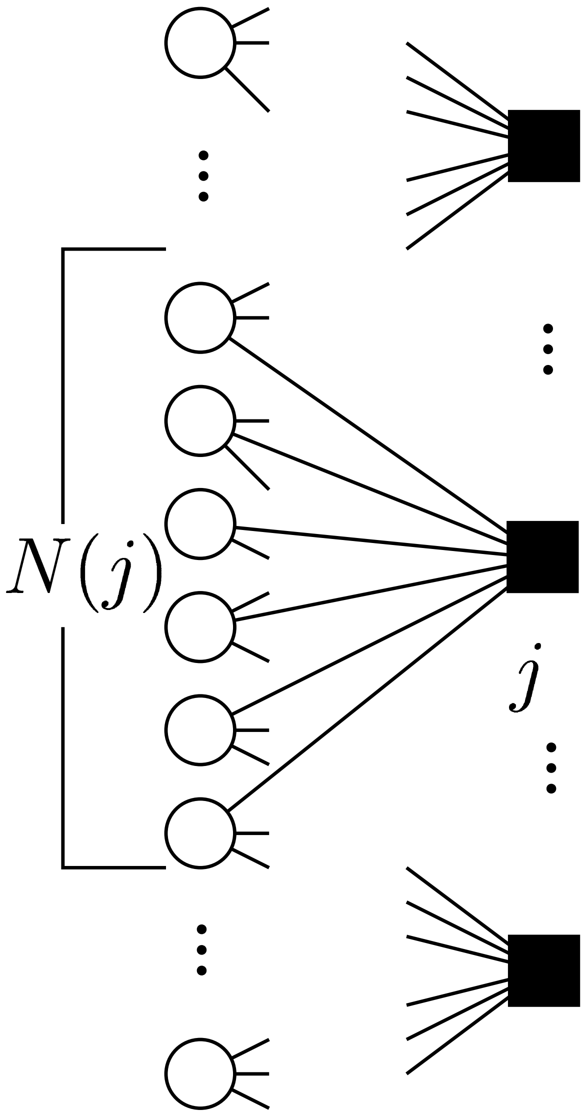
  $$\text{success} \iff \text{Point } 0^n \text{ is LP optimum}$$
  $$\iff \exists \text{ dual feasible point w/ value } 0$$

- Take LP dual, set dual objective = 0: polytope $\hat{P}$.

  $$\text{success} \iff \hat{P} \text{ non-empty}$$

- Buys "analytical slack:"
  - With no noise $\Longrightarrow$ $\hat{P}$ is large.
  - Noise increases $\Longrightarrow$ $\hat{P}$ shrinks.
  - Trade off strength of result with ease of analysis.
  - Prove result for LDPC codes, adversarial channel [FMSSW '04]

Tanner Graph Codes



- Bipartite "Tanner" graph G:
  - n "variable" ○ nodes
  - m "check" ■ nodes
  - Subcode C_j for each check j.
  - Overall codeword: setting of bits to var nodes s.t.:
    - $\forall j$, bits of $N(j)$ in code C_j.
  - LDPC codes: special case w/ const. degree, C_j = single parity check code.
- Ex: G is (3,6)-regular, C_j = $\{000000, 111000, 000111, 111111\}$.

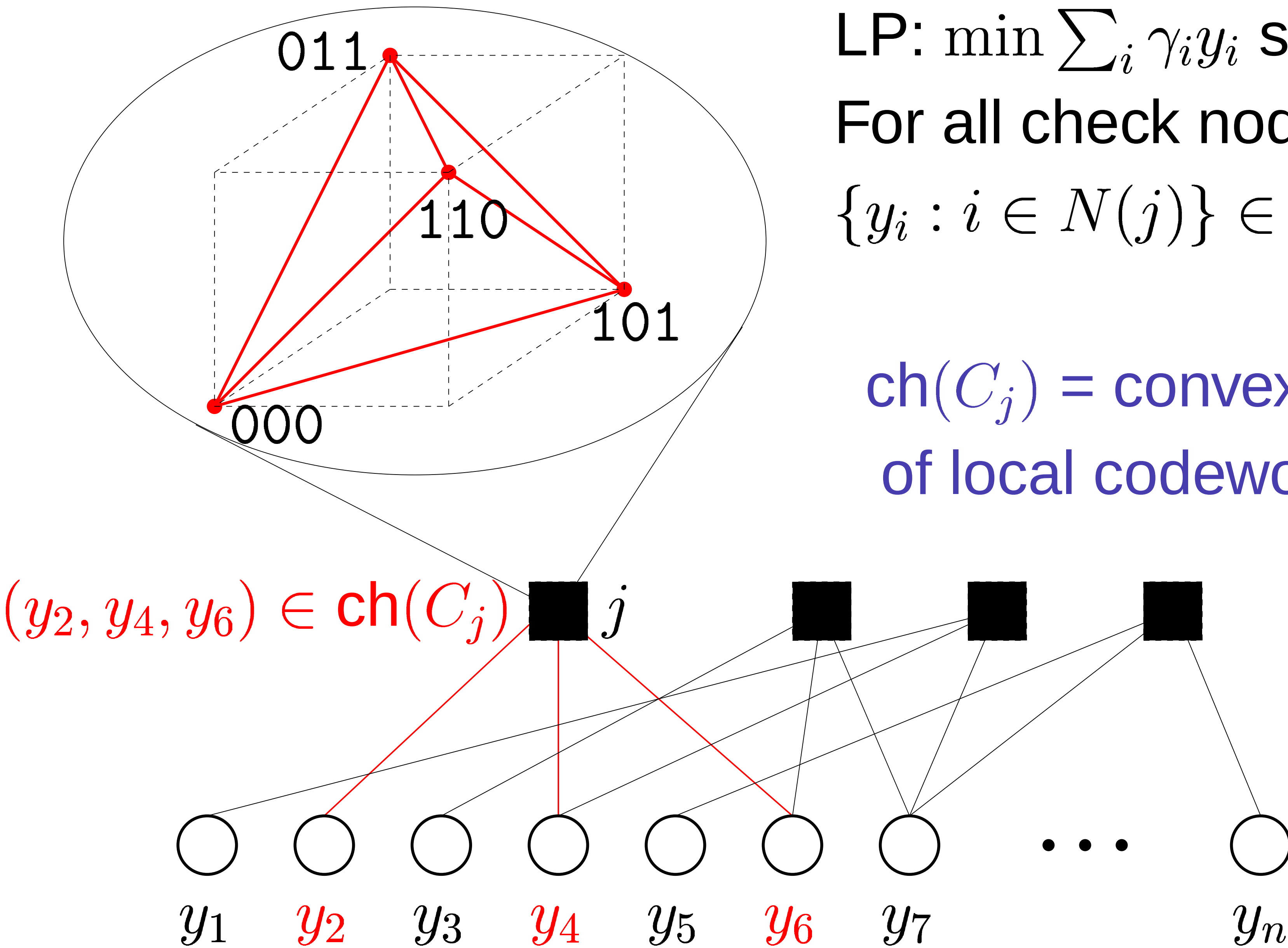
LP Relaxation for Tanner Codes [FWK ’03]

LP: $\min \sum_i \gamma_i y_i$ s.t.

For all check nodes j,

$\{y_i : i \in N(j)\} \in \text{ch}(C_j)$.

ch(C_j) = convex hull of local codewords.

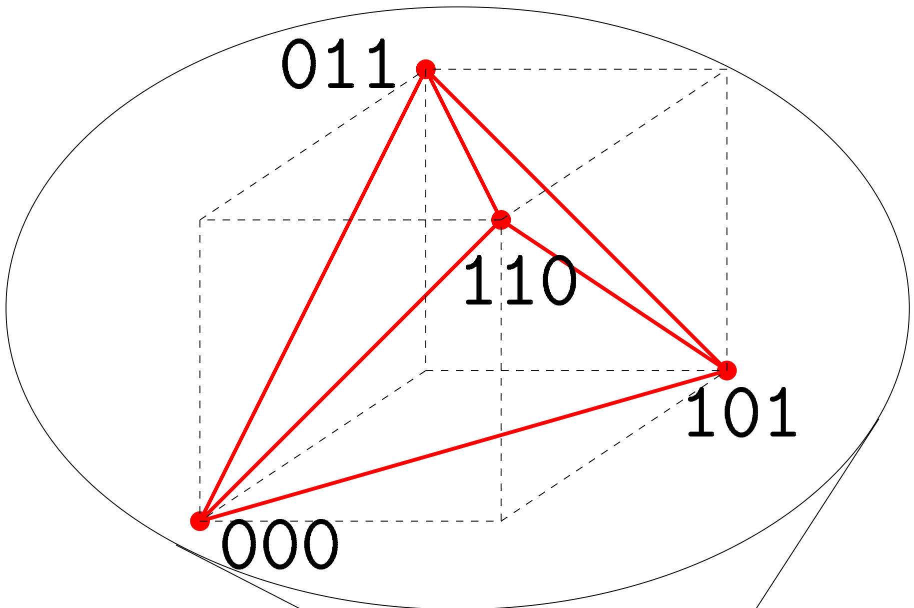


LP Relaxation for Tanner Codes

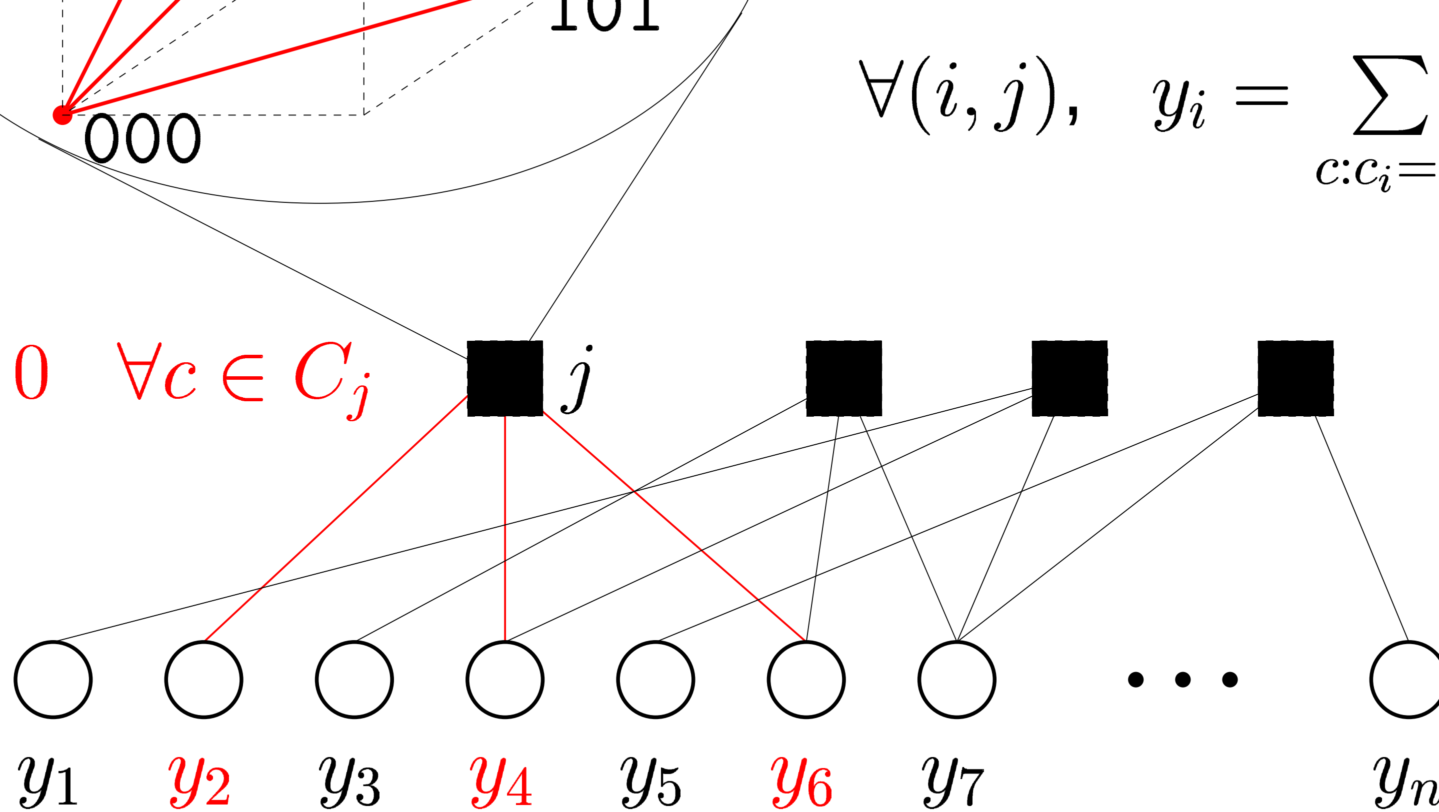
LP: $\min \sum_i \gamma_i y_i$ s.t.

$$\forall j, \qquad \sum_{c \in C_j} w_{j,c} = 1$$

$$\forall (i,j), \quad y_i = \sum_{c : c_i = 1} w_{j,c}$$

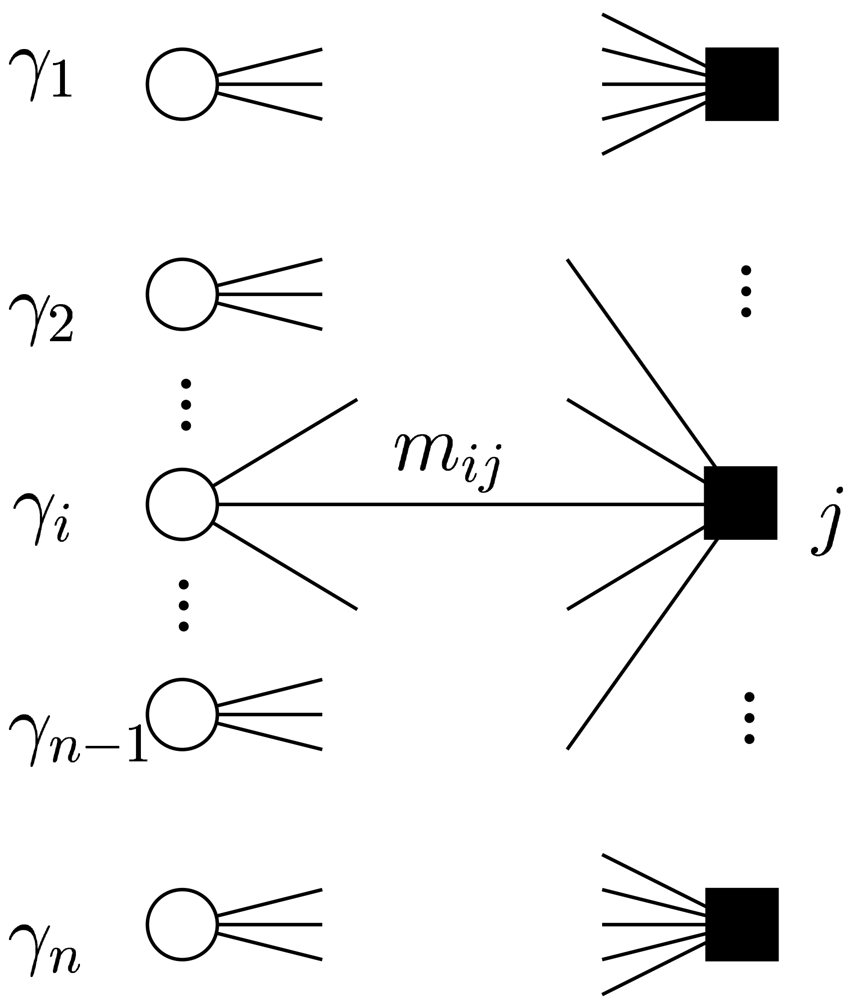


$$w_{j,c} \geq 0 \quad \forall c \in C_j$$



Dual Polytope: Tanner Codes

- Polytope $\hat{P}$ for general Tanner codes:



- Edge weights m_{ij}.
- For all code bits (left nodes) i,

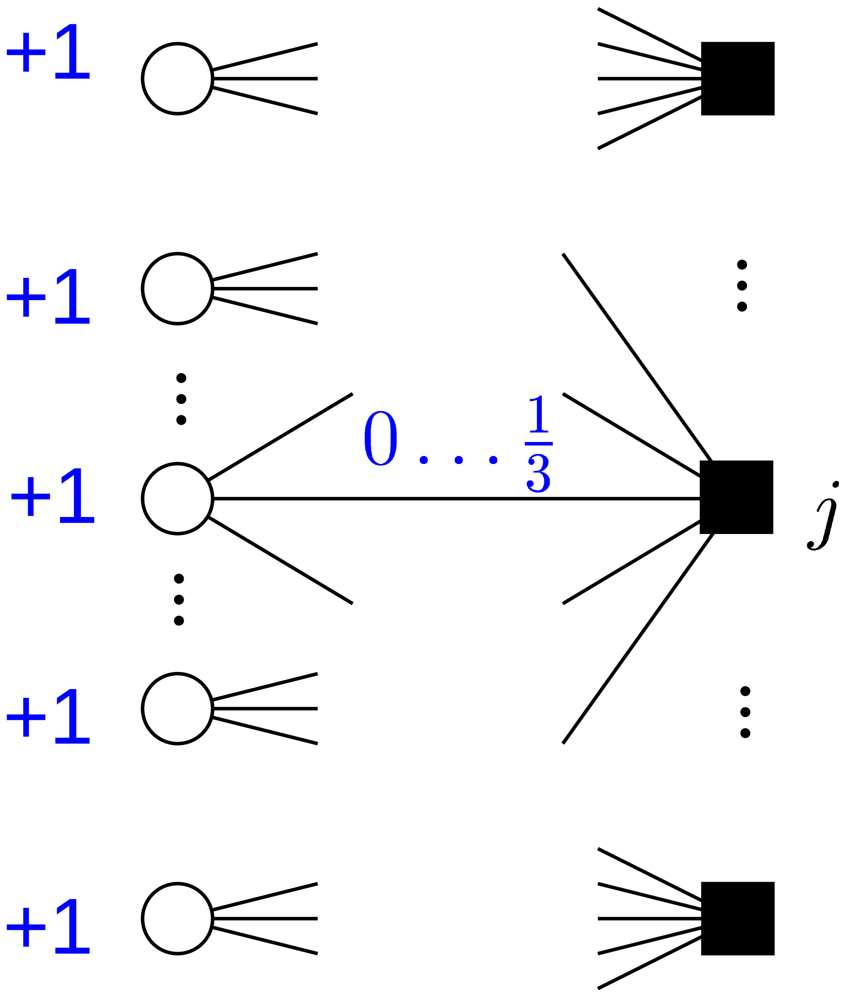
$$\sum_{j \in N(i)} m_{ij} \leq \gamma_i.$$

- For all checks j, codewords $c \in C_j$,

$$\sum_{i \in \text{sup}(c)} m_{ij} \geq 0$$

Dual Polytope: Tanner Codes

- No noise in the binary symmetric channel...



- Edge weights m_{ij}.
- For all code bits (left nodes) i,

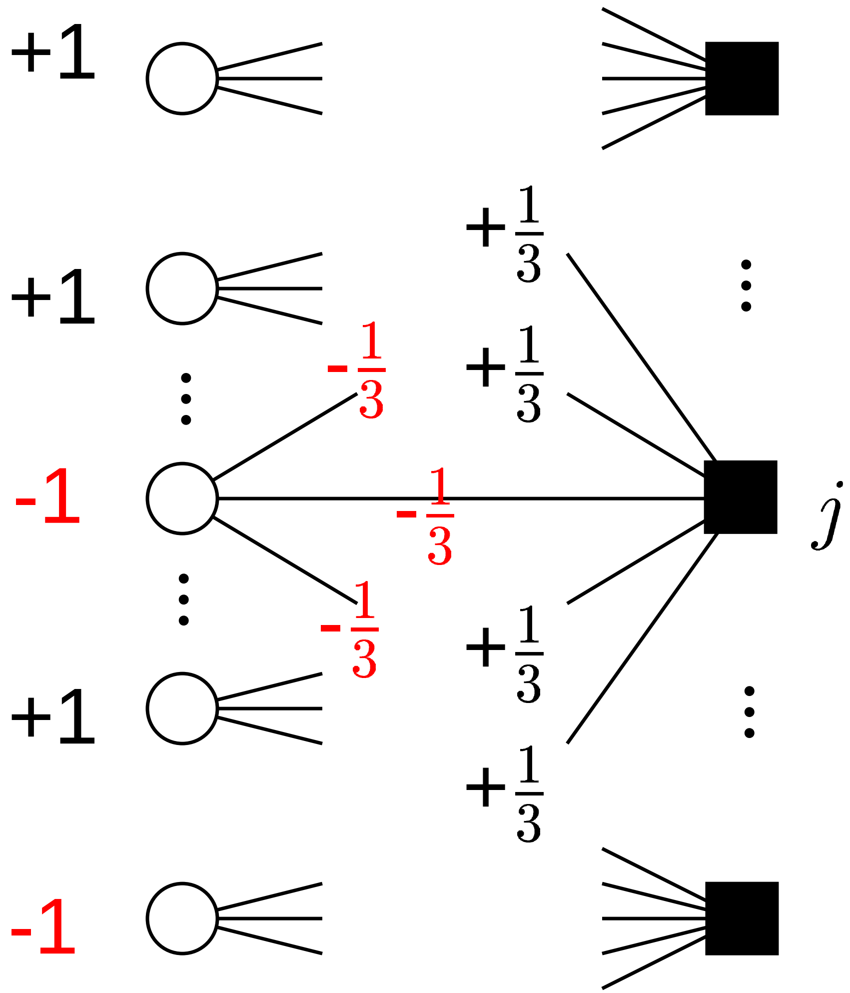
$$\sum_{j \in N(i)} m_{ij} \le \gamma_i.$$

- For all checks j, codewords $c \in C_j$,

$$\sum_{i \in \mathrm{sup}(c)} m_{ij} \ge 0$$

Dual Polytope: Tanner Codes

- A bit of noise...



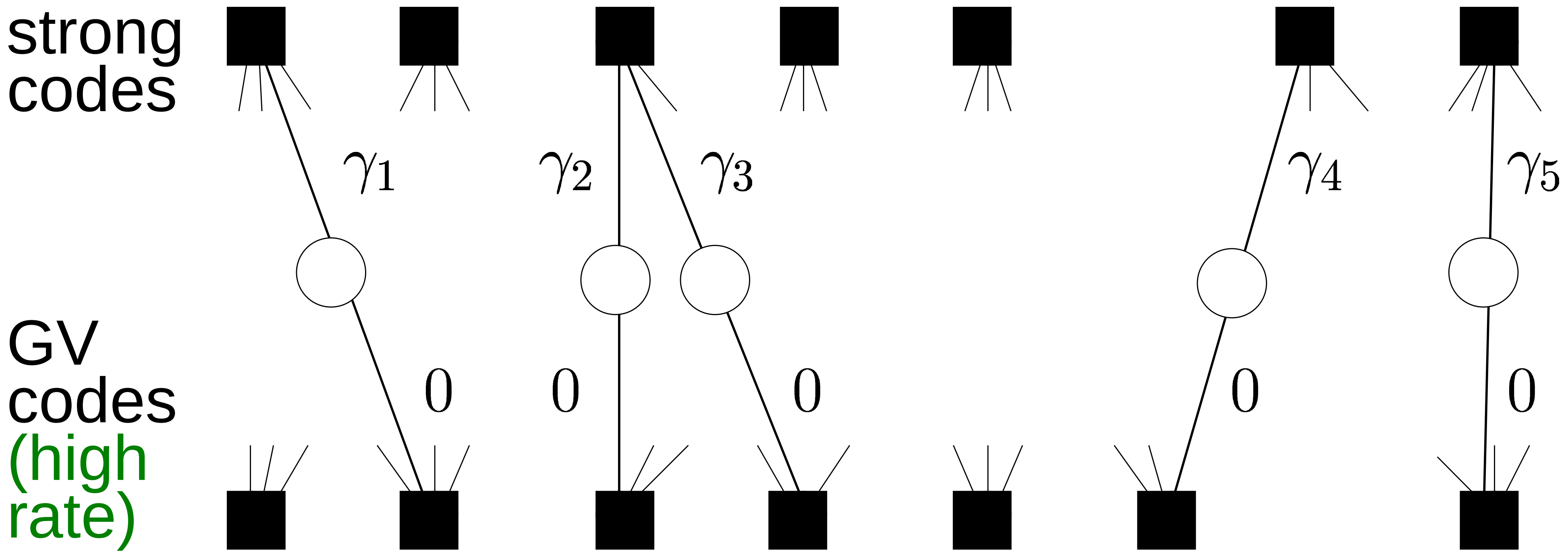
- Edge weights m_{ij}.
- For all code bits (left nodes) i,

$$\sum_{j \in N(i)} m_{ij} \leq \gamma_i.$$

- For all checks j, codewords $c \in C_j$,

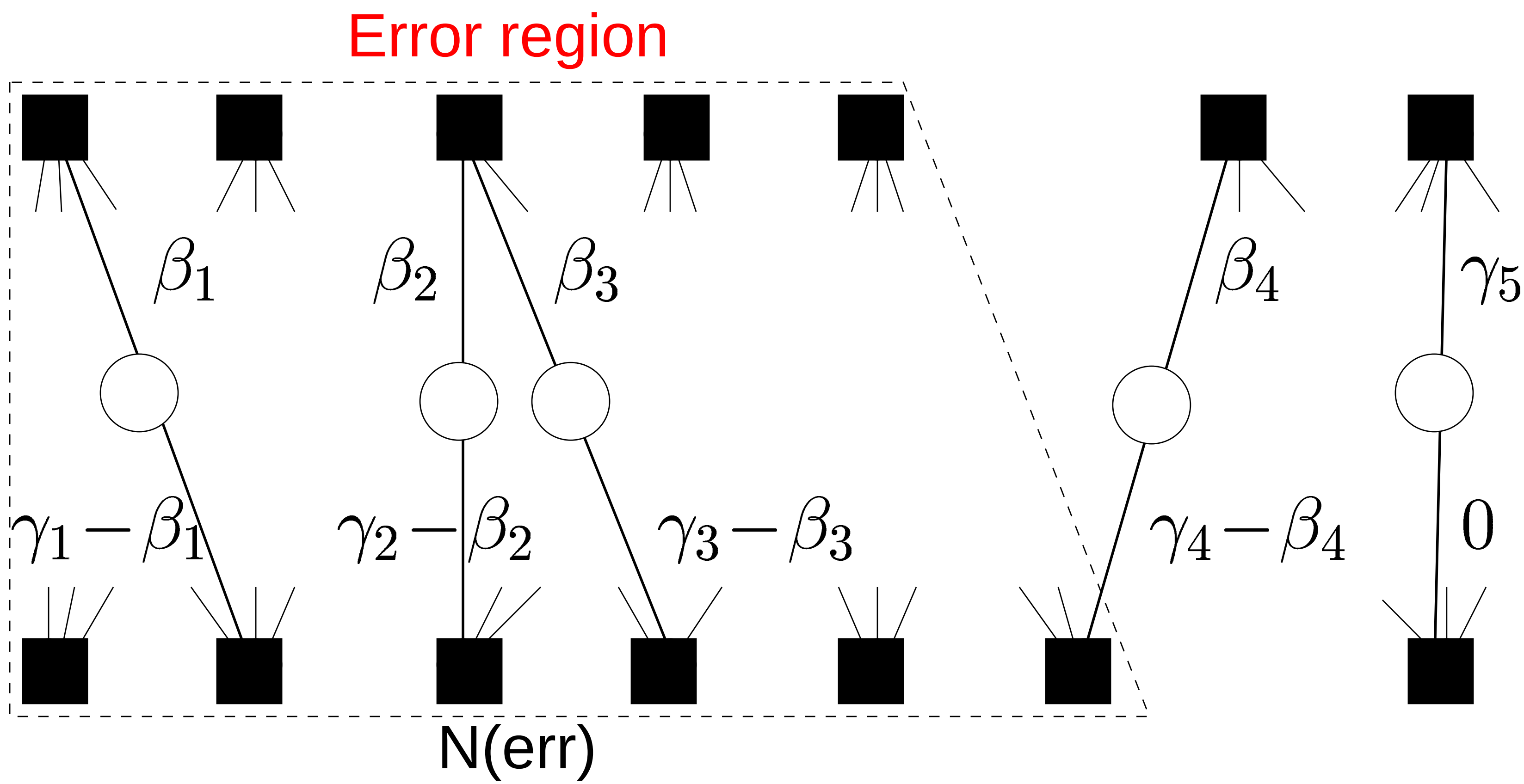
$$\sum_{i \in \text{sup}(c)} m_{ij} \geq 0$$

Using Expansion to Set Edge Weights



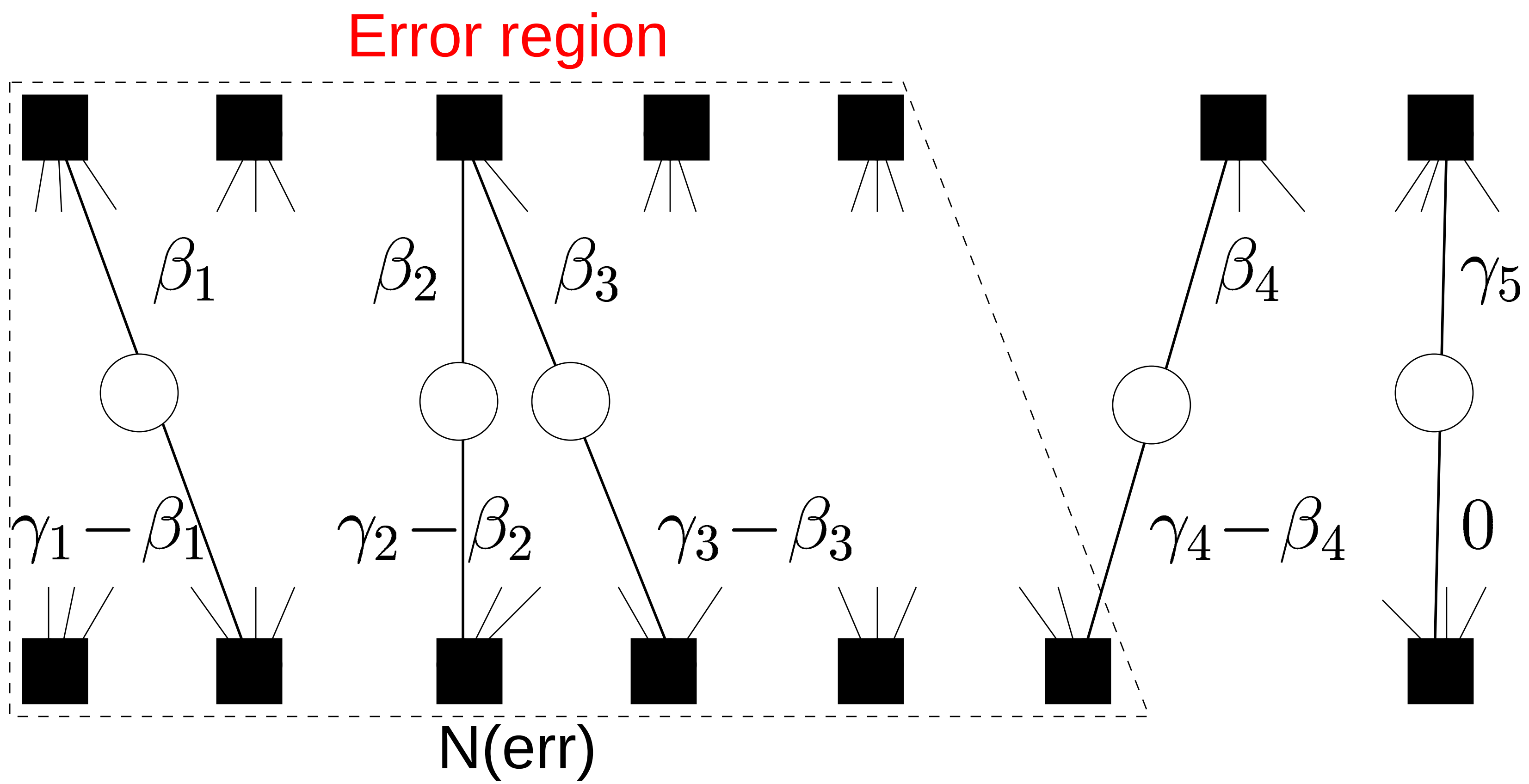
- Code built from Ramanujan graph [SS '96, BZ '02].
- Top nodes: strong codes, rate $r + \epsilon$
- Bottom nodes: GV-bound codes, rate $1 - \epsilon$
- Initial weighting: all "top" edges $m_{ij} = \gamma_i$, bottom edges $m_{ij} = 0$.

Using Expansion to Set Edge Weights



- "Error region:" checks w/ violated dual constraints.
- Use expansion to spread out excess weight among neighbors... details interesting but omitted.
- Maintain $m_{ij} + m_{ij'} = \gamma_i$, check constraints.

Using Expansion to Set Edge Weights



Thm: For any memoryless symmetric LLR-bounded channel with capacity $\mathcal{C}$, and any rate $r < \mathcal{C}$, there exists a code family of rate r for which the word error rate of LP decoding is $2^{-\Omega(n)}$.

Future work: LP Decoder Applications/Extensions

- Turbo codes: natural “flow-like” LP [FK ’02][Fel ’03]
  - RA(1/2) codes: WER $= n^{-\Omega(1)}$ [FK ’02][EH ’03]
  - Can we get WER $= 2^{-\Omega(n^{\epsilon})}$? (e.g., rate 1/3 RA)

Future work: LP Decoder Applications/Extensions

- Turbo codes: natural "flow-like" LP [FK '02][Fel '03]
  - RA(1/2) codes: WER $= n^{-\Omega(1)}$ [FK '02][EH '03]
  - Can we get WER $= 2^{-\Omega(n^{\epsilon})}$? (e.g., rate 1/3 RA)

- Tighter relaxation (lift and project)?
- ML decoding using IP (branch and bound)?
- Deeper connections to message-passing?
- More efficient algorithm to solve LP?

Major Open Questions

- Achieve capacity, decoding time poly$(n, \frac{1}{\mathcal{C}-r})$.
  - All previous capacity results:
    Explicit ML decoding of poly$(\frac{1}{\mathcal{C}-r})$-length codes.
  - This result:
    Representation of ch(poly$(\frac{1}{\mathcal{C}-r})$-length codes)

Major Open Questions

- Achieve capacity, decoding time poly$(n, \frac{1}{\mathcal{C}-r})$.
  - All previous capacity results:
    Explicit ML decoding of poly$(\frac{1}{\mathcal{C}-r})$-length codes.
  - This result:
    Representation of ch(poly$(\frac{1}{\mathcal{C}-r})$-length codes)

- Achieve near-capacity rates for LDPC codes without the tree assumption.
  - "Small cycles are bad": fact of fiction?
  - Expansion: to limited?