Spatially-Coupled LDGM/LDPC codes for Write-Once Memories

Santhosh Kumar[†], **Avinash Vem**[†], Henry D. Pfister[‡], Krishna Narayanan[†]

Texas A&M University[†], Duke University[‡]

Write-Once Memory

- In flash memory, changing a cell from 1 to 0 is easy. 0 to 1 requires rewriting whole block
- Write-once memory(WOM) models such storage system
- Given n memory units with some given state $\{0,1\}^n$, store a message $\in \{1, 2, ...2^{nR}\}$
- $0 \longrightarrow 1$ is allowed. $1 \longrightarrow 0$ is forbidden
- Referred to as rate-R WOM code

Capacity Region

- In 1985, Heegard gave the capacity for t-write system with no read or write errors.
- For the 2-write system it is

$$\Big\{(R_1, R_2) \mid 0 \le R_1 < h(\delta),\ 0 \le R_2 < 1 - \delta\Big\}$$

- WOM with read errors - Message is decoded from a noisy version of the stored vector $Y = X + \text{Ber}(p)$.
- Capacity region is not known.

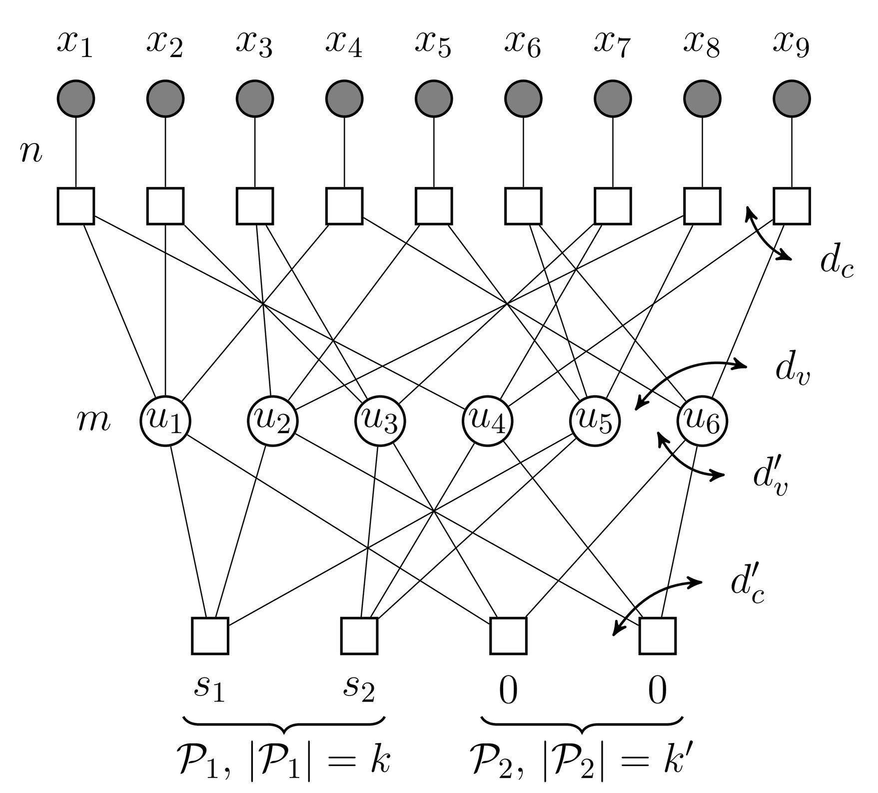
Main Result

- Objective is to construct low complexity encoding and decoding schemes that achieve the capacity region of WOM system.
- Focus mainly on 2-write WOM system
  - We achieve the capacity of the 2-write noiseless WOM system.
  - For WOM system with read errors, achievable rate region is

$$R_1 < h(\delta) - h(p), \qquad R_2 < 1 - \delta - h(p).$$

- Extension to multi-write WOM system seems possible with BPGD.

Compound LDGM/LDPC Codes



- $(n, m - k - k')$ code .
- Message constraints

$$u_1 \oplus u_2 \oplus u_5 = s_1, \qquad u_1 \oplus u_3 \oplus u_6 = 0$$

- Codeword $(x_1, \cdots, x_9)$:

$$x_1 = u_1 \oplus u_4, \qquad x_2 = u_1 \oplus u_3 \cdots$$

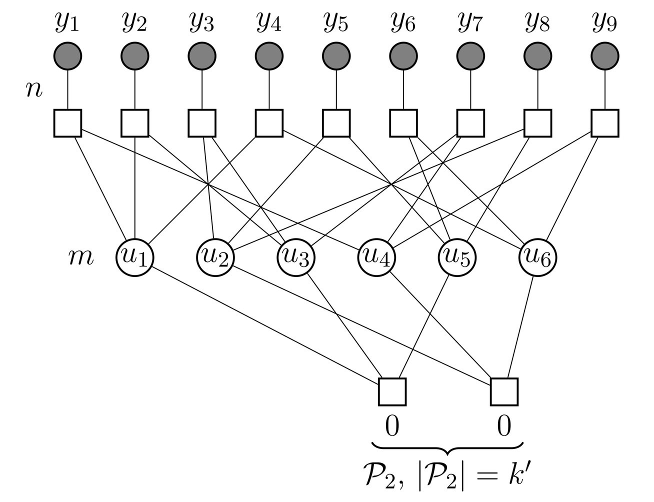
- Parametrized by s^k: $\mathcal{C}(s^k)$
- A natural coset decomposition: $\mathcal{C} = \bigcup_{s^k \in \{0,1\}^k} \mathcal{C}(s^k)$
- "Good" source code under optimal encoding
  - $\exists$ a code of rate $R = 1 - h(\delta) + \varepsilon$
  - Encodes $\text{Ber}(\frac{1}{2})$ source with an average Hamming distortion at most δ
- "Good" channel code under optimal decoding
  - $\exists$ a code of rate $R = 1 - h(p) - \varepsilon$
  - When used for channel coding on $\text{BSC}(p)$, message est. with error probability at most ε

Coding scheme for 2-write: First write

- With message s^k, encode 0^n to x^n (Distortion $\approx \delta$). Store x^n
- Decoder has

$$y_i = x_i \oplus \text{Ber}(p)$$

- Decode x^n using channel code $\mathcal{C}$ and compute s^k

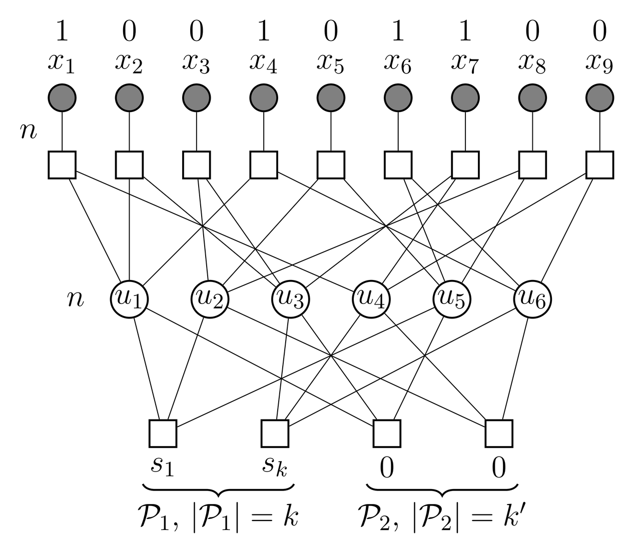


$\frac{m-k-k'}{n} \approx 1 - h(\delta)$ $\qquad$ $\frac{m-k'}{n} \approx 1 - h(p)$

- $R_1 = \frac{k}{n} \approx h(\delta) - h(p)$

Coding scheme for 2-write: Second write

- Need to find a consistent codeword in $\mathcal{C}(s^k)$



$\frac{n-k-k'}{n} \approx \delta$ $\qquad$ $\frac{n-k'}{n} \approx 1 - h(p)$

- Closely related to Binary Erasure Quantization (BEQ) (refer block below)
- To map to BEQ problem , change 0's to $*$
- With message s^k, encode seq. to $\mathcal{C}(s^k)$
- Decoder has

$$y_i = x_i \oplus \text{Ber}(p)$$

- Decode x^n using channel code $\mathcal{C}$ and compute s^k
- $R_2 = \frac{k}{n} \approx 1 - \delta - h(p)$

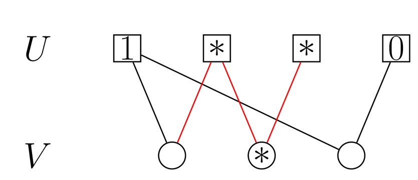
Binary Erasure Quantization

- Quantize a sequence in $\{0, 1, *\}^n$ to $x^n \in \mathcal{C} \subset \{0,1\}^n$
  - 0's and 1's should match exactly
  - $*$'s can be changed to either 0 or 1
- BEQ is the dual of decoding on binary erasure channel
  - Martinian and Yedidia'03
  - Can quan. all seq. with erasure pattern $e^n \in \{0,1\}^n$ to $\mathcal{C}$

    $\Updownarrow$

    Chan. dec. for $\mathcal{C}^\perp$ can correct all vectors with eras. $1^n \oplus e^n$
- Choose a good (dual) code $\mathcal{C}(s^k)$

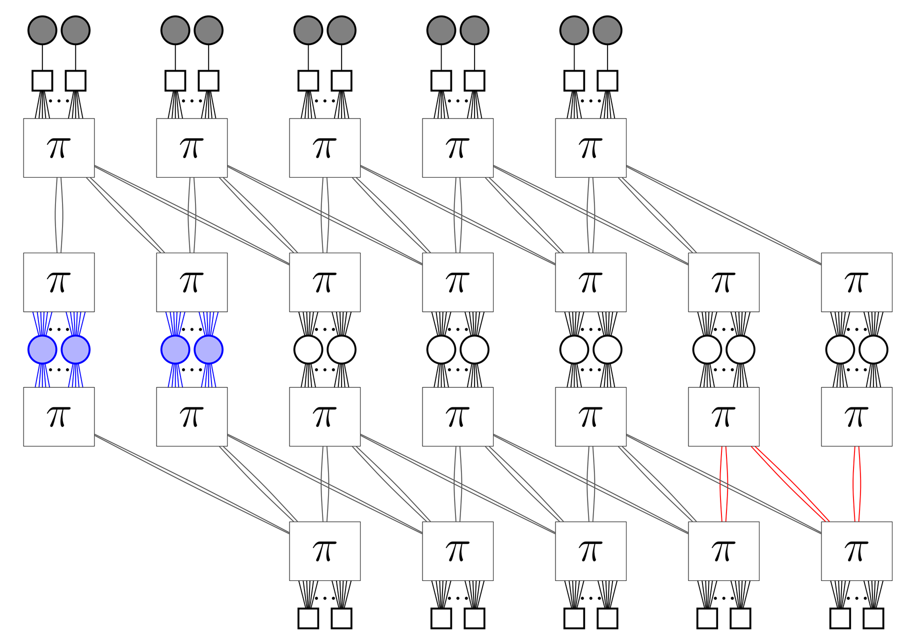
Iterative Erasure Quantization Algorithm



- Peeling type encoder

while $\exists$ non-erasures in V **do**
  if $\exists$ non-erased $u \in U$ such that only one of its neighbors $v \in V$ is not erased **then**
    Pair (u, v).
    Erase u and v.
  else
    FAIL.
    break.
  end if
end while

Spatially-Coupled Compound Codes



Numerical Results

- Noiseless WOM
  - $\delta^* = 1 - R$ is the threshold
  - $L = 30$, Single system length ≈ 30000

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	δ^*	δ $w=2$	δ $w=3$	δ $w=4$
$(4,4,3,6)$	0.500	0.461	0.491	0.492
$(4,4,4,6)$	0.333	0.278	0.323	0.325
$(4,4,5,6)$	0.167	0.086	0.155	0.159
$(5,5,3,6)$	0.500	0.436	0.488	0.491
$(5,5,4,6)$	0.333	0.260	0.320	0.324
$(5,5,5,6)$	0.167	0.079	0.154	0.159

- WOM with Read Errors
  - δ^* and p^* are the achievable thresholds
  - $L = 30$, Single system length ≈ 30000

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	w	(δ^*, p^*)	(δ, p)
$(3,3,4,6)$	3	$(0.333, 0.0615)$	$(0.321, 0.0585)$
$(3,3,4,8)$	3	$(0.500, 0.0417)$	$(0.490, 0.0387)$
$(3,3,6,8)$	4	$(0.250, 0.0724)$	$(0.239, 0.0684)$
$(4,4,4,6)$	4	$(0.333, 0.0615)$	$(0.324, 0.0585)$
$(4,4,4,8)$	4	$(0.500, 0.0417)$	$(0.492, 0.0387)$
$(4,4,6,8)$	4	$(0.250, 0.0724)$	$(0.241, 0.0694)$

Conclusion

- Spatially-coupled compound codes achieve the capacity of 2-write systems
- Coupling structure is crucial
  - to achieve optimum thresholds under practical schemes
  - also for the encoding to succeed
- Will BPGD work for multi-write systems?

References

[1] Heegard, Chris. 'On the capacity of permanent memory.' *IEEE Trans. Inform Theory*, vol. 31, no. 1,pp. 34-42,1985.

[2] E. E. Gad, Eyal En, Y. Li, J. Bruck, 'Rewriting flash memories by message passing.'in *IEEE Int. Symp. Inf. Theory*, June 2015.

[3] Wainwright, Martin J., Elitza Maneva, and E. Martinian. 'Lossy source compression using low-density generator matrix codes: Analysis and algorithms.' *IEEE Trans. Inform Theory*, vol. 56, no. 3, pp. 1351-1368, 2010.

[4] Hsu, Chun-Hao, and A. Anastasopoulos. 'Capacity-achieving codes with bounded graphical complexity and maximum likelihood decoding.' *IEEE Trans. Inform Theory* on vol. 56, no. 3, pp. 992-1006, 2010.

[5] E. Martinian and J. S. Yedidia, 'Iterative quantization using codes on graphs.' in *Proc. Allerton. Conf. Comm., and Comp.,* 2003.