On Codes with Locality

S. B. Balaji, Birenjith Sasidharan, Gaurav Agarwal and P. Vijay Kumar

Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore

Information Theory and Applications Workshop

Feb 4, 2015

Outline of the Talk

1. On Partial Maximum Recoverability
   - a simple, high-rate construction
   - some general comments

2. Hierarchical Codes with Locality
   - d_{min} bound
   - Constructions
   - an example

Related Work (first part of talk)

1. C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," NCA, 2007.
2. M. Blaum, J. L. Hafner, S. Hetzler, "Partial-MDS Codes and Their Application to RAID Type of Architectures," T-IT, July 2013.
3. J. S. Plank, M. Blaum, "Sector-Disk Erasure Codes for Mixed Failure Modes in RAID Systems," TOS 2014.
4. I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," ISIT 2013.
5. I. Tamo and A. Barg, "A family of optimal locally recoverable codes," T-IT, May 2014.
6. P. Gopalan, C. Huang, B. Jenkins, S. Yekhanin, "Explicit Maximally Recoverable Codes With Locality." T-IT Sep. 2014.

Related Work

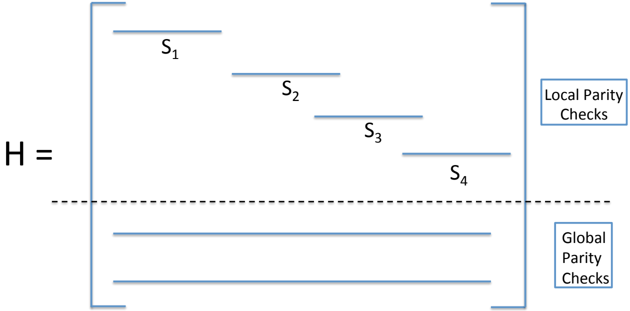
1. P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the Locality of Codeword Symbols," T-IT, Nov. 2012.
2. A. Duminuco and E. Biersack, "A practical study of regenerating codes for peer-to-peer backup systems," IEEE Int. Conf. Distributed Computing Systems, 2009.

Partial Motivation

*				
*				
*				
*		*		
*				
*			*	
*				

Codes with All-Symbol Locality

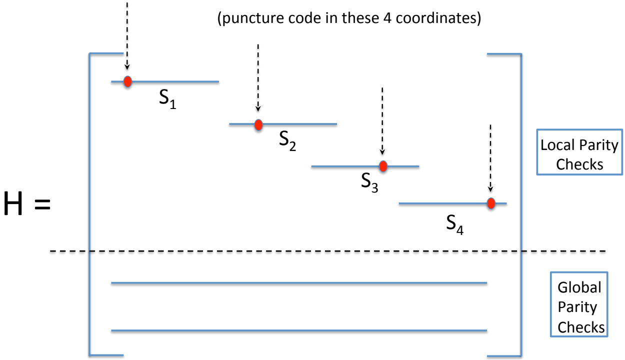
The parity-check matrix of a code with all-symbol locality linear code is of the form:



- S_i is the support of the ith local code.

Codes with All-Symbol Locality

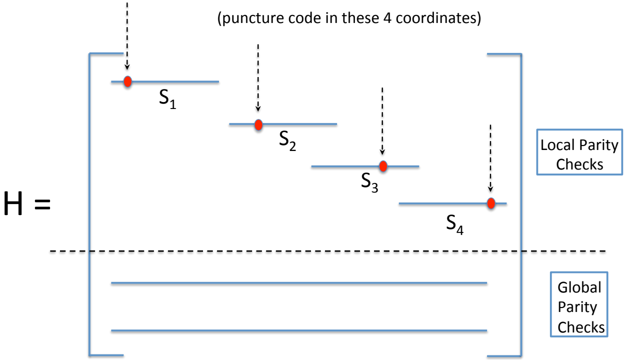
A **maximal recoverable (MR)** code is an all-symbol locality code which becomes an MDS code when one coordinate in every local code is punctured:



(the dual code is shortened).

Codes with All-Symbol Locality

A **partial, maximal recoverable (PMR)** code is an all-symbol locality code which becomes an MDS code when one particular coordinate in every local code is punctured:



(the dual code is shortened).

Canonical Form of Parity-Check Matrix of a PMR Code

The parity-check matrix of a PMR code can be put in the form:

$$H \;=\; \left[\begin{array}{c|c} I_m & F \\ \hline [0] & \underbrace{H_{\mathrm{MDS}}}_{(\Delta\times(n-m))} \end{array}\right],$$

where

1. n is the overall block length of the code
2. there are m local codes
3. the local codes have length $\leq (r+1)$
4. Δ is the number of global parities
5. H_{MDS} is the parity-check matrix of an MDS code
6. the rows of F impose locality and have Hamming weight at most r.

Minimum Distance Bound

$$H = \left[\begin{array}{c|c} I_m & F \\ \hline [0] & \underbrace{H_{\mathrm{MDS}}}_{(\Delta\times(n-m))} \end{array}\right],$$

The minimum distance is given by:

$$\begin{aligned} d_{\min} &\leq (n-k+1) - \left(\lceil \frac{k}{r} \rceil - 1\right) \\ &= \Delta + \lfloor \frac{\Delta}{r} \rfloor + 2. \end{aligned}$$

Hence

$$d_{\min} \leq \begin{cases} \Delta + 2, & \text{for } \Delta \leq (r-1) \\ \Delta + 3, & \text{for } r \leq \Delta \leq 2r-1, \text{ etc.} \end{cases}$$

A Simple, High-Rate Construction

The code $\mathcal{C}$ with parity-check matrix

$$H = \left[\begin{array}{c|c} I_m & F = \begin{bmatrix} \underline{x}_1^T & & \\ & \ddots & \\ & & \underline{x}_m^T \end{bmatrix} \\ \hline [0] & \underbrace{H_{\text{MDS}}}_{(\Delta \times (n-m))} \end{array}\right],$$

where

- $\Delta \leq (r-1)$ and
- $$\left[\begin{array}{ccc} \underline{x}_1^t & \cdots & \underline{x}_m^t \\ \hline & H_{\text{MDS}} & \end{array}\right]$$

  is a Vandermonde matrix
- achieves the bound:

  $$d_{\min} \quad \leq \quad \Delta + 2,$$

  with equality.

An Observation

- The code $\mathcal{C}$ with parity-check matrix

$$H = \left[\begin{array}{c} F \\ \hline H_{\text{MDS}} \end{array}\right],$$

  also defines a code with all-symbol locality.
- However, for the code $\mathcal{C}$ with parity-check matrix

$$H = \left[\begin{array}{c|c} I_m & F \\ \hline [0] & H_{\text{MDS}} \end{array}\right],$$

  to be optimum need that if

$$\nu := \dim\left(\text{Row}(F)\bigcap \text{Row}(H_{\text{MDS}})\right),$$

  then need that when $\Delta = ar + b$,

$$\begin{aligned} \nu &\leq \lfloor \frac{a+b-\nu}{r} \rfloor, \text{ i.e.,} \\ \nu &= 0, \quad (a+b \leq r). \end{aligned}$$

Parity-Splitting Formulation of Tamo-Barg Constructionn

Presented in polynomial form as follows:

$$f(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{k}{r}-1} a_{ij} x^{i+j(r+1)}$$

Can be shown to have parity splitting form:

$$H = \left[\begin{array}{c} F \\ \hline H_{\text{MDS}} \end{array}\right]$$

$$= \left[\begin{array}{ccccccccc} \alpha_1 & \cdots & \alpha_5 & & & & & & \\ & & & \alpha_6 & \cdots & \alpha_{10} & & & \\ & & & & & & \alpha_{11} & \cdots & \alpha_{15} \\ \hline \alpha_1 & \cdots & \alpha_5 & \alpha_6 & \cdots & \alpha_{10} & \alpha_{11} & \cdots & \alpha_{15} \\ \alpha_1^2 & \cdots & \alpha_5^2 & \alpha_6^2 & \cdots & \alpha_{10}^2 & \alpha_{11}^2 & \cdots & \alpha_{15}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_1^8 & \cdots & \alpha_5^8 & \alpha_6^8 & \cdots & \alpha_{10}^8 & \alpha_{11}^8 & \cdots & \alpha_{15}^8 \end{array}\right].$$

Parity-Splitting Formulation of Tamo-Barg Constructionn

$$
H = \left[\begin{array}{c} F \\ \hline H_{\mathsf{MDS}} \end{array}\right]
= \left[\begin{array}{ccccccccc}
\theta_1 & \cdots & \theta_5 & & & & & & \\
 & & & \theta_6 & \cdots & \theta_{10} & & & \\
 & & & & & & \theta_{11} & \cdots & \theta_{15} \\ \hline
\theta_1 & \cdots & \theta_5 & \theta_6 & \cdots & \theta_{10} & \theta_{11} & \cdots & \theta_{15} \\
\theta_1^2 & \cdots & \theta_5^2 & \theta_6^2 & \cdots & \theta_{10}^2 & \theta_{11}^2 & \cdots & \theta_{15}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_1^8 & \cdots & \theta_5^8 & \theta_6^8 & \cdots & \theta_{10}^8 & \theta_{11}^8 & \cdots & \theta_{15}^8
\end{array}\right]
$$

where the ordering is according to a cyclic subgroup and its cosets:

$$
(\theta_1 \cdots \theta_{15}) = \underbrace{(1, \alpha^3, \cdots, \alpha^{12})}_{\text{subgroup}}, \underbrace{\alpha(1, \alpha^3, \cdots, \alpha^{12})}_{\text{first coset}}, \underbrace{\alpha^2(1, \alpha^3, \cdots, \alpha^{12})}_{\text{second coset}}
$$

Parity-Splitting Formulation of Tamo-Barg Constructionn

$$
\begin{aligned}
H &= \left[\begin{array}{c} F \\ \hline H_{\text{MDS}} \end{array}\right] \\
&= \left[\begin{array}{ccccccccc}
\theta_1 & \cdots & \theta_5 & & & & & & \\
& & & \theta_6 & \cdots & \theta_{10} & & & \\
& & & & & & \theta_{11} & \cdots & \theta_{15} \\ \hline
\theta_1 & \cdots & \theta_5 & \theta_6 & \cdots & \theta_{10} & \theta_{11} & \cdots & \theta_{15} \\
\theta_1^2 & \cdots & \theta_5^2 & \theta_6^2 & \cdots & \theta_{10}^2 & \theta_{11}^2 & \cdots & \theta_{15}^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_1^8 & \cdots & \theta_5^8 & \theta_6^8 & \cdots & \theta_{10}^8 & \theta_{11}^8 & \cdots & \theta_{15}^8
\end{array}\right]
\end{aligned}
$$

However, here,

$$
\begin{aligned}
\nu &:= \dim\left(\text{Row}(F)) \bigcap \text{Row}(H_{\text{MDS}}\ )\right), \\
&\geq 1,
\end{aligned}
$$

and this particular construction cannot be used.

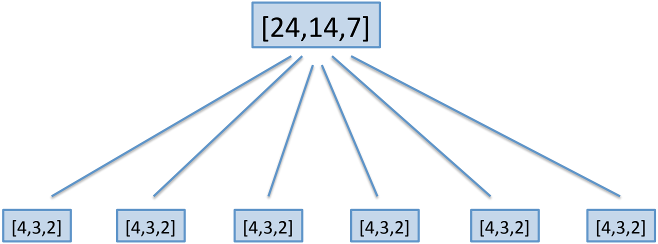
Constructions of Maximum Recoverable Codes

1. Start with a Tamo-Barg Construction for all-symbol ($r = 2$) locality
2. Retain only a small number of cosets of a cyclic group so as to obtain the desired MR property
3. Specific instances of this construction, beats other constructions in terms of field size

Codes with Hierarchical Locality

Birenjith Sasidharan, Gaurav Kumar Agarwal, P. Vijay Kumar, "Codes With Hierarchical Locality," submitted to ISIT 2015, see also arXiv:1501.06683 [cs.IT]

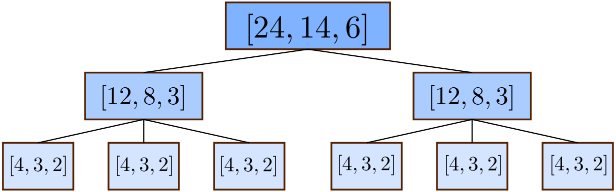
Codes with Locality do not Scale



$$d \leq (n-k+1) - \left(\lceil \frac{k}{r} \rceil - 1\right)(\delta - 1)$$

Codes with Hierarchical Locality

- Each symbol of $[n, k, d]$-code is protected by a $[n_1, r_1, \delta_1]$-code
- Each symbol of $[n_1, r_1, \delta_1]$-code is protected by a $[n_2, r_2, \delta_2]$-code

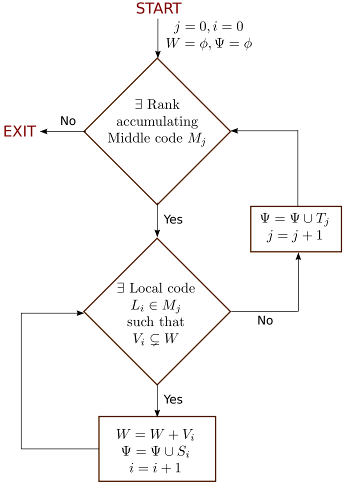


$$d \leq \underbrace{n - k + 1 - \left(\left\lceil \frac{k}{r_2} \right\rceil - 1\right)(\delta_2 - 1)}_{\text{Bound for codes with locality}} - \underbrace{\left(\left\lceil \frac{k}{r_1} \right\rceil - 1\right)(\delta_1 - \delta_2)}_{\text{Extra loss}}$$

Bound on Minimum Distance

- Find a $(k-1)$-dimensional punctured code $\mathcal{C}_s$ with a large support.
- Then, $d_{\text{min}} \leq n - Supp(\mathcal{C}_s)$.

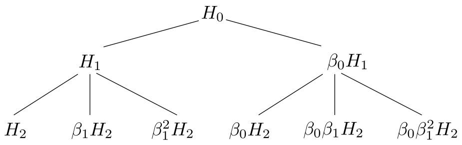
Algorithm used to identify $\mathcal{C}_s$



All-symbol Local Optimal Construction: An Example

- Need to satisfy a divisibility condition $n_2 \mid n_1 \mid n$
- Example: $[n, k] = [24, 14]$, $[n_1, r_1] = [12, 8]$, $[n_2, r_2] = [4, 3]$.

1. Choose $\mathbb{F}_{5^2}$.
2. Identify the subgroup chain: $H_2 \subset H_1 \subset H_0 = \mathbb{F}^*_{5^2}$ s.t. $|H_2| = 4$, $|H_1| = 12$, $|H_0| = 24$.
3. $H_i = \langle \beta_i \rangle$, $i = 0, 1, 2$
4. The subgroups lead to a tree of cosets.



Example Contd.

5. Construct a polynomial for every vertex of the tree of cosets. These polynomials are called lifting polynomials.
6. Polynomial $E_1(X)$ corresponding to H_1 evaluates to 1 at points from H_1, and zero at points from its siblings.

Polynomials for the first level of tree:

$$\begin{aligned} E_1(X) &= \left(\frac{X^{12} - \beta_0^{12}}{1 - \beta_0^{12}}\right), \\ E_2(X) &= \left(\frac{X^{12} - 1}{\beta_0^{12} - 1}\right) \end{aligned}$$

(can find $F_1, \ldots, F_6$ for the second level)

Example Contd.

7. Message polynomials are associated with the leaves of the tree of cosets.
8. They are lifted one-level up in the tree using the lifting polynonials. Coefficients are precoded to adjust the dimension.
9. Lifting is continued upto the root of the tree, resulting in a polynomial $c_0(X)$ associated with the root of the tree.
10. Evaluations of $c_0(X)$ at H_0 give rise to a codeword.

Exponents of Polynomials Develop Gaps that Enforce Locality

~~3~~	2	1	0

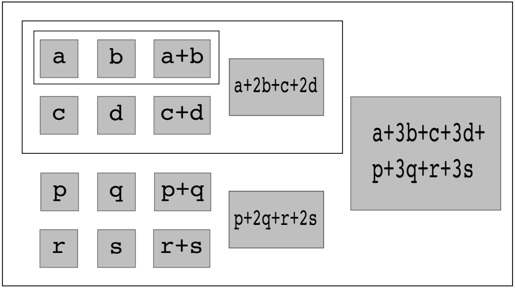
~~11~~	~~10~~	9	8	~~7~~	6	5	4	~~3~~	2	1	0

~~23~~	~~22~~	~~21~~	~~20~~	~~19~~	18	17	16	~~15~~	14	13	12
~~11~~	~~10~~	9	8	~~7~~	6	5	4	~~3~~	2	1	0

Figure: Illustration of the behaviour of exponents in $c_0(X)$.

Information-symbol Local Optimal Construction: Pyramid Codes

- $[n, k] = [15, 8]$, $[n_1, r_1] = [7, 4]$, $[n_2, r_2] = [3, 2]$.
- $\delta_2 = 2$, $\delta_1 = 3$, $d = 4$. (optimal d_{min})



Pyramid Codes (contd.)

- Consider an MDS code with parameter $[k + d - 1, k, d] = [11, 8, 4]$.
- $G_{\text{mds}} = [I_{k\times k} \mid A_{k\times(d-1)}] = [I_{8\times 8} \mid A_{8\times 3}]$.

$$G^{s}_{\text{mds}} = \left[\begin{array}{c|c|c} I_{8\times 8} & \begin{array}{c} B_{4\times 2} \\ C_{4\times 2} \end{array} & D_{8\times 1} \end{array}\right],$$

$$G^{ss}_{\text{mds}} = \left[\begin{array}{c|c|c} I_{8\times 8} & \begin{array}{c|c} \begin{array}{c} E_{2\times 1} \\ F_{2\times 1} \end{array} & G_{4\times 1} \\ \hline \begin{array}{c} H_{2\times 1} \\ J_{2\times 1} \end{array} & K_{4\times 1} \end{array} & D_{8\times 1} \end{array}\right],$$

$$G_{\text{local}} = \left[\begin{array}{c|cc|c|cc|c|c} & E_{2\times 1} & & G_{4\times 1} & & & & \\ I_{8\times 8} & & F_{2\times 1} & & & & & D_{8\times 1} \\ \cline{2-7} & & & & H_{2\times 1} & & K_{4\times 1} & \\ & & & & & J_{2\times 1} & & \end{array}\right]$$

Acknowledgement

Would like to thank Parikshit Gopalan for the introduction to the topic of maximum recoverable codes and helpful discussions.

Thanks!