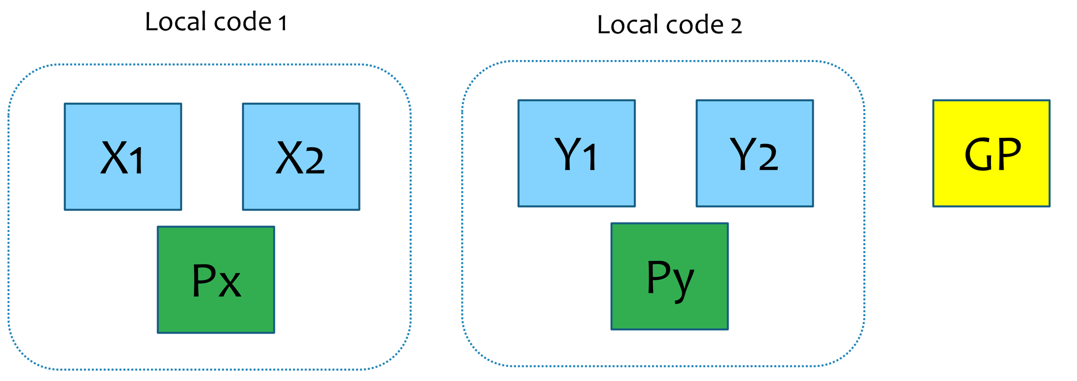
Codes with Locality for Multiple Erasures in Distributed Storage

Author: Balaji.S.B.
Joint work with : Prasanth.K.P. , Ganesh Kini
Advisor: Prof. P. Vijay Kumar

Disk failures are a challenge in distributed storage. Codes with locality can help reduce the number of disks accessed to repair a failed disk. In this poster we present bounds and code constructions for designing codes with locality primarily for correcting multiple erasures.

Codes with Locality

C: Linear [n,k] code over F_q with locality r:
Example [7,4] code with locality r=2:



X1,X2,Y1,Y2 (blue): Information symbols
Px,Py (green): Local parities
GP (yellow): Global parity

Any information symbol on erasure can be recovered by accessing at most 2 other code symbols

Maximum Recoverability

Maximum distance possible for a given block-length and dimension is achieved by codes called MDS codes. In codes with locality, we have locality constraints. Maximum Recoverable codes are those codes which are as MDS as possible along with locality constraints.

Idea : Take an existing construction of a code with locality and puncture or remove local codes one by one until it becomes maximum recoverable.

Consider the code (due to Tamo, Barg): C : [n=q-1, k=2D+1, d=q-1-3D] which is obtained by evaluating a certain set of polynomials at cosets of cube roots of unity.

Result: Given positive integers n, D with $\frac{2D}{n} < \frac{2}{3}$ and q> $\sum_{j=2}^{2D} jC\left(\frac{n}{3}-1, j\right) 3^j + n - 2$, There exists a [n,k=2D+1,r=2] maximum recoverable code over F_q that is obtained from C by puncturing a certain specific $\frac{q-1}{3} - \frac{n}{3}$ cosets from evaluating positions.

An Upper Bound on Rate and a Rate-Optimal Construction for Codes with Locality and Sequential Recovery

Codes with Locality and Sequential Recovery:

An [n, k] code is said to be a locally recoverable code with sequential recovery from t erasures, if for any set of $s \leq t$ erasures, there is an s-step sequential recovery process, in which at each step, a single erased symbol is recovered by accessing at most r other code symbols which are either un-erased or already recovered.

Rate Bound: Let C be an [n,k] code with locality r and sequential recovery from t erasures over a field F_q. Let r≥3. Then:

$$\frac{k}{n} \leq \frac{r^{t/2}}{r^{t/2} + 2\sum_{i=0}^{\frac{t}{2}-1} r^i} \quad \text{for t an even integer}$$

$$\frac{k}{n} \leq \frac{r^s}{r^s + 2\sum_{i=1}^{s-1} r^i + 1} \quad \text{for t an odd integer}$$

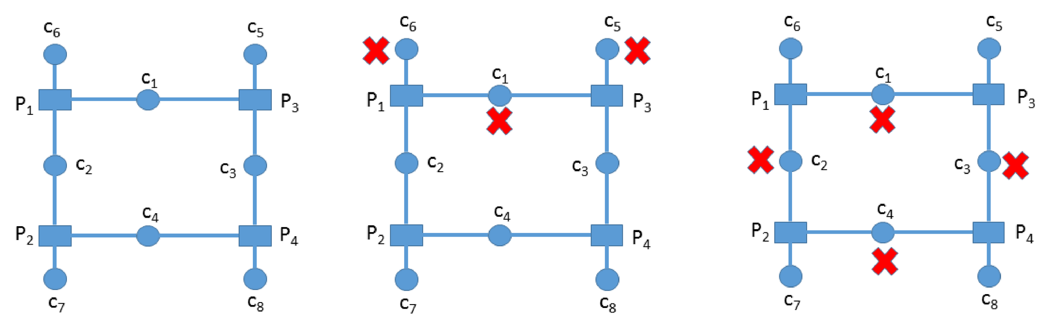
Where s=$\frac{t+1}{2}$. This theorem settles a conjecture by Song, Cai and Yuen.

Proof:

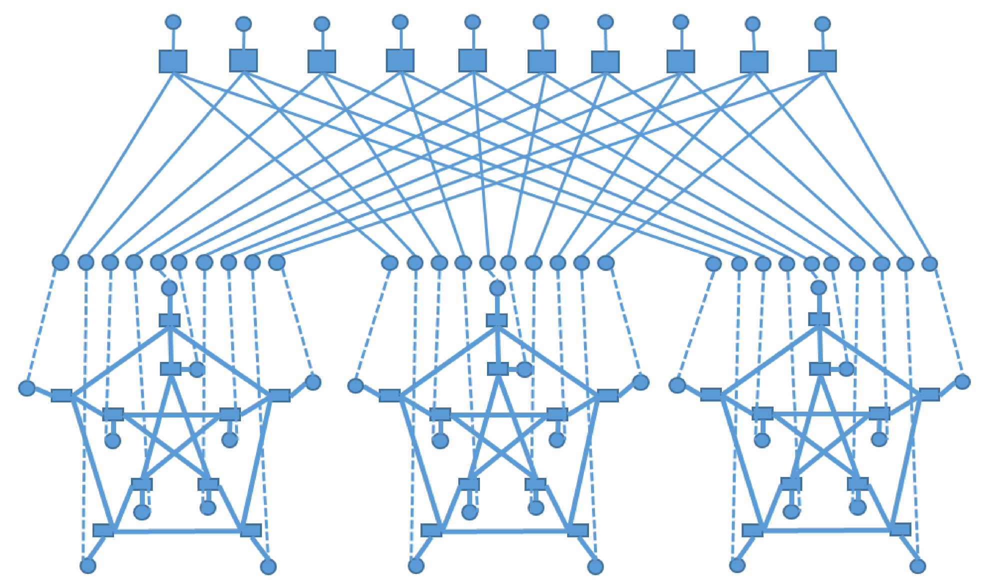
$$H = \left[\begin{array}{c|c|c|c|c|c|c|c} D_0 & A_1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \hline 0 & D_1 & A_2 & 0 & \ldots & 0 & 0 & 0 \\ \hline 0 & 0 & D_2 & A_3 & \ldots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & D_3 & \ldots & 0 & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ \hline 0 & 0 & 0 & 0 & \ldots & A_{\frac{t}{2}-2} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \ldots & D_{\frac{t}{2}-2} & A_{\frac{t}{2}-1} & 0 \\ \hline 0 & 0 & 0 & 0 & \ldots & 0 & D_{\frac{t}{2}-1} & \\ \hline 0 & 0 & 0 & 0 & \ldots & 0 & 0 & C \end{array}\right] D$$

First we observe that the parity check matrix without global parities can be written in the above form and then we observe that $\{D_i\}$ are diagonal matrices as minimum distance is at least t+1 and then we equate row and column weights of different parts of the above matrix H to get the required bound.

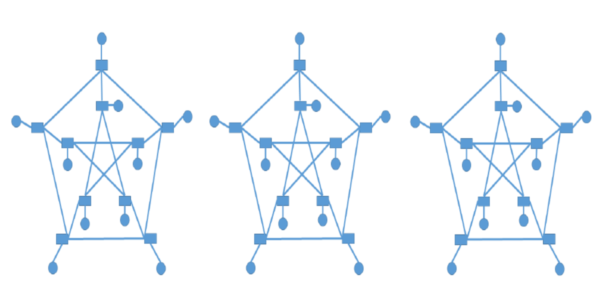
Example rate optimal construction for t=2, r=2 and need of girth and separation of deg 1 nodes:



Example rate optimal construction for t=4, r=3:

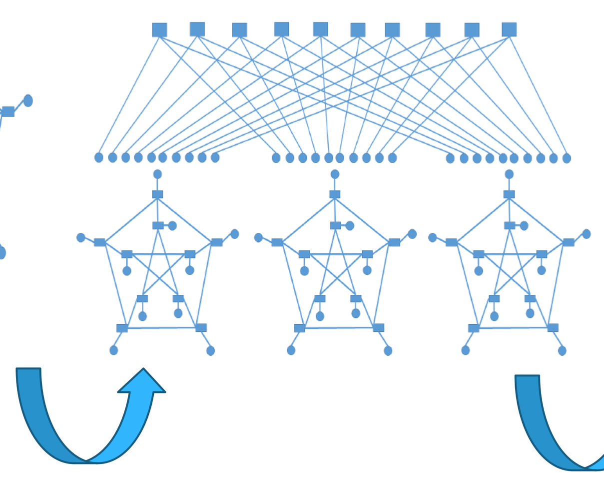


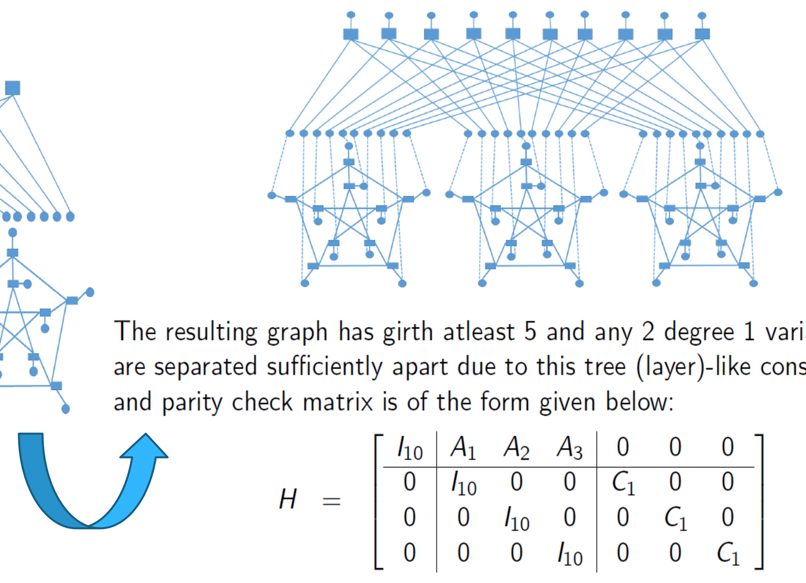
Step by Step explanation of rate optimal construction for t=4, r=3:



Now the resulting code has parity check matrix of the form:

$$H = \left[\begin{array}{ccc|ccc} I_{10} & 0 & 0 & C_1 & 0 & 0 \\ 0 & I_{10} & 0 & 0 & C_1 & 0 \\ 0 & 0 & I_{10} & 0 & 0 & C_1 \end{array}\right]$$





The resulting graph has girth atleast 5 and any 2 degree 1 variable nodes are separated sufficiently apart due to this tree (layer)-like construction and parity check matrix is of the form given below:

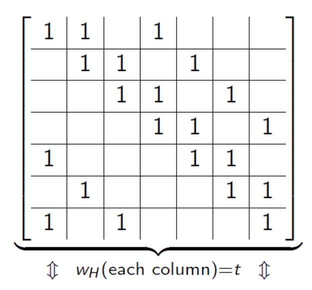
$$H = \left[\begin{array}{c|ccc|ccc} I_{10} & A_1 & A_2 & A_3 & 0 & 0 & 0 \\ \hline 0 & I_{10} & 0 & 0 & C_1 & 0 & 0 \\ 0 & 0 & I_{10} & 0 & 0 & C_1 & 0 \\ 0 & 0 & 0 & I_{10} & 0 & 0 & C_1 \end{array}\right]$$

Codes with Locality and strict Availability:

Codes with strict availability:

Null space of an m x n matrix with each row of weight r+1 and each column of weight t with cardinality of intersection of support sets of any two rows at most one.

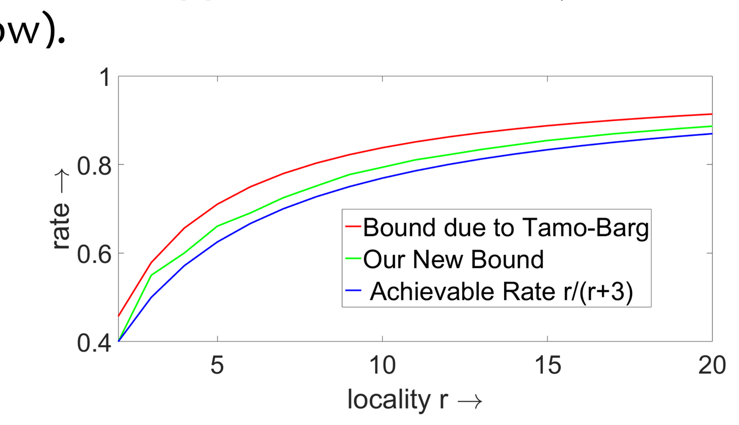
Example parity check matrix of a strict availability code for r=2,t=3 :



Rate Bound on codes with strict availability and minimum distance bound on codes with availability

Rate bound for t=3:

We formed and analyzed a greedy algorithm to calculate an upper bound on rate (shown in plot below).



Rate bound using transpose trick :

Take the example H matrix shown in strict availability definition for r=2,t=3. If we take H^T, it is also a valid parity check matrix of a strict availability code for r=2,t=3. Using this type of relation between H,H^T we get:

$$\begin{aligned} R(r,t) &= 1 - \frac{t}{r+1} + \frac{t}{r+1} R(t-1, r+1)), \\ R(r,t) &\leq 1 - \frac{t}{r+1} + \frac{t}{r+1} \frac{1}{\prod_{j=1}^{r+1}\left(1 + \frac{1}{j(t-1)}\right)}. \end{aligned}$$

Few of Other Results:

- Bound on block length for t=2,3 for code with sequential recovery and almost block length optimal construction for t=3.
- Characterization of rate-optimal code for t=2.
- Low block length – rate optimal construction for r=2 for codes with sequential recovery.
- Bound on minimum distance of a code with availability.
- Construction of codes with partial maximal recoverability.

Codes with Locality for Multiple Erasures in Distributed Storage

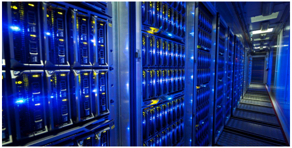
S. B. Balaji
Advisor: P. Vijay Kumar

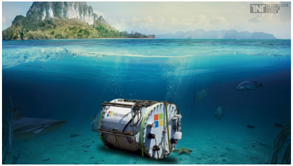
ECE Dept., Indian Institute of Science, Bangalore

EECS Research Students Symposium 2017

Distributed storage

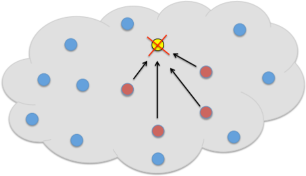
Data center:





- In date center, information is stored in disks.
- It is a challenge to repair a disk when it fails.
- Store redundant information to repair a failed disk.

Disks helping in repair of a failed disk by using redundant information:



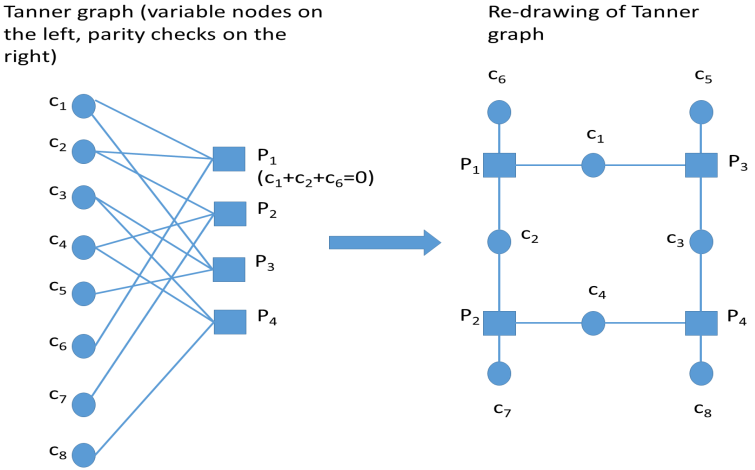
Approaches to Local Multiple-Erasure Recovery

In this presentation we focus on only one of my work:
Focus on two approaches for local recovery from multiple erasures:

- Sequential Recovery from t erasures.
- Recovery from t erasures using availability (similar to parallel recovery).

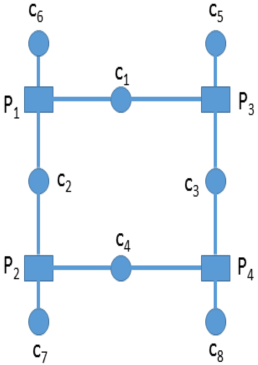
Codes with Locality for Sequential Recovery

An example $[8, 4]$ code for t=2, r=2:



Codes with Locality for Sequential Recovery

Tanner graph:



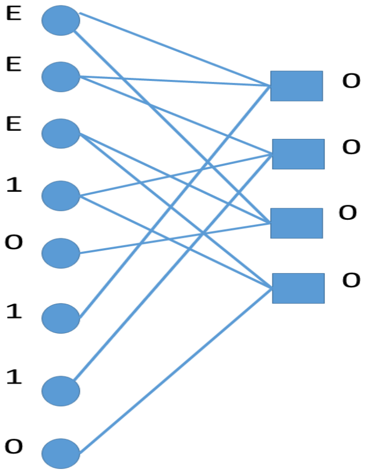
Parity Check matrix (representing equations satisfied by the code symbols):

$$H = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{array}\right]$$

Note that the parity check matrix is of the form $H = [I|C]$, where C has column weight $t = 2$ and row weight $r = 2$.

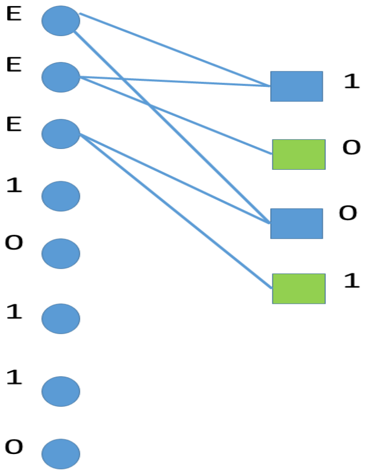
Codes with Locality for Sequential Recovery

Decoding using Peeling Decoder, sequentially from an example of 3 erasures:



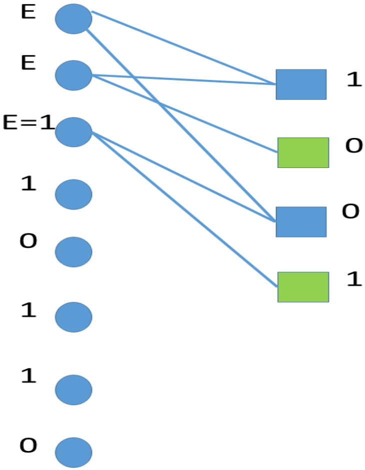
Codes with Locality for Sequential Recovery

Decoding using Peeling Decoder, sequentially from an example of 3 erasures:



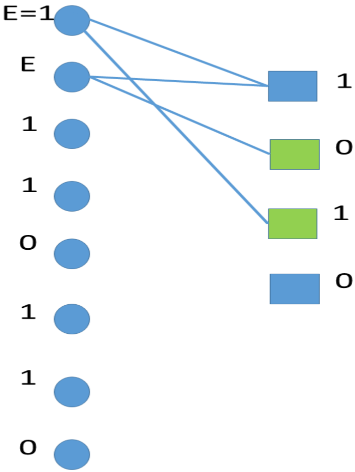
Codes with Locality for Sequential Recovery

Decoding using Peeling Decoder, sequentially from an example of 3 erasures:



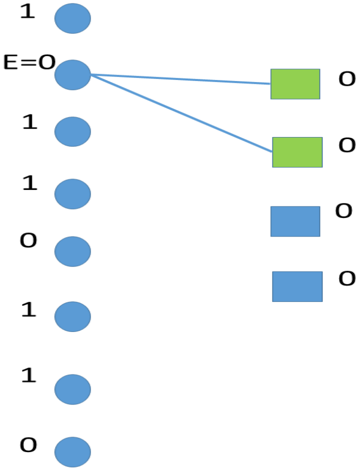
Codes with Locality for Sequential Recovery

Decoding using Peeling Decoder, sequentially from an example of 3 erasures:



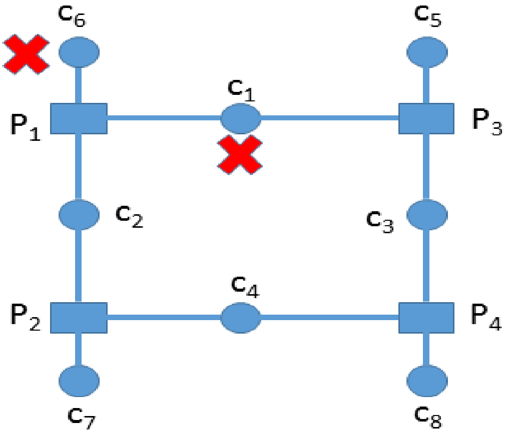
Codes with Locality for Sequential Recovery

Decoding using Peeling Decoder, sequentially from an example of 3 erasures:



Codes with Locality for Sequential Recovery

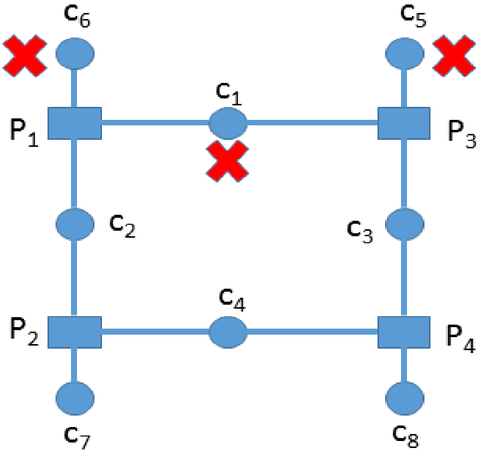
Illustration of Sequential Recovery from 2 erasures :



c_1 is first recovered using pairty check P_3 and then c_6 is recovered using pairty check P_1.

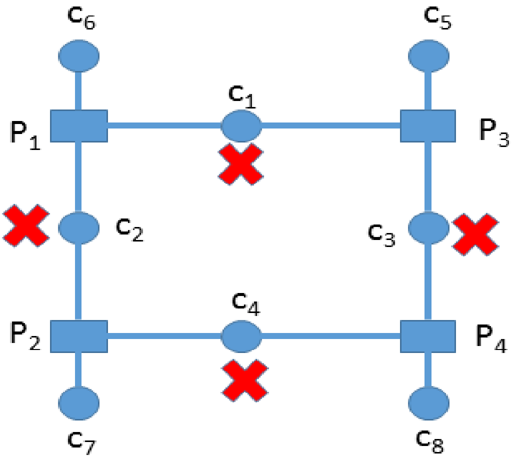
Codes with Locality for Sequential Recovery

Type 1 error pattern: A minimal uncorrectable erasure pattern illustrating the need for large distance between two degree one variable nodes.



Codes with Locality for Sequential Recovery

Type 2 error pattern: A minimal uncorrectable erasure pattern illustrating the need for large cycle length in the graph.



Sequential Recovery Results

- Achievable upper bound on rate R for any $(r \geq 3,\ t \geq 2)$
- Matching binary-code construction.
- The above results settles a conjecture by Song, Cai and Yuen.

The (Tight) Upper Bound on Rate

Theorem

Rate Bound: *Let $\mathcal{C}$ be an $(n, k, r, t)_{seq}$ code over a field $\mathbb{F}_q$. Let $r \geq 3$. Then*

$$\frac{k}{n} \leq \frac{r^{\frac{t}{2}}}{r^{\frac{t}{2}} + 2\sum_{i=0}^{\frac{t}{2}-1} r^i} \tag{1}$$

This theorem settles a conjecture by Song, Cai and Yuen who predicted the achievable rate bound to be of the form below:

$$\frac{k}{n} \leq \frac{1}{1 + \sum_{i=1}^{m} \frac{a_i}{r^i}},$$

$$a_i \geq 0, a_i \in \mathbb{Z}, \sum_{i=1}^{m} a_i = t,$$

$$m = \lceil log_r(k) \rceil.$$

Proof
(after some work, for t even, arrive at form below for parity check matrix of the code)

$$
H = \left[\begin{array}{c|c|c|c|c|c|c|c}
D_0 & A_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \hline
0 & D_1 & A_2 & 0 & \dots & 0 & 0 & 0 \\ \hline
0 & 0 & D_2 & A_3 & \dots & 0 & 0 & 0 \\ \hline
0 & 0 & 0 & D_3 & \dots & 0 & 0 & 0 \\ \hline
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ \hline
0 & 0 & 0 & 0 & \dots & A_{\frac{t}{2}-2} & 0 & 0 \\ \hline
0 & 0 & 0 & 0 & \dots & D_{\frac{t}{2}-2} & A_{\frac{t}{2}-1} & 0 \\ \hline
0 & 0 & 0 & 0 & \dots & 0 & D_{\frac{t}{2}-1} & \\ \cline{1-7}
0 & 0 & 0 & 0 & \dots & 0 & 0 & C
\end{array}\right]
$$

- $\{D_i\}$ are diagonal and each row of $\{A_i\}$ has Hamming weight $\leq r$.
- $\{C\}$ has column weight 2.
- Form equations by equating row weights and column weights of various part of the matrix. These equations lead to the bound.

An example rate optimal construction for t=4

- Parity check matrix of a rate-optimal code for $t = 2$:

$$H = \left[\begin{array}{c|c} D_0 & C \end{array}\right]$$

We have already seen an example construction for $t = 2$ with parity check matrix of the above form.

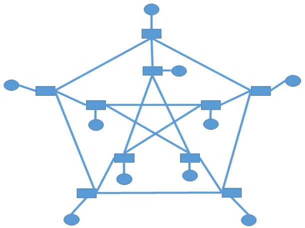
- Parity check matrix of a rate-optimal code for $t = 4$:

$$H = \left[\begin{array}{c|c|c} D_0 & A_1 & 0 \\ \hline 0 & D_1 & C \end{array}\right]$$

We will now construct an example rate-optimal code for $t = 4$ with parity check matrix in the above form.

An example rate optimal construction for t=4: Step 1

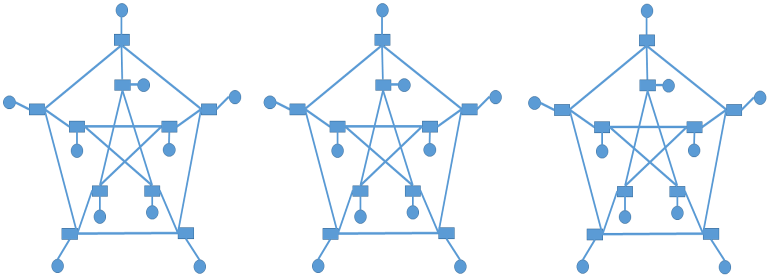
- We have constructed rate-opitmal codes for any r,t. Below we give a simple rate-opimal construction for $t = 4$.
- Below we depict a graph called Petersen graph (it has least number of nodes with girth 5 and degree 3 (Moore graph)).



- This is a representation of Tanner graph for $t = 2, r = 3$ (we did not draw the variable node in the middle of edge to avoid clutter).
- It has parity check matrix of the form $H = [I_{10}|C_1]$, where C_1 has column weight $t = 2$ and row weight $r = 3$.

An example rate optimal construction for t=4: Step 2

Take direct product of the code three times.

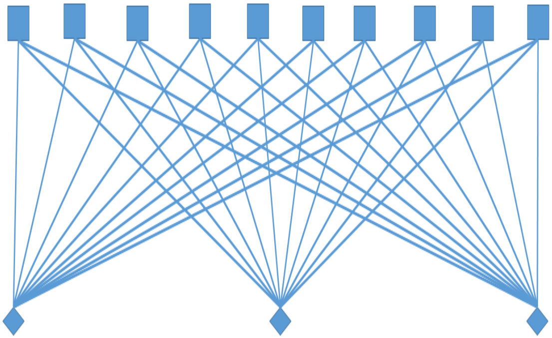


Now the resulting code has parity check matrix of the form:

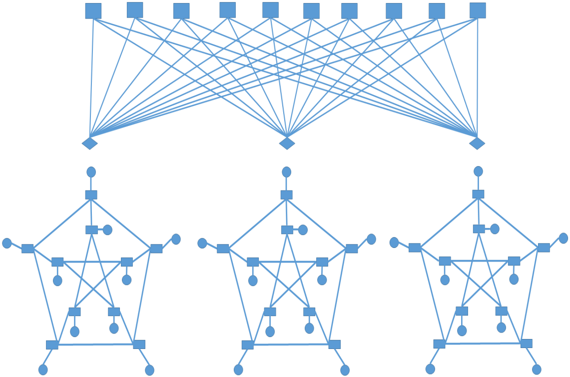
$$H = \left[\begin{array}{ccc|ccc} I_{10} & 0 & 0 & C_1 & 0 & 0 \\ 0 & I_{10} & 0 & 0 & C_1 & 0 \\ 0 & 0 & I_{10} & 0 & 0 & C_1 \end{array} \right]$$

An example rate optimal construction for t=4: Step 3

Take a biregular bipartite graph with top node degree 3 and bottom node degree 10 with girth atleast 4. Here we simply take a complete (3,10)-biregular bipartite graph.

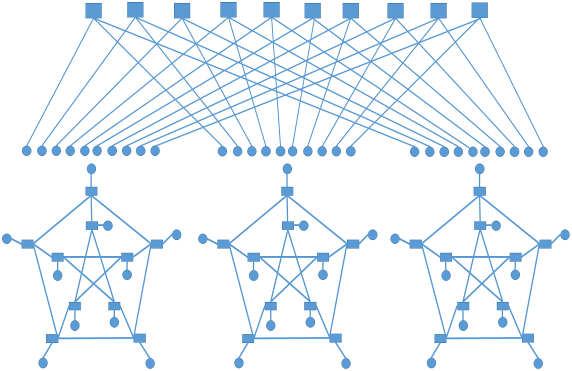


An example rate optimal construction for t=4: Step 4



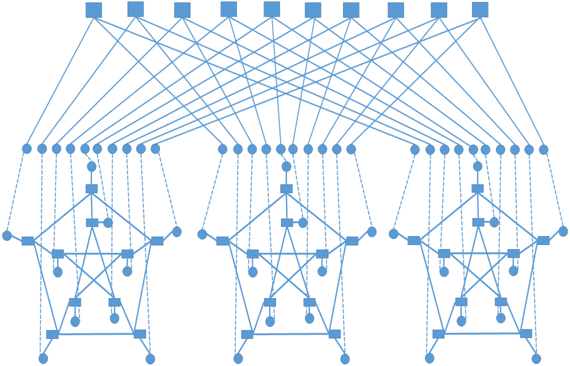
An example rate optimal construction for t=4: Step 5

Split the bottom nodes in the bipartite graph into degree 1 nodes and place it as shown below. Hence each bottom node is split into 10 nodes and the number of nodes in a single petersen graph is also 10.



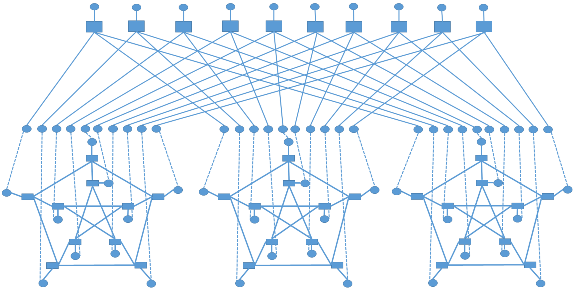
An example rate optimal construction for t=4: Step 6

Merge the nodes as shown below (merging is indicated by dotted lines).



An example rate optimal construction for t=4: Step 7

Attach degree 1 variable nodes to the top nodes of the graph.

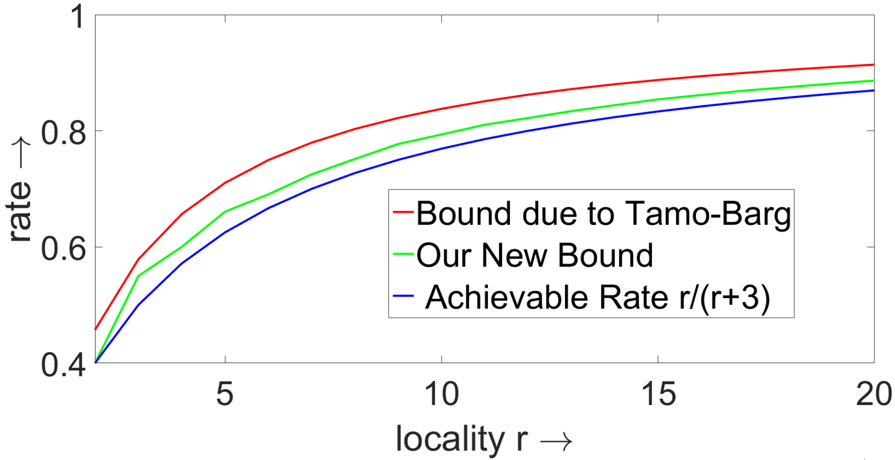


The resulting graph has girth atleast 5 and any 2 degree 1 variable nodes are separated sufficiently apart due to this tree (layer)-like construction and parity check matrix is of the form given below:

$$H = \left[\begin{array}{c|ccc|ccc} I_{10} & A_1 & A_2 & A_3 & 0 & 0 & 0 \\ \hline 0 & I_{10} & 0 & 0 & C_1 & 0 & 0 \\ 0 & 0 & I_{10} & 0 & 0 & C_1 & 0 \\ 0 & 0 & 0 & I_{10} & 0 & 0 & C_1 \end{array}\right]$$

Rate bound for $t = 3$

We derived a rate bound for $t = 3$ by providing and analyzing a greedy algorithm. Below shows the plot of our bound in comparison with existing bound due to Tamo, Barg and known achievable rate.



Other Results

- Bound on block length for t=2,3 for code with sequential recovery and almost block length optimal construction for t=3.
- Low block length rate optimal construction for r=2 for codes with sequential recovery.
- Bound on rate of code with availability.
- Bound on minimum distance of a code with availability.
- Construction of codes with maximum recoverability and partial maximum recoverability.

Thanks!