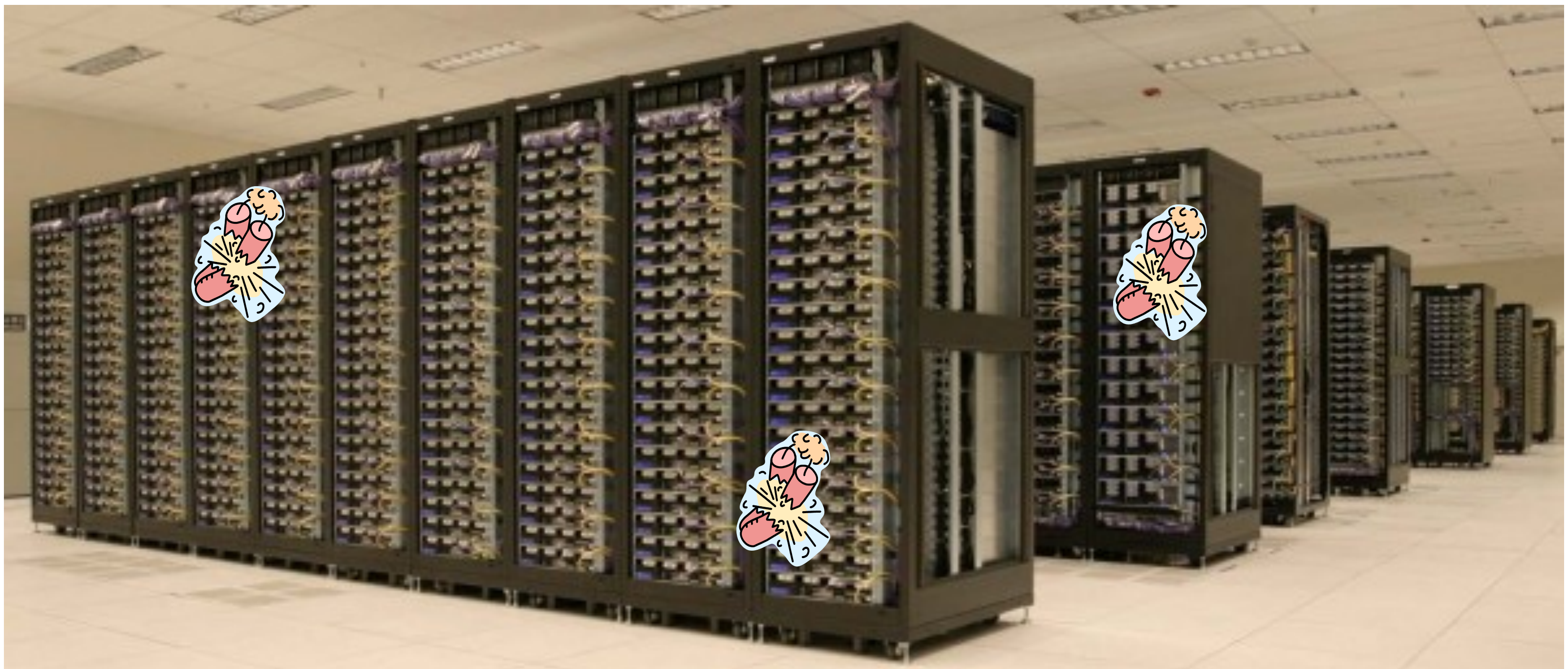
Repairable Fountain Codes

Megasthenis Asteris
Alex Dimakis

Distributed Storage



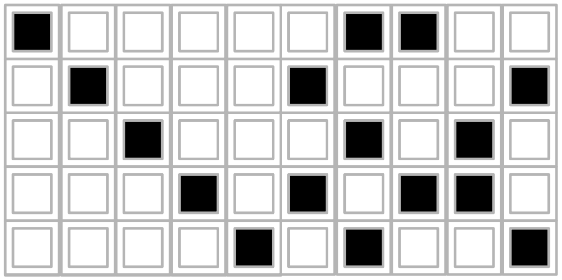
Cluster of machines running Hadoop at Yahoo! (Source: Yahoo!)

- Failures are the **norm.**
- We need to protect the data: **Introduce redundancy**
- Already using **Erasure Codes** (e.g. Reed Solomon)

Problem Description

Create a linear code with the following properties:

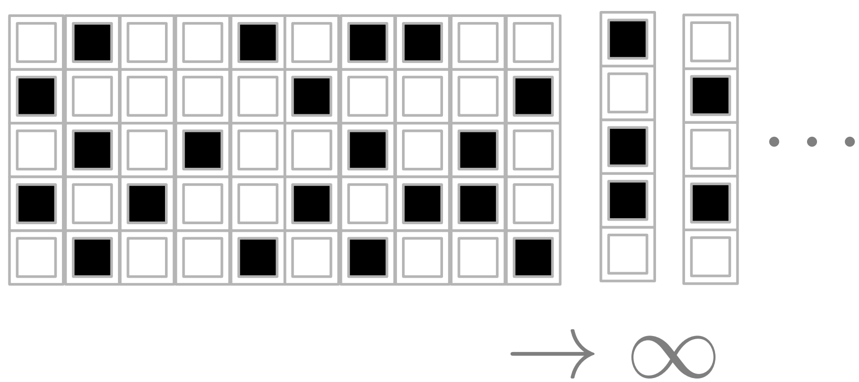
- **Systematic** form
  Input is part of the output



Problem Description

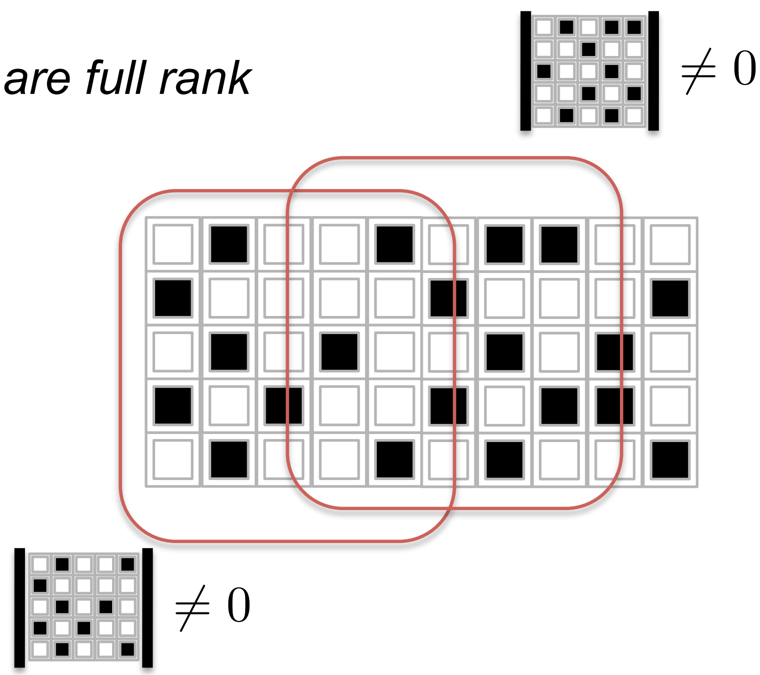
- **Rateless** property

  Columns created independently

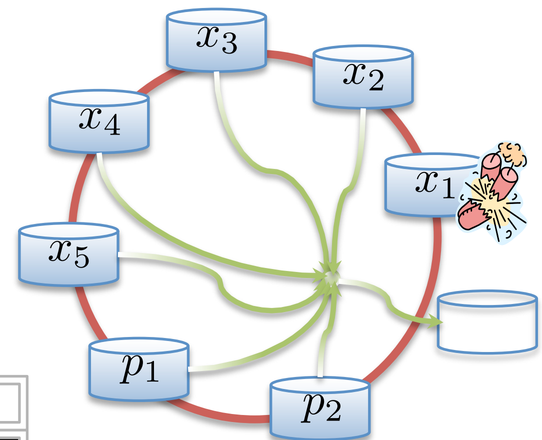


Problem Description

- **MDS** property
  Any k columns are full rank

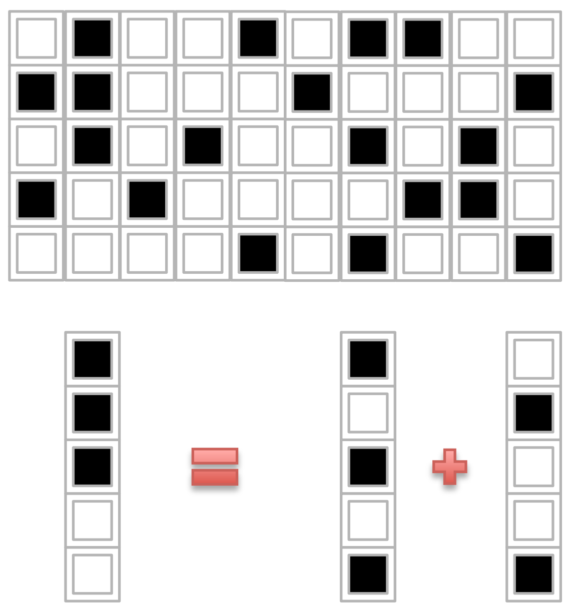


Problem Description



- Good **Locality**

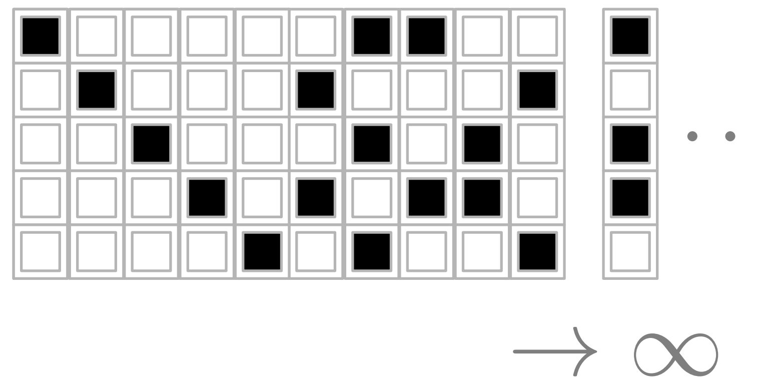
  A column is a linear combination of at most l other columns.



Problem Description

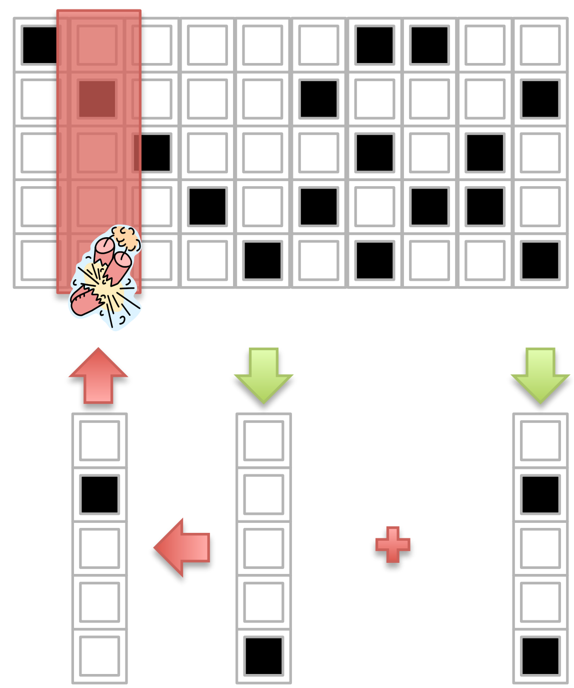
Summarizing…

- **Systematic** form
- **Rateless** property
- **MDS** property
- Good **Locality**



Approach

- **Systematic**
  +
- **Sparse Parities**
  =
- Good **Locality**



- **Sparsity** is a measure of **locality**:
  The sparser the parities the better the locality.

Q How sparse can parities be?

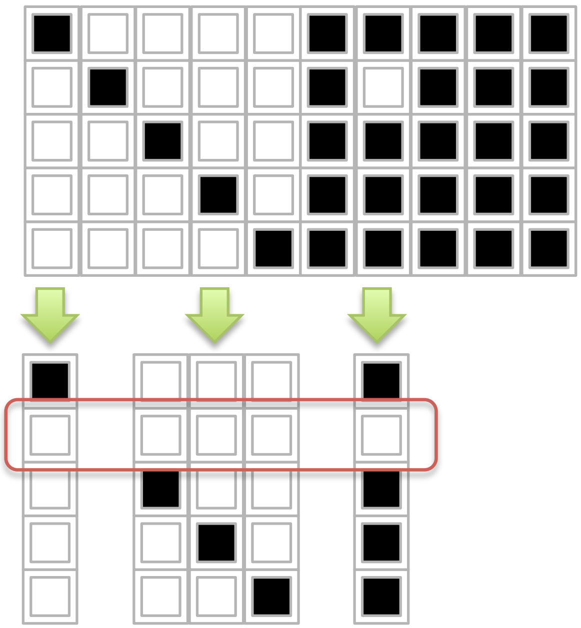
Approach

However...

- **Systematic MDS** codes can afford no zero in the parity columns.

- MDS cannot have **Sparse Parities**.

- Relax **MDS** property:
  - **(1+ε) k** symbols should suffice for decoding.

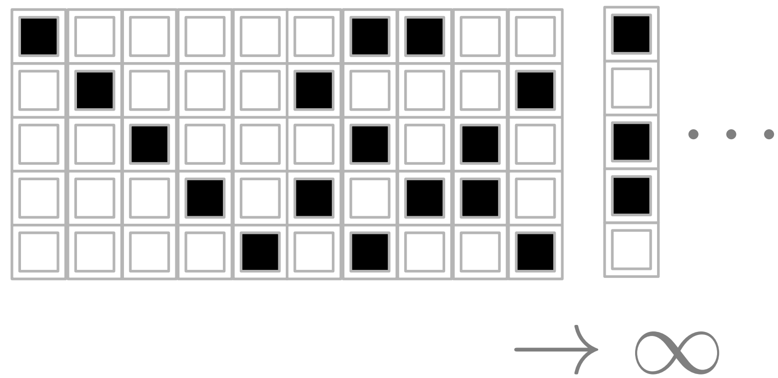


Approach

In light of the observations, we want:

- **Systematic**
- **Rateless**
- **Sparse Parities**
- ~~MDS property~~

(1+ε)k suffice for decoding



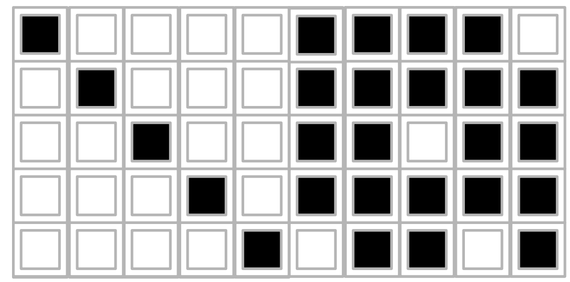
Fountain Codes

ε should be arbitrarily small.

Prior work

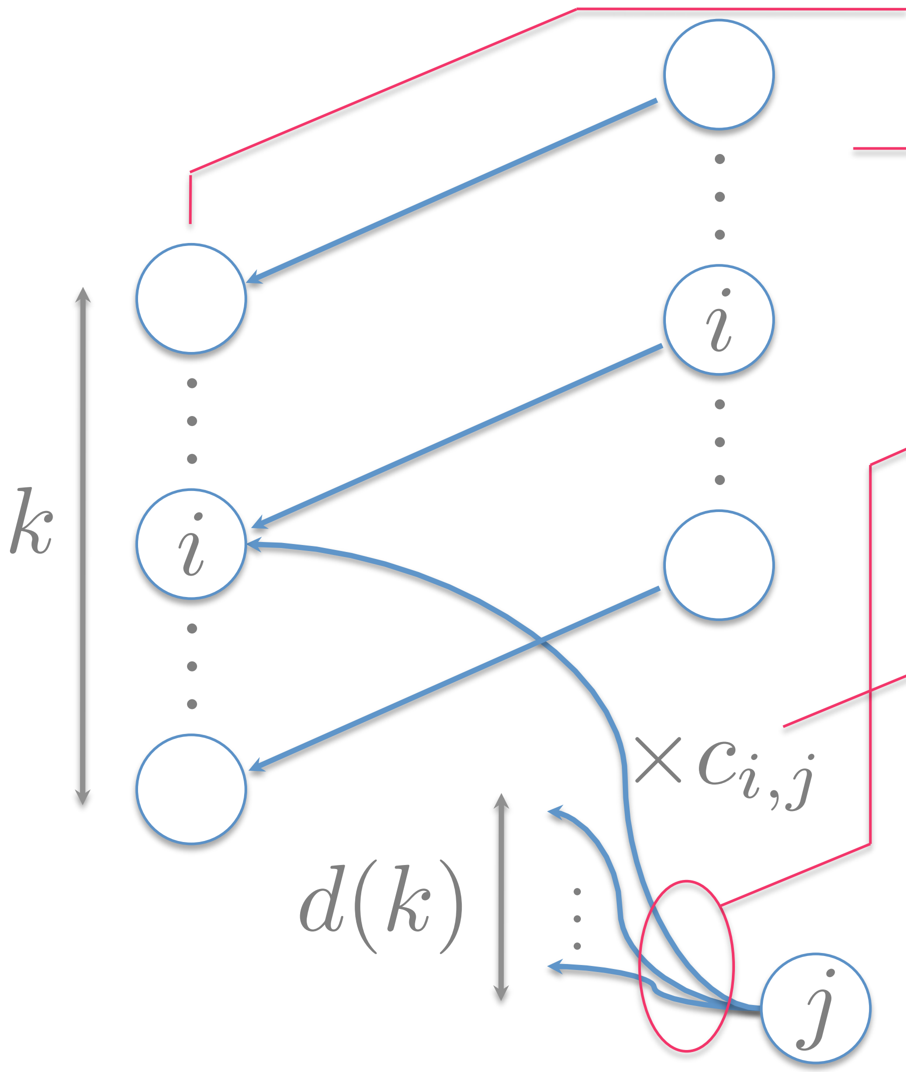
- [Shokrollahi] -Raptor Codes

  - Systematic version
  - Parities no longer sparse in the input



- [Gummadi]– Systematic LT/Raptor codes – ε cannot be arbitrarily small.

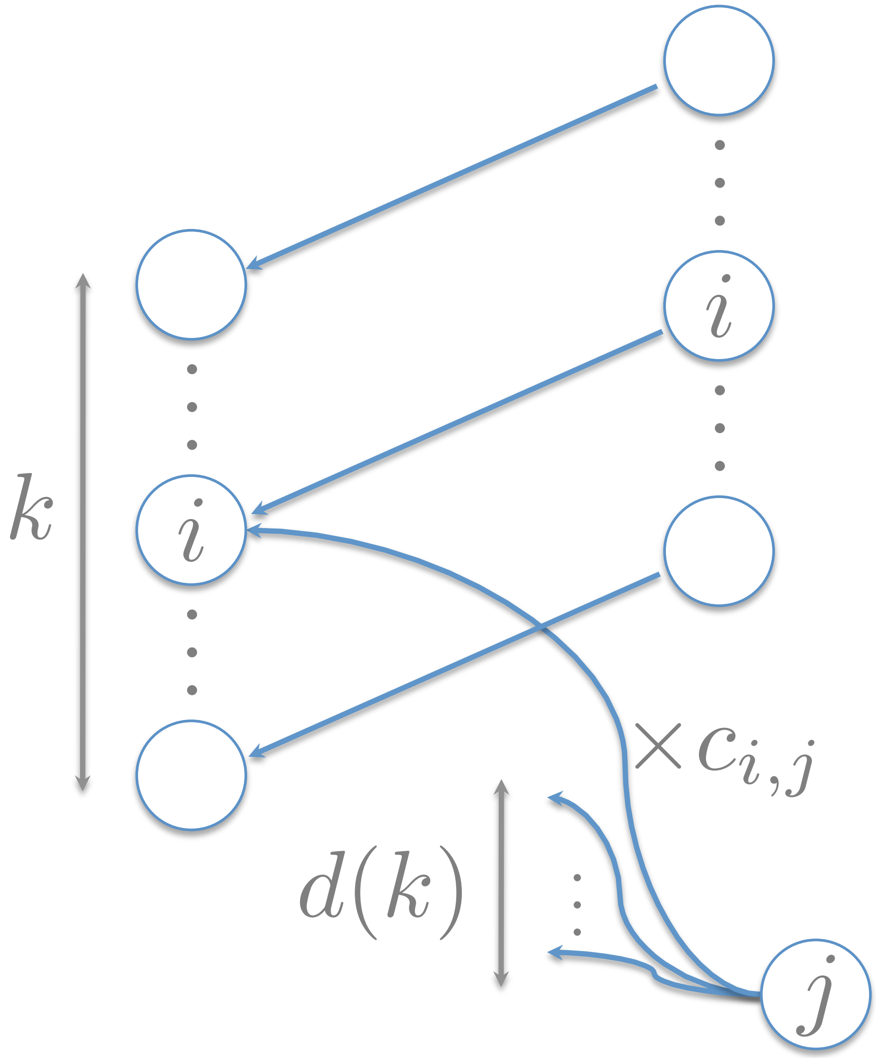
Our construction



- k input symbols
- Systematic part
- For each parity:
  - Choose $d(k)$ symbols.
  - Choose $c_{ij} \sim U[0 \ldots q)$

How small can the degree d(k), of the encoded symbols be?

Results



Theorem 1

Decoding *w.h.p.*

$$\Rightarrow$$

$$d(k) = \Omega(\ln(k))$$

Coupon Collector

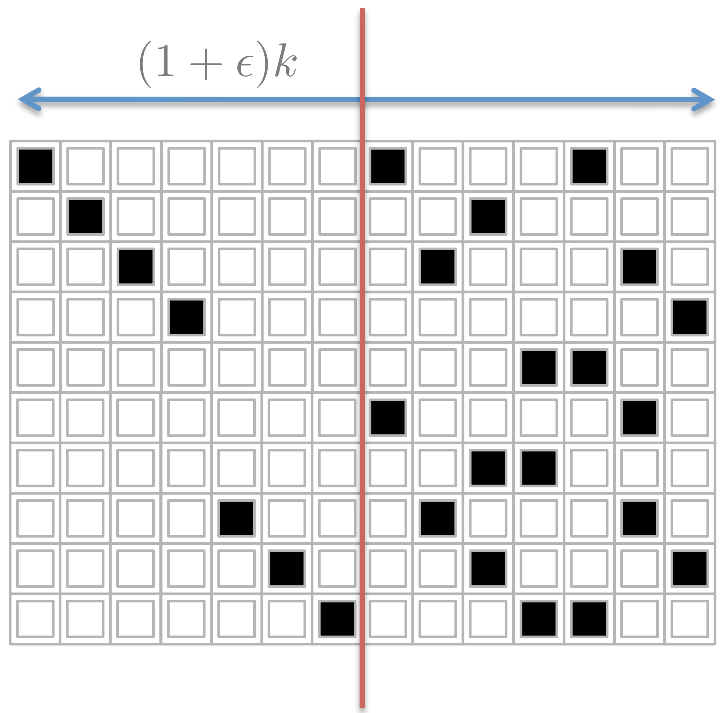
Theorem 2

$$d(k) = c \ln(k) \quad (c \propto 1/\epsilon)$$

$$\Rightarrow$$

$(1+\epsilon)k$ random columns of $\mathbf{G}$ are lin. indep. w.p. k/q close to 1.

Proving Theorem 2



• Analyze the rank of a random matrix.

• But here, we have an arbitrary number of systematic columns and a random part (parities)

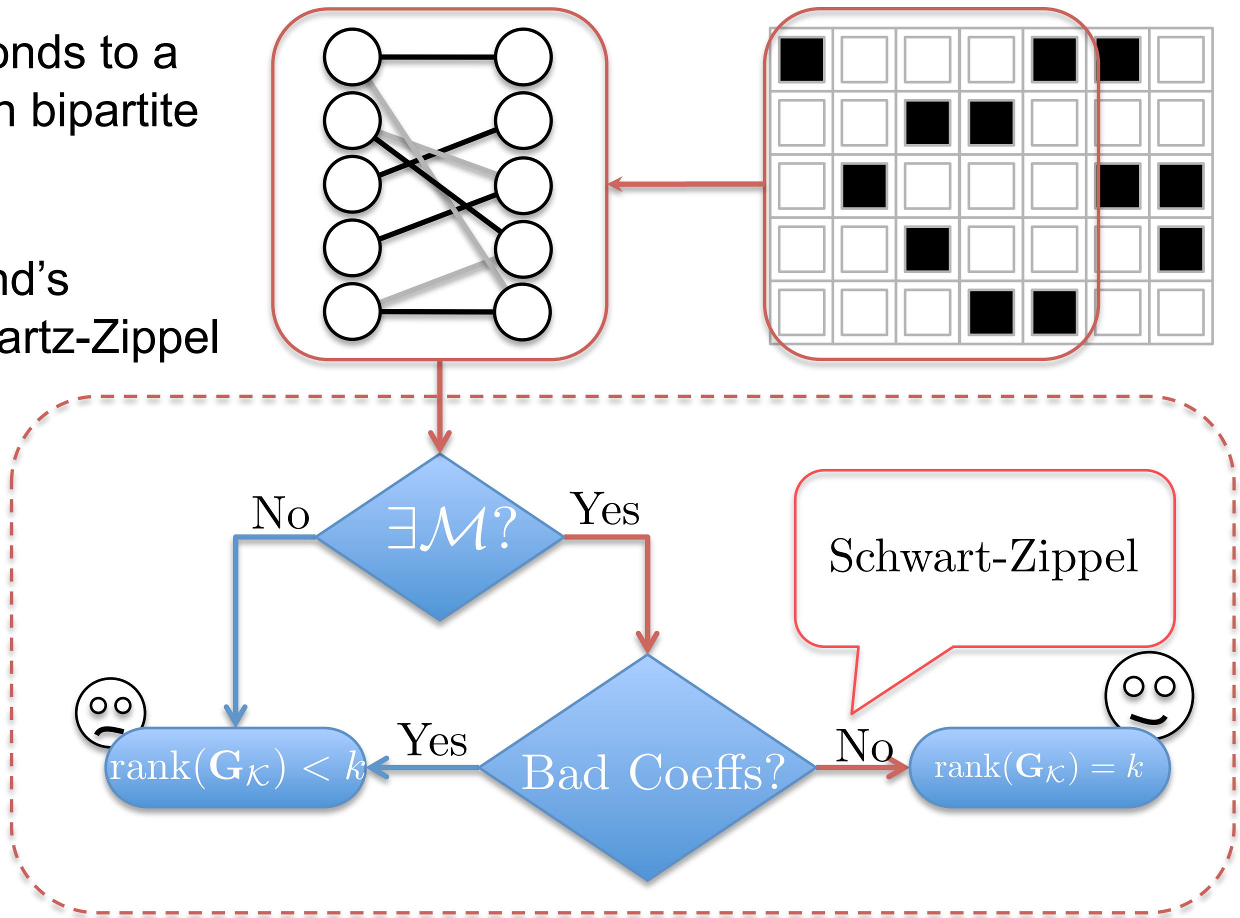
[Kovalenko & Levitskaya, Cooper, Karp, ….]

Proof – Part 1

- Focus on k x k submatrix.

- Full rank corresponds to a perfect matching on bipartite graph

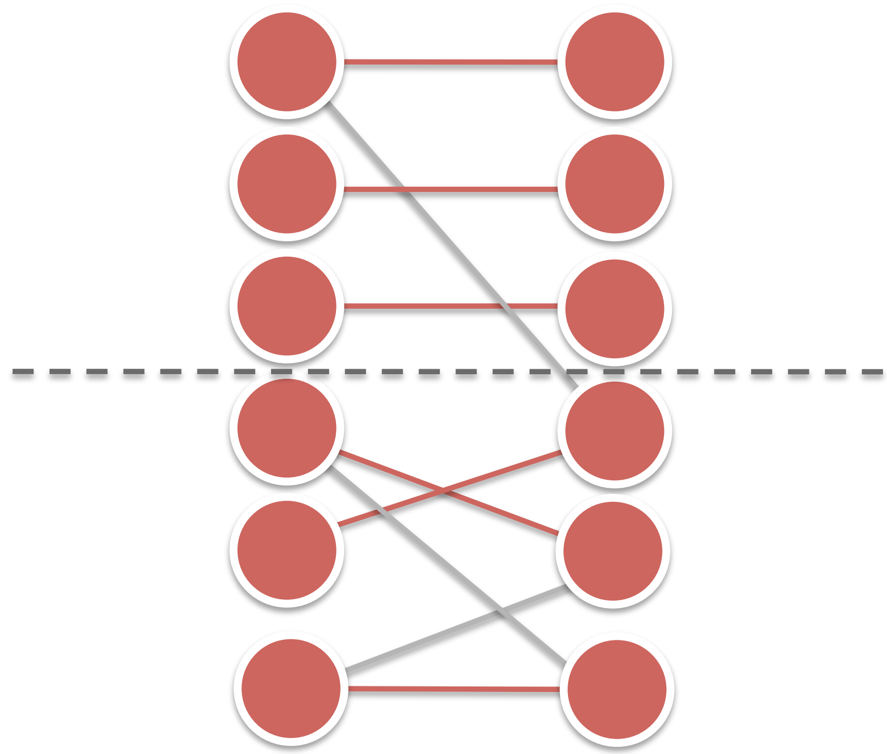
•(whp using Edmond's theorem and Schwartz-Zippel



[Ho et al.]

Proof – Part 2

Key technical result:
- There is a perfect matching whp.

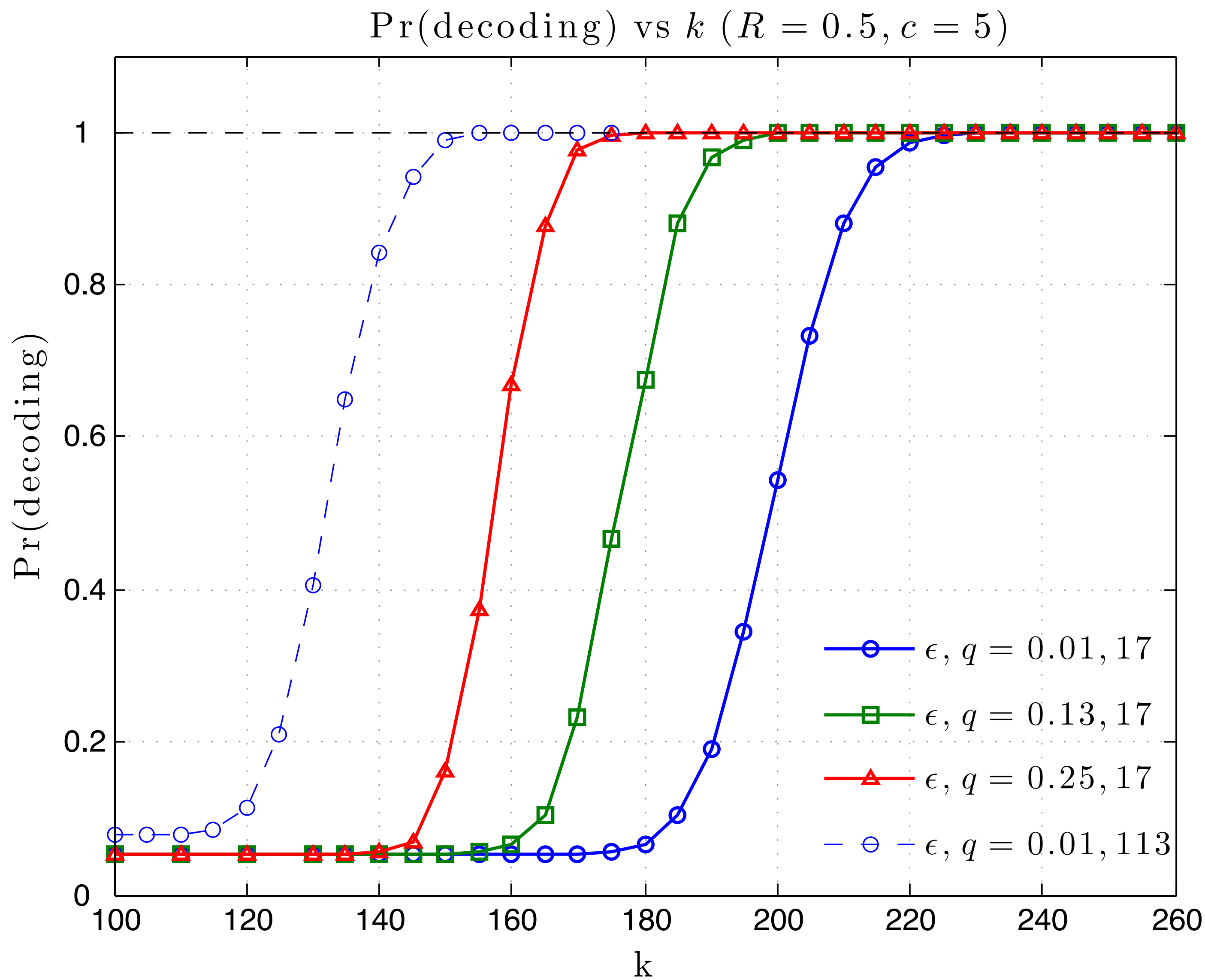


- Erdos-Reyni Random matrix result, Hall's marriage theorem, first moment method.

Conclusions

- We introduced a new family of fountain codes:
  - Systematic
  - Near MDS
  - With logarithmic locality (easy repair of a single symbol failure - very useful in distributed storage systems)
- Our proof involved analyzing a new family of random matrices.
- An interesting open problem: can we use belief propagation decoding for this ensemble?

Simulation



Pr(decoding) vs k ($R = 0.5, c = 5$)
Pr(decoding)
0
0.2
0.4
0.6
0.8
1
100
120
140
160
180
200
220
240
260
k
$\epsilon,\ q = 0.01, 17$
$\epsilon,\ q = 0.13, 17$
$\epsilon,\ q = 0.25, 17$
$\epsilon,\ q = 0.01, 113$