UNIVERSITY OF TARTU

Update and Repair Efficient Storage Codes With Availability via Finite Projective Planes

Junming Ke, Ago-Erik Riet

IEEE ISIT 2022

Projective Planes

Finite projective plane:

Let X be a finite set, and let $\mathcal{L}$ be a system of subsets of X. The pair $(X, \mathcal{L})$ is called a finite projective plane if it satisfies the following axioms.

1. There exists a 4-element set $F \subseteq X$ such that $|L \cap F| \leq 2$ holds for each set $L \in \mathcal{L}$.

2. Any two distinct sets $L_1, L_2 \in \mathcal{L}$ intersect in exactly one element, i.e. $|L_1 \cap L_2| = 1$.

3. For any two distinct elements $x_1, x_2 \in X$, there exists exactly one set $L \in \mathcal{L}$ such that $x_1 \in L$ and $x_2 \in L$.

Two parallel lines will be intersected.

Incidence Matrix

Incidence Matrix:

$$A_q = (a_{ij})$$

$$a_{ij} = \begin{cases} 1, \text{if the point } i \text{ is incident with the hyperplane } j \\ 0, \text{otherwise} \end{cases}$$

***p*-Rank:**

The rank of the incidence matrix of points and hyperplanes in the $\mathrm{PG}(t, p^n)$ is $\binom{p+t-1}{t}^n + 1$.

In $\mathrm{PG}(2, q)$, q odd: $\binom{q+1}{2} + 1 = \frac{q(q+1)}{2} + 1$.

Example: An incidence matrix A_3 of $\mathrm{PG}(2,3)$ is

$$A_3 = \begin{pmatrix}
1&1&0&0&1&0&1&0&0&0&0&0&0\\
0&1&1&0&0&1&0&1&0&0&0&0&0\\
0&0&1&1&0&0&1&0&1&0&0&0&0\\
0&0&0&1&1&0&0&1&0&1&0&0&0\\
0&0&0&0&1&1&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&0&1&0&1&0\\
0&0&0&0&0&0&1&1&0&0&1&0&1\\
1&0&0&0&0&0&0&1&1&0&0&1&0\\
0&1&0&0&0&0&0&0&1&1&0&0&1\\
1&0&1&0&0&0&0&0&0&1&1&0&0\\
0&1&0&1&0&0&0&0&0&0&1&1&0\\
0&0&1&0&1&0&0&0&0&0&0&1&1\\
1&0&0&1&0&1&0&0&0&0&0&0&1
\end{pmatrix}.$$

The rank of A_3 is $\frac{3(3+1)}{2} + 1 = 7$.

Incidence Matrix

Incidence Matrix:

$$A_q = (a_{ij})$$

$$a_{ij} = \begin{cases} 1, \text{if the point } i \text{ is incident with the hyperplane } j \\ 0, \text{otherwise} \end{cases}$$

***p*-Rank:**

The rank of the incidence matrix of points and hyperplanes in the $\mathrm{PG}(t, p^n)$ is $\binom{p+t-1}{t}^n + 1$.

In $\mathrm{PG}(2, q)$, q odd: $\binom{q+1}{2} + 1 = \frac{q(q+1)}{2} + 1$.

Example: An incidence matrix A_3 of $\mathrm{PG}(2,3)$ is

$$A_3 = \begin{pmatrix}
1&1&0&0&1&0&1&0&0&0&0&0&0\\
0&1&1&0&0&1&0&1&0&0&0&0&0\\
0&0&1&1&0&0&1&0&1&0&0&0&0\\
0&0&0&1&1&0&0&1&0&1&0&0&0\\
0&0&0&0&1&1&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&0&1&0&1&0\\
0&0&0&0&0&0&1&1&0&0&1&0&1\\
1&0&0&0&0&0&0&1&1&0&0&1&0\\
0&1&0&0&0&0&0&0&1&1&0&0&1\\
1&0&1&0&0&0&0&0&0&1&1&0&0\\
0&1&0&1&0&0&0&0&0&0&1&1&0\\
0&0&1&0&1&0&0&0&0&0&0&1&1\\
1&0&0&1&0&1&0&0&0&0&0&0&1
\end{pmatrix}.$$

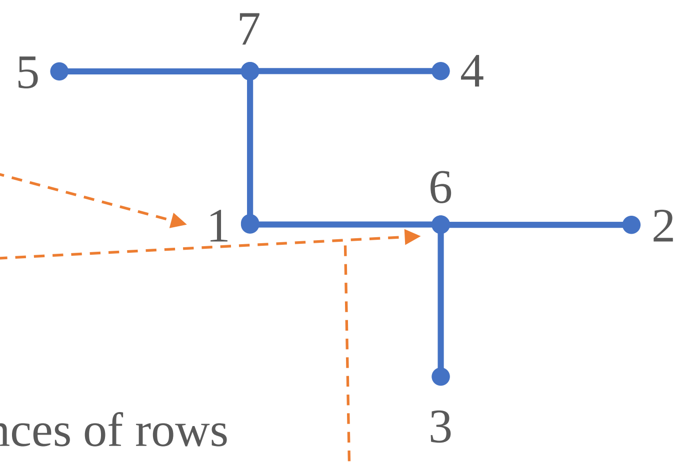
The rank of A_3 is $\frac{3(3+1)}{2} + 1 = 7$.

Moorhouse basis

$$H_3 = \begin{pmatrix}
1&1&0&0&1&0&1&0&0&0&0&0&0\\
0&1&1&0&0&1&0&1&0&0&0&0&0\\
0&0&1&1&0&0&1&0&1&0&0&0&0\\
0&0&0&1&1&0&0&1&0&1&0&0&0\\
0&0&0&0&1&1&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&0&1&0&1&0\\
0&0&0&0&0&0&1&1&0&0&1&0&1
\end{pmatrix}.$$

Generator matrix G

$$H_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \end{pmatrix}.$$

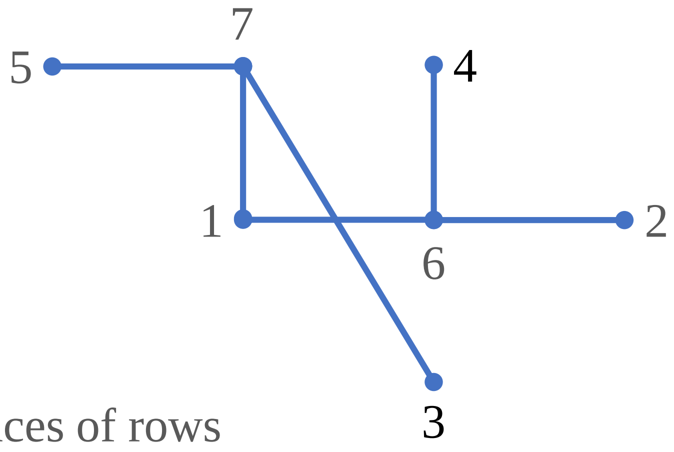


All pairwise differences of rows of H can be expressed via paths between vertices.

$$G_3^{(1)} = \begin{pmatrix} 1&1&0&0&1&-1&0&0&0&-1&0&-1&0 \\ 1&1&0&0&1&0&0&-1&0&0&-1&0&-1 \\ 0&1&1&0&0&0&-1&1&0&-1&0&-1&0 \\ 0&0&1&1&0&-1&0&0&1&-1&0&-1&0 \\ 0&0&0&1&1&0&-1&0&0&1&-1&0&-1 \\ 0&0&0&0&1&1&-1&-1&1&0&0&0&-1 \end{pmatrix}$$

Generator matrix G

$$H_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \end{pmatrix}.$$



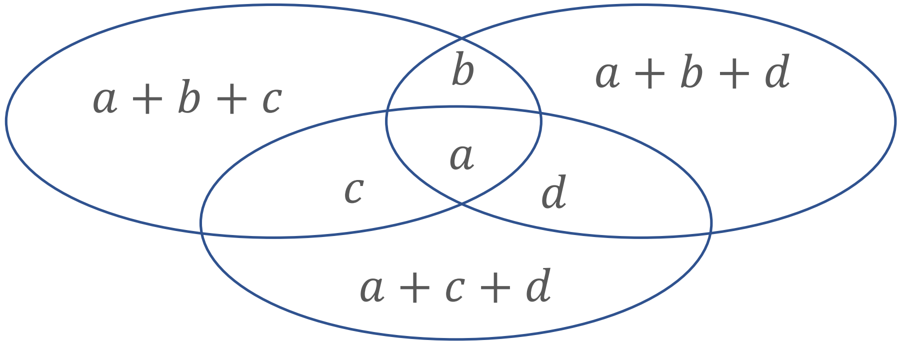
All pairwise differences of rows of H can be expressed via paths between vertices.

$$G_3^{(2)} = \begin{pmatrix} 1&1&0&0&1&-1&0&0&0&-1&0&-1&0 \\ 1&1&0&0&1&0&0&-1&0&0&-1&0&-1 \\ 0&1&1&0&0&0&-1&1&0&-1&0&-1&0 \\ 0&0&1&1&0&0&0&-1&1&0&-1&0&-1 \\ 0&0&0&1&1&-1&-1&1&0&0&0&-1&0 \\ 0&0&0&0&1&1&-1&-1&1&0&0&0&-1 \end{pmatrix}$$

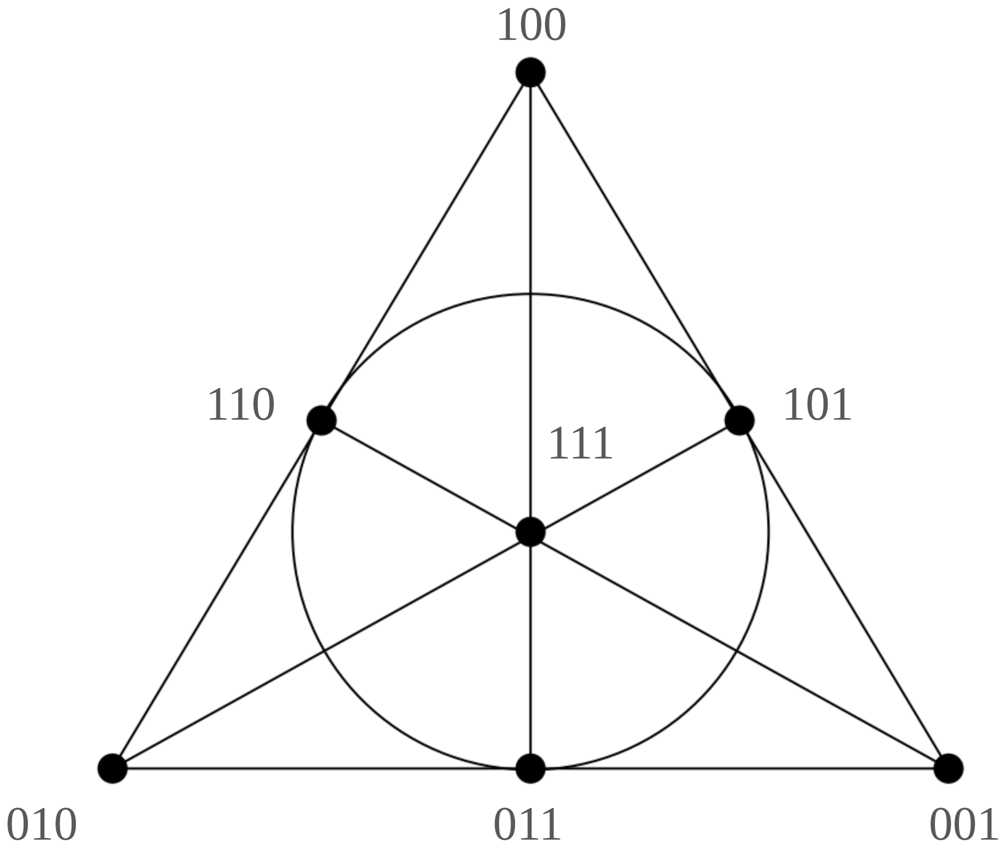
Hamming codes

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The parity check matrix of $[7, 4, 3]$ Hamming codes



Venn diagram of $[7, 4, 3]$ Hamming codes



Labeling the Fano plane

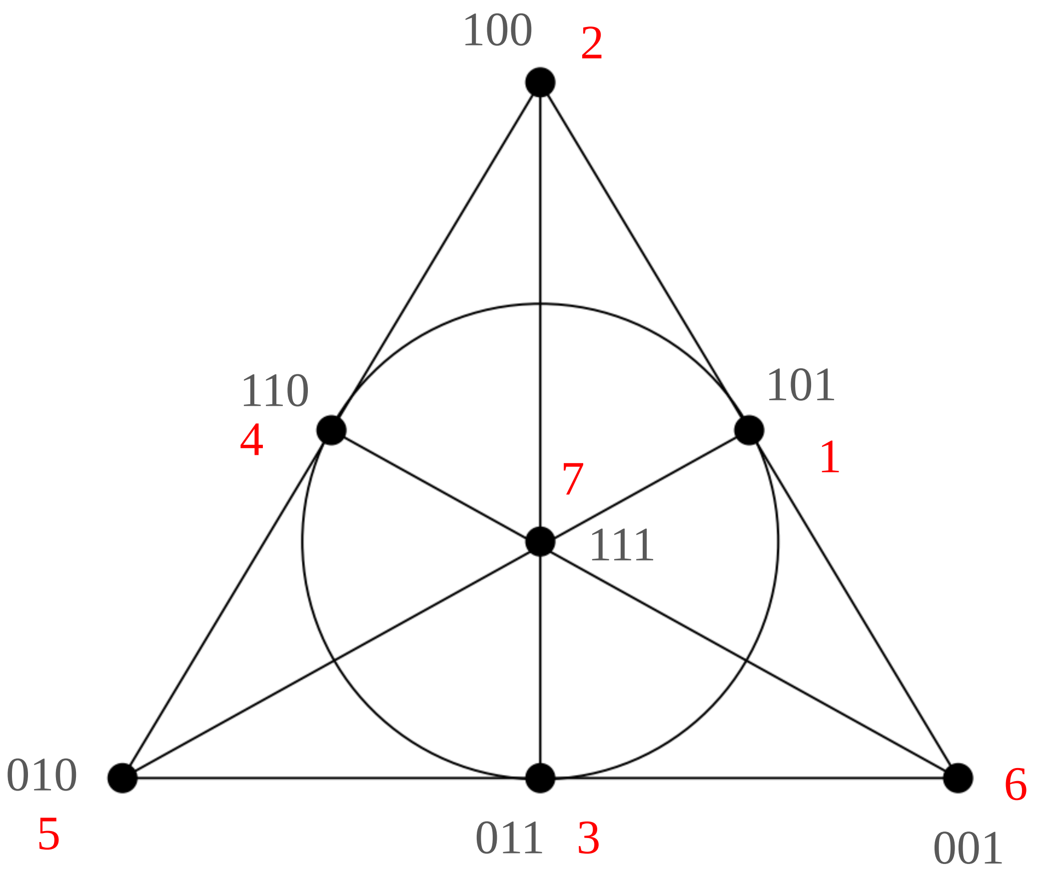
Hamming codes

$$
\begin{array}{c c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
A = & \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
\begin{array}{c} 1,3,4 \\ 2,4,5 \\ 3,5,6 \\ 4,6,7 \\ 1,5,7 \\ 1,2,6 \\ 2,3,7 \end{array}
$$

Incidence matrix of $[7, 4, 3]$ Hamming codes

Reorder the columns of H in order to get the cyclic form of A.

$$
H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \boldsymbol{HA^T = AH^T = 0}
$$



Labeling the Fano plane

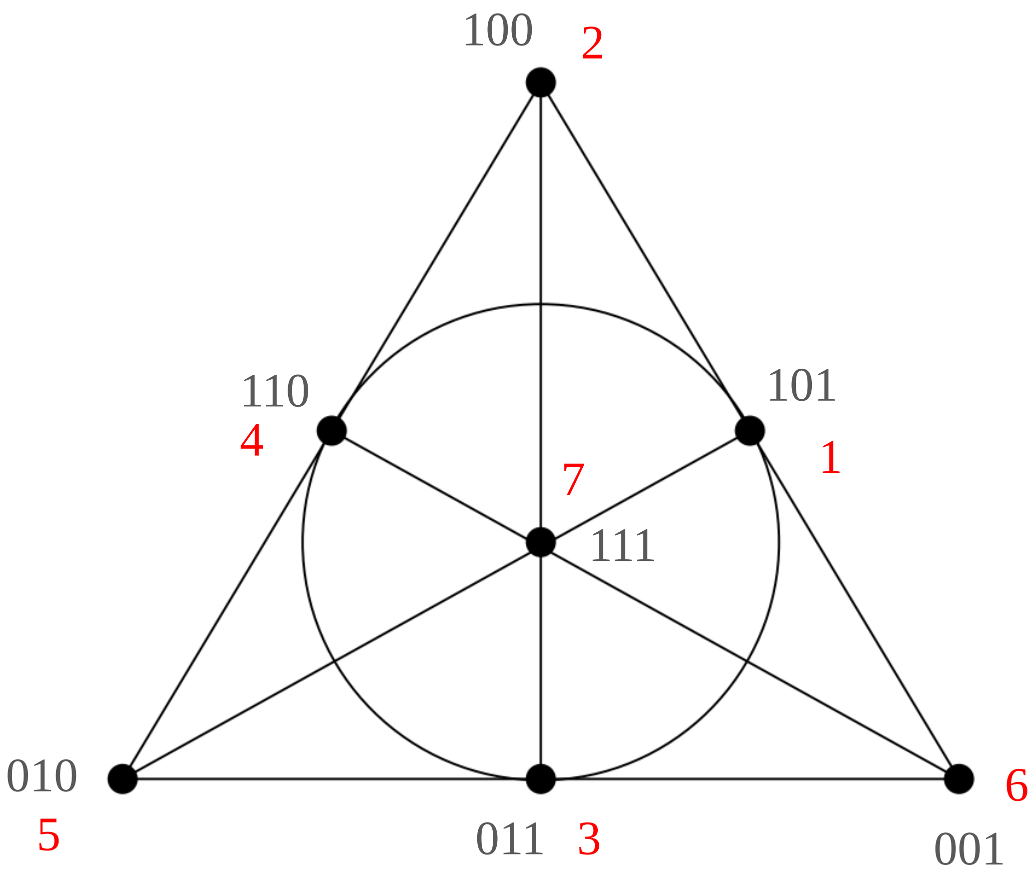
Hamming codes

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

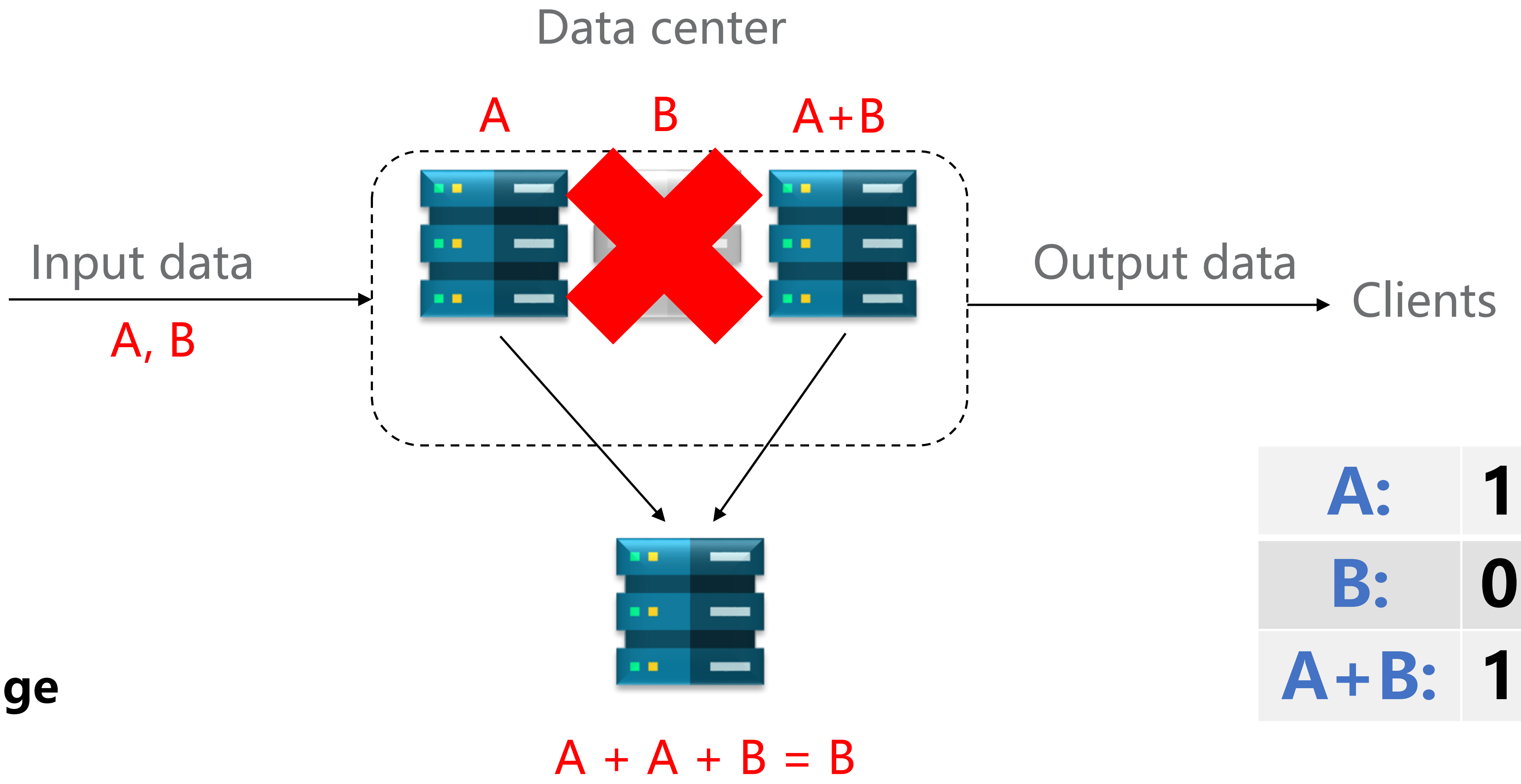
Generator matrix of $[7, 4, 3]$ Hamming codes

Moorhouse basis



Labeling the Fano plane

Repair



A:	10
B:	01
A+B:	11

Coded Storage

Repair locality

The **locality** of a coded symbol b_j is the minimum r_j such that b_j is a function of some other r_j coded symbols

$$b_{i_1}, \dots, b_{i_r} \in \{b_1, \dots, b_n\} \setminus \{b_j\}.$$

Then $\{b_{i_1}, \dots, b_{i_r}\}$ is a **repair group** for b_j.

The **repair locality** r of the code is $r = \max_j r_j$.

The repair locality r of the linear code from finite projective plane $\mathrm{PG}(2, q)$ is q with respect to the codeword length $q^2 + q + 1$.

Note: exactly $q + 1$ points in one line.

Example: An incidence matrix A_3 of $\mathrm{PG}(2,3)$ is

$$A_3 = \begin{pmatrix}
1&1&0&0&1&0&1&0&0&0&0&0&0\\
0&1&1&0&0&1&0&1&0&0&0&0&0\\
0&0&1&1&0&0&1&0&1&0&0&0&0\\
0&0&0&1&1&0&0&1&0&1&0&0&0\\
0&0&0&0&1&1&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&0&1&0&1&0\\
0&0&0&0&0&0&1&1&0&0&1&0&1\\
1&0&0&0&0&0&0&1&1&0&0&1&0\\
0&1&0&0&0&0&0&0&1&1&0&0&1\\
1&0&1&0&0&0&0&0&0&1&1&0&0\\
0&1&0&1&0&0&0&0&0&0&1&1&0\\
0&0&1&0&1&0&0&0&0&0&0&1&1\\
1&0&0&1&0&1&0&0&0&0&0&0&1
\end{pmatrix}.$$

$$H_3 = \begin{pmatrix}
1&1&0&0&1&0&1&0&0&0&0&0&0\\
0&1&1&0&0&1&0&1&0&0&0&0&0\\
0&0&1&1&0&0&1&0&1&0&0&0&0\\
0&0&0&1&1&0&0&1&0&1&0&0&0\\
0&0&0&0&1&1&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&0&1&0&1&0\\
0&0&0&0&0&0&1&1&0&0&1&0&1
\end{pmatrix}.$$

Repair availability

The (repair) **availability** of a coded symbol c_j is its maximum number t_j of pairwise disjoint repair groups;

The **repair availability** of the code given by generator matrix G is $t = \min_j t_j$.

The repair availability of the linear code from finite projective plane $\mathrm{PG}(2, q)$ is $q + 1$ with respect to the codeword length $q^2 + q + 1$.

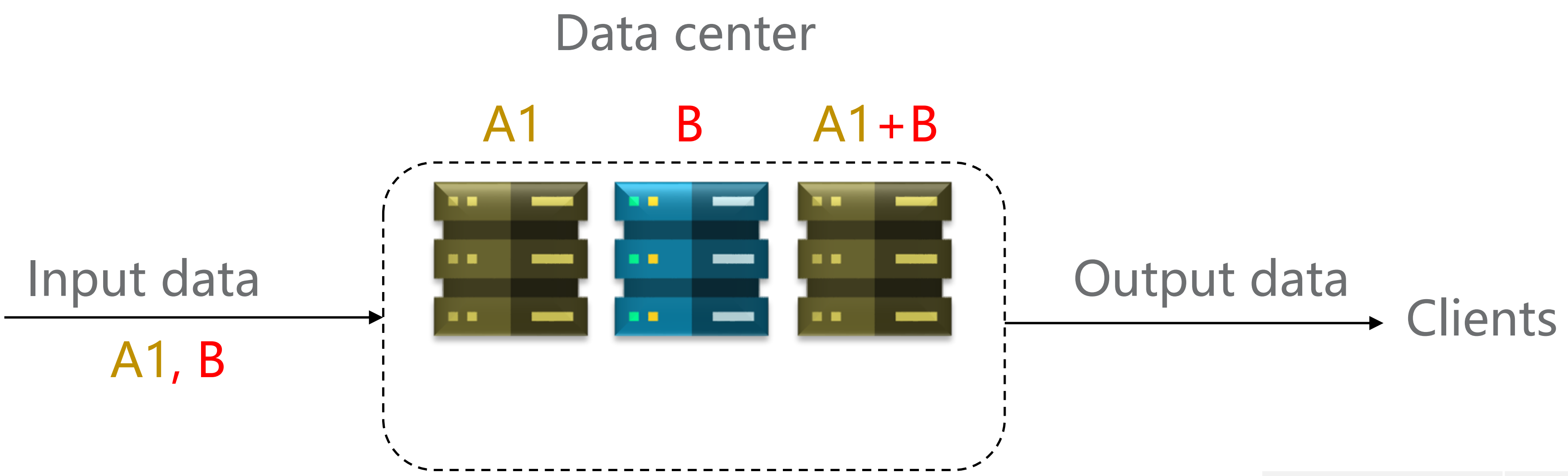
Note: exactly $q + 1$ lines pass through each point.

Example: An incidence matrix A_3 of $\mathrm{PG}(2, 3)$ is

$$A_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \\ 1&0&0&0&0&0&0&1&1&0&0&1&0 \\ 0&1&0&0&0&0&0&0&1&1&0&0&1 \\ 1&0&1&0&0&0&0&0&0&1&1&0&0 \\ 0&1&0&1&0&0&0&0&0&0&1&1&0 \\ 0&0&1&0&1&0&0&0&0&0&0&1&1 \\ 1&0&0&1&0&1&0&0&0&0&0&0&1 \end{pmatrix}.$$

$$G_3^{(1)} = \begin{pmatrix} 1&1&0&0&1&-1&0&0&0&-1&0&-1&0 \\ 1&1&0&0&1&0&0&-1&0&0&-1&0&-1 \\ 0&1&1&0&0&0&-1&1&0&-1&0&-1&0 \\ 0&0&1&1&0&-1&0&0&1&-1&0&-1&0 \\ 0&0&0&1&1&0&-1&0&0&1&-1&0&-1 \\ 0&0&0&0&1&1&-1&-1&1&0&0&0&-1 \end{pmatrix}.$$

Update



A:	10
B:	01
A+B:	11
A1:	11
A1+B:	10

Coded Storage

UNIVERSITY OF TARTU

1632

Update efficiency

The update efficiency u_i of a data symbol a_i is **the number of coded symbols** that need to be updated when updating a_i.

Or, the update efficiency u_i of a data symbol a_i is **the weight of the *i*-th row** of G.

The update efficiency u of the code given by G is

$$\boldsymbol{u} = \max_{\boldsymbol{i}} \boldsymbol{u}_{\boldsymbol{i}}.$$

The update efficiency u of the linear code from finite projective plane $\mathrm{PG}(2, q)$ is $q + 1$ with respect to the codeword length $q^2 + q + 1$.

Points P_j

$$M = \begin{pmatrix} \quad \end{pmatrix}$$

Lines l_i

- $M_{ij} = 1$ iff $P_j \in l_i$,
- $M_{ij} = 0$ iff $P_j \notin l_i$,

The relative Hamming weight of each row:

$$\frac{q+1}{q^2+q+1} \approx \frac{1}{q}$$

Choice of G

$$H_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \end{pmatrix}$$

$$G_3^{(1)} = \begin{pmatrix} 1&1&0&0&1&-1&0&0&0&-1&0&-1&0 \\ 1&1&0&0&1&0&0&-1&0&0&-1&0&-1 \\ 0&1&1&0&0&0&-1&1&0&-1&0&-1&0 \\ 0&0&1&1&0&-1&0&0&1&-1&0&-1&0 \\ 0&0&0&1&1&0&-1&0&0&1&-1&0&-1 \\ 0&0&0&0&1&1&-1&-1&1&0&0&0&-1 \end{pmatrix}$$

$$G_3^{(2)} = \begin{pmatrix} 1&1&0&0&1&-1&0&0&0&-1&0&-1&0 \\ 1&1&0&0&1&0&0&-1&0&0&-1&0&-1 \\ 0&1&1&0&0&0&-1&1&0&-1&0&-1&0 \\ 0&0&1&1&0&0&0&-1&1&0&-1&0&-1 \\ 0&0&0&1&1&-1&-1&1&0&0&0&-1&0 \\ 0&0&0&0&1&1&-1&-1&1&0&0&0&-1 \end{pmatrix}$$

Data symbol:

$$a_1, a_2, a_3, a_4, a_5, a_6$$

Update (supposed):

1) 2, 2, 2, 2, 2, 2
2) 1, 4, 1, 4, 1, 1
3) 4, 1, 1, 4, 1, 1

The maximum update frequencies of coded symbol

$G_3^{(1)}$:

1) 8 (preferable)
2) 7 (preferable)
3) 10

$G_3^{(2)}$:

1) 10
2) 11
3) 8 (preferable)

Circulant structure

Circulant structure:

Codes closed under cyclic shifts of codewords.

Example with $q = 2$:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Short description:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Example with $q = 3$:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We can always find the short description when $n = q^2 + q + 1$ according to the (Singer planar) perfect difference set.

Three properties:

- Cyclic,
- Every two different rows will intersect at exactly one point, $M_i \cdot M_{i'} = 1$ for every $i \neq i'$.
- The Hamming weight of each row is $q + 1$.

Repair algorithm

Algorithm A:

While possible, do:

Find a projective line with exactly one erased point or coded symbol (corresponding to a server that is down), and repair this coded symbol or server from the other points (coded symbols or servers) of the line.

Specifically, the coded symbol is repaired to minus the sum of the other coded symbols of the line (over $\mathbb{F}_q$).

$$A_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \\ 1&0&0&0&0&0&0&1&1&0&0&1&0 \\ 0&1&0&0&0&0&0&0&1&1&0&0&1 \\ 1&0&1&0&0&0&0&0&0&1&1&0&0 \\ 0&1&0&1&0&0&0&0&0&0&1&1&0 \\ 0&0&1&0&1&0&0&0&0&0&0&1&1 \\ 1&0&0&1&0&1&0&0&0&0&0&0&1 \end{pmatrix}.$$

Example:

Suppose the set of erased coded symbols is $\{b_1, b_2, b_3, b_4, b_8\}$. At first, **Algorithm A** cannot repair b_1 or b_4, as each line through b_1 or b_4 (each repair group) has an erased symbol. But after repairing $\{b_2, b_3, b_8\}$ respectively from, say, repair groups $\{b_9, b_{10}, b_{13}\}$, $\{b_5, b_{12}, b_{13}\}$, $\{b_7, b_{11}, b_{13}\}$ (here putting total load 3 on server b_{13}), symbols b_1, b_4 can then be repaired from, say, repair groups $\{b_2, b_5, b_7\}$, $\{b_3, b_8, b_{10}\}$.

Stopping sets

The **stopping sets** here are precisely the **sets without tangents** in geometry, that is, sets of projective points intersecting no line in exactly one point.

Sets without tangents are closed under unions, like stopping sets, there is a unique largest set without tangents S that is a subset of a given set T of failed servers.

Example:

Suppose the set of erased coded symbols contains the set $\{b_2, b_3, b_5, b_6, b_7, b_8\}$. This is a stopping set (as it is the union of two lines through b_1, without b_1), so any line meeting this set meets it in at least 2 points. **Algorithm A** never manages to repair any of those servers.

The smallest size s_q of a set without tangents or stopping set:

$$q + \frac{1}{4}\sqrt{2q} + 2 \leq s_q \leq 2q - 2$$

For small odd q:

$s_3 = 6$,
$s_5 = 10$,
$s_7 = 12$,
$s_9 = 15$,
$s_{11} = 18$. (Experimental: 27)

$$A_3 = \begin{pmatrix} 1&1&0&0&1&0&1&0&0&0&0&0&0 \\ 0&1&1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&1&1&0&0&1&0&1&0&0&0 \\ 0&0&0&0&1&1&0&0&1&0&1&0&0 \\ 0&0&0&0&0&1&1&0&0&1&0&1&0 \\ 0&0&0&0&0&0&1&1&0&0&1&0&1 \\ 1&0&0&0&0&0&0&1&1&0&0&1&0 \\ 0&1&0&0&0&0&0&0&1&1&0&0&1 \\ 1&0&1&0&0&0&0&0&0&1&1&0&0 \\ 0&1&0&1&0&0&0&0&0&0&1&1&0 \\ 0&0&1&0&1&0&0&0&0&0&0&1&1 \\ 1&0&0&1&0&1&0&0&0&0&0&0&1 \end{pmatrix}.$$

UNIVERSITY OF TARTU

Thanks for your attention