Codes with locality: constructions and applications to cryptographic protocols

Julien Lavauzelle

École Polytechnique & INRIA Saclay, Université Paris-Saclay

Séminaire UVSQ

13/11/2018

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Outline

1. Codes with locality
   Locality in coding theory, examples
   Lifted projective Reed-Solomon codes
   A combinatorial point of view

2. Private information retrieval from transversal designs
   Private information retrieval (PIR)
   Transversal designs and codes
   A new PIR construction
   Instances

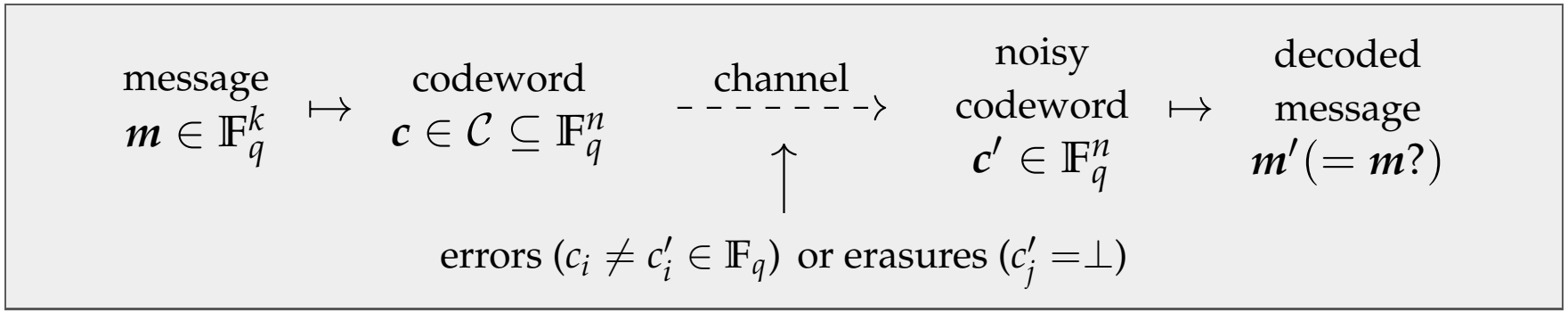
Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

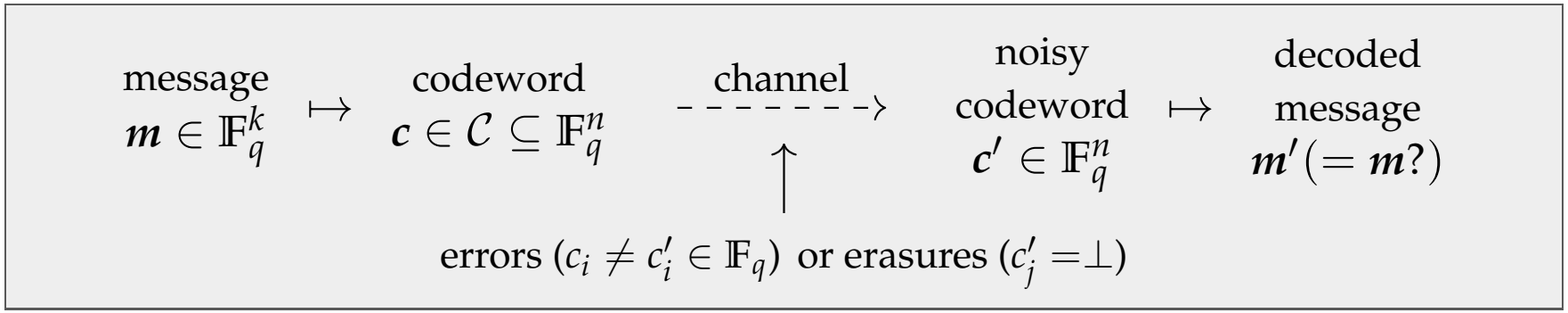
Error-correcting codes

Original goal: transmit information in the presence of noise.



Error-correcting codes

Original goal: transmit information in the presence of noise.



Hamming distance $\mathrm{d}(u, v) := |\{i \in [1, n], u_i \neq v_i\}|$.

Error-correcting codes

Original goal: transmit information in the presence of noise.

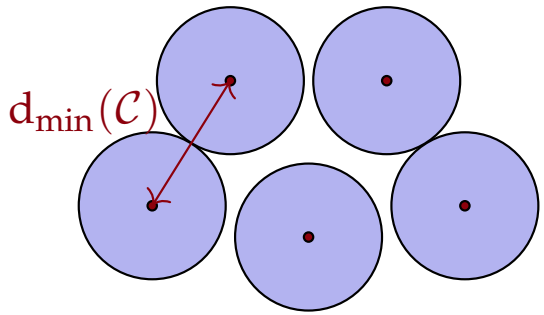
$$\underset{\boldsymbol{m} \in \mathbb{F}_q^k}{\text{message}} \mapsto \underset{\boldsymbol{c} \in \mathcal{C} \subseteq \mathbb{F}_q^n}{\text{codeword}} \xrightarrow{\text{channel}} \underset{\boldsymbol{c}' \in \mathbb{F}_q^n}{\text{noisy codeword}} \mapsto \underset{\boldsymbol{m}' (= \boldsymbol{m}?)}{\text{decoded message}}$$

$$\uparrow$$

errors ($c_i \neq c'_i \in \mathbb{F}_q$) or erasures ($c'_j = \perp$)

Hamming distance $\mathrm{d}(\boldsymbol{u}, \boldsymbol{v}) := |\{i \in [1, n], u_i \neq v_i\}|$.

- $\mathcal{C}$ linear over $\mathbb{F}_q$, with $k = \dim(\mathcal{C})$,
- $d = \mathrm{d}_{\min}(\mathcal{C}) := \min\{\mathrm{d}(\boldsymbol{c}, \boldsymbol{c}'), \boldsymbol{c} \neq \boldsymbol{c}', (\boldsymbol{c}, \boldsymbol{c}') \in \mathcal{C}^2\}$.



Error-correcting codes

Original goal: transmit information in the presence of noise.

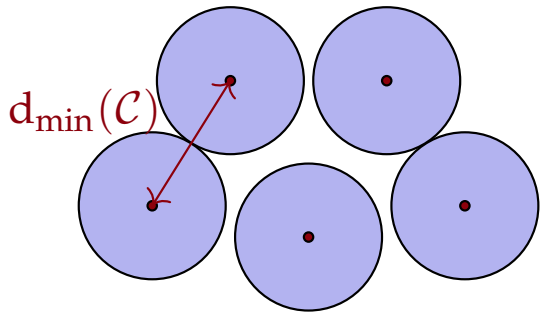
$$\underset{\boldsymbol{m} \in \mathbb{F}_q^k}{\text{message}} \mapsto \underset{\boldsymbol{c} \in \mathcal{C} \subseteq \mathbb{F}_q^n}{\text{codeword}} \xrightarrow{\text{channel}} \underset{\boldsymbol{c}' \in \mathbb{F}_q^n}{\text{noisy codeword}} \mapsto \underset{\boldsymbol{m}' (= \boldsymbol{m}?)}{\text{decoded message}}$$

$$\uparrow$$

errors ($c_i \neq c'_i \in \mathbb{F}_q$) or erasures ($c'_j = \perp$)

Hamming distance $\mathrm{d}(\boldsymbol{u}, \boldsymbol{v}) := |\{i \in [1, n], u_i \neq v_i\}|$.

- $\mathcal{C}$ linear over $\mathbb{F}_q$, with $k = \dim(\mathcal{C})$,
- $d = \mathrm{d}_{\min}(\mathcal{C}) := \min\{\mathrm{d}(\boldsymbol{c}, \boldsymbol{c}'), \boldsymbol{c} \neq \boldsymbol{c}', (\boldsymbol{c}, \boldsymbol{c}') \in \mathcal{C}^2\}$.



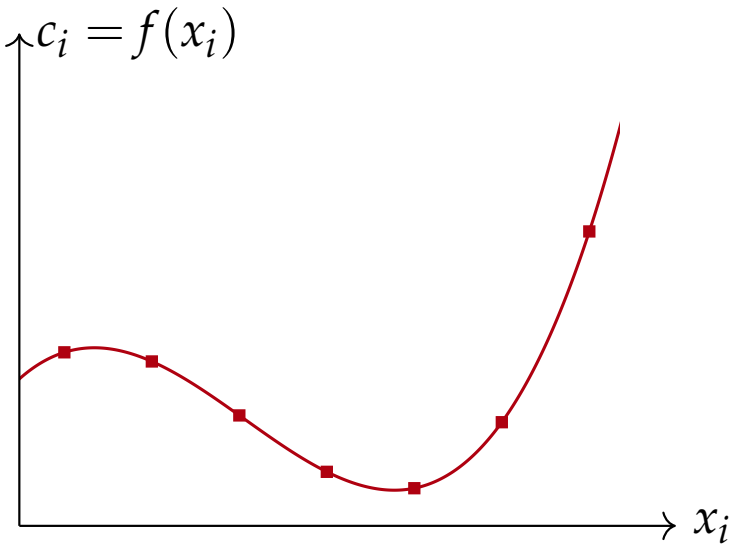
Singleton bound (code is MDS if bound is achieved):

$$k + d \leq n + 1\,.$$

Typical example: Reed-Solomon codes

Definition (Reed-Solomon code). Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$, pairwise distinct.

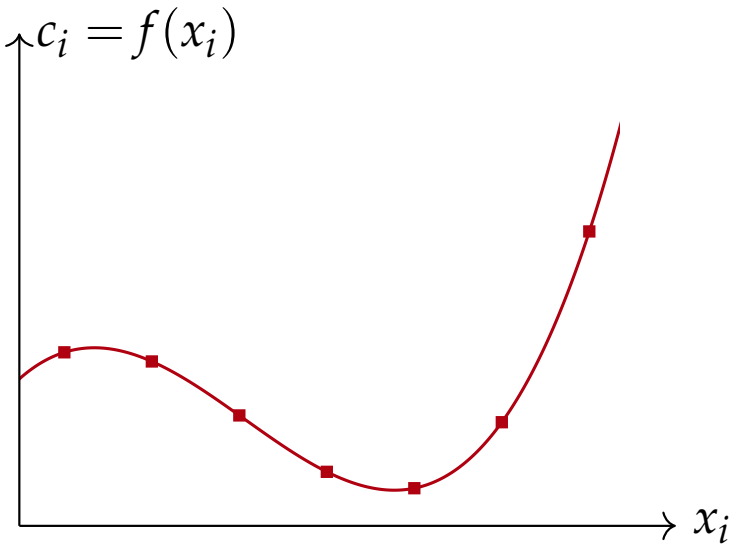
$$\mathrm{RS}_q(r, \boldsymbol{x}) := \{(f(x_1), \ldots, f(x_n)), f \in \mathbb{F}_q[X], \deg f \leq r\}$$



Typical example: Reed-Solomon codes

Definition (Reed-Solomon code). Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$, pairwise distinct.

$$\mathrm{RS}_q(r, \boldsymbol{x}) := \{(f(x_1), \ldots, f(x_n)), f \in \mathbb{F}_q[X], \deg f \leq r\}$$

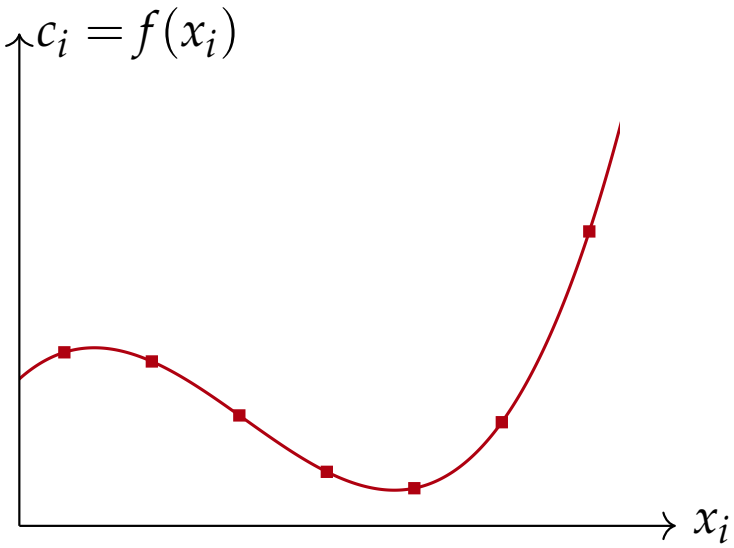


- Dimension $k = r + 1$
- Minimum distance $\mathrm{d}_{\min} = n - r$
  $\Rightarrow$ MDS
- Can decode any b errors and e erasures
  $\rightarrow$ if $e + 2b < \mathrm{d}_{\min}$
  $\rightarrow$ in time $\Theta(n \log^3 n)$.

Typical example: Reed-Solomon codes

Definition (Reed-Solomon code). Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$, pairwise distinct.

$$\mathrm{RS}_q(r, \boldsymbol{x}) := \{(f(x_1), \ldots, f(x_n)), f \in \mathbb{F}_q[X], \deg f \leq r\}$$



- Dimension $k = r + 1$
- Minimum distance $\mathrm{d}_{\min} = n - r$
  $\Rightarrow$ MDS
- Can decode any b errors and e erasures
  $\rightarrow$ if $e + 2b < \mathrm{d}_{\min}$
  $\rightarrow$ in time $\Theta(n \log^3 n)$.

In this talk,

$$\mathrm{RS}_q(r) := \mathrm{RS}_q(r, \mathbb{F}_q)$$

Local correction

Goal: sublinear-time correction of some symbols of $\boldsymbol{c} \in \mathcal{C}$.

Local correction

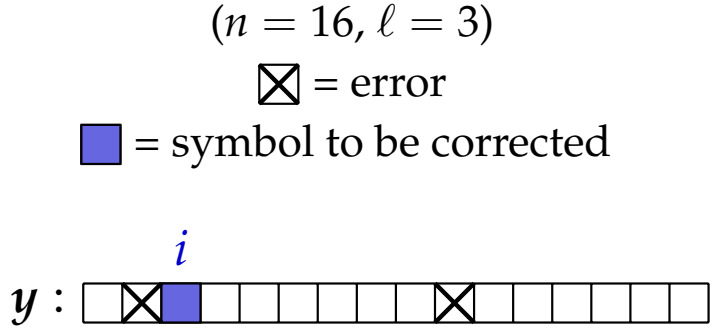
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **locally correctable** with

- **locality** $\ell \leq n$,
- failure probability $\varepsilon \in (0,1)$,
- admissible fraction of errors $\delta \in (0,1)$,

if there exists a **probabilistic algorithm** $\mathcal{D}$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in \mathcal{C}$ satisfying $\mathrm{d}(y,c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(\mathcal{D}^{(y)}(i) = c_i) \geq 1 - \varepsilon$;
- $\mathcal{D}^{(y)}(i)$ queries at most ℓ symbols of y.



Local correction

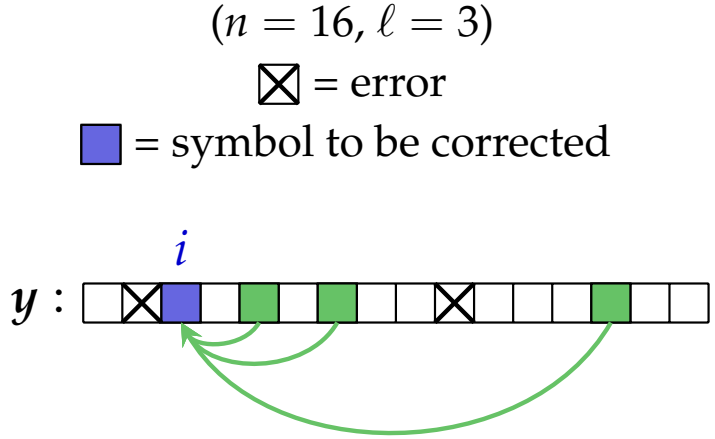
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **locally correctable** with

- **locality** $\ell \leq n$,
- failure probability $\varepsilon \in (0,1)$,
- admissible fraction of errors $\delta \in (0,1)$,

if there exists a **probabilistic algorithm** $\mathcal{D}$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in \mathcal{C}$ satisfying $\mathrm{d}(y,c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(\mathcal{D}^{(y)}(i) = c_i) \geq 1 - \varepsilon$;
- $\mathcal{D}^{(y)}(i)$ queries at most ℓ symbols of y.



Local correction

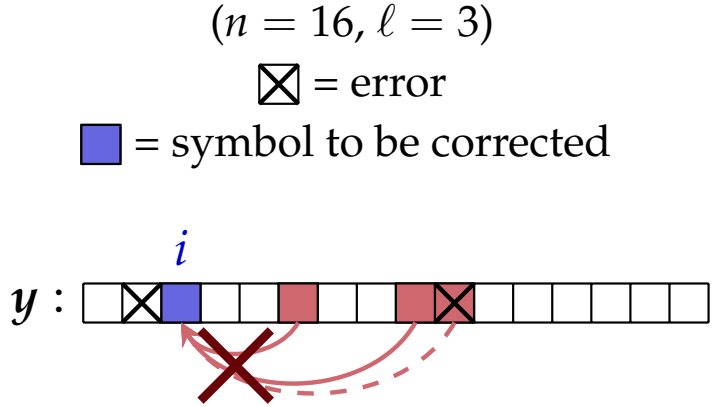
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **locally correctable** with

- **locality** $\ell \leq n$,
- failure probability $\varepsilon \in (0,1)$,
- admissible fraction of errors $\delta \in (0,1)$,

if there exists a **probabilistic algorithm** $\mathcal{D}$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in \mathcal{C}$ satisfying $\mathrm{d}(y,c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(\mathcal{D}^{(y)}(i) = c_i) \geq 1 - \varepsilon$;
- $\mathcal{D}^{(y)}(i)$ queries at most ℓ symbols of y.



Local correction

Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **locally correctable** with

- **locality** $\ell \leq n$,
- failure probability $\varepsilon \in (0,1)$,
- admissible fraction of errors $\delta \in (0,1)$,

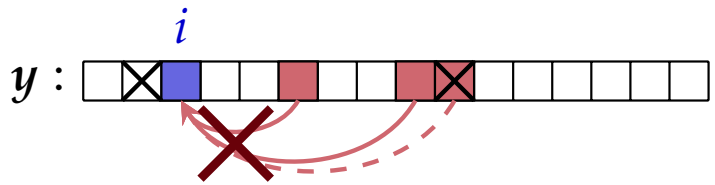
if there exists a **probabilistic algorithm** $\mathcal{D}$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in \mathcal{C}$ satisfying $\mathrm{d}(y,c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(\mathcal{D}^{(y)}(i) = c_i) \geq 1 - \varepsilon$;
- $\mathcal{D}^{(y)}(i)$ queries at most ℓ symbols of y.

$(n = 16, \ell = 3)$

⊠ = error

■ = symbol to be corrected

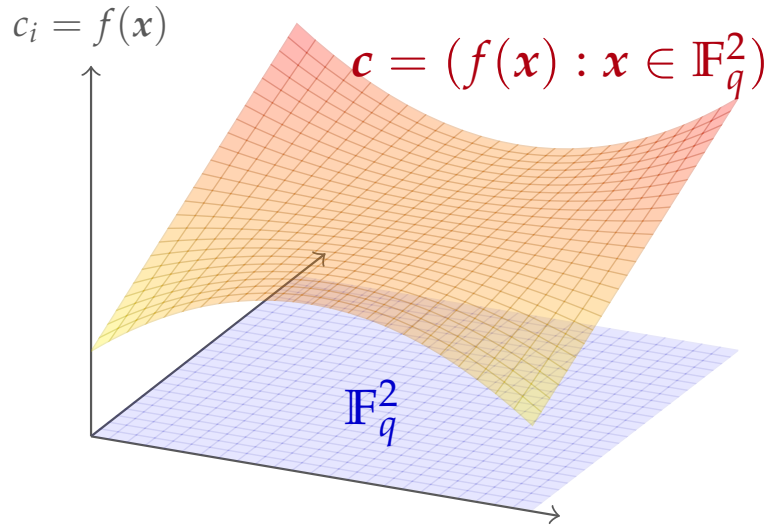


Goals:

- $\ell \ll n$
- $\varepsilon = \mathcal{O}(\delta)$, ideally $\varepsilon = \mathcal{O}(1)$
- $k = \dim \mathcal{C}$ large

Example: Reed-Muller codes

$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

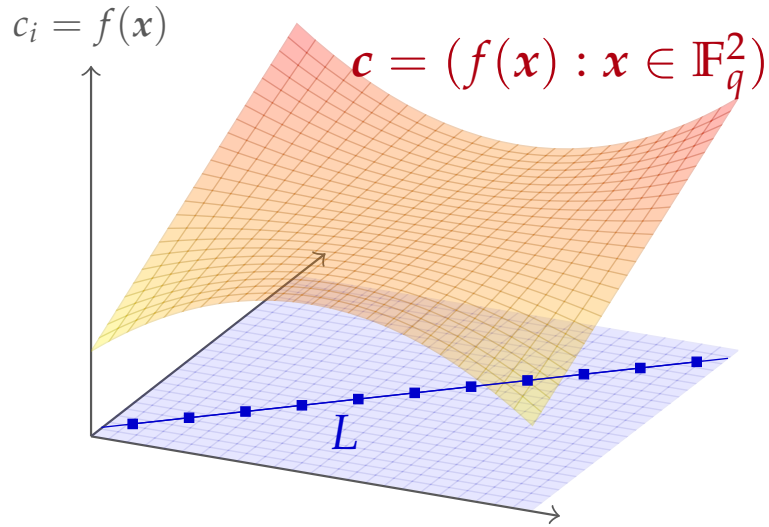


Example: Reed-Muller codes

$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

Assume $r \leq q-2$, and let:

- $\boldsymbol{c} = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathrm{RM}_q(m,r)$
- $\phi : \mathbb{F}_q \to \mathbb{F}_q^m$ affine and injective
- $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$ affine line



Example: Reed-Muller codes

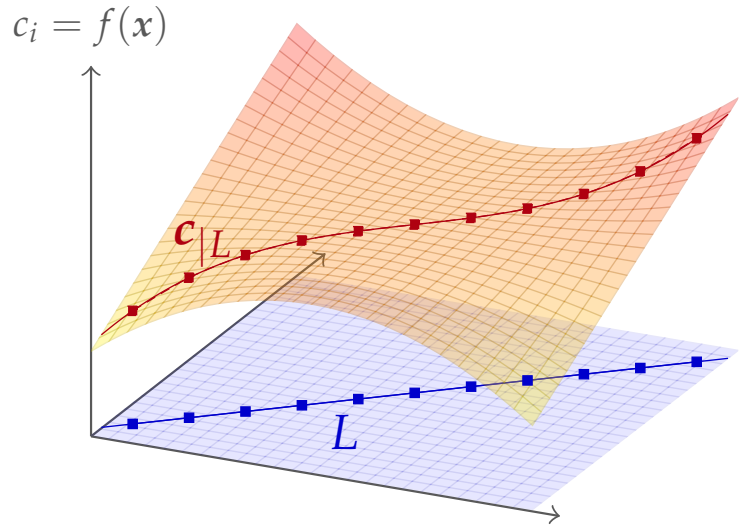
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

Assume $r \leq q-2$, and let:

- $\boldsymbol{c} = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathrm{RM}_q(m,r)$
- $\phi : \mathbb{F}_q \to \mathbb{F}_q^m$ affine and injective
- $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$ affine line

Then, the **restriction** of $\boldsymbol{c}$ to L (*or to* ϕ):

$$\boldsymbol{c}_{|L} := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$



Example: Reed-Muller codes

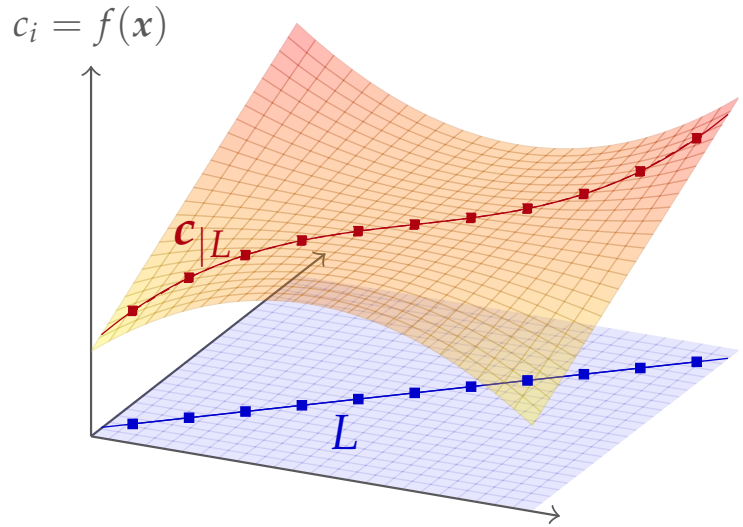
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

Assume $r \leq q-2$, and let:

- $\boldsymbol{c} = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathrm{RM}_q(m,r)$
- $\phi : \mathbb{F}_q \to \mathbb{F}_q^m$ affine and injective
- $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$ affine line

Then, the **restriction** of $\boldsymbol{c}$ to L (*or to* ϕ):

$$\boldsymbol{c}_{|L} := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$



Local correction of $\boldsymbol{y} \in \mathbb{F}_q^{\mathbb{F}_q^m}$ at coordinate $\boldsymbol{i} \in \mathbb{F}_q^m$:

1. Pick at random a line $L \subset \mathbb{F}_q^m$ such that $\boldsymbol{i} \in L$.
2. Correct $\boldsymbol{y}_{|L}$ as a noisy $\mathrm{RS}_q(r)$ codeword, and output $\tilde{y}_{\boldsymbol{i}}$.

Example: Reed-Muller codes

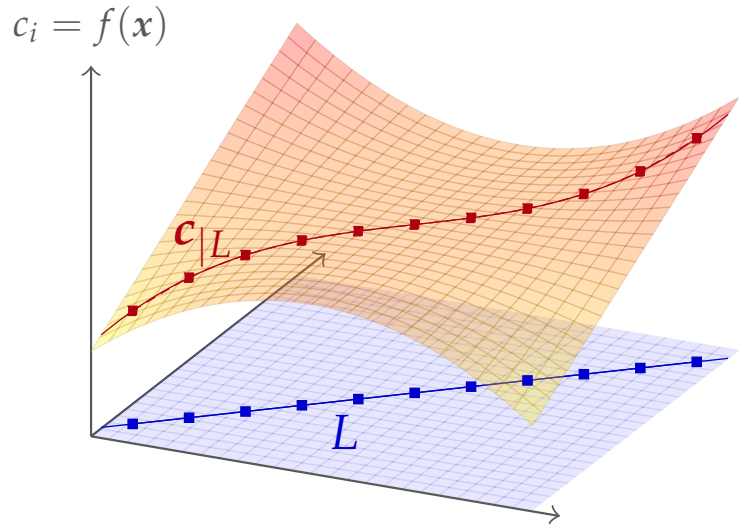
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

Assume $r \leq q-2$, and let:

- $\boldsymbol{c} = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathrm{RM}_q(m,r)$
- $\phi : \mathbb{F}_q \to \mathbb{F}_q^m$ affine and injective
- $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$ affine line

Then, the **restriction** of $\boldsymbol{c}$ to L (*or to* ϕ):

$$\boldsymbol{c}_{|L} := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$



Local correction of $\boldsymbol{y} \in \mathbb{F}_q^{\mathbb{F}_q^m}$ at coordinate $\boldsymbol{i} \in \mathbb{F}_q^m$:

1. Pick at random a line $L \subset \mathbb{F}_q^m$ such that $\boldsymbol{i} \in L$.
2. Correct $\boldsymbol{y}_{|L}$ as a noisy $\mathrm{RS}_q(r)$ codeword, and output $\tilde{y}_{\boldsymbol{i}}$.

$\mathrm{RM}_q(m,r)$ is locally correctable with $\ell = n^{1/m}$ and $\varepsilon = \frac{2\delta}{1-r/q}$.

High-rate construction: lifted codes (1)

Issue: in this setting, rate $\frac{k}{n}$ of RM codes is bounded by $\frac{1}{m!}$.

High-rate construction: lifted codes (1)

Issue: in this setting, rate $\frac{k}{n}$ of RM codes is bounded by $\frac{1}{m!}$.

Idea: consider the set of all polynomials f satisfying the "restriction property":

$$\forall \phi \text{ affine injective}, \quad \big((f \circ \phi)(t) : t \in \mathbb{F}_q\big) \in \mathrm{RS}_q(r)$$

Are there more polynomials than in RM codes?

High-rate construction: lifted codes (1)

Issue: in this setting, rate $\frac{k}{n}$ of RM codes is bounded by $\frac{1}{m!}$.

Idea: consider the set of all polynomials f satisfying the "restriction property":

$$\forall \phi \text{ affine injective}, \quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$

Are there more polynomials than in RM codes?

Example ($q = 4, m = 2, r = 2$). $f(X, Y) = X^2Y^2 \in \mathbb{F}_4[X, Y] \quad \Rightarrow \quad \deg(f) = 4 > r$

$\phi : \mathbb{F}_4 \to \mathbb{F}_4^2, \quad \phi(T) = (aT + b, cT + d)$

High-rate construction: lifted codes (1)

Issue: in this setting, rate $\frac{k}{n}$ of RM codes is bounded by $\frac{1}{m!}$.

Idea: consider the set of all polynomials f satisfying the "restriction property":

$$\forall \phi \text{ affine injective}, \quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$

Are there more polynomials than in RM codes?

Example ($q = 4, m = 2, r = 2$). $f(X, Y) = X^2Y^2 \in \mathbb{F}_4[X, Y] \quad \Rightarrow \quad \deg(f) = 4 > r$

$\phi : \mathbb{F}_4 \to \mathbb{F}_4^2, \quad \phi(T) = (aT + b, cT + d)$

$$\begin{aligned}(f \circ \phi)(T) &= (aT + b)^2(cT + d)^2 \\ &= (a^2T^2 + b^2)(c^2T^2 + d^2) \\ &= (ac)^2T^4 + (ad + bc)^2T^2 + (bd)^2 \\ &= (ad + bc)^2T^2 + (ac)^2T + (bd)^2 \mod (T^4 - T)\end{aligned}$$

$\Rightarrow$ for every ϕ, the "restriction" $(f \circ \phi)(T)$ can be interpolated as a univariate polynomial of degree 2

High-rate construction: lifted codes (2)

- $\mathbb{A}^m := \mathbb{F}_q^m$ $\qquad\qquad$ $\mathrm{ev}_{\mathbb{A}^m}(f) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathbb{F}_q^{\mathbb{A}^m}$
- $\mathrm{Aff}(m) := \{\phi : \mathbb{F}_q \to \mathbb{F}_q^m, \text{injective and affine}\}$

Definition (lifted Reed-Solomon code [GKS13] reformulated).

$$\mathrm{Lift}(\mathrm{RS}_q(r), m) := \{\mathrm{ev}_{\mathbb{A}^m}(f), f \in \mathbb{F}_q[\boldsymbol{X}] \mid \forall \phi \in \mathrm{Aff}(m), \mathrm{ev}_{\mathbb{A}^1}(f \circ \phi) \in \mathrm{RS}_q(r)\}$$

High-rate construction: lifted codes (2)

- $\mathbb{A}^m := \mathbb{F}_q^m$ $\qquad\qquad$ $\mathrm{ev}_{\mathbb{A}^m}(f) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathbb{F}_q^{\mathbb{A}^m}$
- $\mathrm{Aff}(m) := \{\phi : \mathbb{F}_q \to \mathbb{F}_q^m, \text{injective and affine}\}$

Definition (lifted Reed-Solomon code [GKS13] reformulated).

$$\mathrm{Lift}(\mathrm{RS}_q(r), m) := \{\mathrm{ev}_{\mathbb{A}^m}(f), f \in \mathbb{F}_q[\boldsymbol{X}] \mid \forall \phi \in \mathrm{Aff}(m), \mathrm{ev}_{\mathbb{A}^1}(f \circ \phi) \in \mathrm{RS}_q(r)\}$$

$\mathrm{Lift}(\mathrm{RS}_q(r), m)$ is locally correctable with $\ell = n^{1/m}$ and $\varepsilon = \frac{2\delta}{1-r/q}$.

High-rate construction: lifted codes (2)

- $\mathbb{A}^m := \mathbb{F}_q^m$ $\qquad\qquad$ $\mathrm{ev}_{\mathbb{A}^m}(f) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m) \in \mathbb{F}_q^{\mathbb{A}^m}$
- $\mathrm{Aff}(m) := \{\phi : \mathbb{F}_q \to \mathbb{F}_q^m, \text{injective and affine}\}$

Definition (lifted Reed-Solomon code [GKS13] reformulated).

$$\mathrm{Lift}(\mathrm{RS}_q(r), m) := \{\mathrm{ev}_{\mathbb{A}^m}(f), f \in \mathbb{F}_q[\boldsymbol{X}] \mid \forall \phi \in \mathrm{Aff}(m), \mathrm{ev}_{\mathbb{A}^1}(f \circ \phi) \in \mathrm{RS}_q(r)\}$$

$\mathrm{Lift}(\mathrm{RS}_q(r), m)$ is locally correctable with $\ell = n^{1/m}$ and $\varepsilon = \frac{2\delta}{1-r/q}$.

What about the dimension/rate?

Theorem (characteristic 2) [GKS13]. For every $m \geq 2$ and $0 < \alpha < 1$, there exists $0 < \gamma < 1$ and a prime power $q > 0$ such that $\mathrm{Lift}(\mathrm{RS}_q((1-\gamma)q), m)$ is locally correctable with $\ell = n^{1/m}$, $\varepsilon = \Theta_{m,\alpha}(\delta)$, and has rate

$$R \geq 1 - \alpha\,.$$

Rate of lifted codes

Bounds in [GKS13] are **far from being tight**.

- **Ex:** for $m = 2$, GKS' theorem gives $\gamma \leq \alpha^{32}$.

Rate of lifted codes

Bounds in [GKS13] are **far from being tight**.

- **Ex:** for $m = 2$, GKS' theorem gives $\gamma \leq \alpha^{32}$.

Theorem [characteristic 2, finite length $n = q^2 = 2^{2e}$].
For $m = 2$, $q = 2^e$ and $r = (1 - 2^{-c})q - 1$,

$$R = 1 - \frac{5}{4}\left(\frac{3}{4}\right)^c + \frac{1}{4}\left(\frac{1}{4}\right)^c + \frac{1}{2^e}\left(\frac{3^c - 1}{2^{c+2}}\right).$$

- actually, $\gamma \leq \alpha^3$ (roughly) is enough

Rate of lifted codes

Bounds in [GKS13] are **far from being tight**.

- **Ex:** for $m = 2$, GKS' theorem gives $\gamma \leq \alpha^{32}$.

Theorem [characteristic 2, finite length $n = q^2 = 2^{2e}$].
For $m = 2$, $q = 2^e$ and $r = (1 - 2^{-c})q - 1$,

$$R = 1 - \frac{5}{4}\left(\frac{3}{4}\right)^c + \frac{1}{4}\left(\frac{1}{4}\right)^c + \frac{1}{2^e}\left(\frac{3^c - 1}{2^{c+2}}\right).$$

- actually, $\gamma \leq \alpha^3$ (roughly) is enough

Theorem [characteristic p, asymptotic length $n = p^{2e}$, $e \to \infty$].
For $m = 2$, $q = p^e \to \infty$ and $r = (1 - p^{-c})q - 1 \to \infty$,

$$R_{(e\to\infty)} = 1 - \left(1 + \frac{1}{p+2}\right)\left(\frac{1 + 1/p}{2}\right)^c + \frac{1}{p+2}\left(\frac{1}{p^2}\right)^c.$$

Degree sets

Lifted codes are **monomial**, *i.e.* generated by evaluations of monomials

$$\mathrm{ev}_{\mathbb{A}^m}(X_1^{d_1} \dots X_m^{d_m}) = \mathrm{ev}_{\mathbb{A}^m}(\boldsymbol{X}^{\boldsymbol{d}})$$

Degree set of a monomial code:

$$\mathrm{Deg}(\mathcal{C}) := \{\boldsymbol{d} \in [0, q-1]^m, \mathrm{ev}_{\mathbb{A}^m}(\boldsymbol{X}^{\boldsymbol{d}}) \in \mathcal{C}\}$$

Example for $\mathcal{C} = \mathrm{RM}(m, r)$:

$$\mathrm{Deg}(\mathcal{C}) = \{\boldsymbol{d} \in [0, q-1]^m, \sum_{i=1}^{m} d_i \leq r\}$$

Degree sets

Lifted codes are **monomial**, *i.e.* generated by evaluations of monomials

$$\mathrm{ev}_{\mathbb{A}^m}(X_1^{d_1} \dots X_m^{d_m}) = \mathrm{ev}_{\mathbb{A}^m}(\boldsymbol{X}^{\boldsymbol{d}})$$

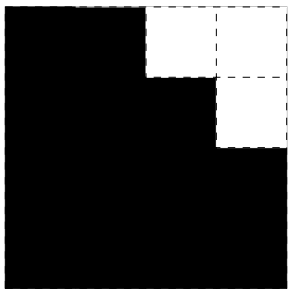
Degree set of a monomial code:

$$\mathrm{Deg}(\mathcal{C}) := \{\boldsymbol{d} \in [0, q-1]^m, \mathrm{ev}_{\mathbb{A}^m}(\boldsymbol{X}^{\boldsymbol{d}}) \in \mathcal{C}\}$$

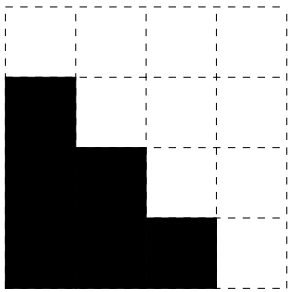
Example for $\mathcal{C} = \mathrm{RM}(m, r)$:

$$\mathrm{Deg}(\mathcal{C}) = \{\boldsymbol{d} \in [0, q-1]^m, \sum_{i=1}^{m} d_i \leq r\}$$

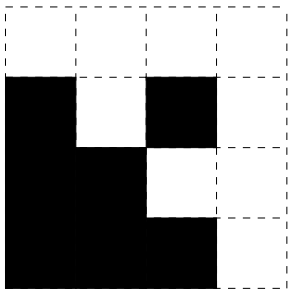
A **representation** for $m = 2$:



$\mathrm{RM}_4(2,4)$

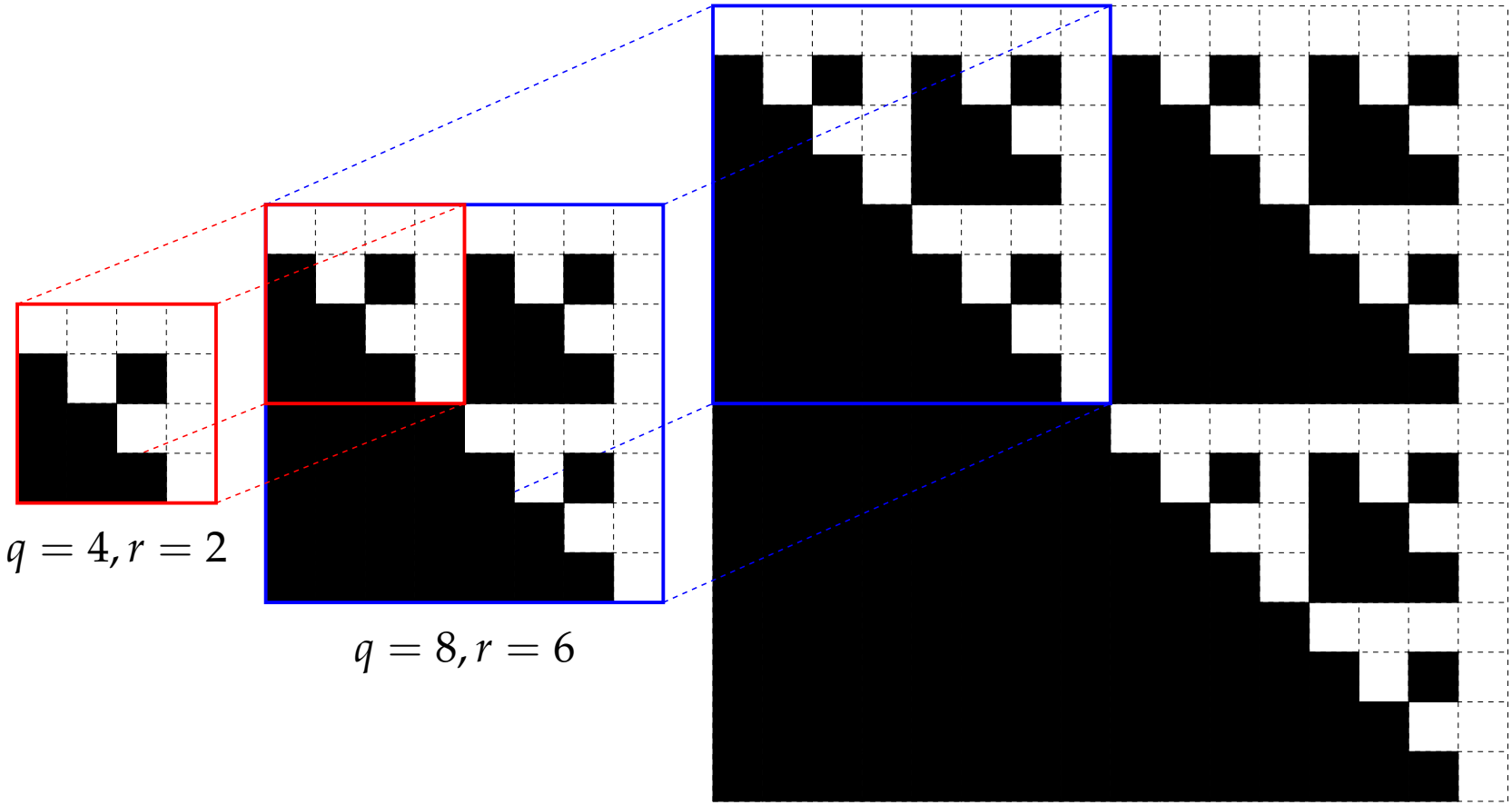


$\mathrm{RM}_4(2,2)$



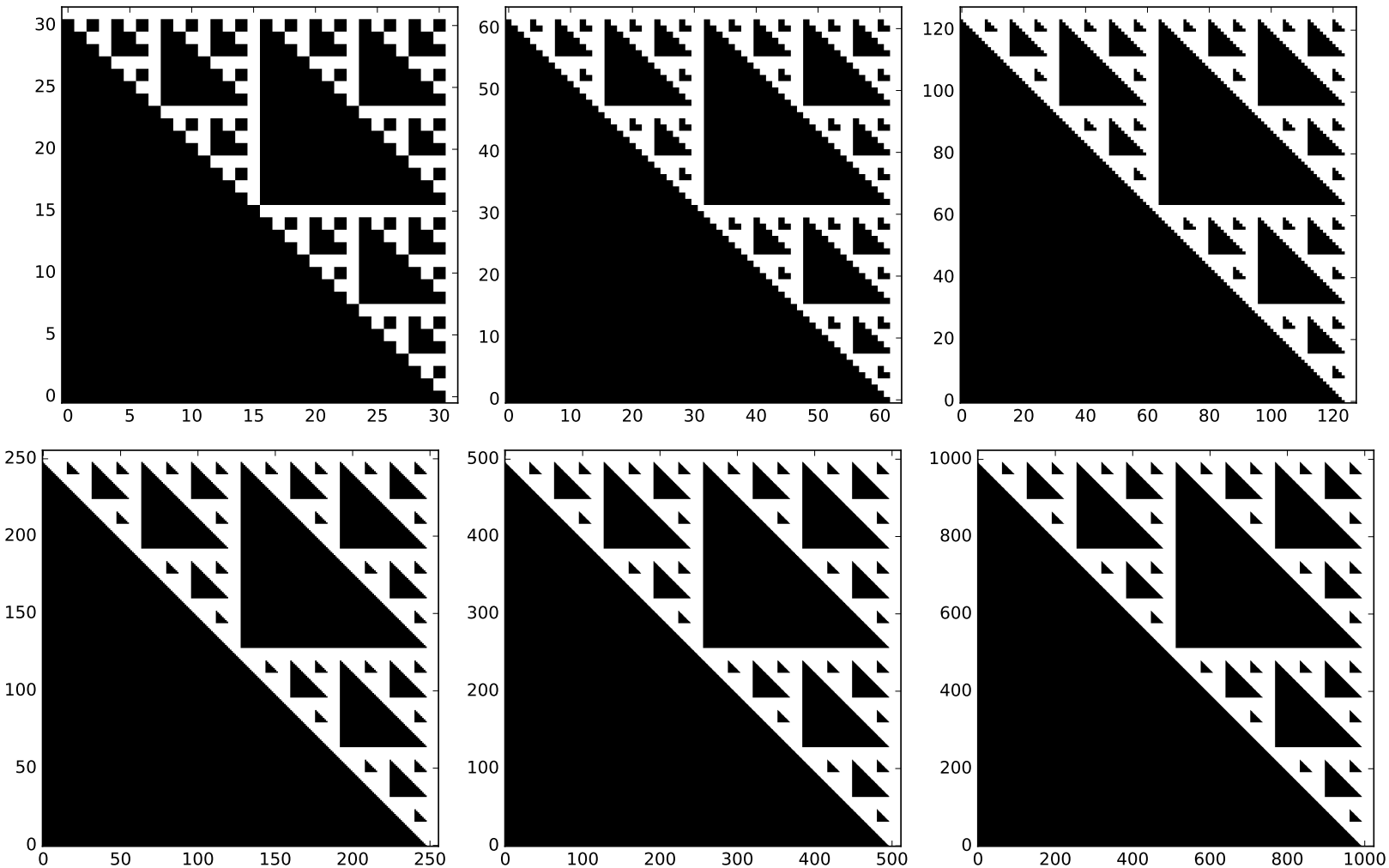
$\mathrm{Lift}(\mathrm{RS}_4(2),2)$

“Fractal” representation of degree sets (1)



“Fractal” representation of degree sets (2)

Degree set of $\mathrm{Lift}(\mathrm{RS}_{2^e}((1-2^{-c})2^e-1),2)$ for fixed $c=5$ and increasing $e \geq 5$.



Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Motivation

Why would we consider lifted codes over **projective spaces**?

- projective versions of Reed-Solomon and Reed-Muller codes **already exist**
- lifted projective RS codes would have slightly **larger length**
- relations between affine and projective RM codes via **puncturing** and **shortening**, *e.g.*

$$0 \to \mathrm{RM}_q(m, k-1) \to \mathrm{PRM}_q(m, k) \xrightarrow{\pi} \mathrm{PRM}_q(m-1, k) \to 0\,.$$

where π is the restriction map $\mathbb{P}^m \to \mathbb{P}^{m-1}$

Projective spaces

Projective space:

$$\mathbb{P}^m := \mathbb{A}^{m+1} / \sim$$

where $\boldsymbol{a} \sim \boldsymbol{b}$ iff $\exists \lambda \in \mathbb{F}_q^\times, \boldsymbol{a} = \lambda \boldsymbol{b}$

Projective spaces

Projective space:

$$\mathbb{P}^m := \mathbb{A}^{m+1} / \sim$$

where $\boldsymbol{a} \sim \boldsymbol{b}$ iff $\exists \lambda \in \mathbb{F}_q^\times, \boldsymbol{a} = \lambda \boldsymbol{b}$

Defining an **evaluation map** over $\mathbb{P}^m$ requires:

- **homogeneous** polynomials $f \in \mathbb{F}_q[\boldsymbol{X}]_v^H$ of fixed degree v,
- to choose a **representative** for every $\boldsymbol{u} \in \mathbb{P}^m$:

$$\boldsymbol{u} = (0 : \cdots : 0 : 1 : * : \cdots : *) \in \mathbb{P}^m$$

Projective spaces

Projective space:

$$\mathbb{P}^m := \mathbb{A}^{m+1} / \sim$$

where $\boldsymbol{a} \sim \boldsymbol{b}$ iff $\exists \lambda \in \mathbb{F}_q^\times, \boldsymbol{a} = \lambda \boldsymbol{b}$

Defining an **evaluation map** over $\mathbb{P}^m$ requires:

- **homogeneous** polynomials $f \in \mathbb{F}_q[X]_v^H$ of fixed degree v,
- to choose a **representative** for every $\boldsymbol{u} \in \mathbb{P}^m$:

$$\boldsymbol{u} = (0 : \cdots : 0 : 1 : * : \cdots : *) \in \mathbb{P}^m$$

We get:

$$f(\boldsymbol{u}) := f(0, \ldots, 0, 1, *, \ldots, *) \in \mathbb{F}_q$$

$$\mathrm{ev}_{\mathbb{P}^m}(f) := (f(\boldsymbol{u}) : \boldsymbol{u} \in \mathbb{P}^m) \in \mathbb{F}_q^{\mathbb{P}^m}$$

The construction

Example. Projective Reed-Solomon code:

$$\mathrm{PRS}_q(r) = \{\mathrm{ev}_{\mathbb{P}^1}(f) = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{P}^1), f \in \mathbb{F}_q[X, Y]_r^H\}$$

The construction

Example. Projective Reed-Solomon code:

$$\mathrm{PRS}_q(r) = \{\mathrm{ev}_{\mathbb{P}^1}(f) = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{P}^1), f \in \mathbb{F}_q[X, Y]_r^H\}$$

Let $\mathrm{Proj}(m) := \{\phi : \mathbb{F}_q^2 \to \mathbb{F}_q^{m+1} \text{ injective}\}$.

Definition (lifted projective RS codes). Let $v = r + (m-1)(q-1)$.

$$\mathrm{Lift}(\mathrm{PRS}_q(r), m) := \{\mathrm{ev}_{\mathbb{P}^m}(f), f \in \mathbb{F}_q[\boldsymbol{X}]_v^H \mid \forall \phi \in \mathrm{Proj}(m), \mathrm{ev}_{\mathbb{P}^1}(f \circ \phi) \in \mathrm{PRS}_q(r)\}$$

Example. Projective Reed-Solomon code:

$$\mathrm{PRS}_q(r) = \{\mathrm{ev}_{\mathbb{P}^1}(f) = (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{P}^1), f \in \mathbb{F}_q[X, Y]_r^H\}$$

Let $\mathrm{Proj}(m) := \{\phi : \mathbb{F}_q^2 \to \mathbb{F}_q^{m+1} \text{ injective}\}$.

Definition (lifted projective RS codes). Let $v = r + (m-1)(q-1)$.

$$\mathrm{Lift}(\mathrm{PRS}_q(r), m) := \{\mathrm{ev}_{\mathbb{P}^m}(f), f \in \mathbb{F}_q[\boldsymbol{X}]_v^H \mid \forall \phi \in \mathrm{Proj}(m), \mathrm{ev}_{\mathbb{P}^1}(f \circ \phi) \in \mathrm{PRS}_q(r)\}$$

Remarks:

- $\mathrm{ev}_{\mathbb{P}^1}(f \circ \phi) \neq \mathrm{ev}_{\mathbb{P}^m}(f)_{|\phi(\mathbb{P}^1)}$ due to the choice of representative
- fortunately $\mathrm{ev}_{\mathbb{P}^1}(f \circ \phi) = \boldsymbol{w} \star \mathrm{ev}_{\mathbb{P}^m}(f)_{|\phi(\mathbb{P}^1)}$, and $\boldsymbol{w}$ is **independent of** f.

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of $\mathrm{Lift}(\mathrm{RS}_q(r-1), m)$, $\mathrm{Lift}(\mathrm{PRS}_q(r), m)$ and $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of $\mathrm{Lift}(\mathrm{RS}_q(r-1), m)$, $\mathrm{Lift}(\mathrm{PRS}_q(r), m)$ and $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$
- satisfy the **puncturing/shortening** relation

$$0 \to \mathrm{Lift}(\mathrm{RS}_q(r-1), m) \to \mathrm{Lift}(\mathrm{PRS}_q(r), m) \xrightarrow{\pi} \mathrm{Lift}(\mathrm{PRS}_q(r), m-1) \to 0\,,$$

where $\pi : \mathbb{P}^m \to \mathbb{P}^{m-1}$.

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of $\mathrm{Lift}(\mathrm{RS}_q(r-1), m)$, $\mathrm{Lift}(\mathrm{PRS}_q(r), m)$ and $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$
- satisfy the **puncturing/shortening** relation

$$0 \to \mathrm{Lift}(\mathrm{RS}_q(r-1), m) \to \mathrm{Lift}(\mathrm{PRS}_q(r), m) \xrightarrow{\pi} \mathrm{Lift}(\mathrm{PRS}_q(r), m-1) \to 0\,,$$

  where $\pi : \mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence) **cyclic codes** if $(q-1)^2 \nmid (q^{m+1}-1)$
  quasi-cyclic codes otherwise

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of $\mathrm{Lift}(\mathrm{RS}_q(r-1), m)$, $\mathrm{Lift}(\mathrm{PRS}_q(r), m)$ and $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$
- satisfy the **puncturing/shortening** relation

$$0 \to \mathrm{Lift}(\mathrm{RS}_q(r-1), m) \to \mathrm{Lift}(\mathrm{PRS}_q(r), m) \xrightarrow{\pi} \mathrm{Lift}(\mathrm{PRS}_q(r), m-1) \to 0\,,$$

  where $\pi : \mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence) **cyclic codes** if $(q-1)^2 \nmid (q^{m+1}-1)$; quasi-cyclic codes otherwise
- admit many explicit and easily computable **information sets**

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of $\mathrm{Lift}(\mathrm{RS}_q(r-1), m)$, $\mathrm{Lift}(\mathrm{PRS}_q(r), m)$ and $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$
- satisfy the **puncturing/shortening** relation

$$0 \to \mathrm{Lift}(\mathrm{RS}_q(r-1), m) \to \mathrm{Lift}(\mathrm{PRS}_q(r), m) \xrightarrow{\pi} \mathrm{Lift}(\mathrm{PRS}_q(r), m-1) \to 0\,,$$

  where $\pi : \mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence) **cyclic codes** if $(q-1)^2 \nmid (q^{m+1}-1)$; quasi-cyclic codes otherwise
- admit many explicit and easily computable **information sets**

Details in:
Lifted Projective Reed-Solomon Codes, L., DCC, to appear
10.1007/s10623-018-0552-8

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Lifted codes when $r = q - 2$

Remark. Assume $r = q - 2$. Then,

$$\boldsymbol{a} \in \mathrm{RS}_q(q-2) \iff \sum_{i=1}^{q} a_i = 0 \iff \langle \mathbf{1}, \boldsymbol{a} \rangle = 0$$

Lifted codes when $r = q - 2$

Remark. Assume $r = q - 2$. Then,

$$\boldsymbol{a} \in \mathrm{RS}_q(q-2) \iff \sum_{i=1}^{q} a_i = 0 \iff \langle \mathbf{1}, \boldsymbol{a} \rangle = 0$$

$$\boldsymbol{c} \in \mathrm{Lift}(\mathrm{RS}_q(q-2), m) \iff \forall L \subseteq \mathbb{F}_q^m, \langle \mathbf{1}, \boldsymbol{c}_{|L} \rangle = 0$$

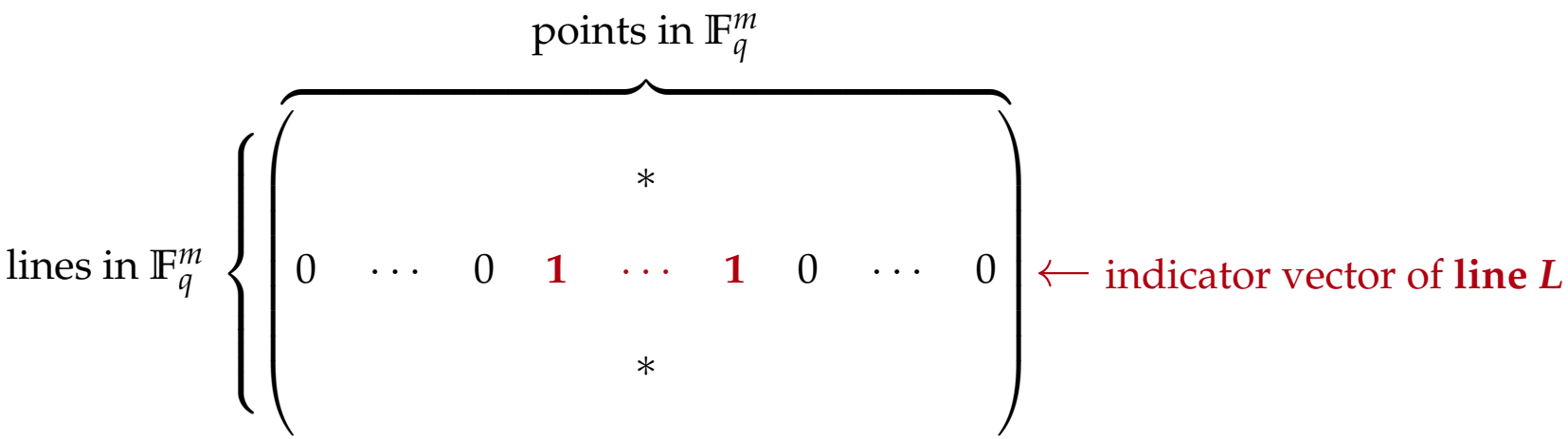
Lifted codes when $r = q - 2$

Remark. Assume $r = q - 2$. Then,

$$\boldsymbol{a} \in \mathrm{RS}_q(q-2) \iff \sum_{i=1}^{q} a_i = 0 \iff \langle \mathbf{1}, \boldsymbol{a} \rangle = 0$$

$$\boldsymbol{c} \in \mathrm{Lift}(\mathrm{RS}_q(q-2), m) \iff \forall L \subseteq \mathbb{F}_q^m, \langle \mathbf{1}, \boldsymbol{c}_{|L} \rangle = 0$$

Parity-check matrix for $\mathrm{Lift}(\mathrm{RS}_q(q-2), m)$:



Block designs

Point-line incidences in the affine space form a 2**-design.**

Definition. A *t-design* of parameters (v, k, λ) consists in:

- a set X of points, $|X| = v$,
- a set $\mathcal{B}$ of blocks $B \subset X$, $|B| = k$

such that every t-set in X appears in exactly λ blocks.

Block designs

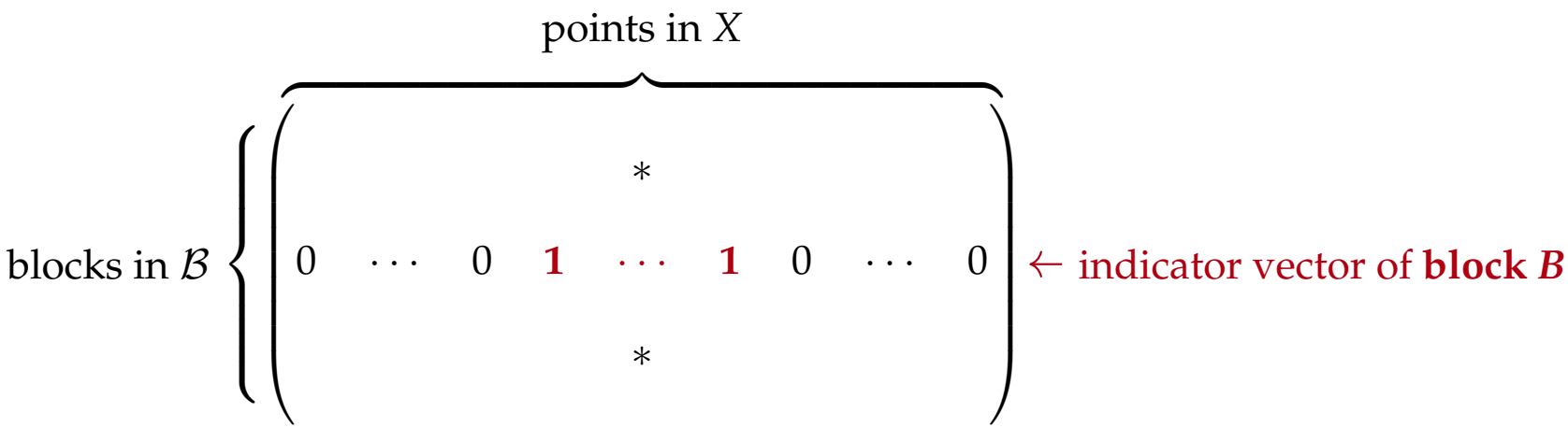
Point-line incidences in the affine space form a 2**-design.**

Definition. A *t-design* of parameters (v, k, λ) consists in:

- a set X of points, $|X| = v$,
- a set $\mathcal{B}$ of blocks $B \subset X$, $|B| = k$

such that every t-set in X appears in exactly λ blocks.

Incidence matrix of a design:



Codes based on designs, and generalisation

The **code based on the design** $\mathcal{D} = (X, \mathcal{B})$ is the code $\mathcal{C} = \mathrm{Code}(\mathcal{D}) \subseteq \mathbb{F}_q^X$ admitting the incidence matrix of $\mathcal{D}$ as a parity-check matrix.

$$\mathrm{Code}(\mathcal{D}) = \{\boldsymbol{c} \in \mathbb{F}_q^X \mid \forall B \in \mathcal{B}, \sum_{x \in B} c_x = 0\}$$

Remark. The dimension of $\mathrm{Code}(\mathcal{D})$ highly depends on the field $\mathbb{F}_q$.

Codes based on designs, and generalisation

The **code based on the design** $\mathcal{D} = (X, \mathcal{B})$ is the code $\mathcal{C} = \mathrm{Code}(\mathcal{D}) \subseteq \mathbb{F}_q^X$ admitting the incidence matrix of $\mathcal{D}$ as a parity-check matrix.

$$\mathrm{Code}(\mathcal{D}) = \{\boldsymbol{c} \in \mathbb{F}_q^X \mid \forall B \in \mathcal{B}, \sum_{x \in B} c_x = 0\}$$

Remark. The dimension of $\mathrm{Code}(\mathcal{D})$ highly depends on the field $\mathbb{F}_q$.

Let $\mathcal{L} = (\mathcal{L}_B : B \in \mathcal{B})$ be a family of codes indexed by blocks $B \in \mathcal{B}$. The **generalised design-based code**, based on $(\mathcal{D}, \mathcal{L})$ is

$$\mathrm{Code}(\mathcal{D}, \mathcal{L}) := \{\boldsymbol{c} \in \mathbb{F}_q^X \mid \forall B \in \mathcal{B}, \boldsymbol{c}_{|B} \in \mathcal{L}_B\}.$$

Remark. $\mathrm{Code}(\mathcal{D}, \mathcal{L}) = \mathrm{Code}(\mathcal{D})$ if every code in $\mathcal{L}$ is a parity-check code.

Design-based codes and LCCs

- $\mathcal{D}$ be a t-$(n, \ell+1, \lambda)$-design
- $0 < \tau < 1$
- $\mathcal{L} = (\mathcal{L}_B : B \in \mathcal{B})$ s.t. every code in $\mathcal{L}$ corrects $\lfloor \tau\ell \rfloor$ errors and 1 erasure.

Algorithm. Local correction of $\boldsymbol{y} \in \mathbb{F}_q^X$ at $i \in X$

- Pick uniformly at random a block $B \in \mathcal{B}$ such that $i \in X$.
- Correct $\boldsymbol{y}_{|B}$ as a noisy codeword from $\mathcal{L}_B$.
- Output the corrected symbol $\tilde{y}_i$.

Design-based codes and LCCs

- $\mathcal{D}$ be a t-$(n, \ell+1, \lambda)$-design
- $0 < \tau < 1$
- $\mathcal{L} = (\mathcal{L}_B : B \in \mathcal{B})$ s.t. every code in $\mathcal{L}$ corrects $\lfloor \tau\ell \rfloor$ errors and 1 erasure.

Algorithm. Local correction of $y \in \mathbb{F}_q^X$ at $i \in X$

- Pick uniformly at random a block $B \in \mathcal{B}$ such that $i \in X$.
- Correct $y_{|B}$ as a noisy codeword from $\mathcal{L}_B$.
- Output the corrected symbol $\tilde{y}_i$.

Proposition [$t = 2$]. For every $\delta < \tau/2$, $\mathrm{Code}(\mathcal{D}, \mathcal{L})$ is a $(\ell, \delta, \varepsilon)$-LCC, where

$$\varepsilon = \delta/\tau\,.$$

Design-based codes and LCCs

- $\mathcal{D}$ be a t-$(n, \ell+1, \lambda)$-design
- $0 < \tau < 1$
- $\mathcal{L} = (\mathcal{L}_B : B \in \mathcal{B})$ s.t. every code in $\mathcal{L}$ corrects $\lfloor \tau\ell \rfloor$ errors and 1 erasure.

Algorithm. Local correction of $y \in \mathbb{F}_q^X$ at $i \in X$

- Pick uniformly at random a block $B \in \mathcal{B}$ such that $i \in X$.
- Correct $y_{|B}$ as a noisy codeword from $\mathcal{L}_B$.
- Output the corrected symbol $\tilde{y}_i$.

Proposition [$t = 2$]. For every $\delta < \tau/2$, $\mathrm{Code}(\mathcal{D}, \mathcal{L})$ is a $(\ell, \delta, \varepsilon)$-LCC, where

$$\varepsilon = \delta/\tau\,.$$

Proposition [$t = 3$]. For every $\delta < \tau - \sqrt{2/\ell}$, $\mathrm{Code}(\mathcal{D}, \mathcal{L})$ is a $(\ell, \delta, \varepsilon)$-LCC where

$$\varepsilon = \frac{\delta(1-\delta)}{(\tau-\delta)^2} \cdot \frac{1}{\ell}\,.$$

Future works

Design-based codes allow to get rid of probabilistic decoders in the definition of locally correctable codes

$\rightarrow$ "combinatorial" coding-theoretic version of LCCs

Future works

Design-based codes allow to get rid of probabilistic decoders in the definition of locally correctable codes

$\rightarrow$ "combinatorial" coding-theoretic version of LCCs

Remaining issues:

- families of 3-designs with high dimension?
- best instances $(\mathcal{D}, \mathcal{L})$ prescribed design parameters (n, ℓ, λ)?

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Informal issue

Given a database $F \in \mathbb{F}_q^k$ and $1 \leq i \leq k$,

can we **retrieve** the entry F_i,

without leaking any information on the index i?

Informal issue

Given a database $F \in \mathbb{F}_q^k$ and $1 \leq i \leq k$,

can we **retrieve** the entry F_i,

without leaking any information on the index i?

Remark:

- PIR $\neq$ anonymity (hidden user)
- PIR $\neq$ encryption (hidden data)

Definition

File F encoded and stored on ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol:

(user U wants to recover F_i privately)

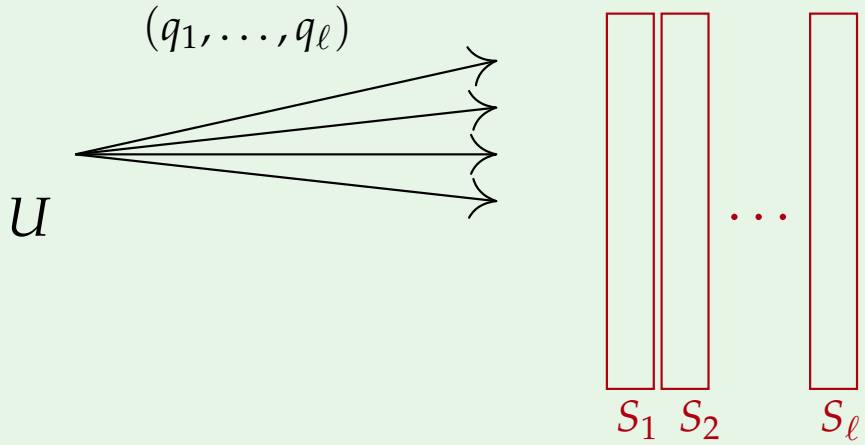
Definition

File F encoded and stored on ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol:

(user U wants to recover F_i privately)

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_\ell) \leftarrow \mathcal{Q}(i)$ and sends q_j to S_j



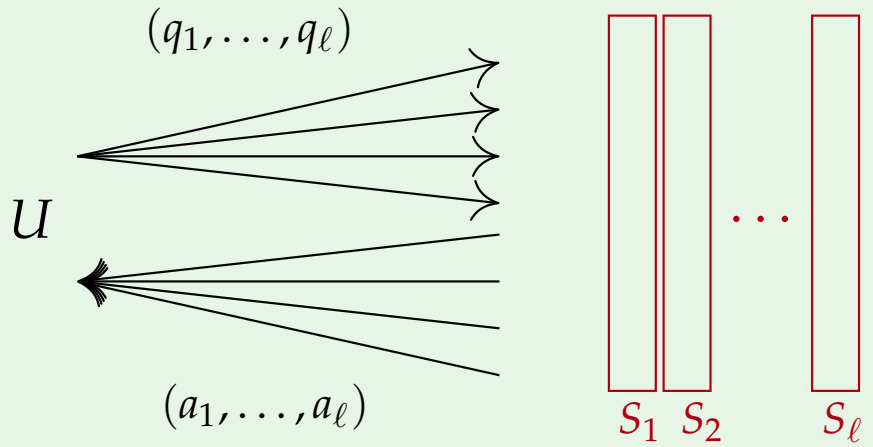
Definition

File F encoded and stored on ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol:

(user U wants to recover F_i privately)

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_\ell) \leftarrow \mathcal{Q}(i)$ and sends q_j to S_j
2. Each server S_j computes $a_j = \mathcal{A}_j(q_j, F_{|S_j})$ and sends it back to U



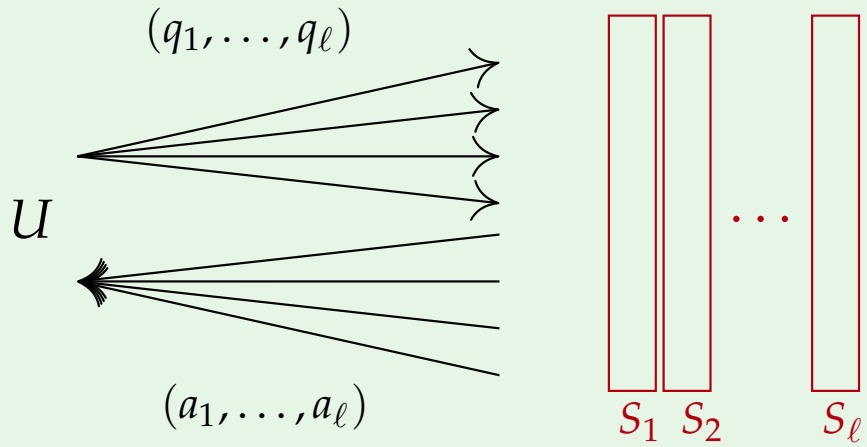
Definition

File F encoded and stored on ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol:

(user U wants to recover F_i privately)

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_\ell) \leftarrow \mathcal{Q}(i)$ and sends q_j to S_j
2. Each server S_j computes $a_j = \mathcal{A}_j(q_j, F_{|S_j})$ and sends it back to U
3. U recovers $F_i = \mathcal{R}(\boldsymbol{q}, \boldsymbol{a}, i)$



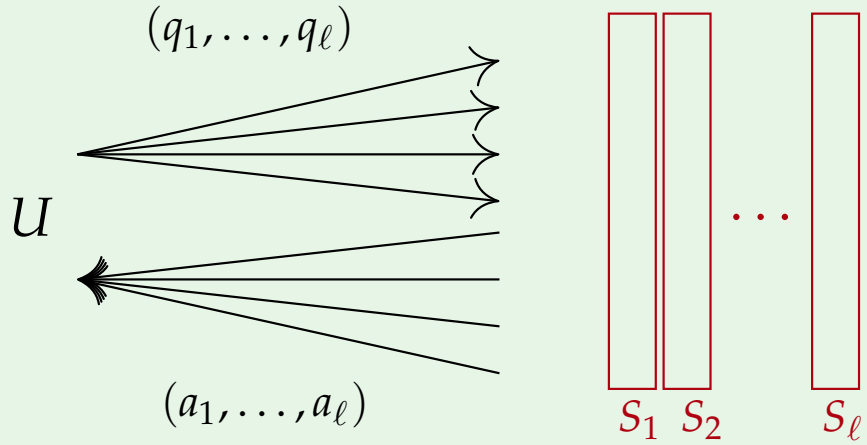
Definition

File F encoded and stored on ℓ servers $S_1, \ldots, S_\ell$.

Private Information Retrieval (PIR) protocol:

(user U wants to recover F_i privately)

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_\ell) \leftarrow \mathcal{Q}(i)$ and sends q_j to S_j
2. Each server S_j computes $a_j = \mathcal{A}_j(q_j, F_{|S_j})$ and sends it back to U
3. U recovers $F_i = \mathcal{R}(\boldsymbol{q}, \boldsymbol{a}, i)$



Information-theoretic privacy: $\mathrm{I}(i\,; q_j) = 0,\ \forall j = 1, \ldots, \ell.$

Motivation

Usual goals for PIR:

- Low communication complexity
- Low storage overhead for the servers (if coded)
- Low computation complexity for algorithms $\mathcal{A}$ (server) and $\mathcal{R}$ (user)

Our context: file F is static and very frequently queried (*e.g.* public database)

Motivation

Usual goals for PIR:

- Low communication complexity
- Low storage overhead for the servers (if coded)
- Low computation complexity for algorithms $\mathcal{A}$ (server) and $\mathcal{R}$ (user)

Our context: file F is static and very frequently queried (*e.g.* public database)

Notion of **price of privacy** for the servers, mainly depends on:

- computational complexity for the servers,
- servers' storage overhead.

Motivation

Usual goals for PIR:

- Low communication complexity
- Low storage overhead for the servers (if coded)
- Low computation complexity for algorithms $\mathcal{A}$ (server) and $\mathcal{R}$ (user)

Our context: file F is static and very frequently queried (*e.g.* public database)

Notion of **price of privacy** for the servers, mainly depends on:

- computational complexity for the servers,
- servers' storage overhead.

Yekhanin (survey, '12): "the **overwhelming computational complexity** of PIR schemes (...) currently presents the **main bottleneck** to their practical deployment".

Outline

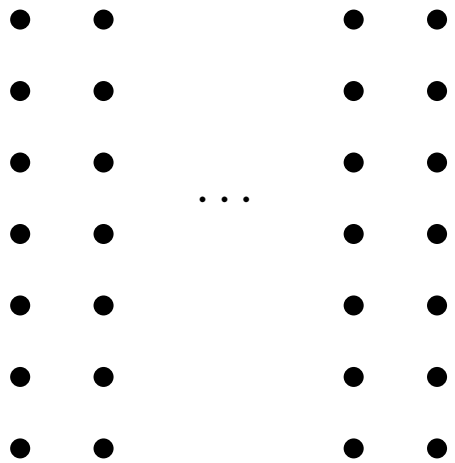
1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Transversal designs

A **transversal design** $\mathrm{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = n = s\ell$,

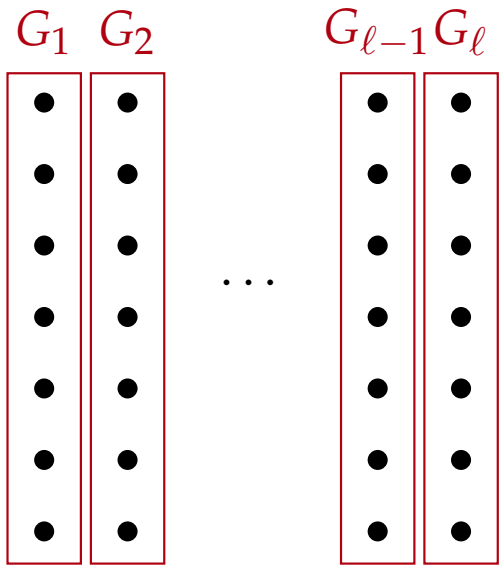


Transversal designs

A **transversal design** $\mathrm{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = n = s\ell$,
- *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy

$$X = \coprod_{j=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$



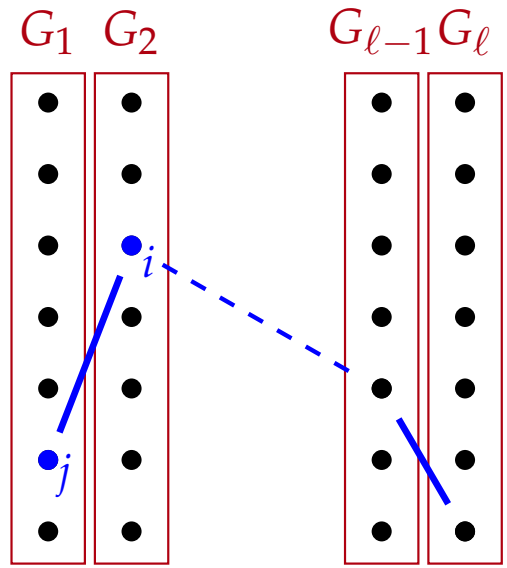
Transversal designs

A **transversal design** $\mathrm{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = n = s\ell$,
- *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy

$$X = \coprod_{j=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$

- *blocks* $B \in \mathcal{B}$ satisfy
  - $B \subset X$ and $|B| = \ell$;
  - for all $\{i, j\} \subset X$, $\{i, j\}$ lie:
    either in the same group $G \in \mathcal{G}$,
    or in a unique block $B \in \mathcal{B}$



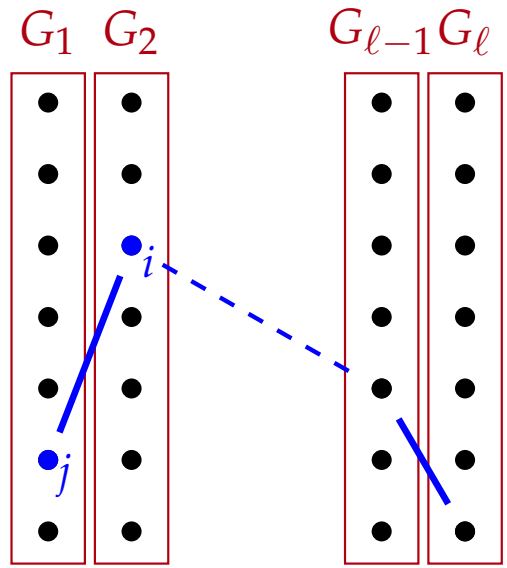
Transversal designs

A **transversal design** $\mathrm{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- X a set of *points*, $|X| = n = s\ell$,
- *groups* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfy

$$X = \coprod_{j=1}^{\ell} G_j \text{ and } |G_j| = s\,,$$

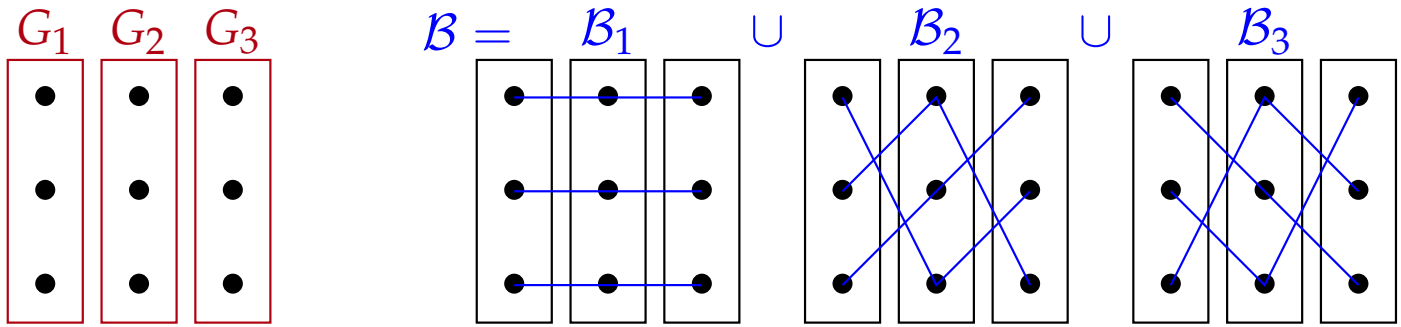
- *blocks* $B \in \mathcal{B}$ satisfy
  - $B \subset X$ and $|B| = \ell$;
  - for all $\{i, j\} \subset X$, $\{i, j\}$ lie:
    either in the same group $G \in \mathcal{G}$,
    or in a unique block $B \in \mathcal{B}$



Its incidence matrix (between points and blocks) defines a code.

Example

The transversal design TD$(3,3)$ represented by:



gives an incidence matrix

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

Its rank over $\mathbb{F}_3$ is 6 $\quad\Longrightarrow\quad$ the associated code $\mathcal{C}$ is a $[9,3]_3$ code.

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

The scheme

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a code based on a TD(ℓ, s).

The scheme

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a code based on a TD(ℓ, s).

- **Initialisation.** User U encodes $F \mapsto \boldsymbol{c} \in \mathcal{C}$, and gives $\boldsymbol{c}_{|G_j}$ to server S_j.
- **To recover** $F_i = c_i$**:**
  1. User U randomly picks a block $B \in \mathcal{B}$ containing i. Then U defines:

$$q_j = \mathcal{Q}(i)_j := \begin{cases} \text{unique } \in B \cap G_j & \text{if } i \notin G_j \\ \text{a random point in } G_j & \text{otherwise.} \end{cases}$$

  2. each server S_j sends back $a_j = \mathcal{A}_j(q_j, \boldsymbol{c}_{|G_j}) := c_{q_j}$
  3. U recovers

$$-\sum_{j:\, i \notin G_j} c_{q_j} = -\sum_{b \in B \setminus \{i\}} c_b = c_i$$

Privacy and parameters

Theorem. If the servers do not collude, then our PIR protocol is information-theoretically private.

Privacy and parameters

Theorem. If the servers do not collude, then our PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Privacy and parameters

Theorem. If the servers do not collude, then our PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Properties. For $|F| = k \log q$ bits, with $k = \dim \mathcal{C} \leq n = s\ell$.

- communication complexity: $\ell(\log s + \log q)$ bits
- computational complexity:
  - $\mathcal{O}(1)$ **for the response algorithm** $\mathcal{A}$ (somewhat optimal)
  - $\mathcal{O}(\ell)$ $\mathbb{F}_q$-operations for $\mathcal{R}$
- storage overhead: $(n-k)\log q$ bits

Privacy and parameters

Theorem. If the servers do not collude, then our PIR protocol is information-theoretically private.

Proof:

- the only server which holds F_i received a random query;
- for each other server S_j, q_j gives no information on the block B which has been picked $\Rightarrow$ no information leaks on i.

Properties. For $|F| = k \log q$ bits, with $k = \dim \mathcal{C} \leq n = s\ell$.

- communication complexity: $\ell(\log s + \log q)$ bits
- computational complexity:
  - $\mathcal{O}(1)$ **for the response algorithm** $\mathcal{A}$ (somewhat optimal)
  - $\mathcal{O}(\ell)$ $\mathbb{F}_q$-operations for $\mathcal{R}$
- storage overhead: $(n-k)\log q$ bits

Question: Transversal designs with good k depending on (ℓ, s)?

Outline

1. Codes with locality
   - Locality in coding theory, examples
   - Lifted projective Reed-Solomon codes
   - A combinatorial point of view

2. Private information retrieval from transversal designs
   - Private information retrieval (PIR)
   - Transversal designs and codes
   - A new PIR construction
   - Instances

Instances with geometric designs

$\mathcal{T}_A$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

Instances with geometric designs

$\mathcal{T}_A$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

Proposition. The code based on $\mathcal{T}_A$ and the code based on $\mathrm{AG}_1(m,q)$ have same length and same dimension.

Instances with geometric designs

$\mathcal{T}_A$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

Proposition. The code based on $\mathcal{T}_A$ and the code based on $\mathrm{AG}_1(m,q)$ have same length and same dimension.

"Practical" instances:

- 3.2% storage overhead if #entries $\leq$ (#servers)^2
- 27% storage overhead if #entries $\leq$ (#servers)^3

Instances with geometric designs

$\mathcal{T}_A$, the **classical affine transversal design**:

- $X = \mathbb{F}_q^m$, $m \geq 2$,
- $\mathcal{G}$ a set of q disjoint hyperplanes partitionning X,
- $\mathcal{B} = \{\text{affine lines } L \text{ secant to each group of } \mathcal{G}\}$.

Proposition. The code based on $\mathcal{T}_A$ and the code based on $\mathrm{AG}_1(m,q)$ have same length and same dimension.

"Practical" instances:

- 3.2% storage overhead if #entries $\leq$ (#servers)^2
- 27% storage overhead if #entries $\leq$ (#servers)^3

Question: are they the best instances?

Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

$$S = \{a, b\}$$

$$\mathrm{OA}(2, 3, 2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

$$S = \{a, b\}$$

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq \ell\}$

$$\mathrm{OA}(2, 3, 2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

$$\begin{matrix} (a,1) \\ (b,1) \end{matrix} \qquad \begin{matrix} (a,2) \\ (b,2) \end{matrix} \qquad \begin{matrix} (a,3) \\ (b,3) \end{matrix}$$

Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

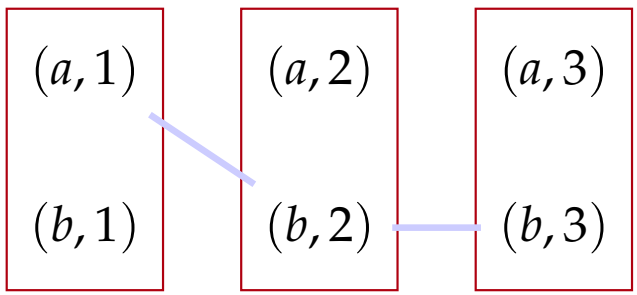
Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq \ell\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \mathrm{OA}\}$

$$S = \{a, b\}$$

$$\mathrm{OA}(2, 3, 2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$



Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

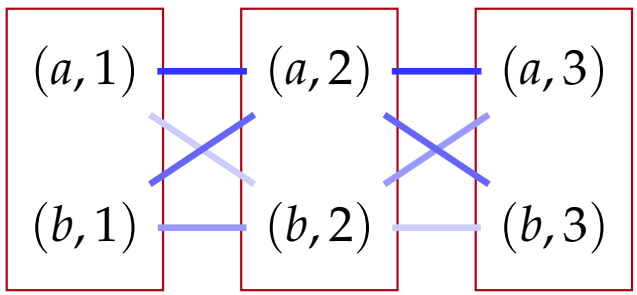
Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

$$S = \{a, b\}$$

$$\mathrm{OA}(2, 3, 2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$

Construction OA $\rightarrow$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq \ell\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \mathrm{OA}\}$



Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t may be seen as a list A of codewords

- over a finite set S, $|S| = s$,
- of length ℓ,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, the code $A \subset S^\ell$ has dual distance $t + 1$.

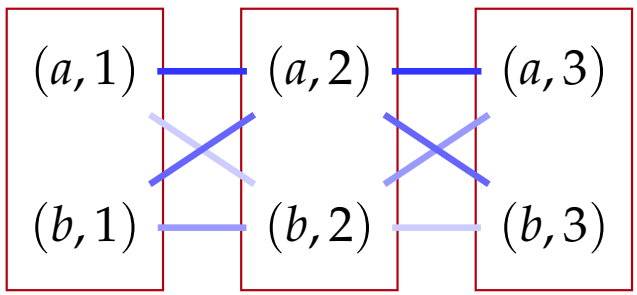
Construction OA $\to$ TD :

- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq \ell\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \mathrm{OA}\}$

Prop. An $\mathrm{OA}(2, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

$S = \{a, b\}$

$$\mathrm{OA}(2,3,2) = \begin{bmatrix} a & b & b \\ b & b & a \\ b & a & b \\ a & a & a \end{bmatrix}$$



Resisting collusions

Experimentally, for $t = 2$ and small ℓ and s, codes based on classical affine TDs have the largest dimension.

Resisting collusions

Experimentally, for $t = 2$ and small ℓ and s, codes based on classical affine TDs have the largest dimension.

Question: what about TDs from $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

We get TDs such that:

for every t-set of points lying in t different groups,
there exists a unique block which contains it.

Resisting collusions

Experimentally, for $t = 2$ and small ℓ and s, codes based on classical affine TDs have the largest dimension.

Question: what about TDs from $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

We get TDs such that:

for every t-set of points lying in t different groups,
there exists a unique block which contains it.

$\Rightarrow$ The PIR protocol resists $t - 1$ colluding servers.

Resisting collusions

Experimentally, for $t = 2$ and small ℓ and s, codes based on classical affine TDs have the largest dimension.

Question: what about TDs from $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

We get TDs such that:

for every t-set of points lying in t different groups,
there exists a unique block which contains it.

$\Rightarrow$ The PIR protocol resists $t - 1$ colluding servers.

- OA with $t > 2$ exist (from Reed-Solomon codes)...
- ... but underlying codes have poor rates except for $t \ll \ell$.

Resisting collusions

Experimentally, for $t = 2$ and small ℓ and s, codes based on classical affine TDs have the largest dimension.

Question: what about TDs from $\mathrm{OA}(t, \ell, s)$ with $t > 2$?

We get TDs such that:

for every t-set of points lying in t different groups,
there exists a unique block which contains it.

$\Rightarrow$ The PIR protocol resists $t - 1$ colluding servers.

- OA with $t > 2$ exist (from Reed-Solomon codes)...
- ... but underlying codes have poor rates except for $t \ll \ell$.

Details in:

Private Information Retrieval from Transversal Designs, L., IEEE TIT, to appear

Conclusion

- Codes with local properties gained interest
  - theoretically: PCP theorem, etc.
  - in practice: storage of large files on distributed storage systems or p2p networks
  - more recently STARKs, etc.
- A combinatorial point of view (through designs) could help their analysis
- Cryptographic applications: private information retrieval (PIR), proofs of retrievability (PoR)