Codes with locality: constructions and applications to cryptographic protocols

Julien Lavauzelle

École Polytechnique & INRIA Saclay, Université Paris-Saclay

PhD defense

30/11/2018

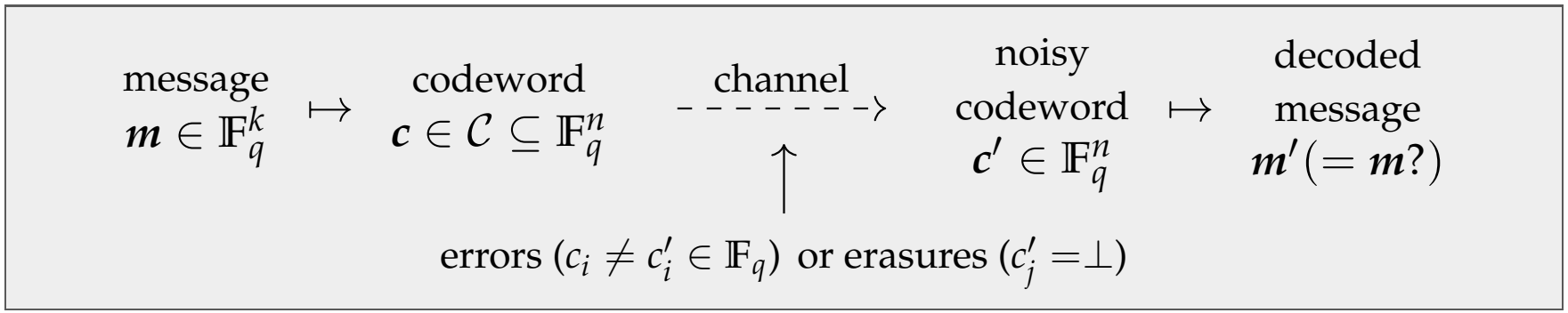
Outline

Outline

Outline

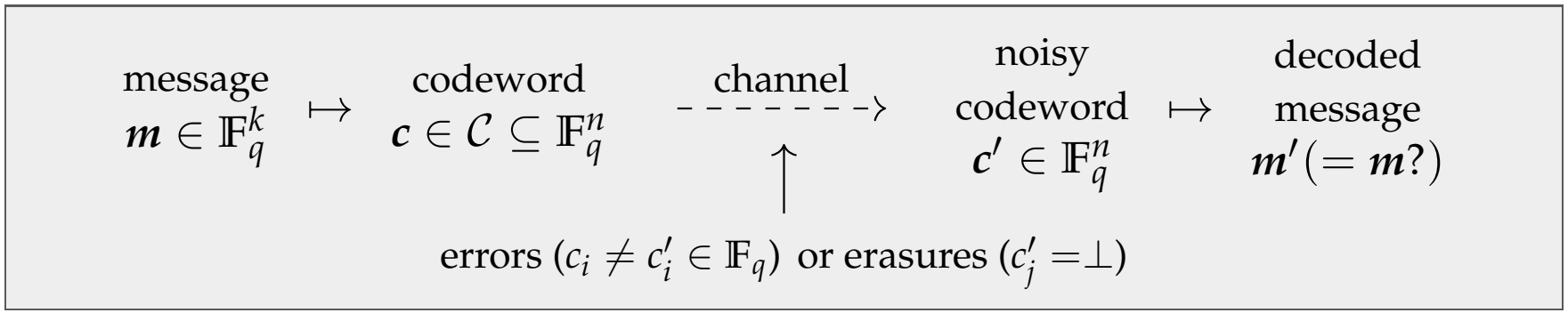
Error-correcting codes

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.

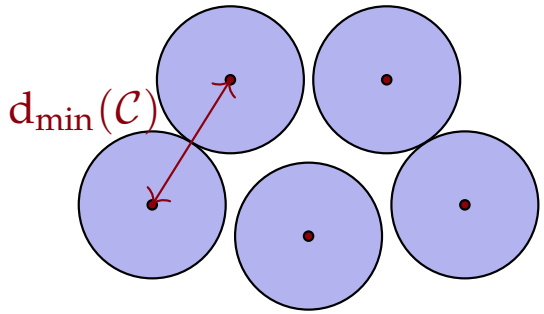
$$\underset{\boldsymbol{m} \in \mathbb{F}_q^k}{\text{message}} \mapsto \underset{\boldsymbol{c} \in \mathcal{C} \subseteq \mathbb{F}_q^n}{\text{codeword}} \xrightarrow{\text{channel}} \underset{\boldsymbol{c}' \in \mathbb{F}_q^n}{\text{noisy codeword}} \mapsto \underset{\boldsymbol{m}'(=\boldsymbol{m}?)}{\text{decoded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

$$\uparrow$$

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

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

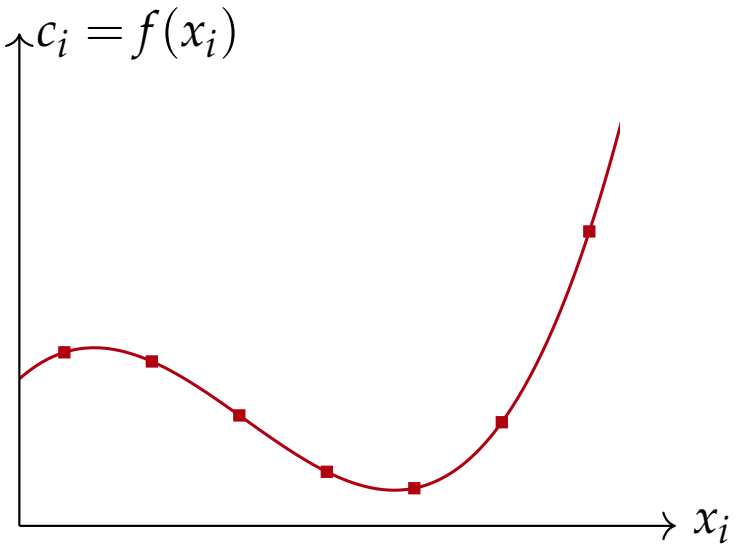
- $d = \mathrm{d}_{\min}(\mathcal{C}) := \min\{\mathrm{d}(\boldsymbol{c}, \boldsymbol{c}'), \boldsymbol{c} \neq \boldsymbol{c}', (\boldsymbol{c}, \boldsymbol{c}') \in \mathcal{C}^2\}$,
- $\mathcal{C}$ linear over $\mathbb{F}_q$, with $k = \dim(\mathcal{C})$.



Typical example: Reed-Solomon codes

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

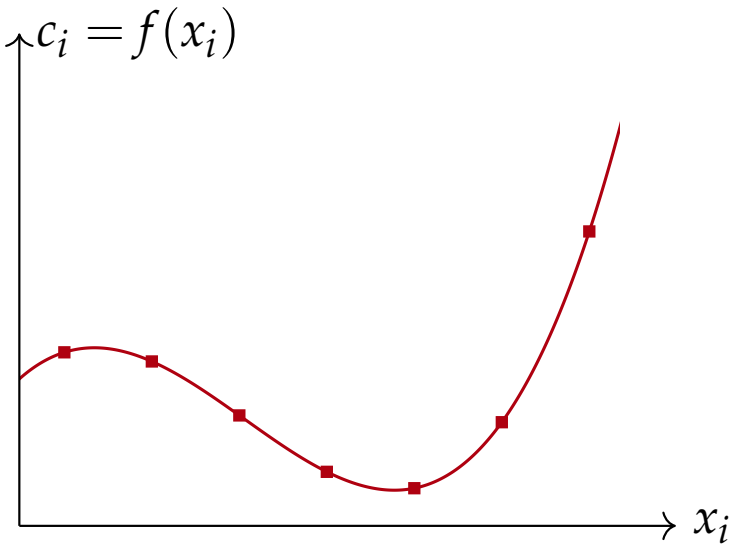
$$\mathrm{RS}_q(r, \boldsymbol{x}) := \{(f(x_1), \ldots, f(x_n)), f \in \mathbb{F}_q[X], \deg f \leq r\}$$



Typical example: Reed-Solomon codes

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

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

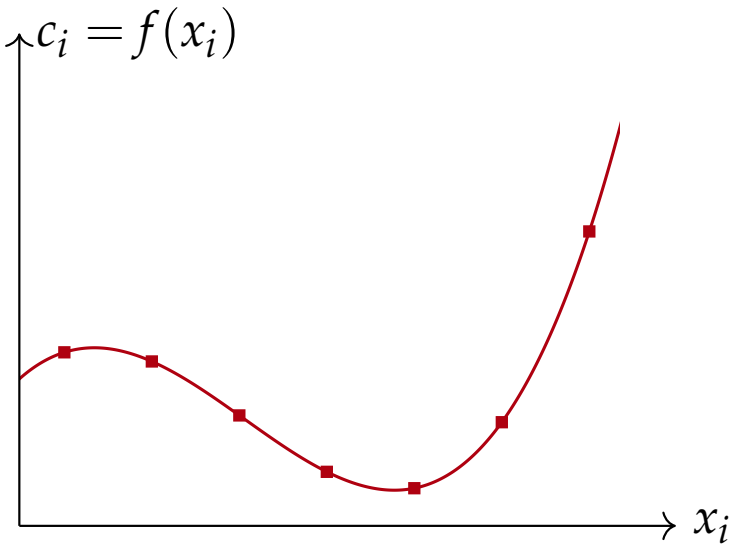


- Dimension $k = r + 1$
- Minimum distance $\mathrm{d}_{\min} = n - r$
- Can decode any b errors and e erasures
  $\rightarrow$ if $e + 2b < \mathrm{d}_{\min}$
  $\rightarrow$ in time $\Theta(n \log^3 n)$.

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 $\mathrm{d}_{\min} = n - r$
- 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 [Katz, Trevisan '00]

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

Local correction [Katz, Trevisan '00]

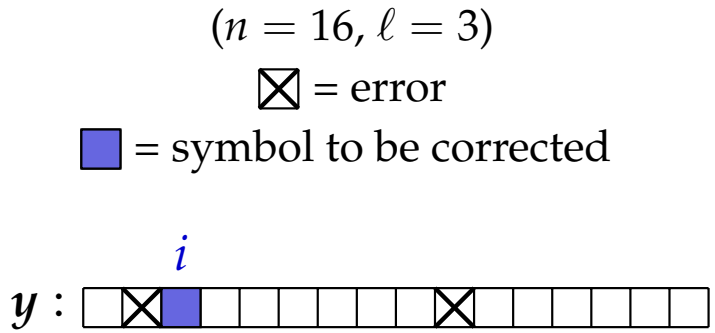
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition [KT00]. 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 poly-time **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 $1 \leq i \leq n$:

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



Local correction [Katz, Trevisan '00]

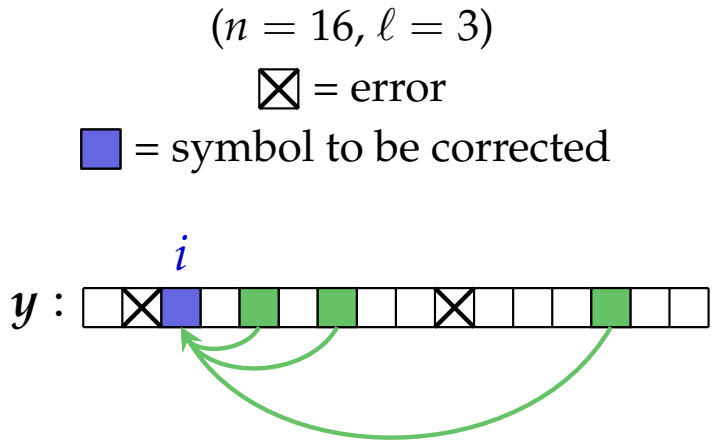
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition [KT00]. 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 poly-time **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 $1 \leq i \leq n$:

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



Local correction [Katz, Trevisan '00]

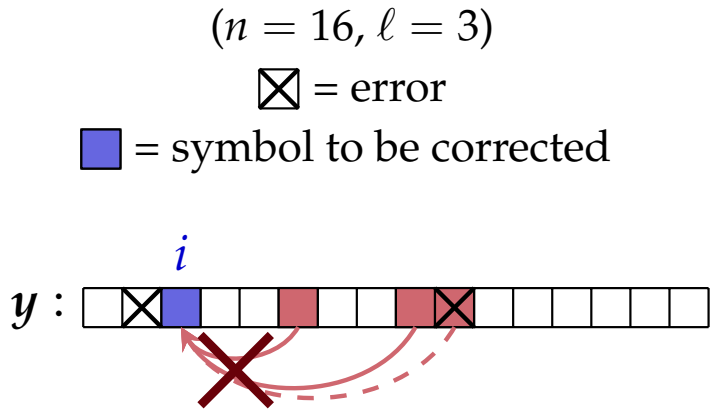
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition [KT00]. 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 poly-time **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 $1 \leq i \leq n$:

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



Local correction [Katz, Trevisan '00]

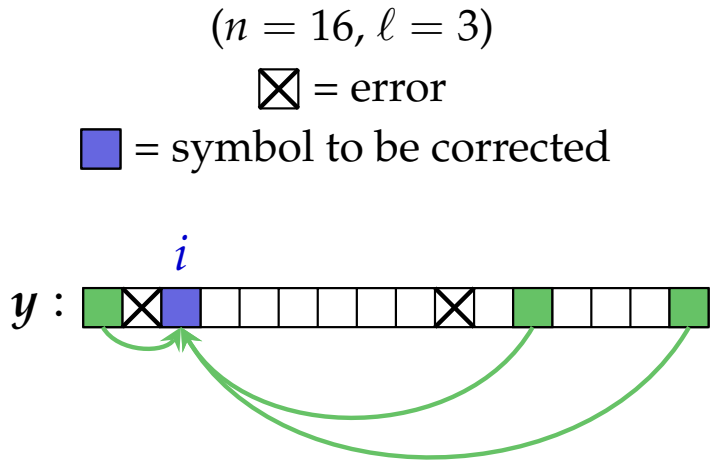
Goal: sublinear-time correction of some symbols of $c \in \mathcal{C}$.

Definition [KT00]. 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 poly-time **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 $1 \leq i \leq n$:

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



LCCs: goals and previous works

Goals:

- failure probability $\varepsilon \leq f(\ell) \cdot \delta$, with $f(\ell) \leq$ cste.
- locality $\ell \ll k$
- large dimension k

LCCs: goals and previous works

Goals:

- failure probability $\varepsilon \leq f(\ell) \cdot \delta$, with $f(\ell) \leq$ cste.
- locality $\ell \ll k$
- large dimension k

Some existing constructions:

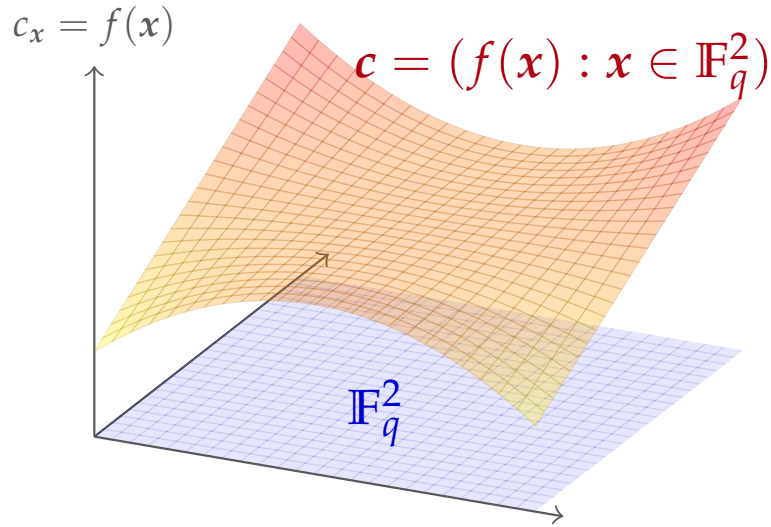
- constant locality ℓ:
  - Hadamard code (folklore) $\ell = 2$ and $k = \log(n)$
  - Matching vector codes [Yek08] k subexponential in $\log(n)$

- constant rate $R = k/n$:
  - Reed-Muller codes (folklore) $\ell = n^{1/m}$ and $k \leq \frac{1}{m!} \cdot n$
  - Multiplicity codes [KSY14], **lifted codes** [GKS13], expander codes [HOW14]

$$\ell \leq n^{\varepsilon} \quad \text{and} \quad k \geq \alpha \cdot n, \quad \forall \varepsilon, \alpha > 0,\ n \to \infty$$

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

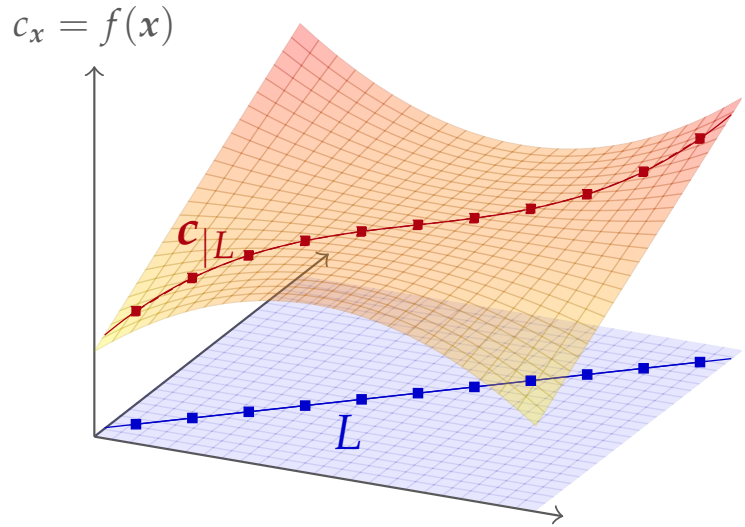
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

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
  $\Rightarrow$ affine line $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$

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

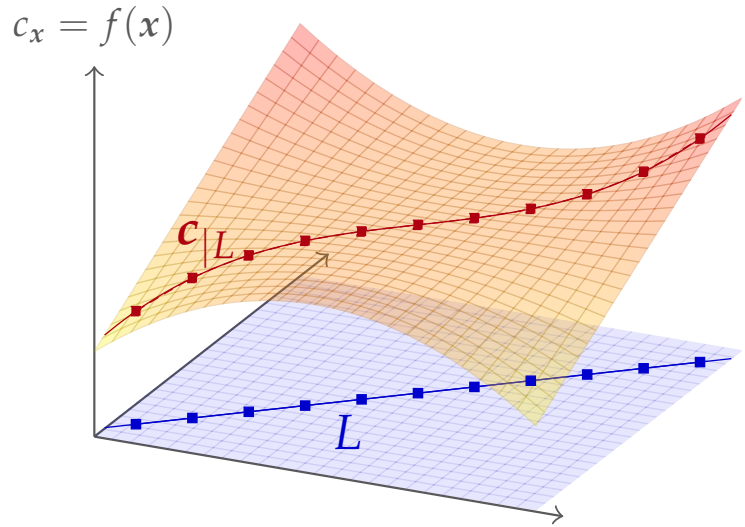
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

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
  $\Rightarrow$ affine line $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$

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

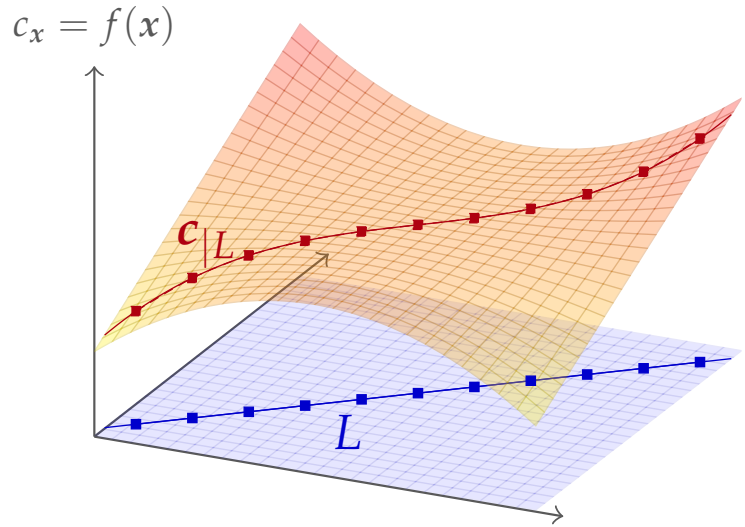
$$\mathrm{RM}_q(m,r) := \{(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1,\ldots,X_m], \deg f \leq r\}$$

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
  $\Rightarrow$ affine line $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$

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 $\varepsilon = \frac{2}{1-r/q} \cdot \delta$

High-rate construction: lifted codes (1)

Issue: if $r \leq q-2$, the rate of $\mathrm{RM}_q(m,r)$ is $\simeq \frac{(r/q)^m}{m!}$.

High-rate construction: lifted codes (1)

Issue: if $r \leq q-2$, the rate of $\mathrm{RM}_q(m,r)$ is $\simeq \frac{(r/q)^m}{m!}$.

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

for every affine line L given by ϕ, $\quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$

Are there more polynomials than in RM codes?

High-rate construction: lifted codes (1)

Issue: if $r \leq q-2$, the rate of $\mathrm{RM}_q(m,r)$ is $\simeq \frac{(r/q)^m}{m!}$.

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

for every affine line L given by ϕ, $\quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$

Are there more polynomials than in RM codes?

Example ($q = 4, m = 2, r = 2$).

$f(X,Y) = X^2Y^2 \in \mathbb{F}_4[X,Y]$, hence $\deg(f) = 4 > 2$

Affine line L given by $\phi(T) = (aT + b, cT + d)$

High-rate construction: lifted codes (1)

Issue: if $r \leq q-2$, the rate of $\mathrm{RM}_q(m,r)$ is $\simeq \frac{(r/q)^m}{m!}$.

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

for every affine line L given by ϕ, $\quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$

Are there more polynomials than in RM codes?

Example ($q = 4, m = 2, r = 2$).

$f(X,Y) = X^2Y^2 \in \mathbb{F}_4[X,Y]$, hence $\deg(f) = 4 > 2$

Affine line L given by $\phi(T) = (aT + b, cT + d)$

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

High-rate construction: lifted codes (1)

Issue: if $r \leq q-2$, the rate of $\mathrm{RM}_q(m,r)$ is $\simeq \frac{(r/q)^m}{m!}$.

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

$$\text{for every affine line } L \text{ given by } \phi, \quad ((f\circ\phi)(t) : t \in \mathbb{F}_q) \in \mathrm{RS}_q(r)$$

Are there more polynomials than in RM codes?

Example ($q=4, m=2, r=2$).

$f(X,Y) = X^2Y^2 \in \mathbb{F}_4[X,Y]$, hence $\deg(f) = 4 > 2$

Affine line L given by $\phi(T) = (aT+b, cT+d)$

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

$\Rightarrow$ 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)

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

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

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

High-rate construction: lifted codes (2)

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

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

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

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

High-rate construction: lifted codes (2)

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

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

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

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

What about the dimension/rate?

Theorem (characteristic 2, simplified from [GKS13]).
For every $m \geq 2$ and $0 < R_0 < 1$, there exists $q > 0$ and $r \leq q - 2$ such that

$$\mathrm{Lift}(\mathrm{RS}_q(r), m)$$

is locally correctable with rate $R \geq R_0$.

Rate of lifted codes

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

- **Ex:** for $m = 2$ and $R_0 = 1/2$, GKS theorem requires $n = q^m \geq 2^{64}$.

Rate of lifted codes

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

- **Ex:** for $m = 2$ and $R_0 = 1/2$, GKS theorem requires $n = q^m \geq 2^{64}$.

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, $n = q^2 \geq 2^6 = 64$ is enough to achieve $R \geq 1/2$.

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 [GKS13]:

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

Degree sets

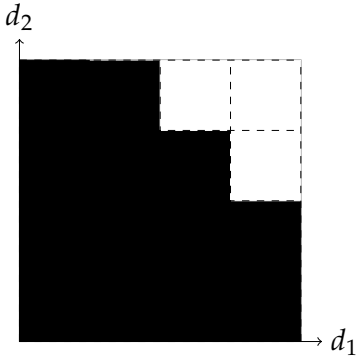
Lifted codes are **monomial**, *i.e.* generated by evaluations of monomials

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

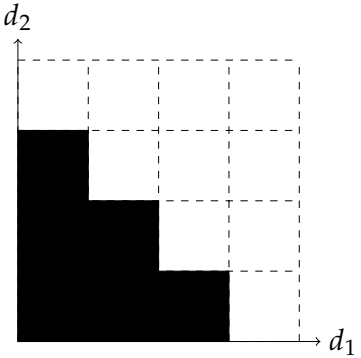
Degree set of a monomial code [GKS13]:

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

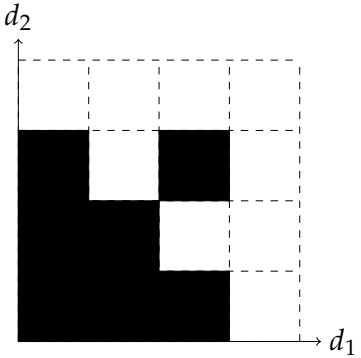
A **representation** for $m = 2$:



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

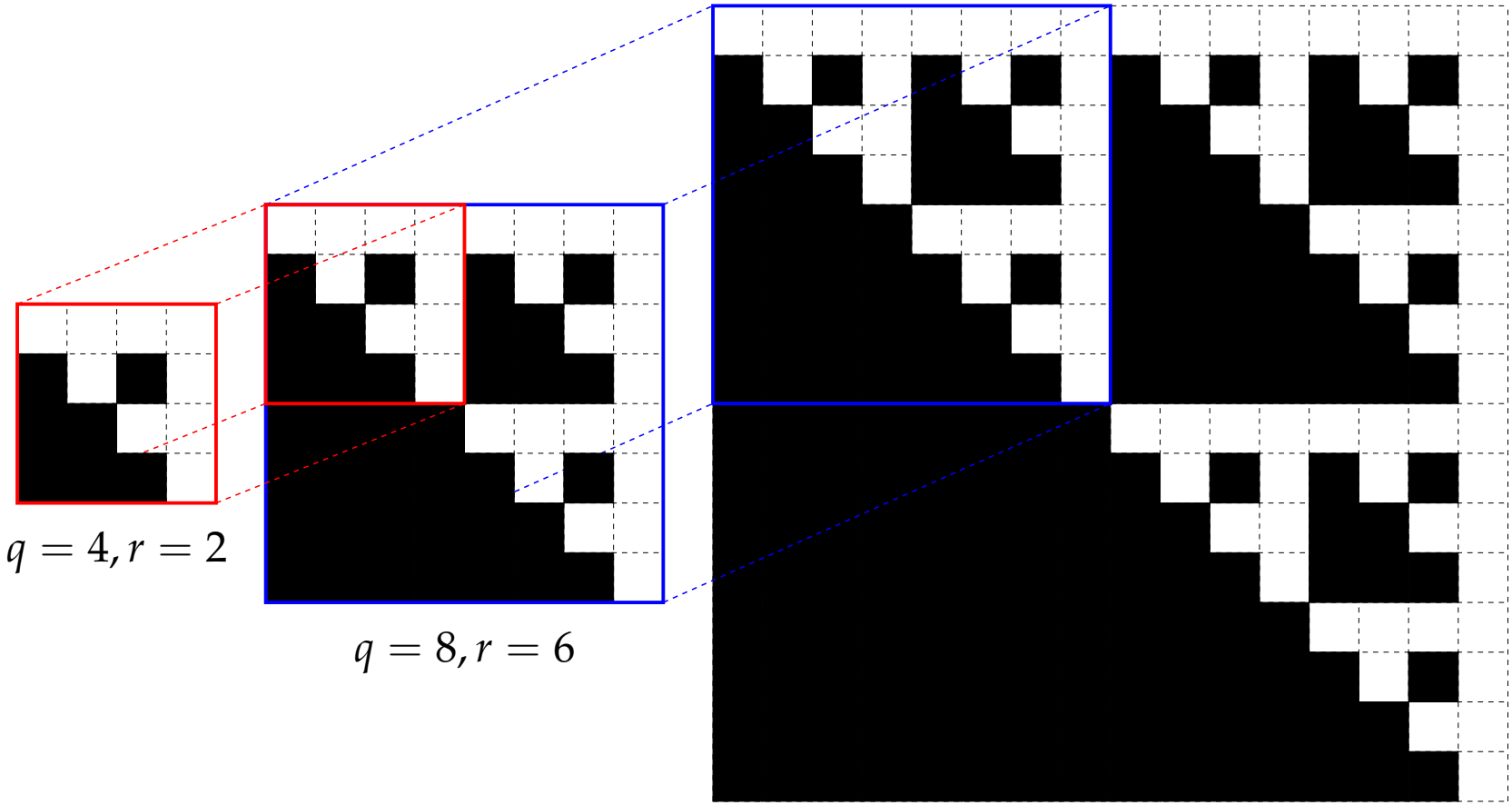


$\mathrm{RM}_4(2,2)$



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

“Fractal” representation of degree sets



Outline

Evaluation on projective spaces

Projective space $\mathbb{P}^m := \left(\mathbb{A}^{m+1} \setminus \{\mathbf{0}\}\right) / \sim$ 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 $\boldsymbol{u} \in \mathbb{P}^m$ (see [Lac86]):

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

We get:

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

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

Projective lifted codes

Example. Projective Reed-Solomon code:

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

Projective lifted codes

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 $\mathrm{Emb}_{\mathbb{P}}(m) := \{\phi : \mathbb{F}_q^2 \to \mathbb{F}_q^{m+1} \text{ linear and injective}\}$.

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

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

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, where τ is the relative correction capability of the small PRS code

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, 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 $\mathrm{Lift}(\mathrm{PRS}_q(r), m-1)$

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, 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 π is induced by $\mathbb{P}^m \to \mathbb{P}^{m-1}$.

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, 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 π is induced by $\mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence)

cyclic codes	if $q-1$ and $n = \frac{q^{m+1}}{q-1}$ are coprime
quasi-cyclic codes	if $q-1$ and $\frac{n}{\gcd(n,q-1)}$ are coprime

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, 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 π is induced by $\mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence)

cyclic codes	if $q-1$ and $n = \frac{q^{m+1}}{q-1}$ are coprime
quasi-cyclic codes	if $q-1$ and $\frac{n}{\gcd(n,q-1)}$ are coprime

- admit many explicit and easily computable **information sets**

Main results on projective lifted codes

Projective lifted codes...

- are **locally correctable**, with parameters $(\ell = q+1, \delta, \varepsilon = \delta/\tau)$, 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 π is induced by $\mathbb{P}^m \to \mathbb{P}^{m-1}$.
- are (up to equivalence)

cyclic codes	if $q-1$ and $n = \frac{q^{m+1}}{q-1}$ are coprime
quasi-cyclic codes	if $q-1$ and $\frac{n}{\gcd(n,q-1)}$ are coprime

- 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

Lifted codes when $r = q - 2$

Remark. Assume $r = q - 2$. Then, $\mathrm{RS}_q(q-2)$ is the parity-check code.

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

$$\boldsymbol{c} \in \mathrm{Lift}(\mathrm{RS}_q(q-2), m) \iff \forall L \subseteq \mathbb{F}_q^m, \sum_{x \in L} c_x = 0$$

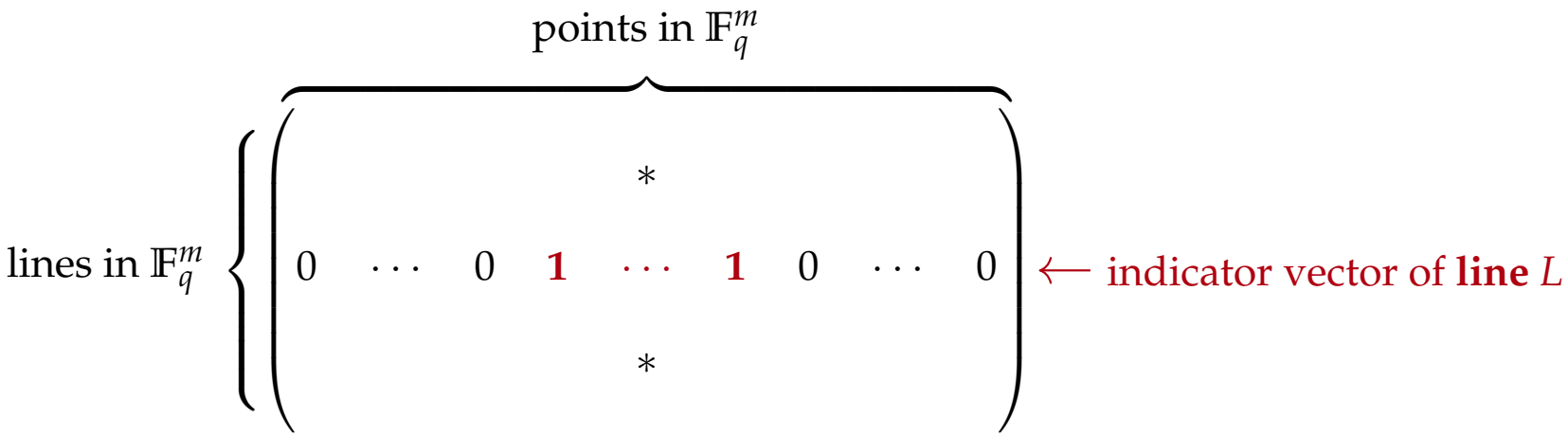
Lifted codes when $r = q - 2$

Remark. Assume $r = q - 2$. Then, $\mathrm{RS}_q(q-2)$ is the parity-check code.

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

$$\boldsymbol{c} \in \mathrm{Lift}(\mathrm{RS}_q(q-2), m) \iff \forall L \subseteq \mathbb{F}_q^m, \sum_{x \in L} c_x = 0$$

A non-full-rank **parity-check matrix** for $\mathrm{Lift}(\mathrm{RS}_q(q-2), m)$:



Block designs

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

Block designs

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

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

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

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

Block designs

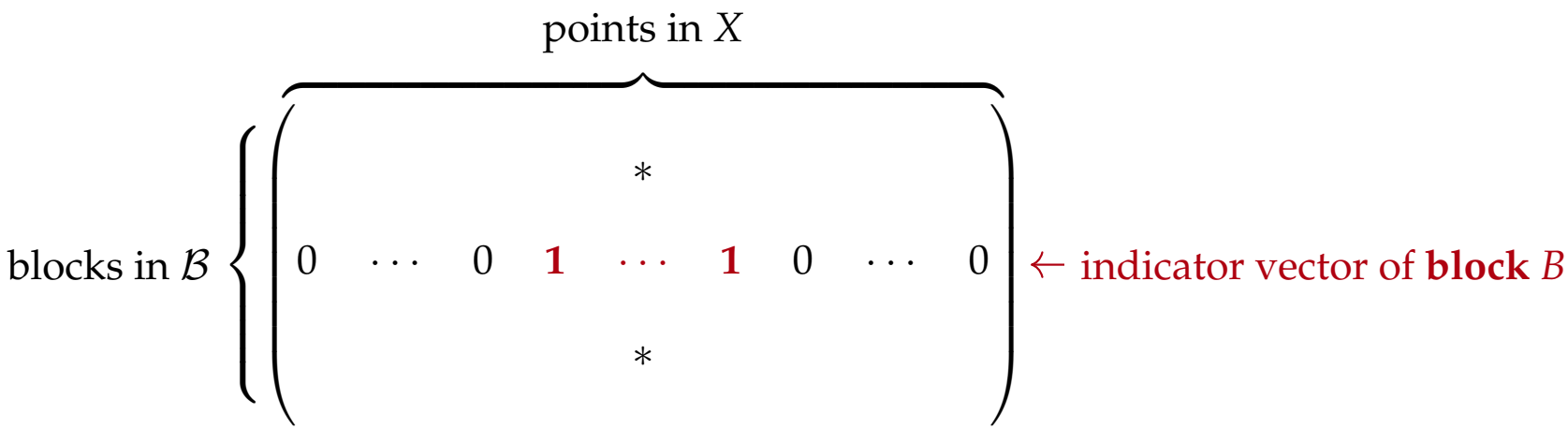
Point-line incidences in the affine space form the **affine geometry 2-design**.

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

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

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.

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

Remark. The dimension of $\mathrm{Code}(\mathcal{D})$ is highly dependent 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.

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

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

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

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

Design-based codes and LCCs

Generalised design-based code $\mathcal{C} = \mathrm{Code}(\mathcal{D}, \mathcal{F})$, where

- $\mathcal{D}$ be a t-$(n, \ell+1, \lambda)$-design
- $\tau \in (0, \frac{1}{2})$ is fixed
- $\mathcal{F} = (\mathcal{F}_B : B \in \mathcal{B})$ s.t. every code in $\mathcal{F}$ corrects a fraction τ of errors

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 $i \in B$.
- Correct $y_{|B}$ as a noisy codeword from $\mathcal{F}_B$, and output $\tilde{y}_i$.

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

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

Proposition [$t = 3$]. For every $\delta < \tau - 1/\sqrt{2\ell}$, $\mathrm{Code}(\mathcal{D}, \mathcal{F})$ is a $(\ell, \delta, \varepsilon)$-LCC where

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

Outline

Outline

Problem statement

Given a remote database $F \in \mathbb{F}_q^k$ and $1 \leq i \leq k$,
can we **retrieve** the entry F_i,
without leaking information on the index i?

Problem statement

Given a remote database $F \in \mathbb{F}_q^k$ and $1 \leq i \leq k$,
can we **retrieve** the entry F_i,
without leaking information on the index i?

Trivial solution: full download.

Problem statement

Given a remote database $F \in \mathbb{F}_q^k$ and $1 \leq i \leq k$,
can we **retrieve** the entry F_i,
without leaking information on the index i?

Trivial solution: full download.

Solutions with **better communication complexity:**

- With 1 server, only **computational** privacy is possible [CGKS95, CG97].
- With $\ell \geq 2$ servers, one can achieve **information-theoretic** privacy [CGKS95-98].

Definition of PIR [CGKS95]

Given a file F and ℓ servers $S_1, \ldots, S_\ell$,
user U wants to recover F_i privately.

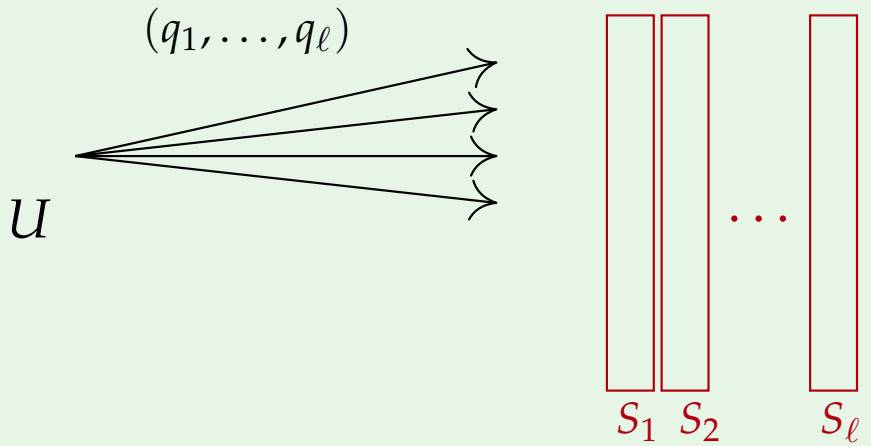
A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

Definition of PIR [CGKS95]

Given a file F and ℓ servers $S_1, \ldots, S_\ell$,
user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

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

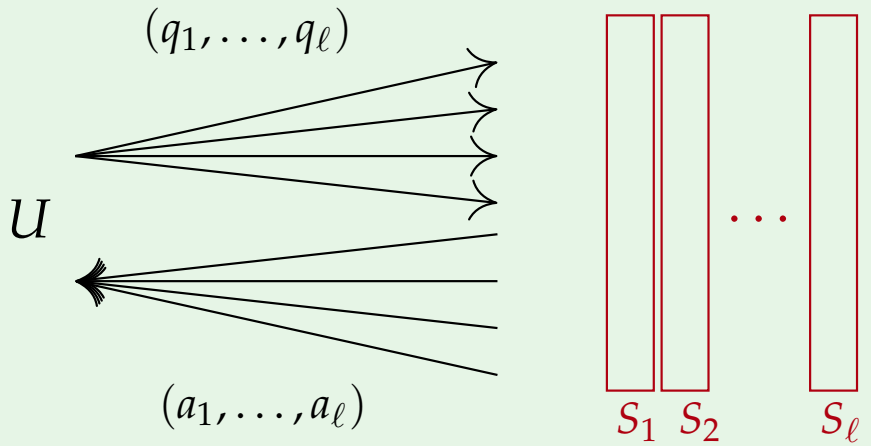


Definition of PIR [CGKS95]

Given a file F and ℓ servers $S_1, \ldots, S_\ell$,
user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

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

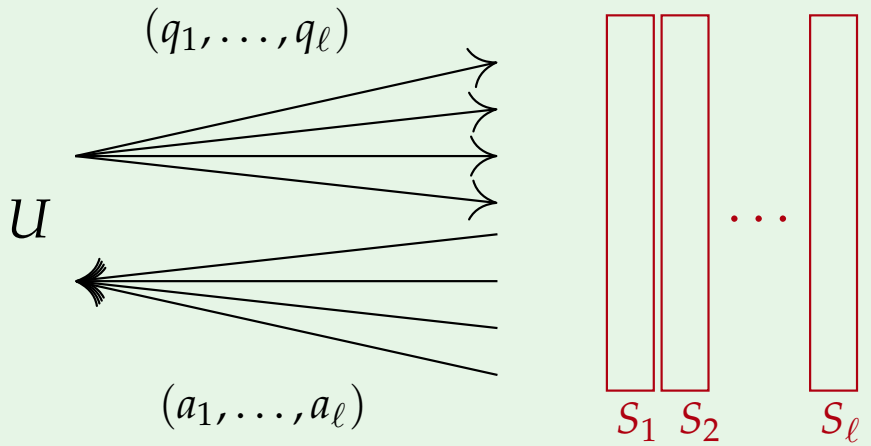


Definition of PIR [CGKS95]

Given a file F and ℓ servers $S_1, \ldots, S_\ell$,
user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

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

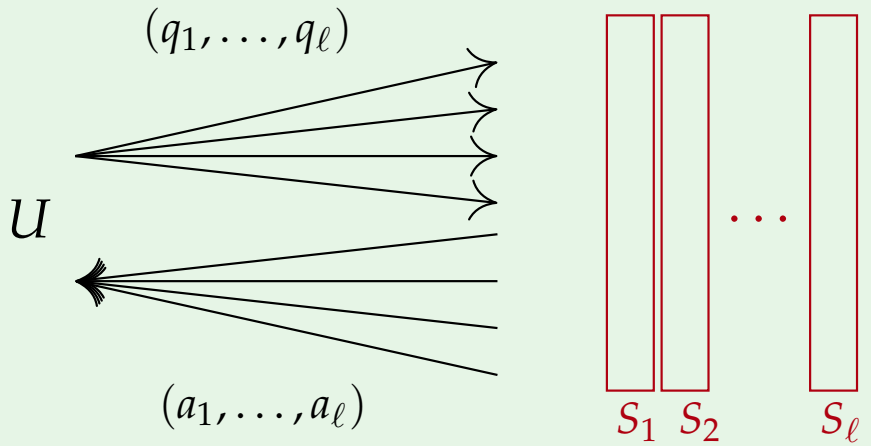


Definition of PIR [CGKS95]

Given a file F and ℓ servers $S_1, \ldots, S_\ell$,
user U wants to recover F_i privately.

A **Private Information Retrieval protocol** is a set of algorithms $(\mathcal{Q}, \mathcal{A}, \mathcal{R})$:

1. U generates a query vector $\boldsymbol{q} = (q_1, \ldots, q_\ell) \leftarrow \mathcal{Q}(i)$ and sends q_j to server S_j
2. Each server S_j computes $a_j = \mathcal{A}(q_j, F_{|S_j})$ 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
- Low computation complexity for algorithms $\mathcal{A}$ (server) and $\mathcal{R}$ (user)

Motivation

Usual goals for PIR:

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

Most constructions focus on the **download communication complexity**

– up to the **PIR capacity** [SJ17]

– but require $\Omega(k)$ computation complexity for each server

Motivation

Usual goals for PIR:

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

Most constructions focus on the **download communication complexity**

– up to the **PIR capacity** [SJ17]

– but require $\Omega(k)$ computation complexity for each server

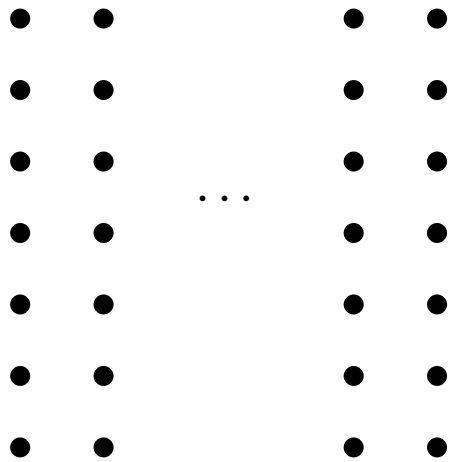
We here focus on the **computation complexity**, crucial for practicality [OG10].

Outline

Transversal designs

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

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

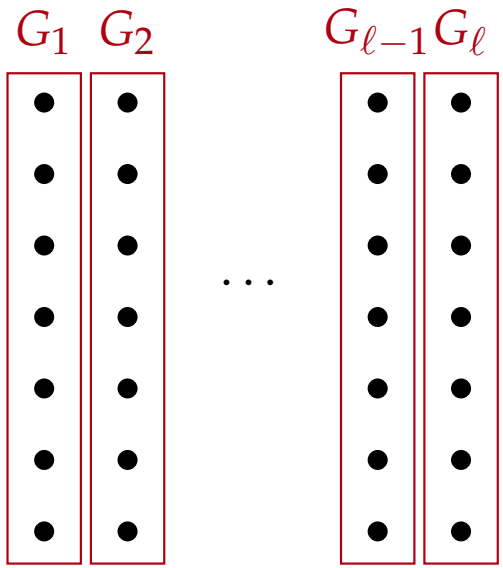


Transversal designs

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* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfying

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



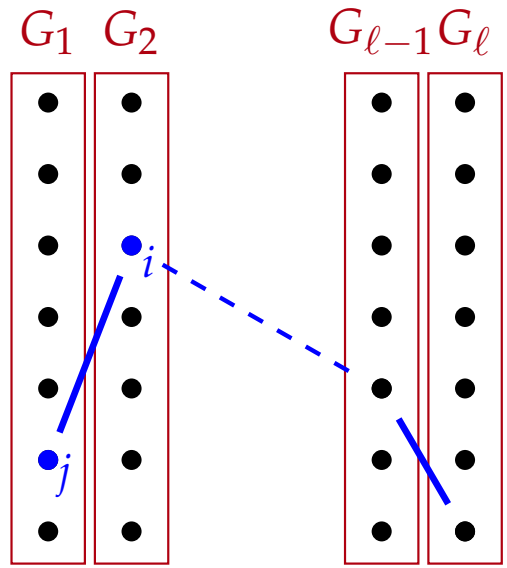
Transversal designs

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* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfying

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

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



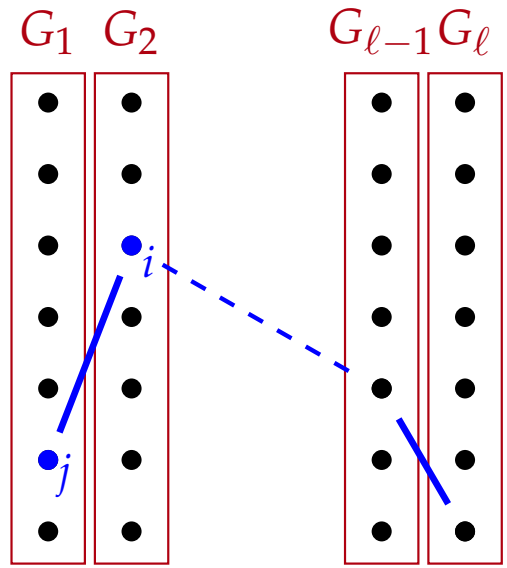
Transversal designs

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* $\mathcal{G} = \{G_j\}_{1 \le j \le \ell}$ satisfying

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

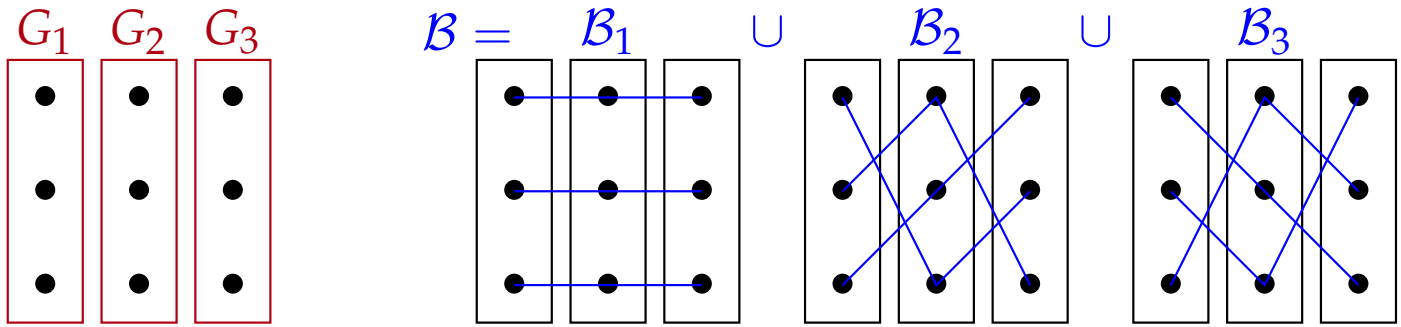
- *blocks* $B \in \mathcal{B}$ satisfying
  - $B \subset X$ and $|B| = \ell$;
  - for all $\{i, j\} \subset X$, $\{i, j\}$ lie:
    either in a single group $G \in \mathcal{G}$,
    or in a unique block $B \in \mathcal{B}$



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

Example

The transversal design $\mathrm{TD}(3,3)$ 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

The PIR scheme

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

The PIR scheme

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

- **Initialisation.** User U encodes $F \mapsto \boldsymbol{c} \in \mathcal{C}$, and gives $\boldsymbol{c}_{|G_j}$ to server S_j.

The PIR 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** $F_i = c_i$, with $i \in X$:
  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 c_{q_j}
  3. U recovers

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

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

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

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

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

Features.

- communication complexity: $\ell \log s$ uploaded bits, $\ell \log q$ dowloaded bits
- computational complexity:
  - **only** 1 **read for each server** (somewhat optimal)
  - $\leq \ell$ additions over $\mathbb{F}_q$ for the user
- storage overhead: $(n-k)\log q$ bits

Privacy and parameters

Theorem. This PIR protocol is information-theoretically private.

Proof:

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

Features.

- communication complexity: $\ell \log s$ uploaded bits, $\ell \log q$ dowloaded bits
- computational complexity:
  - **only 1 read for each server** (somewhat optimal)
  - $\leq \ell$ additions over $\mathbb{F}_q$ for the user
- storage overhead: $(n-k)\log q$ bits

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

Outline

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 $\mathcal{T}_A$ is identical to the code based on the affine geometry design (*i.e.* the lifted code with $r = q - 2$).

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 $\mathcal{T}_A$ is identical to the code based on the affine geometry design (*i.e.* the lifted code with $r = q - 2$).

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 $\mathcal{T}_A$ is identical to the code based on the affine geometry design (*i.e.* the lifted code with $r = q - 2$).

Instances:

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

Question: better instances?

Instances with orthogonal arrays

An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t is a list A of words
- of length ℓ,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t, \ell, s)$ is a code $A \subset S^\ell$ with 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 is a list A of words
- of length ℓ,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t, \ell, s)$ is a code $A \subset S^\ell$ with dual distance $t + 1$.

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

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

Equivalently, an $\mathrm{OA}(t, \ell, s)$ is a code $A \subset S^\ell$ with dual distance $t + 1$.

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

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

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

$(a,1)$	$(a,2)$	$(a,3)$
$(b,1)$	$(b,2)$	$(b,3)$

Instances with orthogonal arrays

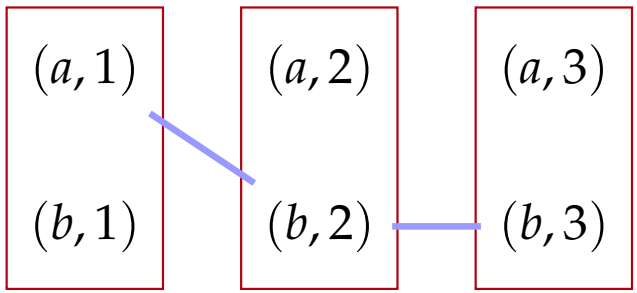
An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t is a list A of words
- of length ℓ,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t, \ell, s)$ is a code $A \subset S^\ell$ with dual distance $t + 1$.

$S = \{a, b\}$

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

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



Instances with orthogonal arrays

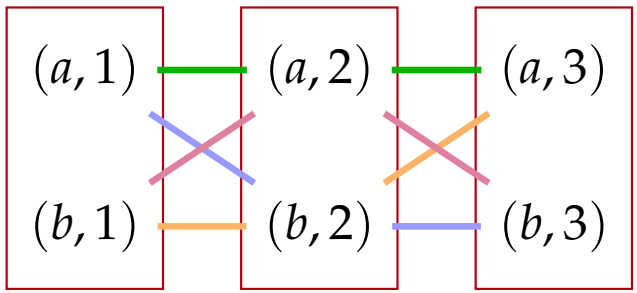
An **orthogonal array** $\mathrm{OA}(t, \ell, s)$ of strength t is a list A of words
- of length ℓ,
- over a finite set S, $|S| = s$,
- such that, for every $I \subset [1, \ell]$ of size t, $A_{|I} = S^t$.

Equivalently, an $\mathrm{OA}(t, \ell, s)$ is a code $A \subset S^\ell$ with dual distance $t + 1$.

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

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

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



Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

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

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

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

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

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

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

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

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

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

- OAs with $t > 2$ exist (*e.g.* from Reed-Solomon codes)
- But associated TDs lead to codes with poor rates except for $t \ll \ell$

Resisting collusions

Proposition. For $t = 2$, an $\mathrm{OA}(t, \ell, s)$ gives a $\mathrm{TD}(\ell, s)$.

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

For $t \geq 3$, we get TDs such that:

for every t-set T of points lying in t different groups,
there exists a unique block $B \in \mathcal{B}$ such that $T \subset B$.

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

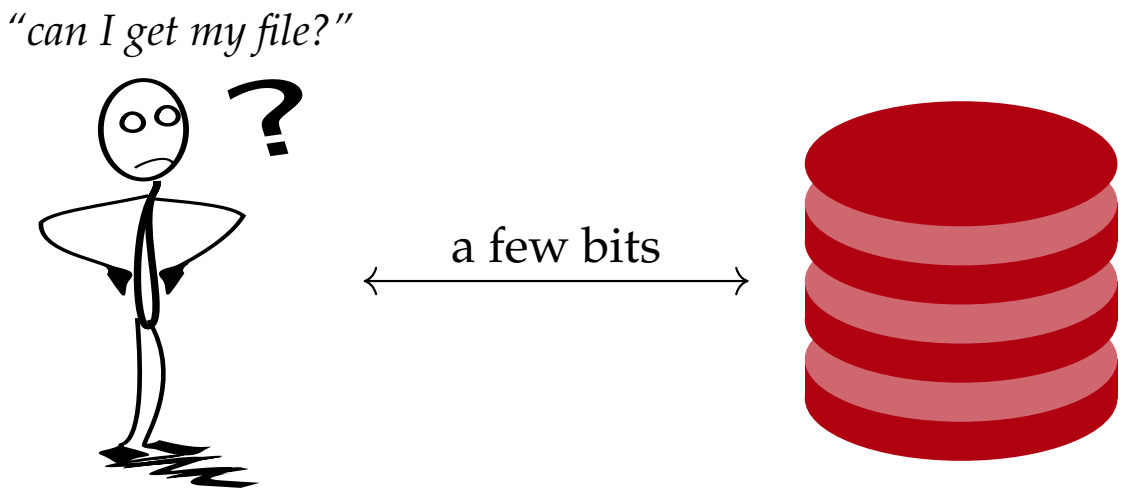
- OAs with $t > 2$ exist (*e.g.* from Reed-Solomon codes)
- But associated TDs lead to codes with poor rates except for $t \ll \ell$

Details in:
Private Information Retrieval from Transversal Designs, L., IEEE TIT, to appear
10.1109/TIT.2018.2861747

Outline

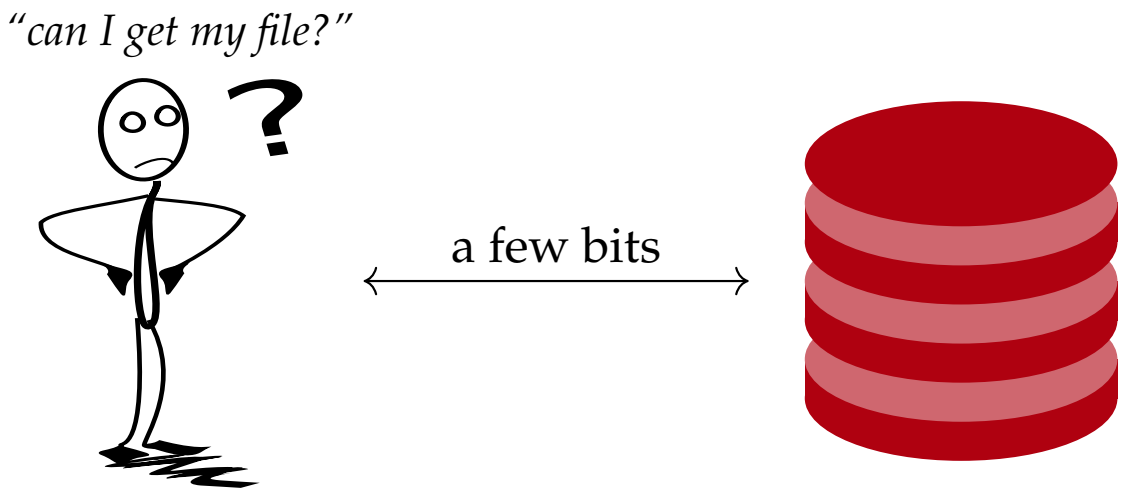
Proofs-of-retrievability [Juels, Kaliski '07]

Issue: a client wants to verify if a file stored on a server is still retrievable, with a low communication challenge-response protocol



Proofs-of-retrievability [Juels, Kaliski '07]

Issue: a client wants to verify if a file stored on a server is still retrievable, with a low communication challenge-response protocol



Additional constraints: unbounded-use, low client storage, low computation

PoR with lifted codes

$$\mathcal{C} = \mathrm{Lift}(\mathrm{RS}_q(r), m)$$

Assumption: one can compute independent pseudo-random permutations

$$\sigma_i^{(\kappa)} \in \mathfrak{S}(\mathbb{F}_q), \quad 1 \leq i \leq n, \ \kappa \in \mathcal{K}$$

Initialisation:

- User picks $\kappa \in \mathcal{K}$ at random
- File F is encoded and permuted as follows:

$$F \ \mapsto \ \boldsymbol{c} \in \mathcal{C} \ \mapsto \ \boldsymbol{w} = \boldsymbol{\sigma}(\boldsymbol{c}) = (\sigma_1^{(\kappa)}(c_1), \ldots, \sigma_n^{(\kappa)}(c_n)) \in \mathbb{F}_q^n$$

- User stores κ, server stores $\boldsymbol{w}$

Verification:

- User picks a line $L \subset \mathbb{F}_q^m$ at random and sends it to the server
- Server reads $\boldsymbol{w}_{|L}$ and sends it back to the user
- User accepts iff $\boldsymbol{\sigma}^{-1}(\boldsymbol{w}_{|L}) \in \mathrm{RS}_q(r)$

Results

Informal result (for the lifted code with $m = 2$):

For every $\varepsilon \leq \varepsilon_0 \simeq 1$, we have:
if the server answers correctly to a fraction $\geq 1 - \varepsilon$ of the challenges,
then with probability $\geq 1 - \mathcal{O}\left(\frac{1}{n(\varepsilon_0 - \varepsilon)^2}\right)$ the file is extractable from the server.

Results

Informal result (for the lifted code with $m = 2$):

For every $\varepsilon \leq \varepsilon_0 \simeq 1$, we have:
if the server answers correctly to a fraction $\geq 1 - \varepsilon$ of the challenges,
then with probability $\geq 1 - \mathcal{O}\left(\frac{1}{n(\varepsilon_0 - \varepsilon)^2}\right)$ the file is extractable from the server.

Details in:
New Proofs of Retrievability using Locally Decodable Codes, L. & Levy-dit-Vehel
IEEE International Symposium on Information Theory, 2016

Results

Informal result (for the lifted code with $m = 2$):

For every $\varepsilon \leq \varepsilon_0 \simeq 1$, we have:
if the server answers correctly to a fraction $\geq 1 - \varepsilon$ of the challenges,
then with probability $\geq 1 - \mathcal{O}\left(\frac{1}{n(\varepsilon_0 - \varepsilon)^2}\right)$ the file is extractable from the server.

Details in:
New Proofs of Retrievability using Locally Decodable Codes, L. & Levy-dit-Vehel
IEEE International Symposium on Information Theory, 2016

This idea can be **generalised** to other codes such as design-based codes.

Details in:
Generic Constructions of PoRs from Codes and Instantiations, L. & Levy-dit-Vehel
submitted, 2018

Outline

Conclusion

- **Analysis** and **generalisation** of a family of high-rate locally correctable codes, namely **lifted Reed-Solomon codes**
- **Combinatorial formalism** for the construction of locally correctable codes, thanks to **block designs**
- Application to **private information retrieval** (PIR)
- Application to **proofs of retrievability** (PoR)

Future works

- PIR with **low server computation complexity**
  - 1 server read $\longrightarrow$ constant/sublinear number of server reads
- Extend the **lifting** process to other geometric varieties
  - *e.g.* the Hermitian variety
- Design-based codes allow us to **remove probabilistic decoders** from a definition of locally correctable codes
  - "usual" combinatorial coding-theoretic version of LCCs
  - new constructions? new bounds?