Codes with locality: constructions and applications to cryptographic protocols

Julien Lavauzelle
École Polytechnique & INRIA Saclay, Université Paris-Saclay
Séminaire UVSQ
13/11/2018
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
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
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
Original goal: transmit information in the presence of noise.

- **message** \( m \in \mathbb{F}_q^k \) \(\rightarrow\) **codeword** \( c \in C \subseteq \mathbb{F}_q^n \)
- **channel** \( \rightarrow \)
- **noisy codeword** \( c' \in \mathbb{F}_q^n \) \(\rightarrow\) **decoded message** \( m' (= m?) \)
- errors \( (c_i \neq c'_i \in \mathbb{F}_q) \) or erasures \( (c'_j = \perp) \)

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

- **C linear over** \( \mathbb{F}_q \), with \( k = \text{dim}(C) \)
- **\( d = d_{\text{min}}(C) \):** \( \min\{d(c,c'), c \neq c', (c,c') \in C^2\} \)

**Singleton bound** (code is MDS if bound is achieved): \( k + d \leq n + 1 \).
**Original goal:** transmit information in the presence of noise.

<table>
<thead>
<tr>
<th>message</th>
<th>codeword</th>
<th>channel</th>
<th>noisy codeword</th>
<th>decoded message</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m \in \mathbb{F}_q^k$</td>
<td>$c \in C \subseteq \mathbb{F}_q^n$</td>
<td></td>
<td>$c' \in \mathbb{F}_q^n$</td>
<td>$m' (= m?)$</td>
</tr>
</tbody>
</table>

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

**Hamming distance** \( d(u, v) := |\{i \in [1, n], u_i \neq v_i\}|. \)
**Original goal:** transmit information in the presence of noise.

| message $m \in \mathbb{F}_q^k$ | codeword $c \in C \subseteq \mathbb{F}_q^n$ | channel | noisy codeword $c' \in \mathbb{F}_q^n$ | decoded message $m' (= m?)$
|------------------------------|--------------------------------|--------|----------------|----------------|
| $\mapsto$ | $\mapsto$ | $\uparrow$ | $\mapsto$ | $\mapsto$
| errors ($c_i \neq c'_i \in \mathbb{F}_q$) or erasures ($c'_j = \perp$)

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

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

- Singleton bound (code is MDS if bound is achieved): $k + d \leq n + 1$. 

![Diagram](attachment:image.png)
Original goal: transmit information in the presence of noise.

Error-correcting codes

**message** \( m \in \mathbb{F}_q^k \) \( \mapsto \) **codeword** \( c \in C \subseteq \mathbb{F}_q^n \) \( \mapsto \) **channel** \( \mapsto \) **noisy** **codeword** \( c' \in \mathbb{F}_q^n \) \( \mapsto \) **decoded** **message** \( m'(=m?) \)

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

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

- \( C \) linear over \( \mathbb{F}_q \), with \( k = \dim(C) \),
- \( d = d_{\text{min}}(C) := \min\{d(c,c'), c \neq c', (c,c') \in C^2\}. \)

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

\[ k + d \leq n + 1. \]
**Definition** (Reed-Solomon code). Let \( x = (x_1, \ldots, x_n) \in \mathbb{F}_q^n \), pairwise distinct.

\[
\text{RS}_q(r, x) := \{(f(x_1), \ldots, f(x_n)), f \in \mathbb{F}_q[X], \deg f \leq r\}
\]
**Definition** (Reed-Solomon code). Let \( x = (x_1, \ldots, x_n) \in \mathbb{F}_q^n \), pairwise distinct.

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

\[ c_i = f(x_i) \]

- **Dimension** \( k = r + 1 \)
- **Minimum distance** \( d_{\text{min}} = n - r \) \( \Rightarrow \) MDS
- **Can decode any** \( b \) errors and \( e \) erasures
  \[ \text{if } e + 2b < d_{\text{min}} \]
  \[ \Rightarrow \text{in time } \Theta(n \log^3 n). \]

In this talk, \( \text{RS}_q(r, x) := \text{RS}_q(r, \mathbb{F}_q) \).
### Definition (Reed-Solomon code)

Let \( x = (x_1, \ldots, x_n) \in \mathbb{F}_q^n \), pairwise distinct.

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

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

In this talk,

\[
\text{RS}_q(r) := \text{RS}_q(r, \mathbb{F}_q)
\]
Goal: sublinear-time correction of some symbols of $c \in C$. 
Local correction

**Goal:** sublinear-time correction of some symbols of $c \in C$.

**Definition.** A code $C \subseteq \mathbb{F}_q^n$ 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** $D$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in C$ satisfying $d(y, c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(D(y)(i) = c_i) \geq 1 - \varepsilon$;
- $D(y)(i)$ queries at most $\ell$ symbols of $y$.

---

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

\[\begin{array}{cccccccccc}
\text{= error} & \text{= symbol to be corrected} \\
\end{array}\]

$y$ : $\ldots\times\square\ldots$
Local correction

**Goal:** sublinear-time correction of some symbols of \( c \in C \).

**Definition.** A code \( C \subseteq \mathbb{F}_q^n \) 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** \( D \) such that, for every \( y \in \mathbb{F}_q^n \) and \( c \in C \) satisfying \( d(y,c) \leq \delta n \) and for every \( 1 \leq i \leq n \):
- \( \Pr(D(y)(i) = c_i) \geq 1 - \varepsilon \);
- \( D(y)(i) \) queries at most \( \ell \) symbols of \( y \).

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

\( \Box = \text{error} \)

\( \square = \text{symbol to be corrected} \)

\( y : \)
**Goal:** sublinear-time correction of some symbols of $c \in C$.

**Definition.** A code $C \subseteq \mathbb{F}_q^n$ 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** $D$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in C$ satisfying $d(y, c) \leq \delta n$ and for every $1 \leq i \leq n$:

- $\Pr(D(y)(i) = c_i) \geq 1 - \varepsilon$;
- $D(y)(i)$ queries at most $\ell$ symbols of $y$. 

$n = 16, \ell = 3$

- $\square = \text{error}$
- $\square = \text{symbol to be corrected}$

$y : \quad \boxed{\times} \boxed{\Box} \boxed{\square} \Box \times \times \times \times \times \times$
Local correction

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

Definition. A code $C \subseteq \mathbb{F}_q^n$ 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 $D$ such that, for every $y \in \mathbb{F}_q^n$ and $c \in C$ satisfying $d(y, c) \leq \delta n$ and for every $1 \leq i \leq n$:
- $\Pr(D(y)(i) = c_i) \geq 1 - \varepsilon$;
- $D(y)(i)$ queries at most $\ell$ symbols of $y$.

(n = 16, $\ell = 3$)
- $\square$ = symbol to be corrected
- $\times$ = error

Goals:
- $\ell \ll n$
- $\varepsilon = O(\delta)$, ideally $\varepsilon = O(1)$
- $k = \dim C$ large
Example: Reed-Muller codes

\[ \text{RM}_q(m,r) := \{ (f(x) : x \in \mathbb{F}_q^m), f \in \mathbb{F}_q[X_1, \ldots, X_m], \deg f \leq r \} \]
Example: Reed-Muller codes

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

Assume \( r \leq q - 2 \), and let:
- \( c = (f(x) : x \in \mathbb{F}_q^m) \in \text{RM}_q(m, r) \)
- \( \phi : \mathbb{F}_q \rightarrow \mathbb{F}_q^m \) affine and injective
- \( L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m \) affine line

Local correction of \( y \in \mathbb{F}_q^m \) at coordinate \( i \in \mathbb{F}_q^m \):
1. Pick at random a line \( L \subset \mathbb{F}_q^m \) such that \( i \in L \).
2. Correct \( y|_L \) as a noisy \( \text{RS}_q(r) \) codeword, and output \( \tilde{y}_i \).
Example: Reed-Muller codes

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

Assume \( r \leq q - 2 \), and let:

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

Then, the restriction of \( c \) to \( L \) (or to \( \phi \)):

\[
c|_L := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \text{RS}_q(r)
\]
Example: Reed-Muller codes

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

Assume \( r \leq q - 2 \), and let:

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

Then, the **restriction** of \( c \) to \( L \) (or to \( \phi \)):

\[ c|_L := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \text{RS}_q(r) \]

**Local correction of** \( y \in \mathbb{F}_q^m \) **at coordinate** \( i \in \mathbb{F}_q^m \):

1. Pick at random a line \( L \subset \mathbb{F}_q^m \) such that \( i \in L \).
2. Correct \( y|_L \) as a noisy \( \text{RS}_q(r) \) codeword, and output \( \tilde{y}_i \).
Example: Reed-Muller codes

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

Assume $r \leq q - 2$, and let:
- $c = (f(x) : x \in \mathbb{F}_q^m) \in \text{RM}_q(m, r)$
- $\phi : \mathbb{F}_q \rightarrow \mathbb{F}_q^m$ affine and injective
- $L := \phi(\mathbb{F}_q) \subset \mathbb{F}_q^m$ affine line

Then, the **restriction** of $c$ to $L$ (or to $\phi$):

$$c|_L := ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \text{RS}_q(r)$$

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

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

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

J. Lavauzelle – Codes with locality: constructions and applications to cryptographic protocols – Sém. UVSQ
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":

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

Are there more polynomials than in RM codes?

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

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

$\phi : \mathbb{F}_4 \rightarrow \mathbb{F}_2^4, \phi(T) = (aT + b, cT + d)$

$((f \circ \phi))(T) = (aT + b)^2 (cT + d)^2 = (ac)^2 T^4 + (ad + bc)^2 T^2 + (bd)^2 \mod (T^4 - T^3)$

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

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

**Are there more polynomials than in RM codes?**
**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”:
\[ \forall \phi \text{ affine injective}, \quad ((f \circ \phi)(t) : t \in \mathbb{F}_q) \in \mathbb{R}\mathbb{S}_q(r) \]

Are there more polynomials than in RM codes?

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

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

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

Are there more polynomials than in RM codes?

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

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

\[
(f \circ \phi)(T) = (aT + b)^2(cT + d)^2
= (a^2T^2 + b^2)(c^2T^2 + d^2)
= (ac)^2T^4 + (ad + bc)^2T^2 + (bd)^2
= (ad + bc)^2T^2 + (ac)^2T + (bd)^2 \mod (T^4 - T)
\]

\[ \Rightarrow \text{for every } \phi, \text{ the “restriction” } (f \circ \phi)(T) \text{ can be interpolated as a univariate polynomial of degree 2} \]
High-rate construction: lifted codes (2)

\[ A^m := \mathbb{F}_q^m \quad \text{ev}_{A^m}(f) := (f(x) : x \in \mathbb{F}_q^m) \in \mathbb{F}_{q^{A^m}} \]
\[ \text{Aff}(m) := \{ \phi : \mathbb{F}_q \to \mathbb{F}_q^m, \text{injective and affine} \} \]

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

\[ \text{Lift}(\text{RS}_q(r), m) := \{ \text{ev}_{A^m}(f), f \in \mathbb{F}_q[X] \mid \forall \phi \in \text{Aff}(m), \text{ev}_{A^1}(f \circ \phi) \in \text{RS}_q(r) \} \]
High-rate construction: lifted codes (2)

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

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

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

$Lift(RS_q(r), m)$ is locally correctable with $\ell = n^{1/m}$ and $\varepsilon = \frac{2\delta}{1-r/q}$.
High-rate construction: lifted codes (2)

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

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

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

\( \text{Lift}(\text{RS}_q(r), m) \) 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 \( \text{Lift}(\text{RS}_q((1-\gamma)q), m) \) is locally correctable with \( \ell = n^{1/m} \), \( \varepsilon = \Theta_{m,\alpha}(\delta) \), and has rate

\[
R \geq 1 - \alpha.
\]
Bounds in [GKS13] are far from being tight.

- Ex: for $m = 2$, GKS' theorem gives $\gamma \leq \alpha^{32}$. 
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 $r = (1 - p^{-c})q - 1 \to \infty$,

$$R_{(e \to \infty)} = 1 - \left( 1 + \frac{1}{p + 2} \right) \left( \frac{1 + 1/p}{2} \right)^c + \frac{1}{p + 2} \left( \frac{1}{p^2} \right)^c.$$
Lifted codes are **monomial**, i.e. generated by evaluations of monomials

\[ \text{ev}^m_X(X_1^{d_1} \ldots X_m^{d_m}) = \text{ev}^m_X(X^d) \]

**Degree set** of a monomial code:

\[ \text{Deg}(C) := \{ d \in [0, q - 1]^m, \text{ev}^m_X(X^d) \in C \} \]

**Example** for \( C = \text{RM}(m, r) \):

\[ \text{Deg}(C) = \{ d \in [0, q - 1]^m, \sum_{i=1}^m d_i \leq r \} \]
Degree sets

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

\[ \text{ev}_{\mathbb{A}^m}(X_1^{d_1} \ldots X_m^{d_m}) = \text{ev}_{\mathbb{A}^m}(X^d) \]

**Degree set** of a monomial code:

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

**Example** for \( C = \text{RM}(m, r) \):

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

**A representation** for \( m = 2 \):

- \( \text{RM}_4(2, 4) \)
- \( \text{RM}_4(2, 2) \)
- \( \text{Lift}(\text{RS}_4(2), 2) \)
“Fractal” representation of degree sets (1)

$q = 16, r = 14$
$q = 8, r = 6$
$q = 4, r = 2$
\[ \text{Degree set of Lift}(\text{RS}_{2^e}((1 - 2^{-c})2^e - 1), 2) \text{ for fixed } c = 5 \text{ and increasing } e \geq 5. \]
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
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.

\[ 0 \to \text{RM}_q(m, k-1) \to \text{PRM}_q(m, k) \xrightarrow{\pi} \text{PRM}_q(m-1, k) \to 0. \]

where \( \pi \) is the restriction map \( \mathbb{P}^m \to \mathbb{P}^{m-1} \)
Projective space:

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

where \( a \sim b \) iff \( \exists \lambda \in \mathbb{F}_q^\times, a = \lambda b \)
Projective spaces

Projective space:

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

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

Defining an evaluation map over \( \mathbb{P}^m \) requires:

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

\[ u = (0 : \cdots : 0 : 1 : * : \cdots : *) \in \mathbb{P}^m \]
Projective spaces

Projective space:

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

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

Defining an evaluation map over \( \mathbb{P}^m \) requires:

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

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

We get:

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

\[ \text{ev}_{\mathbb{P}^m}(f) := (f(u) : u \in \mathbb{P}^m) \in \mathbb{F}_q^{\mathbb{P}^m} \]
Example. Projective Reed-Solomon code:

$$\text{PRS}_q(r) = \{ \text{ev}_{\mathbb{P}^1}(f) = (f(x) : x \in \mathbb{P}^1), f \in \mathbb{F}_q[X,Y]^H \}$$
Example. Projective Reed-Solomon code:

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

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

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

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

**Example.** Projective Reed-Solomon code:

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

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

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

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

**Remarks:**

- \( \text{ev}_{\mathbb{P}^1}(f \circ \phi) \neq \text{ev}_{\mathbb{P}^m}(f)_{|\phi(\mathbb{P}^1)} \) due to the choice of representative
- fortunately \( \text{ev}_{\mathbb{P}^1}(f \circ \phi) = \mathbf{w} \ast \text{ev}_{\mathbb{P}^m}(f)_{|\phi(\mathbb{P}^1)}, \) and \( \mathbf{w} \) is independent of \( f \).
Main results

Projective lifted codes...

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

Projective lifted codes...

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

Projective lifted codes...

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

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

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

Main results

Projective lifted codes...

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

\[
0 \rightarrow \text{Lift}(\text{RS}_q(r-1), m) \rightarrow \text{Lift}(\text{PRS}_q(r), m) \xrightarrow{\pi} \text{Lift}(\text{PRS}_q(r), m-1) \rightarrow 0,
\]

where \(\pi : \mathbb{P}^m \rightarrow \mathbb{P}^{m-1}\).

- are (up to equivalence) **cyclic codes** if \((q - 1)^2 \nmid (q^{m+1} - 1)\)
- quasi-cyclic codes otherwise
Main results

Projective lifted codes...

- are **locally correctable**, with parameters \((\ell = q, \delta, \varepsilon = \delta / \tau)\), where \(\tau\) is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of \(\text{Lift} (\text{RS}_q(r-1), m)\), \(\text{Lift} (\text{PRS}_q(r), m)\) and \(\text{Lift} (\text{PRS}_q(r), m - 1)\)
- satisfy the **puncturing/shortening** relation
  \[
  0 \rightarrow \text{Lift} (\text{RS}_q(r-1), m) \rightarrow \text{Lift} (\text{PRS}_q(r), m) \xrightarrow{\pi} \text{Lift} (\text{PRS}_q(r), m - 1) \rightarrow 0 ,
  \]
  where \(\pi : \mathbb{P}^m \rightarrow \mathbb{P}^{m-1}\).
- 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**
Main results

Projective lifted codes...

- are **locally correctable**, with parameters \((\ell = q, \delta, \varepsilon = \delta / \tau)\), where \(\tau\) is the relative correction capability of the small PRS code
- are **monomial**, with an **explicit bijection** between the degree sets of \(\text{Lift}(\text{RS}_q(r - 1), m)\), \(\text{Lift}(\text{PRS}_q(r), m)\) and \(\text{Lift}(\text{PRS}_q(r), m - 1)\)
- satisfy the **puncturing/shortening** relation
  \[
  0 \rightarrow \text{Lift}(\text{RS}_q(r - 1), m) \rightarrow \text{Lift}(\text{PRS}_q(r), m) \xrightarrow{\pi} \text{Lift}(\text{PRS}_q(r), m - 1) \rightarrow 0 ,
  \]
  where \(\pi : \mathbb{P}^m \rightarrow \mathbb{P}^{m-1}\).
- 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
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
**Remark.** Assume $r = q - 2$. Then,

$$a \in \text{RS}_q(q - 2) \iff \sum_{i=1}^{q} a_i = 0 \iff \langle 1, a \rangle = 0$$
Remark. Assume $r = q - 2$. Then,

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

$$c \in \text{Lift}(\text{RS}_q(q - 2), m) \iff \forall L \subseteq \mathbb{F}_q^m, \langle 1, c|_L \rangle = 0$$
Lifted codes when $r = q - 2$

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

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

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

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

$$\begin{bmatrix}
* & 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 & * & \cdots & * & \cdots & * \\
\end{bmatrix}$$

← indicator vector of line $L$
Point-line incidences in the affine space form a 2-design.

**Definition.** A *t-design* of parameters \((v, k, \lambda)\) consists in:
- a set \(X\) of points, \(|X| = v\),
- a set \(B\) of blocks \(B \subset X\), \(|B| = k\)

such that every \(t\)-set in \(X\) appears in exactly \(\lambda\) blocks.
Point-line incidences in the affine space form a 2-design.

**Definition.** A $t$-design of parameters $(v, k, \lambda)$ consists in:
- a set $X$ of points, $|X| = v$,
- a set $B$ of blocks $B \subset X$, $|B| = k$

such that every $t$-set in $X$ appears in exactly $\lambda$ blocks.

**Incidence matrix of a design:**

\[
\begin{pmatrix}
0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

← indicator vector of block $B$
The **code based on the design** $\mathcal{D} = (X, \mathcal{B})$ is the code $\mathcal{C} = \text{Code}(\mathcal{D}) \subseteq F_q^X$ admitting the incidence matrix of $\mathcal{D}$ as a parity-check matrix.

$$\text{Code}(\mathcal{D}) = \{ c \in F_q^X \mid \forall B \in \mathcal{B}, \sum_{x \in B} c_x = 0 \}$$

**Remark.** The dimension of $\text{Code}(\mathcal{D})$ highly depends on the field $F_q$. 

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

\[
\text{Code}(\mathcal{D}) = \{ c \in \mathbb{F}_q^X | \forall B \in \mathcal{B}, \sum_{x \in B} c_x = 0 \} 
\]

**Remark.** The dimension of \( \text{Code}(\mathcal{D}) \) 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

\[
\text{Code}(\mathcal{D}, \mathcal{L}) := \{ c \in \mathbb{F}_q^X | \forall B \in \mathcal{B}, c_{|B} \in \mathcal{L}_B \}.
\]

**Remark.** \( \text{Code}(\mathcal{D}, \mathcal{L}) = \text{Code}(\mathcal{D}) \) 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 $y \in \mathbb{F}_q^X$ 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 $\tilde{y}_i$. 
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 $\delta < \tau / 2$, 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 $\lceil \tau \ell \rceil$ 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 $\delta < \tau / 2$, Code$(\mathcal{D}, \mathcal{L})$ is a $(\ell, \delta, \varepsilon)$-LCC, where

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

Proposition $[t = 3]$. For every $\delta < \tau - \sqrt{2 / \ell}$, 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
→ “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
→ “combinatorial” coding-theoretic version of LCCs

**Remaining issues:**
- families of 3-designs with high dimension?
- best instances $(D, L)$ prescribed design parameters $(n, \ell, \lambda)$?
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
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
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 $i$?
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)
File $F$ encoded and stored on $\ell$ 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 $\ell$ 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 $q = (q_1, \ldots, q_\ell) \leftarrow Q(i)$ and sends $q_j$ to $S_j$
File $F$ encoded and stored on $\ell$ 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 $q = (q_1, \ldots, q_\ell) \leftarrow Q(i)$ and sends $q_j$ to $S_j$
2. Each server $S_j$ computes $a_j = A_j(q_j, F_i |_{S_j})$ and sends it back to $U$

Information-theoretic privacy: $I(i; q_j) = 0, \forall j = 1, \ldots, \ell$. 
File $F$ encoded and stored on $\ell$ 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 $q = (q_1, \ldots, q_\ell) \leftarrow Q(i)$ and sends $q_j$ to $S_j$
2. Each server $S_j$ computes $a_j = A_j(q_j, F_{|S_j})$ and sends it back to $U$
3. $U$ recovers $F_i = R(q, a, i)$
File $F$ encoded and stored on $\ell$ 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 $q = (q_1, \ldots, q_\ell) \leftarrow Q(i)$ and sends $q_j$ to $S_j$
2. Each server $S_j$ computes $a_j = A_j(q_j, F|_{S_j})$ and sends it back to $U$
3. $U$ recovers $F_i = R(q, a, i)$

**Information-theoretic privacy:** $I(i; q_j) = 0, \forall j = 1, \ldots, \ell$. 
Usual goals for PIR:

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

Our context: file $F$ is static and very frequently queried (e.g. public database)
Usual goals for PIR:

- Low communication complexity
- Low storage overhead for the servers (if coded)
- Low computation complexity for algorithms $A$ (server) and $R$ (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 $A$ (server) and $R$ (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”.
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
A **transversal design** \( \text{TD}(\ell, s) = (X, \mathcal{B}, G) \) is given by:

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

- **groups** \( G = \{G_j\}_{1 \leq j \leq \ell} \) satisfy \( X = \ell \bigsqcup_{j=1}^{\ell} G_j \) 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 \( G \in G \),
    - or in a unique block \( B \in \mathcal{B} \).
A transversal design $\text{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- $X$ a set of points, $|X| = n = s\ell$,
- groups $\mathcal{G} = \{G_j\}_{1 \leq j \leq \ell}$ satisfy
  \[ X = \bigsqcup_{j=1}^{\ell} G_j \] and $|G_j| = s$,
A transversal design $\text{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G})$ is given by:

- $X$ a set of points, $|X| = n = s\ell$,
- groups $\mathcal{G} = \{G_j\}_{1 \leq j \leq \ell}$ satisfy
  $$X = \bigsqcup_{j=1}^{\ell} G_j \text{ and } |G_j| = s,$$
- 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.
A **transversal design** \( \text{TD}(\ell, s) = (X, \mathcal{B}, \mathcal{G}) \) is given by:

- \( X \) a set of **points**, \( |X| = n = s\ell \),
- **groups** \( \mathcal{G} = \{G_j\}_{1 \leq j \leq \ell} \) satisfy
  \[
  X = \bigsqcup_{j=1}^{\ell} G_j \quad \text{and} \quad |G_j| = s ,
  \]
- **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:

\[
\begin{array}{ccc}
G_1 & G_2 & G_3 \\
\bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\end{array}
\]

\[
\begin{array}{ccc}
\mathcal{B} & = & \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \\
\mathcal{B}_1 & \cup & \mathcal{B}_2 \\
\mathcal{B}_3 \\
\end{array}
\]

\[
G_1 = \{1, 2, 3\} \\
G_2 = \{4, 5, 6\} \\
G_3 = \{7, 8, 9\}
\]

\[
B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}
\]

\[
B_1 = \{1, 2, 3\} \\
B_2 = \{4, 5, 6\} \\
B_3 = \{7, 8, 9\}
\]

\[
H = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
\]

Its rank over \(\mathbb{F}_3\) is 6 \implies the associated code \(C\) is a \([9,3]_3\) code.
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
Let $C \subseteq \mathbb{F}_q^n$ be a code based on a $\text{TD}(\ell, s)$. 

The scheme

- **Initialisation.** User $U$ encodes $F^{\rightarrow} \rightarrow c \in C$ and gives $c \mid G$ to server $S_j$.

- **To recover $F_i = c_i$:**
  1. User $U$ randomly picks a block $B \in B$ containing $i$. Then $U$ defines:
     
     
     
     
     

     2. Each server $S_j$ sends back $a_j = A_j(q_j, c \mid G_j)$:

     3. $U$ recovers $c_i = -\sum_j b \in B \setminus \{i\} c_b = c_i$. 

J. Lavauzelle – Codes with locality: constructions and applications to cryptographic protocols – Sém. UVSQ
Let $C \subseteq \mathbb{F}_q^n$ be a code based on a TD($\ell, s$).

- **Initialisation.** User $U$ encodes $F \mapsto c \in C$, and gives $c |_{G_j}$ 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:

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

  2. each server $S_j$ sends back $a_j = A_j(q_j, c |_{G_j}) := c_{q_j}$

  3. $U$ recovers

     $$- \sum_{j: i \notin G_j} c_{q_j} = - \sum_{b \in B \setminus \{i\}} c_b = c_i$$
Theorem. If the servers do not collude, then our PIR protocol is information-theoretically private.
**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 $i$. 
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 $k = \dim C \leq n = s \ell$.

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

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

Question: Transversal designs with good $k$ depending on $(\ell, s)$?
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
$\mathcal{T}_A$, 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 $\text{AG}_1(m,q)$ have the same length and the same dimension.

“Practical” instances:

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

Question: are they the best 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 \( X \),
- \( \mathcal{B} = \{ \text{affine lines } L \text{ secant to each group of } \mathcal{G} \} \).

**Proposition.** The code based on \( \mathcal{T}_A \) and the code based on \( AG_1(m, q) \) 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 partitioning \( X \),
- \( \mathcal{B} = \{ \text{affine lines } L \text{ secant to each group of } \mathcal{G} \} \).

**Proposition.** The code based on \( \mathcal{T}_A \) and the code based on \( AG_1(m, q) \) have same length and same dimension.

“Practical” instances:

- 3.2% storage overhead if \( \#\text{entries} \leq (\#\text{servers})^2 \)
- 27% storage overhead if \( \#\text{entries} \leq (\#\text{servers})^3 \)
\( \mathcal{T}_A \), the classical affine transversal design:

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

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

“Practical” instances:

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

**Question:** are they the best instances?
An **orthogonal array** $\text{OA}(t, \ell, s)$ of strength $t$ may be seen as a list $A$ of code-words
- over a finite set $S$, $|S| = s$,
- of length $\ell$,
- 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 $t + 1$. 
An orthogonal array $OA(t, \ell, s)$ of strength $t$ may be seen as a list $A$ of code-words
- over a finite set $S$, $|S| = s$,
- of length $\ell$,
- 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 $t + 1$. 

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

$$OA(2, 3, 2) = \begin{bmatrix}
a & b & b \\
b & b & a \\
 b & a & b \\
a & a & a
\end{bmatrix}$$
An orthogonal array \( OA(t, \ell, s) \) of strength \( t \) may be seen as a list \( A \) of codewords

- over a finite set \( S, |S| = s \),
- of length \( \ell \),
- 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 \( OA \rightarrow TD : \)

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

\[
\begin{bmatrix}
  a & b & b \\
  b & b & a \\
  b & a & b \\
  a & a & a \\
\end{bmatrix}
\]

\[
\begin{array}{ccc}
(a, 1) & (a, 2) & (a, 3) \\
(b, 1) & (b, 2) & (b, 3) \\
\end{array}
\]

\[ S = \{a, b\} \]
Instances with orthogonal arrays

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

- over a finite set $S$, $|S| = s$,
- of length $\ell$,
- 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 $\text{OA} \rightarrow \text{TD}$:**

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

**Example:**

$S = \{a, b\}$

$\text{OA}(2, 3, 2) = \begin{bmatrix}
  a & b & b \\
  b & b & a \\
  b & a & b \\
  a & a & a \\
\end{bmatrix}$
An **orthogonal array** $OA(t, \ell, s)$ of strength $t$ may be seen as a list $A$ of code-words
- over a finite set $S$, $|S| = s$,
- of length $\ell$,
- 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 $OA \to TD$:
- $X = S \times [1, \ell]$
- $G = \{S \times \{i\}, 1 \leq i \leq \ell\}$
- $B = \{(c_i, i), 1 \leq i \leq \ell\}, c \in OA$

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

$$OA(2, 3, 2) = \begin{bmatrix}
    a & b & b \\
    b & b & a \\
    b & a & b \\
    a & a & a
\end{bmatrix}$$
An **orthogonal array** $\text{OA}(t, \ell, s)$ of strength $t$ may be seen as a list $A$ of code-words
- over a finite set $S$, $|S| = s$,
- of length $\ell$,
- 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 :**
- $X = S \times [1, \ell]$
- $\mathcal{G} = \{S \times \{i\}, 1 \leq i \leq \ell\}$
- $\mathcal{B} = \{\{(c_i, i), 1 \leq i \leq \ell\}, c \in \text{OA}\}$

**Prop.** An OA$(2, \ell, s)$ gives a TD$(\ell, s)$.

\[ S = \{a, b\} \]

\[
\text{OA}(2,3,2) = \begin{bmatrix}
    a & b & b \\
    b & b & a \\
    b & a & b \\
    a & a & a \\
\end{bmatrix}
\]
Experimentally, for $t = 2$ and small $\ell$ and $s$, codes based on classical affine TDs have the largest dimension.
Experimentally, for $t = 2$ and small $\ell$ and $s$, codes based on classical affine TDs have the largest dimension.

**Question:** what about TDs from 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.
Experimentally, for $t = 2$ and small $\ell$ and $s$, codes based on classical affine TDs have the largest dimension.

**Question:** what about TDs from 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 $\ell$ and $s$, codes based on classical affine TDs have the largest dimension.

**Question:** what about TDs from 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 $\ell$ and $s$, codes based on classical affine TDs have the largest dimension.

**Question:** what about TDs from 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)