Algebraic methods in coding theory

Bruxelles Summer School of Mathematics

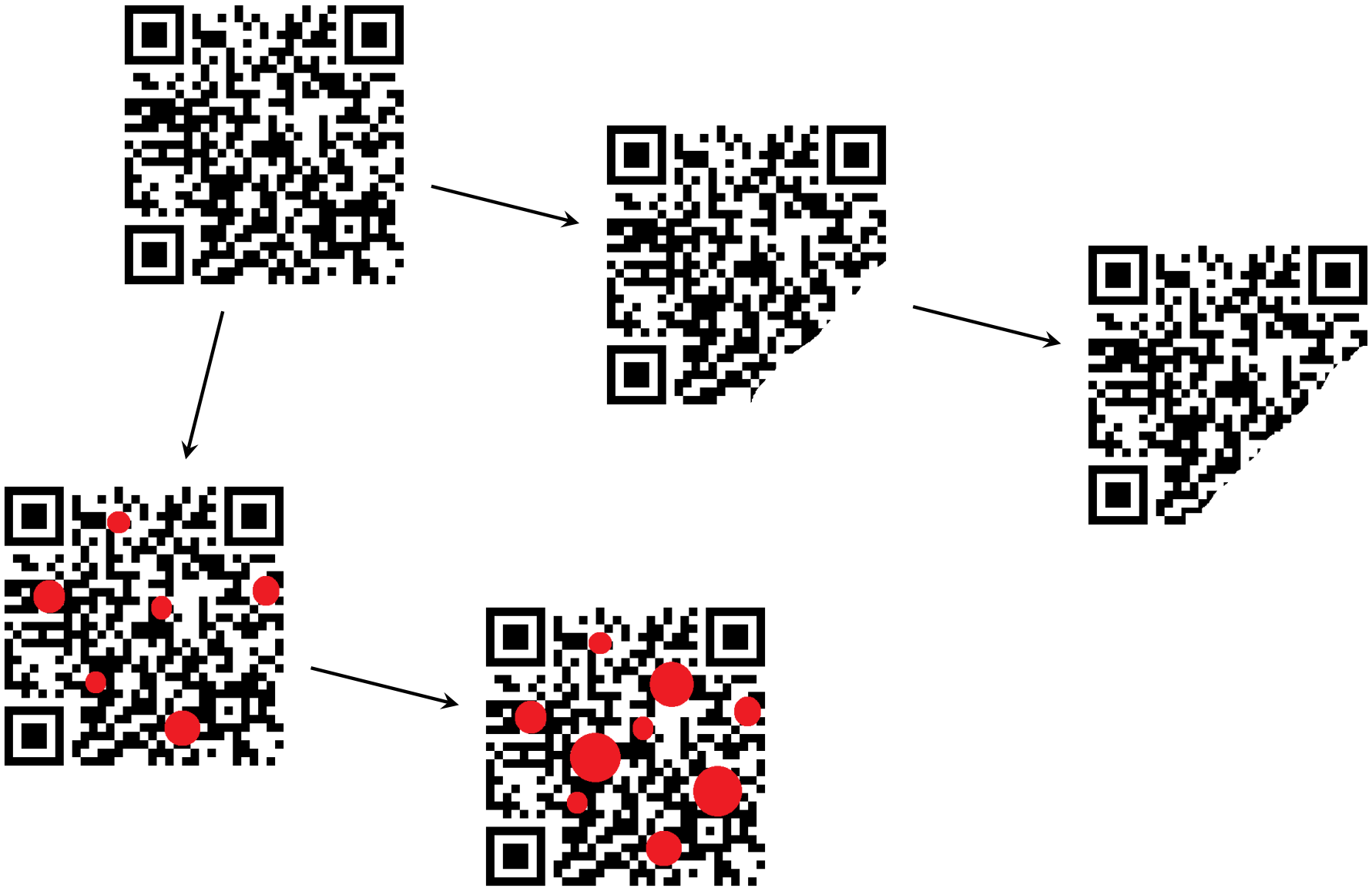
Marta Messia

September 11, 2026

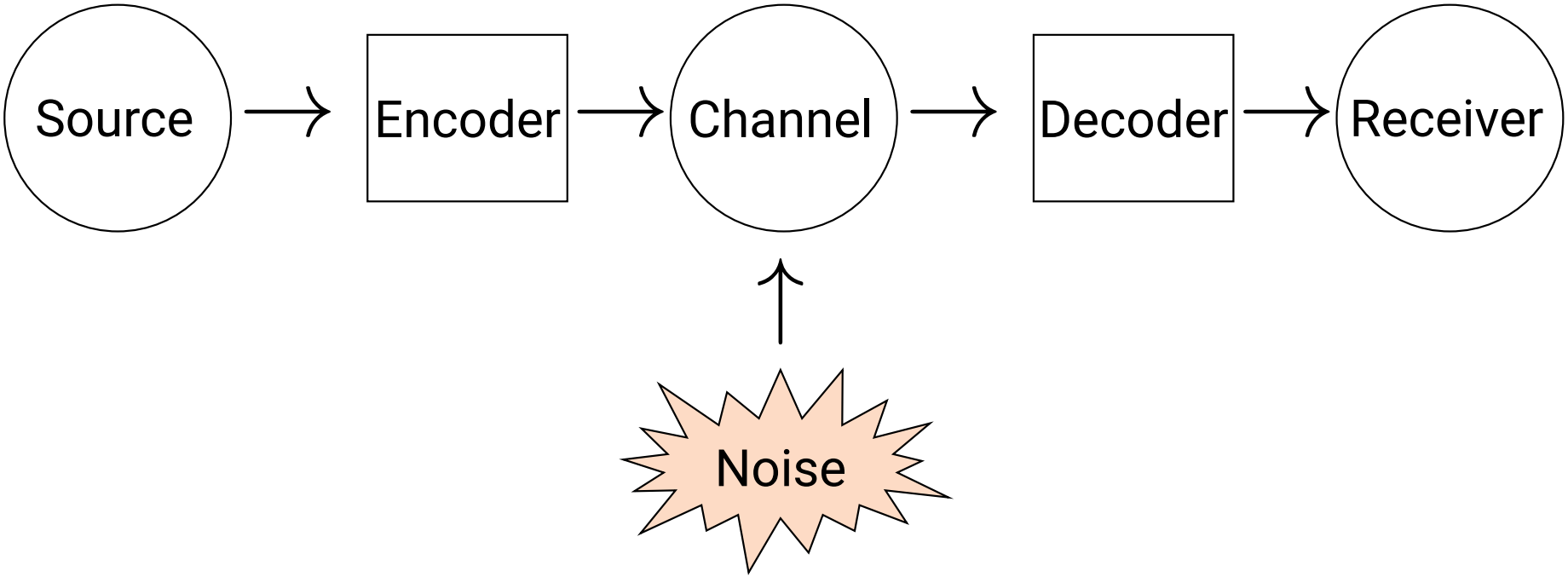
VUB VRIJE UNIVERSITEIT BRUSSEL

Introduction

Introduction



The problem: communication with errors



The challenge

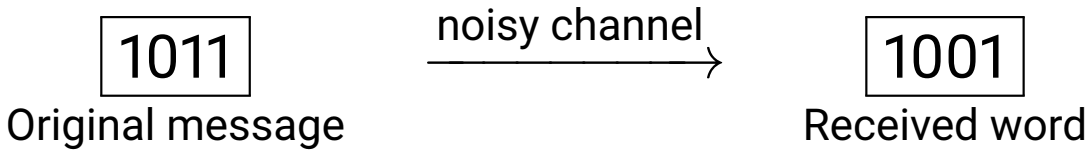
How can we reliably communicate information over a noisy channel?

What is an error

Definition

An **error** is any alteration of a transmitted message caused by noise in the communication channel.

Example



$$\underbrace{1001}_{\text{received}} - \underbrace{1011}_{\text{transmitted}} = \underbrace{0010}_{\text{error vector}}$$

Problem

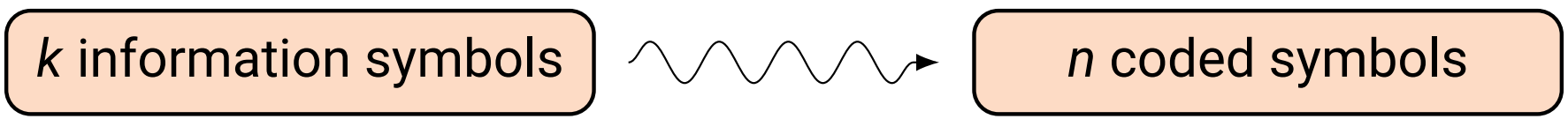
How can we encode information so that the original message can be recovered?

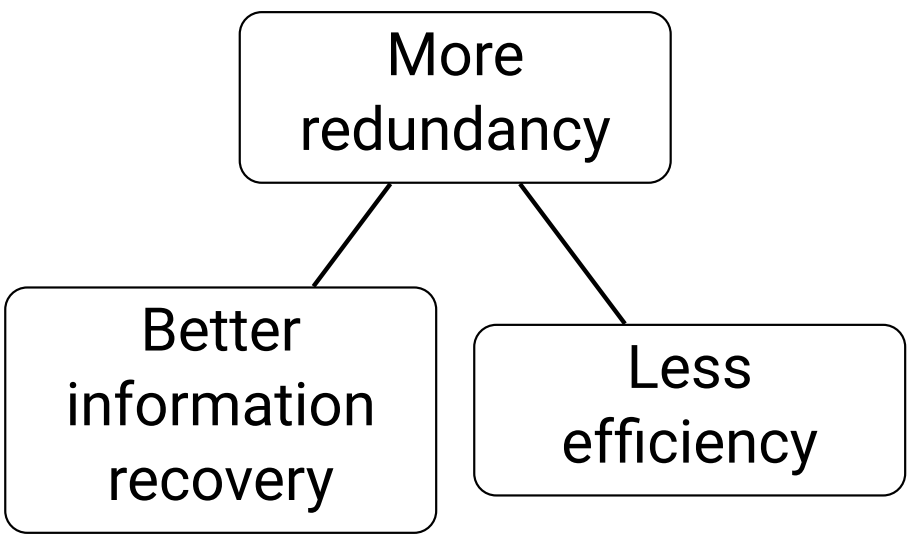
Redundancy

$k \leq n$

Fundamental idea

Add redundancy!





Question

How many errors can we correct?

A quick refresh on finite fields

Base field: $\mathbb{F}_2$.

Extension field structure

- $f(x) \in \mathbb{F}_2[x]$ irreducible, $\deg f = \ell$,
- α a root of f.

Then

$$\mathbb{F}_{2^\ell} = \mathbb{F}_2[\alpha] \simeq \mathbb{F}_2[x]/(f).$$

Example: Let $f(x) = x^2 + x + 1$ and let α be a root. Then

$$\mathbb{F}_4 = \mathbb{F}_2[\alpha] = \{0, 1, \alpha, 1+\alpha\}, \qquad \alpha^2 + \alpha + 1 = 0.$$

The vector space point of view

We will view $\mathbb{F}_{2^\ell}$ as a vector space over $\mathbb{F}_2$.

Vector space structure

- An $\mathbb{F}_2$-basis: $\{1, \alpha, \alpha^2, \ldots, \alpha^{\ell-1}\}$.

Then, as vector spaces,

$$\mathbb{F}_{2^\ell} \simeq \mathbb{F}_2^\ell.$$

Example: $\mathbb{F}_4 = \mathbb{F}_2[\alpha] \simeq \mathbb{F}_2^2$, via

$$\begin{aligned} 0 &= 0 + 0\alpha &\longleftrightarrow\quad & (0,0), \\ 1 &= 1 + 0\alpha &\longleftrightarrow\quad & (1,0), \\ \alpha &= 0 + 1\alpha &\longleftrightarrow\quad & (0,1), \\ 1+\alpha &= 1 + 1\alpha &\longleftrightarrow\quad & (1,1). \end{aligned}$$

Hamming codes

Codes

Notation Let n, M, k be positive integers such that $1 \leq k \leq n$.

Definition

An (n, M) code $\mathcal{C}$ is a subset of $\mathbb{F}_q^n$ of cardinality M.

We can endow $\mathbb{F}_q^n$ with the Hamming distance.

$$\begin{array}{rccc} d: & \mathbb{F}_q^n \times \mathbb{F}_q^n & \to & \mathbb{N} \\ & (x, y) & \mapsto & |\{i : x_i \neq y_i\}| \end{array}$$

The map d is a metric on $\mathbb{F}_q^n$.

Example In $\mathbb{F}_2^5$, $d((0,1,0,0,1),(1,0,0,1,1)) = 3$.

Remark We will focus only on linear codes!

Linear codes

Definition

An $[n, k, d]_q$ *linear code* $\mathcal{C}$ is a k-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_q^n$ endowed with the Hamming metric. In particular

$$d = d(\mathcal{C}) := \min\{d(x, y) \, : \, x, y \in \mathcal{C}, x \neq y\}.$$

We will call n, k, d respectively the *length*, *dimension* and *minimum distance* of the code.

Notation Hamming weight:

$$\mathrm{w}(x) = d(x, 0) = d(0, x).$$

Why linearity is useful

$$d(\mathcal{C}) = \min\{d(x,y) \,:\, x,y \in \mathcal{C}, x \neq y\}$$
$$= \min\{\mathrm{w}(x) \,:\, x \in \mathcal{C}, x \neq 0\}.$$

Why?

$$\boxed{x,y \in \mathcal{C}} \quad \underset{\mathcal{C} \text{ linear}}{\Longrightarrow} \quad \boxed{x-y \in \mathcal{C}}$$

$$d(x,y) = \mathrm{w}(x-y)$$

Key advantage

Determine the minimum distance

$\Downarrow$

Count nonzero components of the **nonzero codewords** and take the minimum

Generator matrix

Definition

Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code. Then, a matrix $G \in \mathbb{F}_q^{k \times n}$ is a *generator matrix* of $\mathcal{C}$ if

$$\mathcal{C} = \{xG \, : \, x \in \mathbb{F}_q^k\}.$$

Remark

- The rows of G are a basis of $\mathcal{C}$,
- the generator matrix is not unique.

Example

Field $\mathbb{F}_4 = \mathbb{F}_2[a]$, where a satisfies $a^2 + a + 1 = 0$.

The linear space

$$\mathcal{C} = \langle (1,0,0,0,1), (0,1,a,a,a^2) \rangle_{\mathbb{F}_4}$$

is a $[5,2]_4$ linear code.

$\mathbf{x}$	$w(\mathbf{x})$	$\mathbf{x}$	$w(\mathbf{x})$
$(a^2,1,a,a,0)$	4	$(1,0,0,0,1)$	2
$(a,1,a,a,1)$	5	$(1,1,a,a,a)$	5
$(1,a^2,1,1,a^2)$	5	$(0,1,a,a,a^2)$	4
$(a^2,0,0,0,a^2)$	2	$(a^2,a^2,1,1,1)$	5
(a^2,a,a^2,a^2,a)	5	$(1,a,a^2,a^2,0)$	4
(a,a,a^2,a^2,a^2)	5	$(0,a,a^2,a^2,1)$	4
$(0,0,0,0,0)$	0	$(a,0,0,0,a)$	2
$(a,a^2,1,1,0)$	4	$(0,a^2,1,1,a)$	4

1. Minimum distance

$$d(\mathcal{C}) = 2.$$

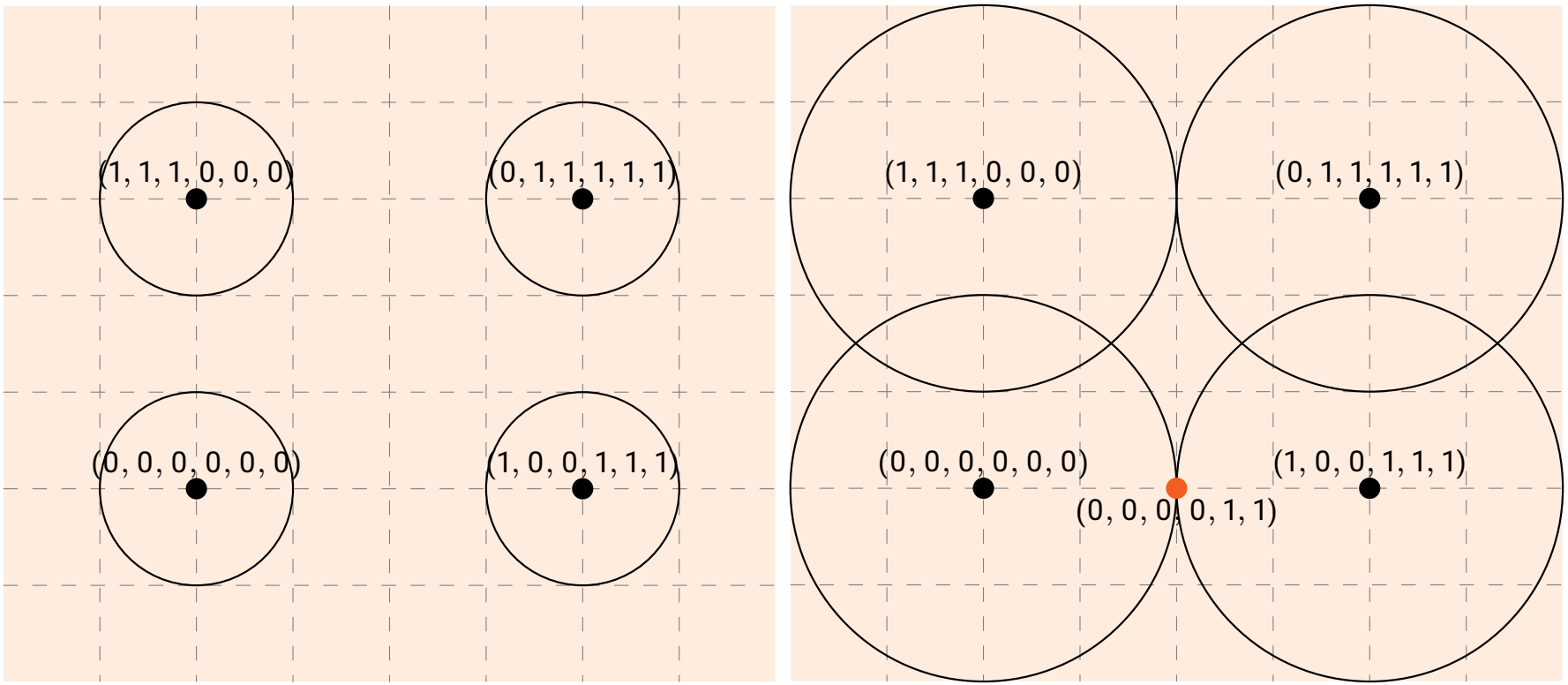
2. Generator matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & a & a & a^2 \end{pmatrix}.$$

Hamming spheres

$$\mathcal{B}(x, r) := \{y \in \mathbb{F}_q^n \ : \ d(x, y) \leq r\}$$

$\mathcal{C} = \{(0,0,0,0,0,0), (1,1,1,0,0,0), (1,0,0,1,1,1), (0,1,1,1,1,1)\}$



Parameters: $[6, 2, 3]_2$, we can correct 1 error.

Error correction capability

Theorem

An $[n, k, d]_q$ linear code can correct up to

$$\left\lfloor \frac{d-1}{2} \right\rfloor$$

errors.

Take away:

Large minimum distance $\Longrightarrow$ Better error correction

The Singleton bound

Question

Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code, if we fix n and k, what is the largest value of d that we can reach?

Theorem (Singleton bound)

Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code, then

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

Definition

An $[n, k, d]_q$ linear code is called *maximum distance separable (MDS)* if $d = n - k + 1$.

A proof of the Singleton bound

Proof.

$[n, k, d]_q$ linear code

$$(\underbrace{c_1, \ldots, c_{n-d+1}}_{n-d+1} \mid \underbrace{?, \ldots, ?}_{d-1})$$

$$\swarrow \qquad \searrow$$

$$(\underbrace{c_1, \ldots, c_{n-d+1}}_{n-d+1} \mid \underbrace{x_1, \ldots, x_{d-1}}_{d-1}) \neq (\underbrace{c_1, \ldots, c_{n-d+1}}_{n-d+1} \mid \underbrace{y_1, \ldots, y_{d-1}}_{d-1}) \quad \text{NO}$$

k	$\leq$	$n - d + 1$
degrees of freedom		fixed coordinates

□

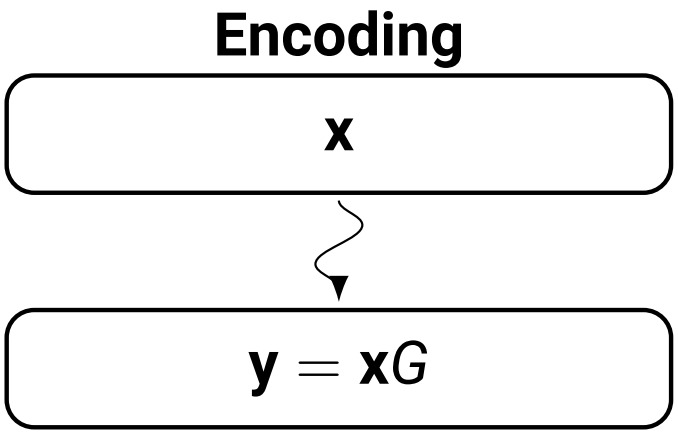
Encoding linear codes

Today's informal definition

The process of converting a message into a codeword of a precise code is called *encoding*.

Ingredients:

1. The message to encode $\mathbf{x} \in \mathbb{F}_q^k$.
2. An $[n, k, d]_q$ linear code with a generator matrix G.



Take away:
Encoding is **easy** for linear codes

From now on,
easy=*polynomial time (P)*.

Decoding linear codes

Today's informal definition

The process of recovering encoded information by correcting all the errors is called *decoding*.

Problem: Nearest Codeword Problem (NCP)

Input:

1. Code parameters: $[n, k, d]_q$.
2. Generator matrix: $G \in \mathbb{F}_q^{k \times n}$.
3. Received word: $\mathbf{y} \in \mathbb{F}_q^n$.
4. Threshold: $t \in \mathbb{Z}_{\geq 0}$.

Output:

Yes if there exists $\mathbf{x} \in \mathbb{F}_q^k$ such that $d(\mathbf{y}, \mathbf{x}G) \leq t$,
No otherwise.

Decoding is difficult

From now on, difficult=*NP-complete*.

Theorem

The Nearest Codeword Problem is NP-complete.

Take away:
Decoding is *in general* **difficult** for linear codes

Reed-Solomon codes

Reed-Solomon codes - Definition

Let $\{\alpha_1, \ldots, \alpha_n\} \in \mathbb{F}_q$ distinct elements,

$$\begin{array}{rccc} \mathrm{ev}_{\boldsymbol{\alpha}}: & \mathbb{F}_q[x] & \to & \mathbb{F}_q^n \\ & f & \mapsto & (f(\alpha_1), \ldots, f(\alpha_n)). \end{array}$$

Definition

The *Reed-Solomon (RS)* code with parameters $[n, k, d]_q$ and evaluation points $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is

$$RS_k(\boldsymbol{\alpha}) = \{\mathrm{ev}_{\boldsymbol{\alpha}}(f) : f \in \mathbb{F}_q[x], \deg(f) < k\}.$$

Example

Field:
$\mathbb{F}_4 = \{0, 1, \alpha, \alpha+1\}$,
$\alpha^2 + \alpha + 1 = 0$

Parameters:

- $n = 3$,
- $k = 2$,
- $\boldsymbol{\alpha} = (1, \alpha, \alpha^2)$.

Polynomials:
$f(x) = a_0 + a_1 x$,
$a_0, a_1 \in \mathbb{F}_4$

Code:
$RS_2((1, \alpha, \alpha^2))$

(a_0, a_1)	f	$\mathrm{ev}_{\boldsymbol{\alpha}}(f)$
$(0,0)$	0	$(0,0,0)$
$(0,1)$	x	$(1, \alpha, \alpha^2)$
$(0,\alpha)$	αx	$(\alpha, \alpha^2, 1)$
$(0,\alpha+1)$	$(\alpha+1)x$	$(\alpha+1, 1, \alpha)$
$(1,0)$	1	$(1,1,1)$
$(1,1)$	$1+x$	$(0, \alpha+1, \alpha)$
$(1,\alpha)$	$1+\alpha x$	$(\alpha+1, \alpha, 0)$
$(1,\alpha+1)$	$1+(\alpha+1)x$	$(\alpha, 0, \alpha^2)$
$(\alpha,0)$	α	(α, α, α)
$(\alpha,1)$	$\alpha+x$	$(\alpha+1, 0, 1)$
(α,α)	$\alpha+\alpha x$	$(0, 1, \alpha+1)$
$(\alpha,\alpha+1)$	$\alpha+(\alpha+1)x$	$(1, \alpha^2, \alpha^2)$
$(\alpha+1,0)$	$\alpha+1$	$(\alpha+1, \alpha+1, \alpha+1)$
$(\alpha+1,1)$	$\alpha+1+x$	$(\alpha, \alpha^2, 0)$
$(\alpha+1,\alpha)$	$\alpha+1+\alpha x$	$(1, 0, \alpha^2)$
$(\alpha+1,\alpha+1)$	$\alpha+1+(\alpha+1)x$	$(0, \alpha, 1)$

Reed-Solomon codes - Parameters I

Theorem

Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ be a tuple of distinct elements of $\mathbb{F}_q$, if $k \leq n$ then $RS_k(\boldsymbol{\alpha})$ is an $[n, k, n-k+1]_q$ MDS code.

Proof.

$RS_k(\boldsymbol{\alpha}) = \{\mathrm{ev}_{\boldsymbol{\alpha}}(f) \ : \ f \in \mathbb{F}_q[x], \deg(f) < k\}$

1. $\dim_{\mathbb{F}_q}(RS_k(\boldsymbol{\alpha})) = k$:
   - $\mathrm{ev}_{\boldsymbol{\alpha}}|_{\mathbb{F}_q[x]_{<k}}$ is injective,
   - $$\mathbb{F}_q[x]_{<k} \simeq \mathrm{Im}_{\mathrm{ev}_{\boldsymbol{\alpha}}}(\mathbb{F}_q[x]_{<k}) = RS_k(\boldsymbol{\alpha})$$

Reed-Solomon codes - Parameters II

Proof.

2. $d(RS_k(\alpha)) = n - k + 1$:

   - $$\begin{aligned} w((f(\alpha_1), \ldots, f(\alpha_n))) &= |\{i : f(\alpha_i) \neq 0\}| \\ &= n - |\{i : f(\alpha_i) = 0\}| \\ &\geq n - k + 1, \end{aligned}$$

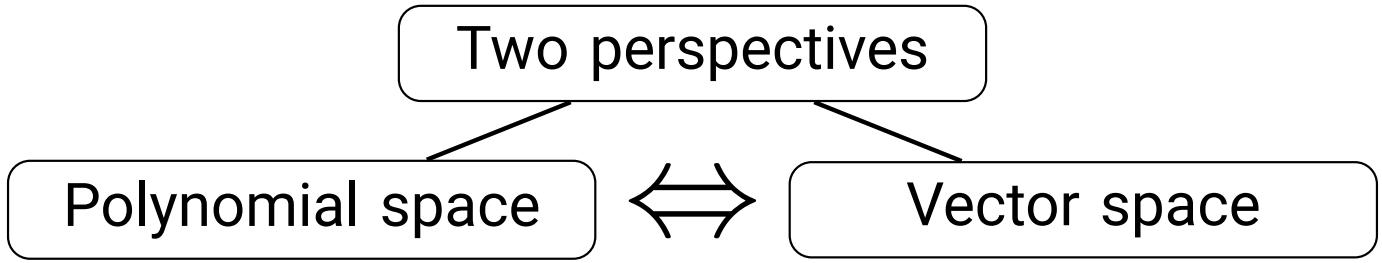
   - $$d(RS_k(\alpha)) \geq n - k + 1,$$

   - Singleton bound:

   $$d(RS_k(\alpha)) \leq n - k + 1.$$

□

Reed-Solomon codes - Generator matrix



$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$,
$c = \mathrm{ev}_{\boldsymbol{\alpha}}(f)$, with $f = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1}$,

$$\begin{cases} f(\alpha_1) = a_0 + \ldots + a_{k-1}\alpha_1^{k-1} \\ f(\alpha_2) = a_0 + \ldots + a_{k-1}\alpha_2^{k-1} \\ \quad\vdots \\ f(\alpha_n) = a_0 + \ldots + a_{k-1}\alpha_n^{k-1} \end{cases} \Leftrightarrow c = (a_0, \ldots, a_{k-1}) \begin{pmatrix} 1 & \ldots & 1 \\ \alpha_1 & \ldots & \alpha_n \\ \vdots & \ldots & \vdots \\ \alpha_1^{k-1} & \ldots & \alpha_n^{k-1} \end{pmatrix}$$

This matrix has rank k and is a generator matrix of $RS_k(\boldsymbol{\alpha})$.

Example

Field: $\mathbb{F}_4 = \{0, 1, \alpha, \alpha + 1\}, \quad \alpha^2 + \alpha + 1 = 0$

Parameters: $n = 3, k = 2, \boldsymbol{\alpha} = (1, \alpha, \alpha^2)$.

Generator matrix:

$$G = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \end{pmatrix}$$

Code

$$RS_2((1, \alpha, \alpha^2)) = \left\{ (a_0, a_1) \begin{pmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \end{pmatrix} \ : \ (a_0, a_1) \in \mathbb{F}_4 \right\}.$$

Decoding Reed-Solomon codes

An $[n, k, d]_q$ RS code can correct up to $\left\lfloor \frac{d-1}{2} \right\rfloor = \left\lfloor \frac{n-k}{2} \right\rfloor$ errors.

Problem: Reed-Solomon Unique Decoding

Input:

1. Code parameters: n, k, q.
2. Evaluation tuple: $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_q^n$.
3. Received word: $\mathbf{y} \in \mathbb{F}_q^n$.

Output:

1. $P \in \mathbb{F}_q[x]_{<k}$ such that $e := |\{i \,|\, y_i \neq P(\alpha_i)\}| \leq \left\lfloor \frac{n-k}{2} \right\rfloor$, if such a polynomial exists and `fail` otherwise.

Decoding and Interpolation

Key idea

A message of k symbols is represented by a polynomial

$$p(x) = m_0 + m_1 x + \cdots + m_{k-1} x^{k-1}, \qquad \deg p < k.$$

Reconstruction

$$\boxed{k \text{ correct points}} \underset{\text{Interpolation}}{\Rightarrow} \boxed{\text{Unique } p(x),\ \deg p < k}$$

Problem:
Identify the errors positions

Error-locator polynomial

Assumption from now on:

- the received word has been encoded correctly,
- $e \leq \left\lfloor \frac{n-k}{2} \right\rfloor$ errors occurred

$\Rightarrow$ there exists a polynomial $P \in \mathbb{F}_q[x]_{<k}$ such that $d(\mathbf{y}, \mathrm{ev}_{\boldsymbol{\alpha}}(P)) \leq e$.

Definition

A non-zero polynomial $E \in \mathbb{F}_q[x]$ is called an *error-locator polynomial* if for all i we have

$$E(\alpha_i) = 0 \text{ if } y_i \neq P(\alpha_i).$$

Remark There exists such a polynomial of degree at most e, for example

$$E(x) = \prod_{i\,:\,y_i \neq P(\alpha_i)} (x - \alpha_i).$$

A system with solution

By definition, for every $1 \leq i \leq n$

$$y_i E(\alpha_i) = P(\alpha_i) E(\alpha_i).$$

System of n equations in $k + e + 1 \leq n$ variables

$\Downarrow$

We can find a solution!

The system is quadratic

$\Downarrow$

Find the solution is hard!

Fundamental trick: Linearization!

Linearization

Define the polynomial $N(x) = P(x)E(x)$.

Remark

1. $\deg(N) \leq e + k - 1$.
2. The new system

   $$y_i E(\alpha_i) = N(\alpha_i), \quad 1 \leq i \leq n$$

   is linear.

Minor issue

We now have a system of possibly $2e + k + 1 \leq n + 1$ variables.

$\Rightarrow$

Solution

We can impose E to be monic and add $E_e = 1$ to the equations.

Welch-Berlekamp Algorithm

Algorithm: Welch-Berlekamp Algorithm

Input: $q > n > k \geq 1$, $0 < e \leq \left\lfloor \frac{n-k}{2} \right\rfloor$, n pairs $\{(\alpha_i, y_i)\}_{i \in \{1,\ldots,n\}}$.
Output: Polynomial $P \in \mathbb{F}_q[x]_{<k}$ or `fail`.

1: Compute a non-zero polynomial E of degree e and a polynomial N of degree at most $e + k - 1$ such that

$$\begin{cases} y_i E(\alpha_i) = N(\alpha_i), & 1 \leq i \leq n \\ E_e = 1 \end{cases}.$$

2: **If** the system has no solution or $E \nmid N$ **then**
3: **return** `fail`
4: $P \leftarrow \frac{N}{E}$.
5: **If** $d(\mathbf{y}, \mathrm{ev}_{\boldsymbol{\alpha}}(P)) > e$ **then**
6: **return** `fail`
7: **Else**
8: **return** P

Complexity

Theorem

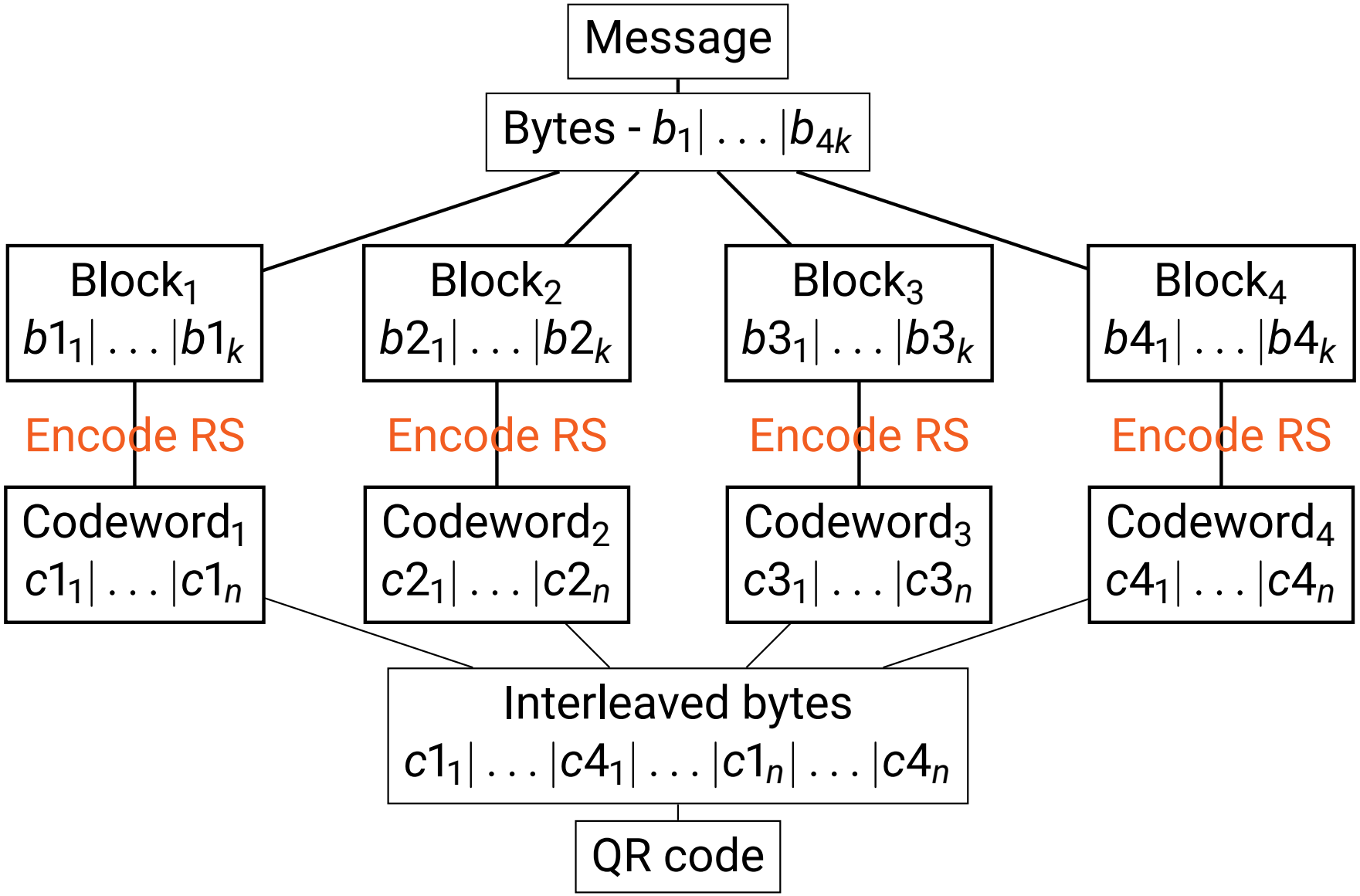
The Welch-Berlekamp algorithm runs in polynomial time.

Take away:

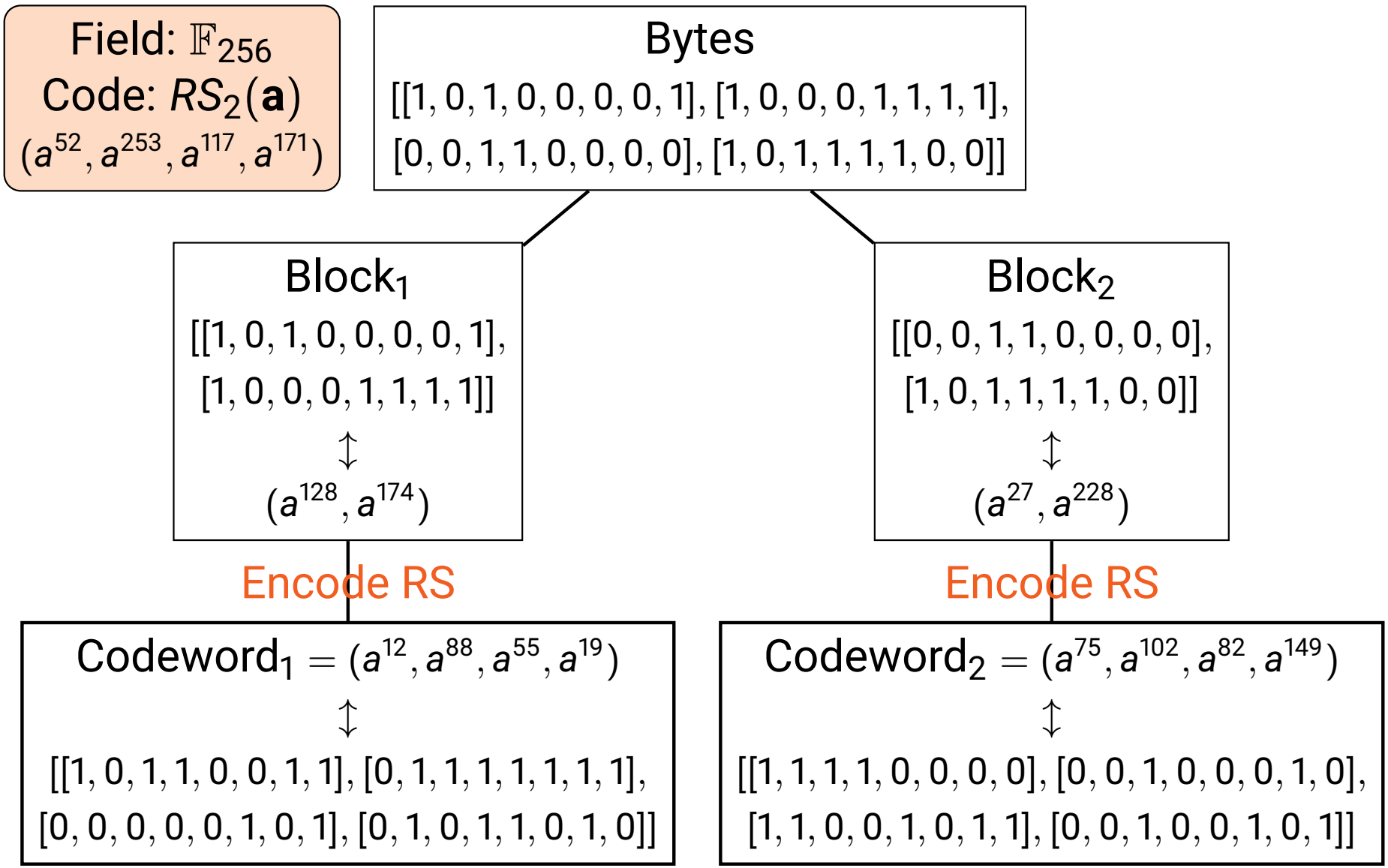
Decoding is **easy** for Reed-Solomon codes due to their rich algebraic structure!

QR-codes

An application - QR codes



From bytes to the QR-code



Why interleaving?

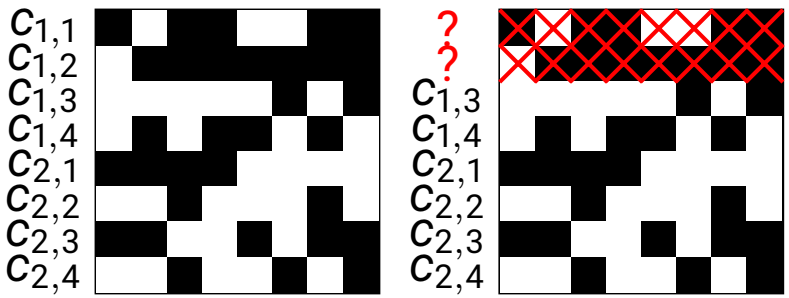
$c_1 = [[1,0,1,1,0,0,1,1],[0,1,1,1,1,1,1,1],[0,0,0,0,0,1,0,1],[0,1,0,1,1,0,1,0]]$
$c_2 = [[1,1,1,1,0,0,0,0],[0,0,1,0,0,0,1,0],[1,1,0,0,1,0,1,1],[0,0,1,0,0,1,0,1]]$

Our code is a $[4,2,3]_{256}$ Reed-Solomon code.

$\Rightarrow$

We can correct $\left\lfloor \frac{d-1}{2} \right\rfloor = 1$ error for each codeword.

Without interleaving



We are left with:

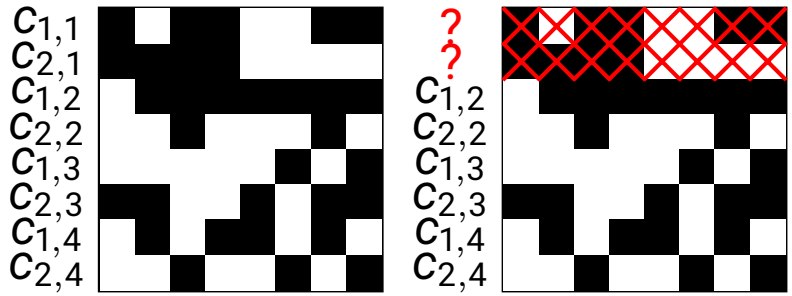
$c_1 = (?, ?, c_{1,3}, c_{1,4}) \rightarrow 2$ errors

$c_2 = (c_{2,1}, c_{2,2}, c_{2,3}, c_{2,4}) \rightarrow 0$ errors

$\Downarrow$

We can not decode c_1 ☹

With interleaving



We are left with:

$c_1 = (?, c_{1,2}, c_{1,3}, c_{1,4}) \rightarrow 1$ error

$c_2 = (?, c_{2,2}, c_{2,3}, c_{2,4}) \rightarrow 1$ error

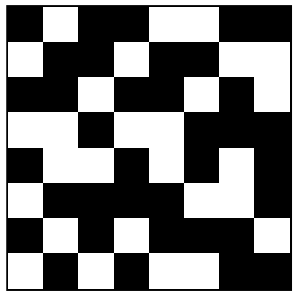
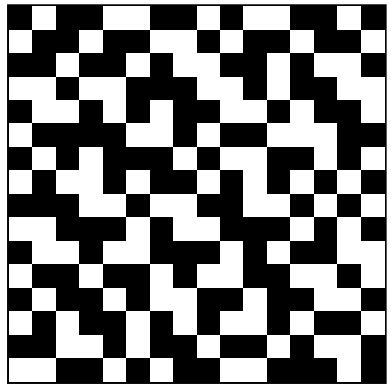
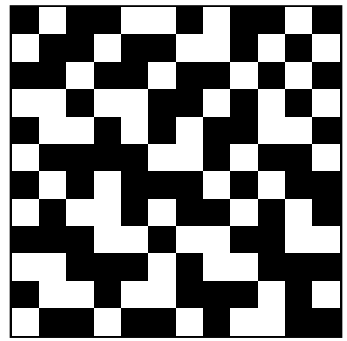
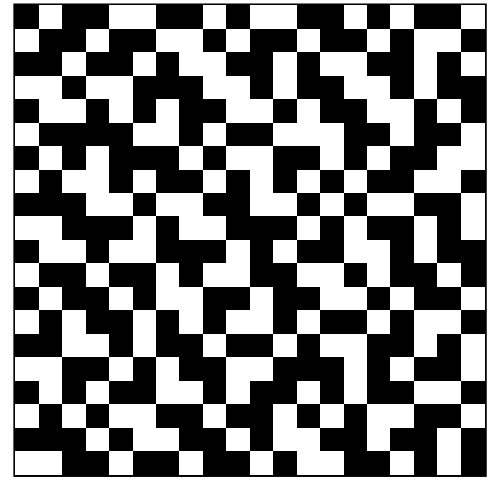
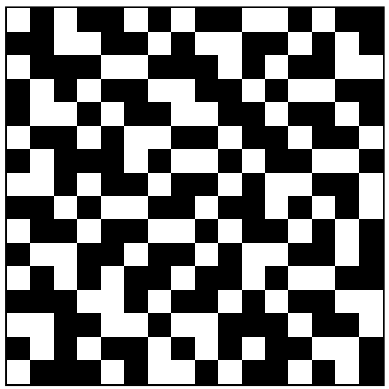
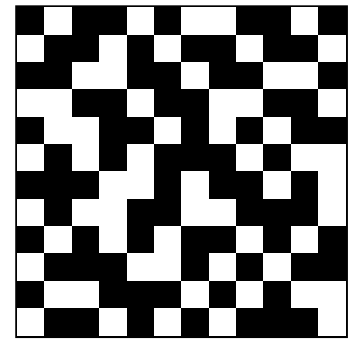
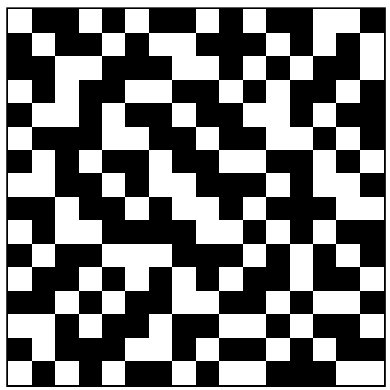
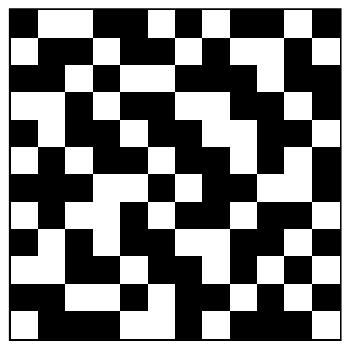
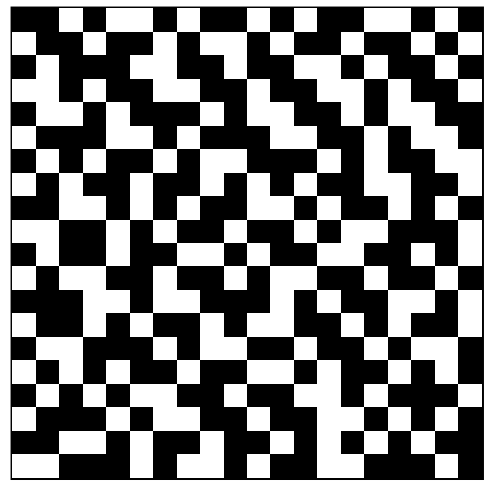
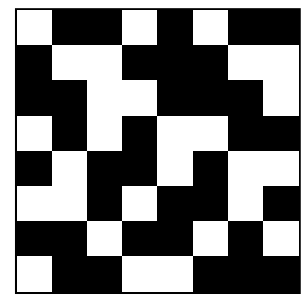
$\Downarrow$

We can decode everything ☺

Take aways

Take aways

1. Coding theory is about recovering information when some noise has occurred during transmission.
2. For linear codes, encoding is easy, and decoding is in general difficult.
3. For fixed length and dimension, Reed-Solomon codes achieve the maximum error-correction capability.
4. Decoding Reed-Solomon codes is easy.
5. QR-codes are an application of Reed-Solomon codes.







Thank you! ☺