Code-Based Cryptography

Error-Correcting Codes and Cryptography

I. Márquez-Corbella

informatiques mathématiques

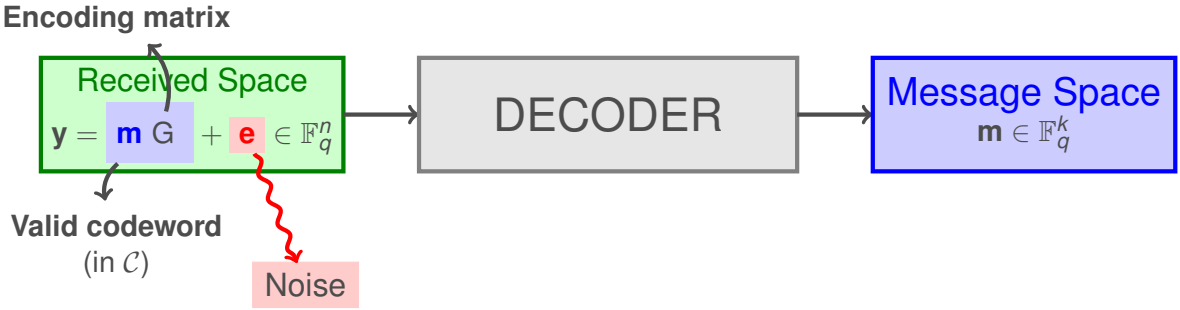
1. Error-Correcting Codes and Cryptography

1. Introduction I - Cryptography
2. Introduction II - Coding Theory
3. Encoding (Linear Transformation)
4. Parity Checking
5. Error Correcting Capacity
6. **Decoding (A Difficult Problem)**
7. Reed-Solomon Codes
8. Goppa Codes
9. McEliece Cryptosystem

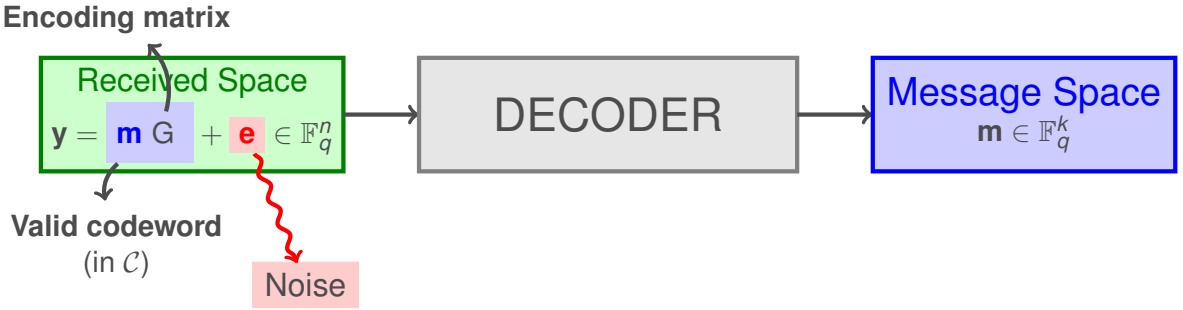
Decoder


Received Space
$\mathbf{y} = \mathbf{m}\,\mathrm{G} + \mathbf{e} \in \mathbb{F}_q^n$
DECODER
Message Space
$\mathbf{m} \in \mathbb{F}_q^k$

Decoder

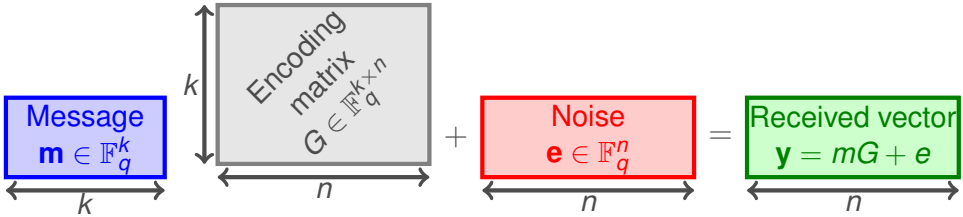

Encoding matrix
Received Space
$\mathbf{y} = \mathbf{m}\,\mathrm{G} + \mathbf{e} \in \mathbb{F}_q^n$
DECODER
Message Space
$\mathbf{m} \in \mathbb{F}_q^k$
Valid codeword
(in $\mathcal{C}$)
Noise

Decoder

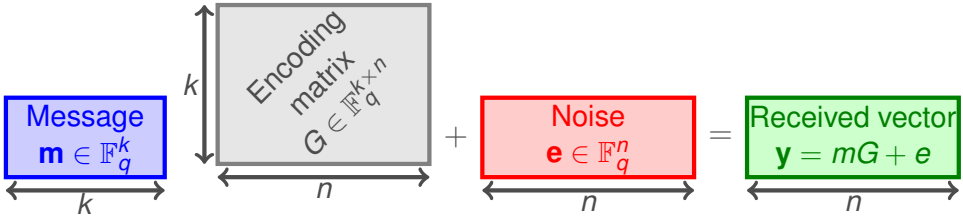


$$\texttt{Decoder}\big(\underbrace{\texttt{Encoder}(\texttt{Message})}_{\text{codeword}} + \texttt{Noise}\big) = \texttt{Message}$$

Minimum Distance Decoding (MDD)



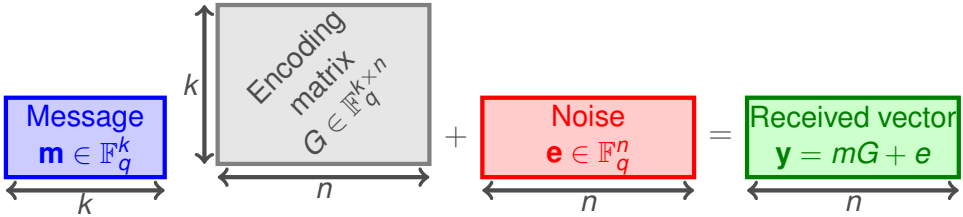
Minimum Distance Decoding (MDD)



Instances:

- A matrix $G \in \mathbb{F}_q^{k \times n}$ (generator matrix for $\mathcal{C}$)
- A received vector $\mathbf{y} \in \mathbb{F}_q^n$

Minimum Distance Decoding (MDD)



Instances:

- A matrix $G \in \mathbb{F}_q^{k \times n}$ (generator matrix for $\mathcal{C}$)
- A received vector $\mathbf{y} \in \mathbb{F}_q^n$

Output:

(Search - MDD): Find $\mathbf{m} \in \mathbb{F}_q^k$ to minimize

$$d_{\mathrm{H}}\left(\mathbf{y}, \mathbf{m}G\right)$$

Brute Force

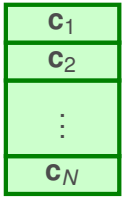
Let **y** be
the received word

First idea: Brute Force

Compute the Hamming distance of the received word with all codewords.

Brute Force

Let $\mathbf{y}$ be
the received word



with $N = q^k$

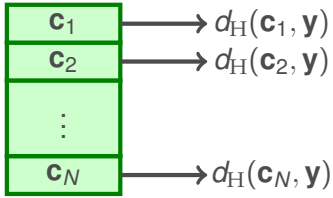
First idea: Brute Force

Compute the Hamming distance of the received word with all codewords.

1. Enumerate all codewords of $\mathcal{C}$.

Brute Force

Let **y** be
the received word



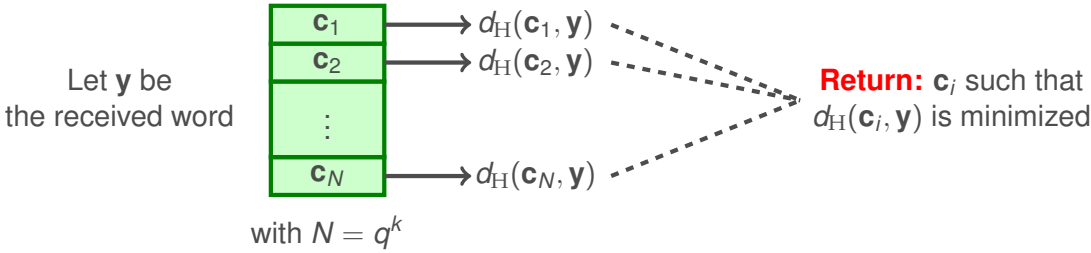
with $N = q^k$

First idea: Brute Force

Compute the Hamming distance of the received word with all codewords.

1. Enumerate all codewords of $\mathcal{C}$.
2. If **y** is the received word.
   Compute the Hamming distance $d_H(\mathbf{c}, \mathbf{y})$, $\forall \mathbf{c} \in \mathcal{C}$

Brute Force

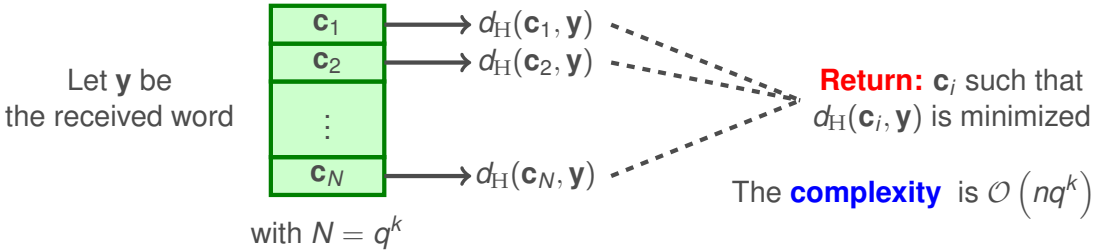


First idea: Brute Force

Compute the Hamming distance of the received word with all codewords.

1. Enumerate all codewords of $\mathcal{C}$.
2. If **y** is the received word.
   Compute the Hamming distance $d_H(\mathbf{c}, \mathbf{y})$, $\forall \mathbf{c} \in \mathcal{C}$
3. Return the codeword that minimizes d_H

Brute Force



First idea: Brute Force

Compute the Hamming distance of the received word with all codewords.

1. Enumerate all codewords of $\mathcal{C}$.
2. If **y** is the received word.
   Compute the Hamming distance $d_H(\mathbf{c}, \mathbf{y})$, $\forall \mathbf{c} \in \mathcal{C}$
3. Return the codeword that minimizes d_H

Syndrome

Let $\mathcal{C}$ be an $[n, k]_q$ code with parity check matrix H

$$\mathbf{c} \in \mathcal{C} \implies H\mathbf{c}^T = \mathbf{0}$$

Syndrome of a vector

The **syndrome of a vector** $\mathbf{x} \in \mathbb{F}_q^n$ is the vector $\mathrm{S}(\mathbf{x}) = H\mathbf{x}^T \in \mathbb{F}_q^{n-k}$

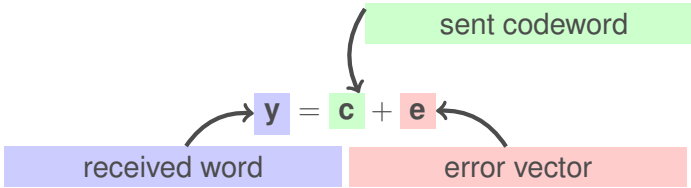
Syndrome

Let $\mathcal{C}$ be an $[n, k]_q$ code with parity check matrix H

$$\mathbf{c} \in \mathcal{C} \implies H\mathbf{c}^T = \mathbf{0}$$

Syndrome of a vector

The **syndrome of a vector** $\mathbf{x} \in \mathbb{F}_q^n$ is the vector $S(\mathbf{x}) = H\mathbf{x}^T \in \mathbb{F}_q^{n-k}$



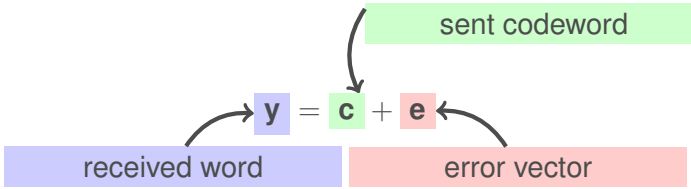
Syndrome

Let $\mathcal{C}$ be an $[n, k]_q$ code with parity check matrix H

$$\mathbf{c} \in \mathcal{C} \implies H\mathbf{c}^T = \mathbf{0}$$

Syndrome of a vector

The **syndrome of a vector** $\mathbf{x} \in \mathbb{F}_q^n$ is the vector $S(\mathbf{x}) = H\mathbf{x}^T \in \mathbb{F}_q^{n-k}$



$$H\mathbf{y}^T = H(\mathbf{c} + \mathbf{e})^T = \underbrace{H\mathbf{c}^T}_{=\mathbf{0}} + H\mathbf{e}^T = H\mathbf{e}^T$$

Syndrome Decoding - Lookup table

Let **y** be
the received word

Suppose we want to correct all patterns of $\leq t$ errors

Syndrome Decoding - Lookup table

Let **y** be
the received word

$S(\mathbf{e}_1) = S_1$
$S(\mathbf{e}_2) = S_2$
$\vdots$
$S(\mathbf{e}_N) = S_N$

Suppose we want to correct all patterns of $\leq t$ errors

1. **Precompute the syndrome corresponding to** $0, 1, \ldots, t$
   Number of Syndromes to pre-compute and store:

$$\binom{n}{0} + (q-1)\binom{n}{1} + (q-1)^2\binom{n}{2} + \ldots + (q-1)^t\binom{n}{t}$$

Syndrome Decoding - Lookup table

Let $\mathbf{y}$ be the received word

$S(\mathbf{e}_1) = S_1$
$S(\mathbf{e}_2) = S_2$
$\vdots$
$S(\mathbf{e}_N) = S_N$

If $S_i = S(\mathbf{y})$, **Return:** $\mathbf{y} - \mathbf{e}_i$

Suppose we want to correct all patterns of $\leq t$ errors

1. **Precompute the syndrome corresponding to** $0, 1, \ldots, t$
   Number of Syndromes to pre-compute and store:

$$\binom{n}{0} + (q-1)\binom{n}{1} + (q-1)^2\binom{n}{2} + \ldots + (q-1)^t\binom{n}{t}$$

2. **Compute the Syndrome of the received word** $S(\mathbf{y})$
   - ➜ If there exists $\mathbf{e} \in \mathbb{F}_q^n$ with $w_H(\mathbf{e}) \leq t : S(\mathbf{e}) = S(\mathbf{y}) \Longrightarrow$ **Return:** $\mathbf{y} - \mathbf{e}$
   - ➜ Otherwise, $\Longrightarrow$ **Return:** `FAILURE`

Gilbert-Varshamov bound

GV bound

$$\sum_{i=0}^{d-2}(q-1)^i\binom{n-1}{i} < q^{n-k} \implies \text{Exists an } [n,k,d]_q \text{ code}$$

Proof:

Gilbert-Varshamov bound

GV bound

$$\sum_{i=0}^{d-2}(q-1)^i\binom{n-1}{i} < q^{n-k} \implies \text{Exists an } [n,k,d]_q \text{ code}$$

Proof:

Let $H \in \mathbb{F}_q^{(n-k)\times n}$ be a parity check matrix of $\mathcal{C}$

Gilbert-Varshamov bound

GV bound

$$\sum_{i=0}^{d-2}(q-1)^i\binom{n-1}{i} < q^{n-k} \implies \text{Exists an } [n,k,d]_q \text{ code}$$

Proof:

Let $H \in \mathbb{F}_q^{(n-k)\times n}$ be a parity check matrix of $\mathcal{C}$

every $d-1$ columns of H are Linear independent

Gilbert-Varshamov bound

GV bound

$$\sum_{i=0}^{d-2}(q-1)^i\binom{n-1}{i} < q^{n-k} \implies \text{Exists an } [n,k,d]_q \text{ code}$$

Proof:

Let $H \in \mathbb{F}_q^{(n-k)\times n}$ be a parity check matrix of $\mathcal{C}$

every $d-1$ columns of H are Linear independent

We construct by induction the columns $h_1, \ldots, h_n \in \mathbb{F}_q^{n-k}$ of H.

Proof (Part II)

We choose:

Proof (Part II)

We choose:

- $h_1 \in \mathbb{F}_q^{n-k}$ any nonzero vector

Proof (Part II)

We choose:

- $h_1 \in \mathbb{F}_q^{n-k}$ any nonzero vector
- $h_2 \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a multiple of h_1

Proof (Part II)

We choose:

- $h_1 \in \mathbb{F}_q^{n-k}$ any nonzero vector
- $h_2 \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a multiple of h_1
- $\ldots$
- $h_j \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a **LC** of $\leq (d-2)$ of $\{h_1, \ldots, h_{j-1}\}$

Proof (Part II)

We choose:

- $h_1 \in \mathbb{F}_q^{n-k}$ any nonzero vector
- $h_2 \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a multiple of h_1
- $\ldots$
- $h_j \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a **LC** of $\leq (d-2)$ of $\{h_1, \ldots, h_{j-1}\}$

Let $j < n$. Exists h_{j+1} with the above property if:

$$\sum_{i=0}^{d-2} \binom{j}{i} (q-1)^i \leq q^{n-k} - 1$$

Number of **LC** of $\leq (d-2)$ of $\{h_1, \ldots, h_j\}$

Total number of vectors in $\mathbb{F}_q^n$ but one

Proof (Part II)

We choose:

- $h_1 \in \mathbb{F}_q^{n-k}$ any nonzero vector
- $h_2 \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a multiple of h_1
- $\ldots$
- $h_j \in \mathbb{F}_q^{n-k}$ any vector that is **NOT** a **LC** of $\leq (d-2)$ of $\{h_1, \ldots, h_{j-1}\}$

Let $j < n$. Exists h_{j+1} with the above property if:

$$\sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i \leq q^{n-k} - 1$$

Number of **LC** of $\leq (d-2)$ of $\{h_1, \ldots, h_j\}$

Total number of vectors in $\mathbb{F}_q^n$ but one

Gilbert Varshamov distance

Gilbert-Varshamov (GV) distance

The GV distance of an $[n, k]_q$ code is the **maximal integer** d_0 such that:

$$\sum_{i=0}^{d_0-1} \binom{n}{i} (q-1)^i \leq q^{n-k}$$

Number of codewords of a given weight

$$A_w(\mathcal{C}) = |\{\mathbf{c} \in \mathcal{C} \mid w_H(\mathbf{c}) = w\}|$$

Distinct codewords in $\mathcal{C}$ of weight exactly **w**

Number of codewords of a given weight

$$A_w(\mathcal{C}) = |\{\mathbf{c} \in \mathcal{C} \mid w_H(\mathbf{c}) = w\}|$$

Distinct codewords in $\mathcal{C}$ of weight exactly **w**

In a binary random code: $\mathbb{E}[A_w(\mathcal{C})] = \frac{\binom{n}{w}|\mathcal{C}|}{2^n} = \frac{\binom{n}{w}}{2^{n-k}}$

Number of codewords of a given weight

$$A_w(\mathcal{C}) = |\{\mathbf{c} \in \mathcal{C} \mid \mathrm{w}_H(\mathbf{c}) = w\}|$$

Distinct codewords in $\mathcal{C}$ of weight exactly **w**

In a binary random code: $\mathbb{E}[A_w(\mathcal{C})] = \frac{\binom{n}{w}|\mathcal{C}|}{2^n} = \frac{\binom{n}{w}}{2^{n-k}}$

In average:

Exists $\mathbf{c} \in \mathcal{C}$ with $\mathrm{w}_H(\mathbf{c}) = w \iff \binom{n}{w} > 2^{n-k}$

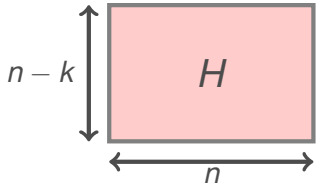
$\iff$ w is closed to the GV distance

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem

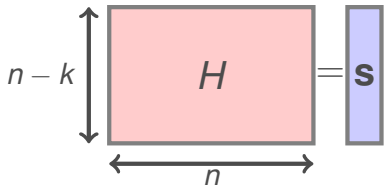


Input:

➜ A matrix $H \in \mathbb{F}_2^{(n-k)\times n}$

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem

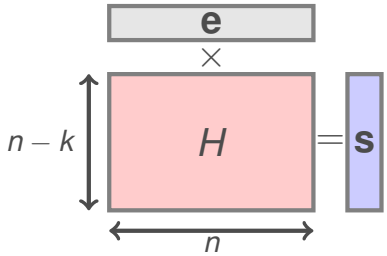


Input:

- A matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
- A syndrome $\mathbf{s} \in \mathbb{F}_2^{n-k}$

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem



Input:

- A matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
- A syndrome $\mathbf{s} \in \mathbb{F}_2^{n-k}$
- A weight $w \in \mathbb{Z}$

The Syndrome Decoding (SD) problem

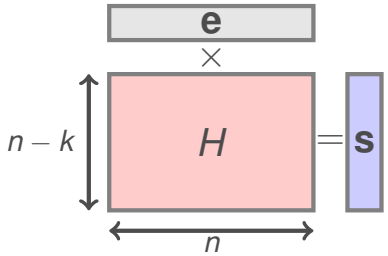
The Syndrome Decoding (SD) problem

Output

(Decision): Does $\mathbf{e} \in \mathbb{F}_2^n$ of $w_H(\mathbf{e}) \leq w$ such that $\mathbf{e}H^T = \mathbf{s}$ exists? NP-complete

E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg.
On the Inherent Intractability of Certain Coding Problems.
IEEE Trans. Inf. Theory. Vol. 24, pp. 384-386, 1978.

A. Barg.
Complexity Issues in Coding Theory.
Chapter 7, in Handbock of Coding Theory, 1998.



Input:

- A matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
- A syndrome $\mathbf{s} \in \mathbb{F}_2^{n-k}$
- A weight $w \in \mathbb{Z}$

The Syndrome Decoding (SD) problem

The Syndrome Decoding (SD) problem

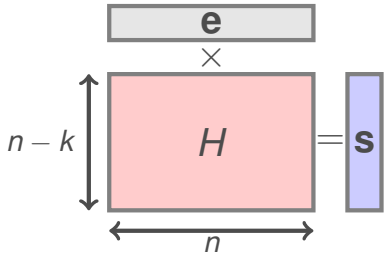
Output

(Decision): Does $\mathbf{e} \in \mathbb{F}_2^n$ of $w_H(\mathbf{e}) \leq w$ such that $\mathbf{e}H^T = \mathbf{s}$ exists? NP-complete

(Computational): Find $\mathbf{e} \in \mathbb{F}_2^n$ of $w_H(\mathbf{e}) \leq w$ such that $\mathbf{e}H^T = \mathbf{s}$ NP-difficult

E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg.
On the Inherent Intractability of Certain Coding Problems.
IEEE Trans. Inf. Theory. Vol. 24, pp. 384-386, 1978.

A. Barg.
Complexity Issues in Coding Theory.
Chapter 7, in Handbock of Coding Theory, 1998.



Input:

- A matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
- A syndrome $\mathbf{s} \in \mathbb{F}_2^{n-k}$
- A weight $w \in \mathbb{Z}$

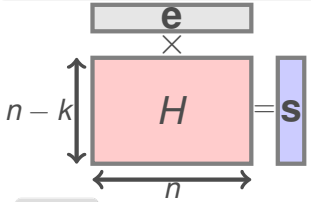
General Decoding

Input:

- A parity-check matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
  - A received vector $\mathbf{y} \in \mathbb{F}_2^{n-k}$
  - A weight $w \in \mathbb{Z}$

SD

Find $\mathbf{e} \in \mathbb{F}_2^n$ of $\mathrm{w}_H(\mathbf{e}) \leq w$ such that
$\mathbf{e}H^T = \mathbf{y}H^T = \mathbf{s}$



General Decoding

Input:

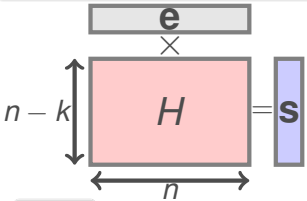
- A parity-check matrix $H \in \mathbb{F}_2^{(n-k)\times n}$
- A generator matrix $G \in \mathbb{F}_2^{k\times n}$
- A received vector $\mathbf{y} \in \mathbb{F}_2^{n-k}$
- A weight $w \in \mathbb{Z}$

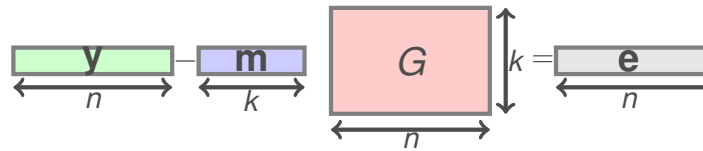
SD

Find $\mathbf{e} \in \mathbb{F}_2^n$ of $\mathrm{w}_H(\mathbf{e}) \leq w$ such that $\mathbf{e}H^T = \mathbf{y}H^T = \mathbf{s}$

MDD

Find $\mathbf{m} \in \mathbb{F}_2^k$ such that $\mathrm{w}_H(\mathbf{y} - m G) \leq w$

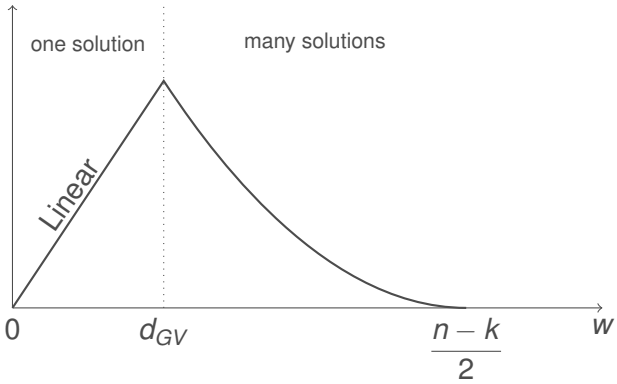




Computational Analysis of Syndrome Decoding

cost (log) of ISD

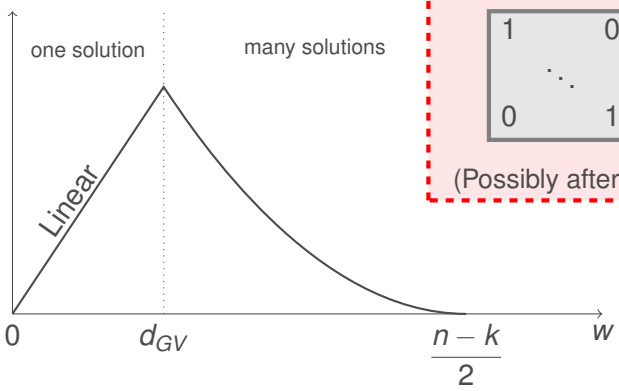
Binary codes



Computational Analysis of Syndrome Decoding

cost (log) of ISD

Binary codes



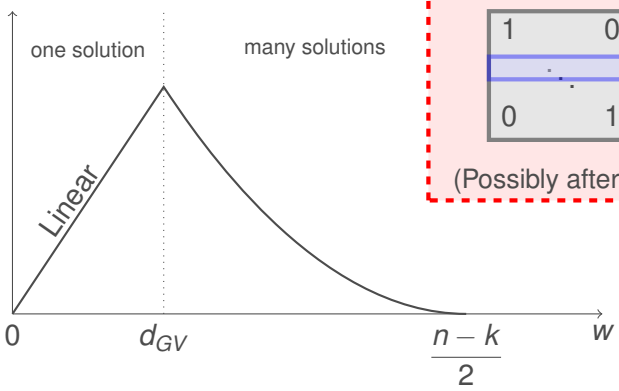
Case $w = \frac{n-k}{2}$

(Possibly after permuting some columns)

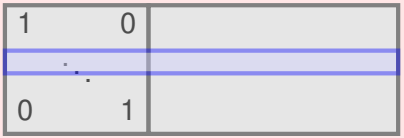
Computational Analysis of Syndrome Decoding

cost (log) of ISD

Binary codes



Case $w = \frac{n-k}{2}$

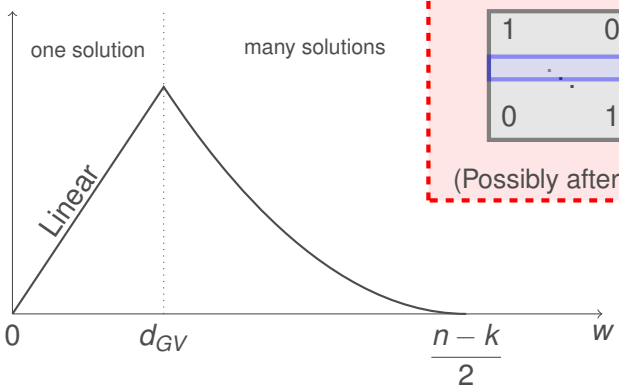


(Possibly after permuting some columns)

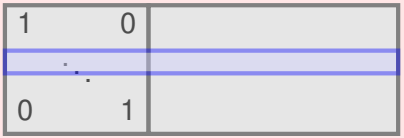
Computational Analysis of Syndrome Decoding

cost (log) of ISD

Binary codes



Case $w = \frac{n-k}{2}$



(Possibly after permuting some columns)

In average:

Exists $\mathbf{c} \in \mathcal{C}$ with $\mathrm{w}_H(\mathbf{c}) = w \iff \binom{n}{w} > 2^{n-k}$

$\iff w$ is closed to the GV distance

Trapdoor one-way functions - Decoder

Encoder = Matrix Multiplication

EASY

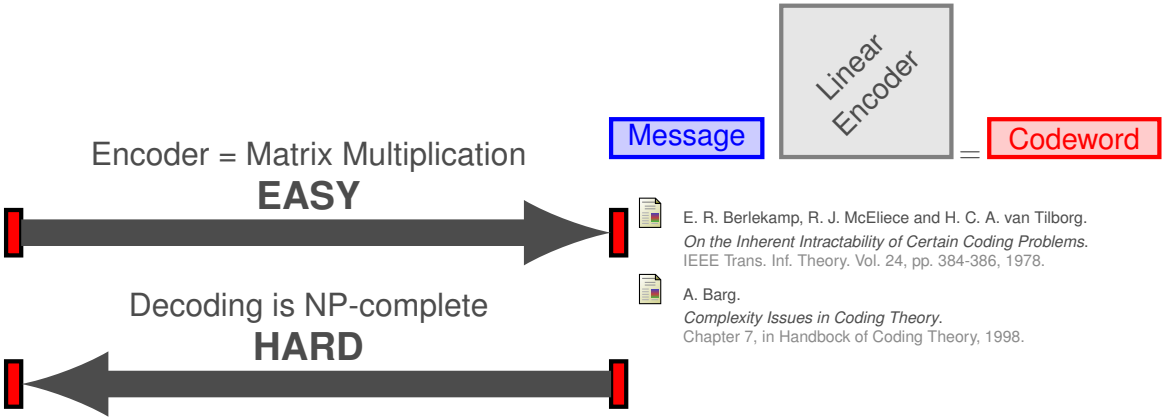
Message

Linear Encoder

$=$

Codeword

Trapdoor one-way functions - Decoder



E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg.
On the Inherent Intractability of Certain Coding Problems.
IEEE Trans. Inf. Theory. Vol. 24, pp. 384-386, 1978.

A. Barg.
Complexity Issues in Coding Theory.
Chapter 7, in Handbock of Coding Theory, 1998.

Trapdoor one-way functions - Decoder

Encoder = Matrix Multiplication

EASY

Decoding is NP-complete

HARD

Efficient decoder for certain families of codes

EASY(with TRAPDOOR information)

E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg.
On the Inherent Intractability of Certain Coding Problems.
IEEE Trans. Inf. Theory. Vol. 24, pp. 384-386, 1978.

A. Barg.
Complexity Issues in Coding Theory.
Chapter 7, in Handbock of Coding Theory, 1998.

1. Error-Correcting Codes and Cryptography

1. Introduction I - Cryptography
2. Introduction II - Coding Theory
3. Encoding (Linear Transformation)
4. Parity Checking
5. Error Correcting Capacity
6. Decoding (A Difficult Problem)
7. **Reed-Solomon Codes**
8. Goppa Codes
9. McEliece Cryptosystem