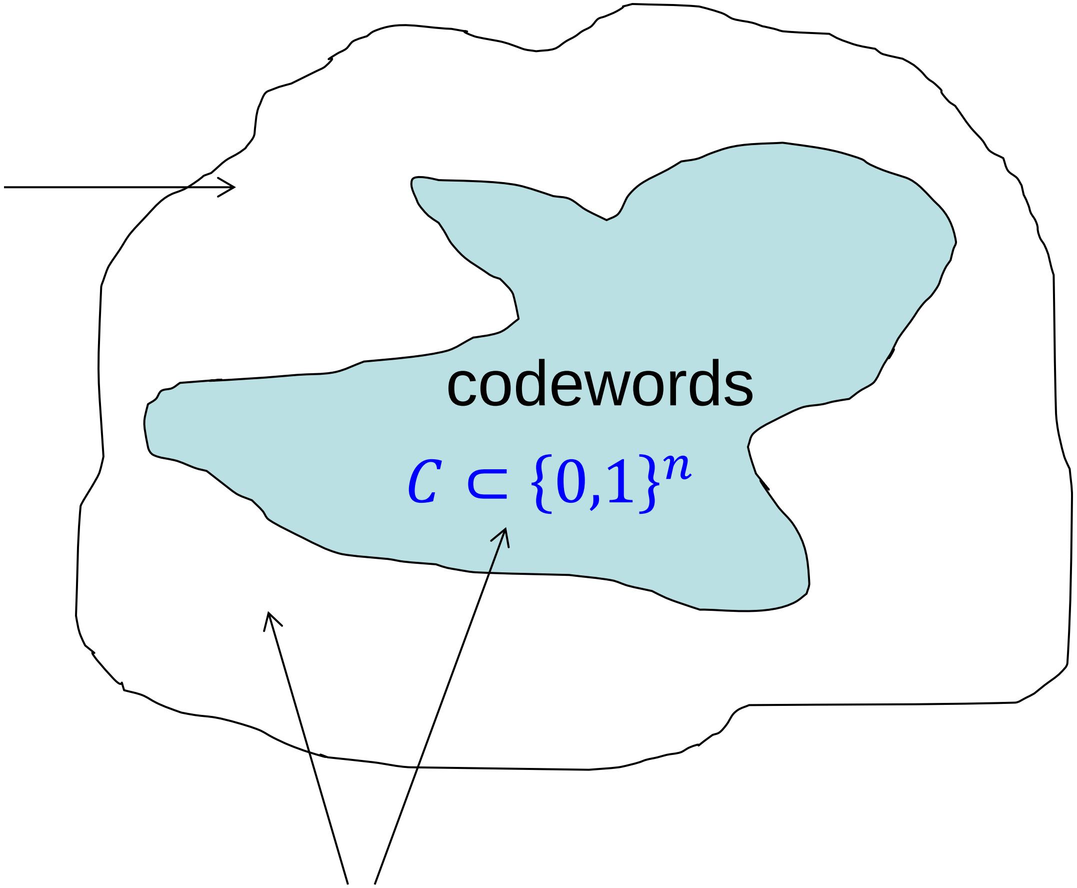
Seminar on Coding for Non-Volatile Memories

236803/048704 – CS/EE Departments, Technion

LINEAR CODES

Codes Need Structure



all words

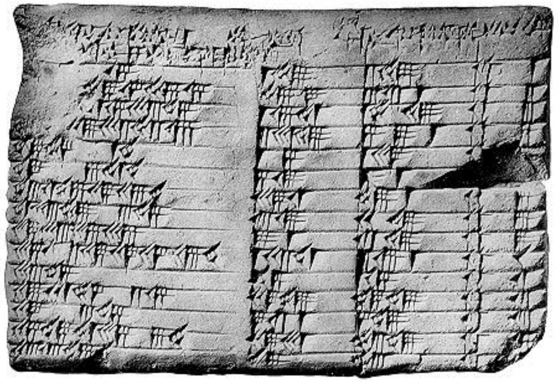
$\{0,1\}^n$

codewords

$C \subset \{0,1\}^n$

Exponential size, need structure!

A 2800 Year Leap



Babylon, 1800BC
Unary: length=O(n)

7

1,000,000

1000AD
Decimal / Positional:
length=O(log n)

Block Codes

$$C = \left\{\underline{c_1}, \underline{c_2}, \ldots, \underline{c_M}\right\}, \qquad \underline{c_i} \in Q^n$$

- Q is the code alphabet
  - $|Q|$ = 2, 4, 8 (SLC, MLC, TLC)

Linear Block Codes

$$C = \left\{\underline{c_1}, \underline{c_2}, \dots, \underline{c_M}\right\}, \qquad \underline{c_i} \in F^n$$

↑

- The code alphabet F is a field
- F is usually a finite field
- In binary codes $F = \{0,1\}$

A linear C satisfies:

1) If $\underline{c_i}, \underline{c_j} \in C$ then $\underline{c_i} + \underline{c_j} \in C$
2) If $\underline{c_i} \in C$ then $\alpha \underline{c_i} \in C$, for all $\alpha \in F$

=> C is a vector (sub)space in F^n

Observation: every linear code contains the all-zero codeword

Linear Codes – Dimension

Definition – code dimension:

The code dimension k is defined as the minimal number of vectors in a basis that spans the entire C.

$Basis = \{\underline{c_1}, \underline{c_2}, \dots, \underline{c_k}\} \subset C$, such that

for every $\underline{c} \in C$

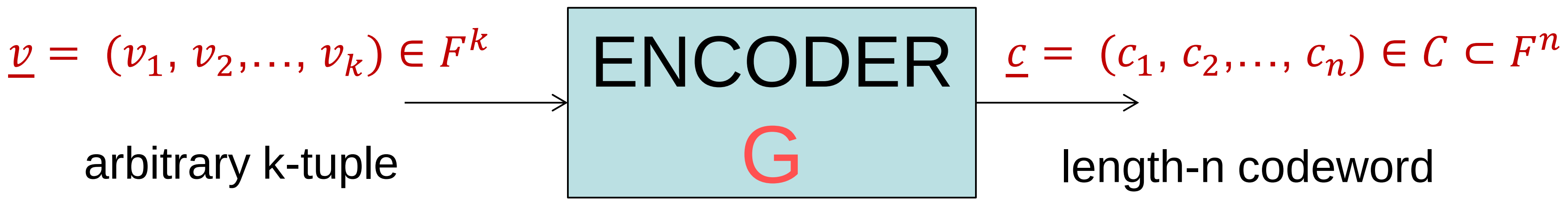
there exist $\alpha_1, \alpha_2, \dots, \alpha_k \in F$ such that

$$\underline{c} = \alpha_1 \underline{c_1} + \alpha_2 \underline{c_2} + \cdots + \alpha_k \underline{c_k}$$

Code Rate

- $R \overset{\text{def}}{=} \frac{log_{|F|}(M)}{n} = \frac{log_{|F|}\left(|F|^k\right)}{n} = \frac{k}{n}$

- Encoder:



$\underline{v} = \ (v_1, \, v_2, \ldots, \, v_k) \in F^k$

arbitrary k-tuple

ENCODER
G

$\underline{c} = \ (c_1, \, c_2, \ldots, \, c_n) \in C \subset F^n$

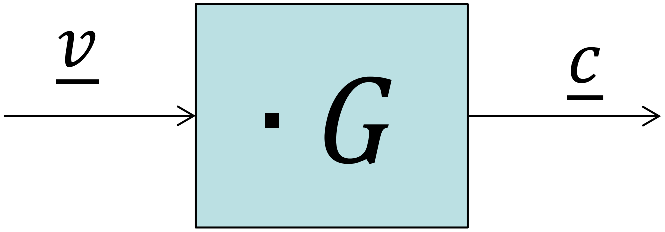
length-n codeword

Generator Matrix

$$G = \begin{bmatrix} \rule[0.5ex]{4em}{0.4pt} & \underline{c_1} & \rule[0.5ex]{4em}{0.4pt} \\ \rule[0.5ex]{4em}{0.4pt} & \underline{c_2} & \rule[0.5ex]{4em}{0.4pt} \\ & \vdots & \\ \rule[0.5ex]{4em}{0.4pt} & \underline{c_k} & \rule[0.5ex]{4em}{0.4pt} \end{bmatrix}$$

- G_{kxn} over F
- Row vectors of G specify a basis for C
- Encoding:

$$\underset{1xn}{\underline{c}} = \underset{1xk}{\underline{v}} \cdot \underset{kxn}{G} \qquad \underline{c} = (c_1, c_2, \ldots, c_n) \in C \subset F^n$$



Generator Matrix

- Examples
  - Repetition code k=1
  - Single parity-check code k=n-1

Hamming Distance of Linear Codes

Theorem: Hamming distance ⇔ Hamming weight:

- The minimum **Hamming distance** between two codewords in a linear code is equal to the minimum **Hamming weight** of a codeword.

Parity-Check Matrix

$$H = \begin{bmatrix} \text{———} & \underline{h_1} & \text{———} \\ \text{———} & \underline{h_2} & \text{———} \\ & \vdots & \\ \text{———} & \underline{h_{n-k}} & \text{———} \end{bmatrix}$$

- $H_{(n-k)xn}$ over F
- Row vectors of H specify parity constraints
- Each codeword $\underline{c}$ satisfies:

$$\underset{(n-k)xn}{H} \cdot \underset{nx1}{\underline{c}^T} = \underset{(n-k)x1}{\underline{0}^T} \qquad \underline{0} \triangleq (0,0,\dots,0)$$

- Algebraically: $C = Ker(H)$

G vs. H

$\underline{v} \in F^k$

$\underline{y} \in F^n$

Systematic Codes

k information bits	n-k check bits

- Systematic G

$$G = \left[\begin{array}{c|c} I_k & A \end{array}\right]$$

- Systematic H

$$H = \left[\begin{array}{c|c} -A^T & I_{n-k} \end{array}\right]$$

- Systematic G ←→ Systematic H

[n,k] Linear Codes - Review

Generator Matrix

$$G = \begin{bmatrix} \underline{c_1} \\ \underline{c_2} \\ \vdots \\ \underline{c_k} \end{bmatrix}$$

$$\boxed{\underline{c} = \underline{v} \cdot G}$$

Parity-check Matrix

$$H = \begin{bmatrix} \underline{h_1} \\ \underline{h_2} \\ \vdots \\ \underline{h_{n-k}} \end{bmatrix}$$

$$\boxed{H \cdot \underline{c}^T = \underline{0}^T}$$

$$\boxed{HG^T = [0]_{(n-k)xk}}$$

Dual Code

Definition – dual code:

C is an [n,k] linear code with generator matrix G.
The dual code of C , denoted $C^{\perp}$, is the dual subspace to the code subspace.

$$C^{\perp} = \{\underline{x} : G \cdot \underline{x}^T = \underline{0}^T\}$$

- In other words:

G is a parity-check matrix for $C^{\perp}$.

What is the dimension of $C^{\perp}$?

Hamming Codes

- For any integer m, there exists a binary Hamming code with parameters:
  - $n=2^m-1$
  - $n-k=m$
- Example: m=3

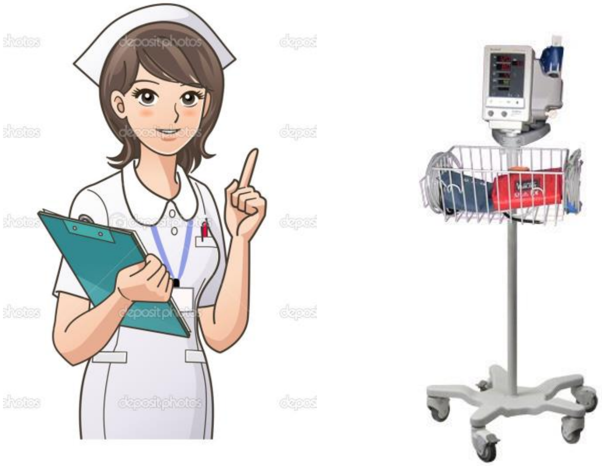
$$H = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Syndrome Decoding

Syndrome:

Multiply noisy codeword by H

1) If 0, release home
2) If not 0, send to “doctor”



Decoder:

Find errors in noisy codeword

Standard Array

Vector space: F^n
Subspace: C

all F^n

$$
\begin{array}{ccccc|l}
c_0 & c_1 & c_2 & \cdots & c_{2^k-1} & \leftarrow \text{the code} \\
e_1 + c_0 & e_1 + c_1 & e_1 + c_2 & \cdots & e_1 + c_{2^k-1} & \leftarrow \text{coset } e_1 \\
e_2 + c_0 & e_2 + c_1 & e_2 + c_2 & \cdots & e_2 + c_{2^k-1} & \leftarrow \text{coset } e_2 \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
e_{2^{n\text{-}k}-1} + c_0 & e_{2^{n\text{-}k}-1} + c_1 & e_{2^{n\text{-}k}-1} + c_2 & & e_{2^{n\text{-}k}-1} + c_{2^k-1} &
\end{array}
$$

Coset Leaders

Definition – coset leader:

The leader of coset e_i is the element in row i with the smallest Hamming weight (arbitrary tie breaking)

$$\begin{array}{ccccc}
\boxed{c_0} & c_1 & c_2 & \cdots & c_{2^k-1} \\
e_1+c_0 & e_1+c_1 & \boxed{e_1+c_2} & \cdots & e_1+c_{2^k-1} \\
e_2+c_0 & \boxed{e_2+c_1} & e_2+c_2 & \cdots & e_2+c_{2^k-1} \\
\vdots & \vdots & \vdots & & \vdots \\
e_{2^{n\text{-}k}-1}+c_0 & e_{2^{n\text{-}k}-1}+c_1 & e_{2^{n\text{-}k}-1}+c_2 & & e_{2^{n\text{-}k}-1}+c_{2^k-1}
\end{array}$$

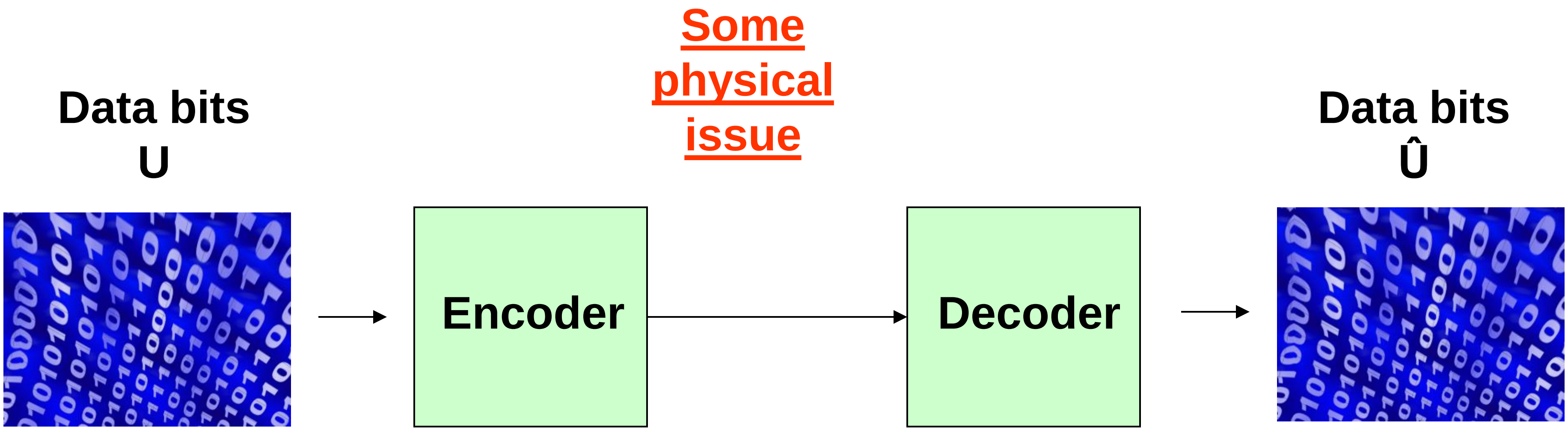
Standard-Array Decoding (ML)

1. Calculate syndrome $H \cdot \underline{y}^T = \underline{\tilde{s}}$
2. Find the coset leader of row $\underline{\tilde{s}}$

$$\begin{array}{ccccccc}
c_0 & c_1 & c_2 & \cdots & c_{2^k-1} & & \\
e_1+c_0 & e_1+c_1 & e_1+c_2 & \cdots & e_1+c_{2^k-1} & & \\
e_2+c_0 & \boxed{e_2+c_1} & e_2+c_2 & \cdots & e_2+c_{2^k-1} & \longleftarrow & \text{coset } \underline{\tilde{s}} \\
\vdots & \vdots & \vdots & & \vdots & & \\
e_{2^{n\text{-}k}-1}+c_0 & e_{2^{n\text{-}k}-1}+c_1 & e_{2^{n\text{-}k}-1}+c_2 & & e_{2^{n\text{-}k}-1}+c_{2^k-1} & &
\end{array}$$

3. Set $\underline{\hat{e}}$ to be the coset leader and output $\underline{\hat{c}} = \underline{y} - \underline{\hat{e}}$

Coding: The General Problem



Data bits
U

Encoder

Some physical issue

Decoder

Data bits
Û

Find good codes that overcome physical issue and give U = Û

The General Coding Method

	Example 1
1. Physical issue	**1. Channel with 0→1 and 1→0**
2. Error model	**2. At most t 0→1 and 1→0 combined**
3. Sufficient + necessary conditions	**3. $D_H(x,y)>2t$ for all x,y in C (necessary and sufficient)**
4. Code constructions vs. upper bounds	**4. Hamming code vs. Hamming bound**
5. Decoding	**5. Find codewords at Hamming distance t or less from received word**

The General Problem

Examples 2,3,4,…

1. **Physical impairment**
2. **Error model**
3. **Sufficient + necessary conditions**
4. **Code constructions vs. upper bounds**
5. **Decoding**

Coding for NVMs