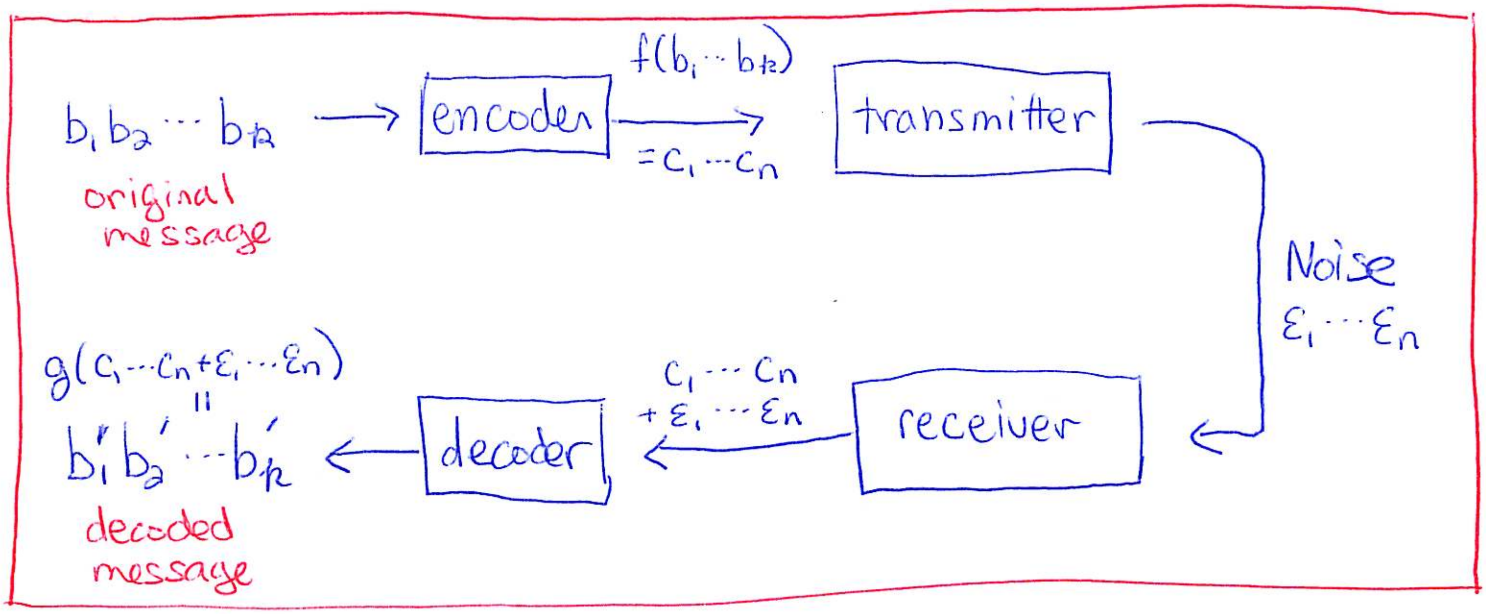
Encoding / Decoding framework

$0 < k \leq n$

Message space $\{0,1\}^k = \{b_1 \cdots b_k \mid b_i \in \mathbb{Z}_2\}$

Codeword space $\{0,1\}^n = \{c_1 \cdots c_n \mid c_i \in \mathbb{Z}_2\}$



$\boxed{\text{encoder}}$ $f : \{0,1\}^k \to \{0,1\}^n$ is a 1-1 function that typically adds redundancy

$\boxed{\text{Noise}}$ a model assumption is made, such as

(i) at most one error, $\varepsilon_1 \cdots \varepsilon_n = 0 \cdots 010 \cdots 0$

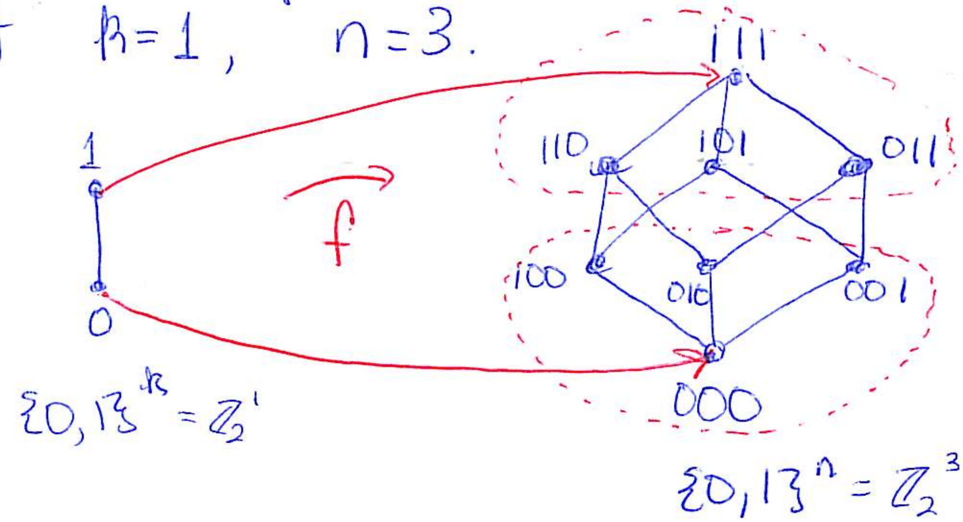
(ii) small independent chance of error in each position.

$\boxed{\text{decoder}}$ a well-defined function

$g : \{0,1\}^n \to \{0,1\}^k \cup$ "resend" $\cup$ "unrecoverable failure"

Example a repetition code.

set $k=1$, $n=3$.



$f(0) = 000$, $f(1) = 111$.

decoding function depends on noise assumption.

assumption 1 ≤ 1 error.

$g(000) = g(100) = g(010) = g(001) = 0$

$g(111) = g(110) = g(101) = g(011) = 1$

assumption 2 ≤ 2 errors

$g(000) = 0$ otherwise, $g(b_1 b_2 b_3) =$ "resend"

$g(111) = 1$

assumption 3 probability of error in any bit is independent with probability .05.

Using g from assumption 1, what is the probability of successful decoding?

Linear Transformation encoders

Any matrix over a field provides an encoding function.

Ex $\quad M: \{0,1\}^k \to \{0,1\}^n$

$$bM = c$$

where M is a $k \times n$ matrix over $\mathbb{Z}_2$.

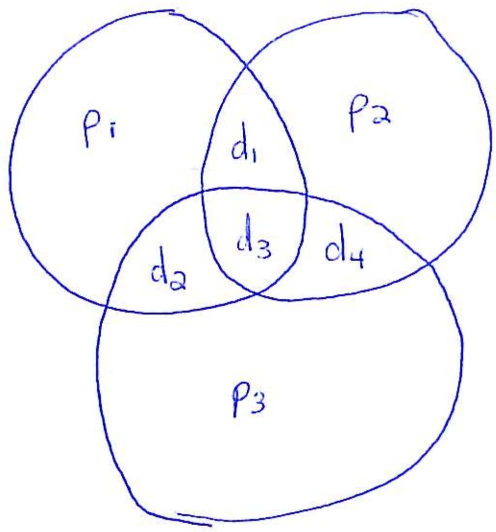
Ex $\quad G: \{0,1\}^4 \to \{0,1\}^7$

$$G = \begin{array}{c} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \\ \begin{array}{ccccccc} d1 & d2 & d3 & d4 & p1 & p2 & p3 \end{array} \end{array}$$

idea: "data" bits are d_1, d_2, d_3, d_4

"redundancy" parity check bits are p_1, p_2, p_3

$b = b_1 b_2 b_3 b_4$	$bG = c_1 c_2 \cdots c_7$ data	parity
0000	0000	000
0001	0001	011
0010	0010	111
0100	0100	101
1000	1000	110
1001	1001	101
1010	1010	001



Venn diagram of G

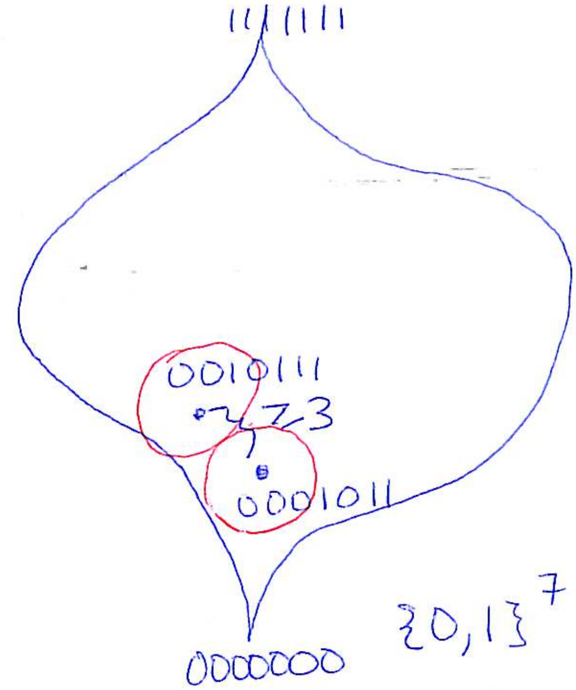
Parity bit p_i is assigned to the data bits in its circle, so that sum $=0$.

Every data bit is covered by
(1) At least 2 parity bits
(2) a unique subset of parity bits.

Up to 1 error can be corrected
Up to 2 errors can be detected.

Hypercube viewpoint

$$\underbrace{\{0,1\}^4}_{\text{message space}} \xrightarrow{f}$$



$\{0,1\}^7$

codeword space

0 errors codeword: 0001011

1 error:
- $\underline{1}001011$
- $0\underline{1}01011$
- $00\underline{1}1011$
- $000\underline{0}011$
- $0001\underline{1}11$
- $00010\underline{0}1$
- $000101\underline{0}$

Radius 1 Hamming ball in $\{0,1\}^7$ with center 0001011

Definition (Linear Code)

An (n, k) linear code over a finite field F is a k-dimensional subspace V of the vector space

$$F^n = \underbrace{F \oplus F \oplus \cdots \oplus F}_{n \text{ copies}}$$

over F. Elements of V are called codewords. When $F = \mathbb{Z}_2$, V is a binary code.

(n, k) linear code $\xleftrightarrow[\text{basis of } V]{\text{via}}$ generator matrix $G : F^k \to F^n$

$$\begin{array}{cc} & \begin{array}{cc} \text{message part} & \text{parity checks} \\ \downarrow & \downarrow \end{array} \\ G = & \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] = \left[I_{k \times k} \mid A_{k \times n-k} \right] \end{array}$$

Generator matrix of a $(7, 4)$ linear code.

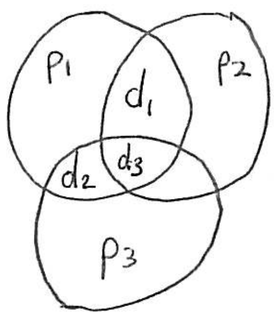
Parity-Check Matrix Decoding (PCMD)

$$G = \left[I_k \mid A_{k\times n-k} \right] \Leftrightarrow H = \left[\begin{array}{c} -A_{k\times n-k} \\ \hline I_{n-k} \end{array} \right]$$

generator matrix — parity-check matrix

Ex

$$G = \begin{array}{c} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array} \right] \\ \begin{array}{cccccc} d_1 & d_2 & d_3 & p_1 & p_2 & p_3 \end{array} \end{array} \qquad H = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right]$$



PCMD algorithm

1. received word $= w$. Find wH.
2. $wH = 0 \Rightarrow$ decode as w itself.
3. $wH = i^{th}$ row of H (and no other) $\Rightarrow$ decode as $w + e_i$.
4. Otherwise ≥ 2 errors. Do not decode.

$$wH = [q_1 \; q_2 \; \cdots \; q_{n-k}]$$

$q_i = 0$ iff i^{th} parity relation holds for w.

Exercise Decode 101011, 111010, 110000

$[101011]H = [010] = 5^{th}$ row of H, so decode as 101001.

$[111010]H = [110] = 1^{st}$ row of H, so decode as 011010

$[110000]H = [011]$ not a row of H. ≥ 2 errors; do not decode.

Lemma (Orthogonality Relation)

Let C be an (n,k) linear code over F with generator matrix G and parity-check matrix H. Then, for any $v \in F^n$,

$$vH = 0 \quad \text{iff} \quad v \in C.$$

Proof $\quad$ H is $n \times (n-k)$. H contains I_{n-k} and so

$$\begin{aligned} \text{rank}(H) + \dim(\ker H) &= n \\ (n-k) + k &= n. \end{aligned}$$

$\dim C = k$, so we just show $C \subseteq \ker H$.
Let $v \in C$, and let m satisfy $v = mG$.
Then

$$\begin{aligned} vH &= mGH \\ &= m[I_k \mid A]\left[\frac{-A}{I_{n-k}}\right] \\ &= m(I_k(-A) + AI_{n-k}) = m \cdot 0 = 0. \end{aligned}$$

$\square$

Exercise Given $r \geq 2$ parity bits, what is the maximum number of data bits

$$C = \{d_1 \cdots d_k \; p_1 \cdots p_r\}$$

in a $(k+r, k)$ linear code for which 1 error can be corrected?

(Hint: Look at rows of H.)

Theorem 31.3 (Parity-check Matrix Decoding)

PCMD will correct any single error iff rows of H are nonzero and no two rows are dependent.

Proof (binary case) independent rows $\Leftrightarrow$ distinct rows

($\Leftarrow$) Assume rows of H are nonzero and distinct.
Assume w is transmitted but received as $w + e_i$.

$$(w+e_i)H = wH + e_iH = e_iH \quad \text{(Orthog. Lemma)}$$
$$= i^{th} \text{ row of } H.$$

The i^{th} row is distinct, and so the error is identified.

($\Rightarrow$) No row of H can be the zero row. Otherwise an error-free codeword w yields $wH = 0$, and an error is reported in the position of the 0 rows of H.
No two rows $i \neq j$ of H are the same. Otherwise if w is a codeword received as $w + e_i$,

$$(w+e_i)H = wH + e_iH = i^{th} \text{ row of } H$$
$$= j^{th} \text{ row of } H.$$

The decoding algorithm reports 2 errors and does not decode. ∎