Intro to ECC

ECC: protect data from noise, using judicious redundancy to enable recovery from corruption

Noise types: Erasure, Error, Deletion

Noise Model: → Random (Stochastic) [Shannon]
→ Worst Case (adversarial) [Hamming]

Alphabet → Bit packet
→ bits $\{0,1\}$

Parity Check Code:

$$PC_n = \{ (c_1, c_2, \dots, c_n) \in \{0,1\}^n \mid \oplus c_i = 0 \}$$

* PC_n is optimal for 2 erasure.

Fact: $C \subseteq \Sigma^n$ of distance d can correct $(d-1)$ erasures.

Hamming Code:

It can correct deletion (can mod $n+1$)

$$VT_n = \{ (c_1, \dots, c_n) \in \{0,1\}^n \mid \sum i c_i \equiv 0 \ [2n+1] \}$$

$$\Rightarrow H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{for } n=7$$

. (as $H \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} =$ (sum $\sum i c_i$ in binary))

* Let C be a code that's linear, H its parity check matrix, then $C^\perp = \{ H^T x \mid x \in \mathbb{F}^n \}$

Simplex Code: it is optimal

$$S_r = a \in \mathbb{F}_2^n \to \langle a, x \rangle_{x \in \mathbb{F}_2^n \setminus \{0\}}$$

r bits → $2^r - 1$ bits

its generator matrix is H^T. (for H the par-check matrix of Hamming code)

Hamming code $\xleftrightarrow{\text{dual}}$ Simplex Code

+Great rate

$[n=2^r-1, k=2^r-1-r, 3]$

-Poor distance

+Optimal trade off

(Meets Hamming bound)

$[n=2^r-1, k=r, d=2^{r-1}]$

-Great distance

-Very poor rate (k/n)

+Optimal trade-off

(meets Plotkin bound)

$$\left(\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\dots+\binom{n}{\lfloor\frac{d-1}{2}\rfloor}\right)|C| \leq 2^n$$

(OUTER BOUND)

$C \subset \{0,1\}^n$, $d > \frac{n}{2}$

Then $|C| \leq \frac{2d}{2d-n} \leq 2d$

Asymptotically good codes:

Code $[n,k,d]_2$

Rate $R = \frac{k}{n} \in [0,1]$ Relative distance $\delta = \frac{d}{n} \in [0,1]$

Hamming code: $\delta \to 0$ as $n \to +\infty$ $(R \to 1)$

Simplex code: $R \to 0$ as $n \to +\infty$ $(\delta \to 1/2)$

Constant code family $\{C_1, C_2, \dots, C_r, \dots\}$
$n_1, n_2, \dots, n_r, \dots$

Does there exist?

s.t R_i, δ_i are bounded away from 0.

Singleton Bound: $k \leq n-d+1 \Rightarrow R+\delta \leq 1+o(1)$

proof by PHP

Reed Solomon Codes

$[n,k,d]_{\mathbb{F}_q}$

$(m_0, m_1, \dots, m_{k-1}) \in \mathbb{F}_q^k \to M(X) = m_0 + m_1 X + m_2 X^2 + \dots + m_{k-1} X^{k-1} \in \mathbb{F}_2[X] \to (M(\alpha_1), M(\alpha_2), \dots, M(\alpha_n)) \in \mathbb{F}_q^n$

$\{\alpha_1, \alpha_2, \dots \alpha_n\} \subseteq \mathbb{F}_q$

Hence the gen matrix is the Vandermonde Matrix:

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{k-1} \\ & & \vdots & & \\ 1 & \alpha_n & \dots & & \alpha_n^{k-1} \end{pmatrix}$$

Obve, requires $q \geq n$

Hence it is better to work on bit packets.

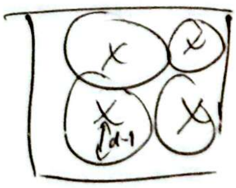
Dist of RS Codes:

As $|\{\alpha \mid M(\alpha) = 0\}| \leq k-1$ as $\deg(M) \leq k-1$, & as the code is linear

then $w(c) \geq n-(k-1) = n-k+1$, $\forall c \in C$.

And thus we can correct $\frac{n-k}{2}$ errors.

Existence of asym good codes. $\boxed{d = \delta n}$

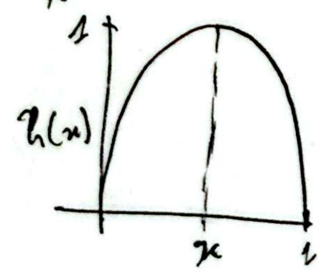


Facts $\exists$ a code of dist d & size $\frac{2^n}{Vol(n,d-1)} \geq 2^{n - h(\delta)n}$

Facts $Vol(n, \delta n) = \sum_{i=0}^{\delta n} \binom{n}{i} \leq 2^{h(\delta)n}$

$2^{h(\delta)n} \geq \binom{n}{\delta n} \geq 2^{h(\delta)n - o(n)}$

$h : [0;1] \to [0;1]$



$$h(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$$

Corollary: $\forall \delta \in [0; \frac{1}{2})$

$\exists$ binary code family of rel dist δ and rate $R \geq R_{GV}(\delta) = 1 - h(\delta)$

↳ Gilbert-Varshamov bound

$\Rightarrow \exists$ rate $\frac{1}{2}$ code family of rel dist ≥ 0.1.

Hamming bds

Rate must be $\leq R_{Hamm}(\delta) = 1 - h(\frac{\delta}{2})$

Is there good binary linear codes

In Parity check matrix H:

Dim of code $= n - m = n - h(\delta)n$

dist $d \Leftrightarrow$ every $(d-1)$ subset of columns are lin indep

If $2^m > \sum_{d=0}^{d-2} \binom{n}{d}$ we can build

RS code (Binary Version):

$M(x) \to (M(\alpha_1), M(\alpha_2), \dots, M(\alpha_n)) \in \mathbb{F}_{2^t}^n$

$\downarrow g(M(\alpha_i))$

for $g : \mathbb{F}_{2^t} \to \mathbb{F}_2^t$ linear 1-to-1

Rate $= \frac{tk}{tn} = \frac{k}{n}$, same as RS

Rel distance:

distance $\geq n - k + 1$

r.distance $\geq \frac{n-k+1}{tn} \leq \frac{1}{t} = \frac{1}{\log_2 n} \to 0$ as $n \to \infty$:/

We can fix RD but decrease the rate:

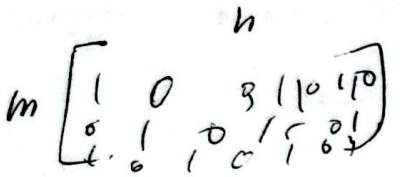
Replace g by $g : \mathbb{F}_{2^t} \to \mathbb{F}_2^{2t}$, linear injective, a $[2t, t]$ code, with good rel. distance, say > 0.

$\Rightarrow$ dist $\geq (n-k+1) * (0.1)2t = 0.2t(n-k+1) \Rightarrow$ rel dist $\geq 0.2 \frac{n-k+1}{n} \geq 0.2$.

Rate $= \frac{tk}{2tn} = \frac{k}{2n}$.

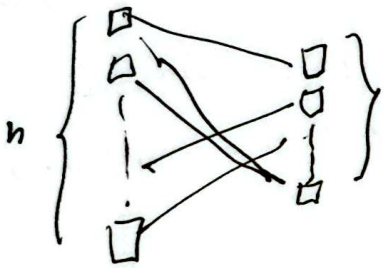
This is "bruteforceable" as t is small enough ($t = \log_2 n$).

Low Density Parity Check (LDPC) Codes:



very few 1's in each row and column
(⇒ sparse parity check matrix)

Graph view of parity checks (Bipartite):



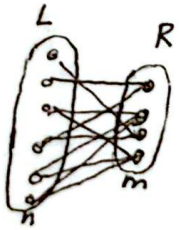
"Factor graph" s.t the parity check matrix is its adj. matrix.

LDPC Code ⇔ Factor graph has low degree (const degree)

Def:

Code defined by G, a bipt graph

$[C(G) \subseteq \mathbb{F}_2^n]$



$$C(G) \triangleq \{(c_1, c_2, \ldots, c_n) \in \{0,1\}^n \mid \forall j \in R, \sum_{i \in N(j)} c_i = 0 [2]\}$$

Rate of $C(G) \geq \frac{n-m}{n}$

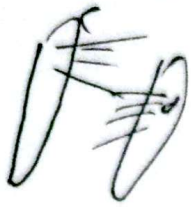
Obs If $\forall S \subseteq L$, $|S| < \delta n$ then is a good test has exactly one neighbor in S, then the distance of $C(G)$ is at least δn.

Magic Expanders graphs

A $n \times m$ bipartite graph is said to be a δ-unique nbr expander

If $\forall S \subset L$, $\exists$ a node $r \in R$ with exactly one nbr in S.
$|S| \leq \delta n$

Unique nbr expanders:



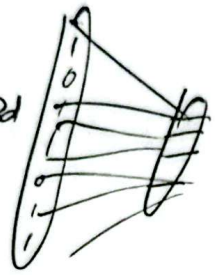
A random (c,d) biregular graph will be a δ-unique nbr expander for some $\delta > 0$ w/ high probability

Also constructions of such graphs are known.

Decoding Algo:

While $\exists$ codeword with #satisfied < #unsatisfied checks, flip that bit.

$(O(nm))$



Thm: If the graph is a sufficiently good expander, the iterative algo will correct γn errors for some γ bd away from 0.

Locally Repairable Codes (LRC) started around 2010

Theorem: "Bad news about RS codes"

In a RS code, the knowledge of any $k-1$ codeword symbols reveals no information about any other symbols.

How to bound locality? Sparse parity checks!

efficient repair = "local" repair

recover c_i from symbols c_j's for every few j's.

Definition (LRC): An $[n,k]_2$ linear code C is an (r,d) locally repairable code (LRC) if following hold:

① C has distance $\geq d$

② $\forall i \in \{1,2,\dots,n\}$ there is a parity check with support $\{i\} \cup R_i$ for some $R_i \subseteq \{1,\dots,n\} \setminus \{i\}$, $|R_i| \leq r$

$$c_i + \sum_{j \in R_i} \lambda_i c_j = 0$$

$\boxed{r=1}$ Repetition code $\begin{cases} c_1 = c_2 \\ c_3 = c_4 \\ \vdots \\ c_{n-1} = c_n \end{cases}$

But $d=2$.

We can improve with: | RS Code | RS Code |

Repetition Code	RS code
Great Locality	Great distance
Terrible distance	Terrible Locality

A Singleton type order bound: Theorem:

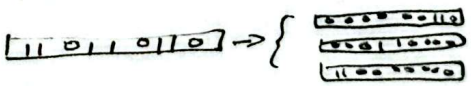
For an $[n,k]_q$ code that is an (r,d)-LRC, then $d \leq n-k+1-\underbrace{\left\lfloor \frac{k-1}{r} \right\rfloor}_{\text{price of locality}}$

How to combine both codes:

Take the Generating matrix in this form:

$$RS = \begin{pmatrix} I_k \\ A \end{pmatrix} \} n$$

Then break every row vector of A into vectors s.t each one has at most r non-zero entries, giving $\lceil \frac{k}{r} \rceil$ rows.

ex:  $(n' := n + \lceil \frac{k}{r} \rceil)$

This will give a new code with $d = n' - k + 1 - \left\lfloor \frac{k-1}{r} \right\rfloor$ of size n', that is an (r,d)-LRC.

Quantum Codes

qubit $\in \mathbb{C}^2$ $\leftrightarrow \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$

Digitization of errors

Two types of errors

BIT FLIP "X"

$|0\rangle \to |1\rangle$
$|1\rangle \to |0\rangle$

$X(\alpha|0\rangle + \beta|1\rangle) = \alpha|1\rangle + \beta|0\rangle$

$X = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$

Phase Flip "Z"

$|0\rangle \to |0\rangle$
$|1\rangle \to -|1\rangle$

$Z(\alpha|0\rangle + \beta|1\rangle) = \alpha|0\rangle - \beta|1\rangle$

$Z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$

$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$, $|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$

$Z|+\rangle = |-\rangle$, $Z|-\rangle = |+\rangle$

Quantum Code: $[[n,k]]$ quantum code

k qubits to encode n qubits

$\mathbb{C}^{2^k} \longrightarrow \mathbb{C}^{2^n} \cong (\mathbb{C}^2)^{\otimes n}$

2^k dimensional subspace of $\mathbb{C}^{2^n}$

Errors on multiple qubits

Only have to consider $E_1 \otimes E_2 \otimes \dots \otimes E_n$

$E_i \in \{I, X, Z\}$

Total # errors $\leq e$ $\to$ $|\{i \mid E_i \neq I\}|$

Repetition Code

~~$|\psi\rangle \to |\psi\rangle \otimes |\psi\rangle$~~ Not possible, "no-cloning"

Baby Step Detect one bit flip (one X)

$\alpha|0\rangle + \beta|1\rangle \to \alpha|00\rangle + \beta|11\rangle$

$R := \text{Code} = \text{span}\{|00\rangle, |11\rangle\} \subseteq \mathbb{C}^4$

"Stabilizer"

$(Z \otimes Z)(\alpha|00\rangle + \beta|11\rangle) = \alpha|00\rangle + \beta|11\rangle$ $[[2,1]]$ code

$Z_1Z_2 \Rightarrow \forall |\psi\rangle \in R,\ Z_1Z_2|\psi\rangle = |\psi\rangle$

$\Rightarrow$ Codeword belongs to eigenspace

$Z_1Z_2(\alpha|10\rangle + \beta|01\rangle) = -(\alpha|10\rangle + \beta|01\rangle)$

$E|\psi\rangle$

$E \in \{I\otimes I, X\otimes I, I\otimes X\}$

$|\psi\rangle|0\rangle \xrightarrow{\text{map}} \left(\frac{I\otimes I + Z_1Z_2}{2}\right)E|\psi\rangle|0\rangle + \left(\frac{I\otimes I - Z_1Z_2}{2}\right)E|\psi\rangle|1\rangle$

And then $E|\psi\rangle|0\rangle \to \begin{cases} E|\psi\rangle|0\rangle \text{ if no bitflip} \\ E|\psi\rangle|1\rangle \text{ if bitflip} \end{cases}$

$\alpha|0\rangle + \beta|1\rangle \xrightarrow{\text{encode}} \alpha|000\rangle + \beta|111\rangle$ $[[3,1]] = R_3$

$I\otimes X\otimes I(\psi) = \alpha|010\rangle + \beta|101\rangle$

Stabilizer of $|\psi\rangle \in R_3$

$$Z_1 Z_2 |\psi\rangle = |\psi\rangle$$
$$Z_2 Z_3 |\psi\rangle = |\psi\rangle$$

$$E \in \{I, X_1, X_2, X_3\}$$

Syndromes: $H = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$

$I : \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
$X_1 : \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
$X_2 : \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$X_3 : \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

} All syndroms are distinct

One phase flip

Same as bit flip but for $\{|+\rangle, |-\rangle\}$ basis

$|+\rangle \to |+\rangle \otimes |+\rangle \otimes |+\rangle = |+++\rangle$

$|-\rangle \to |---\rangle$

$P_3 = \text{Span}\{|+++\rangle, |---\rangle\}$

How to correct both bit flip and phase flip with the same code.

Shor $[[9,1]]$ code

Combine P_3 and R_3 (R_3 on top of P_3)

$$|0\rangle \to \left(\frac{|000\rangle + |111\rangle}{\sqrt{2}}\right)^{\otimes 3} = |\psi_0\rangle$$

$$|1\rangle \to \left(\frac{|000\rangle - |111\rangle}{\sqrt{2}}\right)^{\otimes 3} = |\psi_1\rangle$$

Shor Code $= \text{span}\{|\psi_0\rangle, |\psi_1\rangle\} \subseteq \mathbb{C}^{2^9}$

Measure Syndromes using **Bit Flip:**

Z_1Z_2, Z_2Z_3
Z_4Z_5, Z_5Z_6
Z_7Z_8, Z_8Z_9
$X_1X_2X_3X_4X_5X_6$, $X_1X_2X_3X_7X_8X_9$

} Stabilizers

Error	Syndrom
X_1	10 00 00 00
X_2	11 00 00 00
X_3	01 00 00 00
⋮	
etc	
X_9	00 00 00 01

syndroms are distinct ⇒ we can correct

Phase flip:

Error	Syndrom	
Z_1	10	} Same syndrom but OK (Only OK in quantum codes)
⋮		
Z_3		
Z_4	11	"degeneracy"
⋮		
Z_6		
Z_7	01	
⋮		
Z_5		

[[7,1]] Hamming based code:

$$\begin{bmatrix} 0&0&0&1&1&1&1 \\ 0&1&1&0&0&1&1 \\ 1&0&1&0&1&0&1 \end{bmatrix} \quad H_X = H_Z$$

Stabilizers: $Z_4Z_5Z_6Z_7$, $Z_2Z_3Z_6Z_7$, $Z_1Z_3Z_5Z_7$
$X_4X_5X_6X_7$, $X_2X_3X_6X_7$, $X_1X_3X_5X_7$

All syndroms for 1 bit flip are distinct, so H_Z helps correct any single X_i error, & H_X helps correct every Z_j error.

$$|0\rangle \mapsto \frac{1}{\sqrt{8}} \left(\sum_{\substack{Hc=0 \\ w(c) \equiv 0 [2]}} |c_1c_2 \cdots c_7\rangle \right)$$

$$|1\rangle \to \frac{1}{\sqrt{8}} \left(\sum_{\substack{Hc=0 \\ w(c) \equiv 1 [2]}} |c_1c_2 \cdots c_7\rangle \right)$$