Introduction to Error-correction

- Practical applications of ECC stem from the following situation:

  Alice wants to transmit Bob a string $x \in \{0,1\}^n$ on a channel that corrupts $\leq 10\%$ of bits.

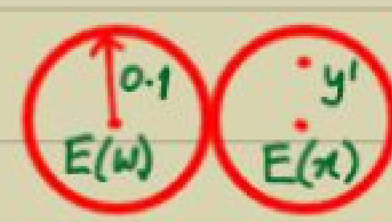
  A $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ B

  $x \in \{0,1\}^n$

  $E(x) \in \{0,1\}^m \xrightarrow{\quad y = E(x) \quad} y'$ (guarantee: $\Delta(y', y) \leq 0.1$

- If E is of distance > 0.2 then $\exists$ unique w s.t. $\Delta(E(w), y') \leq 0.1$,

  $\Rightarrow \quad w = x$.



- This motivates the design of codes with:
  - large distance δ.
  - small length m.
  - efficient encoding & decoding.

- The first two conditions get satisfied for "most" E.

__Lemma__ (Gilbert-Varshamov bound): $\forall \delta \in (0, \frac{1}{2})$ & large enough n, $\exists E : \{0,1\}^n \to \{0,1\}^m$ that is an ECC with distance δ & $m = 2n/(1-H(\delta))$, where $H(\delta) := -\delta \lg \delta - (1-\delta)\lg(1-\delta)$.

[Shannon's entropy. Eg. $0 = H(0) \leq H(\delta) \leq H(\frac{1}{2}) = 1$.]

__Proof:__

- In fact, we show that a "random" E works!
- Pick $y_1, y_2, \ldots, y_{2^n} \in \{0,1\}^m$ at random.

  Define $E : x \mapsto y_x$.
- $\forall i \neq j \in [2^n]$, $\Pr_E[\Delta(y_i, y_j) < \delta]$

  $\leq$ #($\leq \delta m$) places in y_j / # possible y_j

  $$\sum_{i=0}^{\delta m} \binom{m}{i} / 2^m \leq 0.01 \times 2^{-m(1-H(\delta))}$$

  * Stirling's approximation

$$\Rightarrow \Pr_E[\exists i \neq j, \Delta(y_i, y_j) < \delta] \leq 0.01 \times 2^{2n - m(1-H(\delta))} = 0.01$$

$$\Rightarrow \Pr_E[\forall i \neq j, \Delta(y_i, y_j) \geq \delta] > 0.99.$$

$\square$

- It can be seen in the analysis that:
  - For $\delta = 1/2$, $\exists$ code with $m = 2^{\Omega(n)}$.
  - For $\delta > 1/2$, $\nexists$ code for large n.

$\Rightarrow$ These codes might lead to <u>unique</u> decoding up to errors $< \delta/2 \leq 1/4$.

- Can we find encoding & decoding algorithms that run in poly(n)-time?

- We will study four explicit codes: linear!
  - <u>Walsh-Hadamard</u> ($\delta = 1/2$)
  - Reed-Solomon ($\delta < 1/2$ & efficient)
  - <u>Reed-Muller</u> (multivariate generalization)
  - <u>Concatenated codes</u>

- We will strengthen the notion of decoding from: <u>unique</u> $\rightarrow$ <u>local</u> $\rightarrow$ <u>list</u>.

Walsh-Hadamard code (1940s)

Defn: For $x, y \in \{0,1\}^n$ we define $x \odot y = \sum_{i=1}^{n} x_i y_i \pmod 2$. The WH code is
$$WH : \{0,1\}^n \to \{0,1\}^{m=2^n}, \quad x \mapsto z \text{ where}$$
$z_y := x \odot y$, $\forall y \in \{0,1\}^n$. (i.e. all projections of x modulo 2)

Lemma 1: WH is an ecc of distance $1/2$.

Pf:

- $\forall x \neq y$, $WH(x+y) = WH(x) + WH(y)$, where $x+y$ is the coordinate-wise sum mod 2.
- Thus, $wt(WH(x+y)) = \Delta(WH(x), WH(y)) \cdot m$.

- As $x+y \neq \bar{0}$, it is orthogonal to exactly $1/2$ of the vectors in $\{0,1\}^n$.
  $\Rightarrow \Delta(WH(x), WH(y)) = 1/2$. $\square$

– To get a shorter code we look at finite fields other than $\mathbb{F}_2$:

Reed-Solomon code (1960)

- We view the input string as a polynomial & consider all its evaluations.

Defn: Let $\mathbb{F}$ be a field & $n \le m \le |\mathbb{F}|$. RS code is

$$RS : \mathbb{F}^n \longrightarrow \mathbb{F}^m, \ (a_0, \dots, a_{n-1}) \mapsto (z_0, \dots, z_{m-1})$$

where $\forall j$, $z_j = \sum_{0 \le i < n} a_i \cdot f_j^i$ for the j-th element f_j of $\mathbb{F}$.

Lemma 2: RS is an ecc of distance $1 - \frac{n-1}{m}$.

Proof:

- $\forall a \ne b \in \mathbb{F}^n$, $RS(a-b) = RS(a) - RS(b)$, for coordinate-wise sums.
- Thus, $wt(RS(a-b)) = \Delta(RS(a), RS(b)) \cdot m$.

- As $a-b \ne \bar{0}$, $RS(a-b)$ is a set of m evaluations of a nonzero polynomial $\sum_{0 \le i < n} (a_i - b_i) x^i$.

binary-distance is $\frac{m-n+1}{m \cdot \lg |\mathbb{F}|}$

$\Rightarrow$ $< n$ of these could be zero.

$\Rightarrow$ $\Delta(RS(a), RS(b)) \geqslant \frac{m-n+1}{m}$. $\square$