Hardness Amplification

- Our goal is to construct avg-case-hard function using an E-explicit function f that is merely worst-case-hard.

Idea – • View f as a 2^n-length string & apply a map ϕ that "spreads" the hardness throughout the string.
• ϕ will be a very good error-correcting code.
• $\phi(f)$ is still E-explicit.

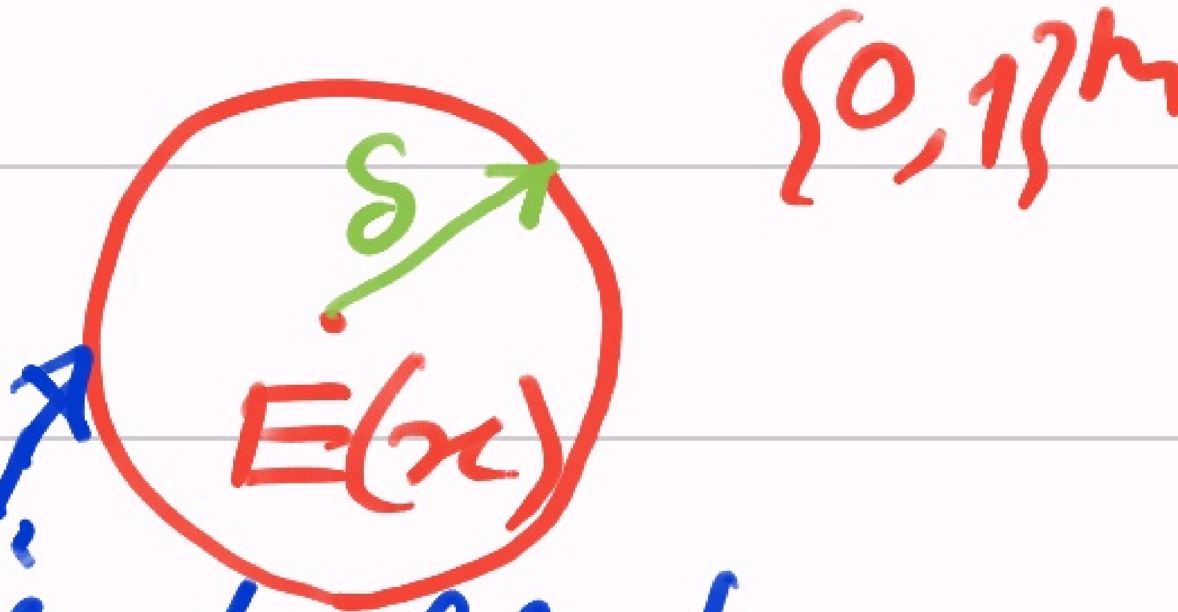
Defn: • For $x, y \in \{0,1\}^m$, the fractional Hamming distance $\Delta(x,y) := \#\{i \mid x_i \neq y_i\}/m$.

• For $\delta \in (0,1)$, function $E : \{0,1\}^n \longrightarrow \{0,1\}^m$ is an error-correcting-code (ecc) with distance δ, if $\forall x \neq y \in \{0,1\}^n$, $\Delta(E(x), E(y)) \geq \delta$.

(so 1-bit difference becomes δm-bits!)

• We call $Im(E) := \{E(x) \mid x \in \{0,1\}^n\}$ codewords.

- These have vast applications. They are used in communication channels & storage media.



this ball has no other codeword!

- For hardness amplification:

Let f be a worst-case-hard function. Let $f' := tt(f)$ be the $N := 2^n$-bit string expressing the truth-table.

Encode f' by an ecc $E : \{0,1\}^N \to \{0,1\}^{N^c}$.

Thus, $E(f')$ is a $N^c = 2^{cn}$-bit string expressing $tt(g)$, for some $g : \{0,1\}^{cn} \to \{0,1\}$.

We'll show later that if E has nice local decoding properties, then g is avg-case hard.

Also, by encoding of E: $f \in Dtime(2^{O(n)})$
$\Rightarrow g \in$ " .

Introduction to Error-Correction

- Practical applications of ecc stem from the example:

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

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

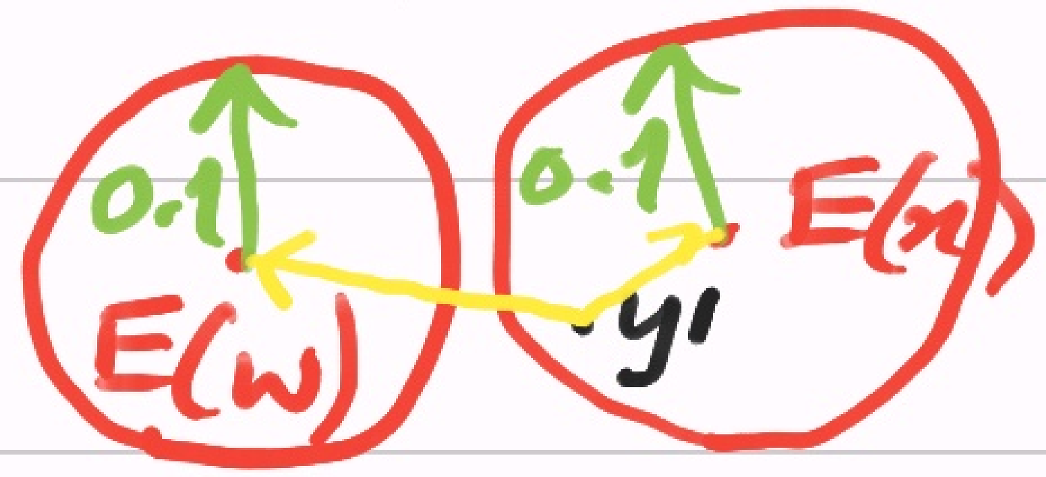
$x \in \{0,1\}^n$

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

$\triangleright$ E has distance $> 0.2 \Rightarrow \exists$ unique $w : \Delta(E(w), y') \leq 0.1$.

$\Rightarrow w = x$.

· Thus, y' is close to only $E(x)$.



- This motivates the design of codes with:
    - large distance δ. $[\delta \leq 1/2]$
    - small length m.
    - efficient encoding & decoding.

- We show that random E satisfies the first two!
  (though encoding takes $> 2^n$ time.)

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

[$H(\delta)$ is Shannon's Entropy function.]

[$0 = H(0) \leq H(\delta) \leq H(\tfrac{1}{2}) = 1$.]

<u>Proof</u>: <u>Idea</u> – Define E by picking random images!

- Pick $y_1, \dots, y_{2^n} \in \{0,1\}^m$ at random.

  Define $E : x \mapsto y_x$.

- $\forall\, i \neq j \in [2^n]$, $\Pr_E[\Delta(y_i, y_j) < \delta] \leq \dfrac{\#(\leq \delta m)\text{-places in } y_i}{\#\text{ possible } y_j}$ (bad event)

$= \binom{m+1}{\delta m+1} / 2^m \leq 0.01 \times 2^{m \cdot H(\delta)} / 2^m$ (Stirling's approximation)

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

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

- Moreover, in the analysis :
  - For $\delta = 1/2$: $\binom{m+1}{\delta m+1} / 2^m \approx \frac{1}{\sqrt{m}}$. Thus, $m \approx 2^{4n}$ required in the proof !
  - For $\delta > 1/2$: E cannot exist, for any m . (Exercise)

▷ This code E allows unique decoding up to # errors $< \delta/2 \approx 1/4$.

Qn: Can encoding/decoding be done in randomized poly(n)-time?

- We will study 4 explicit codes (linear) :

- Walsh-Hadamard ($\delta = 1/2$)
- Reed-Solomon ($\delta < 1/2$ & efficient)
- Reed-Muller (multivariate version of RS)
- Concatenated codes (binary & efficient)

- We'll strengthen the notion of decoding gradually:
  Unique $\longrightarrow$ local $\longrightarrow$ list (i.e. non-unique!)

Walsh-Hadamard Code (1940s)

- **Defn:** • For $x, y \in \{0,1\}^n$ define $x \odot y = \sum_{i=1}^{n} x_i y_i \bmod 2$.
  • WH-code is $WH : \{0,1\}^n \to \{0,1\}^{2^n =: m}$; $x \mapsto z$ where bit $z_y := x \odot y$, for location $y \in \{0,1\}^n$.
  (I.e. all possible projections $x \bmod 2$)

Lemma 1: WH is an ecc with distance 1/2.

Proof: • Note: $WH(x+y) = WH(x) + WH(y)$, where '+' is coordinate-wise sum mod 2.
(as, $\odot$ is bilinear)

$\Rightarrow \quad wt(WH(x+y)) = wt(WH(x)+WH(y)) = \Delta(WH(x), WH(y)) \cdot m.$

* = # nonzero coordinates

- If $x+y \neq \bar{0}$, then $(x+y)$ is orthogonal to exactly $1/2$ of the vectors in $\{0,1\}^n$. (Why?)

$\Rightarrow \quad wt(WH(x+y)) = m/2$

$\Rightarrow \quad \Delta(WH(x), WH(y)) = 1/2$, for $x \neq y$. $\square$

- WH achieves max. distance, but $m = 2^n$.
- To get a <u>shorter</u> code we'll use more algebra: Finite field $\mathbb{F}$ (other than $\mathbb{F}_2$).

Reed-Solomon Code (1960)

Idea - View the string as a polynomial & consider its evaluations (in a finite field).

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

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

where $\forall j$, $z_j = \sum_{i=0}^{n-1} a_i \cdot f_j^i$; for the j-th element f_j in $\mathbb{F}$.

poly. is $\left(\sum_{i<n} a_i x^i\right)$.

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

on $\mathbb{F}$-alphabet.

Pf: • Again, $RS(a-b) = RS(a) - RS(b)$; for coordinatewise difference/sum.

$$\Rightarrow \mathrm{wt}(RS(a-b)) = \mathrm{wt}(RS(a)-RS(b))$$
$$= \Delta(RS(a), RS(b)) \cdot m \, .$$ (Δ for $\mathbb{F}$ alphabet)

- If $a-b \neq \bar{0}$, then $RS(a-b)$ is a set of m evaluations of the nonzero polynomial $\sum_{i<n} (a_i - b_i) X^i$.

$\Rightarrow$ at most $(n-1)$ evaluations vanish.

$$\Rightarrow \Delta(RS(a), RS(b)) \cdot m \geq m-(n-1)$$
$$\Rightarrow \Delta(RS(a), RS(b)) \geq 1 - \frac{n-1}{m} . \qquad \square$$

↳ However, binary-distance is smaller:
$$\frac{m-(n-1)}{m \cdot \lg|\mathbb{F}|} = \frac{1}{\lg|\mathbb{F}|} \cdot \left(1 - \frac{n-1}{m}\right) .$$

Reed-Muller Code (1954)

Idea — View the string as a multivariate polynomial & consider evaluations (in a finite field).

Defn: • Let $\mathbb{F}$ be a finite field; $\ell, d \in \mathbb{N}$ & $d < |\mathbb{F}|$.

• RM-code is $RM : \mathbb{F}^{\binom{\ell+d}{d}} \longrightarrow \mathbb{F}^{|\mathbb{F}|^{\ell}}$; that maps every ℓ-variate d-deg P to all evaluations in $\mathbb{F}^{\ell}$.

• Explicitly, $RM : \{ c_{\bar{i}} \in \mathbb{F} \mid |\bar{i}| \leq d \} \longmapsto$ ← exponent-vector

$$\Big\{ P(x_1, \dots, x_\ell) := \sum_{i_1 + \dots + i_\ell \leq d} c_{\bar{i}} \cdot \bar{x}^{\bar{i}} \;\Big|\; x_1, \dots, x_\ell \in \mathbb{F} \Big\}$$

Observe: • RM with $d=1$ (& $\mathbb{F}=\mathbb{F}_2$) $\Rightarrow$ WH-code.

• RM with $\ell=1$ $\Rightarrow$ RS-code.

Lemma 3: RM is an ecc with distance $1-\frac{d}{|\mathbb{F}|}$.

$\mathbb{F}$-alphabet

Proof: • Again, $wt(RM(a-b))$

$= wt(RM(a)-RM(b)) = \Delta(RM(a), RM(b)) \cdot m$.

$m := |\mathbb{F}|^\ell$

• If $a-b \neq \bar{0}$, then $\sum_{|\bar{i}| \leq d} (a_{\bar{i}} - b_{\bar{i}}) \bar{x}^{\bar{i}}$ has only $d/\mathbb{F}$ fraction of zeros,

by the Polynomial Identity Lemma.

$\Rightarrow \quad \Delta(RM(a), RM(b)) \geq 1 - \frac{d}{|\mathbb{F}|}$.

$\square$

Concatenated Code (Forney 1966)

- WH has a large m, while RS needs a non-binary alphabet. We want to remove both the drawbacks.

So, we first apply RS & then WH.

to spread any change across bits!

Defn: Let $\mathbb{F}$ be a finite field of size q;

$$RS: \mathbb{F}^n \to \mathbb{F}^m \quad \& \quad WH: \mathbb{F} = \{0,1\}^{\lg q} \to \{0,1\}^q.$$

- The concatenated code $WH \circ RS : \mathbb{F}^n = \{0,1\}^{n \lg q} \longrightarrow \{0,1\}^{mq}$ is:

1) $RS(x) =: (RS(x)_1, RS(x)_2, \ldots, RS(x)_m) \in \mathbb{F}^m$

2) $WH \circ RS(x) =: (\underbrace{WH(RS(x)_i)}_{\{0,1\}^q} \mid i \in [m]) \in \{0,1\}^{mq}.$

$\triangleright$ WH$\circ$RS is computable in poly(mq) = poly$(|\mathbb{F}|)$-time.

product

Lemma 4: WH$\circ$RS is ecc of distance $\frac{1}{2}\cdot\left(1-\frac{n-1}{m}\right)$.

Pf: • Let $x \neq y \in \mathbb{F}^n = \{0,1\}^{n\lg q}$.

• We've: #distinct elements in $RS(x)$ & $RS(y)$ is $\geq \left(1-\frac{n-1}{m}\right)\cdot m$

• Moreover, if $x' \neq y' \in \mathbb{F}$ are in the i-th place of $RS(x)$, $RS(y)$ resp., then $\Delta(WH(x'), WH(y')) \geq \frac{1}{2}$.

$$\Rightarrow \Delta(WH\circ RS(x), WH\circ RS(y)) \geq \frac{\left(1-\frac{n-1}{m}\right)m \times \frac{1}{2}\cdot q}{mq}$$

$$= \left(1-\frac{n-1}{m}\right)\cdot\frac{1}{2}.$$

$\square$

– By the prime-number theorem, $\forall k \geq 2$, $\exists$ prime p in $[10k, 11k)$. Let's work over the field $\mathbb{F} := \mathbb{F}_p$.

$\Rightarrow$ WH∘RS is ecc that stretches $\Theta(k \lg k)$-long message to length $10k \cdot 11k = O(k^2)$. With distance $\geq$

$$\frac{1}{2} \cdot \left(1 - \frac{k}{10k}\right) = 0.45$$

▷ $\exists$ poly-time computable ecc $E: \{0,1\}^n \to \{0,1\}^{n^2}$, that can sustain 22% of errors.

in fact, sub-quadratic stretch.

Efficient Decoding

Qn: Can we find the unique x; given a string y' "close to" codeword $E(x)$?

- Decoding WH is trivial. Since WH length is 2^n, we can afford to scan the whole space $\{0,1\}^n$ & find the unique x, given y', in poly(2^n)-time.

Decoding RS

- Setting: Given a list: $(a_1,b_1),\dots,(a_m,b_m) \in \mathbb{F}^2$; for which $\exists$ deg-d (unique) polynomial $G:\mathbb{F}\to\mathbb{F}$ s.t. $G(a_i)=b_i$ for t of the pairs.

▷ Since RS has distance $(1-\frac{d}{m})$, we're guaranteed the existence of a unique G, if
$$t > m - \frac{1}{2}\left(1-\frac{d}{m}\right)m = \frac{m+d}{2} \quad \& \quad |\mathbb{F}| = m > d.$$

- If $t=m$ then we could have just interpolated G from the linear-system: $G(a_i)=b_i$, $\forall i\in[m]$.

Idea – Assume $t < m$: Introduce an auxiliary polynomial – Error-locator polynomial $\mathcal{E}(X)$ of deg $=$ #errors $=$ $(m-d)/2$. Interpolate C & $\mathcal{E}$ from:

$$C(a_i) = b_i \cdot \mathcal{E}(a_i) \quad , \quad \forall i \in [m] \; ;$$

where $\deg \mathcal{E} = (m-d)/2$ & $\deg C = \deg(G \mathcal{E}) = d + \frac{m-d}{2} = (m+d)/2$.

Theorem (Berlekamp-Welch, 1986): $\exists$ poly$(m, \lg|\mathbb{F}|)$-time algorithm to find G from $\{(a_i, b_i) \mid i\}$.

Proof: 1) Find polynomials $C(X), \mathcal{E}(X)$ of deg $= \frac{m+d}{2}, \frac{m-d}{2}$ resp. s.t. $\forall i \in [m]$, $C(a_i) = b_i \cdot \mathcal{E}(a_i)$.

linear-system

2) Output $C(X) / \mathcal{E}(X)$.

- In Step-1, there are m equations & # unknowns $=$ $\left(1+\frac{m+d}{2}\right)+\left(1+\frac{m-d}{2}\right) = m+2$. (linear homogeneous)

- We already "know" a solution by considering:
$$E := \prod_{G(a_i)\neq b_i}(X-a_i) \quad \& \quad C := G\cdot E.$$
$(\Rightarrow \ C(a_i) = b_i\cdot E(a_i), \ \forall i.)$

- Let C & E be the solutions obtained in Step-1.
$\Rightarrow \quad C(a_i) - G(a_i)\cdot E(a_i) = 0$, for t of the i's.

- Note: $\deg(C(X) - G(X)\cdot E(X)) \leq \frac{m+d}{2} < t$.
$\Rightarrow \quad C(X) - G(X)\cdot E(X) = 0. \Rightarrow C/E = G(X)$.

- All steps doable in $\mathrm{poly}(m, \lg|\mathbb{F}|)$-time. $\square$

Decoding WH∘RS

IIT

Theorem: For $WH\circ RS : \{0,1\}^{n\lg q} \to \{0,1\}^{mq}$, $\exists$ poly(q)-time decoder, if error-fraction $< \frac{1}{4}\cdot\left(\frac{1}{2} - \frac{n-1}{2m}\right)$.

Proof:

- Let y' be "close" to $y =: \langle WH(RS(x)_i) \mid i \in [m] \rangle$.
- The hypothesis implies:

$$\#\{i \mid WH(RS(x)_i) \text{ has } \geq q/4 \text{ errors}\} < \left(\frac{1}{2} - \frac{n-1}{2m}\right)\cdot m$$
$$= (m-n+1)/2.$$

$\Rightarrow$ WH-decoding will yield $\langle \tilde{y}_1, \ldots, \tilde{y}_m \rangle =: \tilde{y}$; with $\tilde{y}_i = RS(x)_i$ for $> m - \frac{m-n+1}{2} = \frac{m+n-1}{2}$ of the i's.

$\Rightarrow$ RS-decoding of $\tilde{y}$ yields the unique x. $\square$

▷ WH∘RS is a practical binary-ecc, that handles up to 11% of errors.

- Recall that for hardness-amplification we need stronger forms of decoding.