6.S08 Lecture 6 2/25/08

Note Title 2/25/2008

Today: Algebraic Codes

- Wozencraft's Ensemble
- Reed-Solomon
- Reed-Muller, Hadamard
- Concatenated Codes
- Justesen Codes

Review

- Code parameters: $(n, k, d)_q$; R, δ
- "Random / Greedy / Gilbert / Varshamov":

  $\exists$ codes with $q=2$; $R, \delta > 0$

  $[R = 1 - H(\delta)]$

  "Gilbert Ensemble size": 2^{2^n}

  "Varshamov " " ": 2^{n^2}

WOZENCRAFT ENSEMBLE

- Codes from $\{0,1\}^k \longrightarrow \{0,1\}^{2k}$
- Let $\mathbb{F} = \mathbb{F}_{2^k}$.
- Recall $\mathbb{F}_2^k \longleftrightarrow \mathbb{F}_{2^k}$ preserving addition
- Ensemble $= \{C_\alpha\}_{\alpha \in \mathbb{F}_{2^k}^*}$

$$C_\alpha : \underset{\substack{\Uparrow \\ \mathbb{F}_{2^k}}}{m} \longrightarrow \underset{\substack{\Uparrow \\ \mathbb{F}_{2^k}^2}}{\langle m, \alpha m \rangle}$$

- Lemma: $\exists \alpha$ s.t. $\Delta(C_\alpha) \geq H^{-1}(.5) \cdot n$

In fact $\Pr_\alpha\left[\Delta(C_\alpha) \geq (H^{-1}(.5) - \epsilon)\right] \longrightarrow 1$

- Claim: $\forall \langle x,y\rangle \neq 0$ there is at most one α s.t. $\langle x,y\rangle \in C_\alpha$

Proof: $x \neq 0 \Rightarrow \alpha = x^{-1} y$.

- Say α is bad if $\exists^{0\neq} \langle x,y\rangle \in C_\alpha$ with $wt(\langle x,y\rangle) < H^{-1}(.5) - \epsilon$

- $\#$ bad α's $\leq \# \{\langle x,y\rangle \neq 0$ s.t. $wt(\langle x,y\rangle) \leq H^{-1}(.5) - \epsilon\}$

$$\leq 2^{(.5-\epsilon')\cdot n}$$

- $\Pr_\alpha [\alpha \text{ bad}] \leq 2^{-\epsilon' n}$ $\boxtimes$

Notes:

- Why is this interesting?

    1. Algebraic
    2. Ensemble size even smaller. (2^k)
    3. Can be "computed" in time poly(k).
    4. Can we try to find good α explicitly? Remains open.

- Can extend to larger rates, smaller rates
  $\left(\frac{t-1}{t}\right)$ $\left(\frac{1}{t}\right)$ ...

..

Codes by Polynomials

General Idea

Message $\equiv$ coefficients of polynomial

Encoding $\equiv$ Evaluation

Evaluation $\Rightarrow$ Encoding

Interpolation $\Rightarrow$ Decoding from no errors.

(GENERALIZED) REED SOLOMON CODES :

Specified by $\underline{\Sigma} = \mathbb{F}_q$

$\underline{n} \leq q$, $0 \leq \underline{k} \leq n$, distinct $\underline{\alpha_1 \ldots \alpha_n} \in \mathbb{F}_q$

$$m = (m_0 \ldots m_{k-1}) \longmapsto \langle M(\alpha_1) \ldots M(\alpha_n) \rangle$$

$$M(x) = \sum m_i x^i$$

"Cor" $\Rightarrow$ $\Delta(RS_{\mathbb{F}_q, \alpha_1 \dots \alpha_n, k}) \geq n - (k-1)$

$= n - k + 1$

Matches Singleton !!

[Classical RS : Set $\alpha_1 \dots \alpha_n$ = all non-zero elements of $\mathbb{F}_q$]

Conclusion: if $q \geq n$ & $q = p^t$ then

can achieve "optimal" codes $\underbrace{[n, k, n-k+1]_q}$

MDS - "Maximum Distance Seperable".

What about smaller alphabets ?

Multivariate Polynomials $\Leftrightarrow$ Reed Muller Codes

Fix $\Sigma = \mathbb{F}_q$, degree r,

#variable m.

<u>Then</u>: message = coefficients of deg r poly

$$r < q \Rightarrow k = \binom{m+r}{r}$$

generally $\rightarrow$ $k \geq \left(\frac{r}{m}\right)^m, \binom{m}{r} \ldots$

Encoding $\equiv$ Evaluations

$$n = q^m$$

Distance? :

$r < q$: $\quad \Delta(C) = \left(1 - \frac{r}{q}\right) \cdot n$

$r \geq q$: $\quad \Delta(C) \geq q^{-\frac{r}{q-1}} \cdot n$

Example Choices:

1. Given k

$q = \log^2 k$

$r = \frac{q}{2}$

m s.t. $\binom{m + q/2}{m} = k \Rightarrow m = \frac{\log k}{\log \log k}$

$n = q^m \approx k^2$

$$\Rightarrow \left(k^2, k, \tfrac{1}{2}k^2\right)_{\log^2 k} \quad \text{code}$$

Rate $\longrightarrow 0$; Dist $= \frac{1}{2}$

2 Fix $m = O(1)$

Given k, pick $q = 2m \cdot k^{1/m}$

$r = q/2$

$\vdots$

$$\Rightarrow \left((2m)^m k, k, \tfrac{1}{2}(2m)^m k\right)_{2m k^{1/m}} \quad \text{code}$$

Smaller alphabet than RS, smaller rate.

3 $q=2$; $r=1$; $m=m \to \infty$

coefficients $\triangleq k = m+1$

Gives $[2^k, k+1, 2^{k-1}]_2$ code

$\Downarrow$

$\exists\ [2^k-1, k, 2^{k-1}]_2$ code

Tight for Plotkin $\downarrow$ Simplex Code

Dual $= [2^k-k-1, k, ?]$ code!

$? = 3 \Leftarrow$ Hamming code!!

Sometimes called "Hadamard Code"

Hadamard matrices & Codes

$n \times n$ matrix $H \in \{-1, +1\}^{n \times n}$

is a Hadamard matrix if

$$H \cdot H^T = n \cdot I$$

$H \Rightarrow$ binary codes as follows.

(1) w.l.o.g. first column of H is all $+1$'s

(if not flip entire row).

Drop first column, rest of rows form

$$\left(n-1, \log n, \frac{n}{2}\right)_2 \quad \text{code}$$

(simplex code)

(2) Rows of H & their complements $-H$

form $(n, \log 2n, \frac{n}{2})_2$ code

Hadamard code.

RM with $m = \log n$, $r=1$, $q=2$ is such a code.

Summary

- Algebra leads to nice codes;
- Matches Singleton, Plotkin (ii),
- But hasn't (yet) given $q = O(1)$, $R, \delta > 0$ ....
- But leads to them.

Concatenation Of Codes [Forney]

- A naive idea (to get binary codes):
  - Start with Reed Solomon code over $\mathbb{F}_{2^t}$ $\quad t = \log n$
  - Represent $\mathbb{F}_{2^t}$ as t bits
  - Say RS code was $[n, \frac{n}{2}, \frac{n}{2}]_n$.
    Then we get $[n \log n, \frac{n}{2} \log n, \frac{n}{2}]_2$ code by this process.
  - Rate is still good; Distance suffers because $\mathbb{F}_{2^t}$ represented as t bit string. Poor Redundancy in this rep'n.

Better Idea: Represent $\mathbb{F}_{2^t}$ nicely, using "Error-Correcting Code"

- Say we "know" good code

$$C_{inner} : \{0,1\}^t \to \{0,1\}^{2t}$$

Say $(2t, t, .01t)_2$ code.

- Using C_{inner} to represent elements of $\mathbb{F}_{2^t}$ & "combining" with RS gives

$$\left(2tn, \frac{tn}{2}, \frac{.01tn}{2}\right)_2 \text{ code}$$

$R, \delta > 0$!

CONCATENATED CODES [FORNEY '66]

- Combination technique called "Concatenation"
- Can concatenate

$$(n_1, k_1, d_1)_{2^{k_2}} \circ (n_2, k_2, d_2)_2$$

code to get $(n_1 n_2, k_1 k_2, d_1 d_2)_2$ code.

- Code over big alphabet : Outer code
  Small code over small " : Inner code
- Outer alphabet $\equiv$ Inner message space
- Both Outer, Inner linear & using

  $\mathbb{F}_{2^{k_2}} \longleftrightarrow \mathbb{F}_2^{k_2}$ correspondence yield

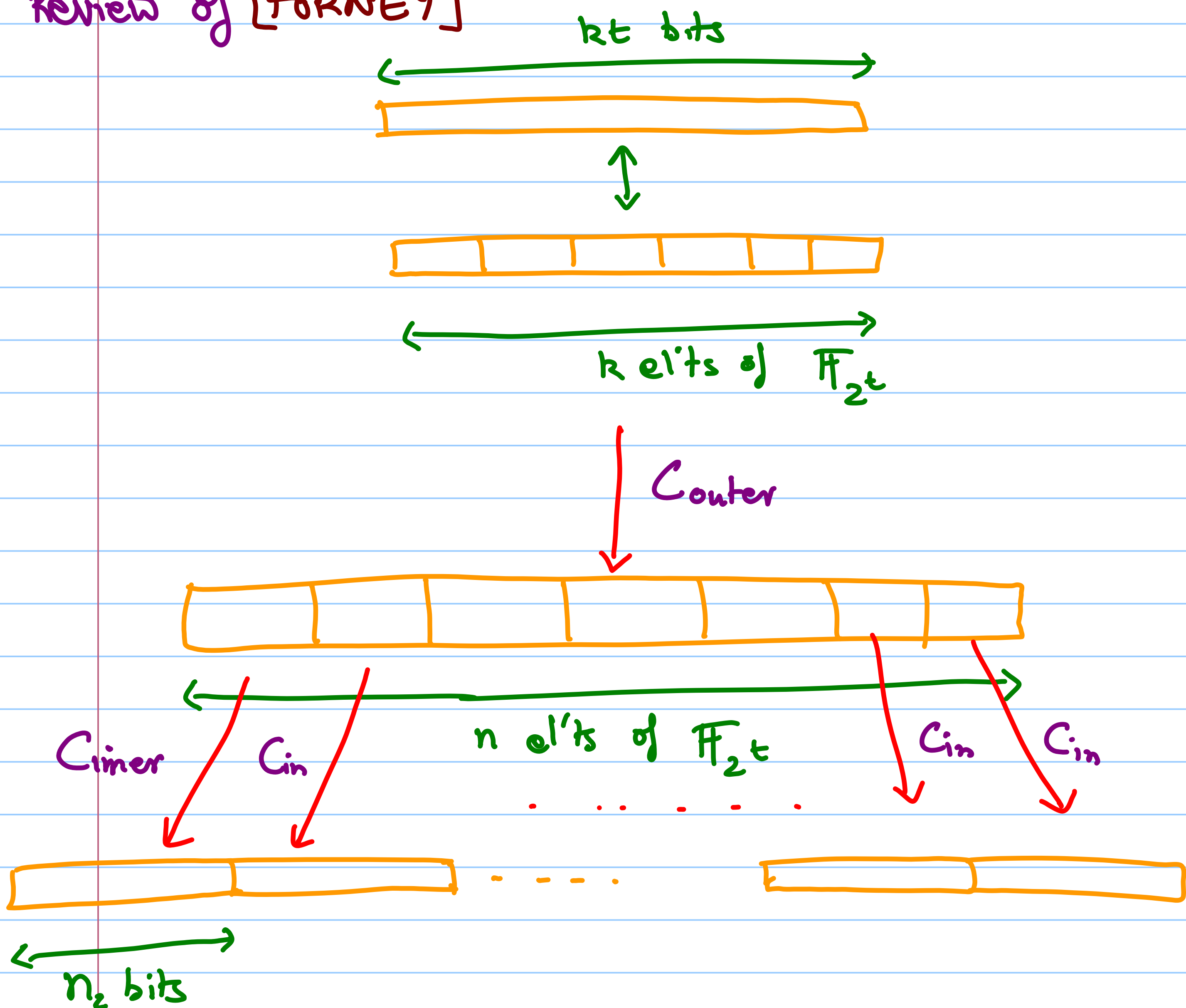
  linear codes.

DOES THIS GIVE EXPLICIT CODES?

- How do you find Outer code? Easy because of larger alphabet (use RS)
- How do you find Inner code?
  - This code is smaller, can try recursion, but hasn't worked ... so far.
  - [FORNEY] Use VARSHAMOV search! Takes time $\text{poly}(2^{k_2}) = \text{poly}(n)$
- Conclusion 1: YES - this gives explicit codes.... Encoding can be done in polynomial time.
- Conclusion 2: NO - this is still "search"...... [Only formalized recently .... e.g. should be able to compute $(i,j)^{th}$ entry of generator in time $\text{poly}(\log n)$.]

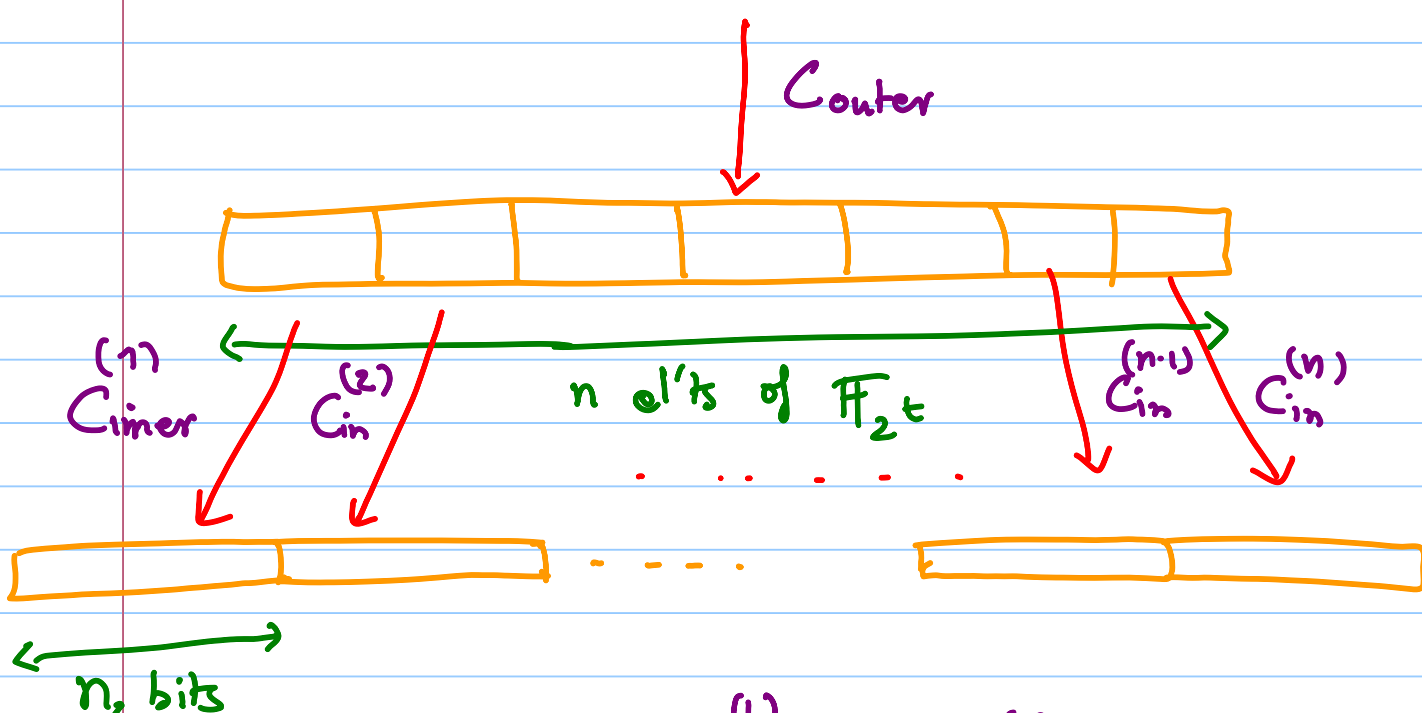
Justesen's Idea

Review of [Forney]



- Search problematic, since we need good C_{inner} so that we use it repeatedly
- But why should we use same C_{inner}? Why not "try" out many different ones, in same code?

(So replace last step of FORNEY with ...



[Using different $C_{in}^{(1)} \dots C_{in}^{(n)} \dots$]

- Construction certainly works if every code in $\{C_{in}^{(1)} \dots C_{in}^{(n)}\}$ good

- But even works if "most" codes are good! As in WOZENCRAFT's ENSEMBLE

- JUSTESEN = REED-SOLOMON $\circ$ {WOZENCRAFT}

EXPLICITLY : Fix integer t

- Compute $\mathbb{F}_{2^t}$.

- Encode : $m_0, \dots m_{k-1} \in \mathbb{F}_{2^t}$

Let $M(x) \triangleq \sum m_i x^i$; $\langle M(\alpha), \alpha \cdot M(\alpha) \rangle_{\alpha \in \mathbb{F}^*}$

- Exercise : Verify this is "explicit".