Lecture 17: Concatenation Codes
A set of codes \( \{ C^{(1)}, C^{(2)}, \ldots, C^{(T)} \} \) is called an ensemble

Wozencraft’s Ensemble is a set of \( T = 2^t - 1 \) \([2t, t, d]\) codes such that a “large” fraction of them have a “large distance”

Let \( \mathbb{F} = \mathbb{GF}(2^t) \)

Let \( C^{(\alpha)} \) be the code that maps \( x \mapsto (x, \alpha x) \), where \( \alpha, x \in \mathbb{F} \) and \( \alpha \neq 0 \)

For any \( \alpha \in \mathbb{F}^* \), note that the code \( C^{(\alpha)} \) is a \([2t, t, 2]_2\) code (here we interchangeable interpret the field elements as \( t \)-bit strings)
Claim

For any $0^{2t} \neq y \in \{0, 1\}^{2t}$, there exists at most one $\alpha \in \mathbb{F}^*$ such that $y \in C^{(\alpha)}$.

Proof.

- Suppose there are two distinct $\alpha, \beta \in \mathbb{F}^*$, such that $y \in C^{(\alpha)}$ and $y \in C^{(\beta)}$

- Suppose $y = (y_1, y_2)$

- This implies $x = y_1$ and $y_2 = \alpha x = \beta x$, that implies $\alpha = \beta$ (a contradiction)
**Claim**

At most $\text{Vol}_2(d - 1, 2t) - 1$ codes in the Wozencraft Ensemble have distance $< d$.

**Proof.**

- Consider any non-zero $y \in \text{Ball}_2(d - 1, 2t) \setminus \{0^{2t}\}$
- There exists at most one code in the Wozencraft Ensemble that contains $y$
- So, there are at most $\text{Vol}_2(d - 1, 2t) - 1$ codes in the Wozencraft Ensemble have distance $< d$
Claim

At least $2^t - \text{Vol}_2(d - 1, 2t) \approx 2^t - 2^{h_2(d/2t) \cdot 2t}$ codes in the Wozencraft Ensemble have distant $\geq d$.

Proof.

- There are $2^t - 1$ codes in the ensemble and the previous claim, the result follows
GV-Bound says that there is an \([n, k, d]_2\) code such that

\[
2^k \geq \frac{2^n}{\text{Vol}_2(d, n)} \approx 2^n(1 - h_2(d/n))
\]

Equivalently

\[
\frac{k}{n} \geq 1 - h_2\left(\frac{d}{n}\right)
\]

Let Rate \(R = k/n\) and relative distance \(\delta = d/n\)

Then, GV-bound says that there exists a binary linear code such that

\[
R \geq 1 - h_2(\delta)
\]

Can we construct one code that (nearly) achieves this?

We will use Reed-Solomon Codes and Wozencraft Ensemble to (nearly) achieve this bound
Let $\mathbb{F} = \text{GF}(2^t)$ and $q = |\mathbb{F}|$
For every $k$, there exists a $[2^t - 1, k, 2^t - k]_q$ code
- Suppose the input message is $(m_0, \ldots, m_{k-1}) \in \mathbb{F}^k$
- Interpret this input message as a polynomial $M(X) = \sum_{i=0}^{k-1} m_i X^i$
- Evaluate the concatenation of $M(X)$, for all $X \in \mathbb{F}^*$
Let $C^{(\text{out})} = [N, K, D]_Q$ code (called, outer code)

Let $C^{(\text{in})} = [n, k, d]_q$ code (called, inner code)

Such that $q^k = Q$

For example, consider $C^{(\text{out})}$ as the $[2^t - 1, k, 2^t - k]_q$ Reed Solomon code in the previous slide and any $C^{(\text{in})} = [n, t, d]_2$ code

The concatenation of $C^{(\text{out})}$ and $C^{(\text{in})}$ is the code where we encode each $Q$-ary alphabet of the codeword in $C^{(\text{out})}$ by the $C^{(\text{in})}$ code

Continuing the example, the concatenation of the Reed-Solon code with the $C^{(\text{in})}$ is the following code. Evaluate the polynomial $M(X)$ at each $X \in \mathbb{F}^*$ and encode $M(X)$ using $C^{(\text{in})}$
The concatenation code $C = C^{(\text{out})} \circ C^{(\text{in})}$ is an $[Nn, Kk]_q$ code (Prove this)

The distance of $C$ is at least $Dd$ (Prove this)

Therefore, $C = C^{(\text{out})} \circ C^{(\text{in})}$ is an $[Nn, Kk, \geq Dd]_q$ code
Continuing the example, the concatenation of the Reed-Solomon with any $C^{(in)} = [n, t, d]_2$ is an $[(2^t - 1)n, kt, (2^t - k)d]_2$ code.
The inner code used to encode each $Q$-ary alphabet on the outer-codeword can be different. As long as they are an $[n, t, d]_q$ code, the resultant concatenation is an $[Nn, Kk, \geq Dd]_q$ code.

Suppose all but $\Lambda$ of the inner codes have distance $d$. Then, the resultant concatenation is an $[Nn, Kk, \geq (D - \Lambda)d]_q$ code.
Recall that each code in the Wozencraft Ensemble is a $[2t, t]_2$ code and all except $\text{Vol}_2(d - 1, n) - 1$ of the codes have distance $\geq d$.

Recall that the Reed-Solomon codeword looks like

$$(M(1), M(2), \ldots, M(2^t - 1))$$

The concatenation with Wozencraft Ensemble implies that the $\alpha$-th $Q$-ary alphabet (here it is, $M(\alpha)$) is encoded with $C^{(\alpha)}$ (i.e., the map $x \mapsto (x, \alpha x)$).

So, the concatenation is

$$((M(1), 1M(1)), (M(2), 2M(2)), \ldots, (M(2^t - 1), (2^t - 1)M(2^t - 1)))$$
The concatenation, therefore, is a

$$[2(2^t - 1)t, kt, (2^t - k - \text{Vol}_2(d - 1, n))d]_2$$-code

How to choose the parameters to beat the GV-bound? (Think)