# Lecture 18: K-fold Repetition Code

**Summary:**
We examine the k-fold repetition code in detail, introducing fundamental concepts of coding theory including code size, length, and rate. We define Hamming distance and weight, proving that Hamming distance forms a metric. We establish conditions for error detection and correction, demonstrating how the minimum distance of a code determines its error-correcting capabilities.

**Topics Covered:** code, Hamming distance, Hamming distance and error correction theorem, Hamming weight, k-fold repetition code

**K-fold repetition code:**

We have a message $n \in \mathbb{F}_2^n$

The code word is

$$y = (\underbrace{x, x, \dots, x}) \in \mathbb{F}_2^{kn}$$

For example, if $x = 01$ and $k = 3$, then, we have

$$y = 010101$$

**Codes:**

$\mathbb{F}_q$ is a finite field with $q$ elements. $q = p^\alpha$ for prime $p$ and $\alpha \in \mathbb{Z}_{>0}$

Messages are all possible words of length $n$ (vectors in $\mathbb{F}_2^n$)

A code is

$$C \subseteq \mathbb{F}_q^m$$

for some $m \geq n$

We want an injective function

$$\mathbb{F}_q^n \to C$$

- $y \in C$ is the encoded word
- $z \in \mathbb{F}^m$ is the word after transmission

When $z \notin C$, an error has been detected.

If $y' = y$, we decoded correctly.

**Definition:**

The size of the code is the number of code words in it.

**Definition:**

The length of a code is the number of bits in a code word. In the case of $C \subseteq \mathbb{F}_q^m$, it is $m$.

**Definition:**

The rate of a code is defined as

$$\text{ratio} = \frac{\text{number of bits in message}}{\text{number of bits in codeword}}$$

**Definition:**

Let $y = (y_1, \ldots, y_m) \in \mathbb{F}_q^m$. We define the hamming weight

$$w(y) = |\{i \in [k]\ ;\ y_i \neq 0\}|$$

**Definition:**

Let $y, z \in \mathbb{F}_q^m$. We define the hamming distance

$$d(y,z) = w(y - z) = w(z - y)$$

The hamming distance counts the number of bits in which $y$ and $z$ differ.

**Definition:**

Let $C \subseteq \mathbb{F}_q^m$. We define the hamming distance of the code to be

$$d(C) = \min\{d(x,y)\ :\ x, y \in C.\ x \neq y\}$$

**Definition:**

Let $y \in C$ (the codeword sent) and $z \in \mathbb{F}_q^m$ (word received).

1. $z \notin C$ we say that an error has been detected
2. If there is a unique $y' \in C$ such that $d(y', z) < d(y'', z),\ \forall y'' \in C, y'' \neq y'$, then we say that $y'$ is the decoding of $z$
   - $y'$ is the unique code word with the closest hamming distance to $z$

**Exercise:**

Show that the Hamming distance is a metric, meaning it satisfies the following axioms:

1. $\forall x.\ d(x,x) = 0$
2. $\forall x, y.\ d(x,y) = d(y,x)$
3. $\forall x, y, z.\ d(x,y) \leq d(x,z) + d(z,y)$

**Theorem:**

Let $C$ be a code with the Hamming distance $d(C)$.

1. $C$ detects up to $s$ errors $\iff d(C) \geq s + 1$.
2. $C$ corrects up to $t$ errors $\iff d(C) \geq 2t + 1$

$(1) \impliedby$ **proof:**

Assume that $d(C) \geq s + 1$. Suppose we send $y \in C$ and we receive $z \in \mathbb{F}_q^n$ and $i$ errors occurred. This means

$$d(y,z) = i$$

We need to show that if $i \leq s$ then the error was detected.

We know that $d(y,z) = i \leq s < s + 1 \leq d(C)$

$$\implies d(y,z) < d(C)$$

Suppose, for contradiction, that $z$ was a code word. Then, we also have

$$d(y,z) \geq d(C)$$

Which is a contradiction. Therefore, we conclude that $z$ is not a code word. So, we have detected the error.

$$\text{QED}$$

$(1) \implies$ **proof:**

$C$ detects $1, 2, \ldots, s$ errors

Let $d(C) = i \implies \exists y_1 \neq y_2$ with $y_1, y_2 \in C$, $d(y_1, y_2) = i$. Note that if we send $y$, make $i$ errors and receive $y_2$, we won't know that errors occurred!

$$d(C) = i \geq s + 1$$

$$\text{QED}$$

$(2) \impliedby$ **proof:**

Assume that $d(C) \geq 2t + 1$. We want to prove that one can correct up to $t$ errors.

We send $y \in C$ and receive $z \in \mathbb{F}_q^m$ with $i$ errors, $i \leq t$, i.e. $d(y, z) = i$. To correct the error means that $y$ is the unique closest codeword to $z$. Suppose that to the contrary, $x \in C$ is such that $d(x, z) < i$.

$$d(x, y) \leq d(x, z) + d(z, y) \leq i + i = 2i < 2t + 1 \leq d(C)$$
$$\implies d(x, y) < d(C)$$

This implies that $x$ and $y$ are not code words. Therefore, $x \notin C$.

$$\text{QED}$$

$(2) \implies$ **proof:**

Assuming our code can correct $1, 2, 3, \ldots., t$ errors. Want to conclude that $d(C) \geq 2t + 1$.

Let $d(C) = i$. This means $\exists y_1 \neq y - 2, y_1, y_2 \in C$ such that $d(y_1, y_2) = i$.

$$f = \left\lfloor \frac{i}{2} \right\rfloor$$

$$c = \left\lceil \frac{i}{2} \right\rceil$$

$$i = f + c$$

Build a new word $z$ be replacing $f$ of the $i$ bits in $y$, in which $y_1$ and $y_2$ differed

$$d(y_1, z) = f$$

$$d(y_2, z) = c$$

$$d(y_1, z) \leq d(y_2, z)$$

Send $y_2$ as our message, receive $z$. We don't know how to correct $\implies i \geq t + 1$.

$$d(C) = i = f + c \geq c - 1 + c = 2c - 1 \geq 2(t + 1) = 2t + 1$$

$$\text{QED}$$