Error-Correcting Codes and the McEliece Cryptosystem

M. P. Nonog

July 22, 2025

Abstract

The goal of this project is to understand the McEliece Cryptosystem, which utilizes Goppa Codes. Goppa Codes are a subclass of Linear Codes, and are related to Cyclic Codes and BCH Codes. This specific cryptosystem is currently one of the four finalist in the Post-Quantum Cryptography standard that is being tested by the National Institute of Standards and Technology (NIST).

1 Introduction

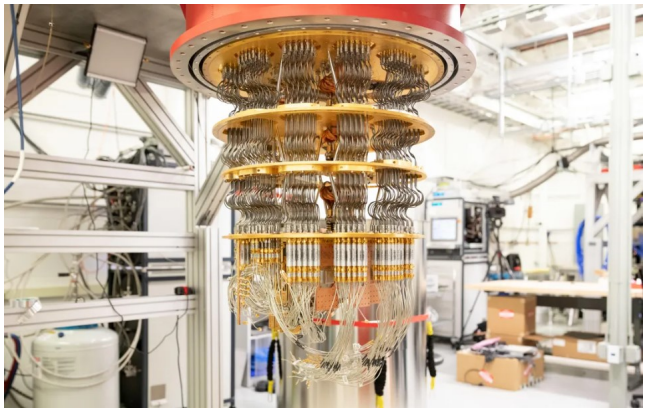


Figure 1: Google Quantum Computer (see `https://www.cnet.com/tech/computing/google-quantum-supremacy-only-first-taste-of-computing-revolution/`)

One of the three quantum computers that U.S. has is the Google Quantum Computer, as shown in Figure 1. This image showcases that Quantum Computer requires not only a space on your desk, but an entire room to establish

your own quantum computer. Therefore, one of the advances we can look forward to, if quantum computer becomes a norm in the future, is to make it compatible and easier to use like the regular computer we use today.

2 Background

Error-correcting code is the process of sending message and encode it to be a codeword. Then, that codeword goes through a noisy channel that will inroduce error and will be transformed to what we will call received word. The received word will be decoded by the recipient through locating where the error is and correct it to obtain the exact message.

RSA was broken using Shor's Algorithm in 1994. However, it took 22 years for the National Institute of Standards and Technology (NIST) to begin the search for a new Post-Quantum Cryptography standard, in order to protect our secrets against adversaries. Since 2016, NIST has evaluated over 50 cryptosystems. As of July 2020, only four are left standing. One of these is the McEliece Cryptosystem, based on error-correcting codes. The goal of this project is to understand the McEliece Cryptosystem, which utilizes Linear Codes, Cyclic Codes, BCH Codes and Goppa Codes.

What follows are examples of repitition code and parity check matrix. Description of what parity check matrix will be discussed in Linear Code section and more examples in each sections of the code for further understanding

Example 1. (Repetition Code) Consider a set $S = \{A, B, C, D\}$. A cryptologist want to send a letter from set S across a noisy channel with a probability of $p = 0.1$ error. Thus, there is 90% that a letter received is right. This leaves too large a chance of error. Instead, we repeat the letter three times. An example is BBB. Then, an error occurs and the received letters were BCB. The received will take the most frequent in the set of letters, which in this case "B". The probability of the correct message being found is calculated as the sum of all letters are correct and the probability of the received letters with one error. This is written as:

$$(0.9)^3 + 2(0.9)^2(0.1) = 0.972$$

This results to a smaller chance of error. This does not work with two repetitions. If you do have AB or BC or BD, we can detect that there is an error but we cannot correct it because we can't decide on what is the most frequent. In binary, let $F = \mathbb{Z}_2$, the integer mod 2, and let $n = 3$ for $X^n - 1$. Let $g(X) = X^2 + X + 1$. Then, using the coefficient of the X's we can write a matrix $(1, 1, 1)$. To get a code C, we need to multiply the $g(x)$ coefficients $(1, 1, 1)$ with 0. Therefore, $C = (0, 0, 0), (1, 1, 1)$. This is an example of binary repetition code. Since we have $n = 3$, let w be a primitive third root of unity. Since it is a root, $g(w) = g(w^2) = 0$.

Example 2. (Parity Check Matrix)
Suppose we want to send a message with 7 bits, which means (1,1,1,1,1,1,1). Then, let us add an eight bit so that the total number of bits are even (1,1,1,1,1,1,1,1). The goal is to make sure the sum of the bits are even. For example, 0111000 is a 7 bit with a total of odd nonzero digit. Therefore the 8th digit will be "1" to make the total an even number of nonzero digit, 01110001. An error of at least one bit is detected if the message receive have an odd number of nonzero bits. However, it is not possible to correct the error because an error in any bit could yield to odd number of nonzero bits.

3 The McEliece Cryptosystem

Bob constructs his public key G_1 by choosing a $k \times n$ generating matrix G for a linear code of dimension k, a $k \times k$ invertible matrix S, and an $n \times n$ permutation matrix P (invertible, one 1 in every row and column). Using this, Bob can create his Public Key $G_1 = SGP$. Next, Alice chooses a $1 \times k$ message $x \in \mathbb{Z}_2^k$, and an error vector e of weight t and length n (i.e., $e \in \mathbb{Z}_2^n$ and has Hamming weight t, or exactly t non-zero entries).

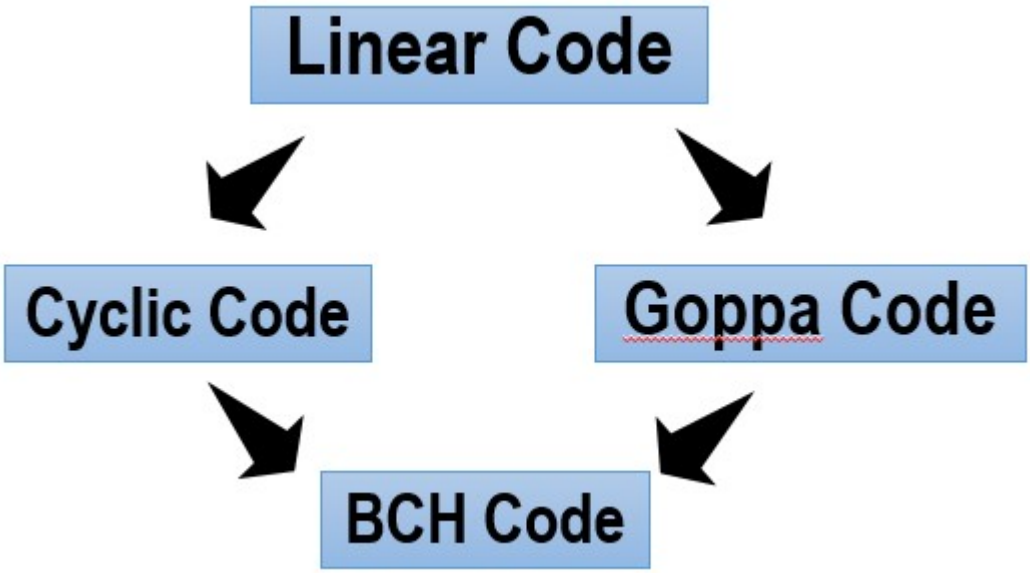
Alice	Public	Bob
	$G_1 \leftarrow$	$G_1 = SGP$
$xG_1 + e = y$	$\rightarrow y \rightarrow$	1) $r_1 = yP^{-1} = (xG_1 + e)P^{-1}$ $= x(SGP)P^{-1} + eP^{-1}$ $= (xS)G + eP^{-1}$
		2) Compute syndrome $r_1 H^T$
		3) Lookup the codeword $c_1 = (xS)G = x_1 G$ associated with received word r_1
		4) Extract message $x_1 = (xS)$ associated with code word c_1 (first k bits)
	Eve	5) Compute $x_1 S^{-1} = (xS)S^{-1} = x$

The McEliece Cryptosystem uses an $[n, k, d] = [1024, 512, 101]$ Goppa code. In this case, Eve has $\binom{1024}{50} \approx 3x10^{85}$ possible locations of the errors. This meets the parameters $n = 2^m$, $d = 2t+1$, and $k = n - mt$, where n is the code length, k is the vector space dimension and d is the minimum distance.

4 Linear Code

A Linear code branches out to Goppa Code and Cyclic Code. Goppa code has only one Cyclic Code, which is the BCH Code. First, the specific linear code we will be discussing in this paper is the linear hamming code and we will show examples to understand it fully. Then, we will follow the discussion with Cyclic

Code and few examples as well. After that, we will learn about Cyclic Code and finally, is the Goppa Code.



- **Definition:** An $[n, k, d]$ linear code is a vector space with dimension k and length n over a field $\mathbb{F}$, where the combination of any two codewords is always a codeword, with minimum distance d between two codewords.
- **Parity Check Matrix (PCM):** Given generating matrix $G = [I_k, P] \in C$, then $H = [-P^T, I_{n-k}]$ is a parity check matrix for C if and only if $vH^T = 0$
- **Coset:** Given linear code C, and n-dimensional vector v, a coset is the set of $v + C$. The vector with minimum Hamming weight is the **coset leader**.
- Given a linear code C, s errors detected if minimum distance $d(C) \geq s+1$
- Given a linear code C, t errors corrected if $d(C) \geq 2t + 1$
- **Syndrome:** This is defined as $S(r) = rH^T$, where r is the received word and H is the the parity check matrix

- **How to DECODE a received word r?**

1. Calculate the syndrome.
2. Find which coset the syndrome belongs.

3. Look for the coset leader.

4. message = r − coset leader.

Example: Consider a $[4, 2, 2]$ linear code with G, H.

$$\text{generating matrix } G = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

$$\text{parity check matrix } H = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

Here is the lookup table for decoding received words.

$$(0,0,0,0)(1,0,0,1)(0,1,0,1)(1,1,0,0)$$
$$(1,0,0,0)(0,0,0,1)(1,1,0,1)(0,1,0,0)$$
$$(0,0,1,0)(1,0,1,1)(0,1,1,1)(1,1,1,0)$$
$$(0,0,1,1)(1,0,1,0)(0,1,1,0)(1,1,1,1)$$

Alice encodes message $x = [1, 1]$ by computing $xG = [1, 1, 0, 0]$. She sends to Bob through a noisy channel, and Bob receives $r = (1, 1, 1, 0)$.

Next, Bob must **DECODE**, by calculating the syndrome $S(r) = rH^T$.

Coset Leader	Syndrome
$(0,0,0,0)$	$(0,0)$
$(1,0,0,0)$	$(0,1)$
$(0,0,1,0)$	$(1,0)$
$(0,0,1,1)$	$(1,1)$

$$S(r) = (1,1,1,0)\begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = (1,0).$$

Therefore, the code word = $\underbrace{(1,1,1,0)}_{\text{received word}} - \underbrace{(0,0,1,0)}_{\text{coset leader}} = (\underbrace{1,1}_{\text{message}}, 0, 0)$.

5 Cyclic Code

A code is said to be a cyclic code if it contains the property

$$(c_1, c_2, ..., c_n) \in C \iff (c_n, c_1, c_2, \ldots, c_{n-1}) \in C$$

Given $g(X) = g_0 + g_1X + \cdots + g_{n-1}X^{n-1} + g_{n-k}X^{n-k}$ (where $g(X) \mid X^n - 1$), we formulate $k \times n$ generating matrix G and $(n-k) \times n$ parity check matrix H. First, Let us create our G with a dimension of $k \times n$. To make the first row, use the coefficients of $g(X)$. Since, it is cyclic, then the second row is the right shift of 1st row, this is the same for the third row and so on. Now, we can obtain our H with a dimension of $(n-k) \times n$, which is as follows:

$$G = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-k} & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \cdots & g_{n-k} & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_0 & g_1 & \cdots & g_{n-k} \end{bmatrix}$$

$$H = \begin{bmatrix} h_k & h_{k-1} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_k & h_{k-1} & \cdots & h_0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_k & h_{k-1} & \cdots & h_0 \end{bmatrix}$$

Example: Here is an $[n, k, d] = [7, 3, 4]$ cyclic code C.

$$X^7 - 1 = g(X)h(X) = \underbrace{(X^4 + X^2 + X + 1)}_{g(X)}\underbrace{(X^3 + X + 1)}_{h(X)}.$$

Then 3×7 generating matrix G and 4×7 parity check matrix H are:

$$G = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \qquad H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

The limitation of cyclic code is that, it is way too complex to decode. Therefore, BCH code was implemented, which is related to both Cyclic Code and Goppa Code.

6 BCH (Bose-Chaudhuri-Hocquenhem) Code

Theorem 1. *Let C be an $[n, k, d]$ cyclic code over $\mathbb{F}_{q=p^m}$, where $p \nmid n$. Let α be a primitive n-th root of unity, and let $g(X)$ be a generating polynomial for C. Suppose there exist integers ℓ and δ such that*

$$g(\alpha^\ell) = g(\alpha^{\ell+1}) = \cdots = g(\alpha^{\ell+\delta}) = 0$$

Then the minimum distance $d \geq \delta + 2$.

Recall:

$$X^n - 1 = \begin{cases} f_1(X)f_2(X)\cdots f_r(X) & \text{over } \mathbb{F}_p \\ (X-1)(X-\alpha)(X-\alpha^2)\cdots(X-\alpha^{n-1}) & \text{over extension field } \mathbb{F}_q \end{cases}$$

Let $q_j(x) = f_i(\alpha^j) = 0$. Then, the generating polynomial of a BCH code is

$$g(x) = \text{LCM}\big(q_{k+1}(X), q_{k+2}(X), \ldots, q_{k+d-1}(X)\big)$$

for some integer k. The parity check matrix H is

$$H = \begin{bmatrix} 1 & \alpha^{k+1} & \alpha^{2(k+1)} & \dots & \alpha^{(n-1)(k+1)} \\ 1 & \alpha^{k+2} & \alpha^{2(k+2)} & \dots & \alpha^{(n-1)(k+2)} \end{bmatrix}$$

Example Given a generating polynomial $g(X)$ that contains a primitive root of unity α

$$g(\alpha^{k+1}) = g(\alpha^{k+1}) = \ldots = g(\alpha^{k+d-1}) = 0$$

where $k \in \mathbb{Z}$. **Example** Let us have

$$X^7 - 1 = \begin{cases} \underbrace{(X+1)}_{f_1(X)}\underbrace{(X^3+X+1)}_{f_2(X)}\underbrace{(X^3+X^2+1)}_{f_3(X)} & \text{over } \mathbb{F}_2 \\ (X+1)(X+\alpha)(X+\alpha^3)\cdots(X+\alpha^6) & \text{over } \mathbb{F}_{2^3} \end{cases}$$

Then, as we define $q_j(X) = f_i(X)$ where $f_j(\alpha^j) = 0$

$$q_1(X) = f_2(X), q_2(X) = f_3(X), q_3(X) = f_3(X),$$
$$q_4(X) = f_2(X), q_5(X) = f_3(X), q_6(X) = f_3(X).$$

The LCM is the generating polynomial $g(X) = X^6+X^5+X^4+X^3+X^2+X+1$.

How to DECODE a received word r for one error?

1. Calculate $rH^T = (s_1, s_2)$.
2. If $s_1 = 0$, then no error (r is a codeword).
3. If $s_1 \neq 0$, compute $\frac{s_2}{s_1} = \alpha^{j-1}$, where j is position of error.
4. $r - e =$ codeword

Example: Consider a $[7, 1, 7]$ BCH code with the generating polynomial $g(X) = X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$. There are two codewords: $(0, 0, 0, 0, 0, 0, 0)$ and $(1, 1, 1, 1, 1, 1, 1)$. Suppose Bob receives $r = (1, 1, 1, 0, 1, 1, 1)$. Now, detect and correct the error.

Solution: Since $rH^T = (s_1, s_2)$, we see

$$rH^T = (1,1,1,0,1,1,1)\begin{bmatrix} 1 & 1 \\ \alpha & \alpha^2 \\ \alpha^2 & \alpha^4 \\ \vdots & \vdots \\ \alpha^6 & \alpha^{12} \end{bmatrix} = (1+\alpha+\alpha^2+\alpha^4+\alpha^5+\alpha^6, 1+\alpha^2+\alpha^4+\alpha^8+\alpha^{10}+\alpha^{12})$$

$rH^T = (\alpha^3, \alpha^6) = (s_1, s_2)$

Since $s_1 \neq 0$, we calculate $\frac{s_2}{s_1} = \frac{\alpha^6}{\alpha^3} = \alpha^3$. Therefore, $j - 1 = 3$, so the error position is at $j = 4$. Finally, we see

$$r - e = \underbrace{(1,1,1,0,1,1,1)}_{\text{received word}} - \underbrace{(0,0,0,1,0,0,0)}_{\text{error vector}} = \underbrace{(1,1,1,1,1,1,1)}_{\text{codeword}}$$

7 Goppa Code

Here is a little introduction about the Goppa Code, for more deeper information see the reference, Let $L = (\gamma_0, \gamma_1, \ldots, \gamma_{n-1})$ be distinct elements in an extension field of $\mathbb{F}_q$, where q is a prime or a power of a prime. Then, the Goppa code $\Gamma(L, G)$ is the set of vectors $c_0 c_1 \ldots c_{n-1} \in \mathbb{F}_q^n$ such that Let $G(x) \in \mathbb{F}_q[x]$ with $F(\gamma_i) \neq 0$ for $0 \leq i \leq n-1$.

$$\sum_{i=0}^{n-1} \frac{c_i}{x - \gamma_i} \equiv 0 \bmod G(x)$$

where $G(\gamma_1) \neq 0$ (thus $x - \gamma_1$ is invertible in this ring). Then, the BCH code of length n and designed distance d is the Goppa code $\Gamma(L, G)$, where $L = (1, \alpha^{-1}, \alpha^{-2}, \ldots, \alpha^{1-n})$ and $G(x) = x^{d-1}$.

8 Conclusion

In this project, we studied the McEliece Cryptosyste, which uses Goppa Code. In order to understand Goppa Code, we needed to study Linear Code. Then, Linear code branched out to Cyclic Code and Goppa Code. Finally, the only Cyclic Code that is Goppa Code is the BCH Code. Comprehending these four gives us a wide understanding on how complicated McEliece Cryptosystem is which is important in error-correcting codes. The importance of learning this specific cryptosystem right now is high because this could be the one replacing the RSA that was published in 1977 that we are still currently using. This might take awhile to find out but if it becomes the one that will be picked by NIST as the Post Quantum Standard, then it might take more than golden years again to actually have a new one. Therefore, learning this one know is valuable and understanding it deeper is important for any future Naval Officer because this could be the system that we will be using in our generation.

References

[1] W. Trappe and L. C. Washington. *Introduction to Cryptography with Coding Theory*, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2nd ed, 2006.

[2] *Fundamentals of Error-Correcting Codes*, by V. Pless, W.C. Huffman, Cambridge University Press, 2003