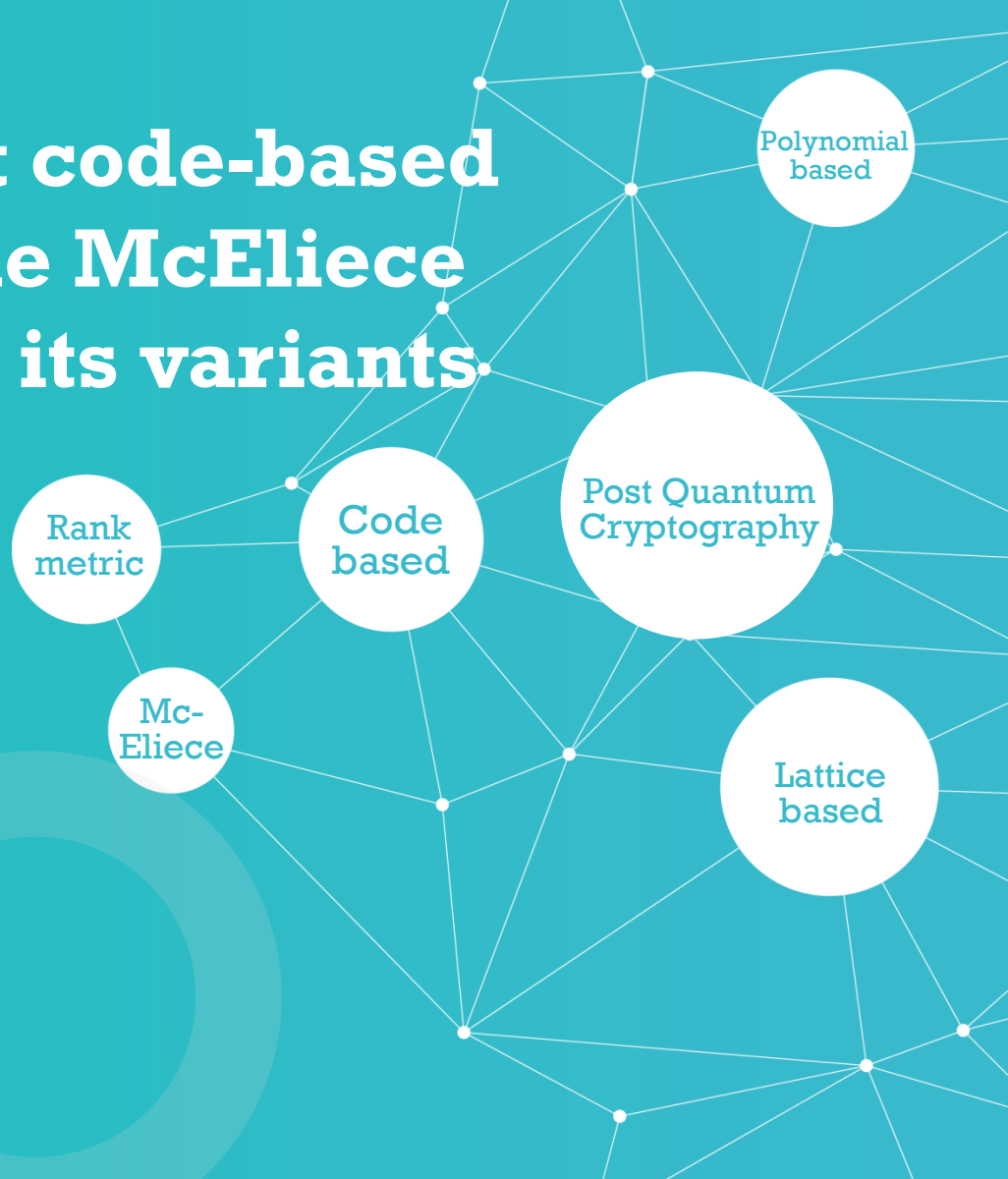
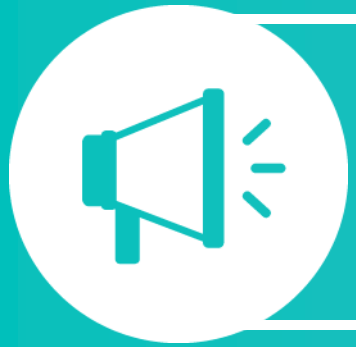
Quantum resistant code-based cryptosystems: the McEliece cryptosystem and its variants



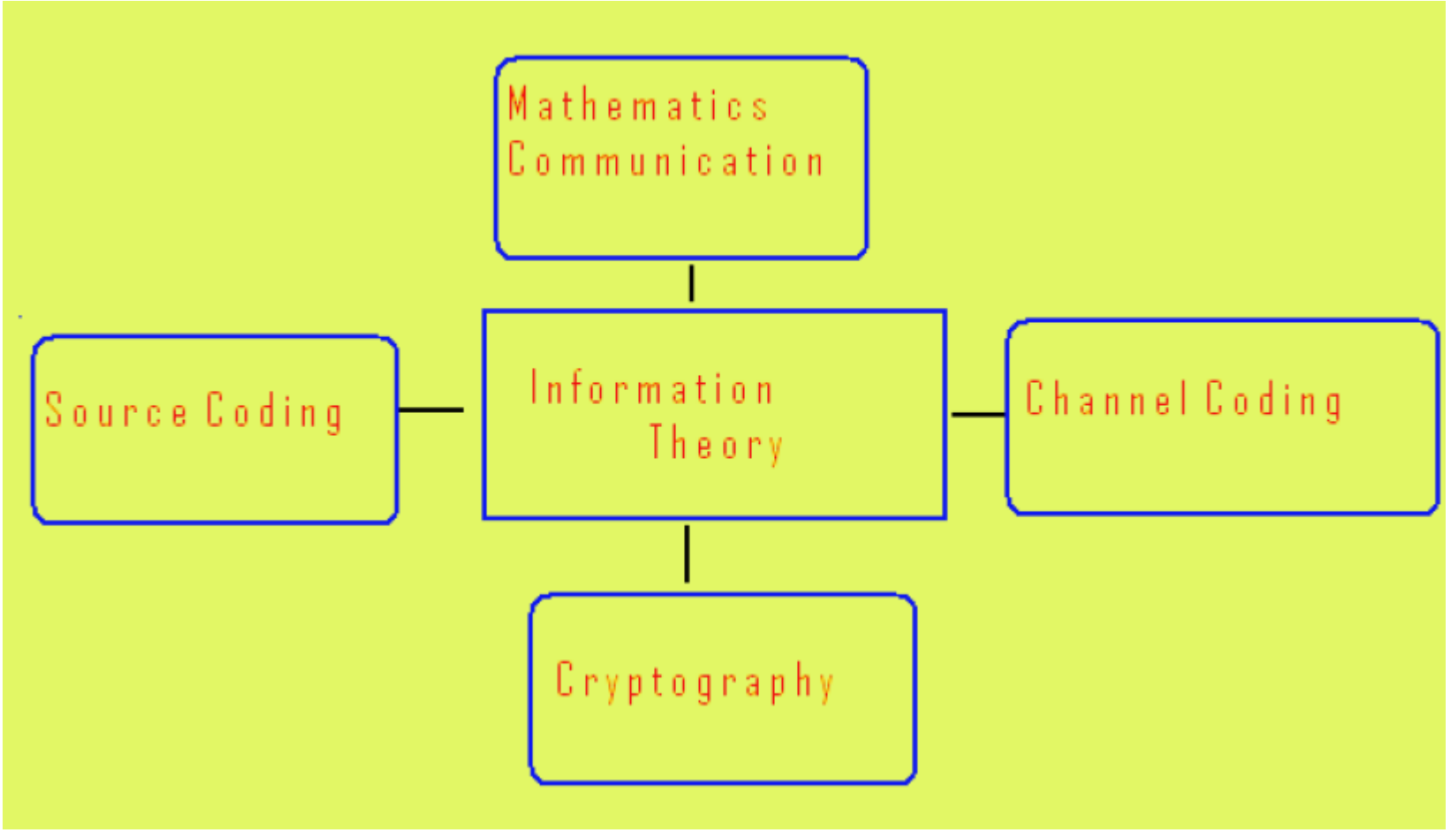
Jon-Lark Kim
Sogang University
3. 1. 2019
U. of California at Irvine
Crypto Seminar

Contents



1. Introduction to Coding Theory

Coding and Other Areas



Father of Information Theory

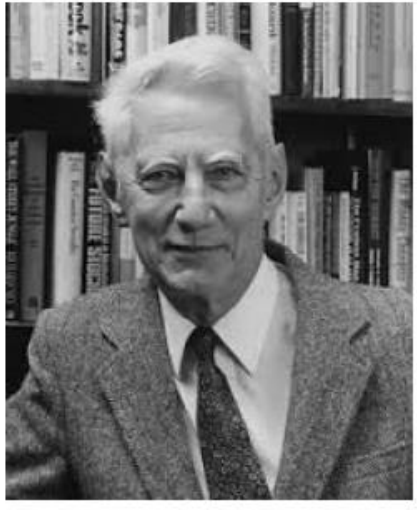
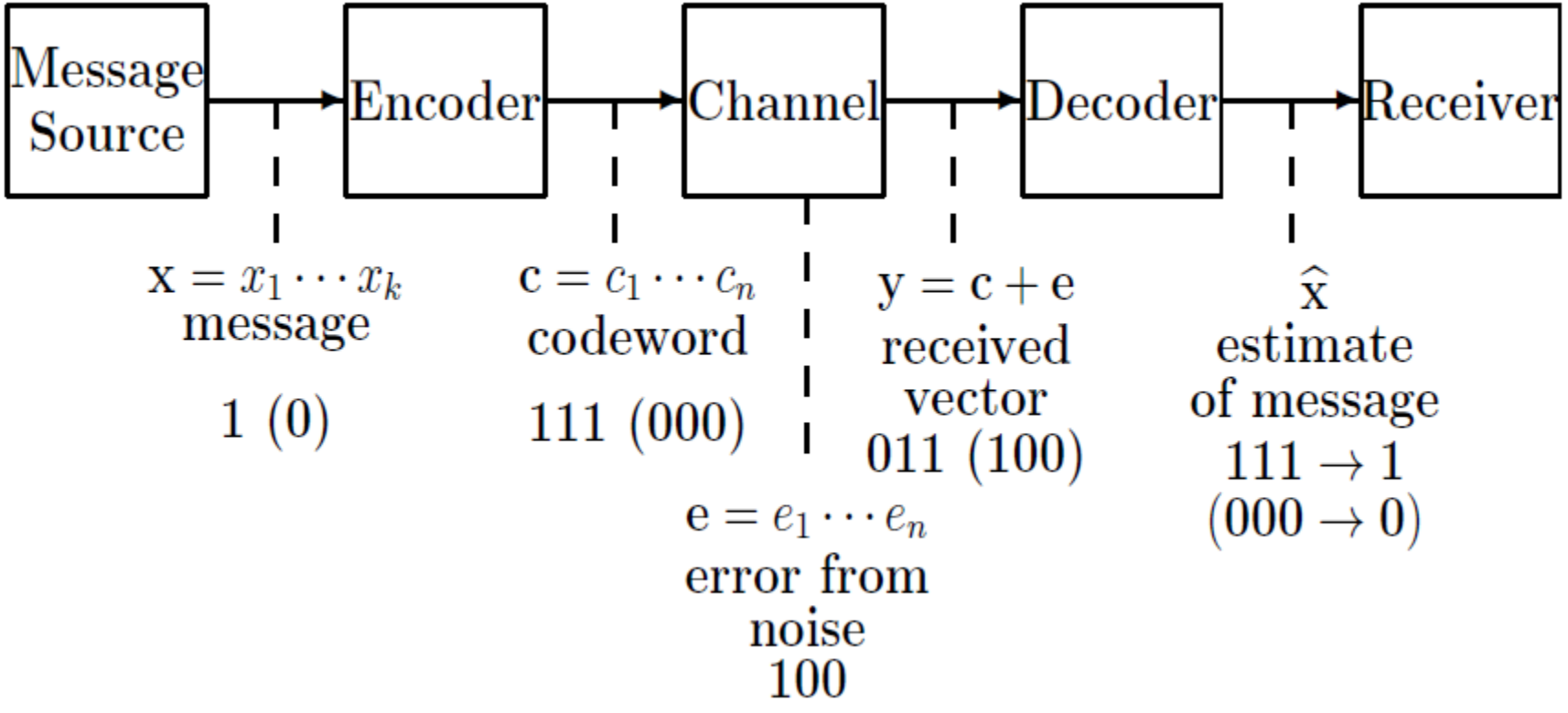


Figure: Claude Shannon (1916-2001)

Shannon's two foundational papers from Bell System Technical Journal:
"A Mathematical Theory of Communication" on Information Theory (1948)
"Communication Theory of Secrecy Systems" on Cryptography (1949)

Shannon's Communication Channel



Inventor of error-correcting codes

Figure: Richard Hamming (1915-1998, Bell Lab)

Two Operations on Finite Alphabets

- Let A be a finite alphabet. Usually $A = \mathbb{Z}_2, \mathbb{Z}_p$ (in general $\mathbb{F}_q$ or $\mathbb{Z}_m$).
- Operation on $\mathbb{Z}_2 = \{0, 1\}$

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

- Operation on $\mathbb{Z}_p = \{0, 1, 2, \cdots, p-1\}$ for a prime p. For $a, b \in \mathbb{Z}_p$, $a + b \pmod p$ and $a \cdot b \pmod p$. For example, $1 + 2 = 3 \equiv 0 \pmod 3$ and $2 \cdot 2 = 4 \equiv 1 \pmod 3$.

What is a code?

- $A^n := \{(x_1, \cdots, x_n) | x_i \in A\}$
- An (error-correcting) code $\mathcal{C}$ over A is a subset of A^n (with at least two elements).
- Elements of $\mathcal{C}$ are called codewords.
- A code over $\mathbb{Z}_2$ is called a binary code.
- The weight of $\mathbf{x} = (x_1, \cdots, x_n)$ is the number of nonzero coordinates, denoted by $\mathrm{wt}(\mathbf{x})$. For example, $\mathrm{wt}(0, 1, 2, 1, 0) = 3$.
- The Hamming distance $d(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}, \mathbf{y} \in A^n$ is $\mathrm{wt}(\mathbf{x} - \mathbf{y})$. For example, if $\mathbf{x} = (1, 0, 0, 1, 0)$ and $\mathbf{y} = (0, 0, 1, 0, 0)$, then their Hamming distance is 3.

Linear code: most useful code

- A linear code $\mathcal{C}$ of length n and dimension k over $\mathbb{Z}_p$:= a k-dimensional subspace of $\mathbb{Z}_p^n$.
- We denote $\mathcal{C}$ by an $[n, k]$ linear code over $\mathbb{Z}_p$.
- The minimum distance (weight) d of a linear code $\mathcal{C}$:=the minimum of wt($\mathbf{x}$), $\mathbf{x} \neq \mathbf{0} \in \mathcal{C}$.
- We denote it by an $[n, k, d]$ code. Given n and k, d can be at most $n - k + 1$ (Singleton' bound).

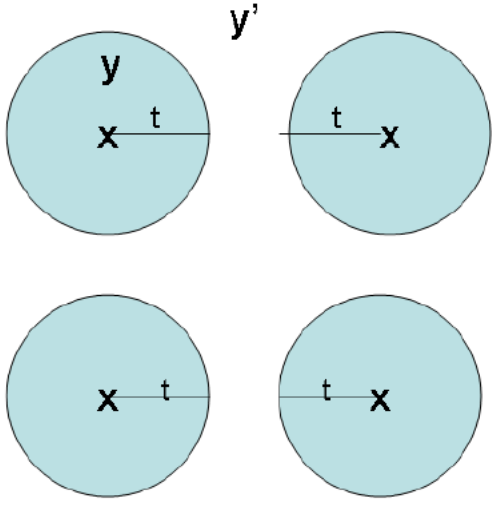
- A linear *[n,k,d]* code with *d=n-k+1* is called an MDS code.

Nearest neighbor decoding

Theorem

Any $[n, k, d]$ linear code can correct up to $t = \lfloor \frac{d-1}{2} \rfloor$ errors (by the nearest neighbor decoding).

Sketch of a geometric proof:



x: codeword
y, y': received vector

2t < d

1. y is uniquely decoded as x.
2. y' is not decoded.

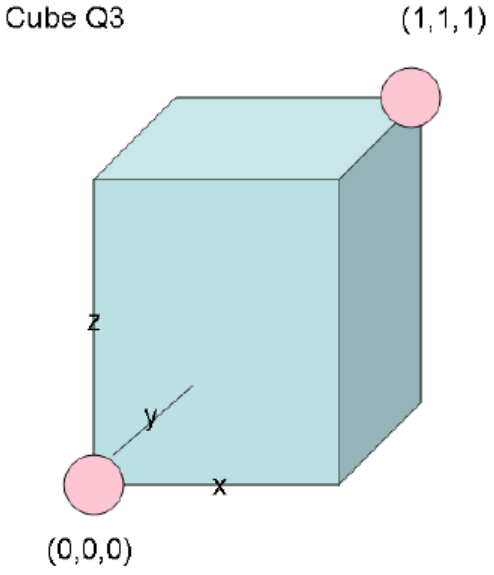
Dual of a linear code

- For $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$ in $\mathbb{Z}_p^n$, the dot product of **x** and **y** is defined as

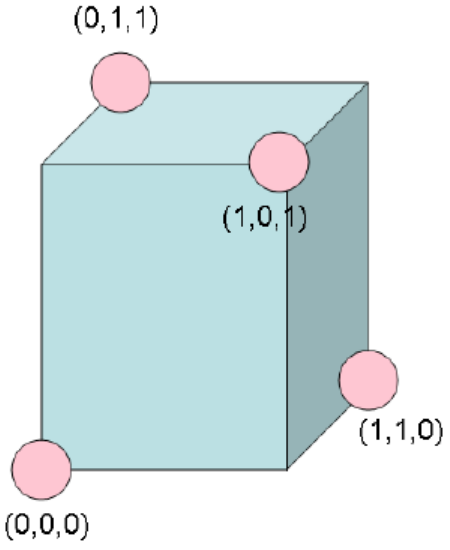
$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n.$$

- The dual of $\mathcal{C}$, denoted by $\mathcal{C}^\perp$, is
$\mathcal{C}^\perp = \{\mathbf{y} \in \mathbb{Z}_p^n \mid \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{x} \in \mathcal{C}\}$

Simple example



C1={(0,0,0),
(1,1,1)}



C2={(0,0,0),
(1,1,0),
(1,0,1),
(0,1,1)}

Relation between C_1 and C_2

- $\mathcal{C}_1$ is a $[3, 1, 3]$ binary linear code.
- $\mathcal{C}_1$ can correct any one error. For example, if (100) is received when (000) is sent, then look at the closest codeword to the received vector (100). This would be (000), hence we can correct one error.
- $\mathcal{C}_1$ has a generator matrix $G_1 = [111]$.
- $\mathcal{C}_2$ is a $[3, 2, 2]$ binary linear code whose generator matrix is

$$G_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- Relation: $\mathcal{C}_2 = \mathcal{C}_1^{\perp}$. Why?

Weight enumerator

- Let $A_i(\mathcal{C})$ be the number of codewords of wt i.
- For an $[n, k, d]$ code, we define the weight enumeratorof $\mathcal{C}$ by

$$W_{\mathcal{C}}(x, y) = x^n + A_d x^{n-d} y^d + A_{d+1} x^{n-(d+1)} y^{d+1} + \cdots + A_n y^n.$$

MacWilliams' identity

- $W_{\mathcal{C}_1}(x,y) = x^3 + y^3$ and $W_{\mathcal{C}_2}(x,y) = x^3 + 3xy^2$
- (MacWilliams' identiy)

$$W_{\mathcal{C}^\perp}(x,y) = \frac{1}{|\mathcal{C}|} W_{\mathcal{C}}(x + (p-1)y, x - y)$$

- Recall $\mathcal{C}_2 = \mathcal{C}_1^\perp$ and $p = 2$ as binary code.

$$\begin{aligned} W_{\mathcal{C}_2}(x,y) &= W_{\mathcal{C}_1^\perp}(x,y) \\ &= \frac{1}{|\mathcal{C}_1|} W_{\mathcal{C}_1}(x+y, x-y) \text{ (by M.I.)} \\ &= \frac{1}{2}\{(x+y)^3 + (x-y)^3\} \\ &= \frac{1}{2}(2x^3 + 6xy^2) = x^2 + 3xy^2. \end{aligned}$$

Encoding linear codes

- Let $\mathcal{C}$ be an $[n, k, d]$ code.
- Let G be a $k \times n$ generator matrix.
- Let $\mathbf{x}$ be a row vector of length k.
- Then the following mapping is an encoding processor:

$$\mathbf{x} \rightarrow \mathbf{x}G$$

- Note that length k vector $\mathbf{x}$ is mapped to a vector of larger length n. So encoding is a process to add redundancy bits (digits) to original message.
- For example, recall that $\mathcal{C}_1$ has $G = [111]$,
  $0 \rightarrow 0[111] = (000)$ and $1 \rightarrow 1[111] = (111)$

Decoding linear codes

- To simplify, we describe how to decode $\mathcal{C}_1$.
- Recall that $\mathcal{C}_2$ is the dual of $\mathcal{C}_1$ and $\mathcal{C}_2$ has a generator matrix

$$G_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- G_2 is also called a parity check matrix of $\mathcal{C}_1$.
- Let $H_1 = G_2$.

Syndrome decoding

- Suppose $\mathbf{y} = (110)$ was received using $G_1 = [111]$.
- Compute the syndrome $H_1\mathbf{y}^T$ of $\mathbf{y}$ using H_1 as follows:

$$H_1\mathbf{y}^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{(error occurred. Why?)}$$

- Note $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is the third column of H_1. Then there is an error in the third position of $\mathbf{y} = (110)$. Decode $\mathbf{y}$ as
- **y=(1 1 0)** $\rightarrow$ **x** $:=$ **y** $- (001) = (111) \in \mathcal{C}_1$.

Example: Hamming [7,4,3] binary code

- Richard Hamming had a brief teaching position at the University of Louisville in the early 1940s.
- The Hamming $[7, 4, 3]$ binary code has a generator matrix G and parity check matrix H as follows.

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

- Question: Decode a received vector $\mathbf{y} = (1\ 1\ 0\ 1\ 0\ 1\ 0)$.

Your answer?

- Compute the syndrome of **y**, which will be the first column of *H*.
- Hence we decode **y** = (1 1 0 1 0 1 0) by **x** = (0 1 0 1 0 1 0)

- Decoding algorithm for the any Hamming [n=2^r, n-r, 0] code:
- If the syndrome of a received vector is i-th column of H, then there is an error in the i-th position.

Hamming code as a perfect code

- Theorem: Any binary vector of length 7 has distance at most 1 from some Hamming codeword.

- Proof: A ball of each code has 8 vectors within distance 1. There are 16 codewords. Any two balls are disjoint since min. distance of the Hamming code is 3.
- Thus the equality holds:
  $8 \times 16 = 2^7$ = the size of all binary vectors of length 7

- Note that there are16 codewords in the Hamming [7,4,3] code:

{1000011, 0100101, 0010110, 0001111, 1100110, 1010101, 1001100, 0110011, 0101010, 0011001, 1110000, 1101001, 1011010, 0111100, 0000000, 1111111}

Hamming code as a perfect code

- Theorem: Any binary vector of length 7 has distance at most 1 from some Hamming codeword.

- Proof: A ball of each code has 8 vectors within distance 1. There are 16 codewords. Any two balls are disjoint since min. distance of the Hamming code is 3.
- Thus the equality holds:
  $8 \times 16 = 2^7$ = the size of all binary vectors of length 7

- Note that there are16 codewords in the Hamming [7,4,3] code:

{1000011, 0100101, 0010110, 0001111, 1100110, 1010101, 1001100, 0110011, 0101010, 0011001, 1110000, 1101001, 1011010, 0111100, 0000000, 1111111}

QUANTUM-RESISTANT CRYPTOGRAPHY

Insecurity of RSA and ECC

- Peter Shor(now at Applied Math Dept at MIT): Quantum algorithm solves factorization and discrete log problem in polynomial time.
- If a quantum computer is built, RSA and Elliptic curve cryptosystem will be broken.
- NIST calls for the 1st round competition of Post-Quantum Cryptography due Nov. 2018.

• Quantum-Resistant Cryptography

Lattice-based cryptography	• based on the NP-hardness of closest vector problem w.r.t. Euclidean metric on R^n • relevant to encryption and signature schemes
Code-based cryptography	• based on the NP-completeness of syndrome decoding w.r.t. Hamming metric(classical) and rank metric(new) on $\mathrm{GF}(q)^n$ • relevant for encryption schemes
Multivariate polynomial cryptography	• based on multivariate polynomials over finite fields • relevant for signature schemes
Hash-based signatures	• relevant for signature schemes

* Recently, supersingular elliptic curve isogeny cryptography was introduced for signature.

Key sizes of post-quantum cryptography

Cryptosystem	Public key size (bits)	Private key size (bits)
Lattice-based	6,956	14,000
Lattice-based NTRU	6,130	6,743
Multivariate cryptography	991,000	740,000
Hash-based cryptography	36,000	36,000
Supersingular elliptic curve isogeny cryptography	6,144	6,144
Code-based	8,373,911	92,027
DC-LRPC	2,809	A random vector can be used to recover the different parameters.

Analogue between Lattice and Code Based Crypto

Lattice based crypto	Code based crypto
$\boldsymbol{R^n}$	$\boldsymbol{GF(q)^n}$
Lattice	Linear code(=a subspace of $\boldsymbol{GF(q)^n}$)
Euclidean distance	Hamming/rank distance
Theta series	Weight enumerator
Gosset/ Leech lattice	Hamming/Golay code
Mathieu groups	Conway Simple groups
SVP, CVP	minimum distance, SDP
Ideal lattice	Cyclic codes
LWE(or LNP) public key(A, b=As+e)	Enc(m)= mG′ + e where G′=MGP
Sparse matrix	Low Density Parity Check Code

MORE ON CODE-BASED CRYPTOGRAPHY

McEliece: The First Code-Based Crypto

McEliece (1978)
The first public key cryptosystem using error correcting codes (Goppa codes)

No efficient structural attacks that might distinguish between a permuted Goppa code used by McEliece and a random code

Parameters of binary Goppa codes: $[n = 2^m, k, 2t + 1]$ **where** $t = \frac{n-k}{m}$

Original parameters: $n = 1024,\ k = 524,\ t = 50$
resulting in over 100k bits of public keysizes

Goppa code-based McEliece crypto with parameters *n=4096, k=3844, t=21*
giving 121086 bytes at the security level of 128 bits is still unbroken!

Due to large keysizes, no practical application of code-based cryptography so far


McEliece

Formal Definition of Goppa codes

The Goppa code $\Gamma(L, g(z))$ is defined by the Goppa polynomial $g(z)$, which is a polynomial of degree t over the extension field $GF(q^m)$, for q a prime, and an accessory subset L of $GF(q^m)$.

$$\begin{aligned} g(z) &= g_0 + g_1 z + \ldots + g_t z^t = \sum_{i=0}^{t} g_i z^i, \\ L &= \{\alpha_1, \ \ldots \ , \alpha_n\} \subseteq GF(q^m), \end{aligned}$$

such that $g(\alpha_i) \neq 0$ for all $\alpha_i \in L$. With a vector $c = (c_1, \ \ldots \ , c_n)$ over $GF(q)$ we associate the function

$$R_c(z) = \sum_{i=1}^{n} \frac{c_i}{z - \alpha_i}, \tag{1}$$

in which $\frac{1}{z-\alpha_i}$ is the unique polynomial with $(z - \alpha_i) \cdot \frac{1}{z-\alpha_i} \equiv 1 \pmod{g(z)}$.

Definition 2.1 *The Goppa code $\Gamma(L, g(z))$ consists of all vectors c such that*

$$R_c(z) \equiv 0 \pmod{g(z)}. \tag{2}$$

Goppa codes as Algebraic Geometry codes

Let $P_i = (\alpha_i : 1)$, $Q = (1 : 0)$ and $D = P_1 + \cdots + P_n$. If we take for E the divisor of zeros of g on the projective line, then $\Gamma(L, g) = C^*(D, E - Q)$ and

$$\mathbf{c} \in \Gamma(L, g) \text{ if and only if } \sum \frac{c_i}{X - \alpha_i} dX \in \Omega(E - Q - D).$$

This is the reason that some authors extend the definiton of geometric Goppa codes to subfield subcodes of codes of the form $C^*(D, G)$.

It is a well-known fact that the parity check matrix of the Goppa code $\Gamma(L, g)$ is equal to the following generator matrix of a generalized RS code

$$\begin{pmatrix} g(\alpha_1)^{-1} & \dots & g(\alpha_n)^{-1} \\ \alpha_1 g(\alpha_1)^{-1} & \dots & \alpha_n g(\alpha_n)^{-1} \\ \vdots & \dots & \vdots \\ \alpha_1^{r-1} g(\alpha_1)^{-1} & \dots & \alpha_n^{r-1} g(\alpha_n)^{-1} \end{pmatrix},$$

where r is the degree of the Goppa polynomial g. So $\Gamma(L, g)$ is the subfield subcode of the dual of a generalized RS code.

Generalized Reed-Solomon (GRS) codes

For some polynomial $f(z) \in \mathbb{F}_{p^m}[z]_{<k}$, pairwise distinct elements $\mathcal{L} = (\alpha_0, \ldots, \alpha_{n-1}) \in \mathbb{F}_{p^m}^n$, non-zero elements $V = (v_0, \ldots, v_{n-1}) \in \mathbb{F}_{p^m}^n$ and $0 \leq k \leq n$, GRS code can be defined as

$$GRS_{n,k}(\mathcal{L}, V) := \{ c \in \mathbb{F}_{p^m}^n | c_i = v_i f(\alpha_i) \} \tag{3.2}$$

Alternant codes are subfield subcodes of a GRS codes, i.e. they can be obtained by restricting GRS-codes to the subfield $\mathbb{F}_p$:

$$Alt_{n,k,p}(\mathcal{L}, v) := GRS_{n,k}(\mathcal{L}, V) \cap \mathbb{F}_p^n \tag{3.3}$$

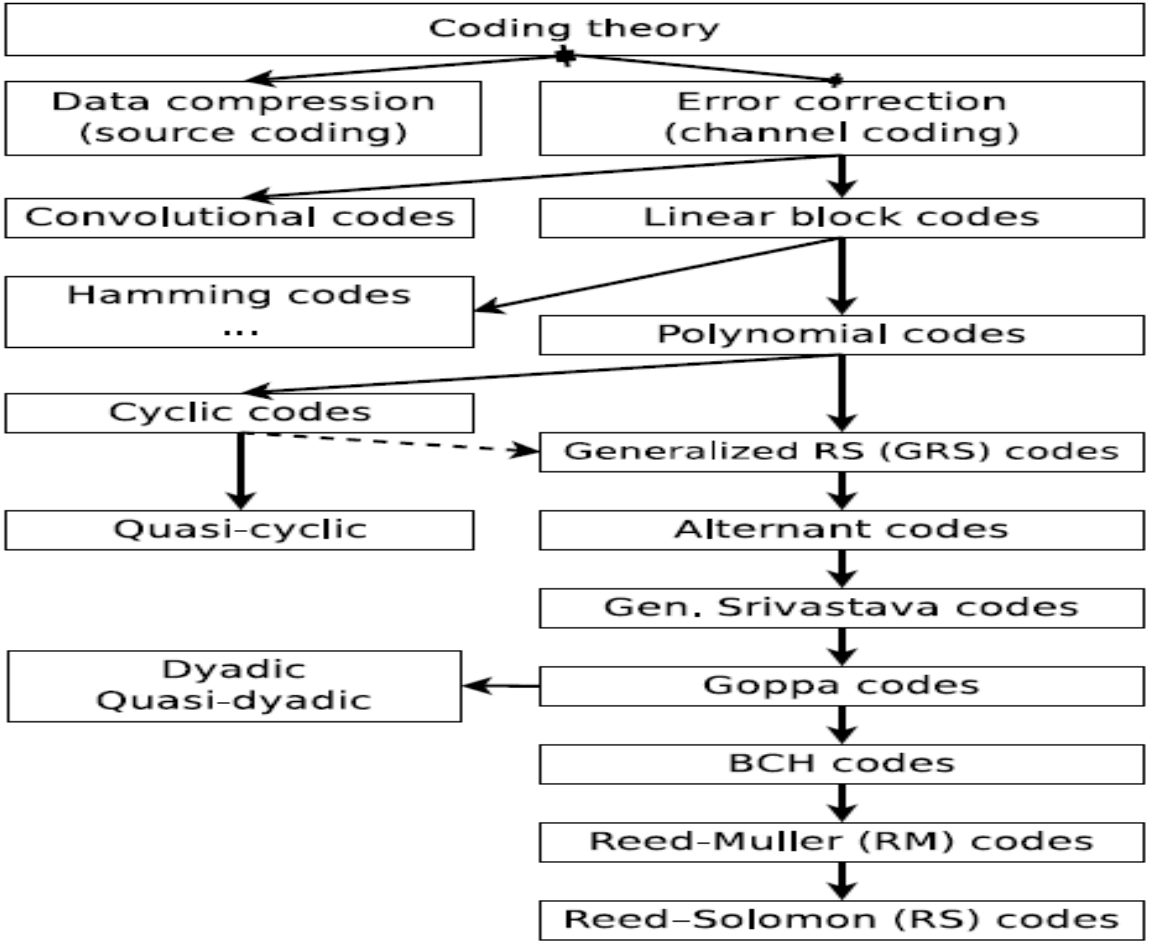


Figure 3.1: Hierarchy of code classes

NP Complete Problems in Coding Theory

- Definition: (Binary Syndrome Decoding (SD) problem; SDP)
  - **Input**: An $r \times n$ matrix H over F_2, a target binary vector $s \in {F_2}^r$, and an integer $t > 0$.
  - **Question**: Is there a binary word $x \in {F_2}^n$ of weight $\leq t$, such that $s = Hx^T$?

  - E. Berlekamp, R. McEliece, H. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Trans. Inf. Theory*, 24(3), pp. 384–386, 1978

NP Complete Problems in Coding Theory

- Definition: (q-ary Syndrome Decoding (q-SD) problem)
  - **Input**: An $r \times n$ matrix H over F_q, a target vector $s \in F_q^r$, and an integer $t > 0$.
  - **Question**: Is there a word $x \in F_q^n$ of weight $\leq t$, such that $s = Hx^T$?

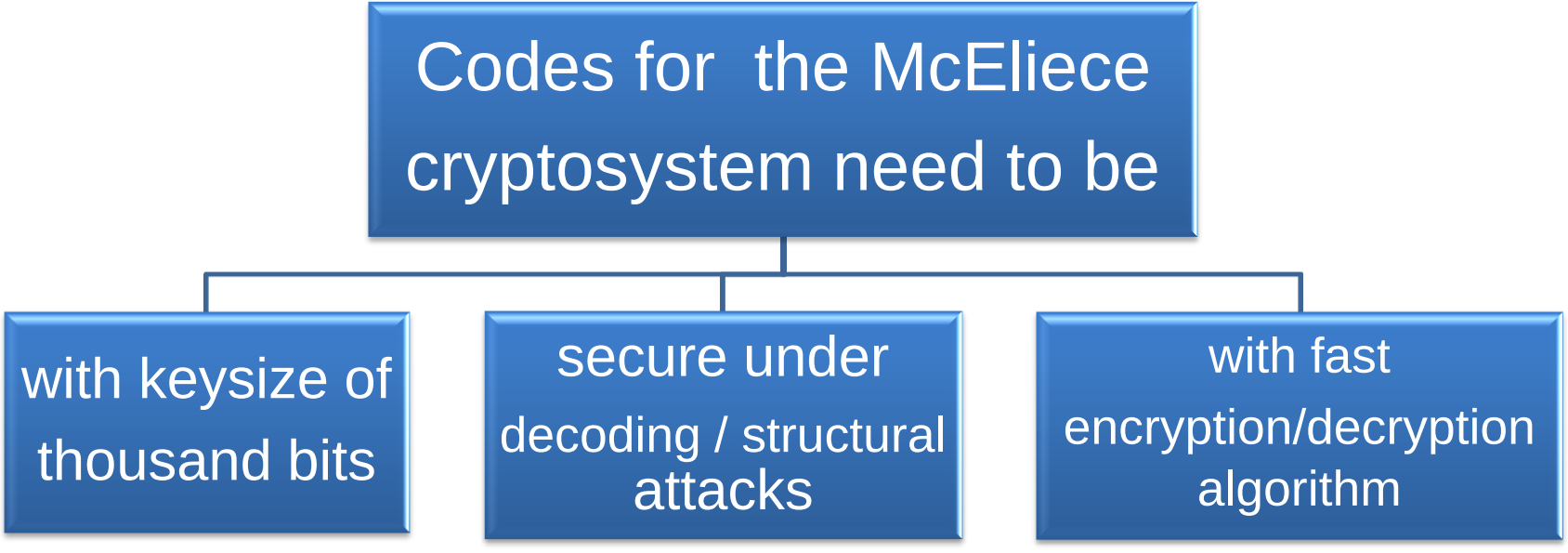
- Definition: (Goppa Code Distinguishing (GD) problem)
  - **Input**: An $(n-k) \times n$ binary matrix H.
  - **Question**: Is H a parity check matrix of a (n, k)-Goppa code or of a random (n, k)-code ?

McEliece and Niederreier PKC

- McEliece's vs. Neiderreiter's Scheme
  - $\boldsymbol{C}$: a binary code of length n and dimension k
  - G': $k \times n$ generating matrix for $\boldsymbol{C}$
  - A *an* invertible matrix and P a permutation matrix
  - H: $(n–k) \times n$ parity check matrix, $GH^T = 0$
  - $s = Hc^T$: syndrome

	McEliece	Niederreiter
Public key	$G=AG'P$	H
Plaintext	$x \in F_2^k$	$x \in F_2^n$, $w_H(x) = t$
Ciphertext	$y = xG + e$, $w_H(e) = t$	$y = Hx^T$
Ciphertext space	F_2^n	F_2^{n-k}
Used codes	*binary Goppa codes*	*generalized Reed-Solomon codes*

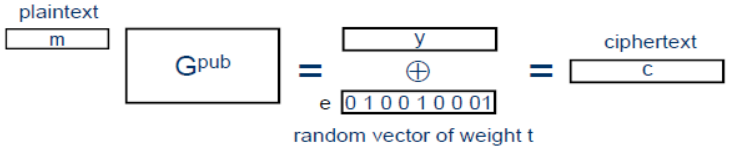
- Main problem in the McEliece cryptosystem



Security of McEliece and Niederreiter PKE is equivalent

- (In some sense)
- [Li et al. IEEE Trans. on Information Theory '94]
- McEliece $\rightarrow$ Niederreiter: Given $c=mG^{pub}+e$
- Compute H such that $H(G^{pub})^T=0$
- $Hc^T = H(G^{pub}+e)^T = H(G^{pub})^T + He^T = He^T$
- Niederreiter $\rightarrow$ McEliece: Given $c=H^{pub}m^T$
- Assume H^{pub} in the systematic form
  - Otherwise, the below argument is easily adjusted
- Then, the corresponding McEliece ciphertext is $(c|0^k)+uG^{pub}$, $\forall u \in F_2^k$
- Indeed, $H^{pub}[\ (c|0^k) + uG^{pub}\ ]^T = c$

Security of McEliece and Niederreiter PKE's



Major attacks:

- Information set decoding [Prange '62], [Lee & Brickell '87], [Leon '88], [Stern '89],…, [May & Ozerov '15]
  - Decode as if the code was random (disregard its structure)
  - Compute "m" by guessing errorless positions in "c"
- Structural attacks [Sendrier IEEE-IT'00], [Faugere, Gauthier-Umana, Otmani, Perret & Tillich, ITW'11]
  - Recover G from G^{pub}
- Running time is exponential in n
- However, G^{pub} can be distinguished from random [FGOPT, ITW'11], if k/n (rate of the code) is "high" (i.e., when k is close to n) – the parameters used for Courtios-Finiasz-Sendrier digital signature [Asiacrypt '01]

Recommended Parameters

- PQCRYPTO EU project (https://pqcrypto.eu.org)
  Initial recommendations of long-term secure post-quantum systems (September 2015)
- McEliece PKE (binary Goppa codes):
  n=6960, k=5413, t=119
  for 128-bit post-quantum security
- Public key size (systematic) ≈ 1 megabyte

Shorter Keys?

- Quasi-cyclic or quasi-dyadic codes and other variants
  - Idea: the generator matrix is defined by the first row
- Security is substantially reduced (makes structural attacks much easier)
  [Faugere et al. Eurocrypt ‘10, Designs Codes & Crypto ‘16]
- Possibly okay, if applied to LDPC/MDPC codes

Using LDPC Codes

- Original idea [Monico et al., ISIT '00]
- Quasi-cyclic variant [Baldi et al., ICC '07], …
  - Springer book by Marco Baldi "QC-LDPC Code-Based Cryptography", 2014
- Idea: In $G^{pub}=SGZ$, take G as an LDPC code, and Z is a sparse invertible matrix
- => eZ^{-1} in decryption will increase # errors
- but G will take care of it, if Z is "reasonably sparse"
- **Problem:** A successful search for low-weight codewords in the dual of G^{pub} will enable one to decode it

Using MDPC Codes

- A new proposal by [Misoczki et al., ISIT '13] is to use Moderate Density Parity Check (MDPC) codes
- Idea: A higher density of parity-check matrix is more important that the error-correcting capability
- Use G^{pub}=SGP, with P as permutation but G is an MDPC code
- Good news: Both structural attacks and decoding attacks boil down to the same problem: Finding minimum weight codewords

Decoding is equivalent to finding low-weight codewords

- Consider $c=mG^{pub}+e$
- Add "c" as a row to G^{pub}:

$$G' = \begin{pmatrix} G^{pub} \\ c \end{pmatrix}$$

- Now, finding "e" is equivalent to finding the minimum weight codeword in G'

QC-MDPC McEliece

- n=65542, k=32771, t=264 [Misoczki et al., ISIT '13] for 128 bit post-quantum security
- Public key size = 32771 bits
- Admits efficient implementations also for reconfigurable hardware and embedded microcontrollers
  [von Maurich et al., ACM Trans. Embedded Computing Systems, '15]

Other Codes?

(Just a few examples)

May be used:

- Algebraic geometric codes
  [Janwa and Moreno, Designs Codes & Crypto '96]
- Generalized Srivastava codes [Cayrel et al., PKC'12]

Should NOT be used:

- Generalized Reed-Solomon codes
  [Sidel'nikov and Shestakov, Discrete Math. & Appl. '92]
- Reed-Muller codes (some parameters)
  [Minder and Shokrollahi, Eurocrypt'07]

ATTACKS ON CODE BASED CRYPTOGRAPHY

Attacks on code based crypto

There are mainly three kinds of attacks.

(1) Structural attacks
- find keys (i.e., private keys from public keys)

(2) Decoding attacks
- find a message from a cipher attack
- related to finding low-weight codewords

(3) Side-channel attacks
- against hardware implementations

Structural attacks

- Structural attacks use the code structure in order to break the code-based crypto.
- Since different codes have different structures, structural attacks are different in general.
- There are many structural attacks.

  1) Sidelnikov-Shestkov's attack against the Niederreiter PKC using GRS(generalized Reed-Solomon) codes
     ["On Cryptosystems based on generalized Reed-Solomon codes", Discrete Math, 1992]

Structural attacks

2) Stern's algorithm on Reed-Muller codes
["A method for finding codewords of small weight", LNCS 388, 1988]

3) Otmani, Tillich, & Dallot's attack against quasi-cyclic codes (e.g, AC LDPC codes)
["Cryptanalysis of two McEliece cryptosystems based on quasi-cyclic codes, preprint 2008]
- idea: exploits the QC structure to find a punctured version of the secret key, and then uses Stern's algorithm to reconstruct the entire secret.

Structural attacks

4) Overbeck's attack against the McEliece PKC based on rank- metric codes
["Structural attacks for public key cryptosystems based on Gabidulin codes", J. of Cryptology, 2008]

5) Faugere, Otmani, Perret, & Tillich's attack against McEliece PKC using non-binary QC & QD(quasi-dyadic) code
["Algebraic cryptanalysis of McEliece variants with compact keys', Eurocrypt, LNCS 6110, 2010 and in SCC '10, 2010 as an extension]

Plain-ISD

Let $\mathbf{u} \in \mathbb{F}_2^k$ be an unknown information sequence and

$$\mathbf{r} = \begin{pmatrix} r_1 & r_2 & \cdots & r_n \end{pmatrix} = \mathbf{uG} + \mathbf{e} \in \mathbb{F}_2^n$$

a received codeword encoded by an $[n, k]$ linear code with generator matrix

$$\mathbf{G} = \begin{pmatrix} \mathbf{g}_1^{\mathrm{T}} & \mathbf{g}_2^{\mathrm{T}} & \cdots & \mathbf{g}_n^{\mathrm{T}} \end{pmatrix}. \tag{2.2}$$

Suppose we pick an information set I in $\{1,2,\ldots,n\}$ of size k.

$$\phi(\mathbf{r}) = (r_i)_{i\in\mathcal{I}} = (\mathbf{uG} + \mathbf{e})_{i\in\mathcal{I}} = (\mathbf{uG})_{i\in\mathcal{I}} + (\mathbf{e})_{i\in\mathcal{I}}.$$

Whenever the selected information set $\mathcal{I}$ is non-corrupted, i.e., $(\mathbf{e})_{i\in\mathcal{I}} = \mathbf{0} \iff \mathcal{I} \cap \mathrm{supp}(\mathbf{e}) = \varnothing$, then the information symbols can be obtained by computing

$$\phi(\mathbf{r})\phi(\mathbf{G})^{-1} = \phi(\mathbf{uG})\phi(\mathbf{G})^{-1} = \mathbf{u}.$$

Plain-ISD

Denote the iteration success probability $P_{\text{Plain-ISD}}$. Then,

$$P_{\text{Plain-ISD}}(n,k,w) \stackrel{\text{def}}{=} \mathbb{P}\left(\mathcal{I} \cap \text{supp}(\mathbf{e}) = \varnothing\right)$$

$$= \frac{\mathcal{S}_q(n-k,w)}{\mathcal{S}_q(n,w)} = \binom{n-k}{w}\binom{n}{w}^{-1}$$

Lee-Brickell's generalized ISD

- Lee-Brickell allows a set of $p \in \{0,1,\ldots,w\}$ errors in the information set.
- If $p = 0$, then it is a plain-ISD.
- It is known that $p = 2$ is optimal in the binary case.
- The probability of success in one iteration is

$$\mathbb{P}\left(|\mathcal{I} \cap \mathrm{supp}(\mathbf{e})| = p\right) = \binom{n-k}{w-p}\binom{k}{p}\binom{n}{w}^{-1}$$

Stern's generalized ISD

- It allows a set of $p \in \{0,1,\ldots,w\}$ errors in the information set and takes $l \in \{0,1,\ldots,n-k\}$.
- Suppose the information set $I = I_1 \cup I_2$.
- The probability of success in one iteration is

$$P_{\text{Stern-ISD}}(n,k,w,p,l) = \binom{k/2}{p}^2 \binom{n-k-l}{w-2p} \binom{n}{w}^{-1}$$

Error distribution in Stern's algorithm

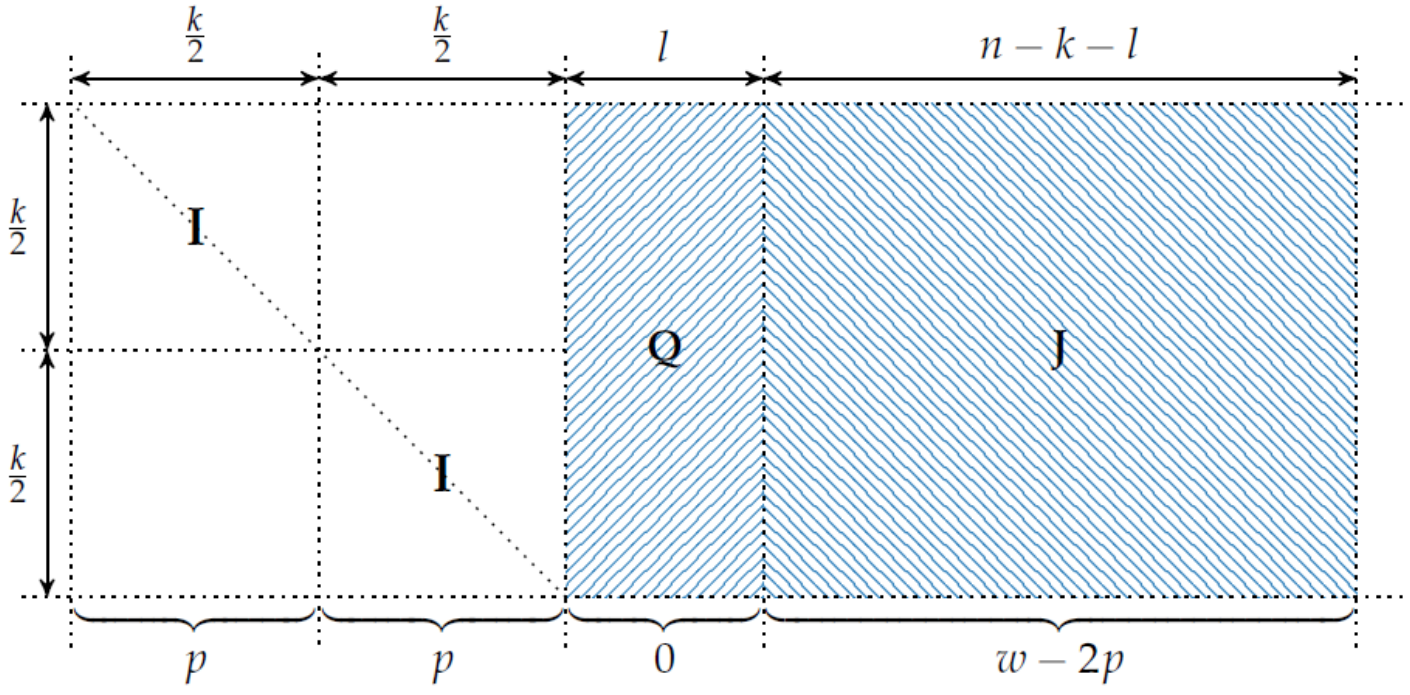


Figure 3.1: The different fields used in Stern's algorithm and their error distributions.

Finiasz-Sendrier-ISD and others

$$P_{\text{FS-ISD}}(n,k,w,p,l) = \binom{(k+l)/2}{p}^2 \binom{n-k-l}{w-2p} \binom{n}{w}^{-1}$$

References

- Finiasz and Sendrier ["Security bounds for the design of code-based cryptosystems", Asiacrypto, 2009]
- Canteaut and Chabaud ["A new algorithm for finding minimum weight words in a linear code", IEEE-IT, 1998]
- Bernstein, Lange, & Peters ["Attacking and defending the McEliece cryptosystem", PQCrypto, Vol. 5299 of LNCS, 2008]
- Peters ["Information-set decoding for linear codes over F_q", In PQCrypto, 2010]

McNie: a new code-based cryptography

- Our McNie is a new code-based public key cryptosystem which is less vulnerable against currently known structural attacks.
- McNie is one of the 64 algorithms which passed round 1 of 2017 NIST Competition for Post-Quantum Cryptography.
- McNie can use Hamming weight or rank weight in general.

McNie: Key generation

- Consider Hamming weight or rank weight.
- **Secret key:** (H, P, S, Φ_H)
  H: a parity check matrix for an $[n, k]$ code C over $\mathbb{F}_{q^m}$
  P: an $n \times n$ permutation matrix
  S: an $(n-k) \times (n-k)$ invertible matrix over $\mathbb{F}_{q^m}$
  Φ_H: an efficient decoding algorithm for C which corrects errors of weight up to r

- **Public key:** (G', F)
  G': Generator matrix for a random $[n, l]$ code over $\mathbb{F}_{q^m}$
  $F = G' P^{-1} H^T S$

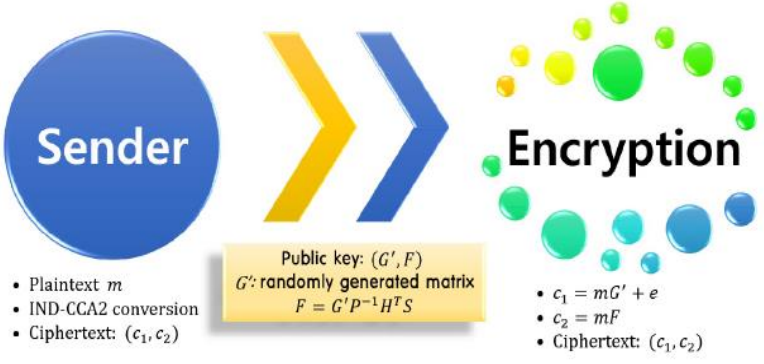
McNie: Encryption

Message: $\mathbf{m} \in \mathbb{F}_{q^m}^{l}$

- Randomly generate $\mathbf{e} \in \mathbb{F}_{q^m}^{n}$ of weight r
- $Enc(\mathbf{m}) = (\mathbf{c_1}, \mathbf{c_2})$

  $\mathbf{c_1} = \mathbf{m}G' + \mathbf{e}$

  $\mathbf{c_2} = \mathbf{m}F = \mathbf{m}G'P^{-1}H^TS$



McNie: Decryption

Received vector: $\mathbf{c} = (\mathbf{c_1}, \mathbf{c_2})$

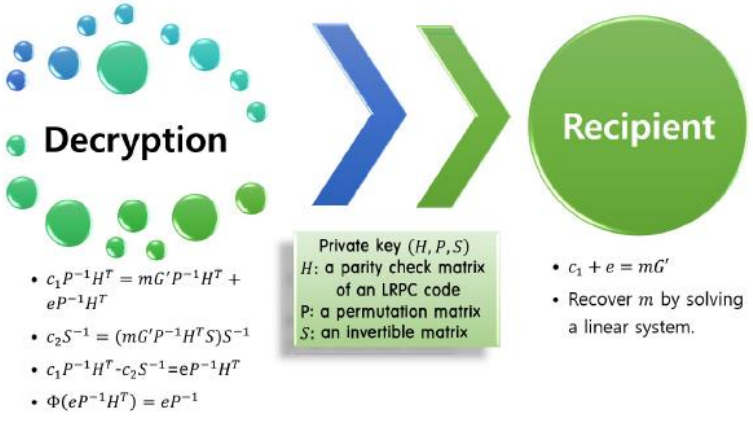
- Compute

$$\begin{aligned} \mathbf{s}' &= \mathbf{c_1} P^{-1} H^T - \mathbf{c_2} S^{-1} \\ &= (\mathbf{m} G' + \mathbf{e}) P^{-1} H^T \\ &\quad -(\mathbf{m} G' P^{-1} H^T S) S^{-1} \\ &= \mathbf{e} P^{-1} H^T \\ \mathbf{e}' &= \Phi_H(\mathbf{s}') = \mathbf{e} P^{-1} \\ \mathbf{e} &= \mathbf{e}' P \end{aligned}$$

- Solve the system

$$\mathbf{m} G' = \mathbf{c_1} - \mathbf{e}$$

to recover $\mathbf{m}$.



Security reduction

The McNie cryptosystem is based on the following hard problem:

(Rank) Syndrome Decoding Problem

Given an $(n-k) \times n$ matrix H, a vector $\mathbf{s}$ of length n and a positive integer r, find a vector $\mathbf{e}$ of (rank) weight r such that $\mathbf{s} = \mathbf{e}H^T$.

Proof.

$$\begin{aligned}(\mathbf{c_1}, \mathbf{c_2}) &= (\mathbf{m}G' + \mathbf{e}, \mathbf{m}F) \\ &= \mathbf{m}[G'|F] + (\mathbf{e}|\mathbf{0}) \\ &= \mathbf{m}G_1 + \mathbf{e_1}\end{aligned}$$

where $G_1 = [G'|F]$ and $\mathbf{e_1} = (\mathbf{e}|\mathbf{0})$.
Since G' and F are public keys, the parity check matrix for the code generated by G_1 can be computed.
This reduces to the (rank) syndrome decoding problem.

Apply McNie to rank metric codes

Let $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be a basis for $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$.

$$c = (c_1, \ldots, c_n) \in \mathbb{F}_{q^m}^n \Leftrightarrow \bar{c} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix}, \quad c_j = \sum_{i=1}^{m} c_{ij}\alpha_i$$

- **rank weight**: $w_R(c) = Rank(\bar{c})$
- **rank distance**: $d_R(c, c') = Rank(\bar{c} - \bar{c}')$

A *rank metric code* is an $[n, k]$ code over $\mathbb{F}_{q^m}$ equipped with the rank metric.

A family of rank metric codes used in McNie:

A *Low Rank Parity Check (LRPC)* code of rank d is an $[n, k]$ code over $\mathbb{F}_{q^m}$ that has for its parity check matrix an $(n-k) \times n$ matrix $H = (h_{ij})$ such that the sub-vector space of $\mathbb{F}_{q^m}$ generated by its coefficients h_{ij} has dimension at most d.

Updated parameters for 3,4-QC LRPC codes

n	l	k	d	r	m	q	failure	Key Size (bytes)	security
120	80	80	3	8	53	2	-23	795	128
138	92	92	3	10	67	2	-25	1156	192
156	104	104	3	12	71	2	-27	1385	256

Table: New suggested parameters for McNie using 3-quasi-cyclic LRPC code

n	l	k	d	r	m	q	failure	Key Size (bytes)	security
92	46	69	3	10	59	2	-36	849	128
112	56	84	3	13	67	2	-38	1173	192
128	64	96	3	16	73	2	-36	1460	256

Table: New suggested parameters for McNie using 4-quasi-cyclic LRPC code

A modified McNie with Gabidulin

- We have modified McNie to make an enhanced McNie with Gabidulin whose public key size is about 1.4 KB at the 128 security level. We will submit our result soon.

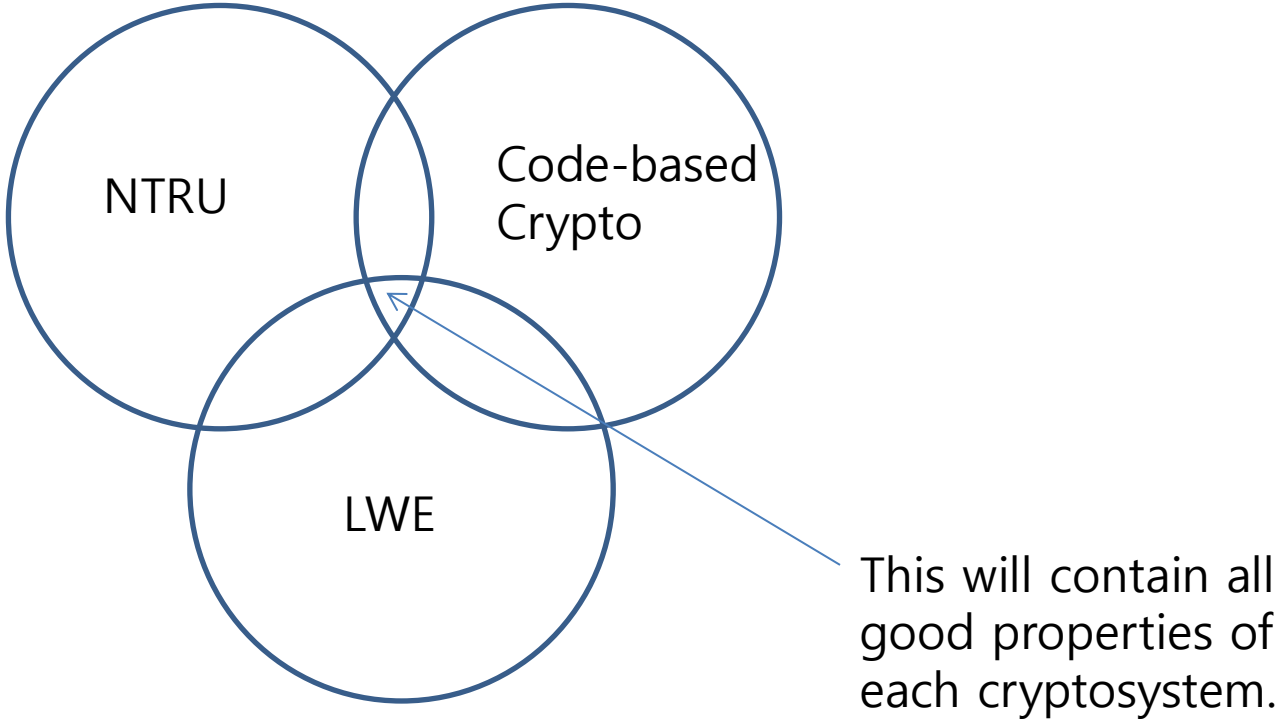
Table 1: Suggested parameters for McNie2-Gabidulin

n	k	l	q	m	r	Sec	PK	SK	CT
24	12	22	2	41	6	128	1.476	0.308	0.185
32	16	24	2	53	8	192	2.756	0.530	0.318
36	18	29	2	59	9	256	4.116	0.664	0.399

Connection between Ouroboros-R and McNie

- Gaborit posted a message recovery attack on the McNie cryptosystem that significantly reduced the security of the original suggested parameters.
- To avoid the this attack, we modify the encryption algorithm by introducing an error $\mathbf{e_2}$ on $\mathbf{c_2}$.
- We submitted a below joint paper.
  P. Gaborit, L. Galvez, A. Hauteville, J.-L. Kim, M. J. Kim, Y.-S. Kim, "Dual-Ouroboros: An improvement of the McNie Scheme", submitted to Advances in Mathematics of Communication.

- My personal opinion



THANK YOU^^

References

- Aragon, N., Gaborit, P., Hauteville, H., Tillich, J.-P.: Improvement of generic attacks on the rank-syndrome decoding problem. 2017. <hal-01618464>
- Bernstein, D.J., Lange, T., and Peters, C.: Attacking and defending the McEliece cryptosystem. In Proceedings of the 2nd International Workshop on Post-Quantum Cryptography, PQCrypto '08, pp. 31–46, Springer-Verlag, Berlin, Heidelberg (2008).
- Gaborit, P., Ruatta, O., Schrek, J., Tillich, J. P., Zémor, G.: Rank based Cryptography: a credible post-quantum alternative to classical crypto. In NIST 2015: Workshop on Cybersecurity in a Post-Quantum World 2015 (2015).
- P. Gaborit, L. Galvez, A. Hauteville, J.-L. Kim, M. J. Kim, Y.-S. Kim, "Dual-Ouroboros: An improvement of the McNie Scheme", submitted to Advances in Mathematics of Communication.
- Lau, T. S. C., Tan, C. H.: Key Recovery Attack on McNie based on Low Rank Parity Check Codes and its Reparation, IWSEC 2018, Sep. 3-5, 2018
- Misoczki, R., Tillich, J. P., Sendrier, N. and Barreto, P. S.: MDPC-McEliece: New McEliece variants from moderate density parity-check codes. IEEE International Symposium on Information Theory - ISIT 2013, pp. 2069-2073 (2013).
- Misoczki, R., and Barreto, P. S.: Compact McEliece keys from Goppa codes. In Selected Areas in Cryptography, pp. 376–392 (2009)