Lattice Coding Theory
Lecture Notes

Brian M. Kurkoski

2023 November 17

Lattice Coding Theory Lecture Notes
2022
Brian M. Kurkoski
kurkoski@jaist.ac.jp

Cover image: Flickr/Matthew T Rader Photography

Contents

Preface

Welcome

Lattices are error-correcting codes over the real numbers. Real numbers describe the physical world, particularly the physical media used in communications. In wireless communications for example, when two signals are transmitted at the same time they will superimpose — that is, they add, making lattice codes a natural fit for wireless communications, particularly multiuser communications. In the physical world, 1 plus 1 is 2, and it is the same for lattices.

Lattices have captured the attention of mathematicians since the late 19th century. Led by Conway and Sloane starting in the 1970s developed lattices, including numerous constructions, particularly in low dimensions with elegant algebraic structure. Throughout the 1990s, Forney, Calderbank and others pioneered the further development of lattices for communications, as well as non-lattice coded modulation. Lattices are useful in the first of their two practical communication applications, to bandwidth efficient communications, which were voice band modems, because telephone bandwidth requires the use of efficient modulation techniques.

Information theoretic results using lattices have appeared, many inspired by the work of Ram Zamir and his colleagues. These results show that lattices can achieve the capacity, or nearly so, for a variety of communication systems. Typical examples are communications scenarios such as point-to-point communications, multi-user communications, multiple-antenna communications and distributed sourcing coding. Characteristic of information theoretic results is the use of lattices which are capacity-achieving, but have little other structure.

For these information theoretic results to be realized in practice, specific constructions of finite dimension are needed. Error-correcting codes underwent a true paradigm shift with the 1993 invention of turbo codes, and the subsequent rediscovery of low-density parity check codes. Since the 2000s, various lattices versions of high dimension have been successfully developed, inheriting the strengths of turbo codes and low-density parity-check codes.

The central goal of *Lattice Coding Theory* is to bring together classical and recent lattice constructions, as many important recent results are distributed in the academic literature. The focus is on lattices for reliable communications, that is, where the lattice is being used as an error-correcting code in a system with noise. Lattices constructions are selected for their importance and promise

in future communications systems.

Reflected in the title, *Lattice Coding Theory* uses the framework of coding theory. Finite-field codes, such as Hamming codes, convolutional codes, BCH codes and LDPC codes, are well-studied and their use in practical communication systems is well established. Many lattices have direct connections with finite-field codes, particularly through Construction A and Construction D/D'. Many properties of codes extend naturally to lattices. Finite-field codes are often written with a matrix representation, and likewise matrix representations for lattices are emphasized here.

By presenting lattice coding theory using finite-field error-correcting codes, those already familiar with error-correcting codes can quickly grasp lattices. Figures of two and three-dimensional lattices can meaningfully and accurately describe important concepts. Group-theoeric structure not needed for applications will be de-emphasized in our drive to develop lattices for wireless communications. The content is not practically itself, but rather presents theory in a way that is understandable to practitioners. As with many books on coding theory, idealized communication channel models are used to illustrate the lattice properties.

Finally, lattices are inherently interesting. Part of my enthusiasm stems from the fact that lattices are elegant and both visually and mathematically pleasing. It is interesting to note that for finite-field codes, two-dimensional figures are more conceptual than accurate

Notation

While the emphasis of these lecture notes is on conceptual understanding, some mathematics is necessary. An attempt has been made to use consistent notation throughout, some of which is summarized in the table below.

$\mathbf{x}$	vectors are indicated by lower case bold face, for example: $$\mathbf{x} = [x_1, x_2]$$
$[\cdot]^{\mathrm{t}}$	vector and matrix transpose, for example: $$\mathbf{x}^{\mathrm{t}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$\mathbf{x} + \mathbf{y}$	vector addition
$\mathbf{G}$	matrices are indicated by upper case bold face, for example: $$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$
$\mathbf{I}_m$	m-by-m identity matrix, for example: $$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$\det(\mathbf{G})$	the determinant of a square matrix $\mathbf{G}$.
$\mathcal{A}$	sets are indicated by calligraphic script, for example $\mathcal{A} = \{0, 1, 2, 3, 4\}$
$\lvert\mathcal{A}\rvert$	cardinality of a set, for example $\lvert\mathcal{A}\rvert = 5$
$\mathcal{A} \setminus a$	set subtraction, for example $\mathcal{A} \setminus \{0, 2\} = \{1, 3, 4\}$
$\mathcal{G}, \mathcal{H}$	a group $\mathcal{G}$ and a subgroup $\mathcal{H}$
$\mathcal{R}$	a ring is indicated
$\mathbb{F}$	a polynomial ring
$\mathbb{F}_q$	a field of size q.
$\mathbb{Z}$	the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{Z}^n$	the set of all integer vectors in n-dimensions, for example $(-4, 0, 1, 3) \in \mathbb{Z}^4$
$\mathbb{R}$	the set of real numbers

Λ	a lattice
$\mathcal{R}$	a shaping region in n dimensions $\mathcal{R} \subset \mathbb{R}^n$. A lattice code is $\Lambda \cap \mathcal{R}$
$\mathbb{R}^n$	the n-dimensional Euclidean space
$\|\|\mathbf{x}\|\|^2$	squared length of vector $\mathbf{x}, \sum_{i=1}^{n} x_i^2$
$\|\|\mathbf{x}-\mathbf{y}\|\|^2$	squared Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$
$\binom{n}{k}$	n choose k, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Linear Algebra

The following are properties of matrices that students should already be familiar with. Let $\mathbf{A}$, etc. be matrices:

1. Matrix transpose: $(\mathbf{AB})^{\mathrm{t}} = \mathbf{B}^{\mathrm{t}}\mathbf{A}^{\mathrm{t}}$
2. Matrix inverses: $(\mathbf{A}^{\mathrm{t}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{t}}$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
3. Orthogonal matrix $\mathbf{Q}$ satisfies $\mathbf{Q}^{\mathrm{t}}\mathbf{Q} = \mathbf{I}$. Equivalently, $\mathbf{Q}^{\mathrm{t}} = \mathbf{Q}^{-1}$.
4. If $\mathbf{A}$ has only real entries, then $\mathbf{A}^{\mathrm{t}}\mathbf{A}$ is a positive semidefinite matrix.

Chapter 1

Error Correcting Codes

This chapter gives an overview of coding theory: abstract algebra, principles of block codes, and Euclidean-space codes. These are the concepts that will be used throughout the course.

1.1 Error Correcting Code Concepts

1.1.1 Communications System

Error-correcting codes provide reliable communications over an unreliable channel. The central idea is that a transmitter wants to send a message to a receiver using a communication channel which causes errors to occur. A model of a communications system is shown in Fig. 1.1. It consists of five parts: an information source, an encoder, a channel, a decoder, and an information sink. The information source produces a message $\mathbf{u}$. An encoder transforms the message to a codeword $\mathbf{c}$, adding redundancy to the message. The codeword $\mathbf{c}$ is transmitted over an unreliable channel, meaning that errors are added to $\mathbf{c}$. The output of the channel is $\mathbf{y}$ and a decoder, with knowledge of the encoding, makes an estimate $\widehat{\mathbf{c}}$ or $\widehat{\mathbf{u}}$ from $\mathbf{y}$. This estimate is delivered to the information sink. The goal is that $\widehat{\mathbf{u}}$ should be the same as $\mathbf{u}$ (or, $\widehat{\mathbf{c}}$ should be the same as $\mathbf{c}$).

To show how error-correcting codes work, two examples are given, a repeat code and a Hamming code.

1.1.2 Example: Repeat Code

Consider the repeat-by-seven code. The information source produces $u = 0$ or $u = 1$, and encodes to $\mathbf{c}$ according to:

u	$\mathbf{c}$
0	0000000
1	1111111

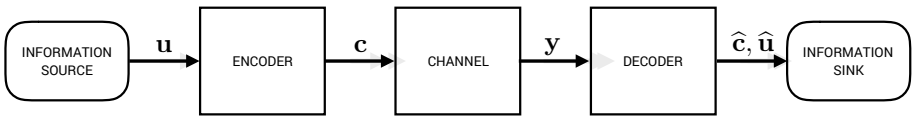


Figure 1.1: Model of a communication system with encoder, channel and decoder.

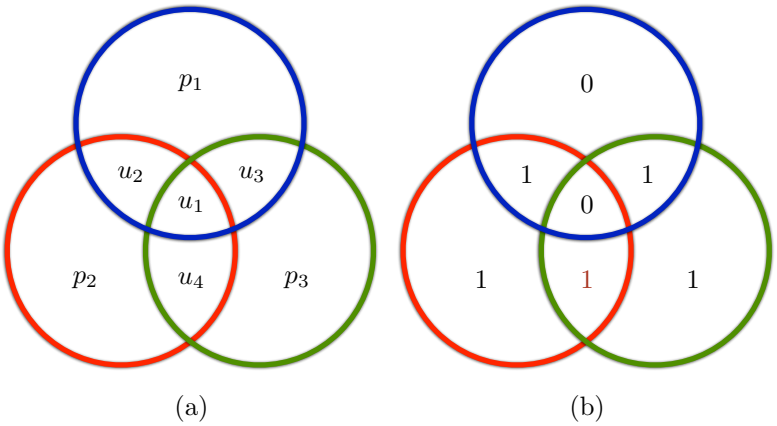


Figure 1.2: Graphical representation of a Hamming code. (a) Information bits are u_1, u_2, u_3, u_4 and parity bits are p_1, p_2, p_3. The number of bits inside each circle must be even. (b) The sequence $\mathbf{y} = [0, 1, 1, 1, 0, 1, 1]$ is received. A single error can be corrected.

Suppose $u = 1$, and the encoder transmits the codeword:

$$\mathbf{c}_1 = [1111111]. \tag{1.1}$$

The channel introduces three errors, such that the channel output is:

$$\mathbf{y} = [0110101]. \tag{1.2}$$

The decoder uses a "majority vote" decoding algorithm. Seeing that there are more 1's than 0's, the decoder chooses:

$$\widehat{\mathbf{c}} = [1111111], \tag{1.3}$$

as the estimated codeword, which corresponds to $\widehat{u} = 1$. In this case, the decoder is correct.

The repeat-by-seven code can correct any pattern of 0, 1, 2 or 3 errors. But, it used the channel seven times to transmit only one bit of information. The following code uses the channel more efficiently.

1.1.3 Example: Hamming Code

Consider the following code, called the (7,4) Hamming code. The source produces one of 16 symbols, with $\mathbf{u} = [u_1, u_2, u_3, u_4]$ where $u_i \in \{0, 1\}$ are four

binary source symbols. This will be encoded to a codeword with 7 bits:

$$\mathbf{c} = [u_1, u_2, u_3, u_4, p_1, p_2, p_3], \tag{1.4}$$

where p_1, p_2 and p_3 are called *parity bits*. A graphical representation of the code is shown in Fig. 1.2-(a), where each circle contains four bits. The seven code bits must satisfy the following condition:

> The number of 1's inside each circle must be even. Equivalently, the modulo-2 sum of the bits inside each circle must be 0.

The parity bits are selected to satisfy this condition.

Suppose $\mathbf{u} = [0, 1, 1, 0]$. Then $p_1 = 0$, $p_2 = 1$, $p_3 = 1$ makes the number of 1's inside each circle even, and the transmitted codeword is:

$$\mathbf{c} = [0, 1, 1, 0, 0, 1, 1]. \tag{1.5}$$

After transmission through the channel, one error occurs:

$$\mathbf{y} = [0, 1, 1, 1, 0, 1, 1], \tag{1.6}$$

which is illustrated in Fig. 1.2-(b). The decoder knows the encoding rule. The blue circle has an even number of ones, and the red and green circles have an odd number of ones. Since only u_4 is inside the red and green circles, but not the blue circle, the decoder changes y_4 from $1 \to 0$, so that the estimated codeword is:

$$\widehat{\mathbf{c}} = [0, 1, 1, 0, 0, 1, 1]. \tag{1.7}$$

In this case, there was only one error, and the decoder produced the correct codeword. See also Example 3.18.

This Hamming code can transmit 4 information source bits for seven channel uses, so it is more efficient than the repeat code. But, the repeat code is more powerful, because it can correct 3 errors, while the Hamming code can only correct one error. This is a fundamental tradeoff in coding theory: the efficiency of the code (later defined as the code's rate) versus its error correction capability (later defined as the code's minimum distance).

SSQ 1.1. For the (7,4) Hamming code in Fig. 1.2, find the estimated codeword $\widehat{\mathbf{c}}$ for each $\mathbf{y}$:

- $\mathbf{y} = [0, 0, 0, 0, 0, 0, 1]$,
- $\mathbf{y} = [0, 0, 0, 1, 0, 0, 0]$

Go to the course website for solutions and more SSQs.

1.2 Communications System Model

A formal description of the system in Fig. 1.1 is given. An *information source* produces one of M symbols $u \in \{1, 2, \ldots, M\}$. Each of the M symbols is equally likely.

Definition 1.1. An *error-correcting code* $\mathcal{C}$ consists of M codewords:

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\} \tag{1.8}$$

where each codeword (or constellation point) is an n-tuple:

$$\mathbf{c} = [c_1, c_2, \ldots, c_n]. \tag{1.9}$$

Each c_i is from a specified alphabet, often the binary alphabet.

Definition 1.2. The *code rate* R of a code is:

$$R = \frac{1}{n} \log_2 M \tag{1.10}$$

Example 1.1. For the repeat code in Subsection 1.1.2, $M = 2$ and $n = 7$, so the rate is $R = \frac{1}{7}$.

For the Hamming code in Subsection 1.1.3, $M = 16$ and $n = 7$, so the rate is $R = \frac{4}{7}$.

The code rate represents the number of information bits per codeword symbol. Consider the case that M is a power of 2, say $M = 2^k$. Then the message can be described using k bits, and the code rate is $R = k/n$, the number of bits of information, divided by the number of times the channel was used. The higher the rate, the more information the code can carry.

An *encoder* is a mapping from information symbols to codewords. If the information alphabet is $\mathcal{U}$, then:

$$\text{Encode} : \mathcal{U} \to \mathcal{C}.$$

The information source symbol u is one-to-one encoded to its codeword $\mathbf{c}_u$. For the repeat code $\mathcal{U} = \{0, 1\}$ and for the Hamming code $\mathcal{U} = \{0000, 0001, \ldots, 1111\}$.

The codeword is transmitted over a noisy *channel*. The output of the channel is a sequence $\mathbf{y}$ of n symbols:

$$\mathbf{y} = [y_1, y_2, \ldots, y_n]. \tag{1.11}$$

The channel output may be continuous, $y_i \in \mathbb{R}$, or discrete, y_i from a finite set. A probabilistic model of the channel is used, meaning that $\Pr[y_i|c_i]$ is given, and examples are given in Section 1.3.

The *decoder* receives this sequence, and then outputs an estimate of the information source $\widehat{\mathbf{u}}$, or an estimate of the transmitted codeword $\widehat{\mathbf{c}} = [\widehat{c}_1, \widehat{c}_2, \ldots, \widehat{c}_n]$. If the channel output alphabet is $\mathcal{Y}$, then the decoder is the mapping:

$$\text{Decode} : \mathcal{Y}^n \to \mathcal{C}. \tag{1.12}$$

Since the encoding is one-to-one, if we know $\widehat{\mathbf{c}}$ then the corresponding $\widehat{\mathbf{u}}$ is easily found. The decoder knows the codebook $\mathcal{C}$, and generally finds the codeword $\widehat{\mathbf{c}} \in \mathcal{C}$ which is "closest" in some sense, to $\mathbf{y}$.

The central objective is that the decoder output $\widehat{\mathbf{u}}$ should be the same as the encoder input $\mathbf{u}$, with high probability. If $\mathbf{u} = \widehat{\mathbf{u}}$, then the decoded succeeded with no errors. If $\mathbf{u} \neq \widehat{\mathbf{u}}$, then a decoding error occurred.

1.3 Channel Models

The channel shown in Fig. 1.1 introduces noise into the communication system. Two channel models are considered. The discrete memoryless channel (DMC) has discrete outputs. The additive white Gaussian noise (AWGN) channel has continuous-valued outputs.

The general channel has input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and output probability $p_{\mathsf{Y}|\mathsf{X}}(y|x)$. In the vector channel model is n uses of the channel, so the channel input sequence is $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and the channel output sequence is $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. The channel is described by joint conditional distribution $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. Because the channel is memoryless, we have:

$$\begin{aligned} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= p_{\mathsf{Y}|\mathsf{X}}(y_1|x_1)p_{\mathsf{Y}|\mathsf{X}}(y_2|x_2)\cdots p_{\mathsf{Y}|\mathsf{X}}(y_n|x_n) \\ &= \prod_{i=1}^{n} p_{\mathsf{Y}|\mathsf{X}}(y_i|x_i). \end{aligned}$$

Note that $\mathcal{X}$ is both the codebook alphabet and the channel input alphabet.

1.3.1 Discrete Memoryless Channel

Definition 1.3. A *discrete memoryless channel* (DMC) consists of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$ and a conditional probability distribution $p_{\mathsf{Y}|\mathsf{X}}(y|x)$.

The codeword has n symbols $\mathbf{c} = [c_1, c_2, \ldots, c_n]$. We use c_i to denote a codeword symbol and x_i to denote the channel input. For DMCs, these two are equal:

$$c_i = x_i, \tag{1.13}$$

for $i = 1, \ldots, n$, so the channel input is $\mathbf{x} = [x_1, x_2, \ldots, x_n]$. The output of the channel is the n symbols $\mathbf{y} = [y_1, y_2, \ldots, y_n]$. To transmit n codeword symbols, we use the channel n times. The DMC is memoryless and the conditional distribution also factors as (1.21). Two important examples are the binary symmetric channel and the binary erasure channel.

Binary Symmetric Channel The binary symmetric channel (BSC) is a discrete memoryless channel where an error occurs with probability α. It has binary inputs $\mathcal{X} = \{0, 1\}$, binary outputs $\mathcal{Y} = \{0, 1\}$. For a parameter α, the probability transition matrix $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is:

$$p_{\mathsf{Y}|\mathsf{Z}}(y|z) = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}. \tag{1.14}$$

That is, the channel makes an error with probability p, where $0 \leq p \leq 1$. For the BSC, an error occurs if an input 0 is changed to a 1, and this happens with probability p. Similarly, an input 1 changes to a 0 with probability p. There is no error with probability $1 - \alpha$.

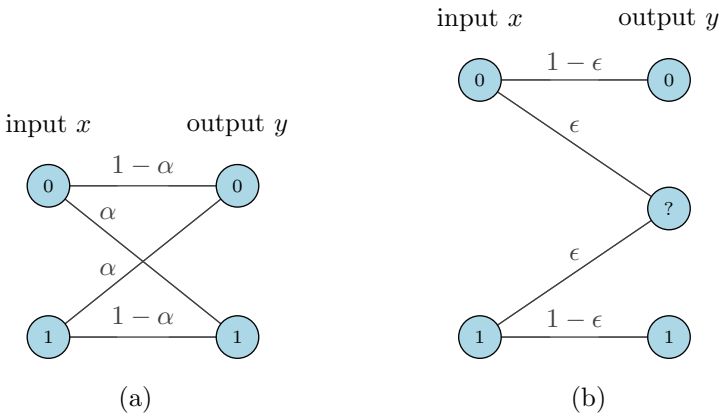


Figure 1.3: (a) Binary symmetric channel (BSC) with error probability α. (b) Binary erasure channel with erasure probability ϵ.

Example 1.2. Consider a BSC with error probability $p = 0.2$. The channel transition probabilities $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ are:

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}. \tag{1.15}$$

If the transmitter sends $x = 0$, then $y = 0$ is correctly received with probability 0.8. An error occurs if $y = 1$ is received; this occurs with probability 0.2. The probabilities on y are reversed if $x = 1$ is transmitted. The following is an example of transmitted $\mathbf{x}$ and received $\mathbf{y}$:

$\mathbf{x} =$	1	1	1	1	1	0	1	0	0	0	1	0	1	1	0	1	0	0	0	1
$\mathbf{y} =$	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	1	0	0	0	1
					E	E		E						E	E					

The errors are marked by "E."

Binary Erasure Channel The binary erasure channel (BEC) is a discrete memoryless channel also has binary inputs $\mathcal{Y} = \{0, 1\}$. The outputs are $\mathcal{Y} = \{0, ?, 1\}$ where ? is an erasure symbol. For a parameter ϵ, the probability transition matrix $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is:

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-\epsilon \end{bmatrix}, \tag{1.16}$$

where $0 \leq \epsilon \leq 1$. For the BEC, an input symbol 0 or 1 is erased with probability ϵ, and received correctly with probability $1 - \epsilon$. It is not possible for an input 0 to become an output 1 (or 1 to become a 0).

1.3.2 Additive White Gaussian Noise (AWGN) Channel

The *Gaussian distribution* (or normal distribution) $p_{\mathsf{Z}}(z)$ with mean m, variance σ^2 is:

$$p_{\mathsf{Z}}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-m)^2}{2\sigma^2}} \text{ for } z \in \mathbb{R}. \tag{1.17}$$

We often abbreviate the Gaussian distribution as $\mathsf{Z} \sim \mathcal{N}(m, \sigma^2)$.

The AWGN channel model is a common model of communications which has Gaussian noise.

Definition 1.4. The *additive white Gaussian noise* (AWGN) channel model with power constraint P and noise power σ^2 has n inputs $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and n outputs $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ where:

$$y_i = x_i + z_i, \tag{1.18}$$

the $z_i \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed Gaussian distributed with mean 0 and variance σ^2, for $i = 1, 2, \ldots, n$. Furthermore, the x_i satsify:

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq P. \tag{1.19}$$

For each $x \in \mathcal{X}$, the probability that symbol x is transmitted is $p_{\mathsf{X}}(x)$. The *average transmit power* E_s is given by:

$$E_s = \sum_{x \in \mathcal{X}} p_{\mathsf{X}}(x) x^2. \tag{1.20}$$

The AWGN channel is memoryless, which means the errors at any time i are independent of errors at any other time. That is, the joint distribution probability on the channel output given the channel input is $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. Because the channel is memoryless, this factors as:

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p_{\mathsf{Y}|\mathsf{X}}(y_i|x_i) \tag{1.21}$$

Due to the Gaussian noise distribution, $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is given by:

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-x)^2/2\sigma^2} \tag{1.22}$$

An example with $n = 1$ is the *binary-input additive white Gaussian noise* (BI-AWGN) channel or the *binary phase-shift keying* (BPSK) channel. This channel has input $\mathcal{X} = \{+1, -1\}$, and $p_{\mathsf{X}}(x) = [\frac{1}{2}, \frac{1}{2}]$. The conditional channel output distribution is:

$$p_{\mathsf{Y}|\mathsf{X}}(y|-1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y+1)^2/2\sigma^2} \tag{1.23}$$

$$p_{\mathsf{Y}|\mathsf{X}}(y|+1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-1)^2/2\sigma^2} \tag{1.24}$$

The average transmit power is:

$$E_s = \frac{1}{2}(-1)^2 + \frac{1}{2}(+1)^2 = 1 \tag{1.25}$$

1.4 Channel Capacity

Now consider using an error-correcting code to communicate over a noisy channel. The error correcting code consists of n symbols $c_1, c_2, \ldots, c_n$ (or $x_1, x_2, \ldots, x_n$), which are input to the channel. The output of the channel are n symbols $y_1, y_2, \ldots, y_n$. The decoder attempts to recover the transmitted data, and produces either an estimate of the codeword $\widehat{\mathbf{c}}$ or an estimate of the information $\widehat{\mathbf{u}}$.

The performance of the decoder is important in evaluating communication systems using error-correcting codes. Two common metrics are the word error rate (WER) and the bit error rate (BER). A word error occurs if $\mathbf{u} \neq \widehat{\mathbf{u}}$, or equivalently if $\mathbf{c} \neq \widehat{\mathbf{c}}$. Then, the word-error rate is:

$$\text{WER} = \Pr(\mathbf{u} \neq \widehat{\mathbf{u}} | \mathbf{u} \text{ was transmitted}, \mathbf{y} \text{ was received}). \tag{1.26}$$

Likewise, a bit error occurs if $u_i \neq \widehat{u}_i$, so the BER is:

$$\text{BER} = \Pr(u_i \neq \widehat{u}_i | u_i \text{ was transmitted}, \mathbf{y} \text{ was received}) \tag{1.27}$$

Finding the WER and BER analytically is impossible is most cases of interest, and Monte Carlo methods are commonly used to estimate WER and BER; see Chapter 4.

1.4.1 Channel Coding Theorem

The channel capacity is the maximum possible communications rate for a channel. Generally speaking, error-correcting codes improve as the block length n gets larger. When discussing channel capacity, reliable communications means we can make the WER go to 0 as $n \to \infty$.

Let X and Y be jointly distributed random variables. The *mutual information* between X and Y is denoted $I(\mathsf{X};\mathsf{Y})$.

Definition 1.5. Consider random variables X and Y with a joint probability distribution function $p_{\mathsf{X},\mathsf{Y}}(x,y)$ and marginal distributions $p_{\mathsf{X}}(x)$ and $p_Y(y)$. Then $I(\mathsf{X};\mathsf{Y})$ is given by:

$$I(\mathsf{X};\mathsf{Y}) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{\mathsf{X},\mathsf{Y}}(x,y) \log \frac{p_{\mathsf{X},\mathsf{Y}}(x,y)}{p_{\mathsf{X}}(x) p_{\mathsf{Y}}(y)}. \tag{1.28}$$

While a channel $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ is fixed, the input distribution $p_{\mathsf{X}}(x)$ is not. The channel capacity is the maximum value of $I(\mathsf{X};\mathsf{Y})$ over all $p_{\mathsf{X}}(x)$.

Definition 1.6. For a discrete memoryless channel $p_{\mathsf{Y}|\mathsf{X}}(y|x)$, the *channel capacity* C of a memoryless channel is:

$$C = \max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y}). \tag{1.29}$$

In addition, an optimal $p^*_{\mathsf{X}}(x)$ is called the capacity-achieving input distribution:

$$p^*_{\mathsf{X}}(x) = \arg\max_{p_{\mathsf{X}}(x)} I(\mathsf{X};\mathsf{Y}). \tag{1.30}$$

An important result in information theory states that for reliable communication over a channel, than the code rate R must be less than the capacity C.

Proposition 1.1. *Channel Coding Theorem* For every rate $R < C$, there exists a sequence of $(2^{nR}, n)$ codes with probability of decoding error going to 0 as $n \to \infty$. Conversely, any sequence of $(2^{nR}, n)$ codes with probability of decoding error going to 0 must have $R \leq C$.

That is, there exists a code with rates $R < C$ for which reliable communications is possible. On the other hand, for any code with $R > C$, reliable communication is not possible.

1.4.2 Capacity of BSC and BEC

Define the binary entropy function as $h(p) = -p\log p - (1-p)\log(1-p)$.

Proposition 1.2. The capacity of the binary symmetric channel (BSC) with error probability α is:

$$C = 1 - h(\alpha)$$

with capacity-achieving input distribution $p^*_{\mathsf{X}}(x) = [\frac{1}{2}, \frac{1}{2}]$.

Proposition 1.3. The capacity of the binary erasure channel (BEC) with erasure probability p is:

$$C = 1 - p$$

with capacity-achieving input distribution $p^*_{\mathsf{X}}(x) = [\frac{1}{2}, \frac{1}{2}]$.

1.4.3 Capacity of the AWGN Channel

For the AWGN channel, the capacity has a closed form solution. This is one of the most elegant results from information theory.

Proposition 1.4. The capacity of AWGN channel with power constraint P and noise variance σ^2 is:

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \text{ bits per transmission} \tag{1.31}$$

The capacity-achieving input distribution is $p^*_{\mathsf{X}}(x)$ is a zero-mean Gaussian with variance P.

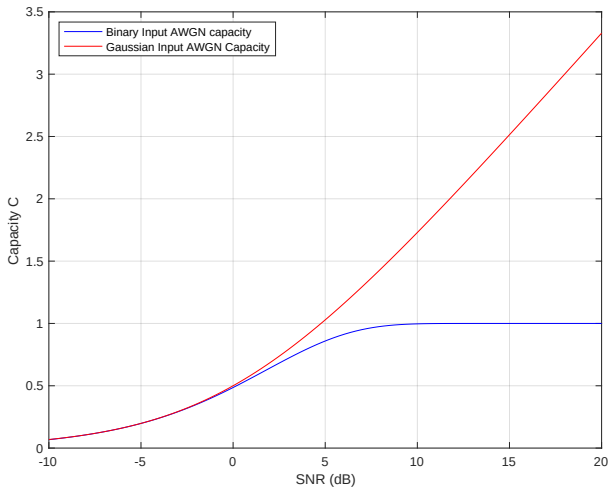


Figure 1.4: Capacity of the AWGN channel, with Gaussian inputs and binary inputs (BI-AWGN).

1.4.4 Capacity of the BI-AWGN Channel

Capacity can be achieved with a Gaussian input distribution. It is convenient to consider the special case when the input alphabet is limited to $\mathcal{X} = \{-1, +1\}$. This is the binary-input AWGN (BI-AWGN) channel. This is particularly convenient using a binary code, using the mapping from the binary code $\{0, 1\}$ to the channel input $\{-1, +1\}$. The capacity of this channel is given by:

$$C = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(y-1)^2/2\sigma^2} \log_2 \frac{2}{1 + e^{-2y/\sigma^2}} dy \tag{1.32}$$

and is achieved by choosing $p_{\mathsf{X}}(0) = p_{\mathsf{X}}(1) = \frac{1}{2}$. Unfortunately, this important channel does not have a closed-form solution for its capacity. To achieve the capacity, we need to use a code with $n \to \infty$. On the other hand, the case of $n = 1$ was considered in Subsection 1.3.2.

The capacity of the AWGN channel and the BI-AWGN channel is plotted in 1.4. For channels with low SNR, the capacity is the same. However, for channels with high SNR, the capacity of the BI-AWGN is limited to 1, since the inputs are binary. On the other hand, the capacity of the AWGN channel is unbounded.

1.4.5 The Challenge of Coding Theory

Now we come to challenge of coding theory — how do we design codes for reliable communications?

While the channel coding theorem perspective is "given a channel, what is the highest possible rate?". However, it is equally valid to ask "give a rate, what is worst channel we can use". The case of $R = \frac{1}{2}$ is considered in Fig. 1.5

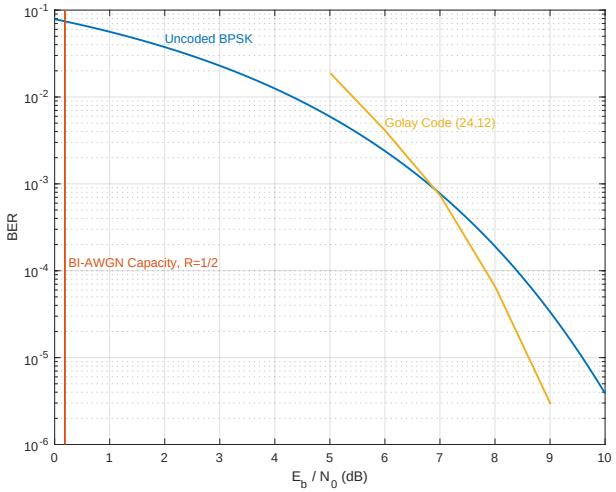


Figure 1.5: Bit-error rate for the uncoded BPSK, the extended Golay code (24,12), and the capacity of the AWGN channel

For the AWGN channel, the signal-to-noise ratio (SNR) is a measure of how noisy the channel is. The definition of the signal-to-noise ratio is quite important, and here E_b/N_0 will be used. The power per information symbol is $E_b = E_s/R$. The noise power is $N_0 = 2\sigma^2$ where σ^2 the the variance of the Gaussian noise. Thus the SNR definition E_b/N_0 is:

$$\frac{E_b}{N_0} = \frac{E_s}{2\sigma^2 R} \tag{1.33}$$

where E_s is given in (1.20). Also, E_b/N_0 is often expressed in dB, that is:

$$E_b/N_0(dB) = 10\log_{10} E_b/N_0. \tag{1.34}$$

Consider first uncoded communication on the BI-AWGN channel, we can find the probability of error. The decoder receives y, and uses the following *hard decision* rule to estimate the transmitted information:

$$\widehat{x} = \begin{cases} -1 & \text{if } y < 0 \\ +1 & \text{if } y \geq 0 \end{cases}. \tag{1.35}$$

An error occurs if $\widehat{x} \neq x$. The exact probability of error has an analytical expression:

$$\Pr(x \neq \widehat{x}) = \frac{1}{2}\Pr(y > 0|x = -1) + \frac{1}{2}\Pr(y < 0|x = 1) \tag{1.36}$$

$$= \Pr(y > 0|x = -1) \tag{1.37}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+1)^2}{2\sigma^2}} dy \tag{1.38}$$

$$= \frac{1}{2}\text{erfc}(\frac{1}{\sqrt{2\sigma^2}}), \tag{1.39}$$

where erfc is the complementary error function.

As can be seen in the figure, there is a large gap between capacity and the uncoded scheme, for any BER. Additionally has been shown the $n = 24$ Golay code, which is an excellent code for this block length. However, in order to get closer to capacity, we need longer and more powerful codes. This is the challenge of coding theory.

1.5 Exercises

1.1 Describe how to cut 88 circles of 1-inch diameter out of a sheet of paper of width 8.5 inches and length 11 inches. Prove that it is not possible to cut on more than 119 circles of 1-inch diameter.

1.2 In $\mathbb{R}^n$, a hypercube of side length ℓ has volume ℓ^n. Give an upper bound on the number of n-balls of radius $r = 1$ that fit inside this hypercube.

1.3 Prove that it is not possible to find 32 binary codewords, each of length 8 bits, such that every word differs from every other word in at least three places.

1.4 For any block code with length n over alphabet size q with minimum distance $2t + 1$ or greater, the number of codewords M satisfies:

$$M \leq \frac{q^n}{\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t\right)}. \tag{1.47}$$

Prove this statement.

1.5 Consider the binary-input AWGN system, with input $x \in \{-1, +1\}$ and Gaussian noise with variance σ^2. The decoder uses the hard decision rule (1.11) to make the estimate $\widehat{x}$.

(a) Using your favorite programming language, write a simulation to estimate the probability of error for this system (an error occurs if $\widehat{c} \neq c$). Transmit the symbol N times. Count the number of times $\widehat{c} \neq c$, call this N_{error}. Using N sufficiently large, use your simulation to estimate the probability of error:

$$\frac{N_{error}}{N} \tag{1.51}$$

for $\sigma^2 = 1, 0.5, 0.15$.

(b) The exact probability of error has an analytical expression given in (1.26). Compare the analytic value (some languages have a built-in erfc function) with the simulation in part (a).

1.6 Solve the following algebraic problems over $\mathbb{F}_2$ $\mathbb{F}_3$ or $\mathbb{F}_4$, as indicated. Use the tables in Section 1.2.

(a) Let $f(x) = x^4 + x^3 + x + 1$ be a polynomial over $\mathbb{F}_2$. Evaluate $f(0)$, $f(1)$

(b) Let $f(x) = x^2 - 1$ be a polynomial over $\mathbb{F}_3$. Evaluate: $f(0)$, $f(1)$, $f(2)$

(c) Let $f(x) = x^3 - 1$ be a polynomial over $\mathbb{F}_4$. Evaluate: $f(0)$, $f(1)$, $f(2)$, $f(3)$,

(d) Solve the following system over $\mathbb{F}_2$ for a, b, c

$$\begin{aligned} a + b + c &= 1 \\ b - c &= 0 \\ a + b &= 1 \end{aligned}$$

(e) Solve the following system over $\mathbb{F}_3$ for a, b, c

$$\begin{aligned} a + 2b + c &= 2 \\ b - 2c &= 1 \\ 2a + b &= 1 \end{aligned}$$

(f) Compute the determinant and the inverse of the following matrix over $\mathbb{F}_4$:

$$\begin{bmatrix} 3 & 3 \\ 2 & 1 \end{bmatrix}$$

1.7 The *extended* Hamming code has a graphical representation created by adding a new parity bit p_4 outside of the three circles of the original Hamming code representation of Fig. 1.2. Add a fourth circle around all 8 bits. The new codeword is:

$$[u_1, u_2, u_3, u_4, p_1, p_2, p_3, p_4]$$

As before, the number of 1's inside each circle, including the new fourth circle, must be even.

(a) Draw a graphical representation of the extended Hamming code.

(b) Encode $[u_1, u_2, u_3, u_4] = [0, 1, 1, 1]$ to its codeword.

(c) Decode $\mathbf{y} = [0, 0, 1, 1, 0, 0, 1, 0]$ and $\mathbf{y} = [0, 1, 1, 0, 0, 1, 1, 1]$ to the Hamming codeword. The extended Hamming code can correct one error.

(d) A codeword was transmitted over the erasure channel and $\mathbf{y} = [?, 1, 0, ?, 0, 1, ?, 0]$ was received. Decode $\mathbf{y}$ to $\widehat{\mathbf{c}}$.

.

Chapter 2

Abstract Algebra

This chapter covers the algebraic structures of groups, rings and fields.

A group $\mathcal{G}$ is a set that has addition and subtraction. A ring $\mathcal{R}$ is a set that has addition, subtraction and multiplication. A field $\mathbb{F}$ is a set that has addition, subtraction, multiplication and division.

If you have a group $\mathcal{G}$ and a subgroup $\mathcal{H}$, then you can form a *quotient group* $\mathcal{G}/\mathcal{H}$. An example is the quotient group $\mathbb{Z}/q\mathbb{Z}$, where $\mathbb{Z}$ is a group and $q\mathbb{Z}$ is a subgroup. For any q, $\mathbb{Z}/q\mathbb{Z}$ is a ring, but if q is prime, then $\mathbb{Z}/q\mathbb{Z}$ is also a field. This idea is used to generate a finite field where the number of elements is a prime number.

If you have a ring $\mathcal{R}$ and a special subset called an ideal $\mathcal{I}$, then you can form a *quotient ring* $\mathcal{R}/\mathcal{I}$. An example is the quotient ring $\mathbb{F}[x]/\langle b(x)\rangle$ where $\mathbb{F}[x]$ is a polynomial ring, and $\langle b(x)\rangle$ is an ideal generated by polynomial $b(x)$. For any polynomial $b(x)$, $\mathbb{F}[x]/\langle b(x)\rangle$ is a ring, but if $b(x)$ is an *irreducible polynomial*, then $\mathbb{F}[x]/\langle b(x)\rangle$ is also a field. This idea is used to generate a finite field where the number of elements is the power of a prime number.

2.1 Overview: Groups, Rings and Fields

Group, rings and fields are algebraic structures that are central to understanding error-correcting codes. They are studied extensively under the heading of "abstract algebra" "group theory" or "field theory". This section has an informal treatment, with more details in Chapter 2. Wikipedia: Abstract algebra

Roughly speaking, a *group* is set where addition and subtraction are defined. A *ring* is a set where addition, subtraction and multiplication are defined. A *field* is a set where addition, subtraction, multiplication and division are defined. Any field is a ring. Any ring is a group. "Addition" and "multiplication" may not be familiar the operations on the numbers, but must be defined with respect to the group, ring or field.

The integer multiples of five, $5\mathbb{Z}$ is an example of a group with respect to

usual addition:

$$5\mathbb{Z} = \{\ldots, -10, -5, 0, 5, 10, \ldots, \}. \tag{2.1}$$

The sum of any two numbers which are a multiple of five, is itself a multiple of five: $10 + 15 = 25$ is a multiple of five. Subtraction is also defined, for example $15 - 25 = -10$. The additive identity is 0, that is $5 + 0 = 5$. (But the multiplicative identity 1 is not in this set, so it is not a ring — see Chatper 2.)

An example of a ring $\mathcal{R}$ is the integers:

$$\mathcal{R} = \{0, 1, 2, 3, 4, 5\} \tag{2.2}$$

with addition and multiplication taken modulo 6. For example, $3+5 \bmod 6 = 2$ is addition in the ring and $2 \cdot 5 \bmod 6 = 4$ is multiplication within the ring. This ring is not a field, because division is not well defined, for example 2^{-1} does not exist, because there is no element a such that $2 \cdot a = 1$. Another example of a ring are the integers $\mathbb{Z}$ — the sum of any two integers is an integer and the product of any two integers is an integer.

The set of real numbers $\mathbb{R}$ is an example of a field, since addition, subtraction, multiplication and division are well defined. However, we are particularly interested in finite fields[1] $\mathbb{F}_q$ of size q, which are fields defined on a finite set of size q. The binary field $\mathbb{F}_2$ has addition and multiplication defined on $\{0, 1\}$:

$+$	0	1
0	0	1
1	1	0

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

Binary arithmetic is the same as conventional arithmetic, except $1 + 1 = 0$. Note that $-1 = 1$ and $-0 = 0$, so for any a, $a = -a$.

The ternary field $\mathbb{F}_3$ on $\{0, 1, 2\}$ has addition and multiplication defined with respect to usual operations modulo 3:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

The additive inverse of a is in the column where row a has a 0. For example,$-2 = 1$, that is $2 + (-2) = 2 + 1 = 0$. Similarly, the multiplicative inverse of a is the column where row a has a 1. For example, $2^{-1} = 2$ since $2 \cdot 2 = 1$ means $2 = 2^{-1}$. If q is a prime number, a field of size q is formed using addition and multiplication modulo q.

Fields exist for some sizes which are not a prime number. Consider the field $\mathbb{F}_4$ of size 4 on the set $\{0, 1, \alpha, \alpha^2\}$. Here α and α^2 are two distinct elements of the field. The addition and multiplication operations are given by:

[1] The finite field $\mathbb{F}_q$ is also called a Galois field, denoted GF(q)

$+$	0	1	α	α^2
0	0	1	α	α^2
1	1	0	α^2	α
α	α	α^2	0	1
α^2	α^2	α	1	0

$\cdot$	0	1	α	α^2
0	0	0	0	0
1	0	1	α	α^2
α	0	α	α^2	1
α^2	0	α^2	1	α

As usual, 0 is the additive identity: $a + 0 = a$, and 1 is the multiplicative identity: $1 \cdot a = a$. Note that addition and multiplication modulo 4 does not give a finite field. Subtraction and division can be found from the tables above. For example, $2 + (-2) = 0$ is an example of subtraction. The multiplicative inverse exists because there is a 1 in each non-zero row and column of the multiplication table, for example $2^{-1} = 3$ because $2 \cdot 3 = 1$.

Fields exist only for sizes that are prime numbers, or a power of a prime number. Since $4 = 2^2$ is a power of a prime number, a field of size 4 exists, as was shown above. But there exist no fields of size 6 or 10, for example, since these numbers cannot be written as the power of a prime number. This is discussed in Chapter 2.

Familiar algebra can be performed using fields. For example, let $f(x) = x^2 - x + 1$ be a polynomial where x and coefficients are from $\mathbb{F}_3$. The polynomial can be evaluated in the field: $f(0) = 0 - 0 + 1 = 1$, $f(1) = 1 - 1 + 1 = 1$ and $f(2) = 2^2 - 2 + 1 = 0$. As an example in $\mathbb{F}_4$, let $f(x) = x^2 + \alpha^2 x + \alpha$. Using the $\mathbb{F}_4$ addition and multiplication tables: $f(0) = \alpha$, $f(1) = 1^2 + \alpha^2 + \alpha = 0$, $f(\alpha) = \alpha^2 + \alpha^2 \cdot \alpha + \alpha = \alpha^2 + 1 + \alpha = 0$ and $f(\alpha^2) = (\alpha^2)^2 + (\alpha^2) \cdot (\alpha^2) + \alpha = \alpha + \alpha + \alpha = \alpha$.

Linear systems of equations can be formed using finite fields. For example, let x, y, z be elements in $\mathbb{F}_2$. The following set of equations can be solved to find the unknowns:

$$x + y \quad = 1 \tag{2.3}$$
$$x \quad + z = 1 \tag{2.4}$$
$$x + y + z = 0 \tag{2.5}$$

It can be solved by subtracting the first equation from the third equation to find $z = 1$, since $-1 = 1$. Using $z = 1$ in the second equation, $x = 0$. Finally using the first equation, $y = 1$.

SSQ 2.1. *Finite field algebra.* Using the table of addition and subtraction for $\mathbb{F}_4$, find:

- -1
- α^{-1}
- $f(\alpha^2)$ where $f(x) = x^2 + 1$.

2.2 Groups and Cosets

2.2.1 Groups

Let $\mathcal{G}$ be a set, and define an operation " $+$ ". Let $a, b, c \in \mathcal{G}$. $\mathcal{G}$ with $+$ forms an *Abelian group* or *commutative group* if the following hold:

- closure under $+$, that is $a + b \in \mathcal{G}$ for all a, b, and
- associativity: $a + (b + c) = (a + b) + c$ for all a, b, c and
- an identity element $e \in \mathcal{G}$, $a + e = e + a = a$ for all a, and
- an inverse element: for each a, there exists an element b, denoted as $-a$, such that $a + b = a + (-a) = e$, where e is the identity element.
- commutativity: $a + b = b + a$ for all a, b.

A group $(\mathcal{G}, +)$ is defined by both a set $\mathcal{G}$ and an operation $+$, but we usually use only $\mathcal{G}$ to denote the group. Non-Abelian groups do not require the commutativity property, but our main interest is in Abelian groups[2]. Since most of the groups of interest are Abelian groups, and we will just say "group" since the distinction is not important.

Example 2.1. The integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}$ form a group under usual addition. It is easy to verify that they satisfy the four properties above.

Example 2.2. The set of integers that are multiples of 5, denoted $5\mathbb{Z}$,

$$5\mathbb{Z} = \{\ldots, -10, -5, 0, 5, 10, 15, \ldots\} \tag{2.6}$$

forms a group under usual addition.

2.2.2 Subgroups and Cosets

Definition 2.1. Let $\mathcal{G}$ be a group and let $\mathcal{H}$ be a subset of $\mathcal{G}$, that is $\mathcal{H} \subseteq \mathcal{G}$. If $\mathcal{H}$ is a group under the same operation as $\mathcal{G}$, then $\mathcal{H}$ is a *subgroup* of $\mathcal{G}$.

Example 2.3. The integers $\mathbb{Z}$ are a group, and $4\mathbb{Z}$ is a subset of $\mathbb{Z}$. Since $4\mathbb{Z}$ forms a group, then $4\mathbb{Z}$ is a subgroup of $\mathbb{Z}$, under usual addition.

Definition 2.2. Let $\mathcal{G}$ be a group and let $\mathcal{H}$ be a subgroup of G. For any $a \in \mathcal{G}$, the set $a + \mathcal{H} = \{a + h \mid h \in \mathcal{H}\}$ is called the *coset of* $\mathcal{H}$ *in* $\mathcal{G}$ *containing* a.

[2] In a non-Abelian group, changing the order of operations may not yield the same value, that is $a + b$ does not have to equal $b + a$.

Example 2.4. The integers $\mathbb{Z}$ form a group under addition, and the integer muliples of 4, $4\mathbb{Z} = \{0, \pm 4, \pm 8, \ldots\}$ form a subgroup, $4\mathbb{Z} \subset \mathbb{Z}$. Here, $\mathbb{Z}$ is a group, and $4\mathbb{Z}$ is a subgroup and there are four cosets of $4\mathbb{Z}$ in $\mathbb{Z}$,

$$0 + 4\mathbb{Z} = \{\ldots, -8, -4, 0, 4, 8, \cdots\} \tag{2.7}$$
$$1 + 4\mathbb{Z} = \{\ldots, -7, -3, 1, 5, 9, \cdots\} \tag{2.8}$$
$$2 + 4\mathbb{Z} = \{\ldots, -6, -2, 2, 6, 10, \cdots\} \tag{2.9}$$
$$3 + 4\mathbb{Z} = \{\ldots, -5, -1, 3, 7, 11, \cdots\} \tag{2.10}$$

The first coset is the coset containing $a = 0$, it is also the coset containing $a = -4$, etc. The second coset $1 + 4\mathbb{Z}$ is the coset containing 1, it is also the coset containing $a = 5$, etc.

2.2.3 Quotient Groups

Let $\mathcal{G}/\mathcal{H}$ be the set of all cosets of $\mathcal{H}$ in $\mathcal{G}$, that is:

$$\mathcal{G}/\mathcal{H} = \{a + \mathcal{H} | a \in \mathcal{G}\} \tag{2.11}$$

Note that $\mathcal{G}/\mathcal{H}$ is a set of sets. The set $\mathcal{G}/\mathcal{H}$ is called a *quotient group*, and $\mathcal{G}/\mathcal{H}$ is itself a group.

Choosing $\mathcal{G} = \mathbb{Z}$ and subgroup $\mathcal{H} = q\mathbb{Z}$ for $q \geq 2$, gives the important quotient group is $\mathbb{Z}/q\mathbb{Z}$. This group is often abbreviated $\mathbb{Z}_q$. The following example shows how to literally apply the definition of quotient group to $\mathbb{Z}/q\mathbb{Z}$.

Example 2.5. Continuing the previous example:

$$\mathbb{Z}/4\mathbb{Z} = \{0 + 4\mathbb{Z}, 1 + 4\mathbb{Z}, 2 + 4\mathbb{Z}, 3 + 4\mathbb{Z}\} \tag{2.12}$$

While not intuitive, this set does indeed form a group, with addition expressed in this table:

$+$	$0+4\mathbb{Z}$	$1+4\mathbb{Z}$	$2+4\mathbb{Z}$	$3+4\mathbb{Z}$
$0+4\mathbb{Z}$	$0+4\mathbb{Z}$	$1+4\mathbb{Z}$	$2+4\mathbb{Z}$	$3+4\mathbb{Z}$
$1+4\mathbb{Z}$	$1+4\mathbb{Z}$	$2+4\mathbb{Z}$	$3+4\mathbb{Z}$	$0+4\mathbb{Z}$
$2+4\mathbb{Z}$	$2+4\mathbb{Z}$	$3+4\mathbb{Z}$	$0+4\mathbb{Z}$	$1+4\mathbb{Z}$
$3+4\mathbb{Z}$	$3+4\mathbb{Z}$	$0+4\mathbb{Z}$	$1+4\mathbb{Z}$	$2+4\mathbb{Z}$

A *coset leader* (or *coset representative*) is a single representative element from each coset. There are generally multiple choices of coset leaders. For $\mathbb{Z}/q\mathbb{Z}$, the set of coset representatives is usually:

$$\{0, 1, 2, \ldots, q-1\}, \tag{2.13}$$

and the group operation is addition modulo q. For example, for $\mathbb{Z}/4\mathbb{Z}$, possible coset leaders are $\{0, 1, 2, 3\}$, and the group operation is addition modulo 4:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Since $\mathcal{G}/\mathcal{H}$ forms a group, and we can perform addition using any coset leader.

However, other sets of coset leaders are possible. For example, for $\mathbb{Z}/4\mathbb{Z}$, $\{-2, -1, 0, 1\}$ is also a set of coset leaders. Note that there is indeed one leader for each set in Example 2.4. In this case, the addition table is:

+	0	1	−2	−1
0	0	1	−2	−1
1	1	−2	−1	0
−2	−2	−1	0	1
−1	−1	0	1	−2

These two groups are isomorphic to each other, and so the more convenient $\{0, 1, 2, \ldots, q-1\}$ is usually used.

2.3 Vector Spaces

.

This section describes vector spaces for finite fields; it is the natural generalization of the more familiar vector spaces for the real numbers.

Vector spaces over the field of real numbers $\mathbb{R}$ are quite familiar. A vector field can be defined over a finite field $\mathbb{F}_q$ as well. In the following description of vector fields, let $\mathbb{F}$ represent either the real numbers $\mathbb{R}$ or a finite field $\mathbb{F}_q$.

A *vector space* $\mathbb{F}_q^n$ in n-dimensions is defined using a field $\mathbb{F}_q$. Form vectors or n-tuples $\mathbf{x}$ and $\mathbf{y}$ of n elements each, that is:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \tag{2.14}$$
$$\mathbf{y} = (y_1, y_2, \ldots, y_n), \tag{2.15}$$

where $x_i, y_i \in \mathbb{F}_q$. Then addition in the vector space is:

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{2.16}$$

Multiplication by a scalar $a \in \mathbb{F}_q$ is given by:

$$a \cdot \mathbf{x} = (a \cdot x_1, a \cdot x_2, \ldots, a \cdot x_n) \tag{2.17}$$

Wikipedia: Vector space

We say k vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are *linearly dependent* if a linear combination of the vectors sums to 0, that is:

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_k\mathbf{x}_k = 0. \tag{2.18}$$

for scalars $a_i \in \mathbb{F}$ not all zero. If this does not hold, then vectors are said to be *linearly independent*. Wikipedia: Linear independence

In other words, linear dependence means at least one non-trivial linear combination of $\mathbf{x}_i$ sum to 0. If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are not dependent, then they are independent.

The *row rank* of a matrix $\mathbf{G}$ is the maximum number of linearly independent columns of $\mathbf{G}$. This is equal to the dimension of the space spanned by its rows. The *column rank* is similarly defined for columns. It can be shown that the row rank and column rank are equal, and are thus the *rank* of the matrix. Wikipedia: Rank

2.4 Exercises

2.1 Consider the group $\mathcal{G} = \{0, 1, 2, \ldots, 11\}$ and its subgroup $\mathcal{H} = \{0, 4, 8\}$ under modulo-12 addition.

(a) How many elements are in the quotient group $\mathcal{G}/\mathcal{H}$? List the elements.

(b) Give two distinct example of coset leaders.

2.2 Consider the group $\mathcal{G} = \mathbb{Z}^2$ and its subgroup $\mathcal{H} = (3\mathbb{Z})^2$, under usual addition.

(a) These groups can be represented in 2 dimensions. Draw a plot of some elements of $\mathcal{G}$ and $\mathcal{H}$.

(b) How many coset leaders in $\mathcal{G}/\mathcal{H}$? Give an example of the set of coset leaders.

2.3 The group $\mathbb{Z}^2$ and its subgroup $(4\mathbb{Z})^2$ forms the quotient group $\mathbb{Z}^2/(4\mathbb{Z})^2$. This has 16 coset leaders, labeled $\mathbf{c}_0$ to $\mathbf{c}_{15}$ as shown in this table:

$\mathbf{c}_0 = [0,0]$	a	$\mathbf{c}_8 = [2,0]$	i
$\mathbf{c}_1 = [0,1]$	b	$\mathbf{c}_9 = [2,1]$	j
$\mathbf{c}_2 = [0,2]$	c	$\mathbf{c}_{10} = [2,2]$	k
$\mathbf{c}_3 = [0,3]$	d	$\mathbf{c}_{11} = [2,3]$	l
$\mathbf{c}_4 = [1,0]$	e	$\mathbf{c}_{12} = [3,0]$	m
$\mathbf{c}_5 = [1,1]$	f	$\mathbf{c}_{13} = [3,1]$	n
$\mathbf{c}_6 = [1,2]$	g	$\mathbf{c}_{14} = [3,2]$	o
$\mathbf{c}_7 = [1,3]$	h	$\mathbf{c}_{15} = [3,3]$	p

(a) Compute the following sums in $\mathbb{Z}^2/(4\mathbb{Z})^2$:

$$\mathbf{c}_3 + \mathbf{c}_2$$
$$\mathbf{c}_2 + \mathbf{c}_6$$
$$\mathbf{c}_{11} + \mathbf{c}_9$$
$$\mathbf{c}_{14} + \mathbf{c}_4$$
$$\mathbf{c}_6 + \mathbf{c}_1$$

(b) For each sum in part (a), find the corresponding letter from the table. The five letters spell a common English word.

2.4 Let $\mathcal{G}$ be a finite group with k elements. Let $\mathcal{H}$ be a subgroup of $\mathcal{G}$ with m elements.

(a) Prove that that $\mathcal{H} + a \to \mathcal{H}$ forms a bijection, $a \in \mathcal{G}$.

(b) Prove that $m|k$, that is, m divides k.

2.5 Let $\mathcal{G}$ be the n roots of $x^n = 1$ in the complex numbers, for $n \geq 1$.

(a) Prove that $\mathcal{G}$ is a group under multiplication.

(b) Prove that $\mathcal{G}$ is not a group under addition.

2.6 It is well known that the polynomial $x^2 + 1$ over the real numbers has no roots in the real numbers, but does have the two roots $\pm\sqrt{-1}$ in the complex numbers. In a similar way, the polynomial $x^2 + x + 1$ over $\mathbb{F}_2$ has no roots in $\mathbb{F}_2$. However, it does have two roots over $\mathbb{F}_4$. Referring to $\mathbb{F}_4$ in the lecture notes, find the two roots of $x^2 + x + 1$ in $\mathbb{F}_4$.

Chapter 3

Linear Block Codes

Almost all error-correction codes used in practice are linear codes. This chapter introduces linear block codes.

3.1 Block Codes

3.1.1 Block Codes and Hamming Distance

This section describes finite-field codes, which are error-correcting codes defined using n symbols from a finite field. These are commonly called block codes. A block code is a set of vectors in $\mathbb{F}_q^n$.

Definition 3.1. A *finite-field code* or *block code* $\mathcal{C}$ with block length n is the set of M codewords

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\} \tag{3.1}$$

and each codeword c has n symbols:

$$\mathbf{c} = (c_1, c_2, \ldots, c_n). \tag{3.2}$$

where each $c_i \in \mathbb{F}_q$.

We say n is the block length of the code. We say the code is defined over the finite field $\mathbb{F}_q$.

Definition 3.2. The rate R of a block length n code with M codewords is:

$$R = \frac{1}{n} \log M \tag{3.3}$$

Example 3.1. The following code has $n = 4, q = 5$ and $M = 3$:

$$\begin{aligned} \mathcal{C} = \{&(0,0,0,0),\\ &(3,0,4,1),\\ &(2,2,2,3)\} \end{aligned}$$

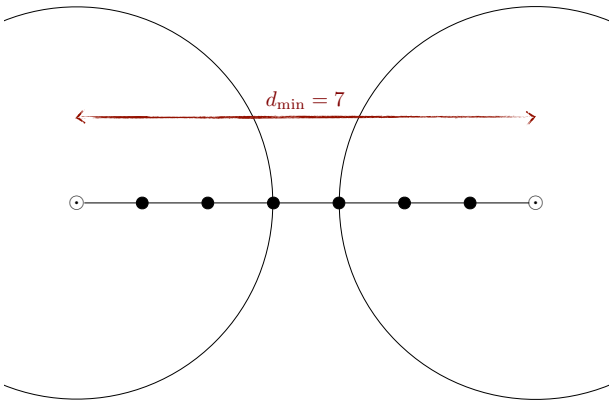


Figure 3.1: Minimum distance between two codewords (denoted $\odot$) is 7. The correctable number of errors is 3.

Definition 3.3. The *Hamming weight* $\mathrm{wt_H}(\mathbf{x})$ of a vector $\mathbf{x}$ is number of non-zero elements in $\mathbf{x}$. Mathematically that is written as:

$$\mathrm{wt_H}(\mathbf{x}) = |\{i|x_i \neq 0\}| \tag{3.4}$$

Definition 3.4. The *Hamming distance* $d(\mathbf{x}, \mathbf{y})$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is the number of positions where $\mathbf{x}$ and $\mathbf{y}$ differ:

$$d(\mathbf{x}, \mathbf{y}) = |\{i|x_i \neq y_i\}| \tag{3.5}$$

Example 3.2. The Hamming weight of the binary vector [1 0 1 1 1 1 0] is 5. The Hamming weight of the non-binary vector [2 0 3 0 0 4] is 3.

Example 3.3. The Hamming distance between the binary vectors:

$$[1\ 0\ 1\ 0\ 0] \text{ and} \tag{3.6}$$
$$[1\ 1\ 0\ 0\ 1] \tag{3.7}$$

is 3. The Hamming distance between the non-binary vectors:

$$[0\ 1\ 5\ 4\ 0] \text{ and} \tag{3.8}$$
$$[0\ 2\ 5\ 1\ 0] \tag{3.9}$$

is 2.

SSQ 3.1. *Hamming Weight and Hamming Distance* What is the Hamming weight of $(1, 2, 3, 0, 0)$? What is the Hamming distance between $(0, 0, 0, 1, 1, 0)$ and $(1, 0, 1, 1, 0, 0)$?

Some properties of Hamming weight and Hamming distance:

$$d(\mathbf{x}, \mathbf{0}) = \mathrm{wt}_{\mathrm{H}}(\mathbf{x}) \tag{3.10}$$

$$d(\mathbf{x}, \mathbf{y}) = \mathrm{wt}_{\mathrm{H}}(\mathbf{x} - \mathbf{y}) = \mathrm{wt}_{\mathrm{H}}(\mathbf{y} - \mathbf{x}) \tag{3.11}$$

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{w}) + d(\mathbf{w}, \mathbf{y}) \qquad \text{the triangle inequality} \tag{3.12}$$

The concept of a Hamming ball and its volume is introduced. "Volume" means number of sequences. In the vector space $\mathbb{F}_q^n$ there are q^n sequences, so we say the volume of $\mathbb{F}_q^n$ is q^n.

Definition 3.5. A *Hamming ball* of radius t around a point $\mathbf{c}$ is the set of sequences with Hamming distance t or less from $\mathbf{c}$:

$$\{\mathbf{y} | d(\mathbf{c}, \mathbf{y}) \leq t\}. \tag{3.13}$$

The number of sequences in the Hamming ball of radius t is its volume $V_n(t)$:

$$V_n(t) = \sum_{i=0}^{t} (q-1)^i \binom{n}{i} \tag{3.14}$$

where $\binom{n}{i} = \frac{n!}{i!(n-i)!}$.

Example 3.4. With $q = 2$ and $n = 4$, the Hamming ball of radius $t = 1$ around the point $(0, 0, 0, 0)$ is:

$$\{(0,0,0,0), (0,0,0,1), (0,0,1,0), (0,1,0,0), (1,0,0,0)\}. \tag{3.15}$$

The volume here is $V_4(1) = 1 + \binom{4}{1} = 5$, because there are 5 sequences. If the radius is increased to $t = 2$, then the Hamming sphere additionally includes:

$$\{(0,0,1,1), (0,1,0,1), (1,0,0,1), (0,1,1,0), (1,0,1,0), (1,1,0,0)\}. \tag{3.16}$$

and the volume is $V_4(2) = 1 + \binom{4}{1} + \binom{4}{2} = 11$.

3.1.2 Minimum Distance

The minimum distance is an important property of a code $\mathcal{C}$. If two codewords $\mathbf{x}$ and $\mathbf{y}$ are close in minimum distance, if $\mathbf{x}$ is transmitted, then noise might cause $\mathbf{y}$ to be decoded erroneously.

Definition 3.6. The *minimum distance* $d_{\min}$ of a block code is the minimum of the Hamming distances between all distinct pairs of codewords $\mathbf{x}$ and $\mathbf{y} \in \mathcal{C}$:

$$d_{\min} = \min_{\mathbf{x}, \mathbf{y} \in \mathcal{C}, \mathbf{x} \neq \mathbf{y}} d(\mathbf{x}, \mathbf{y}) \tag{3.17}$$

Example 3.5. Consider a code with $M = 5$ codewords:

$$\mathbf{c}_1 = (1, 0, 1, 1, 1, 0)$$
$$\mathbf{c}_2 = (1, 0, 0, 1, 0, 1)$$
$$\mathbf{c}_3 = (1, 1, 1, 0, 0, 1)$$
$$\mathbf{c}_4 = (0, 1, 0, 1, 1, 1)$$
$$\mathbf{c}_5 = (0, 0, 0, 0, 0, 0)$$

The distance between all pairs is given by:

$d(\cdot, \cdot)$	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$
$\mathbf{c}_1$	0	3	4	4	4
$\mathbf{c}_2$	3	0	3	3	3
$\mathbf{c}_3$	4	3	0	4	4
$\mathbf{c}_4$	4	3	4	0	4
$\mathbf{c}_5$	4	3	4	4	0

The minimum distance between distinct codewords is 3, so for this code, $d_{\min} = 3$.

Recall the DMC model of a communications channel. The errors due to the channel is modeled using an error vector $\mathbf{e}$:

$$\mathbf{e} = (e_1, e_2, \ldots, e_n) \tag{3.18}$$

Then the received sequence $\mathbf{y}$ is:

$$\mathbf{y} = \mathbf{c} + \mathbf{e}, \tag{3.19}$$

where addition is in the field of the code. The weight of the error vector $\mathrm{wt_H}(\mathbf{e})$ is equal to the number of errors. The error vector from Example 1.1.2 is $\mathbf{e} =$ [1 0 0 1 0 1 0], a three-errors vector; the weight of $\mathbf{e}$ is $\mathrm{wt_H}(\mathbf{e}) = 3$.

Principle of Hamming distance decoding Given a received sequence $\mathbf{y}$, a decoder should choose as its output a codeword $\widehat{\mathbf{c}}$ which is closest in Hamming distance to $\mathbf{y}$:

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c} \in \mathcal{C}} d(\mathbf{y}, \mathbf{c}) \tag{3.20}$$

Example 3.6. Continuing Example 3.5, assume that

$$\mathbf{y} = [0, 0, 0, 1, 1, 1] \tag{3.21}$$

is received. The Hamming distance to each codeword is:

	$d(\mathbf{c}_i, \mathbf{y})$
$\mathbf{c}_1$	3
$\mathbf{c}_2$	2
$\mathbf{c}_3$	5
$\mathbf{c}_4$	1
$\mathbf{c}_5$	3

The closest codeword in the Hamming distance sense is $\mathbf{c}_4$, so the decoder outputs $\widehat{\mathbf{c}} = \mathbf{c}_4$. Note that ties are possible — to resolve the ties, the the decoder may randomly choose one of the minimum distance codewords randomly.

If the number of errors is not too large, then it may be possible to correct the errors, and correctly decode the transmitted codeword. The Hamming distance is important for determining the error-correction capability of the code. *A t-error correcting code* can correctly determine the transmitted codeword if the weight of the error vector is t or smaller. This holds independently of which codeword $\mathbf{c} \in \mathcal{C}$ was transmitted.

Proposition 3.1. A code $\mathcal{C}$ with minimum distance $d_{\min}$ can correct $t \leq \lfloor \frac{d_{\min}-1}{2} \rfloor$ errors[1].

Before giving the proof, Fig. 3.1 illustrates the concepts. If each codeword is surrounded by a sphere of radius t, then any error sequence of weight t or less, will cause the received sequence to be decoded correctly. For successful decoding, the spheres cannot overlap. If $d_{\min}$ is odd, then the radius of this sphere is $t = \frac{1}{2}(d_{\min} - 1)$. If $d_{\min}$ is even, then there is a midway point that cannot be successfully decoded, and $t = \frac{1}{2}(d_{\min} - 2)$.

Proof Let $\mathbf{y} = \mathbf{c} + \mathbf{e}$ and assume $\mathrm{wt}_{\mathrm{H}}(\mathbf{e}) = s \leq t$, that is s errors have occurred. Claim: $\mathbf{y}$ is closer to $\mathbf{c}$ than to any other codeword. Assume the opposite, that is, there exists a $\mathbf{w} \in \mathcal{C}$ with $\mathbf{w} \neq \mathbf{c}$ such that $d(\mathbf{w}, \mathbf{y}) < d(\mathbf{c}, \mathbf{y})$. If so,

$$\begin{aligned} d(\mathbf{c}, \mathbf{w}) &\leq d(\mathbf{c}, \mathbf{y}) + d(\mathbf{y}, \mathbf{w}) && \text{Triangle inequality} \\ &\leq s + s && d(\mathbf{w}, \mathbf{y}) < d(\mathbf{c}, \mathbf{y}) = s \text{ by assumption} \\ &\leq 2t && \text{assumed } s \leq t \\ &\leq d_{\min} - 1. \end{aligned}$$

The last statement is a contradiction, because the minimum distance of the code is $d_{\min}$, that is $d(\mathbf{c}, \mathbf{w}) \geq d_{\min}$ for any two distinct codewords $\mathbf{c}, \mathbf{w}$. □

The code in Example 1.1.2 has $d_{\min} = 7$, so it can correct 3 errors. The code in Example 1.1.3 has $d_{\min} = 3$, so it can can correct 1 error.

SSQ 3.2. *Block Code Minimum Distance* Consider the block linear code

$$\mathcal{C} = \{(0,0,0,0,0,0), (0,0,1,1,1,1), (1,1,1,0,0,0), (1,1,0,1,1,1)\}$$

What is the minimum distance $d_{\min}$ of this code? How many errors can $\mathcal{C}$ correct?

[1] Alternatively, $t < \frac{d_{\min}}{2}$, since $d_{\min}$ is an integer

3.2 Linear Block Codes

3.2.1 Definition

Recall that a vector subspace $\mathcal{V}$ is a vector space which is a subset of $\mathbb{F}_q^n$. For $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, it is closed under addition and multiplication by a scalar.

Definition 3.7. An (n, k) *linear block code* $\mathcal{C}$ is a k-dimensional subspace of a vector space $\mathbb{F}_q^n$.

The number of codewords is $M = q^k$. The dimension of the code $\mathcal{C}$ is k, and its blocklength[2] is n. The code rate is $R = \frac{1}{n} \log M = \frac{k}{n} \log q$ bits. A high-rate code can carry a large amount of information, but has low error-correction capability, compared to a low rate code, which carries less information but may have high error-correction capability. It is common to write (n, k) to indicate a linear block code with block length n and dimension k. A code with minimum distance d may be written (n, k, d).

Since a linear code $\mathcal{C}$ is itself a vector subspace, the following properties hold:

- The all-zeros vector is a codeword. That is, $\mathbf{0} = [0, 0, \ldots, 0] \in \mathcal{C}$ since the subspace must include the all-zeros vector.
- Scalar multiplication of a codeword is a codeword: $a \cdot \mathbf{c} \in \mathcal{C}$ for any $a \in \mathbb{F}$ and $\mathbf{c} \in \mathcal{C}$
- The linear combination of two codewords is a codeword: $a_1\mathbf{c}_1 + a_2\mathbf{c}_2 \in \mathcal{C}$ for $a \in \mathbb{F}$ and $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$.

For binary codes, the last item simplifies to:

- The sum of any two codewords is a codeword: $\mathbf{c}_1 + \mathbf{c}_2 \in \mathcal{C}$ for $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$.

For a binary $q = 2$ code, $M = 2^k$, and its code rate R is $\frac{1}{n} \log M = \frac{k}{n}$. Since $k \leq n$, $R \leq 1$.

3.2.2 Generator Matrix G

Since a linear block code $\mathcal{C}$ is a k-dimensional vector space, this space can be spanned by a set of k linearly independent basis vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_k$, where,

$$\mathbf{g} = [g_1, g_2, \ldots, g_n] \tag{3.22}$$

is a row vector representing a point in $\mathbb{F}_q^n$. A vector $\mathbf{c}$ is a codeword if it can be formed as a linear combination of the basis vectors,

$$\mathbf{c} = u_1\mathbf{g}_1 + u_2\mathbf{g}_2 + \cdots + u_k\mathbf{g}_k. \tag{3.23}$$

[2]It is common to call n the length of the code, since this is the number of symbols — even in Matlab, the function `length` gives the number of elements in a vector. But for a point $\mathbf{x}$ in the n-dimensional real space $\mathbb{R}^n$, the *length* of $\mathbf{x}$ is its distance from the origin, not n. To reduce confusion, let us use "block length."

where $u_i \in \mathbb{F}_q$ for $i = 1, 2, \ldots, k$.

The symbols $u_i \in \mathbb{F}_q$ are called *information symbols* and can be written as a vector

$$\mathbf{u} = [u_1, u_2, \ldots, u_k]. \tag{3.24}$$

This represents the information that is to be transmitted over the unreliable channel. The mapping from information to codewords can be written in matrix form as:

$$\mathbf{c} = \mathbf{uG}, \tag{3.25}$$

where $\mathbf{G}$ is a k-by-n matrix called a *generator matrix*:

$$\mathbf{G} = \begin{bmatrix} — & \mathbf{g}_1 & — \\ — & \mathbf{g}_2 & — \\ & \vdots & \\ — & \mathbf{g}_k & — \end{bmatrix}. \tag{3.26}$$

Any row of $\mathbf{G}$ is a codeword. Any linear combination of the rows of $\mathbf{G}$ is also a codeword. The all-zeros vector is a codeword. If the k vectors $\mathbf{g}_1, \ldots, \mathbf{g}_k$ are linearly independent, then $\mathbf{G}$ is full rank. The entire codebook $\mathcal{C}$ can be found by multiplying all possible sequences $\mathbf{u} \in \mathbb{F}_q^k$ by $\mathbf{G}$.

Example 3.7. Find the code parameters n, k, R, M and the codebook $\mathcal{C}$ for the binary $q = 2$ code with generator matrix $\mathbf{G}$:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{3.27}$$

The code dimension is $k = 3$ and the block length is $n = 5$, the code rate is $R = k/n = 3/5$, and the number of codewords is $M = 2^k = 8$. The codebook $\mathcal{C}$ is shown in the right column of the table below, obtained from $\mathbf{uG}$ for all $\mathbf{u} = (u_1, u_2, u_3) \in \mathbb{F}_2^3$:

information $\mathbf{u}$	codeword $\mathbf{c}$
$[0, 0, 0]$	$[0, 0, 0, 0, 0]$
$[1, 0, 0]$	$[1, 0, 0, 1, 0]$
$[0, 1, 0]$	$[0, 1, 0, 0, 1]$
$[1, 1, 0]$	$[1, 1, 0, 1, 1]$
$[0, 0, 1]$	$[0, 0, 1, 1, 1]$
$[1, 0, 1]$	$[1, 0, 1, 0, 1]$
$[0, 1, 1]$	$[0, 1, 1, 1, 0]$
$[1, 1, 1]$	$[1, 1, 1, 0, 0]$

It is easy to verify that the sum of any two codewords is also a codeword. For example the sum of the last two codewords $[0\ 1\ 1\ 1\ 0] + [1\ 1\ 1\ 0\ 0]$ is the second codeword $[1\ 0\ 0\ 1\ 0]$.

Example 3.8. Find all codewords of the code defined over $\mathbb{F}_3$, with generator matrix:

$$\begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}.$$

The codewords are given by $\mathbf{uG}$ where $\mathbf{u} = [u_1, u_2] \in \{0, 1, 2\}^2$:

$$\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 2 \\ 0 & 2 & 0 & 1 & 1 \\ 1 & 0 & 1 & 2 & 0 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 0 & 1 \\ 2 & 0 & 2 & 1 & 0 \\ 2 & 1 & 2 & 0 & 2 \\ 2 & 2 & 2 & 2 & 1 \end{array}$$

SSQ 3.3. *Generator Matrix G* Consider a code with the following generator matrix $\mathbf{G}$:

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

What is the dimension of this code? What is the block length of this code? How many codewords does the code have?

3.2.3 Parity-Check Matrix

A parity-check matrix $\mathbf{H}$ specifies constraints that all codewords $\mathbf{c} \in \mathcal{C}$ must satisfy, and is an alternative method to describe a linear block code. First, a parity check is described.

Definition 3.8. A *parity check* for a code with k-by-n generator matrix $\mathbf{G}$ is a 1-by-n vector $\mathbf{h}$ that satisfies:

$$\mathbf{G} \cdot \mathbf{h}^{\mathrm{t}} = \mathbf{0}, \tag{3.28}$$

where $\mathbf{0}$ is the k-by-1 all-zeros column vector.

Example 3.9. Consider a dimension $k = 2$ code with $\mathbf{G}$ given below, and $\mathbf{h}_1 = (1, 1, 0, 1, 0)$. Then,

$$\underbrace{\begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}}_{\mathbf{G}} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{3.29}$$

can be confirmed to be equal to $[0\ 0\ 0]^{\mathrm{t}}$ and $\mathbf{h}$ is a parity check for this code.

Definition 3.9. For a block length n code with dimension k, a *parity-check matrix* $\mathbf{H}$ is a matrix with n columns and $n-k$ linearly independent parity checks in its rows. Wikipedia: Parity check matrix

The parity-check matrix can be represented as:

$$\mathbf{H} = \begin{bmatrix} — & \mathbf{h}_1 & — \\ — & \mathbf{h}_2 & — \\ & \vdots & \\ — & \mathbf{h}_m & — \end{bmatrix}, \tag{3.30}$$

where $m \geq (n-k)$. $\mathbf{H}$ has $n-k$ linearly independent parity check vectors. If $m = n-k$, then all the parity check vectors are linearly independent. If $\mathbf{H}$ has more than $n-k$ rows, then at least one row is linearly dependent. Since each row of $\mathbf{H}$ is a parity check:

$$\mathbf{G}\mathbf{H}^{\mathrm{t}} = \mathbf{0}, \tag{3.31}$$

where $\mathbf{0}$ is a $k \times m$ all-zeros matrix.

Example 3.10. Continuing Example 3.9, it is easy to verify that $\mathbf{h}_2 = (0,1,0,0,1)$ and $\mathbf{h}_3 = (1,1,1,0,1)$ are also parity checks. Since the three parity checks $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are linearly independent, they can be used to form a parity-check matrix $\mathbf{H}$, which satisfies (3.31):

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}^{\mathrm{t}}}_{\mathbf{H}^{\mathrm{t}}} = \mathbf{0}, \tag{3.32}$$

where $\mathbf{0}$ is the 2×3 all-zeros matrix.

The syndrome $\mathbf{s}$ of a sequence $\mathbf{y}$ is $\mathbf{s} = \mathbf{y}\mathbf{H}^{\mathrm{t}}$.

Definition 3.10. For a code $\mathcal{C}$ with $m \times n$ parity check matrix $\mathbf{H}$, the *syndrome* $\mathbf{s}$ of a sequence $\mathbf{y}$ is:

$$\mathbf{s} = \mathbf{y}\mathbf{H}^{\mathrm{t}}, \tag{3.33}$$

where $\mathbf{s}$ is a $1 \times m$ vector and $\mathbf{y}$ is a $1 \times n$ vector.

The following proposition shows that a parity check matrix can be used to test whether a given sequence is a codeword or not. The syndrome of a codeword is $\mathbf{0}$. That is, $\mathbf{x} \in \mathcal{C}$ if and only if $\mathbf{x}\mathbf{H}^{\mathrm{t}} = \mathbf{0}$.

Proposition 3.2. A sequence $\mathbf{x}$ is a codeword if and only if $\mathbf{x}\mathbf{H}^{\mathrm{t}} = \mathbf{0}$.

Proof Assume $\mathbf{x}$ is a codeword. There is some $\mathbf{u}$ such that $\mathbf{x} = \mathbf{uG}$. Then, $\mathbf{xH}^{\mathrm{t}} = \mathbf{uGH}^{\mathrm{t}} = \mathbf{x} \cdot \mathbf{0} = \mathbf{0}$. Conversely, if $\mathbf{x}$ is not a codeword, then at least one of $\mathbf{xh}_1^{\mathrm{t}}, \mathbf{xh}_2^{\mathrm{t}}, \ldots, \mathbf{xh}_m^{\mathrm{t}}$ is nonzero and $\mathbf{xH}^{\mathrm{t}} \neq \mathbf{0}$ — otherwise $\mathbf{x}$ would be a codeword (here $\mathbf{h}_i$ are the m parity checks, the rows of $\mathbf{H}$). Thus, $\mathbf{xH}^{\mathrm{t}} = \mathbf{0}$ implies $\mathbf{x} \in \mathcal{C}$. (If $\mathbf{H}$ contains an all-zero column, the corresponding symbol is independent of the codeword.) □

SSQ 3.4. *Test for Codebook Membership* Consider the code $\mathcal{C}$ with parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{3.34}$$

Is $[0, 0, 1, 0, 0, 1]$ a codeword of $\mathcal{C}$? Is $[1, 0, 1, 0, 1, 0]$? Is $[1, 0, 0, 1, 1, 0]$?

3.2.4 Minimum Distance

The minimum distance of an error correcting code was given in Definition 3.6. For a linear block code, the minimum distance is the minimum of the weight of non-zero codewords. That is,

$$d_{\min} = \min_{\mathbf{c} \in \mathcal{C}, \mathbf{c} \neq 0} w(\mathbf{c}) \tag{3.35}$$

This can be seen by observing:

$$d_{\min} = \min_{\mathbf{x}, \mathbf{y} \in \mathcal{C}, \mathbf{x} \neq \mathbf{y}} d(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{c} \in \mathcal{C}, \mathbf{c} \neq 0} d(\mathbf{c}, 0) = \min_{\mathbf{c} \in \mathcal{C}, \mathbf{c} \neq 0} w(\mathbf{c}) \tag{3.36}$$

using $\mathbf{c} = \mathbf{x} - \mathbf{y}$.

In this way, the minimum distance of a code may be found from the minimum of the weights of the codewords, rather than computing the distance between all pairs of codewords.

Example 3.11. The minimum distance of the code in Example 3.7, can be found. The minimum distance is 2, since there are two nonzero codewords of weight two, and none of weight one.

For binary codes, linearly dependent columns sum to the all-zeros column vector.

Proposition 3.3. Let $\mathbf{H}$ be a parity check matrix for a code $\mathcal{C}$. Then the minimum distance of $\mathcal{C}$ is equal to the minimum number of linearly dependent columns of $\mathbf{H}$.

Example 3.12. Find the minimum distance of the code with the following parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{3.37}$$

There are no two columns in $\mathbf{H}$ which sum to 0, so the minimum distance is greater than 2. Columns 1, 2 and 3 sum to $[0\ 1\ 1]^t$, so they are not linearly dependent. But columns 2, 3 and 4 sum to $[0\ 0\ 0]^t$, so they are linearly dependent and the minimum distance of this code is 3.

SSQ 3.5. *Minimum Distance from Parity-Check Matrix* What is the minimum distance of the code with the following parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} ?$$

3.3 Systematic Forms

3.3.1 Systematic Generator Matrix

In a code's *systematic form*, the k information symbols $u_1, u_2, \ldots, u_k$ usually appear in positions 1 to k of the codeword. The remaining $m = n - k$ symbols appear in positions $k+1$ to n and are called *parity symbols*, so the codeword $\mathbf{c}$ is:

$$\mathbf{c} = [u_1\ u_2\ \ldots\ u_k\ p_1\ p_2\ \ldots\ p_m] \tag{3.38}$$

In this case, the generator matrix in systematic form is:

$$\mathbf{G} = \left[\ \mathbf{I}_k\ \middle|\ \mathbf{P}\ \right] \tag{3.39}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix and $\mathbf{P}$ is a $k \times m$ matrix. Any linear block code has a systematic form, although the systematic bits are not necessarily in the first k positions. In this case, the bit positions can be permuted so that the information bits are in the first k positions. This corresponds to swapping columns of the generator matrix.

A code $\mathcal{C}$ has many possible bases and thus many possible generator matrices — two distinct matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ may both generate a code $\mathcal{C}$, although two distinct generator matrices will give distinct mappings from information $\mathbf{u}$ to codewords $\mathbf{c}$. However, a code $\mathcal{C}$ has only one systematic generator matrix, up to permutations of the rows.

A generator matrix for a code $\mathcal{C}$ can be transformed to another generator matrix for the same code $\mathcal{C}$ using standard row operations:

1. For binary codes, replace a row with that row plus another row. (For non-binary codes, replace a row with that row plus another row times a non-zero constant),
2. Two rows may be exchanged.

To obtain the systematic generator matrix, these operations are performed until the k-by-k identity matrix appears on the left.

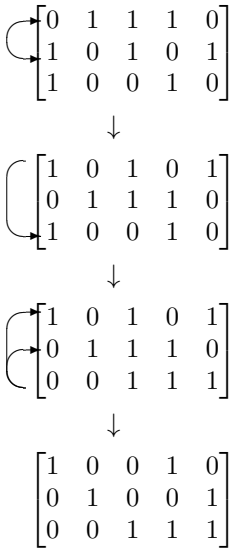


Figure 3.2: Row operations to find the generator matrix in systematic form.

Example 3.13. Consider a non-systematic generator $\mathbf{G}_{\text{non}}$:

$$\mathbf{G}_{\text{non}} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.40}$$

This can be transformed into a systematic generator matrix using row operations as shown in Fig. 3.2. The systematic matrix for the same code is:

$$\mathbf{G}_{\text{sys}} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{3.41}$$

The three information bits are the first three positions of the codeword. Note that $\mathbf{G}_{\text{non}}$ and $\mathbf{G}_{\text{sys}}$ are two generator matrices for the same code.

SSQ 3.6. *SSQ: Systematic Generator Matrix G* Find the systematic generator matrix for a code with generator matrix:

$$\mathbf{G}_1 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{3.42}$$

3.3.2 Systematic Parity-Check Matrix

The parity-check matrix also has a systematic form. If $\mathbf{H}$ is an $m \times n$ matrix, then the systematic parity check form is:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & | & \mathbf{I}_m \end{bmatrix}. \tag{3.43}$$

Row operations may be applied to a parity check matrix $\mathbf{H}$ to find the systematic form. Given an arbitrary matrix $\mathbf{H}$, it may not be possible to put the matrix into systematic form, unless column swaps are allowed.

The systematic forms can be used to convert between the parity-check matrix and generator matrix for a code. In particular, if a code parity-check matrix is $\mathbf{H} = [\mathbf{A} \mid \mathbf{I}_k]$, the generator matrix $\mathbf{G}$ for the code is:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_k & | & -\mathbf{A}^{\mathrm{t}} \end{bmatrix}. \tag{3.44}$$

For binary codes, $-\mathbf{A}^{\mathrm{t}} = \mathbf{A}^{\mathrm{t}}$.

Example 3.14. Put the following binary parity check matrix $\mathbf{H}$ in systematic form using row operations and find the systematic generator matrix for the code:

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

To form the identity matrix on the right-hand side, replace row 3 with the sum of row 2 and row 3. Replace row 2 with the sum of row 1 and row 2, giving:

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.45}$$

which is the systematic form.

Find the generator matrix. Since $\mathbf{H} = [\mathbf{A} \mid \mathbf{I}_k]$ as in (3.45), the generator matrix is given by:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_k & -\mathbf{A}^{\mathrm{t}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}. \tag{3.46}$$

Example 3.15. Put the following ternary $q = 3$ generator matrix $\mathbf{G}$ in systematic form and find the corresponding parity-check matrix $\mathbf{H}$.

$$\mathbf{G} = \begin{bmatrix} 1 & 2 & 1 & 2 & 0 \\ 2 & 2 & 0 & 2 & 2 \end{bmatrix} \tag{3.47}$$

Replace the second row with the sum of the first and second row:

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix} \tag{3.48}$$

And the replace the first row with the sum of the first and second row:

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 2 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix} \tag{3.49}$$

We have a generator matrix in systematic form. The parity check matrix in systematic form is:

$$\mathbf{H} = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{3.50}$$

3.3.3 Encoding Inverse and H Using Matrix Inversion

It is possible to find the inverse of the encoding operation, which maps codewords back to information, using matrix inversion. This same matrix inverse also allows converting between $\mathbf{G}$ and $\mathbf{H}$.

For an (n, k) code $\mathcal{C}$ with generator matrix $\mathbf{G}$, we already know that encoding is a mapping from an information vector $\mathbf{u}$ to a codeword $\mathbf{c} = \mathbf{u} \cdot \mathbf{G}$. The inverse of encoding, mapping from $\mathbf{c}$ to $\mathbf{u}$, is given by the $n \times k$ matrix $\mathbf{E}$:

$$\mathbf{u} = \mathbf{c} \cdot \mathbf{E}. \tag{3.51}$$

If $\mathbf{G}$ is systematic, then the inverse mapping is trivial, as k positions of $\mathbf{c}$ contain the information $\mathbf{u}$. This encoding inverse is of interest for non-systematic generator matrices. Note that $\mathbf{E}$ is a kind of pseudoinverse of $\mathbf{G}$, that is, $\mathbf{G} \cdot \mathbf{E} = \mathbf{I}_k$.

Let $\mathbf{G}$ be a $k \times n$ full-rank generator matrix over $\mathbb{F}$ for code $\mathcal{C}$. Form $\widetilde{\mathbf{G}}$ by adding $n - k$ linearly independent rows $\mathbf{G}'$ to form an $n \times n$ full rank matrix $\widetilde{\mathbf{G}}$. This $\mathbf{G}'$ may be arbitrarily chosen so that $\widetilde{\mathbf{G}}$ is full rank:

$$\widetilde{\mathbf{G}}' = \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix}. \tag{3.52}$$

Let $\widetilde{\mathbf{H}} = \widetilde{\mathbf{G}}^{-1}$, that is $\widetilde{\mathbf{G}} \cdot \widetilde{\mathbf{H}} = \mathbf{I}_n$ where matrix operations are over $\mathbb{F}$. Denote by $\mathbf{E}$ the left k columns of $\widetilde{\mathbf{H}}$ and denote by $\mathbf{H}^{\mathrm{t}}$ the right $n - k$ columns of $\widetilde{\mathbf{H}}$:

$$\widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{E} & \mathbf{H}^{\mathrm{t}} \end{bmatrix}. \tag{3.53}$$

Proposition 3.4. Let $\mathbf{G}$ be a generator matrix for $\mathcal{C}$. If $\widetilde{\mathbf{G}}$ is an $n \times n$ full rank matrix defined as above and $\widetilde{\mathbf{H}} = [\mathbf{E}\ \mathbf{H}^{\mathrm{t}}]$ satisfies $\widetilde{\mathbf{H}} \cdot \widetilde{\mathbf{G}} = \mathbf{I}_n$, then:

1. $\mathbf{E}$ is the encoding inverse operation,
2. $\mathbf{H}$ is a parity check matrix for $\mathcal{C}$.

The proof is given as an exercise.

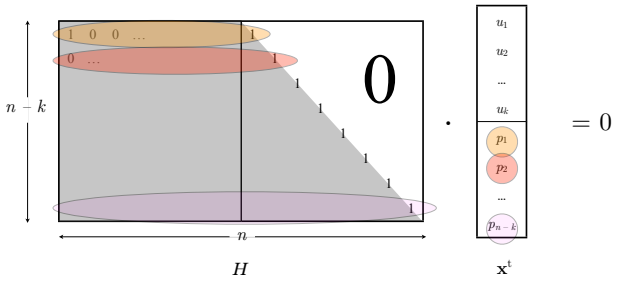


Figure 3.3: Systematic encoding by back-substitution.

Example 3.16. Find the encoding inverse $\mathbf{E}$ and the parity-check matrix $\mathbf{H}$ for the (5,3) code $\mathcal{C}$ with generator matrix $\mathbf{G}$:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Add $n - k = 2$ linearly independent rows to form $\widetilde{\mathbf{G}}$, and find $\widetilde{\mathbf{H}} = \widetilde{\mathbf{G}}^{-1}$:

$$\widetilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \widetilde{\mathbf{H}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Finding linearly independent rows can be aided if $\mathbf{G}$ is in triangular form — add rows so that $\widetilde{\mathbf{G}}$ is triangular with 1's on the diagonal. Such matrices are full rank.

Then we can find $\mathbf{E}$ by taking the left k columns of $\widetilde{\mathbf{H}}$ and find $\mathbf{H}$ by taking the right $n - k$ columns of $\widetilde{\mathbf{H}}$:

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{H} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3.3.4 Encoding Using H

Encoding can be performed using the parity-check matrix $\mathbf{H}$ written as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P} & | & \mathbf{T} \end{bmatrix} \tag{3.54}$$

where $\mathbf{T}$ is a $m \times m$ lower triangular matrix with no zeros on the diagonal. This encoding does not require the generator matrix.

Encoding of data $\mathbf{u}$ to codeword $\mathbf{c}$ can be performed by writing $\mathbf{c} = \begin{bmatrix}\mathbf{u} & \mathbf{p}\end{bmatrix}$; this encoding is systematic. To find the unknown parity $\mathbf{p}$, recognize that any codeword $\mathbf{c}$ must satisfy $\mathbf{H}\mathbf{c}^{\mathrm{t}} = \mathbf{0}$ (Proposition 3.2):

$$\begin{bmatrix}\mathbf{P} & | & \mathbf{T}\end{bmatrix} \cdot \begin{bmatrix}\mathbf{u} \\ p_1 \\ p_2 \\ \vdots \\ p_m\end{bmatrix} = \mathbf{0} \tag{3.55}$$

This can be solved row-by-row, starting with the top row, first finding p_1, then $p_2, p_3, \ldots$ in order, as shown in the following example. This procedure is sometimes called back substitution, see Fig. 3.3.

Example 3.17. Encode information $\mathbf{u} = [1, 1, 0]$ to the systematic codeword of the code with parity-check matrix $\mathbf{H}$ in lower triangular form:

$$\mathbf{H} = \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \end{array}\right]. \tag{3.56}$$

This can be done by solving:

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = \mathbf{0}. \tag{3.57}$$

The top row of the above equality is $1+0+0+p_1 = 0$, which means that $p_1 = 1$. The second row is $0 + 0 + 0 + p_1 + p_2 = 0$; since $p_1 = 1$, we have $p_2 = 1$. The last row is $1 + 1 + 0 + p_1 + 0 + p_3 = 0$, so $p_3 = 1$, giving:

$$[1, 1, 0, 1, 1, 1] \tag{3.58}$$

which is the systematic codeword.

If the parity-check matrix is in nearly lower-triangular form, sometimes called approximate-lower triangular (ALT) form, then efficient encoding is still possible. This is discussed for LDPC codes, see Section 5.4.

3.4 Important Linear Block Codes

This section describes some important linear block codes. Recall that an (n, k, d) linear block code has block length n, dimension k and minimum distance d.

3.4.1 Repeat Code

Definition 3.11. For $n \geq 1$, the *repeat code* is an $(n, 1, n)$ code with generator matrix:

$$\mathbf{G}_{\text{repeat}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \tag{3.59}$$

In the binary case, the codebook has just two codewords, $\mathcal{C} = \{(00\ldots0), (11\ldots1)\}$. For ternary codes, the $\mathcal{C} = \{(00\ldots0), (11\ldots1), (22\ldots2)\}$ The rate of this code is $R = \frac{1}{n}$ and the minimum distance is $d_{\min} = n$. This is a low-rate code with high minimum distance.

3.4.2 Single-Parity Check Code

Definition 3.12. For $n \geq 2$, the *single parity-check code* is an $(n, n-1, 2)$ code with parity-check matrix:

$$\mathbf{H}_{\text{SPC}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{3.60}$$

The rate is $R = \frac{n-1}{n}$ and the minimum distance is $d_{\min} = 2$. This is a high-rate code with low minimum distance.

A binary single-parity check code or SPC code consists of all binary sequences of length n $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ that sum to 0, modulo 2, that is, they have even weight:

$$\mathcal{C} = \{\mathbf{c} : \sum_{i=1}^{n} c_i = 0\} \tag{3.61}$$

When SPC codeword is written in systematic form:

$$\mathbf{c} = (u_1, u_2, \ldots, u_{n-1}, p_1)$$

the parity bit p_1 can be found from the information bits u_1 to u_{n-1}:

$$p_1 = \sum_{i=1}^{n-1} u_i. \tag{3.62}$$

That is, choose p_1 so that there is an even number of ones.

For example if $n = 4$, then the binary SPC codebook consists of all length 4 sequences with even weight:

$$\mathcal{C} = \{0000, 1001, 0101, 1100, 0011, 1010, 0110, 1111\}. \tag{3.63}$$

3.4.3 Hamming Code

Hamming codes are high-rate binary codes with a minimum distance of three. Extended Hamming codes have minimum distance of four.

Definition 3.13. For $m \geq 2$, a binary *Hamming code* is a $(2^m-1, 2^m-m-1, 3)$ code whose parity check matrix consists of all non-zero binary m vectors as columns.

A Hamming code has length $n = 2^m - 1$, for $m = 2, 3, 4, \ldots$. The number of information symbols is $k = 2^m - 1 - m$. The Hamming code has a parity-check matrix of size $m \times 2^m - 1$.

Example 3.18. The parity check matrix for the Hamming code with $m = 3$ is:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{3.64}$$

which has all non-zero binary vectors in columns.

Write the codeword as $\mathbf{c} = [u_1, u_2, u_3, u_4, p_1, p_2, p_3]^{\mathrm{t}}$, where u_i are systematic bits and p_i are parity bits. From $\mathbf{Hc} = \mathbf{0}$, the parity-check equations can be written as:

$$u_1 + u_2 + u_3 + p_1 = 0 \tag{3.65}$$
$$u_1 + u_2 + u_4 + p_2 = 0 \tag{3.66}$$
$$u_1 + u_3 + u_4 + p_3 = 0 \tag{3.67}$$

Given four information bits u_1, u_2, u_3, u_4, the parity bits can be found as $p_1 = u_1+u_2+u_3$ since $u_i = -u_i$ in modulo-2 arithmetic. This corresponds to Example 1.1.3.

The Hamming code has $d_{\min} = 3$, which can be shown using Proposition 3.3. Since the columns of $\mathbf{H}$ consist of all non-zero binary vectors, there are no two columns which sum to the zero column $\mathbf{0}$, so the minimum distance must be greater than 2. But for any two column vectors, there is always a third equal to their sum, so there exist 3 column vectors that sum to $\mathbf{0}$. So, there are 3 linearly dependent column vectors and the minimum distance of the Hamming code is 3.

Extended Codes Any code can be extended. If a code $\mathcal{C}$ has odd minimum distance d, then the extended code $\mathcal{C}'$ will have minimum distance $d + 1$. The parity-check matrix of $\mathcal{C}'$ is obtained by adding the all-zeros column vector to the parity-check matrix of $\mathcal{C}$, then adding the all-ones row vector.

Specifically for Hamming codes, an *extended Hamming code* of length $n = 2^m$ exists for $m = 2, 3, 4, 5, \ldots$. The extended Hamming code has $d_{\min} = 4$.

Example 3.19. The parity check matrix for the $(8, 4, 4)$ extended Hamming code with $m = 3$ is:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{3.68}$$

The elements added to the Hamming code to form the extended Hamming code are colored.

SSQ 3.7. What are the code rates R of the Hamming codes with $m = 4, 5$ and 6?

3.4.4 First-Order Reed-Muller Codes

Reed-Muller codes are family of error-correcting codes. For block length $n \leq 32$, they are among the best-known codes. For $m \geq 1$, let $\mathbf{v}_0 = [1, 1, \cdots, 1]$ be the all-ones vector of block length 2^m. Let $\mathbf{V}$ be the m-by-2^m matrix consisting of all binary m vectors as columns. Let row i of this matrix be $\mathbf{v}_i$, $i = 1, \ldots, m$. For example for $m = 3$:

$$\mathbf{v}_0 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ and}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Wikipedia: Reed–Muller code

Definition 3.14. For $m \geq 1$, a *first-order Reed-Muller code* is a $(2^m, m + 1)$ block code where the first row of the generator matrix is the all-ones vector and rows 2 to $m + 1$ consist of all binary m vectors as columns.

Example 3.20. For $m = 2$, the $(4, 3)$ first-order Reed-Muller code has generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{3.69}$$

For $m = 3$, the $(8, 4)$ first-order Reed-Muller code has generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{3.70}$$

The minimum distance of a first-order Reed-Muller code is 2^{m-1}. In the first-order Reed-Muller code's generator matrix, each row has weight 2^{m-1} except one row of weight 2^m.

3.4.5 Reed–Muller Codes of Higher Order

Reed-Muller codes can be generalized to higher orders. First, for two binary vectors $\mathbf{v}$, and $\mathbf{u}$, let their *binary product* denoted $\mathbf{v} \cdot \mathbf{u}$ be the symbol-by-symbol product. For example, $[0, 0, 1, 1] \cdot [0, 1, 0, 1] = [0, 0, 0, 1]$.

Recall the definition of $\mathbf{V}$ with row vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. Consider choosing r of these m vectors, and then taking the binary product of these r vectors. In total, there are $\binom{m}{r}$ possible products.

Definition 3.15. Let $0 \leq r \leq m$. An order r Reed Muller with block length $n = 2^m$ denoted $\mathrm{RM}(r, m)$, has generator matrix consisting of $\mathbf{v}_0$ and the binary products of up to r of the $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$.

An (r, m) Reed-Muller code has block length 2^m and dimension k given by:

$$k = 1 + \binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{r} \tag{3.71}$$

The rows of its generator matrix is given by binary vectors of order $0, 1, \ldots, r$. The minimum distance is 2^{m-r}.

The block length is $n = 2^m$, for an order m. For $m = 4$, the following gives the vectors for each order $r = 0, 1, 2, 3, 4$.

$$\begin{aligned}
\mathbf{1} &= 1111\ 1111\ 1111\ 1111 \quad \} && \text{order 0} \\
\mathbf{v}_1 &= 0101\ 0101\ 0101\ 0101 \\
\mathbf{v}_2 &= 0011\ 0011\ 0011\ 0011 \\
\mathbf{v}_3 &= 0000\ 1111\ 0000\ 1111 \\
\mathbf{v}_4 &= 0000\ 0000\ 1111\ 1111 \quad \} && \text{order 1} \\
\mathbf{v}_1 \cdot \mathbf{v}_2 &= 0001\ 0001\ 0001\ 0001 \\
\mathbf{v}_1 \cdot \mathbf{v}_3 &= 0000\ 0101\ 0000\ 0101 \\
\mathbf{v}_1 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0101\ 0101 \\
\mathbf{v}_2 \cdot \mathbf{v}_3 &= 0000\ 0011\ 0000\ 0011 \\
\mathbf{v}_2 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0011\ 0011 \\
\mathbf{v}_3 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0000\ 1111 \quad \} && \text{order 2} \\
\mathbf{v}_1 \cdot \mathbf{v}_2 \cdot \mathbf{v}_3 &= 0000\ 0001\ 0000\ 0001 \\
\mathbf{v}_1 \cdot \mathbf{v}_2 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0001\ 0001 \\
\mathbf{v}_1 \cdot \mathbf{v}_3 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0000\ 0101 \\
\mathbf{v}_2 \cdot \mathbf{v}_3 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0000\ 0011 \quad \} && \text{order 3} \\
\mathbf{v}_1 \cdot \mathbf{v}_2 \cdot \mathbf{v}_3 \cdot \mathbf{v}_4 &= 0000\ 0000\ 0000\ 0001 \} && \text{order 4}
\end{aligned}$$

To construct a Reed-Muller code of order r, the generator vectors are the products of order r and lower.

Example 3.21. The RM(4, 2) is a (16,11,4) code with generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1111\ 1111\ 1111\ 1111 \\ 0101\ 0101\ 0101\ 0101 \\ 0011\ 0011\ 0011\ 0011 \\ 0000\ 1111\ 0000\ 1111 \\ 0000\ 0000\ 1111\ 1111 \\ 0001\ 0001\ 0001\ 0001 \\ 0000\ 0101\ 0000\ 0101 \\ 0000\ 0000\ 0101\ 0101 \\ 0000\ 0011\ 0000\ 0011 \\ 0000\ 0000\ 0011\ 0011 \\ 0000\ 0000\ 0000\ 1111 \end{bmatrix} \tag{3.72}$$

Several codes can be seen as instances of Reed-Muller codes, making RM codes a generalization of other codes:

- RM(0,m) codes are Repetition code of length $n = 2^m$, code rate $R = 1/n$ and $d_{\min} = n$.
- For $m = 1, 3, 5, \ldots$, certain RM codes have $k = n/2$ and are self-dual codes.
- RM($m-1$,m) codes are single-parity check code of length 2^m, code rate $R = \frac{n-1}{n}$ and minimum distance $d_{\min} = 2$.
- RM($m-2$,m) codes are extended Hamming codes of length 2^m, and minimum distance $d_{\min} = 4$.

These relationships are illustrated in Fig. 3.4.

3.5 Weight Distribution (Distance Spectrum)

The weight distribution of a code is the number of codewords for each weight w. The weight distribution is sometimes called distance spectrum.

Definition For a block length n linear code, the *weight distribution* or *distance spectrum* is a sequence $A_0, A_1, A_2, \cdots, A_n$ where A_w is the number of codewords with weight w, usually written as a polynomial:

$$A(z) = \sum_{w=0}^{n} A_w z^w. \tag{3.73}$$

This polynomial is called the weight enumerator polynomial or weight distribution polynomial, and is convenient for writing the distance spectrum. Here z is just a placeholder variable or dummy variable. For any linear code, $A_0 = 1$, that is there is one all-zeros codeword. After A_0, the next non-zero term is at the minimum distance, that is $A_{d_{\min}}$. $A(1)$ is the number of codewords, that is $A(1) = \sum_w A_w = q^k$.

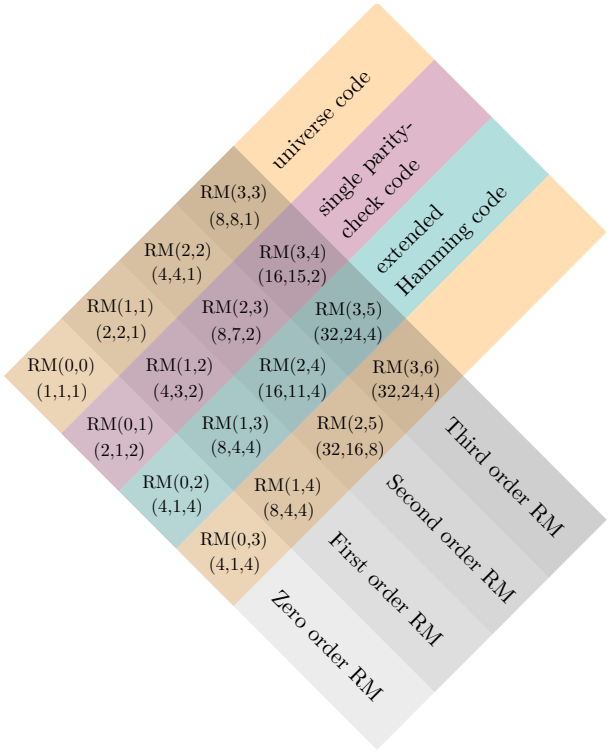


Figure 3.4: Connections between Reed-Muller codes and other codes. The universe code is the set of all binary sequences $\mathbb{F}_2^n$.

The following is a list of all codewords of the $(7, 4)$ Hamming code, and their weights:

codeword $\mathbf{x}$	$\mathrm{wt_H}(\mathbf{x})$	codeword $\mathbf{x}$	$\mathrm{wt_H}(\mathbf{x})$
0 0 0 0 0 0 0	0	1 1 0 1 0 0 1	4
1 1 1 0 0 0 0	3	0 0 1 1 0 0 1	3
1 0 0 1 1 0 0	3	0 1 0 0 1 0 1	3
0 1 1 1 1 0 0	4	1 0 1 0 1 0 1	4
0 1 0 1 0 1 0	3	1 0 0 0 0 1 1	3
1 0 1 1 0 1 0	4	0 1 1 0 0 1 1	4
1 1 0 0 1 1 0	4	0 0 0 1 1 1 1	4
0 0 1 0 1 1 0	3	1 1 1 1 1 1 1	7

From this, $A_0 = 1, A_3 = 7, A_4 = 7, A_7 = 1$ and all other $A_w = 0$. The weight distribution polynomial is:

$$A(z) = 1 + 7z^3 + 7z^4 + z^7 \tag{3.74}$$

3.6 Exercises

3.1 Consider a non-linear code $\mathcal{C}_{\mathrm{NL}}$ with four codewords:

$$\mathbf{c}_1 = [1,0,0,1,0]$$
$$\mathbf{c}_2 = [0,1,0,0,1]$$
$$\mathbf{c}_3 = [1,0,1,0,1]$$
$$\mathbf{c}_4 = [0,1,1,1,0]$$

(a) Find the minimum distance of the non-linear code $\mathcal{C}_{\mathrm{NL}}$ above.
(b) Add the minimal number of codewords to $\mathcal{C}_{\mathrm{NL}}$ to form a binary linear code $\mathcal{C}_{\mathrm{L}}$. Find the systematic generator matrix for $\mathcal{C}_{\mathrm{L}}$. What is the dimension and minimum distance of $\mathcal{C}_{\mathrm{L}}$?

3.2 Using encoding by back-substitution, for the code with parity-check matrix:

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

encode the information sequences $\mathbf{u}_1 = (1,0,1)$ and $\mathbf{u}_2 = (0,1,0)$ to the corresponding codewords $\mathbf{c}_1$ and $\mathbf{c}_2$.

3.3 Find the systematic generator for the code with the parity check matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Note that $\mathbf{H}$ is not full rank.

3.4 What is the minimum distance of the code with the following parity-check matrix? How many codewords does the code have?

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

3.5 For a code over $\mathbb{F}_5$ with generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 3 & 1 \\ 0 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 1 & 3 \end{bmatrix},$$

find the parity check matrix $\mathbf{H}$ in systematic form. For $\mathbb{F}_5$, use modulo 5 addition and multiplication.

3.6 Consider a code $\mathcal{C}$ with generator matrix:

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{3.78}$$

(a) Write the systematic generator matrix for $\mathcal{C}$
(b) How many *distinct* codewords does $\mathcal{C}$ have? Write all the codewords.
(c) What is the block length, dimension and minimum distance of $\mathcal{C}$?
(d) How did you find the minimum distance?

3.7 Prove the following statements.

(a) Let $\mathbf{x}$ be a sequence with even Hamming weight, and let $\mathbf{y}$ be a sequence with Hamming weight 1. Show that $\mathrm{d}(\mathbf{x}, \mathbf{y})$ is odd. Generalize this to $\mathbf{x}$ even Hamming weight and $\mathbf{y}$ odd Hamming weight.
(b) For a linear block code $\mathcal{C}$ over $\mathbb{F}_2$, prove that either (1) $\mathcal{C}$ has no codewords of odd weight, or (2) that half the codewords of $\mathcal{C}$ have odd weight.
(c) For a code $\mathcal{C}$ with parity-check matrix $\mathbf{H}$, its extended code $\mathcal{C}'$ has parity-check matrix $\mathbf{H}'$ obtained by adding one column of zeros and one row of ones:

$$\mathbf{H}' = \begin{bmatrix} & & & 0 \\ & \mathbf{H} & & \vdots \\ & & & 0 \\ 1 & \cdots & 1 & 1 \end{bmatrix}. \tag{3.81}$$

Let the minimum distance $d_{\min}$ of a code $\mathcal{C}$ be odd. Prove that the extended code $\mathcal{C}'$ has minimum distance $d_{\min} + 1$.

3.8 Consider Proposition 3.4 for finding the encoding inverse and parity-check matrix for $\mathcal{C}$.

(a) Let code $\mathcal{C}$ have generator matrix:

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

Using Proposition 3.4, find the encoding inverse $\mathbf{E}$ and the parity-check matrix $\mathbf{H}$.
(b) Prove Proposition 3.4. Hint: what is $\begin{bmatrix} \mathbf{u} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{G} \\ \mathbf{G}' \end{bmatrix}$?

Chapter 4

Decoding Linear Block Codes

4.1 Principles of Decoding

4.1.1 Optimal Decoding

There are various types of decoding algorithms, that depend on the channel model and the code being decoded.

Maximum-likelihood (ML) decoding ML decoding is an optimal decoding metric. It selects the codeword that maximizes the likelihood $\Pr(\mathbf{y}|\mathbf{c})$, that is the output is $\widehat{\mathbf{c}}$:

$$\widehat{\mathbf{c}} = \arg\max_{\mathbf{c}\in\mathcal{C}} \Pr(\mathbf{y}|\mathbf{c}) \tag{4.1}$$

ML decoding is often regarded as optimal decoding. However, for general codes ML decoding has exponential complexity and is too complicated and lower-complexity algorithms are often used.

A posteriori probability (APP) decoding APP decoding outputs soft probabilities, say $r(c)$:

$$r(c) = \Pr(c_i = c|\mathbf{y}). \tag{4.2}$$

APP decoding is a component in other algorithms, where the soft output is passed to another decoder.

Maximum a posteriori (MAP) decoding MAP decoding is symbol-by-symbol (or bit-by-bit, in the binary case). It finds the code symbol c_i which maximizes $\Pr(c_i|\mathbf{y})$, that is $\widehat{c}_i$ is:

$$\widehat{c}_i = \arg\max_{c} \Pr(c_i|\mathbf{y}) \tag{4.3}$$

for $i = 1, 2, \ldots, n$. It gives the highest probability codeword *symbol* c_i, on a symbol-by-symbol basis, whereas ML decoding gives the mostly likely *codeword*. MAP decoding can be seen as applying a hard decision to the output as APP decoding.

Soft-input decoding vs Hard-input decoding A soft-input decoder accepts real values, usually the output of the AWGN channel or probabilities. A hard-input decoder for a binary code accepts only 0 and 1 as inputs (or more generally code symbols as inputs; erasure symbols may also be accepted as input). Hard inputs can be formed from a soft-output channels by making hard decisions symbol-by-symbol. However, discarding soft information by making hard decisions reduces the error-rate performance in general.

Bounded-distance decoding For a code with minimum distance $d_{\min}$, the error-correction capability is $t = \lfloor \frac{d_{\min}-1}{2} \rfloor$, as was shown in Proposition 3.1. Each codeword is at the center of a non-overlapping Hamming ball of radius t. In bounded-distance decoding, the decoder will output the codeword $\widehat{\mathbf{c}}$ if the received sequence $\mathbf{y}$ the Hamming distance between $\mathbf{y}$ and $\widehat{\mathbf{c}}$ is less than or equal to t. But, since spheres do not exactly cover the entire space $\mathbb{F}_q^n$ in general, a failure to decode event E occurs when the received sequence is outside of a sphere. The Berelkamp-Massey algorithm for decoding Reed-Solomon codes and BCH codes is an example of a bound-distance decoder.

4.1.2 Probabilistic Decoder Messages

Probabilities conveniently represent messages used by decoding algorithms. An advantage of probabilistic messages is that they can be used for various types of channels. Many decoding algorithms are probabilistic. Rather than working with the channel output y directly, probabilistic algorithms work with probabilities about y. Messages in both the probability domain and the log probability domain are considered.

Recall from Chapter 1 that channel models such as the BSC and AWGN channel are specified by a conditional probability distribution $p_{\mathsf{Y}|\mathsf{X}}(y|x)$. For an input sequence $\mathbf{x}$ and output sequence $\mathbf{y}$, these channels are memoryless, meaning:

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p_{\mathsf{Y}|\mathsf{X}}(y_i|x_i) \tag{4.4}$$

Consider a single use of a BI-AWGN channel, where the input is $x = 1$ and the output is $y = 0.8$. While it seems natural to represent the channel output by just y, it will be more convenient to instead use a probabilistic message for y.

Probability Domain P0, P1 If x was transmitted and y was received, then we consider

$$r(x) = \Pr(\mathsf{X} = x|\mathsf{Y} = y), \tag{4.5}$$

When x is binary there are two values $r(0)$ and $r(1)$. Since $r(0)+r(1) = 1$, we only need to know one of those two values and there are two probability domains:

$$\text{P0 domain: } r(0) = \Pr(\mathsf{X} = 0|\mathsf{Y} = y) \text{ and} \tag{4.6}$$
$$\text{P1 domain: } r(1) = \Pr(\mathsf{X} = 1|\mathsf{Y} = y). \tag{4.7}$$

Log Domain When x is binary, the log domain can be used. Define:

$$R = \log \frac{r(0)}{r(1)} = \log \frac{\Pr(\mathsf{X} = 0|\mathsf{Y} = y)}{\Pr(\mathsf{X} = 1|\mathsf{Y} = y)} \tag{4.8}$$

It is possible to translate back to the probability domain by:

$$r(0) = \frac{e^R}{1 + e^R} \text{ and } r(1) = \frac{1}{1 + e^R} \tag{4.9}$$

Here, a lower case letter indicates the probability domain, and an upper case letter indicates the log domain.

Example 4.1. Consider the BSC channel with error probability p, the channel output is $y \in \{0, 1\}$ where $\Pr(\mathsf{X} = 1|\mathsf{Y} = 0) = \Pr(\mathsf{X} = 0|\mathsf{Y} = 1) = p$ and $\Pr(\mathsf{X} = 0|\mathsf{Y} = 0) = \Pr(\mathsf{X} = 1|\mathsf{Y} = 1) = 1 - p$. If $y = 0$ is received then the P0 domain message is $r(0) = 1 - p$. If $y = 1$ is received then the P0 domain message is $r(0) = p$. The log domain message L is:

$$R = \begin{cases} \log \frac{1-p}{p} & y = 0 \\ \log \frac{p}{1-p} & y = 1 \end{cases} \tag{4.10}$$

Example 4.2. Consider the binary-input AWGN channel with noise power σ^2. The transmitter maps code bits c to transmitted symbols x as $\{0, 1\} \to \{+1, -1\}$:

$$\Pr[\mathsf{Y} = y|\mathsf{X} = x] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}. \tag{4.11}$$

The channel output is y and the log domain message L is given by:

$$R = \log e^{-\frac{(y-1)^2}{2\sigma^2}} - \log e^{-\frac{(y+1)^2}{2\sigma^2}} \tag{4.12}$$
$$R = \frac{2}{\sigma^2} y. \tag{4.13}$$

That is, for the binary-input AWGN channel, convert from received symbols y to log domain by multiplying by $\frac{2}{\sigma^2}$.

Log-Likelihood Ratio (LLR) A likelihood has the form $\Pr(y|x)$. The log-likelihood ratio (LLR) is:

$$R = \log \frac{\Pr(\mathsf{Y} = y|\mathsf{X} = 0)}{\Pr(\mathsf{Y} = y|\mathsf{X} = 1)}$$

Most commonly, $\Pr(\mathsf{X} = 0) = \Pr(\mathsf{X} = 1) = \frac{1}{2}$, and by Bayes rule, the LLR is equal to the log domain message.

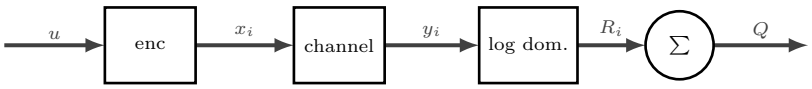


Figure 4.1: Block diagram for repeat code.

SSQ 4.1. *Interpretation of LLR Values* Suppose the log domain value is $R = 8$. Is this high or low probability of a 0 or a 1?

4.1.3 Decoding the Repeat Code

Consider decoding the binary $(n, 1)$ repeat code. The information is $x \in \{0, 1\}$ and codeword $\{000\ldots0, 111\ldots1\}$ is transmitted over the channel. The channel output $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is received. The APP estimate in the log domain is Q:

$$Q = \log \frac{\Pr[\mathsf{X} = 0 | \mathsf{Y} = \mathbf{y}]}{\Pr[\mathsf{X} = 1 | \mathsf{Y} = \mathbf{y}]}$$

Here X and Y_j are random variables corresponding to x and y_j respectively. For the repeat code, the log APP value is:

$$Q = \sum_{i=1}^{n} R_i \tag{4.14}$$

where R_i is the log domain value (4.8). This can be shown using Bayes rule and independence of the channel noise. The MAP estimate is obtained from the hard decision:

$$\widehat{x} = \begin{cases} 0 & \text{if } Q \geq 0 \\ 1 & \text{if } Q < 0 \end{cases} \tag{4.15}$$

If $Q = 0$, in principle one may flip a coin to choose $\widehat{x}$, since $x = 0$ and $x = 1$ are equally likely. But in many cases, one may consistently choose one value, $x = 0$ as above.

Example 4.3. An $n = 3$ binary repeat code was transmitted over the BI-AWGN channel with $\sigma^2 = 0.8$, where the code symbols are mapped to channel inputs $\{0, 1\} \to \{+1, -1\}$. Find the APP and MAP estimates when

$$\mathbf{y} = [0.8, -1.4, 1.2]$$

is received.

Using (4.8), find R_i:

$$R_1 = \log \frac{\Pr[\mathsf{X}=0|\mathsf{Y}_i=y_i]}{\Pr[\mathsf{X}=1|\mathsf{Y}_i=y_i]} = \frac{2}{\sigma^2} y_1 = 2.0$$
$$R_2 = -3.5$$
$$R_3 = \quad 3.0$$
$$\text{and using (4.14) } Q = \sum_{i=1}^{n} R_i = 1.5$$

The MAP decision is $\widehat{x} = 0$ since $Q \geq 0$.

The probability domain decoding for the repeat code is:

$$q(0) = \frac{\prod_{j=1}^{n} r_j(0)}{\prod_{j=1}^{n} r_j(0) + \prod_{j=1}^{n} r_j(1)} \tag{4.16}$$
$$q(1) = \frac{\prod_{j=1}^{n} r_j(1)}{\prod_{j=1}^{n} r_j(0) + \prod_{j=1}^{n} r_j(1)}. \tag{4.17}$$

The probability domain operation multiplies the inputs $r_1(0)r_2(0)\cdots$ to find $q(0)$, and multiplies the inputs $r_1(1)r_2(1)\cdots$ to find $q(1)$. However, because $q(0) + q(1) = 1$ is required, it is necessary to normalize using the term written in the denominator. The MAP decision, or hard decision, is made using the P0 domain APP:

$$\widehat{x} = \begin{cases} 0 & \text{if } q(0) \geq 0.5 \\ 1 & \text{if } q(0) < 0.5 \end{cases}. \tag{4.18}$$

4.1.4 Decoding the Single-Parity Check Code

Consider decoding of the $(3, 2)$ single-parity check code. The 3 code symbols x_1, x_2, x_3 are transmitted over a channel, where $x_1 + x_2 + x_3 = 0$ is satisfied and the codebook is $\mathcal{C} = \{000, 011, 101, 110\}$. Suppose we want to find the APP value for c_1. For the moment, condition only on the channel values y_2, y_3 and not y_1, and call this $s_1(\cdot)$:

$$s_1(0) = \Pr(\mathsf{X}_1 = 0|y_2y_3) = \Pr(\mathsf{X}_2\mathsf{X}_3 = 00 \text{ or } 11|y_2y_3) \tag{4.19}$$
$$s_1(1) = \Pr(\mathsf{X}_1 = 1|y_2y_3) = \Pr(\mathsf{X}_2\mathsf{X}_3 = 01 \text{ or } 10|y_2y_3) \tag{4.20}$$

obtained using the theorem of total probability. Because the channel is memoryless:

$$s_1(0) = q_2(0)q_3(0) + q_2(1)q_3(1) \tag{4.21}$$
$$s_1(1) = q_2(0)q_3(1) + q_2(1)q_3(0) \tag{4.22}$$

For general n and i, the s values are:

$$s_i(0) = \sum_{x_j : \sum_{j \setminus i} x_j = 0} \prod_{k=1}^{n \setminus i} q_k(x_k) \tag{4.23}$$

$$s_i(1) = \sum_{x_j : \sum_{j \setminus i} x_j = 1} \prod_{k=1}^{n \setminus i} q_k(x_k) \tag{4.24}$$

The restriction $x_j : \sum_{j \setminus i} x_j = 0$ means to take the sum over only those code bits except i that sum to 0. Now we want to include $q_i(0), q_i(1)$, the information about y_i. Using the two estimates of x_i: q_i and s_i, we combine them to obtain the APP estimate:

$$r_i(0) = \frac{q_i(0)s_i(0)}{q_i(0)s_i(0) + q_i(1)s_i(1)}$$

$$r_i(1) = \frac{q_i(1)s_i(1)}{q_i(0)s_i(0) + q_i(1)s_i(1)}$$

The MAP decision, or hard decision, for bit position i is made using the APP values:

$$\widehat{x}_i = \begin{cases} 0 & \text{if } r_i(0) \geq 0.5 \\ 1 & \text{if } r_i(0) < 0.5 \end{cases} \tag{4.25}$$

4.1.5 ML Decoding Via Exhaustive Search

Maximum likelihood (ML) decoding is an optimal decoding metric. It is commonly used as a benchmark to express the lowest possible probability of decoder error. ML decoding selects the codeword $\widehat{\mathbf{c}}$ that maximizes the likelihood $\Pr(\mathbf{y}|\mathbf{c})$:

$$\widehat{\mathbf{c}} = \arg\max_{\mathbf{c} \in \mathcal{C}} \Pr(\mathbf{y}|\mathbf{c}) \tag{4.26}$$

One way to perform maximum likelihood decoding is by an exhaustive search — to compare each codeword in the codebook $\mathcal{C}$ to the received sequence $\mathbf{y}$, and select the codeword for which $\Pr(\mathbf{y}|\mathbf{c})$ is greatest. This has high complexity in general, although a few specific codes, notably convolutional codes, have an efficient maximum likelihood decoding algorithm.

Consider the case of transmitting a binary codeword $\mathbf{c}$ over the BI-AWGN channel. Binary code symbols $c \in \{0, 1\}$ are mapped to $x \in \{+1, -1\}$, according to $0 \to +1$ and $1 \to -1$. Convert from x to c using $x = 1 - 2c$. Because the AWGN channel is memoryless:

$$\Pr(\mathbf{y}|\mathbf{c}) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} e^{-\frac{(x_i - y_i)^2}{2\sigma^2}} \tag{4.27}$$

It is convenient to instead maximize $\ln \Pr(\mathbf{y}|\mathbf{c})$:

$$\ln \Pr(\mathbf{y}|\mathbf{c}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - y_i)^2 + n \ln 2\pi\sigma^2 \tag{4.28}$$

Algorithm 4.1 ML Decoding Via Exhaustive Search for BI-AWGN Channel

Require: Received sequence $\mathbf{y}$. Generator matrix $\mathbf{G}$.

Ensure: Estimated codeword $\widehat{\mathbf{c}}$.

$M_{\text{best}} \leftarrow \infty$

Decode:

for $i = 0, 1, \ldots, 2^k - 1$ **do**

$\mathbf{c}_i = \mathbf{uG}$ where $\mathbf{u}$ is binary representation of i

$\mathbf{x} = 1 - 2\mathbf{c}_i$

$M = \sum_{i=1}^{n}(x_i - y_i)^2$

if $M < M_{\text{best}}$ **then**

$\widehat{\mathbf{c}} \leftarrow \mathbf{c}_i$

$M_{\text{best}} \leftarrow M$

end if

end for

In optimizing over $\mathbf{c}$, the independent terms $1/(2\sigma^2)$ and $n \ln 2\pi\sigma^2$ may be ignored. Change the "$-$" maximization problem to a "$+$" minimization problem. Then, the maximum likelihood rule on the BI-AWGN channel is:

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}} ||\mathbf{x} - \mathbf{y}||^2 = \arg\min_{\mathbf{c}\in\mathcal{C}} ||1 - 2\mathbf{c} - \mathbf{y}||^2, \tag{4.29}$$

where

$$||\mathbf{x} - \mathbf{y}||^2 = \sum_{i=1}^{n}(x_i - y_i)^2. \tag{4.30}$$

That is, maximum likelihood decoding for the BI-AWGN channel is minimizing the the Euclidean distance.

Exhaustive search decoding for the BI-AWGN channel is implemented in Algorithm 4.1. This decoding algorithm can be easily applied to decoding on the BSC channel by changing the Euclidean metric $M = \sum_{i=1}^{n}(x_i - y_i)^2$ to the Hamming metric $M = d(\mathbf{c}, \mathbf{y})$.

The algorithm searches over 2^k codewords, and so the complexity is exponential in the number of information symbols k. This is practical when k is small, but as k increases, the computation complexity is prohibitive. Maximum likelihood decoding with lower complexity may be possible for other codes by exploiting their codebook structure, such as convolutional codes.

4.2 Erasure Decoding

Error-correcting codes can be used to correct erasures. For the erasure channel, a transmitted symbol is either received correctly, or erased with a symbol ?. For binary codes, a received 0 or 1 is always correct. The maximum number of erasures that a code can correct can be guaranteed

Proposition 4.1. An error-correcting code $\mathcal{C}$ with minimum distance $d_{\min}$ can correct any pattern of s erasures if:

$$s \leq d_{\min} - 1 \tag{4.31}$$

Let $\mathbf{x}$ be the transmitted codeword and let $\mathbf{y}$ be the received sequence with erasures. If $\mathbf{x}'$ is any other codeword, then $\mathbf{x}$ and $\mathbf{x}'$ differ in at least $d_{\min}$ positions, but the erasures hide at most $d_{\min} - 1$ positions. Thus there is at least one non-erased position where $\mathbf{x}'$ and $\mathbf{x}$ differ. Since $\mathbf{x} \neq \mathbf{x}'$ and $\mathbf{x}'$ is not consistent with $\mathbf{y}$ in at least one position, the decoder can correctly choose $\mathbf{x}$.

This is easy to see with an example. Assume the codebook is $\mathcal{C} = \{0000000, 0110011\}$ with $d_{\min} = 4$. If the received sequence has 3 erasures,

$$\mathbf{y} = [0??00?1]^{\mathrm{t}} \tag{4.32}$$

then decoding is successful – clearly $\widehat{\mathbf{c}} = [0110011]^{\mathrm{t}}$. But if there are 4 erasures $\mathbf{y} = [0??00??]^{\mathrm{t}}$, then it is impossible to distinguish the two codewords.

As suggested by the proof of Proposition 4.1, erasure decoding can be regarded as solving a system of equations. If $\mathbf{y}$ is the received sequence, and an erasure "?" in position i is replaced with an unknown code bit x_i, then a method of decoding is to solve:

$$\mathbf{H}\mathbf{y}^{\mathrm{t}} = \mathbf{0}, \tag{4.33}$$

if possible, as shown in the following example.

Example 4.4. For a code with parity check matrix,

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

suppose that the sequence

$$\mathbf{y} = \begin{bmatrix} ?, 1, 1, 1, ?, ? \end{bmatrix}$$

is received. Perform erasure decoding.

Replace the erasures with variables x_i, so that

$$\mathbf{y} = \begin{bmatrix} x_1, 1, 1, 1, x_5, x_6 \end{bmatrix}$$

and then the system of equations $\mathbf{H}\mathbf{y} = 0$ is:

$$\begin{aligned} x_1 + 1 + 1 &= 0 \\ 1 + x_5 + x_6 &= 0 \\ x_1 + 1 + x_6 &= 0 \\ 1 + 1 + x_5 &= 0 \end{aligned}$$

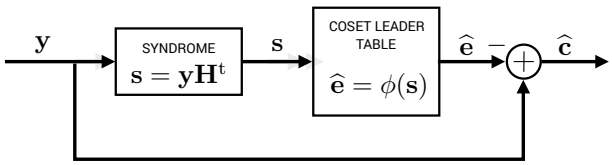


Figure 4.2: Syndrome decoder. The decoder input from the channel is $\mathbf{y}$. The syndrome $\mathbf{s} = \mathbf{y}\mathbf{H}^{\mathrm{t}}$ is computed, the estimated error $\widehat{\mathbf{e}} = \phi(\mathbf{s})$ is found from the coset leader table.

The solution is: $x_1 = 0$ and $x_5 = 0$, and then $x_6 = 1$, and the decoded sequence is:

$$\widehat{\mathbf{c}} = \left[x_1, 1, 1, 1, x_5, x_6\right] = \left[0, 1, 1, 1, 0, 1\right].$$

Note that this code has $d_{\min} = 2$, so it is guaranteed to correct 1 erasure. In this example, it correct a specific erasure pattern of 3 erasures, indicating that some codes may be able to correct certain patterns beyond its guaranteed capability.

SSQ 4.2. *Erasure Decoding* Using the parity check matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \tag{4.34}$$

decode the erasure channel output $\mathbf{y} = [?, ?, 1]$ to a codeword

4.3 Syndrome Decoding

4.3.1 Decoding Cosets and Syndromes

Syndrome decoding is a hard-input method for decoding linear block codes. It works best when $n - k$ is small. For a code $\mathcal{C} \subseteq \mathbb{F}_q^n$, a codeword $\mathbf{c} \in \mathcal{C}$ is transmitted over a channel that adds an error vector $\mathbf{e}$:

$$\mathbf{e} = [e_1\ e_2\ e_3\ \ldots\ e_n], \tag{4.35}$$

to the codeword, with $e_i \in \mathbb{F}$. The channel output is $\mathbf{y} = \mathbf{c} + \mathbf{e}$, and addition is in the field $\mathbb{F}_q^n$. For binary codes in $\mathbb{F}_2$, this corresponds to the binary symmetric channel (BSC).

Recall from Section 2.2 that if $\mathcal{G}$ is a group and $\mathcal{H}$ is a subgroup, then there exists a quotient group $\mathcal{G}/\mathcal{H}$. This partitions $\mathcal{G}$ into non-overlapping subsets called cosets, for example:

$$a + \mathcal{H} = \{a + h \mid h \in \mathcal{H}\} \tag{4.36}$$

is the coset containing a. From each coset, we can select one element to be the coset leader. For linear block codes, the group is the whole space $\mathbb{F}_q^n$ and the subgroup is the code $\mathcal{C}$.

Suppose that $\mathbf{c}$ is transmitted over the channel, noise $\mathbf{e}$ is added, so that the received symbol is $\mathbf{y} = \mathbf{c} + \mathbf{e}$, where addition is in $\mathbb{F}_q$. From this, it can be seen that $\mathcal{C} + \mathbf{e}$ is the coset containing $\mathbf{e}$. The coset containing $\mathbf{e} = \mathbf{0}$ is the code itself. For any other $\mathbf{e} \neq \mathbf{0}$, the coset containing $\mathbf{e}$ does not have a codeword.

The decoder input is $\mathbf{y} = \mathbf{x} + \mathbf{e}$. If the decoder can identify which coset $\mathbf{y}$ belongs to, then it can find a candidate for $\mathbf{e}$. We choose the coset leaders to be low-weight error vectors, since these are most likely. To identify the coset, we can compute a vector called the syndrome.

Definition 4.1. For a code $\mathcal{C}$ with $m \times n$ parity check matrix $\mathbf{H}$, the *syndrome* $\mathbf{s}$ of a sequence $\mathbf{y}$ is:

$$\mathbf{s} = \mathbf{y}\mathbf{H}^{\mathrm{t}}, \tag{4.37}$$

where $\mathbf{s}$ is a $1 \times m$ vector and $\mathbf{y}$ is a $1 \times n$ vector.

The syndrome of a codeword is $\mathbf{0}$. That is, $\mathbf{c} \in \mathcal{C}$ if and only if $\mathbf{c}\mathbf{H}^{\mathrm{t}} = \mathbf{0}$.

Example 4.5. Find the syndrome of $\mathbf{y}_1 = (0, 0, 0, 1, 1)$ and $\mathbf{y}_2 = (1, 0, 1, 1, 0)$ for the parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{4.38}$$

The syndromes are $\mathbf{s}_1 = \mathbf{y}_1\mathbf{H}^{\mathrm{t}} = (0, 1, 1)$ and $\mathbf{s}_2 = \mathbf{y}_2\mathbf{H}^{\mathrm{t}} = (0, 0, 0)$. Here, $\mathbf{y}_2$ is a codeword, but $\mathbf{y}_1$ is not.

Two sequences in the same coset can be described using the same error vector. For two codewords $\mathbf{c}', \mathbf{c}''$, let $\mathbf{e}$ be some fixed error vector, and the received sequences $\mathbf{y}'$ and $\mathbf{y}''$ are:

$$\mathbf{y}' = \mathbf{c}' + \mathbf{e} \text{ and} \tag{4.39}$$

$$\mathbf{y}'' = \mathbf{c}'' + \mathbf{e}. \tag{4.40}$$

Then $\mathbf{y}'$ and $\mathbf{y}''$ are in the same coset, by the definition of a coset. There is a one-to-one correspondence between cosets and syndromes:

Proposition 4.2. Two sequences are in the same coset if and only if they have the same syndrome.

Proof Compute the syndrome $\mathbf{s}'$ of $\mathbf{y}'$ and the syndrome $\mathbf{s}''$ of $\mathbf{y}''$:

$$\mathbf{s}' = \mathbf{y}'\,\mathbf{H}^{\mathrm{t}} = (\mathbf{c}' + \mathbf{e})\mathbf{H}^{\mathrm{t}} = \mathbf{c}'\,\mathbf{H}^{\mathrm{t}} + \mathbf{e}\mathbf{H}^{\mathrm{t}} = \mathbf{e}\mathbf{H}^{\mathrm{t}} \text{ and} \tag{4.41}$$

$$\mathbf{s}'' = \mathbf{y}''\mathbf{H}^{\mathrm{t}} = (\mathbf{c}'' + \mathbf{e})\mathbf{H}^{\mathrm{t}} = \mathbf{c}''\mathbf{H}^{\mathrm{t}} + \mathbf{e}\mathbf{H}^{\mathrm{t}} = \mathbf{e}\mathbf{H}^{\mathrm{t}} \tag{4.42}$$

since $\mathbf{c}\mathbf{H}^{\mathrm{t}} = \mathbf{0}$ for all codewords $\mathbf{c}$. That is, $\mathbf{y}'$ and $\mathbf{y}''$ have the same syndrome and are in the same coset. The syndrome is independent of the codeword. □.

A coset has many members, we will identify the coset by one of its members — in particular, a member with the lowest weight:

Definition 4.2. The one coset word of minimum weight is called the *coset leader*. In the case of a tie, one coset word is arbitrarily designated the coset leader.

The *standard array* for a binary linear code $\mathcal{C}$ is a table with 2^{n-k} rows and 2^k columns such that:

- its 2^n entries are all 2^n vectors in $\mathbb{F}_2^n$,
- each row contains a distinct coset of $\mathcal{C}$,
- the first column is the coset leaders,
- the first row is the codebook $\mathcal{C}$.

Example 4.6. Find the standard array for the binary code $\mathcal{C}$ with generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{4.43}$$

The codebook is $\mathcal{C} = \{00000, 10110, 01110, 11101\}$ and the parity check matrix is:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{4.44}$$

The standard array is given in Fig. 4.3.

The second column shows errors with respect to the all-zeros codeword. These are the correctable errors vectors, which includes all weight-one sequences. All of the sequences in the second row consist of the correctable error vectors. If the sequence (1 0 1 1 1) is received, the nearest codeword is (1 0 1 1 0), corresponding to an error $\mathbf{e} = (0\ 0\ 0\ 0\ 1)$.

Note that in each of the last two rows of the standard array, there are two sequences of weight two. The choice of (00101) and (01100) was made arbitrarily.

4.3.2 Syndrome Decoding

The goal of syndrome decoding is to estimate the error vector. This error vector is then subtracted from the decoder input to obtain the codeword estimate. Since all elements of a coset have the same syndrome, they also have the same error vector. Given the syndrome, the coset leader is interpreted as the *error vector* or *error pattern*, since the coset leader has the lowest weight.

A table $\phi(\mathbf{s}) = \widehat{\mathbf{e}}$ maps the syndrome $\mathbf{s}$ to the coset leader $\widehat{\mathbf{e}}$. The syndrome table should contain one low-weight coset leader for each syndrome. The syndrome table may be found by the following procedure:

syndrome **s**	Standard Array				
0 0 0	0 0 0 0 0	1 0 1 1 0	0 1 0 1 1	1 1 1 0 1	codebook $\mathcal{C}$
0 0 1	0 0 0 0 1	1 0 1 1 1	0 1 0 1 0	1 1 1 0 0	a coset
0 1 0	0 0 0 1 0	1 0 1 0 0	0 1 0 0 1	1 1 1 1 1	
0 1 1	0 1 0 0 0	1 1 1 1 0	0 0 0 1 1	1 0 1 0 1	
1 0 0	0 0 1 0 0	1 0 0 1 0	0 1 1 1 1	1 1 0 0 1	
1 1 0	1 0 0 0 0	0 0 1 1 0	1 1 0 1 1	0 1 1 0 1	
1 0 1	0 0 1 0 1	1 0 0 1 1	0 1 1 1 0	1 1 0 0 0	
1 1 1	0 1 1 0 0	1 1 0 1 0	0 0 1 1 1	1 0 0 0 1	
	coset leaders		decoding region around 01011		

Figure 4.3: Standard array for Example 4.6.

1. Syndrome $\mathbf{s} = \mathbf{0}$ maps to $\widehat{\mathbf{e}} = \mathbf{0}$.
2. For each Hamming weight 1 error vector $\widehat{\mathbf{e}}$, compute the syndrome $\mathbf{s} = \widehat{\mathbf{e}}\mathbf{H}^{\mathrm{t}}$. If this syndrome $\mathbf{s}$ is not yet in ϕ, then add $\mathbf{s}, \widehat{\mathbf{e}}$ to ϕ. Otherwise, do not add this pair to the table.
3. Repeat step 1 for each weight 2 vector.
4. Repeat step 1 for each weight 3 vector, etc. Stop when the syndrome table has all 2^{n-k} entries.

The syndrome table for Example 4.6 is given in Table 4.1. The syndrome table may be generated by finding the syndromes for error patterns of increasing weight. Clearly, the error pattern $\widehat{\mathbf{e}} = \mathbf{0}$ corresponds to $\mathbf{s} = \mathbf{0}$. Then, consider all sequences of weight one, for example $\widehat{\mathbf{e}} = 00001$ corresponds to $\widehat{\mathbf{e}}\mathbf{H}^{\mathrm{t}} = 001$, so $\phi(001) = 000001$. The syndrome table also contains error patterns of weight 2, but a given syndrome may have multiple error patterns of the same weight; in this case, choose one error pattern arbitrarily.

SSQ 4.3. For the (7,4) Hamming code with parity check matrix $\mathbf{H}$:

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

there are 8 possible syndromes. For each syndrome, find the corresponding

Algorithm 4.2 Syndrome Decoding

Require: Received sequence $\mathbf{y}$. Parity-check matrix $\mathbf{H}$ and syndrome table ϕ.

Ensure: Estimated codeword $\widehat{\mathbf{c}}$.

Compute the syndrome $\mathbf{s} = \mathbf{y}\mathbf{H}^{\mathrm{t}}$

Find $\widehat{\mathbf{e}} = \phi(\mathbf{s})$

Output the estimated codeword is $\widehat{\mathbf{c}} = \mathbf{y} - \widehat{\mathbf{e}}$.

Table 4.1: Syndrome table for Example 4.6

syndrome $\mathbf{s}$	coset leader $\widehat{\mathbf{e}} = \phi(\mathbf{s})$
0 0 0	0 0 0 0 0
1 0 0	0 0 1 0 0
0 1 0	0 0 0 1 0
1 1 0	1 0 0 0 0
0 0 1	0 0 0 0 1
1 0 1	0 0 1 0 1
0 1 1	0 1 0 0 0
1 1 1	0 1 1 0 0

error pattern.

To perform syndrome decoding: (1) find the syndrome (2) use ϕ to find the corresponding error vector (3) this error vector is subtracted from the received sequence. Syndrome decoding is described in Algorithm 4.2.

Example 4.7. Continuing Example 4.6, decode the received sequence $\mathbf{y} = (0\ 0\ 1\ 1\ 0)$ using syndrome decoding. The syndrome is $\mathbf{s} = \mathbf{y}\mathbf{H}^t = (1\ 1\ 0)$. Referring to the table, the corresponding error vector is $\widehat{\mathbf{e}} = (1\ 0\ 0\ 0\ 0)$. Then, the estimated codeword is $\widehat{\mathbf{c}} = \mathbf{y} - \widehat{\mathbf{e}} = (1\ 0\ 1\ 1\ 0)$.

4.4 Performance of Error-Correcting Codes

4.4.1 Word error rate and bit error rate

Word-error rate (WER) and bit-error rate (BER) are two measures of a decoders performance. Recall that a codeword $\mathbf{c}$ is transmitted over a channel, the sequence $\mathbf{y}$ is received, and the decoder estimates the transmitted sequence as $\widehat{\mathbf{c}}$. A codeword $\mathbf{c}$ is transmitted, $\mathbf{y}$ is received, and the decoder possibly outputs $\widehat{\mathbf{c}}$. There are three possible decoding results:

Decoder succeeds The output is correct, that is $\widehat{\mathbf{c}} = \mathbf{c}$.

Decoder error The output is incorrect, that is $\widehat{\mathbf{c}} \neq \mathbf{c}$.

Failure to decode The decoder reports that it cannot find a codeword. Certain decoding algorithms recognize that the input sequence cannot be decoded. Denote this as an event E.

The probability of decoding error is an important metric to measure both the goodness of the code $\mathcal{C}$ and its error-correcting algorithm.

The *probability of word error* or word error rate (WER) is the probability that the decoded word $\widehat{\mathbf{c}}$ is not equal to the transmitted word $\mathbf{c}$, that is:

$$\text{WER} = \Pr[\mathbf{c} \neq \widehat{\mathbf{c}}], \tag{4.45}$$

The probability of bit error or *bit error rate* (BER) is the probability that the decoded code bit $\widehat{c}_i$ is not equal to the transmitted code bit c_i:

$$\text{BER} = \Pr[\widehat{c}_i \neq c]. \tag{4.46}$$

When determining WER and BER, the event E is regarded as an error. Usually WER and BER are obtained by Monte Carlo simulations, this is described in a following subsection

4.4.2 Union Bound on Probability of Error

The union bound is technique to upper bound the probability of decoding error for a given error correcting code. Consider decoding of a binary code on the BI-AWGN channel, with the mapping of binary code bits $\{0,1\} \to \{+1,-1\}$. Assume that the codebook $\mathcal{C}$ consists of ± 1 codewords. Recall that the decoder output is $\widehat{\mathbf{c}}$:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{C}} ||\mathbf{x}-\mathbf{y}||^2. \tag{4.47}$$

The space $\mathbb{R}^n$ is partitioned into M *decision regions*, one for each codeword. The decision region $\mathcal{R}_i$ for codeword $\mathbf{x}_i$ is the set of received vectors $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{y}$ is closer to $\mathbf{x}_i$ than any other codeword:

Consider pairwise decision regions $\mathcal{R}_{i,j}$ between $\mathbf{x}_i$ and *one* other codeword $\mathbf{x}_j$:

$$\mathcal{R}_{i,j} = \{y \in \mathbb{R}^n \mid ||\mathbf{y}-\mathbf{x}_i||^2 \le ||\mathbf{y}-\mathbf{x}_j||^2\}$$

That is, codewords other than $\mathbf{x}_i, \mathbf{x}_j$ have been ignored. The decision region $\mathcal{R}_i$ is the intersection of the $M-1$ pairwise decision regions:

$$\mathcal{R}_i = \bigcap_{j=1}^{M-1} \mathcal{R}_{i,j} \tag{4.48}$$

The *pairwise error probability* $\Pr[\mathbf{x}_i \to \mathbf{x}_j]$ is the probability that $\mathbf{x}_i$ was transmitted but $\mathbf{x}_j$ was decoded, assuming that $\mathbf{x}_i, \mathbf{x}_j$ are the only two codewords. The squared distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is $d^2 = 2^2 d_{\mathrm{h}}(\mathbf{c}_i, \mathbf{c}_j)$, where $d_{\mathrm{h}}(\mathbf{c}_i, \mathbf{c}_j)$ is the Hamming distance between two $\{0,1\}$ codewords and 2 accounts for distance between -1 and $+1$. Then, $\Pr[\mathbf{x}_i \to \mathbf{x}_j]$ can be computed exactly because the Gaussian noise is spherically symmetric:

$$\Pr[\mathbf{x}_i \to \mathbf{x}_j] = \frac{1}{2\pi\sigma^2}\int_{d/2}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,dx = \frac{1}{2}\mathrm{erfc}\left(\frac{d/2}{\sqrt{2\sigma^2}}\right) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{d_{\mathrm{h}}(\mathbf{c}_i,\mathbf{c}_j)}{2\sigma^2}}\right)$$

The complementary error function is $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,dt$.

The *union bound* on the probability of decoding error is based on the union bound of events, $\Pr(A \cup B) \le \Pr(A) + \Pr(B)$ for events A and B. The probability of decoder error assuming that $\mathbf{x}_i$ was transmitted $\Pr[\mathsf{E}|\mathbf{x}_i]$ is upper bounded by the sum of the $M-1$ pairwise error probabilities:

$$\Pr[\mathsf{E}|\mathbf{x}_i] \le \sum_{j\neq i} \Pr[\mathbf{x}_i \to \mathbf{x}_j] = \frac{1}{2}\sum_{j\neq i} \mathrm{erfc}\left(\sqrt{\frac{d_{\mathrm{h}}(\mathbf{c}_i,\mathbf{c}_j)}{2\sigma^2}}\right). \tag{4.49}$$

For transmitting a linear block code on the BI-AWGN channel, the probability of error for all codewords is the same. Without loss of generality, assume that the all-zeros codeword was transmitted and find WER $= \Pr[\mathsf{E}|\mathbf{x} = \mathbf{0}]$. Recall the weight distribution for the code is $A(z) = \sum_{w=0}^{n} A_w z^w$, where there are A_w

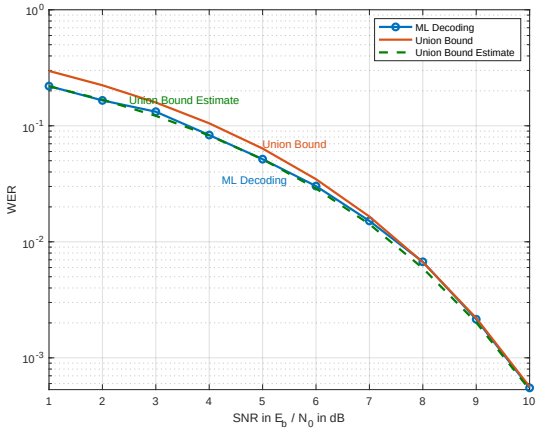


Figure 4.4: Comparison of ML decoding and the union bound for the (5,3) code of Example 4.8.

codewords with weight w. Rather than summing over codewords, sum over the weights and the above equation becomes:

$$\text{WER} \leq \frac{1}{2} \sum_{w=1}^{n} A_w \text{erfc}\left(\sqrt{\frac{w}{2\sigma^2}}\right) \tag{4.50}$$

The *union bound estimate* or *truncated union bound* is based on the idea that the first nonzero term, which has weight $d_{\min}$ will dominate the sum, and the other terms can be removed:

$$\text{WER} \approx \frac{1}{2} A_{d_{\min}} \text{erfc}\left(\sqrt{\frac{d_{\min}}{2\sigma^2}}\right). \tag{4.51}$$

Because we have removed terms from the sum, it is no longer an upper bound but only an approximation.

Example 4.8. Consider the code of Example 4.6 used on the BI-AWGN channel. It has weight enumerator polynomial

$$A(z) = 1 + 2z^3 + z_4.$$

There are 2 codewords of weight 3 and 1 codeword of weight 4. Then the union bound on probability of error is:

$$\text{WER} \leq \text{erfc}\left(\sqrt{\frac{3}{2\sigma^2}}\right) + \frac{1}{2}\text{erfc}\left(\sqrt{\frac{4}{2\sigma^2}}\right) \tag{4.52}$$

and the union bound estimate is:

$$\text{WER} \approx \text{erfc}\left(\sqrt{\frac{3}{2\sigma^2}}\right) \tag{4.53}$$

Fig. 4.4 compares the WER of optimal ML decoding with the union bound. At high SNR, the union bound, the union bound estimate and the ML decoding WER are quite close. The union bound estimate may be lower than the ML decoding WER, since it is not an upper bound. In this case, the union bound estimate gives good agreement even at low SNR, although this may not be true in general.

4.4.3 Evaluating Codes and Decoders Using Monte Carlo Simulations

The performance of an error-correcting code and its decoder is evaluated using WER and BER, as defined in Section 4.4.1. Monte Carlo simulations are most commonly used to evaluate the decoder error rate. The computer simulation models the encoder, the channel and the decoder, that is, the three parts of the communication system shown in Fig. 1.1. For a given channel, the WER and BER depends both on the code and the decoder.

The goal is to express WER and BER as a function of the channel parameter. As the channel gets "better", the WER and BER decrease. Consider the example of the code in Example 4.6, transmitted over a binary symmetric channel (BSC) with error probability p, and is decoded using syndrome decoding. The WER is shown in Fig. 4.5. The "better" channel has smaller p, and the WER for the decoder decreases as p decreases.

A "word error" means even one bit is in error. For example, if

$$\mathbf{u} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{4.54}$$

is the transmitted information and

$$\widehat{\mathbf{u}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{4.55}$$

is the decoded information sequence, then this represents one word error and three bit errors.

4.4.4 Source Code for Syndrome Decoding

The following simple example illustrates the principles of a computer simulation. Source code is provided using Matlab, but the principles apply for any programming language. The following example uses the code in Example 4.6, transmitted over a binary symmetric channel (BSC) with error probability p, and is decoded using syndrome decoding.

The first step is to generate $k = 2$ equally likely bits of information $\mathbf{u}$, from the set $\{0, 1\}$. Matlab has command which does this:

```
1 k = 2;
2 u = randi([0 1],1,k);
```

Encoding may performed using the generator matrix:

```
c = u * G;
```

So that the codeword $\mathbf{c}$ is a 1-by-5 vector.

Channel noise must be simulated. A binary symmetric channel with error probability p can be simulated by generating an error vector $\mathbf{z}$ which has a 1 with probability p. The vector $\mathbf{z}$ is added to $\mathbf{c}$ to obtain the channel output $\mathbf{y}$:

```
p = 0.03;
z = rand(1,n) < p;
y = mod(c + z,2)
```

The decoder is usually the most complicated part of the simulation, but for syndrome decoding is only three lines of source code. The input is the channel sequence $\mathbf{y}$, and the output is the estimated codeword $\widehat{\mathbf{c}}$ or the estimated information $\widehat{\mathbf{u}}$. This is an implementation of the syndrome decoder:

```
s = mod(y * H',2);
ehat = Phi(bi2de(s)+1,:);
chat = mod(y - ehat,2);
```

Finally, the number of bit errors and word errors are counted. The number of bit errors is the number of positions where $\mathbf{c}$ and $\widehat{\mathbf{c}}$ disagree. The number of word errors is either 0 or 1. If the number of bit errors is greater than 0, then a word error occurred:

```
numberOfBitErrors = sum( mod( c - chat, 2) );
numberOfWordErrors = (numberOfBitErrors > 0);
```

Matlab source code for the simulation is shown in Fig. 4.6.

4.4.5 Estimating Error Rates

In order to estimate the WER and BER, Monte Carlo simulations are performed.

That is, the above simulation is repeated for many frames.

After accumulating errors over many frames, the WER can be estimated as:

$$\text{WER} \approx \frac{\text{number of word errors}}{\text{total number of frames}} \tag{4.56}$$

and the BER can be estimated as:

$$\text{BER} \approx \frac{\text{number of bit errors}}{\text{block length} \times \text{total number of frames}} \tag{4.57}$$

More precisely, let N_{word} be the number of word errors and let N be the number of frames. Then,

$$\text{WER} \approx \frac{N_{\text{word}}}{N} \tag{4.58}$$

Let N_{bit} be the number of bit errors. Then,

$$\text{BER} \approx \frac{N_{\text{bit}}}{n \cdot N} \tag{4.59}$$

where n is the code block length.

Each frame is an experiment which either "passes" (decoding succeeds) or "fails" (decoding fails); the probability of failure is WER. Each frame is an independent experiment, and we repeat the experiment many times to obtain an estimate of WER.

How many frames should be simulated? That is, how large should N be? If N is small, we can get a result with a small amount of computer time, but the WER estimate will not be very reliable. In the following example, $N = 7000$ is used. For each of 5 simulations, the computer time is low, but the variation in WER is large, from 0.0001428 to 0.001142.

```
>> syndromeDecodingExample(0.008,7000)
Nerr = 6, WER = 0.0008571, computer time = 0.11 seconds
>> syndromeDecodingExample(0.008,7000)
Nerr = 3, WER = 0.0004285, computer time = 0.11 seconds
>> syndromeDecodingExample(0.008,7000)
Nerr = 2, WER = 0.0002857, computer time = 0.12 seconds
>> syndromeDecodingExample(0.008,7000)
Nerr = 8, WER = 0.001142, computer time = 0.11 seconds
>> syndromeDecodingExample(0.008,7000)
Nerr = 1, WER = 0.0001428, computer time = 0.12 seconds
```

Large variation leads to inconsistent results, and it is better to choose N large enough so that we can have a reliable WER estimate after one simulation. In the following example, N increases to 700000. For each of 5 simulations, the WER variation is smaller, from 0.0004685 to 0.0005328, which indicates a better WER estimate. However, more computer time is required:

```
>> syndromeDecodingExample(0.008,700000)
Nerr = 373, WER = 0.0005328, computer time = 9.57 seconds
>> syndromeDecodingExample(0.008,700000)
Nerr = 337, WER = 0.0004814, computer time = 9.5 seconds
>> syndromeDecodingExample(0.008,700000)
Nerr = 334, WER = 0.0004771, computer time = 9.48 seconds
>> syndromeDecodingExample(0.008,700000)
Nerr = 367, WER = 0.0005242, computer time = 9.45 seconds
>> syndromeDecodingExample(0.008,700000)
Nerr = 361, WER = 0.0005157, computer time = 9.43 seconds
>> syndromeDecodingExample(0.008,700000)
Nerr = 328, WER = 0.0004685, computer time = 9.43 seconds
```

In summary, the value of N should be chosen large enough to obtain a WER with high reliability, but low enough to avoid unnecessary computer time.

Pro Tip When running a simulation, good channels with high SNR require more frames for accurate WER and BER estimates. This is in comparison to bad channels with a low SNR, which require fewer frames. A good approach is to stop the simulation when the number of word errors N_{word} reaches a fixed value, instead of stopping when the number of frames N reaches a fixed value. For example, repeat the simulation until 100 word errors have been accumulated, instead of repeating $100,000$ simulation. This approach gives good reliability for a fixed amount of computer time, whether the channel is good or bad. To implement this in Fig. 4.6, change the while statement to `while numberOfWordErrors < Nstop`. In practice, simulating 5 or 10 frame errors can be used for preliminary results, but publication-quality data should use 50 to 100 word errors for high reliability.

Using N_{word} as the stopping condition is to preferred to using N_{bit} because word errors are statistically independent events, but bit errors are not.

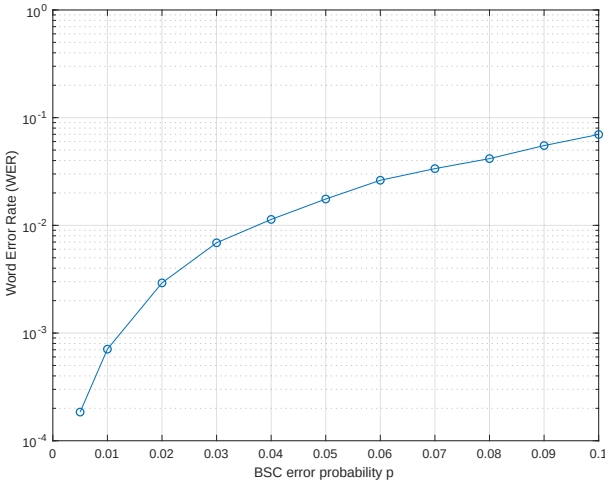


Figure 4.5: Word error rate (WER) for syndrome decoding of the code in Example 4.6 on the BSC with error probability p.

4.4.6 Signal-to-Noise Ratio for Binary-Input AWGN Channels

Consider specifically transmitting a binary codeword $\mathbf{c}$ over the AWGN channel. The binary-input AWGN channel model has inputs $c_1, c_2, \ldots, c_n \in \{0, 1\}$ are mapped to $x \in \{+1, -1\}$, according to $0 \to +1$ and $1 \to -1$. At the transmitter side, convert from x to c using $x = 1 - 2c$.

Consider two definitions of signal-to-noise ratio:

- The energy per symbol is $E_s = 1$, since $+1$ and -1 both have power 1. Thus:
$$\frac{E_s}{N_0} = \frac{1}{2\sigma^2} \tag{4.60}$$
- There are n code symbols but only k information symbols for a rate $R = k/n$ code. If we are interested in the power per information symbol E_b, then we write $E_b = E_s/R$. Thus:
$$\frac{E_b}{N_0} = \frac{1}{2R\sigma^2} \tag{4.61}$$
E_b is the average power of transmitting one information symbol, which is a reasonable goal.

E_b/N_0 is usually expressed in dB, that is:
$$\frac{E_b}{N_0}(\text{dB}) = 10 \log_{10} \frac{E_b}{N_0} \tag{4.62}$$

When you write a simulation, you typically fix E_b/N_0 (dB) for fixed values, for example 1 dB, 2 dB, 3 dB. Then, find σ^2 as:

```
EbNo = 10^(EbNodB / 10)
var = 1 / (2 * EbNo*R)
z = randn(1,n) * sqrt(var)
```

so that `z` is AWGN noise, and the channel output is:

```
y = x + z
```
(4.63)

as your channel output. If you are working with E_s/N_0, then replace `EbNo` with `EsNo` and remove `R`.

4.5 Exercises

4.1 Consider the code with parity-check matrix
$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \tag{4.64}$$
Perform erasure decoding with the received sequences:

```
function syndromeDecodingExample(p,Nstop)

G = [...                    %Generator matrix
     1 0 1 1 0 ; ...
     0 1 0 1 1];
H = [...                    %Parity-check matrix
     1 0 1 0 0 ; ...
     1 1 0 1 0 ; ...
     0 1 0 0 1];
[k,n] = size(G);

Phi = [ ...                 %Syndrome decoder table
     0 0 0 0 0 ; ...
     0 0 1 0 0 ; ...
     0 0 0 1 0 ; ...
     1 0 0 0 0 ; ...
     0 0 0 0 1 ; ...
     0 0 1 0 1 ; ...
     0 1 0 0 0 ; ...
     0 1 1 0 0 ];

numberOfBitErrors  = 0;
numberOfWordErrors = 0;
Nframes = 0;

while Nframes < Nstop
    %encoder
    u = randi([0 1],1,k);
    c = mod(u * G,2);

    %channel
    e = rand(1,n) < p;
    y = mod(c+e,2);

    %syndrome decoding
    s = mod(y * H',2);
    ehat = Phi(bi2de(s)+1,:);  %binary-to-decimal
        conversion
    chat = mod(y - ehat,2);

    nbe                = sum( mod( c - chat, 2) );
    numberOfBitErrors  = numberOfBitErrors + nbe;
    numberOfWordErrors = numberOfWordErrors + (nbe > 0);
    Nframes            = Nframes + 1;
end
WER = numberOfWordErrors / Nframes;
BER = numberOfBitErrors / (n * Nframes);
fprintf('Nerr = %d, WER = %G, BER = %G\n',
    numberOfWordErrors,WER,BER);
save decoding_data  %change filename to avoid overwrite
                    %another .m file loads and plots
```

Figure 4.6: Matlab source code implementing simulation using syndrome decoding.

(a) $\mathbf{y}_1 = [0, 0, 1, ?, ?, ?]$
(b) $\mathbf{y}_2 = [?, ?, ?, 1, 0, 1]$
(c) $\mathbf{y}_3 = [?, ?, ?, 0, 1, ?]$

For each one, give the estimated codeword $\widehat{\mathbf{c}}$, or "failure to decode"

4.2 By computer simulation, find the probability of word error of the $(7, 4)$ Hamming code with:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.65}$$

on the binary symmetric channel with probability of error p, using *syndrome decoding*. Find the generator matrix $\mathbf{G}$. In your simulation, randomly generate data $\mathbf{u}$, and find the corresponding codeword $\mathbf{c} = \mathbf{uG}$. Your syndrome decoder should output an estimated codeword $\widehat{\mathbf{c}}$. Use your simulation to estimate the probability of word error rate WER, for $p = 0.1, 0.025, 0.0025$

4.3 By computer simulation, find the WER and BER of the (7,4) Hamming code given in Question 4.2 on the binary-input AWGN channel, using *maximum likelihood decoding*. Estimate the BER for each $E_s/N_0 = 1$ dB, 2 dB, $\ldots, 9$ dB. Make a plot of your results, with the WER and BER on the vertical axis with a log scale, and the E_s/N_0 in dB on the horizontal axis. On the same graph, plot the bit-error rate for uncoded channel, found in Question 1.5.

4.4 Find the union bound and union bound estimate for the (7,4) Hamming code. Plot the WER on a log scale vs. E_b/N_0 in dB for the BI-AWGN channel.

Chapter 5

Low-Density Parity-Check Codes

Low-density parity-check codes have a sparse parity-check matrix. This enables efficient decoding algorithms which are proportional to the block length. Under certain conditions, the minimum distance increases with the block length. These two properties make LDPC codes very attractive.

5.1 Definition and Graphical Representation

5.1.1 Regular LDPC Codes

A low-density parity-check code is a code whose $m \times n$ parity check matrix $\mathbf{H}$ has a small number of ones. We also say the matrix is sparse or low-density.

Definition 5.1. For a *regular LDPC code*, the parity-check matrix $\mathbf{H}$ has d_{v} ones per column and d_{c} ones per row. We also say the column weight is d_{v} and the row weight is d_{c}.

Example 5.1. The following matrix has $d_{\mathsf{v}} = 2$ and $d_{\mathsf{c}} = 4$, and is a parity-check matrix for a block length $n = 8$ code:

$$\begin{bmatrix} 1 & \mathbf{1} & \mathbf{1} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{5.1}$$

The above matrix does not appear sparse, it has equal numbers of zeros and ones. But as n increases, the sparse structure becomes more clear. The following matrix also has $d_{\mathsf{v}} = 2, d_{\mathsf{c}} = 4$, and is a parity-check matrix for block length

$n = 20$ code:

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.2}$$

This matrix can be more clearly seen to be sparse.

For LDPC codes, a "small number of ones" means the number of ones in $\mathbf{H}$ grows linearly with the block length n. That is, if $\mathbf{H}$ has n columns, the total number of ones in $\mathbf{H}$ is nd_{v} ones. If $\mathbf{H}$ has m columns, then the total number of ones is md_{c}. So, while the number of ones grows linearly in n, the total number of elements grows quadratically in n. This definition can be expanded to nonbinary LDPC codes over a q-ary field $\mathbb{F}_q$ by having a small number of nonzero elements in the parity check matrix; however, this chapter concentrates on binary LDPC code.

The dimension and rate of an LDPC code depends on the rank of $\mathbf{H}$. For a general $m \times n$ matrix $\mathbf{H}$ with $m \leq n$, the rank of $\mathbf{H}$ is less than or equal to m:

$$\operatorname{rank}(\mathbf{H}) \leq m \tag{5.3}$$

For regular LDPC code matrix, there may be linearly dependent rows in $\mathbf{H}$, so the rank of $\mathbf{H}$ is strictly less than m. This means that the dimension of the code is $k = n - \operatorname{rank}(\mathbf{H})$ (not $n - m$) and the code rate is $R = 1 - \frac{\operatorname{rank}(\mathbf{H})}{n}$. The design rate is defined using the size of the parity check matrix, without regard of the rank.

Definition 5.2. The *design rate* of a regular LDPC code with an $m \times n$ parity check matrix is:

$$R_{\mathrm{d}} = 1 - \frac{m}{n}. \tag{5.4}$$

The design rate satisfies $R_{\mathrm{d}} \leq R$. The design rate can be related to d_{v} and d_{c} because there are two ways to compute the number of ones in $\mathbf{H}$, that is $d_{\mathsf{c}}m = d_{\mathsf{v}}n$ and $R_{\mathrm{d}} = 1 - \frac{d_{\mathsf{v}}}{d_{\mathsf{c}}}$.

Example 5.2. In Example 5.1, the check matrix in (5.1) is 4×8 so its design rate is $R_{\mathrm{d}} = \frac{1}{2}$. However, $\mathbf{H}$ is rank 3 with one linearly dependent row, so the code dimension is $8 - 3 = 5$ and the code rate is $R = \frac{5}{8}$.

Similarly, the check matrix in (5.2) has 10 rows but rank 9. The design rate is $R_{\mathrm{d}} = \frac{1}{2}$ and the code rate is $R = \frac{11}{20}$.

SSQ 5.1. LDPC Code Design Rate Given a parity-check matrix $\mathbf{H}$, find the design rate.

5.1.2 Tanner Graph

LDPC codes have a graph representation, which is useful when discussing LDPC codes. A simple undirected graph G consists of a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$, which connect the vertices.

Definition 5.3. A simple undirected graph $G = (\mathcal{V}, \mathcal{E})$ is called a *bipartite graph* if there exists a partition of $\mathcal{V}$ so that both $\mathcal{V}_1$ and $\mathcal{V}_2$ are independent sets. Write $G = (\mathcal{V}_1 + \mathcal{V}_2, \mathcal{E})$ to denote a bipartite graph with partitions $\mathcal{V}_1$ and $\mathcal{V}_2$.

A *Tanner graph* is a bipartite graph corresponding to the $m \times n$ parity-check matrix $\mathbf{H}$. The Tanner graph has: n variable nodes (or bit nodes), represented by circles, m check nodes, represented by squares. There is an edge between variable node i and check node j if there is a one in row i and column j of $\mathbf{H}$. The Tanner graph corresponding to the check matrix in Example 5.1 is shown in Fig. 5.1.

Recall from Subsection 3.2.3 that one row of $\mathbf{H}$ is a parity check $\mathbf{h}$, which corresponds to a parity-check equation $\mathbf{hc}^{\mathrm{t}} = 0$. In the Tanner graph, the row, and thus the parity check constraint, is represented by a check node. For example, the first square in Fig. 5.1 is connected to x_1, x_2, x_3 and x_4, which indicates $x_1 + x_2 + x_3 + x_4 = 0$.

As can be seen from the Tanner graph, each variable node i is connected to several check nodes. Let $\mathcal{M}_i$ denote the set of check nodes that variable node i is connected to. Likewise, each check node j is connected to several variable nodes. Let $\mathcal{N}_j$ denote the set of variable nodes that check node j is connected to. This notation will be useful later on.

Example 5.3. The Tanner graph corresponding to Example 5.1 is shown in Fig. 5.1. In addition, the variable node and check node connectivity sets are given by:

$$\begin{array}{lll} \mathcal{N}_1 = \{1,2,3,4\} & \mathcal{M}_1 = \{1,3\} & \mathcal{M}_5 = \{2,4\} \\ \mathcal{N}_2 = \{5,6,7,8\} & \mathcal{M}_2 = \{1,4\} & \mathcal{M}_6 = \{2,3\} \\ \mathcal{N}_3 = \{1,4,6,8\} & \mathcal{M}_3 = \{1,4\} & \mathcal{M}_7 = \{2,4\} \\ \mathcal{N}_4 = \{2,3,5,7\} & \mathcal{M}_4 = \{1,3\} & \mathcal{M}_8 = \{2,3\} \end{array}$$

If a row of $\mathbf{H}$ has weight d, then the corresponding check node has degree d. Similarly, if a column of $\mathbf{H}$ has weight d, then the corresponding variable node has degree d. Irregular LDPC codes, where each row (and each column) may have varying weights, are described in Section 6.2.

Definition 5.4. A *cycle* of length L in a bipartite graph is a path of L edges which closes back on itself

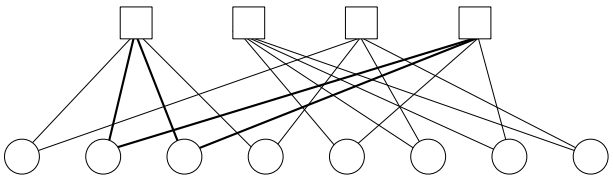


Figure 5.1: Tanner graph corresponding to the first matrix in Example 5.1.

The cycle length L in a bi-partite graph is even. The way we have defined a Tanner graph above, $L = 2$ cycles are not possible, so possible cycle lengths are $4, 6, 8, \ldots$.

Definition 5.5. The *girth* of a bipartite graph is the minimum cycle length of the graph.

Short cycles in a graph can reduce the decoding performance. The length of the shortest cycle to be high, so high girth is desirable. Further, the number of cycles with shortest girth should be small.

5.2 Message-Passing Algorithms

5.2.1 Motivating Example for Message Passing

"Message passing decoding" is a broad class of algorithms on a graph, which includes LDPC decoding algorithms. Before describing message-passing decoding of LDPC codes, we describe another problem which can be solved by message-passing method. Also, this example is used to introduce an important concept, the sum-product update rule.

> Some soldiers are standing in a line. The Commander wants to count them. Calling out the name of each solider is too difficult. And, because there is fog, the commander cannot see very far. How to count the soldiers?

In this problem, each soldier can only communicate integers with his neighbors. Soldiers standing in a line is shown in Fig. 5.2-(a). The following message-passing algorithm allows the commander to obtain the count:

1. The soldiers at each end send a "1" message to their neighbors.
2. A soldier who receives a message, adds one to it, and sends it to his neighbor. If you get a message from the left, send it to the right
3. The commander receives messages from both the left and right, adds one for himself, to obtain the total number of soldier

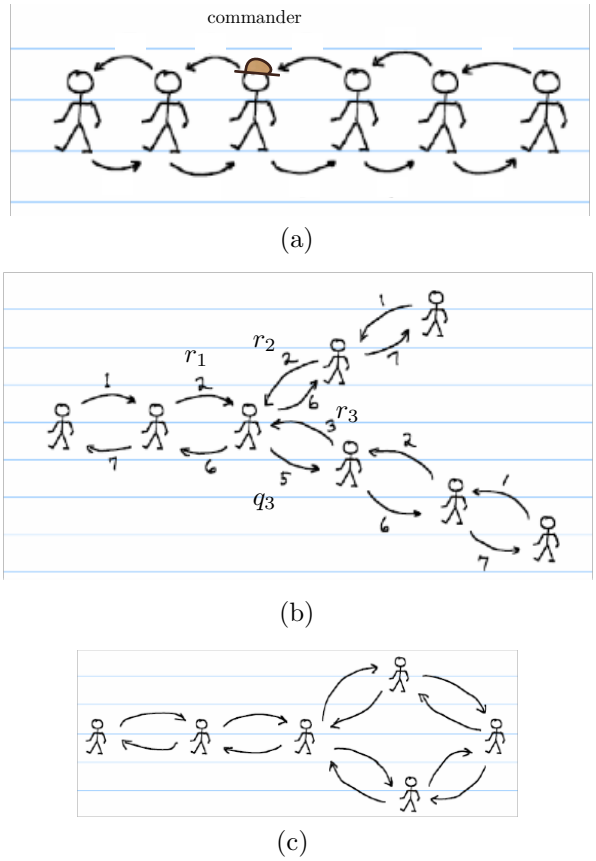


Figure 5.2: (a) Counting soldiers in a line. (b) Counting soldiers in a "Y." (c) There is no easy message-passing approach to counting soldiers in a loop. Figures due to William Ryan.

Not only the commander, but all soldiers can find the total number.

The algorithm works even if the soldiers are standing in a "Y" as shown in Fig. 5.2-(b). The soldier with three neighbors receives two messages $r_1 = 2$ and $r_2 = 2$. This soldier then sends a message $q_3 = r_1 + r_2 = 4$ to the neighbor. Note however, that even though this soldier receives a message $r_3 = 3$, it is not used in computing the outgoing message r_3 on the same edge. This is an example of the sum-product update rule:

Definition 5.6. *Sum-Product Update Rule* For any node t, the outgoing message on edge e depends on the incoming messages on all edges connected to t except e.

However, for soldiers in a loop as shown in Fig. 5.2-(c), there is no simple message-passing solution.

5.2.2 Message-Passing Decoding of LDPC Codes

The following steps express the general idea of LDPC decoding by applying the ideas of message-passing to the Tanner graph.

1. *Initalize* At each variable node, the message from the channel is sent to all connected check nodes
2. *Check Node Function* At each check node, compute each outgoing message using the other incoming messages. These messages are sent to the variable nodes.
3. *Variable Node Function* At each variable node, compute each outgoing message using the other incoming messages. These messages are sent to the check nodes.
4. *Hard Decisions* At each variable node i, make a hard estimate $\widehat{c}_i \in \{0, 1\}$
5. *Check for convergence*
   - If $\widehat{\mathbf{c}}\mathbf{H}^{\mathrm{t}} = 0$, then the decoder has found a valid codeword; output $\widehat{\mathbf{c}}$ and stop.
   - If $\widehat{\mathbf{c}}\mathbf{H}^{\mathrm{t}} \neq 0$ and maximum number of iterations reached, then output $\widehat{\mathbf{c}}$ and stop.
   - Otherwise, go to Step 2.

   There are several variations of LDPC decoding algorithms that follow these steps.

5.2.3 Erasure Decoding of LDPC Codes

LDPC decoding on the erasure channel is particularly simple, and is an example of message-passing algorithm. A codeword $\mathbf{c}$ from an LDPC code $\mathcal{C}$ with parity-check matrix $\mathbf{H}$ is transmitted over the erasure channel. Some of the 0's and 1's are replaced with an erasure symbol ?. Decoding is performed by applying message passing to the Tanner graph. The messages being passed are from the set $\{0, ?, 1\}$.

Check node function The check node finds each outgoing message as follows. If none of the incoming messages is ?, then the outgoing message satisfies the parity check equation. But, if any incoming message is ? then the outgoing message is ?. This follows the sum-product update rule.

For example, a degree 4 check node has $x_1 + x_2 + x_3 + x_4 = 0$. If the inputs are $(q_1, q_2, q_3) = (1, 0, 1)$ then the outgoing message is $r_4 = 0$. But if instead $(q_1, q_2, q_3) = (1, 0, ?)$, then the outgoing message is $r_4 =?$. Another example is shown in Fig. 5.3, where the inputs are $(q_1, q_2, q_3, q_4) = (0, ?, 0, 1)$ and the outputs are $(?, 1, ?, ?)$.

Variable node function If at least one of the incoming messages is a 0 or a 1, then the outgoing message will be that message. But, if all of the incoming messages is ?, then the outgoing message is ?.

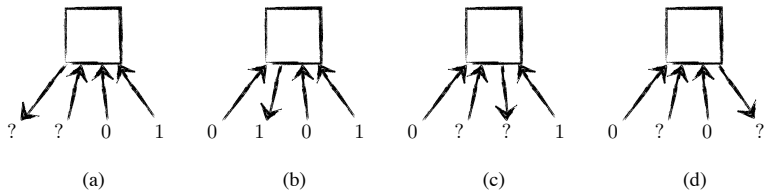


Figure 5.3: Decoding at a degree-4 check node with inputs $(q_1, q_2, q_3, q_4) = (0, ?, 0, 1)$. To compute an output r_i, use all the other inputs, and ignore the input q_i. The output is $(r_1, r_2, r_3, r_4) = (?, 1, ?, ?)$.

For example, a degree 4 variable node, if the input is $(y, r_1, r_2, r_3) = (?, ?, 0, ?)$ then the outgoing message is $q_4 = 0$. But if the input is $(y, r_1, r_2, r_3) = (?, ?, ?, ?)$ then the outgoing message is $q_4 =?$. Another example is shown in Fig. 5.4, where the inputs are $(0, ?, 0, 1)$ and the outputs are $(?, 1, ?, ?)$. Note that specifically for the erasure channel, it will never occur that two incoming messages will be *both* a 0 and a 1, on two incoming messages at one variable node.

Hard decision The variable node makes a hard decision by using all messages, including the channel message. For example, if the input at the variable node is $(y, r_1, r_2, r_3) = (0, ?, ?, ?)$ then the hard decision output is $\widehat{x} = 0$. But, if $(r_1, r_2, r_3, r_4) = (?, ?, ?, ?)$ then $\widehat{x} =?$.

SSQ 5.2. *SSQ: LDPC Erasure Decoding Operations* The input to the LDPC check node is

$$(q_1, q_2, q_3, q_4) = (1, 1, 1, ?)$$

What is the check node output?

Example 5.4. A codeword from a code with parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{5.5}$$

is transmitted over the erasure channel and the received sequence is:

$$\mathbf{y} = (?, 1, 1, 1, ?, ?) \tag{5.6}$$

The decoding process is illustrated in Fig. 5.5, and the final decoding result is:

$$\widehat{\mathbf{c}} = (0, 1, 1, 1, 0, 1)$$

Message passing decoding does not always converge to a solution, even when a solution exists. Because there are loops in the graph, these loops may prevent the iterative approach from finding a solution. While code on the erasure channel decoding can be performed by solving a system of equations as in Section (4.2), the algorithm presented here is both more efficient and demonstrates important points of message passing decoding of LDPC codes.

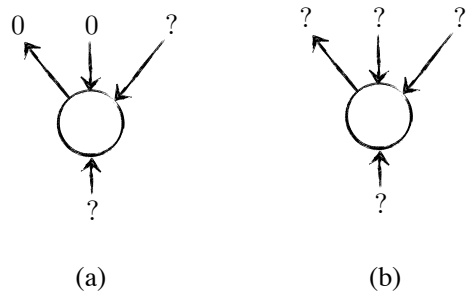


Figure 5.4: Decoding at a degree-3 variable node (a) With inputs $(r_2, r_3) = (0, ?)$ and $y =?$, the output is $q_1 = 0$. (b) With inputs $(r_2, r_3) = (?, ?)$ and $y =?$, the output is $q_1 =?$.

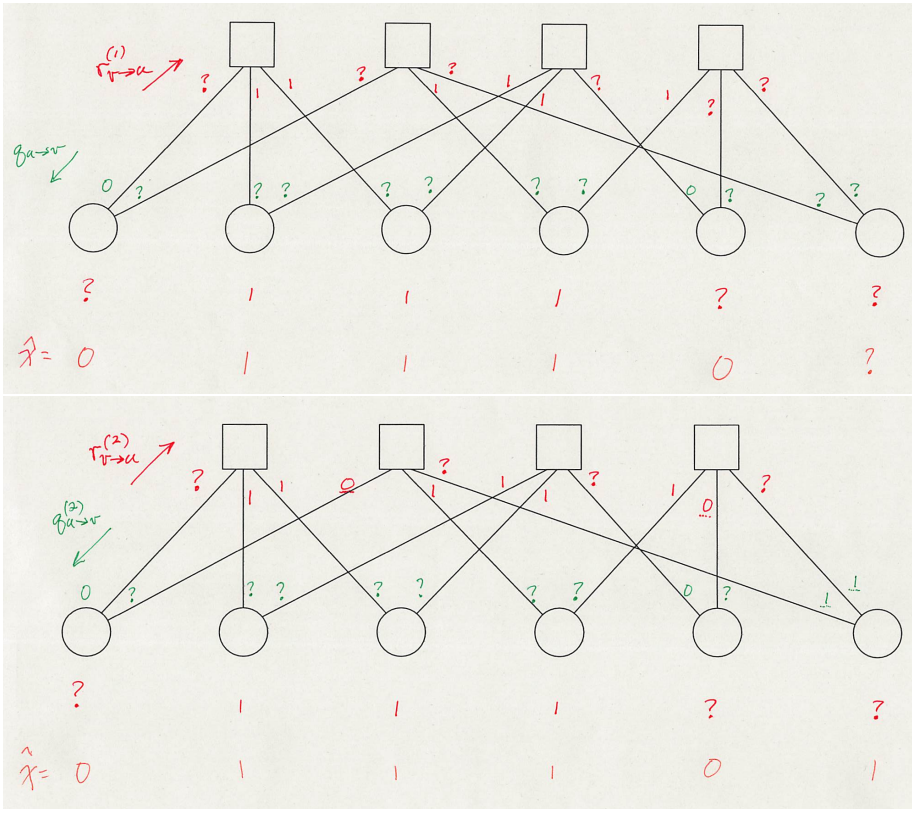


Figure 5.5: Example of decoding an LDPC code on the erasure channel.

5.3 Sum-Product Decoding

Sum-product is a soft-input decoding algorithm for LDPC codes. While not optimal, it performs quite well in practice.

5.3.1 Algorithm

As with erasure decoding, sum-product decoding is message passing on a graph. The decoder has messages, which are passed along edges on graph, and the nodes of the graph represent decoding functions. These are iterative algorithms, that proceed as messages are passed between the nodes. Whereas in erasure decoding the messages were discrete $\{0, ?, 1\}$, in sum-product decoding the messages are real numbers from the log domain. In addition, the sum-product algorithm's check node function and variable node function differ from that of erasure decoding.

To express sum-product decoding, the following definitions are used:

- The n variable nodes are indexed $i \in \{1, 2, \ldots, n\}$,
- The m check nodes are indexed $j \in \{1, 2, \ldots, m\}$,
- The input at variable node i is a message L_i calculated from the channel value y_i,
- The message from variable node i to check node j is denoted $Q_{i\to j}$,
- The message from check node j to variable node i is denoted $R_{j\to i}$.

Also, recall that $\mathcal{M}_i$ denotes the set of check nodes that variable node i is connected to, and $\mathcal{N}_j$ denotes the set of variable nodes that check node j is connected to.

The decoder messages L_i, $Q_{i\to j}$ and $R_{j\to i}$ are in the log domain and have the form:

$$\log \frac{\Pr[c_i = 0]}{\Pr[c_i = 1]} \tag{5.7}$$

It is also possible to implement the decoding algorithm in the probability domain of P0 or P1, but log domain messages have several advantages.

Initialization. Initialization has two parts. (1) Given a channel output $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, the input to the sum-product decoder is the log-likelihood ratio:

$$L_i = \log \frac{\Pr[y_i|c_i = 0]}{\Pr[y_i|c_i = 1]} = \log \frac{\Pr[c_i = 0|y_i]}{\Pr[c_i = 1|y_i]}, \tag{5.8}$$

which follows by Bayes rule and $\Pr[c_i = 1] = \Pr[c_i = 0] = \frac{1}{2}$. Given the channel value, y_i the message L_i is computed; for the BSC channel see (4.10) and for the AWGN channel see (4.13).

Min-Sum Decoding. Min-sum decoding replaces the check node function (5.11) with the following:

$$R_{j\to i} = \left(\prod_{e\in\mathcal{N}_j\setminus i} \text{sign}(Q_{e\to j})\right)\left(\min_{e\in\mathcal{N}_j\setminus i} |Q_{e\to j}|\right) \tag{5.10}$$

The min operation has lower computational complexity than tanh. Complexity can be further reduced because only the minimum and the second-minimum, over all incoming edges, is needed.

While the min-sum decoder error rates are slightly higher than those for sum-product decoding, the decoding complexity is much lower than that of sum-product decoding, so that min-sum decoding is widely used in practical LDPC decoders, particularly in hardware implementations.

(2) To initialize the decoder messages, the channel message at node i is passed to all the connected check nodes:

$$Q_{i\to j} = L_i \text{ for all } j \in \mathcal{M}_i, \tag{5.9}$$

for $i = 1, 2, \ldots, n$.

Check node function Check node j is connected to d variable nodes. The set of connected variable nodes is $\mathcal{N}_j$ and the degree of this node is $d = |\mathcal{N}_j|$. The check node receives d incoming messages and produces d outgoing messages. Check node j performs the following computations:

$$R_{j\to i} = 2\tanh^{-1}\left(\prod_{e\in\mathcal{N}_j\setminus i} \tanh\left(\frac{Q_{e\to j}}{2}\right)\right) \text{ for all } i \in \mathcal{N}_j \tag{5.11}$$

All check nodes $j \in \{1, 2, \ldots, m\}$ perform this operation. The check node is a single parity check code and (5.11) is the log-domain implementation of APP decoding of the single-parity check code, as given in Subsection 4.1.4, including application of the sum-product update rule.

Variable node function Variable node i is connected to d check nodes. The set of connected check nodes $\mathcal{M}_i$ and the degree of this node is $d = |\mathcal{M}_i|$. The variable node receives $d+1$ incoming messages: d from the check nodes and one from the channel. It produces d outgoing messages. Variable node i performs the following computations:

$$Q_{i\to j} = L_i + \sum_{e\in\mathcal{M}_i\setminus j} R_{e\to i} \text{ for all } j \in \mathcal{M}_i \tag{5.12}$$

All variable nodes $i \in \{1, 2, \ldots, n\}$ perform this operation. The variable node can be seen as a repeat code — the variable node represents one bit, and we have various noisy estimates of this bit, similar to decoding of the repeat code. Eq. (5.12) is the log-domain implementation of APP decoding of the repeat code given in Subsection 4.1.3, including application of the sum-product update rule.

Hard Decision and Check for Convergence At variable node i, a hard decision $\widehat{c}_i$ is made using all available inputs:

$$t_i = L_i + \sum_{e \in \mathcal{M}_i} R_{e \to i} \tag{5.13}$$

$$\widehat{c}_i = \begin{cases} 0 & t_i \geq 0 \\ 1 & t_i < 0 \end{cases} \tag{5.14}$$

which is similar to (5.12) but no edge is omitted. It is also similar to the MAP decision for the repeat code. All variable nodes $i \in \{1, 2, \ldots, n\}$ perform this operation.

The hard decision vector $\widehat{\mathbf{c}} = [\widehat{c}_1, \widehat{c}_2, \ldots, \widehat{c}_n]$ is a candidate codeword. If $\widehat{\mathbf{c}}\mathbf{H}^{\mathrm{t}} = \mathbf{0}$ then output $\widehat{\mathbf{c}}$ as a valid codeword and decoding stops. If $\widehat{\mathbf{c}}\mathbf{H}^{\mathrm{t}} \neq \mathbf{0}$, then decoding repeats from the check node step. If no convergence is detected after $\ell_{\max}$ iterations, then output the most recent $\widehat{\mathbf{c}}$; it is also possible for the decoder to declare non-convergence after $\ell_{\max}$ unsuccessful iterations.

Sum-product decoding is given in Algorithm 5.1 on page 98. Matlab source code is in Fig. 5.7 on page 97.

5.3.2 Example

Consider the code with parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{5.15}$$

A codeword $\mathbf{c}$ from this code was transmitted over the binary symmetric channel with $p = 0.1751$ and the sequence:

$$\mathbf{y} = (0, 1, 0, 1, 0, 1) \tag{5.16}$$

was received. Perform sum-product decoding to find the estimated codeword $\widehat{\mathbf{c}}$.

Initialization (1) For the BSC in general:

$$\Pr[\mathsf{Y} = 1 | \mathsf{X} = 0] = \Pr[\mathsf{Y} = 0 | \mathsf{X} = 1] = p \tag{5.17}$$

$$\Pr[\mathsf{Y} = 0 | \mathsf{X} = 0] = \Pr[\mathsf{Y} = 1 | \mathsf{X} = 1] = 1 - p, \tag{5.18}$$

so that:

$$L = \begin{cases} \log \frac{1-p}{p} & \text{if } y = 0 \\ \log \frac{p}{1-p} & \text{if } y = 1 \end{cases} \tag{5.19}$$

Specifically for $p = 0.1751$ and $\mathbf{y} = (0, 1, 0, 1, 0, 1)$ the channel messages L_i are:

$$(L_1, L_2, L_3, L_4, L_5, L_6) = (1.55, -1.55, 1.55, -1.55, 1.55, -1.55) \tag{5.20}$$

(2) These messages are passed on the variable-to-check edges:

$$\begin{aligned} Q_{1\to1} &= 1.55 & Q_{1\to2} &= 1.55 \\ Q_{2\to1} &= -1.55 & Q_{2\to3} &= -1.55 \\ Q_{3\to1} &= 1.55 & Q_{3\to3} &= 1.55 \\ Q_{4\to2} &= -1.55 & Q_{4\to4} &= -1.55 \\ Q_{5\to2} &= 1.55 & Q_{5\to4} &= 1.55 \\ Q_{6\to3} &= -1.55 & Q_{6\to4} &= -1.55 \end{aligned}$$

Check node function Consider the operation at check node 1, which has three incoming messages $Q_{1\to1}$, $Q_{2\to1}$, $Q_{3\to1}$. The check node computes three outputs using (5.11):

$$\begin{aligned} R_{1\to1} &= 2\tanh^{-1}\left(\tanh\left(\frac{Q_{2\to1}}{2}\right)\cdot\tanh\left(\frac{Q_{3\to1}}{2}\right)\right) \\ &= 2\tanh^{-1}\left(\tanh\left(\frac{-1.55}{2}\right)\cdot\tanh\left(\frac{1.55}{2}\right)\right) \\ &= -0.9 \\ R_{1\to2} &= 2\tanh^{-1}\left(\tanh\left(\frac{Q_{1\to1}}{2}\right)\cdot\tanh\left(\frac{Q_{3\to1}}{2}\right)\right) \\ &= 2\tanh^{-1}\left(\tanh\left(\frac{1.55}{2}\right)\cdot\tanh\left(\frac{1.55}{2}\right)\right) \\ &= 0.9 \\ R_{1\to3} &= 2\tanh^{-1}\left(\tanh\left(\frac{Q_{1\to1}}{2}\right)\cdot\tanh\left(\frac{Q_{2\to1}}{2}\right)\right) \\ &= 2\tanh^{-1}\left(\tanh\left(\frac{1.55}{2}\right)\cdot\tanh\left(\frac{-1.55}{2}\right)\right) \\ &= -0.9 \end{aligned}$$

In a similar way, all of the check-to-variable messages are found:

$$\begin{aligned} R_{1\to1} &= -0.9 & R_{1\to2} &= 0.9 & R_{1\to3} &= -0.9 \\ R_{2\to1} &= -0.9 & R_{2\to4} &= 0.9 & R_{2\to5} &= -0.9 \\ R_{3\to2} &= -0.9 & R_{3\to3} &= 0.9 & R_{3\to6} &= -0.9 \\ R_{4\to4} &= -0.9 & R_{4\to5} &= 0.9 & R_{4\to6} &= -0.9 \end{aligned}$$

Variable node function Consider the operation at variable node 1, which has two incoming messages $R_{1\to1}$ and $R_{3\to1}$ and the channel message L_1. The variable node computes two outputs using (5.12):

$$\begin{aligned} Q_{1\to1} &= L_1 + R_{2\to1} \\ &= 1.55 + (-0.9) \\ &= 0.65 \\ Q_{1\to2} &= L_1 + R_{1\to1} \\ &= 1.55 + (-0.9) \\ &= 0.65 \end{aligned}$$

In a similar way, all of the variable-to-check messages are found:

$$\begin{aligned} Q_{1\to 1} &= 0.6491 & Q_{1\to 2} &= 0.6491 \\ Q_{2\to 1} &= -2.451 & Q_{2\to 3} &= -0.6491 \\ Q_{3\to 1} &= 2.451 & Q_{3\to 3} &= 0.6491 \\ Q_{4\to 2} &= -2.451 & Q_{4\to 4} &= -0.6491 \\ Q_{5\to 2} &= 2.451 & Q_{5\to 4} &= 0.6491 \\ Q_{6\to 3} &= -2.451 & Q_{6\to 4} &= -2.451 \end{aligned}$$

Hard Decision Consider the hard decision at variable node 1, which which has two incoming messages $R_{1\to 1}$ and $R_{3\to 1}$ and the channel message L_1. The variable node computes:

$$\begin{aligned} t_1 &= L_1 + R_{1\to 1} + R_{2\to 1} \\ &= 1.55 + (-0.9) + (-0.9) \\ &= -0.25 \end{aligned}$$

which gives the hard estimate $\widehat{\mathbf{c}}_1 = 1$. In a similar way, all of the hard decisions are found:

$$\begin{aligned} t_1 &= -0.2518 & \widehat{c}_1 &= 1 \\ t_2 &= -1.55 & \widehat{c}_2 &= 1 \\ t_3 &= 1.55 & \widehat{c}_3 &= 0 \\ t_4 &= -1.55 & \widehat{c}_4 &= 1 \\ t_5 &= 1.55 & \widehat{c}_5 &= 0 \\ t_6 &= -3.352 & \widehat{c}_6 &= 1 \end{aligned}$$

Giving the estimated codeword $\widehat{\mathbf{c}} = (1, 1, 0, 1, 0, 1)$.

Check for Convergence Compute $\widehat{\mathbf{c}} \cdot \mathbf{H}^{\mathrm{t}}$ and find that this is equal to $\mathbf{0}$. Thus, $\widehat{\mathbf{c}}$ is a valid codeword and the sum-product decoding algorithm outputs $\widehat{\mathbf{c}}$ and stops.

Simulation Results on BSC The WER and BER is of the LDPC code using (5.15) is evaluated by Monte Carlo simulations on the BSC. For each channel error probability p, the WER and BER listed below were obtained.

p	WER	BER
0.1751	0.532198	0.149636
0.01	0.038847	0.0068047
0.001	0.00380992	0.000635621

This LDPC code has poor WER and BER because it is a short code. The next chapter demonstrates good performance for long block length LDPC codes. The reader may use the table above to verify correctness of the sum-product decoding algorithm.

5.4 Encoding LDPC Codes

Section 3.3.4 showed how to perform efficient encoding if $\mathbf{H}$ is in triangular form. But for LDPC codes, it is difficult to find a suitable parity-check matrix which is triangular. However, if $\mathbf{H}$ is in a so-called approximate lower triangular (ALT) form, efficient encoding can still be performed.

5.4.1 Approximate Lower Triangular Encoding

The following technique finds the systematic codeword. Given information $\mathbf{u}$, the codeword is written as:

$$\mathbf{c} = [\mathbf{u}, \mathbf{p}_1, \mathbf{p}_2], \tag{5.21}$$

where $\mathbf{p}_1$ has g symbols and $\mathbf{p}_2$ has $m - g$ symbols. The parity symbols $\mathbf{p}_1$ and $\mathbf{p}_2$ are found by solving:

$$\mathbf{H}\mathbf{c}^{\mathrm{t}} = \mathbf{0}. \tag{5.22}$$

Suppose that $\mathbf{H}$ can be written as an approximately lower triangular matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \mathbf{C} & \mathbf{D} & \mathbf{E} \end{bmatrix}. \tag{5.23}$$

as shown in Fig. 5.6, where $\mathbf{T}$ is a lower-triangular matrix, and

- $\mathbf{T}$ is $(m - g) \times (m - g)$
- $\mathbf{A}$ is $(m - g) \times (n - m)$
- $\mathbf{B}$ is $(m - g) \times g$
- $\mathbf{E}$ is $g \times (m - g)$
- $\mathbf{C}$ is $g \times (n - m)$
- $\mathbf{D}$ is $g \times g$

The term g is called the "gap." If $g = 0$, then encoding by back-substitution is sufficient as described in Subsection 3.3.4. But even if $g > 0$, parity bits $\mathbf{p}_2$ can be encoded by back substitution, if we have already found the parity bits $\mathbf{p}_1$.

Multiply $\mathbf{H}$ on the left by:

$$\begin{bmatrix} \mathbf{I}_{m-g} & \mathbf{0} \\ -\mathbf{E}\mathbf{T}^{-1} & \mathbf{I}_g \end{bmatrix} \tag{5.24}$$

to obtain:

$$\widetilde{\mathbf{H}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \Gamma & \Phi & \mathbf{0} \end{bmatrix} \tag{5.25}$$

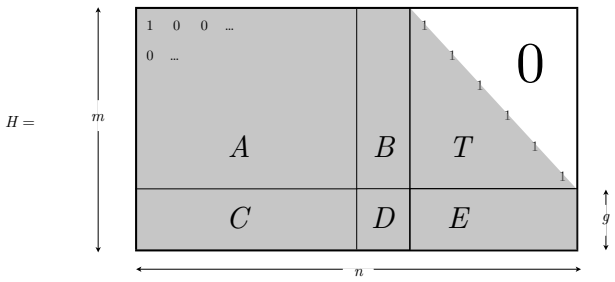


Figure 5.6: Parity-check matrix $\mathbf{H}$ is partially lower-triangular form.

where $\Gamma = -\mathbf{E}\mathbf{T}^{-1}\mathbf{A} + \mathbf{C}$ and $\Phi = -\mathbf{E}\mathbf{T}^{-1}\mathbf{B} + \mathbf{D}$. Multiplying on the left is equivalent to performing row operations — $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ are two parity-check matrices for the same code.

The parity-check equations $\widetilde{\mathbf{H}}\mathbf{c}^{\mathrm{t}} = \mathbf{0}$ become:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \Gamma & \Phi & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}^{\mathrm{t}} \\ \mathbf{p}_1^{\mathrm{t}} \\ \mathbf{p}_2^{\mathrm{t}} \end{bmatrix} = \mathbf{0} \tag{5.26}$$

First, solve the bottom matrix row $\Gamma\mathbf{u}^{\mathrm{t}} + \Phi\mathbf{p}_1^{\mathrm{t}} = \mathbf{0}$ as:

$$\mathbf{p}_1^{\mathrm{t}} = -\Phi^{-1}\Gamma\mathbf{u}^{\mathrm{t}} \tag{5.27}$$

Then, since $\mathbf{u}$ and $\mathbf{p}_1$ are known, $\mathbf{p}_2$ can be found by solving $\mathbf{A}\mathbf{u}^{\mathrm{t}} + \mathbf{B}\mathbf{p}_1^{\mathrm{t}} + \mathbf{T}\mathbf{p}_2^{\mathrm{t}} = \mathbf{0}$. Since $\mathbf{T}$ is lower triangular, $\mathbf{p}_2$ can be found by back-substitution[1]:

$$\mathbf{T}\mathbf{p}_2^{\mathrm{t}} = \mathbf{w}, \tag{5.29}$$

where $\mathbf{w} = -\mathbf{A}\mathbf{u}^{\mathrm{t}} - \mathbf{B}\mathbf{p}_1^{\mathrm{t}}$.

Note that $\mathbf{H}$ should be full rank. Even if $\mathbf{H}$ is full rank, Φ may not be invertible. In this case, swap columns in the left block and center block so that Φ is invertible. Swapping columns changes the position of the code bits.

5.5 Exercises

5.1 Show that if a regular $m \times n$ parity-check matrix $\mathbf{H}$ has d_{v} is even, then the rank cannot be m, that is $\mathrm{rank}(\mathbf{H}) \leq m - 1$.

5.2 *Gallager's check node function* For $\beta > 0$, define $f(\beta)$ as:

$$f(\beta) = \log \frac{e^{\beta} + 1}{e^{\beta} - 1} \tag{5.30}$$

[1] Alternatively, if the matrices are not too large, $\mathbf{p}_2$ may be found by using the inversion of $\mathbf{T}$:

$$\mathbf{p}_2^{\mathrm{t}} = -(\mathbf{u}\mathbf{A}^{\mathrm{t}} + \mathbf{p}_1\mathbf{B}^{\mathrm{t}})(\mathbf{T}^{\mathrm{t}})^{-1} \tag{5.28}$$

(a) Show that $f^{-1}(\beta) = f(\beta)$.
(b) Show that the $d = 3$ check node function can be written as:

$$r_3 = \text{sign}(q_1)\text{sign}(q_2) f\Big(f(|q_1|) + f(|q_2|)\Big) \tag{5.31}$$

5.3 Consider the code with parity-check matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \tag{5.43}$$

Perform erasure decoding with the received sequences:

(a) $\mathbf{y}_1 = [0, 0, 1, ?, ?, ?]$
(b) $\mathbf{y}_2 = [?, ?, ?, 1, 0, 1]$
(c) $\mathbf{y}_3 = [?, ?, ?, 0, 1, ?]$

For each one, give the estimated codeword $\widehat{\mathbf{c}}$, or "failure to decode"

5.4 For the code with parity-check matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix},$$

find systematic codewords corresponding to information sequences:

(a) $\mathbf{u}_1 = (1, 0, 0, 1)$.
(b) $\mathbf{u}_2 = (0, 0, 0, 1)$.
(c) $\mathbf{u}_3 = (1, 1, 1, 1)$,

5.5 Using the code with parity-check matrix in eqn. (4.26) , perform sum-product decoding to find the estimated codeword $\widehat{\mathbf{c}}$ when:

(a) On the BSC with $p = 0.21$ the following $\mathbf{y}$ is received:

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

(b) On the AWGN channel with $\sigma^2 = 1$ and transmit mapping $0 \to +1$ and $1 \to -1$, the following $\mathbf{y}$ is received:

$$\mathbf{y} = \begin{bmatrix} -1.2 & -0.8 & 0.6 & 1 & -1.2 & -0.8 \end{bmatrix}.$$

5.6 Implement sum-product decoding of the $n = 8$ LDPC code in Example 5.1, on the BI-AWGN channel. Transmit the all-zeros codeword, instead of performing encoding. Make a plot of E_b/N_0 vs. WER *and* BER. Your graph should show WER less than 10^{-3}. Submit your source code.

```
function chat = ldpc_decoder(H,Y,ellmax)
[m,n] = size(H);
M     = cell(1,n);
N     = cell(1,m);
for j = 1:m
    N{j} = find( H(j,:) == 1 );
end
for i = 1:n
    M{i} = find( H(:,i) == 1 )'; %transpose is important!
end

%initialize
%R(j,i) is the message r_{j -> i}, Q(j,i) is Q_{i -> j}
R    = zeros(m,n);
Q    = zeros(m,n);
chat = zeros(1,n);
for i = 1:n
    for k = M{i}
        Q(k,i) = Y(i);
    end
end

%begin iterations
for ell = 1:ellmax
    %check node
    for j = 1:m
        for i = N{j}
            e = setdiff(N{j},i);%sum-product update rule
            R(j,i) = 2 * atanh( prod(tanh(Q(j,e) / 2)) );
        end
    end

    %variable node
    for i = 1:n
        for j = M{i}
            e = setdiff(M{i},j);%sum-product update rule
            Q(j,i) = Y(i) + sum( R(e,i) );
        end
    end

    %hard decisions
    for i = 1:n
        e = M{i};
        t(i) = Y(i) + sum( R(e,i) );
        chat(i) = (t(i) < 0);
    end

    %syndrome check
    s = mod(chat * H',2);
    if all(s == 0)
        break;
    end
end
```

Figure 5.7: Matlab source code implementing sum-product decoding.

Algorithm 5.1 Sum-Product Decoding of LDPC Codes

Require: Received sequence $\mathbf{y}$, parity-check matrix $\mathbf{H}$, maximum number of iterations $\ell_{\max}$.

Ensure: Estimated codeword $\widehat{\mathbf{c}}$.

Initialize

for $i = 1, \ldots, n$ **do**

$l_i = \dfrac{2}{\sigma^2} y_i$

$Q_{i \to j} = L_i$ for each $j \in \mathcal{M}_i$.

end for

for $\ell = 1, \ldots, \ell_{\max}$ **do**

Check nodes

for $j = 1, \ldots, m$ **do**

for $i \in \mathcal{N}_j$ **do**

$$R_{j \to i} = 2 \tanh^{-1} \left(\prod_{e \in \mathcal{N}_j \setminus i} \tanh \left(\frac{Q_{e \to j}}{2} \right) \right)$$

end for

end for

Variable nodes

for $i = 1, \ldots, n$ **do**

for $j \in \mathcal{M}_i$ **do**

$$Q_{i \to j} = L_i + \sum_{e \in \mathcal{M}_i \setminus j} R_{e \to i}$$

end for

end for

Hard decisions

for $i = 1, \ldots, n$ **do**

$$t_i = L_i + \sum_{e \in \mathcal{M}_i} R_{e \to i}$$

$$\widehat{c}_i = \begin{cases} 0 & t_i \geq 0 \\ 1 & t_i < 0 \end{cases}$$

end for

Check

if $\widehat{\mathbf{c}} \cdot \mathbf{H}^{\mathrm{t}} = \mathbf{0}$ **then**

Goto End

end if

end for

End: Output estimated codeword $\widehat{\mathbf{c}}$.

Algorithm 5.2 ALT Encoding of LDPC Codes

Require: Full-rank matrix $\mathbf{H}$, with invertible Φ, partitioned according to (5.23). Information sequence $\mathbf{u}$.

Ensure: Systematic codeword satisfying $\mathbf{cH}^{\mathrm{t}}$.

Initialize

Find $\Gamma = -\mathbf{E}\mathbf{T}^{-1}\mathbf{A} + \mathbf{C}$

Find $\Phi = -\mathbf{E}\mathbf{T}^{-1}\mathbf{B} + \mathbf{D}$

Encode

$\mathbf{p}_1^{\mathrm{t}} = -\Phi^{-1}\Gamma\mathbf{u}^{\mathrm{t}}$

$\mathbf{w} = -\mathbf{A}\mathbf{u}^{\mathrm{t}} - \mathbf{B}\mathbf{p}_1^{\mathrm{t}}$

Solve $\mathbf{T}\mathbf{p}_2 = \mathbf{w}$ by back-substitution

Output codeword $\mathbf{c} = [\mathbf{u}, \mathbf{p}_1, \mathbf{p}_2]$.

Chapter 6

Design of LDPC Codes

6.1 Gallager LDPC codes

Given code parameters n, d_{v} and d_{c}, Gallager provided a construction that gives an n-by-m sparse parity-check matrix $\mathbf{H}$ with row degree d_{c}, column degree d_{v} and $m = nd_{\mathsf{v}}/d_{\mathsf{c}}$. The design rate is $R_{\mathrm{d}} = 1 - \frac{d_{\mathsf{v}}}{d_{\mathsf{c}}}$. This is a regular LDPC code. Let $\mathbf{H}_1$ be the matrix consisting of the concatenation of d_{c} identity matrices:

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix}, \tag{6.1}$$

where $\mathbf{I}$ is the $\frac{n}{d_{\mathsf{v}}} \times \frac{n}{d_{\mathsf{v}}}$ identity matrix. The row weight of $\mathbf{H}_1$ is d_{c} and the column weight is 1.

Let π be a pseudo-random[1] column permutation. Then the parity check matrix for the Gallager LDPC construction is:

$$\mathbf{H} = \begin{bmatrix} \pi_1(H_1) \\ \pi_2(H_1) \\ \vdots \\ \pi_{d_{\mathsf{v}}}(H_1) \end{bmatrix}. \tag{6.2}$$

Each of $\pi_1, \pi_2, \ldots,$ are distinct column permutations. So $\mathbf{H}$ has row weight d_{c} and column weight d_{v}.

Example 6.1. Following Gallager, choose π_1 so that $\pi_1(\mathbf{H}_1)$ forms a staircase pattern, although the choice of π_1 is arbitrary. A Gallager construction parity-

[1] Pseudo-random means that it looks random but is "known." Quite often pseudo-random means a random-looking permutation that is known to both the encoder and decoder.

check matrix $\mathbf{H}$ with $n = 20, d_{\mathsf{v}} = 3, d_{\mathsf{c}} = 4$ is:

$$\mathbf{H} = \left[\begin{array}{cccccccccccccccccccc}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\hline
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{array}\right]$$

This code has design rate $R_{\mathrm{d}} = \frac{1}{4}$.

6.2 Irregular LDPC Codes

6.2.1 Degree Distribution

In an irregular LDPC code, the row weight and column weight is not constant. Distributions are used to describe the irregular structure of the parity check matrix. The *node perspective* expresses the fraction of nodes with a given degree. The *edge perspective* expresses the fraction of edges connecting to nodes of a given degree.

For the node perspective, R_j is the fraction of nodes of degree j. Likewise, L_i is the fraction of nodes of degree i. It is convenient to express these distributions as polynomials:

$$R(x) = \sum_{j=1}^{d_{\mathsf{c}}^{\max}} R_j x^j \text{ and } L(x) = \sum_{i=1}^{d_{\mathsf{c}}^{\max}} L_i x^i. \tag{6.3}$$

For the edge perspective, ρ_j is the fraction of edges that connect to a check node of degree j. Likewise, λ_i is the fraction of edges that connect to a variable node of degree i. The sum of the ρ_j is 1 and the sum of the λ_i is also 1. The edge degree distribution polynomials are:

$$\rho(x) = \sum_{j=1}^{d_{\mathsf{c}}^{\max}} \rho_j x^{j-1} \text{ and } \lambda(x) = \sum_{i=1}^{d_{\mathsf{c}}^{\max}} \lambda_i x^{i-1}. \tag{6.4}$$

Note that in $\rho(x)$, the degree of ρ_j is x^{j-1}, whereas in $R(x)$ the degree of R_j is x^j; similarly for the variable node.

The node perspectives and edge perspectives are related by :

$$\rho(x) = \frac{R'(x)}{R'(1)} \tag{6.5}$$

$$\lambda(x) = \frac{L'(x)}{L'(1)} \tag{6.6}$$

The design rate R_{d} is:

$$R_{\mathrm{d}} = 1 - \frac{L'(1)}{R'(1)} = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx} \tag{6.7}$$

Example 6.2. Find $R(x)$, $L(x)$ and $\rho(x)$, $\lambda(x)$ for the following parity-check matrix:

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.8}$$

There are 3 variable nodes of degree 1, 3 variable nodes of degree 2 and 1 variable node of degree 3. Since there are 7 nodes in total:

$$L(x) = \frac{3}{7}x + \frac{3}{7}x^2 + \frac{1}{7}x^3 \tag{6.9}$$

All check nodes have degree 4, so $R(x) = x^4$. For the edge perspective, there are 3 edges connected to degree 1 nodes, 6 edges connected to a degree 2 nodes and 3 edges connected to degree 3 nodes. Since there are 12 edges in total:

$$\lambda(x) = \frac{3}{12} + \frac{6}{12}x + \frac{3}{12}x^2 \tag{6.10}$$

All edges are connected to check nodes of degree 4, so $\rho(x) = x^3$.

6.2.2 Code Ensemble

Analysis of LDPC codes is may be performed not a single code, but on a set or *ensemble* of LDPC codes. All codes in an ensemble have the same degree distribution. They differ in how the variable nodes and check nodes are connected.

Consider for example, the node degree distribution:

$$L(x) = \frac{8}{10}x^2 + \frac{2}{10}x^6 \text{ and } R(x) = x^4. \tag{6.11}$$

The Tanner graph of one possible $n = 10$ code with this degree distribution is shown in Fig. 6.1-(a). Another possible code with the same degree distribution is shown in Fig. 6.1. There is a large but finite possible number of LDPC codes with the same degree distribution.

Now consider to consider not a single code, but the set of all possible codes. Each edge coming from all the variable nodes can be thought of as a “socket”.

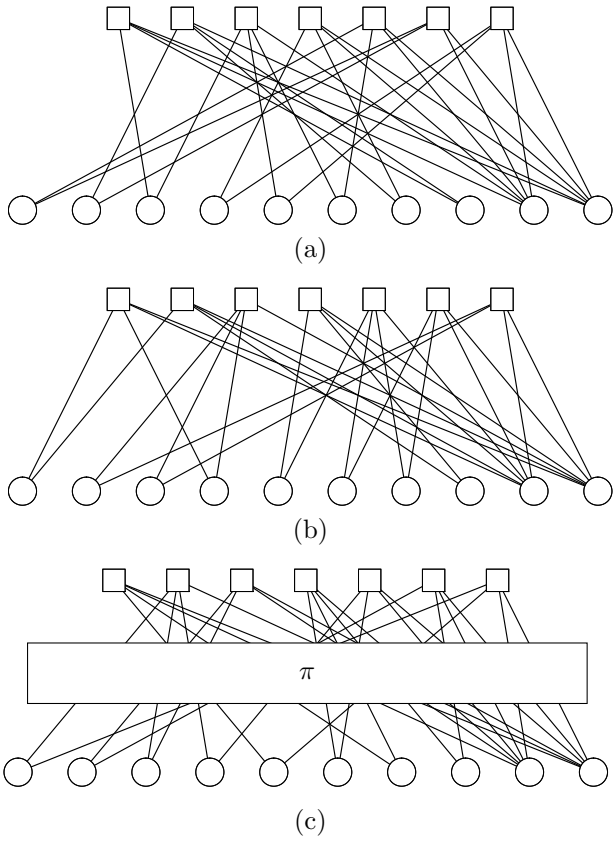


Figure 6.1: (a) One possible LDPC code (b) Another possible LDPC code (c) Ensemble of LDPC codes represented by all possible permutations of edges.

Likewise each edge coming from all the check nodes is also a socket. A pseudorandom permutation connects each variable node socket to each check node socket. For a given degree distribution the *code ensemble* consists of all codes obtained from all distinct permutations, such that there are no cycles of length 2.

When n is small, the structure of the permutation strongly influences the code — some choices of the permutation are good, and some are bad. For example, some permutations may have a large number of short cycles (i.e 4 cycles, 6 cycles), which reduce the performance of the decoder. But as $n \to \infty$ the permutation has less influence on the code. While *some* members of the ensemble will have short cycles, as n gets large, the probability that the members of the ensemble will have a short cycle becomes smaller.

6.2.3 Density Evolution

We already know that increasing the block length n of a code improves its performance. Density evolution is a technique to analyze the decoding of an ensemble of asymptotically long LDPC codes. It analyzes the ensemble with a given degree distribution, rather than a specific code. The analysis ignores any loops in the graph, which corresponds to asymptotically long $n \to \infty$ LDPC codes. We already know that increasing the block length n of a code improves its performance, and the channel capacity is the upper bound on the rate of an asymptotically long code. Density evolution is key to finding the noise threshold described in the next section.

Density evolution tracks the probability distribution of the decoder messages rather than the messages themselves (Probability distributions are alternative called probability density, hence the name of the method). For the considered channels, density evolution will always converge to one of two results: either the probability of decoding error goes to zero (success), or it stops with nonzero probability of decoding error (failure).

The variable-to-check message is a random variable Q and the check-to-variable message is a random variable R. These probability distributions are:

$$f(q) = \Pr(\mathsf{Q} = q) \tag{6.12}$$

$$g(r) = \Pr(\mathsf{R} = r) \tag{6.13}$$

On any given iteration, all the check-to-variable messages are assumed independent and identically distributed — that is, they have the same distribution, in addition to being independent. Similarly for the variable-to-check messages. Rather than one message for each edge, density evolution uses a single distribution, representing all edges.

Density evolution follows the steps of the iterative decoder, replacing messages with their probability densities. On the first iteration, the variable-to-check messages $f^{(1)}(q)$ are initialized with the channel messages. Then the check-to-variable messages $g^{(1)}(r)$ are computed from the variable-to-check messages $f^{(1)}(q)$. This continues as:

$$f^{(1)}(q) \to g^{(1)}(r) \to f^{(2)}(q) \to g^{(2)}(r) \to \cdots .$$

These iterations continue until convergence is detected or non-convergence after a large number of iterations.

Density evolution assumes that the all-zeros codeword was transmitted. This analysis is valid for the BEC, BSC, BI-AWGN channels, and other channels which satisfy a certain symmetry condition. An input to density evolution is the channel parameter ϵ (for BEC), p (for BSC) or σ^2 (for AWGN).

1. Initialization. The distribution $f^{(1)}(q)$ is initalized using the channel output distribution assuming the all-zeros codeword was transmitted, which depends on the channel:

   BEC channel For a BEC with erasure probability ϵ:

   $$f^{(1)}(q) = \begin{cases} 1-\epsilon & q = 0 \\ \epsilon & q = ? \\ 0 & q = 1 \end{cases} \tag{6.14}$$

   $f(1) = 0$ because the all zeros codeword was transmitted

   BSC channel For a BSC channel with error probability p:

   $$f^{(1)}(q) = \begin{cases} 1-p & q = 0 \\ p & q = 1 \end{cases} \tag{6.15}$$

   BI-AWGN channel For the binary-input AWGN channel with noise variance σ^2:

   $$f^{(1)}(q) = \mathcal{N}(-\frac{2}{\sigma^2}, \frac{4}{\sigma^2}), \tag{6.16}$$

   where $\mathcal{N}(m, v)$ is a Gaussian with mean m, variance v. If the all-zeros codeword is transmitted, then the received $y \sim N(-1, \sigma^2)$. Recall that the receiver multiplies y by $\frac{2}{\sigma^2}$ to obtain log-domain, which is used by the decoder.

   This is used to initialize the variable-to-check message f.

2. Check node. A check node of degree d uses the input $f(q)$ $d-1$ times to produce $g_d(r)$. For an irregular code with various edge weights, the check node output $g(r)$ is obtained by weighting each contribution $g_d(r)$ by ρ^d:

   $$g(r) = \sum_{d=1} \rho^d g_d(r). \tag{6.17}$$

   This holds as the fraction of edges connected to a degree d check node is ρ_d.

3. Variable node. A variable node of degree d uses the input $g(r)$ $d-1$ *and* the channel message to produce $f_d(q)$. The variable node output $f(r)$ is obtained similarly to the check node case:

   $$f(q) = \sum_{d=1} \lambda_d f_d(q). \tag{6.18}$$

4. Check for convergence. The probability of error is $\Pr(\mathsf{Q} < 0)$ for log-domain decoding, since the all-zeros codeword was transmitted. If the probability error is sufficiently small (for example, 10^{-10}), then declare convergence and stop. If the number of iterations exceeds some fixed number (for example 200 iterations), then declare non-convergence and stop. Otherwise, go to Step 2.

To implement density evolution in a computer, the distributions g and f must be stored. In the most general case, R and Q are real numbers, and so approximations of g and f are typically formed using a histogram. The challenge is producing $f_d(q)$ from $g(r)$ at check node — a convolution of probability distributions should be computed, as the check node represents a sum of random variables. However, on the BEC, density evolution is considerably simpler and has a pleasing analysis.

6.2.4 Density Evolution for the BEC

Decoding of LDPC codes on the erasure channel was described in Subsection 5.2.3, where the messages were $\{0, ?, 1\}$. The all-zeros codeword is analyzed for density evolution. As a consequence, the messages are from $\{0, ?\}$ only because the 1 message is not possible. This conveniently allows representing the message distribution with a single value. For convenience, let u be the probability the check-to-variable message R is erased, and let v be the probability the variable-to-check message Q is erased:

$$v = f(?) = \Pr(\mathsf{Q} = ?) \tag{6.19}$$

$$u = g(?) = \Pr(\mathsf{R} = ?) \tag{6.20}$$

and $f(0) = 1 - v$ and $g(0) = 1 - u$.

Check node For a check node of degree d, the output is R_d with distribution $u_d = g(?)$. The inputs $\mathsf{Q}_1, \ldots, \mathsf{Q}_{d-1}$, all with the same distribution $v = f(?)$. The probability that the output R_d is erased can be found using the check node decoding rule from Subsection 5.2.3:

$$\begin{aligned} u_d &= \Pr(\text{any of } \mathsf{Q}_1, \mathsf{Q}_2, \ldots, \mathsf{Q}_{d-1} \text{ are } ?) && (6.21)\\ &= 1 - \Pr(\text{none of } \mathsf{Q}_1, \mathsf{Q}_2, \ldots, \mathsf{Q}_{d-1} \text{ are } ?) && (6.22)\\ &= 1 - \Pr(\mathsf{Q}_1 \neq ?)\Pr(\mathsf{Q}_2 \neq ?)\cdots\Pr(\mathsf{Q}_{d-1} \neq ?) && (6.23)\\ &= 1 - (1-v)^{d-1} && (6.24) \end{aligned}$$

To find the probability u, average over all edges according to their weight using eqn. (6.17), so that $u = 1 - \sum_d \rho_d (1-v)^{d-1}$. This may be further simplified using the definition of $\rho(x)$ in (6.4):

$$u(v) = 1 - \rho(1 - v), \tag{6.25}$$

where u is shown as a function of v.

Variable Node For the variable node, the procedure is similar to the check node, but in addition the channel message is used. For a variable node of degree

d, the output is Q_d with distribution $v_d = f(?)$. The inputs are $\mathsf{R}_1, \ldots, \mathsf{R}_{d-1}$, all with the same distribution $u = g(?)$, and Y with $\Pr[\mathsf{Y} = ?] = \epsilon$. The probability that the output Q_d is erased can be found using the check node decoding rule from Subsection 5.2.3:

$$\begin{aligned} v_d &= \Pr(q = ?) && (6.26)\\ &= \Pr(\text{all of } \mathsf{Y}, \mathsf{R}_1, \ldots, \mathsf{R}_{d-1} \text{ are } ?) && (6.27)\\ &= \Pr(\mathsf{Y} = ?) \Pr(\mathsf{R}_1 = ?) \cdots \Pr(\mathsf{R}_{d-1} = ?) && (6.28)\\ &= \epsilon u^{d-1} && (6.29) \end{aligned}$$

Similar to the check node, find v by averaging over all edges according to their weight eqn. (6.18), so that $v = \epsilon \sum_d \rho_d u^{d-1}$. This may be further simplified using the definition of $\lambda(x)$ in (6.4):

$$v(u) = \epsilon \lambda(u). \tag{6.30}$$

where v is shown as a function of u.

Density evolution for the BEC proceeds by tracking the probability that a message is erased as the iterations progress. On iteration ℓ, the check node has input $v^{(\ell)}$ and output $u^{(\ell)}$. The variable node has input $u^{(\ell)}$ and output $v^{(\ell+1)}$. Density evolution is initialized with $v^{(1)} = \epsilon$, the probability of erasure from the channel. Then, the sequence:

$$\epsilon \to u^{(1)} \to v^{(2)} \to u^{(2)} \to v^{(3)} \to u^{(3)} \to \cdots$$

are found by iteratively applying (6.25) and (6.30).

The output of the density evolution procedure is either "converged to 0" or "did not converge to 0 after $\ell_{\max}$ iterations." If during the iterations, $v^{(\ell)} \to 0$ approaches 0, then the probability of decoder error goes to zero, since $v = \Pr[\mathsf{Q} = ?]$, and the decoder succeeded. On the other hand, if after a large number of iterations, $v^{(\ell_{\max})} > 0$, then the decoder failed.

Density evolution for the BEC is conveniently illustrated because the decoder state on any iteration is represented by a single value, either u or v. The check node output $u(v)$ is an function of the input v. On the other hand, the variable node output $v(u)$ is a function of the input u. These two functions are plotted on one graph, where the horizontal axis is v and the vertical axis is u. Beginning with the variable node output initialized with the channel value, the state of the decoder is plotted as a trajectory, found by alternately evaluating $u(v)$ and $v(u)$.

Example 6.3. Consider density evolution for a $d_{\mathsf{v}} = 3, d_{\mathsf{c}} = 6$ regular LDPC code, first on a BEC with $\epsilon = 0.3$. For the check node, $d_{\mathsf{c}} = 6$ means $\rho(x) = x^5$ and applying this to (6.25) gives the check node function:

$$u(v) = 1 - (1 - v)^5.$$

This check node function is plotted in Fig. 6.2-(a), with input v on the horizontal axis and output u on the vertical axis.

For the variable node, $d_v = 3$ means $\lambda(x) = x^2$ and applying this to (6.30) gives the variable node function:

$$v(u) = 0.3u^2.$$

This variable node function is also plotted in Fig. 6.2-(a), but the axes are reversed: the input u is on the vertical axis and the output v is on the horizontal axis.

Density evolution begins by evaluating $u(v)$ with $v = \epsilon = 0.3$, which gives $u^{(1)} = 0.8319$. This is the input to the variable node, giving output $v^{(1)} = 0.2076$. This proceeds iteratively. Because the axes were swapped when plotting $v(u)$, the progress of the decoder states u and v, called the *trajectory* is also drawn in Fig. 6.2-(a), showing the iterative progress of the decoder. The trajectory converges to 0 after about 6 iterations and the decoder succeeds. A long $(3, 6)$ regular LDPC code should have a low probability of decoder error on the BEC with $\epsilon = 0.3$.

Consider instead density evolution on a BEC with $\epsilon = 0.44$, as shown in Fig. 6.2-(b). Because the variable node function depends on ϵ, the curve $v(u)$ has changed, and now the two functions intersect. Because they intersect, the trajectory stops at the first point of intersection. The iterations converge to a point which is not 0, and even after a large number of iterations, convergence to 0 is not achieved and the decoder fails. A $(3, 6)$ regular LDPC code should have a high probability of decoder error when $\epsilon = 0.44$.

6.2.5 Noise Threshold

Example 6.3 in the previous subsection gave two examples: when the channel was good with $\epsilon = 0.3$, the decoder converged to the correct solution. But when the channel was bad with $\epsilon = 0.44$ the decoder did not find the correct solution because the probability of error is nonzero. More generally, if the channel is good, then the decoder should succeed and converge to zero. If the channel is bad (or degraded), then the decoder will fail, converging to nonzero. We want the code of a fixed rate to operate on the worst, or most degraded, channel as possible. The *noise threshold* is the most degraded channel for which the density evolution converges. It is the worst channel for which reliable decoding is possible as the block length $n \to \infty$.

Recall Shannon's channel coding theorem, which states $P_e \to 0$ if and only if $R < C$. In the classical view, the channel is fixed and we want to know the the highest possible code rate with $P_e \to 0$. The alternative view used for discussing LDPC codes, the rate is fixed, and we want to know the worst (most degraded channel) channel with $P_e \to 0$. For example, for a rate $R = \frac{2}{3}$ code on a BEC channel, the capacity is $C = 1 - \epsilon$. Applying the channel coding theorem, $\frac{2}{3} = R < C = 1 - \epsilon$, or $\epsilon < \frac{1}{3}$, so the most degraded channel is $\epsilon^* = \frac{1}{3}$. For density evolution, reliable communications means that the density evolution state predicts $P_e \to 0$.

For the BEC with erasure probability ϵ the noise threshold is the *largest* value ϵ^* such that convergence is obtained . For the code of Fig. 6.2, the noise

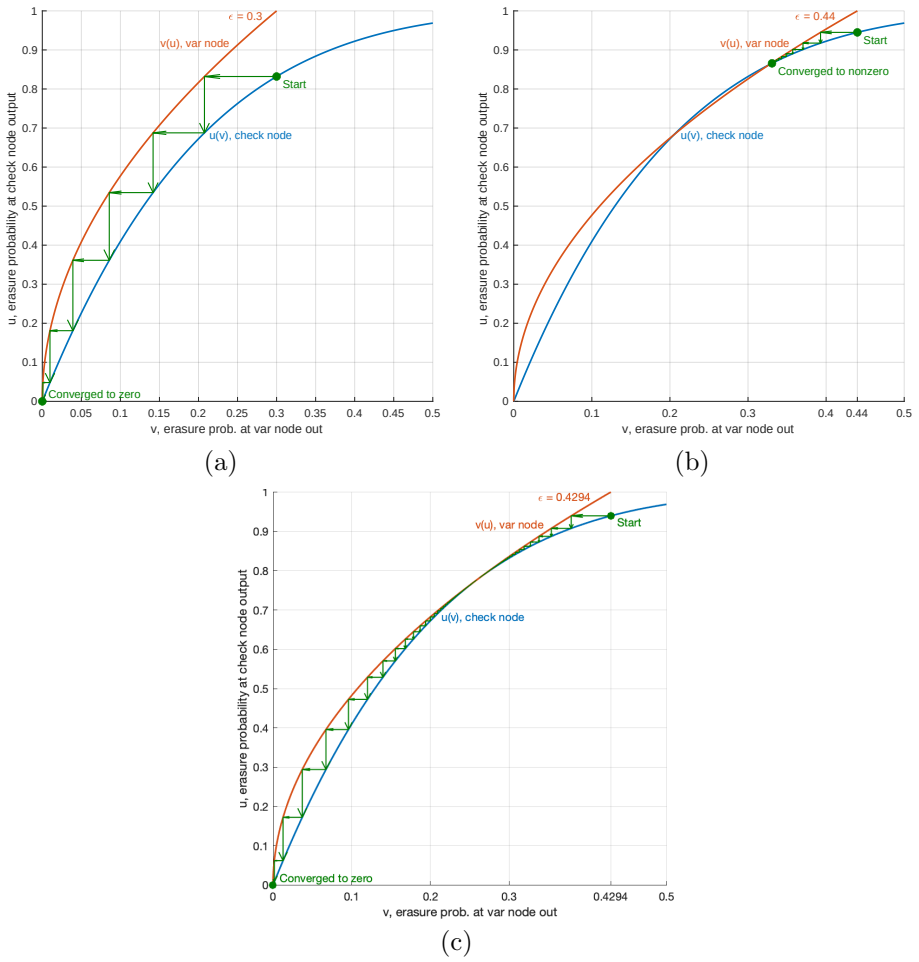


Figure 6.2: Density evolution on BEC for design rate 1/2 LDPC with variable node degree 3, check node degree 6. (a) For BEC erasure probability $\epsilon = 0.3$, the iterations converge to probability of erasure 0 (b) For $\epsilon = 0.44$, the iterations do not converge to 0. (c) $\epsilon = 0.4295$ is the largest value such that convergence to 0 is obtained, and thus is the noise threshold for this code.

threshold is $\epsilon^* = 0.4294$. The trajectory for this channel is shown in Fig. 6.2-(c). It can be seen that the variable node curve nearly touches the check node curve, leaving just enough space for the trajectory to pass through. Although it may not be clear from Fig. 6.2-(c), nearly 500 iterations are needed to converge to 0. The following table shows the noise thresholds for various regular LDPC codes with design rate 1/2.

$(d_\mathrm{v}, d_\mathrm{c})$	R_d	ϵ^*
(2,4)	0.5	0.3333
(3,6)	0.5	0.4294
(4,8)	0.5	0.3834
(5,10)	0.5	0.3415

Table 6.1: Degree distributions for three irregular LDPC codes with design rate $R_{\mathrm{d}} \approx 0.5$.

code	$\lambda(x)$	$\rho(x)$	R_{d}	ϵ^*
1	$\frac{1}{10}x + \frac{1}{2}x^2 + \frac{1}{100}x^{10} + \frac{39}{100}x^{19}$	$\frac{1}{2}x^7 + \frac{1}{2}x^8$	0.502034	0.470846
2	$\lambda_2(x)$	x^5	0.500367	0.480977
3	$\lambda_3(x)$	x^5	0.498958	0.480721

where:

$$\begin{aligned}\lambda_2(x) =& 0.409x + 0.202x^2 + 0.0768x^3 + 0.1971x^6 + 0.1151x^7 \\ \lambda_3(x) =& 0.416x + 0.166x^2 + 0.1x^3 + 0.07x^4 + 0.053x^5 + 0.042x^6 + 0.035x^7 \\ &+ 0.03x^8 + 0.026x^9 + 0.023x^{10} + 0.02x^{11} + 0.0183x^{12}\end{aligned}$$

For the BSC with error probability p, the noise threshold is the largest value p^*. For the AWGN with noise variance σ^2, the noise threshold is the largest value $(\sigma^*)^2$ (or the smallest SNR value).

For the BEC with parameter ϵ, the Shannon capacity is $C = 1 - \epsilon$. For $\epsilon = 0.5$, as the block length $n \to \infty$, there exists a code with $R < C = 1/2$ for which reliable communication on this channel is possible. From density evolution results in the previous section, a column weight 3, rate 1/2 LDPC code has a noise threshold of 0.4294. This is the most degraded channel for which such an LDPC code has reliable decoding. From this, we conclude that regular LDPC codes cannot achieve channel capacity.

Now consider density evolution for irregular LDPC codes. However, irregular LDPC codes can more closely approach the channel capacity. The table below shows 3 irregular LDPC codes, all with design rate close to 0.5. The best irregular code in the table below has a noise threshold of $\epsilon^* = 0.480977$, which is higher than the best regular code. This irregular code, while close to channel capacity, does not achieve the channel capacity, a gap of 0.019023 remains.

6.3 Protograph and Quasi-Cyclic LDPC Codes

We're doing pretty well with regular LDPC codes and further improved noise noise thresholds with irregular LDPC codes. This was infinite-length analysis. Now we want to design finite-length codes. There are a number of approaches to designing LDPC codes, and we choose to focus on one.

Protographs are a method to construct LDPC codes. Many of the codes used in communication standards use LDPC codes based on protographs, including 5G New Radio, WiFi IEEE 802.11n and WiMax IEEE 802.16e. Quasi-cyclic LDPC codes are a specific type of protograph-based codes. QC-LDPC have a number of advantages. A large and random-like parity check matrix $\mathbf{H}$ is specified using a comparatively small matrix of integers. It is efficient to identify

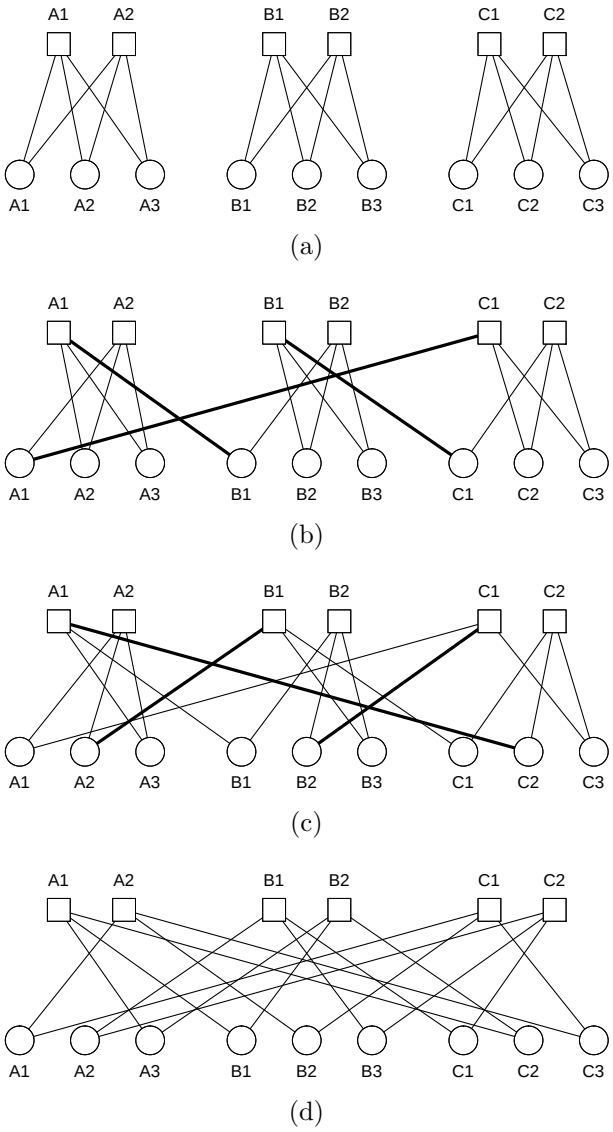


Figure 6.3: Protograph construction for Example 6.4. (a) Base graph repeated $z = 3$ times. (b) The edges connecting variable node 1 and check node 1 are permuted. (c) The edges connecting variable node 2 and check node 1 are permuted. (d) Fully permuted protograph.

and remove 4 and 6 cycles in the corresponding graph. Their structure makes them well suited for hardware implementations.

6.3.1 Protograph Construction

Protograph codes are constructed by repeating a small base graph several times, and then permuting the edges between the copies. In particular begin with a small bipartite graph called a base graph. Duplicate this graph so there are z independent copies. Consider any edge e in the original graph. In the repeated graph e appears z times, and they are connected to z check nodes. To form the protograph code, the connection of the z edges to these z check nodes is permuted. This is repeated for all edges in the base graph. This process of constructing a graph from a protograph is called lifting, and the resulting graph is called the *lifted graph*.

Example 6.4. Lift the following base graph with $z = 3$:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{6.31}$$

Begin by making 3 copies of the base graph, shown in Fig. 6.3-(a). First, the edges connecting variable node 1 and check node 1 are permuted, as shown in Fig. 6.3-(b). This continues for the edges connecting variable node 2 and check node 1, as shown in Fig. 6.3-(c). A final graph with all edges permuted is shown in Fig. 6.3-(d).

The base graph and the lifted graph have a parity-check matrix. Let $\mathbf{H}_{\text{base}}$ be the protograph M-by-N matrix containing 0s and 1s, then parity-check matrix $\mathbf{H}$ for the lifted code has size zM-by-zN. The process of constructing $\mathbf{H}$ from $\mathbf{H}_{\text{base}}$ is called lifting.

After repeating the base graph $z = 3$ times, the lifted graph shown in Fig. 6.3-(a) corresponds to matrix:

$$\begin{array}{c} \begin{array}{ccc|ccc|ccc} \text{A1} & \text{B1} & \text{C1} & \text{A2} & \text{B2} & \text{C2} & \text{A3} & \text{B3} & \text{C3} \end{array} \\ \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \text{A1} \\ \text{B1} \\ \text{C1} \\ \\ \text{A2} \\ \text{B2} \\ \text{C2} \end{array} \end{array}$$

The rows and columns of the matrix have been labeled to show the correspondence with the variable nodes and check nodes, because the order is not the same.

After applying all permutations, the lifted graph shown in Fig. 6.3-(d) has matrix:

$$\begin{array}{c} \begin{array}{ccc|ccc|ccc} \text{A1} & \text{B1} & \text{C1} & \text{A2} & \text{B2} & \text{C2} & \text{A3} & \text{B3} & \text{C3} \end{array} \\ \left[\begin{array}{ccc|ccc|ccc} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ \text{A1} \\ \text{B1} \\ \text{C1} \\ \text{A2} \\ \text{B2} \\ \text{C2} \end{array}$$

The permutation applied for each edge can be seen in each 3-by-3 block shown of the matrix.

6.3.2 Cyclic Permutation Matrix

A $z \times z$ *right-shift cyclic permutation matrix* $\mathbf{P}$ has the property that:

$$(x_z, x_1, x_2, \ldots, x_{z-1}) = (x_1, x_2, \ldots, x_z) \cdot \mathbf{P} \tag{6.32}$$

For example, if $z = 5$ then:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6.33}$$

Note that $\mathbf{P}^2$ is shift-by-two:

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$\mathbf{P}^3$ is shift-by-three, etc. Also $\mathbf{P}^0 = \mathbf{I}_q$, the $q \times q$ identity matrix, that is, no shift. If $t = i \bmod q$, then $\mathbf{P}^i$ is shift-by-t.

6.3.3 Quasi-Cyclic LDPC Codes

A quasi-cyclic LDPC code is a protograph code where the edge permutations are cyclic permutations. If the permutations correspond to a cyclic permutation, then we have a *quasi-cyclic LDPC code.* A protograph code where the random permutation is a permutation matrix is a *quasi-cyclic code.*

The 1s in the protograph matrix are replaced by z-by-z permutation matrices to obtain a zM by zN parity-check matrix $\mathbf{H}$. In the matrix representation of the lifted code, each block was represented by a $z \times z$ permutation. For QC-LDPC codes, we work directly with a parity-check matrix $\mathbf{H}$ formed from

permutation matrices. The parity-check matrix for a QC-LDPC code is given by:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}^{p_{1,1}} & \mathbf{P}^{p_{1,2}} & \cdots & \mathbf{P}^{p_{1,N}} \\ \mathbf{P}^{p_{2,1}} & \mathbf{P}^{p_{2,2}} & \cdots & \mathbf{P}^{p_{2,N}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}^{p_{M,1}} & \mathbf{P}^{p_{M,2}} & \cdots & \mathbf{P}^{p_{M,N}} \end{bmatrix}, \tag{6.34}$$

where $p_{j,i} \geq 0$ or $p_{j,i} = -1$. If $p \geq 0$, then $\mathbf{P}^p$ is the usual shift-by-p right-cyclic shift operation. In particular $\mathbf{P}^0 = \mathbf{I}_z$, the identity matrix. However, it may be necessary to designate an all-zeros matrix, and thus $p = -1$ is thus reserved, that is $\mathbf{P}^{-1} \stackrel{\text{def}}{=} \mathbf{0}$, the $z \times z$ all-zeros matrix.

The above matrix may be more conveniently represented by the $M \times N$ prototype matrix, $\mathbf{H}_{\text{proto}}$:

$$\mathbf{H}_{\text{proto}} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,N} \end{bmatrix}. \tag{6.35}$$

In the example above, a 2×3 base graph was lifted with $z = 3$ to form a 6×9 code using cyclic permutations. The code corresponding to the final graph is given by:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}^1 & \mathbf{P}^2 & \mathbf{P}^0 \\ \mathbf{P}^0 & \mathbf{P}^1 & \mathbf{P}^2 \end{bmatrix} \tag{6.36}$$

where

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \tag{6.37}$$

To represent a QC-LDPC code, only z and the powers of $\mathbf{P}$ are needed. The matrix in (6.36) can be represented by:

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} \tag{6.38}$$

and $z = 3$. This compact representation is useful for large LDPC codes.

Example 6.5. Construct a parity-check matrix $\mathbf{H}$ from the prototype matrix:

$$\mathbf{H}_{\text{proto}} = \begin{bmatrix} 3 & -1 & 0 & -1 \\ -1 & 2 & 0 & 0 \\ 0 & 2 & -1 & 0 \end{bmatrix}$$

using $z = 4$.

The $\mathbf{H}$ matrix is obtained by raising $\mathbf{P}$ to the power of the non-negative integer in $\mathbf{H}_{\text{proto}}$, or using the 4×4 all zeros matrix $\mathbf{0}$ for -1:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}^3 & \mathbf{0} & \mathbf{P}^0 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^2 & \mathbf{P}^0 & \mathbf{P}^0 \\ \mathbf{P}^0 & \mathbf{P}^2 & \mathbf{0} & \mathbf{P}^0 \end{bmatrix} \text{ where } \mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

which is explicitly written as:

$$\mathbf{H} = \left[\begin{array}{cccc|cccc|cccc|cccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

6.3.4 Cycles in QC-LDPC Codes

Consider cycles in the graph. If the base matrix has no cycles of length l, then the lifted graph will also have no cycles of length l. However, even if the base matrix has a cycle of length l, the lifted graph will not necessarily have a cycle of length l.

Consider a protograph which has cycles of length $l = 4$ and length $l = 6$. For $l = 4$, the cycle is:

$$(j_1, i_1) \to (j_1, i_2) \to (j_2, i_2) \to (j_2, i_1) \to (j_1, i_1).$$

where j_1, j_2 are rows and i_1, i_2 are columns. Then, a cycle exists if and only if:

$$p_{j_1,i_1} - p_{j_1,i_2} + p_{j_2,i_2} - p_{j_2,i_1} = 0 \pmod{z} \tag{6.39}$$

For $l = 6$, the cycle is:

$$(j_1, i_1) \to (j_1, i_2) \to (j_2, i_2) \to (j_2, i_3) \to (j_3, i_3) \to (j_3, i_1) \to (j_1, i_1).$$

where j_1, j_2, j_3 are rows and i_1, i_2, i_3 are columns. Then, a cycle exists if and only if:

$$p_{j_1,i_1} - p_{j_1,i_2} + p_{j_2,i_2} - p_{j_2,i_3} + p_{j_3,i_3} - p_{j_3,i_1} = 0 \pmod{z} \tag{6.40}$$

For cycles of length $l = 8$, it is more complicated to identify the cycles. Whereas for $l = 4$ and $l = 6$, each has one protograph pattern which can generate a cycle, for $l = 8$, there are several protograph patterns which can can induce a cycle of length 8.

Table 6.2: Prototype matrices for $n = 648$ LDPC codes with $z = 27$, from the 802.11n-2009 standard.

Code rate $R = \frac{1}{2}$ with $k = 324$

0	-1	-1	-1	0	0	-1	-1	0	-1	-1	0	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
22	0	-1	-1	17	-1	0	0	12	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
6	-1	0	-1	10	-1	-1	-1	24	-1	0	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	0	20	-1	-1	-1	25	0	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
23	-1	-1	-1	3	-1	-1	-1	0	-1	9	11	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1
24	-1	23	1	17	-1	3	-1	10	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1
25	-1	-1	-1	8	-1	-1	-1	7	18	-1	-1	0	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
13	24	-1	-1	0	-1	8	-1	6	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1
7	20	-1	16	22	10	-1	-1	23	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1
11	-1	-1	-1	19	-1	-1	-1	13	-1	3	17	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1
25	-1	8	-1	23	18	-1	14	9	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0
3	-1	-1	-1	16	-1	-1	2	25	5	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0

Code rate $R = \frac{2}{3}$ with $k = 432$

25	26	14	-1	20	-1	2	-1	4	-1	-1	8	-1	16	-1	18	1	0	-1	-1	-1	-1	-1	-1
10	9	15	11	-1	0	-1	1	-1	-1	18	-1	8	-1	10	-1	-1	0	0	-1	-1	-1	-1	-1
16	2	20	26	21	-1	6	-1	1	26	-1	7	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1
10	13	5	0	-1	3	-1	7	-1	-1	26	-1	-1	13	-1	16	-1	-1	-1	0	0	-1	-1	-1
23	14	24	-1	12	-1	19	-1	17	-1	-1	-1	20	-1	21	-1	0	-1	-1	-1	0	0	-1	-1
6	22	9	20	-1	25	-1	17	-1	8	-1	14	-1	18	-1	-1	-1	-1	-1	-1	-1	0	0	-1
14	23	21	11	20	-1	24	-1	18	-1	19	-1	-1	-1	-1	22	-1	-1	-1	-1	-1	-1	0	0
17	11	11	20	-1	21	-1	26	-1	3	-1	-1	18	-1	26	-1	1	-1	-1	-1	-1	-1	-1	0

Code rate $R = \frac{2}{3}$ with $k = 432$

16	17	22	24	9	3	14	-1	4	2	7	-1	26	-1	2	-1	21	-1	1	0	-1	-1	-1	-1
25	12	12	3	3	26	6	21	-1	15	22	-1	15	-1	4	-1	-1	16	-1	0	0	-1	-1	-1
25	18	26	16	22	23	9	-1	0	-1	4	-1	4	-1	8	23	11	-1	-1	-1	0	0	-1	-1
9	7	0	1	17	-1	-1	7	3	-1	3	23	-1	16	-1	-1	21	-1	0	-1	-1	0	0	-1
24	5	26	7	1	-1	-1	15	24	15	-1	8	-1	13	-1	13	-1	11	-1	-1	-1	-1	0	0
2	2	19	14	24	1	15	19	-1	21	-1	2	-1	24	-1	3	-1	2	1	-1	-1	-1	-1	0

Code rate $R = \frac{5}{6}$ with $k = 540$

17	13	8	21	9	3	18	12	10	0	4	15	19	2	5	10	26	19	13	13	1	0	-1	-1
3	12	11	14	11	25	5	18	0	9	2	26	26	10	24	7	14	20	4	2	-1	0	0	-1
22	16	4	3	10	21	12	5	21	14	19	5	-1	8	5	18	11	5	5	15	0	-1	0	0
7	7	14	14	4	16	16	24	24	10	1	7	15	6	10	26	8	18	21	14	1	-1	-1	0

6.3.5 Design of QC-LDPC Codes

The degree distribution of the base matrix is the same as the lifted matrix. This may be used to design irregular QC-LDPC codes.

Design of quasi-cyclic LDPC codes:

- Begin with an irregular degree distribution.
- Create a base graph with that degree distribution
- Lift the graph by z to create a long code
- Choose circulant powers to minimize short cycles in the graph.

6.4 Exercises

6.1 Noise thresholds for regular LDPC codes with rate $\frac{2}{3}$.

(a) What pairs of $(d_{\mathsf{v}}, d_{\mathsf{c}})$ give a regular LDPC code with design rate $\mathcal{R}_d = \frac{2}{3}$?
(b) Plot the check node function $u(v)$ on the BEC with $\epsilon = 0.2$. Plot the variable node function $v(u)$ on the same graph with the axes reversed. What value does the decoder converge to?
(c) Find the noise threshold for the codes from part (a). Which $(d_{\mathsf{v}}, d_{\mathsf{c}})$ pair gives the highest noise threshold for the BEC?

6.2 Consider an irregular LDPC code with node perspective degree distribution:

$$R(x) = x^{11}$$
$$L(x) = \frac{189}{648}x^2 + \frac{216}{648}x^3 + \frac{135}{648}x^4 + \frac{27}{648}x^6 + \frac{81}{648}x^8$$

(a) Find the design rate and the edge perspective distribution.
(b) For a BEC with $\epsilon = 0.2$, plot the check node function $u(v)$. Plot the variable node function $v(u)$ on the same graph with the axes reversed.
(c) Find the noise threshold for this code on the BEC.
(d) What is ϵ^*, the BEC channel whose capacity matches $\mathcal{R}_d$ from part (a)? How close is the noise threshold in part (c) to this ϵ^*?

6.3 Perform sum-product decoding of the $(648, 432)$ IEEE 802.11n-2009 LDPC code, on the BI-AWGN channel. Transmit the all-zeros codeword, instead of performing encoding. Make a plot of E_b/N_0 vs. WER *and* BER. On the BER graph, include the probability of uncoded transmission. Your graph should show WER less than 10^{-3}. Submit your source code.

Chapter 7

Polar Codes

Polar codes were introduced by Erdal Arıkan in 2009. While it has been known for some time that random linear codes can achieve channel capacity, polar codes are the first deterministic construction that can achieve the capacity of binary memoryless symmetric (BMS) channels. They have relatively low decoding complexity, and a flexible construction, in the sense of being able to choose any rate. Polar codes are used for the control channel (as opposed to the data channel) in 5G mobile broadband communications.

7.1 Preliminaries

7.1.1 Kronecker Product

For an $n \times m$ matrix $\mathbf{A}$ and a $p \times q$ matrix $\mathbf{B}$, their *Kronecker product* $\mathbf{A} \otimes \mathbf{B}$ is the $np \times mq$ matrix $\mathbf{A} \otimes \mathbf{B}$ given by:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix}$$

Powers $s \geq 1$ of a Kroneker matrix are:

$$\mathbf{A}^{\otimes s} = \mathbf{A} \otimes \mathbf{A}^{\otimes s-1} = \underbrace{\mathbf{A} \otimes \mathbf{A} \otimes \cdots \otimes \mathbf{A}}_{s \text{ times}} \tag{7.1}$$

The Kronecker product is associative: $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$. Matrix transpose is distributive over the Kroneker product: $(\mathbf{A} \otimes \mathbf{B})^{\mathrm{t}} = \mathbf{A}^{\mathrm{t}} \otimes \mathbf{B}^{\mathrm{t}}$. The *mixed product property* states: If $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are matrices of size such that the products $\mathbf{AC}$ and $\mathbf{BD}$ exist then:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}). \tag{7.2}$$

Wikipedia: Kronecker product

The following matrix $\mathbf{F}^{\otimes n}$ is important in the construction of polar codes:

$$\mathbf{F}^{\otimes n} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} \tag{7.3}$$

for $n \geq 1$, and $N = 2^n$, the matrix is an $N \times N$ square matrix. This matrix is its own inverse over the binary field: $(\mathbf{F}^{\otimes n})^{-1} = \mathbf{F}^{\otimes n}$.

Example 7.1. The $\mathbf{F}^{\otimes n}$ matrices for $N = 2, 4$ and 8 are:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{7.4}$$

$$\mathbf{F}^{\otimes 2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{7.5}$$

$$\mathbf{F}^{\otimes 3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{7.6}$$

The matrix $\mathbf{F}^{\otimes n}$ is related to the Hadamard matrix. Wikipedia: Hadamard matrix

7.1.2 Reverse Shuffle and Bit-Reversal Permutations

A permutation of a vector is a rearrangement of the elements of that vector. A permutation matrix has one 1 in each column and one 1 in each row; all other entries are 0. A permutation matrix $\mathbf{P}$ is an orthogonal matrix, so $\mathbf{P}^{-1} = \mathbf{P}^{\mathrm{t}}$, that is, the inverse permutation may be found using $\mathbf{P}^{\mathrm{t}}$. Wikipedia: Permutation matrix

---

Example 7.2. The permutation:

$$(1, 2, 3, 4) \to (1, 4, 3, 2) \tag{7.7}$$

Can be induced by multiplying on the right with matrix $\mathbf{P}$:

$$(1, 2, 3, 4) \cdot \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{P}} = (1, 4, 3, 2) \tag{7.8}$$

A *perfect shuffle* is a permutation of N elements, $N \geq 4$ and N even, expressed using permutation matrix $\mathbf{S}_N$:

$$(1, 2, 3, 4, \ldots, N) \cdot \mathbf{S}_N = (1, \frac{N}{2} + 1, 2, \frac{N}{2} + 2, 3, \frac{N}{2} + 3, \ldots, \frac{N}{2}, N). \tag{7.9}$$

For example with $N = 8$ the perfect shuffle is:

$$(1, 2, 3, 4, 5, 6, 7, 8) \rightarrow (1, 5, 2, 6, 3, 7, 4, 8). \tag{7.10}$$

For $N = 2$, $\mathbf{S}_2 = \mathbf{I}_2$. Wikipedia: Perfect Shuffle

The *reverse shuffle* $\mathbf{R}_N$ is the inverse of the perfect shuffle:

$$(1, 2, 3, 4, \ldots, N) \cdot \mathbf{R}_N = (1, 3, 5, \ldots, N - 1, 2, 4, 6, \ldots, N), \tag{7.11}$$

For example, with $N = 8$, the reverse shuffle is:

$$(1, 2, 3, 4, 5, 6, 7, 8) \rightarrow (1, 3, 5, 7, 2, 4, 6, 8). \tag{7.12}$$

Since $\mathbf{R}_N$ and $\mathbf{S}_N$ are permutation matrices, $\mathbf{R}_N = \mathbf{S}_N^{-1} = \mathbf{S}_N^{\mathrm{t}}$ and $\mathbf{R}_N \mathbf{S}_N = \mathbf{I}_N$.

Example 7.3. The $N = 8$ reverse shuffle matrix is:

$$\mathbf{R}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7.13}$$

which gives the reverse shuffle:

$$(1, 2, 3, 4, 5, 6, 7, 8) \cdot \mathbf{R}_8 = (1, 3, 5, 7, 2, 4, 6, 8) \tag{7.14}$$

The Kronecker product does not commute in general, but if $\mathbf{A}$ has two rows, and $\mathbf{S}_N$ is the perfect shuffle matrix (described in Subsection 7.1.2) then:

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{S}_N^{\mathrm{t}} (\mathbf{B} \otimes \mathbf{A}) \mathbf{S}_N \tag{7.15}$$

A *bit-reversal shuffle* of N numbers is the permutation induced by reversing the order of bits in an n-bit binary representation. The binary string $(d_0 d_1 \ldots d_{n-1})$ represents the integer $\sum_{i=0}^{n-1} 2^i d_i$, where $n = \log_2 N$; d_0 is the LSB (least significant bit) and d_{n-1} is the MSB.

Consider for example $N = 8$. The bits and their reversals are:

0	000	—	000	0
1	100	—	001	4
2	010	—	010	2
3	110	—	011	6
4	001	—	100	1
5	101	—	101	5
6	011	—	110	3
7	111	—	111	7

This induces the permutation $(0,1,2,3,4,5,6,7) \to (0,4,2,6,1,5,3,7)$, which is equivalent to the 1-index permutation:

$$(1,2,3,4,5,6,7,8) \to (1,5,3,7,2,6,4,8) \tag{7.16}$$

The permutation matrix for bit-reversal can be found recursively for $N = 4, 8, 16, 32, \ldots$,

$$\mathbf{B}_N = \mathbf{R}_N \cdot (\mathbf{I}_2 \otimes \mathbf{B}_{N/2}), \tag{7.17}$$

where $\mathbf{B}_2 = \mathbf{I}_2$.

Example 7.4. The $N = 4$ bit-reversal shuffle is:

$$\begin{aligned}\mathbf{B}_4 &= \mathbf{R}_4 \cdot (\mathbf{I}_2 \otimes \mathbf{B}_2) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\end{aligned}$$

since $\mathbf{I}_2 \otimes \mathbf{B}_2 = \mathbf{I}_4$. This is used to find the $N = 8$ bit-reversal shuffle $\mathbf{B}_8$:

$$\begin{aligned}\mathbf{B}_8 &= \mathbf{R}_8 \cdot (\mathbf{I}_2 \otimes \mathbf{B}_4) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ 0 & 1 & 0 & 0 & & & 0 & \\ 0 & 0 & 0 & 1 & & & & \\ & & & & 1 & 0 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 \\ & & 0 & & 0 & 1 & 0 & 0 \\ & & & & 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}\end{aligned} \tag{7.18}$$

This matrix induces the permutation:

$$(1,2,3,4,5,6,7,8) \cdot \mathbf{B}_8 = (1,5,3,7,2,6,4,8), \tag{7.19}$$

which is the permutation in (7.16).

7.2 Polar Codes

7.2.1 Definition

Definition 7.1. For $N = 2^n, n \geq 1$ and $\mathcal{A} \subseteq \{1,2,3,\ldots,N\}$, a *polar code* is the linear block code with codewords generated by the rows indexed by $\mathcal{A}$ in the matrix:

$$\mathbf{G}_N = \mathbf{B}_N \cdot \mathbf{F}^{\otimes n}, \tag{7.20}$$

where $\mathbf{B}_N$ is the bit-reversal permutation matrix. The set $\mathcal{A}$ is the *information set*.

The dimension of a polar code is $K = |\mathcal{A}|$, the cardinality of the set $\mathcal{A}$, and the rate is $R = K/N$. Since $\mathbf{B}_N$ is a permutation matrix, polar codes could have been defined using $\mathbf{F}^{\otimes n}$ alone. However, the ordering induced by the permutation $\mathbf{B}_N$ will simplify the decoding.

Distinct from other linear block codes, for polar codes we write the information vector $\mathbf{u}$ with N elements rather than K elements:

$$\mathbf{u} = (u_1, u_2, u_3, \ldots, u_N). \tag{7.21}$$

Information is placed into the K positions indicated by $\mathcal{A}$. The remaining $N-K$ bits are called *frozen bits*, and are set to be constant. For convenience frozen bits are set to 0, but in general may be set to 0 or 1 arbitrarily, assuming that the encoder and decoder agree on the value of the frozen bits. Because there are K information positions, the dimension of the code remains K. Thus, codeword $\mathbf{c}$ corresponding to $\mathbf{u}$ is

$$\mathbf{c} = \mathbf{u} \cdot \mathbf{G}_N. \tag{7.22}$$

Example 7.5. Consider the $N = 8$ polar code with information set $\mathcal{A} = \{4,6,7,8\}$. There are $K = |\mathcal{A}| = 4$ information bits so the rate is $R = 0.5$. The information vector is:

$$\mathbf{u} = (0,0,0,u_4,0,u_6,u_7,u_8) \tag{7.23}$$

The generator matrix is $\mathbf{G}_8 = \mathbf{B}_8 \cdot \mathbf{F}^{\otimes 3}$:

$$\mathbf{G}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{7.24}$$

where $\mathbf{B}_8$ is given by (7.18) and $\mathbf{F}^{\otimes 3}$ is given by (7.6). Codewords can be found from $\mathbf{u} \cdot \mathbf{G}_8$.

Note that due to the frozen bits, 4 rows of $\mathbf{G}_8$ do not contribute to the code, so a codeword $\mathbf{c}$ can also be generated by:

$$\mathbf{c} = [u_4, u_6, u_7, u_8] \cdot \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \tag{7.25}$$

which is the more familiar K-by-N form of the generator matrix. This matrix generates the same code as the first-order Reed-Muller code of Example 3.20, after a permutation of the code bits.

7.2.2 Representation of Polar Codes

A graphical representation of polar codes is useful for encoding, decoding, and understanding polar code's recursive structure. First, define a permutation matrix $\mathbf{P}$ which consists of repeats of the reverse shuffle $\mathbf{R}$. In particular, let $n = \log_2 N$ and define $\mathbf{P}_n^{(\ell)}$ as:

$$\mathbf{P}_N^{(\ell)} = \mathbf{I}_{2^{\ell-1}} \otimes \mathbf{R}_{2^{n-\ell+1}}, \tag{7.26}$$

for $1 \leq \ell \leq n$. When $\ell = 1$ the length of the reverse shuffle is equal to N, so no repetition is necessary and $\mathbf{P}_N^{(1)} = \mathbf{R}_N$. When $\ell = 2$, the reverse shuffle $\mathbf{R}_{N/2}$ is repeated twice. When $\ell = n$, $\mathbf{P}_N^{(\ell)}$ is the identity matrix, that is $\mathbf{P}_N^{(n)} = \mathbf{I}_N$.

The following decomposition of $\mathbf{G}_N$ is a useful representation for encoding and decoding.

Proposition 7.1. The polar code generator matrix $\mathbf{G}_N$ as given in Definition 7.1 can be written as:

$$\mathbf{G}_N = (\mathbf{I}_{N/2} \otimes \mathbf{F}) \prod_{i=1}^{n-1} \left(\mathbf{P}_N^{(i)} \cdot (\mathbf{I}_{N/2} \otimes \mathbf{F}) \right) \tag{7.27}$$

Consider this decomposition for various values of N. When $N = 2$:

$$\mathbf{G}_2 = \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{7.28}$$

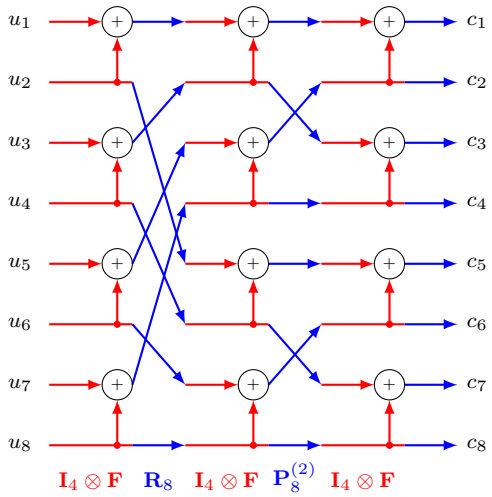


Figure 7.3: Encoder structure for $N = 8$ polar code.

A block diagram representation is useful for expressing the encoder, the decoder as well as the recursive structure of polar codes. Beginning with $N = 2$, the encoding function $\mathbf{c} = \mathbf{u}\mathbf{G}_2$ with $\mathbf{G}_2 = \mathbf{F}$ is $c_1 = u_1 + u_2$ and $c_2 = u_2$. A block diagram for this encoding is shown in Fig. 7.1. This block is a fundamental unit of polar codes, which will be repeated and used recursively for all larger values of N.

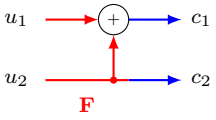


Figure 7.1: Encoder structure when $N = 2$.

When $N = 4$, the decomposition is $\mathbf{G}_4 = (\mathbf{I}_2 \otimes \mathbf{F}) \cdot \mathbf{R}_4 \cdot (\mathbf{I}_2 \otimes \mathbf{F})$. Using this, the encoding $\mathbf{c} = \mathbf{u}\mathbf{G}_4$ can be obtained by performing three operations in sequence: multiply by $(\mathbf{I}_2 \otimes \mathbf{F})$, then multiply by the reverse shuffle $\mathbf{R}_4$, and then multiply by $(\mathbf{I}_2 \otimes \mathbf{F})$ again. Here, we have $\mathbf{P}_4^{(1)} = \mathbf{R}_4$. The block diagram representation of these operations are shown in Fig. 7.2. Here it can be seen that the fundamental unit $\mathbf{F}$ appears four times in the encoder structure for $N = 4$.

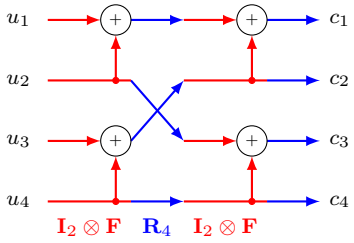


Figure 7.2: Encoder structure when $N = 4$.

When $N = 8$, the decomposition is:

$$\mathbf{G}_4 = (\mathbf{I}_4 \otimes \mathbf{F}) \cdot \mathbf{R}_8 \cdot (\mathbf{I}_4 \otimes \mathbf{F}) \cdot \mathbf{P}_8^{(2)} \cdot (\mathbf{I}_4 \otimes \mathbf{F}) \tag{7.29}$$

Using this, the encoding $\mathbf{c} = \mathbf{u}\mathbf{G}_8$ can be obtained by performing the above operations in sequence. The block diagram representation of these operations are shown in Fig. 7.3. Here, $\mathbf{P}_8^{(1)} = \mathbf{R}_8$. In the figure, the matrix $\mathbf{P}^{(2)}$ can be see to be two reverse shuffles $\mathbf{R}_4$.

Algorithm 7.1 $\mathbf{c} = \text{Enc}(\mathbf{u})$ Recursive Encoding of Polar Codes

Require: Information sequence $\mathbf{u}$ (including frozen bits) of length N.

Ensure: $\mathbf{c} = \mathbf{u}\mathbf{G}_N$

```
N' ← length of u
if N' = 2 then
    c_1 = u_1 + u_2
    c_2 = u_2
else
    r ← 1                                   ▷ r points to top half of reverse shuffle
    s ← N'/2 + 1                            ▷ s points to bottom half of reverse shuffle
    for i = 1, 2, 3, ..., N' do             ▷ Equivalent to t = u · (I_{N'/2} ⊗ F) · R_{N'}
        if i is odd then
            t_r = u_i + u_{i+1}
            r ← r + 1
        else if i is even then
            t_s = u_i
            s ← s + 1
        end if
    end for
    c_1^{N'/2}       = Enc(t_1^{N'/2})      ▷ Recursion
    c_{N'/2+1}^{N'} = Enc(t_{N'/2+1}^{N'})
end if
```

7.2.3 Recursive Encoding

Encoding can be obtained by computing $\mathbf{c} = \mathbf{u}\mathbf{G}_N$, but this requires storage of the $\mathbf{G}_N$ matrix. Instead, a recursive encoding exploits the structure of the polar code. The encoder function is represented by:

$$\mathbf{c} = \text{Enc}(\mathbf{u}) \tag{7.30}$$

A subset of elements a to b from a vector $\mathbf{t}$ are represented as $\mathbf{t}_a^b = (t_a, t_{a+1}, \ldots, t_{b-1}, t_b)$.

The function Enc performs the following steps with some $\mathbf{u}'$ with length N' as an input. If $N' = 2$ the encoding function is:

$$c_1 = u_1 + u_2 \tag{7.31}$$

$$c_2 = u_2 \tag{7.32}$$

and if $N' > 2$ then the encoding is performed recursively:

$$\begin{aligned} \mathbf{t} &= \mathbf{u} \cdot (\mathbf{I}_{N'/2} \otimes \mathbf{F}) \cdot \mathbf{R}_{N'} && \text{one step} \\ \mathbf{c}_1^{N'/2} &= \text{Enc}(\mathbf{t}_1^{N'/2}) && \text{encode upper bits} \\ \mathbf{c}_{N'/2+1}^{N'} &= \text{Enc}(\mathbf{t}_{N'/2+1}^{N'}) && \text{encode lower bits} \end{aligned}$$

which is equivalent to $\mathbf{c} = \mathbf{u}\mathbf{G}_N$. An algorithm implementing the recursive encoding is shown in Algorithm 7.1.

7.2.4 Reed-Muller Codes

Reed-Muller codes given in Section 3.4.5 can be seen as special type of polar code. The matrix $\mathbf{F}^{\otimes n}$ is related to Reed-Muller codes. Selecting certain rows of $\mathbf{F}^{\otimes n}$ gives the generator matrix for Reed-Muller codes. Recall an order-r Reed-Muller code $\mathrm{RM}(r, n)$ has block length 2^n. The generator matrix of an $\mathrm{RM}(r, n)$ code are the rows of of $\mathbf{G}$ with weight greater than or equal to 2^{n-r}. There are $\sum_{i=0}^{r} \binom{n}{i}$ such rows, so this is the dimension K of the code.

7.3 Channel Splitting and Decoding Polar Codess

7.3.1 Successive Cancelation Decoding and Channel Splitting

A codeword from a polar code $\mathbf{c} = (c_1, c_2, \ldots, c_N)$ is transmitted over a binary-input memoryless channel with input c and output y. Usually, the channel is denoted $\Pr(y|c)$, but in the context of polar codes the convention is to instead write $W(y|c)$; $\Pr(\cdot)$ and $W(\cdot)$ mean the same thing. Thus $\mathbf{c}$ is transmitted over N uses of the channel $W(y|c)$ and $\mathbf{y}$ is received.

The principle of successive cancelation decoding is introduced. In successive cancelation decoding of polar codes, the decoder makes a hard decision for bit u_i using all the previous estimates $\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_{i-1}$ as well as the entire received sequence $\mathbf{y} = y_1, \ldots, y_N$. Decoding begins by making an estimate $\widehat{u}_1 \in \{0, 1\}$ using $\mathbf{y}$. Then the decoder makes an estimate $\widehat{u}_2 \in \{0, 1\}$ using $\widehat{u}_1$ and $\mathbf{y}$. Successive cancelation decoder proceeds in this way until it estimates $\widehat{u}_N$ using $\widehat{u}_1, \ldots, \widehat{u}_{N-1}$ and $\mathbf{y}$. Frozen bits are known and do not need to be estimated.

Suppose $W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i)$ is known (or equivalently, $\Pr(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i)$ is known). The hard decision in position i is $\widehat{u}_i = 0$ if:

$$W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 0) > W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 1), \tag{7.33}$$

and otherwise the hard decision is $\widehat{u}_i = 1$. The decision is equivalent to:

$$\widehat{u}_i = \begin{cases} 0 & W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 0) > W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 1) \\ 1 & \text{otherwise} \end{cases} \tag{7.34}$$

In order to make hard decision, the two needed probabilities (or channels) are $W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 0)$ and $W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i = 1)$. These are found by *channel splitting*.

Channel splitting is a key idea in polar codes. In channel splitting, N copies of the channel W are transformed to N synthetic channels, some of which are better than than the original channel W and some of which are worse than W. A recursive channel transformation finds the synthetic channel $W(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i)$

from the physical channel $W(y|c)$; the N synthetic channels are:

$$W_N^{(i)}\big(\underbrace{\mathbf{y}, \mathbf{u}_1^{i-1}}_{\text{output}} | \underbrace{u_i}_{\text{input}}\big) \tag{7.35}$$

for $i = 1, 2, \ldots, N$. As with any channel described by a conditional probability distribution, variables to the left are "outputs" and the variable on the right are "inputs". Here u_1^{i-1} are also regarded as inputs. At step i of successive cancelation decoding, the decoder uses $u_1, u_2, \ldots, u_{i-1}$ and $\mathbf{y}$, to produce an estimate $\widehat{u}_i$. These are called synthetic channels because they do not correspond to physical channels, but indeed are conditional probability distributions.

For example with $N = 4$, the goal is to find the synthetic channels:

$$\begin{aligned}
W_4^{(1)} &= W_4^{(--)}(y_1^4|u_1) \\
W_4^{(2)} &= W_4^{(-+)}(y_1^4, u_1|u_2) \\
W_4^{(3)} &= W_4^{(+-)}(y_1^4, u_1^2|u_3) \\
W_4^{(4)} &= W_4^{(++)}(y_1^4, u_1^3|u_4)
\end{aligned}$$

Here, $+, -$ is used for a binary representation of integers. In what follows, both representations will be used.

If $W_N^{(1)}, W_N^{(2)}, \ldots, W_N^{(N)}$ can be obtained, then these can be used for decoding. The following proposition shows how to compute the synthetic channels recursively.

Proposition 7.2 (Arikan 2008, Proposition 3)**.** For any $n \geq 0$, $N = 2^n$ and $1 \leq i \leq N$,

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2}|u_{2i-1}) = \sum_{u_{2i} \in 0,1} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,\text{odd}}^{2i-2} + u_{1,\text{even}}^{2i-2}|u_{2i-1} + u_{2i}) W_N^{(i)}(y_{N+1}^{2N}, u_{1,\text{even}}^{2i-2}|u_{2i}) \tag{7.36}$$

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1}|u_{2i}) = \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,\text{odd}}^{2i-2} + u_{1,\text{even}}^{2i-2}|u_{2i-1} + u_{2i}) W_N^{(i)}(y_{N+1}^{2N}, u_{1,\text{even}}^{2i-2}|u_{2i}) \tag{7.37}$$

The two colored terms are the same, showing the similarity between the two equations. The proposition gives a recursive means to calculate the synthetic channels. The recursion proceeds until $W_1(y|u)$ is reached — this is the physical channel distribution, which is known.

7.3.2 Decoding $N = 2$ Polar Codes

Polar codes with $N = 2$ give the building block that is used for all other block lengths.

Consider decoding this code. This code is $N = 2$ it is not especially useful as an error-correcting code, but decoding is an important component. Information (u_1, u_2) is encoded to (c_1, c_2), which is transmitted over binary-input memoryless channel $\Pr(y|c)$ and received as (y_1, y_2). In polar code notation, W_1 is the channel itself, and we define $b(c)$ as:

$$b_1(c_1) = W_1^{(1)}(y_1|c_1) = \Pr(y_1|c_1) \tag{7.38}$$

$$b_2(c_2) = W_1^{(2)}(y_2|c_2) = \Pr(y_2|c_2) \tag{7.39}$$

Here we explicitly write c_i for the W_1 channels with the understanding that when we write $W_1^{(1)}(y_1|u_1)$ and $W_1^{(2)}(y_2|u_2)$ that u_1 and u_2 are the inputs to the W_1 channel and are not the same (u_1, u_2) that are being encoded. There are two synthetic channels $W_2^{(1)}(\mathbf{y}|u_1)$ and $W_2^{(2)}(\mathbf{y}, u_1|u_2)$.

Consider the two channels separately, first finding $b^-_{\text{out}}(u_1) = W_2^{(1)}(\mathbf{y}|u_1)$. This is called the upper or "$-$" channel. Here $b_1, b_2, b^-_{\text{out}}$ are probability domain messages. Applying channel splitting Proposition 7.2 gives the following:

$$b^-_{\text{out}}(0) = b_1(0)b_2(0) + b_1(0)b_2(0) \tag{7.40}$$

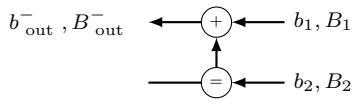
$$b^-_{\text{out}}(1) = b_1(1)b_2(0) + b_1(0)b_2(1) \tag{7.41}$$

It can be seen that this is the APP decoding rule for the single parity-check code from Subsection 4.1.4. As such, the log-domain version is:

$$B^-_{\text{out}} = 2\tanh^{-1}\left(\tanh(\frac{B_1}{2})\tanh(\frac{B_2}{2})\right). \tag{7.42}$$

Find the hard decision estimate $\widehat{u}_1$ by applying a threshold to b^-_{out} or B^-_{out}.

The decoder element can be represented graphically using the $N = 2$ encoder structure, as shown in the figure below.



In this image, the $\oplus$ node can be seen as a (square) check node in the LDPC factor graph and the circle-= node passes the b_2, B_2 values to the check node.

Next, find $b^+_{\text{out}}(u_2) = W_2^{(1)}(\mathbf{y}, u_1, |u_2)$, soft information about u_2. The estimate $\widehat{u}_1$ is used. This is called the lower branch or "+" channel. Again, applying channel splitting Proposition 7.2 gives:

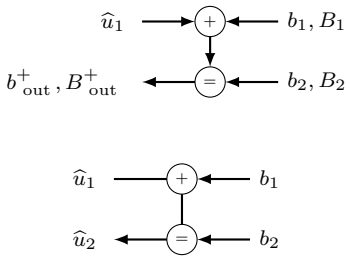
$$b^+_{\text{out}}(u_2) = \frac{b_1(\widehat{u}_1 + u_2)b_2(u_2)}{b_1(\widehat{u}_1 + 0)b_2(0) + b_1(\widehat{u}_1 + 1)b_2(1)} \tag{7.43}$$

If $\widehat{u}_1 = 0$ then this is the APP decoding rule for repeat code. If $\widehat{u}_1 = 1$, then the roles of $b_1(0)$ and $b_1(1)$ switch. The log domain version of this decoding rule is:

$$B^+_{\text{out}} = \begin{cases} B_1 + B_2 & \widehat{u}_1 = 0 \\ -B_1 + B_2 & \widehat{u}_1 = 1 \end{cases} \tag{7.44}$$

which can also be written as $B^+_{\text{out}} = (-1)^{\widehat{u}} B_1 + B_2$.

The decoder element can be represented graphically, on the same $N = 2$ encoder structure as the "$-$" branch:



Example 7.6. For an $N = 2$ polar code, decode $\widehat{u}_1$ then $\widehat{u}_2$ when the log domain input from the channel is $(B_1, B_2) = (1.1, -1.99)$.

For the upper "$-$" branch, find B^-_{out} using (7.42):

$$\begin{aligned} B^-_{\text{out}} &= 2\tanh^{-1}\left(\tanh(\frac{-1.1}{2})\tanh(\frac{1.99}{2})\right) \\ &= -0.800 \end{aligned}$$

The hard decision using B^-_{out} is $\widehat{u}_1 = 1$.

For the lower "$+$" branch, find B^+_{out} using (7.44):

$$\begin{aligned} B^+_{\text{out}} &= (-1)^1 1.1 + (-1.99) \\ &= -3.09 \end{aligned}$$

The hard decision using B^+_{out} is $\widehat{u}_2 = 1$.

SSQ 7.1. *N=2 Polar Code Decoding* For the $N = 2$ polar code with inputs $(B_1, B_2) = (2.244, -1.3)$ find B^-_{out}, $\widehat{u}_1$, B^+_{out} and $\widehat{u}_2$.

7.3.3 Decoding $N = 4$ Polar Codes

The recursive structure of polar codes become clear when considering $N = 4$

Consider Arikan's channel splitting for $N = 4$, giving four channels:

$$W_4^{(1)}(\mathbf{y}|u_1) = \frac{1}{2}W_2^{(1)}(y_1^2|u_1+u_2=0)W_2^{(1)}(y_3^4|u_2=0) + \frac{1}{2}W_2^{(1)}(y_1^2|u_1+u_2=0)W_2^{(1)}(y_3^4|u_2=0)$$

$$W_4^{(2)}(\mathbf{y}, u_1|u_2) = \frac{1}{2}W_2^{(1)}(y_1^2|u_1+u_2)W_2^{(1)}(y_3^4|u_2)$$

$$W_4^{(3)}(\mathbf{y}, u_1^2|u_3) = \frac{1}{2}W_2^{(1)}(y_1^2, u_1+u_2|u_3+0)W_2^{(1)}(y_3^4, u_2|u_4=0) + \frac{1}{2}W_2^{(2)}(y_1^2, u_1+u_2|u_3+1)W_2^{(2)}(y_3^4, u_2|u_4=1)$$

$$W_4^{(4)}(\mathbf{y}, u_1^3|u_4) = \frac{1}{2}W_2^{(2)}(y_1^2, u_1+u_2|u_3+u_4)W_2^{(2)}(y_3^4|u_4)$$

Here, it can be seen that the $W_4^{(i)}$ channels are obtained by transformations of the $W_2^{(i)}$ channels, as illustrated in the figure below:

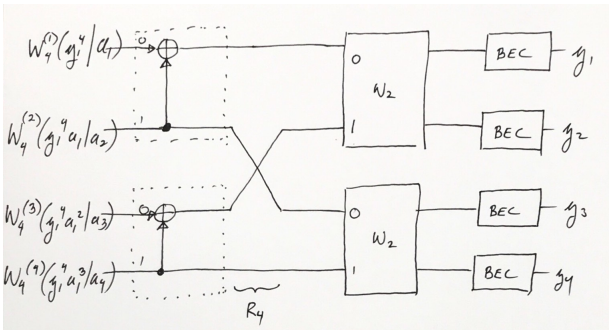


Fig. 7.5 shows the BER for $N = 8$ each channel of polar code on the BI-AWGN channel. Each channel assumes previous channels were decoded successfully. With respect to BER, bit position 1 is very bad (weight 1). Bit positions 2, 3, 5 are bad (weight 2). Bit positions 4, 6, 7 are good (weight 4) Bit position 8 is very good (weight 8). When designing a code, use the best bit positions. Resulting BER suffers from error propagation under successive cancellation decoding.

Arikan's proposition realizes a recursive decoder, with a Matlab implementation shown in Fig. 7.4 This implementation is simple but slow. Recursion has no memory, so many calculations are repeated. Examples of usage

```
>> B = [1.1, -1.99];
>> W(B,[])  %upper branch of N=2 example

ans =

    -0.8004

>> W(B,[1])  %lower branch, uhat=1

ans =

    -3.0900

```

```
function out = W(L,u)

if length(L) == 1
    %physical channel
    out = L;
else
    i   = fix(length(u) / 2) + 1; %get i from length of u
    branch = mod(length(u),2);
    s   = 1:2:2*i-2;
    t   = 2:2:2*i-2;
    N = length(L);                  %N is 2N, as in Arikan
    Na = 1:N/2;
    Nb = (N/2 + 1):N;

    if branch == 0   %upper branch
        L1  = W(L(Na), mod(u(s)+u(t),2) );
        L2  = W(L(Nb), u(t) );
        out = 2*atanh( tanh(L1/2) * tanh(L2/2) );
    else             %lower branch
        L1  = W(L(Na), mod(u(s)+u(t),2) );
        L2  = W(L(Nb), u(t) );
        out = (-1).^u(end) * L1 + L2;
    end
end
```

Figure 7.4: Matlab source code for recursive log-domain decoder

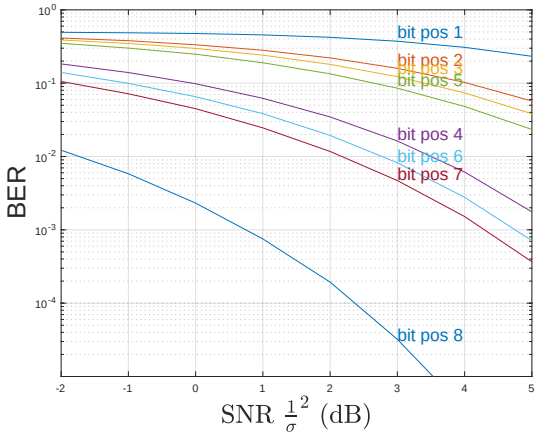


Figure 7.5: BER for $N = 8$ each channel of polar code. Each channel assumes previous channels were decoded successfully.

```
>> W(B,[0 0 0 1 0 1])  %i=7 position of N=8 example

ans =

    9.3673
```

7.4 Channel Splitting for the BEC

This subsection considers the binary erasure channel (BEC), for which computation of the synthetic channels is straightforward.

Consider a BEC with input u, output y and erasure probability ϵ. We can write a decoding table for estimating the input $\widehat{u}$ given the output y:

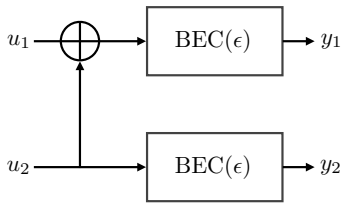
y	$\widehat{u}$	Pr
u	u	$1-\epsilon$
?	?	ϵ

Decoding for the BEC is simple. If the channel output y is u, then the estimate is u. If the channel output is ?, then the estimate is ?. The probability of these two events is also shown. This is an $N = 1$ polar code.

For an $N = 2$ polar code, consider the two synthetic channels $W_2^{(0)}$ and $W_2^{(1)}$ ($W_2^{(-)}$ and $W_2^{(+)}$),

$$W_2^{(-)} = W(y_1, y_2|u_1) \tag{7.45}$$
$$W_2^{(+)} = W(y_1, y_2, u_1|u_2) \tag{7.46}$$

Above, $W_2^{(-)}$ is the same as $W_2^{(1)}$ and $W_2^{(+)}$ is the same as $W_2^{(2)}$. In what follows, both representations will be used.



First consider the upper branch or "-" branch, decoding u_1 using y_1, y_2. The decoding table is:

y_1	y_2	$\widehat{u}_1$	Pr
u_1+u_2	u_2	u_1	$(1-\epsilon)^2$
?	u_2	?	$\epsilon(1-\epsilon)$
u_1+u_2	?	?	$\epsilon(1-\epsilon)$
?	?	?	ϵ^2

For example, if y_1 and y_2 are not both erased, then we have $u_1 + u_2$ and u_2, from which we can compute u_1 (line 1 of the table). On the other hand, if either y_1 or y_2 are erased, then the estimate u_1 is erased — the probability that u_1 is erased is $\epsilon(1-\epsilon) + \epsilon(1-\epsilon) + \epsilon^2 = 2\epsilon - \epsilon^2$.

Second, consider the lower branch or "+" branch, decode u_2 using u_1, y_1, y_2. Since u_1 is known and correct, the decoding table is:

y_1	y_2	u_1	$\widehat{u}_2$	Pr
$u_1 + u_2$	u_2	u_1	u_2	$(1-\epsilon)^2$
?	u_2	u_1	u_2	$\epsilon(1-\epsilon)$
$u_1 + u_2$	?	u_1	u_2	$\epsilon(1-\epsilon)$
?	?	u_1	?	ϵ^2

For example, if y_1 is not erased and y_2 is erased, u_2 can be found, since $y_1 = u_1 + u_2$ and u_1 is known (line 3 of the table). is On the BEC with erasure probability ϵ, the probability that $\widehat{u}_2$ is erased is ϵ^2. The probability that u_2 is erased is ϵ^2.

Combining the upper and lower branch probability of error, the probability of erasure for $N = 2$ is expressed using the function $T^{(s)}$:

$$T^{(s)}(\epsilon) = \begin{cases} 2\epsilon - \epsilon^2 & s = - \\ \epsilon^2 & s = + \end{cases} \tag{7.47}$$

where $s = -$ is the upper branch and $s = +$ is the lower branch.

Now for $N = 4$, we can apply the polar transform of the proposition but it is helpful to write recursively, also as shown in the figure below:

$W_4^{(--)}(y_1^4 \vert u_1)$	Obtained by the $-$ transformation on $W_2^{(-)}$
$W_4^{(-+)}(y_1^4, u_1 \vert u_2)$	Obtained by the $+$ transformation on $W_2^{(-)}$
$W_4^{(+-)}(y_1^4, u_1^2 \vert u_3)$	Obtained by the $-$ transformation on $W_2^{(+)}$
$W_4^{(++)}(y_1^4, u_1^3 \vert u_4)$	Obtained by the $+$ transformation on $W_2^{(+)}$

The four channels are the same as $W_4^{(1)}, W_4^{(2)}, W_4^{(3)}, W_4^{(4)}$. The $+/-$ is a binary string with the LSB on the left.

Now it is possible to find the probability of erasure for $N = 4$ using the lower level. If p is the probability of erasure at level W_2, then the probability of erasure for $N = 4$ is:

$$T^{(s)}(p) = \begin{cases} 2p - p^2 & s = - \\ p^2 & s = + \end{cases} \tag{7.48}$$

This is valid because the probability of erasure for both copies of W_2 are the same. For $N = 8, 16$, etc. on the BEC, the recursion is the same, where p represents the probability of erasure one level lower.

Let $\mathbf{s} = [s_1, s_2, \ldots, s_n]$ be the binary representation of bit position i (indexed from one), so that $1 = ----$, $2 = +---$, $3 = -+--$, $\ldots$, $16 = ++++$.

The probability of erasure for position s on BEC(ϵ) is denoted by $S^{(\mathbf{s})}(\epsilon)$ and can be found recursively as:

$$S^{(\mathbf{s})}(\epsilon) = T^{(s_n)}\left(T^{(s_{n-1})}\left(\cdots\ T^{(s_1)}\left(\epsilon\right)\right)\right). \tag{7.49}$$

For the decoder to succeed, all bits in $\mathcal{A}$ must be decoded correctly. The probability of decoding error under successive cancellation decoding, or the word error rate (WER) can be found analytically as one minus the probability that all synthetic channels decode successfully:

$$\text{WER} = 1 - \prod_{i\in\mathcal{A}}(1 - S^{(i)}(\epsilon)) \tag{7.50}$$

Example 7.7. Consider an $N = 8$ polar code on the BEC with $\epsilon = 0.1$. Evaluate $S^{(1)}(\epsilon)$ and $S^{(7)}(\epsilon)$. For $S^{(1)}(\epsilon)$, evaluate numerically at each recursion step because the analytic form is not simple:

$$\begin{aligned}
S^{(1)} = S^{(---)} &= T^{(-)}\left(T^{(-)}\left(T^{(-)}\left(\epsilon\right)\right)\right) \\
&= T^{(-)}\left(T^{(-)}\left(2\epsilon - \epsilon^2\right)\right) \\
&= T^{(-)}\left(T^{(-)}\left(0.19\right)\right) \\
&= T^{(-)}\left(2\cdot 0.19 - 0.19^2\right) \\
&= T^{(-)}\left(0.3439\right) \\
&= 0.5695
\end{aligned}$$

For $S^{(7)}(\epsilon)$, the analytical form is simple:

$$\begin{aligned}
S^{(7)} = S^{(-++)} &= T^{(-)}\left(T^{(+)}\left(T^{(+)}\left(\epsilon\right)\right)\right) \\
&= T^{(-)}\left(T^{(+)}\left(\epsilon^2\right)\right) \\
&= T^{(-)}\left(\epsilon^4\right) \\
&= 2\epsilon^4 - \epsilon^8 \\
&= 0.0002
\end{aligned}$$

Note that bit position 1 has a higher probability of erasure than ϵ, while bit position 7 has a lower probability of erasure.

7.5 Polarization Phenomenon and Code Design

7.5.1 Polarization

The probability of erasure for the synthetic channels $S^{(i)}$ can be found for $i = 1, 2, \ldots, N$. These can be sorted from smallest to largest, which is shown in Fig. 7.6 for $N = 8, 16, 256, 16384$. As N increases, clear pattern emerges — some

bit positions have probability of erasure close to 0, while other bit positions have probability of erasure close to 1. For moderate values such as $N = 256$, there is a transition region as well. But as N increases, a *polarization* effect occurs – the transition region becomes sharp, and most bits are either highly reliable (probability of erasure close to 0) or are highly unreliable (probability of erasure close to 1). This interesting phenomenon gives polar codes their names.

7.5.2 Polar Code Design

Polar code design means selecting the set of information bits $\mathcal{A}$, for a given block length N and dimension K. This suggests a code design technique — to design a polar code, use the K most reliable synthetic channels to transmit information. For the remaining $N - K$ channels, freeze the information bits to 0. The question is how to find the information set. While various techniques exist, this section describes the polarization weight method for selecting the information set.

Erasure Channel Density Evolution The first method uses density evolution for the erasure channel.

Example 7.8. Design a polar code with $N = 8$ and $K = 4$, continuing Example 7.7. The probability of erasure for all $N = 8$ synthetic channels are:

$$\begin{aligned} S^{(1)} &= 0.569533 & S^{(5)} &= 0.039404 \\ S^{(2)} &= 0.118267 & S^{(6)} &= 0.00039601 \\ S^{(3)} &= 0.0708968 & S^{(7)} &= 0.0002 \\ S^{(4)} &= 0.00130321 & S^{(8)} &= 10^{-8} \end{aligned}$$

The $K = 4$ best channels are 4 and 6 to 8, so we choose the information set to be $\mathcal{A} = \{4, 6, 7, 8\}$.

Polarization Weight Another method selects the information bits according to a metric called the polarization weight. The channels (corresponding to rows of $\mathbf{B}_N\mathbf{F}^{\otimes n}$) are indexed from 1 to N. For index i, write the binary expansion of $i - 1$ as $(d_{n-1}, \ldots, d_1, d_0)$, where d_{n-1} is the most significant bit. For example index 4 has binary expansion $(0, 1, 1)$ when $n = 3$. For some constant β, the *polarization weight* of i is w_i:

$$w_i = \sum_{j=0}^{n-1} d_j \beta^j, \tag{7.51}$$

which is the β expansion, and can be seen as a generalization of the binary expansion. The choice of β influences the resulting polar code design; $\beta = 2^{1/4}$ is suggested. The information set consists of the indices with the K greatest polarization weights.

Example 7.9. Design an $N = 8, K = 4$ polar code using $\beta = 2^{1/4}$. First

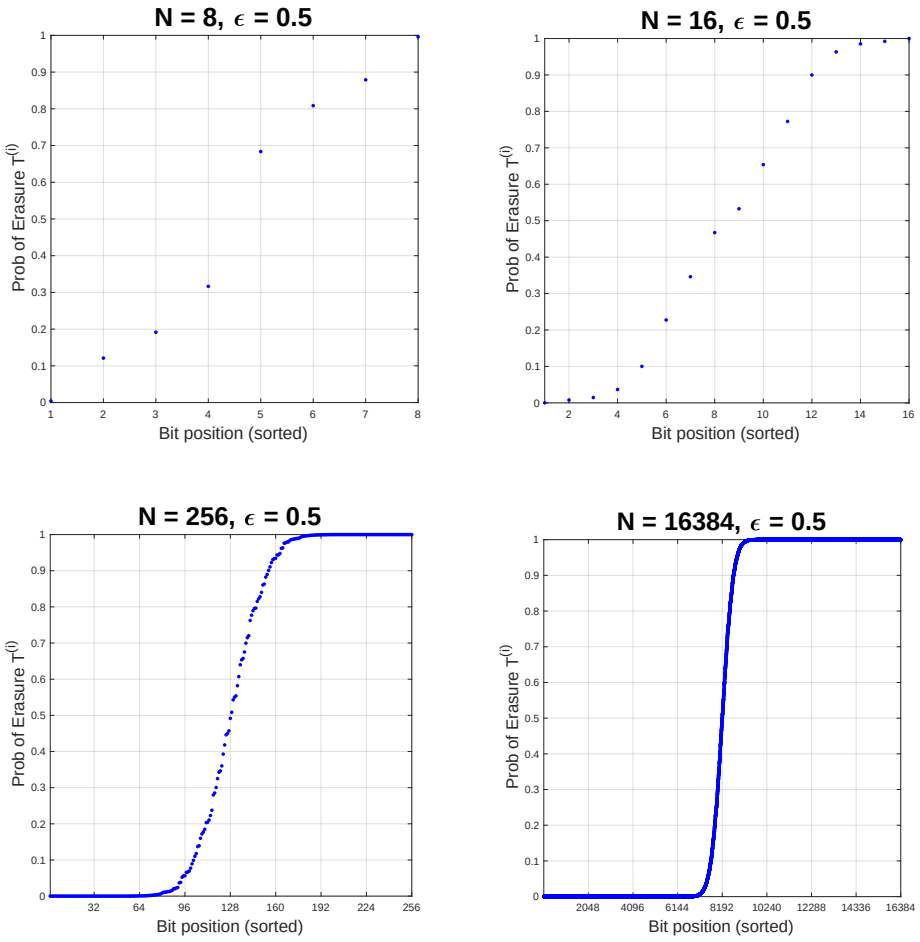


Figure 7.6: Probability of erasure for each bit position. The bit positions have been sorted according to increasing probability of erasure. For $N = 8, 16, 256, 16384$.

consider the index 4 with binary expansion $(0, 1, 1)$, for which the polarization weight w_4 is:

$$w_4 = 0 \cdot 2^{2/4} + 1 \cdot 2^{1/4} + 1 \cdot 2^{0/4} = 2.189$$

The β expansion for all indices 1 to 8 are:

$$\mathbf{w} = [0, 1, 1.18921, \mathbf{2.18921}, 1.41421, \mathbf{2.41421}, \mathbf{2.60342}, \mathbf{3.60342}]$$

The $K = 4$ largest values in bold correspond to indices of the information set, that is $\mathcal{I} = \{4, 6, 7, 8\}$. For this case, polarization weight gives the same design as the erasure channel density evolution in Example 7.8. However, the two techniques give different designs in general.

7.6 Non-Recursive Successive Cancelation Decoding

Using the recursive decoder can be inefficient in computers because function calls in software often have a large overhead. There is also a lot of redundancy in those function calls, when decoding all the bits of a codeword. Use the block diagram for decoding. Note that if the SC decoder makes an error in decoding position i, this error will propagate forward, affecting following decisions.

For each information bit u_i for $i \in \mathcal{A}$:

1. Forward stage: moving left-to-right, compute all possible bits. Use all frozen bits and hard decisions so far.
2. Backward stage: moving right-to-left, compute all possible log values
3. For bit u_i, make a hard decision using its log value:

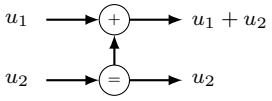
$$\widehat{u}_i = \begin{cases} 0 & \beta_i^{(1)} \geq 0 \\ 1 & \beta_i^{(1)} < 1 \end{cases} \tag{7.52}$$

7.6.1 Decoder Element

The decoding algorithm uses a decoding element, and first "local" operations using this decoding algorithm are discussed. There are two stages: (1) forward computation of hard decisions (2) backward computation of soft values. Each stage has an upper branch and a lower branch computation.

The forward stage is a hard-input, hard-output. Inputs on the left are used to compute outputs on the right.

Forward Stage For the upper branch, if $u_1, u_2 \in \{0, 1\}$ are available at the left-hand side input, then the right-hand side upper output is $u_1 + u_2$. If both u_1 and u_2 are not available, then there is no lower output. For the lower branch, if $u_2 \in \{0, 1\}$ is available at the input, then output is u_2. If u_2 is not available, then there is no lower output.



The backward stage is a soft-input, soft-output. Inputs on the right are used to compute outputs on the left. Consider the basic element of the polar code which represents $(c_1, c_2) = (u_1 + u_2, u_2)$. There are two inputs on the right, and two outputs on the left. In addition, when calculating the lower branch there may be a hard decision on u_1 available. The soft messages are in the log domain.

$$L = \log \frac{\Pr[x = 0]}{\Pr[x = 1]} \tag{7.53}$$

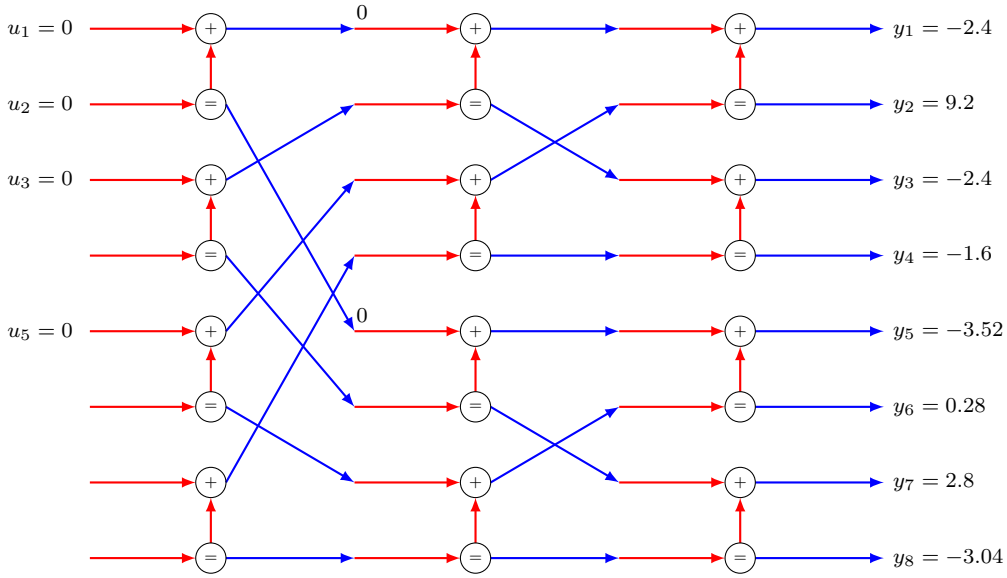


Figure 7.7: Decoding $N = 8$ polar code, information bit position 4, forward part only.

7.6.2 Example

Consider an $N = 8$ polar code with information bits $\mathcal{A} = \{4, 6, 7, 8\}$. The received log-domain messages are:

$$[-2.4, 9.2, -2.4, -1.6, -3.52, 0.28, 2.8, -3.04]; \tag{7.54}$$

Examples of decoding for each bit are illustrated on the following figures.

7.7 References

1. E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, Vol. 55, No. 7, July 2009

2. Alexios Balatsoukas-Stimming, "Efficient Decoding of Polar Codes - Algorithms and Implementations," 2020 European School on Information Theory, video available at https://vimeo.com/showcase/7823297/video/482236401. Use password #StayHomeStaySafe

3. H. D. Pfister, "A brief introduction to polar codes," Lecture notes, Duke University, October 8, 2017. Available at http://pfister.ee.duke.edu/courses/ecen655/polar.pdf

4. S. Aizaz A. Shah, "Polar codes," Hamburg University of Technology.

5. Successive Cancellation(SC) Decoder for Polar Codes: Illustration of its Building Blocks with N=2,4, YouTube https://www.youtube.com/watch?v=wK2KI2LtdQI

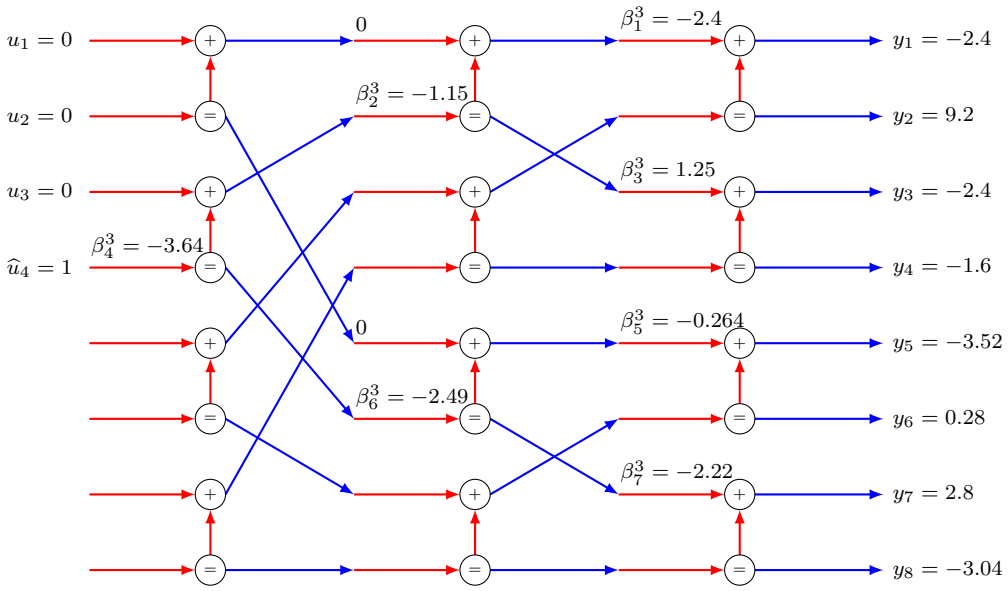


Figure 7.8: Decoding $N = 8$ polar code, information bit position 4, obtaining $\widehat{u}_4$.

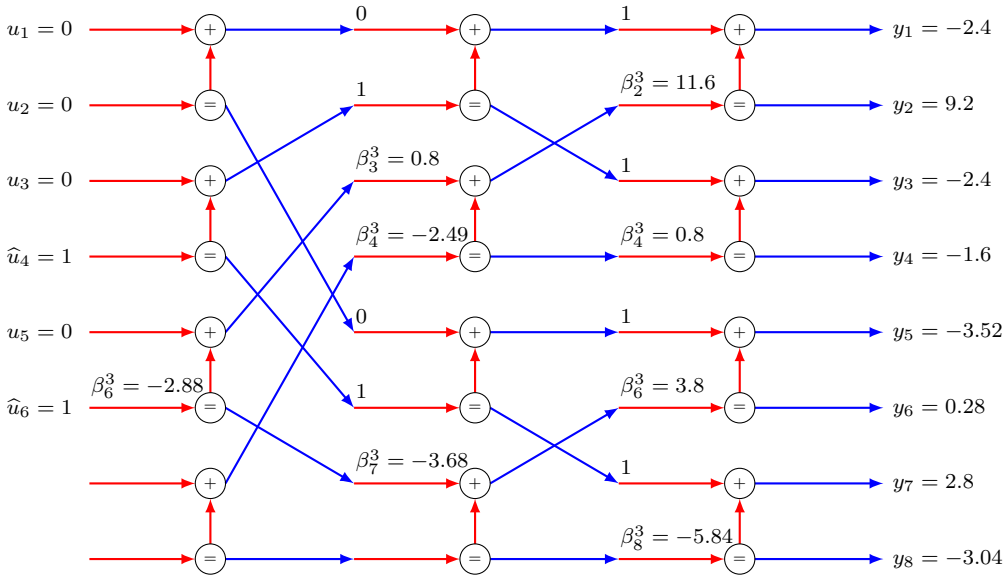


Figure 7.9: Decoding $N = 8$ polar code, information bit position 6, obtaining $\widehat{u}_6$.

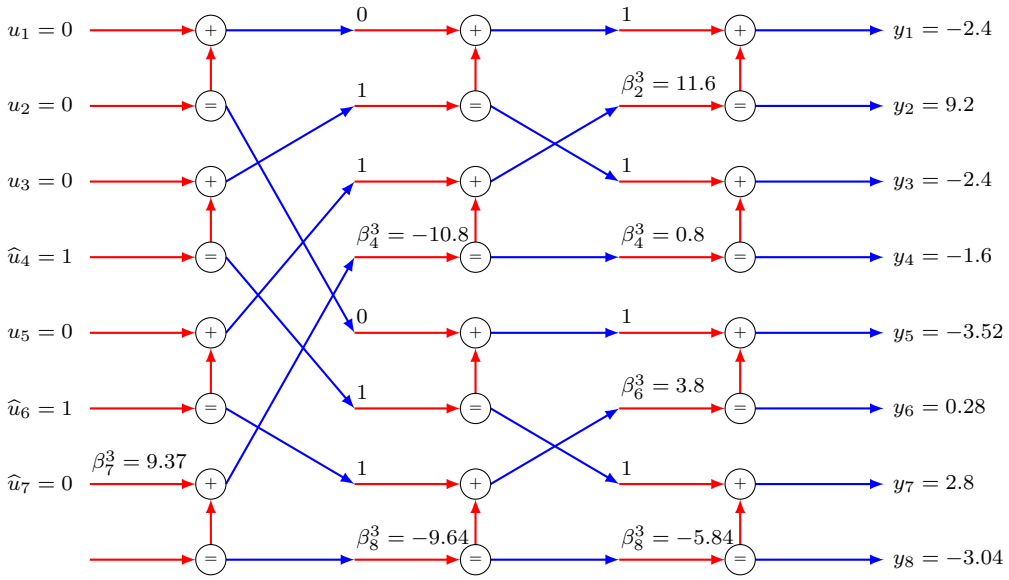


Figure 7.10: Decoding $N = 8$ polar code, information bit position 7, obtaining $\widehat{u}_7$.

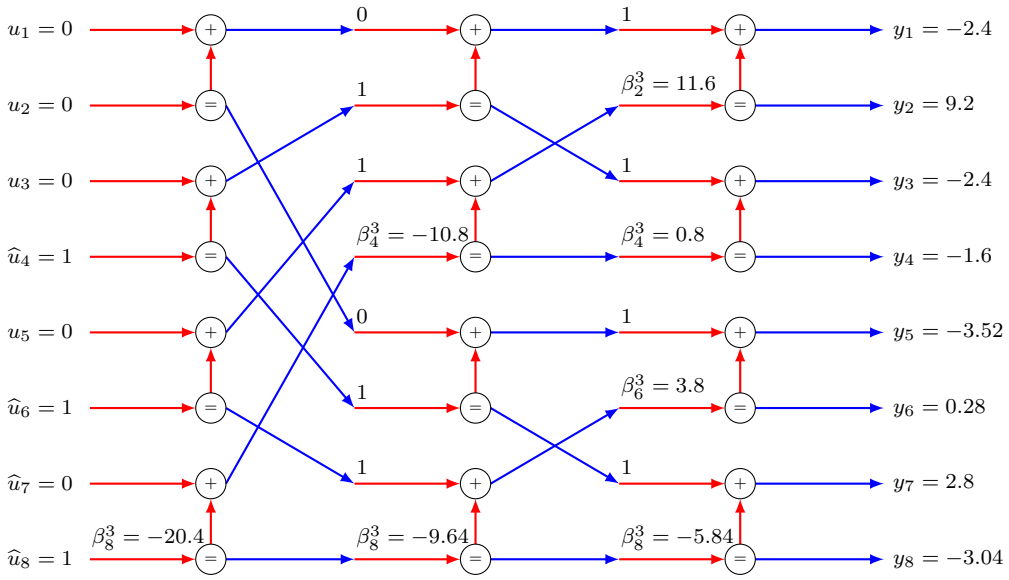


Figure 7.11: Decoding $N = 8$ polar code, information bit position 8, finally obtaining $\widehat{u}_8$.

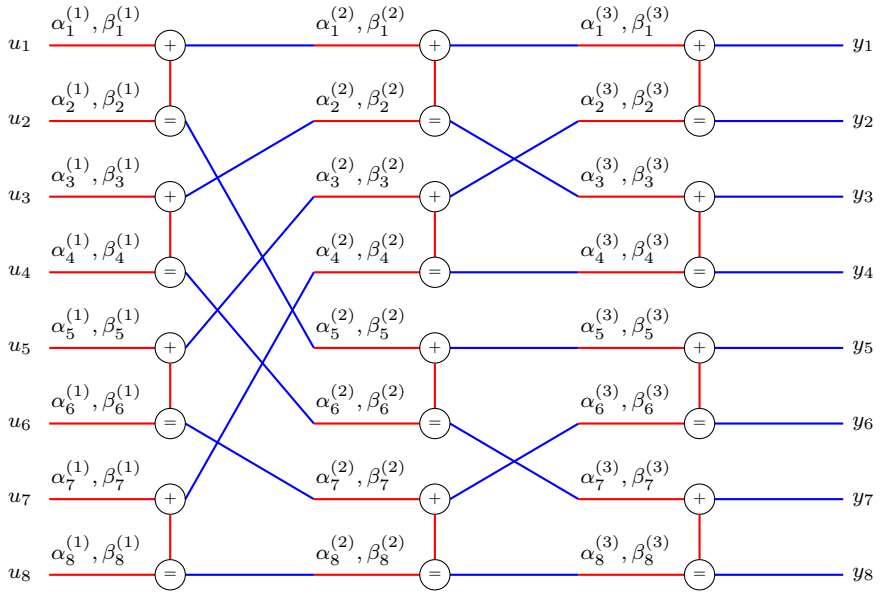


Figure 7.12: Definition of messages for decoding $N = 8$ polar code.

6. A. Montanari, "Polar codes," Lecture Notes, Stanford University, May 2018, https://web.stanford.edu/class/ee388/HOMEWORK2018/lecture-9-10-11.pdf

7. Kai Niu, Kai Chen, Jiaru Lin and Q. T. Zhang, "Polar Codes: Primary Concepts and Practical Decoding Algorithms," *IEEE Communications Magazine*, July 2014. IEEE Xplore Link

7.8 Exercises

7.1 Let $\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Over the field $\mathbb{F}_2$, prove that:

(a) $\left(\mathbf{F}^{\otimes n}\right)^2 = \mathbf{I}_{2^n}$, the identity matrix.
(b) $\left(\mathbf{F}^{\otimes n}\right)^{-1} = \mathbf{F}^{\otimes n}$. This matrix is its own inverse.

7.2 Find the $N = 16$ bit-reversal shuffle $\mathbf{B}_{16}$. Write your answer in the form of (6.14).

7.3 Show the $N = 4$ matrix decomposition for $\mathbf{G}_4$.

(a) What is the relationship between $\mathbf{B}_4$ and $\mathbf{R}_4$?
(b) The matrix decomposition for $N = 4$ is $\mathbf{G}_4 = (\mathbf{I}_2 \otimes \mathbf{F}) \cdot \mathbf{R}_4 \cdot (\mathbf{I}_2 \otimes \mathbf{F})$. Using the perfect shuffle to commute the Kronecker product, equation (10.2) , show that:

$$\mathbf{G}_4 = \mathbf{B}_4 \mathbf{F}^{\otimes 2} \tag{7.56}$$

7.4 (a) Find the probabilities of erasure decoding $S^{(i)}$, $i = 1, 2, \ldots, 8$ for an $N = 8$ polar code on the BEC with $\epsilon = 0.4$. (b) Give the 3×8 generator matrix for an $N = 8$, $K = 3$ polar code for use on this channel. (c) What is the word error rate (WER)?

7.5 (a) Design a (16,5) polar code by taking the 5 rows of $\mathbf{F}^{\otimes 4}$ that have weight 8 or greater. What is the corresponding information set $\mathcal{A}$? (b) Simulate this code on the BI-AWGN channel using successive cancellation decoding, plotting BER and WER vs. E_b/N_0. Show the uncoded performance on the same graph.

7.6 Show that channel splitting for $N = 2$ and $i = 1$ gives APP decoding of the single-parity check code of length 3. (Use Arikan's proposition with $N = 1$).

Chapter 8

Lattices

8.1 Lattices

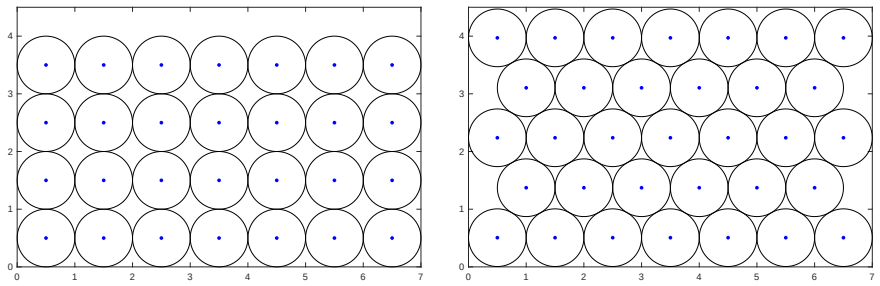
Suppose you want to put oranges in a box, filling the box with as many oranges as possible. Assuming the oranges are ideal spheres and all the same size, there are many ways to place the spherical oranges in the box. What arrangement allows us to put as many spherical oranges into the box as possible? Clearly some arrangements are better than others. There will always be some space remaining between the spherical oranges, but the less this space, the more spheres can be put into the box.

Consider a two-dimensional version of this problem, placing circles of diameter 1 into a 7-by-4.5 rectangle, as shown in Fig. 8.1. The regular packing in Fig. 8.1-(a) places 28 circles inside the box. The gaps between the circles are unused space. Fig. 8.1-(b) places 33 circles in the same space, with fewer gaps between the circles, making it more efficient. This later way is called "hexagonal packing" which will be explained later. Note the boundary effects, unused space at the sides of the box. If the box is very large, then these boundary effects are not significant, and we are primarily interested in reducing the space between the circles.

This problem of packing spheres motivates the study of lattices. The sphere packing problem asks how to arrange spheres of equal size in order to maximize density of the spheres. Each sphere may be represented by its center, so we can talk about a collection of centers rather than spheres. Lattices are one possible way to represent centers, and not all sphere packings can be represented by a lattice. We will concentrate on lattices only, because many good sphere packings are lattices, and lattices have numerous interesting properties.

8.1.1 Definition

Lattices are usually defined in the n-dimensional Euclidean space $\mathbb{R}^n$. Usual vector addition is used, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, then $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n)$.



(a) (b)

Figure 8.1: Circles can be packed into a 7-by-4.5 rectangle using (a) a less efficient pattern which has larger space between the circles or (b) a more efficient hexagonal pattern which has smaller space between the circles. Boundary effects are not usually of concern, and inefficiencies at the boundary can be ignored.

Definition 8.1. An n-dimensional *lattice* Λ is a discrete additive subgroup of $\mathbb{R}^n$.

Importantly, a lattice Λ is group under vector addition. Accordingly, a lattice has the following properties.

1. The zero point is a lattice point.
2. If $\mathbf{x}, \mathbf{y}$ are lattice points, then $\mathbf{x} + \mathbf{y}$ is a lattice point. The linearity property can be stated: if you could stand at any lattice point $\mathbf{x}$, and look around, you would see the same thing, if you were standing at any other point in the lattice.
3. $\mathbf{x} + \mathbf{x} + \cdots + \mathbf{x}$ is a lattice point
4. If $\mathbf{x}$ is a lattice point and a is an integer, then $a \cdot \mathbf{x}$ is also a lattice point.
5. A lattice is an infinite structure.

Note also that since $\mathbb{R}^n$ is a vector space, and Λ is a subgroup, Λ is also vector subspace. Lattices exist for any dimension $n \geq 1$. Lattices of dimension $n = 1, 2\ 3$ can be easily drawn to illustrate lattice properties. Lattices of higher dimension $n = 4, 5, 6, \ldots$ exist, but cannot be easily drawn.

Example 8.1. Using the group properties above, build an $n = 2$ dimension lattice. By the first property $(0, 0)$ is a lattice point. Now choose a non-zero lattice point, say $(3, 1)$ for example. From this, the second property states that $(6, 2), (9, 3), \ldots$ as well as $(-3, -1), (-6, -2), (-9, -3), \ldots$ must be lattice points as well. Now choose another non-zero lattice point, say $(0, 2)$. Again from the second property, $\ldots, (0, -4), (0, -2), (0, 4), (0, 6), \ldots$ are lattice points as well.

From the third property, the sum of any two points is also a lattice point. For example, $(-3, -3)$ is a lattice point, as it is the sum $(-3, -1) + (0, 4)$. Fig. 8.2

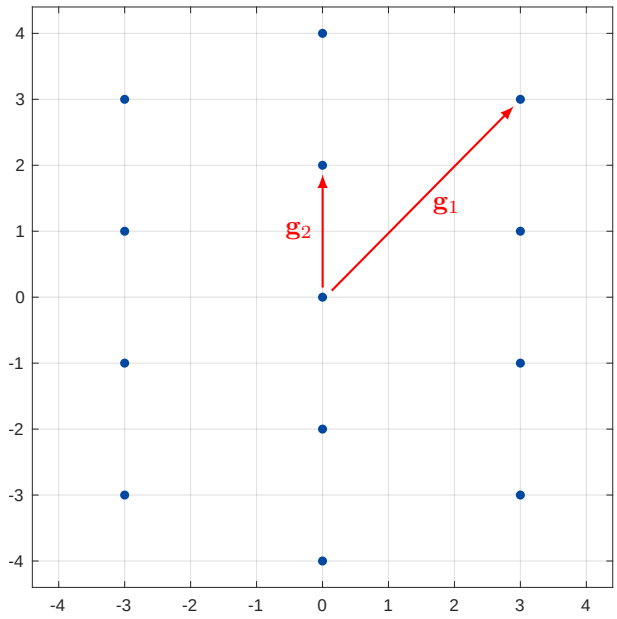


Figure 8.2: Example of a $n = 2$ dimensional lattice.

shows part of the resulting lattice; it is not possible to show the entire infinite lattice structure.

8.1.2 Generator Matrix

Since a lattice Λ is an n-dimensional subspace of $\mathbb{R}^n$, this space can be spanned by a set of n linearly independent basis vectors $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$, where,

$$\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \tag{8.1}$$

is a column vector representing a point in $\mathbb{R}^n$.

A vector $\mathbf{x}$ is a lattice point if it can be formed as a linear combination of the basis vectors scaled by integers $b_i \in \mathbb{Z}$:

$$\mathbf{x} = \mathbf{g}_1 b_1 + \mathbf{g}_2 b_2 + \cdots + \mathbf{g}_n b_n. \tag{8.2}$$

The matrix-form representation[1] is:

$$\mathbf{x} = \mathbf{G}\mathbf{b}, \tag{8.3}$$

[1]Many authors instead write the generator vectors in rows. When reading a paper on lattices, you should ask "are the generator vectors written in rows or in columns?" This text uses the column convention.

where $\mathbf{G}$ is an n-by-n matrix called a *generator matrix*:

$$\mathbf{G} = \begin{bmatrix} | & | & & | \\ \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_n \\ | & | & & | \end{bmatrix}. \tag{8.4}$$

and the integer vector $\mathbf{b}$ is:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \tag{8.5}$$

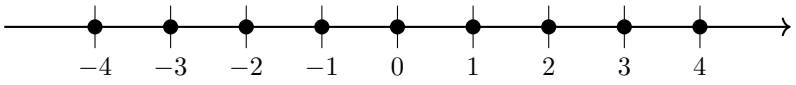
Throughout, $\mathbf{b}$ is used to denote an n-dimensional vector of integers, that is $\mathbf{b} \in \mathbb{Z}^n$. Since the generator vectors $\mathbf{g}_1, \ldots, \mathbf{g}_n$ are linearly independent, the generator matrix $\mathbf{G}$ is full-rank.

The lattice Λ generated by a matrix $\mathbf{G}$ is the set of all integer, linear combinations of the generator vectors. The lattice can be expressed using set builder notation:

$$\Lambda = \{\mathbf{G}\mathbf{b} \mid \mathbf{b} \in \mathbb{Z}^n\}. \tag{8.6}$$

Since there is an infinite number of integers $\mathbf{b}$, the size of Λ is infinite, that is there is an infinite number of lattice points. An *integral lattice* Λ satisfies $\Lambda \subset \mathbb{Z}^n$, that is, all the lattice points are integers. For such a lattice, the generator matrix $\mathbf{G}$ has only integers.

Example 8.2. For dimension $n = 1$, the set of integers scaled by $a\mathbb{Z}$ for $a \neq 0$ form a lattice. The generator is a, so that the lattice points are $x = ab$ for $b \in \mathbb{Z}$. The lattice with $a = 1$ is pictured below.



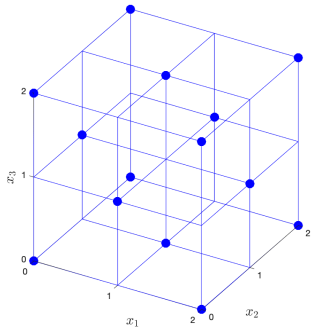
The $n = 1$ integer lattice is a special case of the n-dimensional integer lattice $\mathbb{Z}^n$, see Subsection **??**.

Example 8.3. For the lattice in Example 8.1, two possible basis vectors are $\mathbf{g}_1 = [3, 1]^{\mathrm{t}}$ and $\mathbf{g}_2 = [0, 2]^{\mathrm{t}}$ giving the generator matrix:

$$\mathbf{G} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}. \tag{8.7}$$

This is an integral lattice.

(a)



(b)

Figure 8.3: (a) Stacking of cannonballs (Rye Castle, East Sussex, England) (b) An example of an $n = 3$ lattice. Image credit: Wikipedia/Funkdooby

---

Example 8.4. In $n = 3$ dimensions, stacking cannonballs as shown in Fig. 8.3-(a) is a packing of spheres and is equivalent to putting oranges in a box because both oranges and cannonballs are supposed to be spherical. The packing of cannonballs as shown in Fig. 8.3-(a) can be presented by their centers, as in the 2-dimensional case.

An example of an $n = 3$ dimension lattice is shown in Fig. 8.3-(b) and has a generator matrix given by

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

This lattice is geometrically equivalent to stacking of cannonballs. This is an integral lattice.

Lattice exist in higher dimensions $n = 4, 5, 6, \ldots$ as well. A full-rank n-by-n matrix $\mathbf{G}$ is the generator matrix for an n-dimension lattice. For example:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{8.8}$$

is the generator matrix for an $n = 4$ dimensional lattice. However, lattices in dimension $n > 3$ are difficult to illustrate.

Generally, the n vectors $\mathbf{g}_1, \ldots, \mathbf{g}_n$ are linearly independent, so $\mathbf{G}$ is full rank. However, Definition 8.1 can be generalized by allowing the n dimensional lattice to be in an m-dimensional Euclidian space $\mathbb{R}^m$ with $m \geq n$. In this case, the lattice is an n-dimensional discrete additive subgroup of $\mathbb{R}^m$. For example, the

following generator matrix has dimension $n = 3$ but is in an $m = 5$ Euclidean space:

$$\begin{bmatrix} 6 & 0 & 4 \\ 0 & 4 & 2 \\ 1 & 5 & 0 \\ 0 & 6 & 0 \\ 5 & 0 & 2 \end{bmatrix}. \tag{8.9}$$

Generator matrices for finite-field error-correcting codes usually have more rows than columns (with generator vectors in columns). A code with a square generator matrix of full rank would have minimum Hamming distance of 1 — this is not useful for error correction. On the other hand, even though most lattices have square generator matrix, they are useful for correcting errors. The distinction is because lattices are defined over the real numbers, while error-correcting codes are defined over a finite field. For communications, the additional dimensions does not generally improve the error-correction capability, and we deal primarily with lattices with n-by-n generator matrices of full rank.

A sometimes convenient representation of a lattice is through its *Gram matrix* $\mathbf{A}$, given by $\mathbf{A} = \mathbf{G}^{\mathrm{t}}\mathbf{G}$. The Gram matrix has size n-by-n, even if the generator matrix is m-by-n with $m \geq n$. If $\mathbf{x} = \mathbf{G}\mathbf{b}$, then $||\mathbf{x}||^2 = \mathbf{b}^{\mathrm{t}}\mathbf{A}\mathbf{b}$. An alternative definition of integral lattice requires only that $\mathbf{x}^{\mathrm{t}}\mathbf{x}$ are integers for all $\mathbf{x} \in \Lambda$, or equivalently the Gram matrix $\mathbf{A}$ is integral. This text uses the earlier and more restrictive definition of integral lattice.

8.1.3 Check Matrices and Dual Lattices

Since we deal primarily with lattices with an n-by-n full rank matrix $\mathbf{G}$, this matrix is invertible. The inverse is called the *check matrix*, by analogy with the parity-check matrix of finite-field codes.

Definition 8.2. For a lattice Λ with generator matrix $\mathbf{G}$, the *check matrix* $\mathbf{H}$ is:

$$\mathbf{H} = \mathbf{G}^{-1}. \tag{8.10}$$

A lattice point $\mathbf{x}$ is $\mathbf{G}\mathbf{b}$, so $\mathbf{H}\mathbf{x}$ is a vector of integers if and only if $\mathbf{x}$ is a lattice point. This important idea is expressed in the following proposition:

Proposition 8.1. Let $\mathbf{H}$ be the check matrix for a lattice Λ. Then, for any $\mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{y} \in \Lambda \quad \Leftrightarrow \quad \mathbf{H} \cdot \mathbf{y} \text{ is a vector of integers.} \tag{8.11}$$

Also, $\mathbf{H}\mathbf{G} = \mathbf{G}\mathbf{H} = \mathbf{I}_n$ is the identity matrix.

Example 8.5. The check matrix for the lattice in Example 8.1 is

$$\mathbf{H} = \begin{bmatrix} \frac{1}{3} & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}. \tag{8.12}$$

The point $\mathbf{y}_1 = [-3, -3]^{\mathrm{t}}$ can be verified to be a lattice point since $\mathbf{H}\mathbf{y}_1 = [-1, 0]^{\mathrm{t}}$ is a vector of integers. On the other hand, the point $\mathbf{y}_2 = [3, 2]^{\mathrm{t}}$ is not a lattice point since $\mathbf{H}\mathbf{y}_2 = [1, -\frac{1}{2}]$ is not a vector of integers.

Given a lattice Λ, it has a *dual lattice*, denoted Λ^*. Let Λ have generator matrix $\mathbf{G}$ and check matrix $\mathbf{H} = \mathbf{G}^{-1}$. Then, generator matrix $\mathbf{G}^*$ for the dual lattice Λ^* is:

$$\mathbf{G}^* = \mathbf{H}^{\mathrm{t}}. \tag{8.13}$$

Equivalent to (8.13), if $\mathbf{g}_1, \ldots, \mathbf{g}_n$ is a basis for Λ, then the dual basis $\mathbf{g}_1^*, \ldots, \mathbf{g}_n^*$ satisfies:

$$\mathbf{g}_i \mathbf{g}_i^* = 1 \text{ and } \mathbf{g}_i \mathbf{g}_j^* = 0, i \neq j. \tag{8.14}$$

If the columns of $\mathbf{G}$ are $\mathbf{g}_i$, then the rows of $\mathbf{H}$ are the generator vectors $\mathbf{g}_j^*$ of Λ^*. The dual of Λ^* is Λ.

Example 8.6. Consider the $n = 4$ lattice given by (8.8). Using $\mathbf{G}^* = \mathbf{H}^{\mathrm{t}} = (\mathbf{G}^{-1})^{\mathrm{t}}$, the generator matrix for the dual lattice is:

$$\mathbf{G}^* = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.15}$$

8.2 Fundamental Regions

Definition 8.3. For a lattice Λ, a *fundamental region* $\mathcal{F} \subset \mathbb{R}^n$ is a shape that, if shifted by each lattice point in Λ, will exactly cover the whole real space $\mathbb{R}^n$.

In other words, any point $\mathbf{y} \in \mathbb{R}^n$ is in exactly one fundamental region. This can be expressed as:

$$\mathbb{R}^n = \bigcup_{\mathbf{x} \in \Lambda} \mathcal{F} + \mathbf{x} \text{ and} \tag{8.16}$$

$$\{\mathcal{F} + \mathbf{x}\} \cap \{\mathcal{F} + \mathbf{y}\} = \varnothing, \tag{8.17}$$

for any $\mathbf{x} \neq \mathbf{y}$.

For a given Λ, there are many possible fundamental regions, but they all have the same volume. The volume of the fundamental region is denoted $V(\Lambda)$ and in the case of an $m \times n$ generator matrix $\mathbf{G}$ is $V(\Lambda) = \sqrt{\det(\mathbf{G}^{\mathrm{t}}\mathbf{G})}$. When $\mathbf{G}$ is square, this reduces to:

$$V(\Lambda) = |\det(\mathbf{G})|. \tag{8.18}$$

The determinant could be negative, but volume is never negative[2].

There are many possible fundamental regions for a lattice, so the fundamental region is not unique. Three important fundamental regions are: (1) the Voronoi region, (2) the parallelotope, (3) the hypercube.

(1) The Voronoi region for $\mathbf{x} \in \Lambda$ is the region of space which is closer to $\mathbf{x}$ than to any other lattice point in Λ. Let $\mathbf{x} = Q(\mathbf{y})$ be nearest-neighbor quantization, the lattice point x closest to any $\mathbf{y} \in \mathbb{R}^n$, see also Definition **??** on page **??**.

Definition 8.4. For $\mathbf{x} \in \Lambda$, the *Voronoi region* $\mathcal{V}(\mathbf{x})$ is is the region of space closer to $\mathbf{x}$ than to any other point in Λ:

$$\mathcal{V}(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}^n \mid Q(\mathbf{u}) = \mathbf{x}\} \tag{8.19}$$

The notation $\mathcal{V}$ is the Voronoi region for $\mathbf{0}$, i.e. $\mathcal{V}(\mathbf{0})$, because it is particularly important. There are points in space which are equidistant to two or more lattice points, but each point can belong to only one fundamental region. Tie-breaking in $Q(\mathbf{y})$ is performed in a consistent way, so that $\mathcal{V}(\mathbf{x})$ is indeed a fundamental region. The Voronoi region does not depend on the generator matrix.

Example 8.7. The $n = 2$ hexagonal lattice, denoted A_2, has generator matrix:

$$\mathbf{G} = \begin{bmatrix} \frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \tag{8.20}$$

The A_2 lattice, the basis vectors, and the hexagonal Voronoi region are shown in Fig. 8.4-(a). The volume is $V(\Lambda) = |\det(\mathbf{G})| = \sqrt{3}/2$.

In general, a parallelotope[3] $\mathcal{P} \subset \mathbb{R}^n$ is defined with respect to n linearly independent vectors, the columns of an n-by-n matrix $\mathbf{P}$.

Definition 8.5. Let $\mathbf{P}$ be a matrix of n column vectors $\mathbf{p}_i = \begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix}^{\mathrm{t}}$ for $i = 1, 2, \dots, n$. The *parallelotope* with respect to $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n]$ is given by:

$$\mathcal{P}(\mathbf{P}) = \{s_1\mathbf{p}_1 + s_2\mathbf{p}_2 + \cdots + s_n\mathbf{p}_n | 0 \le s_i < 1\}. \tag{8.21}$$

The volume of the parallelotope is $|\det(\mathbf{P})|$. Wikipedia: Parallelepiped

(2) The parallelotope $\mathcal{P}(\mathbf{G})$ is a fundamental region of a lattice Λ with generator matrix $\mathbf{G}$. The volume of this parallelotope is $|\det(\mathbf{G})|$, and thus the volume $V(\Lambda)$, since all fundamental regions have the same volume. Given $\mathbf{G}$, there are other parallelotopes which also form a fundamental region, see the Appendix.

---

[2] Some texts define the *determinant of a lattice* $\det\Lambda$ as equal to $\det \mathbf{G}^{\mathrm{t}}\mathbf{G}$. This term is $V(\Lambda)^2$ and not $V(\Lambda)$. To avoid confusion, $\det\Lambda$ will not be used.

[3] The term *parallelotope* is attributed to Coxeter [**?**]. Some authors use *parallelepiped* for the general n-dimensional figure, while other authors use parallelepiped for the three-dimensional figure only. In two dimensions, the figure is generally called a parallelogram.

Example 8.8. Continuing Example 8.7 with generator matrix $\mathbf{G}$, the parallelotope $\mathcal{P}(\mathbf{G})$ is:

$$\mathcal{P}(\mathbf{G}) = \{ \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} s_1 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} s_2 \mid 0 \leq s_1, s_2 < 1\} \tag{8.22}$$

and is shown in Fig. 8.4-(b).

---

(3) A hyper-rectangle $\mathcal{H}$ is the generalization of a rectangle to higher dimensions. If Λ has a generator matrix $\mathbf{G}$ in triangular form, then it has a hyperrectangle fundamental region which is easy to find.

Definition 8.6. Let lattice Λ have triangular generator matrix $\mathbf{G}$ with diagonal entries $g_{11}, \ldots, g_{nn}$. The *lattice hyperrectangle* $\mathcal{H}(\mathbf{G})$ is:

$$\mathcal{H}(\mathbf{G}) = \{\mathbf{s} \in \mathbb{R}^n \mid -\frac{|g_{ii}|}{2} \leq s_i < \frac{|g_{ii}|}{2} \text{ for } i = 1, \ldots, n\}. \tag{8.23}$$

If all the $|g_{11}|, |g_{22}|, \ldots, |g_{nn}|$ are equal, then the side lengths are equal and $\mathcal{H}(\mathbf{G})$ is called a hypercube.

Example 8.9. Continuing Example 8.7 with generator matrix $\mathbf{G}$, the hyperrectangle $\mathcal{H}(\mathbf{G})$ is:

$$\mathcal{H}(\mathbf{G}) = \{\mathbf{s} \in \mathbb{R}^2 \mid -\frac{\sqrt{3}}{4} \leq s_1 < \frac{\sqrt{3}}{4}, -\frac{1}{2} \leq s_2 < \frac{1}{2}\}. \tag{8.24}$$

and is shown in Fig. 8.4-(c).

Such a rectangular region exists only when $\mathbf{G}$ is in triangular form:

Proposition 8.2. A lattice Λ has a hypercube fundamental region if and only if Λ has a generator matrix which is triangular.

If $\mathbf{G}$ is triangular, then $\mathcal{H}(\mathbf{G})$ is unique, up to tie-breaking at the boundaries. If $\mathbf{G}$ is not triangular, then it does not have a hyperrectangle fundamental region. However, as described in Section 8.3.2, there exists a congruent lattice which does have a triangular generator matrix.

8.3 Lattice Transformations

This section describes several lattice transformations. Two lattices Λ_1 and Λ_2 are *identical* if all the points are the same. Two lattices Λ_1 and Λ_2 are *congruent* if Λ_2 can be obtained from Λ_1 by an orthogonal transformation. A lattice coset is the shift of a lattice by some vector. A new lattice Λ can be formed by the direct sum of two lattices Λ_1 and Λ_2, that is, $\Lambda = \Lambda_1 \times \Lambda_2$.

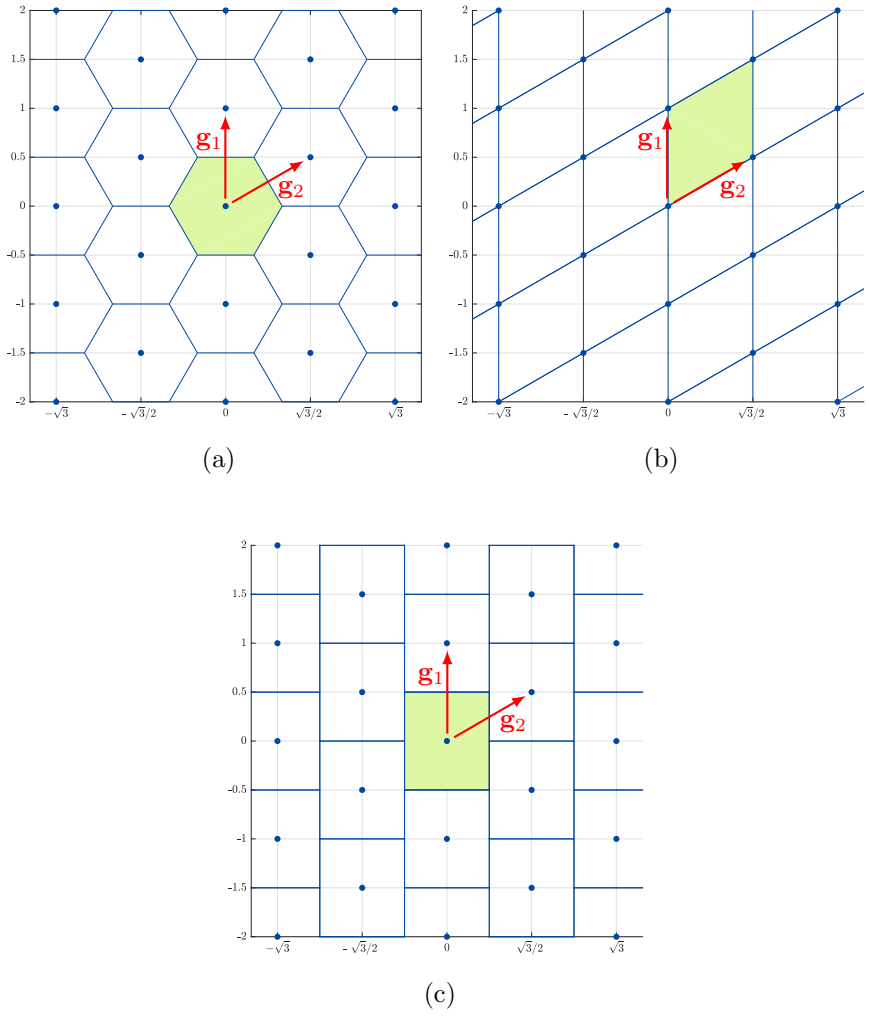


Figure 8.4: For the A_2 lattice: (a) hexagonal Voronoi region which gives the lattice its name, (b) parallelotope fundamental region, (c) hyperrectangle fundamental region.

8.3.1 Basis Transformations — Identical Lattices

Given a lattice Λ, there are many possible generator matrices, that is many possible bases. A unimodular matrix is used to transform one basis to another basis for the same lattice.

Definition 8.7. A *unimodular matrix* $\mathbf{W}$ is a square matrix with integer entries and $\det \mathbf{W} = +1$ or -1.

The inverse of a unimodular matrix is also a unimodular matrix. The product of unimodular matrices, if it exists, is unimodular. The matrix

$$\begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \tag{8.25}$$

has determinant -1 and integer entries, so it is a unimodular matrix.

The basis vectors for a lattice Λ is not unique. Given a generator matrix $\mathbf{G}_1$ for Λ_1, a unimodular transformation can be applied to obtain a distinct matrix $\mathbf{G}_2$ which generates an identical lattice Λ_2.

Proposition 8.3. Two lattices Λ_1 and Λ_2 are identical if the corresponding generator matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ satisfy

$$\mathbf{G}_2 = \mathbf{G}_1 \cdot \mathbf{W} \tag{8.26}$$

where $\mathbf{W}$ is a unimodular matrix.

Equivalently, if the check matrices for Λ_1 and Λ_2 are $\mathbf{H}_1 = \mathbf{G}_1^{-1}$ and $\mathbf{H}_2 = \mathbf{G}_2^{-1}$, then Λ_1 and Λ_2 are identical if

$$\mathbf{H}_2 = \mathbf{W} \cdot \mathbf{H}_1. \tag{8.27}$$

Two generator matrices $\mathbf{G}_1, \mathbf{G}_2$ describe the same lattice if $(\mathbf{G}_1)^{-1}\mathbf{G}_2$ is unimodular. For example, lattice generated by (8.8) and its dual lattice generated by (8.15) are in fact the same lattice since:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

is a unimodular matrix. The lattice and its dual are the same, so this is called a *self-dual lattice*. See also Exercise 8.7.

8.3.2 Congruent Lattices

Two column vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are *orthogonal* (or perpendicular) if $\mathbf{x}^t\mathbf{y} = 0$. The inner product of $\mathbf{x}$ with itself is $\mathbf{x}^{\mathrm{t}}\mathbf{x} = ||\mathbf{x}||^2$. A vector $\mathbf{x}$ is *normalized* if $\mathbf{x}^{\mathrm{t}}\mathbf{x} = ||\mathbf{x}||^2 = 1$.

Definition 8.8. An n-by-n real square matrix $\mathbf{Q}$ is *orthogonal* if

$$\mathbf{Q}^{\mathrm{t}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{t}} = \mathbf{I}_n. \tag{8.28}$$

where $\mathbf{I}_n$ is the identity matrix.

If the columns of $\mathbf{Q}$ are $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$, then $\mathbf{q}_i^{\mathrm{t}} \cdot \mathbf{q}_j$ is 1 if $i = j$ and is 0 if $i \neq j$. For orthogonal matrices, $|\det \mathbf{Q}| = 1$ and $\mathbf{Q}^{\mathrm{t}} = \mathbf{Q}^{-1}$. Wikipedia: Orthogonal matrix

Orthogonal transformation can be applied to the generator matrix:

$$\mathbf{G}_2 = \mathbf{Q}\mathbf{G}_1,$$

where Λ_1 and Λ_2 are two lattices generated by $\mathbf{G}_1$ and $\mathbf{G}_2$. Orthogonal transformation does not change the lattice's minimum distance, volume and kissing number. In two dimensions, orthogonal transformations are rotations and reflections. A rotation is characterized by an angle θ, and the orthogonal rotation matrix is:

$$\mathbf{Q} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{8.29}$$

In dimensions $n \geq 3$, orthogonal matrices can represent not only rotations and reflections, but other transformations as well.

Definition 8.9. Two lattices Λ_1 and Λ_2 are *congruent* if the generator matrix $\mathbf{G}_2$ for Λ_2 can be obtained from the generator matrix $\mathbf{G}_1$ for Λ_1 as:

$$\mathbf{G}_2 = c\mathbf{Q}\mathbf{G}_1\mathbf{W} \tag{8.30}$$

for $c \neq 0$, where $\mathbf{W}$ is unimodular and $\mathbf{Q}$ is orthogonal.

To determine congruence of two lattices, it is not sufficient to test $\mathbf{G}_1^{-1}\mathbf{G}_2$ for orthogonality because $\mathbf{W} \neq \mathbf{I}$ in general.

A triangular generator matrix is useful in many cases, including encoding nested lattice codes and generalized lattice decoding. The QR decomposition is a transformation that puts the generator matrix for a lattice into triangular form. In general, the QR decomposition of a matrix $\mathbf{G}$ is:

$$\mathbf{G} = \mathbf{Q}\mathbf{R} \tag{8.31}$$

where $\mathbf{Q}$ is the orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. If $\mathbf{G}$ is invertible, then this factorization is unique, up to sign changes. It can be see that if $\mathbf{G}$ is the generator matrix for a lattice, then $\mathbf{R}$ is an upper triangular generator matrix for a congruent lattice. Wikipedia: QR Decomposition

However, we prefer working with lower triangular matrices. The QL decomposition of the generator matrix $\mathbf{G}$ is:

$$\mathbf{G} = \mathbf{Q}\mathbf{L} \tag{8.32}$$

where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{L}$ is a lower triangular matrix. Many software packages have a QR decomposition. The QL decomposition may be

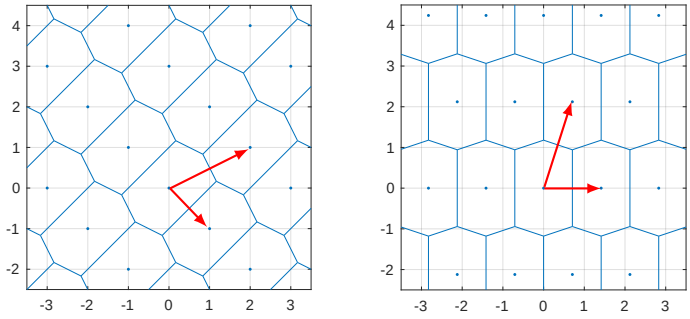


Figure 8.5: Example conversion to a diagonal generator matrix.

obtained from the QR decomposition as follows. For a matrix $\mathbf{A}$, let $\mathbf{A}^{\text{sr}}$ be the matrix obtained by reversing the order of the rows, and reversing the order of the columns. The QR decomposition of $\mathbf{G}^{\text{sr}}$ is $\mathbf{G}^{\text{sr}} = \mathbf{QR}$, where where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. Then, the QL decomposition is $\mathbf{G} = \mathbf{Q}^{\text{sr}}\mathbf{R}^{\text{sr}}$, where the desired orthogonal matrix is $\mathbf{Q}^{\text{sr}}$ and the desired lower triangular matrix is $\mathbf{R}^{\text{sr}}$.

Given a lattice Λ_1, there may be no *identical* lattice with a triangular generator matrix, but there is generally an *equivalent* lattice Λ_2 with a triangular generator matrix. If Λ_1 has generator matrix $\mathbf{G}$ and $\mathbf{G} = \mathbf{QL}$ by the QL decomposition, then $\mathbf{L}$ is a lower-triangular generator matrix for an equivalent lattice Λ_2.

Example 8.10. Consider a lattice with generator matrix

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix}, \tag{8.33}$$

which is shown Fig. 8.5-(a). The QL decomposition gives orthogonal matrix $\mathbf{Q}$ and lower-triangular matrix $\mathbf{G}_2$:

$$\mathbf{Q} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \text{ and } \mathbf{G}_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{18}}{2} & -\sqrt{18} \end{bmatrix}, \tag{8.34}$$

where $\mathbf{G}_1 = \mathbf{QG}_2$. The lattice with generator matrix $\mathbf{G}_2$ is shown in 8.5-(b), although the generator vectors do not correspond to the arrows.

It is also possible to find an orthogonal matrix using the Cholesky decomposition, see Exercise 8.8.

8.3.3 Scaling

Lattices are easily scaled. If Λ is a lattice and $k \in \mathbb{R}$ with $k \neq 0$, then $k\Lambda$ is a scaled lattice. If $\mathbf{x} \in \Lambda$ then $k\mathbf{x} \in k\Lambda$. If Λ has generator matrix $\mathbf{G}$, then $k\Lambda$

has generator matrix $k\mathbf{G}$. The volume of $k\Lambda$ is:

$$V(k\Lambda) = |\det(k\mathbf{G})| = |k^n \det(\mathbf{G})| \tag{8.35}$$
$$= |k^n| V(\Lambda). \tag{8.36}$$

The squared minimum distance of $k\Lambda$ is $k^2 d_{\min}^2$. The kissing number does not change under scaling. Two congruent lattices which are then scaled differently are no longer congruent, but they are equivalent. That is if Λ_1 and Λ_2 are congruent, then $k_1\Lambda_1$ and $k_2\Lambda_2$ are equivalent but not congruent, for $|k_1| \neq |k_2|$.

8.3.4 Lattices by Direct Sum

A new lattice can be made by combining two or more lattices by direct sum. Let Λ_1 and Λ_2 be n-dimensional and m-dimensional lattices with square generator matrices $\mathbf{G}_1$ and $\mathbf{G}_2$, respectively. Then the *direct sum* is an $n+m$ dimensional lattice $\Lambda_1 \oplus \Lambda_2$ with generator matrix $\mathbf{G}$:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix}. \tag{8.37}$$

The volume of $\Lambda_1 \oplus \Lambda_2$ is $|\det \mathbf{G}_1| \cdot |\det \mathbf{G}_2|$. The direct sum generalizes to an arbitrary number of lattices T: $\Lambda_1 \oplus \Lambda_2 \oplus \cdots \oplus \Lambda_T$. Wikipedia: Direct sum

---

Example 8.11. Let the lattice Λ_1 be the direct sum $\sqrt{3}\mathbb{Z} \oplus \mathbb{Z}$ which has generator matrix

$$\mathbf{G}_1 = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}. \tag{8.38}$$

Λ_1 is a scaled version of the integer lattice as is illustrated in Fig. **??**.

Example 8.12. The direct sum of two hexagonal lattices $A_2 \oplus A_2$ has generator matrtix:

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}. \tag{8.39}$$

8.3.5 Lattice Cosets

Lattice cosets are a shift of a lattice by some point.

Definition 8.10. A *lattice coset* of Λ with respect to a vector $\mathbf{s}$ is the set $\Lambda + \mathbf{s}$.

A lattice coset is not a lattice, unless $\mathbf{s} \in \Lambda$, since a lattice must contain the $\mathbf{0}$ point.

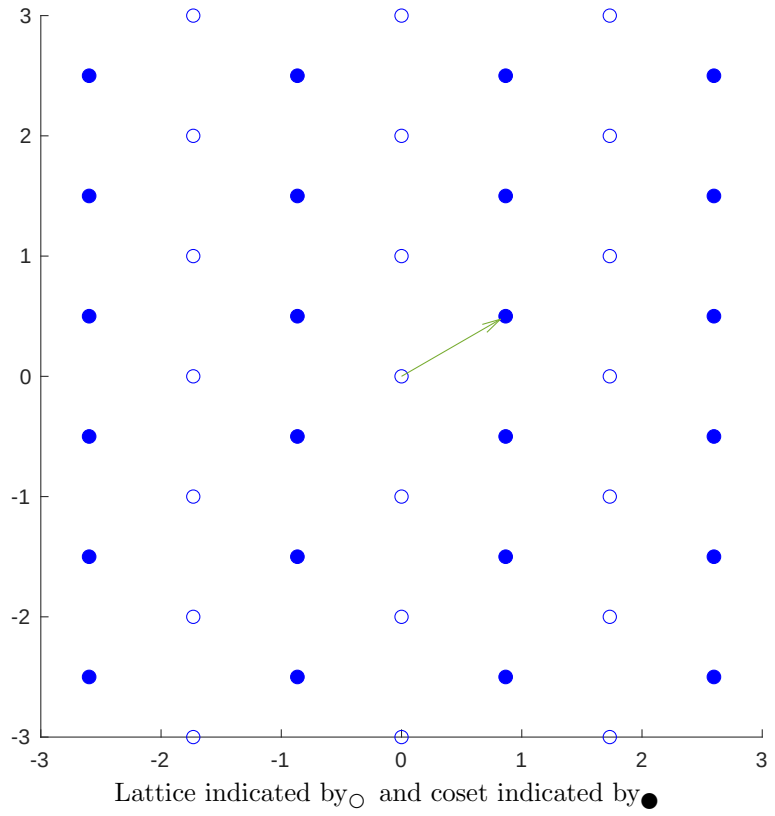


Figure 8.6: Example of a lattice and its coset.

Example 8.13. The lattice coset of the lattice Λ_1 given in Example 8.11 with respect to the coset vector $\mathbf{s}$

$$\mathbf{s} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} \tag{8.40}$$

is the set $\Lambda_1 + \mathbf{s}$. The original lattice Λ_1 and its coset are illustrated in Fig. 8.6.

8.4 Lattice Properties

A lattice Λ has several important parameters, including the squared minimum distance, the packing density and the kissing number. These properties are important when considering the lattice as an error-correcting code. Other properties important for quantization or shaping are considered in Chapter **??**.

8.4.1 Minimum Distance and Packing Sphere

The minimum distance of a lattice is the smallest distance between any two distinct lattice points. We usually work with the squared minimum distance $d^2_{\min}$ as given in Definition **??**. Further, since a lattice is linear, $d^2_{\min}$ is given by:

$$d^2_{\min} = \min_{\mathbf{x} \in \Lambda \setminus \mathbf{0}} ||\mathbf{x}||^2. \tag{8.41}$$

If any two points $\mathbf{x}, \mathbf{y}$ satisfy $||\mathbf{x} - \mathbf{y}||^2 = d^2_{\min}$, then by linearity, another lattice point $\mathbf{z} = \mathbf{x} - \mathbf{y}$ satisfies $||\mathbf{z}||^2 = d^2_{\min}$.

Integrate the following: A vector at the minimum distance from $\mathbf{0}$ is sometimes called the minimal vector.

The squared minimum distance changes under lattice scaling. $d^2_{\min}$ of a lattice Λ, and the squared minimum distance $d'^2_{\min}$ of a scaled version $\Lambda' = k\Lambda$ are related by:

$$d'^2_{\min} = k^2 d^2_{\min}. \tag{8.42}$$

A *packing sphere* is centered on each lattice point with a radius such that these spheres just touch each other. The packing sphere do not overlap. In two dimensions, spheres are circles and are shown in Fig. 8.7-(a) for the hexagonal A_2 lattice. The *packing radius* ρ is the radius of the packing sphere, and is half the minimum distance $d_{\min}$. Dealing with squared distances:

$$\rho^2 = \frac{d^2_{\min}}{4}. \tag{8.43}$$

Recall the problem of packing spherical oranges into a box at the beginning of this chapter. One way to measure the quality of the packing is the packing density. The *packing density* Δ of a lattice is the fraction of the whole space occupied by the spheres:

$$\Delta = \frac{\text{volume of one sphere}}{\text{volume of the fundamental region}} = \frac{V_n \rho^n}{V(\Lambda)}, \tag{8.44}$$

where V_n is the volume of a unit sphere in n dimensions, given by $V_n = \frac{2\pi^{n/2}}{n\Gamma(\frac{n}{2})}$, where Γ is the gamma function. The highest possible packing density is 1, achieved only if the Voronoi region is a sphere.

In n dimensions, the ball $\mathcal{B}_n$ is the set of points in $\mathbb{R}^n$ within a squared radius ρ^2 of $\mathbf{0}$:

$$\mathcal{B}_n = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}||^2 \leq \rho^2\}, \tag{8.45}$$

The volume of an n-ball of radius r is:

$$V_n(r) = \frac{2\pi^{n/2}}{n\Gamma(\frac{n}{2})} r^n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} r^n. \tag{8.46}$$

For n even:

$$V_n(r) = \frac{\pi^{n/2}}{(\frac{n}{2})!} r^n. \tag{8.47}$$

Wikipedia: n-Ball

Example 8.14. Compare the packing density of the A_2 lattice and the $\mathbb{Z}^2$ lattice. The A_2 lattice has $d^2_{\min} = 1$ and squared packing radius $\rho^2 = \frac{1}{4}$ and $|\det \mathbf{G}| = \frac{\sqrt{3}}{2}$.

$$\Delta(A_2) = \frac{\pi\rho^2}{|\det \mathbf{G}|} = \frac{\pi\sqrt{3}}{6} \approx 0.9069 \tag{8.48}$$

The $\mathbb{Z}^2$ lattice also has $d^2_{\min} = 1$ and $\rho^2 = \frac{1}{4}$, but $|\det \mathbf{G}| = 1$, so:

$$\Delta(\mathbb{Z}^2) = \frac{\pi\rho^2}{|\det \mathbf{G}|} = \frac{\pi}{4} \approx 0.7854 \tag{8.49}$$

The higher packing density of the A_2 lattice confirms the visual intuition shown in Fig. 8.1.

The coding gain γ_c measures the increase is density of Λ over the baseline integer lattice $\mathbb{Z}$. The shaping gain $\gamma_s(\mathcal{R})$ measures the decrease in average energy of $\mathcal{R}$ relative to an interval $[-M, M)$. Both contribute a multiplicative factor of gain in the expression for Pr(error).

The coding gain normalizes the minimum distance and is a useful measure of a lattice's suitability for error-correction.

Definition 8.11. For an n-dimensional lattice Λ with squared minimum distance $d^2_{\min}$ the (nominal) *coding gain* is:

$$\gamma = \frac{d^2_{\min}}{V(\Lambda)^{2/n}}. \tag{8.50}$$

The coding gain is independent of lattice scaling.

Table **??** lists the coding gain of some lattices. Coding gain is often expressed in dB. This *nominal* coding gain is predicted from $d^2_{\min}$, however in practice the *effective* coding gain is lower.

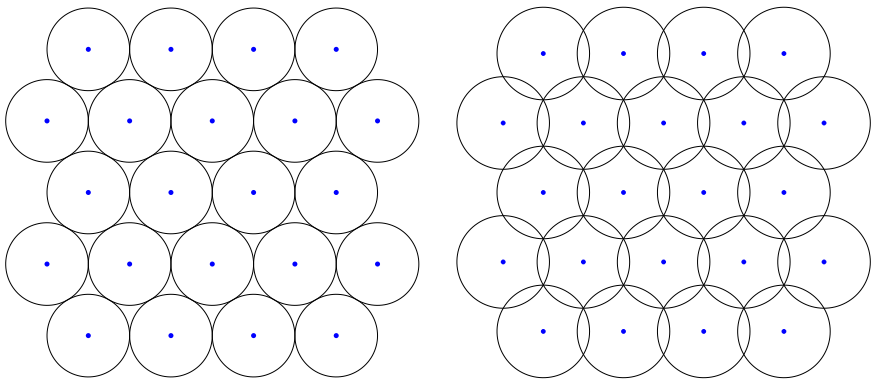
8.4.2 Kissing Number and Theta Function

The kissing number τ for a lattice is the number of lattice points at the minimum distance. It is the number of points at squared distance $d_{\min}$ from the origin (or any other lattice point, by linearity). When lattices are used for communications, the kissing number is used in estimating the probability of decoding error.

The theta function expresses the number of points at each distance from $\mathbf{0}$. Let N_m be the number of lattice points at squared norm m from the origin. The theta function is:

$$\Theta(q) = \sum_{i=1}^{\infty} N_m q^m. \tag{8.51}$$

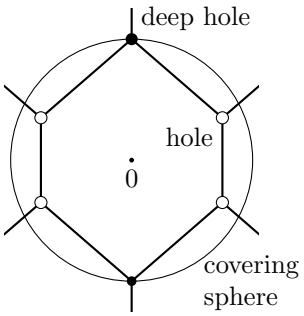
Note that $N_0 = 1$, i.e. the origin itself. The theta function for lattices is analogous to the distance spectrum for error-correcting codes. Since the kissing number τ is the number of lattice points at the squared minimum distance $d^2_{\min}$, $N_{d^2_{\min}} = \tau$ holds.



(a) (b)

Figure 8.7: (a) Packing spheres for the A_2 lattice (b) Covering spheres for the A_2 lattice.

8.4.3 Deep Hole and Covering Radius



The "corners" of a Voronoi region are called holes. These have special characteristics that define the Voronoi region. They are points of the Voronoi region which are further from the center than other points in its local neighborhood, i.e. they are local maximum of the distance from the center. A point of space furthest from any point of a lattice is called a *deep hole*, that is the deep hole is a hole at the greatest distance from its lattice point. The *covering sphere* is the sphere of minimal radius which covers the entire Voronoi region. Its radius is the *covering radius* of a lattice and is the minimum distance from the deep hole to any lattice point. The radius is the worst-case error when the lattice is used for quantization, since no point is further than the deep hole from a lattice point. Hole and deep holes and the covering sphere are illustrated in the figure at the right. Note that this figure does not correspond to a lattice.

Besides the packing sphere and covering sphere, the equivalent sphere is also of interest. The *equivalent sphere* has radius such that the volume of the equivalent sphere is equal to the volume of the Voronoi region.

8.5 Exercises

8.1 Consider the lattice with:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} \tag{8.52}$$

Are $\mathbf{y}_1 = [1, 3]^{\mathrm{t}}$ and $\mathbf{y}_2 = [2, -3]^{\mathrm{t}}$ lattice points?

8.2 Consider the lattice with generator matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 2 \end{bmatrix}. \tag{8.56}$$

For this lattice:

(a) Find and draw (1) the Voronoi fundamental region (2) the parallelotope fundamental region (3) the hyperrectangle fundamental region.

(b) Find the volume $V(\Lambda)$, the squared minimum distance $d^2_{\min}$ and the kissing number τ.

8.3 Consider a lattice Λ_0 generated by

$$\mathbf{G}_0 = \begin{bmatrix} 2 & 1 \\ 4 & 8 \end{bmatrix}. \tag{8.60}$$

(a) Which of the following matrices are also generators for Λ_0?

$$\mathbf{G}_1 = \begin{bmatrix} 4 & 3 \\ 20 & 18 \end{bmatrix}, \mathbf{G}_2 = \begin{bmatrix} 5 & 6 \\ 28 & 36 \end{bmatrix} \mathbf{G}_3 = \begin{bmatrix} 3 & 6 \\ 12 & 24 \end{bmatrix}. \tag{8.61}$$

(b) Let $\mathbf{x} = [3, 3]^{\mathrm{t}}$, and $\mathbf{x} \in \Lambda_0$. Using $\mathbf{G}_0$, what integers $\mathbf{b}_0$ generate this $\mathbf{x}$?

(c) For each generator matrix from part (a) which generates Λ_0, what integers $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$, if any, were used to generate $\mathbf{x}$?

(d) How is it possible that distinct integers $\mathbf{b}_0, \mathbf{b}_1$ generate the same lattice point $\mathbf{x}$?

8.4 The "cannonball lattice" has a rectangular base, scaled by β. It has an additional generator vector $\mathbf{g}$, so the lattice has generator matrix:

$$\left[\begin{array}{ccc} \beta & 0 & | \\ 0 & \beta & \mathbf{g} \\ 0 & 0 & | \end{array}\right].$$

Find β and $\mathbf{g}$ such that the cannonball lattice can be obtained by an orthogonal transformation of:

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

8.5 Some generator matrices can be put into lower triangular form without using the QL decomposition. Using integer column operations only, put the following matrix into lower triangular form:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & -2 \\ -1 & 1 & 3 \\ 0 & -1 & 1 \end{bmatrix} \tag{8.62}$$

An integer column operation is replacing column i with column i plus an integer multiple of column j.

8.6 For dimension $n = 2$, let $\mathbf{g}_1$ and $\mathbf{g}_2$ be linearly independent basis vectors. Consider two lattices with generator matrices $\mathbf{G}_1$ and $\mathbf{G}_2$:

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_2 \end{bmatrix} \text{ and } \mathbf{G}_2 = \begin{bmatrix} \mathbf{g}_1 & \mathbf{g}_1 + \mathbf{g}_2 \end{bmatrix}. \tag{8.65}$$

Show that $\mathbf{G}_1$ and $\mathbf{G}_2$ generate the same lattice.

8.7 A unimodular lattice is a lattice with a unimodular generator matrix. Prove the following statements:

(a) If a lattice is unimodular, then its dual lattice is an integral lattice.

(b) Unimodular lattices are equal to their dual lattices. (For this reason, unimodular lattices are also known as self-dual.)

8.8 For a lattice Λ with generator matrix $\mathbf{G}$, let $\mathbf{A}$ be the Cholesky factorization of the Gram matrix $\mathbf{G}^{\mathrm{t}}\mathbf{G}$. Prove that the Cholesky factorization can be used to find a lattice congruent to Λ which has a triangular generator matrix.