Sample Title

Your Name
Your Affiliation

SYNTHESIS LECTURES ON SAMPLE SERIES #1

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

SAMPLE ABSTRACT

In quantum computing, where algorithms exist that can solve computational problems more efficiently than any known classical algorithms, the elimination of errors that result from external disturbances or from imperfect gates has become the "holy grail," and a worldwide quest for a large scale fault-tolerant and computationally superior quantum computer is currently taking place. Optimists rely on the premise that, under a certain threshold of errors, an arbitrary long fault-tolerant quantum computation can be achieved with only moderate (i.e., at most polynomial) overhead in computational cost. Pessimists, on the other hand, object that there are in principle (as opposed to merely technological) reasons why such machines are still inexistent, and that no matter what gadgets are used, large scale quantum computers will never be computationally superior to classical ones. Lacking a complete empirical characterization of quantum noise, the debate on the physical possibility of such machines invites philosophical scrutiny. Making this debate more precise by suggesting a novel statistical mechanical perspective thereof is the goal of this project.

KEYWORDS

SAMPLE KEYWORDS

computational complexity, decoherence, error-correction, fault-tolerance, Landauer's Principle, Maxwell's Demon, quantum computing, statistical mechanics, thermodynamics

Contents

CHAPTER 1

Information Theory and Rate Distortion Theory

1.1 INTRODUCTION

Thus far in the book, the term *information* has been used sparingly and when it has been used, we have purposely been imprecise as to its meaning. Although, everyone has an intuitive feeling for what information is, it is difficult to attach a meaningful quantitative definition to the term. In the context of communication systems, Claude Shannon was able to do exactly this, and as a result, opened up an entirely new view of communication systems analysis and design Levin (2002). The principal contribution of Shannon's information theory to date has been to allow communication theorists to establish absolute bounds on communication systems performance that cannot be exceeded no matter how ingeniously designed or complex our communication systems are. Fundamental physical limitations on communication systems performance is another topic that has been largely ignored in the preceding chapters, **(OBS: mentions preceding chapters.)** but it is a subject of exceptional practical importance. For example, for any of the numerous communication systems developed thus far in the book, we could decide to design a new system that would outperform the accepted standard for a particular application. The first question that we should ask is: how close is the present system to achieving theoretically optimum performance? If the existing communication system operates at or near the fundamental physical limit on performance, our task may be difficult or impossible. However, if the existing system is far away from the absolute performance bound, this might be an area for fruitful work.

Of course, in specifying the particular communication system under investigation, we must know the important physical parameters, such as transmitted power, bandwidth, type(s) of noise present, and so on, and information theory allows these constraints to be incorporated. However, information theory does not provide a way for communication system complexity to be explicitly included. Although, this is something of a drawback, information theory itself provides a way around this difficulty, since it is generally true that as we approach the fundamental limit on the performance of a communication system, the system complexity increases, sometimes quite drastically. Therefore, for a simple communication system operating far from its performance bound, we may be able to improve the performance with a relatively modest increase in complexity. On the other hand, if we have a rather complicated communication system operating near its fundamental limit, any performance improvement may be possible only with an extremely complicated system.

In this chapter we are concerned with the rather general block diagram shown in Fig. 1.1. Most of the early work by Shannon and others ignored the source encoder/decoder blocks and concentrated on bounding the performance of the channel encoder/decoder pair. Subsequently, the source encoder/decoder blocks have attracted much research attention. In this chapter we consider both topics and expose the reader to the nomenclature used in the information theory literature. Quantitative definitions of information are presented in Sec. 1.2 that lay the foundation for the remaining sections. In Secs. 1.3 and 1.4 we present the fundamental source and channel coding theorems, give some examples, and state the implications of these theorems. Section 1.5 contains a brief development of rate distortion theory, which is the mathematical basis for data compression. A few applications of the theory in this chapter are presented in Sec. 1.6, and a technique for variable-length source coding is given in Sec. 1.7.

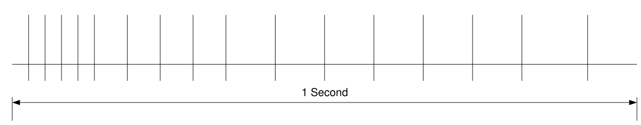

Figure 1.1: Communication system block diagram.

1.2 ENTROPY AND AVERAGE MUTUAL INFORMATION

Consider a discrete random variable U that takes on the values $\{u_1, u_2, \ldots, u_M\}$, where the set of possible values of U is often called the *alphabet* and the elements of the set are called *letters* of the alphabet. Let $P_U(u)$ denote the probability assignment over the alphabet, then we can define the *self-information* of the event $u = u_j$ by

$$I_U\left(u_j\right) = \log \frac{1}{P_U(u_j)} = -\log P_U\left(u_j\right) \ . \tag{1.1}$$

The quantity $I_U(u_j)$ is a measure of the information contained in the event $u = u_j$. Note that the base of the logarithm in Eq. (1.1) is unspecified. It is common to use base e, in which case $I_U(\cdot)$ is in natural units (nats), or base 2, in which case $I_U(\cdot)$ is in binary units (bits). Either base is acceptable since the difference in the two bases is just a scaling operation. We will use base 2 in all of our work, and hence $I_U(\cdot)$ and related quantities will be in bits. The average or expected value of the self-information is called the *entropy,* also discrete entropy or absolute entropy, and is given by

$$H(U) = -\sum_{j=1}^{M} P_U\left(u_j\right) \log P_U\left(u_j\right) \ . \tag{1.2}$$

The following example illustrates the calculation of entropy and how it is affected by probability assignments.

Example 1.1 Given a random variable U with four equally likely letters in its alphabet, we wish to find $H(U)$. Clearly, $M = 4$ and $P_U(u_i) = \frac{1}{4}$ for $i = 1, 2, 3, 4$. Thus, from Eq. (1.2),

$$\begin{aligned} H(U) &= -\sum_{j=1}^{4} \tfrac{1}{4} \log \tfrac{1}{4} \\ &= -\log \tfrac{1}{4} = 2 \text{ bits} . \end{aligned} \tag{1.3}$$

We now consider a discrete random variable X with four letters such that $P_X(x_1) = \frac{1}{2}, P_X(x_2) = \frac{1}{4}, P_X(x_3) = P_X(x_4) = \frac{1}{8}$. Again, we wish to find the entropy of this random variable. Directly,

$$H(X) = -\tfrac{1}{2} \log \tfrac{1}{2} - \tfrac{1}{4} \log \tfrac{1}{4} - \tfrac{1}{8} \log \tfrac{1}{8} - \tfrac{1}{8} \log \tfrac{1}{8} = 1.75 \text{ bits} . \tag{1.4}$$

Comparing Eqs. (1.3) and (1.4), we see that equally likely letters produce a larger entropy. This result is true in general, as we will see shortly.

We now consider two jointly distributed discrete random variables W and X with the probability assignment $P_{WX}(w_j, x_k), j = 1, 2, \ldots, M, k = 1, 2, \ldots, N$. We are particularly interested in the interpretation that w is an input letter to a noisy channel and x is the corresponding output. The information provided about the event $w = w_j$ by the occurrence of the event $x = x_k$ is

$$I_{W;X}\left(w_j; x_k\right) = \log \frac{P_{W|X}\left(w_j | x_k\right)}{P_W\left(w_j\right)} . \tag{1.5}$$

Further, the information provided about the event $x = x_k$ by the occurrence of $w = w_j$ is

$$I_{X;W}\left(x_k; w_j\right) = \log \frac{P_{X|W}\left(x_k | w_j\right)}{P_X\left(x_k\right)} . \tag{1.6}$$

We can show that $I_{W;X}(w_j; x_k) = I_{X;W}(x_k; w_j)$ by starting with Eq. (1.5) and using conditional probability as follows:

$$\begin{aligned} I_{W;X}\left(w_j; x_k\right) &= \log \frac{P_{WX}\left(w_j, x_k\right)}{P_W\left(w_j\right) P_X\left(x_k\right)} \\ &= \log \frac{P_{X|W}\left(x_k | w_j\right)}{P_X\left(x_k\right)} = I_{X;W}\left(x_k; w_j\right) . \end{aligned} \tag{1.7}$$

Because of this symmetry, Eqs. (1.5) and (1.6) are called the *mutual information* between the events $w = w_j$ and $x = x_k$. The *average mutual information* over the joint ensemble is an important quantity defined by

$$I(W;X) = \sum_{j=1}^{M} \sum_{k=1}^{N} P_{WX}(w_j, x_k)\, I_{W;X}(w_j; x_k)$$
$$= \sum_{j=1}^{M} \sum_{k=1}^{N} P_{WX}(w_j, x_k) \log \frac{P_{W|X}(w_j|x_k)}{P_W(w_j)} \ . \tag{1.8}$$

By a straightforward manipulation of Eq. (1.5),

$$I_{W;X}(w_j; x_k) = -\log P_W(w_j) + \log P_{W|X}(w_j|x_k)$$
$$= I_W(w_j) - I_{W|X}(w_j|x_k) \ , \tag{1.9}$$

where

$$I_{W|X}(w_j|x_k) \triangleq -\log P_{W|X}(w_j|x_k) \tag{1.10}$$

is called the *conditional self-information,* and is interpreted as the information that must be supplied to an observer to specify $w = w_j$ after the occurrence of $x = x_k$. Substituting Eq. (1.9) into Eq. (1.8), we find that

$$I(W;X) = H(W) - H(W|X) \ , \tag{1.11}$$

where $H(W|X)$ is the *average conditional self-information.* Since entropy is a measure of uncertainty, we see from Eq. (1.11) that the average mutual information can be interpreted as the average amount of uncertainty remaining after the observation of X.

Example 1.2 Here we wish to calculate the mutual information and the average mutual information for the probability assignments (with $M = 2$ and $N = 2$)

$$P_W(w_1) = P_W(w_2) = \tfrac{1}{2} \tag{1.12}$$

and

$$P_{X|W}(x_1|w_1) = P_{X|W}(x_2|w_2) = 1 - p \tag{1.13}$$
$$P_{X|W}(x_1|w_2) = P_{X|W}(x_2|w_1) = p \ . \tag{1.14}$$

If we interpret W as the input to a channel X as the output, the transition probabilities in Eqs. (1.13) and (1.14) are representative of what is called a *binary symmetric channel* (BSC).

To calculate the mutual information, we could use either Eq. (1.5) or (1.6). From the probabilities given, Eq. (1.6) seems to be slightly simpler. Thus, we need $P_X(x_1)$ and $P_X(x_2)$. Directly,

$$P_{WX}(w_1, x_1) = P_{X|W}(x_1|w_1) P_W(w_1) = \frac{1-p}{2}$$
$$= P_{WX}(w_2, x_2) \tag{1.15}$$

and

$$\begin{aligned} P_{WX}(w_1, x_2) &= P_{X|W}(x_2|w_1)P_W(w_1) \\ &= \frac{p}{2} = P_{WX}(w_2, x_1) \ . \end{aligned} \tag{1.16}$$

Thus,

$$\begin{aligned} P_X(x_1) &= P_{WX}(w_1, x_1) + P_{WX}(w_2, x_1) \\ &= \tfrac{1}{2} = P_X(x_2) \ , \end{aligned} \tag{1.17}$$

so the four mutual information values are

$$I_{X;W}(x_1; w_1) = \log 2(1-p) = I_{X;W}(x_2; w_2) \tag{1.18}$$

and

$$I_{X;W}(x_1; w_2) = \log 2p = I_{X;W}(x_2; w_1) \ . \tag{1.19}$$

The average mutual information follows in a straightforward fashion from Eq. (1.8) as

$$\begin{aligned} I(W;X) = I(X;W) &= \sum_{j=1}^{2} \sum_{k=1}^{2} P_{WX}(w_j, x_k) \log \frac{P_{X|W}(x_k|w_j)}{P_X(x_k)} \\ &= \frac{1-p}{2} \log 2(1-p) + \frac{p}{2} \log 2p + \frac{p}{2} \log 2p + \frac{1-p}{2} \log 2(1-p) \\ &= 1 + (1-p)\log(1-p) + p \log p \ . \end{aligned} \tag{1.20}$$

The average mutual information given by Eq. (1.20) is plotted versus p in Fig. 1.2. The results of this example are discussed more in Sec. 1.4 in a communications context.

Figure 1.2: Average mutual information for a BSC with equally likely inputs (Ex. 1.2).

There are numerous useful properties of entropy and average mutual information. We state two of these properties here without proof.

Property 1.3 Let U be a random variable with possible values $\{u_1, u_2, \ldots, u_M\}$. Then

$$H(U) \le \log M \tag{1.21}$$

with equality if and only if the values of U are equally likely to occur.

Example 1.1 illustrates Pr. 1.3.

Property 1.4 Let W and X be jointly distributed random variables. The average mutual information between W and X satisfies

$$I(W;X) \geq 0 \tag{1.22}$$

with equality if and only if W and X are statistically independent.

Thus, far we have defined the entropy and average mutual information only for discrete random variables. Given an absolutely continuous random variable U with pdf $f_U(u)$ we define the *differential entropy* of U as

$$h(U) = -\int_{-\infty}^{\infty} f_U(u) \log f_U(u)\, du\ . \tag{1.23}$$

Although, Eqs. (1.2) and (1.23) are functionally quite similar, there is a significant difference between the interpretations of absolute or discrete entropy and differencial entropy. While $H(U)$ is an absolute indicator of "randomness," $h(U)$ is only an indicator of randomness with respect to a coordinate system: hence the names "absolute entropy" for $H(U)$ and "differential entropy" for $h(U)$. The following example illustrates the calculation of differential entropy and its property of indicating randomness with respect to a coordinate system.

Example 1.5 Consider an absolutely continuous random variable with uniform pdf

$$f_U(u) = \begin{cases} \dfrac{1}{a}, & \dfrac{-a}{2} \leq u \leq \dfrac{a}{2} \\ 0, & \text{elsewhere}\ . \end{cases} \tag{1.24}$$

(1) Let $a = 1$ in Eq. (1.24) and find $h(U)$. Then

$$h(U) = -\int_{-1/2}^{1/2} \log 1\, du = 0\ . \tag{1.25}$$

(2) Let $a = 32$ and find $h(U)$. We have

$$h(U) = -\int_{-16}^{16} \tfrac{1}{32} \log\left(\tfrac{1}{32}\right) du = 5\ . \tag{1.26}$$

(3) Finally, let $a = \frac{1}{32}$ and find $h(U)$. Here

$$h(U) = -\int_{-1/64}^{1/64} 32 \log(32)\, du = -5\ . \tag{1.27}$$

The fact that differential entropy is a relative indicator of randomness is evident from these three special cases of the uniform distribution. Clearly, $h(U)$ is not an absolute indicator of randomness, since in case (3) $h(U)$ is negative, and negative randomness is difficult to interpret physically! The "reference"distribution is the uniform distribution over a unit interval, with "broader" distributions having a positive entropy and "narrower" distributions having a negative differential entropy.

Differential entropies are unlike absolute entropy in that differential entropy is not always positive, not necessarily finite, not invariant to a one-to-one transformation of the random variable, and not subject to interpretation as an average self-information.

The average mutual information of two jointly distributed continuous random variables, say W and X, can also be defined as

$$\begin{aligned} I(W;X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{WX}(w,x)\log\frac{f_{WX}(w,x)}{f_W(w)f_X(x)}\,dw\,dx \\ &= I(X;W)\ . \end{aligned} \tag{1.28}$$

As in the discrete case, the average mutual information can be expressed in terms of entropies as

$$\begin{aligned} I(W;X) &= h(W) - h(W|X) \\ &= h(X) - h(X|W)\ , \end{aligned} \tag{1.29}$$

where

$$h(W|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{WX}(w,x)\log f_{W|X}(w|x)\,dw\,dx\ . \tag{1.30}$$

Fortunately for our subsequent uses of $I(W;X)$, the average mutual information is invariant under any one-to-one transformation of the variables, even though the individual differential entropies are not.

Example 1.6 In this example we compute the differential entropy of the input and the average mutual information between the input and output of an additive white Gaussian noise channel with zero mean and variance σ_c^2. The input is also assumed to be Gaussian with zero mean and variance σ_s^2. Representing the channel as shown in Fig. 1.3, we thus have that

$$f_W(w) = \frac{1}{\sqrt{2\pi}\sigma_s}e^{-w^2/2\sigma_s^2} \tag{1.31}$$

and

$$f_{X|W}(x|w) = \frac{1}{\sqrt{2\pi}\sigma_c}e^{-(x-w)^2/2\sigma_c^2}\ . \tag{1.32}$$

Figure 1.3: Additive noise channel.

From Eq. (1.23) the differential entropy of the input is

$$\begin{aligned}
h(W) &= -\int_{-\infty}^{\infty} f_W(w) \log f_W(w)\, dw \\
&= \int_{-\infty}^{\infty} f_W(w) \left\{ \log \sqrt{2\pi}\sigma_s + \frac{w^2}{2\sigma_s^2} \log e \right\} dw \\
&= \log \sqrt{2\pi}\sigma_s + \tfrac{1}{2} \log e = \tfrac{1}{2} \log 2\pi e \sigma_s^2 \ . \qquad (1.33)
\end{aligned}$$

To calculate the average mutual information, we choose to employ the expression $I(W; X) = h(X) - h(X|W)$. We already have $f_{X|W}(x|w)$, but we need $f_X(x)$. This follows directly as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\left[\sigma_s^2 + \sigma_c^2\right]^{1/2}} e^{-x^2/2(\sigma_s^2+\sigma_c^2)} \ , \qquad (1.34)$$

since the input and the noise are independent, zero-mean Gaussian processes. By analogy with Eq. (1.33), we have from Eqs. (1.32) and (1.34) that

$$h(X) = \tfrac{1}{2} \log 2\pi e\left[\sigma_s^2 + \sigma_c^2\right] \qquad (1.35)$$

and

$$h(X|W) = \tfrac{1}{2} \log 2\pi e \sigma_c^2 \ . \qquad (1.36)$$

We thus find the average mutual information to be

$$\begin{aligned}
I(W; X) &= \tfrac{1}{2} \log 2\pi e\left[\sigma_s^2 + \sigma_c^2\right] - \tfrac{1}{2} \log 2\pi e \sigma_c^2 \\
&= \tfrac{1}{2} \log \left(1 + \frac{\sigma_s^2}{\sigma_c^2}\right) \ . \qquad (1.37)
\end{aligned}$$

In Eq. (1.37), we see that for $\sigma_s^2 \ll \sigma_c^2, I(W; X) \cong 0$, while if $\sigma_c^2 \to 0$, then $I(W; X) \to \infty$.

1.3 SOURCE CODING THEOREM

We now begin to interpret the quantities developed in Sec. 1.2 within a communications context. For this particular section, we consider the communication system block diagram in Fig. 1.1 under the assumptions that the channel is ideal (no noise or deterministic distortion) and the channel encoder/decoder blocks are identity mappings (a straight wire connection from input to output of each of these blocks). We are left with a communication system block diagram consisting of a source, source encoder, ideal channel, source decoder, and user. Since the channel is ideal and the channel encoder and decoder are identities, we have, with reference to Fig. 1.1, that $V = W = X = Y$. A block diagram of the simplified communication system is shown in Fig. 1.4.

Figure 1.4: Simplified block diagram for the source coding problem.

Just what are the physical meanings of each of the components in Fig. 1.4? The source is some kind of data generation device, such as a computer or computer peripheral, which generates a discrete-valued random variable U with alphabet $\{u_1, u_2, \ldots, u_M\}$ and probability assignment $P_U(\cdot)$. For the development in this section, we assume that the successive letters produced by the source are statistically independent. Such a source is called a *discrete memoryless source* (DMS). The ideal channel can be thought of as some perfectly operating modem or a mass storage device that is error-free. The user is a machine or individual that requires the data accomplish a task. With these descriptions thus far, it is unclear why we need the source encoder/decoder blocks or what utility they might be. The answer is that data at the output of any given discrete source may not be in a form that yields the minimum required transmitted bit rate. The following example illustrates this point.

Example 1.7 Boeringer and Werner (2004) The source output is a ternary-valued random variable that takes on the values $\{u_1, u_2, u_3\}$ with probabilities $P(u_1) = 0.7, P(u_2) = 0.15 = P(u_3)$. The source letter produced at each time instant is assumed to be independent of the letter produced at any other time instant, so that we have a DMS. We wish to find a binary representation of any sequence of ternary source letters such that the source sequence can be recovered exactly (by the decoder) and such that the average number of binary digits per source letter is a minimum.

A straightforward assignment of binary words, called a source code, is to let u_1 be represented by 00, u_2 by 10, and u_3 by 11. This code transmits an average of 2 bits per source letter. Since u_1 occurs much more often than the other two letters, it seems intuitive that we should use a shorter sequence to represent u_1 than those used for u_2 and u_3. One such code assignment is

$$\begin{aligned} u_1 &\to 0 \\ u_2 &\to 10 \\ u_3 &\to 11 \ . \end{aligned}$$

Since this code has the special property (called the *prefix condition*) that no binary word assignment (codeword) is a prefix of any other codeword, the ternary source data can be uniquely recovered from its binary encoding. The average number of bits required per source letter, denoted here by $\bar{n}$, is thus

$$\begin{aligned} \bar{n} &= 1 \cdot P(u_1) + 2 \cdot P(u_2) + 2 \cdot P(u_3) \\ &= 0.7 + 0.3 + 0.3 = 1.3 \text{ bits/source letter} \ . \end{aligned} \tag{1.38}$$

This is a clear improvement over the original 2-bits/source letter code.

To try and reduce $\bar{n}$ further, we encode pairs of source letters, which are listed in Tab. 1.1 together with the probability of each pair. If we now assign a binary word to each pair of source letters as shown in the column labeled "codeword" in Tab. 1.1, we find that the average binary codeword length per source letter is.

$$\begin{aligned} \bar{n} &= \tfrac{1}{2}\{1(0.49) + 3(0.105) + 3(0.105) + 3(0.105) + 4(0.105) + 6(0.0225) \\ &\quad + 6(0.0225) + 6(0.0225) + 6(0.0225)\} \\ &= 1.1975 \text{ bits/source letter} \ . \end{aligned} \tag{1.39}$$

Table 1.1: Velocity autocorrelation function corresponding to different states for the PSO environment.

State	Velocity Autocorrelation
Gas	Damped exponential
Liquid	One period damped sinusoid with single minimum
Solid	Damped sinusoid with possibly multiple-oscillations

State	Velocity Autocorrelation
Gas	Damped exponential
Liquid	One period damped sinusoid with single minimum
Solid	Damped sinusoid with possibly multiple-oscillations

Bit	7	6	5	4	3	2	1	0
0x25	PORTB7	PORTB6	PORTB5	PORTB4	PORTB3	PORTB2	PORTB1	PORTB0
Read/Write	R/W	R/W	R/W	R/W	R/W	R/W	R/W	R/W
Default	0	0	0	0	0	0	0	0

The value in Eq. (1.39) is slightly better than the 1.3 achieved by our second code. The code in Tab. 1.1 is also uniquely decodable back into the original ternary sequence.

Although, we have not described how the binary codes were selected, it is clear that at least in this particular case, source coding allows the ternar data to be represented with a smaller number of bits than one might originally think. Thus, the utility of the source encoder/decoder blocks in Fig. 1.4 is demonstrated.

Example 1.7 demonstrates that source coding techniques can be useful for reducing the bit rate required to represent (exactly) a discrete source, and in Sec. 1.7 we present a constructive technique for designing source codes that is due to Huffman et al. (2005). It is also important to note that in Ex. 1.7 rates are expressed in terms of bits/source letter (or bits/letter). This may be confusing, since prior chapters have discussed data rates primarily in terms of bits/sec. There is no difficulty, however, since bits/letter is just bits/symbol, and hence if we multiply by the number of letters or symbols transmitted per second, the rate in bits/sec results. We thus state all of our rates in this chapter in bits/letter with this understanding in mind. A fundamental question raised by Ex. 1.7 is: what is the minimum bit rate required to represent a given discrete source? The answer is provided by the source coding theorem, which we state in a nonmathematical form here.

Theorem 1.8 (Source Coding Theorem).. *For a DMS with entropy $H(U)$, the minimum average codeword length per source letter ($\bar{n}$) for any code is lower bounded by $H(U)$, that is, $\bar{n} \geq H(U)$, and further, $\bar{n}$ can be made as close to $H(U)$ as desired for some suitably chosen code.*

The (absolute) entropy of a discrete source is thus a very important physical quantity, since it specifies the minimum bit rate required to yield a perfect replication of the original source sequence. Therefore, for the DMS in Ex. 1.7, since

$$\begin{aligned} H(U) &= -0.7 \log 0.7 - 0.15 \log 0.15 - 0.15 \log 0.15 \\ &= 1.18129\ldots\ , \end{aligned} \tag{1.40}$$

we know that $\bar{n} \geq 1.18129\ldots$, and hence only a slight further reduction in rate (from 1.1975 bits/source letter) can be achieved by designing additional source codes (see Sec. 1.7).

1.4 CHANNEL CODING THEOREM

We now turn our attention to the problem of communicating source information over a noisy channel. With respect to the general communication system block diagram in Fig. 1.1, we are presently interested in the channel encoder, channel, and channel decoder blocks. For out current purposes, it is of no interest whether the source encoder/decoder blocks are present or whether the source output is connected directly to the channel encoder input and the channel decoder output is passed directly to the user. What we are interested in here is the transmission of information over a noisy channel. More specifically, we would like to address the question: given the characterization of a communications channel, what is the maximum bit rate that can be sent over this channel with negligibly small error probability? We find that the average mutual information between the channel input and output random variables plays an important role in providing the answer to this question.

Because of the scope of this topic, we limit consideration to *discrete memoryless channels* that have finite input and output alphabets and for which the output letter at any given time depends only on the channel input letter at the same time instant. Therefore, with reference to Fig. 1.1, we define a *discrete memoryless channel* (DMC) with input alphabet $\{w_1, w_2, \ldots, w_M\}$ and probability assignment $P_W(w_j), j = 1, 2, \ldots, M$, and with output alphabet $\{x_1, x_2, \ldots, x_N\}$ and *transition probabilities* $P_{X|W}(x_k|w_j), j = 1, 2, \ldots, M$, and $k = 1, 2, \ldots, N$. From these quantities we can calculate the average mutual information between the channel input W and output X according to Eq. (1.8). We define the capacity of a DMC with input W and output X by

$$\begin{aligned} C &\stackrel{\triangle}{=} \max_{\text{all } P_W(\cdot)} I(W;X) \\ &= \max_{\text{all } P_W(\cdot)} \sum_{j=1}^{M} \sum_{k=1}^{N} P_{WX}(w_j, x_k) \log \frac{P_{WX}(w_j, x_k)}{P_W(w_j) P_X(x_k)} , \end{aligned} \tag{1.41}$$

where the maximum is taken over all channel input probability assignments. We note that $I(W;X)$ is a function of the input probabilities and the transition probabilities, whereas the channel capacity C is a function of the input probabilities only. We have no control over the channel transition probabilities. In words, Eq. (1.41) says that the capacity of a DCM is the largest average mutual information that can be transmitted over the channel in one use. The operational significance of the channel capacity is illustrated further by the following theorem.

Theorem 1.9 (Channel Coding Theorem Liu (2005)). *Given a DMS with entropy H bits/source letter and a DMC with capacity C bits/source letter, if $H \leq C$, the source output can be encoded for transmission over the channel with an arbitrarily small bit error probability. Further, if $H > C$, the bit error probability is bounded away from* 0.

To calculate the channel capacity, it is evident from Eq. (1.41) that we must perform a maximization over M variables, the $P_W(w_j)$, subject to the constraints that $P_W(w_j) \geq 0$ for all j and $\sum_{j=1}^{M} P_W(w_j) = 1$. In general, this is a difficult task. The following example illustrates the calculation of capacity for the special case of what is called a *binary symmetric channel* (BSC).

Example 1.10 A special case of a DMC is the BSC with binary input and output alphabets $\{0, 1\}$ and transition probabilities $P_{X|W}(0|0) = P_{X|W}(1|1) = 1 - p$ and $P_{X|W}(0|1) =$

$P_{X|W}(1|0) = p$, where W is the input random variable and X is the output random variable. A standard diagram for the BSC is shown in Fig. 1.5. We can rewrite the average mutual information in Eq. (1.8) as

$$I(W;X) = \sum_{k=1}^{N} \sum_{j=1}^{M} P_{X|W}\left(x_k|w_j\right) P_W\left(w_j\right) \log \frac{P_{X|W}(x_k|w_j)}{\sum_{l=1}^{M} P_{X|W}(x_k|w_l) P_W(w_l)} . \tag{1.42}$$

Figure 1.5: Binary symmetric channel (see Ex. 1.10).

For the current example $M = N = 2, \{x_1, x_2\} = \{0, 1\}$, and $\{w_1, w_2\} = \{0, 1\}$. We need to write $I(W;X)$ in terms of $P_W(0)$ and $P_W(1) = 1 - P_W(0)$ so that we can perform the maximization. First, we evaluate the denominators of the argument of the logarithm as

$$\begin{aligned}
x_k = 0 : \sum_{l=1}^{2} P_{X|W}(0|w_l) P_W(w_l) &= (1-p)P_W(0) + pP_W(1) \\
&= (1-2p)P_W(0) + p \triangleq \text{den}_0 \\
x_k = 1 : \sum_{l=1}^{2} P_{X|W}(1|w_l) P_W(w_l) &= pP_W(0) + (1-p)P_W(1) \\
&= 1 - p - (1-2p)P_W(0) \triangleq \text{den}_1 \ ,
\end{aligned}$$

since $P_W(1) = 1 - P_W(0)$. Now, expanding Eq. (1.42), we have

$$\begin{aligned}
I(W;X) = {} & (1-p)P_W(0) \log \frac{1-p}{\text{den}_0} + pP_W(1) \log \frac{p}{\text{den}_0} \\
& + pP_W(0) \log \frac{p}{\text{den}_1} + (1-p)P_W(1) \log \frac{1-p}{\text{den}_1} \\
= {} & p \log p + (1-p) \log(1-p) + \big[2pP_W(0) - P_W(0) - p\big] \log(\text{den}_0) \\
& + \big[p + P_W(0) - 2pP_W(0) - 1\big] \log(\text{den}_1) \ ,
\end{aligned} \tag{1.43}$$

where the last equality results from letting $P_W(1) = 1 - P_W(0)$ and simplifying.

Taking the partial derivative of $I(W;X)$ with respect to $P_W(0)$ yields

$$\begin{aligned}\frac{\partial}{\partial P_W(0)} I(W;X) = (2p-1)\log \mathrm{den}_0 + \frac{2pP_W(0)-P_W(0)-p}{\mathrm{den}_0}(1-2p) \\ + (1-2p)\log \mathrm{den}_1 + \frac{p+P_W(0)-2pP_W(0)-1}{\mathrm{den}_1}(2p-1) \\ = -(1-2p)\log \mathrm{den}_0 + (1-2p)\log \mathrm{den}_1 \ , \qquad (1.44)\end{aligned}$$

where the last simplification follows after using the definitions of den_0 and den_1. Equating the partial derivative to 0, we find that $P_W(0) = \frac{1}{2}$. Upon substituting this value back into Eq. (1.43), the capacity of the BSC is found to be

$$C = 1 + p\log p + (1-p)\log(1-p) \qquad (1.45)$$

and is achieved with equally likely inputs. Although, we have not shown that the partial derivative yields a maximum as opposed to a minimum, $I(W;X)$ is a convex $\cap$ (read "cap") function of $P_W(\cdot)$, and hence we have found a maximum. Proofs of such properties of $I(W;X)$ are available elsewhere Ho et al (2005).

We now turn our attention to finding the capacity of discrete-time, memoryless channels with input and output alphabets that consist of the entire set of reals numbers. In other words, the input (W) and output (X) are absolutely continuous random variables here, as opposed to discrete random variables previously. Since the calculation of capacity can be difficult, in general, for continuous-valued random variables, we limit ourselves to the physically important case of the additive Gaussian noise channel with an average power constraint. The channel being considered can thus be represented by the diagram in Fig. 1.6, where the noise ζ has the pdf

$$f_\zeta(\zeta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\zeta^2/2\sigma^2} \ , \qquad (1.46)$$

$-\infty < \zeta < \infty, E[W] = 0$ and $E[W^2] \le S$ (the average power constraint on the input), and each output letter is probabilistically dependent only on the current input letter (the memoryless assumption).

The average mutual information is given by Eq. (1.29). However, for any independent, additive noise channel

$$f_{X|W}(x|w) = f_\zeta(x-w) \ , \qquad (1.47)$$

so

$$\begin{aligned}h(X|W) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_W(w) f_\zeta(x-w)\log f_\zeta(x-w)\,dx\,dw \\ &= -\int_{-\infty}^{\infty} f_W(w)\left\{\int_{-\infty}^{\infty} f_\zeta(\zeta)\log f_\zeta(\zeta)\,d\zeta\right\} dw \\ &= \int_{-\infty}^{\infty} f_W(w) h(\zeta)\,dw = h(\zeta) \ . \qquad (1.48)\end{aligned}$$

Figure 1.6: Additive noise channel.

Therefore, for an additive noise channel,

$$I(W;X) = h(X) - h(\zeta) \ . \tag{1.49}$$

To find the channel capacity, we wish to maximize $I(W;X)$ with respect to the choice of the input pdf $f_W(w)$ subject to the constraint $E[W^2] \le S$. From Eq. (1.49) it is evident that $h(\zeta)$ is independent of the input, and hence only $h(X)$ is affected by the input pdf. Thus, to maximize $I(W;X)$, it is equivalent to maximize $h(X)$.

To complete the calculation of capacity for the additive Gaussian noise channel, the following intermediate results are needed.

Theorem 1.11 (Wang et al (2005).. *For any absolutely continuous random variable ξ, the pdf that maximizes the differential entropy*

$$h(\xi) = -\int f_\xi(\xi)\log f_\xi(\xi)\, d\xi \tag{1.50}$$

subject to the constraint that

$$\int_{-\infty}^{\infty} \xi^2 f_\xi(\xi)\, d\xi \le \sigma_{\max}^2 \tag{1.51}$$

is

$$f_\xi(\xi) = \frac{1}{\sqrt{2\pi}\sigma_{\max}} e^{-\xi^2/2\sigma_{\max}^2} \ . \tag{1.52}$$

Proof. This result can be proved in several ways, including calculus of variations Wang et al (2005) inequality; however, an alternative method is used here. Let $f_\eta(\eta)$ be an arbitrary pdf that satisfies the constraint in Eq. (1.51), and let $f_\xi(\xi)$ be given by Eq. (1.52). Then

$$\begin{aligned} -\int_{-\infty}^{\infty} f_\eta(\alpha)\log f_\xi(\alpha)\, d\alpha &= \int_{-\infty}^{\infty} f_\eta(\alpha)\left\{\log\sqrt{2\pi}\sigma_{\max} + \frac{\alpha^2}{2\sigma_{\max}^2}\log e\right\} d\alpha \\ &= \tfrac{1}{2}\log 2\pi e\sigma_{\max}^2 \ . \end{aligned} \tag{1.53}$$

Now consider

$$h(\eta) - \tfrac{1}{2}\log 2\pi e\sigma_{\max}^2 = \int_{-\infty}^{\infty} f_\eta(\alpha)\log\frac{f_\xi(\alpha)}{f_\eta(\alpha)}\,d\alpha$$
$$\leq \log e \int_{-\infty}^{\infty} f_\eta(\alpha)\left[\frac{f_\xi(\alpha)}{f_\eta(\alpha)} - 1\right] d\alpha = 0\ , \tag{1.54}$$

where the inequality follows from the fact that $\log\beta \leq (\beta-1)\log e$. Thus,

$$h(\eta) \leq \tfrac{1}{2}\log 2\pi e\sigma_{\max}^2 \tag{1.55}$$

with equality if and only if $f_\xi(\alpha)/f_\eta(\alpha) = 1$ for all α. Hence, the theorem follows. □

From Th. 1.11, $I(W;X)$ in Eq. (1.49) is maximized if X is Gaussian with $E[X^2] = S + \sigma^2$ (since $X = W + \zeta$ and $E[W^2] \leq S$); but if X and ζ are Gaussian, then W must be Gaussian. Therefore, the input pdf that achieves channel capacity is

$$f_W(w) = \frac{1}{\sqrt{2\pi S}}\, e^{-w^2/2S}\ , \tag{1.56}$$

so from Eqs. (1.46) and (1.49), the channel capacity of the discrete time, memoryless, additive Gaussian noise channel with an average power constraint on the input is

$$C = \tfrac{1}{2}\log 2\pi e\left(S+\sigma^2\right) - \tfrac{1}{2}\log 2\pi e\sigma^2$$
$$= \tfrac{1}{2}\log\left(1 + \frac{S}{\sigma^2}\right) \text{ bits/source letter}\ . \tag{1.57}$$

The channel capacity given in Eq. (1.57) is a classical, often-quoted result that has considerable intuitive appeal. For instance, if $S/\sigma^2 \ll 1$, then $C \cong 0$, while as $S/\sigma^2 \to \infty, C \to \infty$. Of course, in practical applications, the signal power is constrained, as we assumed in the derivation of Eq. (1.57).

We have only calculated channel capacity for the two special cases of a binary symmetric channel and an additive Gaussian noise channel. The calculation of capacity for other channels can be a tedious and difficult task. A variety of theorems and techniques have been developed to aid in the calculation of capacity and several possibilities are examined in the problems.

1.5 RATE DISTORTION THEORY

In Sec. 1.3 the transmitted data rate required to produce a discrete-time, discrete-amplitude source *exactly* (with no error) is considered, and the minimum rate necessary is shown to be the absolute or discrete entropy. The process of exactly representing a discrete-amplitude source with a reduced or minimum number of binary digits is called *noiseless source coding,* or simply, *source coding.* If the source to be transmitted is a continuous-amplitude random variable or random process, the source has an infinite number of possible amplitudes, and hence the number of bits required to reproduce the source *exactly* at the receiver is infinite. This is indicated by the fact that continuous-amplitude sources have infinite *absolute* entropy. Therefore, to represent continuous-amplitude sources in terms of a finite number of bits/source letter, we must accept the inevitability of some amount of reconstruction error or distortion. We are thus led to the problem of representing a source with a minimum number of bits/source letter subject to a constraint on allowable distortion. This problem is usually called *source coding with a fidelity criterion.* Source coding with respect to some distortion measure may also be necessary for a discrete-amplitude source. For instance, if we are given a DMS with entropy H such that $H > C$ (see Th. 1.9), it may be necessary to accept some amount of distortion in the reproduced version of the DMS in order to reduce the required number of bits/source letter below C.

For source coding with a fidelity criterion, the function of interest is no longer H, but the rate distortion function, denoted $R(D)$. The rate distortion function $R(D)$ with respect to a fidelity criterion is the minimum information rate necessary to represent the source with an average distortion less than or equal to D. To be more specific, we again focus on the block diagram in Fig. 1.4, where the channel is ideal and the channel encoder/decoder blocks are identities, and on discrete memoryless sources. We must choose or be given a meaningful measure of distortion for the source/user pair. If the source generates the output letter u_j and this letter is reproduced at the source decoder output as z_k, we denote the distortion incurred by this reproduction as $d(u_j, z_k)$. The quantity $d(u_j, z_k)$ is sometimes called a *single-letter* distortion measure or fidelity criterion. The average value of $d(\cdot,\cdot)$ over all possible source outputs and user inputs is

$$\bar{d}\left(P_{Z|U}\right) = \sum_{j=1}^{J}\sum_{k=1}^{K} P_U\left(u_j\right) P_{Z|U}\left(z_k|u_j\right) d\left(u_j, z_k\right) \ , \tag{1.58}$$

where the source outputs are $\{u_1, u_2, \ldots, u_j\}$ and the user inputs are $\{z_1, z_2, \ldots, z_k\}$. The average distortion in Eq. (1.58) is a function of the transition probabilities $\{P_{Z|U}(z_k|u_j), j = 1, 2, \ldots, J,\ k = 1, 2, \ldots, K\}$, which are determined by the source encoder/decoder pair. To find the rate distortion function, we wish only to consider those conditional probability assignments $\{P_{Z|U}\}$ that yield an average distortion less than or equal to some acceptable value D, called *D-admissible* transition probabilities, and denoted by

$$\mathcal{P}_D = \left\{P_{Z|U}\left(z_k|u_j\right) : \bar{d}\left(P_{Z|U}\right) \le D\right\} \ . \tag{1.59}$$

For each set of transition probabilities, we have an average mutual information

$$I\left(U;Z\right) = \sum_{j=1}^{J}\sum_{k=1}^{K} P_U\left(u_j\right) P_{Z|U}\left(z_k|u_j\right) \log \frac{P_{Z|U}\left(z_k|u_j\right)}{P_Z\left(z_k\right)} \ . \tag{1.60}$$

We are now able to define the rate distortion function of the source with respect to the fidelity criterion $d(\cdot,\cdot)$ as

$$R(D) = \min_{P_{Z|U} \in \mathcal{P}_D} I\left(U; Z\right) \tag{1.61}$$

for a chosen or given fixed value D. The importance of the rate distortion function is attested to by the fact that for a channel of capacity C, it is possible to reproduce the source at the receiver with an average distortion D if and only if $R(D) \leq C$.

A useful property of $R(D)$ is stated in the following theorem.

Theorem 1.12 *For a DMS with J output letters,*

$$0 \leq R\left(D\right) \leq \log J \ . \tag{1.62}$$

Proof. Using Eqs. (1.11), (1.21), and (1.22),

$$\begin{aligned} 0 \leq I(U;Z) &= H(U) - H(U|Z) \\ &\leq H(U) \leq \log J \ . \end{aligned} \tag{1.63}$$

Now, $R(D)$ is the minimum of $I(U;Z)$ over the admissible conditional probabilities; hence Eq. (1.62) follows. □

The evaluation of the rate distortion function is not straightforward, even for discrete memoryless sources, and therefore we present the following example of a rate distortion function without indicating how it is derived.

Example 1.13 Here we examine a special case of a DMS called a binary symmetric source (BSS) that produces a 0 with probability p and a 1 with probability $1-p$. If we define $u_1 = z_1 = 0$ and $u_2 = z_2 = 1$, the single-letter distortion measure is specified to be

$$d\left(u_j, z_k\right) = \begin{cases} 0, & j = k \\ 1, & j \neq k \ . \end{cases} \tag{1.64}$$

Then, for $p \leq \frac{1}{2}$,

$$\begin{aligned} R(D) = &-p\log p - (1-p)\log(1-p) + D\log D \\ &+ (1-D)\log(1-D), \qquad 0 \leq D \leq p \ . \end{aligned} \tag{1.65}$$

This $R(D)$ is plotted in Fig. 1.7 for $p = 0.1, 0.2, 0.3$, and 0.5. The reader should verify that $R(D) \leq H(U)$ for each p and that $R(p) = 0$.

Figure 1.7: $R(D)$ for a BSS with $p \leq \frac{1}{2}$ (Ex. 1.13). From T. Berger, 1971. *Rate Distortion Theory: A Mathematical Basis for Data Compression, 1971,* Englewood Cliffs, N.J.:Prentice Hall, Inc. Reprinted by permission of the author.

For a discrete-time continuous amplitude source with single-letter distortion measure $d(u, z)$, each conditional pdf relating the source output to the user input produces an average distortion given by

$$\bar{d}\left(f_{Z|U}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_U(u) f_{Z|U}(z|u) d(u,z) du\, dz \tag{1.66}$$

and an average mutual information

$$I(U;Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_U(u) f_{Z|U}(z|u) \log \frac{f_{Z|U}(z|u)}{f_Z(z)} du\, dz \ . \tag{1.67}$$

The admissible pdfs are described by the set

$$\mathcal{P}_D = \left\{ f_{Z|U}(z|u) : \bar{d}\left(f_{Z|U}\right) \leq D \right\} \ .$$

The rate distortion function is then defined as[1]

$$R(D) = \min_{f_{Z|U} \in \mathcal{P}_D} I(U;Z) \ . \tag{1.68}$$

A significant difference between the rate distortion functions for discrete-amplitude and continuous-amplitude sources is that for $R(D)$ in Eq. (1.68), as $D \to 0, R(D) \to \infty$.

Analytical calculation of the rate distortion function for continuous-amplitude sources often is extremely difficult, and relatively few such calculations have been accomplished. We present the results of one such calculation in the following example.

Example 1.14 For the squared-error distortion measure

$$d(u - z) = (u - z)^2 \ , \tag{1.69}$$

[1] Strictly speaking, the "min" in Eq. (1.68) should be replaced with "inf," denoting infimum or greatest lower bound.

a discrete-time, memoryless Gaussian source with zero mean and variance σ_s^2 has the rate distortion function

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma_s^2}{D}, & 0 \le D \le \sigma_s^2 \\ 0, & D \ge \sigma_s^2 \ , \end{cases} \tag{1.70}$$

which is sketched in Fig. 1.8. Note that as $D \to 0, R(D) \to \infty$.

Figure 1.8: $R(D)$ for a memoryless Gaussian source and squared-error distortion (Ex. 1.14).

Equation (1.70) is a relatively simple result that is given added importance by the following theorem.

Theorem 1.15 *Any memoryless, zero-mean, continuous-amplitude source with variance σ_s^2 has a rate distortion function $R(D)$ with respect to the squared-error distortion measure that is upper bounded as*

$$R(D) \le \tfrac{1}{2}\log\frac{\sigma_s^2}{D}, \qquad 0 \le D \le \sigma_s^2 \ . \tag{1.71}$$

Proof. □

Theorem 1.15 thus implies that the Gaussian source is a worst-case source in the sense that it requires the maximum rate of all possible sources to achieve a specified mean square-error distortion. The rate distortion function, when it can be evaluated, provides an absolute lower bound on the performance achievable by these systems.

1.6 APPLICATIONS TO SOURCE AND RECEIVER QUANTIZATION

In this section we present a series of examples that demonstrate how the material in Secs. 1.2 through 1.5 affects realistic communication systems. The first example concerns the coarse quantization of a discrete-time, continuous-amplitude source.

Example 1.16 A memoryless, zero-mean, unit-variance Gaussian source is quantized using a four-level MMSE Gaussian quantizer with the characteristic summarized in Tab. 9.3.1. **(OBS: Needs Table 9.3.1 here.)** This process produces a discrete-time, discrete-amplitude memoryless source with output letters $\{u_1 = -1.510, u_2 = -0.4528, u_3 = 0.4528, u_4 = 1.510\}$. The probability assigned to each of these values must be calculated from the given Gaussian pdf and the quantizer step points in Tab. 9.3.1. That is, since all values of the source from $-\infty$ to -0.9816 are assigned to the output level $-1.510 = u_1$, then

$$P_U(u_1) = \int_{-\infty}^{-0.9816} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = 0.1635\ . \tag{1.72}$$

Similarly, we find that

$$\begin{aligned} P_U(u_2) &= \int_{-0.9816}^{0} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = \tfrac{1}{2} - \int_{-\infty}^{-0.9816} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \\ &= 0.3365 \end{aligned} \tag{1.73}$$

$$P_U(u_3) = \int_{0}^{0.9816} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = 0.3365 \tag{1.74}$$

and

$$P_U(u_4) = \int_{0.9816}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = 0.1635\ . \tag{1.75}$$

Using Eqs. (1.72)–(1.75), we can find the absolute entropy of this newly created DMS from Eq. (1.2) as 1.911 bits/source letter. Three interesting observations can be made concerning these results. First, the quantizer has accomplished entropy reduction in that it has transformed a continuous amplitude source with infinite absolute entropy into a DMS with a finite entropy of 1.911 bits/source letter. Second, the minimum bit rate required to represent the quantizer outputs exactly is 1.911 bits/source letter, which does not seem to be significantly less than the 2 bits/source letter needed by the NBC or FBC. Hence, the extra effort required to achieve the minimum bit rate using coding procedures as illustrated in Ex. 1.7 or in Sec. 1.7 may not be worthwhile. **(OBS: Need Table 9.3.1 here.)** Third, from Tab. 9.3.1, the mean-squared error distortion achieved by this quantizer is $D = 0.1175$ at a minimum rate of 1.911 bits/source letter. This result can be compared to the rate distortion function of a memoryless Gaussian source to ascertain how far away from optimum this quantizer operates.

As a second example, we illustrate the calculation of channel capacity for binary antipodal signals transmitted over an AWGN with a particular receiver.

Example 1.17 Here we examine the transmission of binary messages over an AWGN channel with zero mean and variance $\sigma_c^2 = 1$ using BPSK signals as given by Eq. (10.2.18a) for message m_1 **(OBS: Needs Eqs. 10.2.18a and b here.)** and Eq. (10.2.18b) for message m_2. The optimum receiver in terms of minimizing the probability of error has a threshold at 0 and decides m_1 if the received signal is positive and decides m_2 if the received signal is negative. Thinking of the binary messages as channel inputs, say w_1 and w_2, and the binary decisions of the receiver as channel outputs, say x_1 and x_2, we have a discrete memoryless channel. To complete the specification of the channel, we must determine the input-output transition probabilities.

If we let $A_c = 1$ in **(OBS: Need Eqs. 10.2.18a and b here.)** Eqs. (10.2.18a) and (10.2.18b), the received signal when m_1 is transmitted is $r = 1 + \eta$, where η is the Gaussian noise variable, and when m_2 is transmitted the received signal is $r = -1 + \eta$. The transition probabilities for the DMS are thus

$$P_{X|M}\left(x_1|m_1\right) = \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(r-1)^2/2}\, dr = 0.8413 \tag{1.76}$$

$$P_{X|M}\left(x_1|m_2\right) = \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(r+1)^2/2}\, dr = 0.1587 \tag{1.77}$$

$$P_{X|M}\left(x_2|m_1\right) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-(r-1)^2/2}\, dr = 0.1587 \tag{1.78}$$

and

$$P_{X|M}\left(x_2|m_2\right) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-(r+1)^2/2}\, dr = 0.8413\ . \tag{1.79}$$

From Eqs. (1.76)–(1.79) it is evident that we have a BSC with $p = 0.1587$. It is known from Ex. 1.10 that the capacity for this channel is given by Eq. (1.45), so

$$\begin{aligned} C &= 1 + (0.1587)\log(0.1587) + (0.8413)\log(0.8413) \\ &= 0.369 \text{ bit/source letter} \end{aligned} \tag{1.80}$$

and is achieved by equally likely channel inputs.

In this third example, the rate distortion functions of several memoryless continuous-amplitude sources are presented and the performance of optimum quantizers in comparison to their rate distortion bound is discussed.

Example 1.18 Memoryless, discrete-time, continuous-amplitude sources with uniform, Gaussian, Laplacian, and gamma distributions are of importance in a variety of practical applications. Typical pdfs for these sources are listed below.

Uniform pdf:

$$f_U(u) = \begin{cases} \frac{1}{2}, & -1 \leq u \leq 1 \\ 0 & \text{otherwise} . \end{cases} \tag{1.81}$$

Gaussian pdf:

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \qquad -\infty < u < \infty . \tag{1.82}$$

Laplacian pdf:

$$f_U(u) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|u|}, \qquad -\infty < u < \infty . \tag{1.83}$$

Gamma pdf:

$$f_U(u) = \frac{\sqrt{3}}{4\sqrt{\pi|u|}} e^{-\sqrt{3}|u|/2}, \qquad -\infty < u < \infty . \tag{1.84}$$

The rate distortion functions for sources with the pdfs in Eqs. (1.81)–(1.84) are shown in Fig. 1.9. In Tab. 1.1, the minimum-distortion Gaussian and Laplacian quantizer performance at rates 1, 2, and 3 bits/source letter are compared to the rate distortion functions for Gaussian and Laplacian sources in terms of SQNR $= 10 \log_{10} \sigma_s^2/D$. The quantizer outputs are assumed to be noiselessly encoded at their minimum possible rate (see Sec. 1.7), which is the absolute entropy of the output letters. To close the performance gap further requires more exotic coding techniques. Except for the Gaussian case, the results in Fig. 1.9 and Tab. 1.2 were obtained numerically.

These three examples demonstrate the utility of the theory in Secs. 1.2 to 1.5 to familiar communications problems from earlier chapters. **(OBS: Mentions EARLIER chapters.** As can be seen, information theory and rate distortion theory can be extremely useful for obtaining bounds on communication system performance, and hence can indicate whether further improvements in system design are likely to be worth the time, effort, and/or complexity involved.

1.7 VARIABLE-LENGTH SOURCE CODING

We mention in Sec. 9.5 **(OBS: There is no Section 9.5 in this chapter)** that if the output quantization levels are not equally likely to occur, the average bit rate required may be reduced in comparison to the NBC or FBC by using a variable-length code. Theorem 1.8 states that a lower bound on the average codeword length per source letter, denoted by $\bar{n}$, is the source entropy. Furthermore, Ex. 1.1 illustrates that a DMS with letters that are

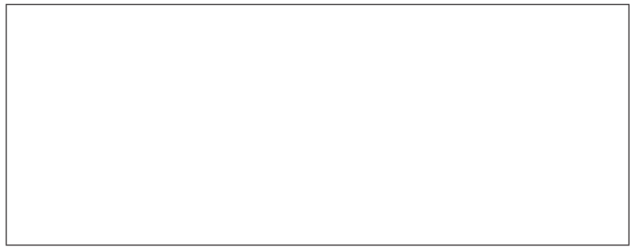

Figure 1.9: Rate distortion functions for (a) Gaussian, (b) Laplacian, (c) Uniform, and (d) Gamma distributed sources. From P. Noll and R. Zelinski, "Bounds on Quantizer Performance in the Low Bit-Rate Region," *IEEE Trans. Commun.,* ©1978 IEEE.

Table 1.2: SQNR (dB) Comparison of $R(D)$ and Optimum Quantizer Performance for Gaussian and Laplacian Sources

Rate (bits/source letter)	Gaussian Source $R(D)$	Gaussian Source Optimum Quantizer	Laplacian Source $R(D)$	Laplacian Source Optimum Quantizer
1	6.02	4.64	6.62	5.76
2	12.04	10.55	12.66	11.31
3	18.06	16.56	18.68	17.20

Source: N. Farvardin and J. W. Modestino, "Optimum Quantizer Performance for a Class of Non-Gaussian Memoryless Sources," *IEEE Trans. Inf. Theory,* ©1984 IEEE.

not equally likely has a smaller entropy than a source with the same number of letters and equally probable outputs. Finally, in Ex. 1.7, a variable-length code is constructed for a DMS with nonequally likely output letters.

The design of variable-length codes with an $\bar{n}$ that approaches the entropy of the DMS is generically referred to as *entropy coding,* and there are several procedures for finding such codes. In this section we present the most familiar and most straightforward of the available techniques for entropy coding, which is due to Huffman and is thus called *Huffman coding.* To specify the encoding procedure, we consider a DMS U with M output letters $\{u_1, u_2, \ldots, u_M\}$ and probabilities $P_U(u_j), j = 1, 2, \ldots, M$. We also assume for simplicity that the letters are numbered such that $P_U(u_1) \geq P_U(u_2) \geq \cdots \geq P_U(u_M)$. Of course, if this property does not hold at the outset, we can always renumber the letters to produce it. The constructive procedure for designing the variable-length code can be described as follows.

The letters and their probabilities are listed in two columns in the order of decreasing probability. The two lowest-probability letters are combined by drawing a straight line out from each and connecting them. The probabilities of these two letters are then summed, and this sum is considered to be the probability of a new letter denoted u'_{M-1}. The next two lowest-probability letters, among $u_1, u_2, \ldots, u_{M-2}$, and u'_{M-1}, are combined to create another letter with probability equal to their sum. This process is continued until only two letters remain and a type of "tree" is generated. Binary codewords are then assigned by moving from right to left in the tree, assigning a 0 to the upper branch and a 1 to the lower branch, where each pair of letters has been combined. The codeword for each letter is read off the tree from right to left. An example will greatly clarify the procedure.

Example 1.19 Given a DMS U with five letters and probabilities $P_U(u_1) = 0.5$, $P_U(u_2) = P_U(u_3) = 0.2$, and $P_U(u_4) = P_U(u_5) = 0.05$, we wish to design a binary variable-length code for this source. With reference to Fig. 1.10, we combine the two least likely letters and sum their probabilities to create a new "letter." The least likely letters are combined again,

Figure 1.10: Huffman encoding for the DMS in Ex. 1.19.

and the procedure is continued until only two letters remain. Codewords are then assigned by moving right to left and assigning a 0 to an upper branch and a 1 to a lower branch. Codewords are also read off in a right-to-left fashion and are shown in the leftmost column of Fig. 1.10. The average codeword length ($\bar{n}$) for this variable-length code is

$$\begin{aligned} \bar{n} &= (4)(0.05) + (4)(0.05) + (3)(0.2) + (2)(0.2) + (1)(0.5) \\ &= 1.9 \text{ bits/source letter} . \end{aligned} \tag{1.85}$$

We know from Eq. (1.21) that $H(U)$ is upper bounded by $\log_2 M = \log_2 5 \cong 2.322$ bits/source letter, and we can show that the entropy of U is $H(U) \cong 1.86$ bits/source letter. Thus, $H(U) < \bar{n} < \log_2 M$.

The codewords in Ex. 1.19 are *uniquely decodable* in that each source letter has a codeword that differs from that assigned to any other letter. *The Huffman procedure yields the smallest average codeword length of any uniquely decodable set of codewords.* Although another code can do as well as the Huffman code, none can be better.

Recall that in Ex. 1.7, a code is designed for single-letter encoding of the source, and then another code is designed to represent pairs of source letters. Should we use Huffman coding on pairs of letters for the source in Ex. 1.19? Probably not, since $\bar{n}$ is already very close to the entropy. However, Ex. 1.7 demonstrates that by encoding blocks of source letters, an average codeword length nearer the source entropy can be obtained (with added complexity). A more general illustration of this property is provided by the following two inequalities. The Huffman encoding procedure generates a code for the DMS U with an $\bar{n}$ that satisfies

$$H(U) \leq \bar{n} < H(U) + 1 \text{ bits/source letter} \tag{1.86}$$

for letter-by-letter (or symbol-by-symbol) encoding. If blocks of L letters are combined before using the Huffman technique, $\bar{n}$ is bounded by

$$H(U) \leq \bar{n} < H(U) + \frac{1}{L} \text{ bits/source letter .} \tag{1.87}$$

Thus, encoding pairs of letters is at least as good as single-letter encoding, and for large $L, \bar{n}$ can be made arbitrarily close to $H(U)$. Whether block encoding makes sense depends on the particular DMS and its letter probabilities (and perhaps, the application of interest).

Only the binary Huffman procedure has been described here, but nonbinary codes can be designed using the Huffman method. The details are somewhat more complicated and nonbinary codes are less commonly encountered than binary ones, so further discussion is left to the problems and the literature.

Table 1.3: Timer0 Compare Output Mode, non-PWM Mode

COM0x1-0	Description
00	Normal port operation
01	Toggle on Compare Match
10	Clear on Compare Match
11	Set on Compare Match

COM0x1-0	Description
00	Normal port operation
01	Toggle on Compare Match
10	Clear on Compare Match
11	Set on Compare Match

SUMMARY

In this chapter we have discussed very briefly some of the salient results from information theory and rate distortion theory and have indicated how these results can be used to bound

Figure 1.11:

communication system performance. It is perhaps surprising to the reader that physically meaningful quantitative measures of information can be defined, but such is the case. It has been impossible to present and to develop the many elegant and insightful results available from information theory, and hence a clear view of the importance of this field may not be available to the reader. However, information theory and rate distortion theory provide the theoretical basis and performance bounds for the practical source coding and channel coding systems in use today.

PROBLEMS

1.1 A random variable U has a sample space consisting of the set of all possible binary sequences of length N, denoted $\{u_j, j = 1, 2, \ldots, 2^N\}$. If each of these sequences is equally probable, so that $P[u_j] = 2^{-N}$ for all j, what is the self-information of any event $u = u_j$?

1.2 Given a random variable U with the alphabet $\{u_1, u_2, u_3, u_4\}$ and probability assignments $P(u_1) = 0.8, P(u_2) = 0.1$, $P(u_3) = 0.05, P(u_4) = 0.05$, calculate the entropy of U. Compare your result to a random variable with equally likely values.

1.3 Given the binary erasure channel (BEC) shown in Fig. 1.11, find an expression for the average mutual information between the input and output $I(W;X)$ if $P_W(w_1) = P_W(w_2) = \frac{1}{2}$. The BEC might be a good channel model for the physical situation where binary antipodal signals are transmitted and the receiver makes a decision if the received signal is much greater than or much less than the threshold, but asks for a retransmission of the received signal if the received signal is very near the threshold.

1.4 For the DMC in Fig. 1.12 with $P_W(0) = \frac{1}{3}$ and $P_W(1) = \frac{2}{3}$, find $H(W)$ and $H(W|X)$. What is the average mutual information for this channel and input probability assignment?

1.5 Calculate the differential entropy for an absolutely continuous random variable with pdf $f_U(u) = (1/\chi)e^{-u/\chi}, 0 < u < \infty$, and $f_U(u) = 0$ for $u \leq 0$.

Figure 1.12:

1.6 Show that the differential entropy for a random variable U with the Laplacian pdf $f_U(u) = (1/\sqrt{2})e^{-\sqrt{2}|u|}$ for $-\infty < u < \infty$ is given by $h(U) = \log(e\sqrt{2})$.

1.7 Given an absolutely continuous random variable X with pdf $f_X(x)$ and the transformation $Y = aX + b$, find an expression for the entropy of Y in terms of $h(X)$. How has the transformation affected the result?

1.8 Consider the discrete random variable X in Ex. 1.1 with $P_X(x_1) = \frac{1}{2}, P_X(x_2) = \frac{1}{4}$, $P_X(x_3) = P_X(x_4) = \frac{1}{8}$, and the linear transformation $Y = aX + b$. Find $H(Y)$. What effect has the transformation had on the entropy of the discrete random variable X?

1.9 For a general multivariated pdf of absolutely continuous random variables denoted by $f_X(x_1, x_2, \ldots, x_N)$, consider the one-to-one transformation represented by $y_i = g_i(X), i = 1, 2, \ldots, N$. Find an expression for the differential entropy of the joint pdf of Y, $f_Y(y_1, y_2, \ldots, y_N)$ in terms of the Jacobian of the transformation. See Eqs. (A.5.4) and (A.5.5). **(OBS: No Eqs. A.5.4 and A.5.5 in this chapter.)**

1.10 Use Eq. (1.29) and the result of Problem 1.9 to show that average mutual information is invariant to one-to-one transformations. That is, given two continuous random vector X and Z with average mutual information $I(X;Z)$ and a one-to-one transformation $y_i = g_i(X), i = 1, 2, \ldots, N$, find $I(Y;Z)$.

1.11 Given a Gaussian source U with mean μ_s and variance σ_s^2, find an expression for its differential entropy $h(U)$.

1.12 A DMS U has $P_U(u_1) = 0.4, P_U(u_2) = 0.3$, and $P_U(u_3) = P_U(u_4) = 0.15$. Construct a source code for U by mimicking the procedure in Ex. 1.7 and Tab. 1.2.

1.13 For the BSC in Ex. 1.10 plot $I(W;X)$ as a function of p when $P_W(0) = \frac{3}{4}$ and $P_W(1) = \frac{1}{4}$ and compare to a plot of C given by Eq. (1.45).

1.14 For a one-sided pdf $f_X(x)$ such that $f_X(x) = 0$ for $x \leq 0$ with mean

$$\mu = \int_0^\infty x f_X(x)\, dx \ ,$$

show that the maximum differential entropy is achieved when

$$f_X(x) = \frac{1}{\mu} e^{-x/\mu} u(x)$$

and that $h(X) = \log \mu e$.

1.15 An often useful approach to finding channel capacity for discrete memoryless channels relies on the following theorem.

Theorem 1.20 *A DMC has input W and output X with transition probabilities $P_{X|W}(x_k|w_j), j = 1, \ldots, M, k = 1, 2, \ldots, N$. Necessary and sufficient conditions for a set of input probabilities $P_W(w_j), j = 1, 2, \ldots, M$, to achieve capacity is for*

$$I\left(w = w_j; X\right) = C \qquad \text{for all } w_j \text{ with } P_W\left(w_j\right) > 0$$

and

$$I\left(w = w_j; X\right) \leq C \qquad \text{for all } w_j \text{ with } P_W\left(w_j\right) = 0 \ ,$$

for some number C, where

$$I\left(w = w_j; X\right) = \sum_{k=1}^{N} P_{X|W}\left(x_k|w_j\right) \log \frac{P_{X|W}(x_k|w_j)}{\sum_{j=1}^{M} P_W(w_j) P_{X|W}(x_k|w_j)} \ .$$

The number C is the channel capacity.

The main use of this theorem is to check the validity of some hypothesized set of input probabilities. Thus, for the BSC in Ex. 1.10, we might guess by symmetry that $P_W(0) = P_W(1) = \frac{1}{2}$ achieves capacity. Substantiate this claim and find the capacity in Eq. (1.45) by using this theorem.

1.16 Use the theorem in Problem 1.15 to show that the capacity of the binary erasure channel in Fig. 1.11 with $p_1 = p_2$, called the binary *symmetric* erasure channel (BSEC), is p_1.

Hint: Guess equally likely inputs.

1.17 For the channel in Fig. 1.12, find channel capacity using the theorem in Problem 1.15.

1.18 We have only calculated the capacity of channels with continuous inputs and outputs when the noise is Gaussian. The following theorem expands the utility of the Gaussian result.

Theorem 1.21 *Consider the additive noise channel in Fig. 1.6, where $E[W^2] \leq S$ and $var(\varsigma) = \sigma^2$. The capacity of this channel is bounded by*

$$\tfrac{1}{2} \log \left(1 + \frac{S}{\sigma^2}\right) \leq C \leq \tfrac{1}{2} \log \left[2\pi e \left(S + \sigma^2\right)\right] - h(\varsigma) \ .$$

In essence, this theorem says that for a fixed noise variance, Gaussian noise is the worst since it lower bounds the channel capacity.

(a) Use Eq. (1.49) to prove the right inequality.

(b) Follow the proof of Th. 1.11 to prove the left inequality.

1.19 Jensen's inequality states that for a random variable W with a distribution defined on an appropriate interval and for a convex $\cup$ function, say $g(x)$, then

$$E[g(W)] \geq g[E(W)]$$

if $E[W]$ exists. If $g(x)$ is convex $\cap$, the inequality is reversed to yield

$$E[g(W)] \leq g[E(W)] \ .$$

Use Jensen's inequality to prove Th. 1.11. Jensen's inequality is particularly useful when working with average mutual information, since $I(W;X)$ is a convex $\cup$ function of the transition probabilities, and $I(W;X)$ is a convex $\cap$ function of the input probabilities.

1.20 The rate distortion function for a Laplacian source U with pdf

$$f_U(u) = \frac{\lambda}{2} e^{-\lambda|u|}, \qquad -\infty < u < \infty$$

subject to the absolute value of the error distortion measure $d(u-z) = |u-z|$ is

$$R(D) = -\log \lambda D, \qquad 0 \leq D \leq \frac{1}{\lambda}$$

Plot this rate distortion function.

1.21 A memoryless, zero-mean, unit-variance Gaussian source is quantized using the MMSE Gaussian quantizer characteristic in Tab. 9.3.2. **(OBS: Need Table 9.3.2)**. Find the equivalent DMS and its entropy. Compare the performance of this quantizer to the rate distortion bound.

1.22 Use the eight-level uniform quantizer in Tab. 9.2.1 **(OBS: Need Table 9.2.1.)** to quantize a memoryless, zero-mean, unit-variance Gaussian source. Find the output entropy and compare its performance to the rate distortion bound.

1.23 Compare the results of Problems 1.21 and 1.22.

1.24 Binary messages are transmitted over an AWGN channel with zero mean and variance $\sigma_c^2 = \frac{1}{2}$ using BPSK signals as given by Eq. (10.2.18a) **(OBS: Need Eqs. 10.2.18a and b.)** for message m_1 and Eq. (10.2.18b) for message m_2. The receiver decides m_1 if the received signal is positive and m_2 if the received signals is negative. Considering the binary messages as channel inputs, w_1 and w_2, and the binary receiver decisions as outputs, x_1 and x_2, specify the equivalent DMC if $A_c = 1$ in Eqs. (10.2.18a,b). **(OBS: Need Eqs. 10.2.18a and b.)** Calculate the channel capacity.

1.25 Repeat Problem 1.24 for the case where the receiver has the four output values $\{x_1, x_2, x_3, x_4\}$ and where x_1 occurs if the received value $r > 1.032$, x_2 occurs if $0 < r < 1.032$, x_3 occurs if $-1.032 < r < 0$, and x_4 occurs if $r < -1.032$.

1.26 Given a DMS U with four letters and probabilities $P_U(u_1) = 0.5$, $P_U(u_2) = 0.25$, and $P_U(u_3) = P_U(u_4) = 0.125$, use the Huffman procedure to design a variable-length code. Find $\bar{n}$ and compare to $H(U)$.

1.27 Use the Huffman procedure to design variable-length codes for the single-letter and paired-letter sources in Ex. 1.7 .

1.28 Plot the upper bound on $\bar{n}$ in Eq. (1.87) as a function of L for the sources in Exs. 1.7 and 1.19.

CHAPTER 2

Jordan Canonical Form

2.1 THE DIAGONALIZABLE CASE

Although, for simplicity, most of our examples will be over the real numbers (and indeed over the rational numbers), we will consider that *all of our vectors and matrices are defined over the complex numbers* $\mathbb{C}$. It is only with this assumption that the theory of Jordan Canonical Form (JCF) works completely. See Remark 2.4 for the key reason why.

Definition 2.1 If $v \neq 0$ is a vector such that, for some λ,

$$Av = \lambda v \ ,$$

then v is an *eigenvector* of A associated to the *eigenvalue* λ.

Example 2.2 Let A be the matrix $A =$ Then, as you can check, if $v_1 =$ then $Av_1 = 3v_1$, so v_1 is an eigenvector of A with associated eigenvalue 3, and if $v_2 =$ then $Av_2 = -2v_2$, so v_2 is an eigenvector of A with associated eigenvalue -2.

We note that the definition of an eigenvalue/eigenvector can be expressed in an alternate form. Here I denotes the identity matrix:

For an eigenvalue λ of A, we let E_λ denote the *eigenspace* of λ,

$$E_\lambda = \{v \mid Av = \lambda v\} = \{v \mid (A - \lambda I)v = 0\} =$$

(The kernel is also known as the nullspace

We also note that this alternate formulation helps us find eigenvalues and eigenvectors. For if $(A - \lambda I)v = 0$ for a nonzero vector v, the matrix $A - \lambda I$ must be singular, and hence its determinant must be 0. This leads us to the following definition.

Definition 2.3 The *characteristic polynomial* of a matrix A is the polynomial

Remark 2.4 This is the customary definition of the characteristic polynomial. But note that, if A is an n-by-n matrix, then the matrix $\lambda I - A$ is obtained from the matrix $A - \lambda I$ by multiplying each of its n rows by -1, and hence In practice, it is most convenient to work with $A - \lambda I$ in finding eigenvectors—this minimizes arithmetic—and when we come to find chains of generalized eigenvectors in Section 1.2, it is (almost) essential to use $A - \lambda I$, as using $\lambda I - A$ would introduce lots of spurious minus signs.

Example 2.5 Returning to the matrix $A =$ of Example 2.2, we compute that $\lambda^2 - \lambda - 6 = (\lambda - 3)(\lambda + 2)$, so A has eigenvalues 3 and -2. Computation then shows that the eigenspace $E_3 =$ and that the eigenspace $E_{-2} =$

Bye.

CHAPTER 3

Citations with Author-Year

An alternative to numerical LaTeX citations makes reference to published works by citing the author's name and year of publication. The entries in the bibliographic listing are not numbered. The citation itself may be either parenthetical like (Carroll et al., 1998) or textual as shown by Carroll et al. (1998).

The `natbib` package by Patrick W. Daly is the most universal package for Author-Year citation and it is compatible with `\bibitem` syntaxes and with the `.bst` files of author-year packages such as `apalike`, `chicago`, and `harvard`.

If you create a bibliography by hand, please see the file `biblio.tex` in the `latex` directory of `bib-AY.zip`. The `biblio.tex` files shows the syntax that `natbib.sty` requires and notice how we use that package in `book.tex` and in `ch01.tex`. You can find the details about `natbib` and Author-Year bibliographies in Kopka and Daly (2004, pages 218-221).

Note that we have bibliographic references like `carroll98`, `KD:2004`, and `weber-97`. Pick a style that works for you and stay with it, then it will be easier for yourself and others to maintain and modify your bibliography and citations.

If you use `.bib` files then you should know about `BibTeX`, `.bst` files, and you should know how to create the `.bbl` files as well. (If you use `.bib` files but you do not know the rest, you should consult a local expert before continuing with `.bib` or create your bibliography by hand.) If you know the steps we mentioned above and you wish to create an Author-Year bibliography and citations, then you are in luck: there are three `.bst` that are provided with `natbib`. They are: `plainnat`, `unsrtnat`, and `abbrvnat`.

Once you create your `.bbl` file(s), please submit them along with your manuscript files, style files, and (preferably) `.eps` figures with `BoundingBoxes` to *Morgan and Claypool Publishers.*

You can get a free copy of the `natbib` package from CTAN at

`http://www.ctan.org/tex-archive/macros/latex/contrib/natbib/`

and download the file `natbib.zip`.

Here are some more samples using the citations that appear in the file `biblio.tex` that we built by hand.

The next paragraph uses text citations using `\citet` (cite as text):
There are several textbooks that give a general introduction to dependency grammar but most of them in other languages than English, for example, Tarvainen (1982) and Weber (1997).

The last paragraph uses text citations with parentheses using `\citetp`:
Tesnière's seminal work was published posthumously as (Tesnière , 1959). Other influential theories in the dependency grammar tradition include Functional Generative Description (Sgall et al., 1986).

Please see the bibliography on the next page and the file `biblio.tex`.

Bibliography

D. Boeringer and D. Werner, “Particle swarm optimization versus genetic algorithms for phased array synthesis,” *IEEE Trans. Antennas Propagat.,* vol. 52, no. 3, pp. 771–779, 2004.

Carroll, John, Briscoe, Edward, and Sanfilippo, Antonio Supertagging: An approach to almost parsing, *Computational Linguistics* 25(2):237–267.

S. Ho, S. Yang, G. Ni, E. W. C. Lo, and H. C. Wong, “A particle swarm optimization-based method for multiobjective design optimizations,” *IEEE Trans. Magn.*, vol. 41, no. 5, pp. 1756–1759, 2005.

Kopka, Helmut and Daly, Patrick W. *Guide to LaTeX,* fourth edition, Addison Wesley.

F. S. Levin, *An Introduction to Quantum Theory*, Cambridge: Cambridge University Press, 2002.

W-C Liu, “A design of a multiband CPW-fed monopole antenna using a particle swarm optimization approach,” *IEEE Trans. Antennas Propagat.,* vol. 53, no. 10, pp. 3273–3279, 2005.

Sgall, Petr, Hajičová, Eva, and Panevová, Jarmila. *The Meaning of the Sentence in Its Pragmatic Aspects*, Reidel.

Tarvainen, Kalevi. *Einführung in die Dependenzgrammatik*, Niemeyer.

Tesnière, Lucien. *Éléments de syntaxe structurale*, Editions Klincksieck.

Weber, H. J. *Dependenzgrammatik. Ein interaktives Arbeitsbuch*, Günter Narr.

W. Wang, Y. Lu, J. S. Fu, and Y. Z. Xiong, “Particle swarm optimization and finite-element based approach for microwave filter design,” *IEEE Trans. Magn.*, vol. 41, no. 5, pp. 1800–1803, 2005.

Author's Biography

YOUR NAME

Your Name Your short bio goes here...