Optics and Information Theory

FRANCIS T. S. YU

Optics and Information Theory

72
4

Optics and Information Theory

FRANCIS T. S. YU
Wayne State University

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, New York · London · Sydney · Toronto

DS56/23

Copyright © 1976 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Yu, Francis T. S. 1934–
Optics and information theory.

"A Wiley-Interscience publication."
Includes bibliographical references and index.
1. Information theory. 2. Optics. 3. Optical data processing. I. Title. [DNLM: 1. Optics. 2. Information theory QC355.2 Y94o]
Q360.Y8 621.38'0414'01 76-23135
ISBN 0-471-01682-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Preface

Recent advances in quantum electronics have brought into use the infrared and visible range of electromagnetic waves. They now permitted us to build new systems for the application of optical information processing and communication. The impact of laser communication and wave-front reconstruction has provided an interesting relationship between optics and information theory, a trend that has grown quite rapidly since the invention of intensive coherent light sources in the early 1960s. Light not only provides a major source of energy, but also is a very important source of information. Therefore it is my purpose in writing this book to bring into closer view this intimate relationship between optics and information theory. The contents of this book were mainly derived from several classical articles, particularly those by Gabor and Brillouin.

The manuscript of this book has been used as lecture notes in my classes in optics and information theory, and the material was chosen to fit the general interest of my students in electrical engineering. However, the book may also serve interested physicists and members of technical staffs. The book's eight chapters range from basic information theory, optics and information to the quantum effect on information transmission. The basic approach centers around the entropy theory of information.

The contents of this book have been used in a one-quarter course in optics and information theory at Wayne State University. Most of the students were in their first year of graduate studies. I have found that it is occasionally possible to teach the whole book without significant omissions, and with very limited additional material the text may be used in a full-semester course. The book in its present form is not intended to cover the vast domain of optics and information theory, but is restricted to an area I consider particularly important and interesting.

I believe that optics and information theory are at the threshold of a major technical revolution. Much remains to be done before optical information and transmission can become a widespread practical reality. The basic requirement for rapid progress in optical information and transmission should be to begin with careful imaginative experimental work based on a deep appreciation of the theoretical foundations that have already been established in part.

In view of the great number of contributors in this area, I apologize for possible omission of appropriate references in various parts of this book. The excellent article "Light and Information" by Gabor and the book *Science and Information Theory* by Brillouin deserve special mention. I am deeply indebted to these two authors.

I am grateful also to Dr. H. K. Dunn, retired member of the technical staff of Bell Telephone Laboratories, for his encouragement, criticism, and technical support during the preparation of the manuscript. I also express my appreciation to Mr. A. Tai and Mr. T. Cheng, for proofreading and for preparing illustrations; Mrs. Sylvia Wasserman and Miss Mai Chen, for their excellent typing of the manuscript; Mrs. K. Y. Ma, for her encouragement and for proofreading most parts of the manuscript; my students, for their constant interest and motivation; and finally, to my wife and children, for their unbounded love, patience, and encouragement.

Detroit, Michigan
June 1976

FRANCIS T. S. YU

试读结束：需要全本请在线购买： www.ertongbook.com

Contents

1

Introduction to Information Transmission

In the physical world light is not only part of the mainstream of energy, that supports life, but also provides us with important sources of information. One can easily imagine that without light present civilization could never exist. Furthermore, humans are equipped with a pair of exceptionally good, although not perfect, eyes. With the combination of an intelligent brain and remarkable eyes, humans were able to advance themselves above the rest of the animals in the world. It is undoubtedly true that, if humans had not been equipped with eyes, they would not have evolved into their present form. In the presence of light, humans are able to search for the food they need and the art they enjoy, and to explore the unknown. Thus light, or rather *optics*, has provided us with a very useful source of information whose application can range from very abstract artistic to very sophisticated scientific uses.

The purpose of this text is to discuss the relationship between optics and information transmission. However, it is emphasized that it is not our intention to consider the whole field of optics and information theory, but rather to center on an area that is important and interesting to us.

Prior to going into a detailed discussion of optics and information, we devote this first chapter to the fundamentals of information transmission. However, it is noted that *information theory* was not originated by optical physicists, but rather by a group of mathematically oriented electrical engineers whose original interest was centered on electrical communication. Nevertheless, from the very beginning of the discovery of information theory, interest in the application has never totally been absent from the optical standpoint. As a result of the recent advances in modern optical information processing and optical communication, the relationship between optics and information theory has grown more rapidly than ever.

Although everyone seems to know the word information, a fundamen-

5506146

tal theoristic concept may not be the case. Let us now define information theory. Actually, information may be defined in relation to several different disciplines. In fact, information may be defined according to its applications but with the identical mathematical formalism as developed in the next few sections. From the viewpoint of pure mathematics, information theory is basically *probability theory*. We see in Sec. 1.1 that without probability there would be no information theory. But, from a physicist's point of view, information theory is essentially an *entropy theory*. In Chapter 4, we see that without the fundamental relationship between entropy and information theory, information theory would have no useful application in physical science. From a communication engineer's standpoint, information theory can be considered an *uncertainty theory*. For example, the more uncertainty there is about a message we have received, the greater the amount of information the message contained.

Since it is not our intention to define information theory for all fields of interest, we quickly summarize: The beauty and greatness of information theory is its' application to all fields of science. Application can range from the very abstract (e.g., music, biology, psychology) to very sophisticated scientific research. However, in our present introductory version, we consider a concept of information from a practical communication standpoint. For example, from the information theory viewpoint, a perfect liar is as good an informant as a perfectly honest person, provided of course that we have the a priori knowledge that the person is a perfect liar or perfectly honest. One should be cautious not to conclude that, if one cannot be an honest person, one should be a liar. For, as we may all agree, the most successful crook is the one that does not look like one. Thus we see that information theory is a guessing game, and is in fact a *game theory*.

In general, an information system can be represented by a block diagram, as shown in Fig. 1.1. For example, in simple optical communication, we have a message (an information source) shown by means of written characters, for example, Chinese, English, French, German. Then we select suitable written characters (a code) appropriate to our communication. After the characters are selected and written on a piece of paper, the information still cannot be transmitted until the paper is illuminated by visible light (the transmitter), which obviously acts as an information carrier. When light reflected from the written characters arrives at your eyes (the receiver), a proper decoding (translating) process takes place, that is, character recognition (decoding) by the user (your mind). Thus, from this simple example, we can see that a suitable encoding process may not be adequate unless a suitable decoding process also takes place.

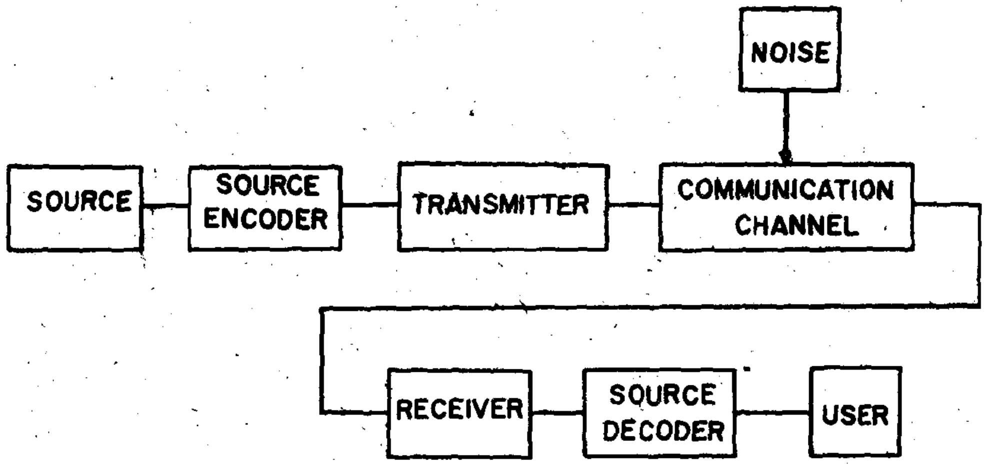

Fig. 1.1 Block diagram of a communication system.

For instance, if I show you a Chinese newspaper you might not be able to decode the language, even if the optical channel is assumed to be perfect (i.e., noiseless). This is because a suitable decoding process requires a priori knowledge of the encoding scheme (i.e., appropriate information storage), for example, a priori knowledge of the Chinese characters. Thus the decoding process can also be called a *recognition process*.

Information theory is a broad subject which can not be fully discussed in a few sections. Although we only investigate the theory in an introductory manner, our discussion in the next few sections provides a very useful application of information theory to optics. Readers who are interested in a rigorous treatment of information theory are referred to the papers by Shannon[1.1–1.3] and the text by Fano[1.4].

Information theory has two general orientations: one developed by Wiener[1.5, 1.6], and the other by Shannon[1.1–1.3]. Although both Wiener and Shannon shared a common probabilistic basis, there is a basic distinction between them.

The significance of Wiener's work is that, if a signal (information) is corrupted by some physical means (e.g., noise, nonlinear distortion), it may be possible to recover the signal from the corrupted one. It is for this purpose that Wiener develops the theories of correlation detection, optimum prediction, matched filtering, and so on. However, Shannon's work is carried a step further. He shows that the signal can be optimally transferred provided it is properly encoded. That is, the signal to be transferred can be processed before and after transmission. In the encoding process, he shows that it is possible to combat the disturbances in the communication channel to a certain extent. Then, by a proper

试读结束：需要全本请在线购买： www.ertongbook.com

decoding process, the signal can be recovered optimally. To do this, Shannon develops the theories of information measure, channel capacity, coding processes, and so on. The major interest in Shannon's theory is efficient utilization of the communication channel.

A fundamental theorem proposed by Shannon can be considered the most surprising result of this work. The theorem can be stated approximately as follows. Given a stationary finite-memory information channel having a channel capacity C, if the binary information transmission rate R of the message is smaller than C, there are channel encoding and decoding processes for which the probability of error in information transmission per digit can be made arbitrarily small. Conversely, if the information transmission rate R is larger than C, there are no encoding and decoding processes with this property, that is, the probability of error in information transmission cannot be made arbitrarily small. In other words, the presence of random disturbances in a communication channel does not, by itself, limit transmission accuracy. Rather, it limits the transmission rate for which arbitrarily high transmission accuracy can be accomplished.

In summarizing this brief introduction to information theory, we point out again the distinction between the viewpoints of Wiener and of Shannon. Wiener assumes in effect that the signal in question can be processed after it has been corrupted by noise. Shannon suggests that the signal can be processed both before and after its transmission through the communication channel. However, the main objectives of these two branches of information theory are basically the same, namely, faithful reproduction of the original signal.

1.1 DEFINITION OF INFORMATION MEASURE

We have in the preceding discussed a general concept of information transmission. In this section, we discuss this subject in more detail. Our first objective is to define a measure of information, which is vitally important in the development of modern information theory. We first consider discrete input and discrete output message ensembles as applied to a communication channel, as shown in Fig. 1.2. We denote the sets of input and output ensembles $A = \{a_i\}$ and $B = \{b_j\}$, respectively, $i = 1, 2, \ldots, M$, and $j = 1, 2, \ldots, N$. It is noted that AB forms a *discrete product space.*

Let us assume that a_i is an input event to the information channel, and b_j is the corresponding output event. Now we would like to define a measure of information in which the received event b_j specifies a_i. In

Fig. 1.2 An input-output communication channel.

other words, we would like to define a measure of the amount of information provided by the output event b_j *about* the corresponding input event a_i. We see that the transmission of a_i through the communication channel causes a change in the probability of a_i, from an a priori $P(a_i)$ to an a posteriori $P(a_i/b_j)$. In measuring this change, we take the logarithmic ratio of these probabilities. It turns out to be appropriate for the definition of information measure. Thus the amount of information provided by the output event b_j about the input event a_i can be defined as

$$I(a_i; b_j) \triangleq \log_2 \frac{P(a_i/b_j)}{P(a_i)} \qquad \text{bits.} \tag{1.1}$$

It is noted that the base of the logarithm can be a value other than 2. However, the base 2 is the most commonly used in information theory. Therefore we adopt this base value of 2 for use in this text. Other base values are also frequently used, for example, $\log_{10}$ and $\ln = \log_e$. The corresponding units of information measure of these different bases are *hartleys* and *nats*. The hartley is named for R. V. Hartley, who first suggested the use of a logarithmic measure of information [1.7], and nat is an abbreviation for *natural unit*. Bit used in Eq. (1.11), is a contraction of *binary unit*.

We see that Eq. (1.1) possesses a symmetric property with respect to input event a_i and output event b_j:

$$I(a_i; b_j) = I(b_j; a_i). \tag{1.2}$$

This symmetric property of information measure can be easily shown:

$$\log_2 \frac{P(a_i, b_j)}{P(b_j)P(a_i)} = \log_2 \frac{P(b_j/a_i)}{P(b_j)}.$$

According to Eq. (1.2), the amount of information provided by event b_j about event a_i is the same as that provided by a_i about b_j. Thus Eq. (1.1) is a measure defined by Shannon as *mutual information* or amount of information transferred between event a_i and event b_j.

It is clear that, if the input and output events are *statistically independent*, that is, if $P(a_i, b_j) = P(a_i)P(b_j)$, then $I(a_i; b_j) = 0$.

Furthermore, if $I(a_i; b_j) > 0$, then $P(a_i, b_j) > P(a_i)P(b_j)$, that is, there is a higher joint probability of a_i and b_j. However, if $I(a_i; b_j) < 0$, then $P(a_i, b_j) < P(a_i)P(b_j)$, that is, there is a lower joint probability of a_i and b_j.

As a result of the conditional probabilities $P(a_i/b_j) \leq 1$, and $P(b_j/a_i) \leq 1$, we see that

$$I(a_i; b_j) \leq I(a_i), \tag{1.3}$$

and

$$I(a_i; b_j) \leq I(b_j), \tag{1.4}$$

where

$$I(a_i) \triangleq -\log_2 P(a_i), \tag{1.5}$$

$$I(b_j) \triangleq -\log_2 P(b_j). \tag{1.6}$$

$I(a_i)$ and $I(b_j)$ are defined as the respective *input* and *output self-information* of event a_i and event b_j. In other words, $I(a_i)$ and $I(b_j)$ represent the amount of information provided at the input and output of the information channel of event a_i and event b_j, respectively. It follows that the mutual information of event a_i and event b_j is equal to the self-information of event a_i if and only if $P(a_i/b_j) = 1$: that is,

$$I(a_i; b_j) = I(a_i). \tag{1.7}$$

It is noted that, if Eq. (1.7) is true for all i, that is, the input ensemble, then the communication channel is *noiseless*. However, if $P(b_j/a_i) = 1$, then

$$I(a_i; b_j) = I(b_j). \tag{1.8}$$

If Eq. (1.8) is true for all the output ensemble, then the information channel is *deterministic*.

It is emphasized that the definition of measure of information can be extended to higher product spaces. For example, we can define the mutual information for a product ensemble ABC:

$$I(a_i; b_j/c_k) \triangleq \log_2 \frac{P(a_i/b_jc_k)}{P(a_i/c_k)}. \tag{1.9}$$

Similarly, one can show that

$$I(a_i; b_j/c_k) = I(b_j; a_i/c_k), \tag{1.10}$$

$$I(a_i; b_j/c_k) \leq I(a_i/c_k), \tag{1.11}$$

and

$$I(a_i; b_j/c_k) \leq I(b_j/c_k), \tag{1.12}$$

where

$$I(a_i/c_k) \triangleq -\log_2 P(a_i/c_k) \tag{1.13}$$

and

$$I(b_j/c_k) \triangleq -\log_2 P(b_j/c_k) \tag{1.14}$$

represent the *conditional self-information.*

Furthermore, from Eq. (1.1) we see that

$$I(a_i;b_j)=I(a_i)-I(a_i/b_j) \tag{1.15}$$

and

$$I(a_i;b_j)=I(b_j)-I(b_j/a_i). \tag{1.16}$$

From the definition of

$$I(a_ib_j)\triangleq -\log_2 P(a_i,b_j), \tag{1.17}$$

the self-information of the point (a_i, b_j) of the product ensemble AB, one can show that

$$I(a_i;b_j)=I(a_i)+I(b_j)-I(a_ib_j). \tag{1.18}$$

Conversely,

$$I(a_ib_j)=I(a_i)+I(b_j)-I(a_i;b_j). \tag{1.19}$$

In concluding this section, we point out that, for the mutual information $I(a_i;b_j)$ (i.e., the amount of information transferred through the channel) there exists an upper bound, $I(a_i)$ or $I(b_j)$, whichever comes first. If the information channel is noiseless, then the mutual information $I(a_i;b_j)$ is equal to $I(a_i)$, the input self-information of a_i. However, if the information channel is deterministic, then the mutual information is equal to $I(b_j)$, the output self-information of b_j. Moreover, if the input-output of the information channel is statistically independent, then no information can be transferred. It is also noted that, when the joint probability $P(a_i;b_j)<P(a_i)P(b_j)$, then $I(a_i;b_j)$ is negative, that is, the information provided by event b_j about event a_i further deteriorates, as compared with the statistically independent case. Finally, it is clear that the definition of the measure of information can also be applied to a higher product ensemble, namely, $ABC\cdots$ produce space.

1.2 ENTROPY AND AVERAGE MUTUAL INFORMATION

In the Sec. 1.1 we defined a measure of information. We saw that information theory is indeed probability theory.

In this section, we consider the measure of information as a random variable, that is, information measure as a random event. Thus the measure of information can be described by a probability distribution $P(I)$, where I is the self-, conditional, or mutual information.

Since the measure of information is usually characterized by an ensemble average, the average amount of information provided can be obtained by the ensemble average

$$E[I] = \sum_I IP(I), \tag{1.20}$$

where E denotes the expectation, and the summation is over all I.

If the self-information a_i in Eq. (1.5) is used in Eq. (1.20), then the average amount of self-information provided by the input ensemble A is

$$E[I(a_i)] = \sum_I IP(I) = \sum_{i=1}^{M} P(a_i)I(a_i), \tag{1.21}$$

where $I(a_i) = -\log_2(a_i)$.

For convenience in notation, we drop the subscript i; thus Eq. (1.21) can be written

$$I(A) \triangleq -\sum_A P(a)\log_2 P(a) \triangleq H(A), \tag{1.22}$$

where the summation is over the input ensemble A

Similarly, the average amount of self-information provided at the output end of the information channel can be written

$$I(B) \triangleq -\sum_B P(b)\log_2 P(b) \triangleq H(B). \tag{1.23}$$

As a matter of fact, Eqs. (1.22) and (1.23) are the starting points of Shannon's[1.1–1.3] information theory. These two equations are in essentially the same form as the *entropy equation* in statistical thermodynamics. Because of the identical form of the entropy expression, $H(A)$ and $H(B)$ are frequently used to describe *information entropy*. Moreover, we see in the next few chapters that Eqs. (1.22) and (1.23) are not just mathematically similar to the entropy equation, but that they represent a profound relationship between science and information theory[1.8–1.10], as well as between optics and information theory [1.11, 1.12].

It is noted that entropy H, from the communication theory point of view, is mainly a measure of *uncertainty*. However, from the statistical thermodynamic point of view, entropy H is a measure of *disorder*.

In addition, from Eqs. (1.22) and (1.23), we see that

$$H(A) \geq 0, \tag{1.24}$$

where $P(a)$ is always a positive quantity. The equality of Eq. (1.24) holds if $P(a) = 1$ or $P(a) = 0$. Thus we can conclude that

$$H(A) \leq \log_2 M, \tag{1.25}$$

where M is the number of different events in the set of input events A, that is, $A = \{a_i\}$, $i = 1, 2, \ldots, M$. We see that the equality of Eq. (1.25) holds if and only if $P(a) = 1/M$, that is, if there is *equiprobability* of all the input events.

In order to prove the inequality of Eq. (1.25), we use the well-known inequality

$$\log_2 u \le u - 1. \tag{1.26}$$

Let us now consider that

$$\begin{aligned} H(A) - \log_2 M &= -\sum_A p(a) \log_2 p(a) - \sum_A p(a) \log_2 M \\ &= \sum_A p(a) \log_2 \frac{1}{Mp(a)}. \end{aligned} \tag{1.27}$$

By the use of Eq. (1.26), one can show that

$$H(A) - \log_2 M \le \sum_A \left[\frac{1}{M} - p(a) \right] = 0. \tag{1.28}$$

Thus we have proved that the equality of Eq. (1.25) holds if and only if the input ensemble is equiprobable, $p(a) = 1/M$. We see that the entropy $H(A)$ is maximum when the probability distribution a is equiprobable. Under the maximization condition of $H(A)$, the amount of information provided is the *information capacity* of A.

To show the behavior of $H(A)$, we describe a simple example for the case of $M = 2$, that is, for a *binary source*. Then the entropy equation (1.22) can be written

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p), \tag{1.29}$$

where p is the probability of one of the events.

From Eq. (1.29) we see that $H(p)$ is maximum if and only if $p = \frac{1}{2}$. Moreover, the variation in entropy as a function of p is plotted in Fig. 1.3. It can be seen that $H(p)$ is a symmetric function, having a maximum value of 1 bit at $p = \frac{1}{2}$.

Similarly, one can extend this concept of ensemble average to the conditional self-information:

$$I(B/A) \triangleq -\sum_B \sum_A p(a, b) \log_2 p(b/a) \triangleq H(B/A). \tag{1.30}$$

We define $H(B/A)$ as the conditional entropy of B given A. Thus the entropy of the product ensemble AB can also be written

$$H(AB) = -\sum_A \sum_B p(a, b) \log_2 p(a, b), \tag{1.31}$$

where $p(a, b)$ is the joint probability of events a and b.

试读结束：需要全本请在线购买： www.ertongbook.com