

Quantum Detection Theory

Carl W. Helstrom

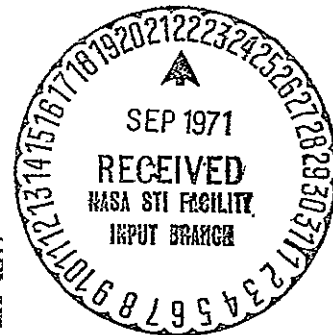
Department of Applied Physics and Information Science

University of California, San Diego

La Jolla, California 92037

NASA NGL 05-009-079

FACILITY FORM 602	<u>N 71-37101</u>	
	(ACCESSION NUMBER)	
	<u>132</u>	(THRU) <u>63</u>
	(PAGES)	(CODE)
	<u>CR-122930</u>	<u>23</u>
	(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)



Reproduced by
**NATIONAL TECHNICAL
INFORMATION SERVICE**
Springfield, Va. 22151

Table of Contents

	Page
1. Detection Theory	4
1. Binary Detection	7
2. Discrete Data and Randomization	11
3. Composite Hypotheses	14
4. Threshold Detection	16
5. Multiple Hypotheses	18
2. Detection Theory in Quantum Mechanics	21
1. Binary Detection	21
2. The Choice Between Pure States	28
3. Threshold Detection	29
4. Multiple Hypotheses	30
3. Detection of a Coherent Signal	32
1. The Transmission-Line Receiver: Classical Analysis	32
2. Quantization of the Receiver	39
3. The Coherent Signal of Known Phase	44
3.1 The Density Operators	44
3.2 The Extreme Quantum Limit	48
3.3 The Threshold Receiver	50
3.4 Approximations to the Optimum Receiver	52
4. Reception of a Signal of Random Phase	54
5. Amplification	57
4. Modal Decomposition of Aperture Fields	61

	Page
5. Detection of Incoherent Light	67
1. The Optical Fields	67
2. The Optimum Receiver	72
3. Detection of Point Sources	74
3.1 Extreme Quantum Limit	74
3.2 Intermediate Range	77
3.3 The Poisson Limit	78
3.4 The Effect of Amplification	79
3.5 The Classical Limit	80
3.6 The Threshold Detector	84
4. Detection of Extended Objects	87
4.1 Extreme Quantum Limit	90
4.2 Intermediate Range	91
6. Estimation Theory	94
1. Classical Parameter Estimation	94
2. Quantum Estimation	97
2.1 Minimum-Mean-Square-Error Estimation	97
2.2 Arbitrary Cost Functions	101
3. The Cramér-Rao Inequality	102
Acknowledgment	107
Bibliography	108
List of Figures	111

1. Detection Theory

To the question "What is the weakest light detectable?" the easy answer is "one photon". Yet it is not the presence of any light whatever that is ordinarily of interest, but the presence of light coming from a particular source or possessing specific characteristics. The astronomer wants to know whether there is a star at a certain point in the sky, the spectroscopist whether light of a definite wavelength is being emitted. In a laser radar the reception of a light pulse from a certain direction reveals the presence of a target. Our question must include the nature of the light to be detected, by for instance specifying the electromagnetic field whose presence or absence at the aperture of some observing instrument is to be determined.

Detection involves decision. Is the specified light field present or not? An observer must decide, but decide he cannot without liability to error. Thermal background light is usually incident on the aperture of his instrument, and because of its random nature, this background field cannot be predicted and subtracted. The light to be detected is itself a random phenomenon, both because of the chaos inherent in natural sources and because of the photonic character shared by all forms of light, coherent or not. When we say that a certain light field is detectable, therefore, we must also state with what probability of error. Our question is complex, and "one photon" is an inadequate response.

A receiver of light is to be viewed as making--or assisting an observer to make---decisions about the presence or absence of a specific light field at its aperture, and these decisions are subject to error. The most effective optical system permits them to be made with minimum probability of error, and

the performance of this optimum system determines the weakest detectable light. How to design the optimum optical receiver is a problem in detection theory, which is basically an application of the statistical testing of hypotheses, or "decision theory". We shall begin with an outline of the detection theory appropriate for classical physics and then show how it is modified to take the laws of quantum mechanics into account. After that a model of an ideal quantum receiver will be proposed and analyzed to determine the detectabilities of both coherent laser light and incoherent natural light. Our main concern will be to discover the fundamental limitations on optical detection imposed by the nature of the wanted light and the unwanted background illumination.

Detailed treatments of quantum detection theory and its applications can be found in papers by the writer [1967-1970]. A review by Helstrom et al. [1970] stresses its relation to optical communications. Articles bringing it into the study of optical communication through the turbulent atmosphere have been published by Kennedy and Hoversten [1968], Kennedy [1970], and Hoversten et al. [1970]. Detection theory as based on classical statistics has been applied to the detection of light by means of photosensitive surfaces in papers by Reiffen and Sherman [1963], Helstrom [1964, 1971], Goodman [1966], Bakut et al. [1966], Stefanyuk [1966], Bakut [1966, 1967], Ginzburg [1966], Bar-David [1969], Karp and Clark [1970], Karp et al. [1970], and Gagliardi [1972]. We shall not attempt to summarize this work here.

A telescope is used not only to discover a star, but to measure its location and radiant power, and a spectrometer measures the wavelengths and radiant fluxes of spectral lines. A laser radar determines the distance to a target by measuring the time elapsed until the arrival of a reflected coherent light pulse. Each of these instruments can be regarded as estimating certain

parameters of the light field at its aperture, and its results are liable to error because of the random nature of that field, subject to quantum fluctuations and corrupted by background light. As the primary field becomes weaker, error increases; and we shall show how the attendant limitations on the measuring power of an optical instrument can be evaluated. To this end we shall appeal to statistical estimation theory. We turn first to detection.

1. Binary Detection

The optical receiver is regarded as processing the light field appearing at its aperture A during an interval of time $(0, T)$, and the fundamental limitations on its ability to detect are brought out by asking what an arbitrary instrument processing this aperture field could achieve. We suppose first that during the observation interval the instrument abstracts from the field certain data, represented by a set of numbers $x_1, x_2, \dots, x_n$, or by a vector $\underline{x} = (x_1, x_2, \dots, x_n)$. These data might, for instance, be the values of one component of the aperture field at various points $\underline{r}_i \in A$ and times $t_i \in (0, T)$. Later the data set will be conceptually expanded until it encompasses all the information available in the field. For the present we shall ignore any quantum-mechanical limitations on the acquisition of the data $\underline{x}$, restricting ourselves thus to the domain of classical physics. Quantum detection theory will be introduced in Chapter 2. Details and proofs of what follows can be found in texts such as those by Middleton [1960], Helstrom [1968c], and Van Trees [1968].

The instrument is to decide, on the basis of the data $\underline{x}$, whether a certain field $\mathcal{F}$ exists as a constituent of the total field at the aperture, or whether the aperture contains only a field $\mathcal{F}_0$ due to the background. It chooses between two hypotheses, (H_0) the field $\mathcal{F}$ is absent, $\mathcal{F}_0$ only being at hand, and (H_1) the field $\mathcal{F}$ is present along with $\mathcal{F}_0$. The field $\mathcal{F}$ is commonly called the *signal*, $\mathcal{F}_0$ the *noise*.

Because of the stochastic nature of the aperture field, the data $x_1, x_2, \dots, x_n$ are random variables, and these two possibilities H_0 and H_1 correspond to distinct probability density functions (p.d.f.'s) of the data $\underline{x}$.

Under hypothesis H_0 the data are distributed according to the p.d.f. $p_0(x_1, x_2, \dots, x_n) = p_0(\underline{x})$, under H_1 according to the p.d.f. $p_1(x_1, x_2, \dots, x_n) = p_1(\underline{x})$. The former describes the statistical properties of the background field $\mathcal{F}_0$, the latter those of the combined fields $\mathcal{F}$ and $\mathcal{F}_0$. We suppose both p.d.f.'s completely known. The instrument or the observer using it chooses in effect one of these p.d.f.'s as the more compatible with the data.

How the choice is made is termed a *strategy* and is best visualized in terms of the n -dimensional Euclidean space of the data $\underline{x} = (x_1, x_2, \dots, x_n)$. The strategy corresponds to a division of this space into two regions R_0 and R_1 . When the data point $\underline{x}$ falls into region R_0 , hypothesis H_0 is selected; when $\underline{x}$ falls into R_1 , H_1 is chosen. Fig. 1.1 illustrates this division of a two-dimensional data space. We envision our strategy as being tried over and over again. The data being random variables, the point $\underline{x}$ falls here and there in the space with probability density $p_0(\underline{x})$ or $p_1(\underline{x})$ depending on which hypothesis happens to be correct.

When hypothesis H_0 is true and the data point $\underline{x}$ falls into region R_1 , hypothesis H_1 is incorrectly chosen, and an error of the first kind is said to occur. Its probability is

$$Q_0 = \int_{R_1} p_0(\underline{x}) d^n \underline{x}, \quad (1.1)$$

where $d^n \underline{x} = dx_1 dx_2 \dots dx_n$ is the rectilinear volume element in the space. In detection theory Q_0 is called the *false-alarm probability*. When on the other hand H_1 is true and $\underline{x}$ falls into R_0 , hypothesis H_0 is incorrectly chosen; and an error of the second kind, or "false dismissal", occurs. Its probability is

$$Q_1 = \int_{R_0} p_1(\underline{x}) d^n \underline{x}. \quad (1.2)$$

As H_1 is associated with the presence of the sought field $\mathcal{F}$ and H_0 with its absence, the complementary probability

$$Q_d = 1 - Q_1 = \int_{R_1} p_1(\underline{x}) d^n \underline{x} \quad (1.3)$$

of choosing H_1 when H_1 is true is called the *probability of detection*.

Statisticians call Q_0 the *size*, Q_d the *power* of the statistical test.

At this point a metaphysical element enters the theory. We must specify how serious these errors are, and this is most conveniently done by introducing the costs C_0 of an error of the first kind and C_1 of an error of the second kind. If in addition we know the prior probability ζ of hypothesis H_0 and $(1 - \zeta)$ of H_1 —that is, if we know how often in a long series of trials each hypothesis occurs—, we can work out the average cost $\bar{C}$ associated with any strategy,

$$\begin{aligned} \bar{C} &= (1 - \zeta)C_1Q_1 + \zeta C_0Q_0 \\ &= (1 - \zeta)C_1 \left[1 - \int_{R_1} p_1(\underline{x}) d^n \underline{x} \right] + \zeta C_0 \int_{R_1} p_0(\underline{x}) d^n \underline{x}, \end{aligned} \quad (1.4)$$

which we can write as

$$\bar{C} = (1 - \zeta) C_1 \left\{ 1 - \int_{R_1} [p_1(\underline{x}) - \Lambda_0 p_0(\underline{x})] d^n \underline{x} \right\} \quad (1.5)$$

by introducing the constant

$$\Lambda_0 = \frac{\zeta C_0}{(1 - \zeta)C_1}. \quad (1.6)$$

The *Bayes criterion* for the optimality of a statistical test requires the average cost $\bar{C}$ to be minimum. The strategy that achieves this is called the *Bayes strategy*, and to find it we move the surface dividing R_1 from R_0 about the data space until the average cost as given by eq. (1.5) is as small as possible.

Minimizing $\bar{C}$ involves maximizing the integral on the right-hand side of eq. (1.5), and this is done by putting into region R_1 all points $\underline{x}$ for which $p_1(\underline{x}) > \Lambda_0 p_0(\underline{x})$ and into region R_0 all the rest. Equivalently, the Bayes strategy chooses hypothesis H_0 whenever $\Lambda(\underline{x}) < \Lambda_0$ and H_1 whenever $\Lambda(\underline{x}) > \Lambda_0$, where

$$\Lambda(\underline{x}) = p_1(\underline{x})/p_0(\underline{x}) \quad (1.7)$$

is called the *likelihood ratio*. The number Λ_0 with which it is compared is called the *decision level*.

As the costs C_0 and C_1 and the prior probability ζ may not always be accurately known, an alternative viewpoint dispenses with them and simply fixes the false-alarm probability Q_0 at a level that the observer can afford. The strategy maximizing the probability Q_d of detection is then sought. This goal is called the Neyman-Pearson criterion. The same strategy results: pick hypothesis H_0 whenever $\Lambda(\underline{x}) < \Lambda_0$ and hypothesis H_1 otherwise. The decision level Λ_0 is set so that the false-alarm probability

$$Q_0 = \Pr [\Lambda(\underline{x}) > \Lambda_0 | H_0] \quad (1.8)$$

equals the pre-assigned value.

The optimum instrument for detecting the field $\mathcal{F}$, then, is one that somehow evaluates the likelihood ratio $\Lambda(\underline{x})$ for the field at its aperture A during the observation interval $(0, T)$, the data set $\underline{x}$ being augmented by ever finer sampling until all the information contained in the field is utilized. Whether the Bayes or the Neyman-Pearson criterion is preferred does not affect the design of the optimum instrument, but only the value Λ_0 with which its output $\Lambda(\underline{x})$ is compared.

2. Discrete Data and Randomization

Practicality may restrict the kind of processing to which the aperture field can be subjected. It may be necessary, for instance, to focus the field onto a photoelectric surface and count the electrons emitted during the observation interval from various regions thereof. A decision strategy utilizing the counts as the data $\underline{x} = (x_1, x_2, \dots, x_n)$ and meeting one of the criteria of optimality is needed. The data are now discrete random variables.

The same type of analysis as what we have just presented can be carried out; it is merely necessary to replace integrals over probability densities as in eqs. (1.1) - (1.5) with sums over probabilities. The best system is now one that bases its decisions on the likelihood ratio

$$\Lambda(\underline{x}) = P_1(\underline{x})/P_0(\underline{x}), \quad (1.9)$$

which is the quotient of the probabilities $P_0(\underline{x})$ and $P_1(\underline{x})$ of the data under the two hypotheses. The likelihood ratio $\Lambda(\underline{x})$ will be a discrete random variable, taking values only on a countable set of numbers. Under the Bayes criterion, hypothesis H_1 is chosen if $\Lambda(\underline{x})$ exceeds the decision level Λ_0 given by eq. (1.6).

When the Neyman-Pearson criterion is applied to discrete data $\underline{x}$, the pre-assigned value of the false-alarm probability Q_0 may not be attainable by setting Λ_0 equal to any of the discrete set of values that the likelihood ratio $\Lambda(\underline{x})$ can take on. It is then necessary to *randomize* the decision.

Imagine the possible values λ_i that the likelihood ratio $\Lambda(\underline{x})$ can assume as arranged in ascending order,

$$0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_{i-1} < \lambda_i < \lambda_{i+1} < \dots < \infty.$$

Each of these has a certain probability

$$P_i = \Pr \{ \Lambda(\underline{x}) = \lambda_i | H_0 \} \quad (1.10)$$

of occurring under hypothesis H_0 . Suppose that setting the decision level at one of the λ_i , say Λ' , yields a false-alarm probability just too large and that taking the next in order, Λ'' , yields one just too small,

$$\sum_{\lambda_i \geq \Lambda'} P_i > Q_0 > \sum_{\lambda_i \geq \Lambda''} P_i, \quad \Lambda'' > \Lambda'. \quad (1.11)$$

The randomized strategy then chooses hypothesis H_0 whenever $\Lambda(\underline{x}) < \Lambda'$ and H_1 whenever $\Lambda(\underline{x}) \geq \Lambda'' > \Lambda'$; but when $\Lambda(\underline{x}) = \Lambda'$, hypothesis H_1 is chosen with a certain probability f , and H_0 with probability $1 - f$, by for instance tossing a properly biased coin. The value of f is picked so that the false-alarm probability

$$Q_0 = \sum_{\lambda_i > \Lambda'} P_i + f \Pr \{ \Lambda(\underline{x}) = \Lambda' | H_0 \} \quad (1.12)$$

takes on the pre-assigned value. The probability of detection is then

$$Q_d = \sum_{\lambda_i > \Lambda'} \Pr \{ \Lambda(\underline{x}) = \lambda_i | H_1 \} + f \Pr \{ \Lambda(\underline{x}) = \Lambda' | H_1 \}. \quad (1.13)$$

We shall need this type of randomized strategy when the receiver is one that counts photons.

As an example, suppose that the decision is based on a single number n of photons whose distribution under each hypothesis has the Bose form, but with different means, $\bar{n}_0$ and $\bar{n}_1$, $\bar{n}_1 > \bar{n}_0$,

$$p_i(n) = (1 - v_i) v_i^n, \quad v_i = \bar{n}_i / (\bar{n}_i + 1), \quad i = 0, 1. \quad (1.14)$$

Basing the decision on the likelihood ratio $\Lambda(n)$ is now equivalent to basing it on the number n of counts. A randomized decision rule would set a certain

integral decision level θ and would choose hypothesis H_1 when $n > \theta$ and H_0 when $n < \theta$, choosing hypothesis H_1 with probability f whenever the number n equals θ exactly. Then eqs. (1.12) and (1.13) become

$$\begin{aligned} Q_0 &= v_0^{\theta+1} + f(1 - v_0)v_0^{\theta}, \\ Q_d &= v_1^{\theta+1} + f(1 - v_1)v_1^{\theta}. \end{aligned} \tag{1.15}$$

From the first of these the values of θ and f , $0 < f \leq 1$, can easily be determined. Further examples, involving photon or photoelectron counting, are found in Helstrom [1969a, 1971].

3. Composite Hypotheses

Up to now we have supposed that the observer or the designer of the optical receiver has complete knowledge of the p.d.f.'s $p_0(\underline{x})$ and $p_1(\underline{x})$ governing the data $\underline{x}$ under the two hypotheses, which are therefore termed *simple*. Although $p_0(\underline{x})$ may be entirely known--it describes the background field in our applications--, it often happens that the p.d.f. $p_1(\underline{x})$ of the data under hypothesis H_1 depends on certain parameters whose values are unknown *a priori*. These are usually parameters of the field $\mathcal{F}$ to be detected, such as an overall amplitude or flux level, the phase of a coherent field, or the arrival time and wavelength of a laser pulse. Hypothesis H_1 is then termed *composite*, and the p.d.f. of the data under H_1 is written as $p_1(\underline{x}; \underline{\theta})$, where $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ stands for the collection of m unknown parameters.

If a prior p.d.f. $z(\underline{\theta})$ of these parameters, representing the relative frequencies with which their values fall into various ranges, is known, and if the error cost C_1 is independent of the parameters $\underline{\theta}$, the Bayes criterion requires the choice between the simple hypothesis H_0 and the composite hypothesis H_1 to be based on the average likelihood ratio

$$\bar{\Lambda}(\underline{x}) = \int z(\underline{\theta}) \Lambda(\underline{x}; \underline{\theta}) d^m \underline{\theta},$$

$$\Lambda(\underline{x}; \underline{\theta}) = p_1(\underline{x}; \underline{\theta}) / p_0(\underline{x}), \quad (1.16)$$

which is compared as before with the decision level Λ_0 given by eq. (1.6). The Neyman-Pearson criterion now specifies that the average detection probability

$$\bar{Q}_d[z] = \iint z(\underline{\theta}) p_1(\underline{x}; \underline{\theta}) d^m \underline{\theta} d^n \underline{x} \quad (1.17)$$

be maximum, the false-alarm probability Q_0 being fixed; and it leads to

comparison of the same average likelihood ratio $\bar{\Lambda}(\underline{x})$ with a decision level Λ_0 set to yield a pre-assigned false-alarm probability.

Occasionally the Neyman-Pearson test of hypothesis H_0 against a hypothesis H_1 with a known set of values $\underline{\theta}$ leads to a dichotomy of the data space into two regions R_0 and R_1 that is independent of the parameters $\underline{\theta}$. This happens when the likelihood ratio $\Lambda(\underline{x}; \underline{\theta})$ depends on the data $\underline{x}$ only through some function $f(\underline{x})$ that does not involve the parameters $\underline{\theta}$. The decision can then be based on the value of $f(\underline{x})$ for the data at hand. Such a function $f(\underline{x})$ is called a *sufficient statistic*. The test is said to be uniformly most powerful and applies whatever the prior p.d.f. $z(\underline{\theta})$ may be.

If the prior p.d.f. $z(\underline{\theta})$ is unknown, and if no uniformly most powerful test exists, a conservative observer may wish to use the Neyman-Pearson test that is based on the least favorable prior p.d.f. $\tilde{z}(\underline{\theta})$. The least favorable prior p.d.f. $\tilde{z}(\underline{\theta})$ is the one for which the maximum average detection probability $\bar{Q}_d[z]$ is minimum, the false-alarm probability being held fixed. The principal application of the concept of a least favorable distribution is to the detection of a coherent signal or field of unknown phase ϕ . The least favorable distribution of ϕ is the uniform one,

$$\tilde{z}(\phi) = (2\pi)^{-1}, \quad 0 \leq \phi < 2\pi, \quad (1.18)$$

and the detection probability attained by the resulting system is independent of the actual phase ϕ of the coherent field that happens to be present. This independence is characteristic of least favorable p.d.f.'s [Helstrom, 1968c, p. 159]. If, on the other hand, the observer knew what phase ϕ the field carries when it is present, he could construct a receiver with a higher probability $Q_d(\phi)$ of detection for the same false-alarm probability. With light fields this knowledge of the phase is seldom to be had.

4. Threshold Detection

When the overall strength S of the field $\mathcal{F}$ to be detected is unknown, the optimum detector can in many cases not be constructed, for the likelihood ratio determining its structure is such a function of S that no uniformly most powerful test exists. There are two courses of action. One is to design the receiver for a standard signal strength S_0 , and for other strengths $S \neq S_0$ to accept a detection probability $Q_d(S)$ less than the maximum possible.

An alternative approach is to postulate the unfavorable situation of very weak signals and design the receiver to be optimum in the limit of vanishing signal strength $S \rightarrow 0$. The likelihood ratio $\Lambda(\underline{x}; S)$ as a function of the strength S of the signal is expanded in a power series about $S = 0$,

$$\Lambda(\underline{x}; S) = 1 + S \left. \frac{\partial \Lambda(\underline{x}; S)}{\partial S} \right|_{S=0} + \frac{1}{2} S^2 \left. \frac{\partial^2 \Lambda(\underline{x}; S)}{\partial S^2} \right|_{S=0} + \dots \quad (1.19)$$

The receiver bases its decisions on the coefficient of the smallest power of S appearing, usually $(\partial \Lambda / \partial S)_{S=0}$ or $(\partial^2 \Lambda / \partial S^2)_{S=0}$, comparing it with a decision level set to yield a pre-assigned false-alarm probability. This scheme is known as *threshold detection* [Middleton, 1960, §19.4; 1966].

When a number M of statistically independent sets $\underline{x}^{(j)}$ of data are available, $j = 1, 2, \dots, M$, all of which either contain the field $\mathcal{F}$ or do not, the decision may be based on the sum Z of a certain function $f(\underline{x}^{(j)})$ of each set of data,

$$Z = \sum_{j=1}^M f(\underline{x}^{(j)}), \quad (1.20)$$

hypothesis H_1 being chosen when Z exceeds a certain decision level Z_0 . The logarithm of the likelihood ratio of the data, $\ln \Lambda(\underline{x}^{(1)}, \underline{x}^{(2)}, \dots, \underline{x}^{(M)})$, will be such a statistic. If $M \gg 1$, Z will have approximately a Gaussian

distribution under both hypotheses H_0 and H_1 , by virtue of the central limit theorem [Cramér, 1946, p. 213], and the false-alarm and detection probabilities will be

$$Q_0 \cong \text{erfc } x = (2\pi)^{-1/2} \int_x^\infty \exp(-t^2/2) dt, \quad (1.21)$$

$$Q_d \cong \text{erfc } (x - \Delta\sqrt{M}). \quad (1.22)$$

Here x is related to the decision level Z_0 , and $\Delta(S)$ is an effective signal-to-noise ratio given by

$$[\Delta(S)]^2 = \frac{[\underline{E}(f|H_1) - \underline{E}(f|H_0)]^2}{\text{Var}_0 f}, \quad (1.23)$$

where $\text{Var}_0 f$ is the variance of the statistic $f(\underline{x})$ when no signal is present, and $\underline{E}$ stands for expected value.

If we now fix the probabilities Q_0 and Q_d and let M increase, the signal strength S required will decrease. The best statistic $f(\underline{x})$ to use will, for $M \gg 1$, be that for which the effective signal-to-noise ratio $\Delta(S)$ is largest in the limit $S \rightarrow 0$. This optimum statistic can easily be shown to be the threshold statistic just described through eq. (1.19) [Rudnick, 1962]. Thus if we compare two detectors forming different functions $f(\underline{x})$ of the data, for the same values of Q_0 , Q_d , and $M \gg 1$, the threshold detector will require the smaller signal strength S .

5. Multiple Hypotheses

A communication system might transmit any one of M laser pulses of different forms or different wavelengths, each corresponding to a symbol of an alphabet of M symbols into which messages have been coded. A pulse is sent forth once every T seconds. During each interval of T seconds' duration when a transmitted signal arrives, the receiver must decide which of the M possible pulses has been sent. Its decisions are based on the data $\underline{x}$ obtained, for instance, by sampling the field at the aperture of the receiver during the interval. The receiving optical system must now permit a choice not between two, but among M hypotheses, which we label H_j , $j = 1, 2, \dots, M$. Under hypothesis H_j , "The j -th signal is present", the data $\underline{x}$ are described by a probability density function $p_j(\underline{x})$. We say that the receiver must carry out a multiple hypothesis test.

The Bayes criterion of minimum average cost can again be applied if we know the prior probabilities ζ_j of each of the M hypotheses, with

$$\sum_{j=1}^M \zeta_j = 1,$$

and the costs C_{ij} of choosing hypothesis H_i when H_j is true. The average cost is then

$$\bar{C} = \sum_{i=1}^M \sum_{j=1}^M \zeta_j C_{ij} \int_{R_i} p_j(\underline{x}) d^n \underline{x}, \quad (1.24)$$

where R_i is the region of the data space for which hypothesis H_i is selected.

The average cost $\bar{C}$ is minimum for a strategy that calculates the M posterior risks

$$r_i = \sum_{j=1}^M C_{ij} p(H_j | \underline{x}) \quad (1.25)$$

and selects the hypothesis with the smallest posterior risk. Here

$$p(H_j | \underline{x}) = \zeta_j p_j(\underline{x}) / p(\underline{x}) \quad (1.26)$$

is the posterior probability of hypothesis H_j upon observation of the data $\underline{x}$, with

$$p(\underline{x}) = \sum_{i=1}^M \zeta_i p_i(\underline{x}) \quad (1.27)$$

the total probability density function of the data $\underline{x}$.

Let us introduce the p.d.f. $p_0(\underline{x})$ of the data when no signal field, but only the background field $\mathcal{F}_0$ is present. It corresponds to a pseudo-hypothesis H_0 that $\mathcal{F}_0$ alone is at hand. Then by dividing both numerator and denominator of eq. (1.26) by $p_0(\underline{x})$, we can write the posterior probabilities upon which the decision depends as

$$\begin{aligned} p(H_j | \underline{x}) &= \zeta_j \Lambda_j(\underline{x}) / \Lambda_a(\underline{x}), \\ \Lambda_j(\underline{x}) &= p_j(\underline{x}) / p_0(\underline{x}), \quad j = 1, 2, \dots, M, \\ \Lambda_a(\underline{x}) &= \sum_{k=1}^M \zeta_k \Lambda_k(\underline{x}), \end{aligned} \quad (1.28)$$

where $\Lambda_j(\underline{x})$ is the likelihood ratio for deciding between hypothesis H_j and the pseudo-hypothesis H_0 . In this form the passage to the limit of exhaustive sampling of the aperture field is most easily carried out.

When the costs of error are equal, and correct decisions cost nothing,

$$C_{ij} = C, \quad i \neq j; \quad C_{ii} = 0, \quad \text{all } i, \quad (1.29)$$

the Bayes strategy requires selection of that hypothesis for which the posterior probability $p(H_i | \underline{x})$ is maximum. This is the familiar Bayes rule for choosing among a number of statistical hypotheses. Equivalently, one chooses the hypothesis H_i for which $\zeta_i \Lambda_i(\underline{x})$ is largest.

Estimating an unknown parameter θ of a field or of its p.d.f. $p(\underline{x}; \theta)$ can be considered a continuous version of the testing of multiple hypotheses. Indeed, if we content ourselves with deciding in which of a number of finite ranges of values the parameter θ lies, estimation becomes equivalent to choosing among a number of hypotheses H_j . We shall see in ch. 6 that the Bayes formulation of multiple hypothesis testing can be carried over directly to estimation.

2. Detection Theory in Quantum Mechanics

1. Binary Detection

In our outline of detection theory in the first chapter, we showed that the best optical instrument for binary detection is one that generates a likelihood ratio $\Lambda(\underline{x})$ for samples $\underline{x}$ of its aperture field during the observation interval; and we presumed that this would eventually be accomplished with a set of samples $\underline{x}$ that exhausts all the information in that field. The field would have to be sampled at points $\underline{r}_i$ in the aperture and at times t_i in the observation interval that are but infinitesimally separated. This infinitely dense sampling needs to be possible only conceptually; the same result can be achieved by appropriate filtering of the aperture field. Thus in classical physics this passage to the limit of an infinite number n of data as required by detection theory poses no serious difficulties.

Light fields, however, are subject to the laws of quantum mechanics, which place limitations on the extent and precision with which the fields can be measured. The amplitude and phase of a coherent field, for instance, are not simultaneously measurable with perfect accuracy; and measurements of a field at points on each other's light cones interfere. Although fields containing many photons can be treated along the lines of classical physics, when as in many optical detection problems few photons may be available, the quantum-mechanical behavior of the fields must be taken into account. Some care must be exercised, therefore, in prescribing how the samples $\underline{x}$ are to be taken in order to use all the information in the aperture field. Indeed, we require a reformulation of detection theory in quantum-mechanical terms.

In quantum mechanics it is difficult to treat measurements of a system at a succession of times, and in order to avoid this difficulty in studying the detection of light, we imagine a Gedankenexperiment. Behind the aperture of our optical instrument we place a large, lossless cavity, initially closed and empty. During the observation interval $(0, T)$ the aperture is open, and the field inside the cavity interacts with the external field, which contains background light and perhaps also the field $\mathcal{F}$ to be detected. At the end of the observation interval the aperture is closed. The decision about the presence or absence of the sought field--that is, the choice between hypotheses H_0 and H_1 --will now be based on measurements of the field inside the cavity at some later time $t > T$.

The cavity field can be described quantum-mechanically by giving its density operators ρ_0 and ρ_1 under the two hypotheses H_0 and H_1 . These operators correspond to the classical p.d.f.'s $p_0(\underline{x})$ and $p_1(\underline{x})$ that we dealt with in ch. 1. Any measurable quantity attached to the field corresponds to an Hermitian operator, say X , and the outcomes of a measurement are the eigenvalues of X . If X has discrete eigenvalues x_k and eigenstates $|x_k\rangle$,

$$X|x_k\rangle = x_k|x_k\rangle,$$

the probability under hypothesis H_i that a measurement yields the value x_k is $\langle x_k|\rho_i|x_k\rangle$, $i = 0, 1$. If the eigenvalues x form a continuous spectrum,

$$X|x\rangle = x|x\rangle,$$

the p.d.f.'s of the outcome of a measurement of X are $\langle x|\rho_i|x\rangle$, $i = 0, 1$, under the two hypotheses. The observer must choose between the hypotheses by making the best possible measurements on the field in the cavity, the best measurements again being defined as those enabling us to choose between H_0 and H_1 with minimum average cost.

If n operators $X_1, X_2, \dots, X_n$ are to be measured simultaneously, they must commute among themselves: $X_i X_j = X_j X_i$, all i and j . Let the outcomes of measuring them be $x_1, x_2, \dots, x_n$, respectively. A decision strategy can be described as a function $f(x_1, x_2, \dots, x_n)$ of the outcomes that takes on only two possible values, 0 and 1. If in a given test of the strategy, $f(x_1, x_2, \dots, x_n) = 0$, hypothesis H_0 is chosen, otherwise H_1 . The optimum classical strategy derived in ch. 1, for instance, can be expressed as the function

$$\begin{aligned} f(\underline{x}) &= U(p_1(\underline{x}) - \Lambda_0 p_0(\underline{x})), \\ \underline{x} &= (x_1, x_2, \dots, x_n), \end{aligned} \quad (2.1)$$

where

$$U(y) = 0, \quad y < 0; \quad U(y) = 1, \quad y > 0, \quad (2.2)$$

is the unit step function. Classically the set $\underline{x}$ can be augmented until all the relevant information in the field is encompassed. Quantum-mechanically this is impossible, and the problem of which operators X_k to measure must be faced.

Since the n operators X_k must commute, we can form the operator $f(X_1, X_2, \dots, X_n)$ and measure it instead. Because only the values 0 and 1 can result from the measurement, this operator must be a projection operator, and we denote it by Π . Among all possible projection operators for the cavity field, we must determine which one yields minimum average cost.

The probability Q_0 of an error of the first kind is the probability that measuring Π yields the value 1 when H_0 is true,

$$Q_0 = \text{Tr} (\rho_0 \Pi), \quad (2.3)$$

where Tr stands for the trace of an operator. The probability Q_1 of an error of the second kind is

$$Q_1 = \text{Tr} [\rho_1 (1 - \Pi)] = 1 - \text{Tr} (\rho_1 \Pi), \quad (2.4)$$

where $\underline{1}$ is the identity operator. The average cost is now, as in eq. (1.5),

$$\bar{C} = \zeta C_0 Q_0 + (1 - \zeta) C_1 Q_1 = (1 - \zeta) C_1 \{1 - \text{Tr}[(\rho_1 - \Lambda_0 \rho_0) \Pi]\}, \quad (2.5)$$

where Λ_0 is again given by eq. (1.6). We must choose the projection operator Π minimizing $\bar{C}$, or equivalently, maximizing $\text{Tr}[(\rho_1 - \Lambda_0 \rho_0) \Pi]$.

Let the eigenvalues and eigenstates of the operator $\rho_1 - \Lambda_0 \rho_0$ be η_k and $|\eta_k\rangle$ respectively, as in

$$(\rho_1 - \Lambda_0 \rho_0) |\eta_k\rangle = \eta_k |\eta_k\rangle. \quad (2.6)$$

Then we must maximize

$$\text{Tr}[(\rho_1 - \Lambda_0 \rho_0) \Pi] = \sum_k \eta_k \langle \eta_k | \Pi | \eta_k \rangle, \quad (2.7)$$

and this we can do by picking Π as an operator projecting the state vector of the system onto the subspace spanned by the eigenstates $|\eta_k\rangle$ associated with positive eigenvalues, $\eta_k > 0$,

$$\Pi = \sum_{k: \eta_k > 0} |\eta_k\rangle \langle \eta_k| = \sum_k U(\eta_k) |\eta_k\rangle \langle \eta_k|. \quad (2.8)$$

The minimum average cost is now

$$\bar{C}_{\min} = (1 - \zeta) C_1 \left[1 - \sum_k \eta_k U(\eta_k) \right]. \quad (2.9)$$

The optimum projection operator can be freely written in terms of the unit step function as

$$\Pi = U(\rho_1 - \Lambda_0 \rho_0),$$

which is the quantum-mechanical counterpart to eq. (2.1).

When the density operators ρ_0 and ρ_1 commute, they possess a common set of eigenstates $|\eta_k\rangle$,

$$\rho_i |\eta_k\rangle = P_i(k) |\eta_k\rangle, \quad i = 0, 1, \quad (2.10)$$

where $P_i(k)$ is the probability under hypothesis H_i that the system is in state

$|\eta_k\rangle$. The eigenvalues of the operator $\rho_1 - \Lambda_0 \rho_0$ are

$$\eta_k = P_1(k) - \Lambda_0 P_0(k), \quad (2.11)$$

and the optimum strategy is therefore to pick hypothesis H_1 when

$$\Lambda(k) = P_1(k)/P_0(k) > \Lambda_0; \quad (2.12)$$

this is the same as in classical detection theory. Any operator with the same eigenstates $|\eta_k\rangle$ as ρ_0 and ρ_1 and having distinct eigenvalues can just as well be measured, and the decision can be based on a likelihood ratio formed from the outcome of the measurement.

The Neyman-Pearson criterion is not so easily handled in quantum mechanics. One defines a randomized decision operator Π_r , the outcome of whose measurement is a number f_r lying between 0 and 1. This outcome is taken as the probability with which one should choose hypothesis H_1 ; that is, a chance device would be constructed that yields a zero with probability $1 - f_r$ and a one with probability f_r , and which of these, 0 or 1, turned up would determine the decision. Measurement of the operator Π_r may yield a different value of f_r on each trial of the system.

The false-alarm and detection probabilities are now

$$Q_0 = \text{Tr}(\rho_0 \Pi_r), \quad Q_d = \text{Tr}(\rho_1 \Pi_r), \quad (2.13)$$

and in order to maximize Q_d for fixed Q_0 we introduce a Lagrange multiplier Λ and maximize

$$\text{Tr}[(\rho_1 - \Lambda \rho_0) \Pi_r].$$

Again Π_r takes the form in eq. (2.8), where the η_k and $|\eta_k\rangle$ are the eigenvalues and eigenstates of the operator $(\rho_1 - \Lambda \rho_0)$,

$$(\rho_1 - \Lambda \rho_0) |\eta_k\rangle = \eta_k |\eta_k\rangle. \quad (2.14)$$

It is now necessary to vary Λ until the false-alarm probability Q_0 takes on exactly the pre-assigned value. As the operator equation (2.14) is generally difficult to solve when ρ_0 and ρ_1 do not commute, finding the optimum operator for the Neyman-Pearson strategy will not be easy. If ρ_0 and ρ_1 do commute, the problem reduces to the classical one, and the randomized strategy described in ch. 1 §2 is optimum.

In many applications of the Neyman-Pearson criterion, as in radar detection, the false-alarm probability Q_0 is very small, $Q_0 \ll 1$, and Λ is thereupon very large, $\Lambda \gg 1$. (If $\Lambda = \infty$, hypothesis H_1 is never chosen, and $Q_0 = 0$.) We are then tempted to define $\Lambda' = 1/\Lambda$ and rewrite the eigenvalue equation (2.14) as

$$(\rho_0 - \Lambda' \rho_1) |\eta_k\rangle = \eta_k' |\eta_k\rangle, \quad \eta_k' = -\eta_k/\Lambda. \quad (2.15)$$

When $\Lambda' \ll 1$ we can treat the term $\Lambda' \rho_1$ as a perturbation and approximate the eigenvalues by the first two terms of the standard perturbation expansion,

$$\eta_k' = P_0(k) - \Lambda' \langle k | \rho_1 | k \rangle, \quad (2.16)$$

where $P_0(k)$ are the eigenvalues and $|k\rangle$ the eigenstates of the density operator ρ_0 ,

$$\rho_0 |k\rangle = P_0(k) |k\rangle. \quad (2.17)$$

Now hypothesis H_1 is chosen when the outcome η_k' of a measurement of $\rho_0 - \Lambda' \rho_1$ is negative, that is, when

$$\langle k | \rho_1 | k \rangle > \Lambda P_0(k) = \Lambda \langle k | \rho_0 | k \rangle.$$

The decision can therefore, in this approximation, be based on a measurement of ρ_0 or of all the projection operators $|k\rangle\langle k|$, after which the classical likelihood ratio

$$\langle k | \rho_1 | k \rangle / P_0(k)$$

is formed for the eigenstate $|k\rangle$ of ρ_0 that turns up. The decision level Λ is set so that the false-alarm probability takes on a pre-assigned value. We shall later apply this approximation to the detection of a coherent signal.

2. The Choice Between Pure States

A simple case in which the eigenvalue equation (2.14) can be solved exactly involves the choice between two pure states, which we denote by $|\psi_0\rangle$ and $|\psi_1\rangle$. The density operators are then

$$\rho_0 = |\psi_0\rangle\langle\psi_0|, \quad \rho_1 = |\psi_1\rangle\langle\psi_1|$$

[Bakut and Shchurov, 1968; Helstrom, 1968d]. There are only two eigenstates with non-zero eigenvalues; we denote them by $|\eta_0\rangle$ and $|\eta_1\rangle$. They are linear combinations of $|\psi_0\rangle$ and $|\psi_1\rangle$,

$$|\eta_i\rangle = z_{i0}|\psi_0\rangle + z_{i1}|\psi_1\rangle, \quad i = 0, 1.$$

Substitution into eq. (2.14) yields simultaneous equations for the z_{ij} 's, and one finds non-zero solutions for them only when $\eta = \eta_0$ or $\eta = \eta_1$ in the determinantal equation

$$\begin{vmatrix} 1 - \eta & \gamma \\ -\Lambda\gamma^* & -\Lambda - \eta \end{vmatrix} = 0$$

where $\gamma = \langle\psi_1|\psi_0\rangle$. Thus one obtains the eigenvalues

$$\begin{aligned} \eta_1 &= \frac{1}{2} (1 - \Lambda) + R > 0, & \eta_0 &= \frac{1}{2} (1 - \Lambda) - R < 0, \\ R &= \left\{ \left[\frac{1}{2} (1 - \Lambda) \right]^2 + \Lambda q \right\}^{\frac{1}{2}}, & q &= 1 - |\gamma|^2. \end{aligned}$$

The false-alarm and detection probabilities are

$$\begin{aligned} Q_0 &= |\langle\eta_1|\psi_0\rangle|^2 = (\eta_1 - q)/2R, \\ Q_d &= |\langle\eta_1|\psi_1\rangle|^2 = (\eta_1 + \Lambda q)/2R. \end{aligned}$$

When the first equation is solved for Λ , and Λ is substituted into the second, we obtain

$$\begin{aligned} Q_d &= \{ [Q_0(1 - q)]^{\frac{1}{2}} + [(1 - Q_0)q]^{\frac{1}{2}} \}^2, & 0 \leq q \leq 1 - Q_0, \\ &= 1, & 1 - Q_0 \leq q \leq 1. \end{aligned} \quad (2.18)$$

If in particular the states are orthogonal, $\gamma = 0$, $q = 1$, and we find $Q_0 = 0$,

$Q_d = 1$; orthogonal states can be distinguished without error.

3. Threshold Detection

The optimum detection operator Π is often difficult to determine, and it will in general depend on the strength of the signal. If that strength is unknown in advance, the designer might consider the threshold approach described in ch. 1 §4, where it is assumed that the signal is very weak. A detector is then termed optimum when it requires the least signal strength to attain a given pair (Q_0, Q_d) of false-alarm and detection probabilities, a great number M of inputs being at hand, all either containing or not containing the signal. Equivalently, the best detector is the one that maximizes the effective signal-to-noise ratio, eq. (1.23), in the limit $S \rightarrow 0$.

Let us suppose that the detector measures an operator X on the field. The effective signal-to-noise ratio is the quantum-mechanical counterpart of eq. (1.23),

$$\Delta^2 = [\text{Tr}(\rho_1 X) - \text{Tr}(\rho_0 X)]^2 / \text{Var}_0 X, \quad (2.19)$$

where $\text{Var}_0 X = \text{Tr}(\rho_0 X^2) - [\text{Tr}(\rho_0 X)]^2$. The operator X for which Δ^2 is maximum can be shown, by using the Schwarz inequality for traces, to be the solution of the operator equation

$$\rho_1 - \rho_0 = \frac{1}{2} (\rho_0 X + X \rho_0) \quad (2.20)$$

[Helstrom, 1969d, p. 241]. Letting S represent the signal strength, with $\rho_1 = \rho_1(S)$, $\rho_0 = \rho_1(0)$, we define the symmetrized logarithmic derivative (s.l.d.) of $\rho_1(S)$ as the solution L of the equation

$$\partial \rho_1 / \partial S = \frac{1}{2} (\rho_0 L + L \rho_0). \quad (2.21)$$

The threshold detection operator Π_0 is then the value of L in the limit $S \rightarrow 0$. Often the threshold operator is much easier to determine than the optimum detection operator $U(\rho_1 - \Lambda_0 \rho_0)$.

4. Multiple Hypotheses

The distinction between the quantum-mechanical detection theory and the classical lies in this, that classically everything about the field can in principle be measured, and there is no ambiguity about carrying the likelihood ratio $\Lambda(\mathbf{x})$ to the limit where all the information in the field is utilized; quantum-mechanically it is necessary also to decide what to measure, for it is impossible to measure simultaneously all the operators corresponding to a complete classical description of the field. In binary detection the quantum theory discovers the best operator to measure in order to choose between two hypotheses H_0 and H_1 with minimum average cost. When decisions among more than two hypotheses are required, however, the optimum operators to measure are as yet unknown, except when the associated density operators commute.

In a laser communication system transmitting every T seconds one of M different coherent signals, the observer is confronted with the choice among M density operators $\rho_1, \rho_2, \dots, \rho_M$ for the field in the lossless cavity representing the ideal receiver. His strategy can be described as one of measuring M projection operators $\Pi_1, \Pi_2, \dots, \Pi_M$, forming a resolution of the identity,

$$\Pi_1 + \Pi_2 + \dots + \Pi_M = \mathbf{1}, \quad (2.22)$$

and he chooses hypothesis H_i when Π_i yields the value 1 and the rest yield 0. In order that the operators Π_k be measurable on the same system, they must commute. In terms of the costs and prior probabilities introduced in ch. 1 §5, the average cost of operation will be, as in eq. (1.24),

$$\bar{C} = \sum_{i=1}^M \sum_{j=1}^M \zeta_j C_{ij} \text{Tr}(\Pi_i \rho_j), \quad (2.23)$$

for $\text{Tr}(\Pi_i \rho_j)$ is the probability under hypothesis H_j that the outcome of measuring Π_i is 1.

When the density operators ρ_k commute among themselves, they possess a common set of eigenstates, and it is necessary simply to determine which of those states the system is in. The classical procedure described in ch. 1 §5 can then be applied to deciding which hypothesis is true. Optical communication systems involving the choice among commuting density operators have been treated by Liu [1970]. For noncommuting density operators, the optimum resolution of the identity $(\Pi_1, \Pi_2, \dots, \Pi_M)$, that is, the one minimizing $\bar{C}$ in eq. (2.23), is unknown. Although some conditions on the solution of this problem have been reported, it is unclear how they can be applied [Yuen et al., 1970].

3. Detection of a Coherent Signal

1. The Transmission-Line Receiver: Classical Analysis

The optimum quantum receiver has been prescribed in terms of the density operators of the electromagnetic field within a lossless cavity, which is exposed to the incident light during an observation interval $(0, T)$. These density operators are evaluated at an arbitrary time $t > T$ after the aperture has been closed, but they depend on the field at the aperture during the interval $(0, T)$. As it is in terms of the aperture field that we want to express the detectability of light signals, we must specify the relation between it and the internal field at $t > T$.

It is instructive to begin with a one-dimensional counterpart of the ideal receiver, a lossless transmission line into which the signal and the noise are introduced by a voltage source having a real internal impedance Z_s (Fig. 3.1). The source might represent an antenna, whose radiation resistance contributes to Z_s and whose terminals exhibit a randomly fluctuating noise voltage $n(t)$ as a result of background fields incident upon it. When the field $\mathcal{F}$ to be detected also falls upon the surface of the antenna, the terminals experience an additional voltage $s(t)$, which here becomes "the signal". The field surrounding the transmission line, or equivalently the voltage $V(x, t)$ and the current $I(x, t)$ at points x along it, corresponds to the field in the lossless cavity, but having a single spatial dimension can be analyzed with less cumbersome mathematics. The transmission-line receiver has previously been treated by She [1965, 1968].

The transmission line will first be studied on the basis of classical

physics, and by forming a likelihood ratio from the amplitudes of its normal modes, we shall determine the detectability of a coherent signal pulse of known form and finite duration. Next, the normal modes will be quantized, and the quantum detection theory of ch. 2 will be applied to the same problem. Finally we shall see how to treat the effect of amplification on signal detectability in this ideal receiver.

As shown in Fig. 3.1, the transmission line, whose length is ℓ , is open at the far end. At first, the line is quiescent. The generator, whose voltage output is $v(t)$, is connected during the observation interval $(0, T)$, which is long enough to contain the entire signal $s(t)$. The line has an inductance L per unit length and a capacitance C per unit length. The velocity of propagation along the line is $c = (LC)^{-1/2}$, and we assume that $\ell = cT$. The characteristic impedance of the line is $Z_0 = (L/C)^{1/2}$.

At time t the voltage between point x of the line and ground is $V(x, t)$, and the current moving in the positive x -direction is $I(x, t)$. These are related by the partial differential equations

$$-\partial V/\partial x = L \partial I/\partial t, \quad -\partial I/\partial x = C \partial V/\partial t, \quad (3.1)$$

with the boundary conditions

$$V(0, t) + Z_0 I(0, t) = v(t), \quad (3.2)$$

$$I(\ell, t) = 0. \quad (3.3)$$

The general solutions of these equations represent waves moving to the left and to the right, and they can be put into the form

$$V(x, t) = f_+(t - x/c) + f_-(t + x/c), \quad (3.4)$$

$$Z_0 I(x, t) = f_+(t - x/c) - f_-(t + x/c). \quad (3.5)$$

The entire system being linear, the principle of superposition applies, and we can treat the components due to signal and noise separately. The rightward

and leftward waves can be expanded in Fourier series

$$f_+(t - x/c) = \sum_{m=-\infty}^{\infty} g_m \exp [-i\omega_m(t - x/c)], \quad (3.6)$$

$$f_-(t + x/c) = \sum_{m=-\infty}^{\infty} h_m \exp [-i\omega_m(t + x/c)], \quad (3.7)$$

where $\omega_m = 2\pi m/T$. The terms of these series correspond to normal modes of the field between the line and ground when the ends of the line are open. Because $V(x, t)$ and $I(x, t)$ are real, $g_{-m} = g_m^*$, $h_{-m} = h_m^*$.

We begin within the framework of classical physics, disregarding quantum fluctuations. When a signal $s(t)$ is present (hypothesis H_1), the input is

$$v(t) = s(t) + n(t), \quad (3.8)$$

where $n(t)$ is a white, Gaussian random noise process with mean value 0 and autocovariance function given by Nyquist's law as

$$\phi(\tau) = E[n(t) n(t + \tau)] = 2K\mathcal{T}Z_S\delta(\tau), \quad (3.9)$$

where K is Boltzmann's constant and $\mathcal{T}$ is the effective absolute temperature of the source. During the interval $(0, T)$ the voltage and current on the line are

$$V(x, t) = Z_0(Z_S + Z_0)^{-1} v(t - x/c), \quad (3.10)$$

$$I(x, t) = (Z_S + Z_0)^{-1} v(t - x/c), \quad (3.11)$$

which satisfy the conditions in eqs. (3.2), (3.3), $v(t)$ being zero for $t < 0$. The line supports a wave moving to the right.

At time $t = T$ we disconnect the source and measure the voltage and current along the line, or what is the same thing, the amplitudes of the normal modes defined in eqs. (3.6) and (3.7). All the h_m 's are zero, there being as yet no leftward waves. The coefficients g_m of the rightward modes, on the other hand, are given by

$$\begin{aligned}
g_m &= g_{mx} + i g_{my} = \ell^{-1} \int_0^{\ell} f_+(T - x/c) \exp [-i\omega_m x/c] dx \\
&= T^{-1} Z_0(Z_0 + Z_S)^{-1} \int_0^T v(t) \exp (i\omega_m t) dt.
\end{aligned} \tag{3.12}$$

As these are linear functionals of the input, they are Gaussian random variables. Because $g_{-m} = g_m^*$, we fix our attention only on the coefficients with $m \geq 0$, and for these it is easily shown that g_{mx} and g_{ny} are statistically independent for all positive values m and n . Under hypothesis H_0 , $v(t) = n(t)$, the coefficients g_{mx} and g_{my} all have mean value zero ; under hypothesis H_1 their mean values are given by

$$\begin{aligned}
E(g_m | H_1) &= \bar{g}_m = \bar{g}_{mx} + i \bar{g}_{my} = \\
&= T^{-1} Z_0(Z_0 + Z_S)^{-1} \int_0^T s(t) \exp (i\omega_m t) dt.
\end{aligned} \tag{3.13}$$

Their variances are, under both hypotheses,

$$\text{Var } g_{mx} = \text{Var } g_{my} = \sigma^2 = K \mathcal{T} Z_S T^{-1} [Z_0 / (Z_0 + Z_S)]^2 \tag{3.14}$$

as a consequence of eq. (3.9) for the autocovariance of the noise.

In general, the total energy in the electromagnetic field of the transmission line is given by

$$\begin{aligned}
\mathcal{E} &= \frac{1}{2} \int_0^{\ell} \{C[V(x, t)]^2 + L[I(x, t)]^2\} dx = \\
&= 2C\ell \sum_{m=0}^{\infty} (|g_m|^2 + |h_m|^2).
\end{aligned} \tag{3.15}$$

At time T , the portion of the energy in the field due to the signal, when present, is

$$\mathcal{E}_S = 2C\ell \sum_{m=0}^{\infty} |\bar{g}_m|^2 = Z_0(Z_0 + Z_S)^{-2} \int_0^T [s(t)]^2 dt, \tag{3.16}$$

in obtaining which we have used $Cc = Z_0^{-1}$.

The observer, having measured the first M of the mode amplitudes g_m , must form their likelihood ratio in order to decide whether a signal is present, and for this he needs the joint probability density functions (p.d.f.'s) of their real and imaginary parts g_{mx} , g_{my} . As these are Gaussian variables, their joint p.d.f. under hypothesis H_1 is

$$\begin{aligned} p_1(\{g_{mx}, g_{my}\}) &= \prod_{m=0}^{M-1} (2\pi\sigma^2)^{-1} \exp\{-[(g_{mx} - \bar{g}_{mx})^2 + (g_{my} - \bar{g}_{my})^2]/2\sigma^2\} \\ &= \prod_{m=0}^{M-1} (2\pi\sigma^2)^{-1} \exp(-|g_m - \bar{g}_m|^2/2\sigma^2); \end{aligned} \quad (3.17)$$

under H_0 it is

$$p_0(\{g_{mx}, g_{my}\}) = \prod_{m=0}^{M-1} (2\pi\sigma^2)^{-1} \exp(-|g_m|^2/2\sigma^2). \quad (3.18)$$

The likelihood ratio is then

$$\begin{aligned} \Lambda(\{g_m\}) &= p_1(\{g_{mx}, g_{my}\})/p_0(\{g_{mx}, g_{my}\}) \\ &= \prod_{m=0}^{M-1} \exp[(\text{Re } \bar{g}_m^* g_m - \frac{1}{2}|\bar{g}_m|^2)/\sigma^2], \end{aligned}$$

where Re indicates the real part of a complex number. The likelihood ratio, as we learned in ch. 1 §1, is to be compared with a pre-assigned decision level Λ_0 , hypothesis H_1 ("signal present") being chosen when $\Lambda > \Lambda_0$. Because the exponential function is monotone, this procedure is equivalent to the simpler prescription to form the sufficient statistic

$$G_M = \text{Re} \sum_{m=0}^{M-1} \bar{g}_m^* g_m = \sum_{m=0}^{M-1} (\bar{g}_{mx} g_{mx} + \bar{g}_{my} g_{my}) \quad (3.20)$$

and compare it with a decision level G_{M0} depending on the pre-assigned false-alarm probability Q_0 .

As the statistic G_M is a linear combination of Gaussian random variables $\{g_{mx}, g_{my}\}$, it also has a Gaussian distribution under both hypotheses. Its

mean values are, by eq. (3.13),

$$\mathbb{E}(G_M|H_0) = 0, \quad \mathbb{E}(G_M|H_1) = \sum_{m=0}^{M-1} |\bar{g}_m|^2, \quad (3.21)$$

and its variance under both hypotheses is, by eq. (3.14) and the independence of the coefficients,

$$\text{Var } G_M = \sigma_G^2 = \sigma^2 \sum_{m=0}^{M-1} |\bar{g}_m|^2. \quad (3.22)$$

The false-alarm probability is the probability under hypothesis H_0 that G_M exceeds its decision level G_{M0} ,

$$Q_0 = \text{erfc } \xi = (2\pi)^{-1/2} \int_{\xi}^{\infty} \exp(-t^2/2) dt, \\ \xi = G_{M0}/\sigma_G, \quad (3.23)$$

and the probability of detection is similarly

$$Q_d = \Pr(G_M > G_{M0}|H_1) = \text{erfc}(\xi - d_M), \\ d_M = \mathbb{E}(G|H_1)/\sigma_G = \sigma^{-1} \left(\sum_{m=0}^{M-1} |\bar{g}_m|^2 \right)^{1/2}. \quad (3.24)$$

The more modes included in the statistic G_M of eq. (3.20), the larger d_M and Q_d . The maximum detection probability is attained by using all the modes and is given by

$$Q_d = \text{erfc}(\xi - d), \quad (3.25)$$

where by eqs. (3.14) and (3.16)

$$d^2 = \sigma^{-2} \sum_{m=0}^{\infty} |\bar{g}_m|^2 = \frac{(Z_0 + Z_S)^2 \mathcal{E}_S}{2K\mathcal{T}Z_0Z_S}, \quad (3.26)$$

$\mathcal{E}_S$ being the energy of the signal component of the field in the transmission line as in eq. (3.16). This energy is maximum when the transmission line is matched to the source, $Z_0 = Z_S$, and the maximum probability of detection is given by eq. (3.25) with the maximum signal-to-noise ratio

$$d^2 = 2E_s/K\mathcal{T}, \quad E_s = (4Z_s)^{-1} \int_0^T [s(t)]^2 dt. \quad (3.27)$$

The parameter ξ is determined by the false-alarm probability Q_0 through $Q_0 = \text{erfc } \xi$.

The energy E_s is the maximum that can be extracted from the signal field, and the probability of detection we have calculated here is identical to that determined by the conventional detection-theoretical analysis involving the likelihood ratio for the input $v(t) = n(t)$ (H_0) versus the input $v(t) = s(t) + n(t)$ (H_1), $0 < t < T$ [Helstrom, 1968c, ch. 4]. That analysis leads to a receiver consisting of a filter matched to the signal and a subsequent decision device. The output of the matched filter at the end of the observation interval $(0, T)$ is compared with a decision level, and if the level is surpassed, hypothesis H_1 is selected. We have shown here that the field of the transmission line contains at time T all the information necessary to detect the signal in the optimum manner, and the transmission-line receiver, when all its mode amplitudes g_m are processed according to eq. (3.20) with $M = \infty$, is equivalent to the standard optimum receiver of classical detection theory.

2. Quantization of the Receiver

In quantum mechanics the voltage $V(x, t)$, the current $I(x, t)$, and all quantities linearly related to them, such as the mode amplitudes g_m and h_m , become operators acting on the state vector $|\psi\rangle$ of the transmission line and its ambient field. Let us consider first the field when the line is open at both ends, as at time $t < 0$, and let us expand it in normal modes as in eqs. (3.6) and (3.7). The energy in the field is given in terms of the mode amplitudes by eq. (3.15), and hence the Hamiltonian operator of the field must be

$$\mathcal{H} = 2C\ell \sum_{m=0}^{\infty} (g_m^+ g_m + h_m^+ h_m), \quad (3.28)$$

where the adjoint operators g_m^+ , h_m^+ take the place of the complex conjugates g_m^* , h_m^* , respectively. Equations (3.6) and (3.7) show that the operators g_m and h_m must have the time dependences

$$\begin{aligned} g_m(t) &= g_m(t_0) \exp [-i\omega_m(t - t_0)], \\ h_m(t) &= h_m(t_0) \exp [-i\omega_m(t - t_0)], \end{aligned} \quad (3.29)$$

and they must hence obey the differential equations

$$dg_m/dt = -i\omega_m g_m, \quad dh_m/dt = -i\omega_m h_m. \quad (3.30)$$

However, these time derivatives must also be given by the commutators with the Hamiltonian,

$$\begin{aligned} dg_m/dt &= (i/\hbar) [\mathcal{H}, g_m] = (i/\hbar) (\mathcal{H} g_m - g_m \mathcal{H}), \\ dh_m/dt &= (i/\hbar) [\mathcal{H}, h_m], \end{aligned} \quad (3.31)$$

where $\hbar$ is Planck's constant $h/2\pi$. By comparison with eq. (3.30), it follows that the commutators of the operators g_m , h_m and their adjoints must be

$$[g_m, g_n^+] = [h_m, h_n^+] = (\hbar\omega_m/2C\ell)\delta_{mn},$$

$$[g_m, h_n^+] = 0,$$

where δ_{mn} is the Kronecker delta, $\delta_{mn} = 1$, $m = n$, $\delta_{mn} = 0$, $m \neq n$. We thus identify g_m and h_m with the annihilation operators a_m and b_m of the modes, which we define by

$$g_m = (\hbar\omega_m/2C\ell)^{1/2} a_m, \quad h_m = (\hbar\omega_m/2C\ell)^{1/2} b_m. \quad (3.32)$$

The adjoint operators a_m^+ , b_m^+ are the associated creation operators. The annihilation and creation operators obey the standard commutation rules

$$[a_m, a_n^+] = [b_m, b_n^+] = \delta_{mn}, \quad (3.33)$$

$$[a_m, a_n] = [a_m, b_n] = [b_m, b_n] = [a_m, b_n^+] = 0 \quad (3.34)$$

[Louisell, 1964, ch. 4]. In terms of them the Hamiltonian operator takes the usual form

$$\mathcal{H} = \sum_{m=0}^{\infty} \hbar\omega_m (a_m^+ a_m + b_m^+ b_m). \quad (3.35)$$

The normal modes of the field behave like quantum-mechanical harmonic oscillators with angular frequencies ω_m . The operator $a_m^+ a_m$ is the number operator for the m th rightward mode; it has integral eigenvalues, which we identify with the number of photons in this m th mode, each photon contributing an energy $\hbar\omega_m$.

The voltage source producing the input $v(t)$ is to be connected to the transmission line at time $t = 0$, and the field of the line is to be observed at time $T = \ell/c$. We must therefore investigate the behavior of the modes during the interim, when one end of the line is connected to the source with its internal impedance Z_S . For later purposes, we suppose also that the far end of the line terminates in a load impedance Z_L , as in Fig. 3.2. Our aim is to show that the mode operators $a_m(T)$, $b_m(T)$ at time T obey the same

commutation rules (3.33), (3.34), despite the altered terminations of the line. The boundary conditions are now, in place of eq. (3.2),

$$\begin{aligned} V(0, t) + Z_S I(0, t) &= v(t), \\ V(\ell, t) - Z_L I(\ell, t) &= 0. \end{aligned} \quad (3.36)$$

The field at time $t = T$ is composed of two parts, the remnant of the original field at time $t = 0^-$ and the field produced by the source impedance Z_S and the load impedance Z_L . To see what happened to the initial field, we decompose the initial voltage and current into rightward and leftward components $f_+^0(-x/c)$ and $f_-^0(x/c)$ through eqs. (3.4), (3.5), which become for $t = 0$

$$\begin{aligned} V(x, 0) &= f_+^0(-x/c) + f_-^0(x/c), \\ Z_0 I(x, 0) &= f_+^0(-x/c) - f_-^0(x/c), \end{aligned} \quad (3.37)$$

and permit $f_+^0(-x/c)$ and $f_-^0(x/c)$ to be determined. The subsequent behavior of the waves $f_+^0(t - x/c)$ and $f_-^0(t + x/c)$ is governed by eqs. (3.4), (3.5), and (3.36), in the last of which we put $v(t) = 0$. A straightforward analysis shows that during the interval $(0, T)$ the rightward wave f_+^0 is reflected at the right-hand end $x = \ell$ of the line and turns into a leftward wave, diminished in amplitude by the reflection coefficient

$$\beta_L = (Z_L - Z_0)/(Z_L + Z_0); \quad (3.38)$$

the leftward wave f_-^0 is similarly reflected at $x = 0$ and diminished by the factor

$$\beta_S = (Z_S - Z_0)/(Z_S + Z_0), \quad (3.39)$$

becoming a rightward wave. As a result, the portion of the field at time $t = T$ due to the initial field at $t = 0$ is made up of the components

$$\begin{aligned} f_+^1(T - x/c) &= \beta_S f_-^0((\ell - x)/c), \\ f_-^1(T + x/c) &= \beta_L f_+^0(-(\ell - x)/c); \end{aligned} \quad (3.40)$$

we denote these portions by single primes. The corresponding Fourier

coefficients g_m and h_m and the annihilation operators a_m and b_m for the rightward and leftward modes undergo a similar exchange and diminution, the latter becoming

$$a_m'(T) = \beta_S b_m(0), \quad b_m'(T) = \beta_L a_m(0). \quad (3.41)$$

Since $|\beta_S| < 1$ and $|\beta_L| < 1$, these new operators $a_m'(T)$ and $b_m'(T)$ no longer obey the commutation rules in eq. (3.33). The deficiency must be made up from the portion of the field resulting from the terminating impedances Z_S and Z_L . These contain an enormous number of atoms and molecules, ions and electrons whose thermal agitation--as we know from classical statistical mechanics--produces across their terminals fluctuating voltages $n_S(t)$ and $n_L(t)$, respectively, having spectral densities $2K\mathcal{T}_S Z_S$ and $2K\mathcal{T}_L Z_L$, where $\mathcal{T}_S$ and $\mathcal{T}_L$ are their respective absolute temperatures. Lax [1966] and Haus [1970] have shown that these voltages $n_S(t)$ and $n_L(t)$ must be treated as quantum-mechanical operators obeying the commutation rules

$$\begin{aligned} [n_S(t_1), n_S(t_2)] &= 2i\hbar Z_S \delta'(t_1 - t_2), \\ [n_L(t_1), n_L(t_2)] &= 2i\hbar Z_L \delta'(t_1 - t_2), \\ [n_S(t_1), n_L(t_2)] &= 0, \end{aligned} \quad (3.42)$$

where $\delta'(t)$ is the derivative of the delta function.

The randomly varying voltage operators $n_S(t)$ and $n_L(t)$ contribute terms $a_m''(T)$ and $b_m''(T)$ to the mode amplitudes, and by virtue of their disparate physical origins, these terms commute with the portions $a_m'(T)$ and $b_m'(T)$ arising from the initial field at $t = 0^-$. By eqs. (3.12) and (3.32) we find

$$\begin{aligned} a_m''(T) &= (2C\ell/\hbar\omega_m)^{1/2} Z_0(Z_S + Z_0)^{-1} T^{-1} \int_0^T n_S(t) \exp(i\omega_m t) dt, \\ b_m''(T) &= (2C\ell/\hbar\omega_m)^{1/2} Z_0(Z_L + Z_0)^{-1} T^{-1} \int_0^T n_L(t) \exp(i\omega_m t) dt. \end{aligned} \quad (3.43)$$

Using eq. (3.42) we evaluate their commutators as

$$\begin{aligned}
[a_m''(T), a_m^{''+}(T)] &= (2C\ell/\hbar\omega_m) Z_0^2(Z_S + Z_0)^{-2} T^{-2} \\
&\times \int_0^T \int_0^T [n_S(t_1), n_S(t_2)] \exp[i\omega_m(t_1 - t_2)] dt_1 dt_2 \\
&= (2C\ell/\hbar\omega_m) Z_0^2(Z_S + Z_0)^{-2} T^{-2} \\
&\times (2i\hbar Z_S) \int_0^T \int_0^T \delta'(t_1 - t_2) \exp[i\omega_m(t_1 - t_2)] dt_1 dt_2 \\
&= 4Z_0 Z_S / (Z_S + Z_0)^2 = 1 - \beta_S^2, \tag{3.44}
\end{aligned}$$

where we have used $\ell/T = c = (LC)^{-1/2} = (CZ_0)^{-1}$, and where the reflection coefficient β_S is defined by eq. (3.39). Thus with $a_m(T) = a_m'(T) + a_m''(T)$, we find from eqs. (3.41) and (3.44) the commutator

$$[a_m(T), a_m^+(T)] = \beta_S^2 [b_m(0), b_m^+(0)] + 1 - \beta_S^2 = 1. \tag{3.45}$$

Similarly, $[b_m(T), b_m^+(T)] = 1$. Because of the orthogonality of the modes, the commutativity of the operators for different modes continues to hold. We can thus treat the transmission line at time $t = T$ in terms of the annihilation operators $a_n(T)$, $b_n(T)$ and their adjoint creation operators $a_n^+(T)$, $b_n^+(T)$; and these have all the ordinary quantum-mechanical properties deriving from the representation of the modes as independent harmonic oscillators.

3. The Coherent Signal of Known Phase

3.1 The Density Operators

Now we can return to the detection of a coherent signal of known form, analyzed from the classical standpoint in §1, and treat it quantum-mechanically. As before, we imagine measuring the field modes at time $t = T$; but now, in place of the p.d.f.'s of the mode amplitudes, we need the density operators ρ_0 and ρ_1 of the modes under the two hypotheses H_0 ("signal absent") and H_1 ("signal present"). The receiver is the same as the transmission line discussed in §2, but the far end $x = \ell$ is open, $Z_L = \infty$. We learned that we can treat the modes at time $t = T$ in terms of the annihilation operators $a_m = a_m(T)$ and their adjoints $a_m^\dagger = [a_m(T)]^\dagger$ for the rightward modes, and that these obey the usual commutation rules, eqs. (3.33), (3.34), for annihilation and creation operators. At $t = T$ the leftward modes contain neither signal nor noise and can be disregarded.

Because the noise arises from the enormous number of particles in the source impedance Z_S --or equivalently, from the thermal background field picked up by the antenna--, its values have, by the central-limit theorem of statistics, Gaussian distributions. Quantum-mechanically this means that the density operators ρ_0 and ρ_1 have Gaussian P-representations and depend only on the expected values and variances of the operators a_m [Glauber, 1963]. We start with a finite number M of modes, allowing M later to go to infinity.

By eqs. (3.32) and (3.12), we see that under hypothesis H_0 the expected value of a_m is

$$\mathbb{E}[a_m | H_0] = \text{Tr}[\rho_0 a_m(T)] = 0; \quad (3.46)$$

under H_1 it is

$$\mu_m = E[a_m | H_1] = \text{Tr}[\rho_1 a_m(T)] = (2Cl/\hbar\omega_m)^{1/2} \bar{g}_m, \quad (3.47)$$

with $\bar{g}_m$ given by eq. (3.13). We can call $|\mu_m|^2$ the average number of signal photons in mode m ; the energy of the signal component of the field is, by eq. (3.16),

$$\mathcal{E}_S = \text{Tr}[(\rho_1 - \rho_0)\mathcal{H}] = \sum_{m=0}^{\infty} \hbar\omega_m |\mu_m|^2. \quad (3.48)$$

Under both hypotheses the mean number of noise photons in the m -th mode is

$$\begin{aligned} \mathcal{N}_m' &= E(a_m^\dagger a_m | H_0) = \text{Tr}(\rho_0 a_m^\dagger a_m) = 4Z_0 Z_S (Z_0 + Z_S)^{-2} \mathcal{N}_m, \\ \mathcal{N}_m &= [\exp(\hbar\omega_m/K\mathcal{T}) - 1]^{-1} = \text{pl}(\hbar\omega_m/K\mathcal{T}), \end{aligned} \quad (3.49)$$

where $\mathcal{T} = \mathcal{T}_S$ is the effective absolute temperature of the source and

$$\text{pl}(x) = (e^x - 1)^{-1} \quad (3.50)$$

is the Planck factor. In the classical limit $\hbar\omega_m \ll K\mathcal{T}$, eq. (3.49) reduces to eq. (3.14) by virtue of eq. (3.32). From eqs. (3.46), (3.47), and (3.49) the Gaussian P-representations of the density operators ρ_0 and ρ_1 can be written down. The optimum processing, which requires solution of eq. (2.14), has not been determined.

A decisive simplification follows from the realistic assumption that the signal has so narrow a bandwidth W that the mean numbers $\mathcal{N}_m'$ of noise photons in all modes affected by the signal are equal. By eq. (3.49) this requires $W \ll K\mathcal{T}/\hbar$, and as $K\mathcal{T}/\hbar = 6.2 \times 10^{12} \text{ sec.}^{-1}$ at $\mathcal{T} = 300^\circ\text{K}$, this condition will usually be met. We can then put for all modes

$$\mathcal{N}_m \equiv \mathcal{N} = \text{pl}(\hbar\Omega/K\mathcal{T}), \quad (3.51)$$

where Ω is the central angular frequency of the signal.

We can now introduce a new set of modes by a unitary transformation of the original amplitudes a_m [Helstrom, 1968b]. We put

$$c_n = \sum_{m=0}^{M-1} U_{nm} a_m,$$

where $U = \|U_{nm}\|$ is a unitary matrix. The operators c_n and their adjoints c_n^+ then obey the same commutation rules as the a_n 's and a_n^+ 's:

$$[c_n, c_m^+] = \delta_{nm}, [c_n, c_m] = 0. \quad (3.52)$$

Furthermore, the new modes will be statistically independent and contain mean numbers

$$\mathcal{N}' = 4Z_0 Z_S (Z_S + Z_0)^{-2} \mathcal{N} \quad (3.53)$$

of noise photons. The unitary matrix is so chosen that

$$c_1 = v_M^{-\frac{1}{2}} \sum_{m=0}^{M-1} \mu_m^* a_m, \quad (3.54)$$

where

$$v_M = \sum_{m=0}^{M-1} |\mu_m|^2; \quad (3.55)$$

i.e., $U_{1m} = v_M^{-\frac{1}{2}} \mu_m^*$. The remaining rows U_{nm} of U are orthogonal to U_{1m} . The expected values of the c_n 's will then all vanish under both hypotheses, except for c_1 , for which

$$\mathbb{E}(c_1 | H_1) = \text{Tr}(\rho_1 c_1) = \Gamma, \quad |\Gamma|^2 = v_M. \quad (3.56)$$

We can thus disregard all the new modes except the first, for they are unaffected by the signal; the optimum receiver works only with the field in the mode whose annihilation operator is c_1 . The resemblance of c_1 to the statistic G_M in eq. (3.20) is apparent; the quantum counterpart to G_M is the Hermitian operator

$$\mathcal{Q} = \frac{1}{2} (c_1 + c_1^+). \quad (3.57)$$

We call this new mode the "matched mode"; in effect, it correlates the amplitudes a_m with the signal components μ_m .

The density operators for the matched mode under the two hypotheses are, in the P-representation,

$$\rho_0 = (\pi\mathcal{N}')^{-1} \int \exp(-|\gamma|^2/\mathcal{N}') |\gamma\rangle\langle\gamma| d^2\gamma, \quad (3.58)$$

$$\rho_1 = (\pi\mathcal{N}')^{-1} \int \exp(-|\gamma - \Gamma|^2/\mathcal{N}') |\gamma\rangle\langle\gamma| d^2\gamma, \quad (3.59)$$

where $|\gamma\rangle$ is a coherent state [Glauber, 1963], $\gamma = \gamma_x + i\gamma_y$, $d^2\gamma = d\gamma_x d\gamma_y$, and the integration is carried out over the entire (γ_x, γ_y) -plane. These density operators must be substituted into eq. (2.14), and the eigenvalues η_k and eigenstates $|\eta_k\rangle$ must be determined.

3.2 The Extreme Quantum Limit

For arbitrary values of $\mathcal{N}'$ and Γ the exact solution of this eigenvalue problem is unknown. For $\mathcal{T} = 0$ and $\mathcal{N}' = 0$, however, the matched mode is in a pure coherent state under each hypothesis, the density operators are simply

$$\rho_0 = |0\rangle\langle 0|, \quad \rho_1 = |\Gamma\rangle\langle \Gamma|, \quad (3.60)$$

and the method of ch. 2 §2 can be applied. (Our assumption that $\hbar W \ll K\mathcal{T}$, where W is the bandwidth of the signal, is unnecessary when as here $\mathcal{N}' = 0$ for all modes.) The false-alarm and detection probabilities can be calculated from eq. (2.18), in which

$$q = 1 - |\langle 0|\Gamma\rangle|^2 = 1 - \exp(-\nu_M).$$

The detection probability is maximum when $M = \infty$; then by eq. (3.48) with $\omega_m \cong \Omega$ for the significant modes

$$q = 1 - \exp(-N_S'), \quad N_S' = \sum_{m=0}^{\infty} |\mu_m|^2 = \mathcal{E}_S/\hbar\Omega, \\ \mathcal{E}_S = 4Z_S Z_0 (Z_S + Z_0)^{-2} E_S, \quad (3.61)$$

where E_S is the available signal energy as defined in eq. (3.27). The detection probability is maximum when the transmission line is matched to the source, $Z_S = Z_0$, $\mathcal{E}_S = E_S$, whereupon $N_S' = N_S = E_S/\hbar\Omega$ equals the average number of photons provided by the source of the signal. In Fig. 3.3 this maximum probability Q_d of detection is plotted versus the equivalent signal-to-noise ratio $D_q = 2N_S^{1/2}$ as the curve marked "optimum"; the false-alarm probability is $Q_0 = 10^{-4}$.

When the false-alarm probability is very small, $Q_0 \ll 1$, the probability of detection is approximately, by eq. (2.18),

$$Q_d \approx q = 1 - \exp(-N_s), \quad (3.62)$$

which--as we shall see--is the detection probability attained by a receiver that disregards the phase of the signal and simply counts the number of photons in the matched mode. The difference between the maximum possible value of Q_d and that in eq. (3.62) is of the order of $Q_0^{\frac{1}{2}}$.

3.3 The Threshold Receiver

The optimum receiver for a coherent signal of known phase depends on the amplitude of the expected signal; the corresponding statistical test is not uniformly most powerful. In view of the added disadvantage that the required eigenvalue equation is difficult to solve, it is natural to turn to the threshold receiver, which was described in ch. 2 §3. It is not hard to show from eqs. (3.58) and (3.59) that the threshold operator Π_θ is proportional to the operator $\mathcal{Q} = \frac{1}{2} (c_1 + c_1^\dagger)$ given in eq. (3.57) as the quantum counterpart of the classical sufficient statistic G_M [Helstrom, 1967b]. The threshold receiver measures the Hermitian operator $\mathcal{Q}$ and compares the outcome $\mathcal{Q}'$ with a decision level $\mathcal{Q}_0$, declaring a signal present when $\mathcal{Q}' > \mathcal{Q}_0$.

This outcome $\mathcal{Q}'$ has a Gaussian distribution by virtue of the Gaussian forms of the density operators ρ_0 and ρ_1 , eqs. (3.58), (3.59). In the limit $M \rightarrow \infty$ the means and variance of $\mathcal{Q}'$ are, by eqs. (3.54), (3.49),

$$\begin{aligned} E[\mathcal{Q}' | H_0] &= 0, \quad E[\mathcal{Q}' | H_1] = \text{Tr}(\rho_1 \mathcal{Q}) = N_S'^{\frac{1}{2}}, \\ \text{Var } \mathcal{Q}' &= \text{Tr}(\rho_0 \mathcal{Q}^2) = \\ &= \lim_{M \rightarrow \infty} \frac{1}{4} v_M^{-1} \sum_{m=0}^{M-1} |u_m|^2 \text{Tr}(a_m a_m^\dagger + a_m^\dagger a_m) = \frac{1}{4} (2\mathcal{N}' + 1), \end{aligned} \quad (3.64)$$

and the false-alarm and detection probabilities are, as in §1,

$$Q_0 = \text{erfc } \xi, \quad Q_d = \text{erfc } (\xi - D_q), \quad (3.65)$$

where D_q is an equivalent signal-to-noise ratio defined by

$$D_q^2 = 4N_S' / (2\mathcal{N}' + 1), \quad (3.66)$$

which again is maximum when $Z_S = Z_0$, $N_S' = N_S$, $\mathcal{N}' = \mathcal{N}$. The probability Q_d

for $Q_0 = 10^{-4}$ is plotted versus D_q in Fig. 3.3 as the dashed line marked "threshold". In the classical limit $K\mathcal{T} \gg \hbar\Omega$, D_q^2 goes to the signal-to-noise ratio $d^2 = 2E_S/K\mathcal{T}$ defined in eq. (3.27). By the correspondence principle, we expect the optimum quantum receiver to tend ever closer to one that measures $\mathcal{Q}$ as the effective temperature $\mathcal{T}$ or the average number $\mathcal{N}$ of noise photons increases.

3.4 Approximations to the Optimum Receiver

For small values of $\mathcal{N}$ the eigenvalue equation (2.6) with ρ_0 and ρ_1 given by eqs. (3.58), (3.59) has been solved by Yoshitani [1970], taking $\Lambda_0 = 1$ as for a binary communication system transmitting 0's and 1's with equal probability. He used a perturbation method starting with the exact solution for $\mathcal{N} = 0$ described in §3.2. An alternative approach is to diagonalize the matrix $\langle m | (\rho_1 - \lambda \rho_0) | n \rangle$ obtained by expressing the density operators in the number representation, with

$$c_1^\dagger c_1 |n\rangle = n |n\rangle, \quad n = 0, 1, 2, \dots, \quad (3.67)$$

defining the eigenstates $|n\rangle$ of ρ_0 . The matrix elements are

$$\langle n | \rho_0 | m \rangle = (1 - v) v^n \delta_{nm}, \quad v = \mathcal{N} / (\mathcal{N} + 1), \quad (3.68)$$

$$\begin{aligned} \langle n | \rho_1 | m \rangle &= (1 - v) (n! / m!)^{\frac{1}{2}} v^m (\Gamma / \mathcal{N})^{m-n} \\ &\times \exp[(1 - v) |\Gamma|^2] L_n^{m-n} [-(1 - v)^2 |\Gamma|^2 / v], \quad m \geq n, \end{aligned}$$

$$\langle n | \rho | m \rangle = \langle m | \rho | n \rangle^*, \quad m < n, \quad (3.69)$$

where $L_n^{m-n}(x)$ is the associated Laguerre polynomial [Louisell and Walker, 1965].

One truncates the matrix to such a size that the neglected elements are insignificant. The average probability of error

$$P_e = \frac{1}{2} (Q_0 + 1 - Q_d),$$

which is given by eq. (2.9) with $\zeta = \frac{1}{2}$, $C_1 = 1$, is plotted in Fig. 3.4 versus the signal-to-noise ratio $D_q = [4N_s / (2\mathcal{N} + 1)]^{\frac{1}{2}}$ for a matched receiver ($Z_0 = Z_s$). The line marked $\mathcal{N} = \infty$ represents the error probability $P_e = \text{erfc} \left(\frac{1}{2} D_q \right)$ attained at all values of $\mathcal{N}$ by a receiver measuring the threshold operator $\mathcal{Q}$ of eq. (3.57). The optimum quantum receiver has a significantly smaller

probability P_e of error only for small values of the average number $\mathcal{N}$ of noise photons per mode.

We remarked in ch. 2 §1 that when the false-alarm probability Q_0 is very small, the optimum receiver is nearly the same as one that measures the operator ρ_0 and bases its decision on the outcome. In the present case, measuring ρ_0 is equivalent to measuring $c_1^\dagger c_1$, that is, to counting the number of photons in the matched mode. We saw in §3.2 that in the extreme quantum limit $\mathcal{N} = 0$ the optimum receiver yields only a slightly larger probability of detection than one that disregards the known phase of the signal and counts photons. We conjecture that the same holds true for small as well as for zero average numbers $\mathcal{N}$ of noise photons per mode.

In Fig. 3.3 we have plotted versus the signal-to-noise ratio D_q the probability of detection attained by a receiver that counts the number of photons in the matched mode. How these were calculated will be shown in the next section. The curve marked "optimum" gives the detection probability attained by the optimum receiver for $\mathcal{N} = 0$. The dashed line in Fig. 3.3 indicates the detection probability for the threshold receiver at all values of $\mathcal{N}$. For an average number $\mathcal{N}$ of noise photons greater than about 0.2 the threshold receiver is superior to the photon counter when $Q_0 = 10^{-4}$, but the difference is small.

4. Reception of a Signal of Random Phase

The receiver just described needs to know the form of the expected signal exactly, and when--as we have assumed--the signal is a pulse modulation of a carrier of frequency Ω , the phase of the carrier upon arrival of the signal at the receiver must be known as well. The distance between transmitter and receiver must therefore be known within a fraction of a wavelength $\lambda = 2\pi c/\Omega$, or else the carrier phase, as in a communication system, must have been as accurately tracked from some initially precise value. At very high frequencies such knowledge of the phase will most often be unavailable. The signal will have the form

$$s(t) = \text{Re} [F(t) e^{i\Omega t + i\psi}], \quad (3.70)$$

where $F(t)$ is a known complex envelope, but the phase ψ will ordinarily be unknown. It is as likely to have one value as another, and the least favorable distribution of the phase is the uniform one,

$$z(\psi) = (2\pi)^{-1}, \quad 0 \leq \psi < 2\pi. \quad (3.71)$$

When this signal appears in the transmission line of our ideal receiver, all its complex mode amplitudes $\bar{g}_m$ will have a common phase factor $e^{i\psi}$. It is still possible to form a matched mode as in eq. (3.54) by taking the coefficients U_{1m} proportional to $\bar{g}_m$ for $\psi = 0$. The remaining new modes can be made independent of the first for all values of ψ , and as the signal does not affect them, it suffices to consider only the field in the matched mode. In place of eq. (3.56) the mean value of the operator c_1 when the signal is present with phase ψ is

$$E(c_1 | H_1, \psi) = \Gamma e^{i\psi}. \quad (3.72)$$

The density operator ρ_0 of the mode in the absence of a signal is the same as in eq. (3.58); when the signal in eq. (3.70) is present, the density operator is

$$\rho_1(\psi) = (\pi\mathcal{N}')^{-1} \int \exp(-|\gamma - \Gamma e^{i\psi}|^2/\mathcal{N}') |\gamma\rangle\langle\gamma| d^2\gamma \quad (3.73)$$

in place of eq. (3.59). Since ψ is unknown and random with the uniform prior p.d.f. $z(\psi)$ of eq. (3.71), the actual density operator under hypothesis H_1 is the average

$$\begin{aligned} \rho_1 &= \int_0^{2\pi} z(\psi) \rho_1(\psi) d\psi \\ &= (\pi\mathcal{N}')^{-1} \int \exp[-(|\gamma|^2 + |\Gamma|^2)/\mathcal{N}'] I_0(2|\Gamma\gamma|/\mathcal{N}') |\gamma\rangle\langle\gamma| d^2\gamma, \end{aligned} \quad (3.74)$$

where $I_0(x)$ is the modified Bessel function of the first kind. Since in the P-representation both ρ_0 and ρ_1 depend on the complex variable γ only through its absolute value $|\gamma|$, they must be diagonal in the number representation. From eq. (3.69) we obtain, by averaging over ψ after replacing Γ by $\Gamma e^{i\psi}$,

$$\rho_i = \sum_{n=0}^{\infty} P_{in} |n\rangle\langle n|, \quad i = 0, 1, \quad (3.75)$$

with

$$\begin{aligned} P_{0n} &= (1-v) v^n, \quad v = \mathcal{N}'/(\mathcal{N}' + 1), \\ P_{1n} &= (1-v) v^n \exp[-(1-v)N_s'] L_n[-(1-v)^2 N_s'/v] \end{aligned} \quad (3.76)$$

[Lachs, 1965]. As the density operators commute, the optimum receiver reduces to one measuring the number n of photons in the matched mode. A receiver based on the Bayes criterion will form the likelihood ratio $\Lambda(n) = P_{1n}/P_{0n}$ for the actual number n of photons counted, and it will decide that a signal is present if $\Lambda(n)$ exceeds a decision level defined as in eq. (1.6).

If the Neyman-Pearson criterion governs the design of the receiver, randomization must be employed, as described in ch. 1 §2. Hypothesis H_1 is chosen

whenever n exceeds a certain integer v , H_0 whenever $n < v$; when $n = v$, H_1 is chosen with a certain probability f . The false-alarm and detection probabilities are now, as in eqs. (1.12) and (1.13),

$$Q_0 = f(1 - v) v^v + v^{v+1}, \quad (3.77)$$

$$Q_d = fP_{1v} + \sum_{n=v+1}^{\infty} P_{1n}. \quad (3.78)$$

Here v is the greatest integer in $\ln Q_0 / \ln v$. From these formulas the curves in Fig. 3.3 were constructed. [See Helstrom, 1969a for additional graphs of Q_d .] In the limit $\mathcal{N} \gg 1$ the false-alarm and detection probabilities are given approximately by

$$\begin{aligned} Q_0 &= \exp(-\beta^2/2), \\ Q_d &= Q(\alpha, \beta) = \int_{\beta}^{\infty} x \exp\left[-\frac{1}{2}(\alpha^2 + x^2)\right] I_0(\alpha x) dx, \\ \alpha &= (2N_S/\mathcal{N}')^{1/2}, \end{aligned} \quad (3.79)$$

where $Q(\alpha, \beta)$ is Marcum's Q -function. These probabilities are the same as for the optimum classical receiver of a signal of random phase, which passes the output of the matched filter through a rectifier and compares the rectified output at time $t = T$ with a decision level related to β [Helstrom, 1968c, pp. 166-171].

5. Amplification

Let us return to our model of the ideal receiver as the transmission line shown in Fig. 3.1. The signal, when present at the input, has come in during the interval $(0, T)$ and at time $t = T$ occupies the entire line, whose length is $\ell = cT$. At time $t = T$ the input with its impedance Z_S is detached, and a load impedance Z_L is attached at the right-hand end ($x = \ell$) of the line. The signal and noise waves occupying the line and moving to the right are reflected and attenuated by the load, which injects additional noise at temperature $\mathcal{T}_L$.

At time $t = 2T$ the signal, when present, again occupies the entire line and is represented by

$$V(x, 2T) = -Z_0 I(x, 2T) = \beta_L Z_0 (Z_S + Z_0)^{-1} s(x/c), \quad (3.80)$$

where β_L is the reflection coefficient defined in eq. (3.38). The field is again expanded in a Fourier series as in eq. (3.7), and classically the leftward modes have signal components

$$\bar{h}_m = \left(\frac{\beta_L Z_0}{Z_S + Z_0} \right) \frac{1}{T} \int_0^T s(t) \exp(i\omega_m t) dt. \quad (3.81)$$

The noise components of $h_m = h_{mx} + ih_{my}$ have the variances

$$\begin{aligned} \text{Var } h_{mx} &= \text{Var } h_{my} = \\ K \mathcal{T}_L \beta_L^2 Z_S [Z_0 / (Z_S + Z_0)]^2 T^{-1} &+ K \mathcal{T}_L Z_L [Z_0 / (Z_L + Z_0)]^2 T^{-1}, \end{aligned} \quad (3.82)$$

as in eq. (3.14). Thus the mean number of signal photons in the matched mode whose annihilation operator is

$$d_1 = \left[\sum_{m=0}^{\infty} |\mu_m|^2 \right]^{-1/2} \sum_{m=0}^{\infty} \mu_m^* b_m,$$

$$b_m = (2C\ell/\hbar\Omega)^{\frac{1}{2}} h_m, \quad \mu_m = (2C\ell/\hbar\Omega)^{\frac{1}{2}} \bar{h}_m, \quad (3.83)$$

is

$$N_S'' = \beta_L^2 N_S', \quad (3.84)$$

where N_S' is given by eq. (3.61). The mean number of noise photons is, by comparison of eqs. (3.82), (3.14), and (3.49),

$$\begin{aligned} \mathcal{N}'' &= \beta_L^2 \mathcal{N}' + (1 - \beta_L^2) \text{pl}(\hbar\Omega/K\mathcal{T}_L), \\ \mathcal{N}' &= (1 - \beta_S^2) \text{pl}(\hbar\Omega/K\mathcal{T}), \end{aligned} \quad (3.85)$$

where we have replaced the classical noise factor $K\mathcal{T}/\hbar\Omega$ by the Planck factor $\text{pl}(\hbar\Omega/K\mathcal{T})$ for both the source and the load. With $|\beta_L| = 1$ ($Z_L = 0$ or $Z_L = \infty$), the processing of the leftward modes can take place at time $t = 2T$ with the same effectiveness as for detecting the signal in the rightward modes at time $t = T$. Otherwise the signal will, with $|\beta_L| < 1$, be weakened, and additional noise will be added by the load.

Let us henceforth for simplicity assume that the source and the transmission line are matched so that $Z_0 = Z_S$, $\beta_S = 0$, and

$$\begin{aligned} N_S'' &= \beta_L^2 N_S, \\ \mathcal{N}'' &= \beta_L^2 \text{pl}(\hbar\Omega/K\mathcal{T}) + (1 - \beta_L^2) \text{pl}(\hbar\Omega/K\mathcal{T}_L), \\ \beta_L^2 &= \left(\frac{Z_L - Z_0}{Z_L + Z_0} \right)^2, \end{aligned} \quad (3.86)$$

and let us ask what happens if Z_L is a negative resistance, $Z_L < 0$, such as might appear in an idealized amplifier. Then $\beta_L^2 > 1$ and $N_S'' > N_S$; the signal in the transmission line at time $t = 2T$ has been amplified, the additional photons having been supplied by the power source driving the negative resistance. We know that a quantum amplifier such as the maser contains a great many atoms or molecules, the population of whose energy levels has been inverted by pumping. As discussed by Takahasi [1965], this reversed distribution can be

idealized as a Planck distribution corresponding to a negative temperature $\mathcal{T}_L < 0$. The Planck factor $\text{pl}(\hbar\Omega/K\mathcal{T}_L)$ is then negative, and the average number of noise photons in the matched mode is

$$\mathcal{N}'' = G\mathcal{N} + (G - 1)\mathcal{N}_a, \quad (3.87)$$

$$\mathcal{N}_a = -\text{pl}(-\hbar\Omega/K|\mathcal{T}_L|) = \exp(\hbar\Omega/K|\mathcal{T}_L|) \text{pl}(\hbar\Omega/K|\mathcal{T}_L|), \quad (3.88)$$

where $G = \beta_L^2$ is the amplification factor, or gain,

$$N_S'' = GN_S. \quad (3.89)$$

The minimum number $(G - 1)\mathcal{N}_a$ of additional photons is added when $\mathcal{T}_L = -\infty$, $\mathcal{N}_a = 1$,

$$\mathcal{N}_{\min}'' = G(\mathcal{N} + 1) - 1. \quad (3.90)$$

Our earlier discussion of how the commutation rules are maintained by the presence of an additional driving voltage in the load Z_L applies here as well; the commutator

$$[b_m''(T), b_m^{''+}(T)] = 4Z_0Z_L/(Z_L + Z_0)^2$$

corresponding to eq. (3.44) takes a negative sign with Z_L . Further details can be found in the article by Takahasi [1965].

For a coherent signal of known phase, the best processing of the amplified mode is, in the limit $G \gg 1$, to measure the operator $\frac{1}{2}(d_1 + d_1^+)$. The detection probability then depends, as in eqs. (3.65) and (3.66), on the equivalent signal-to-noise ratio

$$D_q^{''2} = 4N_S''/(2\mathcal{N}'' + 1) = \frac{4GN_S}{2G(\mathcal{N} + 1) - 1} \rightarrow \frac{2N_S}{\mathcal{N} + 1}, \quad G \gg 1. \quad (3.91)$$

This maximum attainable signal-to-noise ratio is always less than the effective signal-to-noise ratio $D_q^2 = 2N_S/(\mathcal{N} + \frac{1}{2})$ characteristic of the mode before amplification. Thus amplification results in diminishing the detectability of the signal.

When the phase of the signal is unknown, the false-alarm and detection probabilities can be calculated by eqs. (3.77) and (3.78), into which eq. (3.76) has been substituted, with $\mathcal{N}'$ replaced by $\mathcal{N}'' = G(\mathcal{N} + 1) - 1$ and N_s' replaced by $N_s'' = GN_s$. In Figs. 3.5 and 3.6 we plot the detection probability Q_d versus the average initial number N_s of signal photons, for various gains G , taking $\mathcal{N} = 0$ and $\mathcal{N} = 1$, respectively. The effect of amplification is the smaller, the larger the average initial number $\mathcal{N}$ of noise photons per mode.

This decrease of signal detectability due to amplification is to be expected. Detection theory prescribes the best method of processing the input $v(t)$ to the receiver; that is, it gives the strategy yielding minimum average probability of error or maximum detection probability with fixed false-alarm probability. In neither of the cases we have treated does this strategy include an amplification of the data. Hence amplification cannot yield a lower average error probability or higher detection probability, and as it is accompanied by the addition of noise to the data, amplification generally diminishes the detectability of a signal.

4. Modal Decomposition of Aperture Fields

We turn now to the detection of light beams, which we suppose to be falling normally on the aperture of an optical instrument and coming from sources spanning a narrow solid angle as seen by the observer. Because stops might be used to cut out all the incident background light outside a slender cone of directions normal to the aperture of the instrument, without reducing the detectability of the light from the source, all rays can be treated as paraxial, and the light fields can be accurately described by a scalar wave theory [Green and Wolf, 1953; Walther, 1967]. In thus dealing with a scalar field, our subsequent discussions assume that the incident light, both signal and background, is linearly polarized. To treat unpolarized light it is merely necessary to double the number of field modes considered when applying our results, and intermediate degrees of polarization can be similarly handled by breaking the incident light up into statistically independent, linearly polarized components.

Furthermore, the light to be detected occupies a spectral band whose width W is a small fraction of the central carrier frequency $\Omega = 2\pi c/\lambda$; $W \ll \Omega$. The background light is distributed broadly not only in direction, but also in frequency, possessing a Planck spectral density with absolute temperature $\mathcal{T}$. Components of the background light having frequencies far outside the band occupied by the light from the source can be filtered out without affecting the detectability of the signal.

The observer is to decide whether a certain coherent or incoherent light beam is present or not, and the only information on which he can base his decision resides in the scalar field $\psi(\mathbf{r}, t)$ at the aperture A of his optical

instrument during a finite interval $(0, T)$. He must so process that field that his decisions are made with the greatest effectiveness, and to this end he applies the principles of detection theory outlined in the preceding chapters.

The field at the aperture A is expanded in a set of orthonormal functions $\eta_p(\underline{r})$,

$$\int_A \eta_p(\underline{r}) \eta_q^*(\underline{r}) d^2\underline{r} = \delta_{pq} = \begin{cases} 1, & p = q, \\ 0, & p \neq q, \end{cases} \quad (4.1)$$

$d^2\underline{r}$ being the two-dimensional element of area in the aperture. We write the field there as

$$\psi(\underline{r}, t) = \sum_p \psi_p(t) \eta_p(\underline{r}), \quad (4.2)$$

where the component

$$\psi_p(t) = \int_A \eta_p^*(\underline{r}) \psi(\underline{r}, t) d^2\underline{r} \quad (4.3)$$

will be called the p -th aperture mode.

Because the background light is spatially homogeneous, falling on the aperture from a cone of directions much broader than that from which the light to be detected arrives, its spatial coherence function can be assumed to be proportional to the two-dimensional delta-function $\delta(\underline{r}_1 - \underline{r}_2)$, and the separate aperture modes $\psi_p(t)$ will be statistically independent in the absence of a signal field (hypothesis H_0). By appropriately choosing the spatial expansion functions $\eta_p(\underline{r})$ the aperture modes $\psi_p(t)$ can be made independent under hypothesis H_1 as well [Kuriksha, 1968]. This statistical independence requires the mode functions $\eta_p(\underline{r})$ to vary but little over the correlation distance $hc/K\mathcal{T}$ of the background field whenever any background light is present; the source must subtend a solid angle much smaller than $(K\mathcal{T}/h\Omega)^2$.

In our ideal receiver the light passing through the aperture is admitted to a lossless cavity during the observation interval $(0, T)$. We assume that

the aperture A is so much larger than a wavelength of light that all the light falling on the aperture enters the cavity and none is reflected or diffracted backward. Each aperture mode $\psi_p(t)$ can be thought of as exciting a distinct combination of normal modes within the cavity in much the same way as the input source $v(t)$ to the transmission-line receiver in ch. 3 excited the normal modes of the line, and we can regard each of these sets of cavity modes as associated with a separate and independent transmission line matched to the external field. The p -th equivalent transmission line, excited only by the p -th aperture mode, can be imagined as having connected to its input terminals the generator of a voltage

$$v_p(t) = Z_0^{\frac{1}{2}} \psi_p(t), \quad (4.4)$$

the internal impedance Z_s of this generator being equal to the characteristic impedance Z_0 of the line. The energy absorbed by the p -th line is, as in eq. (3.27), equal to

$$E_p = (4Z_0)^{-1} \int_0^T [v_p(t)]^2 dt = \frac{1}{4} \int_0^T [\psi_p(t)]^2 dt, \quad (4.5)$$

and the total energy absorbed by the cavity is

$$E_T = \frac{1}{4} \sum_p \int_0^T [\psi_p(t)]^2 dt = \frac{1}{4} \int_A \int_0^T [\psi(\underline{r}, t)]^2 d^2\underline{r} dt, \quad (4.6)$$

which specifies the normalization of the aperture field $\psi(\underline{r}, t)$.

In ch. 3 we supposed that the input is disconnected from the transmission line at time $t = T$ and that the modes of the line are observed immediately. Actually, their amplitudes could be measured at any subsequent time without changing the attainable false-alarm and detection probabilities, for they possess a known sinusoidal dependence on the time t , as in eq. (3.29). On a more fundamental level we recognize that the eigenvalues η_k of the operator $\rho_1 - \Lambda_0 \rho_0$ in eq. (2.6) are independent of the time of observation of the field

in the ideal receiver, and so are the false-alarm and detection probabilities, as given by eq. (2.13). A change in the time of observation merely imposes a unitary transformation on the detection operator Π or Π_T . It does not matter, therefore, at what time $t \geq T$ the cavity modes associated with the several aperture modes $\psi_p(t)$ are observed, once the aperture has been closed or--equivalently--the sources have been disconnected from the transmission lines.

Because the fields under consideration are quasimonochromatic, occupying a frequency band of width W about a carrier frequency $\Omega = 2\pi c/\lambda$, the aperture modes can be unambiguously divided into their positive- and negative-frequency components,

$$\psi_p(t) = \psi_p^{(+)}(t) + \psi_p^{(-)}(t). \quad (4.7)$$

The former is a carrier $e^{-i\Omega t}$ multiplied by a slowly varying modulation factor. Because $\psi_p(t)$ is real, the latter is the complex conjugate,

$$\psi_p^{(-)}(t) = [\psi_p^{(+)}(t)]^*, \quad (4.8)$$

and represents a corresponding modulation of $e^{i\Omega t}$.

Quantum-mechanically the positive- and negative-frequency components become adjoint noncommuting operators,

$$\psi_p^{(-)}(t) = [\psi_p^{(+)}(t)]^\dagger. \quad (4.9)$$

Their commutators follow from those in eq. (3.42), and when the time-dependences $e^{\pm i\Omega t}$ and the slowly varying nature of the modulations are taken into account, these can be written, by using eq. (4.4) in the conversion,

$$\begin{aligned} [\psi_p^{(+)}(t_1), \psi_q^{(-)}(t_2)] &= 2\hbar\Omega \delta(t_1 - t_2) \delta_{pq}, \\ [\psi_p^{(+)}(t_1), \psi_q^{(+)}(t_2)] &= 0. \end{aligned} \quad (4.10)$$

Distinct aperture modes commute by virtue of the orthogonality of the expansion functions $\eta_p(\mathbf{r})$.

The mode amplitudes g_m introduced in ch. 3 are seen from eq. (3.12) to be proportional to the coefficients of a Fourier series for the input $v(t)$ over the interval $(0, T)$. Because of the limitation of the input to frequencies in the neighborhood of Ω , only the positive-frequency part of $v(t)$ contributes to the Fourier integral for g_m . When we use eqs. (3.32) and (4.4) with $Z_0 = Z_S$ and $C\ell = T/Z_0$, we find that we can write the annihilation operators for mode p as

$$a_{pm} = (2K\Omega)^{-1/2} \int_0^T \gamma_m^*(t) \psi_p^{(+)}(t) e^{i\Omega t} dt, \quad (4.11)$$

$$\gamma_m(t) = T^{-1/2} \exp[-i(\omega_m - \Omega)t], \quad \omega_m = 2\pi m/T, \quad (4.12)$$

where we have replaced the factor $(\hbar\omega_m)^{-1/2}$ with $(\hbar\Omega)^{-1/2}$ because $\omega_m \approx \Omega$. The expansion functions $\gamma_m(t)$ are orthonormal over the observation interval $(0, T)$.

In treating the detection of incoherent light beams, however, we shall find it convenient to expand the aperture modes $\psi_p^{(+)}(t)$ in series of orthonormal functions $\gamma_m(t) e^{-i\Omega t}$ different from the sinusoidal ones in eq. (4.12). In general, then, we shall define the annihilation operators a_{pm} for the p -th aperture mode by eq. (4.11) with

$$\int_0^T \gamma_m^*(t) \gamma_n(t) dt = \delta_{mn}. \quad (4.13)$$

Which expansion functions are used will depend on the spectral density of the light to be detected, and as we shall assume spectral purity, the same set $\gamma_m(t)$ will apply to all the aperture modes $\psi_p(t)$. These annihilation operators a_{pm} and their adjoint creation operators a_{pm}^+ will obey the commutation rules

$$[a_{pm}, a_{qn}^+] = \delta_{pq} \delta_{mn}, \quad [a_{pm}, a_{qn}] = 0, \quad (4.14)$$

by virtue of the commutators given in eq. (4.10). These have also been derived from the commutation rules for a scalar field in free space [Helstrom, 1970b].

We are now writing the p -th aperture mode, as a function of time, in the form

$$\psi_p^{(+)}(t) = (2\kappa\Omega)^{1/2} \sum_{m=0}^{\infty} a_{pm} \gamma_m(t) e^{-i\Omega t}. \quad (4.15)$$

The total energy in the p -th mode is, by eqs. (4.5), (4.7), and (4.13),

$$E_p = \frac{1}{2} \int_0^T \psi_p^{(+)}(t) \psi_p^{(-)}(t) dt = \kappa\Omega \sum_{m=0}^{\infty} a_{pm}^+ a_{pm}, \quad (4.16)$$

and $a_{pm}^+ a_{pm}$ is the number operator for the m -th temporal mode of $\psi_p(t)$.

Once the spatial modes have been properly chosen, the results of our analysis in ch. 3 can be applied to the detection of coherent light beams. Suppose, for example, that the light to be detected is coming from an ideal laser whose emissions are pulses of known temporal form $s(t)$. Let the field created by the laser at point $\underline{r}$ of the aperture be proportional to $f(\underline{r})$, where

$$\int_A |f(\underline{r})|^2 d^2\underline{r} = 1.$$

We then take the first aperture mode function $\eta_1(\underline{r})$ equal to $f(\underline{r})$, and we pick the remaining functions $\eta_p(\underline{r})$ to be orthogonal to $f(\underline{r})$ over the aperture. Our results in ch. 3 determine the detectability of the laser pulse in the presence of spatially and temporally white thermal background light of absolute temperature $\mathcal{T}$. If the carrier phase of the laser pulse is known, the analysis in ch. 3 §3 applies; if it is unknown, that in §4 applies.

5. Detection of Incoherent Light

1. The Optical Fields

By incoherent light we mean light emitted by natural sources, which consist of enormous numbers of randomly emitting atoms and ions. Classically, the fields produced by such sources can be treated as spatio-temporal Gaussian random processes, for they are the sums of a great many small, random components produced by individual atomic radiators. Quantum-mechanically, the density operators of these fields have Gaussian P-representations, and the fields possess only first-order coherence [Glauber, 1963].

When the light to be detected is present (hypothesis H_1), the positive-frequency half $\psi^{(+)}(\underline{r}, t)$ of the field at the aperture of the observing instrument consists of two parts,

$$\psi^{(+)}(\underline{r}, t) = \psi_s(\underline{r}, t) + \psi_n(\underline{r}, t),$$

$\psi_s(\underline{r}, t)$ representing the beam to be detected (the signal), and $\psi_n(\underline{r}, t)$ representing the background light (the noise). (We omit the superscript $(+)$ indicating that ψ_s and ψ_n contain only positive frequencies.) The background light arrives with a distribution in frequency and angle so much broader than that of the signal that the background can be taken as spatially and temporally white; its mutual coherence function has the form

$$\frac{1}{2} E[\psi_n^+(\underline{r}_1, t_1) \psi_n(\underline{r}_2, t_2)] = N_0 \delta(\underline{r}_1 - \underline{r}_2) \delta(t_1 - t_2), \quad (5.1)$$

where the superscript $+$ indicates classically the complex-conjugate, quantum-mechanically the adjoint.

The signal field we assume to be spectrally pure, so that its mutual

coherence function can be factored into a spatial part $\varphi_S(\mathbf{r}_1, \mathbf{r}_2)$ and a temporal part $\chi(t_1 - t_2)$; we write it in the form

$$\frac{1}{2} E[\psi_S^+(\mathbf{r}_1, t_1) \psi_S(\mathbf{r}_2, t_2)] = \varphi_S(\mathbf{r}_1, \mathbf{r}_2) \chi(t_1 - t_2) \exp[-i\Omega(t_1 - t_2)], \quad (5.2)$$

where $\Omega = 2\pi c/\lambda$ is the angular frequency of the carrier. The temporal coherence function $\chi(\tau)$ is so normalized that $\chi(0) = 1$; its Fourier transform

$$X(\omega) = \int_{-\infty}^{\infty} \chi(\tau) e^{i\omega\tau} d\tau \quad (5.3)$$

represents the spectral density of the object light, with angular frequencies ω referred to Ω as origin. The bandwidth W of this light is conveniently defined as

$$W = \left[\int X(\omega) d\omega / 2\pi \right]^2 / \int_{-\infty}^{\infty} [X(\omega)]^2 d\omega / 2\pi = |\chi(0)|^2 / \int_{-\infty}^{\infty} |\chi(\tau)|^2 d\tau; \quad (5.4)$$

W is the reciprocal of the coherence time defined by Mandel [1958]. The signal and noise processes are both circular Gaussian in the sense that $E[\psi^{(+)}(\mathbf{r}_1, t_1) \times \psi^{(+)}(\mathbf{r}_2, t_2)] = 0$ [Helstrom, 1968c, pp. 69-72].

The signal is considered as arising from a distant luminous source, which can be taken as planar and as having a radiance $B(\mathbf{u})$ as a function of position $\mathbf{u} = (u_x, u_y)$. The spatial part of the mutual coherence function at the aperture is then

$$\varphi_S(\mathbf{r}_1, \mathbf{r}_2) = (4\pi R^2)^{-1} \exp\left[\frac{ik}{2R} (\mathbf{r}_1^2 - \mathbf{r}_2^2)\right] \beta(\mathbf{r}_1 - \mathbf{r}_2), \quad (5.5)$$

where $k = \Omega/c = 2\pi/\lambda$ is the propagation constant, R is the distance between object and aperture, and

$$\beta(\mathbf{r}) = \int_0 B(\mathbf{u}) \exp(ik\mathbf{u} \cdot \mathbf{r}/R) d^2\mathbf{u}, \quad (5.6)$$

0 designating an integration over the source or object plane. Equation (5.5) follows directly from the Fresnel-Kirchhoff diffraction formula as applied to the object plane, at which the mutual coherence function of the signal has the

form

$$\frac{1}{2} E[\psi_s^+(u_1, t_1) \psi_s(u_2, t_2)] = \pi k^{-2} B(u_1) \delta(u_1 - u_2) \chi(t_1 - t_2) \exp[-i\Omega(t_1 - t_2)], \quad (5.7)$$

the delta-function indicating the spatial incoherence of the light upon emission by the object plane. The average total energy received from the object is, by eqs. (4.16), (5.2), (5.5), and (5.6),

$$E_S = \int_A \int_0^T \varphi_S(\underline{r}, \underline{r}) \chi(0) d^2 \underline{r} dt = B_T A T / 4 \pi R^2, \\ B_T = \beta(0) = \int_0 B(\underline{y}) d^2 \underline{y}. \quad (5.8)$$

The most useful decomposition of the aperture field into modes as in eq. (4.2) now involves as expansion functions $\eta_p(\underline{r})$ the eigenfunctions of the spatial part $\varphi_S(\underline{r}_1, \underline{r}_2)$ of the mutual coherence function of the object light; they are solutions of the integral equation

$$h_p \eta_p(\underline{r}_1) = (A B_T)^{-1} \int_A \beta(\underline{r}_2 - \underline{r}_1) \eta_p(\underline{r}_2) d^2 \underline{r}_2 \quad (5.9)$$

with eigenvalues h_p that sum to 1,

$$\sum_p h_p = 1, \quad (5.10)$$

and indicate the fraction of the total signal energy going into each aperture mode $\psi_p(t)$. By using these expansion functions, we ensure that the aperture modes $\psi_p(t)$ are statistically independent both in the presence and in the absence of the object light.

For the temporal decomposition of each aperture mode $\psi_p(t)$ we use the eigenfunctions $\gamma_m(t)$ of the temporal part of the mutual coherence function of the object light; these are defined by the integral equation

$$\chi_m \gamma_m(t_1) = T^{-1} \int_0^T \chi(t_1 - t_2) \gamma_m(t_2) dt_2. \quad (5.11)$$

The eigenvalues χ_m sum to 1,

$$\sum \chi_m = 1, \quad (5.12)$$

and specify the fractions of the signal energy in each aperture mode $\psi_p(t)$ that are associated with the terms $\gamma_m(t) e^{-i\Omega t}$ in eq. (4.15). When, as is often the case, the product WT of the signal bandwidth and the observation time is very large, the eigenvalues χ_m are given approximately by

$$\chi_m \cong T^{-1} X(2\pi m/T), \quad (5.13)$$

where $X(\omega)$ is the spectral density of the object light as defined in eq. (5.3), and m takes on both positive and negative integral values. The associated eigenfunctions are approximately the complex exponentials given by eq. (4.12), as can be seen by substitution into the integral equation (5.11) [Grenander and Szegő, 1958, §8.6, pp. 136-139].

We define the combination $\eta_p(\underline{r}) \gamma_m(t) e^{-i\Omega t}$ as a spatio-temporal mode of the aperture field. It possesses on the average a fraction $h_p \chi_m$ of the energy received from the object. All the spatio-temporal modes are statistically independent under both hypotheses H_0 ("signal absent") and H_1 ("signal present").

From eqs. (4.11), (5.2), and (5.3) we can calculate the average number of photons in the spatio-temporal mode (pm) under hypothesis H_1 ,

$$E(a_{pm}^+ a_{pm} | H_1) = \text{Tr}(\rho_1 a_{pm}^+ a_{pm}) = (\hbar\Omega)^{-1} (h_p \chi_m E_s + N_0) = N_{pm} + \mathcal{N}, \quad (5.14)$$

where

$$N_{pm} = h_p \chi_m E_s / \hbar\Omega = h_p \chi_m N_s \quad (5.15)$$

is the average number of signal photons in the mode and $\mathcal{N} = N_0 / \hbar\Omega$ is the average number of noise photons; $N_s = E_s / \hbar\Omega$ is the total average number of

photons received at aperture A from the object during the observation interval $(0, T)$. The number $\mathcal{N} = \text{pl}(\hbar\Omega/K\mathcal{T})$ is given by the Planck law, eq. (3.50), in terms of Boltzmann's constant K and the effective absolute temperature $\mathcal{T}$ of the background light; it is the same for all the significant spatio-temporal modes, for we again assume that $\hbar\omega \ll K\mathcal{T}$ whenever $\mathcal{N} > 0$. Under hypothesis H_0 ,

$$E(a_{\text{pm}}^+ a_{\text{pm}} | H_0) = \text{Tr}(\rho_0 a_{\text{pm}}^+ a_{\text{pm}}) = \mathcal{N}. \quad (5.16)$$

The number of photons in each mode has a Bose distribution under both hypotheses H_0 and H_1 ; the mean numbers given by eqs. (5.14) and (5.16) suffice to determine these distributions.

2. The Optimum Receiver

Because we have arranged for statistically independent modes under both hypotheses, both the density operators ρ_0 and ρ_1 for the field in the ideal receiver factor into products of density operators for the individual spatio-temporal modes. For each mode the density operator has, under both H_0 and H_1 , a Gaussian P-representation with mean zero, like that in eq. (3.48), in which $\mathcal{N}' = \mathcal{N}$ under hypothesis H_0 and $\mathcal{N}' = N_{pm} + \mathcal{N}$ under H_1 . Both these modal density operators are therefore diagonal in the number representation. Let $|v_{pm}\rangle$ be the eigenstate of the number operator $a_{pm}^\dagger a_{pm}$ with eigenvalue v_{pm} ,

$$a_{pm}^\dagger a_{pm} |v_{pm}\rangle = v_{pm} |v_{pm}\rangle. \quad (5.17)$$

Then the density operators for the entire field are

$$\rho_i = \prod_{p,m} P_{pm}^{(i)}(v_{pm}) |v_{pm}\rangle \langle v_{pm}|, \quad (5.18)$$

$$P_{pm}^{(i)}(v) = (1 - v_{pm}^{(i)}) v_{pm}^v, \quad v_{pm}^{(i)} = N_{pm}^{(i)} / (N_{pm}^{(i)} + 1), \quad i = 0, 1, \\ N_{pm}^{(0)} = \mathcal{N}, \quad N_{pm}^{(1)} = N_{pm} + \mathcal{N} = h_p \chi_m N_s + \mathcal{N}. \quad (5.19)$$

Because the density operators ρ_0 and ρ_1 are simultaneously diagonal in the number representation, the optimum receiver is one that measures the numbers v_{pm} of photons in each spatio-temporal mode, forms the likelihood ratio

$$\Lambda(\{v_{pm}\}) = \prod_{p,m} P_{pm}^{(1)}(v_{pm}) / P_{pm}^{(0)}(v_{pm}), \quad (5.20)$$

and compares it with an appropriate decision level Λ_0 , deciding that the signal field is present if $\Lambda > \Lambda_0$. Equivalently, it compares the logarithmic likelihood ratio

$$\ln \Lambda(\{v_{pm}\}) = \sum_{p,m} \{v_{pm} \ln(v_{pm}^{(1)}/v_{pm}^{(0)}) + \ln[(1 - v_{pm}^{(1)})/(1 - v_{pm}^{(0)})]\} \quad (5.21)$$

with $\ln \Lambda_0$ or with a decision level determined by the false-alarm probability. In the subsequent sections we shall for various special cases study this optimum receiver or its approximating threshold receiver and evaluate its performance.

3. Detection of Point Sources

When the light to be detected comes from a point source such as a distant star, its field possesses complete first-order spatial coherence over the aperture, $\beta(\underline{r}) \equiv \beta(0)$, $\underline{r} \in A$. The lowest spatial eigenfunction is then constant,

$$\eta_1(\underline{r}) \equiv A^{-1/2}, \quad \underline{r} \in A, \quad h_1 = 1, \quad (5.22)$$

where A stands for the area of the aperture. The remaining eigenfunctions are orthogonal to $\eta_1(\underline{r})$ over the aperture. Only the aperture mode $\psi_1(t)$ is excited by the light from the object, and the others can be disregarded. We can therefore drop the spatial-mode subscript p in the preceding equations, putting $h_1 = 1$ and restricting our attention to the temporal modes $\gamma_m(t) e^{-i\Omega t}$ into which $\psi_1(t)$ is decomposed.

3.1 Extreme Quantum Limit

In the extreme quantum limit $\mathcal{N} = 0$ ($\mathcal{T} = 0$) no photons are counted in any of the temporal modes in the absence of a signal (hypothesis H_0). It is therefore necessary to use a randomized strategy, as discussed in ch. 1 §2, in order to attain a pre-assigned false-alarm probability Q_0 . The receiver will choose hypothesis H_1 whenever any photons at all are counted; when no photons are observed, it chooses hypothesis H_1 with probability Q_0 . The probability of detection is now

$$Q_d = 1 - p_1(0) + Q_0 p_1(0) = 1 - (1 - Q_0)p_1(0), \quad (5.23)$$

where $p_1(0)$ is the probability that no photons are found in any of the modes under hypothesis H_1 , that is, by eq. (5.19),

$$p_1(0) = \prod_m (1 - v_m^{(1)}) = \prod_m (1 + \chi_m N_s)^{-1} = [f(N_s)]^{-1}, \quad (5.24)$$

where χ_m is an eigenvalue of the integral equation (5.11). Our previous assumption that $hW \ll K\mathcal{T}$ is unnecessary here, the number $\mathcal{N}$ being zero in all modes.

The function

$$f(z) = \prod_m (1 + \chi_m z) \quad (5.25)$$

is the Fredholm determinant of the integral equation (5.11). For a source with a Lorentz spectrum

$$X(\omega) = 2W(\omega^2 + W^2)^{-1}, \quad \chi(\tau) = e^{-W|\tau|}, \quad (5.26)$$

the Fredholm determinant is well known,

$$f(z) = (4\mu b)^{-1} e^{-\mu} [(b + \mu)^2 e^b - (b - \mu)^2 e^{-b}],$$

$$\mu = WT, \quad b = (\mu^2 + 2\mu z)^{\frac{1}{2}} \quad (5.27)$$

[Siegert, 1957]. In Fig. 5.1 we have plotted the probability Q_d of detection for $Q_0 = 10^{-4}$ and various values of the time-bandwidth product $\mu = WT$. For $\mu = 0$, $p_1(0) = (1 + N_s)^{-1}$, and in the limit $\mu \gg 1$, $p_1(0) = \exp(-N_s)$. These curves would be altered only imperceptibly for smaller false-alarm probabilities.

If the incident light is unpolarized, it can be broken up into two statistically independent linearly polarized modes with an average number $\frac{1}{2} N_s$ of signal photons in each. The detection probability is now

$$Q_d = 1 - (1 - Q_0) [f(\frac{1}{2} N_s)]^2 \quad (5.28)$$

and is plotted in Fig. 5.2 for various values of WT , again with $Q_0 = 10^{-4}$.

Whether polarized or unpolarized, the light is the more easily detected, the greater its bandwidth.

The slow descent of the Lorentz spectrum to zero with increasing ω is in strong contrast to the sharp cutoff of the rectangular spectrum,

$$\begin{aligned}
X(\omega) &= W^{-1}, & -\pi W < \omega < \pi W, \\
X(\omega) &= 0, & |\omega| > \pi W,
\end{aligned}
\tag{5.29}$$

for which the temporal coherence function is

$$\chi(\tau) = \text{sinc } W\tau, \tag{5.30}$$

with $\text{sinc } x = \sin(\pi x)/(\pi x)$. The integral equation (5.11) that results has been solved by Slepian and Pollak [1961], and extensive tables of eigenvalues have been provided by Slepian and Sonnenblick [1965]. In their notation,

$$\chi_m = \lambda_m(c)/WT = \pi\lambda_m(c)/2c, \quad c = \pi WT/2; \tag{5.31}$$

the eigenfunctions $\gamma_m(t)$ are prolate spheroidal wave functions. For $WT \gg 1$ there are approximately WT eigenvalues χ_m equal to $(WT)^{-1}$, and the rest are exponentially small; compare eq. (5.13). In order to show the effect of the spectral shape on the detectability of incoherent light at the extreme quantum limit, we have plotted as dashed lines in Figs. 5.1 and 5.2 the detection probability calculated from eqs. (5.24) and (5.28) by using these eigenvalues.

3.2 Intermediate Range

The logarithmic likelihood ratio of eq. (5.21) cannot be converted into a statistic having only integral values when $\mathcal{N} \neq 0$, for the factors $\ln(v_{\text{pm}}^{(1)}/v_{\text{pm}}^{(0)})$ are not in general commensurable. The distributions of $\ln \Lambda(\{v_{\text{pm}}\})$ under the two hypotheses are then very difficult to calculate.

When the spectral density of the signal is rectangular, and the time-bandwidth product WT is large, however, there are approximately WT temporal eigenvalues $\chi_m = (WT)^{-1}$, and the rest are negligible. If we make this approximation, we find that the optimum receiver of the light forms the sum

$$v = \sum_{m=1}^{\mu} v_m, \quad \mu = WT, \quad (5.32)$$

of the numbers v_m of photons in the μ significant temporal modes. The sum v differs now from the logarithmic likelihood ratio $\ln \Lambda(\{v_m\})$ only by an additive constant and a constant factor, which can be absorbed into the decision level. Since v is an integer, a randomized decision rule must be applied as described in ch. 1 §2, and the false-alarm and detection probabilities are calculated by eqs. (1.12) and (1.13). The probability of obtaining a total of v photons is now the generalized Bose-Einstein distribution

$$p_i(v) = \mu!^{-1} (v + \mu - 1)! (1 - v_i)^{\mu} v_i^v, \quad i = 0, 1, \quad (5.33)$$

$$v_0 = \mathcal{N}/(\mathcal{N} + 1), \quad v_1 = (\mathcal{N} + \mu^{-1}N_s)/(\mathcal{N} + \mu^{-1}N_s + 1). \quad (5.34)$$

The mean total numbers of photons are obtained from eq. (5.13),

$$\bar{v}_0 = \mathbb{E}(v|H_0) = \mu\mathcal{N}, \quad \bar{v}_1 = \mathbb{E}(v|H_1) = \mu\mathcal{N} + N_s. \quad (5.35)$$

The probability of detection has been plotted in Fig. 5.3 for $\mathcal{N} = 1$, $Q_0 = 10^{-4}$, and various values of the time-bandwidth product $\mu = WT$. The curve for $\mu = 1$ gives the detection probability for a single temporal mode when $\mathcal{N} = 1$ and can be calculated by eq. (1.15). For $\mu = 2$ and 5 the curves represent only crude approximations.

3.3 The Poisson Limit

If the time-bandwidth product WT increases without limit and at the same time the mean number $\mathcal{N}$ of noise photons per mode decreases in such a way that $E(v|H_0) = \mathcal{N}WT$ remains fixed, the distributions of the total number v of photons counted become in the limit $\mu = \infty$ Poisson distributions, independently of the spectral density of the light from the point source,

$$p_i(v) = \bar{v}_i^v \exp(-\bar{v}_i) / v_i!, \quad i = 0, 1. \quad (5.36)$$

A detector based on observations of the number v is optimum, however, only when the spectral density of the signal is rectangular, as in eq. (5.29).

In Fig. 5.4 we have plotted the probability Q_d of detection in this limit $WT \gg 1$ versus the average number N_s of signal photons for various values of $\mathcal{N}WT \leq 1$. For $\mathcal{N}WT \geq 1$ the detection probability is plotted in Fig. 5.5 versus the signal-to-noise ratio

$$D_n = N_s (\mathcal{N}WT)^{-1/2}. \quad (5.37)$$

For $\mathcal{N}WT \gg 1$ the distribution of the sum v is nearly Gaussian, and the false-alarm and detection probabilities are approximately as in eq. (3.65) with D_n in place of D_q .

3.4 The Effect of Amplification

When the spectral density of the signal is rectangular, the $\mu = WT$ temporal modes correspond to the sinusoidal eigenfunctions $g_m(t) e^{-i\Omega t}$ of eq. (4.12) and can be associated directly with the modes of the transmission-line receiver as treated in ch. 3. We can then envision amplifying the mixture of incoherent signal and background as described in ch. 3 §5. Under hypothesis H_1 the mean number of photons in each of the significant temporal modes before amplification is

$$\bar{v}_{m1} = \mathcal{N} + N_s/\mu; \quad (5.38)$$

after amplification with gain G it is under the best of circumstances given by eq. (3.90) as

$$\bar{v}'_{m1} = G(\bar{v}_{m1} + 1) - 1 = G(\mathcal{N} + \mu^{-1}N_s + 1) - 1, \quad (5.39)$$

and the parameter v_1 in the distribution in eq. (5.34) is

$$v_1 = \bar{v}'_{m1} / (\bar{v}'_{m1} + 1), \quad (5.40)$$

with v_0 obtained by setting $N_s = 0$. In Figs. 5.6 and 5.7 we have plotted the probability of detection calculated from the distributions of eq. (5.33) with these values of v_0 and v_1 , for $WT = 5$ and 10 and for $\mathcal{N} = 0$ and various amplification factors G . A randomized decision procedure leading to an exact false-alarm probability $Q_0 = 10^{-4}$ was postulated. Again the detectability of the signal decreases with increasing amplification. For larger values of $\mathcal{N}$ the effect of amplification is less severe.

3.5 The Classical Limit

For the sake of completeness we analyze the detection of light from an incoherent point source in the classical limit when the average number of photons in each temporal mode is very large under both hypotheses. The problem is then the same as that of detecting a narrowband Gaussian stochastic signal having energy $E_s = \hbar\Omega N_s$ and a normalized temporal mutual coherence function $\chi(\tau) e^{-i\Omega\tau}$ in the presence of white Gaussian noise of unilateral spectral density $K\mathcal{T} = \hbar\Omega\mathcal{N}$. Our results will also apply to the detection of an incoherently emitting point source after extreme amplification, $G \gg 1$, the number of noise and signal photons in each temporal mode $\gamma_m(t) e^{-i\Omega t}$ being assumed given by an expression like eq. (3.90) after amplification; it is only necessary in what follows to replace $\mathcal{N}$ by $\mathcal{N} + 1$.

In the limit $\mathcal{N} \gg 1$, $N_s \gg 1$, the logarithmic likelihood ratio in eq. (5.21) becomes, with the spatial-mode subscript p dropped,

$$\begin{aligned} \ln \Lambda(\{v_m\}) &= \sum_m [v_m(N_m^{(0)})^{-1} - N_m^{(1)-1} - \ln(N_m^{(1)}/N_m^{(0)})] \\ &= \sum_m [D^2 \chi_m (1 + D^2 \chi_m)^{-1} x_m - \ln(1 + D^2 \chi_m)] \end{aligned} \quad (5.41)$$

where $D^2 = N_s/\mathcal{N} = E_s/K\mathcal{T}$ is the signal-to-noise ratio, χ_m is the temporal eigenvalue defined in eq. (5.11), and $x_m = v_m/\mathcal{N}$. Here we have used eq. (5.19). In this classical limit $x_m = v_m/\mathcal{N}$ is a continuous random variable with exponential p.d.f.'s under both hypotheses,

$$\begin{aligned} p_{0m} &= \exp(-x_m), \\ p_{1m} &= (1 + D^2 \chi_m)^{-1} \exp[-(1 + D^2 \chi_m)^{-1} x_m], \quad x_m > 0, \end{aligned} \quad (5.42)$$

and the x_m 's for different modes are statistically independent.

In order to calculate the false-alarm and detection probabilities we first find the moment-generating functions (m.g.f.'s) of the detection statistic

$$L = \sum_m D^2 \chi_m (1 + D^2 \chi_m)^{-1} \chi_m, \quad (5.43)$$

which differs only by a constant from the logarithmic likelihood ratio in eq.

(5.41). Hypothesis H_1 is chosen when L exceeds a decision level L_0 . The m.g.f.'s are the Laplace transforms of the p.d.f.'s of L ,

$$\mathbb{E}[e^{-sL} | H_0] = f(D^2) / f(D^2(1 + s)), \quad (5.44)$$

$$\mathbb{E}[e^{-sL} | H_1] = [f(D^2 s)]^{-1}, \quad (5.45)$$

where $f(z)$ is the Fredholm determinant defined in eq. (5.25). The p.d.f.'s of L are found by taking the inverse Laplace transforms of eqs. (5.44) and (5.45) by means of the residue theorem, and when these are integrated over $L_0 < L < \infty$, the false-alarm and detection probabilities are found to be

$$\begin{aligned} Q_0 &= f(D^2) \sum_{m=1}^{\infty} [f'(-\chi_m^{-1}) (D^2 + \chi_m^{-1})]^{-1} \exp[-(1 + D^{-2} \chi_m^{-1}) L_0], \\ Q_d &= \sum_{m=1}^{\infty} \chi_m [f'(-\chi_m^{-1})]^{-1} \exp(-D^{-2} \chi_m^{-1} L_0), \end{aligned} \quad (5.46)$$

where $f'(z) = df/dz$.

For object light with a Lorentz spectral density defined by eq. (5.26) the Fredholm determinant $f(z)$ was given in eq. (5.27); its zeros χ_m^{-1} are obtained by solving a certain transcendental equation [Helstrom, 1968c, p. 140]. In this way the detection probability Q_d was calculated for $Q_0 = 10^{-4}$ and various values of the time-bandwidth product $\mu = WT$; the results are plotted in Fig. 5.8. The statistic L does not yield a uniformly most powerful test, depending as it does on the signal-to-noise ratio D^2 , and hence each point on a curve in Fig. 5.8 refers to a different detector, specifically the optimum one for that signal-to-noise ratio.

When $WT \gg 1$, the eigenvalues χ_m of the temporal coherence function $\chi(\tau)$ are approximately given by eq. (5.13), and the Fredholm determinant in eq. (5.25) can be written approximately as

$$\begin{aligned} f(z) &= \exp \left[\sum_m \ln(1 + \chi_m z) \right] \\ &\cong \exp \left\{ T \int_{-\infty}^{\infty} \ln[1 + zX(\omega)] d\omega/2\pi \right\}. \end{aligned} \quad (5.47)$$

For the Lorentz spectral density this yields

$$f(z) \cong e^{\mu-b}, \quad b = \sqrt{\mu^2 + 2\mu z}, \quad \mu = WT, \quad (5.48)$$

and by substituting into eqs. (5.44) and (5.45) and taking the inverse Laplace transformation we find the approximate false-alarm and detection probabilities for the optimum classical receiver,

$$\begin{aligned} Q_0 &\cong \operatorname{erfc} [(M/\xi)^{1/2}(\xi - 1)] - e^{2M} \operatorname{erfc} [(M/\xi)^{1/2}(\xi + 1)], \\ M &= (\mu^2 + 2D^2\mu)^{1/2}, \quad \mu \gg 1, \end{aligned} \quad (5.49)$$

$$\begin{aligned} Q_d &\cong \operatorname{erfc} [(\mu/\xi')^{1/2}(\xi' - 1)] - e^{2\mu} \operatorname{erfc} [(\mu/\xi')^{1/2}(\xi' + 1)], \\ \xi' &= \mu\xi/M, \quad \mu \gg 1, \end{aligned} \quad (5.50)$$

where $\operatorname{erfc} \xi$ is the error-function integral defined in eq. (3.23). For each value of D^2 , ξ is determined from eq. (5.49) to obtain the pre-assigned false-alarm probability Q_0 , whereupon Q_d is calculated from eq. (5.50). Good agreement was found between the residue series and these approximations for $\mu = WT \gtrsim 12$.

In Fig. 5.9 the probability of detecting light with a rectangular spectral density--eq. (5.29)-- is presented for various values of the time-bandwidth product WT . The eigenvalues χ_m are given by eq. (5.31), and the Fredholm determinant is used in the form given in eq. (5.25). Only a finite number--of the order of WT --of terms need to be carried because the eigenvalues χ_m

decrease rapidly once the index m exceeds WT . For $WT \geq 7$ the approximation obtained from eq. (5.47) has been used; it leads to a gamma-distribution for the test statistic under each hypothesis. The false-alarm and detection probabilities were then calculated from

$$Q_0 = \Gamma(x, WT),$$

$$Q_d = \Gamma(x(1 + D^2/WT), WT),$$

where

$$\Gamma(x, v) = [\Gamma(v)]^{-1} \int_x^\infty t^{v-1} e^{-t} dt$$

is related to the incomplete gamma-function.

3.6 The Threshold Detector

The structure of the optimum detector of incoherent light in the presence of a thermal background field depends on the strength of the signal to be detected; it does not provide a uniformly most powerful test of the hypothesis "signal present" against the null hypothesis "signal absent". If the signal strength is unknown in advance, therefore, the optimum detector derived in the foregoing sections cannot be applied. As signal detectability is poorest for weak signals, it is natural to use instead the detector that is best in the limit of vanishing signal strength; this is the threshold detector described in ch. 1 §4 and ch. 2 §3. Since the density operators for the field in the receiver under the two hypotheses commute, it suffices here to use the classical form of the threshold detector.

The logarithmic likelihood ratio in eq. (5.21) is easily expanded in powers of the average number N_s of signal photons, and the term of first order in N_s yields, after the part independent of the data is dropped and the remainder is divided by $N_s/\mathcal{N}(\mathcal{N}+1)$,

$$U = \sum_p \sum_m h_p x_m v_{pm}, \quad (5.51)$$

which is the threshold statistic. It is a simpler linear combination of the data v_{pm} than the optimum statistic in eq. (5.21) and has the advantage of being independent of the strength N_s of the light to be detected.

The light signal we are considering here comes from a point source and possesses first-order spatial coherence over the aperture. As in the previous parts of this section, we deal only with a single spatial mode and drop the subscript p , writing the threshold statistic in the form

$$U = \sum_m \chi_m v_m, \quad (5.52)$$

where v_m is the number of photons counted in the m -th temporal mode $\gamma_m(t) e^{-i\Omega t}$.

When the spectral density of the object light is rectangular with bandwidth W , there are--as we have said--approximately WT equal eigenvalues $\chi_m = (WT)^{-1}$, and the rest are negligibly small, provided $WT \gg 1$. The threshold detector in this case performs the same operation on the data v_m as the optimum detector, summing those numbers of photons in the WT significant temporal modes and comparing the sum with an appropriate decision level. The detection probabilities calculated in parts (ii) and (iii) and plotted in Figs. 5.3, 5.4, and 5.5 apply to the threshold detector as well, and for light with a rectangular spectrum the threshold and optimum detectors exhibit nearly the same performance.

For signal spectra other than the rectangular, the threshold statistic U becomes a continuous, rather than a discrete random variable. Its p.d.f.'s under the two hypotheses must be obtained from its moment generating functions (m.g.f.'s).

$$\begin{aligned} f_1(s) &= \mathbb{E}(e^{-Us} | H_1, N_s) = \prod_m \mathbb{E}[\exp(-\chi_m s v_m) | H_1, N_s] \\ &= \prod_m \{1 + (\mathcal{N} + \chi_m N_s) [1 - \exp(-\chi_m s)]\}^{-1}, \end{aligned} \quad (5.54)$$

$$f_0(s) = \mathbb{E}(e^{-Us} | H_0) = \prod_m \{1 + \mathcal{N}[1 - \exp(-\chi_m s)]\}^{-1}, \quad (5.55)$$

which follow from the Bose distributions in eq. (5.19). An inverse Laplace transform is required, and numerical methods must ordinarily be used.

For most observations of natural sources of incoherent light the time-bandwidth product WT is very large. The eigenvalues are then given approximately by eq. (5.13). If we assume $\mathcal{N} \ll 1$, with the product $\mathcal{N}WT$ of the order

of 1, the m.g.f.'s in eqs. (5.54) and (5.55) become approximately

$$\begin{aligned} f_1(s) &= \exp \left\{ -T \int_{-\infty}^{\infty} [\mathcal{N} + N_s T^{-1} X(\omega)] \{1 - \exp[-X(\omega)s/T]\} d\omega/2\pi \right\} \\ f_0(s) &= \exp \left\{ -\mathcal{N}T \int_{-\infty}^{\infty} \{1 - \exp[-X(\omega)s/T]\} d\omega/2\pi \right\}, \end{aligned} \quad (5.56)$$

the sum over modes involved in $\ln f_1(s)$ and $\ln f_0(s)$ having been converted to an integration. For a uniform spectral density these lead to the Poisson distributions found in §3.3.

When the object light has a Lorentz spectral density as given by eq. (5.26), the m.g.f. of the statistic U under hypothesis H_1 becomes

$$\begin{aligned} f_1(s) &= \exp \{ -\mathcal{N}WT \alpha e^{-\alpha} [I_0(\alpha) + I_1(\alpha)] - N_s [1 - e^{-\alpha} I_0(\alpha)] \}, \\ \alpha &= s/WT, \end{aligned} \quad (5.57)$$

with $f_0(s)$ obtained by setting $N_s = 0$; $I_0(\alpha)$ and $I_1(\alpha)$ are modified Bessel functions. It appears impossible to invert these m.g.f.'s analytically in order to obtain the p.d.f.'s of the threshold statistic U , from which the false-alarm and detection probabilities can be calculated; numerical methods described elsewhere are required [Helstrom, 1969c]. It was found that the detection probability is slightly smaller for a Lorentz spectrum than for a rectangular spectrum with the same value of WT .

4. Detection of Extended Objects

When the object is not a point source, but extends over more than one resolution element of the instrument, its light arriving at the aperture does not possess full first-order spatial coherence, and more than one aperture mode $\psi_p(t)$ must be processed for detection. A resolution element in the object plane at distance R has an area of the order of $(\lambda R)^2/A$, where $\lambda = 2\pi c/\Omega$ is the wavelength of the object light and A is the area of the aperture of the observing instrument.

It is now necessary to solve the integral equation (5.9), and this can be done analytically for only a few forms of the radiance distribution $B(\underline{u})$ of the object. Some general statements can be made about the spectrum of eigenvalues h_p , however, if the aperture is rectangular ($a_x \times a_y$) and if $B(\underline{u})$ changes only slightly over a single resolution element of area $(\lambda R)^2/a_x a_y$. An approximation for the spatial eigenvalues h_p can be found that is much like the one in eq. (5.13) for the temporal eigenvalues. It links them with samples of the Fourier transform of the kernel $\beta(\underline{r})$ of eq. (5.9), that is, to samples of the radiance $B(\underline{u})$ at uniformly spaced points in the object plane,

$$h_{\underline{p}} \cong \delta_x \delta_y B(p_x \delta_x, p_y \delta_y) / B_T, \quad (5.58)$$

$$\underline{p} = (p_x, p_y), \quad \delta_x = \lambda R / a_x, \quad \delta_y = \lambda R / a_y,$$

where p_x and p_y are integers, and δ_x and δ_y are the linear resolution elements in the x - and y -directions in the object plane. The spatial mode index p has for convenience been made into a 2-vector $\underline{p}$.

The approximate number of significant eigenvalues $h_{\underline{p}}$ is given by eq. (5.58) as

$$M = AA_0 / (\lambda R)^2, \quad (5.59)$$

where $A = a_x a_y$ is the area of the aperture and A_0 is the area of the planar object at distance R . More generally we can derive eq. (5.59) by observing that by eq. (5.9) M is given approximately by

$$\begin{aligned} M &\cong \left(\sum_{\tilde{p}} h_{\tilde{p}} \right)^2 / \sum_{\tilde{p}} h_{\tilde{p}}^2 = A^2 |\beta(0)|^2 \left[\iint_{A \times A} |\beta(\underline{r}_1 - \underline{r}_2)|^2 d^2 \underline{r}_1 d^2 \underline{r}_2 \right]^{-1} \\ &\cong A |\beta(0)|^2 \left[\int_A |\beta(\underline{r})|^2 d^2 \underline{r} \right]^{-1} = A B_T^2 (\lambda R)^{-2} \left[\int_0 [B(\underline{u})]^2 d^2 \underline{u} \right]^{-1} \\ &= AA_0 / (\lambda R)^2, \end{aligned} \quad (5.60)$$

if we define the effective area of the object as

$$A_0 = B_T^2 \left[\int_0 [B(\underline{u})]^2 d^2 \underline{u} \right]^{-1}, \quad B_T = \int_0 B(\underline{u}) d^2 \underline{u}. \quad (5.61)$$

For a uniformly radiating object A_0 is its geometrical area.

Implicit in our derivation is that $M \gg 1$, so that the kernel $\beta(\underline{r})$ covers a much smaller area than the aperture A . The number M given by eq. (5.59) has been defined by Gabor [1961] as the number of spatial degrees of freedom in the object. If the object has a uniform radiance over the area A_0 , the M significant eigenvalues can be taken as equal, $h_{\tilde{p}} \cong M^{-1}$, when $M \gg 1$, and the rest can be set equal to 0.

When the object is a uniformly radiating circular disk of radius b and area $A_0 = \pi b^2$, and the aperture is circular with radius a , the integral equation (5.9) reduces to one treated by Slepian [1964]. The spatial mode functions $\eta_p(\underline{r})$ are proportional to the generalized prolate spheroidal wave functions, and the eigenvalues can be expressed as

$$h_{N,k} = (4/\alpha^2) \lambda_{N,k}, \quad \alpha = kab/R, \quad (5.62)$$

where the $\lambda_{N,k}$ are the eigenvalues tabulated by Slepian [1964]; our α corresponds

to his parameter c . The eigenfunctions with $N = 0$ exhibit circular symmetry. The eigenvalues with $N > 0$ have multiplicity 2, corresponding to eigenfunctions proportional to $\cos N\phi$ and $\sin N\phi$, ϕ being the angular coordinate in the aperture plane. Hence in all our expressions the terms for modes with $N > 0$ must be taken twice. With this proviso, there are when $M \gg 1$ approximately $M = \alpha^2/4$ significant eigenvalues, and the rest are nearly 0. When a specific set of eigenvalues is needed for calculating detection probabilities, those for the uniform circular object and the circular aperture will be used.

4.1 Extreme Quantum Limit

When there is no background light at all, the method of §3.1 can be applied immediately. For linearly polarized light the detection probability is given by eq. (5.23), where now instead of eq. (5.24) the probability $p_1(0)$ of receiving no photons at all is given by

$$p_1(0) = \prod_{\underline{p}} \prod_m (1 + h_{\underline{p}} \chi_m N_S)^{-1} = \prod_{\underline{p}} [f(h_{\underline{p}} N_S)]^{-1}, \quad (5.63)$$

with $f(z)$ the Fredholm determinant specified by the temporal spectral density $X(\omega)$ as in eq. (5.25). In particular, for a very small time-bandwidth product, $WT \ll 1$,

$$p_1(0) = \prod_{\underline{p}} (1 + h_{\underline{p}} N_S)^{-1}, \quad (5.64)$$

and for $WT \gg 1$,

$$p_1(0) = \prod_{\underline{p}} \exp(-h_{\underline{p}} N_S) = \exp(-N_S). \quad (5.65)$$

Thus when the observation time is very long, $WT \gg 1$, the detectability of the extended object depends only on the total average number N_S of photons received from it, and not at all on its size and shape, provided that the optimum receiver is utilized.

In Fig. 5.10 we have plotted the probability Q_d of detecting a circular object at a circular aperture when the light is linearly polarized and $WT \ll 1$; curves for various values of the parameter $\alpha = 2M^{\frac{1}{2}}$ are exhibited. In Fig. 5.11 we give the detection probabilities for unpolarized light, determined as in eq. (5.28) appropriately modified.

4.2 Intermediate Range

The optimum statistic for detecting the light from an arbitrary extended object is given in eq. (5.21) as a certain linear combination of the numbers v_{pm} of photons observed in the spatio-temporal modes. For most sources of incoherent light $WT \gg 1$, and for ordinary temperatures and wavelengths the mean number $\mathcal{N}$ of thermal photons per mode, given by the Planck formula, eq. (3.49), is very small, $\mathcal{N} \ll 1$. From eq. (5.19) we see that since χ_m is of the order of $(WT)^{-1}$ and very small, $N_{pm}^{(i)} \ll 1$ for $i = 0, 1$, and the logarithm of the likelihood ratio, eq. (5.21), is approximately

$$U' = \sum_{\underline{p}} \sum_{\underline{m}} [v_{\underline{p}\underline{m}} \ln(1 + h_{\underline{p}} \chi_{\underline{m}} N_S / \mathcal{N}) - h_{\underline{p}} \chi_{\underline{m}} N_S]. \quad (5.66)$$

We shall assume here that the temporal spectral density $X(\omega)$ is rectangular, as in eq. (5.29), so that $\chi_m \approx (WT)^{-1}$ for all $\mu = WT$ significant temporal modes. The statistic U' then depends only on

$$v_{\underline{p}} = \sum_{m=1}^{WT} v_{\underline{p}m}, \quad (5.67)$$

which is the sum of the number of photons in all the μ significant temporal modes into which the aperture mode $\psi_{\underline{p}}(t)$ is decomposed. The statistic U' can then be written as

$$U' = \sum_{\underline{p}} [v_{\underline{p}} \ln(1 + h_{\underline{p}} N_S / \mathcal{N} WT) - h_{\underline{p}} N_S]. \quad (5.68)$$

This statistic depends on the expected number N_S of photons from the object, and it cannot provide a uniformly most powerful test for the presence of the object light. A receiver that is less effective, but whose design does not require knowing the number N_S , is the threshold receiver introduced in

ch. 1 §4. Here its statistic is obtained by expanding U' in powers of $N_s/\mathcal{M}WT$ and keeping only the term proportional to the first. Using the threshold receiver involves comparing the statistic

$$U'' = \sum_{\underline{p}} h_{\underline{p}} v_{\underline{p}} \quad (5.69)$$

with a decision level U_0 and deciding that the light from the object is present whenever $U'' > U_0$. When $WT \gg 1$ the total number $v_{\underline{p}}$ of photons in the aperture mode $\psi_{\underline{p}}(t)$ has a Poisson distribution under each hypothesis,

$$P_i(v_{\underline{p}}) = \frac{v_{\underline{p}}^{v_{\underline{p}}}}{v_{\underline{p}}!} \exp(-v_{\underline{p}}), \quad i = 0, 1, \\ v_{\underline{p}0} = \mathcal{M}WT, \quad v_{\underline{p}1} = h_{\underline{p}} N_s + \mathcal{M}WT. \quad (5.70)$$

These distributions have been used to calculate the probability of detecting a circular object at a circular aperture [Helstrom, 1970b]. In Fig. 5.12 this probability has been plotted versus the average number N_s of received photons for $\mathcal{M}WT = 1$ and various values of the parameter $\alpha = 2M^{\frac{1}{2}}$. The larger the object, the more incoherent aperture modes among which the object light is divided, and the smaller the probability of detection.

The optimum detector represented by the statistic U' in eq. (5.68) was also evaluated and found to attain detection probabilities very close to those of the threshold detector. The two detectors are the same for a point source, when only a single aperture mode contributes. When $M \gg 1$, on the other hand, the significant eigenvalues $h_{\underline{p}}$ are nearly equal to M^{-1} , and both U' and U'' weight the associated numbers $v_{\underline{p}}$ in the same way. Thus the optimum and threshold detectors differ only in their treatment of the modes whose eigenvalues $h_{\underline{p}}$ lie between M^{-1} and zero, and these influence the statistics U' and U'' only weakly.

When $M \gg 1$ the optimum and the threshold statistics are nearly proportional to Poisson-distributed random variables with mean values $\mathcal{M}WT$ and

$\mathcal{N}^{\text{MWT}} + N_s$ under the two hypotheses. The probability of detection can then be read from the curves in Figs. 5.4 and 5.5 if $\mathcal{N}^{\text{WT}}$ is replaced by $\mathcal{N}^{\text{MWT}}$.

6. Estimation Theory

1. Classical Parameter Estimation

The field at the aperture of an optical instrument may consist of two parts, the signal component $\mathcal{F}$ due to light from a source or object plane, and the noise component $\mathcal{F}_0$ due to the thermal background. The field $\mathcal{F}$ may depend on certain parameters of the source, such as its radiant power or its wavelength or, in the case of a laser radar echo, its time of arrival; we denote them by $(\theta_1, \theta_2, \dots, \theta_m) = \underline{\theta}$, and we indicate that the field depends on them by writing it $\mathcal{F}(\underline{\theta})$. The observer wishes to estimate the values of these parameters. His estimates $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ will be based on the actual total field $\mathcal{F}(\underline{\theta}) + \mathcal{F}_0$ at the aperture A during an observation interval $(0, T)$. How best to process that field in order to estimate the parameters $\underline{\theta}$ is a problem in statistical estimation theory. As applied to the parameters of signals received in the presence of noise, estimation theory has been discussed in texts such as those by Middleton [1960, ch. 21], Van Trees [1968, §2.4], and Helstrom [1968c, ch. 8].

When classical electromagnetic theory is valid, we can sample the aperture field at various space-time points $(\underline{r}, t)$, $\underline{r} \in A$, $t \in (0, T)$, denoting the samples we obtain by $\underline{x} = (x_1, x_2, \dots, x_n)$, as in ch. 1. Later these n sampling points are taken closer and closer together, and we pass to the limit $n \rightarrow \infty$ of an infinite number of samples. The joint probability density function (p.d.f.) of the data $\underline{x}$, given that the field is $\mathcal{F}(\underline{\theta}) + \mathcal{F}_0$, is designated by $p(\underline{x}|\underline{\theta})$; it embodies the statistical properties of the object light (the signal) and the background (the noise).

The Bayesian attitude toward estimation, as toward detection, asserts that the best scheme is the cheapest, on the average. We define a cost $C(\hat{\underline{\theta}}, \underline{\theta})$

of issuing estimates $\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)$ of the parameters θ_j when the true values are $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$. For a single parameter, the squared error

$$C(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \quad (6.1)$$

is a common and mathematically tractable cost function. In addition, we must specify a prior p.d.f. $z(\underline{\theta})$ of the parameters, representing the relative frequencies with which their values lie in various regions of the parameter space Θ . An estimation strategy is a set of m functions of the data, $\hat{\theta}_i = \hat{\theta}_i(\underline{x})$, $i = 1, 2, \dots, m$, which we collect into a vector estimator $\hat{\underline{\theta}}(\underline{x})$.

Estimation is a continuous version of a multiple-hypothesis test, in which the hypotheses state that the parameters $\underline{\theta}$ of $p(\underline{x}|\underline{\theta})$ lie within one of numerous infinitesimal regions $d^m \underline{\theta}$ of the space Θ . By analogy with eq. (1.24), the average cost of a particular set of strategies $\hat{\underline{\theta}}(\underline{x})$ is

$$\bar{C} = \iint z(\underline{\theta}) C(\hat{\underline{\theta}}(\underline{x}), \underline{\theta}) p(\underline{x}|\underline{\theta}) d^m \underline{\theta} d^n \underline{x}. \quad (6.2)$$

The best estimator $\hat{\underline{\theta}}(\underline{x})$ is the one for which $\bar{C}$ is minimum. As with multiple hypothesis testing, the best strategy picks that set $\hat{\underline{\theta}}$ for which the posterior risk

$$r(\hat{\underline{\theta}}) = \int C(\hat{\underline{\theta}}, \underline{\theta}) p(\underline{\theta}|\underline{x}) d^m \underline{\theta} \quad (6.3)$$

is minimum--cf. eq. (1.25)--, where

$$p(\underline{\theta}|\underline{x}) = z(\underline{\theta}) p(\underline{x}|\underline{\theta})/p(\underline{x}), \quad p(\underline{x}) = \int z(\underline{\theta}) p(\underline{x}|\underline{\theta}) d^n \underline{x}; \quad (6.4)$$

$p(\underline{\theta}|\underline{x})$ is the posterior p.d.f. of the parameter values $\underline{\theta}$, given the observed data $\underline{x}$, and $p(\underline{x})$ is the total p.d.f. of the data $\underline{x}$.

With a quadratic cost function as in eq. (6.1), the Bayes estimate of a single parameter θ is the conditional expected value

$$\hat{\theta} = \int \theta p(\theta|\underline{x}) d\theta, \quad (6.5)$$

as can be shown by substitution into eq. (6.3) and minimization with respect to $\hat{\theta}$. On the other hand, a cost function of the form

$$C(\hat{\theta}, \theta) = A - B \delta(\hat{\theta} - \theta), \quad (6.6)$$

which is the continuous counterpart to the one in eq. (1.29), leads to the maximum-likelihood estimate, which selects those values of the parameters for which the posterior p.d.f. $p(\theta|\underline{x})$ is maximum. When the prior p.d.f. $z(\theta)$ of the parameters is very broad, as when nothing is known about their values in advance, all information about them being necessarily derived from the measurements of $\underline{x}$, the maximum-likelihood strategy is equivalent to choosing the values of $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ for which the conditional p.d.f. $p(\underline{x}|\theta)$ is maximum.

When the parameters refer to the field created by some source, we can generally specify the p.d.f. $p_0(\underline{x})$ that would describe the data $\underline{x}$ were the source inoperative. In optics this p.d.f. embodies the statistical properties of the background light. As it does not involve the parameters of the source field, we can just as well express these estimation strategies in terms of the likelihood ratio $\Lambda(\underline{x}; \theta) = p(\underline{x}|\theta)/p_0(\underline{x})$. Using such a likelihood ratio facilitates passage to the limit of an infinite number of data, in which all the information in the aperture field is exhausted.

2. Quantum Estimation

When the field $\mathcal{F}(\underline{\theta}) + \mathcal{F}_0$ of the incident light must be treated quantum-mechanically, it is again convenient to imagine it as admitted to the lossless cavity of an ideal receiver by opening the aperture during the observation interval $(0, T)$. The density operator ρ of the field inside the receiver at a later time $t > T$ will depend on the parameters $\underline{\theta}$ of the light from the source, $\rho = \rho(\underline{\theta})$. The parameters $\underline{\theta}$ must be estimated by measurements on the field that are consistent with the laws of quantum mechanics. As with multiple-hypothesis testing, if the density operators $\rho(\underline{\theta})$ commute for all pairs of parameter sets $\underline{\theta}$, they all have a common set of eigenstates; and by determining first in which of these states the cavity field actually is, the parameters can be estimated by the classical methods described in §1. For noncommuting density operators $\rho(\underline{\theta})$, on the other hand, optimum estimation poses difficult problems.

2.1 Minimum-Mean-Square-Error Estimation

Let us consider first the estimation of a single parameter θ . The Bayes cost is, as in eq. (6.2),

$$\bar{C} = \int \text{Tr}[C(\tilde{\theta}, \theta) \rho(\theta)] z(\theta) d\theta, \quad (6.7)$$

where as before $z(\theta)$ is the prior p.d.f. of the unknown parameter and $C(\hat{\theta}, \theta)$ is a cost function measuring the seriousness of an error, that is, of issuing an estimate $\hat{\theta}$ when the true value of the parameter is θ . The estimator $\tilde{\theta}$ is an Hermitian operator whose measurement on the system yields the numerical estimate $\hat{\theta}$, leaving the system in an eigenstate $|\hat{\theta}\rangle$ of $\tilde{\theta}$, $\tilde{\theta}|\hat{\theta}\rangle = \hat{\theta}|\hat{\theta}\rangle$.

When the cost of an error is measured by its square, as in eq. (6.1),

the average cost is

$$\bar{C} = \int \text{Tr}[\rho(\theta) (\tilde{\theta} - \theta)^2] z(\theta) d\theta = \text{Tr}(\tilde{\theta}^2 \Gamma_0 - 2\Gamma_1 \tilde{\theta} + \Gamma_2), \quad (6.8)$$

where Γ_0 , Γ_1 , and Γ_2 are operators defined by

$$\Gamma_k = \int \theta^k z(\theta) \rho(\theta) d\theta. \quad (6.9)$$

Personick [1971] has shown that the optimum estimator $\tilde{\theta}$ is the solution of the operator equation

$$\tilde{\theta} \Gamma_0 + \Gamma_0 \tilde{\theta} = 2\Gamma_1 \quad (6.10)$$

and that the minimum mean-square error is

$$\bar{C}_{\min} = \text{Tr}(\Gamma_2 - \Gamma_1 \tilde{\theta}). \quad (6.11)$$

The operator $\tilde{\theta}$ can be written formally as

$$\tilde{\theta} = 2 \int_0^\infty \exp(-\Gamma_0 \alpha) \Gamma_1 \exp(-\Gamma_0 \alpha) d\alpha, \quad (6.12)$$

which can be shown to satisfy eq. (6.10),

$$\begin{aligned} \Gamma_0 \tilde{\theta} + \tilde{\theta} \Gamma_0 &= -2 \int_0^\infty \frac{\partial}{\partial \alpha} [\exp(-\Gamma_0 \alpha) \Gamma_1 \exp(-\Gamma_0 \alpha)] d\alpha \\ &= -2 \exp(-\Gamma_0 \alpha) \Gamma_1 \exp(-\Gamma_0 \alpha) \Big|_0^\infty = 2\Gamma_1. \end{aligned}$$

When the density operators $\rho(\theta)$ commute for all pairs of values of θ , it is only necessary to determine in which of their common set of eigenstates the system actually is, and measuring the operator $\tilde{\theta}$ is equivalent to finding the conditional expected value given by eq. (6.5). A similar approach to optimum filtering of quantum-mechanical variables has been taken by Grishanin and Stratonovich [1970].

As an example, consider estimating the amplitude Γ of a coherent signal of known form, received in the presence of a thermal background. We showed in ch. 3 §3.1 that by making a certain linear combination of the normal modes of

the receiver we can form a "matched mode" that alone contains the signal. The density operator for this matched mode can be written as in eq. (3.59), and if we take as our unknown parameter Γ as defined by eq. (3.56) in the limit $M \rightarrow \infty$, we can write the density operator as

$$\rho(\Gamma) = (\pi\mathcal{N})^{-1} \int \exp(-|\gamma - \Gamma|^2/\mathcal{N}) |\gamma\rangle\langle\gamma| d^2\gamma. \quad (6.13)$$

We allow Γ to be either positive or negative; it is proportional to the classical signal amplitude, with $N_s = |\Gamma|^2$ the mean number of photons supplied by the source. Here $\mathcal{N}$ is the mean number of background photons per mode, as given by the Planck formula, eq. (3.51); a matched receiver ($Z_0 = Z_s$) is assumed.

Suppose that the amplitude Γ has a Gaussian prior p.d.f.,

$$Z(\Gamma) = (2\pi\Gamma^2)^{-1/2} \exp[-(\Gamma - \bar{\Gamma})^2/2\sigma^2], \quad (6.14)$$

where $\bar{\Gamma}$ is the *a priori* expected value of Γ and σ^2 is a variance measuring our uncertainty about the value of Γ before any observations. The estimation operator $\tilde{\Gamma}$ that minimizes the mean-square error is then

$$\tilde{\Gamma} = [\sigma^2 \mathcal{Q} + (\frac{1}{2}\mathcal{N} + \frac{1}{4})\bar{\Gamma}]/(\sigma^2 + \frac{1}{2}\mathcal{N} + \frac{1}{4}), \quad (6.15)$$

where $\mathcal{Q} = \frac{1}{2}(c_1 + c_1^\dagger)$ is the operator related to the "coordinate" of the harmonic oscillator representing the matched mode, as defined in eq. (3.57) in terms of the annihilation and creation operators for that mode [Personick, 1971]. The minimum mean-square error when $\tilde{\Gamma}$ is measured is

$$\bar{C} = \sigma^2(\frac{1}{2}\mathcal{N} + \frac{1}{4})/(\sigma^2 + \frac{1}{2}\mathcal{N} + \frac{1}{4}). \quad (6.16)$$

If the amplitude Γ is most uncertain *a priori*, $\sigma^2 \rightarrow \infty$, $\tilde{\Gamma} = \mathcal{Q}$, and $\bar{C} = \frac{1}{2}\mathcal{N} + \frac{1}{4}$. The relative mean-square error is

$$E(\hat{\Gamma} - \Gamma)^2/\Gamma^2 = (2\mathcal{N} + 1)/4N_s = D_q^{-2}, \quad (6.17)$$

where D_q^2 is the signal-to-noise ratio specifying, through eq. (3.65), the probability of detecting the coherent signal in thermal noise when the threshold operator $\mathcal{Q}$ is measured on the field in the ideal receiver. In the classical limit this estimator $\tilde{r}$ of the signal amplitude becomes equivalent to the one derived by ordinary estimation theory, and its relative mean-square error approaches the classical value $\mathcal{N}/2N_S = K\mathcal{T}/2E_S$, where E_S is the energy in the signal field and $K\mathcal{T}$ is the mean thermal energy per mode [Helstrom, 1968c, pp. 254-256; 1968d]. Although classically the optimum statistic for estimating the amplitude of a coherent signal in the presence of Gaussian noise is the same as the optimum statistic for detecting the signal, this is not the case quantum-mechanically; the optimum detection operator Π , as we learned in ch. 3 §3, is not the same as the operator $\mathcal{Q}$.

When two or more parameters are to be estimated with minimum mean-square error, the cost function takes the form

$$C(\hat{\underline{\theta}}, \underline{\theta}) = \sum_{i=1}^m \alpha_i (\hat{\theta}_i - \theta_i)^2, \quad (6.18)$$

where the α_i are appropriate weights. Although one might proceed formally as for a single parameter, the resulting estimation operators $\tilde{\theta}_i$ may not commute, and hence may not be simultaneously measurable on the same system. A case in point is the estimation of amplitude and phase of a coherent signal in a thermal background, or equivalently, the estimation of $\bar{\gamma}_x$ and $\bar{\gamma}_y$ in the density operator

$$\rho(\bar{\gamma}) = (\pi\mathcal{N})^{-1} \int \exp(-|\gamma - \bar{\gamma}|^2/\mathcal{N}) |\gamma\rangle\langle\gamma| d^2\gamma, \quad \bar{\gamma} = \bar{\gamma}_x + i\bar{\gamma}_y. \quad (6.19)$$

The estimators that would result involve $\frac{1}{2}(c_1 + c_1^\dagger)$ for $\bar{\gamma}_x$ and $\frac{1}{2}i(c_1^\dagger - c_1)$ for $\bar{\gamma}_y$, and these do not commute. How to find commuting operators $\{\tilde{\theta}_i\}$ that minimize the total mean squared error is unknown.

2.2 Arbitrary Cost Functions

When cost functions other than the squared error, eq. (6.1), govern the estimation, the optimum estimation operators are unknown. For a single parameter, it is necessary to discover an operator $\tilde{\theta}$ that minimizes $\bar{C}$ in eq. (6.7). We have noted the similarity between estimation and multiple hypothesis testing, and in the quantum domain the optimum strategies for both remain undiscovered. There appears to be no quantum counterpart for the posterior p.d.f. $p(\theta|\underline{x})$ in the limit where the data $\underline{x}$ encompass all the information available in the field of the receiver.

3. The Cramér-Rao Inequality

Although the optimum Bayes estimator of a parameter θ of a density operator $\rho(\theta)$ is not in general known, it is possible to set a lower bound to the mean-square error attainable by any estimator having a given bias. In some cases the estimator attaining this lower bound can be determined. The way is shown by the Cramér-Rao inequality of classical statistics.

Suppose that a parameter θ of the joint p.d.f. $p(\underline{x}|\theta)$ of a set $\underline{x}$ of data is to be estimated; the data might be samples of the field at the aperture of an optical instrument during an observation interval $(0, T)$, and θ might be the radiant power of a star in the field of view. An estimator of θ is a function $\hat{\theta}(\underline{x})$ of the data, and its bias is

$$b(\theta) = E[\hat{\theta}(\underline{x})|\theta] - \theta = \int \hat{\theta}(\underline{x}) p(\underline{x}|\theta) d^n \underline{x} - \theta. \quad (6.20)$$

The mean square error is

$$\mathcal{E}(\theta) = E\{[\hat{\theta}(\underline{x}) - \theta]^2|\theta\} = \int [\hat{\theta}(\underline{x}) - \theta]^2 p(\underline{x}|\theta) d^n \underline{x}, \quad (6.21)$$

and it has been shown that this mean square error can be no less than the Cramér-Rao bound given by the inequality

$$\mathcal{E}(\theta) \geq [1 + b'(\theta)]^2 [E\{[\frac{\partial}{\partial \theta} \ln p(\underline{x}; \theta)]^2|\theta\}]^{-1}, \quad (6.22)$$

where $b'(\theta) = db(\theta)/d\theta$, all derivatives being evaluated at the true value of the parameter [Cramér, 1946, p. 473 ff; Rao, 1945]. Furthermore, this lower bound is attainable if the derivative appearing in eq. (6.22) has the form

$$\frac{\partial}{\partial \theta} \ln p(\underline{x}|\theta) = k(\theta) [\hat{\theta}(\underline{x}) - \theta], \quad (6.23)$$

where $k(\theta)$ is independent of the data $\underline{x}$. The estimate $\hat{\theta}(\underline{x})$ in eq. (6.23) is unbiased and attains a minimum mean-square error given by

$$\mathcal{E}(\theta) = [k(\theta)]^{-1}; \quad (6.24)$$

it is termed an *efficient* estimator. Efficient estimators exist only in exceptional cases. The Cramér-Rao inequality in this form has been applied to the estimation of the position of a stellar image on a photosensitive surface by Helstrom [1964] and Farrell [1966]; applications to the estimation of other parameters of an optical source are given by Helstrom [1969b, 1970c].

A quantum-mechanical estimator of the parameter θ will be an Hermitian operator $\tilde{\theta}$ that when measured on the system yields an outcome that is taken as the estimate $\hat{\theta}$ of θ . Its expected value is

$$E[\hat{\theta}|\theta] = \text{Tr}[\rho(\theta) \tilde{\theta}], \quad (6.25)$$

and its bias is, as in eq. (6.20),

$$b(\theta) = E[\hat{\theta}|\theta] - \theta = \text{Tr}[\rho(\theta) (\tilde{\theta} - \theta)]. \quad (6.26)$$

The mean-square error is defined by

$$\mathcal{E}(\theta) = E[(\hat{\theta} - \theta)^2|\theta] = \text{Tr}[\rho(\theta) (\tilde{\theta} - \theta)^2], \quad (6.27)$$

and it can be shown to be bounded below by the quantum-mechanical counterpart of eq. (6.22),

$$\mathcal{E}(\theta) \geq [1 + b'(\theta)]^2 / \text{Tr}(\rho L^2), \quad (6.28)$$

where L is the symmetrized logarithmic derivative (s.l.d.) of $\rho(\theta)$ with respect to θ , defined as the solution of the operator equation

$$\frac{\partial \rho}{\partial \theta} = \frac{1}{2} (\rho L + L \rho) \quad (6.29)$$

[Helstrom, 1967c].

The estimator $\tilde{\theta}$ is said to be efficient if the s.l.d. has the form corresponding to eq. (6.23),

$$L = k(\theta) (\tilde{\theta} - \theta),$$

$k(\theta)$ being a c-number function, whereupon the estimator $\tilde{\theta}$ is also unbiased and attains the minimum mean-square error given by eq. (6.28) with $b'(\theta) = 0$. As in eq. (6.24), the minimum mean-square error is $[k(\theta)]^{-1}$.

Multi-dimensional versions of both the classical and the quantum-mechanical Cramér-Rao inequalities exist [Helstrom, 1968a] and permit one to bound the mean-square errors of estimates of more than a single unknown parameter. An extension to parameter estimation when a prior p.d.f. of the unknown parameters is known has been given by Personick [1971], who has applied the bound to estimation in quantum communications.

The estimator $\mathcal{Q} = \frac{1}{2}(c_1 + c_1^\dagger)$ of the amplitude Γ of a coherent signal in thermal noise, which was derived in §2.1, is an efficient estimator. The s.l.d. L has the form

$$L = 4(\mathcal{Q} - \Gamma)/(2\mathcal{N} + 1), \quad (6.31)$$

and $[k(\Gamma)]^{-1} = \frac{1}{4}(2\mathcal{N} + 1)$ is the mean-square error it attains, as in eq. (6.16) for $\sigma^2 \rightarrow \infty$ [Helstrom, 1968a]. A smaller mean-square error is possible, but only for a biased estimator.

If both components of the complex amplitude $\bar{\gamma} = \bar{\gamma}_x + i\bar{\gamma}_y$ of a coherent signal in thermal noise are to be estimated--the density operator is now that given in eq. (6.19)--, the multivariate Cramér-Rao inequality in its quantum-mechanical form sets the low bounds

$$\mathbb{E}(\hat{\gamma}_x - \bar{\gamma}_x)^2 \geq \frac{1}{4}(2\mathcal{N} + 1), \quad \mathbb{E}(\hat{\gamma}_y - \bar{\gamma}_y)^2 \geq \frac{1}{4}(2\mathcal{N} + 1). \quad (6.32)$$

The operators

$$\tilde{\gamma}_x = \frac{1}{2}(c_1 + c_1^\dagger), \quad \tilde{\gamma}_y = \frac{1}{2}i(c_1^\dagger - c_1)$$

are efficient estimators of $\bar{\gamma}_x$ and $\bar{\gamma}_y$, respectively, but they do not commute and hence cannot be measured in the same receiver.

If the field in the matched mode is amplified, as described in ch. 3 §5, the complex amplitude of the matched mode might be estimated and the estimate divided by the square root $G^{\frac{1}{2}}$ of the gain to yield estimates of $\bar{\gamma}_x$ and $\bar{\gamma}_y$. The resulting minimum mean-square errors are obtained by substituting from eq. (3.90) $G(\mathcal{N} + 1) - 1$ for $\mathcal{N}$. In the limit $G \rightarrow \infty$, both $G^{\frac{1}{2}}\bar{\gamma}_x$ and $G^{\frac{1}{2}}\bar{\gamma}_y$ can be measured simultaneously by classical-physical efficient estimators, and the minimum mean-square errors so attained are, after division by G ,

$$\mathbb{E}(\hat{\gamma}_x - \bar{\gamma}_x)^2 = \frac{1}{2}(\mathcal{N} + 1), \quad \mathbb{E}(\hat{\gamma}_y - \bar{\gamma}_y)^2 = \frac{1}{2}(\mathcal{N} + 1), \quad (6.33)$$

which are larger than those in eq. (6.32). These correspond to an uncertainty product

$$[\mathbb{E}(\hat{p} - p)^2 \mathbb{E}(\hat{q} - q)^2]^{\frac{1}{2}} = \hbar(\mathcal{N} + 1)$$

in the coordinate $q = (2\hbar/\Omega)^{\frac{1}{2}}(c_1 + c_1^\dagger)$ and momentum $p = (2\hbar\Omega)^{\frac{1}{2}}i(c_1^\dagger - c_1)$ of the harmonic oscillator representing the matched mode. When $\mathcal{N} = 0$, this is the same uncertainty product as found by Arthurs and Kelly [1965] in a study of the simultaneous measurement of noncommuting observables; see also She and Heffner [1966].

The quantum-mechanical Cramér-Rao inequality has been applied to the estimation of parameters of incoherently radiating objects [Helstrom, 1970a]. It was assumed that the product WT of the observation time T and the bandwidth W of the object light is very large, yet the mean number $\mathcal{N}$ of noise photons per mode is very small, so that $\mathcal{N}WT$ is of the order of 1. For a point source, the mean-square relative error of an unbiased estimate of the radiant power B_0 of a point source is bounded below by

$$\mathbb{E}(\hat{B}_0 - B_0)^2/B_0^2 \geq [N_S f_1(\mathcal{D})]^{-1}$$

where with $X(\omega)$ the spectral density of the object light as defined in eq. (5.2),

$$f_1(\mathcal{D}) = \mathcal{D}W \int_{-\infty}^{\infty} [X(\omega)]^2 [1 + \mathcal{D}WX(\omega)]^{-1} d\omega/2\pi,$$

$$\mathcal{D} = N_S/\mathcal{N}WT.$$

In the extreme quantum limit, $\mathcal{D} \rightarrow \infty$, $f_1(\mathcal{D}) \rightarrow 1$, and the minimum mean-square relative error is bounded below by N_S^{-1} . For $\mathcal{N}WT \gg N_S$, $f_1(\mathcal{D}) \doteq \mathcal{D}$, and the lower bound is $(N_S^2/\mathcal{N}WT)^{-1}$.

For an unbiased estimate of the frequency Ω of incoherent light having bandwidth W and coming from a point source, the following lower bounds were derived from the Cramér-Rao inequality,

$$E(\hat{\Omega} - \Omega)^2/W^2 \geq 2/N_S, \quad \mathcal{N}WT \ll N_S$$

$$E(\hat{\Omega} - \Omega)^2/W^2 \geq 2\mathcal{N}WT/N_S^2, \quad \mathcal{N}WT \gg N_S.$$

For a coordinate u_x of the position of the point source in a plane at distance R , the relative mean-square error is subject to the following bounds,

$$E(\hat{u}_x - u_x)^2/\delta^2 \geq 1/N_S, \quad \mathcal{N}WT \ll N_S$$

$$E(\hat{u}_x - u_x)^2/\delta^2 \geq 2\mathcal{N}WT/N_S^2, \quad \mathcal{N}WT \gg N_S.$$

Here $\delta = \lambda R/2\pi a$, where $\lambda = 2\pi c/\Omega$ is the wave-length of the light and a is the radius of the aperture, taken as circular; thus δ represents the size of a resolution element in the object plane. Bounds on errors in estimates for extended sources were also derived in the reference cited.

Acknowledgment

The writer's research described in this article was partly carried out under Grant NGL 05-009-079 from the National Aeronautics and Space Administration. I wish to thank Mrs. Lily Wang and Mr. Yie-Ming Hong for their assistance with the numerical computations.

BIBLIOGRAPHY

- Arthurs, E. and J. L. Kelly, Jr., 1965, Bell System Tech. J. 44, 725.
- Bakut, P. A., 1966, Radio Eng. Electron. (USSR) 11, 551
- 1967, Radio Eng. Electron. (USSR) 12, 1
- Bakut, P. A. and S. S. Shchurov, 1968, Probl. Peredachi Informatsii 4(1), 77
- Bakut, P. A., V. G. Vygon, A. A. Kuriksha, V. G. Repin, and G. P. Tartakovskii, 1966, Probl. Peredachi Informatsii 2(4), 39
- Bar-David, I., 1969, Trans. IEEE IT-15, 31
- Cramér, H., 1946, Mathematical Methods of Statistics (Princeton Univ. Press, Princeton)
- Farrell, E. J., 1966, J. Opt. Soc. Am. 56, 578
- Gabor, D., 1961, Progress in Optics 1, 138
- Gagliardi, R., 1972, Trans. IEEE IT-18,
- Ginzburg, S. A., 1966, Radio Eng. Electron. (USSR) 11, 1972
- Glauber, R. J., 1963, Phys. Rev. 131, 2766
- Goodman, J. W., 1966, Trans. IEEE AES-2, 526
- Green, H. S. and E. Wolf, 1953, Proc. Phys. Soc. (London) A66, 1129
- Grenander, U. and G. Szegö, 1958, Toeplitz Forms and Their Applications (Univ. of California Press, Berkeley)
- Grishanin, B. A. and R. L. Stratonovich, 1970, Probl. Peredachi Informatsii 6(3), 15
- Haus, H. A., 1970, Proc. IEEE 58, 1599
- Helstrom, C. W., 1964, Trans. IEEE IT-10, 275.
- 1967a, J. Opt. Soc. Am. 57, 353
- 1967b, Infor. & Control 10, 254

- 1967c, Phys. Letters 25A, 101
 - 1968a, Trans. IEEE IT-14, 234
 - 1968b, Int. J. Theor. Phys. 1, 37
 - 1968c, Statistical Theory of Signal Detection, 2nd ed. (Pergamon Press, Oxford)
 - 1968d, Infor. & Control 13, 156
 - 1969a, Trans. IEEE AES-5, 562
 - 1969b, J. Opt. Soc. Am. 59, 164
 - 1969c, J. Opt. Soc. Am. 59, 924
 - 1969d, J. Stat. Phys. 1, 231
 - 1970a, J. Opt. Soc. Am. 60, 233
 - 1970b, J. Opt. Soc. Am. 60, 521
 - 1970c, J. Opt. Soc. Am. 60, 659
 - 1971, Trans. IEEE AES-7, 210
- Helstrom, C. W., J. W. S. Liu, and J. P. Gordon, 1970, Proc. IEEE 58, 1578
- Hoversten, E. V., R. O. Harger, and S. J. Halme, 1970, Proc. IEEE 58, 1626
- Karp, S. and J. R. Clark, 1970, Trans. IEEE IT-16, 672
- Karp, S., E. L. O'Neill, and R. M. Gagliardi, 1970, Proc. IEEE 58, 1611
- Kennedy, R. S., 1970, Proc. IEEE 58, 1651
- Kennedy, R. S and E. V. Hoversten, 1968, Trans. IEEE IT-14, 716
- Kuriksha, A. A., 1968, Radio Eng. Electron. (USSR) 13, 1567
- Lachs, G., 1965, Phys. Rev. 138, B1012
- Lax, M., 1966, Phys. Rev. 145, 110
- Liu, J. W. S., 1970, Trans. IEEE IT-16, 319
- Louisell, W. H., 1964, Radiation and Noise in Quantum Electronics (McGraw-Hill, New York)
- Louisell, W. H. and L. R. Walker, 1965, Phys. Rev. 137, B204

- Mandel, L., 1958, Proc. Phys. Soc. (London) 72, 1037
- Middleton, D., 1960, An Introduction to Statistical Communication Theory
(McGraw-Hill, New York)
- 1966, Trans. IEEE IT-12, 230
- Personick, S., 1971, Trans. IEEE IT-17, 240.
- Rao, C. R., 1945, Bull. Calcutta Math. Soc. 37, 81
- Reiffen, B. and H. Sherman, 1963, Proc. IEEE 51, 1316
- Rudnick, P., 1962, Nature 193, 604
- She, C. Y., 1965, J. Appl. Phys. 36, 3784
- 1968, Trans. IEEE IT-14, 32
- She, C. Y. and H. Heffner, 1966, Phys. Rev. 152, 1103
- Siegert, A. J. F., 1957, Trans. IRE IT-3, 38
- Slepian, D., 1964, Bell System Tech. J. 43, 3009
- Slepian, D. and H. O. Pollak, 1961, Bell System Tech. J. 40, 43
- Slepian, D. and E. Sonnenblick, 1965, Bell System Tech. J. 44, 1745
- Stefanyuk, V. L., 1966, Probl. Peredachi Informatsii 2(1), 58
- Takahasi, H., 1965, Information Theory of Quantum-Mechanical Channels. In:
Balakrishnan, A., ed., Adv. in Communications Systems (Academic Press,
New York), vol. 1, pp. 227-310
- Van Trees, H. L., 1968, Detection, Estimation, and Modulation Theory, pt. 1
(Wiley, New York)
- Walther, A., 1967, J. Opt. Soc. Am. 57, 639
- Yoshitani, R., 1970, On the Detectability Limit of Coherent Optical Signals in
Thermal Radiation, Ph.D. thesis, Univ. of California, Los Angeles
- Yuen, H. P., R. S. Kennedy, and M. Lax, 1970, Proc. IEEE 58, 1770

List of Figures

- 1.1 Decision regions in two dimensions.
- 3.1 Transmission-line receiver.
- 3.2 Transmission-line terminated at both ends.
- 3.3 Probability Q_d of detecting a coherent signal versus the signal-to-noise ratio $D_q = [4N_s/(2\mathcal{N} + 1)]^{1/2}$ for $Q_0 = 10^{-4}$. The curve "optimum" gives Q_d for the optimum receiver when the signal phase is known and $\mathcal{N} = 0$. The dashed curve gives Q_d for the threshold receiver of a signal of known phase for all $\mathcal{N}$. The remaining curves refer to a receiver that counts the number of photons in the matched mode; these are indexed by the value of $\mathcal{N}$.
- 3.4 Probability P_e of error in detection of known signal with prior probability $\zeta = \frac{1}{2}$. D_q = signal-to-noise ratio = $[4N_s/(2\mathcal{N} + 1)]^{1/2}$, N_s = average number of signal photons, $\mathcal{N}$ = average number of noise photons. [From Helstrom et al., 1970.]
- 3.5 Probability Q_d of detecting an amplified coherent signal of random phase versus the initial number N_s of signal photons; $\mathcal{N} = 0$, $Q_0 = 10^{-4}$. The curves are indexed with the gain G .
- 3.6 Probability Q_d of detecting an amplified coherent signal of random phase versus the initial number N_s of signal photons; $\mathcal{N} = 1$, $Q_0 = 10^{-2}$ and 10^{-4} . The curves are indexed with the gain G .
- 5.1 Probability Q_d of detecting linearly polarized incoherent light from a point source versus number N_s of photons; $\mathcal{N} = 0$, $Q_0 = 10^{-4}$. Curves are indexed by the time-bandwidth product WT . Solid lines: Lorentz spectrum;

- dashed lines: rectangular spectrum. The curve for $WT = 6.366$ for the rectangular spectrum nearly coincides with that for $WT = 5$ for the Lorentz spectrum.
- 5.2 Probability Q_d of detecting unpolarized incoherent light from a point source versus number N_s of photons; $\mathcal{N} = 0$, $Q_0 = 10^{-4}$. Curves are indexed by the time-bandwidth product WT . Solid lines: Lorentz spectrum; dashed lines: rectangular spectrum.
- 5.3 Probability Q_d of detecting an incoherent point source with a rectangular spectral density versus number N_s of photons from the source; $\mathcal{N} = 1$, $Q_0 = 10^{-4}$. Curves are indexed with the number $\mu = WT$ of significant modes.
- 5.4 Probability Q_d of detecting an incoherent point source having a rectangular spectral density and supplying a total number N_s of photons in the Poisson limit $\mathcal{N} \ll 1$, $WT \gg 1$. The curves are indexed by the average total number $\mathcal{N}WT$ of thermal photons in the significant modes; $Q_0 = 10^{-4}$.
- 5.5 Probability Q_d of detecting an incoherent point source having a rectangular spectral density and supplying a total number N_s of photons, versus $D_n = N_s(\mathcal{N}WT)^{-1/2}$, in the Poisson limit $\mathcal{N} \ll 1$, $WT \gg 1$. The curves are indexed by the average total number $\mathcal{N}WT$ of thermal photons in the significant modes; $Q_0 = 10^{-4}$.
- 5.6 Probability Q_d of detecting an amplified incoherent signal of random phase and rectangular spectral density versus the mean number N_s of signal photons; $\mathcal{N} = 0$, $Q_0 = 10^{-4}$, $WT = 5$. The curves are indexed with the gain G .

- 5.7 Probability Q_d of detecting an amplified incoherent signal of random phase and rectangular spectral density versus the mean number N_s of signal photons; $\mathcal{N} = 0$, $Q_0 = 10^{-4}$, $WT = 10$. The curves are indexed with the gain G .
- 5.8 Probability Q_d of detecting an incoherent point source with a Lorentz spectral density in the classical limit $\mathcal{N} \gg 1$, versus the signal-to-noise ratio $D^2 = N_s/\mathcal{N}$; $Q_0 = 10^{-4}$. The curves are indexed by the time-bandwidth product WT .
- 5.9 Probability Q_d of detecting an incoherent point source with a rectangular spectral density in the classical limit $\mathcal{N} \gg 1$, versus the signal-to-noise ratio $D^2 = N_s/\mathcal{N}$; $Q_0 = 10^{-4}$. The curves are indexed by the time-bandwidth product WT .
- 5.10 Probability Q_d of detecting a circular object at a circular aperture in the extreme quantum limit, versus the average number N_s of photons received; linearly polarized light, $WT \ll 1$, $Q_0 = 10^{-4}$. Curves are indexed by the parameter $\alpha = 2M^{\frac{1}{2}} = kab/R$.
- 5.11 Probability Q_d of detecting a circular object at a circular aperture in the extreme quantum limit, versus the average number N_s of photons received; unpolarized light, $WT \ll 1$, $Q_0 = 10^{-4}$. Curves are indexed by the parameter $\alpha = 2M^{\frac{1}{2}} = kab/R$.
- 5.12 Probability Q_d of detecting a uniform circular object of radius b by observations at a circular aperture of radius a , versus the average number N_s of photons received from the object; $\mathcal{N}/WT = 1$, $Q_0 = 10^{-2}$. The curves are indexed by the parameter $\alpha = 2M^{\frac{1}{2}} = kab/R$. [From Helstrom, 1970b.]

List of Symbols

Boldface letters:

$\mathbf{\bar{E}}$ $\mathbf{\bar{p}}$ $\mathbf{\bar{r}}$ $\mathbf{\bar{u}}$ $\mathbf{\bar{U}}$ $\mathbf{\bar{x}}$ $\mathbf{\bar{0}}$ (zero) $\mathbf{\bar{1}}$ (one) $\mathbf{\bar{\theta}}$ $\mathbf{\hat{\theta}}$

Symbols:

ϵ $|$ $\rangle$ $\langle$ $|$ $\approx$

Greek letters:

α β γ Γ $\bar{\Gamma}$ $\tilde{\Gamma}$ $\hat{\Gamma}$ δ Δ ζ η

θ $\hat{\theta}$ $\tilde{\theta}$ $\cdot \theta$ λ Λ $\bar{\Lambda}$ μ ν $\bar{\nu}$ ξ

π Π Π_{θ} ρ σ τ ϕ Φ χ ψ Ψ

ω Ω $\bar{\gamma}$ $\hat{\Omega}$

Script letters:

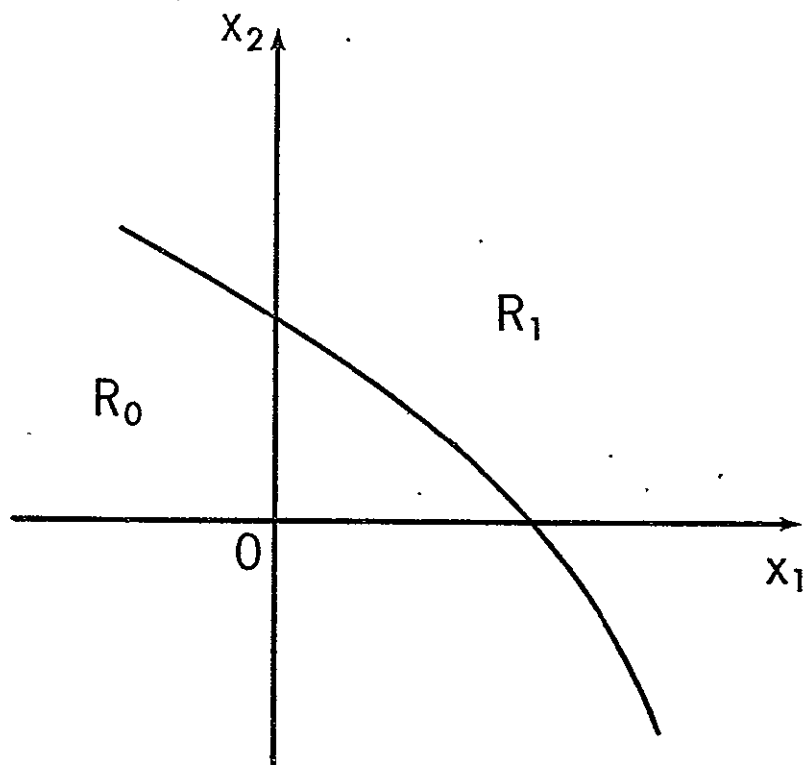
$\mathcal{E}$ $\mathcal{F}$ $\mathcal{H}$ $\mathcal{N}$ $\mathcal{Q}$ $\mathcal{T}$ $\mathcal{D}$

Roman letters

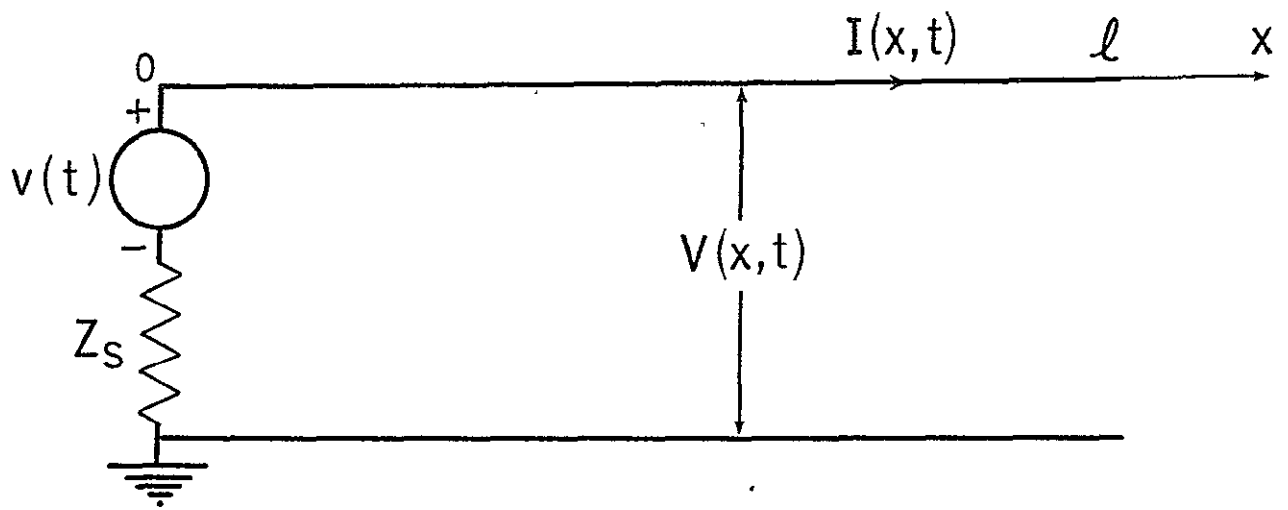
$\bar{C}$ $\bar{g}$ $\bar{h}$ $\bar{n}$ ℓ (ell) $\bar{Q}_d$ $\bar{z}$ $\hat{u}$ $\hat{B}$ $\bar{n}$

Sans-serif letters

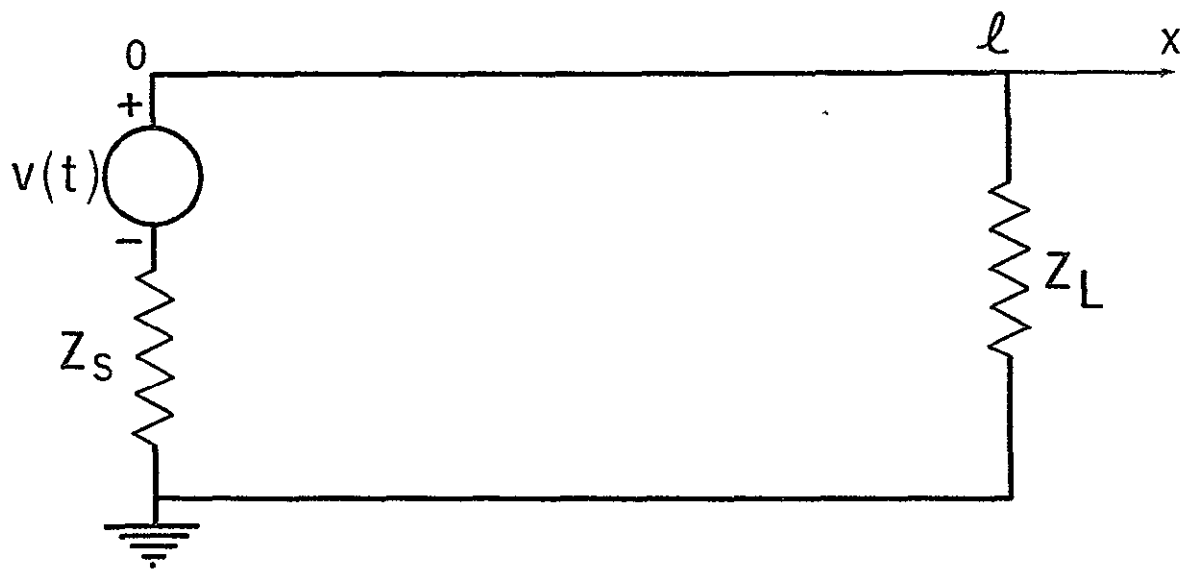
P_0 P_1 P_i



1.1 Decision regions in two dimensions.



3.1 Transmission-line receiver.



3.2 Transmission line terminated at both ends.

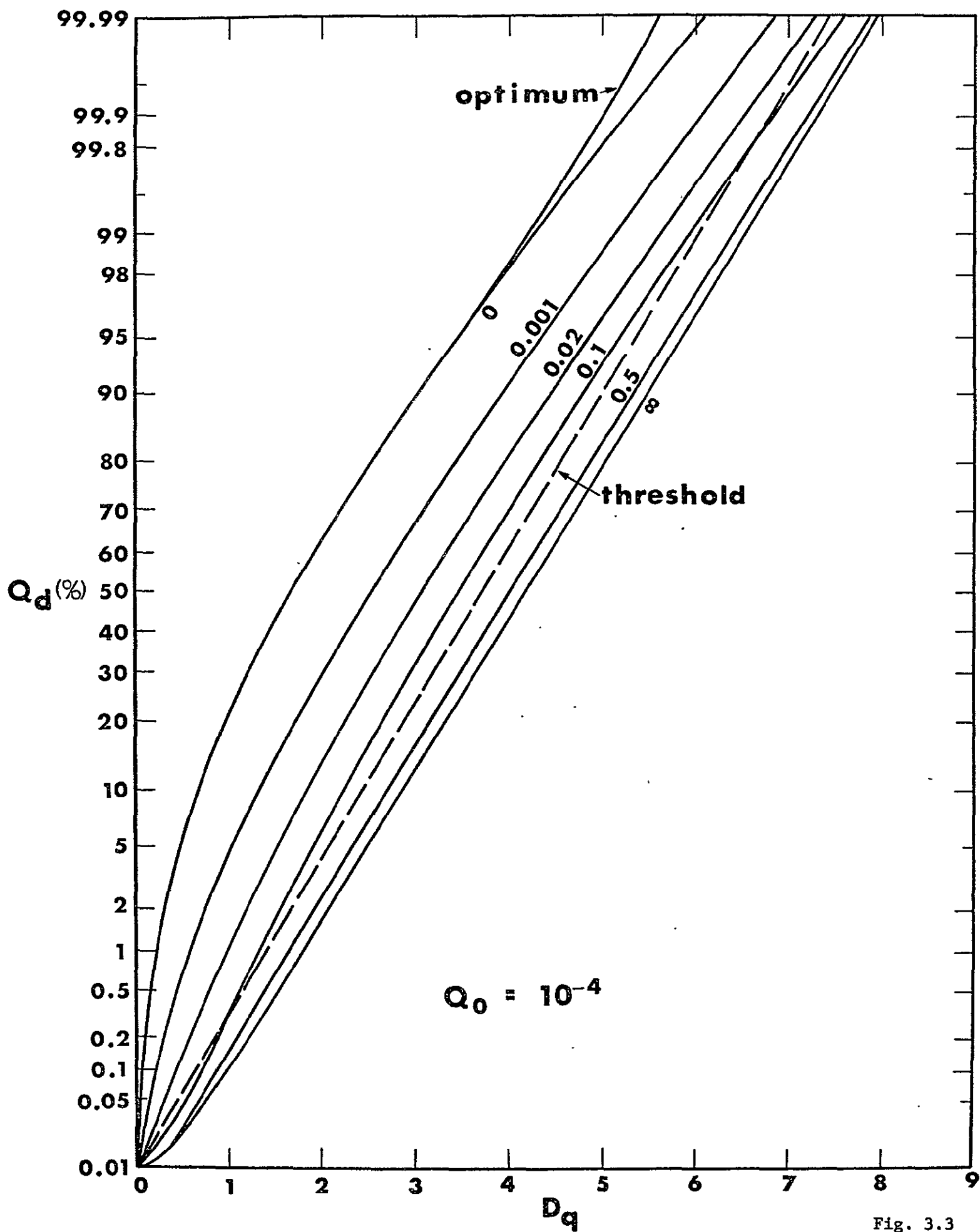


Fig. 3.3

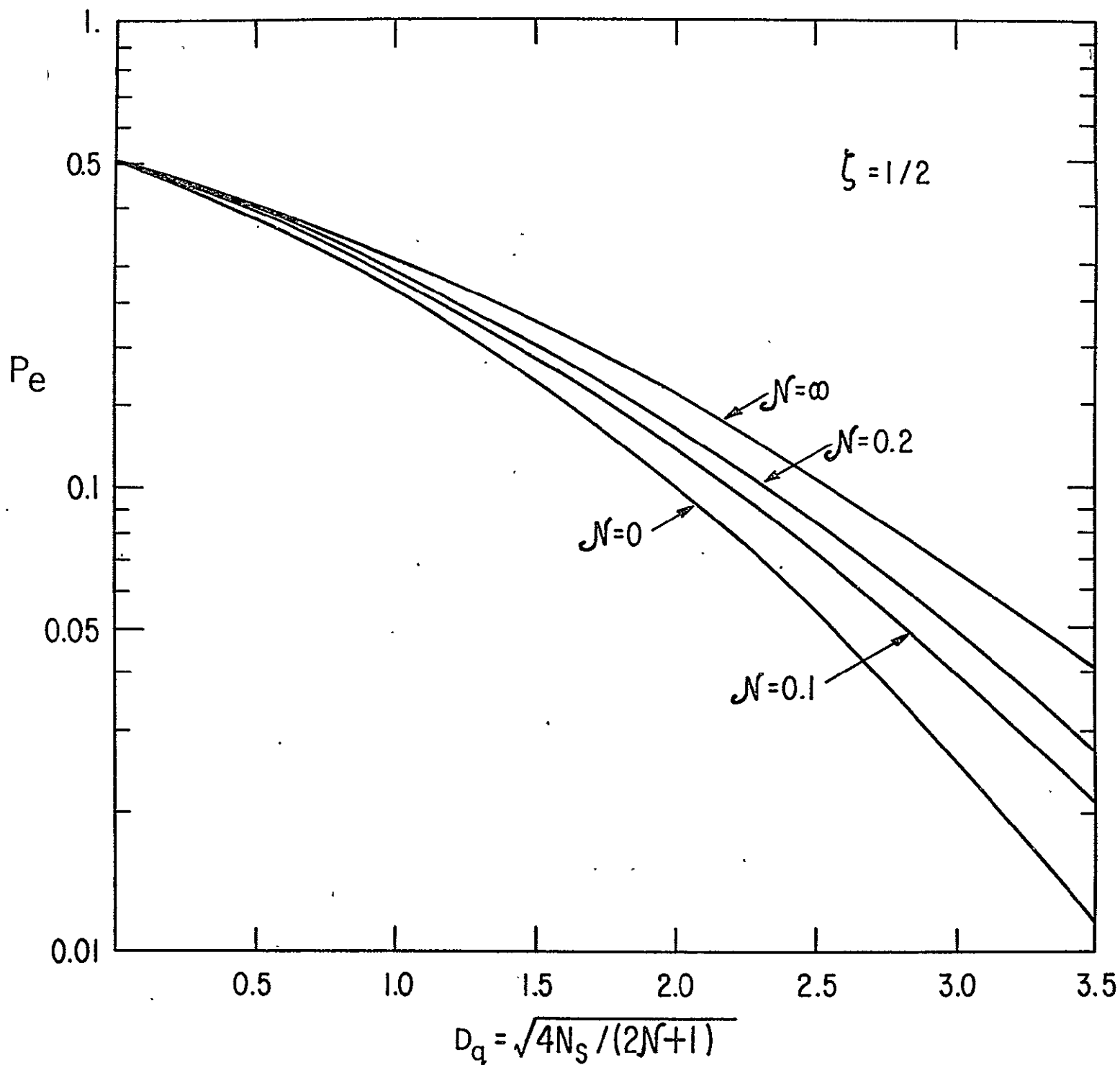


Fig. 3.4 Probability P_e of error in detection of known signal with prior probability $\zeta = \frac{1}{2}$. D_q = signal-to-noise ratio = $[4N_s / (2N + 1)]^{\frac{1}{2}}$, N_s = average number of signal photons, N = average number of noise photons. [From Helstrom et al., 1970.]

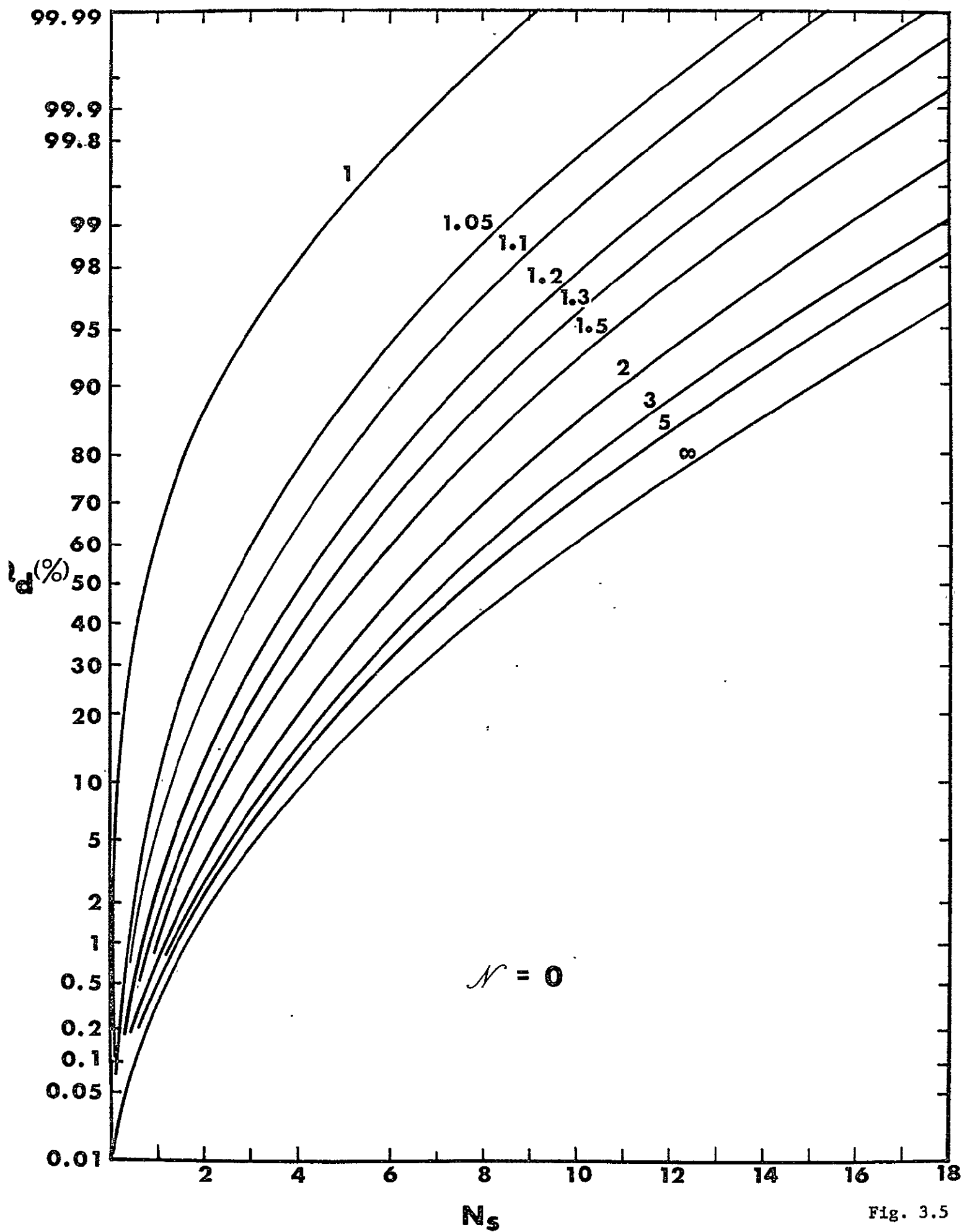


Fig. 3.5

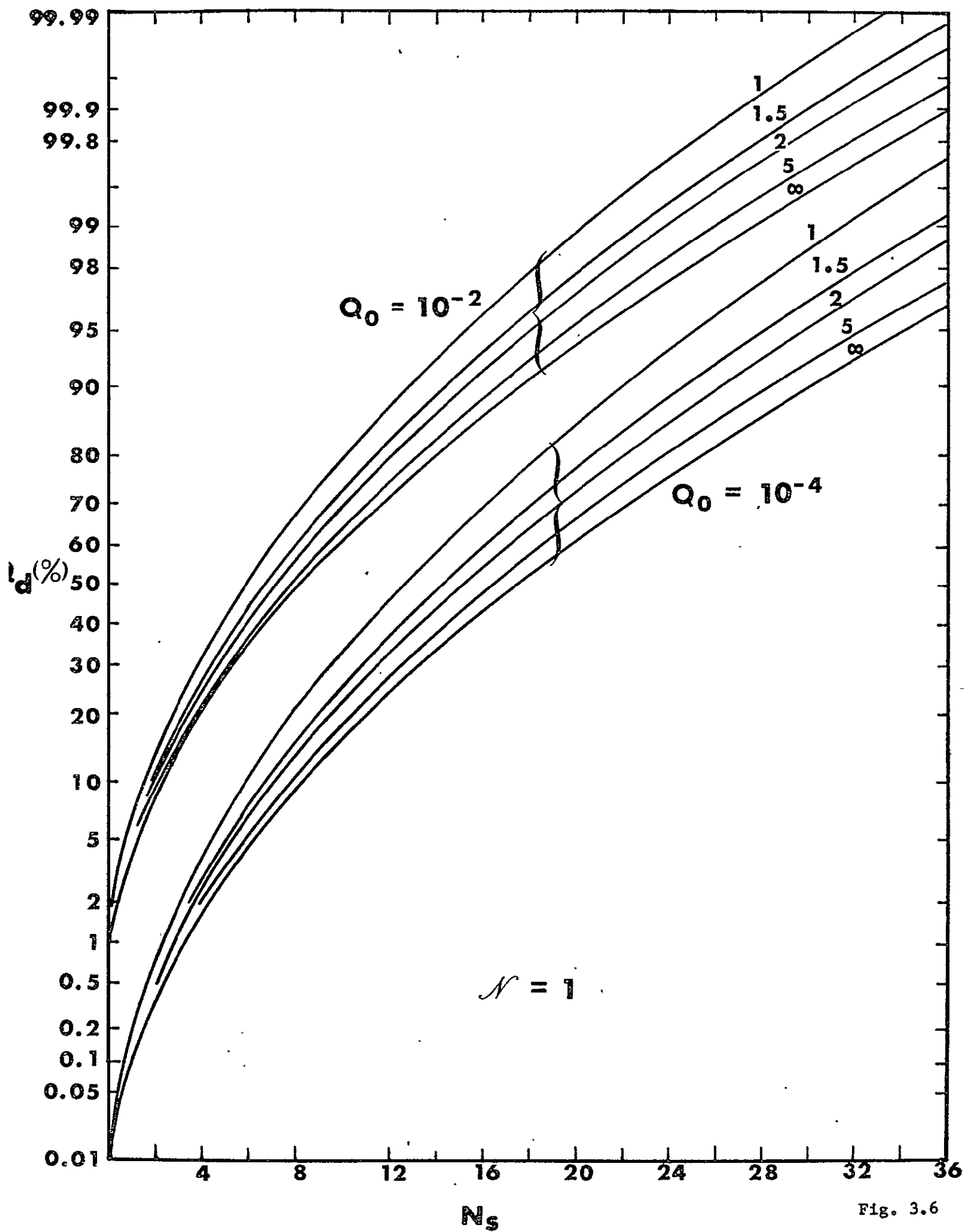


Fig. 3.6

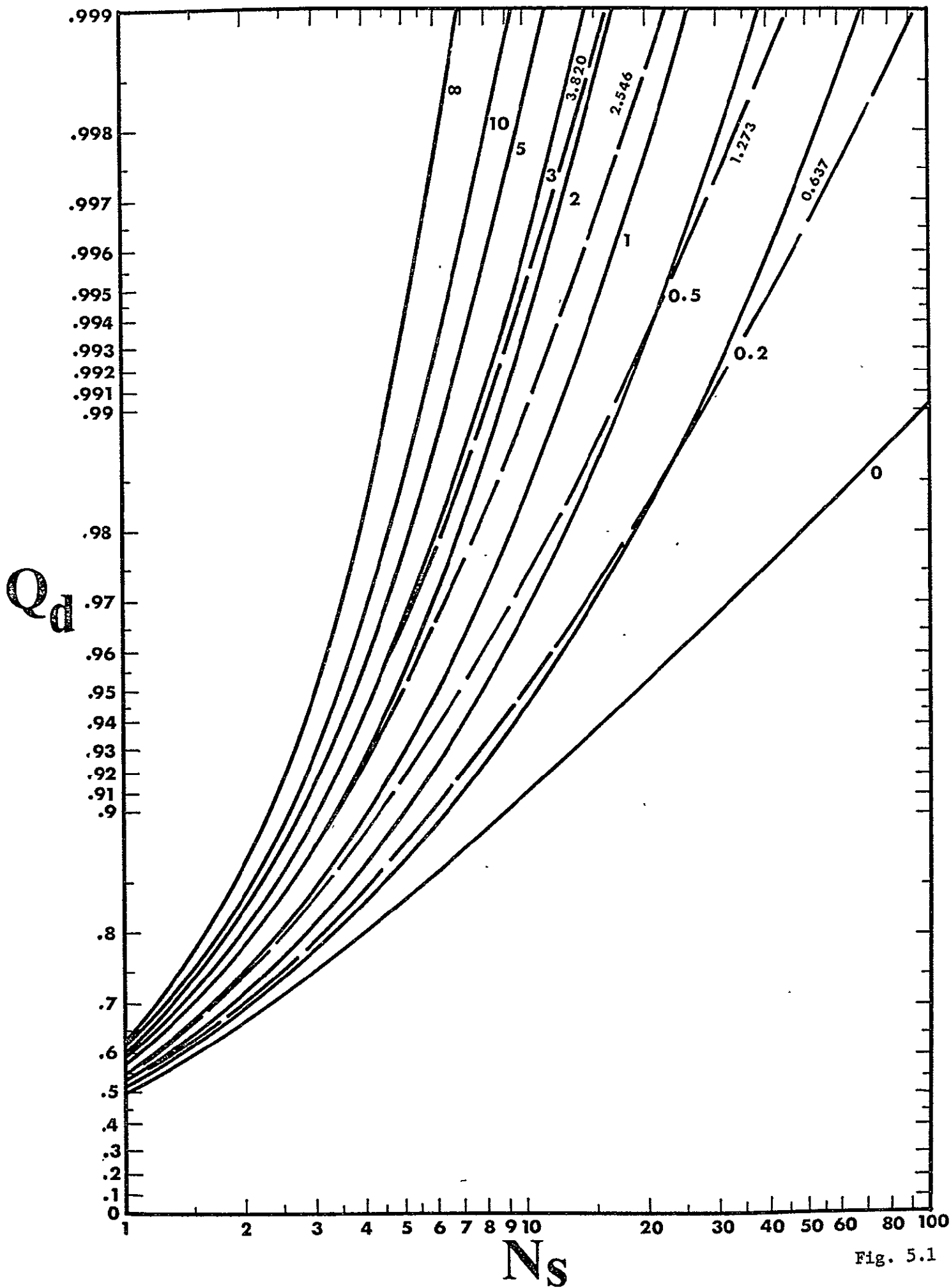


Fig. 5.1

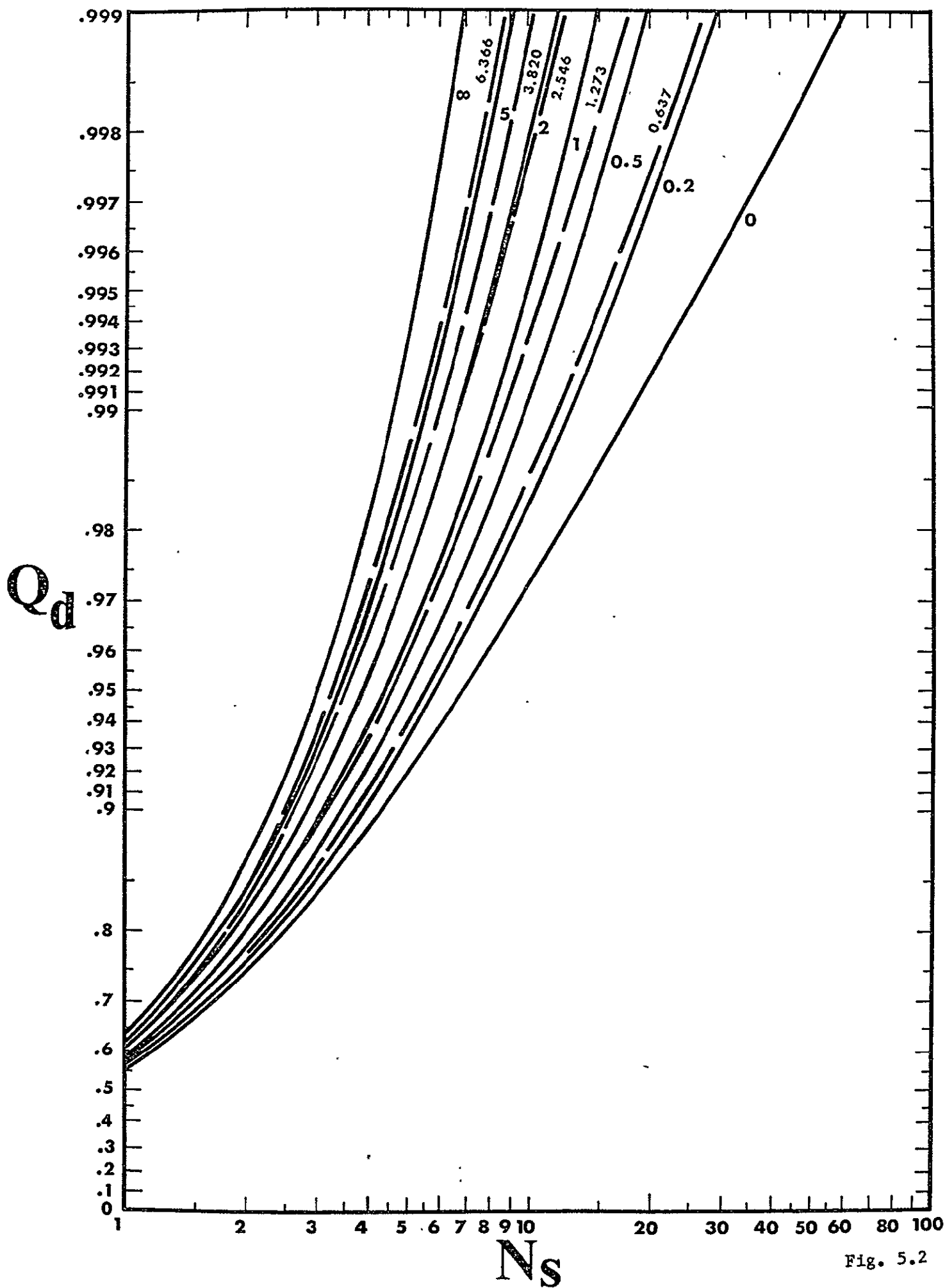
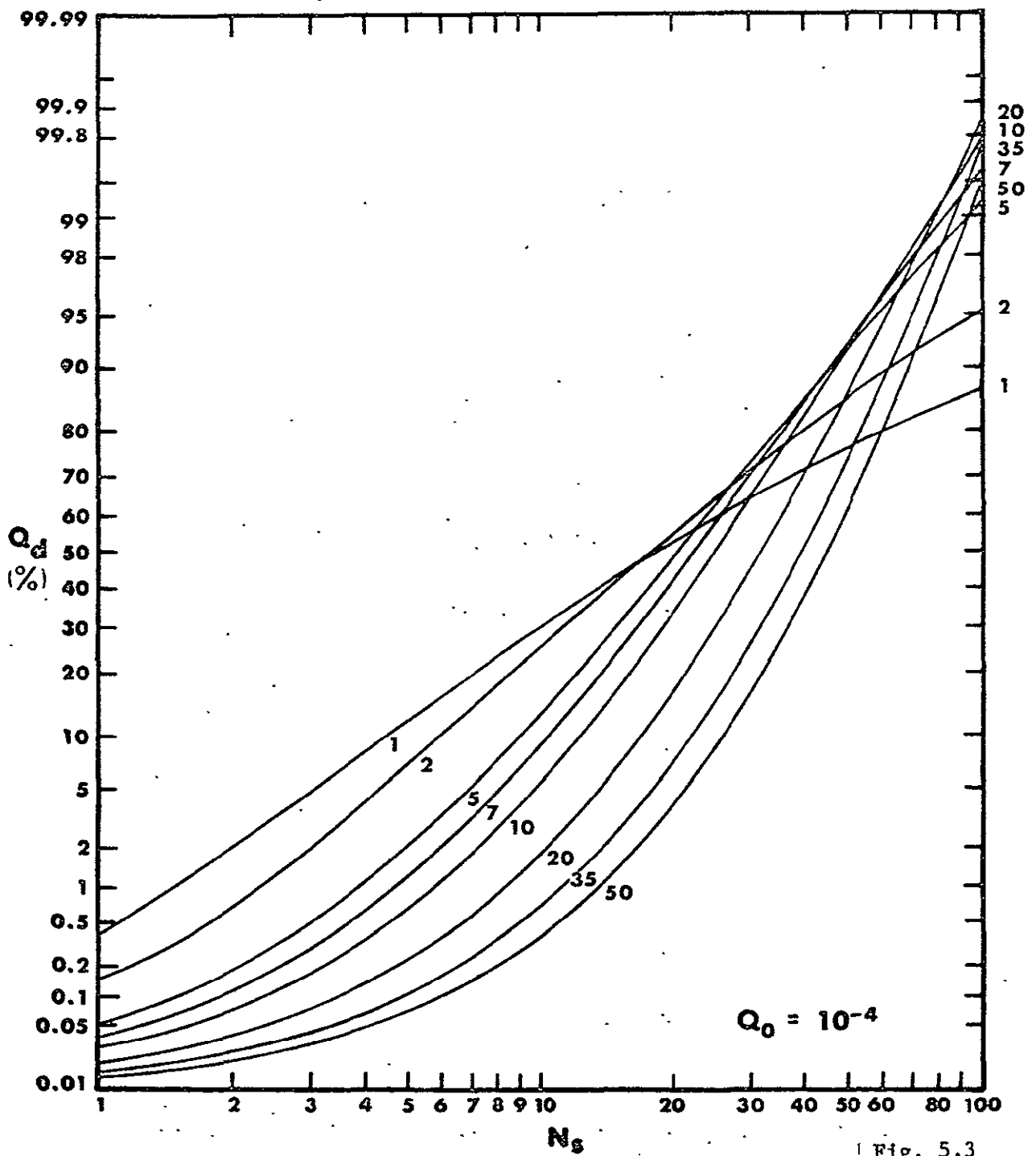


Fig. 5.2



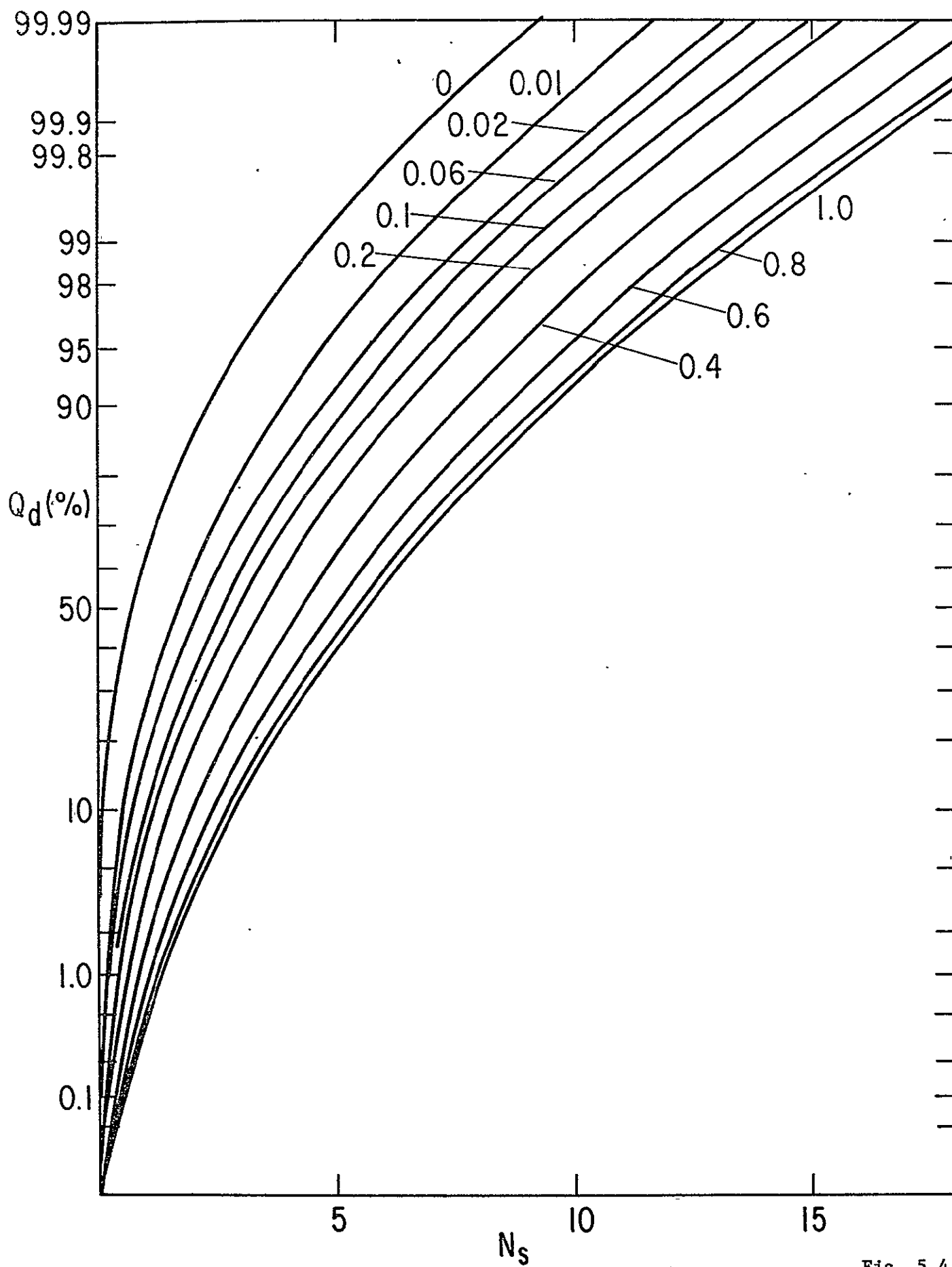


Fig. 5.4

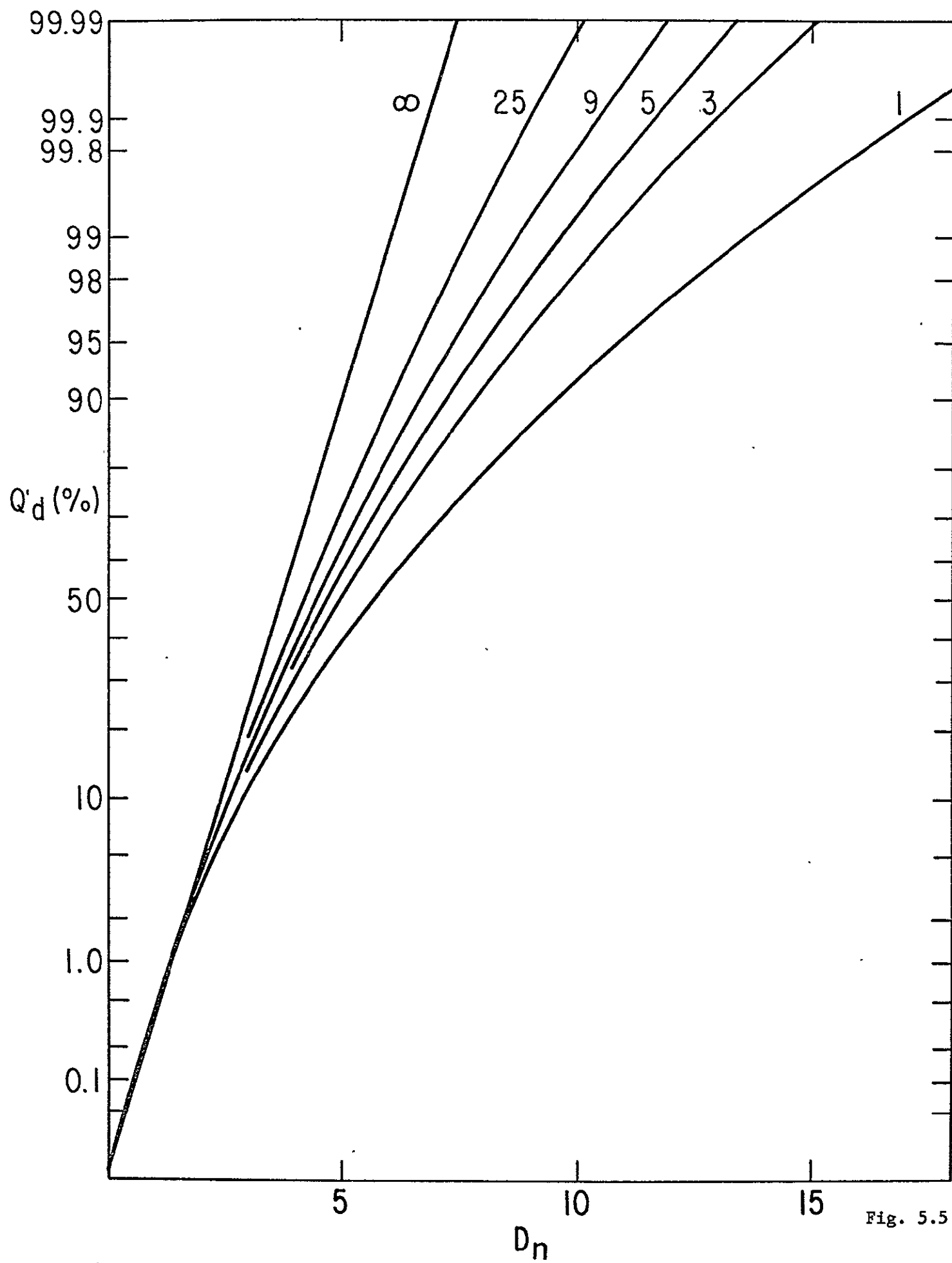


Fig. 5.5

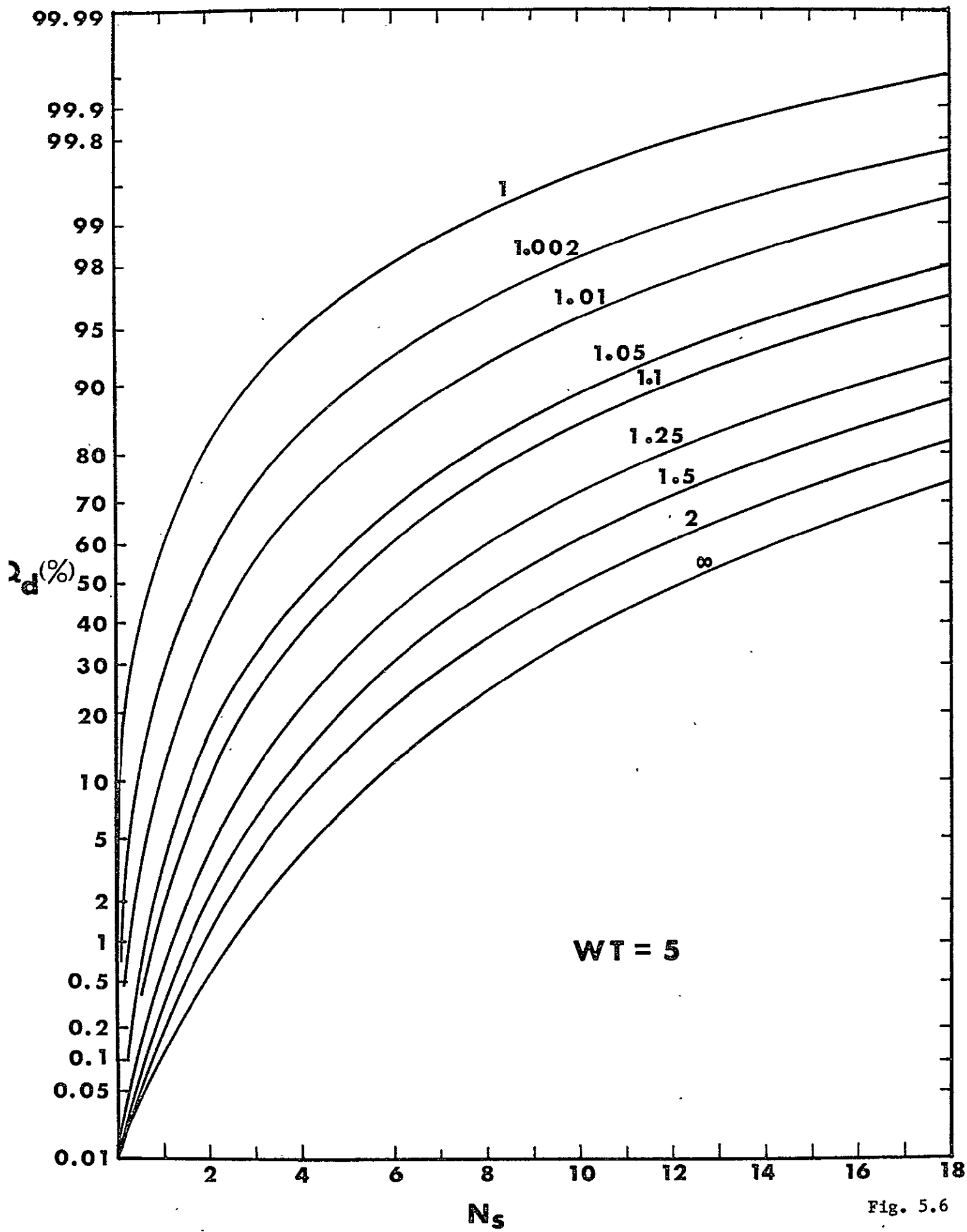


Fig. 5.6

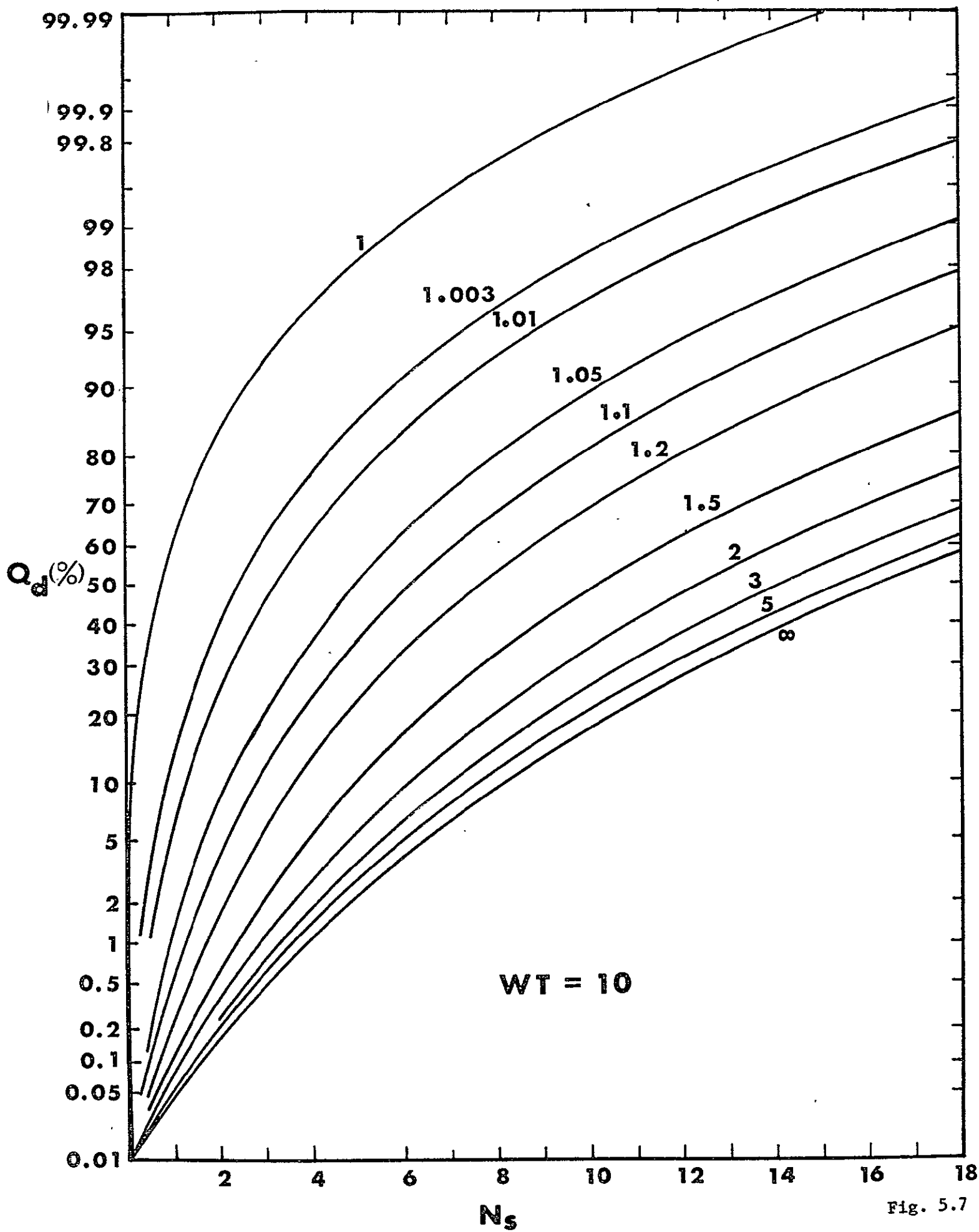
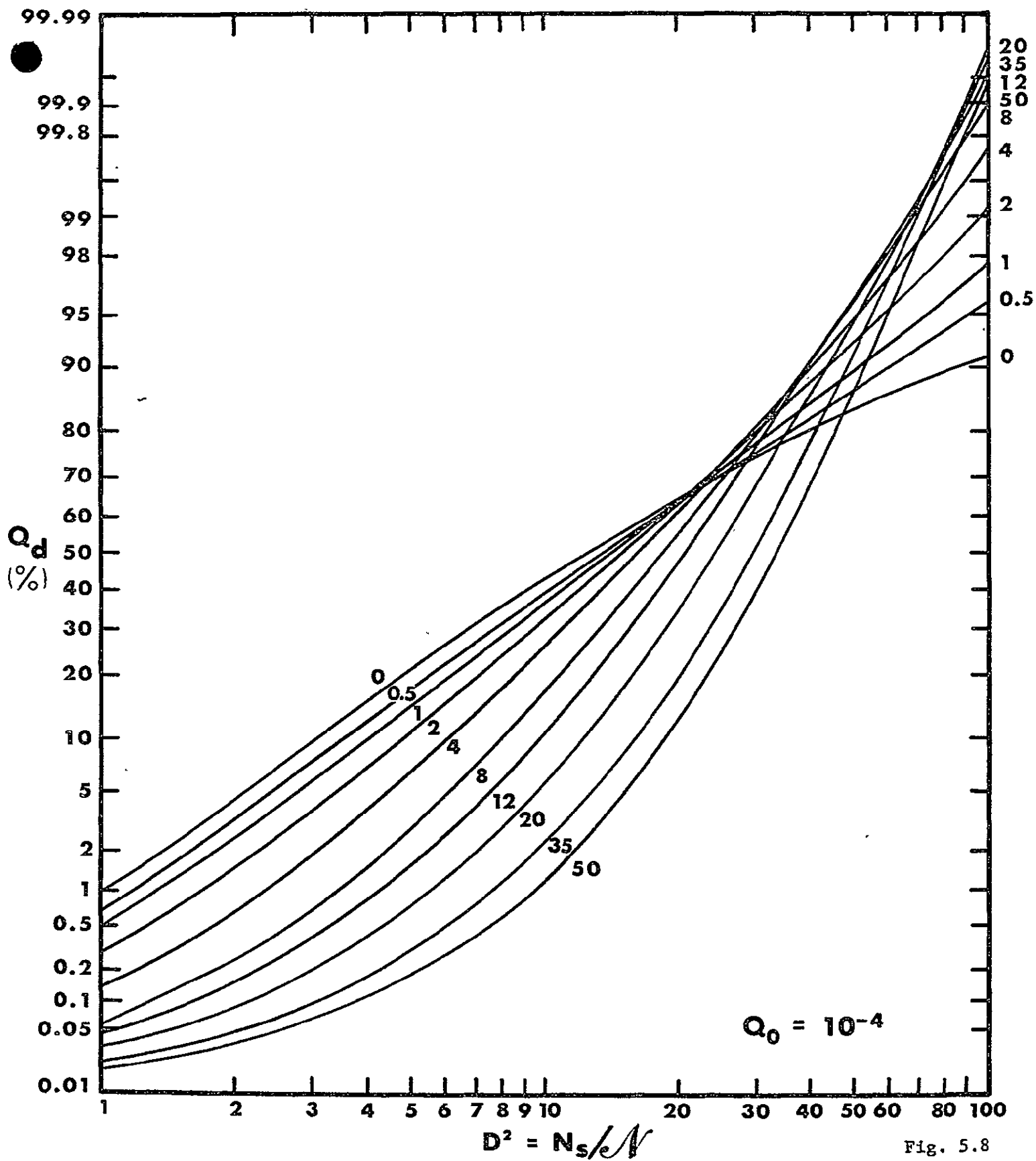


Fig. 5.7



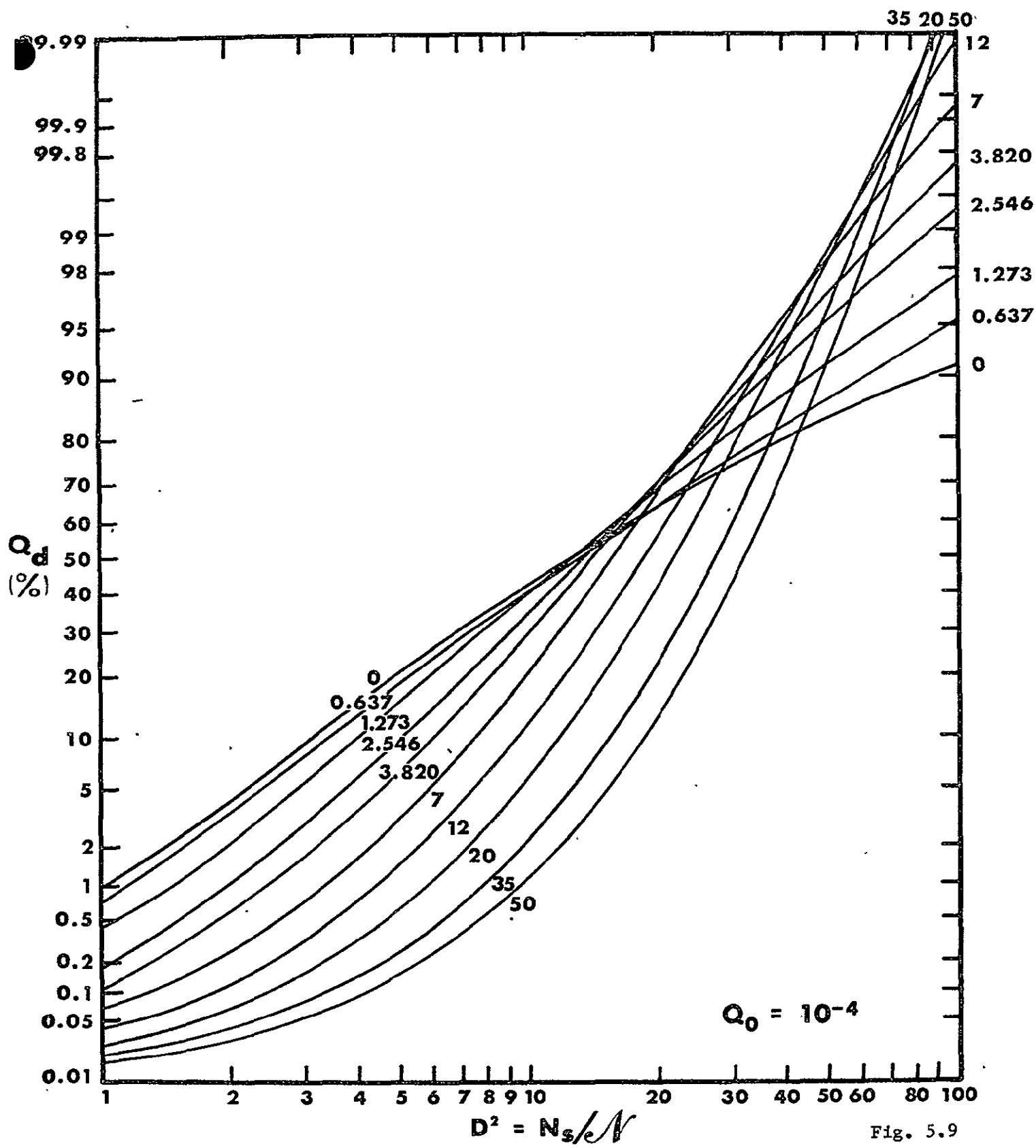


Fig. 5.9

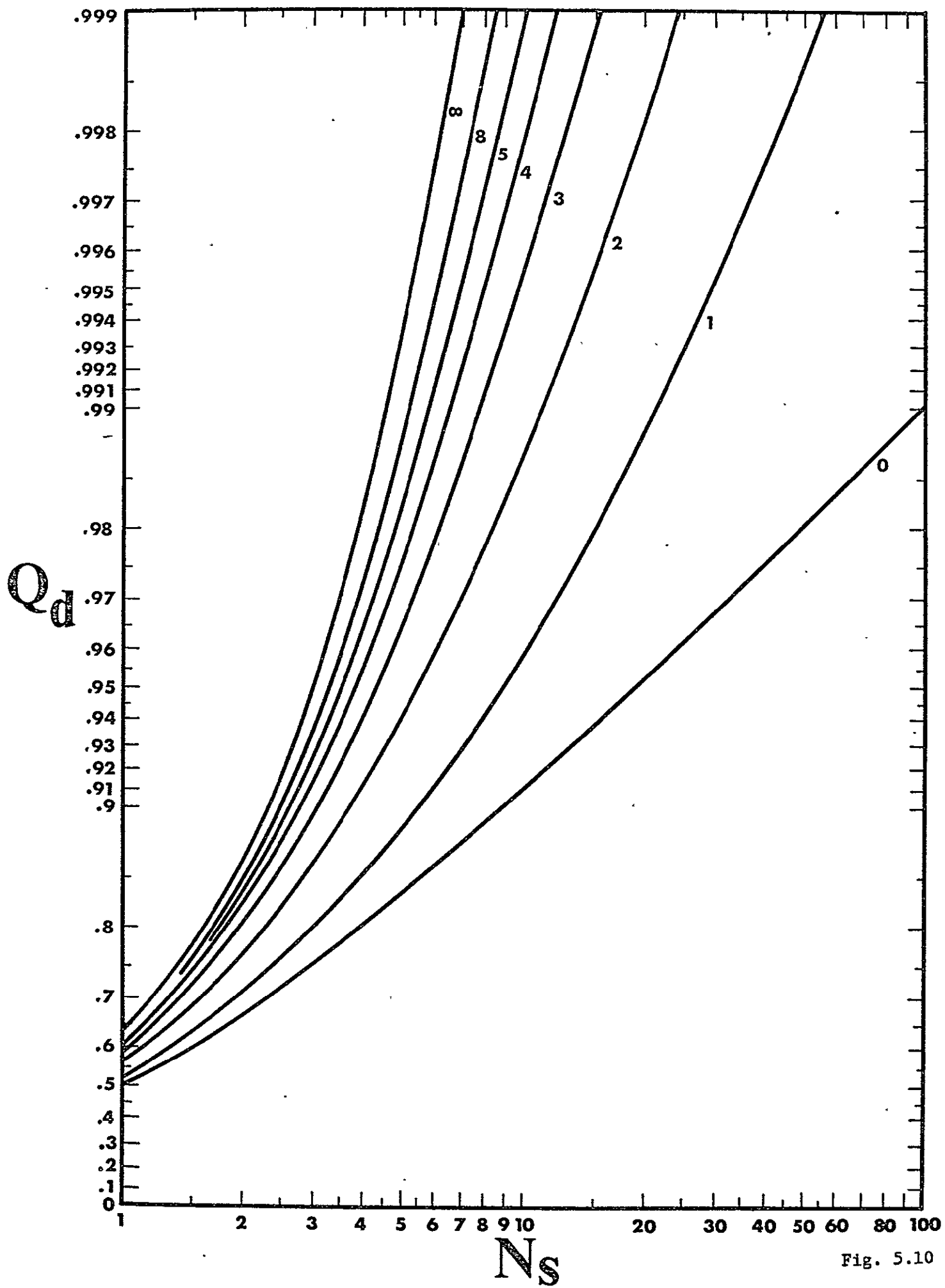


Fig. 5.10

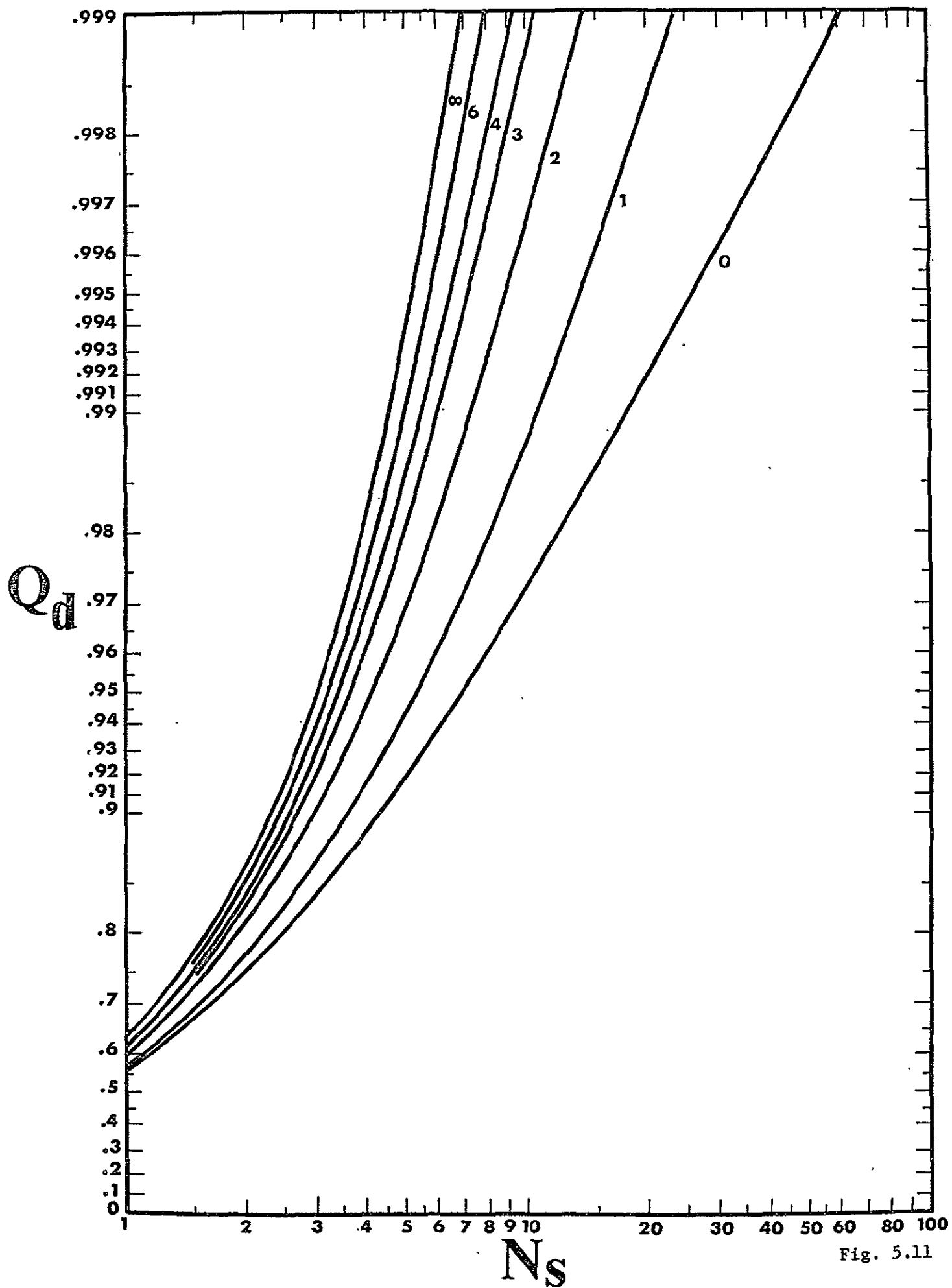


Fig. 5.11

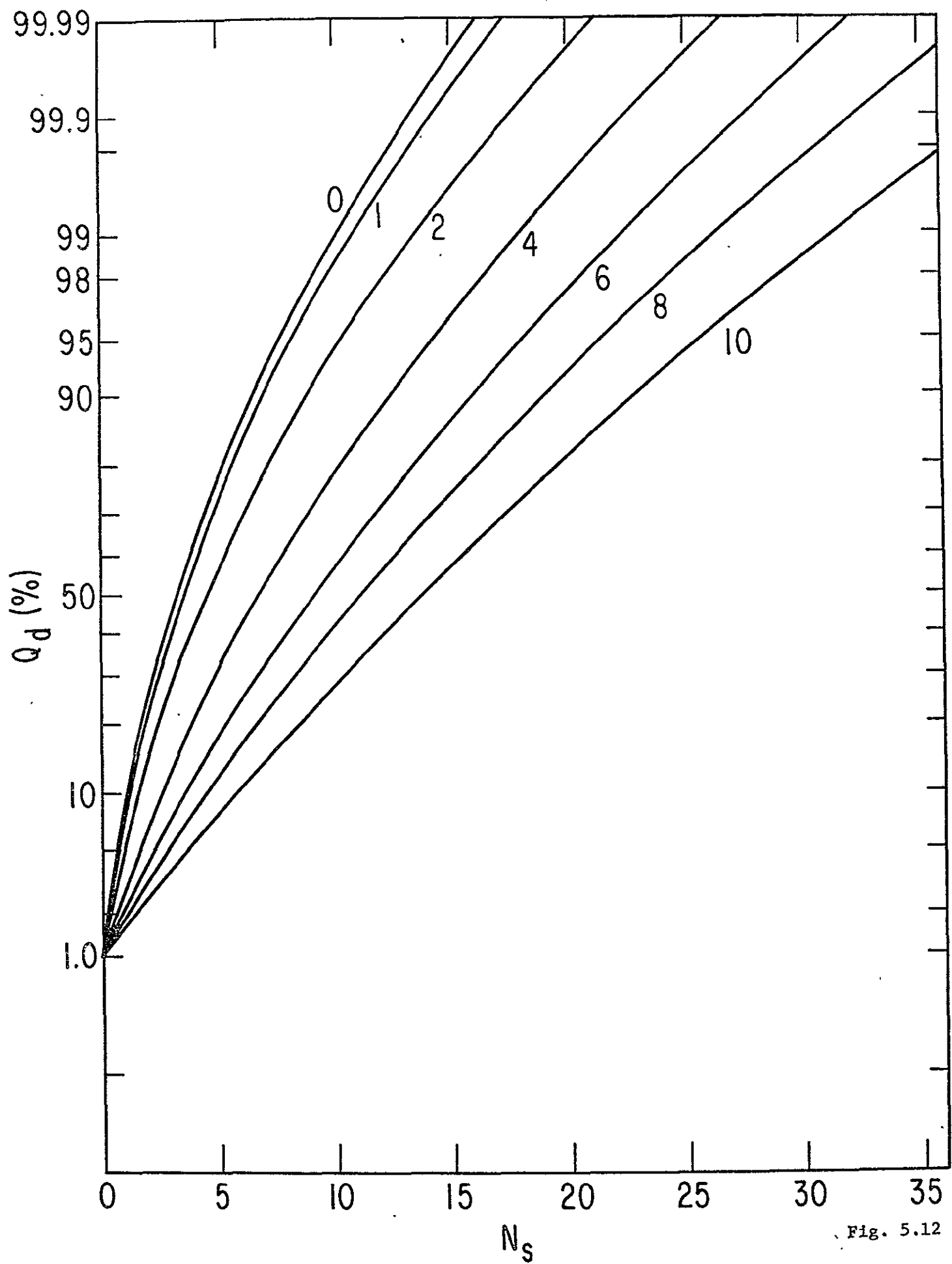


Fig. 5.12