THE UNIVERSITY OF MICHIGAN

06515-1-F

AFAL-TR-66-27

STUDY OF PROBLEM AREAS IN OPTICAL COMMUNICATIONS

Gunnar Hok
M. L. Barasch
Peter Lambropoulos
E. K. Miller

TECHNICAL REPORT AFAL-TR-66-27

February 1966

Contract No. AF 33(615)-1021
Project No. 4335
Task No. 433511

6515-1-F = RL-2136

Air Force Avionics Laboratory
Research and Technology Division
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio 45433

FOREWARD

This report was prepared by The University of Michigan, College of Engineering, Department of Electrical Engineering, The Radiation Laboratory through the Office of Research Administration, Ann Arbor, Michigan under USAF Contract No. AF 33(615)-1021, Project No. 4335, Task No. 433511, and covers work from May 1964 through September 1965. This is a final report. The work was administered under the direction of the Air Force Avionics Laboratory, Research and Technology Division, Air Force Systems Command, Frank M. Polak, Project Engineer.

Publication of this report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

Joseph A. Dombrowski
Lt. Colonel, USAF
Chief, Electronic Warfare Division

ABSTRACT

The purpose of this report is to investigate some of the ultimate performance limits predicted by theory in a few specific problem areas associated with laser communication.

First problem area: The various causes for the deterioriation of an optical signal during its passage from space to a terminal on the earth are considered, such as attenuation by the normal constituents of the atmosphere and by clouds, fog, haze and rain; seeing, scintillation, anomalous disperion and background radiation.

Second problem area: The quantum-mechanical limitations on observations of electromagnetic radiation are reviewed. The statistical theory of communication and of detection of signals is extended to quantum-limited communication channels. A photon counter is shown to be able to recover as much information from a signal as is compatible with the quantum limitations. At low signal levels it is found that binary operation can utilize a high percentage for the channel capacity. A photon counter may be a photomultiplier or a laser amplifier followed by a photomultiplier. A theory of laser amplifiers is presented. The latter type counter, which has not yet been realized, promises the additional advantage of providing discrimination against background radiation outside the channel frequency band. The means for controlling the bandwidth of such an amplifier have not yet been investigated.

Third problem area: For the purpose of providing adequate discrimination against background radiation before detection of signals by a photon counter, various optical filter principles are investigated, in particular the Wernicke prismatic filter and the polarization interference filter. The result is that desired bandwidths and tunability are theoretically obtainable but that state-of-the art limitations are as yet discouraging.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

I
INTRODUCTION: GENERAL OBJECTIVES AND PROBLEM AREAS

In the last few years considerable attention has been directed to the use of optical frequencies for long-distance communication. The continuous pressure toward higher and higher frequencies in the field of electromagnetic communication has traditionally been based in the past on the fact that bandwidth is one of the important factors affecting the capacity of a communication channel; for constant relative bandwidth the absolute bandwidth grows directly with the midband channel frequency. Another motivating circumstance has been the growing realizable antenna gain with increasing frequency for constant aperture; the approximate coherency made possible by the laser has extended the range of this advantage into the optical frequency band.

However, there are also adverse circumstances encountered in the optical range. Every communication channel involves flow of energy from a transmitting system to a receiving system; at optical frequencies the smallest resolvable energy increment or the minimum "grain size" of this flow becomes appreciable and limits severely the amount of information that can be carried by a wave of given bandwidth and power at the point of reception. Since classical field theory and network theory are not capable of providing adequate theoretical models under these conditions, it is necessary to reformulate the statistical theory of communication in terms of quantum mechanics.

This reformulation and the conclusions reached on the basis of resulting quantum theory of communication constitutes the most fundamental problem area in optical communication. The primary targets here are theoretical upper bounds for the rate of transmission of information under various specified ranges of the parameters involved, such as frequency, bandwidth, power, background radiation, etc.

Other problem areas will be concerned with the realization and implementation of channels which approach these upper bounds as closely as physical limitations and the state of the art permits. Particular areas may be defined according to the

different operational transformations which the signal processes pass through in the channel. At the transmitter terminal these include encoding, modulation, and optical processing performed by the antenna system. During their passage through space and atmosphere the signal wave may be subject to diffraction, absorption, scattering, mixing with background radiation, and Doppler frequency shift. At the receiving end there is again processing by antenna optics and finally detection and decoding, the latter two operations subject to additional noise and to the quantum measurement limitations referred to above.

The general objective of the work reported here has not been to delineate the present state of the art but to investigate the ultimate performance limits predicted by theory and possibly to suggest the most promising research strategy for bringing the state of the art rapidly closer to these limits.

The directives for the present study call for work in three specific problem areas, which will now be quoted and given a preliminary discussion.

First Problem Area: Deterioration of Optical Signals Propagating Through The Atmosphere.

> "A possibility which exists for future optical communication links is the employment of direct aerospace-to-earth or earth-to-aerospace transmission. While receiver and transmitter sites can usually be chosen for ideal visibilities, even here a finite probability exists of temporarily poor visibility conditions due to excessive rain, fog, smoke, or haze. Assuming the worst possible conditions which can occur, the possibility of information transmission during these conditions is to be determined. As an example, consider extremely dense fog conditions to exist. The existence of such a fog condition will, at best, cause considerable attenuation of the received signal. However, there is also a possibility of almost total attenuation of the signal energy and/or high information degradation due to molecular scattering and absorption. The basic question which needs to be resolved is whether situations exist where it is impossible to achieve satisfactory information transmission. Possibilities which need to be studied for satisfactory transmission should include large power output

boosts, operation at frequencies avoiding molecular absorption regions, simultaneous transmission at various different frequencies, and choice of modulation techniques. The analysis is to include effects on the coherence properties of the received signal."

This task directive calls for a comprehensive study of all the factors that may affect the power level and information content of an optical wave train between its emission from the transmitter antenna and its acceptance by the receiver antenna. Many aspects of this problem area have been investigated by the astronomers since the beginning of this science. The choice of a site of an astronomical observatory and an evaluation of its merits are very nearly the same problems as those encountered when planning an earth terminal for space communication. The difference lies primarily in the nature, particularly the bandwidth, of the observational data to be collected in the two cases. Consequently a large amount of research results obtained by the astronomers are available and relevant for the purposes of this task (See, for instance, Augason and Spinrad, 1965).

The power accepted by the receiving antenna may be written

$$P_R = P_T A_T A_R \lambda^{-2} R^{-2} \exp\left\{-\int_0^R \gamma(r)dr\right\} \qquad (1.1)$$

where P_T is the transmitted power, A_T and A_R the effective apertures of the transmitter and receiver antennas, respectively, λ the wavelength, R the distance, and $\gamma(r)$ the extinction coefficient at the distance r from the transmitter.

The order of magnitude of R for deep-space communication may vary from a few hundred miles at perigee in an earthbound orbit to interplanetary distances of 10^8 miles or more. The wavelength range specified for this study extends from 0.4 microns in the visible spectrum (violet) to 20 microns in the far infrared. Modulation bandwidths of 1000 megacycles are to be anticipated.

The extinction coefficient $\gamma(x)$ is the combined result of a number of different processes in the atmosphere which partially absorb or scatter the radiation during its passage . The normal constituents of the atmosphere have numerous molecular bands in the infrared part of the spectrum; drops or particles forming haze, fog,

clouds, rain and snow scatter the radiation to an extent depending on the ratio of particle size to wavelength as well as on the density and electromagnetic properties of the particles. There are certainly conditions under which complete opacity is encountered; the maintenance of uninterrupted communication depends ultimately on the choice of sites or the existence of alternative sites and facilities as far as the earth-based antenna is concerned.

The received signal level and its low-frequency modulation are affected by the phenomena of "seeing" and scintillation, associated with turbulence in the lower and higher strata, respectively. The former may be minimized by choice of site; design of an appropriate type of modulation may reduce the interference due to the fluctuations introduced by the latter. The theory of communication channels with randomly varying properties is not yet very far advanced; for that reason we have not been able to include a thorough study of these phenomena.

The modulation of an optical signal may be affected by dispersion in the atmosphere which occurs on the edges of absorption bands. The remedy is obvious: use only wavelengths well within the atmospheric "windows".

The question to what extent the coherence of a laser signal is affected by the passage through the atmosphere requires further study. We have included a brief discussion of this subject.

The background radiation appearing as random noise in the optical channel constitutes another limitation, particularly at longer wavelengths. In addition to the obvious contributions from the galaxy, stars and planets, an appreciable component is added by the atmosphere and by sunlight scattered in the atmosphere and by interplanetary matter. High resolution of the receiver in direction as well as in wavelength reduces this source of noise to its minimum.

Second Problem Area: Choice of Detection System for Optimum Rate of Transmission of Information.

"A second problem to be considered is a study of the desirability of using binary quantum counters as detectors in the ideal laser communication system. Gordon (Proc. IRE, 50, 1898 (Sept. '62)) shows that the binary quantum counter has high

information efficiency when the total received power is less than one quantum per sample. Consequently, the basic question here concerns the utility consistent with the basic tenets of the communication system of operating at information receipt of less than one quantum per sample, where the communication system is subject only to the ultimate theoretical limitations discussed above. Consideration shall also be given to coding techniques allowing minimum information degradation. "

The questions raised in these directives reach the fundamental limitations on physical measurements. The electromagnetic field in the laser beam, reduced in energy density by diffraction, absorption and scattering and possibly affected by background radiation, "seeing", scintillation and anomalous dispersion, carries a limited amount of observable information per unit time. The upper bounds for the rate at which information is recovered at the receiving end of the channel depends also on the kind of detectors and amplifiers used; it is the need for an investigation of these particular relationships which are emphasized in the above directives.

In Chapter III we begin such an investigation by a brief review of the quantum theory of the electromagnetic field. From the minimum-uncertainty specification of the field in terms of energy eigenfunctions, its maximum entropy, i.e. its maximum information content, is obtained. A more recently developed description of the field in "coherent states" is introduced, since it has certain advantages for the subsequent work concerning interaction of radiation with matter in the detection processes.

The maximum entropy or "channel capacity" of an optical wave train of given power and frequency band is discussed and compared with the classical results for microwaves and lower frequencies. If the background radiation is negligible, the entropy is limited by the Heisenberg uncertainty or (equivalently) by the "zero-point energy" or "vacuum fluctuations" associated with each mode of the field. This entropy is calculated from an expansion in energy eigenstates as basis functions, since these form a complete set of mutually exclusive events; there is no implication, however, that the information must be recovered from the wave train by means of energy measurements.

In connection with optical channel capacity we shall discuss the concept of "optimum bandwidth". Since the channel capacity is a monotonic function of the bandwidth, there is no finite bandwidth that maximizes the channel capacity. But beyond a certain point, which will be roughly estimated, the increase in channel capacity is so small that it is by far outweighed by the growing error rate and rapidly increasing complexity of terminal equipment for processing of the large number of samples in encoder and decoder. At a background temperature of $300^{o}K$ and a wavelength of 20μ, this marginal bandwidth is about three times the average number of photons entering the receiving antenna per second, giving an average signal level of 0.3 photons per sample pair. At a wavelength of 0.4μ, the corresponding figures are 10^{22} and 10^{-22}, respectively, which appear to be too extreme to have any practical interest; the scattered sunlight in the atmosphere provides a much larger background spectral density than equilibrium density at $300^{o}K$ leading to a narrower bandwidth and a larger fraction of a photon per sample.

Pulse modulation techniques have been very highly developed for classical communication systems in recent years. Because of the different statistical properties of quantum-limited channels, the application of such techniques to optical channels requires a new performance analysis. The binary pulse code suggested in the directives quoted above is conveniently evaluated by means of statistical decision theory. For this reason the statistical theory of signal detection is in Section 3.4 extended to quantum-limited channels.

One significant difference between the classical background-limited binary channel and the quantum-limited binary channel is the extremely small "false-alarm" or "false-count" probability in the latter at low signal levels. It is consequently possible to design the signal so as to minimize the "miss" probability alone. This leads to a highly asymmetric operation, using high-energy pulses separated by sufficiently long intervals to satisfy the average-power limitation of the transmitter.

There are two serious state-of-the-art problems involved in the realization of such a channel. In the first place, subsequent receiver components have to be

found that do not appreciably raise the false-alarm probability. In the second place, efficient codes have to be found for such highly asymmetric channels.

In Section 3.5 the coding problem is considered. The recently proposed permutation principle in coding lends itself well to the coding of asymmetric channels. A numerical example involving a simple error-correcting code of this kind is given.

In order to give the discussion of the photon-counter channel a good perspective, comparisons with other methods of detection are given. The superheterodyne method is briefly covered in connection with the statistical detection theory in Section 3.4. A laser amplifier as a receiver component is analyzed in Section 3.6. Such an amplifier serves two purposes: to raise the signal level before detection or "counting," and to discriminate against background radiation and possible extraneous signals outside the modulation band of the channel, i.e. to satisfy the need for a narrow optical filter at the input to the receiver. Both purposes are considered in this section.

This laser study was undertaken by a member of the group with a very strong background in quantum mechanics and laser theory. No attempt has been made to coordinate in detail the notation and presentation in this section with the rest of this report. It remains a more or less self-contained contribution to the illumination of an important subject in the second problem area of this report.

The spontaneous-emission noise in the amplifier raises the false-alarm probability and thus changes the output statistics in a Gaussian direction. The unique discrete low-level behavior of the photon counter is lost in this method as well as in the superheterodyne method. The filter properties of a high-gain laser amplifier appear promising; however, it cannot at the present time be stated with certainty that as large bandwidths as 1 Kmc can be achieved with a laser amplifier.

Third Problem Area: Optical Bandpass Filters for Communication Systems.

"The third area to be considered is the determination of the theoretical limitations which establish the narrowness and the transmissivity of bandpass filters, tunable and untunable, within the spectrum specified.

Interference filters are of special interest. The limitations to be considered are only those of the filters themselves and not those due to external system considerations such as Doppler. Recommendations for future and unusual filter research which may approach the established limitations are also of interest."

Motivation for interest in this problem area is derived from the second problem area: In a binary channel employing a simple photon counter as a detector, the reduction of interference from background radiation and extraneous signals by frequency filters must take place at optical frequencies.

Bandpass filters at optical frequencies have the drawback that their geometrical dimensions must be taken very large compared to the wavelength of the radiation. This fact alone largely eliminates the flexibility and easy control of filter parameters over wide ranges which characterizes filter design at conventional communication frequencies.

The filter design problem is further complicated by the tunability requirement, i.e., the desirability that the center of the pass band be adjustable to account for variable conditions, such as the Doppler shift due to relative motion of transmitter and receiver.

A filter which absorbs radiation emits under equilbrium conditions an equal amount of radiation; it is consequently necessary to maintain the filter temperature far below the temperature of the source of background radiation in order to improve the signal to noise ratio. In the visible range, where the background is primarily scattered light from the sun or moon, this condition is usually automatically satisfied, but in the infrared local thermal radiation may become important, and refrigeration of filter as well as of the detector itself may substantially improve the signal-to-noise ratio.

In Chapter IV existing types of optical filters are described and analyzed. There do not appear to be any strict theoretical limitations for the performance of filters operating on the interference principle. However, practical and state-of-the-art considerations suggest that filters – in particular tunable filters – suitable for

optical photon-counter space communication will not be designed for some time to come. The narrow bandwidth and high stop-band attenuation required leads to considerable bulk and very large pass-band attenuation with presently available materials. The former may be tolerated at the ground terminal, but the latter is prohibitive for channels expected to operate with a fraction of a photon per sample signal level at the receiver antenna.

II
FIRST PROBLEM AREA: DETERIORATION OF OPTICAL SIGNALS PROPAGATING THROUGH THE ATMOSPHERE

2.1 INTRODUCTION

After defining the first problem area in the preceding chapter, a brief outline was given of the purpose and contents of this chapter. The first group of phenomena to be considered are various extinction processes in the atmosphere: scattering by clouds, rain, molecular and aerosol particles, and absorption by molecular bands. Furthermore, optical signals are under some circumstances subject to deterioration by "seeing", scintillation and anomalous dispersion. We shall also consider the possible loss of coherence of the signal waves during passage through the atmosphere. The background radiation, its sources and magnitudes, are finally the last subject considered in this chapter.

2.2 EXTINCTION PROCESSES IN THE ATMOSPHERE

2.2.1 Attenuation by Clouds

Clouds are mixtures of water droplets, ice crystals, and some water vapor. There is much variation between clouds, and in the course of evolution of any individual cloud. For this reason, it is felt to be of more value to cite here the results of experiments performed by Gates and Shaw (1960) under a wide variety of cloud-cover conditions than to refer to calculations such as those of Deirmendjian (1964).

A cumulus cloud model was introduced by Deirmendjian (1964) in which the attenuation per unit path corresponding to the liquid water droplets only is reported to be essentially frequency independent for wavelengths between 0.45μ and 16.6μ and has an average value

$$\gamma_{\text{cumulus}} \sim 17\,\text{Km}^{-1} \quad .$$

Since cloud thickness is said to be in the range 230-2100 m, the one-way transmissivity would be between exp (-4) and exp(-34), or roughly between 0 and 2 percent. Models for other types of clouds, and the effect of water vapor were not considered.

The extension of exact computations to longer wavelengths may be difficult, since values of the index of refraction of water at these wavelengths, which would be required, do not seem to be available.

The experimental data presented in Gates and Shaw (1960) were obtained in a spectral region lying between 0.1μ and 15μ for transmission through a large variety of mixtures of cirrus, cumulus, and nimbus cloud covers. Near 0.5μ, the transmissivities varied between roughly 20 percent and 95 percent. Above 10μ, they ranged from 30 percent up.

The conclusion to be drawn from these data is that clouds may present significant attenuation, but not total opacity.

2.2.2 Attenuation by Fog and Haze

For fogs no satisfactory theoretical values of the attenuation coefficient have been located in the literature, although some indication of parameters for a fog has been obtained, namely a water content of 10^{-7} g/cm with radius of droplets distributed according to an inverse square law, $n(r) = Cr^{-2}$. A calculation has been attempted, but remains incomplete. It has not been possible to construct a satisfactory functional form for an approximation to the Mie efficiency factor for extinction which will be valid over the whole range of integration over droplet size, for the index of refraction corresponding to the frequency band of interest, or that part for which the index is avaiable. The theoretical approach is as follows:

Fogs are assumed to consist of a distribution of water droplets whose radii r are taken to obey the law suggested by Kurnick, Zitter, and Williams (1960):

$$\begin{aligned} n(r) &= Cr^{-2} && r_o < r < r_1 \\ &= 0 && r < r_o \quad \text{or} \quad r > r_1 \end{aligned} \tag{2.1}$$

with a total water content w of 10^{-7} g cm^{-3}. The constant C may be related to w since

$$w = \int_{r_o}^{r_1} 1\,g\,cm^{-3}\, n(r)\, \frac{4}{3}\pi r^3\, dr$$

$$\sim \frac{4}{3}\pi C \cdot g\,cm^{-3}\, \frac{r_1^2}{2} = \frac{2}{3}\pi C r_1^2\, g\,cm^{-3}$$

(since $r_1 >> r_o$)

$$w \sim 10^{-7}\, g\,cm^{-3} \tag{2.2}$$

whence

$$C \sim \frac{3}{2} \cdot 10^{-7}\, r_1^2 \tag{2.3}$$

Now, where $Q(x,\lambda)$ is the extinction area-ratio function derived from electromagnetic theory, the attenuation coefficient β is defined as

$$\beta_{fog}(\lambda) \sim \int Q(r,\lambda)\,\pi r^2\, n(r)\, dr$$

$$= C\pi \int_{r_o}^{r_1} Q(r,\lambda)\, dr \tag{2.4}$$

$$= C\pi\lambda \int_{a_o}^{a_1} Q(a,\lambda)\, da, \tag{2.5}$$

where the variable a, or r/λ, or $kr/2\pi$, is convenient. The limits of integration are thus functions of λ. Q is a function of λ through the index of refraction as well as of kr. The index m of refraction has been obtained by many workers, and is tabulated by Deirmendjian (1964):

TABLE II-1. REFRACTIVE INDEX OF WATER IN THE INFRARED

λ (microns)	Re(m)	Im(m)
0.45	1.34	0
0.70	1.33	0
1.61	1.315	0
2.25	1.29	0
3.07	1.525	0.0682
3.90	1.353	0.0059
5.30	1.315	0.0143
6.05	1.315	0.1370
8.15	1.29	0.0472
10.0	1.212	0.0601
11.5	1.111	0.1831
16.6	1.44	0.4000

We have then,

$$\beta_{fog}(\lambda) = 1.5 \cdot 10^{-7} r_1^{-2} \lambda I_1(\lambda) ,$$

$$I_1(\lambda) = \int_{a_o}^{a_1} Q[a, m(\lambda)]\, da , \qquad (2.6)$$

$$a = r/\lambda \qquad 20\mu \leq r_1 \leq 50\mu, \; r_o \leq 10^{-1}\mu$$

and the computational problem is essentially to obtain $I_1(\lambda)$. This has not been carried out, although an approach has been devised, as described below.

An expression is needed for Q which holds for complex values of m. This may be obtained by use of the "forward scattering theorem" to relate Q to the "scattering amplitude" $S(m, \theta, 0)$. This latter quantity may be obtained by analytical continuation to complex m of an empirical expression valid for real m, which is given by van de Hulst (1957).

Haze is an aerosol of solid particles suspended in the atmosphere. A size distribution model commonly employed is that proposed by Junge (1963). Experiments on infrared attenuation reported by Kurnick et al (1960) seem to fit the predictions based on this model, and furthermore a direct count of the particles confirmed the distribution. Based on this evidence

$$n(r) = Cr^{-P}, \quad r_1 < r < r_o \tag{2.7}$$

$$= 0 \quad \text{otherwise}$$

$$\text{with } r_o = 20 \text{ or } 50\mu$$

$$r_1 \leqslant 0.1\mu$$

and P is a function of r rather than a constant, but roughly $2 < P < 4.5$, with the larger values typifying haze and $P = 2$ for water fogs at 100 percent relative humidity. C will depend on the mass of aerosol present. No information on values has been obtained.

With the distribution (2.7), one has, according to Kurnick, Zitter, and Williams (1960), for the attenuation by scattering in transversing path length L, $\exp\left[-\sigma_{sc} L\right]$, the values given by

$$\sigma_{sc} = C'\lambda^{3-P}, \tag{2.8}$$

as long as the condition is satisfied that

$$r_1 \leqslant \lambda . \tag{2.9}$$

Otherwise, a more complicated dependence of σ on λ results. We might investigate the approximate form valid near the limit but exact theoretical treatment is not justified in view of the uncertainty of the meteorological parameters.

The experiments of Kurnick, Zitter and Williams, with L = 200 yards, indicate values of σL consistently near unity for $\lambda = 1\mu$, and decreasing to

about 0.2 as λ increases to 10μ.

The situation with regard to comparison of theory and experiment, or applicability of experimental results as "typical," is unsatisfactory here as for clouds. Therefore, it must suffice to cite the few numbers reported by Curcio, Knestrick, and Cosden (1961), and Kurnick, Zitter, and Williams (1960). Fogs in the Chesapeake Bay region yielded attenuation coefficients (in the spectral region 0.30μ-3μ) which were usually between $0.02\,Km^{-1}$ and $2\,Km^{-1}$. This wide a variation does not permit drawing a practical conclusion. A series of measurements which were made in Chicago on days of fog and generally poor weather, during the course of formation of fog, for the band $1 - 10\mu$ showed the optical density σL to be between 0.1 and 1.0 typically, where the path L was 200 yards. In other words, the scattering attenuation coefficient σ varied between $0.5\,Km^{-1}$ and $5\,Km^{-1}$. Again, no practical conclusion is possible. In both references, the curves reported do exhibit better transmission for the large wavelength end of the band. The latter measurements do not exceed attenuations of $0.4\,Km^{-1}$ for $\lambda > 2\mu$, nor do the former. It appears, then, that this end of the band is preferable for transmissions through fog and haze.

Deirmendjian (1964) reports calculations based on integration of the exact Mie solution over assumed droplet size distributions for fogs, incorporating the frequency dependence of the dielectric constant. The computed attenuation coefficients decrease from roughly $1\,Km^{-1}$ near $.5\mu$ to slightly smaller values near 10μ, in agreement with the trend obtained by experiments. Any conclusion which is to serve as a foundation for important decision must await further experimental investigations which may further clarify the situation.

The model of a clear standard atmosphere published in the U.S. Standard Atmosphere (1962) permits a haze component and is discussed by Elterman (1963). Attenuation of infrared radiation by that aerosol distribution was computed by Elterman, and the attenuation coefficient for wavelengths below 0.5μ was found to be $0.20\,Km^{-1}$ at sea level. For the band 0.5μ - 5μ, an approximately linear

decrease to a value of 0.05 Km^{-1} is anticipated on the basis of experimental results obtained by Curcio and Durbin (1959), Knestrick, Cosden, and Curcio (1961) and Curcio, Knestrick, and Cosden (1961).

Deirmendjian (1964) has computed the attenuation coefficients for continental and coastal types of haze. These results are tabulated below.

TABLE II-2. ATTENUATION BY HAZES (Deirmendjian, 1964)

Wavelength (microns)	$\beta_{\text{continental Haze}}$ (Km^{-1})	$\beta_{\text{coastal Haze}}$ (Km^{-1})
0.45	0.1206	0.1056
0.70	0.0759	0.1055
1.61	0.0312	0.0691
2.25	0.0194	0.0424
3.07	0.0269	0.0602
3.90	0.0128	0.0236
5.30	0.0075	0.0112
6.05	0.0129	0.0189
8.15	0.0050	0.0062
10.0	0.0032	0.0045
11.5	0.0064	0.0097
16.6	0.0082	0.0134

2.2.3 Attenuation by Rain

A distribution function for the droplet radius r in a "typical" rainstorm (10^3 droplets per $meter^3$ with modal radius r_c of 5×10^{-2} mm) has been suggested by Diermendjian (1963). Ideally, the attenuation coefficient would be computed by integrating over this distribution the extinction cross-section. This latter quantity could in theory be derived from the Mie series or one of approximations to it, such as those discussed in van de Hulst (1957). However, no approximation valid for rain drops of the sizes assumed by the distribution function (radii of the order of a few mm) at micron wavelengths, and convenient for accurate calculation, has been discovered. For this reason, we have observed that the droplets are large enough

electrically at all wavelengths we consider to insure the approximate validity of the simple assumption that the extinction cross-section is given by

$$\sigma_E = 2\pi r^2 \quad . \tag{2.10}$$

The "generalized gamma distribution" proposed by Deirmendjian (1963) for a moderate rainstorm reads, for a concentration of 10^3 per m^3 and $r_c = 5 \times 10^{-2}$ mm,

$$\frac{dn}{dr} = 5.333 \times 10^5 \, r \exp\left[-8.944\sqrt{r}\right] m^3 \, mm^{-1} \quad . \tag{2.11}$$

It is clear on further analysis that r here is measured in mm. This distribution may be expressed in terms of the normalized radius $x = r/r_c$ as follows:

$$dn = \frac{16}{6} \times 500 x \exp\left[-2\sqrt{x}\right] dx \text{ meter}^{-3} \tag{2.12}$$

By utilizing (2.10), one is enabled to form the attenuation coefficient

$$\beta = \frac{8}{3} \pi \times 10^3 \, r_c^2 \int x^3 \exp\left[-2\sqrt{x}\right] dx \text{ meter}^{-1} \tag{2.13}$$

$$\sim \frac{2}{3} \pi \; 10^{-2} \text{km}^{-1} \int_0^\infty x^3 \exp\left[-2\sqrt{x}\right] dx = \frac{2}{3} \pi \, 10^{-2} \frac{\Gamma(6)}{2^7} \text{km}^{-1} \tag{2.14}$$

$$\sim 0.2625 \, \pi \text{ km}^{-1} \sim 0.795 \text{ km}^{-1} \quad . \tag{2.15}$$

This attenuation is the equivalent of only 40 percent transmission in 1 km through a moderate rainstorm, for all frequencies in the band we consider. In the absence of parameters for droplet distribution in rains described as other than "moderate", no statement about them can be made.

2.2.4 Rayleigh Molecular Scattering

Another mechanism attenuating infrared radiation even in clear weather is the scattering by molecules of the atmospheric gases. Since the molecules are

much smaller than the wavelength, this scattering is referred to as Rayleigh scattering. A calculation of this effect employing molecular abundances based on the U. S. Standard Atmosphere appears in Elterman (1963). The attenuation at 0.55μ and sea level is roughly $10^{-2}\,\mathrm{Km}^{-1}$. It decreases rapidly with altitude as abundance falls. For larger wavelengths, the attenuation is orders of magnitude smaller. As Elterman states, it is consistent to associate with the "clear standard atmosphere" an aerosol component, forming the "turbid atmosphere." He computes attenuation by this aerosol in the infrared.

2.2.5 Rayleigh and Aerosol Attenuation

These two effects have been combined by Elterman (1963) into an optical thickness for the turbid atmosphere, $\tau_t^{\infty}(h)$. In terms of this quantity, the slant path transmittance at zenith angle θ from altitude h to the edge of the atmosphere, as reduced by Rayleigh scattering and clear atmosphere aerosol attenuation, is given by

$$T_{h-\infty} = \exp\left[-\tau_t^{\infty}(h)\sec\theta\right] . \qquad (2.16)$$

Elterman tabulates this quantity $\tau_t^{\infty}(h)$ for wavelengths between 0.4μ and 4μ; it should not exceed the values computed for 4μ if larger wavelengths are employed. Since the transmittance of a 4μ wavelength from ground to aerospace at angles only 5° above horizontal may be calculated to exceed 90 percent (from Elterman's tabulated value of $t_t^{\infty}(0) = 0.0489$ at 4μ), there appears to be no reason to consider the mechanism for attenuation for $\lambda > 4\mu$. For smaller wavelengths, it can be significant at near-horizontal angles for transmision from ground or low altitudes, and the tabulations are accordingly reproduced here for λ between 0.4μ and 1.67μ.

TABLE II-3. TURBID OPTICAL THICKNESS

h(km)	τ_t^{∞}(h) 0.4μ	τ_t^{∞}(h) 0.5μ	τ_t^{∞}(h) 0.6μ	τ_t^{∞}(h) 0.7μ	τ_t^{∞}(h) 0.9μ	τ_t^{∞}(h) 1.67μ
0	0.4977	0.2661	0.1813	0.1394	0.1057	0.0771
1	0.3707	0.1771	0.1079	0.0761	0.0514	0.0338
2	0.2973	0.1312	0.0727	0.0468	0.0273	0.0151
3	0.2494	0.1051	0.0546	0.0328	0.0165	0.0071
4	0.2134	0.0877	0.0438	0.0250	0.0112	0.0035
5	0.1843	0.0747	0.0364	0.0202	0.0083	0.0019
6	0.1600	0.0645	0.0312	0.0171	0.0068	0.0013
7	0.1389	0.0559	0.0270	0.0148	0.0058	0.0010
8	0.0893	0.0484	0.0234	0.0128	0.0050	0.0009
9	0.0765	0.0418	0.0202	0.0111	0.0044	0.0008

2.2.6 Absorption by Molecular Bands

The simple Lambert-Beer exponential law of attenuation is useful only for monochromatic radiation and homogeneous propagation paths (i.e. paths of constant atmospheric composition, pressure, density and temperature). For other situations it has been necessary, as many authors have realized recently, to develop more convenient formulations in order to predict attenuation. Some of these approaches are discussed briefly here, and the extent to which they promise to be applicable to situations of interest is indicated.

It has proven fruitful to introduce the 'band model' and to deal with averaged absorptances over a spectral interval larger than the width of a molecular absorption band, thus eliminating the computation difficulties associated with the rapidly varying absorption coefficient derived from considering individual lines of the band. In conjunction with the band model approach, a method of introducing the variation of atmospheric properties over the propagation path is needed. Plass (1962) has derived an 'equivalent path' concept. Using this concept and band parameters arising from the theory of molecular spectra, he computed and tabulated a set of transmittances. Since no intermediate steps in the computations are reproduced or tabulated in the referenced reports, these computations cannot easily be extended to

situations other than those for which his results are tabulated. They deal with the effects of CO_2 bands in the wave number range 500 - 10,000 cm^{-1} and of water vapor bands (using both wet and dry stratospheric distributions) in the range 1000-10,000 cm^{-1}. Transmittances (averaged over 50 cm^{-1} regions for slant paths originating at altitudes of 15, 25, 30 and 50 Km and traversing the entire atmospheric thickness at initial angles with the horizontal from 0^o to 90^o have been tabulated. The tabular values have been graphed and are presented in Figs. 2.1 - 2.13, so that implications for this study may be seen.

First of all, the conditions of relatively low transmittance can be seen to occur in isolated spectral regions, such as 550-750 cm^{-1}, 2200-2500 cm^{-1}, 3400-3800 cm^{-1}, 1450-1750 cm^{-1}. At the centers of these regions, essentially complete opacity may be encountered for near-horizontal operation from an initial altitude of 15 Km. One would conclude from Figs. 2.8, 2.9, 2.10, 2.11 and 2.12, on which curves of a given spectral region for the same elevation angle but different transmitter altitudes are displayed, that there is improvement in transmittance, but there is still some loss, even for altitudes like 25 and 30 Km. This loss might present practical difficulties. Transmittance values typically fall below 50 percent near the band centers. Although near-vertical transmittance is more dependable than low-angle (see Fig. 2.7 for an indication), this is a fact of little significance unless the trajectory of the object with which communication is desired can be controlled.

An example of averaging, taken from a table of Plass, in which the bandwidth is 50 cm^{-1}, or 1500 Kmc, is presented in Fig. 2.4. It appears that the transmittance is not improved at the band center, but the region of poor transmittance is widened, and is less sharply bounded than for the non-averaged curves. These conclusions are to be expected from the nature of averaging with bandwidth comparable to the width of the absorption region.

It should be pointed out that any wavelength region where the absorption varies rapidly with wavelength is entirely unsuitable for communication since dispersion will severely distort the modulation of the signal.

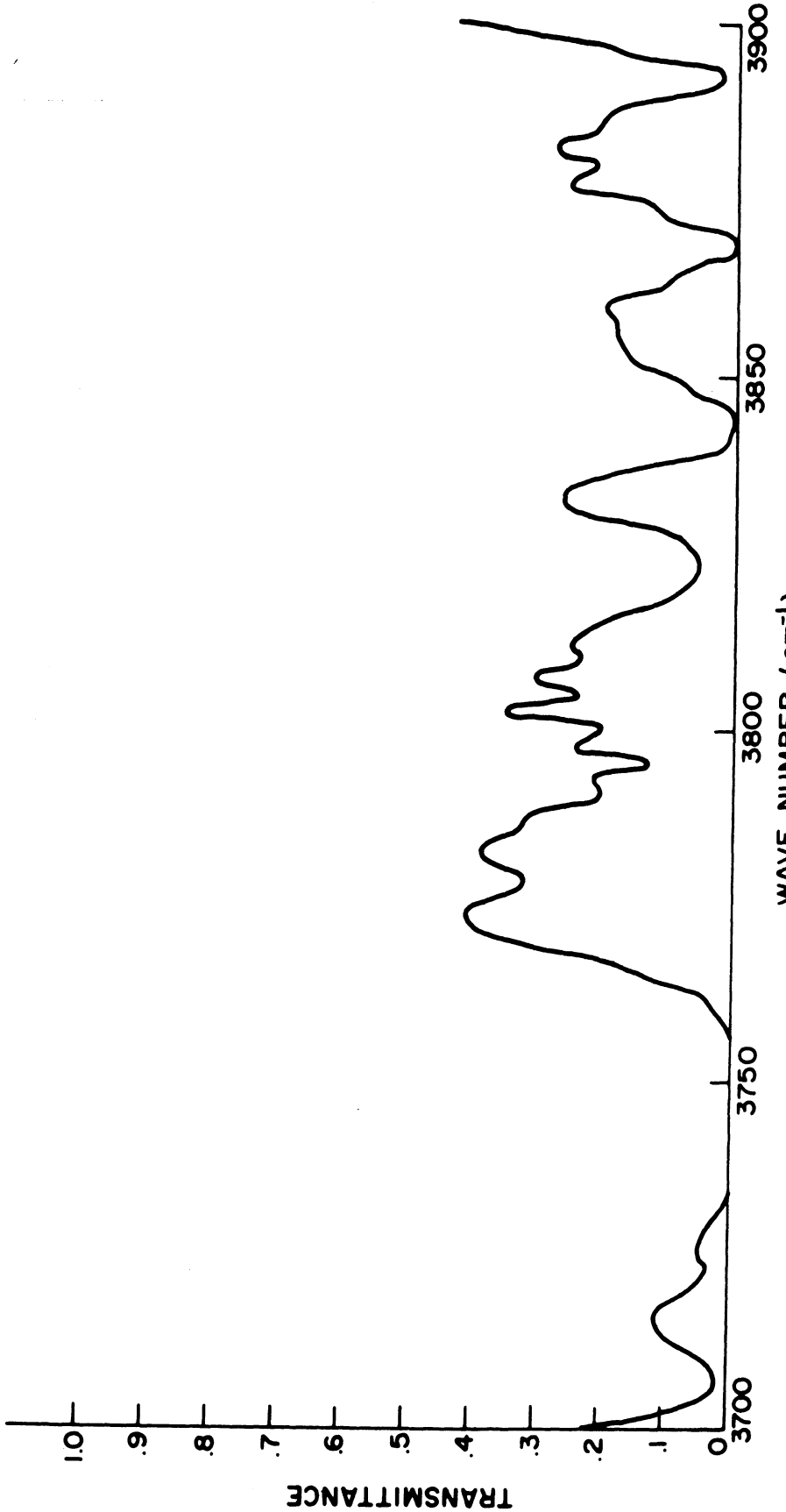

FIG. 2.1. ATMOSPHERIC TRANSMITTANCE CONSIDERING WATER VAPOR, INITIAL ALTITUDE 15 KM, 0° INCLINATION TO HORIZONTAL, DRY STRATOSPHERE MODEL.

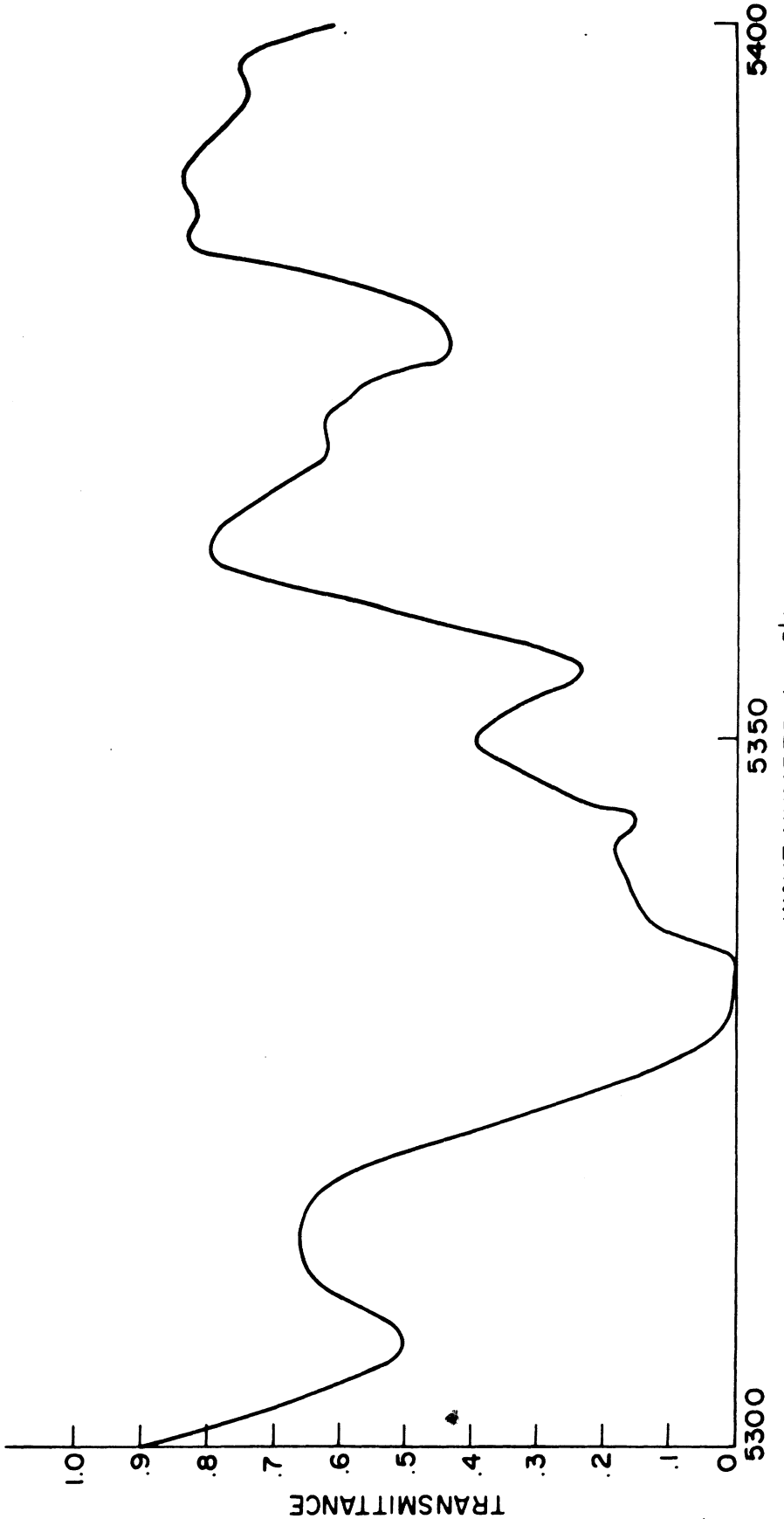

FIG. 2.2. ATMOSPHERIC TRANSMITTANCE CONSIDERING WATER VAPOR, DRY STRATOSPHERE MODEL, 15 KM INITIAL ALTITUDE, 0° INCLINATION TO HORIZONTAL.

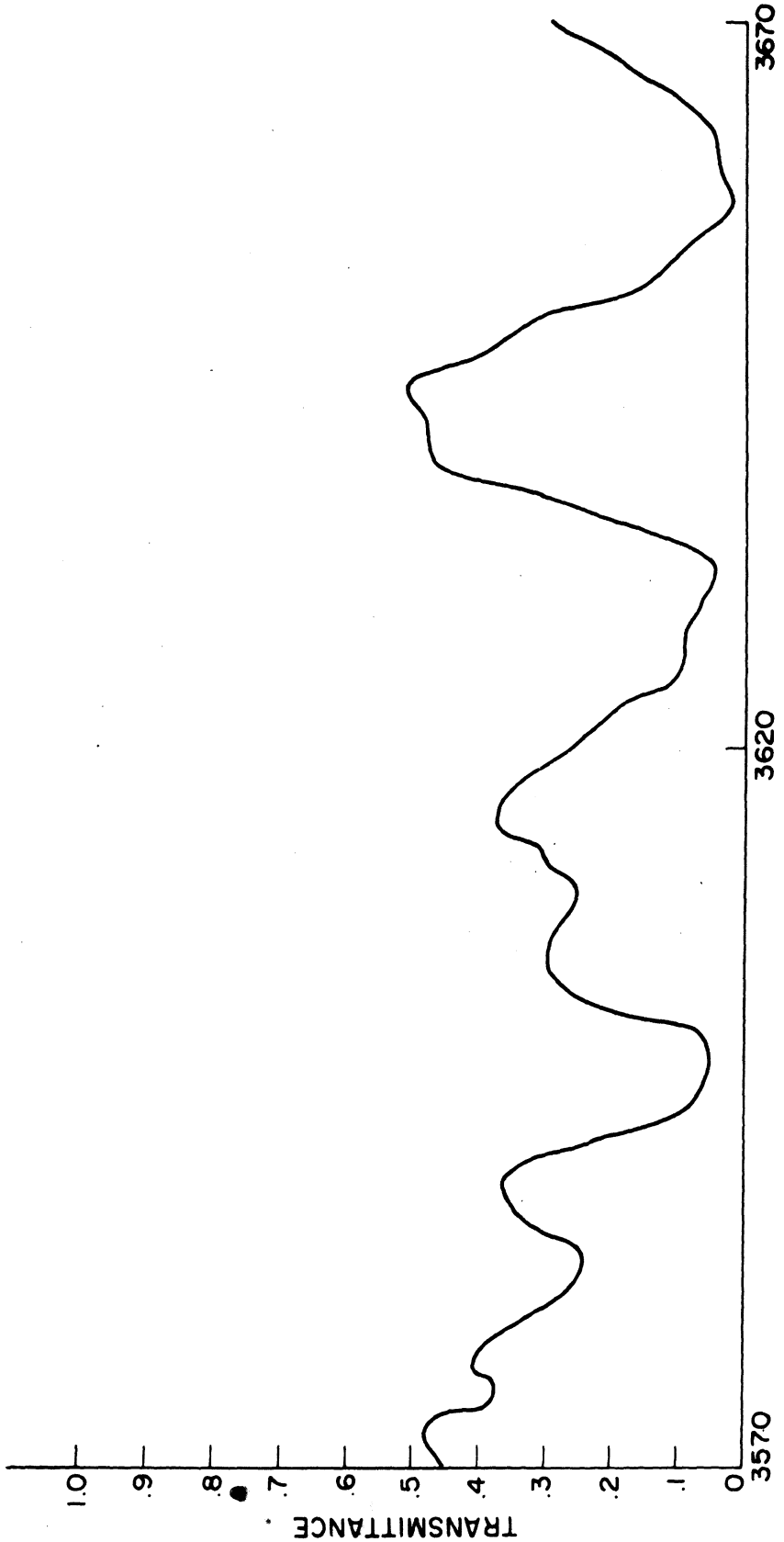

FIG. 2.3. ATMOSPHERIC TRANSMITTANCE CONSIDERING WATER VAPOR, DRY STRATOSPHERE MODEL, 15 KM INITIAL ALTITUDE, 0° INCLINATION TO HORIZONTAL.

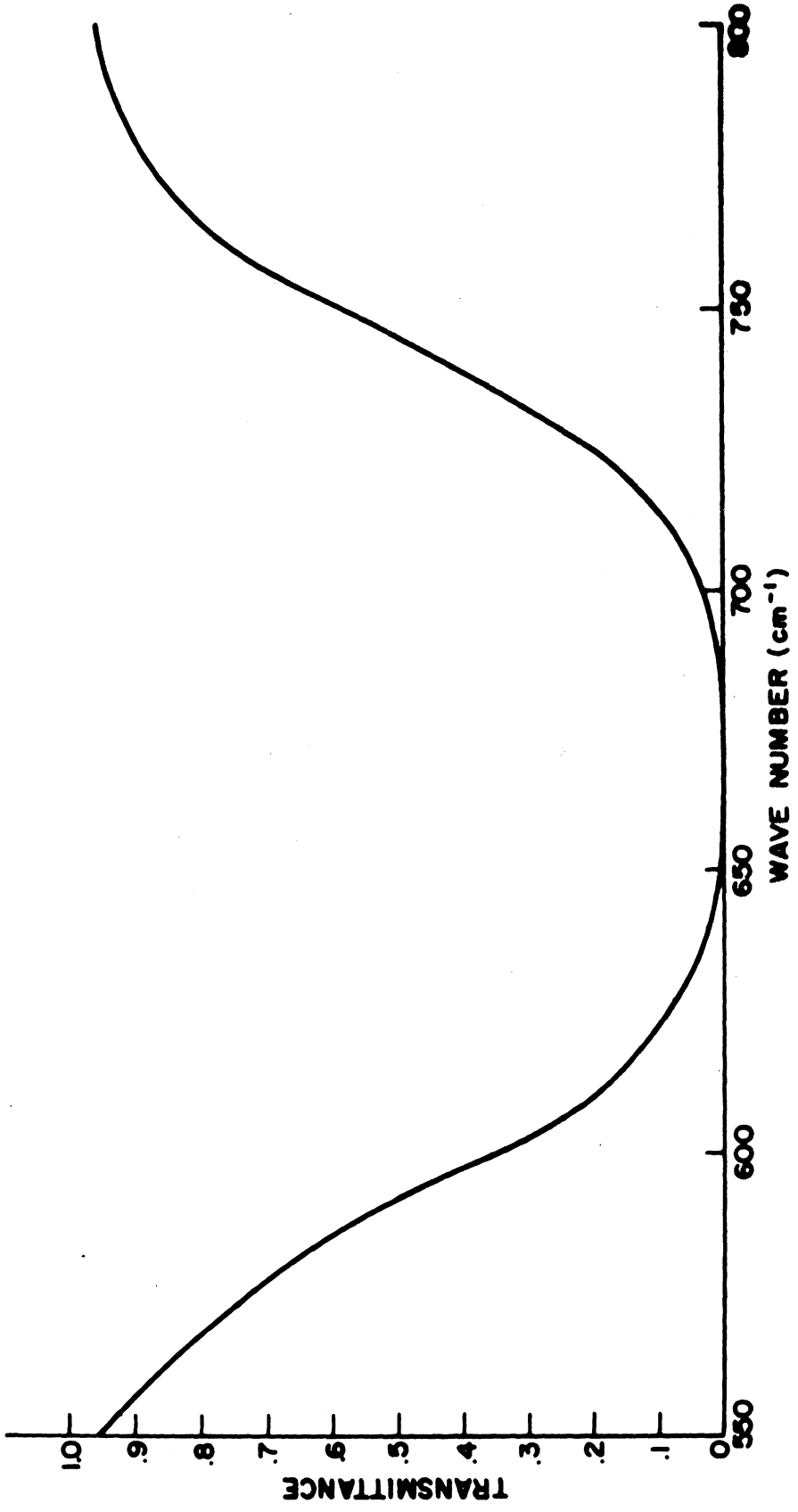

FIG. 2.4. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, AVERAGED OVER 50 CM^{-1} INTERVALS, 15 KM INITIAL ALTITUDE, 0° TO HORIZONTAL.

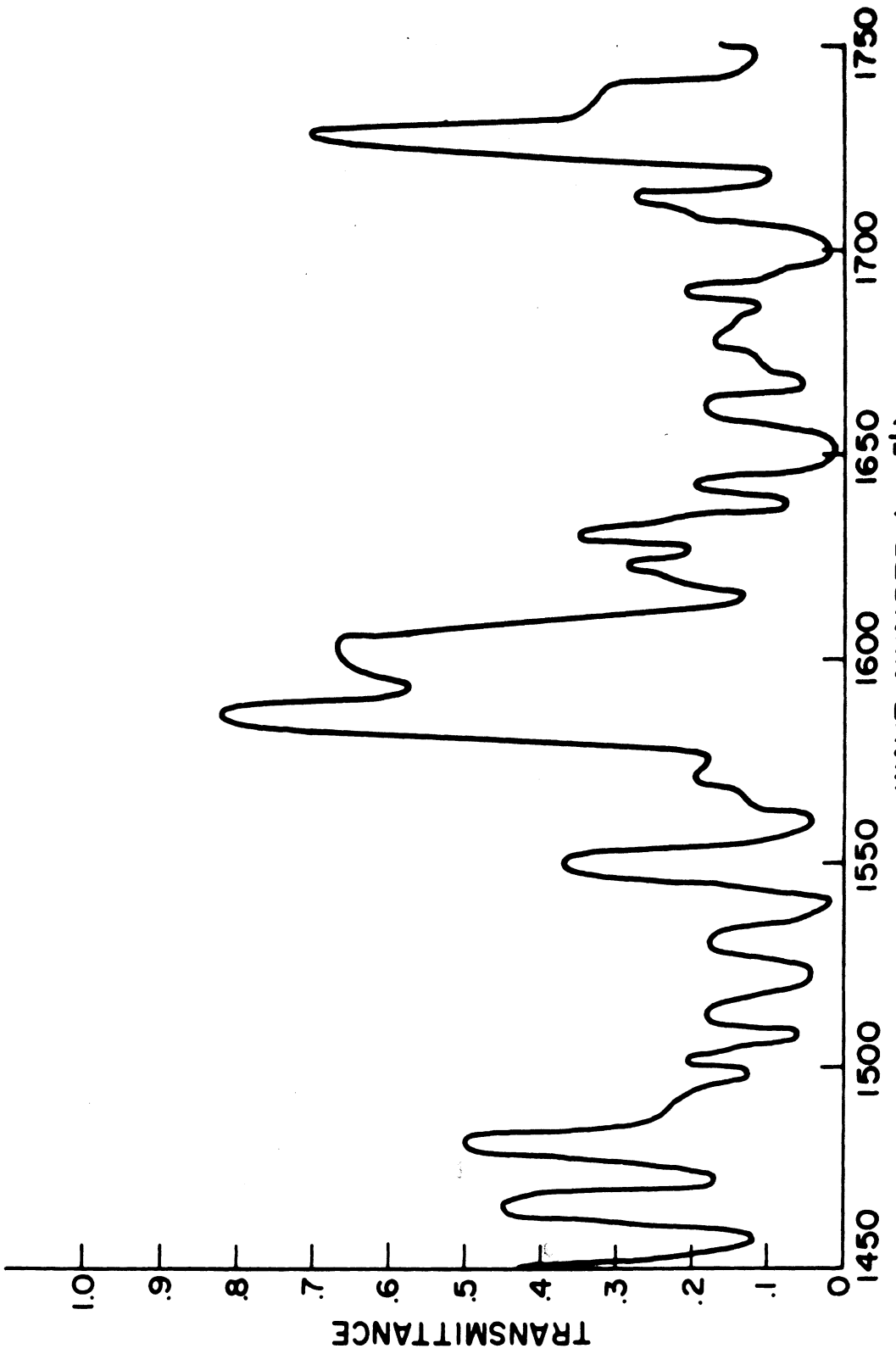

FIG. 2.5. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 15 KM, 0° INCLINATION.

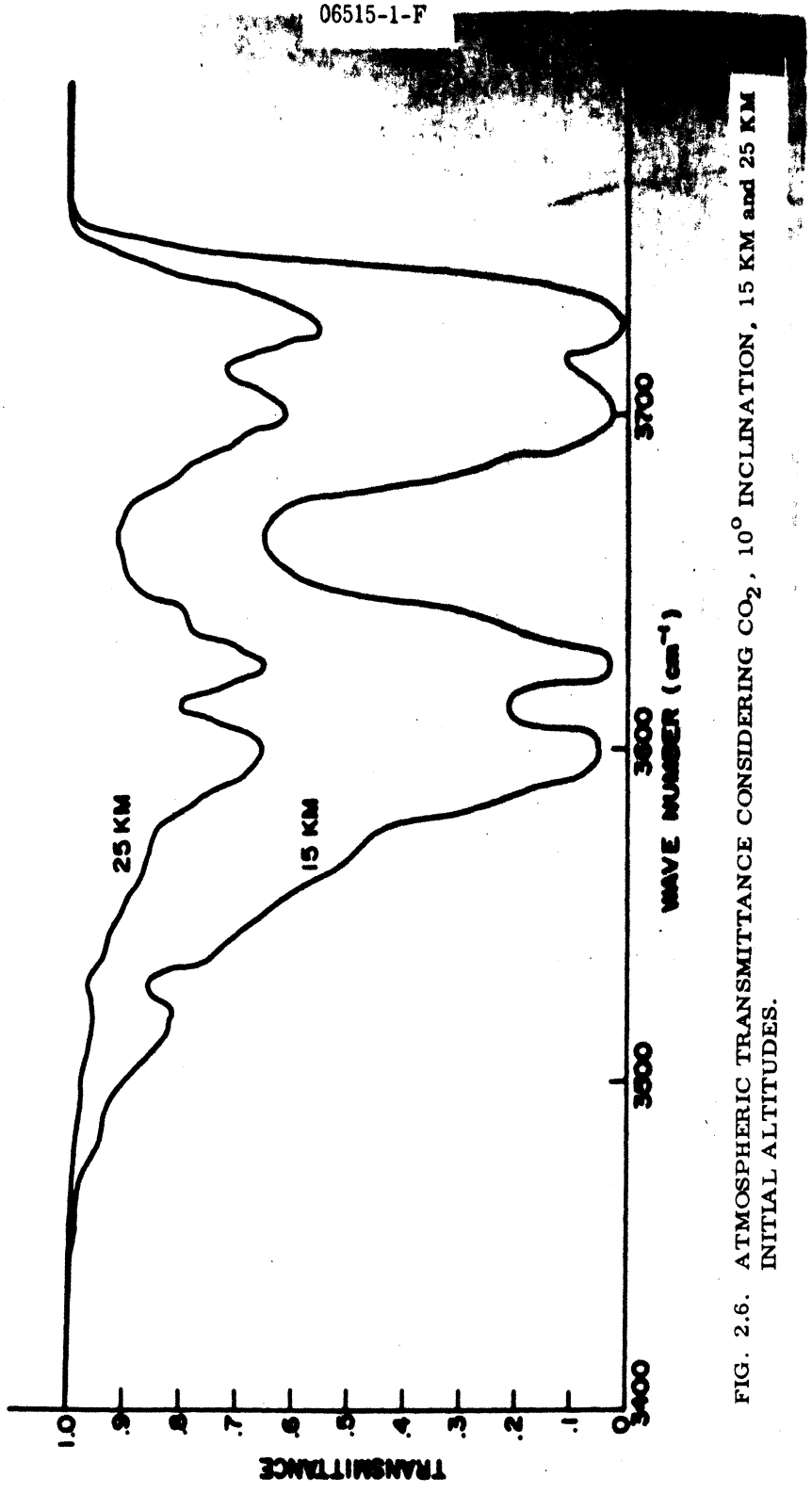

FIG. 2.6. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 10° INCLINATION, 15 KM and 25 KM INITIAL ALTITUDES.

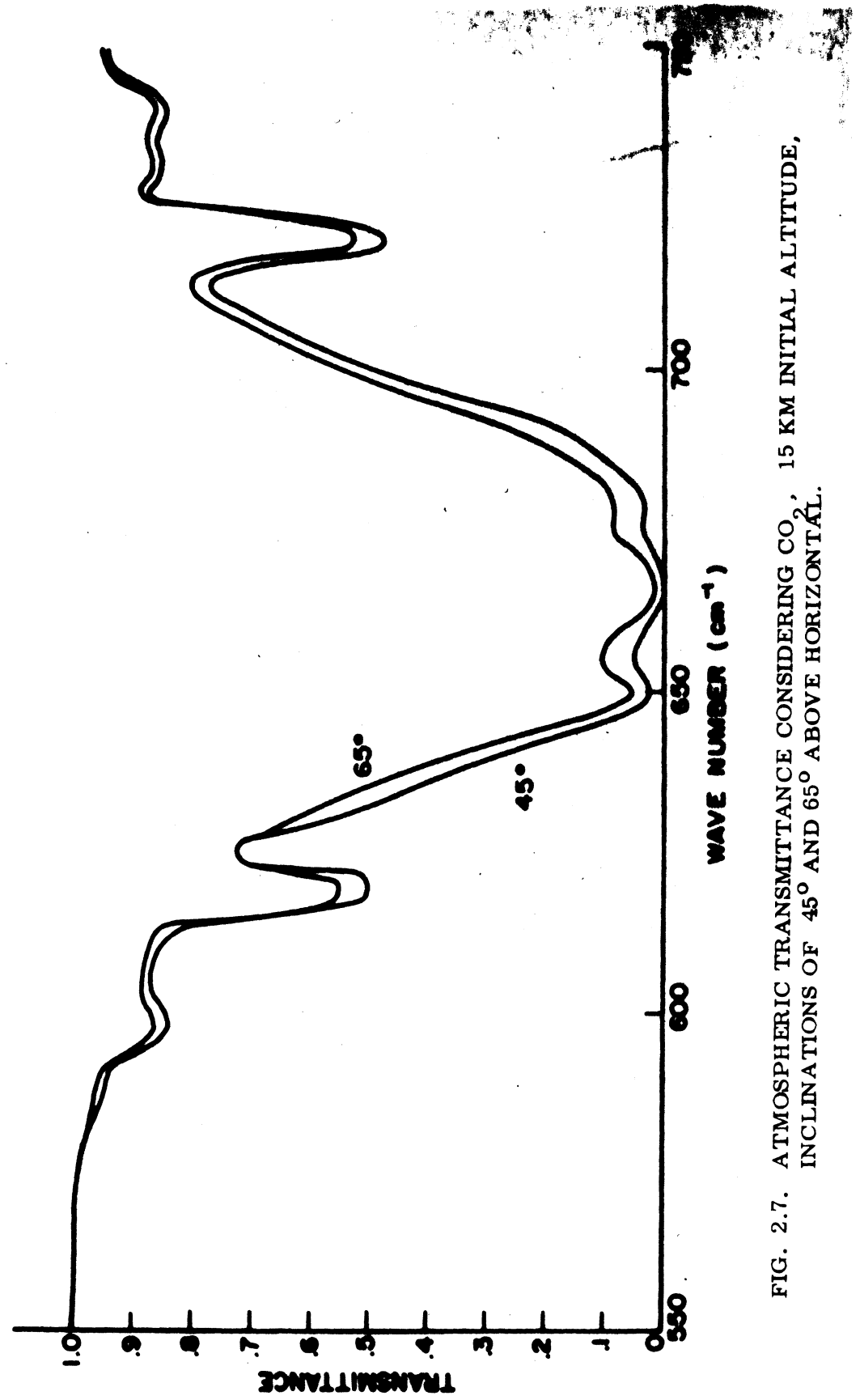

FIG. 2.7. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 15 KM INITIAL ALTITUDE, INCLINATIONS OF 45° AND 65° ABOVE HORIZONTAL.

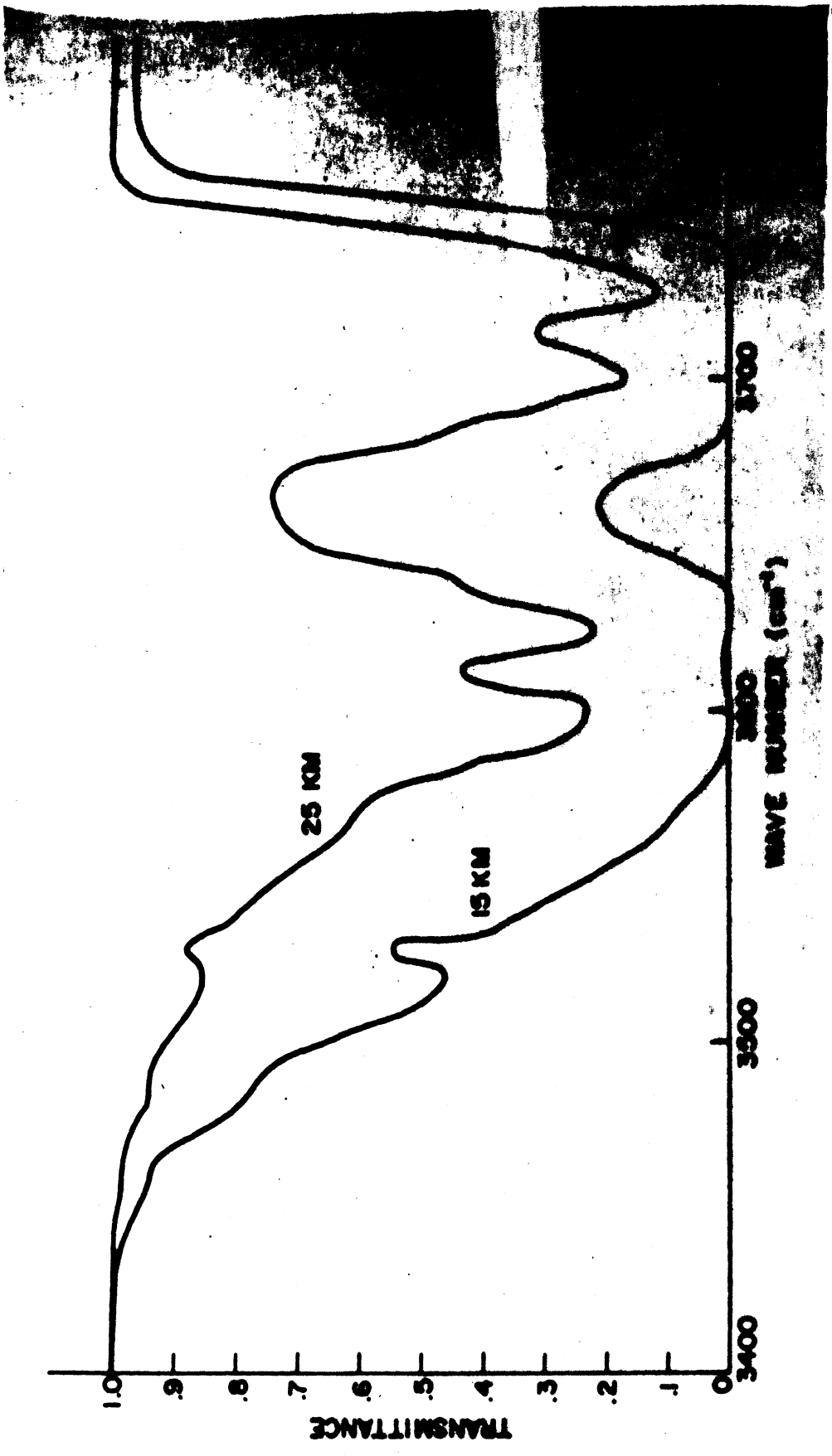

FIG. 2.8. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 0^o TO HORIZONTAL, INITIAL ALTITUDES OF 15 KM and 25 KM.

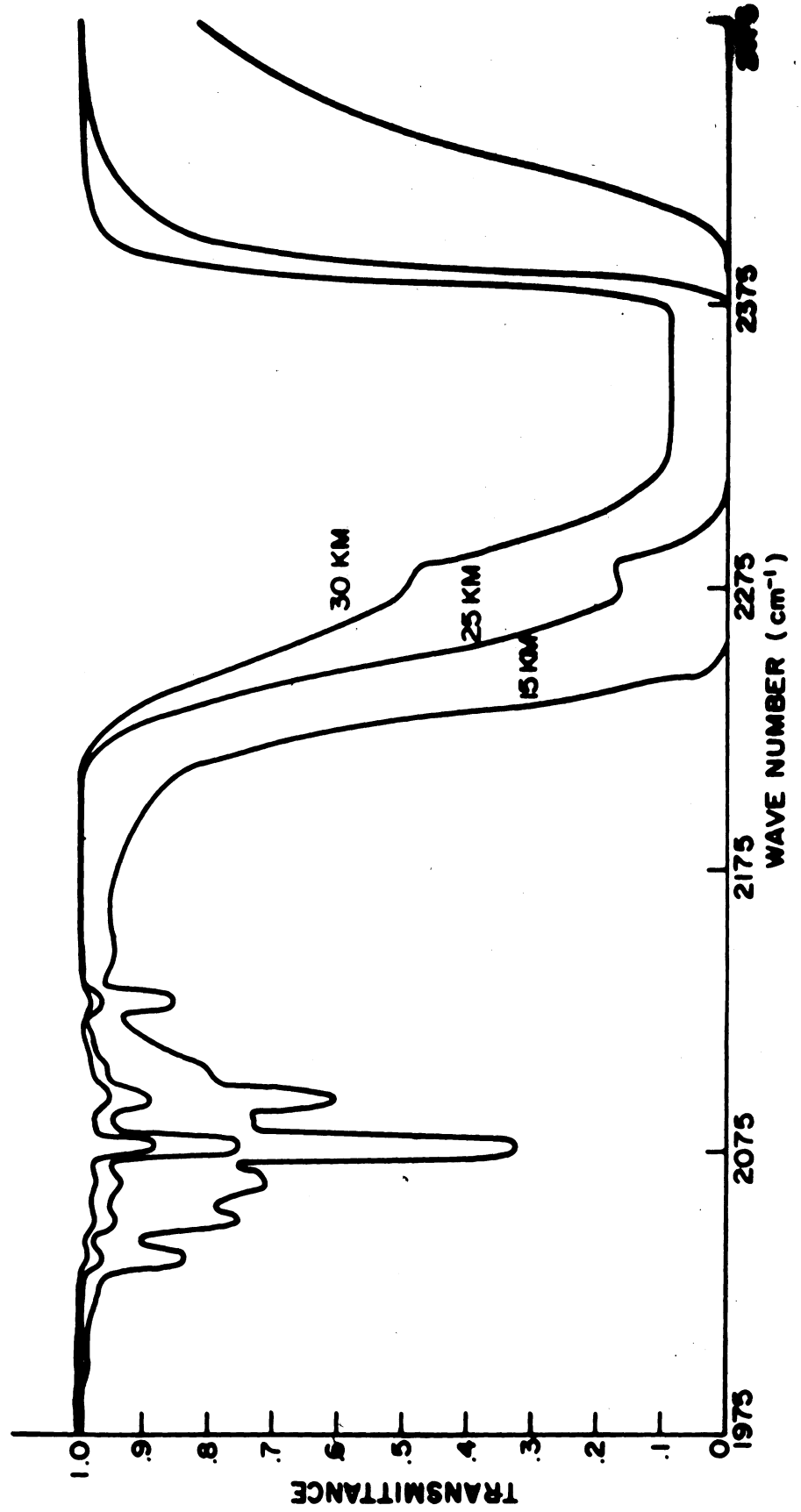

FIG. 2.9. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 0^o INCLINATION TO HORIZONTAL, INITIAL ALTITUDES OF 15 KM, 25 KM, 30 KM.

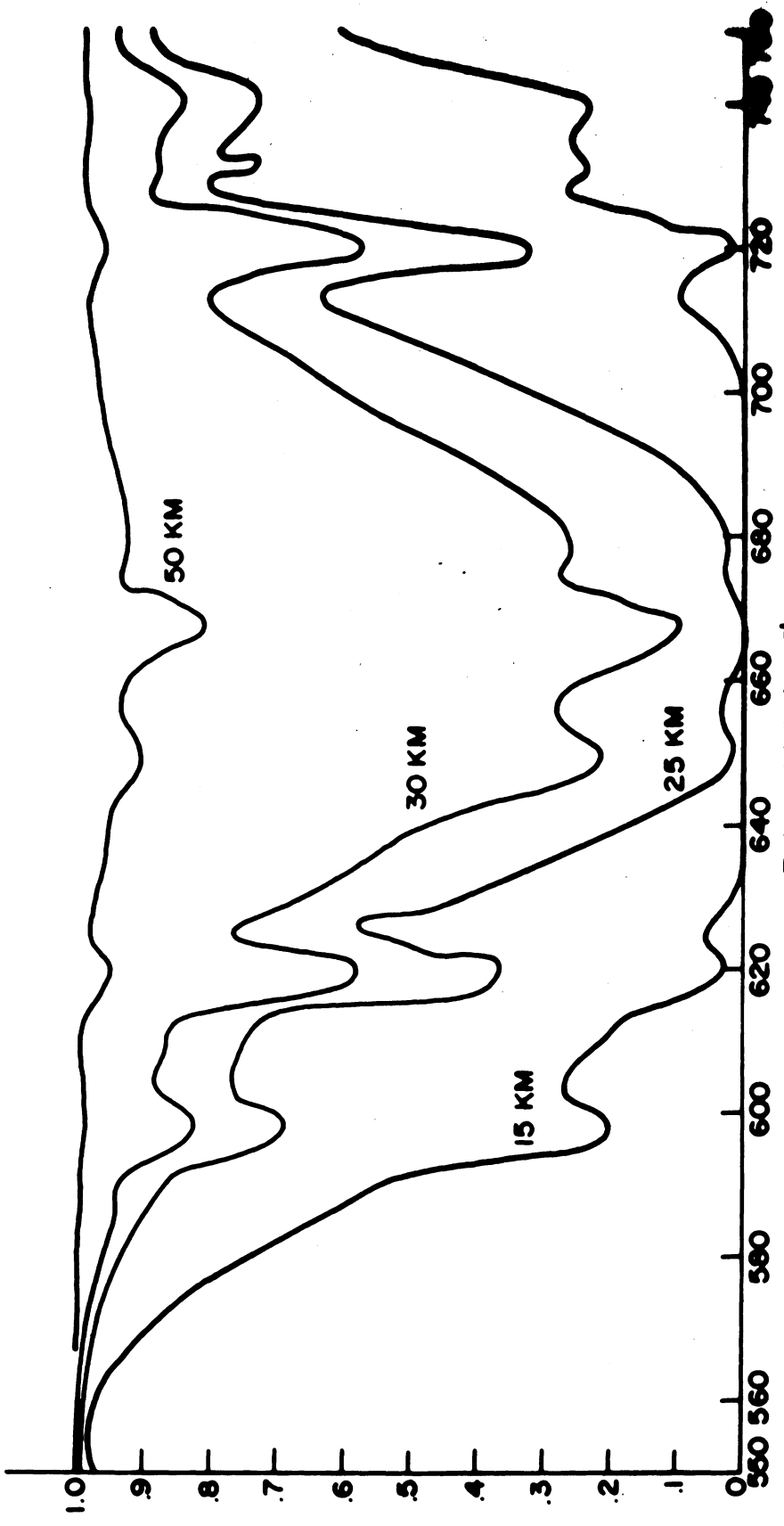

FIG. 2.10. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 0^o INCLINATION TO HORIZONTAL 15 KM, 25 KM, 30 KM, 50 KM INITIAL ALTITUDE.

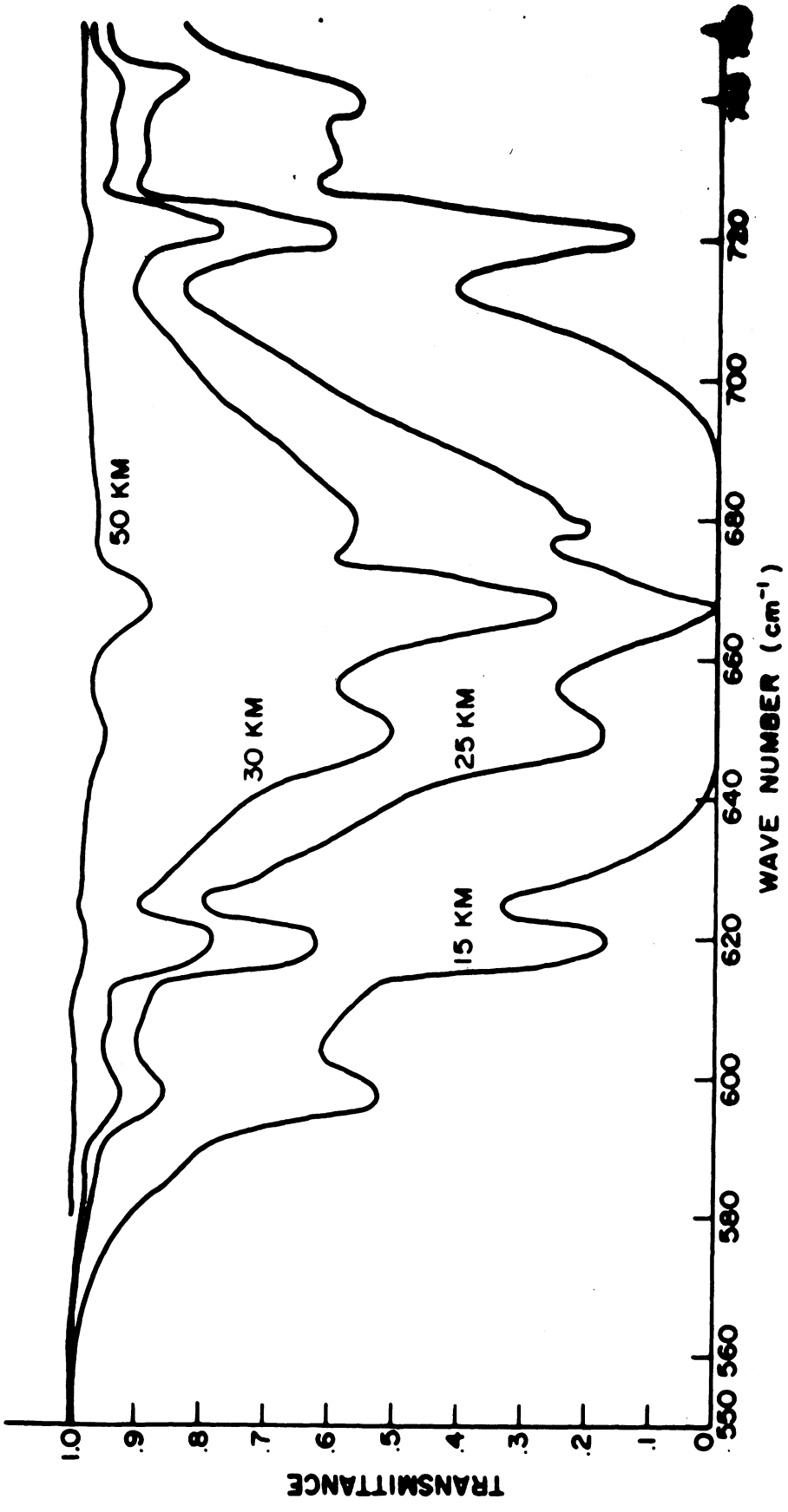

FIG. 2.11. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 5° ELEVATION, 15 KM, 25 KM, 30 KM, 50 KM INITIAL ALTITUDES.

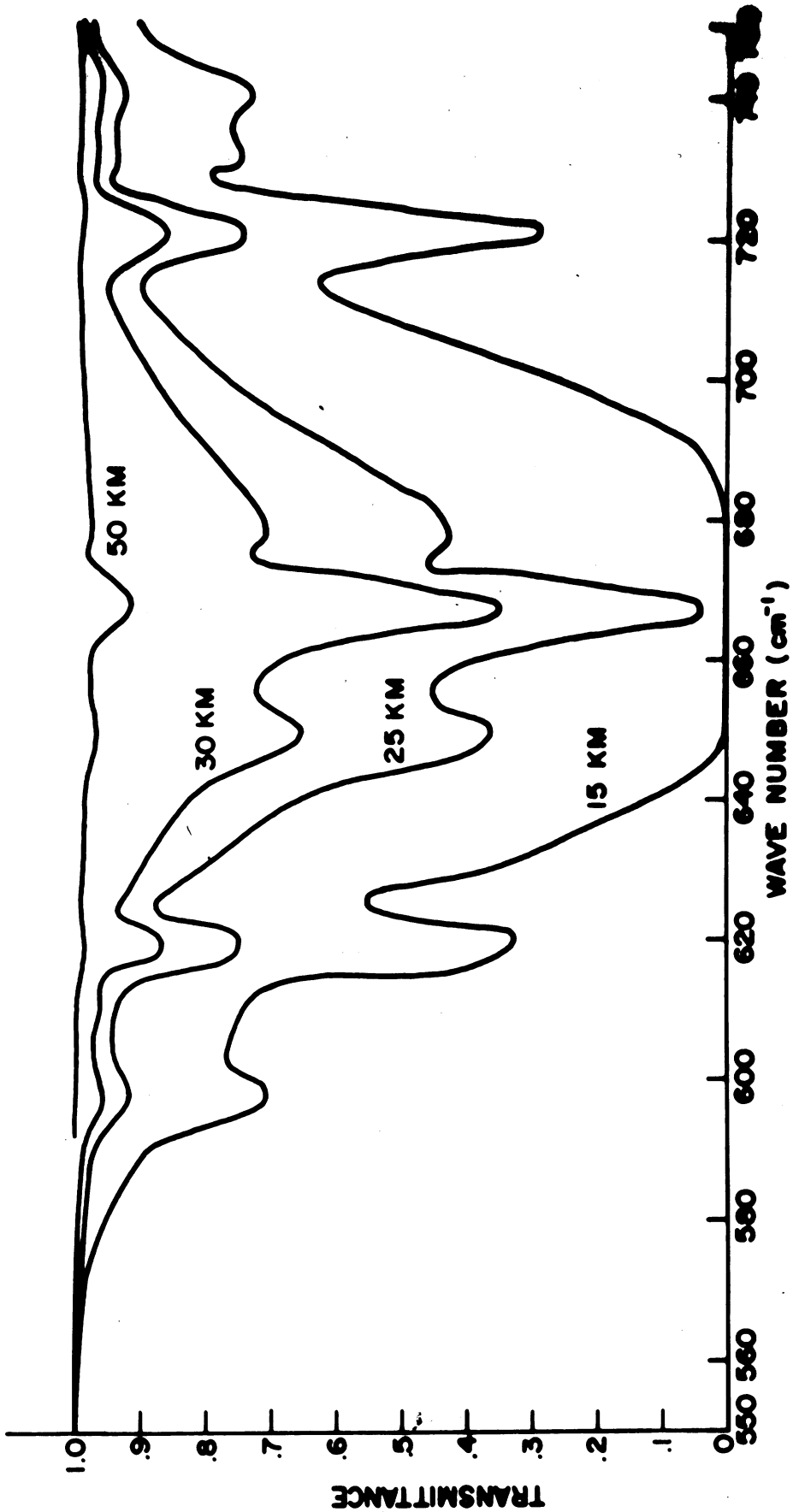

FIG. 2.12. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 15^{o} ELEVATION, INITIAL ALTITUDES OF 15 KM, 25 KM, 30 KM, 50 KM.

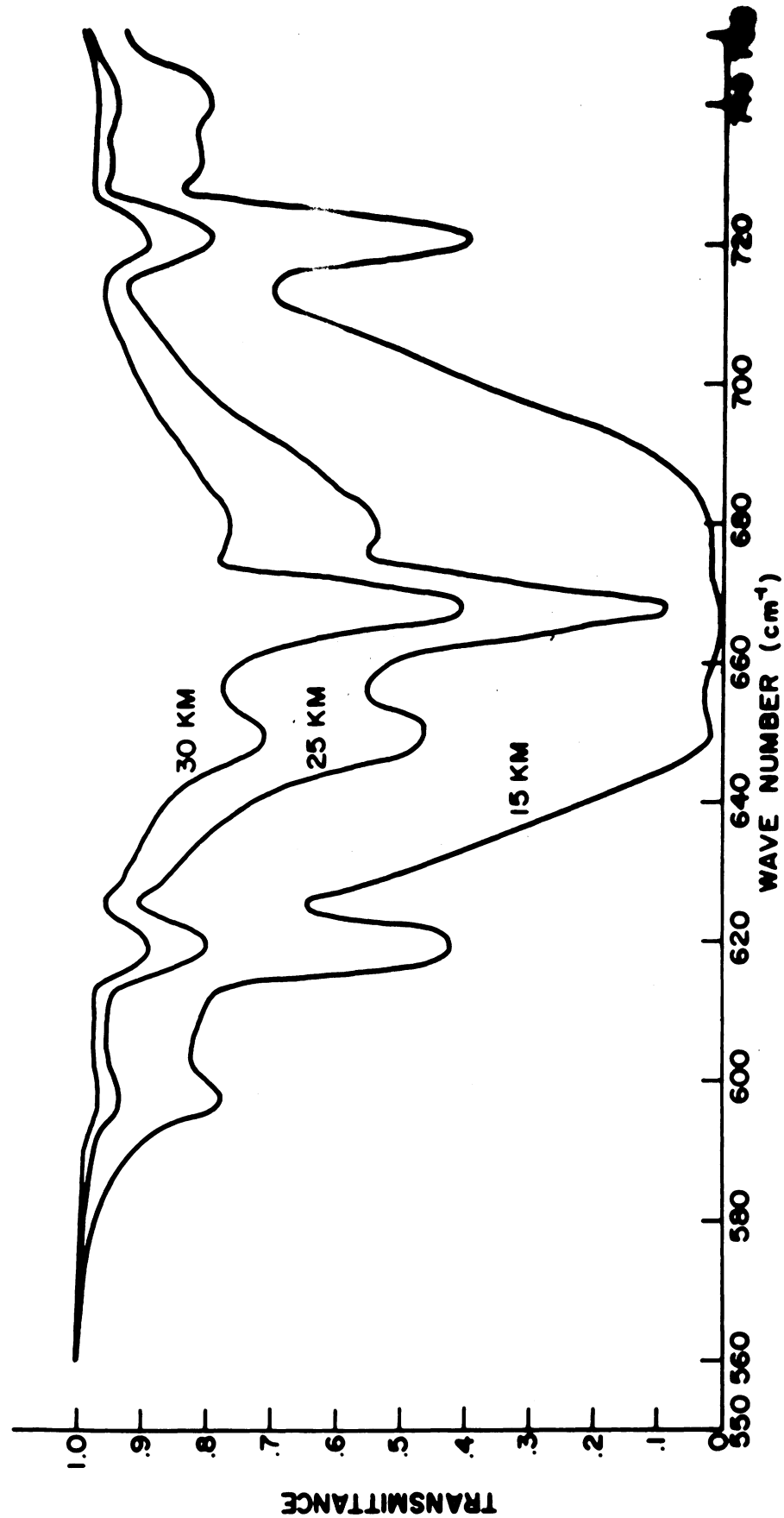

FIG. 2.13. ATMOSPHERIC TRANSMITTANCE CONSIDERING CO_2, 25° ELEVATION, INITIAL ALTITUDES OF 15 KM, 25 KM, 30 KM, 50 KM.

The general character of the curves for water vapor, as displayed in Figs. 2.1-2.3 is the same as for CO_2. High altitude sites and near-vertical inclination improve transmittance. We have chosen to base curves on the calculations made by Plass for the "dry atmosphere" model of Gutnick, since in any case the attenuation should not be less than predicted by this model. Regions of poor transmittance appear to exist for wave numbers 3570-3670 cm^{-1}, 3700-3900 cm^{-1}, and 5300-5400 cm^{-1}.

Because of the paucity of intermediate tabulations, the results of Plass (1963) cannot be applied to transmission over arbitrary slant paths wholly within the atmosphere. For ground-to-aerospace predictions, they require supplementation by a method which treats the segment of the path between ground level and the initial altitudes appearing in the tabulations.

Such a method has been constructed by Zachor (1961) and is detailed in that reference. He starts with band models which contain certain parameters. These are evaluated for the bands important at each wavelength region by fitting laboratory data to the models. Zachor tabulates the appropriate model and parameters as a function of wavelength, essentially from 1.4μ to 10.8μ, for both CO_2 and H_2O. He invokes the Curtis-Godson approximation to formulate his models in terms of mean concentrations and pressures over the propagation path, supplying instructions for obtaining these means associated with arbitrary slant paths (within some limits of location, length, etc.) Summer and winter distributions of water vapor enter the tabulations. Although much laborious computation would be involved, the material of Zachor (1961) is a satisfactory source for a method of predicting CO_2 and water vapor attenuation over paths wholly within the atmosphere. It may, of course, be used for the missing lower part to join with Plass' results in a prediction for ground-to-aerospace applications.

This material deals only with CO_2 and water vapor. It is suggested that

transmitter sites be chosen to avoid sources of atmospheric pollution, since adequate predictions of attenuation due to them are not readily feasible. Alternatively, it will be advantageous to operate within spectral regions which, as far as the spectra of these contaminants are concerned, lie in 'windows'.

No reliable estimate of attenuation due to minor but permanent constituents of the atmosphere is possible at this time.

The strong absorption bands are listed in Table II-4 below, and the corresponding abundance at sea level appears in Table II-6. Table II-4 is based on work reported by Goldberg (1954).

TABLE II-4. ABSORPTION BANDS OF SIGNIFICANT INTRINSIC STRENGTH

Band Center Wave Nr. cm^{-1}	Approximate Wavelength microns	Species	Absorption Cross Section cm^{-1} per cm NTP
3756	2.6	H_2O vapor	180
1597	6.3	H_2O vapor	300
5331	1.9	H_2O vapor	30
7250	1.4	H_2O vapor	20
667	15.0	CO_2	173
2349	4.3	CO_2	2480
3613	2.8	CO_2	28
3714	2.7	CO_2	25
1167	8.5	N_2O	8.5
1285	7.8	N_2O	245
2224	4.5	N_2O	1933
1306	7.7	N_2O	150
3019	3.3	N_2O	300
6005	1.7	N_2O	1

Weak absorption bands would be difficult to avoid, so that in contrast to the strong absorption bands at which the atmosphere may simply be taken as opaque, it is of interest to be able to compute the absorption associated with these bands. Strengths of some weak bands appear in Table II-5 (Goldberg, 1954).

TABLE II-5. SOME WEAK ABSORPTION BANDS

Band Center Wave Nr. cm^{-1}	Wavelength Microns	Constituent	Absorption Coefficient cm^{-1} per cm NTP
2130	4.5	CO_2	0.0037
2093	4.6	CO_2	0.0070
1933	5.2	CO_2	0.0110
6228	1.6	CO_2	0.0077
4260	2.3	CO	1.8000
1868	5.1	NO_2	0.4000
6005	1.5	CH	0.9600

In addition to the absorption due to these constituents, water vapor exhibits a number of weak bands at diverse wavelengths. In the absence of reliable information on the abundance of water vapor (which is in any case highly variable), it does not appear useful to list in Table II-7. Abundances of other constituents have been cited by Altshuler (1961) and Watanabe (1958). They are presented both for the sea-level atmosphere (that table would then pertain to low-altitude situations) and the total atmosphere (which is important for communications with an extra-atmospheric vehicle).

As an indication that even weak bands at sea level can be significant, one may compute the transmission over 1 Km in the 1933 cm^{-1} band of CO_2 at sea level:

$$\ln T_{1Km} = -0.011 \times 32 = -0.352$$

$$T \cong 0.7 \ .$$

Thus, there is non-negligible attenuation for even this illustrative calculation. It might be possible to communicate in spite of this loss. However, if transmission through the entire atmospheric thickness were required, one would predict a greater attenuation:

$$\ln T_{ATM} = -0.011 \times 320 = -3.52$$

$$T \cong 0.3 .$$

TABLE II-6. SEA LEVEL ABUNDANCES

Constituent	Abundance at Sea Level Atm.-cm per Km Path
CO_2	32.0
N_2O	0.027
CH	0.24
CO	0.11

TABLE II-7. SOME TOTAL ABUNDANCES

Constituent	Total Atmospheric Abundance (atm.-cm)
CO_2	320.0
CH	1.2
N_2O	0.4
CO	0.06-0.15
H_2O	$10^3 - 10^4$

2.2.7 General Considerations on Absorption: Windows

We have seen that absorption within bands may be severe. In view of this fact and the difficulty of computing absorption profiles, it is imperative to circumvent the difficulty by deciding to operate in "windows" between the regions of severe absorption. Such a decision is practical only if sufficient bandwidth is available in the windows. Since the relation between frequency and wavelength is

$$\lambda = c/\nu \qquad (2.17)$$

one has for the relation between two descriptions of bandwidth

$$\Delta\lambda = c/\nu^2 \Delta\nu = \frac{\lambda^2}{c} \Delta\nu \qquad (2.18)$$

so that, e.g., for a 1 Kmc bandwidth at wavelength of N microns, the wavelength interval required is given by

$$\Delta\lambda = (N^2/3)\,10^{-5}\ \mu\,. \qquad (2.19)$$

Thus there is adequate bandwidth available in the windows, which are usually determined from the solar spectrum to be (in microns) 0.95 - 1.05, 1.2 - 1.3, 1.5 - 1.8, and 3 - 5. Some authors (Gaertner, 1957) list 8 - 10 as a window but others (Howard and Garing, 1962) consider the O_3 band at 9.6μ to be a serious obstacle to transmission through the atmosphere, and the astronomer Goldberg (1954) refers to it as "strong" so that 8 - 10μ should be rejected as a window until further study. It is clear that the windows listed allow bandwidths in excess of 1 Kmc from application of criterion (2.19)

2.3 Background Radiation from Sky and Astronomical Objects

It is known (see, for example, Augason, Spinrad, 1965) that the sky is a source of radiation in the wavelength region of interest here, commonly designated "atmosphere noise", and that the moon, planets, and many stars emit infrared as well as visible light. Design of a ground-based detector for reception from a satellite transmitter must take this radiation into account. This section contains a brief review of the power levels of these sources, and suggestions for reducing their effects somewhat. The problem of communicating in the reverse direction would necessitate study of radiation from the earth, which is not treated here.

Clear-sky noise radiation originates in two main mechanisms (Bell et al, 1960), scattering of sunlight and emission by atmospheric molecules. Each of these contributes radiation of slightly different characteristics, according to Rollin and Zwas (1965) which combines presentation of experimental data with conclusion from simplified model calculations. According to this publication, the scattered sunlight should not be significant for $\lambda > 4\mu$. For the shorter

wavelengths, and operation during the daylight hours, it may contribute a spectral radiance as large as 10^3 microwatts per cm^2-sterradian-micron, peaking slightly below 1μ wavelength. If this value is intolerably large, there are few measures available. Perhaps operation between 1 and 5μ might allow some reduction of this estimate of spectral radiance from scattered sunlight.

The other source of "sky-noise" is atmospheric emission from molecules, which has been measured by Bell and coworkers and is present both day and night. For a clear sky this seems to be negligible below roughly 5μ. If other considerations dictate use of longer wavelengths, the spectral radiance could be as high as 10^3 microwatts per cm^2 - sterradian-micron, but could be reduced nearly an order of magnitude by choosing a site at high altitude in a dry climate and not subject to air pollution. Even in such a situation, noise can attain the value quoted above during near-horizontal beam orientation, since the effective atmospheric thickness and thus the emissivity are a maximum for the orientation.

Clouds have been found to be good radiators which is reasonable because of their water-vapor content. Their spectral radiance may approach 10^3 $\mu w - cm^{-2}$ $-sterradian^{-1}$ $-\mu^{-1}$, near 10μ wavelength, and again is small for $\lambda < 5\mu$.

Bell reports that an overcast sky radiates like a good blackbody at temperatures near ground ambient. This would imply that little radiation is produced for $\lambda < 5\mu$, and that peak radiance occurs near 10μ, with values of perhaps 600-1200 microwatts per cm^2 - sterradian-micron for ground temperatures of 0^oC to 40^oC.

The moon, planets, and some stars are sources of small amounts of infrared radiation, which is discussed by Rollin and Zwas (1965) on whose tables and curves we base this section. He gives peak spectral irradiances, neglecting absorption in the terrestrial atmosphere which would reduce the already small values. Some of these calculated values are:

TABLE II-8. BACKGROUND FROM ASTRONOMICAL OBJECTS

SOURCE	MAXIMUM SPECTRAL IRRADIANCE, NO ABSORPTION ATM.
	Watts - cm^{-2} - μ^{-1}
Full Moon	10^{-7}
Venus	10^{-10}
Mars, Mercury Jupiter, Saturn	10^{-11} each
Brighter Stars	$< 10^{-10}$ each

2.4 Turbulence, Dispersion and Coherence

Two processes interfering with quantitative astronomical observations are "seeing" and "scintillation," both associated with turbulence in the atmosphere. The former appears as smearing and dancing of images in a telescope and is caused by refraction of the light from a distant source as it passes through turbulent layers of the lower atmosphere. Scintillation appears as fluctuations of the amplifitude of light after passing through turbulent air in the upper strata of the atmosphere. For an optical communication channel the result of those processes is somewhat analagous to the consequences of multipath transmission through the ionosphere or communication by tropospheric scattering in the radio spectrum. The theory of communication channels with randomly varying properties is still in its infancy, and for that reason no complete treatment nor detailed prediction of the reliable communication in the presence of these phenomena can be given here. However, it is suggested that the relatively slow and coarse variations due to seeing can be largely overcome by a choice of site at relatively high altitude. Also the scintillation frequencies are relatively low, as far as available observations indicate; a choice of modulation bands excluding vulnerable frequencies may possibly render scintillation harmless without too radical sacrifice in channel capacity.

Another possible source of distortion of the information-carrying modulation is anomalous dispersion in the atmosphere. However, such dispersion is associated with absorption bands, which furnishes another strong reason for selecting an operating frequency as free as possible from molecular absorption. If this precaution is taken, there should be no disturbance from this source.

The question of whether or not the coherence of the signal waves will be affected by the interaction with the atmosphere is best discussed in a quantum-mechanical context and is therefore postponed to the end of the third chapter.

III
SECOND PROBLEM AREA: CHOICE OF DETECTION SYSTEM FOR OPTIMUM RATE OF TRANSMISSION OF INFORMATION

3.1 Introduction

The fundamental considerations in this general problem area concern the performance limits of a communication channel utilizing a beam of light in the spectrum between 0.4μ and 20μ wavelengths under specified conditions of power, background radiation, distance, antenna optics, etc. A second set of objectives must necessarily be investigations of various ways and means of approaching these performance limits, particularly various methods for modulation and detection, as well as for encoding and decoding.

In the next section we begin this study with a brief review of the quantum theory of the electromagnetic field and the ultimate uncertainties or limits of resolution affecting any observations or measurements of the electromagnetic field. The procedure followed is to expand the field in a set of natural modes, preferably traveling-wave modes, and then to make use of the equivalence of each mode with a quantum-mechanical harmonic oscillator. An arbitrary state of the mode may be expressed alternatively in terms of energy eigenstates or in terms of "coherent states" as basis functions. The energy eigenstates are most convenient for the evaluation of entropy and channel capacity, because they constitute mutually exclusive events. The coherent states, on the other hand, have considerable advantages in describing electromagnetic waves generated at a power level where a classical specification is a good approximation and then reduced to quantum levels by diffraction and attenuation and finally interacting with a detector.

A following section applies the entropy concept to the evaluation of the channel capacity of a beam of light for communication purposes and studies the variation of this capacity with various parameters of the beam.

As an introduction to a digital operation of the channel, a statistical theory of signal detection is developed in close analogy with the classical theory of detectiontion of signals perturbed by white Gaussian noise. The binary channel with an

ideal photon counter is then discussed in the light of these theoretical results, and comparisons are made with other methods of detection. A section on coding and decoding of binary sequences in optical channels for elimination or reduction of errors follows.

A theoretical study of quantum amplifiers is also included in this chapter, since it seems highly likely that laser amplifiers and oscillators may become important components of optical receivers in the future. The motivation for this study is more immediately the inference that an amplifier-counter combination may be designed to count photons as effectively as a photomultiplier and in addition supply the desirable discrimination against background radiation outside the bandwidth of the signal.

3.2 Fundamental Uncertainties in Observations of Electromagnetic Radiation

The usefulness of electromagnetic radiation for communication over great distances depends on the fact that such radiation possesses at a point of observation observable parameters that are related in a causal or statistical manner to the corresponding parameters at the distant source.

The second chapter of this report has discussed some of the transformations to which radiation is exposed during propagation through space and atmosphere primarily from the point of view of classical physics. In the present chapter the objectives are to study the theoretical limits for observations and measurements of electromagnetic fields in the wavelength range of $0.4\,\mu$ to $20\,\mu$; furthermore, to use the results to evaluate the maximum amount of information per second that can be recovered from an electromagnetic wave train of given power and bandwidth and to compare the potential performance of different methods of modulation and detection.

This report is obviously not the place for a complete development of the quantum theory of radiation and the interaction of radiation with matter. A brief review of this subject matter shall be given, however, since a number of the fundamental concepts and mathematical models will necessarily be used in the subsequent analysis. The complete presentation is found in the references (Schiff, 1955,

Louisell, 1964, Glauber, 1963).

The classical electromagnetic field is described by Maxwell's equations and a set of initial and boundary conditions. In order to obtain a discrete set of natural modes, it is convenient to postulate a field confined by a box or waveguide with perfectly reflecting walls. For the box dimension a x b x d the modes are defined by the roots of the dispersion equation,

$$\left(\frac{4\pi}{\ell \cdot a}\right)^2 + \left(\frac{4\pi}{m \cdot b}\right)^2 + \left(\frac{4\pi}{n \cdot d}\right)^2 - \frac{\omega_r^2}{c^2} = k_x^2 + k_y^2 + k_z^2 - \frac{\omega_r^2}{c^2} = 0 , \tag{3.1}$$

where ℓ, m and n are appropriate integers, ω_r is the eigenfrequency of the mode (ℓ, m, n) and c is the velocity of light. Alternatively the reflection in one dimension (say z) may be replaced by periodic boundary conditions, so that the natural modes are defined in terms of traveling waves rather than standing waves. The latter is the more suitable procedure for application to communication channels. In a vacuum (Coulomb gauge) Maxwell's equations may be written

$$\text{curl } E = -\frac{\partial B}{\partial t} = -\frac{\partial}{\partial t} \text{curl } A \tag{3.2}$$

and

$$\text{curl } B = \frac{1}{c^2}\frac{\partial E}{\partial t} = -\frac{1}{c^2}\frac{\partial^2 A}{\partial t^2} = \nabla^2 A , \tag{3.3}$$

where A is the vector potential.

The energy H of the field in the waveguide is

$$H = \frac{1}{2}\int \left(\epsilon_o E^2 + \frac{1}{\mu_o} B^2\right) dV = \frac{1}{2}\int \left[\epsilon_o \left(\frac{\partial A}{\partial t}\right)^2 + \frac{1}{\mu_0} (\text{curl } A)^2\right] dV , \tag{3.4}$$

where dV = dxdydz and integration is extended over the total volume.

Since the natural modes constitute a complete orthogonal set, any field in the waveguide may be expanded in a sum of natural-mode components

$$A(x,y,z,t) = \frac{1}{\sqrt{\epsilon_0}} \sum_r q_r(t) u_r(x,y,z) \tag{3.5}$$

and

$$\frac{\partial A(x,y,z,t)}{\partial t} = \frac{1}{\sqrt{\epsilon_0}} \sum_r p_r(t) u_r(x,y,z) \ , \tag{3.6}$$

where

$$\nabla^2 u_r + \frac{\omega_r^2}{c^2} \cdot u^r = 0 \tag{3.7}$$

and

$$\frac{d^2 q_r}{dt^2} + \omega_r^2 \cdot q_r = 0 \ . \tag{3.8}$$

The quantum-theory modifications of the classical field descriptions may consequently be introduced for each mode at a time. In the following a single mode is referred to and the subscript is dropped from the functions q(t), p(t), and u(x, y, z).

Equations (3.7) and (3.8) describe simple harmonic variation in space and time. A natural mode in the waveguide constitutes a harmonic oscillator, and the theory of the harmonic oscillator according to quantum mechanics applies to it. The time functions p and q introduced above in (3.5) and (3.6) represent the momentum and position of an equivalent mechanical harmonic oscillator of unit mass and radian eigenfrequency ω_r. From (3.2), (3.3) and (3.4) it is easily confirmed that the energy expressed in these variables is

$$H = \frac{1}{2} (p^2 + \omega^2 q^2) \ . \tag{3.9}$$

This result follows from the orthonormality of u(x, y, z), the boundary conditions, the vector identity

$$\text{curl } u \cdot \text{curl } u = u \cdot \text{curl curl } u + \text{div } (u \times \text{curl } u) \tag{3.10}$$

and the separation constant given by (3.7) and (3.8).

In the quantum-mechanical analysis of the oscillator motion H, p and q become Hermitian operators modifying the state vector or wave function of the oscillator. The oscillator is "quantized" in the usual way by postulating the commutation relations

$$\begin{aligned} [q,q] &= [p,p] = 0 , \\ [q,q] &= j\hbar . \end{aligned} \tag{3.11}$$

The operators p and q may be replaced by two linear combinations, the annihilation and creation operators (non-Hermitian), defined by the relations

$$a = \frac{1}{\sqrt{2\hbar\omega}} (\omega q + jp) \tag{3.12}$$

and

$$a^{+} = \frac{1}{\sqrt{2\hbar\omega}} (\omega q - jp) , \tag{3.13}$$

obeying the equivalent commutation relations

$$[a,a] = [a^{+},a^{+}] = 0 \tag{3.14a}$$

$$[a,a^{+}] = 1 . \tag{3.14b}$$

The Hamiltonian energy operator (3.8) can then be written

$$\begin{aligned} H &= \frac{1}{2}\hbar\omega (a^{+}a + aa^{+}) \\ &= \hbar\omega (a^{+}a + \frac{1}{2}) . \end{aligned} \tag{3.15}$$

The eigenstates of this operator constitute a complete orthogonal set of wave functions with the eigenvalues

$$H_n = \hbar\omega(n+\frac{1}{2}) \qquad (n = 0, 1, \ldots) , \tag{3.16}$$

where $\frac{1}{2}\hbar\omega$ constitutes the ground-state energy, the "zero-point" energy or "vacuum fluctuations" related to the basic quantum-mechanical uncertainty

expressed by Heisenberg's uncertainty principle.

An arbitrary (time-varying) state can always be expanded in terms of these functions.

$$| > = \sum_{n=0}^{\infty} c_n | n > \quad . \tag{3.17}$$

The operators a and a^+ make it possible to relate the energy eigenstates to each other by recursion equations:

$$a | n> = \sqrt{n}\, | n - 1 > \tag{3.18}$$

and

$$a^+ | n > = \sqrt{n+1}\, | n + 1 > \quad . \tag{3.19}$$

By repeated application of (3.19), any state $| n >$ may be related to the ground state $| 0 >$

$$| n > = \frac{(a^+)^n}{\sqrt{n!}} \; | 0 > \tag{3.20}$$

and

$$< n | = < 0 | \; \frac{a^n}{\sqrt{n!}} \quad . \tag{3.21}$$

According to the quantum theory, any measurement of an "observable" will as a result indicate one of the eigenstates of the corresponding operator. If the state of an oscillator is represented by an expansion in energy states, and if precise measurements of energy are made, the uncertainty principle states that no information can simultaneously be obtained about the time or phase coordinate. A more desirable procedure would be to minimize the total uncertainty about a pair of conjugate variables, such as position and momentum, or electric and magnetic fields.

Glauber (1963) has suggested than an expansion in terms of "coherent states" and corresponding measurements will accomplish this purpose. The annihilation operator a is a linear combination of the position and momentum operators; the

eigenvalues of this operator will specify jointly position and momentum with an uncertainty which can be shown to be consistent with the uncertainty principle.

The complex eigenvalues α and α^* of the operators a and a^+, respectively, satisfy the equations

$$a|\alpha\rangle = \alpha|\alpha\rangle \tag{3.22}$$

$$\langle\alpha|a^+ = \langle\alpha|\alpha^* \tag{3.23}$$

It is easily shown that a normalized coherent state $|\alpha\rangle$ expanded in energy eigenfunctions is

$$|\alpha\rangle = \exp\left(-\frac{1}{2}|\alpha|^2\right)\sum_n \frac{\alpha^n}{\sqrt{n!}}|n\rangle \ . \tag{3.24}$$

When the expected value $|\alpha|^2$ of the number of photons in the mode is given, it follows that the number of photons in a coherent state of the mode has a Poisson distribution

$$|\langle n|\alpha\rangle|^2 = \frac{|\alpha|^{2n}}{n!}\exp(-|\alpha|^2) \ . \tag{3.25}$$

The ground state $|\alpha|^2 = 0$ is the same unique ground state $|n\rangle$ for $n = 0$.

The coherent states defined by (3.22) form a complete but not orthogonal set of states. Because of the nonorthogonality the expansion of an arbitrary state in terms of coherent states may not be unique, but Glauber has shown a unique procedure that leads to an expansion with suitable properties. The starting point is the expansion (3.17) in energy eigenfunctions. Multiplying both members of (3.17) with the unit operator

$$\frac{1}{\pi}\int |\alpha\rangle\langle\alpha| d^2\alpha = \sum_n |n\rangle\langle n| = 1 \ , \tag{3.26}$$

where integration is extended over the complete complex α-plane, the result is

$$|\rangle = \sum_n c_n |n\rangle = \sum_n c_n \frac{(a^+)^n}{\sqrt{n!}}|0\rangle = \frac{1}{\pi}\int |\alpha\rangle\langle\alpha| \sum_n c_n \frac{(a^+)^n}{\sqrt{n!}}|0\rangle \ . \tag{3.27}$$

In the last member the operators operating on $<\alpha|$ may be replaced by their eigenvalues α^*. The summation is then a function of a complex variable $z = \alpha^*$

$$f(z) = \sum_n c_n \frac{z^n}{\sqrt{n!}} \quad . \tag{3.28}$$

Since

$$\sum_n |c_n|^2 = 1 \tag{3.29}$$

this function is finite for all finite z; it is also analytic over the whole finite z-plane.

Choosing the phase origin to give from (3.25)

$$<\alpha|0> = \exp\left(-\frac{1}{2}|\alpha|^2\right) \tag{3.30}$$

(3.27) is obtained in the more convenient form

$$|f> = \frac{1}{\pi}\int \alpha > f(\alpha^*) \exp(-|\alpha|^2)\, d^2\alpha \, . \tag{3.31}$$

The fact that the expansion functions have been limited to analytic functions of α^* represents the constraint which makes the result unique; otherwise many different expansions in coherent states could be found.

In the case of a mixed coherent state specified statistically by its density operator ρ, Glauber introduces the "P-representation" defined by the relation

$$\rho = \int P(\alpha)\, |\alpha><\alpha|\, d^2\alpha \, . \tag{3.32}$$

Because of the nonorthogonality of the α-states, the integration is not carried out over a set of mutually exclusive events and $P(\alpha)$ is not strictly a probability density, even though it is real and satisfies the equation

$$\text{trace } \rho = \int P(\alpha)\, d^2(\alpha) = 1 \, . \tag{3.33}$$

An interesting special case is the Gaussian density operator

$$\rho = \frac{1}{\pi \langle n \rangle} \int \exp\left\{ - \frac{|\alpha|^2}{\langle n \rangle} \right\} |\alpha\rangle\langle\alpha| d^2\alpha \quad . \tag{3.34}$$

Transformed to the energy representation it proves to be identical with the well-known density operator of a harmonic oscillator in thermal equilibrium, i.e., the operator form of Planck's black-body radiation law. In this orthogonal representation the operator is diagonal and has the form

$$\rho = \frac{1}{1+\langle n \rangle} \sum_m \left\{ \frac{\langle n \rangle}{1+\langle n \rangle} \right\}^m |m\rangle\langle m| \tag{3.35}$$

3.3 Entropy and Channel Capacity

In classical statistical mechanics and in information theory the entropy concept plays a central part (Jaynes I., 1957). When nothing is known about the state of a system, except some average or integrated state variables, the most conservative estimate in statistical terms is the state that maximizes the entropy under the given constraints. This maximum entropy value then is a measure of the amount of information that would be required to specify the state of the system completely. The entropy of a system with n discrete states with probabilities p_i is defined as

$$H = - \sum_{i=1}^{n} p_i \log p_i \tag{3.36}$$

which is a maximum when all the states are equally probably (Shannon, 1949)*.

The entropy of a system described in terms of quantum mechanics (Jaynes II, 1957) is conveniently described in terms of the density operator ρ for the system

$$H = - \text{Trace} (\rho \log \rho) \quad . \tag{3.37}$$

*Boltzmann's symbol H is adopted here for the statistical entropy function. Confusion with the Hamiltonian operator is less likely than confusion with signal power if the symbol S were used for entropy.

The state of a harmonic oscillator in thermal equilibrium with its environment is a maximum entropy state; it is easily shown that the density operator (3.35) maximizes the entropy under the constraint of given temperature or expected value of energy. Equation (3.34) represents the same state in a different mathematical form, but it is not an appropriate basis for calculation of entropy, because the different α-states do not all represent mutually exclusive events.

The maximum entropy of a harmonic oscillator at temperature T or expected value $\langle n \rangle$ of number of quanta above zero-point energy is

$$H = \log (1+\langle n \rangle) + \langle n \rangle \log \left(1+\frac{1}{\langle n \rangle}\right). \tag{3.38}$$

Since both the density function in (3.34) and the wave function for the ground state $|0\rangle$ of the harmonic oscillator have a Gaussian form, it is interesting to compare (3.38) with the classical entropy of a Gaussian variable perturbed by another Gaussian variable. If the variances are $\langle n \rangle$ and N, respectively, the latter is

$$H_G = \log \left(1+\frac{\langle n \rangle}{N}\right) \tag{3.39}$$

for one pair of samples.

Equating (3.38) and (3.39) and solving for the equivalent noise energy leads to the result

$$\frac{1}{N} + \frac{1}{\langle n \rangle} = \left(1+\frac{1}{\langle n \rangle}\right)^{1+\langle n \rangle} \tag{3.40}$$

and

$$N = \left\{\left(1+\frac{1}{\langle n\rangle}\right)^{1+\langle n\rangle} - \frac{1}{\langle n\rangle}\right\}^{-1} . \qquad (3.41)$$

Since the right member of (3.41) is not a constant independent of $\langle n\rangle$, there is no universal level of equivalent classical Gaussian noise which accounts for the fluctuations or uncertainty predicted by the quantum theory.

As expected from the correspondence principle, such a level exists in the limit $\langle n\rangle >> 1$. However, it is not the naively expected zero-point energy, $1/2\hbar\omega$, but $1/e\,\hbar\omega$ per cycle bandwidth. The approximation

$$H \approx \log(1+e\langle n\rangle) \qquad (\langle n\rangle >> 1) \qquad (3.42)$$

holds within one percent for $\langle n\rangle$ as low as four; at $\langle n\rangle = 1$ the error is about -5 percent.

The corresponding limit in the opposite extreme is

$$N \approx \frac{1}{-\log\langle n\rangle} \qquad \text{for } \langle n\rangle << 1 \quad . \qquad (3.43)$$

The second term in (3.38) becomes the dominant one, and the effective signal-to-noise ratio

$$\frac{\langle n\rangle}{N} \approx -\langle n\rangle \log\langle n\rangle \qquad (3.44)$$

decreases more slowly than $\langle n\rangle$, or, in other words, the "noise equivalent" (3.43) is not a constant but decreases logarithmically as $\langle n\rangle$ approaches zero.

These striking differences between the results of Gaussian statistics and those of the quantum distribution (3.3.5) are consequences of the asymmetry and discreteness of the latter. The probabilities p(m) of observing m photons varies exponentially with m

$$p(m) = C\cdot\exp\left[-m\,\frac{hf}{kT}\right] = \frac{1}{1+\langle n\rangle}\left[\frac{\langle n\rangle}{1+\langle n\rangle}\right]^m . \qquad (3.45)$$

This geometric progression of probabilities can be simply illustrated by a lopsided "tree" composed of an infinite number of cascaded identical random binary choices (Fig. 3.1.) The two probabilities in the diagram are:

$$p = p(0) = \frac{1}{1+<n>} \tag{3.46}$$

$$q = 1-p(0) = \frac{<n>}{1+<n>} \tag{3.47}$$

The entropy of m is a weighted sum of the entropy H_o of each single choice

$$H = H_o + qH_o + q^2H_o + \ldots . \tag{3.48}$$

where

$$H_o = -p\log p - q\log q = \frac{1}{1+<n>}\log(1+<n>) + \frac{<n>}{1+<n>}\log\left(1+\frac{1}{<n>}\right). \tag{3.49}$$

The appropriate factoring and substitution then leads to the previous expression (3.38)

$$\begin{aligned} H &= \left\{\log(1+<n>) + <n>\log\left(1+\frac{1}{<n>}\right)\right\}(p(0)+p(1)+p(2)+\ldots.) \\ &= \log(1+<n>) + <n>\log\left(1+\frac{1}{<n>}\right) \\ &= (1+<n>)\log(1+<n>) + <n>\log\frac{1}{<n>} \quad . \end{aligned} \tag{3.50}$$

From the second member we can conclude that in the absence of background radiation and with a perfect photon counter, a binary receiver which can tell the outcome of the first choice only, may utilize the fraction p(0) of the wave entropy. A ternary receiver can utilize the fraction p(0) + p(1), and so on. Thus a 50 percent utilization is indicated with a binary receiver and an average of one photon per sample pair. A higher utilization is obtained at smaller photon counts, but at a considerably lower rate of transmission.

Fig. 3.2 shows the variation of this ideal utilization factor and rate of transmission with frequency in terms of the frequency f_o where

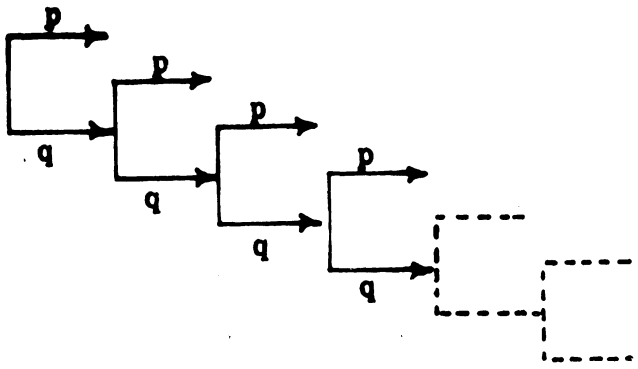

FIG. 3.1. Probability "tree" for consecutive random choices.

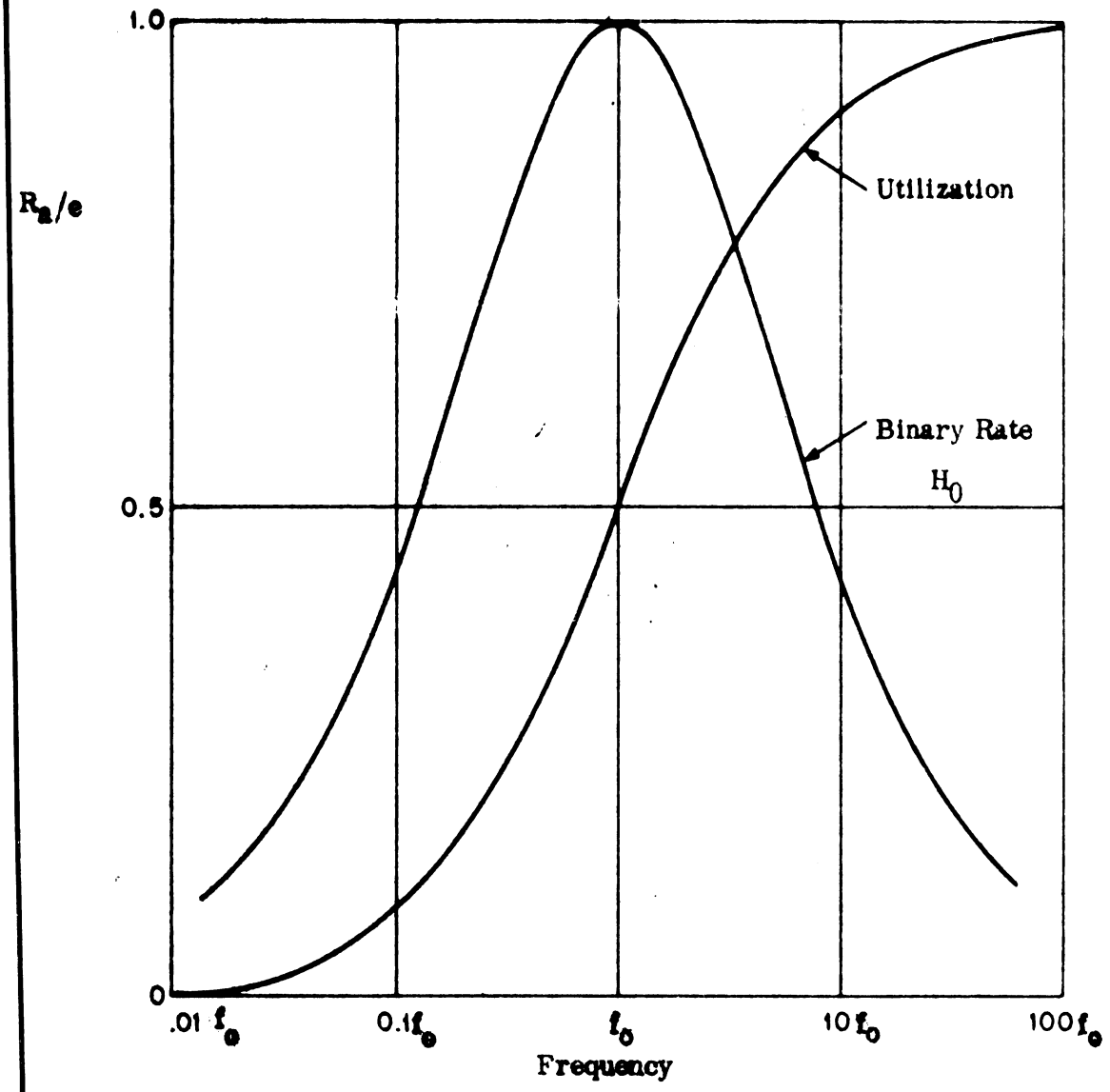

FIG. 3-2. UTILIZATION FACTOR AND RATE OF TRANSMISSION WITH FREQUENCY.

$$\frac{hf_o}{\mathcal{E}_s} = \frac{hf_o}{kT_s} = 1 \qquad (3.51)$$

Here $\mathcal{E}_s$ is the expected value of the energy per sample pair and T_s is the equivalent equilibrium temperature. This unsophisticated discussion, of course, omits the uncertainty and loss of information due to the probabilistic nature of the detection process, which was taken into account in the optimization of the binary channel in Gordon's paper (Gordon, 1962) and which will be considered in the next section of this report.

The square of the envelope of classical Gaussian noise has also an exponential distribution, but it is continuous rather than discrete such as (3.45). The most significant difference between Gaussian noise and a quantum-limited maximum-entropy wave is that envelope and phase are independent random processes in the former case but not in the latter.

Equation (3.38) expresses the maximum total entropy in the latter case. Whatever wave parameters are being measured, the total entropy cannot exceed this value per sample pair. The fact that the entropy is most directly calculated in terms of the energy eigenstates (3.35) is simply a consequence of the orthogonality of these states, their unique property of offering a set of mutally exclusive events. This should not be taken to mean that each mode of the electromagnetic radiation necessarily assumes one of these discret "photon states". The coherent states previously discussed offer an alternative and often preferable way of describing an arbitrary state of the field in terms equally consistent with the uncertainty principle.

The preceding discussion suggests that whenever the average energy per mode or per pair of samples in an electromagnetic wavetrain amounts to at least a few quanta, the fluctuations and uncertainties introduced by the basic quantum phenomena are very nearly equivalent to a perturbation by Gaussian noise with a variance of $h\omega/e$ per cycle bandwidth. In this sense it is sometimes convenient to refer to

a "quantum noise level" which reasonably well above the boundary $\hbar\omega/kT \approx e$ plays a part similar to the thermal noise level far below this boundary.

The maximum-entropy radiation described by the density operator (3.34) and (3.35) for one mode or a pair of samples may represent radiation emitted by a black body in equilibrium or an originally coherent carrier wave modulated to its full informational capacity for given average intensity before being reduced to a low level by diffraction and attenuation.

At microwave frequencies and lower there is no question concerning what is meant by a coherent carrier wave: a pure sine wave in the time domain, or a delta function in the frequency domain, and fair approximations to such waves are readily available. At optical frequencies this is not yet the case. Even the purest "monochromatic" wave so far obtainable from a laser oscillator has a nonzero spectral width depending on spontaneous-emission noise, Doppler broadening, and a number of other physical phenomena.

Such a wave may be characterized by the width and shape of its spectrum or equivalently, in the time domain, by its mutual coherence function (Glauber, 1963, Beran et al, 1964):

$$\Gamma_{12} = \langle V_1(t) V_2^*(t+\tau) \rangle , \tag{3.52}$$

where $V_i(t)\,(i=1,2)$ is a complete representation of a component of the electromagnetic field at the time t a t the point P_i in space. If the negative-frequency range is omitted from the Fourier transform of the field component, the inverse transform multiplied by two defines the "analytic signal" $V_i(t)$. Representing the field by a set of harmonic oscillators, it is found that the annihilation operator of the field modes in the domain of quantum-mechanical operators corresponds, except for a dimensional factor, to this analytic signal $V_i(t)$.

The mutual coherence function is an extremely useful concept in the study of the behavior of light beams in time and space.

Let us however for the moment disregard the imperfections of the carrier

wave and assume that the optical signal at the transmitter can be completely specified classically as a modulated sine wave. At a certain time t_i a pair of time samples specify the instantaneous value of the signal in the following way:

$$s(t_i) = S_{ic} \cos \omega t_i + S_{is} \sin \omega t_i = S_i \cos(\omega t_i - \theta_i) \tag{3.53}$$

where ω is the carrier or midband radian frequency and the pair of samples are alternatively given by (S_{ic}, S_{is}) or (S_i, θ_i). Except for a fixed change of scale and phase the signal will be represented by an expectation value of this same form at a distant observation point, most likely superposed by the expectation value of some background radiation.

If the transmitter makes optimum use of the full width of the frequency band f_1 to f_2, there are

$$n = (f_2 - f_1)T \tag{3.54}$$

independent traveling modes generated by the transmitter and an equal number arriving at the receiver in the observation interval T.

The channel capacity, i.e., the maximum rate of transmission of information, is now the difference between the entropy of the input to the receiver and the entropy of the background noise, multiplied by the number of modes transmitted per second (Gordon 1962).

$$C = \left[H(\text{signal} + \text{noise}) - H(\text{noise})\right] n \quad (\text{for } T = 1 \text{ sec}) \ . \tag{3.55}$$

In terms of the bandwidth $\Delta f = f_2 - f_1$ the average noise power N or noise power per cycle bandwidth $N_o = N/\Delta f$ the expected value of the number of quanta per mode in the absence of a signal at the receiver is

$$\langle n \rangle_{\text{noise}} = \frac{N}{\hbar\omega \Delta f} = \frac{N_o}{\hbar\omega} = \mu \ . \tag{3.56}$$

Similarly in the presence of a signal of average power S

$$\langle n \rangle_{\text{sig+noise}} = \frac{S+N}{\hbar\omega \Delta f} = \frac{S + N_o \Delta f}{\hbar\omega \Delta f} = \frac{S}{\hbar \Delta f} + \mu \ . \tag{3.57}$$

For given average signal power S, noise power N_0 per cycle bandwidth and bandwidth Δf the capacity of an optical communication channel may then be obtained from equations (3.54), (3.55), (3.56) and (3.57). After some simplification the result is

$$C = \left\{\log\left[1+\frac{S}{(N_0+\hbar\omega)\Delta f}\right] + \frac{S+N_0\Delta f}{\hbar\omega\Delta f}\log\left(1+\frac{\hbar\omega\Delta f}{S+N_0\Delta f}\right) - \frac{N_0}{\hbar\omega}\log\left(1+\frac{\hbar\omega}{N_0}\right)\right\}\Delta f \quad . \tag{3.58}$$

If $\hbar\omega\Delta f$ is very small compared to both S and N, this expression reduces to the classical capacity for a channel of given average power perturbed by white Gaussian noise

$$C = \Delta f \log\left(1+\frac{S}{N}\right) . \tag{3.59}$$

In the other extreme, the pure quantum-limited channel ($N_0 << \hbar\omega$), the capacity is

$$C = \Delta f\left\{\log\left(1+\frac{S}{\hbar\omega\Delta f}\right) + \frac{S}{\hbar\omega\Delta f}\log\left(1+\frac{\hbar\omega\Delta f}{S}\right)\right\} . \tag{3.60}$$

These results show that for given bandwidth and average power the channel capacity falls off rapidly toward the higher frequencies where the quantum limitation becomes more and more severe. Optical communication channels are consequently potentially competitive only where extreme bandwidths are required or in cases where other properties than channel capacity are of dominating importance.

If channel capacity is the only criterion, there is no "optimum" bandwidth, since the capacity is a monotone function of bandwidth. However, the channel capacity approaches asymptotically a finite limit with increasing bandwidth. A certain value can therefore be roughly determined beyond which any further gain in capacity is obviously too small to justify the resulting increase in error probability and complexity of terminal equipment.

Such a rough bandwidth measure is in the classical channel perturbed by white Gaussian noise easily obtained by rewriting equation (3.59) in the following manner:

$$\frac{CN_o}{S} = \frac{\Delta f}{\Delta f_m} \log\left(1 + \frac{\Delta f_m}{\Delta f}\right) \tag{3.61}$$

which approaches unity with increasing bandwidth. The right-hand side has one single parameter Δf_m which is the bandwidth for which the signal-to-noise ratio is unity. For $\Delta f = \Delta f_m$ this dimensionless capacity is $\ell n\, 2 = 0.69$ of the asymptotic value one.

For a corresponding analysis of the capacity of a partially quantum-limited channel (3.51), two independent parameters are required:

$$\sigma = \frac{\Delta f_g}{\Delta f} = \frac{S}{\hbar\omega \Delta f} \tag{3.62}$$

$$\mu = \frac{N_o}{\hbar\omega} \tag{3.63}$$

with the result

$$C \cdot \frac{\hbar\omega}{S} = \frac{1}{\sigma}\left\{\log\left(1 + \frac{\sigma}{1+\mu}\right) + (\mu+\sigma)\log\left(1 + \frac{1}{\mu+\sigma}\right) - \mu \log\left(1 + \frac{1}{\mu}\right)\right\} . \tag{3.64}$$

This dimensionless channel capacity increases montonically from

$$0 \text{ at } \Delta f = 0\,(\sigma = \infty) \text{ to } \log\left(1 + \frac{1}{\mu}\right) \text{ for } \Delta f = \infty\,(\sigma = 0) .$$

By algebraic manipulations (3.57) may be transformed to

$$C \cdot \frac{\hbar\omega}{S} = \log\left(1 + \frac{1}{\mu+\sigma}\right) + \frac{1}{\sigma} \log \frac{\left(1 + \frac{1}{\mu+\sigma}\right)^{1+\mu}\left(1 + \frac{\sigma}{\mu}\right)}{\left(1 + \frac{1}{\mu}\right)^{1+\mu}} . \tag{3.65}$$

Here the first term has the same asymptotic values quoted above, while the second term is zero at both limits. The first term reaches a value of half the asymptotic limit

$$\log\left(1 + \frac{1}{\mu+\sigma_h}\right) = \frac{1}{2} \log\left(1 + \frac{1}{\mu}\right) \tag{3.66}$$

for the bandwidth

$$\Delta f_h = \frac{S}{\sigma_h \hbar\omega} = \left[\mu(1+\mu)\right]^{-1/2} \frac{S}{\hbar\omega} = \frac{S}{\sqrt{N_o(\hbar\omega+N_o)}} \tag{3.67}$$

At this point the second term is not negligible; for $\mu \leqslant 1$ it is positive and reaches values up to about half the value of the first term. However, since the first term is by far dominant and the sum is monotone, the bandwidth (3.60) may be taken as a rough indication of the point beyond which the payoff in channel capacity will necessarily become small compared to the cost of coding complexity. Since the measure is rough and of interest primarily in the strongly quantum-limited frequency range where μ is small, it may be more appropriate to write:

$$\Delta f_h = \frac{S}{\sigma_h \hbar\omega} \approx \mu^{-1/2} \cdot \frac{S}{\hbar\omega} = \frac{S}{\sqrt{\hbar\omega N_o}} . \tag{3.68}$$

Since no known practical codes approach the channel capacity with negligible error probability, it is necessary to settle for a rate of transmission of information considerably below the ideal capacity. The discussion above suggests that given a wide band channel the first unsophisticated coding operation that may be performed is to reduce the bandwidth of the signal. Close to the asymptotic capacity a substantial increase in signal-to-noise ratio and reduction in error probability can be obtained with only a moderate loss in theoretical capacity.

If (3.60) is accepted as the upper limit of useful bandwidth, the number of signal photons per sample in a marginal channel is

$$n_s = \sqrt{\mu(1+\mu)} . \tag{3.69}$$

At the crossover point between noise-limited and quantum-limited channels ($\mu = 1$), which for a temperature of 290°K occurs in the infrared at a wavelength of 72μ, $n_s = 1.4$. In the wavelength range of this project 0.4 to 20μ, the marginal number of photons per sample varies from 10^{-22} to 0.3. The theoretical channel capacity under these conditions varies approximately from 45.5 $S/\hbar\omega$ to 1.1 $S/\hbar\omega$ bits, respectively, i.e. roughly fifty to one bit, respectively, multiplied by the average

number of received signal photons per second.

It is doubtful that it will be practically and economically justifiable to design optical communication links for as large bandwidths and as small photon numbers per sample as indicated by the marginal conditions discussed above.

3.4 Statistical Detection Theory for Optical Signals

The derivation of the channel capacity (3.51) implies that in principle an ideal code and an ideal detection process exist which make it possible to receive information at the rate stated by this equation. With an "ideal" code is understood a code that may require infinitely long "words" or "messages" and infinite storage time. An "ideal detection process" measures the signal with no greater uncertainty than required by the postulates of quantum mechanics.

With practically realizable codes and additional uncertainties introduced by the measuring equipment the rate of transmission may fall far short of the channel capacity. The optimum design of an optical channel then involves a judicious compromise between fractional sacrifice of the channel capacity on the one hand and cost and complexity on the other.

Some of the basic design alternatives for a communication system are:

1. Information-carrying wave parameter: Amplitude, phase, frequency, or polarization
2. Digital or analog data transmission
3. Detection method:
   a) Direct transformation of the received light intensity into charge or current, "photon counting,"
   b) Superheterodyne conversion to a microwave frequency band and subsequent classical data processing.
   c) Photon amplifier followed by a).
   d) "Homodyne" detection.
4. Choice of a redundant code for elimination of errors.

An early analysis of optical communication systems (Gordon, 1962) indicated

among other results that for low-level channels where less than one signal photon per sample pair on the average reaches the receiver a remarkably high fraction of the channel capacity appears to be attainable with binary digital signals detected by a photon counter. Under these conditions such a simple system compared favorably with more sophisticated systems. These tentative conclusions furnished the motivation for the directives in the present problem area, which call for an investigation of the theoretical upper limits for the performance of a low-level binary photon-counter communication channel, i.e., a maximum rate of transmission of information with a negligible frequency of errors.

This investigation requires the performance of two tasks:

1. Optimization by means of statistical decision theory* of the detection of binary digits in a quantum-limited channel.

2. Design of an error-correcting binary code to match the channel statistics determined in 1.

The purpose of the present section is to analyze the first task; the subsequent section will discuss the second task.

Figure 3.3 is a schematic presentation of a generalized form of the first task. The space $\mathcal{E}_S$ represents the ensemble of possible signals that may arrive from the transmitter in a given observation interval T. Each signal is specified by the expected values of the field variables at a set of time samples, or equivalently by a set of eigenvalues α_i of coherent states of the field modes incident during T. In the same terms the space $\mathcal{E}_N$ represents the ensemble of all possible background radiation fields in T. The purpose of the counter and decision unit is to produce a decision as to whether or not a signal has been reaching the input during T. In order to simplify the first formulation of the problem, the signal ensemble may be taken to have one single nonzero member S_1 so that the decision is between two subsets of $\mathcal{E}_S$, S_1 and S_o, the latter being the null set for which all samples are zero. Later

*For a presentation of the classical statistical decision theory applied to signals in the presence of noise, see Peterson et al (1954), Van Meter et al (1954).

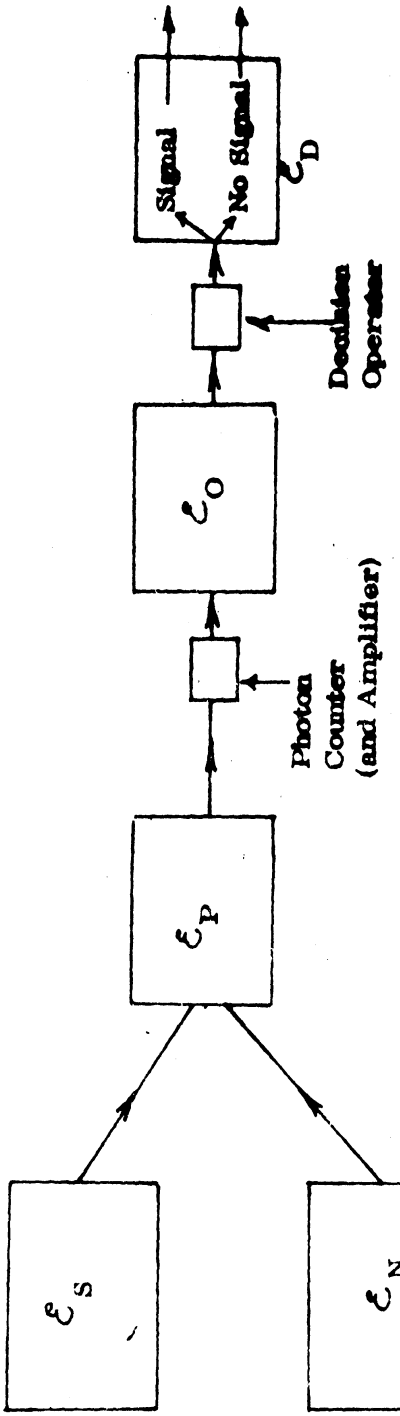

FIG. 3-3: $\mathcal{E}_S$ = Signal Ensemble (Transmitted Radiation); $\mathcal{E}_N$ = Noise Ensemble (Thermal Background Radiation), $\mathcal{E}_P$ = Pre-observation Space, $\mathcal{E}_O$ = Observation Space, and $\mathcal{E}_D$ = Decision Space.

on, the results may be generalized by averaging them over a more realistic signal ensemble, taking into account uncertainties in the phase and amplitude of the carrier wave generated at the transmitter, variations in propagation characteristics, etc.

The superposition of signal and background radiation generates "pre-observation space" $\mathcal{E}_p$ which is essentiallyan α-space in terms of coherent states of the field. Any observation on this space is subject to the Heisenberg uncertainty; an ideal photon counter may be used to project each point in this space onto a set of points in "observation space" $\mathcal{E}_o$, where the points are defined by a stream of electrons observable by classical means. The rest of the receiving equipment serves to implement a certain decision rule, which projects observation spaces into decision space, which has only two elements: "Yes" and "no", signal or no signal.

An ideal photon counter translates each input $|\alpha_i|^2$ into a number of photons n_i according to (3.25). It should be noted that the Poisson distribution governing this random transformation is not an independent postulate but follows directly from the quantum statistics of the radiation field. A simple model of an ideal counter is an ensemble of atomic or molecular systems normally in their ground states; each photon captured raises one of them to a continuum of excited states, where an electron emitted into a vacuum or a conduction band can be observed by classical means. Even in case of 100 percent quantum efficiency and a perfect count of electrons, the counter introduces a certain amount of uncertainty and statistical fluctuations, required by the uncertainty principle, or equivalently, produced by the photon statistics. This "quantum noise" is another manifestation of the same fundamental rules discussed above in terms of entropy (Eqs. 3.38 and 3.39).

Because of the random nature of the background radiation and the "quantum noise" introduced by the counter, it is not possible to find a decision rule that consistently will lead to a correct decision, whether or not a signal was received during the observation interval T. In order to formulate such a rule in a quantitative manner and evaluate the performance its application will lead to, it is first necessary to assign appropriate probabilities to the various points in $\mathcal{E}_p$ and $\mathcal{E}_o$. If a point P

is produced by thermal background radiation alone, each sample in the time interval T may be assigned a Gaussian probability density with zero mean and a variance corresponding to the temperature of the background. In the presence of the signal, the mean distribution of each sample is displaced by the expected value of the signal sample. In the projection of a point in $\mathcal{E}_p$ to a point in $\mathcal{E}_o$ each pair of samples a_i (real and imaginary) of the former changes into a time sample of the latter according to (3.25). The samples of P and S_1 may be written in real rather than complex form as follows:

$$p_i = P_{ic} \cos \omega t_i + P_{is} \sin \omega t_i = P_i \cos(\omega t_i - \Phi_i) \tag{3.70}$$

and

$$s_i = S_{ic} \cos \omega t_i + S_{is} \sin \omega t_i = S_i \cos(\omega t_i - \Phi_i) . \tag{3.71}$$

In the absence of a signal the Gaussian distribution in pre-observation space is then written

$$dF_{S_o}(P, \Phi) = \prod_{i=o}^{n} \frac{d\Phi_i}{2\pi} \quad \frac{P_i dP_i}{N} \exp\left\{-\frac{P_i^2}{2N}\right\}, \tag{3.72}$$

where N is the noise variance and

$$n = (f_2 - f_1) T \tag{3.73}$$

is the number of sample pairs in time T.

In the presence of a signal the Gaussian variable of mean zero and variance N is p - s; the corresponding distribution is, after the above transformation of variables

$$dF_{S_1}(P, \Phi) = \prod_{i=o}^{n} \frac{d\Phi_i}{2\pi} \cdot \frac{P_i dP_i}{N} \exp\left\{-\frac{1}{2N}\left[P_i^2 + S_i^2 - 2P_i S_i \cos(\Phi_i - \theta_i)\right]\right\} \tag{3.74}$$

The operation of the photon counter depends only on the power, not on the phase of the incident radiation. Equations (3.65) and (3.67) may consequently be integrated at once with respect to Φ_i

$$dF_{S_o}(P) = \prod_{i=o}^{n} \exp\left\{-\frac{P_i^2}{2N}\right\} \cdot \frac{P_i dP_i}{N} \tag{3.75}$$

and

$$dF_{S_1}(P) = \exp\left\{-\frac{L(S)}{N_o}\right\} \cdot \prod_{i=o}^{n} \exp\left(-\frac{P_i^2}{2N}\right) \cdot I_o\left(\frac{P_i S_i}{N}\right) \frac{P_i dP_i}{N} \ , \tag{3.76}$$

where N_o is the noise power per cycle bandwidth, i.e., the spectral density of the noise. L(S) is the total energy of the signal, evaluated from the average value of the signal power

$$\bar{S} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} S_i^2 = \frac{1}{T} \int_0^T \left[S(t) \cos(\omega t - \Phi_i)\right]^2 dt = \frac{L(S)}{T} \tag{3.77}$$

and

$$\sum_{i=1}^{n} \frac{1}{2} S_i^2 = (f_2 - f_1) \cdot L(S) \ . \tag{3.78}$$

The symbol $I_o(x)$ denotes as usual the zero-order Bessel function with an imaginary argument.

The expected value of the number of photons in the traveling-wave field mode represented by the ith sample is

$$m_i = \frac{1}{2} \frac{P_i^2 \cdot \tau}{\hbar\omega} = \frac{P_i^2}{2N} \cdot \frac{N_o}{\hbar\omega} = \frac{P_i^2}{2N} \cdot \mu \ , \tag{3.79}$$

defining the new parameters μ, the average number of thermal photons per mode, and τ, the time between consecutive samples or the reciprocal of the bandwidth.

Let r_i be the number of photons counted by the detector in the ith interval. From the Poisson distribution for r_i, given m_i, the observation-space distributions are obtained as follows:

$$P_{S_o}(r) = \int_0^\infty dF_{S_o}(m_i)\, P_{m_i}(r)$$

$$= \prod_{i=o}^{n} \int_0^\infty \frac{dm_i}{\mu} \exp\left(-\frac{m_i}{\mu}\right) \cdot \frac{m_i^{r_1}}{r_i!} \cdot e^{-m_i} = \prod_{i=o}^{n} \frac{\mu^{r_i}}{(1+\mu)^{r_i+1}} \quad , \tag{3.80}$$

which is Planck's well known black-body distribution.

$$P_{S_1}(r) = \int_0^\infty dF_{S_1}(m_i)\, P_{m_i}(r)$$

$$= \prod_{i=o}^{n} \exp\left\{-\frac{S_i^2}{2N}\right\} \int_0^\infty \frac{dm_i}{\mu} \cdot I_o\left(\frac{P_i S_i}{N}\right) \exp\left\{-m_i \frac{1+\mu}{\mu}\right\} \frac{m_i^{r_i}}{r_i!} \quad . \tag{3.81}$$

After the substitutions

$$x^2 = \frac{1+\mu}{\mu} m_i = \frac{P_i^2}{2N}(1+\mu) \tag{3.82}$$

and

$$\beta^2 = \gamma = \frac{S_i^2}{2N(1+\mu)} \quad , \tag{3.83}$$

each integral in (3.81) can be written

$$J = \frac{2}{r!} \frac{\mu^r}{(1+\mu)^{r+1}} \int_0^\infty \epsilon^{-x^2} I_o(2\beta x)\, x^{2r+1}\, dx$$

$$= \frac{\mu^r}{(1+\mu)^{r+1}} \frac{e^\gamma}{r!} L_r(-\gamma) \tag{3.84}$$

where $L_r(x)$ is the rth order Laguerre polynomial.

Consequently, the distribution in the presence of a signal is*

$$P_{S_1}(r) = \prod_{i=0}^{n} e^{-\mu\gamma_i} L_{r_i}(-\gamma_i) \frac{\mu^{r_i}}{(1+\mu)^{r_i+1}} \cdot \frac{1}{r_i!} \tag{3.85}$$

and the likelihood ratio for any point in observation space represented by a radius vector $\vec{r}$

$$\lambda(r) = \frac{P_{S_1}(r)}{P_{S_0}(r)} = \exp\left\{-\mu \sum_i \gamma_i\right\} \prod_i L_{r_i}(-\gamma_i) \frac{1}{r_i!} \quad . \tag{3.86}$$

Here γ_i and μ are known characteristics of the signal and noise ensembles, normalized with respect to the photon energy as specified by (3.72) and (3.75). The components of the random vector occur only in the polynomial $L_r(-\gamma)$.

The logarithm of the likelihood ratio is

$$\log \lambda(r) = -\mu \sum_i \gamma_i + \sum_i \log\left[\frac{1}{r_i!} L_{r_i}(-\gamma_i)\right] = -A + \alpha . \tag{3.87}$$

The first term on the right is independent of r_i; in a previous notation it may also be written

$$A = \frac{L(S)}{N_0} \cdot \frac{\mu}{1+\mu} = \frac{L(S)}{N_0 + \hbar\omega} . \tag{3.88}$$

The last term

$$\alpha = \sum_i \log\left[\frac{1}{r_i!} L_{r_i}(-\gamma_i)\right] = \sum_{i,j} \log\left(\frac{\gamma_i}{\gamma_{jr}} + 1\right), \tag{3.89}$$

*This distribution which was first included in the April 1965 interim report on this contract has later been independently presented by Glauber (1965).

where γ_{jr} is the jth zero of the rth order polynomial, is clearly a monotone function of the likelihood ratio and its value can be used as a basis for a decision on whether or not the observed point in observation space indicates the arrival of a signal. Since all the zeros of the Laguerre polynomials are real and positive, the last member of (3.89) results from a complete factoring of the polynomials in the sum of the second member. It is somewhat more suggestive to write each polynomial in (3.89) as r_i times the average logarithm over all the zeros of the r_i polynomial.

$$\alpha = \sum_i r_i \operatorname{avg}\left[\log\left(\frac{\gamma_i}{\gamma_r} + 1\right)\right] . \tag{3.90}$$

Thus α measures a certain nonlinear correlation between the observed vector $\vec{r}$ and the specified signal S_1. Any term in the summation is zero for r_i and γ_i equal to zero; it grows monotonically but more slowly than linearly with r_i and γ_i.

A very large class of decision functions can be stated in the form

$$D\left[\lambda(r) \gtrless \lambda_c\right] = 1 \quad \text{or} \quad D\left[\alpha \gtrless \alpha_c\right] = 1$$

and

$$D\left[\lambda(r) < \lambda_c\right] = 0 \quad \text{or} \quad D\left[\alpha < \alpha_c\right] = 0 , \tag{3.91}$$

where the threshold values λ_c and α_c may be selected by minimizing a risk function of some kind. According to (3.87) and (3.88) they are related by the equation

$$\log \lambda_c = -\frac{L(S)}{N_o} \cdot \frac{\mu}{\mu+1} + \alpha_c . \tag{3.92}$$

In the analogous case of a classical detector of a known signal in white Gaussian noise the variable corresponding to α is the linear correlation between the observed vector and the known signal; then a matched filter constitutes a perfect analog computer, producing at its output a current or a voltage proportional to the time integral of their product. In the case of the photon counter, the analog computation of an acceptable variable on which to base a decision is not equally straightforward.

The problem can be made more tractable by restricting the class of signals.

Let us limit the choice to binary signals; each sample S_i (γ_i) has only two possible values, zero and S_k (γ_k), respectively. Disregarding the obvious practical obstacles, α can in principle be obtained from $\vec{r}$ in the following way. The receiver input $\vec{r}$ is first transformed by a nonlinear transducer without energy storage, designed to give an output proportional to $L_r(-\gamma_k)$ for a fixed given γ_k. Since all nonzero values of γ are the same, there is no need to compensate for the nonlinearity of the relation between α and γ. The transducer output is applied to a filter matched to the signal, in principle a tapped delay line, giving an output which at $t = T$ is proportional to α. The final component of an automatic receiver is a trigger circuit that observes the filter output at $t = T$ and indicates that a signal was observed, if this output exceeds the value calibrated to correspond to α_c. Otherwise it indicates that no signal was observed.

The performance of a receiver operating on the basis of the decision function (3.91) is evaluated in terms of the relative frequencies of erroneous decisions, "misses" as well as "false alarms". These quantities are simply indicated by the distribution functions of $\lambda(r)$ or $\alpha(r)$ in the presence or absence of a signal, respectively. Since α is the sum of a number of independent random variables of known statistical characteristics, a characteristic-function method may be used to compute these distributions. The characteristic functions are:

$$\phi_{S_o}^{\alpha}(\xi) = \left[\sum_{r=1}^{\infty} \exp \left\{ j\xi \log \left[\frac{1}{r!} L_r(-\gamma_k) \right] \right\} P_{S_o}(r) \right]^{n_k} \tag{3.93}$$

and

$$\phi_{S_1}(\xi) = \left[\sum_{r=1} \exp \left\{ j\xi \log \left[\frac{1}{r!} L_r(-\gamma_k) \right] \right\} P_{S_1}(r) \right]^{n_k}, \tag{3.94}$$

where n_k is a number of equal nonzero signal samples S_k in T. The inverse transformations give the distributions in the respective cases.

For values of n_k between one and ten numerical computations may be performed to give information of the statistical behavior of the channel. In the

particular case $n_k = 1$, i.e., when pulse length, duty cycle and bandwidth are such that only one independent sample pair or channel mode is observed in each observation interval T, the photon count itself is the obvious basis for the decision, and only the distributions $F_{S_0}(r)$ and $F_{S_1}(r)$ are needed for evaluation of the performance. Gordon (1962) presented extensive computations for binary optical channels on this basis, maximizing the rate of transmission of information under the constraint of given average signal power and background radiation. We shall return to a more detailed discussion of the case $n_k = 1$, which can be considerably simplified, after a brief analysis of the opposite extreme, $n_k >> 1$, which permits a few general conclusions of interest.

When n_k is large, the variable α is the sum of a large number of independent random variables with identical distributions. According to the central limit theorem the distribution of α is then approximately standard normal. This permits a rather close comparison with the classical channel perturbed by white Gaussian noise. In the classical case the performance of a likelihood-ratio detector depends on one parameter only, which is the same as A in (3.88) with $\hbar\omega = 0$. In the case of a photon counter for detection of binary signals, on the other hand, there are three independent parameters in (3.80), (3.81), (3.85) and (3.86), which may be chosen:

$$n_k = \eta n \qquad (\eta = \text{duty factor}), \tag{3.95}$$

$$\mu = N_o/\hbar\omega \tag{3.96}$$

and

$$A = \mu\gamma_k n_k = L(S)/(N_o + \hbar\omega). \tag{3.97}$$

In the classical case the single parameter A can alternatively be expressed as the average signal-to-noise ratio multiplied by the number of independent sample pairs. One is tempted to define a tentative effective noise power $(f_2 - f_1)(N_o + \hbar\omega)$ and express A from (3.97) in the analogous manner. However, the signal-to-noise ratio so defined is not equally significant as in the classical case, since the three quantities μ, n_k and γ_k occur not only as a product but also separately in the expression

for the likelihood ratio and its distribution function. That this signal-to-noise ratio under some conditions is unduly pessimistic is suggested by the entropy calculations in the previous section, where the limit of the spectral density of the quantum noise alone was found to be $\hbar\omega/e$ rather than $\hbar\omega$.

For large n_k the approximately Gaussian distribution functions $F_{S_o}(\alpha)$ and $F_{S_1}(\alpha)$ for the variable α in (3.87) and (3.89) have means and variances which we denote by (m_o, σ_o) and (m_1, σ_1), respectively. The logarithm of the likelihood ratio is then

$$\log \frac{dF_{S_1}(\alpha)}{dF_{S_o}(\alpha)} = \frac{(\alpha - m_o)^2}{2\sigma_o^2} - \frac{(\alpha - m_1)^2}{2\sigma_1^2} + \log \frac{\sigma_o}{\sigma_1} = \alpha - A \tag{3.98}$$

where the last member is taken from (3.88). The equality of the second and third members of (3.98) holds for all values of α; consequently equal powers of α must have equal coefficients. The following relations are obtained:

$$\sigma_o = \sigma_1 = \sigma \tag{3.99}$$

$$m_o = A - \frac{1}{2}\sigma^2 \tag{3.100}$$

$$m_1 = A + \frac{1}{2}\sigma^2 \quad . \tag{3.101}$$

Here A is easily evaluated numerically from (3.97); in addition one of the three quantities m_o, m, or σ^2 must be determined in order to specify the approximate distribution functions of α. Figure 3.4 shows a graph of m_o/n_k evaluated from

$$m_o/n_k = \frac{1}{n_k} E_{S_o}(\alpha) = \sum_{r=0}^{\infty} \frac{\mu^r}{(1+\mu)^{r+1}} \log\left[\frac{1}{r!} L_r(-\gamma)\right] \tag{3.102}$$

as a function of $\gamma = \gamma_k$ for various values of μ.

Except for the choice of a threshold value α_c for the decision signal-no signal the error probabilities can be expressed as functions of one single parameter σ, provided that n_k is large enough for the Gaussian approximation to be satisfactory.

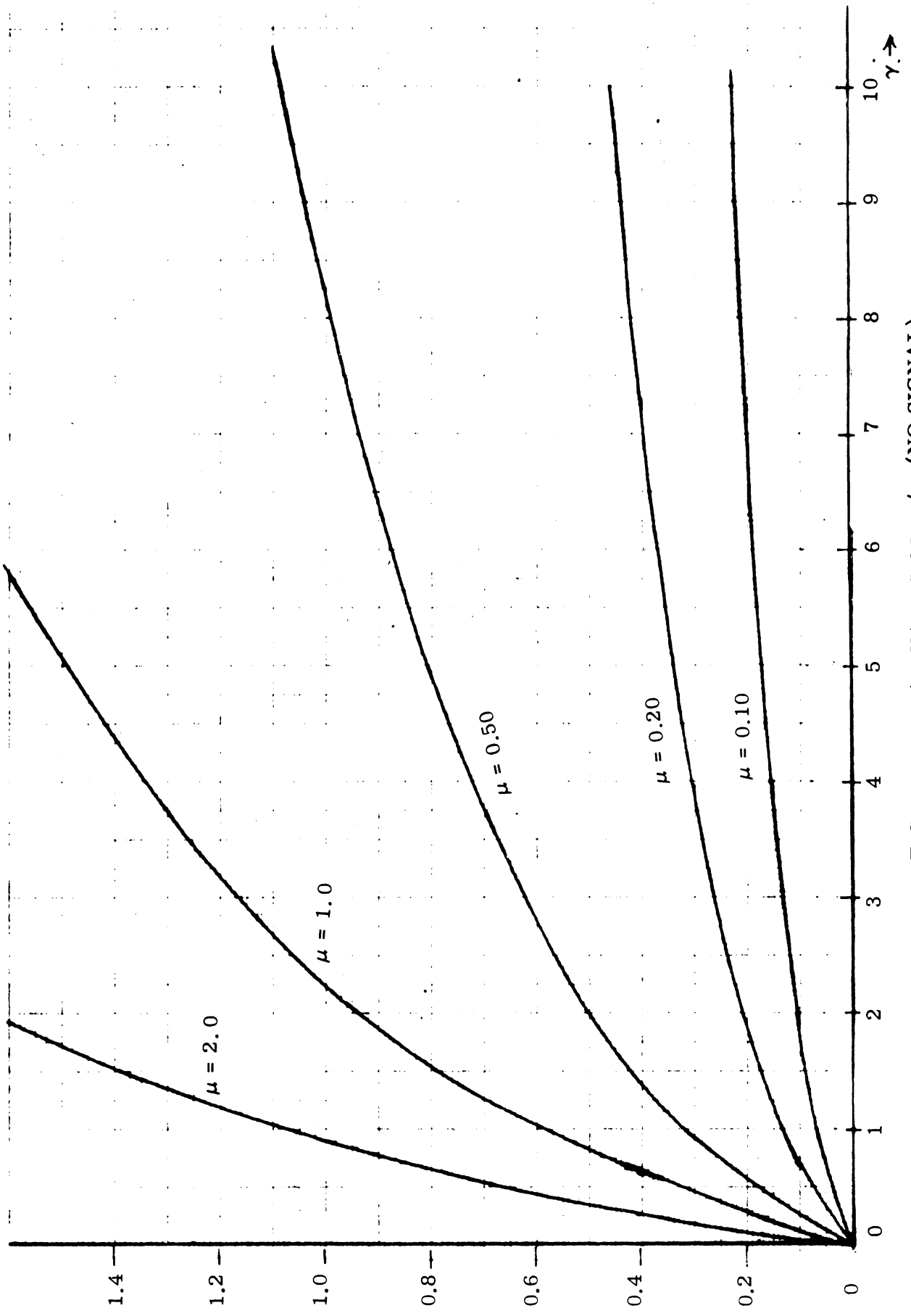

FIG. 3-4. MEAN VALUE OF α/n_k (NO SIGNAL).

Figure 3.5 shows qualitatively the relationship between the two distribution functions in terms of the normalized argument α/σ. The curves are of equal shape and displaced a distance σ from each other. This single parameter is easily found from (3.100).

$$\sigma = \sqrt{2A(1 - \frac{m_o}{A})} = \sqrt{2n_k\gamma\mu(1 - \frac{m_o}{n_k\gamma\mu})}. \tag{3.103}$$

The first factor

$$\sqrt{2A} = \left[\frac{2L(S)}{N_o + \hbar\omega}\right]^{1/2} \tag{3.104}$$

corresponds very closely to the single parameter $\sqrt{2L/N_o}$ in the classical detection problem, the only difference being the $\hbar\omega$ added in the denominator. The important qualitative difference between the classical and the quantum-limited detection problem is expressed by the second factor

$$\sqrt{1 - \frac{m_o}{A}} = \sqrt{1 - \frac{m_o}{n_k\gamma\mu}} \tag{3.105}$$

which severely reduces the distance between the two distributions for small values of γ. Figure 3.6 gives the square of this factor for $\mu = .05$ and 1.0, respectively, plotted vs γ. In Fig. 3.7, $\sigma^2/2n_k$ is shown as a function of γ for a few values of μ.

It is seen from Fig. 3.6 that the factor (3.105) is not very strongly affected by the value of μ except at very small values of γ.

When the Gaussian approximation is good enough, the above results make it now possible to calculate, at least numerically, the error probabilities for any chosen threshold value $\lambda(r) = \lambda_c$. The receiver operating characteristic (Figs. 3.8 and 3.9) is a convenient graph for presenting this information in compact form. The coordinates in this diagram are the conditional complementary distribution functions

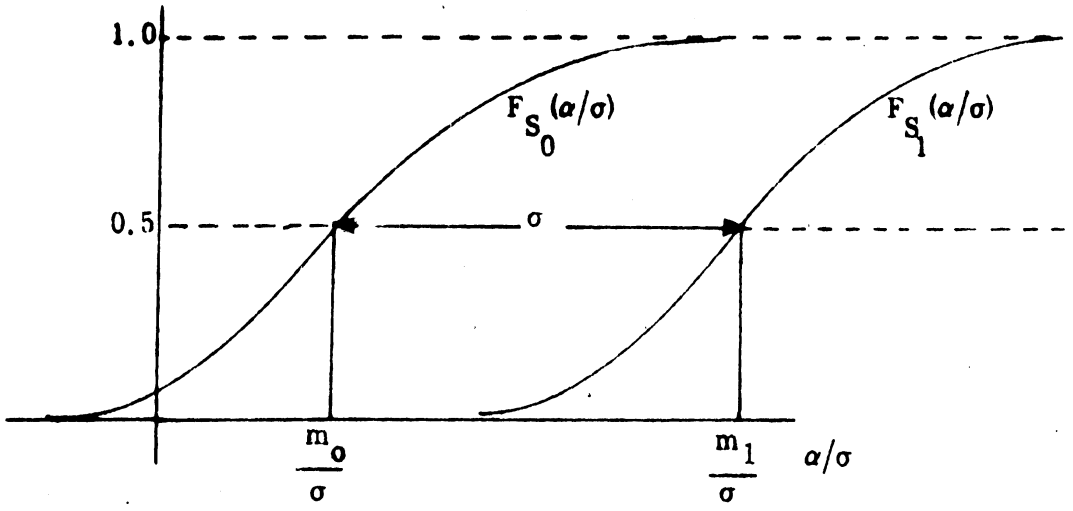

FIG. 3-5. DISTRIBUTION FUNCTIONS α/σ .

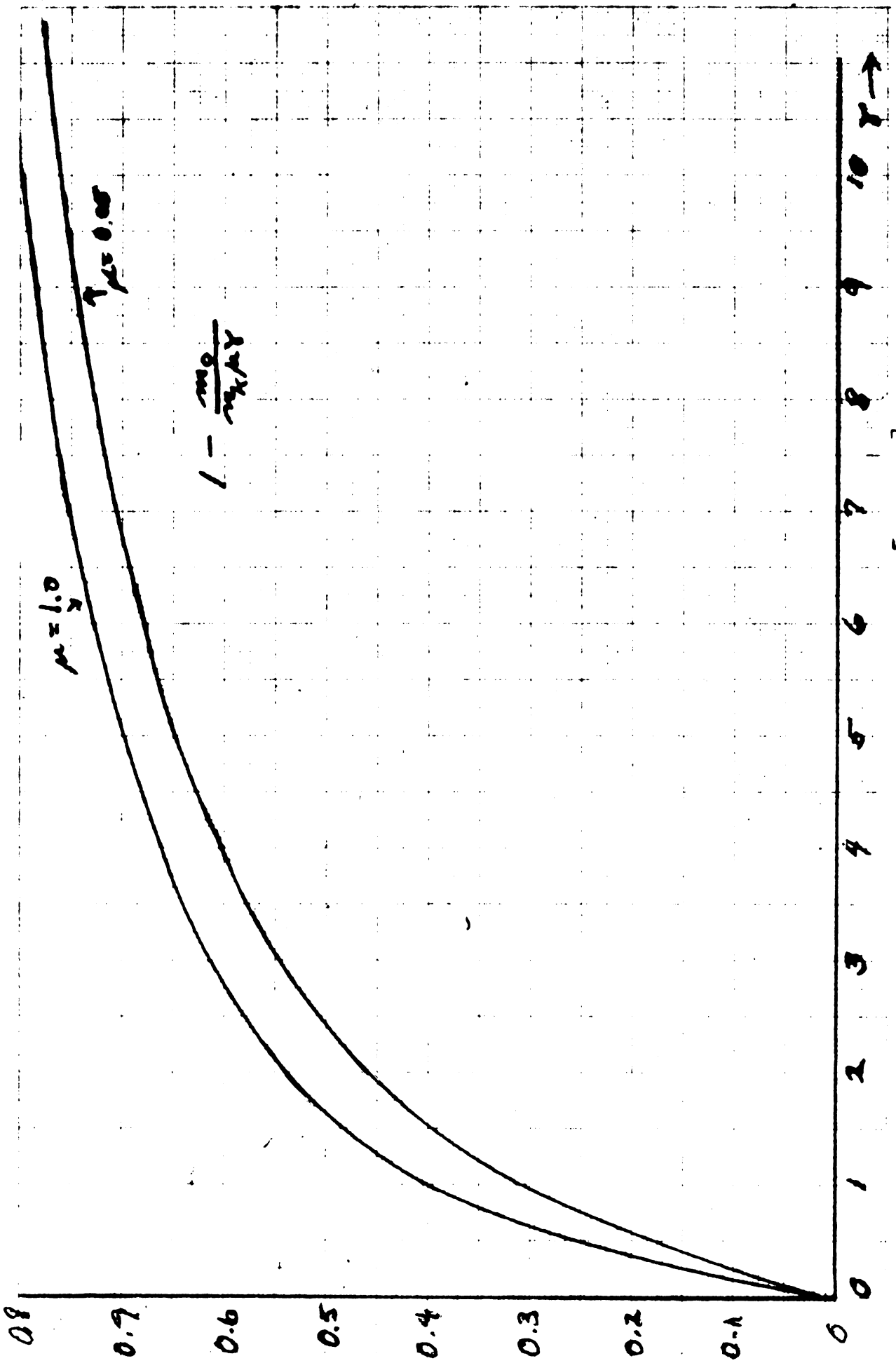

FIG. 3-6. GRAPH OF THE FACTOR $\left[1 - \frac{m_o}{n_L \mu \gamma}\right]$.

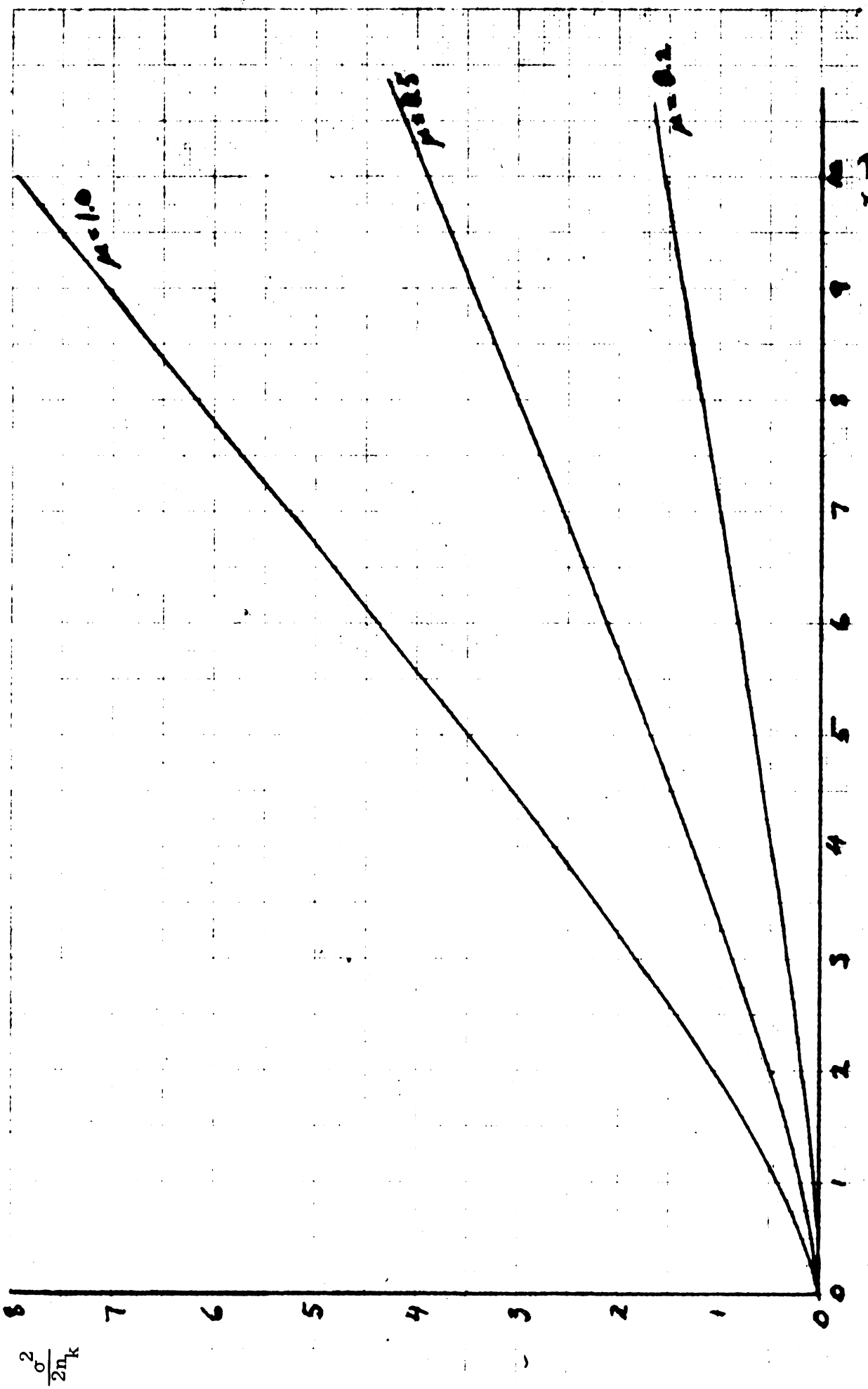

FIG. 3-7. THE VARIANCE OF THE LOGARITHM OF THE LIKELIHOOD RATIO.

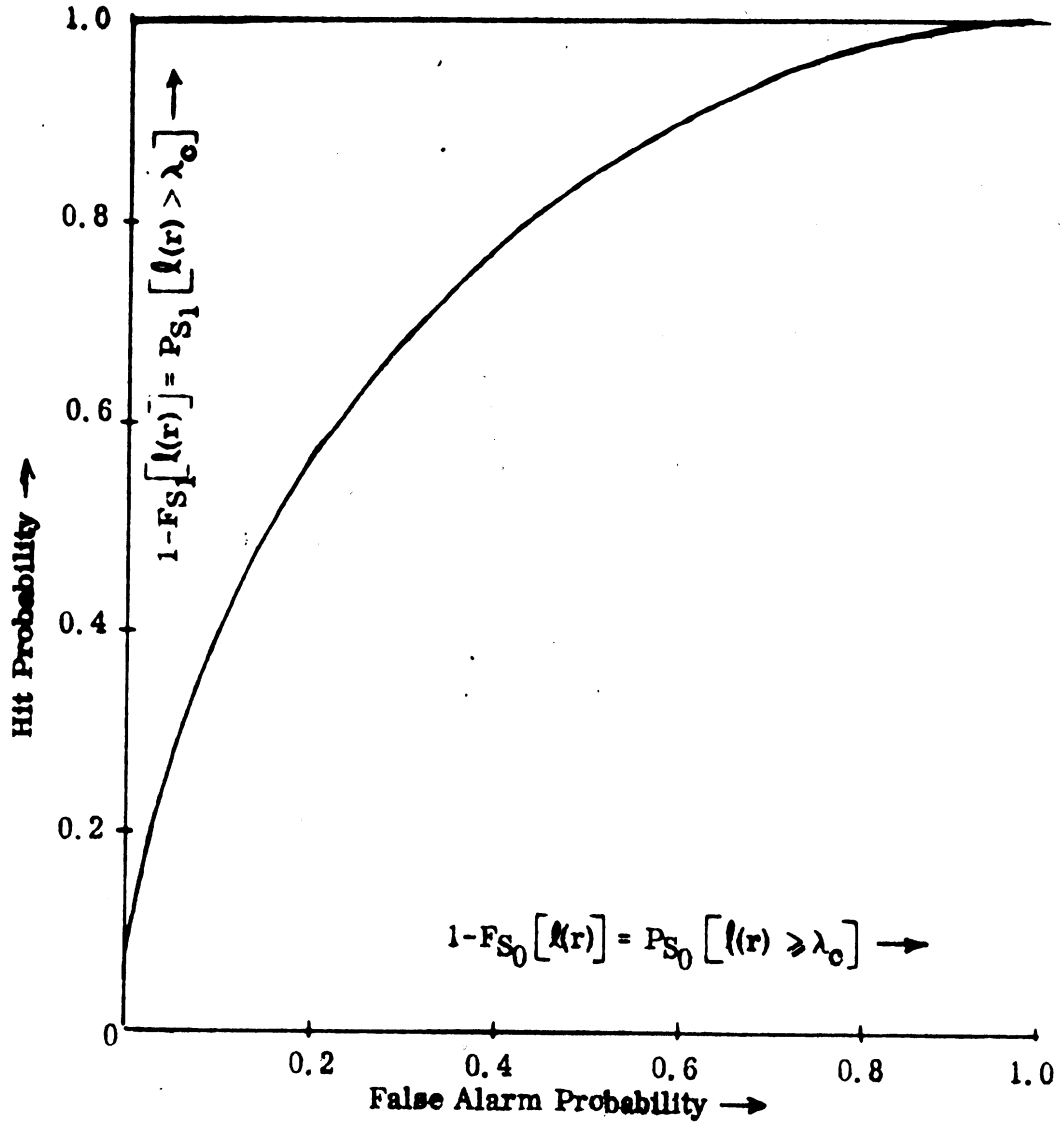

FIG. 3-8: RECEIVER OPERATING CHARACTERISTIC

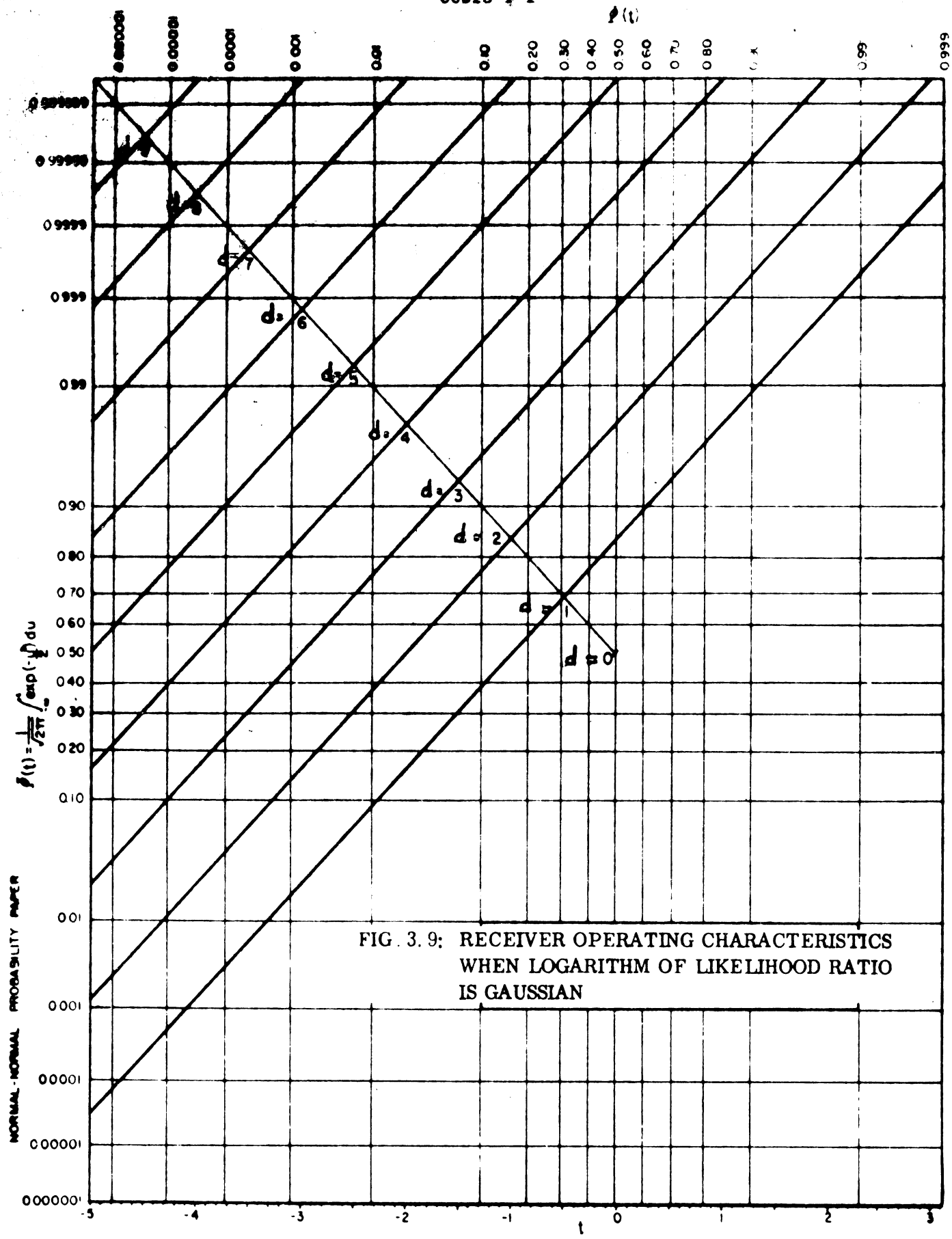

FIG. 3.9: RECEIVER OPERATING CHARACTERISTICS WHEN LOGARITHM OF LIKELIHOOD RATIO IS GAUSSIAN

$$x = P_{S_o}(\mu(r) \geq \lambda_c) = P_{S_o}(F) = 1 - F_{S_o}(\lambda_c) = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha - m_o)^2}{2\sigma^2}\right) d\alpha \qquad (3.106)$$

$$y = P_{S_1}(\mu(r) \geq \lambda_c) = P_{S_1}(H) = 1 - F_{S_1}(\lambda_c) = \int_{\alpha_c}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha - m_1)^2}{2\sigma^2}\right) d\alpha . \qquad (3.107)$$

Because of the relations between m_1, m_2 and σ, this diagram has one single parameter σ which determines the error statistics for any chosen threshold value λ_c. The graph is identical with the standard graph representing detection of a completely specified classical signal in white Gaussian noise, the only difference being the definition of the parameter σ which in Fig. 3.6 is represented by the symbol d. Thus, as soon as σ has been calculated for a given binary channel, the hit and false-alarm probabilities can be chosen from any point on the corresponding curve in the standard graph, each point representing a different value of λ_c (= dy/dx in a linear plot).

The first conclusion suggested by these results is that under the conditions compatible with this Gaussian approximation the optical channel always gives a poorer performance (higher error probabilities) than a classical channel with the same pulse energy and white Gaussian noise of the power

$$N = (N_o + \hbar\omega)\Delta f , \qquad (3.108)$$

since the factor (3.105) in the parameter σ (3.103) only asymptotically reaches the value one. On the other hand, the entropy calculations in the present section indicate that the "quantum noise" under optimum conditions is considerably smaller than this value. The large number n_k of samples in a pulse or code group which is required for the Gaussian approximation consequently appear to be unfavorable.

It is illuminating to state the variation of σ with n_k within the Gaussian domain. The factor $\sqrt{2A}$ depends only on the total pulse energy and on the spectral

density of the "total noise" $N_o + \hbar\omega$. If σ contained this factor alone, it would be immaterial whether the pulse energy were spread over a large number of samples in a wide frequency band or concentrated in a few samples in a narrow band; this would be in complete agreement with the classical Gaussian detection problem. The second factor (3.105) depends on γ which is inversely proportional to the bandwidth but it is independent of n_k. Consequently, if n_k and the bandwidth are reduced proportionally, the first factor remains invariant, while the second increases monotonically. For given pulse energy and noise density the best performance is obtainable with the smallest n_k and the bandwidth compatible with the other constraints of the problem.

It is tempting to generalize this observation by the inference that it is necessary to preserve the characteristic statistics of the photon counter, illustrated by the asymmetric "tree" in Fig. 3.1, in order to optimize the performance. The second term in the entropy formula (3.38) is a component peculiar to such statistics which apparently is lost in any mode of operation that tends to modify the statistics in a Gaussian direction. At low signal levels this second term becomes a substantial fraction of the total and consequently important to recover as far as possible.

The conclusion is then that the first condition toward an optimum mode of operation is the use of pulses, each one formed by a single pair of samples ($n_k = 1$). The variable α in (3.88) is then simply equal to r, the photon count for the pulse. At low signal levels the discrete nature of r reduces the flexibility of the decision process; the threshold value $\alpha_c = r_c = 1$ has to take care of a considerable range of signal levels. The error statistics of the channel for this threshold are given by the following probabilities, illustrated by the diagram Fig. 3.10.

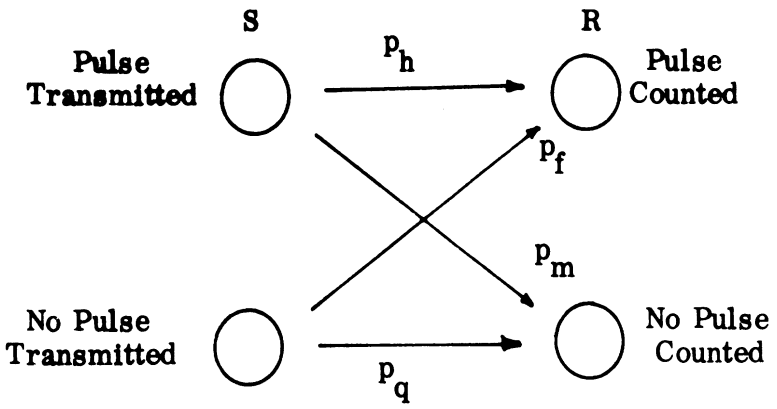

FIG. 3-10. PROBABILITY DIAGRAM FOR BINARY CHANNEL.

"Quiet" $$p_q = P_{S_o}(0) = \frac{1}{1+\mu} = \frac{\hbar\omega}{N_o+\hbar\omega} \tag{3.109}$$

"False Alarm" $$p_f = 1 - P_{S_o}(0) = \frac{\mu}{1+\mu} = \frac{N_o}{N_o+\hbar\omega} \tag{3.110}$$

"Miss" $$p_m = P_{S_1}(0) = \frac{1}{1+\mu}\exp(-\mu\gamma) = \frac{\hbar\omega}{N_o+\hbar\omega}\exp\left[-\frac{L(S)\,N_o}{(N_o+\hbar\omega)\hbar\omega}\right] \tag{3.111}$$

"Hit" $$p_h = 1 - P_{S_1}(0) = 1 - \frac{1}{1+\mu}\exp(-\mu\gamma) = 1 - \frac{\hbar\omega}{N_o+\hbar\omega}\exp\left(-\frac{L(S)\,N_o}{(N_o+\hbar\omega)\hbar\omega}\right) \tag{3.112}$$

If the background noise is thermal radiation of temperature T, these quantities may alternatively be written in terms of the quantity

$$\gamma = \exp\left(-\frac{\hbar\omega}{kT}\right) \tag{3.113}$$

$$p_q = 1 - \gamma = 1 - \exp\left(-\frac{\hbar\omega}{kT}\right) \tag{3.114}$$

$$p_f = \gamma = \exp\left(-\frac{\hbar\omega}{kT}\right) \tag{3.115}$$

$$p_m = p = (1-\gamma)\exp\left\{-\frac{L(S)}{\hbar\omega}(1-\gamma)\right\} = \left[1-\exp\left(-\frac{\hbar\omega}{kT}\right)\right]\exp\left\{-\frac{L(S)}{\hbar\omega}\left[1-\exp\left(-\frac{\hbar\omega}{kT}\right)\right]\right\} \tag{3.116}$$

$$p_h = 1 - p_m = 1 - p \quad . \tag{3.117}$$

These probabilities differ slightly from those presented for the binary channel by Gordon (1962); he postulated Poisson distributions for both signal and noise rather than the distributions derived above (3.81) and (3.82). For the case of negligible noise ($\gamma \approx 0$),, however, the differences vanish.

The asymmetric character of the channel is here indicated by the fact that in

general one type of error is much more probable than the other, i.e., $p_m >> p_f$. It is intuitively obvious that the error rate can be minimized by choice of a binary code that makes a pulse much less frequent than a space. The prior probability Q of a pulse is also the duty cycle of the pulse train, and for given average power the energy per pulse is inversely proportional to Q.

The joint event $[S, R]$ of a pulse or a space being sent and a pulse or a space being received then has the probability matrix

$$\begin{vmatrix} P[0,0] & P[0,1] \\ P[1,0] & P[1,1] \end{vmatrix} = \begin{vmatrix} (1-Q)p_q & (1-Q)p_f \\ Qp_m & Qp_h \end{vmatrix} =$$

$$= \begin{vmatrix} (1-Q)(1-\gamma) & (1-Q)\gamma \\ Qp & Q(1-p) \end{vmatrix} . \qquad (3.118)$$

The rate of transmission of information per digit may be written as the sum of the entropies of input and output minus the joint entropy of input and output.

$$H = H(S) + H(R) - H(S,R)$$

$$= -\log\left[1 - P(S_1)\right] - P(S_1)\log\frac{P(S_1)}{1-P(S_1)}$$

$$-\log\left[1 - P(R_1)\right] - P(R_1)\log\frac{P(R_1)}{1-P(R_1)}$$

$$+ P(0,0)\log P(0,0) + P(0,1)\log P(0,1)$$

$$+ P(1,0)\log P(1,0) + P(1,1)\log P(1,1) . \qquad (3.119)$$

An optimum value of the duty cycle Q may be found by setting the derivative of H with respect to Q equal to zero under the constraint of given average power, i.e., a number of photons per pulse inversely proportional to Q

$$\frac{L(S)}{\hbar\omega} = \frac{\beta}{Q} = r . \qquad (3.120)$$

Thus β is the average number of signal photons per sample and r the number of signal photons per pulse.

The resulting equation for the optimum Q or r may be written as follows:

$$\log\left[\frac{\frac{1}{\beta}\log\frac{1-\gamma}{\gamma}-1+p+\gamma}{1+\frac{\gamma}{1-p}\left(\frac{1}{\beta(1-\gamma)}\log\frac{1-\gamma}{p}-1\right)}\right] = \frac{p\log\frac{1}{p}\log\frac{e(1-\gamma)}{p}}{1-\gamma-p\log\frac{e(1-\gamma)}{p}} = \psi(p) \qquad (3.121)$$

where

$$p = (1-\gamma)\exp\left[-r(1-\gamma)\right] = (1-\gamma)\exp\left[-\frac{\beta}{Q}(1-\gamma)\right] . \qquad (3.122)$$

It is clearly not feasible to express any one of the variables Q, r or p as a single function of the parameters β and γ. However, a graph mapping the numerical solution may equally well be plotted from an inverse function $\beta = \beta(p, \gamma)$ which may be written as follows:

$$\beta = \frac{\log\frac{1-\gamma}{p}\left[1-\frac{\gamma e^{\psi}}{(1-\gamma)(1-p)}\right]}{(1-p-\gamma)\left[1+\frac{e^{\psi}}{1-p}\right]} \qquad (3.123)$$

where ψ is the right-hand side of (3.121).

In the case that the background radiation is negligible ($\gamma << 1$), this expression simplifies to

$$\beta = r\left\{1-e^{-r}+\exp\left[\frac{r(r+1)}{e^{r}-r-1}\right]\right\}^{-1} \qquad (3.124)$$

which is the same as the function given by Gordon (1962). His graph is reproduced in Fig. 3.11.

It is helpful at this point to illustrate the result by some numerical examples. Take for instance a binary channel operating under the conditions: $\beta = 0.06$, $Q = 0.06$, $r = 1$, $\gamma = 0$. The theoretical capacity of a wave with this average energy

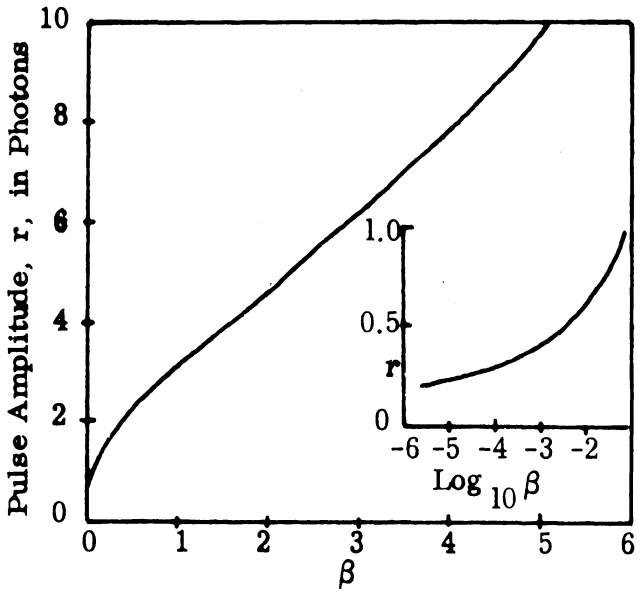

FIG 3.11: OPTIMIZED AVERAGE RECEIVED PULSE AMPLITUDE FOR NOISELESS BINARY CHANNEL AS FUNCTION OF AVERAGE NUMBER OF RECEIVED PHOTONS PER AVAILABLE TIME INTERVAL. The probability of sending a pulse is given by $Q = \beta/r$.

per sample is 0.333 bits per sample pair while the binary channel at the transmitting end has an entropy of 0.328 bits per sample pair. The detection process recovers 0.176 bits per sample pair under ideal circumstances. The difference between the latter two figures 0.152 bits constitutes the equivocation introduced by the random detection process. The recoverable entropy is consequently about 53 percent of the maximum input entropy at the given power level.

The coding problem, which remains to be solved and which will be discussed in the next section, is to design a code that will remove the errors in the received message caused by the missing pulses. In this example the probability of a miss is $p = e^{-r} = 0.368$. In a hundred sample-pair intervals there will be on the average 16.7 pulses transmitted, of which 6.2 are missed by the detector. This is clearly a very high percentage of errors which would require a highly redundant code with very long code sequences and elaborate computer programs for coding and decoding for correction of the errors without substantial loss in rate of transmission.

An extreme example is obtained by taking $r = 0.5$ in (3.125). This represents an average level of $\beta = 0.0032$ photons per sample pair or an optimum Q of 0.0064. The percentage recoverable entropy has increased to 76.5 percent which is, however, only 0.022 of a bit per sample pair. The probability of a miss is now $e^{-0.5} = 0.606$. A sequence of a thousand pulse intervals will on the average have 6.4 transmitted pulses out of which 3.9 will be missed.

Going in the opposite direction to $r = 2$, we obtain $\beta = 0.416$ photons per sample pair and an optimum $Q = 0.208$. The error statistics are now more reasonable: out of a 100 pulse intervals 20.8 pulses are transmitted and 2.8 missed, and 45 percent of the channel capacity is still theoretically obtainable. The transmitted pulse train has a bit rate of 0.92 per pulse, while the theoretical channel capacity at this pulse level is 1.24 bits per pulse. 45 percent of this is 0.56 bit per pulse. At this level the first term of (3.38), which according to Gordon's analysis

is the fraction that theoretically can be recovered by a superheterodyne method of reception, is only 41 percent of the total. The photon counter retains an advantage at this power level, although it is not a very large one. The error frequency decreases here considerably faster than the "efficiency." For this reason a reduction of the bandwidth is a rather inexpensive way of making the coding problem easier and less costly to solve.

Another way of reducing the error frequency at a slight cost in channel capacity is to use a somewhat smaller Q than the optimum value, if the properties of the laser transmitter permit operation with a low duty cycle and corresponding high pulse energy. Because the tangent to the entropy function at the maximum is horizontal, the Q-value is not very critical, as far as the entropy is concerned, while it affects the error probability more rapidly.

The comparison made above with a superheterodyne method of detection can easily be verified by a modification of the statistical detection theory presented earlier in this section.

The high-level local-oscillator field may be taken as a common component of every signal represented by a point in signal space $\mathcal{E}_S$ in Fig. 3.3. For a binary detection problem the no-signal subset S_0 has one point representing a field with the constant expectation amplitude S_0 and the oscillator frequency ω_0. Similarly the signal subset S_1 has one single point representing the sum of the oscillator field and a signal field of frequency ω and expectation amplitude S_1 ($S_1 << S_0$); the resulting envelope squared has a steady component of $\frac{1}{2}(S_0^2 + S_1^2)$ and a sine wave component of amplitude $S_0 S_1$ and the frequency $|\omega_0 - \omega| = \omega_1$. In the pre-observation space the expectation vectors of thermal radiation are added to the signal vectors as in the previous case. The distributions in observation space are then :

$$P_{S_0}(\vec{r}) = \prod_{i=0}^{n} \exp(-\mu\gamma_0) \cdot \frac{1}{r_i!} L_{r_i}(-\gamma_0) \frac{\mu^{r_i}}{(1+\mu)^{r_i+1}} \tag{3.125}$$

$$P_{S_1}(\vec{r}) = \prod_{i=o}^{n} \exp\left[-\mu(\gamma_o+\gamma_i)\right] \frac{1}{r_i!} L_{r_i}(-\gamma_o-\gamma_i) \frac{\mu^{r_i}}{(1+\mu)^{r_i+1}} \tag{3.126}$$

where in analogy with (3.83)

$$\gamma_o = \frac{S_o^2}{2N(1+\mu)} \tag{3.127}$$

$$\gamma_o + \gamma_i = \frac{S_o^2 + 2S_oS_1\cos\omega_1 t_i + S_1^2}{2N(1+\mu)} \quad . \tag{3.128}$$

The number n of envelope samples or normal modes of the field incident on the photon counter must now be chosen with regard to the intermediate frequency rather than the signal bandwidth, since a wave of the frequency ω_1 is to be reproduced in the counter output.

From (3.125) and (3.126) the likelihood ratio is now found to be

$$\lambda(\vec{r}) = \prod_{i=o}^{n} \exp(-\mu\gamma_i) \frac{L_{r_i}(-\gamma_o-\gamma_i)}{L_{r_i}(-\gamma_o)} \tag{3.129}$$

and its logarithm

$$\log\lambda(\vec{r}) = \prod_{i=1}^{n} -\mu\gamma_i + \sum_{i=1}^{n}\sum_{j=1}^{r_i} \log\left(1+\frac{\gamma_i}{\gamma_o+\gamma_{r_j}}\right) \quad . \tag{3.130}$$

As in the previous case (3.89) the Laguerre polynomial in (3.129) has been broken down into factors of the first degree, γ_{rj} being the j^{th} root of $L_r(x) = 0$.

The variable γ_i (3.128) has both a first-order and a second-order small component; to facilitate a consistent approximation, expand the logarithm in (3.130) retaining two terms. The result is

$$\log \lambda(\vec{r}) = \sum_{i=1}^{n} \frac{S_1^2}{2(N_o + \hbar\omega)} + \sum_{i=1}^{n} \frac{2S_1 \cos \omega_1 t_i}{S_o} \sum_{j=1}^{r_i} \frac{1}{1 + \frac{\gamma_{r_j}}{\gamma_o}}$$

$$+ \sum_{i=1}^{n} \frac{S_1^2}{S_o^2} \sum_{j=1}^{r_i} \frac{\gamma_o \gamma_{r_j}}{(\gamma_o + \gamma_{r_j})^2}$$

$$= -A + \alpha \, . \qquad (3.131)$$

Since n in this case is necessarily rather large, the random variable α will be very nearly Gaussian, and we can find its distribution functions if we determine the mean m_o in the absence of a signal. Because of the cosine factors, the second summation has mean zero for identical r_i - statistics, and the last summation is negligible. Consequently $m_o = 0$, and the problem reduces to exactly the same form as a classical known signal in white Gaussian noise of spectral density $N_o + \hbar\omega$, the characteristic parameter being

$$\sigma = \sqrt{2A} = \sqrt{\frac{2L(S)}{N_o + \hbar\omega}} \qquad (3.132)$$

i.e., the square root of twice the pulse energy divided by an equivalent noise power per cycle bandwidth.

Again the result indicates that it is the Gaussian statistics introduced by the necessity of averaging over a large number of events rather than counting single discrete events which reduces the recoverable entropy to the first term in the equation (3.57). Extrapolation of these conclusions down to pulses of only one or two photons, however, does not look reasonable. The correlation between the number of counts r and the known signal measured by α in (3.131) is not very significant when the total count during the whole pulse interval differs only by one or

two units in the presence or absence of a signal, respectively. In any case, the conclusion that the superheterodyne method of detection is unsuitable for reception of marginal optical signals is certainly well justified.

The discussion in this section of detection of small light pulses by photon counters has so far assumed an ideal photon counter, i.e. a device with no other limitations than those imposed by the basic postulates of quantum mechanics. Other limitations are encountered in any practical realization, primarily of the nature of the state of the art, imperfect materials, etc. A vacuum photo-emission tube has a "quantum efficiency" less than 100 percent; this can easily be accounted for simply by reducing the expected values of all signal and noise powers in preobservation space by this factor. More serious is the "dark current" in the photocell caused by thermionic emission, leakage currents, etc. which causes spurious counts and raises the false-alarm probability, which is a very vulnerable point in the processing of signals with average powers of less than one photon per sample. Cooling of the emitter and careful design and processing techniques can keep this source of error to a minimum. The situation is consequently reasonably satisfactory in the wavelength region above about 1μ, where photo-emission cells with good quantum efficiency are available. In the far infrared, however, where semi-conductor devices offer the most promising performance, the reduction of the dark current to the point where single photons can be counted with reasonable accuracy appears to be a much more difficult problem. Otherwise the semiconductor photocells have the advantage of offering very nearly 100 percent quantum efficiency, and recent developments indicate that the bandwidth problem is not as serious as has been anticipated.

In a photoemissive cell the "count" is represented by the emission of an electron. It remains to count the electrons, which on a single-event basis is no trivial task, even if it is a classical task without any absolute limitation similar to the quantum limitations on counting photons.

In the photomultiplier tube each electron emitted by the photocathode is

"multiplied" by a number of cascaded secondary-emission processes. At the output terminals the pulses are large enough to be measured or recorded by ordinary methods. Also the secondary emitters have a dark current, so that the false alarm rate would be very high if all the pulses were counted. However, those pulses originating in the multiplier section are likely to be of smaller amplitude than the original pulses from the photo cathode. The error rate can consequently be kept down, if all pulses below a certain amplitude are ignored. Extremely low dark counting rates as low as one in ten minutes have been reported in the literature for tubes specially designed for counting single-electron events (Eberhardt, 1965, Baum, 1962).

The photomultiplier preserves and explores the digital nature of a low-level optical signal. More conventional broad-band electronic amplifiers have a noise factor appreciably different from unity; they provide a noise background which may add appreciably to the false-alarm rate.

Anticipating a few facts from the last section of this chapter, some remarks may be made about the use of a combination of a laser amplifier and a photocell as a photon counter. The laser amplifies the signal as well as background and quantum noise; in addition it provides noise by spontaneous emission. Theoretically we would expect at least that an ideal amplifier and an ideal counter would perform as well as the counter alone. The situation in the amplifier is somewhat analogous to the electron multiplier: a noise pulse generated somewhere in the active material will not travel the full length of the amplifier; it will not be amplified as much as the light pulses entering from the antenna. Amplitude discrimination should thus be able to eliminate most of the spurious pulses. The remaining advantage of the laser amplifier is its selective gain: only signals and noise within its narrow bandwidth will be amplified.

3.5 Coding and Decoding of Optical Channels

Within the limits of available time and manpower it has not been possible to devote as much attention to the question of coding in optical channels as the subject

deserves. The results developed in the last few years for binary symmetric channels, or more specifically, for classical signals perturbed by white Gaussian noise, are not directly applicable because of the different statistical characteristics of a quantum-limited channel operating under conditions of less than one photon per sample.

The cause of the different behavior of the optical channel is of course the discrete photon statistics that dominates the detection process at low signal levels. The nearly complete absence of false-alarm type errors makes it advantageous to concentrate the signal energy in short pulses rather than to spread it uniformly over the whole signal interval, which in the classical Gaussian channel is equally satisfactory.

The optimum pulse length for given average power was discussed in a previous section. This quantity determines also the bandwidth, since the pulse should also be a single envelope sample. The remaining problem is to design a code which permits the transmission of information with a sufficiently small frequency of errors.

It may be of interest to consider for background information the performance of some standard binary codes for classical channels perturbed by white Gaussian noise. Tables and graphs given by Viterbi (Golomb et al, 1964), show that for $L(S)/N_0 = 1$, which is the marginal cut-off condition discussed in connection with (3.61) in the previous section, the error probability decreases very slowly with increasing code length, from about 0.15 uncoded to about 0.05 for a length of 2^{10} of an orthogonal or transorthogonal code.

In a quantum-limited channel with negligible background radiation the quantity β in (3.120) and (3.124) may be expected to play a similar part to $L(S)/N_0$. The numerical examples given subsequent to (3.124) show a much more favorable situation. With $\beta = 0.06$ the overall error frequency is no more than 6.2 percent but when a pulse is actually transmitted, the pulse energy over $\hbar\omega$ is $r = 1$ and the probability of a miss is 0.368. Considering the last two figures, a rough guess is

that the convergence in error probability for increasing code length is no better than the Gaussian channel for $L(S)/N_0 = 1$.

For $L(S)/N_0 = 2$, the code tables indicate a much better convergence for the Gaussian channel, and our last example of a quantum-limited channel with $r = 2$ also seems more promising. We shall come back later and discuss an error-correcting code for this case.

Slepian's permutation modulation (Slepian, 1965) offers a coding-decoding principle that is in no way restricted to symmetric channel statistics. We shall present a modified version of this principle, adapted to a low-duty-cycle off-on operation.

Consider a sequence of n digit intervals. Fill the first m of them with pulses so that the desired duty cycle $Q = m/n$ is obtained. All different permutations of this basic array of digits now constitute the ensemble of code words, numbering:

$$M = \frac{n!.}{(n-m)!\,m!}$$

Redundancy is introduced by including only words with exactly m pulses rather than all binary words used with such probabilities that the expected value of the number of pulses is m. This constraint serves an analogous purpose to the parity relation in parity codes.

In the absence of background errors, this simple code only detects missing pulses but cannot correct the errors.

In case of multiphoton pulses and background noise, the receiver first records the energy or likelihood ratio for all the n intervals of a sequence, then assigns pulses where the m largest values are located. The result is a member of the code ensemble and can be decoded on an one-to-one basis. This procedure leads to a maximum-likelihood decoding.

By adding digits in more than one dimension, additional constraints and more information about the errors can be achieved. If n/m is a reasonably small integer,

a two-dimensional permutation code may be generated by the matrix shown in Fig. 3.12. All different permutations of rows and columns form the vocabulary of the code.

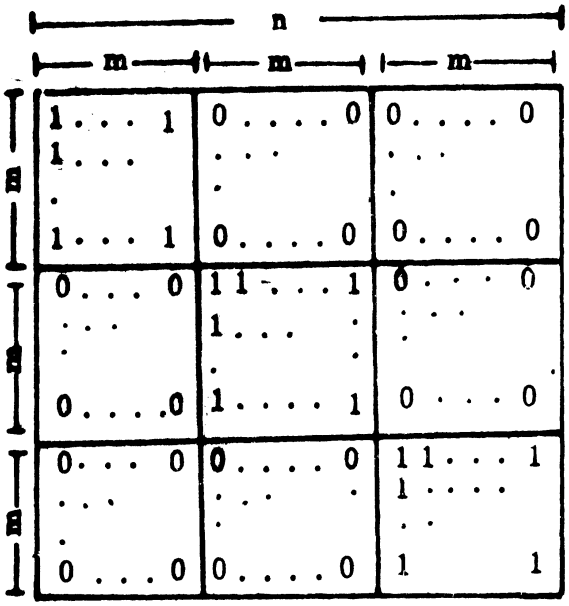

FIG. 3.12. Matrix for Two-Dimensional Code.

With no false alarms, such a code corrects a single error which is located at the intersection of the row and column for which the digit sum is m - 1. It detects the presence of multiple errors, but the intersections cannot unambiguously indicate the location of the errors in that case.

For a channel with multiphoton pulses and background noise the receiver constructs two matrices for each sequence and applies the maximum likelihood procedures to the rows of one and the columns of the other. Lack of coincidence indicates errors. Coincidence should be achieved by erasing one or more of the lowest 1-digits and insert 1-digits for the same number of zeros with the highest likelihood ratio.

As an example, let us consider the last numerical illustration mentioned above. Q is rounded off to 0.20; taking the rows and columns of the length 5 digits, there will be one pulse per row and column, each with a miss probability somewhat smaller than $e^{-2} = 0.135$. Then the probability that the whole square will be received without error is

$$P(0) = \exp(-5 \quad 0.135) = .509 .$$

The probability of one error in a block of 25 digits is

$$P(1) = .675\, e^{-.675} = 0.35 .$$

After correction of the single errors, on the average 86 percent of the blocks are rendered correctly, and the code indicates which blocks belong to the 14 percent containing multiple errors. The price paid for this error reduction is a redundancy of

$$1 - \frac{\log 5!}{-25\left[Q \log Q + (1-Q)\log(1-Q)\right]} = 62.5\,\%$$

in comparison with a binary channel with independent digits, selected at random with the probabilities Q and 1 - Q.

If a channel performance on this basis is not acceptable, it is necessary either to set the sights considerably lower as far as rate of transmissions is concerned, boost the average power, or to develop an asymmetric class of algebraic codes which permits more flexibility than the simple code discussed above, and makes it possible to operate with a higher initial error frequency and negligible errors after decoding. Time and manpower limitations have made it impossible to devote appreciable effort to this general problem.

3.6 A Theory of Quantum Amplifiers

3.6.1 Introduction

Quantum amplifiers have received considerable attention in the literature. Shimoda et al (1957) have presented an analysis of the amplification and fluctuations

of the number of photons. Louisell et al (1961) and Gordon et al (1963a, 1963b) have studied the amplification and fluctuations of the field amplitudes as well as the number of photons, and have derived expressions for the probability distributions of the field amplitudes at the output, for various forms of input fields.

The present analysis is aimed particularly at the traveling-wave quantum amplifier, although the model can be easily adapted to other types of amplifiers. The approach is based on the density operator method. Equations for the density operator of the electromagnetic radiation are derived and solved. The basic approximation is that the amplification is assumed to be linear. This means that the analysis will not be valid if the field becomes strong enough to saturate the gain. Among the novel features of this study is that the broadening of the energy levels of the active material is taken into account. This broadening gives the amplifier a finite bandwidth.

The motivation for this analysis stems from the possibility of using the amplifier in an infrared or optical frequency communication channel. It is known that, at infrared and optical frequencies, the spontaneous emission noise of the amplifier is relatively large. An important question, therefore, is whether and under what conditions the use of the amplifier improves the signal detection despite the spontaneous emission noise. In view of the gain selectivity provided by the narrow amplification bandwidth, it is expected that the amplifier used in front of the detector will improve the signal-to-noise ratio by reducing the fraction of the incident noise reaching the detector. Although our main interest is directed toward a channel based on energy measurements, we do give some consideration to the amplification and fluctuations of field amplitudes.

The present study of the quantum amplifier is by no means complete. A very important aspect, namely the nonlinear behavior of the amplifier, has been left out. This phenomenon will presumably play a significant role in future applications

of laser amplifiers in communications. It is felt that the present approach provides a promising starting point for further understanding of the nonlinear behavior of quantum amplifiers.

3.6.2 Formulation of Problem

The physical problem we wish to study can be represented schematically as in Fig. 3.6.13. Electromagnetic energy carrying information enters the input of the amplifier. The state of the input field is represented by a density operator $\rho^R(t_o)$

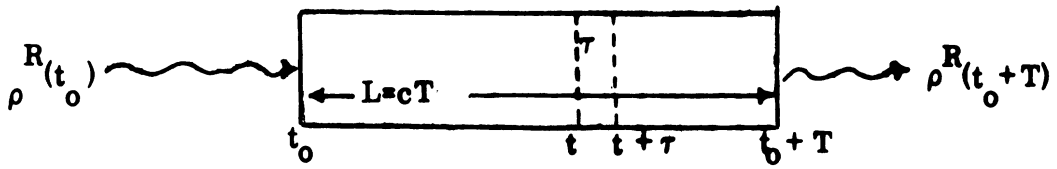

FIG. 3.6.13 QUANTUM AMPLIFIER SCHEMATIC.

and the state of the output by $\rho^R(t_o+T)$. The length of the amplifier is L. The signal enters at time t_o and leaves the amplifier at time t_o+T. Assuming that dispersion can be neglected, we shall have

$$L = cT \tag{3.133}$$

where c is the speed of light. Knowing the density operator of the signal constitutes as complete a knowledge of the state of the signal as is allowed by quantum theory. The problem we wish to solve, therefore, is to determine $\rho^R(t_o+T)$ in terms of $\rho^R(t_o)$ and the characteristics of the amplifier. The latter is an assemblage of material particles grouped in atoms or molecules. For simplicity, we shall use the term atoms to denote both. It is assumed that the atoms do not interact with each other, although they may interact with external fields, such as pumping fields, and with atoms of different species that may be present. In fact, we do assume that there is a second species of atoms present whose Hamiltonian will be denoted by H^P. This species will be referred to as the

"perturber." The first species, whose Hamiltonian will be denoted by H^A will be referred to as the active material. Let V^c be the interaction between the two. This interaction causes the broadening of the levels of H^A. Let furthermore, H^R be the Hamiltonian of the radiation field, and $A(\underline{r})$ the vector potential. Then, the total non-relativisitic Hamiltonian is

$$H = H^R + H^A + H^P + V^c - \sum_\sigma \frac{e_\sigma}{m_\sigma c} \underline{p}_\sigma \cdot A(\underline{r}^\sigma) + \sum_\sigma \frac{e_\sigma^2}{2m_\sigma c^2} \underline{A}^2(\underline{r}^\sigma) , \qquad (3.134)$$

where the index σ refers to the σ^{th} particle, e_σ and m_σ are its charge and mass, and $\underline{r}^\sigma$, $\underline{p}_\sigma$ its position and momentum operators, respectively. The net macroscopic charge density is assumed to vanish so that we need consider only transverse fields.

Let now $\{\underline{u}_k(\underline{r})\}$ be an orthonormal complete set of eigenvectors appropriate to the problem. These vectors satisfy the equations

$$\underline{\nabla} \cdot \underline{u}_k(\underline{r}) = 0 , \qquad (3.135)$$

and

$$\nabla^2 \underline{u}_k(\underline{r}) + \frac{\omega_k^2}{c^2} \underline{u}_k(\underline{r}) = 0 , \qquad (3.136)$$

where

$$\omega_k = ck . \qquad (3.137)$$

If, for example, $\underline{u}_k$ are the free space eigenvectors, we shall have

$$\underline{u}_k(\underline{r}) = (2\pi)^{-3/2} \underline{\epsilon}_k e^{i\underline{k}\cdot\underline{r}} , \qquad (3.138)$$

where the index k is assumed to contain the polarization index as well. In terms of the eigenvectors $\underline{u}_k$, the vector potential operator becomes

$$\underline{A}(\underline{r}) = c \sum_k \left(\frac{\hbar}{2\omega_k}\right)^{1/2} \left\{ a_k \underline{u}_k(\underline{r}) + a_k^\dagger \underline{u}_k^* (\underline{r}) \right\} \tag{3.139}$$

where a_k and $a_k^\dagger$ are the usual annihilation and creation operators, obeying the commutation relations

$$\left[a_k, a_{k'}^\dagger \right] = \delta_{kk'} , \tag{3.140}$$

and

$$\left[a_k, a_{k'} \right] = \left[a_k^\dagger, a_{k'}^\dagger \right] = 0 . \tag{3.141}$$

Processes in which more than one photon are simultaneously emitted or absorbed, or processes in which photons are scattered do not have any importance in laser amplifiers. Thus, we may dispense with the term $\underline{A}^2$. The total Hamiltonian now reads

$$H = H^R + H^A + H^P + V^C + \sum_k d_k (a_k + a_k^\dagger) , \tag{3.142}$$

where

$$H^R = \sum_k \hbar\omega_k (\alpha_k^\dagger \alpha_k + \frac{1}{2}) , \tag{3.143}$$

and d_k is a particle operator defined by

$$d_k \equiv - \sum_\sigma \frac{e_\sigma}{m_\sigma} \left(\frac{\hbar}{2\omega_k} \right)^{1/2} \left(p_\sigma \cdot \underline{\epsilon}_k \right) . \tag{3.144}$$

In defining d_k, we have used the dipole approximation and have replaced $e^{i\underline{k}\cdot\underline{r}}$ by 1. Hence, d_k is the dynamic electric dipole moment operator. Note that it is the collective dipole moment operator since it contains a summation over all particles. This summation extends over electrons as well as nuclei. One should group this sum into partial sums each of them containing the particles of each atom. Since the atoms are assumed to be uncorrelated, one can write

$$d_k = \sum_{j=1}^{N} d_k^{(j)}, \tag{3.145}$$

where $d_k^{(j)}$ is the dipole moment operator of the j^{th} atom. Strictly speaking, $d_k^{(j)}$ is the projection of the dipole operator on the polarization vector ε_k of the k^{th} mode.

The problem has thus been reduced, as usual, to the interaction of an assemblage of harmonic oscillators with an assemblage of atoms. As long as mode coupling does not play a significant role, one can study the amplification of a single mode without losing much of the physics of the problem. Limiting this treatment to the case of a single mode, Eqs. 3.142 and 3.143 becomes

$$H = H^R + H^A + H^P + V^C + d\,(a^\dagger + a) \tag{3.146}$$

and

$$H^R = \hbar\omega(a^\dagger a + \frac{1}{2}), \tag{3.147}$$

where we have dropped the mode subscript k for simplicity. Thus, ω denotes the frequency of the field mode.

3.6.3 Time Evolution of the Density Operator .

Let $\rho(t)$ be the density operator of the compound system (harmonic oscillator plus atoms) at time t. At a later time $t + \tau$, the density operator is given by (see, for example, Margenau and Murphy, 1963)

$$\rho(t+\tau) = U(\tau)\,\rho(t)\,U^\dagger(\tau) \tag{3.148}$$

where $U(\tau)$ is the time evolution operator of the system and is given by

$$U(\tau) = e^{-\frac{i}{\hbar}H\tau}. \tag{3.149}$$

Let us also introduce

$$H^B \equiv H^A + H^P + V^C \,, \tag{3.150}$$

$$H^O \equiv H^R + H^B \,, \tag{3.151}$$

and

$$V \equiv d(a^\dagger + a) \,. \tag{3.152}$$

Then, the total Hamiltonian becomes

$$H = H^O + V \,. \tag{3.153}$$

Now, consider the representations $\{|n\rangle\}$ and $\{|\beta\rangle\}$ defined by

$$a^\dagger a\,|n\rangle = n\,|n\rangle \,, \qquad n = 0, 1, 2 \tag{3.154}$$

and

$$H^B\,|\beta\rangle = E_\beta\,|\beta\rangle \,, \tag{3.155}$$

respectively. The eigenstates $|n\rangle$ are the well known photon number operator eigenstates. The eigenstates $|\beta\rangle$ cannot be found easily. What we know is that the Hamiltonian H^A of the active material can be assumed to have two eigenstates $|1\rangle$ and $|2\rangle$ with energies E_1 and E_2, respectively, such that $E_2 > E_1$. The Hamiltonian H^P may consist of several parts. At least one of them must have a continuous spectrum. Let $|p\rangle$ be its eigenstates. To compute $|\beta\rangle$ one must solve the eigenvalue problem (3.155). Of course this requires a more detailed specification of H^P and the interaction V^C. Since we wish to keep the problem at rather a general level, we shall avoid specifying either H^P or V^C any further. Hence, we shall not attempt to solve the eigenvalue problem (3.155) but we shall only make use of some of the properties that $|\beta\rangle$ can be expected to have. Note that if the interaction V^C is sufficiently weak, the eigenstates $|\beta\rangle$ can be expressed in terms of the eigenstates of H^A and H^P.

The density operator $\rho(t)$ represents the state of the whole system. As the electromagnetic wave progresses from left to right in Fig. 3.6.13, it finds itself in an environment of new atoms each time. The population of the levels

of the atoms are controlled from outside by means of pumping. Of course, this is true only if the field is not so strong as to saturate the amplifier. With this understanding, the level populations are assumed to be constant throughout the amplifier. Then, we may write

$$\rho(t) = \rho^R(t)\,\rho^B(t)\,, \tag{3.156}$$

where $\rho^R(t)$ describes the state of the field made at time t and $\rho^B(t)$ describes the state of the rest of the system at time t. This separation does not imply that field and atoms are constantly uncoupled. In fact, Equation (3.148) describes precisely the coupling between the two. Strictly speaking, $\rho^B(t)$ represents the state of those atoms that have not interacted yet with the field.

Our ultimate goal is to describe the field after amplification. To describe the field means to be able to compute expectation values of field operators. Let Ω be a field operator. Its expectation value $<\Omega>$ is given by

$$<\Omega> = \mathrm{Tr}\,\Omega\rho\,. \tag{3.157}$$

Since Ω, being a field operator, commutes with $|\beta>$, Eq. (3.157) can be written

$$<\Omega> = \mathrm{Tr}\left\{\Omega\left[\mathrm{Tr}_\beta\rho\right]\right\}, \tag{3.158}$$

where Tr_β indicates that the trace with respect to atomic variables is taken. If we introduce the reduced density operator (Fano, 1957)

$$\rho^R(t) \equiv \mathrm{Tr}_\beta\rho(t) \tag{3.159}$$

we shall have

$$<\Omega> = \mathrm{Tr}\,\Omega\,\rho^R(t)\,. \tag{3.160}$$

Note that this definition of ρ^R is consistent with Eqs. (3.157) since $\mathrm{Tr}_\beta\rho^B(t) = 1$. Thus, our objective now is to determine the time evolution of $\rho^R(t)$.

Combining Eqs. (3.148), (3.156) and (3.159), we obtain

$$\rho^R(t+\tau) = \mathrm{Tr}_\beta\, U(\tau)\,\rho^R(t)\,\rho^B(t)\,U^\dagger(\tau)\,. \tag{3.161}$$

This is a field operator. Considering an arbitrary matrix element in the

n-representation we have*

$$\rho^{R}_{mn}(t+\tau) = \langle m \,|\, \mathrm{Tr}_{\beta} U(\tau)\, \rho^{R}(t)\, \rho^{B}(t)\, U^{+}(\tau) \,|\, n \rangle . \tag{3.162}$$

Writing out the trace explicitly, and neglecting the off-diagonal matrix elements of ρ^{B}, by invoking the random phase approximation, we have

$$\rho^{R}_{mn}(t+\tau) = \sum_{m_1 m_2} \sum_{\beta \beta_1} \langle m\beta \,|\, U(\tau) \,|\, m_1\beta_1 \rangle \rho^{R}_{m_1 m_2}(t)\, \rho^{B}_{\beta_1\beta_1}(t) \langle m_2\beta_1 \,|\, U^{+}(\tau) \,|\, n\beta \rangle . \tag{3.163}$$

To proceed further, one needs to calculate the matrix elements of $U(\tau)$. Recall that $U(\tau)$ is given by

$$e^{-\frac{i}{\hbar}(H^{o}+V)\tau}$$

and that H^{o} is diagonal in the representation $\{|n\rangle \; |\beta\rangle\}$. The matrix elements of $U(\tau)$ in this representation can be conveniently calculated by means of damping theory. As discussed by Akcasu (1963) or Messiah (1964), the diagonal matrix elements are given by

$$\langle m\beta \,|U(\tau)|\, m\beta \rangle = e^{-i(m\omega+\omega_{\beta}+S_{m\beta}-i\Gamma_{m\beta})\tau} \tag{3.164}$$

and the off-diagonal matrix elements by

$$\langle m\beta \,|U(\tau)|\, m'\beta' \rangle = \langle m\beta \,|\, V \,|\, m'\beta' \rangle \int_0^{\tau} dt \langle m\beta \,|\, U(\tau-t) \,|\, m\beta \rangle \langle m'\beta' |U(t)|\, m'\beta' \rangle \tag{3.165}$$

where $S_{m\beta'}$, $\Gamma_{m\beta}$ and ω_{β} are defined by

$$S_{m\beta} = \hbar^{-1} \langle m\beta \,|\, V \,|\, m\beta \rangle + \mathcal{P} \sum_{m'\beta' \neq m\beta} \frac{|\langle m'\beta' \,|\, V \,|\, m\beta \rangle|^2}{\hbar(E_{m\beta}-E_{m'\beta'})} , \tag{3.166}$$

*The Dirac ket and the subscript notations will be used interchangeably throughout.

$$\Gamma_{m\beta} = \pi \hbar^{-1} \sum_{m'\beta' \neq m\beta} |\langle m'\beta' | V | m\beta \rangle|^2 \, \delta(E_{m\beta} - E_{m'\beta'}) \tag{3.167}$$

and

$$\omega_\beta \equiv \frac{E_\beta}{\hbar} \quad . \tag{3.168}$$

In the present problem, the diagonal matrix elements of V vanish. Thus $S_{m\beta}$ becomes

$$S_{m\beta} = \mathbb{P} \sum_{m'\beta' \neq m\beta} \frac{|\langle m'\beta' | V | m\beta \rangle|^2}{\hbar (E_{m\beta} - E_{m'\beta'})} \quad , \tag{3.169}$$

where $\mathbb{P}$ indicates the Cauchy principal value. It is essential to keep in mind that both S and Γ are quadratic in V.

Having found expressions for the matrix elements of $U(\tau)$, we now consider Eq.(3.163). The terms resulting from combining diagonal matrix elements of $U(\tau)$ with off-diagonal matrix elements of $U^+(\tau)$, and vice-versa, do not make any contribution of interest. This is easily seen after a straightforward calculation of these terms. Thus, Eq. (3.163) becomes

$$\rho^R_{mn}(t+\tau) = \rho^R_{mn}(t) \sum_\beta \langle m\beta | U(\tau) | m\beta \rangle \, \rho^B_{\beta\beta} \langle n\beta | \, U^+(\tau) | n\beta \rangle + \tag{3.170}$$

$$+ \sum_{m_1 m_2} \rho^R_{m_1 m_2}(t) \sum_{\beta\beta_1}{}' \langle m\beta | U(\tau) | m_1\beta_1 \rangle \rho^B_{\beta_1\beta_1} \langle m_2\beta_1 | \, U^+(\tau) | n\beta \rangle \, ,$$

where the prime on the summation symbol indicates that equal values of the dummy indices are to be excluded. To simplify writing somewhat, we introduce the symbols

$$A^{mn}(\tau) \equiv \sum_\beta \langle m\beta | U(\tau) | m\beta \rangle \rho^B_{\beta\beta} \langle n\beta | U^+(\tau) | n\beta \rangle \quad , \tag{3.171}$$

and

$$C^{mn}_{m_1 m_2}(\tau) \equiv \sum_{\beta\beta_1}{}' \langle m\beta | U(\tau) | m_1\beta_1 \rangle \rho^B_{\beta_1\beta_1} \langle m_2\beta_1 | U^+(\tau) | n\beta \rangle \ . \tag{3.172}$$

In terms of these quantities Eq. (3.170) becomes

$$\rho^R_{mn}(t+\tau) = \rho^R_{mn}(t) A^{mn}(\tau) + \sum_{m_1 m_2} \rho^R_{m_1 m_2} C^{mn}_{m_1 m_2}(\tau) \ . \tag{3.173}$$

Nothing has been said about the time interval, τ thus far. In fact, due to the perturbation calculation we have used (damping theory is indeed a perturbative approach which, however, takes into account the decay of the initial state), τ cannot be made arbitrarily large. It is to be understood as macroscopically small but large compared to characteristic times of atomic processes. This time interval will be used to calculate events per unit time and then develop a differential equation for $\rho^R_{mn}(t)$. This is the subject of the following section.

3.6.4 Differential Equation for the Matrix Elements of ρ^R.

The calculation of $A^{mn}(\tau)$ and $C^{mn}_{m_1 m_2}(\tau)$ is rather lengthy but straightforward. The result is

$$A^{mn}(\tau) = e^{-i(m-n)\omega\tau} \sum_\beta \rho^B_{\beta\beta} e^{-i(\gamma_{m\beta} - \gamma^*_{n\beta})\tau} \tag{3.174}$$

where we have introduced

$$\gamma_{m\beta} \equiv S_{m\beta} - i\Gamma_{m\beta} \ . \tag{3.175}$$

From the term containing $C^{mn}_{m_1 m_2}$, it turns out that only the terms for $(m_1 = m+1,\ m_2 = n+1)$ and $(m_1 = m-1,\ m_2 = n-1)$ do not vanish. Thus, the following equation is obtained:

$$\rho^R_{mn}(t+\tau) = e^{-i(m-n)\omega\tau} \Big[\rho^R_{mn}(t) \sum_\beta \rho^B_{\beta\beta} e^{-i(\gamma_{m\beta} - \gamma^*_{n\beta})\tau} +$$

$$+ \sqrt{(m+1)(n+1)}\, \rho^R_{(m+1)(n+1)}(t) \sum_{\beta\beta_1}{}' \rho^B_{\beta_1\beta_1} |D_{\beta\beta_1}|^2 J^{(+)}_{\beta\beta_1}(\tau)$$

$$+\sqrt{mn}\;\rho^{R}_{(m-1)(n-1)}(t)\sum_{\beta\beta_1}{}'\,\rho^{B}_{\beta_1\beta_1}\left|D_{\beta\beta_1}\right|^2 J^{(-)}_{\beta\beta_1}(\tau)\Bigg] \tag{3.176}$$

where the quantites $J^{(\pm)}$ are defined by

$$J^{(\pm)}_{\beta\beta_1}(\tau) \equiv e^{-i(\gamma_{m\beta}-\gamma^{*}_{n\beta})\tau}$$

$$\frac{\left\{e^{-i\left(\pm\omega-\omega_{\beta\beta_1}-\gamma_{m\beta}+\gamma_{(m\pm1)\beta_1}\right)\tau}-1\right\}\left\{e^{i\left(\pm\omega-\omega_{\beta\beta_1}-\gamma^{*}_{n\beta}+\gamma^{*}_{(n\pm1)\beta_1}\right)\tau}-1\right\}}{\left(\pm\omega-\omega_{\beta\beta_1}-\gamma_{m\beta}+\gamma_{(m\pm1)\beta_1}\right)\left(\pm\omega-\omega_{\beta\beta_1}-\gamma^{*}_{n\beta}+\gamma^{*}_{(n\pm1)\beta_1}\right)} \tag{3.177}$$

For τ sufficiently small, the exponential in the first term of the right side of Eq.(3.176) can be expanded in a series and the first two terms be retained. Recall that γ is quadratic in V and hence quadratic in D. This means that by retaining only the first two terms, we limit the expansion to second-order perturbation theory. Now taking

$$\frac{d}{dt}\rho^{R}_{mn}(t) \cong \frac{\rho^{R}_{mn}(t+\tau)-\rho^{R}_{mn}(t)}{\tau}\,, \tag{3.178}$$

we obtain the following differential equation:

$$\frac{d}{dt}\rho^{R}_{mn}(t) = \rho^{R}_{mn}(t)\left[-i(m-n)\omega - i\sum_{\beta}\rho^{B}_{\beta\beta}(\gamma_{m\beta}-\gamma^{*}_{n\beta})\right] +$$

$$+\sqrt{(m+1)(n+1)}\;\rho^{R}_{(m+1)(n+1)}(t)\sum_{\beta\beta_1}{}'\,\rho^{B}_{\beta_1\beta_1}\left|D_{\beta\beta_1}\right|^2\frac{J^{(+)}_{\beta\beta_1}(\tau)}{\tau} +$$

$$+\sqrt{mn}\;\rho^{R}_{(m-1)(n-1)}(t)\sum_{\beta\beta_1}{}'\,\rho^{B}_{\beta_1\beta_1}\left|D_{\beta\beta_1}\right|^2\frac{J^{(-)}_{\beta\beta_1}(\tau)}{\tau}\,.$$

The quantities $\frac{J^{(\pm)}_{\beta\beta_1}}{\tau}$ are to be interpreted as independent of τ.

To proceed further, one must compute the summations over β and B_1. This in turn implies that the eigenstates and the spectrum of $H^B = H^A + H^P + V^C$ must be found. Of course this problem can be solved by, for example, using damping theory again. However we do not know V^C and H^P, and we do not wish to specify them. Thus, we resort to another alternative.

Recall that H^A represents a two-level system. If the coupling V^C is turned off, that is, if the levels of H^A are not broadened, the summations over β and β_1 simply reduce to summations over the two levels. Also, $\rho^B_{\beta\beta}$ would then assume the values ρ^B_{22} and ρ^B_{11} representing the fractional populations of the upper and lower levels, respectively. The coupling to the perturbing mechanism causes the levels of H^A to be slightly shifted from their previous values. This shift and broadening are statistical in nature. Roughly speaking, one may say that the two-level atoms can now be found to have energies adjacent to the previously sharp levels. We may then use the following formula: Replace the summation over β by

$$\sum_{j=1}^{2} \int \sigma_j(\omega', \omega_j)\, d\omega \,, \tag{3.180}$$

where $\hbar\omega_2 = E_2$ and $\hbar\omega_1 = E_1$ are the unbroadened levels of H^A. The function $\sigma_j(\omega', \omega_j)$ is assumed to be sharply peaked at ω_j so that, although the integration can be extended to infinity formally, the main contribution comes from a small frequency interval around ω_j. The above is to be used as follows: Consider, for example, an expression of the form

$$\sum_{\beta\beta_1}{}' \rho^B_{\beta_1\beta_1} \left| D_{\beta\beta_1} \right|^2 F(\omega_\beta, \omega_{\beta_1}) \tag{3.181}$$

where $F(\omega_\beta, \omega_{\beta_1})$ is some function of ω_β and ω_{β_1}. This expression will be replaced by

$$\rho_{22}^{A} \left|D_{21}\right|^2 \iint \sigma_2(\omega', \omega_2)\, \sigma_1(\omega'', \omega_1)\, F(\omega', \omega'')\, d\omega'\, d\omega'' +$$

$$+ \rho_{11}^{A} \left|D_{21}\right|^2 \iint \sigma_1(\omega', \omega_1)\, \sigma_2(\omega'', \omega_2)\, F(\omega', \omega'')\, d\omega'\, d\omega'' , \qquad (3.182)$$

where ρ_{22}^{A} and ρ_{11}^{A} are the fractional populations of the upper and lower levels of the active material, and D_{21} is the usual dipole moment matrix element. In writing this expression we have assumed that $\rho^{B} = \rho^{A} \rho^{P}$ and ρ^{P} has been absorbed in the function σ_j. Indeed these functions depend on the state of the perturbing mechanism and the nature of interaction between H^{A} and H^{P}. In our case, the validity of the approximation involved in replacing (3.181) by (3.182) is enhanced further by the fact that the functions $F(\omega_\beta, \omega_{\beta_1})$ we shall encounter are themselves peaked functions of ω_β and ω_{β_1}. This approximation does not solve the problem of finding $|\beta>$ but simply shifts it. Nevertheless, this enables us to reach some conclusions without knowing the functions σ_j explicitly. To determine these functions, however, one has no other alternative than solving the eigenvalue problem.

We now use this procedure to compute the summations over β and β_1 in Eq. (3.179). First, we take up the quantity

$$i \sum_{\beta} \rho_{\beta\beta}^{B} (\gamma_{m\beta} - \gamma_{n\beta}^{*}) .$$

The real and imaginary parts of $\gamma_{m\beta}$ are given by Eq. (3.169) and (3.167), respectively. A somewhat lengthy calculation in which use of the peaked character of σ_j is made, gives the following result:

$$i \sum_{\beta} \rho_{\beta\beta}^{B} (\gamma_{m\beta} - \gamma_{n\beta}^{*}) = c_2 \left[(m+1)K^{*}(\omega) + (n+1)K(\omega)\right] + c_1 \left[mK(\omega) + nK^{*}(\omega)\right] , \qquad (3.183)$$

where

$$c_j \equiv \rho_{jj}^{A} \left|D_{21}\right|^2 \hbar^{-2}, \qquad j = 1, 2, \tag{3.184}$$

$$b_j \equiv 2\pi c_j g(\omega, \omega_o) = 2\pi \hbar^{-2} \rho_{jj}^{A} \left|D_{21}\right|^2 g(\omega, \omega_o), \tag{3.185}$$

$$K(\omega) \equiv \pi g(\omega, \omega_o) + i f(\omega, \omega_o), \tag{3.186}$$

$$g(\omega, \omega_o) \equiv \int \sigma_1(\omega', \omega_1)\, \sigma_2(\omega + \omega', \omega_2)\, d\omega', \tag{3.187}$$

$$f(\omega, \omega_o) \equiv \mathcal{P} \iint \frac{\sigma_2(\omega'', \omega_2)\, \sigma_1(\omega', \omega_1)}{\omega' + \omega - \omega''}\, d\omega'\, d\omega'', \tag{3.188}$$

and

$$\omega_o \equiv \frac{E_2 - E_1}{\hbar}. \tag{3.189}$$

Next we consider the terms involving $J^{(\pm)}$. Note that $J^{(\pm)}$ contains γ's in the exponentials as well as in the denominators. However, the terms containing $J^{(\pm)}$ also contain the factor $\left|D_{\beta\beta_1}\right|^2$. Since the γ's are themselves quadratic in D, we should retain only terms of second order in D, if our approximation scheme is to be consistent. Thus we replace all γ's by zero. The resulting expression is substituted into Eq. (3.179) and the summations are then performed. Use of the peaked character of the functions σ_j is again made. The result is

$$\sum_{\beta\beta_1}{}' \rho_{\beta_1\beta_1}^{B} \left|D_{\beta\beta_1}\right|^2 \frac{J_{\beta\beta_1}^{(+)}(\tau)}{\tau} = 2\pi\hbar^{-2} \left|D_{21}\right|^2 \rho_{11}^{A}\, g(\omega, \omega_o) = b_1 \tag{3.190a}$$

$$\sum_{\beta\beta_1}{}' \rho_{\beta_1\beta_1}^{B} \left|D_{\beta\beta_1}\right|^2 \frac{J_{\beta\beta_1}^{(-)}(\tau)}{\tau} = 2\pi\hbar^{-2} \left|D_{21}\right|^2 \rho_{22}^{A}\, g(\omega, \omega_o) = b_2. \tag{3.190b}$$

The quantities b_1 and b_2 are as defined in Eq. (3.185).

Combining now these results with Eq. (3.179), we have

$$\frac{d}{dt}\rho^R_{mn}(t) = -i(m-n)\omega\rho^R_{mn}(t) - \left\{c_2\left[(m+1)K^*(\omega)+(n+1)K(\omega)\right] + c_1\left(mK(\omega)+nK^*(\omega)\right)\right\}\rho^R_{mn}(t)$$

$$+ b_2\sqrt{mn}\,\rho^R_{(m-1)(n-1)}(t) + b_1\sqrt{(m+1)(n+1)}\,\rho^R_{(m+1)(n+1)}(t) \quad . \qquad (3.191)$$

This differential equation describes the time evolution of the state of the field mode. The parameters appearing in this equation have been expressed in terms of the populations of the energy levels of the active material, the coupling constant D and the functions $g(\omega,\omega_o)$ and $f(\omega,\omega_o)$. One can easily see that these functions are peaked about ω_o. The function $g(\omega,\omega_o)$ is to be identified with the spectrum of spontaneous emission from $|2>$ to $|1>$. If one assumes that the levels of the active material are sharply defined (i.e. ignore broadening), then σ_j would be a delta function; i.e.

$$\sigma_j(\omega',\omega_j) = \delta(\omega'-\omega_j) \quad , \qquad j = 1,2. \qquad (3.192)$$

One can then easily verify that

$$g(\omega,\omega_o) = \delta(\omega-\omega_o) \quad , \qquad (3.193a)$$

and

$$f(\omega,\omega_o) = \mathcal{P}\,\frac{1}{\omega-\omega_o} \quad . \qquad (3.193b)$$

If one starts with two sharply defined energy levels and proceeds as we did, the summations over β,β_1 can be carried out rigorously and no need for introducing the functions σ_j exists. Then one finds the results given by Eqs. (3.193a,b). This shows that the procedure we used to perform the summations is self-consistent since it gives the correct results in this limiting case.

All matrix elements of the density operator ρ^R obey Eq. (3.191). For the diagonal matrix elements, however, this equation assumes a simpler form. Indeed, if we take m = n and observe that

$$c_j(K+K^*) = b_j , \qquad j = 1, 2 , \tag{3.194}$$

we obtain

$$\frac{d}{dt}\rho^R_{mm}(t) = -\left[b_2(m+1) + b_1 m\right]\rho^R_{mm}(t) + b_2 m\,\rho^R_{(m-1)(m-1)}(t) +$$

$$+ b_1(m+1)\,\rho^R_{(m+1)(m+1)}(t) \quad . \tag{3.195}$$

This is essentially the equation that Shimoda et al (1957) have obtained on the basis of probabilistic arguments. Here, we have generalized the results of these authors in two respects: we have obtained equations for the off-diagonal matrix elements as well, and we have accounted for broadening so that the amplifier has a finite amplification bandwidth. The off-diagonal matrix elements of ρ^R are necessary in the study of the amplification of the field amplitudes. This aspect has been studied by Louisell et al (1961) and by Gordon et al (1963a, 1963b). The present work differs in the approach and in that we have accounted for broadening. The results of the above references are recaptured if the functions σ_j are replaced by delta functions.

3.6.5 Solution of the Differential Equation

The equation for the diagonal matrix elements of ρ^R has been solved by Shimoda et al (1957). The method they used was to introduce an appropriate generating function, develop a partial differential equation for it and solve the equation. Here, we solve the general equation for arbitrary matrix elements. Due to the different form of the equation the method of the above authors cannot be used directly.

First we observe that according to Eq. (3.191), the (m, n) matrix element

is coupled to the (m+1, n+1) and (m-1, n-1) matrix elements only. Let us consider the subset of matrix elements for which m-n=ℓ has a given value. By letting ℓ vary from $-\infty$ to $+\infty$, over the integers, we obtain the totality of the matrix elements of ρ^R. We observe, furthermore, that the matrix elements of a subset with a given ℓ are coupled to each other and not to matrix elements belonging to any other subset. Thus, let us set

$$m = n + \ell \quad . \tag{3.196}$$

Substituting this into Eq. (3.19) we obtain

$$\frac{d}{dt}\,\rho^R_{(n+\ell)n}(t) = -(bn+c)\rho^R_{(n+\ell)n}(t) + b_2\sqrt{(n+\ell)n}\;\rho^R_{(n+\ell-1)(n-1)}(t) +$$

$$+ b_1\sqrt{(n+\ell+1)(n+1)}\;\rho^R_{(n+\ell+1)(n+1)}(t) \tag{3.197}$$

where

$$b \equiv b_2 + b_1 \tag{3.198a}$$

$$c \equiv b_2 + \gamma_\ell \tag{3.198b}$$

$$\gamma_\ell \equiv \left(\frac{b}{2} + i\nu\right)\ell \tag{3.198c}$$

and

$$\nu \equiv \omega - (c_2 - c_1)\,\mathrm{Im}\,K = \omega - (c_2 - c_1)\,f(\omega, \omega_o). \tag{3.198d}$$

These new parameters have been defined for the sake of reducing the equation to as simple a form as possible.

Now, we introduce a new set of functions $F_{mn}(t)$ defined by

$$F_{mn}(t) = \rho_{mn}(t)\sqrt{\frac{m!}{n!}} \tag{3.199}$$

and substitute into Eq. (3.197). We obtain

$$\frac{d}{dt}\,F^{(\ell)}_m(t) = -(bm+c)\,F^{(\ell)}_m(t) + b_2(m+\ell)\,F^{(\ell)}_{m-1}(t) + b_1(m+1)\,F^{(R)}_{m+1}(t) \tag{3.200}$$

where we have changed the notation slightly. That is, we have written $F^{(\ell)}_m(t)$ instead of $F_{(m+\ell)m}(t)$. The advantage of transformation (3.199) is that it has

removed the square roots and has reduced the equation to a form very similar to the one obeyed by the diagonal matrix elements. We may now proceed by using the technique that Shimoda et al (1957) have used.

Let us introduce a generating function $R^{(\ell)}(x, t)$ defined by

$$R^{(\ell)}(x,t) = \sum_{m=0}^{\infty} F_m^{(\ell)}(t)\, x^m . \tag{3.201}$$

We have a family of differential equations and a corresponding family of generating functions parametrized by ℓ. Again, for $\ell = 0$ we obtain the generating function of Shimoda et al (1957). Since the parameter ℓ is contained in the coefficients of the equations, we shall drop it from here on. Taking now the partial derivative of $R(x, t)$ with respect to time, and using Eq. (68) we obtain

$$\frac{\partial R}{\partial t} + (1 - x)(b_2 x - b_1)\frac{\partial R}{\partial x} + (c - ax)R = 0 , \tag{3.202}$$

where we have introduced

$$a = b_2(\ell + 1) .$$

Note that for $\ell = 0$, we have $a = c = b_2$. The program now is to solve Eq. (3.202), express the solution as a series of powers of x, identify the coefficients with $F_m(t)$ and then determine $\rho^R_{(m+\ell)m}(t)$ from the equation

$$\rho^R_{(m+\ell)m}(t) = F_m(t)\sqrt{\frac{m!}{(m+\ell)!}} . \tag{3.204}$$

The initial condition will be a given set of matrix elements $\rho^R_{mn}(0)$ and hence a known set of numbers $F_{mn}(0)$.

To solve Eq. (3.202) we consider the equivalent system of ordinary differential equations

$$\frac{dx}{(1-x)(b_2 x - b_1)} = dt = \frac{dR}{(ax - c)R} . \tag{3.205}$$

Introducing

$$y \equiv \frac{1}{1-x} \tag{3.206}$$

the first equation gives

$$\frac{dy}{dt} = 2ky - b_2 , \tag{3.207}$$

where k is defined by

$$k = \frac{b_2 - b_1}{2} . \tag{3.208}$$

The solution of Eq. (3.207) is

$$y(t) = y_o e^{2kt} + \lambda , \tag{3.209}$$

where we have introduced

$$\lambda \equiv \frac{b_2}{2k} = \frac{b_2}{b_2 - b_1} . \tag{3.210}$$

Considering now the second differential equation and using the above expression for y(t), one can solve for R(x,t). The result is

$$R(x,t) = C\, e^{-i\nu\ell t} e^{\ell kt} (\lambda e^{-2kt} + y_o)^{\ell+1} \tag{3.211}$$

where C is the constant of integration.

Let now $\rho^R_{mn}(0)$ be the matrix elements of the density operator of the field mode at t = 0. These are assumed to be known. Then, from Eq. (3.199) we have

$$F_m(0) = \sqrt{\frac{(m+\ell)!}{m!}}\, \rho_{(m+\ell)m}(0) , \tag{3.212}$$

and the initial value of R(x,t) is

$$R_o = \sum_{m=o}^{\infty} F_m(0)\, x_o^m . \tag{3.213}$$

Setting t = 0 in Eq. (3.211) and using Eq. (3.213) we find

$$C = (y_o + \lambda)^{-(\ell+1)} \sum_{m=0}^{\infty} F_m(0) x_o^m \tag{3.214}$$

which determines the constant of integration. Substituting into Eq. (3.211), we have

$$R(x,t) = \left(\sum_{m=0}^{\infty} F_m(0) x_o^m \right) e^{\ell(k-i\nu)t} \left(\frac{y_o + \lambda e^{-2kt}}{y_o + \lambda} \right)^{\ell+1} . \tag{3.215}$$

Now introduce the quantity

$$G \equiv e^{2kt} , \tag{3.216}$$

in terms of which Eq. (3.209) becomes

$$y = y_o G + \lambda . \tag{3.217}$$

Solving for y_o we have

$$y_o = \frac{1}{G} (y - \lambda) . \tag{3.218}$$

Using this relation and the fact that

$$x_o = 1 - \frac{1}{y_o} , \tag{3.219}$$

we obtain

$$x_o = \frac{y - (\lambda + G)}{y - \lambda} . \tag{3.220}$$

Recalling the definition of y (see Eq. (3.206)), we have

$$x_o = \frac{1 - (\lambda + G)(1 - x)}{1 - \lambda (1 - x)} . \tag{3.221}$$

Simiarly, using Eqs. (3.216), (3,217) and (3.206), we obtain

$$\frac{y_o + \lambda e^{-2kt}}{y_o + \lambda} = \frac{1}{1 + \lambda (G-1)(1-x)} . \tag{3.222}$$

It will prove to be more convenient to write Eqs. (3.221) and (3.222) as follows:

$$x_o = \frac{G+\lambda-1}{\lambda-1}\,\frac{1-\dfrac{G+\lambda}{G+\lambda-1}\,x}{1-\dfrac{\lambda}{\lambda-1}\,x} \tag{3.223}$$

and

$$\frac{y_o+\lambda e^{-2kt}}{y_o+\lambda} = \frac{1}{1+\lambda(G-1)}\,\frac{1}{1-\dfrac{\lambda(G-1)}{1+\lambda(G-1)}\,x}\,. \tag{3.224}$$

To compress notation, introduce the parameters

$$\xi \equiv \frac{G+\lambda-1}{\lambda-1}\,, \tag{3.225a}$$

$$\eta \equiv \frac{G+1}{G+\lambda-1}\,, \tag{3.225b}$$

$$\mu \equiv \frac{1}{1+\lambda(G-1)} \tag{3.225c}$$

and

$$z \equiv \frac{\lambda(G-1)}{1+\lambda(G-1)} = \mu\lambda(G-1)\,. \tag{3.225d}$$

Also, note that

$$\frac{\lambda}{\lambda-1} = \frac{b_2}{b_1}\,.$$

In terms of these parameters, Eqs. (3.223) and (3.224) become

$$x_o = \frac{1-\eta x}{1-\dfrac{b_2}{b_1}\,x} \tag{3.226}$$

and

$$\frac{y_o+\lambda e^{-2kt}}{y_o+\lambda} = \frac{\mu}{1-zx}\,. \tag{3.227}$$

Substituting into Eq. (3.215), we obtain the solution of the partial differential equation, namely

$$R(x,t) = e^{\ell(k-i\nu)t} \sum_{m=0}^{\infty} F_m(0)\xi^m (1-\eta x)^m \left(1-\frac{b_2}{b_1}x\right)^{-m} \frac{\mu^{\ell+1}}{(1-zx)^{\ell+1}} \quad . \tag{3.228}$$

To express this as a power series in x, we use the identities

$$(1-\alpha)^n = \sum_{r=0}^{n} (-1)^r \frac{n!}{r!\,(n-r)!}\alpha^r \quad , \tag{3.229}$$

and

$$(1-\alpha)^{-n} = \sum_{q=0}^{\infty} \frac{(n-1+q)!}{(n-1)!\,q!}\alpha^q \quad . \tag{3.230}$$

Then, Eq. (3.228) becomes

$$R(x,t) = \mu^{\ell+1} e^{\ell(k-i\nu)t}$$

$$\sum_{m=0}^{\infty}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} (-1)^r F_m(0)\,\xi^m \eta^r \left(\frac{b_2}{b_1}\right)^j z^q$$

$$\frac{m!\,(m-1+j)!\,(\ell+q)!}{r!\,(m-r)!\,(m-1)!\,j!\,\ell!\,q!}\, x^{r+j+q} \quad . \tag{3.231}$$

Setting now $r + j + q = M$, (3.232)

Eq. (3.231) becomes

$$R(x,t) = \mu^{\ell+1} e^{\ell(k-i\nu)t}$$

$$\sum_{j=0}^{\infty}\sum_{q=0}^{\infty}\sum_{m=0}^{\infty}\sum_{M=j+q}^{m+j+q} (-1)^{M-j-q} F_m(0)\,\xi^m \eta^{M-j-q}$$

$$\left(\frac{b_2}{b_1}\right)^j z^q \frac{m!\,(m-1+j)!\,(\ell+q)!}{(M-j-q)!\,(m-M+j+q)!\,(m-1)!\,j!\,\ell!\,q!}\, x^M \quad .$$

(3.233)

Considering the coefficient of a specific power x^M and using Eq. (3.204), we obtain the final result

$$\rho^R_{(M+\ell)M}(t) = (-\eta)^M \sqrt{\frac{M!}{(M+\ell)!}}\, \mu^{\ell+1}\, e^{\ell(k-i\nu)t}$$

$$\sum_{m=0}^{\infty}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} \rho^R_{(m+\ell)m}(0)\, \xi^m\, (-1)^{j+q} \left(\frac{b_2}{\eta b_1}\right)^j \left(\frac{z}{\eta}\right)^q$$

$$\frac{\sqrt{(m+\ell)!\, m!}\, (m-1+j)!\, (\ell+q)!}{(m-1)!\, \ell!\, (M-j-q)!\, (m+j+q-M)!\, j!\, q!}\,, \qquad (3.234)$$

or in a slightly different form

$$\rho^R_{(M+\ell)M}(t) = (-\eta)^M \sqrt{\frac{M!}{(M+\ell)!}}\, \mu^{\ell+1}\, e^{\ell(k-i\nu)t}$$

$$\sum_{m=0}^{\infty} \rho^R_{(m+\ell)m}(0)\, \xi^m\, \frac{\sqrt{(m+\ell)!\, m!}}{(m-1)!\, \ell!} \qquad (3.235)$$

$$\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} (-1)^{j+q} \left(\frac{b_2}{\eta b_1}\right)^j \left(\frac{z}{\eta}\right)^q \frac{(m-1+j)!\, (\ell+q)!}{(M-j-q)!\, (m+j+q-M)!\, j!\, q!}\,.$$

Although the summations extend formally from zero to infinity, in fact they are limited by the requirement that none of the factorials be negative. This restriction stems from the initial equation (3.231).

Letting $\ell = 0$ in Eq. (3.235), we obtain the solution for the diagonal matrix elements, that is,

$$\rho^R_{MM}(t) = (-\eta)^M\, \mu \sum_{m=0}^{\infty} \rho^R_{mm}(0)\, \xi^m\, \frac{m!}{(m-1)!}$$

$$\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} (-1)^{j+q} \left(\frac{b_2}{\eta b_1}\right)^j \left(\frac{z}{\eta}\right)^q \frac{(m-1+j)!}{(M-j-q)!\, (m+j+q-M)!\, j!}\,. \qquad (3.236)$$

This still is not the Shimoda et al (1957) result. These authors assume that the field mode has a precise number of photons in the initial state. This implies that

$$\rho^{R}_{mm}(0) = \delta_{mm_0}, \tag{3.237}$$

where m_0 is the initial number of photons. Therefore, by deleting the summation over m in Eq. (3.236) and replacing m by m_0, we obtain a result equivalent to that of Shimoda et al (1957). Note that all parameters appearing in Eq. (3.235) and (3.236) depend on the two parameters G and λ. The second characterizes the population inversion of the active material. Quite often a maser-laser temperature is introduced in order to characterize the population inversion. This temperature T_m is defined by

$$\frac{b_2}{b_1} = e^{-\hbar\omega_0/kT_m}, \tag{3.238}$$

and is negative for an active material (here k is Boltzmann's constant and should not be confused with the same symbol appearing in previous equations). In terms of this negative temperature one has

$$\lambda = \frac{1}{1-e^{\hbar\omega_0/kT_m}}. \tag{3.239}$$

The other parameter G will be shown to be the gain of the amplifier.

3.6.6 Amplification of the Field Amplitudes

Although our central interest in this study is the amplification of the energy (or number of quanta), we shall discuss somewhat the amplification of the field amplitudes since this aspect sheds some light on the properties of quantum amplifiers. Moreover, it is not necessary that laser communication channels be confined to energy measurements only.

The field amplitude operators are given by

$$q = \sqrt{\frac{\hbar}{2\omega}}\,(a^{+} + a)\,, \tag{3.240a}$$

and

$$p = i\sqrt{\frac{\hbar\omega}{2}}\,(a^{+} - a)\,. \tag{3.240b}$$

It is more convenient to introduce the dimensionless hermitian operators

$$Q \equiv a^{+} + a\,, \tag{3.241a}$$

and

$$P \equiv i(a^{+} - a)\,, \tag{3.241b}$$

which are related to q and p through the equations

$$q = Q\sqrt{\frac{\hbar}{2\omega}} \qquad \text{and} \qquad p = P\sqrt{\frac{\hbar\omega}{2}}\,. \tag{3.242}$$

Let now $\langle a^{+}(t)\rangle = \mathrm{Tr}\left\{\rho^{R}(t)\,a^{+}\right\}$ be the expectation value of a^{+}. The matrix elements of a^{+} are given by

$$\langle n \mid a^{+} \mid m\rangle = \sqrt{m+1}\;\delta_{n,\,m+1}\,. \tag{3.243}$$

Using this relation we obtain

$$\frac{\partial}{\partial t}\langle a^{+}(t)\rangle = \left\{\mathrm{Tr}\;\dot{\rho}^{R}(t)\,a^{+}\right\} = \sum_{m=0}^{\infty}\dot{\rho}^{R}_{m(m+1)}(t)\sqrt{m+1}\,. \tag{3.244}$$

From Eq. (3.90) we find

$$\frac{\partial}{\partial t}\,\rho^{R}_{m(m+1)}(t) = i\nu\,\rho^{R}_{m(m+1)}(t) - \left(b_2+\frac{b}{2}\right)\rho^{R}_{m(m+1)}(t) - bm\,\rho^{R}_{m(m+1)}(t) +$$

$$+ b_2\sqrt{m(m+1)}\,\rho^{R}_{(m-1)m}(t) + b_1\sqrt{(m+1)(m+2)}\,\rho^{R}_{(m+1)(m+2)}(t)\,. \tag{3.245}$$

Combining this with Eq. (3.244) and after some rearrangement and simplification, we obtain

$$\frac{\partial}{\partial t} < a^{+}(t) > = (k+i\nu) < a^{+}(t) > \quad , \tag{3.246}$$

where k and ν have been defined in Eq. (3.208) and (3.198d). Evidently, the solution of Eq. (3.246) is

$$< a^{+}(t) > = < a^{+}(0) > e^{(k+i\nu)t} , \tag{3.247}$$

where $< a^{+}(0) >$ is the expectation value at the initial time which, for our purposes is the time zero at which the signal enters the input. The time that the signal spends inside the amplifier is related to the length of the amplifier through Eq. (3.133) .

Taking now the complex conjugate of Eq. (3.247), we obtain the solution for $< a(t) >$, namely,

$$< a(t) > = < a(0) > e^{(k-i\nu)t} . \tag{3.248}$$

Combining the last two equations, we have

$$< Q(t) > = \left[< Q(0) > \cos \nu t + < P(0) > \sin \nu t \right] e^{kt} , \tag{3.249a}$$

$$< P(t) > = \left[- < Q(0) > \sin \nu t + < P(0) > \cos \nu t \right] e^{kt} . \tag{3.249b}$$

The corresponding equations for the oscillator coordinates q and p are:

$$< q(t) > = \left[< q(0) > \cos \nu t + \frac{1}{\omega} < p(0) > \sin \nu t \right] e^{kt} \tag{3.250a}$$

$$< p(t) > = \left[- \omega < q(0) > \sin \nu t + < p(0) > \cos \nu t \right] e^{kt} . \tag{3.250b}$$

These equations are similar to those obtained by Louisell et al (1961). They differ slightly because here we have accounted for a shift in the oscillator frequency which will be discussed later. The equations of Louisell et al (1961) are obtained if we replace ν by ω, which is equivalent to neglecting the shift represented by $(c_2 - c_1 \operatorname{Im} K$ (see Eq. (3.198d). The exponential represents the gain which increases exponentially with the difference $(b_2 - b_1)$ and the time that the signal travels inside the amplifier. Note that e^{kt} is the square root of G which is defined by Eq. (3.216).

The field coordinates increase as the square root of G. Of course the gain is larger than unity only if the upper level is more populated than the lower i.e. if $b_2 > b_1$. This is understood to be the case throughout this treatment.

5.6.7 Amplification of the Energy

The energy of the field mode is directly proportional to the number of photons present. Thus, we shall study the amplification of the number of photons. The photon number operator is a^+a. If we denote by $f(t)$ the average number of photons at time t we shall have

$$f(t) = \langle a^+(t)\, a(t) \rangle \quad . \tag{3.251}$$

The expected value of the energy then is $\hbar\omega f(t)$. To find an equation for $f(t)$, we proceed exactly as we did for $\langle a^+(t) \rangle$. In the present case only diagonal matrix elements of ρ^R need be considered since the representation $\{|n\rangle\}$ by definition diagonalizes a^+a. Thus, we obtain the equation

$$\frac{\partial}{\partial t} f(t) = 2kf(t) + b_2 , \tag{3.252}$$

which has the solution

$$f(t) = \left(f(0) + \frac{b_2}{b_2 - b_1} \right) e^{2kt} - \frac{b_2}{b_2 - b_1} \quad . \tag{3.253}$$

A more convenient form is

$$f(t) = Gf(0) + (G-1)\lambda , \tag{3.254}$$

where λ and G are as defined by Eqs. (3.210) and (3.216). The quantity λ characterizes the population inversion. For active materials, that is for $b_2 > b_1$, we shall have $\lambda \geq 1$. The equality is attained under complete inversion. To have the maximum possible amplification, we would desire this extreme case. For very large gain, i.e. for $G >> 1$, $f(t)$ can be approximated by

$$f(t) = G(f(0) + \lambda) . \tag{3.255}$$

This shows that there will be an output even when the number of input photons f(0) is zero. Clearly this is due to spontaneous emission.

3.6.8 Fluctuations of the Field Amplitudes.

In a previous section, the expected values of the field amplitudes were expressed in terms of the expected values at the input, and the gain. The outcomes of measurements of these amplitudes at the output will fluctuate about the expected values. It is desirable, therefore to have a quantitative estimate of these fluctuations upon which one can base a criterion for the usefulness of the amplifier. A rather conventional measure of the fluctuations is given by the quantity

$$\mathcal{E}_Q^2 \equiv \frac{\langle \Delta Q^2 \rangle}{\langle Q \rangle^2} \tag{3.256}$$

where $\langle \Delta Q^2 \rangle \equiv \langle Q^2 \rangle - \langle Q \rangle^2$. (3.257)

A similar quantity $\mathcal{E}_P^2$ measures the fluctuations of P. Let us now calculate $\langle Q^2(t) \rangle$ and $\langle P^2(t) \rangle$.

From the definition of Q, we have

$$Q^2 = a^+ a^+ + aa + 2a^+ a + 1 , \tag{3.258}$$

from which we obtain

$$\langle Q^2(t) \rangle = \langle a^+(t) a^+(t) \rangle + \langle a(t) a(t) \rangle + 2f(t) + 1 . \tag{3.259}$$

Observing that

$$\langle a(t) a(t) \rangle = \langle a^+(t) a^+(t) \rangle^* , \tag{3.260}$$

we conclude that we only need to calculate the quantity $\langle a^+(t) a^+(t) \rangle$. Using Eq. (3.243) we obtain

$$\langle a^+(t) a^+(t) \rangle = \mathrm{Tr}\left\{ a^+ a^+ \rho^R(t) \right\} = \sum_{m=0}^{\infty} \rho^R_{m(m+2)}(t) \sqrt{(m+1)(m+2)} , \tag{3.261}$$

from which we have

$$\frac{d}{dt}\langle a^{+}(t)a^{+}(t)\rangle = \sum_{m=o}^{\infty} \frac{d}{dt}\,\rho^{R}_{m(m+2)}(t)\sqrt{(m+1)(m+2)}\ . \qquad (3.262)$$

Now, we make use of the differential equation (3.191) for the matrix elements of ρ^{R}. After some straightforward algebraic manipulations, we obtain

$$\frac{d}{dt}\langle a^{+}(t)\,a^{+}(t)\rangle = 2(k+i\nu)\langle a^{+}(t)\,a^{+}(t)\rangle\ , \qquad (3.263)$$

which has the solution

$$\langle a^{+}(t)\,a^{+}(t)\rangle = \langle a^{+}(0)\,a^{+}(0)\rangle\, e^{2(k+i\nu)t}\ . \qquad (3.264)$$

From this and Eq. (3.260) we have

$$\langle a(t)\,a(t)\rangle = \langle a(0)\,a(0)\rangle\, e^{2(k-i\nu)t}\ . \qquad (3.265)$$

Now Eq. (3.259) becomes

$$\begin{aligned}\langle Q^{(2)}(t)\rangle &= 2G\langle a^{+}(0)\,a^{+}(0)\rangle^{(r)}\cos 2\nu t + \\ &\quad + 2G\langle a^{+}(0)a^{+}(0)\rangle^{(i)}\sin 2\nu t + \\ &\quad + 2Gf(0)+2(G-1)\lambda + 1\ , \end{aligned} \qquad (3.266)$$

where we have used Eq. (3.254). The superscripts (r) and (i) indicate the real and imaginary parts, respectively, of the quantity they qualify. Using the same procedure to compute $\langle P^{2}(t)\rangle$, we find

$$\begin{aligned}\langle P^{2}(t)\rangle &= -2G\langle a^{+}(0)\,a^{+}(0)\rangle^{(r)}\cos 2\nu t - \\ &\quad - 2G\langle a^{+}(0)\,a^{+}(0)\rangle^{(i)}\sin 2\nu t + \\ &\quad + 2Gf(0) + 2(G-1)\lambda + 1\ . \end{aligned} \qquad (3.267)$$

Taking the squares of Eqs. (3.249) we have

$$\langle Q(t)\rangle^2 = G\langle Q(0)\rangle^2 \cos^2 \nu t + G\langle P(0)\rangle^2 \sin^2 \nu t + G\langle Q(0)\rangle\langle P(0)\rangle \sin 2\nu t , \tag{3.268a}$$

and

$$\langle P(t)\rangle^2 = G\langle Q(0)\rangle^2 \sin^2 \nu t + G\langle P(0)\rangle^2 \cos^2 \nu t - G\langle Q(0)\rangle\langle P(0)\rangle \sin 2\nu t . \tag{3.268b}$$

To study the fluctuations of the field amplitudes, it suffices to consider the average values and derivatives at a particular time. In fact, this time should be the time T at which the signal leaves the amplifier. Thereafter, the signal will be free and evolve like a free harmonic oscillator. The phase of the output depends on the phase of the input and the length of the amplifier. Since we are not interested in the phase of the signal at this stage, we may assume that the time T that the signal spends inside the implifier is given by $T = M\frac{2\pi}{\nu}$, where M is a large integer. This assumption simplifies the equations somewhat without affecting the conclusions concerning the statistics of the amplitudes. The equations thus become

$$\langle Q^2(t)\rangle = 2G\langle a^+(0)a^+(0)\rangle^{(r)} + 2Gf(0) + 2(G-1)\lambda + 1, \tag{3.269a}$$

$$\langle P^2(t)\rangle = -2G\langle a^+(0)a^+(0)\rangle^{(r)} + 2Gf(0) + 2(G-1)\lambda + 1, \tag{3.269b}$$

$$\langle Q(t)\rangle^2 = G\langle Q(0)\rangle^2 , \tag{3.270a}$$

$$\langle P(t)\rangle^2 = G\langle P(0)\rangle^2 . \tag{3.270b}$$

Using the definitions of Q and P one can easily show that

$$2\langle a^+(0)a^+(0)\rangle^{(r)} = \frac{1}{2}\left[\langle Q^2(0)\rangle - \langle P^2(0)\rangle\right] \tag{3.271a}$$

and

$$\frac{1}{2}\left[\langle Q^2(0)\rangle + \langle P^2(0)\rangle\right] = 2f(0)+1 . \tag{3.271b}$$

By virtue of these relations, the quantity $2\langle a^{+}(0)\,a^{+}(0)\rangle^{(r)}$ can be eliminated from Eqs. (3.269a,b) which become

$$\langle Q^2(t)\rangle = G\langle Q^2(0)\rangle + (2\lambda - 1)(G-1)\,, \tag{3.272a}$$

and

$$\langle P^2(t)\rangle = G\langle P^2(0)\rangle + (2\lambda - 1)(G-1)\,. \tag{3.272b}$$

Combining these equations with Eqs. (3.270a,b) and (3.256), we obtain

$$\langle \Delta Q^2(t)\rangle = G\langle \Delta Q^2(0)\rangle + (2\lambda - 1)(G-1)\,, \tag{3.273a}$$

$$\langle \Delta P^2(t)\rangle = G\langle \Delta P^2(0)\rangle + (2\lambda - 1)(G-1)\,, \tag{3.273b}$$

and therefore

$$\mathcal{E}_Q^2 = \mathcal{E}_{Q_o}^2 + \frac{2\lambda - 1}{\langle Q(0)\rangle^2}\,\frac{G-1}{G}\,, \tag{3.274a}$$

$$\mathcal{E}_P^2 = \mathcal{E}_P^2 + \frac{2\lambda - 1}{\langle P(0)\rangle^2}\left(\frac{G-1}{G}\right). \tag{3.274b}$$

Recalling the definition of λ (see Eq. (3.210), we have

$$2\lambda - 1 = \frac{b_2 + b_1}{b_2 - b_1} \tag{3.275}$$

For active materials we shall have $2\lambda - 1 \geq 1$, the equality occurring when the lower level is empty. The quantities $\mathcal{E}_Q^2$ and $\mathcal{E}_P^2$ characterize the amplifier noise referring to measurements of Q and P, respectively. The quantities $\mathcal{E}_{Q_o}^2$ and $\mathcal{E}_{P_o}^2$ refer to the input signal. The other terms depend on the parameters of the device, and the input field amplitudes. They decrease as the field amplitudes

at the input increase. These terms are due to spontaneous emission. This follows from the fact that they are present even when the gain of the amplifier is unity, that is, when $b_2 - b_1 = 0$; and, on the other hand, they vanish for $b_2 = 0$. To see this, one recalls that

$$G = e^{(b_2 - b_1)} .$$

Then it follows that

$$\lim_{(b_2 - b_1) \to o} \left\{ (2\lambda - 1) \frac{G-1}{G} \right\} = 2b_2 t, \tag{3.276}$$

and this vanishes for $b_2 = 0$.

For large gain $(G >> 1)$, Equs. (3.274a, b) can be approximated by

$$\mathcal{E}_Q^2 \cong \mathcal{E}_{Q_o}^2 + \frac{2\lambda - 1}{\langle Q(0) \rangle^2} , \tag{3.277a}$$

$$\mathcal{E}_P^2 \cong \mathcal{E}_{P_o}^2 + \frac{2\lambda - 1}{\langle P(0) \rangle^2} \tag{3.277b}$$

It should be noted that large gain does not imply $b_2 >> b_1$ and, therefore, even for large gain $(2\lambda - 1)$ will not be equal to unity, in general. The relative values of b_2 and b_1 are determined by the physical properties of the active material, while the gain can be made as large as desired by increasing the length of the amplifier. From Eqs. (3.274a, b) we conclude that the quantum mechanical fluctuations at the input (which depend on the state of the input signal) go through the amplification process unchanged. In addition, one has the spontaneous emission noise whose relative importance decreases as the field amplitude of the input signal increases. The terms $\mathcal{E}_{Q_o}^2$ and $\mathcal{E}_{P_o}^2$ are due to what one might call quantum noise. They are a manifestation of the fact that a quantum variable cannot be determined precisely unless the system is in an eigenstate of the variable in question. The other terms in Eqs. (3.274a, b) are due to the internal noise of the amplifier that is spontaneous emission. The quantum amplifier, therefore, does not change

the quantum noise of the signal. It only adds to it the noise of spontaneous emission. Recalling the relationship between (Q, P) and (q, p), we can write Eqs. (3.274a, b) in the form

$$\mathcal{E}_q^2 = \mathcal{E}_{q_o}^2 + \frac{\hbar\omega/2}{\omega^2 \langle q(0) \rangle^2} (2\lambda - 1) \frac{G-1}{G} , \tag{3.278a}$$

and

$$\mathcal{E}_p^2 = \mathcal{E}_{p_o}^2 + \frac{\hbar\omega/2}{\langle p(0) \rangle^2} (2\lambda - 1) \left(\frac{G-1}{G}\right) . \tag{3.278b}$$

Again, the quantities $\mathcal{E}_q^2$ and $\mathcal{E}_q^2$ are dimensionless. The above equations indicate that the effect of internal noise on a measurement of q, for example, decreases as $\omega^2 \langle q(0) \rangle^2$ increases in comparison to $\hbar\omega/2$, and similarly for p. The significant conclusion is that the effect of spontaneous emission noise is determined primarily by the state of the signal at the input. To see it more clearly, note that for $b_1 = 0$ and large gain one may replace the quantity $(2\lambda - 1)\frac{G-1}{G}$ by 1. Then, it follows that the effect of spontaneous emission noise on the output signal increases as the average value of the input signal decreases. This is reminiscent of Gordon's (1962) conclusion about energy measurements.

3.6.9 Application to a Special Case.

Let us consider now the case in which the input signal is in a pure coherent state in the Glauber (1963) sense. Such a state is represented by $|\alpha\rangle$ and is defined by $a|\alpha\rangle = \alpha|\alpha\rangle$. Then we shall have

$$\langle Q(0) \rangle = 2 \operatorname{Re} \alpha , \quad \langle P(0) \rangle = 2 \operatorname{Im} \alpha . \tag{3.279}$$

It is straightforward to show that

$$\langle Q^2(0) \rangle = 4 (\operatorname{Re} \alpha)^2 + 1$$

and

$$\langle P^2(0) \rangle = 4 (\operatorname{Im} \alpha)^2 + 1 .$$

Consequently $\langle \Delta Q^2(0) \rangle = \langle \Delta P^2(0) \rangle = 1 .$ (3.280)

Then Eqs. (3.273a,b) give

$$\Delta Q^2(t) > = < \Delta P^2(t)> = G + (2\lambda - 1)(G-1) \quad . \tag{3.281}$$

The uncertainties associated with P and Q are equal because this is a property of the coherent state. And the amplifier preserves this property.

If we consider the limiting case in which $\lambda = 1$ and $G >> 1$, and transform to q and p, we find

$$<\Delta q^2(t)> \cong \frac{\hbar}{\omega} G \ , \tag{3.282a}$$

$$<\Delta p^2(t)> \cong \frac{\hbar}{\omega} G \ . \tag{3.282b}$$

Thus, we recapture the results of Louisell et al (1961) as a special case of a more general set of equations.

3.6.10 Fluctuations of the Number of Photons

The second moment of the distribution of the number of photons is given by $< a^+ a a^+ a>$. To compute this expectation value one proceeds as in the previous sections. Thus, a differential equation is obtained whose solution is

$$<a^+(t)a(t)\,a^+(t)\,a(t)> = \left[<a^+(0)\,a(0)\,a^+(0)\,a(0)> - \right.$$

$$\left. - \frac{b_2(b_2+b_1)}{(b_2-b_1)^2} + \left(f(0) + \frac{b_2}{b_2-b_1}\right)\left(\frac{3b_2+b_1}{b_2-b_1}\right)\right] G^2 -$$

$$- \left(f(0) + \frac{b_2}{b_2-b_1}\right)\left(\frac{3b_2+b_1}{b_2-b_1}\right) G + \frac{b_2(b_2+b_1)}{(b_2-b_1)^2} \ . \tag{3.283}$$

Let us introduce now the quantity

$$\Delta f^2(t) \equiv <a^+(t)\,a(t)\,a^+(t)a(t)> - f^2(t) \ , \tag{3.284}$$

which is the variance of the photon probability distribution. We shall calculate this quantity for the case $G >> 1$. Then one obtains the simpler result

$$\Delta f^2(t) = G^2 \left[\Delta f^2(0) + \frac{b_2 + b_1}{b_2 - b_1} f(0) + \frac{b_2^2}{(b_2 - b_1)^2} \right] \tag{3.285}$$

where only leading terms have been kept. For a more exact calculation, one will have to use all terms in Eq. (3.283).

The term $\Delta f^2(0)$ is the variance at the input and depends on the initial state of the field mode. The remaining terms are minimized for $b_1 = 0$ (i.e. when the lower level is empty) but never become zero. For the ideal amplifier (i.e., $b_1 = 0$), we have

$$\Delta f^2(t) = G^2 \Delta f^2(0) + G^2 \left(f(0) + 1 \right) . \tag{3.286}$$

Let us introduce the quantity

$$\mathcal{E}_f^2 \equiv \frac{\Delta f^2(t)}{f^2(t)} , \tag{3.287}$$

and denote by $\mathcal{E}_{f_o}^2$ its value at $t = 0$. From Eq. (3.255) we have $f(t) = G(f(0)+1)$ under the assumption that $G >> 1$ and the lower level is empty, in which case $\lambda = 1$.

Then we have

$$\mathcal{E}_f^2 = \frac{\Delta f^2(0)}{(f(0)+1)^2} + \frac{1}{f(0)+1} . \tag{3.288}$$

If the input signal is in an energy eigenstate, then $\Delta f^2(0) = 0$, and we obtain

$$\mathcal{E}_f^2 = \frac{1}{f(0)+1} . \tag{3.289}$$

This is due to spontaneous emission and it decreases as the number of input photons increases. Of course, this is valid under the assumption that the input was in an energy eigenstate. Such an eigenstate is not realizable in the form of a propagating signal. Thus, one will always have a finite $\Delta f^2(0)$ in most practical situations and it is Eq. (3.288) that is applicable.

3.6.11 Gain Selectivity of Quantum Amplifiers.

The analysis up to this point has shown that the amplifier gain is an exponential function of $k(\omega)$ and $k(\omega)$ is directly proportional to the line-shape function $g(\omega, \omega_o)$. Usually, $g(\omega, \omega_o)$ is a function of frequency more or less peaked about ω_o. A rather typical example is a Lorentzian centered at ω_o. This gives the amplifier a finite amplification bandwidth. In addition, the gain has a certain shape depending on the function $g(\omega, \omega_o)$. The amplifier will exhibit, therefore, a certain gain selectivity, and the question we attempt to answer in this section is whether or not this selectivity can be used to improve the signal-to-noise ratio at the detector.

Specifically, we shall consider the following situation: A signal with power spectral density $f_{is}(\omega)$ enters the input of the amplifier. Noise with power spectral density $f_{in}(\omega)$ accompanies the signal. These power spectral densities are here understood to represent number of photons per unit frequency. The output signal will be detected by a detector assumed to have a bandwidth $\Delta\omega$. For simplicity, its quantum efficiency is taken to be unity over the whole bandwidth $\Delta\omega$. We now wish to compare the signal-to-noise ratio at the input with the signal-to-noise ratio at the output.

The signal power spectral density $f(\omega)$ in photons per cycle is assumed to be appreciable only inside a bandwidth $\delta\omega$ and negligible otherwise. The bandwidth $\delta\omega$ is assumed to be smaller than $\Delta\omega$. One can easily convince oneself that if this is not the case, the preamplifier will reduce the signal-to-noise ratio because of its spontaneous emission noise.

Let us consider the input conditions first. The total input signal power is

$$S_i = \int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} f_{is}(\omega) \frac{d\omega}{2\pi} \qquad (3.290)$$

where ω_o denotes the carrier frequency which is assumed to coincide with the transition frequency of the two-level active material. The total added (external)

noise power at the input is

$$N_i^e = \int_{\omega_o - \frac{\Delta\omega}{2}}^{\omega_o + \frac{\Delta\omega}{2}} f_{in}(\omega) \frac{d\omega}{2\pi} \quad . \tag{3.291}$$

This noise may be assumed to be black-body thermal radiation of a certain temperature T_o, which arbitrarily will be taken as 290^oK for reference. Then we have, also in photons per cycle:

$$f_{in}(\omega) = \frac{1}{\exp \frac{\hbar\omega}{kT_o} - 1} \quad . \tag{3.292}$$

For the bandwidth of interest in this contract we may assume that $f_{in}(\omega)$ is constant in each application to a very good approximation. Thus, we may take

$$N_i^e = f_{in} \frac{\Delta\omega}{2\pi} \quad . \tag{3.293}$$

A signal generated at classical power level and attenuated coherently can be represented by a "coherent state" according to Glauber (1963). As pointed out in Section 3.6.9, the quantum fluctuations, or the zero point uncertainty, then has a spectral density of one photon per cycle. The same result has been arrived at through different reasoning by a number of authors (Stern, 1960; Gordon, 1962) and is discussed at length in Section 3.4.'

The total noise before amplification may consequently be represented by

$$N_i = (f_{in} + 1) \frac{\Delta\omega}{2\pi} \tag{3.294}$$

The input signal-to-noise ratio, which will be denoted by r_i, now is

$$r_i = \frac{S_i}{N_i} = \frac{\int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} f_{is}(\omega)\,\frac{d\omega}{2\pi}}{(f_{in}+1)\,\frac{\Delta\omega}{2\pi}} \quad . \tag{3.295}$$

Note that when the external noise f_{in} becomes much larger than 1, then the quantum noise can be neglected and one has again the classical formula.

Consider now the output conditions. From Eq. (3.253) we find that the output power spectral density $f_o(\omega)$ is given by

$$f_o(\omega) = G(\omega)\left(f_{is}(\omega) + f_{in}\right) + \left(G(\omega) - 1\right)\lambda \quad . \tag{3.296}$$

This contains the amplified signal, the amplified external noise, and the spontaneous emission noise. Thus the output signal power spectral density is

$$f_{os}(\omega) = G(\omega) f_{is}(\omega) \quad , \tag{3.297}$$

while the output noise power spectral density is

$$f_{on}(\omega) = G(\omega) f_{in} + \left(G(\omega) - 1\right)\lambda + 1 \, , \tag{3.298}$$

where we have again added the quantum noise. For gain much larger than unity, the quantum noise can be neglected. Although it is the case of large gain that we are especially interested in, we shall nevertheless retain the quantum noise in our formulation for the sake of consistency. Note that the quantum noise remains unaffected by the amplification process. Of course this is only a pictorial way of speaking since the quantum noise manifests itself upon detection and has to do with the uncertainty principle.

It must be emphasized that by taking one photon per unit bandwidth as the power spectral density of quantum noise, we do not exhaust all possibilities. In general, the fluctuations of quantum origin depend on the state of the system.

Here, we have in mind a field in a state of maximum entropy in the sense that Gordon (1962) defines it. This is to be considered as a standard case useful in evaluating the quantum amplifier.

Now, the total output signal and noise will be, respectively,

$$S_o = \int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} f_{os}(\omega) \frac{d\omega}{2\pi} = \int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} G(\omega) f_{is}(\omega) \frac{d\omega}{2\pi} , \tag{3.299}$$

and

$$N_o = \int_{\omega_o - \frac{\Delta\omega}{2}}^{\omega_o + \frac{\Delta\omega}{2}} f_{on}(\omega) \frac{d\omega}{2\pi} = \int_{\omega_o - \frac{\Delta\omega}{2}}^{\omega_o + \frac{\Delta\omega}{2}} \left[G(\omega) f_{in} + \left(G(\omega) - 1 \right) \lambda + 1 \right] \frac{d\omega}{2\pi} \tag{3.300}$$

If we denote by r_o the signal-to-noise ratio after amplification, we shall have

$$r_o = \frac{S_o}{N_o} = \frac{\int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} G(\omega) f_{is}(\omega) \frac{d\omega}{2\pi}}{\int_{\omega_o - \frac{\Delta\omega}{2}}^{\omega_o + \frac{\Delta\omega}{2}} \left[G(\omega) f_{in} + \left(G(\omega) - 1 \right) \lambda + 1 \right] \frac{d\omega}{2\pi}} . \tag{3.301}$$

Let us now introduce

$$R \equiv \frac{r_o}{r_i} , \tag{3.302}$$

and investigate the conditions under which R is greater than unity, in which case

the preamplifier improves the signal to noise ratio.

It has already been assumed that the signal power spectral density $f_{is}(\omega)$ is appreciable inside a bandwidth $\delta\omega$. We also have found that the amplifier gain is appreciably larger than unity only inside a certain bandwidth determined by the function $g(\omega, \omega_o)$. At this point we reinterpret $\delta\omega$ as being the larger one of the signal and the amplifier bandwidths. It is again understood that $\delta\omega$ is smaller than the detector bandwidth, otherwise no quantum amplifier can improve the situation. In fact, what we are essentially attempting to determine is how small should $\delta\omega/\Delta\omega$ be in order to achieve a desired value for R.

Let $\bar{G}$ be an average gain defined by

$$\bar{G} = \frac{1}{\delta\omega} \int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} G(\omega) d\omega . \tag{3.303}$$

Recalling that f_{in} is constant inside the bandwidths of interest, and after some straightforward manipulations, we find

$$N_o = (\bar{G} - 1)(f_{in} + \lambda)\delta\omega + (f_{in} + 1)\Delta\omega . \tag{3.304}$$

Let us also define an effective gain G_e given by

$$G_e = \frac{\int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} G(\omega) f_{is}(\omega) d\omega}{\int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} f_{is}(\omega) d\omega} . \tag{3.305}$$

In terms of this effective gain we have

$$S_o = G_e \int_{\omega_o - \frac{\delta\omega}{2}}^{\omega_o + \frac{\delta\omega}{2}} f_{is}(\omega)\, d\omega \,, \tag{3.306}$$

and upon combining this with Eqs. (3.301) (3.302), and (3.305), we obtain

$$R = \frac{G_e}{(\bar{G}-1)\left(\frac{f_{in}+\lambda}{f_{in}+1}\right)\left(\frac{\delta\omega}{\Delta\omega}\right)+1} \,. \tag{3.307}$$

In order to compress notation, let us introduce the quantity

$$\zeta = \left(\frac{f_{in}+\lambda}{f_{in}+1}\right)\left(\frac{\delta\omega}{\Delta\omega}\right) \,, \tag{3.308}$$

in terms of which R becomes

$$R = \frac{G_e}{(\bar{G}-1)\zeta+1} \,. \tag{3.309}$$

From Eq. (3.292) one can easily see that for the range of frequencies of interest to this contract, we shall have either $f_{in} \cong 1$ or $f_{in} < 1$. As for λ, for almost all practical cases one can assume that $\lambda < 10$. Consequently, the order of magnitude of ζ will be determined primarily by the order of magnitude of $\delta\omega/\Delta\omega$. The bandwidth $\Delta\omega$ depends on the detector, while $\delta\omega$ depends on the signal and the amplifier. One can probably take the value 50 Mcps as a rough lower bound for $\delta\omega$. As for $\Delta\omega$, one would expect it to be at least four orders of magnitude larger than $\delta\omega$. Thus, a rough estimate of ζ would be from 10^{-1} to 10^{-4}. Of course these values are not intended to represent precise upper and lower bounds but rather typical values of ζ. It should be evident now that the

magnitude of R will depend substantially on the relative magnitude of G_e and G. Since G_e depends on the spectral shape $f_{1s}(\omega)$ of the signal, so does R. There is a special case, however, in which R is independent of the spectral shape of the signal. This happens when $G(\omega)$ has a square form and we shall examine this case in some detail.

Let us assume that $G(\omega)$ is given by

$$G(\omega) = G_o \quad \text{for} \quad \omega_o - \frac{\delta\omega}{2} < \omega < \omega_o + \frac{\delta\omega}{2} \tag{3.310}$$

$$= 1 \text{ otherwise.}$$

Then, from Eqs. (3.303) and (3.305) we obtain

$$\bar{G} = G_e = G_o \tag{3.311}$$

and consequently

$$R = \frac{G_o}{(G_o - 1)\zeta + 1} \quad . \tag{3.312}$$

If the gain is much larger than unity, R can be approximated by

$$R = \frac{G_o}{G_o \zeta + 1} \quad . \tag{3.313}$$

If, in addition, ζ is such that $G_o \zeta << 1$, we shall have

$$R \cong G \, . \tag{3.314}$$

This means that for gain $G_o = 10$, for example, we have an improvement of 10 dbs. Of course, Eq. (3.314) is also valid whenever $G_o \zeta << 1$ independently of the order of magnitude of G_o.

In the general case in which the conditions $G_o >> 1$ and $G_o \zeta >> 1$ are not necessarily satisfied, one should investigate the behavior of R as a function of G_o and ζ. The results of this investigation are given in Fig. 3.6.14 where we have

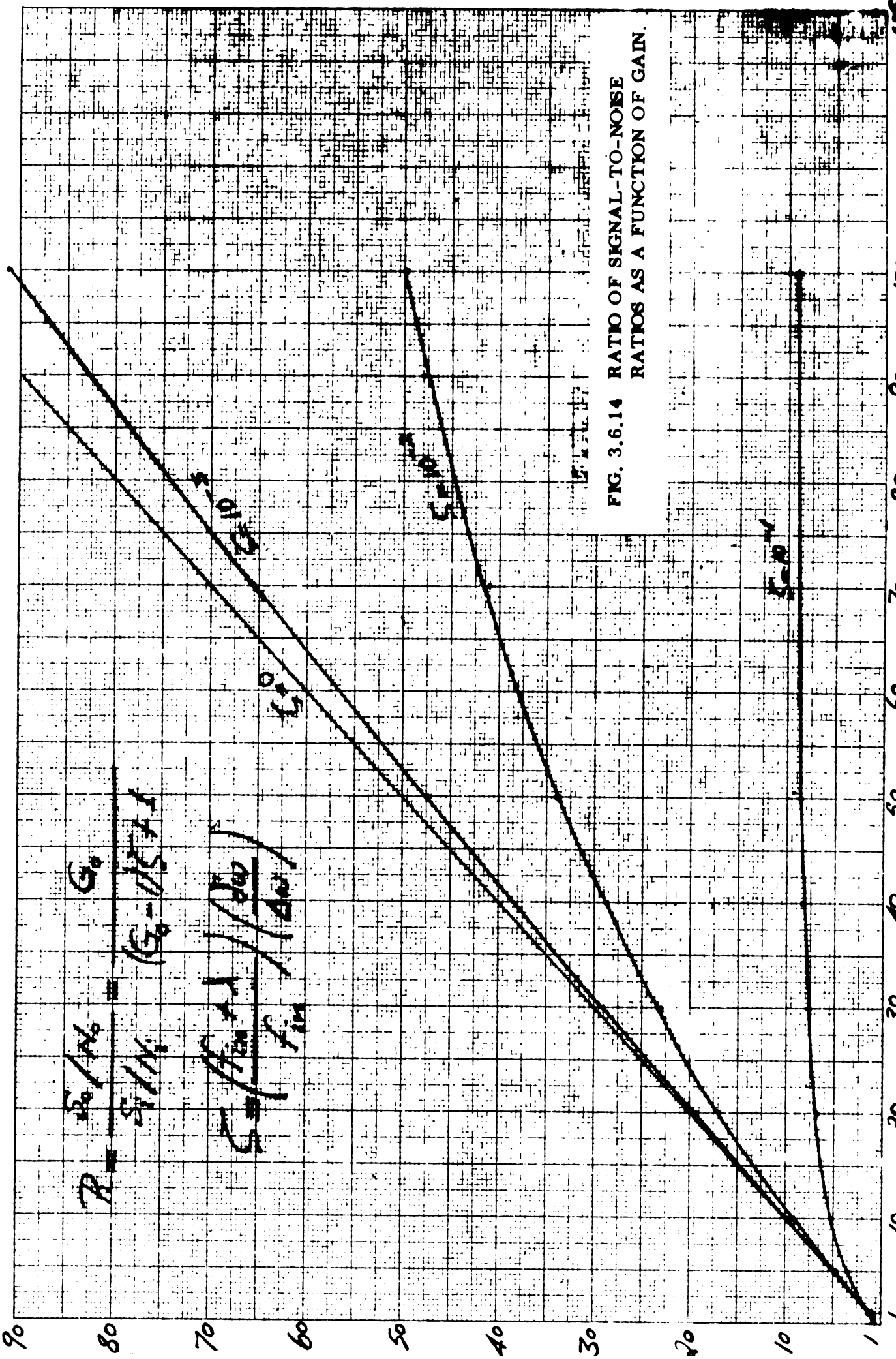

FIG. 3.6.14 RATIO OF SIGNAL-TO-NOISE RATIOS AS A FUNCTION OF GAIN.

plotted R as a function of G_o for three values of ζ. From this Figure it is evident that, for a given ζ, R increases rather fast as the gain increases up to a certain point. Further increase of the gain does not give any substantial improvement. One also sees that improvement of the order of 10 db can be obtained with a rather low gain (approximately $G_o = 10$), as long as ζ is smaller than 10^{-2}. Values of ζ much smaller than 10^{-2} should not be difficult to achieve. Then, one could have, for example, 20 dbs improvement with gain approximately 100.

A comment as to the actual shape of $G(\omega)$ is in place at this point. Recall that

$$G(\omega) = e^{Bk(\omega)} . \tag{3.315}$$

In many practical problems, the shape of $k(\omega)$ can be approximated by a Lorentzian, i.e.

$$k(\omega) \sim \frac{\gamma}{(\omega-\omega_o)^2 + \gamma^2} , \tag{3.316}$$

where B and γ are constants. Then, $G(\omega)$ will be the exponential of a Lorentzian. In Fig. 3.6.15 we have plotted $k(\omega)$ normalized to unity. We have also plotted $G(\omega)$ for maximum gain 20 and 150, again normalized to unity. As is seen from the Figure, considerable narrowing, and hence selectivity, results because of the exponential form of $G(\omega)$. That is, the selectivity of the material is enhanced as the gain is increased. The selectivity is enhanced further when the signal itself is peaked about ω_o.

The results of this section indicate that the use of a quantum amplifier in front of a detector with detection bandwidth larger than the amplification bandwidth, in general improves the signal to noise ratio at the detector. The improvement increases as the gain increases and as the parameter ζ decreases. The analysis of the case of square gain indicates that as the gain increases, one should decrease ζ in order to achieve maximum improvement. That is, increasing the gain beyond a certain point does not improve the situation appreciably without further decrease of ζ.

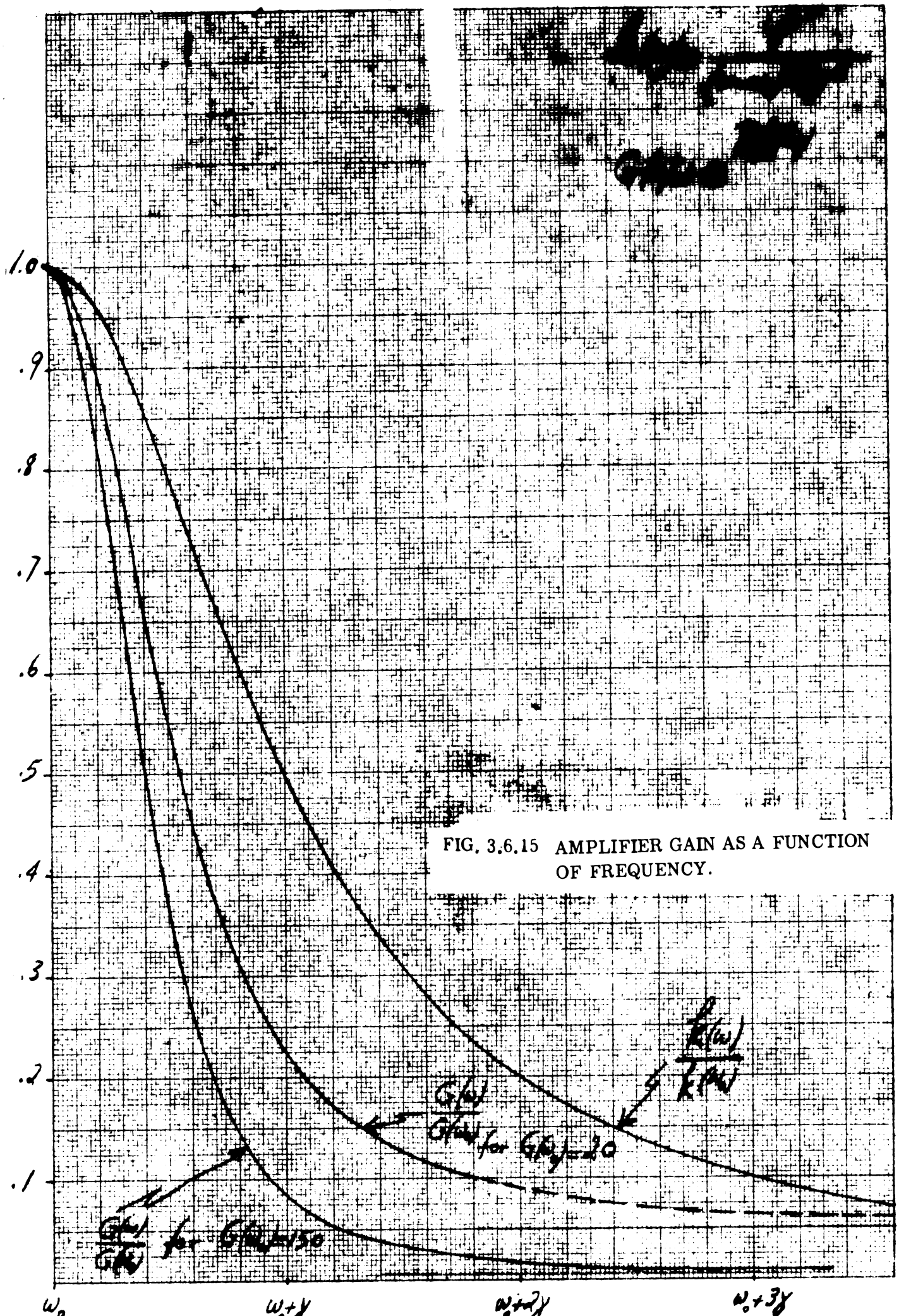

FIG. 3.6.15 AMPLIFIER GAIN AS A FUNCTION OF FREQUENCY.

3.6.12 A Note on the Effect of Attenuation on Signal Coherence.

A pertinent question is whether the coherence properties of a signal are affected by the attenuation that the signal may suffer during transmission. If the attenuation is linear, the question can be partially answered on the basis of the foregoing analysis. By assuming that $b_1 > b_2$, and hence $G < 1$, we have a model for a linear attenuator. Then, according to the results of Section (3.6.8) the quantum uncertainties ξ_q and ξ_p associated with field measurements remain unchanged except for the fact that the noise of spontaneous emission is added (see Eqs. (3.278a, b)). Also, from the analysis of Gordon et al (1963a), follows that the moments of the field distributions are simply multiplied by respective powers of the gain, i.e., if the nth input moment is M_n, the same output moment will be $G^{n/2}M_n$. Thus, as far as the field amplitude distributions and their moments are concerned, one may say that coherence is not destroyed by attenuation. In other words, the phase relationship between field amplitudes will not be destroyed. However, it should be pointed out that this does not imply that coherence in its most general sense is preserved. For example, it is conceivable that a detection scheme might be based on an n fold coincidence measurement. In that case one does not know whether coherence is preserved or not. Further research is necessary before an answer to this question becomes possible. In addition, at the present time it is not known whether scattering will destroy the coherence of the signal either in the limited or the general sense. The previous conclusions were based on the assumption of linear attenuation. If the attenuation is due to absorption, this assumption is satisfied. But it is not clear at this time whether attenuation due to scattering can be considered as a linear process. This is another question which must await further research.

IV

OPTICAL BAND PASS FILTERS FOR COMMUNICATION CHANNELS

4.1 Survey and a Preliminary Evaluation of Known Filter Principles

The third problem area selected for this study is motivated by the importance of minimizing by optical means the background radiation incident on a quantum-counter detector in a communication system, since in this detection scheme no opportunity of intermediate-frequency frequency discrimination is offered.

Filtering of optical and infrared radiation in the wavelength interval 0.4 to 20μ is generally accomplished by the use of interference or absorption type filters. There are, however, other kinds of filters available for special applications. Optical filters can be classified according to the physical mechanism involved in their operation as: (1) selective absorption; (2) selective reflection, (3) scattering; (4) interference; (5) polarization; and (6) refraction (Greenler, 1958). We will briefly discuss some filters now being used, their characteristics and the possibility of changing their center wavelengths or of tuning them.

Glass color filters and gelatin filters are commercially available in the 0.2 to 1.0 micron range. These filters operate on the selective absorption principle and their pass band characteristics are generally not very sharp. They can be useful as bandpass filters in conjunction with some other narrow-band filter to eliminate multiple pass bands of the latter. Tuning of such filters is not possible.

Transmission interference filters (Greenler, 1958), in the simplest form a Fabry-Perot interferometer, are available commercially with transmission peaks of 30 percent and a width of 150 Å in the range 3900 Å to 7800 Å. When the semi-reflecting metal films of the Fabry-Perot interferometer are replaced with quarter-wavelength stacks of alternately high and low indexes of refraction, narrower pass bands and higher peak transmissions can be obtained, while the sidebands are decreased. In this way peak transmission of 73 percent and pass bands of 70 Å are typically achieved; although with some sacrifice in the peak transmission, pass bands as narrow as 20 Å are possible. At these frequencies, a wavelength pass band of 20 Å corresponds to a frequency pass band of about 10^{12} cps. (Fock et al,

1964). Interference filters in the 1.4 to 4.5 micron range are available with pass bands of about 1/50 of the center frequency, or about 10^{13} cps. The interference filter can also be made as a reflection, rather than a transmission filter. Hadley and Dennison (1947) discuss both thoroughly. The interference type filter is feasible to tune; this possibility will be discussed below.

Another filter which is useful in the visible and infrared range is the Christiansen filter (Greenler, 1958, Kruse et al, 1962, Holter et al, 1962). In this filter a powdered dielectric is suspended in an appropriate fluid. The two components are chosen so that their refractive-index vs wavelength curves have different slopes and cross near a desired wavelength. At the point where the refractive indices are the same, the radiation passing through the filter is not deflected. On either side of this point, however, the particles of dielectric scatter the radiation, thereby attenuating it. Pass bands of 200 Å in the range 4000 to 8000 Å have been obtained. The width of the pass band is determined by the relative slopes of the two refractive index curves. The fact that the index of refraction of a liquid changes more rapidly with temperature than that of a solid has been used to make a tunable band-pass filter (Greenler, 1958).

Clever use of the birefringence of crystal quartz plates has been applied in the design of the birefringent filter (Greenler, 1958) It is also called the interference polarization filter or Lyot-Ohman filter. Quartz plates are cut with their optical axes parallel to their large faces, and these plates are mounted between polarizers whose axes are oriented at 45^{o} with respect to the quartz optical axes. The incident radiation is converted to linearly polarized light at the first polarizer, and thereafter only that part of the light whose axis of polarization is parallel to the polarizer axis will pass. If each succeeding quartz plate is twice as thick as that which precedes it, then in a given wavelength interval it will have twice as many maxima and minima. The transmission of the entire assembly of plates and polarizers is the product of the transmission of the individual sections, and there are now sharp transmission peaks separated by λ_o/N where λ_o is the lowest-order wavelength

for maximum transmission, and N is the order of the maxima. Filters using 6 quartz plates have been made with a pass band of 4.1 Å at 6563 A (Greenler, 1958). Some commercial filters are available with pass bands as narrow as 1 Å. These filters have a potential for tuning.

There is another type of interference filter called the frustrated total-reflection filter which has been used in the vicinity of 5000 Å (Fock et al, 1964). Its operation is based on the fact that the reflection is not total when radiation is incident on a boundary between high and low index of refraction materials at an angle greater than the critical angle, if the low-index of refraction material is thin enough. The filter is constructed of two flat triangular pieces of glass joined along their hypotenuses but separated by two thin layers of low refractive index and a high index of refraction spacer in between. The wavelength of maximum transmission is determined by the optical thickness of the spacer, the angle of reflection in it and the phase change on reflection while the width of the pass band depends on the thickness of the low refractive index layer. Bandwidths of 120 Å to 30 Å have been achieved with peak transmission of 93 percent to 12 percent.

Two other filters which are restricted primarily to the infrared range are commonly used. They are the semiconductor filter and the residual-ray plate or reststrahlen filter (Greenler, 1958; Kruse et al, 1962; Holter, et al, 1962). The semiconductor filter is a low-pass filter which absorbs wavelengths shorter than that associated with its energy gap between the valence and conduction bands. The cut-off characteristics are quite sharp, but because impurity atoms can introduce electron energy levels between the valence and conduction bands, the impurity content must be kept low if absorption beyond the cut-off wavelength is to be minimized. The semi-conductor filter has a cut-off wavelength in the range of 1 to 8 microns and can be used with interference filters to obtain a rather narrow pass band and to eliminate the side bands of the interference filter. Since the width of the energy gap of the semi-conductor is temperature-dependent, the cut-off wavelength may exhibit a temperature-dependent shift.

The residual-ray plate filter depends for its operation on the ionic structure of a crystal. When radiation impinges on the atoms in a crystal, the electric field tends to set the ions into motion. A resonance will occur which is related to the restoring-force constant of the crystal and the masses of the two ions. The crystal will re-emit a band of wavelengths about this natural frequency. The principal restrictions on the use of the residual-ray plate filter is that there is only a limited number of center wavelengths available. In addition, they are limited generally to applications at wavelengths greater than 15 microns.

One other choice that is commonly used to filter light or disperse it is the prism or grating spectroscope. Such a device is essentially one which converts a frequency spectrum into a spatial spectrum by means of a dispersive element. The resolving power of a rock-salt prism spectroscope is about 200 at 5 microns and 400 at 14 microns which is equivalent to a pass band on the order of 300 Å at 10 microns wavelength (Harrison et al, 1961). The grating instrument in the wavelength range 3 to 25 microns has no better resolution than a well-designed prism spectroscope, since the theoretical resolving power of the grating is seldom realized in infrared work (Brady, 1950). This is due to the fact that the limit is set not by the grating, but by the slit width which must be employed for adequate illumination. The present practical limit of the resolving power of the grating is 8000 from 1.5 to 7 microns, or a pass band on the order of 6 Å at 5 microns wavelength. At 20 microns wavelength the pass band would be about 160 Å. The spectroscope may be of some interest in designing a tunable filter.

In the application in which we are interested, the ideal filter bandwidth is on the order of 10^9 cps, the expected modulation bandwidth. The corresponding passbands at 0.4 microns and 20 microns wavelength are 5.3×10^{-3} Å and 13 Å, respectively. If we recall the passband widths quoted above for present commercial filters, we observe that these desired bandwidths are 1 to 2 orders of magnitude smaller than that is available with the filters discussed. It appears that the best chance for obtaining the desired bandwidth occurs at the longer wavelengths of the

interval 0.4 to 20 microns.

When the question of tuning a filter is considered, of these filters mentioned above, two kinds seem easily adaptable for tuning, the interference and the Christiansen filter. The spectroscope may also have some potential in this area, as mentioned previously. Tuning of the interference filter is feasible by electro-optic, piezoelectric and perhaps also magneto-optic and magneto-strictive means, while the Christiansen filter is, in principle, at least, capable of electro-optic tuning. A quite extensive amount of work has been published in recent years in this connection concerning the modulation of light by application of electric or magnetic fields to optic materials. We mention briefly here the common electro- and magneto-optic effects and present some examples where they have been used.

The electro-optic effect in a crystal is a change in the refractive index with the strength of the applied electric field. If the change is linearly related to the field strength, it is known as the Pockel's effect. The coherent-light phase modulator using the Pockel's effect was described by Peters, (1963) who used a slow-wave microwave structure to provide a modulating electric field in a crystal of ammonium dihydrogen phosphate. The experimental evidence suggests that a modulation bandwidth several octaves greater than 1 Gc could be obtained. Kaminow (1961) carried out a somewhat similar experiment, using a crystal of potassium dihydrogen phosphate in a cylindrical cavity and a modulation frequency of 9.25 Gc. Gil'varg and Kolesov (1961) reported on the use of these two crystals in a high-speed shutter and Niblack and Wolf (1964) report on a polarization modulation-demodulation system utilizing a Pockel's cell for the modulation of a continuous-wave helium-neon gas laser as the light source.

When the change in the refractive index is proportional to the square of the applied electric field, it is known as the Kerr electro-optic effect, which should be distinguished from the Kerr magneto-optic effect which will be mentioned later. The Kerr electro-optic effect is characteristic of a crystal with a center of symmetry (Nye, 1960); if the crystal has no center of symmetry, then the electro-optic

effect may be linear. Holshouser, et al (1961) constructed a Kerr-cell microwave light modulator which used a re-entrant microwave cavity to provide the modulation field and which achieved up to 80 percent modulation at 3 Gc.

The magneto-optic effect may also be classified as linear or quadratic. The linear magneto-optic effect is called the Faraday effect; when the effect is proportional to the square of the magnetic field it is known as the Cotton-Mouton effect. Schmidt et al (1964) have used the enhanced Faraday effect near the absorption line of sodium vapor to modulate the plane of polarization of light traveling through the vapor. Pulse modulation was carried out up to 698 Mc with no evidence of reduction in frequency response. Fock and Bradley, (1964) reported on the Faraday rotation in mercury vapor near an optical resonance. Since only light in the vicinity of the resonance experiences appreciable rotation, they suggest that by using a pair of crossed polarizers with the cell a narrow-band modulated filter, having a pass band of about 1500 Mc can be realized.

Another magneto-optic effect which may be of interest is the Kerr magneto-optic effect. This term refers to a change in the polarization or intensity of light reflected from the surface of a magnetized medium. When we speak of the Kerr effect hereafter, we will mean the Kerr electro-optic effect, unless we specify otherwise.

The piezoelectric effect is also of interest in this context. It is exhibited by some crystals in which a strain is set up when an electric field is applied; conversely, a stress produces an electric polarization. This effect has been used by Hauser et al (1963) in designing a stressed-plate shutter. Seraphin et al (1963) employed the piezoelectric effect in designing a Fabry-Perot type light modulator using a quartz spacer, the thickness of which was changed piezoelectrically. Another interesting application of this effect was made by Ashtheimer et al (1964) who designed a frustrated total internal reflection infrared modulator. Its principle of operation is similar to that of the frustrated total-reflection filter which was discussed previously. In this case, however, there is only air separating the two

halves of the device, and the separation is controlled piezoelectrically. Since the transmission falls off as $\sinh^{-2}\frac{d}{\lambda}$ where d is the separation and λ the wavelength, we see that decreasing λ or increasing d decreases the transmission. It can thus act as a low-pass filter whose cut-off frequency can be altered by changing d. It could possibly be used with an interference filter to make a tunable band-pass filter.

The magnetostrictive effect, that is distortion produced in certain materials when subjected to a magnetic field, has also been used in this area. Bennet and Kindlemann (1962) report on the use of magnetostriction for making angular and separation adjustments on the end plates of a helium-neon maser. The device was constructed with the eventual use in mind of achieving negative-feedback frequency stabilization.

As discussed above, the electro-optic effect involves a change of the index of refraction with applied electric field. If a Christiansen filter were to be made using an electro-optic fluid, then the cross-over point of the refraction indexes of the two materials composing the filter could be shifted in frequency, providing a tuning mechanism for the filter. From a practical standpoint, however, the relative slopes required for the dispersion relations of the two materials in order to obtain a sufficiently narrow pass band may be unrealistic considering the properties of materials presently used for such filters, particularly since one of them must be an electro-optic material. In addition, the tuning range that can be achieved may be too limited to make such a filter of interest. These questions will be considered more fully later.

The various interference filters such as the polarization, frustrated internal reflection and reflection and transmission filters are tunable in principle, though the techniques of accomplishing the tuning will be different. For the present we discuss briefly the transmission-type filter.

The transmissivity of the interference transmission filter for normal incidence is given by:

$$T = \frac{1}{\left[1+\frac{A}{T}\right]^2} \frac{1}{\left[1+\frac{4R}{1-R}^2 \sin^2\frac{\delta}{2}\right]} \tag{4.1}$$

where R, T, and A are the spectral reflectivity, transmissivity and absorptivity of a single reflection layer of the filter and

$$\delta = \frac{4\pi nd}{\lambda} + 2\phi \quad . \tag{4.2}$$

In this formula, n is the refractive index of the medium between the two layers, d is the layer separation, λ is the free-space wavelength and ϕ is the phase shift for reflection from one layer.

We observe that T is a maximum whenever $\sin\frac{\delta}{2} = 0$, i.e., when

$$\frac{2\pi nd}{\lambda} + \phi = N\pi \quad , \qquad N = 1,2,3 \ldots \tag{4.3}$$

N is referred to as the order of the interference fringe. We see that the wavelength of maximum transmission for a given order of interference N is determined by n, d and ϕ. Thus, the filter should be tunable by adjusting any one or more of these three parameters. If for the moment we neglect changes in ϕ and consider only n and d variables, and if we wish to tune the filter over the interval λ_1 to λ_2 where $\lambda_2 = \Delta\lambda_1$, then

$$\lambda_1(\Delta - 1) = \frac{2}{N}(n_2d_2 - n_1d_1) \tag{4.4}$$

n_1, d_1 and n_2, d_2 are the index of refraction and the separation at the ends of the tuning band. We can rearrange this expression to get

$$n_2d_2 = \Delta n_1d_1 = \frac{\Delta N\lambda_1}{2} \quad . \tag{4.5}$$

So we see that the product of the index of refraction and the separation must be changed in proportion to the desired change in center wavelength λ_1 of the pass

band. In addition, the absolute change in the product of nd is proportional to N, so that the change required is a minimum for a given λ when N=1 .

To get an idea of the order to magnitude of the quantities involved, suppose we consider piezoelectrically tuning an air-filled Fabry-Perot interferometer over an octave bandwidth from 5 to 10 microns wavelength. Then λ_1 = 5 microns, n=1, and N=1, d_2 = 5 microns, d_1 = 2.5 microns. If barium titanate, which has the largest piezoelectric constant known, 3×10^{-4} microns per volt, (Astheimer et al, 1964), were to be used to accomplish the 2.5 micron change in separation, then a voltage of almost 10,000 volts would be required. This large voltage can be decreased by placing slabs of piezoelectric in series mechanically, but in parallel electrically. And since the relative displacement of the plates of the filter could be reversed by changing the direction of the electric field, the 10,000 volt requirement could be decreased to 5,000 volts. The requirements for tuning the simple Fabry-Perot interferometer do not seem unreasonable then.

The frustrated internal reflection filter and the Christiansen filter together suggest another possible filter configuration that would be tunable. We will refer to this configuration, which has the geometry of the Wernicke prism (Strong, 1958) as the Wernicke filter. A cross-section of this filter is shown in Fig. 4.1.

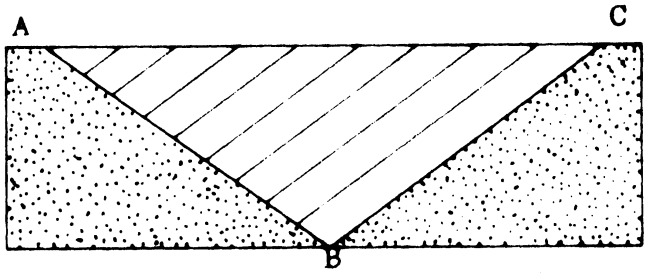

FIG. 4.1. Sketch of a Wernicke Filter.

If the two materials of which this filter is composed have refractive index relations vs wavelength with different slopes, which cross over at some wavelength, then at this particular wavelength, the incident radiation will pass through the filter with no deflection. At wavelengths longer or shorter than this, however, reflections will

occur at interfaces AB and CB. For wavelengths where the indexes of refraction are sufficiently different, total internal reflection will take place at these interfaces, but on opposite sides of the center wavelength. If now, one or the other of the materials is electro-optic in nature, the center wavelength could be changed. The consideration mentioned above with respect to the Christiansen filter apply here also. There is an additional feature here, however, in that the radiation emerging from the Wernicke filter would be spatially dispersed on either side of the center wavelength, providing a potential additional filtering mechanism.

We have reviewed here some common methods used for filtering in the optical and near infrared regions of the spectrum. Some possible mechanisms by which a filter might be tuned were discussed, and some examples of the application of these mechanisms in the modulation of light and infrared radiation were presented. Some possible schemes for constructing a tunable filter were considered.

In the next section we shall investigate some specific filter designs more thoroughly, particularly concerning theoretical limitations on their band pass characteristics and tunability.

4.2 The Wernicke Prismatic Filter

This prism-type filter was mentioned briefly in the previous section. It was suggested that this filter, which has the geometrical configuration of the Wernicke prism, has some possible potential for tuning. This section of the report discusses the transmission characteristics of this filter and the practicality of tuning it.

4.3 Development of Filter Transmissivity and Calculations

The filter configuration is shown in Fig. 4-2 with the dimensions and other parameters indicated. The refractive indices of the two materials of which the prism is made are n_1 and n_2. In this development we will assume there is an entrance slit of height 2H through which passes a parallel beam of light. Light which is reflected out of the primary beam within the prism will be assumed to be absorbed at the walls so that no secondary beams are considered. Absorption within the prism is neglected for simplicity; it will be discussed later. Diffraction effects at the entrance and exit slits are also neglected. The light will be assumed to be polarized with the electric vector parallel to the entrance slit. The Gaussian system of units is used.

Light which is normally incident on the first air-glass interface will be partially reflected, and the amplitude E_{t1} of the transmitted electric vector will be, for unit incident amplitude,

$$E_{t1} = \frac{2n_o}{n_1 + n_o} \tag{4.6}$$

At AB where the light falls obliquely on the interface, the transmitted amplitude is given by

$$E_{t2} = \frac{E_{t1}\, 2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\phi_1} \tag{4.7}$$

and from Snell's law we have

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 . \tag{4.8}$$

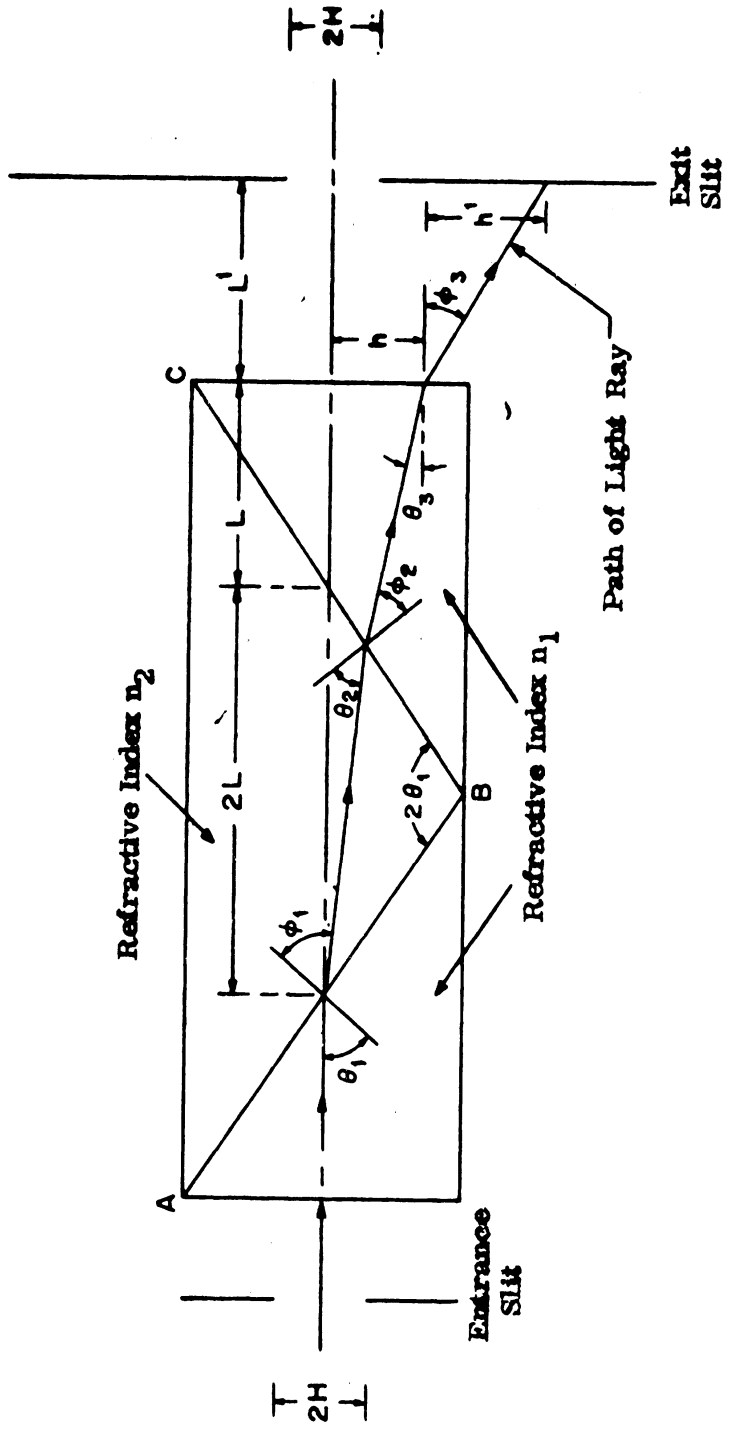

FIG. 4-2: DIAGRAM OF THE WERNICKE FILTER

The amplitude of the transmitted electric vector at the remaining interfaces is found in a similar manner, so that the electric vector of the emerging wave is given by

$$E_{to} = \frac{2^4 n_o n_1^2 n_2 \cos\theta_1 \cos\theta_2 \cos\theta_3}{\left[(n_1+n_o)(n_1\cos\theta_1+n_2\cos\phi_1)(n_2\cos\theta_2+n_1\cos\phi_2)(n_1\cos\theta_3+n_o\cos\phi_3)\right]} \tag{4.9}$$

where

$$\phi_1 = \sin^{-1}\left[\frac{n_1}{n_2}\sin\theta_1\right] \tag{4.10}$$

$$\theta_2 = 2\theta_1 - \phi_1 \tag{4.11}$$

$$\phi_2 = \sin^{-1}\left[\frac{n_2}{n_1}\sin\theta_2\right] \tag{4.12}$$

$$\theta_3 = \theta_1 - \phi_2 \tag{4.13}$$

$$\phi_3 = \sin^{-1}\left[\frac{n_1}{n_o}\sin\theta_3\right] . \tag{4.14}$$

In order to obtain the intensity of the output light, we square the amplitude of the electric vector as given by (4.9) and multiply by the ratio of the cosine of the angle of incidence to the cosine of the angle of transmission and the ratio of the refractive indices at each interface to find the power flow through the interfaces and obtain

$$I_o = \frac{E_{to}^2 \cos\phi_1 \cos\phi_2 \cos\phi_3}{\cos\theta_1 \cos\theta_2 \cos\theta_3} \tag{4.15}$$

The refractive index terms which appear in the power flow expression have cancelled out.

The light power passed by the exit slit will be further decreased due to the dispersive effect of the prism. If h is the total displacement of the beam, then the total light power passed by the exit slit will be

$$P_o = I_o \left(1 - \frac{h}{2H}\right) A \tag{4.16}$$

where we have assumed an exit slit of the same height as the entrance slit and A is the area of both slits. The displacement of the output light beam is found to be

$$h = \frac{L}{\sin\theta_1} \left\{ \left[3\sin\theta_1 - \frac{\cos(\theta_1 - \theta_2 \sin 2\theta_1}{\cos\theta_2} \right] \tan\theta_3 + \frac{\sin 2\theta_1}{\cos\theta_2} \sin\left|\theta_1 - \theta_2\right| \right\} . \tag{4.17}$$

This displacement is measured with respect to the point at which an undeflected ray would emerge from the filter, and is taken to be positive in either direction from the undeflected position. If the exit slit is located a distance L^1 from the last glass-air interface, then an additional displacement h^1 results,

$$h^1 = L^1 \tan\phi_3 \quad . \tag{4.18}$$

Finally then if the ratio of the output to input power is formed, the power transmission coefficient τ is obtained, as

$$\tau = I_o \left[1 - \frac{h+h^1}{2H} \right] \tag{4.19}$$

where now I_o is dimensionless, since we have divided by the power in the incident field. When this expression becomes negative, the output is then, of course, zero.

Some calculations have been carried out for the Wernicke filter as a function of the ratio $\frac{n_1}{n_2}$, for an angle θ_1 of 60^o and a ratio $\frac{n_1}{n_o}$ of 1.5. Figure 4.3 shows the quantity I_o as $\frac{n_1}{n_2}$ is varied. The transmission abruptly goes to zero when $\frac{n_1}{n_2} = 1.152$ since then the light no longer is incident on the exit face of the prism, but strikes the bottom surface. Figures 4.4 and 4.5 show the displacement ratios

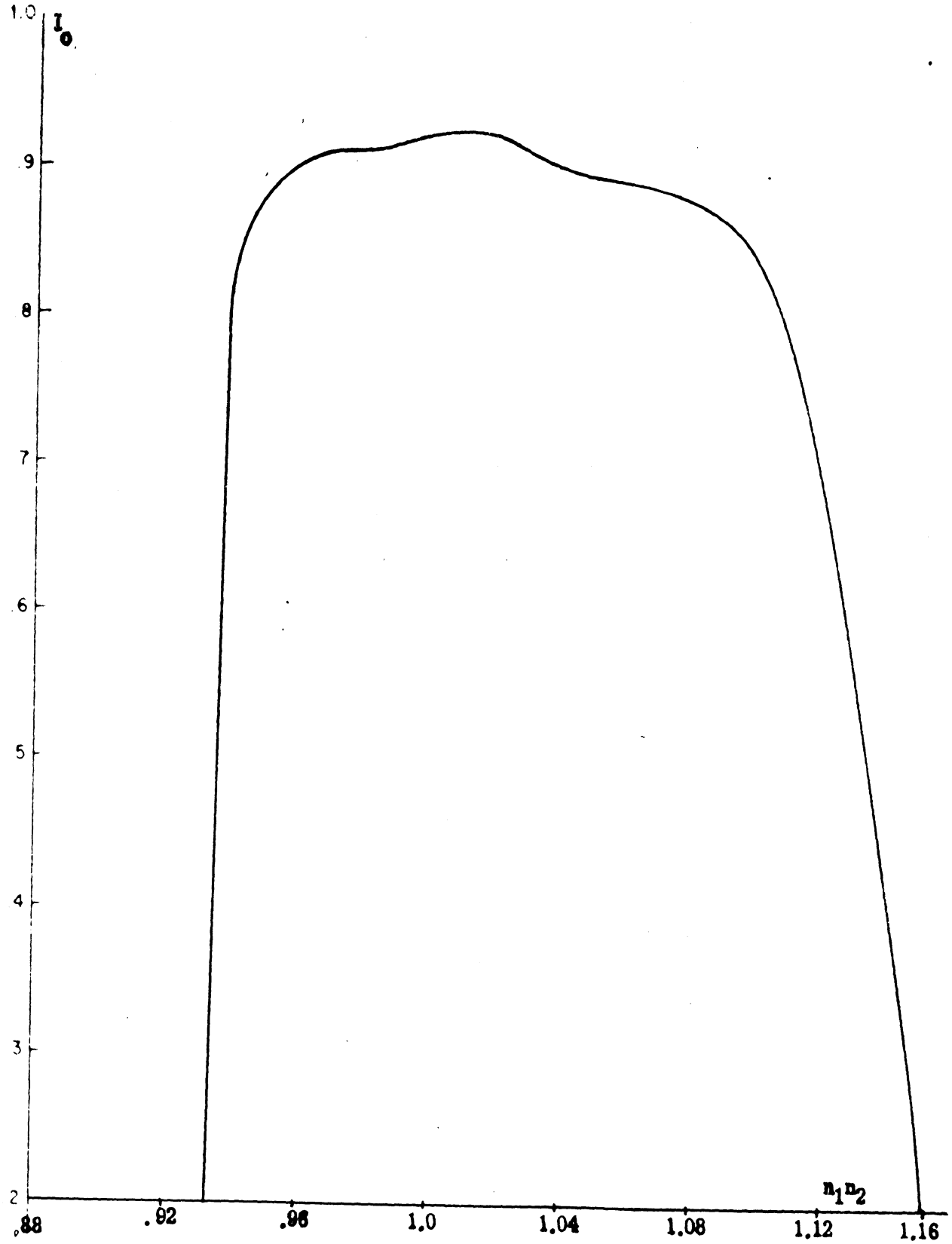

FIG. 4.3: TRANSMITTED LIGHT INTENSITY, I_o, AS FUNCTION OF n_1/n_2=1.5

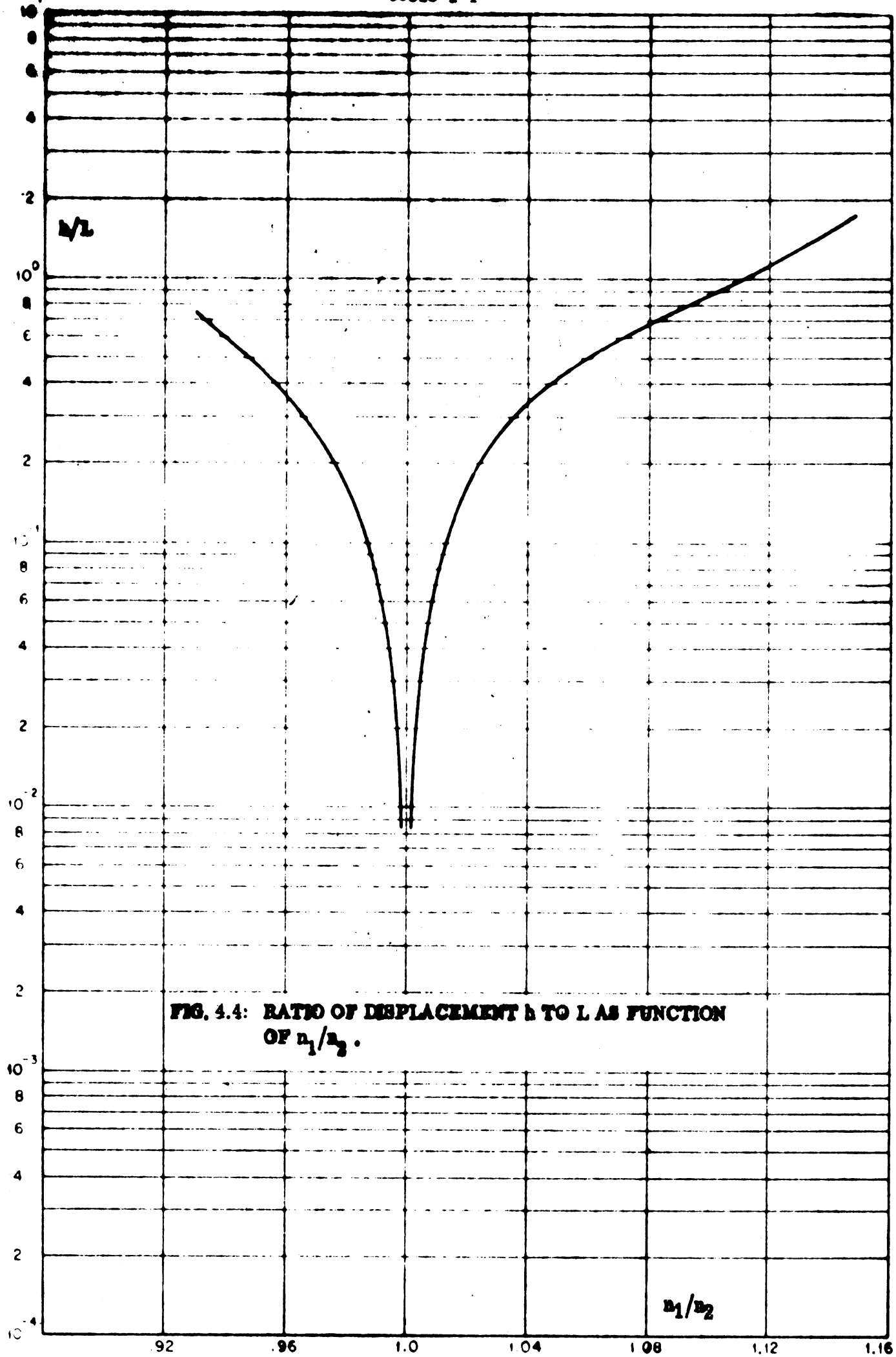

FIG. 4.4: RATIO OF DISPLACEMENT h TO L AS FUNCTION OF n_1/n_2.

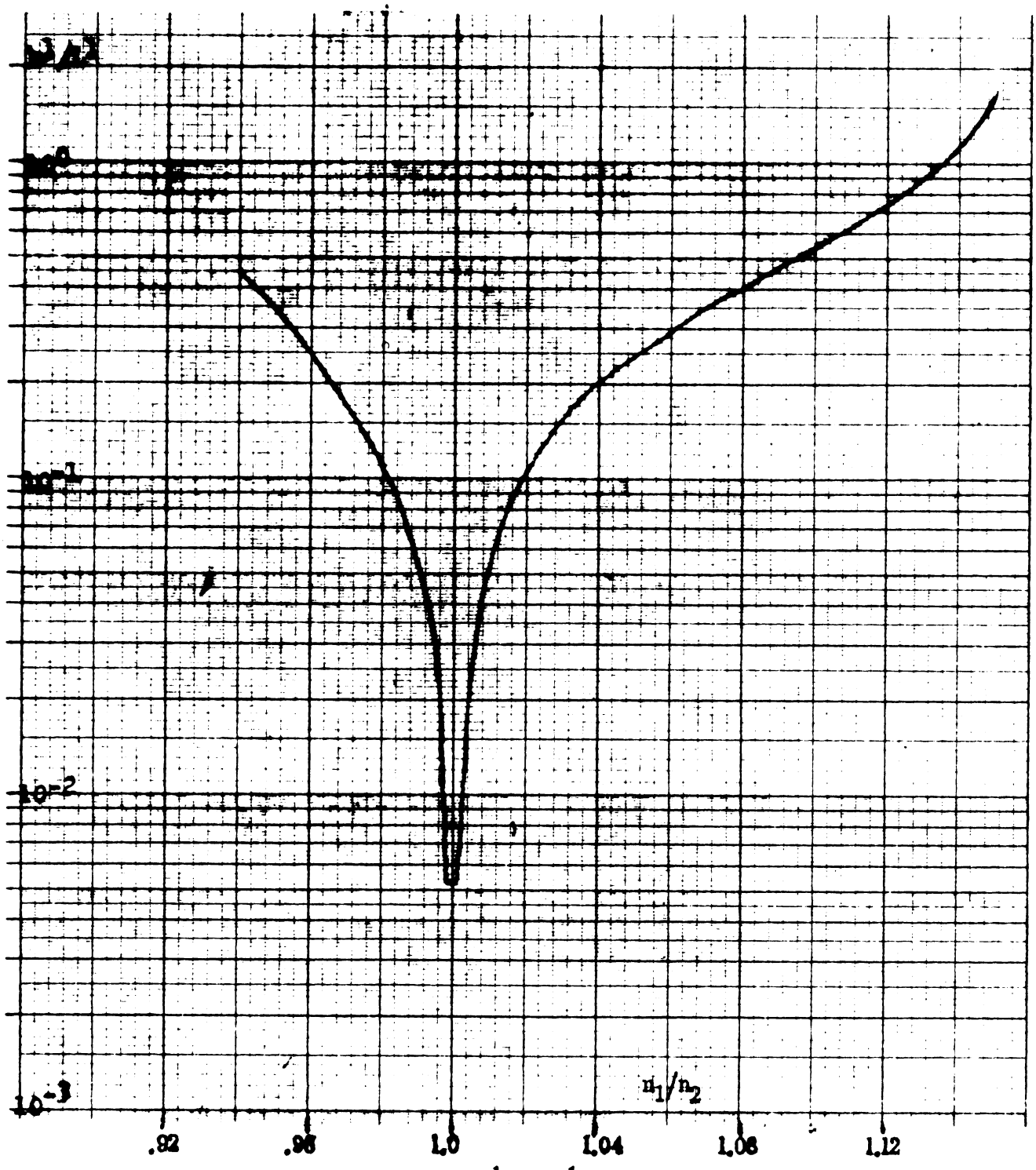

FIG. 4.5: RATIO OF DISPLACEMENT h^1 TO L^1 AS FUNCTION OF n_1/n_2

$\frac{h_o}{L}$ and $\frac{h_o^1}{L^1}$ as $\frac{n_1}{n_2}$ varies. By utilizing the results of these graphs we can find the power transmission coefficient of the filter as a function of $\frac{n_1}{n_2}$. We are primarily interested however in the filter transmission as the frequency is varied, so that we must assume some model for the frequency dependence of n_1 and n_2 on f. For simplicity it has been assumed that

$$n_{1,2}(f) = n_c \left[1 + \Delta \tan \alpha_{1,2}\right] \tag{4.20}$$

where n_c is the common refractive index of both parts of the filter at the center frequency of the passband, and Δ is the fractional change in the frequency about the center frequency, i.e.,

$$\frac{f - f_c}{f_c} = \Delta \quad . \tag{4.21}$$

Since

$$\frac{f_c}{n_c} \frac{dn}{df} = \tan \alpha$$

it is apparent that $\tan \alpha$ is a measure of the sensitivity of n to changes in f.

The bandwidth of the filter is now defined to be the frequency separation of the half-power points with respect to the maximum transmission and the normalized bandwidth N_B is taken to be the bandwidth divided by the center frequency. We can write the normalized bandwidth N_B as

$$N_B = \frac{f_U - f_L}{f_c} = \Delta_U - \Delta_L = \frac{\left(\frac{n_1}{n_2}\right)_U - 1}{\tan\alpha_1 - \left(\frac{n_1}{n_2}\right)_U \tan\alpha_2} - \frac{\left(\frac{n_1}{n_2}\right)_L - 1}{\tan\alpha_1 - \left(\frac{n_1}{n_2}\right)_L \tan\alpha_2} \tag{4.22}$$

where f_U and f_L are the upper and lower frequencies of the pass bands, and $(n_1/n_2)_U$ and $(n_1/n_2)_L$ are the ratios of (n_1/n_2) when the transmitted power

has fallen to one-half at the frequencies f_U and f_L. It should be noted that $-1 < \Delta_L < 0$ and $0 < \Delta_U < \infty$ so that $0 < N_B < \infty$. This presents an unrealistic picture when $N_B > 1$ since the concept of a normalized bandwidth then becomes somewhat meaningless. Actually, a normalized bandwidth greater than 1 is of no interest here, and we will devote our attention primarily to values of $N_B < 1$.

Since n_1/n_2 is symmetric about 1.0 over the range of values used in the calculations below, then (3.333) can be simplified to

$$N_B = \frac{2\delta}{\tan\alpha_1} = \frac{2}{\tan\alpha_1}\left[(n_1/n_2)_U - 1\right] \tag{4.23}$$

where $\tan\alpha_2$ has been set zero for simplicity, and

$$(n_1/n_2)_U = 1.0 + \delta; \qquad (n_1/n_2)_L = 1.0 - \delta \quad . \tag{4.24}$$

The quantity $(n_1/n_2)_U$ can now be obtained from (3.330), noting that τ is reduced to one-half its maximum value when

$$\frac{h+h^1}{2H} = \frac{1}{2} \quad . \tag{4.25}$$

An approximate value for $(n_1/n_2)_U$ can then be calculated for the range $.995 \leq n_1/n_2 \leq 1.005$ by noting that the expressions

$$\frac{h^1}{L^1} = \pm b^1\left(\frac{n_1}{n_2} - 1\right); \qquad \frac{h}{L} = \pm b\left(\frac{n_1}{n_2} - 1\right), \tag{4.26}$$

are approximately true, where b and b^1 are the slopes of the h/L and h^1/L^1 curves, the + sign applying for $n_1/n_2 > 1.0$, and the - sign for $n_1/n_2 < 1.0$. As a result of (3.334) and (3.337) we obtain then

$$Lb\left[(n_1/n_2)_U - 1\right] + L^1b^1\left[(n_1/n_2)_U - 1\right] = H \quad . \tag{4.27}$$

Thus,

$$(n_1/n_2)_U - 1 = \frac{H}{Lb + L^1b^1} \tag{4.28}$$

and

$$N_B = \frac{2}{\tan\alpha_1\left[\frac{Lb}{H} + \frac{L^1 b^1}{H}\right]} \tag{4.29}$$

This expression is valid when (3.337) holds, or when $N_B \tan\alpha < .01$. In that case, since $b \approx 8$ and $b^1 \approx 5$, $8\frac{L}{H} + 5\frac{L^1}{H} > 200$ is required. It may be seen from (3.340) that N_B decreases inversely proportional to L/H and L^1/H so that the ability to achieve sharper filter characteristics by increasing these parameters is limited by the largest practical L and L^1 and smallest practical H which could be used.

The normalized bandwidth for this filter is shown in Figs. 4-6 to 4-8 as a function of L/H, L^1/H and $\tan\alpha_1$, obtained without approximation from Figs 4-4 and 4-5 and expression (3.334). It should be noted that values of L/H less than $\tan\theta_1$ are not meaningful since then, as can be determined by Fig. 4-2, the entrance and exit slits would extend below the apex of the center prism and light could pass through the two slits without having encountered it. Since θ_1 is 60° for these curves, then $L/H > 1.73$.

Recalling that the intended application for this filter would be in a laser communication system with a frequency range from 1.5×10^{13} to 7.5×10^{14} cps and a bandwidth of 10^9 cps, it is apparent that the Wernicke filter is unsuited for this application, since the required value of N_B is 6.6×10^{-5} to 1.33×10^{-6}. This is 3 to 4 orders of magnitude less than what can be achieved with a reasonable value for $\tan\alpha$ and L/H. Actually, almost any value of N_B that was desired could be obtained by choosing $\tan\alpha$ large enough, but some data on materials used in this frequency range given by Kruse[2] indicates that the maximum value for $\tan\alpha$ may be on the order of 0.1. This value leads to normalized bandwidths on the order of 10^{-2}. Usually, materials that are more dispersive are also more lossy so that the insertion loss of the filter could be intolerably large. It should be recalled that Fig. 4-3 shows the maximum transmissivity of the filter considered in these

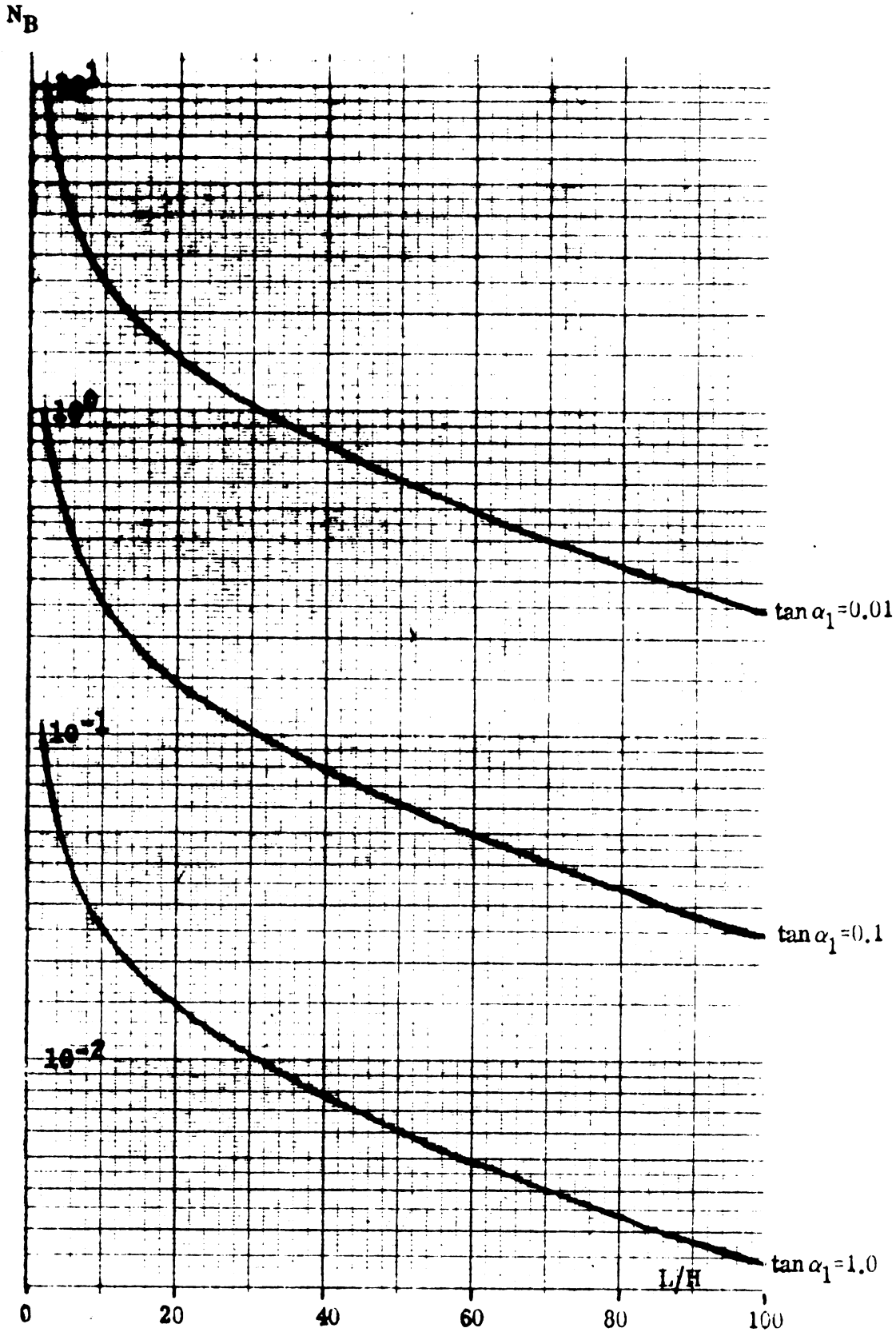

FIG. 4.6: NORMALIZED BANDWIDTH N_B FOR L^1 AND $\tan\alpha_2 = 0$ AS A FUNCTION OF L/H AND $\tan\alpha_1$

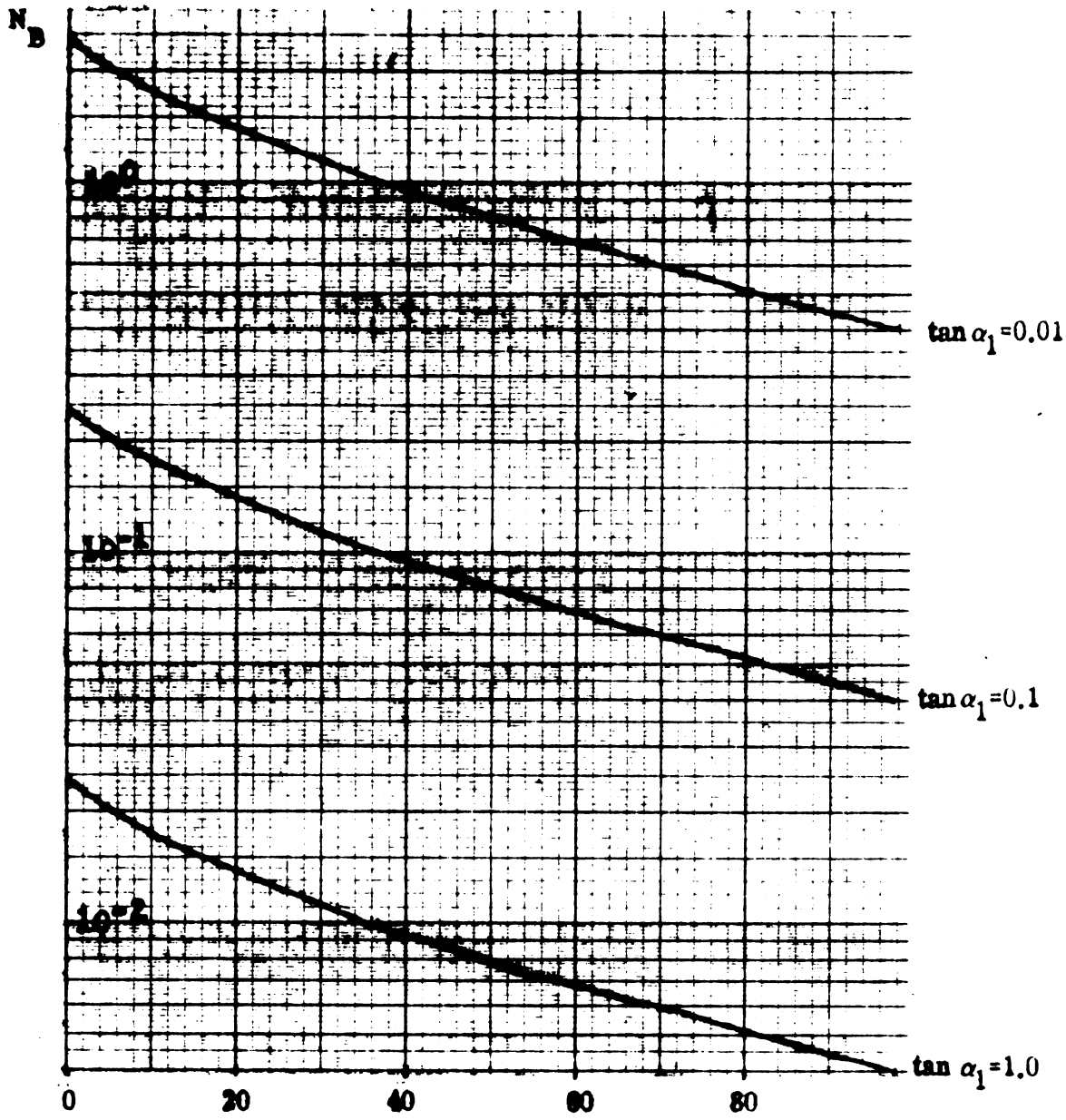

FIG. 4.7: NORMALIZED BANDWIDTH N_B FOR $\tan\alpha_2=0$ AND L/H=10 AS A FUNCTION OF L^1/H AND $\tan\alpha_1$.

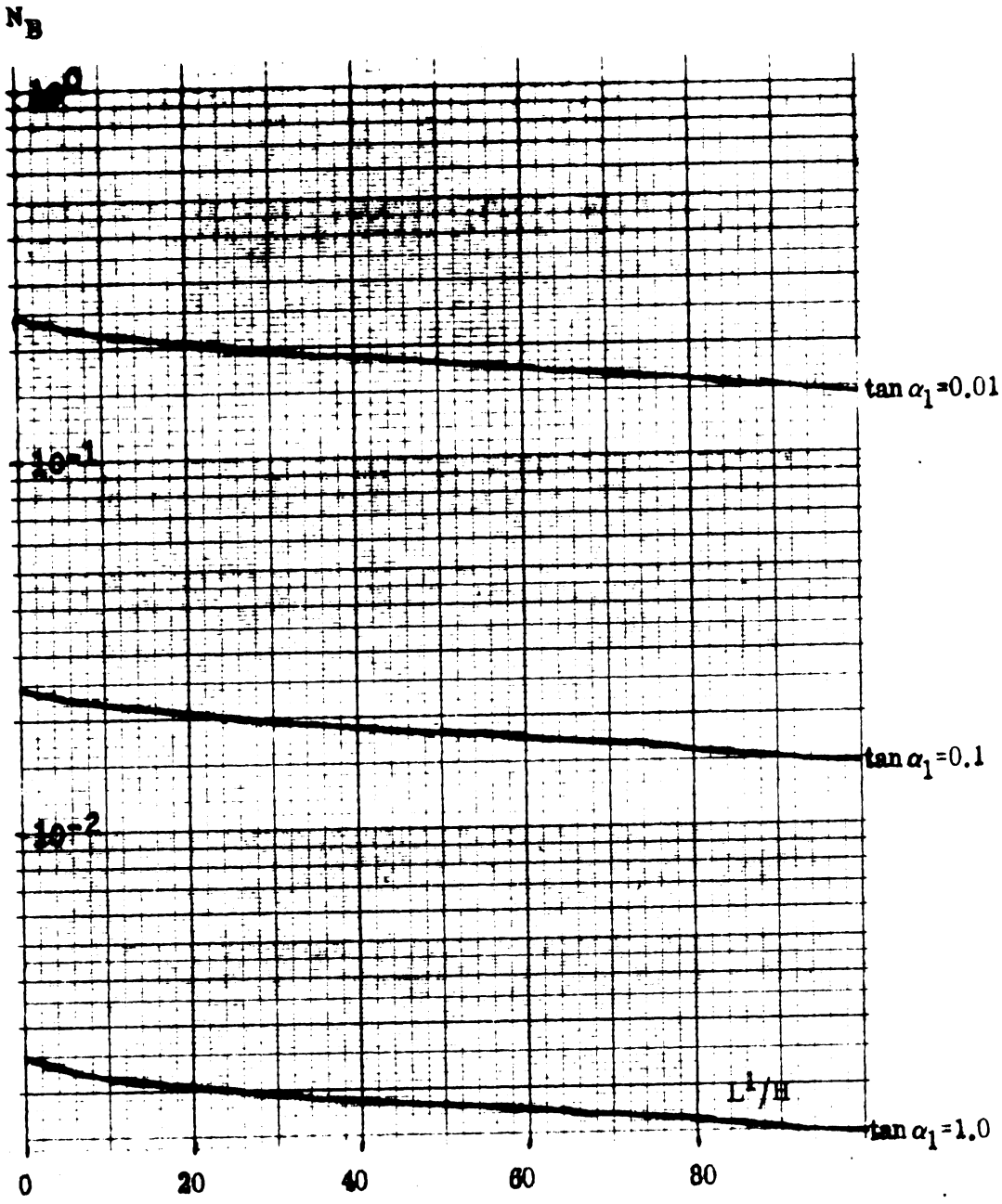

FIG. 4-8: NORMALIZED BANDWIDTH N_B FOR $\tan\alpha_2$=0 and L/H=100 AS A FUNCTION OF L^1/H AND $\tan\alpha_1$

calculations about 90 percent, but absorption has been neglected.

The normalized bandwidth could be decreased by increasing the angle θ_1. In the idealized treatment made here, this would indicate an improved filter performance. However, because the normalized bandwidth when θ_1 is 60° has been found to be 3 to 4 orders of magnitude larger than what is required, and losses in a real filter would further decrease its effectiveness, there seems to be no advantage to calculating the filter performance for other values of θ_1.

Tuning of this filter could be achieved in principle if one of the materials of which the filter is constructed were electro-optic in nature. However, the electro-optic coefficient for potassium dihydrogen phosphate (KDP), a material with a relatively large electro-optic coefficient of 8.47×10^{-12} meter/volt (Peters, 1963) leads to a change in the refractive index of only 25×10^{-12} E, where E is the electric field in volts/meter. Thus, even for very large field strengths, only a very small refractive index change takes place, with negligible change in the center frequency of the filter.

Actually, the requirement for tunability of the filter would then be that the refractive indices of the two materials comprising the filter have very nearly the same frequency dependence, so that the crossover point could be changed greatly by a small change in either refractive index. On the other hand, in order to achieve a narrow band pass the frequency characteristics of the refractive indices of the two materials should be very different. These two requirements are thus not compatible.

4.4 The Polarization Interference Filter

The polarization interference filter, or as it is also known, the Lyot-Öhman filter and birefringent filter is a particularly attractive one from the standpoint of the narrowness of the obtainable bandwidth and the possibility of tuning it. This filter was designed by Lyot (1933) and first built by Öhman (1938). In this report, the various features of this filter are discussed and its applicability in a laser communication channel is considered.

4.5 Bandwidth and Transmission Characteristics of the Polarization Interference Filter

The polarization interference filter may take different forms. Evans (1949) considers several possibilities. The discussion here is limited to the most basic of these forms which consists of sections of birefringent plates cut with their optic axes parallel to the large faces of the plates and placed between polarizers oriented at 45^{o} to the optic axes. A diagram of the filter is shown in Fig. 4.9.

The mechanism on which the filter operation depends is the difference in propagation velocity for ordinary and extraordinary waves in the birefringent material. When a linearly polarized wave is incident at 45^{o} to the optic axis of the birefringent material, the beam is split into two waves, the ordinary and the extraordinary wave. The beams emerge from the birefringent material polarized at right angles and with a phase difference depending on the difference in velocity of propagation. The amplitude of the wave transmitted by a second polarizer depends on the relative retardation of the two beams. If the difference in retardation is an integral number of half waves, the interference between the waves is destructive, and no transmission occurs. For the single birefringent plate of thickness d placed between parallel polarizers, where absorption within the plate and reflection losses are neglected, the transmission is given by Billings, et al (1951) as

$$T_1 = \cos^2\left[\frac{\pi d\mu}{\lambda}\right] \tag{4.30}$$

where $\mu = n_e - n_o$ and n_e and n_o are the extraordinary and ordinary indices of refraction at wavelength λ . When light light is incident on such a plate and polarizer arrangement and the output is observed by a spectrograph, a series of bright and dark bands, which is called a channel spectrum, is obtained. This results from the argument of the cosine above taking on integral values representing the higher order interference fringes.

If several sections of these plates are put together as shown in Fig. 4.9, and

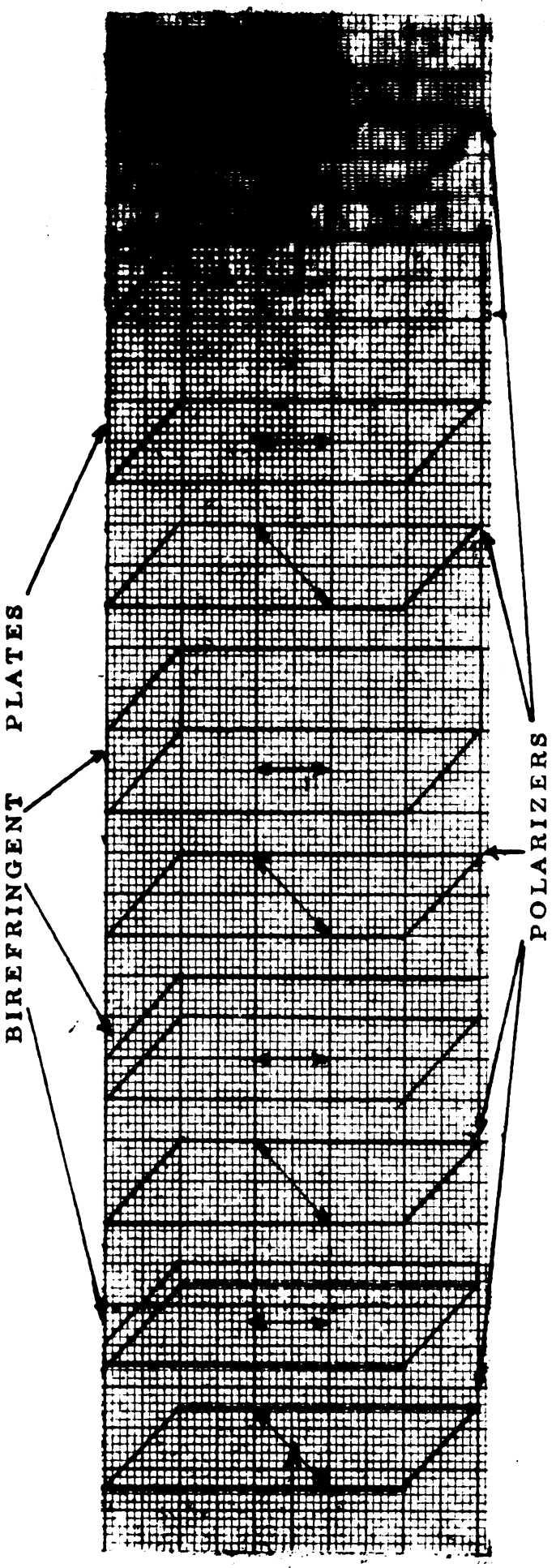

FIG. 4.9: POLARIZATION INTERFERENCE FILTER.
Arrows show orientation of optic axes of birefringent plates and polarizers.

the thickness of birefringent plates are adjusted in the ratios 1, 2, 4 etc., the overall transmission characteristic of such an arrangement exhibits pass bands whose widths are equal to the fringe widths of the thickest plate separated in wave length by the distance between fringe widths of the thickest plate separated in wave length by the distance between fringes of the thinnest plate. If the sections are numbered i = 1, . . ., N, the ith section will have a thickness

$$d_i = 2^{i-1} d_1$$

$$d_N = 2^{N-1} d_1$$

where d_1 is the thickness of section 1. The transmission of the whole system can then be given, when the absorption is ignored, by

$$T = \left[\cos\theta \cos 2\theta \cos 4\theta \ldots \cos 2^{N-1}\theta\right]^2 \qquad (4.31)$$

where

$$\theta = \frac{\pi d \mu}{\lambda}$$

is 1/2 the angular retardation of Section 1. The retardation is also spoken of in waves, and is given by

$$\Gamma_\lambda = \frac{d}{\lambda^\mu} \ .$$

Thus, in order for the whole array to have a common transmission peak, the retardation of each section in waves must be an integer such that with

$$\Gamma_i = n_i; \quad \Gamma_1 = n_1 \ ,$$

$$n_i = 2^{i-1} n_1 \ .$$

There are also side band peaks centered about the main peak and separated from it by the fringe width of the thickest section.

Billings (1947) showed that (3.342) can be written as

$$T = \left[\frac{\sin^{2N}\theta}{2^N \sin\theta} \right]^2 . \qquad (4.32)$$

Figure 4.10 shows a plot of this function. It may be observed that this curve resembles that of the familiar $\left(\frac{\sin x}{x}\right)^2$ function encountered in diffraction theory. When the amount of light energy transmitted by the filter between the pass bands is calculated, it is found to be 0.11 of that contained in the pass band (Evans, 1949). Thus, if an auxiliary wide-band filter were to be used to remove the unwanted transmission bands on either side of the desired one, the background light transmitted by the polarization filter would be essentially confined to the main pass band.

The band width B of the filter is, as mentioned above, determined by the thickest plate in the filter, and is (Billings, 1951) at the half power points given by

$$B_\lambda = \frac{0.5\,\lambda_o^2}{2^{N-1} d_1 \mu} = \frac{0.5\lambda_o^2}{d_N \mu} \qquad (4.33)$$

in terms of the wavelength. λ_o is the position of the pass band. The normalized bandwidth N_B is defined to be the bandwidth to center wavelength ratio, or

$$N_B = \frac{0.5\,\lambda_o}{2^{N-1} d_1 \mu} = \frac{0.5\lambda_o}{d_N \mu} . \qquad (4.34)$$

Thus for a given normalized bandwidth, the filter dimension will be proportional to the wavelength. If, instead, it is desired that the bandwidth be given in terms of the frequency, then with c the velocity of light in free space

$$B_f = \frac{0.5\,c}{d_N \mu} , \qquad (4.35)$$

while the normalized bandwidth from the frequency approach produces the same result as (4.34). Note that the frequency bandwidth for the polarization filter of a given length is independent of frequency, which is desirable in the laser communications channel.

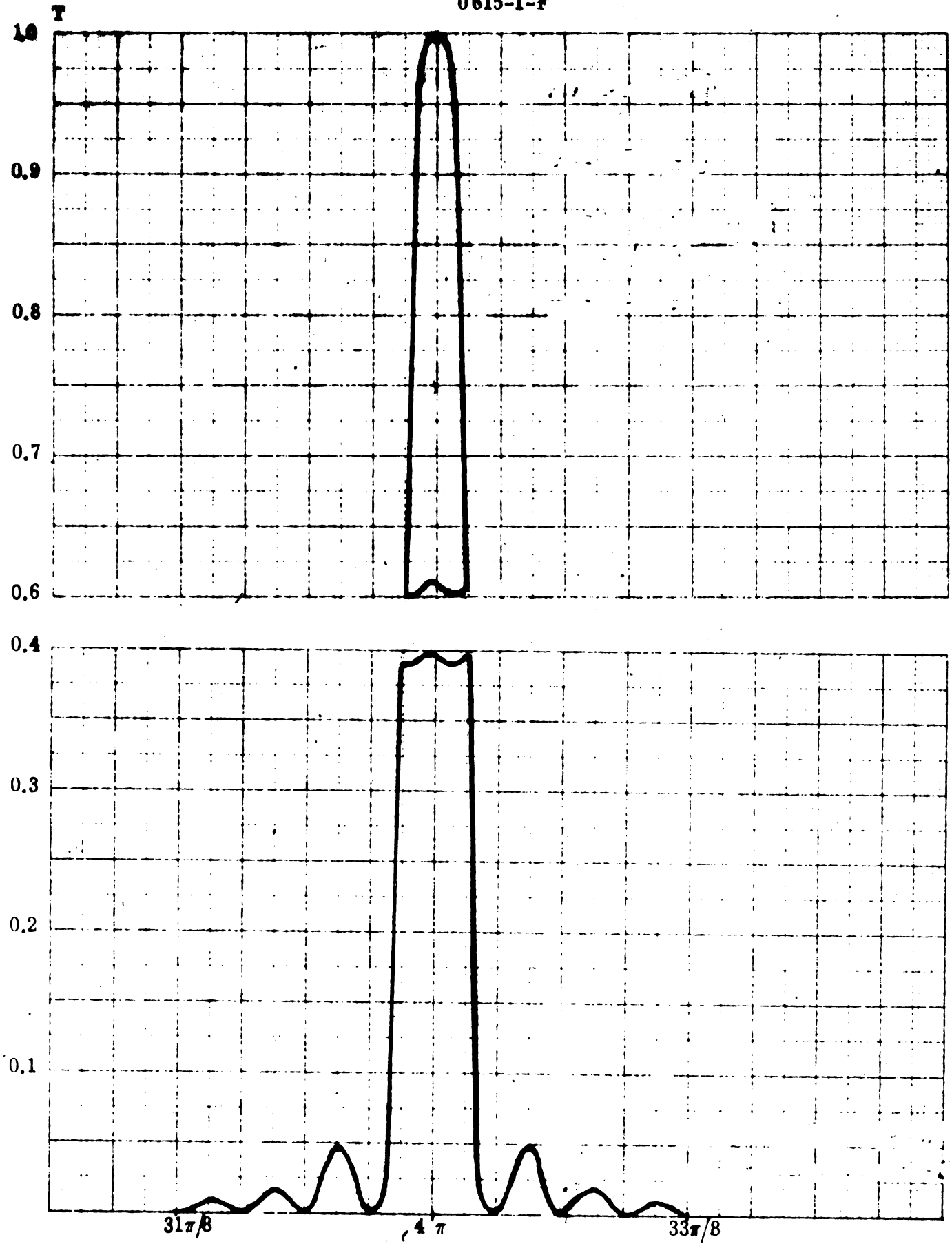

FIG. 4.10: THE FUNCTION $T=(\sin 2^N\theta/2^N\sin\theta)^2$ FOR N=5. ANGULAR RETARDATION OF THINNEST PLATE $=2\theta$.

Quartz has been one of the birefringent materials used in commercial polarization filters. The quantity μ has a value of 0.011 for quartz so that for a bandwidth of 10^9 cps, d_N must be 13.6 meters. This is clearly too great a thickness for a practical filter, quartz plates would limit a filter to a bandwidth of an order of magnitude or two wider than 10^9 cps. Calcite, another material that has been used for the polarization filter, has a μ value of 0.172, so d_N in this case must be 87.2 cm which is a more realistic thickness but still impractically large. A list of values of refractive indices for some birefringent materials is shown in Table IV-1. It can be observed that the μ value for calcite is exceeded by that of only two other materials, but the difference is not enough to make the filter dimension less than 40 cm for a bandwidth of 10^9 cps. It seems then, in light of presently available materials, that the bandwidth requirement of 10^9 cps cannot be achieved in a practical polarization filter.

The absorption within the birefringent material of which the filter is constructed has not been considered above. The attenuation of a wave propagating through a lossy medium can be accounted for by an attenuation constant α such that

$$I(x) = I_o e^{-\alpha x} \tag{4.36}$$

where α can be a function of frequency. $I(x)$ is the intensity at a point x within the medium measured in the direction of propagation from the reference point where the intensity is I_o. In the filter whose Nth element has a thickness $d_N = 2^{N-1} d_1$ where d_1 is the thickness of the first element, the total thickness is

$$d_T = \sum_{i=1}^{N} d_i = (2^N - 1) d_1 .$$

Thus, the output from the last plate of the filter is decreased from that given in (3.342) by the additional factor

TABLE IV-1. VALUES OF REFRACTIVE INDICES (Billings, 1951)

Crystal	Class	Indices
ADP	Tetragonal	$\epsilon = 1.4792$ $\omega = 1.5246$
KDP	Tetragonal	$\epsilon = 1.4684$ $\omega = 1.5095$
Rochelle Salt	Orthorhombic	$\alpha = 1.492$ $\beta = 1.493$ $\gamma = 1.496$
EDT	Monoclinic	$\alpha = 1.5086$ $\beta = 1.5893$ $\gamma = 1.5930$
Quartz	Hexagonal	$\epsilon = 1.553$ $\omega = 1.544$
Calcite	Hexagonal	$\epsilon = 1.486$ $\omega = 1.658$
$NaNO_3$	Hexagonal	$\epsilon = 1.3361$ $\omega = 1.5874$
Rutile	Tetragonal	$\epsilon = 2.903$ $\omega = 2.616$
Mica	Monoclinic	$\alpha = 1.503$ to 1.623 $\beta = 1.545$ to 1.685 $\gamma = 1.545$ to 1.704

ϵ extraordinary refractive index of uniaxial crystal.
ω ordinary refractive index of uniaxial crystal.
α, β, γ smallest, intermediate and largest refractive indices of biaxial crystal.

$$T_\alpha = e^{-\alpha(2^N-1)d_1} \approx e^{-\alpha 2^N d_1}, \tag{4.37}$$

since $2^N >> 1$ for $N \geq 5$. From the expression (3.346), we can obtain then

$$T_\alpha = e^{-[\alpha c/\mu B_f]} \tag{4.38}$$

The transmission is seen to decrease exponentially with decreasing bandwidth, so a further narrowing of the bandwidth is eventually achieved, depending upon the magnitude of α, only through rapid reduction of a peak transmission of the filter.

Another mechanism which decreases the peak filter transmission is the reflection occurring at the various interfaces. This can be reduced in practice by use of refractive index matching cements to join various components of the filter, or by immersing the filter in an oil whose refractive index is near that of the filter components. Reflection losses will generally be less important than absorption losses and will not be dealt with here.

As a practical example, polarization interference filters have been made commercially by Baird Associates (Greenler, 1958) using quartz plates having a bandwidth of 1 Å at 6563 Å and with a peak transmission on the order of 10 percent. Steel et al (1961) describe a filter using quartz and calcite plates with a bandwidth of 1/8 Å at the same wavelength, having a peak transmission of 12 percent.

The aperture and field of view of the filter are also characteristics which are of interest. The field of view is determined by the maximum off-axis effects that can be tolerated in the filter. Evans (1949) considers this problem in detail and so it will not be discussed here.

4.6 Tuning of the Polarization Filter

The utility of the polarization interference filter can be greatly enhanced by devising a means for tuning it. A slight amount of tuning can be obtained by varying the temperature of the filter, since the refractive index and thus the retardation

of the birefringent plate is a function of the temperature. The control which can be obtained in this manner is, however, limited to a few angstroms and may actually become a nuisance in a narrow band filter since then it must be very accurately temperature-regulated.

A second method suggested by Lyot (1944) for tuning the filter is the obvious one of changing the plate thickness. He proposed the scheme of making the individual plates in the form of two wedges which could be slid across each other, thus varying the length of the light path through the wedge combination. This results in the requirement for very fine adjustments to be made in synchronism to each plate of the filter and does not represent a truly variable filter.

A third and most practical method for tuning of the polarization filter is the addition of a variable achromatic phase plate or retardation plate to each fixed element so that one can vary the retardation of each section of the filter. The achromatic phase shifter, which is, by definition capable of producing a phase shift independent of the wavelength, could give results exactly comparable to those obtained by the use of variable thickness plates mentioned above. The retardation plate on the other hand, one example of which that is adjustable being the Kerr cell, produces a retardation dependent on the wavelength and so does not give results as good as those obtainable with an achromatic phase shifter. Billings (1947) and Evans (1949) consider this problem. The discussion below follows Billings' treatment.

The tuning is accomplished by varying the retardation of each section so that its nearest transmission peak is shifted to the place where the overall filter transmission is to be a maximum. If, at the desired wavelength λ_x the retardation of the Nth plate is

$$\Gamma_{Nx} = x + 2^{N-1} n_1 = 2^{N-1} \Gamma_1 ,$$

the argument of the cosine is given by

$$\theta_{Nx} = \pi\left(x + 2^{N-1} n_1\right) = 2^{N-1}\,\theta_{1x},$$

where n_1 is the number of waves retardation for the first plate's nearest peak. The nearest peak of the Nth plate has as the argument of its cosine

$$\theta_{Nn} = \left(n_N + 2^{N-1} n_1\right)\pi$$

where n_N is an integer so chosen as to minimize $|n_N - x|$. It is convenient to number the peaks of each plate starting with zero at the natural filter transmission peak.

The retardation plate is now used to shift θ_{Nx} to the value θ_{Nn} without changing the wavelengths λ_x by adding an additional retardation δ_{Nx} such that

$$2\theta_{Nx} + \delta_{Nx} = 2\theta_{Nn} \quad . \tag{4.39}$$

Thus the transmission peak is shifted to the wavelength λ_x. We can mention here that δ_{Nx} will vary with wavelength for the retardation plate, but for the achromatic phase plate it does not. The relation above can be put into another form, the utility of which will be apparent below. Thus, equivalent to (3.350)

$$\theta_{Nn} = a_N \theta_{Nx} \tag{4.40}$$

so that

$$a_N = \frac{2^{N-1} n_1 + n_N}{2^{N-1} n_1 + x} \quad . \tag{4.41}$$

When the Kth plate is considered, then,

$$a_K = \frac{2^{N-1} n_1 + 2^{N-K} n_K}{2^{N-1} n_1 + x} \quad . \tag{4.42}$$

The relation between θ_{Kx} and θ_{1x} is determined, it should be noted, only by the

fact that $d_K = 2^{K-1} d_1$, so that $\theta_{Kx} = 2^{K-1} \theta_{1x}$. But now, the retardation of each filter section is given by $a_K \theta_{Kx}$ so that the 1:2:4: etc., relation between the sections of the filter is no longer true. Thus,

$$T_x = \left[\cos a_1 \theta_{1x} \cos a_2 2\theta_{1x} \cos a_3 2^2 \theta_{1x} \, , \, , \, . \cos a_N a^{N-1} \theta_{1x}\right]^2 \tag{4.43}$$

so that while the peak of the filter has been shifted to wavelength λ_x, the sideband peaks may occur with different spacing and amplitudes than for the unshifted peak, where T is given by (4.31). Billings (1947) points out, however, that the various a_i values may be the same for several of the filter sections and with an n_1 value of two or more, there is no significant distortion in the filter characteristics or increase in residual light transmission even when the filter is tuned far from a natural peak. It is obvious that the larger the variation of the a_i's from 1, the greater will be the effect on the filter characteristic. At its largest deviation from 1, the value of a_K is

$$a_K = 1 \pm \frac{1}{2^{K} n_1 + 1} \, . \tag{4.44}$$

Thus the variation is relatively small for the thicker sections, and this maximum deviation occurs when the maximum retardation, 1/2 wave, is required to be added to or subtracted from the Kth section.

The phase shift δ_{Kx} may be expressed as

$$\begin{aligned} \delta_{Kx} &= 2\left[\theta_{Kn} - \theta_{Kx}\right] \\ &= 2^K \theta_{1x}\left[a_K - 1\right] \\ &= 2\theta_{Kx}\left[a_K - 1\right], \end{aligned}$$

where, it should be recalled, θ_{1x} is 1/2 the retardation angle of section 1 at λ_x.

If a_K is substituted for from (3.353) and the formula

$$x = 2^{N-1} n_1 \frac{(\lambda_o - \lambda_x)}{\lambda_x}$$

is used, where λ_o is a reference wavelength, then

$$\delta_{Kx} = \frac{\lambda_x n_K + 2^{K-1} n_1 - 2^{K-1} n_1 \lambda_o}{n_1 \lambda_o} 2\pi \qquad (4.45)$$

λ_o is the wavelength at which the Kth section has a retardation angle of $2\theta_{Kn}$, and λ_x is the wavelength at which the Kth section has a retardation angle of $2\theta_{Kx}$. Billings (1947) has plotted the retardation in waves which must be added to the plates of a three element filter in order to shift its pass band continuously from $0.8\lambda_c$ to $1.33\lambda_c$, where the first plate has a retardation of 4π at λ_c.

When the achromatic phase shifter is considered, then the retardation which is added is independent of the wavelength. As a result, the transmission peak of each section is merely shifted along the wavelength scale to the desired point of an overall peak without distorting the transmission characteristics of the filter even when tuned far from a natural peak. As Evans (1949) points out, however, tuning by the method of shifting the nearest peak of each filter section to the desired transmission peak is a practical possibility whether the phase shifter is achromatic or not, since the results obtained with the retardation plate and achromatic phase shifter are comparable.

4.7 Comments and Conclusions

Based on the results of the preceding analysis and calculations, the prismatic or Wernicke type filter appears to be unsuited for application in a laser communication channel due to the limitation on the narrowness of the passband. In addition, the tuning of such a filter is impractical and is incompatible with the requirement of a narrow passband.

It has been shown above that the polarization interference filter is capable

of producing arbitrarily narrow passbands, at least theoretically, when the absorption is ignored. However, when considered from a practical viewpoint, the thickness of material required to produce a passband of 10^9 cps is prohibitive when presently available materials are considered. Tuning of the polarization interference filter is theoretically possible over a 2:1 band using electro-optic techniques.

V

SUMMARY OF CONCLUSIONS

5.1 Introduction

This chapter attempts to summarize the conclusions reached on the basis of the material gathered and analyzed as well as the theoretical investigations presented in this report. All through the work the objectives have been to explore the theoretical limits in each area rather than present boundaries defined by the state of the art. We hope that particularly in the second problem area we have been able to make organized contributions towards a better understanding as well as a better quantitative grasp of the problems involved in optical communication. In the first and the third problem areas the emphasis has been on ariving at conclusions from a search of published theoretical investigations and recorded data.

5.2 First Problem Area

An optical signal from outer space may suffer deterioration before reaching the ground due to

a) extinction (absorption and scattering)

b) "seeing"

c) scintillation

d) "noise" radiation

e) dispersion

f) loss of coherence

In order to maintain communication despite these difficulties a number of countermeasures are available:

a) boosting channel parameters (power, antenna gain)

b) choice of site (high altitude, stable weather)

c) diversity reception

d) alternate terminals (against local and temporary conditions)

e) choice of wavelengths ("windows")

f) choice of code and modulation frequencies

g) selectivity of receiver.

The conclusions are that several of the difficulties are easily minimized; molecular absorption and anomalous dispersion by choice of operating wavelength; meteorological conditions and associated extinction and seeing by choice of site. The literature shows a very wide range of data on extinction. Even for a specific known site it would be difficult to give a firm value of minimum transmittivity. Near opacity over a wide wavelength range is found only in dense fog and under a thick cloud cover.

Background radiation is minimized by receiver selectivity and antenna gain. During daylight scattered sunlight is a severe limiting factor.

We have tentatively concluded that loss of coherence due to attenuation in the atmosphere is likely to be small, but a more careful study of this problem is suggested.

Scintillation is another phenomenon which we have not had opportunity to devote much attention to. Tentative conclusions are that it introduces a modulation of low frequency which may be discriminated against by choices of modulation or code.

5.3 Second Problem Area

At very low power levels, observations of electromagnetic radiation in the optical range become for fundamental physical reasons digital in nature. It should not be surprising then that digital communication channels are found more efficient than analog-type channels and detection methods under such conditions.

Entropy calculations for a binary optical channel using a photon counter as a detector indicated an unexpectedly high "efficiency" (Gordon, 1962) for average signal levels of less than one photon per sample pair, if we with efficiency understand the fraction of the theoretical capacity that can be utilized with an ideal binary code. The efficiency approaches 100% as the signal level approaches zero.

Our conclusion is that such high efficiency figures are of the same nature as the basic theorems in information theory: the length of the code and complexity of the terminal equipment do not show resonable convergence close to this limit. Numerical calculations indicate definitely, however, that there is a range of signal levels below one photon per sample pair where the error probabilities are reasonable, so that a practical solution is possible.

As far as the photon counter is concerned, it recovers ideally as much of the information as quantum mechanics permits; no detector can do better. A photomultiplier is not an ideal photon counter, but such tubes have been made which have an extremely good performance.

The comparison of the photon counter with a superheterodyne detector we find to be based on a somewhat doubtful extrapolation of the latter's performance down to the signal levels in question, but there is little doubt that the photon counter method of detection is superior in this range.

Considerable attention has in this report been devoted to a quantum amplifier as a receiver component. It appears intuitively plausible that an ideal quantum amplifier followed by an ideal photon counter should also be able to extract as much imformation from the signal as quantum mechanics permits. The spontaneous emission noise in the amplifier is in a sense analogous to the dark current in the photocell and the electron multiplier; most of the output pulses due to the latter are smaller than the pulses generated by photons and can be discarded. Similarly, in a single-transit traveling-wave laser amplifier the pulses generated by spontaneous emission travel less than the full length of the active material; thus they are amplified less and can be eliminated by a threshold device.

It is interesting to speculate whether or not similar amplitude discrimination can be made to eliminate the dark-current pulses in a tiny semiconductor junction photocell and whether or not a digital amplifier or multiplier for such a photocell is feasible. If the answer is affirmative, the photon counter operation

of a communication channel could be extended to wavelengths above one micron, for which effective photoemissive devices are not available and are unlikely to become available in the future.

5.4 Third Problem Area

The photon-counter detector discussed above needs a narrow filter to prevent counts of background radiation outside the bandwidth of the signal. The investigation reported here concludes that there is no theoretical difficulty in devising such filters, but the state of the art, as far as materials with desired properties and other technological developments are concerned, does not at the present permit the design of filters that meet simultaneously reasonable requirements of length, weight, transmissivity, tunability etc.

The combination of a laser amplifier and photon counter discussed in the previous section would have the advantage of selectivity as well as gain. The question of whether or not sufficient bandwidth can be obtained for the signals specified in the directives for this contract remains to be answered.

BIBLIOGRAPHY

Akcasu, A. Z., (1963) "A Study of Line Shape with Heitler's Damping Theory," The University of Michigan Technical Report 04836-1-T, (April).

Altshuler, T. L., (1961) "Infrared Transmission and Background Radiation by Clear Atmospheres," General Electric Missile and Space Vehicle Division Report No. 61SD199. UNCLASSIFIED

Astreimer, R. W., G. Falbel and S. Mickowitz, (1964) "Infrared Modulation by Means of Frustrated Total Internal Reflection," Presented at Spring meeting of Opt. Soc. of Amer., Washington, D. C.

Augason, G. C. and H. Spinrad, (1965) "Infrared Astronomy Review for the Astronomy Subcommittee of the National Aeronautics and Space Administration," NASA TMX-1074.

Baum, W. A., (1962) "The Detection and Measurements of Faint Astronomical Sources," in Astronomical Techniques, W. A. Hiltner (Ed.), (The University of Chicago Press).

Bell, E. E., L. Eisner, J. Young and R. Oetjen, (1960) "Infrared Radiance of Sky and Terrain at Wavelengths Between 1 and 20μ," J. Opt. Soc. Am., 50, No. 12, 1313.

Bennett, W. R. and P. J. Kindlmann, (1962) "Magnetostrictively Tuned Optical Maser," Rev. of Sci. Inst., 33, No. 6, 601.

Billings, B. H., (1947) "A Tunable Narrow-Band Optical Filter," J. Opt. Soc. Am., 37, 738.

Billings, B. H., S. Sage and W. Draisin, (1951) "A Narrow Passband Polarization Interference Filter for Hydrogen Alpha," Rev. Sci. Inst., 22, 1009.

Born, M. and E. Wolf, (1959) "Principles of Optics, (Pergamon Press) 348.

Brady, L. J., (1950) "Spectrophotometers, Infrared Region," Chapter 8 of Analytical Absorption Spectroscopy, M. G. Mellon (Ed.), (John Wiley and Sons Inc.).

Curcio, J. A., (1961) "Evaluation of Atmospheric Aerosol Particle Size Distribution from Scattering Measurements," J. Opt. Soc. Am., 51, No. 5, 548.

Curcio, J. A. and K. A. Durbin, (1959) "Atmospheric Transmission in the Visible Region, NRL Report 5368, U. S. Naval Research Laboratory, Washington, D. C.

Curcio, J.A., G.L. Knestrick and T.H. Cosden, (1961) "Atmospheric Scattering in the Visible and Infrared," NRL Report 5567, U.S. Naval Research Laboratory, Washington, D.C.

Deirmendjian, D., (1963) "Complete Microwave Scattering and Extinction Properties of Polydispersed Cloud and Rain Elements," The Rand Corporation Report R-422-PR. UNCLASSIFIED.

Deirmendjian, D., (1964) "Scattering and Polarization of Water Clouds and Hazes in the Visible and Infrared," Appl. Opt., 3, No. 2, 187.

Eberhardt, E.H., (1965) "Multiplier Phototubes for Single Electron Counting," Electrical Communication, 40, 124-133.

Elterman, L., (1963) "A Model of a Clear Standard Atmosphere for Attenuation in the Visible Region and Infrared Windows," AFCRL Optical Physics Laboratory, Report No. AFCRL-63-675.

Elterman, L., (1964) "Parameters for Attenuation in the Atmospheric Windows for Fifteen Wavelengths," Appl. Opts., 3, No. 6, 745.

Evans, J.W., (1949) "The Birefringent Filter," J. Opt. Soc. Am., 39, 229.

Fano, U., (1957) "Description of States in Quantum Mechanics by Density Matrix and Operator Techniques," Rev. Mod. Phys., 29, 74.

Fock, R.L. and L.C. Bradley, (1964) "Dispersion in the Vicinity of an Optical Resonance," Appl. Opts., 3, No. 1, 137.

Gaertner, H., (1957) "The Transmission of Infrared in a Cloudy Atmosphere," NAVORD Report No. 429, U.S. Government Printing Office, Washington, D.C.

Gates, D.M. and C.C. Shaw, (1960) "Infrared Transmissions of Clouds," J. Opt. Soc. Am., 50, No. 9, 876.

Gil'varg, A.B. and G.V. Kolesov, (1961) "Use of the Electro-optical Effect in Crystals for High Speed Shutters," Instr. and Exp. Tech. (Translated from Russian) No. 3, 535.

Glauber, R., (1963) "Quantum Theory of Optical Coherence," Phys. Rev., 130, 2529.

Glauber, R., (1963) "Coherent and Incoherent States of the Radiation Field," Phys. Rev., 131, 2766-2788.

Glauber, R., (1965) "Photon Counting and Field Correlations," Paper presented at the Conference on The Physics of Quantum Electronics, SanJuan, Puerto Rico.

Goldberg, L., (1954) "The Absorption Spectrum of the Atmosphere," Chapter 9 of The Solar System: Vol. II--The Earth As a Planet. G. P. Kuiper (Ed.), (The University of Chicago Press) 434-490.

Golomb, S. W., (1964) "Digital Communications with Space Applications," (Prentice Hall, Inc.).

Gordon, J. P., (1962) "Quantum Effects in Communication Systems," Proc. IRE, 50, 1898-1908.

Gordon, J. P., W. H. Louisell and L. R. Walker, (1963) "Quantum Fluctuations and Noise in Parametric Processes-II," Phys. Rev., 129, 481.

Gordon, J. P., W. H. Louisell and L. R. Walker, (1963) "Quantum Statistics of Masers and Attenuators," Phys. Rev., 130, 806.

Gordon, J. P., (1963) "Quantum Noise in Communication Channels" Proceedings of Quantum Electronics, Paris, Vol. I, 55-64.

Greenler, R. G., (1958) "Optical Filters," Concepts of Classical Optics-Appendix Q, (W. H. Freeman and Co., San Francisco).

Hadley, L. N. and D. M. Dennison, (1947) "Reflection and Transmission Interference Filters," J. Opt. Soc. Am., 37, No. 6, 451-465.

Harrison, G. R., R. C. Lord and J. R. Loofbourow, (1961) "Practical Spectroscopy," (Prentice-Hall Inc., New York).

Hauser, S. M., L. Smith, D. Marlowe and P. Yodes, (1963) "The Stressed Plate Shutter, a New Moderate Speed Electro-Optical Light Modulator," J. Opt. Soc. Am., 52, No. 11, 1175-1179.

Holshouser, D. F., H. VonForester and G. L. Clark, (1961) "Microwave Modulation of Light Using the Kerr Effect," J. Opt. Soc. Am., 51, No. 12, 1360.

Holter, M. R., S. Nudelman, G. Suits, W. Wolfe and G. Zissis, (1962) Fundamentals of Infrared Technology, (The MacMillan Co., New York).

Howard, J. N. and J. S. Garing, (1962) "Transmission of the Atmosphere in the Infrared-A Review," Air Force Surveys in Geophysics, No. 150, AFCRL.

Jaynes, E. T., (1957) "Information Theory and Statistical Mechanics-I," Phys. Rev. 106, 620-630.

Jaynes, E. T., (1957) "Information Theory and Statistical Mechanics-II," Phys. Rev., 108, 171-190.

Junge, C. W., (1963) "Aerosols," Air Chemistry and Radioactivity. (New York and London: Academic Press).

Kaminow, I. P., (1961) "Microwave Modulation of the Electro-Optic Effect in KH_2PO_4," Phys. Rev. Letters, 6, No. 10, 528.

Knestrick, G. L., T. H. Cosden and J. A. Curcio, (1961) "Atmospheric Attenuation Coefficients in the Visible and Infrared," NRL Report 5648, U. S. Naval Research Laboratory.

Kruse, P. W., L. D. McLauchlin and R. B. McQuistan, (1962) Elements of Infrared Technology, (John Wiley and Sons).

Kurnick, S. W., R. N. Zitter and D. B. Williams, (1960) "Attenuation of Infrared Radiation in Fogs," J. Opt. Soc. Am., 50, No. 6, 578-583.

Louisell, W. H., A. Yariv and E. A. Siegman, (1961) "Quantum Fluctuations and Noise in Parametric Processes-I," Phys. Rev., 124, 1646.

Lyot, B., (1933) Comptes Rendus, 197, 1593.

Lyot, B., (1944) American Astrophys., 17, Nos. 1-2.

Messiah, A., (1962) "Mechanique Quantique," Chapter XII, Vol. I, Dunod, Paris.

Messiah, A., (1964) "Mechanique Quantique", Vol. II, Dunod, Paris.

Niblack, W. and E. Wolf, (1964) "Polarization Modulation and Demodulation of Light," Appl. Opt., 3, No. 2, 277.

Nye, J. F., (1960) Physical Properties of Crystals, (Oxford University Press).

Öhman, Y., (1938) "A New Monochromator," Nature, 141, 291.

Peters, C. J., (1963) "Gigacycle Bandwidth Coherent Light Traveling Wave Phase Modulator," Proc. IEEE, 51, No. 1, 147-153.

Peterson, W. W., T. G. Birdsall and W. C. Fox, (1954) "The Theory of Signal Detectability," Trans. IRE, IT-4.

Plass, G., (1962) "Infrared Transmission Studies: I-Spectral Band Absorptance for Atmospheric Slant Path," Aeronutronic Division Report SSD-TDR-62-127.

Plass, G., (1963) "Transmittance Tables for Slant Paths in the Atmosphere-Vol. V," Aeronutronic Division Report SSD-TDR-62-127.

Renk, K. F. and L. Genzei, (1962) "Interference Filters and Fabry-Perot Interferometers for the Far Infrared," Appl. Opt., 1, No. 5, 643-648.

Rollin, R.A. and F. Zwas, "Investigations of Space Communications Systems Using Lasers," Report 5693-9-T, The Optical and Radio Systems Laboratory, Institute of Science and Technology, The University of Michigan (to be published).

Schmidt, R.M., J.M. Williams and D.M. Williams, (1964) "Magneto-Optic Modulation of a Light Beam in Sodium Vapor," J. Opt. Soc. Amer., 54, No. 4, 454.

Seraphin, B.O., D.G. McCauley and L.G. LaMarca, (1963) "Piezoelectric Laser Modulator," Paper presented at the Symposium on Optical Mases, Polytechnique Institute of Brooklyn.

Shannon, C.E. and W. Weaver, (1949) The Mathematical Theory of Communication, (University of Illinois Press, Urbana, Illinois).

Shimoda, K., H. Takahasi and C.H. Townes, (1957) "Fluctuations in the Amplification of Quanta with Applications to Maser Amplifiers," J. Phys. Soc. Japan 12, 685.

Slepian, D., (1965) "Permutation Modulation," Proc. IEEE, 53, 228.

Spencer, P.E., (1960) "Scattering Function for Fogs," J. Opt. Soc. Amer., 50, No. 6, 584-585.

Steel, W.H., R.N. Smartt and R.G. Giovanelli, (1961) "A 1/8Å Birefringent Filter for Solar Research," Australian J. Phys., 14, 201.

Steinberg, H.A., (1964) "Signal Detection with a Laser Amplifier," Proceedings of the Institute of Electrical and Electronics Engineers, 52, 28-32.

Stern, T.E., (1960) "Some Quantum Effects in Communications Channels," Trans. IRE, IT-6, 435-440.

Strong, J., (1958) Concepts of Classical Optics, (W.H. Freeman and Co., San Francisco, California) 305.

U.S. Standard Atmosphere, 1962 (1962) U.S. Government Printing Office, Washington, D.C.

van de Hulst, H.C., (1957) Light Scattering by Small Particles. (John Wiley and Sons).

Van Meter, D. and D. Middleton, (1954) "Modern Statistical Approaches to Reception in Communication Theory," Trans. IRE, IT-4.

Watanabe, K., (1958) "Ultraviolet Absorption Processes in Upper Atmosphere," Advances in Geophysics, -5, (Academic Press, New York).

Zachor, A.S., (1961) "Near Infrared Transmission over Atmosphere Slant Paths," Mithras, Incorporated Report No. MC-61-13.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

The University of Michigan Radiation Laboratory
Department of Electrical Engineering
Ann Arbor, Michigan

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

N/A

3. REPORT TITLE

STUDIES OF PROBLEM AREAS IN OPTICAL COMMUNICATIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report, April 1964 through November 1965

5. AUTHOR(S) (Last name, first name, initial)

Hok, Gunnar; Barasch, M. L.; Lambropoulos, Peter; Miller, E. K.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 1966	189	177

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 33(615)-1021	06515-1-F
b. PROJECT NO. 4335	
c. Task No. 433511	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	AFAL-TR-66-27

10. AVAILABILITY/LIMITATION NOTICES

USC No. 2

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Air Force Avionics Laboratory
Research and Technology Div.
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

13. ABSTRACT

First problem Area: The various causes for the deterioration of an optical signal during its passage from space to a terminal on the earth is considered, such as attenuation by the normal constituents of the atmosphere and by clouds, fog, haze and rain; seeing, scintillation, anomalous dispersion and background radiation. Second problem area: The quantum-mechanical limitations on observations of electromagnetic radiation are reviewed. The statistical theory of communication and of detection of signals is extended to quantum-limited communication channels. A photon counter is shown to be able to recover as much information from a signal as is compatible with the quantum limitations. At low signal levels it is found that binary operation can utilize a high percentage for the channel capacity. A photon counter may be a photomultiplier or a laser amplifier followed by a photomultiplier. A theory of laser amplifiers is presented. The latter type counter, which has not yet been realized, promises the additional advantage of providing discrimination against background radiation outside the channel frequency band. The means for controlling the bandwidth of such an amplifier have not yet been investigated. Third problem area: For the purpose of providing adequate discrimination against background radiation before detection of signals by a photon counter, various optical filter principles are investigated, in particular the Wernicke prismatic filter and the polarization interference filter. The result is that desired bandwidths and tunability are theoretically obtainable but that state-of-the-art limitations are as yet discouraging.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Laser Beam Communication (theoretical Limits) Quantum-Limited Optical Communication Channels Statistical Detection Theory for Quantum-Limited Channels Reception by Photon Counting in Optical Channels						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.