

Chapter 1

Introduction and Definition of Problem

Information theory was initially developed in the context of telephonic communication through copper cables (baseband channels). Radio systems (bandpass channels) correspond to parallel channels with four degrees of freedom corresponding to in-phase and quadrature components relative to the carrier, as well as two orthogonal polarizations. With coherent detection and polarization filters there are four parallel channels, and the basic Shannon results [1] are easily generalized to this case (see section 1.1 below).

In the case of fiber optic systems phase coherence and polarization are not generally maintained, and applications have been primarily focused on incoherent detection without polarization filtering.

The primary focus of this dissertation is to find the channel capacity for the incoherent case involving single (polarization filtering) and two-polarizations (no polarization filtering). In coherent systems, phase modulation may be used and for the ideal coherent system the received signal should not have any phase noise and an exact phase reference needs to be available in the receiver. The incoherent case then becomes important because: a) we may need to use transmitters with insufficient phase stability to permit phase modulation; b) we may be constrained to using incoherent receivers that are incapable of recovering the optical carrier phase, and finally c) the channel might also add a random phase which will make it useless to code phase of the input signal so the phase of the output signal can be measured.

Most of the analysis in this work is done assuming the system to be linear with the main error source being amplified spontaneous emission (ASE) noise from an optical preamplifier in the receiver, as well as the accumulation of ASE from intermediate optical amplifier repeaters. With the use of dispersion management, nonlinearities in the

fiber can be made small and thus this idealized case is representative of many practical systems.

Recently a considerable number of publications [2-8] have appeared dealing with ways to increase the information density or finding upper realistic limits in fiber-optic systems. Some of these publications have dealt mainly with linear systems [2-5] while some work [6-8] has dealt with channel capacity in a non-linear fiber environment. However, apart from upper bounds on capacity, there has been no systematic treatment of the effect of incoherent detection on the capacity of fiber optic systems. Relative to the coherent case, the incoherent channel has two additional constraints: a) the real and imaginary parts of the complex field strength cannot be coded independently, and b) only the magnitude of the received field may be measured.

In addition, the analysis for two-polarizations is important in our work because the single polarization case is impractical in optical systems as it would require use of a polarization tracking filter at the receiver. Although this problem has been considered in the literature, prior work is generally restricted to obtaining bounds on the capacity.

This chapter will summarize the work that has been done in the literature for the linear environment and we will explain our methodology for the following chapters. We will start by giving a brief summary of the Shannon channel capacity limit.

1.1. The Shannon Bound

Claude Shannon derived his capacity formula [1] for a linear baseband channel, but the results are simply generalized to carrier systems with coherent detection which is equivalent to two independent baseband channels. Such a baseband channel is defined by a linear relationship between the output y and the input signal x by the relation:

$$y = x + n \quad (1.1)$$

where n is a Gaussian noise process with a flat power spectrum. The capacity of this channel is achieved when the input signal has also a Gaussian distribution, and is given by

$$C = W \log_2(1 + SNR) \quad (1.2)$$

Here W is the bandwidth of a baseband channel in which $2W$ real numbers are transmitted per unit time. SNR is the ratio of the average transmitted signal power (S) and the variance of the channel noise (N), also known as signal-to-noise ratio.

Although the traditional channel capacity formulas are not valid for non-linear channels (fiber optics communication channels), by using linear channels we can have a foundation for future work for non-linear channels. We will summarize the work done in this area comparing coherent and incoherent (direct) detection for these channels and also extend the work to an ideal fiber channel with the use of 1 and 2 polarizations.

1.2. Receiver Models

Fiber optic systems employ photodetectors that convert optical intensity to an electrical current. Bandpass vector fields may be represented by four orthogonal baseband components, corresponding to two quadrature phases and two orthogonal polarizations. Intensity is then proportional to the sum of the squares of these four components.

In this section we will define the receiver models that we will be using throughout this dissertation and justifying also the importance of each of these models and how they relate to each individual chapter and to the existing literature.

Receivers can then be divided into three basic categories:

1.2.1. Coherent receivers

A strong optical local oscillator (in phase and with same polarization as the signal) is added to the signal prior to the photodetector. This results in the removal of quadrature and orthogonal polarization components, and the problem reduces to a single degree of freedom case of signal plus additive noise, where the noise is local oscillator (LO) shot noise. Electrical noise following the photodetector may be neglected assuming the optical LO is sufficiently strong. The photocurrent is then of the form

$$z_1 = x_1 + n_1 \quad (1.3)$$

where x_1 is proportional to the signal optical field, and n_1 is the effective added noise¹ which is assumed to be Gaussian. We are therefore in the one degree-of-freedom case for which the basic Shannon capacity formula applies. Owing to implementation issues coherent receivers are not generally used in fiber optic systems, but this case serves as a useful benchmark to the best that may be theoretically achieved.

¹ In this dissertation we are concerned with how information capacity depends on signal-to-noise ratio (SNR), but not specifically on how SNR depends on physical parameters. This latter subject is covered adequately in standard texts.

The work we have done in chapter 2 is related to the one degree-of-freedom case. Through Shannon's work we know that the optimal distribution for the input signal to maximize channel capacity is achieved with the Gaussian distribution. In order to get better mathematical insight for cases with higher degrees of freedom we analyzed different input distributions and how they influence the spectral efficiency of the channel.

In Chapter 2 we deal with different possibilities for the input distribution as well as the output of the receiver for the coherent case. We start out with the basic Shannon model

$$y_I = x_I + n_I \quad (1.4)$$

and extend it to non-negative input distributions

$$y_I = |x_I| + n_I \quad (1.5)$$

We then work with the absolute value of the output and see how the mutual information changes.

$$z_I = |y_I| = |x_I + n_I| \quad (1.6)$$

We also extend this work here to non-negative distributions for the input signal

$$z_I = |y_I| = ||x_I| + n_I| \quad (1.7)$$

1.2.2. Optical preamplifier direct detection (incoherent) receivers

Electrical noise following the photodetector may also be neglected if there is an optical amplifier in the receiver prior to the photodetector. However, the amplifier introduces noise (amplified spontaneous emission noise) that is added to the signal. This noise contains both quadrature phase components and both polarizations. Consequently the output of the photodetector is of the form

$$z_4^2 = (x_1 + n_1)^2 + n_2^2 + n_3^2 + n_4^2 \quad (1.8)$$

This is the four DOF case. If a polarization filter were used it would be theoretically possible to remove two of the noise components in which case we would have

$$z_2^2 = (x_1 + n_1)^2 + n_2^2 \quad (1.9)$$

This is the two DOF case. The conditional pdf of z_2 given x_1 is Rician, and this case has been partially studied in the past [9]. However, it is generally not practical to use polarization filters, and one of the objectives of this work is to determine how the inclusion of the orthogonal polarization noise terms affects the capacity.

1.2.3. PIN Direct Detection Receiver

If a simple PIN receiver is used, the dominant noise contribution is electrical thermal noise following the photodetector. In this case the photocurrent is of the form

$$z_e = x_1^2 + n_e \quad (1.10)$$

From an information theory standpoint this is of the same form as (1.3), if in (1.3) we restrict x_1 to be non-negative.

The work done in chapter 2 is valid for the PIN receiver as we have worked with non-negative input distributions such as the Rayleigh, half-Gaussian, modified half-Gaussian (amplitude) and exponential.

Based on the above considerations the principal objectives of this dissertation are to

- a) evaluate the capacity of the one DOF case under the constraint that signal is non-negative, and
- b) evaluate and compare the capacities of the two and four DOF cases.

1.3. Definition of SNR

Before we present the results derived in the literature we need to standardize the definition of the signal-to-noise ratio we will be using so we can have meaningful comparisons between different results.

For the input signal, we take

$$E[x_1^2] = S \quad (1.11)$$

and for the noise components

$$E[n_1^2] = E[n_2^2] = E[n_3^2] = E[n_4^2] = N \quad (1.12)$$

where E is the statistical average. $\text{SNR} = S/N$ which is the average signal power divided by the average noise power in one DOF.

1.4. Parallel channels

For the coherent receiver (one DOF) case the channel capacity is given by

$$C_{1c} = \frac{W}{2} \log\left(1 + \frac{S}{N}\right) \quad (1.13)$$

where W is the optical bandwidth, and the subscripts 1c denote 1 DOF and coherent detection.

The factor of 2 in this equation is a consequence of the fact that in a bandpass channel there are W independent complex numbers that may be transmitted per second, but half of these (say the real components) are used. With two parallel channels in which the real

and imaginary parts (corresponding to components that are in phase and in quadrature with a carrier) are used, the capacity is given by²

$$C_{2c} = W \log \left(1 + \frac{S}{2N} \right) \quad (1.14)$$

Note, however, that S is the total power, but N is the noise power per DOF. If, in addition orthogonal polarizations are used with the total power divided over four parallel independent channels, we obtain

$$C_{4c} = 2W \log \left(1 + \frac{S}{4N} \right) \quad (1.15)$$

In the limit of high SNR, $C_{4c} \approx 4C_{1c}$, whereas in the limit of low SNR, $C_{4c} \approx C_{1c}$.

1.5. Literature on the capacity of incoherent channels

The work done by Jacobs [4] deals with the capacity of the incoherent Gaussian channel. This case is essentially the case which was defined in Section 1.2.2 with 2 degrees of freedom or single polarization given by (1.9).

In the case of incoherent detection there are two constraints:

- i) The signal can not be transmitted independently in each of the four channels
- ii) The four channels can not be separated at the receiver.

If we apply the first but not the second constraint we should get an upper bound to the capacity of the incoherent channel. If the same signal $x_1/2$ is transmitted in each of the four channels (or $x_i/\sqrt{2}$ in each of two channels), and the four (or two) channels are

² Note that (1.14) becomes the conventional Shannon formula if $2N$ is replaced by N , so that N is the total noise power not the noise power per DOF. In comparing alternative systems we chose to fix the total signal power, and the noise power per DOF

averaged at the receiver this results in a single channel with a SNR given by S/N for both the two and four channel cases.

Thus, for both the two and four channel cases we obtain identical upper bounds to the incoherent channel capacity given by

$$C_i \leq \frac{W}{2} \log_2(1 + SNR) \quad (1.16)$$

The fact that the bound on the capacity of the incoherent channel is independent of the number of DOFs is not surprising in the high SNR case. For $SNR \gg 1$, it follows from (1.8) that

$$z_4^2 = (x_1 + n_1)^2 + n_2^2 + n_3^2 + n_4^2 \approx x_1^2 + 2x_1n_1 \quad (1.17)$$

so that the dominant noise term is the noise component that is in phase and of the same polarization as the signal.

1.6. Direct Detection Bound obtained by Shtaif and Mecozzi

Shtaif and Mecozzi [2,3] presented an approximation for the spectral efficiency for the direct (incoherent) detection case for high Signal-Noise Ratio (SNR).

This section will explain the results they obtained and the method used by them. The results obtained by Shtaif and Mecozzi will be used as another benchmark throughout this dissertation.

The major difference between the work done by Shtaif and Mecozzi and the work done by Jacobs described in Section 1.5 is that Jacobs bounded the capacity of an incoherent channel by considering it as two coherent channels in which the real and imaginary signal components are not coded independently, whereas Shtaif considered the actual incoherent channel model, and obtained an approximation to the mutual information in the high SNR limit.

Here, the channel is assumed to be linear with the main source of errors being the amplified spontaneous emission noise. Although this is somewhat of an idealized-case it does provide us an upper bound for the actual capacity that a real system can support.

The starting point for this scenario is the generic expression for the channel capacity given by Shannon

$$C = \frac{1}{T} \max_{P_x(x)} \int dx P_x(x) \int dy P(y|x) \log_2 P(y|x) - \int dy P_y(y) \log_2 P_y(y) \quad (1.18)$$

Here x and y are the transmitted and received signals and the channel is described by the transition probability $P(y/x)$ which is the conditional pdf for receiving y when x has been transmitted.

The probability of y depends on the probability of the input signal, $P_x(x)$, through the following relation

$$P_y(y) = \int dx P(y|x) P_x(x) \quad (1.19)$$

To obtain the capacity the integral has to be maximized over all possible distributions of x . Using calculus of variations combined with the method of Lagrange multipliers, an expression for the optimal input distribution $P_x(x)$ which maximizes the integral was obtained. This distribution has an average power constraint given by

$$\int dx P_x(x) x^2 = S \quad (1.20)$$

For an incoherent detection system the relation between transmitted and received signal is given by $y = |x+n|$. It should be noted that the term n is a complex Gaussian noise term generated by amplified spontaneous emission (ASE). This relation implies that x is the electric field of the transmitted signal which can be assumed to be a real signal and y is the square root of the detected intensity. Amplitudes are used here for x and y instead of transmitted and received intensities as it simplifies the derivation and the physical interpretation of the results that follow.

The channel distribution $P(y/x)$ is related to the non-central chi square distribution with two or four degrees of freedom depending on whether a polarization sensitive or insensitive receiver is used.

A simple solution is obtained in the asymptotic case, when $\text{SNR} \gg 1$. According to Shtaif and Mecozzi, the optimal distribution (the one that maximizes Mutual Information) for x , $P_x(x)$, when the SNR is high is the half-Gaussian distribution (Gaussian for $x > 0$ and 0, otherwise). This is confirmed by recent results published by Ho [5] where by using different numerical algorithms the half-Gaussian was found to be the optimum distribution in the high SNR limit.

By substituting the half-Gaussian distribution as the input distribution into the capacity formula, the following capacity per state of polarization is obtained in the high SNR limit

$$C_I = W \log_2 \left(\frac{SNR}{4} \right) \quad (1.21)$$

An important point to notice is the difference between the expression above and the Shannon Capacity (1.2). With the factor of 4 in (1.21), we have the following relation between the incoherent case and the Shannon formula in the high SNR limit

$$C_I = C - 2W \quad (1.22)$$

The single polarization case can be extended for a higher number of polarizations, the expression for the spectral efficiency is given by

$$n_I \approx \frac{m}{2} \log_2 \left(\frac{SNR}{4} \right) \quad (1.23)$$

where the possibility of polarization multiplexing is taken into account by multiplying the information density by m which represents the number of polarization states. Note that this assumes the use of polarization filters.

In the notation (mainly in regards to the definition of SNR) we have been using throughout this chapter, the results obtained by Shtaif and Mecozzi for the single polarization (or 2 DOFs) can be expressed as

$$C_{i(SM)} \leq \frac{W}{2} \log \left(\frac{S}{2N} \right) = \frac{W}{2} \left[\log \left(\frac{S}{N} \right) - \log 2 \right] \quad (1.24)$$

In the next figure we compare the expressions obtained by Meozzi and Shtaiif for the incoherent case with the classic Shannon coherent case formula. The comparison is for the single polarization case(2 DOFs).

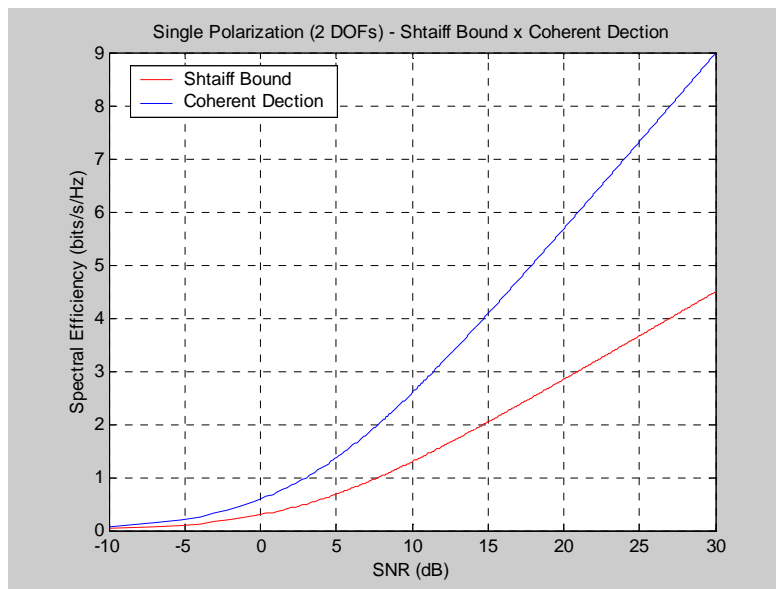


Fig.1.1 Spectral efficiency – Incoherent detection with single polarization (Shtaiif and Meozzi)

The following figure compares the expressions obtained by Shtaiif and Meozzi and the bound obtained by Jacobs which has already been discussed in the first part of this chapter.

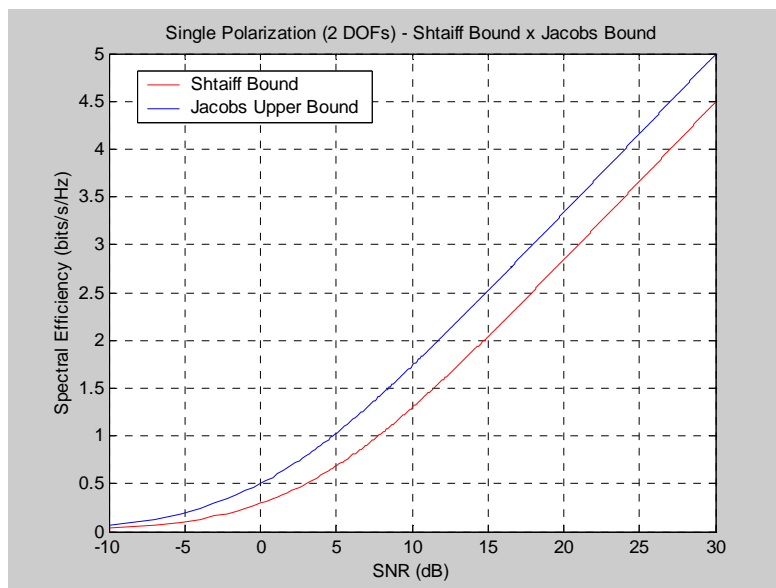


Fig.1.2 Spectral efficiency – Incoherent detection with single polarization (Comparison of Shtaiif and Jacobs bounds)

As can be seen the work done by Shtaiff and Mecozzi gives us a similar but more conservative bound and should be assumed as the better bound. It should also be noted that the results by Shtaiff and Mecozzi are only valid for high values of SNR so for low values of SNR this comparison might not be valid.

1.7. Methodology for following chapters

As mentioned in the previous section, an approximation has been found by Mecozzi and Shtaif in [2,3] when the SNR $\gg 1$ and in this case the optimum distribution is the half-Gaussian distribution. The next few chapters will give us a mathematical basis to derive a more complete expression for any value of SNR for the incoherent case for both single and 2-polarizations.

We will also get valuable insight into how different input distributions behave for different channels, so we can verify if indeed the half-Gaussian distribution is the optimum distribution. In the next few chapters we will concentrate our efforts on distributions which only exist for $x > 0$, such as the half-Gaussian, Rayleigh, modified half-Gaussian (amplitude) and exponential distributions.

We will start in chapter 2 with the basic $Y=X+N$ case (single degree of freedom, X and N real)³ where we know according to Shannon's theorem that the Gaussian distribution is the optimum in order to maximize the channel capacity. Here we will see how different distributions behave and how the mutual information changes according to the input distribution. In the existing literature we were not able to find results for different distributions, only the proof of Shannon's bound with the Gaussian distribution is usually shown. We will then extend this work to the case where we have $Z=|Y|$; i.e., the receiver looks only at the magnitude of the received signal, this will give us insight on how PIN receivers work.

We will numerically evaluate the mutual information $I(X;Y)$ and $I(X;Z)$ for a complete range of SNR including low values of SNR. The motivation for including low values of

³ We will generally use capital letters to represent random variables and the corresponding lower case letter to indicated realizations of that random variable.

SNR is the power efficiency. These results will be used to calculate the minimum number of photons per bit necessary.

In chapter 3 we will extend the work of the previous chapter to two and four degrees of freedom. Work done by Jacobs previously has found bounds for 2 degrees of freedom. We repeat some of that work in chapter 3 and extend it to when we have 4 degrees of freedom (incoherent detection with 2 polarizations). Here we will prove that the bound on the capacity of the incoherent channel is independent of the number of DOFs.

We will start out first by finding the conditional pdf of z_4 given x_1 which is not Rician unlike the 2 DOFs case and prove that both conditional probabilities have the same behavior for high SNR thus validating our assumption that the bounds are also the same.

In addition to bounding the capacity we will numerically obtain the mutual information for all ranges of SNR for 2 and 4 degrees of freedom including low values of SNR.

Also in chapter 3 we will show the exact analytical expressions we have obtained for the spectral efficiency for high values of m (number of degrees of freedom) and compare these expressions with the numerical results we obtained for 2 and 4 degrees of freedom to see if these simpler analytical expressions can be used as approximations for lower values of m .

Finally in chapter 4 we will both summarize the major contributions of this dissertation and outline the open questions that remain in this dissertation and possible ways of solving these open questions for future work.

Chapter 2

Output Distributions and Spectral Efficiencies for Coherent Receivers (1 DOF case)

The reasoning behind the work in this chapter is that in the literature there are results for the optimal input distribution, the Gaussian distribution, which gives us the channel capacity, but there are no results for different input distributions. We need these results for various input distributions to get a better understanding in solving for a generic expression for the channel capacity for the incoherent case for single (2 DOFs) and two polarizations (4 DOFs).

Also, these results are related to the incoherent model with no polarization filter when we consider a PIN receiver where the effective noise power (relative to the signal power at the input of the receiver) is much higher than that in the ASE noise case.

We start by analyzing the mutual information for the simple case of when the output is a sum of the non-negative channel input and the channel noise (assumed to be Gaussian). In other words, we are evaluating the case $y_I = x_I + n_I$ with the restriction that $x_I \geq 0$. We will consider three non-negative input distributions; Rayleigh, half-Gaussian, and exponential.

We use non-negative distributions here because in the existing literature for the incoherent case, non-negative distributions (mainly the half-Gaussian and Rayleigh) were used to find bounds for the channel capacity.

We will start this chapter by calculating the mutual information when the input distribution is Rayleigh. To calculate the mutual information for any input distribution we will initially calculate the distribution of the output signal when we have additive Gaussian noise. Following the Rayleigh case, we will move on to the half-Gaussian input and then the exponential distribution. In the end we will compare the spectral efficiencies obtained with all three distributions, which will give us valuable insight on which input

might not only give us the best bound but also the best general expression for the channel capacity for the incoherent case.

After we have calculated the mutual information for the case of $y_I = x_I + n_I$ we extend our work to when we are restricted to taking the absolute value of the output, in other words, $z_I = |y_I| = |x_I + n_I|$. This is an intermediary step towards finding the channel capacity when incoherent receivers are used. Here we do not restrict ourselves only to non-negative input distributions, and use the Gaussian distribution also, proving that even for non-negative outputs, the Gaussian distribution also maximizes the mutual information.

At the end of this chapter we will show numerical results obtained for low SNR for a half-Gaussian input. The motivation to work with low SNR values is to obtain the minimum number of photons per bit.

2.1. Output PDF f_y with Rayleigh input and Gaussian noise

We will work in this section with the Rayleigh distribution as the input distribution. The first step towards finding the mutual information is to find the output distribution $f_y(y)$. In order to find it we will work with the Fourier transforms of the input distribution and the noise, multiply each other and then find the inverse Fourier transform of the product.

The Fourier transform of the Rayleigh input distribution is given by [10,11]:

$$F_x(w) = \left(1 + j\sqrt{\frac{\pi u}{2}}w\right)e^{-\frac{uw^2}{2}} \quad (2.1)$$

The average signal power S is related to u by the following relation $S=2u$. Thus in terms of total signal power the Fourier transform of the Rayleigh distribution is given by

$$F_x(w) = \left(1 + j\sqrt{0.25\pi S}w\right)e^{-0.25Sw^2} \quad (2.2)$$

The Fourier transform of the Gaussian noise pdf is given by:

$$F_n(w) = e^{-\frac{w^2N}{2}} \quad (2.3)$$

with N representing the average noise power.

$$F_y(w) = F_x(w).F_n(w) \quad (2.4)$$

Taking the inverse Fourier transform the output distribution is found to be a combination of two functions defined as follows:

$$\begin{aligned}
 &-\infty \leq y \leq 0 \\
 &f_y = \frac{1}{2} \left(\sqrt{\frac{0.5S}{0.5S+N}} - 1 \right) \frac{y}{0.5S+N} \exp\left(-\frac{y^2}{2(0.5S+N)}\right) \\
 &0 \leq y \leq \infty \\
 &f_y = \left(1 + \frac{1}{2} \left(\sqrt{\frac{0.5S}{0.5S+N}} - 1 \right) \right) \frac{y}{0.5S+N} \exp\left(-\frac{y^2}{2(0.5S+N)}\right)
 \end{aligned} \tag{2.5}$$

Figures 2.1 and 2.2 show how the output distribution (as a function of the variable y) also becomes Rayleigh as $\text{SNR} \gg 1$ which is something expected as at high SNR values the signal power becomes dominant.

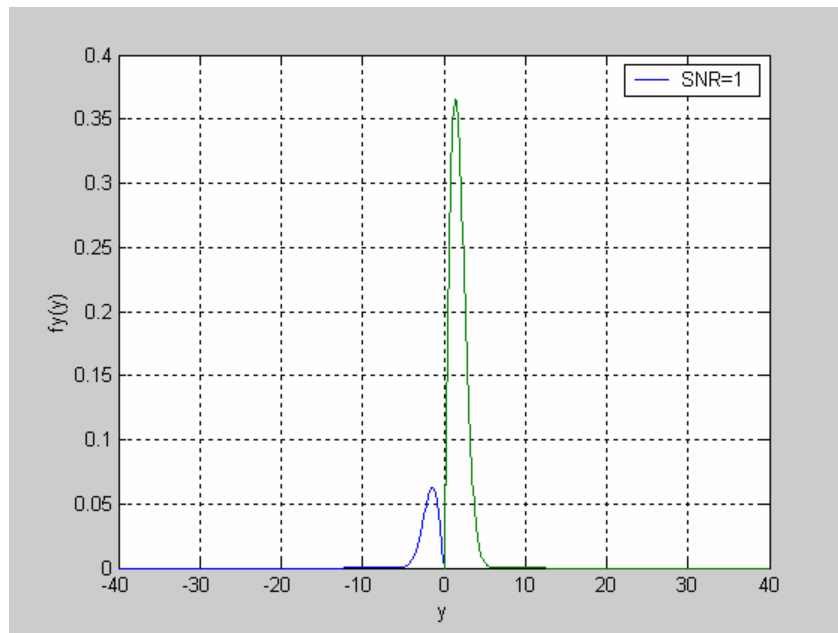


Fig.2.1 Output distribution when the input distribution is Rayleigh and the SNR=1

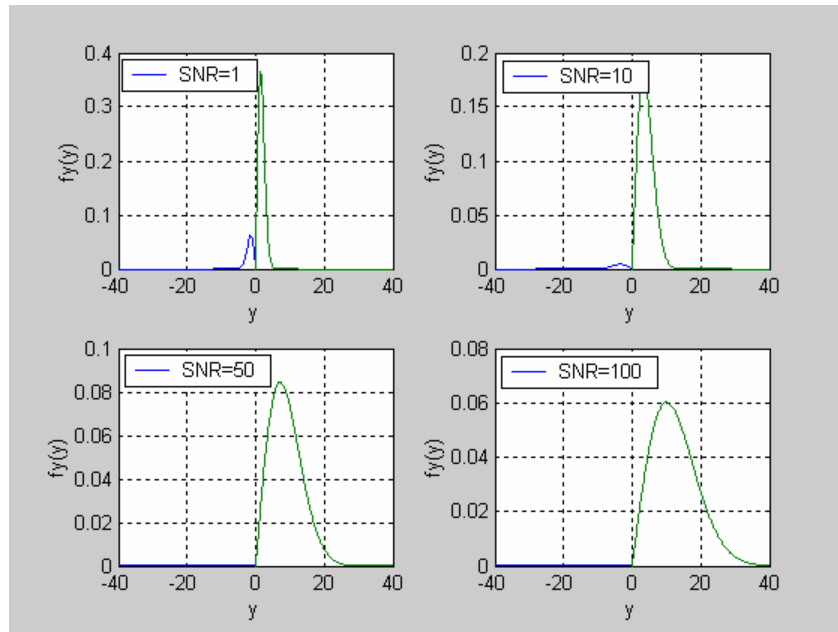


Fig.2.2 Plot of $f_y(y)$ when $f_x(x)$ is Rayleigh for various values of SNR

2.2. Calculation of mutual information expression with f_y as the output distribution and Rayleigh input

Now that we have the output distribution we can continue calculating the mutual information. The next step is to calculate the entropy of y . The entropy is given by

$$H(y_1) = -\int_{-\infty}^0 f_y \log f_y dy - \int_0^{\infty} f_y \log f_y dy \quad (2.6)$$

By evaluating the integral in (2.6), the entropy can be shown to be given by:

$$H(y_1) = -\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \log \left(\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \right) + 1 + \frac{\gamma}{2} + \log \sqrt{\frac{0.5S+N}{2}} \quad (2.7)$$

with γ being the Euler-Gamma constant.

Now that we have the entropy of y , we can calculate the expression for the mutual information $I(x;y)$. By definition it is the difference between the entropy of y (output) and the entropy of the channel noise n .

$$I(x_1; y_1) = H(y_1) - H(n) \quad (2.8)$$

The entropy of n which is Gaussian noise is given by:

$$H(n) = \log(2\pi eN) \quad (2.9)$$

By simplifying $I(x_1; y_1)$, we obtain the following expression for mutual information for a Rayleigh input with Gaussian noise:

$$I(x_1; y_1) = -\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \log \left(\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \right) + \frac{\gamma+1}{2} + \frac{1}{2} \log \left(\frac{1}{4\pi} (0.5SNR+1) \right) \quad (2.10)$$

We can get a simpler expression for high values of the SNR, (2.10) simplifies to:

$$I(x_1; y_1) = \frac{1}{2} \left[\gamma + 1 + \ln \left(\frac{0.5SNR+1}{4\pi} \right) \right] \quad (2.11)$$

It should be noted that by using the natural logarithm the capacity is currently expressed in nits per transmission. To obtain it in bits, we have to divide these two expressions by a factor of $\ln 2$.

Numerical evaluation and plots of these expressions are provided in Section 2.7 where the spectral efficiencies for the three distributions considered in this chapter are compared.

2.3. Output PDF f_y with half-Gaussian input and Gaussian noise

In this section we will work with the half-Gaussian distribution as the input distribution. The probability density function for the half-Gaussian distribution is given by (2.12):

$$f_x(x) = \frac{2}{\sqrt{2\pi S}} e^{-\frac{x^2}{2S}} u(x) \quad (2.12)$$

where $u(x)$ is the unit step function.

And similarly to the previous section the channel is also Gaussian with the noise distribution given by

$$f_n(n) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{n^2}{2N}} \quad (2.13)$$

With the half-Gaussian distribution as the input the approach to obtain the output distribution is different than with the Rayleigh distribution as input. Instead of working with Fourier transforms, it is easier to work with a convolution of the input signal and the noise. This is given by

$$f_y(y) = f_x * f_n = \int_0^{\infty} \frac{2}{\sqrt{2\pi S}} e^{-\frac{\tau^2}{2S}} \cdot \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-\tau)^2}{2N}} \partial \tau \quad (2.14)$$

The following expression is then obtained for the output distribution, which will be used in the calculations to find the mutual information expression:

$$f_y(y) = \underbrace{\sqrt{\frac{1}{2\pi(S+N)}} e^{-\frac{y^2}{2(S+N)}}}_{f_1} \left[\underbrace{1 + \operatorname{erf}\left(\frac{y\sqrt{SNR}}{\sqrt{2S+2N}}\right)}_{f_2} \right] \quad (2.15)$$

Figure 2.3 shows the output distribution (2.15), and how it also becomes a half-Gaussian distribution as the SNR increases. This is a valid pdf; it integrates to 1 and is non-negative.

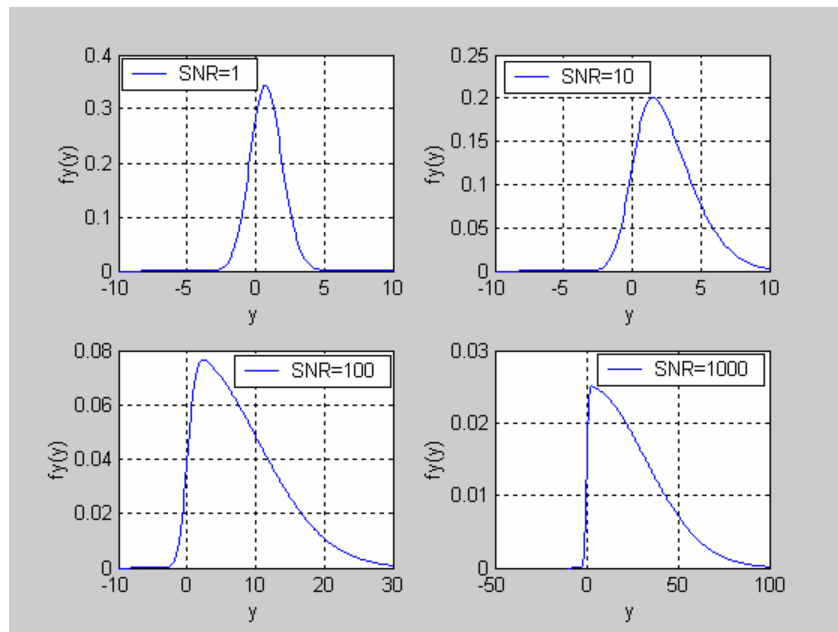


Fig.2.3 Plot of $f_y(y)$ when $f_x(x)$ is half-Gaussian for various values of SNR

2.4. Calculation of mutual information expression with f_y as the output distribution and half-Gaussian input

In a similar manner to the Rayleigh case we start out calculating the mutual information by first calculating the entropy of y . By definition this is given by

$$H(y_1) = \int_{-\infty}^{\infty} f_y \log f_y dy = - \int_{-\infty}^{\infty} f_1 f_2 \log(f_1 f_2) dy = - \int_{-\infty}^{\infty} f_1 f_2 (\log f_1 + \log f_2) dy \quad (2.16)$$

A final simplified expression for $H(y)$ is given by:

$$H(y_1) = -\log\left(\sqrt{\frac{1}{2\pi(S+N)}}\right) + \frac{1}{2} - Int_1(SNR) \quad (2.17)$$

where

$$Int_1 = \sqrt{\frac{1}{\pi SNR}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{SNR}} [1 + erf(x)] \log[1 + erf(x)] dx \quad (2.18)$$

And thus, we are able to get the expression for mutual information between x and y :

$$\Rightarrow I(x_1; y_1) = -\log\left(\sqrt{\frac{1}{2\pi(S+N)}}\right) + \frac{1}{2} - Int_1(SNR) - \frac{1}{2} \log(2\pi e N) \quad (2.19)$$

Simplifying (2.19), we obtain the following final expression for $I(x;y)$ when the input has a half-Gaussian distribution:

$$I(x_1; y_1) = \frac{1}{2} \log(SNR + 1) - Int_1(SNR) \quad (2.20)$$

2.5. Output PDF f_y with exponential input and Gaussian noise

The final input distribution we work with in this chapter is the exponential distribution which is given by (2.21), where again S represents the average signal power.

$$f_x(x) = \lambda e^{-\lambda x} u(x) \quad , S = \frac{2}{\lambda^2} \rightarrow \lambda = \sqrt{2/S} \quad (2.21)$$

The exponential input is treated in a manner similar to the half-Gaussian input where we use a convolution between the input signal and the noise to obtain the output distribution.

The final expression for the output distribution is

$$\therefore f_y(y) = \frac{e^{-\frac{N\lambda^2}{2}}}{2} \lambda e^{-\lambda y} \left[1 + \operatorname{erf}\left(\frac{y - N\lambda}{\sqrt{2N}}\right) \right] \quad (2.22)$$

This is a valid pdf, it integrates to 1 and is non-negative. Also, as the $\text{SNR} \gg 1$ the output pdf becomes an exponential distribution as can be seen in Figure 2.4.

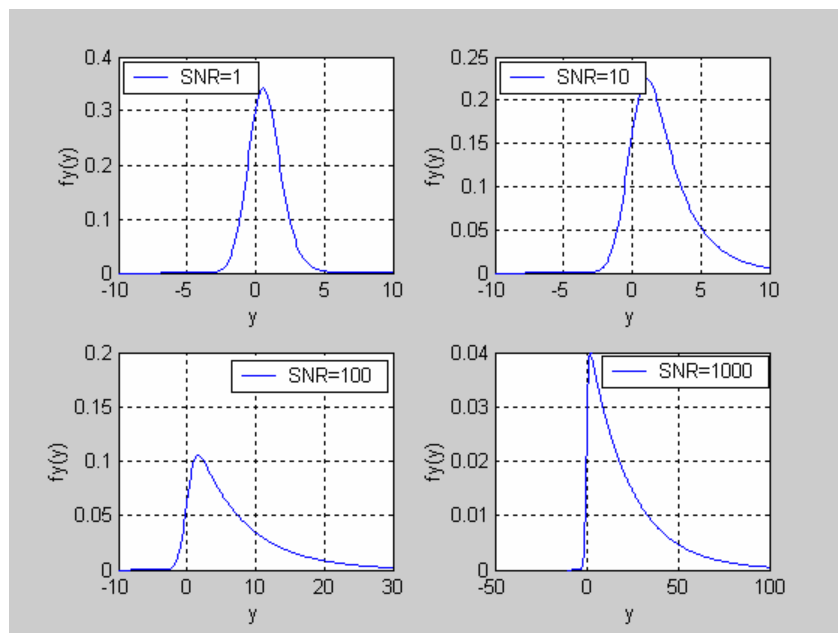


Fig. 2.4 Plot of $f_y(y)$ when $f_x(x)$ is exponential for various values of SNR

2.6. Calculation of mutual information expression with f_y as the output distribution and exponential input

We calculate the entropy of Y by using (2.22) as the output distribution in the definition of entropy:

$$H(y_1) = - \int_{-\infty}^{\infty} \frac{e^{-\frac{N\lambda^2}{2}}}{2} \lambda e^{-\lambda y} \left[1 + \operatorname{erf} \left(\frac{y - N\lambda}{\sqrt{2N}} \right) \right] \left[\log \left(\frac{e^{-\frac{N\lambda^2}{2}}}{2} \lambda e^{-\lambda y} \right) + \log \left(1 + \operatorname{erf} \left(\frac{y - N\lambda}{\sqrt{2N}} \right) \right) \right] \partial y \quad (2.23)$$

The final expression for $I(x_1; y_1)$ for an exponential input is also only dependent on the SNR and given by

$$\boxed{\therefore I(x_1; y_1) = \frac{1}{2} \log \left(\frac{\text{SNR}}{\pi} \right) + \frac{1}{2} - \frac{2}{\text{SNR}} - \text{Int}_2} \quad (2.24)$$

where

$$\boxed{\text{Int}_2 = \frac{1}{\sqrt{\text{SNR}}} e^{-\frac{1}{\text{SNR}}} \int_{-\infty}^{\infty} e^{-\frac{2x}{\sqrt{\text{SNR}}}} [1 + \operatorname{erf}(x)] \log[1 + \operatorname{erf}(x)] \partial x} \quad (2.25)$$

2.7. Comparison of spectral efficiencies for Rayleigh, half-Gaussian and exponential inputs for $y=x+n$

Now that we have expressions for mutual information for all three input distributions, we can compare all three and verify how the spectral efficiency behaves for each one of them for various values of SNR. The following general equations were used for each:

Rayleigh:

$$C = -\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \log \left(\frac{1}{2} \left(\sqrt{\frac{0.5SNR}{0.5SNR+1}} + 1 \right) \right) + \frac{\gamma+1}{2} + \frac{1}{2} \log \left(\frac{1}{4\pi} (0.5SNR+1) \right) \quad (2.26)$$

Half-Gaussian:

$$C = \frac{1}{2} \log(SNR+1) - Int_1(SNR) \quad (2.27)$$

with:

$$Int_1 = \sqrt{\frac{1}{\pi SNR}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{SNR}} [1 + erf(x)] \log[1 + erf(x)] \partial x \quad (2.28)$$

Exponential:

$$C = \frac{1}{2} \log\left(\frac{SNR}{\pi}\right) + \frac{1}{2} - \frac{2}{SNR} - Int_2 \quad (2.29)$$

with:

$$Int_2 = \frac{1}{\sqrt{SNR}} e^{-\frac{1}{SNR}} \int_{-\infty}^{\infty} e^{-\frac{2x}{\sqrt{SNR}}} [1 + erf(x)] \log[1 + erf(x)] \partial x \quad (2.30)$$

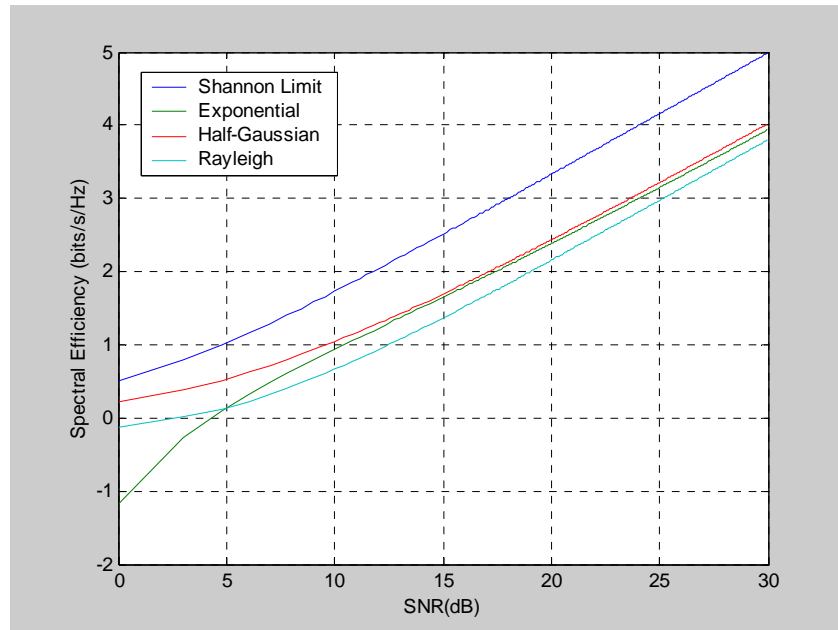


Fig. 2.5 Comparison of spectral efficiencies for various distributions

Figure 2.5 shows the spectral efficiency for all three distributions we worked with in this Chapter. They are compared to the Gaussian distribution, which was demonstrated by Claude Shannon as the optimal distribution to maximize the channel capacity (spectral efficiency). As expected the half-Gaussian has the best results among distributions which are only defined for $x > 0$.

One of the challenges in this work is to find expressions that hold for extremely low values of SNR. As can be seen from the results above the expressions we derived are plotted for values of SNR > 0 dB but not below, even though these expressions should be valid for all SNR. The reason for this is that extending these curves to lower SNR gives rise to some computational difficulties when evaluating the integrals involved, especially for the Rayleigh and exponential distributions. By making some substitutions in the integrals we have been able to numerically extend these curves for the half-Gaussian distribution and this work will be described in detail in section 2.9 where we focus on the channel capacity for low SNR values.

We can also get a lower bound for high SNR values, these are approximated by:

Rayleigh:

$$C > \frac{1}{2} \log \left(\frac{0.5SNR}{4\pi} \right) \quad (2.31)$$

Half-Gaussian:

$$C > \frac{1}{2} \log \left(\frac{SNR}{4} \right) \quad (2.32)$$

Exponential:

$$C > \frac{1}{2} \log \left(\frac{SNR}{4\pi} \right) \quad (2.33)$$

It should be noted that equation (2.32) for the half-Gaussian input distribution gives us the same results obtained by Shtaiif and Mecozzi [3] in their work.

Figure 2.6 compares these upper bounds among themselves and also to the Gaussian input. Once again, the half-Gaussian input is the one that most closely approaches the optimum value.

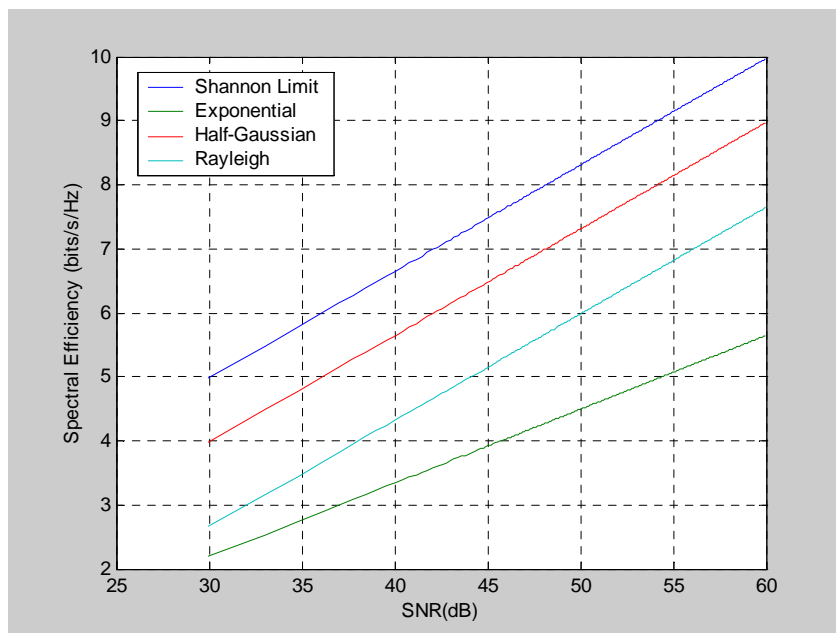


Fig.2.6 Comparison of spectral efficiencies for various distributions with high SNR bound

2.8. Calculation of $I(x;z)$ when $z_I=|y_I|$ by using a Markov chain

We now extend our work to when the receiver is only free to measure the absolute value of the output, in other words, $z_I=|y_I|$.

There are two methods to calculate $I(x_I; z_I)$. The first one is using the definition of mutual information which requires us to find $f_z(z)$, the distribution of z_I , while the second method uses a Markov chain taking advantage of already having $I(x_I; y_I)$.

The Markov chain method is useful when using $f_z(z)$ becomes difficult in the mutual information equation and $f_y(y)$ is available.

The Markov chain relates X , Y , and Z , i.e.

$$X \rightarrow Y \rightarrow Z$$

where Y and Z are defined as

$$Y=X+N$$

$$Z=|Y|$$

From the chain rule theorem found in [12,15] we have that

$$I(X; Y, Z) = I(X; Z) + I(X; Y | Z) = I(X; Y) + I(X; Z | Y) \quad (2.34)$$

Since Z/Y is independent of X

$$I(X; Z | Y) = 0 \quad (2.35)$$

Therefore, using (2.34) and (2.35) we come to the conclusion that

$$I(X;Z) = \underbrace{I(X;Y)}_{\text{already have this}} - I(X;Y|Z) \quad (2.36)$$

We know that by definition $I(X;Y)$ can be written as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (2.37)$$

Thus, in a similar manner we can write $I(X;Y|Z)$ as

$$I(X;Y|Z) = H(X) - H(X|Y|Z) = H(Y|Z) - H(Y|Z|X) \quad (2.38)$$

Substituting (2.37) and (2.38) into (2.36) gives us the following expression for $I(X;Z)$

$$I(X;Y|Z) = H(\eta) - H(\eta|X) \quad (2.39)$$

where $\eta = Y|Z$.

Whereas Y is a continuous random variable, η is a binary random variable taking the values

$$\begin{cases} \eta = +z & \text{when } y > 0 \\ \eta = -z & \text{when } y < 0 \end{cases}$$

The term $I(X;Y|Z)$ can be considered as a correction term which gives us the difference between $I(X;Y)$ and $I(X;Z)$ and in general the loss in channel capacity when the receiver is able to measure only the absolute value of Y .

In (2.39) from the definition of entropy, the terms $H(\eta)$ and $H(\eta|X)$ will be of the form:

$$-p_1 \log p_1 - p_2 \log p_2 \quad (2.40)$$

where in $H(\eta)$, p_1 and p_2 are defined as

$$\begin{aligned} p_1 &= \int_{-\infty}^0 f_y dy \\ p_2 &= \int_0^{\infty} f_y dy \end{aligned} \quad (2.41)$$

from which

$$H(\eta) = - \left[\int_{-\infty}^0 f_y dy \right] \log \left[\int_{-\infty}^0 f_y dy \right] - \left[\int_0^{\infty} f_y dy \right] \log \left[\int_0^{\infty} f_y dy \right] \quad (2.41)$$

To calculate $H(\eta|X)$ for any value of X we need to initially calculate it when $X=x$

In this case the probabilities p_1 and p_2 are given by

$$\begin{aligned} p_1 &= \text{prob}\{y < 0\} = \text{prob}\{n < x\} = \Phi(x) \\ p_2 &= 1 - \Phi(x) \end{aligned} \quad (2.43)$$

Here n is the noise and the function $\Phi(x)$ is defined in terms of the error function, $\text{erf}(x)$ and given by

$$\Phi(x) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right]$$

Thus by putting (2.43) into (2.42) we obtain

$$H(\eta | X = x) = -\Phi(x) \log \Phi(x) - (1 - \Phi(x)) \log(1 - \Phi(x)) \quad (2.44)$$

Thus, for any X we have the following relation

$$H(\eta | X) = \int H(\eta | X = x) f_x(x) dx \quad (2.45)$$

2.8.1. Half-Gaussian input

To test the Markov chain method, we will initially use the half-Gaussian distribution as the input distribution $f_x(x)$

From previous results and using the method outlined in the previous section. $I(x_1; z_1)$ is given by

$$\begin{aligned} I(x_1; z_1) &= I(x_1; y_1) - [H(\eta) - H(\eta | x_1)] \\ &= \left[\frac{1}{2} \log(SNR + 1) - Int_1(SNR) \right] - H(\eta) + H(\eta | x_1) \end{aligned} \quad (2.46)$$

where:

$$Int_1 = \sqrt{\frac{1}{\pi SNR}} \int_{-\infty}^{\infty} e^{-\frac{x_1^2}{SNR}} [1 + erf(x_1)] \log[1 + erf(x_1)] dx_1 \quad (2.47)$$

For high SNR, the term above will go to $\log 2$, and $H(\eta)$ and $H(\eta | x_1)$ will go to zero. Figure 2.7 shows the behavior of each of these terms as the SNR increases.

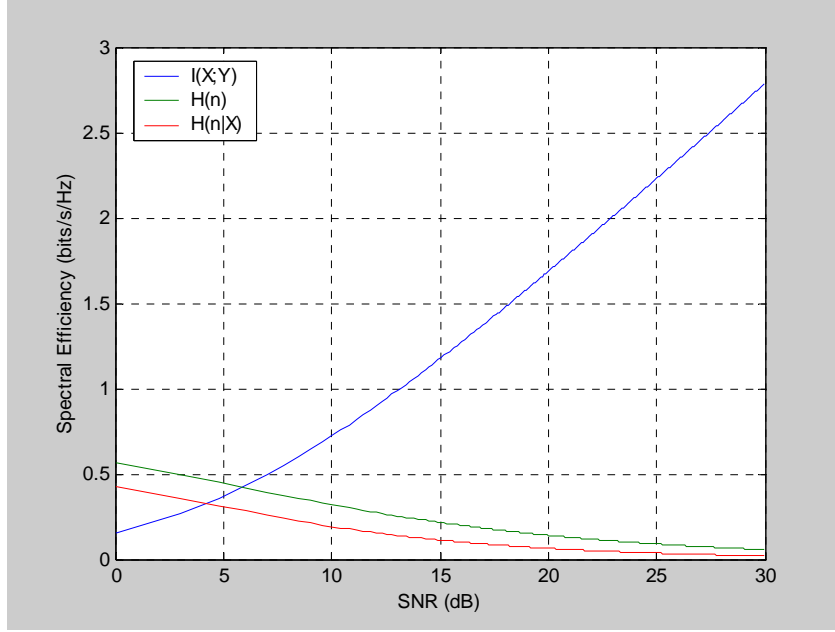


Fig. 2.7 Contributions of each part of $I(x_I; z_I)$ for half-Gaussian input

Thus we can get the following for $I(x_I; z_I)$ in the limit of high SNR

$$I(x_I; z_I) = \frac{1}{2} \log(SNR) - \log 2 = \frac{1}{2} \log\left(\frac{SNR+1}{4}\right) \quad (2.48)$$

which matches the results obtained with the previous method and is the same approximation we obtained for $I(x_I; y_I)$ for high values of SNR.

The simplified expression above was for high values of the Signal-to-Noise Ratio. We can also obtain a closed expression for any value of SNR, this is given by

$$I(x_I; z_I) = \left[\frac{1}{2} \log(SNR+1) - \text{Int}_1(SNR) \right] + H(\eta | x_I) + \left[\frac{1}{2} - \frac{1}{\pi} a \tan(\sqrt{SNR}) \right] \log \left[\frac{1}{2} - \frac{1}{\pi} a \tan(\sqrt{SNR}) \right] \\ + \left[\frac{1}{2} + \frac{1}{\pi} a \tan(\sqrt{SNR}) \right] \log \left[\frac{1}{2} + \frac{1}{\pi} a \tan(\sqrt{SNR}) \right] \quad (2.49)$$

The expression above was evaluated numerically and the results for $I(x_I; z_I)$ are shown later in Figure 2.11 in Section 2.8.5, where this result is also compared to results obtained for other distributions.

2.8.2. Rayleigh input

As for the half-Gaussian case, the relationship we obtained through the Markov chain method to calculate $I(x_I; z_I)$ is given by

$$I(x_I; z_I) = I(x_I; y_I) - [H(\eta) - H(\eta | x_I)] \quad (2.50)$$

So the final expression for $I(x_I; z_I)$ is given by

$$\begin{aligned} I(x_I; z_I) = & -\frac{1}{2} \left(\sqrt{\frac{SNR}{SNR+1}} + 1 \right) \log \left(\frac{1}{2} \left(\sqrt{\frac{SNR}{SNR+1}} + 1 \right) \right) + \frac{\gamma+1}{2} + \frac{1}{2} \log \left(\frac{1}{4\pi} (SNR+1) \right) - H(\eta) \\ & - \left(\frac{1}{S} \int_0^\infty [1 + \operatorname{erf}(\alpha)] \log \left[\frac{1}{2} (1 + \operatorname{erf}(\alpha)) \right] \cdot \alpha \cdot e^{-\frac{\alpha^2}{S}} d\alpha + \frac{1}{S} \int_0^\infty [1 - \operatorname{erf}(\alpha)] \log \left[\frac{1}{2} (1 - \operatorname{erf}(\alpha)) \right] \cdot \alpha \cdot e^{-\frac{\alpha^2}{S}} d\alpha \right) \end{aligned} \quad (2.51)$$

We can now plot the individual contributions of $I(x_I; y_I)$, $H(\eta)$ and $H(\eta | x_I)$ to see how they influence the final results for $I(x_I; z_I)$ and especially how they behave for high values of SNR.

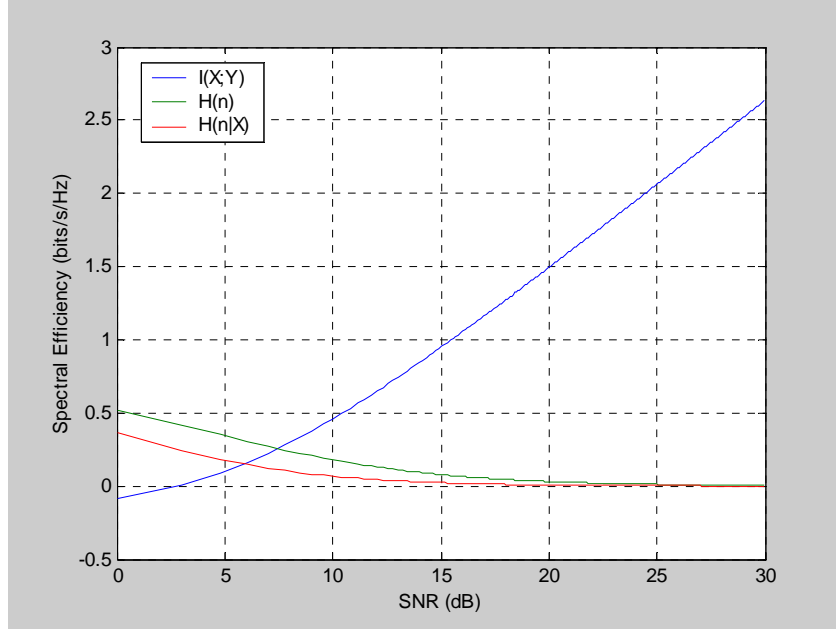


Fig. 2.8 Contributions of each part of $I(x_I; z_I)$ for Rayleigh input

2.8.3. Gaussian input

An interesting but simple case to test the Markov chain method is the Gaussian distribution as the input.

$I(x_I; y_I)$ is well known as the Shannon bound and given by

$$I(X; Y) = \frac{1}{2} \log(SNR + 1) \quad (2.52)$$

And $I(x_I; z_I)$ is given by

$$I(x_I; z_I) = \frac{1}{2} \log(SNR + 1) + \log \frac{1}{2} - (f_1 + f_2) \quad (2.53)$$

where f_1 and f_2 are

$$f_1 = \frac{1}{\sqrt{\pi SNR}} \int_0^{\infty} [1 + \operatorname{erf}(\alpha)] \log \left[\frac{1}{2} (1 + \operatorname{erf}(\alpha)) \right] e^{-\frac{\alpha^2}{SNR}} d\alpha \quad (2.54)$$

$$f_2 = \frac{1}{\sqrt{\pi SNR}} \int_0^{\infty} [1 - \text{erf}(\alpha)] \log \left[\frac{1}{2} (1 - \text{erf}(\alpha)) \right] e^{-\frac{\alpha^2}{SNR}} d\alpha \quad (2.55)$$

with $\alpha = x/\sqrt{2}$

(2.53) is the same result as $I(X;Y)$ when the input distribution is half-Gaussian. When we evaluate both expressions numerically we obtain the same results as it can be seen in Figure 2.9

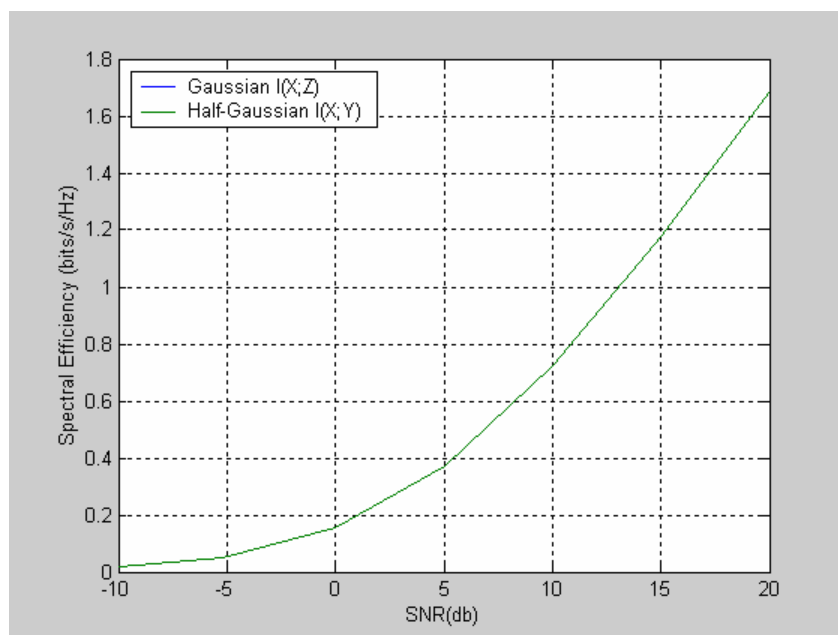


Fig. 2.9 Comparison of $I(X;Z)$ with Gaussian input and $I(X;Y)$ with half-Gaussian input

2.8.4. Exponential input

One of the main reasons we needed a technique like the Markov chain method was for input distributions like the exponential where the direct definition method became very complicated due to the nature of the distribution of z , $f_z(z)$. In this section we use the Markov chain method to obtain a high-SNR approximation for $I(x_1; z_1)$ and also a more general expression. The methodology is straight-forward and follows the same steps we developed in Section 3.3.1 and have used for other distributions so far.

We obtain

$$\begin{aligned}
 I(x_1; z_1) = & -\frac{2}{SNR} + \frac{1}{2} + \frac{1}{2} \log\left(\frac{SNR}{\pi}\right) - \text{Int}_2(SNR) + \\
 & + \frac{e^{-\frac{1}{SNR}}}{\sqrt{SNR}} \left[I_1(SNR) \log\left(\frac{e^{-\frac{1}{SNR}}}{\sqrt{SNR}} I_1(SNR)\right) + I_2(SNR) \log\left(\frac{e^{-\frac{1}{SNR}}}{\sqrt{SNR}} I_2(SNR)\right) \right] + \\
 & + H(\eta | x_1)
 \end{aligned} \tag{2.56}$$

where

$$I_1(SNR) = \int_{-\infty}^{-\sqrt{2/SNR}} e^{-\frac{2\alpha}{\sqrt{SNR}}} [1 + \text{erf}(\alpha)] d\alpha \tag{2.57}$$

$$I_2(SNR) = \int_{-\sqrt{2/SNR}}^{\infty} e^{-\frac{2\alpha}{\sqrt{SNR}}} [1 + \text{erf}(\alpha)] d\alpha \tag{2.58}$$

And $H(\eta | x_1)$ is calculated as

$$H(\eta | x_1) = -(f_1 + f_2)$$

where

$$f_1 = \int_0^{\infty} \frac{1}{2} \left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \log\left(\frac{1}{2} \left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \right) \cdot \sqrt{\frac{2}{S}} e^{-\sqrt{2/S}x} dx \tag{2.59}$$

$$f_2 = \int_0^{\infty} \left(1 - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \right) \log \left(\left(1 - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \right) \right) \sqrt{\frac{2}{S}} e^{-\sqrt{\frac{2}{S}} x} dx \quad (2.60)$$

Here the integral $Int_2(SNR)$ was derived in the previous section in this chapter and is given by:

$$Int_2(SNR) = \frac{e^{-\frac{1}{SNR}}}{\sqrt{SNR}} \int_{-\infty}^{\infty} e^{-\frac{2x}{\sqrt{SNR}}} [1 + \operatorname{erf}(x)] \log[1 + \operatorname{erf}(x)] dx \quad (2.61)$$

For high values of SNR the expression above goes to the value of $\log 2$.

Thus, we can find an approximation for $I(x_I; z_I)$ for high values of SNR

$$\boxed{I(x_I; z_I) = \frac{1}{2} \log \left(\frac{SNR \cdot \exp(1)}{4\pi} \right) \approx \frac{1}{2} \log \left(\frac{SNR}{4\pi} \right)} \quad (2.62)$$

which is the same approximation we obtain for $I(x_I; y_I)$ for high values of SNR.

In the following figure we can see how each individual term contributes to $I(x_I; z_I)$. As can be seen, for high values of SNR $H(\eta)$ and $H(\eta | x_I)$ go to zero giving us only the contribution of $I(x_I; y_I)$.

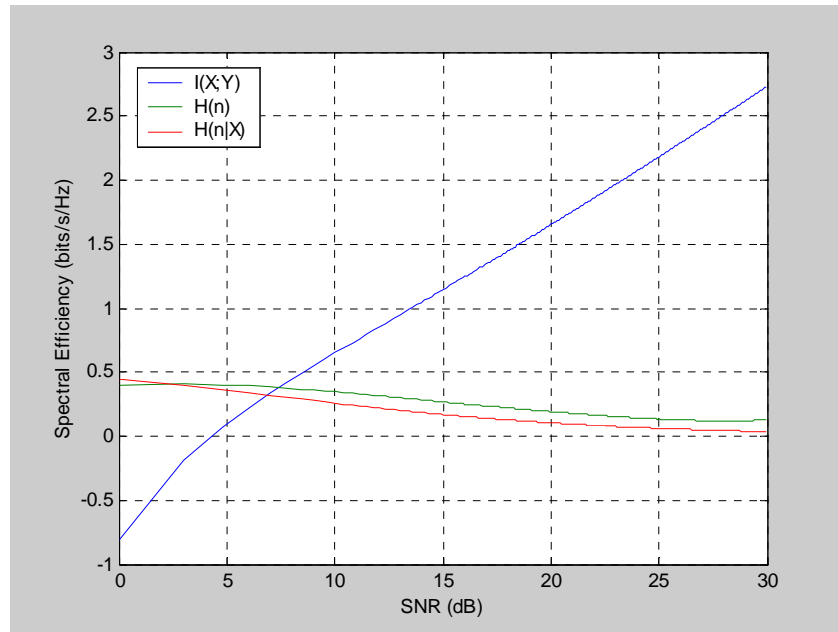


Fig. 2.10 Contributions of each part of $I(x_t; z_t)$ for exponential input

2.8.5. Summary of results for $I(x_I; z_I)$

Figure 2.11 compares the different expressions for $I(x_I; z_I)$ we obtained for the various distributions we worked with in this chapter. As expected the Gaussian distribution gives us the closest result to the Shannon bound.

Also, for distributions which only exist for $x > 0$, the half-Gaussian distribution has the best results. For high values of SNR the spectral efficiencies for all distributions tend to approach that obtained with the Gaussian distribution.

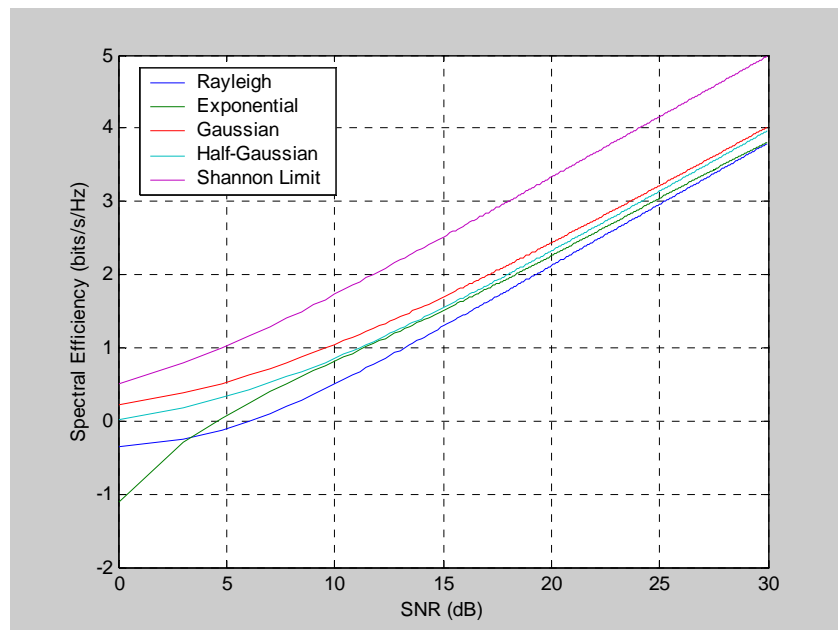


Fig. 2.11 Comparison of $I(x_I; z_I)$ obtained via the Markov chain method for various distributions

Figures 2.11 to 2.14 compare the results obtained for $I(x_I; y_I)$ and $I(x_I; z_I)$ for each distribution considered, showing how taking the absolute value of the output reduces the mutual information.

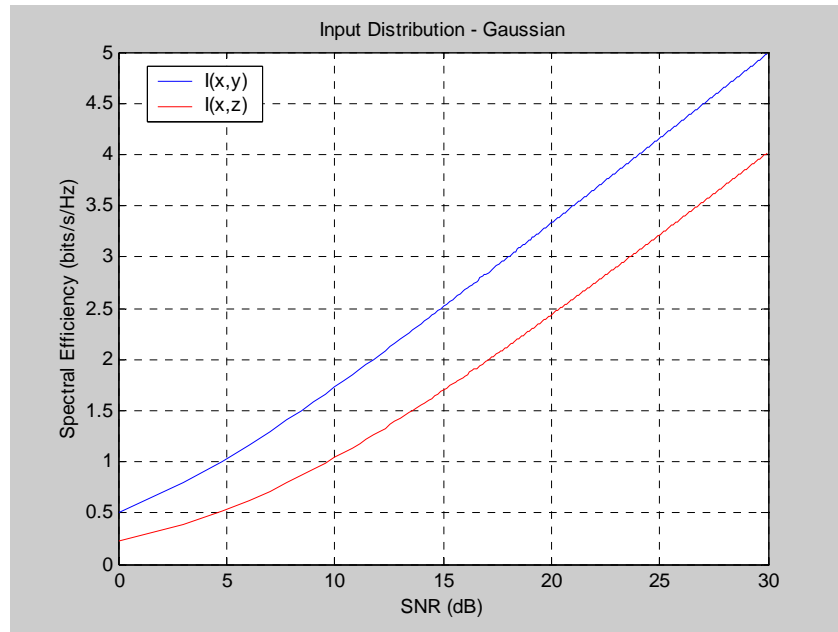


Fig. 2.12 Comparison of $I(x_I; y_I)$ and $I(x_I; z_I)$ with Gaussian input distribution

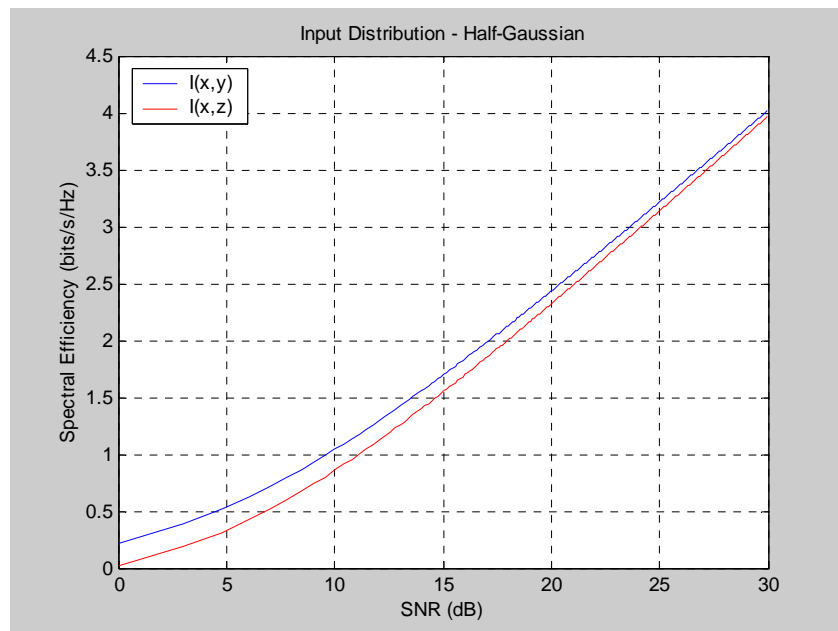


Fig. 2.13 Comparison of $I(x_I; y_I)$ and $I(x_I; z_I)$ with half-Gaussian input distribution

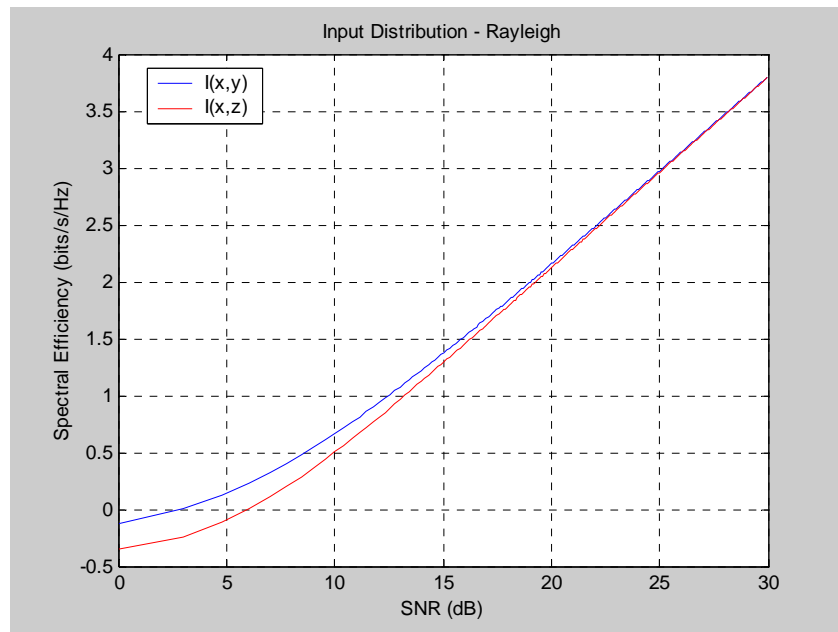


Fig. 2.14 Comparison of $I(x_I; y_I)$ and $I(x_I; z_I)$ with Rayleigh input distribution

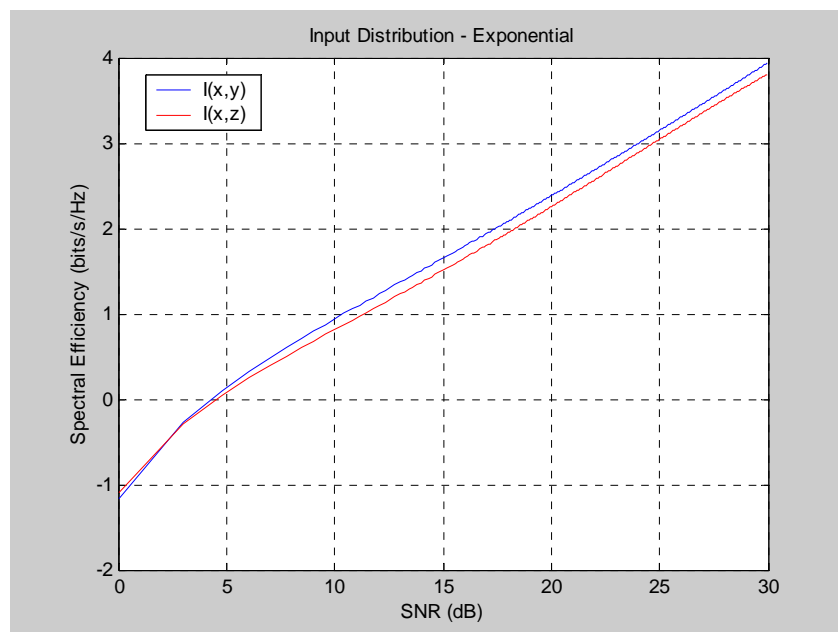


Fig. 2.15 Comparison of $I(x_I; y_I)$ and $I(x_I; z_I)$ with exponential input distribution

In Figure 2.11 we see that there is a slight difference in the numerical results between the half-Gaussian and the Gaussian input for $I(X;Z)$. We believe that these come due to inaccuracies when numerically calculating the integral in (2.47).

We have shown previously by direct evaluation that Z is half-Gaussian distributed if X is half-Gaussian. But Z is also half-Gaussian distributed if X is Gaussian. This follows from the fact that if X is zero mean Gaussian, Y is also zero mean Gaussian, and the absolute value of a zero-mean Gaussian random variable is half-Gaussian distributed. Thus, the distribution of Z is the same for X Gaussian and half-Gaussian. (This follows from the fact that Z is half-Gaussian in both cases with the same $E\{Z^2\} = S+N$.) Consequently, $H(Z)$ is the same for the Gaussian and half-Gaussian cases.

From symmetry it follows that $H(Z|X = x)$ is an even function of x ; i.e.,

$$H(Z|X = x) = H(Z|X = -x) \quad (2.63)$$

Therefore,

$$H(Z|X) = \int_{-\infty}^{\infty} H(Z|X = x) f_X(x) dx = \int_0^{\infty} H(Z|X = x) [f_X(x) + f_X(-x)] dx \quad (2.64)$$

For X Gaussian, this becomes

$$H(Z|X) = 2 \int_0^{\infty} H(Z|X = x) f_G(x) dx = \int_0^{\infty} H(Z|X = x) f_{hG}(x) dx \quad (2.65)$$

Therefore, $H(Z|X)$ is also the same for X Gaussian and half-Gaussian, from which it follows that

$$I(X_G; Z) = I(X_{hG}; Z) \quad (2.66)$$

contrary to the numerical results we obtained.

Note that by replacing $H(Z|X = x)$ by $f(z|X = x)$ in (2.64), the same argument may be used to show that the pdf of Z is the same for X half-Gaussian and Gaussian, thereby more simply showing that Z is half-Gaussian when X is half-Gaussian. Also, the same argument may be used to show that (2.66) also applies in the 2 and 4 DOF cases.

2.9. Channel capacity results for low SNR values for a half-Gaussian input

After some simplifications and substitutions in the integral in (2.27) we have been able to extend our spectral efficiency numerical results for low SNR values for a half-Gaussian input. The motivation to work with low SNR values is to obtain the minimum number of photons per bit.

The numerical results obtained for low SNR values are shown in Figure 2.16.

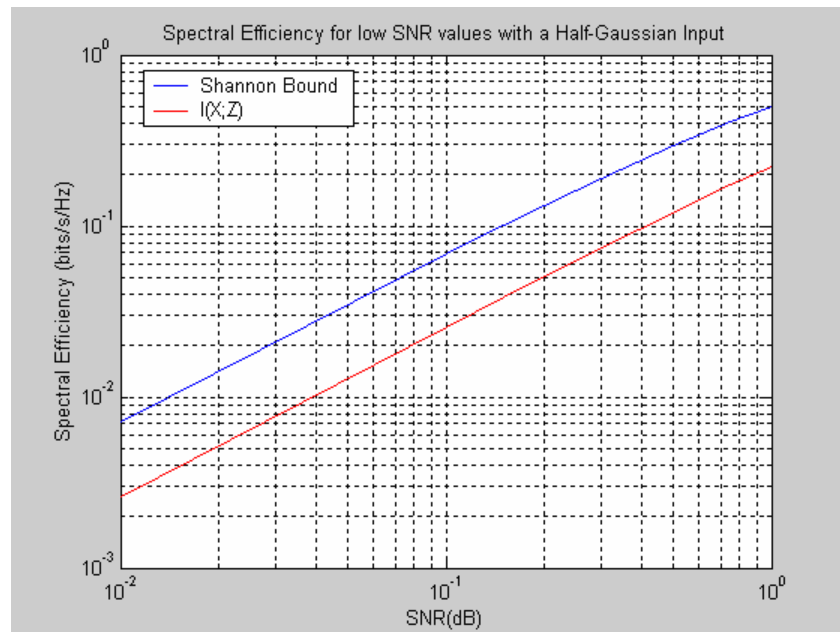


Fig. 2.16 Spectral efficiency for 1 DOF at low SNR values for a half-Gaussian input

The fundamental noise limitation in optical systems is quantum noise associated with the discrete nature of the photon. In the case of coherent systems employing a strong local oscillator in the receiver, the dominant noise is the shot noise introduced by the local oscillator. For a sufficiently strong local oscillator (large number of local oscillator photons in the detection interval) this noise is essentially Gaussian with a noise power per DOF given by

$$N = h\nu W / 2 \quad (2.67)$$

where $h\nu$ is the photon energy. Since the number of photons per bit, N_p , is energy-per-bit divided by the photon energy, it follows that

$$N_p = \frac{S}{R_b h\nu} = \frac{S}{N} \frac{2W}{R_b} \quad (2.68)$$

where W is the channel bandwidth and R_b is the bit rate. This argument is also valid for an incoherent channel where the signal-to-noise ratio is still S/N and the capacity is upper-bounded by the capacity of the coherent channel.

Thus using (2.68), the spectral efficiency results we obtained can be used to calculate the minimum number of photons per bit, which is given by [12]

$$N_p(SNR) = \frac{SNR}{2I(X;Z)} \quad (2.69)$$

Figure 2.17 has the plots of the minimum number of photons per bit as a function of spectral efficiency.

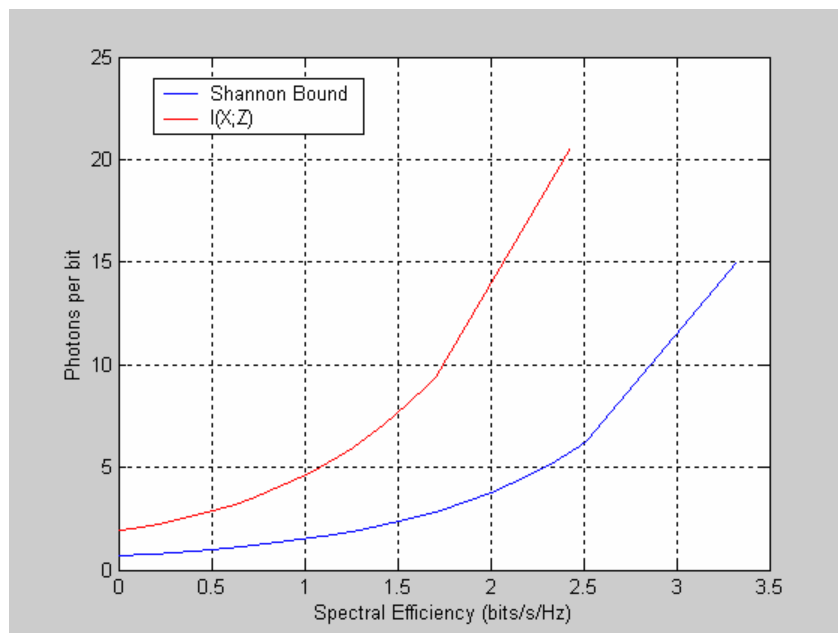


Fig. 2.17 Comparison of power efficiency for 1 DOF coherent channel and the incoherent channel

As SNR goes to zero, the minimum number of photons per bit approaches 0.69 for the $Y=X+N$ channel and for the $Z=X+N$ it approaches 1.9.

Chapter 3

Incoherent Detection with 2 and 4 Degrees of Freedom

In the previous chapter we worked with only 1 degree of freedom which corresponded to coherent receivers or when the input was non-negative, to PIN receivers. In this chapter we extend our work to 2 and 4 degrees of freedom, corresponding to incoherent receivers which we described briefly in our introduction to different receiver models in Chapter 1.

Work done by Jacobs in [9] has bounds for the high SNR and low SNR limits for the 2 DOFs case. In this chapter we review those bounds as well as the methodology used to find them. We then continue by showing numerical results we have obtained for any value of SNR and the methodology used to find these numerical results and from there we continue with the 4 DOFs case.

In Chapter 1 we made a statement that for high SNR the bound for channel capacity would be the same independent of the number of DOFs. This statement will be proved in this chapter.

3.1. Existing bounds for 2 degrees of freedom

In this section we describe the methodology used by Jacobs in finding upper bounds for incoherent receivers for single polarization (or 2 DOFs). For 2 DOFs, the relation between the input signal, the output signal and the noise components is given by

$$z_2^2 = (x_1 + n_1)^2 + n_2^2 \quad (3.1)$$

In such a system we assume that the transmitter is free to code only the amplitude of the input signal, i.e., $A=|x_1|$ and the receiver is free to measure only $R=|z_2|$.

To find the channel capacity the mutual information between the input A and the output R has to be calculated. By definition the mutual information is given by

$$I(A, R) = \int_0^{\infty} dA \int_0^{\infty} dR p(A) p(R | A) \log \left[\frac{p(R | A)}{\int_0^{\infty} dV p(V) p(R | V)} \right] \quad (3.2)$$

where the transition probability is the Rician distribution and given by

$$p(R | A) = \frac{R}{N} \exp\left(-\frac{R^2 + A^2}{2N}\right) I_0\left(\frac{AR}{N}\right) \quad (3.3)$$

where I_0 is the modified Bessel function.

The capacity by definition is obtained when we maximize $I(A, R)$ with respect to the input distribution $p(A)$. The input distribution is subject to the following constraints

$$\int_0^{\infty} dA p(A) = 1 \quad (3.4)$$

Normalization

$$\int_0^{\infty} dA A^2 p(A) = S \quad (3.5)$$

Average power

Although Jacobs in [9] was unable to obtain a general expression for channel capacity analytically, he obtained upper bounds, and furthermore showed that in the high and low SNR limits these bounds were achieved by particular choices of input distribution.

The upper bound is given by

$$C \leq \frac{1}{2} \log_2(1 + SNR) \quad (3.6)$$

Jacobs also showed that the Rayleigh distribution is able to achieve this bound in the high SNR limit.

In the low SNR limit ‘well-behaved’ functions can not be used to achieve this upper bound, but a highly singular amplitude distribution was shown to achieve this bound in the low SNR case. The reason a highly singular distribution was used at low SNR is because for an incoherent channel it is necessary to highly concentrate the signal energy to achieve satisfactory error performance.

3.2. Numerical calculation of mutual information for 2 degrees of freedom with a half-Gaussian input

Since Z is non-negative there is a 1-to-1 correspondence between Z and Z^2 , and it then follows that

$$I(X; Z) = I(X; Z^2) \quad (3.7)$$

Direct evaluation of $I(X; Z^2)$ is easier than that of $I(X; Z)$. The relationship between Z^2 , the input signal and the various noise sources is thus given by

$$Z^2 = (X + N_1)^2 + \sum_{i=2}^m N_i^2 \quad (3.8)$$

From this expression we can obtain the pdf of Z^2 using convolution or Laplace transforms. Also, the pdf of $Z^2/X=x$ is non-central chi-square and we have general forms of it for the cases of interest ($m=2$ and $m=4$).

Although we are still faced with double integrals, the integrands are simpler when compared to those used to calculate $I(X; Z)$ directly.

To calculate the channel capacity we start with the basic formula:

$$I(X; Z^2) = H(Z^2) - H(Z^2 | X) \quad (3.9)$$

3.2.1. Calculation of $H(Z^2)$ for a half-Gaussian input

$$Z^2 = \underbrace{(X + N_1)^2}_{f_1} + \underbrace{N_2^2}_{f_2} \quad (3.10)$$

The distributions f_1 and f_2 are given by:

$$f_1(x) = \frac{1}{\sqrt{\pi k_1 x}} \exp\left(-\frac{x}{k_1}\right) \quad (3.11)$$

, where: $k_1 = 2(S + 1)$. We have assumed here and throughout this dissertation that $N=1$ so we essentially have $S=SNR$.

$$f_2(x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{x}{2}\right) \quad (3.12)$$

To obtain $f_{z^2}(z)$ using the distributions (3.11) and (3.12), we need to calculate the following convolution:

$$f_{z^2}(z) = f_1(x) * f_2(x) \quad (3.13)$$

As an alternative to the convolution we can use Laplace transforms:

$$F_{z^2}(s) = \ell\{f_{z^2}(z)\} = \ell\{f_1(x)\} \cdot \ell\{f_2(x)\} \quad (3.14)$$

These transforms are given by

$$\ell\{f_1(x)\} = \frac{1}{k_1} \left(\frac{\sqrt{k_1}}{\sqrt{s + 1/k_1}} \right) \quad (3.15)$$

$$\ell\{f_2(x)\} = \frac{1}{2} \left(\frac{\sqrt{2}}{\sqrt{s + 1/2}} \right) \quad (3.16)$$

Thus:

$$f_{z^2}(z) = \ell^{-1}\{F_{z^2}(s)\} = \frac{1}{2\sqrt{S+1}} \left[\exp\left(-\frac{z}{4} \left(\frac{S+2}{S+1}\right)\right) I_0\left(\frac{zS}{4(S+1)}\right) \right] \quad (3.17)$$

With this distribution we can then obtain the entropy of Z^2 :

$$H(Z^2) = -\int_0^{\infty} f_{z^2}(z) \log(f_{z^2}(z)) dz \quad (3.18)$$

which can be evaluated numerically without any problems using any mathematical software package.

The plot of the entropy $H(Z^2)$ is shown in Figure 3.1 and compared to the entropy of Z when we have $m=1$:

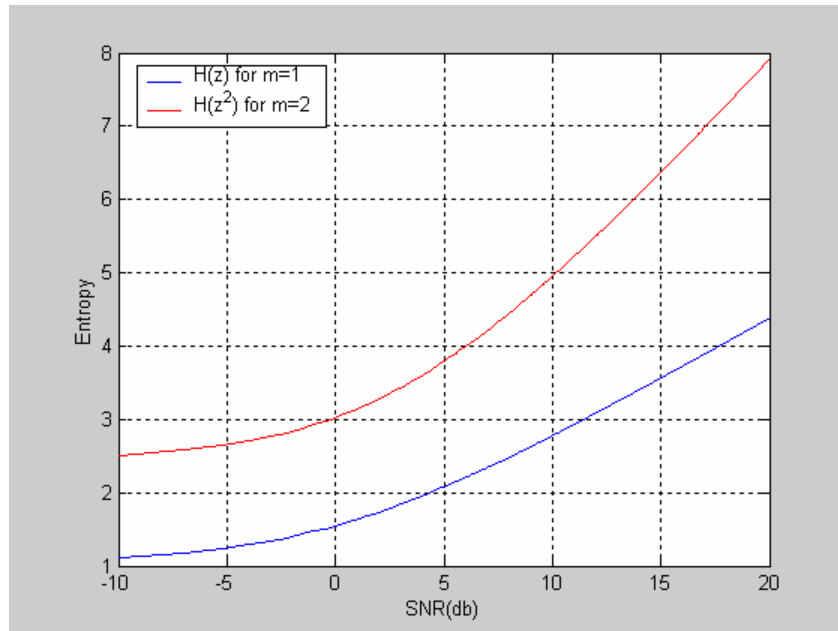


Fig. 3.1 Entropy of Z^2 for a half-Gaussian input

3.2.2. Calculation of $H(Z^2|X)$ for a half-Gaussian input

We start out with the basic formula for calculating $H(Z^2 | X)$ which is similar to how we calculate $H(Z | X)$ as explained in Chapter 2.

$$H(Z^2 | X) = -\int_0^{\infty} H(Z^2 | X = x) f_x(x) dx \quad (3.19)$$

Here the input distribution $f_x(x)$ is half-Gaussian, i.e.

$$f_x(x) = \sqrt{\frac{2}{\pi S}} \exp\left(-\frac{x^2}{2S}\right) \quad (3.20)$$

In order to evaluate (3.19) we need to calculate $H(Z^2 | X = x)$, which is given by

$$H(Z^2 | X = x) = -\int_0^{\infty} f_{Z^2|X=x} \log(f_{Z^2|X=x}) dz \quad (3.21)$$

We know that the conditional pdf $f_{Z^2|X=x}$ is a non-central chi-square distribution function and to use its direct definition would make the integral in (3.21) extremely difficult to be calculated numerically.

For the $m=2$ case from work done by Jacobs in [9] we know that the conditional distribution $f_{Z|X=x}$ is Rician. We can relate the distribution of $Z|X=x$ to the distribution of $Z^2|X=x$, by the following relation [13]:

$$f_{z^2|X=x} = \frac{1}{2z} f_{Z|X=x}(z) \quad (3.22)$$

Thus we obtain

$$f_{z^2|X=x} = \frac{1}{2} \exp\left(-\frac{(x^2 + z)}{2}\right) I_0(x\sqrt{z}) \quad (3.23)$$

Using (3.17), (3.18), (3.21) and (3.23) we are able to find the mutual information for a half-Gaussian input when we have 2 degrees of freedom. These numerical results are shown later in Figures 3.2 and 3.3 where we compare them with the results also obtained numerically for 4 degrees of freedom (Section 3.4) and the numerical results for 1 degree of freedom calculated in Chapter 2.

3.3. Finding the transition probability for two polarizations or 4 degrees of freedom

When dealing with two polarizations in an incoherent channel, the main difference becomes the increased number of DOFs from 2 to 4. To obtain upper bounds or to find an analytical expression for the channel capacity in this case we have to follow the same methodology outlined in the previous section. The relation between the input signal, the output signal and the noise components is given by

$$z_4^2 = \underbrace{(x_1 + n_1)^2 + n_2^2}_{R_1^2} + \underbrace{n_3^2 + n_4^2}_{r^2} \quad (3.24)$$

where n_k has a Gaussian distribution. In a similar fashion to section 3.1, the transmitter is free to code only the amplitude of the input signal $A=|x_1|$ and the receiver is free to measure only $R=|z_4|$.

To find the mutual information between the input and the output we have to initially find the transition probability which is not Rician in this case.

The distribution of R^2 is given by the following convolution operation:

$$f_{R^2} = f_{R_1^2} * f_{r^2} \quad (3.25)$$

From the single polarization case (2 DOFs case) we can get the expression for $f(R_1)$:

$$f_{R_1}(\alpha) = \frac{\alpha}{N} \exp\left(-\frac{\alpha^2 + A^2}{2N}\right) I_0\left(\frac{A\alpha}{N}\right) \quad (3.26)$$

which is the Rician distribution and α is non-negative.

If we have two random variables, X and Y , with pdfs $f_X(x)$ and $f_Y(y)$, and their relation is that $Y=X^2$, then their pdfs are also related by [13] for $X>0$

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) \quad (3.27)$$

Thus we can find the pdf of R_1^2 by using (3.26) and (3.27)

$$f_{R_1^2}(\alpha) = \frac{1}{2N} \exp\left(\frac{-\alpha^2 - A^2}{2N}\right) I_0\left(\frac{A\alpha}{N}\right) \quad (3.28)$$

and the distribution of r^2 is given by

$$f_{r^2}(\beta) = \frac{1}{2N} \exp\left(\frac{-\beta}{2N}\right) \quad (3.29)$$

The distribution above is an exponential distribution.

By convolving (3.28) and (3.29) we obtain

$$f_{R^2}(\gamma) = \frac{\gamma}{2AN} e^{\frac{-\gamma^2 - A^2}{2N}} I_1\left(\frac{A\gamma}{N}\right) \quad (3.30)$$

and thus the transition probability is given by

$$\boxed{f_R = p(R | A) = \frac{R^2}{AN} e^{\frac{-R^2 - A^2}{2N}} I_1\left(\frac{AR}{N}\right)} \quad (3.31)$$

which is a valid pdf.

3.4. Numerical calculation of channel capacity for 4 degrees of freedom with a half-Gaussian input

In this section we extend the work we did in section 3.2 for the 2 degrees of freedom case to 4 degrees of freedom and then compare these numerical results with each other as well as with the 1 degree of freedom case and the Shannon bound. The methodology is identical to the 2 degrees case but the expressions because of the extra noise terms are different.

Because of the extra number of degrees of freedom we have 2 more noise terms in the Z^2 expression:

$$Z^2 = \underbrace{(X + N_1)^2}_{f_1} + \underbrace{N_2^2}_{f_2} + \underbrace{N_3^2}_{f_3} + \underbrace{N_4^2}_{f_4} \quad (3.32)$$

The distribution f_1 is given by and (3.11) and the noise terms f_2, f_3, f_4 have the same distribution

$$f_2(x) = f_3(x) = f_4(x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{x}{2}\right) \quad (3.33)$$

Using the Laplace transform we find the distribution of z^2 to be

$$f_{z^2}(z) = \ell^{-1}\{F_{z^2}(s)\} = \frac{1}{4\sqrt{S+1}} \left[\exp\left(-\frac{z}{4}\left(\frac{S+2}{S+1}\right)\right) \left[I_0\left(\frac{zS}{4(S+1)}\right) - I_1\left(\frac{zS}{4(S+1)}\right) \right] \right] \quad (3.34)$$

where I_0 and I_1 are modified Bessel functions of the zero and first order.

And to calculate $H(Z^2 | X = x)$ and subsequently $H(Z^2 | X)$ we use the transition probability for Z/X found in (3.31) to obtain

$$f_{z^2|X=x} = \frac{\sqrt{z}}{2x} \exp\left(-\frac{(x^2+z)}{2}\right) I_1(x\sqrt{z}) \quad (3.35)$$

Using the distributions (3.34) and (3.35) and the method outlined in section 3.2 we are also able to numerically evaluate the mutual information for 4 degrees of freedom. These numerical results are shown in Figures 3.2 and 3.3 and are compared to the results we obtained previously for 1 DOF in Chapter 2 and 2 DOFs obtained in section 3.2.

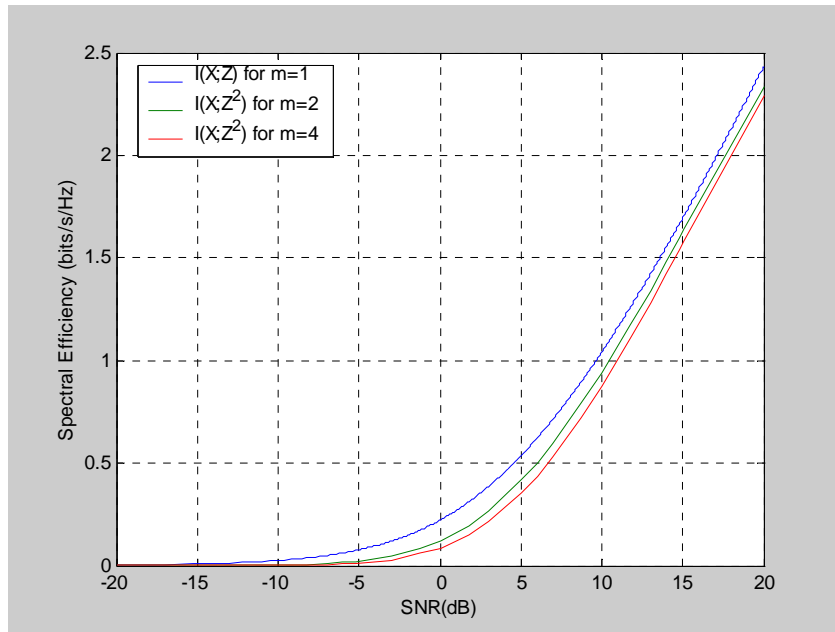


Fig. 3.2 Mutual information calculated numerically for a half-Gaussian input for various values of m
– linear scale

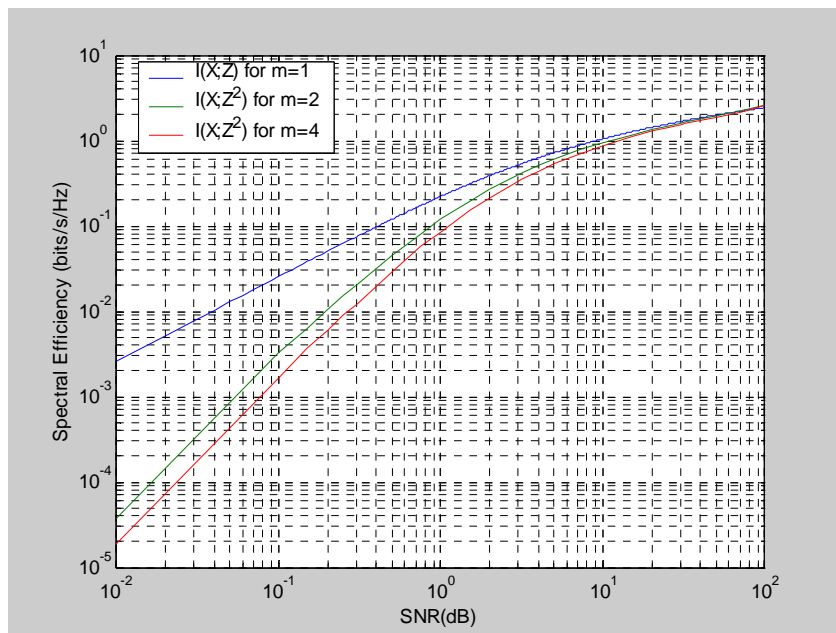


Fig. 3.3 Mutual information calculated numerically for a half-Gaussian input for various values of m
– logarithmic scale

For low values of SNR we can see that the spectral efficiency for the $m=1$ case, the slope is proportional to the SNR, while for the $m=2$ and $m=4$ the slope changes with the square of SNR.

3.5. High m approximation

As the intention of this chapter is to determine the effect of degrees-of-freedom on the channel capacity results, an interesting special case would be that of large m for which $Z^2/X=x$ and Z^2 should be approximately Gaussian for which the entropies are easily obtained. In this section we obtain a closed form expression for $I(X;Z^2)$ in the limit of large m and compare this expression with numerical results that we obtained previously for the cases we are interested in.

3.5.1. Calculation of $H(Z^2)$

We start out with the relation between Z, X and the noise terms previously shown in (3.8):

$$Z^2 = \underbrace{(X + N_1)^2}_{f_1} + \underbrace{\sum_{i=2}^m N_i^2}_{\sum_{i=2}^m f_i} \quad (3.36)$$

By the central limit theorem the pdf of Z^2 will be approximately Gaussian with mean $\mu_{z^2} = \mu_1 + \dots + \mu_m$ and variance $\sigma_{z^2}^2 = \sigma_1^2 + \dots + \sigma_m^2$.

The pdf of each of the squared noise terms (f_2, \dots, f_m) is given by

$$f_{N_i^2}(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, \quad y > 0 \quad (3.37)$$

from which $\mu_2 = \mu_3 = \mu_m = 1$ and $\sigma_2^2 = \sigma_3^2 = \sigma_m^2 = 2$.

And from previous results (Chapter 2), the pdf of f_1 is:

$$f_{f_1}(y) = \frac{1}{\sqrt{2\pi y(S+1)}} e^{-\frac{y}{2(S+1)}}, \quad y > 0 \quad (3.38)$$

for which $\mu_1 = (S+1)$ and $\sigma_1^2 = 2(S+1)^2$.

Thus the mean and variance of Z^2 are given by

$$\mu_{Z^2} = (S + 1) + (m - 1) \quad (3.39)$$

$$\sigma_{Z^2}^2 = 2(S + 1)^2 + 2(m - 1) \quad (3.40)$$

3.5.2. Calculation of $H(Z^2|X=x)$

The pdf of $Z^2|X=x$ is non-central chi-square, but for high m it approaches also a Gaussian distribution. The mean and variance of this Gaussian distribution is given by:

$$\mu_{Z^2|X=x} = m + S \quad (3.41)$$

$$\sigma_{Z^2|X=x}^2 = 2(m + 2S) \quad (3.42)$$

3.5.3. Channel capacity approximation

Using the results in sub-sections 3.5.1 and 3.5.2 we are able to obtain a final expression for channel capacity for high values of m :

$$I(X; Z^2) = \frac{1}{2} \log \left(\frac{S^2 + 2S + m}{2S + m} \right), m \gg 1 \quad (3.43)$$

In Figure 3.4 we see how m influences the spectral efficiency, and independent of the number of degrees of freedom, for high SNR all curves approach a common slope.

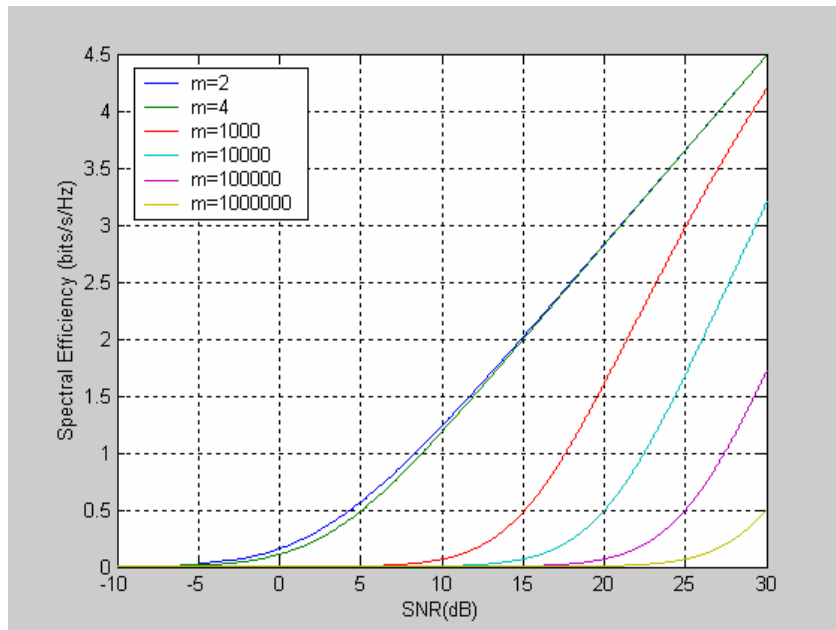


Fig 3.4 High m approximation – for various values of m

In Figure 3.5 we compare the high m approximation we obtained for the case of $m=2$ with the results we obtained numerically in section 3.2 for this case. As it can be seen, the high m approximation when evaluated at $m=2$ overestimates the spectral efficiency.

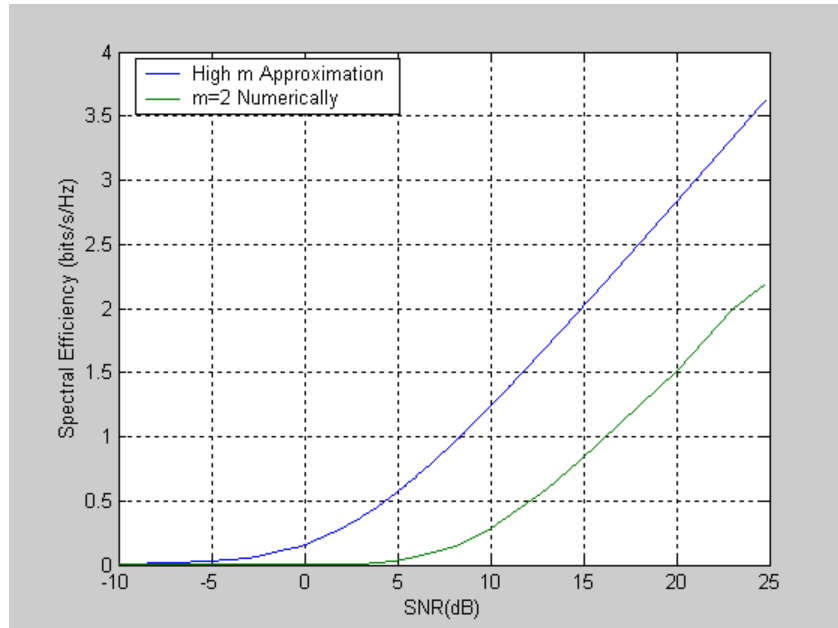


Fig 3.5 Comparison of high m approximation with numerical results for $m=2$

3.6. Summary of results obtained with a half-Gaussian input

Figures 3.6 and 3.7 show the various results we have obtained in this chapter. As it is seen for high values of SNR, the mutual information obtained numerically for 2 and 4 degrees of freedom approaches the same value. Similar results are also seen when we use the high m approximation for these two cases. In section 3.7 we will try to justify this analytically.

Also as expected the results for 1 degree of freedom are higher than those for 2 and 4 degrees. And as mentioned previously, the high m approximation does not give us accurate results but can be used to give us an upper limit for the channel capacity.

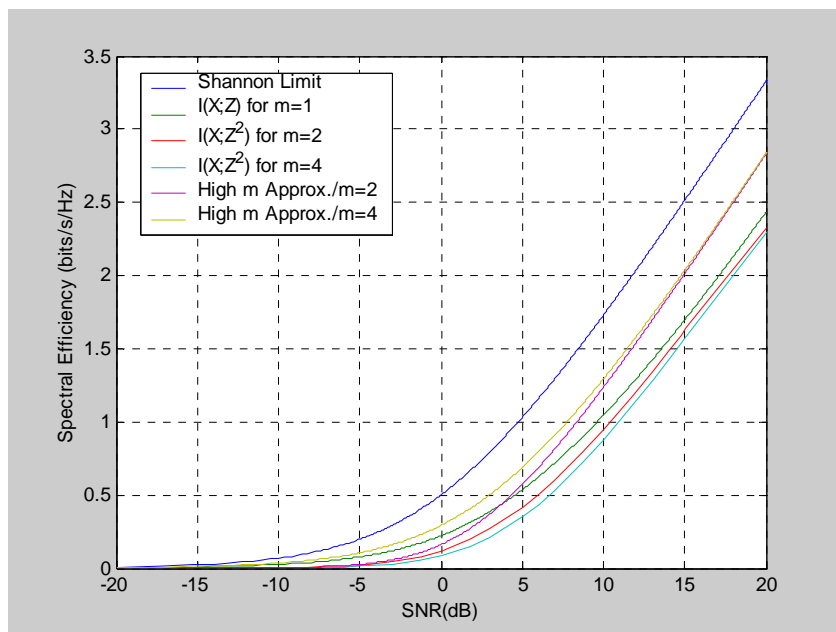


Fig 3.6 Linear plot of various results obtained in chapter 3 for different values of m

Figure 3.7 shows us the mutual information on a logarithmic scale which shows clearly the behavior for low values of SNR. As noted previously, for 1 degree of freedom the mutual information is proportional to the SNR while for 2 and 4 degrees of freedom it slopes as the square of the SNR for a half-Gaussian input. As will be shown later in Section 3.10 when a modified half-Gaussian is used the slope also changes linearly for 2 and 4 degrees of freedom.

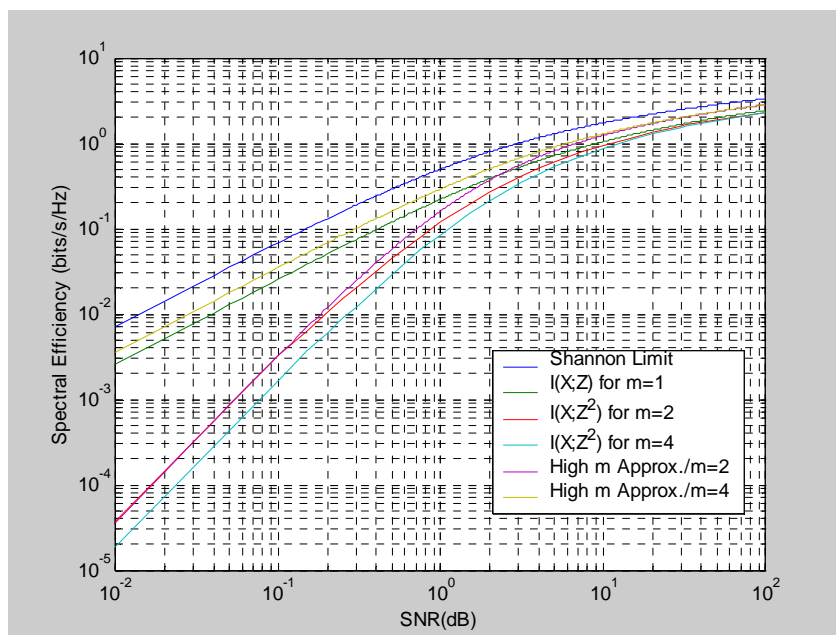


Fig 3.7 Logarithmic plot showing lower SNR levels more clearly for various m

Based on the results we have obtained numerically shown in Figures 3.6 and 3.7 we believe we can obtain a compact equation for spectral efficiency which behaves in a similar way to the Shannon capacity formula but with the SNR scaled by a constant.

For the incoherent case we expect the spectral efficiency to have the following form for high SNR:

$$C_{im} = \frac{1}{2} \log(1 + kSNR) \quad (3.44)$$

As we already have the numerical results for C_{im} , we can call the difference between the Shannon capacity and C_{im} , a Δ (the numerical results have been divided by $\log 2$ to obtain channel capacity in bits/s/Hz)

$$\Delta = \frac{1}{2} \log(1 + SNR) - \frac{1}{2} \log(1 + kSNR) \quad (3.45)$$

We can then isolate k in terms of variables which we already know:

$$k = \frac{\left(\left(\frac{1 + SNR}{\exp(2\Delta \log 2)} \right) - 1 \right)}{SNR} \quad (3.46)$$

In figure 3.8 we plot $1/k$ for the incoherent cases $m=1, 2$ and 4 versus SNR and see how it behaves for low and high values of SNR.

For $m=1, 2$ and 4 , $1/k$ approaches 4 for high SNR and we can thus use a compact linear expression to approximate the channel capacity.

For low SNR values however, we can see that for $m=2$ and $m=4$ we can not approximate the mutual information expression by a linear approximation and that indeed for a half-Gaussian input the channel capacity slope changes quadratically with the SNR unlike the $m=1$ case where it changes in a linear fashion. This corresponds to $1/k$ increasing rapidly as SNR goes to zero as seen in Figure 3.8.

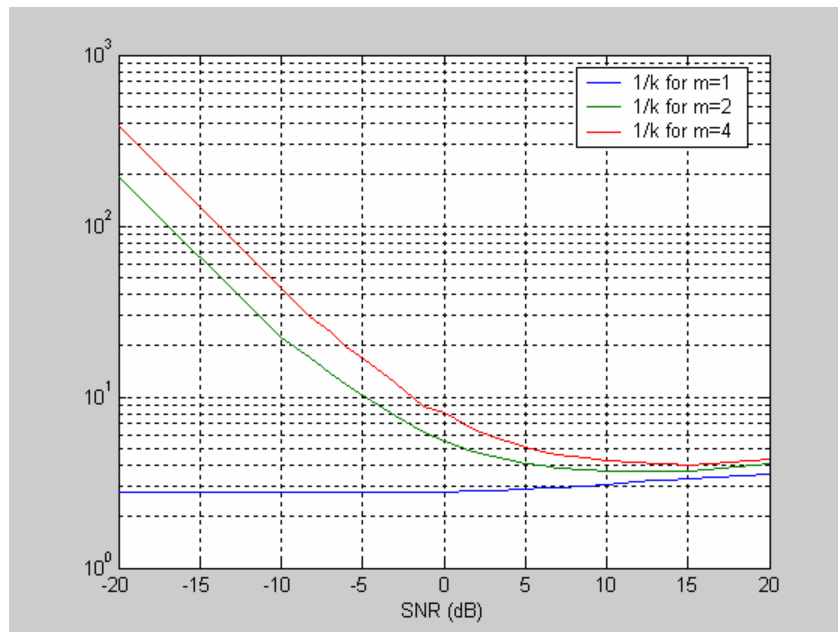


Fig 3.8 Calculation of constant to approximate mutual information with a simple expression for various values of m

3.7. Analytical proof that the transition probability for 2 and 4 DOFs is the same for high SNR

In chapter 1 we mentioned that for both the two and four DOFs cases we obtain identical upper bounds for the channel capacity for high SNR values. One way to prove that is to show that the transition probability $p(R/A)$ for 4 DOFs (2 polarizations) is the same as the $p(R/A)$ for the 2 DOFs case (single polarization) when $A \rightarrow \infty$ (or high values of SNR).

From [14] we have the following definition for the modified Bessel function of the ν -th order:

$$I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} \left(1 - \frac{u-1}{8z} + O\left(\frac{1}{z^2}\right) \right), \text{ with } u = 4\nu^2 \quad (3.47)$$

and the transition probabilities for the 2 and 4 DOFs cases are given by

$$p_2 = \frac{R}{N} \exp\left(\frac{-R^2 - A^2}{2N}\right) I_0\left(\frac{AR}{N}\right) \quad (3.48)$$

$$p_4 = \frac{R^2}{NA} \exp\left(\frac{-R^2 - A^2}{2N}\right) I_1\left(\frac{AR}{N}\right) \quad (3.49)$$

I_0 and I_1 have the same dominant asymptotic term in (3.47). So the dominant term is independent of ν . This means that for $A \gg I$:

$$p_4 \approx \frac{R}{A} p_2 \quad (3.50)$$

The proof to this statement is shown:

$$\begin{aligned}
p_4 - \frac{R}{A} p_2 &= \frac{R^2}{AN} e^{\frac{-R^2-A^2}{2N}} \left[I_1\left(\frac{AR}{N}\right) - I_0\left(\frac{AR}{N}\right) \right] \\
&= \frac{R^2}{AN} e^{\frac{-R^2-A^2}{2N}} \left[\frac{-3}{8\left(\frac{AR}{N}\right)} - \frac{1}{8\left(\frac{AR}{N}\right)} + O\left(\frac{N^2}{A^2 R^2}\right) \right] = -e^{\frac{-R^2-A^2}{2N}} \left[\frac{R}{A^2} + O\left(\frac{1}{A^3}\right) \right]
\end{aligned}
\tag{3.51}$$

And thus we obtain

$$\boxed{\therefore \lim_{A \rightarrow \infty} \left(\frac{R}{A} p_2 - p_4 \right) = 0} \tag{3.52}$$

So for large values of A we have:

$$p_4 \approx \frac{R}{A} p_2 \tag{3.53}$$

For these large values of A the significant values of R (where p_2 and p_4 are greater than zero) will be close to A , thus:

$$\frac{R}{A} \approx 1, \text{ when } A \rightarrow \infty \tag{3.54}$$

Thus:

$$\boxed{p_2 \approx p_4} \tag{3.55}$$

3.8. Numerical calculation of mutual information for 2 degrees of freedom with a modified half-Gaussian input

In this section the input distribution is taken to be a modified half-Gaussian distribution which is similar to the amplitude distribution discussed by Jacobs in [9], where it was shown that an appropriately chosen binary distribution achieves the Shannon capacity for low SNR values instead of the half-Gaussian distribution. Numerical calculations done by Ho [5] also indicate that a binary distribution is the optimum distribution in the low SNR limit.

The modified half-Gaussian distribution we use is given by

$$f_x = \varepsilon\delta(0) + (1 - \varepsilon)\sqrt{\frac{2}{\pi S}}e^{-\frac{x^2}{2S}} \quad (3.56)$$

with $0 \leq \varepsilon \leq 1$.

It should be noted that in this case the average signal power is given by $S(1 - \varepsilon)$ and during our numerical calculations we will consider this when obtaining the spectral efficiency for different values of SNR.

The motivation for considering the modified half-Gaussian is that, like the binary distribution it has a discrete component at the origin and that it enables us to utilize results already obtained. Furthermore, we speculate that choosing ε close to 1 we may obtain much better results with the modified half-Gaussian, and indeed by a proper choice of ε as a function of SNR we might achieve capacity.

3.8.1. Calculation of $H(Z^2)$ for a modified half-Gaussian input

We already have the distribution f_2 from previous work:

$$f_2(x) = \frac{1}{\sqrt{2\pi x}}e^{-\frac{x}{2}} \quad (3.57)$$

To find the distribution f_l we will first find the distribution of $Y=X+N$.

$$f_y(y) = f_x * f_n = \left((1-\varepsilon) \sqrt{\frac{2}{\pi S}} e^{-\frac{x^2}{2S}} * \frac{1}{\sqrt{2\pi}} e^{-\frac{n^2}{2}} \right) + \left(\varepsilon \delta(0) * \frac{1}{\sqrt{2\pi}} e^{-\frac{n^2}{2}} \right) \quad (3.58)$$

The results from the first convolution come from previous results when we used a half-Gaussian input, and the second convolution is solved by using impulse function and convolution properties. Thus we have:

$$f_y(y) = (1-\varepsilon) \sqrt{\frac{1}{2\pi(S+1)}} e^{-\frac{y^2}{2(S+1)}} \left[1 + \operatorname{erf} \left(\frac{y\sqrt{S}}{\sqrt{2(S+1)}} \right) \right] + \varepsilon \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \quad (3.59)$$

And for Y^2 , we have the following distribution:

$$f_2 = f_{y^2}(y) = (1-\varepsilon) \sqrt{\frac{1}{2\pi y(S+1)}} e^{-\frac{y}{2(S+1)}} + \varepsilon \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}} \quad (3.60)$$

Thus for Z^2 we have to use a convolution to obtain its distribution:

$$f_{z^2}(z) = f_1 * f_2 \quad (3.61)$$

Using the Laplace transform this becomes very easy to solve, and a big part of it has already been solved when we did these calculations for a half-Gaussian input. So we have:

$$\boxed{f_{z^2}(z) = \frac{(1-\varepsilon)}{2\sqrt{S+1}} \left[e^{-\frac{z(S+2)}{4(S+1)}} I_0 \left(z \frac{S}{4(S+1)} \right) \right] + \frac{\varepsilon}{2} e^{-\frac{z}{2}}} \quad (3.62)$$

which is a valid pdf with the second term being the only different term as compared to a half-Gaussian input.

3.8.2. Calculation of $H(Z^2|X)$ for a modified half-Gaussian input

Thus using these expressions we obtain:

$$H(Z^2 | X) = \int_0^\infty \underbrace{\left(\varepsilon \delta(0) + (1 - \varepsilon) \sqrt{\frac{2}{\pi S}} e^{-\frac{x^2}{2S}} \right)}_{f_x(x)} \int_0^\infty \underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \log \left(\underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \right) dz dx \quad (3.63)$$

which can be simplified due to the presence of the delta function inside the double integral.

$$H(Z^2 | X) = \left[\int_0^\infty \int_0^\infty \left((1 - \varepsilon) \sqrt{\frac{2}{\pi S}} e^{-\frac{x^2}{2S}} \right) \underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \log \left(\underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \right) dz dx \right] + \left[\int_0^\infty \int_0^\infty (\varepsilon \delta(0)) \underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \log \left(\underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \right) dz dx \right] \quad (3.64)$$

We can simplify the second double integral (*Int*₂) which involves a delta function and it becomes:

$$Int_2 = \varepsilon [1 + \log 2] \quad (3.65)$$

Thus $H(Z^2|X)$ is given by:

$$H(Z^2 | X) = \left[\int_0^\infty \int_0^\infty \left((1 - \varepsilon) \sqrt{\frac{2}{\pi S}} e^{-\frac{x^2}{2S}} \right) \underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \log \left(\underbrace{\frac{1}{2} \exp\left(-\frac{(x^2+z)}{2}\right) I_0(x\sqrt{z})}_{f_{Z^2|X=x}} \right) dz dx \right] + \varepsilon [1 + \log 2] \quad (3.66)$$

where the first double integral can be solved numerically as it is the same one in the half-Gaussian input case with the exception of the $(1-\varepsilon)$ factor.

After using (3.62) to find $H(Z^2)$ and then (3.64) to find the mutual information we are able to compare this modified half-Gaussian distribution with a pure half-Gaussian. The results are shown in Figure 3.9. As SNR decreases the optimum value of ε approaches 1, making the input distribution completely discrete while for higher SNR values, ε approaches 0 giving us the optimum distribution as the half-Gaussian distribution.

Figures 3.10 and 3.11 show how the spectral efficiency changes for different values of ε for fixed values of SNR, -10 and 0 dB respectively. For -10dB the optimum value of ε is higher than that for 0dB, confirming what Jacobs mentioned in [9] and Ho in [5] that for extremely low values of SNR, a discrete amplitude distribution is the optimum.

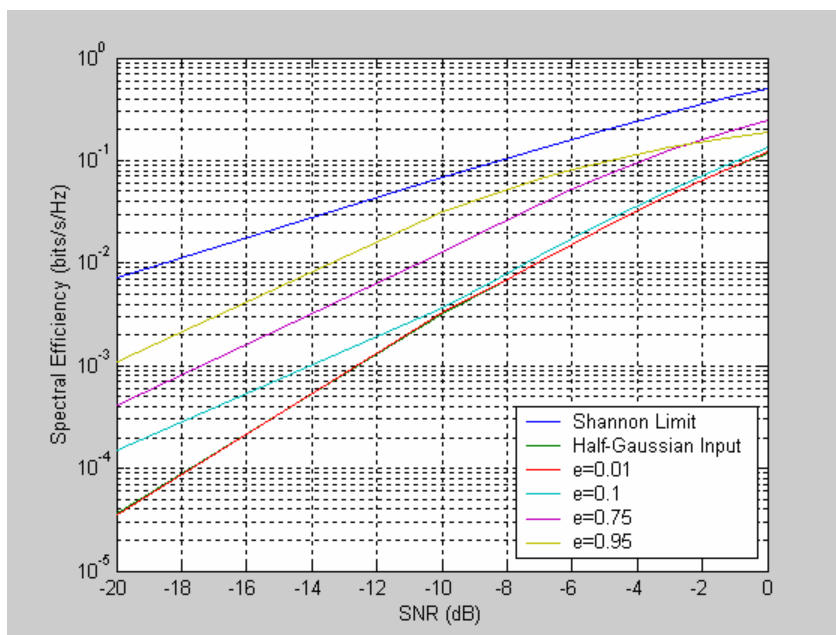


Fig. 3.9 Mutual information for modified half-Gaussian input with $m=2$ compared to a half-Gaussian input and the Shannon bound

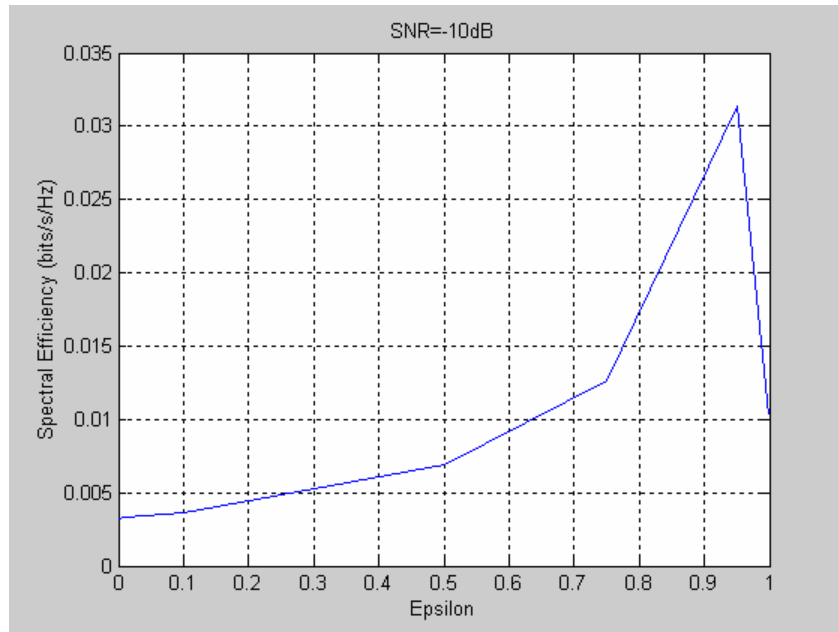


Fig.3.10 Mutual information for increasing ϵ - SNR=-10dB

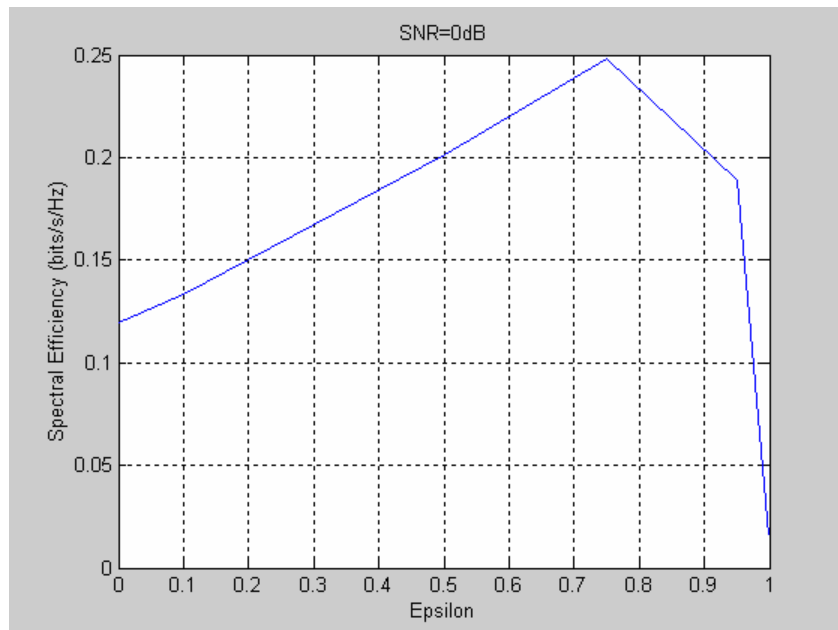


Fig.3.11 Mutual information for increasing ϵ - SNR=0dB

3.8.3. Power efficiency for a modified half-Gaussian input

To calculate the power efficiency we use the method outlined in chapter 2 on how to obtain the power efficiency based on our spectral efficiency results. From our numerical results for the $m=2$ case, the modified half-Gaussian with an $\varepsilon=0.95$ gives us the best results for the minimum number of photons per bit as it closely approaches the limit of 0.693 when spectral efficiency goes to zero, at -20dB the number of photons per bit for this distribution is 1.5. Figure 3.12 shows the results for number of photons per bit for the modified half-Gaussian distribution with $\varepsilon=0.95$.

For purposes of comparison, for the $m=1$ case where a half-Gaussian distribution was used as the input, the minimum number of photons per bit as shown in Chapter 2 is 1.9.

If we use a half-Gaussian distribution for the $m=2$ case however, we will not be able to find numerically a lower bound on number of photons per bit as the slope of the channel capacity changes with the square of SNR.

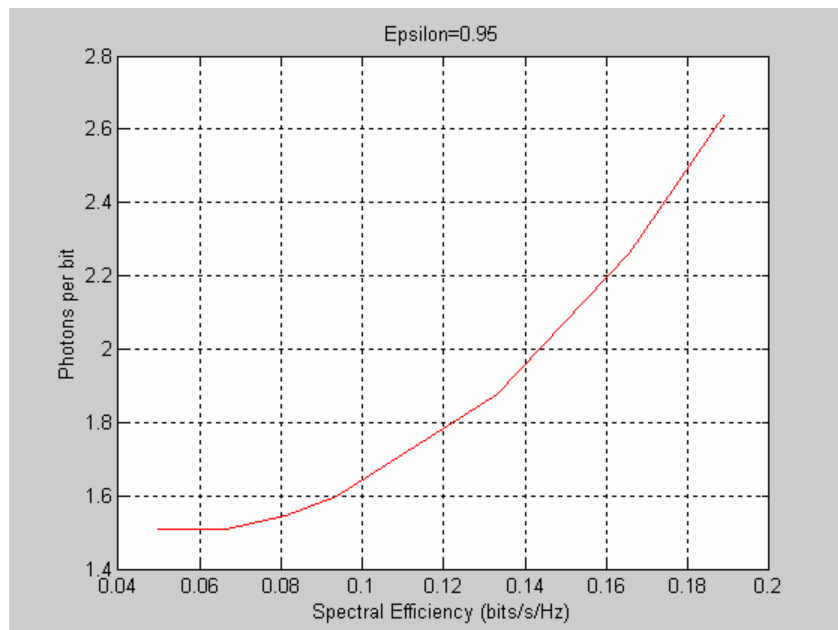


Fig. 3.12 Minimum number of photons per bit for 2 DOF using a modified half-Gaussian distribution with $\varepsilon=0.95$

3.9. Numerical calculation of mutual information for 4 degrees of freedom with a modified half-Gaussian input

In this section we extend the work we did in section 3.8 for the 2 degrees of freedom case to 4 degrees of freedom. The methodology is identical to the 2 degrees case but the expressions because of the extra noise terms are different.

To calculate the mutual information we have to again calculate the entropies $H(Z^2)$ and $H(Z^2|X)$.

Using the Laplace transform and work done previously for the $m=4$ case for a half-Gaussian input we find the distribution of z^2 to be

$$f_{z^2}(z) = \frac{(1-\varepsilon)}{4\sqrt{S+1}} e^{-\frac{z}{4}\left(\frac{S+2}{S+1}\right)} \left[I_0\left(z \frac{S}{4(S+1)}\right) - I_1\left(z \frac{S}{4(S+1)}\right) \right] + \frac{\varepsilon}{4} z e^{-\frac{z}{2}} \quad (3.67)$$

which is used to calculate the entropy of z^2 .

The expression for conditional entropy $H(Z^2|X)$ we obtain by using the method outlined in sections 3.2-3.4 is given by

$$H(Z^2|X) = \left[\int_0^\infty \int_0^\infty \left((1-\varepsilon) \sqrt{\frac{2}{\pi S}} e^{-\frac{x^2}{2S}} \right) \underbrace{\frac{\sqrt{z}}{2x} \exp\left(-\frac{(x^2+z)}{2}\right) I_1(x\sqrt{z})}_{f_{z^2|X=x}} \log \left(\frac{\sqrt{z}}{2x} \exp\left(-\frac{(x^2+z)}{2}\right) I_1(x\sqrt{z}) \right) \underbrace{I_1(x\sqrt{z})}_{f_{z^2|X=x}} dz dx \right] + \frac{\varepsilon}{4} \sqrt{2\pi} \left[\frac{5}{2} - \frac{3}{2} \log(2) + \frac{\text{eulergamma}}{2} \right]$$

(3.68)

, where *eulergamma* is the Eulergamma constant and is approximately 0.577.

Although we can now obtain a final expression for the mutual information for the modified half-Gaussian for the $m=4$ case, calculating direct numerical results from it for a complete range of SNR is difficult, especially at low SNR values. Due to the nature of the

modified Bessel functions involved and also the many exponential functions numerical integration becomes extremely difficult and inaccurate for low SNR values. However, we can obtain fairly accurate results for the mutual information for higher values of SNR and from there we can use an extrapolation method (described in Section 3.10) for lower values of SNR.

Work done by Jacobs in [9] mentioned that the optimum distribution at low SNR leads to mutual information that is linearly dependent on SNR. In the following section we will present a simple proof that the modified half-Gaussian distribution gives mutual information that is linearly dependent on SNR for small SNR. After this proof we will show results we have obtained for the $m=4$ case for different values of ε and compare it to results calculated previously for the half-Gaussian distribution and also the Shannon bound.

3.10. Proof that modified half-Gaussian (MHG) distribution gives mutual information that is linearly dependent on SNR for small SNR

Numerical results in section 3.6 have shown that the mutual information for the half-Gaussian (HG) input distribution is of the form

$$I_{HG}(X;Z) \approx \frac{1}{2} \ln(1+kS) \quad (3.69)$$

,where k is a constant (depending on the number of degrees of freedom) for $S > S_o$ where $S_o \approx 1$. Specifically $k \approx 0.36, 0.18,$ and 0.12 for $m=1,2,$ and 4 respectively at high SNR values. The mutual information for the modified half-Gaussian, for $S < S_o$ should then be of the form

$$I_{MHG}(X;Z) = \frac{1}{2}(1-\varepsilon) \ln(1+kS_o) \quad (3.70)$$

where

$$1-\varepsilon = S/S_o \quad (3.71)$$

Substituting (3.71) in (3.72) gives

$$I_{MHG}(X;Z) = \frac{1}{2} \frac{S}{S_o} \ln(1+kS_o) \quad (3.72)$$

so that the mutual information goes to zero linearly with SNR S .

For the $m=2$ case we have used (3.72) to estimate the results for values of $\varepsilon=0.1$ and 0.75 and compared to the numerical results obtained in section 3.9 as well as the half-Gaussian results obtained previously. Figure 3.13 shows these results. As it can be seen the estimated results give higher values for the mutual information than the direct calculation.

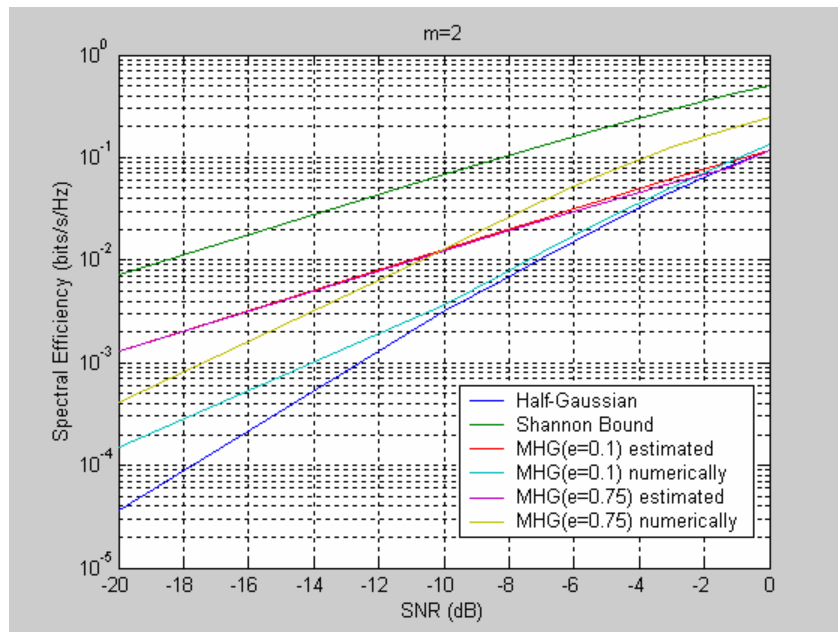


Fig. 3.13 Mutual information for modified half-Gaussian input with $m=2$ compared to a half-Gaussian input and the Shannon bound using a linear approximation

Results obtained using the extrapolation procedure for the $m=4$ case are shown in Figure 3.14. As expected the modified half-Gaussian gives us better results than the half-Gaussian distribution for low values of SNR.

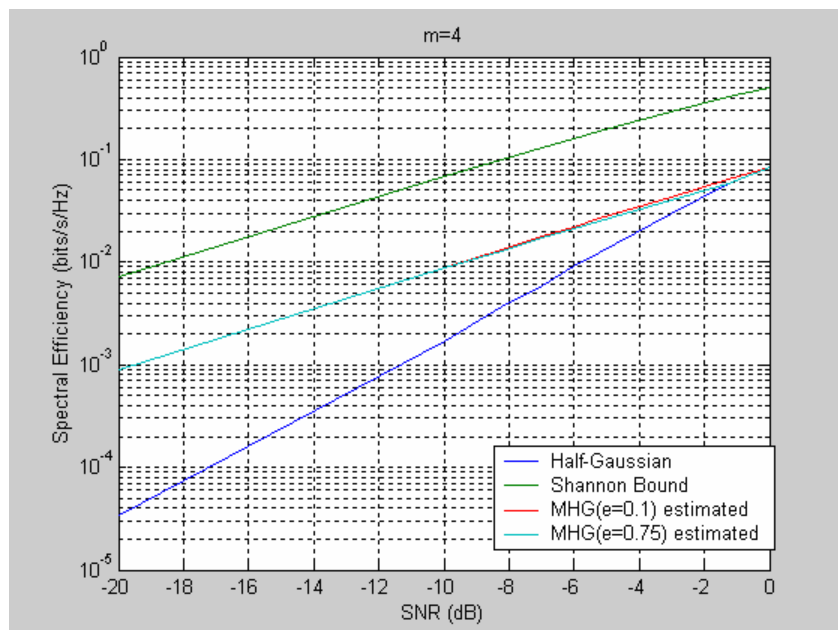


Fig. 3.14 Mutual information for modified half-Gaussian input with $m=4$ compared to a half-Gaussian input and the Shannon bound using the extrapolation method

In Figure 3.15 we have also compared the results for the modified half-Gaussian between the $m=2$ and $m=4$ cases to each other and to the half-Gaussian. As expected the results for the $m=2$ case are slightly better than the $m=4$ case and both give results much closer to the Shannon limit than does the half-Gaussian.

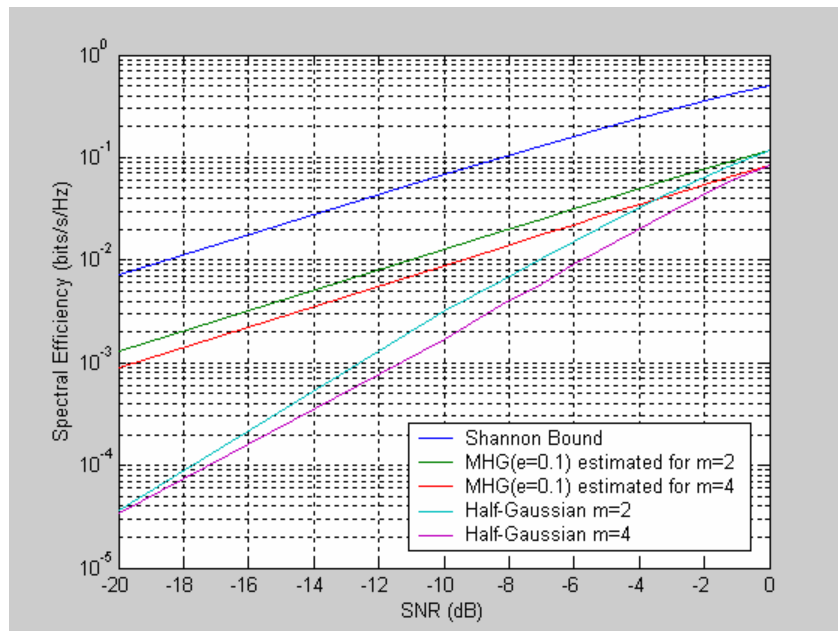


Fig. 3.15 Mutual information for modified half-Gaussian input with $m=2$ and $m=4$ compared to results obtained with the half-Gaussian input

The principal conclusions of this chapter are that for high SNR the spectral efficiency is the same independent of the number of DOFs and that the half-Gaussian is indeed the optimum distribution at high SNR values. For low values of SNR for 2 and 4 DOFs the slope of the mutual information changes with the square of SNR when a half-Gaussian distribution is used thus it becomes necessary to use a different distribution to achieve channel capacity at low SNR values. We used a modified half-Gaussian distribution which had discrete and continuous elements and provided a simple proof that the modified half-Gaussian distribution results in mutual information that goes to zero linearly with SNR

The overall conclusions of the dissertation are considered in the following chapter.

Chapter 4

Conclusions and Open Questions

In the previous 3 chapters we have considered the channel capacity in fiber optic communication systems with incoherent detection involving single (2 DOFs) and two-polarizations (4 DOFs). Some important results have been obtained, and these will be summarized in this chapter. Also, open questions will be described which might lead to future work.

The first important contribution of this dissertation was the evaluation of the mutual information for different input distributions when using a coherent receiver (1 DOF). We know according to Shannon's theorem that the Gaussian distribution is the optimum in order to maximize the mutual information but that is not necessarily true when incoherent detection is used. In the existing literature we were not able to find results for different distributions for coherent detection, only the proof for Shannon's bound. Also, by using non-negative input distributions, we were able to get insight on how PIN receivers work

These results were then extended to when the receiver looks only at the magnitude of the received signal. In this case instead of using the direct definition of mutual information we used a Markov chain which gave us simpler expressions that could be solved numerically, and which allowed us to use the results for the coherent receiver obtained previously. We obtained numerical results for a complete range of SNR including low values of SNR which gave us the minimum number of photons per bit necessary for transmission. In this case the minimum number of photons per bit approaches 1.9 when a half-Gaussian distribution is used compared to the optimum number of 0.69 obtained for a Gaussian distribution.

By obtaining results for the 1 DOF case we were able to form a mathematical framework that later was used for higher DOFs cases.

In Chapter 3, the biggest contributions of this dissertation were made. The prime objective of this dissertation was to obtain the spectral and power efficiencies for incoherent systems using 2 (single polarization) or 4 (two polarizations) DOFs.

For incoherent detection we started out by using the half-Gaussian distribution as the input as in the existing literature [2,3] it was considered as the optimum distribution when a single polarization is used. We proved numerically and analytically that for high SNR the spectral efficiency would be the same independent of the number of DOFs. We were able to obtain a compact equation for spectral efficiency which behaves in a similar way to the Shannon capacity formula but with the SNR scaled by a constant.

We found that for low values of SNR for 2 and 4 DOFs the slope of the mutual information changes with the square of SNR when a half-Gaussian distribution is used which is in contrast to the 1 DOF case where the slope of the mutual information changes linearly with the SNR like the Shannon bound. This fact together with work done by Jacobs in [4] as well as a recent paper by Ho [5] led us to conclude that for low SNR the half-Gaussian distribution is not the optimum distribution and that a binary distribution is needed to obtain channel capacity in the limit of low SNR values.

This led to consideration of a modified half-Gaussian distribution which had a discrete component (an impulse function at the origin) scaled by an ε factor and a continuous component (half-Gaussian distribution) scaled by a $(1-\varepsilon)$ factor. The objective was to have a distribution that gives good results at low SNR, and by proper choice of ε scales to the half Gaussian result at high SNR. Furthermore, this enabled us to utilize results already obtained.

We provided a simple proof that the modified half-Gaussian distribution results in a mutual information that goes to zero linearly with SNR. We found that as SNR decreases the optimum value of ε approaches 1, making the input distribution approach a discrete

binary distribution, while for higher SNR values, ε approaches 0 giving us the optimum distribution as the half-Gaussian distribution.

In terms of power efficiency as ε increased we obtained better results for the minimum number of photons per bit for the 2 DOFs case. The minimum number approached the limit of 0.693 for increasing ε when the spectral efficiency went to zero. For $\varepsilon=0.95$, the minimum number of photons per bit was approximately 1.5. We concluded that with the appropriate choice of input distribution the incoherent channel can achieve the same lower limit for photons per bit as the coherent channel.

For future work, some suggestions can be made based on open questions from this dissertation. The first question is the analytical proof that the half-Gaussian distribution is indeed the optimum distribution for high SNR values for the incoherent channel. Existing literature and this work have been able to demonstrate this only on a numerical basis.

Another possibility for future work would be to explicitly determine the limiting optimum distribution for the 4 DOF case analogous to that obtained by Jacobs for the 2 DOF case.

Finally there are some small discrepancies between results computed in Chapter 2 using the Markov chain method and those computed by direct calculation. It is not clear whether these discrepancies are associated with difficulties in accurate numerical evaluation of the integrals or whether there is some more fundamental reason for these discrepancies.

References

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.
- [2] M.Shtaif & A.Mecozzi, "Limits on the Spectral Efficiency of Intensity Modulated Direct Detection Systems with Optical Amplifiers," *OFC 2001*, Paper MM1, Anaheim, CA, March 2001.
- [3] A.Mecozzi & M.Shtaif, "On the Capacity of Intensity Modulated Systems Using Optical Amplifiers," *IEEE Photonics Technology Letters*, vol.13, no.9, 1029-1031, September 2001.
- [4] Ira Jacobs, "Limits on the Power & Spectral Efficiency of Direct Detection Systems with Optical Amplifiers," Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers, Vol. 1, pp. 8-12, Nov. 4-7, 2001.
- [5] K.P.Ho, "Exact Evaluation of the Capacity for Intensity-Modulated Direct-Detection Channels With Optical Amplifier Noises," *IEEE Photonics Technology Letters*, vol.17, no.4, pp.858-860, April 2005.
- [6] Mitra et.al, "Nonlinear Propagation and Information Theory", *Opt Trends*, pp.S22-S28, March 2002.
- [7] P.P.Mitra & J.B.Stark, "Nonlinear limits to the information capacity of optical fibre communications", *Nature*, vol.411, pp.1027-1030, 28 June 2001.
- [8] E.E.Narimanov & P.Mitra, "The Channel Capacity of a Fiber Optics Communication System: Perturbation Theory", *Journal of Lightwave Technology*, vol.20, no.3, 530-537, March 2002.
- [9] I.Jacobs, "Some Results on the Capacity of the Time-Discrete Incoherent Gaussian Channel," *Bell Telephone Laboratories Internal Memorandum*, August 1966.
- [10] Proakis, J.G. (1995) *Digital Communications*, McGraw-Hill, New York.
- [11] Kokoska, S. and Zwillinger, D. (2000) *Standard Probability and Statistics Tables and Formulae: Student Edition*, CRC, Boca Raton.

- [12] I.Jacobs & S.Teotia, "Bounds on the power and spectral efficiency of optical fiber communication systems," Optics East 2005, Boston, MA, August 2005.
- [13] Papoullis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.
- [14] M.Abramowitz & I.Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965
- [15] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley, 1991.