TIME-VARYING CHANNELS WITH SIDE INFORMATION AND SEPARATION PRINCIPLE

by

Masoud Khansari and Martin Vetterli

Memorandum No. UCB/ERL M94/103

30 December 1994

TIME-VARYING CHANNELS WITH SIDE INFORMATION AND SEPARATION PRINCIPLE

by

Masoud Khansari and Martin Vetterli

Memorandum No. UCB/ERL M94/103

30 December 1994

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Time-Varying Channels with Side Information and Separation Principle

Masoud Khansari and Martin Vetterli
Department of EECS, University of California, Berkeley, CA, 94720

Abstract

The separation principle - since its introduction by Shannon for ergodic sources and discrete memoryless channels - has been a guideline in designing communication systems. It states that separate optimization of source and channel coding modules is optimum and by joint optimization of source and channel coder, the overall performance of the system cannot be improved any further. Therefore, to achieve minimum end-to-end distortion, one has to transmit information at maximum rate at which reliable communication is possible. In this paper, we consider a class of time-varying channels where the impulse response of the channel changes with time. We consider the cases where the side information regarding the state of the channel is available at the receiver and/or the transmitter. Availability of the channel state information at the receiver can result in the possibility of using different channel codes based on the state of the channel. At the same time, having this information at the transmitter can be used to transmit signals over the channel with time-varying power. It has been shown that that availability of this information at both the transmitter and the receiver can increase the rate at which information is transmitted reliably [5]. We show that this is also the case when these information are available at either the receiver or the transmitter alone.

We also demonstrate that if the channel side information is available at the receiver. the separation principle does not hold any longer. Specifically, when the state information is present at both the receiver and the transmitter, then for the time-varying additive noise Gaussian channels with constraint on the average power, the power allocation that achieves maximum capacity is, in general, not the same as when minimum end-to-end distortion is targeted. Explicit power control algorithms for narrow-band Rayleigh fading channels are found and shown that different power allocation strategy and considerable overall improvement can be expected when rate-distortion function of the source is taken into consideration.

We also show that for time-varying channels with informed receiver superposition codes - constructed in a similar fashion as for broadcast channels - can be used to achieve the capacity. Indeed, this class of channels is a generalization of the broadcast channels. Both for the time-varying channels with informed receiver and the broadcast channels, the separation principle fails and different channel codes should be used based on the distortion-rate function of the source. Finally, time-varying channels with informed transmitter are investigated and shown that the capacity of these channels are achieved through allocation of power based on the state of the channel.

1 Introduction

The joint source and channel coding theorem has been stated for different classes of sources and channels. For example, Shannon derived this theorem for ergodic sources with rate-distortion function $R(D)$ and discrete memoryless channels with capacity C [10]. This theorem usually has two components known as positive and converse coding theorems. More precisely, Shannon showed that, for all $\epsilon > 0$, there exists a block code to transmit the information over the channel with distortion $D+\epsilon$ or less through transmitting $\left(\frac{C}{R} - \epsilon\right)$ source letter per use of channel. Conversely, if a coder transmits t message letters through n use of the channel resulting in end-to-end distortion of D, then

$$\frac{n}{t}C \geq R(D) - \epsilon, \tag{1}$$

for all $\epsilon > 0$. Note that the above theorem implies that only two parameters, namely the rate of the source at distortion level D, $R(D)$, and the capacity of the channel C, is sufficient to determine if reliable transmission of the source over the channel is feasible. Therefore. to achieve minimum end-to-end distortion one has to aim for maximum possible channel capacity.

Also, an implicit consequence of the above theorem is what is known as the *separation principle*. Based on this principle, the operation of the transmission of a source over a channel can be split into a source coding module and a channel coding module without loss of optimality. In other words, the channel encoder can be designed -without any consideration of the source- to achieve transmission rate close to the channel capacity, and at the same time, the source encoder should compress the source as much as possible without any consideration of the channel encoder being used.

Even though the separation principle has become a guideline in designing communication systems, one has to be aware of the fact that there are instances where the separation principle does not hold. For example, if the ergodicity assumption of the source is not valid or if it is not possible to characterize the capacity of the channel by only one parameter, then the separation principle is not necessarily valid and further investigation of the statistical characteristics of the source and/or channel become necessary. Also, while transmitting correlated sources over multi-access channels, it can be shown that the separation principle does not hold [2]. In [14], it is shown that source-channel separation principle always holds if the source is stationary or the capacity [1] of the channel is insensitive to the "good" codes being required for all sufficiently long block-lengths or only for infinitely many long block-lengths.

If we assume that the amount of incurred delay [2] should be bounded, then joint source and channel coding design can be advantageous. For example in the case of causal coding of a source transmitted over the binary symmetric channel it has been shown that the separation principle is not valid [16]. Also, for channels with memory, one can show that the quantization of the source considering the channel statistics can improve the overall end-to-end performance of the quantization operation[15].

In this report(paper), we are considering the class of time-varying channels where the statistics of the channel changes with time. These variations are usually characterized by considering the impulse response of the channel to be time-varying or the noise conditions change over time. Time-varying channels can be used to model a large class of channels. An example is the Rayleigh fading channel where due to the multi-path the instantaneous value of the signal to noise ratio (SNR) can be an order of magnitude lower or higher than the nominal SNR value [5]. The information regarding the state of the channel can be used in either the receiver and/or the transmitter to improve the overall performance of the system. It is known that availability of these information at both the receiver and the transmitter can increase the rate at which reliable transmission is possible. We show that availability of these information at either the receiver or the transmitter alone can also increase the overall transmission rate.

Another interesting question with respect to these channels is the validity of the separation principle. Specifically, for the *informed* receiver and/or transmitter, is the two stage processing - separate source and channel coding - still optimum or can further improvement be realized through joint source and channel coding? We show that if the channel state information is available at the receiver then the separation principle is not valid anymore. This is, however, not true in the case of informed transmitter.

In the next section, we first outline and formalize the channel model used throughout the

[1] A different definition of capacity is used in [14] than the usual definition based on the mutual information of the input and output of the channel.

[2] This delay is usually characterized by the block size of the codes used in the source and the channel coders

paper and then characterize four fundamental cases based on the availability of the channel state information at the transmitter and/or the receiver where each case is treated separately in the subsequent sections. We also show that the broadcast channel can be considered as a specific case of the time-varying channel with informed receiver, and similarly to this case the separation principle of the source and channel does not hold any longer.

2 Time-Varying Channels

Time-varying channels are modeled with linear time-varying impulse response, namely $h(t, w(t))$, and additive channel noise, $n(t)$. The channel state is indexed by $w(t) \in \Omega$ which is assumed to be stationary and ergodic and all the states $w(t)$ are assumed to be positive recurrent. Furthermore, it is assumed that the set Ω, which can be uncountable, is known to both the transmitter and the receiver and they both have information about the possible states. It is also assumed that the channel at each state is memoryless for the duration of that state. Based on the availability of $w(t)$, at all time t, at the receiver and/or receiver one can distinguish four different cases:

Case I (Informed transmitter and receiver): Both the receiver and the transmitter are aware of the state of the channel

Case II (Informed receiver): Only the receiver knows about the current state of the channel.

Case III (Informed transmitter): Only the transmitter knows about the current state of the channel.

Case IV : Neither the receiver nor the transmitter knows the channel state.

These classes of channels are known as channels with side information [11], Figure 1. In the case where the state of the channel is available, we assume that it is known instantaneously and without any delay [3].

Availability of the channel state information at the receiver and the transmitter has different consequences. Having this information at the transmitter results in being able to vary the transmitted channel symbol as well as the frequency of choosing each input symbol based on the state of the channel. For example, in the case of channels with constraint on

[3] In the case where the state of the channel needs to be estimated and it is not provided as a side information then this delay cannot be assumed to be zero.

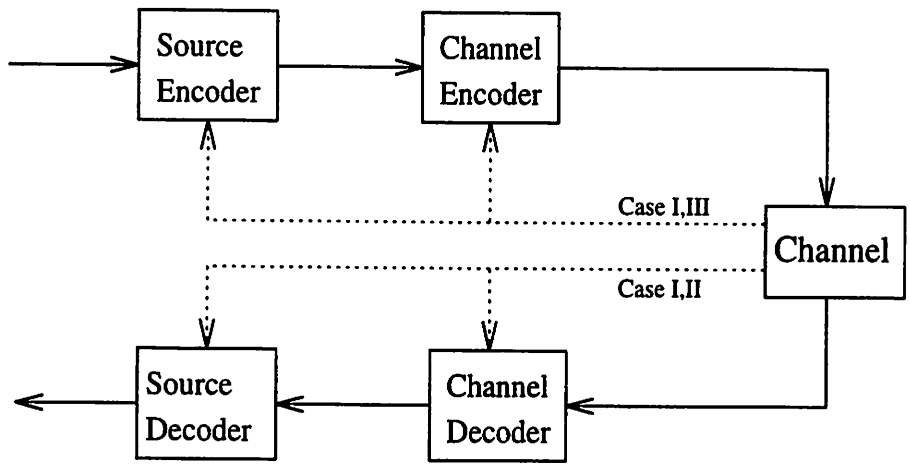

Figure 1: Schematic of a communication system with side information

the transmitted power, the transmitted power can be varied based on the state of the channel. Having this information at the receiver, however, provides the capability of being able to decode the output of the channel based on the state of the channel or to use different channel codes altogether.

An important class of channels are the additive white Gaussian noise (AWGN) channels. that is, channels where the channel noise is assumed to be additive and has Gaussian distribution. If we assume that this noise is white, then the capacity of this channel is given as a function of single parameter, namely the SNR as

$$C(\gamma) = \frac{1}{2}\log(1+\gamma), \tag{2}$$

where SNR is denoted by γ. As far as the capacity is concerned, the effect of time-varying impulse response is then to modulate the value of γ. Therefore, we can index the states of the channel based on the noise power of the states.

From a practical aspect, by monitoring the value of the received SNR, the receiver can get a good indication of the status of the channel. This is true since the state of the channel is classified based on its noise level. If now this information is not sent back to the transmitter then one can assume that the operating situation is similar to that of **Case II** and the transmitted power is constant throughout the transmission [9]. And, if this information is sent back to or is available directly from the channel at the transmitter, then the situation is similar to that of **Case I**. This is very similar in nature to that of closed-loop or fast power control algorithms currently used in wireless transceivers. In such transceivers, the receiver updates the transmitter about the status of the received signal (and hence the channel) and requests for either increasing or decreasing the transmitted signal power [13].

We define the capacity of the channel as the maximum possible information rate that can be transmitted reliably over a channel or

$$C_{\max} \triangleq \max \sum_{i \in \mathcal{I}} p_i R_{i,c} \tag{3}$$

where $\mathcal{I} = \Omega$ is the sample space of the state process, $R_{i,c}$ is the information rate transmitted and p_i is the probability of being at state $i \in \mathcal{I}$. The state process being ergodic, p_i also corresponds to the fraction of time spent at state i. Note that the value of R_i is not necessarily the same for the four cases stated under investigation and clearly the overall capacity can be different.

Our primarily goal is to replicate a source signal while encountering minimum end-to-end distortion. Assuming that the distortion function used is additive we then have

$$D_{\min} \triangleq \min \sum_{i \in \mathcal{I}} p_i D(R_{i,d}), \tag{4}$$

where $D(R)$ is the distortion-rate function of the source being transmitted and $R_{i,d}$ is the transmission rate at state i. Note that

$$D_{\min} \leq \sum_{i \in \mathcal{I}} p_i D(R_{i,c}) \tag{5}$$

or the solution of (3) gives an upper bound for the distortion $D_{\min}$ in (4).

Definition: *The source and channel coding separation holds if*

$$D_{\min} = \sum_{i \in \mathcal{I}} p_i D(R_{i,c}) \tag{6}$$

If the separation principle does not hold then it is advantageous to combine source and channel coding in a sense that the codes used to transmit information over the channel should be designed considering the distortion-rate function $D(R)$ of the source.

Considering the AWGN channel, we assume that there is a constraint on the average power of the transmitted signal. In this case, both (3) and (4) are constrained optimization problems with the constraint being on the average power $\bar{S}$. Let S_i be the transmitted power used at state i, then this constraint can be written as:

$$\sum_{i \in \mathcal{I}} p_i S_i \leq \bar{S}. \tag{7}$$

This constraint defines a polytope $\mathcal{S}$ of all the I-tuple points $(S_1, S_2, \ldots S_I)$ that satisfy the above, where I is the cardinality of the state process $\mathcal{I}$. Note that it is possible to have different transmitted power for each state i only if the transmitter is aware of the current state of the channel.

If the rate R_i at state i is a function of S_i, the power used at that state, then we can define the distortion-power function $\bar{D} \triangleq D \circ C$ as

$$\bar{D}(S_i) = D(C(S_i)) \tag{8}$$

where D is the distortion-rate function and C is the capacity function which in the case of the AWGN channel is $\frac{1}{2}\log(1 + \frac{S_i}{N_i})$. Now since C is a convex increasing function of S_i and D is a convex decreasing function of R_i, it is straightforward to show that $\bar{D}$ is a convex decreasing function. The optimization problem (4) can now be rewritten as

$$D_{\min} \triangleq \min \sum_{i \in \mathcal{I}} p_i \bar{D}(S_i). \tag{9}$$

In the following sections, we look at the **Case I,II,III** outlined previously in more details. **Case IV** is the usual case with no information available at either the transmitter or the receiver and its behavior is similar to that of average channel with transmitted power being constant at $S = \bar{S}$ for all the state. In the case of the AWGN channel, the capacity is then

$$C = \frac{1}{2}\log\left(1 + \frac{S}{N}\right), \tag{10}$$

where $N = \sum_{i \in \mathcal{I}} p_i N_i$, N_i being the noise power at state i. Clearly, in this case the separation principle holds and joint source and channel coding does not provide any improvement over two-step processing (separate source and channel coding).

3 Case I: Informed Transmitter and Receiver

In this case, both the transmitter and the receiver are instantaneously aware of the status of the channel. Therefore, the transmitter has the capability of sending information at variable power and can also use different source and channel coder at different states. The latter is possible since the receiver can switch to the corresponding source and channel encoder simultaneously. Such coding system can be looked at as a "time diversity" system with multiplexed input and de-multiplexed output [5]. It is then straightforward to show that

$$C_{\max} = \sum_{i \in \mathcal{I}} p_i I(X_i; Y_i), \tag{11}$$

where $I(X_i, Y_i)$ is the mutual information between the input and the output random processes of the channel at state i [11]. In the case where there is constraint on the average signal power then the capacity is given by:

$$\begin{aligned} C_{\max} &= \max \sum_{i \in \mathcal{I}} p_i I(X_i; Y_i) \\ \text{s.t.} \quad & \sum_{i \in \mathcal{I}} p_i S_i \leq \bar{S} \end{aligned} \tag{12}$$

where now $I(X_i; Y_i)$ is a function of S_i - the signal transmission power at state i. For example, in the case of the AWGN channel where $I(X_i; Y_i) = 1/2 \log(1 + S_i/N_i)$, it is straightforward to show that

$$\begin{aligned} S_i + N_i = \theta & \quad \text{for all } i \text{ such that } S_i > 0 \\ S_i + N_i > \theta & \quad \text{for all } i \text{ such that } S_i = 0 \end{aligned} \tag{13}$$

where θ is a positive constant number [5].

If the aim is to minimize the distortion of a source signal having distortion-rate function D(R), then using a similar multiplexed input and de-multiplexed output scheme for the channel as well as source encoding and decoding, the minimum distortion is found through solving the following optimization problem

$$D_{\min} = \min \sum_{i \in \mathcal{I}} p_i D(I(X_i; Y_i))$$

$$\text{s.t.} \qquad \sum_{i \in \mathcal{I}} p_i S_i \leq \bar{S}. \tag{14}$$

Clearly, unless $D(R)$ is a linear function, the solution to (14) is not the same as (12). This means that for each channel state i, different channel coder should be used based on the distortion-rate function of the source and as a result the separation principle does not hold. For example, for the AWGN channel the minimum distortion is reached if

$$\begin{array}{ll} \frac{-D'(R)}{S_i+N_i}\Big|_{R=1/2\log(1+S_i/N_i)} = \theta & \text{for all } i \text{ such that } S_i > 0 \\ \frac{-D'(R)}{S_i+N_i}\Big|_{R=1/2\log(1+S_i/N_i)} < \theta & \text{for all } i \text{ such that } S_i = 0 \end{array} \tag{15}$$

An implication of the above is to include more number of states with high noise level and at the same time decrease the allocated power of those the states which have low noise level. Also, the more "convex" the distortion-rate function of the source, the more advantageous the above policy and hence the joint source and channel coding becomes. We can now state the following theorem:

Theorem 1 *For a time-varying channel with informed transmitter and receiver and constraint on the average signal power transmission, the separation does not hold and the minimum end-to-end distortion is not necessarily achieved by transmitting at the channel capacity.*

Note that the above is true even though the source is stationary. This result may seem to be in contrast to that of [14] where it is shown that the stationarity of the source is a sufficient condition for the separation principle to hold, however, there is no contradiction as we are considering sources with fidelity criterion, while [14] considers only the case of $D = 0$.

In the following, we consider the Rayleigh fading channel. We show that we can model this channel as a time-varying channel and joint source and channel coding results in a different power allocation algorithms, hence the separation principle does not hold.

3.1 Narrow-band Rayleigh Fading Channel

In this section, we first discuss the development outlined in [5] to find the optimal power control assignment for narrow-band Rayleigh fading channels. This is done through capacity optimization formalized as in (3). We then show that by using optimization (4), an alternative power control algorithm results which has different characteristics and is *dependent* on the distortion-rate function of the source being transmitted. An implication of this is that the power control algorithm that achieves maximum capacity is not necessarily the same as the case which achieves minimum end-to-end distortion, and as a result, different channel coder should be used at each state of the channel. In other words, the separation principle is *not* valid.

Due to multi-path and constructive or destructive combining of these paths, the envelope of the received signal is shown to vary in time and follows a Rayleigh distribution. It is then possible to define channel gain, G, which corresponds to the received signal power when the transmitted signal, S, is constant and $S = 1$. Then G has exponential distribution given by $f_G(g) = e^{-g}$. Let γ be the received SNR when the transmitted power is constant and equal to S, then

$$\gamma = \frac{SG}{N}. \tag{16}$$

Now, if we let $S(\gamma)$, the transmitted signal power, vary in time as a function of γ with an average value $\bar{S}$, then the received SNR is $\gamma S(\gamma)/\bar{S}$ and the optimization (3) becomes

$$\begin{aligned} C_{\max} &= \max_{S(\gamma))} \int \frac{1}{2} \log\left(1 + \frac{\gamma S(\gamma)}{\bar{S}}\right) f(\gamma) d\gamma \\ \text{s.t.} & \quad \int S(\gamma) f(\gamma) d\gamma = \bar{S} \end{aligned} \tag{17}$$

where $f(\gamma)$ is the distribution of γ which is exponential with mean $\gamma_S = \frac{SG}{N}$. It can then be shown that the following power control algorithm is the solution of the above optimization problem

$$\frac{S(\gamma)}{\bar{S}} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma} & \gamma \geq \gamma_0 \\ 0 & \gamma < \gamma_0 \end{cases} \tag{18}$$

where γ_0 is the so-called cut-off SNR which is bounded above from zero and is found by substituting (18) into (17).

The power control algorithm used in practical applications is usually based on the received SNR. If this SNR is lower than what is expected, then the receiver can request the transmitter to increase its transmitting power and hence increases the received SNR. The power control algorithm proposed by (18) and shown in Figure 2 has the following two distinguished differences from what is usually used in practice.

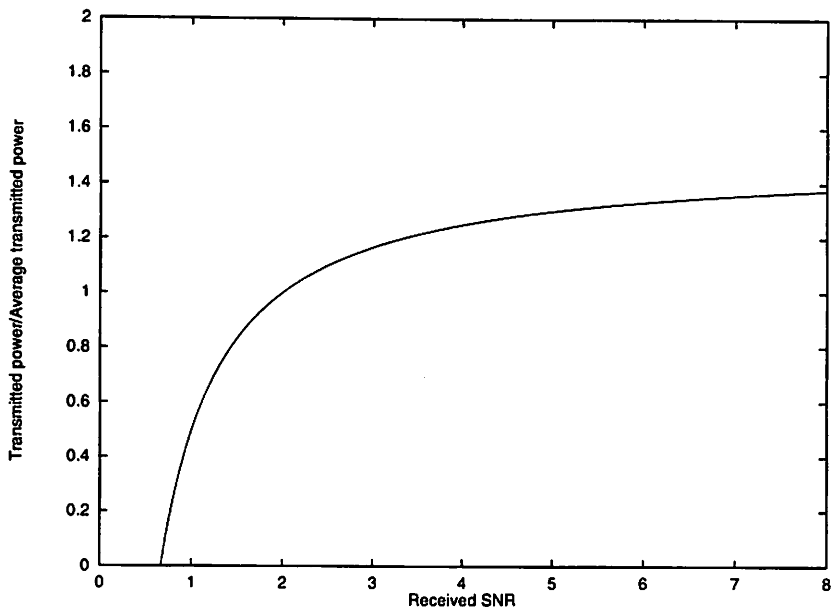

Figure 2: $\frac{S(\gamma)}{\bar{S}}$ vs. received SNR (γ), $\gamma_S = 5, \gamma_0 = 0.667$

- It imposes a cut-off SNR (γ_0) where if the received SNR is below this value. then the transmitted power of that user is set to zero.
- $S(\gamma)$ is a strictly increasing function of γ and approaches $\bar{S}/\gamma_0$ as $\gamma \to \infty$.

An implication of these two properties is that in the multi-user, multi-access environment. where users share transmission medium and each user signal is considered as noise to the other users, the users are split into two groups, those who will be cut-off due to their low signal level (first property) and those who as a result can improve their signal level (second property).

We now show that by considering the optimization (4) a different power control algorithm can be found which is close to what is used in practice and both of the above properties are not necessarily present.

Following the same procedure as above, we have the following optimization problem

$$\begin{aligned} D_{\min} &= \min_{S(\gamma))} \int D\left(\frac{1}{2}\log(1+\frac{\gamma S(\gamma)}{\bar{S}})\right) f(\gamma)d\gamma \\ \text{s.t.} & \quad \int S(\gamma)f(\gamma)d\gamma = \bar{S} \end{aligned} \tag{19}$$

Using the Lagrange multiplier method, the optimal policy has to satisfy

$$\frac{\gamma}{1+\frac{\gamma S(\gamma)}{\bar{S}}} g\left(\frac{S(\gamma)}{\bar{S}}, \gamma\right) = \gamma_0, \tag{20}$$

or

$$\frac{S(\gamma)}{\bar{S}} = \frac{1}{\gamma_0} g\left(\frac{S(\gamma)}{\bar{S}}, \gamma\right) - \frac{1}{\gamma} \tag{21}$$

where g is defined as,

$$g(x, y) \triangleq -D'(R) \mid_{R=\frac{1}{2}\log(1+xy)}. \tag{22}$$

Note that g is a positive increasing convex function in both $S(\gamma)$ and γ [4]. Clearly, γ_0 is a constant greater than zero which can be found by substituting (21) into (19).

From (22) we have

$$g\left(\frac{S(\gamma)}{\bar{S}} = 0, \gamma\right) = -D'(0), \tag{23}$$

where $D'(0)$ is the slope of the distortion-rate function when no information is transmitted. Now since $S(\gamma) \geq 0$, (21) is defined only for

$$\gamma \geq \frac{\gamma_0}{-D'(0)}, \tag{24}$$

where the right hand side is the cut-off SNR which we denote by γ_c. This is a rare event. however, for sources where $D'(0) = -\infty$, $\gamma_c = 0$ and there is no cut-off SNR. This is in contrast to the maximum-capacity power control algorithm of (18).

As $\gamma \to \infty$, from (21), one can show that

$$\frac{S(\gamma)}{\bar{S}} \to 0. \tag{25}$$

This suggest that the asymptotic behavior of (21) is also different from that of the maximum-capacity policy of (18) and is similar to the power control algorithm used in practice. Relations (24) and (25) can be explained intuitively by noting the convexity property of the distortion-rate function, where its behavior around $R = 0$ results in (24) and its asymptotic behavior ($R \to \infty$) results in (25).

---

[4] If $D'(R)$ is constant for all value of R ($D(R)$ is a linear function of R) then (21) is similar to (18) which maximizes the capacity.

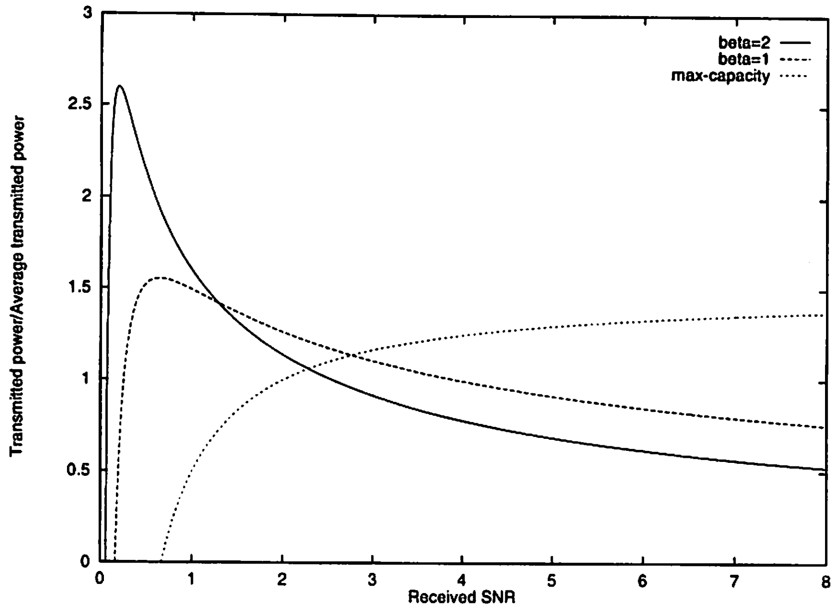

Figure 3: $\frac{S(\gamma)}{\bar{S}}$ vs. received SNR (γ), $\gamma_S = 5$

Example:

Let us assume exponential distortion-rate function $D(R) = 2^{-\beta R}$, $R \geq 0$, where β is a nonzero positive real number. Since $-D'(0) = \beta \ln 2$, then for all $\beta < \infty$, the cut-off SNR, γ_c, is bounded above from zero. Then we can show that the minimum-distortion policy is

$$\frac{S(\gamma)}{\bar{S}} = \begin{cases} \left(\frac{1}{\gamma^{\beta}\gamma_c}\right)^{\frac{1}{\beta+1}} - \frac{1}{\gamma} & \gamma \geq \gamma_c \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

and

$$\gamma_m = \gamma_c \left(1 + \frac{1}{\beta}\right)^{1+\beta}, \tag{27}$$

where γ_m is the γ corresponding to the maximum $\frac{S(\gamma)}{\bar{S}}$. For $\beta = 2$, which corresponds to distortion-rate function of high-rate quantizer [6], we can further simplify the above to

$$\frac{S(\gamma)}{\bar{S}} = \begin{cases} \frac{1}{\gamma^{2/3}\gamma_c^{1/3}} - \frac{1}{\gamma} & \gamma \geq \gamma_c \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

and

$$\gamma_m = \frac{27}{8}\gamma_c. \tag{29}$$

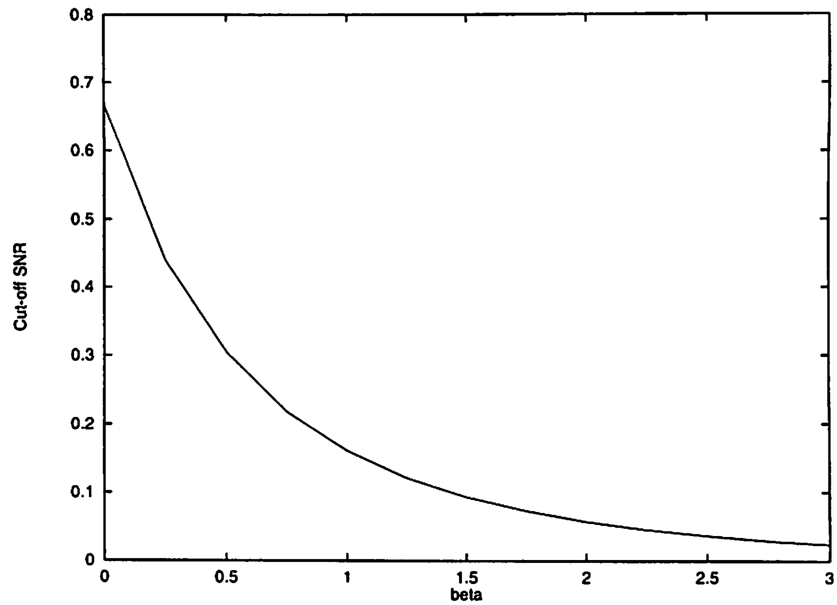

Figure 4: Cut-off SNR (γ_c) vs. β

Figure 3 shows the optimum distortion power allocation policy for different value of β ($\beta = 0, 1, 2$) where the respected cut-off SNR, γ_c, is shown in Figure 4. Note that $\beta = 0$ corresponds to the power allocation for maximum-capacity which is repeated from Figure 2.

Now, as $\beta \rightarrow 0$, $\gamma_m \rightarrow \infty$ and the policy (26) becomes similar to the maximum-capacity policy. This is expected since $D'(R) \rightarrow 0$ is independent of the value of R. At the other extreme, as $\beta \rightarrow \infty$, $\gamma_m \rightarrow e\gamma_c \rightarrow 0$, since $\gamma_c \rightarrow 0$ as shown in Figure 4. Note that the value of β is an indication of the convexity of the distortion-rate function. As a result, as the distortion-rate function becomes more convex, the difference between maximum-capacity and minimum-distortion becomes more pronounced as shown in Figure 5. Note that as the distortion-rate function becomes more convex (β becomes larger), the overall distortion due to the use of minimum-distortion policy decreases whereas maximum-capacity policy distortion flattens and becomes insensitive to the distortion-rate function. Also, for example for $\beta = 2$, the amount of reduction in the overall distortion is significant (more than an order of magnitude) which points to the advantages of joint source and channel coding.

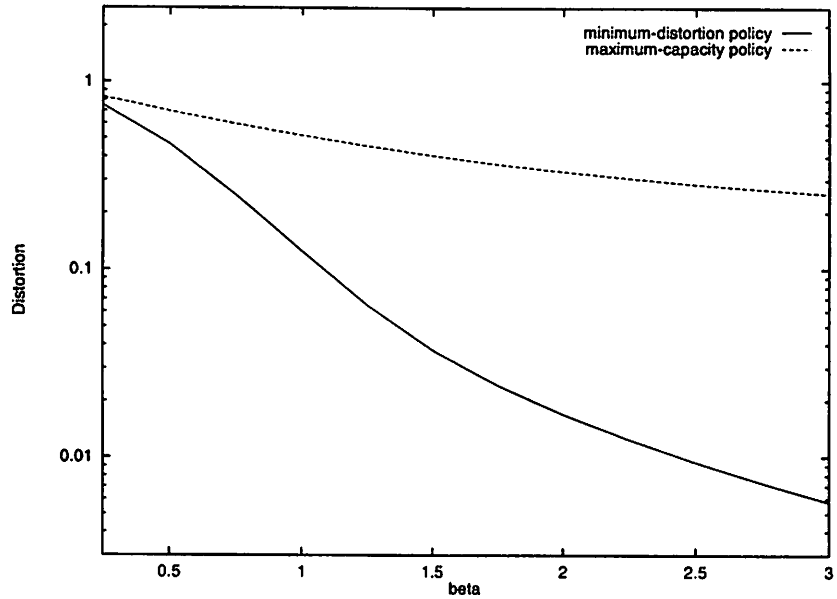

Figure 5: Distortion for minimum-distortion and maximum-capacity policies vs. β in case of informed transmitter and receiver for narrow-band Rayleigh channels with $D(R) = 2^{-\beta R}$.

3.2 Optimization using Distortion-Power Function

If the channel noise is not additive Gaussian then the capacity of the channel cannot be found using (2) and one needs to generalize the derivation of the previous section. As it was stated in the introduction, it is possible to define a distortion-power function for each channel state and show that it is a decreasing convex function of the signal transmission power used at that state. Even though the distortion-rate function is the same for each state, they will have different distortion-power functions since the corresponding noise at each state is different. The higher the noise power, the less convex this function for a given distortion-rate function. The optimum solution can be achieved through minimization of (9) subject to the constraint on the average power. Distortion-power $\bar{D}(S_i)$ being a convex function, the solution of this optimization is given by [7]

$$\begin{aligned} \frac{\partial \bar{D}_i(S_i)}{\partial S_i} &= \lambda \quad \text{all } i \text{ such that } S_i > 0 \\ \frac{\partial \bar{D}_i(S_i)}{\partial S_i} &\leq \lambda \quad \text{all } i \text{ such that } S_i = 0 \end{aligned} \tag{30}$$

where $\bar{D}_i(S_i)$ is the distortion-power function at state i. As a result, the optimum policy is the *constant slope policy* similar to that being proposed to minimize the total rate-constrained distortion [12]. Implicit in the above solution are the following conclusions:

- In general, there is a cut-off state where the noise power is too high to allocate any power to that and all the states with higher noise power.
- As the noise level decreases, the amount of allocated power diminishes to zero.

4 Case II: Informed Receiver

This case corresponds to the situation where the receiver instantaneously becomes aware of the state of the channel but the transmitter does not have this knowledge. In this case the transmitting power cannot be varied by the transmitter but the receiver will have the capability to decode different codes based on the channel state. We assume that both the receiver and the transmitter know the above. In other words, even though the transmitter does not know about the current state of the channel, it knows that the receiver has this knowledge. Therefore, it is possible for the transmitter to present the information in such a way that the receiver can extract the most amount of information.

The situation is similar to the *broadcast* channel problem considered by Cover [3] where a transmitter wants to send information to two different receivers through the same transmission medium but with different channel characteristics - the noise level of the channels are different. The broadcast channel problem is to find the capacity region where each point in this region corresponds to a set of rates that can be simultaneously transmitted to all the receivers. The general two receiver case has been solved by Cover but the capacity region corresponding to three or more receivers is still an open problem. As a result, in this section we only consider a channel with two states, namely *good* and *bad* states where the probability of being in the good state is denoted by q. In general, each two-state time-varying channel can be characterized by a two-receiver broadcast channel with additional parameter q. Even though only the two-state channel case is considered, we must add that the conclusions drawn are general enough to include multi-state channels.

In the appendix, we state some of the formal definitions and theorems related to the broadcast and the *physically degraded* broadcast channels for two receiver case. Superposition codes are usually used to achieve the capacity region, where two codes with the same input alphabet $\mathcal{X}$ but different output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ corresponding to different channels are designed. At each transmission instant the codes intended to both receivers are added and then transmitted. The good receiver first decodes the information intended for the bad receiver, which can be done free of error, and then subtract this from the received code to get the information intended to itself. For the bad receiver, however, the information sent to the good receiver is considered as an additive noise. For example, in the case of the AWGN channel, two codes corresponding to two different SNR values of $\frac{\alpha S}{N_1}$ and $\frac{(1-\alpha)S}{\alpha S+N_2}$ are designed, where as in the appendix, N_1 and N_2 are the noise power of the good and bad receivers, respectively, and $0 \leq \alpha \leq 1$.

In the following sub-sections, we first derive the optimal code design policy to achieve maximum-capacity as well as minimum-distortion. We show that these two policies are not necessarily the same. In other words, the channel coders used to achieve optimum performance are not the same (the separation principle does not hold) and as a result, while designing channel coders, it is necessary to consider the characteristics of the source. The case of the AWGN channel is considered in more details.

4.1 Maximum-Capacity Optimization

Let $X - Y_1 - Y_2$ be the corresponding broadcast channel of the two-state time-varying channel and define

$$C(f(u)) \triangleq I(U;Y_2) + q \max\{I(X;Y_1|U) - I(U;Y_2), 0\}. \tag{31}$$

where $f(u)$ is the probability distribution of random variable U, then the capacity of the time-varying channel is given by

$$C_{\max} = \max_{f(u)} C(f(u)). \tag{32}$$

In the case of the AWGN channel, $C(f(u))$ can be described by a single parameter α as

$$C(\alpha) \triangleq R_2(\alpha) + q \max\{R_1(\alpha) - R_2(\alpha), 0\}, \tag{33}$$

where

$$R_1(\alpha) = \frac{1}{2}\log\left(1 + \frac{\alpha S}{N_1}\right),$$

$$R_2(\alpha) \;=\; \frac{1}{2}\log\left(1+\frac{(1-\alpha)S}{\alpha S+N_2}\right), \tag{34}$$

and the optimization problem (3) becomes

$$C_{\max} = \max_{0\leq\alpha\leq 1} C(\alpha). \tag{35}$$

Note that (33) can be interpreted as the possibility of reliable transmission of information at rate $R_2(\alpha)$ all the time, and extra information at rate $R_1(\alpha) - R_2(\alpha)$ only during the time where the channel is in the good state.

For $\alpha < \alpha_l$ where

$$\alpha_l = \frac{\sqrt{(N_1+N_2)^2+4N_1S}-(N_1+N_2)}{2S} \tag{36}$$

$R_1(\alpha) < R_2(\alpha)$ and as a result, (35) can be rewritten as

$$\begin{aligned} C_{\max} &= \max\left\{R_2(0),\ \max_{\alpha_l\leq\alpha\leq 1} C(\alpha)\right\} \\ &= \max\left\{R_2(0),\ \max_{\alpha_l\leq\alpha\leq 1}\{qR_1(\alpha)+(1-q)R_2(\alpha)\}\right\} \end{aligned} \tag{37}$$

Note that the value of α_l is independent of q and depends only on the noise power of the channel, namely N_1 and N_2, and the transmission power S.

Let us define q_* as

$$\begin{aligned} q_* &\triangleq \frac{R_2(0)}{R_1(1)} \\ &= \frac{\log(1+S/N_2)}{\log(1+S/N_1)}, \qquad \text{(for AWGN channel)} \end{aligned} \tag{38}$$

then only if $q > q_*$, $\alpha = 1$ can result in $C_{\max}$ and similarly, $\alpha = 0$ can result in $C_{\max}$ only if $q < q_*$. Now if we assume $\alpha > \alpha_l$, then taking the derivative of $C(\alpha)$ with respect to α, one arrives at

$$\frac{\partial C(\alpha)}{\partial\alpha} = \frac{q}{2}\frac{S}{\alpha S+N_1} - \frac{1-q}{2}\frac{S}{\alpha S+N_2}. \tag{39}$$

Equating the above to zero, we get the following relation between α that maximizes the capacity ($\alpha_{m,c}$) in term of q

$$\alpha_{m,c} = \frac{q(N_1+N_2)-N_1}{-2qS+S}. \tag{40}$$

In the interval of our interest, namely $\alpha_l \leq \alpha_{m,c} \leq 1$, $\alpha_{m,c}$ is an increasing function of q. Substituting (40) into $C(\alpha)$, we have

$$C(\alpha_{m,c}) = \frac{q}{2}\log\left(\frac{q}{1-2q}\frac{N_2-N_1}{N_1}\right) + \frac{(1-q)}{2}\log\left(\frac{1-2q}{1-q}\frac{S+N_2}{N_2-N_1}\right). \tag{41}$$

Let q_l be the value of q which makes (41) to be equal to $R_2(0)$ and find q_h by equating $\alpha_{m,c}$ to 1 or

$$q_h = \frac{1}{2} - \frac{N_2-N_1}{2(2S+N_2+N_1)}. \tag{42}$$

Clearly, $q_l \leq q_h \leq \frac{1}{2}$ where $q_l = q_h = 1/2$ if and only if $N_2 = N_1$, which is the case if the channel had only one state.

We can now define the *superposition region* (q_l, q_h) only if $q_h > q_*$. Note that only if $q \in (q_l, q_h)$ one can achieve higher capacity through use of superposition codes. Outside of this region, one either designs the channel code for the good state ($q \geq q_h$) or for the bad state ($q \leq q_l$). Also, note that both q_h and q_l are less than 1/2, or superposition codes should not be used if the channel is in the good state at least half of the time. In most time-varying channels encountered in practice, the fraction of time spent in the bad state is small (e.g 1%). For such channels, to achieve maximum capacity, one has to design the channel coder for the good state and consider the information transmitted during the bad state as loss of capacity.

Example:

Let $S/N_1 = 7$ and $S/N_2 = 3$. Then from (42) and (38) $q_h' = 0.462$ and $q_* = 0.666$. Now since $q_* > q_h$, it is not possible to define superposition region and the optimal policy is to use $\alpha = 1$ if $q \geq q_*$, and to use $\alpha = 0$, otherwise.

If $S/N_1 = 7$ and $S/N_2 = 1$, then

$$\begin{aligned} q_* &= 0.333, \\ q_h &= 0.367, \\ q_l &= 0.329. \end{aligned} \tag{43}$$

Since $q_* < q_h$, it is possible to define superposition region as $(0.329, 0.367)$ where superposition codes should be used if q belongs to that interval. In the next sub-section, we show that if we consider the distortion-rate function and aim for minimum-distortion instead of maximum-capacity, the superposition region changes and the separation does therefore not hold.

4.2 Minimum-Distortion Optimization

We assume that *successive refinement* of the source as defined by Equitz and Cover is possible [4]. In other words, it is possible to generate two information streams with rates R_2 and $R_1 - R_2$ such that both $\{R_1, D(R_1)\}$ and $\{R_2, D(R_2)\}$ are on the distortion-rate curve. This condition is not too restrictive as a large class of sources such as Gaussian distributed sources with squared-error and Laplacian distributed sources with absolute error can be successively refined [4]. Similarly to (33), then for the AWGN channels, we define

$$D(\alpha) \triangleq D(R_2(\alpha)) + q \min \{D(R_1(\alpha)) - D(R_2(\alpha)), 0\}, \tag{44}$$

where R_1 and R_2 are defined as (34), and the optimization problem is to find

$$D_{\min} = \min_{0 \leq \alpha \leq 1} D(\alpha). \tag{45}$$

Note that if $R_1(\alpha) \leq R_2(\alpha)$ then $D(R_1) \geq D(R_2)$. Therefore, for the α_l given by (36), we can rewrite the above as

$$D_{\min} = \min \left\{ D(R(0)), \min_{\alpha_l \leq \alpha \leq 1} D(\alpha) \right\} \tag{46}$$

Let us define q_* as

$$q_* \triangleq \frac{1 - D(R_2(1))}{1 - D(R_1(0))}, \tag{47}$$

then only if $q > q_*$, $\alpha = 1$ can result in $D_{\min}$ and similarly, $\alpha = 0$ can result in $D_{\min}$ only if $q < q_*$. Note that, similar to the maximum-capacity optimization, the superposition region can be defined only if $q_h > q_*$. Now if $\alpha > \alpha_l$, then $D(R_1(\alpha)) < D(R_2(\alpha))$ and by taking the derivative of the both sides of (44), we arrive at

$$\frac{\partial D(\alpha)}{\partial \alpha} = \frac{qS}{2(\alpha S + N_1)} D'(R_1) - \frac{(1-q)S}{2(\alpha S + N_1)} D'(R_2), \tag{48}$$

where

$$D'(R_i) \triangleq \frac{\partial D(R)}{\partial R}|_{R=R_i(\alpha)} \qquad i = 1, 2. \tag{49}$$

and equating (48) to zero, we find the following relation between q and $\alpha_{m,d}$,

$$q = \frac{\alpha_{m,d}S + N_1}{\alpha_{m,d}S + N_1 + \frac{D'(R_1)}{D'(R_2)}(\alpha_{m,d}S + N_2)}. \tag{50}$$

Note that since $D(R_1) < D(R_2)$, $D'(R_2(\alpha)) \leq D'(R_1(\alpha)) < 0$, and

$$\frac{D'(R_1)}{D'(R_2)} \leq 1, \tag{51}$$

where the equality holds if and only if the distortion-rate function is linear or $R_1 = R_2$. Note that this constitutes a small class of sources. If we now substitute $\alpha_{m,d}$ found from (50) into $D(\alpha)$, then

$$\begin{aligned} D(\alpha_{m,d}) &= qD\left(\frac{1}{2}\log\left(\frac{Aq}{1-(1+A)q}\frac{N_2 - N_1}{N_1}\right)\right) \\ &+ (1-q)D\left(\frac{1}{2}\log\left(\frac{1-(1+A)q}{1-q}\frac{S+N_2}{N_2-N_1}\right)\right), \end{aligned} \tag{52}$$

where $A = D'(R_1(\alpha))/D'(R_2(\alpha))|_{\alpha=\alpha_{m,d}}$, and q_l can be found by equating the above to $D(R_2(0))$.

The right hand side of the superposition region, q_h, can be found by substituting $\alpha_{m,d} = 1$ into (50). Note that

$$q_h \geq \frac{S + N_1}{2S + N_1 + N_2} \tag{53}$$

with equality holding if and only if $D'(R_1)/D'(R_2) = 1$. More surprisingly,

$$\lim_{D'(R_1)/D'(R_2)\to 0} q_h = 1. \tag{54}$$

Noting that if $\alpha_{m,d} = 1$, then $D'(R_2) = D'(0)$, we can rewrite the above as

$$\lim_{D'(0)\to-\infty} q_h = 1. \tag{55}$$

Therefore, for the sources with $D'(0) = -\infty$, if $q \leq q_l$, the channel codes should be designed based on the noise power of the bad state and if $q > q_l$ superposition codes should be used to achieve minimum end-to-end distortion.

In general, the value of $D'(R_1)/D'(R_2)$ is related to the convexity of the distortion-rate function of the source. The more convex this function, the smaller the value of $D'(R_1)/D'(R_2)$, the closer q_h to 1, and the superposition codes become attractive for the time-varying channels encountered in practice. Also, it becomes more advantageous to jointly design the source and the channel coder. We can now state the following theorem:

Theorem 2 *For a two-state time-varying channel with informed receiver and constraint on the average signal power transmission, the separation does not hold and the minimum end-to-end distortion is not necessarily achieved by transmitting at the channel capacity.*

Example:

Let us assume that the rate distortion of the source is given by $D(R) = e^{-\beta R}$ where β is a positive real number. Then

$$q_* = \frac{1 - (\frac{N_2}{S+N_2})^{\beta/2}}{1 - (\frac{N_1}{S+N_1})^{\beta/2}} \tag{56}$$

and from (50),

$$q = \frac{\alpha_{m,d}S + N_1}{\alpha_{m,d}S + N_1 + \left(\frac{N_1(S+N_2)}{(\alpha_{m,d}S+N_1)(\alpha_{m,d}S+N_2)}\right)^{\beta/2}(\alpha_{m,d}S + N_2)}, \tag{57}$$

where q_h corresponds to $\alpha_{m,d} = 1$ or

$$q_h = \frac{S + N_1}{S + N_1 + \left(\frac{N_1}{S+N_1}\right)^{\beta/2}(S + N_2)}. \tag{58}$$

Superposition region is defined only if $q_* < q_h$. In Figure (6), we show the three operating regions, $\alpha = 0, \alpha = 1, 0 < \alpha < 1$, for the case where $\frac{S}{N_1} = 10$ and $\frac{S}{N_2} = 2$. As $\beta - \infty$, the distortion-rate function becomes more convex and $q_h \rightarrow 1$.

4.3 Broadcast Channel

The two-state time-varying channel with informed receiver and constraint on average transmitted power can be characterized by $\{\frac{S}{N_1}, \frac{S}{N_2}, q\}$. If $q = \frac{1}{2}$, then the corresponding time-varying channel is equivalent to the broadcast channel where one transmitter simultaneously sends information to two receivers. Therefore, the analysis done in the previous section is

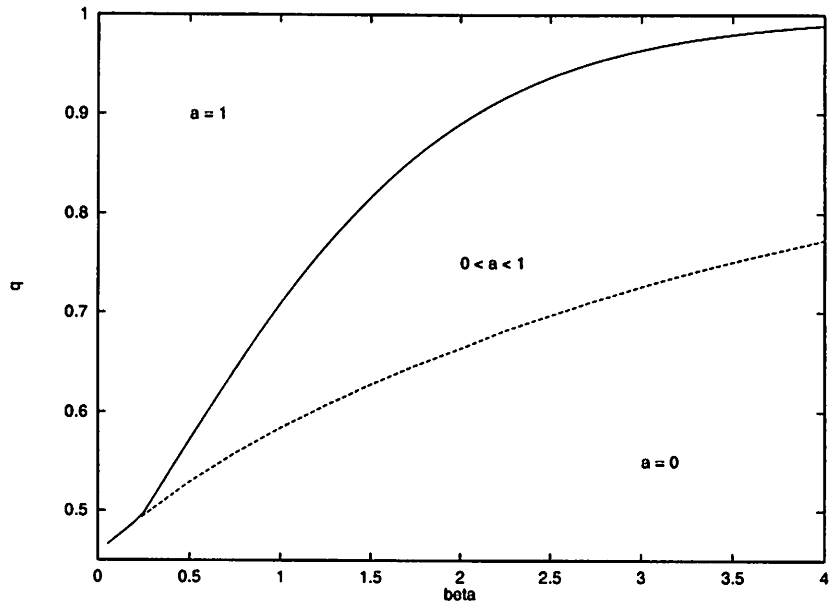

Figure 6: Different operating region for $S/N_1 = 10$ and $S/N_2 = 2$

equally applicable to the broadcast channel. For example, in the case of the broadcast channel, the superposition codes should be used only if 1/2 is in the superposition region found from the analysis of the corresponding time-varying channel. In the case of maximum-capacity analysis, $q_h < 1/2$ and $q = 1/2$ is always outside the superposition region. If $q_* < q_h$, then the capacity is maximized if $\alpha = 1$, otherwise $\alpha = 0$ should be used to achieve maximum capacity.

The situation, however, is not the same if we intend to minimize the total distortion of the signals received at both receivers. This is true since it is now possible for the superposition region to include $q = 1/2$. For this to be true, it is necessary to have $q_* < q_h$ and $q_l < 1/2 < q_h$. The first condition guarantees that superposition region exists where as the second one implies that $q = 1/2$ is in that region.

Example:

In the case of AWGN channel and distortion-rate function $D(R) = 2^{-\beta R}$, q_* and q_h are given by equations 56 and 58, respectively. If we now assume that $\frac{S}{N_1} = 7$ and $\frac{S}{N_2} = 1$, then for

$$0.539 \leq \beta \leq 6.871$$

superposition region exists and $q = 1/2$ is in that region.

5 Case III: Informed Transmitter

This is the case where the transmitter is aware of the state of the channel but the receiver does not have any information on the variations of the channel. Therefore, it is not possible to change the channel coders or to use superposition codes as was the case in the previous two sections. It is however possible to vary the transmitted power based on the channel state due to the knowledge of the transmitter of the current state of the channel.

In his paper, "Channels with side information at the transmitter", Shannon derives the capacity of memoryless discrete channel with side information available at the transmitter [11], where the side information is the current state of the channel. The main theorem of the paper which we partly re-state for the completeness is as follows:

Theorem 3 *The capacity of a memoryless discrete channel K with side state information. defined by p_t and $q_{tj}(i)$* [5] *is equal to the capacity of the memoryless channel K' (without side information) with the same output alphabet and an input alphabet with a^h input letters $X = (x_1, x_2, \cdots, x_h)$ where each $x_i = 1, 2, \cdots, a$. The transition probabilities $r_X(y)$ for the channel K' are given by*

$$q_X(y) = q_{x_1,x_2,\cdots,x_h}(y) = \sum_{t=1}^{h} p_t q_{tx_t}(y). \tag{59}$$

Any code and decoding system for K' can be translated into an equivalent code and decoding system for K with the same probability of error.

The application of this theorem can become clear by the following example. Let us consider a time-varying binary symmetric channel with two states as shown in Figure 7. where α and β are the probabilities of error at each state. The above theorem states that the capacity of this channel is equivalent to the channel K' shown in Figure 7. where for example the input alphabet (A, B) corresponds to transmitter sending A when in state 1 and B when in state 2. The transitional probabilities of this equivalent channel (K') is found using (59).

Since there are only two output channel symbols, two non-zero probability input symbols are sufficient to achieve the capacity. Depending on the value of α or β being smaller or

[5] p_t is the probability of state t and $q_{ti}(j)$ is the conditional probability, if in state t and i is transmitted. that j will be received.

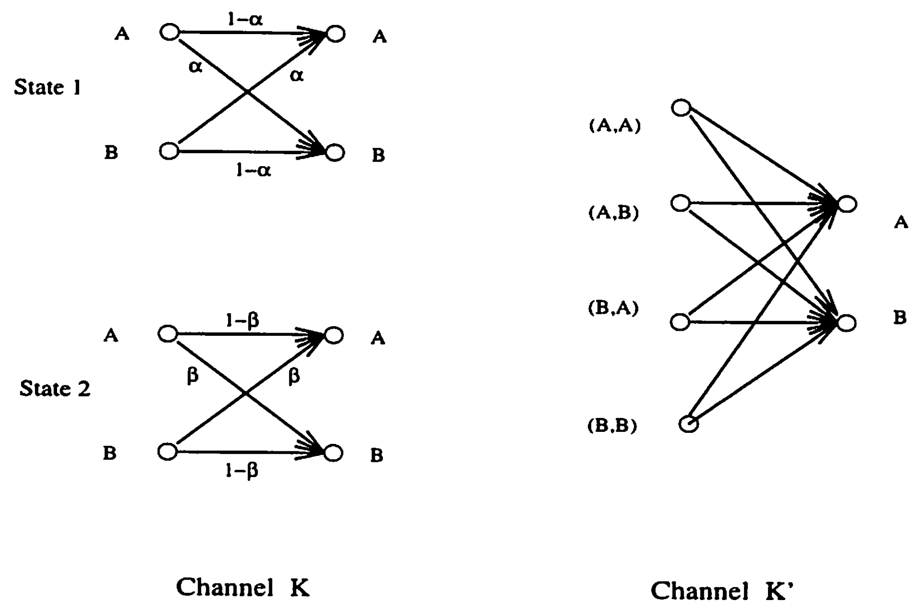

Figure 7: Time-varying binary symmetric channel

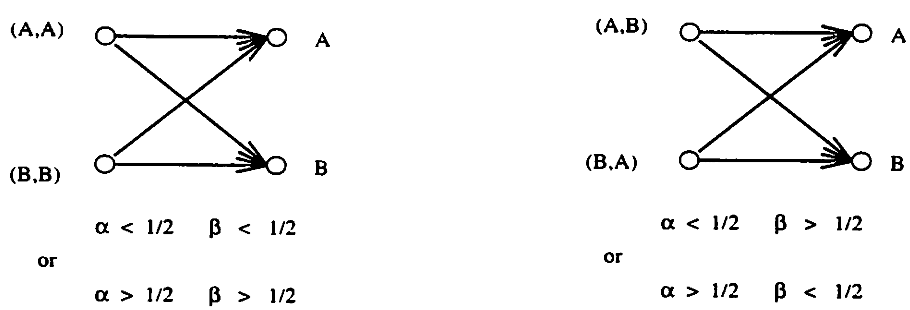

Figure 8: Channel K' for different values of α and β

greater than 1/2, it can be shown that different two input symbols of K' achieves the capacity, as shown in Figure 8.

Let us assume that $P_e(i)$, $i = 1, 2$, are the probabilities of the error at state i which are convex function of S_i and are always smaller than 1/2. Then, the capacity of the channel K is achieved if the left hand coder in Figure 8 is used. This means to transmit the same input symbol of the coder K independent of the current state of the channel. To maximize the overall capacity, one has to therefore minimize the overall probability of the error or to solve the following constrained optimization problem:

$$\begin{aligned} \min_{S_1, S_2} \quad & P_{e,m} = p_1 P_e(1) + p_2 P_e(2) \\ \text{s.t.} \quad & p_1 S_1 + p_2 S_2 = \bar{S}, \end{aligned} \tag{60}$$

where p_i is the probability of being in state i. Then, $P_e(i)$ being a convex function of S_i, the solution to the above problem is given by [7]

$$\begin{array}{ll} \frac{\partial P_e(i)}{\partial S_i} = \theta & \text{all } i \text{ such that } S_i > 0 \\ \frac{\partial P_e(i)}{\partial S_i} < \theta & \text{all } i \text{ such that } S_i = 0 \end{array} \qquad (61)$$

where θ is a constant. Therefore, the solution is the *constant slope* of $P_e(i)$ as a function of S. The maximum achievable capacity is then given by $1 - \mathcal{H}(P_{e,m})$ [6] per transmitted symbol.

Example:

If binary phase shift keying (BPSK) is used to transmit information over the channel [8], then the probability of error at state i ($P_e(i)$) is given by

$$P_e(i) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{S_i}{N_i}}\right) \qquad i = 1,2 \qquad (62)$$

where S_i and N_i are the signal and noise power at state i, respectively and

$$\text{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\zeta^2} d\zeta.$$

Then $P_e(i)$ is convex function of S_i and less than half. The optimum power allocation policy can be shown to be

$$N_i S_i e^{\frac{2S_i}{N_i}} = \theta \qquad i = 1,2, \qquad (63)$$

for all those i such that $S_i \neq 0$, where as before θ is a constant.

End of Example

Direct generalizations of the above derivation is possible. First, it can readily be extended to more that two channel states. Secondly, (61) still results in maximum capacity if $P(S_i)$ is given by a general convex function of S_i which is always smaller than $1/2$ for all i. This does not impose any restriction as $P(S_i)$ is indeed a convex function of S_i.

Intuitively, as the block size of the transmitted symbol increases, the bit error probability in each transmitted symbol approaches the average probability of error. The capacity is

[6] $\mathcal{H}(P_{e,m})$ is defined as $\mathcal{H}(P_{e,m}) \triangleq P_{e,m}\log(P_{e,m}) + (1 - P_{e,m})\log(1 - P_{e,m})$.

then maximized when this average probability of error is minimized through intelligent allocation of signal power based on the state of the channel. Note that even though the behavior of the channel is similar to that of "average channel" - the channel codes used do not change with time - the improvement over the average channel can be expected through power assignment. At the same time, since the channel code used does not change with time, the rate at which the information is being transmitted is constant over the time and hence it is not possible to use similar type of optimization as was done in the previous two cases. Therefore, the solutions to the optimization problem (3) and (4) are the same and the separation principle holds. We conjecture that this is not only true for the binary symmetric channels considered here but also for more general channel classes such as the AWGN time-varying channels.

References

[1] T.M. Cover and J.A. Thomas, *Information Theory*, Wiley series in telecommunications, John Wiley, 1991.

[2] T.M. Cover, A. El Gamal and M. Salehi, "Multiple access channels with arbitrarily correlated sources," IEEE Trans. on Information Theory, Vol. 26, No. 6, November 1980.

[3] T.M. Cover, "Broadcast channel," IEEE Trans. on Information Theory, Vol. 21, No. 1, January 1972.

[4] W.H.R. Equitz and T.M. Cover, "Successive refinement of information," IEEE Trans. on Information Theory, Vol 37, No. 2, March 1991.

[5] A. Goldsmith, *Design and Performance of High-Speed Communication Systems over Time-Varying Radio Channels*, Ph.D. dissertation, University of California at Berkeley, 1994.

[6] N.S. Jayant and P. Noll, *Digital Coding of Waveforms, principles and applications to speech and video*, Prentice-Hall, Signal processing series, 1984.

[7] M. Minoux, *Mathematical Programming: Theory and Algorithms*, Wiley, 1986.

[8] P.Z. Peebles, Jr., *Digital Communication Systems*, Prentice-Hall, Englewood, New Jersey, 1987.

[9] L.H. Ozarow, S. Shamai, A.D. Wyner, "Information theoretic considerations for cellular mobile radio," IEEE Trans. on Vech. Technology, Vol. 43, No. 2, May 1994.

[10] C.E. Shannon, " Coding theorems for a discrete source with a fidelity criterion," IRE National Convention Record, Part 4, pages 142-163, 1959.

[11] C.E. Shannon, "Channels with side information at the transmitter," IBM Journal Research Development, Vol. 2, pp. 289-293, 1958.

[12] Y. Shoham and A. Gresho, "Efficient bit allocation for an arbitrary set of quantizers." IEEE Trans. on Acoust. Speech and Signal Processing, Vol. 36, No. 9, September 1988.

[13] TIA/EIA/IS-95, "Mobile-station-base-station compatibility standard for dual-mode wideband spread spectrum cellular system," Telecommunications Industry Association, July 1993.

[14] S. Vembu, S. Verdu and Y. Steinberg, "The source-channel separation theorem revisited," IEEE Trans. on Info. Theory, Vol. 41, No. 1, Jan. 1995.

[15] V.A. Vishampayan and N. Farvardin, "Joint design of block source codes and modulation signal sets," IEEE Trans, on Info. Theory, Vol. 38, July 1992.

[16] J.C. Walrand, P. Varaiya, "Optimal causal coding-decoding problems," IEEE Trans. on Info. Theory, Vol. 29, No. 6, November 1983.

Appendix

In this appendix some of the formal definitions and theorems related to the broadcast and the *physically degraded* broadcast channels for two receiver case are re-stated from [1].

Definition: *A broadcast channel consists of an input alphabet* $\mathcal{X}$ *and two output alphabets* $\mathcal{Y}_1$ *and* $\mathcal{Y}_2$ *and a probability transition function* $f(y_1, y_2|x)$.

Definition: *A broadcast channel is said to be physically degraded if*

$$f(y_1, y_2|x) = f(y_1|x)f(y_2|y_1).$$

Theorem 4 *If the rate pair* (R_1, R_2) *is achievable for a physically degraded broadcast channel, the rate triple* $(R_0, R_1, R_2 - R_0)$ *is achievable for the channel with common information. provided that* R_0*, the rate of the common information, is less than* R_2.

Note that R_0 can be arbitrary close to R_2.

Theorem 5 *The capacity region for sending independent information over the degraded broadcast channel* $X \to Y_1 \to Y_2$ *is the convex hull of the closure of all* (R_1, R_2) *satisfying*

$$\begin{aligned} R_2 &\leq I(U; Y_2) \\ R_1 &\leq I(X; Y_1|U) \end{aligned} \tag{64}$$

for some joint distribution $f(u)f(x|u)f(y_1, y_2|x)$.

Theorem 6 *All Gaussian broadcast channels are physically degraded and the capacity region* (R_1, R_2) *is given by*

$$\begin{aligned} R_1 &< \frac{1}{2}\log\left(1 + \frac{\alpha S}{N_1}\right) \\ R_2 &< \frac{1}{2}\log\left(1 + \frac{(1-\alpha)S}{\alpha S + N_2}\right) \end{aligned} \tag{65}$$

where $0 \leq \alpha \leq 1$ *is an arbitrary number and* S *is the power of the transmitted signal.*