Contents

Chapter 12

Multi-User Fundamentals

The last 4 chapters of this text investigate the important area of multiple-user communications systems. This chapter (12) establishes a few bounds and basic models and concepts that characterize multi-user transmission. The goal of this chapter is establishment of a familiarity with some basic multi-user procedures in simpler cases, with later chapters exploring more detailed implementation.

Section 12.1 begins with a definition of a user before proceeding with a discussion of the rate region, a generalization to more than one user of Chapter 8's capacity. Multi-user detection is also summarized in Section 12.1. A general characterization of capacity regions $c(\boldsymbol{b})$ concludes Section 12.1, particularly introducing user order as a missing element in previous multi-user capacity-region characterization. The user-order concept allows all multi-user capacity regions to be specified in terms of elementary mutual information and set operations. Section 12.2 proceeds to the important multiple-access channels where several users' signals converge on a single receiver, developing successive decoding as the logical subset of Chapter 5's generalized decision feedback. Successive decoding can be used to simplify the general capacity region of Section 12.1 specifically characterizing the possible sums of users' rates that can be achieved. A dual to the multiple-access channel is the broadcast channel of Section 12.3. Broadcast channels transmit from a common modulator to multiple (often physically separated) receivers. Section 12.3 simplifies the capacity region for the broadcast channel. Section 12.4 concludes with interference channels where multiple users' signals simultaneously occupy a single channel, but both transmitters and receivers are separated for all users.

12.1 Multi-user Channels and Bounds

Transmission channels often accommodate more than one user. Such channels in general are called "multi-user" channels.[1] At the physical layer or modulation layer, there are 3 basic configurations of interest: Figures 12.1, 12.2, and 12.3 illustrate respectively the multiple-access, broadcast, and interference multi-user communication channels for U users, $u = 1, ..., U$. They are respectively distinguished by coordination, or lack thereof, of users signals at the transmitter, receiver, or both. Other multi-user channels can be represented as some combination of these 3 basic types. This text's use of the term "multi-user" forces all user's transmitters to be independent of any receiver's output (decoding) of any user's message. Specifically then "relay" channels are not directly considered where a chain of intermediate users can successively pass the same message from a first user to an ultimate intended recipient of that message. These relay channels are instead a special case of an enlarged multi-user channel as discussed in Section 12.1.4. This user-independent-of-receiver restriction is important as it allows a characterization of the capacity region of all multi-user channels - a characterization that appears elusive without this restriction. This section of this chapter generally characterizes the achievable data rates (Subsections 12.1.1 and 12.1.3) and the detection methods (Subsection 12.1.2) that apply to all 3 types of multi-user channels.

In single-user communication, the concept of a user is obvious, but there are subtleties in multi-user communication as to what constitutes a user. In some sense, failure to recognize the importance of definition of a user has lead many researchers to conclude that some multi-user channels (like the interference channel) have unknown bounds (or unknown capacity regions, where capacity regions are defined in Subsection 12.1.1) – however, with the proper definition of users, these bounds can be determined as in this chapter. A natural tendency would be to associate a transmitter in the MAC or IC or a receiver in the BC or IC with a user, and this is typically the case. However, it may be possible that one or more of the up to U receivers can detect, with vanishingly small error probability, some part (sub-message) of one of up to U transmitters' messages but not the entire message. Should this happen for any given set of inputs and channel, then the corresponding message should be decomposed into multiple users for which the smallest intersections of detectable message components now become these new users. Thus, the number of transmitters or receivers may be a lower bound on the number of users for general multi-user channels. The original users' rates can be recomputed easily as in Section 12.1.5. The concept that a user is either reliably detectable or not is fundamental to the construction of capacity regions in this text. For Gaussian inputs, such decomposition degenerates into the U original Gaussian transmit signals. For other input distributions, the decomposition may be complicated. An example would be the sum of a binary sequence of ± 1 and a Gaussian code. If the variance of the Gaussian code is less than $1/3$ ($\text{SNR} = 3$ or 4.7 dB), then the binary component may be detectable at some receivers, but the Gaussian component not also detectable for at least one receiver. In this case, the transmitter would be decomposed into a simple 2-user sub-broadcast channel.

> **Definition 12.1.1 (user)** *A* **user** *in a multi-user channel is defined by the smallest non-zero information-bearing component, that is $H\boldsymbol{x}_u \neq 0$, that can be reliably detected with vanishingly small probability of error at one or more receivers in the multi-user channel. Thus, this definition of a user can depend on both the channel probability distribution $p_{\boldsymbol{y}/\boldsymbol{x}}$ and the input distribution $p_{\boldsymbol{x}}$.*

The definition of a user is channel- and input-dependent in this text. Again, Section 12.1.5 shows how a single user may be decomposed into several sub-users, increasing U, but then adding these sub-users' rates back into a single user rate.

The single receiver of Figure 12.1's **multiple-access channel (MAC)** accepts signals from U physically separated transmitters, each with its own transmit symbol vector $\boldsymbol{x}_u$, $u = 1, ..., U$. The single receiver processes the single channel-output vector $\boldsymbol{y}$. There are thus multiple users accessing a multi-input channel with a single output, whence the name "multiple-access" channel. The single output vector $\boldsymbol{y}$ (as well as each of the input vectors) may have $N \geq 1$, thus creating a **vector multiple**

[1]This text use a hyphen in "multi-user" instead of writing the compound "multiuser" often found elsewhere as the author could find no evidence that such a word exists.

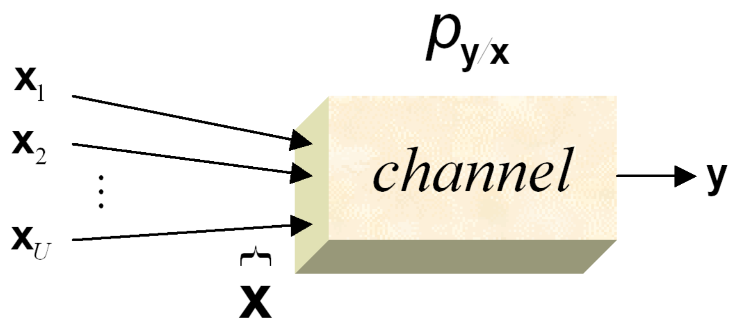

Figure 12.1: The multiple-access channel.

access channel. It will be assumed that all users have the same number of dimensions – when this is not the case physically, dummy unused dimensions can be augmented to transmit symbols or receiver vectors as necessary so that N is determined as the maximum number of dimensions used by any user. In multiple-access channels, the single output allows a single "coordinated" receiver to detect all the transmitted messages. The use of a single receiver also simplifies analysis of a MAC. The inputs cannot be coordinated or co-generated in the multiple-access channel. If a single user could be decomposed into two or more transmit-cooperating users for some $p_{\boldsymbol{x}}$, then the channel ceases to be a MAC. If such a single user decomposes into two or more separated transmitters, then the channel remains a MAC but with a larger value for U.

Multiple-access channels abound in communications, often occuring where several subscribers communicate to a central service provider: For instance, a wireless channel may have several user transmitters that each share a common frequency band in transmitting "uplink" to an "access point" or "base station." Such architectures occur in both cell phones (GSM, CDMA, all generations) and in higher-speed wireless data transmission with WiFi (wireless hi-fidelity) or WiMax, LTE, LT-advanced networks. The upstream[2] direction of a cable or passive-optical network is also an example of a multiple-access channel with various residential customers all sharing a common frequency band for transmissions to a central "hub" receiver. Upstream DSL systems form an interesting vector multiple-access channel when the twisted pair packed together in a binder of wires crosstalk electromagnetically into one another. Yet another example is a disk drive where a single receive (read) head (or an array of such read heads) may sense the signals (simultaneously) of several previously written adjacent tracks.

Of interest in multiple-access channels are the maximum achievable data rates for each user (which may be a function of the data rates selected by the other users, complicating design and analysis). Section 12.2 provides some basic bounds and methods that will be specific to and simplified for the MAC. Section 12.2 also includes a simplified description and construction of the so-called "rate region" for the MAC. Chapter 13 studies the MAC in more detail.

The dual of the multiple-access channel is Figure 12.2's **broadcast channel (BC)** in which a single transmitter generates the channel input $\boldsymbol{x}$. The BC also has U physically separated outputs $\boldsymbol{y}_u$. When the single input is a vector, the channel is **vector broadcast**. In the BC, each of the user signals are embedded within the single transmitter's channel-input symbol $\boldsymbol{x}$, and so "coordination" of transmit signals – i.e., co-generation of the single input with knowledge of all the users' message signals is possible. Coordinated reception of the U outputs is not possible in a broadcast channel. Section 12.3 provides some bounds and techniques for the BC, while Chapter 14 studies the broadcast channel in more detail. If some user associated with a receiver location can be decomposed into more than one user with cooperating receivers, then the channel is no longer a BC. If such a user decomposes into two isolated receivers, then the channel remains BC with a increased value for U.

Examples of broadcast channels would be the opposite direction of transmission for multiple-access

[2]Upstream is the term used for customers' transmissions to a central site in cable and DSL, while "uplink" is the term used in wireless even though it is not a word (upstream is a word).

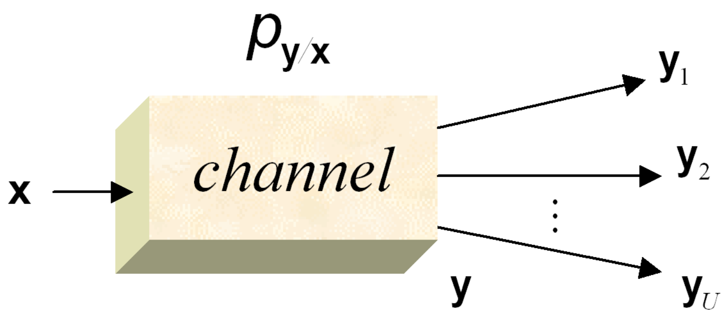

Figure 12.2: The broadcast channel.

channels from a service provider to a customer: For instance a "downlink" in wireless or the "downstream directions" in cable or DSL networks. Broadcast channels also occur for television and radio where no reverse multiple-access-like channel occurs (usually).

The third channel type is the **interference channel (IC)** of Figure 12.3. No coordination of either inputs nor outputs is possible. Each of U possible receivers can attempt independent detection of one or of all the U transmitted messages. In the **vector** interference channel, some of the inputs and/or outputs, are vectors. Interference channels occur when multiple transmitters and receivers share a frequency band (wireless) or medium (wire-line). Some examples include some types of home networks that share unregulated bands and establish links on an ad-hoc basis (some wireless ultra wideband, power-line home-plug, or home-phone that establish connections by ad-hoc exploration of "who is out there?"). Military and espionage channels may also then fall into the category of interference channel where loss of a central coordinating receiver or transmitter would not result in loss of all communication. The IC may be enlarged to U^2 users since any of the U transmit locations might more generally be attempting to transmit non-zero information content to any of the receivers. If any of the users may be decomposed into two or more users that could coordinate at either or both of the transmitter and receiver, then the channel is no longer an interference channel. If such a decomposition retains independence of transmit locations and receive locations for each component user, then the overall channel remains an IC with yet a larger value (maybe in excess of U^*) for the number of users.

This text also defines a **distributed-control interference channel DCIC** for situations with Gaussian noise. The DCIC imposes a design restriction that the design of the transmit signals for any one user may not know the channel descriptions from the other users' inputs to their outputs nor the crosstalking channel descriptions; the modem receiver can only identify its channel, and the other users' signals are viewed only as noise (known only in terms of total received power spectra into that modem's receiver). The DCIC basically models "unbundled" or "unlicensed" situations where several competing services or users may share a common facility, but will not share information. Section 12.4 investigates the IC bounds, while Chapter 15 investigates some approaches to interference channels and DCICs.

Linear Additive-noise multi-user channel models

The linear additive noise channel (particularly with Gaussian noise) is particularly important and merits some additional consideration in this subsection. Table 12.1 lists the possible dimension sizes of inputs and outputs for each of the simple[3] linear multi-user channels of the additive-noise form

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n} \quad . \tag{12.1}$$

[3]Simple means that $N = 1$ or 2 (or in a few cases infinite with no finite values in between). Intermediate values are considered in Chapter 13.

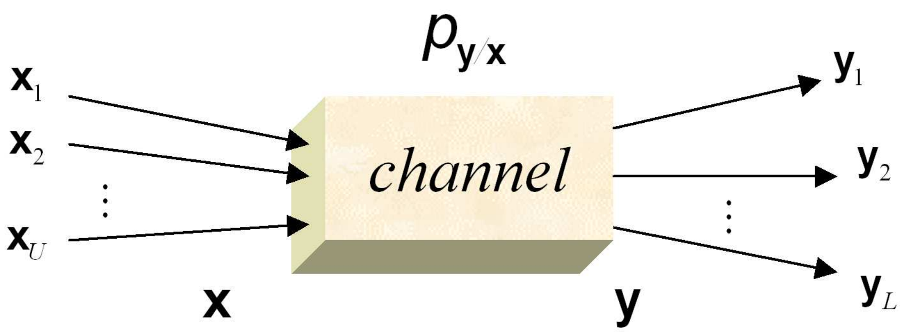

Figure 12.3: The interference channel.

chan	$\boldsymbol{x}$ Number of inputs	$\boldsymbol{y}$ Number of outputs	H
multiple access	U	1	$[H_U \ldots H_2\ H_1]$
broadcast	1	U	$\begin{bmatrix} H_U \\ \vdots \\ H_2 \\ H_1 \end{bmatrix}$
interference	U	U	$\begin{bmatrix} H_{UU} & \ldots & H_{U1} \\ \vdots & \ddots & \vdots \\ H_{2U} & \ldots & H_{21} \\ H_{1U} & \ldots & H_{11} \end{bmatrix}$

Table 12.1: Table of dimensionality for the multi-user Gaussian channel $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n}$.

Chapters 12 through 15 will use the designation H for channels, and this will presume to describe the channel and any part of transmitter or receiver filters that cannot be changed by designers.[4] Later chapters revisit this model with multi-dimensional users. This Chapter (12) largely addresses only simple capacity/bound results for the $U \times 1$ multiple-access, $1 \times U$ broadcast, and $U \times U$ "scalar" interference channels. Some level of intersymbol interference is addressed in an asymptotic sense here, but extensions to the more practical finite-length packets and possibly vector-user inputs and vector-user outputs for all channels occur in Chapters 13-15.

Multi-user transmission channels are characterized in the same way as are single-user channels. The general characterization of any channel (multiple-user or not) remains a conditional probability distribution $p_{\boldsymbol{y}/\boldsymbol{x}}$. The input is characterized by a probability distribution $p_{\boldsymbol{x}}$. All other probability distributions

[4]This "H" was sometimes called a pulse response and the letter "P" used previously in Chapters 3, 4, and 5. There will be potential transmit filter designs for any given H as was seen in Chapters 4 and 5, but the use of "P" was continued there because of the proximity to Chapter 3 and intersymbol-interference discussions. The reader at this point is presumed sufficiently mature to model the channel without continued discussion here of every detail in so doing.

are consequently determined. For the IC and MAC, the input distribution must factor as

$$p_{\boldsymbol{x}} = \prod_{u=1}^{U} p_{\boldsymbol{x}_u} \quad . \tag{12.2}$$

The factorization of input probability distributions for MAC and IC holds for any channel model (and not just additive noise). In some situations for the BC, the channel output distribution may also factor, but this is not necessarily always true. While the users' input information "bit" streams may be independent, their corresponding input components need not necessarily be independent. Independence of $\boldsymbol{x}_u$ for the linear AWGN can happen for certain inputs jointly with what is known as worst-case additive noise of Section 5.5, an important case later developed in Chapter 14.

12.1.1 Data Rates and Rate Regions

This section generalizes the concept of a data rate to the more general concept of a rate region. Two types of mutual information are of interest in a multi-user channel: the **sum of the rates of all the users** $I(\boldsymbol{x};\boldsymbol{y})$ and the **individual user mutual information** $I(\boldsymbol{x}_u;\boldsymbol{y})$, $u = 1, ..., U$.[5] $I(\boldsymbol{x};\boldsymbol{y})$ is thus a **maximum sum rate** for the given input distribution to the channel. If each user has bits per symbol b_u, $u = 1, ..., U$, then

$$\sum_{u=1}^{U} b_u \le I(\boldsymbol{x};\boldsymbol{y}) \quad , \tag{12.3}$$

with the inequality constraint imposed because inputs, outputs, or both may not be coordinated.[6] Also, the **best average rate** is

$$I(\boldsymbol{x}_u;\boldsymbol{y}) \le I(\boldsymbol{x};\boldsymbol{y}) \quad , \tag{12.4}$$

where the other $U-1$ users are averaged over their distributions in calculation of $I(\boldsymbol{x}_u;\boldsymbol{y})$. The distributions used in 12.34 are

$$p_{\boldsymbol{x}_u} = \int_{\boldsymbol{v} \in \boldsymbol{x} \setminus \boldsymbol{x}_u} p_{\boldsymbol{x}}(\boldsymbol{v}) d\boldsymbol{v} \tag{12.5}$$

and

$$p_{\boldsymbol{x}_u,\boldsymbol{y}} = \int_{\boldsymbol{v} \in \boldsymbol{x} \setminus \boldsymbol{x}_u} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{v}) d\boldsymbol{v} \quad . \tag{12.6}$$

The operation indicated by $\setminus$ means to remove an element (or subset) from a larger set. $I(\boldsymbol{x}_u;\boldsymbol{y})$ may be conditioned upon a set of other users $\boldsymbol{u} \subseteq \{1, ..., U\}$ to obtain the **conditional average rate** $I(\boldsymbol{x}_u;\boldsymbol{y}/\boldsymbol{x_u})$. The conditional mutual information is of interest when other users' ($\boldsymbol{u}$'s)signals might be first detected or somehow known and then used to remove those other signals degrading effect upon subsequent detection of the user of interest.

> **Definition 12.1.2 (Detectable prior user set for user u and output $\boldsymbol{y}$)** *The set $\mathcal{U}_u(\boldsymbol{y}) \subset \boldsymbol{U}$ includes any and all other users $u' \neq u \in \boldsymbol{U}$ that are decodable from multi-user channel output vector $\boldsymbol{y}$. (Decodable means that with $P_{e,u} \to 0$.) $\mathcal{U}_u(\boldsymbol{y})$ is tacitly also a function of $p_{\boldsymbol{y}/\boldsymbol{x}}$ and $p_{\boldsymbol{x}}$.*

This subset of users essentially represents at any particular output $\boldsymbol{y}$ (which could be written $\boldsymbol{y}_i, i = 1, ..., U$ for the BC and IC) a set of other users who could be detected reliably and removed before user u is decoded. This subset need not be unique, because it depends on the choice of input distribution. This approach to identifying decodable other users is the reason for the definition of a user in Definition 12.1.1.

[5] The individual mutual information rates for a particular output $I(\boldsymbol{x}_u;\boldsymbol{y}_u)$ are also of interest in the interference channel.

[6] There is a presumption of a common symbol period in multi-user communication that essentially assumes a synchronization that may not be present in practice. However, a sufficiently long symbol interval may always be defined so that essentially all users conform to it. Introduction of multiple symbol rates or actual data rates obfuscates basic principles and adds little to the discussion, but such a constraint would need consideration in practice.

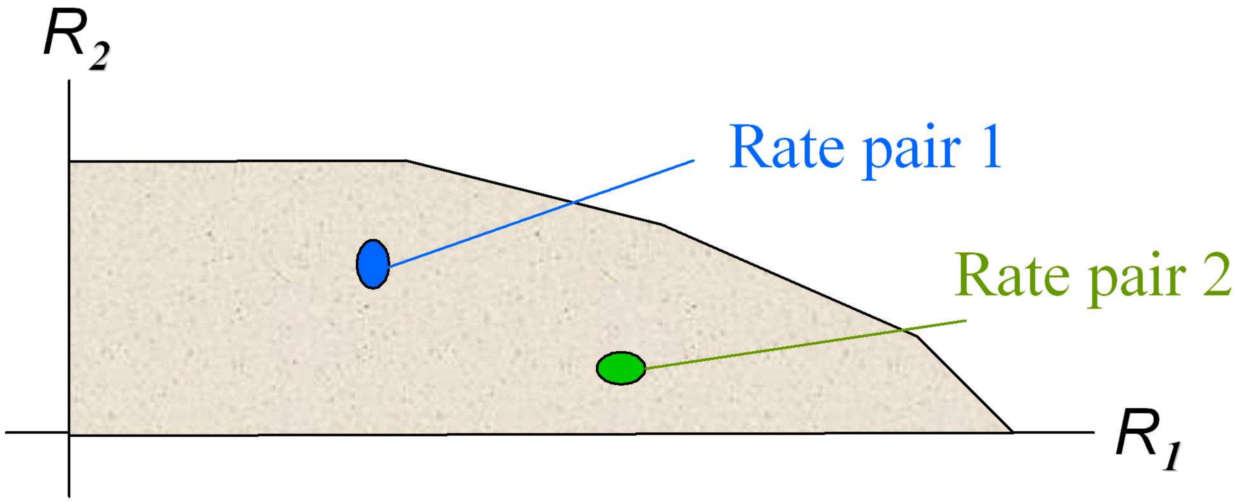

Figure 12.4: Rate Region.

All three mutual-information quantities, $I(\boldsymbol{x};\boldsymbol{y})$, $I(\boldsymbol{x}_u;\boldsymbol{y})$, and $I(\boldsymbol{x}_u;\boldsymbol{y}/\boldsymbol{x}_{\boldsymbol{u}})$, may be of interest to a particular user u. The mutual information $I(\boldsymbol{x}_u;\boldsymbol{y})$ is of interest to the u^{th} particular user, but presumes a probability distribution on all users. It is a data rate average treating all other users as noise. Thus, this data rate $I(\boldsymbol{x}_u;\boldsymbol{y})$ tacitly depends on all the other users and may be significantly affected by the other users' choices of input distributions. The maximum of $I(\boldsymbol{x}_u;\boldsymbol{y})$ found over all choices of input probability distributions for $\boldsymbol{x}_u$ is a capacity, but this capacity remains a function of the choices of the codes/distributions of the other users. Furthermore, this bound does not account for the potential collaboration between users (for instance, one user could be decoded first and its effect subsequently subtracted at the receiver). So it is possible for b_u to exceed $I(\boldsymbol{x}_u;\boldsymbol{y})$, but never to exceed $I(\boldsymbol{x};\boldsymbol{y})$. $I(\boldsymbol{x}_u;\boldsymbol{y}/\boldsymbol{x}_{\boldsymbol{u}})$ attempts to characterize the prior removal of other users ($\boldsymbol{u}$) by a decoder, but there are 2^{U-1} possible choices for the subset $\boldsymbol{u}$ along with a check to see if those other users could be reliably detected. The bound of **best average conditional rate** is

$$\max_{\boldsymbol{u}\subseteq\mathcal{U}_u(\boldsymbol{y})} b_u \leq I(\boldsymbol{x}_u;\boldsymbol{y}/\boldsymbol{x}_{\boldsymbol{u}}) \tag{12.7}$$

where the maximization is over the possible subsets $\mathcal{U}_u(\boldsymbol{y})$ of $\boldsymbol{U}$ that contain only all those other ($u' \neq u$) users that can be first decoded without error. Thus, the discerning of best data rate is more complicated than in the single-user case. Equation (12.7) is key, again depending on Definition 12.1.1, and allows capacity region specification to follow.

Thus, there may be several "capacities" in multi-user transmission, unlike single-user transmission where the single-user's data rate is upper bounded by only one capacity. The interdependencies among users complicates the interpretation of theoretical bounds on maximum performance with a range of consequent contingencies. The general approach to characterizing the proliferation of capacities for a multi-user channel is the capacity rate region, such as generally illustrated for two users in Figure 12.4. The rate region is a plot of all the possible U-tuples of data rates that can be achieved by all the users. The points on the outer boundary represent best multiple-user code designs. Interior points represent systems using codes that have performance below best. Rate regions vary with the type of channel, and may be difficult to construct. The end of this section provides a general capacity rate region that applies to all $U \times 1$ multiple-access, $1 \times U$ broadcast, and $U \times U$ interference channels, as well as any (non-relay) multi-user channels as defined in this text.

Definition 12.1.3 (Capacity Rate Region) *The capacity region $C(\boldsymbol{b})$ of any multiple-user channel is defined as the largest union (or convex hull) of all U-dimensional bits/symbol* *vectors* $\boldsymbol{b} = [b_1,\ b_2,\ ...,\ b_U]$ *such that any of the receiver(s) may decode its user(s) of interest with arbitrarily small probability of error.*

Furthermore, any point outside this region results in at least one of the receivers having probability of error bounded away from 0 no matter what coding method is used.

The capacity region is less formal in definition than the more precise mutual-information definition used for single-user channels earlier in Chapter 8. The following lemma is useful and perhaps immediately obvious:

Lemma 12.1.1 (Convexity of Capacity Rate Region) *The capacity rate region is convex, meaning that any convex combination of two rate tuples within the region produces another rate tuple within the region. Mathematically, if* $\boldsymbol{b}_\alpha \in C(\boldsymbol{b})$, $\boldsymbol{b}_\beta \in C(\boldsymbol{b})$, *and* $\alpha + \beta = 1$, *then* $\alpha \cdot \boldsymbol{b}_\alpha + \beta \cdot \boldsymbol{b}_\beta \in C(\boldsymbol{b})$.

Proof: *The basic proof follows from the definition: The convex combination of two rate tuples corresponds to dimension ("time") sharing of the two codes, and the corresponding receivers, for the same fraction of dimensions (time slots) as used in the convex combination. Such a system is an allowable implementation, and thus corresponds to a point that can be realized within the capacity region.*

Technically, the concept of "time-sharing" or dimensional-sharing of two different single-user message components requires a sub-division of that user into two (or more) users according to the definition of a user in Definition 12.1.1. Subsection 12.1.5 deals with recombining users into larger "macro" users of several independent components.

12.1.2 Optimum multi-user detection

The optimum detector for a multi-user channel is a generalization of the optimum detector for the single-user channel in Chapter 1. The set of all possible multi-user channel inputs is denoted $\mathcal{C}_{\boldsymbol{x}}$, and contains $M = |\mathcal{C}_{\boldsymbol{x}}|$ possible distinct symbols, which may be a large number for the multiple-user channel. $\mathcal{C}_{\boldsymbol{x}}$ is thus a signal constellation, equivalently a code, for the set of all users that represents the aggregation of the individual users' codes, most generally described by $p_{\boldsymbol{x}}$.

Theorem 12.1.1 (Optimum multi-user detection - maximum à posteriori) *The probability of multi-user-symbol error is minimum when the detector selects* $\hat{\boldsymbol{x}} \in \mathcal{C}_{\boldsymbol{x}}$ *to maximize* $p_{\boldsymbol{x}/\boldsymbol{y}}$ *and is known as the* **maximum à posteriori** *multi-user detector. When all possible multi-user input symbol values are equally likely, this optimum detector simplifies to maximization of the conditional probability* $p_{\boldsymbol{y}/\boldsymbol{x}}$ *over the choice for* $\hat{\boldsymbol{x}} \in \mathcal{C}_{\boldsymbol{x}}$, *and is called the* **maximum likelihood** *multi-user detector.*

Proof: See Chapter 1. **QED.**

For the broadcast and interference channels, each physically distinct output $\boldsymbol{y}_l$ replaces the general $\boldsymbol{y}$ in the above Theorem 12.1.1. Thus, each receiver optimally detects each of the inputs (with possibly different performance for each user $\boldsymbol{x}_i, i = 1, ..., U$ and/or at each receiver $\boldsymbol{y}_u$ for $u = 1, ..., U$). Only user $u = i$ may be of interest at receiver $\boldsymbol{y}_u$ in such detection. The probability of error then becomes a function of u that is not shown explicitly here.

The overall probability of error for detecting all users is (as always)

$$P_e = 1 - P_c = 1 - \sum_{i=0}^{M_{\boldsymbol{x}}-1} P_{c/i} \cdot p_{\boldsymbol{x}}(i) \quad . \tag{12.8}$$

This probability has meaning for all multi-user channels, but in particular may be of most interest in the MAC where a single receiver detects all users.

For the BC and the IC, the probability of error for such a MAP receiver would be a function of which receiver; in particular a function of u for the u^{th} channel output. Such a detector observes only $\boldsymbol{y}_u$ then minimizes the likelihood that the receiver for user u may have incorrectly detected user u's message:

$$P_e(u) = 1 - P_c(u) = 1 - \sum_{i=1}^{M} P_{c/i}(u) \cdot p_i(u) \tag{12.9}$$

where $P_{c/i}(u)$ is the probability that user u's i^{th} possible message is correctly detected from $\boldsymbol{y}_u$. Equation (12.9) also could vary with the user index at each output – that is, each receiver may have different $P_c(u, j)$'s for each of the users j, but user u is probably most of interest.

The users are often modeled as being independent in their choice of transmit messages so that

$$\mathrm{p}_{\boldsymbol{x}} = \prod_{u=1}^{U} p_{\boldsymbol{x}_u} \quad . \tag{12.10}$$

Then, a MAP decoder simplifies to an ML decoder if each of the users independently also has a uniform distribution. The independent-inputs assumption is always made on the MAC and IC.

An ML decoder has a probability of error that is approximated by the NNUB for the AWGN channel

$$P_e \leq N_e \cdot Q(\frac{d_{\min}}{2\sigma}) \quad , \tag{12.11}$$

where the number of nearest neighbors, N_e, now includes all multi-user-symbol values in the calculation and similarly the minimum distance is calculated over the entire set of all multi-user symbol values. However, such a decoder only has interest for the MAC. Again (12.11) can vary for each receiver with broadcast and interference channels.

Definition 12.1.4 (Crosstalk-Free Channel) *A* **Crosstalk-free multi-user channel (CFC)** *has a conditional probability distribution that satisfies*

$$p_{\boldsymbol{y}/\boldsymbol{x}} = \prod_{u=1}^{U} p_{\boldsymbol{y}_u/\boldsymbol{x}_u} \quad . \tag{12.12}$$

That is, the channel probability distribution factors into independent terms for each of the users. When the channel is not CFC, it is said to have **crosstalk***.*

A receiver for a CFC trivializes into an independent set of single-user receivers:

Theorem 12.1.2 (Independent Detection) *The ML decoder for the CFC is equivalent to a set of independent optimum decoders for each individual user.*

The proof follows trivially from inspection of $\mathrm{p}_{\boldsymbol{y}/\boldsymbol{x}}$, which factors into U independent terms that can each be independently maximized to implement the ML detector (and ML detection tacitly assumes that (12.10) holds.)

Independent detection means that the designer can use a separate receiver for each user, potentially then enormously simplifying the detector implementation. The absence of crosstalk essentially renders the channel to be a set of independent single-user channels. Such a receiver is analogous to a symbol-by-symbol detector for an ISI-free channel, except now more appropriately termed perhaps "user by user." Orthogonal multiplexing in Chapter 5's Appendix A attempts to ensure an CFC for channels with neither ISI nor crosstalk. Such systems are the norm in early multi-user transmission designs, but the assumption of an CFC may not be true especially when users are not well coordinated or channels are not accurately known during design. Transformation of the channel into a CFC by linear matrix multiplication (like "equalization") is not necessarily a mechanism that enables the best points in the rate region to be achieved, because sharing of dimensions by users may provide larger data rates.

The overall probability of error on the CFC channel for all users is (a receiver index l is not shown but would normally be used in a BC or IC)

$$P_e = 1 - \prod_{u=1}^{U} P_{c,u} \quad , \tag{12.13}$$

and the overall probability of error can never be less than the probability of error for any one of the users

$$P_e \geq P_{e,u} \; \forall \; u \in \{1, ..., U\} \quad . \tag{12.14}$$

The individual user probability distribution can be computed directly from the overall conditional distribution according to

$$\mathrm{p}_{\boldsymbol{x}_u/\boldsymbol{y}} = \int_{\boldsymbol{x} \setminus \boldsymbol{x}_u} \mathrm{p}_{\boldsymbol{x}/\boldsymbol{y}} d\{\boldsymbol{x} \setminus \boldsymbol{x}_u\} \quad . \tag{12.15}$$

The conditional distribution can thus be computed from known quantities, using also the fact that

$$\mathrm{p}_{\boldsymbol{x}/\boldsymbol{y}} = \frac{\mathrm{p}_{\boldsymbol{y}/\boldsymbol{x}} \cdot \mathrm{p}_{\boldsymbol{x}}}{\mathrm{p}_{\boldsymbol{y}}} \quad . \tag{12.16}$$

Equivalently, the individual ML detector for $\boldsymbol{x}_u$ given $\boldsymbol{y}$ uses the distribution

$$\mathrm{p}_{\boldsymbol{y}/\boldsymbol{x}_u} = \int_{\boldsymbol{x} \setminus \boldsymbol{x}_u} \mathrm{p}_{\boldsymbol{y}/\boldsymbol{x}} \cdot \mathrm{p}_{\boldsymbol{x} \setminus \boldsymbol{x}_u} \cdot d\{\boldsymbol{x} \setminus \boldsymbol{x}_u\} \quad . \tag{12.17}$$

Detection might also be given some of the other users signals if these other users' signals could be and were decoded first. Two definitions are necessary to simplify such conditional detection: First, an order of user detection, sometimes denoted π, a one-to-one mapping of user indices onto themselves where $\pi(1)$ is the first decoded user, $\pi(i)$ is the i^{th} decoded user and $\pi(U)$ is the last decoded user. There are $U!$ possible orders of user detection for any particular user's receiver. Second, there is a set of users $X_\pi(u, \boldsymbol{y})$ for user u, output vector $\boldsymbol{y}$, and each order π such that all $\boldsymbol{x}_i \in X_\pi(u)$ that occur before user u in order π (that is, $\pi(i) < \pi(u)$) that can be decoded with arbitrarily small probability of error for the given input multi-user probability distribution $p_{\boldsymbol{x}}$

$$X_\pi(u, \boldsymbol{y}) = \left\{ \boldsymbol{x}_{\pi(i)} \mid \boldsymbol{x}_{\pi(i)} \in \{\mathcal{U}_u(\boldsymbol{y})\} \wedge \pi(i) < \pi(u) \; \right\} \quad . \tag{12.18}$$

$X_\pi(u, \boldsymbol{y})$ differs from $\mathcal{U}_u(\boldsymbol{y})$ in that it is also a function of the given order and includes all decodable users prior to $\pi(u)$. Thus, $X_\pi(u, \boldsymbol{y})$ is one of the all the possible $\mid \mathcal{U}_u(\boldsymbol{y}) \mid$ elements that could occur for user u and channel output $\boldsymbol{y}$. The inclusion of order makes this a single unique set. The set $\boldsymbol{U} \setminus [u \cup X_\pi(u, \boldsymbol{y})]$ is the other-users "noise" in detecting user u from channel output $\boldsymbol{y}$.

EXAMPLE 12.1.1 (sum of 3 Gaussian signals and noise) Three Gaussian signals are summed with noise to form

$$y = x_1 + x_2 + x_3 + n \quad . \tag{12.19}$$

The energies are $\sigma_n^2 = .001$, $\mathcal{E}_1 = 3.072$, $\mathcal{E}_2 = 1.008$, and $\mathcal{E}_3 = .015$. Gaussian codes ($p_{\boldsymbol{x}}$ Gaussian) are selected so that $b_1 = 1$, $b_2 = 3$ and $b_3 = 2$. There are 6 possible orders of decoding in this case as in the table below. For each order π, $X_\pi(1)$, $X_\pi(2)$, and $X_\pi(3)$ are listed. The first number in the ordered triple for order is the first signal that would be attempted to be decoded (and must be decoded in the presence of all other users as noise), the second is the second attempted to be decoded (and for which possibly the first user in the order was decoded if the SNR was high enough), and the last is the last to be decoded (allowing possibly for up to both of the previous users to possibly be decoded).

Order π	$X_\pi(1)$	check decode of $pi(1)$	$X_\pi(2)$	check decode of $\pi(2)$	$X_\pi(3)$
(1,2,3)	$\emptyset$	$0.5 \cdot \log_2\left(1 + \frac{\mathcal{E}_1}{\mathcal{E}_2+\mathcal{E}_3+\sigma^2}\right) = 1 \geq b_1$	1	$0.5 \cdot \log_2\left(1 + \frac{\mathcal{E}_2}{\mathcal{E}_3+\sigma^2}\right) = 3 \geq b_2$	1,2
(1,3,2)	$\emptyset$		1	$0.5 \cdot \log_2\left(1 + \frac{\mathcal{E}_3}{\mathcal{E}_2+\sigma^2}\right) = .016 < b_3$	$\emptyset$
(2,1,3)	$\emptyset$	$0.5 \cdot \log_2\left(1 + \frac{\mathcal{E}_2}{\mathcal{E}_1+\mathcal{E}_3+\sigma^2}\right) = .2 < b_2$	$\emptyset$		$\emptyset$
(2,3,1)	$\emptyset$		$\emptyset$		$\emptyset$
(3,1,2)	$\emptyset$	$0.5 \cdot \log_2\left(1 + \frac{\mathcal{E}_3}{\mathcal{E}_1+\mathcal{E}_2+\sigma^2}\right) < b_1$	$\emptyset$		$\emptyset$
(3,2,1)	$\emptyset$		$\emptyset$		$\emptyset$

For this set of Gaussian signals as characterized by their energies, it is better to have user 1 earlier in the order because only user 1 can be decoded in the presence of both (or either) the other two users as noise.

Thus, the ML decoder then maximizes over the values for $\boldsymbol{x}_u$

$$p_{\boldsymbol{y}/\boldsymbol{x}_u;X_\pi(u)} = \int_{\boldsymbol{x}\setminus X_\pi(u)} \mathrm{p}_{\boldsymbol{y}/\boldsymbol{x}} \cdot p_{\boldsymbol{x}\setminus X_\pi(u)} \cdot d\{\boldsymbol{x} \setminus X_\pi(u)\}\} \quad . \tag{12.20}$$

The detector for the order π^* with smallest $P_{e,u}$ would then be the best for that particular order. The same order, however, is not necessarily the best for all users, so one might write the best order as $\pi^*(u)$.

Detection for the linear multi-user AWGN

The linear multi-user AWGN is

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n} \quad . \tag{12.21}$$

For detection of only user u, it may be that the overall minimum distance in Equation 12.11 is too small. That is, a single fixed value for $\boldsymbol{x}_u$ may correspond to the two multi-user codewords that determine the overall $d_{\min}$. So, this text defines a

$$d_{\min,u} = \min_{\boldsymbol{x}\neq\boldsymbol{x}' \wedge \boldsymbol{x}_u \neq \boldsymbol{x}'_u} \|H(\boldsymbol{x}-\boldsymbol{x}')\| \quad . \tag{12.22}$$

Trivially,

$$d_{\min,u} \geq d_{\min} \tag{12.23}$$

with equality if and only if any codewords in $\mathcal{C}_{\boldsymbol{x}}$ corresponding to the overall $d_{\min}$ also correspond to different values for the u^{th} user's symbol contribution. Also,

$$\min_u d_{\min,u} = d_{\min} \quad . \tag{12.24}$$

This more clearly illustrates how it is possible for a detector extracting a single user to have better performance than one that extracts all users.

Sequences can be handled in a straight forward extension of the channel to

$$\boldsymbol{Y}(D) = H(D) \cdot \boldsymbol{X}(D) + N(D) \tag{12.25}$$

where all the vector/matrix D-transforms are defined by finding the D transform of each element and D corresponds to a delay of one symbol period (which may be a large packet of data).

EXAMPLE 12.1.2 (2 users in 2 dimensions) Two users both use 4-QAM with identical symbol periods, but different energy levels. The combined constellation when these two signals are added is shown in Figure 12.5. The users' symbol contributions are not orthogonal, and both occupy both dimensions. The first signal has the smaller $d_{min,1}$, which is equal to the overall $d_{\min}$. The second signal has a larger $d_{min,2}$ and whose values corresponds to the distance between the 4 colors in the figure.

A detector for both signals simply selects the closest of the 16 points to a received two-dimensional value $\boldsymbol{Y}$ and then outputs the corresponding two messages. The probability of

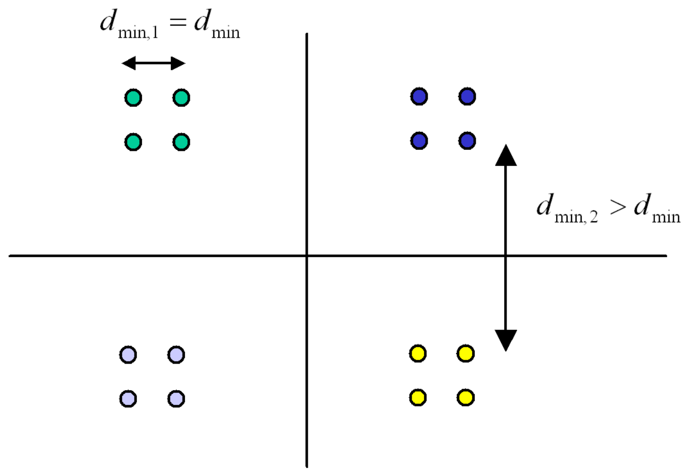

Figure 12.5: Illustration of two-dimensional two-user channel constellation.

error for the first user and for the overall detector are approximately the same. However a detector for only the second user (the different colors) clearly would perform significantly better. In this simple example, a detector that assumed user 1 was Gaussian noise have the same decision regions as the optimum detector for user 2, but such simplification is usually not the case.

12.1.3 A General Multi-User Capacity Region

The general capacity-region specification for a multi-user channel can be stated in terms of a search over all $(U!)^U$ possible orders of users. Such a search may require large computation.

The general specification proceeds in 3 separate optimization/searches as enumerated in this paragraph:

For a general multi-user channel with fixed input distribution $p_{\boldsymbol{x}}$ and channel $p_{\boldsymbol{y}/\boldsymbol{x}}$, there is an order vector let $\boldsymbol{\pi} = [\pi_1, \pi_2, \cdots, \pi_U]$, where $\pi_u, u = 1, \cdots, U$, is the decoding order for the receiver of user u. Thus there are $(U!)^U$ decoding orders for a U-user channel. The capacity region of the general multiple-user channel is described U already corresponds to the situation in Definition 12.1.1, and so U may be more than the "actual set of users" on the channel, see Subsection 12.1.5 to reconstruct the original users). Thus, U may vary for each choice of input distribution, but an exhaustive search over all such distributions still must occur (no matter how U may vary as that search proceeds). For any given order vector $\boldsymbol{\pi}$ and given input probability distribution $p_{\boldsymbol{x}}$, receiver u uses an optimal decoder for its component order π_u to achieve a vector of user bits/symbol denoted $\boldsymbol{b}_u(\boldsymbol{\pi}, p_{\boldsymbol{x}})$. The particular order will determine the receiver's data rates for each of the users with all later users in the order at receiver u viewed as noise and all earlier users decoded first. The notation

$$\boldsymbol{b}_i(u, \boldsymbol{\pi}, p_{\boldsymbol{x}}) \triangleq \mathbf{I}(\mathbf{x}_i; \mathbf{y}_i | \mathbf{x}_{\pi_u(1)}, \mathbf{x}_{\pi_u(2)}, \cdots, \boldsymbol{x}_{\pi_u(i-1)}) \tag{12.26}$$

denotes the u^{th} component of $\boldsymbol{b}_u(\boldsymbol{\pi}, p_{\boldsymbol{x}})$ (user u data rate) at receiver i.

The data rates achievable for this order and input distribution may not be achievable at other receivers $i \neq u$. Equivalently, user u's data rate at receiver u may not be achievable at some or all of the other receivers. Indeed, there will be a minimum data rate for user u across all receivers for the given order vector $\boldsymbol{\pi}$ and the given input distribution $p_{\boldsymbol{x}}$:

$$b_u(\boldsymbol{\pi}, p_{\boldsymbol{x}}) = \min_i \boldsymbol{b}_i(u, \boldsymbol{\pi}, p_{\boldsymbol{x}}) \quad . \tag{12.27}$$

This minimum can be achieved at all receivers for the given order and input distribution. A vector of these minimum rates can be constructed as

$$\boldsymbol{b}(\boldsymbol{\pi}, p_{\boldsymbol{x}}) = \bigotimes_{u=1}^{U} b_u(\boldsymbol{\pi}, p_{\boldsymbol{x}}) \quad , \tag{12.28}$$

where $\bigotimes$ corresponds to Cartesian product (or simply form an ordered U-tuple). Each element data rate of $\boldsymbol{b}(\boldsymbol{\pi}, p_{\boldsymbol{x}})$ can be achieved at all receivers, and any user data rate that exceeds its corresponding entry in $\boldsymbol{b}(\boldsymbol{\pi}, p_{\boldsymbol{x}})$ cannot be achieved for this order $\boldsymbol{\pi}$ and this input $p_{\boldsymbol{x}}$ at one or more receivers (and thus a higher data rate would be a single-user capacity violation at one or more receivers for decoding). There are many $((U!)^U$ orders, and so there are many points $\boldsymbol{b}(\boldsymbol{\pi}, p_{\boldsymbol{x}})$. Any time-sharing of the designs corresponding to these points is allowed; equivalently the convex hull of the region formed by the set of points over all orders for any given $p_{\boldsymbol{x}}$ is achievable (some interpretation according to Subsection 12.1.5 may be necessary here):

$$A(\boldsymbol{b}, p_{\boldsymbol{x}}) = \bigcup_{\boldsymbol{\pi}}^{conv} \boldsymbol{b}(\boldsymbol{\pi}, p_{\boldsymbol{x}}) \quad . \tag{12.29}$$

Any point outside this convex hull has at least one user's data rate, that for the given $p_{\boldsymbol{x}}$, corresponds to at least one receiver that cannot decode that one user, no matter what order is used. Such a point then violates a single-user capacity limit for all orders and the given input $p_{\boldsymbol{x}}$. Finally then,

$$c_{general}(\boldsymbol{b}) = \bigcup_{p_{\boldsymbol{x}}}^{conv} A(\boldsymbol{b}, p_{\boldsymbol{x}}) \tag{12.30}$$

where the convex hull over all possible input spectra allows input distributions for each independent user that each must satisfy the particular user's input constraints.

The unions in Equations (12.29) and (12.30) may cause the value of U to change for each possible element in the set over which the union is taken accordingly the general definition of a user. However, each such $b(\boldsymbol{\pi}, p_{\boldsymbol{x}}$ or $A(\boldsymbol{b}, p_{\boldsymbol{x}}$ can be reconstructed in terms of the original set of users as described in Section 12.1.5.

12.1.4 Enlarged Channels (Relays)

An enlarged multi-user channel is one where the number of users U according to Defintion 12.1.1 exceeds the number of transmit (MAC, IC, or general multi-user) or receive (BC, IC, or general multi-user) locations in the channel description. The results and capacity regions of this chapter then apply to the enlarged number of users. This region could then be called the enlarged capacity region[7]. The next Subsection 12.1.5 discusses computation of the desired capacity region from the enlarged-channel capacity region. Enlarged multi-user channels allow the description of the capacity region for the enlarged number of users and allow the concept of order and successive decoding (or precoding) to be used heavily in the general capacity region description of Subsection 12.1.3.

An example of an enlarged channel is what is sometimes called a "relay channel." In a relay channel, an intermediate user may decode and then forward a message from one user to a third user. For instance, an IC with nominally U users might instead be viewed as a system with $U' = U^2$ users where U is the number of transmitter/receiver locations, but in fact the capacity description for some orders might indeed be described in terms of probability distributions that tacitly imply a larger number of users. Indeed the capacity description could imply up to U^2 users, and presuming any and all messages from one transmit site to another receiver site could not be further decomposed. (For the linear Gaussian IC channel, the maximum number of users need not exceed $U' = U^2$.) When an IC is viewed as a relay channel, the number of users actually becomes (at least) $U' = U^2 + U$ if receiver sites identify with actual users desiring messages. The U^2 component refers to the number of possibly intermediate users' messages. Nonetheless, for this enlarged number of users, the capacity region can be calculated according to Subsection 12.1.3. Relay channels may generalize to have m intermediate "hops," which then means

[7] "Enlarged" means that U is enlarged so that there are more dimensions.

that the number of users would increase to $U' = (U^2)^m + U$. Many of the actual users may have zero transmission paths and thus can be eliminated as not of interest. The capacity region for the enlarged multi-user or relay channel would then need to add an additional calculation to the enlarged-channel's capacity region that would be

$$b_{u \to u'} = \max_{(i,j)^m \in \left(\boldsymbol{U} \otimes \boldsymbol{U}\right)^m} \left\{ \min \left(R_{u \to i}, \ldots.., R_{j \to u'} \right) \right\} \quad . \tag{12.31}$$

Essentially, then the enlarged channel is searched over all possible enlarged rate tuples for the maximum possible transfers over all possible multi-hop paths. While the concept is fairly straightforward, the actual search could be enormously complex.

Reduction to the capacity region for the original actual users then requires the summing of user components as Subsection 12.1.5.

12.1.5 Macro User Groups

Macro user groups are essentially groups of users that correspond to all the decodable components of one user for enlarged multi-user channels. The capacity region for the enlarged channel is consructed. From it, sums of allowed user components in rate tuples can be computed for all rate tuples in a macro-user group, thus "compressing" (reducing the number of dimensions) the length of the rate tuple. Such sums correspond to actual users that happened to decompose into more than one user earlier. Certainly by finding the largest possible enlarged multi-user channel rate region, the compression to sums creates all possible points for the actual users rates (some of which could be repeated by several designs) but the region created would be the largest possible (since any point outside the region would correspond to a point not allowed in the enlarged capacity region).

12.2 Capacity Rate Regions for the multiple-access channel

This section describes the rate region for any multiple-access channel. The general capacity region in Equations (12.27) - (12.30) of Section 12.1.3 can be alternately specified through the use of rate sums for only the MAC for any particular probability distribution. This rate-sum formulation is more common than the general formulation in (12.27) - (12.30) for MACs. This rate-sum formulation is often simpler since there is only one receiver, and the operation in (12.28) is not necessary. Section 12.2.1 handles the general case from the perspective of rate sums and relates the original region specification to the rate-sum version, showing they provide the same region from different perspectives. Section 12.2.2 shows how simple extension of Chapter 4's water-filling can be used to compute the maximum rate sum and corresponding distribution for Gaussian MACs.

12.2.1 The General Multiple-access Rate Region

A MAC analysis often focuses upon the sum of users' data rates: The MAC receiver simply treats all inputs as the collection of individual users' inputs into a single input vector $\boldsymbol{x} = [\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ ...,\ \boldsymbol{x}_U]$, and the mutual information is then the single-user mutual information for the channel from $\boldsymbol{x}$ to $\boldsymbol{y}$. Such a rate-sum treatment shows no preference for the distribution of data rates among the users on the multiple-access channel. As in in (12.34) and (12.7), an individual b_u may exceed $I(\boldsymbol{x}_u;\boldsymbol{y})$ if the other users are known, an important observation for constructing rate regions. A best case would be that all the other users' signals were somehow known to the common receiver of the multiple-access channel output. Then

$$b_u \le I(\boldsymbol{x}_u;\boldsymbol{y}/\boldsymbol{x}_{i\neq u,\ i=1,...,U}) \quad . \tag{12.32}$$

If one or more of the users is inactive, then subsets (indexed here by a vector of indices $\boldsymbol{u}$, which is a subset of the entire set of user indices $\boldsymbol{U} = \{1, 2, ..., U\}$) have mutual information to the channel output

$$I(\boldsymbol{x}_{\boldsymbol{u}};\boldsymbol{y})\ \forall\ \boldsymbol{u} \subseteq \boldsymbol{U} \quad , \tag{12.33}$$

and represent partial maximum rate sums over a subset of user indices that describe the user-index vector $\boldsymbol{u}$ if the other users, denoted in set notation as $\boldsymbol{U} \setminus \boldsymbol{u}$, are unknown.

These partial rate sums are maximums over the average statistics of the excluded users, so the effect of the others is averaged as if they were a "noise" to the decision. A higher rate sum can be achieved for any group of users specified by $\boldsymbol{u}$ by conditioning the mutual information on all the other users, $\{\boldsymbol{U}\setminus\boldsymbol{u}\}$, which essentially means all the other users are known and eliminated or canceled in terms of their effects on the detection of the first set of users. Being given the set $\{\boldsymbol{U}\setminus\boldsymbol{u}\}$ can only help optimal detectors, thus increasing the bound. Since both are mutual information quantities, they represent maximum (sum) data rates for the group of users in $\boldsymbol{u}$ in the respective conditions (from single-user information theory in Chapter 8 where the single user is the set $\boldsymbol{u}$). Then the following bounds hold for all possible $\boldsymbol{u}$ and any given input probability distribution and channel:

$$\sum_{u\in\boldsymbol{u}\subseteq\boldsymbol{U}} b_u \le I(\boldsymbol{x}_{\boldsymbol{u}};\boldsymbol{y}/\{\boldsymbol{U}\setminus\boldsymbol{u}\}) \quad . \tag{12.34}$$

Equations (12.32), (12.33) and all the instances of (12.34) over user subsets will specify an achievable multi-user rate region for the multiple-access channel for any given input distribution $p_{\boldsymbol{x}}$.

The proof that this achievable region is the bound on $\boldsymbol{b}$ for any given probability distribution can use single-user mutual information bounds and the so-called **chain rule** of information theory:

Theorem 12.2.1 (Chain Rule) *The* **chain rule of mutual information** *is*

$$I(\boldsymbol{x};\boldsymbol{y}) = I(\boldsymbol{x}_1;\boldsymbol{y}) + I(\boldsymbol{x}_2;\boldsymbol{y}/\boldsymbol{x}_1) + ... + I(\boldsymbol{x}_U;\boldsymbol{y}/[\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ ...,\boldsymbol{x}_{U-1}]) \tag{12.35}$$

$$= \sum_{u=1}^{U} I(\boldsymbol{x}_u;\boldsymbol{y}/[\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ ...,\boldsymbol{x}_{u-1}]) \tag{12.36}$$

The proof is trivial and follows from factorization of $p_{\boldsymbol{x},\boldsymbol{y}}$ into a chain of conditional probability distributions.

Essentially, the chain rule could apply for any order of users, and the rate sum for all orders is the same. The chain rule could also be applied within any subset of users and their partial rate sums would be the same for the situation that all others are given. So, instead of specifying orders as in Section 12.1.3, an equivalent enumeration of situations to consider is all the possible rate sums for different subsets of user indices.

The chain rule provides a **successive decoding** strategy for achieving any of the points corresponding to a rate sum. The successive-decoding interpretation for the chain rule is that the code for user 1 can be designed as if all other users were "noise" or according to the mutual information averaged over all the other users' distributions. When the single receiver correctly decodes user 1 (with $P_{e,1} \to 0$), user 1 can be accepted as given for the decoding of user 2, which implies a certain "non causality" in decoding: technically $\boldsymbol{x}_1$ should be known over all time to be decoded with zero error. User 2's code is designed for the case where users 3,...,U were "noise" or according to the mutual information averaged over all the other users' distributions, but as if user 1 were known. The successive decoding procedure can thus achieve the maximum sum of data rates $I(\boldsymbol{x};\boldsymbol{y})$. There are $U!$ orderings, and thus $U!$ data rate U-tuples that achieve the same maximum sum. Any set of data rates satisfying $\sum_{u=1}^{U} b_u = I(\boldsymbol{x};\boldsymbol{y})$ is achievable simply by designing a set of codes that time-share the appropriate $U!$ points. However, the rate sum need be computed only once for any of the orders, thus simplifying the need to compute the same rate sum repeatedly if it is the only quantity of interest (a deficiency of the more general capacity-region generating strategy in Section 12.1.3). Successive decoding is essentially a non-causal generalization the GDFE procedure of Chapter 5 when $N = U$ on an AWGN and when the multi-user Gaussian input distribution is such that it has diagonal autocorrelation matrix.

The chain rule can be applied to all the subsets of users denoted by $\boldsymbol{u}$. Any order within each partial rate sum can be used. There are thus $\sum_{u=1}^{U}\binom{U}{u} = 2^U - 1 < U!$ (when $U > 3$) partial rate sums to compute, which may be a substantial reduction in calculations depending on exactly the designer's objectives. The procedure in 12.1.3) also requires forming convex combinations, which may be tedious. However, as this section later shows, essentially the two procedures are equivalent. Once given all the partial rate sums for any probability distribution, then the capacity rate region follows:

Theorem 12.2.2 (Multiple-Access Channel Rate Region) *The general multiple-access rate region $A(\boldsymbol{b})$ for a given user input distribution $p_{\boldsymbol{x}}$ is the set of all U-dimensional rate vectors $\boldsymbol{b}$ that satisfy:*

$$A(\boldsymbol{b}) = \left\{ \boldsymbol{b} \mid \boldsymbol{b} \in \bigcup_{\boldsymbol{u} \subseteq \boldsymbol{U}}^{conv} 0 \le \sum_{\boldsymbol{u}} b_u \le I(\boldsymbol{x}_{\boldsymbol{u}}; \boldsymbol{y}/\boldsymbol{x}_{\tilde{\boldsymbol{u}} \in \boldsymbol{U} \setminus \boldsymbol{u}}) \right\} \quad . \tag{12.37}$$

$\boldsymbol{x}_{\tilde{\boldsymbol{u}}}$ is the given set of all input symbols not in $\boldsymbol{u}$.

Proof: *For achievement of any point in this region, see the discussion leading to this theorem on successive decoding with single-user capacity-achieving codes independently used for each user. Any point outside the region violates at least one of the rate sums and would be a single-user capacity-theorem violation for at least one user corresponding to that rate sum, with other users in $\boldsymbol{U} \setminus \boldsymbol{u}$ given.* **QED**.

The intersection in 12.37 essentially runs through all orders, but recognizing **for the MAC** that the "sub order" used in computing a rate sum does not change the rate sum, nor does the "sub order" within the users excluded for the computation of the rate sum.

Lemma 12.2.1 (Capacity Region for the Multiple Access Channel) *The capacity region $C(\boldsymbol{b})$ is the maximum, or technically the convex hull of the union of the $A(\boldsymbol{b})$, over all the possible joint input probability distributions allowed on the particular channel.*

$$C(\boldsymbol{b}) = \left\{ \boldsymbol{b} \mid \boldsymbol{b} \in \bigcup_{p_{\boldsymbol{x}}}^{conv} A\left(\boldsymbol{b}(p_{\boldsymbol{x}})\right) \right\} \tag{12.38}$$

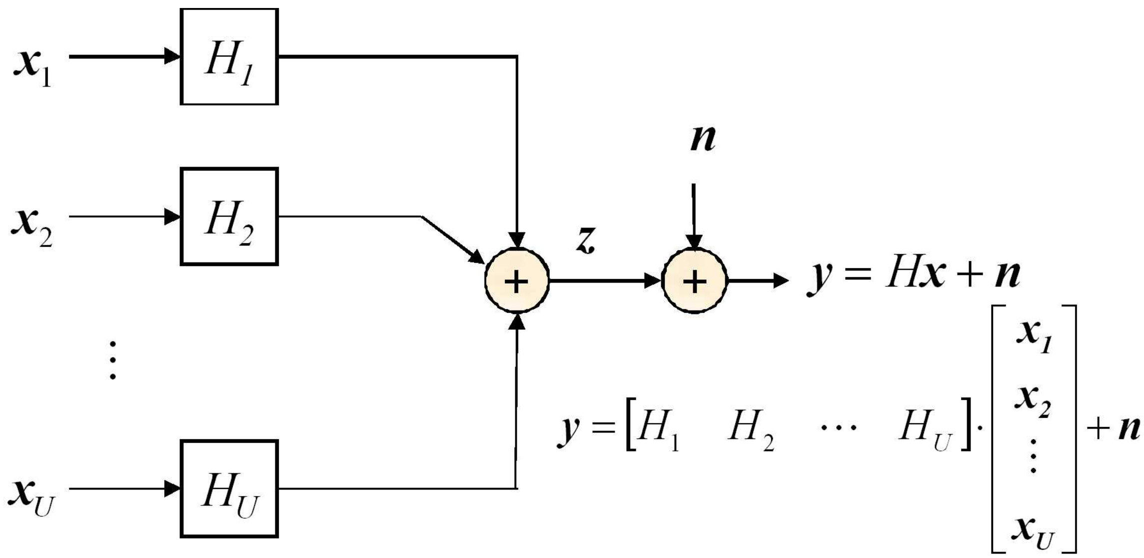

Figure 12.6: Rate Region for two-user multiple-access channel.

where the union is over all allowed probability distributions.

Proof: *Any possible input distribution leads to a valid rate region according to Theorem 12.2.2. The union of such regions is thus also achievable for at least one of the distributions. The convex hull of the union corresponds to points that may time-share (or dimension-share) some of the different possible probability distributions – assuming such a time-sharing is an allowed input probability distribution. (See the qualifications in Subsection 12.1.5.)*

By contrast, any point outside the region $C(F)$ violates at least one single-user capacity theorem and so cannot be achieved with arbitrarily small probability of error. **QED**.

An example of the general rate region for 2 users and a particular $p_{\boldsymbol{x}}$ appears in Figure 12.6 using both the method of subsection 12.1.3's Equations (12.29) - (12.30) and the rate-sum method. This 2-user region is a pentagon with the boundary lines shown. Each of the 3 rate sums has been computed to bound the region, tracing the pentagon for any particular input distribution $p_{\boldsymbol{x}}$. 3 calculations were necessary to compute the rate sums (the points of intersection admittedly require some additional linear-equation solving that is somewhat trivial in the case of $U = 2$). The approach of Section 12.1.3 is also fairly easy to compute for any given $p_{\boldsymbol{x}}$. Points in the left rectangle are those for which user 1 is decoded first and then used to assist the decoding of user 2. The lower rectangle corresponds to decoding first user 2 and then user 1. The lower left corner corresponding to the overlap of the two rectangles admits either order of decoding. The upper right triangular "time-share" region more precisely corresponds to dimension-sharing (convex hull of union) the two codes and decoding orders. Each particular $p_{\boldsymbol{x}}$ again corresponds to a pentagon after the dimension-sharing.

Extending the example to $U = 3$ users, the rate-sum approach must compute 7 rate sums and the associated intersecting lines and points, which are defined by the 6 points corresponding to boxes and each of the 6 orders from the procedure in Section 12.1.3, along with all possible convex combinations. The corners of the boxes are the intersection points of the planes defined by the rate sums. The region is 10 sided in 3 dimensions for any given $p_{\boldsymbol{x}}$. For $U = 4$, there are 15 rate sums plus intersections, or equivalently 24 4D boxes that correspond to 24 orders and the resultant convex combinations. In general, there are $2^U - 1$ partial rate sums to compute or $U!$ boxes corresponding to all the possible orders. The result is a region with $2^U + U - 1$ faces for any given $p_{\boldsymbol{x}}$.

The union of regions for multiple $p_{\boldsymbol{x}}$ possibilities need not be a pentagon with 2 users, nor any $2^U + U - 1$-sided face in general, but will always be convex.

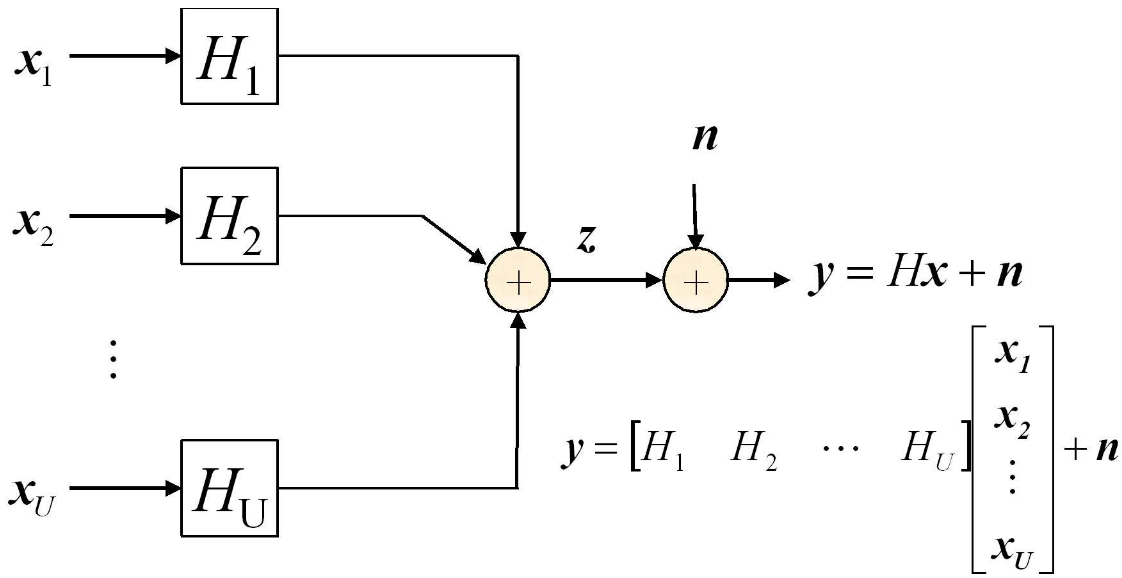

Figure 12.7: Gaussian multiple-access channel.

12.2.2 The Gaussian multiple-access channel.

The Gaussian MAC rate region exactly follows the general region, but can be more directly computed using familiar signal-to-noise ratio quantities. All inputs may be Gaussian for the Gaussian MAC and in particular must be Gaussian at the boundary of the region. The sufficiency and necessity of Gaussian inputs is proved easily by reversing the chain rule and looking at the last user decoded in successive decoding. This last user sees only Gaussian noise, so by single-user capacity has highest rate when Gaussian. Then, since this last user is Gaussian, the noise plus crosstalk seen by the 2nd to last is also Gaussian, that user and all earlier users must also be Gaussian by induction. Figure 12.7 illustrates the Gaussian multiple-access channel. Each MAC user has a gain H_u to the channel output where a common Gaussian noise $\boldsymbol{n}$ is added to construct $\boldsymbol{y}$. Since for each and every rate sum, the users contributing to the sum can be considered as a single "group user," the joint distribution of that single group user must be Gaussian when the noise is Gaussian. Then, each of the marginals or individual users are then also Gaussian. The mutual information quantities of the general MAC capacity region rate-sum descriptions become log-signal-to-noise ratio quantities where the signal power for any user is the product of the input energy $\bar{\mathcal{E}}_u$ and the channel gains $|H_u|^2$, and the noise is the sum of the Gaussian noise and any uncanceled other users. Mathematically,

$$\bar{I}(\boldsymbol{x_u}; \boldsymbol{y}/\boldsymbol{x}_{\tilde{\boldsymbol{u}} \in \boldsymbol{U} \setminus \boldsymbol{u}}) = \frac{1}{2} \log_2 \left(1 + \frac{\bar{\mathcal{E}}_u \cdot |H_u|^2}{\sigma^2 + \sum_{i \in \boldsymbol{U} \setminus \boldsymbol{u}} \bar{\mathcal{E}}_i \cdot |H_i|^2} \right) \quad . \tag{12.39}$$

For example if $U = 2$ and $H = [h_1 \; h_2]$, then

$$y = h_1 \cdot x_1 + h_2 \cdot x_2 + n \quad . \tag{12.40}$$

Each rate sum is associated with an SNR:

$$\text{SNR}_1 = \frac{\mathcal{E}_1 \cdot |h_1|^2}{\sigma^2} \tag{12.41}$$

$$\text{SNR}_2 = \frac{\mathcal{E}_2 \cdot |h_2|^2}{\sigma^2} \tag{12.42}$$

$$\text{SNR} = \frac{\mathcal{E}_1 \cdot |h_1|^2 + \mathcal{E}_2 \cdot |h_2|^2}{\sigma^2} \quad , \tag{12.43}$$

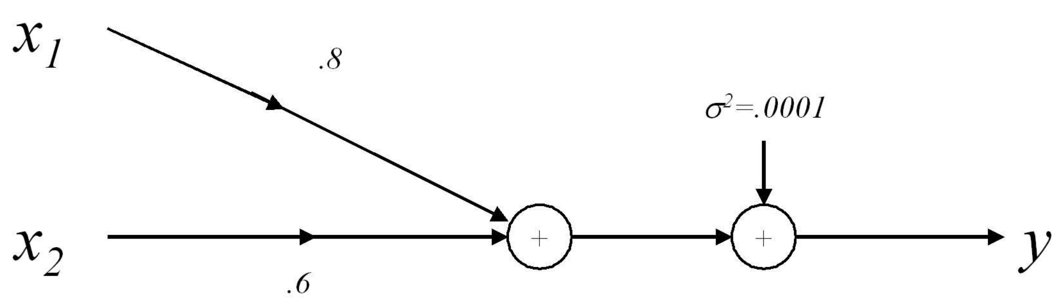

Figure 12.8: Simple example of capacity calculation for 2-user AWGN.

and

$$b_1 \leq \frac{1}{2}\log_2\left(1+\text{SNR}_1\right) \quad (12.44)$$

$$b_2 \leq \frac{1}{2}\log_2\left(1+\text{SNR}_2\right) \quad (12.45)$$

$$b_1+b_2 \leq \frac{1}{2}\log_2\left(1+\text{SNR}\right) \quad . \quad (12.46)$$

The SNR's and thus rate sums for all possible Gaussian input distributions would all be less than those shown because trivially each of the SNR's either stays the same or is reduced for any Gaussian distributions other than those using the maximum energy. Thus, the 3 equations above also specify the capacity region.

EXAMPLE 12.2.1 (AWGN channel with large capacity) Figure 12.8 illustrates a simple 2-user AWGN channel. The overall capacity for this system is easy to compute because the relationship between the channel input and $\boldsymbol{z}$ is one-to-one for discrete on users 1 and 2. Thus, the overall capacity of the channel is $I(\boldsymbol{x};\boldsymbol{y}) = I(\boldsymbol{z};\boldsymbol{y}) = .5\log_2(1+10^4)$ =6.64 bits/dimension. However, consideration of either user's signal as Gaussian noise into the other leads to capacities of

$$I(x_2;y) = \frac{1}{2}\log_2\left(1+\frac{.36\cdot 1}{.0001+.64}\right) = .32 \text{ bits/dimension} \quad (12.47)$$

$$I(x_1;y) = \frac{1}{2}\log_2\left(1+\frac{.64\cdot 1}{.0001+.36}\right) = .74 \text{ bits/dimension} \quad (12.48)$$

$$\text{total} = 1.06 \text{ bits/dimension} \quad , \quad (12.49)$$

which is well below the maximum rate sum.

Decoding x_2 first (point B) leads to

$$I(x_1;y/x_2) = \frac{1}{2}\log_2\left(1+\text{SNR}_1\right) = \frac{1}{2}\log_2\left(1+\frac{.64\cdot 1}{.0001}\right) = 6.32 \text{ bits/dimension} \ , \quad (12.50)$$

or decoding x_1 first (point A) leads to

$$I(x_2;y/x_1) = \frac{1}{2}\log_2\left(1+\text{SNR}_2\right) = \frac{1}{2}\log_2\left(1+\frac{.36\cdot 1}{.0001}\right) = 5.90 \text{ bits/dimension} \quad (12.51)$$

In either case, the sum is the same as .32+6.32=.74+5.90=6.64 bits/dimension.

Figure 12.9 illustrates the specific rate region and the two successive-decoding points. Figure

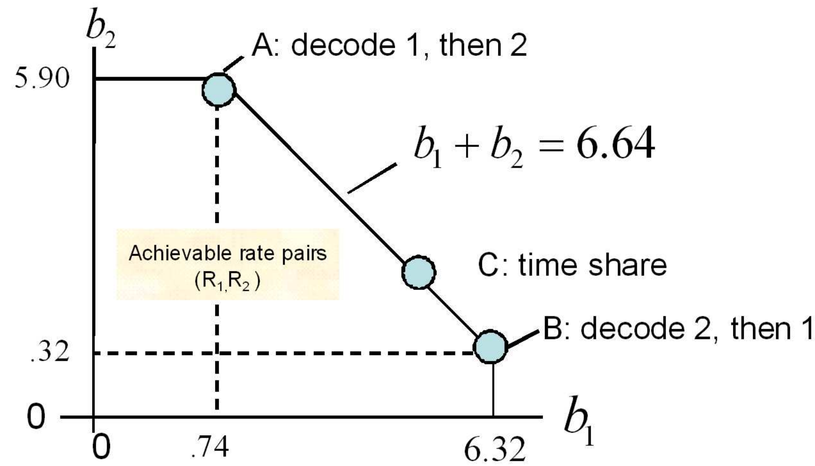

Figure 12.9: 2-user AWGN rate region.

12.10 illustrates the two successive-decoding receivers for this channel. Other points along the slanted line imply time-sharing the two receivers and codes. In the Gaussian case, time-sharing of two different distributions always corresponds to a single vector Gaussian distribution for each user. For instance, a time-sharing of 3/4 point B and 1/4 point A yields point C with

$$\mathcal{E}_{x_1} = \frac{1}{4}(1) + \frac{3}{4}(1) = 1 \quad (12.52)$$

$$\mathcal{E}_{x_2} = \frac{1}{4}(1) + \frac{3}{4}(1) = 1 \quad (12.53)$$

$$b_1 = \frac{1}{4}(.74) + \frac{3}{4}(6.32) = 4.92 \quad (12.54)$$

$$b_2 = \frac{1}{4}(5.90) + \frac{3}{4}(.32) = 1.72 \quad (12.55)$$

$$b_1 + b_2 = 6.64 \quad . \quad (12.56)$$

This means that each user decomposes into 2 sub-users. Technically, this is no longer a MAC. However, the box of (12.28) is now 4-dimensional, and still applies for this $U = 4$ multi-user channel. As in Subsection 12.1.5 the rates of the sub-users can be added to generate a 2-dimensional box corresponding to point C. The receiver is then 3/4, 1/4 time-share of receivers A and B.

Non-Zero Gaps

The rate regions so far have used a $\Gamma = 0$ dB gap. Strictly speaking, the chain rule does not exactly hold when the gap is nonzero. For instance, in the above example with 3 dB gap, $I(x_2; y) \rightarrow 0.1788$ and $I(x_1; y) \rightarrow 0.4587$ while $I(x_1; y/x_2) \rightarrow 5.8222$ and $I(x_2; y/x_1) \rightarrow 5.403$. The two rate sums are no longer equal, $.01788 + 5.8222 = 5.8301 \leq 0.4587 + 5.403 = 5.866$, and there is a higher rate sum when decoding user 1 first. Furthermore, $5.866 < 6.144 = .5 \log_2(1 + 10,000/\Gamma)$, where the latter "overall gap-reduced capacity" really has no direct bearing for non-zero gaps. Any point on the line between [.01788 , 5.8222] and [.4587 , 5.403] could be achieved by time/dimension sharing, but the line does not have slope -1, and is not even necessarily the boundary to the "gap-reduced" rate region. Generally, some caution should be exercised with rate regions with non-zero gaps. Jagannathan investigates the construction of non-zero-gap exact achievable rate regions (p. 164-173 of his 2008 Stanford dissertation). His findings show that the rate-sum-bounding-plane method no longer applies. As the gap grows, so does the deviation. His approach basically follows the more general approach of Equations (12.27) to (12.30)

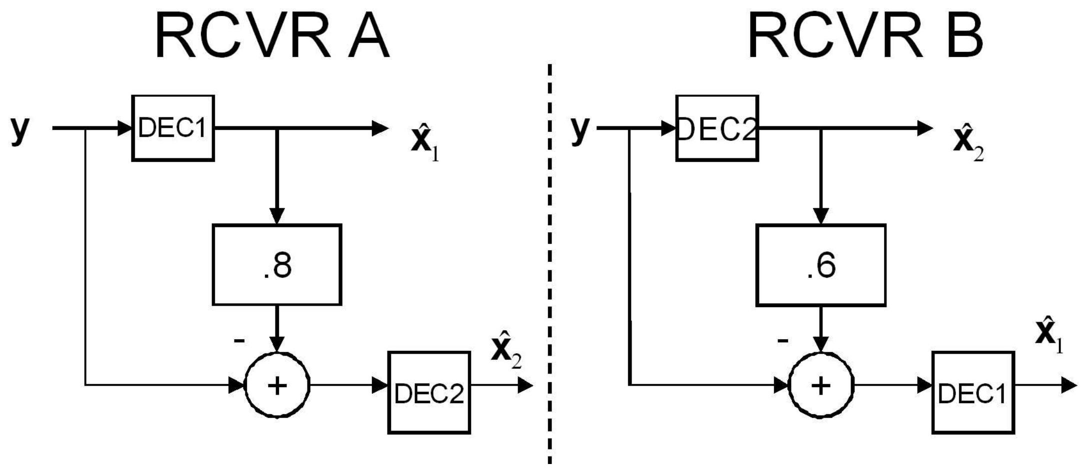

Figure 12.10: Example GDFEs.

in section 12.1.3 where each mutual information is replaced by the corresponding $\frac{1}{2} \cdot log(1+\text{SNR}/\Gamma)$ style term and the convex-hull-union process becomes more involved but still produces an achievable rate region. Even this approach assumes that other users can be decoded without error (which may not be true for non-zero code gaps) and that furthermore the users are all Gaussian codes (which of course they may not be with non-zero gaps). Thus, with sophisticated systems of the type that these later chapters analyze, the price of using a powerful code (with close to zero gap) may be only a small fraction of the multi-user design costs. This text thus returns to the use of zero gaps in these last 4 chapters unless otherwise specifically noted.

In general for $U > 2$ AWGN channels summed together to form a multiple-access channel, the pentagon generalizes to an U-dimensional region with $2^U + U - 1$ sides in U-space for U users. Each axis has a maximum data-rate plane orthogonal to that axis and forms a boundary for the U-dimensional capacity region at $C_u = .5 \log_2(1 + \text{SNR}_u)$ where $\text{SNR}_u = \mathcal{E}_u |H_u|^2/\sigma^2$. There are planes for all possible subsets $\boldsymbol{u}$, each as $\sum_{u \in \boldsymbol{u}} R_u \le C_{\boldsymbol{u}} = 1/2 \log_2(1 + \text{SNR}_{\boldsymbol{u}})$ where $\text{SNR}_{\boldsymbol{u}} = \sum_{u \in \boldsymbol{u}} \mathcal{E}_u |H_u|^2/\sigma^2$.

Non-white (that is, noise that is correlated) noise is handled for the Gaussian multiple-access channel by the same noise-whitening transformation as is used in the single-user case.

GDFEs and relation to Successive Decoding

The GDFE of Chapter 5 could be used with a non-singular $U \times U$ matrix channel H where the specific input autocorrelation matrix $R_{\boldsymbol{xx}} = R_{\boldsymbol{vv}}$ is diagonal. Instead of viewing the input as U dimensions, that same input could be viewed as U independent users, each with its own input dimension, but possibly contributing to as many as all U of the MAC output dimensions. The contributions of all the inputs are summed in the MAC to the common output $\boldsymbol{y}$, upon which the GDFE acts to estimate each of those independent input dimensions (or users). Such a system achieves the cannonical sum-rate equal to the appropriate conditional mutual information for each user/dimension in each of the $U!$ possible GDFE user/dimension orderings. Each successive user/dimension has the benefit of removal of the crosstalk from previous users/dimensions in the triangular back-subsitution GDFE solution of Chapter 5. However, the feedback sections are not quite the same. The GDFE would use a MMSE feedback section formed by a matrix $G \neq H$ in general (for certain trivial H like diagonals, the two are the same after bias removal). Successive decoding directly uses H: for instance, user 2 would use the first column of H, $\boldsymbol{h}_1$, to multiply by x_1 and then subtract from $\boldsymbol{y}$. The difference of these two identically performing systems that both directly correspond to a MMSE solution is subtle: this difference depends upon the type of detector: The GDFE uses a dimension-by-dimension (user-by-user) simple slicing decision on each and every dimension/user of the feedforward section output. The successive decoder would use an N-dimensional decoder for $\boldsymbol{y}$, treating other users as noise on each of those N dimensions. The ML detection for each user is more complex and involves search of closest possible noise-free (this includes

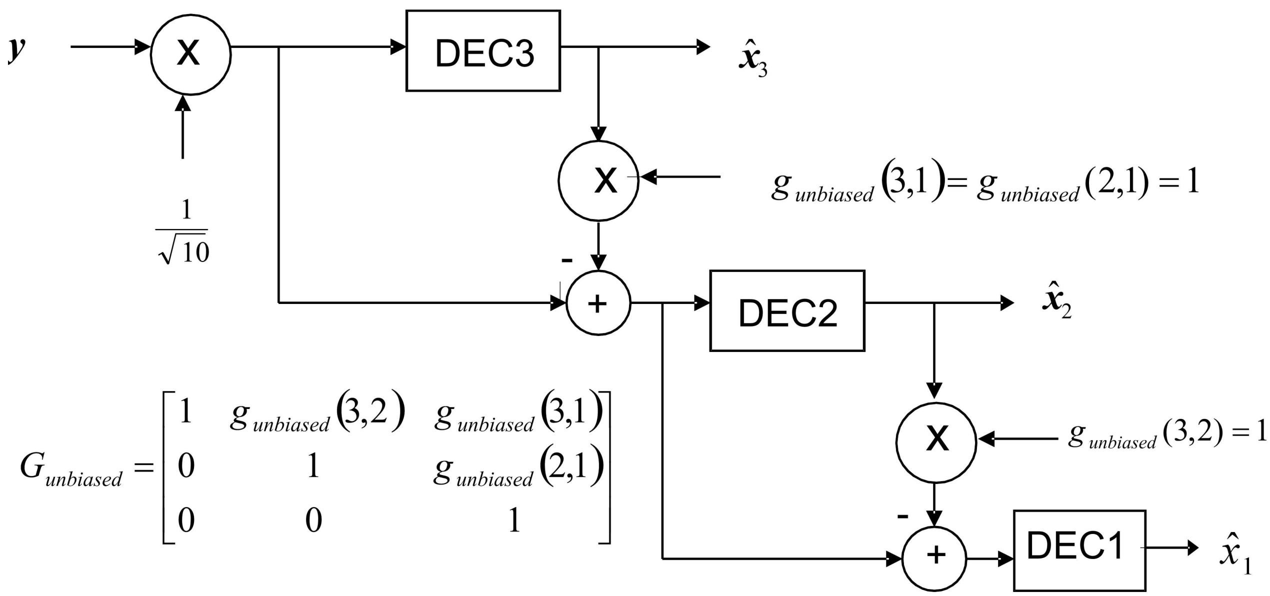

Figure 12.11: GDFE structure for multiple-access receiver.

no other-user noise) channel outputs to the actual $\boldsymbol{y}$ value. Instead, the GDFE with linear processing, can simplify this detection to one dimension (where other users are also noise filtered by feedforward section). This decoder-complexity difference is the reason for the different feedback sections. Otherwise the two perform the same when all users are Gaussian codes with zero-dB gap, and when there is no error propagation. With these powerful (zero-gap) codes, the extra ML or "successive decoding" complexity is higher than the GDFE, but they perform the same. Essentially the GDFE is "focusing" the single dimension of each user that was spread by the channel into N dimensions back to a single dimension. Successive decoding does not so focus on to a single dimension, but performs the same – just more complicated. Each GDFE user decision is largely free (in MMSE sense) of other user noise when $N \geq U$ so that noise from other users is very small tending to zero as true noise goes to zero, when $N \geq U$.

The relationship between DFE's and successive decoding extends in certain interesting ways even when $\rho(H) < N$, and in this case the other-user/dimension noise will be significantly larger. When there is only one output dimension $N = 1$, the GDFE's usual elimination of channel singularity is then not feasible in the MAC with $U > N$ (that is the MAC does not allow coordination of channel-input dimensions so that elimination of channel signularity is not possible physically). Thus many users may share a dimension and the one-to-one association between dimensions and users of the preceding paragraph when $N \geq U$ is lost.

This multi-user approach provides an extension of GDFE's to the situation of a channel rank $\rho(H) < U$. GDFE theory can proceed with this singular channel by forming by forming

$$\boldsymbol{z} = H^* \boldsymbol{y} = (H^* H)\, \boldsymbol{x} + \boldsymbol{n}' \tag{12.57}$$

where

$$R_f = H^* H \tag{12.58}$$

is a singular $N \times U$ matrix of rank $\rho(H)$, and $R_{\boldsymbol{n}'\boldsymbol{n}'} = R_f$. Any input components that are in the null space of the singular H are lost, but each MAC user has at least some non-zero energy in the pass space as long as no column of H is zero. Since these inputs are one dimensional in the MAC, any user's $d_{\min,u}$ must lie upon this one dimension, of which some energy passes to the channel output. (Effectively the MAC's uncoordinated-input restriction prevents placement of two constellation points for any user so that the line between them will align with a singular vector of H that is zeroed in the channel. Equation (12.57) is a forward canonical model since the autocorrelation of the filtered noise is also R_f (assuming the original noise was white, or already whitened, with unit variance on all components). There exists a (not necessarily unique) backward channel where

$$\boldsymbol{x} = R_b \boldsymbol{z} + \boldsymbol{e} \quad , \tag{12.59}$$

where the $U \times U$ R_b is the MMSE estimator of the channel input vector. R_b^{-1} can be computed and factored as

$$R_b^{-1} = R_f + R_{\boldsymbol{xx}}^{-1} = G^* \cdot S_0 \cdot G \quad , \tag{12.60}$$

as in Chapter 5's GDFE theory. Again, a decision-feedback arrangement can be used to estimate x_U first, then x_{U-1}, ..., x_1. The "ISI" or "crosstalk" from other dimensions was minimized in the MMSE approach. This inter-dimensional interference now includes other users' signals, and these signal's MMSE components may be quite large if they've not yet been cancelled. When the number of output dimensions of the channel equals or exceeds the number of input dimensions, then signal components in the minimized mean-square error are usually small (if noise is much less than the signal levels generally), as was the case in Chapter 5 where input and channel singularity were prudently removed to ensure such a condition holds. However, when the number of input dimensions exceeds the number of output dimensions, as often happens in the MAC, the MMSE can contain very significant signal components even when the noise is very small. In the single-user case of Chapter 5, such a situation was viewed as "injecting energy" into the null space of the channel and thus wasting energy that contributed nothing to improvement of performance. In the MAC, such energy loss is often inevitable because of the physical separation of the users. In this case, the loss of signal energy into the null space can also be viewed as adding a large signal component to the MMSE for the GDFE. The need for "non-causal" successive decoding is reduced to a simple "correct" previous decision assumption. From any perspective, the same GDFE approach will now do all the "book-keeping" for any order of the users, as Example 12.2.2 illustrates:

EXAMPLE 12.2.2 (A singular multi-user GDFE) A 3-user channel has $N = 1$ so that

$$y = \sqrt{10} \cdot (x_1 + x_2 + x_3) + n = \sqrt{10} \cdot \underbrace{[1\ 1\ 1]}_{H} \boldsymbol{x} + n \quad . \tag{12.61}$$

The noise variance is $\sigma^2 = 1$, and the input users are independent so $R_{\boldsymbol{xx}} = I$. Clearly this channel has a two-dimensional null space and one-dimensional pass space - any input that does not lie along the vector [1 1 1]' is lost. However, the receiver can still detect all 3 inputs, just with greater MMSE level caused by the large signal-dependent components in the MMSE. The following sequence of Matlab commands designs a MMSE GDFE for this channel:

```
>> h=sqrt(10)*[1 1 1] =  3.1623    3.1623    3.1623
>> Rf=h'*h
                   10.0000   10.0000   10.0000
                   10.0000   10.0000   10.0000
                   10.0000   10.0000   10.0000
>> Rbinv=Rf+eye(3)
                   11.0000   10.0000   10.0000
                   10.0000   11.0000   10.0000
                   10.0000   10.0000   11.0000
>> Gbar=chol(Rbinv)
                    3.3166    3.0151    3.0151
                         0    1.3817    0.6580
                         0         0    1.2150
>> G=inv(diag(diag(Gbar)))*Gbar
                   1.0000    0.9091    0.9091
                   0         1.0000    0.4762
                   0         0         1.0000
```

This monic upper-triangular G matrix illustrates that user 1 is actually detected first, then user 2, and user 3. Another order can be attempted by simply re-indexing inputs. Since this channel is symmetric in all inputs, then all orders are essentially the same. However, the first user detected sees much higher MMSE as the next few equations show:

```
>> S0=diag(diag(Gbar))*diag(diag(Gbar))
                   11.0000         0         0
                         0    1.9091         0
                         0         0    1.4762
>> SNR=det(S0) = 31.0000
```

While the overall SNR may be reasonably high (13 dB), energy has been lost into the null space of the channel. The 3 units of input energy might have been better used if not for the physically separated nature of the 3 users associated with the 3 input dimensions. Each dimension has a biased SNR on the diagonal of S_0, and so for instance user 1 sees an unbiased SNR of .4762, which is also the ratio of 10/(10+10+1) – that is user 1's 10 units of energy divided by the 20 units of combined energy in users 2 and 3 plus the 1 unit of noise energy. Similarly user 2 sees an SNR of .9091 = 10/(10+1), while user 3 sees unbiased SNR = 10/1 = 10. The GDFE automatically does all the mathematical "chain-rule" bookkeeping. The GDFE also produces a forward path calculated as:

```
>> W=inv(S0)*inv(G')
                    0.0909         0         0
                   -0.4762    0.5238         0
                   -0.3226   -0.3226    0.6774
>> unbiasedSNR=S0-eye(3)
                   10.0000         0         0
                         0    0.9091         0
                         0         0    0.4762
>> unbiasedW=S0*inv(unbiasedSNR)*W
                    0.1000         0         0
                   -1.0000    1.1000         0
                   -1.0000   -1.0000    2.1000
>> unbiasedG=S0*inv(unbiasedSNR)*(G-eye(3))+eye(3)
                    1.0000    1.0000    1.0000
                         0    1.0000    1.0000
                         0         0    1.0000
>> W*Rf
                    0.9091    0.9091    0.9091
                    0.4762    0.4762    0.4762
                    0.3226    0.3226    0.3226

>> Wunbiased=S0*inv(unbiasedSNR)*inv(S0)*inv(G')*h'

    0.3162
    0.3162
    0.3162

>> Wunbiased*h
                   1.0000    1.0000    1.0000
                   1.0000    1.0000    1.0000
                   1.0000    1.0000    1.0000
```

This basically shows that the output of the feedforward processing in the GDFE is essentially a reproduction of the same sum-channel 3 times. Thus, the GDFE does nothing in this case but suggest a successive-decoding scheme. However, the GDFE machinary is very powerful as will become evident. Furthermore, the non-causal decoding of the entire sequence of another user reduces to the "correct" instantaneous decisions assumption, well known in DFEs generally and the GDFE specifically.

The corresponding unbiased GDFE is shown in Figure 12.11. The result of the overall channel plus feedforward section is such that the channel does not appear triangularized as in Chapter 5. This is because remaning other-user noise is still present in large measure for user 1 (first to be decoded) and user 2 (second to be decoded). Otherwise the GDFE works and the overall (biased) SNR of 31 is exactly correct (or 30) and corresponds to the rate sum of the users. Each of the 3 users' data rates can be computed as

```
>> (0.5/log(2))*log(diag(S0))

    1.7297
    0.4664
    0.2809    .
```

Any of these 3 data rates could be attained by any of the users by re-ordering. Further, a set of 6 GDFE's could be used to time share according to the different orders to produce any convex combination of these rates. By zeroing any of the inputs energy, a simpler 2-input GDFE problem could also be solved to provide additional planes that bound the MAC rate region.

The example clearly illustrates the utility of GDFE theory. In practice, this example is trivial because the channel has no memory and so a ZF solution corresponds exactly to a MMSE-GDFE for $N = 1$. However, a designer could easily extend to N dimensions and $U \cdot N$ corresponding input contributions and via ordering compute one user's input estimates, then the next, and so forth. Indeed this works for any $\rho(H)$ and thus any Gaussian MAC. Chapter 13 will study this area more completely.

MMSE-DFE

A GDFE-style structure can also be constructed using a MMSE-DFE, and in fact the channels from users to the common output can each have intersymbol interference and crosstalk that corresponds to channel memory (non-flat channels) on all paths to the common receiver. Figure **??** illustrates this successive-decoding or GDFE-like concept. The chain rule suggests that a single first user can be decoded as if all the others were noise, and the result subsequently used to remove its effect from the received signal before proceeding to decode the second user. Thus a MMSE-DFE for this first user treats all other users as part of the noise spectrum used in the design of the MMSE-DFE, proceding as in Chapter 3. The coefficients in the feedback section of this MMSE-DFE are used in causal way to eliminate trailing intersymbol interference from the minimum-phase version of the other-users-as-noise-plus-true-noise-equivalent channel. If the other users are Gaussian signals, this approach does exactly produce a data rate for the first user that is $I(x_1/;\boldsymbol{y})$. The error-free decisions of the first MMSE-DFE then are filtered by a replica of the channel $h_1(D)$ and subtracted from the common channel output before it goes through a second MMSE-DFE, removing all effect of $x_1(D)$. Other than the possibly infinite-length delay implied by the feedforward-filtering of the second MMSE-DFE, the removal can be implemented sample-by-sample. The design then recursively repeats in causal fashion in successively removing each user's effect upon later users. If there were no memory in the channel (thus no ISI within or among any of the users), each MMSE-DFE degenerates into a simple decision device in Figure 12.12, causing a pure successive decoding.

For the $U \times 1$ degenerate MAC channel, there is no intersymbol interference and the noise is white Gaussian. MMSE-DFE's and ZF-DFE's are the same in this special case. More generally, the singular GDFE will turn into a MMSE-DFE where other uncanceled users are treated as noise (whatever their spectra) and any ISI is also handled for the channel in the usual way by the MMSE-DFE for the combined noise. The conditional expectations in the chain rule correspond to MMSE-DFE's and thus the successive

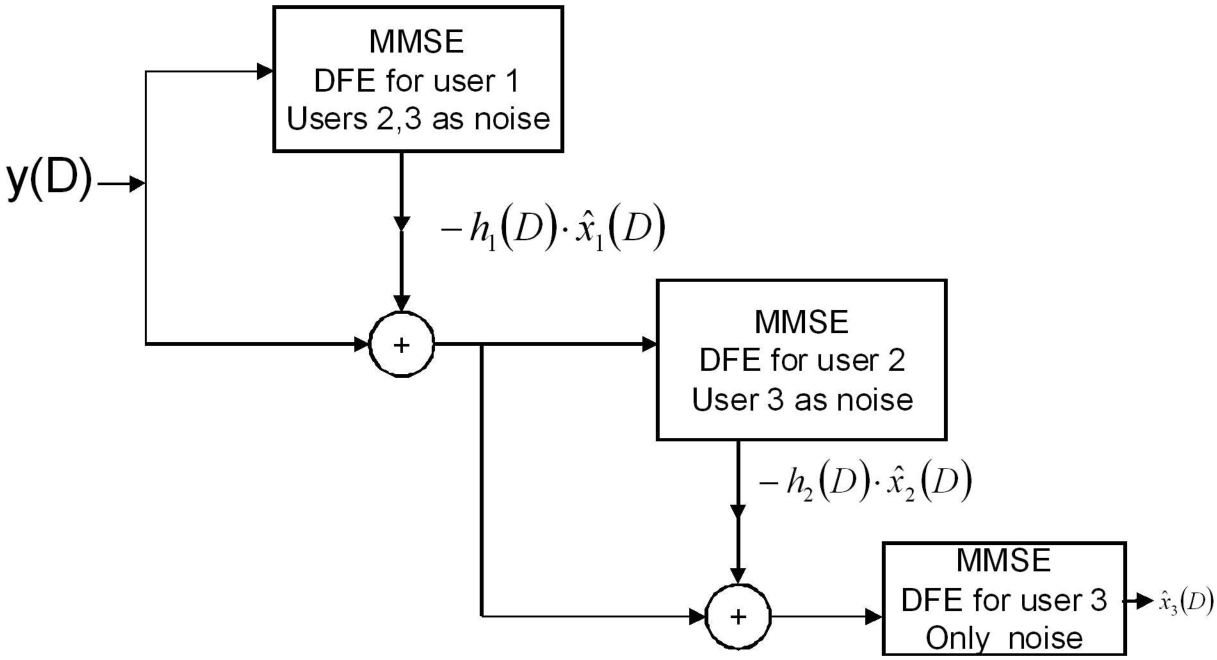

Figure 12.12: MMSE-DFE interpretation of successive cancelation.

decoding steps may be construed in more general cases as using the coefficients of the feedback filter (which are not in general equal to h_u values) for removal of ISI or other users on the same dimension. This observation forebodes some easily made mistakes in successive cancelation that often always use H_u for cancelation when those actually correspond to ZF-DFE solutions and not MMSE-DFE solutions. In the trivial cases considered in this chapter, ZF and MMSE just happen to be the same.

Figure 12.12 illustrates this concept. The cancelation becomes trivial in the case of x_u being white signals (so all noises or users plus noises are white) and the cancelation coefficient is the channel gain. But if the users spectra are not white, or there is any memory in the channel, then MMSE's will estimate an innovations white component of x_u and the feedback sections will be based on that component or equivalently determined by MMSE-DFE feedback sections that have that component on the input. For more, see Chapter 13. This subtle reduction to white innovations essentially allows a causal implementation.

The MMSE-DFE has $SNR = 2^{2\bar{I}(\boldsymbol{x}(D);\boldsymbol{y}(D))} - 1$ and is canonical, and the $I(\boldsymbol{x}(D);\boldsymbol{y}(D))$ can correspond to a chain-rule implementation for any order. If the symbol sequence $\boldsymbol{x}(D)$ has time index k, each symbol could be viewed as a distinct user in a multiple-access system with the order of decoding selected as the order of transmission in time. The successive decoding view would suggest simple removal of the contribution of each user with coefficient $h_{u=k}$, which would correspond to a ZF-DFE and a GDFE. The subtle difference is that for each $\boldsymbol{x}_k$, treating all future (uncanceled) $\boldsymbol{x}_{>k}$ symbols are treated as noise and an ML detector that observes the entire future is necessary given the past. Such an ML detector is more than the feed-forward section of the ZF-DFE. However, the feed forward section of the MMSE-DFE with appropriate use of the coefficients $g_k \neq h_k$ for feedback leads to a situation where an instantaneous (correct) decision has the same SNR as could be obtained by the infinite-length ML detector for Gaussian inputs. Thus, the ZF-DFE misses only in that instantaneous decisions do not have a canonical SNR, but a full future-observing ML detector (and not simply the feed-forward section of the ZF-DFE) would achieve the canonical performance level and the coefficients h_k could be used in such a situation. A similar argument can be made for the ZF-GDFE in that it's instantaneous SNRs are deficient, but a full ML detector could be augmented to obtain the canonical performance more easily obtained with the MMSE-GDFE.

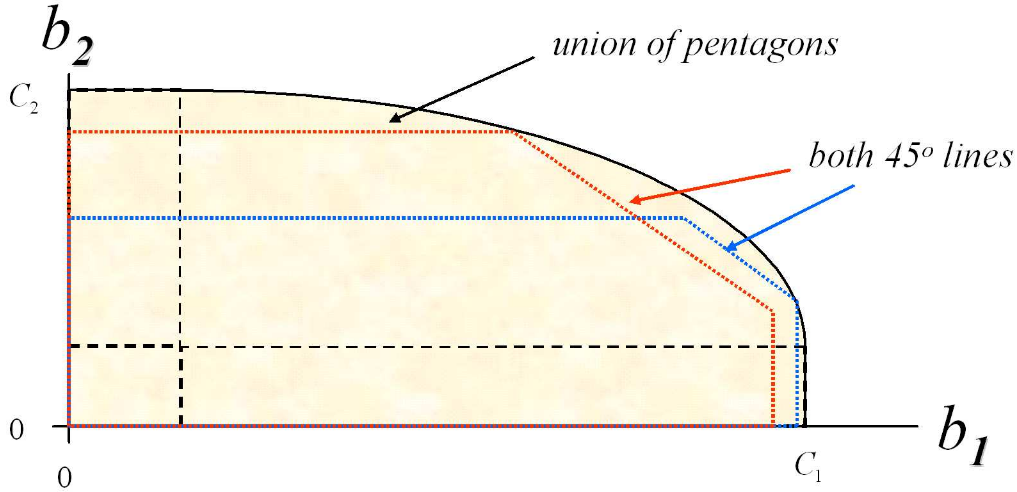

Figure 12.13: 2-user AWGN multiple-access capacity region.

12.2.3 Capacity regions for the $N = 1$ Gaussian Multiple Access Channel with ISI

The development of Subsection 12.2.2 essentially assumes that the energy of each user $\mathcal{E}_u$, or more generally the spectrum of each user, is known and bounded.

For the situation of linear time-invariant channels with pulse responses $p_u(t)$, then $\mathcal{E}_u$ is the individual energy constraint that applies to user u's integrated power spectral density. Such channels could be viewed as limiting cases of discrete time-invariant channels with DMT systems used by every user (with common symbol boundary of all users at the receiver, which is a special case of the GDFE with vector coding used and a cyclic prefix used on each of the users - in particular then the transmit singular matrix becomes constant at the channel input and is thus constant for every user and implementable without coordination).[8] There may be more than one power spectral density that meets the energy constraint for any particular user. Any of these spectra could be used with any of the possible spectra for other users, leading to a large range of possibilities for construction of rate regions. Each different power spectral density is a possible Gaussian distribution for that input that should be considered when forming the convex hull over all input probability densities. Each combination of spectra might have a different rate region.

The capacity region for this case of linear channels is the union or convex hull of all allowed rate regions. Figure 12.13 illustrates such a union of pentagon regions for $U = 2$. For a specific choice of power spectral densities, the capacity region is just one pentagon. Different user power spectral densities lead to possibly different pentagons. The union of pentagons in the general case can provide a smooth convex region that contains all the pentagons (or their convex hull, which is the same as the union for the Gaussian noise case). Each point on the boundary of the capacity region may correspond to a different use of spectra. Any such point is achievable with capacity achieving codes, and a single GDFE. The use of time-sharing is less important for achieving points in this more general capacity region, because the designer can always find a pentagon for which one of the corner points corresponds to the selected point on the capacity region boundary – that is time-sharing of GDFE's is an abberation caused by over-restriction of the input.[9] Equivalently, time-sharing really more generally corresponds to "dimension" sharing where dimensions could be time, frequency, space, or any linear combinations

[8] This tone-by-tone approach is explored further in Chapters 13 and 14.

[9] Even on two flat channels, the designer could decide to divide the spectrum according to rate (and user-channel SNR) such that a GDFE could be used in the receiver, potentially just a simple FDM or vector-coded version, to decode.

thereof. The capacity C_1 or C_2 in Figure 12.13 with all other users zeroed limits the maximum data rate for any particular user – the boundary for the capacity region may be a straight line (vertical or horizontal lines for C_1 and C_2 in Figure 12.13 respectively) at these C_1 and C_2 points. Usually, the region curves gradually to these points for practical channels with $N > 1$ and there are no such straight lines. These straight lines occur only on channels with a user who can use the entire band available or more (which is usually caused by sampling too slow in modern transmission systems and practical channels).

Calculation of the rate region for any channel (for any particular choice of input energies and power-spectral densities) can calculate the maximum rate sums for all subsets of the set of users, as in the beginning of this section. Then if all power spectral densities were somehow enumerated, the convex hull of all $2^U + U - 1$-sided polytopes describes the rate region. The enumeration of power spectral densities for a Gaussian channel is equivalent to enumerating all input probability densities.

EXAMPLE 12.2.3 (Two-user ISI) Two synchronized users use a guard period of $\nu = 1$ sample with $N = 2$ output samples on the two independent paths of a MAC with $H_2(D) = 1 + .9D$ and $H_1(D) = 1 - D$. The noise is white of variance $\sigma^2 = .181$ and added as two dimensional noise to the sum of the two channel outputs for each 2-sample packet at the common MAC receiver. These are not one-dimensional channels but are packet approximations for such channels. There are 6 input dimensions, 3 for each user and 2 output dimensions. Nonetheless, GDFE theory can be applied for any set of input autocorrelation matrices. This example first examines the situation of 6 independent input dimensions, each of energy 1 per sample:

```
>> P =[P2 P1]
    1.0000    0.9000         0    1.0000   -1.0000         0
         0    1.0000    0.9000         0    1.0000   -1.0000
>> bsum=(.5/log(2))*log((det(P*eye(6)*P'+.181*eye(2))/det(.181*eye(2))))

    4.4622
```

In this case, User 1 will be decoded first on each of its 3 dimensions by a GDFE in the receiver. The sum rate is not the highest possible, but corresponds to the choice of input spectra (autocorrelation matrices) for the two users. GDFE design continues as:

```
>> Rf=(1/.181)*P'*P

    5.5249    4.9724         0    5.5249   -5.5249         0
    4.9724   10.0000    4.9724    4.9724    0.5525   -5.5249
         0    4.9724    4.4751         0    4.9724   -4.9724
    5.5249    4.9724         0    5.5249   -5.5249         0
   -5.5249    0.5525    4.9724   -5.5249   11.0497   -5.5249
         0   -5.5249   -4.9724         0   -5.5249    5.5249

>> Rbinv=Rf+eye(6)

    6.5249    4.9724         0    5.5249   -5.5249         0
    4.9724   11.0000    4.9724    4.9724    0.5525   -5.5249
         0    4.9724    5.4751         0    4.9724   -4.9724
    5.5249    4.9724         0    6.5249   -5.5249         0
   -5.5249    0.5525    4.9724   -5.5249   12.0497   -5.5249
         0   -5.5249   -4.9724         0   -5.5249    6.5249

>> Gbar=chol(Rbinv)
```

```
    2.5544    1.9466         0    2.1629   -2.1629         0
         0    2.6853    1.8517    0.2838    1.7737   -2.0575
         0         0    1.4305   -0.3674    1.1801   -0.8127
         0         0         0    1.2772   -0.7177    0.2234
         0         0         0         0    1.5225   -0.4967
         0         0         0         0         0    1.1553

>> G=inv(diag(diag(Gbar)))*Gbar

    1.0000    0.7621         0    0.8467   -0.8467         0
         0    1.0000    0.6896    0.1057    0.6605   -0.7662
         0         0    1.0000   -0.2568    0.8249   -0.5681
         0         0         0    1.0000   -0.5619    0.1749
         0         0         0         0    1.0000   -0.3262
         0         0         0         0         0    1.0000

>> S0=diag(diag(Gbar))*diag(diag(Gbar))

    6.5249         0         0         0         0         0
         0    7.2107         0         0         0         0
         0         0    2.0463         0         0         0
         0         0         0    1.6312         0         0
         0         0         0         0    2.3182         0
         0         0         0         0         0    1.3346

>> bvec=diag((0.5/log(2))*log(S0))

    1.3530
    1.4251
    0.5165
    0.3530
    0.6065
    0.2082
>> sum(bvec)  = 4.4622
```

As expected, the mutual information is the sum of the data rates. User 1 (decoded first) is at a significant disadvantage in terms of data rate and SNR that is evident. The order of the two users could be reversed to get a different decomposition of bit rates where user 2 is at a disadvantage. The feedforward processing is also of interest:

```
>> W=inv(S0)*inv(G')

    0.1533         0         0         0         0         0
   -0.1057    0.1387         0         0         0         0
    0.2568   -0.3370    0.4887         0         0         0
   -0.3870   -0.1734    0.1574    0.6130         0         0
    0.2424   -0.1081   -0.2936    0.2424    0.4314         0
    0.0063    0.2564    0.2256    0.0063    0.2444    0.7493

>> MSWMF=(1/.181)*W*P'

    0.8467         0
    0.1057    0.7662
```

```
   -0.2568    0.5681
    0.3870   -0.1749
   -0.2424    0.3262
   -0.0063   -0.2507

>> MSWMF*P

    0.8467    0.7621         0    0.8467   -0.8467         0
    0.1057    0.8613    0.6896    0.1057    0.6605   -0.7662
   -0.2568    0.3370    0.5113   -0.2568    0.8249   -0.5681
    0.3870    0.1734   -0.1574    0.3870   -0.5619    0.1749
   -0.2424    0.1081    0.2936   -0.2424    0.5686   -0.3262
   -0.0063   -0.2564   -0.2256   -0.0063   -0.2444    0.2507

>> G

    1.0000    0.7621         0    0.8467   -0.8467         0
         0    1.0000    0.6896    0.1057    0.6605   -0.7662
         0         0    1.0000   -0.2568    0.8249   -0.5681
         0         0         0    1.0000   -0.5619    0.1749
         0         0         0         0    1.0000   -0.3262
         0         0         0         0         0    1.0000
```

The overall feedforward processing and G are interesting to compare. Feedback coefficients are identical to the forward channel as they should be. However, it is clear that the contribution of MMSE for User 1 is very significant both from feedforward ISI (GDFE packet not sufficiently long) and from User 2.

Perhaps of yet more interest would be the case where the designer recognizes that energy inserted into the individual null spaces of channel 1 and channel 2 is wasted and can be eliminated (while energy into the null space of the combined channel cannot be so avoided). In this case, the following sequence is obtained:

```
>> P2=[1 .9 0
0 1 .9]

    1.0000    0.9000         0
         0    1.0000    0.9000

>> [F2, L2, M2]=svd(P2)

F2 =   -0.7071   -0.7071
       -0.7071    0.7071

L2 = 1.6462         0         0
          0    0.9539         0


M2 = -0.4295   -0.7412    0.5158
     -0.8161    0.0741   -0.5731
     -0.3866    0.6671    0.6368

>> M2ns=M2(1:3,1:2)

   -0.4295   -0.7412
```

```
   -0.8161    0.0741
   -0.3866    0.6671

>> Ruu2= 1.5*eye(2)

    1.5000         0
         0    1.5000

>> Rxx2=M2ns*Ruu2*M2ns'

    1.1009    0.4434   -0.4927
    0.4434    1.0073    0.5474
   -0.4927    0.5474    0.8918

>> H2=P2*M2ns

   -1.1640   -0.6745
   -1.1640    0.6745

>> P1=[1 -1 0
0 1 -1]

     1    -1     0
     0     1    -1

>> [F1, L1, M1]=svd(P1)

F1 = -0.7071    0.7071
      0.7071    0.7071
```

This F1 is essentially the same as F2, a happy coincidence that will raise some questions about design heading to Chapter 13.

```
L1 =  1.7321         0         0
           0    1.0000         0

M1 = -0.4082    0.7071    0.5774
      0.8165    0.0000    0.5774
     -0.4082   -0.7071    0.5774

>> M1ns=M1(1:3,1:2)

   -0.4082    0.7071
    0.8165    0.0000
   -0.4082   -0.7071

>> Ruu1=1.5*eye(2)

    1.5000         0
         0    1.5000

>> Rxx1=M1ns*Ruu1*M1ns'
```

```
   1.0000   -0.5000   -0.5000
  -0.5000    1.0000   -0.5000
  -0.5000   -0.5000    1.0000

>> H1=P1*M1ns

  -1.2247    0.7071
   1.2247    0.7071

>> P =[P2 P1]

   1.0000    0.9000         0    1.0000   -1.0000         0
        0    1.0000    0.9000         0    1.0000   -1.0000

>> Rxx = [ Rxx2 zeros(2,2)
        zeros(2,2) Rxx1 ];

>> bsum=(.5/log(2))*log((det(P*Rxx*P'+.181*eye(2))/det(.181*eye(2))))

   5.0252
```

This rate sum is the mutual information, now for the different inputs. These inputs waste no energy into the null space of the individual user channels and thus lead to a greater rate sum. A GDFE can be designed for the 4-dimensional white input of u1 and u2.

```
>> Rf=(1/.181)*H'*H

  14.9724   -0.0000    0.0000   -9.0951
  -0.0000    5.0276    9.1286    0.0000
   0.0000    9.1286   16.5746    0.0000
  -9.0951    0.0000    0.0000    5.5249

>> Rbinv=Rf+inv(Ruu)

  15.6390   -0.0000    0.0000   -9.0951
  -0.0000    5.6943    9.1286    0.0000
   0.0000    9.1286   17.2413    0.0000
  -9.0951    0.0000    0.0000    6.1915

>> Gbar=chol(Rbinv)

   3.9546   -0.0000    0.0000   -2.2999
        0    2.3863    3.8255    0.0000
        0         0    1.6147    0.0000
        0         0         0    0.9498

>> G=inv(diag(diag(Gbar)))*Gbar

   1.0000   -0.0000    0.0000   -0.5816
        0    1.0000    1.6031    0.0000
        0         0    1.0000    0.0000
        0         0         0    1.0000
```

```
>> S0=diag(diag(Gbar))*diag(diag(Gbar))

   15.6390         0         0         0
         0    5.6943         0         0
         0         0    2.6072         0
         0         0         0    0.9022

>> (0.5/log(2))*log(det(Ruu)*det(S0))

    5.0252

>> bvec=diag((0.5/log(2))*log(Ruu.*S0))

    2.2760
    1.5472
    0.9837
    0.2182

>> sum(bvec)

    5.0252

>> b2=sum(bvec(1:2))

    3.8233

>> b1=sum(bvec(3:4))

    1.2019

>> W=inv(S0)*inv(G')

    0.0639         0         0         0
    0.0000    0.1756         0         0
   -0.0000   -0.6149    0.3836         0
    0.6446   -0.0000   -0.0000    1.1084

>> MSWMF=(1/.181)*W*H'

   -0.4112   -0.4112
   -0.6545    0.6545
   -0.3039    0.3039
    0.1846    0.1846

>> MSWMF*H

    0.9574   -0.0000    0.0000   -0.5816
    0.0000    0.8829    1.6031    0.0000
   -0.0000    0.4099    0.7443    0.0000
   -0.4297    0.0000    0.0000    0.2611

>> G
    1.0000   -0.0000    0.0000   -0.5816
         0    1.0000    1.6031    0.0000
```

```
0          0     1.0000    0.0000
0          0          0    1.0000
```

The 4-dimensional channel is very interesting. Again User 1 sees significant inteference from User 2 that is not removed. User 2 sees no interference nor ISI beyond that removed from user 1. This happy result occured because the vector coding choice, coincidentally, diagonalized both of the users. Chapter 13 will show a way to induce such an effect on any set of linear-time-invariant channels within a MAC via "vectored DMT."

The maximum rate sum over all power spectra

From Chapters 4 and 8 earlier, a single rate (whether the other users are treated as noise or zeroed) is easily computed according to water-filling for the subsequent imposed single-user channel. A variation of that procedure known as simultaneous water-filling appears in this subsection.

The limiting aggregate MT sum rate for all users in the case of ISI can be written as

$$\bar{b} = \int_{-\infty}^{\infty} \frac{1}{2} \log_2 \left[1 + \frac{\sum_{u=1}^{U} S_{x,u}(f) \cdot |H(f)|^2}{S_n(f)} \right] df \quad , \tag{12.62}$$

where the sum is replaced by an integral over all the infinitesimally small tones of a MT system. The individual energy constraints of each user cause the rate-sum maximization problem to differ slightly from straightforward water-filling. Forming the Lagrangian for each integrand with respect to input user spectra and the side constraint of non-negative power spectral density integrating to the given energy $\mathcal{E}_u$, provides

$$S_u(f) + \frac{\sigma^2 + \sum_{i \neq u} S_i(f) \cdot |H_i(f)|^2}{|H_u(f)|^2} = \lambda_u \quad . \tag{12.63}$$

Since the original function's integrand was convex and all the constraints are convex (linear inequality), then a solution exists. Each instance of (12.62) for $u = 1, ..., U$ is water-filling with all the other users viewed as "noise." This leads to the following lemma:

Lemma 12.2.2 (Simultaneous Water-Filling Optimum) *The maximum of the sum of data rates $\sum_{\boldsymbol{u}} b_{u \in \boldsymbol{u}}$ for the scalar ($N = 1$)* [10] *Gaussian AWGN with input power spectral density $S_u(f)$ for the u^{th} user satisfies*

$$S_u(f) + \frac{\sigma_u^2(f)}{|H_u(f)|^2} = \lambda_u \tag{12.64}$$

where

$$\sigma_u^2(f) = \sigma^2 + \sum_{i \neq u} S_i(f) \cdot |H_i(f)|^2 \tag{12.65}$$

with

$$S_u(f) \geq 0 \quad , \tag{12.66}$$

and

$$\int_{-\infty}^{\infty} S_u(f) \cdot df = P_u \quad . \tag{12.67}$$

$H_u(f)$ is the Fourier transform of the u^{th} channel impulse response, and P_u is the U^{th} user's power constraint. **proof:** *See above paragraph.*

The simultaneous water-filling (SWF) result is a powerful result that finds a particular solution for maximizing a particular MAC rate sum. Chapter 13 returns to find a modification of the SWF criterion that does allow rate-region construction.

[10] Chapter 13, Section 2 generalizes this to any Gaussian MAC channel, but the additional notational burden adds nothing to the present development.

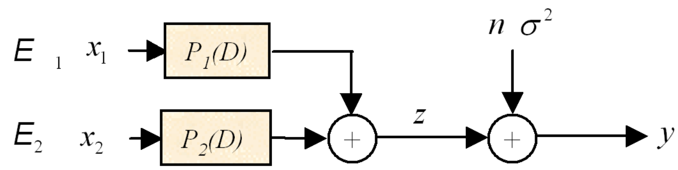

Figure 12.14: 2-user AWGN channel with ISI.

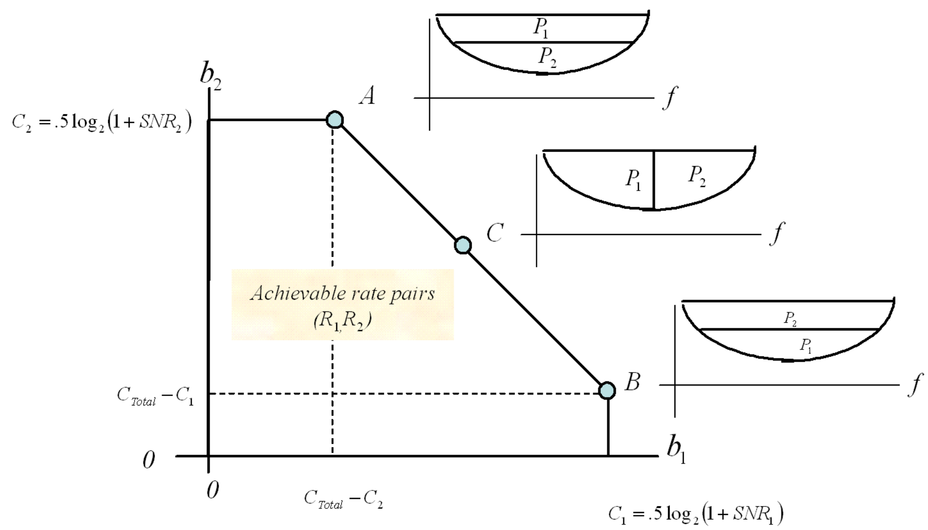

Figure 12.15: 2-user AWGN capacity region for $P_1(D) = P_2(D)$ case.

Figure 12.14 illustrates the more general 2-user AWGN with intersymbol interference. A special case arises when the two channels have equal pulse responses, $H_1(D) = H_2(D)$: In this case, the rate region remains a pentagon as illustrated in Figure 12.15 with C_1, C_2, and C_{Total} all being water-filling capacities for the multiple-access channel. Three simultaneous water-filling spectra situations appear in Figure 12.15. Any division of the water-filling spectra that achieves C_{Total} and also satisfies each individual power constraint corresponds to a valid point on the slanted boundary of the pentagon. Figure 12.15 illustrates the corresponding water-fill spectra for the 3 different situations, A, B, and C. The two corner points need not correspond to the use of a water-fill spectra for one of the two users. However, at point A, user 1 has a WF spectra when user 2 is viewed as noise, while user 2 has WF spectra as if user 1 were zero (i.e., removed by GDFE). The corresponding opposite is true for point B. Both points A and B also satisfy SWF as is evident in the power-spectral-density diagrams in Figure 12.15. Points in between can correspond to frequency-division or any other division of energy in the total water-fill region that satisfies each of the individual water-filling constraints also. Each such division corresponds to a different pair of rates with $R_1 + R_2 = C_{Total}$ (R_1, R_2) along the slanted boundary. At least one point corresponds to non-overlapping spectra (frequency-division multiplexing or FDM) if the channel output is one-dimensional. This FDM point clearly satisfies SWF individually as each user is zero in the others band. At this point, the sum of the individual user rates treating all others as noise is equal to the maximum rate sum (which is not often true for other non FDM points). An FDM point is CFC and has an independent set of optimum receivers.

For situations that are not necessarily the maximum rate sum, Chapter 13 provides methods for computing the best input spectra to achieve a point; said spectra are not necessarily simultaneous water-filling.

12.3 Rate regions for the Gaussian broadcast channel

The general broadcast capacity region is also described by Equations (12.27) - (12.30) in Section 12.1.3. Further simplification awaits invention for the general case of BC or IC. For the additive Gaussian noise case, BC rate regions can be somewhat more succinctly described and an actual algorithm implemented to construct them. A few basic concepts in Sections 12.3.1 allow illustration of BC rate regions for the ISI free channel first in Section 12.3.2. ISI channels and the vector broadcast channel (with inputs and outputs possibly corresponding to more than one dimension) are deferred to Chapter 14. Section 12.3.3 introduces the concept of duality (a concept first emanating from many former students and instantiations of this EE479 class).

12.3.1 Basic concepts for Broadcast Capacity

This subsection reviews some basic aspects of transmission that are useful for simplifying BC capacity rate-region description.

Decomposition of Gaussian process into user message components

Any stationary random process $\boldsymbol{x}_k$ decomposes into and has a one-to-one relationship with its "innovations" process $\boldsymbol{v}_k$. For a Gaussian process, the innovations sequence is found by linear MMSE prediction as discussed earlier in Chapter 5. The innovations process is also Gaussian. The entropy of the process $\boldsymbol{x}_k$, and of the innovations $\boldsymbol{v}_k$ are equal, $H_{\boldsymbol{x}} = H_{\boldsymbol{v}}$, which means their "information content" is equal. All the components of the innovations sequence are independent (which follows directly from Chapter 5's linear-prediction interpretation). If the process $\boldsymbol{x}_k$ is shared to convey simultaneously several users' messages, $\boldsymbol{x}_{u,k}$, each of these messages would have an image component in the innovations sequence, $\boldsymbol{v}_{u,k}$. Thus, for the Gaussian BC channel, at least a Gaussian input could be so decomposed into independent constituents for each of the users. The following theorem addresses the sufficiency of the Gaussian input itself:

Theorem 12.3.1 (Gaussian inputs are sufficient) *For any (linear) multi-user channel with additive Gaussian noise, the capacity region is largest with all users jointly Gaussian in distribution.*

proof:

The general capacity region in (12.27) - (12.30) is revisited here. Each of the individual "boxes" for any single output u, namely $b_u(\boldsymbol{\pi}, p_{\boldsymbol{x}})$ in (12.28) could be considered a MAC for that particular output (without regard to other receivers for other users) and therefore is largest when all users are Gaussian. Similarly, the convex hull operation for that same receiver over all orders in (12.29) can not be any larger than when the inputs are all Gaussian (again by analogy with the MAC for this one receiver). The Cartesean product in (12.28) necessarily then is also largest when all are Gaussian. Finally, then the capacity region in (12.30) need not consider distributions other than joint Gaussian as those lead to the largest set of regions (now for each covariance describing such joint Gaussian inputs) in the previous 3 steps. **QED.**

The above theorem makes the address of capacity regions for Gaussian BC (or any linear additive Gaussian noise channel) simpler. As with the single-user case and the MAC, it is sufficient to consider only Gaussian inputs for all users when the noise is additive Gaussian. Since the messages are independent, and each can be Gaussian without loss

$$H_{\boldsymbol{x}} = \sum_{u=1}^{U} H_{\boldsymbol{v}_u} \quad . \tag{12.68}$$

Furthermore, since the sum

$$\boldsymbol{x} = \sum_{u=1}^{U} \boldsymbol{x}_u \tag{12.69}$$

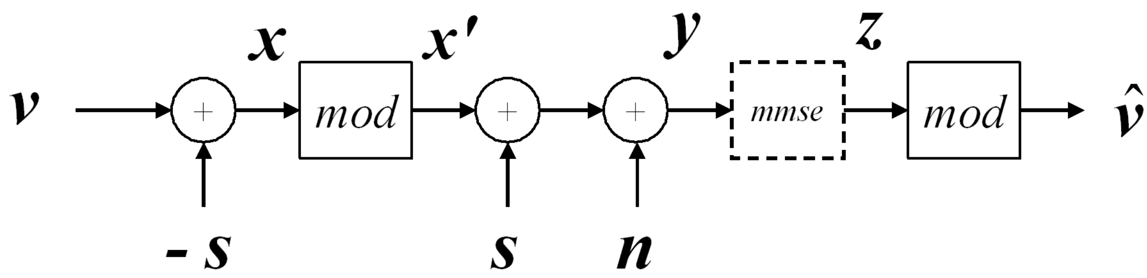

Figure 12.16: Non-causal precoding in place of successive decoding.

has also this same entropy $H_{\boldsymbol{x}}$, then there is no loss whatsoever in investigating an encoder that is the sum of the individual Gaussian components with one-to-one mappings from some white Gaussian sequence $\boldsymbol{v}_u \leftrightarrow \boldsymbol{x}_u$. As shown by Kailath, the innovations is a complete and unique representation of a Gaussian process, and clearly this innovations can be constructed by a sum of the message modulator outputs. No greater entropy or information can be contained in any other representation. Thus, there is no loss for the Gaussian Broadcast channel in assuming that the messages of the individual users are summed into a single transmit vector $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + ... + \boldsymbol{x}_U$. This result holds regardless of the number of dimensions, presence of ISI, etc. In effect, the Gaussian innovations $\boldsymbol{v}_u$ can be viewed as the message content of Gaussian user u.

The Lossless non-causal "dirty paper" precoder

Figure 12.16 illustrates non-causal precoding, which is sometimes also known as "dirty paper" precoding. An interesting result in non-causal precoding is that the addition of a sequence $\boldsymbol{s}$ in the channel known only to the transmitter does not change the capacity of the link. This result was first noted by Costa in the early 1980's for Gaussian channels and eventually extended to all channels by Forney in 2003 through a use of the so-called "crypto-lemma":

> **Theorem 12.3.2 (Forney's Crypto Lemma)** *Given the channel in Figure 12.16 with both $\boldsymbol{v}$ and $\boldsymbol{s}$ uniform in distribution over some Voronoi region Λ, then $\boldsymbol{x}' = mod_{\Lambda}(\boldsymbol{x})$ has energy $\mathcal{E}_{\boldsymbol{v}}$, $H_{\boldsymbol{x}} = H_{\boldsymbol{v}}$, and $\boldsymbol{x}'$ is independent of $\boldsymbol{s}$ and $\boldsymbol{v}$.*
>
> **proof:** *Addition modulo Λ is a group operation so that any two points added together will produce another point within the region. Let $\boldsymbol{v}$ and $\boldsymbol{s}$ first be random and uniform over Λ with $\boldsymbol{s}$ independent of $\boldsymbol{v}$, then $p_{\boldsymbol{x}'/\boldsymbol{v}}(\boldsymbol{x}' \ominus \boldsymbol{v}) = p_{\boldsymbol{s}}(\boldsymbol{x}' \ominus \boldsymbol{v}) \equiv$ constant (since s has uniform distribution) for all $\boldsymbol{x}'$ and any and all particular $\boldsymbol{v}$. Thus $\boldsymbol{x}'$ is uniform and independent of $\boldsymbol{v}$.*
>
> *Continuing if $\boldsymbol{s}$ is deterministic and not random nor uniform, then a precomputed uniform random dither sequence $\boldsymbol{d}$ may be added to $\boldsymbol{s}$ at the transmitter and subtracted at the receiver without loss. The sum $\boldsymbol{s} \oplus \boldsymbol{d}$ is uniform and the theorem then still applies. (Dither is not necessary in realization, but it helps prove the uniformity and independence.)* **QED.**

Essentially the crypto lemma describes a block precoder that executes modulo arithmetic over a block of N sample periods, where N is the number of dimensions per symbol (as usual) or the dimensionality of Λ. Because of the second modulo device in the receiver, the precoder completely removes the effect of the channel's sequence $\boldsymbol{s}$, and a detector for $\boldsymbol{v}$ can be applied at the receiver modulo's output to recover $\boldsymbol{v}$. This decoder need not know anything about $\boldsymbol{s}$. The name "crypto" arises from the independence of $\boldsymbol{v}$ from $\boldsymbol{x}'$, which basically means the input is disguised if somehow $\boldsymbol{s}$ was used as a key in encryption.

The crytpo lemma has a use in BCs: This use is for asymptotic results on the BC channel as $N \to \infty$. In the infinite-length case, the uniform distribution for a constrained average total energy per dimension over an infinite number of dimensions is Gaussian in any particular dimension. Essentially then all quantities become Gaussian, and the modulo device conceptually observes the entire sequence before reflecting it inside a infinite-dimensional hypersphere. This proves Costa's Gaussian result as a special case – the so-called "dirty paper" precoder, but it also allows one to consider $\boldsymbol{s}$ as another known user, previously encoded. Such encoding requires infinite delay, but then allows a subsequent user to be encoded with its level of energy (Gaussian case), and no more, without any loss in capacity as if the previous known user were not present or zero. This is a transmitter version of infinite delay in successive decoding, now using the modulo device at the transmitter and receiver to replace the sequence of decodings. However, there is now an order in encoding (precoding) at the transmitter, with the last user to be encoded being in the most favored position. However, earlier users in the precoder order must treat later users as noise for encoding. Broadcast channels, for reasons that will become apparent later in the discussion of MAC/BC duality, typically consider user $u = 1$ to be the one that sees only channel noise (so really last in the order since all other users $u > 1$ must therefore already be known). The "causality" of the GDFE will again be useful in BCs to transform a non-causal precoder into effectively an implementable precoder.

Essentially when the number of users is less than N, the non-causal precoder turns into the block "Tomlinson/Laroia/flexible" precoders of Chapter 5 over a region Λ selected from the various constellations. While a real "causal finite-dimensional" precoder will increase energy slightly depending on the constellation choices and energies, the reader will recall such energy loss is small. Thus, for at least the situation of more dimensions than users, a simple precoder may suffice to handle broadcast channels. Precoders are revisited again in Chapter 14.

Of interest to the user, but not further used in this text, is the curious result of Forney that extends the CDEF result for finite block lengths via the crypto lemma, namely that the highest rate achievable for any situation

$$\bar{c}(\Lambda) \geq \bar{c} - \frac{1}{2}\log_2\left(2\pi e G(\Lambda)\right) \tag{12.70}$$

where $G(\Lambda)$ is the so-called normalized second moment of the lattice Λ's Voronoi (or decision) region,

$$G(\Lambda) = \frac{\mathcal{E}(\Lambda)}{V(\Lambda)^{2/N}} \quad . \tag{12.71}$$

and $\mathcal{E}(\Lambda)$ is the energy of a continuous uniform distribution over the Voronoi region. Essentially, then the non-causal precoder achieves a capacity less than the full capacity by the amount shown in Equation (12.70), which goes to zero as $N \to \infty$. Basically, the MMSE-GDFE structure can be implemented as a precoder without loss in the Gaussian case.

precoding

The crypto lemma is used in BC's by letting $\boldsymbol{x} = \boldsymbol{x}_2$ so effectively user 2 is encoded first (with user 1 viewed as noise or random), but then user 1 treats the ultimate user-2 encoder output $\boldsymbol{x}_2$ (over the entire block of N samples) as known à priori for the encoding of x_1, which is achieved modulo Λ. User 1 essentially is then free of the effects of user 2. Precoding via the crytpo lemma concept is an alternative transmitter-side implementation of successive decoding. Figure 12.17 illustrates the transmitter and the corresponding user receivers for precoding.

Thus, with this interpretation, precoding can be used mod-sphere without loss, allowing the GDFE theory to also apply to the BC without loss. This application of GDFE theory lead to what is often called BC/MAC duality.

Successive Decoding

Successive decoding is the receiver's equivalent of non-causal precoding. A Gaussian BC system can use either successive decoding or non-causal precoding to achieve best performance. Thus, the precoder circuitry prior to and including mod at the transmitter in Figure 12.17 can be avoided if successive decoding at each receiver is preferred as the implementation where those users of higher index in the current order are decoded first (as they can be, since they were encoded viewing this user and all ahead of themselves as noise). Figure 12.18 illustrates successive decoding for the same situation as the non-causal precoder of Figure 12.17. The overall complexity, as well as the complexity of each receiver is higher with successive decoding than with precoding, but the transmitter is simpler. Overall, precoding is simpler to implement and often preferred.

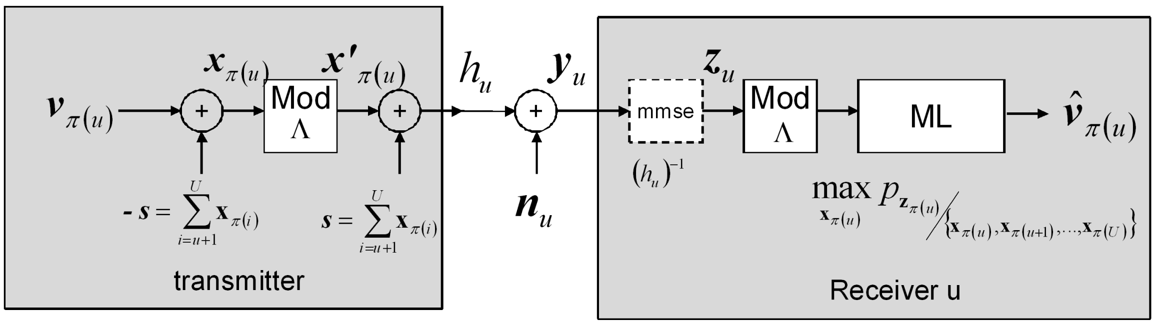

Figure 12.17: Transmitter and receivers for BC with precoders.

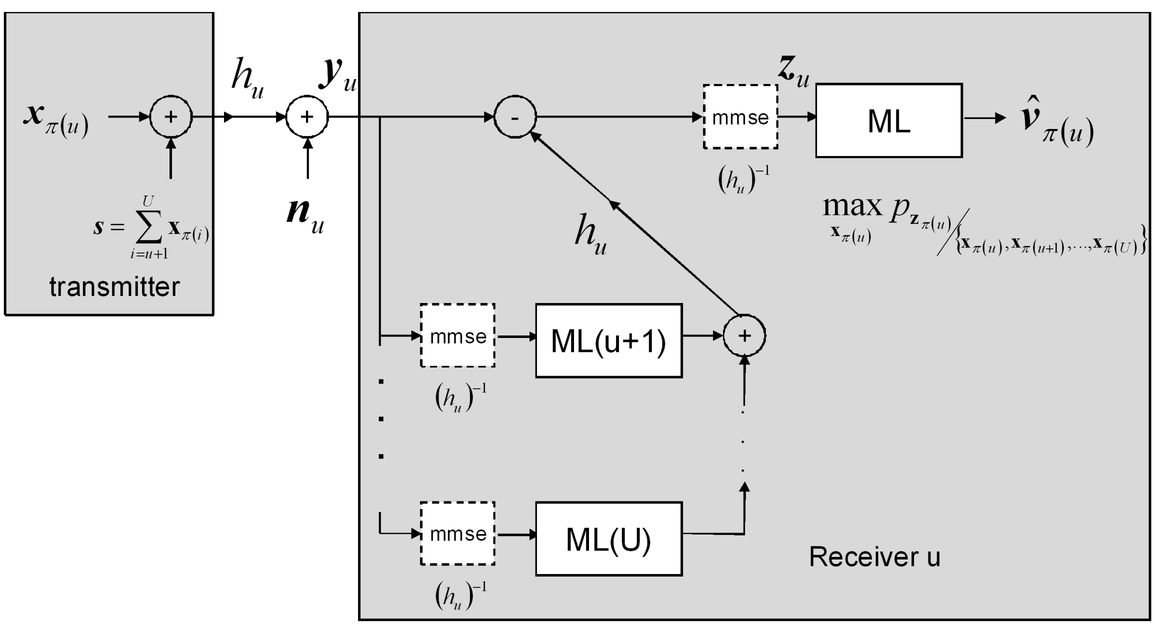

Figure 12.18: Transmitter for BC with successive decoding receivers.

Worst-Case Noise Autocorrelation Diagonalizes the MS-WMF of a GDFE

For any linear Gaussian noise channel

$$\boldsymbol{y} = P\boldsymbol{x} + \boldsymbol{n} \tag{12.72}$$

with a rank-U channel, Chapter 5 proved that the performance of a MMSE-GDFE will have the lowest possible SNR (biased or unbiased) when the Gaussian noise autocorrelation matrix is such that the MS-WMF matrix/filter of a GDFE is diagonal, and furthermore there exists a specific implementation of the input autocorrelation matrix that led to the worst-case noise for which a canonical GDFE (i.e., with white diagonal input $\boldsymbol{v}$) can be implemented for this diagonal MS-WMF (or worst-case noise) situation. A restriction noted in Chapter 5 was that the channel rank $\rho(H)$ must equal the number of output dimensions – BCs can easily violate this condition and thus can be rank deficient. When the channel is rank deficient, the input energy lost in singular dimensions should be eliminated from the transmit signal or better reallocated to the useful dimensions. It is then possible to determined a GDFE for the remaining channel-pass-space input dimensions for which the overall feedforward processing is diagonal. The diagonal MS-WMF is pertinent to the BC channel because it corresponds to "no coordination" among the receivers, but is also a highest performance MMSE GDFE with maximum SNR for the given worst-case noise.

As early in Sections 12.2.2 and 12.2.3, the GDFE extends readily to the case of all ranks of the MAC by recognizing the signal-energy component of the MMSE as other users' energy. This same concept extends to the BC channel where the feedback section coefficients are implemented as part of the precoder, possibly now adding other-user energy on to shared dimensions in the common case where $\rho(H) < U$. Diagonalization of the GDFE receiver forward-processing corresponds to no coordination among the receivers, or equivalent satisfies the BCs' no-coordination constraint. For broadcast channels with rank less than the number of outputs, this diagonizalization however may be somewhat trivial in scope as an example later in this section shows. The worst-case noise equation (see Chapter 5, Sections 5.3.6 and 5.5.5) of

$$R_{wcn}^{-1} - \left[HR_{\boldsymbol{xx}}H^* + R_{wcn}\right]^{-1} = -D \text{ (diagonal)} \tag{12.73}$$

can be satisfied for any dimensionality of the channel and input, subject to the maintenance of the indivdual noise covariance (power) constraints for each of the separate receiver-input noises.

12.3.2 Gaussian BC Channel Capacity regions without ISI

The two-user scalar AWGN BC appears in Figure 12.19. For this simple case, no generality is lost in assuming the noise variances on the two channels are equal and that $|h_1| \geq |h_2|$.[11] More generally, the channel SNR quantities

$$g_u = \frac{|h_u|^2}{\sigma_u^2} \tag{12.74}$$

are reintroduced without loss of generality assuming $g_1 \geq g_2$, or essentially $|h_u| = \sigma_u \cdot \sqrt{g_u}$. (Effectively, the noise variances are set equal because any difference can be accommodated in redefining the scale factors h_1 and h_2.) The energy is partitioned between the two users who share the common signal $\boldsymbol{x}$ so that

$$\mathcal{E}_1 = \alpha \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \tag{12.75}$$

$$\mathcal{E}_2 = (1-\alpha) \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \quad , \tag{12.76}$$

where $0 \leq \alpha \leq 1$. A successive-decoding approach to this channel would decode the entire signal $\boldsymbol{x}_2$ first in receiver 1, and thus remove its effect prior to decoding $\boldsymbol{x}_1$. The receiver of user 2 treats user 1's signal as Gaussian noise. Then the data rates are bounded by the achievable combinations

$$\bar{b}_1 \leq I(x_1; y_1/x_2) = \frac{1}{2}\log_2\left(1 + \alpha \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_1\right) \tag{12.77}$$

[11] This relation leads eventually to user 1 being last in a best decoding/precoding order, unlike this text's usual convention of the last user being last. The reason is this order will be convenient for reasons not yet apparent until the duality of Section 12.3.3.

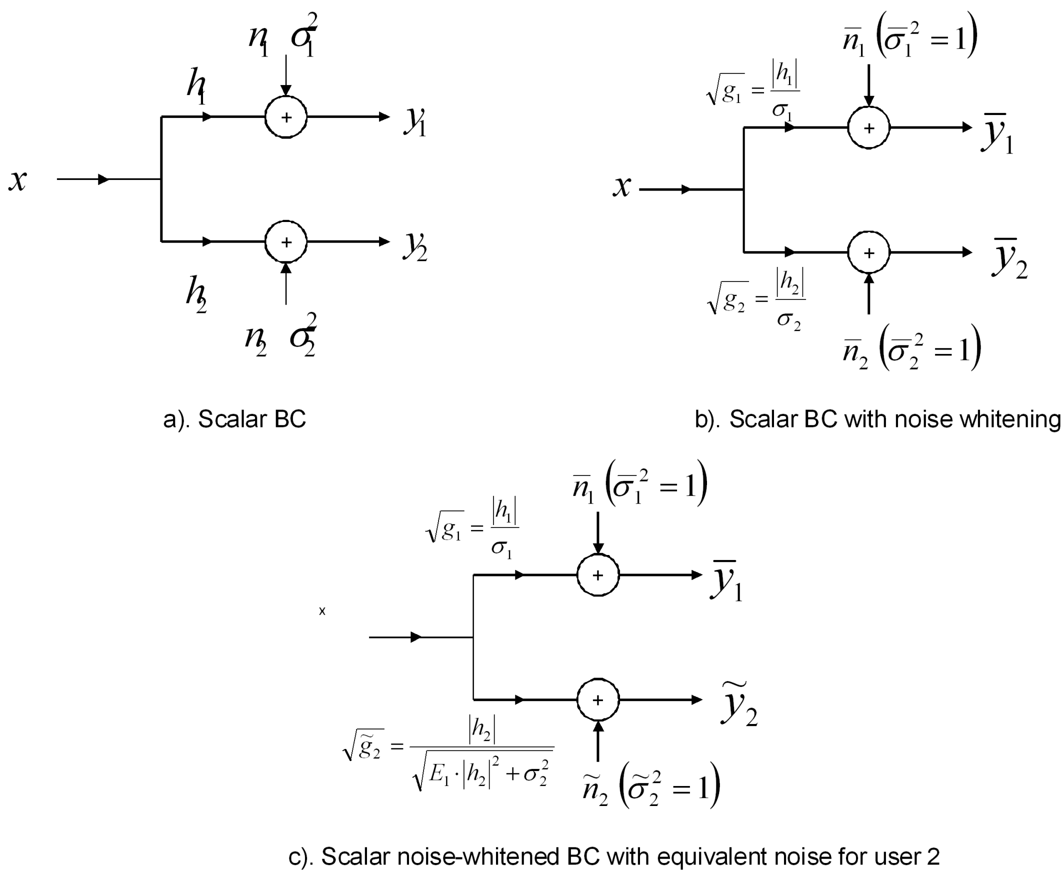

Figure 12.19: Simple two-user broadcast AWGN channel with $g_1 \geq g_2$.

$$\bar{b}_2 \quad \leq \quad I(x_2; y_2) = \frac{1}{2}\log_2\left(1 + \frac{(1-\alpha)\cdot\bar{\mathcal{E}}_{\boldsymbol{x}} g_2}{1+\alpha\cdot\bar{\mathcal{E}}_{\boldsymbol{x}}\cdot g_2}\right) \tag{12.78}$$

Equivalently, user 2 could be side information for a noiseless precoder that is used in encoding user 1 (no successive decoding then at receiver 1). Clearly the data rate of user 1 cannot be higher by single-user capacity results for any value of α. For each such value of α, the $\bar{b}_2$ that corresponds to this same value of α is also the highest because $\boldsymbol{x}_1$ will be Gaussian and because $\boldsymbol{x}$ itself is Gaussian. Then $\boldsymbol{x}_2$ can also use a Gaussian code since (from earlier sections) there is then no loss since user 1 and the noise are Gaussian, and the channel is linear. Furthermore, since $g_2 < g_1$, then receiver 2 cannot decode $\boldsymbol{x}_1$ (even if $\boldsymbol{x}_2$ were known, which it is not). Thus, any higher rate than that in Equation 12.78 would be a violation of the single-user capacity-theorem converse for user 2 with user1-plus-noise as the noise. In effect, no other probability distributions than those with the full sum energy of 1 need to be considered in constructing the general Gaussian capacity region's boundary.

This result extends by induction for $U > 2$, by numbering the users (without loss of generality) so that $g_1 > g_2 > g_3 > ... > g_U$. Then user U is decoded first at receiver U with all the other signals considered as Gaussian noise, then receiver $U-1$ decodes user $U-1$ with users $1...U-2$ treated as noise but user U removed (non-causally) and so forth. (A sequence of lossless precoders could also instead be used at the transmitter.) U, $U-1$, ..., 1 is the only order that needs to be considered in this simple BC channel because, by induction from the case of 2 users, all other rate points for other orders will lie within this region (as long as energies are such that their sum is the maximum sum energy allowed). To trace the boundary of the capacity rate region, then the designer needs to compute rate tuples for all possible combinations of U-way energy assignments that sum to the total energy. This requires in general $U-1$ energy factors α_u such that

$$\mathcal{E}_u = \alpha_u \cdot \mathcal{E} \;\forall\; u = 1, ..., U \tag{12.79}$$

where $\alpha_U = 1 - \sum_{i=1}^{U-1} \alpha_i$ and

$$0 \leq \alpha_u \leq 1 \quad . \tag{12.80}$$

The following numerical example illustrates the BC's data-rate loss caused by not being able to coordinate at the receiver in the capacity region for a simple 2-user broadcast channel

EXAMPLE 12.3.1 (Simple Broadcast Channel) Returning to Figure 12.19, let $h_1 = .8$ and $h_2 = .5$ and $\sigma^2 = .0001$. The mutual information upper bound (that may require receiver coordination and not be attainable) is

$$I(x; \boldsymbol{y}) = \frac{1}{2}\log_2\left(\frac{|R_{\boldsymbol{yy}}|}{|R_{\boldsymbol{nn}}|}\right) = \frac{1}{2}\log_2\left(\frac{(.6401)\cdot(.2501) - .4^2}{.0001^2}\right) = 6.56 \text{ bits/dimension.} \tag{12.81}$$

The sum of the data rates thus cannot exceed 6.56 bits/dimension for the BC. Using the exact formulas in (12.77) and (12.78) for various α produces the following table:

α	b_1	b_2	$b_1 + b_2$
1.0	6.32	0	6.32
.75	6.12	.20	6.32
.50	5.82	.50	6.32
.25	5.32	1.0	6.32
.10	4.66	1.66	6.32
.05	4.16	2.20	6.26
0	0	5.64	5.64

The corresponding rate region appears in Figure 12.20. The maximum rate sum in (12.81) is never attained, indicating the loss from the inability to coordinate the two receivers. The rate sum in this example is relatively fixed except for very small values of α. Such a nearly fixed rate sum need not be the case in general.

For this channel, all input energy allocations in the table correspond to a GDFE with $\mathcal{E}_x = 1$. There is a worst-case noise and the steps to the diagonalization are as below.

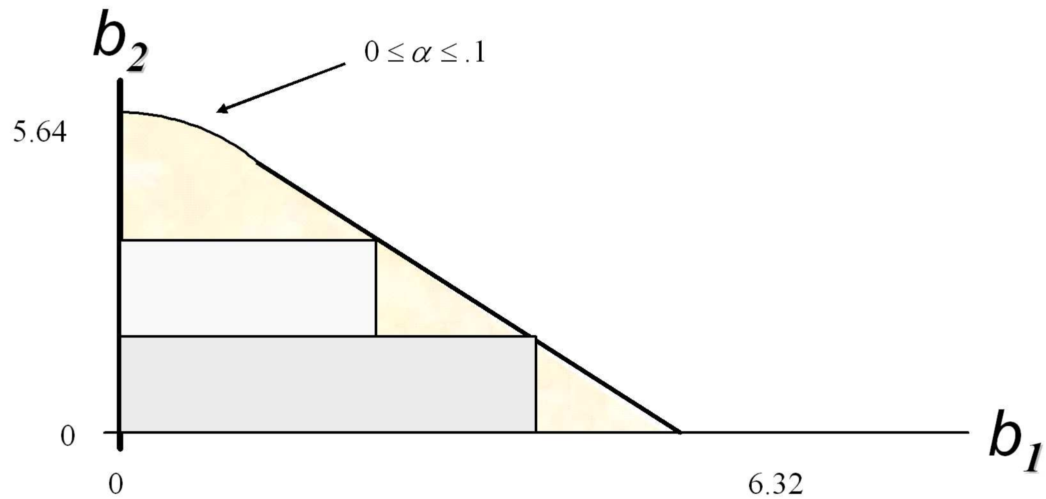

Figure 12.20: Rate region for broadcast channel

```
>> H=100*[.5
.8] =

    50
    80
>> [Rwcn,rate]=wcnoise(1,H,1)=

    1.0000    0.6250
    0.6250    1.0000

rate =  6.3220

>> Htilde=inv(Rwcn)*H =

    0.0002
   79.9999

>> J2=[ 0 1
1 0];
>> J1=1;
>> [Q,R]=qr(J1*Htilde'*J2)

Q =  1
R = 79.9999    0.0002

>> Q=J1*Q*J1;
>> R=(J1*R*J2)' =

    0.0002
   79.9999

>> R*Q' =
```

```
    0.0002
   79.9999   (checks)
>> U=1;
>> D=inv(Rwcn)-inv(H*H' + Rwcn) =

    0.0000    0.0000
    0.0000    0.9998

>> A=1;
>> Rbinv=H'*inv(Rwcn)*H + 1 =

  6.4010e+003

>>  G=1;
>> S0=Rbinv;
>> W=inv(S0)*inv(G')*A'*H'*inv(Rwcn) =

    0.0000    0.0125

>> snr=10*log10(det(S0)-1) =  38.0618

>> .5/log10(2)*log10(1+10^(snr/10)) =  6.3220
```

The value of W clearly indicates the GDFE is telling us the best energy assignment is to allocate all energy to user 1. While a GDFE can be defined to estimate each of the two inputs x_1 and x_2 for any modulation input matrix A by simply augmenting the channel to

$$H = \begin{bmatrix} .5 & .5 \\ .8 & .8 \end{bmatrix} \quad , \tag{12.82}$$

this singular channel has a rank less than the number of outputs, the condition in Chapter 5 for which a diagonalization of the GDFE foward processing does not exist. One can check the worst-case noise does exist (and in fact is the same as for the original channel), but the diagonalizing input cannot exist. Under such situations, a component of the input that does not pass the channel should actually be considered as noise (or a contribution to a MMSE as we saw for the GDFE on MAC channels), and this new set of noise variances is then used to determine the worst-case noise/MSE correlation. This annomaly lead to the concept of duality, first posed and addressed in an earlier version of these course notes, and to be addressed shortly.

Figure 12.21 illustrates the 2-user precoder instead of a successive decoder, as opposed to the GDFE implied in the above example. The signals are Gaussian and the modulo device is presumed the infinite-dimensional hypershpherical device with output power $\mathcal{E}_1$. The total energy is $\mathcal{E}_x = \mathcal{E}_1 + \mathcal{E}_2$. The precoding gain $\tilde{g}_2 = \frac{g_2}{1+\alpha_1 \cdot \bar{\mathcal{E}} \cdot g_2}$ presumes the use of scalar noise whitening (scaling) for the equivalent noise that is the sum of a scaled user 1 (encoded last in the order) and noise 2. Thus, x_2 is the "side" information in this precoder.

The broadcast capacity region for general $1 \times U$ Gaussian channel is then (assuming the ordering $|h_1| \geq |h_2| \geq |h_3| \geq ... \geq |h_U|$) can be written using mutual information as

$$b_1 \leq I(x_1; y_1/x_2, x_3, ..., x_U) = \frac{1}{2}\log_2\left(1 + \alpha_1 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} g_1\right) \tag{12.83}$$

$$b_2 \leq I(x_2; y_2/x_3, ..., X_U) = \frac{1}{2}\log_2\left(1 + \frac{\alpha_2 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_2}{1 + \alpha_1 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_2}\right) \tag{12.84}$$

$$b_3 \leq I(x_3; y_3/x_4, ..., X_U) = \frac{1}{2}\log_2\left(1 + \frac{\alpha_3 \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_3}{1 + (\alpha_1 + \alpha_2) \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_3}\right) \tag{12.85}$$

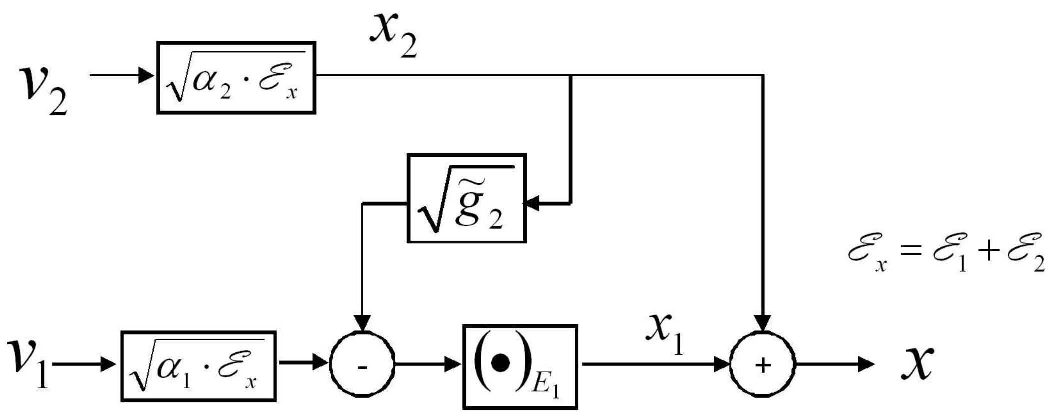

Figure 12.21: Precoder for 2-user case with Gaussian signals.

$$\vdots \tag{12.86}$$

$$b_U \quad \leq \quad I(x_U; y_U) = \frac{1}{2}\log_2\left(1 + \frac{\alpha_U \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_U}{1 + (\sum_{u=1}^{U-1} \alpha_u) \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot g_U}\right) \tag{12.87}$$

The rate regions in (12.77) and (12.78) follow in the obvious direct way for nonzero gaps. As with the MA case, the regions are approximate because the gap approximation is not exact:

$$b_1 \quad \leq \quad I(x_1; y_1/x_2) = \frac{1}{2}\log_2\left(1 + \frac{\alpha \cdot \bar{\mathcal{E}}_{\boldsymbol{x}}|h_1|^2}{\Gamma \cdot \sigma^2}\right) \tag{12.88}$$

$$b_2 \quad \leq \quad I(x_2; y_2) = \frac{1}{2}\log_2\left(1 + \frac{(1-\alpha) \cdot \bar{\mathcal{E}}_{\boldsymbol{x}}|h_2|^2}{\Gamma \cdot (\sigma^2 + \alpha \cdot \bar{\mathcal{E}}_{\boldsymbol{x}} \cdot |h_2|^2)}\right) \quad . \tag{12.89}$$

This can be expanded to more than two users in the obvious way. Further, as with the MAC, the results confuse the gap concept. An investigation (such as done by S. Jagannathan in Problem 12.14) awaits the efforts of a motivated engineer).

12.3.3 Scalar duality for Gaussian MAC/BC Channels

Duality is a concept that arises from the GDFE approach to multi-user transmission (and was first posed in an earlier on-line edition of this book). The interchange of successive decoding and non-causal precoding (both as instances of a possibly degenerate GDFE) essentially allows movement of the "coordinated end" of a multi-user Gaussian channel from transmitter to receiver and vice-versa, just as they are moved in GDFE theory. The "other-user" noise into receiver u in the BC case is filtered by the same filter h_u for all other users; while for the common MAC receiver, the "other-user" noise from user i component is filtered by h_i for detection of user u. The BC has a single common energy constraint

$$\sum_{u=1}^{U} \mathcal{E}_u \leq \mathcal{E}_{\boldsymbol{x}} \quad , \tag{12.90}$$

which more directly follows the GDFE single energy constraint, while the MAC has additional restrictions of energy per component $\mathcal{E}_u$ individually bounded or an energy vector $\boldsymbol{E}$ such that

$$\begin{bmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_U \end{bmatrix} \preceq \boldsymbol{E} \tag{12.91}$$

where $\preceq$ denotes component-wise $\leq$. Duality envisions a set of MAC's, each with its own $\boldsymbol{E}$ such that any and all sum-of-components is less than $\mathcal{E}_{\boldsymbol{x}}$ or

$$\{\boldsymbol{E} \mid \mathbf{1}^* \boldsymbol{E} \leq \mathcal{E}_{\boldsymbol{x}}\} \quad . \tag{12.92}$$

There is also an order reversal in duality that will allow the energy-sum constraint to be retained for both MAC and BC energies (as is shown momentarily). The gap is 0 dB, and thus Gaussian capacity-achieving codes are used by all users.

The following table lists the number of bits per user for the MAC and the BC with the order reversal in most preferred position evident (User 1 is most preferred on the BC while user U is most preferred on the MAC. The reason for the earlier order reversal is now apparent as it simplifies the indexing in the ensuing table of dualities.) The bits per user symbol in the table have been assumed equal, thus imposing a constraint on the energies selected[12]

MAC	BC
$b_1 = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_1^{MAC}\cdot g_1}{1+\mathcal{E}_2^{MAC}\cdot g_2 + ... + \mathcal{E}_U^{MAC}\cdot g_U}\right)$	$b_1 = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_1^{BC}\cdot g_1}{1}\right)$
$b_2 = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_2^{MAC}\cdot g_2}{1+\mathcal{E}_3^{MAC}\cdot g_3 + ... + \mathcal{E}_U^{MAC}\cdot g_U}\right)$	$b_2 = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_2^{BC}\cdot g_2}{1+\mathcal{E}_1^{BC}\cdot g_2}\right)$
$\vdots$	$\vdots$
$b_U = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_U^{MAC}\cdot g_U}{1}\right)$	$b_U = \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_U^{BC}\cdot g_U}{1+\left[\mathcal{E}_1^{BC}+...+\mathcal{E}_{U-1}^{BC}\right]\cdot g_U}\right)$

For equality of bit rates, these equations follow

$$\mathcal{E}_1^{BC} = \mathcal{E}_1^{MAC} \cdot \frac{1}{1 + \mathcal{E}_2^{MAC}\cdot g_2 + ... + \mathcal{E}_U^{MAC}\cdot g_U} \tag{12.93}$$

$$\mathcal{E}_2^{BC} = \mathcal{E}_2^{MAC} \cdot \frac{1 + \mathcal{E}_2^{BC}\cdot g_2}{1 + \mathcal{E}_3^{MAC}\cdot g_3 + ... + \mathcal{E}_U^{MAC}\cdot g_U} \tag{12.94}$$

$$\vdots = \vdots \tag{12.95}$$

$$\mathcal{E}_U^{BC} = \mathcal{E}_U^{MAC} \cdot \left(1 + \left[\mathcal{E}_1^{BC} + ... + \mathcal{E}_{U-1}^{BC}\right]\cdot g_U\right) \tag{12.96}$$

$$\tag{12.97}$$

Theorem 12.3.3 (Equal Sum Energy for Duality) *The sum of the energies in scalar duality is the same for the MAC and the BC when the bit rates are set equal, namely*

$$\sum_{u=1}^{U} \mathcal{E}_u^{BC} = \sum_{u=1}^{U} \mathcal{E}_u^{MAC} \quad . \tag{12.98}$$

Proof: *The sum of the equations:*

$$\mathcal{E}_1^{BC}\cdot\left[1 + \mathcal{E}_2^{MAC}\cdot g_2 + ... + \mathcal{E}_U^{MAC}\cdot g_U\right] = \mathcal{E}_1^{MAC} \tag{12.99}$$

[12] While the bit rates could always be set equal and the consequent energies of MAC derived from those of BC, the reversed-order of preference allows the sum of the energies to be the same.

$$\begin{aligned} \mathcal{E}_2^{BC} \cdot \left[1 + \mathcal{E}_3^{MAC} \cdot g_3 + ... + \mathcal{E}_U^{MAC} \cdot g_U\right] &= \mathcal{E}_2^{MAC} \cdot \left[1 + \mathcal{E}_2^{BC} g_2\right] && (12.100) \\ \vdots &= \vdots && (12.101) \\ \mathcal{E}_U^{BC} &= \mathcal{E}_U^{MAC} \cdot \left(1 + \left[\mathcal{E}_1^{BC} + ... + \mathcal{E}_{U-1}^{BC}\right] \cdot g_U\right) \end{aligned}$$

equals

$$\sum_{i=1}^{U} \mathcal{E}_i^{BC} + \sum_{i=1}^{U} \mathcal{E}_i^{BC} \cdot \sum_{k=i+1}^{U} g_k \cdot \mathcal{E}_k^{MAC} = \sum_{i=1}^{U} \mathcal{E}_i^{MAC} + \sum_{i=2}^{U} \mathcal{E}_i^{MAC} \cdot g_i \cdot \sum_{k=1}^{i-1} \mathcal{E}_k^{BC} \quad (12.102)$$

Inspection of the result of the 2^{nd} terms on the left and on the right in (12.102) when $U = 3$ provides insight:

$$\mathcal{E}_1^{BC}\left(g_2 \cdot \mathcal{E}_2^{MAC} + g_3 \cdot \mathcal{E}_3^{MAC}\right) + \mathcal{E}_2^{BC} \cdot g_3 \mathcal{E}_3^{MAC} = g_2 \cdot \mathcal{E}_1^{BC} \cdot \mathcal{E}_2^{MAC} + g_3 \cdot \left(\mathcal{E}_1^{BC} \cdot \mathcal{E}_3^{MAC} + \mathcal{E}_2^{BC} \cdot \mathcal{E}_3^{MAC}\right) \quad : \quad (12.103)$$

They are equal, and thus the remaining terms on the left and right of (12.102) are also equal. In general, the 2nd term on the left then can be rewritten

$$\sum_{i=2}^{U} \cdot \mathcal{E}_i^{MAC} \cdot g_i \cdot \sum_{k=1}^{i-1} \mathcal{E}_k^{BC} \quad (12.104)$$

and thus again is equal to the 2nd term on the right, leaving the energy sums equal for the BC and the MAC. **QED.**

A simpler and more intuitive proof appears in Chapter 14, Section 5, but requires a vector generalization and some nomenclature and defined quantities not yet introduced nor otherwise needed here in the scalar case.

EXAMPLE 12.3.2 (Example 12.3.1 revisited with duality) Returning again to Figure 12.19, let $h_1 = .8$ and $h_2 = .5$ and $\sigma^2 = .0001$. As an example point for duality, let $\mathcal{E}_1^{BC} = \mathcal{E}_2^{BC} = .5$, or equivalently $\alpha = .5$. Then, $g_1 = 80^2$, $g_2 = 50^2$ and

$$\mathcal{E}_2^{MAC} = \frac{\mathcal{E}_2^{BC}}{1 + \mathcal{E}_1^{BC} \cdot g_2} = \frac{1}{2}\left(\frac{1}{1 + 2500(.5)}\right) = \frac{1}{2502} \quad , \quad (12.105)$$

while then

$$\mathcal{E}_1^{MAC} = \mathcal{E}_1^{BC} \cdot (1 + g_2 \cdot \mathcal{E}_2^{MAC}) = (.5) \cdot \left(1 + \frac{2500}{2502}\right) = \frac{2501}{2502} = 1 - \mathcal{E}_2^{MAC} \quad . \quad (12.106)$$

The sum of energies is again 1 as it should be.

Then the two rates for the dual MAC channel are

$$\begin{aligned} b_1 &= \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_1^{MAC} \cdot g_1}{1 + \mathcal{E}_2^{MAC} \cdot g_2}\right) = 5.82 && (12.107) \\ b_2 &= \frac{1}{2}\log_2\left(1 + \frac{\mathcal{E}_2^{MAC} \cdot g_2}{1}\right) = .50 \quad , && (12.108) \end{aligned}$$

which are the same as for this energy point in the original instance of this example. Thus, the rate region could be traced for set of dual MAC channels for which the sum of the energy constraints are 1.

The GDFE for this dual MAC channel could be designed similar to the example in Section, where the diagonalization of the receiver corresponds to the separation of the transmitters inherent in the "choice" of the input (really the restrictions of the input to no collaboration). As channels with full rank occur in Chapters 13 and 14, a richer relationship of the duals and GDFE evolves.

Duality makes use of an existing MAC capacity-region program generator (for instance an easy pentagon generator for $U = 2$ and the Gaussian scalar MAC) and then forms the union of such regions for all possible energies that sum to the total allowed. The vector of energy constraints for a MAC is

$$\boldsymbol{E} = \begin{bmatrix} \mathcal{E}_{1,max} \\ \vdots \\ \mathcal{E}_{U,max} \end{bmatrix} \quad . \tag{12.109}$$

The capacity rate region of the linear Gaussian BC would then be traced by the union of all the rate regions for the MAC:

1. Initialize capacity set to $\{c(\boldsymbol{b})\} = \emptyset$.
2. for all $\boldsymbol{E}$ such that $\boldsymbol{1}^*\boldsymbol{E} = \mathcal{E}_{\boldsymbol{x}}$
   - Compute $a(\boldsymbol{b})$ for the MAC defined by g_1, ..., g_U with $\boldsymbol{E}$ as energy constraint.
   - Form $\{c(\boldsymbol{b})\} = \bigcup \{\{c(\boldsymbol{b})\}, a(\boldsymbol{b})\}$.

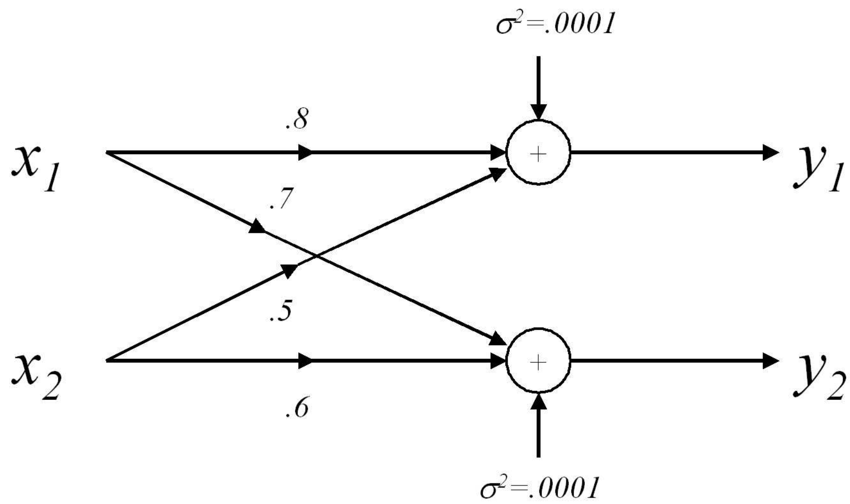

Figure 12.22: Interference-channel example.

12.4 The Interference Channel

This section provides a 2×2 AWGN example of IC capacity region construction using the general method of Section 12.1.3. Very little more is known about simplification of the region's construction, unlike the BC and MAC situations. This section further addresses only the Gaussian situation.

Figure 12.22 shows the IC example. There are only two orders that need to be considered $\pi = \{(1,2)\ ,\ (2,1)\}$. Some useful expressions are

$$\frac{1}{2}\log_2\left(1+\frac{.64}{.0001}\right) = 6.3220 \tag{12.110}$$

$$\frac{1}{2}\log_2\left(1+\frac{.36}{.0001}\right) = 5.9071 \tag{12.111}$$

$$\frac{1}{2}\log_2\left(1+\frac{.64}{.2501}\right) = 0.9157 \tag{12.112}$$

$$\frac{1}{2}\log_2\left(1+\frac{.36}{.4901}\right) = 0.3973 \tag{12.113}$$

$$\frac{1}{2}\log_2\left(1+\frac{.49}{.3601}\right) = 0.6196 \tag{12.114}$$

$$\frac{1}{2}\log_2\left(1+\frac{(.5).64}{.0001}\right) = 5.8222 \tag{12.115}$$

$$\frac{1}{2}\log_2\left(1+\frac{(.5).36}{.0001}\right) = 5.4073 \tag{12.116}$$

$$\frac{1}{2}\log_2\left(1+\frac{.64}{.1251}\right) = 1.3063 \tag{12.117}$$

$$\frac{1}{2}\log_2\left(1+\frac{.18}{.4901}\right) = 0.2257 \tag{12.118}$$

$$\frac{1}{2}\log_2\left(1+\frac{.49}{.1801}\right) = 0.9478 \tag{12.119}$$

$$\frac{1}{2}\log_2\left(1+\frac{.36}{.2451}\right) = 0.6519 \tag{12.120}$$

$$\frac{1}{2}\log_2\left(1+\frac{.5(.64)}{.2501}\right) = 0.5944 \quad (12.121)$$

$$\frac{1}{2}\log_2\left(1+\frac{.5(.49)}{.3601}\right) = 0.3744 \quad (12.122)$$

$$\frac{1}{2}\log_2\left(1+\frac{.5(.25)}{.6401}\right) = 0.1287 \quad (12.123)$$

$$\frac{1}{2}\log_2\left(1+\frac{.25}{.3201}\right) = 0.4163 \quad (12.124)$$

Again from Section 12.1.1, $X_\pi(u)$ is defines a set of users in order π that can be decoded prior to u. This set is either the null set $\emptyset$, or $\{1\}$ when $u = 2$ and $\pi = (1, 2)$, and either $\emptyset$ ir $\{2\}$ when $u = 1$ and $\pi = (2, 1)$. The following table considers the order (1,2) for both users:

$\mathcal{E}_1$	$\mathcal{E}_2$	$X_{(1,2)}(1)$	$X_{(1,2)}(2)$	b_1	b_2
1	1	$\emptyset$	$\emptyset$	.9157	.3973
1	1	$\emptyset$	1	.6196	5.9071
1	0	$\emptyset$	$\emptyset$	6.3220	0
0	1	$\emptyset$	$\emptyset$	0	5.9071
1	.5	$\emptyset$	$\emptyset$	1.3063	.2257
1	.5	$\emptyset$	1	.9478	5.4073
.5	1	$\emptyset$	$\emptyset$	.5944	.6519
.5	1	$\emptyset$	1	.3744	5.9071

The following table considers the order (2,1) for both users:

$\mathcal{E}_1$	$\mathcal{E}_2$	$X_{(2,1)}(1)$	$X_{(2,1)}(2)$	b_1	b_2
1	1	$\emptyset$	$\emptyset$	.9157	.3973
1	1	2	$\emptyset$	6.322	.2378
1	0	$\emptyset$	$\emptyset$	6.3220	0
0	1	$\emptyset$	$\emptyset$	0	5.9071
1	.5	$\emptyset$	$\emptyset$	1.3063	.2257
1	.5	2	$\emptyset$	6.3220	.1287
.5	1	$\emptyset$	$\emptyset$	.5944	.6519
.5	1	2	$\emptyset$	5.8222	.4163

One could investigate the situations of different orders on the two different receivers but would find easily that the points generated describe rectangles interior to the ones in the above two tables.

Figure 12.23 sketches the corresponding rate region. Rectangles trace the region, but not so simply as in the BC or MAC cases in that points with less energy than the maximum on one of the two users can correspond to boundary points in the interference channel. The enumeration of all energies for more than 2 users and a simple channel could be complex.

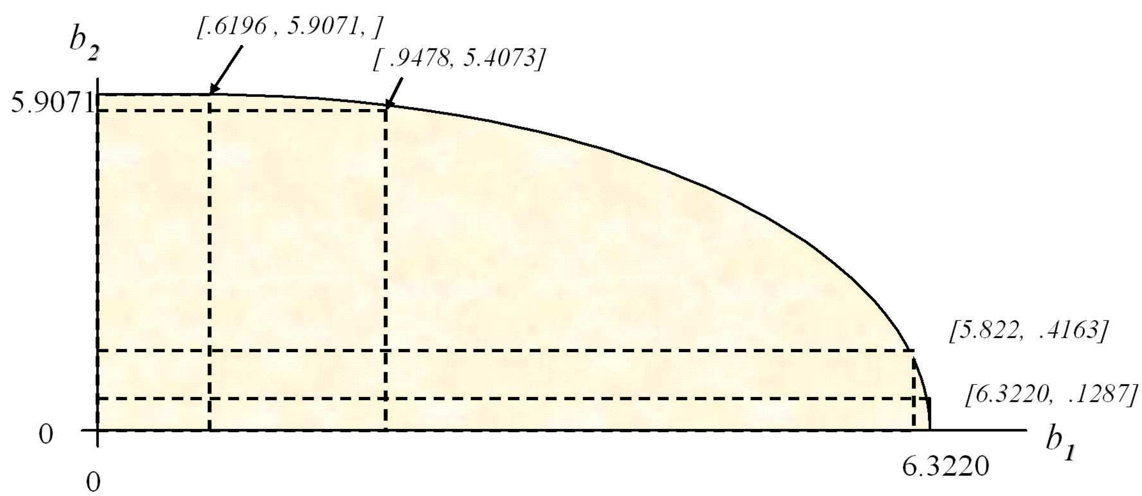

Figure 12.23: Interference-channel example.

Exercises - Chapter 12

12.1 *AWGN's*
Derive for the complex additive Gaussian noise channel $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n}$ that

$$I(\boldsymbol{x};\boldsymbol{y}) = \log_2 \frac{|R_{\boldsymbol{yy}}|}{|R_{\boldsymbol{nn}}|} = \log_2 \frac{|R_{\boldsymbol{xx}}|}{|R_{\boldsymbol{ee}}|} \quad . \tag{12.125}$$

Find equivalent expressions when the channel is real baseband.
(Hint: ok to use Chapter 5 results.)

12.2 *multiple-user channel types - row vector*
A multiple-user channel with additive Gaussian noise of power spectral density $\sigma^2 = .001$ on all outputs has channel matrix $H = [1\ 2]$. The input user-energy constraints are $\mathcal{E}_{x,1} = 1$ and $\mathcal{E}_{x,2} = 1$.

a. What type of multi-user channel is this?

b. Find $I(\boldsymbol{x};\boldsymbol{y})$.

c. Can the rate in part b be achieved? If so, describe how you would achieve it.

12.3 *multiple-user channel types - column vector*
A multiple-user channel with additive Gaussian noise of power spectral density $\sigma^2 = .001$ on all outputs has channel matrix $H = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. The noises at each output are independent. The input user-energy constraints are $\mathcal{E}_{x,1} = 1$ and $\mathcal{E}_{x,2} = 1$.

a. What type of multi-user channel is this?

b. Find $I(\boldsymbol{x};\boldsymbol{y})$ (Linear modulation may be assumed.)

c. Can the rate in part b be achieved? If so, describe how you would achieve it.

12.4 *2 way or not 2 way? - 10 pts*
Figure 12.24 shows a wireless cellphone system with 2 users. The "downlink" channel allows communication from one base station to two mobile customers who may share the wireless frequency band's dimensions. The "uplink" channel is the path shared from these two users to the base station in a different frequency band than the downlink band.

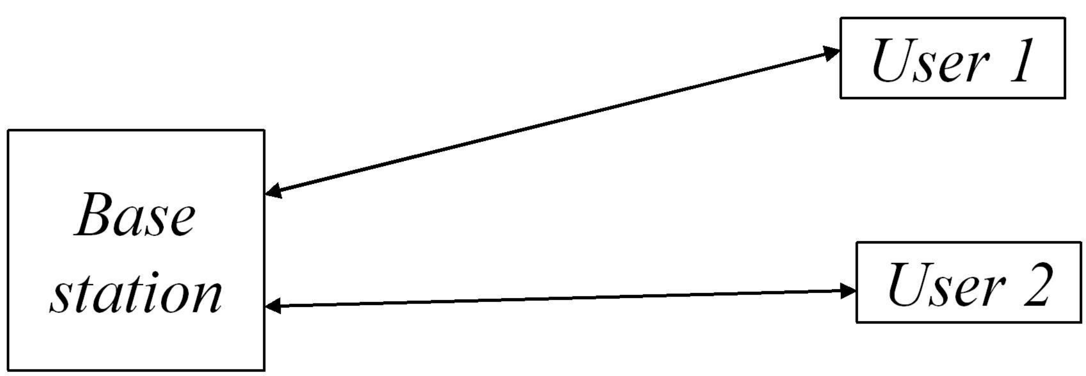

Figure 12.24: One way or two ways.

a. What type of channel is the downlink transmission system and at least how many users? (2 pts)

b. What type of channel is the uplink transmission system and at least how many users? (2 pts)

c. Suppose the downlink and uplink transmissions can share the same dimensions (frequency or time). At least how many users are there? (2 pts)

d. Using only IC, BC, and MAC as components describe the channel of part c as recursive representation of the 3 basic multi-channel types assuming that all input signals are Gaussian. (4 pts)

12.5 *multiple-user channel types - matrix*

A multiple-user channel with additive Gaussian noise of power spectral density $\sigma^2 = .001$ on all outputs has channel matrix $H = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. The input user-energy constraints are $\mathcal{E}_{x,1} = 1$ and $\mathcal{E}_{x,2} = 1$.

a. What type of multi-user channel is this if the transmitters and receivers cannot cooperate?

b. What other types of multi-user channels are possible for this H? Describe them.

c. Find $I(\boldsymbol{x}; \boldsymbol{y})$.

d. Which multi-user systems can always achieve the data rate in part c?

12.6 *3 × 3 MAC*

For an IC with 3 users:

a. How many user decoding orders are possible?

b. Enumerate the orders for part a.

c. For any given order and $p_{\boldsymbol{x}}$, what is the shape of the achievable region?

d. For the convex combinations of only any 2 orders, what are the types of shapes of achievable regions?

e. For the convex combinations of only any 3 orders, what are the types of shapes of achievable regions? (extra credit).

Hint: For parts d and e, you may pick 2 or 3 arbitrary rate-tuples corresponding to different decoding orders and plot achievable rate regions for all convex combinations of them.

12.7 *Detectors*

For the additive Gaussian channel $y = x_1 + x_2 + n$ with x_1 as uncoded 4 QAM with $d_{\min} = 4$ and x_2 as uncoded 16 QAM with $d_{\min} = 1$:

a. Show decision regions for joint detection of user 1 and user 2.

b. Show decision regions if user 1 is given and your corresponding receiver.

c. Show decision regions if user 2 is given and your corresponding receiver.

d. If user 2 is more important in terms of reliable decisions, should it go first or last in decoding order of successive decoding?

e. If user 1 is more important in terms of reliable decisions, should it go first or last in decoding order of successive decoding?

12.8 *Simple MAC*

An scalar multiple-access AWGN channel has two scalar inputs with unit eneries, gains .7 and 1.5 with noise variance .0001.

a. Find the maximum sum rate. (2 pts)

b. Find the maximum rates of the two users if the other user is treated as noise. (2 pts)

c. Find the capacity rate region for this channel. (2 pts)

d. Find any corner-point GDFE receiver. (1 pt)

e. Design a GDFE receiver and transmitter for the situation where the the two users have equal data rates. (3 pts)

f. Find a simultaneous water-filling spectrum solution for the two users that has maximum sum rate. (3 pts)

12.9 *Nonsingular GDFE.*

A MAC has channel matrix

$$H = \begin{bmatrix} 1 & .9 & 0 \\ 0 & 1 & .9 \\ a & 0 & 1 \end{bmatrix} \tag{12.126}$$

with additive Gaussian noise of autocorrelation matrix $R_{\boldsymbol{nn}} = 0.1 \cdot I$.

For both $a = 0$ and $a = .9$, answer each of the following questions.

a. (2 pt) How many users are there?

b. (4 pts) Find the highest rate sum for the users.

c. (4 pts) What can you say about the 6 possible orderings in terms of combinations that lead to the maximum sum rate?

d. (6 pts) Design a GDFE receiver for this channel showing G, and then unbiased feed-forward matrix processing and unbiased feedback sections.

e. (4 pts) Suppose input coordination were allowed for a single-user vector channel - what would the maximum data rate be?

12.10 *Singular GDFE.*

A MAC has channel matrix

$$H = [1 \; .9 \; .81] \tag{12.127}$$

with additive Gaussian noise of autocorrelation matrix $R_{\boldsymbol{nn}} = 0.1 \cdot I$.

a. (2 pt) How many users are there?

b. (4 pts) Find the highest rate sum for the users.

c. (4 pts) Design a GDFE receiver for this channel showing G, and then unbiased feed-forward matrix processing and unbiased feedback sections.

d. (5 pts) Find the possible rate-triples for each of the other 5 possible orderings.

12.11 *Infinite-finite mixed GDFE.*

A MAC has channel has $N = 1$ with ISI on each of two users who have pulse responses in discrete time $H_1(D) = 1 + .9D^{-1}$ and $H_2(D) = \frac{1}{\sqrt{10}} \cdot (1 + .8D)$, The AWGN noise variance is 0.017.

a. (4 pts) Find the highest rate sum for the users.

b. (4 pts) Design a GDFE receiver using two infinite-length MMSE-DFEs and also showing successive elimination of user 2 first in order from user 1.

c. (2 pts) What happens to the rate sum if the order is reversed.

12.12 *Simultaneous Water-Fill*

For the multiple access channel with $P_1 = (1 + D)$ and $P_2 = 1 - D$ with AWGN variance $\sigma^2 = 1$ and $\mathcal{E}_1 = \mathcal{E}_2 = 1$ ($T = 1$):

a. Find the maximum sum rate. (3 pts)

b. Sketch the capacity rate region for this channel. (3 pts)

c. Find an FDM simulataneously water-filling solution for this channel that achieves maximum sum rate. (Hint: you probably found this in part a - 1 pt.)

d. Design a single GDFE receiver and two transmitters for the situation in part c where the two users have equal data rates. (3 pts)

12.13 *Rate region for a 2-user multiple access example*

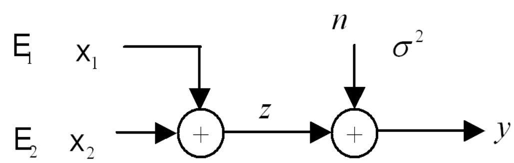

Figure 12.25: Figure for Problem 12.13

This problem considers 2 users, with energies $\mathcal{E}_1$ and $\mathcal{E}_2$ respectively as shown in Figure 12.25. The additive noise, $n \sim \mathcal{N}(0, \sigma^2)$. Let $C(\frac{\mathcal{E}}{N}) = \frac{1}{2}\log\left(1+\frac{\mathcal{E}}{N}\right)$ denote the capacity of a single user Gaussian channel with signal to noise ratio $\frac{\mathcal{E}}{\sigma^2}$. Let $T = 1$.

a. (1 pt) The rates for user 1 (R_1) and user 2 (R_2) are bounded by C_1 and C_2 respectively. Find the expressions for C_1 and C_2.

b. (2 pts) Similarly, $R_1 + R_2$ is bounded by C_3. We proceed to find C_3.

$$R_1 + R_2 \leq I(x_1, x_2; y) \tag{12.128}$$
$$= I(x_1 + x_2; y) \tag{12.129}$$
$$\leq \frac{1}{2}\log\left(1+\frac{\mathcal{E}_1+\mathcal{E}_2}{\sigma^2}\right) \tag{12.130}$$

(i) Explain why (2) is true.
(ii) Explain why (3) is true.

c. (2 pts) Using the bounds found in (a) and (b), sketch the capacity region for the 2-user example on an R_1-R_2 plane.

d. (5 pts)
(i) By letting user 1 achieve C_1, what's the maximum rate that user 2 can achieve? Describe a decoding scheme for this transmission. Mark this (R_1, R_2) point on the rate region.
(ii) By letting user 2 achieve C_2, what's the maximum rate that user 1 can achieve? Mark this (R_1, R_2) point on the rate region.
(iii) How can we achieve any point on the rate-sum curve given by $R_1 + R_2 = C(\frac{\mathcal{E}_1+\mathcal{E}_2}{\sigma^2})$? Hint: think GDFEs.

12.14 *Jagannathan's Non-Zero Gaps (Best achievable rate region of a scalar MAC with non-zero gap)*
Consider a scalar 2-user multiple access channel (with gains h_1 and h_2 and AWGN σ^2). Let the gap of the codes for the two users be $\Gamma > 1$ (i.e. > 0 dB).

a. Find the rate pairs that are achieved using successive decoding. Assume that both users completely use their energies $\mathcal{E}_1$ and $\mathcal{E}_2$.

b. (Bonus) Prove the claim that when either or both users do not use their energies completely, the resulting rate pairs achieved by successive decoding lie within the pentagon defined by the rate pairs found in a.

c. Find the rate pair that maximizes the sum rate for the MAC when a FDM scheme is used. (Hint: Think of the best FDM scheme when the gap is 0 dB and the energies of the two users are scaled by $1/\Gamma$).

d. Using the results of parts a, b and c, determine the scheme that achieves the maximum sum rate. Comment about the optimality of successive decoding for this channel.

e. The boundary of the best achievable rate region for this channel can be shown to be characterized by the convex combination of the successive decoding rate pairs and the rate pairs achieved by FDM. You need not prove this result for this problem. Assuming this result, compute and plot the best achievable rate region for the scalar 2-user MAC when $h_1 = 1$, $h_2 = 2$, $\mathcal{E}_1 = \mathcal{E}_2 = 1$, $\sigma^2 = 1$ and $\Gamma = 4$ dB. Compare this with the capacity region of the the scalar MAC when the gap is 0 dB but the energies of the 2 users are scaled by $1/\Gamma$ (in linear units).

12.15 *FDMA scheme for the 2-user Multiple Access example*

The multiple access scheme used in Problem 12.13(d) is called Code Division Multiple Access (CDMA). In Frequency Division Multiple Access (FDMA) scheme, disjoint frequency bands are allocated to different users so that they don't interfere with each other. This simplifies the decoding as compared to the decoding for CDMA. We will now find the rate region for the FDMA scheme and compare it to the one for CDMA.
Consider again the 2-user example shown in Fig. 1 where $\mathcal{E}_1, \mathcal{E}_2$ and σ^2 are as defined in Problem 2. Let α be the fraction of bandwidth assigned to user 1.

a. (3 pts) Write down the rate expressions R_1, R_2 for user 1 and 2.

b. (2 pts) Sketch the rate region (R_1,R_2) for $0 \leq \alpha \leq 1$ and compare it to the rate region for CDMA. You can use Matlab.

c. (2 pts) Notice that the FDMA curve lies *inside* the rate region for CDMA, and only touches the rate-sum curve at one point. Again, the rate-sum curve is given by $R_1 + R_2 = C(\frac{\mathcal{E}_1+\mathcal{E}_2}{\sigma^2})$.We will now find the value of α that achieves the rate-sum point for the FDMA scheme.
Let $C(\alpha) = R_1 + R_2$ where R_1, R_2 are found in (a). Show that $\alpha^\star = \frac{\mathcal{E}_1}{\mathcal{E}_1+\mathcal{E}_2}$ achieves the maximum and find $C(\alpha^\star)$.

12.16 *FDMA scheme for unequal 2-user multiple access channels with ISI*

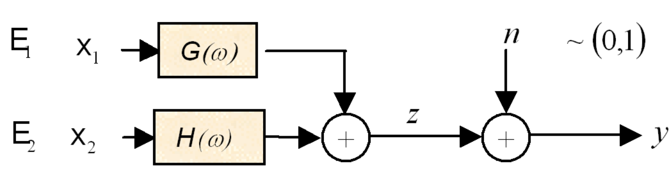

Figure 12.26: Figure for Problem 12.16

Figure 12.26 shows 2 users with energies $\mathcal{E}_1$ and $\mathcal{E}_2$ respectively and a noise power spectral density of 1.

a. (2 pts) Let $G(\omega)$ and $H(\omega)$ be two bandpass filters with *non-overlapping* passbands. Describe how the optimal input PSD's for both users can be determined.

b. (11 pts total) Now, let $G(\omega) = 1/\sqrt{2}$ and $H(\omega) = 1/\sqrt{\omega+1}$. Also, normalize the total bandwidth to 1, i.e., $0 \le \omega \le 1$. In this case, the channel frequency responses overlap with one another. The optimal input PSD's for both users for an FDMA scheme can be found using the *scaling* approach described in Section 12.2.3. Specifically, such scaling arbitrarily fixes the water-filling level to be 1 for both users, and finds the scaling factors of b_1 and b_2 such that the individual power constraints are met. Note that the individual power constraints are now $b_1 \cdot \mathcal{E}_1$ and $b_2 \cdot \mathcal{E}_2$ with the scaling as in Figure 12.27. Assume $T = 1$.

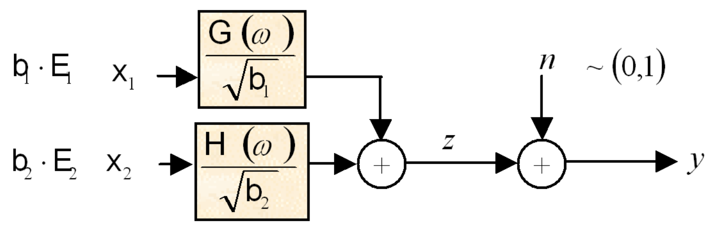

Figure 12.27: 2-user scaled multiaccess channel with ISI.

(i) (2 pts) Sketch the water-filling diagram that describes how the optimal PSD's are split for the FDMA scheme for the 2-user scaled multiaccess channel with ISI. Annotate your diagram with the scaling factors b_1 and b_2.

(ii) (1 pt) Let $w^\star$ denote the frequency where the scaled, inverted SNR curves of both users intersect. Find $w^\star$ in terms of b_1 and b_2.

(iii) (1 pt) Write down the power constraint for user 1 in terms of b_1, b_2 and P_1.

(iv) (1 pt) Write down the power constraint for user 2 in terms of b_1, b_2 and P_2.

(v) (2 pts) Letting $P_1 = P_2 = 1$, solve for b_1, b_2 and $w^\star$ (*Hint:* use *conv* and *roots* in Matlab).

(vi) (2 pts) What are the respective rates achieved by user 1 and user 2?
(*Hint:* $\int \ln x = x \ln x - x$).

(vii) (2 pts) Sketch the water-filling diagram *without* the scaling factors. (The water-filling levels for both users will be different).

12.17 *Simple broadcast rate region*

The scalar broadcast AWGN channel of Figure 12.19 has two scalar outputs with unit energy input. Both noise variances are equal to $\frac{\mathcal{N}_0}{2} = .0001$.

a. Sketch the rate region for gains $p_1 = 2$ and $p_2 = .4$. (2 pts)

b. Sketch the rate region for gains $p_1 = .4$ and $p_2 = 2$. (1 pt) (2 pts)

c. Sketh the rate region for gains $p_1 = 1$ and $p_2 = 1$. (2 pts)

d. Estimate the maximum rate sum for each of parts a, b, and c (3 pts).

12.18 *Yu's Region*

This question investigates various aspects of the capacity rate region for the broadcast channel.

a. Describe the capacity rate region with equations for the 3-user AWGN broadcast channel with gains $p_1 = 1$, $p_2 = .5$ and $p_3 = .25$. (3 pts)

b. For the general AWGN broadcast channel, relate the mutual information $I([\boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\boldsymbol{x}_u]\,; \boldsymbol{y})$ and the two autocorrelation matrices $R_{\boldsymbol{y}/[\boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\boldsymbol{x}_u]}$ and $R^{\perp}_{\boldsymbol{y}/[\boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\boldsymbol{x}_u]}$ where $R_{\boldsymbol{y}/[\boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\boldsymbol{x}_u]} + R^{\perp}_{\boldsymbol{y}/[\boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\boldsymbol{x}_u]} = R_{\boldsymbol{yy}}$. (2 pts)

c. Use your expressions in part c to write a bound on the maximum rate sum. (2 pts)

d. Describe the ordering that produces the largest bound achievable rate region? (2 pts)

e. Is the rate sum bound in part c achievable? Compare this rate sum bound to the rate sum bound of the multiple access channel – how do they differ?(3 pts)

12.19 *Simple IC rate region*

The scalar IC AWGN channel of Figure 12.22 has two scalar inputs with unit energy input. The gain from user 1 to its output is changed to 1. The gain from user 2 to its output is changed to 2. The gain from user 1 to user 2 is now labelled $a < 1$, a variable. Similarly, the gain from user 2 to user 1 is $b < a$.

a. Sketch the rate region for gains $a = .4$ and $b = .1$. (6 pts)

b. Describe what happens to the rate region as b approaches $a = .4$ in magnitude. (2 pts)

c. What happens to the rate region when $a >> b$? (2 pts)

12.20 *Kim's MAC - 10 pts*

Consider the following 1×2 single-user channel with a total transmitting energy constraint of $\mathcal{E}_x = \mathcal{E}(|x_1|^2 + |x_2|^2) \leq 2$.

$$y = [2 \quad i] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + n, \qquad (12.131)$$

where $n \sim \mathcal{CN}(0, 1)$ is an additive complex Gaussian noise at the receiver.

a. Calculate the capacity of this channel in bits per one complex dimensional symbol. (2 pts)

b. Suppose that there is an additional per-antenna energy constraint, i.e., each antenna should use energy of $\mathcal{E}_{u,n} \leq 1$. Calculate the capacity under these energy constraints. (3 pts)

c. Suppose this system is instead a two-user-multiple-access-channel (MAC), where each user occupies each antenna and sends its own independent signal. Specifically, no coordination is allowed between the transmitted signals. If each user has individual transmit energy constraint of 1, calculate the maximum rate sum for this MAC channel. (3 pts)

d. Discuss your results. (2 pts)

12.21 *Worst-Case Noise for Broadcast Channels - Mohseni1 - 14 pts*

Consider the following scalar broadcast channel with $U = 2$ users,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \qquad (12.132)$$

where $z_1, z_2 \sim \mathcal{N}(0, 1)$ are additive Gaussian noises at each receiver. Let $P = 1$ be the average transmit power constraint, $\mathcal{E}(x^2) \leq P$.

a. Plot the capacity region for this broadcast channel. By examining different pre-coding orders show that the capacity region is obtained when the better user (user 2) is pre-coded second. Find the maximum rate-sum point in the capacity region and prove this point is achieved by only transmitting to the best user. (4 pts)

b. Since for broadcast channels, coordination is not possible at receivers' side, the capacity region only depends on the marginal probability densities, $p(y_1|x)$ and $p(y_2|x)$. Thus the capacity region does not depend on the correlation coefficient among receiver noises, $\mathcal{E}(z_1 z_2) = \alpha$. Hence for any $-1 \le \alpha \le 1$, $I(x; y_1, y_2)$ is an upper bound on the maximum sum-rate or equivalently

$$R_1 + R_2 \le \min_{|\alpha| \le 1} I(x; y_1, y_2). \tag{12.133}$$

Find $\alpha \in [-1, 1]$ that minimizes

$$I(x; y_1, y_2) = \frac{1}{2} \log \frac{\left| \begin{bmatrix} 1 \\ 2 \end{bmatrix} P \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \right|}{\left| \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \right|}. \tag{12.134}$$

show that for this value of α, the matrix,

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}^{-1} - \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} P \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \right)^{-1} \tag{12.135}$$

is a diagonal matrix. This minimizing noise covariance matrix corresponds to the *Worst Case noise*. (4 pts)

c. Show that the minimum value of $I(x; y_1, y_2)$ obtained in part (b) is actually equal to maximum $R_1 + R_2$ from part (a). (2 pts)

d. Now suppose the receivers could some how cooperate with each other and also assume the noise covariance matrix is equal to the one obtained in part (b). Show that the feed forward whitening matched filter of a GDFE receiver only operates on received symbol of user 2. In other words, $I(x; y_1, y_2)$ can be achieved by a GDFE receiver without any required coordination and only user 2 communicates with the transmitter, the fact that you have already shown in part (c). (4 pts)

In general, it can be shown that this filter is diagonal and can be employed even without coordination at receivers. Also the successive decoding at the receiver can be transferred to transmitter side by performing *Dirty Paper Pre-coding* at the transmitter. As a result, the upper bound of part (b) on maximum sum-rate is achievable, and is known as the sum-capacity of Gaussian broadcast channels.

(For those of you who have not had a chance to take EE379C, GDFE receiver is an optimal receiver that could achieve the capacity of a single user point-to-point channel. This receiver consists of a feed forward filter that whitens the receiver noise and converts the channel into an upper triangular one and a successive decoder that decodes each symbol after subtracting the effects of previously decoded symbols from current symbol. Since the channel matrix is converted into an upper triangular one, the first symbol to be decoded does not suffer from interference caused by other symbols and by subtracting this symbol after decoding, its effect is removed from decoding of successive ones. For this receiver, feed forward filter is given by $L^{-1} H^T S_z^{-1}$, where S_z is the noise covariance matrix, H is the channel matrix including all transmit covariance shaping matrices and L is a lower triangular matrix such that $LL^T = I + H^T S_z^{-1} H$ and can be found by cholesky factorization.)

12.22 *Very Strong and Strong Interference Channels (Mohseni 2): - 10 pts*
Consider the following scalar interference channel with $U = 2$ users,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & a_1 \\ a_2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \tag{12.136}$$

where $z_1, z_2 \sim \mathcal{N}(0, 1)$ are additive Gaussian noises at each receiver. Let P_1 and P_2 be the average transmit power constraints for user 1 and user 2 respectively, $\mathcal{E}(x_1^2) \le P_1$ and $\mathcal{E}(x_2^2) \le P_2$.

a. Suppose $a_1^2 \geq 1 + P_1$ and $a_2^2 \geq 1 + P_2$. This scenario is usually referred to as *Very Strong Interference Channel*. Show that the capacity region of a very strong interference channel is the same as the capacity of an interference channel with $a_1 = a_2 = 0$. Please provide justifications for both achievability and converse parts.

b. Now suppose $a_1 \geq 1$ and $a_2 \geq 1$. This interference channel is known as *Strong Interference Channel*. Show that any rate pair $(b_1, b_2) \in \mathcal{R}_+^2$ satisfying the following properties is achievable,

$$\begin{aligned} b_1 &\leq \frac{1}{2}\log(1 + P_1) \\ b_2 &\leq \frac{1}{2}\log(1 + P_2) \\ b_1 + b_2 &\leq \min\{\frac{1}{2}\log(1 + P_1 + a_1^2 P_2), \frac{1}{2}\log(1 + P_2 + a_2^2 P_1)\}. \end{aligned}$$

It can be shown that this region is in fact the capacity region of a very strong interference channel.