Topics in Multi-User Information Theory

Topics in Multi-User Information Theory

Gerhard Kramer

Bell Laboratories, Alcatel-Lucent
Murray Hill
New Jersey, 07974
USA
gkr@bell-labs.com

the essence of knowledge

Boston – Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is G. Kramer, Topics in Multi-User Information Theory, Foundations and Trends® in Communications and Information Theory, vol 4, nos 4–5, pp 265–444, 2007

ISBN: 978-1-60198-148-6

Foundations and Trends® in Communications and Information Theory
Volume 4 Issue 4–5, 2007

Editorial Board

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2007, Volume 4, 6 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Foundations and Trends® in
Communications and Information Theory
Vol. 4, Nos. 4–5 (2007) 265–444

DOI: 10.1561/0100000028

the essence of knowledge

Topics in Multi-User Information Theory

Gerhard Kramer

Bell Laboratories, Alcatel-Lucent, 600 Mountain Avenue, Murray Hill, New Jersey, 07974, USA, gkr@bell-labs.com

Abstract

This survey reviews fundamental concepts of multi-user information theory. Starting with typical sequences, the survey builds up knowledge on random coding, binning, superposition coding, and capacity converses by introducing progressively more sophisticated tools for a selection of source and channel models. The problems addressed include: Source Coding; Rate-Distortion and Multiple Descriptions; Capacity-Cost; The Slepian–Wolf Problem; The Wyner-Ziv Problem; The Gelfand-Pinsker Problem; The Broadcast Channel; The Multiaccess Channel; The Relay Channel; The Multiple Relay Channel; and The Multiaccess Channel with Generalized Feedback. The survey also includes a review of basic probability and information theory.

Contents

Notations and Acronyms

We use standard notation for probabilities, random variables, entropy, mutual information, and so forth. Table 1 lists notation developed in the appendices of this survey, and we use this without further explanation in the main body of the survey. We introduce the remaining notation as we go along. The reader is referred to the appendices for a review of the relevant probability and information theory concepts.

Table 1 Probability and information theory notation.

Sequences, Vectors, Matrices	
x^n	the finite sequence $x_1, x_2, \ldots, x_n$
$x^n y^m$	sequence concatenation: $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m$
$\underline{x}$	the vector $[x_1, x_2, \ldots, x_n]$
$\mathbf{H}$	a matrix
$\|\mathbf{Q}\|$	determinant of the matrix $\mathbf{Q}$

(*Continued*)

Table 1 (*Continued*)

Probability	
$\Pr[\mathcal{A}]$	probability of the event $\mathcal{A}$
$\Pr[\mathcal{A}\vert\mathcal{B}]$	probability of event $\mathcal{A}$ conditioned on event $\mathcal{B}$
$P_X(\cdot)$	probability distribution of the random variable X
$P_{X\vert Y}(\cdot)$	probability distribution of X conditioned on Y
$\text{supp}(P_X)$	support of P_X
$p_X(\cdot)$	probability density of the random variable X
$p_{X\vert Y}(\cdot)$	probability density of X conditioned on Y
$E[X]$	expectation of the real-valued X
$E[X\vert\mathcal{A}]$	expectation of X conditioned on event $\mathcal{A}$
$\mathsf{Var}[X]$	variance of X
$\mathbf{Q}_{\underline{X}}$	covariance matrix of $\underline{X}$
Information Theory	
$H(X)$	entropy of the discrete random variable X
$H(X\vert Y)$	entropy of X conditioned on Y
$I(X;Y)$	mutual information between X and Y
$I(X;Y\vert Z)$	mutual information between X and Y conditioned on Z
$D(P_X\Vert P_Y)$	informational divergence between P_X and P_Y
$h(X)$	differential entropy of X
$h(X\vert Y)$	differential entropy of X conditioned on Y
$H_2(\cdot)$	binary entropy function

1

Typical Sequences and Source Coding

1.1 Typical Sequences

Shannon introduced the notion of a "typical sequence" in his 1948 paper "A Mathematical Theory of Communication" [55]. To illustrate the idea, consider a discrete memoryless source (DMS), which is a device that emits symbols from a discrete and finite alphabet $\mathcal{X}$ in an independent and identically distributed (i.i.d.) manner (see Figure 1.1). Suppose the source probability distribution is $P_X(\cdot)$ where

$$P_X(0) = 2/3 \quad \text{and} \quad P_X(1) = 1/3. \tag{1.1}$$

Consider the following experiment: we generated a sequence of length 18 by using a random number generator with the distribution (1.1). We write this sequence below along with three other sequences that we generated artificially.

(a) 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
(b) 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0
(c) 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0
(d) 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1.

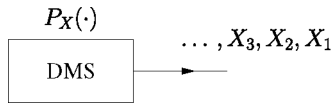

Fig. 1.1 A discrete memoryless source with distribution $P_X(\cdot)$.

If we compute the probabilities that these sequences were emitted by the source (1.1), we have

(a) $(2/3)^{18} \cdot (1/3)^0 \approx 6.77 \cdot 10^{-4}$
(b) $(2/3)^9 \cdot (1/3)^9 \approx 1.32 \cdot 10^{-6}$
(c) $(2/3)^{11} \cdot (1/3)^7 \approx 5.29 \cdot 10^{-6}$
(d) $(2/3)^0 \cdot (1/3)^{18} \approx 2.58 \cdot 10^{-9}$.

Thus, the first sequence is the most probable one by a large margin. However, the reader will likely *not* be surprised to find out that it is sequence (c) that was actually put out by the random number generator. Why is this intuition correct? To explain this, we must define more precisely what one might mean by a "typical" sequence.

1.2 Entropy-Typical Sequences

Let x^n be a finite sequence whose ith entry x_i takes on values in $\mathcal{X}$. We write $\mathcal{X}^n$ for the Cartesian product of the set $\mathcal{X}$ with itself n times, i.e., we have $x^n \in \mathcal{X}^n$. Let $N(a|x^n)$ be the number of positions of x^n having the letter a, where $a \in \mathcal{X}$.

There are several natural definitions for typical sequences. Shannon in [55, Sec. 7] chose a definition based on the entropy of a random variable X. Suppose that X^n is a sequence put out by the DMS $P_X(\cdot)$, which means that $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$ is the probability that x^n was put out by the DMS $P_X(\cdot)$. More generally, we will use the notation

$$P_X^n(x^n) = \prod_{i=1}^n P_X(x_i). \tag{1.2}$$

We further have

$$P_X^n(x^n) = \begin{cases} \prod_{a \in \mathrm{supp}(P_X)} P_X(a)^{N(a|x^n)} & \text{if } N(a|x^n) = 0 \text{ whenever } P_X(a) = 0 \\ 0 & \text{else} \end{cases} \tag{1.3}$$

and intuitively one might expect that the letter a occurs about $N(a|x^n) \approx nP_X(a)$ times, so that $P_X^n(x^n) \approx \Pi_{a\in\text{supp}(P_X)} P_X(a)^{nP_X(a)}$ or

$$-\frac{1}{n}\log_2 P_X^n(x^n) \approx \sum_{a\in\text{supp}(P_X)} -P_X(a)\log_2 P_X(a).$$

Shannon therefore defined a sequence x^n to be typical with respect to ϵ and $P_X(\cdot)$ if

$$\left|\frac{-\log_2 P_X^n(x^n)}{n} - H(X)\right| < \epsilon \tag{1.4}$$

for some small positive ϵ. The sequences satisfying (1.4) are sometimes called *weakly* typical sequences or *entropy*-typical sequences [19, p. 40]. We can equivalently write (1.4) as

$$2^{-n[H(X)+\epsilon]} < P_X^n(x^n) < 2^{-n[H(X)-\epsilon]}. \tag{1.5}$$

Example 1.1. If $P_X(\cdot)$ is uniform then for any x^n we have

$$P_X^n(x^n) = |\mathcal{X}|^{-n} = 2^{-n\log_2|\mathcal{X}|} = 2^{-nH(X)} \tag{1.6}$$

and *all* sequences in $\mathcal{X}^n$ are entropy-typical.

Example 1.2. The source (1.1) has $H(X) \approx 0.9183$ and the above four sequences are entropy-typical with respect to $P_X(\cdot)$ if

(a) $\epsilon > 1/3$
(b) $\epsilon > 1/6$
(c) $\epsilon > 1/18$
(d) $\epsilon > 2/3$.

Note that sequence (*c*) requires the smallest ϵ.

We remark that *entropy* typicality applies to *continuous* random variables with a density if we replace the probability $P_X^n(x^n)$ in (1.4) with the density value $p_X^n(x^n)$. In contrast, the next definition can be used only for discrete random variables.

1.3 Letter-Typical Sequences

A perhaps more natural definition for *discrete* random variables than (1.4) is the following. For $\epsilon \geq 0$, we say a sequence x^n is ϵ-*letter* typical with respect to $P_X(\cdot)$ if

$$\left| \frac{1}{n} N(a|x^n) - P_X(a) \right| \leq \epsilon \cdot P_X(a) \quad \text{for all } a \in \mathcal{X} \tag{1.7}$$

The set of x^n satisfying (1.7) is called the ϵ-*letter-typical set* $T_\epsilon^n(P_X)$ with respect to $P_X(\cdot)$. The letter typical x^n are thus sequences whose *empirical* probability distribution is close to $P_X(\cdot)$.

Example 1.3. If $P_X(\cdot)$ is uniform then ϵ-letter typical x^n satisfy

$$\frac{(1-\epsilon)n}{|\mathcal{X}|} \leq N(a|x^n) \leq \frac{(1+\epsilon)n}{|\mathcal{X}|} \tag{1.8}$$

and if $\epsilon < |\mathcal{X}| - 1$, as is usually the case, then *not* all x^n are letter-typical. The definition (1.7) is then more restrictive than (1.4) (see Example 1.1).

We will generally rely on letter typicality, since for discrete random variables it is just as easy to use as entropy typicality, but can give stronger results.

We remark that one often finds small variations of the conditions (1.7). For example, for *strongly* typical sequences one replaces the $\epsilon P_X(a)$ on the right-hand side of (1.7) with ϵ or $\epsilon/|\mathcal{X}|$ (see [19, p. 33], and [18, pp. 288, 358]). One further often adds the condition that $N(a|x^n) = 0$ if $P_X(a) = 0$ so that typical sequences cannot have zero-probability letters. Observe, however, that this condition is included in (1.7). We also remark that the letter-typical sequences are simply called "typical sequences" in [44] and "robustly typical sequences" in [46]. In general, by the label "letter-typical" we mean any choice of typicality where one performs a per-alphabet-letter test on the empirical probabilities. We will focus on the definition (1.7).

We next develop the following theorem that describes some of the most important properties of letter-typical sequences and sets.

Let $\mu_X = \min_{x \in \mathrm{supp}(P_X)} P_X(x)$ and define

$$\delta_\epsilon(n) = 2|\mathcal{X}| \cdot e^{-n\epsilon^2 \mu_X}. \tag{1.9}$$

Observe that $\delta_\epsilon(n) \to 0$ for fixed ϵ, $\epsilon > 0$, and $n \to \infty$.

Theorem 1.1. Suppose $0 \le \epsilon \le \mu_X$, $x^n \in T_\epsilon^n(P_X)$, and X^n is emitted by a DMS $P_X(\cdot)$. We have

$$2^{-n(1+\epsilon)H(X)} \le P_X^n(x^n) \le 2^{-n(1-\epsilon)H(X)} \tag{1.10}$$

$$(1 - \delta_\epsilon(n))\, 2^{n(1-\epsilon)H(X)} \le |T_\epsilon^n(P_X)| \le 2^{n(1+\epsilon)H(X)} \tag{1.11}$$

$$1 - \delta_\epsilon(n) \le \Pr[X^n \in T_\epsilon^n(P_X)] \le 1. \tag{1.12}$$

Proof. Consider (1.10). For $x^n \in T_\epsilon^n(P_X)$, we have

$$\begin{aligned} P_X^n(x^n) &= \prod_{a \in \mathrm{supp}(P_X)} P_X(a)^{N(a|x^n)} \\ &\le \prod_{a \in \mathrm{supp}(P_X)} P_X(a)^{nP_X(a)(1-\epsilon)} \\ &= 2^{\sum_{a \in \mathrm{supp}(P_X)} nP_X(a)(1-\epsilon)\log_2 P_X(a)} \\ &= 2^{-n(1-\epsilon)H(X)}, \end{aligned} \tag{1.13}$$

where the inequality follows because, by the definition (1.7), typical x^n satisfy $N(a|x^n)/n \ge P_X(a)(1-\epsilon)$. One can similarly prove the left-hand side of (1.10).

Next, consider (1.12). In the appendix of this section, we prove the following result using the Chernoff bound:

$$\Pr\left[\left|\frac{N(a|X^n)}{n} - P_X(a)\right| > \epsilon P_X(a)\right] \le 2 \cdot e^{-n\epsilon^2 \mu_X}, \tag{1.14}$$

where $0 \le \epsilon \le \mu_X$. We thus have

$$\begin{aligned} \Pr[X^n \notin T_\epsilon^n(P_X)] &= \Pr\left[\bigcup_{a \in \mathcal{X}} \left\{\left|\frac{N(a|X^n)}{n} - P_X(a)\right| > \epsilon P_X(a)\right\}\right] \\ &\le \sum_{a \in \mathcal{X}} \Pr\left[\left|\frac{N(a|X^n)}{n} - P_X(a)\right| > \epsilon P_X(a)\right] \\ &\le 2|\mathcal{X}| \cdot e^{-n\epsilon^2 \mu_X}, \end{aligned} \tag{1.15}$$

where we have used the union bound (see (A.5)) for the second step. This proves the left-hand side of (1.12).

Finally, for (1.11) observe that

$$\begin{aligned}\Pr\left[X^n \in T_\epsilon^n(P_X)\right] &= \sum_{x^n \in T_\epsilon^n(P_X)} P_X^n(x^n) \\ &\le |T_\epsilon^n(P_X)|\, 2^{-n(1-\epsilon)H(X)}, \end{aligned} \tag{1.16}$$

where the inequality follows by (1.13). Using (1.15) and (1.16), we thus have

$$|T_\epsilon^n(P_X)| \ge (1 - \delta_\epsilon(n))\, 2^{n(1-\epsilon)H(X)}. \tag{1.17}$$

We similarly derive the right-hand side of (1.11). □

1.4 Source Coding

The source coding problem is depicted in Figure 1.2. A DMS $P_X(\cdot)$ emits a sequence x^n of symbols that are passed to an encoder. The source encoder "compresses" x^n into an index w and sends w to the decoder. The decoder reconstructs x^n from w as $\hat{x}^n(w)$, and is said to be successful if $\hat{x}^n(w) = x^n$.

The source encoding can be done in several ways:

- Fixed-length to fixed-length coding (or block-to-block coding).
- Fixed-length to variable-length coding (block-to-variable-length coding).
- Variable-length to fixed-length coding (variable-length-to-block coding).
- Variable-length to variable-length coding.

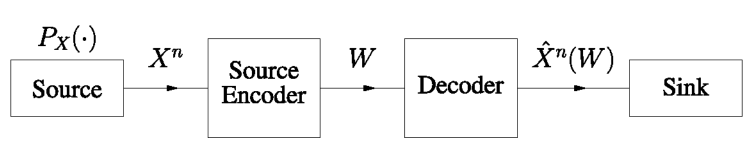

Fig. 1.2 The source coding problem.

We will here consider only the first two approaches. For a block-to-variable-length scheme, the number of bits transmitted by the encoder depends on x^n. We will consider the case where every source sequence is assigned a *unique* index w. Hence, one can reconstruct x^n perfectly. Let $L(x^n)$ be the number of bits transmitted for x^n. The goal is to minimize the *average* rate $R = E\left[L(X^N)\right]/n$.

For a block-to-block encoding scheme, the index w takes on one of 2^{nR} indexes w, $w = 1, 2, \ldots, 2^{nR}$, and we assume that 2^{nR} is a positive integer. The encoder sends exactly nR bits for every source sequence x^n, and the goal is to make R as small as possible. Observe that block-to-block encoding might require the encoder to send the *same* w for two *different* source sequences.

Suppose first that we permit no error in the reconstruction. We use the block-to-variable-length encoder, choose an n and an ϵ, and assign each sequence in $T_\epsilon^n(P_X)$ a unique positive integer w. According to (1.11), these indexes w can be represented by at most $n(1+\epsilon)H(X) + 1$ bits. Next, the encoder collects a sequence x^n. If $x^n \in T_\epsilon^n(P_X)$, then the encoder sends a "0" followed by the $n(1+\epsilon)H(X) + 1$ bits that represent this sequence. If $x^n \notin T_\epsilon^n(P_X)$, then the encoder sends a "1" followed by $n\log_2|\mathcal{X}| + 1$ bits that represent x^n. The average number of bits per source symbol is the compression rate R, and it is upper bounded by

$$\begin{aligned}
R &\le \Pr\left[X^n \in T_\epsilon^n(P_X)\right]\left[(1+\epsilon)H(X) + 2/n\right] \\
&\quad + \Pr\left[X^n \notin T_\epsilon^n(P_X)\right]\left(\log_2|\mathcal{X}| + 2/n\right) \\
&\le (1+\epsilon)H(X) + 2/n + \delta_\epsilon(n)(\log_2|\mathcal{X}| + 2/n). \qquad (1.18)
\end{aligned}$$

But since $\delta_\epsilon(n) \to 0$ as $n \to \infty$, we can transmit at any rate above $H(X)$ bits per source symbol. For example, if the DMS is binary with $P_X(0) = 1 - P_X(1) = 2/3$, then we can transmit the source outputs in a lossless fashion at any rate above $H(X) \approx 0.9183$ bits per source symbol.

Suppose next that we must use a block-to-block encoder, but that we permit a small error probability in the reconstruction. Based on the above discussion, we can transmit at any rate above $(1+\epsilon)H(X)$ bits per source symbol with an error probability $\delta_\epsilon(n)$. By making n large, we can make $\delta_\epsilon(n)$ as close to zero as desired.

But what about a converse result? Can one compress with a small error probability, or even zero error probability, at rates below $H(X)$? We will prove a converse for block-to-block encoders only, since the block-to-variable-length case requires somewhat more work.

Consider Fano's inequality (see Section A.10) which ensures us that

$$H_2(P_e) + P_e \log_2(|\mathcal{X}|^n - 1) \geq H(X^n|\hat{X}^n), \tag{1.19}$$

where $P_e = \Pr[\hat{X}^n \neq X^n]$. Recall that there are at most 2^{nR} different sequences $\hat{x}^n$, and that $\hat{x}^n$ is a function of x^n. We thus have

$$\begin{aligned} nR &\geq H(\hat{X}^n) \\ &= H(\hat{X}^n) - H(\hat{X}^n|X^n) \\ &= I(X^n;\hat{X}^n) \\ &= H(X^n) - H(X^n|\hat{X}^n) \\ &= nH(X) - H(X^n|\hat{X}^n) \\ &\geq n\left[H(X) - \frac{H_2(P_e)}{n} - P_e \log_2|\mathcal{X}|\right], \end{aligned} \tag{1.20}$$

where the last step follows by (1.19). Since we require that P_e be zero, or approach zero with n, we find that $R \geq H(X)$ for block-to-block encoders with arbitrarily small positive P_e. This is the desired converse.

1.5 Jointly and Conditionally Typical Sequences

Let $N(a,b|x^n,y^n)$ be the number of times the pair (a,b) occurs in the sequence of pairs $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$. The *jointly* typical set with respect to $P_{XY}(\cdot)$ is simply

$$\begin{aligned} T_\epsilon^n(P_{XY}) = \Big\{(x^n,y^n) : &\left|\frac{1}{n}N(a,b|x^n,y^n) - P_{XY}(a,b)\right| \\ &\leq \epsilon \cdot P_{XY}(a,b) \quad \text{for all } (a,b) \in \mathcal{X} \times \mathcal{Y}\Big\}. \end{aligned} \tag{1.21}$$

The reader can easily check that $(x^n,y^n) \in T_\epsilon^n(P_{XY})$ implies both $x^n \in T_\epsilon^n(P_X)$ and $y^n \in T_\epsilon^n(P_Y)$.

Consider the conditional distribution $P_{Y|X}(\cdot)$ and define

$$P_{Y|X}^n(y^n|x^n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \tag{1.22}$$

$$T_\epsilon^n(P_{XY}|x^n) = \{y^n : (x^n,y^n) \in T_\epsilon^n(P_{XY})\}. \tag{1.23}$$

Observe that $T_\epsilon^n(P_{XY}|x^n) = \emptyset$ if x^n is not in $T_\epsilon^n(P_X)$. We shall further need the following counterpart of $\delta_\epsilon(n)$ in (1.9):

$$\delta_{\epsilon_1,\epsilon_2}(n) = 2|\mathcal{X}||\mathcal{Y}| \exp\left(-n \cdot \frac{(\epsilon_2 - \epsilon_1)^2}{1+\epsilon_1} \cdot \mu_{XY}\right), \tag{1.24}$$

where $\mu_{XY} = \min_{(a,b)\in \mathrm{supp}(P_{XY})} P_{XY}(a,b)$ and $0 \le \epsilon_1 < \epsilon_2 \le 1$. Note that $\delta_{\epsilon_1,\epsilon_2}(n) \to 0$ as $n \to \infty$. In the Appendix, we prove the following theorem that generalizes Theorem 1.1 to include conditioning.

Theorem 1.2. Suppose $0 \le \epsilon_1 < \epsilon_2 \le \mu_{XY}$, $(x^n,y^n) \in T_{\epsilon_1}^n(P_{XY})$, and (X^n,Y^n) was emitted by the DMS $P_{XY}(\cdot)$. We have

$$2^{-nH(Y|X)(1+\epsilon_1)} \le P_{Y|X}^n(y^n|x^n) \le 2^{-nH(Y|X)(1-\epsilon_1)} \tag{1.25}$$

$$(1-\delta_{\epsilon_1,\epsilon_2}(n))\, 2^{nH(Y|X)(1-\epsilon_2)} \le |T_{\epsilon_2}^n(P_{XY}|x^n)| \le 2^{nH(Y|X)(1+\epsilon_2)} \tag{1.26}$$

$$1-\delta_{\epsilon_1,\epsilon_2}(n) \le \Pr\left[Y^n \in T_{\epsilon_2}^n(P_{XY}|x^n) \,\middle|\, X^n = x^n\right] \le 1. \tag{1.27}$$

The following result follows easily from Theorem 1.2 and will be extremely useful to us.

Theorem 1.3. Consider a joint distribution $P_{XY}(\cdot)$ and suppose $0 \le \epsilon_1 < \epsilon_2 \le \mu_{XY}$, Y^n is emitted by a DMS $P_Y(\cdot)$, and $x^n \in T_{\epsilon_1}^n(P_X)$. We have

$$\begin{aligned}(1-\delta_{\epsilon_1,\epsilon_2}(n))\ & 2^{-n[I(X;Y)+2\epsilon_2 H(Y)]} \\ & \le \Pr\left[Y^n \in T_{\epsilon_2}^n(P_{XY}|x^n)\right] \le 2^{-n[I(X;Y)-2\epsilon_2 H(Y)]}.\end{aligned} \tag{1.28}$$

Proof. The upper bound follows by (1.25) and (1.26):

$$\begin{aligned}\Pr\left[Y^n \in T^n_{\epsilon_2}(P_{XY}|x^n)\right] &= \sum_{y^n \in T_{\epsilon_2}(P_{XY}|x^n)} P^n_Y(y^n) \\ &\le 2^{nH(Y|X)(1+\epsilon_2)}\, 2^{-nH(Y)(1-\epsilon_2)} \\ &\le 2^{-n[I(X;Y)-2\epsilon_2 H(Y)]}. \end{aligned} \tag{1.29}$$

The lower bound also follows from (1.25) and (1.26). □

For small ϵ_1 and ϵ_2, large n, typical (x^n, y^n), and (X^n, Y^n) emitted by a DMS $P_{XY}(\cdot)$, we thus have

$$P^n_{Y|X}(y^n|x^n) \approx 2^{-nH(Y|X)} \tag{1.30}$$

$$|T^n_{\epsilon_2}(P_{XY}|x^n)| \approx 2^{nH(Y|X)} \tag{1.31}$$

$$\Pr\left[Y^n \in T^n_{\epsilon_2}(P_{XY}|x^n) \,\middle|\, X^n = x^n\right] \approx 1 \tag{1.32}$$

$$\Pr\left[Y^n \in T^n_{\epsilon_2}(P_{XY}|x^n)\right] \approx 2^{-nI(X;Y)}. \tag{1.33}$$

We remark that the probabilities in (1.27) and (1.28) (or (1.32) and (1.33)) differ only in whether or not one conditions on $X^n = x^n$.

Example 1.4. Suppose X and Y are independent, in which case the approximations (1.32) and (1.33) both give

$$\Pr\left[Y^n \in T^n_{\epsilon_2}(P_{XY}|x^n)\right] \approx 1. \tag{1.34}$$

Note, however, that the precise version (1.28) of (1.33) is trivial for large n. This example shows that one must exercise caution when working with the approximations (1.30)–(1.33).

Example 1.5. Suppose that $X = Y$ so that (1.33) gives

$$\Pr\left[Y^n \in T^n_{\epsilon_2}(P_{XY}|x^n)\right] \approx 2^{-nH(X)}. \tag{1.35}$$

This result should not be surprising because $|T^n_{\epsilon_2}(P_X)| \approx 2^{nH(X)}$ and we are computing the probability of the event $X^n = x^n$ for some $x^n \in T^n_{\epsilon_1}(P_{XY})$ (the fact that ϵ_2 is larger than ϵ_1 does not play a role for large n).

1.6 Appendix: Proofs

Proof of Inequality (1.14)

We prove the bound (1.14). Consider first $P_X(a) = 0$ for which we have

$$\Pr\left[\frac{N(a|X^n)}{n} > P_X(a)(1+\epsilon)\right] = 0. \tag{1.36}$$

Next, suppose that $P_X(a) > 0$. Using the Chernoff bound, we have

$$\begin{aligned}
\Pr\left[\frac{N(a|X^n)}{n} > P_X(a)(1+\epsilon)\right] &\le \Pr\left[\frac{N(a|X^n)}{n} \ge P_X(a)(1+\epsilon)\right] \\
&\le E\left[e^{\nu N(a|X^n)/n}\right] e^{-\nu P_X(a)(1+\epsilon)} \\
&= \left[\sum_{m=0}^{n} \Pr\left[N(a|X^n) = m\right] e^{\nu m/n}\right] e^{-\nu P_X(a)(1+\epsilon)} \\
&= \left[\sum_{m=0}^{n} \binom{n}{m} P_X(a)^m (1-P_X(a))^{n-m} e^{\nu m/n}\right] e^{-\nu P_X(a)(1+\epsilon)} \\
&= \left[(1-P_X(a)) + P_X(a) e^{\nu/n}\right]^n e^{-\nu P_X(a)(1+\epsilon)}. &(1.37)\\
& &(1.38)
\end{aligned}$$

Optimizing (1.38) with respect to ν, we find that

$$\begin{array}{ll} \nu = \infty & \text{if } P_X(a)(1+\epsilon) \ge 1 \\ e^{\nu/n} = \frac{(1-P_X(a))(1+\epsilon)}{1-P_X(a)(1+\epsilon)} & \text{if } P_X(a)(1+\epsilon) < 1. \end{array} \tag{1.39}$$

In fact, the Chernoff bound correctly identifies the probabilities to be 0 and $P_X(a)^n$ for the cases $P_X(a)(1+\epsilon) > 1$ and $P_X(a)(1+\epsilon) = 1$, respectively. More interestingly, for $P_X(a)(1+\epsilon) < 1$ we insert (1.39) into (1.38) and obtain

$$\Pr\left[\frac{N(a|X^n)}{n} \ge P_X(a)(1+\epsilon)\right] \le 2^{-nD(P_B\|P_A)}, \tag{1.40}$$

where A and B are binary random variables with

$$\begin{aligned} P_A(0) &= 1 - P_A(1) = P_X(a) \\ P_B(0) &= 1 - P_B(1) = P_X(a)(1+\epsilon). \end{aligned} \tag{1.41}$$

We can write $P_B(0) = P_A(0)(1+\epsilon)$ and hence

$$D(P_B \| P_A) = P_A(0)(1+\epsilon)\log_2(1+\epsilon) + [1 - P_A(0)(1+\epsilon)]\log_2\left(\frac{1-P_A(0)(1+\epsilon)}{1-P_A(0)}\right). \quad (1.42)$$

We wish to further simplify (1.42). The first two derivatives of (1.42) with respect to ϵ are

$$\frac{dD(P_B\|P_A)}{d\epsilon} = P_A(0)\log_2\left(\frac{(1-P_A(0))(1+\epsilon)}{(1-P_A(0))(1+\epsilon)}\right) \quad (1.43)$$

$$\frac{d^2D(P_B\|P_A)}{d\epsilon^2} = \frac{P_A(0)\log_2(e)}{(1+\epsilon)[1-P_A(0)(1+\epsilon)]}. \quad (1.44)$$

We find that (1.43) is zero for $\epsilon = 0$ and we can lower bound (1.44) by $P_X(a)\log_2(e)$ for $0 \le \epsilon \le \mu_X$. The second derivative of $D(P_B\|P_A)$ with respect to ϵ is thus larger than $P_X(a)\log_2(e)$ and so we have

$$D(P_B\|P_A) \ge \epsilon^2 \cdot P_A(0)\log_2(e) \quad (1.45)$$

for $0 \le \epsilon \le \mu_X$. Combining (1.40) and (1.45) we arrive at

$$\Pr\left[\frac{N(a|X^n)}{n} \ge P_X(a)(1+\epsilon)\right] \le e^{-n\epsilon^2 P_X(a)}. \quad (1.46)$$

One can similarly bound

$$\Pr\left[\frac{N(a|X^n)}{n} \le P_X(a)(1-\epsilon)\right] \le e^{-n\epsilon^2 P_X(a)}. \quad (1.47)$$

Note that (1.46) and (1.47) are valid for all $a \in \mathcal{X}$ including a with $P_X(a) = 0$. However, the event in (1.14) has a strict inequality so we can improve the above bounds for the case $P_X(a) = 0$ (see (1.36)). This observation lets us replace $P_X(a)$ in (1.46) and (1.47) with μ_X and the result is (1.14).

Proof of Theorem 1.2

Suppose that $(x^n, y^n) \in T^n_{\epsilon_1}(P_{XY})$. We prove (1.25) by bounding

$$\begin{aligned} P^n_{Y|X}(y^n|x^n) &= \prod_{(a,b)\in \text{supp}(P_{XY})} P_{Y|X}(b|a)^{N(a,b|x^n,y^n)} \\ &\le \prod_{(a,b)\in \text{supp}(P_{XY})} P_{Y|X}(b|a)^{nP_{XY}(a,b)(1-\epsilon_1)} \\ &= 2^{n(1-\epsilon_1)\sum_{(a,b)\in \text{supp}(P_{XY})} P_{XY}(a,b)\log_2 P_{Y|X}(b|a)} \\ &= 2^{-n(1-\epsilon_1)H(Y|X)}. \end{aligned} \tag{1.48}$$

This gives the lower bound in (1.25) and the upper bound is proved similarly.

Next, suppose that $(x^n, y^n) \in T^n_\epsilon(P_{XY})$ and (X^n, Y^n) was emitted by the DMS $P_{XY}(\cdot)$. We prove (1.27) as follows.

Consider first $P_{XY}(a,b) = 0$ for which we have

$$\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} > P_{XY}(a,b)(1+\epsilon)\right] = 0. \tag{1.49}$$

Now consider $P_{XY}(a,b) > 0$. If $N(a|x^n) = 0$, then $N(a,b|x^n,y^n) = 0$ and

$$\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} > P_{XY}(a,b)(1+\epsilon)\middle| X^n = x^n\right] = 0. \tag{1.50}$$

More interestingly, if $N(a|x^n) > 0$ then the Chernoff bound gives

$$\begin{aligned} &\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} > P_{XY}(a,b)(1+\epsilon)\middle| X^n = x^n\right] \\ &\le \Pr\left[\frac{N(a,b|X^n,Y^n)}{n} \ge P_{XY}(a,b)(1+\epsilon)\middle| X^n = x^n\right] \\ &= \Pr\left[\frac{N(a,b|X^n,Y^n)}{N(a|x^n)} \ge \frac{P_{XY}(a,b)}{N(a|x^n)/n}(1+\epsilon)\middle| X^n = x^n\right] \end{aligned}$$

$$
\begin{aligned}
&\le E\left[e^{\nu N(a,b|X^n,Y^n)/N(a|x^n)} \middle| X^n = x^n \right] e^{-\nu \frac{P_{XY}(a,b)(1+\epsilon)}{N(a|x^n)/n}} \\
&= \left[\sum_{m=0}^{N(a|x^n)} \binom{N(a|x^n)}{m} P_{Y|X}(b|a)^m (1-P_{Y|X}(b|a))^{N(a|x^n)-m} \right. \\
&\qquad \left. e^{\nu m/N(a|x^n)} \right] e^{-\nu \frac{P_{XY}(a,b)(1+\epsilon)}{N(a|x^n)/n}} \\
&= \left[(1-P_{Y|X}(b|a)) + P_{Y|X}(b|a) e^{\nu/N(a|x^n)} \right]^{N(a|x^n)} e^{-\nu \frac{P_{XY}(a,b)(1+\epsilon)}{N(a|x^n)/n}}.
\end{aligned} \tag{1.51}
$$

Minimizing (1.51) with respect to ν, we find that

$$
\begin{aligned}
&\nu = \infty && \text{if } P_{XY}(a,b)(1+\epsilon) \ge N(a|x^n)/n \\
&e^{\nu/N(a|x^n)} = \frac{P_X(a)(1-P_{Y|X}(b|a))(1+\epsilon)}{N(a|x^n)/n - P_{XY}(a,b)(1+\epsilon)} && \text{if } P_{XY}(a,b)(1+\epsilon) < N(a|x^n)/n.
\end{aligned} \tag{1.52}
$$

Again, the Chernoff bound correctly identifies the probabilities to be 0 and $P_{Y|X}(b|a)^n$ for the cases $P_{XY}(a,b)(1+\epsilon) > N(a|x^n)/n$ and $P_{XY}(a,b)(1+\epsilon) = N(a|x^n)/n$, respectively. More interestingly, for $P_{XY}(a,b)(1+\epsilon) < N(a|x^n)/n$ we insert (1.52) into (1.51) and obtain

$$
\Pr\left[\frac{N(a,b|X^n)}{n} \ge P_{XY}(a,b)(1+\epsilon) \middle| X^n = x^n \right] \le 2^{-N(a|x^n) D(P_B \| P_A)}, \tag{1.53}
$$

where A and B are binary random variables with

$$
\begin{aligned}
P_A(0) &= 1 - P_A(1) = P_{Y|X}(b|a) \\
P_B(0) &= 1 - P_B(1) = \frac{P_{XY}(a,b)}{N(a|x^n)/n}(1+\epsilon).
\end{aligned} \tag{1.54}
$$

We would like to have the form $P_B(0) = P_A(0)(1+\tilde{\epsilon})$ and compute

$$
\tilde{\epsilon} = \frac{P_X(a)}{N(a|x^n)/n}(1+\epsilon) - 1. \tag{1.55}
$$

We can now use (1.41)–(1.46) to arrive at

$$
\begin{aligned}
&\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} \ge P_{XY}(a,b)(1+\epsilon) \middle| X^n = x^n \right] \\
&\qquad \le e^{-N(a|x^n)\tilde{\epsilon}^2 P_{Y|X}(b|a)}
\end{aligned} \tag{1.56}
$$

as long as $\epsilon \leq \min_{b:(a,b)\in\mathrm{supp}(P_{XY})} P_{Y|X}(b|a)$. Now to guarantee that $\tilde{\epsilon}^2$ is positive, we must require that x^n is "more than" ϵ-letter typical, i.e., we must choose $x^n \in T_{\epsilon_1}(P_X)$, where $0 \leq \epsilon_1 < \epsilon$. Inserting $N(a|x^n)/n \geq (1+\epsilon_1)P_X(a)$ into (1.56), we have

$$\begin{aligned}\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} \geq P_{XY}(a,b)(1+\epsilon)\middle| X^n = x^n\right] \\ \leq e^{-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}P_{XY}(a,b)}\end{aligned} \tag{1.57}$$

for $0 \leq \epsilon_1 < \epsilon \leq \mu_{XY}$ (we could allow ϵ to be up to $\min_{b:(a,b)\in\mathrm{supp}(P_{XY})}$ $P_{Y|X}(b|a)$ but we ignore this subtlety). One can similarly bound

$$\begin{aligned}\Pr\left[\frac{N(a,b|X^n,Y^n)}{n} \leq P_{XY}(a,b)(1-\epsilon)\middle| X^n = x^n\right] \\ \leq e^{-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}P_{XY}(a,b)}.\end{aligned} \tag{1.58}$$

As for the unconditioned case, note that (1.57) and (1.58) are valid for all (a,b) including (a,b) with $P_{XY}(a,b) = 0$. However, the event we are interested in has a strict inequality so that we can improve the above bounds for the case $P_{XY}(a,b) = 0$ (see (1.49)). We can thus replace $P_{XY}(a,b)$ in (1.57) and (1.58) with μ_{XY} and the result is

$$\begin{aligned}\Pr\left[\left|\frac{N(a,b|X^n,Y^n)}{n} - P_{XY}(a,b)\right| > \epsilon P_{XY}(a,b)\middle| X^n = x^n\right] \\ \leq 2\cdot e^{-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{XY}}.\end{aligned} \tag{1.59}$$

for $0 \leq \epsilon_1 < \epsilon \leq \mu_{XY}$ (we could allow ϵ to be up to $\mu_{Y|X} = \min_{(a,b)\in\mathrm{supp}(P_{XY})} P_{Y|X}(b|a)$ but, again, we ignore this subtlety). We thus have

$$\begin{aligned}&\Pr\left[Y^n \notin T_\epsilon^n(P_{XY}|x^n) \middle| X^n = x^n\right] \\ &\quad = \Pr\left[\bigcup_{a,b}\left\{\left|\frac{N(a,b|X^n)}{n} - P_{XY}(a,b)\right| > \epsilon P_{XY}(a,b)\right\}\middle| X^n = x^n\right]\end{aligned}$$

$$\leq \sum_{a,b} \Pr\left[\left| \frac{N(a,b|X^n,Y^n)}{n} - P_{XY}(a,b) \right| > \epsilon P_{XY}(a,b) \middle| X^n = x^n \right]$$

$$\leq 2|\mathcal{X}||\mathcal{Y}| \cdot e^{-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{XY}}, \tag{1.60}$$

where we have used the union bound for the last inequality. The result is the left-hand side of (1.27).

Finally, for $x^n \in T^n_{\epsilon_1}(P_X)$ and $0 \leq \epsilon_1 < \epsilon \leq \mu_{XY}$ we have

$$\begin{aligned} \Pr\left[Y^n \in T^n_\epsilon(P_{XY}|x^n) \,|\, X^n = x^n\right] &= \sum_{y^n \in T^n_\epsilon(P_{XY}|x^n)} P^n_{Y|X}(y^n|x^n) \\ &\leq |T^n_\epsilon(P_{XY}|x^n)|\, 2^{-n(1-\epsilon)H(Y|X)}, \end{aligned} \tag{1.61}$$

where the inequality follows by (1.48). We thus have

$$|T^n_\epsilon(P_{XY}|x^n)| \geq (1 - \delta_{\epsilon_1,\epsilon}(n))\, 2^{n(1-\epsilon)H(Y|X)}. \tag{1.62}$$

We similarly have

$$|T^n_\epsilon(P_{XY}|x^n)| \leq 2^{n(1+\epsilon)H(Y|X)}. \tag{1.63}$$

References

[1] R. Ahlswede, "Multi-way communication channels," in *Proceedings of 2nd International Symposium Information Theory (1971)*, pp. 23–52, Tsahkadsor, Armenian S.S.R.: Publishing House of the Hungarian Academy of Sciences, 1973.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, New Jersey: Prentice Hall, 1993.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method*. New York: Wiley, Second ed., 2000.

[4] A. Amraoui, S. Dusad, and R. Urbanke, "Achieving general points in the 2-user Gaussian MAC without time-sharing or rate-splitting by means of iterative coding," in *Proceedings of IEEE International Symposium on Information Theory*, p. 334, Lausanne, Switzerland, June 30–July 5 2002.

[5] M. R. Aref, *Information Flow in Relay Networks*. PhD thesis, Stanford, CA: Stanford University, October 1980.

[6] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, (G. Longo, ed.), pp. 171–231, Berlin, Germany: Springer Verlag, 1978.

[7] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, March 1973.

[8] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 279–280, March 1974.

[9] N. Blachman, "The convolution inequality for entropy powers," *IEEE Transactions on Information Theory*, vol. 11, no. 2, pp. 267–271, April 1965.

[10] S. I. Bross, A. Lapidoth, and M. A. Wigger, "The Gaussian MAC with conferencing encoders," in *Proceedings of IEEE International Symposium on Information Theory*, Toronto, Canada, July 6–11 2008.

[11] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 841–850, November 1982.

[12] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

[13] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[14] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, March 1975.

[15] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, January 1972.

[16] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.

[17] T. M. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[19] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Budapest: Akadémiai Kiadó, 1981.

[20] R. L. Dobrushin, "Information transmission in a channel with feedback," *Theory of Probabilistic Applications*, vol. 34, pp. 367–383, December 1958.

[21] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Transactions on Information Theory*, vol. 28, no. 3, p. 536, May 1982.

[22] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, November 1982.

[23] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 120–122, January 1981.

[24] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[25] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 100–102, January 1975.

[26] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[27] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, July–September 1974.

[28] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[29] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple-access for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 873–890, March 2001.

[30] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1877–1894, August 2003.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[32] O. Johnson, "A conditional entropy power inequality for dependent random variables," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1581–1583, August 2004.

[33] O. Johnson, *Information Theory and the Central Limit Theorem*. London, UK: Imperial College Press, 2004.

[34] R. C. King, *Multiple Access Channels with Generalized Feedback*. PhD thesis, Stanford, CA: Stanford University, March 1978.

[35] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 60–64, January 1977.

[36] G. Kramer, *Directed Information for Channels with Feedback,* volume ETH Series in Information Processing. Vol. 11, Konstanz, Germany: Hartung-Gorre Verlag, 1998.

[37] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, January 2003.

[38] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, September 2005.

[39] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, no. 3–4, pp. 271–425, 2006.

[40] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *Journal of Network and Systems Management*, vol. 14, no. 1, pp. 49–67, March 2006.

[41] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3517–3535, October 2007.

[42] H. Liao, "A coding theorem for multiple access communications," in *Proceedings of IEEE International Symposium on Information Theory*, Asilomar, CA, 1972.

[43] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[44] J. L. Massey, *Applied Digital Information Theory*. Zurich, Switerland: ETH Zurich, 1980–1998.

[45] J. L. Massey, "Causality, feedback and directed information," in *Proceedings of IEEE International Symposium on Information Theory Applications*, pp. 27–30, Hawaii, USA, November 1990.

[46] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[47] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, December 1980.

[48] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, July 1984.

[49] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[50] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple-descriptions — Part I: (n,k) source-channel erasure codes," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 47–61, January 2004.

[51] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple-descriptions — Part II: an achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1377–1392, April 2005.

[52] T. J. Richardson, A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[53] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, March 1996.

[54] H. Sato, "An outer bound to the capacity region of broadcast channels," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 374–377, May 1978.

[55] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July and October 1948, (Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 5–83, (N. J. A. Sloane and A. D. Wyner, eds.) Piscataway: IEEE Press, 1993).

[56] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transaction Information Theory*, vol. 2, pp. 221–238, September 1956, (Reprinted in *Claude Elwood Shannon: Collected Papers*, (N. J. A. Sloane and A. D. Wyner, eds.) pp. 221–238, Piscataway: IEEE Press, 1993).

[57] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE International Convention Record*, pp. 142–163, March 1959. (Reprinted in *Claude Elwood Shannon: Collected Papers*, (N. J. A. Sloane and A. D. Wyner, eds.) pp. 325–350, Piscataway: IEEE Press, 1993.

[58] C. E. Shannon, "Two-way communication channels," in *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, (J. Neyman, ed.), pp. 611–644, Berkeley, CA: University California Press, 1961. (Reprinted in *Claude Elwood Shannon: Collected Papers*, (N. J. A. Sloane and A. D. Wyner, eds.), pp. 351–384, Piscataway: IEEE Press, 1993).

[59] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Technical Journal*, vol. 52, pp. 1037–1076, September 1973.

[60] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 9, pp. 471–480, July 1973.

[61] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information Control*, vol. 2, pp. 101–112, July 1959.

[62] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunication*, vol. 10, no. 6, pp. 585–595, November–December 1999.

[63] E. C. van der Meulen, *Transmission of Information in a T-Terminal Discrete Memoryless Channel*. PhD thesis, Berkeley, CA: University of California, January 1968.

[64] R. Venkataramani, G.Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, September 2003.

[65] H. Wang and P. Viswanath, "Vector Gaussian multiple-description for individual and central receivers," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2133–2153, June 2007.

[66] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, September 2006.

[67] F. M. J. Willems, *Information Theoretical Results for the Discrete Memoryless Multiple Access Channel*. PhD thesis, Leuven, Belgium: Katholieke Universiteit, October 1982.

[68] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[69] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 772–777, November 1973.

[70] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769–772, November 1973.

[71] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[72] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[73] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1348–1358, April 2005.

[74] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in *Proceedings of 2001 IEEE International Symposium Information Theory*, p. 302, Washington, DC, June 24–29 2001.

[75] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Transactions on Information Theory*, vol. 33, no. 4, pp. 502–521, July 1987.