ing, IEEE Trans. Commun., vol. COM-26, pp. 1819–1821, 1978.
[15] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The capacity of
[17] T. E. Stern, "Some quantum effects in information channels," IEEE Trans. Inform. Theory, vol. IT-6, pp. 435–440, 1960.

To Get a Bit of Information May Be As Hard As to Get Full Information

RUDOLF AHLSWEDE AND IMRE CSISZÁR

Abstract—**The following coding problem for correlated discrete memoryless sources is considered. The two sources can be separately block encoded, and the values of the encoding functions are available to a decoder who wants to answer a certain question concerning the source outputs. Typically, this question has only a few possible answers (even as few as two). The rates of the encoding functions must be found that enable the decoder to answer this question correctly with high probability. It is proven that these rates are often as large as those needed for a full reproduction of the outputs of both sources. Furthermore, if one source is completely known at the decoder, this phenomenon already occurs when what is asked for is the joint type (joint composition) of the two source output blocks, or some function thereof such as the Hamming distance of the two blocks or (for alphabet size at least three) just the parity of this Hamming distance.**

I. Introduction

WE ARE given a discrete memoryless double source (DMDS) with alphabets $\mathcal{X}$, $\mathcal{Y}$, and generic variables X, Y, i.e., a sequence of independent replicas (X_i, Y_i), $i = 1, 2, \cdots$, of the pair of random variables (X, Y) taking values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Slepian and Wolf [9] considered the problem of encoding the source output blocks $X^n \triangleq X_1 \cdots X_n$ resp. $Y^n \triangleq Y_1 \cdots Y_n$ by two separate encoders in such a way that a common decoder could reproduce both blocks with small probability of error. They proved that such an encoding is possible with rates (R_1, R_2) if and only if

$$R_1 \geq H(\hat{X}|Y), \quad R_2 \geq H(Y|X), \quad R_1 + R_2 \geq H(X, Y). \tag{1.1}$$

Manuscript received March 19, 1980.

R. Ahlswede is with the Department of Mathematics at the University of Bielefeld, 4800 Bielefeld, West Germany.

I. Csiszár was with the University of Bielefeld, on leave from the Mathematical Institute of the Hungarian Academy of Sciences, 1053 Budapest, Hungary.

It may happen, however, that what is actually required at the decoder is to answer a certain question concerning (X^n, Y^n). Such a question can of course be described by a function F of (X^n, Y^n). We are interested in those functions for which the number k_n of possible values of $F(X^n, Y^n)$ satisfies

$$\lim_{n \to \infty} \frac{1}{n} \log k_n = 0. \tag{1.2}$$

This means that the questions asked have only "a few" possible answers. For example, X_i and Y_i may be the results of two different quality control tests performed on the ith item of a lot. Then for certain purposes, e.g., for determining the price of the lot, one may be interested only in the frequencies of the various possible pairs (x, y) among the results, their order, i.e., the knowledge of the individual pairs (X_i, Y_i), being irrelevant. In this case $k_n \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}$, and (1.2) holds. A natural first question is whether or not it is always true in this case that, for large n, arbitrarily small encoding rates permit the decoder to determine $F(X^n, Y^n)$. To our knowledge, even this seemingly simple question had not been answered prior to this paper, except for the particular case of independent binary X and Y, where one of them takes the values 0, 1 with equal probabilities. In this particular case, Körner [6] showed the necessity of positive rates if both entropies are positive.

We also consider here other choices of F and first obtain the following result. For every DMDS with

$$H(X|Y) > 0, \quad H(Y|X) > 0$$

there exists a binary question (function F with only two possible values) such that in order to answer this question

0018-9448/81/0700-0398$00.75

(determine $F(X^n, Y^n)$) one needs encoding rates as specified in (1.1).

As a matter of fact, almost all randomly selected functions F are of this kind. Since the reason for this unexpected phenomenon might be that randomly selected functions are very irregular, we next study more regular functions. A function F of special interest is the joint type of the two source blocks hinted at in the quality control example. In this respect our main result is that for determining the joint type of X^n and Y^n when Y^n is completely known at the decoder, X^n must be encoded with just as large a rate as if X^n were to be fully reproduced, except for (exactly specified) singular cases. Actually, we shall prove an analogous result for a class of functions F which include, in addition to the joint type, the Hamming distance and—for alphabet size at least three—the parity of the Hamming distance.

As a consequence of these results one obtains that in the case of encoding both X^n and Y^n, the rates must satisfy

$$R_1 \geq H(X|Y), \quad R_2 \geq H(Y|X), \tag{1.3}$$

in order that the joint type or the Hamming distance of X^n and Y^n can be determined by the decoder. In particular, it follows that for a DMDS with independent components (i.e., when X and Y are independent random variables (RV's)) nothing can be gained in rates, if instead of (X^n, Y^n) only the joint type or the Hamming distance of X^n and Y^n is to be determined by the decoder. For a DMDS with dependent components such a rate gain is possible, although it remains to be seen whether this always happens and to what extent. At present a complete solution to this problem is available only in the binary symmetric case. In fact, it readily follows from a result of Körner and Marton, published in [5], that our necessary conditions (1.3) are also sufficient. Let us emphasize that their result concerns "componentwise" functions F

$$F(X^n, Y^n) \triangleq (F_1(X_1, Y_1), F_1(X_2, Y_2), \cdots, F_1(X_n, Y_n)), \tag{1.4}$$

where F_1 is defined on $\mathcal{X} \times \mathcal{Y}$.

In the binary symmetric case (i.e. $\Pr\{X = Y = 0\} = \Pr\{X = Y = 1\}$, $\Pr\{X = 0, Y = 1\} = \Pr\{X = 1, Y = 0\}$), they proved for the particular F with $F_1(x, y) \triangleq x + y \pmod 2$ that (R_1, R_2) is an achievable rate pair for determining $F(X^n, Y^n)$ if and only if (1.3) holds. Now observe that the types of X^n and of Y^n can be encoded with arbitrarily small rates and that those two types and the mod 2 sum $F(X^n, Y^n)$ determine the Hamming distance and also the joint type of X^n, Y^n.

II. Notation

We use the notation of [4]. Script capitals denote finite sets. The cardinality of a set $\mathcal{A}$ resp. of the range of a function f is denoted by $|\mathcal{A}|$ resp. $\|f\|$. The letters P, Q always stand for probability distributions (PD's) on finite sets. A stochastic matrix $W = \{W(y|x)\colon x \in \mathcal{X}, y \in \mathcal{Y}\}$ with rows $W(\cdot|x)$ which are PD's on $\mathcal{Y}$ is referred to as $W\colon \mathcal{X} \to \mathcal{Y}$. The discrete memoryless channel (DMC) with this transition probability matrix is denoted by $\{W\colon \mathcal{X} \to \mathcal{Y}\}$. Given a PD P on $\mathcal{X}$ resp. a stochastic matrix $W\colon \mathcal{X} \to \mathcal{Y}$, for $x = x_1 \cdots x_n \in \mathcal{X}^n$, $y = y_1 \cdots y_n \in \mathcal{Y}^n$ we write

$$P^n(x) \triangleq \prod_{i=1}^{n} P(x_i), \quad W^n(y|x) \triangleq \prod_{i=1}^{n} W(y_i|x_i).$$

We denote the distribution of a RV X by P_X and the conditional distribution of Y given X by $P_{Y|X}$. For $W\colon \mathcal{X} \to \mathcal{Y}$, the equality $P_{Y|X} = W$ means that

$$P_{Y|X}(y|x) = W(y|x)$$

whenever the left side is defined, i.e., $P_X(x) \neq 0$. The *type* P_x of a sequence $x \in \mathcal{X}^n$ is a PD on $\mathcal{X}$ where $P_x(x)$ is the relative frequency of x in x. The *joint type* $P_{x,y}$ of two sequences $x \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$ is the PD on $\mathcal{X} \times \mathcal{Y}$ defined similarly.

Given any PD P on $\mathcal{X}$, we designate by $\mathcal{T}_P^n$ the set of those sequences $x \in \mathcal{X}^n$ which have type $P_x = P$. Further, given an RV X and a positive number η, we denote by $\mathcal{T}_{X,\eta}^n$ the set of (X, η)-*typical sequences* in $\mathcal{X}^n$, i.e.,

$$\mathcal{T}_{X,\eta}^n \triangleq \{x\colon x \in \mathcal{X}^n, |P_x(x) - P_X(x)| \leq \eta \text{ for every } x \in \mathcal{X}\} = \bigcup_{P\colon \max_{x \in \mathcal{X}} |P(x) - P_X(x)| \leq \eta} \mathcal{T}_P^n. \tag{2.1}$$

All exponents and logarithms in this paper are to the base two. The following well-known facts about types and typical sequences will be used:

$$|\{P\colon \mathcal{T}_P^n \neq \varnothing\}| \leq (n+1)^{|\mathcal{X}|}, \tag{2.2}$$

$$(n+1)^{-|\mathcal{X}|} \exp[nH(P)] \leq |\mathcal{T}_P^n| \leq \exp[nH(P)], \quad \text{if } \mathcal{T}_P^n \neq \varnothing, \tag{2.3}$$

$$P_X^n(\mathcal{T}_{X,\eta}^n) \geq 1 - \frac{|\mathcal{X}|}{n\eta^2}, \tag{2.4}$$

and for every $\alpha > 0$, $\delta > 0$, and $n \geq n_0 = n_0(|\mathcal{X}|, \alpha, \delta)$,

$$\mathcal{A} \subset \mathcal{X}^n, \quad P_X^n(\mathcal{A}) \geq \alpha \text{ imply } |\mathcal{A}| \geq \exp[n(H(X) - \delta)]. \tag{2.5}$$

The inequalities (2.2) and (2.4) are obvious while (2.5) follows from (2.4), (2.1), (2.3), and the continuity of the entropy function. Equation (2.3) can be checked with Stirling's formula; a simple direct proof is given in [4].

The *Hamming distance* of two sequences $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$ will be denoted by $d(x, y)$:

$$d(x, y) \triangleq |\{i\colon x_i \neq y_i, 1 \leq i \leq n\}|.$$

III. The Results

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets and F an arbitrary function on $\cup_{n=1}^{\infty}(\mathcal{X}^n \times \mathcal{Y}^n)$.

Definition 1: A *length-n block-F-code* for a double source with alphabets $\mathcal{X}$, $\mathcal{Y}$ (or, briefly, a source F-code) is a triple of mappings (f, g, φ) where the *encoders* f and g map $\mathcal{X}^n$ resp. $\mathcal{Y}^n$ into some finite sets whereas the *decoder* φ maps the Cartesian product of the latter into the range of F.

Given a DMDS with alphabets $\mathcal{X}, \mathcal{Y}$, the *probability of error* of this F-code is

$$e \triangleq e(f, g, \varphi; F)$$
$$= \Pr\{\varphi(f(X^n), g(Y^n)) \neq F(X^n, Y^n)\}. \quad (3.1)$$

Further, a pair of nonnegative numbers (R_1, R_2) is an *achievable F rate pair* if for every $\epsilon > 0$, $\delta > 0$, and sufficiently large n there exists a source F-code (f, g, φ) of block length n with

$$\frac{1}{n}\log\|f\| < R_1 + \delta,$$
$$\frac{1}{n}\log\|g\| < R_2 + \delta,$$
$$e(f, g, \varphi; F) < \epsilon. \quad (3.2)$$

When not mentioning F we shall always understand that F is the identity mapping. In this case Definition 1 reduces to the familiar one of block codes and achievable rate pairs for double sources given by Slepian and Wolf [9].

Theorem 1: For every DMDS satisfying the conditions

$$H(X|Y) > 0, \quad H(Y|X) > 0 \quad (3.3)$$

there exists a binary valued function F on $\cup_{n=1}^{\infty}(\mathcal{X}^n \times \mathcal{Y}^n)$ such that only those (R_1, R_2) are achievable F rate pairs which satisfy the Slepian–Wolf condition (1.1).

This theorem will be proved in the next section by randomly selecting the value of F for every (x, y). In this way, for every possible joint type P of pairs (x, y), $x \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$, the set $\mathcal{T}_P^n$ of pairs having joint type P will be partitioned very irregularly by F. One might think that this irregularity is the reason for the unexpected result. Hence, to get more insight into the problem, we now focus attention to such functions F which are constant on each $\mathcal{T}_P^n \subset \mathcal{X}^n \times \mathcal{Y}^n$, i.e., $F(x, y)$ is a function of the joint type $P_{x,y}$.

Example 1: Let $\mathcal{X} \triangleq \{0, 1\}$, $\mathcal{Y} \triangleq \{0, 1, 2\}$, $F(x, y) \triangleq P_{x,y}$. Consider a DMDS with generic variables X, Y such that

$$P_{XY}(0, 1) = P_{XY}(1, 0) = 0. \quad (3.4)$$

Thus, since no pairs $(0, 1)$ or $(1, 0)$ are possible, the joint type of X^n and Y^n is uniquely determined (with probability one) if the types of X^n and of Y^n are known. This can be achieved with encoders of rates approaching zero. Hence, for DMDS's satisfying (3.4), even $(0, 0)$ is an achievable F rate pair.

We shall see that this example is quite atypical. For most DMDS's positive rates are needed in order to determine the joint type of X^n and Y^n. Conclusive answers (coinciding direct and converse results) will be obtained only for the projections of achievable F rate regions to the R_1-axis, or what is the same, for codes with Y^n completely known at the decoder. Our results will be of the kind that if Y^n is completely known at the decoder, then as large a rate of the X encoder is needed to determine $F(X^n, Y^n)$ as to reproduce X^n itself, i.e., $H(X|Y)$, except for singular cases such as Example 1.

We say that R_1 is an *achievable F rate in the knowledge of Y* if for every $\epsilon > 0$, $\delta > 0$ and sufficiently large n there exist length-n block-F-codes (f, g, φ) such that g is the identity mapping on Y^n and (3.2) holds with $R_2 \triangleq \log|\mathcal{Y}|$. Clearly this happens if and only if there exists any R_2 such that (R_1, R_2) is an achievable F rate pair in the sense of Definition 1. A central result is the following

Theorem 2: Let $F(x, y) \triangleq P_{x,y}$ and let us be given an arbitrary DMDS with generic variables X, Y such that for every $x_1 \neq x_2$ in $\mathcal{X}$, the number of elements $y \in \mathcal{Y}$ with

$$P_{XY}(x_1, y) \cdot P_{XY}(x_2, y) > 0$$

is different from one. Then R is an achievable F rate in the knowledge of Y if and only if

$$R \geq H(X|Y). \quad (3.5)$$

Remark: It is easy to see that the condition of Theorem 2 on the joint distribution P_{XY} is necessary for the assertion to hold. In fact, if to some $x_1 \neq x_2$ there exists exactly one $y^* \in \mathcal{Y}$ for which $P_{XY}(x_1, y^*)$ and $P_{XY}(x_2, y^*)$ are both positive, then write

$$\pi(x) \triangleq \begin{cases} x, & \text{if } x \in \mathcal{X} \setminus \{x_1, x_2\}, \\ \{x_1, x_2\}, & \text{if } x = x_1, \text{ or } x = x_2. \end{cases}$$

Notice that $\pi(X_1)\cdots\pi(X_n)$ and Y^n uniquely determine the frequency of each pair (x, y) among the random pairs (X_i, Y_i), $i = 1, \cdots, n$, with the exception of the pairs (x_1, y^*) and (x_2, y^*). For the latter, only the sum of their frequencies is determined, but this ambiguity will be removed if the frequency of x_1 in X^n is known. Applying the theorem of Slepian and Wolf [9] for reproducing $\pi(X_1), \cdots, \pi(X_n)$ and using the fact that the frequency of x_1 in X^n can be communicated with asymptotically zero rate, it follows that $H(\pi(X)|Y) < H(X|Y)$ is an achievable F rate in the knowledge of Y.

The proof of Theorem 2 will be based on the property of the function $F(x, y) \triangleq P_{x,y}$ that changing y in one component drastically changes the partition of $\mathcal{X}^n$ whose atoms are the sets on which $F(x, y)$ is constant (for the given y). We shall refer to this property as high sensitivity.

Definition 2: A function F on $\cup_{n=1}^{\infty}(\mathcal{X}^n \times \mathcal{Y}^n)$ is *highly sensitive* if for every $x_1 \neq x_2$ in $\mathcal{X}$ and $y_1 \neq y_2$ in $\mathcal{Y}$ the following holds. Whenever $x \in \mathcal{X}^n$, $x' \in \mathcal{X}^n$, $y \in \mathcal{Y}^n$ have ith component x_1, x_2 and y_1, respectively, and $F(x, y) = F(x', y)$, then for the sequence $y' \in \mathcal{Y}^n$ obtained from y by replacing the ith component by y_2 we always have $F(x, y') \neq F(x', y')$. Further, F will be called *sensitive*, if for every $x_1 \neq x_2$ in $\mathcal{X}$ and y_1 in $\mathcal{Y}$ there exists a y_2 in $\mathcal{Y}$ for which the above statement holds.

Remark: Clearly $F(x, y) \triangleq P_{x,y}$ is highly sensitive, and if $\mathcal{X} \subset \mathcal{Y}$, then $F(x, y) \triangleq d(x, y)$ is sensitive. Further, if

$\mathcal{X} \subset \mathcal{Y}$ and $|\mathcal{Y}| \geq 3$, then the binary valued function

$$F(x, y) \triangleq \begin{cases} 0, & \text{if } d(x, y) \text{ is even,} \\ 1, & \text{if } d(x, y) \text{ is odd,} \end{cases} \tag{3.6}$$

is sensitive.

Instead of Theorem 2 we shall actually prove the more general Theorem 3 below. In addition to Theorem 2, it also contains the results that an achievable F rate in the knowledge of Y for a DMDS with strictly positive P_{XY} must satisfy (3.5), if: a) $\mathcal{X} \subset \mathcal{Y}$, and F is the Hamming distance; or b) $\mathcal{X} \subset \mathcal{Y}$, $|\mathcal{Y}| \geq 3$, and F is the parity of the Hamming distance.

Theorem 3: If for a DMDS with generic variables X, Y, all probabilities $P_{XY}(x, y)$ $(x \in \mathcal{X}, y \in \mathcal{Y})$ are positive and F is sensitive, then R is an achievable F rate in the knowledge of Y if and only if $R \geq H(X|Y)$. If F is highly sensitive then the same is true even under the weaker condition that for every $x_1 \neq x_2$ in $\mathcal{X}$ the number of elements $y \in \mathcal{Y}$ with

$$P_{XY}(x_1, y) \cdot P_{XY}(x_2, y) > 0$$

is different from one.

This theorem will be a consequence of Theorem 4 below, which states a result for channel F-codes. From a mathematical point of view, channel F-codes are a natural counterpart to source F-codes. We shall not enter the question whether channel F-codes also correspond to some real communication situations since for us their primary role will be to provide a tool for proving Theorem 3 (via Theorem 4).

Definition 3: A *length-n block-F-code* for a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ (or briefly, a channel F-code) is a pair $(\mathcal{C}, \psi)$, where the *codeword set* $\mathcal{C}$ is a subset of $\mathcal{X}^n$ and the *decoder* ψ is a mapping of $\mathcal{Y}^n$ into the range of F. Given a DMC $\{W: \mathcal{X} \to \mathcal{Y}\}$, the *average probability of error* of this F-code is

$$\begin{aligned} \bar{e} &= \bar{e}(\mathcal{C}, \psi; F) \\ &\triangleq \frac{1}{|\mathcal{C}|} \sum_{x \in \mathcal{C}} W^n\big(\{y: \psi(y) \neq F(x, y)\} | x\big). \end{aligned} \tag{3.7}$$

Remark: Clearly, if $F(x, y) \triangleq x$, then Definition 3 reduces to the usual definition of channel block codes and their average probability of error. In the general case, F-codes may be interpreted as *list codes*, putting for a received sequence $y \in \mathcal{Y}^n$ exactly those codewords $x \in \mathcal{C}$ on the list which satisfy $F(x, y) = \psi(y)$. Our present point of view is, however, quite different from that of previous papers dealing with list codes, e.g., [1], [8], for we are interested in lists of a specific structure and pay no attention to list size. A connection between source and channel F-coding problems is established by the following simple lemma.

Lemma 1: Consider a DMDS with generic variables X, Y and a DMC $\{W: \mathcal{X} \to \mathcal{Y}\}$ with $W = P_{Y|X}$. Then for every $\delta > 0$, $\eta > 0$ and block length $n \geq n_0(|\mathcal{X}|, \delta, \eta)$, to each source F-code (f, g, φ) there exists a channel F-code $(\mathcal{C}, \psi)$ such that

$$\bar{e}(\mathcal{C}, \psi; F) \leq 2e(f, g, \varphi; F), \tag{3.8}$$

$$\frac{1}{n} \log |\mathcal{C}| \geq H(X) - \frac{1}{n} \log \|f\| - \delta, \tag{3.9}$$

$$\mathcal{C} \subset \mathcal{T}^n_{X, \eta}, \tag{3.10}$$

$$d(x, x') \geq \tau\eta, \text{ if } x \neq x' \text{ are in } \mathcal{C}, \tag{3.11}$$

where τ is a positive number depending only on $|\mathcal{X}|$ and δ.

This Lemma and the following theorem will be proved in Section V.

Theorem 4: Let a DMC $\{W: \mathcal{X} \to \mathcal{Y}\}$ have positive transition probabilities, i.e., $W(y|x) > 0$ for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and let F be a sensitive function defined on $\cup_{n=1}^{\infty}(\mathcal{X}^n \times \mathcal{Y}^n)$. Then to every $\tau > 0$ and $\lambda > 0$ there is an $\epsilon > 0$ such that a channel F-code $(\mathcal{C}, \psi)$ of any block length n can satisfy

$$\bar{e}(\mathcal{C}, \psi; F) < \epsilon \tag{3.12}$$

and

$$d(x, x') \geq \tau n, \qquad \text{for } x \neq x' \text{ in } \mathcal{C} \tag{3.13}$$

only if there exists a decoder $\tilde{\psi}: \mathcal{Y}^n \to \mathcal{C}$ such that the channel code $(\mathcal{C}, \tilde{\psi})$ in the usual sense has average probability of error less than λ, i.e.,

$$\bar{e}(\mathcal{C}; \tilde{\psi}) \triangleq \frac{1}{|\mathcal{C}|} W^n\big(\{y: \tilde{\psi}(y) \neq x\} | x\big) < \lambda. \tag{3.14}$$

Further, if F is highly sensitive, then the same holds also under the weaker condition on W that for every $x_1 \neq x_2$ in $\mathcal{X}$ the number of elements $y \in \mathcal{Y}$ with

$$W(y|x_1) W(y|x_2) > 0$$

is different from one.

It is easy to see that Lemma 1 and Theorem 4 do imply Theorem 3. In fact, because of the theorem of Slepian and Wolf [9], only the converse part of Theorem 3 needs a proof. In other words, the statement to be verified is that, if for every $\delta > 0$, $\epsilon' > 0$ and sufficiently large block length n there exists a source F-code (f, g, φ) such that g is the identity mapping on $\mathcal{Y}^n$ and

$$\frac{1}{n} \log \|f\| < R + \delta, \quad e(f, g, \varphi; F) < \epsilon', \tag{3.15}$$

then necessarily $R \geq H(X|Y)$.

To deduce this from Lemma 1 and Theorem 4, consider a DMC $\{W: \mathcal{X} \to \mathcal{Y}\}$ with $W = P_{Y|X}$, fix an $\eta > 0$ to be specified later, and find for the above source F-code (f, g, φ) a channel F-code $(\mathcal{C}, \psi)$ as in Lemma 1. Then by (3.9) and (3.15) we have

$$\frac{1}{n} \log |\mathcal{C}| \geq H(X) - R - 2\delta. \tag{3.16}$$

Further, to the τ of Lemma 1 and an arbitrary $\lambda > 0$,

choose $\epsilon > 0$ as in Theorem 4. Since we may assume that $\epsilon' \le \epsilon/2$, we then see from (3.15), (3.8), and (3.11), that Theorem 4 is applicable to the channel F-code $(\mathcal{C}, \psi)$.

There follows the existence of an ordinary channel code $(\mathcal{C}, \tilde{\psi})$ with the same codeword set $\mathcal{C}$ which has average probability of error less than λ. Since $\mathcal{C} \subset \mathcal{T}^n_{X,\eta}$ by (3.10), this implies

$$\frac{1}{n}\log|\mathcal{C}| \le I(X \wedge Y) + \delta$$

provided that λ and η have been chosen sufficiently small (depending on δ). Comparing this with (3.16) and remembering that $\delta > 0$ can be arbitrarily small, we get the desired inequality $R \ge H(X|Y)$.

IV. Proof of Theorem 1

Theorem 1 will easily follow from a more abstract result, which we now formulate and prove. We are given finite sets $\mathcal{V} = \{1, \cdots, M\}$, $\mathcal{W} = \{1, \cdots, N\}$, and a subset $\mathcal{E}$ of $\mathcal{V} \times \mathcal{W}$. Also μ denotes the probability distribution on $\mathcal{V} \times \mathcal{W}$ specified by

$$\mu(v, w) \triangleq |\mathcal{E}|^{-1}, \qquad \text{for all } (v, w) \in \mathcal{E}.$$

Let $F: \mathcal{V} \times \mathcal{W} \to \{0, 1\}$ and $G = (f, g)$ with $f: \mathcal{V} \to \{1, \cdots, K\}$, $\mathcal{W} \to g\{1, \cdots, L\}$ be two functions that can be considered as random variables defined on $(\mathcal{V} \times \mathcal{W}, \mu)$.

Suppose we are interested in the value of F but can observe only G. We describe now a (decoding) function $\varphi: \{1, \cdots, K\} \times \{1, \cdots, L\} \to \{0, 1\}$ for which $\Pr(\varphi(G) \ne F)$ is minimal. Define

$$\mathcal{E}_{kl} \triangleq \mathcal{E} \cap G^{-1}(kl) \quad \text{and} \quad \mathcal{F}_{kl} \triangleq \mathcal{E}_{kl} \cap F^{-1}(1).$$

Clearly, an optimal choice of φ is

$$\varphi(k, l) = \begin{cases} 1, & \text{if } |\mathcal{F}_{kl}| \ge |\mathcal{E}_{kl}| - |\mathcal{F}_{kl}|, \\ 0, & \text{if } |\mathcal{F}_{kl}| < |\mathcal{E}_{kl}| - |\mathcal{F}_{kl}|, \end{cases} \tag{4.1}$$

for $1 \le k \le K$, $1 \le l \le L$. Its error probability $e(G, F) \triangleq \Pr(\varphi(G) \ne F)$ is given by

$$e(G, F) = \sum_{k,l} \frac{|\mathcal{E}_{k,l}|}{|\mathcal{E}|} \frac{\min\left(|\mathcal{E}_{kl}| - |\mathcal{F}_{kl}|, |\mathcal{F}_{kl}|\right)}{|\mathcal{E}_{kl}|}. \tag{4.2}$$

We use the abbreviation

$$t_{kl} \triangleq |\mathcal{E}_{kl}||\mathcal{E}|^{-1}. \tag{4.3}$$

Later we consider also a function $G': \mathcal{V} \times \mathcal{W} \to \{1, \cdots, K\} \times \{1, \cdots, L\} \times (\mathbb{N} \cup \{0\})$. To all quantities defined by means of G and having indices (k, l) there will correspond quantities defined by means of G' and having indices (k, l, m). Next replace F by a randomly chosen function Z, defined as follows. Let X_{vw} $((v, w) \in \mathcal{V} \times \mathcal{W})$ be independent identically distributed (i.i.d.) RV's with $\Pr(X_{vw} = 1) = \Pr(X_{vw} = 0) = 1/2$ and set

$$Z(v, w) = \begin{cases} 1, & \text{if } X_{vw} = 1, \\ 0, & \text{if } X_{vw} = 0 \text{ for } (v, w) \in \mathcal{V} \times \mathcal{W}. \end{cases} \tag{4.4}$$

Lemma 2: Assume that for some $\beta > 0$

$$|\mathcal{E}| \ge \max(M, N)(MN)^{2\beta}, \tag{4.5}$$

and that

$$\|G\| = \|f\|\,\|g\| \le K \cdot L \le |\mathcal{E}|(MN)^{-\beta}. \tag{4.6}$$

Then for any $\lambda \in (0, 1/8)$ there exists a constant $c_1(\lambda, \beta)$ such that for $MN \ge c_1(\lambda, \beta)$:

a) $\Pr(e(G, Z) \le \lambda) \le \exp\left\{-\frac{1}{6}\max(M, N)(MN)^{\beta}\right\}$;

b) $\Pr(e(G, Z) \le \lambda$ for some G satisfying (4.6)$)$

$$\le \exp\left\{-\frac{1}{8}\max(M, N)(MN)^{\beta}\right\}.$$

Proof: Set $Z_{kl} \triangleq \mathcal{E}_{kl} \cap Z^{-1}(1)$. Since $|Z_{kl}| = \Sigma_{(v,w)\in\mathcal{E}_{kl}} X_{vw}$, we can write

$$e(G, Z) = \sum_{k,l} t_{kl} \min\Bigg(1 - |\mathcal{E}_{kl}|^{-1} \sum_{(v,w)\in\mathcal{E}_{kl}} X_{vw},\ |\mathcal{E}_{kl}|^{-1} \cdot \sum_{(v,w)\in\mathcal{E}_{kl}} X_{vw}\Bigg).$$

Define

$$Y_{kl} \triangleq \begin{cases} 1, & \text{if } 4\lambda \le |\mathcal{E}_{kl}|^{-1} \cdot \sum_{(v,w)\in\mathcal{E}_{kl}} X_{vw} \le 1 - 4\lambda, \\ 0, & \text{else.} \end{cases} \tag{4.7}$$

Clearly,

$$e(G, Z) \ge 4\lambda \sum_{k,l} t_{kl} Y_{kl} \tag{4.8}$$

and therefore

$$\Pr(e(G, Z) \le \lambda) \le \Pr\left(\sum_{k,l} t_{kl} Y_{kl} \le \frac{1}{4}\right). \tag{4.9}$$

For an arbitrary $\alpha > 0$ and with the abbreviation

$$p_{kl} \triangleq \Pr(Y_{kl} = 0) \tag{4.10}$$

we can upperbound the right-side expression in (4.9) as follows:

$$\Pr\left(\sum_{k,l} t_{kl} Y_{kl} \le \frac{1}{4}\right) = \Pr\left(\exp\left\{-\alpha \sum_{k,l} t_{kl} Y_{kl}\right\} \ge \exp\left(-\frac{\alpha}{4}\right)\right)$$

$$\le \exp\left(\frac{\alpha}{4}\right) \prod_{k,l} \mathbf{E} \exp\{-\alpha t_{kl} Y_{kl}\}.$$

Recalling that in this paper the exponents are to the base 2, we have

$$1 - u \le \exp(-u) \le 1 - \frac{u}{2}, \qquad \text{if } 0 \le u \le 1.$$

Thus if $\alpha t_{kl} \le 1$ for every k, l, then

$$\prod_{k,l} \mathbf{E} \exp\{-\alpha t_{kl} Y_{kl}\} = \prod_{k,l} \left(p_{kl} + (1 - p_{kl}) \exp\{-\alpha t_{kl}\} \right)$$
$$\le \prod_{k,l} \left(p_{kl} + (1 - p_{kl}) \left(1 - \frac{\alpha}{2} t_{kl}\right) \right)$$
$$= \prod_{k,l} \left(1 - \frac{1}{2}(1 - p_{kl}) \alpha t_{kl}\right)$$
$$\le \exp\left\{ -\frac{\alpha}{2} \sum_{k,l} (1 - p_{kl}) t_{kl} \right\},$$

so that

$$\Pr(e(G, Z) \le \lambda) \le \Pr\left(\sum_{k,l} t_{kl} Y_{kl} \le \frac{1}{4} \right)$$
$$\le \exp\left\{ -\frac{\alpha}{2} \left(-\frac{1}{2} + \sum_{k,l} (1 - p_{kl}) t_{kl} \right) \right\},$$
$$\text{if } \max_{k,l} t_{kl} \le \alpha^{-1}. \quad (4.11)$$

This bound is good unless there are relatively large $\mathcal{E}_{kl}$. In fact, setting

$$\alpha \triangleq \max(M, N)(MN)^{\beta}, \quad (4.12)$$

assume that

$$t_{kl} = |\mathcal{E}_{kl}| |\mathcal{E}^{-1}| \le \alpha^{-1} \quad \text{for every } k, l. \quad (4.13)$$

To see that in this case (4.11) implies assertion a), notice that from (4.7) and (4.10) it follows that p_{kl} can be fairly large only if $|\mathcal{E}_{kl}|$ is very small. In particular, for a suitable constant $c = c(\lambda)$,

$$p_{kl} < \frac{1}{8}, \quad \text{whenever } |\mathcal{E}_{kl}| > c. \quad (4.14)$$

Further, by (4.6),

$$\sum_{k,l:\,|\mathcal{E}_{kl}| \le c} t_{kl} \le cKL|\mathcal{E}|^{-1} \le c(MN)^{-\beta}. \quad (4.15)$$

Since $\sum_{k,l} t_{kl} = 1$, (4.11), (4.12), and (4.15) give

$$\Pr(e(G, Z) \le \lambda) \le \exp\left\{ -\frac{\alpha}{2} \left(-\frac{1}{2} + \frac{7}{8} - c(MN)^{-\beta} \right) \right\}$$
$$\le \exp\left\{ -\frac{\alpha}{6} \right\}, \quad (4.16)$$

if $MN \ge c_1(\alpha, \beta)$, proving assertion a) (under the condition (4.13)).

If (4.13) does not hold, then the bound in (4.11) need not be valid. We can overcome this difficulty by replacing G by a suitable "refinement" G'. For this, partition each $\mathcal{E}_{kl}$ of size greater than $\alpha^{-1}|\mathcal{E}|$ into sets $\mathcal{E}_{klm}$, $0 \le m \le r_{kl}$, such that

$$c < |\mathcal{E}_{klm}| \le \alpha^{-1}|\mathcal{E}| \quad (4.17)$$

(recall that $\alpha^{-1}|\mathcal{E}| \ge (MN)^{\beta}$ by (4.5)). If $|\mathcal{E}_{kl}| \le \alpha^{-1}|\mathcal{E}|$, then write $r_{kl} \triangleq 0$, $\mathcal{E}_{kl0} \triangleq \mathcal{E}_{kl}$. Define G' by

$$G'(v, w) \triangleq \begin{cases} (k, l, m), & \text{if } (v, w) \in \mathcal{E}_{klm} \\ (k, l, 0), & \text{if } (v, w) \notin \mathcal{E} \\ & \text{and } G(v, w) = (k, l). \end{cases} \quad (4.18)$$

Then for the (decoding) function

$$\varphi'(k, l, m) \triangleq \begin{cases} 1, & \text{if } |\mathcal{F}_{klm}| \ge |\mathcal{E}_{klm}| - |\mathcal{F}_{klm}|, \\ 0, & \text{otherwise,} \end{cases} \quad (4.19)$$

obviously

$$e(G', F) \triangleq \Pr(\varphi'(G') \ne F) \le e(G, F). \quad (4.20)$$

By the very same arguments which led to (4.11), we obtain

$$\Pr(e(G, Z) \le \lambda) \le \Pr(e(G', Z) \le \lambda)$$
$$\le \exp\left\{ -\frac{\alpha}{2} \left(-\frac{1}{2} + \sum_{k,l,m} (1 - p_{klm}) t_{klm} \right) \right\}. \quad (4.21)$$

The condition $t_{klm} \le \alpha^{-1}$ is automatically met by (4.17). Since $|\mathcal{E}_{klm}| \le c$ holds only in the case when $m = 0$, $|\mathcal{E}_{kl0}| = |\mathcal{E}_{kl}| \le c$, cf. (4.17), the bound (4.15) applies also for t_{klm} instead of t_{kl}. Thus (4.20) implies (4.16) in the same way as (4.11) did. This proves a) without any additional condition.

The number of functions $G = (f, g)$ with $f: \{1, \cdots, M\} \to \{1, \cdots, M\}$ and $g: \{1, \cdots, N\} \to \{1, \cdots, N\}$ equals $M^M \cdot N^N = \exp\{M \log M + N \log N\}$. This and a) imply b). □

Remarks

1) The main reason for the Lemma to hold can best be understood in the special case $\mathcal{E} = \mathcal{W} \times \mathcal{V}$. The number of functions $F: \mathcal{W} \times \mathcal{V} \to \{0, 1\}$ equals then 2^{MN}, which is much larger than the number $N^N \cdot M^M$ of functions of type $G = (f, g)$. In the "one-dimensional" case the number of functions $F: \mathcal{W} \to \{0, 1\}$ equals 2^N, which is smaller than the number N^N of functions $f: \mathcal{W} \to \mathcal{W}$. In this case one can of course always choose $f = F$ and get exact reproduction.

2) Condition (4.5) guarantees that we are sufficiently far away from the one-dimensional case. In the terminology of [2], the pair $(\mathcal{V} \times \mathcal{W}, \{\mathcal{E}\})$ is a rectangular hypergraph with one edge $\mathcal{E}$. $G = (f, g)$ is called an orthogonal coloring. It is of type ρ_λ if in $\mathcal{E}$ at least $(1 - \lambda)|\mathcal{E}|$ different colors occur. One readily verifies that condition (4.5) implies that $\mathcal{E}$ can be partitioned into two sets $\mathcal{E}_1, \mathcal{E}_2$ with $|\mathcal{E}_2||\mathcal{E}_1|^{-1} \to 0$ $(MN \to \infty)$ such that $\mathcal{E}_1$ is of rectangular type in the sense of [2, sec. 2, pt. II]. The results there imply that (4.6) is "essentially" necessary, because otherwise $(\|G\| \ge |\mathcal{E}|(MN)^{\beta})$ there exists a $G = (f, g)$ reproducing all but a small fraction of elements in $\mathcal{E}$ exactly.

Proof of Theorem 1: For a DMDS with generic variables X, Y, consider the family of sets

$$\Omega^n_{X,Y,\eta} = \{\mathcal{T}^n_P : |P(x, y) - P_{X,Y}(x, y)| \le \eta,$$
$$\text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\}.$$

Then

$$\mathcal{T}^n_{X,Y,\eta} = \cup\{\mathcal{E}: \mathcal{E} \in \Omega^n_{X,Y,\eta}\}, \quad \text{(by (2.1))}, \quad (4.22)$$

$$P^n_{X,Y}(\mathcal{T}^n_{X,Y,\eta}) \geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{n\eta^2}, \quad \text{(by (2.4))}, \quad (4.23)$$

$$\mathcal{T}^n_{X,Y,\eta} \subset \mathcal{T}^n_{X,\eta|\mathcal{Y}|} \times \mathcal{T}^n_{Y,\eta|\mathcal{X}|}, \quad \text{(by (2.1))}, \quad (4.24)$$

and therefore $\Omega^n_{X,Y,\eta}$ is a set of subsets of $\mathcal{T}^n_{X,\eta|\mathcal{Y}|} \times \mathcal{T}^n_{Y,\eta|\mathcal{X}|}$; that is, $(\mathcal{T}^n_{X,\eta|\mathcal{Y}|} \times \mathcal{T}^n_{Y,\eta|\mathcal{X}|}, \Omega^n_{X,Y,\eta})$ is a hypergraph. Apply Lemma 2 to the triple $(\mathcal{V}, \mathcal{W}, \mathcal{E}) = (\mathcal{T}^n_{X,\eta|\mathcal{Y}|}, \mathcal{T}^n_{Y,\eta|\mathcal{X}|}, \mathcal{E})$, $\mathcal{E} \in \Omega^n_{X,Y,\eta}$. By (2.3) and the continuity of the entropy function, there exists a positive function $c(\eta)$ with $\lim_{\eta\to 0} c(\eta) = 0$ such that for $n \geq n_0(\eta)$

$$\exp\left[n(H(X) - c(\eta))\right] \leq M \leq \exp\left[n(H(X) + c(\eta))\right], \quad M \triangleq |\mathcal{V}|, \quad (4.25)$$

$$\exp\left[n(H(Y) - c(\eta))\right] \leq N \leq \exp\left[n(H(Y) + c(\eta))\right], \quad N \triangleq |\mathcal{W}|, \quad (4.26)$$

$$\exp\left[n(H(X,Y) - c(\eta))\right] \leq |\mathcal{E}|. \quad (4.27)$$

Choose η so small that

$$H(X|Y) - c(\eta) > 0, \quad H(Y|X) - c(\eta) > 0. \quad (4.28)$$

The inequalities (4.25)–(4.28) imply that for $\beta \leq \beta_0(\eta)$, sufficiently small and $n \geq n_0(\eta)$,

$$\begin{aligned} |\mathcal{E}| &\geq \exp\left[n(H(X,Y) - c(\eta))\right] \\ &\geq \exp\left[n(\max(H(X), H(Y)) \right. \\ &\qquad \left. + 2\beta(H(X) + H(Y) + 2c(\eta))\right] \\ &\geq \max(M, N)(MN)^{2\beta}, \end{aligned}$$

which is (4.5). Fix $\lambda \in (0, 1/8)$, $\delta > 0$, and choose η and β so small that

$$H(X,Y) - \delta \leq H(X,Y) - c(\eta) - \beta(H(X) + H(Y) + 2c(\eta)).$$

Since the expression to the right is smaller than $|\mathcal{E}|(MN)^{-\beta}$, Lemma 2 applied to the class of functions $\mathcal{G}^n \triangleq \{G = (f, g): (1/n)\log\|f\| + (1/n)\log\|g\| \leq H(X,Y) - \delta\}$ yields for $e(G, Z, \mathcal{E}) \triangleq e(G, Z)$,

$$\Pr\left(e(G, Z, \mathcal{E}) > \lambda, \text{ for all } G \in \mathcal{G}^n\right) \geq 1 - \exp\left[-\frac{1}{8}\max(M, N)(MN)^{\beta}\right], \quad (4.29)$$

and therefore also

$$\Pr\left(e(G, Z, \mathcal{E}) > \lambda, \text{ for all } \mathcal{E} \in \Omega^n_{X,Y,\eta} \text{ and all } G \in \mathcal{G}^n\right) \geq 1 - (n+1)^{|\mathcal{X}||\mathcal{Y}|}\exp\left[-\frac{1}{8}\max(M, N)(MN)^{\beta}\right].$$

Since the right side expression is obviously positive for $n > n_1(\eta, \beta)$ suitable, there exists a function $F_n: \mathcal{T}^n_{X,|\mathcal{Y}|\eta} \times \mathcal{T}^n_{Y,|\mathcal{X}|\eta} \to \{0, 1\}$ with

$$e(G, F_n, \mathcal{E}) > \lambda, \quad \text{for all } G \in \mathcal{G}^n, \ \mathcal{E} \in \Omega^n_{X,Y,\eta}. \quad (4.30)$$

Define now $F: \cup_{n=1}^{\infty}(\mathcal{X}^n \times \mathcal{Y}^n) \to \{0, 1\}$ by

$$F(x, y) \triangleq \begin{cases} 0, & \text{for } (x, y) \in \bigcup_{n=1}^{n_1}(\mathcal{X}^n \times \mathcal{Y}^n), \\ F_n(x, y), & \text{for } (x, y) \in \mathcal{T}^n_{X,Y,\eta}, \ n > n_1, \\ 0, & \text{otherwise.} \end{cases} \quad (4.31)$$

Equations (4.30), (4.22), and (4.23) imply that the error probability is bounded away from zero, because the sets in $\Omega^n_{X,Y,\eta}$ are disjoint and their elements are equiprobable. We have thus seen that rates R_1, R_2 with $R_1 + R_2 \leq H(X,Y) - \delta$ are too small if $R_1 \leq H(X) + c(\eta)$ and $R_2 \leq H(Y) + c(\eta)$. There is, however, no point in choosing for instance $R_1 > H(X) + c(\eta)$, because the projection of $\mathcal{T}^n_{X,Y,\eta}$ on $\mathcal{V}$ has rate less than $H(X) - c(\eta)$. □

Remark: From here it is just an exercise to show the existence of a binary (universal) function F such that for *every* DMDS with $H(X|Y) > 0$, $H(Y|X) > 0$ one needs encoding rates as in (1.1) in order to decode F with small error probability.

V. Proof of the Results on F-Codes for Sensitive Functions F

We have seen in Section III that Theorem 3 is a consequence of Lemma 1 and Theorem 4, while Theorem 2 is a special case of Theorem 3. Thus we have to prove Lemma 1 and Theorem 4.

Proof of Lemma 1: Consider an arbitrary source F-code (f, g, φ) of block length n. Our first claim is that it can be replaced by another F-code (f', g', φ') such that

$$e(f, g, \varphi; F) = e(f', g', \varphi'; F) \quad (5.1)$$

where f' maps $\mathcal{X}^n$ into $\{1, \cdots, N\}$ with

$$N \leq \|f\|\exp\left(\frac{n\delta}{2}\right), \quad (5.2)$$

each set

$$\mathcal{C}_i \triangleq \{x: f'(x) = i\}, \quad (i = 1, \cdots, N),$$

consisting of sequences of the same type and satisfying

$$d(x, x') \geq \tau n, \quad \text{if } x \neq x' \text{ are in the same } \mathcal{C}_i, \quad (5.3)$$

whereas g' is the identity mapping on $\mathcal{Y}^n$. To verify this, denote by $V_n(|\mathcal{X}|, \tau)$ the cardinality of a Hamming sphere in $\mathcal{X}^n$ with radius τn (and arbitrary center $x_0 \in \mathcal{X}^n$), i.e.,

$$V_n(|\mathcal{X}|, \tau) \triangleq |\{x: d(x, x_0) \leq \tau n\}|.$$

Then the sequences $x \in \mathcal{X}^n$ can be partitioned into $V_n(|\mathcal{X}|, \tau)$ classes in a successive manner so that each $x \in \mathcal{X}^n$ is put into a class different from the at most $V_n(|\mathcal{X}|, \tau) - 1$ classes which already contain sequences of Hamming distance less than τn from x. Subpartitioning each of these classes according to types we get, cf. (2.2), a partition of $\mathcal{X}^n$ into at most

$$M \triangleq (n+1)^{|\mathcal{X}|}V_n(|\mathcal{X}|, \tau) \quad (5.4)$$

subsets, each consisting of sequences of the same type and

neither containing distinct sequences of Hamming distance less than τn.

Now let $(\mathcal{Q}_1, \mathcal{Q}_1, \cdots, \mathcal{Q}_N)$ be the coarsest joint refinement of the latter partition of $\mathcal{X}^n$ and of the one defined by the encoder f. Since clearly

$$V_n(|\mathcal{X}|, \tau) \le \exp\left(\frac{n\delta}{3}\right),$$

if τ is sufficiently small (depending only on $|\mathcal{X}|$ and δ), it follows from (5.4) that (5.2) and (5.3) are valid for this τ if n is sufficiently large. Defining

$$f'(x) \triangleq i, \qquad \text{if } x \in \mathcal{Q}_i, \quad g'(y) \triangleq y,$$

it is obvious that to the encoders f' and g' there exists a decoder φ' satisfying (5.1). This establishes our first claim.

Since each $\mathcal{Q}_i$ consists of sequences of the same type, $P_X^n(x)$ is constant on each $\mathcal{Q}_i$, and thus

$$\begin{aligned} &e(f', g', \varphi'; \mathcal{F}) \\ &= \sum_{x \in \mathcal{X}^n} P_X^n(x) \Pr\{\varphi'(f'(x), Y^n) \ne F(x, Y^n) | X^n = x\} \\ &= \sum_{i=1}^{N} P_X^n(\mathcal{Q}_i) \frac{1}{|\mathcal{Q}_i|} \sum_{x \in \mathcal{Q}_i} W^n\big(\{y: \varphi'(i, y) \ne F(x, y)\} | x\big) \end{aligned}$$

Denoting by $\mathcal{J}$ the set of those indices $1 \le i \le N$ for which

$$\frac{1}{|\mathcal{Q}_i|} \sum_{x \in \mathcal{Q}_i} W^n\big(\{y: \varphi'(i, y) \ne F(x, y)\} | x\big) \le 2e(f', g', \varphi'; F), \quad (5.5)$$

it follows that

$$P_X^n\Big(\bigcup_{i \in \mathcal{J}} \mathcal{Q}_i\Big) = \sum_{i \in \mathcal{J}} P_X^n(\mathcal{Q}_i) \ge \frac{1}{2}.$$

For any fixed $\eta > 0$, this implies by (2.4) that

$$P_X^n\Big(\bigcup_{i \in \mathcal{J}} \mathcal{Q}_i \cap \mathcal{T}_{X,\eta}^n\Big) \ge \frac{1}{4}$$

provided that $n \ge (|\mathcal{X}|/4\eta^2)$. In turn, one gets by (2.5) that

$$\begin{aligned} \sum_{i \in \mathcal{J}, \mathcal{Q}_i \subset \mathcal{T}_{X,\eta}^n} |\mathcal{Q}_i| &= \Big|\Big(\bigcup_{i \in \mathcal{J}} \mathcal{Q}_i\Big) \cap \mathcal{T}_{X,\eta}^n\Big| \\ &\ge \exp\left\{n\left(H(X) - \frac{\delta}{2}\right)\right\} \\ &\qquad \text{if } n \ge n_0 = n_0(|\mathcal{X}|, \eta, \delta). \end{aligned}$$

Consider an $i_0 \in \mathcal{J}$ for which $|\mathcal{Q}_i|$ is maximal subject to $\mathcal{Q}_i \subset \mathcal{T}_{X,\eta}^n$. Then the last inequality and (5.2) result in

$$\begin{aligned} \frac{1}{n}\log|\mathcal{Q}_{i_0}| &\ge \frac{1}{n}\log\left[\frac{1}{N}\exp\left\{n\left(H(X) - \frac{\delta}{2}\right)\right\}\right] \\ &\ge H(X) - \frac{1}{n}\log\|f\| - \delta. \end{aligned} \quad (5.6)$$

Now a channel F-code $(\mathcal{C}, \psi)$ with the required properties (3.8)–(3.11) can be given by $\mathcal{C} \triangleq \mathcal{Q}_{i_0}$, $\psi(y) \triangleq \varphi'(i_0, y)$. In fact, (5.6) means that this $\mathcal{C}$ satisfies (3.9), whereas (3.8) follows from (5.5) and (5.1), for $\bar{e}(\mathcal{C}, \psi; F)$ is just the left side of (5.5) if $i = i_0$. Of course (3.10) and (3.11) are valid for $\mathcal{C}$ by construction. □

For the proof of Theorem 4 we require two simple lemmas.

Lemma 3: Let $\mathcal{Q}$ be any subset of $\mathcal{X}^n$ with the property that $d(x, x') \ge \tau n$ for every $x \ne x'$ in $\mathcal{Q}$. Let P be an arbitrary PD on $\mathcal{X}^n$ and denote by $P_i(x)$ the P-probability of the set of those sequences $x \in \mathcal{Q}$ which have ith component x. Then

$$\frac{1}{n}\sum_{i=1}^{n} \sum_{x \ne x'} P_i(x)P_i(x') \ge P(\mathcal{Q})\Big(P(\mathcal{Q}) - \max_{x \in \mathcal{Q}} P(x)\Big)\tau. \quad (5.7)$$

Remark: The special case of inequality (5.7) when P is the uniform distribution on $\mathcal{Q}$, is familiar from the derivation of Plotkin's [7] bound, cf. Berlekamp [3, p. 311.]

Proof: Consider two independent random sequences $X^n = X_1 \cdots X_n$ and $\tilde{X}^n = \tilde{X}_1 \cdots \tilde{X}_n$ both having distribution P, and define

$$Z_i \triangleq \begin{cases} 1, & \text{if } X^n \in \mathcal{Q}, \tilde{X}^n \in \mathcal{Q}, X_i \ne \tilde{X}_i, \\ 0, & \text{else.} \end{cases}$$

Then

$$\sum_{i=1}^{n} Z_i = \begin{cases} d(X^n, \tilde{X}^n), & \text{if } X^n \in \mathcal{Q}, \tilde{X}^n \in \mathcal{Q}, \\ 0, & \text{else,} \end{cases}$$

and therefore

$$\begin{aligned} &\sum_{i=1}^{n} \sum_{x \ne x'} P_i(x)P_i(x') \\ &= E\sum_{i=1}^{n} Z_i \ge \tau n \Pr\{X^n \in \mathcal{Q}, \tilde{X}^n \in \mathcal{Q}, X^n \ne \tilde{X}^n\}. \end{aligned} \quad (5.8)$$

But

$$\begin{aligned} \Pr\{X^n \in \mathcal{Q}, \tilde{X}^n \in \mathcal{Q}, X^n \ne \tilde{X}^n\} &= \sum_{x \in \mathcal{Q}} P(x) \sum_{x' \in \mathcal{Q} \setminus \{x\}} P(x') \\ &\ge P(\mathcal{Q})\Big(P(\mathcal{Q}) - \max_{x \in \mathcal{Q}} P(x)\Big). \end{aligned}$$

Thus (5.8) gives (5.7). □

Lemma 4: Consider a directed graph with vertex set $\mathcal{V}$. Denote by $d_i(v)$ the $d_o(v)$ resp. in- resp. out-degree of a vertex $v \in \mathcal{V}$, i.e., the number of edges leading to resp. starting from v. Let the vertices $v \in \mathcal{V}$ have weights $\mu(v) \ge 0$ such that for some $\xi > 0$,

$$\mu(v_1) \le \xi\mu(v_2) \quad (5.9)$$

if there is an edge from v_1 to v_2. Then if $d_i(v) \le k$ for every $v \in \mathcal{V}$, and $d_o(v) \ge l$ whenever $d_o(v) > 0$, we have

$$\sum_{v: d_o(v) > 0} \mu(v) \le \frac{\xi k}{l} \sum_{v \in \mathcal{V}} \mu(v). \quad (5.10)$$

Proof: Summing the inequalities (5.9) for all edges, we get

$$\sum_{v \in \mathcal{V}} d_o(v)\mu(v) \le \xi \sum_{v \in \mathcal{V}} d_i(v)\mu(v).$$

By the assumption on the degrees this yields

$$l \sum_{v:\, d_o(v)>0} \mu(v) \le \xi k \sum_{v \in \mathcal{V}} \mu(v).$$

□

A sketch of the proof of Theorem 4 is the following. Given a "good" channel F-code $(\mathcal{C}, \psi)$, we shall consider the ordinary channel code $(\mathcal{C}, \tilde{\psi})$ where $\tilde{\psi}$ is a maximum likelihood decoder. Supposing that the codewords $x \in \mathcal{C}$ are equiprobable, we shall look at the (posterior) probabilities $e(y)$ resp. $\tilde{e}(y)$ that ψ resp. $\tilde{\psi}$ makes an error if y is the received sequence. The main idea will be that if $e(y)$ is small but $\tilde{e}(y)$ is not, then sufficiently many sequences y' with $d(y, y') = 1$ can be found such that either y has much smaller probability than y' or else $e(y')$ is large and the probability of y is not much larger than that of y'. Then Lemma 4 will enable us to conclude that the set of sequences $y \in \mathcal{Y}^n$ with large $\tilde{e}(y)$ must have a small probability.

Formal Proof of Theorem 4: Consider an arbitrary channel F-code $(\mathcal{C}, \psi)$ as in Theorem 4, i.e., satisfying (3.12), (3.13), with $0 < \epsilon \le 1/4$ to be specified later. Define a PD $\tilde{Q}$ on $\mathcal{C} \times \mathcal{Y}^n$ by

$$\tilde{Q}(x, y) \triangleq \frac{1}{|\mathcal{C}|} W^n(y|x) \qquad x \in \mathcal{C}, y \in \mathcal{Y}^n \tag{5.11}$$

and denote by $Q(y)$ resp. $Q(x|y)$ the corresponding marginal resp. conditional probabilities, i.e.,

$$Q(y) \triangleq \frac{1}{|\mathcal{C}|} \sum_{x \in \mathcal{C}} W^n(y|x), \tag{5.12}$$

$$Q(x|y) \triangleq \frac{\tilde{Q}(x, y)}{Q(y)} = \frac{W^n(y|x)}{\sum_{x' \in \mathcal{C}} W^n(y|x')}. \tag{5.13}$$

Let $\tilde{\psi}: \mathcal{Y}^n \to \mathcal{C}$ be a maximum likelihood decoder, i.e., let $\tilde{\psi}(y)$ equal an $x \in \mathcal{C}$ maximizing $Q(x|y)$. We denote by $e(y)$ resp. $\tilde{e}(y)$ the $Q(\cdot|y)$-probability that ψ resp. $\tilde{\psi}$ makes an error, i.e.,

$$e(y) \triangleq Q\big(\{x: \psi(y) \ne F(x, y)\}|y\big), \tag{5.14}$$

$$\tilde{e}(y) \triangleq Q\big(\{x: \tilde{\psi}(y) \ne x\}|y\big) = 1 - \max_{x \in \mathcal{C}} Q(x|y). \tag{5.15}$$

Then by definition,

$$\bar{e}(\mathcal{C}, \psi; F) = \tilde{Q}\big(\{(x, y): \psi(y) \ne F(x, y)\}\big) = \sum_{y \in \mathcal{Y}^n} Q(y)\tilde{e}(y). \tag{5.16}$$

$$\bar{e}(\mathcal{C}, \tilde{\psi}) = \tilde{Q}\big(\{(x, y): \psi(y) \ne x\}\big) = \sum_{y \in \mathcal{Y}^n} Q(y)\tilde{e}(y). \tag{5.17}$$

On account of the assumption $\bar{e}(\mathcal{C}, \psi; F) < \epsilon$, the identity (5.16) implies that the set

$$\mathcal{B} \triangleq \big\{y: e(y) \le \sqrt{\epsilon}\big\} \tag{5.18}$$

has probability

$$Q(\mathcal{B}) \ge 1 - \sqrt{\epsilon}. \tag{5.19}$$

Now consider an arbitrary $y \in \mathcal{B}$ for which

$$\tilde{e}(y) \ge e(y) + \delta, \tag{5.20}$$

where $\delta > 0$ will be specified later. We claim that there exist at least $(\delta\tau n/4)$ indices $i \in \{1, \cdots, n\}$ such that changing just the ith component of y, one can get sequences y' for which either

$$Q(y) \le \xi Q(y'), \tag{5.21}$$

or

$$y' \notin \mathcal{B}, \quad Q(y) \le \xi \frac{Q(y')}{2\sqrt{\epsilon}}, \tag{5.22}$$

where ξ is a small positive number to be specified later. To establish this, we apply Lemma 3 to the set

$$\mathcal{A}(y) \triangleq \{x: \psi(y) = F(x, y)\} \subset \mathcal{X}^n, \tag{5.23}$$

and the distribution $Q(\cdot|y)$ on $\mathcal{X}^n$.

Notice that by (5.14), (5.18), and the assumption $\epsilon \le 1/4$, we have

$$Q(\mathcal{A}(y)|y) = 1 - e(y) \ge 1 - \sqrt{\epsilon} \ge \frac{1}{2};$$

further, using (5.20) and (5.15),

$$Q(\mathcal{A}(y)|y) - \max_{x \in \mathcal{C}} Q(x|y) = 1 - e(y) - (1 - \tilde{e}(y)) \ge \delta.$$

Thus Lemma 3 gives for

$$P_i(x) \triangleq Q\big(\{x: x \in \mathcal{A}(y), i\text{th component of } x \text{ is } x\}|y\big) \tag{5.24}$$

that

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{x \ne x'} P_i(x)P_i(x') \ge \frac{\delta\tau}{2}.$$

Let $\mathcal{J}(y)$ be the set of those indices $i \in \{1, \cdots, n\}$ for which

$$\sum_{x \ne x'} P_i(x)P_i(x') \ge \frac{\delta\tau}{4}. \tag{5.25}$$

Then the previous inequality implies

$$|\mathcal{J}(y)| \ge \frac{\delta\tau n}{4}. \tag{5.26}$$

Fix an arbitrary $i \in \mathcal{J}(y)$. By (5.25), there exist elements $x_1 \ne x_2$ of $\mathcal{X}$ such that

$$P_i(x_1)P_i(x_2) \ge \eta, \qquad \text{where } \eta \triangleq \frac{\delta\tau}{4|\mathcal{X}|^2}. \tag{5.27}$$

Let $\mathcal{A}_1$ resp. $\mathcal{A}_2$ denote the set of those sequences $x \in \mathcal{A}(y)$ the ith component of which is x_1 resp. x_2. Since $\mathcal{A}_1$ and $\mathcal{A}_2$ are subsets of $\mathcal{A}(y)$, we have

$$F(x, y) = F(x', y), \qquad \text{whenever } x \in \mathcal{A}_1, \ x' \in \mathcal{A}_2. \tag{5.28}$$

Further, (5.27) and (5.24) give that

$$Q(\mathcal{A}_1|y) \geq \eta, \quad Q(\mathcal{A}_2|y) \geq \eta. \tag{5.29}$$

Denote the ith element of y by y_1. By (5.28) and our assumption that F is sensitive, cf. Definition 2, there exists a $y_2 \in \mathcal{Y}$ such that for the sequence $y' \in \mathcal{Y}^n$ obtained from y by replacing its ith component by y_2, we have

$$F(x, y') \neq F(x', y'), \qquad \text{whenever } x \in \mathcal{A}_1,\ x' \in \mathcal{A}_2. \tag{5.30}$$

This means, in particular—cf. (5.14)—that

$$e(y') \geq \min\big(Q(\mathcal{A}_1|y'), Q(\mathcal{A}_2|y')\big). \tag{5.31}$$

Now suppose that for the present y' (5.21) does not hold. Then writing $W_{\min} \triangleq \min_{x,y} W(y|x)$, from (5.13) and (5.29) we obtain

$$Q(\mathcal{A}_1|y') = \frac{\sum_{x\in\mathcal{A}_1} W^n(y'|x)}{|\mathcal{C}|Q(y')} > \frac{\xi \sum_{x\in\mathcal{A}_1} W^n(y|x) W_{\min}}{|\mathcal{C}|Q(y)}$$
$$= \xi W_{\min} Q(\mathcal{A}_1|y) \geq \xi\eta W_{\min},$$

and similarly

$$Q(\mathcal{A}_2|y') > \xi\eta W_{\min}.$$

Now choose

$$\xi \triangleq \frac{\sqrt{\epsilon}}{\eta W_{\min}}. \tag{5.32}$$

With this ξ, upon substituting the last bounds into (5.31) we get that if (5.21) does not hold, then $e(y') > \sqrt{\epsilon}$, i.e., $y' \notin \mathcal{B}$. Further, since $Q(\mathcal{A}_1 \cup \mathcal{A}_2|y) \geq 2\eta$ by (5.29), using (5.12) and (5.13) we also have

$$Q(y) \leq \frac{Q(\mathcal{A}_1 \cup \mathcal{A}_2|y)}{2\eta} Q(y) = \frac{\sum_{x\in\mathcal{A}_1\cup\mathcal{A}_2} W^n(y|x)}{2\eta|\mathcal{C}|}$$
$$\leq \frac{\sum_{x\in\mathcal{A}_1\cup\mathcal{A}_2} W^n(y'|x)}{2\eta|\mathcal{C}|W_{\min}} \leq \frac{\sum_{x\in\mathcal{C}} W^n(y'|x)}{2\eta|\mathcal{C}|W_{\min}} = \frac{Q(y')}{2\eta W_{\min}}.$$

On account of the choice (5.32), this proves the inequality in (5.22).

Thus we have established our claim stated in the paragraph containing (5.20) for the case when F is sensitive and $W(y|x) > 0$ for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$. If F is highly sensitive, then replacing the ith component y_1 of y by any $y_2 \neq y_1$, for the resulting y' (5.30) holds. If W satisfies the weaker hypothesis postulated in Theorem 4 for the case when F is highly sensitive, this y_2 can be chosen so that $W(y_2|x_1)$ and $W(y_2|x_1)$ are both positive (for $W(y_1|x_1)$ and $W(y_1|x_2)$ are positive by (5.29)). Then the previous argument applies word for word, with the only difference that now $W_{\min}$ should stand for the smallest positive element of the matrix W.

To complete the proof of Theorem 4 we apply Lemma 4 to the graph with vertex set $\mathcal{Y}^n$, drawing an edge from y to y' if and only if $y \in \mathcal{B}$, y satisfies (5.20), and y' is obtained from y in the same way as above for some index $i \in \mathcal{J}(y)$. Let the weights be defined by

$$\mu(y) \triangleq \begin{cases} Q(y), & \text{if } y \in \mathcal{B}, \\ \dfrac{1}{2\sqrt{\epsilon}} Q(y), & \text{if } y \notin \mathcal{B}. \end{cases} \tag{5.33}$$

Then (5.9) holds according to (5.21), (5.22), and the out-degree of each $y \in \mathcal{B}$ satisfying (5.20) is at least $\delta\tau n/4$ by (5.26). Of course, every $y \in \mathcal{Y}^n$ has in-degree less than $n|\mathcal{Y}|$. Thus (5.10) gives

$$Q\big(\{y: y \in \mathcal{B}, \tilde{e}(y) \geq e(y) + \delta\}\big) \leq \frac{\xi n|\mathcal{Y}|}{\delta\tau n/4}\left(1 + \frac{Q(\mathcal{Y}^n \setminus \mathcal{B})}{2\sqrt{\epsilon}}\right) \leq \frac{6\xi|\mathcal{Y}|}{\delta\tau},$$

where the second inequality follows from (5.19). Substituting the value

$$\xi = \frac{\sqrt{\epsilon}}{\eta W_{\min}} = \frac{4|\mathcal{X}|^2\sqrt{\epsilon}}{\delta\tau W_{\min}},$$

cf. (5.32), (5.27), we get

$$Q\big(\{y: y \in \mathcal{B}, \tilde{e}(y) \leq e(y) + \delta\}\big) \leq K\frac{\sqrt{\epsilon}}{\delta^2}, \qquad \text{with } K \triangleq \frac{24|\mathcal{X}|^2|\mathcal{Y}|}{\tau^2 W_{\min}}. \tag{5.34}$$

On account of (5.16), (5.17), and (5.19), this results in

$$\tilde{e}(\mathcal{C}, \tilde{\psi}) = \sum_{y\in\mathcal{Y}^n} Q(y)\tilde{e}(y)$$
$$\leq Q\big(\{y: \tilde{e}(y) \geq e(y) + \delta\}\big) + \sum_{y\in\mathcal{Y}^n} Q(y)(e(y) + \delta)$$
$$\leq K\frac{\sqrt{\epsilon}}{\delta^2} + \sqrt{\epsilon} + \bar{e}(\mathcal{C}, \psi; F) + \delta.$$

Thus, choosing the so far unspecified $\delta > 0$ as $\delta \triangleq K^{1/3}\epsilon^{1/6}$, we have proved that

$$\tilde{e}(\mathcal{C}, \tilde{\psi}) \leq 2K^{1/3}\epsilon^{1/6} + \epsilon^{1/2} + \epsilon, \qquad \text{if } \bar{e}(\mathcal{C}, \psi; F) \leq \epsilon, \tag{5.35}$$

where K is the constant of (5.34). □

VI. Discussion

For a discrete memoryless double source we have shown that in order for a decoder to determine a function

$F(X^n, Y^n)$ of the length-n messages of the component sources, for "most"—even binary valued—functions F, the encoders of the component sources must have as large rates as if (X^n, Y^n) were to be determined. Next, we have considered specific functions F such as the joint type or the Hamming distance of X^n and Y^n or just the parity of the Hamming distance. For a class of functions F containing the mentioned ones we have shown that for determining $F(X^n, Y^n)$ in the knowledge of Y^n, the X-encoder typically must have as large a rate as for determining X^n itself. This implies that all achievable F rate pairs must satisfy

$$R_1 \geq H(X|Y), \quad R_2 \geq H(Y|Y).$$

The problem of describing the achievable F rate region (for either F in the mentioned class) remains open unless the double source has independent components or is binary symmetric (in the latter cases the above necessary conditions are also sufficient).

Notice that the problem of F-codes is outside the usual framework of rate-distortion theory except for "componentwise" functions F, cf. (1.4). Still, a complete description of the achievable F rate region, e.g., for $F(x, y) \triangleq P_{x,y}$, may be as hard a problem as to determine the achievable rate region for reproducing X^n, Y^n within a prescribed distortion measure. We draw attention to the fact that in analogy to our Theorem 3, for the latter problem it is also the projection of the achievable rate region to the R_1-axis which could be determined (Wyner–Ziv, [10]).

As a tool for proving our results concerning F-codes we have considered also channel F-codes. From a mathematical point of view the latter are natural counterparts of source F-codes. One could also define the F-capacity of a DMC in an obvious manner. Under the conditions of Theorem 4 one then easily gets the corollary that the F-capacity equals the ordinary capacity. We do believe that Theorem 4 and this corollary are also of independent interest and that communication situations can be found for which the concept of F-capacity and these results are relevant. This, however, still remains to be explored.

References

[1] R. Ahlswede, "Channel capacities for list codes," *J. Appl. Prob.*, vol. 10, pp. 824–836, 1973.

[2] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding," pt. I, *J. Comb. Inform. Syst. Sci.*, vol. 4, no. 1, pp. 76–115, 1979, pt. II, *J. Comb. Inform. Syst. Sci.*, vol. 5, no. 3, pp. 220–268, 1980.

[3] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[5] J. Körner, "Some methods in multi-user communication: a tutorial survey," *Information Theory, New Trends and Open Problems*, G. Longo ed., CISM courses and lectures no. 219. Wien: Springer, 1975.

[6] J. Körner, personal communication.

[7] M. Plotkin, "Binary codes with specified minimum distance," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 445–450, 1960.

[8] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding in discrete memoryless channels," pt. I, II, *Inform. Contr.*, vol. 10, pp. 65–103, 522–552, 1967.

[9] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, 1973.

[10] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–1, 1976.