On Source Coding with Side Information via a Multiple-Access Channel and Related Problems in Multi-User Information Theory

RUDOLF AHLSWEDE AND TE SUN HAN, MEMBER, IEEE

Abstract—A simple proof of the coding theorem for the multiple-access channel (MAC) with arbitrarily correlated sources (DMCS) of Cover–El Gamal–Salehi, which includes the results of Ahlswede for the MAC and of Slepian–Wolf for the DMCS and the MAC as special cases, is first given. A coding theorem is introduced and established for another type of source-channel matching problem, i.e., a system of source coding with side information via a MAC, which can be regarded as an extension of the Ahlswede–Körner–Wyner type noiseless coding system. This result is extended to a more general system with several principal sources and several side information sources subject to cross observation at the encoders in the sense of Han. The regions are shown to be optimal in special situations. Dueck's example shows that this is in general not the case for the result of Cover–El Gamal–Salehi and the present work. In another direction, the achievable rate region for the modulo-two sum source network found by Körner–Marton is improved. Finally, some ideas about a new approach to the source-channel matching problem in multi-user communication theory are presented. The basic concept is that of a correlated channel code. The approach leads to several new coding problems.

I. Introduction

It is well-known that Shannon's paper [1] was the starting point of multi-user information theory and it still seems, at least to us, that this paper is not given the attention that it deserves. In our judgment one of the most important problems raised is that of transmitting correlated messages over a noisy channel with two (or more) senders. This problem is still far from being completely understood. For the ease of our later reference and discussions we repeat here what Shannon wrote in the section *Attainment of the Outer Bound with Dependent Sources* [1, sect. 14, p. 636] (the numbers were inserted in the text by the authors for later reference):

> With regard to the outer bound there is an interesting interpretation relating to a somewhat more general communication system. Suppose that the message sources at the two ends of our channel are not independent but statistically dependent. Thus, one might be sending weather information from Boston to New York and from New York to Boston. The weather at these cities is of course not statistically independent. 1) *If the dependence were of just the right type for the channel or if the messages could be transformed so that this were the case, then it may be possible to attain transmission at the rates given by the outer bound.* For example, in the multiplying channel just discussed, suppose that the messages at the two ends consist of streams of binary digits which occur with the dependent probabilities given by Table III. Successive x_1, x_2 pairs are assumed independent. Then by merely sending these streams into the channel (without processing) the outer bound curve is achieved at its midpoint.
>
> It is not known whether this is possible in general. 2) Does there always exist a suitable pair of dependent sources that can be coded to give rates R_1, R_2 within ϵ of any point in the outer bound. 3) This is at least often possible in the noiseless memoryless case, that is, when y_1 and y_2 are strict functions of x_1 and x_2 (no channel noise). 4) The source pair defined by the assignment $p\{x_1, x_2\}$ that produces the point in question is often suitable in such a case without coding as in the above example.

Now, in Shannon's notation, x_i (resp. y_i) are the inputs (resp. outputs) at terminal i ($i = 1, 2$) for the two-way channel (TWC) and the outer bound G_0 is the set of all pairs (R_1, R_2) with

$$R_1 = I(X_1; Y_2|X_2), \quad R_2 = I(X_2; Y_1|X_1), \quad (1.1)$$

where X_1, X_2 are dependent input variables and Y_1, Y_2 the output variables induced by the channel.

Shannon showed that the inner bound G_I, defined as convex hull of the set of rate pairs obtained in (1.1) for independent X_1, X_2, is an achievable rate region in the case of *independent messages*. Recently Dueck [2] showed that G_I is in general not the capacity region $\mathcal{C}$ in the case of independent messages and $\mathcal{C}$ is still not known.

New progress in multi-user communication began by considering a simpler channel model, namely, that of a

Manuscript received June 25, 1980; revised September 22, 1982. This work was partially presented at the IEEE International Symposium on Information Theory, Santa Monica, CA, 1981.

R. Ahleswede is with Department of Mathematics, University of Bielefeld, 4800 Bielefeld 1, West Germany.

T. Sun Han is with the Department of Information Science, Sagami Institute of Technology, Tsujido Nishikaigan 1-1-25, Fujisawa, Japan 251.

0018-9448/83/0500-0396$01.00

multiple-access channel (MAC; also mentioned in [1]). Ahlswede [3] was the first to establish the capacity region of this channel in case of *independent messages* and subsequently Slepian and Wolf [4] found the region of this channel for the following situation:

a) The *correlated* message statistic is described by a correlated memoryless source;
b) the MAC is *noiseless*.

They also found the region for an arbitrary MAC and a certain special correlated message statistic in [5]. The result of [4] can also be viewed as solution to sentence 3) mentioned in the quote. In case of a noiseless TWC formula (1.1) gives

$$R_1 = H(X_1|X_2), \quad R_2 = H(X_2|X_1). \qquad (1.2)$$

Notice that by the result of [4] sender 2, who knows X_2, can be informed about X_1 with a rate $H(X_1|X_2)$, and sender 1, who knows X_1, can be informed about X_2 with a rate $H(X_2|X_1)$.

Considering again the MAC, which is better understood than the TWC because the feedback problem is not present, Cover, El Gamal, and Salehi [6] recently found a way of using the dependency structure of the correlated message source for the channel coding and thus obtained a general coding theorem for the MAC which includes as special cases the results of [3], [4], [5]. (The close connections between the results of [3] and [4] are explained for instance in [7, part I, sect. 6].)

Dueck recently showed [8] that the approach of [6] does not always give the full capacity region; it only does if the dependency structure of the source fits nicely with the dependency structure of the channel.

It seems that Shannon gave not only the direction but also understood the situation quite well in the remarkable sentence 3) of the quote. In the last Section VII of this paper we will present some results and ideas, which we hope will be helpful in making some further progress in the direction indicated in sentence 1).

But the larger part of this paper (Sections II, III, and IV) relates to the approach to the correlated source-multiple-access channel matching problem given by Cover, El Gamal, and Salehi [6]. They introduced an interesting coding technique based on a kind of *correlation-preserving mapping*.

In Section II we look at their coding theorem from the viewpoint of *cross observation* at the encoders (Han [9], [10]), revealing that the heart of their theorem consists in a simpler but elegant version and that the theorem itself has a simple proof.

In Section III we introduce another type of source-channel matching problem, i.e., a system of source coding with side information via a multiple-access channel, which may be regarded as an extension of the Ahlswede–Körner–Wyner type noiseless source coding system ([11], [12]). For this system we establish a matching condition.

In Section IV we consider a more general system with several principal sources and several side information sources subject to cross observation at the encoders, and we establish a sufficient condition for the matching of this system.

In Section V we present an achievable rate region for a channel with side information at the decoder, which is perhaps one of the simplest cases for which the converse is presently not known.

In Section VI we describe a new achievable rate region for the modulo-two sum source network considered by Körner and Marton [13], which is a special but very instructive case of the general two-helper source network introduced in [24].

Even though most of our results are (except for special cases) incomplete in the sense that no converses are proved, we still feel that the results obtained, the problems proposed, and the handy formalism provided will be of some benefit for the advancement of the subject of multi-user communication theory.

II. A New Look at and Simple Derivation of the Cover–El Gamal–Salehi Coding Theorem

We establish first the fairly simple theorem 1 below and then show that it contains the result of [6] as a special case. Consider a memoryless multiple-access channel (MAC) M_2 with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ (finite), an output alphabet $\mathcal{Y}$ (finite), and the transmission probabilities

$$w(y|x_1, x_2), \qquad \text{for } y \in \mathcal{Y},\ x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2.$$

Let $S = (S_1, S_2, S_3)$ be a multiple information source, where S_1, S_2, S_3 are arbitrarily *correlated* random variables with values in finite sets $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$, respectively. Denote by S_i^n an independent identically distributed (i.i.d.) n-sequence of S_i $(i = 1, 2, 3)$.

Let us consider the following joint system of the source S and the channel M_2 with cross observation at the encoders (Fig. 1). The encoder ϕ_1 observes a pair (S_1^n, S_2^n) and maps it to an input n-sequence X_1^n:

$$\phi_1: \mathcal{S}_1^n \times \mathcal{S}_2^n \to \mathcal{X}_2^n.$$

Similarly, the encoder ϕ_2 observes (S_2^n, S_3^n) and maps it to an input n-sequence X_2^n:

$$\phi_2: \mathcal{S}_2^n \times \mathcal{S}_3^n \to \mathcal{X}_2^n.$$

The decoder ψ observes an output n-sequence Y^n and maps it to an element $(\hat{S}_1^n, \hat{S}_2^n, \hat{S}_3^n)$:

$$\psi: \mathcal{Y}^n \to \mathcal{S}_1^n \times \mathcal{S}_2^n \times \mathcal{S}_3^n.$$

The probability of error P_e is given by

$$P_e = \Pr\{\hat{S}_1^n\hat{S}_2^n\hat{S}_3^n \neq S_1^nS_2^nS_3^n\}.$$

Definition 1: The source $S = (S_1, S_2, S_3)$ is said to be *admissible* for the channel M_2 if for any $0 < \lambda < 1$ and sufficiently large n there exist encoding functions ϕ_1, ϕ_2, and a decoding function ψ for which $P_e < \lambda$.

In order to obtain a sufficient condition for admissibility, it will be convenient to consider an associated test channel as follows (cf. Han [9]).

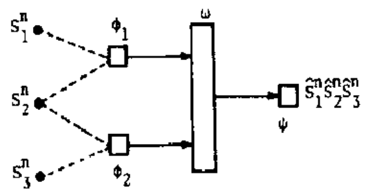

Fig. 1. Joint system of source and channel with cross observation.

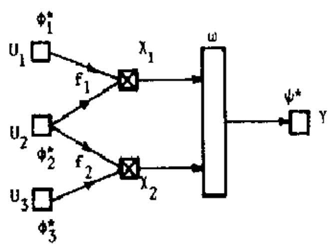

Fig. 2. Test channel M_2^*.

Let $\mathfrak{U}_1, \mathfrak{U}_2, \mathfrak{U}_3$ be finite sets and consider two deterministic functions

$$f_1: \mathfrak{U}_1 \times \mathfrak{U}_2 \to \mathfrak{X}_1,$$
$$f_2: \mathfrak{U}_2 \times \mathfrak{U}_3 \to \mathfrak{X}_2.$$

Define an associated test channel M_2^* as the channel with the input alphabets $\mathfrak{U}_1, \mathfrak{U}_2, \mathfrak{U}_3$, the output alphabet $\mathfrak{Y}$, and the transmission probabilities $w^*(y|u_1, u_2, u_3) = w(y|f_1(u_1, u_2), f_2(u_2, u_3))$ for $y \in \mathfrak{Y}$, $u_i \in \mathfrak{U}_i$ $(i = 1, 2, 3)$. (See Fig. 2.)

Let U_1, U_2, U_3 be random variables on $\mathfrak{U}_1, \mathfrak{U}_2, \mathfrak{U}_3$, respectively, such that

$$\Pr\{S_1 = s_1, S_2 = s_2, S_3 = s_3, U_1 = u_1, U_2 = u_2, U_3 = u_3\} = p(s_1, s_2, s_3)p(u_1|s_1)p(u_2|s_2)p(u_3|s_3).$$

Denote by X_1, X_2, Y the random variables with values in $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{Y}$ induced by U_1, U_2, U_3 via the test channel M_2^*. In particular,

$$X_1 = f_1(U_1, U_2), \quad X_2 = f_2(U_2, U_3). \tag{2.1}$$

Theorem 1: If there exist some U_1, U_2, U_3, f_1, f_2, X_1, X_2, Y such that for all A with $A \subset \Sigma \equiv \{1, 2, 3\}$, $A \neq \phi$,

$$H(S_A|S_{\bar{A}}) < I(U_A; Y|U_{\bar{A}}S_{\bar{A}}), \tag{2.2}$$

where $\bar{A}$ is the complement of A in Σ, and $S_A = (S_i)_{i \in A}$, *then* $S = (S_1, S_2, S_3)$ is admissible for the channel M_2.

Remark 1: If in (2.2) $H(S_A|S_{\bar{A}})$ vanishes for an $A \subset \Sigma$, then for that A the inequality "$<$" may be replaced by "$\leqslant$".

Proof: First, consider the problem of encoding S_1^n, S_2^n, S_3^n for the test channel M_2^* using three encoders ϕ_i^*: $\mathfrak{S}_i^n \to \mathfrak{U}_i^n$ $(i = 1, 2, 3)$ and one decoder ϕ^*: $\mathfrak{Y}^n \to \mathfrak{S}_1^n \times \mathfrak{S}_2^n \times \mathfrak{S}_3^n$ (cf. Fig. 2). Then, entirely paralleling the argument of Cover–El Gamal–Salehi used only for demonstrating their simpler case (i.e., a case without "common information" and hence the "time-sharing" parameter $\tilde{Q}$ set constant; see [6, p. 648]), it immediately follows that under condition (2.2) there exist encoding functions ϕ_i^* $(i = 1, 2, 3)$ and a decoding function ψ^* yielding the probability of error $\lambda \to 0$.

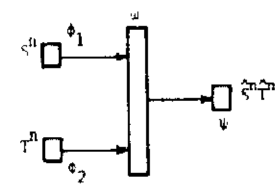

Fig. 3. Source coding via a multiple-access channel.

Next, for the original channel M_2, define encoding functions ϕ_1, ϕ_2 and a decoding function ψ by

$$\phi_1(S_1^n, S_2^n) = f_1(\phi_1^*(S_1^n), \phi_2^*(S_2^n)),$$
$$\phi_2(S_2^n, S_3^n) = f_2(\phi_2^*(S_2^n), \phi_3^*(S_3^n)),$$
$$\psi = \psi^*.$$

where $f_i(v, w) = (f_i(v_1, w_1), \cdots, f_i(v_n, w_n))$ for $v = (v_1, \cdots, v_n)$, $w = (w_1, \cdots, w_n)$ $(i = 1, 2)$. Clearly, these ϕ_1, ϕ_2, ψ yield the same probability of error λ for the channel M_2 too. Q.E.D.

Now let us apply Theorem 1 to the following special case: let S, T be arbitrarily correlated sources, let K be the common variable of S and T in the sense of Gács and Körner [14], and set $S_1 = S$, $S_2 = K$, $S_3 = T$. In this case our system (Fig. 1) is equivalent to the system considered by Cover–El Gamal–Salehi (Fig. 3). Thus, as a consequence of Theorem 1, we have the following result.

Theorem 2: (Cover–El Gamal–Salehi [6]) If there exist some $\tilde{Q}$, X_1, X_2, Y such that

$$\Pr\{S = s, T = t, \tilde{Q} = \tilde{q}, X_1 = x_1, X_2 = x_2, Y = y\} = p(s, t)p(\tilde{q})p(x_1|s, \tilde{q})p(x_2|t, \tilde{q})w(y|x_1, x_2) \tag{2.3}$$

and such that

$$H(S|T) < I(X_1; Y|X_2T\tilde{Q}), \tag{2.4}$$
$$H(T|S) < I(X_2; Y|X_1S\tilde{Q}), \tag{2.5}$$
$$H(STK) < I(X_1X_2; Y|K\tilde{Q}), \tag{2.6}$$
$$H(ST) < I(X_1X_2; Y), \tag{2.7}$$

then the source (S, T) is admissible for the channel M_2.

Proof: With the choice $U_2 = \tilde{Q}$ condition (2.2) is in the present situation equivalent with the following seven inequalities:

$$H(S|TK) < I(U_1; Y|U_3TK\tilde{Q}), \tag{2.8}$$
$$H(T|SK) < I(U_3; Y|U_1SK\tilde{Q}), \tag{2.9}$$
$$H(K|ST) < I(\tilde{Q}; Y|U_1U_3ST), \tag{2.10}$$
$$H(SK|T) < I(U_1\tilde{Q}; Y|U_3T), \tag{2.11}$$
$$H(TK|S) < I(U_3\tilde{Q}; Y|U_1S), \tag{2.12}$$
$$H(ST|K) < I(U_1U_3; Y|K\tilde{Q}), \tag{2.13}$$
$$H(STK) < I(U_1U_3\tilde{Q}; Y). \tag{2.14}$$

It is easy to check that (2.8) and (2.9) imply (2.11) and (2.12), respectively, and (2.10) is trivial (cf. Remark 1).

On the other hand, the right-hand side of (2.8) can be rewritten using the assumed Markov chain properties and (2.1) as

$$\begin{aligned} I(U_1; Y|U_3TK\tilde{Q}) &= I(U_1X_1; Y|U_3T\tilde{Q}) \\ &= I(X_1; Y|U_3T\tilde{Q}) \\ &= I(X_1; Y|U_3X_2T\tilde{Q}) \\ &= I(X_1; Y|X_2T\tilde{Q}). \end{aligned}$$

Similarly, for (2.9), (2.13), and (2.14) we have

$$I(U_3; Y|U_1SK\tilde{Q}) = I(X_2; Y|X_1S\tilde{Q}),$$
$$I(U_1U_3; Y|K\tilde{Q}) = I(X_1X_2; Y|K\tilde{Q}),$$
$$I(U_1U_3\tilde{Q}; Y) = I(X_1X_2; Y).$$

Q.E.D.

Remark 2: The way of deriving the theorem from Theorem 1 reveals that the heart of it consists in its special but elegant case with $\tilde{Q}$ being a constant:

$$H(S|T) < I(X_1; Y|X_2T), \tag{2.15}$$
$$H(T|S) < I(X_2; Y|X_1S), \tag{2.16}$$
$$H(ST) < I(X_1X_2; Y). \tag{2.17}$$

Note here that (2.7) implies (2.6) if $\tilde{Q}$ is a constant.

III. Source Coding with Side Information via a Multiple-Access Channel

The system considered by Cover–El Gamal–Salehi may be regarded as an extension of the Slepian–Wolf type noiseless source coding system [4]. In this section we consider a parallel extension of the Ahlswede–Körner–Wyner type noiseless source coding system with side information [11], [12].

Let M_2 be the multiple-access channel as specified in Section II, and let $S = (S, T)$ be an arbitrarily correlated source with alphabets $\mathcal{S}, \mathcal{T}$ (finite sets), respectively. We shall consider here the following joint system of S and M_2 as depicted in Fig. 4. The encoders ϕ_1, ϕ_2 are defined by

$$\phi_1: \mathcal{S}^n \to \mathcal{X}_1^n,$$
$$\phi_2: \mathcal{T}^n \to \mathcal{X}_2^n.$$

The decoder ψ observes an output n-sequence Y^n and maps it to an element $\hat{S}^n$ of $\mathcal{S}^n$: $\psi: \mathcal{Y}^n \to \mathcal{S}^n$. Since in this system the purpose of the decoder ψ is to reliably reproduce the source S^n alone, the probability of error $P_e(S)$ is defined by

$$P_e(S) = \Pr\{\hat{S}^n \neq S^n\}.$$

Definition 2: The source $S = (S, T)$ is said to be $\mathcal{S}$-*admissible* for the channel M_2 if for any $0 < \lambda < 1$ and sufficiently large n there exist encoding functions ϕ_1, ϕ_2 and a decoding function ψ such that $P_e(S) < \lambda$.

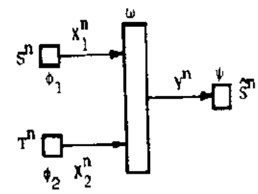

Fig. 4. Source coding with side information via a multiple-access channel.

Let Q, V be any random variables with values in finite sets $\mathcal{Q}, \mathcal{V}$, respectively, such that

$$\begin{aligned} \Pr\{S = s, T = t, Q = q, V = v, \\ X_1 = x_1, X_2 = x_2, Y = y\} \\ = p(s,t)p(q)p(x_1|s,q)p(v|t,q) \\ \cdot p(x_2|t,v,q)w(y|x_1,x_2), \end{aligned} \tag{3.1}$$

where X_1, X_2 take values in the input alphabets $\mathcal{X}_1, \mathcal{X}_2$, respectively, and Y in the output alphabet $\mathcal{Y}$ (Q is the time-sharing parameter).

Then we have Theorem 3.

Theorem 3: If there exist some Q, V, X_1, X_2, Y satisfying (3.1) for which

$$H(S|VQ) < I(SX_1; Y|VQ), \tag{3.2}$$
$$H(S|VQ) + I(T; V|Q) < I(SX_1V; Y|Q), \tag{3.3}$$

then the source (S, T) is S-admissible for the channel M_2. Here it is sufficient to consider only Q, V such that the cardinalities $\|Q\|, \|V\|$ of the ranges of Q, V are bounded by

$$\|V\| \leqslant |\mathcal{T}| \cdot |\mathcal{X}_2| + 3, \qquad \|Q\| \leqslant 4.$$

Remark 3: Conditions (3.2), (3.3) are equivalent to the following seemingly stronger conditions:

$$H(S|VQ) < I(SX_1; Y|VQ), \tag{3.4}$$
$$I(T; V|SQ) < I(V; Y|SX_1Q), \tag{3.5}$$
$$H(S|VQ) + I(T; V|Q) < I(SX_1V; Y|Q). \tag{3.6}$$

In fact, suppose that Q, V, X_1, X_2, Y satisfy (3.2), (3.3). If $I(T; V|SQ) < I(V; Y|SX_1Q)$ for those variables, then (3.4)–(3.6) follows. On the other hand, if $I(T; V|SQ) \geqslant I(V; Y|SX_1Q)$, by rewriting (3.3) we have

$$\begin{aligned} H(S) &< I(SX_1; Y|Q) + \{I(V; Y|SX_1Q) - I(T; V|SQ)\} \\ &\leqslant I(SX_1; Y|Q). \end{aligned}$$

This coincides with (3.4)–(3.6) with V set constant (cf. Remark 1).

Proof of Theorem 3: In view of Remark 3, it suffices to prove the admissibility under conditions (3.4)–(3.6). In proving the theorem we use the fundamental properties of jointly typical sequences (cf. Berger [15], Han and Kobayashi [16]. The notion used is slightly different from the one of Wolfowitz [17]). The set of all ϵ-typical sequences for a random variable Z is denoted by $T_\epsilon(Z)$, and, for $w \in \mathcal{W}^n$,

$$T_\epsilon(Z|w) = \{z | zw \in T_\epsilon(ZW)\}.$$

1) Auxiliary Code: First generate one random n-sequence $q = (q_1, \cdots, q_n)$ according to $\prod P(q_k)$.

Next, fix any R such that $R > I(T; V|Q)$ and take $L = \exp[nR]$ mutually independent n-sequences $\tilde{V}_1, \cdots, \tilde{V}_L$ with values taken equiprobably in $T_\epsilon(V|q)$, and set $\tilde{\mathcal{V}} = \{\tilde{V}_1, \cdots, \tilde{V}_L\}$.

Note here that $\tilde{V}_1, \cdots, \tilde{V}_L$ depend on the value of q. Since (3.1) implies that S, T, V (given Q), form a Markov chain in this order, there exists for sufficiently large n a function $V^* = g(T^n; \tilde{V}_1, \cdots, \tilde{V}_L)$ such that $V^* = \tilde{V}_i$ for $i = 1, \cdots, L$ and

$$\Pr\{(S^n, T^n, V^*) \in T_\epsilon(STV|q)\} \geq 1 - \delta, \quad (3.7)$$

where $\delta = \delta(\epsilon) \to 0$ as $\epsilon \to 0$ (a conditional version of Lemma 3.3 of Han and Kobayashi [16]).

2) Random Code Generation: For each $s = (s_1, \cdots, s_n) \in \mathcal{S}^n$ generate one random n-sequence $x_1(s) = (x_{11}, \cdots, x_{1n}) \in \mathcal{X}_1^n$ according to

$$\prod_{k=1}^{n} p(x_{1k}|s_k, q_k).$$

For each $t = (t_1, \cdots, t_n) \in \mathcal{T}^n$ generate one random n-sequence $x_2(t, v^*) = (x_{21}, \cdots, x_{2n}) \in \mathcal{X}_2^n$ according to

$$\prod_{k=1}^{n} p(x_{2k}|t_k, v_k^*, q_k),$$

where $v^* = (v_1^*, \cdots, v_n^*) = g(t; \tilde{V}_1, \cdots, \tilde{V}_L)$.

3) Encoding: Define the encoding functions $\phi_1: \mathcal{S}^n \to \mathcal{X}_1^n$, $\phi_2: \mathcal{T}^n \to \mathcal{X}_2^n$ by

$$\phi_1(s) = x_1(s), \quad (3.8)$$

$$\phi_2(t) = x_2(t, V^*) = x_2\big(t, g(t; \tilde{V}_1, \cdots, \tilde{V}_L)\big). \quad (3.9)$$

4) Decoding: Let $X_1^n = x_1(S^n)$, $X_2^n = x_2(T^n, V^*)$, and indicate by Y^n the output n-sequence induced on $\mathcal{Y}^n$ from X_1^n, X_2^n via the channel M_2. If (s, v^*) is the only element of $\mathcal{S}^n \times \tilde{\mathcal{V}}$ such that

$$(s, x_1(s), v^*, Y^n) \in T_\epsilon(SX_1VY|q). \quad (3.10)$$

Then define the decoding function $\psi: \mathcal{Y}^n \to \mathcal{S}^n$ by $\psi(Y^n) = s$, otherwise let $\psi(Y^n)$ be arbitrary (the decoder is to be informed about the value of q).

5) Probability of Error: Denoting by $E(s, v^*)$ the event (3.10) for $(s, v^*) \in \mathcal{S}^n \times \tilde{\mathcal{V}}$, we have the following expression for the probability of error:

$$\begin{aligned} P_e &= \Pr\Big\{E^c(S^n, V^*) \text{ or } \bigcup_{sv^* \neq S^nV^*} E(s, v^*)\Big\} \\ &\leq \Pr\{(S^n, T^n, V^*) \notin T_\epsilon(STV|q)\} \\ &\quad + \Pr\{E^c(S^n, V^*)|(S^n, T^n, V^*) \in T_\epsilon(STV|q)\} \\ &\quad + \mathrm{Ex}\Big(\sum_{sv^* \neq S^nV^*} \Pr\{E(s, v^*)|F_0(S^n, T^n, V^*)\}\Big), \end{aligned} \quad (3.11)$$

where "c" indicates the complement; Ex( ) denotes expectation, and $F_0(S^n, T^n, V^*)$ denotes the event

$$\{(S^n, T^n, V^*) \in T_\epsilon(STV|q)\} \cap E(S^n, V^*).$$

Note that the range of v^* in the above sums is restricted to be within $\tilde{\mathcal{V}}$.

From (3.7) we have

$$\Pr\{(S^n, T^n, V^*) \notin T_\epsilon(STV|q)\} \leq \delta. \quad (3.12)$$

Next, by the Markov properties among S, V, X_1, X_2, Y given Q (derived from (3.1)) and from the way of generating the random sequences $x_1(s)$, $x_2(t, v^*)$, it follows that

$$\Pr\{E^c(S^n, V^*)|(S^n, T^n, V^*) \in T_\epsilon(STV|q)\} \leq \epsilon. \quad (3.13)$$

On the other hand, the last term in (3.11) can be decomposed as follows:

$$\begin{aligned} &\mathrm{Ex}\Big(\sum_{sv^* \neq S^nV^*} \Pr\{E(s, v^*)|F_0(S^n, T^n, V^*)\}\Big) \\ &= \mathrm{Ex}\Big(\sum_{s \neq S^n} \Pr\{E(s, V^*)|F_0(S^n, T^n, V^*)\}\Big) \\ &\quad + \mathrm{Ex}\Big(\sum_{v^* \neq V^*} \Pr\{E(S^n, v^*)|F_0(S^n, T^n, V^*)\}\Big) \\ &\quad + \mathrm{Ex}\Big(\sum_{s \neq S^n} \sum_{v^* \neq V^*} \Pr\{E(s, v^*)|E_0(S^n, T^n, V^*)\}\Big). \end{aligned} \quad (3.14)$$

Denote the first, second, and third terms on the right-hand side of (3.14) by P_1, P_2, P_3, respectively.

a) Evaluation of P_1: Set

$$\mathcal{A}_1 = \{s | s \neq S^n, s \in T_\epsilon(S|V^*Y^nq)\},$$

and for each $s \in \mathcal{A}_1$,

$$\mathcal{B}_1(s) = T_\epsilon(X_1|sV^*Y^nq)\},$$

$$q_1(s) = \max_{x_1 \in \mathcal{B}_1(s)} \Pr\{x_1(s) = x_1\} \leq \exp[-n(H(X_1|SQ) - 2\epsilon)].$$

Then, for any $s \in \mathcal{A}_1$,

$$\begin{aligned} P_1(s) &\equiv \Pr\{E(s, V^*)|F_0(S^n, T^n, V^*)\} \\ &\leq \mathrm{Ex}\Big(\max_{s \in \mathcal{A}_1}(|\mathcal{B}_1(s)| \cdot q_1(s))\Big) \\ &\leq \exp[n(H(X_1|SVYQ) + 2\epsilon)] \\ &\quad \cdot \exp[-n(H(X_1|SQ) - 2\epsilon)] \\ &= \exp[-n(I(X_1; Y|SVQ) - 4\epsilon)] \\ &= \exp[-n(I(SX_1; Y|VQ) - I(S; Y|VQ) - 4\epsilon)]. \end{aligned} \quad (3.15)$$

On the other hand,

$$\begin{aligned} |\mathcal{A}_1| &\leq \exp[n(HS|VYQ) + 2\epsilon)] \\ &= \exp[n(H(S|VQ) - I(S; Y|VQ) + 2\epsilon)]. \end{aligned} \quad (3.16)$$

From (3.15) and (3.16),

$$P_1 = \mathrm{Ex}\left(\sum_{s\in\mathcal{C}_1} P_1(s)\right) \leqslant \mathrm{Ex}\left(|\mathcal{C}|_1 \cdot \max_{s\in\mathcal{C}_1} P_1(s)\right)$$

$$\leqslant \exp\left[-n(I(SX_1;Y|VQ) - H(S|VQ) - 6\epsilon)\right].$$

Consequently, condition (3.4) yields

$$P_1 < \epsilon \tag{3.17}$$

for sufficiently large n.

b) Evaluation of P_2: Noting that $\tilde{V}_1,\cdots,\tilde{V}_L$ ($\neq V^*$) are all uniformly distributed on $T_\epsilon(V|\boldsymbol{q})$, we have

$$P_2 = \mathrm{Ex}\left(\sum_{v^*\neq V^*} \Pr\{(E(S^n,v^*)|F_0(S^n,T^n,V^*))\}\right)$$

$$\leqslant \frac{(L-1)\exp\left[n(H(V|SX_1YQ)+2\epsilon)\right]}{(1-\epsilon)\exp\left[n(H(V|Q)-2\epsilon)\right]}$$

$$\leqslant (L-1)\exp\left[-n(I(V;SX_1Y|Q)-5\epsilon)\right] \tag{3.18}$$

for sufficiently large n. Since R ($L = \exp[nR]$) can be arbitrarily chosen so long as $R > I(T;V|Q)$, we may set $R = I(T;V|Q) + \epsilon$.

Therefore,

$$P_2 \leqslant \exp\left[-n(I(V;SX_1Y|Q) - I(T;V|Q) - 6\epsilon)\right]$$

$$= \exp\left[-n(I(V;Y|SX_1Q) - I(T;V|SQ) - 6\epsilon)\right], \tag{3.19}$$

which implies by condition (3.5) that

$$P_2 < \epsilon \tag{3.20}$$

for sufficiently large n.

c) Evaluation of P_3: Set

$$\mathcal{C}_3 = \{(s,v^*)|s\neq S^n, v^*\neq V^*, (s,v^*)\in T_\epsilon(SV|Y^n\boldsymbol{q})\},$$

and for each $(s,v^*)\in\mathcal{C}_3$,

$$\mathcal{B}_3(s,v^*) = T_\epsilon(X_1|sv^*Y^n\boldsymbol{q}),$$

$$q_3(s,v^*) = \max_{x_1\in\mathcal{B}_3(s,v^*)} \Pr\{x_1(s)=x_1\}.$$

Then for every $(s,v^*)\in\mathcal{C}_3$,

$$P_3(s,v^*) \equiv \Pr\{E(s,v^*)|E(S^n,V^*)\}$$

$$\leqslant \mathrm{Ex}\left(\max_{sv^*\in\mathcal{C}_3}(|\mathcal{B}_3(s,v^*)|\cdot q_3(s,v^*))\right)$$

$$\leqslant \exp\left[n(H(X_1|SVYQ)+2\epsilon)\right]$$

$$\cdot\exp\left[-n(H(X_1|SVQ)-2\epsilon)\right]$$

$$= \exp\left[-n(I(X_1;Y|SVQ)-4\epsilon)\right].$$

Therefore,

$$P_3 = \mathrm{Ex}\left(\sum_{sv^*\in\mathcal{C}_3} P_3(s,v^*)\right)$$

$$= \mathrm{Ex}\{|\mathcal{C}_3|\}\exp\left[-n(I(X_1;Y|SVQ)-4\epsilon)\right]$$

$$= \mathrm{Ex}\{|\mathcal{C}_3|\}\exp\left[-n(I(SX_1V;Y|Q) - I(SV;Y|Q) - 4\epsilon)\right]. \tag{3.21}$$

On the other hand, again from the uniform distribution property of $\tilde{V}_1,\cdots,\tilde{V}_L$ ($\neq V^*$) on

$$T_\epsilon(V|\boldsymbol{q})(L = \exp[n(I(T;V|Q)+\epsilon)]),$$

we have

$$\mathrm{Ex}\{|\mathcal{C}_3|\}$$

$$\leqslant \frac{(L-1)\exp\left[n(H(V|YQ)+2\epsilon)\right]}{(1-\epsilon)\exp\left[n(H(V|Q)-2\epsilon)\right]}$$

$$\cdot\exp\left[n(H(S|YVQ)+2\epsilon)\right]$$

$$\leqslant (L-1)\exp\left[n(H(S|VQ) - I(SV;Y|Q) + 7\epsilon)\right]$$

$$\leqslant \exp\left[n(I(T;V|Q) + H(S|VQ) - I(SV;Y|Q) + 8\epsilon)\right]. \tag{3.22}$$

From (3.21) and (3.22),

$$P_3 \leqslant \exp\left[-n(I(SX_1V;Y|Q) - H(S|VQ) - I(T;V|Q) - 12\epsilon)\right].$$

Hence, condition (3.6) yields

$$P_3 < \epsilon \tag{3.23}$$

for sufficiently large n.

Summarizing (3.11)–(3.14), (3.17), (3.20), (3.23), we can conclude that $P_e < \delta + 4\epsilon$. Q.E.D.

We present now several special cases of Theorem 3.

Corollary 1: If there exist X_1, X_2, Y such that

$$\Pr\{S=s, T=t, X_1=x_1, X_2=x_2, Y=y\} = p(s,t)p(x_1|s)p(x_2|t)w(y|x_1,x_2) \tag{3.24}$$

and

$$H(S) < I(SX_1;Y), \tag{3.25}$$

then the source (S,T) is S-admissible.

Proof: Let $Q = V = \phi$ (ϕ is a constant variable) in (3.2), (3.3) of Theorem 3 (also, cf. Remark 1). Q.E.D.

Remark 4: The right-hand side of (3.25) may be interpreted as follows. First, decompose $I(SX_1;Y)$ as

$$I(SX_1;Y) = I(S_1;Y) + I(S;Y|X_1).$$

The first term $I(X_1;Y)$ represents the information passing directly from the input X_1 to the output Y when we use the given multiple-access channel M_2 as a single-user channel with the "random state T" correlated to the source S; the second term $I(S;Y|X_1)$ represents the information passing from the input S via the intermediate terminal T to the output Y, without passing through X_1.

Here, in order to establish some application of Corollary 1, let us consider a single-user channel with "random state T" as follows. The transition probabilities are specified by

$$w(y|x,t), \qquad x\in\mathcal{X},\ y\in\mathcal{Y},\ t\in\mathcal{T},$$

where $\mathcal{X}$, $\mathcal{Y}$ are the input and output alphabets, respectively, and t indicates a value of random state distributed according to the random variable T. We are required to reliably send a source S (taking values in $\mathfrak{S}$) which is correlated to T. The encoder $\phi: \mathfrak{S}^n \to \mathfrak{X}^n$ maps $S^n = (S_1, \cdots, S_n)$ to the input n-sequence X^n, and the decoder $\psi: \mathfrak{Y}^n \to \mathfrak{S}^n$ maps the output n-sequence Y^n to an estimate $\hat{S}^n$ of S^n. The random state T_i for the ith channel use is correlated only to the ith source output S_i, and the joint distribution of (S_i, T_i) is i.i.d. of the generic random variable (S, T). We designate this kind of channel by $C(T)$.

Theorem 4 (Coding Theorem for C(T)): If there exists X, Y such that

$$\Pr\{S = s, T = t, X = x, Y = y\} = P(s,t)p(x|s)w(y|x,t) \quad (3.26)$$

and

$$H(S) < I(SX; Y), \quad (3.27)$$

then the source S is admissible for the channel $C(T)$ with correlated random state T.

Conversely, if the source S is admissible for $C(T)$, then (3.27) with $\leqslant$ replacing $<$ holds for some X, Y satisfying (3.26).

Proof: See Appendix I.

Remark 5: Theorem 4 establishes a necessary and sufficient condition for the admissibility of the source for the channel with random state, but with no statement on the admissibility for a trivial situation $H(S) = I(SX; Y|Q)$. This situation should be further examined not in general but depending on specific characteristics of each particular channel.

Corollary 1 treats a case where the encoder ϕ_2 for the side source T^n no longer carries out block encoding. On the other hand, the following example demonstrates another case where the encoder ϕ_1 for the principal source S^n attains no block encoding, but componentwise random encoding. (That is, $\phi_2: \mathfrak{T}^n \to \mathfrak{X}_2^n$ is called a *componentwise random encoder* if there exist n independent random functions $h_1, \cdots, h_n$ from $\mathfrak{T}$ to $\mathfrak{X}_2$ such that $\phi_2(t) = (h_1(t_1), \cdots, h_n(t_n))$, where $t = (t_1, \cdots, t_n)$.) In this latter case, the auxiliary variable V actually intervenes and hence block encoding is essentially needed for the source T^n.

To show an example, let us consider a special multiple-access channel M_2^0 with some deterministic function f: $\mathfrak{Y} \to \mathfrak{X}_2$ such that $w(y|x_1, x_2) = 0$ for $x_2 \neq f(y)$. In other words, the channel M_2^0 is such that one of the inputs x_2 is noiselessly transmitted to the receiver.

Theorem 5 (Coding Theorem for M_2^0): If there exist some Q, V, X_1, X_2, Y such that

$$\Pr\{S = s, T = t, Q = q, X_1 = x_1, X_2 = x_2, Y = y\} = p(s,t)p(q)p(v|t,q)p(x_1|s,q) \cdot p(x_2|v,q)w(y|x_1,x_2) \quad (3.28)$$

for which

$$H(S|VQ) < I(X_1; Y|X_2VQ), \quad (3.29)$$

$$H(S|VQ) + I(T; V|Q) < I(X_1X_2; Y|Q), \quad (3.30)$$

then the source (S, T) is S-admissible for M_2^0 by a componentwise random encoder $\phi_1: \mathfrak{S}^n \to \mathfrak{X}_1^n$. Conversely, if the source (S, T) is S-admissible for M_2^0 with a componentwise random encoder $\phi_1: \mathfrak{S}^n \to \mathfrak{X}^n$, then conditions (3.29), (3.30) with " $<$ " replaced by " $\leqslant$ " have to be satisfied for some Q, V, X_1, X_2, Y satisfying (3.28). Here it is sufficient to consider only V, Q such that

$$\|V\| \leqslant |\mathfrak{T}| \cdot |\mathfrak{X}_1| + 3, \quad \|Q\| \leqslant 4.$$

Proof: See Appendix II.

Example 1: Consider the case where M_2^0 is a pair of noiseless channels, i.e., $Y = X_1X_2$. If we put in (3.28) of Theorem 5 $Q = \phi$, $p(x_1|s,q) = p(x_1)$, $p(x_2|v,q) = p(x_2)$, then conditions (3.29), (3.30) are reduced to

$$H(S|V) < R_1,$$

$$H(S|V) + I(T; V) < R_1 + R_2,$$

where $R_1 = H(X_1)$, $R_2 = H(X_2)$. Clearly, these conditions are implied by the conditions $H(S|V) < R_1$, $I(T; V) < R_2$ established by Ahlswede–Körner–Wyner [11], [12]. Therefore, Theorem 5 may be regarded as an extension of the *noiseless* source coding theorem with side information.

Example 2: Consider a case where S, T are independent. If in Theorem 3 we put $p(x_1|s,q) = p(x_1)$, $p(x_2|v,q) = p(x_2)$, replace V by VX_2, and then set $V = \phi$, conditions (3.2), (3.3) reduce to $H(S) < I(X_1; Y|X_2)$, and can therefore be replaced by

$$H(S) < \max_{x_2 \in \mathfrak{X}_2} I(X_1; Y|X_2 = x_2). \quad (3.31)$$

Example 3: In the case $S = T$ (total cooperation), Theorem 3 does *not cover* the optimal condition

$$H(S) < \max_{p(x_1, x_2)} I(X_1X_2; Y). \quad (3.32)$$

This optimal condition *is covered* by Corollary 2 to appear in Section IV.

We conclude this section by giving a limiting expression of the condition for the admissibility. For any interger $m = 1, 2, \cdots$, let $\phi_1: \mathfrak{S}^m \to \mathfrak{X}_1^m$, $\phi_2: \mathfrak{T}^m \to \mathfrak{X}_2^m$ be any random function such that $X_1^m = \phi_1(S^m)$, S^m, T^m, $\phi_2(T^m) = X_2^m$ form a Markov chain in this order, and indicate by Y^m the corresponding output variable on $\mathfrak{Y}^m$.

Theorem 6: If

$$H(S) < \sup_{m, \phi_1, \phi_2} \frac{1}{m} I(S^m X_1^m; Y^m),$$

then the source (S, T) is S-admissible for M_2. Conversely, if the source (S, T) is S-admissible for M_2, then

$$H(S) \leqslant \sup_{m, \phi_1, \phi_2} \frac{1}{m} I(S^m X_1^m; Y^m).$$

Proof: The former part immediately follows from Corollary 1 by considering S^m, T^m as "supersources" of length m instead of S, T. The latter part is derived as follows: put

$$r(\lambda) = n\lambda \log|\mathcal{S}| + h(\lambda)$$
$$\times (h(\lambda) = -\lambda \log \lambda - (1-\lambda)\log(1-\lambda)),$$

then by Fano's inequality,

$$mH(S) - r(\lambda) \leqslant H(S^m) - H(S^m|Y^m)$$
$$= I(S^m; Y^m) = I(S^m X_1^m; Y^m),$$

where λ is the probability of error, m is the block length of the code, and $X_1^m Y^m$ are defined with respect to the encoders ϕ_1, ϕ_2 under consideration. Note that $r(\lambda)/m \to 0$ as $\lambda \to 0$. Q.E.D.

Remark 6: Of course the "naive" characterisation given in Theorem 6 cannot be used even in principal for numerical evaluations. However, since the so called single letter characterization involving auxiliary variables is often hard to get in multi-user theory, one should try as an *alternate approach to get estimates on the speed of convergence in the "naive" approach*. This may be a hard task, but there are no obvious reasons why this should be impossible.

IV. Results for a More General System

In the field of multiterminal noiseless source coding theory various kinds of source coding systems have been devised and studied. After Slepian–Wolf [4] treated the most basic and simplest case, a new dimension was added by Ahlswede–Körner [11] and Wyner [12] (and earlier, but in a weaker form, by Gray–Wyner [18]) by considering not only principal sources but also side information sources. In channel coding a certain kind of side information had been studied much earlier already by Shannon [21]. Various extensions of those results have been found and presently the most general ones seem to be those of Csiszár–Körner [19] and Han–Kobayashi [16]. A new direction in source coding was recently provided by Ahlswede [7], where multisources are described by specifying conditional distributions rather than joint distributions, that is, less knowledge about the source is assumed.

Here we stick to classical correlated source systems with several principal sources and several side information sources (helpers) as components whose coded versions are to be transmitted over a multiple-access channel with a "single" receiver.

Let $\boldsymbol{M}_r$ be a multiple-access channel with r input alphabets $\mathcal{X}_1, \cdots, \mathcal{X}_r$ (finite), an output alphabet $\mathcal{Y}$ (finite), and the transition probability

$$w(y|x_1, \cdots, x_r), \qquad \text{for } y \in \mathcal{Y},\ x_1 \in \mathcal{X}_1, \cdots, x_r \in \mathcal{X}_r. \tag{4.1}$$

Let $\boldsymbol{S}_p = (S_1, \cdots, S_a, S_{a+1}, \cdots, S_p)$ be a collection of correlated sources with values in $\mathcal{S}_1, \cdots, \mathcal{S}_a, \mathcal{S}_{a+1}, \cdots, \mathcal{S}_p$ (finite sets), where $S_1, \cdots, S_a$ are the *principal* sources and $S_{a+1}, \cdots, S_p$ the *side* sources. Let

$$\Sigma^{(1)} = \{1, \cdots, a\}, \quad \Sigma^{(2)} = \{a+1, \cdots, p\}.$$

The encoders $\phi_1, \cdots, \phi_r$ with cross observation at the encoders are defined as follows. For each $j = 1, \cdots, r$, the jth encoder ϕ_j observes a subset $S_{\Sigma_j}^n$ of $\{S_1^n, \cdots, S_p^n\}$ and maps the $S_{\Sigma_j}^n$ to an input n-sequences $X_j^n \in \mathcal{X}_j^n$:

$$\phi_j: \mathcal{S}_{\Sigma_j}^n \to \mathcal{X}_j^n,$$

where Σ_j is a prescribed subset of $\Sigma \equiv \{1, \cdots, p\}$, and $S_{\Sigma_j}^n = (S_k^n)_{k \in \Sigma_j}$, $\mathcal{S}_{\Sigma_j}^n = \prod_{k \in \Sigma_j} \mathcal{S}_k^n$. The decoder ψ observes an output n-sequence $Y^n \in \mathcal{Y}^n$ and maps it to an element $(\hat{S}_1^n, \cdots, \hat{S}_a^n) \in \mathcal{S}_1^n \times \cdots \times \mathcal{S}_a^n$:

$$\psi: \mathcal{Y}^n \to \mathcal{S}_1^n \times \cdots \times \mathcal{S}_a^n.$$

The purpose of the decoder ψ is to reliably reproduce the principal source informations $S_1^n, \cdots, S_a^n$ alone ($0 \leqslant a \leqslant p$).

Definition 3: We shall say that the source $\boldsymbol{S}_p$ is $(S_1, \cdots, S_a)$-admissible for the channel $\boldsymbol{M}_r$ if there exist $\phi_1, \cdots, \phi_r$ and ψ for which the probability of error $\Pr\{\hat{S}_1^n \cdots \hat{S}_a^n \neq S_1^n \cdots S_a^n\}$ approaches zero as $n \to \infty$.

To describe a sufficient condition for the $(S_1, \cdots, S_a)$-admissibility, we introduce many auxiliary variables as follows. Let $Q, U_1, \cdots, U_p; V_{a+1}, \cdots, V_p$ be any random variables with values in $\mathcal{Q}; \mathcal{U}_1, \cdots, \mathcal{U}_p; \mathcal{V}_{a+1}, \cdots, \mathcal{V}_p$, respectively, such that

$$\Pr\{Q = q; S_i = s_i, U_i = u_i\ (i \in \Sigma); V_k = v_k\ (k \in \Sigma^{(2)})\}$$
$$= p(q)p(s_1, \cdots, s_p) \prod_{i \in \Sigma} p(v_i|s_i, q) \prod_{i \in \Sigma} p(u_i|s_i, v_i, q), \tag{4.2}$$

where for notational simplicity we have set

$$V_1 = S_1, \cdots, V_a = S_a. \tag{4.3}$$

Next, choose r arbitrary deterministic functions

$$f_j: \mathcal{U}_{\Sigma_j} \to \mathcal{X}_j, \qquad j = 1, \cdots, r. \tag{4.4}$$

and define the input variables X_j on $\mathcal{X}_j$ of the channel $\boldsymbol{M}_r$ by $X_j = f_j(U_{\Sigma_j})$ $(j = 1, \cdots, r)$. Denote by Y the output variable with values in $\mathcal{Y}$ induced form the X_j via the channel $\boldsymbol{M}_r$. The relation among S_i, V_k, U_i, X_j, Y is illustrated in Fig. 5 (Q is the time-sharing parameter).

Theorem 7: If there exist some $Q; V_{a+1}, \cdots, V_p; U_1, \cdots, U_p; f_1, \cdots, f_r; X_1, \cdots, X_r; Y$ as above for which for all A with $A \subset \Sigma$, $A \cap \Sigma^{(1)} \neq \phi$,

$$I(S_A; V_A | V_{\bar{A}} Q) < I(W_A; Y | W_{\bar{A}} Q), \tag{4.5}$$

where we have put $W_j = S_j U_j$ for $j \in \Sigma^{(1)}$, $W_j = V_j$ for $j \in \Sigma^{(2)}$, then the source $\boldsymbol{S}_p$ is $(S_1, \cdots, S_a)$-admissible for the channel $\boldsymbol{M}_r$.

Remark 7: In (4.5) we may omit the corresponding condition if its left-hand side vanishes.

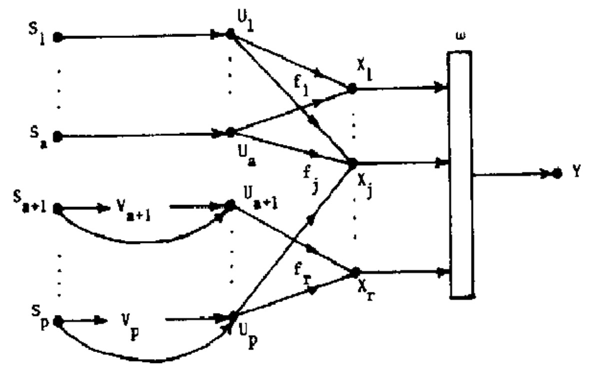

Fig. 5. Region for modulo-two sum problem.

Proof: Both the test channel argument in Section II and the argument used in the proof of Theorem 2 can be extended to the general case under consideration (cf. Han [9]), from which condition (4.5) follows. Q.E.D.

We shall next give a limiting expression for the condition of the admissibility for this general system. Consider any integer $m = 1, 2, \cdots$, and let ϕ_i: $\mathcal{S}_i^m \to \mathcal{U}_i^m$ be any random functions ($i \in \Sigma$) such that

$$\Pr\{S_i^m = s_i,\ U_i^m = u_i,\ i \in \Sigma\} = p(s_i, i \in \Sigma) \prod_{i \in \Sigma} p(u_i|s_i),$$

where $U_i^m = \phi_i(S_i^m)$. Let f_j: $\mathcal{U}_{\Sigma_j}^m \to \mathcal{X}_j^m$ be any deterministic functions ($j = 1, \cdots, r$) and set $X_j^m = f_j(U_{\Sigma_j}^m)$. Denote by Y^m the output variable on $\mathcal{Y}^m$ induced from X_j^m ($j = 1, \cdots, r$). Indicate by (σ_A) the vector with components σ_A indexed by $A \subset \Sigma^{(1)}$ ($A \neq \phi$) and define

$$\mathcal{C}_m = \left\{ (\sigma_A) | \sigma_A \leqslant \frac{1}{m} I(S_A^m U_A^m;\ Y^m | S_{A^c}^m U_{A^c}^m) \text{ for some } \phi_i, f_j \right\},$$

where A^c indicates the complement of A in $\Sigma^{(1)}$; and

$$\mathcal{C} = \bigcup_{m=1}^{\infty} \mathcal{C}_m.$$

Note that $\mathcal{C}$ is a bounded set because $\mathcal{C}_m$ ($m = 1, 2, \cdots$) are all within a bounded set. Denote by $\overline{\mathcal{C}}$ the closure of $\mathcal{C}$.

Theorem 8 (Limiting Expression): Let $h_A = H(S_A|S_{A^c})$ for A with $\phi \neq A \subset \Sigma^{(1)}$. If (h_A) is an internal point of $\overline{\mathcal{C}}$, then the source $(S_1, \cdots, S_a, S_{a+1}, \cdots, S_p)$ is $(S_1, \cdots, S_a)$-admissible. Conversely, if the source $(S_1, \cdots, S_a, S_{a+1}, \cdots, S_p)$ is $(S_1, \cdots, S_a)$-admissible then $(h_A) \in \overline{\mathcal{C}}$.

Proof: The former part is an immediate consequence of Theorem 7: Let $V_i = \phi$ ($i \in \Sigma^{(2)}$) and apply to the super sources S_i^m instead of S_i. The latter part follows analogously to the proof of Theorem 6. Q.E.D.

We are now in the position to prove a stronger version of Theorem 3. Let $S_1 = S$, $S_2 = K$, $S_3 = T$ (K is the common variable of S and T in the sense of Gács and Körner) and apply Theorem 7 to the special system depicted in Fig. 1 with $(\hat{S}_1^n, \hat{S}_2^n, \hat{S}_3^n)$ replaced by $(\hat{S}_1^n, \hat{S}_2^n)$, where S_1, S_2 are principal; S_3 is a helper ($\Sigma^{(1)} = \{1, 2\}$, $p = 3$, $a = 2$, $r = 2$). Then we have the following result for the equivalent side information system depicted in Fig. 4.

Theorem 9: If there exists some Q, $\tilde{Q}$, V, X_1, X_2, Y such that

$$\begin{aligned}\Pr\{S = s, T = t, K = k, Q = q, \tilde{Q} = \tilde{q}, \\ X_1 = x_1, X_2 = x_2, Y = y\} \\ = p(s,t)p(q)p(\tilde{q}|k)p(x_1|s,\tilde{q},q)p(v|t,q) \\ \cdot p(x_2|t,v,\tilde{q},q)w(y|x_1,x_2)\end{aligned} \tag{4.6}$$

for which

$$H(S|KVQ) < I(SX_1;\ Y|KV\tilde{Q}Q), \tag{4.7}$$

$$H(S|VQ) < I(SX_1\tilde{Q};\ Y|VQ), \tag{4.8}$$

$$H(S|KVQ) + I(T;V|KQ) < I(SX_1V;\ Y|K\tilde{Q}Q), \tag{4.9}$$

$$H(S|VQ) + I(T;V|Q) < I(SX_1V\tilde{Q};\ Y|Q), \tag{4.10}$$

then the source (S, T) is S-admissible for the channel M_2. Here it is sufficient to consider only $V, Q, \tilde{Q}$ such that

$$\|V\| \leqslant |T| \cdot |X_2| + 7, \qquad \|Q\| \leqslant 8,$$

$$\|\tilde{Q}\| \leqslant |\mathcal{X}_1| \cdot |\mathcal{X}_2| \cdot |K| + 4.$$

Remark 8:

1) It is easy to see that Theorem 3 is a special case of Theorem 9. In fact, if we set $\tilde{Q} = \phi$ (a constant) in (4.7)–(4.10) then (4.8), (4.10) imply (4.7), (4.9), respectively, yielding Theorem 3.
2) The variable $\tilde{Q}$ seems somewhat similar to the time-sharing parameter Q, but is different in that $\tilde{Q}$ is used here also to carry the common information K as is the case in the theorem of Cover–El Gamal–Salehi.

Corollary 2: Suppose that there exists a deterministic function $T = f(S)$, i.e., $K = T$; then the following holds. If there exist $\tilde{Q}$, X_1, X_2, Y such that

$$\begin{aligned}\Pr\{S = s, T = t, \tilde{Q} = \tilde{q}, X_1 = x_1, X_2 = x_2, Y = y\} \\ = p(s,t)p(q)p(x_1|s,q)p(x_2|t,q)w(y|x_1,x_2),\end{aligned} \tag{4.11}$$

for which

$$H(S|K) < I(X_1;\ Y|K\tilde{Q}), \tag{4.12}$$

$$H(S) < I(SX_1\tilde{Q};\ Y), \tag{4.13}$$

then the source (S, T) is S-admissible.

Conversely, if the source (S, T) is S-admissible, the conditions (4.12), (4.13) with "$<$" replaced by "$\leqslant$" have to be satisfied for some $\tilde{Q}$, X_1, X_2, Y satisfying (4.11). Here it is sufficient to consider only $\tilde{Q}$ such that $\|\tilde{Q}\| \leqslant |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 2$.

Proof: See Appendix III.

Remark 9: If $S = T = K$, conditions (4.12), (4.13) reduce to the optimal condition (3.32) in Example 3: for any distribution $p(x_1, x_2)$ we can set $\tilde{Q} = X_1X_2$.

V. A Channel with Side Information at the Decoder

Shannon introduced in [21] a one sender–one receiver channel with finitely many states $\mathcal{T}$ selected from use to use independently at random and also independently of the letters sent (cf. channels with random state described in Section III). This channel can be viewed as being composed of a multiple-access channel with transmission probabilities $w(y|x,t)$ for $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $t \in \mathcal{T}$ and a random mechanism selecting letters (states) $t_1, t_2, \cdots \in \mathcal{T}$ that is a sequence of i.i.d. random variables $T_1, T_2, \cdots$ with values in $\mathcal{T}$. When neither the $\mathcal{X}$-sender nor the $\mathcal{Y}$-receiver know anything about the outcome of $T_1, T_2, \cdots, T_n$ we are just dealing with an ordinary discrete memoryless channel.

Shannon investigated a case of side information at the $\mathcal{X}$-sender: we know the outcomes of $T_1, \cdots, T_n$ before we send X_{n+1}. Wolfowitz [17] studied this and other cases.

Here it is assumed that the $\mathcal{X}$-sender has no side information, but that the $\mathcal{Y}$-decoder can be informed about the outcomes of $T_1, \cdots, T_n$ via a separate noiseless channel at a fixed rate R (partial side information). Denote the capacities of this channel in dependence of R by $C(R)$.

Conjecture: For $R \geqslant 0$,

$$C(R) = \max\{I(X;Y|V) | I(T;V|Y) \leqslant R,\ V \to T \to Y, \|V\| \leqslant \|T\| + 1\}.$$

It is not hard to show by the argument exploited in Section III that the expression at the right-hand side is an achievable rate; the problem is to prove optimality, that is, a converse.

This channel is somewhat related to, but much simpler than, the interference channel (see [22], [23]) and represents one of the simplest cases in channel coding where the converse is still not known. A fairly simple example of this kind in source coding was formulated in [7, part II, sect. VII].

VI. A New Achievable Rate Region for the Binary Modulo-Two Sum Two Helper Problem

Let us consider a correlated source with three generic random variables X, Y, Z taking values in $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$, respectively, such that $Z = X \oplus Y$, where $X \oplus Y$ is 0 if $X = Y$, and 1 otherwise.

Suppose there are two encoders

$$\phi_1: \mathcal{X}^n \to \mathcal{M}_1 = \{1, 2, \cdots, M_1\},$$
$$\phi_2: \mathcal{Y}^n \to \mathcal{M}_2 = \{1, 2, \cdots, M_2\}$$

and the decoder

$$\psi: \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{Z}^n,$$

which is required to reliably reproduce Z^n based on the knowledge of $\phi_1(X^n)$ and $\phi_2(Y^n)$.

Definition 4: (R_1, R_2) is said to be an achievable pair of rates, if for any $0 < \lambda < 1$, $0 < \eta$ and all sufficiently large n there exist some ϕ_1, ϕ_2, ψ such that

$$\frac{1}{n}\log\|\phi_1\| \leqslant R_1 + \eta,$$
$$\frac{1}{n}\log\|\phi_2\| \leqslant R_2 + \eta,$$

and $\Pr\{\hat{Z}^n \neq Z^n\} \leqslant \lambda$, where $Z^n = \psi(\phi_1(\hat{X}^n), \phi_2(Y^n))$, and $\|\phi_i\|$ denotes the cardinality of the range of the function ϕ_i.

Denote the closure of the set of all achievable pairs of rates by $R_\oplus$. The problem of determinng $R_\oplus$ is a very special but also very interesting case of the two-helper side information problem stated by Ahlswede–Körner in [24]. Clearly, if the Slepian–Wolf conditions

$$R_1 \geqslant H(H|Y),$$
$$R_2 \geqslant H(Y|X),$$
$$R_1 + R_2 \geqslant H(X,Y) \tag{6.1}$$

are satisfied, then $(R_1, R_2) \in \mathcal{R}_\oplus$. Körner–Marton proved in [13] that $(R_1, R_2) \in \mathcal{R}_\oplus$, if

$$R_1 \geqslant H(Z), \quad R_2 \geqslant H(Z). \tag{6.2}$$

Moreover, they showed that (6.2) gives the exact rate region if $\Pr(X = 0) = \Pr(X = 1)$ and Y is the output variable of a binary symmetric channel with input variable X.

Here we present an achievable region $\mathcal{R}$ which contains the regions described by (6.1) and (6.2), and is in general larger than $\mathcal{R}_0$, the convex hull of both of them.

Let U, V be finite-valued random variables such that U, X, Y, V form a Markov chain in this order:

$$U \to X \to Y \to V, \tag{6.3}$$

and let $\mathcal{R}(U, V)$ be the set of all (R_1, R_2) satisfying

$$R_1 \geqslant I(U; X|V) + H(Z|UV), \tag{6.4}$$
$$R_2 \geqslant I(V; Y|U) + H(Z|UV), \tag{6.5}$$
$$R_1 + R_2 \geqslant I(UV; XY) + 2H(Z|UV). \tag{6.6}$$

Denote by $\mathcal{R}$ the convex closure of $\cup_{U,V}\mathcal{R}(U, V)$, where the union is taken over all U, V satisfying (6.3).

Theorem 10: $\mathcal{R} \subset \mathcal{R}_\oplus$.

The proof is given in Appendix IV, which is based on a combination of a standard technique in source coding [15] and the method of Elias [26] (cf. Gallager [20]) for finding linear codes, which was previously used by Körner–Marton for proving their result stated above.

Remark 10: Theorem 10 contains the previous results for the modulo-two sum problem. In fact, if we put $U = X$, $V = Y$ in (6.4)–(6.6) we obtain the Slepian–Wolf condition (6.1).

Next, if we put $U = V = \phi$, we obtain the Körner–Marton condition (6.2).

The region $\mathcal{R}$ established in Theorem 10 strictly extends $\mathcal{R}_0$ for general binary sources, as is seen from the following example.

Example 4: We choose X, Y with the probability distribution:

$$\Pr\{X=0, Y=0\} = 0.003920,$$
$$\Pr\{X=0, Y=1\} = 0.976080,$$
$$\Pr\{X=1, Y=0\} = 0.019920,$$
$$\Pr\{X=1, Y=1\} = 0.000080;$$
$$\Pr\{X=0\} = 0.980000, \quad \Pr\{Y=0\} = 0.023840,$$
$$\Pr\{Z=0\} = 0.004000,$$

for which we have in bits

$$H(X) = 0.1414405, \quad H(Y) = 0.1624894,$$
$$H(Z) = 0.0376223, \quad H(XY) = 0.1790629.$$

Choose auxiliary variables U, V taking values in $\{0,1\}$ with

$$\Pr\{U=0|X=0\} = 0.55, \quad \Pr\{U=0|X=1\} = 0.45,$$
$$\Pr\{V=0|Y=0\} = 0.95, \quad \Pr\{V=0|Y=1\} = 0.05,$$

then the point $P = (R_1, R_2)$ with

$$R_1 = I(U; X) + H(Z|UV),$$
$$R_2 = I(V; Y|U) + H(Z|UV)$$

has the value $R_1 = 0.0251286$, $R_2 = 0.1096321$. The region $\mathcal{R}_0$ is illustrated in Fig. 6, in which the boundary line AB is specified by

$$f(R_1, R_2) \equiv (R_1 - H(Z))(H(Y) - H(Z)) + (R_2 - H(Z))(H(X) - H(X|Y)) = 0.$$

For the point P we have

$$f(R_1, R_2) = -0.0000443$$

which implies that at least the point P lies outside $\mathcal{R}_0$.

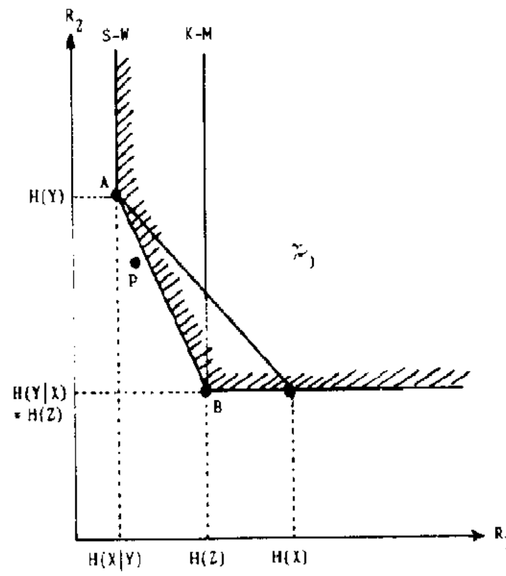

Fig. 6. Region $\mathcal{R}_0$.

VII. Correlated Codes: An Alternate Approach to the Problem of Transmitting Correlated Messages

We first take a closer look at the quote from [1] given in the Introduction and emphasize the following observations.

a) In sentence 1) Shannon asks whether it is always possible to achieve the outer bound G_0 to the capacity region of the two-way channel for messages with a suitable dependency structure.

b) Sentence 4) seems to indicate that he has in mind that the message structure is that of a memoryless correlated source, but in sentence 1) the message structure is not specified.

3) In sentence 2) Shannon assigns rates R_1, R_2 to whatever "source-channel-code" he has in mind. It seems to us that only with two parameters such as R_1, R_2 no reasonable code of this kind can be satisfactorily described. Note that Cover–El Gamal–Salehi avoid talking about rates at all in their approach (Definition 1, also cf. Han [27]).

d) It is by now well-known that in the problem of transmitting correlated messages over multiway channels the classical *separation principle* of Shannon (separate coding of the source and the channel for single-user channels) fails in the sense that separate coding results in loss of efficiency (in rates, if defined). The approach indicated in [1] as well as the particular "source-channel-code" (or matching) chosen in [6] take this into account. *An obvious drawback of such an approach is that it is only meaningful if the message statistics are known exactly.* We tend to share the opinion expressed by Gallager in [20, p. 14]: "In many data transmission systems the probabilities with which messages are to be used are either unknown or unmeaningful." Therefore the results of [6] and also our results in the previous sections should probably be considered to be more of the nature of providing theoretical insight than of direct practical importance. The least one should try is to safeguard against classes of message distributions (see [25] and [7]). Another way of coming closer to a real communication situation with our models consists of *enforcing the separation principle* (in spite of its suboptimality in an ideal situation) and investigating what can be done (also optimally) if source and channel coding are carried out separately.

A first approach is to study *correlated codes* for channels *without* any reference to sources. In particular it is interesting to know whether the question raised by Shannon, namely, the achievability of the outer bound, can be formulated for correlated *codes* and whether then the answer is positive. We shall focus here on the multiple-access channel, because it is better understood than the two-way channel. We find it now more convenient to denote the input alphabets by $\mathcal{X}, \mathcal{Y}$, and the output alphabet by $\mathcal{Z}$.

Definition 5: An $(n, M_x, M_y, M_{xy}, \alpha, \lambda)$ *correlated code* for the MAC M_2 is a system $(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$, where

$$\mathcal{M}_x = \{u_1, \cdots, u_{M_x}\} \subset \mathcal{X}^n, \quad |\mathcal{M}_x| = M_x \quad (7.1)$$

$$\mathcal{M}_y = \{v_1, \cdots, v_{M_y}\} \subset \mathcal{Y}^n, \quad |\mathcal{M}_y| = M_y \quad (7.2)$$

$$\alpha: \mathcal{M}_x \times \mathcal{M}_y \to \{0,1\}; \quad \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} \alpha(u_i, v_j) = M_{xy} \quad (7.3)$$

$$\mathfrak{D} = \{D_{ij} \subset \mathcal{Z}^n | \alpha(u_i, v_j) = 1\},$$

$$D_{ij} \cap D_{i'j'} = \phi, \qquad \text{for } (i,j) \neq (i',j'), \tag{7.4}$$

$$w(D_{ij}|u_i, v_j) \geq 1 - \lambda, \qquad \text{for every } u_i, v_j \text{ with } \alpha(u_i, v_j) = 1. \tag{7.5}$$

Other quantities of interest are

$$M_{x|j} = \sum_i \alpha(u_i, v_j), \quad M_{y|i} = \sum_j \alpha(u_i, v_j), \tag{7.6}$$

$$M_{x|y} = \frac{1}{M}\sum_j M_{x|j} = M_{xy}/M_y,$$

$$M_{y|x} = \frac{1}{M}\sum_i M_{y|i} = M_{xy}/M_x, \tag{7.7}$$

$$M_0 = M_{xy}(M_{x|y}M_{y|x})^{-1} = M_y M_{y|x}^{-1} = M_x M_{x|y}^{-1}. \tag{7.8}$$

Clearly, $M_{x|y}$, $M_{y|x}$, and M_{xy} uniquely determine M_x, M_y and M_{xy}, and vice versa.

Definition 6: An $(n, M_x, M_y, M_{xy}, \alpha, \lambda)_X$ *correlated side-information code* for M_2 is a system satisfying (7.1)-(7.3) and

$$\mathfrak{D} = \{D_i \subset \mathcal{Z}^n | 1 \leq i \leq M_x\},$$

$$D_i \cap D_{i'} = \phi, \qquad \text{for } i \neq i', \tag{7.9}$$

$$w(D_i|u_i, v_j) \geq 1 - \lambda, \qquad \text{for every } u_i, v_j \text{ with } \alpha(u_i, v_j) = 1. \tag{7.10}$$

This code concept is for the case where only the messages in $\mathfrak{M}_x$ are to be reproduced. Note that these code definitions do not necessarily imply that $u_i \neq u_{i'}$ for $i \neq i'$ (resp. $v_j \neq v_{j'}$ for $j \neq j'$).

To the parameters M_{xy}, M_x, M_y, $M_{x|y}$, $M_{y|x}$, we can assign the rates $R_{xy} = (1/n)\log M_{xy}$, $R_x = (1/n)\log M_x$, $R_y = (1/n)\log M_y$, $R_{x|y} = (1/n)\log M_{x|y}$, $R_{y|x} = (1/n)\log M_{y|x}$. Clearly, $R_{x|y} + R_{y|x} \leq R_{xy} \leq R_x + R_y$.

Definition 7: (R_x, R_y, R_{xy}) is an *achievable* rate-triple for the correlated code problem (resp. correlated side-information code problem) for M_2, if for any $\eta > 0$, $0 < \lambda < 1$, and every sufficiently large n there exists, for some $\tilde{R}_x, \tilde{R}_y, \tilde{R}_{xy}$ such that $|\tilde{R}_x - R_x| < \eta$, $|\tilde{R}_y - R_y| < \eta$, $|\tilde{R}_{xy} - R_{xy}| < \eta$, an $(n, \exp[n\tilde{R}_x], \exp[n\tilde{R}_y], \alpha, \lambda)$ correlated code (resp. correlated side-information code). Denote the region of achievable rate triples (R_x, R_y, R_{xy}) by $\mathfrak{R}_{\text{cor}}$ (resp. $\mathfrak{R}_{\text{cor-side}}$).

A. A Partial Result for R_{cor}

Using the identities

$$R_{x|y} = R_{xy} - R_y, \quad R_{y|x} = R_{xy} - R_x, \tag{7.11}$$

one can equivalently transform $\mathfrak{R}_{\text{cor}}$ into the set $\overline{\mathfrak{R}}_{\text{cor}}$ of achievable triples $(R_{x|y}, R_{y|x}, R_{xy})$, and vice versa. It is often more convenient to work with $\overline{\mathfrak{R}}_{\text{cor}}$.

Define $\mathfrak{R}^*$ as the set of triples $(R_{x|y}, R_{y|x}, R_{xy})$ satisfying

$$R_{x|y} = I(X; Z|YQ),$$

$$R_{y|x} = I(Y; Z|XQ),$$

$$R_{xy} = I(XY; Z|Q),$$

for some correlated input variable pair (X, Y) and the corresponding output variable Z, where Q is the time-sharing parameter. With a slight modification of the standard argument used to establish the coding theorem on MAC, we can easily prove the following theorem.

Theorem 11:

$$\overline{R}_{\text{cor}} \supset \mathfrak{R}^*. \tag{7.12}$$

It should be noted that the result of Slepian–Wolf [5] for the MAC with correlated sources can be viewed as a coding theorem for correlated codes with an α having the 1 in disjoint equal-sized rectangles placed along a diagonal. For *those* α they show that $(R_{x|y}, R_{y|x}, R_{xy})$ is achievable if

$$R_{x|y} \leq I(X; Z|Y\tilde{Q}), \tag{7.13}$$

$$R_{y|x} \leq I(Y; Z|X\tilde{Q}), \tag{7.14}$$

$$R_{x|y} + R_{y|x} \leq I(XY; Z|\tilde{Q}), \tag{7.15}$$

$$R_{x|y} \leq I(XY; Z), \tag{7.16}$$

for some $\tilde{Q}$, X, Y, Z with $X \to \tilde{Q} \to Y$, $\tilde{Q} \to XY \to Z$. This region is in general smaller than $\overline{\mathfrak{R}}_{\text{cor}}$ for the multiple-access channel.

B. Remarks about $\mathfrak{R}_{\text{cor-side}}$

In this case of side codes the achievability of $(R_x, R_{x|y}, R_{y|x})$ implies that $(R_x, R_{x|y}, cR_{y|x})$ is also achievable for $c \geq 1$, because one can always take the v_j with multiplicity. Therefore one should look for the smallest $R_{y|x}$ such that the triple is achievable.

One may also study first the projection $(R_x, R_{x|y})$. This already constitutes an amazing new problem. Let $\mathcal{P}_{\text{cor-side}}$ be the region of those achievable pairs.

Obviously, $(0,0)$, $(\max_Y I(X; Z|Y), \max_Y I(X; Z|Y))$, $(\max_{X,Y} I(XY; Z), 0)$ and their convex combinations are in $\mathcal{P}_{\text{cor-side}}$.

Problem 1: What is the exact region of $\mathfrak{R}_{\text{cor-side}}$ (resp. of its projection)?

With this short discussion we hope to have made clear that there are several new problems for the code concepts defined (and others one might think of) if used for the MAC, and there is a whole collection of problems if one considers various multiway channels such as the TWC, broadcast channel, interference channel, etc. Instead of going into further details, we now give some ideas about source-channel matching.

C. The Subcode and Minor Problem

The following problems strikingly demonstrate the richness of multi-user information theory in comparison to the classical single-user theory. First notice that for the two code concepts defined above one can already pass from average error to maximal error codes without essential loss in rate and hence the rate (capacity) regions are the same. We can therefore pass from a code $(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$ to a (minor) subcode $(\mathcal{M}'_x, \mathcal{M}'_y, \alpha', \mathcal{D}')$ without increase in error probability, where $\mathcal{M}'_x \subset \mathcal{M}_x$, $\mathcal{M}'_y \subset \mathcal{M}_y$, α' is the restriction of α on $\mathcal{M}'_x \times \mathcal{M}'_y$, and $\mathcal{D}'$ is a new decoding rule, which can, but need not, be obtained by restricting $\mathcal{D}$. Denote the collection of all such subcodes by $\mathcal{C}(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$.

A similar but larger class of subcodes is the collection of all subcodes, denoted by $\mathcal{C}^*(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$, obtained if we replace α' in the above by any α'' with $\alpha'' \leqslant \alpha$, where "$\leqslant$" means that $\alpha(u_i, v_j) = 1$ whenever $\alpha''(u_i, v_j) = 1$. Clearly, $\mathcal{C}(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D}) \subset \mathcal{C}^*(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$. Assigning rates (R'_x, R'_y, R'_{xy}) to these subcodes, one obtains two spectra of rate triples from a single correlated code, which we denote by $\mathcal{SPC}(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$ and $\mathcal{SPC}^*(\mathcal{M}_x, \mathcal{M}_y, \alpha, \mathcal{D})$, respectively.

Problem 2: Which spectra $\mathcal{SPC}^*$ (resp. $\mathcal{SPC}$) are achievable with arbitrarily small error probabilities for n large?

Note that the same correlated code can be used for several message sources by suitably selecting its subcodes (in this sense it may be called a *multi-user code*) and the richness of the spectrum $\mathcal{SPC}^*$ (resp. $\mathcal{SPC}$) is a criterion for its capability to meet various demands.

Some hard problems arise. For instance, to what extent does the spectrum of a code allow to reconstruct this code (at least in an approximate sense)? There is a relation to reconstruction problems of bipartite graphs from subgraphs, because both problems are based on the same kind of incidence structure α.

D. Matching

An *abstract* correlated source is simply a pair of random variables (S, T) where S, T take values in finite sets $\mathcal{S}$ resp. $\mathcal{T}$.

We say that this source can be *α-matched* with the correlated code $(\mathcal{M}_x, \mathcal{M}_y, \mathcal{D}, \alpha)$ if there exist functions $f: \mathcal{S} \to \mathcal{M}_x$, $g: \mathcal{T} \to \mathcal{M}_y$, $\psi: \mathcal{M}_x \times \mathcal{M}_y \to \mathcal{S} \times \mathcal{T}$ such that

$$\Pr\{\psi(f(S), g(T)) = (S, T)\} \geqslant 1 - \epsilon \quad (7.17)$$

and

$$\Pr\{\alpha(f(S), g(T)) = 1\} \geqslant 1 - \epsilon. \quad (7.18)$$

Now imagine that the component sources S resp. T are not directly placed at the channel $\mathcal{X}$-encoder resp. $\mathcal{Y}$-encoder, but are linked to them, respectively, by the channel which may be assumed to be noiseless. Then the cardinalities of the ranges of f and g become an issue, because the channel can transmit only at limited rates R_1 resp. R_2. We say that (S, T) can be (R_1, R_2, ϵ)-*matched* with $(\mathcal{M}_x, \mathcal{M}_y, \mathcal{D}, \alpha)$ if, in addition to (7.17), (7.18), also $\|f\| \leqslant R_1$, $\|g\| \leqslant R_2$. Finally, we say that (S, T) is $(R_1, R_2, \epsilon, \lambda)$-*transmissible* over the channel M_2, if it can be (R_1, R_2, ϵ)-matched with some correlated code $(\mathcal{M}_x, \mathcal{M}_y, \mathcal{D}, \alpha, \lambda)$.

Remark 11: Cover, El Gamal, and Salehi [6] assumed in their approach that the source outputs of the component sources are available at the corresponding channel encoders. An interesting more general problem arises if, as above, the outputs of the component sources are to be transmitted via channels of rates R_1, resp. R_2 to the corresponding encoders of M_2.

E. Continuous Transmission and the Embedding Problem

Suppose that the correlated code $(\mathcal{M}_x, \mathcal{M}_y, \mathcal{D}, \alpha, \lambda)$ is to be used repeatedly. If we place a probability distribution β on $\mathcal{M}_x \times \mathcal{M}_y$ such that for instance

$$\beta(u_i, v_j) = \left(\sum_{i', j'} \alpha(u_{i'}, v_{j'})\right)^{-1} \qquad \text{if } \alpha(u_i, v_j) = 1,$$

then we obtain a correlated code $(\mathcal{M}_x, \mathcal{M}_y, \mathcal{D}, \alpha, \beta, \lambda)$ with an average (taken with respect to β) error probability less than λ.

If we denote by I, J random variables with values in $\mathcal{M}_x$ resp. $\mathcal{M}_y$ and joint distribution β, then the repeated use of the code can be described by $(I_t, J_t)_{t=1}^{\infty}$ where the (I_t, J_t) are independent copies of (I, J). For simplicity let us assume that the message statistic is that of a correlated source $(S_t, T_t)_{t=1}^{\infty}$, which is to be transmitted and reproduced according to some fidelity criterion. This situation is quite complex and as a first step for its analysis one may consider the following problem, which is also of interest in itself.

Embedding Problem: For two (correlated) sources $(X_t, Y_t)_{t=1}^{n}$, $(U_t, V_t)_{t=1}^{m}$ and $0 < \epsilon < 1$ it is to be decided whether there exist functions $f: \mathcal{X}^n \to \mathcal{U}^m$, $g: \mathcal{Y}^n \to \mathcal{V}^m$, $\psi: \mathcal{U}^m \times \mathcal{V}^m \to \mathcal{X}^n \times \mathcal{Y}^n$ with

$$\Pr\{\psi(f(X^n), g(Y^n)) = (X^n, Y^n)\} \geqslant 1 - \epsilon, \quad (7.19)$$

and

$$\sum_{(u^m, v^m) \in f(\mathcal{X}^n) \times g(\mathcal{Y}^n)} |P_1(u^m, v^m) - P_2(u^m, v^m)| \leqslant \epsilon,$$

where

$$P_1(u^m, v^m) = \Pr\{(f(X^n), g(Y^n)) = (u^m, v^m)\},$$

$$P_2(u^m, v^m) = \frac{\Pr\{(U^m, V^m) = (u^m, v^m)\}}{\Pr\{(U^m, V^m) \epsilon f(\mathcal{X}^n) \times g(\mathcal{Y}^n)\}}.$$

Further, what are the minimal rates of f and g?

Appendix I

Proof of Theorem 4

As will be seen from the way of generating the random codes in the proof of Theorem 3, the admissibility of S based on condition (3.25) can be attained even if we use instead of ϕ_2 the encoder $\tilde{\phi}_2: \mathfrak{T}^n \to \mathfrak{X}_2^n$ as follows, i.e., $\tilde{\phi}_2$ maps $T^n = (T_1, \cdots, T_n)$ to $X_2^n = (X_{2i}, \cdots, X_{2n})$ in such a way that X_{2i} is randomly generated from T_i according to the probability $p(X_{2i} = x_2 | T_i = t)$ as specified in (3.24).

This implies that block coding ϕ_2 is no longer needed in this special situation. Thus, if we replace X_1 (resp. x_1) by X (resp. x) and interpret $w(y|x_1, t) \equiv \sum_{x_2} p(x_2|t) w(y|x_1, x_2)$ as the transition probability of a channel $\tilde{C}(T)$ with random state T, then the direct part of Theorem 4 immediately follows.

Now consider the converse part of Theorem 4. Suppose that we have an encoder ϕ and a decoder ψ for $C(T)$ yielding the probability of error $0 < \lambda < 1$. Set $X^n = (X_1, \cdots, X_n) = \phi(S^n)$, $S^n = (S_1, \cdots, S_n)$, and denote by $Y^n = (Y_1, \cdots, Y_n)$ the output n-sequence induced by X^n, T^n.

Using Fano's lemma $H(S^n|Y^n) \leqslant n\lambda \log|\mathfrak{S}| + h(\lambda) \equiv r(\lambda)$, where $h(\lambda) = -\lambda \log \lambda - (1-\lambda)\log(1-\lambda)$, we have

$$\begin{aligned} nH(S) &\leqslant H(S^n) - H(S^n|Y^n) + r(\lambda) \\ &= I(S^n; Y^n) + r(\lambda) = I(S^n X^n; Y^n) + r(\lambda) \\ &= H(Y^n) - H(Y^n|S^n X^n) + r(\lambda) \\ &\leqslant \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S^n X^n Y^{i-1}) + r(\lambda), \end{aligned}$$

where $Y^{i-1} = (Y_1, \cdots, Y_{i-1})$. Note that $Y_i, S_i X_i, S^{[i]} X^{[i]} Y^{i-1}$ form a Markov chain in this order, where

$$S^{[i]} = (S_1, \cdots, S_{i-1}, S_{i+1}, \cdots, S_n),$$
$$X^{[i]} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n).$$

Hence

$$\begin{aligned} nH(S) &\leqslant \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S_i X_i) + r(\lambda) \\ &= \sum_{i=1}^n I(S_i X_i; Y_i) + r(\lambda) \\ &\leqslant nI(S_k X_k; Y_k) + r(\lambda), \end{aligned} \tag{A1}$$

where $I(S_k X_k; Y_k) = \max_i I(S_i X_i; Y_i)$. Set $S = S_k$, $X = X_k$, $Y = Y_k$, then (A1) can be rewritten as

$$H(S) \leqslant I(SX; Y) + r(\lambda).$$

Letting $\lambda \to 0$ we have the desired converse. Q.E.D.

Appendix II

Proof of Theorem 5

1) The Direct Part: Replace V by VX_2 in (3.2), (3.3) and use the relations

$$I(SX_1; Y|X_2 VQ) = I(X_1; Y|X_2 VQ),$$
$$I(SX_1 VX_2; Y|Q) = I(X_1 X_2; Y|Q).$$

Note that the replacement of V by VX_2 does not violate the form of condition (3.28) because $p(x_2|v, q)$ is not of the form which depends on t.

2) The Converse Part: We use the same notation as that in the proof of Theorem 4 (Appendix I). Set $V_i = X_2^n T^{i-1} \tilde{S}^{i+1}$, where $\tilde{S}^{i+1} = (S_{i+1}, \cdots, S_n)$ Then, by Fano's lemma,

$$\begin{aligned} r(\lambda) \geqslant H(S^n|Y^n X_2^n) &= \sum_{i=1}^n H(S_i|Y^n X_2^n \tilde{S}^{i+1}) \geqslant \sum_{i=1}^n H(S_i|Y^n X_2^n T^{i-1} \tilde{S}^{i+1}) \\ &\overset{(1)}{=} \sum_{i=1}^n H(S_i|Y_i X_2^n T^{i-1} \tilde{S}^{i+1}) = \sum_{i=1}^n H(S_i|Y_i X_{2i} V_i), \end{aligned}$$

where equality (1) follows from the fact that $Y^{i-1}\tilde{Y}^{i+1}$, $X_2^n T^{i-1} \tilde{S}^{i+1}$, S_i from a Markov chain in this order, indicated as $Y^{i-1}\tilde{Y}^{i+1} \to X_2^n T^{i-1}\tilde{S}^{i+1} \to S_i$, which holds because ϕ_1 is componentwise.

Since V_i contains X_{2i}, we have

$$\begin{aligned} \sum_{i=1}^n H(S_i|V_i) &\leqslant \sum_{i=1}^n H(S_i|V_i X_{2i}) \\ &\quad - \sum_{i=1}^n H(S_i|Y_i X_{2i} V_i) + r(\lambda) \\ &= \sum_{i=1}^n I(S_i; Y_i|X_{2i} V_i) + r(\lambda) \\ &\leqslant \sum_{i=1}^n I(S_i X_{1i}; Y_i|X_{2i} V_i) + r(\lambda) \\ &= \sum_{i=1}^n I(X_{1i}; Y_i|X_{2i} V_i) + r(\lambda), \end{aligned} \tag{A2}$$

where in the last step we have used the memoryless character of the channel.

Next, noting that Y^n contains X_2^n by assumption, we have $H(S^n X_2^n|Y^n) \leqslant r(\lambda)$. Hence,

$$\begin{aligned} H(S^n X_2^n) - r(\lambda) &\leqslant H(S^n X_2^n) - H(S^n X_2^n|Y^n) \\ &= I(S^n X_2^n; Y^n) = H(Y^n) - H(Y^n|S^n X_2^n) \\ &\leqslant \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S^n X_2^n Y^{i-1}) \\ &\leqslant \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S^n X_1^n X_2^n Y^{i-1}) \\ &\overset{(2)}{=} \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|X_{1i} X_{2i}) \\ &= \sum_{i=1}^n I(X_{1i} X_{2i}; Y_i), \end{aligned} \tag{A3}$$

where step (2) follows from the memoryless character of the channel. On the other hand,

$$\begin{aligned} H(S^n X_2^n) &= H(S^n|X_2^n) + H(X_2^n) \\ &= H(S^n|X_2^n) + I(T^n; X_2^n); \end{aligned}$$

$$\begin{aligned} H(S^n|X_2^n) &= \sum_{i=1}^n H(S_i|X_2^n \tilde{S}^{i+1}) \\ &= \sum_{i=1}^n H(S_i|X_2^n T^{i-1} \tilde{S}^{i+1}) \\ &\quad + \sum_{i=1}^n I(S_i; T^{i-1}|X_2^n \tilde{S}^{i+1}) \\ &= \sum_{i=1}^n H(S_i|V_i) + \sum_{i=1}^n I(S_i; T^{i-1}|X_2^n \tilde{S}^{i+1}), \end{aligned}$$

$$\begin{aligned} I(T^n; X_2^n) &= H(T^n) - H(T^n|X_2^n) \\ &= \sum_{i=1}^n H(T_i) - \sum_{i=1}^n H(T_i|X_2^n T^{i-1}) \\ &= \sum_{i=1}^n H(T_i) - \sum_{i=1}^n H(T_i|X_2^n T^{i-1}\tilde{S}^{i+1}) \\ &\quad - \sum_{i=1}^n I(T_i; \tilde{S}^{i+1}|X_2^n T^{i-1}) \\ &= \sum_{i=1}^n I(T_i; V_i) - \sum_{i=1}^n I(T_i; \tilde{S}^{i+1}|X_2^n T^{i-1}). \end{aligned}$$

Therefore, using the lemma of Csiszár and Körner [28]:

$$\sum_{i=1}^n I(S_i; T^{i-1}|X_2^n \tilde{S}^{i+1}) = \sum_{i=1}^n I(T_i; \tilde{S}^{i+1}|X_2^n T^{i-1}),$$

we have

$$H(S^n X_2^n) = \sum_{i=1}^n H(S_i|V_i) + \sum_{i=1}^n I(T_i; V_i). \tag{A4}$$

Hence, from (A3) and (A4),

$$\sum_{i=1}^n H(S_i|V_i) + \sum_{i=1}^n I(T_i; V_i) \leq \sum_{i=1}^n I(X_{1i} X_{2i}; Y) + r(\lambda). \tag{A5}$$

We notice here that the property

$$X_{1i} \to S_i \to T_i \to V_i \to X_{2i}$$

holds because ϕ_1 is componentwise.

Finally, define Q, S, T, V, X_1, X_2, Y by

$$\Pr\{Q = i\} = 1/n, \qquad i = 1, \cdots, n;$$

$$S = S_i, \quad T = T_i, \quad V = V_i, \quad X_1 = X_{1i},$$

$$X_2 = X_{2i}, \quad Y = Y_i \qquad \text{given } Q = i,$$

then (A2) and (A5) yield the required result if we note that $\|Q\|, \|V\|$ can be finitely bounded. Q.E.D.

Appendix III

Proof of Corollary 2

Letting $V = Q = \phi$ in (4.7)–(4.10) yields (4.12), (4.13). Note that $I(SX_1; Y|K\tilde{Q}) = I(X_1; Y|K\tilde{Q})$, because $K = T$, yielding the former part of the corollary. Consider the latter part.

With the same notation as in the proof of Theorem 4 of (Appendix I), we have

$$\begin{aligned} H(S^n|K^n) - r(\lambda) &\leq I(S^n; Y^n|K^n) = I(S^n X_1^n; Y^n|K^n) \\ &= H(Y^n|K^n) - H(Y^n|S^n X_1^n K^n) \\ &= \sum_{i=1}^n H(Y_i|K^n Y^{i-1}) - \sum_{i=1}^n H(Y_i|S^n X_1^n K^n Y^{i-1}) \\ &\leq \sum_{i=1}^n H(Y_i|K_i T^{[i]}) - \sum H(Y_i|X_{1i} K_i T^{[i]}) \\ &= \sum_{i=1}^n I(X_{1i}; Y_i|K_i T^{[i]}), \end{aligned}$$

where $T^{[i]} = (T_1, \cdots, T_{i-1}, T_{i+1}, \cdots, T_n)$. Hence, defining Q, Q', S, T, X_1, X_2, Y by

$$\Pr\{Q = i\} = 1/n, \qquad i = 1, \cdots, n;$$

$$S = S_i, \quad T = T_i, \quad K = K_i, \quad Q' = T^{[i]},$$

$$X_1 = X_{1i}, \quad X_2 = X_{2i}, \quad Y = Y_i \qquad \text{given } Q = i,$$

we have

$$H(S|K) - \frac{1}{n} r(\lambda) \leq I(X_1; Y|KQ'Q). \tag{A6}$$

On the other hand,

$$\begin{aligned} H(S^n) - r(\lambda) &\leq I(S^n; Y^n) = I(S^n X_1^n; Y^n) \\ &= H(Y^n) - H(Y^n|S^n X_1^n) \\ &\leq \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i S|^n X_1^n Y^{i-1}) \\ &= \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S^n X_1^n T^{[i]}) \\ &= \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i|S_i X_{1i} T^{[i]}) \\ &= \sum_{i=1}^n I(S_i X_{1i} T^{[i]}; Y_i) \\ &= nI(SX_1 Q'; Y|Q) \leq nI(SX_1 Q'Q; Y). \end{aligned}$$

Hence,

$$H(S) - \frac{1}{n} r(\lambda) \leq I(SX_1 Q'Q; Y). \tag{A7}$$

Inequalities (A6), (A7) with $\tilde{Q} = Q'Q$ yield the latter part. Q.E.D.

Appendix IV

Proof of Theorem 10

By the time-sharing argument it suffices to show $\mathfrak{R}(U, V) \subset \mathfrak{R}_{\oplus}$ for any (U, V) and for this we have to consider only the extreme points of $\mathfrak{R}(U, V)$, say,

$$R_1 = I(U; X) + H(Z|UV), \tag{A8}$$

$$R_2 = I(V; Y|U) + H(Z|UV). \tag{A9}$$

Decompose the rates (R_1, R_2) as follows:

$$R_1 = R_1^{(1)} + R_1^{(2)}, \quad R_2 = R_2^{(1)} + R_2^{(2)}, \tag{A10}$$

where

$$R_1^{(1)} = I(U; X), \quad R_1^{(2)} = H(Z|UV), \tag{A11}$$

$$R_2^{(1)} = I(V; Y|U), \quad R_2^{(2)} = H(Z|UV). \tag{A12}$$

We use the rates $(R_1^{(1)}, R_2^{(1)})$ in the first step, and the rates $(R_1^{(2)}, R_2^{(2)})$ in the second step.

Step 1: Denote by $\tilde{U}_1, \cdots, \tilde{U}_L$ mutually independent random variables distributed uniformly in $T_\epsilon(U)$, and by $\tilde{V}_1, \cdots, \tilde{V}_J$ mutually independent random variables distributed uniformly in $T_\epsilon(V)$, where $\eta > 0$ is a positive number, and

$$\begin{aligned} L_1 &= \exp[n(I(U; X) + \eta/2)], \\ &= \exp\left[n\left(R_1^{(1)} + \eta/2\right)\right], \end{aligned} \tag{A13}$$

$$J = \exp[n(I(V; Y) + \eta/4)]. \tag{A14}$$

Then, there exist functions $U^* = U^*(X^n; \tilde{U}_1, \cdots, \tilde{U}_L)$ $V^* = V^*(Y^n; \tilde{V}_1, \cdots, \tilde{V}_J)$ such that $U^* = \tilde{U}_i$ for some $i = 1, \cdots, L_1$; and $V^* = \tilde{V}_j$ for some $j = 1, \cdots, J$; and for sufficiently large n

$$\Pr\{(U^*, V^*, X^N, Y^n)\} \in T_\epsilon(UVXY)\} \geq 1 - \delta_1, \quad (A15)$$

where $\delta_1 = \delta_1(\epsilon, \nu) \to 0$ as $\epsilon \to 0$ (cf. [16]).

Put $\mathcal{L}_1 = \{1, 2, \cdots, L_1\}$, $\mathcal{L}_2 = \{1, 2, \cdots, L_2\}$ where $L_2 = \exp[n(R_2^{(1)} + \eta/2)]$, and partition $\{\tilde{V}_1, \cdots, \tilde{V}_J\}$ into L_2 subclasses of the same cardinality. Let us define the subencoders $h_1^{(1)}: \mathcal{X}^n \to \mathcal{L}_1$, $h_2^{(1)}: \mathcal{Y}^n \to \mathcal{L}_2$ by $h_1^{(1)}(X^n) = i$ if $U^* = \tilde{U}_i$; and by $h_2^{(1)}(Y^n) = j$ if V^* belongs to the jth subclass of $\{\tilde{V}_1, \cdots, \tilde{V}_J\}$. The corresponding subdecoder

$$k: \mathcal{L}_1 \times \mathcal{L}_2 \to \{\tilde{U}_1, \cdots, \tilde{U}_L\} \times \{\tilde{V}_1, \cdots, \tilde{V}_J\} \quad (A16)$$

is defined by $k(i, j) = (\tilde{U}_i, \tilde{V}_m)$ if $\tilde{V}_m$ is one and only one element of the jth subclass such that $(\tilde{U}_i, \tilde{V}_m) \in T_\epsilon(UV)$; otherwise let $k(i, j)$ be an arbitrary element. Then, by virtue of the standard evaluation technique in multiterminal source coding,

$$\Pr\{k(i, j) = (U^*, V^*) | (U^*, V^*, X^n, Y^n) \in T_\epsilon(UVX)\} \geq 1 - \delta_2, \quad (A17)$$

where $\delta_2 = \delta_2(\epsilon, \nu) \to 0$ as $\epsilon \to 0$.

Step 2: Let

$$m = n(H(Z|UV) + \eta/2)/\log 2 \quad (A18)$$

and consider the random linear mapping $f: \{0, 1\}^n \to \{0, 1\}^m = \mathfrak{M}$ such that $f(s) = As$, where A is an $m \times n$ matrix whose elements are all independently and uniformly distributed in $\{0, 1\}$. Here we regard $\{0, 1\}$ as forming the Galois field GF(2).

Denote other subencoders $h_1^{(2)}: \mathcal{X}^n \to \mathfrak{M}$, $h_2^{(2)}: \mathcal{Y}^n \to \mathfrak{M}$ by $h_1^{(2)}(X^n) = f(X^n)$, $h_2^{(2)}(Y^n) = f(Y^n)$. Note here that

$$(1/n)\log|\mathfrak{M}| = R_1^{(2)} + \eta/2 = R_2^{(2)} - \eta/2. \quad (A19)$$

Step 3: Let us now define the composite encoders $\phi_1: \mathcal{X}^n \to \mathcal{L}_1 \times \mathfrak{M} \equiv \mathfrak{M}_1$, $\phi_2: \mathcal{Y}^n \to \mathcal{L}_2 \times \mathfrak{M} \equiv \mathfrak{M}_2$ by $\phi_1 = (h_1^{(1)}, h_1^{(2)})$, $\phi_2 = (h_2^{(1)}, h_2^{(2)})$, and the composite decoder $\psi: \mathfrak{M}_1 \times \mathfrak{M}_2 \to \mathcal{Z}^n$ by $\psi(i, j, f(X^n), f(Y^n)) = z$ if z is one and only element of $T_\epsilon(Z|k(i, j))$ such that $f(z) = f(X^n) + f(Y^n) \equiv f(Z^n)$, where $i \in \mathcal{L}_1, j \in \mathcal{L}_2$.

To evaluate the probability of decoding error, consider the following exhaustive error events:

$$E_1: \quad (U^*, V^*, X^n) \notin T_\epsilon(UVXY),$$
$$E_2: \quad k(i, j) \neq (U^*, V^*),$$
$$E_3: \quad (k(i, j), Z^n) \notin T_\epsilon(UVZ),$$
$$E_4: \quad f(z) = f(Z^n), \quad \text{for some } z \neq Z^n \text{ such that } (k(i, j), z) \in T_\epsilon(UVZ).$$

Thus, by putting $Z^n = \psi(\phi_1(X^n), \phi_2(Y^n))$, we have

$$\begin{aligned} P_e &= \Pr\{Z^n \neq \hat{Z}^n\} \\ &\leq \Pr\{E_1\} - \Pr\{E_2|E_1^c\} + \Pr\{E_3|E_1^c E_2^c\} \\ &\quad + \Pr\{E_4|E_1^c \cap E_2^c \cap E_3^c\}. \end{aligned}$$

By (A15) and (A17),

$$\Pr\{E_1\} \leq \delta_1, \quad (A20)$$

$$\Pr\{E_2|E_1^c\} \leq \delta_2. \quad (A21)$$

Since $UVXY$ uniquely determines the value of $Z = X \oplus Y$ the event $E_1^c \cap E_2^c$ implies the event E_3^c, so that

$$\Pr\{E_3|E_1^c \cap E_2^c\} = 0. \quad (A22)$$

On the other hand,

$$\begin{aligned} &\Pr\{E_4|E_1^c \cap E_2 \cap E_3^c\} \\ &\leq \mathrm{Ex}\left(\sum_{z \neq Z^n} \Pr\{f(z) = f(Z^n) | E_1^c \cap E_2^c \cap E_3^c\} \right), \\ &z \in T_\epsilon(Z|k(i, j)), \end{aligned} \quad (A23)$$

where the random variables in (A23) are A in addition to X^n, Y^n, Z^n, $\tilde{U}_i$, $\tilde{V}_j$.

As all the elements of A are independently and uniformly distributed in $\{0, 1\}$, by counting all the cases satisfying $f(z) = f(Z^n)$ it follows that for any $z \neq Z^n$

$$\Pr\{f(z) = f(Z^n) | E_1^c \cap E_2^c \cap E_3^c\} = \left(2^{(n-1)}/2^n\right)^m = 2^{-m}.$$

Hence, by (A18),

$$\begin{aligned} \Pr\{E_4|E_1^c \cap E_2^c \cap E_3^c\} &\leq |T_\epsilon(Z|UV)| \cdot 2^{-m} \\ &\leq \exp[n(H(Z|UV) + 2\epsilon)] \cdot 2^{-m} \\ &= \exp[-n(\eta/2 - 2\epsilon)] \leq \delta_3, \end{aligned} \quad (A24)$$

where $\delta_3 = \delta_3(\eta, \epsilon, n)$ can be made arbitrarily small by choosing ϵ small and then n large.

Consequently, we have

$$P_e \leq \delta_1 + \delta_2 + \delta_3. \quad (A25)$$

Finally, note that the total rates $(\bar{R}_1, \bar{R}_2)$ used for the composite encoders ϕ_1, ϕ_2 are

$$\bar{R}_1 = R_1^{(1)} + R_1^{(2)} + \eta = R_1 + \eta$$
$$\bar{R}_2 = R_2^{(1)} + R_2^{(2)} + \eta = R_2 + \eta,$$

and hence the rate given by (A8) and (A9) is achievable. Q.E.D.

REFERENCES

[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Univ. California, pp. 611–644, 1961.

[2] G. Dueck, "The capacity of the two-way channel can exceed the inner bound," *Inform. Contr.*, vol. 40, pp. 258–266, 1979.

[3] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Inform. Theory*, Tsahkadsor, 1971; Publishing House of the Hungarian Academy of Science, pp. 23–52, 1973.

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

[5] ——, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, 1973.

[6] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple-access channel with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 648–659, 1980.

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

[8] G. Dueck, "A note on the multiple access channel with correlated sources," *IEEE Trans. Inform. Theory*, to appear.

[9] T. S. Han, "The capacity region of a general multiple-access channel with certain correlated sources," *Inform. Contr.*, vol. 40, no. 1, pp. 37–60, 1979

[10] ——, "Source coding with cross observation at the encoders," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 360–361, May 1979.

[11] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, 1975.

[12] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 294–300, 1975.
[13] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 219–221, 1979.
[14] P. Gács and J. Körner, "Common information is far less than mutual information," *Prob. Contr. Inform.*, vol. 2, pp. 149–162, 1973.
[15] T. Berger, "Multiterminal source coding," Notes from a lecture presented at CISM Summer School, Udine, Italy, 1977.
[16] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 277–288, 1980.
[17] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer-Verlag, 1978.
[18] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, pp. 1681–1721, 1974.
[19] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 155–165, 1980.
[20] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[21] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Dev.*, vol. 2, pp. 289–293, 1958.
[22] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, 1974.
[23] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, 1978.
[24] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," presented at the *7th Prague Conf. Inform. Theory, Stat. Dec. Fct's, and Rand. Proc.*, Sept. 1974.
[25] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 44, pp. 159–175, 1978.
[26] P. Elias, "Coding for noisy channels," in *IRE Convention Record, Part 4*, pp. 37–46, 1955.
[27] T. S. Han, "Slepian-Wolf-Cover theorem for networks of channels," presented at the *1979 IEEE Int. Symp. Inform. Theory*, Grignano, Italy, 25–29 June, 1979; also, *Inform. Contr.*, vol. 47, pp. 67–83, 1980.
[28] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 339–348, 1978.

Estimation of Spatial and Spectral Parameters of Multiple Sources

BOAZ PORAT, MEMBER, IEEE, AND BENJAMIN FRIEDLANDER, SENIOR MEMBER, IEEE

Abstract—**The problem of estimating spatial and spectral parameters of multiple sources by sensor arrays is considered. A parametric model is derived for the multichannel measurement vector. The model parameters are the time delays and the autoregressive moving-average parameters of all the sources. A suboptimal parameter estimation procedure is proposed, and simulation results are presented to illustrate its performance.**

I. INTRODUCTION

PASSIVE surveillance systems are often required to detect sources of acoustic or electromagnetic energy and to estimate their location and spectral characteristics. Examples include sonar systems, acoustic arrays for detection of low-flying aircraft, and geophones for detection and localization of seismic events. These systems consist of multiple sensors arranged in some pattern such as a linear array. The relative time delays between the arrival of signals from a point source to the various sensors contain information about the location (bearing and range) of the source. For moving sources and sensors additional location information is contained in the differential Doppler shifts of the signals arriving at different sensors.

The estimation of time delay and Doppler shift has been studied extensively in the past two decades. The subject of optimum (or maximum-likelihood) processing has received particular attention [1]–[4]. The structure of the maximum-likelihood estimator (MLE) of delay and Doppler has been developed for various situations. Some of the key results are discussed briefly in the following paragraphs.

The MLE for the time delay between signals arriving at two sensors (with no differential Doppler shift) can be implemented by a generalized correlation procedure [5]: the sensor data are filtered and then cross-correlated; the location of the peak of the cross correlation function provides the delay estimate. The optimal filters depend on the spectral characteristics of the signal and noise and on the signal-to-noise ratio (SNR).

When more than two sensors are available it is possible to combine the pairwise delay estimates (obtained by generalized cross correlation) to get a global delay estimate [6]. An alternative implementation of the delay estimator is provided by a conventional beamformer configuration involving a set of steering delays followed by filtering,

Manuscript received March 16, 1982; revised July 6, 1982. This work was supported by the Office of Naval Research under Contract N00014-79-C-0743.

B. Porat is with RAFAEL, P. O. Box 2250, Haifa, Israel.
B. Friedlander is with Systems Control Technology, 1801 Page Mill Road, Palo Alto, CA 94304.

0018-9448/83/0500-0412$01.00 ©1983 IEEE