Finally, with (A.20), (A.22), and (A.23), we conclude that

$$\begin{aligned}
&\sum_{x,x',y,y',z} P_X(x)P_{X'}(x')P_Y(y)P_{Y'}(y') \\
&\quad\times\left|\sum_s P_{S|X'Y'}(s|x',y')W(z|x,y,s)\right. \\
&\quad\left.-\sum_s P_{S'|XY}(s|x,y)W(z|x',y',s)\right| \\
&\le \sum_{x,x',y,y',z} P_X(x)P_Y(y)\left(\sum_s P_{S|X',Y'}(s|x',y')W(z|x,y,s)\right) \\
&\quad\cdot|P_{X'}(x')P_{Y'}(y')-P_{X'Y'}(x',y')| \\
&\quad+\sum_{x,x',y,y',z}\left|P_X(x)P_Y(y)P_{X'Y'}(x',y')\right. \\
&\quad\cdot\sum_s P_{S|X',Y'}(s|x',y')W(z|x,y,s) \\
&\quad\left.-P_{X'}(x')P_{Y'}(y')P_{XY}(x,y)\sum_s P_{S'|XY}(s|x,y)W(z|x',y',s)\right| \\
&\quad+\sum_{x,x',y,y',z} P_{X'}(x)P_{Y'}(y)\left(\sum_s P_{S'|XY}(s|x,y)W(z|x',y',s)\right) \\
&\quad\cdot|P_X(x)P_Y(y)-P_{XY}(x,y)| \\
&\le \sum_{x,x',y,y',z}|P_{X'}(x')P_{Y'}(y')-P_{X'Y'}(x',y')|+2c\sqrt{\xi+\zeta} \\
&\quad+\sum_{x,x',y,y',z}|P_X(x),P_Y(y)-P_{XY}(x,y)|\le 2c\sqrt{\xi}|\mathcal{X}||\mathcal{Y}||Z| \\
&\quad+2c\sqrt{\xi+\zeta}\le 4c|\mathcal{X}||\mathcal{Y}||Z|\sqrt{\xi+\zeta} \qquad \text{(A.24)}
\end{aligned}$$

which, together with (A.15) (for $\tau' = P_{S|X'Y'}$, $\tau = P_{S|XY}$), $P_X = P_{X'}$, $P_Y = P_{Y'}$, $\min_x P_X(x) \ge \alpha$, and $\min_y P_Y(y) \ge \beta$, implies

$$\xi+\zeta \ge (4c|\mathcal{X}||\mathcal{Y}||Z|)^{-2}\alpha^4\beta^4\theta^2.$$

■

References

[1] R. Ahlswede, "Multiway communication channels," in *Proc. 2nd Int. Symp. Inform. Theory,* Akademiai Kiado, Budapest, Hungary, 1973, pp. 23–52 (Thakadsor, Armenian SSR 1971).

[2] ——, "The capacity region of a channel with two senders and two receivers," *Ann. Probability,* vol. 2, no. 5, pp. 805–814, 1974.

[3] ——, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrsch. Verw. Geb.,* vol. 44, pp. 159–175, 1978.

[4] ——, "A method of coding and its application to arbitrarily varying channels," *J. Combin. Inform. Syst. Sci.,* vol. 5, no. 1, pp. 10–35, 1980.

[5] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Trans. Inform. Theory,* vol. IT-31, pp. 42–48, 1985.

[6] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory,* vol. 34, pp. 181–193, 1988.

[7] J. A. Gubner, "On the deterministic-code capacity of the multiple-access arbitrarily varying channel," *IEEE Trans. Inform. Theory,* vol. 36, pp. 262–275, 1990.

[8] J. H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 212–226, 1981.

[9] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channels for maximum probability of error," *Z. Wahrsch. Verw. Geb.,* vol. 57, pp. 87–101, 1981.

[10] M. S. Pinsker, "Information and information stability of random variable and processes," (in Russian), vol. 7 of the series, *Problemy Peredacii.* San Francisco, CA: Holden-Day, 1964 (SSSR Moscow, Engl. transl.).

Arbitrarily Varying Multiple-Access Channels—Part II: Correlated Senders' Side Information, Correlated Messages, and Ambiguous Transmission

Rudolf Ahlswede and Ning Cai

***Abstract*—We consider an arbitrarily varying multiple-access channel (AVMAC) $\mathcal{W}$ in which the two senders $\mathcal{X}$ and $\mathcal{Y}$ observe, respectively, the components K^m and L^m of a memoryless correlated source (MCS) $\{(K^m, L^m)\}_{m=1}^{\infty}$ with generic rv's (K, L). In Part I of this work [16], it has been shown for the AVMAC *without* the MCS that in order for the achievable rate region for deterministic codes and the average probability of error criterion to be nonempty, it was sufficient if the AVC were $\mathcal{X}$ nonsymmetrizable, $\mathcal{Y}$ nonsymmetrizable, and $\mathcal{XY}$ nonsymmetrizable. (The necessity of these conditions had been shown earlier by Gubner [7].)**

Let $\mathcal{R}_R(\mathcal{W})$ denote the random code achievable rate region of the AVMAC $\mathcal{W}$. In the present paper, the authors, in effect, trade the loss in achievable rates due to symmetrizability off the gains provided by the MCS. Let $\mathcal{R}(\mathcal{W}, (K, L))$ represent the achievable rate region of the AVC $\mathcal{W}$ *with* MCS, for deterministic codes and the average probability of error criterion. There are two main results:

1) **if $I(K \wedge L) > 0$, then $\mathcal{R}(\mathcal{W}, (K, L))$ has a nonempty interior iff $\mathcal{R}_R(\mathcal{W})$ does too and $\mathcal{W}$ is $\mathcal{XY}$ nonsymmetrizable;**
2) **if $I(K \wedge L) > 0, H(K|L) > 0, H(L|K) > 0$, then the MCS can be transmitted over the AVMAC iff $\mathcal{R}_R(\mathcal{W})$ has a nonempty interior and $\mathcal{W}$ is $\mathcal{XY}$ nonsymmetrizable.**

***Index Terms*— Ahlswede's dichotomy, correlated messages, Ericson's symmetrizability, multiple-access AVC, senders' side information.**

I. Introduction

In [10], we have shown how a memoryless correlated source (MCS) helps the transmission over an arbitrarily varying channel (AVC). Precisely, we established the following result.

Theorem AC_1: For an AVC $\mathcal{W}$ let the sender observe $K^n = K_1, \cdots, K_n$ and let the receiver observe $L^n = L_1, \cdots, L_n$ where $(K^n, L^n)_{n=1}^{\infty}$ is a MCS with generic pair of RV's (K, L) having mutual information $I(K \wedge L) > 0$. Then the capacity $C(\mathcal{W}, (K, L))$ for deterministic codes and the average error criterion equals the random capacity $C_R(\mathcal{W})$.

It serves here as a guide to establish coding theorems for arbitrarily varying multiple-access channels (MAC).

For an arbitrarily varying MAC, abbreviated as AVMAC and defined by a set $\mathcal{W} = \{W(\cdot|\cdot,\cdot,s): s \in \mathcal{S}\}$ of stochastic $(\mathcal{X} \times \mathcal{Y} \to \mathcal{Z})$ matrices, it is natural to investigate the effect of the MCS

$$(K^n, L^n)_{n=1}^{\infty} \text{ with } I(K \wedge L) > 0, \qquad (1)$$

when the $\mathcal{X}$-encoder observes K^n and the $\mathcal{Y}$-encoder observes L^n.

Manuscript received January 7, 1997; revised May 15, 1998. Presented at the IEEE International Symposium on Information Theory, Ulm, Germany, June 29–July 4, 1997.

The authors are with the Fakultät für Mathematik, Universität Bielefeld, Bielefeld, 33501 Germany.

Communicated by S. Shamai, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(99)01408-X.

But let us first recall what has been done in Part I (see [16]). Gubner [7] extended Ericson's symmetrizability to the following conditions.

1) $\mathcal{W}$ is $(\mathcal{X}, \mathcal{Y})$ symmetrizable iff for a stochastic $\sigma: \mathcal{X}\times\mathcal{Y} \to \mathcal{S}$
$$\sum_s W(z|x, y, s)\sigma(s|x', y') = \sum_s W(z|x', y', s)\sigma(s|x, y)$$
for all $x, x' \in \mathcal{X}$, $y, y' \in \mathcal{Y}$, and $z \in \mathcal{Z}$.
2) $\mathcal{W}$ is $\mathcal{X}$ symmetrizable iff for a stochastic $\sigma_1: \mathcal{X} \to \mathcal{S}$
$$\sum_s W(z|x, y, s)\sigma_1(s|x') = \sum_s W(z|x', y, s)\sigma_1(s|x)$$
for all $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$.
3) $\mathcal{W}$ is $\mathcal{Y}$ symmetrizable iff for a stochastic $\sigma_2: \mathcal{Y} \to \mathcal{S}$
$$\sum_s W(z|x, y, s)\sigma_2(s|y') = \sum_s W(z|x, y', s)\sigma_2(s|y)$$
for all $y, y' \in \mathcal{Y}$, $x \in \mathcal{X}$, and $z \in \mathcal{Z}$.

He showed the necessity for non-i), non-ii), and non-iii) for $\mathcal{R}(\mathcal{W})$ to have nonempty interior. He conjectured also sufficiency, which we proved in Part I.

Theorem 1: The achievable rate region $\mathcal{R}(\mathcal{W})$ of $\mathcal{W}$ has nonempty interior iff none of the conditions i)–iii) holds.

$\mathcal{R}(\mathcal{W})$ is now completely known, because we have also Jahn's result [8].

Theorem J: If $\text{int}(\mathcal{R}(\mathcal{W})) \neq \emptyset$, then $\mathcal{R}(\mathcal{W})$ equals the random code capacity region (described also in Part I) $\mathcal{R}_R(\mathcal{W})$.

We can summarize these two results.

Theorem AC_2: For every AVC $\mathcal{W}$ $\text{int}(\mathcal{R}(\mathcal{W})) = \emptyset$ if one of the conditions i)–iii) holds and else $\mathcal{R}(\mathcal{W}) = \mathcal{R}_R(\mathcal{W})$.

Inspection of the proof of Theorem 1, especially *Lemma 1(b)* in Part I, shows that the condition non-iii) is only needed in the decoding rule (III) to decode the message from the $\mathcal{Y}$-decoder. In other words, with the same sets of codewords and the decoding rules (0), (I), and (II) as in Theorem 1, the decoder is still able to decode the message from the $\mathcal{X}$ encoder even when only the condition non-i) and non-ii) hold. It certainly needs the cooperation of the $\mathcal{Y}$ encoder. In this case, the $\mathcal{Y}$-encoder may not or does not want to send any message, and so he only *randomly* chooses a codeword from his codebook. (We must point out that, according to the proof of Theorem 1, Part I, this random choice plays an important role against "the jammer".)

Since we shall use this fact in the sequel, we state it as a theorem.

Theorem 1': Suppose that only the $\mathcal{X}$-encoder wants to send his message and the $\mathcal{Y}$-encoder sends a codeword randomly out of a codebook, a subset of $\mathcal{Y}^n$, which may be undecodable, to help the transmission between the $\mathcal{X}$ encoder and the receiver. Then the $\mathcal{X}$ encoder can send messages with positive rate and arbitrarily small average error probability if neither of the conditions i) and ii) holds. (One of course can interchange the roles of the two encoders.)

Non-i) and non-ii) are actually necessary for the positivity (cf. the analogous results Theorem 3.4 and Lemma 3.5 of [7]), but this is not needed here.

The issue is now to understand what happens in the presence of (1) if one or more of the conditions i)–iii) hold. Are here positive rates possible?

The answer is positive if condition i) does not hold and we have a complete characterization. But first let us give the formal definition of the achievable rate region $\mathcal{R}(\mathcal{W}, (K, L))$.

The $\mathcal{X}$-encoder (or sender) observes the source output $K^m = k^m$ and the $\mathcal{Y}$-encoder observes $L^m = \ell^m$. They encode the message $u \in \mathcal{U}$, respectively, $v \in \mathcal{V}$ into codewords
$$\Phi_u(k^m) \in \mathcal{X}^n \text{ resp. } \Psi_v(\ell^m) \in \mathcal{Y}^n.$$

We call $(\{\Phi_u\}_{u\in\mathcal{U}}, \{\Psi_v\}_{v\in\mathcal{V}}, \{\mathcal{D}_{uv}\}_{u\in\mathcal{U}, v\in\mathcal{V}})$ an $(m, n, M_1, M_2, \lambda)$-code for $\mathcal{W}$ [with side information $(K^m, L^m)_{m=1}^{\infty}$], if $|\mathcal{U}| = M_1$, $|\mathcal{V}| = M_2$, the $\mathcal{D}_{uv} \subset \mathcal{Z}^n$ are pairwise disjoint, and for all $s^n \in \mathcal{S}^n$
$$\frac{1}{|\mathcal{U}||\mathcal{V}|}\sum_{k^m, \ell^m} P_{KL}^m(k^m, \ell^m)\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}} W^n \cdot (\mathcal{D}_{uv}|\Phi_u(k^m), \Psi_v(\ell^m), s^n) > 1 - \lambda. \quad (2)$$

Letting m increase with n proportionally and neglecting the ratio (cf. the next section), we define $\mathcal{R}(\mathcal{W}, (K, L))$ as set of achievable rate pairs.

Theorem 2: Suppose that $I(K \wedge L) > 0$, then $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$ iff $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$ and $\mathcal{W}$ is not $(\mathcal{X}, \mathcal{Y})$-symmetrizable.

Notice that in contrast to Theorem 1 $\mathcal{X}$- and $\mathcal{Y}$ nonsymmetrizability are not necessary for a nonempty interior here!

The proof is done in two steps, where in a first step it is shown that one of the conditions, non-ii) and non-iii), can be dropped.

For people who are not familiar with the elimination technique, to understand our proofs, we explain its role here.

First, let us recall that in [8] Jahn applied the elimination technique of [3] to AVMAC and proved the following. For all $(R_1', R_2') \in \mathcal{R}_R(\mathcal{W})$, arbitrarily small $\lambda, \varepsilon > 0$, and sufficiently large n', one can always find sets of codewords $\mathcal{U}^\gamma \subset \mathcal{X}^{n'}$ for $\gamma \in \Gamma$ and $\mathcal{V}^{\gamma'} \subset \mathcal{Y}^{n'}$ for $\gamma' \in \Gamma'$ and a family of decoding sets
$$\{\mathcal{D}_{u^\gamma, v^{\gamma'}}: u^\gamma \in \mathcal{U}^\gamma, v^{\gamma'} \in \mathcal{V}^{\gamma'}, \gamma \in \Gamma \text{ and } \gamma' \in \Gamma'\}$$
such that $|\Gamma| = |\Gamma'| = n^2$, for fixed (γ, γ'), $\{\mathcal{D}_{u^\gamma, v^{\gamma'}}\}_{u^\gamma\in\mathcal{U}^\gamma, v^{\gamma'}\in\mathcal{V}^{\gamma'}}$ are pairwise disjoint, for all γ, γ', and $s^n \in \mathcal{S}^{n'}$
$$|\mathcal{U}^\gamma| \triangleq M_1' \geq 2^{n'(R_1'-\varepsilon)}$$
$$|\mathcal{V}^{\gamma'}| \triangleq M_2' \geq 2^{n'(R_2'-\varepsilon)}$$
$$|\Gamma|^{-1}|\Gamma'|^{-1}\sum_{\gamma,\gamma'} M_1'^{-1} M_2'^{-1}\sum_{u^\gamma, v^{\gamma'}} W^{n'} \cdot \left(\mathcal{D}_{u^\gamma, v^{\gamma'}}|u^\gamma, v^{\gamma'}, s^{n'}\right) > 1 - \lambda.$$

In fact, one can require the sizes of index sets to be bounded by a constant depending on the probability of error λ (cf. [16]). Let us call this code an eliminated correlated code.

Thus, whenever the achievable rate region has nonempty interior the following strategy applies. For a sufficiently large n' the $\mathcal{X}$ encoder and $\mathcal{Y}$ encoder first pick a $\gamma \in \Gamma\gamma' \in \Gamma'$ randomly and independently and then send them by a code with length $n \triangleq \alpha \log n'$ (for a suitable constant α). Next they encode the message sets $\{1, \cdots, M_1'\}$ and $\{1, \cdots, M_2'\}$ into the sets of codewords $\mathcal{U}^\gamma$ and $\mathcal{V}^{\gamma'}$, respectively, for the chosen γ and γ' and send the corresponding codewords. At the same time the decoder, knowing the indexes γ and γ' (correctly with high probability) uses the decoding sets $\{\mathcal{D}_{u^\gamma, v^{\gamma'}}\}_{u^\gamma\in\mathcal{U}^\gamma, v^{\gamma'}\in\mathcal{V}^{\gamma'}}$ to decode.

Let us return to $\mathcal{R}(\mathcal{W}(K, L))$. If we can show that there exist positive reals R_1, R_2 and A such that for all $\lambda \in (0, 1)$ and sufficiently large n, an $(m, n, M_1, M_2, \lambda)$-code exists with $m \leq An$ and $\frac{1}{n}\log M_i \geq R_1 (i = 1, 2)$ [under the conditions of non-i) and $\text{int}(\mathcal{R}(\mathcal{W}(K, L)) \neq \emptyset)$], then we can use this code to the indices of the above eliminated correlated code with length $n' \sim 2^{n^\beta}$ (β suitable) and then follow the same procedure as before.

We summarize our discussion as follows.

1) Necessary and sufficient for $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$ is the condition (*): There exist positive constants R_1, R_2, and A such that for all $\lambda \in (0, 1)$, and all sufficiently large n $(m, n, M_1, M_2, \lambda)$ codes exist with
$$\geq R_i, \qquad \text{for } i = 1, 2. \quad (3)$$

2)

$$m \le An \quad \text{and} \quad \tfrac{1}{n}\log M_i \\ \text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \supset \text{int}(\mathcal{R}_R(\mathcal{W})) \text{ under } (*). \quad (4)$$

Moreover, the parameter m (or A) is not essential for obtaining a nonempty $\text{int}(R(\mathcal{W}(K, L)))$.

Next, let us consider the transmission of the outputs of a memoryless source $(K^m, L^m)_{m=1}^{\infty}$ with generic (K, L) via an AVMAC. It is already known from [7] that one can never transmit the given length outputs of a memoryless source $(K^m, L^m)_{m=1}^{\infty}$ with arbitrarily small probability of error via an AVMAC no matter how long a channel code one uses whenever K and L are independent and one of the conditions i)–iii) holds. In this case, we say that one cannot transmit the output of the source via the AVMAC or the output is intransmittable; otherwise we say that it is transmittable. On the other hand, is well known, e.g., from [11] and [12], that the dependency structure of a MCS may enlarge the achievable region of a MAC. It is natural to ask whether the dependency structure can change the outputs of a MCS from intransmittable to transmittable over an AVMAC. Our second contribution concerns the question, whether we can transmit the outputs of a MCS with arbitrarily small probability of error via some AVMAC satisfying one or two of the symmetrizability conditions i)–iii). The answer is again positive if condition i) does not hold and again we have a complete characterization.

Theorem 3: Assume that $I(K \wedge L) > 0$ and also $H(K|L), H(L|K) > 0$. Then $(K^m, L^m)_{m=1}^{\infty}$ is transmittable iff $\mathcal{W}$ is not $(\mathcal{X}, \mathcal{Y})$-symmetrizable and $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$.

The readers should notice that we are concerned here only with the possibility of the transmission but not with the achievable rates. The reason is as follows.

Equation (4) shows that $\mathcal{R}_R(\mathcal{W})$ is an inner bound of $\mathcal{R}(\mathcal{W}(K, L))$. In the last section we shall see that $\mathcal{R}(\mathcal{W}(K, L))$ is equal to the achievable rate region of the corresponding compound channel $\{W(\cdot|\cdot, \cdot, \bar{s}) = \sum_s \pi(s)W(\cdot|\cdot, \cdot, s)\colon \pi \in \mathcal{P}(\mathcal{S})\}$ if $K = L$ (almost surely [a.s.]), namely, in this case, $\mathcal{R}(\mathcal{W}(K, L))$ may strictly contain $\mathcal{R}_R(\mathcal{W})$ if the achievable rate region of the compound channel is not equal to $\mathcal{R}_R(\mathcal{W})$. Again, for an n'-length code to achieve this region the outputs $K^m = k^m$, $L^m = \ell^m$ with much smaller length $m \sim \alpha \log n'$ are sufficient. However, in general we do not know $\mathcal{R}(\mathcal{W}(K, L))$ and we even do not know whether the ratio of the lengths of outputs of MCS and the channel code makes a difference. In this sense, even the notation $\mathcal{R}(\mathcal{W}(K, L))$ is not quite precise. However, it is sufficient for our goal to determine whether int$\mathcal{R}(\mathcal{W}(K, L))$ is empty. The situation in the model of Theorem 3 is even more complicated. We notice the achievable region for ordinary MAC is still unknown when it is connected with an MCS (cf. [11] and [12]). So, in both models, we keep the problems to determine the achievable region open.

Theorem 2 is proved in Sections II–IV. We first prove the necessity of $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq 0$ and non-i) in Section II. Then, in Section III, we show that not both non-$\mathcal{X}$ and non-$\mathcal{Y}$ symmetrizability are necessary. Finally, we prove that non-$(\mathcal{X}, \mathcal{Y})$ symmetrizability and $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$ are sufficient for $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$ in Section IV, which finishes our proof of Theorem 2. Applying Theorem 2, we show Theorem 3 also and not in Section IV. Section V contains the discussion about conditions on MCS and the relation between $\mathcal{R}(\mathcal{W}, (K, L))$ and $\mathcal{R}_R(\mathcal{W})$.

II. Necessary Conditions for $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$

We show first that

$$(*) \text{ implies } \text{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset. \quad (5)$$

Indeed, assuming to the opposite that $\text{int}(\mathcal{R}_R(\mathcal{W})) = \emptyset$, then, by convexity of $\mathcal{R}_R(\mathcal{W})$ the intersection with at least one of the axes, say the R_1 axis, equals $\{0\}$.

Let $\mathcal{P}(\mathcal{A})$ denote the set of probability distributions over a finite set $\mathcal{A}$.

We choose now any $P_X \in \mathcal{P}(\mathcal{X})$ and any $P_Y \in \mathcal{P}(\mathcal{Y})$ with the properties

$$P_X(x) > 0 \text{ for all } x \in \mathcal{X} \text{ and } P_Y(y) > 0 \text{ for all } y \in \mathcal{Y}. \quad (6)$$

By the previous intersection property and by the definition of $\mathcal{R}_R(\mathcal{W})$ there exists a $P_S \in \mathcal{P}(\mathcal{S})$ such that

$$I(X \wedge Z|Y) = 0, \qquad \text{if } P_{XYSZ} = P_X P_Y P_S W. \quad (7)$$

However, this implies that (X, Y, Z) forms a Markov chain or, in other terms, for some channel $V\colon \mathcal{Y} \to \mathcal{Z}$ and all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$

$$\begin{aligned} &P_X(x)P_Y(y)\sum_{s\in\mathcal{S}} P_S(s)W(z|x, y, s) \\ &\quad = P_X(x)P_{Y|X}(y|x)V(z|y) \\ &\quad = P_X(x)P_Y(y)V(z|y), \text{ i.e.,} \\ &\sum_{s\in\mathcal{S}} P_S(s)W(z|x, y, s) = V(z|y), \qquad \text{for all } x, y, z. \end{aligned} \quad (8)$$

Let $\overline{W}(\cdot|\cdot, \cdot) = \sum_{s\in\mathcal{S}} P_S(s)W(\cdot|\cdot, \cdot, s)$, then $\overline{W}(\cdot|\cdot, \cdot)$ is in the convex hull $\overline{\mathcal{W}}$ of $\mathcal{W}$ and LHS of (8) is $\overline{W}(z|x, y)$.

Averaging over $\mathcal{S}^n$ with weight P_S^n we verify that an $(m, n, M_1, M_2, \lambda)$ code $(\{\Phi_u\}_{u\in\mathcal{U}}, \{\Psi_v\}_{v\in\mathcal{V}}, \{\mathcal{D}_{uv}\}_{u\in\mathcal{U}, v\in\mathcal{V}})$ for $\mathcal{W}$ satisfies (9), shown at the bottom of the page. Thus, there exists a pair (k^m, ℓ^m) such that $(\{\Phi_u(k^m)\}_{u\in\mathcal{U}}, \{\Psi_v(\ell^m)\}_{v\in\mathcal{V}}, \{\mathcal{D}_{uv}\}_{u\in\mathcal{U}, v\in\mathcal{V}})$ is an (n, M_1, M_2, λ)-code for the MAC $\overline{W}(\cdot|\cdot, \cdot)$. On the other hand by (8) we have for all $P_{XYZ} = P_{XY}\overline{W}(\cdot|\cdot, \cdot)$, $I(X \wedge Z|Y) = 0$ and a positive $\mathcal{X}$-rate would contradict the coding theorem for the MAC in [1]. We have proved (5).

Next we strengthen (5) as follows.

Lemma 3: Condition $(*)$ implies

a) $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$

b) $\mathcal{W}$ is not $(\mathcal{X}, \mathcal{Y})$ symmetrizable.

Proof: It suffices to show that b) must hold even if both senders are combined to one. After this combination we have an ordinary AVC and know from Theorem 2 in [3] that randomization in the encoding (especially the one from our correlated source) does not increase capacity for the average error criterion. In particular it cannot cause an increase from zero to a positive value. So it must be positive without randomization and therefore $\mathcal{W}$ cannot be $(\mathcal{X}, \mathcal{Y})$-symmetrizable (by [5]).

III. Not Both, Non-$\mathcal{X}$- and Non-$\mathcal{Y}$ Symmetrizability Are Necessary for $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$

Our investigation of the side information problem is guided by Theorem 1, which concerns the case without side information. From the three conditions non-i) to non-iii) we have to keep—as shown in Section II–non-i).

$$\frac{1}{M_1 M_2}\sum_{k^m, \ell^m} P_{KL}^m(k^m, \ell^m)\sum_u\sum_v \overline{W}^n(\mathcal{D}_{uv}|\Phi_u(k^m), \Psi_v(\ell^m)) > 1 - \lambda \quad (9)$$

We show in this section that at least one of the two others, say non-ii), can be omitted. The reasoning is this.

We speak of the common randomness A between two persons if both of them know the outcome of a RV A (with a probability close to one) [17]. According to the discussion in the previous section, to transmit messages via an AVMAC with an eliminated correlated code, one only has to establish (independently) common randomnesses between the $\mathcal{X}$ sender and the receiver and between the $\mathcal{Y}$ sender and the receiver, respectively.

We now assume that $\mathrm{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$ and iii) does not hold, then by Theorem $1'$ with the help of $\mathcal{X}$ encoder the $\mathcal{Y}$ encoder can send messages to the receiver. So there is no question to establish the common randomness between the y encoder and the receiver. The only thing that we have to face is to establish the other common randomness. To do this, we can use our AVMAC as an AVC (with two terminals, the $\mathcal{X}$ encoder, and the receiver). On the other hand, in this case, the $\mathcal{Y}$ encoder can send the source output ℓ^m over the channel with high probability correctly to the receiver. This brings us (with high probability) into the situation of side information K^m at the $\mathcal{X}$ sender and L^m at $\mathcal{Y}$ sender and the receiver. Now we just apply the following Lemma to obtain $\mathrm{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$.

Lemma 4: Suppose that $\mathrm{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$ and that $I(K \wedge L) > 0$. Then, for all $(R_1, R_2) \in \mathcal{R}_R(\mathcal{W})$, δ, $\lambda \in (0, 1)$, $A > 0$, there exists for sufficiently large n a code of length n

$$
\begin{aligned}
&(\{\Phi_u\}_{u\in\mathcal{U}}, \{\Psi_v\}_{v\in\mathcal{V}}, \{\mathcal{D}^{\mathcal{Y}}_{uv}(\ell^m)\}_{u\in\mathcal{U},\, v\in\mathcal{V},\, \ell^m\in L^m}) \\
&\quad \text{such that } |\mathcal{U}| = M_1,\ |\mathcal{V}| = M_2 \\
&\frac{1}{n}\log M_1 > R_1 - \delta,\ \frac{1}{n}\log M_2 > R_2 - \delta \\
&\quad m \le An,\ D^{\mathcal{Y}}_{uv}(\ell^m) \cap D^{\mathcal{Y}}_{u'v'}(\ell^m) = \emptyset \text{ for} \\
&(u, v) \neq (u', v'),\ \Phi_u\colon \mathcal{K}^m \to \mathcal{X}^n,\ \Psi_v\colon \mathcal{L}^m \to \mathcal{Y}^m \\
&\quad \text{and for all } s^n \in \mathcal{S}^n \\
&M_1^{-1}M_2^{-1} \sum_{k^m,\,\ell^m} P^m_{KL}(k^m, \ell^m) \sum_u \sum_v W^n \\
&\quad \cdot (D^{\mathcal{Y}}_{uv}(\ell^m) | \Phi_u(k^m), \Psi_u(\ell^m), s^n) > 1 - \lambda. \qquad (10)
\end{aligned}
$$

By symmetry, the role of the $D^{\mathcal{Y}}_{uv}(\ell^m)$'s can be played by the $D^{\mathcal{X}}_{uv}(k^m)$'s.}

Proof: We now are in the situation where the $\mathcal{X}$-sender observes an output $K^m = k^m$ and the $\mathcal{Y}$-sender and the receiver observe another output $L^m = \ell^m$. This is so, because there exists already a common randomness between the $\mathcal{Y}$ sender and the receiver. To apply the elimination technique, we only have to establish common randomness between the $\mathcal{X}$ sender and the receiver, which is done by the communication between them. Thus, the whole procedure of transmission is divided into two blocks. In the first block, only the $\mathcal{X}$ sender transmits messages and the transmission of course needs help from the side information of the MCS and through the cooperation of the $\mathcal{Y}$ sender. Having established the common randomness between the $\mathcal{X}$ sender and the receiver, in the second block the communicators use a much longer eliminated correlated code with average probability of error close to zero. Here, the $\mathcal{Y}$ sender and the receiver partition the source output space with (nearly) equal probabilities as their common randomness, which of course is independent of the common randomness between the $\mathcal{X}$ sender and the receiver. So the main problem is in the first block. The main tool is Theorem AC_1 [10] (see also Section I).

The source output of length m is also divided into two blocks with lengths m_1 and m_2, respectively. (K^{m_1}, L^{m_1}) serves as correlated source between the $\mathcal{X}$ sender and the receiver (see Theorem CA) and L^{m_2} serves as common randomness between the $\mathcal{Y}$ sender and the receiver. They are independent, since the source is memoryless.

We now assume that $\mathrm{int}(\mathcal{R}_R(\mathcal{W})) \neq \emptyset$. Then there are $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y \in \mathcal{P}(\mathcal{Y})$ such that for all $P_S \in \mathcal{P}(\mathcal{S})$, $P_{XYSZ} = P_X P_Y P_S W$

$$
I(X \wedge Z|Y) > 0. \qquad (11)
$$

Since $\max_y \min_{P_S} I(X \wedge Z|Y = y)$ may be zero, the $\mathcal{Y}$-receiver may not always send a fixed codeword. To help the transmission, he has to send a codeword chosen randomly from a code like in Theorem $1'$ in Section I. This seemingly draws us back to the decoding against the average over the messages of the $\mathcal{Y}$-sender as in [1] and [7], which is suboptimal (see the Introduction of Part I and [2]). Consequently, it may seem that we need $\min_{P_S} I(Y \wedge Z) > 0$. However, this is not necessary for (11)!

Fortunately, we have common randomness between the $\mathcal{Y}$-sender and the receiver. It is the time for it to play its role. Since $I(X \wedge YZ) = I(X \wedge Y) + I(X \wedge Z|Y) = I(X \wedge Z|Y)$, (11) implies that the random code capacity $C_R(\mathcal{W}_Y)$ of the AVC $\mathcal{W}_Y\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ is positive, if we define (12), shown at the bottom of the page. By Theorem AC_1 there is an m_1-length code $(\hat{\Phi}_u(k^{m_1}), \hat{\mathcal{D}}_u(\ell^{m_1}))_{u\in\hat{\mathcal{U}}}$ [for all ℓ^{m_1}, $u \neq u'$ $\mathcal{D}_u(\ell^{m_1}) \cap \mathcal{D}_{u'}(\ell^{m_1}) = \emptyset$] with positive rate such that for some $\hat{\eta} > 0$ [see (13) at the bottom of the page. Denote by $\mathcal{E}_u(\ell^{m_1}, y^{m_1}) \triangleq \{z^{m_1}\colon (y^{m_1}, z^{m_1}) \in \hat{\mathcal{D}}_u(\ell^{m_1})\}$ for all $y^{m_1} \in \mathcal{Y}^{m_1}$. Then (12) and (13) imply that

$$
\begin{aligned}
&|\hat{\mathcal{U}}|^{-1} \sum_{u\in\hat{\mathcal{U}}} \sum_{k^{m_1},\,\ell^{m_1}} P^{m_1}_{KL}(k^{m_1}, \ell^{m_1}) \sum_{y^{m_1}} P^{m_1}_Y(y^{m_1}) W^{m_1} \\
&\quad \cdot (\mathcal{E}_u(\ell^{m_1}, y^{m_1}) | \hat{\Phi}_u(k^{m_1}), y^{m_1}, s^{m_1}) \\
&\quad > 1 - 2^{-m_1\hat{\eta}} \text{ for all } s^{m_1} \in \mathcal{S}^{m_1}. \qquad (14)
\end{aligned}
$$

Next, we partition the output space $\mathcal{L}^{m_2}$ of L^{m_2}, the second block of L^m, into $|\mathcal{Y}^{m_1}|$ parts $\mathcal{L}^{m_2}(y^{m_1})$ $(y^{m_1} \in \mathcal{Y}^{m_1})$ for suitable m_1 and m_2 such that for some positive $\tilde{\eta}$

$$
|P^{m_2}_L(\mathcal{L}^{m_2}(y^{m_1})) - P^{m_1}_Y(y^{m_1})| < 2^{-m_2\tilde{\eta}} P^{m_1}_Y(y^{m_1}). \qquad (15)
$$

For an output $L^n = (\ell^{m_1}, \ell^{m_2})$, the $\mathcal{Y}$-sender sends y^{m_1}, if $\ell^{m_2} \in \mathcal{L}^{m_2}(y^{m_1})$. At the same time the $\mathcal{X}$ sender, who observes $K^m = (k^{m_1}, k^{m_2})$, sends $\hat{\Phi}_u(k^{m_1})$ for the message $u \in \hat{\mathcal{U}}$. The receiver knows the output $L^m = (\ell^{m_1}, \ell^{m_2})$ and he decodes accordingly to the decoding sets $\{\mathcal{E}_{u'}(\ell^{m_1}, y^{m_1}) : u' \in \hat{\mathcal{U}}\}$, when $\ell^{m_2} \in \mathcal{L}^{m_2}(y^{m_1})$. Thus, the probability of decoding error is

$$
\begin{aligned}
&1 - |\hat{\mathcal{U}}|^{-1} \sum_{u\in\hat{\mathcal{U}}} \sum_{k^{m_1},\,\ell^{m_1}} P^{m_1}_{KL}(k^{m_1}, \ell^{m_1}) \sum_{y^{m_1}} P^{m_2}_L \\
&\quad \cdot (\mathcal{L}^{m_2}(y^{m_1})) W^{m_1}(\mathcal{E}_u(\ell^{m_1}, y^{m_1}) | \hat{\Phi}_u(k^{m_1}), y^{m_1}, s^{m_1}) \qquad (16)
\end{aligned}
$$

which, by (14) and (15), does not exceed $2^{-m_1\hat{\eta}} + 2^{-m_2\tilde{\eta}}$.

$$
\mathcal{W}_Y \triangleq \{W_Y\colon W_Y(y, z|x, s) = P_Y(y)W(z|x, y, s) \text{ for } x \in \mathcal{X},\ y \in \mathcal{Y},\ z \in \mathcal{Z} \text{ and } s \in \mathcal{S}\} \qquad (12)
$$

$$
|\hat{\mathcal{U}}|^{-1} \sum_{u\in\hat{\mathcal{U}}} \sum_{k^{m_1},\,\ell^{m_1}} P^{m_1}_{KL}(k^{m_1}, \ell^{m_1}) W^{m_1}_Y(\hat{\mathcal{D}}_u(\ell^{m_1}) | \hat{\Phi}_u(k^{m_1}), s^{m_1}) > 1 - 2^{-m_1\hat{\eta}} \text{ for all } s^{m_1} \in \mathcal{S}^{m_1} \qquad (13)
$$

Thus, we build a code of length m_1 with side information (K^m, L^m) for $m = m_1 + m_2$ and positive rate that can be used to send a $u \in \hat{\mathcal{U}}$ from the $\mathcal{X}$ encoder to the receiver. Then the following procedure works. For an $n' \sim 2^{m_1\beta}$ and an eliminated correlated code of length n' as in the discussion at the beginning of the previous section:

1) In the first block, the $\mathcal{X}$ encoder picks up an index $\gamma \in \Gamma(\triangleq \hat{\mathcal{U}})$ randomly and sends it to the receiver.
2) In the second block, the $\mathcal{Y}$ encoder first sends an index $\gamma' \in \Gamma$, which is randomly chosen, with a code of length $m' \sim \alpha' \log n'$. (This can be done by Theorem 1'). After this, the two encoders know γ and γ', respectively, and the receiver knows both. Then an eliminated code (of length n') can be applied. This finishes our proof.

IV. Neither Non-$\mathcal{X}$ nor Non-$\mathcal{Y}$ Symmetrizability Is Necessary for $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \emptyset$

A. Heuristics

This last reduction is based on an idea, which we find most exciting. Without side information from the correlated source none of the senders can transmit messages to the receiver if ii) and iii) hold. So, how can we get started?

Well, under condition non-i) it can still be arranged with sets of codewords $\mathcal{U} \subset \mathcal{X}^n$ and $\mathcal{V} \subset \mathcal{Y}^n$ that the receiver knows with high probability *u or v*, if (u, v) was sent! It follows from Lemma 1(a) in Part I [16] that for a set of codewords in this lemma the condition non-i) and decoding rules (0) and (I) (in Part I) imply that $\mathcal{D}_{uv} \cap \mathcal{D}_{u',v'} = \emptyset$ if $u \neq u'$ and $v \neq v'$.

We speak of *ambiguous* transmission. It can be explained already for the standard MAC. For sets of codewords $\mathcal{U} \subset \mathcal{X}^n$ and $\mathcal{V} \subset \mathcal{Y}^n$ we use for the decoding sets below

$$\mathcal{D}_{uv} = \{z^n: W^n(z^n|u, v) > W^n(z^n|u', v'), \text{ for all } (u', v') \text{ with } u' \neq u \text{ and } v' \neq v\} \tag{17}$$

and notice first that

$$\mathcal{D}_{uv} \cap \mathcal{D}_{u'v'} \neq \emptyset \text{ implies either } u' = u \text{ or } v' = v. \tag{18}$$

Therefore, by forming the list of pairs $\mathcal{F}(z^n) = \{(u, v) \in \mathcal{U} \times \mathcal{V}: z^n \in \mathcal{D}_{uv}\}$ we realize that for any $(u, v), (u', v') \in \mathcal{F}(z^n)$ either $u = u'$ or $v = v'$.

Set $\mathcal{D}_{u,\mathcal{V}\setminus\{v\}} = \bigcup_{v' \neq v} \mathcal{D}_{uv'}$, $\mathcal{D}_{\mathcal{U}\setminus\{u\},v} = \bigcup_{u' \neq u} \mathcal{D}_{u'v}$.

We can partition D_{uv} into $\mathcal{D}_{uv} \cap \mathcal{D}_{u,\mathcal{V}\setminus\{v\}}$, $\mathcal{D}_{uv} \cap \mathcal{D}_{\mathcal{U}\setminus\{u\},v}$, and the rest. In the first set decode for u, in the second decode for v, and make any decision in the rest. (Could vote for both.)

The senders do not know the decision of the decoder, not even with high probability (in contrast to transmission).

We now turn to AVMAC and describe the idea of the proof for Theorem 2. We proceed in two steps.

Step 1: Let the $\mathcal{X}$-sender observe k^m and the $\mathcal{Y}$ sender observe ℓ^m. We first transmit (k^m, ℓ^m) via $\mathcal{W}$ ambiguously, for which non-i) is sufficient. It can be done for a pair $(\mathcal{M}_1, \mathcal{M}_2)$ of sets of independent messages by Lemma 1(a) in Part I with a set of codewords $(\mathcal{U}, \mathcal{V})$ chosen as in Lemma 2 in Part I and decoding rules (O) and (I) in Section II of Part I (cf. the proof of Theorem 1). However, here K^m and L^m are dependent. To match the source with an ambiguous channel code we have to remove the dependency. That is done in Lemma 5 below.

Step 2: Both senders know that with high probability the receiver knows one of the two, of course he knows whether it is k^m or ℓ^m. Therefore the communicators agree on two further blocks of transmission over $\mathcal{W}$. In the first block the $\mathcal{X}$ sender assumes that his k^m is known to the receiver. If true the code of Lemma 4 works and if not we have another chance. In the second block of transmission the $\mathcal{Y}$ encoder operates on the assumption that his ℓ^m is known to the receiver and uses also a code described in Lemma 4. His assumption is now correct with high probability and the receiver knows which of the two codes he should use in his decoding for the messages.

B. Matching the Source

Lemma 5: For any MCS $(K^m, L^m)_{m=1}^{\infty}$ with $I(K \wedge L) > 0$ and AVMAC $\mathcal{W}$, which is not $(\mathcal{X}, \mathcal{Y})$ symmetrizable, there exists a positive constant c such that for all $c^* > c$ and sufficiently large m, $cm \leq n_1 \leq c^* m$, one can find a code of length n_1

$$(\{f(k^m)\}_{k^m \in \mathcal{K}^m}, \{g(\ell^m)\}_{\ell^m \in \mathcal{L}^m} \{\mathcal{D}^*(k^m, \ell^m)\}_{k^m \in \mathcal{K}^m, \ell^m \in \mathcal{L}^m})$$

with the following properties:

1) $f(k^m) \in \mathcal{X}^{n_1}$, $g(\ell^m) \in \mathcal{Y}^{n_1}$ and $\mathcal{D}^*(k^m, \ell^m) \subset \mathcal{Z}^{n_1}$ for all $k^m \in \mathcal{K}^m$ and $\ell^m \in \mathcal{L}^m$ (19)

2) $\mathcal{D}^*(k^m, \ell^m) \cap \mathcal{D}^*(k'^m, \ell'^m) \neq \emptyset$ implies

$$k^m = k'^m \tag{20}$$

or

$$\ell^m = \ell'^m \tag{21}$$

3) for a positive constant θ independent of n_1

$$\sum_{k^m, \ell^m} P_{KL}^m(k^m, \ell^m) W^{n_1}(\mathcal{D}^*(k^m, \ell^m)|f(k^m), g(\ell^m), s^{n_1}) > 1 - 2^{-\theta n_1}, \quad \text{for all } s^{n_1} \in \mathcal{S}^{n_1}. \tag{22}$$

Proof: First of all, the assumption in the proof of Theorem 1 in Part I that two message sets have the same size is not essential. Therefore, following the proof of Theorem 1, especially the application of Lemma 1(a), we obtain that for a non-$(\mathcal{X}, \mathcal{Y})$-symmetrizable $\mathcal{W}$ there exists a $\tilde{R} > 0$ such that for all $0 < \delta < \tilde{R}$, sufficiently large n_1 and integers M_1 and M_2 with $\frac{1}{n_1} \log M_1$, $\frac{1}{n_1} \log M_2 \in [\delta, \tilde{R}]$ there is a code $(\mathcal{U}', \mathcal{V}', \{\mathcal{D}'_{uv}\}_{u \in \mathcal{U}', v \in \mathcal{V}'})$ with $|\mathcal{U}'| = M_1$, $|\mathcal{V}'| = M_2$ such that

$$\mathcal{D}'_{uv} \cap \mathcal{D}'_{u'v'} = \emptyset, \quad \text{for } u \neq u', v \neq v' \tag{23}$$

and for some $\eta' > 0$ and all $s^{n_1} \in \mathcal{S}^{n_1}$

$$M_1^{-1} M_2^{-1} \sum_{u, v} W^{n_1}(\mathcal{D}'^c_{uv}|u, v, s^{n_1}) < 2^{-n_1 \cdot \eta'}. \tag{24}$$

Next, we assume that without loss of generality $H(K|L) > 0$, because otherwise $H(K|L) = H(L|K) = 0$, or $K = L$ (a.s.) and the standard AVC coding theorem in [6] settles our lemma in this case.

Define now

$$h = \min(H(L), H(K|L)) \text{ and } H = \max(H(L), H(K|L)). \tag{25}$$

Then

$$0 < h \leq H. \tag{26}$$

By the coding theorem for correlated sources for any $\delta_1, \delta_2, \delta', \delta'' > 0$ and a sufficiently large m we can find

encoding functions a: $\mathcal{K}^m \to \{0\} \cup \mathcal{U}$ and b: $\mathcal{L}^m \to \{0\} \cup \mathcal{V}$ and a subset $\mathcal{N} \subset \mathcal{U} \times \mathcal{V}$ such that

$$a(k^m) = 0 \text{ iff } k^m \notin \mathcal{T}^m_{K,\delta_1}$$
$$b(\ell^m) = 0 \text{ iff } \ell^m \notin \mathcal{T}^m_{L,\delta_2}$$
$$\Pr((a(K^m), b(L^m)) \in \mathcal{N}) > 1 - 2^{-m\hat{\theta}} \tag{27}$$

for all $(u, v) \in \mathcal{N}$ there is a unique $(k^m, \ell^m) \in \mathcal{T}^m_{K,\delta_1} \times \mathcal{T}^m_{L,\delta_2}$ with $a(k^m) = u$ and $b(\ell^m) = v$,

$$\Pr((a(K^m), b(L^m)) = (u, v)) < 2^{m\delta'} M_1^{-1} M_2^{-1} \tag{28}$$

and

$$H(K|L) \le \frac{1}{m} \log M_1 \le H(K|L) + \delta'' \tag{29}$$
$$H(L) \le \frac{1}{m} \log M_2 \le H(L) + \delta''. \tag{30}$$

Now we choose $\delta'' < \frac{1}{2}$, $\delta' < \frac{1}{2}(H + 1/\tilde{R})\eta'$, $n_1 = \lceil (m(H + 1)/\tilde{R}) \rceil$, $\delta < (mh/n_1)$, and sufficiently large m and n_1. Then for M_1 and M_2 in (29), (30) we have

$$\frac{1}{n_1} \log M_1, \frac{1}{n_1} \log M_2 \in [\delta, \tilde{R}]. \tag{31}$$

Therefore a code satisfying (23) and (24) exists.

Now we define

$$f(k^m) = \begin{cases} u, & \text{if } k^m \in \mathcal{T}^m_{K,\delta_1} \\ & \text{and } a(k^m) = u \\ \text{any fixed } x^{n_1} \in \mathcal{X}^{n_1}, & \text{if } k^m \notin \mathcal{T}^m_{K,\delta_1} \end{cases} \tag{32}$$

$$g(\ell^m) = \begin{cases} v, & \text{if } \ell^m \in \mathcal{T}^m_{L,\delta_2} \\ & \text{and } b(\ell^m) = v \\ \text{any fixed } y^{n_1} \in \mathcal{Y}^{n_1}, & \text{if } \ell^m \notin \mathcal{T}^m_{L,\delta_2} \end{cases} \tag{33}$$

$$\mathcal{D}^*(k^m, \ell^m) = \begin{cases} \mathcal{D}'_{uv}, & \text{if } a(k^m) = u, b(\ell^m) = v \\ & \text{and } (u, v) \in \mathcal{N} \\ \varnothing & \text{otherwise.} \end{cases} \tag{34}$$

Clearly, (19)–(21) hold, where (20), (21) follow from (23).

Finally, by (24), (27), (28), and (32)–(34)

$$\sum_{k^m, \ell^m} P^m_{KL}(k^m, \ell^m) W^{n_1}(\mathcal{D}^{*c}(k^m, \ell^m) | f(k^m), g(\ell^m), s^n)$$
$$\le \sum_{u, v \in \mathcal{N}} \Pr((a(k^m), b(\ell^m)) = (u, v)) W^{n_1}(\mathcal{D}'^c_{uv} | u, v, s^{n_1})$$
$$+ 2^{-m\hat{\theta}} \le 2^{m\delta' - n_1\eta'} + 2^{-m\hat{\theta}} \le 2^{-(1/2)n_1\eta'} + 2^{-m\hat{\theta}} \tag{35}$$

where the last inequality follows from

$$\delta' < \frac{1}{2} \frac{H + 1}{\tilde{R}} \eta' \quad \text{and} \quad n_1 = \left\lceil \frac{m(H + 1)}{\tilde{R}} \right\rceil.$$

In conclusion, (35) yields (22).

Proof of Theorem 2

By Lemma 3 in Section II it remains to be shown that for a not $(\mathcal{X}, \mathcal{Y})$-symmetrizable $\mathcal{W}$ under the assumption $I(K \wedge L) > 0$ $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \varnothing$ implies $\text{int}(\mathcal{R}(\mathcal{W}, (K, L)) \neq \varnothing$. For this we start with an "ambiguous source matching code" of Lemma 5 (Step 1).

Now we proceed with two codes as in Lemma 4 (Step 2). Specifically, the codes are

$$(\{f(k^m)\}_{k^m \in \mathcal{K}^m}, \{g(\ell^m)\}_{\ell^m \in \mathcal{L}^m} \{\mathcal{D}^*(k^m, \ell^m)\}_{k^m \in \mathcal{K}^m, \ell^m \in \mathcal{L}^m})$$

with blocklength n_1, and (with message sets $\mathcal{U}$, $\mathcal{V}$)

$$(\{\Phi^{(1)}_u\}_{u \in \mathcal{U}} \{\Psi^{(1)}_v\}_{v \in \mathcal{V}}, \{\mathcal{D}^{\mathcal{X}}_{uv}(k^m)\}_{u \in \mathcal{U}, v \in \mathcal{V}, k^m \in \mathcal{K}^m})$$

of blocklength n_2 and

$$(\{\Phi^{(2)}_u\}_{u \in \mathcal{U}}, \{\Psi^{(2)}_v\}_{v \in \mathcal{V}}, \{\mathcal{D}^{\mathcal{Y}}_{uv}(\ell^m)\}_{u \in \mathcal{U}, v \in \mathcal{V}, \ell^m \in \mathcal{L}^m})$$

of blocklength n_3.

After (k^m, ℓ^m) has been sent in blocklength n_1 (in Step 1) and the receiver recovered one of them with probability close to 1, in the second block of length n_2 the message at hand, say (u, v), is encoded by our second code and thus transmitted. Finally in the third block of length n_3, the same pair (u, v) is processed by the third code. The total block length is $n = n_1 + n_2 + n_3$ and the total average error probability does not exceed the sum of the error probabilities for the three codes.

Proof of Theorem 3

Next we show Theorem 3, i.e., we have to prove the following. Assuming

$$I(K \wedge L) \neq 0, H(K|L) \neq 0, \text{ and } H(L|K) \neq 0 \tag{36}$$

there is a positive r such that for all $\lambda > 0$ there exists a code

$$(\{\{\Phi(k^m)\}_{k^m \in \mathcal{K}^m}, \{\Psi(\ell^m)\}_{\ell^m \in \mathcal{L}^m} \{\mathcal{D}(k^m, \ell^m)\}_{k^m \in \mathcal{K}^m, \ell^m \in \mathcal{L}^m})$$

of length n, for sufficiently large m with $n \le rm$, $\mathcal{D}(k^m, \ell^m) \cap \mathcal{D}(k'^m, \ell'^m) = \varnothing$ for $(k^m, \ell^m) \neq (k'^m, \ell'^m)$, and for all $s^n \in \mathcal{S}^n$

$$\sum_{k^m \ell^m} P^m_{KL}(k^m, \ell^m) W^n(\mathcal{D}(k^m, \ell^m) | \Phi(k^m), \Psi(\ell^m), s^n) > 1 - \lambda \tag{37}$$

iff non-i) holds and $\text{int}(\mathcal{R}_R(\mathcal{W})) \neq \varnothing$.

The "if" part follows from Theorem 2 and the result of elimination in [8].

Here is our coding procedure.

1) Find an eliminated correlated code with index sets Γ and Γ' as at beginning of Section II. Take a small segment $(K^{m'}, L^{m'})$ of source output sequence (K^n, L^n) as the correlated senders' side information in Theorem 2 and apply this theorem to send $\gamma \in \Gamma$ and $\gamma' \in \Gamma'$, the indices which are randomly and independently chosen by the two encoders.
2) After the receiver learns the indices γ and γ', send (K^m, L^m) by the (γ, γ')th component code of the eliminated correlated code.

For the "only if" part notice first that by the same reason as in Lemma 3, the condition non-i) is necessary. Let us assume $\text{int}(\mathcal{R}_R(\mathcal{W})) = \varnothing$, and without loss of generality assume that the intersection of $\mathcal{R}_R(\mathcal{W})$ with the R_1 axis is $\{0\}$. Let $(\{\Phi(k^m)\}_{k^m \in K^m}, \{\Psi(\ell^m)\}_{\ell^m \in L^m}, \{\mathcal{D}(k^m, \ell^m)\}_{k^m \in \mathcal{K}^m, \ell^m \in \mathcal{L}^m})$ be a code with a λ-probability of error for transmission via $\mathcal{W}$ and let $\overline{W}(\cdot|\cdot, \cdot)$ be the channel in (8) and (9). Then $(\{\Phi(k^m)\}_{k^m \in K^m}, \{\Psi(\ell^m)\}_{\ell^m \in L^m}, \{\mathcal{D}(k^m, \ell^m)\}_{k^m, \ell^m})$ is a code with λ-probability of error for $\overline{W}(\cdot|\cdot, \cdot)$, too. Introduce $X^n = \Phi(K^m)$, $Y^n = \Psi(L^m)$ and let Z^n be the output RV of $\overline{W}^n(\cdot|\cdot, \cdot)$. Then it follows by Fano's Lemma and standard calculations that

$$\begin{aligned} mH(K|L) = H(K^m|L^m) &\le H(K^m|Y^n) \\ &\le [I(K^m \wedge Z^n|Y^n) + \lambda \log |\mathcal{L}| + 1 \\ &\le I(X^n \wedge Z^n|Y^n) + \lambda \log |\mathcal{L}| + 1 \\ &\le \sum_{t=1}^{n} I(X_t \wedge Z_t|Y_t) + \lambda \log |\mathcal{L}| + 1 \\ &= \lambda \log |\mathcal{L}| + 1 \end{aligned}$$

which contradicts (36).

V. Discussion

A. Conditions for the MCS

The problem of correlated side information and correlation in messages effects on the AVMAC to have positive rates have been solved except for a complete analysis of the conditions

$$I(K \wedge L) > 0 \tag{38}$$

and

$$H(K|L) > 0,\ H(L|K) > 0. \tag{39}$$

Actually, we used them in order to avoid trivial cases.

Another issue is a comparison of the effects on positivity of rates due to the dependency in messages and the dependency in source helpers.

For a classical MAC, we know from [11] and [12] that dependency of messages enlarges the rate regions and that obviously the side information even of completely correlated sources at two senders does not.

Similarly, we encounter that for the AVMAC dependency of messages gives more help for obtaining regions with nonempty interior. Here are specific observations.

1) If (38) is violated, then K and L are independent and the problem addressed in Theorems 2 and 3 reduces to that in Theorem 1. So the conditions in Theorems 2 and 3 are no longer sufficient, the additional conditions non-ii) and non-iii) are needed. Indeed the code in (5) becomes a code for $\mathcal{W}$ with independent randomizations of the encoders, which can be shown—in the same way as in [3]—not to enlarge the rate region. Therefore non-ii) and non-iii) are needed.
2) Assume that (39) does not hold. Then Theorem 2 is still true, but Theorem 3 will change. It can easily be seen that the MCS $(K^m, L^m)_{m=1}^{\infty}$ can be transmitted via $\mathcal{W}$ iff i) does not hold and the intersection of $\mathcal{R}_R(\mathcal{W})$ with the R_1 axis (R_2 axis) is unequal $\{0\}$ under the assumption $H(K|L) > 0$ and $H(L|K) = 0$ ($H(L|K) > 0$ and $H(K|L) = 0$). The proof of sufficiency follows from the proof of sufficiency for Theorem 3. Let now the intersection of $\mathcal{R}_R(\mathcal{W})$ with the R_1 axis be not $\{0\}$ and $H(K|L) > 0$, $H(L|K) = 0$. Then $\max_{P_{XY}} \min_S I(XY \wedge Z) > 0$, $\max_{P_X, P_Y} \min_S I(X \wedge Z|Y) > 0$, and L^m is a function of K^m, where the first maximum is taken over all input distributions, the second one is taken over all independent distributions and both minima are taken over the convex hull of $\mathcal{W}$. Now both senders know L^m. They first cooperate to send L^m using $\mathcal{W}$ as an AVC with two terminals. Thus the common randomness (between the senders and receiver) has been established. So with the help of the common randomness and the second sender, the first sender can send K^m with a code like that in Lemma 4.

Finally, when $H(K|L) = H(L|K) = 0$, i.e., $K^m = L^m$ (a.s.) the two senders "become one," and the problem in Theorem 3 becomes a two-terminal AVC problem. Here the condition non-i) is sufficient and necessary for positive capacity. The following simple example shows that in this case when $\mathcal{R}_R(\mathcal{W}) = \{(0, 0)\}$ even a zero-error code with "full rate" may exist. This demonstrates that dependency of messages can be more useful than side information.

Example—$\mathcal{K} = \mathcal{L} = \mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{S} = \{0\}$: $\Pr(K = L = 0) = p$, $\Pr(K = L = 1) = 1 - p$, $p \in (0, 1)$, $W(z|x, y, s) = 1$ if $z = x$ and $s = 0$ or $z = y$ and $s = 1$. Then $\mathcal{R}_R(\mathcal{W}) = \{(0, 0)\}$ but $\Phi(k^m) = k^m$, $\Psi(\ell^m) = \ell^m$

$$\mathcal{D}(k^m, \ell^m) = \begin{cases} k^m, & \text{if } k^m = \ell^m \\ \varnothing & \text{otherwise} \end{cases} \tag{40}$$

is an m-length error-free code.

B. The Rate Region

Assume now $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \varnothing$. By (4) $\mathcal{R}(\mathcal{W}, (K, L)) \subset \mathcal{R}_R(\mathcal{W})$. A natural question is whether they are always equal. The answer is negative. Let us consider the extremal case that $K = L$ (a.s.). First we notice that $\mathcal{R}_R(\mathcal{W})$ may not be equal to (or be contained properly by) the rate region $\mathcal{R}_C(\mathcal{W})$ of the corresponding compound channel (cf. [2] and [8] to see that they are formally different). This is different from the two-terminal case.

Next, we show that in case $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \varnothing$ and $K = L$ (a.s), we have $\mathcal{R}_R(\mathcal{W}) = \mathcal{R}_C(\mathcal{W})$. By the robustification technique in [13], one can obtain from a code $(\mathcal{U}, \mathcal{V}, \{\mathcal{D}_{uv}\}_{u \in \mathcal{U}, v \in \mathcal{V}})$ for the compound channel a family $\mathcal{C} = \{(\mathcal{U}^\gamma, \mathcal{V}^\gamma, \{\mathcal{D}^\gamma_{u,v}\}_{u \in \mathcal{U}, v \in \mathcal{V}}) : \gamma \in \Gamma\}$ of codes for the AVMAC $\mathcal{W}$ with the same rate and a relatively small $|\Gamma|$, such that for all s^n

$$|\Gamma|^{-1}|\mathcal{U}|^{-1}|\mathcal{V}|^{-1} \sum_{\gamma \in \Gamma} \sum_{u^\gamma, v^\gamma} W^n(\mathcal{D}^\gamma_{u^\gamma, v^\gamma} | u^\gamma, v^\gamma, s^n)$$

is close to one, if the error probability of the code $(\mathcal{U}, \mathcal{V}, \{\mathcal{D}_{uv}\}_{u \in \mathcal{U}, v \in \mathcal{V}})$ is close to zero.

This is done by a random permutation. We omit the details because they are exactly the same as in Section II of [14]. Notice that an analogous result for random correlated codes was obtained in [15]. Now let us return to our question. The two senders have a common randomness since $K = L$ (a.s.) and they can send it to the receiver since $\text{int}(\mathcal{R}(\mathcal{W}, (K, L))) \neq \varnothing$. Thus, this family $\mathcal{C}$ of codes can be used and we are done.

The opposite containment is obvious.

References

[1] R. Ahlswede, "Multy-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory*, Akademiai Kiado, Budapest, Hungary, 1973, pp. 23–52 (Thakadsor, Armenian SSR 1971).

[2] ——, "The capacity region of a channel with two senders and two receivers," *Ann. Probability*, vol. 2, no. 5, pp. 805–814, 1974.

[3] ——, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrsch. Verw. Geb.*, vol. 44, pp. 159–175, 1978.

[4] ——, "A method of coding and its application to arbitrarily varying channels," *J. Combin. Inform. Syst. Sci.*, vol. 5, no. 1, pp. 10–35, 1980.

[5] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 42–48, 1985.

[6] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, 1988.

[7] J. A. Gubner, "On the deterministic-code capacity of the multiple-access arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 262–275, 1990.

[8] J. H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 212–226, 1981.

[9] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channels for maximum probability of error," *Z. Wahrsch. Verw. Geb.*, vol. 57, pp. 87–101, 1981.

[10] R. Ahlswede and N. Cai, "Correlated sources help the transmission over AVC," preprint 95–106, SFB 343, "Diskrete strukturen in der mathematik," Universität Bielefeld, 1995, *IEEE Trans. Inform. Theory*, vol. 135, pp. 37–67, 1997.

[11] 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.

[12] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 396–412, 1983.

[13] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user source coding," *J. Combin. Inform. Syst. Sci.—Part II*, vol. 5, pp. 220–268, 1980.

[14] R. Ahlswede and N. Cai, "Two proofs of Pinsker's conjecture concerning arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1647–1649, 1991.

[15] J. A. Gubner and B. L. Hughes, "Dependent codebooks can increase capacity," preprint, 1995.
[16] R. Ahlswede and N. Cai, "Arbitrarily varying multiple-access channels—Part I. Ericson's symmetrizability is adequate, Gubner's conjecture is true," preprint 96–068, SFB 343, "Diskrete strukturen in der mathematik," Universität Bielefeld, *IEEE Trans. Inform. Theory*, to be published.
[17] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography—Part I: Secret sharing," *IEEE Trans. Information Theory*, vol. 39, pp. 1121–1132, 1993; "Part II: CR capacity," preprint 95–101, SFB 343, "Diskrete strukturen in der mathematik," Universität Bielefeld, *IEEE Trans. Inform. Theory*, vol. 44, pp. 55–62, 1998.

Disjointness of Random Sequence Sets with Respect to Distinct Probability Measures

Te Sun Han, *Fellow, IEEE*, and Mitsuru Hamada

***Abstract*—It is shown that the set of deterministic random sequences (of symbols from a finite alphabet) with respect to a computable probability measure μ, in Martin-Löf's sense, and the set of deterministic random sequences with respect to another computable probability measure ν are disjoint if μ and ν are different and the measures are either i.i.d. or homogeneous finite-order irreducible Markov measures.**

***Index Terms*—Disjointness, Kullback–Leibler information, mth-order composition (type), probability measure, random sequence, stochastic typicality.**

I. Introduction

This correspondence shows random sequences possess *stochastic typicality* and, as a consequence, that the set of deterministic random sequences with respect to a computable probability measure μ and the set of deterministic random sequences with respect to another computable probability measure ν are disjoint if μ and ν are *distinct* in the sense to be specified below. Here the randomness is defined in an equivalent way to Martin-Löf's [1], where he defined it using the notion of statistical hypothesis testing. As a simple example, consider sequences from two-symbol alphabet $\{0,1\}$ and let P be a probability distribution on $\{0,1\}$. Intuitively, a deterministic random sequence with respect to P is one that looks as if the symbols in the sequence were drawn independent and identically distributed (i.i.d.) according to P. Therefore, random sequences with respect to P should have stochastic typicality, i.e., the property that the relative frequency of occurrences of the symbol "0" in the first n symbols approaches $P(0)$ as n goes to infinity. Consequently, if P and Q are distinct distributions, the random sequence set with respect to P and that with respect to Q ought to be disjoint. Our assertion extends to the case where the probability measure is "Markov." In this case, it looks less apparent that distinctness between two measures results in the disjointness of the two random sequence sets.

In what follows, we first discuss the i.i.d. case, where the basic idea we use to derive the typicality is attributed to Cover and Thomas [2, Theorem 7.5.2]. They defined the randomness ("incompressibility" in their words) in terms of Kolmogorov complexity and showed the typicality of random sequences by using the compositions (or types) of sequences as well as their empirical entropies in order to estimate Kolmogorov complexity. While they treated only the uniform distribution case, we generalize their result to the case where the sequences are random with respect to arbitrary distributions. In our argument, the Kullback–Leibler information (divergence) plays an important role. Secondly, we treat the Markov case, where the first half of the discussion to derive the typicality is parallel to that in the i.i.d. case. In the Markov case, we use the method for counting the sequences of a fixed "Markov type" introduced by Davisson, Longo, and Sgarro [3]. Then, our main assertion, i.e., the disjointness of the random sequence set with respect to a Markov measure and that with respect to another Markov measure, follows from the typicality with a short elementary proof, although not very immediately.

We conventionally denote by $\boldsymbol{N}$ and $\boldsymbol{R}$ the set of natural numbers and the set of real numbers, respectively. Throughout this correspondence, $\mathcal{X}$ is an arbitrary finite set (alphabet) with r elements, $r \geq 2$, and $\mathcal{X}^n, \mathcal{X}^*, \mathcal{X}^\infty$ are the set of all n-length sequences, the set of all finite-length sequences of symbols from $\mathcal{X}$, and the set of all one-way infinite sequences of the form $x_1 x_2 \cdots$ $(x_i \in \mathcal{X}, i = 1, 2, \cdots)$, respectively. For a given infinite sequence $x \in \mathcal{X}^\infty$ we denote by x_i the ith symbol of x. We use the notation that $x_i^j = x_i x_{i+1} \cdots x_j$ for $j \geq i$. We fix a sigma-algebra $\mathcal{F}$ on $\mathcal{X}^\infty$ generated by all cylinder sets of the form

$$\{x \in \mathcal{X}^\infty | x_1^n = a_1^n\} \quad (\forall a_1^n \in \mathcal{X}^n, \forall n \in \boldsymbol{N}).$$

In our following discussion we are only interested in measures of cylinder sets of this form so that we write $\mu(a_1^n)$ for

$$\mu(\{x \in \mathcal{X}^\infty | x_1^n = a_1^n\}), \qquad n \in \boldsymbol{N}$$

where μ is a probability measure on $(\mathcal{X}^\infty, \mathcal{F})$. This notation suggests that a measure μ can be identified with a function μ: $\mathcal{X}^* \to [0,1]$. The computability of a measure μ is understood as that of the corresponding function μ: $\mathcal{X}^* \to [0,1]$. A probability distribution on a finite set $\mathcal{A}$ is a function p defined on $\mathcal{A}$ such that $p(a) \geq 0$ for all $a \in \mathcal{A}$ and $\Sigma_{a \in \mathcal{A}} p(a) = 1$. We denote by $K(x_1^n)$ the Kolmogorov complexity of a sequence $x_1^n \in \mathcal{X}^n$ with respect to a *universal* Turing machine. We restrict ourselves to treating Turing machines of the following property: the programs that lead to halting computation when they are fed into a fixed machine form a prefix-free set (i.e., a set each element of which is not a prefix of another). Namely, we consider the same machines as described in [2, Sec. 7.1] except that the alphabet for writing on the input and output tape is $\mathcal{X}$ instead of $\{0,1\}$.

Probably the most popular notion of randomness is the one defined by Martin-Löf [1]. The following definition due to Chaitin [4] is known to be equivalent to Martin-Löf's (see, [5, Corollary 4.5]; see also [6, Ch. 7] for the proof in the case of i.i.d. uniform measures; in fact, Chaitin's original definition is for this special case).

Manuscript received December 20, 1997; revised June 10, 1998. The work of M. Hamada was supported by a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists. The material in this correspondence was presented in part at the 1997 IEEE International Symposium on Information Theory, Ulm, Germany, July 1997.

The authors are with the Graduate School of Information Systems, The University of Electro-Communications, Chofugaoka 1-5-1, Chofu, Tokyo 182-8585, Japan.

Communicated by K. Zeger, Associate Editor at Large.
Publisher Item Identifier S 0018-9448(99)00058-9.