An Improved Achievable Region for the Discrete Memoryless Two-User Multiple-Access Channel With Noiseless Feedback

Shraga I. Bross, *Member, IEEE,* and Amos Lapidoth, *Fellow, IEEE*

Abstract—An achievable region for the two-user discrete memoryless multiple-access channel (DMMAC) with noiseless feedback is proposed. The proposed region includes the Cover–Leung region, with the inclusion being, for some channels, strict. This inner bound is demonstrated for the ideal two-user Poisson multiple-access channel with noiseless feedback, in which case it is shown to improve on the Cover–Leung rate-sum.

Index Terms—Feedback capacity, ideal Poisson multiple-access channel (MAC), optical code-division multiple access (CDMA), two-user discrete memoryless multiple-access channel (DMMAC) with noiseless feedback.

I. INTRODUCTION

NO single-letter expression for the capacity region of a general discrete memoryless (DM) multiple-access channel (MAC) with feedback is known. The observation that this region can be strictly larger than the capacity region of the MAC without feedback is due to Gaarder and Wolf [1]. Cover and Leung established an inner bound to the capacity region for the MAC with feedback in [2]. However, as demonstrated by Ozarow [7], this inner bound is not always tight. Nevertheless, Willems has shown [4] that the Cover–Leung inner bound is tight if the channel satisfies the condition that at least one of the encoders can determine the symbol produced by the other encoder based on the channel output and the symbol it produced itself. Willems's condition was subsequently somewhat weakened by Hekstra and Willems [5].

Recently, Kramer [8], [9] used the notion of *directed information* to derive an expression for the capacity region of the MAC with feedback. This expression, however, is an incomputable non-single-letter expression.

In this paper, we shall propose a new achievable region with a single-letter characterization. The proposed region contains the Cover–Leung region with the inclusion being, for some channels, strict.

An example of a MAC for which the proposed region is strictly larger than the Cover–Leung region is the binary MAC

where the probability of receiving the symbol "1" is proportional to the sum of the channel inputs

$$p(1|1,0) = p(1|0,1) = \frac{1}{2}p(1|1,1) = q$$
$$p(1|0,0) = 0 \qquad (1)$$

where $0 < q < 1/2$. Notice that for this channel Willems's condition is not met: no encoder can determine the symbol produced by the other encoder based on the channel output and the symbol it produced itself. Indeed, if an encoder transmits "1" then it cannot—even with the aid of the output—infer what was transmitted by the other encoder. Thus, the fact that for this channel the Cover–Leung region is not tight does not contradict Willems's result [4]. Observe, however, that Willems's condition is "partially fulfilled" in the sense that if the encoder produces a "0" and if it observes that the channel output is "1," it can infer that the other encoder produced a "1." This observation makes this channel particularly suitable for demonstrating that our proposed region may be strictly larger than the Cover–Leung region.

The rest of this paper is organized as follows. In Section II, we introduce the two-user channel with feedback and recall some basic results that will be needed to establish an achievable rate region. In Section III, we present our main results and in Sections IV and V we prove the achievability of these results. Finally, Section VI demonstrates how our code construction can be applied to the ideal two-user Poisson MAC to improve on the Cover–Leung inner bound.

II. PRELIMINARIES

A. Notation

Henceforth, we adopt the following notation conventions. Random variables will be denoted by capital letters, while their realizations will be denoted by the respective lower case letters. Whenever the dimension of a random vector is clear from the context the random vector will be denoted by a bold face letter, that is, $\boldsymbol{X}$ denotes the random vector $(X_1, X_2, \ldots, X_n)$, and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ will designate a specific sample value of $\boldsymbol{X}$. However, in those cases where we find it important to emphasize explicitly the dimension of a random vector—$X_1^{n_1}$ shall denote the random vector $(X_{1,1}, X_{1,2}, \ldots, X_{1,n_1})$. The alphabet of a scalar random variable X will be designated by a calligraphic letter $\mathcal{X}$. The n-fold Cartesian power of a generic alphabet $\mathcal{V}$, that is, the set of all n-vectors over $\mathcal{V}$, will be denoted $\mathcal{V}^n$.

The transmitters in a two-user MAC will be usually indexed by $\mu \in \{1, 2\}$. We shall denote by $\bar{\mu}$ the element in $\{1, 2\}$ that is not equal to μ.

B. Achievable Rates

Definition 1: A two-user DMMAC is a quadruple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ where $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite sets corresponding to the input alphabets, the finite set $\mathcal{Y}$ is the output alphabet, and $p(\cdot|x_1, x_2)$ is a collection of probability mass functions on $\mathcal{Y}$ indexed by the input symbols $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. These probability mass functions extend to n-tuples according to the memoryless law

$$P(y_k|x_{1,1}, \ldots, x_{1,k}, x_{2,1}, \ldots, x_{2,k}, y_1, \ldots, y_{k-1}) = \\ p(y_k|x_{1,k}, x_{2,k})$$

where $x_{1,k}$, $x_{2,k}$, and y_k denote the inputs and output of the channel at time k.

Feedback is incorporated into the DMMAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ via the definition of a code. An (M_1, M_2, n) code for the MAC with feedback is defined as follows.

1) A collection of encoding functions

$$f_{\mu,k} : \{1, \ldots, M_\mu\} \times \mathcal{Y}^{k-1} \to \mathcal{X}_\mu, \\ \mu = 1, 2, \ k = 1, 2, \ldots, n$$

where $f_{\mu,k}(w_\mu, y^{k-1})$ is the symbol transmitted at time k by Encoder μ when attempting to convey Message w_μ to the receiver after obtaining the previous $k-1$ output symbols $y^{k-1} = (y_1, \ldots, y_{k-1})$. Here M_μ denotes the number of different messages that Encoder μ can transmit.

2) A decoding function

$$\phi : \mathcal{Y}^n \to \{1, \ldots, M_1\} \times \{1, \ldots, M_2\}.$$

We shall use the average probability of error criterion assuming that the messages (W_1, W_2) are drawn according to a uniform distribution over $\{1, \ldots, M_1\} \times \{1, \ldots, M_2\}$. Consequently, the error probability for the code is defined as

$$P_e = \Pr\{\phi(Y_1, Y_2, \ldots, Y_n) \neq (W_1, W_2)\} \\ = \frac{1}{M_1 M_2} \sum_{w_1=1}^{M_1} \sum_{w_2=1}^{M_2} \\ \cdot \Pr\{\phi(\boldsymbol{Y}) \neq (w_1, w_2)|W_1 = w_1, W_2 = w_2\}.$$

A rate pair (R_1, R_2) nats per channel use is achievable for the memoryless MAC with feedback if there exists a sequence of (M_1, M_2, n) codes with

$$R_1 \leq \frac{1}{n} \ln M_1, \quad R_2 \leq \frac{1}{n} \ln M_2$$

such that $P_e \to 0$ as $n \to \infty$. (Throughout, we use natural logarithms). The feedback *capacity region* $\mathcal{C}_f$ is the closure of the set of all achievable rates.

C. Weak Typicality

Let $\{X_1, X_2, \ldots, X_k\}$ denote a finite collection of discrete random variables with some joint distribution $P(x_1, x_2, \ldots, x_k)$ with $(x_1, x_2, \ldots, x_k) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$. Let Z denote an ordered nonempty subset of these random variables and consider n independent copies of Z. Thus, with $\boldsymbol{Z} \triangleq (Z_1, Z_2, \ldots, Z_n)$

$$\Pr\{\boldsymbol{Z} = \boldsymbol{z}\} = \prod_{j=1}^{n} \Pr\{Z_j = z_j\}.$$

Furthermore, let

$$H(Z) \triangleq \sum_z - \Pr\{Z = z\} \ln(\Pr\{Z = z\}).$$

Definition 2: The set $\mathcal{A}_\epsilon$ of ϵ-typical n-sequences is defined by (see [3, Ch. 3])

$$\mathcal{A}_\epsilon \triangleq \mathcal{A}_\epsilon(X_1, X_2, \ldots, X_k) \\ \triangleq \Bigg\{ (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) : \\ \left| -\frac{1}{n} \ln(\Pr\{\boldsymbol{Z} = \boldsymbol{z}\}) - H(Z) \right| \leq \epsilon, \\ \forall Z \subseteq \{X_1, X_2, \ldots, X_k\} \Bigg\}.$$

Let $\mathcal{A}_\epsilon(Z)$ be defined similar to $\mathcal{A}_\epsilon$ but now with constraints corresponding to all nonempty subsets of Z. We recall now three basic lemmas (for the proofs we refer to [3]).

Lemma 1: For any $\epsilon > 0$ the following statements hold for every integer $n \geq 1$:

1) If $\boldsymbol{z} \in \mathcal{A}_\epsilon(Z)$, then

$$\exp(-n(H(Z) + \epsilon)) \leq \Pr\{\boldsymbol{Z} = \boldsymbol{z}\} \\ \leq \exp(-n(H(Z) - \epsilon)).$$

2) If $Z_1, Z_2 \subseteq \{X_1, X_2, \ldots, X_k\}$ and $(\boldsymbol{z}_1, \boldsymbol{z}_2) \in \mathcal{A}_\epsilon(Z_1 \cup Z_2)$, then

$$\exp(-n(H(Z_1|Z_2) + 2\epsilon)) \leq \Pr\{\boldsymbol{Z_1} = \boldsymbol{z_1}|\boldsymbol{Z_2} = \boldsymbol{z_2}\} \\ \leq \exp(-n(H(Z_1|Z_2) - 2\epsilon)).$$

Moreover, the following statements hold for every sufficiently large n:

3) $\Pr\{\mathcal{A}_\epsilon(Z)\} \geq 1 - \epsilon$;

4) $(1 - \epsilon) \exp(n(H(Z) - \epsilon)) \leq |\mathcal{A}_\epsilon(Z)| \leq \exp(n(H(Z) + \epsilon))$.

Lemma 2: Let the discrete random variables X, Y have joint distribution $P_{X,Y}(x, y)$. Let X' and Y' be independent with the marginals

$$P_{X'}(x) = \sum_y P_{X,Y}(x, y) \\ P_{Y'}(y) = \sum_x P_{X,Y}(x, y).$$

Let the pair of n-length random vectors $(\boldsymbol{X}, \boldsymbol{Y})$ be drawn

$$(\boldsymbol{X}, \boldsymbol{Y}) \sim \prod_{k=1}^{n} P_{X,Y}(x_k, y_k)$$

and the pair of n-length vectors $(\boldsymbol{X'}, \boldsymbol{Y'})$ be drawn

$$(\boldsymbol{X'}, \boldsymbol{Y'}) \sim \prod_{k=1}^{n} P_{X'}(x_k) P_{Y'}(y_k).$$

Then

$$\Pr\{(\boldsymbol{X'}, \boldsymbol{Y'}) \in \mathcal{A}_\epsilon(X,Y)\} \leq \exp\{-n[I(X;Y) - \epsilon]\}.$$

Lemma 3: Let the discrete random variables X, Y, Z have joint distribution $P_{X,Y,Z}(x, y, z)$. Let X' and Y' be conditionally independent given Z, with the marginals

$$P_{X'|Z}(x|z) = \sum_{y} P_{X,Y,Z}(x, y, z)/P_Z(z)$$

$$P_{Y'|Z}(y|z) = \sum_{x} P_{X,Y,Z}(x, y, z)/P_Z(z).$$

Let

$$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \sim \prod_{k=1}^{n} P_{X,Y,Z}(x_k, y_k, z_k)$$

and

$$(\boldsymbol{X'}, \boldsymbol{Y'}, \boldsymbol{Z}) \sim \prod_{k=1}^{n} P_{X'|Z}(x_k|z_k) P_{Y'|Z}(y_k|z_k) P_Z(z_k).$$

Then

$$\Pr\{(\boldsymbol{X'}, \boldsymbol{Y'}, \boldsymbol{Z}) \in \mathcal{A}_\epsilon(X,Y,Z)\} \leq \exp\{-n[I(X;Y|Z) - \epsilon]\}.$$

III. MAIN RESULTS

The achievable region for the MAC with feedback that we present is based on the analysis of a transmission system that can be roughly divided into "two-way" phases and "multiple-access" phases. To simplify the exposition we shall first present a result that is related to the two-way phase and only later present the main result regarding the achievable region.

Consider a MAC $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2))$ and assume that before the two-way phase begins the n-length sequences (S_1^n, S_2^n, S_d^n) are generated independently and identically distributed (i.i.d.) according to some joint law $P_{S_1,S_2,S_d}(s_1, s_2, s_d)$ so that

$$(S_1^n, S_2^n, S_d^n) \sim \prod_{k=1}^{n} P_{S_1,S_2,S_d}(s_{1,k}, s_{2,k}, s_{d,k}) \qquad (2)$$

where $s_{1,k}$ (respectively, $s_{2,k}$, and $s_{d,k}$) denotes the time-k component of the sequence S_1^n (respectively, S_2^n and S_d^n) and is assumed to take value in some finite set $\mathcal{S}_1$ (respectively, $\mathcal{S}_2$ and $\mathcal{S}_d$). The sequences (S_μ^n, S_d^n) are then made available to Encoder μ, $\mu = 1, 2$, whereas only the sequence S_d^n is made

available to the decoder. Encoder μ then applies componentwise some deterministic function

$$g_\mu : \mathcal{S}_\mu \times \mathcal{S}_d \to \mathcal{V}_\mu, \qquad \mu = 1, 2 \qquad (3)$$

(where $\mathcal{V}_\mu$ is some finite set) to produce the sequence

$$\boldsymbol{v}_\mu(S_\mu^n, S_d^n) = (g_\mu(S_{\mu,1}, S_{d,1}), g_\mu(S_{\mu,2}, S_{d,2}), \dots, \\ g_\mu(S_{\mu,n}, S_{d,n})), \qquad \mu = 1, 2.$$

In the two-way phase, Encoder μ, whose side information consists of (S_μ^n, S_d^n), wishes to provide Encoder $\bar{\mu}$ with the sequence $\boldsymbol{v}_\mu(S_\mu^n, S_d^n)$ in order to convey something (possibly everything) about its private side information S_μ^n. The functions $\{\boldsymbol{v}_\mu\}$ allow for the possibility that the two encoders may wish to exchange not S_1^n and S_2^n but some functions thereof.

In the two-way phase, the MAC with feedback is used n_{TW} times in order to exchange the sequence $\boldsymbol{v}_1(S_1^n, S_d^n)$ and the sequence $\boldsymbol{v}_2(S_2^n, S_d^n)$ between the two encoders. After the exchange (which should be error free with high probability) the decoder forms a list

$$\mathcal{L}_{1,2} = \{(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))\}$$

of possible sequence pairs, based on its knowledge of the sequence S_d^n and upon the outputs that were produced during the exchange $\tilde{y}_1, \dots, \tilde{y}_{n_{TW}}$. Notice that this list will typically be of an exponential size even for a successful exchange of information, because compared to the encoders the decoder is hindered: at the beginning of the two-way phase only the sequence S_d^n is made available to the decoder.

We are interested in the asymptotic ratio of the number of channel uses n_{TW} to n that are needed to guarantee a reliable exchange of the sequences $(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))$ and in the exponential rate of growth R_L of the list size needed by the decoder to guarantee that the correct pair $(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))$ is (with high probability) in the list $\mathcal{L}_{1,2}$.

More formally, for a MAC with feedback $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2))$ and sequences (S_1^n, S_2^n, S_d^n) generated i.i.d. from the law $P_{S_1,S_2,S_d}(s_1, s_2, s_d)$ we consider encoding functions

$$\tilde{f}_{\mu,k} : \mathcal{S}_\mu^n \times \mathcal{S}_d^n \times \tilde{\mathcal{Y}}^{k-1} \to \tilde{\mathcal{X}}_\mu, \quad \mu = 1, 2, \ k = 1, \dots, n_{TW} \tag{4}$$

that given the sequences available to Encoder μ before the two-way phase began and the channel outputs that are available via the feedback link, produces the next channel input.

We shall say that the pair (η, R_L) is achievable for the law $P_{S_1,S_2,S_d}(s_1, s_2, s_d)$ and the MAC with feedback $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2)$ when it is used to convey the sequences determined by the functions g_1, g_2, if for any $\epsilon > 0$ and all sufficiently large n, there exists some $n_{TW} \leq \eta \cdot n$; a pair of encoders as in (4); a pair of decoders

$$\tilde{\phi}_\mu : \mathcal{S}_\mu^n \times \mathcal{S}_d^n \times \tilde{\mathcal{Y}}^{n_{TW}} \to \mathcal{V}_{\bar{\mu}}^n, \qquad \mu = 1, 2 \qquad (5)$$

and a list decoder $\tilde{\phi}_L$ from $\mathcal{S}_d^n \times \tilde{\mathcal{Y}}^{n_{TW}}$ to the set of all subsets of $\mathcal{V}_1^n \times \mathcal{V}_2^n$ of cardinality not exceeding $\exp(nR_L)$; such that
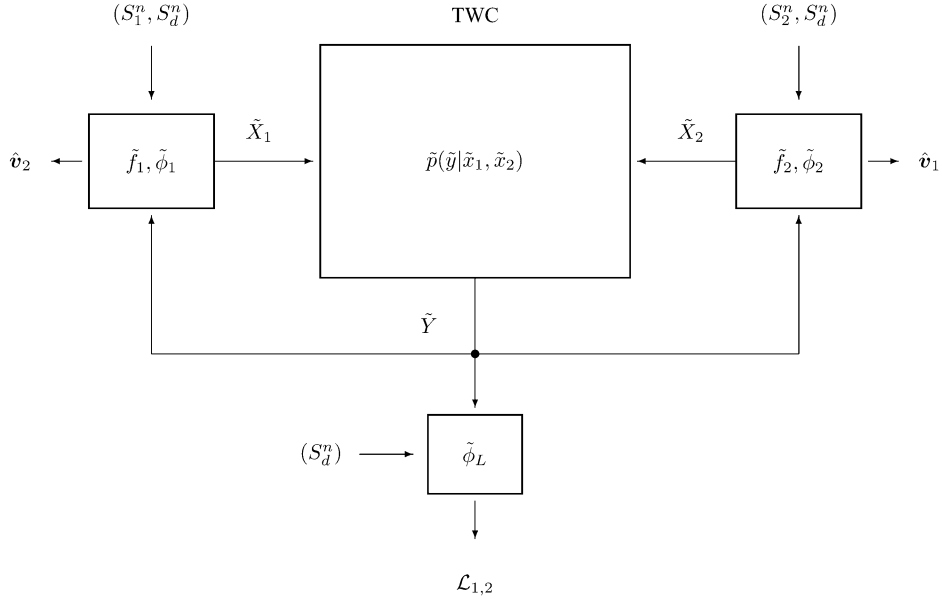
Fig. 1.   Single-output two-way channel for the exchange of $(\boldsymbol{v}_1\,(S_1^n, S_d^n)\,, \boldsymbol{v}_2\,(S_2^n, S_d^n))$.

the probability that either $\tilde{\phi}_1$ produces a sequence other than $\boldsymbol{v}_2(S_2^n, S_d^n)$ or $\tilde{\phi}_2$ produces a sequence other than $\boldsymbol{v}_1(S_1^n, S_d^n)$ or $\tilde{\phi}_L$ produces a list of pairs $\mathcal{L}_{1,2}$ that does not contain $(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))$ is smaller than ϵ. The single-output two-way channel with the encoding/decoding scheme thus defined is illustrated in Fig. 1.

We denote the set of achievable (η, R_L) pairs by

$$\mathcal{T}(P_{S_1, S_2, S_d}(s_1, s_2, s_d), g_1, g_2, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2)).$$

Having defined the set $\mathcal{T}$ we now proceed to inner bound it.

Theorem 1: Consider a MAC with feedback $(\tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \tilde{\mathcal{Y}}, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2))$ and let (S_1, S_2, S_d) take value in the finite set $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_d$ according to the given law P_{S_1, S_2, S_d}. Let the functions g_1, g_2 (3) also be given, and set $V_\mu = g_\mu(S_\mu, S_d)$, $\mu = 1, 2$.

Let $(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$ take value in $\tilde{\mathcal{X}}_1 \times \tilde{\mathcal{X}}_2 \times \tilde{\mathcal{Y}}$ according to a joint law $P_{\tilde{X}_1, \tilde{X}_2, \tilde{Y}}$ of the form

$$P_{\tilde{X}_1, \tilde{X}_2, \tilde{Y}}(\tilde{x}_1, \tilde{x}_2, \tilde{y}) = P_{\tilde{X}_1}(\tilde{x}_1) \cdot P_{\tilde{X}_2}(\tilde{x}_2) \cdot \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2) \quad (6)$$

where $P_{\tilde{X}_1}$ and $P_{\tilde{X}_2}$ are arbitrary laws on $\tilde{\mathcal{X}}_1$ and $\tilde{\mathcal{X}}_2$, respectively, and where $\tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2)$ is the MAC law, i.e., the probability that the MAC emits the symbol $\tilde{y}$ when its inputs are $\tilde{x}_1$ and $\tilde{x}_2$, respectively.

Let the functions $(\tilde{g}_1, \tilde{g}_2)$

$$\tilde{g}_\mu : \tilde{\mathcal{X}}_\mu \times \tilde{\mathcal{Y}} \to \tilde{\mathcal{V}}_\mu, \qquad \mu = 1, 2 \quad (7)$$

(where $\tilde{\mathcal{V}}_\mu$ is some arbitrary finite set) be chosen such that with

$$\tilde{V}_\mu = \tilde{g}_\mu(\tilde{X}_\mu, \tilde{Y}), \qquad \mu = 1, 2 \quad (8)$$

the following holds:

$$\max_{\mu \in \{1,2\}} \left\{ \frac{H(\tilde{V}_\mu | \tilde{X}_{\bar{\mu}}, \tilde{Y})}{I(\tilde{X}_\mu; \tilde{Y} | \tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\} < 1. \quad (9)$$

Define the quantities $\eta^{(TW)}$, ϖ_μ, and $R_L^{(TW)}$ by

$$\eta^{(TW)} \triangleq \max_{\mu \in \{1,2\}} \left\{ \frac{H(V_\mu | S_{\bar{\mu}}, S_d)}{I(\tilde{X}_\mu; \tilde{Y} | \tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\}$$

$$\cdot \left[1 - \max_{\mu \in \{1,2\}} \left\{ \frac{H(\tilde{V}_\mu | \tilde{X}_{\bar{\mu}}, \tilde{Y})}{I(\tilde{X}_\mu; \tilde{Y} | \tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\} \right]^{-1}$$

$$\varpi_\mu \triangleq \frac{I(\tilde{X}_\mu; \tilde{Y} | \tilde{V}_1, \tilde{V}_2)}{I(\tilde{X}_\mu; \tilde{Y} | \tilde{V}_\mu, \tilde{X}_{\bar{\mu}})}$$

$$R_L^{(TW)} \triangleq \min_{\mu \in \{1,2\}} \left\{ H(V_\mu | S_d) - \varpi_\mu H(V_\mu | S_{\bar{\mu}}, S_d) \right.$$

$$+ H(V_{\bar{\mu}} | V_\mu, S_d) - \varpi_{\bar{\mu}} H(V_{\bar{\mu}} | S_\mu, S_d)$$

$$+ \eta^{(TW)} \cdot [H(\tilde{V}_\mu | \tilde{Y}) - \varpi_\mu H(\tilde{V}_\mu | \tilde{X}_{\bar{\mu}}, \tilde{Y})$$

$$\left. + H(\tilde{V}_{\bar{\mu}} | \tilde{V}_\mu, \tilde{Y}) - \varpi_{\bar{\mu}} H(\tilde{V}_{\bar{\mu}} | \tilde{X}_\mu, \tilde{Y})] \right\}.$$

$$(10)$$

Then any pair (η, R_L) satisfying $\eta \geq \eta^{(TW)}$, $R_L \geq R_L^{(TW)}$ is achievable, i.e., it satisfies

$$(\eta, R_L) \in \mathcal{T}(P_{S_1, S_2, S_d}(s_1, s_2, s_d), g_1, g_2, \tilde{p}(\tilde{y}|\tilde{x}_1, \tilde{x}_2)).$$

Moreover, the set $\mathcal{T}$ is convex.

Remark: The preferable mappings $\tilde{g}_\mu(\tilde{X}_\mu, \tilde{Y})$ in (7) are those for which

$$I(\tilde{X}_\mu; \tilde{Y} | \tilde{X}_{\bar{\mu}}, \tilde{V}_\mu) > I(\tilde{X}_\mu; \tilde{Y} | \tilde{X}_{\bar{\mu}}), \qquad \mu = 1, 2.$$

The proof of this theorem appears in Section IV.

We next describe the proposed achievable region for a MAC with feedback.

Theorem 2: Consider a MAC with feedback $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$. Let the auxiliary random variable U take value in some finite set $\mathcal{U}$ according to some arbitrary law P_U. Let the random variables (S_1, S_2, S_d) take

value in $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ and assume that the joint law of (U, S_1, S_2, S_d), which we denote by P_{U,S_1,S_2,S_d}, is of the form

$$P_{U,S_1,S_2,S_d}(u, s_1, s_2, s_d)$$
$$= P_U(u)P_{S_1|U}(s_1|u)P_{S_2|U}(s_2|u)p(s_d|s_1, s_2)$$

where $P_{S_1|U}$ and $P_{S_2|U}$ are arbitrary conditional laws and where $p(s_d|s_1, s_2)$ is the MAC law, i.e., the probability that the MAC produces at its output the symbol s_d when its inputs are s_1 and s_2, respectively. Denote the joint law of (S_1, S_2, S_d) by P_{S_1,S_2,S_d} so that

$$P_{S_1,S_2,S_d}(s_1, s_2, s_d) = \sum_{u \in \mathcal{U}} P_{U,S_1,S_2,S_d}(u, s_1, s_2, s_d).$$

Let the functions (g_1, g_2) as defined by (3) be arbitrary, and let $V_\mu = g_\mu(S_\mu, S_d), \mu = 1, 2$.

Then, if the pair (η, R_L) satisfies

$$(\eta, R_L) \in \mathcal{T}(P_{S_1,S_2,S_d}(s_1, s_2, s_d), g_1, g_2, p(y|x_1, x_2))$$

and if the rates R_1, R_2 satisfy

$$R_1 \leq (1 + \eta)^{-1}I(S_1; S_d|S_2, U, V_1)$$
$$R_2 \leq (1 + \eta)^{-1}I(S_2; S_d|S_1, U, V_2)$$
$$R_1 + R_2 \leq (1 + \eta)^{-1}[I(S_1, S_2; S_d|V_1, V_2) - R_L]$$

then the rate pair (R_1, R_2) is achievable on the MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ with feedback.

Moreover

- Any rate-pair that can be demonstrated to be achievable using this theorem can also be demonstrated to be achievable using this theorem with the auxiliary random variable U taking value in a finite set $\mathcal{U}$ of cardinality

$$\min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1, |\mathcal{Y}| + 2\}.$$

- The capacity region of the MAC with feedback is topologically closed and is convex.

The proof of this theorem appears in Section V.

Combining this result with the inner bound of Theorem 1 on $\mathcal{T}$ we obtain the following.

Corollary 1: Consider a MAC with feedback $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$. Let the auxiliary random variable U take value in some finite set $\mathcal{U}$, according to some arbitrary law P_U, where $|\mathcal{U}| \leq \min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1, |\mathcal{Y}| + 2\}$. Let the random variables (X_1, X_2, Y) take value in $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ and assume that the joint law of (U, X_1, X_2, Y), which we denote by $P_{U,X_1,X_2,Y}$, is of the form

$$P_{U,X_1,X_2,Y}(u, x_1, x_2, y)$$
$$= P_U(u)P_{X_1|U}(x_1|u)P_{X_2|U}(x_2|u)p(y|x_1, x_2) \quad (11)$$

where $P_{X_1|U}$ and $P_{X_2|U}$ are arbitrary conditional laws. Let $(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$ take value in $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ according to a joint law $P_{\tilde{X}_1,\tilde{X}_2,\tilde{Y}}$ of the form

$$P_{\tilde{X}_1,\tilde{X}_2,\tilde{Y}}(\tilde{x}_1, \tilde{x}_2, \tilde{y}) = P_{\tilde{X}_1}(\tilde{x}_1) \cdot P_{\tilde{X}_2}(\tilde{x}_2) \cdot p(\tilde{y}|\tilde{x}_1, \tilde{x}_2)$$

where $P_{\tilde{X}_1}$ and $P_{\tilde{X}_2}$ are arbitrary laws on $\tilde{\mathcal{X}}_1$ and $\tilde{\mathcal{X}}_2$, respectively, and where $p(\tilde{y}|\tilde{x}_1, \tilde{x}_2)$ is the probability that the MAC emits the symbol $\tilde{y}$ when its inputs are $\tilde{x}_1$ and $\tilde{x}_2$, respectively.

For arbitrary deterministic functions (g_1, g_2)

$$g_\mu : \mathcal{X}_\mu \times \mathcal{Y} \to \mathcal{V}_\mu, \qquad \mu = 1, 2$$

let $V_\mu = g_\mu(X_\mu, Y)$, and any deterministic functions $(\tilde{g}_1, \tilde{g}_2)$, which are defined as in (7), and satisfy (9) with $\tilde{V}_\mu$ as in (8).

Finally, with ϖ_μ as defined in (10), define the quantities $\tilde{\eta}^{(TW)}$ and $\tilde{R}_L^{(TW)}$ by

$$\tilde{\eta}^{(TW)} \triangleq \max_{\mu \in \{1,2\}} \left\{ \frac{H(V_\mu|X_{\bar{\mu}}, Y)}{I(\tilde{X}_\mu; \tilde{Y}|\tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\}$$
$$\cdot \left[1 - \max_{\mu \in \{1,2\}} \left\{ \frac{H(\tilde{V}_\mu|\tilde{X}_{\bar{\mu}}, \tilde{Y})}{I(\tilde{X}_\mu; \tilde{Y}|\tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\} \right]^{-1}$$

$$\tilde{R}_L^{(TW)} \triangleq \min_{\mu \in \{1,2\}} \left\{ H(V_\mu|Y) - \varpi_\mu H(V_\mu|X_{\bar{\mu}}, Y) \right.$$
$$+ H(V_{\bar{\mu}}|V_\mu, Y) - \varpi_{\bar{\mu}} H(V_{\bar{\mu}}|X_\mu, Y)$$
$$+ \tilde{\eta}^{(TW)} \cdot [H(\tilde{V}_\mu|\tilde{Y}) - \varpi_\mu H(\tilde{V}_\mu|\tilde{X}_{\bar{\mu}}, \tilde{Y})$$
$$\left. + H(\tilde{V}_{\bar{\mu}}|\tilde{V}_\mu, \tilde{Y}) - \varpi_{\bar{\mu}} H(\tilde{V}_{\bar{\mu}}|\tilde{X}_\mu, \tilde{Y})] \right\}.$$

Then whenever the quadruple (η, R_L, R_1, R_2) satisfies

$$\eta \geq \tilde{\eta}^{(TW)}$$
$$R_L \geq \tilde{R}_L^{(TW)}$$
$$R_1 \leq (1 + \eta)^{-1}I(X_1; Y|X_2, U, V_1)$$
$$R_2 \leq (1 + \eta)^{-1}I(X_2; Y|X_1, U, V_2)$$
$$R_1 + R_2 \leq (1 + \eta)^{-1}[I(X_1, X_2; Y|V_1, V_2) - R_L]$$

the rate pair (R_1, R_2) is achievable on the MAC with feedback. Moreover, the region is topologically closed and convex.

We next demonstrate that the above region always contains the Cover–Leung region. This can be seen as follows. If g_μ and $\tilde{g}_\mu, \mu = 1, 2$ are chosen to take on exactly one value each, then $V_\mu = g_\mu(X_\mu, Y)$ and $\tilde{V}_\mu = \tilde{g}_\mu(\tilde{X}_\mu, \tilde{Y})$ are deterministic. In this case, $\tilde{\eta}^{(TW)} = 0$ and $\tilde{R}_L^{(TW)} = 0$ in which case we obtain (upon setting $\eta = R_L = 0$) the achievability of any rate pair (R_1, R_2) satisfying

$$R_1 \leq I(Y; X_1|X_2, U)$$
$$R_2 \leq I(Y; X_2|X_1, U)$$
$$R_1 + R_2 \leq I(Y; X_1, X_2) \quad (12)$$

which coincides with the Cover–Leung region [2, Theorem 1].

An example with a strict inclusion

To demonstrate a case with a strict inclusion of the Cover–Leung region we shall exhibit, for the channel (1), a symmetric rate pair that belongs to the region defined by Corollary 1 but is outside the region (12). To that end, we let the mappings (g_1, g_2) for the channel (1) be defined as follows:

$$g_\mu(x_\mu, y) = \begin{cases} 1, & x_\mu = 0 \text{ and } y = 1 \\ e, & \text{otherwise} \end{cases} \quad (13)$$

and similarly, the mappings $(\tilde{g}_1, \tilde{g}_2)$ are defined by

$$\tilde{g}_\mu(\tilde{x}_\mu, \tilde{y}) = \begin{cases} 1, & \tilde{x}_\mu = 0 \text{ and } \tilde{y} = 1 \\ e, & \text{otherwise.} \end{cases} \quad (14)$$

Here "e" denotes an "erasure" meaning that Encoder μ cannot infer what Encoder $\bar{\mu}$ has sent during that particular time.

Next, consider the case where $q \ll 1$, and let us look at the quantity $(R_1 + R_2)/q$. Then, the choice $\mathcal{U} = \{0,1\}$; $P_U(1) = 0.0045$

$$P_{X_\mu|U}(x_\mu|u) = \begin{cases} (0|0), & 0.731 \\ (0|1), & 1 - \epsilon, \quad \epsilon \ll 1, \ \mu = 1, 2 \end{cases}$$

and $P_{\tilde{X}_1}(1) = P_{\tilde{X}_2}(1) = 0.2601$; yields a symmetric sum rate of $(R_1 + R_2)/q = 0.553 + o(1)$, where the $o(1)$ term tends to zero as $q \to 0$.

On the other hand, the maximal symmetric sum rate in the Cover–Leung region is $(R_1 + R_2)/q = 0.499426 + o(1)$. This rate pair is attained with $\mathcal{U} = \{0,1,2\}$; $P_U(0) = 0.446808$, $P_U(1) = 0.09$, $P_U(2) = 0.463192$, and

$$\Pr[X_1 = 1|U = 0] = \Pr[X_2 = 1|U = 0] = 0.214$$
$$\Pr[X_1 = 1|U = 1] = \Pr[X_2 = 1|U = 1] = 1 - \epsilon, \quad \epsilon \ll 1$$
$$\Pr[X_1 = 1|U = 2] = \Pr[X_2 = 1|U = 2] = 0.242.$$

IV. THE TWO-WAY PHASE: ACHIEVABILITY PROOF OF THEOREM 1

The two-way phase is composed of L recursive encoding steps and it is decoded via the backward decoding technique [6], [12]. When backward decoding is used, the decoder as well as each of the encoders starts decoding only after all L blocks (corresponding to the L encoding steps) have been received. In block l, $l = 2, \ldots, L$ each encoder transmits the resolution information his partner needs in order to decode the previous block. Assuming the resolution information in block L is sent at a low enough rate, it is then used to transform the high-rate information in block $L - 1$ into low-rate information. Consequently, this permits the simultaneous decoding of the aggregate information sent in block $L - 1$. Next, this resolution information transforms the high-rate information in block $L - 2$ into decodable low-rate information, wherein simultaneous decoding can be applied again. Continuing this way, each encoder as well as the the decoder (with the help of the extra resolution information sent to him by both encoders during the following transmission phase as described later on) successively decode the information in all blocks in a block-simultaneous way, with block 1 decoded last.

Before getting into the details of the two-way phase, we define explicitly what is our aim during this phase.

Let both encoders and the decoder observe an n-length sequence S_d^n that is generated i.i.d. according to some joint law P_{S_1, S_2, S_d}, wherein Encoder μ has as well access to the side information S_μ^n. As a result, Encoder μ forms the corresponding vector $V_\mu^n = (V_{\mu,1}, V_{\mu,2}, \ldots, V_{\mu,n})$ and now both encoders wish to exchange the pair (V_1^n, V_2^n) between them.

For Encoder μ who knows V_μ^n, the ambiguity regarding $V_{\bar{\mu}}^n$ is given by the *conditional* (on its private side information S_μ^n as well as on the public side information S_d^n) entropy

$$H\left(V_{\bar{\mu}}^n|S_\mu^n, S_d^n\right) = nH(V_{\bar{\mu}}|S_\mu, S_d), \qquad \mu = 1, 2 \quad (15)$$

where (15) follows by the fact that conditional on S_d^n the outcome (V_1^n, V_2^n) is a realization of n i.i.d. drawings of a pair of random variables (V_1, V_2). At the same time for the decoder, who is ignorant of (S_1^n, S_2^n), the ambiguity regarding V_μ^n is given by the *conditional* (just on the public side information S_d^n) entropy

$$H\left(V_\mu^n|S_d^n\right) = nH(V_\mu|S_d), \qquad \mu = 1, 2. \quad (16)$$

Thus, the problem of exchanging the correlated pair of sequences (V_1^n, V_2^n) between the encoders via the two-way channel can be envisaged as follows.

Consider a bipartite graph with $N_{2,1}^{(1)} \triangleq \exp\{nH(V_2|S_d)\}$ right nodes and $N_{1,1}^{(1)} \triangleq \exp\{nH(V_1|S_d)\}$ left nodes (later on we set $S_d = Y$). A right node represents an outcome of the random vector V_2^n, while a left node represents an outcome of the random vector V_1^n. Each left node has

$$N_{1,2}^{(1)} \triangleq \exp\{nH(V_2|S_1, S_d)\}$$

outgoing edges connected to the corresponding right nodes representing the possible outcomes of V_2^n as viewed by Encoder 1 (who has access to its private side information S_1). At the same time each right node has

$$N_{2,2}^{(1)} \triangleq \exp\{nH(V_1|S_2, S_d)\}$$

outgoing edges connected to the corresponding left nodes representing the possible outcomes of V_1^n as viewed by Encoder 2 (later on we set $S_\mu = X_\mu$). Encoder 1 who knows the correct left node seeks to resolve the correct right node, while Encoder 2 who knows the correct right node seeks to determine the correct left node.

A two-way communication phase consists of a sequence of L two-way steps, wherein during step l, $l = 1, 2, \ldots, L$ the encoders use an n_l-length two-user code $\mathcal{C}_{TW}^{(l)}$ that is generated according to the joint distribution $P_{\tilde{X}_1, \tilde{X}_2}(\tilde{x}_1, \tilde{x}_2)$ induced by the law (6) as follows.

- Generate $N_{1,1}^{(l)}$ sequences $\tilde{\boldsymbol{x}}_1 = (\tilde{x}_{1,1}, \tilde{x}_{1,2}, \ldots, \tilde{x}_{1,n_l})$, each with probability

$$\Pr\{\tilde{\boldsymbol{x}}_1\} = \prod_{k=1}^{n_l} P_{\tilde{X}_1}(\tilde{x}_{1,k}).$$

  Label them $\tilde{\boldsymbol{x}}_1(w_1^{(l)})$, $w_1^{(l)} \in \{1, 2, \ldots, N_{1,1}^{(l)}\}$.
- Generate $N_{2,1}^{(l)}$ sequences $\tilde{\boldsymbol{x}}_2 = (\tilde{x}_{2,1}, \tilde{x}_{2,2}, \ldots, \tilde{x}_{2,n_l})$, each with probability

$$\Pr\{\tilde{\boldsymbol{x}}_2\} = \prod_{k=1}^{n_l} P_{\tilde{X}_2}(\tilde{x}_{2,k}).$$

  Label them $\tilde{\boldsymbol{x}}_2(w_2^{(l)})$, $w_2^{(l)} \in \{1, 2, \ldots, N_{2,1}^{(l)}\}$.

Consider next the decoding of a typical two-way communication phase.

During the first step of the two-way phase the encoders send the blocklength n_1 codeword pair $(\tilde{\boldsymbol{x}}_1(w_1^{(1)}), \tilde{\boldsymbol{x}}_2(w_2^{(1)}))$ and as a result both encoders and the decoder observe the channel output $\tilde{Y}^{n_1}$. Thus, Encoder μ can generate

$$\tilde{V}_\mu^{n_1} = \tilde{g}_\mu\left(\tilde{\boldsymbol{x}}_\mu\left(w_\mu^{(1)}\right), \tilde{Y}^{n_1}\right)$$

and the decoding of this step is accomplished assuming that both encoders have (somehow) exchanged first the pair $(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1})$ thereby obtaining the rates $I(\tilde{X}_\mu; \tilde{Y}|\tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)$, $\mu = 1, 2$.

Consequently, to decode two-way step l, $l = 1, 2, \ldots, L-1$, presumably via joint typicality considerations, it is assumed that both encoders exchange first the pair of vectors $(\tilde{V}_1^{n_l}, \tilde{V}_2^{n_l})$. Therefore, two-way step 1, which is intended for the exchange of (V_1^n, V_2^n) (the information regarding (S_1^n, S_2^n, S_d^n)), entails the execution of two-way step 2 during which the encoders exchange $(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1})$ (the information regarding $(\tilde{X}_1^{n_1}, \tilde{X}_2^{n_1}, \tilde{Y}^{n_1})$), and so on. As a result of this recursive encoding procedure, both encoders end up executing a series of two-way steps with corresponding codewords' block lengths $n_1 > n_2 > \cdots > n_{L-1}$, where this chain of inequalities is implied by the technical condition (9) as shown in the sequel.

The encoders decode the information as follows. Suppose that in two-way step L both encoders use standard multiple-access transmission and as a result decode the "resolution information" $(\tilde{V}_1^{n_{L-1}}, \tilde{V}_2^{n_{L-1}})$ which is needed to interpret the realization of two-way step $L-1$. Based on $\tilde{\boldsymbol{x}}_1(w_1^{(L-1)})$, $\tilde{Y}^{n_{L-1}}$, and $\tilde{V}_2^{n_{L-1}}$ Encoder 1 decodes $w_2^{(L-1)}$ taking into account that $w_2^{(L-1)}$ resides within a block length n_{L-1} codebook of size $\exp\{H(\tilde{V}_2^{n_{L-2}}|\tilde{Y}^{n_{L-2}})\}$, while actually its ambiguity (for Encoder 1) is just within a subcode (of this codebook) the size of which equals

$$\exp\left\{H\left(\tilde{V}_2^{n_{L-2}}|\tilde{\boldsymbol{x}}_1\left(w_1^{(L-2)}\right), \tilde{Y}^{n_{L-2}}\right)\right\}.$$

Similarly, based on $\tilde{\boldsymbol{x}}_2(w_2^{(L-1)})$, $\tilde{Y}^{n_{L-1}}$, and $\tilde{V}_1^{n_{L-1}}$ Encoder 2 decodes $w_1^{(L-1)}$ taking into account that $w_1^{(L-1)}$ resides within a block length n_{L-1} codebook of size $\exp\{H(\tilde{V}_1^{n_{L-2}}|\tilde{Y}^{n_{L-2}})\}$, while actually its ambiguity (for Encoder 2) is just within a subcode of size

$$\exp\left\{H\left(\tilde{V}_1^{n_{L-2}}|\tilde{\boldsymbol{x}}_2\left(w_2^{(L-2)}\right), \tilde{Y}^{n_{L-2}}\right)\right\}.$$

As a result, both encoders acquire $(\tilde{V}_1^{n_{L-2}}, \tilde{V}_2^{n_{L-2}})$. This in turn enables the decoding of two-way step $L-2$ and so on, until each encoder recovers (V_1^n, V_2^n) which concludes the exchange of the sequences $(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))$.

We can now choose L such that the list sizes corresponding to two-way step $L-1$ are exponentially small when measuring them in reference to the initial block length n, yet at the same time n_{L-1} is sufficiently large. Thus, the information transmitted during two-way step L lies within a set that is sufficiently small compared to the direct product of the sets corresponding to two-way steps $1, 2, \ldots, L-1$. Consequently, the duration consumed by two-way step L is negligible.

During two-way step l, $l = 1, 2, \ldots, L-1$ Encoder 1 estimates the message of Encoder 2 and declares $\hat{w}_2^{(l)} = w_2^{(l)}$ if and only if there is a unique $w_2^{(l)}$ such that

$$\left(\tilde{\boldsymbol{x}}_1\left(w_1^{(l)}\right), \tilde{\boldsymbol{x}}_2\left(w_2^{(l)}\right), \tilde{\boldsymbol{y}}^{(l)}\right) \in \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{X}_2, \tilde{Y}). \quad (17)$$

Using Lemmas 1 and 3, it can be shown that Encoder 1's decision will be correct with arbitrarily small probability of error $P_{e2,1}^{(l)} < \epsilon/(8BL)$ if (here B is some predetermined positive integer)

$$nH(V_2|S_1, S_d) \leq n_1[I(\tilde{X}_2; \tilde{Y}|\tilde{X}_1, \tilde{V}_2) - \epsilon]$$
$$n_{l-1}H(\tilde{V}_2|\tilde{X}_1, \tilde{Y}) \leq n_l[I(\tilde{X}_2; \tilde{Y}|\tilde{X}_1, \tilde{V}_2) - \epsilon],$$
$$l = 2, 3, \ldots, L-1 \quad (18)$$

and n, n_l are sufficiently large.

Similarly, Encoder 2 declares $\hat{w}_1^{(l)} = w_1^{(l)}$ if and only if there is a unique $w_1^{(l)}$ such that (17) is satisfied. Using Lemmas 1 and 3, it can be shown that Encoder 2's decision will be correct with arbitrarily small probability of error $P_{e1,2}^{(l)} < \epsilon/(8BL)$ if

$$nH(V_1|S_2, S_d) \leq n_1[I(\tilde{X}_1; \tilde{Y}|\tilde{X}_2, \tilde{V}_1) - \epsilon]$$
$$n_{l-1}H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) \leq n_l[I(\tilde{X}_1; \tilde{Y}|\tilde{X}_2, \tilde{V}_1) - \epsilon],$$
$$l = 2, 3, \ldots, L-1 \quad (19)$$

and n, n_l are sufficiently large.

Thus, the combination of (18), (19) together with the technical condition (9) ensures that $n_1 > n_2 > \cdots > n_{L-1}$ is admissible.

In conclusion—upon the termination of the two-way phase both encoders know the exact realization of $(\tilde{V}_1^{n_l}, \tilde{V}_2^{n_l})$, $l = 1, 2, \ldots, L-1$, as well as (V_1^n, V_2^n).

Turning to the decoder, we recall that the decoder seeks to determine both sequences $\boldsymbol{v}_1(S_1^n, S_d^n)$ and $\boldsymbol{v}_2(S_2^n, S_d^n)$, based on its knowledge of the sequence S_d^n and upon the outputs that were produced during the exchange $\tilde{y}_1, \ldots, \tilde{y}_{n_{TW}}$. To this end, we consider two possible decoders for the situation at hand.

Decoder 1—joint decoder.

The decoder decodes first two-way step L which provides him with $(\tilde{V}_1^{n_{L-1}}, \tilde{V}_2^{n_{L-1}})$, then it sets $l = L-1$.

- Based on $\tilde{Y}^{n_l-1}$ Decoder 1 forms an *a priori* list of possible transmitted block length n_l codeword pairs, $\{(\tilde{\boldsymbol{x}}_1(w_1^{(l)}), \tilde{\boldsymbol{x}}_2(w_2^{(l)}))\}$, the size of which equals $\exp\{H(\tilde{V}_1^{n_{l-1}}, \tilde{V}_2^{n_{l-1}}|\tilde{Y}^{n_{l-1}})\}$.
- Based on $(\tilde{V}_1^{n_l}, \tilde{V}_2^{n_l})$ and its acquaintance with the sequence $\tilde{Y}^{n_l}$, Decoder 1 forms the *a posteriori* list

$$\mathcal{L}_{1,2} = \left\{\left(\tilde{V}_1^{n_{l-1}}, \tilde{V}_2^{n_{l-1}}\right)\right\} = \left\{\left(w_1^{(l)}, w_2^{(l)}\right)\right\}$$

of all message pairs such that (17) is satisfied.

Using the resolution information, that both encoders communicate later on, Decoder 1 resolves the correct message pair $(w_1^{(l)}, w_2^{(l)})$ within the list $\mathcal{L}_{1,2}$.

- Having acquired $(\tilde{V}_1^{n_{l-1}}, \tilde{V}_2^{n_{l-1}})$, Decoder 1 sets $l = l-1$ and, provided that $l \geq 2$, proceeds backward to decode two-way step l.

Following the above procedure Decoder 1 obtains

$$\left(w_1^{(2)}, w_2^{(2)}\right) = \left(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1}\right).$$

- Based on S_d^n, Decoder 1 forms an *a priori* list of possible transmitted block length n_1 codeword pairs, $\{(\tilde{\boldsymbol{x}}_1(w_1^{(1)}), \tilde{\boldsymbol{x}}_2(w_2^{(1)}))\}$, the size of which equals $\exp\{H(V_1^n, V_2^n | S_d^n)\}$.
- Based on $(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1})$ and its acquaintance with the sequence $\tilde{Y}^{n_1}$, Decoder 1 forms the *a posteriori* list

$$\mathcal{L}_{1,2} = \{(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))\} = \left\{\left(w_1^{(1)}, w_2^{(1)}\right)\right\}$$

of all message pairs such that

$$\left(\tilde{\boldsymbol{x}}_1\left(w_1^{(1)}\right), \tilde{\boldsymbol{x}}_2\left(w_2^{(1)}\right), \tilde{\boldsymbol{y}}^{(1)}\right) \in \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{X}_2, \tilde{Y}). \quad (20)$$

Using the resolution information that both encoders communicate later on, Decoder 1 resolves the correct message pair (V_1^n, V_2^n) within $\mathcal{L}_{1,2}$.

Decoder 2—successive decoder.

The decoder decodes first two-way step L which provides him with $(\tilde{V}_1^{n_L-1}, \tilde{V}_2^{n_L-1})$, then it sets $l = L - 1$.

- Based on $\tilde{Y}^{n_l-1}$ Decoder 2 forms an *a priori* list of possible transmitted block length n_l codewords, $\{\tilde{\boldsymbol{x}}_1(w_1^{(l)})\}$, the size of which equals $\exp\{H(\tilde{V}_1^{n_l-1} | \tilde{Y}^{n_l-1})\}$.
- Based on $(\tilde{V}_1^{n_l}, \tilde{V}_2^{n_l})$ and $\tilde{Y}^{n_l}$ Decoder 2 forms the *a posteriori* list $\mathcal{L}_1 = \{w_1^{(l)}\}$ of all messages such that

$$\left(\tilde{\boldsymbol{x}}_1\left(w_1^{(l)}\right), \tilde{\boldsymbol{y}}^{(l)}\right) \in \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{Y}). \quad (21)$$

Using the resolution information that both encoders communicate later on, Decoder 2 resolves the correct message $w_1^{(l)} = \tilde{V}_1^{n_l-1}$ within $\mathcal{L}_1$.
- Based on $(\tilde{Y}^{n_l-1}, \tilde{V}_1^{n_l-1})$ Decoder 2 forms an *a priori* list of possible transmitted block length n_l codewords, $\{\tilde{\boldsymbol{x}}_2(w_2^{(l)})\}$, the size of which equals $\exp\{H(\tilde{V}_2^{n_l-1} | \tilde{V}_1^{n_l-1}, \tilde{Y}^{n_l-1})\}$.
- Based on $(\tilde{V}_1^{n_l}, \tilde{V}_2^{n_l})$ and $\tilde{Y}^{n_l}$ Decoder 2 forms the *a posteriori* list $\mathcal{L}_2 = \{w_2^{(l)}\}$ of all messages such that (17) is satisfied. Using the resolution information that both encoders communicate later on, Decoder 2 resolves the correct message $w_2^{(l)}$ within $\mathcal{L}_2$.
- Having acquired $(\tilde{V}_1^{n_l-1}, \tilde{V}_2^{n_l-1})$, Decoder 2 sets $l = l-1$ and, provided that $l \geq 2$, proceeds backward to decode two-way step l.

Following the above procedure Decoder 2 obtains

$$\left(w_1^{(2)}, w_2^{(2)}\right) = \left(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1}\right).$$

- Based on S_d^n Decoder 2 forms an a-priori list of possible transmitted blocklength n_1 codewords, $\left\{\tilde{\boldsymbol{x}}_1\left(w_1^{(1)}\right)\right\}$, the size of which equals $\exp\left\{H\left(V_1^n | S_d^n\right)\right\}$.
- Based on $(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1})$ and its acquaintance with the sequence $\tilde{Y}^{n_1}$, Decoder 2 forms the *a posteriori* list

$$\mathcal{L}_1 = \{\boldsymbol{v}_1(S_1^n, S_d^n)\} = \left\{w_1^{(1)}\right\}$$

of all messages such that

$$\left(\tilde{\boldsymbol{x}}_1\left(w_1^{(1)}\right), \tilde{\boldsymbol{y}}^{(1)}\right) \in \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{Y}). \quad (22)$$

Using the resolution information, that both encoders communicate later on, Decoder 2 resolves the correct sequence V_1^n within $\mathcal{L}_1$.
- Based on (S_d^n, V_1^n) Decoder 2 forms an *a priori* list of possible transmitted block length n_1 codewords, $\{\tilde{\boldsymbol{x}}_2(w_2^{(1)})\}$, the size of which equals $\exp\{H(V_2^n | V_1^n, S_d^n)\}$.
- Based on $(\tilde{V}_1^{n_1}, \tilde{V}_2^{n_1})$ and $\tilde{Y}^{n_1}$, Decoder 2 forms the *a posteriori* list

$$\mathcal{L}_2 = \{\boldsymbol{v}_2(S_2^n, S_d^n)\} = \left\{w_2^{(1)}\right\}$$

of all messages such that (20) is satisfied. Using the resolution information that both encoders communicate later on, Decoder 2 resolves the correct sequence V_2^n within $\mathcal{L}_2$.

In conclusion—following the above procedures either Decoder 1 or Decoder 2 acquire $(w_1^{(1)}, w_2^{(1)})$ which determine the pair (V_1^n, V_2^n) that corresponds to (S_1^n, S_2^n, S_d^n).

In what follows, we upper-bound the list sizes of $\mathcal{L}_1$ and $\mathcal{L}_2$. Since

$$|\mathcal{L}_{1,2}| \leq |\mathcal{L}_1| \cdot |\mathcal{L}_2|$$

this provides an upper bound R_L on the exponential growth of the list size for either decoder type.

Before that, we argue that when analyzing the successive list decoder performance one may assume that during the list decoding of $\mathcal{L}_2$ (the successive decoder's second step), the decoder exhibits the same *a priori* statistics as during the list decoding of $\mathcal{L}_1$. This is true because the first step terminates with the list $\mathcal{L}_1$, then Decoder 2 acquires the correct message from somewhere else via the resolution information sent to him by both encoders later on. As a result, there is no propagation from the first decoding step to the second, as is the case, e.g., with the successive decoder for the Gaussian MAC [16, Sec. II]. In this sense, the successive list decoder is a genie-aided decoder which "forgets" the list $\mathcal{L}_1$, assuming that nature has provided him with the result of the first decoding step, and proceeds to the second decoding step as if it were the first task being executed.

Performance evaluation for the successive decoder.

Consider first the set $T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_1)$ of codewords which at the end of two-way step l, $l = 1, 2, \ldots, L - 1$ are jointly typical with $\tilde{\boldsymbol{y}}^{(l)} \triangleq \tilde{Y}^{n_l}$. From Lemma 1, we know that $(\tilde{\boldsymbol{x}}_1(w_1^{(l)}), \tilde{\boldsymbol{x}}_2(w_2^{(l)}))$ will be jointly typical with $\tilde{\boldsymbol{y}}^{(l)}$ with high probability, say $P_c^{(l)} > 1 - \epsilon/(8BL)$.

Let

$$\Psi_k(\tilde{\boldsymbol{y}}^{(l)}) = \begin{cases} 1, & \left(\tilde{\boldsymbol{x}}_1(k), \tilde{\boldsymbol{y}}^{(l)}\right) \in \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{Y}) \\ 0, & \text{otherwise.} \end{cases}$$

Then the cardinality of $T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_1)$ is the random variable

$$\left|T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_1)\right| = \sum_k \Psi_k(\tilde{\boldsymbol{y}}^{(l)})$$

and

$$\mathbb{E}\left|T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_1)\right| = \mathbb{E}\Psi_{w_1^{(l)}}(\tilde{\boldsymbol{y}}^{(l)}) + \sum_{k \neq w_1^{(l)}} \mathbb{E}\Psi_k(\tilde{\boldsymbol{y}}^{(l)})$$

where $\mathbb{E}$ denotes the expectation operator.

We now bound the random variable $|T_{\tilde{\boldsymbol{Y}}}^{(l)}(\mathcal{L}_1)|$.

Lemma 4: Define

$$\varpi_1 \triangleq \frac{I(\tilde{X}_1; \tilde{Y}|\tilde{V}_1, \tilde{V}_2)}{I(\tilde{X}_1; \tilde{Y}|\tilde{V}_1, \tilde{X}_2)}. \tag{23}$$

Then, for any $\epsilon > 0$, there exists an n_0 such that for $n \geq n_0$ and $n_l \geq n_0$, $l = 1, 2, \ldots, L-1$ the list sizes $\{|\mathcal{L}_1^{(l)}|\}_{l=1}^{L-1}$ accumulated by the successive list decoder are bounded by

$$\left|\mathcal{L}_1^{(1)}\right| < \exp\{n[H(V_1|Y) - \varpi_1 H(V_1|X_2, Y) + \epsilon]\}$$
$$= \exp\{n[H(V_1|S_d) - \varpi_1 H(V_1|S_2, S_d) + \epsilon]\}$$
$$\left|\mathcal{L}_1^{(l)}\right| < \exp\{n_{l-1}[H(\tilde{V}_1|\tilde{Y}) - \varpi_1 H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) + \epsilon]\},$$
$$l = 2, \ldots, L-1 \tag{24}$$

with probability larger than $1 - \epsilon/(16BL)$.

In particular, if $\varpi_1 = 1$—i.e., whenever there *exists* a mapping $\psi : \mathcal{Y} \times \mathcal{V}_1 \times \mathcal{V}_2 \to \mathcal{X}_2$ such that $\psi(y, v_1, v_2) = x_2$ for any pair $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ with $p(y|x_1, x_2) > 0$, then (24) reduces to

$$\left|\mathcal{L}_1^{(1)}\right| < \exp\{n[I(V_1; X_2|Y) + \epsilon]\}$$
$$= \exp\{n[I(V_1; S_2|S_d) + \epsilon]\}$$
$$\left|\mathcal{L}_1^{(l)}\right| < \exp\{n_{l-1}[I(\tilde{V}_1; \tilde{X}_2|\tilde{Y}) + \epsilon]\}, \qquad l = 2, \ldots, L-1 \tag{25}$$

with probability larger than $1 - \epsilon/(16BL)$.

Apart from providing upper bounds on the list sizes, Lemma 4 characterizes a pair of "good" mappings $\tilde{V}_\mu = \tilde{g}_\mu(\tilde{X}_\mu, \tilde{Y})$, $\mu = 1, 2$ in the sense that the corresponding list sizes are minimized. Specifically, this happens when the triple (y, v_1, v_2) uniquely determines x_2 and in this case the list sizes are determined by the corresponding mutual information functionals as per (25).

Proof: Using Lemma 3, it follows that

$$\mathbb{E}\Psi_k(\tilde{\boldsymbol{Y}}^{(l)}) \leq \exp\{-n_l[I(\tilde{X}_1; \tilde{Y}|\tilde{V}_1, \tilde{V}_2) - \epsilon]\}, \quad k \neq w_1^{(l)}.$$

Therefore,

$$\mathbb{E}\left|T_{\tilde{\boldsymbol{Y}}}^{(l)}(\mathcal{L}_1)\right| \leq 1 + \left(N_{1,1}^{(l)} - 1\right)\exp\{-n_l[I(\tilde{X}_1; \tilde{Y}|\tilde{V}_1, \tilde{V}_2) - \epsilon]\}.$$

Recall that the relevant codebook sizes during the first $L-1$ coding steps executed during the two-way phase are determined as follows:

$$N_{1,1}^{(1)} = \exp\{nH(V_1|S_d)\}$$
$$N_{1,1}^{(l)} = \exp\{n_{l-1}H(\tilde{V}_1|\tilde{Y})\}, \qquad l = 2, 3, \ldots, L-1$$
$$N_{2,1}^{(1)} = \exp\{nH(V_2|S_d)\}$$
$$N_{2,1}^{(l)} = \exp\{n_{l-1}H(\tilde{V}_2|\tilde{Y})\}, \qquad l = 2, 3, \ldots, L-1. \tag{26}$$

Combining (19), (23), and (26) we get (27) (at the bottom of the page) for n, n_l sufficiently large, and any $\epsilon > 0$.

Using Markov's inequality, we obtain for any $\epsilon_1 > 0$

$$P_{LD}^{(1)}(\mathcal{L}_1) \triangleq \Pr\left\{\left|T_{\tilde{\boldsymbol{Y}}}^{(l)}(\mathcal{L}_1)\right| > \exp\{n[H(V_1|S_d) - \varpi_1 H(V_1|S_2, S_d) + \epsilon_1]\}\right\} \leq e^{-n\epsilon_1}$$
$$\leq \epsilon/(16BL) \text{ for sufficiently large } n$$
$$P_{LD}^{(l)}(\mathcal{L}_1) \triangleq \Pr\left\{\left|T_{\tilde{\boldsymbol{Y}}}^{(l)}(\mathcal{L}_1)\right| > \exp\{n_{l-1}[H(\tilde{V}_1|\tilde{Y}) - \varpi_1 H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) + \epsilon_1]\}\right\} \leq e^{-n_{l-1}\epsilon_1}$$
$$\leq \epsilon/(16BL) \text{ for sufficiently large } n_{l-1},$$
$$l = 2, \ldots, L-1. \tag{28}$$

This completes the proof of Lemma 4.

Consider next the set $T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_2)$ of codewords which at the end of two-way step l, $l = 1, 2, \ldots, L-1$ are jointly typical with $(\tilde{\boldsymbol{x}}_1(w_1^{(l)}), \tilde{\boldsymbol{y}}^{(l)})$. To this end note that, having acquired $w_1^{(l)}$, the decoder's ambiguity regarding $w_2^{(l)}$ is

$$H\left(w_2^{(l)}|w_1^{(l)}, Y^n\right) = nH(V_2|V_1, S_d), \qquad l = 1$$
$$H\left(w_2^{(l)}|w_1^{(l)}, \tilde{Y}^{n_{l-1}}\right) = n_{l-1}H(\tilde{V}_2|\tilde{V}_1, \tilde{Y}),$$
$$l = 2, 3, \ldots, L-1. \tag{29}$$

Lemma 5: Define

$$\varpi_2 \triangleq \frac{I(\tilde{X}_2; \tilde{Y}|\tilde{V}_1, \tilde{V}_2)}{I(\tilde{X}_2; \tilde{Y}|\tilde{V}_2, \tilde{X}_1)}. \tag{30}$$

Then, for any $\epsilon > 0$, there exists an n_0 such that for $n \geq n_0$ and $n_l \geq n_0$, $l = 1, 2, \ldots, L-1$ the list sizes $\{|\mathcal{L}_2^{(l)}|\}_{l=1}^{L-1}$ accumulated by the successive list decoder are bounded by

$$\left|\mathcal{L}_2^{(1)}\right| < \exp\{n[H(V_2|V_1, Y) - \varpi_2 H(V_2|X_1, Y) + \epsilon]\}$$
$$= \exp\{n[H(V_2|V_1, S_d) - \varpi_2 H(V_2|S_1, S_d) + \epsilon]\}$$
$$\left|\mathcal{L}_2^{(l)}\right| < \exp\{n_{l-1}[H(\tilde{V}_2|\tilde{V}_1, \tilde{Y}) - \varpi_2 H(\tilde{V}_2|\tilde{X}_1, \tilde{Y}) + \epsilon]\},$$
$$l = 2, \ldots, L-1 \tag{31}$$

with probability larger than $1 - \epsilon/(16BL)$.

In particular, if $\varpi_2 = 1$—i.e., whenever there *exists* a mapping $\psi : \mathcal{Y} \times \mathcal{V}_1 \times \mathcal{V}_2 \to \mathcal{X}_1$ such that $\psi(y, v_1, v_2) = x_1$ for any pair $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ with $p(y|x_1, x_2) > 0$, then (31) reduces to

$$\left|\mathcal{L}_2^{(1)}\right| < \exp\{n[H(V_2|V_1, Y) - H(V_2|X_1, Y) + \epsilon]\}$$
$$= \exp\{n[H(V_2|V_1, S_d) - H(V_2|S_1, S_d) + \epsilon]\}$$
$$\left|\mathcal{L}_2^{(l)}\right| < \exp\{n_{l-1}[H(\tilde{V}_2|\tilde{V}_1, \tilde{Y}) - H(\tilde{V}_2|\tilde{X}_1, \tilde{Y}) + \epsilon]\},$$
$$l = 2, \ldots, L-1 \tag{32}$$

with probability larger than $1 - \epsilon/(16BL)$.

$$\mathbb{E}\left|T_{\tilde{\boldsymbol{Y}}}^{(l)}(\mathcal{L}_1)\right| \leq 1 + \begin{cases} \exp\{n[H(V_1|S_d) - \varpi_1 H(V_1|S_2, S_d) + \epsilon]\}, & l = 1 \\ \exp\{n_{l-1}[H(\tilde{V}_1|\tilde{Y}) - \varpi_1 H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) + \epsilon]\}, & l = 2, \ldots, L-1 \end{cases} \tag{27}$$

Proof: The proof follows the same lines as the proof of Lemma 4 taking into account (18), (29), and (30).

Thus, with high probability

$$1 - \sum_{l=1}^{L-1} \left(P_{LD}^{(l)}(\mathcal{L}_1) + P_{LD}^{(l)}(\mathcal{L}_2) \right),$$

the index required (in a typical two-way phase) in order to resolve the decoder's residual uncertainty, about the two-way phase, takes on no more than

$$\prod_{l=1}^{L-1} \left| \mathcal{L}_1^{(l)} \right| \cdot \left| \mathcal{L}_2^{(l)} \right|$$

$$= \exp\Bigg\{ n[H(V_1|S_d) - \varpi_1 H(V_1|S_2, S_d) \\ + H(V_2|V_1, S_d) - \varpi_2 H(V_2|S_1, S_d) + \epsilon] \\ + \sum_{l=1}^{L-1} n_l[H(\tilde{V}_1|\tilde{Y}) - \varpi_1 H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) \\ + H(\tilde{V}_2|\tilde{V}_1, \tilde{Y}) - \varpi_2 H(\tilde{V}_2|\tilde{X}_1, \tilde{Y})] \Bigg\} \quad (33)$$

values.

Noting that (18) and (19) imply that

$$n_1 = n \max_{\mu_1 \in \{1,2\}} \left\{ \frac{H(V_{\mu_1}|S_{\bar{\mu}_1}, S_d)}{I(\tilde{X}_{\mu_1}; \tilde{Y}|\tilde{X}_{\bar{\mu}_1}, \tilde{V}_{\mu_1})} \right\}$$

$$n_l = n_{l-1} \max_{\mu_2 \in \{1,2\}} \left\{ \frac{H(\tilde{V}_{\mu_2}|\tilde{X}_{\bar{\mu}_2}, \tilde{Y})}{I(\tilde{X}_{\mu_2}; \tilde{Y}|\tilde{X}_{\bar{\mu}_2}, \tilde{V}_{\mu_2})} \right\}, \\ l = 2, 3, \ldots, L-1 \quad (34)$$

we can upper-bound the *rate* $R_L^{(TW)}$—i.e., the exponential growth of the list size—implied by the right-hand side (RHS) of (33) as

$$R_L^{(TW)} \triangleq \frac{1}{n} \sum_{l=1}^{L-1} \ln\left(\left| \mathcal{L}_1^{(l)} \right| \cdot \left| \mathcal{L}_2^{(l)} \right| \right) < H(V_1|S_d) \\ - \varpi_1 H(V_1|S_2, S_d) + H(V_2|V_1, S_d) \\ - \varpi_2 H(V_2|S_1, S_d) \\ + \max_{\mu_1 \in \{1,2\}} \left\{ \frac{H(V_{\mu_1}|S_{\bar{\mu}_1}, S_d)}{I(\tilde{X}_{\mu_1}; \tilde{Y}|\tilde{X}_{\bar{\mu}_1}, \tilde{V}_{\mu_1})} \right\} \\ \cdot \left[1 - \max_{\mu_2 \in \{1,2\}} \left\{ \frac{H(\tilde{V}_{\mu_2}|\tilde{X}_{\bar{\mu}_2}, \tilde{Y})}{I(\tilde{X}_{\mu_2}; \tilde{Y}|\tilde{X}_{\bar{\mu}_2}, \tilde{V}_{\mu_2})} \right\} \right]^{-1} \\ \cdot [H(\tilde{V}_1|\tilde{Y}) - \varpi_1 H(\tilde{V}_1|\tilde{X}_2, \tilde{Y}) + H(\tilde{V}_2|\tilde{V}_1, \tilde{Y}) \\ - \varpi_2 H(\tilde{V}_2|\tilde{X}_1, \tilde{Y})]. \quad (35)$$

Remark: The rate $R_L^{(TW)}$ has been obtained assuming the decoder performs successive list decoding first on $w_1^{(l)}$, $l = 1, \ldots, L-1$, and then on $w_2^{(l)}$ given $w_1^{(l)}$, during the two-way phase. Obviously, an expression similar to (35) can be obtained assuming the decoder performs successive list decoding first on $w_2^{(l)}$ and then on $w_1^{(l)}$ given $w_2^{(l)}$, and the minimal rate should be selected as $R_L^{(TW)}$.

Bounding the probability of error. For the above scheme, we declare an error in the two-way communication phase if at least one of the following events happens.

- E_1: Encoder 1 estimates $w_2^{(l)}$ incorrectly during two-way step l, $l = 1, 2, \ldots, L-1$.
- E_2: Encoder 2 estimates $w_1^{(l)}$ incorrectly during two-way step l.
- E_3:

$$\left(\tilde{\boldsymbol{x}}_\mu\left(w_\mu^{(l)}\right), \tilde{\boldsymbol{y}}^{(l)} \right) \notin \mathcal{A}_\epsilon(\tilde{X}_\mu, \tilde{Y})$$

or

$$\left(\tilde{\boldsymbol{x}}_1\left(w_1^{(l)}\right), \tilde{\boldsymbol{x}}_2\left(w_2^{(l)}\right), \tilde{\boldsymbol{y}}^{(l)} \right) \notin \mathcal{A}_\epsilon(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$$

  $l = 1, 2, \ldots, L-1$. Here $w_\mu^{(l)}$ is the message decoded first by the list decoder.
- E_4: Either $|T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_1)|$ or $|T_{\tilde{\boldsymbol{y}}}^{(l)}(\mathcal{L}_2)|$ exceed their typical size as implied by Lemmas 4 and 5.

Using the union of events bound, we can upper-bound the probability of error for a typical two-way phase $\overline{P}_{e,b,TW}$ (averaged over the choice of codebooks and possible input messages) by

$$\overline{P}_{e,b,TW} \leq (L-1) \sum_{i=1}^{4} \Pr\{E_i\} < \epsilon/(2B). \quad (36)$$

This proves the existence of at least one code with an average probability of error in a typical two-way phase of less than $\epsilon/(2B)$. The union bound then implies that the average probability of error while executing a sequence of B two-way phases is less than $\epsilon/2$.

Finally, we compute the aggregate block length of our code, namely

$$n_{TW} \triangleq \sum_{l=1}^{L-1} n_l \leq n_1 \left[1 - \max_{\mu \in \{1,2\}} \left\{ \frac{H(\tilde{V}_\mu|\tilde{X}_{\bar{\mu}}, \tilde{Y})}{I(\tilde{X}_\mu; \tilde{Y}|\tilde{X}_{\bar{\mu}}, \tilde{V}_\mu)} \right\} \right]^{-1}. \quad (37)$$

Combining (34) with (37) our construction thus demonstrates that the ratio $n_{TW}/n \leq \eta^{(TW)}$, with $\eta^{(TW)}$ defined in (10), is achievable with list decoder's exponential list size not exceeding (35).

Moreover, the achievability of $(\eta_1, R_L^{(1)})$ and $(\eta_2, R_L^{(2)})$ implies the achievability of

$$\left(\alpha\eta_1 + (1-\alpha)\eta_2, \alpha R_L^{(1)} + (1-\alpha)R_L^{(2)} \right)$$

for any $0 \leq \alpha \leq 1$, therefore, the set $\mathcal{T}$ is convex.

This completes the proof of Theorem 1. $\qquad \square$

Remark: Let $\left(\tilde{R}_1, \tilde{R}_2 \right)$ denote the information rate pair which is attained between both encoders during the two-way phase. Defining $Z \triangleq \left(\tilde{V}_1, \tilde{V}_2 \right)$ it follows from (18) and (19) that

$$\tilde{R}_1 \leq I(\tilde{X}_1; \tilde{Y}|\tilde{X}_2, \tilde{V}_1) \overset{(a)}{=} I(\tilde{X}_1; \tilde{Y}|\tilde{X}_2, \tilde{V}_1, \tilde{V}_2) \\ = I(\tilde{X}_1; \tilde{Y}|\tilde{X}_2, Z) \\ \tilde{R}_2 \leq I(\tilde{X}_2; \tilde{Y}|\tilde{X}_1, \tilde{V}_2) \overset{(b)}{=} I(\tilde{X}_2; \tilde{Y}|\tilde{X}_1, \tilde{V}_1, \tilde{V}_2) \\ = I(\tilde{X}_2; \tilde{Y}|\tilde{X}_1, Z). \quad (38)$$

Here (a) and (b) follow since the pair $(\tilde{X}_\mu, \tilde{Y})$ determine $\tilde{V}_\mu$ by (7).

Furthermore, by way of our construction

$$I(\tilde{X}_1; \tilde{X}_2|Z) \le I(\tilde{X}_1; \tilde{X}_2|\tilde{Y}, Z)$$

and the rate pair in (38) is obtained for a joint distribution

$$P_{Z,\tilde{X}_1,\tilde{X}_2,\tilde{Y}}(z, \tilde{x}_1, \tilde{x}_2, \tilde{y})$$
$$= P_{Z,\tilde{X}_1,\tilde{X}_2}(z, \tilde{x}_1, \tilde{x}_2) P_{\tilde{Y}|\tilde{X}_1,\tilde{X}_2}(\tilde{y}|\tilde{x}_1, \tilde{x}_2).$$

Comparing this expression with the *dependence balance* outer bound for the capacity region of the single-output two-way channel (TWC) provided by [5, Theorem 1] we may infer that our two-way scheme is "optimal" in the sense that the information exchange between the encoders conforms with a law which characterizes an outer bound for the capacity region. It is not claimed, however, that the specific choice of Z herein is the capacity achieving one.

V. Achievability Proof for the Two-User DMMAC With Feedback

In this section, we describe a coding scheme for the MAC with feedback and show that it achieves the rate region defined in Theorem 2.

Main idea behind our coding strategy.

The key idea in the Cover–Leung coding scheme is the use of superposition block Markov encoding (SBME). In this method, transmission occurs in B blocks wherein in block b, $b = 1, 2, \ldots, B-1$ both encoders send a certain amount of "fresh information" reliably to each other at the maximum possible rate. Since the transmission rate at this stage is too high for reliable transmission to the receiver, all the receiver can do is reduce the set of possible transmitted messages to a considerably smaller set of "typical" messages. Since this set contains the correct message with probability very close to one, both encoders can cooperate during the next block to send the final resolution information to the decoder. Superimposed on this resolution information, the encoders again send fresh information at high rate.

In our scheme, we use as a starting point the Cover–Leung construction; however, before the encoders decode block b they exchange some extra information (V_1, V_2) that each one can tell about himself to his partner in order to improve the information rate his partner can obtain. This exchange of information is accomplished via an additional two-way communication phase which is appended to each "fresh information" block. Only after both encoders have acquired the extra information from the two-way phase they decode the "fresh information" that has been transmitted during the current block. At the same time, the receiver performs list decoding over the two-way phase and resolves its remaining ambiguity via the resolution information which both encoders send during the next block. Having acquired the extra information from the two-way phase, the receiver decodes the "fresh information" as well as the resolution information contained in the current block, given this extra information, in the very same way as in the Cover–Leung scheme.

The reason for exchanging the extra information (V_1, V_2) between the encoders can be explained as follows. Let $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ be a DMMAC and suppose that when given the output y and the input x_μ, the input $x_{\bar{\mu}}$ is not uniquely determined—i.e., there *does not* exist for any $\mu \in \{1, 2\}$ a function $\psi_\mu : \mathcal{Y} \times \mathcal{X}_\mu \to \mathcal{X}_{\bar{\mu}}$ such that $\psi_\mu(y, x_\mu) = x_{\bar{\mu}}$ whenever there is an $x_\mu \in \mathcal{X}_\mu$ with $p(y|x_\mu, x_{\bar{\mu}}) > 0$. Instead, consider any pair of functions $V_\mu = g_\mu(X_\mu, Y)$, $\mu = 1, 2$ that map for each of the encoders an input–output pair to an element in a corresponding set $\mathcal{V}_\mu$. As a result of observing the feedback link, Encoder μ knows V_μ yet it needs to acquire $V_{\bar{\mu}}$. Assume further that somehow Encoder μ has acquired $V_{\bar{\mu}}$ and the decoder has acquired the pair (V_1, V_2). Presumably, there *exists* a function $\psi : \mathcal{Y} \times \mathcal{V}_1 \times \mathcal{V}_2 \to \mathcal{X}_1 \times \mathcal{X}_2$ such that $\psi(y, v_1, v_2) = (x_1, x_2)$ for a subset of triples $(y, x_1, x_2) \in \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2$ with $p(y|x_1, x_2) > 0$, and consequently

$$I(X_1; Y|X_2, U, V_1) > I(X_1; Y|X_2, U)$$
$$I(X_2; Y|X_1, U, V_2) > I(X_2; Y|X_1, U). \qquad (39)$$

Inequality (39) is the key for our motivation to exchange the pair (V_1, V_2) between the encoders since having done so they can cooperate in order to "shift" this information to the decoder (to use the terminology of [13]) and this offers an improvement on the Cover–Leung rate-sum.

To be consistent with the notation in Theorem 2, recall that the random variables (S_1, S_2, S_d) take value in $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, so that without loss of generality (w.l.o.g.), we may interpret X_μ as S_μ and Y as S_d.

Provided that

$$H(V_\mu|S_d) > H(V_\mu|S_{\bar{\mu}}, S_d), \qquad \mu = 1, 2, \qquad (40)$$

it is conceivable that it pays to invest the extra time consumed by the two-way phase, in order to exchange the sequences $(\boldsymbol{v}_1(S_1^n, S_d^n), \boldsymbol{v}_2(S_2^n, S_d^n))$, as its impact on the improved rate-sum (39) would not be that significant, and the net information rate to the receiver would still be better than the Cover–Leung result.

The structure of our coding scheme.

In our coding scheme, transmission consists of a sequence of B generic transmission phases, wherein a generic transmission phase consists of a multiple-access phase succeeded by a two-way phase. During the MAC phase of generic transmission phase b, $b = 1, 2, \ldots, B-1$, each encoder transmits a new information block combined with a resolution information the decoder needs to decode generic phase $b - 1$. Following the MAC phase, both encoders execute a two-way phase in order to exchange the (V_1^n, V_2^n) realization of the MAC phase. Upon the termination of the two-way phase, having acquired common knowledge, both encoders can cooperate to send the resolution information for the decoder. Consequently, the resolution information for the decoder is sent during the MAC phase of the next generic transmission phase in the form of a common time-sharing random variable. During generic phase B just the resolution information for phase $B-1$ is sent. Fig. 2 shows schematically the structure of the proposed coding scheme—specifically blocks b and $b+1$ are illustrated.
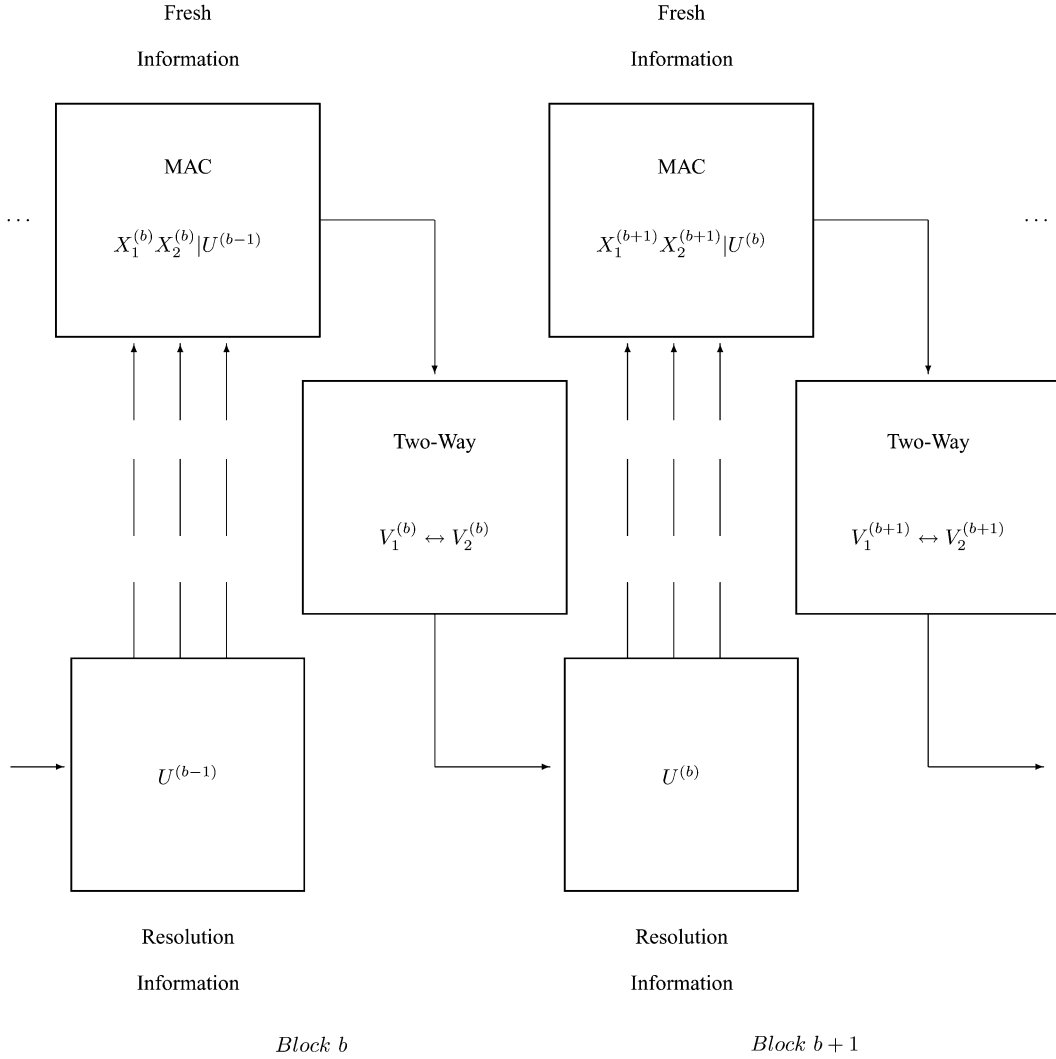
Fig. 2.   The proposed coding scheme.

The MAC Phase: Achievability proof of Theorem 2

During the multiple-access phase of generic transmission phase b, $b = 1, \ldots, B - 1$ the encoders use a two-user code $\mathcal{C}^{(b)}$ that is generated according to the joint distribution $P_{X_1, X_2}(x_1, x_2)$ induced by the law (11) as follows.

- For $b = 1$ generate a single sequence

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$$

with probability

$$\Pr\{\boldsymbol{u}\} = \prod_{k=1}^{n} P_U(u_k).$$

Label this sequence $\boldsymbol{u}^{(1)}(1)$.

- For $b \geq 2$, let $\mathcal{U}^{(b)}$ be a collection of $|\Lambda|$ binary n-length sequences drawn independently according to the product distribution of marginal $P_U(u)$. Index $\mathcal{U}^{(b)}$ by Λ, thus, a generic element of $\mathcal{U}^{(b)}$ can be expressed as $\boldsymbol{u}^{(b)}(\lambda)$, $\lambda \in \Lambda$.

Using the terminology of Bergmans [10, Sec. III], the sequences $\boldsymbol{u}^{(b)}(\lambda)$ are now used as "cloud centers" to generate

the actual transmitted codewords. The codebooks are generated as follows.

1) Based on every $\boldsymbol{u}^{(b)}(\lambda)$ generate M_1 sequences $\boldsymbol{x}_1^{(b)} = (x_{1,1}^{(b)}, x_{1,2}^{(b)}, \ldots, x_{1,n}^{(b)})$, each with probability

$$\Pr\left\{\boldsymbol{x}_1^{(b)} | \boldsymbol{u}^{(b)}(\lambda)\right\} = \prod_{k=1}^{n} P_{X_1|U}\left(x_{1,k}^{(b)} | u_k^{(b)}\right).$$

Label them $\boldsymbol{x}_1^{(b)}(\boldsymbol{u}^{(b)}(\lambda), w_1) = \boldsymbol{x}_1^{(b)}(\lambda, w_1)$ for $\lambda \in \Lambda$ and $w_1 \in \mathcal{M}_1 = \{1, \ldots, M_1\}$.

2) Similarly, based on every $\boldsymbol{u}^{(b)}(\lambda)$ generate M_2 sequences $\boldsymbol{x}_2^{(b)} = (x_{2,1}^{(b)}, x_{2,2}^{(b)}, \ldots, x_{2,n}^{(b)})$, each with probability

$$\Pr\left\{\boldsymbol{x}_2^{(b)} | \boldsymbol{u}^{(b)}(\lambda)\right\} = \prod_{k=1}^{n} P_{X_2|U}\left(x_{2,k}^{(b)} | u_k^{(b)}\right).$$

Label them $\boldsymbol{x}_2^{(b)}(\boldsymbol{u}^{(b)}(\lambda), w_2) = \boldsymbol{x}_2^{(b)}(\lambda, w_2)$ for $\lambda \in \Lambda$ and $w_2 \in \mathcal{M}_2 = \{1, \ldots, M_2\}$.

Henceforth, as we refer to any specific codeword $\boldsymbol{x}_\mu^{(b)}(\boldsymbol{u}^{(b)}(\lambda), w_\mu)$, $\mu = 1, 2$, we refer to $\boldsymbol{u}^{(b)}(\lambda)$ as the cloud center of this codeword.

During the MAC phase of generic transmission phase b, $b = 1, 2, \ldots$ the message $(w_1^{(b)}, w_2^{(b)}) \in \mathcal{M}_1 \times \mathcal{M}_2$ is *encoded* in the following way.

Let

$$(V_1^n, V_2^n)^{(b)} \triangleq \left(V_{1,1}^{(b)}, V_{1,2}^{(b)}, \ldots, V_{1,n}^{(b)}, V_{2,1}^{(b)}, V_{2,2}^{(b)}, \ldots, V_{2,n}^{(b)}\right)$$

denote the realization for the MAC block of generic transmission phase b. Furthermore, let $\lambda^{(b)}$ denote the common random variable furnished by both encoders upon the termination of generic transmission phase b.

In generic transmission phase 1, the encoders send during the MAC phase the codewords

$$\boldsymbol{x}_1^{(1)} = \boldsymbol{x}_1^{(1)}\left(\boldsymbol{u}^{(1)}(1), w_1^{(1)}\right)$$
$$\boldsymbol{x}_2^{(1)} = \boldsymbol{x}_2^{(1)}\left(\boldsymbol{u}^{(1)}(1), w_2^{(1)}\right).$$

Upon the termination of generic phase b, $b = 1, \ldots, B-1$

- Encoder 1, who knows $(V_{1,1}^{(b)}, V_{1,2}^{(b)}, \ldots, V_{1,n}^{(b)})$ decodes $(V_{2,1}^{(b)}, V_{2,2}^{(b)}, \ldots, V_{2,n}^{(b)})$ via the two-way phase that followed the MAC phase. Knowing $(V_{2,1}^{(b)}, V_{2,2}^{(b)}, \ldots, V_{2,n}^{(b)})$, $\boldsymbol{y}^{(b)} \triangleq (Y^n)^{(b)}$ and $\lambda^{(b-1)}$ the cloud center of $\boldsymbol{x}_2^{(b)}$ Encoder 1 can now decode the information $w_2^{(b)}$. Specifically, Encoder 1 estimates the message of Encoder 2 by considering the set of codewords $\{\boldsymbol{x}_2^{(b)}(\lambda^{(b-1)}, w_2^{(b)})\}$ and declaring $\hat{w}_2^{(b)} = w_2^{(b)}$ if and only if there is a unique $w_2^{(b)}$ such that

$$\left(\boldsymbol{x}_1^{(b)}\left(\lambda^{(b-1)}, w_1^{(b)}\right), \boldsymbol{x}_2^{(b)}\left(\lambda^{(b-1)}, w_2^{(b)}\right), \boldsymbol{y}^{(b)}\right)$$
$$\in \mathcal{A}_\epsilon(U, X_1, X_2, Y). \quad (41)$$

Using Lemmas 1 and 3, it can be shown that Encoder 1's decision will be correct with arbitrarily small probability of error $P_{e2,1}^{(b)} < \epsilon/(10B)$ if

$$R_2 \leq I(X_2; Y | X_1, U, V_2) - \epsilon \quad (42)$$

and n is sufficiently large.

- Similarly, Encoder 2 who knows $(V_{2,1}^{(b)}, V_{2,2}^{(b)}, \ldots, V_{2,n}^{(b)})$ decodes $(V_{1,1}^{(b)}, V_{1,2}^{(b)}, \ldots, V_{1,n}^{(b)})$ via the two-way phase. Knowing $(V_{1,1}^{(b)}, V_{1,2}^{(b)}, \ldots, V_{1,n}^{(b)})$, $\boldsymbol{y}^{(b)}$ and $\lambda^{(b-1)}$ the cloud center of $\boldsymbol{x}_1^{(b)}$, Encoder 2 can now decode the information $w_1^{(b)}$. Specifically, Encoder 2 estimates the message of Encoder 1 by considering the set of codewords $\{\boldsymbol{x}_1^{(b)}(\lambda^{(b-1)}, w_1^{(b)})\}$ and declaring $\hat{w}_1^{(b)} = w_1^{(b)}$ if and only if there is a unique $w_1^{(b)}$ such that (41) is satisfied. Using Lemmas 1 and 3, it can be shown that Encoder 2's decision will be correct with arbitrarily small probability of error $P_{e1,2}^{(b)} < \epsilon/(10B)$ if

$$R_1 \leq I(X_1; Y | X_2, U, V_1) - \epsilon \quad (43)$$

and n is sufficiently large.

- Both encoders emulate the decoding procedure performed by the decoder during the two-way phase and obtain the final set of lists $(T_{\hat{\boldsymbol{y}}}^{(1)}, \ldots, T_{\hat{\boldsymbol{y}}}^{(L-1)})$ that the decoder is left

with when trying to decode $(V_1^n, V_2^n)^{(b)}$. This forms the *first part* of the resolution information that will be sent by both encoders via the random variable $\lambda^{(b)} \in \Lambda$ which serves as the cloud center of the multiple-access block during the next generic transmission phase. The *second part* of the resolution information $\lambda^{(b)}$ will be defined subsequently.

In generic transmission phase b, $b = 2, 3, \ldots, B-1$, the encoders transmit during the MAC phase the codewords

$$\boldsymbol{x}_1^{(b)} = \boldsymbol{x}_1^{(b)}\left(\boldsymbol{u}^{(b)}(\lambda^{(b-1)}), w_1^{(b)}\right)$$
$$\boldsymbol{x}_2^{(b)} = \boldsymbol{x}_2^{(b)}\left(\boldsymbol{u}^{(b)}(\lambda^{(b-1)}), w_2^{(b)}\right).$$

We see that $w_1^{(b)}$ and $w_2^{(b)}$ contain new information, while the rest consists of resolution information.

In generic phase B, the encoders do not send new information and just cooperate to inform the decoder on $\lambda^{(B-1)}$.

The decoder uses backward decoding to find the transmitted information. First it decodes generic phase B thereby acquiring $\lambda^{(B-1)}$. Subsequently, it sets $b = B - 1$ and performs the following procedure.

Decoding Procedure of $\lambda^{(b-1)}$ and $(w_1^{(b)}, w_2^{(b)})$

The decoder uses $\lambda^{(b)}$ to obtain the complete resolution information needed in order to decode generic phase b.

First, the decoder decodes the two-way phase of generic transmission phase b using the first part of the resolution information. This provides the decoder with $(V_1^n, V_2^n)^{(b)}$. Turning to the MAC block of generic transmission phase b the decoder declares $\hat{\lambda}^{(b-1)} = \lambda^{(b-1)}$ if and only if there is a unique $\lambda^{(b-1)}$ such that

$$(\boldsymbol{u}^{(b)}(\lambda^{(b-1)}), \boldsymbol{y}^{(b)}) \in \mathcal{A}_\epsilon(U, Y). \quad (44)$$

Using Lemmas 1 and 3, it can be shown that the decoder's decision will be correct with arbitrarily small probability of error $P_{ed,1}^{(b)} < \epsilon/(10B)$ if

$$|\Lambda| < \exp\{n[I(U; Y | V_1, V_2) - \epsilon_2]\} \quad (45)$$

and n is sufficiently large.

Having determined $\lambda^{(b-1)}$ the decoder forms a list $\mathcal{L}_{1,2} = \{(w_1^{(b)}, w_2^{(b)})\}$ of all message pairs such that (41) is satisfied. Using the second part of the resolution information that both encoders send via $\lambda^{(b)}$ the decoder resolves the correct pair within $\mathcal{L}_{1,2}$.

We consider now the set $T_{\boldsymbol{y}}^{(b)}(\mathcal{L}_{1,2})$ of codewords which are jointly typical with $\boldsymbol{y}^{(b)}$. Following the same procedure as Cover–Leung did in [2] one concludes that, provided that (42) and (43) are satisfied

$$\mathbb{E}\left|T_{\boldsymbol{y}}^{(b)}(\mathcal{L}_{1,2})\right| \leq \exp\left\{n\left[R_L^{(\text{MAC})} - \epsilon\right]\right\}$$
$$R_L^{(\text{MAC})} = R_1 + R_2 - I(X_1, X_2; Y | U, V_1, V_2) \quad (46)$$

for n sufficiently large and $R_L^{(\text{MAC})} > \epsilon$.

By Markov's inequality, it follows that with high probability $1 - P_{ed,2}^{(b)} \geq 1 - \epsilon/(10B)$, the index required in order to resolve

the decoder's residual uncertainty about MAC block b takes no more than $\exp\{n(R_L^{(\text{MAC})} + \epsilon_1)\}$, $\epsilon_1 > 0$ values. This index forms the *second part* of the resolution information that is sent by both encoders via the random variable $\lambda^{(b)} \in \Lambda$ which serves as the cloud center of the multiple-access block during generic transmission phase $b + 1$.

End Decoding Procedure—Output $\lambda^{(b-1)}, (w_1^{(b)}, w_2^{(b)})$

Next, the decoder sets $b = b - 1$ and performs the decoding procedure again.

We see that the decoder starts decoding only after all B blocks have been received. Assuming that the resolution information in block B is sent at low enough rate, it is then used to resolve the decoder's ambiguity while decoding block $B - 1$. Continuing this way, the decoder successively decodes all blocks, with block 1 decoded last. This is in sharp contrast to the Cover–Leung scheme wherein the decoder is capable of decoding block b once it has decoded the cloud center $U^{(b+1)}$, and this can be accomplished upon the termination of block $b + 1$. Thus, the decoding delay of the scheme proposed herein is strictly larger than that of the Cover–Leung scheme.

Let us now take stock at where we are. The aggregate resolution information should satisfy (45), while on the other hand, we know that it consists of two components; the first having list-size exponent not larger than $R_L^{(TW)}$ as per (35) and the second having list-size exponent not larger than $R_L^{(\text{MAC})}$ as per (46). Thus, we need

$$
\begin{aligned}
R_L^{(TW)} &+ R_L^{(\text{MAC})} + \epsilon_1 \\
&= R_1 + R_2 + R_L^{(TW)} - I(X_1, X_2; Y|U, V_1, V_2) + \epsilon_1 \\
&\overset{(45)}{<} I(U; Y|V_1, V_2) - \epsilon_2
\end{aligned}
\tag{47}
$$

or equivalently

$$
\begin{aligned}
R_1 + R_2 &+ R_L^{(TW)} \\
&< I(X_1, X_2; Y|U, V_1, V_2) + I(U; Y|V_1, V_2) - (\epsilon_1 + \epsilon_2) \\
&= I(U, X_1, X_2; Y|V_1, V_2) - (\epsilon_1 + \epsilon_2) \\
&= I(X_1, X_2; Y|V_1, V_2) - (\epsilon_1 + \epsilon_2)
\end{aligned}
\tag{48}
$$

where the last step follows since $U \oplus (X_1, X_2) \oplus Y$ is a Markov chain.

Bounding the probability of error. For the above scheme, we declare an error in block b if at least one of the following events happens.
- E_1: Encoder 1 estimates $w_2^{(b)}$ incorrectly during the MAC phase.
- E_2: Encoder 2 estimates $w_1^{(b)}$ incorrectly during the MAC phase.
- E_3: The decoder estimates the cloud center $\boldsymbol{u}^{(b)}(\lambda^{(b-1)})$ incorrectly during the MAC phase.
- E_4:

$$
\left(\boldsymbol{x}_1^{(b)}\left(\lambda^{(b-1)}, w_1^{(b)}\right), \boldsymbol{x}_2^{(b)}\left(\lambda^{(b-1)}, w_2^{(b)}\right), \boldsymbol{y}^{(b)} \right) \\ \notin \mathcal{A}_\epsilon(U, X_1, X_2, Y).
$$

- E_5:

$$
\left| T_{\boldsymbol{y}}^{(b)}(\mathcal{L}_{1,2}) \right| > \exp\left\{ n\left(R_L^{(\text{MAC})} + \epsilon \right) \right\}, \qquad \epsilon > 0.
$$

- E_6: An error has occurred during the two-way phase.

Using the union of events bound, we can upper-bound the probability of error $\overline{P}_{e,b}$ (averaged over the choice of codebooks and possible input messages) in generic block b by

$$
\overline{P}_{e,b} \leq \sum_{i=1}^{5} \Pr\{E_i\} + \overline{P}_{e,b,TW} < \epsilon/B.
\tag{49}
$$

This proves the existence of at least one code with an average probability of error in generic transmission phase b of less than ϵ/B. The union bound then implies that the average probability of error in B blocks is less than ϵ.

Finally, we compute the aggregate block length of our code, namely

$$
\begin{aligned}
n_t &\triangleq n + \sum_{l=1}^{L-1} n_l = n\left(1 + \frac{n_{TW}}{n}\right) \leq n(1 + \eta) \\
&= n\Bigg\{ 1 + \max_{\mu_1 \in \{1,2\}} \left\{ \frac{H(V_{\mu_1}|S_{\bar{\mu}_1}, S_d)}{I(\tilde{X}_{\mu_1}; \tilde{Y}|\tilde{X}_{\bar{\mu}_1}, \tilde{V}_{\mu_1})} \right\} \\
&\quad \cdot \left[1 - \max_{\mu_2 \in \{1,2\}} \left\{ \frac{H(\tilde{V}_{\mu_2}|\tilde{X}_{\bar{\mu}_2}, \tilde{Y})}{I(\tilde{X}_{\mu_2}; \tilde{Y}|\tilde{X}_{\bar{\mu}_2}, \tilde{V}_{\mu_2})} \right\} \right]^{-1} \Bigg\}.
\end{aligned}
$$

Our construction thus demonstrates that the rate pair $(\frac{n}{n_t} R_1, \frac{n}{n_t} R_2)$ is achievable.

Since this part of the coding scheme is similar to the Cover–Leung scheme, then using the same arguments regarding cardinality bounds as in [4] (that further refers to [14], [15]), it follows that for the auxiliary random variable U it suffices to take $|\mathcal{U}| \leq \min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1, |\mathcal{Y}| + 2\}$ without affecting the region in Theorem 2.

This completes the proof of Theorem 2. $\qquad\square$

This construction demonstrates the following.
- The capacity region for the MAC with feedback is very sensitive to the joint probability structure of the component channels.
- Two-way channels which provide equal outputs on both terminals exhibit an improved capacity region. Moreover, the solution to this problem is closely related to the solution of the feedback problem for the MAC. This observation has been made previously by Dueck in [11].

We demonstrate the results of this section by considering next the continuous-time ideal two-user Poisson MAC with noiseless feedback.

VI. THE IDEAL TWO-USER POISSON MAC

The model for the two-user continuous-time Poisson MAC studied here is described as follows. Two independent users generate inputs $x_\mu(t)$, $\mu = 1, 2$, $0 \leq t \leq \infty$, that determine the

rates of two corresponding doubly stochastic Poisson processes $\{D_\mu(t)\}$. The observation is

$$\nu(t) = \sum_{\mu=1}^{2} D_\mu(t) + D(t) \qquad (50)$$

which is also a Poisson process with instantaneous rate $\lambda_0 + \sum_{\mu=1}^{2} x_\mu(t)$. The dark current represented by $D(t)$ is a homogeneous Poisson process of rate λ_0. It is further assumed that a causal feedback link informs both encoders at every time t of the channel output $\nu(t)$ at all times prior to t. Our main focus here will be the case where $\lambda_0 = 0$ which henceforth will be referred as the ideal Poisson MAC.

An (M_1, M_2, T) code for the continuous-time Poisson MAC with feedback is defined as follows.

1) A collection of encoding functions

$$x_{\mu,t} = x_{\mu,t}\left(w_\mu, \nu_0^{t^-}\right), \qquad \mu = 1, 2, \ 0 \le t \le T$$

where $x_{\mu,t}\left(w_\mu, \nu_0^{t^-}\right)$ is the waveform transmitted at time t by Encoder μ when attempting to convey Message w_μ to the receiver after observing the channel output $\nu(s)$, $0 \le s < t$. Here M_μ denotes the number of different messages that Encoder μ can transmit.

2) A decoding function

$$\phi : \nu_0^T \to \{1, \ldots, M_1\} \times \{1, \ldots, M_2\}.$$

We shall use the average probability of error criterion assuming that the messages (W_1, W_2) are drawn according to a uniform distribution over $\{1, \ldots, M_1\} \times \{1, \ldots, M_2\}$. Consequently, the error probability for the code is defined as

$$\begin{aligned} P_e&(M_1, M_2, T) \\ &= \Pr\left\{(W_1, W_2) \ne \phi\left(\nu_0^T\right)\right\} \\ &= \frac{1}{M_1 M_2} \sum_{w_1=1}^{M_1} \sum_{w_2=1}^{M_2} \\ &\quad \times \Pr\left\{\phi\left(\nu_0^T\right) \ne (w_1, w_2) | W_1 = w_1, W_2 = w_2\right\}. \end{aligned}$$

A rate pair (R_1, R_2) nats per second is achievable for the MAC (50) with feedback if there exists a sequence of (M_1, M_2, T) codes with $R_1 \le (\ln M_1)/T$, $R_2 \le (\ln M_2)/T$ such that $P_e(M_1, M_2, T) \to 0$ as $T \to \infty$. The feedback *capacity region* is the closure of the set of all achievable rates.

The information capacity of the one-way Poisson MAC was determined by Lapidoth and Shamai in [20], while the reliability function (i.e., the exponential behavior of the probability of error for the best code, as the coding delay increases while the transmission rate is held fixed) of the one-way Poisson MAC was computed in [21]. To the best of our knowledge, no result regarding the feedback capacity of this channel has been reported. This is in contrast to the single-user Poisson channel for which the effect of feedback on the capacity has been studied in detail by Frey [18] who showed that as long as the dark current is deterministic, capacity is not increased by feedback, whereas feedback may improve the capacity in the case of a random dark current. Additionally, the impact that feedback has on the reliability of the single-user Poisson channel has been determined by Lapidoth in [19].

Our interest here is in fixed-transmission-time coding schemes (as opposed to sequential coding schemes which adhere to a random transmission time) which meet the peak and average-power constraints

$$x_\mu(t) \le A, \quad \mathbb{E}\left[\frac{1}{T}\int_0^T x_\mu(t)dt\right] \le \sigma A,$$
$$\mu = 1, 2, \ 0 \le t \le T \quad (51)$$

where $[0, T]$ is the transmission interval and the expectation is with respect to (w.r.t.) the codebook message distribution. Except for these input constraints, it is assumed throughout that the channel is unlimited in bandwidth.

The main contribution of this section is the introduction of a feedback code and explicit decoding rules that combine the ingredients of SBME together with backward decoding to improve on the Cover–Leung result. As a consequence, the Poisson MAC with noiseless feedback forms a nontrivial example for a MAC for which SBME *per se* is suboptimal.

A. The Cover–Leung Inner Bound on the Capacity With Noiseless Feedback

We begin by recalling some known results which will be useful later on. Following Wyner [17], the assumption that the channel is unlimited in bandwidth is made explicit by the following.

a) We fix a small $\Delta > 0$, partition the time interval $[0, T]$ into equal segments of length Δ/A, and consider input waveforms that take just the extreme values on each of those segments. Specifically, the channel input waveforms $x_\mu(t)$, $\mu = 1, 2$, are constant for $(k-1)\Delta/A < t \le k\Delta/A$, $k = 1, \ldots, n = AT/\Delta$, and $x_\mu(t)$ take only the values 0 or A. As a result, the receiver may collect the sufficient statistics n-vector $\hat{\boldsymbol{y}}$, formed by the increments $\hat{y}_k = \nu(k\Delta/A) - \nu((k-1)\Delta/A)$.

b) The receiver interprets $\hat{y}_k \ge 2$ as being the same as $\hat{y}_k = 0$. Thus, it bases its decision on the n-vector $\boldsymbol{y}$, where

$$y_k \triangleq \begin{cases} 1, & \hat{y}_k = 1 \\ 0, & \text{otherwise.} \end{cases}$$

In the limit as $\Delta \to 0$ these assumptions do not imply any loss in generality as shown in [17].

Subject to the assumptions a), b), the channel reduces to a two-user binary-input binary-output (input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, output alphabet $\mathcal{Y}$) DMMAC with transition probability $p(y|x_1, x_2)$ given by ($\lambda_0 = sA$)

$$\begin{aligned} p(1|0, 0) &= s\Delta \exp^{-s\Delta} = s\Delta + O(\Delta^2) \\ p(1|A, 0) &= p(1|0, A) = (1+s)\Delta \exp^{-(1+s)\Delta} \\ &= (1+s)\Delta + O(\Delta^2), \\ p(1|A, A) &= (2+s)\Delta \exp^{-(2+s)\Delta} \\ &= (2+s)\Delta + O(\Delta^2). \end{aligned} \quad (52)$$

Thus, in the absence of dark current, the continuous-time bandwidth-unlimited Poisson MAC can be modeled as a DMMAC which falls within the category of MACs described by (1).

TABLE I
ATTAINABLE SYMMETRIC RATES VIA THE COVER–LEUNG CONSTRUCTION AS A FUNCTION OF $|\mathcal{U}|$

$|\mathcal{U}|$	$p(0)$	$p(1)$	$p(2)$	$p(3)$	q_0	q_1	q_2	q_3	$R_1 = R_2$	$UB/2$
2	0.099	0.901	0.0	0.0	0.9	0.2	0.0	0.0	0.246	0.247203
3	0.446808	0.09	0.463192	0.0	0.214	1.0	0.242	0.0	0.249713	0.249728
4	0.811	0.0901	0.0889	0.01	0.22	0.27	1.0	0.276	0.249592	0.249840

1) One-Way Capacity Region: When $\lambda_0 = 0$ one may verify that

$$I(Y; X_1, X_2) = \Delta \left[-2q \ln(2q) + 2q^2 \ln 2 \right]$$

$$I(Y; X_2|X_1) = \Delta \left[-q \ln(q(1+q)) - q^2 \ln\left(1 + \frac{1}{q}\right) \right.$$
$$\left. + 2q^2 \ln 2 \right]$$

implying that a one-way rate-sum capacity of $C_{nf}^{(sum)} = 0.4338A$ (nats per second) is attainable with $q = 0.2659$.

2) The Cover–Leung Inner Bound on the Feedback Capacity Region: Having obtained a discrete-time decomposition for our continuous-time model, we compute now an inner bound for the feedback capacity region based on (12).

For general $P_{UX_1X_2Y}$, we define $q_{1u} = \Pr(X_1 = 1|U = u)$ and $q_{2u} = \Pr(X_2 = 1|U = u)$, then straightforward computation yields

$$\frac{I(Y; X_1X_2)}{\Delta} = \sum_u p(u) \left\{ -(q_{1u} + q_{2u}) \right]$$
$$\cdot \ln \left[\sum_u p(u)(q_{1u} + q_{2u}) \right] + 2q_{1u}q_{2u} \ln 2 \right\} \tag{53}$$

$$\frac{I(Y; X_2|X_1U)}{\Delta} = \sum_u p(u) \left[-(1 - q_{1u})q_{2u} \ln q_{2u} \right.$$
$$\left. - q_{1u}(1 + q_{2u}) \ln(1 + q_{2u}) + 2q_{1u}q_{2u} \ln 2 \right]$$

$$\frac{I(Y; X_1|X_2U)}{\Delta} = \sum_u p(u) \left[-(1 - q_{2u})q_{1u} \ln q_{1u} \right.$$
$$\left. - q_{2u}(1 + q_{1u}) \ln(1 + q_{1u}) + 2q_{1u}q_{2u} \ln 2 \right]. \tag{54}$$

We seek to maximize (53) subject to the average-power constraint $\sum_u p(u)[q_{1u} + q_{2u}] \leq 2\sigma$. Fixing for any $U = u$ the sum $q_{1u} + q_{2u}$ it follows immediately that

$$I(Y; X_1X_2, U = u)/\Delta$$
$$= -(q_{1u} + q_{2u}) \ln(q_{1u} + q_{2u}) + 2q_{1u}q_{2u} \ln 2$$

is maximized by letting $q_{1u} = q_{2u} \triangleq q_u$. Turning to (54), fixing for any $U = u$ the sum $q_{1u} + q_{2u}$, the function

$$[I(Y; X_2|X_1, U = u) + I(Y; X_1|X_2, U = u)]/\Delta$$
$$= -(1 - q_{1u})q_{2u} \ln q_{2u} - q_{1u}(1 + q_{2u}) \ln(1 + q_{2u})$$
$$- (1 - q_{2u})q_{1u} \ln q_{1u} - q_{2u}(1 + q_{1u}) \ln(1 + q_{1u})$$
$$+ 4q_{1u}q_{2u} \ln 2$$

is maximized by letting $q_{1u} = q_{2u} \triangleq q_u$. It follows that

$$I(Y; X_1X_2)/\Delta$$
$$= 2 \sum_u p(u) \left\{ -q_u \ln \left[2 \sum_u p(u)q_u \right] + q_u^2 \ln 2 \right\} \tag{55}$$

$$I(Y; X_2|X_1U)/\Delta = I(Y; X_1|X_2U)/\Delta$$
$$= \sum_u p(u) \left[-(1 - q_u)q_u \ln q_u \right.$$
$$\left. - q_u(1 + q_u) \ln(1 + q_u) + 2q_u^2 \ln 2 \right]. \tag{56}$$

Table I summarizes the attainable symmetric rates $R_1 = R_2$ via (56), and the upper bound on the rate-sum as given by (55), for $|\mathcal{U}| = 2, 3, 4$. These results have been obtained by numerical search over general $p(u)$, q_u. The search does not give any improvement in R_1 for the case $|\mathcal{U}| = 4$; in fact, when forcing all four mass points to have strictly positive probability (as we do) the rate slightly decreases.

B. A Coding Scheme

Transmission consists of a sequence of B generic transmission phases, each of which is executed during a length T interval. Furthermore, each generic transmission phase is split into the following three parts.

- A multiple-access phase the duration of which is bounded by $(1 - \beta_1 - \beta_2)T$. This interval is further partitioned into equi-length arbitrarily small subintervals labeled as $I_{0,\vartheta}$, $\vartheta = 1, 2, \ldots, K_0$, by taking K_0 to be arbitrarily large.

- A two-way phase the duration of which is bounded by $\beta_1 T$. This interval is further partitioned into equi-length arbitrarily small subintervals labeled as $I_{1,\vartheta}$, $\vartheta = 1, 2, \ldots, K_1$, by taking K_1 to be arbitrarily large.

- A total-cooperation phase the duration of which is bounded by $\beta_2 T$. This interval is further partitioned into equi-length arbitrarily small subintervals labeled as $I_{2,\vartheta}$, $\vartheta = 1, 2, \ldots, K_2$, by taking K_2 to be arbitrarily large.

In fact, the total cooperation phase does not conform with the coding scheme presented in Section V. However, for the sake of brevity we allow both encoders to split the resolution information they wish to send to the decoder into two parts; the first part is conveyed via the random variable (r.v.) U during the MAC phase while the rest is conveyed during the total cooperation phase. Thus, in the forthcoming analysis we shall take into account the impact of the total cooperation phase with the aim of showing that the optimal solution renders this phase superfluous and the resolution information is communicated just via U.

A binary n-length codeword $\boldsymbol{x}_\mu$, $\mu = 1, 2$ is mapped (by Encoder μ) for transmission during the MAC phase as follows. An

interval $I_{0,\vartheta}$ is partitioned into n equi-length half-open closed on the right subintervals I_ℓ, $\ell = 1, 2, \ldots, n$, such that $I_\ell \triangleq \left(\frac{I_{0,\vartheta}}{n}(\ell-1), \frac{I_{0,\vartheta}}{n}\ell\right]$ and the waveform $x_\mu(t)$, $t \in I_{0,\vartheta}$ is defined as

$$x_\mu(t) = \begin{cases} A, & (t \in I_\ell) \cap (\boldsymbol{x}_{\mu,\ell} = 1) \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{x}_{\mu,\ell}$ denotes the ℓth component of $\boldsymbol{x}_\mu$. Thus, for example, the aggregate block length of the MAC phase is nK_0. Similar mappings, of binary n-length codewords into waveforms over the subintervals $I_{1,\vartheta}$ and $I_{2,\vartheta}$, are used during the two-way and total-cooperation phases, respectively.

During each of the aforementioned phases both encoders execute sequences of *coding steps*. A coding step consists of the generation of a pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ of binary n-length codewords (as explained in the sequel) and a repeated transmission of the corresponding waveforms $(x_1(t), x_2(t))$ by sequentially consuming $I_{j,\vartheta}$, $j = 0, 1, 2$ subintervals until the first arrival is registered at the receiver.

The Multiple-Access Phase

During the MAC phase of generic transmission phase b, $b = 1, \ldots, B-1$, the encoders execute κ coding steps while using codebooks generated via the joint distribution (11) as follows.

- For $b = 1$, generate a single sequence

$$\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$$

  with probability

$$\Pr\{\boldsymbol{u}\} = \prod_{k=1}^n P_U(u_k).$$

  Label this sequence $\boldsymbol{u}^{(1)}(1)$.

- For $b \geq 2$, let $\mathcal{U}^{(b)}$ be a collection of $|\Lambda|$ binary n-length sequences drawn independently according to the product distribution of marginal $P_U(u)$. Index $\mathcal{U}^{(b)}$ by Λ, thus, a generic element of $\mathcal{U}^{(b)}$ can be expressed as $\boldsymbol{u}^{(b)}(\lambda)$, $\lambda \in \Lambda$.

Now perform the following.

1) Based on every $\boldsymbol{u}^{(b)}(\lambda)$ generate per each coding step ℓ, $\ell = 1, \ldots, \kappa$, M_1 sequences

$$\boldsymbol{x}_1^{(b,\ell)} = \left(x_{1,1}^{(b,\ell)}, x_{1,2}^{(b,\ell)}, \ldots, x_{1,n}^{(b,\ell)}\right)$$

  each with probability

$$\Pr\left\{\boldsymbol{x}_1^{(b,\ell)} | \boldsymbol{u}^{(b)}(\lambda)\right\} = \prod_{k=1}^n P_{X_1|U}\left(x_{1,k}^{(b,\ell)} | u_k^{(b)}\right)$$

  Label them

$$\boldsymbol{x}_1^{(b,\ell)}\left(\boldsymbol{u}^{(b)}(\lambda), w_1\right) = \boldsymbol{x}_1^{(b,\ell)}(\lambda, w_1)$$

  for $\lambda \in \Lambda$ and $w_1 \in \mathcal{M}_1$.

2) Similarly, based on every $\boldsymbol{u}^{(b)}(\lambda)$, generate per each coding step ℓ, $\ell = 1, \ldots, \kappa$, M_2 sequences

$$\boldsymbol{x}_2^{(b,\ell)} = \left(x_{2,1}^{(b,\ell)}, x_{2,2}^{(b,\ell)}, \ldots, x_{2,n}^{(b,\ell)}\right)$$

each with probability

$$\Pr\left\{\boldsymbol{x}_2^{(b,\ell)} | \boldsymbol{u}^{(b)}(\lambda)\right\} = \prod_{k=1}^n P_{X_2|U}\left(x_{2,k}^{(b,\ell)} | u_k^{(b)}\right).$$

Label them

$$\boldsymbol{x}_2^{(b,\ell)}\left(\boldsymbol{u}^{(b)}(\lambda), w_2\right) = \boldsymbol{x}_2^{(b,\ell)}(\lambda, w_2)$$

for $\lambda \in \Lambda$ and $w_2 \in \mathcal{M}_2$.

With a slight abuse of notation we shall henceforth refer just to $\boldsymbol{x}_\mu^{(b)}\left(\boldsymbol{u}^{(b)}(\lambda), w_\mu\right)$, $\mu = 1, 2$, while keeping in mind that per each arrival the relevant codebook is considered.

We next define the deterministic mappings (g_1, g_2) for the channel (52) (with $s = 0$) by (13).

This choice and the fact that the event $\{Y_k = 0\}$ does not convey any information for the MAC (52), imply a "thinning" on the time-discrete process $\{Y_k\}$ leaving just those coordinates in which $y_k = 1$ relevant. As a result, the effective block length of the MAC phase, for the purpose of computing the expressions in (10), reduces to $\kappa = nK_0 \Pr[Y_k = 1]$, i.e., the number of coding steps. Consequently, the case at hand—for which we shall analyze the region defined by Corollary 1—corresponds to the MAC (1) where $q = \Delta \ll 1$; the MAC phase block length is $nK_0 = (1 - \beta_1 - \beta_2)TA/\Delta$; the effective MAC phase block length is $\kappa = 2(1 - \beta_1 - \beta_2)TA\mathbb{E}(X_\mu)$ (as shown later on). Thus, while the ratio $\kappa/(nK_0) = 2\mathbb{E}(X_\mu)\Delta$ is permitted to be arbitrarily small, the product qnK_0 is held fixed.

With a slight abuse of notation, we shall henceforth denote the restriction of Y^{nK_0} to these κ coordinates by Y^κ.

Consider the κ arrivals registered during a MAC phase with the encoders signaling at the same information rate and same average power.

Let $A_{01} \subset \{1, \ldots, \kappa\}$ be the set of photons that arrived while Encoder 1 was off and Encoder 2 was on. Similarly, let $A_{10} \subset \{1, \ldots, \kappa\}$ be the set of photons that arrived while Encoder 2 was off and Encoder 1 was on. Let $\zeta_1 \triangleq \Pr\{(1,0)\} = \Pr\{(0,1)\}$ be the probability that just one encoder was on when an arrival was registered, and let $\mathcal{A}_\delta$ be the set of all subsets of $\{1, \ldots, \kappa\}$ of cardinality $\kappa(\zeta_1 \pm \delta)$. We note that for sufficiently large κ with high probability A_{01} and A_{10} will fall within $\mathcal{A}_\delta$ (which is the set of typical sequences of A_{01} and A_{10}). If A_{01} or A_{10} does not belong to $\mathcal{A}_\delta$ an error is declared by the corresponding encoder.

It follows that

$$\begin{aligned} H\left(A_{01}^{(b)} | Y^\kappa\right) &= H\left(A_{10}^{(b)} | Y^\kappa\right) \\ &= |\mathcal{A}_\delta| = \binom{\kappa}{\kappa(\zeta_1 \pm \delta)} \approx \exp\{\kappa h(\zeta_1)\} \\ &\triangleq N_{\mu,1}^{(1)}, \qquad \mu = 1, 2 \end{aligned} \qquad (57)$$

where the superscript $^{(1)}$ indicates that $N_{\mu,1}^{(1)}$ (which is the relevant size of $\mathcal{A}_\delta$ during the MAC phase) is the message set size for the information exchange during two-way step 1, and $h(\cdot)$ is the binary entropy function.

Next, we claim that given A_{01}, the set A_{10} is actually located within a strictly smaller subset: $\mathcal{D}_{10}(A_{01})$ of $\mathcal{A}_\delta$. To see this,

note that since Encoder 1 already knows A_{01}, and since $A_{01} \cap A_{10} = \emptyset$, A_{10} should be somewhere within

$$\mathcal{D}_{10}(A_{01}) \triangleq \{\{1, \ldots, \kappa\} \setminus A_{01}\}.$$

It follows that given A_{01}, any typical member of A_{10} can be described as a $\kappa(\zeta_1 \pm \delta)$ subset within a set of size $\kappa(1 - \zeta_1 \pm \delta)$. That is,

$$
\begin{aligned}
H\left(A_{10}^{(b)} | A_{01}^{(b)}, Y^\kappa\right) &= H\left(A_{01}^{(b)} | A_{10}^{(b)}, Y^\kappa\right) \\
&= |\mathcal{D}_{10}(A_{01})| = |\mathcal{D}_{01}(A_{10})| \\
&= \binom{\kappa(1 - \zeta_1 \pm \delta)}{\kappa(\zeta_1 \pm \delta)} \\
&\approx \exp\left\{\kappa(1 - \zeta_1) h\left(\frac{\zeta_1}{1 - \zeta_1}\right)\right\} \\
&\triangleq N_{\mu,2}^{(1)}, \qquad \mu = 1, 2 \qquad (58)
\end{aligned}
$$

as similarly, given A_{10} any typical member of A_{01} is located within

$$\mathcal{D}_{01}(A_{10}) \triangleq \{\{1, \ldots, \kappa\} \setminus A_{10}\} \text{ and } |\mathcal{D}_{01}(A_{10})| = N_{\mu,2}^{(1)}.$$

Upon the termination of the MAC block of generic transmission phase b, the decoder performs *joint decoding* as follows. It generates an $(M_1 M_2, |\Lambda|)$ matrix C_d the columns of which correspond to the $|\Lambda|$ possible cloud centers $\lambda^{(b-1)}$ while the rows of which correspond to the $M_1 M_2$ possible pairs $(w_1^{(b)}, w_2^{(b)})$. Initially, all entries in C_d are "active"—i.e., they are set to "1."

The decoder considers first all arrivals ℓ which support just $\boldsymbol{x}_2^{(b,\ell)}$ (it knows these arrivals from $A_{01}^{(b)}$). For each such arrival the decoder looks at any active entry (i, j) in C_d and considers the codeword pair

$$\left(\boldsymbol{x}_1^{(b,\ell)}\left(\boldsymbol{u}^{(b)}(j), w_1^{(b)}(i)\right), \boldsymbol{x}_2^{(b,\ell)}\left(\boldsymbol{u}^{(b)}(j), w_2^{(b)}(i)\right)\right) \quad (59)$$

where $(w_1^{(b)}(i), w_2^{(b)}(i))$ denotes the fact that $(w_1^{(b)}, w_2^{(b)})$ is determined by the row index i. If the pair (59) supports the arrival just with $\boldsymbol{x}_2$, entry (i, j) remains active, otherwise, it is "deactivated"—i.e., it is set to "0." Having considered the $A_{01}^{(b)}$ arrivals, the decoder turns to the $A_{10}^{(b)}$ arrivals that support just $\boldsymbol{x}_1^{(b,\ell)}$. For each such arrival ℓ and an active entry (i, j), the decoder considers the pair (59) and if it supports the arrival just with $\boldsymbol{x}_1$, entry (i, j) remains active, otherwise it is deactivated. Turning to those arrivals supporting both $\boldsymbol{x}_1^{(b,\ell)} \cap \boldsymbol{x}_2^{(b,\ell)}$, for each such arrival ℓ and a remaining active entry (i, j), the decoder considers the pair (59) and if both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ support the arrival, entry (i, j) remains active, otherwise it is deactivated.

Upon the termination of this procedure, the decoder should find but a *single* surviving active column j^* in C_d that defines (with high probability) the correct cloud center $\lambda^{(b-1)}$ (if more than one column survives an error is declared). The decoder sets $\lambda^{(b-1)} = j^*$, and forms a list $\mathcal{L}_{1,2} = \{(w_1^{(b)}(i), w_2^{(b)}(i))\}$ of the remaining active entries in the surviving column j^* of C_d. This forms the *second part* of the resolution information that the decoder needs and which is sent via $\lambda^{(b)}$.

The Two-Way Phase

The communication instance faced by both encoders upon the termination of the multiaccess phase can be stated as follows. Consider a bipartite graph with $N_{2,1}^{(1)}$ right nodes and $N_{1,1}^{(1)}$ left

nodes. A right node represents an outcome of the random variable A_{10}, while a left node represents an outcome of the r.v. A_{01}. Each left node has $N_{1,2}^{(1)}$ outgoing edges connected to the corresponding $|\mathcal{D}_{10}(A_{01})|$ right nodes, and each right node has $N_{2,2}^{(1)}$ outgoing edges connected to the corresponding $|\mathcal{D}_{01}(A_{10})|$ left nodes. Encoder 1 who knows the correct left node seeks to resolve the correct right node, while Encoder 2 who knows the correct right node seeks to determine the correct left node. Moreover, the r.v. pair $(A_{01}^{(b)}, A_{10}^{(b)})$ satisfies

$$
\begin{aligned}
H\left(A_{01}^{(b)} | Y^\kappa\right) &> H\left(A_{01}^{(b)} | A_{10}^{(b)}, Y^\kappa\right) \\
H\left(A_{10}^{(b)} | Y^\kappa\right) &> H\left(A_{10}^{(b)} | A_{01}^{(b)}, Y^\kappa\right) \\
I\left(X_1^\kappa; Y^\kappa | A_{01}^{(b)}, A_{10}^{(b)}\right) &= I\left(X_1^\kappa; Y^\kappa | A_{01}^{(b)}, X_2^\kappa\right) \\
I\left(X_2^\kappa; Y^\kappa | A_{01}^{(b)}, A_{10}^{(b)}\right) &= I\left(X_2^\kappa; Y^\kappa | A_{10}^{(b)}, X_1^\kappa\right) \\
H\left(A_{10}^{(b)} | A_{01}^{(b)}, Y^\kappa\right) &= H\left(A_{10}^{(b)} | X_1^\kappa, Y^\kappa\right) \\
H\left(A_{01}^{(b)} | A_{10}^{(b)}, Y^\kappa\right) &= H\left(A_{01}^{(b)} | X_2^\kappa, Y^\kappa\right). \qquad (60)
\end{aligned}
$$

Consequently, the mappings (13) which generate $(A_{01}^{(b)}, A_{10}^{(b)})$ satisfy (40) and furthermore this is a pair of "good" mappings in the sense that (25) of Lemma 4 holds. Furthermore, as a consequence of the equalities in (60), the list size $\mathcal{L}_2 < \exp\{n\epsilon\}$ since having acquired $A_{01}^{(b)}$ the decoder is in the same position as Encoder 1 was before the execution of the two-way phase, a fact which renders the second stage of the successive decoding unnecessary.

For the two-way phase we let $\tilde{v}_\mu = \tilde{g}_\mu(\tilde{x}_\mu, \tilde{y})$ be defined as in (14).

A two-way phase consists of a sequence of L two-way steps, wherein during step l, $l = 1, 2, \ldots, L$, the encoders execute κ_l coding steps (and decode them) as explained subsequently.

Consider the first step of the two-way phase of generic transmission phase b. Let both encoders use an $(N_{1,1}^{(1)}, q, N_{2,1}^{(1)}, q)$ two-user code which is expressed via the joint distribution (6) with

$$P_{\tilde{X}_1}(\tilde{x} = 1) = P_{\tilde{X}_2}(\tilde{x} = 1) = q$$

(not to be confused with q in (1) which in the model analyzed here is replaced by Δ), and is generated independently κ_1 times (i.e., once per each coding step) as follows.

- Generate $N_{1,1}^{(1)}$ sequences

$$\tilde{\boldsymbol{x}}_1 = (\tilde{x}_{1,1}, \tilde{x}_{1,2}, \ldots, \tilde{x}_{1,n})$$

each with probability

$$\Pr\{\tilde{\boldsymbol{x}}_1\} = \prod_{k=1}^{n} P_{\tilde{X}_1}(\tilde{x}_{1,k}).$$

Label them $\tilde{\boldsymbol{x}}_1(w_1)$, $w_1 \in \{1, 2, \ldots, N_{1,1}^{(1)}\}$.

- Generate $N_{2,1}^{(1)}$ sequences

$$\tilde{\boldsymbol{x}}_2 = (\tilde{x}_{2,1}, \tilde{x}_{2,2}, \ldots, \tilde{x}_{2,n})$$

each with probability

$$\Pr\{\tilde{\boldsymbol{x}}_2\} = \prod_{k=1}^{n} P_{\tilde{X}_2}(\tilde{x}_{2,k}).$$

Label them $\tilde{\boldsymbol{x}}_2(w_2)$, $w_2 \in \{1, 2, \ldots, N_{2,1}^{(1)}\}$.

Consider (for example) the decoding of W_1 as performed by Encoder 2 based on knowing the realization of the κ_1 arrivals. Suppose now that arrival ℓ, $\ell \in \{1, 2, \ldots, \kappa_1\}$ falls where the transmitted message is on. Let $S^{(\ell-1)}$ be the survivors set before arrival ℓ has been considered and let $s^{(\ell-1)}$ denote the survivors set size. Let Q_ℓ be the random variable that equals the number of messages within $S^{(\ell-1)}$ which have "1" on the instant of arrival ℓ. Given that arrival ℓ, $\ell \in \{1, 2, \ldots, \kappa_1\}$ falls where the transmitted message is on we have

$$Q_\ell \sim B[s^{(\ell-1)} - 1, q] + 1$$

where $B[m, p]$ is a binomial random variable with parameters m and p.

It follows that

$$\frac{1}{s^{(\ell-1)}} \mathbb{E}[Q_\ell] = q + \frac{1-q}{s^{(\ell-1)}}$$

and provided that $s^{(\ell-1)}$ is sufficiently large the reduction factor of $S^{(\ell-1)}$ is q. Similarly, given that arrival ℓ falls where the transmitted message is off we have

$$Q_\ell \sim B[s^{(\ell-1)} - 1, q]$$

and the reduction factor of $S^{(\ell-1)}$ is

$$1 - \frac{1}{s^{(\ell-1)}} \mathbb{E}[Q_\ell] = 1 - q + \frac{q}{s^{(\ell-1)}}.$$

Let $\{1, 2, \ldots, \kappa_1\} = \mathcal{G}_0 \cup \mathcal{G}_1$, where $\mathcal{G}_0$ is the subset of arrivals which fall where the transmitted message is off and $\mathcal{G}_1$ is the subset of arrivals which fall where the transmitted message is on. Furthermore, let q_ℓ and $1 - q_\ell$ be the reduction factors when $\ell \in \mathcal{G}_1$ or $\ell \in \mathcal{G}_0$, respectively, then correct decoding of a codebook with size M_μ is assured provided that

$$\Pr\left[M_\mu \prod_{\ell \in \mathcal{G}_1} q_\ell \prod_{\ell \in \mathcal{G}_0} (1 - q_\ell) > 1 \right] \leq \epsilon. \quad (61)$$

By Markov's inequality and the fact that given $s^{(\ell-1)}$ the random variable q_ℓ is independent of q_k, $k < \ell$

$$\Pr\left[M_\mu \prod_{\ell \in \mathcal{G}_1} q_\ell \prod_{\ell \in \mathcal{G}_0} (1 - q_\ell) > 1 \right]$$

$$\leq \mathbb{E}\left[M_\mu \prod_{\ell \in \mathcal{G}_1} q_\ell \prod_{\ell \in \mathcal{G}_0} (1 - q_\ell) \right]$$

$$= M_\mu \prod_{\ell \in \mathcal{G}_1} \mathbb{E}[q_\ell] \prod_{\ell \in \mathcal{G}_0} \mathbb{E}[1 - q_\ell].$$

Hence, if we take

$$\prod_{\ell \in \mathcal{G}_1} q \prod_{\ell \in \mathcal{G}_0} (1 - q) \leq M_\mu^{-1} e^{-\kappa_1 \epsilon_1}$$

inequality (61) is satisfied for κ_1 sufficiently large and $\epsilon_1 > 0$.

Upon the termination of the two-way phase, both encoders know the exact realization of $(A_{01}^{(b)}, A_{10}^{(b)})$ during the MAC phase (via backward decoding two-way step L, then two-way step $L - 1$ up to two-way step 1). Had the decoder been aware of the sequence, say

$$\overbrace{A_{01}^{(b)}}^{\text{MAC}}, \overbrace{A_{01}^{(1)}, A_{01}^{(2)}, \ldots, A_{01}^{(L-1)}}^{\text{two-way}}$$

this would put him precisely in the same position as Encoder 1 was, and consequently, he could as well decode the two-way phase in the same way as Encoder 1 does. To this end, it is assumed that knowing the realization of $(A_{01}^{(l)}, A_{10}^{(l)})$, $l = L, L - 1, \ldots, 2$ (i.e., the realization of the κ_l arrivals during step l of the two-way phase), the decoder performs list decoding on $A_{01}^{(l-1)}$ as explained in the sequel. Thus, it remains for both encoders to provide the decoder with the exact index of $A_{01}^{(l-1)}$ within the decoder's final list.

Suppose that a genie informs the decoder on $(A_{01}^{(l)}, A_{10}^{(l)})$, that is, which $\kappa_l(1+q)/2$ arrivals support Encoder 1's message and which $\kappa_l(1+q)/2$ support Encoder 2's message during step l of the two-way phase. The decoder performs a similar decoding procedure as Encoder 2 does starting with $H(A_{01}^{(l-1)} | \tilde{Y}^{\kappa_l - 1}) = N_{1,1}^{(l)}$ possible messages and reducing the surviving messages set size by a factor of q whenever the corresponding arrival supports the transmitted waveform, and by a factor of $1 - q$ whenever the arrival does not support the transmitted waveform. Taking into account all κ_l arrivals this provides a reduction by factor of $1/N_{2,2}^{(l)}$. Thus, eventually the list decoder ends with a list of possible messages that could have been sent by Encoder 1 the size of which is $N_{1,1}^{(l)}/N_{2,2}^{(l)}$.

Combining (57) and (58), it follows that the list size exponent, corresponding to step 1 of the two-way phase equals

$$\ln \left| \mathcal{L}_1^{(1)} \right| = \ln \frac{N_{1,1}^{(1)}}{N_{2,2}^{(1)}}$$

$$= \kappa \left[h(\zeta_1) - (1 - \zeta_1) h\left(\frac{\zeta_1}{1 - \zeta_1} \right) \right] \triangleq r_1 \quad (62)$$

while the list size exponent, corresponding to step l, $l = 2, \ldots, L - 1$, of the two-way phase equals

$$\ln |\mathcal{L}_1^{(l)}| = \ln \frac{N_{1,1}^{(l)}}{N_{2,2}^{(l)}}$$

$$= \kappa_{l-1} \left[h\left(\frac{1-q}{2} \right) - \frac{(1+q)}{2} h\left(\frac{2q}{1+q} \right) \right]$$

$$\triangleq r_l. \quad (63)$$

Performance evaluation

For the MAC phase consider a joint distribution P_{U, X_1, X_2} where all random variables are binary. Let U be distributed according to the law $\Pr\{U = 1\} = t$ and let (X_1, X_2) be a pair of binary random variables that are conditionally independent given U and satisfy

$$P_{X_\mu | U}(x_\mu | u) = \begin{cases} (0|0), & 1 - r_0 \\ (0|1), & 1 - r_1, \end{cases} \mu = 1, 2. \quad (64)$$

For this code one obtains

$$\mathbb{E}(X_\mu) = \Pr\{X_\mu = 1\} = (1 - t)r_0 + t r_1 \triangleq \tilde{q}$$

where $\tilde{q}$ denotes the average power transmitted by each encoder.

Any transmitted two-user waveform on $I_{0,\vartheta}$ induces three independent Poisson processes with corresponding rates

$$\Lambda_1 = \Lambda_2 = \Pr[X_1 X_2 = 01](1 - \beta_1 - \beta_2) AT / K_0$$

and

$$\Lambda_0 = 2 \Pr[X_1 X_2 = 11](1 - \beta_1 - \beta_2)AT/K_0$$

(the joint statistics $P_{X_1 X_2}$ is given in (74)). Furthermore, our coding scheme satisfies for any positive ϑ

$$\mathbb{E}\left[\int_{I_{0,\vartheta}} \boldsymbol{x}_\mu(\lambda, m_\mu)dt\right] = \tilde{q}A(1 - \beta_1 - \beta_2)T/K_0, \quad \mu = 1, 2.$$

Given that a count was registered, let Z_μ, $\mu = 1, 2$, be the binary random variable indicating whether the arrival falls within the support of Encoder μ's message. The joint distribution of (Z_1, Z_2) equals

$$P_{Z_1 Z_2}(z_1, z_2) = \begin{cases} (0, 1), & \frac{\Pr[X_1 X_2 = 01]}{2\tilde{q}} \triangleq \zeta_1 \\ (1, 0), & \zeta_1 \\ (1, 1), & 1 - 2\zeta_1 \triangleq \zeta_2. \end{cases}$$

Suppose now that κ arrivals have been registered during the MAC phase of generic transmission phase b, then (57) and (58) hold with ζ_1 as defined above.

These κ arrivals are used by the encoders to decode $(w_1^{(b)}, w_2^{(b)})$ and it is shown in Appendix I that

$$M_\mu = \left[\eta_{e,01}^{\zeta_1} \eta_{e,11}^{\zeta_2} \eta_{e,10}^{\zeta_1}\right]^{-\kappa}, \qquad \mu = 1, 2$$

where $\eta_{e,01}$, $\eta_{e,10}$, and $\eta_{e,11}$ are defined by (75), (76), and (77), respectively. Assuming that this amount is communicated per an interval of duration T, the information rate-sum equals

$$\begin{aligned} R &= R_1 + R_2 \\ &= \frac{\kappa}{T}[-2\zeta_2 \ln \eta_{e,11} - 2\zeta_1(\ln \eta_{e,10} + \ln \eta_{e,01})] \\ &\triangleq \frac{\kappa}{T}\tilde{f}(t, \vec{r}). \end{aligned} \quad (65)$$

Define the random variable $\tau_\ell^{(K_0)}$ $\ell = 1, 2, \ldots, \kappa$ as the duration time for the completion of coding step ℓ of a MAC phase. The distribution of $\tau_\ell^{(K_0)}$ converges weakly (via similar arguments as in [19]), as K_0 goes to infinity, to the exponential distribution with parameter $2\tilde{q}A$. Since $\mathbb{E}(\tau_\ell^{(K_0)}) = 1/(2\tilde{q}A)$, the probability that this phase of our coding scheme fails on an interval of length $(1 - \beta_1 - \beta_2)T$ will be exponentially small once

$$\mathbb{E}\left(\sum_{\ell=1}^{\kappa} \tau_\ell^{(K_0)}\right) = \frac{\kappa}{2\tilde{q}A} \le (1 - \beta_1 - \beta_2)T. \quad (66)$$

In fact, the last inequality expresses the effective block length $\kappa = nK_0 \Pr[Y_k = 1]$ in terms of the MAC phase duration $(1 - \beta_1 - \beta_2)T$ and $\mathbb{E}(X_\mu)$.

So far, the RHS of (65) provides us with an expression for

$$R_1 + R_2 = I(X_1; Y | X_2, U, V_1) + I(X_2; Y | X_1, U, V_2)$$

in terms of the parameters of our coding scheme, while the maximization of this quantity is to be carried out under the constraint (66).

Recall that for the two-way phase we choose a joint distribution

$$P_{\tilde{X}_1, \tilde{X}_2}(\tilde{x}_1, \tilde{x}_2) = P_{\tilde{X}}(\tilde{x}_1)P_{\tilde{X}}(\tilde{x}_2)$$

where $P_{\tilde{X}}(\tilde{x} = 1) = q$.

Consider next the sequences of coding steps $\kappa_1, \kappa_2, \ldots,$ κ_{L-1} executed during steps $1, 2, \ldots, L - 1$ of a two-way phase.

In this case, any transmitted two-user waveform on $I_{1,\vartheta}$ induces three independent Poisson processes with corresponding rates $\Lambda_1 = \Lambda_2 = q(1 - q)\beta_1 AT/K_1$ and $\Lambda_0 = 2q^2 \beta_1 AT/K_1$. Furthermore, our coding scheme satisfies for any positive ϑ

$$\mathbb{E}\left[\int_{I_{1,\vartheta}} \tilde{\boldsymbol{x}}_\mu(w_\mu)dt\right] = q\beta_1 T/K_1, \qquad \mu = 1, 2.$$

Given that a count was registered let Z_μ, $\mu = 1, 2$ be the binary random variable indicating whether the arrival falls within the support of sender μ's message. The joint distribution of (Z_1, Z_2) equals

$$P_{Z_1 Z_2}(z_1, z_2) = \begin{cases} (0, 1), & \frac{1-q}{2} \\ (1, 0), & \frac{1-q}{2} \\ (1, 1), & q. \end{cases}$$

Consequently, (61) implies that

$$N_{\mu,2}^{(1)} = q^{-\frac{(1+q)}{2}\kappa_1}(1 - q)^{-\frac{(1-q)}{2}\kappa_1}, \qquad \mu = 1, 2$$

which combined together with (58) yields

$$\begin{aligned} \kappa_1 &= \frac{2(1 - \zeta_1)h(\frac{\zeta_1}{1-\zeta_1})}{g(q)}\kappa \\ g(q) &\triangleq -\{(1 + q)\ln q + (1 - q)\ln(1 - q)\}. \end{aligned} \quad (67)$$

As a result of the first step of the two-way phase, both encoders need to exchange the new correlated information that is contained in $\{1, \ldots, \kappa_1\}$. This time, the relevant sets sizes are

$$\begin{aligned} |\mathcal{A}_\delta| &= \binom{\kappa_1}{\kappa_1 \left(\frac{1-q}{2} \pm \delta\right)} \\ &\approx \exp\left\{\kappa_1 h\left(\frac{1-q}{2}\right)\right\} \triangleq N_{\mu,1}^{(2)} \\ \left|\mathcal{D}_{10}\left(A_{01}^{(1)}\right)\right| &= \binom{\kappa_1 \left(\frac{1+q}{2} \pm \delta\right)}{\kappa_1 \left(\frac{1-q}{2} \pm \delta\right)} \\ &\approx \exp\left\{\kappa_1 \frac{1+q}{2}h\left(\frac{2q}{1+q}\right)\right\} \triangleq N_{\mu,2}^{(2)} \end{aligned}$$

where it is implicitly assumed (with a slight abuse of notation) that $\mathcal{A}_\delta$ refers now to the set of all typical subsets of $\{1, \ldots, \kappa_1\}$ of cardinality $\kappa_1(\frac{1-q}{2} \pm \delta)$ and the notation $\mathcal{D}_{10}(A_{01}^{(1)})$ is interpreted similarly.

Again, let both encoders use during step 2 of the two-way phase an $(N_{1,1}^{(2)}, q, N_{2,1}^{(2)}, q)$ two-user code and feedback decode it this time based on κ_2 photon arrivals. This implies that

$$N_{\mu,2}^{(2)} = q^{-\frac{(1+q)}{2}\kappa_2}(1 - q)^{-\frac{(1-q)}{2}\kappa_2}, \qquad \mu = 1, 2$$

which yields

$$\kappa_2 = \frac{(1 + q)h(\frac{2q}{1+q})}{g(q)}\kappa_1.$$

Continuing this way with steps $3, 4, \ldots, L-1$ of the two-way phase, the number of required coding steps is upper-bounded by

$$
\begin{aligned}
\kappa_{TW} &< \sum_{\ell=1}^{\infty} \kappa_\ell \\
&= \kappa \frac{2(1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)}{g(q)}\left\{1 + \frac{(1+q)h\left(\frac{2q}{1+q}\right)}{g(q)}\{1 + \cdots \right. \\
&= \frac{2(1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)}{g(q)-(1+q)h\left(\frac{2q}{1+q}\right)}\kappa \triangleq \tilde{\kappa}_1
\end{aligned} \tag{68}
$$

Furthermore, recalling (62) and (63), the list-size exponent imposed by the two-way phase can be upper-bounded as

$$
\begin{aligned}
TR_L^{(TW)} \\
< \sum_{l=1}^{\infty} r_l &= \kappa\left[h(\zeta_1) - (1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)\right] \\
&+ \left[h\left(\frac{1-q}{2}\right) - \frac{(1+q)}{2}h\left(\frac{2q}{1+q}\right)\right]\cdot\sum_{\ell=1}^{\infty}\kappa_\ell \\
&= \kappa\left\{h(\zeta_1) - (1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)\right. \\
&\left. + \frac{2(1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)\left[h\left(\frac{1-q}{2}\right) - \frac{1+q}{2}h\left(\frac{2q}{1+q}\right)\right]}{g(q)-(1+q)h\left(\frac{2q}{1+q}\right)}\right\} \\
&\triangleq \kappa\psi(\zeta_1, q).
\end{aligned}
$$

Thus, the RHS of (68) provides us with an asymptotic upper bound on the ratio of the number of coding steps—κ_{TW}—that have been executed during the two-way phase to—κ—the number of coding steps executed during the MAC phase. Combined with (69) this yields an attainable pair $(\kappa_{TW}/\kappa, TR_L^{(TW)})$ for our scheme.

Define the random variable $\xi_\ell^{(K_1)}$ $\ell = 1, 2, \ldots, \tilde{\kappa}_1$ as the duration time for the completion of coding step ℓ of the two-way phase. The probability that this phase of our coding scheme fails on an interval of length $\beta_1 T$ will be exponentially small once

$$
\mathbb{E}\left(\sum_{\ell=1}^{\tilde{\kappa}_1}\xi_\ell^{(K_1)}\right) = \frac{\tilde{\kappa}_1}{2qA} \leq \beta_1 T. \tag{69}
$$

We consider next the rate-sum obtained by our decoder during the MAC phase. Recall from (47) and (48) that if the random variable U provides an amount of resolution information the exponential size of which is upper-bounded by $A^{(\text{MAC})} + A^{(TW)}$ whereas

$$
A^{(\text{MAC})} = T[R_1 + R_2 - I(X_1, X_2; Y|U, V_1, V_2)]
$$

then

$$
A^{(TW)} < T[I(X_1, X_2; Y|V_1, V_2) - (R_1 + R_2)]. \tag{70}
$$

Consequently, we turn our attention to the evaluation of the RHS of (70) for our coding scheme. Insofar as we assume that part of the resolution information $R_L^{(TW)}$ (for the two-way phase) is

communicated during the total cooperation phase it follows that this quantity is given by

$$
A^{(TC)} = TR_L^{(TW)} - T[I(X_1, X_2; Y|V_1, V_2) - (R_1 + R_2)]. \tag{71}
$$

In order to evaluate the RHS of (70), we consider the reduction factor that the joint decoder exhibits, while considering the $(|\Lambda| - 1)|\mathcal{M}_1||\mathcal{M}_2|$ entries of the matrix C_d which correspond to all noncorrect cloud centers, as a result of the κ arrivals registered during a MAC phase. This reflects

$$
I(U, X_1, X_2; Y|V_1, V_2) = I(X_1, X_2; Y|V_1, V_2).
$$

Equation (65) implies that the encoders' incremental rate in response to a pair of arrivals, the first supporting just x_1 and the second supporting just x_2, is

$$
-2\left[\frac{(1-t)r_0(1-r_0)}{P_{01}}\ln[r_0(1-r_0)] + \frac{tr_1(1-r_1)}{P_{01}}\ln[r_1(1-r_1)]\right] \triangleq \delta R_{e,01}
$$

while as shown in Appendix I, the decoder's incremental rate in response to this pair of arrivals, while performing joint decoding of the triple $(w_1^{(b)}, w_2^{(b)}, U^{(b)})$ is

$$
-2[(1-t)\ln[r_0(1-r_0)] + t\ln[r_1(1-r_1)]] \triangleq \delta R_{d,01}.
$$

Similarly, the encoders' incremental rate per an arrival which supports both message waveforms is

$$
-2\left[\frac{(1-t)r_0^2}{P_{11}}\ln r_0 + \frac{tr_1^2}{P_{11}}\ln r_1\right] \triangleq \delta R_{e,11}
$$

while that of the decoder is

$$
-2[(1-t)\ln r_0 + t\ln r_1] \triangleq \delta R_{d,11}.
$$

Define the indicator functions

$$
\chi_{01} = \begin{cases} 1, & \delta R_{d,01} - \delta R_{e,01} > 0 \\ 0, & \text{otherwise} \end{cases}
$$

$$
\chi_{11} = \begin{cases} 1, & \delta R_{d,11} - \delta R_{e,11} > 0 \\ 0, & \text{otherwise}. \end{cases}
$$

It is obvious that we shall be interested in those cases where $\chi_{01} = 1$ and $\chi_{11} = 1$.

We can now express the aggregate rate communicated to the decoder per the duration T interval as a function of the information rate-sum exchanged between both encoders

$$
\begin{aligned}
R_d &= I(U, X_1, X_2; Y|V_1, V_2) \\
&= \frac{\kappa}{T}[\tilde{f}(t, \vec{r}) + \zeta_2(\delta R_{d,11} - \delta R_{e,11})\chi_{11} \\
&\quad + \zeta_1(\delta R_{d,01} - \delta R_{e,01})\chi_{01}].
\end{aligned}
$$

It follows that the decoder's list-size exponent per a generic transmission phase, upon subtraction of the rate it obtains via $U^{(b+1)}$, equals

$$
\begin{aligned}
\ln L_d &= T\left\{R_L^{(TW)} - [I(X_1, X_2; Y|V_1, V_2) - (R_1 + R_2)]\right\}_+ \\
&= \kappa \max\{0, \psi(\zeta_1, q) - \zeta_2(\delta R_{d,11} - \delta R_{e,11})\chi_{11} \\
&\quad - \zeta_1(\delta R_{d,01} - \delta R_{e,01})\chi_{01}\} \\
&\triangleq \kappa\tilde{\psi}(t, \vec{r}, q)
\end{aligned} \tag{72}
$$

where $\{z\}_+ = \max\{0, z\}$.

The quantity in (72) is the remaining ambiguity to be resolved via the total cooperation phase of generic transmission phase b—i.e., it is the expression for $A^{(TC)}$ as defined in (71) via the parameters of our coding scheme.

The total cooperation is performed via the single-user feedback decoding scheme of [19]. To this end, assume that during this phase the average power of each encoder is γ and $\tilde{\kappa}_2$ coding steps are executed. Thus, we have $L_d = \gamma^{-\tilde{\kappa}_2}$, implying that $\tilde{\kappa}_2 = \tilde{\psi}(t, \vec{r}, q)\kappa/(-\ln\gamma)$.

Define the random variable $\tilde{\xi}_\ell^{(K_2)}$ $\ell = 1, 2, \ldots, \tilde{\kappa}_2$ as the duration time for the completion of coding step ℓ of the total cooperation phase. Then the probability that this phase fails on an interval of length $\beta_2 T$ will be exponentially small once

$$\mathbb{E}\left(\sum_{\ell=1}^{\tilde{\kappa}_2} \tilde{\xi}_\ell^{(K_2)}\right) = \frac{\tilde{\kappa}_2}{2\gamma A} \leq \beta_2 T. \tag{73}$$

Combining (65)–(73), the following set of inequalities must be satisfied:

$$\frac{\kappa}{2\tilde{q}AT} = \frac{R}{2\tilde{q}\tilde{f}(t,\vec{r})A} \leq 1 - \beta_1 - \beta_2$$

$$\frac{\tilde{\kappa}_1}{2qAT} = \frac{(1-\zeta_1)h\left(\frac{\zeta_1}{1-\zeta_1}\right)R}{\tilde{f}(t,\vec{r})q\left[g(q)-(1+q)h\left(\frac{2q}{1+q}\right)\right]A} \leq \beta_1$$

$$\frac{\tilde{\kappa}_2}{2\gamma AT} = \frac{\tilde{\psi}(t,\vec{r},q)R}{2(-\gamma\ln\gamma)\tilde{f}(t,\vec{r})A} \leq \beta_2.$$

Setting all inequalities to hold with equality and taking $\gamma = e^{-1}$, we conclude that the attainable rate-sum for this scheme is

$$C_f^{(\text{sum})} = \max_{t,\vec{r},q} 2\tilde{f}(t,\vec{r})\left\{\frac{1}{\tilde{q}} + \frac{2(1-\zeta_1)h(\frac{\zeta_1}{1-\zeta_1})}{q[g(q)-(1+q)h(\frac{2q}{1+q})]} + e\tilde{\psi}(t,\vec{r},q)\right\}^{-1} A.$$

It can be verified numerically that for $t = 0.0045$, $r_0 = 0.269$, $r_1 = 10^{-32} = \epsilon \ll 1$, and $q = 0.2601$ one obtains $C_f^{(\text{sum})} = 0.553009A$, thus demonstrating an improvement on the Cover–Leung rate-sum. This rate-sum is attained with $\delta R_{d,01} - \delta R_{e,01} = 0.6485$, $\delta R_{d,11} - \delta R_{e,11} = 0.651327$, and $\zeta_1 = 0.3655$, $\zeta_2 = 0.269$ in which case $\tilde{\psi}(t,\vec{r},q) = 0$. Thus, cooperation between the encoders is maintained just via the random variable $U^{(b)}$ that furnishes the cloud center during the MAC phase. As a result, $\beta_2 = 0$ while $\beta_1 = 0.45521$—i.e., the two-way interaction between both encoders in order to exchange (A_{01}, A_{10}) consumes 45.5% of the communication time. Furthermore, each encoder transmits with average power $\sigma A = [(1-\beta_1)\tilde{q} + \beta_1 q]A = 0.2643A$. It is thus demonstrated that it suffices for the encoders to build correlation just via the random variable U as in Theorem 2.

Remark: It is not claimed that our particular choice of $P_{UX_1X_2}$ (i.e., $|\mathcal{U}| = 2$) is optimal. This choice just serves as an example to show the performance of this scheme as compared to the Cover–Leung inner bound.

APPENDIX I

Consider a joint distribution P_{U,X_1,X_2} where all random variables are binary. Let U be distributed according to the law $\Pr\{U = 1\} = t$ and let (X_1, X_2) be a pair of binary random variables that are conditionally independent given U and satisfy

$$P_{X_\mu|U}(x_\mu|u) = \begin{cases} (0|0), & 1 - r_0 \\ (0|1), & 1 - r_1, \end{cases} \mu = 1, 2.$$

The joint statistics of (X_1, X_2) for this code is

$$\Pr[X_1 = 0, X_2 = 1] = (1-t)r_0(1-r_0) + tr_1(1-r_1)$$
$$\triangleq P_{01}$$
$$\Pr[X_1 = 1, X_2 = 1] = (1-t)r_0^2 + tr_1^2 \triangleq P_{11}. \tag{74}$$

To analyze the reduction factor when Encoder 1 decodes $w_2^{(b)}$ consider first the $\zeta_1\kappa$ arrivals that fall where $(\boldsymbol{x}_1 = 0, \boldsymbol{x}_2 = 1)$. This group is split into two subsets

$$\Pr[U = 0|X_1 = 0, X_2 = 1]\zeta_1\kappa$$

arrivals which fall where $U = 0$, in which case the reduction factor for Encoder 1 equals r_0, while the rest

$$\Pr[U = 1|X_1 = 0, X_2 = 1]\zeta_1\kappa$$

arrivals fall where $U = 1$, in which case the reduction factor for Encoder 1 equals r_1. In summary

$$\mathbb{E}\left[\prod_{\ell=1}^{\zeta_1\kappa} q_\ell\right] = \left[r_0^{\Pr[U=0|X_1=0,X_2=1]}r_1^{\Pr[U=1|X_1=0,X_2=1]}\right]^{\zeta_1\kappa}$$
$$\triangleq \eta_{e,01}^{\zeta_1\kappa}$$

$$\Pr[U = 0|X_1 = 0, X_2 = 1] = \frac{(1-t)r_0(1-r_0)}{P_{01}}$$

$$\Pr[U = 1|X_1 = 0, X_2 = 1] = \frac{tr_1(1-r_1)}{P_{01}}. \tag{75}$$

Consider next the $\zeta_1\kappa$ arrivals for which $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 0)$. By similar arguments it follows that the reduction factor for Encoder 1 is

$$\mathbb{E}\left[\prod_{\ell=1}^{\zeta_1\kappa} q_\ell\right] = \left[(1-r_0)^{\Pr[U=0|X_1=1,X_2=0]}\right.$$
$$\left.\cdot (1-r_1)^{\Pr[U=1|X_1=1,X_2=0]}\right]^{\zeta_1\kappa} \triangleq \eta_{e,10}^{\zeta_1\kappa}$$
$$\Pr[U = 0|X_1 = 1, X_2 = 0] = \Pr[U = 0|X_1 = 0, X_2 = 1]. \tag{76}$$

Finally, on the the $\zeta_2\kappa$ arrivals for which $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 1)$, the reduction factor for *both* encoders is

$$\mathbb{E}\left[\prod_{\ell=1}^{\zeta_2\kappa} q_\ell\right] = \left[r_0^{\Pr[U=0|X_1=1,X_2=1]}r_1^{\Pr[U=1|X_1=1,X_2=1]}\right]^{\zeta_2\kappa}$$
$$\triangleq \eta_{e,11}^{\zeta_2\kappa}$$

$$\Pr[U = 0|X_1 = 1, X_2 = 1] = \frac{(1-t)r_0^2}{P_{11}}$$

$$\Pr[U = 1|X_1 = 1, X_2 = 1] = \frac{tr_1^2}{P_{11}}. \tag{77}$$

To analyze the decoder's performance while decoding the triple $(\lambda^{(b-1)}(w_1^{(b)}, w_2^{(b)}))$ note that any random variable U' that

is generated independently of U under the same marginal P_U satisfies

$$\Pr[U' = 0 | X_1 = 0, X_2 = 1] = (1 - t)$$
$$\Pr[U' = 1 | X_1 = 0, X_2 = 1] = t$$
$$\Pr[X_1' = 0, X_2' = 1 | U' = 0] = r_0(1 - r_0)$$
$$\Pr[X_1' = 0, X_2' = 1 | U' = 1] = r_1(1 - r_1).$$

Consequently, fixing a pair of messages $(w_1^{(b)}, w_2^{(b)})$ as the decoder considers the possible codeword pairs in (59) for say n_1 arrivals that satisfy $(\boldsymbol{x}_1 = 0, \boldsymbol{x}_2 = 1)$ one obtains

$$\mathbb{E}\left[\prod_{\ell=1}^{n_1} q_\ell\right] = \left\{[r_0(1 - r_0)]^{(1-t)}[r_1(1 - r_1)]^t\right\}^{n_1}$$

and the same reduction is obtained for any n_1 arrivals that satisfy $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 0)$. On the other hand, if the decoder decodes the cloud center based on say n_2 arrivals that satisfy $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 1)$ one obtains

$$\mathbb{E}\left[\prod_{\ell=1}^{n_2} q_\ell\right] = \left[r_0^{2(1-t)}r_1^{2t}\right]^{n_2}.$$

In summary, suppose the decoder combines n_1 arrivals for which $(\boldsymbol{x}_1 = 0, \boldsymbol{x}_2 = 1)$ with another n_1 arrivals for which $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 0)$ plus additional n_2 arrivals for which $(\boldsymbol{x}_1 = 1, \boldsymbol{x}_2 = 1)$ and performs joint decoding, his reduction factor due to these $2n_1 + n_2$ arrivals equals

$$\mathbb{E}\left[\prod_{\ell=1}^{2n_1+n_2} q_\ell\right] = \left\{[r_0(1 - r_0)]^{1-t}[r_1(1 - r_1)^t\right\}^{2n_1}\left[r_0^{1-t}r_1^t\right]^{2n_2}$$
$$\triangleq \mathcal{Q}_d(n_1, n_2).$$

At the same time, the aggregate reduction factor obtained by both encoders on the same $2n_1 + n_2$ arrivals equals

$$\mathbb{E}\left[\prod_{\ell=1}^{2n_1+n_2} q_\ell\right]$$
$$= \left\{[r_0(1 - r_0)]^{\frac{(1-t)r_0(1-r_0)}{P_{01}}}[r_1(1 - r_1)^{\frac{tr_1(1-r_1)}{P_{01}}}\right\}^{2n_1}$$
$$\cdot \left[r_0^{\frac{(1-t)r_0^2}{P_{11}}}r_1^{\frac{tr_1^2}{P_{11}}}\right]^{2n_2} \triangleq \mathcal{Q}_e(n_1, n_2).$$

In particular, we shall be interested in those cases $(t, \vec{r})$ where

$$\mathcal{Q}_d(n_1, n_2) < \mathcal{Q}_e(n_1, n_2).$$

ACKNOWLEDGMENT

REFERENCES

[1] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 100–102, Jan. 1975.

[2] T. M. Cover and S. K. Leung, "An achievable rate region for the multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 3, pp. 292–298, May 1981.

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

[4] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple-access channels," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 93–95, Jan. 1982.

[5] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, Jan. 1989.

[6] F. M. J. Willems and E. C. Van Der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[7] L. H. Ozarow, "The capacity of the white Gaussian multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–629, Jul. 1984.

[8] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[9] ——, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Inst. Technol., Zurich, Switzerland, 1998.

[10] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 197–207, Mar. 1973.

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

[12] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[13] R. Ahlswede and N. Cai, "Seminoisy deterministic multiple-access channels: Coding theorems for list codes and codes with feedback," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2153–2162, Aug. 2002.

[14] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[15] M. Salehi, "Cardinality Bounds on Auxiliary Variables in Multi-User Theory via the Method of Ahlswede and Körner," Dep. Statist., Stanford Univ., Stanford, CA, Tech. Rep. 33, 1978.

[16] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.

[17] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Parts I and II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1471, Nov. 1988.

[18] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 244–256, Mar. 1991.

[19] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 491–503, Mar. 1993.

[20] A. Lapidoth and S. Shamai (Shitz), "The Poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 488–502, Mar. 1998.

[21] S. I. Bross, M. V. Burnashev, and S. Shamai (Shitz), "Error exponents for the two-user Poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1999–2016, Jul. 2001.