

Sum-Rate Capacity for Symmetric Gaussian Multiple Access Channels with Feedback

Erixhen Sula, *Student Member, IEEE*, Michael Gastpar, *Fellow, IEEE*, Gerhard Kramer, *Fellow, IEEE*

Abstract

The feedback sum-rate capacity is established for the symmetric J -user Gaussian multiple-access channel (GMAC). The main contribution is a converse bound that combines the dependence-balance argument of Hekstra and Willems (1989) with a variant of the factorization of a convex envelope of Geng and Nair (2014). The converse bound matches the achievable sum-rate of the Fourier-Modulated Estimate Correction strategy of Kramer (2002).

I. INTRODUCTION

THE feedback capacity of the *two-user* Gaussian multiple-access channel (GMAC) was established by Ozarow [1]. The coding theorem was based on extending feedback strategies of Elias [2] and Schalkwijk and Kailath [3] (see [4]), while the converse followed from a cut-set argument. Kramer [5] extended Ozarow's scheme to more than two users by using a method he called Fourier-Modulated Estimate Correction, or Fourier-MEC. For the symmetric GMAC and sufficiently large signal-to-noise ratio (SNR), the Fourier-MEC sum-rate meets the cut-set bound and is thus optimal. However, the problem remains open for low and intermediate SNRs.

The coding schemes in [1], [2], [3], [5] start by mapping the message onto a point on the real line or complex plane, and they iteratively correct the receiver's estimate of this point by using linear minimum mean squared error (LMMSE) estimation. There are many variants of the schemes. For example, one can convert complex-channel strategies to real-channel strategies [5], one can interpret the LMMSE step as posterior matching [6], [7], and one can use multi-dimensional Fourier transforms for Fourier-MEC, e.g., a Hadamard transform [5], [7].

Progress on improving capacity outer bounds was made in [8] by applying the Hekstra-Willems dependence-balance argument [9], and in [10] that studies linear feedback strategies. Other results regarding GMACs with imperfect and/or noisy feedback are presented in [11], [12], [13]. Inner and outer bounds for other classes of multiple-access channels (MACs) with feedback are derived in [14], [15], [16].

A. Contribution

We derive a converse bound that establishes the following capacity result. Achievability follows from [5, Sec. V].

Theorem 1: The feedback sum-rate capacity of the J -user symmetric GMAC is

$$C_{\text{sum}} = \frac{1}{2} \log_2(1 + PJ\beta) \quad \text{bits/channel use} \quad (1)$$

where β is the unique solution satisfying $\beta \in [1, J]$ and

$$(1 + PJ\beta)^{J-1} = (1 + P\beta(J - \beta))^J. \quad (2)$$

The proof of Theorem 1 combines the Lagrange-duality approach of [17] with a variant of the factorization of a convex envelope used in [18], [19], and that was inspired by work on functional inequalities [20], [21]. One difference to [18] is that in our problem the matrix that describes the channel is rank-deficit. Besides establishing the optimality of Gaussian signaling, we transform a non-convex Lagrangian dual problem into a convex problem.

B. Notation

We use the following notation. Random variables are denoted by uppercase letters and their realizations by lowercase letters. Random column vectors are denoted by boldface uppercase letters and their realizations by boldface lowercase letters. The i -th entry of the column vector \mathbf{X} is denoted by X_i or $[\mathbf{X}]_i$. We denote matrices with uppercase letters, e.g., A, B, C . The (i, j) element of matrix A is denoted by A_{ij} or $[A]_{ij}$. For the cross-covariance matrix of two random vectors \mathbf{X} and \mathbf{Y} , we use the shorthand notation $K_{\mathbf{X}\mathbf{Y}}$, and for the covariance matrix of a random vector \mathbf{X} we use the shorthand notation $K_{\mathbf{X}} := K_{\mathbf{X}\mathbf{X}}$. Calligraphic letters denote sets, e.g., $\mathcal{A}, \mathcal{B}, \mathcal{C}$. The expression $X \sim \mathcal{N}(m, \sigma^2)$ denotes a Gaussian random variable with mean m and variance σ^2 . We denote the convergence in distribution (weak convergence) by \xrightarrow{w} .

Erixhen Sula is with École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, e-mail: erixhen.sula@epfl.ch.

Michael Gastpar is with École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, e-mail: michael.gastpar@epfl.ch

Gerhard Kramer is with Technical University of Munich (TUM), Germany, e-mail: gerhard.kramer@tum.de

C. Organization

This paper is organized as follows. In Section II, we present the system model and review existing capacity bounds. In Section III, we give an upper bound on the sum rate for general GMACs with feedback. In Section IV, we prove Theorem 1. Section V concludes the paper and the appendices provide supporting results and proofs.

II. SYSTEM MODEL AND RELATED WORK

Consider a GMAC with J transmitters (called users) with channel input symbols X_1, X_2, \dots, X_J , and a receiver with the channel output symbol Y . The received signal at time instant i is

$$Y_i = Z_i + \sum_{j=1}^J g_j X_{j,i} \quad (3)$$

for $i = 1, 2, \dots, n$, where Z_1, Z_2, \dots, Z_n is a string of independent and identically distributed (i.i.d.) zero-mean Gaussian noise variables with unit variance and g_1, g_2, \dots, g_J are channel gains. The J channel inputs have the block power constraints

$$\sum_{i=1}^n E \left[X_{j,i}^2 \right] \leq nP_j, \quad j = 1, 2, \dots, J. \quad (4)$$

The SNR of user j is thus $P_j g_j^2$. If $P_1 g_1^2 = P_2 g_2^2 = \dots = P_J g_J^2$, then the transmitters can be swapped without changing the problem. For such models, we may as well set $P_j = P$ and $g_j = 1$ for all j , and we refer to this channel as the *symmetric GMAC*.

Let W_j with nR_j bits be the message of user j . The transmitted signal at time instant i is

$$X_{j,i} = f_{j,i}(W_j, Y^{i-1}), \quad j = 1, 2, \dots, J \quad (5)$$

where the $f_{j,i}(\cdot)$ are encoding functions to be optimized. The receiver puts out the estimates

$$\left(\hat{W}_1, \hat{W}_2, \dots, \hat{W}_J \right) = g(Y^n) \quad (6)$$

where $g(\cdot)$ is a decoding function. The event that the receiver makes an error is

$$\mathcal{E} = \bigcup_{j=1}^J \{ \hat{W}_j \neq W_j \}. \quad (7)$$

The rate-tuple $\mathbf{R} = (R_1, R_2, \dots, R_J)$ is said to be *achievable* if, for any specified positive error probability P_e and sufficiently large n , there are encoding functions and a decoder such that $\Pr[\mathcal{E}] \leq P_e$. The closure of the set of achievable \mathbf{R} is called the *capacity region* $C_{\text{MAC-FB}}$. We are interested in characterizing the *sum-rate capacity* C_{sum} , i.e., the maximum sum of the entries of any \mathbf{R} in $C_{\text{MAC-FB}}$.

A. Capacity without Feedback

The capacity region C_{MAC} *without* feedback is the set of $\mathbf{R} = (R_1, R_2, \dots, R_J)$ satisfying (see [22, Section 14.3])

$$R_S \leq \frac{1}{2} \log(1 + P_S) \quad (8)$$

for all $S \subseteq \mathcal{J} = \{1, 2, \dots, J\}$, where $R_S = \sum_{j \in S} R_j$ and $P_S = \sum_{j \in S} P_j$. The sum-rate capacity without feedback is therefore

$$\frac{1}{2} \log(1 + P_{\mathcal{J}}). \quad (9)$$

B. Two-User Capacity with Feedback

Feedback allows the users cooperate to increase rates. For $J = 2$ the capacity region is known to be [1]

$$C_{\text{MAC-FB}} = \bigcup_{0 \leq \rho \leq 1} \mathcal{R}(\rho) \quad (10)$$

where $\mathcal{R}(\rho)$ is the set of rate pairs (R_1, R_2) that satisfy

$$\begin{aligned} 0 &\leq R_1 \leq \frac{1}{2} \log \left(1 + P_1(1 - \rho^2) \right) \\ 0 &\leq R_2 \leq \frac{1}{2} \log \left(1 + P_2(1 - \rho^2) \right) \\ R_1 + R_2 &\leq \frac{1}{2} \log \left(1 + P_1 + P_2 + 2\rho\sqrt{P_1 P_2} \right). \end{aligned} \quad (11)$$

The parameter ρ is the correlation coefficient of X_1 and X_2 , and the optimal X_1 and X_2 are zero-mean Gaussian with second moments P_1 and P_2 , respectively. $C_{\text{MAC-FB}}$ is here the same as a standard cut-set bound. However, we show that cut-set bounds are loose for $J > 2$, and that dependence balance bounds can characterize the fundamental limits of communication.

C. Two-User Dependence Balance Bounds

Dependence balance bounds were introduced by Hekstra and Willems [9] for single output two-way channels. The tool generalizes to other models such as MACs with feedback. For example, for the two-user MAC with feedback, the achievable (R_1, R_2) must satisfy

$$\begin{aligned} 0 &\leq R_1 \leq I(X_1; Y|X_2, T) \\ 0 &\leq R_2 \leq I(X_2; Y|X_1, T) \\ R_1 + R_2 &\leq I(X_1, X_2; Y|T) \end{aligned} \quad (12)$$

for some $p(t, x_1, x_2, y)$ for which

$$T \leftrightarrow [X_1, X_2] \leftrightarrow Y \text{ forms a Markov chain} \quad (13)$$

$$I(X_1; X_2|T) \leq I(X_1; X_2|Y, T). \quad (14)$$

In [9, Section 7], the term $I(X_1; X_2|T)$ is interpreted as the amount of dependence *consumed*, and $I(X_1; X_2|Y, T)$ as the amount of dependence *produced* by communication. An interpretation of the inequality (14) is thus that dependence consumed cannot exceed the dependence produced, i.e., communication is limited by *dependence balance*.

Other interpretations of this bound are described in [8]. Observe that (14) can be rewritten in the following two ways:

$$I(X_1; Y|T) + I(X_2; Y|T) \leq I(X_1, X_2; Y|T) \quad (15)$$

$$I(X_1, X_2; Y|T) \leq I(X_1; Y|X_2, T) + I(X_2; Y|X_1, T). \quad (16)$$

The bound (16) requires the set function $f : 2^{\{1,2\}} \rightarrow \mathbb{R}$ defined by

$$f(\{1\}) = I(X_1; Y|X_2, T) \quad (17)$$

$$f(\{2\}) = I(X_2; Y|X_1, T) \quad (18)$$

$$f(\{1, 2\}) = I(X_1, X_2; Y|T) \quad (19)$$

to be *submodular*. In other words, for valid choices of $p(t, x_1, x_2, y)$, the rate region defined by (12) is a *polymatroid*. We may thus interpret dependence balance as a submodularity (or polymatroid) constraint, i.e., communication is limited by submodularity.

We remark that for two-user GMACs the dependence balance bound yields the same rate region as the standard cut-set bound. However, the dependence balance bound is more informative in the following sense. Consider jointly Gaussian $p(t, x_1, x_2, y)$. The optimal correlation coefficient ρ^* in (11) is the one that satisfies (14)-(16) with equality. However, dependence balance (or submodularity) limits ρ to the range $[0, \rho^*]$, whereas the cut-set bound permits all ρ in $[0, 1]$.

D. Multi-User Dependence Balance Bounds

The two-user dependence balance concept was generalized to J users in [15, Thm. 4] and more dependence balance bounds are derived in [8, Thm. 1]. The capacity region of the J -user MAC with feedback is a subset of the set of rate-tuples (R_1, R_2, \dots, R_J) satisfying

$$R_S \leq I(X_S; Y|X_{S^c}, T) \quad (20)$$

for all $S \subseteq \mathcal{J}$, where S^c is the complement of S , and where

$$T \leftrightarrow [X_1, X_2, \dots, X_J] \leftrightarrow Y \text{ forms a Markov chain} \quad (21)$$

$$I(X_1, X_2, \dots, X_J; Y|T) \leq \frac{1}{M-1} \sum_{m=1}^M I(X_{S_m^c}; Y|X_{S_m}, T), \quad (22)$$

for any partition $\{S_m\}_{m=1}^M$ of \mathcal{J} into $M \geq 2$ subsets. One may again interpret (22) as a submodular constraint. For example, for the partition $S_1 = \{1\}$, $S_2 = \{2\}$, \dots , $S_J = \{J\}$ the dependence balance constraint (22) becomes

$$I(X_1, X_2, \dots, X_J; Y|T) \leq \frac{1}{J-1} \sum_{j=1}^J I(X_{\mathcal{J} \setminus \{j\}}; Y|X_j, T) \quad (23)$$

where $\mathcal{J} \setminus \{j\}$ is the set $\{1, 2, \dots, j-1, j+1, \dots, J\}$. As usual, one can add the power constraints $E[X_j^2] \leq P_j$, $j \in \mathcal{J}$, to these bounds. Also, as for (15) for the two-user case, the bound (22) can be written as

$$\sum_{m=1}^M I(X_{S_m}; Y|T) \leq I(X_1, X_2, \dots, X_J; Y|T). \quad (24)$$

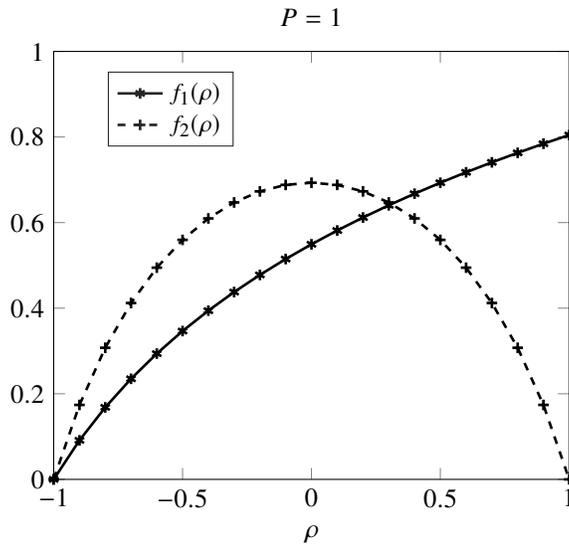


Fig. 1. Cut-set bounds for the sum-rate of a two-user symmetric GMAC with feedback.

E. Cut-set Bound

The cut-set bounds give the following result, see [1] and [22, Theorem 15.10.1].

Proposition 1: For the two-user symmetric GMAC with feedback, we have

$$C_{\text{sum}} \leq \max_{-1 \leq \rho \leq 1} \min \left\{ \underbrace{\frac{1}{2} \log(1 + 2P(1 + \rho))}_{f_1(\rho)}, \underbrace{\log(1 + P(1 - \rho^2))}_{f_2(\rho)} \right\}. \quad (25)$$

The sum-rate on the RHS of (25) turns out to be achievable [1], and it is depicted in Figure 1. Similarly, again starting from [22, Theorem 15.10.1], we obtain the following result.

Proposition 2: For the three-user symmetric GMAC with feedback, we have

$$C_{\text{sum}} \leq \max_{-1/2 \leq \rho \leq 1} \min \left\{ \underbrace{\frac{1}{2} \log(1 + 3P(1 + 2\rho))}_{g_1(\rho)}, \underbrace{\frac{3}{4} \log(1 + 2P(1 - \rho)(1 + 2\rho))}_{g_2(\rho)}, \underbrace{\frac{3}{2} \log \left(1 + \frac{P(1 + 2\rho)(1 - \rho)}{1 + \rho} \right)}_{g_3(\rho)} \right\}. \quad (26)$$

The sum rate (26) is not generally achievable. Figure 2 illustrates the situation for the special case $P = 0.3$. The cut-set bound of Proposition 2 leads to an upper bound on the sum rate given by the intersection point of the curves g_2 and g_3 . However, we show that the capacity is given by the intersection of the curves g_1 and g_2 ; this intersection point is achieved by Fourier-MEC [5].

For the symmetric GMAC and large SNR, i.e., more than a certain threshold, the Fourier-MEC sum-rate meets the cut-set bound. For example, a sufficient condition for the cut-set bound to give the sum-rate capacity is that the SNR is greater than or equal to $2^{J+1}/J^2$ [5]. Observe that this threshold grows almost exponentially with the number of users.

III. GENERAL CONVERSE BOUND FOR J USERS

We derive the following upper bounds on the feedback sum-rate capacity of the general J -user GMAC.

Theorem 2: For any $\lambda \geq 0$ and any partition $\{\mathcal{S}_m\}_{m=1}^M$ of \mathcal{J} into $M \geq 2$ subsets, we have

$$C_{\text{sum}} \leq \max_{p(x) \in G_{\mathcal{G}}} (1 - \lambda) I(X_1, X_2, \dots, X_J; Y) + \frac{\lambda}{M-1} \sum_{m=1}^M I(X_{\mathcal{S}_m^c}; Y | X_{\mathcal{S}_m}) \quad (27)$$

where $G_{\mathcal{G}}$ is the set of zero-mean Gaussian distributions satisfying $E[X_j^2] \leq P_j$ for $j = 1, 2, \dots, J$.

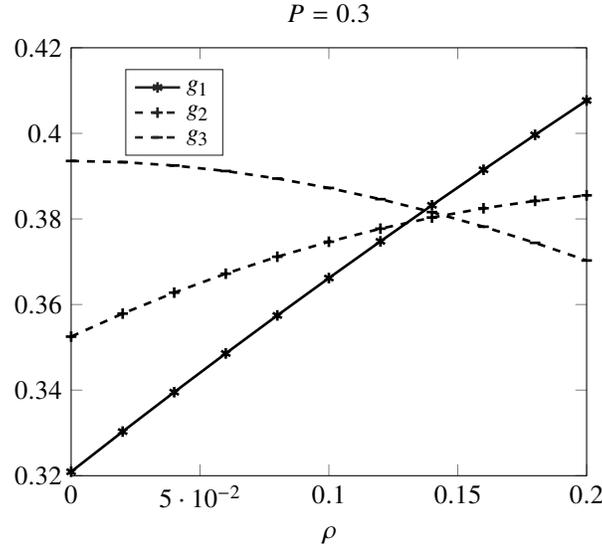


Fig. 2. Cut-set bounds for the sum-rate of a three-user symmetric GMAC with feedback.

A. Proof of Theorem 2

Our converse bound starts from Section II-D and we use the shorthand $\mathbf{X} = (X_1, X_2, \dots, X_J)$. We find it convenient to express our problem as a minimization, i.e., we seek to minimize $-I(\mathbf{X}; Y|T)$ over the input distributions $p(t, \mathbf{x})$ that satisfy the dependence-balance constraint

$$I(\mathbf{X}; Y|T) \leq \frac{1}{M-1} \sum_{m=1}^M I(X_{S_m^c}; Y|X_{S_m}, T) \quad (28)$$

for (T, \mathbf{X}) such that $T \leftrightarrow \mathbf{X} \leftrightarrow Y$ forms a Markov chain. We will treat the power constraints in two steps. First, for any fixed covariance matrix K , we will optimize over all distributions satisfying $E[\mathbf{X}\mathbf{X}^T] = K$. Then, we will optimize over all K whose diagonal entries are at most P .

Given the covariance matrix K , we form the Lagrangian for our optimization problem as

$$s_\lambda(\mathbf{X}|T) := (\lambda - 1)I(\mathbf{X}; Y|T) - \frac{\lambda}{M-1} \sum_{m=1}^M I(X_{S_m^c}; Y|X_{S_m}, T). \quad (29)$$

This can be rewritten as

$$s_\lambda(\mathbf{X}|T) = - \left(\frac{\lambda}{M-1} + 1 \right) I(\mathbf{X}; Y|T) + \frac{\lambda}{M-1} \sum_{m=1}^M I(X_{S_m}; Y|T). \quad (30)$$

We define

$$S_\lambda(\mathbf{X}) = \inf_{\substack{p(t|\mathbf{x}): \\ T \leftrightarrow \mathbf{X} \leftrightarrow Y}} \{s_\lambda(\mathbf{X}|T)\}$$

and note that $S_\lambda(\mathbf{X})$ is a convex function of $p(\mathbf{x})$ because $S_\lambda(\mathbf{X})$ is the lower convex envelope of $s_\lambda(\mathbf{X})$, which is defined by dropping the random variable T in (29). In addition, we define

$$S_\lambda(\mathbf{X}|T) = \sum_t p(t) S_\lambda(\mathbf{X}|T=t). \quad (31)$$

The dual function of our problem for $K \geq 0$ is

$$V_\lambda(K) := \inf_{p(\mathbf{x}): E[\mathbf{X}\mathbf{X}^T]=K} \{S_\lambda(\mathbf{X})\}. \quad (32)$$

Alternatively, we have

$$V_\lambda(K) = \inf_{\substack{p(t, \mathbf{x}): E[\mathbf{X}\mathbf{X}^T]=K \\ T \leftrightarrow \mathbf{X} \leftrightarrow Y}} \{s_\lambda(\mathbf{X}|T)\}. \quad (33)$$

By the standard Lagrangian duality we bound the original optimization problem as follows

$$C_{\text{sum}} \leq -\max_{\lambda} \min_{\substack{p(t, \mathbf{x}): E[X_j^2] \leq P_j \\ T \leftrightarrow \mathbf{X} \leftrightarrow Y}} s_{\lambda}(\mathbf{X}|T) \quad (34)$$

$$= -\max_{\lambda} \min_{\substack{K \geq 0: \\ [K]_{jj} \leq P_j}} \underbrace{\min_{\substack{p(t, \mathbf{x}): E[\mathbf{X}\mathbf{X}^T] = K \\ T \leftrightarrow \mathbf{X} \leftrightarrow Y}} s_{\lambda}(\mathbf{X}|T)}_{V_{\lambda}(K)}. \quad (35)$$

From now on we mainly deal with the dual function $V_{\lambda}(K)$. For $0 < \lambda \leq 1$, the minimization problem (33) is a convex problem, and it follows from (29) and maximum entropy results that the optimizing distribution $p(t, x_1, x_2, x_3)$ is jointly Gaussian. The more difficult case is $\lambda > 1$. Our approach for all cases will be to establish that the distribution attaining the minimum in (33) must make the *channel input* Gaussian. This follows from a novel variant of the factorization of convex envelopes.

Consider two independent uses of the GMAC:

$$\begin{aligned} Y_1 &= G\mathbf{X}_1 + Z_1 \\ Y_2 &= G\mathbf{X}_2 + Z_2 \end{aligned} \quad (36)$$

where $G = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$, \mathbf{X}_1 and \mathbf{X}_2 are independent and identically distributed, and where $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ are independent. One key difference to [18] is that G is *not* an invertible matrix. We define

$$\mathbf{X}_{\theta_1} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 + \mathbf{X}_2), \quad \mathbf{X}_{\theta_2} = \frac{1}{\sqrt{2}}(\mathbf{X}_1 - \mathbf{X}_2), \quad (37)$$

$$Y_{\theta_1} = \frac{1}{\sqrt{2}}(Y_1 + Y_2), \quad Y_{\theta_2} = \frac{1}{\sqrt{2}}(Y_1 - Y_2). \quad (38)$$

We thus have

$$Y_{\theta_1} = G\mathbf{X}_{\theta_1} + \tilde{Z}_1, \quad Y_{\theta_2} = G\mathbf{X}_{\theta_2} + \tilde{Z}_2 \quad (39)$$

where $\tilde{Z}_1, \tilde{Z}_2 \sim \mathcal{N}(0, 1)$ are independent. Moreover, we generalize the definition (30) to the two-letter extension as

$$s_{\lambda}(\mathbf{X}_1, \mathbf{X}_2|T) := -\left(\frac{\lambda}{M-1} + 1\right)I(\mathbf{X}_1, \mathbf{X}_2; Y_1, Y_2|T) + \frac{\lambda}{M-1} \sum_{m=1}^M I([\mathbf{X}_1]_{S_m}, [\mathbf{X}_2]_{S_m}; Y_1, Y_2|T). \quad (40)$$

Next, we state and prove four key propositions.

Proposition 3: $I(\mathbf{X}_1, \mathbf{X}_2; Y_1, Y_2) = I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_1}, Y_{\theta_2})$.

Proof: The function $f(x, y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right)$ is bijective. □

Proposition 4: The chain $Y_{\theta_1} \leftrightarrow \mathbf{X}_{\theta_1} \leftrightarrow \mathbf{X}_{\theta_2} \leftrightarrow Y_{\theta_2}$ is Markov and we have

$$I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_1}, Y_{\theta_2}) = I(\mathbf{X}_{\theta_1}; Y_{\theta_1}) + I(\mathbf{X}_{\theta_2}; Y_{\theta_2} | Y_{\theta_1}). \quad (41)$$

Proof: The Markovity follows by (39). We further compute

$$\begin{aligned} I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_1}, Y_{\theta_2}) &= I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_1}) + I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_2} | Y_{\theta_1}) \\ &= I(\mathbf{X}_{\theta_1}; Y_{\theta_1}) + I(\mathbf{X}_{\theta_2}; Y_{\theta_1} | \mathbf{X}_{\theta_1}) + I(\mathbf{X}_{\theta_2}; Y_{\theta_2} | Y_{\theta_1}) + I(\mathbf{X}_{\theta_1}; Y_{\theta_2} | Y_{\theta_1}, \mathbf{X}_{\theta_2}) \\ &\stackrel{(a)}{=} I(\mathbf{X}_{\theta_1}; Y_{\theta_1}) + I(\mathbf{X}_{\theta_2}; Y_{\theta_2} | Y_{\theta_1}) \end{aligned} \quad (42)$$

where (a) is consequence of the Markov chain. □

Proposition 5: For any $\lambda \geq 0$ we have

$$s_{\lambda}(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}|T) \geq s_{\lambda}(\mathbf{X}_{\theta_1}|T) + s_{\lambda}(\mathbf{X}_{\theta_2}|Y_{\theta_1}, T) \quad (43)$$

with equality if and only we have

- $I(Y_{\theta_1}; [\mathbf{X}_{\theta_2}]_{S_m} | [\mathbf{X}_{\theta_1}]_{S_m}, T) = 0$
- $I(Y_{\theta_2}; [\mathbf{X}_{\theta_1}]_{S_m} | Y_{\theta_1}, [\mathbf{X}_{\theta_2}]_{S_m}, T) = 0$

for $m = 1, \dots, M$.

Proof: We compute

$$\begin{aligned}
 & s_\lambda(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2} | T) - s_\lambda(\mathbf{X}_{\theta_1} | T) - s_\lambda(\mathbf{X}_{\theta_2} | Y_{\theta_1}, T) \\
 &= \left(\frac{\lambda}{M-1} + 1 \right) (-I(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2}; Y_{\theta_1}, Y_{\theta_2} | T) + I(\mathbf{X}_{\theta_1}; Y_{\theta_1} | T) + I(\mathbf{X}_{\theta_2}; Y_{\theta_2} | Y_{\theta_1}, T)) \\
 &+ \frac{\lambda}{M-1} \left(\sum_{m=1}^M I([\mathbf{X}_{\theta_1}]_{S_m}, [\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_1}, Y_{\theta_2} | T) - I([\mathbf{X}_{\theta_1}]_{S_m}; Y_{\theta_1} | T) - I([\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_2} | Y_{\theta_1}, T) \right) \\
 &\stackrel{(a)}{=} \frac{\lambda}{M-1} \sum_{m=1}^M (I([\mathbf{X}_{\theta_1}]_{S_m}, [\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_1} | T) + I([\mathbf{X}_{\theta_1}]_{S_m}, [\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_2} | Y_{\theta_1}, T) - I([\mathbf{X}_{\theta_1}]_{S_m}; Y_{\theta_1} | T) - I([\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_2} | Y_{\theta_1}, T)) \\
 &= \frac{\lambda}{M-1} \sum_{m=1}^M (I([\mathbf{X}_{\theta_2}]_{S_m}; Y_{\theta_1} | [\mathbf{X}_{\theta_1}]_{S_m}, T) + I([\mathbf{X}_{\theta_1}]_{S_m}; Y_{\theta_2} | Y_{\theta_1}, [\mathbf{X}_{\theta_2}]_{S_m}, T)) \geq 0
 \end{aligned}$$

where (a) follows from Proposition 4. The last step follows from the non-negativity of conditional mutual information terms, and we have equality if and only if all the terms are zero. \square

Proposition 6: There is a pair of random variables (T_*, \mathbf{X}_*) with $|\mathcal{T}_*| \leq \frac{J(J+1)}{2} + 1$ and $E[\mathbf{X}_* \mathbf{X}_*^T] = K$ such that

$$V_\lambda(K) = s_\lambda(\mathbf{X}_* | T_*). \quad (44)$$

Proof: The existence of a maximizer and the cardinality bound on T_* are established in the Appendix A by using a similar argument as in [18, Appendix 2A]. \square

We can now establish the desired result.

Lemma 1: Let $p_*(t, \mathbf{x})$ attain $V_\lambda(K)$ and let $(T_1, T_2, \mathbf{X}_1, \mathbf{X}_2) \sim p_*(t_1, \mathbf{x}_1)p_*(t_2, \mathbf{x}_2)$. Let \mathbf{X}_t denote the conditional distribution $p_*(\mathbf{x} | T = t)$ and define

$$\begin{aligned}
 \mathbf{X}_{\theta_1} | ((T_1, T_2) = (t_1, t_2)) &\sim \frac{1}{\sqrt{2}}(\mathbf{X}_{t_1} + \mathbf{X}_{t_2}), & Y_{\theta_1} | ((T_1, T_2) = (t_1, t_2)) &\sim \frac{1}{\sqrt{2}}(Y_{t_1} + Y_{t_2}), \\
 \mathbf{X}_{\theta_2} | ((T_1, T_2) = (t_1, t_2)) &\sim \frac{1}{\sqrt{2}}(\mathbf{X}_{t_1} - \mathbf{X}_{t_2}), & Y_{\theta_2} | ((T_1, T_2) = (t_1, t_2)) &\sim \frac{1}{\sqrt{2}}(Y_{t_1} - Y_{t_2}).
 \end{aligned}$$

Then:

- 1) $(T, \mathbf{X}_{\theta_1})$ also attains $V_\lambda(K)$.
- 2) $(T, \mathbf{X}_{\theta_2})$ also attains $V_\lambda(K)$.
- 3) The joint distribution $(T, \mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2})$ must satisfy

- $I(Y_{\theta_1}; [\mathbf{X}_{\theta_2}]_{S_m} | [\mathbf{X}_{\theta_1}]_{S_m}, T) = 0$
- $I(Y_{\theta_2}; [\mathbf{X}_{\theta_1}]_{S_m} | Y_{\theta_1}, [\mathbf{X}_{\theta_2}]_{S_m}, T) = 0$

for $m = 1, \dots, M$.

Proof: Consider the steps:

$$\begin{aligned}
 2V_\lambda(K) &\stackrel{(a)}{=} s_\lambda(\mathbf{X}_1 | T_1) + s_\lambda(\mathbf{X}_2 | T_2) \\
 &\stackrel{(b)}{=} s_\lambda(\mathbf{X}_1, \mathbf{X}_2 | T_1, T_2) \\
 &\stackrel{(c)}{=} s_\lambda(\mathbf{X}_{\theta_1}, \mathbf{X}_{\theta_2} | T_1, T_2) \\
 &\stackrel{(d)}{\geq} s_\lambda(\mathbf{X}_{\theta_1} | T_1, T_2) + s_\lambda(\mathbf{X}_{\theta_2} | Y_{\theta_1}, T_1, T_2) \\
 &\stackrel{(e)}{\geq} S_\lambda(\mathbf{X}_{\theta_1}) + S_\lambda(\mathbf{X}_{\theta_2} | Y_{\theta_1}) \\
 &\stackrel{(f)}{\geq} S_\lambda(\mathbf{X}_{\theta_1}) + S_\lambda(\mathbf{X}_{\theta_2}) \\
 &\stackrel{(g)}{\geq} 2V_\lambda(K).
 \end{aligned} \quad (45)$$

Here (a) holds for the distribution $p_*(t, \mathbf{x})$ that attains $V_\lambda(K)$; (b) holds since (T_1, \mathbf{X}_1) and (T_2, \mathbf{X}_2) are independent by assumption; (c) follows by Proposition 3; (d) follows by Proposition 5; (e) follows from

$$\begin{aligned}
 s_\lambda(\mathbf{X}_{\theta_1} | T, Y_{\theta_2}) &= \sum_{y_{\theta_2}} p(y_{\theta_2}) s_\lambda(\mathbf{X}_{\theta_1} | T, Y_{\theta_2} = y_{\theta_2}) \\
 &\stackrel{(h)}{\geq} \sum_{y_{\theta_2}} p(y_{\theta_2}) S_\lambda(\mathbf{X}_{\theta_1} | Y_{\theta_2} = y_{\theta_2}) \\
 &\stackrel{(i)}{=} S_\lambda(\mathbf{X}_{\theta_1} | Y_{\theta_2})
 \end{aligned} \quad (46)$$

where (h) holds because $S_\lambda(\mathbf{X}_{\theta_1}|Y_{\theta_2} = y_{\theta_2})$ is the lower convex envelope of $s_\lambda(\mathbf{X}_{\theta_1}|Y_{\theta_2} = y_{\theta_2})$ and the chain $T \leftrightarrow \mathbf{X}_{\theta_1} \leftrightarrow Y_{\theta_1}$ conditioned on $Y_{\theta_2} = y_{\theta_2}$ is Markov (this Markov chain is an immediate implication of (39) where $T = (T_1, T_2)$) and (i) is the definition of $S_\lambda(\cdot)$; (g) holds since $S_\lambda(\mathbf{X}_{\theta_1})$ is convex in $p(\mathbf{x}_{\theta_1})$ and by Jensen's inequality $S_\lambda(\mathbf{X}_{\theta_1}|Y_{\theta_2}) \geq S_\lambda(\mathbf{X}_{\theta_1})$; (i) follows from definition of $V_\lambda(K)$ and by checking the constraint

$$\begin{aligned} E[\mathbf{X}_{\theta_1} \mathbf{X}_{\theta_1}^T] &= \sum_{t_1, t_2} p_*(t_1) p_*(t_2) \frac{1}{2} \left(E[\mathbf{X}_{t_1} \mathbf{X}_{t_1}^T] + E[\mathbf{X}_{t_2} \mathbf{X}_{t_2}^T] \right) \\ &= \sum_t p_*(t) E[\mathbf{X}_t \mathbf{X}_t^T] = K. \end{aligned}$$

We now see that all inequalities of (45) are equalities, and step (d) combined with Proposition 5 proves the claim. \square

Corollary 1: For every $\ell \in \mathbb{N}$, let $n = 2^\ell$ and $(T^n, \mathbf{X}_n) \sim \prod_{i=1}^n p_*(t_i, \mathbf{x}_i)$. Then $(T^n, \tilde{\mathbf{X}}_n)$ achieves $V_\lambda(K)$ where $\tilde{\mathbf{X}}_n|(T_n = (t_1, t_2, \dots, t_n)) \sim \frac{1}{\sqrt{n}}(\mathbf{X}_{t_1} + \mathbf{X}_{t_2} + \dots + \mathbf{X}_{t_n})$. We choose $\mathbf{X}_{t_1}, \mathbf{X}_{t_2}, \dots, \mathbf{X}_{t_n}$ to be independent random variables.

Proof: The proof follows by induction using Lemma 1. \square

Now we state the main Lemma which shows that one of the optimizers is Gaussian.

Lemma 2: There is a single Gaussian distribution (i.e., the random variable T can be chosen to be a constant) that achieves $V_\lambda(K)$.

Proof: See Appendix B. \square

Note that our approach does not establish the uniqueness of the maximizing distribution.

IV. FEEDBACK SUM-RATE CAPACITY FOR SYMMETRIC GMACS

The proof of Theorem 1 is based on Theorem 2 with the partition $\mathcal{S}_1 = \{1\}, \mathcal{S}_2 = \{2\}, \dots, \mathcal{S}_J = \{J\}$. We tackle the resulting (non-convex) optimization problem with Lagrange duality.

Consider the covariance matrix

$$K = \begin{pmatrix} Q_1 & \rho_{12}\sqrt{Q_1 Q_2} & \dots & \rho_{1J}\sqrt{Q_1 Q_J} \\ \rho_{21}\sqrt{Q_2 Q_1} & Q_2 & \dots & \rho_{2J}\sqrt{Q_2 Q_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1}\sqrt{Q_J Q_1} & \rho_{J2}\sqrt{Q_J Q_2} & \dots & Q_J \end{pmatrix}. \quad (47)$$

Lemma 3: The original problem is bounded as follows

$$-C_{\text{sum}} = \min_{\substack{p(t, \mathbf{x}): E[X_i^2] \leq P_j \\ T \leftrightarrow \mathbf{X} \leftrightarrow Y \\ \text{subject to (28)}}} -I(\mathbf{X}; Y|T) \geq \max_{\lambda} \min_{K \geq 0: Q_j \leq P_j} h(\lambda, K) \quad (48)$$

where

$$h(\lambda, K) = \frac{(\lambda - 1)}{2} \log \left(1 + \sum_{j,k=1}^J [K]_{jk} \right) - \frac{\lambda}{2(J-1)} \sum_{j=1}^J \log \left(1 + \sum_{\ell, k=1}^J [K]_{\ell k} - \frac{(\sum_{k=1}^J [K]_{jk})^2}{P_j} \right). \quad (49)$$

Proof: See Appendix C. \square

We have shown that the optimal input distributions are Gaussian. At this point the problem is similar to the one in [10], and we can use Lemmas 4-5 from [10] to complete the optimization. However, the converse in [10] relies on a specific covariance matrix form with only two variables, and this does not necessarily work for asymmetric power constraints. Therefore, we provide a different analysis that applies to asymmetric power constraints.

Consider the covariance matrix

$$M = \begin{pmatrix} P_1 & \rho_{12}\sqrt{P_1 P_2} & \dots & \rho_{1J}\sqrt{P_1 P_J} \\ \rho_{21}\sqrt{P_2 P_1} & P_2 & \dots & \rho_{2J}\sqrt{P_2 P_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1}\sqrt{P_J P_1} & \rho_{J2}\sqrt{P_J P_2} & \dots & P_J \end{pmatrix}. \quad (50)$$

Lemma 4: For every $\lambda \geq 0$, we have

$$\max_{\lambda} \min_{K \geq 0: Q_j \leq P_j} h(\lambda, K) \geq \max_{\lambda} \min_{\rho_{12}, \dots, \rho_{N(N-1)}: M \geq 0} h(\lambda, M). \quad (51)$$

Proof: See Appendix D. \square

Remark 1: The argument in Lemma 4 is valid for any dependence balance constraint in (24).

We now set all power constraints to be the same, i.e., $P_1 = P_2 = \dots = P_J = P$.

Lemma 5: For every $\lambda \geq 0$, we have

$$\max_{\lambda} \min_{\rho_{12}, \dots, \rho_{J(J-1)}: M \geq 0} h(\lambda, M) \geq \max_{\lambda} \min_{\beta \in [0, J]} \left\{ \frac{(\lambda - 1)}{2} \log(1 + JP\beta) - \frac{J\lambda}{2(J-1)} \log(1 + P\beta(J - \beta)) \right\}. \quad (52)$$

Proof: See Appendix E. □

Remark 2: The paper [10] (see also [5]) shows that the optimal covariance matrix is of a simpler form with only two degrees of freedom. Our argument is based on the Cauchy-Schwarz inequality, as shown in the proof of Lemma 5.

We define the function

$$\ell(\beta, J, P) = \frac{1}{2} \log(1 + JP\beta) - \frac{J}{2(J-1)} \log(1 + P\beta(J - \beta)). \quad (53)$$

Lemma 6: There exists $\lambda^* \geq 0$ such that

$$\max_{\lambda} \min_{\beta \in [0, J]} \left\{ -\frac{1}{2} \log(1 + JP\beta) + \lambda \ell(\beta, J, P) \right\} \quad (54)$$

$$= -\frac{1}{2} \log(1 + JP\beta) \quad (55)$$

where $\beta \in [1, J]$ is the unique solution to $(1 + JP\beta)^{J-1} = (1 + P\beta(J - \beta))^J$.

Proof: We have

$$\max_{\lambda} \min_{\beta \in [0, J]} \left\{ -\frac{1}{2} \log(1 + JP\beta) + \lambda \ell(\beta, J, P) \right\} \quad (56)$$

$$\stackrel{(a)}{=} \min_{\beta \in [1, J]: \ell(\beta, J, P) \leq 0} \left\{ -\frac{1}{2} \log(1 + JP\beta) \right\} \quad (57)$$

$$\stackrel{(b)}{=} -\frac{1}{2} \log(1 + JP\beta) \quad (58)$$

where β is the unique solution satisfying $\beta \in [1, J]$ and $(1 + P\beta)^{J-1} = (1 + P\beta(J - \beta))^J$; step (a) follows from strong duality as the problem is convex from Lemma 7 in Appendix F and satisfy Slater's condition. Slater's condition holds because there exists a β such that $\ell(\beta, J, P) < 0$, i.e., $\beta = 1$ so the primal problem is strictly feasible. Step (b) follows from the Karush-Kuhn-Tucker (KKT) conditions for a convex problem which satisfy the Slater's condition to find the primal and dual optimal β^* and λ^* . We start by showing that $\lambda^* \neq 0$. Suppose for now that $\lambda^* = 0$, then from the KKT conditions we have

$$\frac{\partial}{\partial \beta} \left\{ -\frac{1}{2} \log(1 + JP\beta) + \lambda \ell(\beta, J, P) \right\} \Big|_{\lambda=0} = -\frac{JP}{2(1 + JP\beta)} = 0 \quad (59)$$

which implies that $P = 0$. This is impossible, so by contradiction we have $\lambda^* \neq 0$. Now by using complementary slackness condition $\lambda^* \cdot \ell(\beta^*, J, P) = 0$ we deduce that $\ell(\beta^*, J, P) = 0$, which is equivalent to

$$(1 + JP\beta^*)^{J-1} = (1 + P\beta^*(J - \beta^*))^J. \quad (60)$$

This equation has a unique solution for $\beta \in [1, J]$, see [5, Lemma 1], [10, Appendix A]. Finally, the value

$$\lambda^* = \left(1 - \frac{(J - 2\beta)(1 + JP\beta)}{(J - 1)(1 + P\beta(J - \beta))} \right)^{-1} \quad (61)$$

is a valid choice since λ^* is positive because of (59), which concludes the proof of the converse. □

V. CONCLUSIONS

We have derived a new converse bound that combines the Lagrange duality approach of [8] with a novel variation of the factorization of a convex envelope [18]. The new converse bound meets the achievable sum-rate of the Fourier MEC scheme, thus establishing the sum-rate capacity for the J -user symmetric GMAC with feedback.

It remains to see whether Fourier MEC can achieve all rate points in the capacity region of the symmetric GMAC with feedback. For asymmetric transmit power constraints, however, it is known that Fourier-MEC can be improved by using modulation frequencies other than the uniformly-spaced frequencies $(j - 1)/J$ for $j \in \mathcal{J}$. A few more variations of MEC strategies are described in [5, Sec. VIII].

Example 1: Fourier MEC does not meet the dependence-balance bound under asymmetric power constraints. For example, consider three users with the power constraints $P_1 \leq 1, P_2 \leq 4$ and $P_3 \leq 9$, for which Fourier-MEC achieves the sum-rate $R_{\text{sum}} = 1.6215$ bits/use, see [5, Sec. III]. However, the dependence balance bound permits a larger sum-rate, since the choice $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.5, 0.44, 0.58)$ satisfies the dependence balance constraints and permits $R_{\text{sum}} = 1.6427$.

APPENDIX A
EXISTENCE OF MINIMIZING DISTRIBUTION

Proposition 7 ([23, Lemma 1]): Suppose that Y_n and Y have continuous densities $f_n(y)$, $f(y)$ with respect to the Lebesgue measure on \mathbb{R} . If $Y_n \xrightarrow{w} Y$ and

$$\sup_n |f_n(y)| \leq M(y) < \infty, \quad \forall y \in \mathbb{R} \quad (62)$$

and f_n is equicontinuous, i.e., $\forall y, \epsilon > 0, \exists \delta(y, \epsilon), n(y, \epsilon)$ such that $\|y - y_1\| < \delta(y, \epsilon)$ implies that $|f_n(y) - f_n(y_1)| < \epsilon$ $\forall n \geq n(y, \epsilon)$, then for any compact subset C of \mathbb{R} we have

$$\sup_{y \in C} |f_n(y) - f(y)| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (63)$$

If $\{f_n\}$ is uniformly equicontinuous, i.e., $\delta(y, \epsilon), n(y, \epsilon)$ do not depend on y , and $f(y_n) \rightarrow 0$ whenever $\|y_n\| \rightarrow \infty$ then

$$\sup_{y \in \mathbb{R}} |f_n(y) - f(y)| = \|f_n(y) - f(y)\|_\infty \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (64)$$

Proposition 8 ([18, Proposition 16]): Let $\{X_n\}$ be any sequence of random variables satisfying $Y_n = GX_n + Z$ where $Z \sim \mathcal{N}(0, I)$ is independent of $\{X_n\}$ and $f_n(y)$ represent the density of Y_n . Then the collection of functions $\{f_n(y)\}$ is uniformly bounded and uniformly equicontinuous.

Definition 1: A collection of random variables X_n on \mathbb{R}^N is said to be tight if for every $\epsilon > 0$ there is a compact set $C_\epsilon \subset \mathbb{R}^N$ such that $P(X_n \notin C_\epsilon) \leq \epsilon, \forall n$.

Proposition 9 ([18, Proposition 17]): Consider a sequence of random variables $\{X_n\}$ that satisfies the covariance constraint $E[X_n X_n^T] = K, \forall n$. Then the sequence is tight.

Theorem 3 (Prokhorov): If $\{X_n\}$ is a tight sequence of random variables in \mathbb{R}^N then there exists a subsequence $\{X_{n_i}\}$ and a limiting probability distribution X_* such that $X_{n_i} \xrightarrow{w} X_*$.

Proposition 10 ([18, Proposition 18]): Let $X_n \xrightarrow{w} X_*$ and let $Z \sim \mathcal{N}(0, 1)$ be pairwise independent of $\{X_n\}, X_*$. Let $Y_n = GX_n + Z, Y_* = GX_* + Z$. Further let $E[X_n X_n^T] = K, E[X_* X_*^T] = K$. Let $f_n(y)$ denote the density of Y_n and $f_*(y)$ denote the density of Y_* . Then we have

- 1) $Y_n \xrightarrow{w} Y_*$,
- 2) $f_n(y) \rightarrow f_*(y)$ for all y ,
- 3) $h(Y_n) \rightarrow h(Y_*)$.

Proposition 11 (Lower Semi-continuity): Let $X_n \xrightarrow{w} X_*$ and $Y_n = GX_n + Z, Y_* = GX_* + Z$, where $Z \sim \mathcal{N}(0, 1)$ is pairwise independent of $\{X_n\}, X_*$. Let $s_\lambda(X_n) = (\lambda - 1)h(Y_n) + \left(\frac{\lambda}{J-1} + 1\right)h(Z) - \frac{\lambda}{J-1} \sum_{j=1}^J h(Y_n|X_{j_n})$ and $s_\lambda(X_*)$ similarly. Then

- 1) $(Y_n, X_{1n}) \xrightarrow{w} (Y_*, X_{1*})$,
- 2) $\liminf_{n \rightarrow \infty} s_\lambda(X_n) \geq s_\lambda(X_*)$.

Proof: The first part follows from pointwise convergence of characteristic functions (which is equivalent to weak convergence by Levy's continuity theorem) since $\Phi_{(X_n, Z)}(\mathbf{u}, v) = E[e^{i\mathbf{u}X_n + ivZ}] = E[e^{i\mathbf{u}X_n}]E[e^{ivZ}] = \Phi_{X_n}(\mathbf{u})\Phi_Z(v)$ then by letting $n \rightarrow \infty$ we have $\Phi_{X_*}(\mathbf{u})\Phi_Z(v) = E[e^{i\mathbf{u}X_*}]E[e^{ivZ}] = E[e^{i\mathbf{u}X_* + ivZ}] = \Phi_{(X_*, Z)}(\mathbf{u}, v)$. To relate (Y_n, X_{1n}) with (X_n, Z) we use the linear transformation $(Y_n, X_{1n})^T = A(X_n, Z)^T$ for a deterministic matrix A . By using the previous steps and the linear dependence we obtain $\lim_{n \rightarrow \infty} \Phi_{(Y_n, X_{1n})}(\mathbf{t}) = \lim_{n \rightarrow \infty} \Phi_{(X_n, Z)}(A\mathbf{t}) = \Phi_{(X_*, Z)}(A\mathbf{t}) = \Phi_{(Y_*, X_{1*})}(\mathbf{t})$.

For the second part fix $\delta > 0$ and define $N_\delta \sim \mathcal{N}(0, \delta)$, pairwise independent of $\{X_n\}, X_*$. From the Markov chain $X_{1n} + N_\delta \leftrightarrow X_{1n} \leftrightarrow Y_n$ and the data processing inequality we have $h(Y_n|X_{1n}) \leq h(Y_n|X_{1n} + N_\delta)$. By the third claim of proposition 10, we obtain

$$(\lambda - 1)h(Y_n) + \left(\frac{\lambda}{J-1} + 1\right)h(Z) - \frac{\lambda}{J-1} \sum_{j=1}^J h(Y_n|X_{j_n} + J_\delta) \rightarrow (\lambda - 1)h(Y_*) + \left(\frac{\lambda}{J-1} + 1\right)h(Z) - \frac{\lambda}{J-1} \sum_{j=1}^J h(Y_*|X_{j_*} + N_\delta) \quad (65)$$

as $n \rightarrow \infty$. Thus, we have

$$\liminf_{n \rightarrow \infty} s_\lambda(X_n) \geq (\lambda - 1)h(Y_*) + \left(\frac{\lambda}{J-1} + 1\right)h(Z) - \frac{\lambda}{J-1} \sum_{j=1}^J h(Y_*|X_{j_*} + N_\delta). \quad (66)$$

Since the RHS of (66) is continuous in δ , we take $\delta \downarrow 0$ and prove the second claim. \square

Theorem 4 ([24, Theorem 1]): Let $\{Y_i \in C\}$ be a sequence of continuous random variables with pdf's $\{f_i\}$ and Y_* be a continuous random variable with pdf f_* such that $f_i \rightarrow f_*$ pointwise. Let $\|y\| = \sqrt{y^\dagger y}$ denote the Euclidean norm of $y \in C$. If the conditions

$$\max\{\sup_y f_i(y), \sup_y f_*(y)\} \leq F, \quad (67)$$

$$\max\left\{\int \|y\|^\kappa f_i(y) dy, \int \|y\|^\kappa f_*(y) dy\right\} \leq L \quad (68)$$

hold for some $\kappa > 1$ and for all i then $h(Y_i) \rightarrow h(Y_*)$.

Remark 3: This theorem is relatively straightforward. We have $\liminf_i h(Y_i) \geq h(Y_*)$ due to the upper bound on the densities and $\limsup_i h(Y_i) \leq h(Y_*)$ due to the moment constraints.

Proof of Proposition 6: Define

$$v_\lambda(\hat{K}) = \inf_{p(x): E[XX^T] = \hat{K}} s_\lambda(X). \quad (69)$$

Let X_n be a sequence of random variables such that $E[X_n X_n^T] = \hat{K}$ and $s_\lambda(X_n) \downarrow v_\lambda(\hat{K})$. By the covariance constraint (Proposition 9) we know that the sequence of random variables X_n forms a tight sequence and by Theorem 3 there exists $X_{\hat{K}}^*$ and a convergent subsequence such that $X_{n_i} \xrightarrow{w} X_{\hat{K}}^*$. From Proposition 10 and 11 we have $s_\lambda(X_{\hat{K}}^*) = v_\lambda(\hat{K})$. For $\lambda \geq 0$ we have the following trivial bound

$$v_\lambda(\hat{K}) = s_\lambda(X_{\hat{K}}^*) \geq -\left(\frac{\lambda}{J-1} + 1\right) I(X_{\hat{K}}^*; Y) \geq \left(\frac{-\lambda + J - 1}{2(J-1)}\right) \log(1 + G\hat{K}G^T) = C_\lambda. \quad (70)$$

Recall that $V_\lambda(K)$ is defined using a convex combination:

$$V_\lambda(K) = \inf_{\substack{(V, X): E[XX^T] = K \\ T \leftrightarrow X \leftrightarrow Y}} s_\lambda(X|V). \quad (71)$$

Hence to obtain the best convex combination subject to the covariance constraint it suffices to consider the family of maximizers $X_{\hat{K}}^*$ for $\hat{K} \geq 0$. Thus, we have

$$V_\lambda(K) = \inf_{\substack{\alpha_i, \hat{K}_i: \alpha_i \geq 0, \sum_i \alpha_i = 1 \\ \sum_i \alpha_i \hat{K}_i = K}} \sum_i \alpha_i v_\lambda(\hat{K}_i). \quad (72)$$

It takes $\frac{J(J+1)}{2}$ constraints to preserve the covariance matrix and one constraint to preserve $\sum_i \alpha_i v_\lambda(\hat{K}_i)$. Hence, by using the Bunt-Caratheodory theorem, we can consider convex combinations of at most $m := \frac{J(J+1)}{2} + 1$ points, i.e.,

$$V_\lambda(K) = \inf_{\substack{\alpha_i, \hat{K}_i: \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1 \\ \sum_{i=1}^m \alpha_i \hat{K}_i = K}} \sum_{i=1}^m \alpha_i v_\lambda(\hat{K}_i). \quad (73)$$

Consider any sequence of convex combinations $(\{\alpha_i^n\}, \{K_i^n\})$ that approaches the supremum as $n \rightarrow \infty$. Using compactness of the m -dimensional simplex, we can assume w.l.o.g. that $\alpha_i^n \xrightarrow{n \rightarrow \infty} \alpha_i^*$, $i = 1, \dots, m$. If any $\alpha_i^* = 0$, since $\alpha_i^n K_i^n = K$ and $v_\lambda(K_i^n) \geq C_\lambda$ it is easy to see that $\alpha_i^n v_\lambda(K_i^n) \xrightarrow{n \rightarrow \infty} 0$. Thus we can assume that $\min_{i=1, \dots, m} \alpha_i^* = \alpha^* > 0$. This implies that $K_i^n \leq \frac{2}{\alpha^* K}$ for large enough n uniformly in i . Hence we can find a convergent subsequence for each i , $1 \leq i \leq m$, so that $K_i^{n_k} \xrightarrow{k \rightarrow \infty} K_i^*$. We thus have

$$V_\lambda(K) = \sum_{i=1}^m \alpha_i^* v_\lambda(K_i^*). \quad (74)$$

In other words, we can find a pair of random variables (T_*, X_*) with $|\mathcal{T}| \leq \frac{J(J+1)}{2} + 1$ such that $V_\lambda(K) = s_\lambda(X_*|T_*)$. \square

APPENDIX B PROOF OF LEMMA 2

We define the set of typical sequences as $\mathcal{T}^{(n)}(T) := \{t^n : |\{i : t_i = t\} - np_*(t)| \leq n\omega_n p_*(t), \forall t \in [1 : m]\}$, where ω_n is any sequence such that $\omega_n \rightarrow 0$ as $n \rightarrow \infty$ and $\omega_n \sqrt{n} \rightarrow \infty$ as $n \rightarrow \infty$. For instance $\omega_n = \frac{\log n}{\sqrt{n}}$. By using Chebyshev's inequality, we have

$$P(|\{i : t_i = t\} - np_*(t)| \leq n\omega_n p_*(t)) \leq \frac{1 - p_*(t)}{p_*(t)\omega_n^2 n}.$$

Hence $P(t^n \notin \mathcal{T}^{(n)}(T)) \rightarrow 0$ as $n \rightarrow \infty$. Consider a sequence of induced distributions $\hat{X}_n \sim \tilde{X}_n|t^n$, where $\tilde{X}_n|t^n$ and $\tilde{X}_n|(T^n = t^n)$ denote the same thing.

Proposition 12: $\hat{X}_n \stackrel{w}{\Rightarrow} \mathcal{N}(0, \sum_{t=1}^m p_*(t)K_t)$.

Proof: For given t^n , let $A_n(t) = |\{i : t_i = t\}|$. We know that $A_n(t) \in np_*(t)(1 \pm \omega_n)$, $\forall t$. Consider a \mathbf{c} with entries in real number and $\|\mathbf{c}\| = 1$. Let $\hat{X}_{n,i}^c \sim \frac{1}{\sqrt{n}}\mathbf{c}^T \cdot X_{t_i}$ and $\hat{X}_{n,i}^c$ be independent random variable over i . Observe that $\sum_{i=1}^n \hat{X}_{n,i}^c \sim \mathbf{c}^T \hat{X}_n$. Note that

$$\begin{aligned} \sum_{i=1}^n E[(\hat{X}_{n,i}^c)^2] &= \frac{1}{n} \sum_t A_n(t) \mathbf{c}^T K_t \mathbf{c} \\ &\rightarrow \mathbf{c}^T \left(\sum_t p_*(t) K_t \right) \mathbf{c}. \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^n E[(\hat{X}_{n,i}^c)^2; |\hat{X}_{n,i}^c| > \epsilon_1] &= \frac{1}{n} \sum_t A_n(t) E[\mathbf{c}^T X_t X_t^T \mathbf{c}; \mathbf{c}^T X_t X_t^T \mathbf{c} > n\epsilon_1^2] \\ &\leq \sum_t p_*(t)(1 + \omega_n) E[\mathbf{c}^T X_t X_t^T \mathbf{c}; \mathbf{c}^T X_t X_t^T \mathbf{c} > n\epsilon_1^2] \rightarrow 0. \end{aligned}$$

For the last step, we used that the K_t 's are bounded, and hence $\mathbf{c}^T X_t$ has a bounded second moment. The Lindeberg-Feller Central Limit Theorem gives $\sum_{i=1}^n \hat{X}_{n,i}^c \stackrel{w}{\Rightarrow} \mathcal{N}(0, \mathbf{c}^T \sum_t p_*(t) K_t \mathbf{c})$. Hence $\hat{X}_n \stackrel{w}{\Rightarrow} \mathcal{N}(0, \sum_t p_*(t) K_t)$ from Kramer-Wold. \square

Proposition 13: Given any $\delta > 0$, there exists N_0 such that $\forall n > N_0$ we have for all $t^n \in \mathcal{T}^{(n)}(T)$

$$s_\lambda(\tilde{X}_n|t^n) - s_\lambda(\mathbf{X}^*) \leq \delta,$$

where $\mathbf{X}^* \sim \mathcal{N}(0, \sum_t p_*(t) K_t)$.

Proof: Assume the claim is not true. Then we have a subsequence $t^{n_k} \in \mathcal{T}^{n_k}(T)$ and distributions $\tilde{X}_{n_k}|t^{n_k}$ such that

$$s_\lambda(\tilde{X}_{n_k}|t^{n_k}) > s_\lambda(\mathbf{X}^*) + \delta, \forall k.$$

However from Proposition 12 we know that $\tilde{X}_{n_k}|t^{n_k} \stackrel{w}{\Rightarrow} \mathbf{X}^*$ and from Proposition 10 we have $s_\lambda(\tilde{X}_{n_k}|t^{n_k}) \rightarrow s_\lambda(\mathbf{X}^*)$, a contradiction. \square

Proof of Lemma 2: We know from Corollary 1 that for every $\ell \in \mathbb{N}$ and $n = 2^\ell$, the pair (T^n, \tilde{X}_n) achieves $V_\lambda(K)$. Hence we have

$$\begin{aligned} V_\lambda(K) &= \sum_{t^n} p_*(t^n) s_\lambda(\tilde{X}_n|t^n) \\ &= \sum_{t^n \in \mathcal{T}^{(n)}(T)} p_*(t^n) s_\lambda(\tilde{X}_n|t^n) + \sum_{t^n \notin \mathcal{T}^{(n)}(T)} p_*(t^n) s_\lambda(\tilde{X}_n|t^n). \end{aligned}$$

For a given t^n , let $\hat{X} \sim \tilde{X}_n|t^n$ so that $E[\hat{X}\hat{X}^T] \leq \sum_{t=1}^m K_t$. Thus $s_\lambda(\hat{X}) \leq C_\lambda$ for some fixed constant that is independent of t^n . Using Proposition 13 we can upper bound $V_\lambda(K)$ for large n by

$$\begin{aligned} V_\lambda(K) &= \sum_{t^n \in \mathcal{T}^{(n)}(T)} p_*(t^n) s_\lambda(\tilde{X}_n|t^n) + \sum_{t^n \notin \mathcal{T}^{(n)}(T)} p_*(t^n) s_\lambda(\tilde{X}_n|t^n) \\ &\leq \sum_{t^n \in \mathcal{T}^{(n)}(T)} p_*(t^n) (s_\lambda(\mathbf{X}^*) + \delta) + C_\lambda \sum_{t^n \notin \mathcal{T}^{(n)}(T)} p_*(t^n) \\ &= P(t^n \in \mathcal{T}^{(n)}(T))(s_\lambda(\mathbf{X}^*) + \delta) + C_\lambda P(t^n \notin \mathcal{T}^{(n)}(T)). \end{aligned}$$

Here $\mathbf{X}^* \sim \mathcal{N}(0, \sum_t p_*(t) K_t)$. Since $P(t^n \in \mathcal{T}^{(n)}(T)) \rightarrow 1$ as $n \rightarrow \infty$ we have $V_\lambda(K) \leq s_\lambda(\mathbf{X}^*) + \delta$. However, $\delta > 0$ is arbitrary, and hence $V_\lambda(K) \leq s_\lambda(\mathbf{X}^*)$. The other direction $V_\lambda(K) \geq s_\lambda(\mathbf{X}^*)$ follows from the definition of $V_\lambda(K)$ and $\sum_t p_*(t) K_t \leq K$. \square

APPENDIX C
PROOF OF LEMMA 3

By the standard Lagrangian duality we have

$$\begin{aligned}
-C_{\text{sum}} &= \min_{\substack{p(t,x): E[X_j^2] \leq P_j \\ T \leftrightarrow X \leftrightarrow Y \\ \text{subject to (22)}}} -I(\mathbf{X}; Y|T) \stackrel{(a)}{\geq} \max_{\lambda} \min_{p(\mathbf{x}) \in G_{\mathcal{G}}} s_{\lambda}(\mathbf{X}) \\
&= \max_{\lambda} \min_{K \geq 0: Q_j \leq P_j} \left\{ \frac{(\lambda-1)}{2} \log \left(1 + \sum_{j,k=1}^J [K]_{jk} \right) - \frac{\lambda}{2(J-1)} \sum_{j=1}^J \log \left(1 + \sum_{\ell,k=1}^J [K]_{\ell k} - \frac{(\sum_{k=1}^J [K]_{jk})^2}{P_j} \right) \right\}
\end{aligned} \tag{75}$$

where (a) follows from Theorem 2 and last step by inserting an optimal Gaussian *input* distribution. \square

APPENDIX D
PROOF OF LEMMA 4

For the Gaussian MAC with $g_j = 1$ for all j we have

$$K_{XY} = \begin{pmatrix} K_X & \text{Cov}(\mathbf{X}, Y) \\ \text{Cov}(\mathbf{X}, Y)^T & K_Y \end{pmatrix} = \begin{pmatrix} K_X & K_X \mathbf{1} \\ (K_X \mathbf{1})^T & 1 + \mathbf{1}^T K_X \mathbf{1} \end{pmatrix} \tag{76}$$

where $K_Y = 1 + \mathbf{1}^T K_X \mathbf{1}$, $\text{Cov}(\mathbf{X}, Y) = K_X \mathbf{1}$ and $\mathbf{1}$ is column vector of all ones. The function $h(\cdot)$ can be rewritten as follows:

$$2h(\lambda, K_X) = -\log \det K_Y - \frac{\lambda}{J-1} \log \frac{\prod_{j=1}^J \det K_{X_j Y}}{(\det K_Y)^{J-1} \prod_{j=1}^J \det K_{X_j}}. \tag{77}$$

We define K'_X to be the same as K_X except that the (1,1) entry of K'_X is P_1 rather than Q_1 . Then from (76) we have

$$K'_{X_1 Y} = K_{X_1 Y} \circ \begin{pmatrix} D & F \\ F & E \end{pmatrix}, \tag{78}$$

$$K'_{X_j Y} = K_{X_j Y} \circ \begin{pmatrix} 1 & 1 \\ 1 & E \end{pmatrix}, \quad j \neq 1 \tag{79}$$

where ‘ \circ ’ denotes Hadamard multiplication, and where

$$D = \frac{P_1}{Q_1}, \quad E = \frac{K_Y + P_1 - Q_1}{K_Y}, \quad F = \frac{\text{Cov}(X_1, Y) + P_1 - Q_1}{\text{Cov}(X_1, Y)}. \tag{80}$$

Observe that $D > 1$ and $E > 1$. Now by using Oppenheim’s inequality ($\det K'_{X_1 Y} \geq DE \det K_{X_1 Y}$) [25, p. 480] we have

$$2h(\lambda, K'_X) = -\log EK_Y - \frac{\lambda}{J-1} \log \frac{\left(\prod_{j=2}^J \det K'_{X_j Y} \right) \det K'_{X_1 Y}}{E^{J-1} DK_Y^{J-1} \prod_{j=1}^J K_{X_j}} \tag{81}$$

$$\leq -\log EK_Y - \frac{\lambda}{N-1} \log E \frac{\prod_{j=1}^J \det K_{X_j Y}}{K_Y^{J-1} \prod_{j=1}^J K_{X_j}} \tag{82}$$

$$= -(1 + \frac{\lambda}{J-1}) \log E + 2h(\lambda, K_X) \leq 2h(\lambda, K_X). \tag{83}$$

The minimum is reached for K'_X which is a covariance matrix from the argument below

$$K'_X = \begin{pmatrix} P_1 & \rho_{12} \sqrt{Q_1 Q_2} & \dots & \rho_{1J} \sqrt{Q_1 Q_J} \\ \rho_{21} \sqrt{Q_2 Q_1} & Q_2 & \dots & \rho_{2J} \sqrt{Q_2 Q_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1} \sqrt{Q_J Q_1} & \rho_{J2} \sqrt{Q_J Q_2} & \dots & Q_J \end{pmatrix} = \begin{pmatrix} P_1 & \rho_{12} \sqrt{\frac{Q_1}{P_1}} \sqrt{P_1 Q_2} & \dots & \rho_{1J} \sqrt{\frac{Q_1}{P_1}} \sqrt{P_1 Q_J} \\ \rho_{21} \sqrt{\frac{Q_1}{P_1}} \sqrt{Q_2 P_1} & Q_2 & \dots & \rho_{2J} \sqrt{Q_2 Q_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1} \sqrt{\frac{Q_1}{P_1}} \sqrt{Q_J P_1} & \rho_{J2} \sqrt{Q_J Q_2} & \dots & Q_J \end{pmatrix}. \tag{84}$$

With the same approach we attain the minimum for $Q_2 = P_2, Q_3 = P_3, \dots, Q_J = P_J$. Thus, we can get the desired lower bound on the original problem.

APPENDIX E
PROOF OF LEMMA 5

Consider the arithmetic mean

$$\rho = \frac{1}{J(J-1)} \left(\sum_{\substack{j,k=1 \\ j \neq k}}^J \rho_{jk} \right). \quad (85)$$

For the inequality in Lemma 5 to hold we need to show the following inequality

$$\prod_{j=1}^J \left(1 + \sum_{\ell,k=1}^J M_{\ell k} - \frac{\left(\sum_{k=1}^J M_{jk} \right)^2}{P} \right) \leq (1 + P\beta(J - \beta))^J. \quad (86)$$

We prove it as follows:

$$\prod_{j=1}^J \left(1 + \sum_{\ell,k=1}^J M_{\ell k} - \frac{\left(\sum_{k=1}^J M_{jk} \right)^2}{P} \right) \stackrel{(a)}{\leq} \left(1 + \sum_{\ell,k=1}^J M_{\ell k} - \sum_{j=1}^J \frac{\left(\sum_{k=1}^J M_{jk} \right)^2}{JP} \right)^J \quad (87)$$

$$\stackrel{(b)}{\leq} \left(1 + \sum_{\ell,k=1}^J M_{\ell k} - \left(\sum_{j,k=1}^J \frac{M_{jk}}{J\sqrt{P}} \right)^2 \right)^J \quad (88)$$

$$= (1 + P\beta(J - \beta))^J \quad (89)$$

where (a) follows from arithmetic-geometric mean (AM-GM) which is valid for non-negative real numbers, and (b) follows from Cauchy-Schwarz inequality

$$(1^2 + 1^2 + \dots + 1^2) \left(\sum_{j=1}^J \left(\sum_{k=1}^J \frac{M_{jk}}{\sqrt{P}} \right)^2 \right) \geq \left(\sum_{j,k=1}^J \frac{M_{jk}}{\sqrt{P}} \right)^2. \quad (90)$$

For equality in both (a) and (b) a sufficient and necessary condition is $\rho_{12} = \rho_{13} = \dots = \rho_{(J-1)J} = \rho$. Define $\beta = 1 + (J-1)\rho$, therefore to obtain the expression in (89) we will use the identity $\sum_{\ell,k=1}^J M_{\ell k} = PJ\beta$. Since $-1/(J-1) \leq \rho \leq 1$, we have $0 \leq \beta \leq J$.

APPENDIX F
CONVEXITY

Lemma 7: The problem $\min_{\beta \in [0, J]: \ell(\beta, J, P) \leq 0} \left\{ -\frac{1}{2} \log(1 + JP\beta) \right\}$ is convex, where

$$\ell(\beta, J, P) = \frac{1}{2} \log(1 + JP\beta) - \frac{J}{2(J-1)} \log(1 + P\beta(J - \beta)). \quad (91)$$

Proof: For a fixed P and J , the term $-\log(1 + JP\beta)$ is convex on β . We now show that, for a fixed P and J , $\ell(\beta, J, P)$ is convex in β . We split the derivations into two steps. In the first step, we show that $\frac{\partial^2 \ell(\beta, J, P)}{\partial \beta^2} \geq 0$ for $\beta \geq 1$. In the second step, we show that $\beta \in [1, J]$ are the only possible maximizers for the original problem.

1) We compute $\frac{\partial^2 \ell(\beta, J, P)}{\partial \beta^2}$ as

$$(1 \quad \gamma \quad \gamma^2 \quad \gamma^3 \quad \gamma^4) \begin{pmatrix} M^3 P^4 + 8M^2 P^4 + 2M^2 P^3 + 22MP^4 + 14MP^3 \\ + 3MP^2 + 21P^4 + 24P^3 + 11P^2 + 2P \\ 2M^3 P^4 + 16M^2 P^4 + 2M^2 P^3 + 46MP^4 + 20MP^3 \\ + 2MP^2 + 48P^4 + 42P^3 + 10P^2 \\ M^3 P^4 + 9M^2 P^4 + 33MP^4 + 10MP^3 + 45P^4 \\ + 30P^3 + 2P^2 \\ 2M^2 P^4 + 16MP^4 + 4MP^3 + 30P^4 + 12P^3 \\ M^2 P^4 + 7MP^4 + 12P^4 \end{pmatrix} > 0$$

where $M = J - 3 \geq 0$ and $\gamma = \beta - 1 \geq 0$ for the case of three or more users ($J \geq 3$). For $P > 0$ and $\gamma \geq 0$ the above inequality holds from the addition of only positive terms, thus the problem is convex. To show that the problem is convex for all admissible values of β , we need to rule out some values and that is done in the next part.

2) Assume that $\beta_0 \in [0, 1[$ is the minimizer of the problem defined above, so that the following holds

$$\min_{\beta \in [0, J]: \ell(\beta, J, P) \leq 0} \left\{ -\frac{1}{2} \log(1 + JP\beta) \right\} = -\frac{1}{2} \log(1 + JP\beta_0). \quad (92)$$

For $\beta_1 = 1$, the constraint $\ell(1, J, P) < 0$ is satisfied, which is equivalent to showing that

$$(1 + JP)^{J-1} < (1 + (J - 1)P)^J \quad (93)$$

for $P > 0$. But we have

$$\left(\frac{1 + (J - 1)P}{1 + JP} \right)^{J-1} = \left(1 - \frac{P}{1 + JP} \right)^{J-1} \quad (94)$$

$$\geq \frac{2}{1 + JP} \quad (95)$$

$$> \frac{1}{1 + (J - 1)P} \quad (96)$$

where (95) follows by Bernoulli's inequality ($(1 + x)^r \geq 1 + rx$ for $r \geq 1, x \geq -1$). The new minimizer is $\beta_1 = 1$ instead of $\beta_0 < \beta_1$ based on

$$-\frac{1}{2} \log(1 + JP\beta_1) < -\frac{1}{2} \log(1 + JP\beta_0). \quad (97)$$

This leads to a contradiction, and therefore we rule out the minimizers β_0 , and the original problem becomes equivalent to the problem

$$\min_{\beta \in [1, J]: \ell(\beta, J, P) \leq 0} \left\{ -\frac{1}{2} \log(1 + JP\beta) \right\}. \quad (98)$$

□

ACKNOWLEDGMENTS

The authors wish to thank O. Lévêque for helpful discussions and M. Wigger for pointing out an issue in an earlier version of the proof. This work was supported in part by the Swiss National Science Foundation under Grant 169294, Grant P2ELP2_165137. G. Kramer was supported in part by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research.

REFERENCES

- [1] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623 – 629, July 1984.
- [2] P. Elias, "Channel capacity without coding," MIT Research Laboratory of Electronics, Cambridge, MA, Quarterly Progress Report, October 1956.
- [3] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, pp. 172 – 182, April 1966.
- [4] R. G. Gallager and B. Nakiboglu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, January 2010.
- [5] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, June 2002.
- [6] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1221, Mar. 2011.
- [7] L. V. Truong, "Posterior matching scheme for Gaussian multiple access channel with feedback," *IEEE Inf. Theory Workshop*, 2014.
- [8] G. Kramer and M. Gastpar, "Dependence balance and the Gaussian multiaccess channel with feedback," *IEEE Inf. Theory Workshop*, pp. 198–202, March 2006.
- [9] A. Hekstra and F. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, pp. 44 – 53, January 1989.
- [10] E. Ardestanizadeh, M. A. Wigger, Y.-H. Kim, and T. Javidi, "Linear-feedback sum-capacity for Gaussian multiple access channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 224–236, January 2012.
- [11] A. Lapidoth and M. A. Wigger, "On the Gaussian MAC with imperfect feedback," *IEEE Convention of Electrical and Electronics Engineers in Israel*, 2006.
- [12] R. Tandon and S. Ulukus, "On the capacity region of the Gaussian multiple access channel with noisy feedback," *IEEE International Conference on Communications (ICC)*, 2009.
- [13] —, "Dependence balance based outer bounds for gaussian networks with cooperation and feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4063–4086, June 2011.
- [14] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.
- [15] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, pp. 4–21, January 2003.
- [16] R. Tandon and S. Ulukus, "Outer bounds for multiple-access channels with feedback using dependence balance," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4494–4507, October 2009.
- [17] M. Gastpar and G. Kramer, "On noisy feedback for interference channels," *Asilomar Conf. Signals, Systems, Computers*, Nov. 2006.

- [18] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IWCIT*, vol. 60, no. 4, April 2014.
- [19] T. Courtade, "Strengthening the entropy power inequality," *IEEE Int. Symp. Inf. Theory*, 2016.
- [20] E. Lieb, "Gaussian kernels have only Gaussian maximizers," *Inventiones mathematicae*, pp. 179–208, 1990.
- [21] E. Carlen, "Superadditivity of fisher's information and logarithmic sobolev inequalities," *J. Functional Analysis*, vol. 101, no. 1, pp. 194–211, Oct. 1991.
- [22] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [23] D. D. Boos, "On a converse to Scheffe's theorem," *The Annals of Statistics*, vol. 13, no. 1, pp. 423–427, 1985.
- [24] M. Godavarti and A. Hero, "Convergence of differential entropies," *IEEE Int. Symp. Inf. Theory*, vol. 50, no. 1, pp. 171–176, January 2004.
- [25] R. A. Horn and C. R. Jonson, *Matrix Analysis*. Cambridge University Press, 1985.