Linear-Feedback Sum-Capacity for Gaussian Multiple Access Channels

Ehsan Ardestanizadeh, *Member, IEEE*, Michèle Wigger, *Member, IEEE*, Young-Han Kim, *Member, IEEE*, and Tara Javidi, *Member, IEEE*

***Abstract*—The capacity region of the N-sender Gaussian multiple access channel with feedback is not known in general. This paper studies the class of *linear-feedback codes* that includes (nonlinear) nonfeedback codes at one extreme and the linear-feedback codes by Schalkwijk and Kailath, Ozarow, and Kramer at the other extreme. The *linear-feedback sum-capacity* $C_{\mathrm{L}}(N,P)$ under symmetric power constraints P is characterized, the maximum sum-rate achieved by linear-feedback codes when each sender has the equal block power constraint P. In particular, it is shown that Kramer's code achieves this linear-feedback sum-capacity. The proof involves the dependence balance condition introduced by Hekstra and Willems and extended by Kramer and Gastpar, and the analysis of the resulting nonconvex optimization problem via a Lagrange dual formulation. Finally, an observation is presented based on the properties of the *conditional maximal correlation*—an extension of the Hirschfeld-Gebelein-Rényi maximal correlation—which reinforces the conjecture that Kramer's code achieves not only the linear-feedback sum-capacity, but also the sum-capacity itself (the maximum sum-rate achieved by arbitrary feedback codes).**

***Index Terms*—Feedback, Gaussian multiple access channel, Kramer's code, linear-feedback codes, maximal correlation, sum-capacity.**

I. Introduction

FEEDBACK from the receivers to the senders can improve the performance of the communication systems in various ways. For example, as first shown by Gaarder and Wolf [1], feedback can enlarge the capacity region of memoryless multiple access channels by enabling the distributed senders to cooperate via coherent transmissions.

In this paper, we study the sum-capacity of the additive white Gaussian noise multiple access channel (Gaussian multiple access channel in short) with feedback depicted in Fig. 1. For $N = 2$ senders, Ozarow [2] established the capacity region which—unlike for the point-to-point channel—is strictly larger than the one without feedback. The capacity-achieving code proposed by Ozarow is an extension of the Schalkwijk-Kailath code [3], [4] for Gaussian point-to-point channels.

For $N \geq 3$, the capacity region is not known in general. Thomas [5] proved that feedback can at most double the sum capacity, and later Ordentlich [6] showed that the same bound holds for the entire capacity region even when the noise sequence is not white (cf. Pombra and Cover [7]). More recently, Kramer [8] extended Ozarow's linear-feedback code to $N \geq 3$ senders, and proved that this code achieves the sum-capacity under symmetric block power constraints P on all the senders, when the power P is above a certain threshold (see (4) in Section II) that depends on the number of senders N.

In this paper, we focus on the class of *linear-feedback codes*, where the feedback signals are incorporated linearly into the transmit signals (see Definition 1 in Section II). This class of codes includes the linear-feedback codes by Schalkwijk and Kailath [3], Ozarow [2], and Kramer [8] as well as arbitrary (nonlinear) nonfeedback codes.

We characterize the *linear-feedback sum-capacity* $C_{\mathrm{L}}(N,P)$ under symmetric block power constraints P, which is the maximum sum-rate achieved by linear-feedback codes under equal block power constraints P at all the senders. Our main contribution is the proof of the converse. We first prove an upper bound on $C_{\mathrm{L}}(N,P)$, which is a multiletter optimization problem over Gaussian distributions satisfying a certain functional relationship (cf. Cover and Pombra [9]). Next, we relax the functional relationship by considering a dependence balance condition, introduced by Hekstra and Willems [10] and extended by Kramer and Gastpar [11], and derive an optimization problem over the set of positive semidefinite (covariance) matrices. Lastly, we carefully analyze this nonconvex optimization problem via a Lagrange dual formulation [12].

The linear-feedback sum-capacity $C_{\mathrm{L}}(N,P)$ is achieved by Kramer's linear-feedback code. Hence, this rather simple code, which iteratively refines the receiver's knowledge about the messages, is sum-rate optimal among the class of linear-feedback codes. For completeness, we briefly describe Kramer's linear-feedback code and analyze it via properties of discrete algebraic Riccati recursions (cf. Wu *et al.* [13]). This analysis differs from the original approaches by Ozarow [2] and Kramer [8].

The complete characterization of the *sum-capacity* $C(N,P)$ under symmetric block power constraints P, i.e., the maximum sum-rate achieved by arbitrary feedback codes, still remains

Manuscript received February 09, 2010; revised June 02, 2011; accepted August 10, 2011. Date of current version January 06, 2012. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

E. Ardestanizadeh was with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093-0407 USA. He is now with ASSIA, Inc., Redwood City, CA 94065 USA (e-mail: eardestani@assia-inc.com).

M. Wigger was with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093-0407 USA. She is now with the Department of Communications and Electronics, Telecom ParisTech, Paris Cedex 13, France (e-mail: michele.wigger@telecom-paristech.fr).

Y.-H. Kim and T. Javidi are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093-0407 USA (e-mail: yhk@ucsd.edu; tjavidi@ucsd.edu).

Communicated by M. C. Gastpar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2011.2169307

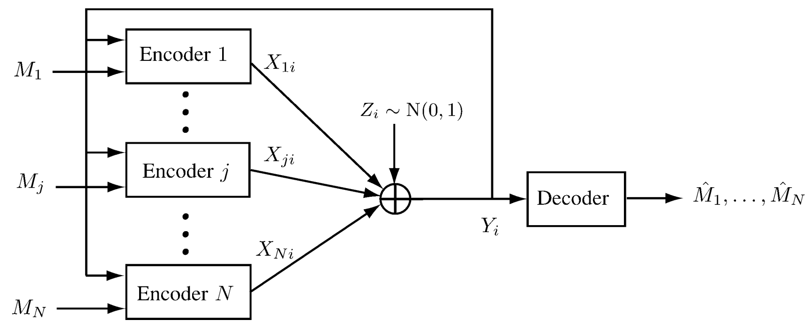

Fig. 1. N-sender Gaussian multiple access channel.

open. However, it has been commonly believed (cf. [11], [13]) that linear-feedback codes achieve the sum-capacity, i.e., $C(N,P) = C_{\mathrm{L}}(N,P)$. We offer an observation that further supports this conjecture. By introducing and analyzing the properties of *conditional maximal correlation*, which is an extension of the Hirschfeld-Gebelein-Rényi maximal correlation [14] to the case where an additional common random variable is shared, we show in Section V that the linear-feedback codes are *greedy optimal* for a multiletter optimization problem that upper bounds $C(N,P)$.

The rest of the paper is organized as follows. In Section II we formally state the problem and present our main result. Section III provides the proof of the converse and Section IV gives an alternative proof of achievability via Kramer's linear-feedback code. Section V concludes the paper with a discussion on potential extensions of the main ideas to nonequal power constraints and arbitrary feedback codes, and with a proof that linear-feedback codes are greedy optimal for a multiletter optimization problem that upper bounds $C(N,P)$.

We closely follow the notation in [15]. In particular, a random variable is denoted by an upper case letter (e.g., X, Y, Z) and its realization is denoted by a lower case letter (e.g., x, y, z). The shorthand notation X^n is used to denote the tuple (or the column vector) of random variables $(X_1, \ldots, X_n)$, and x^n is used to denote their realizations. A random column vector and its realization are denoted by boldface letters (e.g., $\mathbf{X}$ and $\mathbf{x}$) as well. Uppercase letters (e.g., A, B, C) also denote deterministic matrices, which can be distinguished from random variables based on the context. The (i,j) element of a matrix A is denoted by A_{ij}. The conjugate transpose of a real or complex matrix A is denoted by A' and the determinant of A is denoted by $|A|$. For the cross-covariance matrix of two random vectors $\mathbf{X}$ and $\mathbf{Y}$, we use the shorthand notation $K_{\mathbf{XY}} := \mathsf{E}(\mathbf{XY}') - \mathsf{E}(\mathbf{X})\mathsf{E}(\mathbf{Y}')$ and for the covariance matrix of a random vector $\mathbf{X}$ we use $K_{\mathbf{X}} := K_{\mathbf{XX}}$. Calligraphic letters (e.g., $\mathcal{A}, \mathcal{B}, \mathcal{C}$) denote discrete sets. Let $(X_1, \ldots, X_N)$ be a tuple of N random variables and $\mathcal{A} \subseteq \mathcal{S} := \{1, \ldots, N\}$. The subtuple of random variables with indices from $\mathcal{A}$ is denoted by $X(\mathcal{A}) := (X_j : j \in \mathcal{A})$. For every positive real number m, the shorthand notation $[1 : 2^m]$ is used to denote the set of integers $\{1, \ldots, 2^{\lceil m \rceil}\}$.

II. Problem Setup and the Main Result

Consider the communication problem over a Gaussian multiple access channel with feedback depicted in Fig. 1. Each sender $j \in \{1, \ldots, N\}$ wishes to transmit a message M_j reliably to the common receiver. At each time $i = 1, \ldots, n$, the output of the channel is

$$Y_i = \sum_{j=1}^{N} X_{ji} + Z_i \tag{1}$$

where $\{Z_i\}$ is a discrete-time zero-mean white Gaussian noise process with unit average power, i.e., $\mathsf{E}(Z_i^2) = 1$, and is independent of $(M_1, \ldots, M_N)$. We assume that the output symbols are causally fed back to each sender, and that the transmitted symbol X_{ji} from sender j at time i can thus depend on both the previous channel output sequence $Y^{i-1} := (Y_1, Y_2, \ldots, Y_{i-1})$ and the message M_j. We define a $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ feedback code as

1) N message sets $\mathcal{M}_1, \ldots, \mathcal{M}_N$, where $\mathcal{M}_j := [1 : 2^{nR_j}]$ for $j = 1, \ldots, N$;
2) a set of N encoders, where encoder $j \in \{1, \ldots, N\}$ assigns a symbol $x_{ji}(m_j, y^{i-1})$ to its message $m_j \in \mathcal{M}_j$ and the past channel output sequence $y^{i-1} \in \mathbb{R}^{i-1}$ for $i \in \{1, \ldots, n\}$; and
3) a decoder that assigns message estimates $\hat{m}_j \in [1 : 2^{nR_j}]$, $j \in \{1, \ldots, N\}$, to each received sequence y^n.

We assume throughout that $M(\mathcal{S}) := (M_1, \ldots, M_N)$ is uniformly distributed over $[1 : 2^{nR_1}] \times \cdots \times [1 : 2^{nR_N}]$. The probability of error is defined as

$$P_e^{(n)} := \mathsf{P}\{\hat{M}(\mathcal{S}) \neq M(\mathcal{S})\}.$$

A rate tuple $(R_1, \ldots, R_N)$ and its corresponding sum-rate $R = \sum_{j=1}^{N} R_j$ are said to be achievable under the power constraints $(P_1, \ldots, P_N)$ if there exists a sequence of $(2^{nR_1}, \ldots, 2^{nR_N}, n)$ feedback codes such that the *expected block power constraints*

$$\frac{1}{n} \sum_{i=1}^{n} \mathsf{E}\left(X_{ji}^2(M_j, Y^{i-1})\right) \leq P_j, \quad j = 1, \ldots, N$$

are satisfied and $\lim_{n\to\infty} P_e^{(n)} = 0$. The supremum over all achievable sum-rates is referred to as the *sum-capacity*. In most of the paper, we will be interested in the case of symmetric power constraints $P_1 = P_2 = \cdots = P_N = P$. In this case we denote the sum-capacity by $C(N,P)$.

Our focus will be on the special class of linear-feedback codes defined as follows.

Definition 1: A $(2^{nR_1}, \dots, 2^{nR_N}, n)$ feedback code is said to be a *linear-feedback code* if the encoder $x_{ji}(m_j, y^{i-1})$ has the form

$$x_{ji} = L_{ji}(\boldsymbol{\theta}_j(m_j), y^{i-1})$$

where

1) the (potentially nonlinear) *nonfeedback mapping* $\boldsymbol{\theta}_j$ is independent of i and maps the message m_j to a k-dimensional real vector (message point) $\boldsymbol{\theta}_j$ for some $k \in \{1, \dots, n\}$; and
2) the *linear-feedback mapping* L_{ji} maps the message point $\boldsymbol{\theta}_j(m_j)$ and the past feedback output sequence y^{i-1} to the channel input symbol x_{ji}.

The class of linear-feedback codes includes as special cases the feedback codes by Schalkwijk and Kailath [3], Ozarow [2], and Kramer [8], and all nonfeedback codes. To recover the codes by Schalkwijk and Kailath [3] and Ozarow [2] it suffices to choose $k = 1$; for Kramer's code [8] we need $k = 2$; and to recover all nonfeedback codes we have to choose $k = n$ and each message point $\boldsymbol{\theta}_j$ equal to the codeword sent by encoder j.

The *linear-feedback sum-capacity* is defined as the maximum achievable sum-rate using only linear-feedback codes. Under symmetric block power constraints $P_1 = \cdots = P_N = P$, we denote the linear-feedback sum-capacity by $C_{\mathrm{L}}(N, P)$.

We are ready to state the main result of this paper.

Theorem 1: For the Gaussian multiple access channel with symmetric block power constraints P, the linear-feedback sum-capacity is

$$C_{\mathrm{L}}(N, P) = \frac{1}{2}\log(1 + NP\phi(N, P)) \tag{2}$$

where $\phi(N, P)$ is the unique solution to

$$(1 + NP\phi)^{N-1} = (1 + P\phi(N - \phi))^N \tag{3}$$

in the interval $[1, N]$.

The proof of Theorem 1 has several parts. The converse is proved in Section III. The proof of achievability follows by [8, Theorem 2] and can be proved based on Kramer's linear-feedback code [8]. For completeness, we present a simple description and analysis of Kramer's code in Section IV. Finally, the property that (3) has a unique solution in $[1, N]$ is proved in Appendix A.

Remark 1: Kramer showed [8] that when the power constraint P exceeds the threshold $P_c(N)$, which is the unique positive solution to

$$(1 + N^2P/2)^{N-1} = (1 + N^2P/4)^N \tag{4}$$

then the sum-capacity $C(N, P)$ is given by the right-hand side (RHS) of (2). Thus, for this case Theorem 1 follows directly from Kramer's more general result. Consequently, when $P \geq P_c(N)$, then the linear-feedback sum-capacity coincides with the sum-capacity, i.e., $C_{\mathrm{L}}(N, P) = C(N, P)$. It is not known whether this equality holds for all powers P; see also our discussion in Section V-B.

Remark 2: Since $\phi(N, P) \in [1, N]$, we can define a parameter $\rho \in [0, 1]$ so that $\phi(N, P) = 1 + (N - 1)\rho$. Intuitively, ρ measures the correlation between the transmitted signals. For example, when $N = 2$, the corresponding ρ coincides with the optimal correlation coefficient ρ^* in [2]. Thus, $\phi(N, P) \in [1, N]$ captures the amount of cooperation (coherent power gain) that can be established among the senders using linear-feedback codes, where $\phi = 1$ corresponds to no cooperation and $\phi = N$ corresponds to full cooperation. For a fixed $N \geq 2$, $\phi(N, P)$ is strictly increasing (see Appendix A); thus, more power allows for more cooperation. Moreover, $\phi(N, P) \to 1$ as $P \to 0$ and $\phi(N, P) \to N$ as $P \to \infty$, which is seen as follows. We rewrite (3) as

$$\left(1 + \frac{P\phi^2}{1 + P\phi(N - \phi)}\right)^{N-1} = 1 + P\phi(N - \phi) \tag{5}$$

and notice that the left-hand side (LHS) of (5) can be written as $1 + P\phi^2(N - 1) + o(P)$, where $o(P)$ tends to 0 faster than P. Thus, the LHS of (5) can equal its RHS only if $\phi^2(N - 1) - \phi(N - \phi) \to 0$ as $P \to 0$, or equivalently, $\phi(N, P) \to 1$ as $P \to 0$. On the other hand, as $P \to \infty$, the LHS tends to a constant while the RHS tends to infinity unless $N - \phi$ tends to 0. Thus, by contradiction, $\phi(N, P) \to N$ as $P \to \infty$.

By the above observation, we have the following two corollaries to Theorem 1 for the low and high signal-to-noise ratio (SNR) regimes.

Corollary 1: In the low SNR regime, almost no cooperation is possible and the linear-feedback sum-capacity approaches the sum-capacity without feedback

$$\lim_{P \to 0} \left(C_{\mathrm{L}}(N, P) - \frac{1}{2}\log(1 + NP) \right) = 0.$$

Corollary 2: In the high SNR regime, the linear-feedback sum-capacity approaches the sum-capacity with full cooperation where all the transmitted signals are coherently aligned with combined SNR equal to N^2P:

$$\lim_{P \to \infty} \left(C_{\mathrm{L}}(N, P) - \frac{1}{2}\log(1 + N^2P) \right) = 0.$$

III. Proof of the Converse

In this section, we show that under the symmetric block power constraints P, the linear-feedback sum-capacity $C_{\mathrm{L}}(N, P)$ is upper bounded as

$$C_{\mathrm{L}}(N, P) \leq \frac{1}{2}\log(1 + NP\phi(N, P)) \tag{6}$$

where $\phi(N, P) \in [1, N]$ is defined in (3).

The proof involves five steps. First, we derive an upper bound on the linear-feedback sum-capacity based on Fano's inequality and the maximum entropy property of Gaussian distributions (see Lemma 1). Second, we relax the problem by replacing the functional structure in the optimizing Gaussian input distributions (8) with a dependence balance condition [10], [11], and we rewrite the resulting nonconvex optimization problem as one over positive semidefinite matrices (see Lemma 2). Third, we consider the Lagrange dual function $J(\lambda, \gamma)$, which yields an upper bound on $C_{\mathrm{L}}(N, P)$ for every $\lambda, \gamma \geq 0$ (see Lemma 3). Fourth, by exploiting the convexity and symmetry of the problem, we simplify the upper bound $J(\lambda, \gamma)$ into an unconstrained optimization problem (which is still nonconvex) that involves only two optimization variables (see Lemma 4). Fifth and last, using brute-force calculus and strong duality, we show that there exist $\lambda^*, \gamma^* \geq 0$ such that the corresponding upper bound $J(\lambda^*, \gamma^*)$ coincides with the RHS of (6) (see Lemma 5).

The details are as follows.

Lemma 1: The linear-feedback sum-capacity $C_{\mathrm{L}}(N, P)$ is upper bounded as

$$C_{\mathrm{L}}(N,P) \leq \limsup_{n\to\infty} C_{\mathrm{L}}^{(n)}(P)$$

where[1]

$$C_{\mathrm{L}}^{(n)}(P) := \max \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},\ldots,X_{Ni};Y_i \,|\, Y^{i-1}) \tag{7}$$

and the maximum is over all inputs X_{ji} of the form

$$X_{ji} = L_{ji}(V_{ji}, Y^{i-1}), \quad i=1,\ldots,n,\ j=1,\ldots,N \tag{8}$$

such that the function L_{ji} is linear, the vector $\mathbf{V}_j \in \mathbb{R}^n \sim \mathrm{N}(0, K_{\mathbf{V}_j})$ is Gaussian, independent of the noise vector Z^n and the tuple $(\mathbf{V}_{j'} : j' \neq j)$, and the power constraint $\sum_{i=1}^n \mathsf{E}(X_{ji}^2) \leq nP$ is satisfied.

Proof: By Fano's inequality [16]

$$H(M(\mathcal{S}) \,|\, Y^n) \leq 1 + nP_e^{(n)} \sum_{j=1}^{N} R_j =: n\epsilon_n$$

for some ϵ_n that tends to zero along with $P_e^{(n)}$ as $n \to \infty$. Thus, for any achievable rate tuple $(R_1, \ldots, R_N)$, the sum-rate R can be upper bounded as follows:

$$\begin{aligned} nR &= n\sum_{j=1}^{N} R_j \\ &= H(M(\mathcal{S})) \\ &\leq I(M(\mathcal{S}); Y^n) + n\epsilon_n && (9)\\ &\leq I(\mathbf{\Theta}(\mathcal{S}); Y^n) + n\epsilon_n && (10)\\ &\leq \sum_{i=1}^{n} I(X_i(\mathcal{S}); Y_i \,|\, Y^{i-1}) + n\epsilon_n && (11) \end{aligned}$$

where (10) and (11) follow by the data processing inequality and the memoryless property of the channel, respectively. Therefore, the linear-feedback sum-capacity is upper bounded as

$$C_{\mathrm{L}}(N,P) \leq \limsup_{n\to\infty} \frac{1}{n} \max \sum_{i=1}^{n} I(X_i(\mathcal{S}); Y_i \,|\, Y^{i-1}), \tag{12}$$

where the maximum is over all input distributions induced by a linear-feedback code satisfying the symmetric power constraints P, i.e., over all choices of independent random vectors $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_N$ and linear functions L_{ji} such that the inputs $X_{ji} = L_{ji}(\mathbf{\Theta}_j, Y^{i-1})$ satisfy the power constraints $\sum_{i=1}^n \mathsf{E}(X_{ji}^2) \leq nP$. Now let

$$\tilde{\mathbf{V}}_j \sim \mathrm{N}(0, K_{\mathbf{\Theta}_j}), \quad j=1,\ldots,N$$

be a Gaussian random vector with the same covariance matrix as $\mathbf{\Theta}_j$, independent of $(\tilde{\mathbf{V}}_{j'} : j' \neq j)$. Using the same linear functions L_{ji} as in the given code, define

$$\tilde{X}_{ji} = L_{ji}(\tilde{\mathbf{V}}_j, \tilde{Y}^{i-1}) \tag{13}$$

where $\tilde{Y}_i$ is the channel output of a Gaussian MAC corresponding to the input tuple $\tilde{X}_i(\mathcal{S})$. It is not hard to see that $(\tilde{X}_i(\mathcal{S}), \tilde{Y}^i)$ is jointly Gaussian with zero mean and of the same covariance matrix as $(X_i(\mathcal{S}), Y^i)$. Therefore, by the conditional maximum entropy theorem [5, Lemma 1] we have

$$I(X_i(\mathcal{S}); Y_i \,|\, Y^{i-1}) \leq I(\tilde{X}_i(\mathcal{S}); \tilde{Y}_i \,|\, \tilde{Y}^{i-1}). \tag{14}$$

Combining (12) and (14) and appropriately defining $\mathbf{V}_j$ in (8) from $\tilde{\mathbf{V}}_j$ in (13) completes the proof of Lemma 1. ■

We define the following functions on N-by-N covariance matrices K:

$$f_1(K) = \frac{1}{2}\log\left(1 + \sum_{j,j'} K_{jj'}\right) \tag{15a}$$

$$f_2(K) = \frac{1}{2(N-1)} \sum_{j=1}^{N} \log\left[1 + \sum_{j',j''} K_{j'j''} - \frac{\left(\sum_{j'} K_{jj'}\right)^2}{K_{jj}}\right]. \tag{15b}$$

It can be readily checked that both functions are concave in K (see Appendix B).

Lemma 2: The linear-feedback sum-capacity $C_{\mathrm{L}}(N, P)$ is upper bounded as

$$C_{\mathrm{L}}(N,P) \leq \limsup_{n\to\infty} \max_{K_1,\ldots,K_N} \frac{1}{n}\sum_{i=1}^{n} f_1(K_i) \tag{16}$$

where the maximum is over N-by-N covariance matrices $\{K_i \succeq 0\}_{i=1}^n$ such that

$$\sum_{i=1}^{n} (K_i)_{jj} \leq nP, \quad j=1,\ldots,N \tag{17}$$

$$\sum_{i=1}^{n} f_1(K_i) - f_2(K_i) \leq 0. \tag{18}$$

Proof: Since X_{ji} is defined by the (causal) functional relationship in (8), by [10], [11, Theorem 1] we have the dependence balance condition

$$\sum_{i=1}^{n} I(X_i(\mathcal{S}); Y_i \,|\, Y^{i-1}) \leq \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{N} I(X_i(\mathcal{S}\backslash\{j\}); Y_i \,|\, Y^{i-1}, X_{ji}). \tag{19}$$

[1]For simplicity of notation we do not include the parameter N explicitly in most functions that we define in this section, e.g., $C_{\mathrm{L}}^{(n)}(P)$.

Furthermore, recall that $(X^n(\mathcal{S}), Y^n)$ is jointly Gaussian. Therefore, for every $i \in \{1,\ldots,n\}$, conditioned on $Y_{i-1} = y^{i-1}$, the input (column) vector $\mathbf{X}_i := (X_{1i},\ldots,X_{Ni})$ is zero-mean Gaussian with covariance matrix

$$K_i := K_{\mathbf{X}_i} - K_{\mathbf{X}_i Y^{i-1}} K_{Y^{i-1}}^{-1} K_{Y^{i-1}\mathbf{X}_i} \succeq 0$$

irrespective of y^{i-1}. Now consider

$$\begin{aligned} I(X_i(\mathcal{S}); Y_i \,|\, Y^{i-1}) &= h(Y_i \,|\, Y^{i-1}) - h(Z_i) \\ &= \frac{1}{2}\log(\mathrm{Var}(Y_i \,|\, Y^{i-1})) \\ &= \frac{1}{2}\log\left(1 + \sum_{j,j'} (K_i)_{jj'}\right) \\ &= f_1(K_i). \end{aligned} \tag{20}$$

Also consider

$$\mathrm{Var}(Y_i \,|\, X_{ji}, Y^{i-1}) = 1 + \sum_{j',j''} (K_i)_{j'j''} - \frac{\left(\sum_{j'} (K_i)_{jj'}\right)^2}{(K_i)_{jj}}$$

which implies that

$$\frac{1}{N-1}\sum_{j=1}^{N} I(X_i(\mathcal{S}\backslash\{j\}); Y_i \,|\, Y^{i-1}, X_{ji}) = f_2(K_i). \tag{21}$$

Hence, (19) reduces to (18). Rewriting (7) in terms of covariance matrices K_i via (20) and relaxing the functional relationship (8) by the dependence balance condition (18) completes the proof of Lemma 2. ■

Remark 3: Although both functions $f_1(K)$ and $f_2(K)$ are concave, their difference $f_2(K) - f_1(K)$ is neither concave nor convex. Hence, the optimization problem in (16) is nonconvex.

Lemma 3: Let $f_1(K)$ and $f_2(K)$ be defined as in (15a) and (15b). Then for every $\lambda, \gamma \geq 0$

$$C_{\mathrm{L}}(N,P) \leq J(\lambda,\gamma) \tag{22}$$

where

$$J(\lambda,\gamma) := \max_{K\succeq 0}\left[(1-\gamma)f_1(K) + \gamma f_2(K) + \lambda\sum_{j=1}^{N}(P - K_{jj})\right]. \tag{23}$$

Proof: By the standard Lagrange duality [12], for any $\lambda_1,\ldots,\lambda_N,\gamma \geq 0$, the maximum in (16) is upper bounded as

$$\begin{aligned} &\max_{K_1,\ldots,K_N} \frac{1}{n}\sum_{i=1}^{n} f_1(K_i) \\ &\leq \max_{K_1,\ldots,K_N} \frac{1}{n}\sum_{i=1}^{n}\left[f_1(K_i) + \gamma(f_2(K_i) - f_1(K_i)) + \sum_{j=1}^{N}\lambda_j(P - (K_i)_{jj})\right] \end{aligned}$$

where the maximum is over $K_1,\ldots,K_N \succeq 0$ (without any other constraints). Here, $\lambda_1,\ldots,\lambda_N \geq 0$ are the Lagrange multipliers corresponding to the power constraints (17) and $\gamma \geq 0$ is the Lagrange multiplier corresponding to the dependence balance constraint (18). Finally, we choose $\lambda_1 = \cdots = \lambda_N = \lambda$, which yields

$$\begin{aligned} &\max_{K_1,\ldots,K_N} \frac{1}{n}\sum_{i=1}^{n} f_1(K_i) \\ &\leq \max_{K_1,\ldots,K_N} \frac{1}{n}\sum_{i=1}^{n}\left[f_1(K_i) + \gamma(f_2(K_i) - f_1(K_i)) + \sum_{j=1}^{N}\lambda(P - (K_i)_{jj})\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\max_{K_i}\left[f_1(K_i) + \gamma(f_2(K_i) - f_1(K_i)) + \sum_{j=1}^{N}\lambda(P - (K_i)_{jj})\right] \\ &= J(\lambda,\gamma) \end{aligned}$$

and completes the proof of Lemma 3. ■

Lemma 4: For every $\lambda, \gamma \geq 0$,

$$J(\lambda,\gamma) = \max_{x\geq 0}\max_{0\leq\phi\leq N} g(\gamma,x,\phi) + \lambda N(P - x) \tag{24}$$

where

$$g(\gamma,x,\phi) := (1-\gamma)C_1(x,\phi) + \gamma C_2(x,\phi) \tag{25}$$

and

$$C_1(x,\phi) := \frac{1}{2}\log(1 + Nx\phi) \tag{26a}$$

$$C_2(x,\phi) := \frac{N}{2(N-1)}\log(1 + (N-\phi)x\phi). \tag{26b}$$

Proof: Suppose that a covariance matrix K attains the maximum in (23). For each permutation π on $\{1,\ldots,N\}$, let $\pi(K)$ be the covariance matrix obtained by permuting the rows and columns of K according to π, i.e., $(\pi(K))_{jj'} = K_{\pi(j)\pi(j')}$ for $j, j' \in \{1,\ldots,N\}$. Let

$$\bar{K} := \frac{1}{N!}\sum_{\pi}\pi(K)$$

be the arithmetic average of K over all $N!$ permutations. Clearly, $\bar{K}$ is positive semidefinite and of the form

$$\bar{K} = x\cdot\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \tag{27}$$

for some $x \geq 0$ and $-1/(N-1) \leq \rho \leq 1$. (The conditions on x and ρ assure that $\bar{K}$ is positive semidefinite.) We now show that also $\bar{K}$ attains the maximum in (23). First, notice that the function $f_1(K)$ depends on the matrix K only via the sum of its entries and hence

$$f_1(K) = f_1(\pi(K)) = f_1(\bar{K}).$$

Similarly

$$\sum_{j=1}^{N} K_{jj} = \sum_{j=1}^{N} (\pi(K))_{jj} = \sum_{j=1}^{N} \bar{K}_{jj}.$$

Also, by symmetry we have $f_2(K) = f_2(\pi(K))$. Hence, by the concavity of $f_2(K)$ (see Appendix B) and Jensen's inequality, $f_2(K) \leq f_2(\bar{K})$. Therefore

$$\begin{aligned}(1-\gamma)f_1(K) + \gamma f_2(K) + \lambda \sum_{j=1}^{N}(P - K_{jj}) \\ \leq (1-\gamma)f_1(\bar{K}) + \gamma f_2(\bar{K}) + \lambda \sum_{j=1}^{N}(P - \bar{K}_{jj})\end{aligned}$$

and the maximum of (23) is also attained by $\bar{K}$. Finally, defining $\phi := 1 + (N-1)\rho \in [0, N]$ and simplifying (15a) and (15b) yields

$$\begin{aligned}f_1(\bar{K}) &= C_1(x, \phi) \\ f_2(\bar{K}) &= C_2(x, \phi)\end{aligned}$$

which completes the proof of Lemma 4. ■

Remark 4: The symmetric $\bar{K}$ in (27) was also considered in [5] and [8] to evaluate the cutset upper bound, which corresponds to taking $\gamma \leq 1$.

Lemma 5: There exist $\lambda^*, \gamma^* \geq 0$ such that

$$\begin{aligned}J(\lambda^*, \gamma^*) &\leq C_1(P, \phi(N, P)) \\ &= \frac{1}{2}\log(1 + NP\phi(N, P))\end{aligned}$$

where $\phi(N, P)$ is defined in (3).

Proof: Consider the optimization problem over (x, ϕ), which defines $J(\lambda, \gamma)$ in (24). Note that $g(\gamma, x, \phi)$ given by (25) is neither concave or convex in (x, ϕ) for $\gamma > 1$. However, $g(\gamma, x, \phi)$ is concave in $\phi \geq 0$ for fixed $x, \gamma \geq 0$ as shown in Appendix C.

Let $\phi^* = \phi^*(\gamma, x)$ be the unique nonnegative solution to

$$\left.\frac{\partial g(\gamma, x, \phi)}{\partial \phi}\right|_{\phi=\phi^*} = 0$$

or equivalently to

$$\frac{(1-\gamma)(N-1)}{1 + Nx\phi^*} = \frac{\gamma(2\phi^* - N)}{1 + x\phi^*(N - \phi^*)}. \tag{28}$$

(That such a unique solution exists is easily verified considering the equivalent quadratic equation; see (70) in Appendix D.) Then, by the concavity of $g(\gamma, x, \phi)$ in ϕ for fixed γ and x

$$\begin{aligned}J(\lambda, \gamma) &= \max_{x \geq 0} \max_{0 \leq \phi \leq N} g(\gamma, x, \phi) + \lambda N(P - x) \\ &\leq \max_{x \geq 0} g(\gamma, x, \phi^*(\gamma, x)) + \lambda N(P - x) \end{aligned} \tag{29}$$

for any $\gamma \geq 0$. (The inequality follows because $\phi^*(\gamma, x)$ might be larger than N.)

Now let $g^*(\gamma, x) = g(\gamma, x, \phi^*(\gamma, x))$. Then, $g^*(\gamma, x)$ is nondecreasing and concave in x for fixed γ as shown in Appendix D. Thus

$$\begin{aligned}\min_{\lambda \geq 0} J(\lambda, \gamma) &\leq \min_{\lambda \geq 0} \max_{x} g^*(\gamma, x) + \lambda N(P - x) \\ &= \max_{x \leq P} g^*(\gamma, x) \\ &= g^*(\gamma, P)\end{aligned} \tag{30}$$

where the first equality follows by Slater's condition [12] and strong duality, and the last equality follows by the monotonicity of $g^*(\gamma, x)$ in x. Alternatively, the equality in (30) can be viewed as the complementary slackness condition [12]. Indeed, since $g^*(\gamma, x)$ is not bounded from above, the optimal Lagrange multiplier $\lambda^* > 0$ must be positive. Therefore, the corresponding constraint $x \leq P$ is active at the optimum, i.e., $x^* = P$.

Finally, we choose $\gamma = \gamma^*$, where

$$\gamma^* = \left(1 - \frac{(N - 2\phi(N,P))(1 + NP\phi(N,P))}{(N-1)(1 + P\phi(N,P)(N - \phi(N,P)))}\right)^{-1}$$

which assures that $\phi^*(\gamma^*, P)$ coincides with $\phi(N, P)$; see (28). Since γ^* is nonnegative by (57) in Appendix A and thus is a valid choice

$$\begin{aligned}g^*(\gamma^*, P) &= g(\gamma^*, P, \phi(N, P)) \\ &= (1-\gamma^*)C_1(P, \phi(N, P)) + \gamma^* C_2(P, \phi(N, P)) \\ &= C_1(P, \phi(N, P))\end{aligned}$$

which, combined with (30), concludes the proof of Lemma 5 and of the converse. ■

IV. Achievability via Kramer's Code

We present (a slightly modified version of) Kramer's linear-feedback code and analyze it based on the properties of discrete algebraic Riccati equations (DARE). In particular, we establish the following:

Theorem 2: Suppose that $\beta_1, \ldots, \beta_N > 1$ are real numbers and $\omega_1, \ldots, \omega_N$ are distinct complex numbers on the unit circle. Let $A = \mathrm{diag}(\beta_1\omega_1, \ldots, \beta_N\omega_N)$ be a diagonal matrix, $\mathbf{1} = (1, \ldots, 1)$ be the all-one column vector, and K^* be the unique positive-definite solution to the discrete algebraic Riccati equation (DARE)

$$K = AKA' - (AK\mathbf{1})(1 + \mathbf{1}'K\mathbf{1})^{-1}(AK\mathbf{1})'. \tag{31}$$

Then, a rate tuple $(R_1, \ldots, R_N)$ is achievable under power constraints $(P_1, \ldots, P_N)$, provided that $R_j < \log \beta_j$ and $P_j > K^\star_{jj}$, $j = 1, \ldots, N$.

Achievability of Theorem 1 will be proved in Subsection IV-C as a corollary to Theorem 2.

A. Kramer's Linear-Feedback Code

Following [8], we represent a pair of consecutive uses of the given real Gaussian MAC as a single use of a complex Gaussian MAC. We represent the message point of sender j by the complex scalar Θ_j (corresponding to $k = 2$ in the original real channel) and let $\mathbf{\Theta} := (\Theta_1, \dots, \Theta_N)$ be the (column) vector of message points.

The coding scheme has the following parameters: real coefficients $\beta_1, \dots, \beta_N > 1$ and distinct complex numbers $\omega_1, \dots, \omega_N$ on the unit circle.

Nonfeedback Mappings: For $j = 1, \dots, N$, we divide the square with corners at $\{\pm 1 \pm \sqrt{-1}\}$ on the complex plane into 2^{2nR_j} equal subsquares. We then assign a different message $m_j \in [1 : 2^{2nR_j}]$ to each subsquare and denote the complex number in the center of the subsquare by $\theta_j(m_j)$. The message point Θ_j of sender j is then $\Theta_j = \theta_j(M_j)$.

Linear-Feedback Mappings: Let $\mathbf{X}_i := (X_{1i}, \dots, X_{Ni})$ denote the (column) vector of channel inputs at time i. We use the linear-feedback mappings

$$\begin{aligned}\mathbf{X}_1 &= \mathbf{\Theta}, \\ \mathbf{X}_i &= A \cdot (\mathbf{X}_{i-1} - \hat{\mathbf{X}}_{i-1}(Y_{i-1})), \quad i > 1 \end{aligned} \tag{32}$$

where

$$A = \mathrm{diag}(\beta_1\omega_1, \beta_2\omega_2, \dots, \beta_N\omega_N) \tag{33}$$

is a diagonal matrix with $A_{jj} = \beta_j\omega_j$ and

$$\hat{\mathbf{X}}_{i-1}(Y_{i-1}) = \frac{\mathsf{E}(\mathbf{X}_{i-1}Y'_{i-1})}{\mathsf{E}(|Y_{i-1}|^2)} Y_{i-1}$$

is the linear minimum mean squared error (MMSE) estimate of $\mathbf{X}_{i-1}$ given Y_{i-1}.

Decoding: Upon receiving Y^n, the decoder forms a message estimate vector

$$\hat{\mathbf{\Theta}} := (\hat{\Theta}_1, \dots, \hat{\Theta}_N) = \sum_{i=0}^{n-1} A^{-i}\hat{\mathbf{X}}_i \tag{34}$$

and for each $j = 1, \dots, N$ chooses $\hat{M}_j$ such that $\theta_j(\hat{M}_j)$ is the center point of the subsquare containing $\hat{\Theta}_j$.

B. Analysis of the Probability of Error

Our analysis is based on the following auxiliary lemma. We use the short-hand notation $K_i = K_{\mathbf{X}_i}$.

Lemma 6:

$$\lim_{n\to\infty} K_n = K^\star \tag{35}$$

where $K^\star$ is the unique positive-definite solution to the DARE (31).

Proof: We rewrite the channel outputs in (1) as

$$Y_i = \mathbf{1}'\mathbf{X}_i + Z_i. \tag{36}$$

From (32) we have

$$K_{i+1} = AK_{\mathbf{X}_i - \hat{\mathbf{X}}_i}A' \tag{37}$$

where $K_{\mathbf{X}_i - \hat{\mathbf{X}}_i} = K_{\mathbf{X}_i} - K_{\mathbf{X}_iY_i}K_{Y_i}^{-1}K'_{\mathbf{X}_iY_i}$ is the error covariance matrix of the linear MMSE estimate of $\mathbf{X}_i$ given Y_i. Combining (36) and (37) we obtain the Riccati recursion [17]

$$K_{i+1} = AK_iA' - (AK_i\mathbf{1})(1 + \mathbf{1}'K_i\mathbf{1})^{-1}(AK_i\mathbf{1})' \tag{38}$$

for $i = 1, \dots, n-1$. Since A has no unit-circle eigenvalue and the pair $(A, \mathbf{1})$ is detectable,[2] we use Lemma 2.5 in [19] to conclude (35). ■

We now prove that Kramer's code achieves any rate tuple $(R_1, \dots, R_N)$ such that

$$R_j < \log\beta_j, \quad j = 1, \dots, N. \tag{39}$$

Define the difference vector $\mathbf{D}_n := \mathbf{\Theta} - \hat{\mathbf{\Theta}}_n$. Since the minimum distance between message points is $\Delta = 2 \cdot 2^{-nR_j}$, by the union of events bound and the Chebyshev inequality, the probability of error of Kramer's code is upper bounded as

$$\begin{aligned} P_e^{(n)} &\le \mathsf{P}\left(\bigcup_j \{|\mathbf{D}_n(j)| > \Delta/2\}\right) \\ &\le \sum_{j=1}^{N} 2^{2nR_j}\mathsf{E}(|\mathbf{D}_n(j)|^2). \end{aligned} \tag{40}$$

Rewriting the encoding rule in (32) as

$$\mathbf{X}_n = A^n\mathbf{\Theta} - \sum_{i=0}^{n-1} A^{n-i}\hat{\mathbf{X}}_i$$

and comparing it with the decoder's estimation rule in (34) we have $\mathbf{D}_n = A^{-n}\mathbf{X}_n$. Hence, $K_{\mathbf{D}_n} = A^{-n}K_n(A')^{-n}$ with diagonal elements $\mathsf{E}(|\mathbf{D}_n(j)|^2) = \beta_j^{-2n}K_n(j,j)$ and (40) can be written as

$$P_e^{(n)} \le \sum_{j=1}^{N} K_n(j,j) \cdot 2^{2n(R_j - \log\beta_j)}. \tag{41}$$

But by Lemma 6, $\limsup_{n\to\infty} K_n(j,j) < \infty$. Therefore, $P_e^{(n)} \to 0$ as $n \to \infty$.

Finally, by Lemma 6 and the Césaro mean lemma [20], the asymptotic power of sender j satisfies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}(X_{ji}^2) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(K_i)_{jj} = K_{jj}^\star$$

Hence, Kramer's code satisfies the power constraints $P_1, \dots, P_N$ for sufficiently large n, provided that

$$K_{jj}^\star < P_j, \quad j = 1, \dots, N. \tag{42}$$

This completes the proof of Theorem 2.

[2]A pair $(A, \mathbf{b})$ is said to be detectable if there exists a column vector $\mathbf{c}$ such that all the eigenvalues of $A - \mathbf{b}\mathbf{c}'$ lie inside the unit circle. For a diagonal matrix $A = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$, the pair $(A, \mathbf{1})$ is detectable if and only if all the unstable eigenvalues λ_j, i.e., the ones on or outside the unit-circle, are distinct [18, Appendix C].

C. Achievability Proof of Theorem 1

Fix any $\beta > 1$ such that

$$N \log \beta < C_{\mathrm{L}}(P, N), \tag{43}$$

and choose

$$\beta_j = \beta, \tag{44a}$$
$$\omega_j = e^{2\pi\sqrt{-1}\frac{(j-1)}{N}} \tag{44b}$$

for $j = 1, \ldots, N$. Under this choice of parameters, by Theorem 2, Kramer's code achieves any sum-rate $R < N \log \beta < C_{\mathrm{L}}(P, N)$ provided that (42) holds. To show (42) we use the following lemma (see Appendix E for a proof).

Lemma 7: When A is defined through (33) and (44), then the unique positive-definite solution $K^\star$ to the DARE (31) is circulant with all real eigenvalues satisfying $\lambda_j = \lambda_{j-1}/\beta^2$, $j = 2, \ldots, N$, and with the largest eigenvalue λ_1 satisfying

$$1 + N\lambda_1 = \beta^{2N} \tag{45}$$
$$1 + \lambda_1 \left(N - \frac{\lambda_1}{K^\star_{jj}} \right) = \beta^{2(N-1)}. \tag{46}$$

Now by the lemma and the standing assumption (43) on β, we have

$$\begin{aligned} \frac{1}{2}\log(1 + N\lambda_1) &< C_{\mathrm{L}}(N, P) \\ &= \frac{1}{2}\log\left(1 + NP\phi(N, P)\right). \end{aligned}$$

Thus

$$\lambda_1 < P\phi(N, P). \tag{47}$$

On the other hand, from (45) and (46) we have

$$(1 + N\lambda_1)^{N-1} = \left(1 + \lambda_1 \left(N - \frac{\lambda_1}{K^\star_{jj}}\right)\right)^N.$$

Hence, by the definition of the function $\phi(N, P)$ in (3),

$$\lambda_1 = K^\star_{jj}\phi(N, K^\star_{jj}). \tag{48}$$

Combining (47) and (48), we obtain $K^\star_{jj}\phi(N, K^\star_{jj}) < P\phi(N, P)$. Finally, by the monotonicity of $\phi(N, \cdot)$ (see Appendix A), we conclude that $K^*_{jj} < P$, $j = 1, \ldots, N$, which completes the achievability proof of Theorem 1.

V. Discussion

In this paper, we established the linear-feedback sum-capacity $C_{\mathrm{L}}(N, P)$ for symmetric power constraints P. Below, we discuss the complications in extending our proof technique to establish the linear-feedback sum-capacity under asymmetric power constraints or the sum-capacity $C(N, P)$.

A. General Power Constraints

The main difficulty in generalizing our proof to asymmetric power constraints $(P_1, \ldots, P_N)$ lies in extending Lemma 4. The proof of Lemma 4 heavily relies on the fact that covariance matrices of the form (27) are optimal for the optimization problem in (23). This allows us to reduce the optimization problem (23) over covariance matrices to the much simpler optimization problem in (24) over only two variables x and ϕ. However, covariance matrices of the form (27) are not necessarily optimal for the equivalent optimization problem under asymmetric power constraints.

B. Sum-Capacity

It is commonly believed that under symmetric power constraints the linear-feedback sum-capacity generally equals the sum-capacity, i.e., $C(N, P) = C_{\mathrm{L}}(N, P)$ for all values of P and N (cf. [11]). However, currently a proof is only known when the power constraint P is larger than a certain threshold—the unique positive solution to (4)—that depends on N [8]. The main difficulty in establishing this conjecture for all values of $P \geq 0$ lies in proving that Lemma 1 also holds for $C(N, P)$. The rest of the proof remains valid even for arbitrary (nonlinear) feedback codes.

Below, we provide an observation based on the properties of Hirschfeld-Gebelein-Rényi maximal correlation [14], which further supports the conjecture that $C(N, P) = C_{\mathrm{L}}(N, P)$.

C. Greedy Optimality of Linear-Feedback Codes

Let

$$C^{(n)}(P) := \max \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, \ldots, X_{Ni}; Y_i \,|\, Y^{i-1}) \tag{49}$$

where the maximum is over the set of *arbitrary* functions $\{X_{ji}(V_j, Y^{i-1})\}$ satisfying the symmetric block power constraint P and $V_1, \ldots, V_N$ are independent standard (real) Gaussian random variables. As shown in Appendix G, the sum-capacity is upper bounded as

$$C(P) \leq \limsup_{n\to\infty} C^{(n)}(P). \tag{50}$$

We introduce a new notion of conditional maximum correlation to show that for every n linear functions are *greedy* optimal for the optimization problem defining $C^{(n)}(P)$ in (49).

Recall that the maximal correlation $\rho^*(V_1, V_2)$ between two random variables V_1 and V_2 is defined [14] as

$$\rho^*(V_1, V_2) := \sup_{g_1, g_2} \mathsf{E}(g_1(V_1)g_2(V_2)) \tag{51}$$

where the supremum is over all functions $g_1(v_1)$ and $g_2(v_2)$ such that $\mathsf{E}(g_1(V_1)) = \mathsf{E}(g_2(V_2)) = 0$ and $\mathsf{E}(g_1^2(V_1)) = \mathsf{E}(g_2^2(V_2)) = 1$. We extend this notion of maximal correlation to a conditional one. The *conditional maximal correlation* between V_1 and V_2 given another random variable (or vector) Y is defined as

$$\rho^*(V_1, V_2 \,|\, Y) := \sup_{g_1, g_2} \mathsf{E}(g_1(V_1, Y)g_2(V_2, Y)) \tag{52}$$

where the supremum is over all functions $g_1(v_1, y)$ and $g_2(v_2, y)$ such that $\mathsf{E}(g_1(V_1, Y) \,|\, Y) = \mathsf{E}(g_2(V_2, Y) \,|\, Y) = 0$ and $\mathsf{E}(g_1^2(V_1, Y) \,|\, Y) = \mathsf{E}(g_2^2(V_2, Y) \,|\, Y) = 1$ almost surely. The assumption that $g_1(V_1, Y)$ and $g_2(V_2, Y)$ are orthogonal to

Y is crucial; otherwise, both could be chosen as functions only of Y and $\rho^*(V_1, V_2|Y) = 1$ trivially.

Let $\rho(V_1, V_2)$ denote the correlation between V_1 and V_2. We define the (expected) *conditional correlation* between V_1 and V_2 given Y as

$$\rho(V_1, V_2 \,|\, Y) := \int \rho(V_1, V_2 \,|\, Y = y)\, dF(y),$$

where $\rho(V_1, V_2 \,|\, Y = y)$ denotes the correlation between V_1 and V_2 conditioned on $Y = y$. It can be shown (see Appendix F) that if (V_1, V_2, Y) is jointly Gaussian, then

$$\rho^*(V_1, V_2 \,|\, Y) = \rho(V_1, V_2 \,|\, Y)$$

and linear functions g_1 and g_2 of the form

$$g_j(V_j, Y) = \frac{V_j - \mathsf{E}(V_j|Y)}{\sqrt{\mathsf{E}((V_j - \mathsf{E}(V_j \,|\, Y))^2)}}, \quad j = 1, 2, \tag{53}$$

attain $\rho^*(V_1, V_2 \,|\, Y)$.

Back to our discussion on $C^{(n)}(P)$, consider the case $N = 2$ for simplicity. Then, $C^{(n)}(P)$ is upper bounded (see Appendix H) by

$$\max_{\{P_{ji}\}} \max \frac{1}{2n} \sum_{i=1}^{n} \log(1 + P_{1i} + P_{2i} + 2\sqrt{P_{1i}P_{2i}}\, \rho(\tilde{X}_{1i}, \tilde{X}_{2i})) \tag{54}$$

where $\tilde{X}_{ji} = X_{ji} - \mathsf{E}(X_{ji} \,|\, Y^{i-1})$, $j = 1, 2$, $i = 1, \ldots, n$; the inner maximum is over the set $\{X_{ji}(V_j, Y^{i-1})\}$ satisfying $\mathsf{E}(X_{ji}^2) = P_{ji}$; and the outer maximum is over the set $\{P_{ji}\}$ satisfying $\sum_{i=1}^{n} P_{ji} \leq nP$. Suppose that linear functions $X_{ji} = L_{ji}(V_j, Y^{i-1})$ are used up to time $i-1$ and therefore (V_1, V_2, Y^{i-1}) is jointly Gaussian. By definition, $\rho(\tilde{X}_{1i}, \tilde{X}_{2i}) \leq \rho^*(V_1, V_2 \,|\, Y^{i-1})$, which by Appendix F equals $\rho(V_1, V_2 \,|\, Y^{i-1})$ and is attained by linear functions L_{1i} and L_{2i}. In this sense, choosing X_{ji} linear is greedy optimal for the inner maximization in (54). Note that when (V_1, V_2, Y^{i-1}) is jointly Gaussian, then a linear choice of X_{1i} and X_{2i} implies that also (V_1, V_2, Y^i) is jointly Gaussian. This observation, which can be easily extended to any number of senders N, further corroborates the conjecture that $C_{\mathrm{L}}(N, P) = C(N, P)$ for all symmetric power constraints $P \geq 0$.

Incidentally, global optimality of linear-feedback codes of the form $X_{ji} = L_{ji}(V_j, Y^{i-1})$ would also imply that the performance of Kramer's code, which uses complex signaling ($k = 2$), can be achieved by real signaling. In this case, the optimal real signaling would involve nonstationary or cyclostationary operations, because a stationary extension of Ozarow's scheme to $N \geq 3$ senders is strictly suboptimal [21].

Appendix A
Properties of $\phi(N, P)$

We fix the integer $N \geq 2$ and prove that for $P > 0$ the solution $\phi(N, P)$ to (3) is unique and increasing in P. Note that the identity in (3) is equivalent to

$$f(P, \phi) := C_2(P, \phi) - C_1(P, \phi) = 0 \tag{55}$$

where $C_1(P, \phi)$ and $C_2(P, \phi)$ are defined in (26). We prove the uniqueness of $\phi(N, P)$ by showing that $f(P, 1) \geq 0$, $f(P, N) < 0$, and $\partial f(P, \phi)/\partial \phi < 0$ for $1 \leq \phi \leq N$. The fact that $f(P, N) < 0$ is immediate. For $f(P, 1) \geq 0$, note that $(1 - 1/N)^k \geq 1 - k/N$ for $N \geq 1$, or equivalently

$$\binom{N}{k}(N-1)^k \geq \binom{N-1}{k} N^k, \quad 1 \leq k \leq N-1.$$

Thus

$$\sum_{k=1}^{N} \binom{N}{k}(N-1)^k P^k \geq \sum_{k=1}^{N-1} \binom{N-1}{k} N^k P^k \tag{56}$$

which implies that $(N-1)C_2(P, 1) \geq (N-1)C_1(P, 1)$ and thus that $f(P, 1) \geq 0$. The condition $\partial f(P, \phi)/\partial \phi < 0$ is equivalent to

$$\frac{N - 2\phi}{1 + P\phi(N - \phi)} - \frac{N-1}{1 + NP\phi} < 0. \tag{57}$$

Rearranging terms in (57) we have $1 + NP\phi - (2\phi + P\phi^2 + NP\phi^2) < 0$ which holds for all $\phi \geq 1$. This completes the proof of the uniqueness. We next prove the monotonicity of $\phi(N, P)$ in P. By (3), we have

$$\frac{1 + NP\phi}{1 + P\phi(N - \phi)} = (1 + NP\phi)^{1/N}, \tag{58}$$

or equivalently

$$P\phi(N - \phi) = (1 + NP\phi)^{(N-1)/N} - 1. \tag{59}$$

Moreover, since $1 + P\phi > (1 + NP\phi)^{1/N}$ for $N > 1$

$$1 + NP\phi - (1 + NP\phi)^{1/N} > P\phi(N-1). \tag{60}$$

Multiplying (58) by (59) and considering (60), we obtain

$$(N - \phi) \cdot \frac{1 + NP\phi}{1 + P\phi(N - \phi)} > N - 1. \tag{61}$$

From (61), it is straightforward to verify that

$$\left.\frac{\partial f}{\partial P}\right|_{P, \phi(N,P)} > 0. \tag{62}$$

Finally, by differentiating (55), we have

$$\left.\frac{\partial f}{\partial P}\right|_{P, \phi(N,P)} dP + \left.\frac{\partial f}{\partial \phi}\right|_{P, \phi(N,P)} d\phi = 0. \tag{63}$$

Combining (62), (63), and the fact that $\partial f/\partial \phi < 0$ [shown in (57)], we conclude that $d\phi/dP > 0$ for $(P, \phi(N, P))$.

Appendix B
Concavity of $f_1(K)$ and $f_2(K)$

Our proof is based on the following general lemma.

Lemma 8: Let $(\mathbf{U}, \mathbf{V})$ be a Gaussian random vector with covariance matrix $A\Sigma A' + BB'$. Let $f(\Sigma) := h(\mathbf{U} \,|\, \mathbf{V})$. Then, $f(\Sigma)$ is concave in $\Sigma \succeq 0$.

Proof: Fix A and B. Let Σ_1, Σ_2, and $\lambda \in [0,1]$ be given, and $\Sigma := \lambda\Sigma_1 + (1-\lambda)\Sigma_2$. For $q = 1,2$, let $(\mathbf{U}_q, \mathbf{V}_q) \sim \mathrm{N}(0, A\Sigma_q A' + BB')$ and $(\mathbf{U}, \mathbf{V}) \sim \mathrm{N}(0, A\Sigma A' + BB')$, and let Q be a binary random variable with $\mathsf{P}\{Q=1\} = \lambda = 1 - \mathsf{P}\{Q=2\}$. Assume that $(\mathbf{U}_1, \mathbf{V}_1)$, $(\mathbf{U}_2, \mathbf{V}_2)$, and Q are independent. Then

$$\begin{aligned}\lambda f(\Sigma_1) + (1-\lambda) f(\Sigma_2) &= h(\mathbf{U}_Q \,|\, \mathbf{V}_Q, Q) \\ &\le h(\mathbf{U}_Q \,|\, \mathbf{V}_Q) \\ &\le h(\mathbf{U} \,|\, \mathbf{V}) \\ &= f(\Sigma)\end{aligned}$$

where the last inequality follows by the conditional maximum entropy theorem [5, Lemma 1] and the fact that $(\mathbf{U}_Q, \mathbf{V}_Q)$ has the covariance matrix $A\Sigma A' + BB'$. ■

Now let $X(\mathcal{S}) \sim \mathrm{N}(0, K)$ and $Y = \sum_{j=1}^N X_j + Z$, where $Z \sim \mathrm{N}(0,1)$ is independent of $X(\mathcal{S})$. Then

$$\begin{aligned} f_1(K) &= h(Y), \\ f_2(K) &= \frac{1}{N-1}\sum_{j=1}^N h(Y \,|\, X_j)\end{aligned}$$

and the concavity of f_1 and f_2 in K follows immediately from Lemma 8.

APPENDIX C
CONCAVITY OF $g(\gamma, x, \phi)$ IN ϕ

Comparing the definitions of $f_1(K)$ and $f_2(K)$ in (15) with the definitions of $C_1(x,\phi)$ and $C_2(x,\phi)$ in (26), respectively, we see that when K has the symmetric form in (27) with $\rho = \frac{\phi-1}{N-1}$, then $f_1(K) = C_1(x,\phi)$ and $f_2(K) = C_2(x,\phi) = f_2(K)$. We prove in the following that for every $\gamma \ge 0$ the function $(1-\gamma)f_1(K) + \gamma f_2(K)$ is concave in K over the set of positive semi-definite matrices $K \succeq 0$ with fixed diagonal elements. This implies the concavity of $g(\gamma, x, \phi)$ in ϕ for fixed x, γ.

Let $X(\mathcal{S}) \sim \mathrm{N}(0,K)$ and $Y = \sum_{j=1}^N X_j + Z$, where $Z \sim \mathrm{N}(0,1)$ is independent of $X(\mathcal{S})$. Then,

$$\begin{aligned}&(1-\gamma)f_1(K) + \gamma f_2(K) \\ &= (1-\gamma)h(Y) + \frac{\gamma}{N-1}\sum_{j=1}^N h(Y \,|\, X_j) \\ &= (1-\gamma)h(Y) + \frac{\gamma}{N-1}\sum_{j=1}^N (h(Y) + h(X_j \,|\, Y) - h(X_j)) \\ &= h(Y)\left(1 + \frac{\gamma}{N-1}\right) + \frac{\gamma}{N-1}\sum_{j=1}^N h(X_j \,|\, Y) - h(X_j).\end{aligned}$$

By Lemma 8 in Appendix B, $h(Y)$ and $h(X_j \,|\, Y)$ are concave in K. Since $h(X_j) = \frac{1}{2}\log(2\pi e K_{jj})$ depends only on the diagonal elements of K, the claim follows.

APPENDIX D
PROPERTIES OF $g(\gamma, x, \phi^*(\gamma, x))$ IN x

For simplicity, we do not include γ explicitly in our notation: $g(x,\phi) := g(\gamma, x, \phi)$ and $\phi^*(x) := \phi^*(\gamma, x)$. We first show that $g(x, \phi^*(x))$ is monotonically nondecreasing in x. Since $\phi^*(x)$ satisfies (28) and $\left.\frac{\partial g(x,\phi)}{\partial \phi}\right|_{x,\phi^*(x)} = 0$, we obtain

$$\begin{aligned}&\frac{dg(x,\phi^*(x))}{dx} \\ &= \left.\frac{\partial g(x,\phi)}{\partial x} + \frac{\partial g(x,\phi)}{\partial \phi}\,\frac{d\phi}{dx}\right|_{x,\phi^*(x)} \\ &= \left.\frac{\partial g(x,\phi)}{\partial x}\right|_{x,\phi^*(x)} \\ &= \left.\frac{(1-\gamma)N\phi}{2(1+Nx\phi)} + \frac{\gamma N\phi(N-\phi)}{2(N-1)(1+x\phi(N-\phi))}\right|_{x,\phi^*(x)} \\ &= \frac{N(\gamma-1)(\phi^*(x))^2}{2(1+Nx\phi^*(x))(N-2\phi^*(x))} \qquad (64)\\ &\ge 0 \qquad (65)\end{aligned}$$

where (64) follows by (28) and (65) follows since $(\gamma - 1)$ and $(N - 2\phi^*(x))$ have the same sign; see (28). Thus, $g(x, \phi^*(x))$ is nondecreasing in x.

We now show that $g(x,\phi^*(x))$ is concave in x. We first note that for $0 \le \gamma \le 1$ the function $g(x,\phi) = (1-\gamma)C_1(x,\phi) + \gamma C_2(x,\phi)$ is concave in (x,ϕ) because for symmetric matrices K of the form in (27) with $\rho = \frac{\phi-1}{N-1}$ both $C_1(x,\phi) = f_1(K)$ and $C_2(x,\phi) = f_2(K)$ are concave in K (see Appendix B). Thus, for any $\nu \in [0,1]$, $x_1, x_2 > 0$, and $x = \nu x_1 + (1-\nu)x_2$,

$$\begin{aligned}&\nu g(x_1, \phi^*(x_1)) + (1-\nu) g(x_2, \phi^*(x_2)) \\ &\quad \le g(x, \nu\phi^*(x_1) + (1-\nu)\phi^*(x_2)) \qquad (66)\\ &\quad \le g(x, \phi^*(x)) \qquad (67)\end{aligned}$$

where (66) follows by the concavity of $g(x,\phi)$ and (67) follows by the definition of $\phi^*(x)$. This establishes the concavity of $\phi(x,\phi^*(x))$ for $0 \le \gamma \le 1$.

To prove the concavity for $\gamma > 1$, we show that the second derivative $d^2 g(x,\phi^*(x))/dx^2$ is negative. Define

$$h(x,\phi) := \frac{\phi^2}{(1+Nx\phi)(N-2\phi)}. \qquad (68)$$

Then, by (64)

$$\begin{aligned}&\frac{d^2 g(x,\phi^*(x))}{d^2 x} \cdot \frac{2}{N(\gamma-1)} \\ &= \left.\frac{\partial h(x,\phi)}{\partial x}\right|_{x,\phi^*(x)} + \left.\frac{\partial h(x,\phi)}{\partial \phi}\right|_{x,\phi^*(x)} \frac{d\phi^*(x)}{dx} \\ &= \left.\frac{-N\phi^3}{(1+Nx\phi)^2(N-2\phi)}\right|_{x,\phi^*(x)} \\ &\quad + \left.\frac{\phi(N^2x\phi + 2(N-\phi))}{(1+Nx\phi)^2(N-2\phi)^2}\right|_{x,\phi^*(x)} \frac{d\phi^*(x)}{dx} \\ &= \left.\frac{\frac{d\phi^*(x)}{dx}(N^2x\phi + 2(N-\phi)) - N\phi^2(N-2\phi)}{(1+Nx\phi)^2(N-2\phi)^2}\phi\right|_{x,\phi^*(x)}.\end{aligned}$$

Since the denominator and $\phi^*(x)$ are positive, the following inequality concludes the proof of concavity for $\gamma > 1$:

$$\frac{d\phi^*(x)}{dx} < \frac{N\phi^*(x)^2(N - 2\phi^*(x))}{N^2 x\phi^*(x) + 2(N - \phi^*(x))}. \tag{69}$$

We now establish (69). Rearranging terms in (28), we obtain that $\phi^*(x)$ is the solution to the quadratic equation

$$a\phi^2 + b\phi + c = 0, \tag{70}$$

where $a = (N + \gamma - 1 + \gamma N)x$, $b = -N(N + \gamma - 1)x + 2\gamma$, and $c = -(N + \gamma - 1)$. Since $ac < 0$, there is a unique positive solution $\phi^*(x) = (-b + \sqrt{b^2 - 4ac})/2a$. Taking the derivative of (70) with respect to x, we find

$$\begin{aligned}\frac{d\phi^*(x)}{dx} &= \frac{-(\phi^*(x))^2(a'\phi^*(x) + b')}{a(\phi^*(x))^2 - c}\\ &= \frac{N(\phi^*(x))^2(N - \alpha\phi^*(x))}{\alpha N x(\phi^*(x))^2 + N}\end{aligned} \tag{71}$$

where $a' = N + \gamma - 1 + \gamma N$ and $b' = -N(N + \gamma - 1)$ are derivatives of a and b with respect to x, respectively, and $\alpha := 1 + \gamma N/(N + \gamma - 1) \in (2, N+1)$. Note by simple algebra that $a(b'/a')^2 - b(b'/a') + c > 0$. Because $\phi^*(x)$ is the unique positive solution to (70) with $a > 0$, we have $\phi^*(x) < -b'/a'$, or equivalently, $a'\phi^*(x) + b' < 0$ for every $x \geq 0$. Hence, $\phi^*(x)$ is strictly increasing in $x \geq 0$ and $\phi^*(x) > \phi^*(0) = (N + \gamma - 1)/2\gamma$. Therefore

$$\frac{N - (\alpha - 2)\phi^*(x)}{N} < \frac{\alpha\phi^*(x)}{N}. \tag{72}$$

On the other hand, since $\alpha > 2$ and for $q, s > 0$

$$\frac{p}{q} < \frac{r}{s} \quad \text{if and only if} \quad \frac{p}{q} < \frac{p + r}{q + s} \tag{73}$$

we have

$$\frac{N - \alpha\phi^*(x)}{N - 2\phi^*(x)} < \frac{N - (\alpha - 2)\phi^*(x)}{N},$$

which, combined with (72), implies

$$\frac{N - \alpha\phi^*(x)}{N - 2\phi^*(x)} < \frac{\alpha\phi^*(x)}{N} \cdot \frac{Nx\phi^*(x) + 1}{Nx\phi^*(x) + 1}. \tag{74}$$

Applying (73) to (74) once again, we obtain

$$\frac{N - \alpha\phi^*(x)}{N - 2\phi^*(x)} < \frac{\alpha N x(\phi^*(x))^2 + N}{N^2 x\phi^*(x) + 2(N - \phi^*(x))}$$

which, combined with (71), establishes (69).

Appendix E
Proof of Lemma 7

We first show that the circulant matrix K with all real eigenvalues satisfying $\lambda_i = \lambda_{i-1}/\beta^2$ for $i = 2, \ldots, N$, and with λ_1 satisfying (45) is a solution to the DARE (31). We then show that this also implies (46).

Recall that every circulant matrix can be written as $Q\Lambda Q'$, where Q is the N-point discrete Fourier transform (DFT) matrix with $Q_{jk} = \frac{1}{\sqrt{N}} e^{-2\pi\sqrt{-1}(j-1)(k-1)/N}$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ is a diagonal matrix. We can therefore write $K = Q\Lambda Q'$, and rewrite the DARE (31) as $\Lambda = (Q'AQ)\Lambda(Q'AQ)' - ((Q'AQ)\Lambda(Q'\mathbf{1}))(1 + \mathbf{1}'Q\Lambda Q'\mathbf{1})^{-1}((Q'AQ)\Lambda(Q'\mathbf{1}))'$. By our choice of A in (33) and (44), and since Q is the N-point DFT matrix

$$(Q'AQ)\Lambda(Q'AQ)' = \beta^2 \begin{pmatrix} \lambda_2 & 0 & \ldots & 0 \\ 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_1 \end{pmatrix}$$

$$(Q'AQ)\Lambda(Q'\mathbf{1}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \beta\lambda_1\sqrt{N} \end{pmatrix}$$

and the DARE in (31) can be expressed in terms of diagonal matrices only. Thus, in this case the DARE is equivalent to a set of N equations, where the first $N - 1$ equations are

$$\lambda_j = \beta^2\lambda_{j+1}, \quad j = 1, \ldots, N - 1 \tag{75}$$

and the N-th equation is

$$\lambda_N = \beta^2\lambda_1 - \frac{\beta^2\lambda_1^2 N}{1 + N\lambda_1}. \tag{76}$$

By (45) and since $\lambda_i = \lambda_{i-1}/\beta^2$ for $i = 2, \ldots, N$, we conclude that K satisfies (75) and (76), and hence is a solution to the DARE (31).

To prove (46), we notice that by the DARE (31) the diagonal entries of K must satisfy

$$K_{jj} = \beta^2 K_{jj} - \beta^2 \frac{\left(\sum_{k=1}^N K_{jk}\right)^2}{(1 + \mathbf{1}'K\mathbf{1})}. \tag{77}$$

Also, since Q is the N-point DFT matrix, $\lambda_1 = \sum_{k=1}^N K_{1k}$, and since K is circulant, $\sum_{k=1}^N K_{jk} = \sum_{k=1}^N K_{1k}$ for $j = 1, \ldots, N$. Thus, $\mathbf{1}'K\mathbf{1} = N\lambda_1$. Combining these two observations with (77), we obtain

$$\beta^2 = (1 + N\lambda_1)/(1 + \lambda_1(N - \lambda_1/K_{jj}))$$

which, combined with (45), yields (46) (with K replaced by $K^\star$).

Appendix F
Conditional Maximal Correlation

Let (V_1, V_2, Y) be jointly Gaussian. Then, the pair (V_1, V_2) is jointly Gaussian also when conditioned on $\{Y = y\}$, and the conditional correlation $\rho(V_1, V_2 \,|\, Y = y)$ does not depend on y and

$$\rho(V_1, V_2 \,|\, Y = y) = \rho(V_1, V_2 \,|\, Y) \tag{78}$$

where we recall that $\rho(V_1,V_2 \,|\, Y) = \int \rho(V_1,V_2 \,|\, Y = y)\, dF(y)$. Moreover, by the maximal correlation property of jointly Gaussian random variables [22], for every y

$$\sup_{g_1, g_2} \mathsf{E}(g_1(V_1) g_2(V_2) \,|\, Y = y) = \rho(V_1, V_2 \,|\, Y = y) \tag{79}$$

when the supremum on the RHS is over all functions $g_1(v_1)$ and $g_2(v_2)$ (implicitly dependent on y) that are of zero mean and unit variance with respect to the conditional distribution of (V_1, V_2) given $\{Y = y\}$. Hence

$$\begin{aligned}\rho^*(V_1, V_2 \,|\, Y) &= \sup_{g_1, g_2} \int \mathsf{E}(g_1(V_1, y) g_2(V_2, y) \,|\, Y = y) dF_Y(y) \\ &= \int \left(\sup_{g_1, g_2} \mathsf{E}(g_1(V_1) g_2(V_2) \,|\, Y = y) \right) dF_Y(y) \\ &= \rho(V_1, V_2 \,|\, Y) \end{aligned} \tag{80}$$

where the equality in (80) follows by (78) and (79), and because g_1 and g_2 are zero-mean for each y.

Verifying that the linear functions g_1 and g_2 in (53) satisfy $\mathsf{E}(g_1 \,|\, Y) = \mathsf{E}(g_2 \,|\, Y) = 0$, $\mathsf{E}(g_1^2 \,|\, Y) = \mathsf{E}(g_2^2 \,|\, Y) = 1$, and $\mathsf{E}(g_1(V_1, Y) g_2(V_2, Y)) = \rho(V_1, V_2 \,|\, Y)$ concludes the proof. Note that the proof remains valid also when Y is a Gaussian vector (instead of a scalar).

APPENDIX G
UPPER BOUND ON $C(P)$

By the standard arguments, we have

$$C(P) \le \limsup_{n \to \infty} \max \frac{1}{n} \sum_{i=1}^{n} I(X_{1i}, \ldots, X_{Ni}; Y_i \,|\, Y^{i-1}) \tag{81}$$

where the maximum is over the set of arbitrary functions $\{X_{ji}(M_j, Y^{i-1})\}$. Define now for each n an N-tuple of independent auxiliary random variables $U_1, \ldots, U_N$, where U_j is uniformly distributed over $[0, 1]$. Also, let

$$V_j := \Phi^{-1} \left(\frac{M_j - 1 + U_j}{2^{\lceil nR_j \rceil}} \right)$$

where Φ denotes the cumulative distribution function of a standard Gaussian random variable. Since $(M_j - 1 + U_j)/2^{\lceil nR_j \rceil}$ is uniformly distributed over $[0, 1]$, $V_j \sim \mathrm{N}(0, 1)$. Furthermore, by the strict monotonicity of Φ, it is possible to reconstruct M_j from V_j. Hence, the set of feasible functions in (81) can only increase if we consider $\{X_{ji}(V_j, Y^{i-1})\}$ instead of $\{X_{ji}(M_j, Y^{i-1})\}$. This establishes the upper bound in (50).

APPENDIX H
UPPER BOUND ON $C^{(n)}(P)$

Let $\tilde{X}_{ji} := X_{ji} - \mathsf{E}(X_{ji} \,|\, Y^{i-1})$ and $\tilde{Y}_i := \tilde{X}_{1i} + \tilde{X}_{2i} + Z_i$. It is not hard to see that $\mathsf{E}(\tilde{X}_{ji}^2) \le \mathsf{E}(X_{ji}^2) \le P_{ji}$ and $\mathsf{E}(\tilde{X}_{ji} \,|\, Y^{i-1}) = 0$ for $i = 1, \ldots, n$ and $j = 1, 2$. To establish the upper bound (54), consider

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i \,|\, Y^{i-1}) \\ &\quad = \frac{1}{n} \sum_{i=1}^{n} I(\tilde{X}_{1i}, \tilde{X}_{2i}; \tilde{Y}_i \,|\, Y^{i-1}) && (82) \\ &\quad \le \frac{1}{n} \sum_{i=1}^{n} I(\tilde{X}_{1i}, \tilde{X}_{2i}; \tilde{Y}_i) && (83) \\ &\quad \le \frac{1}{2n} \sum_{i=1}^{n} \log \left(1 + P_{1i} + P_{2i} + 2\sqrt{P_{1i} P_{2i}}\, \rho(\tilde{X}_{1i}, \tilde{X}_{2i}) \right) && (84) \end{aligned}$$

where the equality in (82) holds because $\mathsf{E}(X_{ji} \,|\, Y^{i-1})$ is a function of Y^{i-1}; the inequality in (83) follows since $\tilde{Y}_i \to (\tilde{X}_{1i}, \tilde{X}_{2i}) \to Y^{i-1}$ form a Markov chain; and the inequality in (84) follows by the maximum entropy theorem [16] and the fact that $\mathsf{E}(\tilde{X}_{ji}^2) \le P_{ji}$.

REFERENCES

[1] N. A. Gaarder and J. K. Wolf, "The capacity of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol. 21, pp. 100–102, 1975.

[2] L. H. Ozarow, "The capacity of the white Gaussian multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, pp. 623–629, 1984.

[3] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—I: No bandlimited constraint," *IEEE Trans. Inf. Theory*, vol. 12, pp. 172–182, 1966.

[4] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback—II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. 12, pp. 183–189, 1966.

[5] J. A. Thomas, "Feedback can at most double Gaussain multiple access channel capacity," *IEEE Trans. Inf. Theory*, vol. 33, pp. 711–716, 1987.

[6] E. Ordentlich, "On the factor-of-two bound for Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 42, pp. 2231–2235, 1996.

[7] S. Pombra and T. M. Cover, "Non white Gaussian multiple access channels with feedback," *IEEE Trans. Inf. Theory*, vol. 40, pp. 885–892, 1994.

[8] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1423–1438, 2002.

[9] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, pp. 37–43, 1989.

[10] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, pp. 44–53, 1989.

[11] G. Kramer and M. Gastpar, "Dependece balance and the Gaussian multiaccess channel with feedback," in *Proc. IEEE Inf. Theory Workshop*, 2006, pp. 198–202.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[13] W. Wu, S. Vishwanath, and A. Arapostathis, "Gaussian interference networks with feedback: Duality, sum capacity and dynamic team problem," in *Proc. 44th Ann. Allerton Conf. Commun. Control Comput.*, Monticello, IL, 2005.

[14] A. Réyni, "On measures of dependence," *Acta Math. Hungarica*, vol. 10, pp. 441–451, 1959.

[15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[17] P. Lancaster and L. Rodman, *Algebraic Riccati Equations*. New York: Oxford Univ. Press, 1995.
[18] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[19] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 57–85, 2010.
[20] G. H. Hardy, *Divergent Series*, 2nd ed. New York: Amer. Math. Soc., 1992.
[21] M. S. Iacobucci and M. G. DiBenedetto, "A feedback code for the multiple access channel (MAC): A case study," in *Proc. IEEE Global Commun. Conf., Commun. Theory Mini Conf.*, Phoenix, AZ, Nov. 1997, pp. 128–132.
[22] H. O. Lancaster, "Some properties of the bivariate normal distribution considered in the form of a contingency table," *Biometrika*, vol. 44, pp. 289–292, 1957.

Ehsan Ardestanizadeh (S'08–M'11) received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2004, and the M.S. and the Ph.D. degrees in electrical engineering from the University of California, San Diego, in 2007 and 2010, respectively.

Since February 2011, he has been with ASSIA Inc., Redwood City, CA. His research interests are in information theory, feedback communication, and networked systems.

Dr. Ardestanizadeh is the recipient of the 2010 Shannon Fellowship awarded by the Center for Magnetic Recording at the University of California, San Diego.

Michèle Wigger (S'05–M'09) received the M.Sc. degree in electrical engineering (with distinction) and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively.

In 2009, she was a Postdoctoral Researcher at the ITA Center, University of California, San Diego. Since December 2009, she has been an Assistant Professor at Telecom ParisTech, France. Her research interests are in information and communications theory; in particular in wireless networks, feedback channels, and channels with states.

Young-Han Kim (S'99–M'06) received the B.S. degree with honors in electrical engineering from Seoul National University, Korea, in 1996 and the M.S. degrees in electrical engineering and statistics, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2001, 2006, and 2006, respectively.

In July 2006, he joined the University of California, San Diego, where he is currently an Assistant Professor of Electrical and Computer Engineering. His research interests are in statistical signal processing and information theory, with applications in communication, control, computation, networking, data compression, and learning.

Dr. Kim is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award and the 2009 U.S.—Israel Binational Science Foundation Bergmann Memorial Award.

Tara Javidi (S'96–M'02) studied electrical engineering at the Sharif University of Technology, Tehran, Iran, from 1992 to 1996. She received the M.S. degrees in electrical engineering (systems) and applied mathematics (stochastics) as well as the Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 1998, 1999, and 2002, respectively.

From 2002 to 2004, she was an Assistant Professor with the Electrical Engineering Department, University of Washington, Seattle. She joined the University of California, San Diego, in 2005, where she is currently an Associate Professor of Electrical and Computer Engineering. Her research interests are in communication networks, stochastic resource allocation, and wireless communications.

Dr. Javidi was a Barbour Scholar during the 1999–2000 academic year and received an NSF CAREER Award in 2004.