On the Computation of the Capacity Region of the Discrete MAC

Eduard Calvo, *Student Member, IEEE*, Daniel P. Palomar, *Senior Member, IEEE*, Javier Rodríguez Fonollosa, *Senior Member, IEEE*, and Josep Vidal, *Member, IEEE*

***Abstract*—The computation of the channel capacity of discrete memoryless channels is a convex problem that can be efficiently solved using the Arimoto-Blahut (AB) iterative algorithm. However, the extension of this algorithm to the computation of capacity regions of multiterminal networks is not straightforward since it gives rise to non-convex problems. In this context, the AB algorithm has only been successfully extended to the calculation of the sum-capacity of the discrete memoryless multiple-access channel (DMAC). Thus, the computation of the whole capacity region still requires the use of computationally demanding search methods.**

In this paper, we first give an alternative reformulation of the capacity region of the DMAC which condenses all the non-convexities of the problem into a single rank-one constraint. Then, we propose efficient methods to compute outer and inner bounds on the capacity region of the two-user DMAC by solving a relaxed version of the problem and projecting its solution onto the original feasible set. Targeting numerical results, we first take a randomization approach. Focusing on analytical results, we study projection via minimum divergence, which amounts to the marginalization of the relaxed solution. In this case we derive sufficient conditions and necessary and sufficient conditions for the bounds to be tight. Furthermore, we are able to show that the class of channels for which the marginalization bounds match exactly the capacity region includes all the two-user binary-input deterministic DMACs as well as other non-deterministic channels. In general, however, both methods are able to compute very tight bounds as shown for various examples.

***Index Terms*—Multiple-access channel (MAC), algorithms, nonconvex optimization.**

I. Introduction

THE characterization of the capacity of an arbitrary single-user memoryless channel is a problem that admits a single-letter representation in the form of a maximization of a concave function over a convex set, e.g., a probability simplex for the Discrete Memoryless Channel (DMC). This is a convex problem that can be efficiently solved in practice (i.e., with polynomial time worst-case complexity) [1]. For example, for the continuous Gaussian channel the solution admits a simple closed-form characterization [2] whereas for the DMC the popular practical Arimoto-Blahut (AB) algorithm [3], [4] can be used.

In contrast, for the general multiuser case we do not even have a characterization of the capacity region. Although many major breakthroughs in the field have been achieved (see [5, Ch. 14] and [6] and references therein), there are many open problems on single-letter characterizations of capacity regions. Fortunately, for the multiple-access channel (MAC) we also have a single-letter representation of the capacity region [7], [8]. However, this characterization is not generally in the form of a convex optimization problem. While for the continuous Gaussian channel convexity holds and the capacity region can be numerically evaluated in an efficient way [9], for the discrete memoryless MAC (DMAC) the lack of convexity prevents us from finding an efficient algorithm to compute the capacity region in practice. In this context, many authors have recently contributed toward the computation of the sum-capacity (or total capacity) of an arbitrary DMAC [10], [11], and an algorithm for its exact computation has been found [12].

It was shown in [10] that any two-user DMAC can be decomposed into a finite number of elementary (2-user binary-input and binary-output) DMACs for which their total capacity can be computed using a necessary and sufficient optimality condition of the input probability distributions. In addition, [10] showed that for any 2-user non-binary inputs and binary output DMAC, the total capacity can be determined by computing the total capacities of the elementary DMACs of its decomposition, and provided an iterative algorithm. In a later work, [11] extended the decomposition result for the K-user DMAC (with arbitrary input and output alphabets), which allowed [12] to propose an algorithm for the computation of the total capacity based on a generalization of the AB algorithm. Other applications of generalizations of the AB algorithm can be found in the context of the computation of channel capacity with side information [13].

Regarding the computation of the whole capacity region of the DMAC, not much work has been done due to the intractability of the problem because of its non-convexity. As a consequence, brute-force algorithms or random search methods seem to be the only alternative to compute inner bounds on the capacity region with no quantification on the suboptimality incurred. In addition to the theoretical relevance of finding efficient numerical methods for the computation of

Paper approved by T.-S. P. Yum, the Editor for Packet Access and Switching of the IEEE Communications Society. Manuscript received May 13, 2009; revised October 9, 2009 and March 1, 2010.

This work has been partially funded by the Spanish Science and Technology Commissions, FEDER funds from the EC (TEC2006-06481/TCM, TEC2010-19171/TCM and CONSOLIDER INGENIO CSD2008-00010 COMONSENS), and the NSFC/RGC N_HKUST604/08 research grant.

E. Calvo is with the Department of Production, Technology, and Operations Management, IESE Business School, University of Navarra, Barcelona, Spain (e-mail: ecalvo@iese.edu).

D. P. Palomar is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: palomar@ust.hk).

J. Rodríguez Fonollosa and J. Vidal are with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Barcelona, Spain (e-mail: {javier.fonollosa, josep.vidal}@upc.edu).

Digital Object Identifier 10.1109/TCOMM.2010.091710.090239

0090-6778/10$25.00 © 2010 IEEE

the capacity region of the DMAC, it is worth mentioning also some implications on the design of practical multiple-access communication systems:

- First, the DMAC has an operational meaning. Regardless of the underlying physical channel model, the inherently digital nature of communication systems renders multiple-access channels discrete. In other words, the use of deterministic constellations at the transmitters plus quantization of the channel output at the receiver results in an equivalent end-to-end channel that can be solely described using a transition probability matrix.
- Often, the evaluation of the virtues of practical multiple-access schemes involves extensive simulation over a large set of scenarios accounting for different values of system parameters and/or uncertainties. If computation of the capacity region were efficient, the computational burden associated to this performance quantification could be greatly alleviated.
- Finally, since the computation of the capacity region is frequently related to searching optimal input probability distributions, these can be used to guide the design of practical schemes aiming at the maximization of the achievable rates via structured coding.

In this work, we show that the key difficulty in computing the capacity region of an arbitrary DMAC can be identified as a rank-one constraint (a non-convex constraint) in an otherwise convex optimization problem. Optimization problems with this kind of constraint arise in areas such as control theory [14]–[16] and signal processing [17], [18], and cannot be solved optimally in polynomial time with state of the art knowledge. Hence, alternative suboptimal methods must be used to obtain good approximations of the capacity region. One approach that has reported near optimal performance when dealing with rank-one constraints in maximum-likelihood single-user [17], [19] and multiuser [18] detection is the use of relaxation methods. Relaxation methods are based on i) replacing the rank-one constrained problem by an approximate (not equivalent) tractable convex problem and ii) generating a potential solution to the original problem from the solution to the relaxed problem. This way, the use of computationally demanding algorithms is avoided since efficient interior point methods can be used to solve the convex approximation of the problem in polynomial time.

We propose two efficient methods for the computation of both an inner and an outer bound of the capacity region of any two-user DMAC. The outer bound follows from removing the rank-one constraint and corresponds to the achievable rates in the situation of full transmit cooperation, since user codewords can thus be arbitrarily correlated. To generate potential solutions to the original problem, we first focus on randomization, an approach that has shown near optimal numerical performance in the previously mentioned areas. In essence, several rank-one input probability distributions are generated close (in the mean sense) to the optimal solution of the relaxed problem and the one yielding the largest achievable rates is kept. This can be viewed as a random search algorithm with guidance on the correlation matrix of the potential solutions.

Pursuing a simpler method that allows for performance analysis, we then study a deterministic alternative in which the solution to the relaxed problem is projected onto the feasible set via a minimum divergence criterion. This criterion yields a candidate solution which turns out to be the marginalization of the relaxed solution, a very simple operation scalable with the number of users. Regarding analytical results, there exists a class of channels for which this algorithm is able to compute exactly the capacity region. It comprises the subclass of channels with identical inner and outer bounds and the subclass of channels with strict outer bound and tight inner bound. Given a channel, we derive necessary and sufficient conditions for checking whether it belongs to the first subclass and sufficient conditions for verifying whether it belongs to the second subclass. These conditions are used to show that for all the two-user binary-input deterministic DMACs as well as for some non-deterministic channels simple marginalization of the full cooperation solution achieves capacity. Although we have not been able to fully characterize analytically the class of channels for which marginalization is optimal, numerical simulations for various channels show that both randomization and marginalization perform indistinguishably to the optimal solution obtained with a computationally intensive brute-force full search.

The structure of this paper is as follows. Section II introduces the problem of the computation of the capacity region of the DMAC and reformulates it as a rank-one constrained optimization problem. Section III describes the proposed relaxation-based methods for the computation of inner and outer bounds on the capacity region: randomization and marginalization. Analytical optimality conditions that determine when the marginalization bounds are tight are provided in Section IV. Then, the performance of the proposed algorithms among various channels is numerically compared to that of a random search method in Section V. Finally, Section VI concludes the paper.

II. The Capacity Region as a Rank-One Constrained Optimization Problem

The computation of the capacity region of an arbitrary DMAC (a convex set) is a non-convex problem. It can be formulated in a matrix form that reveals the non-convexity of the problem as a rank-one constraint.

A. The Problem of the Capacity Region for Two Users

The capacity region $\mathcal{C}$ of the 2-user DMAC is the convex hull of the set[1] of rate pairs (R_1, R_2) satisfying

$$0 \leq R_1 \leq I(X_1; Y|X_2) \quad (1)$$

$$0 \leq R_2 \leq I(X_2; Y|X_1) \quad (2)$$

$$R_1 + R_2 \leq I(X_1 X_2; Y) \quad (3)$$

for a distribution of the form $P_{X_1X_2Y} = P_{X_1}P_{X_2}P_{Y|X_1X_2}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, where the input alphabets can be characterized as $\mathcal{X}_k = \{x_k^{(1)}, \ldots, x_k^{(|\mathcal{X}_k|)}\}$, with $|\mathcal{X}_k|$ denoting the cardinality of $\mathcal{X}_k$, $k = 1, 2$. P_{X_k} is the input probability distribution of the k-th user ($k = 1, 2$), and $P_{Y|X_1X_2}$ is the given conditional distribution that characterizes the channel. It is well known that $\mathcal{C}$ is a convex set [5, Thm. 14.3.2] and hence,

[1]The convex hull is strictly necessary for convexification of $\mathcal{C}$ since otherwise it may not be convex in general [20].

$$\underset{R_1,R_2,P_{X_1},P_{X_2}}{\text{maximize}} \quad \theta R_1 + (1-\theta) R_2 \tag{4}$$

$$\text{subject to} \quad 0 \leq R_1 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1) P_{X_2}(x_2) P_{Y|X_1X_2}(y|x_1x_2) \log \frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_1'} P_{X_1}(x_1') P_{Y|X_1X_2}(y|x_1'x_2)} \tag{5}$$

$$0 \leq R_2 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1) P_{X_2}(x_2) P_{Y|X_1X_2}(y|x_1x_2) \log \frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_2'} P_{X_2}(x_2') P_{Y|X_1X_2}(y|x_1x_2')} \tag{6}$$

$$R_1 + R_2 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1) P_{X_2}(x_2) P_{Y|X_1X_2}(y|x_1x_2) \log \frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_1',x_2'} P_{X_1}(x_1') P_{X_2}(x_2') P_{Y|X_1X_2}(y|x_1'x_2')} \tag{7}$$

$$\sum_{x_k} P_{X_k}(x_k) = 1 \ , \ P_{X_k}(x_k) \geq 0 \ \forall x_k \in \mathcal{X}_k \quad k = 1,2 \tag{8}$$

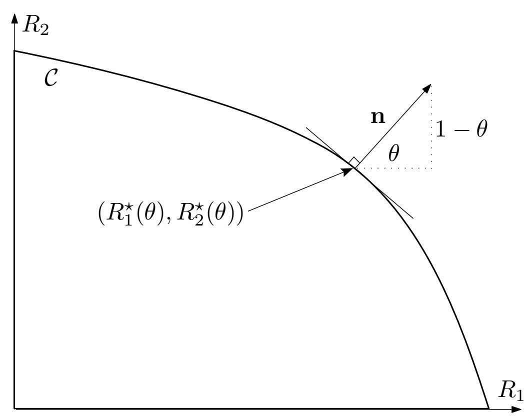


Fig. 1. The boundary of $\mathcal{C}$ is obtained solving (4)-(8) for each $\theta \in [0,1]$.

by applying the supporting hyperplane theorem [1, Sec. 2.5.2], the computation of the capacity region can be parameterized for $\theta \in [0,1]$ as[2] (4)-(8) at the top of the page, where the expressions (5)-(7) correspond to (1)-(3), respectively, instantiated for the DMAC. Note that the solutions $P^\star_{X_1}(\theta)$ and $P^\star_{X_2}(\theta)$ generally depend on θ. For each θ, the problem (4)-(8) computes the intersection between the contour of the capacity region and a tangent hyperplane with normal vector $\mathbf{n} = [\theta, \ 1-\theta]^T$, as illustrated in Fig. 1. Hence, the capacity region is computed when (4)-(8) is solved for all $\theta \in [0,1]$ and the convex hull of $\{R_1^\star(\theta), R_2^\star(\theta)\}_{\forall\theta\in[0,1]}$ is taken; in other words, the solutions $(R_1^\star(\theta), R_2^\star(\theta))$ are rate pairs lying in the boundary of $\mathcal{C}$.

B. A Rank-one Constrained Optimization Problem

The problem (4)-(8) of the computation of the capacity region is non-convex because the constraints (5)-(7) are not jointly convex in P_{X_1} and P_{X_2}. For instance, the right hand side of the constraint in (7) is not concave (note that it should be concave for the problem to be convex). To see this, observe that even though $x\log(1/x)$ is concave, the composition with a linear combination of terms of the form xy is not[3]. Similar reasonings may be applied to the constraints (5) and (6) to obtain again that the lack of convexity follows from the product terms $P_{X_1}(x_1)P_{X_2}(x_2)$.

Although the problem (4)-(8) is not jointly convex in (P_{X_1}, P_{X_2}), it is separately convex in each of the input probability distributions. This would allow us to perform an alternate optimization procedure: $P^{(0)}_{X_1} \to P^{(0)}_{X_2} \to P^{(1)}_{X_1} \to P^{(1)}_{X_2} \to \ldots$, where $P^{(n)}_{X_k}$ denotes the optimal solution P_{X_k} at the n-th iteration. However, alternate optimization procedures applied to non-convex problems do not generally converge to global maxima of the cost function, and for this particular problem they do not yield acceptable results (as can be verified by numerical simulations).

Interestingly, we shall see that if we allow the variables X_1 and X_2 to be dependent on each other with joint distribution $P_{X_1X_2}$, then problem (4)-(8) becomes convex (recall that $x\log(1/x)$ is a concave function). Using a matrix-vector notation, each of the input probability distributions P_{X_k} admits a vector representation of the form $\mathbf{p}_k$, where $[\mathbf{p}_k]_i = P_{X_k}(x_k^{(i)})$, $1 \leq i \leq |\mathcal{X}_k|$, $k = 1,2$, while the joint distribution admits a matrix representation of the form $\mathbf{P}$, where $[\mathbf{P}]_{i,j} = P_{X_1X_2}(x_1^{(i)}, x_2^{(j)})$, $1 \leq i \leq |\mathcal{X}_1|$, $1 \leq j \leq |\mathcal{X}_2|$. Then, we define $\mathcal{P}_{\mathsf{prod}}$ as the subset containing all the *product* distributions $P_{X_1}P_{X_2}$ of X_1 and X_2, see (9) at the top of the next page. For any joint probability matrix $\mathbf{P} \in \mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|}$, $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$ is equivalent to rank$(\mathbf{P}) = 1$, and hence the following simpler equivalent description of $\mathcal{P}_{\mathsf{prod}}$ can be given as (10) at the top of the next page, where $\geq$ denotes component-wise as well as scalar inequality indistinctly and $\mathbf{1}$ is an all-one column vector of appropriate length. The problem (4)-(8) can now be expressed in terms of $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$, the joint distribution of X_1 and X_2, and its marginals $\mathbf{p}_1$ and $\mathbf{p}_2$, making use of expression (10) and the fact that

$$\mathbf{P}\mathbf{1} = \mathbf{p}_1, \quad \mathbf{P}^T\mathbf{1} = \mathbf{p}_2, \tag{11}$$

i.e., that P_{X_1} and P_{X_2} are the marginal distributions of $P_{X_1X_2}$. The following reformulation of the problem is the key point of the identification of (4)-(8) as a rank-one non-convex optimization problem.

[2]Unless the logarithm basis is indicated, it can be chosen arbitrarily as long as both sides of the equation have the same units.

[3]It is sufficient to note that the Hessian of $f(x,y) = xy\log(xy)$ has one positive and one negative eigenvalue at $(x,y) = (1/\sqrt{2}, 1/\sqrt{2})$.

$$\mathcal{P}_{\mathsf{prod}} = \left\{ \mathbf{P} \in \mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|} \;\middle|\; [\mathbf{P}]_{i,j} = P_{X_1}(x_1^{(i)})P_{X_2}(x_2^{(j)}) \text{ for some feasible } (P_{X_1}, P_{X_2}) \text{ on } \mathcal{X}_1 \times \mathcal{X}_2 \right\} \quad (9)$$

$$\mathcal{P}_{\mathsf{prod}} = \left\{ \mathbf{P} \in \mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|} \;\middle|\; \mathrm{rank}(\mathbf{P}) = 1 \;,\; \mathbf{P} \geq \mathbf{0} \;,\; \mathbf{1}^T\mathbf{P}\mathbf{1} = 1 \right\} \quad (10)$$

Proposition 1: The problem (4)-(8) of the computation of the capacity region of an arbitrary two-user DMAC is equivalent to the following *rank-one non-convex optimization problem*

$$\begin{aligned} \underset{R_1,R_2,\mathbf{P},\mathbf{p}_1,\mathbf{p}_2}{\text{maximize}} \quad & \theta R_1 + (1-\theta)R_2 && (12)\\ \text{subject to} \quad & 0 \leq R_1 \leq f_1(\mathbf{P}, \mathbf{p}_2) && (13)\\ & 0 \leq R_2 \leq f_2(\mathbf{P}, \mathbf{p}_1) && (14)\\ & R_1 + R_2 \leq f_{12}(\mathbf{P}) && (15)\\ & \mathbf{P}\mathbf{1} = \mathbf{p}_1 \;,\; \mathbf{P}^T\mathbf{1} = \mathbf{p}_2 && (16)\\ & \mathbf{P} \geq \mathbf{0} \;,\; \mathbf{1}^T\mathbf{P}\mathbf{1} = 1 && (17)\\ & \mathrm{rank}(\mathbf{P}) = 1, && (18) \end{aligned}$$

where f_1, f_2 , and f_{12} at the top of the next page are concave in $(\mathbf{P}, \mathbf{p}_2)$, $(\mathbf{P}, \mathbf{p}_1)$, and $\mathbf{P}$, respectively.

Proof: See Appendix I. ■

Observe that concavity of f_1, f_2, and f_{12} and linearity of (12) and (16)-(17) imply that, if (18) were ignored, the resulting problem would be convex. While (17) ensures that $\mathbf{P}$ is a feasible probability matrix, (18) constrains it to $\mathcal{P}_{\mathsf{prod}}$, and (16) relates $\mathbf{P}$ with its marginals.

C. Extension to K Users

The formulation of the computation of the capacity region as a rank-one constrained problem introduced in Section II-B for two users can be extended to the K-user case. Using similar equivalences now involving tensors [21], the rank-one constraint applies to any number of users.

The capacity region $\mathcal{C}$ of the K-user DMAC is the convex hull of the set of rate tuples $(R_1, \ldots, R_K)$ satisfying

$$0 \leq R_{\mathcal{S}} < I(X_{(\mathcal{S})}; Y | X_{(\mathcal{S}^c)}), \quad \forall \mathcal{S} \subseteq \mathcal{N} \quad (22)$$

for a distribution of the form $P_{X_1 \ldots X_K Y} = P_{X_1} \cdot \ldots \cdot P_{X_K} P_{Y|X_1 \ldots X_K}$ on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_K \times \mathcal{Y}$, where $\mathcal{N} \triangleq \{1, 2, \ldots, K\}$. By defining $\mathcal{S}^c = \mathcal{N} \setminus \mathcal{S}$ the complement set of $\mathcal{S}$, $X_{(\mathcal{S})} = \{X_k : k \in \mathcal{S}\}$, and $R_{\mathcal{S}} = \sum_{k\in\mathcal{S}} R_k$, the computation of the capacity region can be parameterized as (23)-(26) at the top of the next page. The solution to (23)-(26) for any given $\boldsymbol{\theta} = [\theta_1 \ldots \theta_K]^T$ such that $\boldsymbol{\theta} \geq \mathbf{0}$ and $\mathbf{1}^T\boldsymbol{\theta} = 1$ is a point $(R_1^\star(\boldsymbol{\theta}), \ldots, R_K^\star(\boldsymbol{\theta}))$ of the boundary of the capacity region $\mathcal{C}$.

Similarly to what happened in the two-user case, the problem (23)-(26) is not convex because the constraints in (25) are not jointly convex in $(P_{X_1}, \ldots, P_{X_K})$. However, (23)-(26) can be also reformulated as a rank-one non-convex optimization problem if we allow X_1, …, X_K to be dependent with distribution $P_{X_{(\mathcal{N})}}$. In doing so, it is useful to extend the matrix-vector notation of Section II-B using the tensor $\mathbf{P}_{(\mathcal{S})}$ to denote $P_{X_{(\mathcal{S})}}$, where $[\mathbf{P}_{(\mathcal{S})}]_{i_1,i_2,\ldots,i_{|\mathcal{S}|}} = P_{X_{(\mathcal{S})}}\left(x_{k_1}^{(i_1)}, x_{k_2}^{(i_2)}, \ldots, x_{k_{|\mathcal{S}|}}^{(i_{|\mathcal{S}|})}\right)$ $\forall 1 \leq i_j \leq |\mathcal{X}_{k_j}|$, $1 \leq j \leq |\mathcal{S}|$, $\forall \mathcal{S} = \{k_1, k_2, \ldots, k_{|\mathcal{S}|}\} \subseteq \mathcal{N}$. $P_{X_{(\mathcal{S})}}$ is the marginalization of $P_{X_{(\mathcal{N})}}$ into the codeword set $X_{(\mathcal{S})}$, i.e., $P_{X_{(\mathcal{S})}}(x_{(\mathcal{S})}) = \sum_{x_{(\mathcal{S}^c)}} P_{X_{(\mathcal{N})}}(x_{(\mathcal{N})})$ or, equivalently in tensor notation, $\sum_{i_{(\mathcal{S}^c)}} \mathbf{P}_{(\mathcal{N})} = \mathbf{P}_{(\mathcal{S})}$.

Proposition 2: The problem (23)-(26) of the computation of the capacity region of an arbitrary K-user DMAC is equivalent to the following rank-one[4] non-convex optimization problem:

$$\begin{aligned} \underset{\{R_k\},\{\mathbf{P}_{(\mathcal{S})}\}_{\forall\mathcal{S}\subseteq\mathcal{N}}}{\text{maximize}} \quad & \sum_{k=1}^{K} \theta_k R_k && (28)\\ \text{subject to} \quad & 0 \leq R_{\mathcal{S}} \leq f_{\mathcal{S}}(\mathbf{P}_{(\mathcal{N})}, \mathbf{P}_{(\mathcal{S}^c)}), \quad \forall \mathcal{S} \subseteq \mathcal{N} && (29)\\ & \sum_{i_{(\mathcal{S}^c)}} \mathbf{P}_{(\mathcal{N})} = \mathbf{P}_{(\mathcal{S})} \quad \forall \mathcal{S} \subseteq \mathcal{N} && (30)\\ & \mathbf{P}_{(\mathcal{N})} \succeq 0 \;,\; \sum_{i_{(\mathcal{N})}} \mathbf{P}_{(\mathcal{N})} = 1 && (31)\\ & \mathrm{rank}(\mathbf{P}_{(\mathcal{N})}) = 1, && (32) \end{aligned}$$

where the functions in (33) (see next page), are concave in $(\mathbf{P}_{(\mathcal{N})}, \mathbf{P}_{(\mathcal{S}^c)})$.

Proof: This follows from extending the definition of $\mathcal{P}_{\mathsf{prod}}$ (9) to suit the K-dimensional tensor $\mathbf{P}_{(\mathcal{N})}$ and noticing that the functions $f_{\mathcal{S}}$ are the generalizations of f_1, f_2, and f_{12} in (19)-(21) to the K-user case. Hence, concavity of $f_{\mathcal{S}}$ also follows from Proposition 1. ■

III. Relaxation Methods

Propositions 1 and 2 identified the non-convexity of the problem of the computation of the capacity region of an arbitrary DMAC as a rank-one constrained optimization problem, enabling the design of efficient numerical approaches aiming at upper and lower bounding the capacity region. For the sake of simplicity in notation and presentability of subsequent results, we shall concentrate on the two-user case in the remainder of the paper. It is understood, however, that the techniques and results presented next extend naturally to the general K-user case with the appropriate considerations.

Rank-one constrained optimization problems, even with linear matrix inequalities, are non-convex problems that cannot be solved optimally in polynomial time. Thus, we first choose to relax (12)-(18) by removing the rank-one constraint (18) to obtain a tractable, convex problem equivalent to solving the capacity region of a DMAC with arbitrarily dependent codewords (full transmitter cooperation). We will denote by $\mathcal{R}^\circ$ the outer bound on the true capacity region obtained with the relaxed problem (12)-(17).

[4]The K-dimensional tensor $\mathbf{P}_{(\mathcal{N})} \in \mathbb{R}_+^{|\mathcal{X}_1|\times\ldots|\mathcal{X}_K|}$ has rank one if and only if it can be written as

$$\mathbf{P}_{(\mathcal{N})} = \mathbf{p}_1 \otimes \mathbf{p}_2 \otimes \ldots \otimes \mathbf{p}_K, \quad (27)$$

where $\otimes$ denotes outer product and the vectors $\mathbf{p}_k \in \mathbb{R}_+^{|\mathcal{X}_k|}$ admit a vector equivalence similar to that of Section II-B. Similarly to Lemma 1, $\mathrm{rank}(\mathbf{P}_{(\mathcal{N})}) = 1$ is equivalent to imposing $P_{X_1 \ldots X_K} = P_{X_1} \cdot \ldots \cdot P_{X_K}$.

$$f_1(\mathbf{P},\mathbf{p}_2) \triangleq \sum_{i,j,y}[\mathbf{P}]_{i,j}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})\log\frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})[\mathbf{p}_2]_j}{\sum_{i'}[\mathbf{P}]_{i',j}P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j)})} \quad (19)$$

$$f_2(\mathbf{P},\mathbf{p}_1) \triangleq \sum_{i,j,y}[\mathbf{P}]_{i,j}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})\log\frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})[\mathbf{p}_1]_i}{\sum_{j'}[\mathbf{P}]_{i,j'}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j')})} \quad (20)$$

$$f_{12}(\mathbf{P}) \triangleq \sum_{i,j,y}[\mathbf{P}]_{i,j}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})\log\frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{i',j'}[\mathbf{P}]_{i',j'}P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j')})} \quad (21)$$

$$\underset{\{R_k\},\{P_{X_{(\mathcal{S})}}\}_{\forall\mathcal{S}\subseteq\mathcal{N}}}{\text{maximize}} \quad \sum_{k=1}^{K}\theta_k R_k \quad (23)$$

$$\text{subject to} \quad 0 \le R_{\mathcal{S}} \le \sum_{x_{(\mathcal{N})},y}P_{X_{(\mathcal{N})}}(x_{(\mathcal{N})})P_{Y|X_{(\mathcal{N})}}(y|x_{(\mathcal{N})})\log\frac{P_{Y|X_{(\mathcal{N})}}(y|x_{(\mathcal{N})})}{\sum_{x'_{(\mathcal{S})}}P_{X_{(\mathcal{S})}}(x'_{(\mathcal{S})})P_{Y|X_{(\mathcal{N})}}(y|x'_{(\mathcal{S})},x_{(\mathcal{S}^c)})} \quad (24)$$

$$P_{X_{(\mathcal{S})}}(x_{(\mathcal{S})}) = \prod_{k\in\mathcal{S}}P_{X_k}(x_k),\ \forall\mathcal{S}\subseteq\mathcal{N} \quad (25)$$

$$P_{X_k}(x_k)\ge 0\ \forall x_k\in\mathcal{X}_k,\ \sum_{x_k}P_{X_k}(x_k)=1,\quad \forall k\in\mathcal{N} \quad (26)$$

$$f_{\mathcal{S}}(\mathbf{P}_{(\mathcal{N})},\mathbf{P}_{(\mathcal{S}^c)}) = \sum_{i_{(\mathcal{N})},y}[\mathbf{P}_{(\mathcal{N})}]_{i_{(\mathcal{N})}}P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{N})}^{i_{(\mathcal{N})}}\big)\log\frac{P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{N})}^{i_{(\mathcal{N})}}\big)[\mathbf{P}_{(\mathcal{S}^c)}]_{i_{(\mathcal{S}^c)}}}{\sum_{i'_{(\mathcal{S})}}[\mathbf{P}_{(\mathcal{N})}]_{i'_{(\mathcal{S})},i_{(\mathcal{S}^c)}}P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{S})}^{i'_{(\mathcal{S})}},x_{(\mathcal{S}^c)}^{i_{(\mathcal{S}^c)}}\big)} \quad (33)$$

If the optimal solution to the relaxed problem happens to be rank one, then it will also be an optimal (capacity-achieving) solution to the original problem. Clearly, this happens in channels where transmit cooperation does not increase the achievable rates, rendering cooperation useless. Necessary and sufficient conditions for this phenomenon are provided in Section IV-A (Corollary 1) which are essentially derived from the Karush-Kuhn-Tucker (KKT) conditions of the relaxed problem.

When the solution to the relaxed problem is not rank one, it has to be projected onto $\mathcal{P}_{\text{prod}}$ to obtain a candidate solution (not necessarily optimal). Many different ad-hoc approaches can be applied to approximate an arbitrary joint distribution $\mathbf{P}$ by a reduced rank distribution of the form $\mathbf{q}_1\mathbf{q}_2^T$; based on simulation and mathematical amenability, however, this paper will deal with two of them only: randomization and marginalization.

A. Randomization

A randomization approach generates random samples of candidate probability vectors $(\mathbf{q}_1,\mathbf{q}_2)$ such that $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}^\star$, thus approximating the solution to the relaxed problem in the mean sense. For each of the generated pairs, the achievable rates (4)-(8) are evaluated and the pair of distributions yielding the largest objective value is kept as the solution. This is equivalent to performing a random search on the original problem with guidance on the correlation matrix of the distributions under test taken from the relaxed problem.

The nature of the random pairs $(\mathbf{q}_1,\mathbf{q}_2)$, which are input probability distributions, prevents us from easily finding statistics for generating them yielding $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}$ directly. Instead, we first choose to approximate $\mathbf{P}$ by the convex combination of n rank-one distributions under a minimum divergence criterion, i.e.,

$$\mathbf{P} \approx \sum_{i=1}^{n}\lambda_i\mathbf{a}_i\mathbf{b}_i^T, \quad (34)$$

where n is fixed, and

$$\{\lambda_i,\mathbf{a}_i,\mathbf{b}_i\} = \underset{\{\lambda_i,\mathbf{a}_i,\mathbf{b}_i\}}{\arg\min}\ D(\mathbf{P}\|\sum_{i=1}^{n}\lambda_i\mathbf{a}_i\mathbf{b}_i^T) \quad (35)$$

$$\text{subject to} \sum_{i=1}^{n}\lambda_i = 1,\ \lambda_i\ge 0 \quad 1\le i\le n \quad (36)$$

$$\mathbf{1}^T\mathbf{a}_i = 1,\ \mathbf{a}_i\ge\mathbf{0} \quad 1\le i\le n \quad (37)$$

$$\mathbf{1}^T\mathbf{b}_i = 1,\ \mathbf{b}_i\ge\mathbf{0} \quad 1\le i\le n, \quad (38)$$

where $D(\cdot\|\cdot)$ in (35) refers to Kullback-Leibler divergence[5]. The problem (35)-(38) is not jointly convex in $\{\lambda_i,\mathbf{a}_i,\mathbf{b}_i\}$ but is separately convex in $\{\lambda_i\}$, $\{\mathbf{a}_i\}$, and $\{\mathbf{b}_i\}$. Thus, a practical approximation as in (34) can be obtained through an alternate optimization $\{\lambda_i^{(0)}\}\to\{\mathbf{a}_i^{(0)}\}\to\{\mathbf{b}_i^{(0)}\}\to\{\lambda_i^{(1)}\}\to\ldots$

[5] The Kullback-Leibler divergence of two probability distributions P_X, Q_X on $\mathcal{X}$ is defined as

$$D(P_X\|Q_X) = \sum_{x\in\mathcal{X}}P_X(x)\log\frac{P_X(x)}{Q_X(x)}$$

and satisfies $D(P_X\|Q_X)\ge 0$, where $D(P_X\|Q_X)=0$ if and only if $P_X = Q_X$.

until convergence of the objective is achieved[6] ($(.)^{(r)}$ denotes value at the r-th iteration).

Second, given that N random samples $(\mathbf{q}_1, \mathbf{q}_2)$ must be generated, the approximation in the mean sense $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}$ can be achieved using (34) by drawing, $i = 1 \ldots n$, $N\lambda_i$ pairs from a pair of independent distributions such that $\mathbb{E}\{\mathbf{q}_1\} = \mathbf{a}_i$, $\mathbb{E}\{\mathbf{q}_2\} = \mathbf{b}_i$ (statistical independence implies $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{a}_i\mathbf{b}_i^T$). To generate a random vector $\mathbf{q}$ with prescribed average $\mathbb{E}\{\mathbf{q}\} = \overline{\mathbf{q}}$ we use a distribution whose support $\Omega_{\mathbf{q}}$ is the largest sphere centered at $\overline{\mathbf{q}}$ such that all its boundary lies within the probability simplex, as illustrated in Fig. 2 for the 3-dimensional case. While the radius r of such sphere can be analytically determined resorting to the point-line and point-plane distance formulas[7], its expression is omitted here for the sake of brevity. As for the distribution of $\mathbf{q}$, we choose to project a random vector $\mathbf{x}$ drawn uniformly over $\{\mathbf{x} \in \mathbb{R}^{|\mathcal{X}|\times 1} \ : \ \|\mathbf{x} - \overline{\mathbf{q}}\|^2 \le r^2\}$[8] onto the probability simplex[9], i.e.,

$$\mathbf{q} = \mathbf{x} + \frac{1 - \mathbf{1}^T\mathbf{x}}{|\mathcal{X}|}\mathbf{1}, \tag{39}$$

which results in a circularly symmetric distribution with average $\overline{\mathbf{q}}$.

The evaluation of the achievable rates (4)-(8) yields the randomization inner bound, $\mathcal{R}^{\mathsf{i}}_{\mathsf{rand}}$ (see Algorithm 1 for a pseudocode description).

B. Marginalization

While the numerical performance of randomization is good, it is at the price of generating many potential solutions at random. Therefore, it is desirable to explore other simpler (deterministic) methods retaining most of the accuracy and also allowing for performance analysis.

To this end, we adopt a projection criterion based also on the Kullback-Leibler divergence. It has been used in [23] and [24] as the criterion for approximating joint discrete probability distributions given a dependence tree, although here the purpose is different. The use of the information divergence as the measure that quantifies the quality of the approximation offers several advantages, the most useful one being that, for some fixed $\mathbf{P}$ (with marginals $\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$), the pair $(\mathbf{p}_1, \mathbf{p}_2)$ that minimizes $D(\mathbf{P}||\mathbf{p}_1\mathbf{p}_2^T)$ follows easily from

$$\begin{aligned} D(\mathbf{P}||\mathbf{p}_1\mathbf{p}_2^T) &= D(\mathbf{P}||\tilde{\mathbf{p}}_1\tilde{\mathbf{p}}_2^T) + D(\tilde{\mathbf{p}}_1||\mathbf{p}_1) \\ &+ D(\tilde{\mathbf{p}}_2||\mathbf{p}_2) \ge D(\mathbf{P}||\tilde{\mathbf{p}}_1\tilde{\mathbf{p}}_2^T), \end{aligned} \tag{40}$$

which shows that $(\mathbf{p}_1^\star, \mathbf{p}_2^\star) = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2)$. Therefore, marginalization is the solution to the minimum divergence criterion[11]. In order to obtain an approximation of $\mathcal{C}$, it is sufficient to solve (12)-(17), take the marginal distributions of the solution, plug them into (4)-(8), and evaluate (R_1, R_2) (the problem (4)-(8) is convex for fixed input probability distributions). The solution to (4)-(8) in terms of (R_1, R_2) defines the marginalization inner bound $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ (see Algorithm 2).

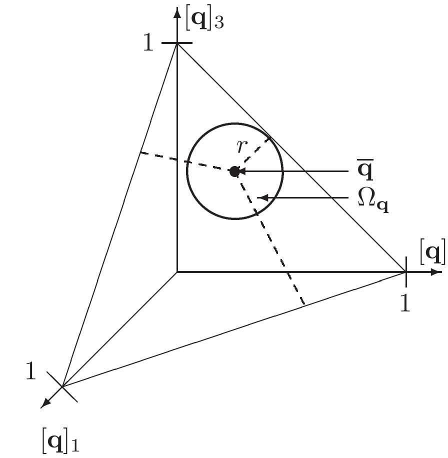


Fig. 2. The support of the randomly generated probability distributions $\mathbf{q}$ is the largest circle centered at $\mathbb{E}\{\mathbf{q}\} = \overline{\mathbf{q}}$ that fits within the probability simplex.

Algorithm 1 Randomization

```
for each value of θ ∈ [0,1] do
    Solve (12)-(17): (R°_1(θ), R°_2(θ)) = (R*_1, R*_2) and P(θ) = P*.
    Approximate P(θ) using (34) for some specified n: {λ_i, a_i, b_i}_{i=1}^n.
    for i = 1 ... n do
        for j = 1 ... Nλ_i do
            Generate according to (39) a random pair (q_1, q_2) such that E{q_1} = a_i, E{q_2} = b_i.
            Evaluate (4)-(8) using (q_1, q_2): (R_1^(i,j), R_2^(i,j)).
        end for
    end for
    Choose the best pair: (R^i_rand,1(θ), R^i_rand,2(θ)) = max_{i,j} θR_1^(i,j) + (1-θ)R_2^(i,j)
end for
Randomization inner bound[10]: R^i_rand = Co({(R^i_rand,1(θ), R^i_rand,2(θ)), ∀θ}).
Outer bound: R° = Co({(R°_1(θ), R°_2(θ)), ∀θ}).
```

[6]Convergence of the objective value (35) to a number (albeit not necessarily the optimum value) follows since it cannot increase in the iterations of alternate maximization and it is lower-bounded by zero. We shall use the limit value as an approximation of the global minimum of (35)-(38), since its lack of convexity prevents us from ensuring convergence to the global (and even to a local) minimum. Notice also that convergence in the objective does not imply convergence in the sequence $\{\lambda_i^{(t)}, \mathbf{a}_i^{(t)}, \mathbf{b}_i^{(t)}\}_{i=1}^n$. However, once convergence in the objective is reached to some prescribed tolerance, any of the potential solutions achieving it is equally useful.

[7]For instance, when $|\mathcal{X}| = 3$, the radius r is set to

$$r = \min\left\{ \frac{[\overline{\mathbf{q}}]_1 + [\overline{\mathbf{q}}]_2 - 1}{\sqrt{2}}, \frac{[\overline{\mathbf{q}}]_1 + [\overline{\mathbf{q}}]_3 - 1}{\sqrt{2}}, \frac{[\overline{\mathbf{q}}]_2 + [\overline{\mathbf{q}}]_3 - 1}{\sqrt{2}}, ||\overline{\mathbf{q}}||_2^2 \right\},$$

where three point-line distances and one point-point distance formula has been used. Point-plane formulas arise in higher dimensions.

[8]To generate a vector $\mathbf{x}$ with uniform distribution on the sphere, we generate random vectors drawn uniformly on the hypercube $\left[[\overline{\mathbf{q}}]_1 - r, [\overline{\mathbf{q}}]_1 + r\right] \times \ldots \times \left[[\overline{\mathbf{q}}]_{|\mathcal{X}|} - r, [\overline{\mathbf{q}}]_{|\mathcal{X}|} + r\right]$ until the condition $\|\mathbf{x} - \overline{\mathbf{q}}\|^2 \le r^2$ is satisfied.

[9]The projection onto a simplex usually has a water-filling form [22], but, since by construction $\mathbf{x}$ belongs to the non-negative orthant, it reduces to (39), where the water level has been analytically found.

[10]$\mathrm{Co}(.)$ denotes convex hull.

[11]Another deterministic strategy is to use the singular value decomposition (SVD) of $\mathbf{P}$ to choose $\mathbf{q}_1$ and $\mathbf{q}_2$ as the suitably normalized left and right singular vectors associated to the largest singular value. However, numerical simulations of this method performed over various channels have shown that it is outperformed by marginalization.

Algorithm 2 Marginalization

```
for each value of θ ∈ [0,1] do
    Solve (12)-(17): (R°_1(θ), R°_2(θ)) = (R*_1, R*_2) and (p_1(θ), p_2(θ)) = (p*_1, p*_2).
    Evaluate (R_1, R_2) (4)-(8) for fixed distributions (p_1(θ), p_2(θ)): (R^i_marg,1(θ), R^i_marg,2(θ)) = (R*_1, R*_2).
end for
Marginalization inner bound: R^i_marg = Co({(R^i_marg,1(θ), R^i_marg,2(θ)), ∀θ}).
Outer bound: R° = Co({(R°_1(θ), R°_2(θ)), ∀θ}).
```

Remark 1: The outer bound $\mathcal{R}^{\circ}$ can be tightened using the algorithm of [12] for the exact computation of the sum-capacity, denoted by C^{sum}. In particular, a tighter outer bound is $\mathcal{R}^{\circ} \cap \mathcal{C}^{\mathsf{sum}}$, where

$$\mathcal{C}^{\mathsf{sum}} = \{(R_1, R_2) \in \mathbb{R}^2 \mid R_1 + R_2 \leq C^{\mathsf{sum}}\}. \tag{41}$$

IV. Performance Analysis of Marginalization

A. Analytical Results

There exists a class of channels for which marginalization computes optimally the capacity region. Although we have not been able to fully characterize this class analytically, we have been able to show that some specific channels belong to it. For some channels, this can be proved by showing $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{R}^{\circ} = \mathcal{C}$, while for others $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} \subset \mathcal{R}^{\circ}$ or $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} \subset \mathcal{C} = \mathcal{R}^{\circ}$. We restrict our attention to the two-user case for the sake of simplicity of the expressions.

It is worth to point out that $\mathcal{R}^{\circ}$ is also an outer bound of the capacity region of the two-user DMAC with feedback (see [25, Sec. IV]), and it follows that the class of channels for which $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}\mathcal{R}^{\circ}$ is a subset of the class of DMACs for which feedback does not increase the capacity region. Although the capacity region of the discrete memoryless MAC with feedback is not known in general (it is known in the continuous memoryless Gaussian (scalar) case [25]), there are some achievability results [26], [27] and a class of DMACs for which the achievable region of [27] is tight (see [28]). Let us start first with some optimality conditions that will be key for obtaining subsequent results.

Lemma 1: Consider a joint probability distribution $P_{X_1X_2}$ satisfying

$$I(X_1;Y|X_2) + I(X_2;Y|X_1) \geq I(X_1X_2;Y). \tag{42}$$

A necessary and sufficient condition for optimality of such $P_{X_1X_2}$ with respect to the relaxed problem (12)-(17) for any fixed $(\theta_1, \theta_2) = (\theta, 1-\theta)$, assuming $\theta_2 \geq \theta_1$, is[12] (43) at the top of the next page, for all $(x_1, x_2) \in (\mathcal{X}_1, \mathcal{X}_2)$ and some $L_{\circ}(\theta_1, \theta_2) \geq 0$. In case (42)-(43) is satisfied by some $P_{X_1X_2}$, all the other optimal distributions, if any, also satisfy (42)-(43) and share the same objective value $L_{\circ}(\theta_1, \theta_2) = R^{\star}_{\circ}(\theta_1, \theta_2) = \max_{(R_1,R_2)\in\mathcal{R}^{\circ}} \theta_1 R_1 + \theta_2 R_2$.

Proof: See Appendix II. ■

Lemma 2: A sufficient condition for optimality of the input probability distributions P_{X_1} and P_{X_2} with respect to the capacity region for any fixed $(\theta_1, \theta_2) = (\theta, 1-\theta)$, assuming $\theta_2 \geq \theta_1$, is[12] (44) at the top of the next page, for all $x_k \in \mathcal{X}_k$, $k = 1, 2$. If (44) is satisfied by some P_{X_1}, P_{X_2} all the other optimal distributions, if any, also satisfy (44) and share the same objective value $L(\theta_1, \theta_2) = C^{\star}(\theta_1, \theta_2) = \max_{(R_1,R_2)\in\mathcal{C}} \theta_1 R_1 + \theta_2 R_2$.

Proof: Lemma 2 follows from the particularization of Lemma 1 to a product distribution of the form $P_{X_1X_2} = P_{X_1}P_{X_2}$, which always satisfies (42). Since such product distribution is a solution to the relaxed problem (12)-(17), it is capacity-achieving and hence optimal. ■

Corollary 1: $\mathcal{R}^{\circ} = \mathcal{C}$ if and only if for each (θ_1, θ_2) there exists at least one pair of distributions satisfying the conditions in Lemma 2.

Proof: The 'if' part is proved by noticing that existence of input distributions satisfying Lemma 2 for all (θ_1, θ_2) is equivalent to $R^{\star}_{\circ}(\theta_1, \theta_2) = C^{\star}(\theta_1, \theta_2)\ \forall(\theta_1, \theta_2)$, which implies $\mathcal{R}^{\circ} = \mathcal{C}$, since both $\mathcal{C}$ and $\mathcal{R}^{\circ}$ are convex sets. The 'only if' part follows from the fact that $\mathcal{R}^{\circ} = \mathcal{C}$ implies that for each (θ_1, θ_2) there must exist at least one product distribution which is optimal with respect to the relaxed problem. Since such product distribution satisfies (42), it must also satisfy (43) and, consequently, Lemma 2. ■

Corollary 2: $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$ if for each (θ_1, θ_2) there exist one or more joint distributions satisfying Lemma 1, all of them with product distributions induced by their marginals satisfying Lemma 2.

Proof: While the relaxed problem (12)-(17) may have multiple solutions not all of them product distributions, Corollary 2 implies that the inner bound $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ of the capacity region provided by the relaxation method in Algorithm 1 is obtained with capacity achieving product distributions for all (θ_1, θ_2). ■

Corollary 3: $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} = \mathcal{R}^{\circ}$ if and only if for each (θ_1, θ_2) there exists at least one pair of distributions satisfying the conditions in Lemma 2 and all the joint (non-product) distributions satisfying Lemma 1 (if any such distributions exist) have product distributions induced by their marginals satisfying Lemma 2.

Proof: This follows from corollaries 1 and 2. ■

B. Applications

The previous results describe the conditions under which the bounds provided by the relaxation method in Algorithm 1 are tight[13], and provide us with a test for quantifying optimality of the marginalization approach. This test can be performed either analytically through any of the corollaries, or numerically (by checking if the product distribution that achieves $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ satisfies the conditions in Lemma 2, for example). In addition, they can also be used as a tool for deriving new capacity results or for re-deriving existing ones. In this respect, the application of Corollary 3 to the binary switching MAC (BS-MAC) [29], [30] shows that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} = \mathcal{R}^{\circ}$, giving an alternative proof of the capacity region of this channel. Additionally, it was shown in [31] that the capacity region of the BS-MAC with and without feedback were identical. Since this result is a necessary condition for allowing the inner and

[12] If $\theta_2 \leq \theta_1$ the user indexes of θ and X in (43) or (44) must be swapped.

[13] Note that the inner and outer bounds are always tight for $\theta = 0$ and $\theta = 1$.

$$\theta_1 D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (\theta_2-\theta_1)D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1})\begin{cases} = L_{\circ}(\theta_1,\theta_2) & \text{if } P_{X_1X_2}(x_1,x_2)>0 \\ \leq L_{\circ}(\theta_1,\theta_2) & \text{if } P_{X_1X_2}(x_1,x_2)=0 \end{cases} \tag{43}$$

$$\theta_1 D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (\theta_2-\theta_1)D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1})\begin{cases} = L(\theta_1,\theta_2) & \text{if } P_{X_1}(x_1)P_{X_2}(x_2)>0 \\ \leq L(\theta_1,\theta_2) & \text{if } P_{X_1}(x_1)P_{X_2}(x_2)=0 \end{cases} \tag{44}$$

outer bounds to coincide, it could have been also inferred from the application of Corollary 3 to this channel.

However, the derived analytical conditions for tightness of the marginalization bounds do not exhaust all the situations for which the marginalization approach is optimal (Lemma 2 is sufficient but not necessary for a capacity-achieving product distribution). Thus, it can happen for some channels that none of the previous results applies but marginalization is still optimal: it may occur that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} \subset \mathcal{R}^{\circ}$ without satisfying the conditions in Corollary 2. A representative example of this subclass of channels is the Binary Adder MAC [8] (BA-MAC or binary erasure MAC, as named in [5, Example 14.3.3]).

The BS-MAC and BA-MAC are the only two non-trivial channels that characterize the rest of binary-input ternary-output[14] deterministic DMACs, which can be obtained through isomorphisms of the input and/or output alphabets and/or user indices in one of these two canonical channels [30]. Hence, marginalization is tight for all of them and, as shown next, for a wider class of channels also.

Theorem 1: Marginalization is tight, i.e. $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$, for all binary-input deterministic DMACs.

Proof: See Appendix III. ∎

Finally, to show that not all the channels for which marginalization is tight are deterministic, we study next a non-deterministic extension of the BS-MAC, the noisy BS-MAC, for which the outer bound is shown to be tight through Corollary 1.

Example - The noisy binary-switching MAC: We denote by noisy binary-switching MAC (nBS-MAC) the binary-input ternary-output ($\mathcal{Y} = \{0, 1, \infty\}$) multiple-access channel characterized by the transition probability distribution

$$P^{\text{nBS-MAC}}_{Y|X_1X_2} \equiv \begin{bmatrix} \delta/2 & \delta/2 & 1-\delta \\ \delta/2 & \delta/2 & 1-\delta \\ 1-\epsilon & \epsilon & 0 \\ \epsilon & 1-\epsilon & 0 \end{bmatrix}, \tag{45}$$

where the columns represent the different elements of $\mathcal{Y}$ and the rows correspond to the natural ordering of the inputs (X_1, X_2). This non-deterministic extension of the BS-MAC adds two random behaviors to the channel: i) a noisy switch, that with probability δ is not able to maintain open the circuit and outputs equally likely bits, and ii) a binary symmetric channel with error probability ϵ when the switch is closed. The BS-MAC is hence obtained by particularizing $(\delta, \epsilon) = (0, 0)$.

Proposition 3: The outer bound is tight for the nBS-MAC, $\mathcal{R}^{\circ} = \mathcal{C}$, and one capacity-achieving pair of distributions achieving the boundary point $(R_1^{\star}(\theta), R_2^{\star}(\theta)) =$ $\arg\max_{(R_1,R_2)\in\mathcal{C}} \ \theta R_1 + (1-\theta)R_2$ is

$$\begin{aligned} P^{\star}_{X_1}(1;\theta) &= p(\delta,\epsilon;\theta) = 1 - P^{\star}_{X_1}(0;\theta), \\ P^{\star}_{X_2}(0;\theta) &= P^{\star}_{X_2}(1;\theta) = 1/2, \end{aligned} \tag{46}$$

where

$$p(\delta,\epsilon;\theta) = \frac{A(\delta,\epsilon;\theta) - \delta/(1-\delta)}{A(\delta,\epsilon;\theta)+1} \tag{47}$$

if $0 < \theta \leq 1/2$ and

$$p(\delta,\epsilon;\theta) = \{0 \leq p < 1 : B(p,\delta,\epsilon;\theta) = \theta(h(\delta) - h(\epsilon) + \log(2))\} \tag{48}$$

otherwise, see (49)-(51) on the next page, where $h(x) \triangleq -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function.

Proof: The proof follows from the application of Corollary 1 to this channel. See Appendix IV. ∎

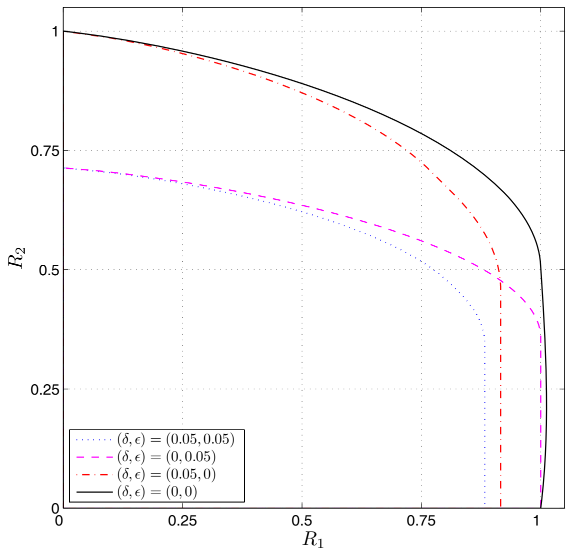


Fig. 3. The capacity region $\mathcal{C}$ of the nBS-MAC, in [bit/ch. use], for different values of δ and ϵ. Note that $(\delta, \epsilon) = (0, 0)$ corresponds to the BS-MAC.

Fig. 3 shows the capacity region of the nBS-MAC for several values of δ and ϵ. Typically, when δ is small, sender 1 has access to the channel in much better conditions than sender 2, which is reflected in the shape of the capacity regions. Thus, when θ is small, the rate of sender 2 is prioritized and hence sender 1 "opens the tap" for the transmission of information of X_2 by setting a large value for $p(\delta, \epsilon; \theta)$. When θ increases, the tap is progressively closed towards a value that maximizes R_1.

V. Numerical Results

To start with, we analyze the numerical performance of randomization and marginalization over three different binary-inputs ternary-output non-deterministic two-user DMACs,

[14]When we refer to M-ary output DMACs we assume that $|\mathcal{Y}| = M$ and all the M values of the output alphabet can be exhausted by at least one input (i.e., there are not dummy output letters).

$$A(\delta,\epsilon;\theta) = \left(\exp(h(\delta)+(1-1/\theta)h(\epsilon))2^{(1/\theta-1)}\right)^{\frac{1}{1-\delta}} \tag{49}$$

$$B(p,\delta,\epsilon;\theta) = (1-\theta)(1-\delta)\log\frac{\delta+(1-\delta)p}{(1-\delta)(1-p)} \tag{50}$$

$$+ (2\theta-1)\Big((1-\epsilon-\delta/2)\log\frac{\delta+(2-2\epsilon-\delta)p}{(1-\delta)(1-p)} + (\epsilon-\delta/2)\log\frac{\delta+(2\epsilon-\delta)p}{(1-\delta)(1-p)}\Big) \tag{51}$$

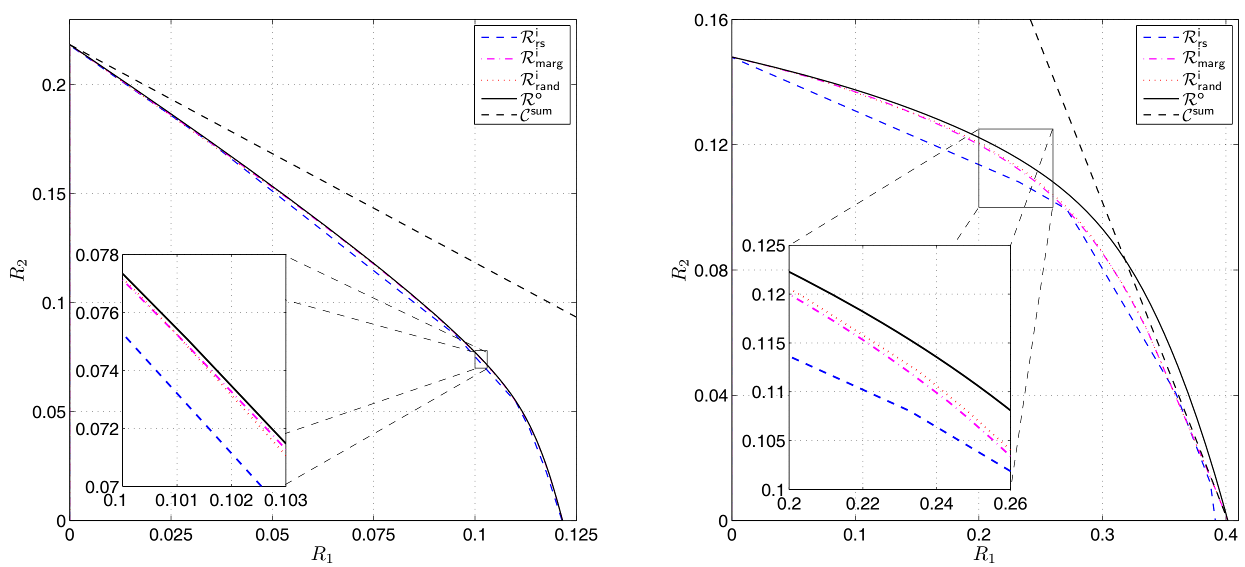


Fig. 4. Bounds of the capacity region for DMAC_1 (left) and DMAC_2 (right). Units are [bit/ch. use].

which we name DMAC_1, DMAC_2, and DMAC_3, characterized respectively by the transition probability distributions in (52). The columns represent the different elements of $\mathcal{Y}=\{0,1,2\}$, and the rows correspond to the natural ordering of the inputs. While DMAC_1 is the multiple-access channel example used in [12] to illustrate the behavior of the algorithm for the computation of the sum-rate capacity, DMAC_2 and DMAC_3 have been chosen randomly. For each of the channels we compute the randomization and marginalization bounds described in Section III, and use algorithm in [12] to compute $\mathcal{C}^{\mathsf{sum}}$. As for the randomization bound, $N=500$ randomly generated product distributions have been tested for each θ using the approximation (34) with $n=4$. Additionally, we consider the achievable region of a random search algorithm, denoted by $\mathcal{R}^{\mathsf{i}}_{\mathsf{rs}}$, as a benchmark (see Algorithm 3).

Fig. 4 shows the bounds for DMAC_1 and DMAC_2. DMAC_1 is another example of a non-deterministic channel for which $\mathcal{R}^{\mathsf{o}}$, $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$, and $\mathcal{R}^{\mathsf{i}}_{\mathsf{rand}}$ coincide, and hence $\mathcal{C}$ can be effectively computed with the proposed methods. As for DMAC_2 and DMAC_3, which represent a more general situation, the bounds

Algorithm 3 Random search

Set $(N,\sigma^2)=(500,1/9)$.
for each value of $\theta\in[0,1]$ **do**
  Set $(R_1(\theta),R_2(\theta),f^\star,\mathbf{p}_1(\theta),\mathbf{p}_2(\theta))=(0,0,0,\mathbf{0},\mathbf{0})$.
  Set $(\mathbf{p}_1^{(0)},\mathbf{p}_2^{(0)})=(\mathbf{1}/|\mathcal{X}_1|,\mathbf{1}/|\mathcal{X}_2|)$.
  for $j=1\ldots N$ **do**
    Generate $(\mathbf{r}_1^{(j)},\mathbf{r}_2^{(j)})$ i.i.d. $\sim\mathcal{N}(0,\sigma^2)$ of lengths $|\mathcal{X}_1|$ and $|\mathcal{X}_2|$, respectively.
    Update[15] $\mathbf{p}_k^{(j)}=[\mathbf{p}_k^{(j-1)}+\mathbf{r}_k^{(j)}]^+$ and normalize $\mathbf{p}_k^{(j)}:=\mathbf{p}_k^{(j)}/(\mathbf{1}^T\mathbf{p}_k^{(j)})$, $k=1,2$.
    Evaluate (4)-(8) using $(\mathbf{p}_1^{(j)},\mathbf{p}_2^{(j)})$: $(R_1^\star(\theta),R_2^\star(\theta))$.
    if $\theta R_1^\star+(1-\theta)R_2^\star>f^\star$ **then**
      $(R_1(\theta),R_2(\theta),f^\star,\mathbf{p}_1(\theta),\mathbf{p}_2(\theta)) = (R_1^\star,R_2^\star,\theta R_1^\star+(1-\theta)R_2^\star,\mathbf{p}_1^{(j)},\mathbf{p}_2^{(j)})$.
    end if
  end for
end for
$\mathcal{R}^{\mathsf{i}}_{\mathsf{rs}}=\mathrm{Co}(\{(R_1(\theta),R_2(\theta)),\ \forall\theta\})$.

[15] $[\mathbf{x}]^+$ denotes the component-wise application of the operator $[x]^+\triangleq\max\{x,0\}$.

$$P^{(1)}_{Y|X_1X_2}\equiv\begin{bmatrix}0.2&0.3&0.5\\0.7&0.2&0.1\\0.5&0.1&0.4\\0.3&0.4&0.3\end{bmatrix},\quad P^{(2)}_{Y|X_1X_2}\equiv\begin{bmatrix}0.4&0.1&0.5\\0.3&0.2&0.5\\0.5&0.4&0.1\\0.2&0.8&0\end{bmatrix},\quad P^{(3)}_{Y|X_1X_2}\equiv\begin{bmatrix}0.1&0.2&0.7\\0.3&0.5&0.2\\0.3&0.4&0.3\\0.8&0.1&0.1\end{bmatrix} \tag{52}$$

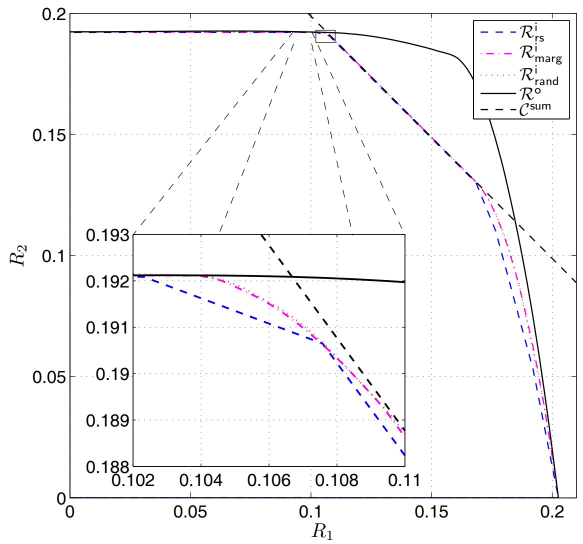


Fig. 5. Bounds of the capacity region for DMAC_3. Units are [bit/ch. use].

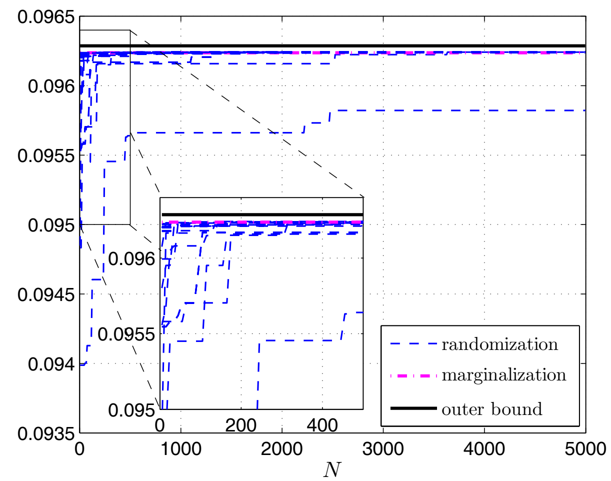


Fig. 6. Weighted sum-rate $\theta R_1 + (1-\theta)R_2$ in [bit/ch. use] achieved by the outer bound, the marginalization inner bound, and several tries of the randomization inner bound as a function of the number of samples, N, for DMAC_1 and $\theta = 0.75$.

do not coincide and $\mathcal{C}$ cannot be directly evaluated. However, if we consider $\mathcal{R}^{\text{o}} \cap \mathcal{C}^{\text{sum}}$ we are able to obtain a much tighter outer bound. Regardless of the tightness of $\mathcal{R}^{\text{o}}$, the accuracy of the sum-capacity offered by the proposed relaxation methods is remarkable for all the channels. The performance of marginalization and randomization is indistinguishable, as shown by appropriate zooms in Figs. 4 and 5, while the behavior of the random search is irregular and in general much worse. To achieve similar performance, it requires a number of distributions under test orders of magnitude above that of randomization.

Next, Fig. 6 explores the performance of randomization as a function of the number of samples, N. By fixing $\theta = 0.75$ in DMAC_1, we evaluate the weighted-sum rate achieved by several tries of the randomization method as a function of the number of samples. While no performance guarantee can be a priori given, $N = 500$ samples seems to be a reasonable tradeoff between performance and complexity in this case. Nevertheless, this has to be compared to marginalization, which is able to provide excellent performance at no computational cost, hence rendering it a reasonable choice to be used in practical applications.

Finally, we explore the effect of quantization of the channel output on a Gaussian MAC of the form

$$Y = h_1 X_1 + h_2 X_2 + Z, \tag{53}$$

where the users' pathloss is different, $h_1 = 1$ and $h_2 = 0.35$. To model the situation where users can either transmit a BPSK symbol or remain silent, with every action carrying information, we choose to use ternary alphabets, i.e. $\mathcal{X}_1 = \mathcal{X}_2 = \{-1, 0, 1\}$. Thus, the AWGN is drawn $Z \sim \mathcal{N}(0, 1/\sqrt{\mathsf{snr}})$, incurring in the slight abuse of notation of using snr to denote the signal-to-noise ratio experienced by either user *when* a BPSK symbol is transmitted. Quantization of the channel output Y to b bits results in a discrete end-to-end channel where the size of the output alphabet is $|\mathcal{Y}| = 2^b$ and the conditional probability distribution depends on the quantization bins. In particular, we perform uniform quantization of Y over the interval $[-2 - 3/\sqrt{\mathsf{snr}}, 2 + 3/\sqrt{\mathsf{snr}}]$. For each choice of snr and b, the resulting transition probability distribution is used to compute the achievable rates.

In this setup, Fig. 7 (left) studies the behavior of the weighted sum rate $\theta R_1 + (1-\theta)R_2$ achieved by the marginalization method when user 2 has a larger weight, $\theta = 0.2$. While the tightness of marginalization depends on snr and b, we can say that, in practical terms, marginalization is optimal for $b \geq 4$. Next, we focus on the optimal codeword distributions resulting from the marginalization method. Given the symmetry of (53), their structure is always of the form

$$\begin{aligned} P_{X_k}(0;\theta) &= \gamma_k, \\ P_{X_k}(-1;\theta) &= P_{X_k}(1;\theta) \\ &= (1-\gamma_k)/2, \quad k = 1,2 \end{aligned} \tag{54}$$

for some $0 \leq \gamma_k \leq 1$. It hence suffices to study the behavior of $P_{X_k}(0,\theta)$, $k = 1, 2$, only. From the curves of Fig. 7 (right), it follows that i) for low snr, the weighted sum rate is maximized when both users are always active; ii) as snr increases, user 1 has to remain silent a fraction of channel uses that decreases with the number of quantization bits, b. This can be used as a guideline to design explicit multiple-access techniques that tune the degree of activity of user 1 as a function of snr and b.

VI. Conclusions

The computation of the capacity region $\mathcal{C}$ of the DMAC is a nonconvex problem. A matrix formulation has been used to concentrate the non-convexity into a matrix rank-one constraint. Since problems with this type of constraint cannot be solved optimally with the current state of the art, suboptimal yet practical methods for the computation of $\mathcal{C}$ have been proposed: randomization and marginalization. These methods, which provide inner and outer bounds of $\mathcal{C}$, are optimal for a class of DMACs for which the capacity region can then be efficiently computed. In particular, analytical optimality conditions for marginalization have been obtained which have

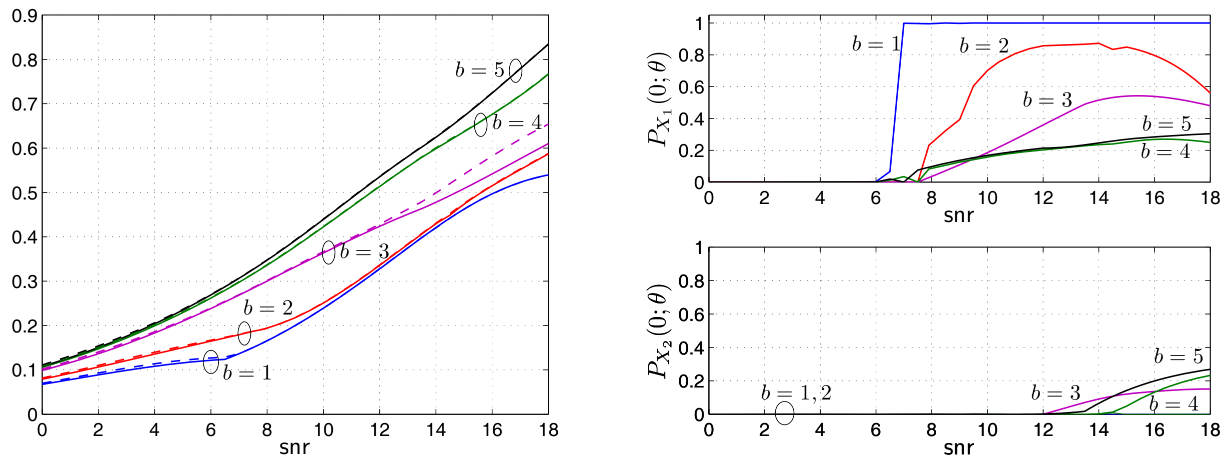


Fig. 7. Weighted sum-rate $\theta R_1 + (1-\theta)R_2$ in [bit/ch. use] for $\theta = 0.2$ achieved by the outer bound (dashed lines) and the marginalization inner bound (solid lines) (left), and the resulting optimal probability of the idle state for both users (right) as a function of the signal-to-noise ratio snr and the number of quantization bits b.

been used to show tightness of the bounds for all the binary-input deterministic DMACs but also other non-deterministic channels such as the nBS-MAC.

Appendix A
Proof of Proposition 1

First, note that the functions f_1, f_2, and f_{12} in (19)-(21) simplify to the right hand sides of (5), (6), and (7), respectively, when the vector-matrix formulation is used for some $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$ with marginal distributions $\mathbf{p}_1$ and $\mathbf{p}_2$. Equivalence between (4)-(8) and (12)-(18) is hence proved thanks to constraint (16) and the equivalence of (17)-(18) and $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$. Regarding the concavity of the function $f_1(\mathbf{P}, \mathbf{p}_2)$ we can rewrite (19) as (55). See next page. The term $H(Y|X_1, X_2)$ is linear in $\mathbf{P}$ and thus concave. The second term of (55) is jointly concave in $(P_{X_2Y}, \mathbf{p}_2)$ (by the same arguments that ensure the convexity of the divergence [5, Thm. 2.7.2]), but since P_{X_2Y} is linear in $\mathbf{P}$, it is also concave in $(\mathbf{P}, \mathbf{p}_2)$, so is the function $f_1(\mathbf{P}, \mathbf{p}_2)$.

It follows by symmetry of (19) and (20) that the function $f_2(\mathbf{P}, \mathbf{p}_1)$ is jointly concave in $(\mathbf{P}, \mathbf{p}_1)$. Finally, the function f_{12} (21) can be shown to be concave resorting to the concavity of the mutual information of a DMAC with respect to the input distribution, in this case represented by $\mathbf{P}$. □

Appendix B
Proof of Lemma 1

Consider the relaxed problem (12)-(17). Since it is convex and satisfies Slater's conditions, the KKT conditions are necessary and sufficient for optimality of any $(\mathbf{P}, \mathbf{p}_1, \mathbf{p}_2)$ [1]. Taking its partial Lagrangian without relaxing the constraints (16)-(17),

$$\begin{aligned}\widetilde{\mathcal{L}}(R_1, R_2, \mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; \boldsymbol{\lambda}) &= \theta R_1 + (1-\theta)R_2 \\ +\lambda_1(f_1(\mathbf{P}, \mathbf{p}_2) - R_1) &+ \lambda_2(f_2(\mathbf{P}, \mathbf{p}_1) - R_2) \\ &+ \lambda(f_{12}(\mathbf{P}) - R_1 - R_2), \quad (56)\end{aligned}$$

and setting its derivatives with respect to R_1 and R_2 equal to zero we obtain

$$\lambda_1^\star = \theta - \lambda \ , \ \lambda_2^\star = (1-\theta) - \lambda. \quad (57)$$

Using (57), (56) admits a simplified form that does not show dependency on (R_1, R_2),

$$\begin{aligned}\widetilde{\mathcal{L}}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; \lambda) &\equiv \widetilde{\mathcal{L}}(R_1, R_2, \mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; [\lambda_1^\star \ \lambda_2^\star \ \lambda]) \\ &= \lambda(f_{12}(\mathbf{P}) - f_1(\mathbf{P}, \mathbf{p}_2) - f_2(\mathbf{P}, \mathbf{p}_1)) \\ &+ \theta f_1(\mathbf{P}, \mathbf{p}_2) + (1-\theta) f_2(\mathbf{P}, \mathbf{p}_1). \quad (58)\end{aligned}$$

By grouping the primal optimization variables in $\mathbf{y} \triangleq (\mathbf{p}_1, \mathbf{p}_2, \mathbf{P})$, an explicit definition of the feasibility domain $\mathcal{D} \triangleq \{\mathbf{y} \mid \text{(16)-(17) are satisfied}\}$ simplifies the application of the saddle-point property as

$$\min_{0 \le \lambda \le \min\{\theta, 1-\theta\}} \max_{\mathbf{y} \in \mathcal{D}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda) = \max_{\mathbf{y} \in \mathcal{D}} \min_{0 \le \lambda \le \min\{\theta, 1-\theta\}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda), \quad (59)$$

where we have used $0 \le \lambda \le \min\{\theta, 1-\theta\}$ according to dual feasibility ($\boldsymbol{\lambda} \succeq \mathbf{0}$) and (57). The dependence of $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda)$ on λ in the inner minimization of the RHS of (59) is linear (58) and hence its optimal value satisfies

$$\lambda^\star(\mathbf{y}) \equiv \lambda^\star(\varphi(\mathbf{y})) = \begin{cases} 0 & \text{if } \varphi(\mathbf{y}) > 0 \\ \min\{\theta, 1-\theta\} & \text{if } \varphi(\mathbf{y}) < 0 \end{cases}, \quad (60)$$

where $\varphi(\mathbf{y}) \triangleq f_{12}(\mathbf{P}) - f_1(\mathbf{P}, \mathbf{p}_2) - f_2(\mathbf{P}, \mathbf{p}_1)$[16]. Thus, the optimal value of the problem is

$$\max_{\mathbf{y} \in \mathcal{D}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star(\mathbf{y})) = \widetilde{\mathcal{L}}(\mathbf{y}^\star; \lambda^\star(\mathbf{y}^\star)). \quad (61)$$

Let us now restrict the proof to the $0 \le \theta < 1/2$ case for the sake of simplicity (similar results are obtained for the case $1/2 \le \theta < 1$, and are included in the conditions of Lemma 2). To compute the optimal value (61) we need to know the sign of $\varphi(\mathbf{y}^\star)$ and adjust $\lambda^\star$ accordingly. However, $\mathbf{y}^\star$ depends in turn on $\lambda^\star$ through the maximization of $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star(\mathbf{y}))$. Therefore, to obtain the optimal $\mathbf{y}^\star$ we should first maximize $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = \min\{\theta, 1-\theta\} = \theta)$ subject to $\varphi(\mathbf{y}) \le 0$, then maximize $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = 0)$ subject to $\varphi(\mathbf{y}) > 0$, and select the $\mathbf{y}$ yielding the maximum objective value among both hypothesis.

Let us start hypothesizing $\varphi(\mathbf{y}^\star) \le 0$, which implies $\lambda^\star = \theta$ (60) and simplifies (58) to

$$\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = \theta) = \theta f_{12}(\mathbf{P}) + (1-2\theta) f_2(\mathbf{P}, \mathbf{p}_1), \quad (62)$$

[16] If $\varphi(\mathbf{y}) = 0$, $\lambda^\star(\mathbf{y})$ can take any value, but it is irrelevant since it does not affect the result.

$$\begin{aligned} f_1(\mathbf{P}, \mathbf{p}_2) &= \sum_{i,j} [\mathbf{P}]_{i,j} \underbrace{\left(\sum_y P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \log P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \right)}_{-H(Y|X_1=x_1^{(i)}, X_2=x_2^{(j)})} \\ &+ \sum_{j,y} \underbrace{\left(\sum_i [\mathbf{P}]_{i,j} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \right)}_{\equiv P_{X_2Y}(x_2^{(j)}, y)} \log \frac{[\mathbf{p}_2]_j}{\sum_{i'} [\mathbf{P}]_{i',j} P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j)})} \\ &= -H(Y|X_1X_2) - \sum_{j,y} P_{X_2Y}(x_2^{(j)}, y) \log \frac{P_{X_2Y}(x_2^{(j)}, y)}{[\mathbf{p}_2]_j} \end{aligned} \tag{55}$$

which should be maximized under the constraint $\varphi(\mathbf{y}) \leq 0$. Instead, we shall perform an unconstrained maximization of (62) and later impose that the solution satisfies non-positivity of φ. Thus, we aim at finding a solution to the problem

$$\underset{\mathbf{p}_1, \mathbf{p}_2, \mathbf{P}}{\text{maximize}} \quad \theta f_{12}(\mathbf{P}) + (1 - 2\theta) f_2(\mathbf{P}, \mathbf{p}_1) \tag{63}$$

$$\text{subject to} \quad \mathbf{P1} = \mathbf{p}_1 \ , \ \mathbf{P}^T\mathbf{1} = \mathbf{p}_2 \tag{64}$$

$$\mathbf{P} \succeq \mathbf{0} \ , \ \mathbf{1}^T\mathbf{P1} = 1, \tag{65}$$

whose Lagrangian is

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \mathbf{\Phi}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \eta) &= \theta f_{12}(\mathbf{P}) + (1 - 2\theta) f_2(\mathbf{P}, \mathbf{p}_1) \\ &+ \mathbf{1}^T(\mathbf{\Phi} \odot \mathbf{P})\mathbf{1} \qquad (66) \\ &+ \boldsymbol{\nu}_1^T(\mathbf{P1} - \mathbf{p}_1) + \boldsymbol{\nu}_2^T(\mathbf{P}^T\mathbf{1} - \mathbf{p}_2) \\ &+ \eta(\mathbf{1}^T\mathbf{P1} - 1), \qquad (67) \end{aligned}$$

where $\mathbf{\Phi} \in \mathbb{R}_+^{|\mathcal{X}_1| \times |\mathcal{X}_2|}$, $\boldsymbol{\nu}_k \in \mathbb{R}^{|\mathcal{X}_k|}$ for $k = 1, 2$, $\eta \in \mathbb{R}$, and $\odot$ denotes Hadamard (element-wise) product. Setting the derivatives of (66)-(67) with respect to $\mathbf{p}_1$ and $\mathbf{p}_2$ equal to zero, the optimal $\{\boldsymbol{\nu}_k\}$ become

$$\boldsymbol{\nu}_1^\star = (1 - 2\theta) \log(e)\mathbf{1}, \ \boldsymbol{\nu}_2^\star = \mathbf{0}. \tag{68}$$

Finally, plugging (68) into (66)-(67) and setting its derivative with respect to $[\mathbf{P}]_{i,j}$ equal to zero, (69) at the top of the next page follows. Now, by complementary slackness, we know that $[\mathbf{\Phi}]_{i,j}[\mathbf{P}]_{i,j} = 0$ which forces

$$[\mathbf{\Phi}]_{i,j} \begin{cases} = 0 & \text{if } [\mathbf{P}]_{i,j} > 0 \\ \geq 0 & \text{if } [\mathbf{P}]_{i,j} = 0 \end{cases}, \tag{70}$$

and allows us to rewrite (69) as (71) (see next page). It can be shown that any $\mathbf{P}$ with marginals $\mathbf{p}_1, \mathbf{p}_2$ satisfying (71) for some $\eta \in \mathbb{R}$ has an associated objective value $\theta \log(e) - \eta$. Let us impose that such a solution to the unconstrained maximization of (62) satisfies also $\varphi(\mathbf{y}) \leq 0$ (the constraint of the current hypothesis under test). We now show by contradiction that the objective value of any distribution with $\varphi(\mathbf{y}) > 0$ is strictly lower. Suppose that the optimal distribution, with optimal objective value $R_o^\star$, is such that $\varphi(\mathbf{y}^\star) > 0$. In this case, $\lambda^\star = 0$ from (60), which results in the objective function

$$\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = 0) = \theta f_1(\mathbf{P}, \mathbf{p}_2) + (1 - \theta) f_2(\mathbf{P}, \mathbf{p}_1), \tag{72}$$

which is maximized by $\mathbf{y}^\star$ under the constraint $\varphi > 0$. Then, since

$$f_1(\mathbf{P}, \mathbf{p}_2) = f_{12}(\mathbf{P}) - f_2(\mathbf{P}, \mathbf{p}_1) - \varphi(\mathbf{y}^\star) < f_{12}(\mathbf{P}) - f_2(\mathbf{P}, \mathbf{p}_1) \tag{73}$$

(74) on next page follows by assumption, which contradicts optimality. Therefore, a probability distribution $\mathbf{y}$ with $\varphi(\mathbf{y}) \leq 0$ satisfying (71) is optimal, its objective value is $\theta \log(e) - \eta = R_o^\star$, and invalidates the existence of other optimal solutions with $\varphi(\mathbf{y}) > 0$. Thus, provided such distribution exists (71) and $\varphi(\mathbf{y}) \leq 0$ becomes necessary for optimality. As a final remark, note that (42) and $\varphi(\mathbf{y}) \leq 0$ are equivalent statements thanks to the equivalence of the functions f_1, f_2, and f_{12} and mutual information. □

APPENDIX C
PROOF OF THEOREM 1

Consider each feasible value of the size of the output alphabet $|\mathcal{Y}| \leq 4$:

- *Binary output* - None of the binary-input binary-output deterministic DMACs has a capacity region dominating the timesharing line joining the points $(\log 2, 0)$ and $(0, \log 2)$[17]. Since marginalization is tight for $\theta \in \{0, 1\}$ and $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ is convexified using the convex hull operation it follows that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$.
- *Ternary output* - The arguments of Section IV-B and the results of [30] allow us to extrapolate the tightness of marginalization from the BS-MAC and the BA-MAC to all the ternary output DMACs.
- *Quaternary output* - The unique arbitrary distribution maximizing simultaneously $I(X_1; Y|X_2), I(X_2; Y|X_1)$, and $I(X_1X_2; Y)$ is $P_{X_1X_2}(x_1, x_2) = 1/4 \ \forall x_1, x_2 \in \{0, 1\}$, which is a product distribution. Hence, $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{R}^{\circ} = \mathcal{C}$. □

APPENDIX D
PROOF OF PROPOSITION 3

It is sufficient to show that there exists a product distribution satisfying Lemma 2 for any $0 < \theta < 1$. Let us simplify notation by using $P_{X_k}(1) = p_k = 1 - \overline{p}_k$, $k = 1, 2$, and assume arbitrarily that $p_2 = \overline{p}_2 = 1/2$.

In the $0 < \theta \leq 1/2$ case the distributions $P_Y = \left\{ \frac{\delta}{2}\overline{p}_1 + p_1/2, \frac{\delta}{2}\overline{p}_1 + p_1/2, (1 - \delta)\overline{p}_1 \right\}$, $P_{Y|X_1=0} = \{\delta/2, \delta/2, 1 - \delta\}$, and $P_{Y|X_1=1} = \{1/2, 1/2, 0\}$ with the arbitrary assumption $p_1, \overline{p}_1 > 0$ reduce (44) to (75) on next page. By setting the left hand sides of (75) equal to each other, we arrive at

$$\theta(1 - \delta) \log \frac{\delta\overline{p}_1 + p_1}{(1 - \delta)p_1} = \theta h(\delta) + (1 - \theta)(\log 2 - h(\epsilon)), \tag{76}$$

[17]The capacity region is upper-bounded by $R_k \leq H(Y|X_k) \leq H(Y) \leq \log 2$, $k = 1, 2$, and $R_1 + R_2 \leq H(Y) \leq \log 2$, which defines the triangle joining the points $(0, 0)$, $(\log 2, 0)$, and $(0, \log 2)$.

$$\sum_y P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \left(\theta \log \frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{i',j'}[\mathbf{P}]_{i',j'}P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j')})} + (1-2\theta)\log\frac{[\mathbf{p}_1]_i P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{j'}[\mathbf{P}]_{i,j'}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j')})}\right) = \theta\log(e) - \eta - [\mathbf{\Phi}]_{i,j} \quad (69)$$

$$\theta D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (1-2\theta)D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1})\begin{cases} = \theta\log(e)-\eta \text{ if } P_{X_1X_2}(x_1,x_2)>0 \\ \leq \theta\log(e)-\eta \text{ if } P_{X_1X_2}(x_1,x_2)=0\end{cases} \quad (71)$$

$$R_\circ^\star = \widetilde{\mathcal{L}}(\mathbf{y}^\star;\lambda^\star=0) = \max_{\mathbf{y}\in\mathcal{D}:\varphi(\mathbf{y})>0}\widetilde{\mathcal{L}}(\mathbf{y};\lambda^\star=0) < \max_{\mathbf{y}\in\mathcal{D}}(\theta f_{12}(\mathbf{P}) + (1-2\theta)f_2(\mathbf{P},\mathbf{p}_1)) = \theta\log(e)-\eta \quad (74)$$

$$C^\star(\theta) = \begin{cases} \theta\big(\delta\log\frac{\delta}{\delta\overline{p}_1+p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)\overline{p}_1}\big) & (x_1,x_2)\in\{(0,0),(0,1)\} \\ \theta\big(\epsilon\log\frac{2\epsilon}{\delta\overline{p}_1+p_1} + (1-\epsilon)\log\frac{2(1-\epsilon)}{(1-\delta)\overline{p}_1}\big) + (1-2\theta)(\log 2 - h(\epsilon)) & (x_1,x_2)\in\{(1,0),(1,1)\}\end{cases} \quad (75)$$

which is satisfied by $p_1 = 1-\overline{p}_1 = p(\delta,\epsilon;\theta)$ as defined in (47). For the case $1/2 < \theta < 1$, the conditional distributions $P_{Y|X_2=0} = \left\{\frac{\delta}{2}\overline{p}_1 + (1-\epsilon)p_1, \frac{\delta}{2}\overline{p}_1 + \epsilon p_1, (1-\delta)\overline{p}_1\right\}$, $P_{Y|X_2=1} = \left\{\frac{\delta}{2}\overline{p}_1 + \epsilon p_1, \frac{\delta}{2}\overline{p}_1 + (1-\epsilon)p_1, (1-\delta)\overline{p}_1\right\}$, and the same arbitrary assumption $p_1, \overline{p}_1 > 0$ allow us to rephrase (44) as

$$(1-\theta)\big(\delta\log\frac{\delta}{\delta\overline{p}_1+p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)\overline{p}_1}\big) + (2\theta-1)\big(\frac{\delta}{2}\log\frac{\delta}{\delta\overline{p}_1+2(1-\epsilon)p_1} + \frac{\delta}{2}\log\frac{\delta}{\delta\overline{p}_1+2\epsilon p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)p_1}\big) = C^\star(\theta) \quad (77)$$

for $(x_1,x_2)\in\{(0,0),(0,1)\}$ and

$$(1-\theta)\big((1-\epsilon)\log\frac{2(1-\epsilon)}{\delta\overline{p}_1+p_1} + \epsilon\log\frac{2\epsilon}{(1-\delta)\overline{p}_1}\big) + (2\theta-1)\big((1-\epsilon)\log\frac{2(1-\epsilon)}{\delta\overline{p}_1+2(1-\epsilon)p_1} + \epsilon\log\frac{2\epsilon}{\delta\overline{p}_1+2\epsilon p_1}\big) = C^\star(\theta) \quad (78)$$

for $(x_1,x_2)\in\{(1,0),(1,1)\}$. By setting the left hand sides of (77)-(78) equal to each other and using $\overline{p}_1 = 1-p_1$ we obtain

$$B(p_1,\delta,\epsilon;\theta) = \theta(h(\delta)-h(\epsilon)+\log 2), \quad (79)$$

where $B(p_1,\delta,\epsilon;\theta)$ is defined in (51). Since

$$B(0,\delta,\epsilon;\theta) = \theta(1-\delta)\log\frac{\delta}{1-\delta} \leq \theta h(\delta) \overset{(a)}{\leq} \theta(h(\delta)-h(\epsilon)+\log 2), \quad (80)$$

where (a) follows from the entropy upper bound $h(\epsilon)\leq\log 2$, and

$$\lim_{p_1\to 1} B(p_1,\delta,\epsilon;\theta) = +\infty > \theta(h(\delta)-h(\epsilon)+\log 2), \quad (81)$$

it follows by continuity that an optimum $p_1\in[0,1)$ satisfying (79) exists for any $1/2\leq\theta<1$. □

REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379-423, pt. II, pp. 623-656, 1948.

[3] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, pp. 14-20, Jan. 1972.

[4] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, pp. 460-473, July 1972.

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

[6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936-3964, Sep. 2006.

[7] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inform. Theory*, 1971.

[8] H. J. Liao, "Multiple access channels," Ph.D. diss., Dept. of Elec. Eng., University of Hawaii, Honolulu 1972.

[9] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 145-152, Jan. 2004.

[10] Y. Watanabe, "The total capacity of two-user multiple-access channel with binary output," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1453-1465, Sep. 1996.

[11] Y. Watanabe and K. Kamoi, "The total capacity of multiple-access channel," in *Proc. IEEE Int. Symp. Information Theory*, June/July 2002, p. 308.

[12] M. Rezaeian and A. Grant, "Computation of total capacity for discrete memoryless multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2779-2784, Nov. 2004.

[13] F. Dupuis, W. Yu, and F. M. J. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information," in *Proc. IEEE Int. Symp. Information Theory*, June/July 2004, p. 181.

[14] P. Apkarian and H. D. Tuan, "A sequential SDP/Gauss-Newton algorithm for rank-constrained LMI problems," in *Proc. IEEE Conf. Decision Control*, Dec. 1999, pp. 2328-2333.

[15] D. Henrion and G. Meinsma, "Rank-one LMIs and Lyapunov's inequality," *IEEE Trans. Autom. Control*, vol. 46, pp. 1285-1288, Aug. 2001.

[16] R. Orsi, U. Helmke, and J. B. Moore, "A Newton-like method for solving rank contrained linear matrix inequalities," in *Proc. IEEE Conf. Decision Control*, Dec. 2004, pp. 3138-3144.

[17] J. Jaldén, C. Martin, and B. Ottersten, "Semidefinite programming for detection in linear systems—optimality conditions and space-time decoding," in *Proc. IEEE ICASSP*, Apr. 2003, vol. 4, pp. 9-12.

[18] W. Ma, T. N. Davidson, K. M. Wong, Z. Luo, and P. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, pp. 912-922, Apr. 2002.

[19] A. Wiesel, Y. C. Eldar, and S. Shamai, "Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 653-656, Sep. 2005.

[20] M. Bierbaum and H. M. Wallmeier, "A note on the capacity region of the multi-access channel," *IEEE Trans. Inf. Theory*, vol. 25, p. 484, July 1979.

[21] J. R. Ruíz-Tolosa and E. Castillo, *From Vectors to Tensors*. Springer, 2005.

[22] D. P. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4661-4674, Dec. 2005.

[23] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Trans. Inf. Theory*, vol. 14, pp. 462-467, May 1968.

[24] H. Ku and S. Kullback, "Approximating discrete probability distributions," *IEEE Trans. Inf. Theory*, vol. 15, pp. 444-447, July 1969.

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

[26] S. I. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 51, pp. 811-833, Mar. 2005.

[27] 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, pp. 292-298, May 1981.

[28] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Inf. Theory*, vol. 28, pp. 93-95, Jan. 1982.

[29] P. Vanroose and E. C. van der Meulen, "Coding for the binary switching multiple access channel," in *Proc. 7th Symp. Inform. Theory Benelux*, 1986, pp. 183-189.

[30] A. J. Vinck, "On the multiple access channel," in *Proc. 2nd Joint Swedish-Soviet Int. Workshop Inform. Theory*, 1985, pp. 24-29.

[31] A. J. Vinck, W. Hoeks, and K. Post, "On the capacity of the 2-user M-ary multiple access channel with feedback," *IEEE Trans. Inf. Theory*, pp. 540-543, July 1985.

Eduard Calvo (S'02) received the Electrical Engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 2004 and 2009, respectively.

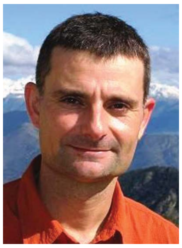
He is an Assistant Professor in the Department of Production, Technology, and Operations Management at IESE Business School, University of Navarra, Spain, which he joined in 2009. In 2004 and 2008, he held research appointments at the Massachusetts Institute of Technology (MIT). He has participated in several EC-funded projects. His research interests lie in the intersection of information theory, optimization theory, statistical signal processing, and their applications to wireless networks and operations research. Dr. Calvo was given the Salvà i Campillo award to the Best Young Engineer by the Guild of Electrical Engineers of Catalonia in 2005.

Daniel P. Palomar (S'99-M'03-SM'08) received the Electrical Engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is an Associate Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. He had previously held several research appointments, namely, at King's College London (KCL), London, UK; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to signal processing and communications.

Dr. Palomar has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of *IEEE Signal Processing Magazine* 2010 Special Issue on "Convex Optimization for Signal Processing," a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on "Game Theory in Communication Systems," and the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks." He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He was the General Co-Chair of the 2009 IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP).

He is a recipient of a 2004/06 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC); the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the best doctoral thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.



Javier Rodríguez Fonollosa (S'90-M'92-SM'98) received the Ph.D. degree in electrical and computer engineering at Northeastern University, Boston, MA, in 1992. In 1993 he joined the Department of Signal Theory and Communications of UPC where he became Associate Professor in 1996, Professor in 2003, and Department Head from 2006 until 2010. Since March 2010, he is manager of the Communications and Electronic Technologies (TEC) area of the National Research Plan of the Ministry of Science and Innovation of Spain.

He is the author of more than 100 papers in the area of signal processing and communications. Since 1995, he has actively participated in European Commission (EC) funded projects. In 1995 he lead UPC's participation in the projects TSUNAMI(II) and SUNBEAM, that included the analysis of adaptive antennas in 2on and 3rd generation mobile communication systems. Since January 2000, he acted as project coordinator of the projects METRA and I-METRA, dedicated to the introduction of multi-antenna terminals in UMTS and systems beyond 3G. From January 2006 to December 2008, he coordinated the EC funded project SURFACE, which evaluated the performance of a generalized air interface with self-configuration capabilities. He has also been actively engaged in the National Research Plan of Spain. Since October 2006, he is project coordinator of the Type C project Fundamental Bounds in Network Information Theory, and since December 2008, he is project coordinator of the CONSOLIDER-INGENIO project COMONSENS on Foundations and Methodologies for Future Communication and Sensor Networks, a 5 year, 3.5 million euro effort of 86 researchers belonging to 10 universities and research centers in Spain.

In June 1995 and September 2001 he was co-chairman and organizer of the IEEE Signal Processing/ATHOS Workshop on Higher-Order Statistics held in Begur, Girona, Spain, and of the IST Mobile Communications Summit 2001 held in Sitges, Barcelona, Spain. He was elected member of the Signal Processing for Communications (SPCOM) Technical Committee of the IEEE Signal Processing Society in January 1999. His research interests include many different aspects of statistical signal processing for communications and information theory. Since May 2005, he is member of the Editorial Board of the *EURASIP Signal Processing Journal*.

Josep Vidal (M'91) received the Telecommunication Engineering and Ph.D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. He was awarded the Premio Extraordinario de Doctorado in 1996.

From 1989 to 1990, he was with the LTS at the Ecole Polytechnique de Lausanne as an Associate Researcher. Since 1993, he has been teaching different topics in telecommunication engineering at UPC. His current research interests are in statistical signal processing and information and communication theory. He has authored more than 120 journal and conference papers in those areas. Since 2000, he has led UPC participation in the EC-funded projects SATURN, FIREWORKS, ROMANTIK, ROCKET adn FREEDOM, and has served as Project Coordinator in the last three.

He was co-organizer of the IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics (Begur, 1995), the IEEE International Symposium on Image/Video Communications over Fixed and Mobile Networks (Rabat, 2000), the IST Mobile Communications Summit, (Sitges, 2001), and the Workshop on Multihop and Adhoc Networking (Brussels, 2003).

He has held research appointments with INP Grenoble (1992), INP Toulouse (2006), and the University of Hawaii (2007).