On MMSE Crossing Properties and Implications in Parallel Vector Gaussian Channels

Ronit Bustin, *Student Member, IEEE*, Miquel Payaró, *Member, IEEE*, Daniel P. Palomar, *Fellow, IEEE*, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*—The scalar additive Gaussian noise channel has the "single crossing point" property between the minimum mean square error (MMSE) in the estimation of the input given the channel output, assuming a Gaussian input to the channel, and the MMSE assuming an arbitrary input. This paper extends the result to the parallel vector additive Gaussian channel in three phases. 1) The channel matrix is the identity matrix, and we limit the Gaussian input to a vector of Gaussian i.i.d. elements. The "single crossing point" property is with respect to the signal-to-noise ratio (as in the scalar case). 2) The channel matrix is arbitrary, and the Gaussian input is limited to an independent Gaussian input. A "single crossing point" property is derived for each diagonal element of the MMSE matrix. 3) The Gaussian input is allowed to be an arbitrary Gaussian random vector. A "single crossing point" property is derived for each eigenvalue of the difference matrix between the two MMSE matrices. These three extensions are then translated to new information theoretic properties on the mutual information, using the I-MMSE relationship, a fundamental relationship between estimation theory and information theory revealed by Guo and coworkers. The results of the last phase are also translated to a new property of Fisher information. Finally, the applicability of all three extensions on information theoretic problems is demonstrated through a proof of a special case of Shannon's vector entropy power inequality, a converse proof of the capacity region of the parallel *degraded* broadcast channel (BC) under an input per-antenna power constraint and under an input covariance constraint, and a converse proof of the capacity region of the compound parallel *degraded* BC under an input covariance constraint.**

***Index Terms*—Entropy power inequality (EPI), Gaussian broadcast channel, Gaussian compound broadcast channel, Gaussian noise, I-MMSE, minimum mean square error (MMSE), multiple-input multiple-output (MIMO), mutual information, parallel vector channel, single crossing point.**

Manuscript received March 26, 2012; accepted September 20, 2012. Date of publication October 18, 2012; date of current version January 16, 2013. R. Bustin was supported in part by the Israel Science Foundation and in part by the Adams Fellowship of the Israel Academy of Sciences and Humanities. S. Shamai (Shitz) was supported by the Israel Science Foundation and the CORNET Consortium and Network of Excellence in Wireless COMmunications (NEWCOM#). M. Payaró was supported in part by the Spanish Ministry of Economy and Competitiveness under project TEC2011-29006-C03-01 (GRE3N-PHY) and in part by the Catalan Government under Grant 2009 SGR 1046. D. P. Palomar was supported by the Hong Kong RGC under Grant 617911. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: bustin@tx.technion.ac.il; sshlomo@ee.technion.ac.il).

M. Payaró is with the Engineering Unit, Centre Tecnològic de Telecomunicacions de Catalunya, Castelldefels 08860, Barcelona, Spain (e-mail: miquel.payaro@cttc.es).

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

Communicated by D. Guo, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2225405

I. Introduction

THIS paper considers parallel vector channels, with an arbitrary input distribution and additive standard Gaussian noise. These channels are a subset of the important family of multiple-input multiple-output (MIMO) additive Gaussian noise channels, which have been extensively investigated in the literature. For most Gaussian channel models studied in information theory, Gaussian signaling happens to be optimal, from point-to-point channels, to multiple-access channels (MAC), and broadcast channels (BC) [1, Chs. 9 and 15], [2], [3]. The methods used to prove this optimality were not easy to come across, even when considering scalar Gaussian channels. For example, in order to prove that Gaussian inputs are optimal for the scalar Gaussian BC, Bergmans employed Shannon's entropy power inequality (EPI) [4]. The solution for the MIMO Gaussian BC came only 30 years later in [2], using a new enhancement approach. Since then, several other proofs were derived, using different tools, such as the extremal inequality in [5], the de Bruijn identity in coordination with Dembo's inequality in [6], and the "single crossing point" property presented by Guo *et al.* in [7]. The "single crossing point" stemmed from the I-MMSE relationship, a fundamental relationship between estimation theory and information theory revealed by Guo *et al.* in [8].

The relationship between estimation theory and information theory goes back to the late 1950s, when Stam [9] used the de Bruijn's identity to prove Shannon's EPI, and then in the early 1970s when the mutual information was represented as a function of the causal filtering error by Duncan [10] and Kadota *et al.* [11]. The I-MMSE relationship, given for discrete-time and continuous-time, scalar and vector additive Gaussian noise channels, deepens the connection between these two fields [12]. Specifically, for a scalar additive Gaussian noise channel

$$Y = \sqrt{\mathsf{snr}}X + N \tag{1}$$

where N is standard Gaussian additive noise, then, regardless of the input distribution of X, the mutual information, $I(X;Y)$, and minimum mean square error (MMSE) in the estimation of X given the observation Y, $\mathsf{mmse}(X, \mathsf{snr})$, are related (assuming real-valued inputs and outputs) by

$$\mathsf{D}_{\mathsf{snr}} I(X;Y) = \frac{1}{2}\mathsf{mmse}(X, \mathsf{snr}) \tag{2}$$

where $\mathsf{D}_{\mathsf{snr}}$ is the derivative with respect to snr, and

$$\mathsf{mmse}(X,\mathsf{snr}) = \mathsf{E}\left\{(X - \mathsf{E}\left\{X \mid \sqrt{\mathsf{snr}}X + N\right\})^2\right\}. \quad (3)$$

The work in [8] has been extended in several directions, among which we have the additive Poisson noise channel [13], [14], the general additive noise channel [15], arbitrary channels [16], representation of the relative entropy as a function of the difference between the mismatched MMSE and the matched MMSE in [17] and [18], and others. One important extension, on which we heavily rely, is the one done by Palomar and Verdú in [19], where they obtain the gradient of the mutual information with respect to different parameters of the MIMO channel.

Going back to the "single crossing point" property, one of the goals in [7] was to show the applicability of the I-MMSE relationship as a tool to solve information-theoretic problems. Specifically, the authors of [7] examined the scalar Gaussian BC and gave an alternative proof for the optimality of Gaussian inputs. In order to show this, Guo *et al.* defined the following function in [7]:

$$f(X,\gamma) = (1+\gamma)^{-1} - \mathsf{mmse}(X,\gamma) \quad (4)$$

where the simplified notation $f(\gamma)$ will be used when there is no confusion about the distribution of X. It was shown that $f(\gamma)$ has *at most* a single crossing point of the horizontal axis. In other words, the first term, which is the MMSE assuming a standard Gaussian input, may be smaller than the second term in some range of SNR values (note that the parameter γ is the SNR); however, once the two terms are equal, at some γ_0, the MMSE of the standard Gaussian input remains greater than the MMSE of the arbitrary input for all $\gamma > \gamma_0$, and the function remains nonnegative. This property, together with the I-MMSE relationship, provides the missing link to derive a simple and elegant converse proof of the capacity region of the scalar Gaussian BC.

The "single crossing point" was derived only for the scalar additive Gaussian channel, as can be seen from the definition of the function $f(\gamma)$. The motivation of this study is to extend this property to the vector Gaussian channel. We consider the following general channel model:

$$\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \boldsymbol{N} \quad (5)$$

where $\boldsymbol{N}$ is a standard Gaussian random vector and $\mathbf{H}$ is a square and diagonal channel matrix known to the receiver(s). In the vector case, the scalar MMSE does not capture all the needed information, and we need to resort to the matrix extension, the MMSE matrix defined as

$$\mathbf{E}_{\boldsymbol{X}} = \mathsf{E}\left\{(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \mathbf{H}\boldsymbol{X} + \boldsymbol{N}\right\})(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \mathbf{H}\boldsymbol{X} + \boldsymbol{N}\right\})^{\mathsf{T}}\right\} \quad (6)$$

from which we can see that, in general, the MMSE matrix $\mathbf{E}_{\boldsymbol{X}}$ depends on the channel $\mathbf{E}_{\boldsymbol{X}} = \mathbf{E}_{\boldsymbol{X}}(\mathbf{H})$, but whenever the channel coefficients depend on other parameters $\mathbf{H} = \mathbf{H}(\boldsymbol{\phi})$, we will write $\mathbf{E}_{\boldsymbol{X}}(\boldsymbol{\phi})$. Observe that the standard scalar MMSE value in the vector case can be easily recovered from the MMSE matrix as follows:

$$\begin{aligned}\mathsf{mmse}(\boldsymbol{X},\mathsf{snr}) &= \mathsf{E}\left\{\|\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \sqrt{\mathsf{snr}}\boldsymbol{X} + \boldsymbol{N}\right\}\|^2\right\} \\ &= \mathsf{Tr}(\mathbf{E}_{\boldsymbol{X}}(\sqrt{\mathsf{snr}}\mathbf{I}_n)). \quad (7)\end{aligned}$$

TABLE I
THREE PHASES OF THE "SINGLE CROSSING POINT" PROPERTY EXTENSION DONE IN THIS PAPER

Phase	$\mathbf{H}$	$\mathbf{R}_{\boldsymbol{G}}$
I	$\mathbf{H} = \sqrt{\mathsf{snr}}\mathbf{I}_n$	i.i.d. , $\mathbf{R}_{\boldsymbol{G}} = \sigma^2\mathbf{I}_n$
II	$\mathbf{H}$ diagonal	independent $\mathbf{R}_{\boldsymbol{G}} = \mathbf{\Lambda}_{\boldsymbol{G}}$
III	$\mathbf{H}$ diagonal	general $\mathbf{R}_{\boldsymbol{G}}$

For the important case when the input distribution of $\boldsymbol{X}$ is Gaussian with covariance matrix $\mathbf{R}_{\boldsymbol{G}}$, we will use the following notation:

$$\mathbf{E}_G(\mathbf{R}_{\boldsymbol{G}},\mathbf{H}) = (\mathbf{R}_{\boldsymbol{G}}^{-1} + \mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1} \quad (8)$$

where we assumed that $\mathbf{R}_{\boldsymbol{G}}$ is of full rank. As in the case of $\mathbf{E}_{\boldsymbol{X}}$, whenever the channel coefficients depend on other parameters $\mathbf{H} = \mathbf{H}(\boldsymbol{\phi})$, we will write $\mathbf{E}_G(\mathbf{R}_{\boldsymbol{G}},\boldsymbol{\phi})$. Another important quantity is the MMSE given for a specific output, $\boldsymbol{Y} = \boldsymbol{y}$, defined as

$$\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{y}) = \mathsf{E}\left\{(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \boldsymbol{y}\right\})(\boldsymbol{X} - \mathsf{E}\left\{\boldsymbol{X} \mid \boldsymbol{y}\right\})^{\mathsf{T}} | \boldsymbol{y}\right\}. \quad (9)$$

Although not specified explicitly, $\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ depends on the channel matrix/parameters. Note that $\mathbf{E}_{\boldsymbol{X}}(\mathbf{H}) = \mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}$. Interestingly, when the input distribution of $\boldsymbol{X}$ is Gaussian, $\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ is independent of $\boldsymbol{y}$ and the following equality holds for all $\boldsymbol{y}$: $\mathbf{E}_G(\mathbf{R}_{\boldsymbol{G}},\mathbf{H}) = \mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$. Finally, given all these quantities we can define the main player in this study: the MMSE matrix difference (analog to $f(\gamma)$ in the scalar case)

$$\begin{aligned}\mathbf{Q}(\boldsymbol{X},\mathbf{R}_{\boldsymbol{G}},\boldsymbol{\phi}) &= \mathbf{E}_G(\mathbf{R}_{\boldsymbol{G}},\boldsymbol{\phi}) - \mathbf{E}_{\boldsymbol{X}}(\boldsymbol{\phi}) \quad (10)\\ &= \mathbf{E}_{\boldsymbol{X}_{\boldsymbol{G}}}(\boldsymbol{\phi}) - \mathbf{E}_{\boldsymbol{X}}(\boldsymbol{\phi}) \quad (11)\\ &= (\mathbf{R}_{\boldsymbol{G}}^{-1} + (\mathbf{H}(\boldsymbol{\phi}))^{\mathsf{T}}\mathbf{H}(\boldsymbol{\phi}))^{-1} - \mathbf{E}_{\boldsymbol{X}}(\boldsymbol{\phi}) \quad (12)\end{aligned}$$

where, similarly to the scalar case in (4), we will use the simplified notation $\mathbf{Q}(\boldsymbol{\phi})$ when the distribution of $\boldsymbol{X}$ and the covariance matrix of the Gaussian distribution $\mathbf{R}_{\boldsymbol{G}}$ are clear from the context. Note that there is *no* requirement that the covariance of the random vector $\boldsymbol{X}$ be equal to $\mathbf{R}_{\boldsymbol{G}}$.

The extension of the "single crossing point" to the vector Gaussian channel (5) is done in three phases, summarized in Table I. Note that in all three phases, the arbitrary input distribution over $\boldsymbol{X}$ is general.

In the first phase, the dependence remains on a scalar parameter—the snr. This is obtained by setting $\mathbf{H} = \sqrt{\mathsf{snr}}\mathbf{I}$ in the general MIMO model in (5). We further limit our observation to the comparison of an arbitrary input distribution with the subset of Gaussian random vectors with i.i.d. elements. For this case, we show that the "single crossing point" property extends smoothly to any linear function of the form $\mathsf{Tr}\left(\mathbf{A}\mathbf{Q}(\boldsymbol{X},\sigma^2\mathbf{I}_n,\gamma)\right)$, where $\mathbf{A}$ is a positive-semidefinite matrix. Although this is the simplest scalar-to-vector extension, the proof is not straightforward. In order to demonstrate the applicability of this result, we extend the proof of the special case of Shannon's EPI, where one of the two random variables is Gaussian, done in [7], to the vector case.

Proceeding with the scalar-to-vector extension, we assume that the channel matrix $\mathbf{H}$ is diagonal; thus, our dependence is now on a vector parameter. In this setting, we have two distinct results, given in phases 2 and 3, that cannot be trivially deduced from each other. In phase 2, we limit the Gaussian distribution, to which we compare, to any independent Gaussian distribution characterized by its diagonal covariance matrix, $\mathbf{\Lambda}_{\boldsymbol{G}}$. Under this assumption, we show that a "single crossing point" property exists for each and every diagonal element of $\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, \boldsymbol{\phi})$. This is not a straightforward extension of the scalar property, since the elements of the random input vector $\boldsymbol{X}$ are, in general, not independent. Together with the I-MMSE relationship, this result provides some interesting properties of the mutual information, and its applicability is demonstrated by providing a simple converse proof for the parallel Gaussian BC capacity region under an input per-antenna power constraint.

The third phase, which is the main result of this study, does not require any further assumptions (apart from the diagonal channel matrix). That is, we compare an arbitrary input distribution with any general Gaussian input distribution, with covariance $\mathbf{R}_{\boldsymbol{G}}$. In this setting, we show that a "single crossing point" property exists for each and every eigenvalue of the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, \boldsymbol{\phi})$. Surely, this is not a straightforward extension of any of the previous results. Moreover, the results of phase 2 cannot be trivially deduced from the results of phase 3, since restricting only the Gaussian covariance to be diagonal does not guarantee that the eigenvalues of $\mathbf{Q}$ will be on its diagonal. The applicability of this result is demonstrated with two information-theoretic problems: the converse proof of the parallel Gaussian BC capacity region under an input covariance constraint and the converse proof of the compound parallel Gaussian BC capacity region under an input covariance constraint.

All three results fall back to the scalar "single crossing point" property result [7], [20] when both the arbitrary input vector $\boldsymbol{X}$ and the Gaussian input random vector are restricted to have independent elements.

Much of this work regards the behavior of functions around zeros, the existence and amount of actual crossings of the horizontal axis. Thus, before proceeding with the technical content of this paper and, in order to make these observations rigorous, we require the next definitions which will be used throughout this paper.

Definition 1: Given a function $h(t)$ continuous within a neighborhood of t_0, we say that a negative-to-nonnegative zero crossing occurs at $t = t_0$ if, and only if, $h(t_0) = 0$ and there exists a positive value ϵ such that $h(t) < 0$ for $t \in (t_0 - \epsilon, t_0)$ and $h(t) \geq 0$ for $t \in (t_0, t_0 + \epsilon)$.

Definition 2: Given a function $h(t)$ continuous within a neighborhood of t_0, we say that a nonnegative-to-negative zero crossing occurs at $t = t_0$ if, and only if, $h(t_0) = 0$ and there exists a positive value ϵ such that $h(t) \geq 0$ for $t \in (t_0 - \epsilon, t_0)$ and $h(t) < 0$ for $t \in (t_0, t_0 + \epsilon)$.

Similar definitions can be given for positive-to-nonpositive and nonpositive-to-positive zero crossings. Another required definition is the following:

Definition 3: Given a function $h(t)$ continuous within a neighborhood of t_0, we say that a negative–zero–positive crossing occurs at $t = t_0$ if, and only if, a negative-to-nonnegative zero crossing occurs at $t = t_0$ and there exists a positive δ such that $h(t) = 0$ for $t \in (t_0, t_0 + \delta)$ and a nonpositive-to-positive zero crossing occurs at $t_0 + \delta$.

Similarly, we can define a positive–zero–negative crossing.

The remaining of this paper is organized as follows. Section II considers the first phase of our extension from scalar to vector, in which case the dependence is on the scalar parameter, snr. In Section III, we provide the framework in which we handle the assumption of a diagonal channel matrix, $\mathbf{H}$. This framework is relevant for phases 2 and 3 of our scalar-to-vector extension. In Section IV, we consider phase 2 of our extension, where we limit our observations to an independent Gaussian input distribution. Section V considers phase 3, where we compare the arbitrary input to any general Gaussian input distribution.

Notation: Straight boldface denotes multivariate quantities such as vectors (lowercase) and matrices (uppercase). Uppercase italic denotes random variables (boldface if we consider random vectors rather than random variables), and their realizations are represented by lowercase italics. The set of n-dimensional positive-semidefinite matrices is denoted by $\mathbb{S}_{+}^{n}$. The elements of a matrix $\mathbf{A}$ are represented by $[\mathbf{A}]_{ij}$. The operator $\mathbf{diag}\,(\mathbf{A})$ represents a column vector with the diagonal entries of matrix $\mathbf{A}$, and $\mathbf{Diag}\,(\mathbf{a})$ represents a diagonal matrix whose nonzero elements are given by the elements of vector $\mathbf{a}$. The superscript $(\cdot)^{\mathsf{T}}$ denotes the transpose. The operator $\mathsf{Tr}(\cdot)$ denotes the trace function, and $|\cdot|$ denotes the determinant function. The operator $\circ$ denotes the Schur product, that is $[\mathbf{A} \circ \mathbf{B}]_{ij} = [\mathbf{A}]_{ij}[\mathbf{B}]_{ij}$. The operator $\mathsf{D}_{\boldsymbol{\gamma}}\mathbf{A}$ denotes the Jacobian matrix of $\mathbf{A}$ with respect to $\boldsymbol{\gamma}$ [21]. The operator $\nabla_{\mathbf{A}} f(\mathbf{A})$, for any differentiable function $f : \mathbb{R}^{k \times n} \to \mathbb{R}$, and $k \times n$ matrix $\mathbf{A}$, is a $k \times n$ matrix with the following (i, j) elements: $\mathsf{D}_{[\mathbf{A}]_{ij}} f(\mathbf{A})$.

Note that we also consider the conditioned version of the previously defined quantities. That is, when the random vector $\boldsymbol{X}$ depends on the random vector $\boldsymbol{U}$, we require, for example, a conditioned version for the MMSE and the matrix $\mathbf{Q}$ given for a specific value of $\boldsymbol{U} = \boldsymbol{u}$. In this case, both quantities depend on an additional parameter $\boldsymbol{u}$, i.e., $\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(\boldsymbol{\phi}, \boldsymbol{u})$ and $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U} = \boldsymbol{u}, \mathbf{R}_{\boldsymbol{G}}, \boldsymbol{\phi})$ (the precise definitions given in Section IV-B).

II. Scalar MIMO Channel

As pointed out in Section I, we begin our study with the simplest multivariate extension of the result in [7, Prp. 16], that is, we consider that the scalar random variables involved in the model in (1) become random vectors. In other words, in this section, we consider the following model:

$$\boldsymbol{Y} = \sqrt{\mathsf{snr}}\boldsymbol{X} + \boldsymbol{N} \tag{13}$$

where the input random vector $\boldsymbol{X} \in \mathbb{R}^n$ is arbitrarily distributed and $\boldsymbol{N} \in \mathbb{R}^n$ follows a standard Gaussian distribution. Observe that (13) is obtained by setting $\mathbf{H} = \sqrt{\mathsf{snr}}\mathbf{I}_n$ in the vector model in (5).

Moreover, we further limit our discussion in this section to the comparison with a Gaussian input with i.i.d. elements, i.e., we assume that $\mathbf{R}_{\boldsymbol{G}} = \sigma^2\mathbf{I}_n$.

Thus, for the settings in this section, the general MMSE matrix difference function in (12) simplifies to

$$\mathbf{Q}(\boldsymbol{X}, \sigma^2\mathbf{I}_n, \gamma) = \frac{\sigma^2}{1 + \sigma^2\gamma}\mathbf{I}_n - \mathbf{E}_{\boldsymbol{X}}(\gamma) \tag{14}$$

where γ plays the role of the estimation SNR.

A. Single Crossing Point

Motivated by the "single crossing property" of $f(X,\gamma)$ presented in [7, Prp. 16], an immediate question that comes to mind is "does this property extend to the MIMO scenario?" Our hypothesis was that for the setting in (13) this property will have a simple extension. Thus, we examine the simplest scalar function of the MMSE matrix difference function of (14), that is, we consider some linear combination of it. Accordingly, we define

$$q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma) = \mathsf{Tr}\left(\mathbf{A}\mathbf{Q}(\boldsymbol{X},\sigma^2\mathbf{I}_n,\gamma)\right) \tag{15}$$

$$= \frac{\sigma^2}{1+\sigma^2\gamma}\mathsf{Tr}\left(\mathbf{A}\right) - \mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right) \tag{16}$$

where $\mathbf{A}$ is a weighting matrix.

The "single crossing point" property of $f(\gamma)$ extends naturally to the function $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$, for a specific subset of matrices $\mathbf{A}$. This result is given in the next theorem.

Theorem 1: Let $\mathbf{A} \in \mathbb{S}_+^n$ be a positive-semidefinite matrix. Then, the function $\gamma \mapsto q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$, defined in (16), has no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $\gamma \in [0,\infty)$.

Moreover, assume that $\mathsf{snr}_0 \in [0,\infty)$ is a negative-to-nonnegative crossing point. Then:

1) $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,0) \leq 0$;
2) $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ is a strictly increasing function in the range $\gamma \in [0,\mathsf{snr}_0)$;
3) $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma) \geq 0$ for all $\gamma \in [\mathsf{snr}_0,\infty)$;
4) $\lim_{\gamma\to\infty} q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma) = 0$.

Proof: We start with the following three lemmas that are instrumental for this proof.

Lemma 1: Let $\mathbf{A} \in \mathbb{S}_+^n$ be a positive-semidefinite matrix and let the random vector $\boldsymbol{X} \in \mathbb{R}^n$ be arbitrarily distributed. Then, we can always find a random vector $\hat{\boldsymbol{X}} \in \mathbb{R}^n$ such that the number of nonnegative-to-negative and negative-to-nonnegative zero crossings of $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ is the same as those of $q_{\mathbf{I}_n}(\hat{\boldsymbol{X}},\sigma^2,\gamma)$.

Proof: See Appendix A1. ■

Lemma 2: Let $\boldsymbol{X} \in \mathbb{R}^n$ be a random vector such that $\mathsf{Tr}(\mathbf{R}_{\boldsymbol{X}})/n \leq \sigma^2$. Then, for every $\gamma \geq 0$, we have

$$\frac{\mathsf{Tr}(\mathbf{E}_{\boldsymbol{X}}(\gamma))}{n} \leq \frac{\sigma^2}{1+\sigma^2\gamma} \tag{17}$$

with equality if and only if $\boldsymbol{X}$ is a Gaussian vector with i.i.d. elements of variance σ^2.

Proof: See Appendix A2. ■

Lemma 3: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. The derivative of the function $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ with respect to γ is given by

$$\mathsf{D}_\gamma q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma) = \mathsf{Tr}\left(\mathbf{A}\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})^2\right\}\right) - \frac{\sigma^4}{(1+\sigma^2\gamma)^2}\mathsf{Tr}\left(\mathbf{A}\right). \tag{18}$$

Proof: See Appendix A3. ■

With these three lemmas at hand, we are now ready to continue with the proof of Theorem 1.

Since we are assuming that the matrix $\mathbf{A}$ is positive semidefinite and the distribution of $\boldsymbol{X}$ is arbitrary, from Lemma 1, we see that we can restrict our study of $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ to that of $q_{\mathbf{I}_n}(\hat{\boldsymbol{X}},\sigma^2,\gamma)$. For the sake of simplicity, throughout this proof we will use $q(\sigma^2,\gamma) = q_{\mathbf{I}_n}(\hat{\boldsymbol{X}},\sigma^2,\gamma)$.

Now, according to Lemma 2, for the case where $\mathsf{Tr}(\mathbf{R}_{\boldsymbol{X}})/n < \sigma^2$, the function $q(\sigma^2,\gamma)$ has no zeros and the statement in Theorem 1 is true. In addition, if $\boldsymbol{X}$ is Gaussian distributed with covariance matrix equal to $\sigma^2\mathbf{I}_n$, then $q(\sigma^2,\gamma) = 0\ \forall\gamma$, which is also consistent with Theorem 1.

Thus, from this point, we can assume that $\mathsf{Tr}(\mathbf{R}_{\boldsymbol{X}})/n \geq \sigma^2$ and that $\boldsymbol{X}$ is not a Gaussian vector with covariance matrix $\sigma^2\mathbf{I}_n$. Now, for $\gamma = 0$ we have $q(\sigma^2,0) = \sigma^2 - \mathsf{Tr}(\mathbf{R}_{\boldsymbol{X}})/n \leq 0$ as required.

From the smoothness of $q(\sigma^2,\gamma)$ as a function of γ, as done in [7, Prp. 16], in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $q(\sigma^2,\gamma)$ can occur, we only need to show that the derivative of $q(\sigma^2,\gamma)$ is positive for all values of γ for which $q(\sigma^2,\gamma) < 0$. Observe that $q(\sigma^2,\gamma) < 0$ implies that

$$n\frac{\sigma^2}{1+\sigma^2\gamma} < \mathsf{Tr}(\mathbf{E}_{\boldsymbol{X}}(\gamma)) = \mathsf{Tr}\left(\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\right). \tag{19}$$

Now, particularizing Lemma 3 for $\mathbf{A} = \mathbf{I}_n$, we have that

$$\begin{aligned}
&\mathsf{D}_\gamma q(\sigma^2,\gamma)\\
&= \mathsf{Tr}\left(\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})^2\right\}\right) - n\frac{\sigma^4}{(1+\sigma^2\gamma)^2} &(20)\\
&> \mathsf{Tr}\left(\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})^2\right\}\right) - \frac{\left(\mathsf{Tr}\left(\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\right)\right)^2}{n} &(21)\\
&= \mathbf{1}^\mathsf{T}\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\circ\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\mathbf{1}\\
&\quad - \mathbf{1}^\mathsf{T}\frac{\mathsf{E}\left\{\mathbf{diag}\left(\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right)\right\}\mathsf{E}\left\{\mathbf{diag}\left(\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right)^\mathsf{T}\right\}}{n}\mathbf{1} &(22)\\
&\geq \mathsf{E}\left\{\mathbf{1}^\mathsf{T}\left(\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\circ\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right.\right.\\
&\quad \left.\left. - \frac{\mathbf{diag}\left(\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right)\mathbf{diag}\left(\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right)^\mathsf{T}}{n}\right)\mathbf{1}\right\} &(23)\\
&\geq 0 &(24)
\end{aligned}$$

where (21) follows directly from (19); in (22), we have defined $\mathbf{1}$ as the column vector whose entries are all 1s; (23) follows from Jensen's inequality; finally, (24) follows from [22, Prp. H.9].

Observe that the inequality in (24), which holds for values of γ such that $q(\sigma^2,\gamma) < 0$, also proves the second item in Theorem 1 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Furthermore, regarding the fourth item, it is clear that $\lim_{\gamma\to\infty} q(\sigma^2,\gamma) = 0$, as both terms in $q(\sigma^2,\gamma)$ tend to zero. ■

Remark 1: Note that the aforementioned theorem also holds for the normalized function, $\frac{1}{n}q(\sigma^2,\gamma)$. Specifically, for the case of $\mathbf{A} = \mathbf{I}_n$, this is simply the difference between the MMSE of a general Gaussian random variable, with variance σ^2, and the *average* MMSE of the n elements of the random vector $\boldsymbol{X}$.

Remark 2: For negative-semidefinite $\mathbf{A}$, it can easily be seen from the proof of Lemma 1 that $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ has the inverse properties, since it is a mirroring of some $q_{\mathbf{I}_n}(\hat{\boldsymbol{X}},\sigma^2,\gamma)$ over the *x*-axis. This is to say that it has at most a single positive-to-nonpositive zero crossing and, if such a crossing exists, $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ will be nonnegative at $\gamma = 0$, strictly decreasing up to the crossing, nonpositive after the crossing, and will tend to zero as $\gamma \to \infty$.

Remark 3: For indefinite $\mathbf{A}$, "single crossing point" properties, such as those shown in Theorem 1, do not hold in general.

B. Application: A Proof of the Special Case of Shannon's Vector EPI, Where One of the Two Random Vectors Is Gaussian

We now show that Theorem 1 can be used to prove the special case of Shannon's EPI, where one of the two random vectors is Gaussian [1, Th. 17.7.3], similarly as it was done in [7] for the scalar case. Precisely, we will show that

$$\exp\left(\frac{2}{n}h\left(\boldsymbol{X}+\boldsymbol{N}\right)\right) \geq \exp\left(\frac{2}{n}h\left(\boldsymbol{X}\right)\right) + 2\pi\mathrm{e}|\mathbf{R}_{\boldsymbol{N}}|^{\frac{1}{n}} \quad (25)$$

for any independent n-dimensional vectors $\boldsymbol{X}$ and $\boldsymbol{N}$ as long as the differential entropy of $\boldsymbol{X}$ is well defined and $\boldsymbol{N}$ is Gaussian distributed with a positive-definite covariance matrix $\mathbf{R}_{\boldsymbol{N}}$.

We define $\boldsymbol{Z}$ to be an n-dimensional Gaussian vector with covariance $\mathbf{R}_{\boldsymbol{Z}} = \mathbf{R}_{\boldsymbol{N}}$ and independent of both $\boldsymbol{X}$ and $\boldsymbol{N}$. Thus, without making any assumptions on the covariance matrix of $\boldsymbol{X}$, we can find $\alpha \in [0,\infty)$ such that the following equality holds:

$$h\left(\boldsymbol{X}\right) = h\left(\alpha\boldsymbol{Z}\right) = \frac{1}{2}\log\left((2\pi\mathrm{e})^n\alpha^{2n}|\mathbf{R}_{\boldsymbol{N}}|\right). \quad (26)$$

Since $\mathbf{R}_{\boldsymbol{N}}$ is positive definite, there exists an invertible matrix $\mathbf{V}$ such that $\mathbf{R}_{\boldsymbol{N}} = \mathbf{V}\mathbf{V}^{\mathsf{T}}$. Defining $\tilde{\boldsymbol{X}} = \mathbf{V}^{-1}\boldsymbol{X}$, $\tilde{\boldsymbol{Z}} = \mathbf{V}^{-1}\boldsymbol{Z}$, and $\tilde{\boldsymbol{N}} = \mathbf{V}^{-1}\boldsymbol{N}$, we have the following chain of equalities:

$$\begin{aligned}
\Delta I(\mathsf{snr}) &= I\left(\alpha\boldsymbol{Z};\sqrt{\mathsf{snr}}\alpha\boldsymbol{Z}+\boldsymbol{N}\right) - I\left(\boldsymbol{X};\sqrt{\mathsf{snr}}\boldsymbol{X}+\boldsymbol{N}\right) && (27)\\
&= I\left(\alpha\tilde{\boldsymbol{Z}};\sqrt{\mathsf{snr}}\alpha\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{N}}\right) - I\left(\tilde{\boldsymbol{X}};\sqrt{\mathsf{snr}}\tilde{\boldsymbol{X}}+\tilde{\boldsymbol{N}}\right) && (28)\\
&= h\left(\sqrt{\mathsf{snr}}\alpha\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{N}}\right) - h\left(\sqrt{\mathsf{snr}}\tilde{\boldsymbol{X}}+\tilde{\boldsymbol{N}}\right) && (29)\\
&= \frac{1}{2}\int_0^{\mathsf{snr}} \left(\mathsf{mmse}(\alpha\tilde{\boldsymbol{Z}},\gamma) - \mathsf{mmse}(\tilde{\boldsymbol{X}},\gamma)\right)\mathrm{d}\gamma && (30)\\
&= \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{Tr}(\mathbf{E}_{\alpha\tilde{\boldsymbol{Z}}}(\gamma) - \mathbf{E}_{\tilde{\boldsymbol{X}}}(\gamma))\,\mathrm{d}\gamma && (31)\\
&= \frac{1}{2}\int_0^{\mathsf{snr}} q_{\mathbf{I}_n}(\tilde{\boldsymbol{X}},\alpha^2,\gamma)\,\mathrm{d}\gamma && (32)
\end{aligned}$$

where we have used the mmse function defined in (7) and the integral expression for the entropy function in [8].

Now, from (29) together with (26), it follows that

$$\lim_{\mathsf{snr}\to\infty} \Delta I(\mathsf{snr}) = 0 \quad (33)$$

which, from the integral expression in (32), further implies that the (smooth) integrand must have, at least, one zero crossing. However, from Theorem 1, we know that $q_{\mathbf{I}_n}(\tilde{\boldsymbol{X}},\alpha^2,\gamma)$ can have, at most, one zero crossing. Consequently, in this case, $q_{\mathbf{I}_n}(\tilde{\boldsymbol{X}},\alpha^2,\gamma)$ must have exactly one zero crossing. Also, from Theorem 1 and (33), we can infer that there exists some $\mathsf{snr}_0 \in (0,\infty)$ such that $q_{\mathbf{I}_n}(\tilde{\boldsymbol{X}},\alpha^2,\gamma) < 0\ \forall\gamma \in [0,\mathsf{snr}_0)$, and $q_{\mathbf{I}_n}(\tilde{\boldsymbol{X}},\alpha^2,\gamma) \geq 0\ \forall\gamma \in [\mathsf{snr}_0,\infty)$. Thus, it immediately follows that for finite snr, $\Delta I(\mathsf{snr}) \leq 0$ and

$$\begin{aligned}
\Delta I(\mathsf{snr}) &= I\left(\alpha\boldsymbol{Z};\sqrt{\mathsf{snr}}\alpha\boldsymbol{Z}+\boldsymbol{N}\right) - I\left(\boldsymbol{X};\sqrt{\mathsf{snr}}\boldsymbol{X}+\boldsymbol{N}\right) && (34)\\
&= h\left(\sqrt{\mathsf{snr}}\alpha\boldsymbol{Z}+\boldsymbol{N}\right) - h\left(\sqrt{\mathsf{snr}}\boldsymbol{X}+\boldsymbol{N}\right) \leq 0. && (35)
\end{aligned}$$

It is now straightforward to see that

$$\begin{aligned}
&\exp\left(\frac{2}{n}h\left(\sqrt{\mathsf{snr}}\boldsymbol{X}+\boldsymbol{N}\right)\right)\\
&\geq \exp\left(\frac{2}{n}h\left(\sqrt{\mathsf{snr}}\alpha\boldsymbol{Z}+\boldsymbol{N}\right)\right) && (36)\\
&= \exp\left(\frac{1}{n}\log\left((2\pi\mathrm{e})^n(\mathsf{snr}\alpha^2+1)^n|\mathbf{R}_{\boldsymbol{N}}|\right)\right) && (37)\\
&= (2\pi\mathrm{e})(\mathsf{snr}\alpha^2+1)|\mathbf{R}_{\boldsymbol{N}}|^{\frac{1}{n}} && (38)\\
&= \exp\left(\frac{2}{n}h\left(\sqrt{\mathsf{snr}}\alpha\boldsymbol{Z}\right)\right) + (2\pi\mathrm{e})|\mathbf{R}_{\boldsymbol{N}}|^{\frac{1}{n}} && (39)\\
&= \exp\left(\frac{2}{n}h\left(\sqrt{\mathsf{snr}}\boldsymbol{X}\right)\right) + (2\pi\mathrm{e})|\mathbf{R}_{\boldsymbol{N}}|^{\frac{1}{n}} && (40)
\end{aligned}$$

which is exactly (25) up to scaling in $\sqrt{\mathsf{snr}}$, which we can always take equal to 1. We note here that the I-MMSE relationship was used in [23] to prove Shannon's EPI, Costa's EPI, and also the generalized EPI for linear transformations of a random vector.

III. From Scalar to Vector Channels: Definitions and Preliminaries

In the previous section, we discussed the simple model presented in (13). We have shown that the "single crossing point" property initially proved for the scalar channel in [7] extends very smoothly and intuitively to this model. The reason for the smooth transition is that even though we are considering a multivariate scenario, all elements of the input vector undergo the same effect in the channel. They are all amplified by snr and distorted by additive standard Gaussian noise. From a more technical viewpoint, when one wants to search for a "single crossing point" property, one must define some scalar function of some scalar parameter, for which the property holds. In the model of (13), the intuitive choice is simply to take the trace of the MMSE as a function of snr. And indeed, this is just one possible linear combination included in Theorem 1, for which we have shown that the property can be extended.

Taking the next step, from this initial extension to the general model of (5), is a harder task. Moreover, there is no single method of doing so. In fact, there are two degrees of freedom in this transition. First of all, there is a need for some scalar parameter that will define $\mathbf{H}$. This parameter will be equivalent to the snr parameter in the scalar case or the simple model of (13). Second, there is a need for some scalar function of the matrix $\mathbf{Q}$. In the simple model of (13), we defined the function $q_{\mathbf{A}}(\boldsymbol{X},\sigma^2,\gamma)$ which was simply taking some linear (positive-semidefinite) combination of the elements of the matrix. The trace function is one example of such a combination, which is also the most intuitive extension; however, in the general model (or even the parallel model, which we will discuss later) the "single crossing point" property does not hold, in general, for the trace function. Thus, our goal is to find a "single crossing point" property that will be both elegant and, more importantly, useful and applicable.

As such, in this study we narrowed our investigation to the subset of parallel channels or diagonal matrices $\mathbf{H}$, for which we have the following result.

Lemma 4: For any two diagonal channel matrices, $\mathbf{H}_1$ and $\mathbf{H}_2$, such that $\mathbf{0} \preceq \mathbf{H}_1 \preceq \mathbf{H}_2$, there exists a path $\mathbf{H}(t)$ such that the following holds.

1 For all t, $\mathbf{H}(t) \succeq \mathbf{0}$ and is a diagonal matrix.
2 For all t, $\mathsf{D}_t\mathbf{H}(t) \succeq \mathbf{0}$ and is a diagonal matrix.
3 $\mathbf{H}(0) = \mathbf{0}$.
4 $\mathbf{H}(t_1) = \mathbf{H}_1$ and $\mathbf{H}(t_2) = \mathbf{H}_2$ where $0 \leq t_1 \leq t_2$.
5 The diagonal elements of $\mathbf{H}(t)$ go to ∞ in a linear rate.

Proof: We need to define a function $g_i(t)$ for each diagonal element of the matrix $\mathbf{H}(t)$. It suffices to choose any nonnegative function $h_i(t)$ such that the area from 0 to t_1 will equal $[\mathbf{H}_1]_{ii}$ and the area from t_1 to t_2 will equal $[\mathbf{H}_2]_{ii} - [\mathbf{H}_1]_{ii}$. Given that, we can set the function to be $g_i(t) = \int_0^t h_i(\tau)\,\mathrm{d}\tau$. The entire path $\mathbf{H}(t)$ will be given by

$$\mathbf{H}(t) = \mathbf{Diag}\left(\{g_i(t)\}\right). \tag{41}$$

As required, this path passes between the zero matrix at $t = 0$, $\mathbf{H}_1$ at t_1, and $\mathbf{H}_2$ at t_2. Since $h_i(t)$ are chosen nonnegative for all i, we have a nonnegative and monotonically nondecreasing path for all t. The aforementioned construction guarantees that both $\mathbf{H}(t)$ and $\mathsf{D}_t\mathbf{H}(t)$ will be diagonal matrices for all t. Moreover, we may also assume that the functions $h_i(t)$ plateau after complying with all other requirements, that is, from t_2 onward. This assures that $g_i(t)$ goes to ∞ in a linear rate. ■

Note that the aforementioned lemma can be extended to M matrices $\mathbf{H}_j \preceq \mathbf{H}_{j+1}$ for $j = 1, \ldots, M-1$, using a similar construction.

Under the aforementioned detailed limitation, of restricting ourselves to parallel channels, we examine two different cases: phases 2 and 3 of our extension.

Before proceeding to examine these two cases, we require a preliminary result. The basis for the applicability of the "single crossing point" property in the scalar case and in the simple model of (5) is the I-MMSE relationship [8]. This is still the case in the extensions we are considering next; however, we require also an extension of the I-MMSE result, derived by Palomar and Verdú in [19], valid for any general deterministic channel matrix, $\mathbf{H}$

$$\nabla_{\mathbf{H}} I\left(\boldsymbol{X}; \mathbf{H}\boldsymbol{X} + \boldsymbol{N}\right) = \mathbf{H}\mathbf{E}_{\boldsymbol{X}} \tag{42}$$

where $\mathbf{E}_{\boldsymbol{X}}$ is the MMSE matrix defined in (6). This relationship was derived for complex-valued variables; however, it holds verbatim for real-valued variables. Assuming that the channel coefficients can be written as a function of a single parameter t and using line integral of a vector field [24], we can rewrite the aforementioned relationship as an integral over this parameter, which results with the following expression:

$$\begin{aligned} I\left(\boldsymbol{X}; \boldsymbol{Y}(t)\right) &= I\left(\boldsymbol{X}; \mathbf{H}(t)\boldsymbol{X} + \boldsymbol{N}\right) \\ &= \int_{\tau=0}^{t} \mathbf{1}^{\mathsf{T}} \left(\mathbf{H}(\tau)\mathbf{E}_{\boldsymbol{X}}(\tau) \circ \mathsf{D}_\tau \mathbf{H}(\tau)\right) \mathbf{1}\,\mathrm{d}\tau \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\left(\mathbf{H}(\tau)\mathbf{E}_{\boldsymbol{X}}(\tau)\right)^{\mathsf{T}} \mathsf{D}_\tau \mathbf{H}(\tau)\right)\mathrm{d}\tau \quad (43) \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{E}_{\boldsymbol{X}}(\tau)\right)\mathrm{d}\tau \quad (44) \end{aligned}$$

where we have used the following definition:

$$\mathbf{B}(t) \equiv \mathbf{H}(t)\left(\mathsf{D}_t\mathbf{H}(t)\right)^{\mathsf{T}}. \tag{45}$$

This also carries over to the conditioned case as follows:

$$\begin{aligned} I\left(\boldsymbol{X}; \boldsymbol{Y}(t)|\boldsymbol{U}\right) &= I\left(\boldsymbol{X}; \mathbf{H}(t)\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U}\right) \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\left(\mathbf{H}(\tau)\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(\tau)\right)^{\mathsf{T}} \mathsf{D}_\tau \mathbf{H}(\tau)\right)\mathrm{d}\tau \quad (46) \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(\tau)\right)\mathrm{d}\tau. \quad (47) \end{aligned}$$

IV. Vector Channel: Comparing With an Independent Gaussian Distribution

We begin our analysis of the extended model (5), limited to parallel channel matrices, by assuming that the Gaussian covariance matrix, defining the matrix $\mathbf{Q}$, is that of an independent distribution, that is, $\mathbf{R}_{\boldsymbol{G}} = \mathbf{\Lambda}_{\boldsymbol{G}}$, throughout this section. Recall, nonetheless, that $\boldsymbol{X}$ remains completely arbitrary. More precisely, we consider the following matrix:

$$\begin{aligned} \mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t) &= \mathbf{E}_G(\mathbf{\Lambda}_{\boldsymbol{G}}, t) - \mathbf{E}_{\boldsymbol{X}}(t) \quad (48) \\ &= \mathbf{Diag}\left(\left\{\frac{[\mathbf{\Lambda}_{\boldsymbol{G}}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\mathbf{\Lambda}_{\boldsymbol{G}}]_{ii}}\right\}\right) - \mathbf{E}_{\boldsymbol{X}}(t). \quad (49) \end{aligned}$$

Under these assumptions we will see, in Section IV-A, that a "single crossing point" property occurs for each and every diagonal element of the matrix $\mathbf{Q}$. After extending this result to the conditioned case, in Section IV-B, we will use the I-MMSE relationship, in Section IV-C, to show the effect of this property on information-theoretic quantities, and more specifically on the mutual information. Finally, in Section IV-D, we will put these results to use on a variant of the *degraded* BC, in order to show their applicability to information theory problems.

A. Single Crossing Point Property on the Diagonal Elements of Q

As pointed out previously, our main result, in this section, is an extension of the "single crossing point" property. Precisely, we show that the property extends on each and every diagonal element of the matrix $\mathbf{Q}$. This result is given in the next theorem.

Theorem 2: The diagonal entries of the matrix-valued function $t \mapsto \mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)$, defined in (48), have no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $t \in [0, \infty)$. Moreover, let $t_{0,i} \in [0, \infty)$ be the negative-to-nonnegative crossing point for $[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$. Then:

1) $[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, 0)]_{ii} \leq 0$;
2) $[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ is a strictly increasing function in the range $t \in [0, t_{0,i})$;
3) $[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)]_{ii} \geq 0$ for all $t \in [t_{0,i}, \infty)$;
4) assuming $\lim_{t\to\infty}[\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t\to\infty}[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)]_{ii} = 0$.
5) $[\mathbf{Q}(\boldsymbol{X}, \mathbf{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ is a continuous and monotonically increasing function in $[\mathbf{\Lambda}_{\boldsymbol{G}}]_{ii}$.

Proof: Before giving the actual proof, let us first present an intermediate result.

Lemma 5: Let $\boldsymbol{X} \in \mathbb{R}^n$ be a random vector such that $[\mathbf{R}_{\boldsymbol{X}}]_{ii} \leq [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$, where $i \in \{1, \dots, n\}$. Then, for every $t \geq 0$, we have

$$[\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} \leq [\mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} = \frac{[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}} \tag{50}$$

with equality if and only if $[\boldsymbol{X}]_i$ is Gaussian distributed, independent of the other entries of $\boldsymbol{X}$ and such that $[\mathbf{R}_{\boldsymbol{X}}]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$.

Proof: See Appendix A4. ■

Now, according to Lemma 5, for the case where $[\mathbf{R}_{\boldsymbol{X}}]_{ii} < [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$, the function $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ has no zeros and the statement in Theorem 1 is true. In addition, if $[\boldsymbol{X}]_i$ is Gaussian distributed (and independent of the other entries of the vector $\boldsymbol{X}$) with variance equal to $[\mathbf{R}_{\boldsymbol{X}}]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$, then $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} = 0 \ \forall t$, which is also consistent with Theorem 1.

Thus, from this point, we can assume that $[\mathbf{R}_{\boldsymbol{X}}]_{ii} \geq [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$ and that $[\boldsymbol{X}]_i$ is not Gaussian distributed, independent of the other entries of $\boldsymbol{X}$, and with $[\mathbf{R}_{\boldsymbol{X}}]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$. Now, for $t = 0$ we have $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, 0)]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii} - [\mathbf{R}_{\boldsymbol{X}}]_{ii} \leq 0$ as required.

Similarly as was done in the proof of Theorem 1, in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ can occur, we only need to show that the derivative of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ with respect to t is positive for all values of t for which $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} < 0$. Observe that $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} < 0$ implies

$$\frac{[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2 [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}} = [\mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} < [\mathbf{E}_{\boldsymbol{X}}(t)]_{ii}. \tag{51}$$

Now, from (48), it is clear that, in order to compute the derivative of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$, we first need the derivative of $[\mathbf{E}_{\boldsymbol{X}}(t)]_{ii}$:

$$\mathsf{D}_t [\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} = \mathsf{D}_{\mathbf{H}(t)} [\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} \mathsf{D}_t \mathbf{H}(t) \tag{52}$$

$$= \sum_{j=1}^{n} \mathsf{D}_{[\mathbf{H}(t)]_{jj}} [\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} \mathsf{D}_t [\mathbf{H}(t)]_{jj} \tag{53}$$

where, in the last step, we have used the assumption that $\mathbf{H}(t)$ is a diagonal matrix for all t. From [22, eq. (131)], we have

$$\mathsf{D}_{[\mathbf{H}(t)]_{jj}} [\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} = -2 \mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ij} [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y}) \mathbf{H}(t)^{\mathsf{T}}]_{ij} \right\} \tag{54}$$

$$= -2 [\mathbf{H}(t)]_{jj} \mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ij}^2 \right\}. \tag{55}$$

Recalling the definition $[\mathbf{B}(t)]_{ii} = [\mathbf{H}(t)]_{ii} \mathsf{D}_t [\mathbf{H}(t)]_{ii}$ in (45), we are now ready to compute the derivative of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$, which reads as

$$\mathsf{D}_t [\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$$

$$= 2 \sum_{j=1}^{n} [\mathbf{B}(t)]_{jj} \left(\mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ij}^2 \right\} - [\mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ij}^2 \right) \tag{56}$$

$$= 2 [\mathbf{B}(t)]_{ii} \left(\mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ii}^2 \right\} - [\mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}^2 \right) + 2 \sum_{j \neq i} [\mathbf{B}(t)]_{jj} \mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ij}^2 \right\} \tag{57}$$

$$\geq 2 [\mathbf{B}(t)]_{ii} \left(\mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ii}^2 \right\} - [\mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}^2 \right) \tag{58}$$

$$> 2 [\mathbf{B}(t)]_{ii} \Big(\mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ii}^2 \right\} - \left(\mathsf{E}\left\{ [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{ii} \right\} \right)^2 \Big) \tag{59}$$

$$\geq 0 \tag{60}$$

where (56) follows from the fact that for Gaussian input distributions (not necessarily i.i.d.), the conditional MMSE matrix $\boldsymbol{\Phi}_{\boldsymbol{X_G}}(\boldsymbol{y})$ does not depend on the observation $\boldsymbol{y}$, i.e., $\mathbf{E}_G(\mathbf{R}_{\boldsymbol{X_G}}, t) = \boldsymbol{\Phi}_{\boldsymbol{X_G}}$. Equation (57) is due to the fact that the entries of the Gaussian input distribution $\boldsymbol{X_G}$ are independent, and thus, its MMSE matrix is diagonal; (58) is due to the fact that $[\mathbf{B}(t)]_{ii} \geq 0$, as shown in Lemma 4; (59) follows from the assumption that $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} < 0$ and (60) can be derived from Jensen's inequality.

Observe that the inequality in (60), which holds for values of t such that $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} < 0$, also proves the second item in Theorem 2 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Regarding the fourth item, it is clear that $\lim_{t \to \infty} [\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} = 0$, as both terms in the expression of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$ in (49) tend to zero, when $\lim_{t \to \infty} [\mathbf{H}(t)]_{ii} = \infty$. Finally, the last property is a direct consequence of the definition of the function $\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ (49). ■

We now define the following function:

$$\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) = [\mathbf{B}(t)]_{ii} [\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii} \tag{61}$$

and also

$$\mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) = \sum_{i=1}^{n} \mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) \tag{62}$$

for which we can give the following two corollaries.

Corollary 1: Let $\boldsymbol{X} \in \mathbb{R}^n$ be any random vector. The function $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ has the following properties.

1) $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, 0) = 0$.
2) It has at most a single negative–zero–positive crossing in the range $t \in (0, \infty)$.
3) When $\lim_{t \to \infty} [\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t \to \infty} \mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) = 0$.
4) If $[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii} = [\mathbf{R}_{\boldsymbol{X}}]_{ii}$, then $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) \geq 0$ for all t. Furthermore, $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ is a continuous and monotonically increasing function in $[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$.

Proof: The first three properties follow from Theorem 2 and the fact that $[\mathbf{B}(t)]_{ii}$ is zero at $t = 0$, is nonnegative for all other values of $t \in (0, \infty)$, and goes to ∞ in a linear rate, as shown in Lemma 4. The fourth property is a direct result of Lemma 5 and the fifth item of Theorem 2. ■

Fig. 1 illustrates this property, in which the negative–zero–positive crossing of $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ is simply a negative-to-nonnegative zero crossing and, thus, agrees with the negative-to-nonegative zero crossing of $[\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)]_{ii}$.

Corollary 2: Let $\boldsymbol{X} \in \mathbb{R}^n$ be any random vector. The function $\mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ is either negative for all t, or there exists $t' \in [0, \infty)$ such that for all $t > t'$ the function $\mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t)$ is nonnegative. Moreover, when $\lim_{t \to \infty} [\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t \to \infty} \mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) = 0$, and if $[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii} = [\mathbf{R}_{\boldsymbol{X}}]_{ii}$ for all i, then $\mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda}_{\boldsymbol{G}}, t) \geq 0$ for all t.

B. Conditioned Case

Before proceeding to understanding the implications of the aforementioned results on information-theoretic quantities, we would like to extend these results to the conditioned case.

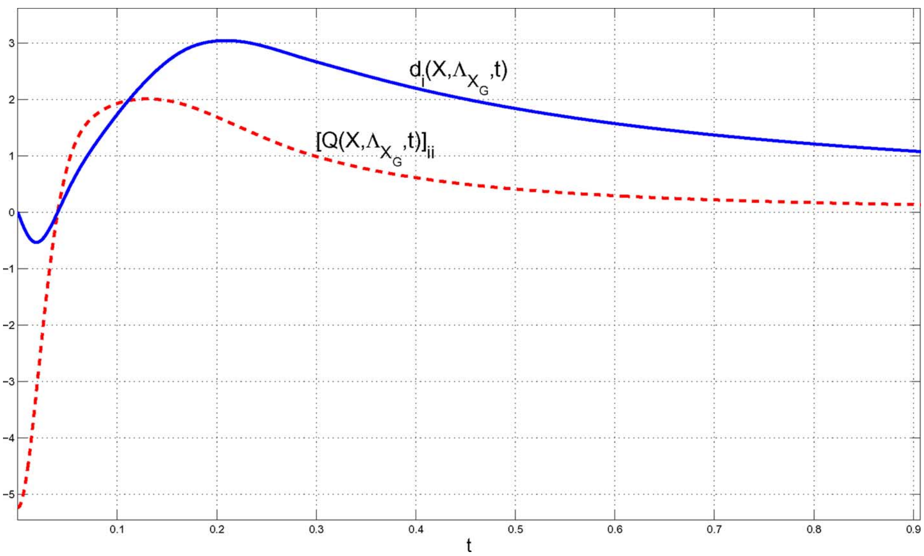

Fig. 1. Example of the function $[\mathbf{Q}(\boldsymbol{X},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ (dashed) and the matching function $\mathsf{d}_i(\boldsymbol{X},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)$ (solid). Both have the same single negative-to-nonnegative zero crossing in the range $t \in (0,\infty)$.

Let us begin with the conditioned MMSE matrix. We first consider the following matrix quantity:

$$\begin{aligned}
&\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{u})\\
&\quad= \mathsf{E}\{(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\mathbf{H}(t)\boldsymbol{X}+\boldsymbol{N},\boldsymbol{U}=\boldsymbol{u}\})\\
&\qquad(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\mathbf{H}(t)\boldsymbol{X}+\boldsymbol{N},\boldsymbol{U}=\boldsymbol{u}\})^{\mathsf{T}}\,\big|\,\boldsymbol{U}=\boldsymbol{u}\} \quad (63)\\
&\quad= \mathsf{E}\{(\boldsymbol{X}_{\boldsymbol{u}}-\mathsf{E}\{\boldsymbol{X}_{\boldsymbol{u}}\,|\,\mathbf{H}(t)\boldsymbol{X}_{\boldsymbol{u}}+\boldsymbol{N}\})\\
&\qquad(\boldsymbol{X}_{\boldsymbol{u}}-\mathsf{E}\{\boldsymbol{X}_{\boldsymbol{u}}\,|\,\mathbf{H}(t)\boldsymbol{X}_{\boldsymbol{u}}+\boldsymbol{N}\})^{\mathsf{T}}\} \quad (64)
\end{aligned}$$

where $\boldsymbol{X}_{\boldsymbol{u}}$ is a random vector distributed according to $P_{\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}}$. The conditioned MMSE matrix is simply the expectation of (63) according to the distribution of the random vector $\boldsymbol{U}$

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t) &= \mathsf{E}\{\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{U})\}\\
&= \mathsf{E}\{(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\mathbf{H}(t)\boldsymbol{X}+\boldsymbol{N},\boldsymbol{U}\})\\
&\qquad(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\mathbf{H}(t)\boldsymbol{X}+\boldsymbol{N},\boldsymbol{U}\})^{\mathsf{T}}\}. \quad (65)
\end{aligned}$$

Another important quantity that needs to be extended to the conditioned case is

$$\begin{aligned}
\boldsymbol{\Phi}_{\boldsymbol{X}_{\boldsymbol{u}}}(\boldsymbol{y}) &= \mathsf{E}\{(\boldsymbol{X}_{\boldsymbol{u}}-\mathsf{E}\{\boldsymbol{X}_{\boldsymbol{u}}\,|\,\boldsymbol{y}\})(\boldsymbol{X}_{\boldsymbol{u}}-\mathsf{E}\{\boldsymbol{X}_{\boldsymbol{u}}\,|\,\boldsymbol{y}\})^{\mathsf{T}}|\boldsymbol{y}\}\\
&= \mathsf{E}\{(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\boldsymbol{y},\boldsymbol{U}=\boldsymbol{u}\})\\
&\qquad(\boldsymbol{X}-\mathsf{E}\{\boldsymbol{X}\,|\,\boldsymbol{y},\boldsymbol{U}=\boldsymbol{u}\})^{\mathsf{T}}|\boldsymbol{y},\boldsymbol{U}=\boldsymbol{u}\} \quad (66)\\
&= \boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{U}=\boldsymbol{u}) \quad (67)
\end{aligned}$$

where, as in the unconditioned case, this function, in general, depends on both $\boldsymbol{u}$ and $\boldsymbol{y}$; thus, we have $\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{u}) = \mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{u}=\boldsymbol{u})\}$, where the expectation is over $\boldsymbol{Y}$. However, when the input distribution of $\boldsymbol{X}_{\boldsymbol{u}}$ is Gaussian, $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y},\boldsymbol{u}=\boldsymbol{u})$ is independent of $\boldsymbol{y}$. In a similar manner, we have the following:

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u},\boldsymbol{\Lambda}_{\boldsymbol{G}},t) = \mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}},t) - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{u}) \quad (68)$$

and, thus, we also have

$$\begin{aligned}
\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t) &= \mathsf{E}_{\boldsymbol{U}}\{\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)\}\\
&= \mathbf{E}_G(\boldsymbol{\Lambda}_{\boldsymbol{G}},t) - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t). \quad (69)
\end{aligned}$$

Using these definitions, we can now extend the results of Theorem 2 to the conditioned case in the following theorem.

Theorem 3: Let $\boldsymbol{U}-\boldsymbol{X}-\boldsymbol{Y}$ form a Markov chain. Then, the diagonal entries of the matrix-valued function $t \mapsto \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)$, defined in (69), have no nonnegative-to-negative zero crossings and, at most, a single negative-to-nonnegative zero crossing in the range $t \in [0,\infty)$. Moreover, let $t_{0,i} \in [0,\infty)$ be the negative-to-nonnegative crossing point for $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$. Then:

1) $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},0)]_{ii} \le 0$;
2) $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ is a strictly increasing function in the range $t \in [0,t_{0,i})$;
3) $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii} \ge 0$ for all $t \in [t_{0,i},\infty)$;
4) when $\lim_{t\to\infty}[\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t\to\infty}[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii} = 0$;
5) $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ is a continuous and monotonically increasing function in $[\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$.

Proof: If $[\boldsymbol{X}]_i$ is Gaussian distributed (independent of $\boldsymbol{U}$ and independent of the other entries of the vector $\boldsymbol{X}$) with variance equal to $[\mathbf{R}_{\boldsymbol{X}}]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$, then $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii} = 0\ \forall t$, which is also consistent with Theorem 3. Thus, from this point, we can assume that $[\boldsymbol{X}]_i$ is not "Gaussian distributed, independent of $\boldsymbol{U}$ and independent of the other entries of $\boldsymbol{X}$, and such that $[\mathbf{R}_{\boldsymbol{X}}]_{ii} = [\boldsymbol{\Lambda}_{\boldsymbol{G}}]_{ii}$."

In this conditioned case, it is harder to determine, up front, all cases in which the function $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ has no zeros. Thus, contrary to the approach used in the proof of Theorem 2, we first prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ can occur. The first property is a direct consequence of this, and there is no need to determine the exact conditions under which the function has no zeros. This approach could have also been used in proving Theorem 2; however, in the unconditioned case we can easily determine the set of cases in which $[\mathbf{Q}(\boldsymbol{X},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ has no zeros.

Similarly to the proof of Theorem 2, in order to prove that no nonnegative-to-negative and at most one negative-to-nonnegative zero crossings of $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda}_{\boldsymbol{G}},t)]_{ii}$ can occur, we only need to show that the derivative of $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii}$ with respect to t is positive for all values of t for which $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} < 0$. According to (56) and (58), we have the following lower bound:

$$\begin{aligned} &\mathsf{D}_t[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{\Lambda_G}, t)]_{ii} \\ &= 2\sum_{j=1}^{n}[\mathbf{B}(t)]_{jj}\left(\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X_u}}(\boldsymbol{Y})]_{ij}^2\right\} - [\mathbf{E}_G(\boldsymbol{\Lambda_G}, t)]_{ij}^2\right) \quad (70)\\ &\geq 2[\mathbf{B}(t)]_{ii}\left(\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X_u}}(\boldsymbol{Y})]_{ii}^2\right\} - [\mathbf{E}_G(\boldsymbol{\Lambda_G}, t)]_{ii}^2\right). \quad (71)\end{aligned}$$

Now we can take expectation over $\boldsymbol{U}$ on both sides and attain the following:

$$\begin{aligned} &\mathsf{D}_t[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} \\ &= \mathsf{E}_{\boldsymbol{U}}\left\{\mathsf{D}_t[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii}\right\} \quad (72)\\ &\geq \mathsf{E}_{\boldsymbol{U}}\left\{2[\mathbf{B}(t)]_{ii}\left(\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X_u}}(\boldsymbol{Y})]_{ii}^2\right\} - [\mathbf{E}_G(\boldsymbol{\Lambda_G}, t)]_{ii}^2\right)\right\} \quad (73)\\ &= 2[\mathbf{B}(t)]_{ii}\left(\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_{ii}^2\right\} - [\mathbf{E}_G(\boldsymbol{\Lambda_G}, t)]_{ii}^2\right) \quad (74)\\ &> 2[\mathbf{B}(t)]_{ii}\left(\mathsf{E}\left\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_{ii}^2\right\} - \left(\mathsf{E}\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_{ii}\}\right)^2\right) \quad (75)\\ &\geq 0 \quad (76)\end{aligned}$$

where (73) is due to (71), (75) follows from the assumption $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} < 0$, and (76) can be derived from Jensen's inequality.

Observe that the inequality in (75), which holds for values of t such that $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} < 0$, also proves the second item in Theorem 3 and the third one follows directly from the inexistence of nonnegative-to-negative zero crossings. Regarding the fourth item, it is clear that $\lim_{t\to\infty}[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} = 0$, as both terms in the expression of $[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii}$ in (69) tend to zero, when $\lim_{t\to\infty}[\mathbf{H}(t)]_{ii} = \infty$. Finally, the last property is a direct consequence of the definition of the function $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)$ in (69). ■

We now extend the definition of the function $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda_G}, t)$ (61) and the function $\mathsf{d}(\boldsymbol{X}, \boldsymbol{\Lambda_G}, t)$ (62) to the conditioned case

$$\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{\Lambda_G}, t) = [\mathbf{B}(t)]_{ii}[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{\Lambda_G}, t)]_{ii} \quad (77)$$

$$\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) = [\mathbf{B}(t)]_{ii}[\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)]_{ii} \quad (78)$$

and also

$$\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{\Lambda_G}, t) = \sum_{i=1}^{n}\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}, \boldsymbol{\Lambda_G}, t) \quad (79)$$

$$\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) = \sum_{i=1}^{n}\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) \quad (80)$$

for which we can extend corollaries 1 and 2 as follows.

Corollary 3: Let $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}$ form a Markov chain such that the random vector $\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u} \in \mathbb{R}^n$ has covariance matrix $\mathbf{R}_{\boldsymbol{X}|\boldsymbol{U}=\boldsymbol{u}}$. The function $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)$ has the following properties:

1) $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, 0) = 0$.
2) It has at most a single negative–zero–positive crossing in the range $t \in (0, \infty)$.
3) When $\lim_{t\to\infty}[\mathbf{H}(t)]_{ii} = \infty$, we have that $\lim_{t\to\infty}\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) = 0$.
4) If $[\boldsymbol{\Lambda_G}]_{ii} = [\mathbf{R}_{\boldsymbol{X}}]_{ii}$, then $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) \geq 0$ for all t. Furthermore, $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)$ is a continuous and monotonically increasing function in $[\boldsymbol{\Lambda_G}]_{ii}$.

Proof: The first three properties follow directly from Theorem 3 and the fact that $[\mathbf{B}(t)]_{ii}$ is zero at $t = 0$ and nonnegative for all other values of $t \in (0, \infty)$ and $[\mathbf{B}(t)]_{ii}$ goes to ∞ in a linear rate, as shown in Lemma 4. The fourth property is a direct result of Lemma 5, and the fifth property in Theorem 3. ■

Corollary 4: Let $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}$ form a Markov chain. The function $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)$ is either negative for all t, or there exists $t' \in [0, \infty)$ such that for all $t > t'$ the function $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t)$ is nonnegative. Moreover, when $\lim_{t\to\infty}[\mathbf{H}(t)]_{ii} = \infty$ we have that $\lim_{t\to\infty}\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) = 0$, and if $[\boldsymbol{\Lambda_G}]_{ii} = [\mathbf{R}_{\boldsymbol{X}}]_{ii}$ for all i, then $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) \geq 0$ for all t.

C. Properties of the Mutual Information

So far, we have seen properties of the matrix $\mathbf{Q}(\boldsymbol{X}, \boldsymbol{\Lambda_G}, t)$ or, more precisely, of its diagonal elements. We have seen that these properties extend naturally to the conditioned case, and also to the function $\mathsf{d}_i(\boldsymbol{X}, \boldsymbol{\Lambda_G}, t)$ and its conditioned version. In this section, our goal is to use these results to derive new properties on the mutual information between the input and the output of parallel Gaussian channels. In order to derive these results, we put to use the I-MMSE relationship, as given in (43)–(44) and (46)–(47).

For the sake of compactness, we will write the properties in this section only for the more general, conditioned case, from which one can easily derive the respective unconditioned theorems.

Theorem 4: Let $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}$ form a Markov chain. Assume an independent Gaussian input $\boldsymbol{X_G}$, with covariance $\boldsymbol{\Lambda_G}$, such that for all i

$$I([\boldsymbol{X}]_i; [\boldsymbol{Y}(t_e)]_i|\boldsymbol{U}) = I([\boldsymbol{X_G}]_i; [\boldsymbol{Y_G}(t_e)]_i) \quad (81)$$

where

$$\boldsymbol{Y}(t_e) = \mathbf{H}(t_e)\boldsymbol{X} + \boldsymbol{N} \quad \text{and} \quad (82)$$

$$\boldsymbol{Y_G}(t_e) = \mathbf{H}(t_e)\boldsymbol{X_G} + \boldsymbol{N}. \quad (83)$$

Then, $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) \geq 0$ for all $t \geq t_e$.

Proof: Let us define $\boldsymbol{X}^{\text{ind cond.}} \in \mathbb{R}^n$ as a random vector with independent elements when conditioned on $\boldsymbol{U}$, and with distribution of each pair $\left([\boldsymbol{X}^{\text{ind cond.}}]_i, \boldsymbol{U}\right)$ being the same as the marginal distribution of the corresponding pair $([\boldsymbol{X}]_i, \boldsymbol{U})$. Thus, $\left[\mathbf{E}_{\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U}}\right]_{ii}$ is basically the MMSE of $[\boldsymbol{X}^{\text{ind cond.}}]_i$ from $\boldsymbol{U}$ and $[\boldsymbol{Y}(t)]_i$, which is

$$[\boldsymbol{Y}(t)]_i = \left[\mathbf{H}(t)\boldsymbol{X}^{\text{ind cond.}} + \boldsymbol{N}\right]_i = [\mathbf{H}(t)]_{ii}[\boldsymbol{X}^{\text{ind cond.}}]_i + [\boldsymbol{N}]_i$$

where the equality holds due to the fact that the channel matrix $\mathbf{H}(t)$ is diagonal for all t and $\boldsymbol{N}$ is standard Gaussian. Using these definitions, we can give the following special case of (47):

$$\begin{aligned} &I([\boldsymbol{X}]_i; [\boldsymbol{Y}(t)]_i|\boldsymbol{U}) \\ &= I\left([\boldsymbol{X}]_i; [\mathbf{H}(t)]_{ii}[\boldsymbol{X}^{\text{ind cond.}}]_i + [\boldsymbol{N}]_i|\boldsymbol{U}\right) \quad (84)\\ &= \int_{\tau=0}^{t}[\mathbf{H}(\tau)]_{ii}\left[\mathbf{E}_{\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U}}(\tau)\right]_{ii}[\mathsf{D}_\tau\mathbf{H}(\tau)]_{ii}\,\mathrm{d}\tau \quad (85)\\ &= \int_{\tau=0}^{t}[\mathbf{B}(\tau)]_{ii}\left[\mathbf{E}_{\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U}}(\tau)\right]_{ii}\mathrm{d}\tau. \quad (86)\end{aligned}$$

Putting this together with the assumption, we have

$$0 = I([\boldsymbol{X_G}]_i;[\boldsymbol{Y_G}(t_e)]_i) - I([\boldsymbol{X}]_i;[\boldsymbol{Y}(t_e)]_i|\boldsymbol{U}) = \int_{\tau=0}^{t_e} \mathsf{d}_i(\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U},\boldsymbol{\Lambda_G},\tau)\,\mathrm{d}\tau. \quad (87)$$

Now, due to Corollary 4 we can conclude that there exists $t_0 \in [0, t_e]$ such that $\mathsf{d}_i(\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0$ for all $t > t_0$ and as a result, $\mathsf{d}_i(\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0$ for all $t > t_e$. Now, for all t we have that $[\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t)]_{ii} \leq [\mathbf{E}_{\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U}}(t)]_{ii}$. Thus, if the negative–zero–positive crossing of $\mathsf{d}_i(\boldsymbol{X}^{\text{ind cond.}}|\boldsymbol{U},\boldsymbol{\Lambda_G},t)$ is at t_0, the negative–zero–positive crossing of $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t)$ is at $t_0' \leq t_0$. From this, we can conclude also that $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0$ for all $t > t_e$. Finally, since this holds for every i, it also holds for the summation over i, i.e., for the function $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t)$, concluding the proof. ■

We are now ready to give the main theorem of this section.

Theorem 5: Let $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}$ form a Markov chain. For any $t_e \in [0,\infty)$, there exists an independent Gaussian input $\boldsymbol{X_G}$ with covariance $\boldsymbol{\Lambda_G}$ such that the following properties hold.

1) $\mathsf{d}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0$ for all $t \geq t_e$.
2) $I(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}) = I(\boldsymbol{X_G};\boldsymbol{Y_G}(t_e))$, where $\boldsymbol{Y}(t_e)$ and $\boldsymbol{Y_G}(t_e)$ are as defined in (82) and (83), respectively.
3) $[\boldsymbol{\Lambda_G}]_{ii} \leq [\mathbf{R}_{\boldsymbol{X}}]_{ii}$ for all i.

Proof: We provide a constructive proof, and show how one can build an independent Gaussian input distribution complying with all three requirements. We begin by examining the meaning of the second requirement. First, recall the I-MMSE relationship in the parallel setting, given in (47)

$$I(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}) = \int_{\tau=0}^{t_e} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(\tau)\right)\mathrm{d}\tau. \quad (88)$$

Now, the second requirement is equivalent to the following equality:

$$\begin{aligned} 0 &= I(\boldsymbol{X_G};\boldsymbol{Y_G}(t_e)) - I(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}) \\ &= \int_{\tau=0}^{t_e} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},\tau)\right)\mathrm{d}\tau \\ &= \int_{\tau=0}^{t_e} \sum_{i=1}^{n} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},\tau)\,\mathrm{d}\tau \\ &= \sum_{i=1}^{n} \int_{\tau=0}^{t_e} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},\tau)\,\mathrm{d}\tau. \end{aligned} \quad (89)$$

Thus, we wish to show the existence of an independent Gaussian input distribution which complies with requirements 1, 3, and (89). There are different ways to attain equality in (89); however, since we need only to show the existence of a specific independent Gaussian distribution, we follow one possible approach, which is to require the following:

$$\int_{\tau=0}^{t_e} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},\tau)\,\mathrm{d}\tau = 0 \quad \forall i. \quad (90)$$

Now, according to the fourth property in Corollary 3 we know that

$$\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0 \quad (91)$$

for all t when $[\boldsymbol{\Lambda_G}]_{ii} = [\mathbf{R}_{\boldsymbol{X}}]_{ii}$, and that it is continuous and monotonically increasing in the value of $[\boldsymbol{\Lambda_G}]_{ii}$ (and trivially negative, for all t, when $[\boldsymbol{\Lambda_G}]_{ii} = 0$). Thus, there exists a number $\eta_i \in [0,1]$ such that setting $[\boldsymbol{\Lambda_G}]_{ii} = \eta_i[\mathbf{R}_{\boldsymbol{X}}]_{ii}$ results with the equality in (90). Due to the second property in Corollary 3, we know that either $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) = 0$ for all t or that there exists a single negative–zero–positive crossing in the range $[0,t_e]$. In both cases, the setting $[\boldsymbol{\Lambda_G}]_{ii} = \eta_i[\mathbf{R}_{\boldsymbol{X}}]_{ii}$ results with $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U},\boldsymbol{\Lambda_G},t) \geq 0$ for all $t > t_e$. Since there exists such η_i for every i, we comply also with requirements 1 and 3, and conclude the proof. ■

Remark 4: Note that the aforementioned choice of $\boldsymbol{\Lambda_G}$ does not necessarily imply $\boldsymbol{\Lambda_G} \preceq \mathbf{R}_{\mathbf{X}}$. However, we can conclude that $\boldsymbol{\Lambda_G} \not\succ \mathbf{R}_{\mathbf{X}}$.

The following is a simple corollary of the aforementioned theorem.

Corollary 5: Given any arbitrary *independent* input distribution over $\boldsymbol{X} \in \mathbb{R}^n$, with covariance $\boldsymbol{\Lambda_X}$, and any t_e, there exists an independent Gaussian input $\boldsymbol{X_G}$ with covariance $\boldsymbol{\Lambda_G}$ such that

$$I(\boldsymbol{X};\mathbf{H}(t_e)\boldsymbol{X}+\boldsymbol{N}) = I(\boldsymbol{X_G};\mathbf{H}(t_e)\boldsymbol{X_G}+\boldsymbol{N}) \quad (92)$$
$$\boldsymbol{\Lambda_G} \preceq \boldsymbol{\Lambda_X} \quad (93)$$
$$\text{and} \quad \mathbf{E}_G(\boldsymbol{\Lambda_G},t_e) \preceq \mathbf{E}_{\boldsymbol{X}}(t_e). \quad (94)$$

D. Application: The Degraded Parallel Gaussian BC Capacity Region Under Per-Antenna Power Constraint

We now show that Theorem 5 can be used in providing a converse proof for the *degraded* parallel Gaussian BC capacity region under an input per-antenna power constraint. We consider the following model:

$$\begin{aligned} \boldsymbol{Y}_1[m] &= \mathbf{H}_1\boldsymbol{X}[m] + \boldsymbol{N}_1[m] \\ \boldsymbol{Y}_2[m] &= \mathbf{H}_2\boldsymbol{X}[m] + \boldsymbol{N}_2[m] \end{aligned} \quad (95)$$

where $\boldsymbol{N}_1[m]$ and $\boldsymbol{N}_2[m]$ are standard additive Gaussian noise vectors independent of different time indices m, and $\mathbf{H}_1$ and $\mathbf{H}_2$ are diagonal positive-semidefinite matrices such that $\mathbf{H}_1 \preceq \mathbf{H}_2$. $\boldsymbol{X} \in \mathbb{R}^n$ is the random input vector, and it is assumed independent of different time indices m. Note that m is the time index and should not be confused with the scalar parameter t which is used as a "MIMO snr parameter," i.e., the parameter t determines the channel matrix $\mathbf{H}(t)$.

We consider an input per-antenna power constraint

$$\left[\mathsf{E}\left\{\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}}\right\}\right]_{ii} \leq P_i \quad \forall i, 1 \leq i \leq n. \quad (96)$$

Since we have a *degraded* BC, we can use the single-letter expression given in [25]

$$\begin{aligned} \mathsf{R}_1 &\leq I(\boldsymbol{U};\boldsymbol{Y}_1) \\ \mathsf{R}_2 &\leq I(\boldsymbol{X};\boldsymbol{Y}_2|\boldsymbol{U}) \end{aligned} \quad (97)$$

where $\boldsymbol{U}$ is an auxiliary random vector over a certain alphabet that satisfies the Markov relation $\boldsymbol{U} - \boldsymbol{X} - (\boldsymbol{Y}_1,\boldsymbol{Y}_2)$. The following proof was originally given for the scalar Gaussian BC in [7] and [20] and we now extend it to the *degraded* parallel

Gaussian channel. Using Lemma 4, we can construct a path such that

$$\begin{aligned}\mathbf{H}(t_2) &= \mathbf{H}_2\\ \mathbf{H}(t_1) &= \mathbf{H}_1\\ \mathbf{H}(0) &= \mathbf{0}\end{aligned} \tag{98}$$

where $0 \leq t_1 \leq t_2$ and $\mathbf{H}(t)$ is diagonal for all $t \in [0, t_2]$.

Now, assume a pair $(\boldsymbol{U}, \boldsymbol{X})$ such that $\boldsymbol{X}$ has covariance $\mathbf{R}_{\boldsymbol{X}}$. According to Theorem 5, there exists an independent Gaussian vector $\boldsymbol{X_G}$ with covariance matrix $\boldsymbol{\Lambda_G}$ such that the following properties hold:

$$\begin{aligned}I(\boldsymbol{X}; \boldsymbol{Y}_1|\boldsymbol{U}) &= I(\boldsymbol{X}; \mathbf{H}(t_1)\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U})\\ &= I(\boldsymbol{X_G}; \boldsymbol{Y_G}(t_1))\\ &= I(\boldsymbol{X_G}; \mathbf{H}(t_1)\boldsymbol{X_G} + \boldsymbol{N})\end{aligned} \tag{99}$$

$$\mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, t) \geq 0 \quad \forall t \geq t_1 \tag{100}$$

$$[\boldsymbol{\Lambda_G}]_{ii} \leq [\mathbf{R}_{\boldsymbol{X}}]_{ii} \quad \forall i. \tag{101}$$

Using the I-MMSE relationship (47), we can write

$$I(\boldsymbol{X_G}; \mathbf{H}(t)\boldsymbol{X_G} + \boldsymbol{N}) - I(\boldsymbol{X}; \mathbf{H}(t)\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U})$$

$$= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\right) \mathrm{d}\tau \tag{102}$$

$$= \int_{\tau=0}^{t} \sum_{i=1}^{n} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\, \mathrm{d}\tau \tag{103}$$

$$= \int_{\tau=0}^{t} \mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\, \mathrm{d}\tau. \tag{104}$$

Using the aforementioned properties on (104), we have that for any $t' > t_1$

$$I(\boldsymbol{X_G}; \mathbf{H}(t')\boldsymbol{X_G} + \boldsymbol{N}) - I(\boldsymbol{X}; \mathbf{H}(t')\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U})$$

$$= \int_{\tau=0}^{t_1} \mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\, \mathrm{d}\tau + \int_{\tau=t_1}^{t'} \mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\, \mathrm{d}\tau$$

$$= 0 + \int_{\tau=t_1}^{t'} \mathsf{d}(\boldsymbol{X}|\boldsymbol{U}, \boldsymbol{\Lambda_G}, \tau)\, \mathrm{d}\tau \geq 0 \tag{105}$$

where the second transition is due to (99) and the inequality is due to (100). Thus, we have shown the existence of an independent Gaussian vector $\boldsymbol{X_G}$ with covariance matrix $\boldsymbol{\Lambda_G}$, with the following properties:

$$I(\boldsymbol{X}; \mathbf{H}_1\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U}) = \frac{1}{2}\mathsf{log}|\mathbf{I} + \mathbf{H}_1\boldsymbol{\Lambda_G}\mathbf{H}_1^{\mathsf{T}}| \tag{106}$$

$$I(\boldsymbol{X}; \mathbf{H}_2\boldsymbol{X} + \boldsymbol{N}|\boldsymbol{U}) \leq \frac{1}{2}\mathsf{log}|\mathbf{I} + \mathbf{H}_2\boldsymbol{\Lambda_G}\mathbf{H}_2^{\mathsf{T}}| \tag{107}$$

$$\text{and} \quad [\boldsymbol{\Lambda_G}]_{ii} \leq [\mathbf{R}_{\boldsymbol{X}}]_{ii} \quad \forall i. \tag{108}$$

Using these properties on the single-letter expression (97), we obtain the following outer bound:

$$\begin{aligned}\mathsf{R}_1 &\leq I(\boldsymbol{U}; \boldsymbol{Y}_1) = I(\boldsymbol{X}; \boldsymbol{Y}_1) - I(\boldsymbol{X}; \boldsymbol{Y}_1|\boldsymbol{U})\\ &\leq \frac{1}{2}\mathsf{log}|\mathbf{I} + \mathbf{H}_1\mathbf{P}\mathbf{H}_1^{\mathsf{T}}| - \frac{1}{2}\mathsf{log}|\mathbf{I} + \mathbf{H}_1\boldsymbol{\Lambda_G}\mathbf{H}_1^{\mathsf{T}}|\\ &= \frac{1}{2}\mathsf{log}\frac{|\mathbf{I} + \mathbf{H}_1\mathbf{P}\mathbf{H}_1^{\mathsf{T}}|}{|\mathbf{I} + \mathbf{H}_1\boldsymbol{\Lambda_G}\mathbf{H}_1^{\mathsf{T}}|}\end{aligned} \tag{109}$$

$$\mathsf{R}_2 \leq I(\boldsymbol{X}; \boldsymbol{Y}_2|\boldsymbol{U}) \leq \frac{1}{2}\mathsf{log}|\mathbf{I} + \mathbf{H}_2\boldsymbol{\Lambda_G}\mathbf{H}_2^{\mathsf{T}}| \tag{110}$$

where $\mathbf{P}$ is a diagonal matrix with $[\mathbf{P}]_{ii} = P_i$ for all i. This outer bound is tight and the achievability is well known using superposition coding. This approach can be extended to the M-user scenario as shown in Appendix B.

V. Vector Channel: Comparing With a General Gaussian Distribution

This section provides phase 3 of our "single crossing point" extension (see Table I), and extends the analysis of the previous section. More precisely, we continue looking into the model given in (5), limited to parallel channel matrices; however, we now allow the Gaussian covariance matrix, defining the matrix $\mathbf{Q}$, to be any proper covariance matrix. In other words, we no longer limit ourselves to independent Gaussian inputs. For this, more general setting, we will see in Section V-A that a "single crossing point" property occurs for each and every eigenvalue of the matrix $\mathbf{Q}$. After extending this result to the conditioned case, in Section V-B, we will use the I-MMSE relationship, in Section V-C, to show the effect of this property on information-theoretic quantities, and more specifically on the mutual information. We will relate these results to the Fisher information in Section V-D. Finally, in Sections V-E and V-F we will put these results to use in the *degraded* BC capacity converse proof, for both the compound and noncompound scenarios.

A. Single Crossing Point for Each Eigenvalue of $\mathbf{Q}(t)$

In this section, we prove the main result of this paper: showing that each eigenvalue of the matrix $\mathbf{Q}$ has at most a single negative-to-nonnegative zero crossing. This is, to our understanding, not an intuitive extension of the "single crossing point" property, which emphasizes the importance of the eigenvalues in the analysis of MIMO scenarios.

For the proof of the main theorem, we require the following lemma, which is also interesting on its own.

Lemma 6: The following lower bound holds:

$$\mathsf{D}_t\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t) \succeq 2\left(\mathbf{E}_{\boldsymbol{X}}(t)\mathbf{B}(t)\mathbf{E}_{\boldsymbol{X}}^{\mathsf{T}}(t) - \mathbf{E}_G(t)\mathbf{B}(t)\mathbf{E}_G^{\mathsf{T}}(t)\right) \tag{111}$$

where $\mathbf{B}(t)$ was defined in (45) and assumed a positive-semidefinite diagonal matrix for all t (see Lemma 4).

Proof: See Appendix A5. ■

We are now ready to proceed to the main result of this paper:

Theorem 6: Each eigenvalue of $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ has, *at most*, a single negative-to-nonnegative zero crossing of the horizontal axis.

Proof: Loosely speaking, the proof is based on proving that once an eigenvalue has become (or is) nonnegative, it cannot become negative. Thus, from the (weak) continuity of the eigenvalues as a function of t, that follows from [26, App. D], the eigenvalues can cross the horizontal axis, at most, once. Also from continuity arguments, it is easy to see that we must limit our study of the eigenvalues of $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{X_G}}, t)$ to the values of t where the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{X_G}}, t)$ is singular (i.e., a subset of its eigenvalues are zero) as it is the only possible situation where a zero crossing can occur. Finally, throughout this proof and for the sake of simplicity we will use the simplified notation $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t) = \mathbf{Q}(t) \stackrel{\triangle}{=} \mathbf{E}_G(t) - \mathbf{E}_{\boldsymbol{X}}(t)$ because the entire proof is given for any constant setting of the input random vector $\boldsymbol{X}$ and the Gaussian covariance $\mathbf{R}_{\boldsymbol{G}}$.

We begin by stating a few supporting results and giving some preliminary definitions.

Lemma 7: Let $\mathbf{A}$ and $\mathbf{B}$ be two n-dimensional positive-semidefinite matrices, i.e., $\mathbf{A} \succeq \mathbf{0}$, $\mathbf{B} \succeq \mathbf{0}$. Then, there exists an invertible matrix $\mathbf{S}$ such that both $\mathbf{SAS}^T$ and $\mathbf{SBS}^T$ are diagonal matrices.

Proof: See Appendix A6. ∎

Let us consider the simultaneous decomposition of $\{\mathbf{E}_G(t), \mathbf{E}_{\boldsymbol{X}}(t)\}$ according to Lemma 7 as

$$\begin{aligned}\mathbf{E}_G(t) &= \mathbf{V}(t)^{\mathsf{T}}\boldsymbol{\Sigma}_G(t)\mathbf{V}(t)\\ \mathbf{E}_{\boldsymbol{X}}(t) &= \mathbf{V}(t)^{\mathsf{T}}\boldsymbol{\Sigma}_{\boldsymbol{X}}(t)\mathbf{V}(t)\end{aligned} \tag{112}$$

where $\mathbf{V}(t)$ is an invertible matrix and $\boldsymbol{\Sigma}_G(t)$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}(t)$ are diagonal matrices. It will be convenient to define $\widetilde{\mathbf{Q}}(t,\tau)$, for $\tau \geq 0$, according to

$$\widetilde{\mathbf{Q}}(t,\tau) = \mathbf{V}(\tau)^{-\mathsf{T}}\mathbf{Q}(t)\mathbf{V}(\tau)^{-1} \tag{113}$$

where $\mathbf{V}(\tau)$ is the same as defined in (112).

The remainder of the proof is split into two parts. In the first part, we will prove that each eigenvalue of $\widetilde{\mathbf{Q}}(t,\tau)$ has at most a single negative-to-nonnegative zero crossing. In the second part, we will show that this property transfers to $\mathbf{Q}(t)$, thus completing the proof. Coincidentally, both parts of the proof will be based on contradiction arguments, i.e., we assume that the opposite of what we want to prove is true and, then, end up with an inconsistency.

1) Single Crossing Point for the Eigenvalues of $\widetilde{\mathbf{Q}}(t,\tau)$: Let us start by presenting a result on the differentiability of the eigenvalues of a symmetric matrix with respect to some scalar parameter t, which was studied by Rellich in [27, Ch. 1].[1]

Lemma 8 [27, Theorem on p. 57]: Suppose that $\mathbf{A}(t)$ is an n-dimensional symmetric matrix defined on some open interval $t \in (t_1, t_2)$. Suppose that the derivative $\mathsf{D}_t\mathbf{A}(t)$ exists and it is continuous for each $t \in (t_1, t_2)$. Then, there exist n functions $\lambda_i(t)$, $i = 1, \ldots, n$, with continuous derivatives in $t \in (t_1, t_2)$, such that

$$\mathbf{A}(t)\mathbf{u}_i(t) = \lambda_i(t)\mathbf{u}_i(t), \quad i = 1, \ldots, n \tag{114}$$

for some properly chosen orthonormal system of vectors $\mathbf{u}_i(t)$, $i = 1, \ldots, n$.

Since $\widetilde{\mathbf{Q}}(t,\tau)$ is a symmetric matrix whose derivative $\mathsf{D}_t\widetilde{\mathbf{Q}}(t,\tau)$ exists, Lemma 8 ensures the existence of n continuous and differentiable functions such that they are equal to the eigenvalues of the matrix $\widetilde{\mathbf{Q}}(t,\tau)$, for any choice of τ. These functions will be denoted from now on by $\lambda_i(t,\tau)$, for $i = 1, \ldots, n$.

Now, let us assume that, at $t = t_0$, k of these eigenvalues (with $k \leq n$) are equal to zero, i.e., $\lambda_i(t_0,\tau) = 0$, for $i = 1, \ldots, k$. Furthermore, we also assume that, from these k eigenvalues that are zero at $t = t_0$, s of them (with $s \leq k$) have a nonnegative-to-negative zero crossing at $t = t_0$. To sum up, we assume that the differentiable functions $\lambda_i(t,\tau)$ with $i = 1, \ldots, s$ have a nonnegative-to-negative zero crossing at $t = t_0$.

Let us now present a property of differentiable functions that contain nonnegative-to-negative zero crossings.

Lemma 9: Assume that $f(t)$ has a nonnegative-to-negative zero crossing at $t = t_0$ and that $f(t)$ has a continuous derivative with respect to t. Then, there exists a positive value ε such that

$$f(t) < 0, \quad t \in (t_0, t_0 + \varepsilon) \tag{115}$$

$$\mathsf{D}_t f(t) < 0, \quad t \in (t_0, t_0 + \varepsilon). \tag{116}$$

Proof: From Definition 2, (115) follows immediately for any $\varepsilon \leq \epsilon$. The proof for (116) follows easily from the mean value theorem and elementary calculus. ∎

Applying Lemma 9 to the set of functions $\lambda_i(t,\tau)$, with $i = 1, \ldots, s$, we readily obtain

$$\left.\begin{aligned}\lambda_i(t,\tau) < 0, &\quad t \in (t_0, t_0 + \varepsilon_i(\tau))\\ \mathsf{D}_t\lambda_i(t,\tau) < 0, &\quad t \in (t_0, t_0 + \varepsilon_i(\tau))\end{aligned}\right\} \quad i = 1, \ldots, s \tag{117}$$

where we have written $\varepsilon_i(\tau)$ to make explicit the dependence of ε_i on the specific value of τ. For the sake of convenience, we want to eliminate the dependence of ε_i on τ. A possible method to eliminate this dependence is to define

$$\varepsilon_i^\star = \inf_{\tau \in [t_0, t_0+M]} \varepsilon_i(\tau) = \min_{\tau \in [t_0, t_0+M]} \varepsilon_i(\tau) > 0 \tag{118}$$

where, for the sake of convenience, we have restricted the values of τ in the interval $[t_0, t_0 + M]$, with M being an arbitrary fixed positive value (observe that since $\varepsilon_i(\tau)$ can be made arbitrarily small, we can always guarantee that $M > \varepsilon_i(\tau) \geq \varepsilon_i^\star$), and where the second equality follows from the fact that the optimization set is a closed interval and the third one follows from $\varepsilon_i(\tau) > 0\ \forall\tau$.

Consequently, after this simplification, we have that, assuming that the differentiable functions $\lambda_i(t,\tau)$, with $i = 1, \ldots, s$, have a nonnegative-to-negative zero crossing at $t = t_0$, they must fulfill

$$\left.\begin{aligned}\lambda_i(t,\tau) < 0, &\quad t \in (t_0, t_0 + \varepsilon_i^\star)\\ \mathsf{D}_t\lambda_i(t,\tau) < 0, &\quad t \in (t_0, t_0 + \varepsilon_i^\star)\end{aligned}\right\} \quad i = 1, \ldots, s. \tag{119}$$

Now, we can particularize the aforementioned expression for the case where $t = t_0 + \varepsilon_i^\star/2 \stackrel{\triangle}{=} t^\star$ and where we also choose $\tau = t^\star$. We obtain

$$\left.\begin{aligned}\lambda_i(t^\star, t^\star) &< 0\\ \mathsf{D}_t\lambda_i(t, t^\star)\big|_{t=t^\star} &< 0\end{aligned}\right\} \quad i = 1, \ldots, s. \tag{120}$$

From this point, our goal is to prove that the two conditions in (120) cannot hold at the same time. For that purpose, we need an expression for the derivative of the eigenvalue function $\mathsf{D}_t\lambda_i(t,t^\star)$. Since we have that $\lambda_i(t_0,\tau) = 0$ for $i = 1, \ldots, k$ (i.e., the multiplicity of the zero eigenvalue is k), we cannot guarantee that the multiplicity of the eigenvalue $\lambda_i(t^\star, t^\star)$ is equal to 1. From this point, we assume that the multiplicity of $\lambda_i(t^\star, t^\star)$ is l.

Consequently, we now require the following result by Lancaster in [28, Th. 7] (it is also reproduced in [21, Ch. 8, Sec. 12,

[1]Rellich studied the eigenvalue differentiability for Hermitian matrices. We specialized his result for the real case studied in this paper.

Th. 13]), which gives us an expression for the derivatives of the multiple eigenvalues.[2]

Lemma 10 [28, Th. 7]: Under the assumptions in Lemma 8, let us consider the case where $\mathbf{A}(t)$ has a repeated eigenvalue λ_0 with multiplicity l, i.e., $\lambda_1(t) = \lambda_2(t) = \ldots = \lambda_l(t) = \lambda_0$. Assume further that the $n \times l$ matrix $\mathbf{U}(t)$ spans the space associated with the repeated eigenvalues (i.e., $\mathbf{U}(t)$ contains one particular set of eigenvectors associated with the l repeated eigenvalue). Then, the l derivatives of the eigenvalues, which coincide at λ_0, are the eigenvalues of the matrix

$$\mathbf{U}(t)^{\mathsf{T}} \mathsf{D}_t \mathbf{A}(t) \mathbf{U}(t). \tag{121}$$

Using Lemma 10 and denoting by $[\mathbf{A}]_{1:l,1:l}$ the upper-left $l \times l$ submatrix of matrix $\mathbf{A}$, we can write

$$\begin{aligned} &\mathsf{D}_t \lambda_i(t, \tau = t^\star)\big|_{t=t^\star} \\ &= \mu_i \left(\mathbf{1}_{n,l}^{\mathsf{T}}\, \mathsf{D}_t \widetilde{\mathbf{Q}}(t, \tau = t^\star)\Big|_{t=t^\star} \mathbf{1}_{n,l} \right) \qquad (122) \\ &= \mu_i \left(\mathbf{1}_{n,l}^{\mathsf{T}} \mathbf{V}(t^\star)^{-\mathsf{T}}\, \mathsf{D}_t \mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)\big|_{t=t^\star} \mathbf{V}(t^\star)^{-1} \mathbf{1}_{n,l} \right) \\ &\geq \mu_i \left(\left[\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star) \mathbf{C}(t^\star) \boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star) - \boldsymbol{\Sigma}_G(t^\star) \mathbf{C}(t^\star) \boldsymbol{\Sigma}_G(t^\star) \right]_{1:l,1:l} \right) \\ &= \mu_i \Big([\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)]_{1:l,1:l} [\mathbf{C}(t^\star)]_{1:l,1:l} [\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)]_{1:l,1:l} - \\ &\qquad [\boldsymbol{\Sigma}_G(t^\star)]_{1:l,1:l} [\mathbf{C}(t^\star)]_{1:l,1:l} [\boldsymbol{\Sigma}_G(t^\star)]_{1:l,1:l} \Big) \qquad (123) \end{aligned}$$

where $\mu_i(\mathbf{A})$ denotes the eigenvalue function of a generic matrix $\mathbf{A}$. Observe that, thanks to the fact that $\widetilde{\mathbf{Q}}(t^\star, t^\star) = \boldsymbol{\Sigma}_G(t^\star) - \boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)$ is a diagonal matrix, in (122) we have chosen

$$\mathbf{U} = \mathbf{1}_{n,l} \triangleq \begin{pmatrix} \mathbf{I}_l \\ \mathbf{0}_{n-l,l} \end{pmatrix} \tag{124}$$

with $\mathbf{I}_l$ being the $l \times l$ identity matrix and $\mathbf{0}_{n-l,l}$ being the $(n - l) \times l$ zero matrix. Moreover, the inequality in (123) is due to the fact that $\mathbf{A} \succeq \mathbf{B}$ implies both that $\mathbf{C}^{\mathsf{T}}\mathbf{A}\mathbf{C} \succeq \mathbf{C}^{\mathsf{T}}\mathbf{B}\mathbf{C}$ and that $\mu_i(\mathbf{A}) \geq \mu_i(\mathbf{B})$ [26, Cor. 7.7.4(c)] and the lower bound on the derivative of the matrix $\mathbf{Q}(t)$ given in Lemma 6. We further used the definition

$$\mathbf{C}(t^\star) = \mathbf{V}(t^\star)\mathbf{B}(t^\star)\mathbf{V}(t^\star)^{\mathsf{T}}. \tag{125}$$

Observe that since $\mathbf{B}(t^\star)$ is a positive-semidefinite diagonal matrix (see Lemma 4), we have $\mathbf{C}(t^\star) \succeq \mathbf{0}$, which further implies that $[\mathbf{C}(t^\star)]_{ii} \geq 0$, for all i. Finally, the upper-left $l \times l$ submatrix of matrix $\mathbf{A}$ has been denoted by $[\mathbf{A}]_{1:l,1:l}$, and the last transition in (123) is due to the fact that both $\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)$ and $\boldsymbol{\Sigma}_G(t^\star)$ are diagonal matrices.

In order to proceed with the proof, we require the following lemma.

Lemma 11: Let us consider a positive-semidefinite matrix $\mathbf{A}$ and two diagonal positive-semidefinite matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ such that $\mathbf{D}_1 \succeq \mathbf{D}_2 \succeq \mathbf{0}$. Then, we have that

$$\mu_{\max}(\mathbf{D}_1\mathbf{A}\mathbf{D}_1 - \mathbf{D}_2\mathbf{A}\mathbf{D}_2) \geq 0 \tag{126}$$

where $\mu_{\max}$ denotes the maximum eigenvalue function.

Proof: See Appendix A7. ■

[2]The assumptions [28, Th. 7] are different from those in Lemma 8, but, once existence of the derivatives of the eigenvalues has been established, their expression has to be the same.

Now, using the fact that $\mathbf{C}(t^\star)$ is positive semidefinite, and the first condition in (120) that $\lambda_i(t^\star, t^\star) < 0$ for $i = 1, \ldots, l$, which further implies that $[\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)]_{1:l,1:l} \succ [\boldsymbol{\Sigma}_G(t^\star)]_{1:l,1:l} \succeq \mathbf{0}$ we can use Lemma 11 to conclude that

$$\begin{aligned} &\mu_{\max} \Big([\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)]_{1:l,1:l} [\mathbf{C}(t^\star)]_{1:l,1:l} [\boldsymbol{\Sigma}_{\boldsymbol{X}}(t^\star)]_{1:l,1:l} - \\ &\qquad [\boldsymbol{\Sigma}_G(t^\star)]_{1:l,1:l} [\mathbf{C}(t^\star)]_{1:l,1:l} [\boldsymbol{\Sigma}_G(t^\star)]_{1:l,1:l} \Big) \geq 0. \qquad (127) \end{aligned}$$

Last result together with (122)–(123) implies that there exists some $i \in [1, l]$ such that $\lambda_i(t^\star, t^\star) < 0$ and $\mathsf{D}_t \lambda_i(t, \tau = t^\star)\big|_{t=t^\star} \geq 0$, which clearly contradicts the conditions in (120).

Since the contradiction described previously holds for any arbitrary values for k, s, and l (under the condition $l \leq s \leq k \leq n$), we have thus proved that no nonnegative-to-negative zero crossing can occur for the eigenvalues of $\widetilde{\mathbf{Q}}(t, \tau)$ or, equivalently, we have proved that the eigenvalues of $\widetilde{\mathbf{Q}}(t, \tau)$ have at most a single negative-to-nonnegative zero crossing of the horizontal axis.

2) Single Crossing Point for the Eigenvalues of $\mathbf{Q}(t)$: The relation between the sign of the eigenvalues of $\mathbf{Q}(t)$ and those of $\widetilde{\mathbf{Q}}(t, \tau)$ is stated in the following lemma.

Lemma 12: For all τ as a function of t, the number of positive, zero, and negative eigenvalues of $\mathbf{Q}(t)$ and $\widetilde{\mathbf{Q}}(t, \tau)$ coincide.

Proof: The proof follows straightforwardly from the definition of $\widetilde{\mathbf{Q}}(t, \tau)$, given in (113), and Sylvester's law of inertia for congruent matrices [29, p. 5]. ■

In the first part of the proof, we have shown that $\widetilde{\mathbf{Q}}(t, \tau)$ has, for each eigenvalue, at most, a single negative-to-nonnegative zero crossing. From this and Lemma 12, we can conclude that the number of negative eigenvalues of both functions cannot increase. Now, let us assume that $\mathbf{Q}(t)$ has an eigenvalue of multiplicity s with a nonnegative-to-negative zero crossing at t_0, i.e., $\mu_i(\mathbf{Q}(t_0)) = 0$ and $\mu_i(\mathbf{Q}(t)) < 0$ for $t \in (t_0, t_0 + \varepsilon)$, for some positive ε and for $i = 1, \ldots, s$. In order to refrain from increasing the number of negative eigenvalues, s negative eigenvalues at t_0 must become zero. However, if we examine the number of eigenvalues at $t_0 + \Delta$ for a sufficiently small Δ, the eigenvalues that were negative at t_0 are still negative at $t_0 + \Delta$, and the total number of negative eigenvalues has increased, thus contradicting the possibility of a nonnegative-to-negative zero crossing of the multiplicity s eigenvalue of $\mathbf{Q}(t)$. This is valid for any arbitrary t_0, thus concluding our proof. ■

The following corollary is a simple consequence from Theorem 6.

Corollary 6: If for a given t' the function $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t') \succeq \mathbf{0}$, then for all $t \geq t'$ the function $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t) \succeq \mathbf{0}$.

B. Conditioned Case

The results of the previous section can be simply extended to the conditioned case. Given an extension of the lower bound on the derivative of $\mathbf{Q}$, the extension of all other results is trivial. Thus, we briefly give the extension of the lower bound with a full proof (given in Appendix A8) and then for completeness we restate the main result of this paper, for the conditioned case, without detailing the proof, which follows precisely the proof given previously.

Lemma 13: The following lower bound holds:

$$\mathsf{D}_t\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t) \succeq 2\left(\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t)\mathbf{B}(t)\mathbf{E}^{\mathsf{T}}_{\boldsymbol{X}|\boldsymbol{U}}(t) - \mathbf{E}_G(t)\mathbf{B}(t)\mathbf{E}^{\mathsf{T}}_G(t)\right) \quad (128)$$

where $\mathbf{B}(t)$ was defined in (45), and assumed a positive-semidefinite diagonal matrix for all t (see Lemma 4).

Proof: See Appendix A8. ■

Thus, the following theorem follows.

Theorem 7: Each eigenvalue of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ has, *at most*, a single negative-to-nonnegative zero crossing of the horizontal axis.

Proof: The proof follows the same steps as those in the proof of Theorem 6. ■

C. Properties of the Mutual Information

So far we have seen the "single crossing point" property of the matrix $\mathbf{Q}(\boldsymbol{X},\mathbf{R}_{\boldsymbol{G}},t)$, or more precisely, of its eigenvalues. As seen, this property also extends naturally to the conditioned case. In this section, our goal is to relate this result to the mutual information between the input and the output of a parallel Gaussian channel. As expected, the advantage of this result is in the comparison between the mutual information assuming that the input to the channel has an arbitrary distribution and the mutual information assuming that it has a Gaussian distribution with an arbitrary covariance, $\mathbf{R}_{\boldsymbol{G}}$. Our goal is to make use of this result through the I-MMSE relationship, as given in (43)–(44) and (46)–(47). The results given in this section can be viewed as supporting theorem/lemmas that make our "single crossing point" property applicable through the use of the I-MMSE relationship.

For clarity, we will write the results in this section only for, the more general, conditioned case, from which one can easily derive the respective unconditioned theorems.

According to (47), the difference between the mutual information assuming that the input to the channel has an arbitrary distribution and the mutual information assuming that it has a Gaussian distribution with an arbitrary covariance $\mathbf{R}_{\boldsymbol{G}}$ is

$$\begin{aligned} I\left(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}(t)\right) &- I\left(\boldsymbol{X};\boldsymbol{Y}(t)|\boldsymbol{U}\right) \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)(\mathbf{E}_G(\tau)-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(\tau))\right)\mathrm{d}\tau \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau)\right)\mathrm{d}\tau. \quad (129)\end{aligned}$$

Thus, we are interested in the properties of

$$\mathsf{Tr}\left(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)\right) = \sum_{i=1}^{n}\lambda_i(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t))$$

where we have used the fact that the trace of a matrix $\mathbf{A}$ is the sum of its eigenvalues [26, Th. 1.2.12]. The following theorem extends the "single crossing point" property of the eigenvalues of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ to the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$.

Theorem 8: Each eigenvalue of $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ has, *at most*, a single negative-to-nonnegative zero crossing of the horizontal axis. Moreover, the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ have the following property:

$$\mathrm{sign}\left\{\lambda_i(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t))\right\} \in \left\{0,\mathrm{sign}\left\{\lambda_i(\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t))\right\}\right\}.$$

Proof: For a nonsingular $\mathbf{B}(t)$ and due to similarity [26, Cor. 1.3.4], we can write the following:

$$\lambda_i(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)) = \lambda_i(\mathbf{B}^{\frac{1}{2}}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)\mathbf{B}^{\frac{1}{2}}(t)).$$

Recalling that $\mathbf{B}(t)$ is a positive-semidefinite diagonal matrix, we have an eigenvalue of a congruent transformation. Thus, the proof follows similarly to the second part of the proof of Theorem 6 (given in Section V-A2), concluding the preservation of the signs of the eigenvalues of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ in $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ and, as a result, concluding that all eigenvalues have, *at most*, a single, negative-to-nonnegative zero crossing of the horizontal axis.

If $\mathbf{B}(t)$ is singular, we can assume without loss of generality that the $i\mathrm{th}$ diagonal element is zero. Due to that, the $i\mathrm{th}$ row of $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ is all zeros, that is, one of the eigenvalues of $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ is zero (and its sign is also zero). The rest of the eigenvalues can be calculated from the reduced problem, the matrix $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ without the $i\mathrm{th}$ row and column. Recalling that $\mathbf{B}(t)$ is a diagonal matrix, this is simply the product of $\mathbf{B}(t)$ and $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ both without the $i\mathrm{th}$ row and column. This procedure can be repeated as long as the reduced $\mathbf{B}(t)$ matrix is singular. When the reduced matrix is nonsingular, we again follow the proof of Theorem 6.

Thus, we have shown that the eigenvalues preserve the sign of the eigenvalues of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ with the additional possibility of falling to zero when $\mathbf{B}(t)$ becomes singular. ■

The next two lemmas provide the link between the aforementioned results, regarding the behavior of the eigenvalues of the matrix $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ and the matrix $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$, and the mutual information. Thus, they facilitate the usage of these results on information theory problems, as will be shown in the sequel. More particularly, so far we discussed the behavior of each and every eigenvalue of the matrix $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$ and the matrix $\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t)$, which holds true for any proper choice of $\mathbf{R}_{\boldsymbol{G}}$ with no regards to the random vector $\boldsymbol{X}$. The next two lemmas identify the existence of specific Gaussian inputs which have unique properties with respect to the given random vector $\boldsymbol{X}$.

Lemma 14: Assume that $\boldsymbol{X} \in \mathbb{R}^n$ is an arbitrary distributed random vector. For any $t_e \in [0,\infty)$, there exists a Gaussian input covariance matrix $\mathbf{R}_{\boldsymbol{G}}$ such that the following hold:

1) $\mathbf{R}_{\boldsymbol{G}} \preceq \mathbf{R}_{\boldsymbol{X}}$
2) $I\left(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}\right) = I\left(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t_e)\right)$;
3) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e) \succeq \mathbf{0}$.

Proof: See Appendix A9. ■

Note that the aforementioned claim can be extended to a general nonsingular $\mathbf{H}(t_e)$, that is, not necessarily diagonal, by defining $\widetilde{\boldsymbol{X}} \equiv \mathbf{H}(t_e)\boldsymbol{X}$. Due to the nonsingularity of $\mathbf{H}(t_e)$, the mutual information is unchanged, i.e., $I\left(\widetilde{\boldsymbol{X}};\boldsymbol{Y}(t_e)|\boldsymbol{U}\right) = I\left(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}\right)$. Requirements 1 and 3 are preserved under any congruent transformation, specifically under the transformation $\mathbf{H}^{-1}(t_e)$.

The next lemma is an extension of Lemma 14 that will be prove useful in the sequel.

Lemma 15: Assume that for a given input distribution on the pair $(\boldsymbol{U}, \boldsymbol{X})$ there exists a Gaussian random vector, $\boldsymbol{X_G}^{\mathsf{ub}}$, with covariance $\mathbf{R}_{\boldsymbol{G}}^{\mathsf{ub}}$ such that for some $t_e \in [0, \infty)$ we have that

1) $I\left(\boldsymbol{X}; \boldsymbol{Y}(t_e) | \boldsymbol{U}\right) \le I\left(\boldsymbol{X_G}^{\mathsf{ub}}; \boldsymbol{Y_G}^{\mathsf{ub}}(t_e)\right)$;
2) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}^{\mathsf{ub}}, t_e) \succeq \mathbf{0}$.

Thus, there exists a Gaussian random vector $\boldsymbol{X_G}$ with covariance $\mathbf{R}_{\boldsymbol{G}}$ such that the following holds:

1) $\mathbf{R}_{\boldsymbol{G}} \preceq \mathbf{R}_{\boldsymbol{G}}^{\mathsf{ub}}$;
2) $I\left(\boldsymbol{X}; \boldsymbol{Y}(t_e) | \boldsymbol{U}\right) = I\left(\boldsymbol{X_G}; \boldsymbol{Y_G}(t_e)\right)$;
3) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \succeq \mathbf{0}$.

Proof: The proof follows the proof of Lemma 14, where instead of using $\mathbf{E}_{\boldsymbol{X}}^{\mathsf{lin}}(t_e)$ (196) as a trivial upper bound we use

$$\mathbf{E}_G^{\mathsf{ub}}(t_e) = \mathbf{I} - (\mathbf{R}_{\boldsymbol{G}}^{\mathsf{ub}} + \mathbf{I})^{-1} \tag{130}$$

and the assumptions stated previously. ■

D. Connections to Fisher Information

In addition to the MMSE matrix, another important quantity in estimation theory is the Fisher information matrix [30]. Its connection to information theory has been established in the late 1950s and has been attributed to de Bruijn [9]. The de Bruijn identity relates the derivative of the differential entropy to the Fisher information matrix defined as

$$\mathbf{J}(\boldsymbol{Y}) = \mathsf{E}\left\{\left[\nabla \log P_{\boldsymbol{Y}}(\boldsymbol{Y})\right]\left[\nabla \log P_{\boldsymbol{Y}}(\boldsymbol{Y})\right]^{\mathsf{T}}\right\} \tag{131}$$

where the expectation is over $\boldsymbol{Y}$. Note that this is a special form of the Fisher Information matrix (with respect to a translation parameter) which does not involve an explicit parameter as in its most general definition [30]. In [8], the authors have shown that the de Bruijn identity is equivalent to the I-MMSE relationship. Using this connection, the de Bruijn identity has been extended to a multivariate version in [19, Th. 4]. For our purposes, we will use the following notation:

$$\mathbf{J}_{\boldsymbol{X}}(\mathbf{H}) = \mathbf{J}(\mathbf{H}\boldsymbol{X} + \boldsymbol{N}) \tag{132}$$

when we have some arbitrary input distribution on the random vector $\boldsymbol{X}$. For the case of a Gaussian distribution on $\boldsymbol{X}$ with covariance matrix $\mathbf{R}_{\boldsymbol{G}}$, we will write $\mathbf{J}_G(\mathbf{R}_{\boldsymbol{G}}, \mathbf{H})$. We further note that, as in the case of the MMSE matrix, whenever the channel coefficients depend on other parameters, $\mathbf{H} = \mathbf{H}(\phi)$, we will write $\mathbf{J}_{\boldsymbol{X}}(\phi)$. We can now extend the idea of the the matrix $\mathbf{Q}$ to the Fisher Information, using the following definition:

$$\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, \phi) = \mathbf{J}_{\boldsymbol{X}}(\phi) - \mathbf{J}_G(\mathbf{R}_{\boldsymbol{G}}, \phi). \tag{133}$$

As in the case of the matrix $\mathbf{Q}$, the matrix $\mathbf{W}$ has some distinct properties. Using the relationship between the two matrices, we can derive these properties directly from the results of the previous sections. We first require the following lemma, given by Palomar and Verdú in [19].

Lemma 16 [19, App. E]: Assuming the Gaussian additive noise channel (5), the following connection between the Fisher Information matrix and the MMSE matrix holds:

$$\mathbf{J}_{\boldsymbol{Y}} = \mathbf{I}_n - \mathbf{H}\mathbf{E}_{\boldsymbol{X}}\mathbf{H}^{\mathsf{T}}. \tag{134}$$

Proof: The result follows directly from (106) in [19] by setting $\tilde{\boldsymbol{\Sigma}}_n$ equal to the identity matrix and recalling that the MMSE matrix in (106) is the MMSE matrix of $\boldsymbol{Z} = \mathbf{H}\boldsymbol{X}$, from which it follows that $\mathbf{E}_{\boldsymbol{Z}} = \mathbf{H}\mathbf{E}_{\boldsymbol{X}}\mathbf{H}^{\mathsf{T}}$. ■

We can now state the main result of this section.

Theorem 9: The matrix $\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ is related to the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ as follows:

$$\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t) = \mathbf{H}(t)\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)\mathbf{H}(t)^{\mathsf{T}}. \tag{135}$$

Moreover, the properties given in Sections IV and V for the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ transfer to the matrix $\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$.

Proof: Equation (135) is obtained through the use of Lemma 16. The properties given in Section IV regarding the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ transfer to the matrix $\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$, due to the fact that $\mathbf{H}(t)$ is a diagonal positive-semidefinite matrix for all t. The properties given in Section V regarding the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ transfer to the matrix $\mathbf{W}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$, since it is simply a congruent transformation of $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$ (this was explained in detail in part two of the proof of Theorem 6). ■

E. Application: The Degraded Parallel Gaussian BC Capacity Region Under a Covariance Constraint

In this section, we show that the result of Section V-C can be used to provide a converse proof for the *degraded* parallel Gaussian BC capacity region under an input covariance constraint. We consider the following model:

$$\begin{aligned}\boldsymbol{Y}_1[m] &= \mathbf{H}_1\boldsymbol{X}[m] + \boldsymbol{N}_1[m]\\ \boldsymbol{Y}_2[m] &= \mathbf{H}_2\boldsymbol{X}[m] + \boldsymbol{N}_2[m]\end{aligned} \tag{136}$$

where $\boldsymbol{N}_1[m]$ and $\boldsymbol{N}_2[m]$ are standard additive Gaussian noise vectors independent of different time indices m, and $\mathbf{H}_1$ and $\mathbf{H}_2$ are diagonal positive-semidefinite matrices such that $\mathbf{H}_1 \preceq \mathbf{H}_2$. $\boldsymbol{X} \in \mathbb{R}^n$ is the random input vector, and it is assumed independent of different time indices m.

We consider an input covariance constraint

$$\mathbf{R}_{\boldsymbol{X}} \preceq \mathbf{S} \tag{137}$$

where $\mathbf{S}$ is some positive-definite matrix.

Since we have a *degraded* BC, we can use the single-letter expression as given in (97). As in Section IV-D, we will follow the proof given for the scalar Gaussian BC in [7] and [20]. Using Lemma 4, we can construct a path such that

$$\begin{aligned}\mathbf{H}(t_2) &= \mathbf{H}_2\\ \mathbf{H}(t_1) &= \mathbf{H}_1\\ \mathbf{H}(0) &= \mathbf{0}\end{aligned} \tag{138}$$

where $0 \le t_1 \le t_2$ and $\mathbf{H}(t)$ is diagonal for all $t \in [0, t_2]$.

Now, assume a pair $(\boldsymbol{U}, \boldsymbol{X})$ with covariance $\mathbf{R}_{\boldsymbol{X}}$ for $\boldsymbol{X}$. According to Lemma 14, there exists a Gaussian random vector with covariance $\mathbf{R}_{\boldsymbol{G}}$ such that the following properties hold:

1) $\mathbf{R}_{\boldsymbol{G}} \preceq \mathbf{R}_{\boldsymbol{X}}$;
2) $I(\boldsymbol{X};\boldsymbol{Y}(t_1)|\boldsymbol{U}) = I(\boldsymbol{X_G};\boldsymbol{Y_G}(t_1))$;
3) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t) \succeq \mathbf{0}$ for all $t \geq t_1$.

Using the I-MMSE relationship (47), we can write

$$\begin{aligned} I(\boldsymbol{X_G};\mathbf{H}(t)\boldsymbol{X_G}+\boldsymbol{N}) &- I(\boldsymbol{X};\mathbf{H}(t)\boldsymbol{X}+\boldsymbol{N}|\boldsymbol{U}) \\ &= \int_{\tau=0}^{t} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau)\right)\mathrm{d}\tau \qquad (139)\\ &= \int_{\tau=0}^{t}\sum_{i=1}^{n}\lambda_i\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau)\right)\mathrm{d}\tau. \qquad (140)\end{aligned}$$

Using the aforementioned properties on (140), we have that for any $t' > t_1$,

$$\begin{aligned} I(\boldsymbol{X_G};\mathbf{H}(t')\boldsymbol{X_G}+\boldsymbol{N}) &- I(\boldsymbol{X};\mathbf{H}(t')\boldsymbol{X}+\boldsymbol{N}|\boldsymbol{U}) \\ &= \int_{\tau=0}^{t_1} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau)\right)\mathrm{d}\tau \\ &\quad + \int_{\tau=t_1}^{t'} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau)\right)\mathrm{d}\tau \\ &= 0+\int_{\tau=t_1}^{t'}\sum_{i=1}^{n}\lambda_i(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},\tau))\,\mathrm{d}\tau \geq 0 \qquad (141)\end{aligned}$$

where (141) follows from property 2, and the inequality follows from property 3 and Theorem 8.

Thus, we have shown the existence of a Gaussian random vector $\boldsymbol{X_G}$ with covariance matrix $\mathbf{R}_{\boldsymbol{G}}$, with the following properties:

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}(t_1)|\boldsymbol{U}) &= I(\boldsymbol{X_G};\boldsymbol{Y_G}(t_1)) \\ I(\boldsymbol{X};\boldsymbol{Y}(t_2)|\boldsymbol{U}) &\leq I(\boldsymbol{X_G};\boldsymbol{Y_G}(t_2)) \\ \mathbf{R}_{\boldsymbol{G}} &\preceq \mathbf{R}_{\boldsymbol{X}}. \end{aligned} \qquad (142)$$

Using these properties on the single-letter expression (97), we obtain the following outer bound:

$$\begin{aligned} \mathsf{R}_1 &\leq I(\boldsymbol{U};\boldsymbol{Y}_1) = I(\boldsymbol{X};\boldsymbol{Y}_1) - I(\boldsymbol{X};\boldsymbol{Y}_1|\boldsymbol{U}) \\ &\leq \frac{1}{2}\mathsf{log}|\mathbf{I}+\mathbf{H}_1\mathbf{S}\mathbf{H}_1^{\mathsf{T}}| - \frac{1}{2}\mathsf{log}|\mathbf{I}+\mathbf{H}_1\mathbf{R}_{\boldsymbol{G}}\mathbf{H}_1^{\mathsf{T}}| \\ &= \frac{1}{2}\mathsf{log}\frac{|\mathbf{I}+\mathbf{H}_1\mathbf{S}\mathbf{H}_1^{\mathsf{T}}|}{|\mathbf{I}+\mathbf{H}_1\mathbf{R}_{\boldsymbol{G}}\mathbf{H}_1^{\mathsf{T}}|} \qquad (143)\\ \mathsf{R}_2 &\leq I(\boldsymbol{X};\boldsymbol{Y}_2|\boldsymbol{U}) \leq \frac{1}{2}\mathsf{log}|\mathbf{I}+\mathbf{H}_2\mathbf{R}_{\boldsymbol{G}}\mathbf{H}_2^{\mathsf{T}}|. \qquad (144)\end{aligned}$$

This outer bound is tight and the achievability is well known using superposition coding. This approach can be extended to the M-user scenario as shown in Appendix C.

F. Application: The Compound Degraded Parallel Gaussian BC Capacity Region Under a Covariance Constraint

In this section, we show that the results of Section V-A can also be used to provide a converse proof for the compound *degraded* parallel Gaussian BC capacity region under an input covariance constraint. We consider the following model:

$$\boldsymbol{Y}_{i_j}^{j}[m] = \mathbf{H}_{i_j}^{j}\boldsymbol{X}[m]+\boldsymbol{N}_{i_j}^{j}[m], \quad j=1,\ldots,M, \quad i_j=1,\ldots,K_j \qquad (145)$$

where $\boldsymbol{N}_{i_j}^{j}$, $j=1,..,M, i_j=1,\ldots,K_j$, are standard additive Gaussian noise vectors independent of different time indices m, and $\mathbf{H}_{i_j}^{j}$, $j=1,\ldots,M, i_j=1,\ldots,K_j$, are diagonal positive-definite matrices such that

$$\mathbf{H}_{i_j}^{j} \preceq \mathbf{H}_{i_{(j+1)}}^{j+1} \quad \forall j=1,\ldots,M, i_j\in\{1,\ldots,K_j\}, i_{j+1}\in\{1,\ldots,K_{j+1}\}. \qquad (146)$$

Since these matrices are diagonal, there exist matrices $\mathbf{H}_{(j+1)j}^{\star}$ for $j=1,\ldots,M-1$ such that

$$\mathbf{H}_{i_j}^{j} \preceq \mathbf{H}_{(j+1)j}^{\star} \preceq \mathbf{H}_{i_{(j+1)}}^{j+1} \quad \forall j=1,\ldots,M-1, i_j\in\{1,\ldots,K_j\}, i_{j+1}\in\{1,\ldots,K_{j+1}\}. \qquad (147)$$

Note that the equivalence between conditions (146) and (147) is not true in general (for nondiagonal matrices), as explained in [31]. $\boldsymbol{X}\in\mathbb{R}^n$ is the random input vector, and it is assumed independent of different time indices m. We consider an input covariance constraint

$$\mathbf{R}_{\boldsymbol{X}} \preceq \mathbf{S} \qquad (148)$$

where $\mathbf{S}$ is some positive-definite matrix.

Before proceeding, we provide the following single-letter expression for the capacity region of this M user memoryless channel. This is a simple extension of [31, Lem. 4].

Lemma 17: Consider a memoryless compound BC with input $\boldsymbol{X}$, M outputs $\boldsymbol{Y}_{i_j}^{j}$, $j=1,\ldots,M, i_j=1,\ldots,K_j$, and auxiliary random outputs $\boldsymbol{Y}_{(j+1)j}^{\star}$ with $j\in\{1,\ldots,M-1\}$. All outputs are defined by their conditional probability functions: $P_{\boldsymbol{Y}_{i_j}^{j}|\boldsymbol{X}}$ and $P_{\boldsymbol{Y}_{(j+1)j}^{\star}|\boldsymbol{X}}$. Furthermore, assume that these outputs are stochastically *degraded* such that there exists some distribution such that

$$\boldsymbol{X}-\boldsymbol{Y}_{i_M}^{M}-\boldsymbol{Y}_{M(M-1)}^{\star}-\boldsymbol{Y}_{i_{M-1}}^{M-1}-\boldsymbol{Y}_{(M-1)(M-2)}^{\star}- \ldots-\boldsymbol{Y}_{i_2}^{2}-\boldsymbol{Y}_{21}^{\star}-\boldsymbol{Y}_{i_1}^{1}$$

form a Markov chain for every choice of $i_1,i_2,\ldots,i_M$. The capacity region of this channel is given by the union of the rate tuples satisfying

$$\mathsf{R}_j \leq \min_{i_j=1,\ldots,K_j} I\left(\boldsymbol{V}_j;\boldsymbol{Y}_{i_j}^{j}|\boldsymbol{V}_{j-1}\right) \qquad (149)$$

where $\boldsymbol{V}_0\equiv\emptyset$, $\boldsymbol{V}_M\equiv\boldsymbol{X}$, and the union is over all probability distributions satisfying

$$\boldsymbol{V}_0-\boldsymbol{V}_1-\ldots-\boldsymbol{V}_{M-1}-\boldsymbol{V}_M-\boldsymbol{X}-\boldsymbol{Y}_{i_M}^{M}-\boldsymbol{Y}_{M(M-1)}^{\star}- -\boldsymbol{Y}_{i_{M-1}}^{M-1}-\boldsymbol{Y}_{(M-1)(M-2)}^{\star}-\ldots\boldsymbol{Y}_{i_2}^{2}-\boldsymbol{Y}_{21}^{\star}-\boldsymbol{Y}_{i_1}^{1}. \qquad (150)$$

Proof: See Appendix D. ■

Using Lemma 17, we prove the following theorem.

Theorem 10: The capacity region of the compound *degraded* parallel Gaussian BC (145) is given by the following expression:

$$\begin{aligned} \mathsf{R}_M &\leq \min_{i_M=1,\ldots,K_M} \frac{1}{2}\log\left|\mathbf{H}_{i_M}^M \mathbf{R}_{\boldsymbol{G}M}\left(\mathbf{H}_{i_M}^M\right)^{\mathsf{T}}+\mathbf{I}\right| \\ \mathsf{R}_j &\leq \min_{i_j=1,\ldots,K_j} \frac{1}{2}\log\frac{\left|\mathbf{H}_{i_j}^j \sum_{l=j}^M \mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_{i_j}^j\right)^{\mathsf{T}}+\mathbf{I}\right|}{\left|\mathbf{H}_{i_j}^j \sum_{l=j+1}^M \mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_{i_j}^j\right)^{\mathsf{T}}+\mathbf{I}\right|} \\ &\forall j=1,\ldots,M-1 \end{aligned} \tag{151}$$

where $\mathbf{R}_{\boldsymbol{G}j}$ are some positive-semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^M \mathbf{R}_{\boldsymbol{G}l} \preceq \mathbf{S}$.

Proof: According to Lemma 4 (and the remark after this lemma) for any set of $\{i_1, i_2, \ldots, i_M\}$ where $i_j \in K_j$, we can construct a diagonal path such that

$$\begin{aligned} \mathbf{H}(t_{i_j}) &= \mathbf{H}_{i_j}, \quad j=1,\ldots,M \\ \mathbf{H}(t_{(j+1)j}) &= \mathbf{H}^\star_{(j+1)j}, \quad j=1,\ldots,M-1 \\ \mathbf{H}(t=0) &= \mathbf{0} \end{aligned} \tag{152}$$

with

$$\begin{aligned} 0 \leq t_{i_1} &\leq t_{21} \leq t_{i_2} \leq \ldots \leq t_{i_j} \\ &\leq t_{(j+1)j} \leq t_{i_{j+1}} \leq \ldots \leq t_{i_M}. \end{aligned}$$

Now, let us examine a tuple of rates on the boundary of the capacity region: $(\mathsf{R}_1^{\mathrm{opt}}, \mathsf{R}_2^{\mathrm{opt}}, \ldots, \mathsf{R}_M^{\mathrm{opt}})$. Assume that this tuple has been attained by the joint distribution $P_{\boldsymbol{V_1},\ldots,\boldsymbol{V_{M-1}},\boldsymbol{X}}$ on the tuple with covariance $\mathbf{R}_{\boldsymbol{X}} \preceq \mathbf{S}$ as required by the constraint (148).

We begin by looking at the following partial Markov chain:

$$\begin{aligned} \boldsymbol{V}_0 - \boldsymbol{V}_1 - \ldots - \boldsymbol{V}_{M-1} - \boldsymbol{V}_M - \boldsymbol{X} - \boldsymbol{Y}^\star_{M(M-1)} - \\ -\boldsymbol{Y}^\star_{(M-1)(M-2)} - \ldots - \boldsymbol{Y}^\star_{21}. \end{aligned} \tag{153}$$

Now, assuming that $\boldsymbol{Y}^\star_{(j+1)j}$ are the outputs, we can use Lemma 22 (given in Appendix C) which states that there exist M Gaussian inputs $\boldsymbol{X}_{G_j}$, with covariance matrices $\mathbf{R}_{\boldsymbol{G}j}$ such that

$$\int_0^{t_{(j+1)j}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau = 0 \tag{154}$$

$$\begin{aligned} \int_0^{t_{(j+2)j}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \geq 0 \\ \forall j=1,\ldots,M-2 \end{aligned} \tag{155}$$

and such that $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}j} \preceq \mathbf{R}_{\boldsymbol{G}j-1}$, for $j=2,\ldots,M-1$ and $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}1} \preceq \boldsymbol{S}$. Furthermore,

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t_{(j+1)j}) \succeq \mathbf{0} \tag{156}$$

for all $j=1,\ldots,M-1$.

Using this result, and according to Corollary 6 we know that $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t) \succeq \mathbf{0}$ for all $t \geq t_{(j+1)j}$. This holds for any diagonal path, such that $\mathbf{H}(t_{(j+1)j}) = \mathbf{H}^\star_{(j+1)j}$. Now, using Theorem 8 and (154) we can conclude that

$$\begin{aligned} \int_0^t \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \leq 0 \quad \forall t \leq t_{(j+1)j} \\ \int_0^t \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \geq 0 \quad \forall t \geq t_{(j+1)j}. \end{aligned} \tag{157}$$

Due to the Markov chain

$$\boldsymbol{V}_j - \boldsymbol{V}_{j+1} - \boldsymbol{X} - \boldsymbol{Y}^{j+1}_{i_{j+1}} - \boldsymbol{Y}^\star_{(j+1)j} - \boldsymbol{Y}^j_{i_j}. \tag{158}$$

Equation (157) is particularly valid for

$$\begin{aligned} \int_0^{t_{i_j}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \leq 0 \quad \forall i_j \in K_j \\ \int_0^{t_{i_{j+1}}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \geq 0 \quad \forall i_{j+1} \in K_{j+1} \end{aligned} \tag{159}$$

for any $j=1,\ldots,M-1$. Equation (149) can be written explicitly, as follows:

$$\begin{aligned} \mathsf{R}_M &\leq \min_{i_M=1,\ldots,K_M} I\left(\boldsymbol{X}; \boldsymbol{Y}^M_{i_M}|\boldsymbol{V}_{M-1}\right) \\ \mathsf{R}_j &\leq \min_{i_j=1,\ldots,K_j} I\left(\boldsymbol{X}; \boldsymbol{Y}^j_{i_j}|\boldsymbol{V}_{j-1}\right) \\ &\quad - I\left(\boldsymbol{X}; \boldsymbol{Y}^j_{i_j}|\boldsymbol{V}_j\right), \quad j=2,\ldots,M-1 \\ \mathsf{R}_1 &\leq \min_{i_1=1,\ldots,K_1} I\big(\boldsymbol{X}; \boldsymbol{Y}^1_{i_1}|\boldsymbol{V}_0 \equiv \emptyset\big) - I\big(\boldsymbol{X}; \boldsymbol{Y}^1_{i_1}|\boldsymbol{V}_1\big). \end{aligned} \tag{160}$$

Using (159) and the trivial bound on $I\left(\boldsymbol{X}; \boldsymbol{Y}^1_{i_1}\right)$, we can upper bound these expressions as follows:

$$\begin{aligned} \mathsf{R}_M &\leq \min_{i_M=1,\ldots,K_M} I\left(\boldsymbol{X}_{G_{M-1}}; \mathbf{H}^M_{i_M}\boldsymbol{X}_{G_{M-1}}+\boldsymbol{N}\right) \\ \mathsf{R}_j &\leq \min_{i_j=1,\ldots,K_j} I\left(\boldsymbol{X}_{G_{j-1}}; \mathbf{H}^j_{i_j}\boldsymbol{X}_{G_{j-1}}+\boldsymbol{N}\right) - \\ &\quad I\left(\boldsymbol{X}_{G_j}; \mathbf{H}^j_{i_j}\boldsymbol{X}_{G_j}+\boldsymbol{N}\right), \quad j=2,\ldots,M-1 \\ \mathsf{R}_1 &\leq \min_{i_1=1,\ldots,K_1} \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}^1_{i_1}\boldsymbol{S}\left(\mathbf{H}^1_{i_1}\right)^{\mathsf{T}}\right| - \\ &\quad I\left(\boldsymbol{X}_{G_1}; \mathbf{H}^1_{i_1}\boldsymbol{X}_{G_1}+\boldsymbol{N}\right). \end{aligned} \tag{161}$$

Defining

$$\begin{aligned} \mathbf{R}_{\boldsymbol{G}1} &= \boldsymbol{S} - \mathbf{R}_{\boldsymbol{G}1} \\ \mathbf{R}_{\boldsymbol{G}j} &= \mathbf{R}_{\boldsymbol{G}j-1} - \mathbf{R}_{\boldsymbol{G}j} \quad \forall j=2,\ldots,M-1 \\ \mathbf{R}_{\boldsymbol{G}M} &= \mathbf{R}_{\boldsymbol{G}M-1} \end{aligned} \tag{162}$$

(161) becomes the following set of upper bound:

$$\begin{aligned} \mathsf{R}_M &\leq \min_{i_M=1,\ldots,K_M} \frac{1}{2}\log\left|\mathbf{I}+\mathbf{H}^M_{i_M}\mathbf{R}_{\boldsymbol{G}M}\left(\mathbf{H}^M_{i_M}\right)^T\right| \\ \mathsf{R}_j &\leq \min_{i_j=1,\ldots,K_j} \frac{1}{2}\log\frac{\left|\mathbf{I}+\mathbf{H}^j_{i_j}\sum_{l=j}^M \mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}^j_{i_j}\right)^T\right|}{\left|\mathbf{I}+\mathbf{H}^j_{i_j}\sum_{l=j+1}^M \mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}^j_{i_j}\right)^T\right|} \\ &\forall j=1,\ldots,M-1 \end{aligned} \tag{163}$$

where $\mathbf{R}_{\boldsymbol{G}j}$ are some positive-semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^{M} \mathbf{R}_{\boldsymbol{G}l} = \boldsymbol{S}$.

The aforementioned upper bounds can be attained simultaneously using a joint Gaussian distribution on the tuple

$$(\boldsymbol{V}_0 \equiv \emptyset, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{M-1}, \boldsymbol{V}_M \equiv \boldsymbol{X}) \tag{164}$$

as follows:

$$\boldsymbol{V}_j = \boldsymbol{V}_{j-1} + \boldsymbol{U}_j \tag{165}$$

where $\boldsymbol{U}_j \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{\boldsymbol{G}j}\right)$ for $j = 1, \ldots, M$, independent of each other, and where $\mathbf{R}_{\boldsymbol{G}j}$ are positive-semidefinite matrices such that $\sum_{l=1}^{M} \mathbf{R}_{\boldsymbol{G}l} \preceq \boldsymbol{S}$. This concludes the proof of the capacity region. ■

VI. Summary

In this study, we extended the "single crossing point" property from the scalar setting to the parallel vector setting. We have shown three different "single crossing point" properties, given in three phases of extension from scalar to vector. These properties cannot be trivially deduced from each other. All three emphasize the basic optimality of the Gaussian input distribution in the Gaussian regime. The most general of these properties, given in the third phase, shows a "single crossing point" property for each of the eigenvalues of the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$, the difference between the MMSE matrix assuming an arbitrary Gaussian input, and the MMSE matrix assuming an arbitrary input distribution. We demonstrate the applicability of these properties on several information theoretic problems: a proof of a special case of Shannon's vector EPI, where one of the two random vectors is Gaussian, a converse proof of the capacity region of the parallel *degraded* BC under per-antenna power constraint and under an input covariance constraint, and a converse proof of the capacity region of the compound parallel *degraded* BC under an input covariance constraint.

An open question is: Can we extend the "single crossing point" property to the general MIMO channel? Note that although the optimality of the Gaussian input is known for several MIMO Gaussian multiterminal problems, we cannot necessarily conclude the existence of a "single crossing point" property. However, the implications of a general "single crossing point" property go beyond the specific applications shown here, and are also of interest on their own [12].

Appendix

A) Proofs of Lemmas:

1) Proof of Lemma 1: Since $\mathbf{A}$ is positive semidefinite, we can always write $\mathbf{A} = \alpha \bar{\mathbf{A}} \bar{\mathbf{A}}^{\mathsf{T}}$ such that $\mathsf{Tr}\left(\bar{\mathbf{A}}\bar{\mathbf{A}}^{\mathsf{T}}\right) = n$ and $\alpha \geq 0$. Then, it can be checked that

$$\begin{aligned} q_{\mathbf{A}}(\boldsymbol{X}, \sigma^2, \gamma) &= \frac{\sigma^2}{1+\sigma^2\gamma}\mathsf{Tr}\left(\mathbf{A}\right) - \mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right) && (166)\\ &= \alpha\left(n\frac{\sigma^2}{1+\sigma^2\gamma} - \mathsf{Tr}\left(\bar{\mathbf{A}}^{\mathsf{T}}\mathbf{E}_{\boldsymbol{X}}(\gamma)\bar{\mathbf{A}}\right)\right) && (167)\\ &= \alpha\left(n\frac{\sigma^2}{1+\sigma^2\gamma} - \mathsf{Tr}\left(\mathbf{E}_{\bar{\mathbf{A}}^{\mathsf{T}}\boldsymbol{X}}(\gamma)\right)\right) && (168)\\ &= \alpha q_{\mathbf{I}_n}(\hat{\boldsymbol{X}}, \sigma^2, \gamma) && (169) \end{aligned}$$

where we have defined $\hat{\boldsymbol{X}} = \bar{\mathbf{A}}^{\mathsf{T}}\boldsymbol{X}$. Now, from (169) and the fact that $\alpha \geq 0$, the desired result follows. ■

2) Proof of Lemma 2: Let us consider the random vector $\boldsymbol{X} \in \mathbb{R}^n$, whose covariance is given by $\mathbf{R}_{\boldsymbol{X}}$ and denote its eigenvalues by $\lambda_{\boldsymbol{X},i}$. Recalling the model in (13), it is well known that $\mathbf{E}_{\boldsymbol{X}}(\gamma) \preceq \mathbf{R}_{\boldsymbol{X}} - \gamma\mathbf{R}_{\boldsymbol{X}}(\gamma\mathbf{R}_{\boldsymbol{X}} + \mathbf{I}_n)^{-1}\mathbf{R}_{\boldsymbol{X}}$ [30]. Thus, we have that

$$\begin{aligned} \mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{X}}(\gamma)\right) &\leq \mathsf{Tr}\left(\mathbf{R}_{\boldsymbol{X}} - \gamma\mathbf{R}_{\boldsymbol{X}}(\gamma\mathbf{R}_{\boldsymbol{X}}+\mathbf{I}_n)^{-1}\mathbf{R}_{\boldsymbol{X}}\right) && (170)\\ &= \sum_{i=1}^{n}\left(\lambda_{\boldsymbol{X},i} - \frac{\gamma\lambda_{\boldsymbol{X},i}^2}{1+\gamma\lambda_{\boldsymbol{X},i}}\right) && (171)\\ &= \sum_{i=1}^{n}\frac{\lambda_{\boldsymbol{X},i}}{1+\gamma\lambda_{\boldsymbol{X},i}}. && (172) \end{aligned}$$

Now, realizing that the right-hand side in (172) is a Schur-concave function (it follows directly from the concavity of $\frac{\lambda}{1+\gamma\lambda}$) and that, from the statement of Lemma 2, we have $\sum_{i=1}^{n}\lambda_{\boldsymbol{X},i} \leq n\sigma^2$, it follows directly from majorization theory [32] that the right-hand side in (172) is maximized when $\lambda_{\boldsymbol{X},i}$ are uniformly distributed, i.e., $\lambda_{\boldsymbol{X},i} = \sigma^2$. ■

3) Proof of Lemma 3: From the definition in (16), it follows that

$$\mathsf{D}_\gamma q_{\mathbf{A}}(\boldsymbol{X}, \sigma^2, \gamma) = -\frac{\sigma^4}{(1+\sigma^2\gamma)^2}\mathsf{Tr}(\mathbf{A}) - \mathsf{D}_\gamma \mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right). \tag{173}$$

The expression for $\mathsf{D}_\gamma\mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right)$ can be computed from the results in [22] and applying the chain rule as

$$\begin{aligned} &\mathsf{D}_\gamma\mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right)\\ &= \mathsf{D}_{\mathbf{E}_{\boldsymbol{X}}(\gamma)}\mathsf{Tr}\left(\mathbf{A}\mathbf{E}_{\boldsymbol{X}}(\gamma)\right)\cdot\mathsf{D}_{\mathbf{H}}\mathbf{E}_{\boldsymbol{X}}(\gamma)\cdot\mathsf{D}_\gamma\mathbf{H}\\ &= \mathsf{vec}^{\mathsf{T}}\left(\mathbf{A}^{\mathsf{T}}\right)\mathbf{D}_n\Big(-2\mathbf{D}_n^{+}\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\otimes\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\\ &\quad\left(\mathbf{I}_n\otimes\mathbf{H}^{\mathsf{T}}\right)\Big)\frac{1}{2\sqrt{\gamma}}\mathsf{vec}(\mathbf{I}_n)\\ &= -\mathsf{vec}^{\mathsf{T}}\left(\mathbf{A}^{\mathsf{T}}\right)\mathbf{N}_n\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\otimes\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\mathsf{vec}(\mathbf{I}_n)\\ &= -\mathsf{Tr}\left(\mathbf{A}\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})^2\right\}\right) && (174) \end{aligned}$$

where we have used that $\mathbf{H} = \sqrt{\gamma}\mathbf{I}_n$, $\mathbf{N}_n\mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\otimes\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\} = \mathsf{E}\left\{\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\otimes\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\right\}\mathbf{N}_n$, and $\mathbf{N}_n\mathsf{vec}(\mathbf{I}_n) = \mathsf{vec}(\mathbf{I}_n)$ (see [22, App. A] for the definitions of the matrices $\mathbf{D}_n$ and $\mathbf{N}_n$ and some of their properties).

Plugging (174) into (173), the desired result follows. ■

4) Proof of Lemma 5: For any arbitrarily distributed random vector $\boldsymbol{X}$, with zero mean (assumed w.l.o.g.) and covariance matrix given by $\mathbf{R}_{\boldsymbol{X}}$, it is well known that $\mathbf{E}_{\boldsymbol{X}}(t) \preceq \mathbf{E}_G(\mathbf{R}_{\boldsymbol{X}}, t)$, from which it follows that [26, Obs. 7.1.2]

$$[\mathbf{E}_{\boldsymbol{X}}(t)]_{ii} \leq [\mathbf{E}_G(\mathbf{R}_{\boldsymbol{X}}, t)]_{ii} \tag{175}$$

where we recall that $\mathbf{E}_G(\mathbf{R}_{\boldsymbol{X}}, t)$ is the MMSE matrix attained assuming a zero mean Gaussian input with covariance matrix equal to $\mathbf{R}_{\boldsymbol{X}}$. Observe that equality in (175) is attained if and only if $\boldsymbol{X} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{\boldsymbol{X}}\right)$.

Furthermore, from the fact that dependence among entries can only improve the MMSE, we have

$$[\mathbf{E}_G(\mathbf{R}_{\boldsymbol{X}}, t)]_{ii} \leq [\mathbf{E}_G(\mathbf{I}_n \circ \mathbf{R}_{\boldsymbol{X}}, t)]_{ii} = \frac{[\mathbf{R}_{\boldsymbol{X}}]_{ii}}{1 + [\mathbf{H}(t)]_{ii}^2[\mathbf{R}_{\boldsymbol{X}}]_{ii}} \tag{176}$$

where $\mathbf{E}_G(\mathbf{I}_n \circ \mathbf{R}_{\boldsymbol{X}}, t)$ represents the MMSE matrix when the entries of the input vector are independent Gaussian random variables (thus, with diagonal covariance matrix). Observe that equality in (176) is obtained if and only if the entries of the Gaussian distribution on the left-hand side are independent.

Now, the desired result follows immediately from the fact that the right-hand side in (176) is an increasing function of $[\mathbf{R}_{\boldsymbol{X}}]_{ii}$. ■

5) Proof of Lemma 6: We first provide the derivative of the MMSE with respect to the parameter t. Using (52), we have

$$\mathsf{D}_t\,[\mathbf{E}_{\boldsymbol{X}}(t)]_{ij} = \sum_l \mathsf{D}_{[\mathbf{H}(t)]_{ll}}\,[\mathbf{E}_{\boldsymbol{X}}(t)]_{ij}\,[\mathsf{D}_t\mathbf{H}(t)]_{ll}\,. \tag{177}$$

Using the result ([22, eq. (131)]),

$$\begin{aligned}\mathsf{D}_{[\mathbf{H}(t)]_{ll}}\,[\mathbf{E}_{\boldsymbol{X}}(t)]_{ij} &= -\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,\big[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\mathbf{H}(t)^{\mathsf{T}}\big]_{il} + \\ &\qquad [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\,\big[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})\mathbf{H}(t)^{\mathsf{T}}\big]_{jl}\Big\} \\ &= -\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\,[\mathbf{H}(t)]_{ll} + \\ &\qquad [\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,[\mathbf{H}(t)]_{ll}\Big\} \\ &= -\,2\,[\mathbf{H}(t)]_{ll}\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\Big\}\end{aligned} \tag{178}$$

where $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ was defined in (9). The second equality in (178) is due to the fact that $\mathbf{H}(t)$ is diagonal. Thus, we can write the derivative of $[\mathbf{E}_{\boldsymbol{X}}(t)]_{ij}$ as

$$\begin{aligned}&\mathsf{D}_t\,[\mathbf{E}_{\boldsymbol{X}}(t)]_{ij} \\ &= -2\sum_l [\mathbf{H}(t)]_{ll}\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\Big\}\,[\mathsf{D}_t\mathbf{H}(t)]_{ll} \\ &= -2\sum_l [\mathbf{B}(t)]_{ll}\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{jl}\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_{il}\Big\}\end{aligned} \tag{179}$$

since $[\mathbf{B}(t)]_{ll} = [\mathbf{H}(t)]_{ll}\,[\mathsf{D}_t\mathbf{H}(t)]_{ll}$ (45). We can put this expression into a matrix form as follows:

$$\mathsf{D}_t\mathbf{E}_{\boldsymbol{X}}(t) = -2\sum_l [\mathbf{B}(t)]_{ll}\,\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l^{\mathsf{T}}\Big\} \tag{180}$$

where $[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})]_l$ is the lth column of the matrix $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$. Using the fact that for a Gaussian input distribution $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})$ does not depend on $\boldsymbol{Y}$ and thus $\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y}) = \mathsf{E}\{\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{y})\} = \mathbf{E}_G(t)$ [22], we can obtain the following lower bound on the derivative of the matrix $\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t)$:

$$\begin{aligned}&\mathsf{D}_t\mathbf{Q}(\boldsymbol{X}, \mathbf{R}_{\boldsymbol{G}}, t) \\ &= 2\sum_l [\mathbf{B}(t)]_{ll}\left(\mathsf{E}\Big\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l\,[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l^{\mathsf{T}}\Big\} - [\mathbf{E}_G]_l\,[\mathbf{E}_G]_l^{\mathsf{T}}\right) \\ &\succeq 2\sum_l [\mathbf{B}(t)]_{ll}\Big(\mathsf{E}\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l\}\,\mathsf{E}\{[\boldsymbol{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y})]_l\}^{\mathsf{T}} - [\mathbf{E}_G]_l\,[\mathbf{E}_G]_l^{\mathsf{T}}\Big) \\ &= 2\sum_l [\mathbf{B}(t)]_{ll}\left(\mathbf{E}_{\boldsymbol{X}l}\mathbf{E}_{\boldsymbol{X}l}^{\mathsf{T}} - [\mathbf{E}_G]_l\,[\mathbf{E}_G]_l^{\mathsf{T}}\right) \\ &= 2\left(\mathbf{E}_{\boldsymbol{X}}(t)\mathbf{B}(t)\mathbf{E}_{\boldsymbol{X}}^{\mathsf{T}}(t) - \mathbf{E}_G(t)\mathbf{B}(t)\mathbf{E}_G^{\mathsf{T}}(t)\right)\end{aligned}$$

where the inequality is due to Jensen. This concludes the proof of the lemma. ■

6) Proof of Lemma 7: Since $\mathbf{A}$ and $\mathbf{B}$ are two general positive-semidefinite matrices, the dimension of the intersection of their null spaces denoted by $N(\cdot)$ fulfills

$$\dim N(\mathbf{A}) \cap N(\mathbf{B}) = k, \quad 0 \le k \le n. \tag{181}$$

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis of the n-dimensional space such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis of $N(\mathbf{A}) \cap N(\mathbf{B})$ and define $\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_n]$. We thus have

$$\mathbf{U}^{\mathsf{T}}\mathbf{A}\mathbf{U} = \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}'\end{pmatrix}, \quad \mathbf{U}^{\mathsf{T}}\mathbf{B}\mathbf{U} = \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}'\end{pmatrix}, \tag{182}$$

where $\mathbf{A}'$ and $\mathbf{B}'$ are the nonzero $(n-k) \times (n-k)$ lower right square submatrices of $\mathbf{U}^{\mathsf{T}}\mathbf{A}\mathbf{U}$ and $\mathbf{U}^{\mathsf{T}}\mathbf{B}\mathbf{U}$, respectively. Observe that now we have $N(\mathbf{A}') \cap N(\mathbf{B}') = \{\emptyset\}$.

Now, from [26, Sec. 4.5,Prob. 8(e)], we have that $\mathbf{A}$ and $\mathbf{B}$ are simultaneously diagonalizable by an invertible matrix $\mathbf{S}$ if and only if $\mathbf{A}'$ and $\mathbf{B}'$ are also simultaneously diagonalizable. Consequently, we have reduced our proof to showing the simultaneous diagonalization of two positive-semidefinite matrices such that the dimension of the intersection of their null spaces is 0.

From this point, we can thus assume the following:

$$\mathbf{A} = \widetilde{\mathbf{A}}^{\mathsf{T}}\widetilde{\mathbf{A}} \succeq \mathbf{0} \tag{183}$$

$$\mathbf{B} = \widetilde{\mathbf{B}}^{\mathsf{T}}\widetilde{\mathbf{B}} \succeq \mathbf{0} \tag{184}$$

$$\dim N(\mathbf{A}) \cap N(\mathbf{B}) = 0. \tag{185}$$

The next step is to prove that $\mathbf{A}$ and $\mathbf{B}$ have no common isotropic vector, which is defined in [33, Def. 1.7.14] as a vector $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = 0$ and $\mathbf{x}^{\mathsf{T}}\mathbf{B}\mathbf{x} = 0$ are both simultaneously fulfilled.

Using the expression in (183), we have that

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = 0 \Leftrightarrow \mathbf{x}^{\mathsf{T}}\widetilde{\mathbf{A}}^{\mathsf{T}}\widetilde{\mathbf{A}}\mathbf{x} = 0 \Leftrightarrow \widetilde{\mathbf{A}}\mathbf{x} = 0 \tag{186}$$

which can also be applied to $\mathbf{x}^{\mathsf{T}}\mathbf{B}\mathbf{x} = 0$. Consequently, if a vector $\mathbf{x}$ fulfills $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = 0$ and $\mathbf{x}^{\mathsf{T}}\mathbf{B}\mathbf{x} = 0$, we have necessarily that $\mathbf{x} \in N(\widetilde{\mathbf{A}}) \cap N(\widetilde{\mathbf{B}})$. However, since $N(\widetilde{\mathbf{A}}) \subseteq N(\mathbf{A})$ and, similarly, $N(\widetilde{\mathbf{B}}) \subseteq N(\mathbf{B})$, from (185) we have that $\dim N(\widetilde{\mathbf{A}}) \cap N(\widetilde{\mathbf{B}}) = 0$, which implies that $\mathbf{A}$ and $\mathbf{B}$ have no common isotropic vector. Now, from [33, Th. 1.7.17] we have that $\mathbf{A}$ and $\mathbf{B}$ are simultaneously diagonalizable. ■

7) Proof of Lemma 11: For this proof, we require the following result.

Lemma 18: Let us consider two positive-semidefinie matrices $\mathbf{A}$ and $\mathbf{B}$. Then, we have

$$\mu_{\max}(\mathbf{A} - \mathbf{B}) \ge \mu_{\max}(\mathbf{A}) - \mu_{\max}(\mathbf{B}) \tag{187}$$

where we recall that $\mu_{\max}(\mathbf{A})$ denotes the maximum eigenvalue of matrix $\mathbf{A}$.

Proof: The proof follows directly from [26, Th. 4.3.1] recalling that $\mu_{\min}(-\mathbf{B}) = -\mu_{\max}(\mathbf{B})$. ■

Now, it is clear that for a positive-semidefinite matrix $\mathbf{A}$ and two positive-semidefinite diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ we have that $\mathbf{D}_i\mathbf{A}\mathbf{D}_i \succeq \mathbf{0}, i = 1, 2$. Now, from the aforementioned lemma we have

$$\begin{aligned}&\mu_{\max}(\mathbf{D}_1\mathbf{A}\mathbf{D}_1 - \mathbf{D}_2\mathbf{A}\mathbf{D}_2)\\ &\geq \mu_{\max}(\mathbf{D}_1\mathbf{A}\mathbf{D}_1) - \mu_{\max}(\mathbf{D}_2\mathbf{A}\mathbf{D}_2)\\ &= \mu_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{D}_1^2\mathbf{A}^{\frac{1}{2}}) - \mu_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{D}_2^2\mathbf{A}^{\frac{1}{2}})\end{aligned} \tag{188}$$

where the last equality follows from [26, Th. 1.3.20] and the remark, for square matrices, in the paragraph preceding it. Finally, since $\mathbf{D}_1 \succeq \mathbf{D}_2 \succeq \mathbf{0}$ and they are both diagonal, we have that $\mathbf{D}_1^2 \succeq \mathbf{D}_2^2 \succeq \mathbf{0}$ and, using [26, Obs. 7.7.2] and [26, Cor. 7.7.4] we can write

$$\mu_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{D}_1^2\mathbf{A}^{\frac{1}{2}}) \geq \mu_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{D}_2^2\mathbf{A}^{\frac{1}{2}}) \tag{189}$$

from which the desired result follows. ■

8) Proof of Lemma 13: We extend the lower bound derived in Lemma 6 to the conditioned case, that is, we assume $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}$. From (180), for the conditioned case we have the following:

$$\begin{aligned}&\mathsf{D}_t\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{u})\\ &= -2\sum_l [\mathbf{B}(t)]_{ll}\,\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}_{\boldsymbol{u}}}(\boldsymbol{Y})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}_{\boldsymbol{u}}}(\boldsymbol{Y})]_l^{\mathsf{T}}\right\}\\ &= -2\sum_l [\mathbf{B}(t)]_{ll}\,\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U}=\boldsymbol{u})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U}=\boldsymbol{u})]_l^{\mathsf{T}}\right\}.\end{aligned} \tag{190}$$

Taking expectation according to $\boldsymbol{U}$ on both sides, we have

$$\begin{aligned}&\mathsf{D}_t\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t) = \mathsf{E}\left\{\mathsf{D}_t\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t,\boldsymbol{U})\right\} =\\ &\quad -2\sum_l [\mathbf{B}(t)]_{ll}\,\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l^{\mathsf{T}}\right\}.\end{aligned} \tag{191}$$

The derivative of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{Y},\mathbf{R}_{\boldsymbol{G}},t)$ is then given in (192), shown at the bottom of the page, where the inequality is due to Jensen. This completes the proof of the lemma. ■

9) Proof of Lemma 14: We first claim that w.l.o.g. we can restrict the proof to $\mathbf{H}(t_e) = \mathbf{I}$. This is shown by redefining $\widetilde{\boldsymbol{X}} = \mathbf{H}(t_e)\boldsymbol{X}$. Now, if $\mathbf{H}(t_e)$ is nonsingular, then this redefinition does not change the mutual information i.e., $I\left(\widetilde{\boldsymbol{X}};\boldsymbol{Y}(t_e)|\boldsymbol{U}\right) = I\left(\boldsymbol{X};\boldsymbol{Y}(t_e)|\boldsymbol{U}\right)$, and requirements 1 and 3 are preserved under any congruent transformation. If $\mathbf{H}(t_e)$ is singular, the problem can first be reduced in size, since $\mathbf{H}(t_e)$ is diagonal for all t. Thus, from this point on, we will assume $\mathbf{H}(t_e) = \mathbf{I}$.

We provide a constructive proof, and show how one can build a Gaussian input distribution such that all three requirements are fulfilled. We begin by rewriting requirement 1 as a condition on the matrix $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$ rather than on the covariance matrix $\mathbf{R}_{\boldsymbol{G}}$. We do so by defining a new matrix, which is the distance of the MMSE matrix $\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e)$ from the linear MSE matrix $\mathbf{E}_{\boldsymbol{X}}^{\mathsf{lin}}(t_e)$. We proceed by showing that there exists a fraction such that by defining $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$ to be that fraction of the newly defined matrix, we comply also with requirement 2.

As explained previously, we begin by rewriting requirement 1 in terms of the matrix $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$. Requirement 3 is already a requirement on the matrix $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$ and is as follows:

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e) = \mathbf{E}_G(t_e) - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) \succeq \mathbf{0}. \tag{193}$$

The MMSE for the Gaussian input is

$$\begin{aligned}\mathbf{E}_G(t_e) &= \mathbf{R}_{\boldsymbol{G}} - \mathbf{R}_{\boldsymbol{G}}(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}\mathbf{R}_{\boldsymbol{G}}\\ &= \mathbf{R}_{\boldsymbol{G}} - \mathbf{R}_{\boldsymbol{G}}(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I}) +\\ &\quad \mathbf{R}_{\boldsymbol{G}}(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}\\ &= \mathbf{R}_{\boldsymbol{G}}(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}\\ &= (\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1} - (\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}\\ &= \mathbf{I} - (\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1}.\end{aligned} \tag{194}$$

From (193) $\mathbf{R}_{\boldsymbol{G}}$ complies with the following:

$$(\mathbf{R}_{\boldsymbol{G}}+\mathbf{I})^{-1} = \mathbf{I} - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) - \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e). \tag{195}$$

Note that the aforementioned equation connects $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$ with $\mathbf{R}_{\boldsymbol{G}}$. Thus, given a specific substitution of $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}_{\boldsymbol{G}},t_e)$ we have a complete definition of the Gaussian input distribution. Similarly, the MMSE assuming an optimal linear estimator of $\boldsymbol{X}$ (only from $\boldsymbol{Y}(t_e)$) is given by

$$\mathbf{E}_{\boldsymbol{X}}^{\mathsf{lin}} = \mathbf{I} - (\mathbf{R}_{\boldsymbol{X}}+\mathbf{I})^{-1} \tag{196}$$

and we have that

$$\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) \preceq \mathbf{E}_{\boldsymbol{X}}^{\mathsf{lin}}(t_e) \quad \forall t. \tag{197}$$

$$\begin{aligned}\mathsf{D}_t\mathbf{Q}(\boldsymbol{X}|\boldsymbol{Y},\mathbf{R}_{\boldsymbol{G}},t) &= 2\sum_l [\mathbf{B}(t)]_{ll}\left(\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l^{\mathsf{T}}\right\} - \mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}_{\boldsymbol{G}}}(\boldsymbol{Y})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}_{\boldsymbol{G}}}(\boldsymbol{Y})]_l^{\mathsf{T}}\right\}\right)\\ &= 2\sum_l [\mathbf{B}(t)]_{ll}\left(\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l\,[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l^{\mathsf{T}}\right\} - [\mathbf{E}_G(t)]_l\,[\mathbf{E}_G(t)]_l^{\mathsf{T}}\right)\\ &\succeq 2\sum_l [\mathbf{B}(t)]_{ll}\left(\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l\right\}\mathsf{E}\left\{[\mathbf{\Phi}_{\boldsymbol{X}}(\boldsymbol{Y},\boldsymbol{U})]_l^{\mathsf{T}}\right\} - [\mathbf{E}_G(t)]_l\,[\mathbf{E}_G(t)]_l^{\mathsf{T}}\right)\\ &= 2\sum_l [\mathbf{B}(t)]_{ll}\left(\left[\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t)\right]_l\left[\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t)\right]_l^{\mathsf{T}} - [\mathbf{E}_G(t)]_l\,[\mathbf{E}_G(t)]_l^{\mathsf{T}}\right)\\ &= 2\left(\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t)\mathbf{B}(t)\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}^{\mathsf{T}}(t) - \mathbf{E}_G(t)\mathbf{B}(t)\mathbf{E}_G^{\mathsf{T}}(t)\right)\end{aligned} \tag{192}$$

Thus, we can define

$$\mathbf{C} \equiv \mathbf{E}_{\boldsymbol{X}}^{\text{lin}}(t_e) - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) = \mathbf{I} - (\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) \succeq \mathbf{0}$$
$$\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t_e) = \mathbf{I} - (\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} - \mathbf{C}, \quad \mathbf{C} \succeq \mathbf{0}. \tag{198}$$

Note that $\mathbf{C}$ is completely defined by the input random vector $\boldsymbol{X}$. Inserting (198) into (195), we have

$$\begin{aligned}(\mathbf{R}_{\boldsymbol{G}} + \mathbf{I})^{-1} &= \mathbf{I} - \left[\mathbf{I} - (\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} - \mathbf{C}\right] - \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \\ &= (\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e), \\ &\quad \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \succeq \mathbf{0}, \quad \mathbf{C} \succeq \mathbf{0}.\end{aligned} \tag{199}$$

We now require the following supporting lemma.

Lemma 19: Assume $\boldsymbol{X} \in \mathbb{R}^n$ is an arbitrary distributed random vector. For any $t' \in [0, \infty)$, there exists a Gaussian random vector $\boldsymbol{X_G}$ with covariance matrix $\mathbf{R}_{\boldsymbol{G}}$ such that:

1) $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}} \preceq \mathbf{R}_{\boldsymbol{X}}$;
2) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t') = \mathbf{0}$;
3) $I(\boldsymbol{X_G}; \boldsymbol{Y_G}(t')) \leq I(\boldsymbol{X}; \boldsymbol{Y}(t')|\boldsymbol{U})$.

Proof: See Appendix A10. ■

Note that according to Lemma 19, we have that for $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) = \mathbf{0}$ there exists a Gaussian random vector $\boldsymbol{X_G}$, which ensures that $I(\boldsymbol{X_G}; \boldsymbol{Y_G}(t_e)) \leq I(\boldsymbol{X}; \boldsymbol{Y}(t_e)|\boldsymbol{U})$. On the other hand, if $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) = \mathbf{C}$ we have, according to (199), that $\mathbf{R}_{\boldsymbol{G}} = \mathbf{R}_{\boldsymbol{X}}$ in which case we have $I(\boldsymbol{X_G}; \boldsymbol{Y_G}(t_e)) \geq I(\boldsymbol{X}; \boldsymbol{Y}(t_e)|\boldsymbol{U})$. Moreover, from (199) we can observe that instead of requirement 1, i.e., $\mathbf{R}_{\boldsymbol{G}} \preceq \mathbf{R}_{\boldsymbol{X}}$, we may simply require $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \preceq \mathbf{C}$ (199); thus, requirements 1 and 3 can be written as follows:

$$\mathbf{0} \preceq \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \preceq \mathbf{C} \tag{200}$$

where $\mathbf{C}$ is defined in (198). The question is whether there exists such $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e) \preceq \mathbf{C}$ that will also attain requirement 2, i.e., $I(\boldsymbol{X_G}; \boldsymbol{Y_G}(t_e)) = I(\boldsymbol{X}; \boldsymbol{Y}(t_e)|\boldsymbol{U}) \equiv \alpha$. From the aforementioned, we know that

$$\frac{1}{2}\log|\mathbf{I} + \mathbf{R}_{\boldsymbol{G}}^1| \leq \alpha \leq \frac{1}{2}\log|\mathbf{I} + \mathbf{R}_{\boldsymbol{G}}^2| \tag{201}$$

where

$$\mathbf{I} + \mathbf{R}_{\boldsymbol{G}}^1 = \left((\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C}\right)^{-1} \tag{202}$$

$$\mathbf{I} + \mathbf{R}_{\boldsymbol{G}}^2 = \mathbf{I} + \mathbf{R}_{\boldsymbol{X}}. \tag{203}$$

Thus, (201) can be rewritten as

$$\begin{aligned}\frac{1}{2}\log\left|\left((\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C}\right)^{-1}\right| \leq \alpha \leq \frac{1}{2}\log|\mathbf{I} + \mathbf{R}_{\boldsymbol{X}}| \\ \frac{1}{2}\log\frac{|\mathbf{I}|}{|(\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C}|} \leq \alpha \leq \frac{1}{2}\log\frac{|\mathbf{I}|}{|(\mathbf{I} + \mathbf{R}_{\boldsymbol{X}})^{-1}|}.\end{aligned} \tag{204}$$

We now need the following result.

Lemma 20: Let us define the function

$$r(\nu) = \frac{1}{2}\log\frac{|\mathbf{A}|}{|\mathbf{B} + \boldsymbol{\Delta}\nu|}. \tag{205}$$

For $\mathbf{A} \succ \mathbf{0}, \mathbf{B} \succ \mathbf{0}$, and $\boldsymbol{\Delta} \succeq \mathbf{0}$, the function, $r(\nu)$ is continuous and monotonically decreasing in ν for $0 \leq \nu \leq 1$.

Proof: The proof is similar to the proof of Lemma 10 in [6]. ■

In our case, we have

$$\mathbf{A} = \mathbf{I} \succ \mathbf{0} \tag{206}$$

$$\mathbf{B} = (\mathbf{I} + \mathbf{R}_{\boldsymbol{X}})^{-1} \succ \mathbf{0} \tag{207}$$

$$\boldsymbol{\Delta} = \mathbf{C} \succeq \mathbf{0} \tag{208}$$

and

$$\frac{1}{2}\log\frac{|\mathbf{A}|}{|\mathbf{B} + \nu\boldsymbol{\Delta}|}\Bigg|_{\nu=1} \leq \alpha \leq \frac{1}{2}\log\frac{|\mathbf{A}|}{|\mathbf{B} + \nu\boldsymbol{\Delta}|}\Bigg|_{\nu=0}. \tag{209}$$

Thus, according to Lemma 20, there exists $\nu^\star$ such that $r(\nu^\star) = \alpha$. That is

$$\begin{aligned}\alpha &= \frac{1}{2}\log\frac{|\mathbf{A}|}{|\mathbf{B} + \nu^\star\boldsymbol{\Delta}|} = \frac{1}{2}\log\frac{|\mathbf{I}|}{|(\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \nu^\star\mathbf{C}|} \\ &= \frac{1}{2}\log|\mathbf{I} + \mathbf{R}_{\boldsymbol{X}_{\boldsymbol{G}^\star}}| \\ &= \frac{1}{2}\log\frac{|\mathbf{I}|}{|(\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e)^\star|}\end{aligned} \tag{210}$$

where the last equality is due to (199). That is,

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e)^\star = (1 - \nu^\star)\mathbf{C} \tag{211}$$

and since $0 \leq \nu^\star \leq 1$ we have that $\mathbf{0} \preceq \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e)^\star \preceq \mathbf{C}$, as required. To conclude, we can construct a Gaussian input distribution, complying with all three requirements, as follows:

$$\begin{aligned}\mathbf{I} + \mathbf{R}_{\boldsymbol{X}_{\boldsymbol{G}^\star}} &= \left((\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \mathbf{C} - \mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t_e)^\star\right)^{-1} \\ &= \left((\mathbf{R}_{\boldsymbol{X}} + \mathbf{I})^{-1} + \nu^\star\mathbf{C}\right)^{-1}\end{aligned} \tag{212}$$

where $\nu^\star$ is derived from the equality in (210). This completes the proof of the lemma. ■

10) Proof of Lemma 19: We first show that there exists a covariance matrix $\mathbf{R}_{\boldsymbol{G}}$ such that requirements 1 and 2 are fulfilled. Then, we will show, using contradiction, that requirement 3 is also fulfilled.

First note that requirement 2, i.e., $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U}, \mathbf{R}_{\boldsymbol{G}}, t') = \mathbf{0}$, completely defines $\mathbf{R}_{\boldsymbol{G}}$

$$\begin{aligned}\mathbf{I} - (\mathbf{R}_{\boldsymbol{G}} + \mathbf{I})^{-1} &= \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t') \\ (\mathbf{I} - \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t'))^{-1} - \mathbf{I} &= \mathbf{R}_{\boldsymbol{G}}\end{aligned} \tag{213}$$

where we have used the expression in (194), and using the expression in (196) we can show that $\mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t')$ is an invertible matrix since

$$\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t') \preceq \mathbf{E}^{\text{lin}}_{\boldsymbol{X}}(t') = \mathbf{I}-(\mathbf{R}_{\boldsymbol{X}}+\mathbf{I})^{-1} \prec \mathbf{I}. \tag{214}$$

We now need to check that the first requirement holds

$$\begin{aligned} \mathbf{R}_{\boldsymbol{G}} = (\mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t'))^{-1}-\mathbf{I} &\succeq \mathbf{0}\\ \mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t') &\preceq \mathbf{I}\\ \mathbf{0} &\preceq \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t') \end{aligned} \tag{215}$$

and

$$\begin{aligned} \mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t') \preceq \mathbf{E}^{\text{lin}}_{\boldsymbol{X}}(t') &= \mathbf{I}-(\mathbf{R}_{\boldsymbol{X}}+\mathbf{I})^{-1}\\ (\mathbf{R}_{\boldsymbol{X}}+\mathbf{I})^{-1} &\preceq \mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t')\\ \mathbf{R}_{\boldsymbol{X}}+\mathbf{I} &\succeq (\mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t'))^{-1}\\ \mathbf{R}_{\boldsymbol{X}}-(\mathbf{I}-\mathbf{E}_{\boldsymbol{X}|\boldsymbol{U}}(t'))^{-1}+\mathbf{I} &\succeq \mathbf{0}\\ \mathbf{R}_{\boldsymbol{X}} &\succeq \mathbf{R}_{\boldsymbol{G}}. \end{aligned} \tag{216}$$

Thus, we have shown that given an arbitrary input we can find the required $\mathbf{R}_{\boldsymbol{G}}$.

We now want to show that $I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t')) \le I(\boldsymbol{X};\boldsymbol{Y}(t')|\boldsymbol{U})$. For any Gaussian random vector $\boldsymbol{X}_{\boldsymbol{G}}{}^{\star}$ with covariance $\mathbf{R}^{\star}_{\boldsymbol{G}}$ such that $\mathbf{R}^{\star}_{\boldsymbol{G}} \prec \mathbf{R}_{\boldsymbol{G}}$, we have $\mathbf{E}^{\star}_{G}(t') \prec \mathbf{E}_{G}(t')$, and thus, $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{U},\mathbf{R}^{\star}_{\boldsymbol{G}},t') \prec \mathbf{0}$. Using Theorem 8, assuming that we do not have $\mathbf{B}(t)=\mathbf{0}$ for all $0\le t\le t'$,[3] and the I-MMSE relationship (47), we have

$$I(\boldsymbol{X}_{\boldsymbol{G}}{}^{\star};\boldsymbol{Y}_{\boldsymbol{G}}{}^{\star}(t')) < I(\boldsymbol{X};\boldsymbol{Y}(t')|\boldsymbol{U}). \tag{217}$$

Now, let us assume that

$$I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t')) > I(\boldsymbol{X};\boldsymbol{Y}(t')|\boldsymbol{U}). \tag{218}$$

The function $I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t'))$ is continuous in the value of its eigenvalues, since

$$I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t')) = \frac{1}{2}\sum_{i=1}^{n}\log(1+\lambda_i(\mathbf{R}_{\boldsymbol{G}})). \tag{219}$$

We can construct $\mathbf{R}^{\star}_{\boldsymbol{G}}$ by reducing by ϵ the value of all eigenvalues of $\mathbf{R}_{\boldsymbol{G}}$. According to (218), we can find a small enough ϵ, such that the following inequality still holds:

$$I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t')) > I(\boldsymbol{X}_{\boldsymbol{G}}{}^{\star};\boldsymbol{Y}(t')) \ge I(\boldsymbol{X};\boldsymbol{Y}(t')|\boldsymbol{U}) \tag{220}$$

but this contradicts (217) and by that proves that

$$I(\boldsymbol{X}_{\boldsymbol{G}};\boldsymbol{Y}_{\boldsymbol{G}}(t')) \le I(\boldsymbol{X};\boldsymbol{Y}(t')|\boldsymbol{U}). \tag{221}$$

This concludes the proof of the lemma. ∎

[3] $\mathbf{B}(t)=\mathbf{0}$ for all $0\le t\le t'$; then all mutual information equals zero regardless of the input distribution and the lemma holds trivially.

B) Converse Proof of BC Capacity Under Per-Antenna Constraints for M*-Users:* We consider the *degraded* parallel Gaussian BC channel:

$$\boldsymbol{Y}_j[m] = \mathbf{H}_j\boldsymbol{X}[m]+\boldsymbol{N}_j[m] \quad j=1,\ldots,M \tag{222}$$

where $\boldsymbol{N}_j[m]$, $j=1,..,M$, are standard additive Gaussian noise vectors independent of different time indices m (and can be considered independent of each other), and $\mathbf{H}_j$, $j=1,..,M$, are diagonal positive-semidefinite matrices such that $\mathbf{H}_j \preceq \mathbf{H}_{j+1}$, for all $j=1,\ldots,M-1$. $\boldsymbol{X}\in\mathbb{R}^n$ is the random input vector and it is assumed independent of different time indices m.

We consider an input per-antenna power constraint:

$$\left[\mathsf{E}\left\{\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}}\right\}\right]_{ii} \le P_i \quad \forall i, 1\le i\le n. \tag{223}$$

Since we have a *degraded* BC, we can use the single-letter expression given in [25]

$$\mathsf{R}_j \le I(\boldsymbol{V}_j;\boldsymbol{Y}_j|\boldsymbol{V}_{j-1}) \quad j=1,\ldots,M \tag{224}$$

where $\boldsymbol{V}_j$ are auxiliary random variables, $\boldsymbol{V}_M\equiv\boldsymbol{X}$, $\boldsymbol{V}_0\equiv\emptyset$, and the union is over all probability distributions satisfying

$$\begin{aligned}\boldsymbol{V}_0-\ldots-\boldsymbol{V}_{M-1}-\boldsymbol{V}_M-\boldsymbol{X}-\boldsymbol{Y}_M-\boldsymbol{Y}_{M-1}-\\ \ldots-\boldsymbol{Y}_2-\boldsymbol{Y}_1.\end{aligned} \tag{225}$$

This is an extension of the proof given for the two-user case. We begin by rewriting the single-letter expression (224) as follows:

$$\mathsf{R}_j \le I(\boldsymbol{X};\boldsymbol{Y}_j|\boldsymbol{V}_{j-1}) - I(\boldsymbol{X};\boldsymbol{Y}_j|\boldsymbol{V}_j) \quad j=1,..,M \tag{226}$$

and more explicitly

$$\begin{aligned} \mathsf{R}_1 &\le I(\boldsymbol{X};\boldsymbol{Y}_1) - I(\boldsymbol{X};\boldsymbol{Y}_1|\boldsymbol{V}_1)\\ \mathsf{R}_2 &\le I(\boldsymbol{X};\boldsymbol{Y}_2|\boldsymbol{V}_1) - I(\boldsymbol{X};\boldsymbol{Y}_2|\boldsymbol{V}_2)\\ &\vdots\\ \mathsf{R}_{M-2} &\le I(\boldsymbol{X};\boldsymbol{Y}_{M-2}|\boldsymbol{V}_{M-3}) - I(\boldsymbol{X};\boldsymbol{Y}_{M-2}|\boldsymbol{V}_{M-2})\\ \mathsf{R}_{M-1} &\le I(\boldsymbol{X};\boldsymbol{Y}_{M-1}|\boldsymbol{V}_{M-2}) - I(\boldsymbol{X};\boldsymbol{Y}_{M-1}|\boldsymbol{V}_{M-1})\\ \mathsf{R}_M &\le I(\boldsymbol{X};\boldsymbol{Y}_M|\boldsymbol{V}_{M-1}). \end{aligned} \tag{227}$$

According to Lemma 4 (and the remark after this lemma), we can construct a diagonal path such that

$$\begin{aligned} \mathbf{H}(t_j) &= \mathbf{H}_j \quad j=1,\ldots,M\\ \mathbf{H}(0) &= \mathbf{0} \end{aligned} \tag{228}$$

with $0\le t_1\le t_2\le\ldots\le t_M$. Now, assume a distribution $P_{\{\boldsymbol{V}_0\equiv\emptyset,\boldsymbol{V}_1,\ldots,\boldsymbol{V}_{M-1},\boldsymbol{V}_M\equiv\boldsymbol{X}\}}$ on the tuple $(\boldsymbol{V}_0\equiv\emptyset,\boldsymbol{V}_1,\ldots,\boldsymbol{V}_{M-1},\boldsymbol{V}_M\equiv\boldsymbol{X})$ with covariance matrix $\mathbf{R}_{\boldsymbol{X}}$. We begin by proving the following lemma.

Lemma 21: There exist M independent Gaussian inputs $\boldsymbol{X}_{G_j}$, with covariance matrices $\boldsymbol{\Lambda}_{G_j}$ such that

$$\int_0^{t_j} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, \tau)\,\mathrm{d}\tau = 0$$
$$\int_0^{t_{j+1}} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, \tau)\,\mathrm{d}\tau \geq 0 \quad \forall i \quad \forall j = 1, \ldots, M-1 \tag{229}$$

and such that $\left[\boldsymbol{\Lambda}_{\boldsymbol{G}_j}\right]_{ii} \leq \left[\boldsymbol{\Lambda}_{\boldsymbol{G}_{j-1}}\right]_{ii}$, for $j = 2, .., M-1$ and $[\boldsymbol{\Lambda}_{\boldsymbol{G}_1}]_{ii} \leq [\mathbf{R}_{\boldsymbol{X}}]_{ii}$.

Proof: We will prove the above using induction.

The case of $j = 1$: This is identical to the proof given in Section IV-D.

For a general j: We assume that the above holds for j and prove for $j+1$. Due to the Markov relation (225), we have that

$$\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t) = \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t) \quad \forall t \tag{230}$$

and thus

$$\begin{aligned}\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t) &= \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t)\\ &\geq \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t) \quad \forall t\end{aligned} \tag{231}$$

where the inequality is, again, due to the Markov relation (225) and the definition of the function $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, t)$, given in (77). This provides us with the following inequality:

$$\begin{aligned}\int_0^{t_{j+1}} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, \tau)\,\mathrm{d}\tau \geq\\ \int_0^{t_{j+1}} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_j, \boldsymbol{\Lambda}_{\boldsymbol{G}_j}, \tau)\,\mathrm{d}\tau \geq 0\end{aligned} \tag{232}$$

where the first inequality is due to (231) and the second is due to the induction assumption on j (229). Again, following the same derivation as in the proof in Section IV-D, we know that there exists an independent Gaussian input with covariance $\boldsymbol{\Lambda}_{\boldsymbol{G}_{j+1}}$ such that

$$\int_0^{t_{j+1}} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_{j+1}}, \tau)\,\mathrm{d}\tau = 0 \tag{233}$$
$$\int_0^{t'} \mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_{\boldsymbol{G}_{j+1}}, \tau)\,\mathrm{d}\tau \geq 0 \quad \forall t' > t_{j+1} \quad \forall i \tag{234}$$

where (234) is true specifically for $t' = t_{j+2}$. Finally, from (233) and (232) and the monotonically increasing property of $\mathsf{d}_i(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \boldsymbol{\Lambda}_G, \tau)$ in $[\boldsymbol{\Lambda}_G]_{ii}$ (fourth property of Corollary 3), and the fact that it is independent of all other entries in $\boldsymbol{\Lambda}_G$, we can conclude that $\left[\boldsymbol{\Lambda}_{\boldsymbol{G}_{j+1}}\right]_{ii} \leq \left[\boldsymbol{\Lambda}_{\boldsymbol{G}_j}\right]_{ii}$. This concludes the proof of the induction. ∎

Now, inserting the aforementioned bounds (229) [with the addition of the trivial bound on $I(\boldsymbol{X};\boldsymbol{Y}_1)$, under the input per-antenna constraint (223)] into the single-letter expression in (227), we obtain the following outer bound:

$$\begin{aligned}\mathsf{R}_1 &\leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_1\boldsymbol{P}\mathbf{H}_1^T| - \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_1\boldsymbol{\Lambda}_{\boldsymbol{G}_1}\mathbf{H}_1^T|\\ \mathsf{R}_2 &\leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_2\boldsymbol{\Lambda}_{\boldsymbol{G}_1}\mathbf{H}_2^T| - \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_2\boldsymbol{\Lambda}_{\boldsymbol{G}_2}\mathbf{H}_2^T|\\ &\vdots\\ \mathsf{R}_{M-2} &\leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_{M-2}\boldsymbol{\Lambda}_{\boldsymbol{G}_{M-3}}\mathbf{H}_{M-2}^T| -\\ &\quad \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_{M-2}\boldsymbol{\Lambda}_{\boldsymbol{G}_{M-2}}\mathbf{H}_{M-2}^T|\\ \mathsf{R}_{M-1} &\leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_{M-1}\boldsymbol{\Lambda}_{\boldsymbol{G}_{M-2}}\mathbf{H}_{M-1}^T| -\\ &\quad \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_{M-1}\boldsymbol{\Lambda}_{\boldsymbol{G}_{M-1}}\mathbf{H}_{M-1}^T|\\ \mathsf{R}_M &\leq \frac{1}{2}\log|\mathbf{I} + \mathbf{H}_M\boldsymbol{\Lambda}_{\boldsymbol{G}_{M-1}}\mathbf{H}_M^T|.\end{aligned} \tag{235}$$

where $\boldsymbol{P}$ is a diagonal matrix with $[\boldsymbol{P}]_{ii} = P_i$, and $\boldsymbol{\Lambda}_{\boldsymbol{G}_j}$ are positive-semidefinite diagonal matrices such that $\mathbf{0} \preceq \boldsymbol{\Lambda}_{\boldsymbol{G}_M} \preceq \boldsymbol{\Lambda}_{\boldsymbol{G}_{M-1}} \preceq \ldots \preceq \boldsymbol{\Lambda}_{\boldsymbol{G}_2} \preceq \boldsymbol{\Lambda}_{\boldsymbol{G}_1} \preceq \boldsymbol{P}$. The achievability of this outer bound is well known using superposition coding. ∎

C) Converse Proof of BC Capacity Under Covariance Constraints for M-*Users:* We consider the same setting as in Appendix B, given in (222), but now with an input covariance constraint

$$\mathbf{R}_{\boldsymbol{X}} \preceq \mathbf{S} \tag{236}$$

where $\mathbf{S}$ is some positive-definite matrix.

As in Appendix B, since we have a *degraded* BC, we can use the single-letter expression given explicitly in (227), with auxiliary random variables complying with the Markov chain as detailed in (225). Furthermore, we construct a path as was done in (228). Now, assume distribution $P_{\{\boldsymbol{V}_0\equiv\emptyset, \boldsymbol{V}_1,\ldots,\boldsymbol{V}_{M-1},\boldsymbol{V}_M\equiv\boldsymbol{X}\}}$ on the tuple $(\boldsymbol{V}_0 \equiv \emptyset, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{M-1}, \boldsymbol{V}_M \equiv \boldsymbol{X})$ with covariance matrix $\mathbf{R}_{\boldsymbol{X}}$. We begin by proving the following lemma.

Lemma 22: There exist M Gaussian inputs $\boldsymbol{X}_{G_j}$, with covariance matrices $\mathbf{R}_{\boldsymbol{G}j}$ such that

$$\int_0^{t_j} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right)\mathrm{d}\tau = 0$$
$$\int_0^{t_{j+1}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right)\mathrm{d}\tau \geq 0 \quad \forall j = 1, \ldots, M-1 \tag{237}$$

and such that $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}j} \preceq \mathbf{R}_{\boldsymbol{G}j-1}$, for $j = 2, \ldots, M$ and $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}1} \preceq \boldsymbol{S}$. Furthermore, $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t_j) \succeq \mathbf{0}$, for all $j = 1, \ldots, M$.

Proof: We will prove the above using induction.

The case of $j = 1$: This is identical to the proof given in Section V-E.

For a general j: We assume that the above holds for j and prove for $j+1$. Due to the Markov relation (225), we have that

$$\mathbf{E}_{\boldsymbol{X}|\boldsymbol{V}_{j+1}}(t) = \mathbf{E}_{\boldsymbol{X}|\boldsymbol{V}_{j+1},\boldsymbol{V}_j}(t) \preceq \mathbf{E}_{\boldsymbol{X}|\boldsymbol{V}_j}(t) \quad \forall t \tag{238}$$

from which we can conclude that

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t) = \mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t) \quad \forall t \tag{239}$$

and thus

$$\begin{aligned}\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t) &= \mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t) \\ &\succeq \mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t) \quad \forall t.\end{aligned} \tag{240}$$

Since $\mathbf{B}(t)$ is a diagonal positive-semidefinite matrix for all t, this leads to

$$\begin{aligned}&\mathsf{Tr}\left(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t)\right) \geq \\ &\qquad \mathsf{Tr}\left(\mathbf{B}(t)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t)\right) \quad \forall t.\end{aligned} \tag{241}$$

Now, taking into account the induction assumptions on j, together with (240) and (241), we have

$$\begin{aligned}&\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t_{j+1}) \succeq \mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, t_{j+1}) \succeq \mathbf{0} \\ &\int_0^{t_{j+1}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \geq \\ &\qquad \int_0^{t_{j+1}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_j, \mathbf{R}_{\boldsymbol{G}j}, \tau)\right) \mathrm{d}\tau \geq 0\end{aligned} \tag{242}$$

$$\begin{aligned}
\mathsf{R}_j &\leq \frac{1}{n} I\left(W_j; \bar{\boldsymbol{Y}}_{i_j}^j | \boldsymbol{W}^{j-1}\right) + \delta(n) \\
&= \frac{1}{n}\sum_{l=1}^{n} I\left(W_j; \boldsymbol{Y}_{i_j}^j(l) | \boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1)\right) + \delta(n) && (253)\\
&= \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1)\right) - \mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1)\right)\right) + \delta(n) \\
&= \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right. \\
&\qquad \left. -\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1)\right)\right) + \delta(n) && (254)\\
&\leq \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right. \\
&\qquad \left. -\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{(j+1)j}^{\star}(1,\ldots,l-1), \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1)\right)\right) + \delta(n) && (255)\\
&= \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right. \\
&\qquad \left. -\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{(j+1)j}^{\star}(1,\ldots,l-1)\right)\right) + \delta(n) && (256)\\
&= \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{i_j}^j(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right. \\
&\qquad \left. -\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{(j+1)j}^{\star}(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right) + \delta(n) && (257)\\
&\leq \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j-1}, \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right. \\
&\qquad \left. -\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{W}^{j}, \boldsymbol{Y}_{(j+1)j}^{\star}(1,\ldots,l-1), \boldsymbol{Y}_{j(j-1)}^{\star}(1,\ldots,l-1)\right)\right) + \delta(n) && (258)\\
&= \frac{1}{n}\sum_{l=1}^{n} \left(\mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{V}_{j-1}(l)\right) - \mathsf{h}\left(\boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{V}_{j-1}(l), \boldsymbol{V}_j(l)\right)\right) + \delta(n) && (259)\\
&= \frac{1}{n}\sum_{l=1}^{n} I\left(\boldsymbol{V}_j(l); \boldsymbol{Y}_{i_j}^j(l)|\boldsymbol{V}_{j-1}(l)\right) + \delta(n) && (260)
\end{aligned}$$

which can also be written as

$$\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j}, t_{j+1}) \succeq \mathbf{0}$$
$$I\left(\boldsymbol{X}_{\boldsymbol{G}j}; \boldsymbol{Y}_{\boldsymbol{G}j}(t_{j+1})\right) \geq I\left(\boldsymbol{X}; \boldsymbol{Y}(t_{j+1})|\boldsymbol{V}_{j+1}\right). \quad (243)$$

These are the two conditions required for Lemma 15, with $\boldsymbol{X_G}^{ub} \equiv \boldsymbol{X}_{\boldsymbol{G}j}$. Thus, according to Lemma 15 there exists a Gaussian random vector with covariance $\mathbf{R}_{\boldsymbol{G}j+1}$ such that:

1) $\mathbf{R}_{\boldsymbol{G}j+1} \preceq \mathbf{R}_{\boldsymbol{G}j}$;
2) $I\left(\boldsymbol{X}_{\boldsymbol{G}j+1}; \boldsymbol{Y}_{\boldsymbol{G}j+1}(t_{j+1})\right) = I\left(\boldsymbol{X}; \boldsymbol{Y}(t_{j+1})|\boldsymbol{V}_{j+1}\right)$;
3) $\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, t_{j+1}) \succeq \mathbf{0}$.

Property 2 is equivalent to

$$\int_0^{t_{j+1}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, \tau)\right) \mathrm{d}\tau = 0 \quad (244)$$

and from property 3, Corollary 6, and Theorem 8 we can conclude the following:

$$\begin{aligned}
&\int_0^{t_{j+2}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, \tau)\right) \mathrm{d}\tau \\
&= \int_0^{t_{j+1}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, \tau)\right) \mathrm{d}\tau \\
&\quad + \int_{t_{j+1}}^{t_{j+2}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, \tau)\right) \mathrm{d}\tau \\
&= 0 + \int_{t_{j+1}}^{t_{j+2}} \mathsf{Tr}\left(\mathbf{B}(\tau)\mathbf{Q}(\boldsymbol{X}|\boldsymbol{V}_{j+1}, \mathbf{R}_{\boldsymbol{G}j+1}, \tau)\right) \mathrm{d}\tau \geq 0.
\end{aligned} \quad (245)$$

Together with property 1, this concludes the proof of the induction. ■

Lemma 22 provides us with Gaussian random vectors $\boldsymbol{X}_{\boldsymbol{G}j}$ with covariance matrices $\mathbf{R}_{\boldsymbol{G}j}$ with the following properties:

1) $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}j} \preceq \mathbf{R}_{\boldsymbol{G}j-1}$, for $j = 2, \ldots, M-1$ and $\mathbf{0} \preceq \mathbf{R}_{\boldsymbol{G}1} \preceq \boldsymbol{S}$;
2) $I(\boldsymbol{X}; \boldsymbol{Y}_j|\boldsymbol{V}_j) = \frac{1}{2}\mathsf{log}\left|\mathbf{I} + \mathbf{H}_j\mathbf{R}_{\boldsymbol{G}j}\mathbf{H}_j^{\mathsf{T}}\right|$, for $j = 1, \ldots, M-1$;
3) $I(\boldsymbol{X}; \boldsymbol{Y}_{j+1}|\boldsymbol{V}_j) \leq \frac{1}{2}\mathsf{log}\left|\mathbf{I} + \mathbf{H}_{j+1}\mathbf{R}_{\boldsymbol{G}j}\mathbf{H}_{j+1}^{\mathsf{T}}\right|$, for $j = 1, \ldots, M-1$.

Substituting these results into the single-letter expression (227) and defining

$$\begin{aligned}
\mathbf{R}_{\boldsymbol{G}1} &= \boldsymbol{S} - \mathbf{R}_{\boldsymbol{G}1} \\
\mathbf{R}_{\boldsymbol{G}j} &= \mathbf{R}_{\boldsymbol{G}j-1} - \mathbf{R}_{\boldsymbol{G}j} \quad \forall j = 2, \ldots, M-1 \\
\mathbf{R}_{\boldsymbol{G}M} &= \mathbf{R}_{\boldsymbol{G}M-1}
\end{aligned} \quad (246)$$

provides the following upper bound:

$$\begin{aligned}
\mathsf{R}_M &\leq \frac{1}{2}\mathsf{log}\left|\mathbf{H}_M\mathbf{R}_{\boldsymbol{G}M}\left(\mathbf{H}_M\right)^{\mathsf{T}} + \mathbf{I}\right| \\
\mathsf{R}_j &\leq \frac{1}{2}\mathsf{log}\frac{\left|\mathbf{H}_j\sum_{l=j}^{M}\mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_j\right)^{\mathsf{T}} + \mathbf{I}\right|}{\left|\mathbf{H}_j\sum_{l=j+1}^{M}\mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_j\right)^{\mathsf{T}} + \mathbf{I}\right|} \quad \forall j = 1, \ldots, M-1
\end{aligned} \quad (247)$$

where $\mathbf{R}_{\boldsymbol{G}j}$ are some positive-semidefinite matrices such that $\mathbf{0} \preceq \sum_{l=1}^{M}\mathbf{R}_{\boldsymbol{G}l} = \boldsymbol{S}$. This completes the converse proof.

The aforementioned upper bounds can be attained simultaneously using a joint Gaussian distribution on the tuple

$$(\boldsymbol{V}_0 \equiv \emptyset, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{M-1}, \boldsymbol{V}_M \equiv \boldsymbol{X}) \quad (248)$$

as follows:

$$\boldsymbol{V}_j = \boldsymbol{V}_{j-1} + \boldsymbol{U}_j \quad (249)$$

where $\boldsymbol{U}_j \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{\boldsymbol{G}j}\right)$ for $j = 1, \ldots, M$, independent of each other, and where $\mathbf{R}_{\boldsymbol{G}j}$ are positive-semidefinite matrices such that $\sum_{l=1}^{M}\mathbf{R}_{\boldsymbol{G}l} \preceq \mathbf{S}$. Thus, we attain the upper bounds for $j = 1, \ldots, M-1$ as follows:

$$\begin{aligned}
\mathsf{R}_j &\leq I\left(\boldsymbol{V}_j; \boldsymbol{Y}_j|\boldsymbol{V}_{j-1}\right) \\
&= I\left(\boldsymbol{X}; \boldsymbol{Y}_j|\boldsymbol{V}_{j-1}\right) - I\left(\boldsymbol{X}; \boldsymbol{Y}_j|\boldsymbol{V}_j\right) \\
&= \frac{1}{2}\mathsf{log}\left|\mathbf{H}_j\sum_{l=j}^{M}\mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_j\right)^T + \mathbf{I}\right| \\
&\quad - \frac{1}{2}\mathsf{log}\left|\mathbf{H}_j\sum_{l=j+1}^{M}\mathbf{R}_{\boldsymbol{G}l}\left(\mathbf{H}_j\right)^T + \mathbf{I}\right|.
\end{aligned} \quad (250)$$

For $j = M$, we obtain the following:

$$\begin{aligned}
\mathsf{R}_M &\leq I\left(\boldsymbol{X}; \boldsymbol{Y}_M|\boldsymbol{V}_{M-1}\right) \\
&= \frac{1}{2}\mathsf{log}\left|\mathbf{H}_M\mathbf{R}_{\boldsymbol{G}M}\left(\mathbf{H}_M\right)^T + \mathbf{I}\right|.
\end{aligned} \quad (251)$$

Thus, we have shown that (247) is the capacity region under the input covariance constraint. ■

D) Proof of Lemma 17: The proof of this lemma follows the proof of [31, Lem. 4], which is very similar to the well-known proof for the capacity region of a *degraded* BC in [1]. The proof of the direct part relies on successive decoding at the stronger user and is practically identical to that found in [1]. We will detail the converse proof only.

Let $\bar{\boldsymbol{Y}}_{i_j}^j$ denote a sequence of n channel outputs of the i_jth realization of user j. Let W_j for $j = 1, \ldots, M$ denote the message indices, and $\boldsymbol{W}^j$ denote the vector $(W_1, \ldots, W_j)$. Furthermore, let $\boldsymbol{Y}_{i_j}^j(l)$ be the lth sample of $\bar{\boldsymbol{Y}}_{i_j}^j$ and $\boldsymbol{Y}_{i_j}^j(1, \ldots, l-1)$ be the set of all samples up to $l-1$ (including). We use similar notation for all other random variables. As the capacity region depends only on the marginals $P_{\boldsymbol{Y}_{i_j}^j|\boldsymbol{X}}$, we may assume without loss of generality that indeed the mutual distribution is such that

$$\begin{aligned}
&\boldsymbol{W}^M - \boldsymbol{X} - \boldsymbol{Y}_{i_M}^M - \boldsymbol{Y}_{M(M-1)}^\star - \boldsymbol{Y}_{i_{M-1}}^{M-1} - \\
&\quad - \boldsymbol{Y}_{(M-1)(M-2)}^\star - \ldots - \boldsymbol{Y}_{i_2}^2 - \boldsymbol{Y}_{21}^\star - \boldsymbol{Y}_{i_1}^1
\end{aligned} \quad (252)$$

form a Markov chain for every choice of $i_1, i_2, \ldots, i_M$.

Using Fano's inequality and the fact that W_j are independent messages, an upper bound of R_j for any $j = 1, \ldots, M$ which holds for every $i_j \in \{1, \ldots, K_j\}$ is given in (253)–(260) shown at the bottom of the previous page, where $\delta(n) \to 0$ as $n \to \infty$. The equality in (253) is due to the chain rule of mutual information. The equality in (254) is due to the Markov chain $\boldsymbol{W}^M - \boldsymbol{X} - \boldsymbol{Y}_{i_j}^j - \boldsymbol{Y}_{j(j-1)}^\star$ and the memoryless nature of the channel, as can be seen in the identity given in (261) at the top of the next page. The inequality in (255) follows from the fact that conditioning decreases entropy. Equations (256) and (257)

$$
\begin{aligned}
&P\left\{\boldsymbol{Y}^j_{i_j}(l)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1),\boldsymbol{Y}^{\star}_{j(j-1)}(1,\ldots,l-1)\right\}\\
&=\frac{P\left\{\boldsymbol{Y}^j_{i_j}(l),\boldsymbol{Y}^{\star}_{j(j-1)}(1,\ldots,l-1)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}}{P\left\{\boldsymbol{Y}^{\star}_{j(j-1)}(1,\ldots,l-1)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}}\\
&=\frac{P\left\{\boldsymbol{Y}^j_{i_j}(l)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}P\left\{\boldsymbol{Y}^{\star}_{j(j-1)}(1,\ldots,l-1)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1),\boldsymbol{Y}^j_{i_j}(l)\right\}}{P\left\{\boldsymbol{Y}^{\star}_{j(j-1)}(1,\ldots,l-1)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}}\\
&=\frac{P\left\{\boldsymbol{Y}^j_{i_j}(l)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}P\left\{\boldsymbol{Y}^{\star}_{j(j+1)}(1,\ldots,l-1)|\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}}{P\left\{\boldsymbol{Y}^{\star}_{j(j+1)}(1,\ldots,l-1)|\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}}\\
&=P\left\{\boldsymbol{Y}^j_{i_j}(l)|\boldsymbol{W}^{j-1},\boldsymbol{Y}^j_{i_j}(1,\ldots,l-1)\right\}.
\end{aligned}
\tag{261}
$$

follow, again, from the Markov chain $\boldsymbol{W}^M - \boldsymbol{X} - \boldsymbol{Y}^{\star}_{(j+1)j} - \boldsymbol{Y}^j_{i_j} - \boldsymbol{Y}^{\star}_{j(j-1)}$ and the memoryless nature of the channel. Equation (258) follows from the fact that conditioning decreases entropy. In (259), we used the following definition of auxiliary random variables:

$$\boldsymbol{V}_j(l) = \left(\boldsymbol{W}^j, \boldsymbol{Y}^{\star}_{(j+1)j}(1,\ldots,l-1)\right). \tag{262}$$

Next, we replace the index l with a random variable I which is uniformly distributed over the integers $1,\ldots,n$ and define $\boldsymbol{V}_j = (\boldsymbol{V}_j(I), I), \boldsymbol{X} = \boldsymbol{X}(I), \boldsymbol{Y}^j_{i_j} = \boldsymbol{Y}^j_{i_j}(I)$. As the channel is memoryless, we get

$$\mathsf{R}_j \le I\left(\boldsymbol{V}_j; \boldsymbol{Y}^j_{i_j}|\boldsymbol{V}_{j-1}\right) + \delta(n) \tag{263}$$

for all j and for all i_j. Note that as the channel is memoryless, these auxiliary random variables satisfy the Markov chain defined in (150). Moreover, from this definition one can easily see that $\boldsymbol{V}_0 \equiv \emptyset$ and the largest region will be attained when $\boldsymbol{V}_M \equiv \boldsymbol{X}$. Finally, as the aforementioned inequalities hold for every $j = 1,\ldots,M$ and every $i_j = 1,\ldots,K_j$, we complete the proof by taking n to infinity. ■

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements in Information Theory*, 2nd ed. New York: Wiley-Interscience, 2006.

[2] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[3] Y. Geng and C. Nair, "The capacity region of the two-receiver vector Gaussian broadcast channel with private and common messages," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, Jul. 1–6, 2012, pp. 586–590.

[4] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 279–280, Mar. 1974.

[5] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information theoretic problems," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 9–14, 2006, pp. 1016–1020.

[6] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receive wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2083–2114, Apr. 2011.

[7] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[8] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[9] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inf. Control*, vol. 2, pp. 101–112, Jun. 1959.

[10] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, no. 1, pp. 215–220, Jul. 1970.

[11] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information fo the white Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 17, no. 4, pp. 368–371, Jul. 1971.

[12] S. Shamai (Shitz), "From constrained signaling to netwrok interference alignmnet via an information-estimation prespective," *IEEE Inf. Theory Soc. Newslett.*, vol. 62, no. 7, pp. 6–24, Sep. 2012.

[13] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.

[14] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1302–1318, Mar. 2012.

[15] D. Guo, S. Shamai (Shitz), and S. Verdú, "Additive non-Gaussian noise channels: Mutual information and conditional mean estimation," presented at the IEEE Int. Symp. Inf. Theory, Adelaide, Australia, Sep. 4–9, 2005.

[16] D. P. Palomar and S. Verdú, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 453–470, Feb. 2007.

[17] S. Verdú, "Mismatch estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712–3720, Aug. 2010.

[18] T. Weissman, "The relationship between causal and non-causal mismatched estimation in continuous-time awgn channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4256–4273, Sep. 2010.

[19] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[20] D. Guo, Y. Wu, D. Guo, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[21] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. New York: Wiley, 2007.

[22] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.

[23] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 9–14, 2006.

[24] T. M. Apostol, *Calculus, Multi-Variable Calculus and Linear Algebra, With Applications to Differential Equations and Probability*, 2nd ed. New York: Wiley, 1969, vol. 2.

[25] T. M. Cover, “Comments on broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.
[26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[27] F. Rellich, *Perturbation Theory of Eigenvalue Problems*. New York: Courant Institute of Mathematical Sciences, New York Univ., 1954 [Online]. Available: http://www.archive.org/details/perturbationtheo00rell
[28] P. Lancaster, “On eigenvalues of matrices dependent on a parameter,” *Numer. Math.*, vol. 6, no. 1, pp. 377–387, 1964.
[29] R. Bhatia, *Positive Definite Matrices*. Princeton, NJ: Princeton Univ. Press, 2007.
[30] S. M. Kay, *Fundamentals of Statistical Signal processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall.
[31] H. Weingarten, T. Liu, S. Shamai (Shitz), Y. Steinberg, and P. Viswanath, “The capacity region of the degraded multiplt-input-multiple-output compound broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011–5023, Nov. 2009.
[32] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[33] R. A. Horn and C. R. Johnson, *Topics on Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

Ronit Bustin (S’09) received a B.Sc. degree in electrical engineering and computer science and an M.Sc. degree in electrical engineering in 2004 and 2006, respectively, from Tel-Aviv University, Israel. She is currently pursuing the Ph.D. degree in the Department of Electrical Engineering at the Technion, Israel Institute of Technology, Haifa. Her research interests include multiuser information theory, secrecy constraints, Gaussian MIMO channels, estimation theory, and channel coding.

Ronit Bustin received the Irwin and Joan Jacobs scholarship for excellence in graduate studies and research, in January 2010. She is a recipient of the Adams fellowship from the Israel Academy of Sciences and Humanities, April 2010, and an Andrew and Erna Finci Viterbi graduate fellow in the faculty of electrical engineering at the Technion for the fall semester 2010–2011.

Miquel Payaró (S’03–M’08) was born in Barcelona, Spain, in 1979. He received the degree in Electrical Engineering and the Ph.D. degree from the Universitat Politècnica de Catalunya (UPC), Barcelona, in 2002 and 2007, respectively.

He received a predoctoral grant from the Generalitat de Catalunya and from the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) predoctoral fellowships program. During his Ph.D. studies Miquel held a research appointment at the University of New South Wales, Sydney, Australia. From October 2002 to February 2007, he was involved in several research projects in the field of wireless communications, such as MARQUIS included in the European program EUREKA-MEDEA+ or NEWCOM financed by the Information Society Technologies program of the European Commission (EC).

From February 2007 to December 2008 he held a postdoctoral position with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong.

From January 2009, he has been a Research Associate with the CTTC where, in January 2011, he was appointed the head of the Engineering Unit. In his second involvement with CTTC he has participated in several research contracts with the industry as well as in research projects, such as BuNGee or BeFEMTO from the Seventh Framework Program of the EC.

His primary research interests include information-theoretic and signal processing aspects of MIMO channels, with emphasis on the impact of the channel state information and on robust designs.

Daniel P. Palomar (S’99–M’03–SM’08–F’12) received the Electrical Engineering and Ph.D. degrees 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 is also consulting for the research center CTTC-HK. 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; 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 financial systems and communication systems.

Dr. Palomar is an IEEE Fellow, 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 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.

He serves as an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY, and has been an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on “Convex Optimization for Signal Processing,” the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on “Game Theory in Communication Systems,” and 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).

Shlomo Shamai (Shitz) (S’80–M’82–SM’89–F’94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow and a member of the Israeli Academy of Sciences and Humanities. He is the recipient of the 2011 Claude E. Shannon Award. He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.