

On the Compound Capacity of a Class of MIMO Channels Subject to Normed Uncertainty

Sergey Loyka, *Senior Member, IEEE*, and Charalambos D. Charalambous, *Senior Member, IEEE*

Abstract—The compound capacity of uncertain multiple-input multiple-output channels is considered, when the channel is modeled by a class described by a (known) nominal channel and a constrained-norm (unknown) uncertainty. Within this framework, two types of classes are investigated with additive and multiplicative uncertainties subject to a spectral norm constraint, using the singular value decomposition and related singular value inequalities as the main tools. The compound capacity is a maxmin mutual information, representing the capacity of the class, in which the minimization is done over the class of channels while the maximization is done over the transmit covariance. Closed-form solutions for the compound capacity of the classes are obtained and several properties related to transmit and receive eigenvectors are presented. It is shown that, under certain conditions, the compound capacity of the class is equal to the worst-case channel capacity, thus establishing a saddle-point property. Explicit closed-form solutions are given for the worst-case channel uncertainty and the capacity-achieving transmit covariance matrix: the best transmission strategy achieving the compound capacity is a multiple beamforming on the nominal (known) channel eigenmodes with the beam power distribution via the water filling at a degraded SNR. As the uncertainty increases, fewer eigenmodes are used until only the strongest one remains active so that transmit beamforming is an optimal robust transmission strategy in this large-uncertainty regime, for which explicit conditions are given. Using these results, upper and lower bounds of the compound capacity are constructed for other bounded uncertainties and some generic properties are pointed out. The results are extended to compound multiple-access and broadcast channels. In all considered cases, the price to pay for channel uncertainty is an SNR loss (or, equivalently, the nominal channel degradation) commensurate with the uncertainty set radius measured by the spectral norm and the optimal signaling strategy is the transmission on the degraded nominal channel.

Index Terms—Broadcast channel (BC), channel uncertainty, compound channel, multiple-input multiple-output (MIMO) capacity, multiple-access channel (MAC), optimum transmission, saddle point.

I. INTRODUCTION

SINCE the pioneering work of the authors in [1] and [2], multiple-input multiple-output (MIMO) wireless systems have attracted a significant attention due to a promise of high

spectral efficiency. A lot of research activities have been concentrated on both the information-theoretic limits (e.g., capacity) and practical transmission schemes to approach those limits [3]. Performance of such transmission schemes depends heavily on channel state information (CSI) available at the receiver (Rx)/transmitter (Tx) or both. In addition, channel knowledge, either complete or partial, and its accuracy affect also the channel capacity [11]. In wireless systems, imperfect CSI may be due to time-varying nature of the channel as well as due to limitations of the estimation technique and feedback channel [12]. A number of models have been developed to take into account such effects [11], [12], but most of them were not specifically tailored for MIMO systems. A concise review of recent results on the impact of imperfect CSI on MIMO system performance can be found in [15]. The models of channel with imperfect CSI can be classified into statistical (when the true channel is considered to be random with given mean and covariance) and deterministic (when the channel is deterministic (fixed), but only known to belong to a certain class, i.e., compound channel [11]) [15]. Deterministic channel uncertainty models have been used to evaluate the performance of zero-forcing precoding and detection techniques [13], of orthogonal space-time block codes [14], and to evaluate the compound capacity of uncertain rank-one (Ricean) MIMO channels [15].

In this paper, we introduce uncertainty models for the channel matrix based on an induced (spectral) norm constraint, since for practical purposes the spectral norm of the channel matrix has the interpretation of the maximum input/output transfer gain. Following this framework, we develop generic (e.g., any rank) multiplicative and additive channel uncertainty models with the spectral norm constraint specifically tailored for MIMO capacity analysis. In our approach, we follow the ideas of compound channel capacity pioneered in [4]–[6]; a comprehensive review of more recent research activities can be found in [11] and a code construction achieving the compound capacity can be found in [20]. We consider an MIMO channel matrix \mathbf{H} as consisting of two parts: the nominal channel \mathbf{H}_0 , which is known at both the Tx and Rx, and the uncertainty (perturbation) $\Delta\mathbf{H}$, which is not known, so that $\mathbf{H} = \mathbf{H}_0 + \Delta\mathbf{H} \in S_H$, where S_H is a known (bounded) class of channels; both \mathbf{H}_0 and $\Delta\mathbf{H}$ are assumed to be fixed during the transmission interval, while $\Delta\mathbf{H}$ is unknown at both the Tx and Rx ends. Unlike previous work, the size of $\Delta\mathbf{H}$ is measured by the spectral norm, which is an induced norm. This is fundamentally different from the trace norm considered in [15], which is equivalent to a Frobenius norm that is not an induced norm and, thus, does not possess a number of important properties of the latter (see Section II-A for details). Furthermore, while the analysis in [15] was limited to a rank-one nominal channel, we

Manuscript received June 24, 2010; revised July 06, 2011; accepted August 29, 2011. Date of current version March 13, 2012. The material in this paper was presented in part at the 2008 IEEE International Symposium on Information Theory.

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: sloyka@site.uottawa.ca).

C. D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: chadcha@ucy.ac.cy).

Communicated by G. Taricco, Associate Editor for Communications.

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.2011.2173727

consider arbitrary-rank nominal channels. Following [4]–[6], [11], we consider two types of problems: 1) the capacity of the worst-case channel in the class and 2) the capacity of the class of channels also known as the compound channel capacity [11] that is achievable by a single code on any channel in the class. The former is formulated as a minimax optimization problem while the later is formulated as a maxmin optimization. Via explicit evaluation of these two capacities, we demonstrate that, for both the multiplicative and additive uncertainty models under the spectral norm constraint, the saddle point property holds: the capacity of the class equals to the worst-case channel capacity. Additionally, we give explicit expressions for the optimum (capacity-achieving) covariance of the Tx signal and the worst-case uncertainty of the channel, both of which depend only on the nominal channel and the radius of the uncertainty set. In all the cases considered, the best transmit strategy is multiple beamforming on the nominal channel (one beam per each active eigenmode of the nominal channel) plus water-filling power allocation among the beams at the degraded SNR level (reduced according to the uncertainty set radius), i.e., signaling on the degraded nominal channel, and the worst channel perturbation is “antiparallel” to the nominal channel (i.e., each singular vector of the worst channel perturbation is the corresponding singular vector of the nominal channel times a negative scalar). The effect of uncertainty with bounded spectral norm is shown to be equivalent to an SNR loss (in terms of the total SNR for the multiplicative uncertainty and the per-eigenmode SNR for the additive uncertainty). Based on these results, we propose an adaptive transmission strategy that takes into account the “size” of channel variations. As the uncertainty increases, fewer eigenmodes are used until only one remains active. In this large-uncertainty regime, for which explicit conditions are given, the optimal robust transmission strategy is beamforming on the strongest eigenmode of the nominal channel.

Using the fact that any two norms are equivalent, we construct upper and lower bounds of the compound capacity under other normed or bounded channel uncertainties and point out some generic properties. In Sections VI and VII, single-user results are extended to multiple-access channel (MAC) and broadcast channel (BC).

A similar uncertainty model has been also considered in [16] and [17], where, however, a solution was obtained when the uncertainty was limited to the singular values of the channel only (no uncertainty in singular vectors, see e.g., [17, Th. 3.3]), which is unlikely from the physical perspective: random perturbations of the channel due to e.g., mobility or measurement error are likely to affect both the singular values and the singular vectors. Our results, here, are more general as they do not limit uncertainty to the singular values only. In particular, it follows from our Theorems 3 and 4 that [17, Th. 3.3] also applies to this generic setting.

In a related line of work, Vorobyov *et al.* [30] and Lorenz and Boyd [31] consider an uncertain single-input multiple-output channel and obtain, based on worst-case SNR optimization, a robust receive beamformer, which turns out to be equivalent to the classical Capon beamformer with diagonal loading [32]. An extension of this work to an uncertain MIMO channel subject

to the trace and spectral norm constraints are robust transmit/receive beamformers obtained in [33] and [35], which also maximize the worst-case received SNR. Note, however, that maximizing the total received SNR in an MIMO channel is not equivalent in general to maximizing the mutual information, and hence, those designs are not optimal from the information-theoretic perspective, i.e., when the transmission rate is of concern. In this paper, we give explicit conditions (of the large uncertainty regime) when such robust beamformers are information-theoretic optimal. Generic conditions for the optimality of beamforming in terms of the ergodic capacity in an MIMO channel with limited/imperfect feedback have been obtained in [34].

Finally, we would like to point out that the problem we are studying here is not a convex or quasiconvex one, so that standard tools like the Von Neumann minimax theorem [18] or Lagrange multiplier technique with associated Karush–Kuhn–Tucker (KKT) conditions [19] cannot be used to find the optimum solution and a new technique has to be developed. Our approach is based on the singular value decomposition (SVD) and associated singular value inequalities for products and sums of two matrices.

II. UNCERTAIN CHANNEL MODEL

Let us consider the following baseband discrete-time MIMO channel model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \xi \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_m]^T \in \mathcal{C}^{m,1}$ and $\mathbf{y} = [y_1, y_2, \dots, y_n]^T \in \mathcal{C}^{n,1}$ are the vectors representing the Tx and Rx symbols, respectively, “T” denotes transposition, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m] \in \mathcal{C}^{n,m}$ is the $n \times m$ matrix of the complex channel gains between each Tx and each Rx antenna, where \mathbf{h}_i denotes i th column of \mathbf{H} , n and m are the numbers of Rx and Tx antennas, respectively; without loss of generality, we further assume $n \geq m$; ξ is the vector of circularly symmetric additive white Gaussian noise (AWGN), which is independent and identically distributed (i.i.d.) in each receiver¹. The channel is assumed to be quasistatic (i.e., constant for a sufficiently long period of time so that the infinite horizon information theory assumption holds) and frequency flat, with partial CSI at the Rx and Tx ends, as described in the following.

A. Multiplicative Uncertainty

We consider first the multiplicative channel uncertainty model

$$\mathbf{H} = (\mathbf{I} + \mathbf{E})\mathbf{H}_0 \quad (2)$$

where \mathbf{H}_0 is the nominal channel (without uncertainty) known at the Tx end and, possibly, at the Rx end (the Rx CSI is known to have no effect on the capacity of quasi-static channels [10], since the receiver can always learn the channel via a training sequence), and $\mathbf{E} \in \mathcal{C}^{m,n}$ is the multiplicative uncertainty. We

¹the case of unequal noise power per Rx can also be considered within the present framework.

assume \mathbf{E} , and hence \mathbf{H} , belong to a limited uncertainty set (i.e., limited measurement error²)

$$\mathbf{H} \in S_H = \{\mathbf{H} : \|\mathbf{E}\|_2 \leq \varepsilon\}$$

$$\|\mathbf{E}\|_2 = \max_{|\mathbf{x}| \neq 0} \frac{|\mathbf{E}\mathbf{x}|}{|\mathbf{x}|} = \sigma_1(\mathbf{E}) \leq \varepsilon < 1 \quad (3)$$

where S_H is the multiplicative uncertainty set, $\sigma_i(\mathbf{E})$ is the i th singular value of $\mathbf{E} = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^+$, \mathbf{u}_i is the i th left singular vector, \mathbf{v}_i is the i th right singular vector, the singular values are ordered as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$, and $|\mathbf{x}|^2 = \mathbf{x}^+ \mathbf{x} = \sum_i |x_i|^2$ is the vector length squared, where $+$ denotes Hermitian conjugation. Using (2) and (3), the norm of the channel uncertainty $\Delta\mathbf{H} = \mathbf{H} - \mathbf{H}_0$ can be bounded as $\|\Delta\mathbf{H}\|_2 \leq \varepsilon \|\mathbf{H}_0\|_2$. A multiplicative channel uncertainty may be due to a calibration inaccuracy or the dynamic nature of the channel (when the scattering environment changes after the channel estimation has been done).

The use of the spectral norm as a measure of uncertainty has a number of advantages.

- 1) From the input (\mathbf{x})–output ($\mathbf{E}\mathbf{x}$) point of view, \mathbf{v}_1 is the highest gain input (transmitting) direction while \mathbf{u}_1 is the highest gain observing (receiving) direction, and the highest gain is $\sigma_1(\mathbf{E})$.
- 2) It is the only unitary-invariant³ vector-induced⁴ norm [7]. Unitary invariance represents the fact that errors in measurements are statistically equal in all directions, i.e., no preferred direction or bias.
- 3) It lower bounds any unitary-invariant matrix norm (including Frobenius) (see [8, Corollary 5.6.35]) and, thus, gives the largest uncertainty set S_H for any unitary-invariant matrix norm. It is also an indicator of the strongest eigenmode of the uncertainty.
- 4) For a unit energy Rx signal coming from the nominal channel, $|\mathbf{H}_0 \mathbf{x}|^2 = 1$, the spectral norm of \mathbf{E} limits the energy in the uncertain part of the Rx signal, $|\mathbf{E} \mathbf{H}_0 \mathbf{x}|^2 \leq \sigma_1^2(\mathbf{E})$, so that the power ratio of uncertain and certain portions of the Rx signal is upper bounded by $\sigma_1^2(\mathbf{E})$.
- 5) If the uncertain part of the Rx signal is modeled as AWGN, then the corresponding degradation in SNR can be easily evaluated and the new degraded SNR is not less than $\gamma_0 / (1 + \gamma_0 \sigma_1^2(\mathbf{E}))$, where γ_0 is the SNR in the nominal channel. In this model, the effect of uncertainty can be neglected if $\sigma_1(\mathbf{E}) \ll 1/\sqrt{\gamma_0}$, which immediately gives a rough idea as to how good the channel estimation should be: Under the model in (2) and (3), $1 - \sigma_1(\mathbf{E}) \leq \sigma_i(\mathbf{H})/\sigma_i(\mathbf{H}_0) \leq 1 + \sigma_1(\mathbf{E})$, where $\sigma_i(\mathbf{H})$ is the i th singular value of \mathbf{H} . Combining the last two

inequalities, $|1 - \sigma_i(\mathbf{H})/\sigma_i(\mathbf{H}_0)| \ll 1/\sqrt{\gamma_0}$, i.e., the normalized uncertainty in the channel singular values should be much less than $1/\sqrt{\text{SNR}}$ for its effect to be negligible.

B. Additive Uncertainty

In this model, the nominal channel \mathbf{H}_0 experiences an additive perturbation $\Delta\mathbf{H}$

$$\mathbf{H} = \mathbf{H}_0 + \Delta\mathbf{H} \quad (4)$$

where we also assume that $\Delta\mathbf{H}$, and hence \mathbf{H} , belong to a limited uncertainty set

$$\mathbf{H} \in S_{\mathbf{H}} = \{\mathbf{H} : \|\mathbf{H} - \mathbf{H}_0\|_2 = \sigma_1(\Delta\mathbf{H}) \leq \varepsilon\} \quad (5)$$

where $S_{\mathbf{H}}$ is the additive uncertainty set. Similarly to multiplicative uncertainty, the spectral norm bound in (5) can be interpreted in terms of the uncertain signal power, the corresponding degradation in SNR and the condition for its negligible effect can be evaluated. An additive uncertainty mimics the additive noise and may be caused by its presence (i.e., limited SNR available for the channel estimation).

We note that the additive and multiplicative uncertainty models are related, albeit in a nonsymmetric way: for any multiplicative uncertainty \mathbf{E} , there exists an equivalent additive uncertainty $\Delta\mathbf{H} = \mathbf{E}\mathbf{H}_0$; the converse is not always true: for given additive uncertainty $\Delta\mathbf{H}$, there exists an equivalent multiplicative uncertainty \mathbf{E} if and only if the system of linear equations $\Delta\mathbf{H} = \mathbf{E}\mathbf{H}_0$ has a solution, i.e., if the rows of the uncertainty $\Delta\mathbf{H}$ are in the row space of the nominal channel \mathbf{H}_0 .

III. CAPACITY OF MIMO CHANNELS

For fixed channel \mathbf{H} and given covariance of the Tx vector signal $\mathbf{R} = \overline{\mathbf{x}\mathbf{x}^+}$, where $\overline{\cdot}$ denotes expectation of \mathbf{x} , the mutual information between \mathbf{x} and \mathbf{y} when \mathbf{x} is Gaussian (i.e., capacity-achieving) is given by the celebrated Foschini–Telatar formula

$$I = \log |\mathbf{I} + \gamma \mathbf{H} \mathbf{R} \mathbf{H}^+| = \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2(\mathbf{H} \mathbf{R}^{1/2}) \right)$$

$$= I(\mathbf{H}, \mathbf{R}) \quad (6)$$

where γ is the SNR per antenna, $\sigma_i(\mathbf{H} \mathbf{R}^{1/2})$ are singular values of $\mathbf{H} \mathbf{R}^{1/2}$, and, due to the total Tx power constraint, $\text{tr}(\mathbf{R}) \leq m$, where the equality provides the maximum mutual information. If no CSI is available at the Tx end, the popular choice is $\mathbf{R} = \mathbf{I}$. When CSI is available at the Tx end, the capacity can be found as

$$C(\mathbf{H}) = \max_{\text{tr}(\mathbf{R}) \leq m} I(\mathbf{H}, \mathbf{R}) \quad (7)$$

The maximum in (7) has a well-known water-filling solution [2]

$$C(\mathbf{H}) = \sum_i \max_{\lambda_i(\mathbf{R}) \leq m} \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2(\mathbf{H}) \lambda_i(\mathbf{R}) \right)$$

$$= \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2(\mathbf{H}) \lambda_i(\mathbf{R}_b) \right)$$

$$= \sum_{i+} \log \left(\mu \gamma \sigma_i^2(\mathbf{H}) \right) \quad (8)$$

²If the limit $\varepsilon < 1$ is not set, the worst-case capacity becomes zero.

³A norm $\|\cdot\|$ is unitary invariant if $\|\mathbf{U} \mathbf{A} \mathbf{V}\| = \|\mathbf{A}\|$ for any matrix \mathbf{A} and any unitary matrices \mathbf{U}, \mathbf{V} . Any unitary-invariant matrix norm is a function of its singular values only and is independent of the singular vectors [7].

⁴A matrix norm $\|\cdot\|$ is induced by vector norm $N(\cdot)$ if $\|\mathbf{A}\| = \max_{N(\mathbf{x})=1} N(\mathbf{A}\mathbf{x})$ for any matrix \mathbf{A} [8]. For example, the spectral norm $\|\cdot\|_2$ is induced by Euclidean vector norm $|\cdot|$ (the vector length) according to (3).

where

$$\lambda_i(\mathbf{R}_b) = \left[\mu - \frac{1}{\gamma \sigma_i^2(\mathbf{H})} \right]_+ \quad (9)$$

is the i th eigenvalue of the best (maximizing) covariance matrix \mathbf{R}_b , $[x]_+ = x$ if $x > 0$ and 0 otherwise, $i_+ = \{i : \lambda_i(\mathbf{R}) > 0\}$ is the index of active modes, and the constant μ is found from the total power constraint, $\sum_{i_+} \lambda_i(\mathbf{R}_b) = m$.

IV. COMPOUND CAPACITY OF A CLASS OF MIMO CHANNELS WITH UNCERTAINTY

We consider in the following the capacity of channels with uncertainty models in (2)–(5). Following the framework developed in [4]–[6], we distinguish between the capacity $C_{\min \max}$ of the worst-case channel and the capacity $C_{\max \min}$ of the class of channels (compound capacity) in S_H

$$C_{\min \max} = \min_{\mathbf{H} \in S_H} \max_{\text{tr}(\mathbf{R}) \leq m} I(\mathbf{H}, \mathbf{R}) \quad (10)$$

$$C_{\max \min} = \max_{\text{tr}(\mathbf{R}) \leq m} \min_{\mathbf{H} \in S_H} I(\mathbf{H}, \mathbf{R}) \quad (11)$$

Note that the compound capacity of the class $C_{\max \min}$ is achievable using a single (universal) code on any channel in the class [4], [6] (i.e., there exists a code such that the decoding error tends to zero uniformly over the class of channels) so that there is no need to know what particular channel is currently in effect. On the other hand, the worst-case capacity is an appropriate characteristic of an adaptive system, which selects an individual code for each channel realization (so that accurate channel estimation is essential for the latter). In general, the following inequality holds [4]–[6]:

$$C_{\min \max} \geq C_{\max \min} \quad (12)$$

i.e., the compound capacity of a class is never higher than the worst-case channel capacity. In the following, we obtain stronger results for the uncertainty models in (2) and (4).

A. Multiplicative Uncertainty

The capacities in (10) and (11) can be characterized in a simple way, using the following Lemma.

Lemma 1: For given \mathbf{R} , the worst-case mutual information in (6) for the class of channels in (2) and (3) is

$$I_{\min}(\mathbf{H}_0, \mathbf{R}) = \min_{\mathbf{H} \in S_H} I(\mathbf{H}, \mathbf{R}) = \sum_{i=1}^m \log \left(1 + \gamma(1-\varepsilon)^2 \sigma_i^2(\mathbf{H}_0 \mathbf{R}^{1/2}) \right) \quad (13)$$

i.e., the multiplicative uncertainty in (2) and (3) results in relative SNR loss $(1-\varepsilon)^2$ compared to the nominal channel case, for any covariance \mathbf{R} .

Proof: Using the following singular value inequalities [7]

$$\sigma_n(\mathbf{A})\sigma_i(\mathbf{B}) \leq \sigma_i(\mathbf{AB}), \quad \sigma_i(\mathbf{B}) - \sigma_1(\mathbf{C}) \leq \sigma_i(\mathbf{B} + \mathbf{C}), \quad (14)$$

where $\mathbf{A} \in C^{n,n}$, $\mathbf{B}, \mathbf{C} \in C^{m,m}$, $i = 1, \dots, \min(m, n)$, together with (2) and (3) results in the lower bound on the mutual

information $\forall \mathbf{H} \in S_H$

$$I(\mathbf{H}, \mathbf{R}) \geq \sum_{i=1}^m \log \left(1 + \gamma(1-\varepsilon)^2 \sigma_i^2(\mathbf{H}_0 \mathbf{R}^{1/2}) \right). \quad (15)$$

The worst-case channel perturbation \mathbf{E}_w , which achieves the lower bound in (15), is of the form

$$\mathbf{E}_w = -\mathbf{U}'_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{U}'_0{}^+ \quad (16)$$

where $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\varepsilon \dots \varepsilon)$, and the columns of \mathbf{U}'_0 are the left singular vectors of $\mathbf{H}'_0 = \mathbf{H}_0 \mathbf{R}^{1/2}$, which can be found from its SVD, $\mathbf{H}'_0 = \mathbf{U}'_0 \boldsymbol{\Sigma}'_0 \mathbf{V}'_0{}^+$. Hence, (13) follows since the inequality in (15) holds with equality under (16). ■

Theorem 1: The compound capacity of the class of channels in (3) is

$$\begin{aligned} C_{\max \min} &= \sum_{i=1}^m \log \left(1 + \gamma(1-\varepsilon)^2 \sigma_i^2(\mathbf{H}_0) \lambda_i(\mathbf{R}_b) \right) \\ &= \sum_{i_+} \log \left(\mu \gamma(1-\varepsilon)^2 \sigma_i^2(\mathbf{H}_0) \right) \\ &= C((1-\varepsilon)\mathbf{H}_0) \end{aligned} \quad (17)$$

and

$$\mathbf{R}_b = \mathbf{R}_b(\mathbf{H}_0) = \mathbf{V}_0 \mathbf{A}_b \mathbf{V}_0{}^+, \quad \mathbf{A}_b = \text{diag}(\lambda_1(\mathbf{R}_b) \dots \lambda_m(\mathbf{R}_b)) \quad (18)$$

is the best (i.e., capacity-achieving) covariance matrix,

$$\lambda_i(\mathbf{R}_b) = \left[\mu - \frac{1}{\gamma(1-\varepsilon)^2 \sigma_i^2(\mathbf{H}_0)} \right]_+ \quad (19)$$

are the eigenvalues of \mathbf{R}_b (i.e., the optimum power allocation to the eigenmodes via water filling); $i_+ = \{i : \lambda_i(\mathbf{R}_b) > 0\}$ is the index set of active modes; the constant μ is found from the total power constraint $\sum_{i_+} \lambda_i(\mathbf{R}_b) = m$; the columns of \mathbf{V}_0 are right singular vectors of the nominal channel \mathbf{H}_0 found from its SVD [7], [9]

$$\mathbf{H}_0 = \mathbf{U}_0 \boldsymbol{\Sigma}_0 \mathbf{V}_0{}^+ \quad (20)$$

where the columns of \mathbf{U}_0 are the left singular vectors of \mathbf{H}_0 , and the diagonal entries of $\boldsymbol{\Sigma}_0$ are its singular values. Under the best covariance in (18), the worst-case channel perturbation (achieving the minimum in $C_{\max \min}$) takes the form

$$\mathbf{E}_w = -\mathbf{U}_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{U}_0{}^+ \quad (21)$$

where $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\varepsilon \dots \varepsilon)$, so that the worst-case channel is $(1-\varepsilon)\mathbf{H}_0$, i.e., the degraded nominal channel.

Proof: Combining the minimum mutual information in (13) with the water-filling solution in (8) results in (17). It is straightforward to see that (18) and (21) achieve the maxmin capacity in (17). ■

Similarly to (13), the effect of channel uncertainty in this maxmin problem is the SNR loss of $(1-\varepsilon)^2$. For example, a 3 dB loss occurs when $\varepsilon \approx 1 - 1/\sqrt{2} \approx 0.3$, i.e., 30% inaccuracy in the channel knowledge, and the condition for a negligible effect of uncertainty is $\varepsilon \ll 0.3$. Note that the best Tx

strategy in Theorem 1 is the transmission on the degraded nominal channel $(1 - \varepsilon)\mathbf{H}_0$, i.e., multiple beamforming on the right singular vectors of the nominal channel with the power distribution among the beams given by the water-filling algorithm applied to $(1 - \varepsilon)\mathbf{H}_0$ or, equivalently, to the nominal channel \mathbf{H}_0 at the degraded SNR $(1 - \varepsilon)^2\gamma$. Under this strategy, the worst channel perturbation in (21) is antiparallel of the nominal channel, i.e., its singular vectors are antiparallel to those of the nominal channel and its singular values are of the maximum allowed amplitude. The intuition behind this result is that the jammer (uncertainty) tries to reduce the channel singular values as much as possible (since the capacity is a monotonically increasing function of the singular values) and (17) is the best it can achieve (using (21)), under the best Tx strategy.

It follows from (19) that increasing the uncertainty radius ε results in fewer modes being used for the transmission until only the strongest mode remains active, which occurs when

$$\varepsilon \geq 1 - \sqrt{\frac{\sigma_2^{-2}(\mathbf{H}_0) - \sigma_1^{-2}(\mathbf{H}_0)}{m\gamma}} \quad (22)$$

and the transmit beamforming becomes the optimal strategy in this large-uncertainty regime. One further concludes that transmit beamforming is always the optimal robust strategy for rank-one nominal channel.

If the transmitter is not allowed to use nonuniform power allocation among the beams⁵, i.e., if $\lambda_1(\mathbf{R}) = \dots = \lambda_m(\mathbf{R}) = 1$, then $\mathbf{R}_b = \mathbf{V}_0\mathbf{V}_0^+ = \mathbf{I}$ and

$$C_{\max \min} = I_{\min}(\mathbf{H}_0, I) \leq \max_{tr(\mathbf{R}) \leq m} I_{\min}(\mathbf{H}_0, \mathbf{R}).$$

Thus, the nonuniform power allocation is essential in achieving the capacity of the class of channels.

Theorem 2: For the class of channels in (2) and (3), the capacity of the class equals to the worst-case channel capacity (the minimax and maxmin solutions are the same)

$$C_{\min \max} = C_{\max \min} = C((1 - \varepsilon)\mathbf{H}_0) \quad (23)$$

so that there is a saddle point in $I(\mathbf{H}, \mathbf{R})$ for any admissible \mathbf{H} and \mathbf{R} , $\forall \mathbf{H} \in S_H, \forall \mathbf{R} : tr \mathbf{R} \leq m$

$$I(\mathbf{H}_w, \mathbf{R}) \leq I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}, \mathbf{R}_b) \quad (24)$$

where $\mathbf{H}_w = (\mathbf{I} + \mathbf{E}_w)\mathbf{H}_0$ is the worst channel in the class; the worst perturbation \mathbf{E}_w is as in (21) and the best covariance \mathbf{R}_b is as in (18).

Proof:

$$\begin{aligned} C_{\min \max} &= \min_{\mathbf{H} \in S_H} \max_{tr(\mathbf{R}) \leq m} I(\mathbf{H}, \mathbf{R}) \\ &\stackrel{(a)}{=} \min_{\mathbf{H} \in S_H} \sum_{i=1}^m \log(1 + \gamma \sigma_i^2(\mathbf{H}) \lambda_i(\mathbf{R}_b)) \\ &\stackrel{(b)}{\geq} \sum_{i=1}^m \log(1 + \gamma(1 - \varepsilon)^2 \sigma_i^2(\mathbf{H}_0) \lambda_i(\mathbf{R}_b)) \\ &= C_{\max \min} \end{aligned} \quad (25)$$

⁵for example, due to the practical constrain of using the same fixed-power amplifiers.

where (a) follows from the water-filling solution in (8), and (b) follows from the singular-value inequalities in (14). The equality is achieved by $\mathbf{E}_w = -\mathbf{U}_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{U}_0^+$ as in (21), and thus, (23) follows. The saddle-point property in (24) follows from (23) via [18, Corollary 9.16] (in our context, this corollary states that (24) follows from (23) and vice-versa). ■

It follows from Theorem 2 that there exists a single code that achieves the capacity of the worst-case channel on any channel in the uncertainty class. Theorem 2 says in fact that in the matrix game between the transmitter (\mathbf{R}) and the jammer (\mathbf{E}), there is an optimum strategy for both players and each get penalized if it deviates from this strategy, provided that the other follows it. It also follows that the knowledge of \mathbf{E} does not help the transmitter to increase the capacity provided the jammer follows the best (i.e., capacity-minimizing) strategy, i.e., the optimization of \mathbf{R} for the worst (\mathbf{H}_w) and true (\mathbf{H}) channels give the same capacity under the best jamming strategy (see (15); this is reminiscent of the no-uncertainty case where the Rx CSI does not increase the capacity of quasi-static channels [10]).

As a side remark, we note that a direct application of Von Neumann minimax theorem (see e.g., [18]) to prove (23) is not possible here since $I(\mathbf{H}, \mathbf{R})$ is neither convex nor quasiconvex in \mathbf{H} . Using the Lagrange multiplier techniques to solve this problem (e.g., as in [17]) encounters the same difficulty: since the problem is not convex, the KKT conditions are not sufficient for optimality [19].

B. Additive Uncertainty

The results of the previous section need some modifications to be adapted to the additive uncertainty model in (4) and (5), which is done in the following.

Theorem 3: The capacity of the worst-case channel in the class in (4) and (5) is

$$\begin{aligned} C_{\min \max} &= \min_{\sigma_1(\Delta \mathbf{H}) \leq \varepsilon} \max_{tr(\mathbf{R}) \leq m} I(\mathbf{H}_0 + \Delta \mathbf{H}, \mathbf{R}) \\ &= \sum_{i_+} \log(1 + \gamma(\sigma_i(\mathbf{H}_0) - \varepsilon)^2 \lambda_i(\mathbf{R}_b)) \\ &= C(\mathbf{H}_w) \end{aligned} \quad (26)$$

where the summation is over $i_+ = \{i : \sigma_i(\mathbf{H}_0) > \varepsilon\}$, and

$$\lambda_i(\mathbf{R}_b) = \left[\mu - \frac{1}{\gamma(\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2} \right]_+ \quad (27)$$

are the eigenvalues of the capacity-achieving covariance matrix \mathbf{R}_b given by

$$\mathbf{R}_b = \mathbf{V}_0 \mathbf{A}_b \mathbf{V}_0^+, \quad \mathbf{A}_b = \text{diag}(\lambda_1(\mathbf{R}_b), \dots, \lambda_m(\mathbf{R}_b)) \quad (28)$$

i.e., its eigenvectors are the right singular vectors of the nominal channel \mathbf{H}_0 , and the “water level” μ is found from the total power constraint $\sum_i \lambda_i(\mathbf{R}_b) = m$. The worst-case channel is

$$\mathbf{H}_w = \mathbf{H}_0 + \Delta \mathbf{H}_w = \mathbf{U}_0 \boldsymbol{\Sigma}_w \mathbf{V}_0^+, \quad \boldsymbol{\Sigma}_w = \text{diag}\{(\sigma_i(\mathbf{H}_0) - \varepsilon)_+\} \quad (29)$$

where $\Delta \mathbf{H}_w = -\mathbf{U}_0 \boldsymbol{\Sigma}_{\varepsilon+} \mathbf{V}_0^+$ is the worst-case perturbation and $\boldsymbol{\Sigma}_{\varepsilon+} = \text{diag}\{\min(\sigma_i(\mathbf{H}_0), \varepsilon)\}$ is the diagonal matrix of its singular values. Equation (26) implies that for any admissible

$$\mathbf{R}, \forall \mathbf{R} : \text{tr} \mathbf{R} \leq m$$

$$I(\mathbf{H}_w, \mathbf{R}) \leq I(\mathbf{H}_w, \mathbf{R}_b). \tag{30}$$

Proof: The max part is as in (8); applying the singular value inequalities in (14) to it results in the lower bound on $C(\mathbf{H})$ equal to the right-hand side of (26), which is achieved by the worst-case channel in (29) with the best covariance matrix as in (28) and its eigenvalues as in (27). Since $I(\mathbf{H}_w, \mathbf{R}_b) = C_{\min \max}$ and \mathbf{R}_b achieves the maximum in (26), (30) follows. ■

Note that the worst channel perturbation in (29) is “antiparallel” to the nominal channel, so that its singular values are reduced most. The best covariance \mathbf{R}_b has the same structure as in (18), i.e., multibeamforming on the right singular vectors of the nominal channel with water filling in (27) applied to the gains $(\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2$ of the degraded nominal channel.

Theorem 4: The capacity of the class of channels in (4) and (5) is upper bounded by the capacity of the worst-case channel

$$C_{\max \min} \leq C_{\min \max} \tag{31}$$

The equality is achieved

$$C_{\min \max} = C_{\max \min} \tag{32}$$

so that there is a saddle point in $I(\mathbf{H}, \mathbf{R})$ for any admissible \mathbf{H} and $\mathbf{R}, \forall \mathbf{H} \in S_H, \forall \mathbf{R} : \text{tr} \mathbf{R} \leq m$

$$I(\mathbf{H}_w, \mathbf{R}) \leq I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}, \mathbf{R}_b) \tag{33}$$

when at least one of the following conditions hold.

- 4.1. The SNR is low, $m\gamma(\sigma_1(\mathbf{H}_0) + \varepsilon)^2 \ll 1$.
- 4.2. The SNR is high

$$\gamma \gg \frac{2r}{m(\sigma_r(\mathbf{H}_0) - \varepsilon)^2} - \frac{1}{m} \sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^{-2}$$

where r is the rank of the worst-case channel in (29), $r = \text{rank}(\mathbf{H}_w)$, also equal to the number of active eigenmodes at high SNR, $\text{rank}(\mathbf{R}_b) = r$, so that $\sigma_i(\mathbf{H}_0) > \varepsilon, \lambda_i(\mathbf{R}_b) > 0, i = 1, \dots, r$ (weaker singular values do not contribute anything to the compound capacity).

- 4.3. The nominal channel has identical nonzero singular values, $\Sigma_0 = \text{diag}(\sigma_0, \sigma_0, \dots, \sigma_0, 0, \dots, 0)$.
- 4.4. The nominal channel rank is not greater than 2, $\text{rank}(\mathbf{H}_0) \leq 2$. In particular, this holds when the number of antennas at least in one end does not exceed 2, $m \leq 2$, or/and $n \leq 2$.

In all cases, the worst channel and the best Tx covariance (achieving min and max, respectively, in $C_{\max \min}$) are as in Theorem 3.

Proof: See the Appendix. ■

Thus, there exists a single code that achieves the worst-case channel capacity for any channel in the uncertainty class under conditions of Theorem 4. Neither the transmitter (\mathbf{R}) nor the jammer ($\Delta \mathbf{H}$) can deviate from the optimal strategy without incurring a loss. This result parallels one for the multiplicative uncertainty case. It can also be seen that the effect of uncertainty is to reduce each eigenmode gain from

$\sigma_i(\mathbf{H}_0)^2$ to $(\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2$, i.e., an SNR loss on each eigenmode individually, so that the degraded nominal channel is $\mathbf{U}_0(\Sigma_0 - \text{diag}(\varepsilon))_+ \mathbf{V}_0^+$. Unlike the case of multiplicative uncertainty model, the SNR loss effect cannot be expressed solely in terms of the aggregate SNR γ . Only sufficiently strong eigenmodes are used by the optimal transmission strategy

$$\lambda_i(\mathbf{R}_b) > 0 \text{ only if } \sigma_i(\mathbf{H}_0) > \varepsilon \tag{34}$$

and weaker ones do not contribute anything to the compound capacity, so that the effect of uncertainty is to make certain eigenmodes unusable (if $\sigma_i(\mathbf{H}_0) \leq \varepsilon$). At high SNR, “only if” in (34) becomes “if and only if.” As a side remark, we note that when $\gamma \rightarrow \infty$, all the eigenmodes in (34) receive the same power and the rest is not used, and when $\gamma \rightarrow 0$, only the strongest eigenmode is used. Thus, the low-SNR behavior is the same as without uncertainty, and the high-SNR one makes use of a smaller number of eigenmodes (recall that all nonzero eigenmodes are used at high SNR in the channel without uncertainty), which is different from the multiplicative uncertainty in Theorem 2, where all nonzero eigenmodes are used at high SNR.

When the uncertainty radius ε increases, the water filling in Theorems 3 and 4 will use fewer and fewer modes until only the strongest mode of the nominal channel remains active⁶, which occurs when

$$[\sigma_2(\mathbf{H}_0) - \varepsilon]_+^{-2} - [\sigma_1(\mathbf{H}_0) - \varepsilon]_+^{-2} \geq m\gamma \tag{35}$$

and the transmit beamforming becomes the optimal robust strategy in this large-uncertainty regime. Under this condition, the results in [35] are optimal from the *information-theoretic* perspective. Note that, unlike (22), (35) holds for any SNR when $\varepsilon \geq \sigma_2(\mathbf{H}_0)$, and also for any ε when the nominal channel is of rank one (so that the results in [30] and [31] are also optimal from the information-theoretic perspective).

A possible condition for the negligible effect of uncertainty is that $\varepsilon \ll \sigma_r(\mathbf{H}_0)$, but it may be too conservative as small singular values may contribute little to the capacity. Further note that, in order to achieve the optimum, the transmitter needs to know only \mathbf{H}_0 (but not $\Delta \mathbf{H}$), since $\mathbf{R}_b = \mathbf{R}_b(\mathbf{H}_0)$ and the knowledge of $\Delta \mathbf{H}$ does not increase the capacity.

Remark 1: In general, the worst-case channel is not unique (but the compound capacity $C_{\max \min} = C_{\min \max}$ is). Indeed, when \mathbf{R}_b allocates zero power to certain signaling subspace, $\Delta \mathbf{H}$ may take arbitrary values on that subspace without affecting the capacity. Specifically, \mathbf{R}_b can be represented as

$$\mathbf{R}_b = \sum_{i_+} \lambda_i(\mathbf{R}_b) \mathbf{v}_{0i} \mathbf{v}_{0i}^+ \tag{36}$$

where $i_+ = \{i : \lambda_i(\mathbf{R}_b) > 0\}$ is the index set of active eigenmodes and \mathbf{v}_{0i} is i th right singular vector of the nominal channel \mathbf{H}_0 , so that the SVD of the generic worst-case channel perturbation is

$$\Delta \mathbf{H}_w = - \sum_{i_+} \min(\sigma_i(\mathbf{H}_0), \varepsilon) \mathbf{u}_{0i} \mathbf{v}_{0i}^+ + \sum_{i_c^c} a_i \mathbf{u}_{0i} \mathbf{v}_{0i}^+ \tag{37}$$

⁶In this sense, increasing uncertainty is equivalent to increasing dispersion of the eigenmodes, making some of them unusable.

where i_+^c is the complement of i_+ , \mathbf{u}_{0i} is i th left singular vector of the nominal channel \mathbf{H}_0 , and a_i take arbitrary values ($|a_i|$ are the singular values of $\Delta\mathbf{H}_w$ corresponding to signaling directions with zero power; (29) corresponds to $a_i = 0$). Indeed

$$\begin{aligned} & (\mathbf{H}_0 + \Delta\mathbf{H}_w)\mathbf{R}_b(\mathbf{H}_0 + \Delta\mathbf{H}_w)^+ \\ &= (\mathbf{H}_0 + \Delta\mathbf{H}'_w)\mathbf{R}_b(\mathbf{H}_0 + \Delta\mathbf{H}'_w)^+ \end{aligned} \quad (38)$$

where

$$\Delta\mathbf{H}'_w = - \sum_{i_+} \min(\sigma_i(\mathbf{H}_0), \varepsilon) \mathbf{u}_{0i} \mathbf{v}_{0i}^+ \quad (39)$$

so that the capacity is not affected by a_i .

Let us compare the performance of the best transmission strategy in Theorem 4 to that of the isotropic signaling with uniform power allocation across transmitters⁷, i.e., when $\mathbf{R} = \mathbf{I}$, which achieves the following mutual information on the compound channel:

$$\begin{aligned} I(\mathbf{H}_w, \mathbf{I}) &= \log |\mathbf{I} + \gamma \mathbf{H}_w \mathbf{H}_w^+| \\ &= \sum_{i=1}^r \log \left(1 + \gamma (\sigma_i(\mathbf{H}_0) - \varepsilon)^2 \right) \end{aligned} \quad (40)$$

where $r \geq 1$ is the worst-case channel rank ($\sigma_i(\mathbf{H}_0) > \varepsilon$). At low SNR, this becomes

$$I(\mathbf{H}_w, \mathbf{I}) \approx \gamma \sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^2 \quad (41)$$

which should be contrasted with the optimal signaling in Theorem 4.1 (only the strongest eigenmode is active at low SNR), which is

$$C_{\max \min} \approx m\gamma (\sigma_1(\mathbf{H}_0) - \varepsilon)^2 \quad (42)$$

and the SNR loss factor is

$$\frac{I(\mathbf{H}_w, \mathbf{I})}{C_{\max \min}} \approx \frac{\sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^2}{m (\sigma_1(\mathbf{H}_0) - \varepsilon)^2} \quad (43)$$

i.e., there is no SNR loss for isotropic signaling only when the worst-case channel is of full rank and its singular values are identical; otherwise, the isotropic signaling is strictly suboptimal. At sufficiently high SNR, $\gamma \rightarrow \infty$, the optimal signaling of Theorem 4 allocates equal powers to all active eigenmodes, so that

$$C_{\max \min} \approx \sum_{i=1}^r \log \left(1 + \frac{m}{r} \gamma (\sigma_i(\mathbf{H}_0) - \varepsilon)^2 \right). \quad (44)$$

Compared to isotropic signaling in (40), the latter suffers m/r -fold SNR loss over the former, and the loss disappears if $m = r$, i.e., the worst-case channel is of the full rank. Under the

⁷which is the optimum signaling when the uncertainty set is isotropically unconstrained and only the lower bound on the channel power gain is known [28]; this technique is also appealing due to its simplicity: no need for channel estimation and feedback, no adaptive transmission.

condition of Theorem 4.3, the optimal signaling achieves the same capacity as in (44), with $\sigma_i(\mathbf{H}_0) = \sigma_0$, $i = 1, \dots, r$, and therefore, the isotropic signaling suffers the same performance loss as at high SNR above.

To get some insight into the structure of the worst-case channel, in the following we give some properties of the optimal channel perturbation $\Delta\mathbf{H}_*$, i.e., one that achieves the minimum in $C_{\max \min}$ in the general case, even when the conditions 4.1–4.4 in Theorem 4 do not hold.

Proposition 1: The optimal channel perturbation does not reduce the null space of the nominal channel matrix

$$\mathcal{N}(\mathbf{H}_0) \in \mathcal{N}(\Delta\mathbf{H}_*) \text{ and } \mathcal{N}(\mathbf{H}_0) \in \mathcal{N}(\mathbf{H}_0 + \Delta\mathbf{H}_*) \quad (45)$$

where $\mathcal{N}(\mathbf{H}) = \{\mathbf{x} : \mathbf{H}\mathbf{x} = \mathbf{0}\}$ is the null space of matrix \mathbf{H} . When \mathbf{H} is a channel matrix, the null space $\mathcal{N}(\mathbf{H})$ represents the set of all transmit directions for which the channel gain is zero, so that the optimal channel perturbation preserves the existing null directions but may add new ones. Furthermore, the eigenvectors of the optimal transmit covariance \mathbf{R}_* (i.e., one that achieves the maximum in $C_{\max \min}$) are orthogonal to $\mathcal{N}(\mathbf{H}_0)$, i.e., no transmission on null directions of the nominal channel.

Proof: See the Appendix. \blacksquare

Corollary 1: The rank of the nominal channel upper bounds that of the optimal one, of the optimal channel perturbation, and of the optimal covariance matrix

$$\text{rank}(\mathbf{H}_0 + \Delta\mathbf{H}_*), \text{rank}(\Delta\mathbf{H}_*), \text{rank}(\mathbf{R}_*) \leq \text{rank}(\mathbf{H}_0). \quad (46)$$

Proof: Follows directly from Proposition 1. \blacksquare

Corollary 1 states that the nominal channel rank bounds the rank of all other matrices in the problem, which is intuitively appealing as the optimal perturbation should disrupt the transmission as much as possible and introducing new eigendirections would not serve this purpose.

It is straightforward to see that the properties in Proposition 1 and Corollary 1 also hold for the multiplicative channel uncertainty of the previous section (and, thus, have a significant degree of universality), and also under the conditions of Theorem 4, where the optimal channel perturbation is explicitly available.

We further observe that, under all cases considered, the best covariance and the worst channel are as in (18) and (29), i.e., that the best Tx strategy is multibeamforming on the degraded nominal channel plus water filling to distribute the power among the beams, and the worst channel perturbation is “antiparallel” to the nominal channel. Based on this, a conjecture follows.

Conjecture 1: Theorem 4 holds in general (for any SNR) and not only under conditions 4.1–4.4.

Remark 2: The compound capacity result in Theorem 4 suggests the following mixed transmission strategy in a fading channel. Since the capacity loss is small when $\varepsilon \ll \sigma_{\min}(\mathbf{H}_0)$, one may use a universal code [20] (tailored to \mathbf{H}_w) and not to track channel variations (e.g., with adaptive coding) when they are small so that the demand on system resources is reduced. On the contrary, an adaptive coding scheme is required for

large channel variations since a universal code incurs a large loss in performance.

Remark 3: It follows from earlier Theorems 3 and 4 that [17, Th. 3.3] also applies when there is uncertainty in the singular vectors, i.e., when singular vectors of $\Delta\mathbf{H}$ and \mathbf{H}_0 are not necessarily the same, and that an additional noise term in [17, Th. 5.2] is not necessary under the conditions of Theorem 4.

Remark 4: Following the same approach as in [15], it can be shown that the earlier compound capacity serves as a lower bound to the outage capacity C_δ of a random nonergodic channel with the mean \mathbf{H}_0 and $\Delta\mathbf{H}$ representing the random part, $C_\delta \geq C_{\max \min}$, where the uncertainty set is selected in such a way that $\Pr\{\mathbf{H} \notin S_{\mathbf{H}}\} = \delta$, i.e., the probability that channel realization is not in the uncertainty set equals to the target outage probability δ . While evaluation and optimization of the outage capacity is a very challenging task in general, this relationship, which holds for any fading distribution, allows one to use the compound capacity instead, albeit as a lower bound.

Finally, we note that the optimal covariance matrices in Theorems 1 and 4 not only maximize the mutual information and thus achieve the compound capacity, but also solve the dual problem of minimizing the transmit power required to achieve a given compound capacity.

V. RELATIONSHIP TO OTHER UNCERTAINTIES

Since all norms are equivalent, in the sense that any norm can be upper and lower bounded by any other norm [see (47)], the earlier results can be used to upper and lower bound compound capacity subject to other normed uncertainties and also to uncertainty constraints that are not norms, which we consider in this section.

A. Arbitrary Normed Uncertainty

Consider an uncertainty set subject to an arbitrary normed constraint $\|\Delta\mathbf{H}\| \leq \varepsilon$, where $\|\cdot\|$ is any matrix norm. Since any two norms are equivalent [8], [9]

$$\alpha_l \|\Delta\mathbf{H}\|_2 \leq \|\Delta\mathbf{H}\| \leq \alpha_u \|\Delta\mathbf{H}\|_2 \tag{47}$$

where α_l, α_u are constants independent of $\Delta\mathbf{H}$ but which depend on the norm $\|\cdot\|$ (see Table I; the bounds are tight in a sense that there are matrices that achieve them). Using this, the arbitrary normed uncertainty set can be upper and lower bounded via that of the spectral norm in (5), so that

$$C_2 \left(\frac{\varepsilon}{\alpha_l} \right) \leq C_{\text{any}} \leq C_2 \left(\frac{\varepsilon}{\alpha_u} \right) \tag{48}$$

where

$$C_{\text{any}} = \max_{\text{tr}(\mathbf{R}) \leq m} \min_{\mathbf{H} \in S_{\text{any}}} I(\mathbf{H}, \mathbf{R})$$

is the compound capacity of the arbitrary normed uncertainty set and C_2 is that for the spectral norm in (5). Note that the multibeamforming on the nominal channel in combination with water-filling power allocation across the beams achieves the compound capacity within these limits. It follows from Table I that when m, n are not large, the bounds in (48) are reasonably

TABLE I
EQUIVALENCE OF NORMS [8].

Norm $\ \cdot\ $	α_l	α_u
$\ \mathbf{H}\ _{\text{col}} = \max_j \sum_i h_{ij} $	$1/\sqrt{m}$	\sqrt{n}
$\ \mathbf{H}\ _{\text{row}} = \max_i \sum_j h_{ij} $	$1/\sqrt{n}$	\sqrt{m}
$\ \mathbf{H}\ _1 = \sum_{ij} h_{ij} $	1	$\sqrt{m \cdot n \cdot \min(n, m)}$
$\ \mathbf{H}\ _F = \sqrt{\sum_{ij} h_{ij} ^2}$	1	$\sqrt{\min(n, m)}$
$\ \mathbf{H}\ _\infty = \max_{ij} h_{ij} $	$1/\sqrt{n \cdot m}$	1

tight. In the following, we specialize this generic result to some particular cases of interest. For the sake of completeness, we also note a few generic properties of the compound capacity $C_{\text{any}}(\varepsilon)$ as a function of the uncertainty radius ε .

Property 1: $C_{\text{any}}(0) = C(\mathbf{H}_0)$.

Property 2: $C_{\text{any}}(\varepsilon \geq \|\mathbf{H}_0\|) = C_{\text{any}}(\|\mathbf{H}_0\|) = 0$.

Property 3: $dC_{\text{any}}(\varepsilon)/d\varepsilon < 0$ for $\varepsilon < \|\mathbf{H}_0\|$ and 0 otherwise.

B. Trace Norm

The uncertainty is constrained via $\text{tr}(\Delta\mathbf{H}\Delta\mathbf{H}^+) = \|\Delta\mathbf{H}\|_F^2 \leq \varepsilon^2$, which can be upper and lower bounded using

$$\sigma_1^2(\Delta\mathbf{H}) \leq \text{tr}(\Delta\mathbf{H}\Delta\mathbf{H}^+) = \sum_i \sigma_i^2(\Delta\mathbf{H}) \leq m\sigma_1^2(\Delta\mathbf{H})$$

so that the compound capacity C_{tr} for this norm satisfies

$$C_2(\varepsilon) \leq C_{tr} \leq C_2 \left(\frac{\varepsilon}{\sqrt{m}} \right). \tag{49}$$

The bounds are reasonably tight when m is not too large. The lower bound is achieved when the optimal channel perturbation (under the spectral norm constraint) is rank-one (this happens when $\text{rank}(\mathbf{H}_0) = 1$ or at low SNR, when $\text{rank}(\mathbf{R}_b) = 1$). The upper bound is achieved when the optimal channel perturbation (under the trace norm) has identical singular values. We note that both the spectral norm and the trace constraint bound the singular values of the channel uncertainty and put no limitations on its singular vectors.

Fig. 1 compares the upper and lower bounds in (49) to the nominal channel capacity of the following channel:

$$\begin{aligned} \mathbf{H}_0 &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \\ \Delta\mathbf{H}_w &= - \begin{bmatrix} \min(2, \varepsilon) & 0 \\ 0 & \min(1, \varepsilon) \end{bmatrix} \\ \mathbf{H}_w &= \begin{bmatrix} (2 - \varepsilon)_+ & 0 \\ 0 & (1 - \varepsilon)_+ \end{bmatrix}. \end{aligned} \tag{50}$$

Clearly, the bounds in (49) allow one to estimate the compound capacity under the trace constraint reasonably well over the whole SNR range. While this example may appear to be rather restrictive, it is generic enough since the compound capacity and its bounds do not depend on singular vectors but only on the singular values and their maximum uncertainty, as Theorem 4 indicates, so that the same result as in Fig. 1 applies to any other \mathbf{H}_0 with the same singular values as in (50) (in this generic setting, the diagonal entries of the matrices in (50)

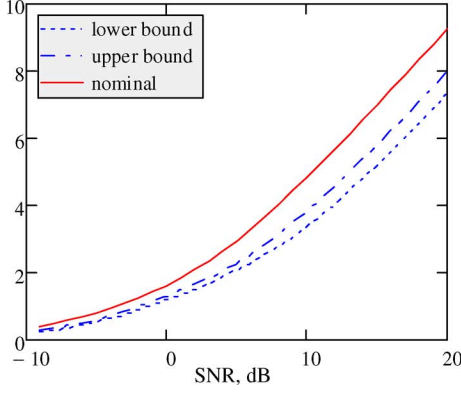


Fig. 1. Nominal channel capacity and the lower and upper bounds of the compound capacity in (49) for the channel in (50), $\varepsilon = 0.5$.

should be thought of as singular values of the corresponding nondiagonal matrices).

C. Weighted Trace

A generalization of the aforementioned case is the following constraint: $\sum_i \alpha_i \sigma_i^2(\Delta \mathbf{H}) \leq \varepsilon^2$, $\sum_i \alpha_i = m$, $\alpha_i \geq 0$, for which the compound capacity is bounded as

$$C_2\left(\frac{\varepsilon}{\sqrt{\alpha_1}}\right) \leq C_{wtr} \leq C_2\left(\frac{\varepsilon}{\sqrt{m}}\right). \quad (51)$$

Note that the upper bound is the same as in (49) and the lower bound depends on α_1 only; both bounds are achieved under the same conditions as in the previous section.

D. Individual Singular Value Constraints

Consider the case when individual singular values of $\Delta \mathbf{H}$ are constrained as $\sigma_i(\Delta \mathbf{H}) \leq \varepsilon_i$, $\varepsilon_1 \geq \varepsilon_2 \geq \dots \geq \varepsilon_m \geq 0$, in which case the compound capacity C_{ind} is bounded as

$$C_2(\varepsilon_1) \leq C_{\text{ind}} \leq C_2(\varepsilon_m) \quad (52)$$

This case includes, as a special case, the relative constraint of the form $\sigma_i(\Delta \mathbf{H})/\sigma_i(\mathbf{H}_0) \leq \varepsilon$, by using $\varepsilon_i = \varepsilon \sigma_i(\mathbf{H}_0)$. In this special case, C_{ind} is as in Section IV-A.

E. Weighted Trace Constraint of Wiesel

The uncertainty is bounded as $\text{tr}(\Delta \mathbf{H} \mathbf{W} \Delta \mathbf{H}^+) \leq 1$, where \mathbf{W} is positive semidefinite. Using the inequalities

$$\lambda_{\min} \|\Delta \mathbf{H}\|_F^2 \leq \text{tr}(\Delta \mathbf{H} \mathbf{W} \Delta \mathbf{H}^+) \leq \lambda_{\max} \|\Delta \mathbf{H}\|_F^2$$

where λ_{\min} , λ_{\max} are the minimum and maximum eigenvalues of \mathbf{W} , and the compound capacity C_{ws} can be bounded as

$$C_2\left(\frac{\varepsilon}{\sqrt{\lambda_{\min}}}\right) \leq C_{ws} \leq C_2\left(\frac{\varepsilon}{\sqrt{m \lambda_{\max}}}\right). \quad (53)$$

The bounds in (53) are reasonably tight provided that $\lambda_{\max}/\lambda_{\min}$ and m are not too large. The lower bound is achieved when $\text{rank}(\Delta \mathbf{H}_w) = 1$ (under the spectral norm) and its right singular vector is the eigenvector of \mathbf{W} corresponding to the minimum eigenvalue. The upper bound is achieved when $\Delta \mathbf{H}_w$ has identical singular values and \mathbf{W} has identical eigenvalues.

We would like to emphasize that, unlike the results in [15] that hold for rank-one \mathbf{H}_0 , the bounds in (53) hold for arbitrary-rank \mathbf{H}_0 and the compound capacity within these bounds is achieved by multibeamforming on \mathbf{H}_0 with optimal power allocation to the beams via water filling.

In the special case of rank-one \mathbf{H}_0 considered in [15], $\mathbf{H}_0 = \mathbf{h}_{r0} \mathbf{h}_{t0}^+$, where \mathbf{h}_{r0} and \mathbf{h}_{t0}^+ are the Rx and Tx array response vectors, and when $\mathbf{W} = \alpha \mathbf{I}$, the upper and lower bounds in (53) coincide and the spectral norm compound capacity in Theorem 3 reduces to that in [15], with $\mathbf{R}_b = m |\mathbf{h}_{t0}|^{-2} \mathbf{h}_{t0} \mathbf{h}_{t0}^+$, i.e., the best Tx strategy is the beamforming tailored to the nominal channel, and the worst-case channel uncertainty is opposite of the nominal channel, $\Delta \mathbf{H}_w = -a \mathbf{H}_0$, where $a = \varepsilon |\mathbf{h}_{t0}|^{-1} |\mathbf{h}_{r0}|^{-1}$. Thus, the spectral and Frobenius-normed channel uncertainties give the same result in this case.

VI. COMPOUND MULTIPLE ACCESS CHANNEL

Let us now consider the MIMO MAC, where multiple transmitters (mobile users) communicate to a single receiver (a base station)

$$\mathbf{y} = \sum_k \mathbf{H}_k \mathbf{x}_k + \xi \quad (54)$$

where \mathbf{H}_k is the channel matrix of user k (mobile) and \mathbf{x}_k is its transmitted signal, which has covariance matrix $\mathbf{R}_k = \mathbf{x}_k \mathbf{x}_k^+$, and the summation is over all users. For given covariance matrices, the sum rate R_{Σ} of this channel is [21]

$$R_{\Sigma} = \log \left| \mathbf{I} + \gamma \sum_k \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+ \right| \quad (55)$$

where $\gamma = 1/\sigma_0^2$ and σ_0^2 is the noise variance in each receive antenna, and this rate is achieved by Gaussian signaling for each user. The sum capacity of this channel can be obtained by optimizing over all possible covariance matrices subject to per-user power constraints [21]

$$C_{\Sigma}(\gamma) = \max_{\{\mathbf{R}_k\}} \log \left| \mathbf{I} + \gamma \sum_k \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+ \right|, \quad (56)$$

s.t. $\text{tr} \mathbf{R}_k \leq P_k, \mathbf{R}_k \geq 0$

where P_k is k th user power, and the optimal covariance \mathbf{R}_{bk} is given by single-user water filling on k th user channel \mathbf{H}_k when the equivalent noise covariance (including interference from other users) is $\mathbf{I} + \gamma \sum_{j \neq k} \mathbf{H}_j \mathbf{R}_j \mathbf{H}_j^+$, so that Gaussian signaling and treating multiple-access interference as noise is sum-rate optimal. While an analytical solution of (56) is not known, a convergent iterative algorithm to solve it is given in [21].

In the compound MAC channel subject to multiplicative uncertainty

$$\mathbf{H}_k = (\mathbf{I} + \mathbf{E}_k) \mathbf{H}_{0k}, \sigma_1(\mathbf{E}_k) \leq \varepsilon_k \quad (57)$$

where \mathbf{H}_{0k} , \mathbf{H}_k , and \mathbf{E}_k are the nominal and actual channels, and the uncertainty, all of k th user. In the following, we give its

compound capacity when all users have the same channel uncertainty ⁸ by generalizing the corresponding result in Theorems 1 and 2.

A. Compound MAC With Identical Uncertainties

Proposition 2: Consider the compound MIMO MAC channels as in (54) and (57) when all users experience the same uncertainty, $\mathbf{E}_k = \mathbf{E}$, $\sigma_1(\mathbf{E}) \leq \varepsilon \forall k$. Its compound capacity is given by (58), as shown at the bottom of the page, where $C_{0\Sigma}(\gamma)$ is the capacity in (56) when all channels are nominal, $\mathbf{H}_k = \mathbf{H}_{0k}$, so that the effect of multiplicative uncertainty is the relative SNR loss of $(1 - \varepsilon)^2$ (compared to the nominal channel). The optimal covariance matrices are as in (56) for $\mathbf{H}_k = \mathbf{H}_{0k}$, where the water filling is done at the effective SNR $(1 - \varepsilon)^2\gamma$, which is equivalent to the optimal transmission on the degraded nominal channels $(1 - \varepsilon)\mathbf{H}_{0k}$.

Proof: Using the singular value inequalities in (14), one obtains (59), and the lower bound is achieved by the worst channel perturbation $\mathbf{E}_w = -\text{diag}(\varepsilon \dots \varepsilon)$, which proves (a). To obtain (b), observe that the chain argument in (60) holds, so that all inequalities hold with equality. ■

Thus, similarly to the single-user MIMO compound channel with multiplicative uncertainty, $C_{\max \min} = C_{\min \max} = C_{0\Sigma}((1 - \varepsilon)^2\gamma)$ and the saddle-point property holds for the compound MIMO MAC channel when all users experience the same uncertainty. The iterative water-filling algorithm in [21]

⁸such uncertainty may be introduced by an imperfect calibration of the channel estimation algorithm at the base station and, hence, will affect all users equally. It may also be introduced by local scatterers around the base station, not accounted for in the calibration process (due to say dynamic nature of the environment), with similar effect. This is reminiscent of the popular Kronecker channel correlation model.

also applies to this compound channel by transmitting on the degraded nominal channels $(1 - \varepsilon)\mathbf{H}_{0k}$.

B. Low SNR Regime

To obtain further results, let us consider the MAC channel in the low-SNR regime $\gamma \rightarrow 0$ (also known as a wideband regime, in which many practical systems, e.g., CDMA, operate [22], [23]). In this regime, (55) simplifies to $R_\Sigma \approx \gamma \sum_k \text{tr}(\mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+)$ and the solution to the optimization problem in (56) follows immediately as the optimization problem decouples into single-user optimizations

$$C_\Sigma(\gamma) = \gamma \sum_k \max_{\{\mathbf{R}_k\}} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) = \gamma \sum_k \sigma_1^2(\mathbf{H}_k) P_k \quad (61)$$

where second equality follows from Von Neumann's trace inequality $|\text{tr}(\mathbf{A}\mathbf{B})| \leq \sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{B})$ [24] applied to $\mathbf{A} = \mathbf{H}_k^+ \mathbf{H}_k$, $\mathbf{B} = \mathbf{R}_k$, and observing that

$$\sum_i \sigma_i^2(\mathbf{H}_k) \sigma_i(\mathbf{R}_k) \leq \sigma_1^2(\mathbf{H}_k) \sum_i \sigma_i(\mathbf{R}_k) = \sigma_1^2(\mathbf{H}_k) P_k.$$

The equality is achieved by $\mathbf{R}_{bk} = P_k \mathbf{v}_{1k} \mathbf{v}_{1k}^+$, where \mathbf{v}_{1k} is the right singular vector of \mathbf{H}_k corresponding to the largest singular value, i.e., by the beamforming on the strongest eigenmode of each user. We are now in a position to characterize the sum capacity of the compound MIMO MAC channel when multiplicative uncertainties are not the same.

Proposition 3: In the low SNR regime, the sum-rate capacity of a compound MIMO MAC channel with multiplicative

$$C_{\max \min} = \max_{\{\text{tr} \mathbf{R}_k \leq P_k\}} \min_{\sigma_1(\mathbf{E}) \leq \varepsilon} \log \left| \mathbf{I} + \gamma(\mathbf{I} + \mathbf{E}) \sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ (\mathbf{I} + \mathbf{E})^+ \right| \stackrel{(a)}{=} C_{0\Sigma}((1 - \varepsilon)^2\gamma) \stackrel{(b)}{=} C_{\min \max} \quad (58)$$

$$\sigma_i \left((\mathbf{I} + \mathbf{E}) \left(\sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ \right)^{1/2} \right) \geq (1 - \varepsilon) \sigma_i \left(\left(\sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ \right)^{1/2} \right) \quad (59)$$

$$\begin{aligned} C_{\max \min} \leq C_{\min \max} &= \min_{\sigma_1(\mathbf{E}) \leq \varepsilon} \max_{\{\text{tr} \mathbf{R}_k \leq P_k\}} \log \left| \mathbf{I} + \gamma(\mathbf{I} + \mathbf{E}) \sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ (\mathbf{I} + \mathbf{E})^+ \right| \quad (60) \\ &\leq \max_{\{\text{tr} \mathbf{R}_k \leq P_k\}} \log \left| \mathbf{I} + \gamma(\mathbf{I} + \mathbf{E}_w) \sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ (\mathbf{I} + \mathbf{E}_w)^+ \right| = C_{0\Sigma}((1 - \varepsilon)^2\gamma) = C_{\max \min} \end{aligned}$$

uncertainty as in (57) is

$$\begin{aligned} C_{\max \min} &= \gamma \sum_k \max_{\text{tr} \mathbf{R}_k \leq P_k} \min_{\sigma_1(\mathbf{E}_k) \leq \varepsilon_k} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) \\ &= \gamma \sum_k (1 - \varepsilon_k)_+^2 \sigma_1^2(\mathbf{H}_{0k}) P_k \\ &= C_{\min \max} \end{aligned} \quad (62)$$

where the best covariance of k th user is $\mathbf{R}_{bk} = P_k \mathbf{v}_{1k} \mathbf{v}_{1k}^+$, \mathbf{v}_{1k} is the right singular vector of its nominal channel \mathbf{H}_{0k} corresponding to the largest singular value, i.e., the best transmission strategy is beamforming on the largest eigenmode of the nominal channel, and the worst-case channel perturbation is $\mathbf{E}_{wk} = -\text{diag}\{\varepsilon_k, \times \cdots \times\}$, where \times means ‘‘any’’ subject to $|\times| \leq \varepsilon_k$.

Proof: The minmax part follows from (61) observing that $\sigma_1^2(\mathbf{H}_k) \geq (1 - \varepsilon_k)_+^2 \sigma_1^2(\mathbf{H}_{0k})$ and the equality is achieved by $\mathbf{E}_{wk} = -\text{diag}\{\varepsilon_k, \times \cdots \times\}$. The maxmin part follows from the following:

$$\begin{aligned} C_{\max \min} &\geq \gamma \sum_k \min_{\sigma_1(\mathbf{E}_k) \leq \varepsilon_k} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_{bk}) \\ &\stackrel{(a)}{\geq} \gamma \sum_k (1 - \varepsilon_k)_+^2 \sigma_1^2(\mathbf{H}_{0k}) P_k \\ &= C_{\min \max} \geq C_{\max \min} \end{aligned} \quad (63)$$

where (a) follows from

$$\text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_{bk}) \geq (1 - \varepsilon_k)_+^2 \text{tr}(\mathbf{H}_{0k}^+ \mathbf{H}_{0k} \mathbf{R}_{bk})$$

which in turn follows from (d) in (76) (since minimization is done for each user separately, single-user result applies), so that all inequalities hold with equality. ■

Let us now consider the additive uncertainty model

$$\mathbf{H}_k = \mathbf{H}_{0k} + \Delta \mathbf{H}_k, \quad \sigma_1(\Delta \mathbf{H}_k) \leq \varepsilon_k. \quad (64)$$

Proposition 4: The sum-rate capacity of the compound MIMO MAC with additive uncertainty in (64) in the low SNR regime is

$$\begin{aligned} C_{\max \min} &= \gamma \sum_k \max_{\text{tr} \mathbf{R}_k \leq P_k} \min_{\sigma_1(\Delta \mathbf{H}_k) \leq \varepsilon_k} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) \\ &= \gamma \sum_k (\sigma_1(\mathbf{H}_{0k}) - \varepsilon_k)_+^2 P_k = C_{\min \max} \end{aligned} \quad (65)$$

and the capacity-achieving transmission strategy is beamforming on the strongest eigenmode of each user.

Similarly to the single-user channel, the effect of uncertainty on the MAC is an average (for multiplicative uncertainty) or per-eigenmode (for additive uncertainty) SNR loss for each user, which is determined by the uncertainty set size. The optimal transmission strategies here are the same as with no uncertainty, except for the rates adjusted according to the SNR loss. Also note that, unlike the generic case where the optimal covariance matrices are not known in a closed form (even without channel uncertainty) but have to be found via a numerical algorithm (see [21]), Propositions 3 and 4 give an explicit form of the optimal signaling at low SNR.

Similar results hold true not only for the sum-rate capacity but also for the capacity region of the compound MIMO MAC. Indeed, each achievable rate point of the MIMO MAC for given $\{\mathbf{R}_k\}$, $\{\mathbf{H}_k\}$ satisfies

$$\sum_{k \in S} R_k \leq \log \left| \mathbf{I} + \gamma \sum_{k \in S} \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+ \right| \quad \forall S \in \{1, \dots, K\} \quad (66)$$

where R_k is the k th user rate and K is the number of users, and the capacity region is obtained via optimization over all possible covariance matrices subject to the individual power constraints $\text{tr} \mathbf{R}_k \leq P_k$ [25]. Since the only difference between the right-hand side in (66) and in (55) is that the summation in the former is limited to a subset S of $\{1, \dots, K\}$, the results in Propositions 2–4 also extend to the capacity region of the compound MAC by limiting the summation to the corresponding subsets, and the effect of uncertainty is an SNR loss measured by the spectral norm of the uncertainty set. The optimal signaling is the same as for the nominal channel applied at the degraded SNR (reduced by uncertainty).

While the Root and Varayia compound channel capacity theorem [6] was proved for the single-user channel only and, thus, does not apply directly to the multiuser scenario here, Propositions 2–4 demonstrate that there exists a transmission strategy that achieves the minmax capacity (i.e., the worst-case channel capacity) on all channels within the uncertainty class (achievability). Since the compound capacity does not exceed the worst-case channel capacity (converse), the latter is the compound capacity under the conditions in Propositions 2–4.

VII. COMPOUND BROADCAST CHANNEL

Let us consider an MIMO BC (dual to the MAC channel in (54))

$$\mathbf{y}_k = \mathbf{H}_k^+ \mathbf{u} + \mathbf{w}_k \quad (67)$$

where \mathbf{y}_k and \mathbf{w}_k are the m -dimensional received signal and noise of k th user (mobile), and \mathbf{u} is the n -dimensional transmitted signal (base station); all \mathbf{w}_k are assumed to be i.i.d. complex Gaussian and independent of each other. Using the MAC–BC duality, the sum-rate capacity is obtained by maximizing (56) over all possible $\{\mathbf{R}_k\}$ subject to the total power constrain

$$\begin{aligned} C_{\Sigma}(\gamma) &= \max_{\{\mathbf{R}_k\}} \log \left| \mathbf{I} + \gamma \sum_k \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+ \right| \\ \text{s.t.} \quad &\sum_k \text{tr} \mathbf{R}_k \leq P, \mathbf{R}_k \geq 0 \end{aligned} \quad (68)$$

where $\{\mathbf{R}_k\}$ are covariance matrices in the dual MAC (not in the BC), and Gaussian signaling in combination with dirty paper coding is optimal [25]. A set of transformations between the dual MAC and BC covariance matrices (achieving the same rates under the same sum power) and also an iterative algorithm to find the optimal matrices that achieve the sum-rate capacity are given in [26] and [29]. Due to significant similarity between (68) and (56), the results of the previous section also apply to the BC after proper modifications to account for the total rather than individual power constraint (we omit the proofs for brevity).

Proposition 5: Consider the compound MIMO BC channel, $\mathbf{H}_k = (\mathbf{I} + \mathbf{E})\mathbf{H}_{0k}$, $\sigma_1(\mathbf{E}) \leq \varepsilon$, where all users experience the same uncertainty \mathbf{E} . Its compound sum-rate capacity is given by (69), as shown at the bottom of the page, where $C_{0\Sigma}(\gamma)$ is the capacity in (68) when all channels are nominal, $\mathbf{H}_k = \mathbf{H}_{0k}$, so that the effect of multiplicative uncertainty is the relative SNR loss of $(1 - \varepsilon)^2$, i.e., the same as for the compound MIMO MAC. The optimal covariance matrices are also as in (68) for $\mathbf{H}_k = \mathbf{H}_{0k}$, where the water filling is done at the degraded SNR $(1 - \varepsilon)^2\gamma$, which is equivalent to the transmission on the degraded nominal channels $(1 - \varepsilon)\mathbf{H}_{0k}$.

In the low SNR regime, (68) reduces to

$$C_{\Sigma}(\gamma) = \gamma \max_{\text{tr} \mathbf{R}_k \leq P} \sum_k \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) \stackrel{(a)}{=} \gamma P \max_k \sigma_1^2(\mathbf{H}_k) \quad (70)$$

where (a) follows in the same way as in (61), and the optimal transmission strategy is beamforming on the strongest eigenmode of the best user (as opposed to the compound MAC in (61), where the optimal strategy is beamforming on the strongest eigenmode of *each* user). The corresponding compound capacities follow.

Proposition 6: In the low SNR regime, the sum-rate capacity of a compound MIMO BC channel with multiplicative uncertainty $\mathbf{H}_k = (\mathbf{I} + \mathbf{E}_k)\mathbf{H}_{0k}$, $\sigma_1(\mathbf{E}_k) \leq \varepsilon_k$, is

$$\begin{aligned} C_{\max \min} &= \gamma \max_{\text{tr} \mathbf{R}_k \leq P} \sum_k \min_{\sigma_1(\mathbf{E}_k) \leq \varepsilon_k} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) \\ &= \gamma P \max_k (1 - \varepsilon_k)_+^2 \sigma_1^2(\mathbf{H}_{0k}) \\ &= C_{\min \max} \end{aligned} \quad (71)$$

and the optimal transmission strategy is beamforming on the largest eigenmode of the best degraded (by the uncertainty size) nominal channel. A worst-case channel perturbation is opposite of this largest eigenmode (it is not unique: channel perturbation along all other eigendirections can be arbitrary).

Proposition 7: The sum-rate capacity of a compound MIMO BC channel with additive uncertainty, $\mathbf{H}_k = \mathbf{H}_{0k} + \Delta \mathbf{H}_k$, $\sigma_1(\Delta \mathbf{H}_k) \leq \varepsilon_k$, in the low SNR regime is

$$\begin{aligned} C_{\max \min} &= \gamma \max_{\text{tr} \mathbf{R}_k \leq P} \sum_k \min_{\sigma_1(\mathbf{E}_k) \leq \varepsilon_k} \text{tr}(\mathbf{H}_k^+ \mathbf{H}_k \mathbf{R}_k) \\ &= \gamma P \max_k (\sigma_1(\mathbf{H}_{0k}) - \varepsilon_k)_+^2 \\ &= C_{\min \max} \end{aligned} \quad (72)$$

and the capacity-achieving transmission strategy is beamforming on the strongest eigenmode of the best degraded (by the uncertainty) nominal channel.

Similarly to the MAC in the previous section, Propositions 5–7 ensure that $C_{\max \min} = C_{\min \max}$ is the operational compound channel capacity since the transmission on the worst-case channels is optimal for the whole class of channels. Note that the optimal transmission strategies at low SNR are given in an explicit form in Propositions 6 and 7, unlike the generic case (even without channel uncertainty), where the optimal covariance matrices follow from an elaborate set of transformations of those in the dual MAC [see [26]] and ultimately have to be evaluated numerically, providing only limited insight about optimal signaling.

Finally, similar results hold also true for the capacity region of the MIMO BC channel, whose capacity region coincides with that of the dirty paper coding [27], which is equal to that of the dual MAC [25], so it includes all the rate points satisfying

$$\sum_{k \in S} R_k \leq \log \left| \mathbf{I} + \gamma \sum_{k \in S} \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^+ \right| \quad \forall S \in \{1, \dots, K\} \quad (73)$$

when optimized over all possible covariance matrices $\{\mathbf{R}_k\}$ subject to the total power constraint $\sum_k \text{tr} \mathbf{R}_k \leq P$. In particular, in the low-SNR regime this simplifies to

$$\sum_{k \in S} R_k \leq \gamma P \max_{k \in S} (1 - \varepsilon_k)_+^2 \sigma_1^2(\mathbf{H}_{0k}) \quad (74)$$

for the multiplicative uncertainty, or to

$$\sum_{k \in S} R_k \leq \gamma P \max_{k \in S} (\sigma_1(\mathbf{H}_{0k}) - \varepsilon_k)_+^2 \quad (75)$$

for the additive uncertainty and Propositions 6 and 7 apply with minor modifications to the compound BC capacity region as well by using (74) and (75), so that the effect of uncertainty on the capacity region is a degradation of the nominal channels by the uncertainty size.

VIII. CONCLUSION

The capacity of a class of uncertain MIMO channels (compound capacity) subject to the spectral norm constraint has been considered for both multiplicative and additive uncertainty models. The compound capacity, achievable by a single code for all channels in the class, has been shown to be equal to the worst-case channel capacity in all considered cases, via explicit closed-form evaluation of the two. Thus, the saddle-point property holds and, in terms of the game theory, neither player can deviate from the optimum strategy without incurring loss. The best transmission strategy and the worst-case channel uncertainties are given in a simple form: the former is the transmission on the eigenmodes of the nominal channel (i.e., multiple beamforming + water filling to distribute the Tx power among the beams), and the latter is opposite of the nominal channel. The

$$C_{\max \min} = \max_{\text{tr} \mathbf{R}_k \leq P} \min_{\sigma_1(\mathbf{E}) \leq \varepsilon} \log \left| \mathbf{I} + \gamma (\mathbf{I} + \mathbf{E}) \sum_k \mathbf{H}_{0k} \mathbf{R}_k \mathbf{H}_{0k}^+ (\mathbf{I} + \mathbf{E})^+ \right| \stackrel{(a)}{=} C_{0\Sigma}((1 - \varepsilon)^2 \gamma) \stackrel{(b)}{=} C_{\min \max} \quad (69)$$

effect of uncertainty is equivalent to an SNR loss commensurable with the size of the uncertainty set and decreases with the latter. Based on this, the following adaptive transmission strategy is suggested: the system does not track the channel and uses a universal code for small channel variations (capacity loss is not large); the system tracks the channel and uses a channel-specific code when the channel variations are large (avoids large loss in capacity due to using a universal code).

In the large-uncertainty regime or in the case of rank-one nominal channel, the best robust transmit strategy is the beamforming on the strongest eigenmode of the nominal channel, the latter being in agreement with [15] and the former indicates the information-theoretic optimality of the results in [35]. The compound channel capacity serves as a lower bound on the outage capacity (at appropriately defined outage probability), for any fading distribution, of a nonergodic fading channel.

The spectral norm compound capacity is used to construct upper and lower bounds of the compound capacity in the case of other bounded uncertainties. Multibeamforming with optimal power allocation across the beams achieves the compound capacity within these bounds.

These results have been extended to compound MAC and BC. While the results have been obtained for frequency-flat channels, they can also be extended to frequency-selective ones via the standard orthogonal frequency-division multiplexing-type argument.

APPENDIX PROOF OF THEOREM 4

The first part is as in (12) and holds in full generality. To prove (32), we have to prove (33) (following [4] and [18], they are equivalent). Keeping in mind (30), we need to prove only $I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}, \mathbf{R}_b) \forall \mathbf{H} \in S_H$. We proceed in the following on a case-by-case basis.

The Low-SNR Regime: To prove claim 4.1 note that the mutual information $I(\mathbf{H}, \mathbf{R}_b)$ can be approximated and lower bounded at low SNR as follows:

$$\begin{aligned}
I(\mathbf{H}, \mathbf{R}_b) &= \log |\mathbf{I} + \gamma \mathbf{H} \mathbf{R}_b \mathbf{H}^+| \\
&\stackrel{(a)}{=} \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2((\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}}) \mathbf{A}_b^{1/2}) \right) \\
&\stackrel{(b)}{\approx} \gamma \cdot \text{tr}((\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}})^+ (\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}}) \mathbf{A}_b) \\
&\stackrel{(c)}{=} \gamma \sum_i \lambda_i(\mathbf{R}_b) |\tilde{\mathbf{h}}_i|^2 \\
&\stackrel{(d)}{\geq} \gamma \sum_i \lambda_i(\mathbf{R}_b) (\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2 \\
&\approx I(\mathbf{H}_w, \mathbf{R}_b)
\end{aligned} \tag{76}$$

where $\mathbf{H}_w = \mathbf{H}_0 + \Delta \mathbf{H}_w = \mathbf{U}_0 \boldsymbol{\Sigma}_w \mathbf{V}_0^+$ is the worst-case channel, $\Delta \tilde{\mathbf{H}} = \mathbf{U}_0^+ \Delta \mathbf{H} \mathbf{V}_0$, $\tilde{\mathbf{h}}_i$ is the i th column of $\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}}$,

$$\begin{aligned}
|\tilde{\mathbf{h}}_i|^2 &= |\sigma_i(\mathbf{H}_0) + \Delta \tilde{h}_{ii}|^2 + \sum_{j \neq i} |\Delta \tilde{h}_{ji}|^2 \\
&\geq (\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2
\end{aligned} \tag{77}$$

and $\Delta \tilde{h}_{ij}$ is i, j th entry of $\Delta \tilde{\mathbf{H}}$; (a) follows from unitary invariance of singular values, (b) follows from $\log(1+x) \approx x$ for $x \ll 1$, which holds at low SNR

$$\gamma \sigma_1^2((\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}}) \mathbf{A}_b^{1/2}) \leq \gamma \lambda_1(\mathbf{R}_b) (\sigma_1(\mathbf{H}_0) + \varepsilon)^2 \ll 1 \tag{78}$$

and the fact that $\text{tr}(\mathbf{H}^+ \mathbf{H}) = \sum_i \sigma_i^2(\mathbf{H})$, (c) is a standard matrix manipulation, and (d) follows from the singular value inequalities in (14). On the other hand, $I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}_w, \mathbf{R}_b)$ from Theorem 3, so that

$$I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}_w, \mathbf{R}_b) \leq I(\mathbf{H}, \mathbf{R}_b) \tag{79}$$

for any admissible \mathbf{H} and \mathbf{R}_b , and (32) follows from it (see e.g., [4]). The low-SNR condition in (78) is not explicit since $\lambda_i(\mathbf{R}_b)$ depend on the SNR. An explicit condition in 4.1 can be obtained by observing that $\lambda_1(\mathbf{R}_b) \leq m$ due to the power constraint.

The High SNR Regime: Let $\text{rank}(\mathbf{H}_w) = r$. It follows then from the properties of the water filling in (27) that $\text{rank}(\mathbf{R}_b) = r$ at sufficiently high SNR (see (82) in the following), so that

$$\begin{aligned}
I(\mathbf{H}, \mathbf{R}_b) &= \log |\mathbf{I} + \gamma \mathbf{H}^+ \mathbf{H} \mathbf{R}_b| \\
&\stackrel{(a)}{=} \log \left| \mathbf{I} + \gamma \underbrace{(\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}})^+ (\boldsymbol{\Sigma}_0 + \Delta \tilde{\mathbf{H}})}_{\mathbf{W}} \mathbf{A}_b \right| \\
&\stackrel{(b)}{=} \log \left| \mathbf{I} + \gamma \begin{bmatrix} \mathbf{W}_r & \mathbf{W}_1 \\ \mathbf{W}_2 & \mathbf{W}_3 \end{bmatrix} \begin{bmatrix} \mathbf{A}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right| \\
&\stackrel{(c)}{=} \log |\mathbf{I}_r + \gamma \mathbf{W}_r \mathbf{A}_r| \\
&\stackrel{(d)}{\geq} \log |\gamma \mathbf{W}_r \mathbf{A}_r| \\
&\stackrel{(e)}{=} \log |\gamma \mathbf{A}_r| + \log |\mathbf{W}_r| \\
&\stackrel{(f)}{=} \sum_{i=1}^r \log(\gamma \lambda_i(\mathbf{R}_b)) + \log \sigma_i^2(\boldsymbol{\Sigma}_{0r} + \Delta \tilde{\mathbf{H}}_r) \\
&\stackrel{(g)}{\geq} \sum_{i=1}^r \log \left(\gamma \lambda_i(\mathbf{R}_b) (\sigma_i(\mathbf{H}_0) - \varepsilon)^2 \right) \\
&\stackrel{(h)}{\approx} I(\mathbf{H}_w, \mathbf{R}_b)
\end{aligned} \tag{80}$$

where \mathbf{A}_b is the diagonal matrix of eigenvalues of \mathbf{R}_b

$$\mathbf{R}_b = \mathbf{V}_0 \mathbf{A}_b \mathbf{V}_0^+, \quad \mathbf{A}_b = \text{diag}(\underbrace{\lambda_1(\mathbf{R}_b) \dots \lambda_r(\mathbf{R}_b)}_r, 0, \dots, 0)$$

and $\mathbf{A}_r = \text{diag}(\lambda_1(\mathbf{R}_b) \dots \lambda_r(\mathbf{R}_b))$ is its principal $r \times r$ submatrix (of positive eigenvalues); $\Delta \tilde{\mathbf{H}}$ is as in (76); $\boldsymbol{\Sigma}_{0r}$ and $\Delta \tilde{\mathbf{H}}_r$ are obtained from $\boldsymbol{\Sigma}_0$ and $\Delta \tilde{\mathbf{H}}$, respectively, by keeping their first r columns and eliminating the rest; \mathbf{W}_r is the principal $r \times r$ submatrix of \mathbf{W} ; \mathbf{I}_r is the $r \times r$ identity matrix. (a) follows in the same way as (a) in (76); (b) make uses of the fact that $\text{rank}(\mathbf{R}_b) = r$ to block-partition \mathbf{W} and \mathbf{A}_b ; (c) follows from block-matrix multiplication rules and properties of the determinant; (d) follows from the fact that $|\mathbf{A}| \geq |\mathbf{B}|$ if $\mathbf{A} \geq \mathbf{B} \geq 0$ [8], where $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite, $\mathbf{A} - \mathbf{B} \geq 0$; (e) follows from $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$ for square matrices \mathbf{A}, \mathbf{B} ; (f) follows from the fact that $\lambda_i(\mathbf{W}_r) =$

$\sigma_i^2(\Sigma_{0r} + \Delta \tilde{\mathbf{H}}_r)$ (by construction); (g) follows from the singular value inequalities in (14) applied to $\sigma_i(\Sigma_{0r} + \Delta \tilde{\mathbf{H}}_r)$ and the following argument:

$$\sigma_1(\Delta \tilde{\mathbf{H}}_r) \stackrel{(j)}{\leq} \sigma_1(\Delta \tilde{\mathbf{H}}) \stackrel{(k)}{=} \sigma_1(\Delta \mathbf{H}) \leq \varepsilon$$

where (j) is due to the interlacing theorem for singular values ([7, Corollary 3.1.13]—removing columns from a matrix cannot increase its spectral norm), and (k) is due to unitary invariance of singular values; (i) follows from the fact that $\sigma_i(\mathbf{H}_w) = (\sigma_i(\mathbf{H}_0) - \varepsilon)_+$ and $\log(1 + x) \approx \log x$ for $x \gg 1$, which is equivalent to

$$\gamma \lambda_r(\mathbf{R}_b) (\sigma_r(\mathbf{H}_0) - \varepsilon)^2 \gg 1. \quad (81)$$

This condition is not explicit in the SNR γ since the power allocation $\lambda_r(\mathbf{R}_b)$ also depends on the SNR. To obtain an explicit condition, we observe from (27) that all eigenmodes are active provided that

$$\gamma > \frac{r}{m(\sigma_r(\mathbf{H}_0) - \varepsilon)^2} - \frac{1}{m} \sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^{-2} \quad (82)$$

where we used the fact that the water-level μ is in this case (from the power constraint $\sum_i \lambda_i(\mathbf{R}_b) = m$)

$$\mu = \frac{m}{r} + \frac{1}{r\gamma} \sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^{-2} \quad (83)$$

and, using (27), the high-SNR condition can be explicitly expressed as

$$\gamma \gg \frac{2r}{m(\sigma_r(\mathbf{H}_0) - \varepsilon)^2} - \frac{1}{m} \sum_{i=1}^r (\sigma_i(\mathbf{H}_0) - \varepsilon)^{-2}. \quad (84)$$

Thus, using Theorem 3 and (80), (79) and thus (32) follow. Note that the max and min operations effectively decouple in the high-SNR regime, so that the result is not surprising. The active eigenmodes are those for which $\sigma_i(\mathbf{H}_0) > \varepsilon$ and all the weaker modes are not used (this is the effect of uncertainty).

Identical Non-Zero Singular Values: Let us consider the case of $\Sigma_0 = \text{diag}(\underbrace{\sigma_0, \dots, \sigma_0}_r, 0 \dots 0)$

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}_b) &= \log |\mathbf{I} + \gamma \mathbf{H} \mathbf{R}_b \mathbf{H}^+| \\ &= \sum_{i=1}^r \log \left(1 + \gamma \lambda_b \sigma_i^2(\widehat{\mathbf{H}}_0 + \Delta \widehat{\mathbf{H}}) \right) \\ &\geq \sum_{i=1}^r \log \left(1 + \gamma \lambda_b (\sigma_i(\mathbf{H}_0) - \varepsilon)_+^2 \right) \\ &= I(\mathbf{H}_w, \mathbf{R}_b) \end{aligned} \quad (85)$$

where $\widehat{\mathbf{H}}_0, \Delta \widehat{\mathbf{H}}$ are obtained from $\mathbf{H}_0 \mathbf{V}_0, \Delta \mathbf{H} \mathbf{V}_0$ by keeping only first r columns (corresponding to nonzero singular values); first equality follows from the fact that

$$\mathbf{R}_b = \mathbf{V}_0 \text{diag}(\underbrace{\lambda_b, \dots, \lambda_b}_r, 0, \dots, 0) \mathbf{V}_0^+$$

i.e., all nonzero eigenvalues of the optimal covariance \mathbf{R}_b are identical for given Σ_0 [as follows from (27)]. Combining (85) with (30), (33) and thus (32) follow.

Low-Rank (1 or 2) Case: To prove the claim 4.4, consider first the case of $m = 2$ (the case of $m = 1$ is trivial)

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}_b) &= \log |\mathbf{I} + \gamma \mathbf{H} \mathbf{R}_b \mathbf{H}^+| \\ &= \log |\mathbf{I} + \gamma \mathbf{H}^+ \mathbf{H} \mathbf{R}_b| \\ &\stackrel{(a)}{=} \log (1 + \gamma \cdot \text{tr}(\mathbf{H}^+ \mathbf{H} \mathbf{R}_b) + \gamma^2 |\mathbf{H}^+ \mathbf{H} \mathbf{R}_b|) \\ &\stackrel{(b)}{\geq} \log (1 + \gamma \text{tr}(\mathbf{H}_w \mathbf{R}_b \mathbf{H}_w^+) + \gamma^2 |\mathbf{H}_w^+ \mathbf{H}_w \mathbf{R}_b|) \\ &= I(\mathbf{H}_w, \mathbf{R}_b) \end{aligned} \quad (86)$$

where (a) follows from the fact that $|\mathbf{I} + \mathbf{A}| = 1 + \text{tr} \mathbf{A} + |\mathbf{A}|$ for a 2×2 matrix \mathbf{A} and (b) follows from that facts that

$$\begin{aligned} \text{tr}(\mathbf{H}^+ \mathbf{H} \mathbf{R}_b) &\geq \text{tr}(\mathbf{H}_w \mathbf{R}_b \mathbf{H}_w^+) \\ |\mathbf{H}^+ \mathbf{H} \mathbf{R}_b| &\geq |\mathbf{H}_w^+ \mathbf{H}_w \mathbf{R}_b| \end{aligned}$$

which follow from (d) in (76) and (g) in (80), respectively. Consider now the case of $\text{rank}(\mathbf{H}_0) \leq 2$

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}_b) &= \log |\mathbf{I} + \gamma \mathbf{H} \mathbf{R}_b \mathbf{H}^+| \\ &= \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2((\Sigma_0 + \Delta \tilde{\mathbf{H}}) \mathbf{A}_b^{1/2}) \right) \\ &\stackrel{(a)}{=} \sum_{i=1}^m \log \left(1 + \gamma \sigma_i^2(\underbrace{(\Sigma_{02} + \Delta \tilde{\mathbf{H}}_2)}_{\mathbf{W}} \mathbf{A}_{b2}^{1/2}) \right) \\ &\stackrel{(b)}{=} \log (1 + \gamma \cdot \text{tr}(\mathbf{W} \mathbf{W}^+) + \gamma^2 |\mathbf{W} \mathbf{W}^+|) \\ &\stackrel{(c)}{\geq} \log (1 + \gamma \cdot \text{tr}(\mathbf{W}_w \mathbf{W}_w^+) + \gamma^2 |\mathbf{W}_w \mathbf{W}_w^+|) \\ &= I(\mathbf{H}_w, \mathbf{R}_b) \end{aligned} \quad (87)$$

where

$$\Sigma_{02} = \text{diag} \{ \sigma_1(\mathbf{H}_0), \sigma_2(\mathbf{H}_0) \}, \mathbf{A}_{b2} = \text{diag} \{ \lambda_1(\mathbf{R}_b), \lambda_2(\mathbf{R}_b) \}$$

and $\Delta \tilde{\mathbf{H}}_2$ are submatrices of Σ_0, \mathbf{A}_b and $\Delta \tilde{\mathbf{H}}$ containing first two columns, and

$$\mathbf{W}_w = \text{diag} \{ (\sigma_1(\mathbf{H}_0) - \varepsilon)_+, (\sigma_2(\mathbf{H}_0) - \varepsilon)_+ \} \mathbf{A}_{b2}^{1/2}$$

(a) follows from the matrix multiplication rules using the fact that $\Sigma_0 = \text{diag} \{ \sigma_1(\mathbf{H}_0), \sigma_2(\mathbf{H}_0), 0 \dots 0 \}$ and $\mathbf{A}_b = \text{diag} \{ \lambda_1(\mathbf{R}_b), \lambda_2(\mathbf{R}_b), 0 \dots 0 \}$; (b) follows in the same way as (a) in (86) since \mathbf{W} is a 2×2 matrix; (c) follows from the same argument as (b) in (86) observing that $\sigma_1(\Delta \tilde{\mathbf{H}}_2) \leq \sigma_1(\Delta \tilde{\mathbf{H}}) = \sigma_1(\Delta \mathbf{H}) \leq \varepsilon$, where first inequality is due to [7, Corollary 3.1.13] (known as an interlacing theorem for singular values) and the equality is due to unitary invariance of singular values, and the equality is achieved in both inequalities when $\Delta \tilde{\mathbf{H}} = \text{diag} \{ \min(\sigma_1(\mathbf{H}_0), \varepsilon)_+, \min(\sigma_2(\mathbf{H}_0), \varepsilon)_+, 0 \dots 0 \}$, so that $\Delta \mathbf{H}_w$ is as in (29) specialized to the $\text{rank}(\mathbf{H}_0) \leq 2$ case. As before, combination with Theorem 3 results in (33) and (32). Note that in all four cases, the best covariance and worst channel are the same as in Theorem 3.

$$\begin{aligned} \mathbf{H}\mathbf{R}\mathbf{H}^+ &= (\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel} + \Delta\mathbf{H}_{\perp})(\mathbf{R}_{\parallel} + \mathbf{R}_{\perp})(\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel} + \Delta\mathbf{H}_{\perp})^+ \\ &= (\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})\mathbf{R}_{\parallel}(\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})^+ + \Delta\mathbf{H}_{\perp}\mathbf{R}_{\perp}\Delta\mathbf{H}_{\perp}^+ \geq (\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})\mathbf{R}_{\parallel}(\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})^+ \geq 0 \end{aligned} \quad (91)$$

APPENDIX
PROOF OF PROPOSITION 1

Split the channel perturbation matrix as follows:

$$\Delta\mathbf{H} = \Delta\mathbf{H}_{\parallel} + \Delta\mathbf{H}_{\perp} \quad (88)$$

where the rows of $\Delta\mathbf{H}_{\parallel}$ are in $\text{span}\{\mathbf{H}_0^T\}$ ($\text{span}\{\mathbf{A}\}$ denotes the space spanned by the columns of matrix \mathbf{A}) and the rows of $\Delta\mathbf{H}_{\perp}$ are orthogonal to it. This can be accomplished by projecting the rows of $\Delta\mathbf{H}$ on $\text{span}\{\mathbf{H}_0^T\}$ and orthogonally to it. Likewise, present the covariance matrix \mathbf{R} in the form

$$\mathbf{R} = \mathbf{R}_{\parallel} + \mathbf{R}_{\perp} \quad (89)$$

where the eigenvectors of \mathbf{R}_{\parallel} corresponding to its nonzero eigenvalues are in $\text{span}\{\mathbf{H}_0^T\}$ and the eigenvectors of \mathbf{R}_{\perp} corresponding to its nonzero eigenvalues are orthogonal to it, so that

$$\mathbf{H}_0\mathbf{R}_{\perp} = \Delta\mathbf{H}_{\parallel}\mathbf{R}_{\perp} = \Delta\mathbf{H}_{\perp}\mathbf{R}_{\parallel} = 0. \quad (90)$$

Now, consider the matrix $\mathbf{H}\mathbf{R}\mathbf{H}^+$ in (91), shown at the top of the page, where the inequality in (91) follows from the fact that $\Delta\mathbf{H}_{\perp}\mathbf{R}_{\perp}\Delta\mathbf{H}_{\perp}^+ \geq 0$. Using this, one obtains

$$\begin{aligned} I(\mathbf{H}, \mathbf{R}) &= \log |\mathbf{I} + \gamma\mathbf{H}\mathbf{R}\mathbf{H}^+| \\ &\geq \log |\mathbf{I} + \gamma(\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})\mathbf{R}_{\parallel}(\mathbf{H}_0 + \Delta\mathbf{H}_{\parallel})^+| \end{aligned} \quad (92)$$

where the inequality follows from the fact that if $\mathbf{A} \geq \mathbf{B} \geq 0$, then $|\mathbf{A}| \geq |\mathbf{B}|$ [8]. Therefore, we conclude that the optimal channel perturbation $\Delta\mathbf{H}_{*} = \Delta\mathbf{H}_{*\parallel}$ and $\Delta\mathbf{H}_{*\perp} = \mathbf{0}$, i.e., if $\mathbf{H}_0\mathbf{x} = 0$, then $\Delta\mathbf{H}_{*}\mathbf{x} = \mathbf{0}$ and also $(\mathbf{H}_0 + \Delta\mathbf{H}_{*})\mathbf{x} = \mathbf{0}$, from which $\mathcal{N}(\mathbf{H}_0) \in \mathcal{N}(\Delta\mathbf{H}_{*})$ and $\mathcal{N}(\mathbf{H}_0) \in \mathcal{N}(\mathbf{H}_0 + \Delta\mathbf{H}_{*})$ follow, i.e., all zero-gain directions of the nominal channel \mathbf{H}_0 are preserved by the optimal perturbation $\Delta\mathbf{H}_{*}$ and possibly new zero gain directions are introduced. Also notice that \mathbf{R}_{\perp} does not affect the lower bound in (92), so that the optimal covariance satisfies $\mathbf{R}_{*} = \mathbf{R}_{*\parallel}$, $\mathbf{R}_{*\perp} = \mathbf{0}$, i.e., no signaling on the zero gain directions of the nominal channel \mathbf{H}_0 (this result would be trivial if there were no channel perturbation, i.e., full channel knowledge at the transmitter, but is not trivial in our setting, where this condition is not satisfied).

ACKNOWLEDGMENT

The authors would like to thank F. Kschischang, S. Denic, and N. Ahmed for many useful discussions and the reviewers for fruitful suggestions.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [2] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels", Internal Tech. Memo. Murray Hill, NJ, 1995, AT&T Bell Labs.
- [3] H. Bolcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds., *Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge, U.K., Cambridge Univ. Press, 2006.
- [4] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radiotekhnika i Elektronika*, vol. 4, no. 12, pp. 1951–1956, Dec. 1959.
- [5] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, pp. 1229–1241, 1959.
- [6] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.
- [7] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [8] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [9] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ: Princeton Univ. Press, 2005.
- [10] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [11] A. Lapidoth and P. Narayan, "Reliable communications under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [12] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [13] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 543–552, Mar. 2006.
- [14] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maxmin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.
- [15] A. Wiesel, Y. C. Eldar, and S. Shamai, "Optimization of the MIMO compound capacity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1094–1101, Mar. 2007.
- [16] C. D. Charalambous, S. Z. Denic, and S. Djouadi, "Robust capacity of white noise channels with uncertainty," in *Proc. 43rd IEEE Conf. Decis. Control*, Paradise Island, Bahamas, Dec. 14–17, 2004, pp. 4880–4884.
- [17] S. Z. Denic, C. D. Charalambous, and S. M. Djouadi, "Information theoretic bounds for compound MIMO Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1603–1617, Apr. 2009.
- [18] E. Zeidler, *Nonlinear Functional Analysis and Its Applications*. New York: Springer-Verlag, 1986, vol. I, Fixed-Point Theorems.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [20] C. Kose and R. D. Wesel, "Universal space-time trellis codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2717–2727, Oct. 2003.
- [21] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.
- [22] S. Verdu, "Recent results on the capacity of wideband channels in the low-power regime," *IEEE Wireless Commun.*, vol. 9, no. 4, pp. 40–45, Aug. 2002.
- [23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [24] J. Von Neumann, "Some matrix-inequalities and metrization of matric-space," *Tomsk University Review*, vol. 1, pp. 286–300, 1937.
- [25] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [26] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.
- [27] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

- [28] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1707–1727, Jul. 2003.
- [29] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [30] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
- [31] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.
- [32] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
- [33] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen-beamforming based on imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1596–1609, May 2006.
- [34] S. Srinivasa and S. A. Jafar, "The optimality of transmit beamforming: A unified view," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1558–1564, Apr. 2007.
- [35] J. Wang and M. Payaro, "On the robustness of transmit beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5933–5938, Nov. 2010.

Sergey Loyka (M'96–SM'04) was born in Minsk, Belarus. He received the Ph.D. degree in Radio Engineering from the Belorussian State University of Informatics and Radioelectronics (BSUIR), Minsk, Belarus in 1995 and the M.S. degree with honors from Minsk Radioengineering Institute, Minsk, Belarus in 1992. Since 2001 he has been a faculty member at the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. Prior to that, he was a research fellow in the Laboratory of Communications and Integrated

Microelectronics (LACIME) of Ecole de Technologie Supérieure, Montreal, Canada; a senior scientist at the Electromagnetic Compatibility Laboratory of BSUIR, Belarus; an invited scientist at the Laboratory of Electromagnetism and Acoustic (LEMA), Swiss Federal Institute of Technology, Lausanne, Switzerland. His research areas include wireless communications and networks, MIMO systems and smart antennas, RF system modeling and simulation, and electromagnetic compatibility, in which he has published extensively. Dr. Loyka is a technical program committee member of several IEEE conferences and a reviewer for numerous IEEE periodicals and conferences. He received a number of awards from the URSI, the IEEE, the Swiss, Belarus and former USSR governments, and the Soros Foundation.

Charalambos D. Charalambous (SM'05) received the B.S. degree in electrical engineering, the M.E. degree, and the Ph.D. degree from the Department of Electrical Engineering, Old Dominion University, Norfolk, VA, in 1987, 1988, and 1992, respectively. In 2003, he joined the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus, where he is currently Professor. He was an Associate Professor at the School of Information Technology and Engineering, University of Ottawa, Ottawa, ON, Canada, from 1999 to 2003. He has served on the faculty of the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, as a nontenure faculty member, from 1995 to 1999. From 1993 to 1995, he was a Postdoctoral Fellow at the Engineering Department, Idaho State University, Pocatello. His research group is interested in theoretical and technological developments concerning large scale distributed communication and control systems and networks in science and engineering. Dr. Charalambous is currently an associate editor for *Systems and Control Letters*, for *Mathematics of Control, Signals, and Systems*, and the Chair of the IFAC technical committee of *Stochastic Systems*. He served as associate editor for *IEEE COMMUNICATIONS LETTERS*, and for *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*.