Outage Probability Under Channel Distribution Uncertainty

Ioanna Ioannou, Charalambos D. Charalambous, and Sergey Loyka

Abstract—Outage probability and capacity of a class of block-fading MIMO channels are considered under partial channel distribution information. Specifically, the channel or its distribution is not known but the latter is known to belong to a class of distributions where each member is within a certain distance (uncertainty) from a nominal distribution. Relative entropy is used as a measure of distance between distributions. Compound outage probability defined as min (over the transmitted signal distribution) -max (over the channel distribution class) outage probability is introduced and investigated. This generalizes the standard outage probability to the case of partial channel distribution information. Compound outage probability characterization (via 1-D convex optimization and in a closed form), its properties, and approximations are given. It is shown to have two-regime behavior: when the nominal outage probability decreases (e.g., by increasing the SNR), the compound outage first decreases linearly down to a certain threshold (related to the relative entropy distance; this is the nominal outage-dominated regime) and then only logarithmically (i.e., very slowly; this is the uncertainty-dominated regime) so that no significant further decrease is possible. This suggests the following design guideline: the outage probability is decreased by increasing the SNR or optimizing the transmitted signal distribution (both decrease nominal outage) in the first regime and by reducing the channel distribution uncertainty (e.g., via better estimation) in the second one. The compound outage depends on the relative entropy distance and the nominal outage only, all other details (nominal fading and noise distributions) being irrelevant. The transmit signal distribution optimized for the nominal channel distribution is shown to be also optimal for the whole class of distributions. The effect of swapping the distributions in relative entropy is investigated and an error floor effect is established. The compound outage probability under L_p distance constraint is also investigated. The obtained results hold in full generality, i.e., for the general channel model with arbitrary nominal fading and noise distributions.

Index Terms—Channel distribution uncertainty, compound multiple-input multiple-output (MIMO) channel, outage probability/capacity, relative entropy distance.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) wireless systems have received significant attention due to the promise of high spectral efficiency [1], [2], which has been extensively investigated. As with any wireless system, their channel capacity depends significantly on the channel state information (CSI) available at the transmitter and the receiver as well as the fading statistics experienced by the channel [3]. When the fading process is egrodic (i.e., the channel "reveals" its statistics to a single codeword), an appropriate performance indicator is ergodic capacity [1], [3]. On the other hand, when the channel is block-fading (or quasi-static), i.e., stays fixed during a codeword transmission and changes from codeword to codeword, its Shannon capacity is zero in many cases of practical interest (i.e., Rayleigh fading) so that outage capacity (capacity versus outage) and outage probability (for a given target rate) are appropriate performance indicators [1]–[4]. In the block-fading (quasi-static) regime, the channel capacity is not affected by the receiver CSI[1] but depends significantly on the CSI available at the transmitter [3], [9], [12]. Since the CSI is obtained via channel measurements, its accuracy may be limited due to variability and difficult propagation conditions (e.g., low SNR) in a wireless channel. The CSI at the transmitter is further limited due to limitations of the feedback channel (if any). This situation can be modeled via a compound channel model, where the true channel is not known but it is known to belong to a certain (limited) class of channels and the corresponding compound channel capacity theorems have been established [5]–[8]. While these theorems treat all channels in the class equally and build a code that performs well on any such channel, the corresponding capacity is typically limited by the worst channel in the class and may be low, even though most channels in class are good and the worst channel is realized with low probability, i.e., it is a conservative performance indicator. To avoid this problem, a concept of composite channel has been introduced [8], [20], where each channel in a class has associated probability measure so that bad low-probability channels do not penalize significantly the performance metric. The corresponding channel capacity theorems can be proved via the concept of information density [18], [20] or using the compound channel approach [3], [12].

Another possibility to model the uncertainty of CSI is to assume that the transmitter knows only the channel distribution but not the channel itself. A number of results on MIMO channel capacity have been obtained under this assumption [9]–[12]. A comprehensive review of the impact of channel uncertainty on its performance and corresponding coding/decoding strategies can be found in [8]. A concise review of more recent activities on MIMO channels is available in [13]. The compound MIMO channel capacities under the trace and spectral norm constraints have been studied recently in [13]–[16]. A construction of a code approaching the compound channel capacity can be found in [17].

I. Ioannou and C. D. Charalambous are with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: aioannak@yahoo.gr; chadcha@ucy.ac.cy).

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: sergey.loyka@ieee.org).

---

[1]Since the receiver can always learn the channel via a training sequence, which results in asymptotically negligible loss in the capacity in the quasi-static mode (with sufficiently long coherence time) [3], full CSI at the receiver can be assumed that significantly simplifies the analysis.

In this paper, we consider a situation where even the channel distribution information is not available at the transmitter; rather, the transmitter knows that the channel distribution belongs to a certain class centered around a nominal distribution. This models a practical scenario where the channel distribution information is obtained from multiple but limited measurements so that the true distribution is known only with finite accuracy (typically related to the number of independent samples used for estimation). This also models a dynamic scenario where the channel distribution information obtained from past measurements may be outdated. The uncertainty in the channel distribution information at the transmitter may also be related to the limitation of the feedback channel used to supply this information (so that only a quantized version of the true distribution is fed back to the transmitter due to the limited feedback bandwidth, which entails a quantization error). We assume a quasi-static (block-fading) scenario with sufficiently long coherence time so that CSI at the receiver is irrelevant. Our channel model is quite general: we do not assume any particular nominal channel distribution and even the channel noise can be arbitrary and its distribution uncertainty can be included as well (except for examples, where a particular distribution and noise are considered) so that the results are general too. Relative entropy between two distributions is used as a measure of distance so that the distribution uncertainty class includes all distributions within certain relative entropy distance of the nominal one. Similar approach was adopted in [21] to study the ergodic capacity under channel distribution uncertainty[2] and in [22] to investigate an optimal control of stochastic uncertain systems. A justification of relative entropy as a measure of distance between distributions can be found in, e.g., [31] and [32]. Our results on the compound outage probability provide further justification, as they indicate that the relative entropy distance limits the achievable outage probability (capacity) via the error floor effect. When the nominal outage probability is negligible and the distance is small, the compound outage probability equals to the relative entropy distance (regardless of all other details).

In a related line of research, the impact of noise distribution uncertainty and optimal (robust) detection schemes have been investigated in [23]–[25]. While the latter concentrates on error rate performance of uncoded systems, our study here deals with the channel outage probability, which is an ultimate bound on error rate performance of any (coded or not) system and is also achievable using capacity-approaching codes [18].

Since the channel is block-fading, the outage probability and capacity are considered as main performance metrics, which we term "compound outage probability/capacity" to emphasize that it applies to a class of fading distributions (i.e., "compound distribution") rather than any particular one. This parallels the concept of compound channel, where a code is designed to operate on any member in the class. In our case, a code is designed to operate for any channel distribution in the class so that the compound outage probability involves maximization over all feasible channel distributions and minimization over the transmitted signal distribution (subject to the power constraint), and the corresponding compound outage capacity is derived from it.

We also consider a scenario where the transmitted signal distribution is fixed *a priori* (e.g., universal code design).

The system/channel model and the performance metrics (outage probability and capacity) are introduced in Section II. Compound outage probability is defined and investigated in Section III, which includes its closed-form characterization in Theorem 1 (as 1-D convex optimization problem) and the worst channel distribution (which is a piecewise constant scaling of the nominal distribution). Remarkably, the compound outage probability depends only on the nominal one and the relative entropy distance, all other details (e.g., nominal fading and noise distributions) being irrelevant. Properties of the compound outage probability are given in Propositions 1–3, and its two-regime asymptotic behavior is identified in Section III-B. Specifically, as the nominal outage probability decreases (say by increasing the SNR), the compound outage probability first decreases linearly too (the nominal outage-dominated regime), but after a certain threshold (equal to about the relative entropy distance when the latter is small), it decreases only logarithmically (the uncertainty-dominated regime), i.e., very slowly so that significant decrease is not possible anymore. Optimizing the transmitted signal distribution in the latter regime does not bring in significant improvement either so that any reasonable distribution (e.g., isotropic signaling) will do as well. This suggests the following design guideline to reduce the outage probability efficiently: via increasing the SNR or optimizing the transmitted signal distribution in the first regime (both decrease the nominal outage probability) and via reducing the channel distribution uncertainty (e.g., by more accurate estimation) in the second one. Compact, closed-form approximations are obtained for the compound outage probability in these two regimes using the tools of asymptotic analysis. Theorem 2 shows that the transmit signal distribution optimal for the nominal channel distribution is also optimal for the whole class so that, e.g., known optimal transmit covariance matrices for Gaussian MIMO channels (see e.g., [1], [9]–[11]) can be "recycled."

Since relative entropy is not symmetric, Section IV investigates the impact of this asymmetry on the outage probability. Swapping the distributions (nominal and true) is shown to result in the error floor effect: the compound outage probability is bounded away from zero, does not matter how low the nominal outage (or how high the SNR) is. The error floor depends on the relative entropy distance: it increases with it; when it is small, they are equal so that the relative entropy distance also serves as the error floor (irreducible compound outage probability) in this small-uncertainty regime. The uncertainty-induced error floor studied in this paper supplements the three error-floor generating mechanisms in mobile wireless channels (due to temporal and frequency dispersions of the channel and multiuser interference), which have been well studied in the past [33], [34]. In general, a sound system design should take into account all of them. Theorem 3 provides a closed-form expression for the compound outage probability that depends on a unique solution of a single nonlinear equation. An alternative characterization is via 1-D convex optimization. The worst case channel distribution and properties of the compound outage probability are also given, including its two-regime behavior and compact approximations via asymptotic analysis.

---

[2]While the impact of channel distribution on the capacity is quite mild in the ergodic regime (due to averaging over the channel statistics), it is much stronger in the nonergodic regime (no averaging) [4].

Based on the aforementioned results, compound outage capacity is studied in Section V. In Section VI, the class of distributions is considered where the distance is defined via L_p norm (a particular case of $p = 2$ corresponds to popular mean-square-error (MSE) estimation) and its outage probability is related to the aforementioned results. In particular, the compound outage with relative entropy distance serves as a lower bound for this case and the error floor effect is present as well. Section VII concludes this paper. Proofs are given in the Appendix.

II. SYSTEM MODEL, OUTAGE PROBABILITY, AND CAPACITY

Let us consider a general discrete-time baseband MIMO channel as shown in Fig. 1, where $\mathbf{x}$ and $\mathbf{y}$ are the input (transmitted) and output (received) vectors (or sequences), and $\mathbf{H}$ denotes channel state. In the general case, the channel is described by the conditional probability distribution $W(\mathbf{y}|\mathbf{x}, \mathbf{H})$ of $\mathbf{y}$ given $\mathbf{x}$ and $\mathbf{H}$, and the mutual information (per channel use) supported by the channel for a given distribution of $\mathbf{x}$ and a given channel state $\mathbf{H}$ is $I(\mathbf{x}; \mathbf{y}|\mathbf{H})$[3]. We assume that the channel is block-fading (nonergodic), i.e., a particular channel realization $\mathbf{H}$ is selected in the beginning and stays fixed for the whole duration of codeword transmission; next codeword will see a different channel realization.[4] Channel fading distribution is described by its probability density function (pdf) $f(\mathbf{H})$. Most of our results will hold in this general scenario, which includes as special cases frequency-selective (intersymbol interference) or frequency-flat (no ISI) Gaussian MIMO channels.

We will not assume any particular fading and noise distribution (except for examples) so that our results are general and apply to *any* such distribution. The transmitted signal, receiver noise, and the channel are assumed to be independent of each other. We also assume that the transmitter does not known the channel but only has a partial knowledge of its distribution (as explained later on); channel knowledge at the receiver is irrelevant in the block-fading environment with sufficiently long coherence time (i.e., quasi-static).

Main performance metrics in the block-fading regime are outage probability and outage capacity [3], [4], [12].[5] Outage probability is the probability that the channel is not able to support the target rate R. When the transmitter knows the channel distribution (but not the channel itself), the outage probability is

$$P_{\text{out}}(R) = \min_{\rho(\mathbf{x})} \Pr\{I(\mathbf{x}; \mathbf{y}|\mathbf{H}) < R\} \tag{1}$$

where $\rho(\mathbf{x})$ is the distribution of $\mathbf{x}$ subject to the total power constraint $\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] \leq P_T$, $\Pr\{I(\mathbf{x}; \mathbf{y}|\mathbf{H}) < R\}$ is the outage

---

[3]The channel state $\mathbf{H}$ may include channel gain(s) as well as channel noise distribution so that different channel realizations can "see" different noise distributions, which can also model the noise distribution uncertainty, e.g., as in [23]–[25]. The two extreme scenarios are when all channel realizations "see" 1) the same noise distribution but different channel gains or 2) the same gain but different noise distributions.

[4]With a slight modification in notations, this block-fading model can also be extended to the case where each codeword sees a finite number of channel realizations, e.g., as in [12] and [19], and our results will hold in that case as well.

[5]It can be further shown that the outage probability is the best achievable average codeword error probability [18]–[20].
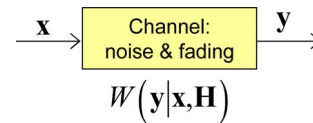

Fig. 1. General discrete-time baseband MIMO block-fading channel model. No assumptions on noise and fading distributions are made.

probability for a given $\rho(\mathbf{x})$ and the minimization is over all possible distributions of the input $\mathbf{x}$ satisfying the power constraint. Outage capacity is defined as the largest possible rate such that the outage probability does not exceed the target value δ

$$C_{0\delta} = \max\{R : P_{\text{out}}(R) \leq \delta\}. \tag{2}$$

Clearly, $P_{\text{out}}(C_{0\delta}) = \delta$. Finally, one may also consider the outage probability and capacity for a given (fixed) $\rho(\mathbf{x})$. Operational meaning of the outage capacity/probability follows from the compound channel capacity theorems [6]–[8], [12], [19] (which guarantee an existence of a code that works on every channel in the no-outage set); see also [18] and [20] for a modern treatment using the concept of information density.

III. COMPOUND OUTAGE PROBABILITY FOR A CLASS OF CHANNELS

Consider the scenario where the transmitter has only partial channel distribution information. Namely, it knows that the channel pdf $f(\mathbf{H})$ is within a certain distance of the nominal distribution $f_0(\mathbf{H})$. We use the relative entropy as a measure of the distance between two distributions so that all feasible distributions f satisfy the following inequality:

$$D(f\|f_0) = \int f \ln \frac{f}{f_0} d\mathbf{H} \leq d \tag{3}$$

where $D(f\|f_0)$ is the relative entropy or Kullback–Leibler distance between the distributions, and d is the maximum possible distance in the uncertainty set to which f belongs; both d and $f_0(\mathbf{H})$ are known to the transmitter. Throughout this paper, we assume that $d < \infty$. In this scenario, the definition in (1) does not apply (since the true distribution f is not known) but can be generalized to

$$P_{\text{out}}^* = \min_{\rho(\mathbf{x})} \max_{D(f\|f_0) \leq d} \Pr\{I(\mathbf{x}; \mathbf{y}|\mathbf{H}) < R\} \tag{4}$$

and the outage capacity can be defined as in (2) with the substitution $P_{\text{out}} \rightarrow P_{\text{out}}^*$. Its operational meaning also follows from the compound channel capacity theorems [7], [8], [12] or from [18] and [20], since the optimal signaling does not depend on the true channel distribution but only on the nominal one and also the relative entropy distance d, both known to the transmitter. This problem setup models a practical situation where the channel distribution information is obtained from measurements or physical modeling, which are never perfect. It also accounts for the fact that the estimated channel distribution may change with time in dynamic scenarios or that the bandwidth-limited feedback provides only a quantized version of the distribution which entails the quantization error. We term

P_{out}^* in (4)"compound outage probability" since it is a performance measure of a class of channel distributions rather than a single distribution. This approach parallels the work on compound channel capacity [5]–[8], [13]–[16] where the channel is not known to the transmitter but it is known to belong to a certain class.

To characterize the compound outage probability P_{out}^*, we adopt a two-step approach: first, we characterize the outage probability for a given input distribution, i.e., no minimization in (4), which also represents a practical situation where this distribution is set *a priori*; then, it is minimized over all feasible input distributions.

A. Step 1: Compound Outage for a Given Input Distribution

When the input distribution $\rho(\mathbf{x})$ is fixed *a priori*, the compound outage probability is

$$P_{\text{out}} = \max_{D(f\|f_0)\leq d} \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R\}. \tag{5}$$

Its characterization is strikingly simple in the general case, i.e., for any noise and nominal fading distributions.

Theorem 1: For a given input distribution $\rho(\mathbf{x})$ and arbitrary nominal fading distribution f_0, the outage probability in (5) can be expressed as

$$P_{\text{out}} = \min_{s\geq 0}[s \ln(1 + (e^{1/s} - 1)\varepsilon) + sd] \tag{6}$$

where

$$\varepsilon = \int_{I(\mathbf{x};\mathbf{y}|\mathbf{H})<R} f_0 d\mathbf{H} \tag{7}$$

is the nominal outage probability (i.e., the outage probability under the nominal channel distribution). The worst channel distribution f^* (the maximizer in (5)) is given by

$$f^* = \frac{(e^{1/s^*} - 1)\ell(\mathbf{H}) + 1}{(e^{1/s^*} - 1)\varepsilon + 1} f_0 \tag{8}$$

where s^* is the minimizing s in (6), and $\ell(\mathbf{H})$ is the indicator of the outage set: $\ell(\mathbf{H}) = 1$ if $I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R$ and 0 otherwise. An alternative characterization of P_{out} is as follows:

$$P_{\text{out}} = \frac{e^{1/s^*}\varepsilon}{1 + \varepsilon(e^{1/s^*} - 1)} \tag{9}$$

where s^* is a unique solution of

$$-\ln(1 + \varepsilon(e^{1/s} - 1)) + \frac{e^{1/s}\varepsilon}{1 + \varepsilon(e^{1/s} - 1)}\frac{1}{s} = d \tag{10}$$

if $d < \ln(1/\varepsilon)$ and $s^* = 0$ otherwise. In the latter case, $P_{\text{out}} = 1$ and a maximizing channel distribution is

$$f^* = \ell(\mathbf{H})f_0/\varepsilon. \tag{11}$$
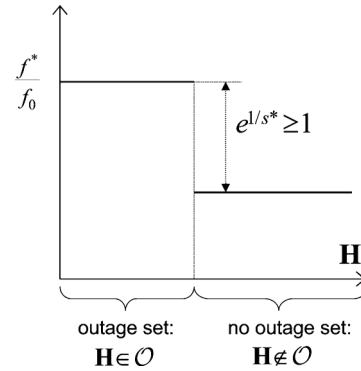
Proof: see the Appendix. ∎

Fig. 2. Worst channel distribution over the nominal one. Note that the worst distribution is a piecewise constant scaling of the nominal one.

Note that Theorem 1 effectively reduces the infinite-dimensional optimization problem in (5) (the optimization there is over the set of all admissible distributions f) to 1-D convex optimization in (6), which can be effectively solved using numerical algorithms [26]. This is accomplished using Lagrange duality theory (see the Appendix for details). As we will see in the following, this is not the only advantage: (6) and (9) also provide a number of insights unavailable from (5). The alternative closed form in (9) requires only a unique solution of the scalar nonlinear equation in (10), which can be found using any suitable numerical technique (e.g., Newton–Raphson method; no convergence problems are expected since the left-hand side of (10) is monotonically decreasing in s).

It is remarkable that the nominal outage distribution enters the compound outage probability in (6) only via the nominal outage probability ε, all other its details being irrelevant, i.e., two different nominal distributions with the same nominal outage probability will produce the same compound outage probability.[6]

Note that the maximizing density f^* in (8) mimics the nominal one f_0 in a piecewise constant manner

$$\frac{f^*}{f_0} = \begin{cases} \frac{e^{1/s^*}}{(e^{1/s^*}-1)\varepsilon+1}, & \text{if } \mathbf{H} \in \mathcal{O} \\ \frac{1}{(e^{1/s^*}-1)\varepsilon+1}, & \text{if } \mathbf{H} \notin \mathcal{O} \end{cases} \tag{12}$$

where $\mathcal{O} = \{\mathbf{H} : \ell(\mathbf{H}) = 1\}$ is the outage set so that the right-hand side of (12) is independent of $\mathbf{H}$ in each set and f^* is a scaled up version of f_0 in the outage set and scaled down otherwise—see Fig. 2.

An additional advantage of (6) is that the resulting optimization problem there is convex, i.e., the function

$$L(s) = s \ln(1 + (e^{1/s} - 1)\varepsilon) + sd \tag{13}$$

is convex in $s > 0$ (see the Appendix) so that $P_{\text{out}} = \min_{s\geq 0} L(s)$ in (6) can be solved efficiently using any known numerical algorithm; the solution is unique and it satisfies $dL(s)/ds = 0$. Alternatively, the tools of asymptotic analysis (see, e.g., [28] and [29]) can be used to obtain approximations

---

[6]We emphasize here that the knowledge of f_0 does not allow one to obtain a better result for a given $\rho(\mathbf{x})$, beyond that available from the knowledge of ε. On the other hand, f_0 is required to optimize P_{out} over $\rho(\mathbf{x})$ in the same way as it is done for ε.
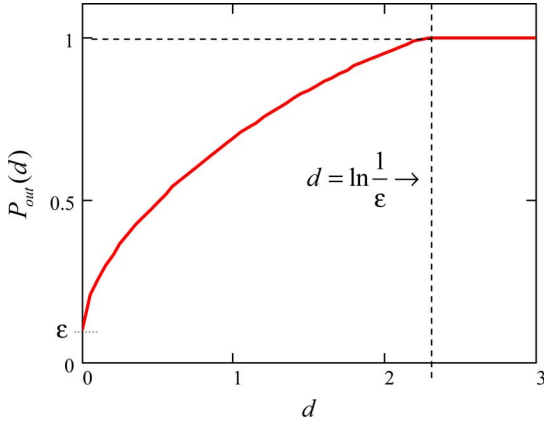
Fig. 3. Properties of P_{out} as a function of d: concave, strictly increasing function, which saturates at $d = \ln(1/\varepsilon)$; $\varepsilon = 0.1$.

(see Section III-B). This dual representation also provides a number of insights, as indicated in the following.

Let us now consider the outage probability in (6) as a function $P_{\text{out}}(d)$ of the distance d. A number of its properties follow.

Proposition 1: For a given input distribution $\rho(\mathbf{x})$, the compound outage probability $P_{\text{out}}(d)$ as a function of distance d has the following properties.

1) $P_{\text{out}}(d)$ is concave in $d \geq 0$.
2) $P_{\text{out}}(d = 0) = \varepsilon$, i.e., the compound outage probability equals the nominal one when $d = 0$.
3) $P_{\text{out}}(d)$ is a nondecreasing function of d, that is

$$P_{\text{out}}(d_1) \leq P_{\text{out}}(d_2), \quad 0 \leq d_1 < d_2 < \infty. \qquad (14)$$

and the equality holds if and only if $P_{\text{out}}(d_1) = P_{\text{out}}(d_2) = 1, 0$, i.e., $P_{\text{out}}(d)$ is a strictly increasing function of the distance d unless $\varepsilon = 0$ or $d \geq \ln(1/\varepsilon)$.
   Proof: see the Appendix. ■

Fig. 3 exemplifies the properties of $P_{\text{out}}(d)$.

Proposition 2: The compound outage probability P_{out} has the following properties.

1) $P_{\text{out}} = 1$ if and only if $d \geq \ln(1/\varepsilon)$.
2) $P_{\text{out}} = 0$ if and only if $\varepsilon = 0$.
3) $P_{\text{out}} \geq \varepsilon$, and the equality holds if and only if $d = 0$ or $\varepsilon = 0, 1$.

   Proof: see the Appendix. ■

While in general the compound and nominal outage probabilities can differ significantly, the former equals 0 if and only if the latter does so. They are also equal when $\varepsilon = 1$.

Proposition 3: The compound outage probability in (6) is an increasing, concave function of the nominal outage ε, i.e.,

$$P_{\text{out}}(\varepsilon_1) \leq P_{\text{out}}(\varepsilon_2), \quad 0 \leq \varepsilon_1 < \varepsilon_2 \leq 1 \qquad (15)$$

with the equality iff $d \geq \ln(1/\varepsilon_1)$ and the boundary conditions $P_{\text{out}}(\varepsilon = 0) = 0$ and $P_{\text{out}}(\varepsilon \geq e^{-d}) = 1$.
   Proof: see the Appendix. ■

Proposition 4: The compound outage probability P_{out} in Theorem 1 can be bounded as follows:

$$\varepsilon \leq P_{\text{out}} \leq \min[d + (e - 1)\varepsilon, 1]. \qquad (16)$$

   Proof: see the Appendix. ■

B. Asymptotic Regimes

We now consider the compound outage in (6) in two limiting regimes.

1) The uncertainty-dominated regime $\varepsilon \to 0$ and fixed d, i.e., the dominant source of outage events is from significant deviation of the true channel distribution from the nominal one (outage events under the nominal distribution can be neglected).
2) The nominal outage-dominated regime $d \to 0$ and fixed ε, i.e., when the impact of channel distribution uncertainty is negligible as outage events under the nominal distribution dominate the performance.

Proposition 5: The outage probability P_{out} in (6) in the low nominal outage regime, $\varepsilon \to 0$ and fixed $d > 0$, is as follows:

$$P_{\text{out}} = \frac{d}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}}(1 + o(1)) \qquad (17)$$

and the optimal (minimizing) s^* in (6) is given by

$$s^* = \frac{1}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}}(1 + o(1)).$$

This is the uncertainty-dominated regime (the main contribution to P_{out} is coming from d rather than ε).
   Proof: see the Appendix. ■

Further analysis shows that the aforementioned approximations (without $o(1)$ term) are accurate provided that $\varepsilon \ll d < 1$. Note from (17) that the main contribution to P_{out} is coming from d (i.e., the uncertainty) rather than ε (i.e., the nominal outage) since $\ln(d/\varepsilon)$ is a slowly varying function of ε so that variations from the nominal channel distribution dominate the outage events. Also, note that the relative entropy distance d is directly related to the compound outage probability, which indicates that it is this distance that should be used as a measure of accuracy in estimating the channel distribution from measurements or physical modeling since it is directly related to the system performance (outage probability and capacity).

Let us now consider the nominal outage-dominated regime (i.e., fixed ε and $d \to 0$).

Proposition 6: In the low channel distribution uncertainty regime, $d \to 0$ and fixed ε, the compound outage probability is

$$P_{\text{out}} = \varepsilon + \sqrt{2d(1 - \varepsilon)\varepsilon} + o(\sqrt{d}) \qquad (18)$$

and the optimal s^* is given by

$$s^* = \sqrt{\varepsilon(1 - \varepsilon)/(2d)}(1 + o(1)).$$

   Proof: see the Appendix. ■

Further analysis shows that the aforementioned approximation is accurate when $d \ll \varepsilon < 1$ and that the impact of uncertainty is negligible, $P_{\text{out}} \approx \varepsilon$, when $\sqrt{d} \ll \sqrt{\varepsilon}$. Comparing Propositions 5 and 6, one concludes that indeed there are two regimes in the behavior of $P_{\text{out}}(\varepsilon)$, as illustrated in Fig. 3.

1) The uncertainty-dominated regime (nominal outage is negligible), when $\varepsilon \ll d < 1$ so that

$$P_{\text{out}} \approx \frac{d}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}} \sim \frac{d}{\ln(1/\varepsilon)} \qquad (19)$$

where $\sim$ means "scales as" so that P_{out} depends linearly on d but only logarithmically (i.e., very slowly) on ε.

2) The nominal outage-dominated regime (uncertainty is negligible), when $d \ll \varepsilon < 1$ and

$$P_{\text{out}} \approx \varepsilon + \sqrt{2d(1-\varepsilon)\varepsilon} \sim \varepsilon \qquad (20)$$

i.e., d contributes very little to the outage probability.

These two regimes immediately suggest some design guidelines related to the outage probability. In the uncertainty-dominated regime, the main way to reduce outage probability is via decreasing the uncertainty of the channel distribution, e.g., via improved channel measurements or modeling; reducing the nominal outage probability is not efficient here so that minimizing it via the optimal transmitted signal distribution is not worth the effort—any reasonable distribution (e.g., isotropic Gaussian signaling in MIMO channels) will do as well. This approach, however, will bring little improvement in the nominal outage-dominated regime, where the only way to reduce the outage probability is via improving systems performance under the nominal fading, e.g., by increasing the SNR or optimizing the input distribution. Note that these conclusions hold for any nominal channel distribution (e.g., not limited to i.i.d. Rayleigh) and for any noise (not only Gaussian).

As an example, let us consider a 1×1 Rayleigh-fading channel with Gaussian noise, in which case $\varepsilon \sim 1/\gamma$, when $d \ll 1$ and $\gamma \gg 1$ so that

$$P_{\text{out}}(\gamma) \sim \begin{cases} \frac{1}{\gamma} & \text{in regime 1 } (\gamma \ll 1/d) \\ \frac{d}{\ln \gamma} & \text{in regime 2 } (\gamma \gg 1/d) \end{cases} \qquad (21)$$

i.e., the outage probability scales with SNR γ as $1/\text{SNR}$ in regime 1 but only as $1/\ln \text{SNR}$ in regime 2 so that increasing the SNR is only efficient in the former case. When $d \ll 1$, increasing the SNR will first decrease P_{out}, but only down to about d and after that point the decrease becomes logarithmically slow. From a practical perspective, it means that P_{out} cannot be reduced significantly beyond d by increasing the SNR. This observation indicates that the relative entropy distance d is indeed an appropriate measure of channel distribution uncertainty in the nonergodic (block-fading) mode. Fig. 4 illustrates the two-regime behavior of $P_{\text{out}}(\gamma)$.

This two-regime behavior can also be linked to the way channel distribution is obtained from measurements: a finite number of fading channel realizations are measured and the empirical channel distribution is derived based on it. However, the relative accuracy of this empirical distribution is always lower at the distribution tails, where fewer measurement points
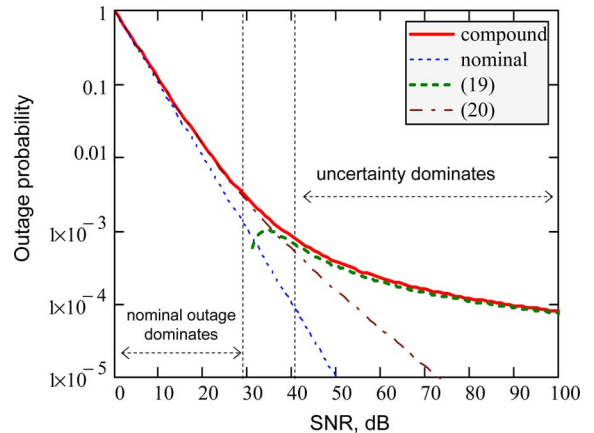

Fig. 4. Two-regime behavior of the compound outage probability. Its approximations in (19) and (20) and nominal outage $\varepsilon = 1/\text{SNR}$ (set for convenience) are also shown; $d = 10^{-3}$. Note that both approximations are accurate in their respective regimes. Decreasing the compound outage probability beyond about $10^{-3} (\approx d)$ requires exponentially high SNR and is not practical (it takes 60 dB extra to go from 10^{-3} to 10^{-4}, while normally, i.e., without uncertainty, it would take only 10 dB).

are available. On the other hand, when the average SNR is high, as in regime 2, an outage event takes place when the channel is very weak, i.e., at the distribution tail so that the inaccuracy in the channel distribution estimation plays a dominant role there. Ultimately, low compound outage probability can only be achieved by ensuring sufficiently high accuracy of the estimated distribution tail (small d), i.e., when a sufficient number of independent measurements fall into that region.

C. Step 2: Minimizing Over the Input Distribution

Using Theorem 1, we are now in a position to characterize the compound outage probability in (4).

Theorem 2: Consider a class of fading channels in (3). Its compound outage probability in (4) can be found from

$$P_{\text{out}}^* = \min_{s \geq 0}[s \ln(1 + (e^{1/s} - 1)\varepsilon^*) + sd] \qquad (22)$$

where $\varepsilon^* = \min_{\rho(\mathbf{x})} \varepsilon$ is the optimized nominal outage probability so that the outage-minimizing input distribution for the class of channel distributions in (3) and for the nominal distribution f_0 are the same

$$\arg \min_{\rho(\mathbf{x})} P_{\text{out}} = \arg \min_{\rho(\mathbf{x})} \varepsilon. \qquad (23)$$

Proof:

$$P_{\text{out}}^* = \min_{\rho(\mathbf{x})} P_{\text{out}}$$

$$\overset{(a)}{=} \min_{\rho(\mathbf{x})} \min_{S \geq 0}[s \ln(1 + (e^{1/s} - 1)\varepsilon) + sd]$$

$$\overset{(b)}{=} \min_{S \geq 0}[s \ln(1 + (e^{1/s} - 1)\varepsilon^*) + sd]$$

where (a) follows from (6) and (b) follows from the fact that $\min_{\rho(\mathbf{x})}$ and $\min_{s \geq 0}$ can be swapped and $\ln(\cdot)$ is a monotonic

function so that the minimization of the compound outage over $\rho(\mathbf{x})$ is equivalent to the minimization of the nominal outage. ∎

A significance of Theorem 2 is that the infinite-dimensional optimization in (4) is reduced to a 1-D convex optimization in (22) and, furthermore, the optimal input distribution is the same as for the nominal channel so that a significant number of known results [1], [9]–[12] apply directly to the compound fading channel as well, i.e., no new search of optimal input distribution/covariances is required.

When the compound outage P_{out}^* in (22) is considered as a function of the distance d, $P_{\text{out}}^*(d)$, its properties mimic those in Proposition 1 with the substitution $\varepsilon \rightarrow \varepsilon^*$. Also, the results and conclusions in Section III-B hold under this substitution. In particular, optimizing the input distribution is worth the effort only in the nominal outage-dominated regime.

IV. IMPACT OF ASYMMETRY

Since relative entropy is not symmetric, i.e., $D(f\|f_0) \neq D(f_0\|f)$, we consider in this section the constraint

$$D(f_o\|f) = \int f_0 \ln \frac{f_0}{f} d\mathbf{H} \leq d \tag{24}$$

to see the impact of the order on the obtained results. One property of the compound outage probability P_{out} for this class of distributions is immediate.

Proposition 7: For a given input distribution, the compound outage P_{out} under the distribution class in (24) is bounded as follows:

$$P_{\text{out}} = \max_{D(f_0\|f) \leq d} \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R\}$$
$$\geq 1 - e^{-d} + e^{-d}\varepsilon \tag{25}$$
$$\geq 1 - e^{-d} \tag{26}$$

and the bounds are achievable. When $d \ll 1$

$$P_{\text{out}} \geq d + \varepsilon. \tag{27}$$

Proof: Equation (25) follows from the fact that the lower bound is achievable by

$$f(\mathbf{H}) = p\delta(\mathbf{H}) + (1-p)f_0(\mathbf{H})$$

where $p = 1 - e^{-d}$ and $\delta(\cdot)$ is the Dirac delta function, i.e., by the distribution that mimics the nominal one except for placing mass p at zero (where the mutual information is zero). The bound in (26) is trivial and is achievable when $e^d - 1 \gg \varepsilon$. Equation (27) follows from $e^{-d} \approx 1 - d$ for $d \ll 1$. ∎

An important conclusion is immediate from (26): $P_{\text{out}}(\varepsilon = 0) \geq 1 - e^{-d}$, i.e., there is a saturation (error floor) effect in the behavior of $P_{\text{out}}(\varepsilon)$: even though $\varepsilon \rightarrow 0$ (e.g., by SNR $\rightarrow \infty$), $P_{\text{out}} \nrightarrow 0$. From (27), $P_{\text{out}} \geq d$ when $d \ll 1$ in this regime, i.e., cannot be made smaller than the relative entropy distance d, does not matter how large the SNR (or how small the nominal outage) is. This is in contrast to (17), where $P_{\text{out}} \rightarrow 0$ when $\varepsilon \rightarrow 0$, even though logarithmically slowly (i.e., no error floor). The absence of error floor in the latter case should not however be overestimated, since the convergence $P_{\text{out}} \rightarrow 0$ is

logarithmically slow in ε, i.e., requires exponentially large SNR so that for all practical purposes, P_{out} also saturates around d, as indicated in Section III-B. Note also that (27) places d and ε on equal footing, re-enforcing our earlier conclusion that d is an adequate measure of fading uncertainty in the nonergodic regime.

Let us now characterize precisely the compound outage probability for the distribution class in (24).

Theorem 3: Consider the compound outage probability for the distribution class in (24) and a given input distribution

$$P_{\text{out}} = \max_{D(f_0\|f) \leq d} \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R\}. \tag{28}$$

It can be expressed in the following form:

$$P_{\text{out}} = \frac{\mu\varepsilon}{\mu - 1 + \varepsilon} \tag{29}$$

where $\mu \geq 1$ is a unique solution of

$$\frac{\mu^\varepsilon(\mu - 1)^{1-\varepsilon}}{\mu - 1 + \varepsilon} = e^{-d}. \tag{30}$$

The maximizing density f^* in (28) is given by

$$f^* = \frac{\lambda^* f_0}{\mu - \ell(\mathbf{H})} \tag{31}$$

where $\lambda^* = e^{-d}(\mu - 1)^\varepsilon\mu^{1-\varepsilon}$. An alternative characterization of P_{out} is

$$P_{\text{out}} = \min_{\lambda \geq 0} \left[\lambda\varepsilon \left(\frac{1}{\mu - 1} + \ln \frac{\mu}{\mu - 1} \right) + \lambda \left(d - \ln \frac{\mu}{\lambda} \right) \right] \tag{32}$$

where

$$\mu = \frac{1}{2} \left(1 + \lambda + \sqrt{(1 - \lambda)^2 + 4\lambda\varepsilon} \right). \tag{33}$$

Proof: see the Appendix. ∎

Note that Theorem 3 reduces the infinite-dimensional optimization problem in (28) to the closed-form solution in (29) that depends on a unique solution of a single monotonic nonlinear equation in (30) (which can be efficiently found using any of the known numerical techniques). The alternative representation in (32) is a 1-D convex optimization problem. It follows from (31) that the maximizing density f^* mimics the nominal one f_0, albeit with different constants in the outage and no outage sets

$$f^* = \begin{cases} \frac{\lambda^*}{\mu - 1} f_0, & \text{if } \mathbf{H} \in \mathcal{O} \\ \frac{\lambda^*}{\mu} f_0, & \text{if } \mathbf{H} \notin \mathcal{O} \end{cases} \tag{34}$$

where $\mathcal{O} = \{\mathbf{H} : \ell(\mathbf{H}) = 1\}$ is the outage set so that f^* is a scaled up version of f_0 in the outage set and scaled down otherwise[7] (to keep the normalization fixed), which is what one would intuitively expect to maximize the outage probability. Fig. 5 illustrates this behavior.

When P_{out} is considered as a function of d or ε, its properties mimic those for the other uncertainty class, which is summarized in the following.

---

[7]Note that $\lambda^* \leq \mu \leq 1 + \lambda^*$, with first equality iff $\varepsilon = 0$ and second one iff $\varepsilon = 1$.
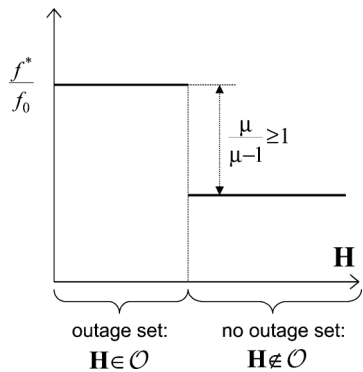
Fig. 5. Worst channel distribution over the nominal one. Note that the worst one is a piecewise constant scaled version of the nominal one, which follows the same tendency as for the other uncertainty set.
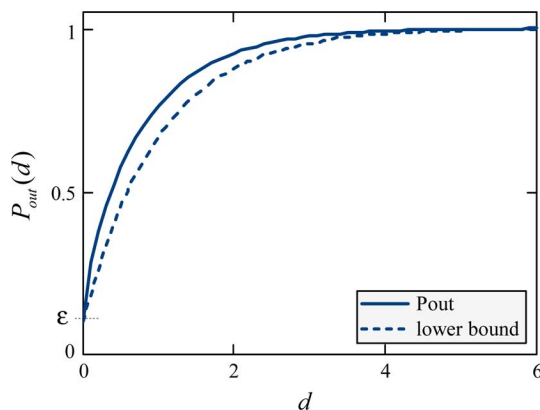

Fig. 6. Typical behavior of P_{out} as a function of d: strictly increasing, concave function. The lower bound in (25) possesses the same properties and closely approximates P_{out}.

Proposition 8: For a given input distribution $\rho(\mathbf{x})$, the compound outage probability $P_{\text{out}}(d)$ as a function of distance d has the following properties.
1) $P_{\text{out}}(d)$ is concave in $d \geq 0$.
2) $P_{\text{out}}(d)$ is a strictly increasing function of d

$$P_{\text{out}}(d_1) > P_{\text{out}}(d_2), \quad 0 \leq d_2 < d_1 < \infty \qquad (35)$$

unless $\varepsilon = 1$, in which case $P_{\text{out}}(d) = 1 \ \forall d$.
3) It is bounded as follows:

$$\varepsilon \leq P_{\text{out}}(d) \leq 1$$

and the bounds are tight: $P_{\text{out}}(d = 0) = \varepsilon, P_{\text{out}}(d = \infty) = 1$, or $P_{\text{out}}(d) = 1 \ \forall d$ iff $\varepsilon = 1$.

Proof: see the Appendix.   ∎

Fig. 6 shows this typical behavior of $P_{\text{out}}(d)$.

Proposition 9: The compound outage probability P_{out} takes on the extreme values as follows.
1) $P_{\text{out}} = 1$ iff $\varepsilon = 1$.
2) $P_{\text{out}} = 0$ iff $\varepsilon = 0$ and $d = 0$.
3) $P_{\text{out}} = \varepsilon$ iff $d = 0$ or $\varepsilon = 1$.

Proof: see the Appendix.   ∎

Proposition 10: $P_{\text{out}}(\varepsilon)$ is a concave, strictly increasing function of the nominal outage ε

$$P_{\text{out}}(\varepsilon_1) < P_{\text{out}}(\varepsilon_2), \quad 0 \leq \varepsilon_1 < \varepsilon_2 \leq 1 \qquad (36)$$

which is bounded as follows:

$$1 - e^{-d} \leq P_{\text{out}}(\varepsilon) \leq 1 \qquad (37)$$

and the bounds are tight: $P_{\text{out}}(\varepsilon = 0) = 1 - e^{-d}$ and $P_{\text{out}}(\varepsilon = 1) = 1$.
Proof: see the Appendix.   ∎

Comparing the properties of P_{out} in Propositions 9–10 to those in Propositions 1–3, one observes that most of the properties are the same so that the relative entropy asymmetry does not have much impact on the compound outage probability. One notable difference is the error floor effect in (37).

Let us now consider asymptotic regimes.

Proposition 11: The compound outage probability P_{out} in Theorem 3 behaves in the low nominal outage regime $\varepsilon \to 0$ and fixed d as follows:

$$P_{\text{out}} = 1 - e^{-d} + o(1). \qquad (38)$$

Proof: see the Appendix.   ∎

Note that (38) implies that $P_{\text{out}}(\varepsilon = 0) = 1 - e^{-d}$, i.e., the lower bound in (37) is tight. It also reaffirms the error floor effect discussed previously. Note that the error floor $1 - e^{-d}$ depends on the distance d only, all other details (e.g., the nominal distribution f_0) being irrelevant. Further analysis shows that the approximation in (38) is accurate, i.e., $P_{\text{out}} \approx 1 - e^{-d}$, provided that $\varepsilon \ll e^d - 1$. This is also consistent with the lower bound in (25) (third term is negligible under the latter condition).

Proposition 12: The compound outage probability P_{out} in Theorem 3 behaves in the small-uncertainty regime $d \to 0$ and fixed ε as follows:

$$P_{\text{out}} = \varepsilon + \sqrt{2\varepsilon(1 - \varepsilon)d} + o(\sqrt{d}) \qquad (39)$$

Proof: see the Appendix.   ∎

It is clear from (39) that $P_{\text{out}}(d = 0) = \varepsilon$, so that the lower bound in (25) is tight in this case. It is also obviously tight when $\varepsilon = 1$ and, as we have seen in Proposition 11, in the $\varepsilon = 0$ case. In fact, numerical analysis shows that this lower bound is a good approximation of P_{out} over the whole range of ε—see Fig. 7. Comparing (39) with (18), we conclude that the compound outage probability is the same for the $D(f\|f_0) \leq d$ and $D(f_0\|f) \leq d$ uncertainty sets in the low-uncertainty regime (i.e., the asymmetry of relative entropy does not have any effect here) while the same cannot be said about the low nominal outage regime (compare (38) with (17)).

It can be shown (via numerical analysis—see Fig. 7) that the approximation in (39) is accurate, i.e.,

$$P_{\text{out}} \approx \varepsilon + \sqrt{2\varepsilon(1 - \varepsilon)d} \qquad (40)$$
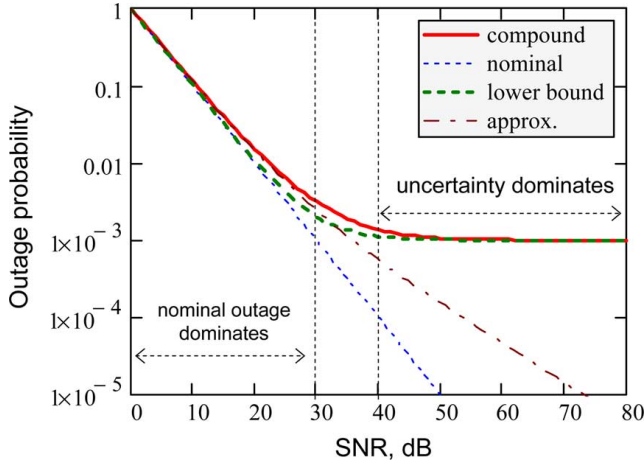
Fig. 7. Two-regime behavior of the compound outage probability and its lower bound in (25), the approximation in (40), and the nominal outage $\varepsilon = 1/\mathrm{SNR}$; $d = 10^{-3}$. Note the saturation (error floor) effect: the compound outage cannot be reduced below $10^{-3}(= d)$, in agreement with (26), does not matter how high the SNR (or how low the nominal outage is).

when $d < \varepsilon$ and that the effect of uncertainty is negligible, i.e., $P_{\mathrm{out}} \approx \varepsilon$, when $d \ll \varepsilon$, which parallels the conclusions made in Section III-B.

Fig. 7 illustrates the two-regime behavior of the compound outage probability and validates the aforementioned approximation. It also demonstrates that the lower bound in (25) is quite tight over the whole SNR (nominal outage) range and thus can be used instead of the true compound outage probability for design/analysis purposes to estimate the impact of channel distribution uncertainty. The uncertainty-induced error floor clearly visible in Fig. 7 resembles and supplements the error-floor effect in mobile wireless channels, which is due to temporal and frequency dispersions of the channel and multiuser interference. Unlike the former, the latter has been well studied in the past [33], [34]. In general, a sound system design should take into account both of them.

We would like to emphasize that the aforementioned results hold for an arbitrary nominal channel distribution f_0 (e.g., Rayleigh, Rician, Nakagami, log-normal, correlated and/or nonidentically distributed, etc.)—it enters into the compound outage probability only via the nominal outage ε, and also for arbitrary noise. When the compound outage probability is minimized over the input distribution

$$P_{\mathrm{out}}^* = \min_{\rho(\mathbf{x})} \max_{D(f_0\|f)\leq d} \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R\} \quad (41)$$

some of the aforementioned results (namely, (25)–(27), (38), and (39), (40) when $\varepsilon < 1/2$) apply directly via the substitution $\varepsilon \to \varepsilon^*$, i.e., the optimal input distribution is the same as for the nominal outage so that (23) holds. Furthermore, while optimizing the input distribution to reduce the nominal outage also reduces the compound outage probability in the low-uncertainty regime (i.e., in (39)), it has negligible impact in the uncertainty-dominated regime in (38) so that any reasonable input distribution (e.g., isotropic signaling in MIMO channels) will work equally well provided that $\varepsilon \ll e^d - 1$.

V. COMPOUND OUTAGE CAPACITY

Let us now consider the compound outage capacity, which extends the definition of outage capacity in (2) to a class of distributions. Consider first the case when the input distribution is given

$$C_\delta = \max\{R : \max_f \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}) < R\} \leq \delta\} \quad (42)$$

where f belongs to a distribution uncertainty class, $D(f\|f_0) \leq d$ or $D(f_0\|f) \leq d$.

It is clear that $C_\delta \leq C_{0\delta}$ in general, where the optimization over the input distribution is either used or not in both cases simultaneously. Using the compound outage probability results obtained previously, the compound outage capacity can be characterized more precisely.

Proposition 13: The compound outage capacity in the low-uncertainty regime ($d \to 0$ and fixed target outage probability δ) approaches the nominal outage capacity for both uncertainty sets, $D(f\|f_0) \leq d$ and $D(f_0\|f) \leq d$

$$C_\delta = C_{0\delta'} \quad (43)$$

where

$$\delta' = \delta - \sqrt{2d(1-\delta)\delta} + o(\sqrt{d}) \quad (44)$$

is an adjusted target outage probability.

Proof: Follows directly from (18) and (39). ∎

Note that the impact of channel distribution uncertainty on the outage capacity is a tighter requirement on the nominal target outage probability via the second term in (44), which is negligible when $\sqrt{2d} \ll \sqrt{\delta/(1-\delta)}$.

Proposition 14: The compound outage capacity for the $D(f_0\|f) \leq d$ uncertainty set satisfies the following inequality,

$$C_\delta = 0 \; \forall \delta < 1 - e^{-d}. \quad (45)$$

In particular, no transmission is possible at $\delta < d$: $C_\delta = 0 \; \forall \delta < d$.

Proof: Follows from (25) and from $1 - e^{-d} \leq d$. ∎

Proposition 15: The compound outage capacity for the $D(f_0\|f) \leq d$ uncertainty set in the $d \ll 1$ regime satisfies the following inequality:

$$C_\delta \leq C_{0(\delta-d)}$$

i.e., d serves as a measure of loss in the targeted outage probability due to the channel distribution uncertainty.

Proof: Follows from (25) and $1 - e^{-d} + e^{-d}\delta \approx d + \delta$ for $d \ll 1$. ∎

Proposition 16: The compound outage capacity in the low target outage regime $\delta \to 0$ and fixed d for the first uncertainty set $D(f\|f_0) \leq d$ is

$$C_\delta = C_{0\delta'} \quad (46)$$

where

$$\delta' = d \exp\{-d/\delta(1 + o(1))\} \quad (47)$$

is the nominal target outage probability.

Proof: Follows from (17) using the standard tools of asymptotic analysis [28], [29].

Note that, in this regime, the compound outage capacity at given target outage δ requires exponentially low (in $1/\delta$) nominal target outage probability δ', i.e., a very significant loss, which mimics the result in Proposition 14 (where the transition is sharp unlike the present case where it is gradual).

When optimization over the input distribution is done, the aforementioned results also apply with the substitution $\varepsilon \to \varepsilon^*$ and $C_{0\delta}$ meaning the optimized nominal outage capacity.

VI. Compound Outage Probability Via L_p Distance Constraint

Let us now consider the channel distribution uncertainty class of the form $\|f - f_0\|_p \leq d$, where $\|f\|_p = (\int |f|^p d\mathbf{H})^{1/p}$ is L_p norm, $p \geq 1$. It can be characterized as follows:

Proposition 17: The compound outage probability for the $\|f - f_0\|_p \leq d$ uncertainty class, $p \geq 1$, can be bounded as follows:

$$P_p = \max_{\|f-f_0\|_p \leq d} \Pr\{I(\mathbf{x}; \mathbf{y}|\mathbf{H}) < R\} \qquad (48)$$

$$\geq P_1 \qquad (49)$$

$$\geq \max[P_{\text{out}}^1(d^2/2), P_{\text{out}}^2(d^2/2)] \qquad (50)$$

$$\geq 1 - e^{-d^2/2} + e^{-d^2/2}\varepsilon \qquad (51)$$

$$\geq 1 - e^{-d^2/2} \qquad (52)$$

where $P_{\text{out}}^1(d)$ and $P_{\text{out}}^2(d)$ are the compound outage probabilities as functions of distance d for the $D(f\|f_0) \leq d$ and $D(f_0\|f) \leq d$ classes.

Proof: Equation (49) follows from the norm inequality $\|f - f_0\|_p \leq \|f - f_0\|_1$. Equation (50) follows from Pinsker inequality $D(f\|f_0) \geq \frac{1}{2}\|f - f_0\|_1^2$ [30] and the fact that $\|f - f_0\|_1 = \|f_0 - f\|_1$. Equations (51) and (52) follow from (25). ∎

Note that there is an error floor effect here as well: $P_p(\varepsilon = 0) \geq 1 - e^{-d^2/2} > 0 \ \forall d > 0$. The $p = 2$ case corresponds to the widely used MSE criterion (including channel estimation and measurements) so that there is an error floor for the MSE uncertainty class as well. Finally, when the compound outage probability is minimized over the input distribution, the same inequalities hold (where $P_{\text{out}}^{1(2)}$ and ε are also minimized), and the error floor is not affected by this. When d is small (so that the deviation of f from f_0 is also small), the L_p and relative entropy distances can be shown to have the same order of magnitude so that the corresponding compound outage probabilities will scale similarly.

VII. Conclusion

Compound outage probability and capacity of a class of fading MIMO channels with partial channel distribution information have been introduced and studied. These concepts generalize well known and widely used concepts of outage probability and capacity of fading channels with completely known distribution to the case where only partial knowledge of distribution is available. Relative entropy distance is used as a measure of uncertainty, which is shown to be related directly to the compound outage probability. Since relative entropy distance is not symmetric, two uncertainty classes are considered and worst case fading distributions are identified for both. A number of properties, bounds, and approximations of the compound outage probability are given. The input distribution optimized for the nominal outage probability is shown to be also optimal for the compound one. The nominal fading distribution enters into the compound outage probability only via the nominal outage probability, all other details being irrelevant, i.e., two different nominal distributions having the same nominal outage will also produce the same compound outage probability. Behavior of the compound outage probability reveals two distinct regimes: uncertainty-dominated regime and nominal outage-dominated one. While increasing the average SNR or optimizing the input distribution to reduce the outage probability is effective for the latter, it has only negligible effect in the former, which immediately suggest a design alternative (via reducing uncertainty rather than increasing the SNR or optimizing the input distribution). All these results are general as they hold for arbitrary nominal fading distribution and also for arbitrary noise (i.e., not only Gaussian).

Appendix

A. Proof of Theorem 1

The problem in (5) can be presented in this form[8]:

$$\max_f \int \ell(\mathbf{H})f d\mathbf{H}, \text{s.t.} \ \int f \ln \frac{f}{f_0} d\mathbf{H} \leq d, \int f d\mathbf{H} = 1. \qquad (53)$$

The problem is clearly convex (since the objective is linear and the constraint set is convex). Furthermore, strong duality holds (i.e., the duality gap is zero), since Slater condition (see, e.g., [26]) is satisfied. Therefore, the Karush–Kuhn–Tucker conditions are sufficient for optimality [26]. The Lagrangian of this problem is

$$L = \int \ell(\mathbf{H})f d\mathbf{H} - s \left(\int f \ln \frac{f}{f_0} d\mathbf{H} - d \right) - \mu \left(\int f d\mathbf{H} - 1 \right) \qquad (54)$$

and, taking the variational derivative (see, e.g., [27]) of L with respect to f, one obtains the KKT conditions

$$\ell(\mathbf{H}) - s \left(\ln \frac{f}{f_0} + 1 \right) - \mu = 0 \qquad (55)$$

$$\int f d\mathbf{H} - 1 = 0 \qquad (56)$$

$$s \left(\int f \ln \frac{f}{f_0} d\mathbf{H} - d \right) = 0 \qquad (57)$$

$$s \geq 0. \qquad (58)$$

The complementary slackness condition (57) implies that the second term is zero, i.e., the optimum is achieved on the boundary, if $s > 0$. In this case, using (55) and (56), one obtains, after some manipulations

$$f^* = f_0 \frac{e^{\ell(\mathbf{H})/s^*}}{1 + \varepsilon(e^{1/s^*} - 1)} \qquad (59)$$

---

[8]An additional constraint $f \geq 0$ is not included since our solution explicitly satisfies it.

where f^* and s^* are the solutions of (55)–(58), from which (8) follows. Integrating f^*, one obtains (9). s^* is found from (57), which results in (10). The latter equation has a unique solution $s^* > 0$ if $d < \ln(1/\varepsilon)$, since $0 \le F(s) < \ln(1/\varepsilon)$ and $F(s)' < 0$ if $s > 0$, where $F(s)$ is the left-hand side of (10). On the other hand, $P_{\text{out}} = 1$ when $d \ge \ln(1/\varepsilon)$, which is achieved by f^* in (11) and also $s^* = 0$. When $d > \ln(1/\varepsilon)$, maximizing density f^* is not unique.

To obtain (6), note that, due to zero duality gap, the solution to (53) equals to that of the dual problem

$$P_{\text{out}} = \min_{s \ge 0} L(s) \tag{60}$$

where $L(s)$ is the dual function

$$L(s) = \max_f \{L\} \text{ s.t. (56)}$$
$$= s\ln(1 + (e^{1/s} - 1)\varepsilon) + sd \tag{61}$$

and μ was eliminated using (56). Since $L(s)$ is the dual function, it is convex. Furthermore, when $0 < \varepsilon < 1$

$$\frac{d^2 L(s)}{ds^2} = \frac{\varepsilon(1 - \varepsilon)e^{1/s}}{s^2(1 + (e^{1/s} - 1)\varepsilon)^2} > 0$$

for $0 < s < \infty$, i.e., strictly convex so that the problem in (60) has a unique solution. Since $P_{\text{out}} = \varepsilon$ when $\varepsilon = 1, 0$ (see Proposition 2), the problem in (60) has always a unique solution, which can be found from $dL(s)/ds = 0$ using any known numerical algorithm. The tools of asymptotic analysis (e.g., [28] and [29]) allow one to obtain a number of approximations.

B. Proof of Proposition 1

1) To prove item 1, note that P_{out} in (6) is a pointwise minimum of a set of affine functions of d (indexed by s) and therefore is concave (see, e.g., [26]).
2) Using Lyapounov inequality

$$(\mathbb{E}|x|^p)^{1/p} \le (\mathbb{E}|x|^q)^{1/q}$$

where $0 < p \le q$ and $\mathbb{E}$ denotes expectation, for random variable $x = e^{\ell(\mathbf{H})}$ and $p = 1/s_1, q = 1/s_2, 0 \le s_2 \le s_1$, one obtains

$$\left(\mathbb{E}e^{\ell(\mathbf{H})/s_1}\right)^{s_1} \le \left(\mathbb{E}e^{\ell(\mathbf{H})/s_2}\right)^{s_2}$$

and taking log

$$s_1 \ln \mathbb{E}e^{\ell(\mathbf{H})/s_1} \le s_2 \ln \mathbb{E}e^{\ell(\mathbf{H})/s_2}$$

so that $F(s) = s \ln \mathbb{E}e^{\ell(\mathbf{H})/s} = s\ln(1 + (e^{1/s} - 1)\varepsilon)$ is a nonincreasing function of s. Therefore

$$P_{\text{out}}(d = 0) = \min_{s \ge 0} F(s) = \lim_{s \to \infty} F(s) = \varepsilon. \tag{62}$$

3) Define $L_i(s) = s\ln(1 + (e^{1/s} - 1)\varepsilon) + sd_i, i = 1, 2, d_1 < d_2$. Then, $L_1(s) \le L_2(s)$ with equality iff $s = 0$, in which case $L_1(s) = L_2(s) = 1$ or 0 (if $\varepsilon > 0$ or $= 0$). Taking $\min_{s \ge 0}$ of both sides, one obtains $P_{\text{out}}(d_1) \le P_{\text{out}}(d_2)$

with equality iff $P_{\text{out}}(d_1) = P_{\text{out}}(d_2) = 1, 0$. The latter equality holds when $d \ge \ln(1/\varepsilon)$ or $\varepsilon = 0$.

C. Proof of Proposition 2

1) Using (59)

$$P_{\text{out}} = \frac{e^{1/s^*}\varepsilon}{1 + \varepsilon(e^{1/s^*} - 1)} \tag{63}$$

from which it follows that $P_{\text{out}} = 1$ iff $\varepsilon = 1$ or $s^* = 0$. In both cases, $d \ge \ln(1/\varepsilon)$.
2) Item 2 is proved in the same way as item 1 earlier.
3) The inequality is due to the fact that the distribution uncertainty set always includes the nominal distribution. The "if" part follows from aforementioned items 1 and 2 and from item 2 of Proposition 1. The "only if" part is verified using (63).

D. Proof of Proposition 3

It mimics the proof of item 3 in Proposition 1. Define $L_i(s) = s\ln(1 + (e^{1/s} - 1)\varepsilon_i) + sd, i = 1, 2, \varepsilon_1 < \varepsilon_2$. Then, $L_1(s) \le L_2(s)$ with equality iff $s = 0$, in which case $L_1(s) = L_2(s) = 1$. Taking $\min_{s \ge 0}$ of both sides, one obtains $P_{\text{out}}(\varepsilon_1) \le P_{\text{out}}(\varepsilon_2)$ with equality iff $P_{\text{out}}(\varepsilon_1) = P_{\text{out}}(\varepsilon_2) = 1$, which is possible iff $d \ge \ln(1/\varepsilon)$. The concavity of $P_{\text{out}}(\varepsilon)$ follows from the fact that it is a pointwise minimum of a set of concave functions of ε (indexed by s), see (60).

E. Proof of Proposition 4

The lower bound was proved earlier. The upper bound follows from the following:

$$P_{\text{out}} = \min_{s \ge 0}[s\ln(1 + (e^{1/s} - 1)\varepsilon) + sd]$$
$$\le \ln(1 + (e - 1)\varepsilon) + d$$
$$\le (e - 1)\varepsilon + d$$

where the first inequality is obtained by setting $s = 1$ and the second one follows from $\ln(1 + x) \le x$ for $x \ge 0$.

F. Proof of Proposition 5

Since $L(s)$ in (13) is convex, a unique minimum in (6) can be found by setting the derivative to zero, $dL(s)/ds = 0$, which can be expressed as

$$d + \ln(1 + y) = \frac{y + \varepsilon}{1 + y}\ln\left(1 + \frac{y}{\varepsilon}\right) \tag{64}$$

where $y = \varepsilon(e^{1/s} - 1)$. It is straightforward to see that $y \to 0$ as $\varepsilon \to 0$ so that (64) can be transformed to

$$\frac{d}{\varepsilon} + o\left(\frac{d}{\varepsilon}\right) = \left(1 + \frac{y}{\varepsilon}\right)\ln\left(1 + \frac{y}{\varepsilon}\right) \tag{65}$$

where we have used $\ln(1 + y) = y + o(y)$ and $1/(1 + y) = 1 - y + o(y)$. Using the techniques of asymptotic analysis (see, e.g., [28] and [29]), the following equation

$$x\ln x = u \tag{66}$$

has the solution when $u \to \infty$

$$x = \frac{u}{\ln u - \ln \ln u} \left(1 + o\left(\frac{1}{\ln u}\right)\right). \tag{67}$$

Setting $x = 1 + y/\varepsilon = e^{1/s}$ and $u = (1 + o(1))d/\varepsilon$, one obtains, after some lengthy but straightforward manipulations, the solution y^* of (65) and corresponding s^*

$$y^* = \frac{d}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}} (1 + o(1))$$

$$s^* = \frac{1}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}} (1 + o(1))$$

so that after some further manipulations

$$P_{\text{out}} = L(s^*) = \frac{d}{\ln \frac{d}{\varepsilon} - \ln \ln \frac{d}{\varepsilon}} (1 + o(1)). \tag{68}$$

G. Proof of Proposition 6

In the $d \to 0$ and fixed ε regime, the minimizer s^* in (6) is $s^* \to \infty$ (as follows from (62)) so that $y^*/\varepsilon = e^{1/s^*} - 1 \to 0$ and (64) can be transformed to

$$\frac{1}{2} \left(\frac{y}{\varepsilon}\right)^2 (1 - \varepsilon) + o(y^2) = \frac{d}{\varepsilon}$$

from which one obtains

$$y^* = \sqrt{\frac{2\varepsilon d}{1 - \varepsilon}} (1 + o(1))$$

so that

$$s^* = \sqrt{\frac{\varepsilon(1 - \varepsilon)}{2d}} (1 + o(1))$$

and

$$P_{\text{out}} = \varepsilon + \sqrt{2d(1 - \varepsilon)\varepsilon} + o(\sqrt{d}). \tag{69}$$

Further analysis shows that the aforementioned approximation (without $o(\sqrt{d})$ term) is accurate when $d \ll \varepsilon < 1$.

H. Proof of Theorem 3

This proof follows along the same lines as that of Theorem 1, which is summarized in the following. The optimization problem is

$$\max_f \int \ell(\mathbf{H}) f d\mathbf{H}, \text{ s.t.} \quad \int f_0 \ln \frac{f_0}{f} d\mathbf{H} \le d \tag{70}$$

$$\int f d\mathbf{H} = 1.$$

The problem is also convex and strong duality holds so that the KKT conditions are sufficient for optimality. The Lagrangian is

$$L = \int \ell(\mathbf{H}) f d\mathbf{H} - \lambda \left(\int f_0 \ln \frac{f_0}{f} d\mathbf{H} - d\right) - \mu \left(\int f d\mathbf{H} - 1\right) \tag{71}$$

and the KKT conditions are

$$\ell(\mathbf{H}) + \lambda \frac{f_0}{f} - \mu = 0 \tag{72}$$

$$\int f d\mathbf{H} - 1 = 0 \tag{73}$$

$$\lambda \left(\int f_0 \ln \frac{f_0}{f} d\mathbf{H} - d\right) = 0 \tag{74}$$

$$\lambda \ge 0. \tag{75}$$

After some manipulations, one obtains

$$f^* = \frac{\lambda^* f_0}{\mu - \ell(\mathbf{H})} \tag{76}$$

where f^* and λ^* are the solutions of (72)–(75). λ^* is found from (74)

$$\lambda^* = e^{-d} (\mu - 1)^\varepsilon \mu^{1-\varepsilon} \tag{77}$$

and μ is found from (73)

$$\frac{\lambda^* \varepsilon}{\mu - 1} + \frac{\lambda^* (1 - \varepsilon)}{\mu} = 1 \tag{78}$$

which can be transformed, after some manipulations, to (30).

The alternative characterization of P_{out} follows from the dual problem

$$P_{\text{out}} = \min_{\lambda \ge 0} L(\lambda) \tag{79}$$

where $L(\lambda)$ is the dual function

$$L(\lambda) = \min_f \{L\} \text{ s.t. } (73)$$
$$= \lambda \varepsilon \left(\frac{1}{\mu - 1} + \ln \frac{\mu}{\mu - 1}\right) + \lambda \left(d - \ln \frac{\mu}{\lambda}\right) \tag{80}$$

and μ is found from (78)

$$\mu = \frac{1}{2} \left(1 + \lambda + \sqrt{(1 - \lambda)^2 + 4\lambda \varepsilon}\right). \tag{81}$$

I. Proof of Proposition 8

Consider the following function:

$$F(\lambda, d) = \lambda \varepsilon \left(\frac{1}{\mu - 1} + \ln \frac{\mu}{\mu - 1}\right) + \lambda \left(d - \ln \frac{\mu}{\lambda}\right) \tag{82}$$

so that

$$P_{\text{out}}(d) = \min_{\lambda \ge 0} F(\lambda, d) \tag{83}$$

i.e., $P_{\text{out}}(d)$ is a pointwise minimum of affine functions $F(\lambda, d)$ of d (indexed by λ) and thus concave [26].

To prove 2), observe that

$$F(\lambda, d_1) \ge F(\lambda, d_2) \text{ if } d_1 > d_2 \tag{84}$$

with the equality iff $\lambda = 0$, so that

$$P_{\text{out}}(d_1) = F(\lambda_1^*, d_1) \ge F(\lambda_1^*, d_2) \ge \min_{\lambda \ge 0} F(\lambda, d_2) \tag{85}$$
$$= P_{\text{out}}(d_2)$$

where $\lambda_i^* = \arg\min_{\lambda \geq 0} F(\lambda, d_i)$. If $P_{\text{out}}(d_1) = P_{\text{out}}(d_2)$, then all inequalities in (85) hold with equality, which implies $\lambda_1^* = 0$ and thus $\varepsilon = 1$. The reverse implication is obvious.

The bounds in 3) are obvious. To prove achievability conditions, note that $F(\lambda \to \infty, d = 0) = \varepsilon$, which proves $P_{\text{out}}(d = 0) = \varepsilon$, and $P_{\text{out}}(d \to \infty) = 1$ follows from (26). ∎

J. Proof of Proposition 9

$\varepsilon = 1 \to P_{\text{out}} = 1$ is obvious. The other way follows from (29) after some manipulations.

$\varepsilon, d = 0 \to P_{\text{out}} = 0$ follows from Proposition 8. The other way follows from (25).

$d = 0 \to P_{\text{out}} = \varepsilon$ is in 3) of Proposition 8. The other way follows from 2) of the same Proposition. The case $\varepsilon = 1$ is obvious. ∎

K. Proof of Proposition 10

Consider the following function:

$$F(\lambda, \varepsilon) = \lambda\varepsilon\left(\frac{1}{\mu - 1} + \ln\frac{\mu}{\mu - 1}\right) + \lambda\left(d - \ln\frac{\mu}{\lambda}\right) \quad (86)$$

so that

$$P_{\text{out}}(\varepsilon) = \min_{\lambda \geq 0} F(\lambda, \varepsilon) \quad (87)$$

i.e., $P_{\text{out}}(\varepsilon)$ is a pointwise minimum of affine functions $F(\lambda, \varepsilon)$ of ε (indexed by λ) and thus concave [26].

To prove that it is strictly increasing, follow the same steps as in the proof of Proposition 8.

The lower bound in (37) is in (26) and the upper bound is obvious. $P_{\text{out}}(\varepsilon = 1) = 1$ follows from (25). $P_{\text{out}}(\varepsilon = 0) = 1 - e^{-d}$ follows from Proposition 11. ∎

L. Proof of Proposition 11

It is straightforward to see that the left-hand side of (30) is a strictly increasing function of μ, $\forall \mu \geq 0$, $\forall \varepsilon \in (0, 1)$, which takes values in $[0, 1]$ so that a solution is unique. It also follows that $\mu \to 1$ as $\varepsilon \to 0$ so that expanding $\mu(\varepsilon)$ in Taylor series, $\mu = 1 + a_1\varepsilon + o(\varepsilon)$, one obtains from (30)

$$\ln\frac{a_1}{1 + a_1} + o(1) = -d \quad (88)$$

where we have used $\ln\mu = a_1\varepsilon + o(\varepsilon)$, from which it follows that

$$\mu = 1 + \frac{\varepsilon}{e^d - 1} + o(\varepsilon) \quad (89)$$

and also $\lambda = e^{-d}(1 + o(1))$ so that

$$P_{\text{out}} = \frac{\lambda^*}{\mu - 1}\varepsilon = 1 - e^{-d} + o(1). \quad (90)$$

Equation (89) suggests that the approximation in (90) (without $o(1)$ term) is accurate when $\varepsilon \ll e^d - 1$, which is also confirmed by numerical analysis.

M. Proof of Proposition 12

It follows from (30) that $\mu \to \infty$ as $d \to 0$, so we use the expansion

$$\ln\left(1 - \frac{1}{\mu}\right) = -\frac{1}{\mu} + \frac{1}{2\mu^2} + o\left(\frac{1}{\mu^2}\right) \quad (91)$$

to transform (30) to

$$\frac{\varepsilon(1 - \varepsilon)}{2\mu^2} + o\left(\frac{1}{\mu^2}\right) = d \quad (92)$$

from which it follows that

$$\mu = \sqrt{\frac{\varepsilon(1 - \varepsilon)}{2d}}(1 + o(1)) \quad (93)$$

and also that $\lambda^* = \mu(1 + o(1))$. Using these, one finally obtains

$$P_{\text{out}} = \frac{\lambda^*}{\mu - 1}\varepsilon = \varepsilon + \sqrt{2\varepsilon(1 - \varepsilon)d} + o(\sqrt{d}). \quad (94)$$

∎

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[2] G. J. Foshini 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.

[3] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[4] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[5] R. L. Dobrushin, "Optimum information transmission through a channel with unknown parameters," *Radiotechnika I Electronika*, vol. 4, no. 12, pp. 1951–1956, Dec. 1959.

[6] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, pp. 1229–1241, 1959.

[7] W. L. Root and P. P. Varaya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[8] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[9] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[10] H. Boche and E. A. Jorswieck, "Outage probability of multiple antenna systems: Optimal transmission and impact of correlation," in *Proc. IEEE Int. Zurich Semin. Commun.*, Feb. 2004, pp. 116–119.

[11] E. A. Jorswieck and H. Boche, "Outage probability in multiple antenna systems," *Eur. Trans. Telecommun.*, vol. 18, no. 3, pp. 217–233, Jun. 2006.

[12] G. Caire and K. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[13] 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.

[14] S. Loyka and C. D. Charalambous, "On the capacity of a class of MIMO channels subject to normed uncertainty," presented at the IEEE Int. Symp. Inf. Theory, Toronto, ON, Canada, Jul. 2008.

[15] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[16] 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.

[17] C. Kose and R. D. Wesel, "Universal space-time trellis codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2717–2727, Oct. 2003.

[18] S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[19] N. Prasad and M. K. Varanasi, "Outage theorems for MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5284–5296, Dec. 2006.

[20] M. Effros, A. Goldsmith, and Y. Liang, "Generalizing capacity: New definitions and capacity theorems for composite channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3069–3087, Jul. 2010.

[21] C. D. Charalambous, D. Stojan, and C. Constantinou, "Capacity of the class of MIMO channels with incomplete CDI: Properties of mutual information for a class of channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3725–3734, Aug. 2009.

[22] C. D. Charalambous and F. Rezaei, "Stochastic uncertain system subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games," *IEEE Trans. Autom. Control*, vol. 52, no. 4, pp. 647–663, Apr. 2007.

[23] A. H. El-Sawy and V. D. Vandelinde, "Robust detection of known signals," *IEEE Trans. Inf. Theory*, vol. 23, no. 6, pp. 722–727, Nov. 1977.

[24] O. Hossjer and M. Mettiji, "Robust multiple classification of known signals in additive noise—An asymptotic weak signal approach," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 594–608, Mar. 1993.

[25] G. Shevlyakov and K. Kim, "Robust minimax detection of a weak signal in noise with a bounded variance and density value at the center of symmetry," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1206–1211, Mar. 2006.

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

[27] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Moscow, Russia: GIFML, 1961.

[28] M. A. Efgrafov, *Asymptotic Expansions and Entire Functions*. Moscow, Russia: GITTL, 1957.

[29] F. W. J. Olver, *Asymptotics and Special Functions*. Moscow, Russia: Nauka, 1990.

[30] A. A. Fedotov, P. Harremoes, and F. Topsoe, "Refinements of Pinsker's inequality," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1491–1498, Jun. 2003.

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

[32] S. Verdu, "Relative entropy," presented at the Neural Inf. Process. Syst. Conf., Vancouver, BC, Canada, Dec. 2009.

[33] W. C. Jakes Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.

[34] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Hoboken, NJ: Prentice-Hall, 2002.

Ioanna Ioannou received the B.S. degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Greece, in 2006. During her studies, she has concentrated in the general area of telecommunications. Since 2006, she has been a doctoral candidate and a researcher at the University of Cyprus supported by a scholarship from Cyprus State Scholarship Foundation. Her research interests include wireless communications, MIMO systems and information theory.

Charalambos D. Charalambous 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 the IEEE COMMUNICATIONS LETTERS and for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Sergey Loyka 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 Superieure, 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.