Conference Paper

The Operational Capacity of Compound Uniformly-Ergodic Fading Channels

Author(s):
Loyka, Sergey; Charalambous, Charalambos D.

Publication Date:
2018-02

Permanent Link:
https://doi.org/10.3929/ethz-b-000245068

The Operational Capacity of Compound Uniformly-Ergodic Fading Channels

Sergey Loyka, Charalambos D. Charalambous

***Abstract*—The impact of distribution uncertainty on the performance of compound fading channels is studied. To this end, a new class of fading channels, termed "uniformly-ergodic", is introduced and several of its equivalent (easy-to-use) characterizations and examples are presented. A single-letter expression for the operational capacity of this class of channels is obtained under the full Rx CSI using the recent general formula for compound channel capacity and the information spectrum approach. The saddle-point property is established, whereby the compound channel capacity is the same as the worst-case capacity so that the full knowledge of the fading distribution at the transmitter does not increase the capacity of this class of channels.**

I. Introduction

The impact of channel uncertainty on its capacity and system design has been extensively studied since late 1950s, see [1] for an extensive literature review up to late 1990s and [2] for a more recent albeit brief review. A widely-accepted approach to the channel uncertainty problem is via the compound channel model, where the channel is assumed to be unknown but is known to belong to a certain class (set) of channels [1]. Since channel estimation is done at the receiver (Rx) and then send back to the transmitter (Tx) via a limited feedback link, many studies concentrate on limited channel state information (CSI) available at the Tx end and assume full CSI at the Rx end.

Fading represents one of the most significant obstacles to reliable wireless communications and respective system design, affecting its performance in a dramatic way [3]. It also makes channel estimation a challenging problem, due to significant channel dynamics, low SNR, limitations of a feedback link etc. In this context, incomplete CSI can also be modelled by assuming that the channel is not known but its distribution is known, the so-called channel distribution information (CDI) [3]. However, complete knowledge of the CDI, which is essential for capacity evaluation and system design, can be questioned on the same grounds as complete CSI: when only a limited sample set is available (always a practicality), the CDI can be obtained with limited accuracy only (especially at the distribution tails); limited feedback link dictates quantization of the estimated CDI before transmission, thus introducing the quantization noise; presence of noise and channel dynamics makes any estimate inaccurate to a certain degree. This motivates us to study the impact of inaccurate CDI on system performance and design.

For quasi-static (and hence non-ergodic) fading channels, the key performance metrics are outage probability/capacity [3]. The impact of CDI uncertainty on these metrics was studied in [5]. In particular, it was shown that the CDI uncertainty induces an error floor effect: increasing the SNR over a certain threshold does not reduce the outage probability and the error floor is determined by the size of the uncertainty set.

For ergodic-fading channels (where the fading process is allowed to have memory provided it is still ergodic), a single-letter capacity expression has been established in [4] under complete CDI at the Tx end and full CSI at the Rx end. However, the standard results on ergodic capacity [3][4] do not apply when only incomplete CDI is available and hence certain performance has to be demonstrated for the whole class of fading distributions, not just for a single one, and, in addition, the Tx does not know the true fading distribution and hence cannot design a codebook using this knowledge (as was done in [3][4]).

The information capacity of ergodic-fading channels under CDI uncertainty, formally defined via the standard max-min expression (of ergodic mutual information), has been studied in [6]. However, its operational meaning as the largest achievable rate subject to the reliability criterion has not been established so it is not clear whether this quantity has practical relevance (while the max-min MI is often the compound channel capacity, it is not always the case [1]). The main difficulty was the lack of general-enough tools for compound channels that would allow one to incorporate CDI uncertainty. Such tools have been recently presented in [7], which are based on the information spectrum approach of Verdu and Han [8][9]. Using these tools, we prove here that the above "max-min" information capacity has the operational meaning of maximum achievable rate under the CDI uncertainty. This is accomplished by introducing a new concept of "uniformly-ergodic compound channel" and applying the general formula for compound channel capacity in [7] to such channel, which results in a compact single-letter expression for the capacity of uniformly-ergodic compound channels, subject to the sets of feasible input and fading distributions being convex but otherwise arbitrary. To facilitate applications, we develop several equivalent and easy-to-use criteria for compound channels to be uniformly-ergodic and give some practically-relevant examples. Apart from the single-letter capacity expression, the key contribution of this paper is the recognition of importance of uniform ergodicity for compound fading channels.

II. Channel Model

To isolate and study the impact of CDI uncertainty, we adopt the conditionally-memoryless channel model of [4], where the

S. Loyka is with the School of EECS, University of Ottawa, Canada, sergey.loyka@ieee.org

C.D. Charalambous is with the ECE Department, University of Cyprus, Cyprus, chadcha@ucy.ac.cy

channel is memoryless conditioned on its state sequence s^n:

$$p(y^n|x^n, s^n) = \prod_{i=1}^{n} p(y_i|x_i, s_i) \quad (1)$$

where x, y, s are the input, output and state; n is the block-length, $x^n = \{x_1, .., x_n\}$ and likewise for y^n, s^n; capitals denote random variables while lower-case letters - their realizations or arguments. The random state sequence $S^n = \{S_1, .., S_n\}$ represents the fading process and is assumed to be stationary and ergodic but not necessarily memoryless - it can have memory provided that the ergodicity assumption still holds, so that a correlated fading process is allowed (see Section III for details). Assuming that the receiver has the full CSI, i.e. the state sequence s^n, but the Tx knows only the fading distribution (i.e. the full Tx CDI, see e.g. [3] for a detailed motivation of this assumption), a single-letter ergodic capacity $C[f]$ was obtained in [4] for this ergodic-fading channel:

$$C[f] = \sup_{p(x)} I(X;Y|f) \quad (2)$$

where $I(X;Y|f)$ is the ergodic mutual information under fading distribution $f(s)$ and i.i.d. input:

$$I(X;Y|f) = \sum_{s} f(s) I(X;Y|s) \quad (3)$$

and $I(X;Y|s)$ is the MI under channel state s, and where all alphabets are assumed to be discrete and finite; under some regularity assumptions, this can also be extended to infinite and continuous alphabets. The optimal input is i.i.d. [4]. The maximization over the input distribution $p(x)$ is subject to a suitable constraint, e.g. maximum or average power, and is independent of channel state s (due to no Tx CSI) but may depend on the fading distribution f. We emphasize, for future use, that the ergodic MI $I(X;Y|f)$ as well as the capacity $C[f]$ also depend on the fading distribution. Note that even though the fading process is allowed to have memory (i.e. does not have to be i.i.d.), the ergodic MI as well as the capacity depend only on the marginal fading distribution $f(s)$, not on the joint one (which is ultimately due to the conditionally-memoryless nature of the channel). This makes the analysis much simpler.

Ergodic channel model is suitable in scenarios with significant channel dynamics so that a single codeword spans many different channel realizations and an encoder can take advantage of it [3]. However, in many practical scenarios, complete knowledge of $f(s)$ may be not available at the transmitter, due to e.g.

- inaccuracy in estimating $f(s)$ at the receiver (due to finite sample size or estimation noise);
- limited/quantized feedback link (quantization noise);
- outdated estimate,

so that the true fading distribution f differs from its estimate f_0 available at the transmitter. To model this fading distribution uncertainty (inaccuracy), we consider the scenario where the transmitter has only partial CDI. Namely, it knows that $f \in \mathcal{F}_1$, where $\mathcal{F}_1$ is the uncertainty set known to the Tx, which is further assumed to be convex; the state sequence S^n is not available to the Tx, while the Rx has the full CSI, i.e. the sequence S^n. This forms a compound channel model where the fading distribution f is a (meta) state. Its respective compound channel capacity is defined in the standard way as the maximum achievable rate subject to the reliability criterion, where the error probability converges to zero uniformly over the whole uncertainty set and where the codebooks are independent of the actual channel state s or its fading distribution f (see e.g. [1] for more details and formal definitions).

The following section presents key definitions and properties of ergodic-fading channels in the compound setting, i.e. when the fading distribution is not known exactly.

III. Compound Ergodic-Fading Channels

In order to simplify notations, we use f to refer to marginal $f(s)$ as well as joint distribution $f(s^n)$, which should be clear from the context. If $\{s_1, s_2, ...\}$ is an ergodic process, we call the joint distribution $f(s^n)$ ergodic as well, with understanding that ergodicity reveals itself as $n \to \infty$. $\mathcal{F}$ denotes a set of joint distributions while $\mathcal{F}_1$ – a set of respective marginal distributions. Since the joint fading distribution completely characterises fading channel (in combination with (1)), we will refer to $\mathcal{F}$ as "channels" as well. $\mathbb{E}\{\cdot\}$ denotes expectation over relevant random variables.

We begin with a standard definition of an ergodic (discrete-time) random process [11]-[14].

Definition 1. *A stationary random process $\{S_1, S_2, ...\}$ is (mean-) ergodic if, for any $g(s)$ such that $\mathbb{E}\{|g(S)|\} < \infty$,*

$$\frac{1}{n} \sum_{i=1}^{n} g(S_i) \to \mathbb{E}\{g(S)\} \quad (4)$$

as $n \to \infty$, where the convergence is either in mean-square, or in probability, or with probability 1.

A few modifications to this definition are in order to accommodate the compound channel setting here: (i) we need to consider a class of distributions $\mathcal{F}$ rather than a single distribution f, (ii) there is no need to consider all absolute-integrable/summable functions $g(s)$; instead, we need to consider only the mutual information $I(X;Y|s)$ under channel state s and i.i.d. input as a function of interest; (iii) we will use convergence in the mean-square sense (since it is needed in the proof of coding theorem); this implies convergence in probability but the converse is not true in general; however, when $I(X;Y|s)$ is uniformly bounded (e.g. when either input or output alphabet is of finite cardinality), they are equivalent.

The following definition extends the standard definition of ergodic channels to the compound setting.

Definition 2. *A class of stationary fading channels $\mathcal{F}$ is uniformly (mean-) ergodic if it is ergodic for each $f \in \mathcal{F}$ under i.i.d. input, i.e. as $n \to \infty$,*

$$\frac{1}{n} \sum_{i=1}^{n} I(X;Y|S_i) \to I(X;Y|f) \quad (5)$$

where the convergence is in the mean-square sense, and, in addition, it is uniform over the whole class $\mathcal{F}$, i.e. $\forall \delta > 0$ $\exists n_0(\delta)$ *such that* $\forall n > n_0(\delta)$

$$\sigma_{nf}^2 \triangleq \mathbb{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} I(X;Y|S_i) - I(X;Y|f)\right)^2\right\} < \delta \quad (6)$$

where n_0 *depends on* δ *but not* f*;* δ *is also independent of* f*.*

It is straightforward to verify that (6) is equivalent to

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \sigma_{nf}^2 = 0 \quad (7)$$

(note that $\lim$ and $\sup$ cannot be swapped). It should be emphasized that the uniform ergodicity property in (5), (6) as well as the ergodicity property in (4) depend on the joint distribution $f(s^n)$, not just marginal $f(s)$, even though the limits depend only on the marginal.

To facilitate applications, we give equivalent criteria of the uniform ergodicity and provide several examples. To simplify notations, let $I_{s_i} = I(X;Y|S_i)$ and $I_f = I(X;Y|f)$ (all under i.i.d. inputs) and let c_{ijf} be the covariance of I_{s_i} and I_{s_j} under fading distribution f,

$$c_{ijf} \triangleq \mathbb{E}\{(I_{s_i} - I_f)(I_{s_j} - I_f)\} \quad (8)$$

Since the channel is stationary, c_{ijf} depends only on $i-j$: $c_{ijf} = c_{(i-j)f}$. We assume below that the variance is uniformly bounded:

$$c_{0f} \leq A < \infty \ \forall f \in \mathcal{F} \quad (9)$$

(note that A is independent of f), which is equivalent to $\sup_{f\in\mathcal{F}} c_{0f} < \infty$. This is the case when e.g. the alphabets are discrete (see e.g. [9]) and also holds in many cases for continuous alphabets as well (e.g. Gaussian).

The following proposition is an extension of Slutsky's Theorem (see e.g. [11][13][14]) to the compound setting here.

Proposition 1. *A compound stationary-fading channel is uniformly mean-ergodic iff*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{l=0}^{n-1}\left(1-\frac{l}{n}\right) c_{lf} = 0 \quad (10)$$

Equivalently,

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{l=0}^{n-1} c_{lf}\right| = 0 \quad (11)$$

Proof. See Appendix. □

The following condition, which follows from (11), is easier to verify in many cases.

Corollary 1.1. *The condition in* (11) *holds if* $c_{lf} \to 0$ *for each* f *as* $l \to \infty$ *and the convergence is uniform over the set* $\mathcal{F}$*:*

$$\lim_{l\to\infty} \sup_{f\in\mathcal{F}} |c_{lf}| = 0 \quad (12)$$

This condition essentially means that the channel is asymptotically uncorrelated for any possible fading distribution and also uniformly so over the uncertainty set $\mathcal{F}$.

Many special cases can be derived from (12).

1. Assume that the fading process is i.i.d. for each f, in which case $c_{lf} = 0$ for any $l \neq 0$ so that (12) holds if the variance is uniformly bounded: $c_{0f} \leq A < \infty \ \forall f \in \mathcal{F}$. The condition of i.i.d. process is trivially extended to a broader condition of uncorrelated fading process.

2. An extension of the previous case is a finite-memory process: $c_{lf} = 0$ for any $|l| > L_f$, where L_f is the memory under fading distribution f, which is uniformly bounded: $L_f \leq L < \infty$ for any $f \in \mathcal{F}$.

3. Infinite-memory processes are also allowed provided that the correlation decays to zero asymptotically, e.g. an exponential correlation model: $c_{lf} = c_{0f} r_f^{|l|}$, where r_f is the correlation coefficient under distribution f and $0 \leq r_f \leq B < 1$ for each $f \in \mathcal{F}$ (i.e. uniformly bounded away from unity), in addition to the standard requirement $c_{0f} \leq A < \infty$.

4. Condition (12) is satisfied if the fading process is uniformly, asymptotically independent:

$$\lim_{l\to\infty} \sup_{f\in\mathcal{F}} \sup_{s_1,s_l} \left|\frac{f(s_1,s_l)}{f(s_1)f(s_l)} - 1\right| = 0 \quad (13)$$

i.e. $f(s_1,s_l) \to f(s_1)f(s_l)$ uniformly over $s_1, s_l, f \in \mathcal{F}$, which is equivalent to $f(s_l|s_1) \to f(s_l)$ so that the process forgets its past asymptotically (and uniformly).

5. Cases when c_{lf} does not decay to zero can be included too, e.g. $c_{lf} = (-1)^l$.

6. Any compound fading channel where each $f \in \mathcal{F}$ is ergodic and $\mathcal{F}$ is of finite cardinality is automatically uniformly-ergodic.

One can also construct examples whereby the channel is not uniformly ergodic while being ergodic for each $f \in \mathcal{F}$. Let $1 \leq k < \infty$ be an integer index specifying an ergodic distribution f from the uncertainty set $\mathcal{F}$ and consider example 2 with $L_f = k$ and $c_{lf} = c_{0f} > 0$ for any $|l| \leq L_f$, or example 3 with $r_f = 1 - 1/k$. In both cases, the uniform convergence condition is broken and the corresponding compound channels are not uniformly ergodic while being ergodic for each $f \in \mathcal{F}$.

IV. The Capacity of Uniformly-Ergodic Channels

Let C be the information capacity of the compound ergodic-fading channel above:

$$C \triangleq \sup_{p(x)} \inf_{f\in\mathcal{F}_1} I(X;Y|f) \quad (14)$$

Note that such $\sup - \inf$ expression appears often in the theory of compound channels and is, in many cases, the operational capacity. However, this is not the case in general [1]. It is the purpose of this section to show that this is indeed the case for the uniformly-ergodic compound channel above. First, we establish the following saddle-point property of the information capacity.

Proposition 2. *Consider the compound ergodic-fading channel in* (1)–(3)*, where* $f \in \mathcal{F}_1$*. Assume that the set of feasible input distributions* $p(x)$ *is convex (e.g. average or maximum power constraint) and that* $\mathcal{F}_1$ *is convex. The information capacity* C *of this compound ergodic-fading channel satisfies the saddle-point property,*

$$C = \sup_{p(x)} \inf_{f\in\mathcal{F}_1} I(X;Y|f) = \inf_{f\in\mathcal{F}_1} \sup_{p(x)} I(X;Y|f) = C_w \quad (15)$$

where C_w is the capacity of worst-case channel in the uncertainty set, i.e. the information capacity equals to the worst-case channel capacity C_w. If the $\inf$ *and* $\sup$ *are achieved, then the following saddle-point inequalities holds for any feasible $p(x)$ and $f(s)$,*

$$I(X;Y|f^*) \le C = I(X^*;Y|f^*) \le I(X^*;Y|f) \tag{16}$$

where X^* denotes the input under its optimal distribution $p^*(x)$ and (p^*, f^*) is a saddle point.

Proof. The saddle point property follows from the fact that $I(X;Y|f)$ is concave in $p(x)$ and linear (and thus convex) in f; since the sets of feasible f and $p(x)$ are convex, von Neumann mini-max Theorem [10] guarantees the existence of a saddle point. The saddle-point inequalities in (16) follow from 2nd equality in (15). ☐

The inequalities in (16) have a well-known game-theoretic interpretation: the Tx chooses $p^*(x)$ and the adversary (nature) chooses f^*; neither player can deviate from this optimal strategy without incurring a penalty.

In the rest of this section, we demonstrate that the information capacity C of the compound ergodic-fading channel has the operational meaning of a maximum achievable rate i.e. the compound channel capacity C_c for the class of uniformly-ergodic fading channels defined above.

Our approach applies to any convex uncertainty set $\mathcal{F}_1$ and also to any convex set of possible input distributions (which may also include a power constraint).

Theorem 1. *Consider a compound uniformly-ergodic fading channel. Let $\mathcal{F}_1$ be a convex set of its marginal fading distributions and $\mathcal{F}$ be a set of its joint fading distributions. Assume that the Rx has the full CSI (i.e. the state sequence s^n) while the Tx has only partial CDI: it knows $\mathcal{F}$ and hence $\mathcal{F}_1$ but neither s^n nor its fading distribution f. Let the set of feasible input distributions $p(x)$ be convex. The operational capacity C_c of this compound channel is*

$$C_c = \sup_{p(x)} \inf_{f \in \mathcal{F}_1} I(X;Y|f) = C = C_w \tag{17}$$

i.e. the same as the worst-case channel capacity C_w in (15).

Proof. See Appendix. ☐

Note that, from this Theorem, (i) full knowledge of the fading distribution at the Tx does not increase the capacity, and (ii) a code designed for the worst-case fading distribution also works for the whole class of distributions (and hence much smaller amount of feedback is needed).

We remark that this result cannot be established using Theorem 3.3.5 in [9] and considering fading distribution as a meta-state since there are uncountably many possible distributions in $\mathcal{F}_1$ (since this set is continuous) while Theorem 3.3.5 requires the number of states to be finite - see [7] for details. It should also be noted that while the definition of uniform ergodic channels and the respective uncertainty set $\mathcal{F}$ as well as the error probability depend on the joint fading distribution $f(s^n)$, the capacity depends only on the marginal distribution $f(s)$ and its uncertainty set $\mathcal{F}_1$. This fact, which results in the single-letter capacity expression, cannot be inferred from Theorem 3.3.5 either. In the proof, we make use of the general formula for compound channel capacity in [7], which does not have the restrictions of Theorem 3.3.5, by applying it to the ergodic scenario of the present paper.

It can be further shown, using Fano's inequality, that the first two equalities in (17) do hold even if $\mathcal{F}_1$ is not convex [15][1]. However, the saddle point property and hence the last equality do not need to hold in this case.

V. Appendix

Proof of Proposition 1: It is straightforward to verify (by direct computations) that

$$\sigma_{nf}^2 = \frac{1}{n}\sum_{l=1-n}^{n-1}\left(1-\frac{|l|}{n}\right)c_{lf} = \frac{2}{n}\sum_{l=0}^{n-1}\left(1-\frac{l}{n}\right)c_{lf} - \frac{1}{n}c_{0f}$$

(since $c_{(-l)f} = c_{lf}$) and that $\lim_{n\to\infty}\sup_{f\in\mathcal{F}}\sigma_{nf}^2 = 0$ if and only if (10) holds provided that c_{0f} is uniformly bounded: $c_{0f} \le A < \infty\ \forall f \in \mathcal{F}$.

To establish (11), observe that

$$\left|\frac{1}{n}\sum_{l=0}^{n-1}c_{lf}\right| = \left|\mathbb{E}\left\{\left(\frac{1}{n}\sum_{l=1}^{n}I_{s_l} - I_f\right)(I_{s_1} - I_f)\right\}\right| \le \sigma_{nf}\sqrt{c_{0f}} \tag{18}$$

where the inequality follows from Cauchy-Schwartz inequality, so that (11) follows provided that c_{0f} is uniformly bounded. This establishes the "only if" part.

To establish the "if" part of (11), let

$$z_{nf} = \frac{1}{n}\sum_{l=0}^{n-1}\left(1-\frac{l}{n}\right)c_{lf} = \frac{1}{n^2}\sum_{l=0}^{n-1}\sum_{i=1}^{n-l}c_{lf} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=0}^{n-i}c_{lf}$$

Observe that (11) implies that for any $\delta > 0$ there exists such $n_0(\delta)$ that for any $n \ge n_0(\delta)$

$$\left|\frac{1}{n}\sum_{l=0}^{n-1}c_{lf}\right| \le \delta\ \forall f \in \mathcal{F} \tag{19}$$

Set $n \ge n_0^2(\delta)$ and let $L_n = n - \sqrt{n}$ (round off if not integer) so that

$$\begin{aligned} z_{nf} &= \frac{1}{n^2}\sum_{i=1}^{L_n}\sum_{l=0}^{n-i}c_{lf} + \frac{1}{n^2}\sum_{i=L_n+1}^{n}\sum_{l=0}^{n-i}c_{lf} \\ &\le \frac{1}{n}\sum_{i=1}^{L_n}\left|\frac{1}{n-i}\sum_{l=0}^{n-i}c_{lf}\right| + \frac{1}{n^2}\sum_{i=L_n+1}^{n}\sum_{l=0}^{n-i}c_{0f} \\ &\le \frac{1}{n}\sum_{i=1}^{L_n}\delta + \frac{(n-L_n)^2}{n^2}c_{0f} \le \delta + \frac{c_{0f}}{n} \end{aligned} \tag{20}$$

where 1st inequality is from $c_{lf} \le c_{0f}$ and 2nd one is from $n - i \ge n - L_n \ge n_0(\delta)$. Since $\delta > 0$ is arbitrary, $z_{nf} \ge 0$ and c_{0f} is uniformly bounded, the "if" part follows by taking $\lim_{n\to\infty}\sup_{f\in\mathcal{F}}$:

$$0 \le \lim_{n\to\infty}\sup_{f\in\mathcal{F}} z_{nf} \le \delta\ \forall\delta > 0 \tag{21}$$

[1]The authors greatly appreciate the very insightful comments by an anonymous reviewer.

□

Proof of Theorem 1: Let $X^n = \{X_1...X_n\}$, $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ and likewise for $\boldsymbol{Y}$. Following Theorem 5 in [7], the capacity of general compound channels (e.g. not necessarily information-stable and where the uncertainty set can be arbitrary) with full Rx CSI but no Tx CSI is given by

$$C_c = \sup_{\boldsymbol{X}} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \quad (22)$$

where the supremum is over all sequences of finite-dimensional input distributions and $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$ is the compound inf-information rate,

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \sup_R \left\{ R : \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\{Z_{ns} \le R\} = 0 \right\} \quad (23)$$

where $Z_{ns} = n^{-1} i(X^n;Y^n|s)$ is the normalized information density under channel state s; $\mathcal{S}$ is the (arbitrary) uncertainty set.

To prove (17), first observe that $C_c \le C_w$ holds in full generality[2] and, using (15),

$$C_c \le C_w = C = \sup_{p(x)} \inf_{f\in\mathcal{F}_1} I(X;Y|f) \quad (24)$$

It remains to show that the inequality is actually equality. To this end, apply the general formula in (22) by considering the fading distribution f as a (meta) state s, and restrict the optimization to i.i.d. inputs $\tilde{\boldsymbol{X}}$ to obtain a lower bound

$$C_c \ge \sup_{\tilde{\boldsymbol{X}}} \underline{\underline{I}}(\tilde{\boldsymbol{X}};\tilde{\boldsymbol{Y}}) \quad (25)$$

where $\tilde{\boldsymbol{Y}}$ is the output under i.i.d. input $\tilde{\boldsymbol{X}}$. The following propositions evaluate $\underline{\underline{I}}(\tilde{\boldsymbol{X}};\tilde{\boldsymbol{Y}})$.

Proposition 3. *The compound inf-information rate* $\underline{\underline{I}}(\tilde{\boldsymbol{X}};\tilde{\boldsymbol{Y}})$ *can be upper bounded as follows:*

$$\underline{\underline{I}}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}) \le \inf_{f\in\mathcal{F}_1} I(X;Y|f) \quad (26)$$

where X has the same distribution as the marginals of $\tilde{\boldsymbol{X}}$.

Proof. Let $I_f = I(X;Y|f)$, $i_k = i(X_k;Y_k|S_k)$, $I_{s_k} = \mathbb{E}_{X,Y}\{i_k\}$, $z_n = n^{-1}\sum_{k=1}^n i_k$. From Proposition 1 in [7],

$$\underline{\underline{I}}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}) \le \inf_{f\in\mathcal{F}} \underline{I}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}|f) \quad (27)$$

where $\underline{I}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}|f)$ is the inf-information rate under (meta) state f:

$$\underline{I}(\tilde{\boldsymbol{X}};\tilde{\boldsymbol{Y}}|f) = \sup_R \left\{ R : \lim_{n\to\infty} \Pr\{z_n \le R\} = 0 \right\} \quad (28)$$

Note that $\mathbb{E}\{i_k\} = I_f$ and

$$\mathbb{E}\left\{|z_n - I_f|^2\right\} = \frac{1}{n^2}\sum_{k,l} \mathbb{E}\{(I_{s_k} - I_f)(I_{s_l} - I_f)\} = \sigma_{nf}^2$$

where 1st equality follows from the fact that (X_i,Y_i) and (X_j,Y_j) are independent of each other ($i \ne j$) given the state sequence $\{s_1, s_2, ...\}$, so that, from Chebychev inequality,

$$\Pr\{|z_n - I_f| \ge \delta\} \le \sigma_{nf}^2/\delta^2 \to 0 \ \forall\ f \in \mathcal{F} \quad (29)$$

[2]the compound capacity never exceeds the worst-case one since a code that works for the whole uncertainty set has also to work on the worst-case channel in the set [1].

for any $\delta > 0$ as $n \to \infty$ since, due to (7), $\lim_{n\to\infty} \sigma_{nf} = 0$. Therefore, $z_n = n^{-1}\sum_{k=1}^n i_k \to I_f$ in probability, from which it follows that

$$\underline{I}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}|f) = I(X;Y|f) \quad (30)$$

Combining this with (27), one obtains (26). □

Proposition 4. *The compound inf-information rate* $\underline{\underline{I}}(\tilde{\boldsymbol{X}};\tilde{\boldsymbol{Y}})$ *can be lower bounded as follows:*

$$\underline{\underline{I}}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}) \ge \inf_{f\in\mathcal{F}_1} I(X;Y|f) \quad (31)$$

Proof. Observe that, for each $\delta > 0$,

$$\Pr\left\{ z_n \le \inf_{f\in\mathcal{F}_1} I_f - \delta \right\} \le \Pr\{|z_n - I_f| \ge \delta\} \le \frac{\sigma_{nf}^2}{\delta^2}$$

Applying $\lim - \sup$ and using (7), one obtains

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Pr\left\{ z_n \le \inf_{f\in\mathcal{F}_1} I_f - \delta \right\} = 0 \quad (32)$$

i.e. $\underline{\underline{I}}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}) \ge \inf_{f\in\mathcal{F}_1} I_f - \delta$. Since this holds for any $\delta > 0$, (31) follows. □

Combining Propositions 3 and 4,

$$\underline{\underline{I}}(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}) = \inf_{f\in\mathcal{F}_1} I(X;Y|f) \quad (33)$$

for any i.i.d. input. Applying $\sup_{\tilde{\boldsymbol{X}}}$ to this equality in combination with (24) and (25), one obtains the desired result.

References

[1] A. Lapidoth and P. Narayan, "Reliable Communication Under Channel Uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, No. 6, Oct. 1998.

[2] S. Loyka, C.D. Charalambous, Novel Matrix Singular Value Inequalities and Their Applications to Uncertain MIMO Channels, IEEE Trans. Info. Theory, v.61, n.12, pp. 6623 - 6634, Dec. 2015.

[3] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, No. 6, pp. 2619-2692, Oct. 1998.

[4] G. Caire, S. Shamai, On The Capacity of Some Channels With Channel State Information, IEEE Transactions on Information Theory, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[5] I. Ioannou, C.D. Charalambous, S. Loyka, "Outage Probability Under Channel Distribution Uncertainty", *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6825-6838, Nov. 2012.

[6] S. Loyka, C. D. Charalambous, Ergodic Capacity Under Channel Distribution Uncertainty, 52nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 1-3 2014.

[7] S. Loyka, C.D. Charalambous, A General Formula for Compound Channel Capacity, IEEE Transactions on Information Theory, v. 62, no.7, pp.3971-2991, July 2016.

[8] S. Verdu, T.S. Han, "A General Formula for Channel Capacity", *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.

[9] T. S. Han, Information-Spectrum Method in Information Theory, New York: Springer, 2003.

[10] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[11] A.M. Yaglom, An Introduction to the Theory of Stationary Random Functions, Uspehi Matematicheskih Nauk, v. 7, no. 5(51), pp. 3–168, 1952.

[12] D.R. Brillinger, Time Series: Data Analysis and Theory, Holt, Rinehart and Winston Inc., New York, 1975.

[13] S.M. Rytov, An Introduction to Statistical Radio Physics, v.1: Stochastic Processes, Nauka, Moscow, 1976.

[14] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw Hill, Boston, 2002.

[15] A Private Communication by An Anonymous Reviewer, 2017.