Strong Converse for General Compound Channels

Sergey Loyka, Charalambos D. Charalambous

***Abstract*—A general compound channel is considered, where no stationarity, ergodicity or information stability is required. Following the recent result on the capacity of this channel under the full Rx CSI, sufficient and necessary conditions are obtained for the strong converse to hold. In a nutshell, even though no information satiability is required upfront, the conditions imply that there exists a sub-sequence of (bad) channel states (indexed by the blocklength) for which the respective information density rates converge in probability to the compound channel capacity, i.e. this sub-sequence is information stable.**

I. Introduction

It is well-known that channel state information (CSI) affects significantly system performance and respective channel capacity. It can be rather limited in many scenarios, especially for wireless systems, where low SNR, interference and channel dynamics are significant, and where the feedback (if any) is also limited [1]. A popular approach to model the impact of limited CSI is to assume that the receiver (Rx) and transmitter (Tx) know that the unknown channel is fixed and belongs to a certain class of channels (uncertainty set), which is known as the compound channel model [2]-[6]. The capacity of compound channels has been extensively studied since late 1950s [2]-[5]; see [6] for an extensive literature review up to late 1990s, and [9] for more recent results.

All of these studies assume that each channel in the uncertainty set is information-stable (in the sense of Dobrushin [10] or Pinsker [11]), e.g. stationary and ergodic. However, there are many scenarios (especially in wireless communications) where the channels are not stationary, ergodic or information-stable. This setting was recently studied in [14], where the capacity of general (information-unstable) compound channels was established under the full Rx CSI using the information density (spectrum) approach of [7][8]. The assumption of full Rx CSI is motivated by the fact that channel estimation is done at the Rx so that full Rx CSI may be available if the SNR is high enough but limited (if any) feedback to the Tx makes full Tx CSI unfeasible.

While the channel capacity theorem ensures the achievability of any rate below the capacity with arbitrary low error probability, there exists a hope to achieve higher rates by allowing slightly higher error probability, since the transition from arbitrary low to high error probability may be slow. Strong converse ensures that this transition is very sharp (for any rate above the capacity, the error probability converges to 1) and hence dispels the hope. In this paper, we extend the study in [14] by establishing the sufficient and necessary conditions for the strong converse to hold for the general compound channel. In a nutshell, the conditions require the existence of an information-stable sub-sequence of (bad) channel states (indexed by the blocklength) such that the respective sub-sequence of information densities converges in probability to the compound channel capacity. No assumptions of stationarity, ergodicity or information stability are made for the members of the uncertainty set.

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ontario, Canada, e-mail: sergey.loyka@ieee.org
C.D. Charalambous is with the ECE Department, University of Cyprus, Nicosia, Cyprus, e-mail: chadcha@ucy.ac.cy

II. Channel Model

Let us consider a generic discrete-time channel model where $X^n = \{X_1...X_n\}$ is a (random) sequence of n input symbols, $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ denotes all such sequences, and Y^n is the corresponding output sequence; $s \in \mathcal{S}$ denotes the channel state (which may also be a sequence) and $\mathcal{S}$ is the (arbitrary) uncertainty set; $p_s(y^n|x^n)$ is the channel transition probability; $p(x^n)$ and $p_s(y^n)$ are the input and output distributions under channel state s.

Let us assume that the full CSI is available at the receiver (Rx) but not the transmitter (Tx) (see e.g. [1] for a detailed motivation of this assumption; when the channel is quasi-static, this assumption is not necessary) and that the channel input $\boldsymbol{X}$ and state s are independent of each other. Following the standard approach (see e.g. [1]), we augment the channel output with the state: $Y^n \to (Y^n, s)$. The information density [10]-[13] between the input and output for a given channel state s and a given input distribution $p(x^n)$ is

$$i(x^n; y^n, s) = \ln \frac{p_s(x^n, y^n)}{p(x^n)p_s(y^n)} = i(x^n; y^n|s) \qquad (1)$$

where we have used the fact that the input X^n and channel state s are independent of each other. Note that we make no assumptions of stationarity, ergodicity or information stability in this paper, so that the normalized information density $n^{-1}i(X^n; Y^n|s)$ does not have to converge to the respective mutual information rate as $n \to \infty$. There is no need for the consistency assumption on $p_s(y^n|x^n)$ either (e.g. the channel may behave differently for even and odd n).

For future use, we give the formal definitions of information stability following [10]-[12] (with a slight extension to the compound setting).

Definition 1. *Two random sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$ are information-stable if*

$$\frac{i(X^n; Y^n|s)}{I(X^n; Y^n|s)} \xrightarrow{\Pr} 1 \text{ as } n \to \infty \qquad (2)$$

i.e. the normalized information density converges in probability to the respective mutual information rate $\frac{1}{n}I(X^n; Y^n|s)$.

Definition 2. *Channel state s is information stable if there exists an input $\boldsymbol{X}$ such that*

$$\frac{i(X^n;Y^n|s)}{I(X^n;Y^n|s)} \xrightarrow{\Pr} 1, \quad \frac{I(X^n;Y^n|s)}{C_{ns}} \to 1 \text{ as } n \to \infty, \tag{3}$$

where $C_{ns} = \sup_{p(x^n)} I(X^n;Y^n|s)$ is the information capacity.

Note that the 2nd definition requires effectively the channel to behave ergodically under the optimal input only, and tells us nothing about its behaviour under other inputs (e.g. a practical code) and, in this sense, is rather limiting. To characterize the channel behaviour under different inputs (not only the optimal one), we will consider the information stability of its input $\boldsymbol{X}$ and the induced output $\boldsymbol{Y}$ following Definition 1. Further note that, for the compound channel, some channel states may be information stable while others are not.

III. Capacity of the General Compound Channel

We define an (n, r_n, ε_n)-code for a compound channel in the standard way, where n is the blocklength, $r_n = \ln M_n/n$ is the code rate and M_n is the number of codewords, and ε_n is the compound error probability,

$$\varepsilon_n = \sup_{s\in\mathcal{S}} \varepsilon_{ns} \tag{4}$$

where ε_{ns} is the error probability under channel state s. Rate R is achievable if $\liminf_{n\to\infty} r_n \geq R$ and $\lim_{n\to\infty} \varepsilon_n = 0$, which ensures arbitrary low error probability for any channel in the uncertainty set for sufficiently large n [1]-[6]. The capacity is the supremum of all achievable rates. Codebooks are required to be independent of the actual channel state s while the decision regions are allowed to depend on s (due to full Rx CSI).

Below, we briefly review the relevant results in [14], which are instrumental for further development here.

Theorem 1 ([14])**.** *Consider a general compound channel where the channel state $s \in \mathcal{S}$ is known to the receiver but not the transmitter and is independent of the channel input; the transmitter knows the (arbitrary) uncertainty set $\mathcal{S}$. Its compound channel capacity is given by*

$$C_c = \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \tag{5}$$

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,

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \sup_R \left\{ R : \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\{Z_{ns} \leq R\} = 0 \right\} \tag{6}$$

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

This theorem was proved using the Verdu-Han and Feinstein Lemmas properly extended to the compound channel setting.

Lemma 1 (Feinstein Lemma for compound channels [14])**.** *For arbitrary input X^n and uncertainty set $\mathcal{S}$ and any r_n, there exists a (n, r_n, ε_n)-code (where the codewords are independent of channel state s), satisfying the following inequality,*

$$\varepsilon_n \leq \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \leq r_n + \gamma \right\} + e^{-\gamma n} \tag{7}$$

for any $\gamma > 0$. □

Lemma 2 (Verdu-Han Lemma for compound channels [14])**.** *For any uncertainty set $\mathcal{S}$, every (n, r_n, ε_n)-code satisfies the following inequality,*

$$\varepsilon_n \geq \sup_{s\in\mathcal{S}} \Pr\left\{ n^{-1} i(X^n;Y^n|s) \leq r_n - \gamma \right\} - e^{-\gamma n} \tag{8}$$

for any $\gamma > 0$, where X^n is uniformly distributed over all codewords and Y^n is the corresponding channel output under channel state s. □

IV. Strong Converse for the General Compound Channel

Strong converse ensures that slightly larger error probability cannot be traded off for higher data rate (since the transition from arbitrary low to high error probability is sharp).

Definition 3. *A compound channel is said to satisfy strong converse if*

$$\lim_{n\to\infty} \varepsilon_n = 1 \tag{9}$$

for any code satisfying

$$\liminf_{n\to\infty} r_n > C_c \tag{10}$$

To obtain conditions for strong converse, let $\check{I}(\boldsymbol{X};\boldsymbol{Y})$ be the "worst-case" sup-information rate,

$$\check{I}(\boldsymbol{X};\boldsymbol{Y}) = \inf_R \left\{ R : \lim_{n\to\infty} \inf_{s\in\mathcal{S}} \Pr\{Z_{ns} > R\} = 0 \right\} \tag{11}$$

where $Z_{ns} = n^{-1} i(X^n;Y^n|s)$ is the information density rate, and $I_{ns}(a)$ be the truncated mutual information,

$$I_{ns}(a) = E\{Z_{ns} 1[Z_{ns} \leq a]\}, \; I_{ns} = \lim_{a\to\infty} I_{ns}(a) \tag{12}$$

where $1[\cdot]$ is the indicator function and $I_{ns} = I(X^n;Y^n|s)$ is the mutual information under channel state s. The compound sup-information rate $\overline{\overline{I}}(\boldsymbol{X};\boldsymbol{Y})$ and the sup-information rate $\bar{I}(\boldsymbol{X};\boldsymbol{Y}|s)$ under channel state s are defined as

$$\overline{\overline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \inf_R \left\{ R : \lim_{n\to\infty} \sup_{s\in\mathcal{S}} \Pr\{Z_{ns} \geq R\} = 0 \right\} \tag{13}$$

$$\bar{I}(\boldsymbol{X};\boldsymbol{Y}|s) = \inf_R \left\{ R : \lim_{n\to\infty} \Pr\{Z_{ns} \geq R\} = 0 \right\} \tag{14}$$

The following Proposition establishes an ordering of various information rates.

Proposition 1. *The following inequalities hold for any input*

$$\begin{aligned} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) &\leq \check{I}(\boldsymbol{X};\boldsymbol{Y}) \\ &\leq \inf_s \bar{I}(\boldsymbol{X};\boldsymbol{Y}|s) \\ &\leq \sup_s \bar{I}(\boldsymbol{X};\boldsymbol{Y}|s) \\ &\leq \overline{\overline{I}}(\boldsymbol{X};\boldsymbol{Y}) \end{aligned} \tag{15}$$

Proof. see the Appendix. ☐

It can be shown, via examples, that all inequalities can be strict. Using this Proposition, sufficient and necessary conditions for the strong converse to hold can be established.

Theorem 2. *A sufficient and necessary condition for the general compound channel to satisfy strong converse is*

$$\sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) \tag{16}$$

If this holds and the convergence $I_{ns}(a) \to I_{ns}$ is uniform in n, s for any input $\boldsymbol{X}^*$ satisfying $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) > C_c - \delta$ for some $\delta > 0$ (i.e. the input $\boldsymbol{X}^*$ is δ-suboptimal), then

$$C_c = \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) = \liminf_{n\to\infty} \sup_{p(x^n)} \inf_s \frac{1}{n} I(X^n;Y^n|s) \tag{17}$$

The condition (16) *is equivalent to:*

1) for any $\delta > 0$ and any input $\boldsymbol{X}^*$ satisfying $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) > C_c - \delta$,

$$\lim_{n\to\infty} \inf_s \Pr\{|Z_{ns}^* - C_c| > \delta\} = 0 \tag{18}$$

where $Z_{ns}^* = \frac{1}{n} i(X^{n*};Y^{n*}|s)$ is the normalized information density under input $\boldsymbol{X}^*$.

2) for any input $\boldsymbol{X}$ and any $\delta > 0$,

$$\lim_{n\to\infty} \inf_s \Pr\{Z_{ns} > C_c + \delta\} = 0 \tag{19}$$

Proof. see the Appendix. ☐

Remark 1. *In the case of a single-state channel,*

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \underline{I}(\boldsymbol{X};\boldsymbol{Y}),\ \check{I}(\boldsymbol{X};\boldsymbol{Y}) = \overline{I}(\boldsymbol{X};\boldsymbol{Y}) \tag{20}$$

where $\underline{I}(\boldsymbol{X};\boldsymbol{Y})$, $\overline{I}(\boldsymbol{X};\boldsymbol{Y})$ are inf and sup-information rates for the regular (single-state) channel, and Theorem 2 reduces to the corresponding Theorem in [7][8].

Remark 2. *Note that, under the conditions of Theorem 2 that lead to* (17)*, the compound channel behaves ergodically even though no assumption of ergodicity (or information stability) was made upfront.*

Below, we consider a special case when the supremum in (5) is achieved.

Corollary 1. *If the channel satisfies strong converse and the supremum in $\sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$ is achieved, i.e.*

$$\exists \boldsymbol{X}^* :\ \underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) = C_c \tag{21}$$

then $\check{I}(\boldsymbol{X}^*;\boldsymbol{Y}^*) = C_c$ and there exists such sequence of channel states $s(n)$ that the corresponding sequence of normalized information densities $Z_{ns(n)}^*$ (under input $\boldsymbol{X}^*$) converges in probability to the compound channel capacity C_c,

$$\lim_{n\to\infty} \Pr\{|Z_{ns(n)}^* - C_c| > \delta\} = 0\ \forall \delta > 0 \tag{22}$$

i.e. this sequence (which represents worst-case channels in the uncertainty set) is information-stable.

Proof. Observe that $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) = C_c$ implies

$$C_c = \underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) \le \check{I}(\boldsymbol{X}^*;\boldsymbol{Y}^*) \le \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) = C_c \tag{23}$$

so that $\check{I}(\boldsymbol{X}^*;\boldsymbol{Y}^*) = C_c$ follows, which also implies that

$$\lim_{n\to\infty} \inf_s \Pr\{Z_{ns}^* > C_c + \delta\} = 0\ \forall\ \delta > 0 \tag{24}$$

On the other hand, $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) = C_c$ implies

$$\lim_{n\to\infty} \sup_s \Pr\{Z_{ns}^* < C_c - \delta\} = 0\ \forall\ \delta > 0 \tag{25}$$

and hence

$$\lim_{n\to\infty} \inf_s \Pr\{|Z_{ns}^* - C_c| > \delta\} = 0\ \forall \delta > 0 \tag{26}$$

follows. Next, we need the following technical Lemma.

Lemma 3. *Let $\{x_{ns}\}$ be a non-negative compound sequence such that*

$$\lim_{n\to\infty} \inf_s x_{ns} = 0 \tag{27}$$

Then, there exists such sequence of states $s(n)$ that

$$\lim_{n\to\infty} x_{ns(n)} = 0 \tag{28}$$

Proof. When $\inf_s$ is achieved, the statement is trivial. To prove it in the general case, observe that, from the definition of $\inf_s$ and for any n, there always exists such $s(n)$ that

$$x_{ns(n)} < \inf_s x_{ns} + 1/n \tag{29}$$

so that taking $\lim_{n\to\infty}$ of both sides, one obtains (28)[1]. ☐

Using this Lemma, (26) implies the existence of a sequence of channel states $s(n)$ such that (22) holds. ☐

Remark 3. *Note that, under the conditions of Corollary 1, the sequence $s(n)$ of worst-case channel states is information-stable even though no assumption of information stability was made upfront.*

Remark 4. *In light of Lemma 3, condition* (19) *means that there exists such sequence of (bad) channel states $s(n)$ that the information spectrum of the corresponding sequence of normalized information densities $Z_{ns(n)}$ does not exceed C_c under any input, i.e.*

$$\exists s(n) :\ \lim_{n\to\infty} \Pr\{Z_{ns(n)} > C_c + \delta\} = 0\ \forall \delta > 0 \tag{30}$$

V. Appendix

A. Proof of Proposition 1

The 1st inequality is proved by contradiction. Let $\underline{\underline{I}} = \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$, $\check{I} = \check{I}(\boldsymbol{X};\boldsymbol{Y})$, assume $\underline{\underline{I}} - \check{I} = 2\delta > 0$ and set

$$R = (\underline{\underline{I}} + \check{I})/2 = \underline{\underline{I}} - \delta = \check{I} + \delta \tag{31}$$

so that

$$\begin{aligned} 0 &= \lim_{n\to\infty} \sup_s \Pr\{Z_{ns} < \underline{\underline{I}} - \delta\} \\ &= \lim_{n\to\infty} \sup_s \Pr\{Z_{ns} < R\} \\ &= 1 - \lim_{n\to\infty} \inf_s \Pr\{Z_{ns} \ge R\} \\ &= 1 - \lim_{n\to\infty} \inf_s \Pr\{Z_{ns} \ge \check{I} + \delta\} = 1 \end{aligned} \tag{32}$$

[1]this way of proof was suggested by a reviewer.

i.e. a contradiction.

The 2nd inequality is also proved by contradiction. Let $\bar{I} = \inf_s \bar{I}(\boldsymbol{X};\boldsymbol{Y}|s)$, assume $\check{I} - \bar{I} = 2\delta > 0$ and set

$$R = (\bar{I} + \check{I})/2 = \bar{I} + \delta = \check{I} - \delta \tag{33}$$

so that, from the definition of $\check{I}$,

$$\begin{aligned} 0 < \epsilon &= \limsup_{n\to\infty} \inf_s \Pr\{Z_{ns} > \check{I} - \delta\} \\ &\le \inf_s \limsup_{n\to\infty} \Pr\{Z_{ns} > \check{I} - \delta\} \\ &= \inf_s \limsup_{n\to\infty} \Pr\{Z_{ns} > \bar{I} + \delta\} \\ &\le \limsup_{n\to\infty} \Pr\{Z_{ns^*} > \bar{I} + \delta\} \\ &\le \limsup_{n\to\infty} \Pr\{Z_{ns^*} > \bar{I}(\boldsymbol{X};\boldsymbol{Y}|s^*) + \delta/2\} = 0 \end{aligned} \tag{34}$$

i.e. a contradiction, where s^* is such channel state that

$$\bar{I}(\boldsymbol{X};\boldsymbol{Y}|s^*) \le \inf_s \bar{I}(\boldsymbol{X};\boldsymbol{Y}|s) + \delta/2 \tag{35}$$

The last inequality can be proved in a similar way.

B. *Proof of Theorem 2*

To prove sufficiency, let the equality in (16) to hold and select a code satisfying

$$\liminf_{n\to\infty} r_n = R = C_c + 3\delta \tag{36}$$

for some $\delta > 0$, so that

$$r_n \ge R - \delta = C_c + 2\delta = \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) + 2\delta \tag{37}$$

for sufficiently large n. Using Lemma 2 for this code, one obtains:

$$\begin{aligned} \lim_{n\to\infty} \varepsilon_n &\ge \lim_{n\to\infty} \sup_s \Pr\left\{Z_{ns} \le r_n - \delta\right\} \\ &\ge \lim_{n\to\infty} \sup_s \Pr\left\{Z_{ns} \le \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) + \delta\right\} \\ &\ge \lim_{n\to\infty} \sup_s \Pr\left\{Z_{ns} \le \check{I}(\boldsymbol{X};\boldsymbol{Y}) + \delta\right\} \\ &= 1 - \lim_{n\to\infty} \inf_s \Pr\left\{Z_{ns} > \check{I}(\boldsymbol{X};\boldsymbol{Y}) + \delta\right\} \\ &= 1 \end{aligned} \tag{38}$$

so that (9) holds, where the last equality is due to

$$\lim_{n\to\infty} \inf_s \Pr\left\{Z_{ns} > \check{I}(\boldsymbol{X};\boldsymbol{Y}) + \delta\right\} = 0 \tag{39}$$

which follows from (11).

To prove the necessary part, assume that (9) holds and, using Lemma 1, select a code satisfying

$$\lim_{n\to\infty} r_n = R = C_c + \delta \tag{40}$$

for some $\delta > 0$. This implies that

$$r_n \le C_c + 2\delta \tag{41}$$

for any sufficiently large n. Applying Lemma 1, one obtains

$$\begin{aligned} 1 = \lim_{n\to\infty} \varepsilon_n &\le \lim_{n\to\infty} \sup_s \Pr\left\{Z_{ns} \le r_n + \delta\right\} \\ &\le \lim_{n\to\infty} \sup_s \Pr\left\{Z_{ns} \le C_c + 3\delta\right\} \\ &= 1 \end{aligned} \tag{42}$$

from which it follows that

$$\lim_{n\to\infty} \inf_s \Pr\left\{Z_{ns} > C_c + 3\delta\right\} = 0 \tag{43}$$

which implies (19) and $\check{I}(\boldsymbol{X};\boldsymbol{Y}) \le C_c$ (under any input) so that, from Proposition 1,

$$C_c = \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \le \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) \le C_c \tag{44}$$

from which (16) follows.

To establish the sufficiency of (19), observe that it implies the 2nd inequality in (44) from which (16) follows, which is sufficient.

To establish (18), observe that $C_c = \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$ implies that there exists such input $\boldsymbol{X}^*$ that $\underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) > C_c - 2\delta$ so that, for any such $\boldsymbol{X}^*$,

$$\begin{aligned} 0 &= \lim_{n\to\infty} \sup_s \Pr\left\{\frac{1}{n} i(X^{n*};Y^{n*}|s) < \underline{\underline{I}}(\boldsymbol{X}^*;\boldsymbol{Y}^*) - \delta\right\} \\ &\ge \lim_{n\to\infty} \sup_s \Pr\left\{\frac{1}{n} i(X^{n*};Y^{n*}|s) < C_c - 3\delta\right\} = 0 \end{aligned} \tag{45}$$

Combining this with (43) applied to input $\boldsymbol{X}^*$, one obtains

$$\begin{aligned} \lim_{n\to\infty} \inf_s \Pr\{|Z_{ns}^* - C_c| > 3\delta\} &\le \lim_{n\to\infty} \inf_s \Pr\{Z_{ns}^* > C_c + 3\delta \\ &+ \lim_{n\to\infty} \sup_s \Pr\{Z_{ns}^* < C_c - 3\delta\} = 0 \end{aligned} \tag{46}$$

from which (18) follows.

To establish last equality in (17), let $\check{I} = \check{I}(\boldsymbol{X};\boldsymbol{Y})$ and observe that

$$\begin{aligned} I_{ns}(a) = &\underbrace{E\{Z_{ns} 1[Z_{ns} \le \check{I} + \delta]\}}_{e_1} \\ &+ \underbrace{E\{Z_{ns} 1[\check{I} + \delta < Z_{ns} \le a]\}}_{e_2} \end{aligned} \tag{47}$$

for some $\delta > 0$, where $1[\cdot]$ is the indicator function. The two expectation terms can be upper bounder as

$$\begin{aligned} e_1 &\le (\check{I} + \delta) \Pr\{Z_{ns} \le \check{I} + \delta\} \\ e_2 &\le a \cdot \Pr\{Z_{ns} > \check{I} + \delta\} \end{aligned} \tag{48}$$

so that

$$\begin{aligned} \liminf_{n\to\infty} \inf_s \frac{1}{n} I(X^n;Y^n|s) &= \liminf_{n\to\infty} \inf_s \lim_{a\to\infty} I_{ns}(a) \\ &= \lim_{a\to\infty} \liminf_{n\to\infty} \inf_s I_{ns}(a) \\ &\le \lim_{a\to\infty} \liminf_{n\to\infty} \inf_s ((\check{I} + \delta) \Pr\{Z_{ns} \le \check{I} + \delta\} \\ &\quad + a \cdot \Pr\{Z_{ns} > \check{I} + \delta\}) \\ &\le \lim_{a\to\infty} ((\check{I} + \delta) \limsup_{n\to\infty} \sup_s \Pr\{Z_{ns} \le \check{I} + \delta\} \\ &\quad + a \cdot \liminf_{n\to\infty} \inf_s \Pr\{Z_{ns} > \check{I} + \delta\}) \\ &= \check{I} + \delta \end{aligned} \tag{49}$$

where the 2nd equality is due to uniform convergence and the last equality is due to

$$\liminf_{n\to\infty} \inf_s \Pr\{Z_{ns} > \check{I} + \delta\}) = 0 \tag{50}$$

$$\begin{aligned} &\limsup_{n\to\infty} \sup_s \Pr\{Z_{ns} \le \check{I} + \delta\} \\ &\quad = 1 - \liminf_{n\to\infty} \inf_s \Pr\{Z_{ns} > \check{I} + \delta\}) = 1 \end{aligned} \tag{51}$$

Since (49) holds for arbitrary small $\delta > 0$, it follows that

$$\liminf_{n\to\infty} \inf_s \frac{1}{n} I(X^n; Y^n|s) \leq \check{I} \tag{52}$$

for any input. Taking $\sup_{p(\boldsymbol{x})}$ on both sides, one obtains:

$$\begin{aligned} C_c &= \sup_{p(\boldsymbol{x})} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \\ &\leq \liminf_{n\to\infty} \sup_{p(x^n)} \inf_s \frac{1}{n} I(X^n; Y^n|s) \\ &\leq \sup_{p(\boldsymbol{x})} \check{I}(\boldsymbol{X};\boldsymbol{Y}) = C_c \end{aligned} \tag{53}$$

from which the desired result follows, where the 1st inequality is due to Proposition 2 below.

Proposition 2. *Consider the general compound channel. Its compound inf-information rate is bounded as follows:*

$$\underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y}) \leq \liminf_{n\to\infty} \inf_s \frac{1}{n} I(X^n; Y^n|s) \leq \check{I}(\boldsymbol{X};\boldsymbol{Y}) \tag{54}$$

Proof. Let $Z_{ns} = \frac{1}{n} i(X^n; Y^n|s)$ and observe that

$$\begin{aligned} &\frac{1}{n} I(X^n; Y^n|s) = E\{Z_{ns}\} \\ &\quad \geq E\{Z_{ns} 1[Z_{ns} \leq 0]\} + E\{Z_{ns} 1[Z_{ns} \geq \underline{\underline{I}} - \delta]\} \end{aligned} \tag{55}$$

for any $0 < \delta < \underline{\underline{I}}$, where $1[\cdot]$ is the indicator function and $\underline{\underline{I}} = \underline{\underline{I}}(\boldsymbol{X},\boldsymbol{Y})$. The 1st term t_1 can be lower bounded as follows:

$$\begin{aligned} t_1 &= E\{Z_{ns} 1[Z_{ns} \leq 0]\} \\ &= \frac{1}{n} \sum_{x^n, y^n : z_{ns} \leq 0} p_s(y^n) p(x^n) w_{ns} \ln w_{ns} \\ &\geq -\frac{1}{ne} \sum_{x^n, y^n : z_{ns} \leq 0} p_s(y^n) p_s(x^n) \\ &\geq -\frac{1}{ne} \end{aligned} \tag{56}$$

where $w_{ns} = p_s(y^n|x^n)/p_s(y^n)$ and the 1st inequality follows from $w \ln w \geq -1/e$. The 2nd term t_2 can be lower bounded as follows:

$$\begin{aligned} t_2 &= E\{Z_{ns} 1[Z_{ns} \geq \underline{\underline{I}} - \delta]\} \\ &= \sum_{x^n, y^n : z_{ns} \geq \underline{\underline{I}} - \delta} z_{ns} p_s(y^n|x^n) p(x^n) \\ &\geq (\underline{\underline{I}} - \delta) \Pr\{Z_{ns} \geq \underline{\underline{I}} - \delta\} \end{aligned} \tag{57}$$

Combining these two bounds, one obtains:

$$\begin{aligned} &\liminf_{n\to\infty} \inf_s \frac{1}{n} I(X^n; Y^n|s) \\ &\quad \geq (\underline{\underline{I}} - \delta) \lim_{n\to\infty} \inf_s \Pr\{Z_{ns} \geq \underline{\underline{I}} - \delta\} \\ &\quad = \underline{\underline{I}} - \delta \end{aligned} \tag{58}$$

where the equality follows from

$$\begin{aligned} 0 &= \lim_{n\to\infty} \sup_s \Pr\{Z_{ns} < \underline{\underline{I}} - \delta\} \\ &= 1 - \lim_{n\to\infty} \inf_s \Pr\{Z_{ns} \geq \underline{\underline{I}} - \delta\} \end{aligned} \tag{59}$$

Since the inequality in (58) holds for each $\delta > 0$, one obtains the 1st inequality in (54) by taking $\delta \to 0$; the 2nd one has been already established in (52). □

REFERENCES

[1] 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.

[2] R.L. Dobrushin, “Optimal information Transmission through a channel with unknown parameters,” *Radiotekhnika i Electronika*, vol. 4, pp. 1951-1956, 1959.

[3] D. Blackwell, L. Breiman, and A. J. Thomasian, “The capacity of a class of channels,” *Ann. Math. Statist.*, vol. 30, pp. 1229-1241, December 1959.

[4] J. Wolfowitz, “Simultaneous channels,” *Arch. Rat. Mech. Anal.*, vol. 4, pp. 371-386, 1960.

[5] W. L. Root, P. P. Varaya, ”Capacity of Classes of Gaussian Channels”, *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350-1393, Nov. 1968.

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

[7] 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.

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

[9] M. Effros, A. Goldsmith, Y. Liang, ”Generalizing Capacity: New Definitions and Capacity Theorems for Composite Channels,” *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3069-3087, July 2010.

[10] R. L. Dobrushin, ”A general formulation of the fundamental theorem of Shannon in information theory”, Uspekhi Mat. Nauk, v. 14, no. 6(90), Nov.-Dec. 1959, pp.3–104.

[11] M. S. Pinsker, Information and Information Stability of Random Variables and Processes. San Francisco: Holden-Day, 1964.

[12] H.K. Ting, ”On the Information Stability of a Sequence of Channels”, Theory of Probability and Its Applications, v.7, N. 3, pp. 258-269, 1962.

[13] R.L. Stratonovich, Information Theory, Moscow: Sovetskoe Radio, 1974.

[14] S. Loyka, C. D. Charalambous, A General Formula for Compound Channel Capacity, IEEE Int. Symp. on Information Theory (ISIT-15), Hong Kong, June 14-19, 2015.