Feedback Does Not Increase the Capacity of Compound Channels with Additive Noise

Sergey Loyka, Charalambos D. Charalambous

***Abstract*—A discrete compound channel with memory is considered, where no stationarity, ergodicity or information stability is required, and where the uncertainty set can be arbitrary. When the discrete noise is additive but otherwise arbitrary and there is no cost constraint on the input, it is shown that the causal feedback does not increase the capacity. This extends the earlier result obtained for general channels with full transmitter (Tx) channel state information (CSI). It is further shown that, for this compound setting and under a mild technical condition on the additive noise, the addition of the full Tx CSI does not increase the capacity either, so that the worst-case and compound channel capacities are the same, thus revealing a saddle-point property.**

I. Introduction

Many channels, especially wireless ones, are non-erogodic, non-stationary in nature [1] so that the standard tools developed for stationary ergodic channels do not apply and new methods are needed for such channels. A powerful method to deal with general channels, for which stationarity, ergodicity or information stability are not required, is the information density (spectrum) approach [2][3]. In this method, the key quantity is the inf-information rate rather than the mutual information since the latter does not have operational meaning for information-unstable channels.

In real systems, channel state information (CSI) may be inaccurate or limited due to a number of reasons such as the limitations of channel estimation or feedback link [1]. The concept of compound channel is one way to address this issue whereby a codebook is designed to work for any channel in the uncertainty set, without any knowledge of what channel state is currently in effect [4]. A number of results have been obtained for the capacity of compound channels, see [4] for a detailed review. While most of the studies do not consider feedback, the compound capacity of a class of finite-state memoryless (and hence information-stable) channels with deterministic feedback was established in [5].

While most of the known results require some form of information stability for any channel in the uncertainty set, a general formula for compound channel capacity has been established in [6] where no stationarity, ergodicity or information stability is required, and the uncertainty set can be arbitrary. The key quantity in this setting is the compound inf-information rate, which is an extension of the inf-information rate of [2][3] to the compound setting.

In this paper, we extend the study in [6][8] and consider a general compound channel with memory and additive noise (no information stability is required so that the channel can be non-stationary, non-ergodic; the uncertainty set can be arbitrary), where all alphabets are discrete, there is no cost constraint and a noiseless, causal feedback link is present, where all past channel outputs are fed back to the transmitter. We consider a scenario where no CSI is available at the transmitter but full CSI is available to the receiver. Under this setting, we demonstrate that the feedback does not increase the compound channel capacity[1]. This extends the earlier result in [7] established for known channels (full CSI available at both ends). Since noisy feedback cannot outperform noiseless one, this also holds for the former case. Under a mild technical condition on the additive noise, we further show that the availability of the full Tx CSI does not increase the capacity either: the worst-case and compound channel capacities are the same. This fact is remarkable since achieving the worst-case capacity allows for the codebooks to depend on the channel state while the compound channel capacity requires the codebooks to be independent of the channel state.

II. Channel Model

Let us consider the following discrete-time model of a compound discrete channel with additive noise:

$$y_k = \sum_{i=0}^{l} x_{k-i} + \xi_{sk} \quad (1)$$

where x_k, y_k, ξ_{ks} are the input, output and noise at time k; without loss of generality, we set $x_i = 0$ for $i < 0$; all alphabets as well as operations are M-ary, $s \in \mathcal{S}$ denotes the channel (noise) state, and $\mathcal{S}$ is the (arbitrary) channel uncertainty set (this can also be extended to the case where channel memory l depends on its state s). The compound sequence $\xi_s^n = \{\xi_{1s}, .., \xi_{ns}\}$ represents arbitrary additive noise, e.g. non-ergodic, non-stationary in general. Note that the channel is not memoryless, it includes inter-symbol interference of depth l and the noise is allowed to have memory as well. The channel is not required to be information stable (in the sense of Dobrushin [11] or Pinsker [12]). We assume that s is known to the receiver but not the transmitter, who knows the (arbitrary) uncertainty set $\mathcal{S}$. This is motivated by the fact that channel estimation is done at the receiver; M may be small, e.g. binary alphabets, while the cardinality of $\mathcal{S}$ can be very large (in fact, $\mathcal{S}$ can be a continuous set) so it is not feasible in practice to feed s back to the transmitter via e.g. a binary feedback channel.

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

[1]Here, we consider the classical compound setting [1][4][5] where a fixed-rate code is designed to operate on any channel in the uncertainty set and its decoding regions are allowed to depend on the state (but not the encoding process); variable-rate coding is beyond the paper's scope.

978-1-5090-1806-2/16/$31.00 ©2016 Crown

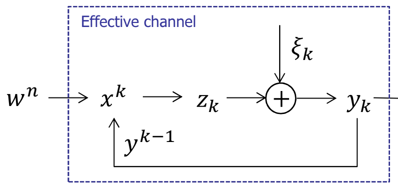

Fig. 1. A general channel with additive noise and causal feedback $\{w^n y^{k-1}\} \rightarrow x^k \rightarrow y_k$ and the effective channel $w^n \rightarrow y^n$ (dashed box), where $z_k = \sum_{i=0}^{l} x_{k-i}$, $x^k = f^k(w^n y^{k-1})$.

The channel has noiseless feedback with 1-symbol delay, so that the transmitted symbol x_k at time $k = 1..n$ is selected as $x_k^{(n)} = f_k^{(n)}(w^n y^{k-1})$[2], where n is the blocklength, w^n denotes the message to be transmitted via n-symbol block, $x^n = \{x_1^{(n)} ... x_n^{(n)}\}$ and likewise for y^n and ξ_s^n; $f_k^{(n)}$ denotes the encoding function at time k, which depends on the selected message w^n and past channel outputs y^{k-1} (due to the feedback); $f^n = \{f_1^{(n)} ... f_n^{(n)}\}$. This induces the input distribution of the form:

$$p(x^n || y^{n-1}) = \prod_{k=1}^{n} p(x_k | x^{k-1} y^{k-1}) \quad (2)$$

where $||$ denotes causal conditioning [9]. No cost constraint is imposed on the input.

Notations: To simplify notations, we use $p(x|y)$ to denote conditional distribution $p_{x|y}(x|y)$ when this causes no confusion (and likewise for joint and marginal distributions) and shortcut $x_k^{(n)}$ as x_k with understanding that all sequences and distributions depend on blocklength n and may be different for different blocklengths. Capitals (X) denote random variables while lower-case letters (x) denote their realizations or arguments of functions; $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$.

III. Capacity without Feedback

Below, we briefly review the relevant results in [6][8], which apply to general compound channels $p_s(y^n|x^n)$, not only those in (1); channels can be information-unstable, e.g. non-stationary, non-ergodic, but without feedback, i.e. $x^k = f^k(w^n)$ (the input depends only on the message and the past inputs, not the outputs).

Theorem 1 ([6][8]). *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 (without feedback) is given by*

$$C_{NFB} = \sup_{\boldsymbol{X}} \underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \quad (3)$$

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,

$$\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) = \sup_{R} \left\{ R : \lim_{n \to \infty} \sup_{s \in \mathcal{S}} \Pr\{Z_{ns} \leq R\} = 0 \right\} \quad (4)$$

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

This theorem was proved in [6][8] using the Verdu-Han and Feinstein Lemmas properly extended to the compound channel setting.

For future use, we need the following definitions.

Definition 1. *Let $\{X_{sn}\}_{n=1}^{\infty}$ be an arbitrary compound random sequence where s is a state (i.e. a random sequence indexed by the state s). The compound infimum $\underline{\underline{\{\cdot\}}}$ and supremum $\overline{\overline{\{\cdot\}}}$ operators are defined as follows:*

$$\underline{\underline{\boldsymbol{X}}} = \underline{\underline{\{X_{sn}\}}} = \sup \left\{ x : \lim_{n \to \infty} \sup_{s} \Pr\{X_{sn} \leq x\} = 0 \right\} \quad (5)$$

$$\overline{\overline{\boldsymbol{X}}} = \overline{\overline{\{X_{sn}\}}} = \inf \left\{ x : \lim_{n \to \infty} \sup_{s} \Pr\{X_{sn} \geq x\} = 0 \right\} \quad (6)$$

These operators generalize the respective inf $\overline{\boldsymbol{X}}$ and sup $\underline{\boldsymbol{X}}$ operators for regular (single-state) sequences. The following definitions extend the respective information-theoretic quantities in [2] to the compound setting.

Definition 2. *Let $\boldsymbol{X} = \{X_s^n\}_{n=1}^{\infty}$ and $\boldsymbol{Y} = \{Y_s^n\}_{n=1}^{\infty}$ be two compound random sequences with distributions p_{sx^n} and p_{sy^n} where s is a state. The compound inf-divergence rate is defined as*

$$\underline{\underline{D}}(\boldsymbol{X}; \boldsymbol{Y}) = \underline{\underline{\{d_{sn}(X_s^n || Y_s^n)\}}} \quad (7)$$

where $d_{sn}(x^n || y^n) = \frac{1}{n} \ln \frac{p_{sx^n}(x^n)}{p_{sy^n}(x^n)}$ is the divergence density rate. The compound inf and sup-entropy rates $\underline{\underline{H}}(\boldsymbol{X})$ and $\overline{\overline{H}}(\boldsymbol{X})$ are defined as

$$\underline{\underline{H}}(\boldsymbol{X}) = \underline{\underline{\{h_{sn}(X_s^n)\}}}, \ \overline{\overline{H}}(\boldsymbol{X}) = \overline{\overline{\{h_{sn}(X_s^n)\}}} \quad (8)$$

where $h_{sn}(x^n) = -n^{-1} \ln p_s(x^n)$ is the entropy density rate. The compound conditional inf-entropy rate $\underline{\underline{H}}(\boldsymbol{Y}|\boldsymbol{X})$ and sup-entropy rate $\overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{X})$ are defined analogously via the conditional entropy density rate $h_{sn}(y^n|x^n) = -n^{-1} \ln p_s(y^n|x^n)$ (with respect to the joint distribution $p_s(x^n, y^n)$), and $\overline{\overline{I}}(\boldsymbol{X}; \boldsymbol{Y})$ is similarly defined.

The proposition below gives the properties useful in evaluation of the compound inf-information rate $\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y})$.

Proposition 1. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be (arbitrary) compound random sequences. The following holds:*

$$\underline{\underline{D}}(\boldsymbol{X} || \boldsymbol{Y}) \geq 0 \quad (9)$$

$$\overline{\overline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \geq \underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \geq 0 \quad (10)$$

$$\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) = \underline{\underline{I}}(\boldsymbol{Y}; \boldsymbol{X}) \quad (11)$$

$$\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \leq \overline{\overline{H}}(\boldsymbol{Y}) - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \quad (12)$$

$$\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \leq \underline{\underline{H}}(\boldsymbol{Y}) - \underline{\underline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \quad (13)$$

$$\underline{\underline{I}}(\boldsymbol{X}; \boldsymbol{Y}) \geq \underline{\underline{H}}(\boldsymbol{Y}) - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \quad (14)$$

$$\overline{\overline{H}}(\boldsymbol{Y}) \geq \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \quad (15)$$

$$\overline{\overline{H}}(\boldsymbol{Y}) \geq \underline{\underline{H}}(\boldsymbol{Y}) \geq \underline{\underline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \quad (16)$$

[2] our result will also hold for a more general feedback of the form $u_k = g_k(y^k)$, where $\{g_k\}$ are arbitrary feedback functions, see Remark 4.

If the alphabets are discrete, then

$$0 \le \underline{\underline{H}}(\boldsymbol{X}|\boldsymbol{Y}) \le \underline{\underline{H}}(\boldsymbol{X}) \le \overline{\overline{H}}(\boldsymbol{X}) \le \ln N_x \tag{17}$$

$$0 \le \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \le \min\{\underline{\underline{H}}(\boldsymbol{X}), \underline{\underline{H}}(\boldsymbol{Y})\} \le \min\{\ln N_x, \ln N_y\} \tag{18}$$

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \min\{\underline{\underline{H}}(\boldsymbol{X}), \underline{\underline{H}}(\boldsymbol{Y})\} \quad \text{if } \min\{\overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{X}), \overline{\overline{H}}(\boldsymbol{X}|\boldsymbol{Y})\} = 0 \tag{19}$$

$$0 \le \overline{\overline{I}}(\boldsymbol{X};\boldsymbol{Y}) \le \min\{\overline{\overline{H}}(\boldsymbol{X}), \overline{\overline{H}}(\boldsymbol{Y})\} \le \min\{\ln N_x, \ln N_y\} \tag{20}$$

where the last inequalities in (17)-(20) *hold if the alphabets are of finite cardinality* N_x, N_y.

Proof. See Appendix. □

Note that many of these properties mimic the respective properties of mutual information and entropy, e.g. "conditioning cannot increase the entropy" and "mutual information is non-negative, symmetric and bounded by the entropy of the alphabet".

IV. Capacity with Feedback

In this section, we consider a discrete compound channel with feedback and general additive noise. Instead of dealing with the feedback channel $\{w^n y^{k-1}\} \to x^k \to y_k$, $k = 1...n$ directly, one can consider an effective channel $w^n \to y^n$ without feedback, see Fig. 1. Applying Theorem 1 to the effective channel, the capacity with the feedback can be expressed as[3]

$$C_{FB} = \sup_{\boldsymbol{W},\boldsymbol{F}} \underline{\underline{I}}(\boldsymbol{W};\boldsymbol{Y}) \tag{21}$$

where $\boldsymbol{Y} = \{Y^n\}_{n=1}^{\infty}$ and likewise for $\boldsymbol{W}$, $\underline{\underline{I}}(\boldsymbol{W};\boldsymbol{Y})$ is the compound inf-information rate:

$$\underline{\underline{I}}(\boldsymbol{W};\boldsymbol{Y}) = \underline{\underline{\{n^{-1} i(W^n;Y^n|s)\}}} \tag{22}$$

where $i(W^n;Y^n|s)$ is the information density:

$$i(w^n;y^n|s) = \log \frac{p_s(y^n|w^n)}{p_s(y^n)} \tag{23}$$

The maximization in (21) is over all possible encoding functions $\boldsymbol{F} = \{f^n\}_{n=1}^{\infty}$ and all possible message distributions. Unfortunately, this maximization is difficult to perform in general. Therefore, we proceed in a different way. Let

$$\overline{\overline{H}}(\boldsymbol{\Xi}) = \overline{\overline{\{n^{-1} h(\Xi_s^n|s)\}}} \tag{24}$$

be the compound sup-entropy rate of the compound noise $\boldsymbol{\Xi} = \{\Xi_s^n\}_{n=1}^{\infty}$, $\Xi_s^n = \{\Xi_{1s}...\Xi_{ns}\}$, $h(\xi^n|s) = -\log p_s(\xi^n)$.

Theorem 2. *The capacity of the compound discrete channel with (arbitrary) additive noise in* (1) *is not increased by the causal feedback:*

$$C_{FB} = C_{NFB} = \sup_{\boldsymbol{X}} \underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \log M - \overline{\overline{H}}(\boldsymbol{\Xi}) \tag{25}$$

where C_{NFB} is the capacity without feedback, and $\sup_{\boldsymbol{X}}$ is over all sequences of input distributions.

Proof. Let us consider the no-feedback case first. The 2nd equality follow from Theorem 1. The following Lemma is needed to prove the last equality.

Lemma 1. *Let $z_k = \sum_{i=0}^{l} x_{k-i}$, $k = 1...n$. If $p(x^n) = 1/M^n$, then $p(z^n) = 1/M^n$, i.e. equiprobable X^n generates equiprobable Z^n.*

Now, since the mapping $x^n \to z^n$ is invertible, $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) = \underline{\underline{I}}(\boldsymbol{Z};\boldsymbol{Y})$. It follows from (12) that

$$\underline{\underline{I}}(\boldsymbol{Z};\boldsymbol{Y}) \le \overline{\overline{H}}(\boldsymbol{Y}) - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{Z}) \tag{26}$$

Using this inequality, one obtains:

$$\underline{\underline{I}}(\boldsymbol{Z};\boldsymbol{Y}) \le \log M - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{Z}) \tag{27}$$

$$= \log M - \overline{\overline{H}}(\boldsymbol{\Xi}) \tag{28}$$

where 1st inequality is due to M-ary alphabets, so that $\overline{\overline{H}}(\boldsymbol{Y}) \le \log M$ (see (17)), and the equality is due to $\overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{Z}) = \overline{\overline{H}}(\boldsymbol{Z}+\boldsymbol{\Xi}|\boldsymbol{Z}) = \overline{\overline{H}}(\boldsymbol{\Xi})$, since the noise is additive and independent of the input (no feedback). Finally,

$$\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y}) \le \log M - \overline{\overline{H}}(\boldsymbol{\Xi}) \tag{29}$$

and the equality is achieved by equiprobable input due to Lemma 1, under which the output is also equiprobable. This proves the last equality in (25).

To prove 1st equality, $C_{FB} = C_{NFB}$, observe that feedback cannot decrease the capacity,

$$\log M - \overline{\overline{H}}(\boldsymbol{\Xi}) = C_{NFB} \le C_{FB} \tag{30}$$

To prove the converse,

$$C_{FB} \le \log M - \overline{\overline{H}}(\boldsymbol{\Xi}) \tag{31}$$

use (21) to conclude

$$C_{FB} = \sup_{\boldsymbol{W},\boldsymbol{F}} \underline{\underline{I}}(\boldsymbol{W};\boldsymbol{Y}) \tag{32}$$

$$\le \sup_{\boldsymbol{W},\boldsymbol{F}} [\overline{\overline{H}}(\boldsymbol{Y}) - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{W})] \tag{33}$$

$$\le \log M - \inf_{\boldsymbol{W},\boldsymbol{F}} \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{W}) \tag{34}$$

where 1st inequality is due to $\underline{\underline{I}}(\boldsymbol{W};\boldsymbol{Y}) \le \overline{\overline{H}}(\boldsymbol{Y}) - \overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{W})$ and 2nd inequality is due to $\overline{\overline{H}}(\boldsymbol{Y}) \le \log M$ (since the alphabet is M-ary).

To evaluate $\overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{W})$, note that

$$p_s(y^n|w^n) = \prod_{k=1}^{n} p_s(y_k|y^{k-1}w^n) \tag{35}$$

and

$$p_s(y_k|y^{k-1}w^n) = p_s(y_k|y^{k-1}x^k w^n) \tag{36}$$

$$= p_s(y_k|y^{k-1}x^k \xi^{k-1} w^n) \tag{37}$$

$$= p_{sy}\left(\sum_{i=0}^{l} x_{k-i} + \xi_k | x^k \xi^{k-1} w^n\right) \tag{38}$$

$$= p_{s\xi}(\xi_k|\xi^{k-1}w^n) \tag{39}$$

$$= p_{s\xi}(\xi_k|\xi^{k-1}) \tag{40}$$

[3] see also [10] for a formulation based on the directed information for the case of full CSI and a proof of equivalence of these two formulations in the latter case.

where $\xi_k = y_k - \sum_{i=0}^{l} f_{k-i}(w^n y^{k-i-1})$, $\xi^n = \{\xi_k\}_{k=1}^n$. 1st equality is due to $x^k = f^k(w^n y^{k-1})$; 2nd and 3rd equalities are due to the channel model $y_k = z_k + \xi_k$; 4th equality is due to $x^k = \check{f}^k(w^n \xi^{k-1})$, where $\check{f}^k$ is a function which depends on encoding functions f^k; last equality is due to independence of noise and message. Thus,

$$p_{sy}(y^n|w^n) = p_{s\xi}(\xi^n) \tag{41}$$

and therefore

$$\overline{\overline{H}}(\boldsymbol{Y}|\boldsymbol{W}) = \overline{\overline{H}}(\boldsymbol{\Xi}) \tag{42}$$

Combining this with (34), one obtains (31) and hence the desired result follows.

Equality in (31) is achieved by i.i.d. equiprobable input. □

Remark 1. *Note that in both feedback and no-feedback systems, the optimizing input is i.i.d. equiprobable, under which the output is also i.i.d. equiprobable under any noise, which explains why feedback is not helpful in this setting.*

Remark 2. *Setting $l = 0$, one obtains the channel without intersymbol interference. When, in addition, the uncertainty set $\mathcal{S}$ is singleton (single-state channel with no uncertainty), Theorem 2 above reduces to the corresponding result in [7] obtained for fully known (no uncertainty) channels.*

Remark 3. *Since noisy feedback cannot perform better than noiseless, this result also implies that noisy feedback cannot increase the compound capacity in this setting either.*

Remark 4. *One may consider a more general feedback of the form $u_k = g_k(y^k)$, where $\{g_k\}$ are arbitrary (possibly random) feedback functions (which account for e.g. quantization of feedback signals), and the corresponding encoding of the form $x_k = f_k(w^n u^{k-1})$. Since the capacity with this form of feedback cannot exceed the capacity with the full feedback of y^{k-1}, Theorem 2 still holds for this setting as well.*

Remark 5. *The above result also applies to a more general channel model of the form $y_k = z_k + \xi_{sk}$, where $z_k = g_k(x^k)$ and the functions g_k, $k = 1..n$, are such that the mapping $x^n \to z^n$ is one-to-one and z_k, ξ_{sk} have the same alphabets*[4].

V. Impact of the Tx CSI

Let us consider the case where channel state s is known at the transmitter, so that codewords can be selected as functions of the channel state. In this case, the worst-case channel capacity is a proper performance metric and it can be expressed as

$$C_w = \inf_s \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) \tag{43}$$

$$= \inf_s(\log M - \overline{H}(\boldsymbol{\Xi}|s)) \tag{44}$$

$$= \log M - \sup_s \overline{H}(\boldsymbol{\Xi}|s) \tag{45}$$

$$\geq \log M - \overline{\overline{H}}(\boldsymbol{\Xi}) = C_{FB} \tag{46}$$

[4]This was pointed out by a reviewer.

where $\underline{I}(\boldsymbol{X};\boldsymbol{Y}|s)$ is the inf-information rate under channel state s:

$$\underline{I}(\boldsymbol{X};\boldsymbol{Y}|s) = \sup_R \left\{ R : \lim_{n\to\infty} \Pr\left\{ \frac{1}{n} i(W^n;Y^n|s) \leq R \right\} = 0 \right\}, \tag{47}$$

$\overline{H}(\boldsymbol{\Xi}|s)$ is the sup-entropy rate of the noise:

$$\overline{H}(\boldsymbol{\Xi}|s) = \inf_R \left\{ R : \lim_{n\to\infty} \Pr\left\{ \frac{1}{n} h(\Xi_s^n|s) > R \right\} = 0 \right\} \tag{48}$$

(43) follows from the general formula in [2]; (44) follows from the Theorem in [7]; (46) follows from the Lemma 2 below, so that the impact of Tx CSI can be characterized by

$$\Delta C = C_w - C_{FB} = \overline{\overline{H}}(\boldsymbol{\Xi}) - \sup_s \overline{H}(\boldsymbol{\Xi}|s) \geq 0 \tag{49}$$

Note that, similarly to the compound capacity, C_w is not increased by the feedback either.

To proceed further, we need the following definition.

Definition 3. *The compound noise sequence $\{\Xi_s^n\}_{n=1}^\infty$ is uniform if the convergence in*

$$\Pr\left\{ \frac{1}{n} h(\Xi_s^n|s) > \sup_s \overline{H}(\boldsymbol{\Xi}|s) + \delta \right\} \to 0 \tag{50}$$

as $n \to \infty$ is uniform in $s \in \mathcal{S}$ for any $\delta > 0$.

Note that, while the convergence to zero in (50) for each $\delta > 0$ and $s \in \mathcal{S}$ is guaranteed from the definition of $\sup_s \overline{H}(\boldsymbol{\Xi}|s)$, this convergence does not have to be uniform in general.

Lemma 2. *The following inequality holds for the general compound noise sequence:*

$$\overline{\overline{H}}(\boldsymbol{\Xi}) \geq \sup_s \overline{H}(\boldsymbol{\Xi}|s) \tag{51}$$

with equality when the compound noise is uniform.

Proof. The proof is by contradiction. Assume that $\overline{\overline{H}}(\boldsymbol{\Xi}) < \sup_s \overline{H}(\boldsymbol{\Xi}|s)$ which implies that

$$\exists s_0 : \overline{\overline{H}} = \overline{\overline{H}}(\boldsymbol{\Xi}) < \overline{H} = \overline{H}(\boldsymbol{\Xi}|s_0) \tag{52}$$

and set

$$R = (\overline{\overline{H}} + \overline{H})/2 = \overline{\overline{H}} + \Delta = \overline{H} - \Delta \tag{53}$$

where $\Delta = (\overline{H} - \overline{\overline{H}})/2 > 0$. Note that

$$\lim_{n\to\infty} \Pr\left\{ \frac{1}{n} h(\Xi_{s_0}^n|s_0) > \overline{H} - \Delta \right\} > 0 \tag{54}$$

from the definition of $\overline{H}$. However,

$$0 = \lim_{n\to\infty} \sup_s \Pr\left\{ \frac{1}{n} h(\Xi_s^n|s) > \overline{\overline{H}} + \Delta \right\} \tag{55}$$

$$\geq \lim_{n\to\infty} \Pr\left\{ \frac{1}{n} h(\Xi_{s_0}^n|s_0) > \overline{\overline{H}} + \Delta \right\} \tag{56}$$

$$= \lim_{n\to\infty} \Pr\left\{ \frac{1}{n} h(\Xi_{s_0}^n|s_0) > \overline{H} - \Delta \right\} > 0 \tag{57}$$

where 1st equality is due to the definition of $\overline{\overline{H}}$, i.e. a contradiction, from which the desired inequality follows.

The equality case is also proved by contradiction: assume that, under the uniform convergence,

$$\overline{\overline{H}} > \overline{H} = \sup_s \overline{H}(\Xi|s) \tag{58}$$

and set

$$R = (\overline{\overline{H}} + \overline{H})/2 = \overline{\overline{H}} - \Delta = \overline{H} + \Delta \tag{59}$$

where $\Delta = (\overline{\overline{H}} - \overline{H})/2 > 0$, and hence

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n}h(\Xi_s^n|s) > \overline{H} + \Delta\right\} = 0 \ \forall s \in \mathcal{S} \tag{60}$$

from the definition of $\overline{H}$, so that a contradiction follows

$$0 = \sup_s \lim_{n\to\infty} \Pr\left\{\frac{1}{n}h(\Xi_s^n|s) > \overline{H} + \Delta\right\} \tag{61}$$

$$= \lim_{n\to\infty} \sup_s \Pr\left\{\frac{1}{n}h(\Xi_s^n|s) > \overline{H} + \Delta\right\} \tag{62}$$

$$= \lim_{n\to\infty} \sup_s \Pr\left\{\frac{1}{n}h(\Xi_s^n|s) > \overline{\overline{H}} - \Delta\right\} > 0 \tag{63}$$

where 2nd equality is due to uniform convergence and the last inequality is from the definition of $\overline{\overline{H}}$. □

Combining Lemma 2 with (49), one obtains the following.

Theorem 3. *Consider the discrete compound channel with additive noise as in* (1). *When the compound noise is uniform, neither Tx CSI nor causal noiseless (or noisy) feedback increase its capacity, i.e.*

$$C_{NFB} = C_{FB} = C_w \tag{64}$$

The last equality states that there exists a saddle point: the worst-case channel capacity (achievable by codebooks tailored to the channel state) is the same as the compound channel capacity (where the codebooks are independent of channel states).

VI. Appendix: Proof of Proposition 1

To prove (9), observe that

$$\lim_{n\to\infty} \sup_s \Pr\left\{\frac{1}{n}\ln\frac{p_{sx^n}(X^n)}{p_{sy^n}(X^n)} \le -\delta\right\}$$

$$= \lim_{n\to\infty} \sup_s \sum_{x^n: p_{sx^n}(x^n) \le p_{sy^n}(x^n)e^{-\delta n}} p_{sx^n}(x^n) \tag{65}$$

$$\le \lim_{n\to\infty} \sup_s \sum_{x^n} p_{sy^n}(x^n)e^{-\delta n}$$

$$= \lim_{n\to\infty} e^{-\delta n} = 0 \ \forall \delta > 0 \tag{66}$$

from which (9) follows.

(10) follows by observing that $\underline{\underline{I}}(\boldsymbol{X};\boldsymbol{Y})$ is the compound inf-divergence rate between $(\boldsymbol{X},\boldsymbol{Y})$ and $(\boldsymbol{X}',\boldsymbol{Y}')$, where $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ are independent of each other and have the same distributions as $\boldsymbol{X}$ and $\boldsymbol{Y}$.

(11) follows from the symmetry of information density: $i(x^n;y^n|s) = i(y^n;x^n|s)$.

(12)-(14) follow from using $\underline{\underline{(\cdot)}}$ on

$$i(x^n;y^n|s) = \ln\frac{1}{p_s(y^n)} - \ln\frac{1}{p_s(y^n|x^n)} \tag{67}$$

with $\underline{\underline{(-\boldsymbol{X})}} = -\overline{\overline{(\boldsymbol{X})}}$, and applying the inequalities in the following Lemma in [8].

Lemma 3. *Let $\{X_{ns}\}_{n=1}^{\infty}$ and $\{Y_{ns}\}_{n=1}^{\infty}$ be two (arbitrary) compound random sequences and s is a (common) state. Then,*

$$\underline{\underline{\boldsymbol{X}}} + \underline{\underline{\boldsymbol{Y}}} \le \underline{\underline{\{\boldsymbol{X}+\boldsymbol{Y}\}}} \le \underline{\underline{\boldsymbol{X}}} + \overline{\overline{\boldsymbol{Y}}} \tag{68}$$

(15)-(16) follow from (12)-(13).

1st inequality in (17) follows from $p_s(x^n|y^n) \le 1$ when the alphabet is discrete. To prove the last inequality, let $Z_{ns} = -n^{-1}\ln p_s(X^n)$ and observe the following:

$$\Pr\{Z_{ns} \ge \ln N_x + \delta\} = \sum_{x^n: p_s(x^n) \le e^{-n(\ln N_x+\delta)}} p_s(x^n)$$

$$\le \sum_{x^n} e^{-n(\ln N_x+\delta)}$$

$$= e^{-n(\ln N_x+\delta)} N_x^n = e^{-n\delta} \tag{69}$$

so that

$$\lim_{n\to\infty} \sup_s \Pr\{Z_{ns} \ge \ln N_x + \delta\} = 0$$

and therefore $\underline{\underline{H}}(\boldsymbol{X}) \le \overline{\overline{H}}(\boldsymbol{X}) \le \ln N_x + \delta$ for any $\delta > 0$, from which the desired inequality follows. This also implies the last inequalities in (18)-(20).

2nd inequality in (18) follows from $\underline{\underline{H}}(\boldsymbol{Y}|\boldsymbol{X}) \ge 0$ and (12), (11).

2nd inequality in (20) can be obtained via similar reasoning using

$$\overline{\overline{I}}(\boldsymbol{X};\boldsymbol{Y}) \le \overline{\overline{H}}(\boldsymbol{X}) - \underline{\underline{H}}(\boldsymbol{X}|\boldsymbol{Y}) \tag{70}$$

(19) follow from (14).

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] S. Verdu, T.S. Han, "A General Formula for Channel Capacity", *IEEE Trans. Info. Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.

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

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

[5] B. Shrader, H. Permuter, Feedback Capacity of the Compound Channel, IEEE Trans. Info. Theory, v. 55, no. 8, pp. 3629-3644, Aug. 2009.

[6] 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.

[7] F. Alajaji, Feedback Does Not Increase the Capacity of Discrete Channels with Additive Noise, IEEE Trans. Information Theory, v. 41, N. 2, pp. 546-549, Mar. 1995.

[8] S. Loyka, C. D. Charalambous, A General Formula for Compound Channel Capacity, IEEE Trans. IT, accepted, Apr. 2016. Available: https://arxiv.org/abs/1604.01434

[9] G. Kramer, Capacity Results for the Discrete Memoryless Network, IEEE Trans. Info. Theory, v. 49, no. 1, pp 4-21, Jan. 2003.

[10] S. Tatikonda, S. Mitter, The Capacity of Channels with Feedback, IEEE Trans. Info. Theory, v.55, no. 1, pp. 323-349, Jan. 2009.

[11] 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.

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