Multiplexing Zero-Error and Rare-Error Communications over a Noisy Channel with Feedback

Tibor Keresztfalvi and Amos Lapidoth,
Signal and Information Processing Laboratory
ETH Zurich
Email: {keresztfalvi, lapidoth}@isi.ee.ethz.ch

***Abstract*—Two independent data streams—the "zero-error stream" and the "rare-error stream"—are to be transmitted over a noisy discrete memoryless channel with feedback. Errors are tolerated only in the rare-error stream, provided that their probability tends to zero. Clearly the rate of the error-free stream cannot exceed the channel's zero-error feedback capacity, and the sum of the streams' rates cannot exceed the channel's Shannon capacity. Using a suitable coding scheme, these necessary conditions are shown to characterize all the achievable rate pairs. Planning for the worst—as is needed to achieve zero-error communication—and planning for the true channel—as is needed to communicate near the Shannon limit—are thus not incompatible.**

***Index Terms*—Feedback, multiplexing, Shannon capacity, Zero-error capacity.**

I. Introduction

Two independent data streams of rates R_0 and R_ε are to be transmitted over a noisy discrete memoryless channel (DMC) $\mathsf{W}(y|x)$ with feedback subject to two different performance requirements. The decoder is required to reconstruct the "zero-error stream" error-free and the "rare-error stream" with a probability of error smaller than some prespecified (arbitrarily small, but positive) $\epsilon > 0$. We wish to characterize the pairs (R_0, R_ε) that can be supported by the channel.

If we set R_ε to zero, then R_0 can be as high as the channel's zero-error feedback capacity $C_{0,\text{FB}}$, which was computed by Shannon in his 1956 paper on feedback [1]–[4]. When positive, it is given in Shannon's form as

$$C_{0,\text{FB}} = \max_{\mathsf{Q}} \left\{ -\log \max_{y\in\mathcal{Y}} \sum_{x:\mathsf{W}(y|x)>0} \mathsf{Q}(x) \right\}, \tag{1}$$

or in Ahlswede's form [5]

$$C_{0,\text{FB}} = \max_{\mathsf{Q}} \min_{\mathsf{V}\ll\mathsf{W}} I(\mathsf{Q};\mathsf{V}), \tag{2}$$

where in both (1) and (2) Q is a probability mass function (PMF) on $\mathcal{X}$, and where in the minimization in (2) the channel V is over the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, and $\mathsf{V} \ll \mathsf{W}$ indicates that V is absolutely continuous with respect to the channel W in the sense that

$$\Big(\mathsf{W}(y|x) = 0\Big) \implies \Big(\mathsf{V}(y|x) = 0\Big), \quad (x,y) \in \mathcal{X}\times\mathcal{Y}. \tag{3}$$

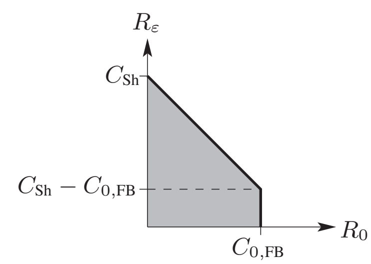

Fig. 1. An illustration of the region $\mathcal{C}(\mathsf{W})$.

If we set R_0 to zero, then R_ε can be as high as the Shannon capacity

$$C_{\text{Sh}} = \max_{\mathsf{Q}} I(\mathsf{Q};\mathsf{W}), \tag{4}$$

which Shannon derived in 1948 [6] and which is unaffected by feedback [1]. In fact, since both streams are *a fortiori* recovered with arbitrarily small probability of error, a necessary condition for the pair (R_0, R_ε) to be achievable is

$$R_0 + R_\varepsilon \le C_{\text{Sh}}. \tag{5a}$$

And since restricting the receiver to recover an additional stream cannot help,

$$R_0 \le C_{0,\text{FB}}. \tag{5b}$$

Here we will show that strict inequalities in (5) suffice to guarantee achievability and that Conditions (5) thus characterize the capacity region. A generic capacity region may thus look like the one depicted in Figure 1. This is true even when—as for the channel depicted in Figure 2—no single PMF Q on $\mathcal{X}$ achieves both the maximum in (1) and in (4). This result is formalized in the next section and is proved in Section III via a coding scheme. Section IV presents an instructive example, and Section V discusses the importance of the feedback link.

II. The Main Result

Consider a DMC with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and having the transition law $\mathsf{W}(y|x)$. Denote its Shannon capacity C_{Sh} $(= \max I(X;Y))$ [6] and its zero-error feedback capacity $C_{0,\text{FB}}$ [1]. Given a blocklength n, let $\mathcal{M}_0$

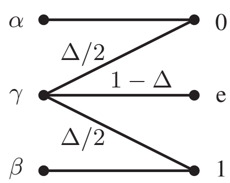

Fig. 2. When $0 < \Delta \ll 1$, the PMF Q maximizing $I(\mathsf{Q};\mathsf{W})$ is nearly uniform on $\{\alpha, \beta, \gamma\}$, whereas the Q achieving the maximum in (1) is uniform over $\{\alpha, \beta\}$.

denote a set of 2^{nR_0} messages from which some message m_0 is to be transmitted error free, and let $\mathcal{M}_\varepsilon$ denote a set of 2^{nR_ε} messages from which some message m_ε is to be transmitted with probability of error smaller than ϵ. A blocklength-n encoder with feedback for the two messages is a collection of n mappings

$$f_i \colon \mathcal{M}_0 \times \mathcal{M}_\varepsilon \times \mathcal{Y}^{i-1} \to \mathcal{X}, \quad i = 1, \dots, n \tag{6}$$

with the understanding that, to convey the pair (m_0, m_ε), the encoder produces at Time i the channel input

$$X_i\bigl(m_0, m_\varepsilon, Y^{i-1}\bigr) = f_i\bigl(m_0, m_\varepsilon, Y^{i-1}\bigr), \tag{7}$$

where Y^{i-1} stands for the vector $(Y_1, \dots, Y_{i-1})$ and denotes the $i-1$ channel outputs that are available to the encoder at Time i thanks to the feedback link. A decoder $\phi = (\phi_0, \phi_\varepsilon)$ is a pair of mappings

$$\phi_0 \colon \mathcal{Y}^n \to \mathcal{M}_0 \tag{8a}$$

$$\phi_\varepsilon \colon \mathcal{Y}^n \to \mathcal{M}_\varepsilon. \tag{8b}$$

We require that the message m_0 be transmitted error-free, i.e.,

$$\max_{m_0, m_\varepsilon} \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \phi_0(\mathbf{y}) \neq m_0}} \prod_{k=1}^{n} \mathsf{W}\bigl(y_k \big| x_k(m_0, m_\varepsilon, y^{k-1})\bigr) = 0. \tag{9}$$

The maximal probability of error associated with the rare-error stream is

$$\Pr(\text{error}) = \max_{m_0, m_\varepsilon} \sum_{\substack{\mathbf{y} \in \mathcal{Y}^n: \\ \phi_\varepsilon(\mathbf{y}) \neq m_\varepsilon}} \prod_{k=1}^{n} \mathsf{W}\bigl(y_k \big| x_k(m_0, m_\varepsilon, y^{k-1})\bigr). \tag{10}$$

We say that a rate pair (R_0, R_ε) is *achievable* if for every $\epsilon > 0$ there exists some $n_0(\epsilon)$ such that for all blocklengths n exceeding $n_0(\epsilon)$ there exist message sets as above and encoder/decoder pairs as above with rates $n^{-1} \log|\mathcal{M}_0| \geq R_0$ and $n^{-1} \log|\mathcal{M}_\varepsilon| \geq R_\varepsilon$ and with $\Pr(\text{error}) < \epsilon$. Here and throughout we use $|\mathcal{A}|$ to denote the cardinality of the set $\mathcal{A}$.

The *multiplexing capacity region* $\mathcal{C}(\mathsf{W})$ is the closure of the set of all achievable pairs.

Remark 1 ($\mathcal{C}(\mathsf{W})$ is a convex set)**.** *The multiplexing capacity region $\mathcal{C}(\mathsf{W})$ is a convex set containing the rate pairs $(0, C_{\mathrm{Sh}})$ and $(C_{0,\mathrm{FB}}, 0)$.*

Proof. The convexity can be established using a time-sharing argument. The details are omitted. □

We are now ready to state our main result.

Theorem 1 (The Multiplexing Capacity)**.** *The multiplexing capacity $\mathcal{C}(\mathsf{W})$ of a DMC W with feedback is the set of rate pairs (R_0, R_ε) satisfying* (5)*, where C_{Sh} is the channel's Shannon capacity and $C_{0,\mathrm{FB}}$ is the channel's zero-error feedback capacity.*

To prove this theorem, it suffices to consider the case where $C_{0,\mathrm{FB}}$ is positive. In this case the zero-error capacity is also positive in the absence of feedback and there exist two inputs $x', x'' \in \mathcal{X}$ such that the product $\mathsf{W}(y|x')\,\mathsf{W}(y|x'')$ is zero for all possible outputs $y \in \mathcal{Y}$ [1]. Such inputs can be used to send a single bit error free in one channel use.

At the end of the transmission the transmitter (which knows the state of the receiver via the feedback link) can send a bit indicating whether the receiver's guess for m_ε is correct or not. In the latter case the receiver can declare an erasure for m_ε and in this way avoid any errors. Consequently, Theorem 1 can be strengthened to address the *erasure-only* requirement as follows:

Remark 2. *The rate pairs in Theorem 1 can also be achieved if we do not allow errors in guessing m_ε and instead allow erasures (with arbitrarily small but positive probability).*

(For more on the erasures-only requirement, see [7]–[14] and references therein.)

Theorem 1 clearly holds when $C_{0,\mathrm{FB}}$ equals C_{Sh}. Hence, we shall henceforth focus on channels for which

$$0 < C_{0,\mathrm{FB}} < C_{\mathrm{Sh}}. \tag{11}$$

In Section III we will present a coding scheme that can achieve the pair $(C_{0,\mathrm{FB}} - \delta, C_{\mathrm{Sh}} - C_{0,\mathrm{FB}} - \delta)$ whenever δ is positive. This will demonstrate that the pair $(C_{0,\mathrm{FB}}, C_{\mathrm{Sh}} - C_{0,\mathrm{FB}})$ is in $\mathcal{C}(\mathsf{W})$ and hence, by Remark 1, also all the rate pairs satisfying (5).[1]

III. A Coding Scheme

In this section we fix some $\delta > 0$ and describe a coding scheme that achieves the pair $(C_{0,\mathrm{FB}} - \delta, C_{\mathrm{Sh}} - C_{0,\mathrm{FB}} - \delta)$. We denote the blocklength by n, and the maximal-allowed probability of error in decoding the rare-error stream by $\epsilon > 0$.

The zero-error message set $\mathcal{M}_0$ will have the form

$$\mathcal{M}_0 = \mathcal{M}_{0,1} \times \cdots \times \mathcal{M}_{0,\mathsf{K}}, \tag{12a}$$

and a generic zero-error message m_0 the form

$$m_0 = \bigl(m_{0,1}, \dots, m_{0,\mathsf{K}}\bigr), \tag{12b}$$

where

$$m_{0,\kappa} \in \mathcal{M}_{0,\kappa}, \quad \kappa \in \{1, \dots, \mathsf{K}\}. \tag{12c}$$

[1]The example in Section IV shows that—while always in $\mathcal{C}(\mathsf{W})$—the pair $(C_{0,\mathrm{FB}}, C_{\mathrm{Sh}} - C_{0,\mathrm{FB}})$ need not be achievable. Backing off by an arbitrarily small but positive δ may be necessary.

We refer to $m_{0,\kappa}$ as the κ-th zero-error packet and to K as the number of packets. The zero-error packets are of equal size, and we denote the number of values each of them can take by $\tilde{\mathsf{M}}_0$,

$$\tilde{\mathsf{M}}_0 = |\mathcal{M}_{0,\kappa}| \quad \kappa \in \{1, \dots, \mathsf{K}\}. \tag{12d}$$

The zero-error rate R_0 of the scheme is thus

$$R_0 = \frac{1}{n} \mathsf{K} \log \tilde{\mathsf{M}}_0. \tag{12e}$$

The rare-error message m_ε is also divided into K packets, so

$$\mathcal{M}_\varepsilon = \mathcal{M}_{\varepsilon,1} \times \cdots \times \mathcal{M}_{\varepsilon,\mathsf{K}}, \tag{13a}$$

with the generic rare-error message m_ε having the form

$$m_\varepsilon = \bigl(m_{\varepsilon,1}, \dots, m_{\varepsilon,\mathsf{K}}\bigr), \tag{13b}$$

where

$$m_{\varepsilon,\kappa} \in \mathcal{M}_{\varepsilon,\kappa}, \quad \kappa \in \{1, \dots, \mathsf{K}\} \tag{13c}$$

is the κ-th rare-error packet. All the rare-error packets are of equal size $\tilde{\mathsf{M}}_\varepsilon$ (possibly different from the size of the zero-error packets),

$$\tilde{\mathsf{M}}_\varepsilon = |\mathcal{M}_{\varepsilon,\kappa}| \quad \kappa \in \{1, \dots, \mathsf{K}\}. \tag{13d}$$

The scheme's rare-error rate is thus

$$R_\varepsilon = \frac{1}{n} \mathsf{K} \log \tilde{\mathsf{M}}_\varepsilon. \tag{13e}$$

The transmission is split into $\mathsf{K}+1$ phases, each comprising $\eta+1$ channel uses, except for the last phase whose length η_F is allowed to exceed $\eta+1$ (by at most η) in order to account for blocklengths that are not divisible by $(\mathsf{K}+1)$. Thus, given a number K of packets and a blocklength n (exceeding K),

$$\eta + 1 = \left\lfloor \frac{n}{\mathsf{K}+1} \right\rfloor, \tag{14a}$$

and

$$\eta_F = n - \mathsf{K}(\eta + 1), \tag{14b}$$

so

$$\eta + 1 \le \eta_F < 2\eta + 2. \tag{14c}$$

In Phase 1 we use the first η channel uses to transmit the pair $m_{0,1}, m_{\varepsilon,1}$ using a code of maximal probability of error smaller than ϵ/K. The $(\eta+1)$-th channel use is utilized as a flag to indicate to the receiver whether or not the tentative decision that the receiver forms based on the phase's first η channel outputs, is correct. (Thanks to the feedback link, the transmitter is cognizant of these channel outputs, and by applying the receiver's decoding mappings to these outputs, it can compute the receiver's tentative decision. Since we are considering the case where $C_{0,\text{FB}}$ is positive, there exist inputs $x', x'' \in \mathcal{X}$ that can be used to send a single bit error free. The transmitter can use such inputs to indicate whether or not the tentative decision is correct.) If not, the receiver discards the tentative decision and we say that a transmission failure occurred in Phase 1. Else, the receiver turns the tentative decision into a final one and we move on to send the pair $m_{0,2}, m_{\varepsilon,2}$ in Phase 2 in a similar fashion.

This continues until either a failure occurs at some Phase $\kappa \le \mathsf{K}$ or until Phase K terminates without failure. In the latter case m_0 and m_ε were transmitted error-free in the first K phases, and what we send in the last phase is immaterial. In the former case—which occurs by the union-of-events bound with probability smaller than ϵ—we give up on sending m_ε and we focus on m_0. Since no failures occurred prior to Phase κ, the first $\kappa-1$ zero-error packets $m_{0,1}, \dots, m_{0,\kappa-1}$ were transmitted error free. The remaining $(\mathsf{K}-\kappa+1)$ zero-error packets $m_{0,\kappa}, \dots, m_{0,\mathsf{K}}$ are then transmitted error-free, using a zero-error feedback scheme, in the remaining $\mathsf{K}-\kappa+1$ phases, namely Phases $(\kappa+1)$ through $(\mathsf{K}+1)$. Here it is critical that the number of phases $(\mathsf{K}+1)$ exceeds the number of packets (K).

We next describe the choice of the scheme's parameters, beginning with K. Its choice will depend only on C_{Sh}, $C_{0,\text{FB}}$, and δ. Once chosen, it will be held fixed as we let the blocklength tend to infinity. We choose K large enough so that the rate-loss resulting from the last phase, which does not always convey information, be negligible. The $\eta/(\eta+1)$ rate loss resulting from the flags will vanish when n and hence also η will tend to infinity. We thus choose K so that R_0 and R_ε of (12e) and (13e) be only slightly lower than $\eta^{-1} \log \tilde{\mathsf{M}}_0$ and $\eta^{-1} \log \tilde{\mathsf{M}}_\varepsilon$ respectively. More specifically, starting from (12e),

$$\begin{aligned} R_0 &= \frac{\log \tilde{\mathsf{M}}_0}{\eta} \frac{\mathsf{K}\eta}{n} \\ &= \frac{\log \tilde{\mathsf{M}}_0}{\eta} \frac{n - \mathsf{K} - \eta_F}{n} \\ &> \frac{\log \tilde{\mathsf{M}}_0}{\eta} \Bigl(1 - \frac{\mathsf{K}}{n} - \frac{2\eta+2}{n}\Bigr) \\ &> \frac{\log \tilde{\mathsf{M}}_0}{\eta} \Bigl(1 - \frac{\mathsf{K}}{n} - \frac{2}{\mathsf{K}+1} - \frac{2}{n}\Bigr) \\ &= \frac{\log \tilde{\mathsf{M}}_0}{\eta} \Bigl(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Bigr), \end{aligned} \tag{15a}$$

where the second line follows from (14b); the third from (14c); and the fourth from (14a) and the inequality $\lfloor \xi \rfloor > \xi - 1$. Analogously,

$$R_\varepsilon > \frac{\log \tilde{\mathsf{M}}_\varepsilon}{\eta} \Bigl(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Bigr). \tag{15b}$$

We choose K sufficiently large so that whenever n exceeds some n_0

$$\Bigl(C_{0,\text{FB}} - \frac{\delta}{2}\Bigr)\Bigl(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Bigr) \ge C_{0,\text{FB}} - \delta \tag{16a}$$

and

$$\Bigl(C_{\text{Sh}} - C_{0,\text{FB}} - \frac{\delta}{2}\Bigr)\Bigl(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Bigr) \ge C_{\text{Sh}} - C_{0,\text{FB}} - \delta. \tag{16b}$$

We choose $\tilde{\mathsf{M}}_0$ and $\tilde{\mathsf{M}}_\varepsilon$ based on C_{Sh}, $C_{0,\text{FB}}$, δ, and η so that (for all sufficiently large η)

$$C_{0,\text{FB}} - \frac{\delta}{2} < \frac{\log \tilde{\mathsf{M}}_0}{\eta} < C_{0,\text{FB}} - \frac{\delta}{4} \tag{17a}$$

and

$$C_{\text{Sh}} - \frac{\delta}{2} < \frac{\log(\tilde{\mathsf{M}}_0 \, \tilde{\mathsf{M}}_\varepsilon)}{\eta} < C_{\text{Sh}} - \frac{\delta}{4}, \tag{17b}$$

which combine to imply that

$$\begin{aligned} \frac{\log \tilde{\mathsf{M}}_\varepsilon}{\eta} &> C_{\text{Sh}} - C_{0,\text{FB}} - \frac{\delta}{4} \\ &> C_{\text{Sh}} - C_{0,\text{FB}} - \frac{\delta}{2}. \end{aligned} \tag{17c}$$

Here (17a) guarantees that—when n is sufficiently large so as to imply by (14a) that η is sufficiently large—each zero-error packet is transmittable error-free in η channel uses using the feedback link. Similarly, (17b) guarantees that, for sufficiently large blocklengths, each pair of zero-error and rare-error packets is transmittable in η channel uses with maximal probability of error smaller that ϵ/K.

As to the scheme's rates, note that by (15), (16), and (17), the rates of our scheme satisfy (for all sufficiently large blocklengths)

$$R_0 > C_{0,\text{FB}} - \delta, \tag{18a}$$

$$R_\varepsilon > C_{\text{Sh}} - C_{0,\text{FB}} - \delta. \tag{18b}$$

This concludes the proof of the achievability part of Theorem 1.

IV. Example

An example of a channel where the input distributions achieving (1) and (4) differ is depicted in Figure 2. When $0 < \Delta \ll 1$, the Shannon capacity is nearly $\log 3$, and it is achieved by an input distribution that is nearly uniform over the entire input alphabet $\mathcal{X} = \{\alpha, \beta, \gamma\}$. The zero-error feedback capacity, however, is $\log 2$, and the unique PMF Q that achieves the maximum in (1) avoids the input γ and is uniform over $\{\alpha, \beta\}$. Nevertheless, Theorem 1 promises that the pair

$$(R_0, R_\varepsilon) = (\log 2 - \delta, \log 3 - \log 2 - \delta)$$

is achievable for any $\delta > 0$. Backing off by δ is crucial: as we next show, if—as opposed to $2^{n(1-\delta)}$—we insist that $\mathcal{M}_0$ be of size 2^n, and if the tolerated rare-error probability ϵ is smaller than $1/2$, then R_ε must be zero. In fact, $\mathcal{M}_\varepsilon$ cannot contain more than one message.

Proof. Let the decoding set $\mathcal{D}(m_0) \subseteq \mathcal{Y}^n$ comprise the output sequences that result in the zero-error decoder declaring that m_0 was sent. We claim that for each $m_0 \in \mathcal{M}_0$, the set $\mathcal{D}(m_0)$ contains exactly one sequence from $\{0,1\}^n$. To see why, fix some message $m_\varepsilon^\star$ from $\mathcal{M}_\varepsilon$, and let $\mathbf{y}'(m_0, m_\varepsilon^\star)$ be the output sequence that results when the pair $(m_0, m_\varepsilon^\star)$ is transmitted and the channel produces the output 0 whenever it is fed the input symbol γ. (Such a channel behavior occurs with positive probability, because $\Delta > 0$.) The sequence $\mathbf{y}'(m_0, m_\varepsilon^\star)$ is in $\mathcal{D}(m_0)$ (to avoid an error in recovering m_0) and is also in $\{0,1\}^n$ (because we assumed that the channel produces the output 0 whenever γ is transmitted). Thus, for each $m_0 \in \mathcal{M}_0$, the decoding set $\mathcal{D}(m_0)$ contains *at least one* sequence from $\{0,1\}^n$. Since there are 2^n sequences in $\{0,1\}^n$, there are $|\mathcal{M}_0|$ $(= 2^n)$ decoding sets, and since the decoding sets are disjoint, each decoding set must contain *exactly one* sequence from $\{0,1\}^n$.

Define $\mathbf{y}''(m_0, m_\varepsilon^\star)$ analogously to $\mathbf{y}'(m_0, m_\varepsilon^\star)$, but with the channel now producing the output 1 whenever it is fed the input symbol γ. Like $\mathbf{y}'(m_0, m_\varepsilon^\star)$, it is $\{0,1\}^n$-valued and must be in $\mathcal{D}(m_0)$. Since $\mathcal{D}(m_0)$ can contain only one such sequence, the two must be the same, which is only possible if γ is never transmitted. We conclude that, with probability one, γ is never transmitted and the output sequence is thus $\{0,1\}^n$-valued. To guarantee that it be in $\mathcal{D}(m_0)$, it must equal $\mathbf{y}'(m_0, m_\varepsilon^\star)$, so the Time-$i$ channel input must be α whenever the i-th component of $\mathbf{y}'(m_0, m_\varepsilon^\star)$ is 0 and β when it is 1. This is also true if the rare-error message m_ε we wish to send is not $m_\varepsilon^\star$! The sequence transmitted to convey the pair (m_0, m_ε) does not depend on m_ε. When the tolerated probability of error ϵ in the recovery of m_ε is smaller than $1/2$, this implies that $\mathcal{M}_\varepsilon$ cannot contain more than one rare-error message. □

V. How Crucial Is Feedback?

The multiplexing scheme we presented in Section III relies heavily on the feedback link. And well it should, because, in general, the region $\mathcal{C}(\mathsf{W})$ cannot be achieved without feedback. Indeed, since feedback can increase the zero-error capacity, the pair $(C_{0,\text{FB}}, 0)$ need not be achievable without feedback. But what about channels, such as that of Figure 2, whose zero-error capacity is not increased by feedback? Can feedback still increase the multiplexing capacity? The answer is "yes." This can be shown using the following simple outer bound on the multiplexing capacity in the absence of feedback [15].

Proposition 2. *If (R_0, R_ε) can be multiplexed on the channel W without feedback, then for every channel $\mathsf{V} \ll \mathsf{W}$ there must correspond some auxiliary chance-variable U and some joint distribution $p_{U,X}$ such that*

$$R_0 \le I(U;Z) \tag{19a}$$

$$R_\varepsilon \le I(X;Y|U) \tag{19b}$$

$$R_0 + R_\varepsilon \le I(X;Y), \tag{19c}$$

where the mutual informations are computed with respect to the joint distribution

$$p_{U,X,Y,Z}(u,x,y,z) = p_{U,X}(u,x)\,\mathsf{W}(y|x)\,\mathsf{V}(z|x), \tag{19d}$$

and where the cardinality of the set $\mathcal{U}$ in which U takes values can be restricted to satisfy

$$|\mathcal{U}| \le \min\{|\mathcal{X}|, 2\,|\mathcal{Y}|\} + 1. \tag{19e}$$

In fact, for the channel in Figure 2 (with $\Delta \ll 1$) the multiplexing capacity without feedback is achieved by time-sharing between the rate pairs $(0, C_{\text{Sh}})$ and $(C_{0,\text{FB}}, 0)$ [15].

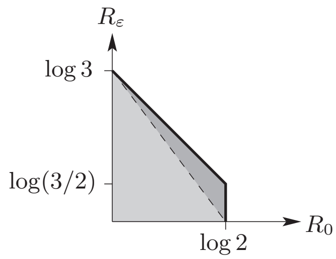

Fig. 3. The Multiplexing Region of the channel in Figure 2 when $\Delta \downarrow 0$ with and without feedback . Lightly shaded is the region without feedback.

For this channel, the multiplexing capacities with and without feedback are thus as depicted in Figure 3, with the latter being strictly included in the former. We prove this in [15] using the following more elaborate outer bound on the multiplexing capacity without feedback.

For finite sets $\mathcal{U}$ and $\mathcal{Q}$, let $\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})$ denote the set of conditional PMFs $\mathsf{V}(y|u,x,q)$ for which the implication

$$\big(\mathsf{W}(y|x)=0\big) \implies \big(\mathsf{V}(y|u,x,q)=0\big) \tag{20}$$

holds for every $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $u \in \mathcal{U}$, and $q \in \mathcal{Q}$.

Proposition 3. *If (R_0, R_ε) can be multiplexed without feedback on the channel W, then for some PMF $p_{U,X,Q}$*

$$R_0 \le \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), \tag{21a}$$

$$R_\varepsilon \le I(X;Y|U,Q), \tag{21b}$$

where the mutual informations are computed w.r.t. the joint PMF

$$p_{U,X,Y,Z,Q}(u,x,y,z,q) = p_{U,X,Q}(u,x,q)\,\mathsf{W}(y|x)\,\mathsf{V}(z|u,x,q). \tag{21c}$$

This bound need not be tight. Computing the multiplexing capacity without feedback is probably very difficult, because it is at least as difficult as computing the zero-error capacity without feedback.

References

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 2, pp. 8–19, 1956.

[2] P. Elias, "Zero error capacity under list decoding," *IEEE Trans. on Inform. Theory*, vol. 34, no. 5, pp. 1070–1074, Sept. 1988.

[3] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2207–2229, 1998.

[4] L. Zhao and H. H. Permuter, "Zero-error feedback capacity of channels with state information via dynamic programming," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2640–2650, June 2010.

[5] R. Ahlswede, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift f. Wahrscheinlichkeitstheorie and verw. Gebiete*, vol. 25, no. 3, pp. 239–252, Sep. 1973.

[6] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[7] G. D. Forney, "Exponential error bounds for erasure, list and decision feedback systems," *IEEE Trans. on Inform. Theory*, vol. 14, no. 2, pp. 206–220, March 1968.

[8] M. Pinsker and A. Y. Sheverdyaev, "Transmission capacity with zero error and erasure," *Probl. Inform. Transm. (Probl. Pered. Inform.),*, vol. 6, no. 1, pp. 20–24, 1970.

[9] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. on Inform. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[10] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. on Inform. Theory*, vol. 42, no. 1, pp. 55–62, Jan 1996.

[11] İ. E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1977–1982, Nov. 1997.

[12] C. Bunte and A. Lapidoth, "The zero-undetected-error capacity of discrete memoryless channels with feedback," in *Proceedings 50th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 1–5, 2012*, 2012.

[13] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. on Inform. Theory*, vol. 58, no. 1, pp. 24–49, Jan. 2012.

[14] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity approaches the Sperner capacity," *IEEE Trans. Information Theory*, vol. 60, no. 7, pp. 3825–3833, 2014.

[15] T. Keresztfalvi and A. Lapidoth, "Multiplexing zero-error and rare-error communications over a noisy channel," *In preparation*, 2017.