Multiplexing Zero-Error and Rare-Error Communications Over a Noisy Channel

Tibor Keresztfalvi and Amos Lapidoth, *Fellow, IEEE*

***Abstract*— Two independent data streams are to be transmitted over a noisy discrete memoryless channel with noiseless (ideal) feedback. Errors are tolerated only in the second stream, provided that they occur with vanishing probability. The rate of the error-free stream cannot, of course, exceed the channel's zero-error feedback capacity, and nor can the sum of the streams' rates 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 channel behavior—as is needed to achieve zero-error communication—and planning for the typical channel behavior—as is needed to communicate near the Shannon limit—are thus not incompatible. It is further shown that feedback may be beneficial for the multiplexing problem even on channels on which it does not increase the zero-error capacity.**

***Index Terms*— Broadcast channel, erasures-only, 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 noiseless (ideal) feedback subject to two different performance requirements. The decoder is required to reconstruct the first stream error-free and the second with 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], [2]. When positive, it is given in Shannon's form as[1]

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

or in Ahlswede's form [3]

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

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

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

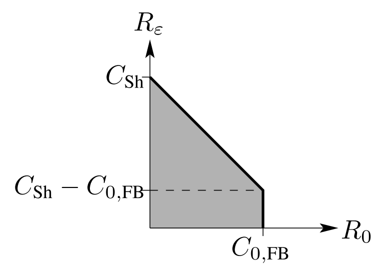

Fig. 1. An illustration of the region $\mathcal{C}_{\text{FB}}(\mathsf{W})$.

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 [4] 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 the conditions in (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. The scheme relies heavily on the feedback link, and for a good reason: the multiplexing capacity region with feedback need not be achievable without feedback even for channels whose zero-error feedback capacity is achievable without feedback. This is shown in Section IV, which addresses the no-feedback case.

Manuscript received March 22, 2018; revised October 2, 2018; accepted November 6, 2018. Date of publication November 16, 2018; date of current version April 19, 2019. This paper was presented at the 2017 International Symposium on Information Theory.

The authors are with ETH Zurich, 8092 Zurich, Switzerland (e-mail: tibork@isi.ee.ethz.ch; lapidoth@isi.ee.ethz.ch).

Communicated by M. Costa, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2018.2881758

[1] All logarithms in this paper are base-2 logarithms, and all rates are in bits per channel use.

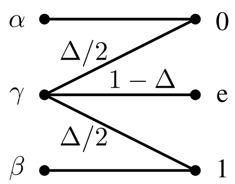

Fig. 2. When $0 < \Delta \ll 1$, the PMF Q maximizing $I(\mathsf{Q};\mathsf{W})$ is nearly uniform over $\{\alpha, \beta, \gamma\}$, whereas the Q achieving the maximum in (1) is uniform over $\{\alpha, \beta\}$.

II. The Main Result

Consider a DMC with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and of transition law $\mathsf{W}(y|x)$. Denote its Shannon capacity C_{Sh} or $C_{\text{Sh}}(\mathsf{W})$ $(= \max I(X;Y))$ [4] and its zero-error feedback capacity $C_{0,\text{FB}}$ or $C_{0,\text{FB}}(\mathsf{W})$ [1]. Given a blocklength n, let $\mathcal{M}_0$ 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

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

where Y^{i-1} denotes the tuple $(Y_1, \dots, Y_{i-1})$ and stands for the $i-1$ channel outputs that, thanks to the feedback link, are available to the encoder at Time i. 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.,

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

The maximal probability of error $\lambda_{\max}$ associated with the rare-error stream is

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

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 $\lambda_{\max} < \epsilon$. Here and throughout we use $|\mathcal{A}|$ to denote the cardinality of the set $\mathcal{A}$. The *multiplexing capacity region* $\mathcal{C}_{\text{FB}}(\mathsf{W})$ is the closure of the set of all achievable pairs.

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

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 2 (The Multiplexing Capacity): The multiplexing capacity $\mathcal{C}_{\text{FB}}(\mathsf{W})$ 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 $C_{0,\text{FB}}$ is the channel's zero-error feedback capacity.*

To prove this theorem, it suffices to consider the case where $C_{0,\text{FB}}$ 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 outputs $y \in \mathcal{Y}$ [1]. Such inputs can be used to send a single bit error free in one channel use. Theorem 2 clearly holds when $C_{0,\text{FB}}$ equals C_{Sh}, so we shall henceforth focus on channels for which

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

In Section III we will present a coding scheme that can achieve the pair $(C_{0,\text{FB}} - \delta, C_{\text{Sh}} - C_{0,\text{FB}} - \delta)$ whenever δ is positive (and the pair is positive). This will demonstrate that the pair $(C_{0,\text{FB}}, C_{\text{Sh}} - C_{0,\text{FB}})$ is in $\mathcal{C}_{\text{FB}}(\mathsf{W})$

$$(C_{0,\text{FB}}, C_{\text{Sh}} - C_{0,\text{FB}}) \in \mathcal{C}_{\text{FB}}(\mathsf{W}) \tag{12}$$

and hence, by Remark 1, also all the rate pairs satisfying (5).[2]

III. A Coding Scheme

In this section we fix some $\delta > 0$ and describe a coding scheme that achieves the pair $(C_{0,\text{FB}} - \delta, C_{\text{Sh}} - C_{0,\text{FB}} - \delta)$. 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

$$\mathcal{M}_0 = \mathcal{M}_{0,1} \times \cdots \times \mathcal{M}_{0,\mathsf{K}} \tag{13a}$$

and a generic zero-error message m_0 the form

$$m_0 = \big(m_{0,1}, \dots, m_{0,\mathsf{K}}\big), \tag{13b}$$

where

$$m_{0,\kappa} \in \mathcal{M}_{0,\kappa}, \quad \kappa \in \{1, \dots, \mathsf{K}\}. \tag{13c}$$

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 $\tilde{\mathsf{M}}_0$,

$$\tilde{\mathsf{M}}_0 = |\mathcal{M}_{0,\kappa}| \quad \kappa \in \{1, \dots, \mathsf{K}\}. \tag{13d}$$

The zero-error rate R_0 of the scheme is thus

$$R_0 = \frac{1}{n} \mathsf{K} \log \tilde{\mathsf{M}}_0. \tag{13e}$$

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

$$\mathcal{M}_\varepsilon = \mathcal{M}_{\varepsilon,1} \times \cdots \times \mathcal{M}_{\varepsilon,\mathsf{K}}, \tag{14a}$$

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

with the generic rare-error message m_ε having the form

$$m_\varepsilon = \big(m_{\varepsilon,1}, \ldots, m_{\varepsilon,\mathsf{K}}\big), \tag{14b}$$

where

$$m_{\varepsilon,\kappa} \in \mathcal{M}_{\varepsilon,\kappa}, \quad \kappa \in \{1, \ldots, \mathsf{K}\} \tag{14c}$$

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

$$\tilde{\mathsf{M}}_\varepsilon = |\mathcal{M}_{\varepsilon,\kappa}| \quad \kappa \in \{1, \ldots, \mathsf{K}\}. \tag{14d}$$

The scheme's rare-error rate is thus

$$R_\varepsilon = \frac{1}{n}\, \mathsf{K} \log \tilde{\mathsf{M}}_\varepsilon. \tag{14e}$$

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),

$$\eta + 1 = \left\lfloor \frac{n}{\mathsf{K}+1} \right\rfloor \tag{15a}$$

and

$$\eta_F = n - \mathsf{K}(\eta + 1), \tag{15b}$$

so

$$\eta + 1 \le \eta_F < 2\eta + 2. \tag{15c}$$

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. For large η this is possible if

$$\frac{1}{\eta} \log\big(\tilde{\mathsf{M}}_0 \cdot \tilde{\mathsf{M}}_\varepsilon\big) < C_{\mathrm{Sh}} - \frac{\delta}{4}. \tag{16}$$

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 $C_{0,\mathrm{FB}}$ 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 enter "panic mode," where we give up on sending m_ε and focus only on m_0. Since no failures occurred prior to Phase κ, the first $\kappa - 1$ zero-error packets $m_{0,1}, \ldots, m_{0,\kappa-1}$ were transmitted error free. The remaining $(\mathsf{K} - \kappa + 1)$ zero-error packets $m_{0,\kappa}, \ldots, m_{0,\mathsf{K}}$ 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). The panic mode is guaranteed to succeed if η is large enough and

$$\frac{1}{\eta} \log \tilde{\mathsf{M}}_0 < C_{0,\mathrm{FB}} - \frac{\delta}{4}. \tag{17}$$

We next sketch the choice of the scheme's parameters, beginning with K. Its choice will depend only on C_{Sh}, $C_{0,\mathrm{FB}}$, 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 (13e) and (14e) 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 (13e),

$$\begin{aligned} R_0 &= \Big(\frac{\log \tilde{\mathsf{M}}_0}{\eta}\Big)\Big(\frac{\mathsf{K}\eta}{n}\Big) \\ &= \Big(\frac{\log \tilde{\mathsf{M}}_0}{\eta}\Big)\Big(\frac{n - \mathsf{K} - \eta_F}{n}\Big) \\ &> \frac{\log \tilde{\mathsf{M}}_0}{\eta}\Big(1 - \frac{\mathsf{K}}{n} - \frac{2\eta + 2}{n}\Big) \\ &> \frac{\log \tilde{\mathsf{M}}_0}{\eta}\Big(1 - \frac{\mathsf{K}}{n} - \frac{2}{\mathsf{K}+1} - \frac{2}{n}\Big) \\ &= \frac{\log \tilde{\mathsf{M}}_0}{\eta}\Big(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Big), \end{aligned} \tag{18a}$$

where the second line follows from (15b); the third from (15c); and the fourth from (15a) and the inequality $\lfloor \xi \rfloor > \xi - 1$. Analogously,

$$R_\varepsilon > \frac{\log \tilde{\mathsf{M}}_\varepsilon}{\eta}\Big(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Big). \tag{18b}$$

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

$$\Big(C_{0,\mathrm{FB}} - \frac{\delta}{2}\Big)\Big(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Big) \ge C_{0,\mathrm{FB}} - \delta \tag{19a}$$

and

$$\Big(C_{\mathrm{Sh}} - C_{0,\mathrm{FB}} - \frac{\delta}{2}\Big)\Big(1 - \frac{\mathsf{K}+2}{n} - \frac{2}{\mathsf{K}+1}\Big) \ge C_{\mathrm{Sh}} - C_{0,\mathrm{FB}} - \delta. \tag{19b}$$

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

$$C_{0,\mathrm{FB}} - \frac{\delta}{2} < \frac{\log \tilde{\mathsf{M}}_0}{\eta} < C_{0,\mathrm{FB}} - \frac{\delta}{4} \tag{20a}$$

(with the upper bound thus guaranteeing (17)) and

$$C_{\mathrm{Sh}} - \frac{\delta}{2} < \frac{\log(\tilde{\mathsf{M}}_0\, \tilde{\mathsf{M}}_\varepsilon)}{\eta} < C_{\mathrm{Sh}} - \frac{\delta}{4}, \tag{20b}$$

(with the upper bound thus guaranteeing (16)) which together imply that

$$\frac{\log \tilde{\mathsf{M}}_\varepsilon}{\eta} > C_{\text{Sh}} - C_{0,\text{FB}} - \frac{\delta}{4} > C_{\text{Sh}} - C_{0,\text{FB}} - \frac{\delta}{2}. \tag{20c}$$

Here (20a) guarantees that—when n is sufficiently large so as to imply by (15a) that η is sufficiently large—each zero-error packet be transmittable error-free in η channel uses using the feedback link. Similarly, (20b) guarantees that, for sufficiently large blocklengths, each pair of zero-error and rare-error packets be transmittable in η channel uses with maximal probability of error smaller that ϵ/K.

As to the scheme's rates, the inequalities (18), (19), and (20) imply that the rates of our scheme satisfy (for all sufficiently large blocklengths)

$$R_0 > C_{0,\text{FB}} - \delta, \tag{21a}$$

$$R_\varepsilon > C_{\text{Sh}} - C_{0,\text{FB}} - \delta. \tag{21b}$$

This concludes the proof of the achievability part of Theorem 2.

IV. No Feedback

Also of interest is the multiplexing capacity in the absence of feedback, $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$, which is defined like $\mathcal{C}_{\text{FB}}(\mathsf{W})$ but with the feedback encoder (22) replaced by its no-feedback counterpart

$$f: \mathcal{M}_0 \times \mathcal{M}_\varepsilon \to \mathcal{X}^n. \tag{22}$$

Computing $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$ is at least as difficult as computing the zero-error capacity in the absence of feedback, which is a longstanding open problem [1], [5], [6]. Here we shall present two outer bounds. Those will show, *inter alia*, that $\mathcal{C}_{\text{FB}}(\mathsf{W})$ can be strictly larger than $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$ even when the zero-error capacities with and without feedback are identical.

The first bound, the "simple outer bound," is based on a connection that we establish next between our problem and Körner and Marton's work on the broadcast channel with degraded message sets [7], [8, Sec. 8.1]. Let $\mathsf{W}(y|x)$ and $\mathsf{V}(y|x)$ be DMCs over the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. If an encoder/decoder pair never errs on the channel W, and if $\mathsf{V} \ll \mathsf{W}$, then the pair will also not err on the channel V, because feeding any $\mathbf{x}$ to V can only produce outputs that could have also been produced had $\mathbf{x}$ been fed to W. This observation allows us to obtain outer bounds on $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$ in terms of the capacity region of an auxiliary broadcast channel that is constructed from W and V as follows: its input X takes values in the set $\mathcal{X}$; its two outputs Y, Z take values in the sets $\mathcal{Y}$ and $\mathcal{Z}$, with $\mathcal{Z}$ being identical to $\mathcal{Y}$; and—conditional on the transmitted symbol being $x \in \mathcal{X}$—the output Y is distributed according to $\mathsf{W}(y|x)$ and the output Z according to $\mathsf{V}(z|x)$. If the pair (m_0, m_ε) can be multiplexed (without feedback) on the channel W (with m_0 recovered error-free and m_ε with probability of error smaller than ϵ), then the pair can also be transmitted over the broadcast channel (without feedback) with the observer of Y recovering the pair (with probability of error smaller than ϵ) and the observer of Z recovering m_0. Any pair (R_0, R_ε) that can be multiplexed on W without feedback must hence lie in the *degraded message sets* capacity region $\mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V})$ of this broadcast channel [7], [8, Sec. 8.1].

This region is convex and equals the union, over all choices of the auxiliary chance-variable U and over all the joint distributions $p_{U,X}$, of the set of pairs (R_0, R_ε) satisfying

$$R_0 \le I(U; Z) \tag{23a}$$

$$R_\varepsilon \le I(X; Y|U) \tag{23b}$$

$$R_0 + R_\varepsilon \le I(X; Y), \tag{23c}$$

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

$$p_{U,X,Y,Z}(u, x, y, z) = p_{U,X}(u, x)\,\mathsf{W}(y|x)\,\mathsf{V}(z|x), \tag{24}$$

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

$$|\mathcal{U}| \le \min\{|\mathcal{X}|, 2\,|\mathcal{Y}|\} + 1. \tag{25}$$

Alternatively, $\mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V})$ can be expressed as the union over all choices of $p_{U,X}$ of the set of rate pairs (R_0, R_ε) that satisfy

$$R_0 \le \min\{I(U; Y), I(U; Z)\} \tag{26a}$$

$$R_\varepsilon \le I(X; Y|U), \tag{26b}$$

where the mutual informations are computed w.r.t. the joint PMF of (24). This alternative characterization is due to Nair and El-Gamal [9]. (A proof of equivalence can by found in [10].)

Any choice of V for which $\mathsf{V} \ll \mathsf{W}$ yields the outer bound

$$\mathcal{C}_{\text{No-FB}}(\mathsf{W}) \subseteq \mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}), \quad \mathsf{V} \ll \mathsf{W}. \tag{27}$$

Intersecting these outer bounds yields the simple outer bound:

Proposition 3 (Simple Outer Bound: No Feedback): In the absence of feedback, the multiplexing capacity region $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$ *of the DMC* W *satisfies*

$$\mathcal{C}_{\text{No-FB}}(\mathsf{W}) \subseteq \mathcal{R}_{\text{Sim}}(\mathsf{W}), \tag{28}$$

where

$$\mathcal{R}_{\text{Sim}}(\mathsf{W}) = \bigcap_{\mathsf{V} \ll \mathsf{W}} \mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}). \tag{29}$$

One could ostensibly improve this bound by considering a number of auxiliary channels concurrently and by studying the rates at which one can communicate reliably with all of them simultaneously. This would preclude the coding scheme from depending on the auxiliary channel V: we would have to use the same scheme for all the channels under consideration. In the limit, we could insist that the scheme allow reliable communication to all the members of the family of channels that are absolutely continuous with respect to W. This, however, does not improve the bound, because the latter family is convex [11]. To improve on Proposition 3 we shall need a different approach, which is described in Section V.

The rest of this section is dedicated to proving that feedback can increase the multiplexing region even when it does not increase the zero-error capacity. To this end, we first provide an alternative characterization of the zero-error feedback capacity $C_{0,\text{FB}}$. We do so by swapping the max and the min in Ahlswede's expression (2) and expressing $C_{0,\text{FB}}$, when positive, as

$$C_{0,\text{FB}} = \min_{\mathsf{V} \ll \mathsf{W}} \max_{\mathsf{Q}} I(\mathsf{Q}; \mathsf{V}) \tag{30}$$
$$= \min_{\mathsf{V} \ll \mathsf{W}} C_{\text{Sh}}(\mathsf{V}), \tag{31}$$

where $C_{\text{Sh}}(\mathsf{V})$ denotes the Shannon capacity of the channel V. This swap can be justified using the Minimax Theorem [12, Th. 1.1.5], because $I(\mathsf{Q}; \mathsf{V})$ is concave in Q and convex in V, the set of input distributions is convex and compact, and so is the set of channels that are absolutely continuous with respect to W.

Proposition 4: Consider a DMC W *whose zero-error feedback capacity can also be achieved without feedback, so that the two can be denoted* C_0*:*

$$C_0 = C_{0,\text{No-FB}}(\mathsf{W}) = C_{0,\text{FB}}(\mathsf{W}). \tag{32}$$

Let the channel $\mathsf{V}^\star \ll \mathsf{W}$ *achieve the minimum in* (31) *(not necessarily uniquely),*

$$C_0 = C_{\text{Sh}}\big(\mathsf{V}^\star\big). \tag{33}$$

If the distributions $\{\mathsf{V}^\star(\cdot|x)\}_{x \in \mathcal{X}}$ *corresponding to the different inputs to* $\mathsf{V}^\star$ *are distinct, then, unlike* $\mathcal{C}_{\text{FB}}(\mathsf{W})$*, the region* $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$ *contains no rate pairs of the form* (C_0, R_ε) *with* R_ε *positive.*

Proof: To prove this result we use the simple outer bound (Proposition 3) to infer that

$$\mathcal{C}_{\text{No-FB}}(\mathsf{W}) \subseteq \mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}^\star) \tag{34}$$

and study the behavior of $\mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}^\star)$ near the point $(C_0, 0)$. More specifically, we show that—subject to the hypothesis that the distributions $\{\mathsf{V}^\star(\cdot|x)\}_{x \in \mathcal{X}}$ are distinct—the region $\mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}^\star)$ contains no rate pairs of the form (C_0, R_ε) with R_ε positive.

If (C_0, R_ε) is in $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$, then substituting $\mathsf{V}^\star$ for V in (27) establishes that for some PMF $p_{U,X}$,

$$C_0 \leq I(U; Z) \tag{35a}$$
$$R_\varepsilon \leq I(X; Y|U) \tag{35b}$$
$$C_0 + R_\varepsilon \leq I(X; Y), \tag{35c}$$

where all the mutual informations are computed with respect to the joint PMF

$$p_{U,X,Y,Z}(u, x, y, z) = p_{U,X}(u, x)\, \mathsf{W}(y|x)\, \mathsf{V}^\star(z|x). \tag{36}$$

Let $\mathsf{V}_{Z|U}$ denote the channel from U to Z induced by $p_{X|U}$ and $\mathsf{V}^\star$

$$\mathsf{V}_{Z|U}(z|u) = \sum_{x \in \mathcal{X}} p_{X|U}(x|u)\, \mathsf{V}^\star(z|x). \tag{37}$$

By (33) and the Data-Processing Inequality [13, Lemma 3.11], (35a) must hold with equality, p_U must achieve $C_{\text{Sh}}\big(\mathsf{V}_{Z|U}\big)$, and p_X must achieve $C_{\text{Sh}}\big(\mathsf{V}^\star\big)$. In Appendix A we show that this and the proposition's assumptions imply that $H(X|U)$ must be zero. Consequently, by (35b), R_ε must also be zero. ∎

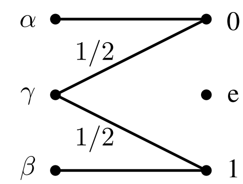

Fig. 3. A channel $\mathsf{V}^\star$ of Shannon capacity 1 bit that is absolutely continuous with respect to the channel W of Figure 2. Its different inputs induce different distributions on its output.

Corollary 5: Feedback can increase the multiplexing region even when it does not increase the zero-error capacity.

Proof: Consider the channel in Figure 2 and the auxiliary channel $\mathsf{V}^\star$ of Figure 3. From (31) we deduce that $C_{0,\text{FB}}$ is upper bounded by $C_{\text{Sh}}(\mathsf{V}^\star)$ and hence by 1. And, using the inputs α and β, we can send 1 bit error free without feedback, so $C_{0,\text{No-FB}} \geq 1$. Consequently the zero-error capacity of this channel is not increased by feedback and $C_{0,\text{No-FB}} = C_{0,\text{FB}} = 1$.

When Δ is sufficiently small, $C_{\text{Sh}}(\mathsf{W})$ exceeds 1, and $C_{\text{Sh}}(\mathsf{W}) - C_{0,\text{FB}}$ is positive. Since $\mathsf{V}^\star$'s different inputs induce different distributions on its output, we can apply Proposition 4 and infer that, for sufficiently small Δ, the pair $\big(1, C_{\text{Sh}}(\mathsf{W}) - 1\big)$ is not in $\mathcal{C}_{\text{No-FB}}(\mathsf{W})$. And yet, by Theorem 2, it is in $\mathcal{C}_{\text{FB}}(\mathsf{W})$. ∎

V. Improved Outer Bound

Given a finite set $\mathcal{U}$, let $\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U}, \mathcal{X})$ denote the set of conditional PMFs $\mathsf{V}(y|u, x)$ satisfying $\mathsf{V}(\cdot|u, \cdot) \ll \mathsf{W}(\cdot|\cdot)$ for every $u \in \mathcal{U}$, i.e., for which the implication

$$\big(\mathsf{W}(y|x) = 0\big) \implies \big(\mathsf{V}(y|u, x) = 0\big) \tag{38}$$

holds for every $y \in \mathcal{Y}$, $x \in \mathcal{X}$, and $u \in \mathcal{U}$. Let $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ be the closure of the union over all PMFs $p_{U,X}$ of the set of rate pairs (R_0, R_ε) that satisfy

$$R_0 \leq \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U}, \mathcal{X})} I(U; Z) \tag{39a}$$
$$R_\varepsilon \leq I(X; Y|U), \tag{39b}$$

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

$$p_{U,X,Y,Z}(u, x, y, z) = p_{U,X}(u, x)\, \mathsf{W}(y|x)\, \mathsf{V}(z|u, x). \tag{40}$$

Two equivalent characterizations of $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ are presented in Appendix B. The first of those shows that $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ is convex (Corollary 13 in Appendix B).

Proposition 6: Any rate pair (R_0, R_ε) *that can be multiplexed on the channel* W *without feedback must lie in* $\mathcal{R}_{\text{Imp}}(\mathsf{W})$

$$\mathcal{C}_{\text{No-FB}}(\mathsf{W}) \subseteq \mathcal{R}_{\text{Imp}}(\mathsf{W}). \tag{41}$$

Proof: See Appendix C. ∎

Proposition 6 improves on Proposition 3 in the following sense.

Remark 7: For any channel W,

$$\mathcal{R}_{\text{Imp}}(\mathsf{W}) \subseteq \mathcal{R}_{\text{Sim}}(\mathsf{W}), \tag{42}$$

with the inclusion being strict for some channels.

Proof: We first prove the inclusion. Let (R_0, R_ε) be in $\mathcal{R}_{\text{Imp}}(\mathsf{W})$. As such, there exists a PMF $p_{U,X}$ for which the inequalities in (39) hold. Fix this PMF. We will show that for this PMF and every $\mathsf{V}' \ll \mathsf{W}$, the pair (R_0, R_ε) satisfies (26) (when we substitute V' for V) and hence lies in $\mathcal{C}_{\text{BC-DM}}(\mathsf{W}; \mathsf{V}')$. Fix some such V'.

Since the pair (R_0, R_ε) satisfies the inequalities in (39), the pair must also satisfy these inequalities when we restrict the minimization over $\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U}, \mathcal{X})$ to those V's for which $\mathsf{V}(y|u, x)$ does not depend on u. (This restriction cannot decrease the right-hand side (RHS) of (39a).) Thus,

$$R_0 \le I_{\mathsf{V}}(U; Z), \quad \mathsf{V} \ll \mathsf{W} \tag{43a}$$

$$R_\varepsilon \le I(X; Y|U), \tag{43b}$$

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

$$p_{U,X,Y,Z}(u, x, y, z) = p_{U,X}(u, x)\, \mathsf{W}(y|x)\, \mathsf{V}(y|x), \tag{43c}$$

and the subscript in I_{V} indicates the channel w.r.t. which the mutual information is calculated. We now complete the proof of the inclusion by proving that (43) implies that (26) holds when we substitute V' for V.

Since the terms $I(X; Y|U)$ appearing on the RHS of (43b) and (26b) depend only on $p_{U,X}$ and W, they are identical and the former inequality thus implies the latter. It thus remains to establish (26a) (when we substitute V' for V). Substituting V' for V in (43) shows that $R_0 \le I_{\mathsf{V}'}(U; Z)$, and substituting W for V shows that $R_0 \le I_{\mathsf{W}}(U; Z)$. Since the latter is just $I(U; Y)$,

$$R_0 \le \min\{I_{\mathsf{V}'}(U; Z), I(U; Y)\}, \tag{44}$$

and (26a) (with the substitution of V' for V) holds. This concludes the proof of the inclusion.

To show that the inclusion can be strict, we consider the channel W in Figure 2 with $\Delta > 0$. As we have seen in the proof of Corollary 5, feedback does not increase its zero-error capacity, which equals 1 both with and without feedback, and which we denote $C_0(\mathsf{W})$. We will show that if $\Delta > 0$ is small enough so that the channel's Shannon capacity is at least 1.5, then $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ is equal to the time-sharing region between the rate pairs $\big(C_{\text{Sh}}(\mathsf{W}), 0\big)$ and $(0, C_0(\mathsf{W}))$, whereas $\mathcal{R}_{\text{Sim}}(\mathsf{W})$ contains rate pairs outside this time-sharing region. We will then be able to infer that, for this channel and such Δ,

$$\mathcal{C}_{\text{No-FB}}(\mathsf{W}) = \mathcal{R}_{\text{Imp}}(\mathsf{W}) \subset \mathcal{R}_{\text{Sim}}(\mathsf{W}), \tag{45}$$

where the inclusion is strict, and the equality holds because the time-sharing region is always achievable.[3]

To show that the simple bound contains pairs outside the time-sharing region, we fix some joint PMF $p_{U,X}$ (thus also fixing $I(U; Y)$ and $I(X; Y|U)$) and note that, since W is absolutely continuous w.r.t. itself, (29) and (26) imply that

$$\Big(\min_{\mathsf{V} \ll \mathsf{W}} I(U; Z),\ I(X; Y|U)\Big) \in \mathcal{R}_{\text{Sim}}(\mathsf{W}), \tag{46}$$

where the mutual informations are computed w.r.t. the PMF in (24).

We next claim that the minimization over V in (46) can be restricted to the family of channels V that, in addition to satisfying $\mathsf{V} \ll \mathsf{W}$, also satisfy $\mathsf{V}(\mathrm{e}|\gamma) = 0$. Since this family comprises the channels V satisfying $\mathsf{V} \ll \mathsf{V}^\star$, where $\mathsf{V}^\star$ is the channel in Figure 3, our claim can be restated as

$$\min_{\mathsf{V} \ll \mathsf{W}} I(U; Z) = \min_{\mathsf{V} \ll \mathsf{V}^\star} I(U; Z). \tag{47}$$

To prove it, we need to show that to every $\mathsf{V} \ll \mathsf{W}$ there corresponds some channel $\tilde{\mathsf{V}} \ll \mathsf{V}^\star$ such that $I(U; Z)$ under $\tilde{\mathsf{V}}$ is no larger than under V. To this end we can choose $\tilde{\mathsf{V}}$ as $\mathsf{V}\,\mathsf{V}^\star$, i.e. as the concatenation of V with $\mathsf{V}^\star$. This concatenation is depicted in Figure 4. It is then readily verified that $\mathsf{V}\,\mathsf{V}^\star \ll \mathsf{V}^\star$ and that, by the Data Processing Inequality, $I(U; Z)$ under $\mathsf{V}\,\mathsf{V}^\star$ is no larger than under V.

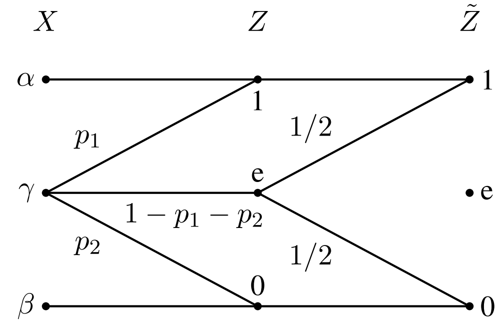

Fig. 4. The concatenation of a generic channel $\mathsf{V} \ll \mathsf{W}$ and the channel $\mathsf{V}^\star$ of Figure 3. The resulting channel between X and $\tilde{Z}$ is absolutely continuous w.r.t. $\mathsf{V}^\star$.

Having established (47), we next study the minimization on its RHS. We will show that, when $p_{U,X}$ has certain symmetry properties, the minimum is achieved when V is equal to $\mathsf{V}^\star$. To state these properties, we introduce the following notation: for $u \in \{0, 1\}$ and $x \in \{\alpha, \beta, \gamma\}$, we define $\bar{u}$ and $\bar{x}$ as

$$\bar{u} = 1 - u \tag{48a}$$

$$\bar{x} = \begin{cases} \beta, & \text{if } x = \alpha \\ \gamma, & \text{if } x = \gamma \\ \alpha, & \text{if } x = \beta. \end{cases} \tag{48b}$$

The properties of p_{UX} that we henceforth assume are that its marginal p_U is uniform over $\{0, 1\}$

$$p_U(0) = p_U(1) = \frac{1}{2} \tag{49a}$$

and that

$$p_{X|U}(x|u) = p_{X|U}(\bar{x}|\bar{u}). \tag{49b}$$

We will show that if these hold, then

$$\min_{\mathsf{V} \ll \mathsf{V}^\star} I_{\mathsf{V}}(U; Z) = I_{\mathsf{V}^\star}(U; Z). \tag{50}$$

[3]In this case the no-feedback multiplexing capacity region is achievable by simple time-sharing. This does not hold in general.

To establish (50), we will now show that, when $p_{U,X}$ satisfies (49),

$$I_{\mathsf{V}^\star}(U;Z) \le I_{\mathsf{V}}(U;Z), \quad \mathsf{V} \ll \mathsf{V}^\star. \tag{51}$$

To this end, let us introduce the notation

$$\bar{z} = \begin{cases} 1, & \text{if } z = 0 \\ \mathrm{e}, & \text{if } z = \mathrm{e} \\ 0, & \text{if } z = 1, \end{cases} \tag{52}$$

and let $\bar{\mathsf{V}}$ denote the channel that results when the output of V is flipped, i.e.,

$$\bar{\mathsf{V}}(z|x) = \mathsf{V}(\bar{z}|x). \tag{53}$$

With this notation,

$$\mathsf{V}^\star(z|x) = \frac{1}{2}\,\mathsf{V}(z|x) + \frac{1}{2}\,\bar{\mathsf{V}}(z|x), \quad \mathsf{V} \ll \mathsf{V}^\star. \tag{54}$$

It follows from (49) and (53) that

$$I_{\mathsf{V}}(U;Z) = I_{\bar{\mathsf{V}}}(U;Z), \quad \mathsf{V} \ll \mathsf{V}^\star. \tag{55}$$

Inequality (51) now follows from the convexity of mutual information in the channel law when the input distribution is fixed: for every $\mathsf{V} \ll \mathsf{V}^\star$,

$$\begin{aligned} I_{\mathsf{V}^\star}(U;Z) &= I_{\frac{1}{2}\mathsf{V}+\frac{1}{2}\bar{\mathsf{V}}}(U;Z) && (56a)\\ &\le \frac{1}{2}\,I_{\mathsf{V}}(U;Z) + \frac{1}{2}\,I_{\bar{\mathsf{V}}}(U;Z) && (56b)\\ &= I_{\mathsf{V}}(U;Z), && (56c)\end{aligned}$$

where (56a) follows from (54); Inequality (56b) holds by the aforementioned convexity; and (56c) follows from (55).

We conclude that, as long as p_{UX} satisfies (49), there is no need to carry out a numerical optimization in order to calculate the rate pair in (46). To obtain a rate pair that is in $\mathcal{R}_{\mathrm{Sim}}(\mathsf{W})$ but not in $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$, we now choose p_{UX} as

$$p_U(0) = p_U(1) = 0.5 \tag{57a}$$

and

$$\begin{aligned} p_{X|U}(\alpha|0) &= 0.85, \quad p_{X|U}(\alpha|1) = 0.02, && (57b)\\ p_{X|U}(\gamma|0) &= 0.13, \quad p_{X|U}(\gamma|1) = 0.13, && (57c)\\ p_{X|U}(\beta|0) &= 0.02, \quad p_{X|U}(\beta|1) = 0.85, && (57d)\end{aligned}$$

which satisfies the symmetry properties in (49), and for which we can thus evaluate the rate pair in (46) using (50).

The term $I_{\mathsf{V}^\star}(U;Z)$ does not depend on Δ and is given by

$$\begin{aligned} I_{\mathsf{V}^\star}(U;Z) &= 1 - H_{\mathrm{b}}\Big(p_{X|U}(\alpha|0) + \frac{1}{2}\,p_{X|U}(\gamma|0)\Big)\\ &\approx 0.5804435, \quad \Delta > 0, && (58a)\end{aligned}$$

where $H_{\mathrm{b}}(\cdot)$ denotes the binary entropy function. As to $I(X;Y|U)$, it does depend on Δ. For our purposes it suffices to study its limit as $\Delta \downarrow 0$, i.e., as W approaches a noise-free ternary channel. In this limit,

$$\lim_{\Delta\downarrow 0} I(X;Y|U) = H(X|U) \approx 0.6948167. \tag{58b}$$

Since $C_{\mathrm{Sh}}(\mathsf{W}) \to \log_2(3)$ as $\Delta \downarrow 0$, and since $C_0(\mathsf{W}) = 1$ for all $\Delta > 0$, it follows from (58a) and (58b) that, for the above choice of p_{UX},

$$\lim_{\Delta\downarrow 0} \frac{\min_{\mathsf{V}\ll\mathsf{W}} I(U;Z)}{C_0(\mathsf{W})} + \frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})} \approx 1.0188. \tag{58c}$$

Since this limit is larger than 1, we conclude that—for all sufficiently small (but positive) Δ—the simple outer bound $\mathcal{R}_{\mathrm{Sim}}(\mathsf{W})$ contains rate pairs outside the time-sharing region.

To prove that, for sufficiently small $\Delta > 0$, every rate pair (R_0, R_ε) in $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$ is also in the time-sharing region, we will show that any such pair must satisfy

$$\frac{R_0}{C_0(\mathsf{W})} + \frac{R_\varepsilon}{C_{\mathrm{Sh}}(\mathsf{W})} \le 1, \quad (R_0, R_\varepsilon) \in \mathcal{R}_{\mathrm{Imp}}(\mathsf{W}). \tag{59}$$

We shall do so by exhibiting for every choice of $p_{U,X}$ a conditional PMF $\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X})$ for which

$$\frac{I(U;Z)}{C_0(\mathsf{W})} + \frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})} \le 1, \tag{60}$$

where the mutual informations are computed w.r.t. to the PMF in (40), namely,

$$p_{U,X,Y,Z}(u,x,y,z) = p_{U,X}(u,x)\,\mathsf{W}(y|x)\,\mathsf{V}(z|u,x). \tag{61}$$

The conditional PMF $\mathsf{V}(z|u,x)$ will be chosen so that the probability of the output e be zero

$$\mathsf{V}(\mathrm{e}|u,x) = 0, \quad \forall u, x \tag{62}$$

and so that, conditional on u, the other outputs, namely 0 and 1, be "as uniformly distributed as possible."

To guarantee that V be in $\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X})$, we set

$$\mathsf{V}(0|u,\alpha) = 1, \quad \mathsf{V}(1|u,\beta) = 1, \ \forall u \in \mathcal{U}. \tag{63}$$

The probability $\mathsf{V}(0|u,\gamma) = 1 - \mathsf{V}(1|u,\gamma)$ corresponding to the input γ will be chosen depending on $p_{U,X}(u,x)$.

For each $u \in \mathcal{U}$,

1) if $p_{X|U}(\alpha|u) \le 1/2$ and $p_{X|U}(\beta|u) \le 1/2$, then $\mathsf{V}(\cdot|u,\gamma)$ is chosen so that, conditional on U, the chance variable Z be uniformly distributed over $\{0,1\}$. More specifically,[4]

$$\mathsf{V}(0|u,\gamma) = \frac{1/2 - p_{X|U}(\alpha|u)}{p_{X|U}(\gamma|u)}; \tag{64a}$$

2) if $p_{X|U}(\alpha|u) \ge 1/2 > p_{X|U}(\beta|u)$, then $\mathsf{V}(\cdot|u,\gamma)$ is chosen as

$$\mathsf{V}(0|u,\gamma) = 1 - \mathsf{V}(1|u,\gamma) = 0; \tag{64b}$$

3) and if $p_{X|U}(\alpha|u) < 1/2 \le p_{X|U}(\beta|u)$, then $\mathsf{V}(\cdot|u,\gamma)$ is chosen as

$$\mathsf{V}(0|u,\gamma) = 1 - \mathsf{V}(1|u,\gamma) = 1. \tag{64c}$$

[4] If $p_{X|U}(\alpha|u) = p_{X|U}(\beta|u) = 1/2$, then $p_{X|U}(\gamma|u) = 0$ and we choose $\mathsf{V}(0|u,\gamma) = 1/2$ for concreteness.

We next upper-bound the left-hand side (LHS) of (60) under the joint PMF (61) corresponding to this choice of $\mathsf{V}(z|u,x)$:

$$\begin{aligned}\frac{I(U;Z)}{C_0(\mathsf{W})}&+\frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})}\\&=I(U;Z)+\frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})} &(65a)\\&=H(Z)-H(Z|U)+\frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})} &(65b)\\&\leq 1-\left(H(Z|U)-\frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})}\right), &(65c)\end{aligned}$$

where (65a) holds because $C_0(\mathsf{W})$ is 1, and (65c) holds because Z takes on only two different values with positive probability. We next show that if

$$C_{\mathrm{Sh}}(\mathsf{W})\geq 1.5, \tag{66}$$

then

$$H(Z|U)-\frac{I(X;Y|U)}{C_{\mathrm{Sh}}(\mathsf{W})}\geq 0. \tag{67}$$

This and (65c) will imply (60).

To see why (67) holds for any $p_{U,X}$ whenever $C_{\mathrm{Sh}}(\mathsf{W})$ exceeds 1.5, we express its LHS as

$$\sum_{u\in\mathcal{U}}p_U(u)\left(H(Z|U=u)-\frac{I(X;Y|U=u)}{C_{\mathrm{Sh}}(\mathsf{W})}\right) \tag{68}$$

and show that each term in the sum is nonnegative, i.e., that, for every $u\in\mathcal{U}$,

$$H(Z|U=u)-\frac{I(X;Y|U=u)}{C_{\mathrm{Sh}}(\mathsf{W})}\geq 0. \tag{69}$$

We show this separately for the three cases defining V in (64). For every $u\in\mathcal{U}$:

1) If $p_{X|U}(\alpha|u)\leq 1/2$ and $p_{X|U}(\beta|u)\leq 1/2$, then by (64a), conditional on U, the chance variable Z has a uniform distribution and thus $H(Z|U=u)$ is 1. As to $I(X;Y|U=u)$, it is upper-bounded by $C_{\mathrm{Sh}}(\mathsf{W})$, so (69) holds.
2) If $p_{X|U}(\alpha|u)\geq 1/2>p_{X|U}(\beta|u)$, we prove (69) by showing that
$$H(Z|U=u)-\frac{H(X|U=u)}{1.5}\geq 0. \tag{70}$$
(This implies (69) because $I(X;Y|U=u)\leq H(X|U=u)$ and $C_{\mathrm{Sh}}(\mathsf{W})\geq 1.5$.) By (64b),
$$H(Z|U=u)=H_{\mathrm{b}}(p), \tag{71a}$$
where we define
$$p\triangleq p_{X|U}(\alpha|u)\geq\frac{1}{2}. \tag{71b}$$
As to $H(X|U=u)$, we upper-bound it in terms of p as follows:
$$\begin{aligned}H(X|U=u)&\\&=H\big((p_{X|U}(\alpha|u),p_{X|U}(\gamma|u),p_{X|U}(\beta|u))\big)\\&\leq H\big((p,(1-p)/2,(1-p)/2)\big) &(72a)\\&=H_{\mathrm{b}}(p)+1-p, &(72b)\end{aligned}$$
where (72a) holds because the entropy $H\big((p_1,p_2,p_3)\big)$ of any probability vector (p_1,p_2,p_3) satisfies
$$H\big((p_1,p_2,p_3)\big)\leq H\left(\left(p_1,\frac{1-p_1}{2},\frac{1-p_1}{2}\right)\right). \tag{73}$$
The LHS of (70) is thus, by (71a) and (72b), lower-bounded by
$$H_{\mathrm{b}}(p)-\frac{H_{\mathrm{b}}(p)+1-p}{1.5}, \tag{74}$$
where $p\geq 1/2$. It can now be verified numerically that (74) is nonnegative whenever $p\geq 1/2$.
3) If $p_{X|U}(\alpha|u)<1/2\leq p_{X|U}(\beta|u)$, a similar argumentation holds as in the previous case.

■

VI. A Back-Off in the Rates May Be Essential

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 2 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.

Proposition 8: Although always in $\mathcal{C}_{\mathrm{FB}}(\mathsf{W})$, the rate pair $(C_{0,\mathrm{FB}},C_{\mathrm{Sh}}-C_{0,\mathrm{FB}})$ need not be achievable.

Proof: Consider the above example. 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 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 $\mathcal{D}(m_0)$ contains 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 therefore 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. The rate R_ε must therefore be zero. ■

VII. Summary and Discussion

Theorem 2 establishes the multiplexing capacity region of zero-error and rare-error data streams over a noisy memoryless channel with feedback. It essentially says that, in the presence of feedback, n channel uses suffices to send $n\,C_{\text{Sh}}$ data bits with arbitrarily-small probability of error even if $n\,C_{0,\text{FB}}$ of them are designated to be transmitted error free. This is possible even when the input distributions that achieve C_{Sh} and $C_{0,\text{FB}}$ differ.

The coding scheme that was used to prove this result can also be employed in some multi-user settings (such as the multiple-access channel), leading to analogous results [10, Ch. 5].

It can also be modified to address scenarios where errors are not allowed in guessing m_ε, and in their stead we allow erasures [14]–[21] (and references therein.) To see how, note that 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 2 can be strengthened to address the *erasure-only* requirement as follows:

Remark 9: For channels with positive zero-error capacity, the rate pairs in Theorem 2 can also be achieved if we do not allow errors in guessing m_ε and instead allow erasures (with arbitrarily small but positive probability).

In the absence of feedback the problem of computing the multiplexing capacity region is at least as difficult as that of computing the zero-error capacity. Our outer bounds of Proposition 3 and Proposition 6 showed that feedback can increase the multiplexing capacity region even on channels where it does not increase the zero-error capacity.

Appendix A
Proof of Proposition 4 Concluded

To conclude the proof of Proposition 4, we shall need the following standard results.

Lemma 10 (Log-Sum Inequality): If the vectors $\mathbf{a} = (a_1,\dots,a_m)$ and $\mathbf{b} = (b_1,\dots,b_m)$ in $\mathbb{R}^m$ have nonnegative components, then

$$\sum_{i=1}^m a_i \log\frac{a_i}{b_i} \geq \left(\sum_{i=1}^m a_i\right)\log\frac{\sum_{i=1}^m a_i}{\sum_{i=1}^m b_i}, \tag{75a}$$

with equality if, and only if,

$$a_i\left(\sum_{j=1}^m b_j\right) = b_i\left(\sum_{j=1}^m a_j\right), \quad \forall i \in [1:m]. \tag{75b}$$

Here we adopt the convention that $0\log\frac{0}{\alpha} = 0$ for all $\alpha \in \mathbb{R}$, and $\alpha\log\frac{\alpha}{0} = \infty$ whenever $\alpha > 0$.

Proof: See [13, Lemma 3.1]. ■

The relative entropy $\mathsf{D}(\mathsf{P}\|\mathsf{Q})$ between two probability mass functions on a finite set $\mathcal{Y}$ is defined as

$$\mathsf{D}(\mathsf{P}\|\mathsf{Q}) = \sum_{y\in\mathcal{Y}} \mathsf{P}(y)\log\frac{\mathsf{P}(y)}{\mathsf{Q}(y)} \tag{76}$$

with the above convention. The next lemma addresses its convexity with emphasis on conditions for equality.

Lemma 11 (Convexity of Relative Entropy): Let R be a PMF on the finite set $\mathcal{Y}$ satisfying

$$\mathsf{R}(y) > 0, \quad \forall y \in \mathcal{Y}. \tag{77}$$

Given some positive integer m, let $\mathsf{P}_1,\dots,\mathsf{P}_m$ be PMFs on $\mathcal{Y}$ and $\alpha_1,\dots,\alpha_m$ positive real numbers summing to one. Then

$$\mathrm{D}\left(\sum_{i=1}^m \alpha_i \mathsf{P}_i \middle\| \mathsf{R}\right) \leq \sum_{i=1}^m \alpha_i\,\mathrm{D}(\mathsf{P}_i\|\mathsf{R}), \tag{78}$$

with equality if, and only if,

$$\mathsf{P}_1 = \mathsf{P}_2 = \ldots = \mathsf{P}_m. \tag{79}$$

Proof: Following the proof of [13, Lemma 3.5],

$$\begin{aligned}
&\mathrm{D}\left(\sum_{i=1}^m \alpha_i \mathsf{P}_i \middle\| \mathsf{R}\right)\\
&= \mathrm{D}\left(\sum_{i=1}^m \alpha_i \mathsf{P}_i \middle\| \sum_{i=1}^m \alpha_i \mathsf{R}\right) &\text{(80a)}\\
&= \sum_{y\in\mathcal{Y}}\left(\left(\sum_{i=1}^m \alpha_i \mathsf{P}_i(y)\right)\log\frac{\sum_{i=1}^m \alpha_i \mathsf{P}_i(y)}{\sum_{i=1}^m \alpha_i \mathsf{R}(y)}\right) &\text{(80b)}\\
&\leq \sum_{y\in\mathcal{Y}}\sum_{i=1}^m \alpha_i \mathsf{P}_i(y)\log\frac{\alpha_i\mathsf{P}_i(y)}{\alpha_i\mathsf{R}(y)} &\text{(80c)}\\
&= \sum_{i=1}^m \alpha_i \sum_{y\in\mathcal{Y}} \mathsf{P}_i(y)\log\frac{\mathsf{P}_i(y)}{\mathsf{R}(y)} &\text{(80d)}\\
&= \sum_{i=1}^m \alpha_i\,\mathrm{D}(\mathsf{P}_i\|\mathsf{R}), &\text{(80e)}
\end{aligned}$$

where in (80c) we applied the log-sum inequality once for each $y \in \mathcal{Y}$.

For equality we must have equality in the log-sum inequality in each of its applications, i.e. for each $y \in \mathcal{Y}$ and $i \in [1:m]$ we must have (*c.f.* (75b))

$$\alpha_i \mathsf{P}_i(y)\left(\sum_{j=1}^m \alpha_j \mathsf{R}(y)\right) = \alpha_i \mathsf{R}(y)\left(\sum_{j=1}^m \alpha_j \mathsf{P}_j(y)\right). \tag{81a}$$

Since the Lemma's hypotheses guarantee that neither $\mathsf{R}(y)$ nor α_i is zero, we can divide by $\alpha_i \mathsf{R}(y)$ and recall that the α_j's sum to 1 to obtain that

$$\mathsf{P}_i(y) = \sum_{j=1}^{m} \alpha_j \mathsf{P}_j(y) \tag{81b}$$

must be satisfied for every $y \in \mathcal{Y}$ and $i \in [1:m]$. Since the RHS does not depend on i, this implies that

$$\mathsf{P}_1(y) = \ldots = \mathsf{P}_m(y) = \sum_{j=1}^{m} \alpha_j \mathsf{P}_j(y) \tag{81c}$$

must be satisfied for every $y \in \mathcal{Y}$. This establishes that (79) is necessary for equality in (78). ■

Proof of Proposition 4 Concluded: Recall that using (33), the Data-Processing inequality, and (35a) we established that (35a) holds with equality; that p_U must achieve $C_{\mathrm{Sh}}(\mathsf{V}_{Z|U})$; that p_X must achieve $C_{\mathrm{Sh}}(\mathsf{V}^\star)$; and that $C_{\mathrm{Sh}}(\mathsf{V}_{Z|U}) = C_{\mathrm{Sh}}(\mathsf{V}^\star)$. What remains to show is that this and the proposition's assumptions imply that $H(X|U)$ is zero.

Let $\mathsf{R}^\star$ be the unique output distribution on $\mathcal{Z}$ that achieves the Shannon capacity $C_{\mathrm{Sh}}(\mathsf{V}^\star)$ of the channel $\mathsf{V}^\star$. Since p_X achieves $C_{\mathrm{Sh}}(\mathsf{V}^\star)$ and since the capacity-achieving output distribution is unique,

$$p_Z = \mathsf{R}^\star. \tag{82}$$

By possibly redefining the alphabet $\mathcal{Z}$, we may assume without loss of generality that

$$\mathsf{R}^\star(z) > 0, \quad \forall z \in \mathcal{Z}. \tag{83}$$

The input distribution p_U to the channel $\mathsf{V}_{Z|U}$ achieves its capacity $C_{\mathrm{Sh}}(\mathsf{V}_{Z|U})$ (which is equal to $C_{\mathrm{Sh}}(\mathsf{V}^\star)$) and induces the output distribution $\mathsf{R}^\star$. It therefore follows from the Karush-Kuhn-Tucker (KKT) conditions [22] that

$$\mathrm{D}\big(\mathsf{V}_{Z|U}(\cdot|u)\big\|\mathsf{R}^\star\big) \le C_{\mathrm{Sh}}(\mathsf{V}^\star), \quad \forall u \in \mathcal{U}, \tag{84}$$

with equality whenever $p_U(u) > 0$. For every $\tilde{u} \in \mathcal{U}$ with $p_U(\tilde{u}) > 0$ we thus have

$$\begin{aligned} C_{\mathrm{Sh}}(\mathsf{V}^\star) &= \mathrm{D}\big(\mathsf{V}_{Z|U}(\cdot|\tilde{u})\big\|\mathsf{R}^\star\big) \\ &= \mathrm{D}\Bigg(\sum_{x\in\mathcal{X}} p_{X|U}(x|\tilde{u})\,\mathsf{V}^\star(\cdot|x)\Bigg\|\mathsf{R}^\star\Bigg) \\ &= \mathrm{D}\Bigg(\sum_{\substack{x\in\mathcal{X}\\ p_{X|U}(x|\tilde{u})>0}} p_{X|U}(x|\tilde{u})\,\mathsf{V}^\star(\cdot|x)\Bigg\|\mathsf{R}^\star\Bigg) \\ &\le \sum_{\substack{x\in\mathcal{X}\\ p_{X|U}(x|\tilde{u})>0}} p_{U|X}(x|\tilde{u})\,\mathrm{D}\big(\mathsf{V}^\star(\cdot|x)\big\|\mathsf{R}^\star\big) \qquad (85) \\ &= C_{\mathrm{Sh}}(\mathsf{V}^\star), \qquad (86) \end{aligned}$$

where (85) follows from Lemma 11; and (86) follows by noting that for $x \in \mathcal{X}$

$$\Big(p_{X|U}(x|\tilde{u}) > 0\Big) \implies \Big(\mathrm{D}\big(\mathsf{V}^\star(\cdot|x)\big\|\mathsf{R}^\star\big) = C_{\mathrm{Sh}}(\mathsf{V}^\star)\Big) \tag{87}$$

because if $p_{X|U}(x|\tilde{u})$ is positive, then so is $p_X(x)$, and the implication follows from the KKT conditions on p_X for the channel $\mathsf{V}^\star$.

Since the LHS and RHS of (86) are identical, (85) must hold with equality. Consequently, by Lemma 11, all the conditional PMFs $\{\mathsf{V}(\cdot|x)\}_{p_{X|U}(x|\tilde{u})}$ corresponding to inputs x for which $p_{X|U}(x|\tilde{u})$ is positive must be identical. But, by the proposition's hypotheses, the PMFs $\{\mathsf{V}^\star(\cdot|x)\}_{x\in\mathcal{X}}$ are distinct, so there can be only one $x \in \mathcal{X}$ for which $p_{X|U}(x|\tilde{u})$ is positive. This is true for every $u \in \mathcal{U}$ for which $p_U(u) > 0$ and consequently $H(X|U) = 0$. ■

Appendix B
Equivalent Characterizations of $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$

This appendix provides two alternative characterizations of $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$ of (39). 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

$$\big(\mathsf{W}(y|x) = 0\big) \implies \big(\mathsf{V}(y|u,x,q) = 0\big) \tag{88}$$

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

Proposition 12: The region $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$ can also be expressed as the closure of the union over all PMFs $p_{U,X,Q}$ of the set of rate pairs (R_0, R_ε) that satisfy

$$R_0 \le \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q) \tag{89a}$$

$$R_\varepsilon \le I(X;Y|U,Q), \tag{89b}$$

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

$$p_{U,X,Y,Z,Q}(u,x,y,z,q) = p_{U,X,Q}(u,x,q)\,\mathsf{W}(y|x)\,\mathsf{V}(z|u,x,q). \tag{90}$$

Proof: One inclusion follows by choosing Q to be deterministic. We therefore focus on the other, namely, on showing that if there exists some joint PMF $p_{U,X,Q}$ under which the rate pair (R_0, R_ε) satisfies (89), then there exists some auxiliary chance variable $\tilde{U}$ and a PMF $p_{\tilde{U},X}$ under which the pair satisfies (39) when we substitute $\tilde{U}$ for U. To this end we choose $\tilde{U} \triangleq (U,Q)$ taking values in $\tilde{\mathcal{U}} \triangleq \mathcal{U}\times\mathcal{Q}$ and show that the result of substituting $\tilde{U}$ for U on the RHS of each of the inequalities in (39) is at least as high as the RHS of the corresponding inequality in (89). Starting with the first,

$$\begin{aligned} &\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\tilde{\mathcal{U}},\mathcal{X})} I(\tilde{U};Z) \\ &\quad= \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U,Q;Z) \\ &\quad= \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} \big(I(U;Z|Q) + I(Q;Z)\big) \\ &\quad\ge \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), \end{aligned}$$

and continuing with the second,

$$I(X;Y|\tilde{U}) = I(X;Y|U,Q).$$

■

Corollary 13: The rate region $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$ is a convex set containing the rate pairs $\big(0, C_{\mathrm{Sh}}(\mathsf{W})\big)$ and $\big(C_{0,\mathrm{FB}}(\mathsf{W}), 0\big)$.

Proof: Convexity follows by choosing Q to be a time-sharing auxiliary chance variable. To see that $\big(0, C_{\mathrm{Sh}}(\mathsf{W})\big)$ is in $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$, consider choosing U and Q to be deterministic. To see that $\big(C_{0,\mathrm{FB}}(\mathsf{W}), 0\big)$ is in $\mathcal{R}_{\mathrm{Imp}}(\mathsf{W})$, consider choosing Q deterministic and U to equal X. ■

The second alternative characterization of $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ follows by replacing the rare-error rate constraint with a sum-rate constraint:

Proposition 14: The region $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ *can also be expressed as the closure of the union over all PMFs* $p_{U,X,Q}$ *of the set of rate pairs* (R_0, R_ε) *that satisfy*

$$R_0 \le \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), \tag{91a}$$

$$R_\varepsilon + R_0 \le I(X;Y|U,Q) + \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), \tag{91b}$$

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

$$p_{U,X,Y,Z,Q}(u,x,y,z,q) = p_{U,X,Q}(u,x,q)\,\mathsf{W}(y|x)\,\mathsf{V}(z|u,x,q). \tag{92}$$

Proof: We prove this result using the characterization of $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ of Proposition 12. The region of Proposition 12 is contained in that of Proposition 14 because, for any fixed PMF $p_{U,X,Q}$, every rate pair that satisfies (89) must also satisfy (91) (because (91b) is the result of adding (89a) and (89b)).

We next establish the reverse inclusion. We fix some $p_{U,X,Q}$ and show that the trapezoid defined by (91) of vertices

$$\begin{aligned}&(0,0),\\ &\Big(0, I(X;Y|U,Q) + \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q)\Big),\\ &\Big(\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q)\,,\, I(X;Y|U,Q)\Big),\\ &\Big(\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), 0\Big)\end{aligned} \tag{93}$$

is included in the region defined by Proposition 12. This trapezoid is contained in the trapezoid of vertices

$$\begin{aligned}&(0,0),\quad (0, C_{\text{Sh}}(\mathsf{W})),\\ &\Big(\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q)\,,\, I(X;Y|U,Q)\Big),\\ &\Big(\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q), 0\Big)\end{aligned} \tag{94}$$

because, as we next argue,

$$I(X;Y|U,Q) + \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q) \le C_{\text{Sh}}(\mathsf{W}). \tag{95}$$

Indeed, the channel W is an element of $\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})$ (with no dependence on u and q), so

$$I(X;Y|U,Q) + \min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q)$$

$$\le I(X;Y|U,Q) + I_{\mathsf{W}}(U;Y|Q) \tag{96}$$

$$= I_{\mathsf{W}}(U,X;Y|Q) \tag{97}$$

$$= I_{\mathsf{W}}(X;Y|Q) \tag{98}$$

$$\le C_{\text{Sh}}(\mathsf{W}). \tag{99}$$

We next argue that the trapezoid of the vertices in (94) is contained in the region defined by Proposition 12. The latter is a convex set containing the vertex $(0, C_{\text{Sh}}(\mathsf{W}))$ (Corollary 13 and Proposition 12), and it contains the vertex/pair

$$\Big(\min_{\mathsf{V}\in\mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U;Z|Q)\,,\, I(X;Y|U,Q)\Big)$$

because this pair satisfies (89). Since these latter two vertices dominate the other vertices in (94), the entire trapezoid is contained in the region defined by Proposition 12. ■

Appendix C
Proof of the Improved Bound

This appendix proves that no rate pair outside $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ can be multiplexed over the channel W without feedback. It uses the characterization of $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ that is given in Proposition 14 in terms of (91).

Proof: Fix some finite message sets $\mathcal{M}_0$, $\mathcal{M}_\varepsilon$, a blocklength n, a blocklength-n no-feedback encoder $f\colon \mathcal{M}_0 \times \mathcal{M}_\varepsilon \to \mathcal{X}^n$, and $|\mathcal{M}_0||\mathcal{M}_\varepsilon|$ disjoint decoding sets $\{\mathcal{D}_{m_0,m_\varepsilon} \subseteq \mathcal{Y}^n\}_{(m_0,m_\varepsilon)\in\mathcal{M}_0\times\mathcal{M}_\varepsilon}$. Let

$$\mathcal{Q} \triangleq \{1,\ldots,n\}. \tag{100}$$

Draw the message pair (M_0, M_ε) uniformly over $\mathcal{M}_0 \times \mathcal{M}_\varepsilon$, and denote its distribution p_{M_0,M_ε}. Let P denote the joint PMF of $(M_0, M_\varepsilon, X^n, Y^n)$ induced by p_{M_0,M_ε}, the encoder f, and the channel W. Thus, for every $(m_0, m_\varepsilon, \mathbf{x}, \mathbf{y}) \in \mathcal{M}_0 \times \mathcal{M}_\varepsilon \times \mathcal{X}^n \times \mathcal{Y}^n$,

$$\begin{aligned}&\mathsf{P}[(M_0, M_\varepsilon, X^n, Y^n)\\ &\quad = (m_0, m_\varepsilon, \mathbf{x}, \mathbf{y})]\\ &\quad = p_{M_0,M_\varepsilon}(m_0,m_\varepsilon)\, p_{X^n|M_0,M_\varepsilon}(\mathbf{x}|m_0,m_\varepsilon)\,\mathsf{W}^n(\mathbf{y}|\mathbf{x}),\end{aligned} \tag{101a}$$

where

$$p_{X^n|M_0,M_\varepsilon}(\mathbf{x}|m_0,m_\varepsilon) = \begin{cases}1 & \text{if } \mathbf{x} = f(m_0,m_\varepsilon)\\ 0 & \text{otherwise,}\end{cases} \tag{101b}$$

and

$$\mathsf{W}^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n \mathsf{W}(y_i|x_i). \tag{101c}$$

Since M_0 is decoded error-free,

$$\mathsf{P}\Big[Y^n \in \bigcup_{m_\varepsilon\in\mathcal{M}_\varepsilon} \mathcal{D}_{M_0,m_\varepsilon}\Big] = 1. \tag{102}$$

For each $i \in \mathcal{Q}$, let Z_i be a $\mathcal{Y}$-valued chance variable whose conditional PMF given $\big(m_0, y_{i+1}^n, z_1^{i-1}, x_i\big)$

$$\tilde{\mathsf{V}}_i(\cdot|m_0, y_{i+1}^n, z_1^{i-1}, x_i), \tag{103}$$

(to be specified later) satisfies

$$\begin{aligned}&\tilde{\mathsf{V}}_i(\cdot|m_0, y_{i+1}^n, z_1^{i-1}, x_i) \ll \mathsf{W}(\cdot|x_i),\\ &\quad \forall (m_0, y_{i+1}^n, z_1^{i-1}, x_i) \in \mathcal{M}_0 \times \mathcal{Y}^{n-i} \times \mathcal{Y}^{i-1} \times \mathcal{X}.\end{aligned} \tag{104}$$

The n conditional PMFs $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$ can also be viewed as single conditional PMF $\tilde{\mathsf{V}}(\cdot|m_0, y_{i+1}^n, z_1^{i-1}, x_i, i)$ with the understanding that, for every $i \in \mathcal{Q}$,

$$\tilde{\mathsf{V}}\big(\,\cdot\,|m_0, y_{i+1}^n, z_1^{i-1}, x_i, i\big) = \tilde{\mathsf{V}}_i(\cdot|m_0, y_{i+1}^n, z_1^{i-1}, x_i). \tag{105}$$

We shall therefore use $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$ and $\tilde{\mathsf{V}}$ interchangeably.

Draw Q independently of $(M_0, M_\epsilon, X^n, Y^n, Z^n)$ uniformly over $\mathcal{Q}$, and denote the joint PMF of $(Q, M_0, M_\varepsilon, X^n, Y^n, Z^n)$ also P, so

$$\begin{aligned}\mathsf{P}[(&Q, M_0, M_\varepsilon, X^n, Y^n, Z^n)\\ &= (q, m_0, m_\varepsilon, \mathbf{x}, \mathbf{y}, \mathbf{z})]\\ &= p_{M_0,M_\varepsilon}(m_0, m_\varepsilon)\, p_{X^n|M_0,M_\varepsilon}(\mathbf{x}|m_0, m_\varepsilon)\, \mathsf{W}^n(\mathbf{y}|\mathbf{x})\\ &\quad\times \frac{1}{n}\prod_{i=1}^{n} \tilde{\mathsf{V}}_i(z_i|m_0, y_{i+1}^n, z_1^{i-1}, x_i).\end{aligned} \tag{106}$$

It follows from (104) and (106) that, for every $\xi^n \in \mathcal{Y}^n$,

$$\Big(\mathsf{P}\big(Y^n = \xi^n\big) = 0\Big) \implies \Big(\mathsf{P}\big(Z^n = \xi^n\big) = 0\Big) \tag{107}$$

and hence, by (102),

$$\mathsf{P}\Bigg[Z^n \in \bigcup_{m_\varepsilon \in \mathcal{M}_\varepsilon} \mathcal{D}_{M_0,m_\varepsilon}\Bigg] = 1, \tag{108}$$

so M_0 is a deterministic function of Z^n. This and Fano's inequality imply

$$\log|\mathcal{M}_0| = I(M_0; Z^n), \tag{109a}$$

$$\log\big(|\mathcal{M}_0||\mathcal{M}_\varepsilon|\big) \le I(M_0; Z^n) + I(M_\epsilon; Y^n) + n\epsilon_n, \tag{109b}$$

where ϵ_n tends to zero as n goes to infinity.

For each $i \in \mathcal{Q}$ define

$$U_i = (M_0, Y_{i+1}^n, Z_1^{i-1}), \tag{110}$$

and define the chance variables

$$U = (U_Q, Q), \quad X = X_Q, \quad Y = Y_Q, \quad Z = Z_Q. \tag{111}$$

Let $\mathsf{P}_{U,X,Q,Y,Z}$ denote the joint distribution of (U, X, Q, Y, Z) induced by P, and let $\mathsf{P}_{U,X,Q}$ denote its (U, X, Q)-marginal, with similar notation for its other marginals. Under P, the conditional distribution of Z given (U, X, Q) is $\tilde{\mathsf{V}}$, so

$$\mathsf{P}_{U,X,Q,Z}(u, x, q, z) = \mathsf{P}_{U,X,Q}(u, x, q)\, \tilde{\mathsf{V}}(z|u, x, q) \tag{112}$$

and

$$\tilde{\mathsf{V}}\big(\cdot\,|(m_0, y_{i+1}^n, z_1^{i-1}, i), x_i, i\big) = \tilde{\mathsf{V}}_i(\cdot|m_0, y_{i+1}^n, z_1^{i-1}, x_i) \tag{113}$$

for every $i \in \mathcal{Q}$ and $(m_0, y_{i+1}^n, z_1^{i-1}, x_i) \in \mathcal{M}_0 \times \mathcal{Y}^{n-i} \times \mathcal{Y}^{i-1} \times \mathcal{X}$.

Continuing from (109), we upper-bound $R_0 = \frac{1}{n}\log|\mathcal{M}_0|$ and $R_0 + R_\varepsilon = \frac{1}{n}\log\big(|\mathcal{M}_0||\mathcal{M}_\epsilon|\big)$ by carrying out the following steps under P of (106):

$$\begin{aligned}\frac{1}{n}\log|\mathcal{M}_0| &= \frac{1}{n} I(M_0; Z_1^n)\\ &= \frac{1}{n}\sum_{i=1}^{n} I(M_0; Z_i|Z_1^{i-1}) &\text{(114a)}\\ &\le \frac{1}{n}\sum_{i=1}^{n} I(M_0, Y_{i+1}^n, Z_1^{i-1}; Z_i) &\text{(114b)}\\ &= I(U; Z|Q), &\text{(114c)}\end{aligned}$$

where (114a) and (114b) follow from the chain rule and the nonnegativity of mutual information, and where (114c) follows from the definitions in (111).

Similarly, for the sum of the rates,

$$\begin{aligned}&\frac{1}{n}\log\big(|\mathcal{M}_0||\mathcal{M}_\varepsilon|\big) - \epsilon_n\\ &\le \frac{1}{n} I(M_0; Z_1^n) + \frac{1}{n} I(M_\varepsilon; Y_1^n|M_0) &\text{(115a)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0; Z_i|Z_1^{i-1}) + I(M_\varepsilon; Y_i|M_0, Y_{i+1}^n)\Big) &\text{(115b)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0; Z_i|Z_1^{i-1}) + I(X_i, M_\varepsilon; Y_i|M_0, Y_{i+1}^n)\Big) &\text{(115c)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0; Z_i|Z_1^{i-1}) + I(X_i; Y_i|M_0, Y_{i+1}^n)\Big) &\text{(115d)}\\ &\le \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0, Z_1^{i-1}; Z_i) + I(X_i, Z_1^{i-1}; Y_i|M_0, Y_{i+1}^n)\Big) &\text{(115e)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0, Y_{i+1}^n, Z_1^{i-1}; Z_i) - I(Y_{i+1}^n; Z_i|M_0, Z_1^{i-1})\\ &\qquad + I(Z_1^{i-1}; Y_i|M_0, Y_{i+1}^n) + I(X_i; Y_i|M_0, Y_{i+1}^n, Z_1^{i-1})\Big) &\text{(115f)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\Big(I(M_0, Y_{i+1}^n, Z_1^{i-1}; Z_i) + I(X_i; Y_i|M_0, Y_{i+1}^n, Z_1^{i-1})\Big) &\text{(115g)}\\ &= I(U; Z|Q) + I(X; Y|U, Q), &\text{(115h)}\end{aligned}$$

where (115a) follows from (109b), because M_0 is independent of M_ε; (115b) follows from the chain rule; (115c) holds because, in the absence of feedback, X_i is computable from (M_ε, M_0); (115d) holds because

$$M_\varepsilon \multimap\!\!- (X_i, M_0, Y_{i+1}^n) \multimap\!\!- Y_i \tag{116}$$

forms a Markov chain; (115e) and (115f) follow from the chain rule and the nonnegativity of mutual information; (115g) follows from Csiszár's Sum Identity [8, Sec. 2.4]; and (115h) follows from the definitions in (111). Inequalities (114) and (115) hold for any choice of the conditional PMFs $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$ in (103) subject to (104). Different choices will merely induce different P's.

Our choice offers no control over the (U, X, Q)-marginal of P, and this is fine because the characterization of $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ provided by Proposition 14 involves a union over all the possible marginals. Let $\tilde{p}_{U,X,Q}$ be the (U, X, Q)-marginal of the joint PMF P that our choice induces. Likewise, let $\tilde{p}_{U_i,X_i}$

denote the (U_i, X_i)-marginal of the joint PMF P that our choice induces. With this notation,

$$\tilde{p}_{U,X|Q=i} = \tilde{p}_{U_i,X_i}, \quad i \in \mathcal{Q}. \tag{117}$$

We will choose $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$ (or, equivalently, $\tilde{\mathsf{V}}$) so that the RHS of (114c), namely $I(U; Z|Q)$, equal the RHS of (91a), i.e., so that

$$I(U; Z|Q) = \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U; Z|Q), \tag{118}$$

where the mutual information on the LHS is computed w.r.t. the joint PMF $\tilde{p}_{U,X,Q}(u, x, q)\, \tilde{\mathsf{V}}(z|u, x, q)$, and the mutual information on the RHS w.r.t. the joint PMF $\tilde{p}_{U,X,Q}(u, x, q)\, \mathsf{V}(z|u, x, q)$. The combination of (118) with (114c) and (115h) will establish (91) (by letting n tend to infinity).

To describe our choice, we first note that, for each $i \in \mathcal{Q}$, the (U_i, X_i)-marginal $\tilde{p}_{U_i,X_i}$ of P is unaffected by the choice of $\tilde{\mathsf{V}}_i, \tilde{\mathsf{V}}_{i+1}, \ldots, \tilde{\mathsf{V}}_n$. Indeed, $\tilde{\mathsf{V}}_i$ influences Z_i^n and not Z_1^{i-1} (106). Consequently, since U_i only involves Z_1^{i-1} (110), it is not influenced by the choice of $\tilde{\mathsf{V}}_i, \ldots, \tilde{\mathsf{V}}_n$. We can now describe our choice of $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$ by expressing $I(U; Z|Q)$ as a sum

$$\begin{aligned} I(U; Z|Q) &= \frac{1}{n}\sum_{i=1}^n I(U; Z|Q = i) \\ &= \frac{1}{n}\sum_{i=1}^n I(U_i; Z_i) \end{aligned} \tag{119}$$

and by considering each of the terms in increasing order, starting with $i = 1$. The joint distribution $\tilde{p}_{U_1,X_1}$ is determined by the messages' distribution p_{M_0,M_ε}, the encoder f, and the channel W. We choose $\tilde{\mathsf{V}}_1$ so that

$$I(U_1; Z_1) = \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X})} I(U_1; Z), \tag{120a}$$

where the mutual information on the LHS is computed w.r.t. to the joint PMF $\tilde{p}_{U_1,X_1}(u_1, x_1)\, \tilde{\mathsf{V}}_1(z_1|u_1, x_1)$ and on the RHS w.r.t the joint PMF $\tilde{p}_{U_1,X_1}(u_1, x_1)\, \mathsf{V}(z|u_1, x_1)$. In general, for Term-$i \in [2 : n]$, the joint distribution $\tilde{p}_{U_i,X_i}$ is determined by the messages' distribution p_{M_0,M_ε}, the encoder f, the channel W, and our previous choices of the conditional PMFs $\{\tilde{\mathsf{V}}_l\}_{l=1}^{i-1}$. We then choose $\tilde{\mathsf{V}}_i$ so that

$$I(U_i; Z_i) = \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X})} I(U_i; Z), \tag{120b}$$

where the mutual information on the LHS is computed w.r.t. the joint PMF $\tilde{p}_{U_i,X_i}(u_i, x_i)\, \tilde{\mathsf{V}}_i(z_i|u_i, x_i)$ and on the RHS w.r.t the joint PMF $\tilde{p}_{U_i,X_i}(u_i, x_i)\, \mathsf{V}_i(z|u_i, x_i)$. After choosing all the conditional PMFs, the joint distributions $\{\tilde{p}_{U_i,X_i}\}_{i=1}^n$ are fully determined and hence also $\tilde{p}_{U,X,Q}$ via (117).

We next show that, under $\tilde{p}_{U,X,Q}(u, x, q)\, \tilde{\mathsf{V}}(z|u, x, q)$ (which is the P corresponding to our choice of $\{\tilde{\mathsf{V}}_i\}_{i=1}^n$), the rates $(R_0, R_\varepsilon - \epsilon_n)$ must satisfy (91). Indeed, under this joint distribution, $I(U; Z|Q)$ can be written as

$$\begin{aligned} I(U; Z|Q) &\\ &= \frac{1}{n}\sum_{i=1}^n I(U; Z|Q = i) \\ &= \frac{1}{n}\sum_{i=1}^n I(U_i; Z_i) && \text{(121a)} \\ &= \frac{1}{n}\sum_{i=1}^n \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X})} I(U_i; Z) && \text{(121b)} \\ &= \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} \frac{1}{n}\sum_{i=1}^n I(U; Z|Q = i) && \text{(121c)} \\ &= \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U; Z|Q), && \text{(121d)} \end{aligned}$$

where $I(U_i; Z_i)$ in (121a) is calculated w.r.t. $\tilde{p}_{U_i,X_i}(u_i, x_i)\, \tilde{\mathsf{V}}_i(z_i|u_i, x_i)$; where $I(U_i; Z)$ in (121b) is calculated w.r.t. $\tilde{p}_{U_i,X_i}(u_i, x_i)\, \mathsf{V}(z|u_i, x_i)$, and the equality follows from (120); and where (121c) holds because the minimization can be carried out independently for each realization of $Q \in \mathcal{Q}$ and can therefore be viewed as a minimization over $\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})$.

The upper bounds (114) and (115) together with (121) yield that, under $\tilde{p}_{U,X,Q}$ of (117), the rates of the coding scheme are upper-bounded by

$$\begin{aligned} R_0 &\le \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U; Z|Q) && \text{(122a)} \\ R_\varepsilon + R_0 &\le I(X; Y|U, Q) \\ &\quad + \min_{\mathsf{V} \in \mathcal{P}_{\mathsf{W}}(\mathcal{Y}|\mathcal{U},\mathcal{X},\mathcal{Q})} I(U; Z|Q) + \epsilon_n, && \text{(122b)} \end{aligned}$$

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

$$\begin{aligned} p_{U,X,Y,Z,Q}&(u, x, y, z, q) \\ &= \tilde{p}_{U,X,Q}(u, x, q)\, \mathsf{W}(y|x)\, \mathsf{V}(z|u, x, q). \end{aligned} \tag{123}$$

Having established that $(R_0, R_\varepsilon - \epsilon_n)$ must satisfy (91), we now recall that ϵ_n tends to zero and that $\mathcal{R}_{\text{Imp}}(\mathsf{W})$ is closed and conclude that (R_0, R_ε) must be in $\mathcal{R}_{\text{Imp}}(\mathsf{W})$. ■

ACKNOWLEDGMENT

The authors thank the Associate Editor and the anonymous referees for their constructive comments.

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.
[2] P. Elias, "Zero error capacity under list decoding," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 5, pp. 1070–1074, Sep. 1988.
[3] R. Ahlswede, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 25, no. 3, pp. 239–252, Sep. 1973.
[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
[5] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 1–7, Jan. 1979.
[6] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. Inf. Theory*, vol. IT-44, no. 6, pp. 2207–2229, Oct. 1998.

[7] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 60–64, Jan. 1977.
[8] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[9] C. Nair and A. El Gamal, "The capacity region of a class of three-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, Oct. 2009.
[10] T. Keresztfalvi, "Some data are more important than others," Ph.D. dissertation, Dept. Inf. Technol. Elect. Eng., ETH Zurich, Zurich, Switzerland, 2018.
[11] T. Keresztfalvi and A. Lapidoth, "Semi-robust communications over a broadcast channel," *IEEE Trans. Inf. Theory*, submitted for publication.
[12] S. Karlin, *Mathematical Methods and Theory in Game, Programming and Economics*. Reading, MA, USA: Addison-Wesley, 1966.
[13] I. Csiszár and J. Körner, *Information Theory: Coding theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[14] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 206–220, Mar. 1968.
[15] M. S. Pinsker and A. Y. Sheverdyaev, "Transmission capacity with zero error and erasure," *Problems Inf. Transmiss.*, vol. IT-6, no. 1, pp. 20–24, 1970.
[16] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.
[17] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, Jan. 1996.
[18] İ. E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1977–1982, Nov. 1997.
[19] C. Bunte and A. Lapidoth, "The zero-undetected-error capacity of discrete memoryless channels with feedback," in *Proc. 50th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Monticello, IL, USA, Oct. 2012, pp. 1838–1842.
[20] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 24–49, Jan. 2012.
[21] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity approaches the Sperner capacity," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3825–3833, Jul. 2014.
[22] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

Tibor Keresztfalvi was born on September 24th, 1991, in Budapest, Hungary. He received his Bachelor of Science (B.Sc.) degree in the field of Electrical Engineering from ETH Zurich, Switzerland, in 2013. He completed his Master of science (M.Sc.) in the field of Electrical Engineering with a focus on Digital Communication at ETH Zurich, in 2015. After that, he joined the Signal and Information Processing Laboratory (ISI) at ETH Zurich, where he pursued a PhD degree in Information Theory, which he acquired in 2018.

His research interests lie in the area of zero-error information theory, and multiterminal information theory.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in Mathematics (*summa cum laude*, 1986), the B.Sc. degree in Electrical Engineering (*summa cum laude*, 1986), and the M.Sc. degree in Electrical Engineering (1990) all from the Technion–Israel Institute of Technology. He received the Ph.D. degree in Electrical Engineering from Stanford University in 1995. In the years 1995–1999 he was an Assistant and Associate Professor at the department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. He served in the years 2003–2004 and 2009 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY. Dr. Lapidoth's research interests are in Digital Communications and Information Theory. He is the author of the textbook *A Foundation in Digital Communication*, second edition, published by Cambridge University Press in 2017.