The Zero-Error Capacity of the Gelfand-Pinsker Channel with a Feedback Link

Annina Bracher and Amos Lapidoth
ETH Zurich, Switzerland
{bracher,lapidoth}@isi.ee.ethz.ch

***Abstract*—The zero-error feedback capacity of the Gelfand-Pinsker channel is established. It can be positive even if the channel's zero-error capacity is zero in the absence of feedback. Moreover, the error-free transmission of a single bit may require more than one channel use.**

I. Introduction

Motivated by Shannon's characterization of the zero-error capacity of the discrete memoryless channel (DMC) with a feedback link from the channel output to the encoder [1], we compute the corresponding capacity for the state-dependent channel whose state is revealed acausally to the transmitter. This "Gelfand-Pinsker channel," which was introduced by Gelfand and Pinsker in [2], [3], is more general than the channel studied by Shannon, and, indeed, when the state is deterministic we recover Shannon's result. But, more interestingly, this channel's zero-error feedback capacity exhibits phenomena that are not observed on the state-less channel: it can be positive even if it is zero in the absence of feedback; the error-free transmission of a single bit may require more than one channel use; and Shannon's sequential coding technique cannot be applied.

Like Shannon's, our coding scheme is a two-phase scheme where the first phase reduces the receiver's ambiguity to a manageable size, and the second removes it entirely. But our first phase differs from Shannon's sequential approach and draws instead on Dueck's scheme for zero-error communication over the multiple-access channel with feedback [4], which in turn draws on Ahlswede's work [3], [5], [6]. The second phase is tricky, because sending a single bit reliably may require more than one channel use, so "uncoded" transmission need not work.

There are interesting connections between the problem of computing the zero-error capacity of a DMC and that of computing the m-capacity (the capacity under the maximal-probability-of-error criterion) of an arbitrarily-varying channel (AVC) [7]. Indeed, given a DMC $W(y|x)$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, the following construction produces an AVC $\widetilde{W}(y|x,\sigma)$ whose m-capacity is equal to the zero-error capacity of the channel $W(y|x)$ [7, Section 2]. To construct the AVC we consider the functions $\sigma\colon \mathcal{X} \to \mathcal{Y}$ that satisfy that $W(\sigma(x)|x)$ is positive for all $x \in \mathcal{X}$. With each such function $\sigma(\cdot)$ we associate a state σ and the transition law

$$\widetilde{W}(y|x,\sigma) = \begin{cases} 1 & \text{if } y = \sigma(x), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The constructed AVC has two important properties. The first is that to every pair of input and output sequences $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ for which $\prod W(y_k|x_k)$ is positive, there corresponds a sequence of states $\sigma_1, \ldots, \sigma_n$ such that $y_k = \sigma_k(x_k)$ for $k = 1, \ldots, n$. The second is that $\widetilde{W}(y|x,\sigma)$ is $\{0,1\}$-valued. This latter property guarantees that the conditional probability of error over the AVC (conditional on the transmitted message and the state sequence) is $\{0,1\}$-valued and thus small (say, smaller than $1/2$) only if it is zero.

This relationship between the zero-error capacity and the m-capacity fails when the original channel whose zero-error capacity we seek is state-dependent and the state is revealed to the encoder. To see the difficulty, let us denote by $W(y|x,s)$ the transition law of the state-dependent channel whose zero-error capacity we seek when the state is revealed to the encoder, and suppose we want to construct an AVC $\widetilde{W}(y|x,\sigma)$ whose m-capacity when the state σ is revealed to the encoder is equal to the zero-error capacity we seek. We have intentionally used different letters s and σ for the state of the original channel and of the AVC because the two need not *prima facie* be the same. For example, if there is only one state $s^\star$, then we are back to the state-less case and the construction we described above in (1) results in the number of AVC states being equal to the number of functions $\sigma\colon \mathcal{X} \to \mathcal{Y}$ that satisfy that $W(\sigma(x)|x,s^\star)$ is positive for all $x \in \mathcal{X}$. However, in this case the m-capacity of the AVC $\widetilde{W}(y|x,\sigma)$ is equal to the zero-error capacity we seek only if the state σ is *not* revealed to the encoder. In attempting to construct the AVC we are faced with two conflicting requirements. For the state information (SI) that is revealed to the encoder in the two scenarios to be identical, the states s and σ should be identical. But for the AVC to have a $\{0,1\}$-law, the number of AVC states σ should typically be larger than the number of states s.

The construction does go through in the special case where the original state-dependent transition law $W(y|x,s)$ happens to be $\{0,1\}$-valued. In this special case we can choose σ to equal s, and the m-capacity equals the zero-error capacity. In this case feedback is superfluous, because from the state (which is revealed to the encoder) and from the input (that it produces) the encoder can compute the output. We thus see that when $W(y|x,s)$ is $\{0,1\}$-valued our results can be inferred from Ahlswede's results on the feedback-less AVC with SI at the encoder [8]; but in the general case they cannot.

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

A. Notation and Terminology

We consider a state-dependent DMC (SD-DMC) of transition law $W(y|x,s)$, which is governed by an IID $\sim Q$ state-process. The input alphabet $\mathcal{X}$, the state alphabet $\mathcal{S}$, and the output alphabet $\mathcal{Y}$ are all finite. By possibly redefining $\mathcal{S}$, we can assume without loss of generality that

$$Q(s) > 0, \quad s \in \mathcal{S}. \tag{2}$$

Subject to (2), the exact nature of the PMF Q is immaterial.

For an SD-DMC $W(y|x,s)$ we denote by $\mathscr{P}(W)$ the set of transition laws $P_{Y|X,S}$ from $\mathcal{X} \times \mathcal{S}$ to $\mathcal{Y}$ for which

$$\bigl(W(y|x,s) = 0\bigr) \implies \bigl(P_{Y|X,S}(y|x,s) = 0\bigr), \ \forall\, x,\, s,\, y.$$

For a state-less DMC $W(y|x)$ we drop S, and $\mathscr{P}(W)$ denotes the set of transition laws $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$ for which

$$\bigl(W(y|x) = 0\bigr) \implies \bigl(P_{Y|X}(y|x) = 0\bigr), \ \forall\, x,\, y.$$

II. Problem Description and Main Result

We consider an SD-DMC $W(y|x,s)$ with feedback whose encoder is furnished with the state sequence acausally. Using n channel uses, the encoder wants to convey to the receiver error-free a message m from some finite set of messages $\mathcal{M}$. To this end, it uses an $(n,\mathcal{M})$ zero-error code:

Definition 1. *Given a finite set $\mathcal{M}$ and a positive integer $n \in \mathbb{N}$, an $(n,\mathcal{M})$ zero-error feedback code for the SD-DMC $W(y|x,s)$ with acausal SI to the encoder consists of n encoding mappings $f_i \colon \mathcal{M} \times \mathcal{S}^n \times \mathcal{Y}^{i-1} \to \mathcal{X}$, $i \in [1:n]$ and $|\mathcal{M}|$ disjoint decoding sets $\mathcal{D}_m \subseteq \mathcal{Y}^n$, $m \in \mathcal{M}$ such that for every $m \in \mathcal{M}$ the probability of a decoding error is zero:*

$$\mathbb{P}[Y^n \notin \mathcal{D}_m | M = m, S^n = \mathbf{s}] = 0, \ \forall\, m \in \mathcal{M},\, \mathbf{s} \in \mathcal{S}^n,$$

where

$$\mathbb{P}[Y^n \notin \mathcal{D}_m | M = m, S^n = \mathbf{s}] = \sum_{\mathbf{y} \in \mathcal{Y}^n \setminus \mathcal{D}_m} \prod_{i=1}^{n} W\bigl(y_i \big| f_i(m, \mathbf{s}, y^{i-1}), s_i\bigr). \tag{3}$$

A rate R is achievable if for every sufficiently-large block-length n there exists an $(n,\mathcal{M})$ zero-error feedback code with $\log|\mathcal{M}| \geq nR$, where all logarithms are base-2. The zero-error feedback capacity is the supremum of all achievable rates and is denoted $C_{\mathrm{f},0}$.

Note that the PMF Q governing the state does not appear in Definition 1 and therefore does not affect the zero-error feedback capacity. Also note that our definition assumes deterministic encoders. This assumption is not restrictive:

Remark 1. *Allowing stochastic encoders does not increase the zero-error feedback capacity with acausal SI.*

Our main result is presented in the following two theorems, which together provide a single-letter characterization of $C_{\mathrm{f},0}$. The first characterizes the channels for which it is positive, and the second provides a formula for $C_{\mathrm{f},0}$ when it is positive.

Theorem 1. *A necessary and sufficient condition for $C_{\mathrm{f},0}$ to be positive is*

$$\begin{aligned} &\forall\, s,\, s' \in \mathcal{S} \quad \exists\, x,\, x' \in \mathcal{X} \colon \\ &\quad \bigl(W(y|x,s)\, W(y|x',s') = 0, \ \forall\, y \in \mathcal{Y}\bigr). \end{aligned} \tag{4}$$

Theorem 2. *If $C_{\mathrm{f},0}$ is positive, then*

$$C_{\mathrm{f},0} = \min_{P_S} \max_{P_{U,X|S}} \min_{\substack{P_{Y|U,X,S}\colon \\ P_{Y|U=u,X,S} \in \mathscr{P}(W),\ \forall\, u \in \mathcal{U}}} I(U;Y) - I(U;S), \tag{5}$$

where U is an auxiliary chance variable taking values in a finite set $\mathcal{U}$, and the mutual informations are computed w.r.t. the joint PMF $P_S \times P_{U,X|S} \times P_{Y|U,X,S}$. Restricting X to be a function of U and S, i.e., $P_{U,X|S}$ to be of the form $P_{U|S} \times P_{X|U,S}$, where $P_{X|U,S}$ is $\{0,1\}$-valued, does not change the RHS of (5), nor does restricting the size of $\mathcal{U}$ to $|\mathcal{U}| \leq |\mathcal{X}|^{|\mathcal{S}|}$.

Remark 2. *The hypothesis in Theorem 2 that $C_{\mathrm{f},0}$ be positive is essential: the RHS of (5) may be positive even when $C_{\mathrm{f},0}$ is zero. In fact, the RHS of (5) is positive if, and only if, (iff)*

$$\forall\, (s,y) \in \mathcal{S} \times \mathcal{Y} \quad \exists\, x \in \mathcal{X} \colon\ W(y|x,s) = 0. \tag{6}$$

By considering the case of a single state, i.e., $|\mathcal{S}| = 1$, and invoking Shannon's result [1] that feedback can increase the zero-error capacity of a DMC, we readily obtain that feedback can also increase the zero-error capacity of an SD-DMC with acausal SI. But, in the presence of acausal SI, more is true. Unlike the stateless channel, here feedback can increase the capacity from zero:

Theorem 3. *The zero-error capacity of an SD-DMC with acausal SI can be positive with feedback yet zero without it.*

Because feedback can help only if the encoder uses the channel more than once, we obtain the following corollary:

Corollary 4. *On the SD-DMC with acausal SI, the error-free transmission of a single bit may require more than one channel use.*

III. Discussion

Shannon showed in [1] that the zero-error capacity of the (state-less) DMC $W(y|x)$ (with or without feedback) is positive iff

$$\exists\, x,\, x' \in \mathcal{X} \colon \bigl(W(y|x)\, W(y|x') = 0, \ \forall\, y \in \mathcal{Y}\bigr). \tag{7}$$

When the channel satisfies (7), then the error-free transmission of a single bit requires one channel use. Theorem 1 (cf. (4)) generalizes this to the SD-DMC with feedback and acausal SI. Unlike (7), Condition (4) is only for channels with feedback: the no-feedback zero-error capacity of the SD-DMC $W(y|x,s)$ with acausal SI can be zero also when the channel satisfies (4) (see Theorem 3). Moreover, (4) does not guarantee that one channel use suffices to transmit a single bit error-free (see Corollary 4).

In [1] Shannon also showed that, when it is positive, the zero-error feedback capacity of the DMC $W(y|x)$ is

$$\max_{P_X} \min_{y \in \mathcal{Y}} -\log \sum_{x \in \mathcal{X} \colon W(y|x) > 0} P_X(x). \tag{8}$$

Theorem 2 generalizes this result (cf. (5)) to the SD-DMC with feedback and acausal SI. That (5) reduces to (8) when $|\mathcal{S}| = 1$ becomes evident when we recall from [5] Ahlswede's alternative form for (8),

$$\max_{P_X} \min_{P_{Y|X} \in \mathscr{P}(W)} I(X;Y), \tag{9}$$

where the mutual information is computed w.r.t. the joint PMF $P_X \times P_{Y|X}$.

As we have seen in Section I, if the transition law $W(y|x,s)$ of the SD-DMC happens to be $\{0,1\}$-valued, then $C_{\text{f},0}$ is related to Ahlswede's AVC with acausal SI. In this case Theorems 1 and 2 can be greatly simplified:

Example 1. *If $W(y|x,s)$ is $\{0,1\}$-valued, then*

$$C_{\text{f},0} = \min_{s \in \mathcal{S}} \log \bigl| \bigl\{ y \in \mathcal{Y} \colon \exists\, x \in \mathcal{X} \text{ s.t. } W(y|x,s) > 0 \bigr\} \bigr|. \tag{10}$$

Remark 2 not withstanding, if $W(y|x,s)$ is $\{0,1\}$-valued, then the RHS of (5)—which in this case is equal to the RHS of (10)—is positive iff $C_{\text{f},0}$ is positive. This agrees with Ahlswede's observation in [8] that the formula for the (a- and m-) capacity of the general AVC $W(y|x,s)$ whose state is revealed acausally to the encoder not only applies when the capacity is positive but also determines whether it is positive.

IV. Selected Proofs

We prove Theorem 1 in Section IV-A and Theorem 3 in Section IV-B. The proofs of Theorem 2, Remarks 1 and 2, and Example 1 can be found in [9].

A. A Proof of Theorem 1

The direct part of Theorem 1 follows from the following:

Remark 3. *Consider an SD-DMC $W(y|x,s)$ with feedback and acausal SI. If (4) holds, then n_{bit} channel uses suffice for the error-free transmission of a bit, where n_{bit} is 1 if $|\mathcal{S}| = 1$, and is otherwise upper-bounded by*[1]

$$\frac{2\,|\mathcal{Y}| \log |\mathcal{S}| - \log |\mathcal{Y}|}{\log |\mathcal{Y}| - \log\bigl(|\mathcal{Y}| - 1\bigr)} + 1 + 2\,|\mathcal{Y}|. \tag{11}$$

In proving Remark 3 we focus on the case $|\mathcal{S}| \geq 2$, because the case $|\mathcal{S}| = 1$ follows directly from Shannon [1]. (In this case (4) is equivalent to (7).)

Before we prove Remark 3, we briefly describe the coding scheme that we propose. The zero-error capacity of the SD-DMC $W(y|x,s)$ with acausal SI can be zero without feedback but positive with feedback (Theorem 3), and it is not always possible to transmit a single bit error-free in only one channel use (Corollary 4). Our scheme thus requires more than one channel use, and it utilizes the feedback link.

The scheme has two phases. Phase 1 is not used to convey the bit but rather to reduce the decoder's ambiguity about the Phase-2 state-sequence. This is attained with an adaptive feedback code reminiscent of the one used in the first phase of Shannon's coding scheme for the stateless DMC [1]. But in our Phase 1, the encoder utilizes the Phase-1 state-sequence (albeit only causally). After Phase 1 the decoder computes the set of Phase-2 state-sequences of positive posterior probability given the Phase-1 outputs. This set can also be computed by the encoder thanks to the Phase-1 feedback. This enables the encoder to transmit the bit error-free in Phase 2. The feedback link is not used in Phase 2.

The condition in Theorem 1 ensures that Phase 1 and 2 are feasible. As we shall see, Phase 1 is feasible iff (6) holds, whereas Phase 2 is feasible iff (4) holds, where by Remark 2

$$(4) \implies (6) \quad \text{and} \quad (4) \;\not\Longleftarrow\; (6),$$

so feasibility is easier to attain in Phase 1 than in Phase 2.

Proof of Remark 3. The case $|\mathcal{S}| = 1$ follows from Shannon [1], and we hence assume that $|\mathcal{S}| \geq 2$. To transmit a single bit $m \in \{0,1\}$, we divide the blocklength-n_{bit} transmission into Phase 1 and Phase 2 of n_1 and n_2 channel uses, where

$$n_{\text{bit}} = n_1 + n_2. \tag{12}$$

A choice of $(n_{\text{bit}}, n_1, n_2)$ will be presented after we describe the two phases, beginning with Phase 1.

Let $\mathcal{S}^{n_1+n_2}$ denote the set of possible length-(n_1+n_2) state-sequences, and let $\mathcal{S}^{n_2}$ denote the set of possible state sequences occurring during Phase 2. Before the transmission begins, the encoder observes the entire state sequence $S^{n_1+n_2}$. The goal of Phase 1 is to produce a random subset $\boldsymbol{\mathcal{S}}_{n_1} \subseteq \mathcal{S}^{n_2}$ with the following three properties: 1) $\boldsymbol{\mathcal{S}}_{n_1}$ is determined by the Phase-1 outputs $Y_1, \ldots, Y_{n_1}$, so both encoder and decoder know $\boldsymbol{\mathcal{S}}_{n_1}$ before Phase 2 begins; 2) with probability one $\boldsymbol{\mathcal{S}}_{n_1}$ contains the Phase-2 state-sequence $S_{n_1+1}^{n_1+n_2}$; and 3) the cardinality of $\boldsymbol{\mathcal{S}}_{n_1}$ is upper-bound by

$$|\boldsymbol{\mathcal{S}}_{n_1}| \leq \left(\frac{|\mathcal{Y}| - 1}{|\mathcal{Y}|} \right)^{n_1} |\mathcal{S}|^{n_2} + |\mathcal{Y}|. \tag{13}$$

To that end, we partition the set $\boldsymbol{\mathcal{S}}_0 = \mathcal{S}^{n_2}$ into $|\mathcal{Y}|$ different subsets whose size is between $\bigl\lfloor |\boldsymbol{\mathcal{S}}_0| / |\mathcal{Y}| \bigr\rfloor$ and $\bigl\lceil |\boldsymbol{\mathcal{S}}_0| / |\mathcal{Y}| \bigr\rceil$. We index the $|\mathcal{Y}|$ subsets by the output alphabet $\mathcal{Y}$ and reveal the result to the encoder and decoder. To every pair $(s,y) \in \mathcal{S} \times \mathcal{Y}$ we assign an input $x(s,y) \in \mathcal{X}$ for which

$$W\bigl(y \big| x(s,y), s\bigr) = 0. \tag{14}$$

Such an $x(s,y)$ exists, because substituting s for both s and s' in (4) demonstrates that (4) implies that there exists a pair of inputs $x', x'' \in \mathcal{X}$ for which

$$W(y|x',s)\,W(y|x'',s) = 0, \ \forall\, y \in \mathcal{Y}, \tag{15}$$

i.e., for which for every $y \in \mathcal{Y}$ either $W(y|x',s)$ or $W(y|x'',s)$ is zero. We can thus choose $x(s,y)$ to be x' when $W(y|x',s)$ is zero and to be x'' when it is not.[2] If, thanks to its acausal SI, the encoder knows that the Time-1 state S_1 is s and that $S_{n_1+1}^{n_1+n_2}$ is in the subset of $\boldsymbol{\mathcal{S}}_0$ indexed by y, then at Time 1 it transmits $x(s,y)$. This choice guarantees by (14) that upon

[1] All logarithms in (11) are nonnegative, because (4) implies that $|\mathcal{Y}| \geq 2$.

[2] This is nothing else but (4) $\implies$ (6), which follows from Remark 2.

observing the Time-1 output Y_1 the decoder will know that the Phase-2 state-sequence is not an element of the subset of $\boldsymbol{\mathcal{S}}_0$ indexed by Y_1, and that it is thus in the $\boldsymbol{\mathcal{S}}_0$-complement of this subset, which we denote $\boldsymbol{\mathcal{S}}_1$. Note that: 1) both encoder and decoder know $\boldsymbol{\mathcal{S}}_1$ after Channel-Use 1; 2) $\boldsymbol{\mathcal{S}}_1$ contains $S_{n_1+1}^{n_1+n_2}$; and 3) the cardinality of $\boldsymbol{\mathcal{S}}_1$ is upper-bounded by

$$|\boldsymbol{\mathcal{S}}_1| \leq |\boldsymbol{\mathcal{S}}_0| - \left\lfloor \frac{|\boldsymbol{\mathcal{S}}_0|}{|\mathcal{Y}|} \right\rfloor \leq \frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\,|\boldsymbol{\mathcal{S}}_0| + 1. \tag{16}$$

Phase 1 continues in the same fashion: Let $i \in [2:n_1]$, and assume that the first $i-1$ channel uses have produced a random subset $\boldsymbol{\mathcal{S}}_{i-1}$ of $\mathcal{S}^{n_2}$ with the following three properties: 1) both encoder and decoder know $\boldsymbol{\mathcal{S}}_{i-1}$ after Channel-Use $(i-1)$; 2) $\boldsymbol{\mathcal{S}}_{i-1}$ contains $S_{n_1+1}^{n_1+n_2}$; and 3) the cardinality of $\boldsymbol{\mathcal{S}}_{i-1}$ is upper-bounded by

$$|\boldsymbol{\mathcal{S}}_{i-1}| \leq \frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\,|\boldsymbol{\mathcal{S}}_{i-2}| + 1. \tag{17}$$

After Channel-Use $(i-1)$, we partition $\boldsymbol{\mathcal{S}}_{i-1}$ into $|\mathcal{Y}|$ different subsets whose size is between $\lfloor |\boldsymbol{\mathcal{S}}_{i-1}|/|\mathcal{Y}| \rfloor$ and $\lceil |\boldsymbol{\mathcal{S}}_{i-1}|/|\mathcal{Y}| \rceil$. We index the subsets by the elements of the output alphabet $\mathcal{Y}$ and reveal the result to the encoder and decoder. If, thanks to its acausal SI, the encoder knows that the Time-i state S_i is s and that $S_{n_1+1}^{n_1+n_2}$ is an element of the subset of $\boldsymbol{\mathcal{S}}_{i-1}$ indexed by y, then at Time i it transmits $x(s,y)$. This choice guarantees by (14) that upon observing the Time-i channel output Y_i the decoder will know that the Phase-2 state-sequence is not an element of the subset indexed by Y_i, and that it is thus in the $\boldsymbol{\mathcal{S}}_{i-1}$-complement of this subset, which we denote $\boldsymbol{\mathcal{S}}_i$. Note that: 1) both encoder and decoder know $\boldsymbol{\mathcal{S}}_i$ after Channel-Use i; 2) $\boldsymbol{\mathcal{S}}_i$ contains $S_{n_1+1}^{n_1+n_2}$; and 3) the cardinality of $\boldsymbol{\mathcal{S}}_i$ is upper-bounded by

$$|\boldsymbol{\mathcal{S}}_i| \leq |\boldsymbol{\mathcal{S}}_{i-1}| - \left\lfloor \frac{|\boldsymbol{\mathcal{S}}_{i-1}|}{|\mathcal{Y}|} \right\rfloor \leq \frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\,|\boldsymbol{\mathcal{S}}_{i-1}| + 1. \tag{18}$$

Since this holds for every $i \in [1:n_1]$, the goal of Phase 1 is attained, and the first n_1 channel uses produce a random subset $\boldsymbol{\mathcal{S}}_{n_1}$ of $\mathcal{S}^{n_2}$ with the following three properties: 1) both encoder and decoder know $\boldsymbol{\mathcal{S}}_{n_1}$ before Phase 2 begins; 2) $\boldsymbol{\mathcal{S}}_{n_1}$ contains the Phase-2 state-sequence $S_{n_1+1}^{n_1+n_2}$; and 3) the cardinality of $\boldsymbol{\mathcal{S}}_{n_1}$ is upper-bound by

$$\begin{aligned}
|\boldsymbol{\mathcal{S}}_{n_1}| &\leq \left(\frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\right)^{n_1} |\boldsymbol{\mathcal{S}}_0| + \sum_{i=0}^{n_1-1} \left(\frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\right)^{i} && (19)\\
&= \left(\frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\right)^{n_1} |\mathcal{S}|^{n_2} + \frac{|\mathcal{Y}|^{n_1} - \big(|\mathcal{Y}|-1\big)^{n_1}}{|\mathcal{Y}|^{n_1-1}} && (20)\\
&\leq \left(\frac{|\mathcal{Y}|-1}{|\mathcal{Y}|}\right)^{n_1} |\mathcal{S}|^{n_2} + |\mathcal{Y}|. && (21)
\end{aligned}$$

We next turn to Phase 2 whose goal is to transmit the bit error-free. To that end, the encoder allocates to every bit value $m \in \{0,1\}$ and every state sequence $\mathbf{s}$ in $\boldsymbol{\mathcal{S}}_{n_1}$ a length-n_2 codeword $\mathbf{x}(m,\mathbf{s})$, where the codewords are chosen so that

$$\begin{aligned}
&\forall\, \mathbf{s},\, \mathbf{s}' \in \boldsymbol{\mathcal{S}}_{n_1} \quad \exists\, i \in [1:n_2]: \\
&\Big(W\big(y\big|x_i(0,\mathbf{s}),s_i\big)\, W\big(y\big|x_i(1,\mathbf{s}'),s_i'\big) = 0,\ \forall\, y \in \mathcal{Y}\Big).
\end{aligned} \tag{22}$$

(We will shortly show how this can be done.) If the bit to be sent is $m \in \{0,1\}$ and if the Phase-2 state-sequence is $\mathbf{s}$, then the encoder transmits in Phase 2 the codeword $\mathbf{x}(m,\mathbf{s})$. Condition (22) implies that, upon observing the realization $\mathbf{y} \in \mathcal{Y}^{n_2}$ of the Phase-2 output-sequence $Y_{n_1+1}^{n_1+n_2}$, the decoder, who knows $\boldsymbol{\mathcal{S}}_{n_1}$ and the codewords $\{\mathbf{x}(m,\mathbf{s})\}$, can determine the value of m error-free, because for the true realization $\mathbf{s} \in \boldsymbol{\mathcal{S}}_{n_1}$ of the Phase-2 state-sequence

$$\prod_{i=1}^{n_2} W\big(y_i\big|x_i(m,\mathbf{s}),s_i\big) > 0, \tag{23}$$

whereas (22) implies for $m' \neq m$

$$\prod_{i=1}^{n_2} W\big(y_i\big|x_i(m',\mathbf{s}'),s_i'\big) = 0,\ \forall\, \mathbf{s}' \in \boldsymbol{\mathcal{S}}_{n_1}. \tag{24}$$

The decoder can thus calculate $\prod W\big(y_i\big|x_i(\tilde{m},\mathbf{s}'),s_i'\big)$ for each $\tilde{m} \in \{0,1\}$ and $\mathbf{s}' \in \boldsymbol{\mathcal{S}}_{n_1}$ and produce the $\tilde{m}$ for which this product is positive for some $\mathbf{s}' \in \boldsymbol{\mathcal{S}}_{n_1}$.

One (inefficient) way to achieve (22) is the following. Let $x^\star$ be an arbitrary fixed element of $\mathcal{X}$, and for every pair $s,\, s' \in \mathcal{S}$ choose a pair $x(s,s'),\, x'(s,s') \in \mathcal{X}$ for which

$$W\big(y\big|x(s,s'),s\big)\, W\big(y\big|x'(s,s'),s'\big) = 0,\ \forall\, y \in \mathcal{Y}. \tag{25}$$

By (4) such a pair $x(s,s'),\, x'(s,s')$ exists. Now choose $n_2 \geq |\boldsymbol{\mathcal{S}}_{n_1}|^2$; allocate to every ordered pair $(\mathbf{s},\mathbf{s}') \in \boldsymbol{\mathcal{S}}_{n_1} \times \boldsymbol{\mathcal{S}}_{n_1}$ a different index $i \in \big[1 : |\boldsymbol{\mathcal{S}}_{n_1}|^2\big]$; and for the allocated index i choose $x_i(0,\mathbf{s}) = x(s_i,s_i')$ and $x_i(1,\mathbf{s}') = x'(s_i,s_i')$, and thus guarantee, by (25), that

$$\Big(W\big(y\big|x_i(0,\mathbf{s}),s_i\big)\, W\big(y\big|x_i(1,\mathbf{s}'),s_i'\big) = 0,\ \forall\, y \in \mathcal{Y}\Big). \tag{26}$$

The above specifies $|\boldsymbol{\mathcal{S}}_{n_1}|$ out of $n_2 \geq |\boldsymbol{\mathcal{S}}_{n_1}|^2$ symbols of each codeword $\mathbf{x}(m,\mathbf{s})$. How we choose the other $n_2 - |\boldsymbol{\mathcal{S}}_{n_1}|$ symbols is immaterial. To be explicit, we choose each of them to be $x^\star$. The described choice of the codewords $\{\mathbf{x}(m,\mathbf{s})\}$ clearly satisfies (22). Hence, it would only remain to exhibit some choice of the triple $(n_{\text{bit}}, n_1, n_2)$ satisfying (12) and $n_2 \geq |\boldsymbol{\mathcal{S}}_{n_1}|^2$. This can be done using (13), but the resulting value of n_{bit} need not be upper-bounded by (11). To fix this, we allocate the indices more efficiently. Note that for every $i \in \big[1 : |\boldsymbol{\mathcal{S}}_{n_1}|^2\big]$ the above choice of the codewords $\{\mathbf{x}(m,\mathbf{s})\}$ allocates meaningful values to the i-th symbols of only two codewords, namely $\mathbf{x}(0,\mathbf{s})$ and $\mathbf{x}(1,\mathbf{s}')$, where $(\mathbf{s},\mathbf{s}')$ is the ordered pair to which we allocated Index i. More efficiently, we can allocate the same index i to several distinct pairs $(\mathbf{s},\mathbf{s}')$. (Still, we let $x_i(0,\mathbf{s}) = x(s_i,s_i')$ and $x_i(1,\mathbf{s}') = x'(s_i,s_i')$ when Index i has been allocated to the ordered pair $(\mathbf{s},\mathbf{s}')$, and we choose each codeword symbol that has not been assigned a value to be $x^\star$.) This works whenever any two distinct pairs $(\mathbf{s},\mathbf{s}')$, $(\tilde{\mathbf{s}},\tilde{\mathbf{s}}')$ that are allocated the same index i satisfy $\mathbf{s} \neq \tilde{\mathbf{s}}$ and $\mathbf{s}' \neq \tilde{\mathbf{s}}'$, because then every codeword symbol $x_i(m,\mathbf{s})$ is assigned exactly one value. An efficient way to allocate the indices is the following. Choose any integer n_2 that satisfies

$$n_2 \geq |\boldsymbol{\mathcal{S}}_{n_1}|. \tag{27}$$

(An explicit choice for which n_{bit} is upper-bounded by (11) will be given in (29).) Index the elements of $\boldsymbol{\mathcal{S}}_{n_1}$ by $\bigl[1 : |\boldsymbol{\mathcal{S}}_{n_1}|\bigr]$, where $\mathbf{s}(j)$ denotes the element of $\boldsymbol{\mathcal{S}}_{n_1}$ indexed by j, and allocate to every ordered pair $\bigl(\mathbf{s}(k), \mathbf{s}(\ell)\bigr)$, where $k, \ell \in \bigl[1 : |\boldsymbol{\mathcal{S}}_{n_1}|\bigr]$, the index

$$i(k,\ell) = \bigl(\ell - k \mod |\boldsymbol{\mathcal{S}}_{n_1}|\bigr) + 1 \quad \in \bigl[1 : |\boldsymbol{\mathcal{S}}_{n_1}|\bigr]. \tag{28}$$

By (28) any two distinct pairs $\bigl(\mathbf{s}(k), \mathbf{s}(\ell)\bigr)$, $\bigl(\mathbf{s}(k'), \mathbf{s}(\ell')\bigr)$ that are allocated the same index i satisfy $k \neq k'$ and $\ell \neq \ell'$, so $\mathbf{s}(k) \neq \mathbf{s}(k')$ and $\mathbf{s}(\ell) \neq \mathbf{s}(\ell')$.

To conclude the direct part, it remains to exhibit some choice of the triple $(n_{\text{bit}}, n_1, n_2)$ satisfying (12) and (27). By (13) these are satisfied if $n_{\text{bit}} = n_1 + n_2$, where

$$n_1 = \left\lceil \frac{2\,|\mathcal{Y}| \log|\mathcal{S}| - \log|\mathcal{Y}|}{\log|\mathcal{Y}| - \log\bigl(|\mathcal{Y}| - 1\bigr)} \right\rceil \quad \text{and} \quad n_2 = 2\,|\mathcal{Y}|, \tag{29}$$

and for this choice n_{bit} is upper-bounded by (11). □

We next prove the converse of Theorem 1:

Converse. To show that (4) is necessary for $C_{\mathrm{f},0}$ to be positive, we need to prove that if (4) does not hold, i.e., if there exists a pair of states $s,\, s' \in \mathcal{S}$ such that

$$\nexists\, x,\, x' \in \mathcal{X} \colon \bigl(W(y|x,s)\, W(y|x',s') = 0,\ \forall\, y \in \mathcal{Y}\bigr), \tag{30}$$

then it is impossible to transmit a single bit error-free. Condition (30) can be alternatively expressed as

$$\forall\, x,\, x' \in \mathcal{X} \quad \exists\, y \in \mathcal{Y} \colon\ W(y|x,s)\, W(y|x',s') > 0, \tag{31}$$

which makes the claim almost obvious. Indeed, (31) implies that if the state sequence is all s or all s' then—during every channel use and irrespective of the inputs x, x' that we choose—the pairs (x,s) and (x',s') can produce the same output. This implies that for every pair of messages $m, m' \in \mathcal{M}$ and every encoding mappings there exists an output sequence of positive probability conditional on each of the following two events: 1) the message is m, and the state sequence is all s; or 2) the message is m', and the state sequence is all s'. □

B. A Proof of Theorem 3

We use the following lemma to establish Theorem 3:

Lemma 1. *Without feedback, a sufficient condition for the zero-error capacity of the SD-DMC $W(y|x,s)$ with acausal SI to be zero is*

$$\begin{gathered}\exists\, s \in \mathcal{S} \quad \forall\, x \in \mathcal{X} \quad \exists\, s' \in \mathcal{S} \quad \forall\, x' \in \mathcal{X} \quad \exists\, y \in \mathcal{Y} \colon \\ W(y|x,s)\, W(y|x',s') > 0.\end{gathered} \tag{32}$$

Proof. We prove that if (32) holds, then without feedback it is impossible to transmit a single bit error-free. Let $\mathcal{M} = \{0,1\}$ be the set of possible values for the bit to be transmitted, and fix a blocklength n, an encoding mapping $f \colon \mathcal{M} \times \mathcal{S}^n \to \mathcal{X}^n$, and two disjoint decoding sets $\mathcal{D}_m \subseteq \mathcal{Y}^n$, $m \in \{0,1\}$. By (32) there exists some state $s \in \mathcal{S}$ for which

$$\begin{gathered}\forall\, x \in \mathcal{X} \quad \exists\, s' \in \mathcal{S} \quad \forall\, x' \in \mathcal{X} \quad \exists\, y \in \mathcal{Y} \colon \\ W(y|x,s)\, W(y|x',s') > 0.\end{gathered} \tag{33}$$

Let $\mathbf{s} \in \mathcal{S}^n$ be the all s state sequence, and let $\mathbf{x} = f(0, \mathbf{s})$. By (33) we obtain that for every $i \in [1:n]$ there exists some $s' \in \mathcal{S}$, say $s'(i)$, for which

$$\forall\, x' \in \mathcal{X} \quad \exists\, y \in \mathcal{Y} \colon\ W(y|x_i,s_i)\, W\bigl(y|x', s'(i)\bigr) > 0. \tag{34}$$

Let $\mathbf{s}' \in \mathcal{S}^n$ be the state sequence for which $s'_i = s'(i)$, $i \in [1:n]$, and let $\mathbf{x}' = f(1, \mathbf{s}')$. By (34)

$$\exists\, \mathbf{y} \in \mathcal{Y}^n \colon \prod_{i=1}^{n} \bigl(W(y_i|x_i,s_i)\, W(y_i|x'_i,s'_i)\bigr) > 0. \tag{35}$$

This concludes the proof, because it implies that the output sequence $\mathbf{y}$ has a positive posterior probability conditional on any one of the following two events: 1) the value of the transmitted bit is 0 and the state sequence $\mathbf{s}$; or 2) the value of the transmitted bit is 1 and the state sequence $\mathbf{s}'$. □

Theorem 3 follows from Theorem 1, Lemma 1, and the following example:

Example 2. *Suppose $\mathcal{X} = \{0,1\}$ and $\mathcal{S} = \mathcal{Y} = \{1,2,3,4,5\}$. For every $x \in \mathcal{X}$ and $s \in \mathcal{S}$ define $\mathcal{Y}_{x,s}$ according to Table I, and let $W(y|x,s)$ be such that*

$$\bigl\{y \in \mathcal{Y} \colon W(y|x,s) > 0\bigr\} = \mathcal{Y}_{x,s},\ \forall\, (x,s) \in \mathcal{X} \times \mathcal{S}. \tag{36}$$

Then, the SD-DMC $W(y|x,s)$ satisfies both (4) *and* (32).

TABLE I
NONZERO TRANSITIONS OF THE SD-DMC IN EXAMPLE 2.

$\mathcal{Y}_{x,s}$		s 1	 2	 3	 4	 5
x	0	{ 2,3 }	{ 1,5 }	{ 1,2 }	{ 2,3 }	{ 1,2 }
	1	{ 4,5 }	{ 3,4 }	{ 4,5 }	{ 1,5 }	{ 3,4 }

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] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control Theory*, Vol. 9, No. 1, pp. 19–31, 1980.

[3] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source," *Proc. of IEEE Int. Symp. on Inf. Theory (ISIT)*, pp. 1506–1510, Sep. 2005.

[4] G. Dueck, "The zero error feedback capacity region of a certain class of multiple-access channels," *Problems of Control and Inf. Theory*, Vol. 14, No. 2, pp. 89–103, 1985.

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

[6] J. M. Ooi and G. W. Wornell, "Fast iterative coding techniques for feedback channels," *IEEE Trans. Inf. Theory*, IT-44, No. 7, pp. 2960–2976, Nov. 1998.

[7] R. Ahlswede, "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity," *Ann. of Math. Stat.*, Vol. 41, No. 3, pp. 1027–103, Jun. 1970.

[8] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inf. Theory*, IT-32, No. 5, pp. 621–629, Sep. 1985.

[9] A. Bracher and A. Lapidoth, The Zero-Error Feedback Capacity of the Gelfand-Pinsker Channel, *draft*, 2016.