Error-Free Communication Over State-Dependent Channels with Variable-Length Feedback

Mladen Kovačević, Carol Wang, and Vincent Y. F. Tan

Abstract

This paper studies the fundamental limits of communication over state-dependent discrete memoryless channels with noiseless feedback, under the assumption that the communicating parties are allowed to use variable-length coding schemes. Various cases are analyzed, with the employed coding schemes having either bounded or unbounded codeword lengths, and with state information revealed to the encoder and/or decoder in a strictly causal, causal, or non-causal manner. In each of these settings, necessary and sufficient conditions for positivity of the zero-error capacity are obtained and it is shown that, whenever the zero-error capacity is positive, it equals the conventional vanishing-error capacity. Moreover, it is shown that the vanishing-error capacity of state-dependent channels is not increased by the use of feedback and variable-length coding. Both these kinds of capacities of state-dependent channels with feedback are thus fully characterized. A comparison of the results with the recently solved fixed-length case is also given.

Index Terms

Channels with states, Gelfand–Pinsker, Shannon strategy, feedback, variable-length code, zero-error code, channel capacity.

I. INTRODUCTION

This work considers the zero-error capacity of state-dependent channels. In contrast to the standard notion of capacity, which allows an asymptotically vanishing probability of error at the decoder, the output of a zero-error decoder must always be correct. We focus on variable-length coding schemes with access to noiseless feedback at the encoder, and derive the zero-error capacity under different models of state information availability.

At first glance, the requirement of zero-error decoding is quite stringent. Indeed, the zero-error capacity of many channels, including the binary symmetric channel (BSC), is 0, and this holds even for the relatively benign binary erasure channel (BEC). However, it became evident after Burnashev’s work on the error exponent of discrete memoryless channels (DMCs) [3] that, with variable-length encoding and noiseless feedback, error-free communication is not only possible over a large class of DMCs, but the zero-error capacity of such channels is equal to their (vanishing-error) Shannon capacity. The fact that the channel capacity can in many cases be achieved with error probability being fixed to zero is quite interesting, and this is what motivated us to study the corresponding problem in the more general context of channels with states. Furthermore, analyzing fundamental limits of communication under variable-length coding is important in systems with feedback, and even more so in systems where side information about the channel is available as well. In fact, most communication schemes in such systems, e.g., the ARQ mechanism,
are adaptive in nature and the length of transmission depends on the side information and the information obtained through the feedback link.

Communicating with zero error has been considered previously in many settings, starting with Shannon’s work [12] on the zero-error capacity of DMCs with and without feedback, and under fixed-length encoding. We refer the reader to [3] for a review of this area. The work most closely related to ours is that of Bracher and Lapidoth [2] where the zero-error feedback capacity of state-dependent channels was determined, under the assumptions that fixed-length encoding is being used and that state information is available only at the encoder. Our work extends these results to the variable-length case, while also analyzing other models of state information availability. We also mention here the work of Zhao and Permuter [16] where the authors characterized the zero-error feedback capacity under fixed-length encoding of channels with state information at both the encoder and the decoder, but in which the state process is not necessarily memoryless and is even allowed to depend on the channel inputs.

The main contributions of this work are threefold. First, we show how to relate the zero-error capacity with variable-length encoding and feedback (“zero-error VLF capacity”) to the standard vanishing-error capacity with fixed-length encoding and no feedback (Theorem 1 and Theorem 3). Second, we give necessary and sufficient conditions for the zero-error VLF capacity to be positive under different models of state information availability at the encoder and the decoder, including none, strictly causal (only past states are available), causal (past and current states available), and non-causal (all states including future states available). These results are summarized in Theorem 2. These theorems together completely characterize the zero-error VLF capacity of state-dependent channels. And third, we obtain analogous results for the zero-error feedback capacity under bounded-length coding (Theorem 5 and Theorem 6). We show that the conditions for positivity are in this case the same as for fixed-length coding schemes, although the final capacity expressions may be different.

In addition, as part of our proofs, we derive an upper bound on the $\epsilon$-error capacity for state-dependent channels (Lemma 10), which may be of independent interest.

A. The Channel Model, Definitions, and Notation

Let $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ denote the set of channel input letters, the set of channel output letters, and the set of channel states, respectively, all of which are assumed finite. A state-dependent discrete memoryless channel (SD-DMC) is described by conditional probability distributions $W(y|x, s)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $s \in \mathcal{S}$, where the states are drawn i.i.d. across all channel uses according to a distribution $Q(\cdot)$ on $\mathcal{S}$. To avoid discussing trivial cases, we assume throughout the paper that $|\mathcal{X}| \geq 2$, $|\mathcal{Y}| \geq 2$, $|\mathcal{S}| \geq 1$; that all states in $\mathcal{S}$ have positive probability:

$$\forall s \in \mathcal{S} \quad Q(s) > 0;$$

and that every channel output is reachable from at least one input in at least one state:

$$\forall y \in \mathcal{Y} \quad \exists x \in \mathcal{X}, s \in \mathcal{S} \quad W(y|x, s) > 0.$$

We use the symbol $\mathcal{M}$ to denote the set of messages to be transmitted in a particular communication setting. The symbols $M, X, Y, S$ denote random variables taking values in $\mathcal{M}, \mathcal{X}, \mathcal{Y}, \mathcal{S}$, respectively, and the lower case versions $m, x, y, s$ denote their realizations, i.e., elements of $\mathcal{M}, \mathcal{X}, \mathcal{Y}, \mathcal{S}$. The random variable $M$ representing the transmitted message is always assumed to be uniform over $\mathcal{M}$. $X^\infty$ is a shorthand for a random infinite sequence $(X_1, X_2, \ldots)$, $X^n$ for a random finite sequence $(X_1, \ldots, X_n)$, and $X^n_k$ for a subsequence $(X_k, \ldots, X_n)$ (hence $X^n_1 \equiv X^n$).

We say that the encoder (resp. decoder) has causal state information if, before the $n$’th channel use, it can see all the past channel states as well as the current—$n$’th—state, i.e., it is given the state sequence $S^n$ and can use it in the $n$’th time slot for the encoding (resp. decoding) operation. State information is said to be strictly causal if only past states ($S^{n-1}$) are available at time instant $n$, and it is said to be
non-causal if all the states \((S^\infty)\) are available at any time instant. We consider the following cases of state information availability:

\[
SI := \left\{ (-, -), (sc, -), (nc, -), (sc, c), (c, c), (nc, c), (nc, nc) \right\},
\]

where the first (resp. second) coordinate of \(s_i \in SI\) denotes state information available at the encoder (resp. decoder) and -/sc/c/nc stand for none/strictly-causal/causal/non-causal. The cases that have been omitted from SI are discussed in Remark 1 to follow.

**Definition 1.** Consider an SD-DMC with causal state information at both the encoder and the decoder. An \((\ell, |\mathcal{M}|, \epsilon)\) variable-length feedback (VLF) code for the message set \(\mathcal{M}\), where \(\ell\) is a positive real and \(0 \leq \epsilon \leq 1\), is defined by:

1. A sequence of encoders \(f_n : \mathcal{M} \times Y^{n-1} \times S^n \rightarrow \mathcal{X}, n \geq 1\), defining channel inputs \(X_n = f_n(M, Y^{n-1}, S^n)\);
2. A sequence of decoders \(g_n : Y^n \times S^n \rightarrow \mathcal{M}, n \geq 1\), defining decoder’s estimates of the transmitted message, \(g_n(Y^n, S^n)\);
3. A positive integer-valued random variable \(\tau\) (a stopping time of the receiver filtration \(\{\sigma(Y^n, S^n)\}_{n=0}^\infty\)) representing the code length and satisfying \(E[\tau] \leq \ell\).

The decoder’s final decision is computed at time \(\hat{M} = g_\tau(Y^{\tau}, S^{\tau})\), and it must satisfy \(\mathbb{P}[\hat{M} \neq M] \leq \epsilon\).

When \(\epsilon = 0\), such a code is called a zero-error VLF code.

If there exists a constant \(b < \infty\) such that \(\tau \leq b\), such a code is called a bounded-length feedback code, and if \(\tau = b = \ell\), it is called a fixed-length feedback code.

Definitions for the cases of state information availability \(s_i \in SI\) are the same except \(S^n\) in 1)–3) is replaced by \(S^0, S^{n-1}, or S^\infty\) accordingly.

The rate of an \((\ell, |\mathcal{M}|, \epsilon)\) code is defined as \(\frac{1}{\ell} \log_2 |\mathcal{M}|\). The vanishing-error capacity of a given channel is defined in the usual way as the supremum of the code rates that are asymptotically achievable (as \(\ell \to \infty\)) with arbitrarily small error probability. The zero-error capacity of a given channel is the supremum of the rates of all zero-error codes for that channel [12]. Capacity is always denoted by \(C\), with subscripts and superscripts indicating the channel and the coding schemes with respect to which it is defined as follows:

- The first subscript is either “0” or “v” and serves to distinguish between the zero-error and the vanishing-error case;
- The second subscript is either “f” or “-)” depending on whether or not the feedback link is present;
- The third subscript is \(vl, bl, or fl\), indicating that the capacity in question is defined with respect to variable-length, bounded-length, or fixed-length codes;
- Superscripts from the set \(SI\) (see (3)) are used to denote state information availability at the encoder and the decoder.

For example, \(C_{0,v,fl}^{nc}\) is the vanishing-error capacity under variable-length feedback codes, where the encoder is given state information in a non-causal manner and the decoder is given no state information; \(C_{0,c,bl}^{sc}\) is the zero-error capacity under bounded-length coding without feedback, and with state information revealed both to the encoder and to the decoder in a causal manner; etc.

**Remark 1.** To conclude this section we explain briefly why, of the sixteen possible cases in \(\{-, sc, c, nc\}^2\), only the cases in SI in (3) are being considered.

First, leaving out the four cases with strictly causal state information at the decoder, \(\{-, sc, c, nc\} \times \{sc\}\), is not a loss in generality. This is because strictly causal state information at the decoder can be “made causal” by simply delaying the decoding process by one time slot. Hence, from the viewpoint of capacity issues, \((*, sc)\) is equivalent to \((*, c)\) for any \(* \in \{-, sc, c, nc\}\).

Second, of the four possible cases where the decoder has non-causal state information \(\{\{-, sc, c, nc\} \times \{nc\}\}\) we only consider one—\((nc, nc)\). This is also not a loss in generality because knowing future states can be helpful to the decoder only if the encoder also knows future states and is using them in the
encoding operation. Otherwise, these states are independent of the channel inputs and no information can be extracted from them. Hence, for our purposes, (*, nc) is equivalent to (*, c) for any * ∈ {/, sc, c}.

Finally, note that (·, c) has not been included in SI either. This case is quite subtle and we choose to discuss it separately in Section [V]. The main issue here is that it is not clear how to define the code length, i.e., the stopping time τ (see Definition [I]). Namely, the decoder making a decision at time instant τ' does not necessarily mean that the transmission is over from the encoder’s perspective. This is because the decoder’s decision is based on the channel outputs and the channel states that it sees, and hence the encoder, not knowing the states, may actually never realize that the decoding is completed and may continue transmitting. As we shall see, this is especially important in the zero-error setting. Despite this difficulty, we shall give in Section [V] a sufficient condition for the zero-error capacity to be positive, and a necessary and sufficient condition for \( C_{0,f,VL} \) to be positive. ▲

II. Vanishing-Error Capacity

As already mentioned in Section [I] one of the results of this paper is the statement that the zero-error VLF capacity of an SD-DMC, whenever positive, equals the vanishing-error capacity of the same channel. For this reason, we first study the vanishing-error capacity and show that this fundamental limit remains unchanged even if the transmitter and the receiver are allowed to use variable-length feedback-dependent coding schemes. For SD-DMCs with state information only at the transmitter, the fact that feedback does not increase the capacity under fixed-length coding was shown in [10].

**Theorem 1.** For every \( si \in SI \), \( C_{si,f,VL}^c = C_{si,f,VL} = C_{si,f,fl} = C_{si,f,sc} \).

**Proof:** Deferred to the Appendix.

Consequently, we shall denote the vanishing-error capacity (with or without feedback) simply by \( C_{si} \) in the rest of the paper.

By Theorem [I] and the known expressions for the capacity \( C_{si} \) we get:

\[
C_{\downarrow, f, VL}^c = \max_{P_X} I(X; S) = \max_{P_\mathcal{U}} I(U; X) = \max_{P_{U|S}, f: U \times S \to \mathcal{X}} I(U; Y) = \max_{P_{U|S}, f: U \times S \to \mathcal{X}} I(U; Y) - I(U; S)
\]

\[
C_{\downarrow, f, VL}^c = \max_{P_X} I(X; S) = \max_{P_{U|S}, f: U \times S \to \mathcal{X}} I(U; Y) - I(U; S)
\]

\[
C_{\downarrow, f, VL}^c = \max_{P_X} I(X; S) = \max_{P_{U|S}, f: U \times S \to \mathcal{X}} I(U; Y) - I(U; S)
\]

\[
C_{\downarrow, f, VL}^c = \max_{P_X} I(X; S) = \max_{P_{U|S}, f: U \times S \to \mathcal{X}} I(U; Y) - I(U; S)
\]

\[
C_{\downarrow, f, VL}^c = \max_{P_X} I(X; Y|S) = \max_{P_X} I(X; Y|S)
\]

where \( U \) denotes an auxiliary random variable with alphabet \( \mathcal{U} \) of cardinality \( |\mathcal{U}| \leq |\mathcal{X}| |\mathcal{S}| \).

III. Zero-Error Capacity: Variable-Length Codes

For a DMC \( W(\cdot|\cdot) \), a necessary and sufficient condition for \( C_{0,f,VL} > 0 \) is

\[
\exists x \in \mathcal{X}, y \in \mathcal{Y} \quad W(y|x) = 0.
\]

This can be concluded from Burnashev’s characterization of the error exponent of DMCs under VLF coding schemes [8]—the corresponding error exponent is infinite if and only if (9) holds. When \( W(y|x) = 0 \), \( y \) is said to be a *disprover* for \( x \). Whenever such a disprover exists, one bit can be transmitted error free in a finite expected number of channel uses as follows. In the first two channel uses the transmitter sends \( x, x' \) for 0 and \( x', x \) for 1, where \( x' \) is any other input letter with \( W(y|x') > 0 \) (such an input letter necessarily exists by our assumption (2)). Due to the fact that \( W(y|x) = 0 \), if the letters obtained at the output are

\[\text{This terminology is from [9] and reflects the fact that such an output } y \text{ “disproves” the possibility that } x \text{ was the channel input.}\]
\( \neg y, y \), where \( \neg y \) denotes an arbitrary letter from \( \mathcal{Y} \setminus \{y\} \), the receiver concludes that 0 must have been transmitted; similarly, if the letters obtained at the output are \( y, \neg y \), then 1 must have been transmitted; finally, if the letters obtained at the output are \( \neg y, \neg y \), the procedure is repeated (the transmitter sees the output letters through feedback and knows whether or not it should repeat the transmission). The expected number of retransmissions needed to complete the protocol is finite because the event \( x' \rightarrow y \) has positive probability, i.e., \( W(y|x') > 0 \). Therefore, in a finite expected number of channel uses the receiver will recover the bit correctly.

For a study of zero-error VLF communication over a special kind of DMC—the so-called Z-channel—see also [14].

In the following statement we identify generalizations of the notion of disprover for state-dependent channels. The idea is that, in each case, there is a disprover output \( y \) which the sender can choose to avoid (using the disproved input \( x \)), so that seeing \( y \) gives error-free information that \( y \) was intended as a possible output.

**Theorem 2.** Necessary and sufficient conditions for positivity of the zero-error VLF capacity of SD-DMCs are as follows:

(a) \( C_{0,f,\text{VL}}^{z} > 0 \) if and only if there is a pair \( (x, y) \in \mathcal{X} \times \mathcal{Y} \) such that \( y \) disproves \( x \) in all states:

\[
\exists x \in \mathcal{X}, \ y \in \mathcal{Y} \quad \forall s \in \mathcal{S} \quad W(y|x, s) = 0. \quad (10)
\]

(b) \( C_{0,f,\text{VL}}^{\text{sc},z} > 0 \) if and only if \( (10) \) holds.

(c) \( C_{0,f,\text{VL}}^{\text{nc},z} > 0 \) if and only if there is an output letter \( y \) which is a disprover in each state:

\[
\exists y \in \mathcal{Y} \quad \forall s \in \mathcal{S} \quad \exists x \in \mathcal{X} \quad W(y|x, s) = 0. \quad (11)
\]

(d) \( C_{0,f,\text{VL}}^{\text{sc},\text{nc}} > 0 \) if and only if \( (11) \) holds.

(e) \( C_{0,f,\text{VL}}^{\text{nc},\text{nc}} > 0 \) if and only if there is at least one state with a disprover:

\[
\exists x, x' \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S} \quad W(y|x, s) = 0 \land W(y|x', s) > 0. \quad (12)
\]

(f) \( C_{0,f,\text{VL}}^{\text{nc},c} > 0 \) if and only if \( (12) \) holds.

(g) \( C_{0,f,\text{VL}}^{\text{nc},c} > 0 \) if and only if \( (12) \) holds.

(h) \( C_{0,f,\text{VL}}^{\text{nc},\text{nc}} > 0 \) if and only if \( (12) \) holds.

**Proof:** (a) In the case \( \text{si} = (-,-) \) when neither side has any state information, the channel is equivalent to the DMC \( \tilde{W}(y|x) := \sum_{s \in \mathcal{S}} Q(s)W(y|x, s) \). The condition \( (10) \) is the condition for positivity of the zero-error VLF capacity of this DMC; see (9) (recall that \( Q(s) > 0 \) for all \( s \in \mathcal{S} \) by our assumption \( (11) \)).

(b) Let \( \text{si} = (\text{sc},-) \). Since \( C_{0,f,\text{VL}}^{\text{sc},z} \geq C_{0,f,\text{VL}}^{z} \), we only need to show the “only if” part. So suppose that \( (10) \) is not satisfied, meaning that for every input-output pair \( (x, y) \in \mathcal{X} \times \mathcal{Y} \) there exists a state \( s_{x,y} \) where \( W(y|x, s_{x,y}) > 0 \). Then every output sequence \( y_{1} \cdots y_{n} \) can be produced by every input sequence \( x_{1} \cdots x_{n} \) with positive probability (if the state sequence happens to be \( s_{x_{1},y_{1}} \cdots s_{x_{n},y_{n}} \)). This means that the decoder cannot decide with certainty at any point in time what the transmitted message was. Therefore, zero-error communication in a finite average number of channel uses is impossible.

The main point in this argument, stated informally, is that the encoder cannot avoid the “bad” states because it only sees the past ones, i.e., no matter which letter it chooses to send in the current time slot, the channel state can always be such that the decoder remains confused.

(c) Now consider the case \( \text{si} = (c,-) \). We first prove sufficiency of \( (11) \) using a procedure analogous to the one for DMCs outlined in the paragraph following (9). Let \( y \) be an output letter claimed to exist in \( (11) \). For every \( s \in \mathcal{S} \) choose an input letter \( x_{s} \) such that \( W(y|x_{s}, s) = 0 \). Also, for every \( s \in \mathcal{S} \) define \( x'_{s} \) to be some input letter with \( W(y|x'_{s}, s) > 0 \), if it exists; otherwise, pick \( x'_{s} \) to be any letter different from \( x_{s} \) (note that a letter \( x' \) with \( W(y|x'_s, s) > 0 \) exists for at least one state \( s \); see (2)). If the states realized in the first two time slots are \( s_{1}, s_{2} \), the transmitter sends \( x_{s_{1}}, x'_{s_{2}} \) for 0 and \( x'_{s_{1}}, x_{s_{2}} \) for 1 (the transmitter knows the channel state before it sends a letter). Since \( W(y|x_{s}, s) = 0 \), if the letters obtained at the output
are \(-y, y\), the receiver concludes that 0 must have been transmitted; if the letters obtained at the output are \(y, -y\), then 1 must have been transmitted; if the letters obtained at the output are \(-y, -y\), the procedure is repeated in the next two slots, and so on. In this way the receiver will recover the transmitted bit in a finite expected number of channel uses \(k\), implying that \(C_{6,f,\text{VLF}}^{sc,c}\geq \frac{1}{k} > 0\).

To prove the converse, suppose that (11) does not hold, i.e., for every output letter \(y\) there exists a state \(s_y\) such that \(W(y|x, s_y) > 0\) for all input letters \(x\). Then for any output sequence \(y_1 \cdots y_n\) the state sequence \(s_{y_1} \cdots s_{y_n}\) produces \(y_1 \cdots y_n\) with positive probability on any input \(x_1 \cdots x_n\). This means that the decoder cannot be certain, at any time instant \(n\), what was the transmitted message, and hence zero-error communication in a finite average number of channel uses is impossible.

\((d)\) The above proof for the \((c, -)\) case goes through for the \((nc, -)\) case as well.

\((e)\) Let \(s = (sc, c)\). Suppose that (12) holds and let \(x, x', y, s\) be input letters, an output letter, and a state satisfying \(W(y|x, s) = 0\), \(W(y|x', s) > 0\). Let also \(* \in \mathcal{X}'\) be an arbitrary input letter. In the first two channel uses the transmitter sends \(x, x', *\) for 0 and \(x', x, *\) for 1. (The third letter is irrelevant and will be ignored by the receiver. It is sent only because state information is delivered to the transmitter with a one-slot delay, so in order for it to learn the states in the first two slots, a dummy letter is transmitted in the third slot.) Now, if the letters obtained at the output are \(-y, y\) and the channel states in these two slots are \(s, s\), the receiver concludes that 0 must have been transmitted (the receiver can see the states); if the letters obtained at the output are \(-y, -y\) and the channel states in these two slots are \(s, s\), then 1 must have been transmitted; if none of the above two situations occurred, the procedure is repeated. In a finite expected number of channel uses one of the above two situations will happen and the receiver will recover the transmitted bit, implying that \(C_{6,f,\text{VLF}}^{sc,c} > 0\).

For the converse, notice that if (12) is not satisfied, then in every state an arbitrary output \(y\) is reachable from either all inputs, or from none of them. Clearly, any such state is useless for zero-error communication.

\((f)-(h)\) The proof of (e) goes through in these cases as well (the proof of achievability can in fact be slightly simplified here—there is no need for sending the dummy letter). We should note that, for these models of state information availability, the necessity and sufficiency of (12) also follows from the results in [4], where the error exponent of finite-state ergodic Markov channels with causal state information at both sides has been characterized. Namely, the error exponent of an SD-DMC with causal state information at both sides is infinite at all rates below capacity if and only if (12) holds.

**Remark 2** (Shannon strategy). The usual way of proving achievability results for channels with causal state information at the encoder, \(si = (c, -)\), is via the so-called Shannon strategy [13], [5]. In this approach, one considers the set \(U := \mathcal{X}^S\) of all functions from \(S\) to \(\mathcal{X}\) and a related DMC with input alphabet \(U\) and output alphabet \(\mathcal{X}\) defined by \(W'(y|u) := \sum_{s \in S} Q(s)W(y|u(s), s)\). A code is then defined over the alphabet \(U\) and the communication over the original SD-DMC \(W(\cdot, \cdot)\) proceeds as follows: if the channel state in the current—\(n'\)th—time slot is \(s\), and the \(n'\)th symbol of the codeword is \(u\), then the transmitter sends \(x = u(s)\). It was shown by Shannon that this strategy achieves the capacity of SD-DMCs with causal state information at the encoder [13]. The strategy also achieves the zero-error capacity under fixed-length coding [2]. Thus, in these settings, an SD-DMC \(W\) (with \(si = (c, -)\)) is essentially equivalent to a stateless DMC \(W'\).

We wish to point out here that optimality of the Shannon strategy continues to hold in the zero-error VLF setting as well. To see this, recall that a necessary and sufficient condition for positivity of the zero-error VLF capacity of the DMC \(W'\) is the existence of an input \(u \in U\) and an output \(y \in \mathcal{Y}\) such that \(W'(y|u) = 0\) (see (9)). This condition is equivalent to (11) because:

\[
\exists u \in U \quad W'(y|u) = 0 \iff \exists u \in U \quad \forall s \in S \quad W(y|u(s), s) = 0
\]

\[
\iff \forall s \in S \quad \exists x \in \mathcal{X} \quad W(y|x, s) = 0,
\]

where (13) follows from the definition of \(W'\), and (14) holds because \(u\) is a function from \(S\) to \(\mathcal{X}\). Notice that the condition in (14) is precisely (11).
Remark 3. Note that the condition \((\text{i})\) for positivity of the zero-error VLF capacity is the same for causal and non-causal state information at the transmitter. This is not the case in the fixed-length and bounded-length settings; see \([2]\) and Theorem 5 ahead.

Likewise, the condition \((\text{ii})\) states that the zero-error VLF capacities for the \((sc, c)\) and \((c, c)\) cases are either both positive or both zero. This is not the case when fixed-length or bounded-length codes are being used; see Theorem 5.

We next characterize the value of the zero-error VLF capacity of SD-DMCs. The statement is that, whenever this quantity is positive, it equals the vanishing-error capacity of the same channel. The analogous result for DMCs is known and can be inferred from Burnashev’s characterization of the error exponent of such channels under VLF coding (as mentioned above, the error exponent is infinite at all rates below \(C_\downarrow\) if and only if \((\text{ii})\) holds). However, there is a simpler and more direct way of proving this statement which can be extended to channels for which error exponents are not known. One such proof was given by Han and Sato for DMCs and bounded-length codes \([7]\), but virtually no changes are required to extend it to the setting we are interested in.

**Theorem 3.** Let \(si \in \text{SI}\). If \(C_{0,f,vl}^{si} > 0\), then \(C_{0,f,vl}^{si} = C_{\downarrow}^{si}\).

**Proof:** Obviously, \(C_{0,f,vl}^{si} \leq C_{\downarrow}^{si} \equiv C_{\downarrow}^{si}\). To demonstrate that the reverse inequality also holds whenever \(C_{0,f,vl}^{si} > 0\), we shall use a slight modification of the Han–Sato coding scheme \([7]\). Assume that \(C_{0,f,vl}^{si} > 0\). Let \(\mathcal{M}\) be the message set. Let \(\mathcal{C}\) be a fixed-length feedback code for this message set having the following properties: length \(n\), rate \(\frac{1}{n} \log_2 |\mathcal{M}|\) with \(C_{\downarrow}^{si} - \delta \leq \frac{1}{n} \log_2 |\mathcal{M}| \leq C_{\downarrow}^{si}\), and error probability \(\leq \epsilon\) (we know that such a code exists for any \(\epsilon, \delta > 0\) and for \(n\) large enough). Let \(\mathcal{C}_0\) be a zero-error VLF code with the same capacity at the block length \(n\). Let \(R_0\) be the average length \(\frac{1}{R_0} \log_2 |\mathcal{M}|\) with \(\frac{1}{R_0} (C_{\downarrow}^{si} - \delta)n \leq \frac{1}{R_0} \log_2 |\mathcal{M}| \leq \frac{1}{R_0} C_{\downarrow}^{si} n\). Based on the codes \(\mathcal{C}\) and \(\mathcal{C}_0\) we shall devise another variable-length zero-error coding scheme of rate arbitrarily close to \(C_{\downarrow}^{si}\), which will prove the desired claim. The communication protocol is as follows. To send a message \(m \in \mathcal{M}\), the transmitter first sends the corresponding codeword from \(\mathcal{C}\). Depending on whether or not the receiver has decoded the received sequence correctly, something that the transmitter knows because it can simulate the decoding process after receiving feedback, the transmitter then sends one bit of information through the channel. This bit has the meaning of an ACK/NACK signal that informs the receiver about the correctness of decoding, and can be transmitted \textit{error free} in a finite expected number of channel uses because the zero-error capacity is positive by assumption. Now, if the sent signal is ACK, meaning that the decoding was correct and that both the transmitter and the receiver are aware of that, the protocol stops. If on the other hand the signal was NACK, meaning that the decoding was incorrect, the transmitter sends the same message again, but this time it encodes the message using a zero-error code \(\mathcal{C}_0\), rather than the code \(\mathcal{C}\). This ensures that the receiver will decode the received sequence correctly with probability 1 and the coding scheme just described is therefore zero-error. Moreover, the overall rate of the scheme is approximately equal to the rate of the code \(\mathcal{C}\) used in the first phase of the protocol, because the second phase of the protocol is active only with probability \(\leq \epsilon\) (the probability that the transmission in the first phase fails), and this can be made arbitrarily small. ▲

We conclude this section with a theorem that is meant to demonstrate the power of variable-length coding compared to fixed-length coding in channels with feedback—with fixed-length codes, information obtained by the transmitter through the feedback link is not fully utilized.

**Theorem 4.** There exists an SD-DMC for which \(C_{0,f,fl}^{nc,nc} = 0\) and yet \(C_{0,f,vl}^{r} > 0\).

**Proof:** Consider the following binary-input-binary-output channel with two states: in state \(s_0\), we have the so-called Z-channel with \(W(0|0, s_0) = 1\) and \(W(1|1, s_0) = 1 - p, 0 < p < 1\), and in state \(s_1\) the channel is noiseless, \(W(0|0, s_1) = W(1|1, s_1) = 1\).

Zero-error communication with fixed-length feedback codes through this channel is not possible, even if
both the transmitter and the receiver have non-causal state information. This is because the state sequence may happen to be $s_0 \cdots s_0$ in which case every two input sequences of length $n$ are confusable, meaning that they can produce the same output sequence with positive probability. Hence, $C_{0,f,l}^{nc,nc} = 0$.

However, zero-error communication with variable-length feedback codes is possible even if neither the transmitter nor the receiver have any state information, as one can verify from (11) ($y = 1$ is a disprover for $x = 0$ in both states). In fact, not only is it possible, but the zero-error VLF capacity is by Theorem 5 equal to the vanishing-error capacity of the corresponding channel, $C_{0,f,vl}^{nc} = C_{\gamma}^{nc}$. 

### IV. Zero-Error Capacity: Bounded-Length Codes

In the previous section we have demonstrated how variable-length encoding can significantly increase the zero-error feedback capacity of an SD-DMC. We now investigate the same problem in the situation where one wishes to impose a fixed and deterministic upper bound on the codeword lengths, or equivalently on the stopping time of transmission. Variable-length codes in general have no such bound—even though their average length is finite, each message is mapped to possibly infinitely many codewords of different lengths, which means that the decoding delay can in general be arbitrarily large. It is therefore natural, especially from the practical point of view, to consider the case where the duration of transmission is upper bounded and to investigate the corresponding fundamental limits.

Zero-error feedback capacity of DMCs under bounded-length coding was first studied by Han and Sato [7]. In particular, it was shown in [7] that the condition for positivity of $C_{0,f,bl}$ is the same as those for fixed-length coding; however, the values of the zero-error capacities in the two settings are in general different (see Theorem 6 below).

**Theorem 5.** For every $s_i \in S_l$, $C_{0,f,bl}^{si} > 0$ if and only if $C_{0,f,bl}^{nc} > 0$. In particular:

(a) $C_{0,f,bl}^{nc} > 0$ if and only if

$$\exists x, x' \in \mathcal{X} \quad \forall y \in \mathcal{Y} \quad \left( \forall s \in \mathcal{S} \quad W(y|x, s) = 0 \right) \lor \left( \forall s \in \mathcal{S} \quad W(y|x', s) = 0 \right).$$

(b) $C_{0,f,bl}^{nc} > 0$ if and only if (16) holds.

(c) $C_{0,f,bl}^{nc} > 0$ if and only if there exists a partition $\mathcal{Y}_0, \mathcal{Y}_1$ of $\mathcal{Y}$ such that

$$\forall s \in \mathcal{S} \quad \exists x, x' \in \mathcal{X} \quad W(\mathcal{Y}_0|x, s) = W(\mathcal{Y}_1|x', s) = 1.$$  

(d) $C_{0,f,bl}^{nc} > 0$ if and only if

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

(e) $C_{0,f,bl}^{nc} > 0$ if and only if there exist two input letters that are non-confusable in each state:

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

(f) $C_{0,f,bl}^{nc} > 0$ if and only if in each state there exist two non-confusable input letters:

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

(g) $C_{0,f,bl}^{nc} > 0$ if and only if (20) holds.

(h) $C_{0,f,bl}^{nc} > 0$ if and only if (20) holds.
Proof: We first show that fixed-length and bounded-length zero-error feedback capacities are either both positive or both zero. Since \( C_{0,f,bl}^{\text{si}} \geq C_{0,f,fl}^{\text{si}} \), the “if part” is clear. Conversely, suppose that \( C_{0,f,bl}^{\text{si}} > 0 \). Then, for some \( 0 < k \leq n < \infty \), there exists a zero-error code of cardinality at least 2, average length \( k \), and maximum length \( n \). By “zero-padding” the codewords we can then construct a fixed-length zero-error code of length \( n \) having the same cardinality, which implies that \( C_{0,f,fl}^{\text{si}} \geq \frac{1}{n} > 0 \).

By the observation from the previous paragraph, we can focus on fixed-length codes in the rest of the proof.

(a)–(d) The conditions (16)–(18) for positivity of \( C_{0,f,fl}^{\text{si}} \) in cases when only the transmitter has state information were derived in [2] Thm 3, Thm 10, Rem. 17. Note that (16) is the condition for positivity of the zero-error fixed-length feedback capacity of the DMC \( \tilde{W}(y|x) = \sum_{s \in S} Q(s) W(y|x,s) \); see (15).

(e) The case \( \text{si} = (sc,c) \) is solved by using the standard technique of treating the output \( Y \) and the state \( S \) as a joint output \( (Y,S) \) of the DMC \( \overline{W}(y,s|x) := Q(s) W(y|x,s) \) with noiseless feedback. (Before the \( n \)th channel use the receiver obtains \( Y^{n-1} \) through feedback and \( S^{n-1} \) as side information, which is equivalent to saying that it obtains the previous outputs of \( \overline{W} \), namely \( (Y,S)^{n-1} \), through feedback.) The condition (19) is the condition for positivity of the zero-error capacity of this DMC; see (15). It is therefore easy to see that this condition is sufficient for \( C_{0,f,fl}^{\text{sc,c}} > 0 \).

To show that (19) is also necessary, suppose for the sake of contradiction that every two input letters \( x, x' \) are confusable in some state \( s_{x,x'} \). Then, for every \( n \), every two input sequences \( x_1 \cdots x_n, x'_1 \cdots x'_n \) can produce the same channel output if the state sequence realized in the channel happens to be \( s_{x_1,x'_1} \cdots s_{x_n,x'_n} \), meaning that error-free communication with fixed-length codes is impossible. Stated informally, since the transmitter does not know the current state, no matter which letter \( x \) it chooses to send in the current time slot there is a chance that the state will be one in which \( x \) is confusable with another letter \( x' \). Thus, error-free communication cannot be guaranteed, no matter how long the codewords are.

(f) Consider now the case \( \text{si} = (c,c) \). If for every state \( s \) there exists a pair of non-confusable inputs \( x_s, x'_s \), which is what the condition (20) means, then the transmitter and the receiver can agree beforehand for \( x_s \) to mean 0 and \( x'_s \) to mean 1. In this way, one bit can be transmitted error free in one channel use and so \( C_{0,f,fl}^{c,c} > 0 \).

Conversely, if (20) is not satisfied, then there exists a state \( s \) for which every two inputs are confusable. If this is the case, then for the state sequence \( s^n = s \cdots s \) it is not possible to transmit one bit error-free in any number of channel uses \( n \), and hence \( C_{0,f,fl}^{c,c} = 0 \).

(g), (h) The previous argument holds in these cases too. Namely, knowing the future states cannot help the encoder/decoder if these states remain unfavorable throughout the entire transmission.

As in the (unbounded) variable-length case, whenever the zero-error feedback capacity under bounded-length coding is positive, it equals the vanishing-error capacity of the same channel.

Theorem 6. Let \( \text{si} \in \text{SI} \). If \( C_{0,f,bl}^{\text{si}} > 0 \), then \( C_{0,f,fl}^{\text{si}} = C_{0,f,bl}^{\text{si}} \).

Proof: The proof is analogous to the proof of Theorem 5 for the variable-length case; the only difference is that the zero-error code that is used in the second phase of the protocol is now required to be of bounded length.

V. STATE INFORMATION AT THE DECODER ONLY

In this section we discuss SD-DMCs with state information available only at the decoder, the case that has been left out of the discussion thus far. As pointed out in Remark 1, in this channel model it is not quite clear how to define the code length for variable-length codes, i.e., the stopping time of the transmission (see Definition 1). Since the decoder makes a decision based on the outputs \( Y^n \) and states \( S^n \), the encoder, not knowing the states, is not able to exactly simulate the decoding process and to determine the moment when the decision has been made. It can only provide an estimate of this moment based on the outputs \( Y^n \) which it obtains through feedback. As we shall see, this estimate is good enough when one considers coding with asymptotically vanishing error probability (in fact, it is not even necessary as
the vanishing-error capacity can be achieved with fixed-length codes, with or without feedback). However, in the case of zero-error communication the encoder has to be certain that the decoding was successful before it stops transmitting a given message and starts transmitting the next message. It is in this case that the effects of the mismatch in state information at the two sides are most apparent.

A. Vanishing-Error Capacity

We know from [5, Ch. 7.4] and (7) that:

\[ C_{V\ell} = \max_{P_X} I(X; Y|S) = C_{V\ell}^{sc,c}. \]  

Note that the issue with the stopping time mentioned above does not arise when defining the two capacities in (21): for \( C_{V\ell} \) because in the fixed-length setting the stopping time is fixed in advance, and for \( C_{V\ell}^{sc,c} \) because in this case both sides have state information.

The quantity we are interested in here, \( C_{V\ell}^{sc,c} \), has not been formally defined. However, it is clear that, for any reasonable definition, the following chain of inequalities must hold: \( C_{V\ell}^{c,c} \leq C_{V\ell}^{c,c} \leq C_{V\ell}^{c,c} \leq C_{V\ell}^{sc,c} \). From this and (21) we then conclude that

\[ C_{V\ell}^{c,c} = C_{V\ell}^{c,c} = \max_{P_X} I(X; Y|S) = C_{V\ell}^{sc,c}. \]  

In particular, the VLF capacity of the \((-c)\) channel can be achieved by using fixed-length codes.

B. Zero-Error Capacity: Bounded-Length Codes

We now turn to the zero-error problems and start with the bounded-length case. The following theorem gives a necessary and sufficient condition for positivity of the zero-error capacity in this setting.

**Theorem 7.** The following statements are equivalent: (a) \( C_{0,f,fl}^{c,c} > 0 \); (b) \( C_{0,f,bl}^{c,c} > 0 \); (c) \( C_{0,f,bl}^{sc,c} > 0 \); (d) \( \text{(19)} \) holds.

**Proof:** (a) \( \Leftrightarrow \) (b): This can be shown by “zero-padding” bounded-length codes to obtain fixed-length codes, as for the other models of state information availability (see Theorem [5]).

(b) \( \Rightarrow \) (c): This follows from \( C_{0,f,bl}^{sc,c} \geq C_{0,f,bl}^{c,c} \).

(c) \( \Leftrightarrow \) (d): This was shown in Theorem [5].

(d) \( \Rightarrow \) (a): Suppose that (d) holds, i.e., there exist two input letters \( x, x' \) that are non-confusable in every state. Then, if the transmitter sends \( x \) for 0 and \( x' \) for 1, the receiver will be able to tell from the output which of the two possibilities is the correct one because it knows the channel state. Therefore, one bit can be transmitted in one channel use, and hence \( C_{0,f,fl}^{c,c} \geq 1 > 0 \). \( \square \)

We now know from Theorem [7] that \( C_{0,f,fl}^{c,c} > 0 \Leftrightarrow C_{0,f,bl}^{sc,c} > 0 \), from Theorem [6] that \( C_{0,f,bl}^{sc,c} = C_{0,f,bl}^{c,c} \) whenever \( C_{0,f,bl}^{c,c} > 0 \), and from (22) that \( C_{V\ell}^{sc,c} = C_{V\ell}^{c,c} \). It is then natural to ask if it is also true that \( C_{0,f,bl}^{c,c} = C_{V\ell}^{c,c} \) whenever \( C_{0,f,bl}^{c,c} > 0 \)? The corresponding equality for the other models of state information availability \( s_i \in SI \) has been established in Theorem [6] but the proof technique used there does not apply when \( s_i = (-c) \). The difficulty is precisely due to the fact we mentioned at the beginning of this section—in this model the transmitter cannot exactly simulate the decoding process because it does not know the states. Hence, the transmitter is in general not able to decide with certainty whether the receiver has decoded the received sequence correctly, and whether it should retransmit the same codeword or start sending the next codeword. Therefore, for \( s_i = (-c) \) it is not clear whether the vanishing-error capacity can always be achieved with zero-error bounded-length codes. We next give an example of a channel for which the answer to this question is affirmative.

**Example 1.** Consider the following binary-input-binary-output channel with two states: in state \( s_0 \) the channel flips the input bit, \( W(1|0, s_0) = W(0|1, s_0) = 1 \), and in state \( s_1 \) it leaves the bit intact, \( W(0|0, s_1) = W(1|1, s_1) = 1 \).
Suppose that the transmitter sends $x$ in the $n$'th slot and that $y$ is produced at the channel output. After the transmitter obtains $y$ through the feedback link, it can easily determine the $n$'th state: the state is $s_1$ if and only if $x = y$. This means that the transmitter effectively obtains (strictly causal) state information through feedback and therefore $C^c_{\text{0f,bl}} = C^\text{sc,c}_{\text{0f,bl}} = C^\text{sc,c}_\circ$, where the second equality holds because $C^\text{sc,c}_{\text{0f,bl}} > 0$ (see (19)).

The main point in Example 1 is the following: if the states are uniquely determined by the channel inputs and outputs, then the problem of calculating $C^c_{\text{0f,bl}}$ is reduced to the (easier) problem of calculating $C^\text{sc,c}_{\text{0f,bl}}$, which was solved in Theorems 5(e) and 6. Whether such a reduction is possible in general is an interesting question that we leave for future work.

We end this subsection with a remark that the problem of determining the value of $C^c_{\text{0f,bl}}$ is related to the problem of determining the zero-error capacity under bounded-length coding of DMC’s with noisy feedback. (The latter problem is also unsolved except in some special cases, see [9], [11].) Namely, as already noted in the proof of Theorem 5(e), the standard trick of dealing with state information at the decoder is to think of it as part of a joint channel output $(Y, S)$ of the DMC $\overline{W}(y, s|x) := Q(s)W(y|x, s)$. Noiseless feedback in the DMC $\overline{W}$ would mean that the transmitter is given $(Y, S)^{n-1}$ before the $n$'th channel use. This implies that an SD-DMC $W$ with $s_i = (\text{sc}, c)$ and noiseless feedback is equivalent to the DMC $\overline{W}$ with noiseless feedback. However, in the case $s_i = (-, c)$ this equivalence fails as the transmitter now obtains only a “degraded” version $(Y^{n-1})$ of the joint output $((Y, S)^{n-1})$ through the feedback link.

C. Zero-Error Capacity: Variable-Length Codes

For variable-length codes, we can only give a sufficient condition for positivity of the zero-error capacity at this point. The idea behind this condition is based on the fact that, for some channels, the transmitter obtains state information “for free” through the feedback link, in which case the $(-, c)$ model is effectively reduced to the $(\text{sc}, c)$ model (see Example 1). As we show in Theorem 8, it is in fact not necessary that all states be uniquely determined by inputs and outputs in order for this reduction to work—it is enough that only a group of states exists that contains a “disprover” and that is discernible from other states with positive probability.

**Theorem 8.** A sufficient condition for $C^c_{\text{0f,vl}} > 0$ is the following:

$$\exists x, x' \in X, y \in Y, S^* \subseteq S, S^* \neq \emptyset$$

$$\left( \forall s \in S^* \right) W(y|x', s) > 0 \land W(y|x, s) = 0 \land \left( \forall s \in S \setminus S^* \right) W(y|x', s) = 0. \tag{23}$$

**Proof:** Let us first parse the condition (23). The meaning of the statement $W(y|x', s) > 0 \land W(y|x, s) = 0$ is the same as before: $y$ is a disprover for $x$ (in the group of states $S^*$). Further, we require the existence of an event $(x' \rightarrow y)$ that has positive probability in the group of states $S^*$, but is impossible in other states. The occurrence of this event can serve to the transmitter as an identifier of the group of states $S^*$.

Now, assuming that (23) holds, one bit of information can be sent with zero error as follows. In the first two channel uses the transmitter sends $x, x'$ for 0 and $x', x$ for 1. If the letters obtained at the output are $\neg y, y$ and the state in the second slot is from $S^*$, then the receiver concludes that 0 must have been sent. The reason is the following: since the received symbol in the second slot is $y$ and the state is from $S^*$ (the receiver can see the states), the transmitted symbol must have been $x'$, and then it automatically follows that the symbol sent in the first slot is $x$. Furthermore, the transmitter is also assured that the state in the second slot is from $S^*$ and that the receiver has received the bit correctly because the transition $x' \rightarrow y$ is only possible in states from $S^*$. Similarly, if the letters obtained at the output are $y, \neg y$ and the state in the first slot is from $S^*$, then the receiver concludes that 1 must have been sent, and the transmitter is assured that the receiver has received the bit correctly. In summary, if the channel output is either $\neg y, y$ or $y, \neg y$ and the state in the slot in which the output is $y$ is from $S^*$, then the protocol stops and both parties agree on the value of the transmitted bit. If any other situation occurs, the procedure is repeated.
The expected number of steps needed to complete the transmission of the bit is finite because the event that both parties are waiting for \((x' \text{ produces } y \text{ in a state from } S^\ast)\) has positive probability. Therefore, \(C_{0,f,\text{VL}}^{r,c} > 0\).

Note that the condition (10) for \(C_{0,f,\text{VL}}^{r,c} > 0\), together with (2), implies the condition (23). To see that it does, suppose that (10) holds and let \(S^\ast = \{s : W(y|x', s) > 0\}\). This is kind of a sanity check because clearly \(C_{0,f,\text{VL}}^{r,c} > 0\) implies \(C_{0,f,\text{VL}}^{r,c} > 0\). The reverse implication does not hold because in (10) we require that \(W(y|x, s) = 0\) in all states, rather than just in states from \(S^\ast\) as in (23).

Note also that the sufficient condition for \(C_{0,f,\text{VL}}^{r,c} > 0\) given in (23) is different from the necessary and sufficient condition for \(C_{0,f,\text{VL}}^{s,c} > 0\) given in (13). Based on Theorem 7, one might wonder whether it holds that \(C_{0,f,\text{VL}}^{r,c} > 0 \Leftrightarrow C_{0,f,\text{VL}}^{s,c} > 0\), i.e., whether (12) is also necessary and sufficient for \(C_{0,f,\text{VL}}^{r,c} > 0\)? This is, however, not the case. In the proof of the following theorem we give an example of a channel for which \(C_{0,f,\text{VL}}^{r,c} = 0\) and yet \(C_{0,f,\text{VL}}^{s,c} > 0\). In other words, strictly causal state information at the transmitter can in some cases enable the parties to communicate error free, even if they were not able to do that in the absence of this information. This is a somewhat curious fact having in mind that in most other settings studied so far, strictly causal state information at the transmitter has been shown equivalent (in the sense of achievable rates) to no state information: \(C_{0,f,\text{VL}}^{r,c} = C_{0,f,\text{VL}}^{s,c} = C_{0,f,\text{BL}}^{s,c}\) (see also Theorem 7).

**Theorem 9.** There exists an SD-DMC for which \(C_{0,f,\text{VL}}^{r,c} = 0\) and \(C_{0,f,\text{VL}}^{s,c} > 0\).

**Proof:** Consider the following binary-input-binary-output channel with two states: in state \(s_0\) the channel is a BSC \((p)\), where \(0 < p < 1\), i.e., \(W(0|0, s_0) = W(1|1, s_0) = 1 - p\), and in state \(s_1\) the channel is noiseless, \(W(0|0, s_1) = W(1|1, s_1) = 1\).

By Theorem 2(e) we know that \(C_{0,f,\text{VL}}^{s,c} > 0\). It is easy to see why this is the case—the transmitter and the receiver can agree on the codeword termination by using the noiseless state since they both know the states (the transmitter obtains the state information with a one-slot delay, but this can be circumvented by sending a dummy letter; see the proof of Theorem 2(e)).

However, if the transmitter has no state information, i.e., if \(s_i = (-, c)\), then it is not possible to communicate with zero-error in a finite expected number of channel uses. To see this, suppose that the transmitter is trying to send one bit through the channel by using a repetition code. The receiver can see the channel states and will therefore recover the bit correctly as soon as the state happens to be \(s_1\), and it will be in a finite expected number of channel uses. However, the transmitter can never be sure whether this state has occurred and whether it should stop transmitting. This is because all the channel transitions that are possible in state \(s_1\) are also possible in state \(s_0\) and therefore, based on the channel inputs (which it knows) and the channel outputs (which it sees through feedback) alone, the transmitter cannot determine with certainty whether any of the states so far was actually \(s_1\). In other words, if the transmitter wants to be sure that the bit was received correctly, it must continue transmitting indefinitely. (Note that the transmitter will occasionally recognize state \(s_0\)—if the channel input is 0 and the output is 1, or vice versa—but this state is useless for zero-error communication.) It should be clear that the above conclusion holds for any coding scheme, not just for repetition codes, and hence \(C_{0,f,\text{VL}}^{r,c} = 0\).

VI. **CONCLUDING REMARKS AND FURTHER WORK**

As we have seen, with fixed-length coding the information obtained by the transmitter through the feedback link, as well as the side information about the channel states, are not fully utilized. For instance, by allowing the possibility of variable and adaptive transmission times it is possible, under certain conditions, to achieve channel capacity with error probability being fixed to zero. Furthermore, for any fixed error probability \(\epsilon > 0\), one can achieve higher rates with variable-length coding compared to those achievable with fixed-length coding (see the discussion in the Appendix). For these reasons, variable-length coding schemes are a natural choice in systems with feedback (whenever one is willing
to tolerate random decoding delays), and it is therefore important to study the corresponding fundamental limits of communication.

To conclude the paper, we state several problems, related to those we have analyzed here, as pointers for further work:

- Derive necessary and sufficient conditions for positivity of the zero-error VLF capacity of SD-DMCs with state information available only at the decoder, \( C^{V_L}_{0, f, VL} \) (see Section V-C).
- Determine the values of the zero-error capacities \( C^{V_L}_{0, f, BL} \) and \( C^{V_L}_{0, f, VL} \) whenever they are positive (see Section V-B and V-C).
- Investigate the corresponding questions about the zero-error VLF capacity when feedback is noisy, or incomplete, or when the receiver can also send coded information over the feedback link (i.e., a function of the received sequence, rather than the sequence itself). See [9], [1] for some results in this direction for DMCs.
- Investigate the corresponding questions in more general settings—multi-user state-dependent channels, channels with non-i.i.d. states, etc.
- Derive necessary and sufficient conditions for positivity of the zero-error capacity of SD-DMCs without feedback. (This was solved in [2] for \( s = (sc, -) \) and \( s = (c, -) \), but the case \( s = (nc, -) \) was left open, see [2] Thm 7.)

ACKNOWLEDGMENT

The authors would like to thank Lan Vinh Truong (National University of Singapore) on insightful discussions regarding VLF codes.

REFERENCES

[1] M. Asadi and N. Devroye, “On the Zero-Error Capacity of Channels with Noisy Feedback,” in Proc. 55th Annual Allerton Conference, pp. 642–649, UIUC, Illinois, USA, Oct. 2017.
[2] A. Bracher and A. Lapidoth, “The Zero-Error Feedback Capacity of State-Dependent Channels,” IEEE Trans. Inf. Theory, vol. 64, no. 5, pp. 3538–3578, May 2018.
[3] M. V. Burnashev, “Data Transmission Over a Discrete Channel with Feedback. Random Transmission Time,” Probl. Inf. Transm., vol. 12, no. 4, pp. 10–30, 1976.
[4] G. Como, S. Yüksel, and S. Tatikonda, “The Error Exponent of Variable-Length Codes Over Markov Channels With Feedback,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2139–2160, May 2009.
[5] A. El Gamal and Y.-H. Kim, Network Information Theory, Cambridge University Press, 2011.
[6] S. I. Gel’fand and M. S. Pinsker, “Coding for Channels with Random Parameters,” Probl. Contr. Inf. Theory, vol. 9, no. 1, pp. 19–31, 1980.
[7] T. S. Han and H. Sato, “Feedback Codes with Uniformly Bounded Codeword Lengths and Zero-Error Capacities,” IEEE Trans. Inf. Theory, vol. 37, no. 3, pp. 655–660, May 1991.
[8] J. Körner and A. Orlitsky, “Zero-Error Information Theory,” IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2207–2229, Oct. 1998.
[9] J. Massey, “Zero Error,” invited talk at the 2007 Winter School of Information Theory, La Colle sur Loup, France, Mar. 2007.
[10] N. Merhav and T. Weissman, “Coding for the Feedback Gel’fand–Pinsker Channel and the Feedforward Wyner–Ziv Source,” IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 4207–4211, Sep. 2006.
[11] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Feedback in the Non-Asymptotic Regime,” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.
[12] C. E. Shannon, “The Zero Error Capacity of a Noisy Channel,” IRE Trans. Inf. Theory, vol. 2, no. 3, pp. 8–19, Sep. 1956.
[13] C. E. Shannon, “Channels with Side Information at the Transmitter,” IBM J. Res. Dev., vol. 2, no. 4, pp. 289–293, Oct. 1958.
[14] L. G. Tallini, S. Al-Bassam, and B. Bose, “Feedback Codes Achieving the Capacity of the Z-Channel,” IEEE Trans. Inf. Theory, vol. 54, no. 3, pp. 1357–1362, Mar. 2008.
[15] C. Wang, M. Kovačević, and V. Y. F. Tan, “Error-Free Communication Over State-Dependent Channels with Variable-Length Feedback,” in Proc. 2018 IEEE Intl. Symp. Inf. Theory (ISIT), pp. 286–290, Vail, CO, USA, Jun. 2018.
[16] L. Zhao and H. H. Permuter, “Zero-Error Feedback Capacity of Channels with State Information via Dynamic Programming,” IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 2640–2650, Jun. 2010.

APPENDIX

PROOF OF THEOREM 1

The theorem states that, for every \( s \in SI \)

\[ C^{VI}_{i, f, VL} = C^{VI}_{i, f, BL} = C^{VI}_{i, f, FL} = C^{VI}_{i, f, FL}. \]

We prove this in two steps: we first argue that \( C^{VI}_{i, f, BL} = C^{VI}_{i, f, VL} \), and then we show in Lemma 10 that \( (1 - \epsilon)C^{VI}_{i, f, BL} \leq C^{VI}_{i, f, FL}, \forall \epsilon > 0 \). This, together with the obvious fact that \( C^{VI}_{i, f, BL} \geq C^{VI}_{i, f, VL} \geq C^{VI}_{i, f, FL} \), will conclude the proof.
To see that $C_{i,f,bl}^{si} = C_{i,f,VL}^{si}$, consider an arbitrary $(\ell, |\mathcal{M}|, \epsilon)$ variable-length code $C_{VL}$ whose codeword lengths are described by the stopping time $\tau$ satisfying $E[\tau] \leq \ell$. One can then define a bounded-length code $C_{bl}$ which uses the same encoding and decoding procedures as $C_{VL}$, except that the decoder is forced to make a decision at time $b$, if it hasn’t already done so. The error probability of the resulting code $C_{bl}$ is at most $\epsilon + P[\tau > b] \leq \epsilon + \frac{E[\tau]}{b} \leq \epsilon + \frac{\ell}{b}$, where we have used Markov’s inequality. Since $b$ can be taken arbitrarily large, this shows that there exists a bounded-length code whose rate and error probability are arbitrarily close to the rate and error probability of a given variable-length code.

We next prove that $(1 - \epsilon)C_{i,f,bl}^{si} \leq C_{i,v,PL}^{si}$ for any $\epsilon > 0$. For concreteness, we consider the case $\epsilon = (nc,-)$ (the Gel’fand–Pinsker channel [6]); the other cases can be obtained in a similar way. The proof combines the approach from [11] for DMCs with feedback, the derivation of the capacity of the Gel’fand–Pinsker channel without feedback [5 Sec. 7.6], and a certain inequality that is derived here and that is needed as a replacement for the so-called Csiszár sum identity.

**Remark 4.** The proof of Lemma 10 in fact implies a slightly stronger statement that the $\epsilon$-error capacity of an SD-DMC under variable-length coding is upper-bounded by $\frac{1}{1-\epsilon} C_{i,v,PL}^{si}$. For some models of state information availability $\epsilon \in SI$ it is not difficult to prove that this value is also achievable, implying that the $\epsilon$-error capacity of the corresponding SD-DMC equals $\frac{1}{1-\epsilon} C_{i,v,PL}^{si}$ (as is the case for DMCs [11]). However, since we have not established this for all $\epsilon \in SI$, and since this is not the focus of the present paper, we do not give here proofs of achievability.

**Lemma 10.** For every $\epsilon \in (0, 1)$, $(1 - \epsilon)C_{i,f,bl}^{si} \leq C_{i,v,PL}^{si}$.

**Proof:** Let $\epsilon = (nc,-)$ and consider a particular bounded-length code $(f_n, g_n, \tau)$ for this channel; see Definition [11] Following [11] Thm 4, we define an extended channel:

$$\tilde{X} := \mathcal{X} \cup \{T\}$$

$$\tilde{Y} := \mathcal{Y} \cup \{T\}$$

$$\tilde{S} := S$$

$$\tilde{W}(\tilde{y}|\tilde{x}, \tilde{s}) := \begin{cases} W(y|x, s), & \tilde{x} \neq T \\ \mathbb{1}\{y = T\}, & \tilde{x} = T, \end{cases}$$

and the corresponding code $(\tilde{f}_n, \tilde{g}_n, \tilde{\tau})$:

$$\tilde{f}_n(M, \hat{Y}^{n-1}, S^\infty) := \begin{cases} f_n(M, \hat{Y}^{n-1}, S^\infty), & \tilde{\tau} \geq n \\ T, & \tilde{\tau} < n, \end{cases}$$

$$\tilde{\tau} := \inf \left\{ n : \hat{Y}_n = T \right\} = \tau + 1$$

$$\tilde{g}_n(\hat{Y}^n) := \begin{cases} g_n(\hat{Y}^n), & \tilde{\tau} > n \\ g_n(\hat{Y}_{\tilde{\tau}-1}), & \tilde{\tau} \leq n. \end{cases}$$

Here $T \notin \mathcal{X} \cup \mathcal{Y}$ is the “termination” symbol which is transmitted noiselessly and serves for the transmitter to inform the receiver that the transmission is over. Apart from this addition, the two channels behave the same. If $(f_n, g_n, \tau)$ is an $(\ell - 1, |\mathcal{M}|, \epsilon)$-bounded-length code for the original channel satisfying $\tau \leq b - 1$, then $(\tilde{f}_n, \tilde{g}_n, \tilde{\tau})$ is an $(\ell, |\mathcal{M}|, \epsilon)$ code for the extended channel with $\tilde{\tau} \leq b$; thus, upper bounding the rate of the former is essentially equivalent to upper bounding the rate of the latter.

To derive the desired upper bound, recall that the Fano inequality asserts that for any code $(\tilde{f}_n, \tilde{g}_n, \tilde{\tau})$ for the message set $\mathcal{M}$ it holds that:

$$(1 - \epsilon) \log |\mathcal{M}| \leq I(M; \hat{Y}^b) + h(\epsilon).$$

The $\epsilon$-error capacity of a channel is the largest rate that is asymptotically achievable with error probability not exceeding $\epsilon$. 

\[\text{Proof:}\]

\[\text{Defin} \]

\[\text{Thm 4}, \text{we define an extended channel:}\]

\[\tilde{X} := \mathcal{X} \cup \{T\}\]

\[\tilde{Y} := \mathcal{Y} \cup \{T\}\]

\[\tilde{S} := S\]

\[\tilde{W}(\tilde{y}|\tilde{x}, \tilde{s}) := \begin{cases} W(y|x, s), & \tilde{x} \neq T \\ \mathbb{1}\{y = T\}, & \tilde{x} = T, \end{cases}\]

\[\tilde{f}_n(M, \hat{Y}^{n-1}, S^\infty) := \begin{cases} f_n(M, \hat{Y}^{n-1}, S^\infty), & \tilde{\tau} \geq n \\ T, & \tilde{\tau} < n, \end{cases}\]

\[\tilde{\tau} := \inf \left\{ n : \hat{Y}_n = T \right\} = \tau + 1\]

\[\tilde{g}_n(\hat{Y}^n) := \begin{cases} g_n(\hat{Y}^n), & \tilde{\tau} > n \\ g_n(\hat{Y}_{\tilde{\tau}-1}), & \tilde{\tau} \leq n. \end{cases}\]
In the following we show that the mutual information term in (31) can be upper bounded as $I(M; \hat{Y}^b) \leq \ell \cdot C_{\vartriangleleft,\vartriangleright_{\mathrm{PL}}}$ + $o(\ell)$. Plugging this back into (31) would yield $\frac{1}{\tau} \log |\mathcal{M}| \leq \frac{1}{\tau} C_{\vartriangleleft,\vartriangleright_{\mathrm{PL}}} + o(1)$ and would thus complete the proof of the lemma.

The following inequality was derived in the proof of [11, Thm 4]:

$$I(M; \hat{Y}^b) \leq H(\tau) + \sum_{k=1}^{b} I(M; \hat{Y}_k|V_k, \hat{Y}_k^{k-1}),$$

where (33) is shown below; (37) holds because $C_k < b$ values of

$$V_n := \mathbb{1}\{\hat{\tau} \leq n\}$. The derivation holds unchanged in our case too so we shall not repeat it here. It was also shown there that $H(\tau) \leq (\ell + 1)h\left(\frac{1}{\tau + 1}\right) = o(\ell)$, so we only need to upper bound the second summand on the right-hand side of (32). First notice that

$$\sum_{k=1}^{b} I(M; \hat{Y}_k|V_k, \hat{Y}_k^{k-1}) \leq \sum_{k=1}^{b} I(M, \hat{Y}_k^{k-1}; \hat{Y}_k|V_k)$$

$$= \sum_{k=1}^{b} \mathbb{P}[V_k = 0] I(M, Y^{k-1}; Y_k),$$

where (33) is by the chain rule for mutual information and (34) is obtained by conditioning on the possible values of $V_k$. Namely, 1.) conditioned on $V_k = 1$ we have $\hat{Y}_k = T$ and hence the corresponding mutual information term is zero, and 2.) conditioned on $V_k = 0$, the statistics of the extended channel is identical to that of the original channel and hence the mutual information term can be computed for the latter. Therefore, $I(M, \hat{Y}_k^{k-1}; \hat{Y}_k|V_k) = \mathbb{P}[V_k = 1] \cdot 0 + \mathbb{P}[V_k = 0] \cdot I(M, Y^{k-1}; Y_k)$. We further have:

$$\sum_{k=1}^{b} \mathbb{P}[V_k = 0] I(M, Y^{k-1}; Y_k) = \sum_{k=1}^{b} \mathbb{P}[V_k = 0] \left( I(M, Y^{k-1}, S_{k+1}^b; Y_k) - I(Y_k; S_{k+1}^b|M, Y^{k-1}) \right)$$

$$\leq \sum_{k=1}^{b} \mathbb{P}[V_k = 0] \left( I(M, Y^{k-1}, S_{k+1}^b; Y_k) - I(Y^{k-1}; S_k|M, S_{k+1}^b) \right)$$

$$= \sum_{k=1}^{b} \mathbb{P}[V_k = 0] \left( I(M, Y^{k-1}, S_{k+1}^b; Y_k) - I(M, Y^{k-1}, S_{k+1}^b; S_k) \right)$$

$$= \sum_{k=1}^{b} \mathbb{P}[V_k = 0] \left( I(U_k; Y_k) - I(U_k; S_k) \right)$$

$$\leq \sum_{k=1}^{b} \mathbb{P}[V_k = 0] C_{\vartriangleleft,\vartriangleright_{\mathrm{PL}}}$$

where (36) is shown below; (37) holds because $(M, S_{k+1}^b)$ is independent from $S_k$; in (38) we have denoted $U_k = (M, Y^{k-1}, S_{k+1}^b)$; (39) follows from the expression for the capacity of the Gel’fand-Pinsker channel, $C_{\vartriangleleft,\vartriangleright_{\mathrm{PL}}} = \sup \left( I(U; Y) - I(U; S) \right)$, see [5, sec. 7.6.2]; and (40) follows from the fact that $\sum_{k=1}^{b} \mathbb{P}[V_k = 0] = \sum_{k=1}^{b} \mathbb{P}[\tau \geq k] = \mathbb{E}[\tau] \leq \ell$.

It is left to justify (36). For fixed-length codes (for which $\hat{\tau} = \ell = b$ and hence $\mathbb{P}[V_k = 0] = 1$ for all $k < b$), the so-called Csiszár sum identity [5, p. 25] can be used in this step to establish equality in (36).

In our notation this identity has the form: $ \sum_{k=1}^{b} I(Y^{k-1}; S_k|M, S_{k+1}^b) = \sum_{k=1}^{b} I(Y_k; S_{k+1}^b|M, Y^{k-1})$. It does not apply in our case due to the factors $\mathbb{P}[V_k = 0]$ appearing in the sums (35) and (36). However, one can establish the inequality in (36) by using the following monotonicity property of the coefficients $\mathbb{P}[V_k = 0]$, which is evident from the definition of $V_k$:

$$j < k \quad \Rightarrow \quad \mathbb{P}[V_j = 0] \geq \mathbb{P}[V_k = 0].$$

(41)
We have:
\[
\sum_{k=1}^{b} \mathbb{P}[V_k = 0] I(Y^{k-1}; S_k|M, S_{k+1}^b) = \sum_{k=1}^{b} \mathbb{P}[V_k = 0] \sum_{j=1}^{k-1} I(Y_j; S_k|M, Y^{j-1}, S_{k+1}^b) \quad (42)
\]
\[
= \sum_{k=1}^{b} \sum_{j=1}^{k-1} \mathbb{P}[V_k = 0] I(Y_j; S_k|M, Y^{j-1}, S_{k+1}^b) \quad (43)
\]
\[
\leq \sum_{k=1}^{b} \sum_{j=1}^{k-1} \mathbb{P}[V_j = 0] I(Y_j; S_k|M, Y^{j-1}, S_{k+1}^b) \quad (44)
\]
\[
= \sum_{j=1}^{b} \sum_{k=j}^{b} \mathbb{P}[V_j = 0] I(Y_j; S_k|M, Y^{j-1}, S_{k+1}^b) \quad (45)
\]
\[
= \sum_{j=1}^{b} \mathbb{P}[V_j = 0] \sum_{k=j+1}^{b} I(Y_j; S_k|M, Y^{j-1}, S_{k+1}^b) \quad (46)
\]
\[
= \sum_{j=1}^{b} \mathbb{P}[V_j = 0] I(Y_j; S_{j+1}^b|M, Y^{j-1}) \quad (47)
\]
where (42) and (47) are obtained from the chain rule for mutual information; (44) follows from (41); and it is understood that \( S_{b+1}^b = Y_0 = \emptyset \), as usual. This implies (36) and concludes the proof. \(\blacksquare\)