Reliability function for streaming over a DMC with feedback

Nian Guo, *Member, IEEE,* Victoria Kostina, *Senior Member, IEEE*

***Abstract*—Conventionally, posterior matching is investigated in channel coding and block encoding contexts – the source symbols are equiprobably distributed and are entirely known by the encoder before the transmission. In this paper, we consider a streaming source, whose symbols progressively arrive at the encoder at a sequence of deterministic times. We derive the joint source-channel coding (JSCC) reliability function for streaming over a discrete memoryless channel (DMC) with feedback. We propose a novel *instantaneous encoding phase* that operates during the symbol arriving period and achieves the JSCC reliability function for streaming when followed by a block encoding scheme that achieves the JSCC reliability function for a classical source whose symbols are fully accessible before the transmission. During the instantaneous encoding phase, the evolving message alphabet is partitioned into groups whose priors are close to the capacity-achieving distribution, and the encoder determines the group index of the actual sequence of symbols arrived so far and applies randomization to exactly match the distribution of the transmitted index to the capacity-achieving one. Surprisingly, the JSCC reliability function for streaming is equal to that for a fully accessible source, implying that the knowledge of the entire symbol sequence before the transmission offers no advantage in terms of the reliability function. For streaming over a symmetric binary-input DMC, we propose a one-phase *instantaneous small-enough difference (SED) code* that not only achieves the JSCC reliability function, but also, thanks to its single-phase time-invariant coding rule, can be used to stabilize an unstable linear system over a noisy channel. For equiprobably distributed source symbols, we design low complexity algorithms to implement both the instantaneous encoding phase and the instantaneous SED code. The algorithms group the source sequences into sets we call types, which enable the encoder and the decoder to track the priors and the posteriors of source sequences jointly, leading to a log-linear complexity in time. While the reliability function is derived for non-degenerate DMCs, i.e., DMCs whose transition probability matrix has all positive entries, for degenerate DMCs, we design a code with instantaneous encoding that achieves zero error for all rates below Shannon's joint source-channel coding limit.**

***Index Terms*—Channels with feedback, reliability function, joint source-channel coding, variable-length codes, streaming, causal coding, posterior matching, anytime codes, control over noisy channels.**

I. Introduction

This paper considers joint source-channel coding of streaming data over a DMC with full feedback using variable-length feedback codes. With the emergence of the Internet of Things, communication systems, such as those employed in distributed control and tracking scenarios, are becoming increasingly dynamic, interactive, and delay-sensitive. The source symbols in such real-time systems arrive at the encoder in a streaming fashion. For example, the height and the speed data of an unmanned aerial vehicle stream into the encoder in real time. An intriguing question is: What codes can transmit streaming data with both high reliability and low latency over a channel with feedback? Classical posterior matching schemes [1]–[10] can reliably transmit messages over a channel with feedback but under the assumption that the source sequence is fully accessible to the encoder before the transmission. One can simply buffer the arriving data into a block and then transmit the data block using a classical posterior matching scheme. Intuitively, the buffer-then-transmit code is a good choice if the buffering time is negligibly short, i.e., if data packets arrive at the encoder at an extremely fast rate. However, if data packets arrive at the encoder steadily rather than in a burst, the buffer-then-transmit code becomes ill-suited due to the delay introduced by collecting data into a block before the transmission [11]. The encoder in this paper performs *instantaneous* encoding: it starts transmitting as soon as the first message symbol arrives and incorporates new message symbols into the continuing transmission on the fly. Like classical posterior matching schemes, it relies on full channel feedback.

Designing good channel block encoding schemes with feedback is a classical problem in information theory [1]–[10], since feedback, though unable to increase the capacity of a memoryless channel [12], can simplify the design of capacity-achieving codes [1]–[4] and improve achievable delay-reliability tradeoffs [5][13]. The underlying principle behind capacity-achieving block encoding schemes with feedback [1]–[10], termed posterior matching [4], is to transmit a channel input that has two features. First, the channel input is independent of the past channel outputs, representing the new information in the message that the decoder has not yet observed. Second, the probability distribution of the channel input is matched to the capacity-achieving one using the posterior of the message.

While asymptotically achieving the channel capacity ensures the best possible transmission rates in the limit of large delay, optimizing the tradeoff between delay and reliability is critical for time-sensitive applications. The delay-reliability tradeoff is often measured by the reliability function (a.k.a. optimal error exponent), which is defined as the maximum rate of the exponential decay of the error probability at a rate strictly below the channel capacity as the blocklength is taken to infinity. It is a classical fundamental limit that helps to gain insight into the finite blocklength performance of codes via large deviations theorems in probability. In the context

N. Guo and V. Kostina are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, 91125 USA. E-mail: {nguo,vkostina}@caltech.edu. This work was supported in part by the National Science Foundation (NSF) under grants CCF-1751356 and CCF-1956386.

of channel coding, the reliability function of a DMC with feedback was first shown by Burnashev [5]. Variable-length channel codes with block encoding that achieve Burnashev's reliability function are proposed in [5]–[8], [10]. Burnashev's [5] and Yamamoto and Itoh (Y-I)'s schemes [6] are structurally similar in that they both have two phases. In the communication phase, the encoder matches the distribution of its output to the capacity-achieving input distribution, while aiming to increase the decoder's belief about the true message. In the confirmation phase, the encoder repeatedly transmits one of two symbols indicating whether or not the decoder's estimate at the end of the communication phase is correct. Caire et al. [7] showed that the code transmitted in the communication phase of the Y-I scheme can be replaced by any non-feedback block channel code, provided that the error probability of the block code is less than a constant determined by the code rate as the blocklength goes to infinity. Naghshvar et al. [8] challenged the convention of using a two-phase code [5]–[7] to achieve Burnashev's reliability function by proposing the MaxEJS code, which searches for the deterministic encoding function that maximizes an extrinsic Jensen-Shannon (EJS) divergence at each time. Since the MaxEJS code has a double exponential complexity in the length of the message sequence k, for symmetric binary-input DMCs, Naghshvar et al. [8] proposed a simplified encoding function that is referred to as the *small-enough difference* (SED) rule in [9]. The SED encoder partitions the message alphabet into two groups such that the difference between group posteriors and the Bernoulli$\left(\frac{1}{2}\right)$ capacity-achieving distribution is small. While the SED rule still has an exponential complexity in the length of the message, Antonini et al. [9] designed a systematic variable-length code for transmitting k bits over a binary symmetric channel (BSC) with feedback that has complexity $O(k^2)$. The complexity reduction is realized by grouping messages with the same posterior. Yang et al. [10] generalized Naghshvar et al.'s SED rule-based code [8] to binary-input binary-output asymmetric channels.

While the message in [5]–[10] is equiprobably distributed on its alphabet, the JSCC reliability function for transmitting a non-equiprobable discrete-memoryless source (DMS) over a DMC has also been studied [14]–[17]. For fixed-length almost lossless coding without feedback, Gallager [14] derived an achievability bound on the JSCC reliability function, which indicates that JSCC leads to a strictly larger error exponent than separate source and channel coding in some cases; Csiszàr [15] provided achievability and converse bounds on the JSCC reliability function using random coding and type counting; Zhong et al. [16] showed that Csiszàr's achievability bound [15] is tighter than Gallager's bound [14] and provided sufficient conditions for the JSCC reliability function to be strictly larger than the separate source and channel coding reliability function. For variable-length lossy coding with feedback, Truong and Tan [17] derived the JSCC excess-distortion reliability function under the assumption that 1 source symbol is transmitted per channel use on average. To achieve the excess-distortion reliability function, Truong and Tan [17] used separate source and channel codes: the source is compressed down to its rate-distortion function, and the compressed symbols are transmitted using the Y-I communication phase, while the Y-I confirmation phase is modified to compare the uncompressed source and its lossy estimate instead of the compressed symbol and the estimate thereof. Due to the modification, some channel coding errors bear no effect on the overall decoding error, and the overall decoding error is dominated by the decoding error of the repetition code in the confirmation phase.

While most feedback coding schemes in the literature considered block encoding of a source whose outputs are accessible in their entirety before the transmission [1]–[10], [17], several existing works considered instantaneous encoding of a streaming source [18]–[24], [11]. A large portion of them [18]–[22] explores instantaneous (causal) encoding schemes for stabilizing a control system. The evolving system state is considered as a streaming data source, the observer instantaneously transmits information about the state to the controller, and the controller injects control signals into the plant. Sahai and Mitter [18] defined the *anytime* capacity at anytime reliability α as the maximum transmission rate R (nats per channel use) such that the decoding error of the first k R-nat symbols at time t decays as $e^{-\alpha(t-k)}$ for any $k \leq t$; they showed that the scalar linear system can be stabilized provided that the logarithm of its unstable coefficient is less than the anytime capacity; they suggested that codes that lead to an exponentially decaying error have a natural tree structure (similar to Schulman's code [19] for interactive computing) that tracks the state evolution over time. Tree coding schemes for stabilizing control systems have been studied in [20]–[21]. Assuming that the inter-arrival times of message bits are known by the decoder and that the channel is a BSC, Lalitha et al. [22] proposed an anytime code [18] that achieves a positive anytime reliability and derived a lower bound on the maximum rate that leads to an exponentially vanishing error probability. Instantaneous encoding schemes have also been studied in pure communication settings, where one may evaluate the error exponent [23][11], consider a streaming source with finite length [24][11], and allow non-periodic deterministic [22] or random [11] streaming times. Chang and Sahai [23] considered instantaneous encoding of i.i.d. message symbols that arrive at the encoder at consecutive times for the transmission over a binary erasure channel (BEC) with feedback, and showed the zero-rate JSCC error exponent of erroneously decoding the k-th message symbol at time t for fixed k and $t \to \infty$. Antonini et al. [24] designed a causal encoding scheme for $k < \infty$ streaming bits with a fixed arrival rate over a BSC and showed by simulation that the code rate approaches the channel capacity as the bit arrival rate approaches the transmission rate. In our previous work [11], we proposed a code that uses an adapted SED rule [8] to instantaneously transmit $k < \infty$ randomly arriving bits and that leads to an achievability bound on the reliability function for binary-input DMCs with instantaneous encoding, and we designed a polynomial-time version of it. While the instantaneous encoding schemes in [18]–[24], [11] employ feedback, transmission schemes for streaming data *without* feedback have been investigated for finite memory encoders [25], for distributed sources [26], and for point-to-

point channels in the moderate deviations [27] and the central limit theorem [28] regimes.

In this paper, we propose a novel coding phase – the instantaneous encoding phase – for transmitting a sequence of k source symbols over a DMC with feedback. It performs instantaneous encoding during the arriving period of the symbols. At time t, the encoder and the decoder calculate the priors of all possible symbol sequences using the source distribution and the posteriors at time $t-1$. Then, they partition the evolving message alphabet into groups, so that the group priors are close to the capacity-achieving distribution. In contrast to Naghshvar et al.'s SED rule [8] for symmetric binary-input channels, our partitioning rule applies to any DMCs, and it uses group priors instead of group posteriors for the partitioning. Using group priors is necessary because if a new symbol arrives at time t, the posteriors at time $t-1$ are insufficient to describe the symbol sequences at time t. Feedback codes with block encoding [1]–[10], [17] only need to consider the posteriors, since block encoding implies that the priors at time t are equal to the posteriors at time $t-1$. Once the groups are partitioned, the encoder determines the index of the group that contains the true symbol sequence it received so far and applies randomization to match the distribution of the transmitted index to the capacity-achieving one.

We derive the JSCC reliability function for the almost lossless transmission of a discrete streaming source over a DMC with feedback. Since allowing the encoder to know the entire source sequence before the transmission will not decrease the reliability function, converse bounds for a classical fully accessible source pertain. We extend Berlin et al.'s converse bound [31] for Burnashev's reliability function to JSCC. For fully accessible sources, we show that the converse is achievable by a variable-length joint source-channel code with block encoding – the MaxEJS code [8]. For a source whose symbols arrive at the encoder with an infinite arriving rate (symbols per channel use) as the source length goes to infinity, we show that the converse is achievable by the buffer-then-transmit code that buffers the arriving symbols during the symbol arriving period and implements a block encoding scheme that achieves the JSCC reliability function for a fully accessible source after the arriving period. For example, a classical fully accessible source has an infinite symbol arriving rate because its symbols arrive all at once. Yet, this buffer-then-transmit code fails to achieve the JSCC reliability function for streaming if the source symbols arrive at the encoder with a finite arriving rate of symbols per channel use. For streaming symbols with an arriving rate greater than $\frac{1}{\underline{H}}\left(H(P_Y^*)-\log\frac{1}{p_{\max}}\right)$, we show that preceding any code with block encoding that achieves the JSCC reliability function for a fully accessible source by our instantaneous encoding phase will make it achieve the block encoding error exponent as if the encoder knew the entire source sequence before the transmission. Here $\underline{H}$ is a lower bound on the information in the streaming source and is equal to the source entropy rate if the source is information stable, $H(P_Y^*)$ is the entropy of the channel output distribution induced by the capacity-achieving channel input distribution, and $p_{\max}$ is the maximum channel transition probability. Thus, surprisingly, the JSCC reliability function for streaming is equal to that for a fully accessible source. Furthermore, we show via simulations that the reliability function gives a surprisingly good approximation to the delay-reliability tradeoffs attained by the JSCC reliability function-achieving codes in the ultra-short blocklength regime.

The above discussion highlights the existence of a sequence of codes with instantaneous encoding indexed by the length of the source sequence k that achieves the JSCC reliability function as $k\to\infty$. However, in the remote tracking and control scenarios, a single code that can choose to decode any k symbols of a streaming source at any time t with an error probability that decays exponentially with the decoding delay (i.e., an anytime code [18]) is desired. To this end, we design the *instantaneous small-enough difference (SED) code*. The instantaneous SED code is similar to the instantaneous encoding phase except that it continues the transmissions after the symbol arriving period, drops the randomization step, and specifies the group partitioning rule to the instantaneous SED rule. The instantaneous SED code is also similar to the instantaneous encoding scheme in our previous work [11] designed for transmitting a streaming source with random symbol arriving times unknown to the decoder, except that [11] used an instantaneous *smallest-difference* rule. The instantaneous smallest-difference rule minimizes the difference between the group priors and the capacity-achieving probabilities, whereas the instantaneous SED rule only drives their difference small enough. The instantaneous SED rule reduces to Naghshvar et al.'s [8] SED rule if the source is fully accessible before the transmission. In contrast to the instantaneous encoding phase followed by a block encoding scheme, the instantaneous SED code only has one phase, namely, it follows the same transmission strategy at each time. For transmitting i.i.d. Bernoulli$\left(\frac{1}{2}\right)$ bits that arrive at the encoder at consecutive times over a BSC(0.05), simulations of the instantaneous SED code show that the error probability of decoding the first $k=[4\colon 4\colon 16]$ bits at times $t\in[4,64]$, $t\geq k$, decreases exponentially with an anytime reliability $\alpha\simeq 0.172$, outperforming the theoretical anytime reliability of Lalitha et al's anytime code [22]. This implies that the binary instantaneous SED code can be used to stabilize an unstable linear system with bounded noise [18]–[22]. Although the achievability of a positive anytime reliability is evidenced by the simulation, it is difficult to prove this analytically since one cannot leverage the submartingales and the bounds on the expected decoding time of a block encoding scheme in [5], [8]. Nevertheless, we show that a sequence of instantaneous SED codes indexed by the length of the symbol sequence k achieves the JSCC reliability function for streaming over a Gallager-symmetric [14, p. 94] binary-input DMC. This result is based on our finding that, after dropping the randomization step, the instantaneous encoding phase continues to achieve the JSCC reliability function when followed by a reliability function-achieving block encoding scheme, but at a cost of increasing the lower bound on the symbol arriving rate to $\frac{1}{\log\frac{1}{p_{S,\max}}}\left(\log\frac{1}{p_{\min}}-\log\frac{1}{p_{\max}}\right)$. Here, $p_{S,\max}$ is the maximum symbol arriving probability and $p_{\min}$

is the minimum channel transition probability.

Since the size of the evolving source alphabet grows exponentially in time t, the complexities of the instantaneous encoding phase and the instantaneous SED code are exponential in time t. For the source symbols that are equiprobably distributed, we design low-complexity algorithms for both codes that we term *type-based* codes. The complexity reduction is achieved by judiciously partitioning the evolving source alphabet into *types*. The cardinality of the partition is $O(t)$, i.e., it is exponentially smaller than the size of the source alphabet. The type partitioning enables the encoder and the decoder to update the priors and the posteriors of the source sequences as well as to partition source sequences in terms of types rather than individual sequences. Since the prior and the posterior updates have a linear complexity in the number of types, and the type-based group partitioning rule has a log-linear complexity in the number of types due to type sorting, our type-based codes only have a log-linear complexity $O(t \log t)$. Although Antonini et al.'s block encoding scheme for BSCs [9] attains a reduction in complexity also by grouping message sequences, the types in Antonini et al.'s scheme [9] are generated all at once by grouping the message sequences that have the same Hamming distance to the received channel outputs, while the types in our type-based codes evolve with the arrival of source symbols.

For the transmission over a degenerate DMC, i.e., a DMC whose transition matrix contains a zero, we propose a code with instantaneous encoding that achieves zero error for all rates asymptotically below Shannon's JSCC limit. Wh back codes in most prior literature [6]–[10], [17] are for non-degenerate DMCs, i.e., a DMC whose transiti ability matrix has all positive entries, Burnashev [5, constructed a channel code for degenerate DMCs that zero error for all rates asymptotically below the capacity. Our code extends Burnashev's code [5, Se JSCC and to the streaming source. Similar to [5]–[our code is divided into blocks, and each block cons communication phase and a confirmation phase. Bur [5, Sec. 6] communication phases use a block e scheme that can transmit reliably for all rates be channel capacity. The communication phase in the fi of our scheme uses a code with instantaneous encoc can transmit reliably for all rates below Shannon's JSC our ℓ-th communication phase transmits the uncon source sequence to avoid compression errors, and uses coding to establish an analyzable probability distributi decoding time. Our confirmation phase is the same a Burnashev's code [5, Sec. 6]: the encoder repeatedly t a pre-selected symbol that never leads to channel ou the decoder's estimate at the end of the communicati is wrong, and transmits another symbol that can le if the estimate is correct. The confirmation phases rel degenerate nature of the channel to ensure zero error: r a y secures an error-free estimate of the source.

The rest of the paper is organized as follows. In Se we formulate the problem and define the variabl joint source-channel codes with instantaneous enco Section III, we present the instantaneous encoding p Section IV, we show the JSCC reliability function for ing. In Section V, we present the instantaneous SED code. In Section VI, we present the type-based codes with log-linear complexity. In Section VII, we display the simulations of the instantaneous encoding phase, the instantaneous SED code, and their corresponding type-based codes. In Section VIII, we present the zero-error code for degenerate DMCs.

A part of this work is presented at the 2022 IEEE International Symposium on Information Theory [44]. The conference version does not contain Sections V–VIII or any proofs.

Notation: $\log(\cdot)$ is the natural logarithm. Notation $X \leftarrow Y$ reads "replace X by Y". For any positive integer q, we denote $[q] \triangleq \{1, 2, \ldots, q\}$. We denote by $[q]^k$ the set of all q-ary sequences of length equal to k. For a possibly infinite sequence $x = \{x_1, x_2, \ldots\}$, we write $x^n = \{x_1, x_2, \ldots, x_n\}$ to denote the vector of its first n elements, and we write $\{x_n\}_{n=n_1}^{n_2} = \{x_{n_1}, x_{n_1+1}, \ldots, x_{n_2}\}$ to denote the vector formed by its $n_1, n_1+1, \ldots, n_2$-th elements. For a sequence of random variables X_k, $k = 1, 2, \ldots$ and a real number $a \in \mathbb{R}$, we write $X_k \xrightarrow{\text{i.p.}} a$ to denote that X_k converges to a in probability, i.e., $\lim_{k\to\infty} \mathbb{P}[|X_k - a| \geq \epsilon] = 0$, $\forall \epsilon > 0$. For any set $\mathcal{A}$, we denote by $\mathbb{1}_{\mathcal{A}}(x)$ an indicator function that is equal to 1 if and only if $x \in \mathcal{A}$. For two positive functions $f, g\colon \mathbb{Z}_+ \to \mathbb{R}_+$, we write $f(k) = o(g(k))$ to denote $\lim_{k\to\infty} \frac{f(k)}{g(k)} = 0$; we write $f(k) = O(g(k))$ to denote $\limsup_{k\to\infty} \frac{f(k)}{g(k)} < \infty$; we write $f(k) = \Omega(g(k))$ to denote $\liminf \frac{f(k)}{g(k)} > 0$.

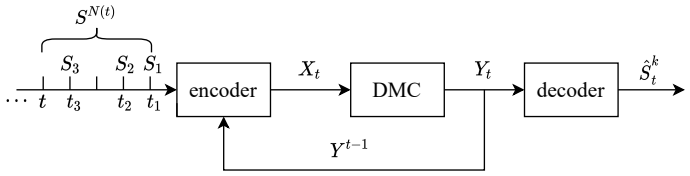


Fig. 1. Real-time feedback communication system with a streaming source.

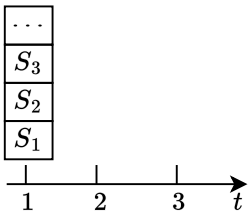
Definition 1 (A $(q, \{t_n\}_{n=1}^{\infty})$ discrete streaming source (DSS))**.** *We say that a source is a DSS if it emits a sequence of discrete source symbols $S_n \in [q]$, $n = 1, 2, \ldots$ at times $t_1 \leq t_2 \leq \ldots$, where symbol S_n that arrives at the encoder at time t_n is distributed according to the source distribution*

$$P_{S_n|S^{n-1}}, n = 1, 2, \ldots \tag{1}$$

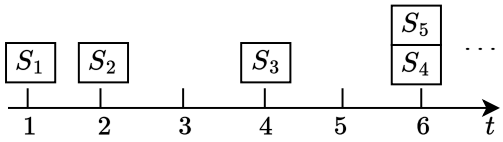
Throughout, we assume that the entropy rate of the DSS

$$H \triangleq \lim_{n\to\infty} \frac{H(S^n)}{n} \text{ (nats per symbol)} \tag{2}$$

is well-defined and positive; the first symbol S_1 arrives at the encoder at time $t_1 \triangleq 1$; both the encoder and the decoder know the symbol alphabet $[q]$, the arrival times $t_1, t_2, \ldots$, and the source distribution (1). The DSS reduces to the classical



(a) fully accessible: $1 = t_1 = t_2 = \dots$



(b) streaming: $t_1 = 1, t_2 = 2, t_3 = 4, t_4 = t_5 = 6, \dots$

Fig. 2. A fully accessible source vs. a streaming source. A fully accessible source emits all symbols at $t = 1$. A streaming source emits symbols progressively.

discrete source (DS) that is fully accessible to the encoder before the transmission if

$$t_n = 1, \ \forall n = 1, 2, \dots \tag{3}$$

Fig. 2 displays a fully accessible source and a streaming source. Operationally, symbol S_n represents a data packet. We denote the number of symbols that the encoder has received by time t by

$$N(t) \triangleq \max\{n \colon t_n \le t, n = 1, 2, \dots\}. \tag{4}$$

Given a DSS (Definition 1) with symbol arriving times $t_1, t_2, \dots$, we denote its *symbol arriving rate* by, assuming that the limit exists

$$f \triangleq \lim_{n\to\infty} \frac{n}{t_n} \text{ (symbols per unit time)} \in [0, \infty]. \tag{5}$$

The symbol arriving rate $f = \infty$ implies that the source symbols arrive at the encoder so frequently that the number of channel uses increases slower than the source length. For example, the DS (3) has $f = \infty$. The symbol arriving rate $f < \infty$ implies that the number of channel uses goes to infinity as the source length goes to infinity. For example, if one source symbol arrives at the encoder every $\lambda \ge 1$ channel uses, $\lambda \in \mathbb{Z}_+$, i.e.,

$$t_n = \lambda(n-1) + 1, \tag{6}$$

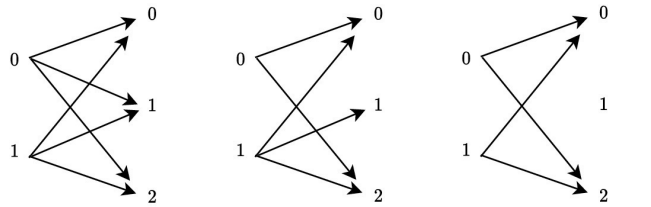
then

$$f = \frac{1}{\lambda}. \tag{7}$$

We assume that the channel is a DMC with a single-letter transition probability distribution $P_{Y|X} \colon \mathcal{X} \to \mathcal{Y}$.

Definition 2 (Non-degenerate and degenerate DMCs)**.** *A DMC is non-degenerate if it satisfies*

$$P_{Y|X}(y|x) > 0, \forall x \in \mathcal{X}, y \in \mathcal{Y}. \tag{8}$$



(a) non-degenerate (b) degenerate (c) neither non-degenerate nor degenerate

Fig. 3. A DMC $P_{Y|X} \colon \{0,1\} \to \{0,1,2\}$. An arrow from channel input $x \in \{0,1\}$ to channel output $y \in \{0,1,2\}$ signifies $P_{Y|X}(y|x) > 0$. Channel (a) is a non-degenerate DMC that satisfies (8). Channel (b) is a degenerate DMC that satisfies (9) with $y = 1$, $x = 1$, $x' = 0$. Channel (c) does not satisfy (8)–(9) since $y = 1$ is not reachable.

A DMC is degenerate if there exist $y \in \mathcal{Y}$, $x \in \mathcal{X}$, $x' \in \mathcal{X}$, such that

$$P_{Y|X}(y|x) > 0, \tag{9a}$$

$$P_{Y|X}(y|x') = 0. \tag{9b}$$

A non-degenerate DMC is considered in [5]–[10], e.g., a BSC. A degenerate DMC is considered in [5, Sec.6], e.g., a BEC. Fig. 3 display examples of DMCs. We denote the capacity of the DMC by

$$C \triangleq \max_{P_X} I(X;Y), \tag{10}$$

and we denote the maximum Kullback–Leibler (KL) divergence between its transition probabilities by

$$C_1 \triangleq \max_{x,x' \in \mathcal{X}} D(P_{Y|X=x} || P_{Y|X=x'}). \tag{11}$$

Assumption (8) posits that C_1 (11) is finite.

A DMC is *symmetric* (Gallager-symmetric [14, p. 94]) if the columns in its channel transition probability matrix can be partitioned so that within each partition, all rows are permutations of each other, and all columns are permutations of each other.

Throughout, we measure the symbol arriving rate (5) with a unit time equal to a channel use.

We proceed to define the codes that we use to transmit a DSS over a DMC with feedback. All the codes in this paper are *variable-length joint source-channel codes with feedback*. We distinguish two classes of codes, one is called a code with instantaneous encoding, and the other is called a code with block encoding. Next, we define the code with instantaneous encoding designed to recover the first k symbols of a DSS at rate R symbols per channel use and error probability ϵ.

Definition 3 (A (k, R, ϵ) code with instantaneous encoding)**.** *Fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS and fix a DMC with a single-letter transition probability distribution $P_{Y|X} \colon \mathcal{X} \to \mathcal{Y}$. An (k, R, ϵ) code with instantaneous encoding consists of:*
1. a sequence of (possibly randomized) encoding functions $\mathsf{f}_t \colon [q]^{N(t)} \times \mathcal{Y}^{t-1} \to \mathcal{X}$, $t = 1, 2, \dots$ that the encoder uses to form the channel input

$$X_t \triangleq \mathsf{f}_t\left(S^{N(t)}, Y^{t-1}\right); \tag{12}$$

2. a sequence of decoding functions $\mathsf{g}_t\colon \mathcal{Y}^t \to [q]^k$, $t = 1, 2, \ldots$ *that the decoder uses to form the estimate*

$$\hat{S}_t^k \triangleq \mathsf{g}_t(Y^t); \tag{13}$$

3. a stopping time η_k *adapted to the filtration generated by the channel output* $Y_1, Y_2, \ldots$ *that determines when the transmission stops and that satisfies*

$$\frac{k}{\mathbb{E}[\eta_k]} \geq R \text{ (symbols per channel use)}, \tag{14}$$

$$\mathbb{P}[\hat{S}_{\eta_k}^k \neq S^k] \leq \epsilon. \tag{15}$$

For any rate $R > 0$, the minimum error probability achievable by rate-R codes with instantaneous encoding and message length k is given by

$$\epsilon^*(k, R) \triangleq \inf\{\epsilon\colon \exists \ (k, R, \epsilon) \text{ code with instantaneous encoding}\}. \tag{16}$$

For transmitting a DSS over a non-degenerate DMC with noiseless feedback via a code with instantaneous encoding, we define the *JSCC reliability function for streaming* as

$$E(R) \triangleq \lim_{k\to\infty} \frac{R}{k} \log \frac{1}{\epsilon^*(k, R)}. \tag{17}$$

If a DSS satisfies (3), i.e., a DS, a code with instantaneous encoding (i.e., causal code) in Definition 3 reduces to a code with block encoding (i.e., non-causal code), and the JSCC reliability function for streaming (17) reduces to the JSCC reliability function for a fully accessible source.

We use $E(R)$ (17) to quantify the fundamental delay-reliability tradeoff achieved by codes with instantaneous encoding. The reliability function is a classical performance metric that can be used to approximate that tradeoff as $\epsilon \simeq e^{-\frac{k}{R}E(R)}$. Although this approximation ignores the subexponential terms, it still sheds light on the finite-blocklength performance as our numerical simulations in Section VII demonstrate.

Similar to classical codes with block encoding, a (k, R, ϵ) code with instantaneous encoding in Definition 3 is designed to recover only the first k symbols of a DSS, and $E(R)$ (17) is achieved by a sequence of codes with instantaneous encoding indexed by the length of the symbol sequence k as $k \to \infty$. We proceed to define a code with instantaneous encoding that decodes the first k symbols at a time $t \geq t_k$ with an error probability that decays exponentially with delay $t - t_k$, for all k and t. Because the decoding time and the number of symbols to decode can be chosen on the fly, this code is referred to as an *anytime* code and can be used to stabilize an unstable linear system with bounded noise over a noisy channel with feedback [18]. We formally define anytime codes as follows.

Definition 4 (A (κ, α) anytime code)**.** *Fix a* $(q, \{t_n\}_{n=1}^\infty)$ *DSS and fix a DMC with a single-letter transition probability distribution* $P_{Y|X}\colon \mathcal{X} \to \mathcal{Y}$*. A* (κ, α) *anytime code consists of:*

1. a sequence of (possibly randomized) encoding functions defined in Definition 3-1;

2. a sequence of decoding functions $\mathsf{g}_{t,k}\colon \mathcal{Y}^t \to [q]^k$ *indexed both by the decoding time* t *and the length of the decoded symbol sequence* k *that the decoder uses to form an estimate* $\hat{S}_t^k \triangleq \mathsf{g}_{t,k}(Y^t)$ *of the first* k *symbols at time* t*.*

For all $k = 1, 2, \ldots$, $t = 1, 2, \ldots$, $t \geq t_k$*, the error probability of decoding the first* k *symbols at time* t *must satisfy*

$$\mathbb{P}[\hat{S}_t^k \neq S^k] \leq \kappa e^{-\alpha(t - t_k)} \tag{18}$$

for some $\kappa, \alpha \in \mathbb{R}_+$*.*

The exponentially decaying rate α of the error probability in (18) is referred to as the anytime reliability. While Sahai and Mitter's anytime code in [18, Definition 3.1] is defined to transmit a DSS that emits source symbols one by one at consecutive times, Definition 4 slightly extends [18, Definition 3.1] to a general DSS in Definition 1.

In this paper, we aim to find $E(R)$ (17), the codes with instantaneous encoding that achieve $E(R)$, and an anytime code.

III. Instantaneous Encoding Phase

With the aim of transmitting the first k source symbols of a DSS, we present our instantaneous encoding phase, which specifies the encoding functions $\{\mathsf{f}_t\}_{t=1}^{t_k}$ in Definition 3. We fix a DMC with a single-letter transition probability distribution $P_{Y|X}\colon \mathcal{X} \to \mathcal{Y}$ and capacity-achieving distribution P_X^*, and we fix a $(q, \{t_n\}_{n=1}^\infty)$ DSS with distribution (1). We denote the following functions of the channel outputs,

$$\rho_i(Y^t) \triangleq P_{S^{N(t)}|Y^t}(i|Y^t), \tag{19}$$

$$\theta_i(Y^{t-1}) \triangleq P_{S^{N(t)}|Y^{t-1}}(i|Y^{t-1}), \tag{20}$$

$$\pi_x(Y^{t-1}) \triangleq \sum_{i \in \mathcal{G}_x(Y^{t-1})} \theta_i(Y^{t-1}), \tag{21}$$

where we refer to $\rho_i(Y^t)$ and $\theta_i(Y^t)$ as the posterior and the prior of source sequence $i \in [q]^{N(t)}$, respectively; we refer to $\pi_x(Y^{t-1})$ as the prior of the group $\mathcal{G}_x(Y^{t-1})$ corresponding to channel input $x \in \mathcal{X}$ that we specify in (24) below. The probability distributions $P_{S^{N(t)}|Y^t}$ and $P_{S^{N(t)}|Y^{t-1}}$ are determined by the code below.

Algorithm: The instantaneous encoding phase operates during times $t = 1, 2, \ldots, t_k$.

At each time t, the encoder and the decoder first update the priors $\theta_i(y^{t-1})$ for all $i \in [q]^{N(t)}$. At symbol arriving times $t = t_n$, $n = 1, 2, \ldots, k$ the prior $\theta_i(y^{t-1})$, $i \in [q]^{N(t)}$ is updated using the posterior $\rho_{i^{N(t-1)}}(y^{t-1})$ and the source distribution (1), i.e.,

$$\theta_i(y^{t-1}) = P_{S^{N(t)}|S^{N(t-1)}}\left(i|i^{N(t-1)}\right) \rho_{i^{N(t-1)}}(y^{t-1}), \tag{22}$$

where $i^{N(t-1)}$ is the length-$N(t-1)$ prefix of sequence i. At times in-between arrivals, i.e., at $t \in (t_n, t_{n+1})$, $n = 1, 2, \ldots, k-1$, the prior $\theta_i(y^{t-1})$ is equal to the posterior $\rho_i(y^{t-1})$ for all $i \in [q]^{N(t)}$, i.e.,

$$\theta_i(y^{t-1}) = \rho_i(y^{t-1}). \tag{23}$$

At each time t, once the priors are updated, the encoder and the decoder partition the message alphabet $[q]^{N(t)}$ into $|\mathcal{X}|$ disjoint groups $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ such that for all $x \in \mathcal{X}$,

$$\pi_x(y^{t-1}) - P_X^*(x) \leq \min_{i \in \mathcal{G}_x(y^{t-1})} \theta_i(y^{t-1}). \tag{24}$$

The partitioning rul
$\{\pi_x(y^{t-1})\}_{x\in\mathcal{X}}$ are
distribution $\{P_X^*(x)\}$
$\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ of $[q$
(24), since the partitio
[35] satisfies it, see th

Using the partitio
decoder construct tw
ors $\{\pi_x(y^{t-1})\}_{x\in\mathcal{X}}$
$\{P_X^*(x)\}_{x\in\mathcal{X}}$:

$$\underline{\mathcal{X}}(y^{t-1}) \triangleq$$
$$\overline{\mathcal{X}}(y^{t-1}) \triangleq$$

Then, the encoder
probabilities $\{p_{\overline{x}\to\underline{x}}\}_{\overline{x}}$
channel input, such th
holds that

$$\pi_{\overline{x}}(y^{t-1})$$

$$\pi_{\underline{x}}(y^{t-1})$$

The set $\{p_{\overline{x}\to\underline{x}}\}_{\overline{x}\in\overline{\mathcal{X}}(y}$
algorithm in Appendi

The output of the
follows. The encoder
the sequence $S^{N(t)}$ it

$$Z_t \triangleq$$

Then, the encoder ou

$$P_{X_t|Z_t,Y^{t-1}}(x$$
$$= \begin{cases} \frac{P_X^*(z)}{\pi_z(y^{t-1})}, \\ \frac{p_{z\to x}}{\pi_z(y^{t-1})}, \\ \mathbb{1}_{\{z\}}(x), \\ 0, \end{cases}$$

The decoder also
$P_{X_t|Z_t,Y^{t-1}}$ (30), sinc
(24), sets $\overline{\mathcal{X}}(y^{t-1})$ ar
$\{p_{\overline{x}\to\underline{x}}\}_{\overline{x}\in\overline{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}}$
channel input distribu
the capacity-achievin
$y^{t-1}\in\mathcal{Y}^{t-1}$,

$$P_{X_t|}$$

See the proof of (31)
an example of group
ization.

Upon receiving the
encoder and the decc
possible sequences o

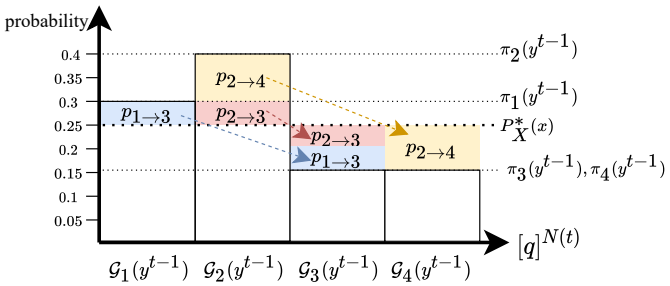


Fig. 4. An example of group partitioning and channel input randomization for a DMC with uniform capacity-achieving distribution $P_X^*(x) = 0.25$, $\mathcal{X} = [4]$. The horizontal axis represents a partition of 4 groups. The vertical axis represents the prior probabilities of the groups. The source alphabet $[q]^{N(t)}$ is partitioned into $\{\mathcal{G}_x(y^{t-1})\}_{x\in[4]}$ such that the partitioning rule (24) is satisfied. Groups $\mathcal{G}_x(y^{t-1}), x\in\{1,2\}$ constitute $\overline{\mathcal{X}}(y^{t-1})$ (26) and groups $\mathcal{G}_x(y^{t-1}), x\in\{3,4\}$ constitute $\underline{\mathcal{X}}(y^{t-1})$ (25). The probabilities $\{p_{\overline{x}\to\underline{x}}\}_{\overline{x}\in\{1,2\},\underline{x}\in\{3,4\}}$ (27)–(28) used to randomize transmitted group indices are colored. The randomization matches the probability of transmitting group index $x\in[4]$ to $P_X^*(x)$.

prior $\theta_i(y^{t-1})$, the channel output y_t, and the randomization probability (30), i.e.,

$$\rho_i(y^t) = \frac{\sum_{x\in\mathcal{X}} P_{Y|X}(y_t|x) P_{X_t|Z_t,Y^{t-1}}(x|z(i),y^{t-1})}{P_Y^*(y_t)}\theta_i(y^{t-1}), \tag{32}$$

where $z(i)$ is the index of the group that contains sequence i, i.e., it is equal to the right side of (29) with $S^{N(t)}\leftarrow i$; P_Y^* is the channel output distribution induced by the capacity-achieving distribution P_X^*; (32) holds due to (31) and the Markov chain $Y_t - X_t - (Z_t, Y^{t-1}) - S^{N(t)}$.

We conclude the presentation of the instantaneous encoding phase with several remarks.

The randomization (25)–(30) of the instantaneous encoding phase is only used for analysis: Theorem 1 in Section IV continues to hold if the randomization step (25)–(30) is dropped and the deterministic group index Z_t (29) is transmitted, but at a cost of imposing assumptions on the DSS that are stricter than assumptions (a)–(b) in Theorem 1. See Remark 1 in Section IV for details. From the perspective of encoding, the randomization (30) turns the encoding function f_t into a stochastic kernel $P_{X_t|S^{N(t)},Y^{t-1}}$. From the perspective of the channel, the randomization $P_{X_t|Z_t,Y^{t-1}}$ (30) together with the DMC $P_{Y|X}$ can be viewed as a cascaded DMC with channel input (Z_t, Y^{t-1}). The randomness in (29) is not common with the decoder as it only needs to know the distribution $P_{X_t|Z_t,Y^{t-1}}$ to update posterior $\rho_i(y^t)$ in (32).

The complexity of the instantaneous encoding phase is $O\left(q^{N(t)}\log q^{N(t)}\right)$ if the classical greedy heuristic algorithm (Appendix A) is used for group partitioning (24). For equiprobably distributed source symbols, we design an efficient algorithm that reduces the complexity down to $O(t\log t)$ in Section VI.

IV. JSCC Reliability Function

In this section, we show the JSCC reliability function for streaming $E(R)$ (17) using the instantaneous encoding phase

introduced in Section III. For brevity, we denote the maximum and the minimum channel transition probabilities of a DMC $P_{Y|X}\colon \mathcal{X}\to\mathcal{Y}$ by

$$p_{\max} \triangleq \max_{x\in\mathcal{X},y\in\mathcal{Y}} P_{Y|X}(y|x), \tag{33}$$

$$p_{\min} \triangleq \min_{x\in\mathcal{X},y\in\mathcal{Y}} P_{Y|X}(y|x), \tag{34}$$

and we denote the maximum symbol arriving probability of the DSS (1) by

$$p_{S,\max} \triangleq \max_{n\in\mathbb{N},s\in[q],s'\in[q]^{n-1}} P_{S_n|S^{n-1}}(s|s'). \tag{35}$$

Theorem 1. *Fix a non-degenerate DMC with capacity C (10), maximum KL divergence C_1 (11), and maximum channel transition probability $p_{\max}$ (33). Fix a $(q,\{t_n\}_{n=1}^{\infty})$ DSS with entropy rate $H>0$ (2) and symbol arriving rate f (5). If the DSS has $f<\infty$ (5), then we further assume that*

(a) *the information in the DSS is asymptotically lower bounded as*

$$\lim_{n\to\infty}\mathbb{P}\left[\frac{1}{n}\log\frac{1}{P_{S^n}(S^n)}\geq \underline{H}\right]=1 \tag{36}$$

for some $\underline{H}\in(0,\infty)$;

(b) *the symbol arriving rate is large enough:*

$$f>\frac{1}{\underline{H}}\left(H(P_Y^*)-\log\frac{1}{p_{\max}}\right). \tag{37}$$

Then, the JSCC reliability function for streaming (17) is equal to

$$E(R)=C_1\left(1-\frac{H}{C}R\right),\ 0<R<\frac{C}{H}. \tag{38}$$

Proof sketch. The converse proof is in Appendix D: allowing the encoder to know the entire source sequence before the transmission will not reduce the JSCC reliability function, therefore converse bounds for a (fully accessible) DS apply. Namely, we lower bound the expected decoding time for any code with block encoding to attain a target error probability using Fano's inequality and a binary hypothesis test. This extends Berlin et al.'s [31] converse bound on Burnashev's reliability function applicable to the channel coding setting to the JSCC setting.

The achievability proof is in Appendix E: for any (fully accessible) DS, the JSCC reliability function (38) is achievable by the MaxEJS code [8, Sec. IV-C], and is achievable by the SED code [8, Sec. V-B] if the channel is a symmetric binary-input DMC (Appendix E-A).

For any DSS with $f=\infty$, including the DS (3), the buffer-then-transmit code for k source symbols that achieves $E(R)$ (38) operates as follows. It waits until the k-th symbol arrives at time t_k, and at times $t\geq t_k+1$, applies a JSCC reliability function-achieving code with block encoding for k symbols S^k of a (fully accessible) DS with prior P_{S^k} (1) (e.g., the MaxEJS code or the SED code [8]). The buffer-then-transmit code achieves (see details in Appendix E-B)

$$E(R)\geq C_1\left(1-\left(\frac{H}{C}+\frac{1}{f}\right)R\right), \tag{39}$$

which reduces to $E(R)$ (38) for $f=\infty$. Indeed, $f=\infty$ means that the arrival time t_k is negligible compared to the blocklength. The buffer-then-transmit code fails to achieve $E(R)$ (38) if $f<\infty$.

For any DSS with $f<\infty$ that satisfies the assumptions (a)–(b) in Theorem 1, the code with instantaneous encoding for k source symbols that achieves $E(R)$ (38) implements the instantaneous encoding phase (Section III) at times $t=1,2,\dots,t_k$ and operates as a JSCC reliability function-achieving code with block encoding for k symbols S^k of a (fully accessible) DS with prior $P_{S^k|Y^{t_k}}$ at times $t\geq t_k+1$, where $Y_1,\dots,Y_{t_k}$ are the channel outputs generated in the instantaneous encoding phase. For example, we can insert the instantaneous encoding phase before the MaxEJS code (or the SED code for symmetric binary-input DMCs). See Appendix E-C. ■

Assumption (a) holds with $\underline{H}=H$ for any information stable source since such sources satisfy $\frac{1}{n}\log\frac{1}{P_{S^n}(S^n)}\xrightarrow{\text{i.p.}}H$ [29]. For example, $\underline{H}=H(S)$ if the source emits i.i.d. symbols. Assumption (b) in Theorem 1 implies

$$f\geq\frac{C}{H} \tag{40}$$

since $H(Y|X)\geq\log\frac{1}{p_{\max}}$ and $H\geq\underline{H}$. The symbol arriving rate constraint (40) ensures that all coding rates $R<\frac{C}{H}$ are achievable. Otherwise, if (40) is not satisfied and the DSS has $p_{S,\max}<1$, the rate region achievable by any code with instantaneous encoding is limited to $R\leq f$. The limitation arises because decoding S^k before the final arrival time t_k results in a non-vanishing error probability (Appendix F). For example, if the DSS emits i.i.d. symbols with entropy rate $H=1$ nat per symbol arriving at the encoder every 1000 channel uses, and the DMC has capacity $C=1$ nat per channel use, then the achievable rate is limited by $\frac{1}{1000}$ symbols per channel use, which is far less than Shannon's JSCC limit $\frac{C}{H}=1$ symbol per channel use. While (40) gives a converse bound on the symbol arriving rate and (37) guarantees achievability of $E(R)$ (38), the existence of a critical symbol arriving rate f_{cr} such that for all $f>f_{\text{cr}}$, $E(R)$ (38) is achievable, and for all $f<f_{\text{cr}}$, $E(R)$ (38) is not achievable, remains open. While $E(R)$ (38) is not a function of f, it is conceivable that for $f<f_{\text{cr}}$, the reliability function (17) will depend on f. This is reminiscent of the channel reliability function for transmitting over a DMC without feedback via a fixed-length block code, which is known only for rates greater than a critical value where its converse bound (sphere-packing exponent [32]) coincides with its achievability bound (random-coding exponent [33]).

Since the (fully accessible) DS (3) is a special DSS, Theorem 1 gives the JSCC reliability function (38) for a fully accessible source. It generalizes Burnashev's reliability function [5] to the classical JSCC context, and generalizes Truong and Tan's excess-distortion reliability function [17] at zero distortion to the DS with memory and to all rates $R<\frac{C}{H}$.

Remarkably, Theorem 1 establishes that the JSCC reliability function for a streaming source (satisfying assumptions (a)–(b)) is equal to that for a fully accessible source. This is

surprising as this means that revealing source symbols only causally to the encoder has no detrimental effect on the reliability function.

While the instantaneous encoding phase in Section III achieves $E(R)$ (38), in fact, any coding strategy during the symbol arriving period that satisfies

$$\lim_{k\to\infty} \frac{I(S^k;Y^{t_k})}{t_k} = C \tag{41}$$

achieves $E(R)$ (38) when followed by a JSCC reliability function-achieving code with block encoding. This is because (99b)–(99c) in the achievability proof in Appendix E-C always hold for such a coding strategy. For equiprobably distributed q-ary source symbols that arrive at the encoder one by one at consecutive times $t = 1, 2, \ldots, k$ and a symmetric q-input DMC, uncoded transmission during the symbol arriving period $t = 1, 2, \ldots, k$ satisfies (41) and thus constitutes an appropriate instantaneous encoding phase for that scenario. If $q = 2$, this corresponds to the systematic transmission phase in [9]. Furthermore, even if the instantaneous encoding phase in Section III drops the randomization (25)–(30) and transmits Z_t (29) as the channel input, it continues to satisfy the sufficient condition (41) under a more conservative condition than (37) (see Remark 1 below).

Remark 1. *Fix a non-degenerate DMC with the maximum and the minimum channel transition probabilities $p_{\max}$ and $p_{\min}$, and fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with maximum symbol arriving probability $p_{S,\max} < 1$ and symbol arriving rate $f < \infty$. If the DSS satisfies*

(b′) *the symbol arriving rate is large enough:*

$$f > \frac{1}{\log\frac{1}{p_{S,\max}}}\left(\log\frac{1}{p_{\min}} - \log\frac{1}{p_{\max}}\right), \tag{42}$$

then the instantaneous encoding phase in Section III that transmits the non-randomized Z_t (29) as the channel input at each time $t = 1, 2, \ldots, t_k$ satisfies (41), which means that it achieves $E(R)$ (38), the JSCC reliability function for streaming, when followed by a JSCC reliability function-achieving code with block encoding.

Proof sketch. We show that under assumption (b′), all source priors $\theta_i(y^{t-1})$, $i \in [q]^{N(t)}$, converge pointwise to zero in t during the symbol arriving period $t \in [1, t_k]$ as $k \to \infty$. The convergent source priors and the partitioning rule (24) imply that the group priors converge pointwise to the capacity-achieving distribution P_X^*. Since the encoder transmits a group index without randomization as the channel input, the channel input distribution converges to the capacity-achieving distribution, yielding (41). See Appendix G for details. ■

Note that the result of Remark 1 does not require assumption (a) since $P_{S^n}(s^n) \le (p_{S,\max})^n$, $\forall s^n \in [q]^n$, already implies that it holds with $\underline{H} \leftarrow \log\frac{1}{p_{S,\max}}$.

Since $\underline{H} \ge \log\frac{1}{p_{S,\max}}$ and $\log\frac{1}{p_{\min}} \ge H(P_Y^*)$, assumption (b′) is stricter than assumption (b). The increase of the threshold is because 1) the channel output distribution P_Y^* in (109b) is replaced by $P_{Y_t|Y^{t-1}}(\cdot|\cdot) \ge p_{\min}$ (121); 2) in the proof of Remark 1, we show that all the source priors converge *pointwise* to zero (124) during the symbol arriving period as $k \to \infty$ using the upper bound $P_{S_n|S^{n-1}}(\cdot|\cdot) \le p_{S,\max}$, whereas in Theorem 1, we only need that the the source prior of the true symbol sequence converges *in probability* to zero (112).

V. Instantaneous SED code

While the JSCC reliability function-achieving codes with instantaneous encoding in Section III are designed to transmit the first k symbols of a DSS, and a sequence of such codes indexed by the source length k achieves $E(R)$ (38) as $k \to \infty$, we now show an anytime code (Definition 4) termed the instantaneous SED code. In Section V-A, we present the algorithm of the instantaneous SED code for a symmetric binary-input DMC. In Section V-B, we show by simulations that the instantaneous SED code empirically achieves a positive anytime reliability, and thus can be used to stabilize an unstable linear system with bounded noise over a noisy channel. In Section V-C, we show that if the instantaneous SED code is restricted to transmit the first k symbols of a DSS, a sequence of instantaneous SED codes indexed by the length of the symbol sequence k also achieves $E(R)$ (38) for streaming over a symmetric binary-input DMC.

A. Algorithm of the instantaneous SED code

The instantaneous SED code is almost the same as the instantaneous encoding phase in Section III, expect that 1) it particularizes the partitioning rule (24) to the instantaneous SED rule in (43)–(44) below; 2) its encoder does not randomize the channel input and transmits Z_t (29) at each time t; 3) it continues to operate after the symbol arriving period. Fixing a symmetric binary-input DMC $P_{Y|X}\colon \{0,1\} \to \mathcal{Y}$ and fixing a $(q, \{t_n\}_{n=1}^{\infty})$ DSS, we present the algorithm of the instantaneous SED code.

Algorithm: The instantaneous SED code operates at times $t = 1, 2, \ldots$

At each time t, the encoder and the decoder first update the priors $\theta_i\left(y^{t-1}\right)$ for all possible sequences $i \in [q]^{N(t)}$ that the source could have emitted by time t. If $t = t_n$, $n = 1, 2, \ldots$, the prior is updated using (22); otherwise, the prior is equal to the posterior (23).

Once the priors are updated, the encoder and the decoder partition the source alphabet $[q]^{N(t)}$ into 2 disjoint groups $\{\mathcal{G}_x\}_{x\in\{0,1\}}$ according to the *instantaneous SED rule*, which says the following: if $x, x' \in \{0,1\}$ satisfy

$$\pi_x(y^{t-1}) \ge \pi_{x'}(y^{t-1}), \tag{43}$$

then they must also satisfy

$$\pi_x(y^{t-1}) - \pi_{x'}(y^{t-1}) \le \min_{i\in\mathcal{G}_x(y^{t-1})} \theta_i(y^{t-1}). \tag{44}$$

There always exists a partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\{0,1\}}$ that satisfies the instantaneous SED rule (43)–(44) since the partition that attains the smallest difference $|\pi_0(y^{t-1}) - \pi_1(y^{t-1})|$ satisfies it [8, Appendix III-E].

Once the source alphabet is partitioned, the encoder transmits the index Z_t (29) of the group that contains the true source sequence $S^{N(t)}$ as the channel input.

Upon receiving the channel output $Y_t = y_t$ at time t, the encoder and the decoder update the posteriors $\rho_i\left(y^t\right)$ for all $i \in [q]^{N(t)}$ using the priors $\theta_i\left(y^{t-1}\right)$ and the channel output y_t, i.e.,

$$\rho_i\left(y^{t-1}\right) = \frac{P_{Y|X}(y_t|z(i))}{\sum_{x\in\mathcal{X}} P_{Y|X}(y|x)\pi_x(y^{t-1})}\theta_i(y^{t-1}), \tag{45}$$

where $z(i)$ is the index of the group that contains sequence i, i.e., it is equal to the right side of (29) with $S^{N(t)} \leftarrow i$.

The maximum a posteriori (MAP) decoder estimates the first k symbols at time t as

$$\hat{S}_t^k \triangleq \arg\max_{i\in[q]^k} P_{S^k|Y^t}(i|Y^t). \tag{46}$$

We conclude the presentation of the algorithm with several remarks.

We call the group partitioning rule in (43)–(44) instantaneous small-enough difference (SED) rule since it reduces to Naghshvar et al.'s SED rule [8] if the source is fully accessible to the encoder before the transmission. The rule ensures that the difference between a group prior $\pi_x(y^{t-1})$ and its corresponding capacity-achieving probability $P_X^*(x) = \frac{1}{2}$, $x \in \{0, 1\}$ is bounded by the source prior on the right side of (44).

Remark 2. *Even though the algorithm of the instantaneous SED code is presented for a DSS with deterministic symbol arriving times, it can be used to transmit a DSS with* random *symbol arriving times. In that case, the number of symbols $N(t)$ that have arrived by time t is a random variable, and the decoder only knows the symbol arriving distribution $\{P_{S^{N(t)}|S^{N(t-1)}}\}_{t=1}^{\infty}$ rather than the exact symbol arriving times. The instantaneous SED code can be used to transmit such a streaming source as long as the encoder and the decoder keep updating the source priors, partitioning the groups, and updating the posteriors at times $t = 1, 2, \ldots$ for* all possible *source sequences that can arrive at the encoder by time t, see our work [11] for the algorithm with the instantaneous SED rule replaced by an instantaneous smallest-difference rule.*

Designing a code with instantaneous encoding whose decoder knows neither the symbol arriving times nor the symbol arriving distribution remains an open problem. In this case, the decoder needs to learn the symbol arriving distribution online using the past symbol arriving times. Online learning of the distribution of source symbols is also an interesting research direction.

B. Instantaneous SED code is an anytime code

We first provide numerical evidence showing that the instantaneous SED code is an anytime code: it empirically attains an error probability that decreases exponentially as (18). We then determine which unstable scalar linear systems can be stabilized by the instantaneous SED code.

In Fig. 5, we display the error probability (18) of the instantaneous SED code, where the y-axis corresponds to the error probability of decoding the length-k prefix of a DSS at time t (18). At each time t, we generate a Bernoulli$\left(\frac{1}{2}\right)$ source bit and a realization of a BSC(0.05), run these experiments for 10^5 trials, and obtain the error probability (18) by dividing the total number of errors by the total number of trials. To reduce the implementation complexity, we simulate the type-based version of the instantaneous SED code in Section VI-B, which has a log-linear complexity. The type-based version is an approximation of the exact instantaneous SED code since it uses an approximating instantaneous SED rule and an approximating decoding rule to mimic the instantaneous SED rule (43)–(44) and the MAP decoder (46), respectively, however, it performs remarkably close to the original instantaneous SED code. See Section VI-B for details. The slope of

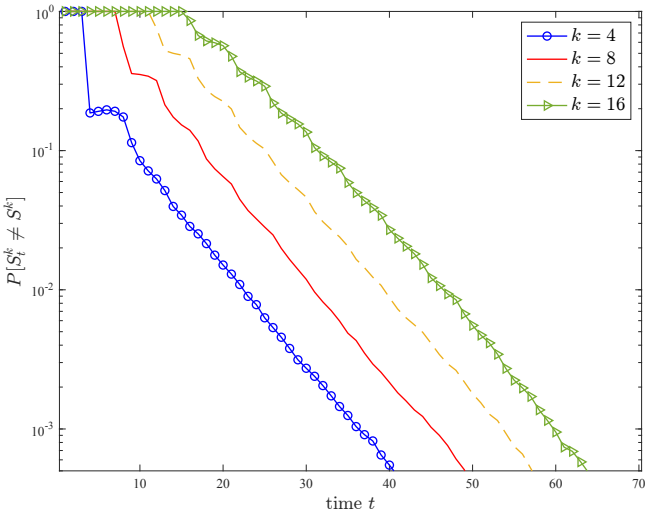


Fig. 5. The error probability $\mathbb{P}[\hat{S}_t^k \neq S^k]$ of decoding the first k symbols of a DSS at time t achieved by the type-based instantaneous SED code (Section VI-B). The DSS emits a Bernoulli$\left(\frac{1}{2}\right)$ bit at times $t = 1, 2, \ldots$. The channel is a BSC(0.05).

the curves corresponds to the anytime reliability α (18) of the instantaneous SED code. The anytime reliability for the source and the channel in Fig. 5 is approximately equal to $\alpha \approx 0.172$. The simulation results in Fig. 5 align with our expectation: the error probability decays exponentially with delay $t - k$ (18), implying that the instantaneous SED code is an anytime code.

We proceed to display the unstable scalar linear system that can be stabilized by the instantaneous SED code. Consider the scalar linear system in Fig. 6, $Z_{t+1} = \lambda Z_t + U_t + W_t$, where $\lambda > 1$, Z_t is the real-valued state, U_t is the real-valued control signal, $|W_t| \leq \frac{\Omega}{2}$ is the bounded noise, and the initial state is $Z_1 \triangleq 0$. At time t, the observer uses the observed states Z^t as well as the past channel feedback Y^{t-1} to form a channel input X_t; the controller uses the received channel outputs Y^t to form a control signal U_t. For a $(q, \{t_n\}_{n=1}^{\infty})$ DSS that emits source symbols one by one at consecutive times $t_n = n$, $n = 1, 2, \ldots$, the anytime rate of a (κ, α) anytime code in Definition 4 is defined as $R_{\text{any}} = \log q$ nats per channel use, e.g., for the DSS in Fig. 5, $R_{\text{any}} = \log 2$; the α-anytime capacity $C_{\text{any}}(\alpha)$ is defined as the least upper bound on the anytime rates R_{any} such that the anytime reliability α is achievable [18]. For such a DSS, Sahai and Mitter [18, Lemma 4.1 in Sec. IV-D] showed that the unstable scalar linear system with bounded noise in Fig. 6 can be stabilized so that η-th moment $\mathbb{E}[|Z_t|^{\eta}]$ stays

finite at all times, provided that $C_{\text{any}}(\alpha) > \log\lambda$, $\alpha > \eta \log\lambda$. Thus, the instantaneous SED code can be used to stabilize the η-th moment of the unstable scalar linear system in Fig. 6 over a BSC(0.05) for any coefficient

$$\lambda < e^{\min\{R_{\text{any}}, \frac{\alpha}{\eta}\}} \tag{47}$$

$$= \min\left\{2, e^{\frac{0.172}{\eta}}\right\}. \tag{48}$$

E.g., if $\eta = 2$, then $\lambda < 1.09$. In comparison, the theoretical results in [22, Corollary 1, Fig. 2] with $n = 1$ show that, for the control over a BSC(0.05) in Fig. 6, Lalitha et al.'s anytime code is only guaranteed to stabilize the η-th moment of a linear system with $\lambda = 1$.

The control scheme [18, Sec. IV] that stabilizes the system in Fig. 6 employs an anytime code and operates as follows. At each time t, the observer computes an R_{any}-nat virtual control signal $\bar{U}_t$ and acts as an anytime encoder to transmit $\bar{U}_t$ as the t-th symbol of a DSS over a noisy channel with feedback. Here, $\bar{U}_t$ controls a virtual state $\bar{Z}_{t+1} = \lambda\bar{Z}_t + W_t + \bar{U}_t$, and is equal to the negative of the R_{any}-nat quantization of $\lambda\bar{Z}_t$. It ensures the boundedness of $\bar{Z}_{t+1}$. Upon receiving the channel output, the controller acts as an anytime decoder to refresh its estimate $\hat{\bar{U}}_t^t$ of $\bar{U}^t$ and forms a control signal U_t that compensates the past estimation errors of the virtual control signals as if the plant $\{Z_s\}_{s=1}^{t+1}$ was controlled by $\hat{\bar{U}}_t^t$ heretofore. As a result of applying U_t, the actual state Z_{t+1} is forced close to the bounded virtual state $\bar{Z}_{t+1}$ with the difference $|Z_{t+1} - \bar{Z}_{t+1}|$ governed by the difference between $\bar{U}_t^t$ and $\hat{\bar{U}}_t^t$. The exponentially decaying with $t-k$ error probability of decoding $\bar{U}_t^k$ achieved by the anytime code together with the bounded $\bar{Z}_{t+1}$ ensures a finite $\mathbb{E}[|Z_{t+1}|^\eta]$. In fact, the full feedback channel in Fig. 6 can be replaced by a channel that only feeds the control signal from the controller to the observer, since Z_t, Z_{t-1}, U_{t-1} suffice to compute W_{t-1}.

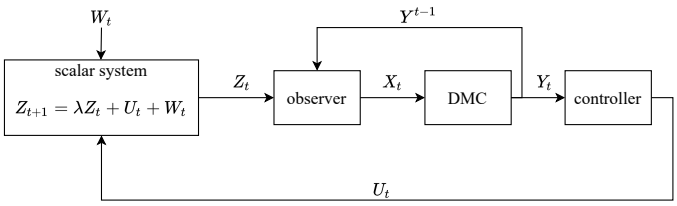


Fig. 6. A scalar linear system controlled over a noisy channel with noiseless feedback.

As verified by the simulations in Fig. 5, the instantaneous SED code achieves a positive anytime reliability, however, it is difficult to extend our analysis for $E(R)$ (38) to show that the instantaneous SED code satisfies (18) analytically. The submartingales in [5][8] used to compute the upper bound on the expected decoding time for a block encoding scheme to attain a target error probability fail to hold if the encoder keeps incorporating newly arrived symbols after time t_k. Therefore, we cannot directly use Lemma 5 in Appendix E-A to upper bound the expected decoding time, and different tools are needed for the analysis of anytime reliability.

C. Instantaneous SED code achieves $E(R)$

We first restrict the instantaneous SED code in Section V-A to transmit only the first k source symbols of a DSS, and we form a sequence of instantaneous SED codes indexed by the length of the symbol sequence k. We then show that the code sequence achieves the JSCC reliability function (38) for streaming over a symmetric binary-input DMC as $k \to \infty$.

We restrict the instantaneous SED code in Section V-A to transmit the first k symbols of a $(q, \{t_n\}_{n=1}^\infty)$ DSS as follows.

1) The alphabet $[q]^{N(t)}$ that contains all possible sequences that could have arrived by time t is replaced by the alphabet $[q]^{\min\{N(t),k\}}$ that stops evolving and reduces to $[q]^k$ after all k symbols arrive at time t_k. As a consequence, for $t \geq t_k + 1$ and all $i \in [q]^k$, the priors $\theta_i(y^{t-1})$ are equal to the corresponding posteriors $\rho_i(y^{t-1})$, the encoder and the decoder partition $[q]^k$ to obtain $\{\mathcal{G}_x(y^{t-1})\}_{x\in\{0,1\}}$, the encoder transmits the index of the group (29) that contains S^k, and only the posteriors $\rho_i(y^t)$ are updated.
2) The transmission is stopped and the MAP estimate (46) of S^k is produced at the stopping time

$$\eta_k \triangleq \min\left\{t\colon \max_{i\in[q]^k} P_{S^k|Y^t}(i|Y^t) \geq 1-\epsilon\right\}, \epsilon \in (0,1). \tag{49}$$

The MAP decoder (46) together with the stopping rule (49) ensures the error constraint in (15), since the MAP decoder (46) implies $\mathbb{P}[\hat{S}_{\eta_k}^k = S^k] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{\hat{S}_{\eta_k}^k\}}(S^k)\middle|Y^{\eta_k}\right]\right] = \mathbb{E}\left[\max_{i\in[q]^k} P_{S^k|Y^{\eta_k}}(i|Y^{\eta_k})\right]$, which is lower bounded by $1-\epsilon$ due to the stopping time (49).

Theorem 2. *Fix a non-degenerate symmetric binary-input DMC and a $(q, \{t_n\}_{n=1}^\infty)$ DSS satisfying assumption (b′) in Remark 1. The sequence of instantaneous SED codes for transmitting the first k symbols of the DSS achieves $E(R)$ (38) as $k \to \infty$.*

Proof. First, we observe that after the symbol arriving period $t \geq t_k + 1$, the instantaneous SED code reduces to the SED code [8, Sec. V-B] because the instantaneous SED rule (43)–(44) reduces to the SED rule [8, Eq. (50)] if all k source symbols are fully accessible (3) to the encoder. The SED code [8] achieves the JSCC reliability function (38) for transmitting a fully accessible source over a non-degenerate symmetric binary-input DMC (Appendix E-A).

Second, we observe that during the symbol arriving period $t = 1, 2 \ldots, t_k$, the instantaneous SED code corresponds to dropping the randomization step of the instantaneous encoding phase in Section III. This is because for a symmetric binary-input DMC, (43) implies $\pi_{x'}(y^{t-1}) \leq P_X^*(x') = \frac{1}{2}$, thus any partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\{0,1\}}$ that satisfies the instantaneous SED rule (43)–(44) also satisfies the partitioning rule in (24). Therefore, Remark 1 implies that the instantaneous SED code at times $t = 1, 2, \ldots, t_k$ satisfies the sufficient condition (41) under assumption (b′).

As we have discussed in the proof sketch of Theorem 1, a JSCC reliability function-achieving code with instantaneous encoding can be obtained by preceding a JSCC reliability

function-achieving code with block encoding by an instantaneous encoding phase that satisfies (41). The two observations above imply that the instantaneous SED code achieves $E(R)$ (38) in the setting of Theorem 2. ■

VI. Low-complexity codes with instantaneous encoding

We present the type-based algorithms for the instantaneous encoding phase in Section III, for the instantaneous SED code as an anytime code in Section V-A, and for the instantaneous SED code restricted to transmit k symbols only in Section V-C. The type-based instantaneous encoding phase is the exact phase in Section III, whereas the type-based instantaneous SED codes are approximations of the original codes in Sections V-A and V-C. All the type-based codes that we are about to see have a log-linear complexity $O(t \log t)$ in time t.

We assume that the source symbols of the DSS are equiprobably distributed, i.e., the source distribution (1) satisfies

$$P_{S_n|S^{n-1}}(a|b) = \frac{1}{q}, \tag{50}$$

for all $a \in [q]$, $b \in [q]^{n-1}$, $n = 1, 2, \dots$

In our type-based codes, the evolving source alphabet is judiciously divided into disjoint sets that we call *types*, so that the source sequences in each type share the same prior and the same posterior. Here, the same prior is guaranteed by the equiprobably distributed symbols (50), and the same posterior is guaranteed by moving a whole type to a group during the group partitioning process (see step (iii) below). As a consequence of classifying source sequences into types, the prior update, the group partitioning, and the posterior update can be implemented in terms of types rather than individual source sequences, which results in an exponential reduction of complexity.

We denote by $\mathcal{S}_1, \mathcal{S}_2, \dots$ a sequence of types. We slightly abuse the notation to denote by $\theta_{\mathcal{S}_j}(Y^{t-1})$ and $\rho_{\mathcal{S}_j}(Y^t)$ the prior and the posterior of a single source sequence in type $\mathcal{S}_j$ at time t rather than the prior and the posterior of the whole type. We fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS that satisfies (50) and fix a DMC with a single-letter transition probability $P_{Y|X} \colon \mathcal{X} \to \mathcal{Y}$.

A. Type-based instantaneous encoding phase

The type-based instantaneous encoding phase operates at times $t = 1, 2, \dots, t_k$, where k is the number of source symbols of a DSS that we aim to transmit.

(i) *Type update:* At each time t, the algorithm first updates the types. At $t = 1$, the algorithm is initialized with one type $\mathcal{S}_1 \triangleq [q]^{N(1)}$. At $t = t_n$, $n = 2, \dots, k$, the algorithm updates all the existing types by appending every sequence in $[q]^{N(t)-N(t-1)}$ to every sequence in the type. After the update, the length of the source sequences in each type is equal to $N(t)$; the cardinality of each type is multiplied by $q^{N(t)-N(t-1)}$; the total number of types remains unchanged. At $t \neq t_n$, $n = 1, 2, \dots, k$, the algorithm does not update the types.

(ii) *Prior update:* Once the types are updated, the algorithm proceeds to update the prior of the source sequences in each existing type. The prior $\theta_{\mathcal{S}_j}(y^{t-1})$, $j = 1, 2, \dots$ of the source sequences in type $\mathcal{S}_j$ is fully determined by (22) with $\theta_i(y^{t-1}) \leftarrow \theta_{\mathcal{S}_j}(y^{t-1})$, $P_{S^{N(t)}|S^{N(t-1)}}(\cdot|\cdot) \leftarrow \left(\frac{1}{q}\right)^{N(t)-N(t-1)}$, and $\rho_{i^{N(t-1)}}(y^{t-1}) \leftarrow \rho_{\mathcal{S}_j}(y^{t-1})$. If the types are not updated, the priors are equal to the posteriors, i.e., $\theta_{\mathcal{S}_j}(y^{t-1}) \leftarrow \rho_{\mathcal{S}_j}(y^{t-1})$, $j = 1, 2, \dots$

(iii) *Group partitioning:* Using all the existing types and their priors, the algorithm determines a partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ that satisfies the partitioning rule (24) via a *type-based greedy heuristic algorithm*. It operates as follows. It initializes all the groups $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ by empty sets and initializes the group priors $\{\pi_x(y^{t-1})\}_{x\in\mathcal{X}}$ by zeros. It forms a queue by sorting all the existing types according to priors $\theta_{\mathcal{S}_j}(y^{t-1})$, $j = 1, 2, \dots$ in a descending order. It moves the types in the queue one by one to one of the groups $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$. Before each move, it first determines a group $\mathcal{G}_{x^*}(y^{t-1})$ whose current prior $\pi_{x^*}(y^{t-1})$ has the largest gap to the corresponding capacity-achieving probability $P_X^*(x^*)$,

$$x^* \triangleq \arg\max_{x\in\mathcal{X}} P_X^*(x) - \pi_x(y^{t-1}). \tag{51}$$

Suppose the first type in the sorted queue, i.e., the type whose sequences have the largest prior, is $\mathcal{S}_j$. It then proceeds to determine the number of sequences that are moved from type $\mathcal{S}_j$ to group $\mathcal{G}_{x^*}(y^{t-1})$ by calculating

$$n \triangleq \left\lceil \frac{P_X^*(x^*) - \pi_{x^*}(y^{t-1})}{\theta_{\mathcal{S}_j}(y^{t-1})} \right\rceil. \tag{52}$$

If $n \geq |\mathcal{S}_j|$, then it moves the whole type $\mathcal{S}_j$ to group $\mathcal{G}_{x^*}(y^{t-1})$; otherwise, it splits $\mathcal{S}_j$ into two types by keeping the smallest or the largest n consecutive[1] (in lexicographic order) sequences in $\mathcal{S}_j$ and transferring the rest into a new type, and it moves type $\mathcal{S}_j$ to group $\mathcal{G}_{x^*}(y^{t-1})$ and moves the new type to the beginning of the queue. It updates the prior $\pi_{x^*}(y^{t-1})$ after each move.

(iv) *Randomization:* The type-based instantaneous encoding algorithm implements the randomization in (25)–(30) with respect to a partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$.

(v) *Posterior update:* Upon receiving the channel output $Y_t = y_t$, the algorithm updates the posterior of the source sequences in each existing type. The posterior $\rho_{\mathcal{S}_j}(y^t)$, $j = 1, 2, \dots$ of the source sequences in type $\mathcal{S}_j$ is fully determined by (32) with $\rho_i(y^t) \leftarrow \rho_{\mathcal{S}_j}(y^t)$, $\theta_i(y^{t-1}) \leftarrow \theta_{\mathcal{S}_j}(y^{t-1})$.

Using (52) and Appendix A, we conclude that the type-based greedy heuristic algorithm ensures (24).

We show that the complexity of the type-based instantaneous encoding phase is log-linear $O(t \log t)$ at times $t = 1, 2, \dots, t_k$. We first show that the number of types grows linearly, i.e., $O(t)$. Since the type update in step (i) does not add new types, the number of types increases only due to the split of types during group partitioning in step (iii). At most $|\mathcal{X}|$ types are split at each time. This is because the ceiling in (52) ensures that the group that receives the n sequences from a split type will have a group prior no smaller than the

[1]This step ensures that all sequences in a type are consecutive. Thus, as we will discuss in the last paragraph in Section VI-A, it is sufficient to store two sequences, one with the smallest and one with the largest lexicographic orders, in a type to fully specify that type.

corresponding capacity-achieving probability, thus the group will no longer be the solution to the maximization problem (51) and will not cause the split of other types. We proceed to analyze the complexity of each step of the algorithm. Step (i) (type update) has a linear complexity in the number of types, i.e., $O(t)$. This is because the methods of updating and splitting a type in steps (i) and (iii) ensure that the sequences in any type are consecutive, thus it is sufficient to store the starting and the ending sequences in each type to fully specify all the sequences in that type. As a result, updating a type is equivalent to updating the starting and the ending sequences of that type. Step (ii) (prior update) and step (v) (posterior update) have a linear complexity in the number of types, i.e., $O(t)$. Step (iii) (group partitioning) has a log-linear complexity in the number of types due to type sorting, i.e., $O(t \log t)$. This is because the average complexity of sorting a sequence of numbers is log-linear in the size of the sequence [34]. Step (iv) (randomization) has complexity $O(1)$ due to determining $\{p_{\overline{x}\to\underline{x}}\}_{\overline{x}\in\overline{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$ in (27)–(28).

B. Type-based instantaneous SED codes

We present type-based codes for the anytime instan SED code in Section V-A and for the instantaneous SE restricted to transmit k symbols in Section V-C, respe

The type-based anytime instantaneous SED code symmetric binary-input DMC operates at times $t = 1,$

(i′) *Type update:* At each time t, the algorithm update as in step (i) with $k = \infty$.

(ii′) *Prior update:* The algorithm updates the prior source sequences in each existing type as in step (i $k = \infty$.

(iii′) *Group partitioning:* Using all the existing and their priors, the algorithm determines a p $\{\mathcal{G}_x(y^{t-1})\}_{x\in\{0,1\}}$ using an *approximating* instantaneo rule that mimics the exact rule in (43)–(44) as foll forms a queue by sorting all the existing types ac to priors $\theta_{\mathcal{S}_j}(y^{t-1})$, $j = 1, 2, \ldots$ in a descending o moves the types in the queue one by one to $\mathcal{G}_0(y^{t-}$ $\pi_0(y^{t-1}) \geq P_X^*(0) = 0.5$ for the first time. Suppose type moved to $\mathcal{G}_0(y^{t-1})$ is $\mathcal{S}_j$. To make the group prio even, it then calculates the number of sequences n to be away from $\mathcal{S}_j$ as

$$
\begin{aligned}
n &\triangleq \arg\min_{n\in\{\underline{n},\bar{n}\}} \Big| \big(\pi_0(y^{t-1}) - n\theta_{\mathcal{S}_j}(y^{t-1})\big) \\
&\quad - \big(\pi_1(y^{t-1}) + n\theta_{\mathcal{S}_j}(y^{t-1})\big) \Big|, \\
\underline{n} &\triangleq \left\lfloor \frac{\pi_0(y^{t-1}) - 0.5}{\theta_{\mathcal{S}_j}(y^{t-1})} \right\rfloor, \\
\bar{n} &\triangleq \left\lceil \frac{\pi_0(y^{t-1}) - 0.5}{\theta_{\mathcal{S}_j}(y^{t-1})} \right\rceil.
\end{aligned}
$$

It splits $\mathcal{S}_j$ into two types by transferring the first or n (53a) lexicographically ordered sequences in $\mathcal{S}_j$ to type. It moves the new type and all the remaining type queue to $\mathcal{G}_1(y^{t-1})$.

(iv′) The randomization step in (iv) is dropped.

(v′) *Posterior update:* The algorithm updates the posteriors of the source sequences in each existing type. The posterior $\rho_{\mathcal{S}_j}(y^t)$, $j = 1, 2, \ldots$, is fully determined by (45) with $\rho_i(y^t) \leftarrow \rho_{\mathcal{S}_j}(y^t)$, $\theta_i(y^{t-1}) \leftarrow \theta_{\mathcal{S}_j}(y^{t-1})$.

(vi′) *Decoding at time t:* To decode the first k symbols at time t, where k can be any integer that satisfies $t_k \leq t$, the algorithm first finds the type whose source sequences have the largest posterior. Then, it searches for the most probable length-k prefix in that type by relying on the fact that sequences in the same type share the same posterior; thus, the prefix shared by the maximum number of sequences is the most probable one. Namely, the algorithm extracts the length-k prefixes of the starting and the ending sequences, denoted by i^k_{start} and i^k_{end}, respectively. If $i^k_{\text{start}} = i^k_{\text{end}}$ (Fig. 7-a), then the decoder outputs $\hat{S}^k_t = i^k_{\text{start}}$. If i^k_{start} and i^k_{end} are not lexicographically consecutive (Fig. 7-b), then the decoder outputs a length-k prefix in between the two prefixes. If i^k_{start} and i^k_{end} are lexicographically consecutive (Fig. 7-c), then the algorithm computes the number of sequences in the type that have prefix i^k_{start} and the number of sequences in the type that have prefix i^k_{end} using the last $N(t)-k$ symbols of the starting

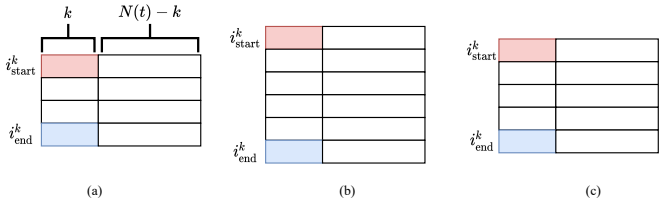


Fig. 7. Tables (a), (b), (c) represent three types at time t. Each row represents a source sequence in the type. The first row and the last row in each type represent the starting sequence and the ending sequence in that type, respectively. The first column represents the length-k prefix of sequences in the type. The source sequences in a type are lexicographically consecutive due to the methods of updating and splitting a type in steps (i′) and (iii′). In (a), since $i^k_{\text{start}} = i^k_{\text{end}} = 000$, the most probable sequence is 000. In (b), since $i^k_{\text{start}} = 000$ and $i^k_{\text{end}} = 010$ are not lexicographically consecutive, the most probable prefix is 001. In (c), since $i^k_{\text{start}} = 010$ and $i^k_{\text{end}} = 011$ are lexicographically consecutive, the number of sequences with prefix i^k_{start} can be computed by subtracting 1111110, the last $N(t) - k$ symbols of the starting sequence, from 1111111 and adding 1; the number of sequences with prefix i^k_{end} is equal to the last $N(t)-k$ symbols of the ending sequence plus 1. Since (c) contains more sequences with prefix 011, this is the most probable prefix.

We proceed to show that the complexity of the type-based anytime instantaneous SED code is $O(t \log t)$. Similar to the type-based instantaneous encoding phase in Section VI-A, the number of types grows linearly with time t since the number of types increases only if a type is split in step (iii′), and at most 1 type is split at each time t. The complexities of steps (i′), (ii′), (v′) are all linear in the number of types $O(t)$ due to the discussion at the end of Section VI-A. The complexity of step (iii′) is log-linear in the number of types $O(t \log t)$ due to sorting the types. Since the sequences in a type are lexicographically consecutive due to the updating and the splitting methods in steps (i′) and (iii′), it suffices to use the starting and the ending sequences in a type to determine the most probable prefix in that type. Thus, the complexity of

step (vi′) is linear in the number of types due to searching for the type whose sequences have the largest posterior.

Restricting the type-based anytime instantaneous SED code described above to transmit only the first k symbols of a DSS is equivalent to implementing steps (i), (ii), (iii′), (v) one by one, and performing decoding as follows.

(vi″) *Decoding and stopping:* If there exists a type $\mathcal{S}_j$ that satisfies $\rho_{\mathcal{S}_j}(y^t) \geq 1-\epsilon$ and contains a source sequence of length k, then the decoder stops and outputs a sequence in that type as the estimate $\hat{S}^k_{\eta_k}$.

The complexity of the type-based instantaneous SED code for transmitting k symbols remains log-linear, $O(t \log t)$, since the complexity of step (vi″) is $O(t)$ due to searching for the type that satisfies the requirements.

While the type-based instantaneous encoding phase in Section VI-A is the exact algorithm of the instantaneous encoding phase in Section III, the type-based anytime instantaneous SED code and the type-based instantaneous SED code for transmitting k symbols are *approximations* of the original algorithms in Sections V-A and V-C due to two reasons below:

First, in step (iii′) (group partitioning), we use the approximating instantaneous SED rule to mimic the exact rule in (43)–(44). The minimum of the objective function in (53a) is equal to the difference $|\pi_0(y^{t-1}) - \pi_1(y^{t-1})|$ between the group priors of the partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\{0,1\}}$ obtained by the approximating rule in step (iii′). The difference is upper bounded as (Appendix H)

$$|\pi_0(y^{t-1}) - \pi_1(y^{t-1})| \leq \theta_{\mathcal{S}_j}(y^{t-1}), \tag{54}$$

where $\mathcal{S}_j$ is the last type moved to $\mathcal{G}_0(y^{t-1})$ so that its group prior exceeds 0.5 for the first time. If $\pi_0(y^{t-1}) \geq \pi_1(y^{t-1})$, (54) recovers (44) since $\theta_{\mathcal{S}_j}(y^{t-1})$ is the smallest prior in $\mathcal{G}_0(y^{t-1})$, thus the approximating instantaneous SED rule recovers the exact rule. If $\pi_0(y^{t-1}) < \pi_1(y^{t-1})$, $\theta_{\mathcal{S}_j}(y^{t-1})$ on the right side of (54) is the largest prior in $\mathcal{G}_1(y^{t-1})$, violating the right side of (44).

We use the approximating algorithm of the instantaneous SED rule (43)–(44) since it is unclear how to implement the exact instantaneous SED rule with polynomial complexity. In the worst case, the complexity of the latter is as high as double exponential $O\left(2^{q^{N(t)}}\right)$ due to solving a minimization problem via an exhaustive search [8, Algorithm 1]. An exact algorithm for the SED rule with exponential complexity in the source length is given by [8, Algorithm 2].

Second, in step (vi′) (decoding at time t) of the type-based anytime instantaneous SED code, we only find the most likely length-k prefix in the type that achieves $\max_j \rho_{\mathcal{S}_j}(y^t)$, yet it is possible that this prefix is not the one that has the globally largest posterior (46). To search for the most probable length-k prefix, one needs to compute the posteriors for all q^k prefixes of length k using $O(t)$ types, resulting in an exponential complexity $O(q^k t)$ in the length of the prefix k, whereas the complexity of step (vi′) is only $O(t)$ independent of k.

Although the type-based instantaneous SED code is an approximation, as we are about to see in Fig. 9 Section VII, it is almost as good as the exact code.

VII. Simulations

Fig. 8 shows the performance of our new instantaneous encoding schemes. Namely, we fix an error probability $\epsilon = 10^{-6}$, a BSC(0.05), and a DSS that emits i.i.d. Bernoulli$\left(\frac{1}{2}\right)$ bits one by one at consecutive times. We display the rate $R_k \triangleq \frac{k}{\mathbb{E}[\eta_k]}$ as a function of source length k empirically attained by the instantaneous encoding phase followed by the SED code [8, Algorithm 2] and the instantaneous SED code in Section V-C, and we compare achievable rates to that of the SED code for a fully accessible source, as well as to that of a buffer-then-transmit code that implements the SED code during the block encoding phase. We also plot the rate R_k obtained from the reliability function approximation (17):

$$E(R_k) \simeq \frac{R_k}{k} \log \frac{1}{\epsilon}. \tag{55}$$

Due to the discussions in the proof sketch of Theorem 1, the instantaneous encoding phase followed either by the MaxEJS code or by the SED code achieves the JSCC reliability function for streaming (38). For the simulations in Fig. 8, we choose the SED code since it applies to a BSC and its complexity, exponential in the source length, is lower than the double-exponential complexity of the MaxEJS code. To obtain the empirical rate in Fig. 8, at each source length k, we run the experiments for every code for 10^5 trials, and we obtain the denominator $\mathbb{E}[\eta_k]$ of the empirical rate by averaging the stopping times in all the experiments.

We observe from Fig. 8 that the achievable rate of the instantaneous encoding phase followed by the SED code is significantly larger than that of the buffer-then-transmit code, and approaches that of the SED code as k increases even though the SED encoder knows the entire source sequence before the transmission. The instantaneous SED code demonstrates an even better performance: it is essentially as good as the SED code. The rate obtained from reliability function approximation (55) is remarkably close to the empirical achievable rates of our codes with instantaneous encoding even for very short source length $k \simeq 16$. For example, at $k = 16$, the rate obtained from approximation (55) is 0.58 (symbols per channel use) and the empirical rate of the instantaneous SED code is 0.59 (symbols per channel use). This means that the reliability function (17), an inherently asymptotic notion, accurately reflects the delay-reliability tradeoffs attained by the JSCC reliability function-achieving codes in the ultra-short blocklength regime. The achievable rate corresponding to the buffer-then-transmit code is limited by (39).

Fig. 9 shows the performance of the type-based instantaneous SED code. We fix an error probability $\epsilon = 10^{-6}$ (15), a BSC(p) with $p = 0.05, 0.03, 0.01$, and a DSS that emits i.i.d. Bernoulli$\left(\frac{1}{2}\right)$ bits one by one at consecutive times. We plot rate $R_k = \frac{k}{\mathbb{E}[\eta_k]}$ as a function of source length k empirically achieved by the instantaneous SED code in Section V-C and its corresponding type-based code in Section VI-B, as well as the rate obtained from the reliability function approximation (55). At each source length k, we run the experiments using the same method as in Fig. 8. The rate gap between the instantaneous SED code and the type-based instantaneous SED code is negligible, meaning that the type-based instantaneous SED

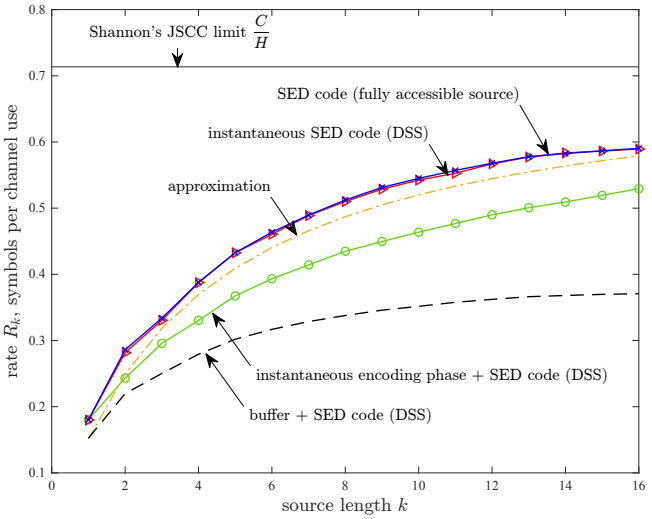


Fig. 8. Rate R_k (symbols per channel use) vs. source length k. The error probability is constrained by $\epsilon = 10^{-6}$ (15). The DMC is a BSC(0.05). Naghshvar et al.'s SED code [8, Algorithm 2] operates on a fully accessible block S^k of independent Bernoulli$\left(\frac{1}{2}\right)$ bits. The instantaneous encoding phase followed by the SED code, the instantaneous SED code, and the buffer-then-transmit code operate on k i.i.d. Bernoulli$\left(\frac{1}{2}\right)$ source bits emitted one by one at times $t = 1, 2, \ldots, k$. The curves are displayed for the range of k's where the complexities of the SED code and the instantaneous SED code are not prohibitive.

code with only log-linear complexity is a good approximation to the exact code in Section V-C. Furthermore, it is interesting to see that even though the DSS has symbol arriving rate $f = 1$ symbol per channel use, which is far less than that required in assumption (b'), the achievable rates of the instantaneous SED code stay very close to the rates obtained from the reliability function approximation. This suggests that assumption (b') on the symbol arriving rate, sufficient for the instantaneous SED code to achieve $E(R)$, could be conservative.

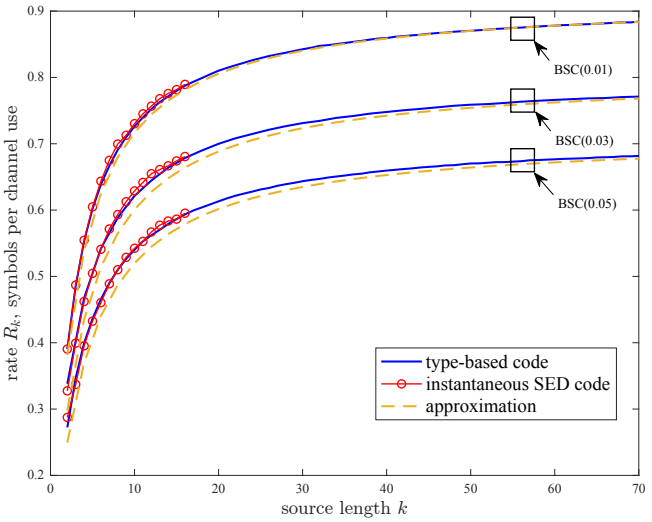


Fig. 9. Rate R_k (symbols per channel use) vs. source length k. The error probability is constrained by $\epsilon = 10^{-6}$ (15). The type-base instantaneous SED code in Section VI-B and the instantaneous SED code in Section V-C operate on k i.i.d. Bernoulli$\left(\frac{1}{2}\right)$ source bits emitted one by one at times $t = 1, 2, \ldots, k$.

VIII. Streaming over a degenerate DMC with zero error

In this section, we propose a code with instantaneous encoding for a degenerate DMC (9) that achieves zero decoding error at any rate asymptotically below $\frac{C}{H}$. Here, our code does not exactly follow Definition 3 since it generalizes the code with instantaneous encoding in Definition 3 by allowing *common randomness* $U \in \mathcal{U}$, which is a random variable that is revealed to the encoder and the decoder before the transmission. With common randomness U, the encoder f_t (12) can use U to form X_t, and the decoder g_t (13) can use U to decide the stopping time η_k and the estimate $\hat{S}^k_{\eta_k}$. We refer to such a code as a $\langle k, R, \epsilon\rangle$ *code with instantaneous encoding and common randomness* if it achieves rate R (14) and error probability ϵ (15) for transmitting k symbols of a DSS. Common randomness is widely used to specify a random codebook in the scenario where multiple constraints on expectations of quantities that depend on the codebook must be satisfied simultaneously and where Shannon's probabilistic method is not sufficient to claim the existence of a deterministic codebook satisfying all constraints, e.g., [6][13][17][30]. Since for a fixed k, we seek to satisfy two constraints, on the rate and on the error probability, the cardinality of $\mathcal{U}$ can be restricted as $|\mathcal{U}| \le 2$ (Appendix I).

Theorem 3, stated next, establishes the existence of zero-error codes for the transmission over a degenerate DMC at any rate asymptotically below $\frac{C}{H}$.

Theorem 3. *Fix a degenerate DMC with capacity C (10), fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with entropy entropy rate $H > 0$ (2) satisfying assumptions* (a)–(b) *in Theorem 1, and fix any $R < \frac{C}{H}$. There exists a sequence of $\langle k, R_k, 0\rangle$ codes with instantaneous encoding and common randomness that satisfies*

$$\lim_{k\to\infty} R_k = R. \tag{56}$$

Proof sketch. Our zero-error code for degenerate DMCs extends Burnashev's scheme [5, Sec. 6] to JSCC and to streaming sources: to achieve Shannon's JSCC limit $\frac{C}{H}$, a Shannon limit-achieving code is used in the first communication phase to compress the source; to transmit streaming sources, we combine an instantaneous encoding phase that satisfies (41) with a Shannon limit-achieving block encoding scheme to form a Shannon limit-achieving instantaneous encoding scheme. To achieve zero error, we employ confirmation phases similar to those in Burnashev's scheme [5]. We say that a $\langle k, R, \epsilon_k\rangle$ code with instantaneous encoding and common randomness achieves Shannon's JSCC limit $\frac{C}{H}$ if for all $R < \frac{C}{H}$, a sequence of such codes indexed by k satisfies $\epsilon_k \to 0$ as $k \to \infty$. Our zero-error code includes such Shannon limit-achieving codes as a building block. Note that in contrast to the discussions in Sections IV–V focused on the exponential rate of decay of ϵ_k to 0 (17) over non-degenerate DMCs, here merely having ϵ_k decrease to 0 suffices. The following argument shows the existence of such codes for the class of channels that includes both non-degenerate and degenerate DMCs.

We employ the joint source-channel code in [30, Theorem 2] due to the simplicity of the error analysis it affords. The code in [30, Theorem 2] is a $\langle k, R, \epsilon_k\rangle$ Shannon limit-achieving

code with block encoding and common randomness because its expected decoding time to attain error probability ϵ is upper bounded as (87) in Appendix E-A with $C_1 \leftarrow C$ [30, Eq. (16)], implying that it achieves a positive error exponent that is equal to (38) with $C_1 \leftarrow C$ for all $R < \frac{C}{H}$. The block encoding scheme in [30, Theorem 2] is a stop-feedback code, meaning that the encoder uses channel feedback only to decide whether to stop the transmission but not to form channel inputs. If the DSS has an infinite symbol arriving rate $f = \infty$ (5), a buffer-then-transmit code using the block encoding scheme in [30, Theorem 2] achieves the Shannon limit since it achieves the same error exponent as the code in [30, Theorem 2]. To see this, one can simply invoke (87) in Lemma 5 with $C_1 \leftarrow C$ and follow the proofs in Appendix E-B. By the same token, if the DSS has a finite symbol arriving rate $f < \infty$ (5), a code implementing an instantaneous encoding phase that satisfies (41) followed by the block encoding scheme in [30, Theorem 2] for k source symbols with prior $P_{S^k|Y^{t_k}}$ achieves the Shannon limit with the same error exponent as the code in [30, Theorem 2].

Our zero-error code with instantaneous encoding and common randomness for transmitting k symbols over a degenerate DMC operates as follows (details in Appendix J-A). Similar to [5]–[7], [17], our code is divided into blocks. Each block contains a communication phase and a confirmation phase. In the first block, the communication phase uses a $\langle k, R, \epsilon_k \rangle$ Shannon limit-achieving code with instantaneous encoding and common randomness. The confirmation phase selects two symbols x (9a) and x' (9b) as the channel inputs (i.e., x' never leads to channel output y); the encoder repeatedly transmits x if the decoder's estimate of the source sequence at the end of the communication phase is correct, and transmits x' otherwise. If the decoder receives a y in the confirmation phase, meaning that the encoder communicated its knowledge that the decoder's estimate is correct with zero error, then it outputs its estimate, otherwise, the next block is transmitted. The ℓ-th block, $\ell \geq 2$, differs from the first block in that it does not compress the source to avoid errors due to an atypical source realization and in that it uses random coding whereas the first block can employ any Shannon-limit achieving code.

We proceed to discuss the error and the rate achievable by our code (details in Appendix J-B).

Our code achieves zero error by employing confirmation phases that rely on the degenerate nature of the channel: receiving a y in the confirmation phase guarantees a correct estimate.

Our code achieves all rates asymptotically below $\frac{C}{H}$ because 1) the first block employs a Shannon limit-achieving code in the communication phase, 2) the length of the confirmation phase is made negligible compared to the length of the communication phase as the source length $k \to \infty$, meaning that the length of the first block asymptotically equals the length of its communication phase, and 3) subsequent blocks asymptotically do not incur a penalty on rate, as we discuss next. Since the length of each block is comparable to the length of the first block, it is enough to show that the expected number of blocks T_k transmitted after the first block converges to zero. The refreshing of random codebook for all uncompressed source sequences in every block after the first block ensures that the channel output vectors in these subsequent blocks are i.i.d. and are independent of the channel outputs in the first block. Conditioned on $T_k > 0$, the i.i.d. vectors give rise to a geometric distribution of T_k with failure probability converging to 0, which implies $\mathbb{E}[T_k] \to 0$ as $k \to \infty$. ■

A stop-feedback code with block encoding that retransmits blocks with the overall rate asymptotically equal to the rate of the first block is also used by Forney [43, p. 213] for deriving a lower bound on the reliability function of a DMC.

IX. Conclusion

In this paper, we have derived the reliability function for transmitting a discrete streaming source over a DMC with feedback using variable-length joint source-channel coding with instantaneous encoding under regularity conditions (Theorem 1). Since a classical fully accessible DS is a special DSS (see (3)), Theorem 1 extends Burnashev's reliability function to the classical JSCC scenario with block encoding, as well as to a streaming scenario. The most surprising observation in this paper is that the JSCC reliability function for a streaming source is equal to that for a fully accessible source. A naive buffer-then-transmit code that idles the transmission during the symbol arriving period does not achieve the JSCC reliability function for a non-trivial streaming source (see (39)). To achieve the JSCC reliability function for such sources, we have proposed a novel instantaneous encoding phase (Section III). We have shown that preceding a JSCC reliability function-achieving code with block encoding, e.g., the MaxEJS code or the SED code [8], by our instantaneous encoding phase (Section III) will make it overcome the detrimental effect due to the streaming nature of the source and make it achieve the same error exponent as if the encoder knew the entire source sequence before the transmission. The instantaneous encoding phase (Section III) achieves the JSCC reliability function because it satisfies the sufficient condition (41) on the statistics of the encoder outputs during the symbol arriving period, for example, the instantaneous encoding phase continues to achieve the sufficient condition (41) after it drops the randomization step, but at a cost of increasing the threshold for the symbol arriving rate (Remark 1). While our JSCC reliability function-achieving codes are designed to transmit k symbols of a streaming source and stop, we have also designed an instantaneous SED code (Section V) that can choose the decoding time and the number of symbols to decode on the fly. It empirically attains a positive anytime reliability (Fig. 5), thus it can be used to stabilize an unstable scalar linear system with a bounded noise over a noisy channel. A sequence of such codes indexed by the source length to decode also achieves the JSCC reliability function for streaming in the limit of large source length (Theorem 3). For practical implementations, we have designed type-based log-linear complexity algorithms for the instantaneous encoding phase and the instantaneous SED code (Section VI) that apply to streaming sources with equiprobable symbols. While the codes that achieve the JSCC reliability function are designed for non-degenerate DMCs, we have also designed zero-error codes with instantaneous

encoding for degenerate DMCs (Section VIII), extending Burnashev's zero-error channel code to the JSCC and to the streaming scenarios.

Future research directions include the following. First, it would be interesting to find the JSCC reliability function for a wider class of channels and streaming sources (e.g., sources without a valid f, with a small f, or with an infinite alphabet). Second, it would be interesting to extend $E(R)$ to lossy JSCC by inserting an appropriate instantaneous encoding phase before a lossy JSCC reliability function-achieving block encoding scheme. Third, it is practically important to design instantaneous encoding schemes using limited or noisy feedback. Finally, it would be interesting to find codes that can learn the streaming source distribution on the fly.

ACKNOWLEDGEMENT

Insightful comments from Dr. Oron Sabag are gratefully acknowledged.

APPENDIX A
A PARTITION THAT SATISFIES (24)

For any $t = 1, 2, \dots$ and any $y^{t-1} \in \mathcal{Y}^{t-1}$, we show that the greedy heuristic algorithm [35] yields a partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ that satisfies the partitioning rule (24).

The greedy heuristic algorithm operates as follows. At time t, it initializes all the groups $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ by empty sets and initializes all the group priors $\{\pi_x(y^{t-1})\}_{x\in\mathcal{X}}$ by zeros. It sorts all the source sequences in $[q]^{N(t)}$ according to their priors $\theta_i(y^{t-1})$, $i \in [q]^{N(t)}$ in a descending manner. Starting from the sequence with the largest prior, it moves the sequence in the sorted list to the group $\mathcal{G}_{x^*}(y^{t-1})$ whose current group prior has the largest gap to the corresponding capacity-achieving probability, i.e.,

$$x^* \triangleq \arg\max_{x\in\mathcal{X}} P_X^*(x) - \pi_x(y^{t-1}). \tag{57}$$

The group prior $\pi_{x^*}(y^{t-1})$ is updated after each move. The partitioning process repeats until all the source sequences have been classified.

We show that the resulting partition $\{\mathcal{G}_x(y^{t-1})\}_{x\in\mathcal{X}}$ satisfies (24). We first notice that the maximization problem on the right side of (57) must be strictly larger than zero before all source sequences have been classified. If moving sequence i to group $\mathcal{G}_{x^*}(y^{t-1})$ leads to

$$\pi_{x^*}(y^{t-1}) - P_X^*(x^*) < 0, \tag{58}$$

then (24) is obviously satisfied. If moving sequence i to group $\mathcal{G}_{x^*}(y^{t-1})$ leads to

$$\pi_{x^*}(y^{t-1}) - P_X^*(x^*) \geq 0, \tag{59}$$

then (24) is satisfied since (58) holds before the move, and sequence i has the smallest prior in $\mathcal{G}_{x^*}(y^{t-1})$ after the move. Furthermore, if the group $\mathcal{G}_{x^*}(y^{t-1})$ satisfies (59), it will no longer be the solution to the maximization problem in (57) and thus will no longer accept new sequences. This means that (24) holds for all $x \in \mathcal{X}$ at the end of the greedy heuristic partitioning.

Algorithm 1: Determine $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$ that satisfies (27)–(28)

Data: $\pi_x(y^{t-1})_{x\in\mathcal{X}}$, $\bar{\mathcal{X}}(y^{t-1})$, $\underline{\mathcal{X}}(y^{t-1})$
Result: $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$

1 $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})} = 0$;
2 $\hat{\pi}_x(y^{t-1}) \leftarrow \pi_x(y^{t-1}), \forall x \in \mathcal{X}$;
3 **for** $\underline{x} \in \underline{\mathcal{X}}(y^{t-1})$ **do**
4 **while** $\hat{\pi}_{\underline{x}}(y^{t-1}) < P_X^*(\underline{x})$ **do**
5 $\bar{x} \leftarrow \bar{\mathcal{X}}(1)$;
6 $p_{\bar{x}\to\underline{x}} \leftarrow \min\{\hat{\pi}_{\bar{x}}(y^{t-1}) - P_X^*(\bar{x}), P_X^*(\underline{x}) - \hat{\pi}_{\underline{x}}(y^{t-1})\}$;
7 $\hat{\pi}_{\bar{x}}(y^{t-1}) \leftarrow \hat{\pi}_{\bar{x}}(y^{t-1}) - p_{\bar{x}\to\underline{x}}$;
8 $\hat{\pi}_{\underline{x}}(y^{t-1}) \leftarrow \hat{\pi}_{\underline{x}}(y^{t-1}) + p_{\underline{x}\to\underline{x}}$;
9 **if** $\hat{\pi}_{\underline{x}}(y^{t-1}) = P_X^*(\underline{x})$ **then**
10 $\underline{\mathcal{X}}(y^{t-1}) \leftarrow \underline{\mathcal{X}}(y^{t-1}) \setminus \underline{x}$;
11 **if** $\hat{\pi}_{\bar{x}}(y^{t-1}) = P_X^*(\bar{x})$ **then**
12 $\bar{\mathcal{X}}(y^{t-1}) \leftarrow \bar{\mathcal{X}}(y^{t-1}) \setminus \bar{x}$;

APPENDIX B
AN ALGORITHM TO DETERMINE $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$

We design Algorithm 1 to determine a set of probabilities $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$ that satisfies (27)–(28). We denote by $\bar{\mathcal{X}}(1)$ the first element in set $\bar{\mathcal{X}}(y^{t-1})$. The order of the elements in $\bar{\mathcal{X}}(y^{t-1})$ is irrelevant.

For every group $\mathcal{G}_{\underline{x}}(y^{t-1})$ with $\underline{x} \in \underline{\mathcal{X}}(y^{t-1})$, Algorithm 1 goes through groups $\mathcal{G}_{\bar{x}}(y^{t-1})$ with $\bar{x} \in \bar{\mathcal{X}}(y^{t-1})$ to transfer probability $p_{\bar{x}\to\underline{x}}$ to $\mathcal{G}_{\underline{x}}(y^{t-1})$. The amount of probability $p_{\bar{x}\to\underline{x}}$ to transfer from $\mathcal{G}_{\bar{x}}(y^{t-1})$ to $\mathcal{G}_{\underline{x}}(y^{t-1})$ is the smallest of $\hat{\pi}_{\bar{x}}(y^{t-1}) - P_X^*(\bar{x})$ and $P_X^*(\underline{x}) - \hat{\pi}_{\underline{x}}(y^{t-1})$. After the update, if the new prior $\hat{\pi}_{\bar{x}}(y^{t-1})$ (or $\hat{\pi}_{\underline{x}}(y^{t-1})$) is equal to its target value $P_X^*(\bar{x})$ (or $P_X^*(\underline{x})$), the corresponding group will be removed from the set $\bar{\mathcal{X}}(y^{t-1})$ (or $\underline{\mathcal{X}}(y^{t-1})$). In this way, $\{p_{\bar{x}\to\underline{x}}\}_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1}),\underline{x}\in\underline{\mathcal{X}}(y^{t-1})}$ are determined. At the end of the algorithm, $\hat{\pi}_{\bar{x}}(y^{t-1})$ and $\hat{\pi}_{\underline{x}}(y^{t-1})$ indeed represent the left sides of (27) and (28), respectively.

We show that (27)–(28) hold: (28) holds by lines 4, 6, 8, 9–10 of Algorithm 1 and the fact that line 9 must be satisfied during the while loop since

$$\sum_{\underline{x}\in\underline{\mathcal{X}}(y^{t-1})} P_X^*(\underline{x}) - \pi_{\underline{x}}(y^{t-1}) = \sum_{\bar{x}\in\bar{\mathcal{X}}(y^{t-1})} \pi_{\bar{x}}(y^{t-1}) - P_X^*(\bar{x}) \tag{60}$$

ensures that there are enough probabilities $p_{\bar{x}\to\underline{x}}$ to transfer from groups in $\bar{\mathcal{X}}(y^{t-1})$ to $\mathcal{G}_{\underline{x}}(y^{t-1})$; (27) holds by lines 6, 7, 11–12 of Algorithm 1 and the facts that 1) (28) holds, i.e., $\hat{\pi}_{\underline{x}}(y^{t-1}) = P_X^*(\underline{x})$, 2) the minimum in line 6 ensures that $\hat{\pi}_{\bar{x}}(y^{t-1}) \geq P_X^*(\bar{x})$ at the end of the algorithm, 3) (60) implies that $\hat{\pi}_{\bar{x}}(y^{t-1}) > P_X^*(\bar{x})$ is impossible, otherwise $\sum_{x\in\mathcal{X}} \hat{\pi}_x(y^{t-1}) > \sum_{x\in\mathcal{X}} P_X^*(x) = 1$.

APPENDIX C
CHANNEL INPUT DISTRIBUTION IS EQUAL TO THE CAPACITY-ACHIEVING DISTRIBUTION

We show that (31) holds, i.e., the channel input distribution is equal to the capacity-achieving distribution. For any $x \in \mathcal{X}$

and $y^{t-1} \in \mathcal{Y}^{t-1}$, we expand right side of (31) as

$$\begin{aligned} & P_{X_t|Y^{t-1}}(x|y^{t-1}) \\ =& \sum_{z\in\mathcal{X}} P_{X_t|Z_t,Y^{t-1}}(x|z,y^{t-1}) P_{Z_t|Y^{t-1}}(z|y^{t-1}) && \text{(61a)} \\ =& \sum_{z\in\mathcal{X}} P_{X_t|Z_t,Y^{t-1}}(x|z,y^{t-1}) \pi_z(y^{t-1}) && \text{(61b)} \\ =& P_{X_t|Z_t,Y^{t-1}}(x|x,y^{t-1}) \pi_x(y^{t-1}) && \text{(61c)} \\ +& \sum_{z\neq x} P_{X_t|Z_t,Y^{t-1}}(x|z,y^{t-1}) \pi_z(y^{t-1}), && \text{(61d)} \end{aligned}$$

where (61a) holds by the law of total probability and (61b) holds by the definition of Z_t in (29).

By the randomization distribution in (30), if $x \in \overline{\mathcal{X}}(y^{t-1})$, then (61c) is equal to $P_X^*(x)$ and (61d) is equal to 0, and if $x \in \underline{\mathcal{X}}(y^{t-1})$, then (61c) is equal to $\pi_x(y^{t-1})$ and (61d) is equal to

$$\sum_{z\in\overline{\mathcal{X}}(y^{t-1})} \frac{p_{z\to x}}{\pi_x(y^{t-1})} \pi_x(y^{t-1}) = P_X^*(x) - \pi_x(y^{t-1}), \quad (62)$$

where (62) uses (28).

Appendix D
Converse Proof of Theorem 1

A. Converse proof

Inspired by Berlin et al.'s converse proof [31] for Burnashev's reliability function, we provide a converse bound on the JSCC reliability function for a fully accessible source by lower bounding the expected stopping time of an arbitrary code with block encoding using the error probability at the stopping time. The converse bound continues to apply for the JSCC reliability function for streaming, since given a DMC, every code with instantaneous encoding for transmitting the first k symbols of a $(q, \{t_n\}_{n=1}^{\infty})$ DSS in Definition 3 is a special code with block encoding for transmitting the first k symbols $S^k \in [q]^k$ of a DS (3).

We consider k symbols $S^k \in [q]^k$ of a DS with source distribution P_{S^k}, and we fix a non-degenerate DMC with a single-letter transition probability $P_{Y|X}\colon \mathcal{X} \to \mathcal{Y}$. We fix an arbitrary code with block encoding with a stopping time η_k for transmitting S^k over the non-degenerate DMC with feedback. We assume that the decoder is a MAP decoder (46), since given any encoding function and any stopping time in Definition 3, the MAP decoder (46) achieves the minimum error probability (15). For brevity, we denote the error probability of a MAP decoder given channel outputs $y^t \in \mathcal{Y}^t$ by

$$P_e(y^t) \triangleq 1 - \max_{s\in[q]^k} P_{S^k|Y^t}(s|y^t), \quad (63)$$

and we denote the error probability of a MAP decoder at the stopping time η_k by

$$P_e \triangleq \mathbb{E}[P_e(Y^{\eta_k})]. \quad (64)$$

We define stopping time τ_δ as

$$\tau_\delta \triangleq \min\{t\colon P_e(y^t) \le \delta \text{ or } t = \eta_k\}. \quad (65)$$

To obtain the converse bound on the JSCC reliability function for a fully accessible source, we establish a lower bound on the expected decoding time $\mathbb{E}[\eta_k]$ using the error probability P_e and and source distribution P_{S^k}. To this end, we lower bound $\mathbb{E}[\tau_\delta]$ and $\mathbb{E}[\eta_k - \tau_\delta]$, respectively. The lower bound on $\mathbb{E}[\tau_\delta]$ is stated below.

Lemma 1 (Modified Lemma 2 in [31])**.** *Consider k symbols $S^k \in [q]^k$ of a DS with source distribution P_{S^k} (1) and fix a non-degenerate DMC with capacity C (10). For any $\delta \in \left(0, \frac{1}{2}\right]$, it holds that*

$$\mathbb{E}[\tau_\delta] \ge \frac{H(S^k)}{C}\left(1 - \left(\delta + \frac{P_e}{\delta}\right)\frac{\log q^k}{H(S^k)}\right) - \frac{h(\delta)}{C}. \quad (66)$$

Proof. Appendix D-B. ∎

The lower bound on $\mathbb{E}[\eta_k - \tau_\delta]$ is stated below.

Lemma 2 (Modified Eq. (17) [31])**.** *Consider k symbols $S^k \in [q]^k$ of a DS with source distribution P_{S^k} and fix a non-degenerate DMC with transition probability $P_{Y|X}\colon \mathcal{X} \to \mathcal{Y}$ and maximum KL divergence C_1 (11). For any $\delta \in \left(0, \frac{1}{2}\right]$, it holds that*

$$\mathbb{E}[\eta_k - \tau_\delta] \ge \frac{\log\frac{1}{P_e} - \log 4 + \log(\min\left\{p_{\min}\delta, 1 - \max_{s\in[q]^k} P_{S^k}(s)\right\})}{C_1}, \quad (67)$$

where $p_{\min}$ in (67) is the minimum channel transition probability (34).

Proof. Appendix D-C. ∎

Summing up the right sides of (66) and (67), we obtain the following lower bound on the expected decoding time η_k of an arbitrary code with block encoding:

$$\mathbb{E}[\eta_k] \ge \frac{H(S^k)}{C}\left(1 - \left(\delta + \frac{P_e}{\delta}\right)\frac{\log q^k}{H(S^k)}\right) + \frac{\log\frac{1}{P_e}}{C_1} + \frac{-\log 4 + \log(\min\left\{p_{\min}\delta, 1 - \max_{s\in[q]^k} P_{S^k}(s)\right\})}{C_1} - \frac{h(\delta)}{C}. \quad (68)$$

The asymptotic performance of the lower bound (68) relies on two properties of the DS in Lemma 3, stated next.

Lemma 3. *Consider a DS with a well-defined and positive entropy rate H (2) and a finite single-letter alphabet $[q]$. Then,*

$$\lim_{k\to\infty} \frac{\log q^k}{H(S^k)} = \frac{\log q}{H} < \infty, \quad (69)$$

$$\liminf_{k\to\infty} \left(1 - \max_{s\in[q]^k} P_{S^k}(s)\right) > 0. \quad (70)$$

Proof. The proof of (70) is in Appendix D-D. ∎

Plugging (69)–(70) and $\delta = -\frac{1}{\log P_e}$ into the right side of (68), we obtain

$$\mathbb{E}[\eta_k] \ge \left(\frac{H(S^k)}{C} + \frac{\log\frac{1}{P_e}}{C_1}\right)(1 - o(1)), \quad (71)$$

where $o(1)$ in (71) is a positive term that converges to 0 as both $P_e \to 0$ and $k \to \infty$. Rearranging terms of (71), we

conclude that $E(R)$ is upper bounded by the right side of (38). Similar to [5], [31, Eq. (5)], [17, Proposition 9], here we need not consider the case where P_e does not converge to zero since this means $E(R) = 0$.

B. Proof of Lemma 1

We follow Berlin et al.'s notations [31]: we denote by $\mathcal{H}(S^k|Y^t)$ a random variable that satisfies $\mathcal{H}(S^k|y^t) = H(S^k|Y^t = y^t)$. Note that $\mathbb{E}[\mathcal{H}(S^k|Y^t)] = H(S^k|Y^t)$. If any step below has already been proved in [31], we avoid repeated reasoning by referring to the proof in [31]. Compared to Berlin et al.'s proof in [31, Sec. IV], the proof below does not assume that the source is equiprobably distributed – it keeps the generic form of the source prior P_{S^k}.

The sequence $\{\mathcal{H}(S^k|Y^t) + tC\}_{t=0,1,\dots}$ is a submartingale ([31, Lemma 2]) with respect to the filtration generated by the channel outputs. Using Doob's optional stopping theorem [38], the initial state of the submartingale $\{\mathcal{H}(S^k|Y^t) + tC\}_{t=0,1,\dots}$ is upper bounded as

$$H(S^k) \le H(S^k|Y^{\tau_\delta}) + \mathbb{E}[\tau_\delta]C. \tag{72}$$

For $\delta \in \left(0, \frac{1}{2}\right]$, the conditional entropy on the right side of (72) is upper bounded as

$$H(S^k|Y^{\tau_\delta}) \le h(\delta) + \left(\delta + \frac{P_e}{\delta}\right) \log q^k \tag{73}$$

using Fano's inequality (in the same manner as in [31, Eq. (14)–(16)]). Plugging (73) to the right side of (72) and rearranging terms, we obtain (66).

C. Proof of Lemma 2

We obtain the lower bound on $\mathbb{E}[\eta_k - \tau_\delta]$ in Lemma 2 by constructing a binary hypothesis test performed over a non-degenerate DMC with feedback. We first state a lower bound on the error probability of such a test. Consider a binary hypothesis test (H_0, H_1) performed over a DMC with feedback via a variable-length code with block encoding. The encoder sends a sequence of symbols $X_1, X_2, \dots,$ over the given DMC with feedback, such that at the stopping time T, if H_0 is true, then the channel output vector Y^T is distributed according to Q_{H_0}, otherwise, the channel output vector Y^T is distributed according to Q_{H_1}. At the stopping time, the decoder uses the decoding function $\hat{W}\colon \mathcal{Y}^T \to \{H_0, H_1\}$ to form a decoded hypothesis. We denote the set of channel outputs y^T that leads to decoded hypothesis H_i, $i \in \{0, 1\}$ by

$$\mathcal{Y}_{H_i} \triangleq \{y^T \in \mathcal{Y}^T \colon \hat{W}(y^T) = H_i\}, i \in \{0, 1\}. \tag{74}$$

We denote by p_{H_i}, $i \in \{0, 1\}$, the prior probability of hypothesis H_i, $i \in \{0, 1\}$ before the transmission. We denote the error probability of the binary hypothesis test at the stopping time T by

$$P_b \triangleq p_{H_0} Q_{H_0}(\mathcal{Y}_{H_1}) + p_{H_1} Q_{H_1}(\mathcal{Y}_{H_0}). \tag{75}$$

Lemma 4 (Lemma 1 in [31])**.** *Consider a binary hypothesis test with hypotheses H_0 and H_1 performed over a non-degenerate DMC with feedback that has maximum KL divergence C_1 (11) via a variable-length code with block encoding. The error probability of the binary hypothesis test P_b at stopping time T is lower bounded as*

$$P_b \ge \frac{\min\{p_{H_0}, p_{H_1}\}}{4} e^{-C_1 \mathbb{E}[T]}, \tag{76}$$

where p_{H_0} and p_{H_1} are the prior probabilities of the hypotheses[2]*.*

We employ the same hypothesis test as that in [31, Section V]. Compared to Berlin et al.'s proof [31, Sec. V], the proof below lower bounds the priors of the hypotheses differently since we consider a generic source distribution whereas Berlin at al.'s [31] considered equiprobable source symbols.

The binary hypothesis test (cf. [31, Section V]) starts at time $\tau_\delta + 1$ and operates as follows. Given any Y^{τ_δ}, we partition the alphabet $[q]^k$ into two sets $\mathcal{G}(Y^{\tau_\delta})$ and $[q]^k \setminus \mathcal{G}(Y^{\tau_\delta})$ (we will specify $\mathcal{G}$ in the sequel). The two hypotheses are $H_0\colon S^k \in \mathcal{G}(Y^{\tau_\delta})$ and $H_1\colon S^k \in [q]^k \setminus \mathcal{G}(Y^{\tau_\delta})$. At the stopping time η_k, the MAP decoder outputs the estimate of the source $\hat{S}^k_{\eta_k}$ using the channel outputs Y^{η_k}. If the estimate satisfies $\hat{S}^k_{\eta_k} \in \mathcal{G}(Y^{\tau_\delta})$, then we declare H_0, otherwise, we declare H_1. The error probability of decoding S^k is lower bounded by the error probability of the binary hypothesis test ([31, the second paragraph below Prop. 2]), i.e., given any $t \ge 0$, $y^t \in \mathcal{Y}^t$,

$$\begin{aligned} &\mathbb{P}[\hat{S}^k_{\eta_k} \ne S^k | Y^{\tau_\delta} = y^t] \\ \ge\ &\mathbb{P}[\hat{S}^k_{\eta_k} \notin \mathcal{G}(Y^{\tau_\delta}), S^k \in \mathcal{G}(Y^{\tau_\delta}) | Y^{\tau_\delta} = y^t] \\ &+ \mathbb{P}[\hat{S}^k_{\eta_k} \in \mathcal{G}(Y^{\tau_\delta}), S^k \notin \mathcal{G}(Y^{\tau_\delta}) | Y^{\tau_\delta} = y^t]. \end{aligned} \tag{77}$$

We invoke Lemma 4 with $p_{H_0} \leftarrow \mathbb{P}[H_0|Y^{\tau_\delta} = y^t]$, $p_{H_1} \leftarrow \mathbb{P}[H_1|Y^{\tau_\delta} = y^t]$, $\mathbb{E}[T] \leftarrow \mathbb{E}[\eta_k - \tau_\delta | Y^{\tau_\delta} = y^t]$ to further lower bound the left side of (77) and obtain

$$\begin{aligned} &\mathbb{P}[\hat{S}^k_{\eta_k} \ne S^k | Y^{\tau_\delta} = y^t] \ge \\ &\frac{\min\{\mathbb{P}[H_0|Y^{\tau_\delta} = y^t], \mathbb{P}[H_1|Y^{\tau_\delta} = y^t]\}}{4} e^{-C_1 \mathbb{E}[\eta_k - \tau_\delta | Y^{\tau_\delta} = y^t]}. \end{aligned} \tag{78}$$

To lower bound the minimization function on the right side of (78), we show that alphabet $[q]^k$ can always be partitioned into two groups $\mathcal{G}(Y^{\tau_\delta})$ and $[q^k] \setminus \mathcal{G}(Y^{\tau_\delta})$ such that for all $t \ge 0$, $y^t \in \mathcal{Y}^t$, the priors of the hypotheses are lower bounded as

$$\mathbb{P}[H_0|Y^{\tau_\delta} = y^t] \ge \min\left\{p_{\min}\delta, 1 - \max_{s \in [q]^k} P_{S^k}(s)\right\}, \tag{79a}$$

$$\mathbb{P}[H_1|Y^{\tau_\delta} = y^t] \ge \min\left\{p_{\min}\delta, 1 - \max_{s \in [q]^k} P_{S^k}(s)\right\}, \tag{79b}$$

where $p_{\min}$ is defined in (34). The priors of the hypotheses are both lower bounded by $p_{\min}\delta$ for any $\delta \in \left(0, \frac{1}{2}\right]$ if either event $A_1 \triangleq \{\tau_\delta \ge 1\}$ or event $A_2 \triangleq \{\tau_\delta = 0, \max_{s \in [q]^k} P_{S^k}(s) \le 0.5\}$ occurs, see [31, Section V]. The threshold 0.5 defining event A_2 corresponds to Berlin et al.'s reasoning in [31, the second case in the third paragraph after Prop. 2], which says at time τ_δ, if the posteriors[3] of all source sequences in $[q]^k$ are upper bounded by $1 - \delta \in [0.5, 1]$, then $[q]^k$ can be

[2]Notice that in the proof of [31, Lemma 1], Berlin et al. invoked [31, Proposition 1], which relies on $\mathbb{E}[T] < \infty$. Yet, (76) also trivially holds for $\mathbb{E}[T] = \infty$.

[3]The source posterior at time 0 is equal to the source prior P_{S^k}.

divided into two groups with both hypotheses priors lower bounded by $p_{\min}\delta$. In Berlin et al.'s [31] channel coding context where the source symbols are equiprobably distributed, the union of the events $A_1 \cup A_2$ occurs almost surely, since $P_{S^k}(s) = \frac{1}{q^k}$. Yet, in the JSCC context, it is possible that event $A_3 \triangleq \{\tau_\delta = 0, \max_{s\in[q]^k} P_{S^k}(s) > 0.5\}$ occurs. When A_3 occurs, we move the sequence $s \in [q]^k$ that attains the maximum in event A_3 to $\mathcal{G}(Y^{\tau_\delta})$, and move the remaining sequences to the other group. This group partitioning rule implies (79). Plugging (79) into (78), taking an expectation of (78) over Y^{τ_δ}, and applying Jensen's inequality to e^{-x} on the right side of (78), we obtain

$$P_e \geq \frac{\min\{p_{\min}\delta, 1 - \max_{s\in[q]^k} P_{S^k}(s)\}}{4} e^{-C_1 \mathbb{E}[\eta_k - \tau_\delta]}. \quad (80)$$

Rearranging terms in (80), we obtain (67).

D. Proof of Lemma 3

We show that (70) holds. We upper bound the entropy rate as

$$\lim_{k\to\infty} \frac{H(S^k)}{k}$$

$$\leq \liminf_{k\to\infty} \frac{h\left(\max_{s\in[q]^k} P_{S^k}(s)\right)}{k} + \left(1 - \max_{s\in[q]^k} P_{S^k}(s)\right) \log q \quad (81)$$

$$= \liminf_{k\to\infty} \left(1 - \max_{s\in[q]^k} P_{S^k}(s)\right) \log q, \quad (82)$$

where (81) holds since fixing the probability of the source sequence that attains $\max_{s\in[q]^k} P_{S^k}(s)$, the equiprobable distribution on the rest of $q^k - 1$ sequences maximizes the concave entropy function; (82) holds since the binary entropy function in (81) is bounded between $[0, 1]$. Finally, (70) holds since the entropy rate is positive by assumption.

Appendix E
Achievability Proof of Theorem 1

In the achievability proof, we fix a sequence of codes with instantaneous encoding for transmitting the first k symbols of a DSS, $k = 1, 2, \ldots$, over a non-degenerate DMC with feedback, evaluate the asymptotic behavior of the code sequence as $k \to \infty$, and conclude the achievability of $E(R)$ (38). In Appendix E-A, we particularize the DSS in Theorem 1 to the DS (3), and we show that both the MaxEJS code and the SED code [8] achieve $E(R)$ (38). In Appendix E-B, we consider a DSS with $f = \infty$, and we show that $E(R)$ is achievable by a buffer-then-transmit code that idles the transmissions and only buffers the arriving symbols during the symbol arriving period and implements a JSCC reliability function-achieving code with block encoding after the symbol arriving period. In Appendix E-C, we consider a DSS with $f < \infty$ that satisfies (a)–(b), and we show that $E(R)$ (38) is achievable by a code with instantaneous encoding that implements the instantaneous encoding phase in Section III during the symbol arriving period and a JSCC reliability function-achieving code with block encoding after the symbol arriving period.

A. A (fully accessible) DS

We show that both the MaxEJS code for all non-degenerate DMCs [8, Sec. IV-C] and the SED code for non-degenerate symmetric binary-input DMCs [8, Sec. V-B] achieve $E(R)$ (38) for a DS. We denote a deterministic encoding function at time t by

$$\gamma_t \colon [q]^k \to \mathcal{X}, \quad (83)$$

we denote the vector of the message posteriors at time t by

$$\boldsymbol{\rho}(Y^t) \triangleq [P_{S^k|Y^t}(1|Y^t), P_{S^k|Y^t}(2|Y^t), \ldots, P_{S^k|Y^t}(q^k|Y^t)], \quad (84)$$

and we denote the extrinsic Jensen-Shannon (EJS) divergence [8] at time t by

$$\mathrm{EJS}(\boldsymbol{\rho}(Y^{t-1}), \gamma_t) \triangleq \sum_{i=1}^{q^k} P_{S^k|Y^{t-1}}(i|Y^{t-1})$$
$$D\left(P_{Y|X=\gamma_t(i)} \middle|\middle| \sum_{j\neq i} \frac{P_{S^k|Y^{t-1}}(j|Y^{t-1})}{1 - P_{S^k|Y^{t-1}}(i|Y^{t-1})} P_{Y|X=\gamma_t(j)}\right). \quad (85)$$

The MaxEJS code [8, Section IV.C] sets its encoding function γ_t^* at time t by solving the maximization problem:

$$\gamma_t^* \triangleq \arg\max_{\gamma_t\in\mathcal{E}} \mathrm{EJS}(\boldsymbol{\rho}(Y^{t-1}), \gamma_t), \quad (86)$$

where $\mathcal{E}$ is the set of all possible deterministic functions γ_t (83). The SED code [8] corresponds to the instantaneous SED code in Section V-C for a fully accessible source.

Lemma 5, stated next, will be used to examine whether a code with block encoding achieves the JSCC reliability function for a fully accessible source.

Lemma 5. *Consider k symbols $S^k \in [q]^k$ of a DS with prior probability P_{S^k} and fix a non-degenerate DMC with capacity C (10) and the maximum KL divergence C_1 (11). A code with block encoding achieves the JSCC reliability function (38) for the fully accessible source if and only if its stopping time η_k and its error probability ϵ (15) at the stopping time η_k satisfy*

$$\mathbb{E}[\eta_k] \leq \left(\frac{H(P_{S^k})}{C} + \frac{\log\frac{1}{\epsilon}}{C_1}\right)(1 + o(1)), \quad (87)$$

where $o(1) \to 0$ as $k \to \infty$.

Proof. If a code with block encoding satisfies (87), then it achieves $E(R)$ (38) because plugging (87) into (17) gives (38). Conversely, if a code with block encoding achieves $E(R)$ (38), then $\mathbb{E}[\eta_k]$ is upper bounded by the right side of (87). This is because any achievability bound on $\mathbb{E}[\eta_k]$ that is asymptotically larger than the right side of (87) cannot achieve (38). ∎

We show that the MaxEJS code and the SED code both satisfy (87). While [8, Eq. (32)] in [8, Theorem 1] is obtained by plugging a uniform prior of the message to the entropy function in [8, Appendix II, Eq. (71)], we leave the prior in its generic form and obtain a modified version of [8, Theorem 1] as follows.

Lemma 6 (Modified Theorem 1 in [8]). *Fix a non-degenerate DMC with capacity C (10) and maximum KL divergence C_1 (11), and consider k symbols $S^k \in [q]^k$ of a DS with source distribution P_{S^k}. If the encoding functions γ_t, $t = 1, \ldots, \eta_k$ of a code with block encoding with the MAP decoder (46) and the ϵ-stopping rule (49) satisfy*

$$\mathrm{EJS}(\boldsymbol{\rho}(Y^{t-1}), \gamma_t) \geq C, \tag{88}$$

$$\mathrm{EJS}(\boldsymbol{\rho}(Y^{t-1}), \gamma_t) \geq \left(1 - \frac{1}{1 + \max\{\log q^k, \log \frac{1}{\epsilon}\}}\right) C_1,$$

$$\textit{if } \max_{i \in [q]^k} \rho_i(Y^{t-1}) \geq 1 - \frac{1}{1 + \max\{\log q^k, \log \frac{1}{\epsilon}\}}, \tag{89}$$

then the expected decoding time of the code with block encoding is upper bounded as

$$\mathbb{E}[\eta_k] \leq \frac{H(P_{S^k}) + \log\log \frac{q^k}{\epsilon}}{C} + \frac{\log \frac{1}{\epsilon} + 1}{C_1} + \frac{6(4C_2)^2}{CC_1}, \tag{90}$$

where $C_2 \triangleq \max_{y \in \mathcal{Y}} \frac{\max_{x \in \mathcal{X}} P_{Y|X}(y|x)}{\min_{x \in \mathcal{X}} P_{Y|X}(y|x)}$.

Since the MaxEJS code satisfies (88)–(89) for all non-degenerate DMCs by [8, Proposition 2], and the SED code [8, Sec. V-B] satisfies (88)–(89) for non-degenerate symmetric binary-input DMCs by [8, Proposition 4], we conclude from (90) and (69) that they satisfy (87).

B. A DSS with $f = \infty$

Fixing a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with $f = \infty$, we show that $E(R)$ (38) is achievable by a buffer-then-transmit code that buffers the arriving symbols at times $t = 1, \ldots, t_k$ and operates as a JSCC reliability function (38)-achieving code with block encoding for k symbols S^k of a (fully accessible) DS with prior P_{S^k} at times $t \geq t_k + 1$ (e.g., the MaxEJS code [8]). To this end, we show an achievability (upper) bound on the expected stopping time of the buffer-then-transmit code.

We denote by η'_k the stopping time of the buffer-then-transmit code. We denote by η_k the stopping time of a code with block encoding that achieves the JSCC reliability function (38) for a fully accessible source, and we denote by ϵ_k its error probability at η_k (15). Since the decoding starts after time t_k, we have

$$\eta'_k = t_k + \eta_k. \tag{91}$$

We invoke Lemma 5 with $\epsilon \leftarrow \epsilon_k$ to upper bound $\mathbb{E}[\eta_k]$ on the right side of (91) and obtain an achievability bound on the expected decoding time $\mathbb{E}[\eta'_k]$ of the buffer-then-transmit code:

$$\mathbb{E}[\eta'_k] \leq \left(\frac{H(P_{S^k})}{C} + \frac{\log \frac{1}{\epsilon_k}}{C_1}\right)(1 + o(1)) + t_k. \tag{92}$$

For any DSS satisfying assumptions in Theorem 1, plugging (92) into (17), we obtain (39). Since $f = \infty$, the achievability bound (39) is equal to (38).

C. A DSS with $f < \infty$

Fixing a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with $f < \infty$, we show that $E(R)$ (38) is achievable by a code with instantaneous encoding that implements the instantaneous encoding phase at times $t = 1, 2, \ldots, t_k$ and operates as a JSCC reliability function (38)-achieving code with block encoding for k symbols S^k of a (fully accessible) DS with prior $P_{S^k|Y^{t_k}}$ at times $t \geq t_k + 1$, where $Y_1, \ldots, Y_{t_k}$ are the channel outputs generated in the instantaneous encoding phase. To this end, we will use Lemmas 7–9, stated below, together with Lemma 5 in Appendix E-A to obtain an achievability (upper) bound on the expected stopping time of the code.

We fix an error probability ϵ_k (15). We denote by η_k the stopping time that ensures ϵ_k of a JSCC reliability function-achieving code with block encoding for k symbols with prior $P_{S^k|Y^{t_k}}$. The stopping time η'_k of the code with instantaneous encoding described above is

$$\eta'_k = t_k + \eta_k \tag{93}$$

and its error probability is ϵ_k.

The *directed information* $I(A^n \to B^n)$ from a sequence A^n to a sequence B^n is defined as [42]

$$I(A^n \to B^n) = \sum_{i=1}^{n} I(A^i; B_i | B^{i-1}). \tag{94}$$

The directed information captures the information due to the causal dependence of B^n on A^n.

To upper bound the expected decoding time $\mathbb{E}[\eta'_k]$, it suffices to upper bound $\mathbb{E}[\eta_k]$ (93). Lemmas 7–9, stated below, show the behavior of the mutual information $I(S^k; Y^{t_k})$ as $k \to \infty$ generated by the instantaneous encoding phase in Section III.

Lemma 7. *Fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS, and fix a non-degenerate DMC with capacity C (10) and the maximum KL divergence C_1 (11). The instantaneous encoding phase that operates at times $t = 1, 2, \ldots, t_k$ in Section III gives rise to*

$$I(S^k; Y^{t_k}) = t_k C - I(X^{t_k} \to Y^{t_k} | S^k). \tag{95}$$

Proof. Appendix E-D. ∎

Lemma 8, stated next, displays the implications of assumption (b) in Theorem 1. Given a DSS, we extract all the *distinct* symbol arriving times from $t_1 \leq t_2 \leq \ldots$, and we denote the sequence of distinct symbol arriving times by

$$d_1 < d_2 < \ldots \tag{96}$$

For example, if a DSS emits a source symbol every $\lambda \geq 1$ channel uses (6), then the symbol arriving times are equal to the distinct symbols arriving times, i.e., $t_n = d_n$, and Lemma 8 below trivially holds.

Lemma 8. *Fix a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with $f < \infty$ and f satisfying assumption (b) in Theorem 1. Then,*

(i) *The time interval between consecutive symbol arriving times satisfies*

$$t_{n+1} - t_n = o(n), n = 1, 2, \ldots; \tag{97}$$

(ii) *The DSS has an infinite number of distinct symbol arriving times* $d_{n'}$, $n' = 1, 2, \ldots$.

Proof. (i) Assumption (b) and $f < \infty$ ensure that $f \in \left(\frac{1}{H}\left(H(P_Y^*) - \log\frac{1}{p_{\max}}\right), \infty\right)$. Thus, $\{\frac{t_n}{n}\}, n = 1, 2, \ldots$ is a Cauchy sequence, and (97) follows.

(ii) The DSS has an infinite number of distinct symbol arriving times since $0 < f < \infty$ implies that there exist two positive functions g_1, g_2 with $g_1(n) = \Omega(n)$ and $g_2(n) = O(n)$ such that the symbol arriving time is bounded between $g_1(n) \leq t_n \leq g_2(n)$, and the symbol arriving interval is constrained by (97). ∎

Lemma 9, stated next, shows the asymptotic behavior of $I(X^{t_k} \to Y^{t_k}|S^k)$ in (95).

Lemma 9. *Fix a* $(q, \{t_n\}_{n=1}^{\infty})$ *DSS that satisfies* (a)–(b) *and* $f < \infty$, *and fix a non-degenerate DMC with capacity* C (10) *and the maximum KL divergence* C_1 (11). *The instantaneous encoding phase that operates at times* $t = 1, 2, \ldots, t_k$ *in Section III satisfies*

$$I(X^{t_k} \to Y^{t_k}|S^k) = o(t_k), \tag{98}$$

where $\lim_{k\to\infty} \frac{o(t_k)}{t_k} = 0$.

Proof. Appendix E-E. ∎

Using Lemmas 5, 7–9, we obtain an achievability bound on the expected decoding time $\mathbb{E}[\eta_k]$ (93):

$$\mathbb{E}[\eta_k'] \leq \left(\frac{H(S^k|Y^{t_k})}{C} + \frac{\log\frac{1}{\epsilon_k}}{C_1}\right)(1 + o(1)) + t_k \tag{99a}$$

$$= \left(\frac{H(S^k) - I(S^k; Y^{t_k})}{C} + \frac{\log\frac{1}{\epsilon_k}}{C_1}\right)(1 + o(1)) + t_k \tag{99b}$$

$$= \left(\frac{H(S^k)}{C} + \frac{\log\frac{1}{\epsilon_k}}{C_1}\right)(1 + o(1)) \tag{99c}$$

where (99a) holds by upper bounding $\mathbb{E}[\eta_k]$ in (93) using (87) with $P_{S^k} \leftarrow P_{S^k|Y^{t_k}=y^{t_k}}$ and taking an expectation with respect to Y^{t_k}; (99b) holds by expanding $H(S^k|Y^{t_k})$ in (99a); (99c) holds by plugging Lemmas 7 and 9 into $I(S^k; Y^{t_k})$ in (99b) and using the fact that $\frac{o(t_k)}{H(S^k)} \leq \frac{o(t_k)}{t_k}\frac{1}{fH} = o(1)$, true due to the assumptions that the entropy rate H and the symbol arriving rate f are both positive. Plugging the achievability bound (99) into (17), we conclude that the code with instantaneous encoding achieves (38).

D. Proof of Lemma 7

We first write the mutual information $I(S^k, X^t; Y_t|Y^{t-1})$ in two ways:

$$I(S^k, X^t; Y_t|Y^{t-1}) = I(S^k; Y_t|Y^{t-1}) + I(X^t; Y_t|Y^{t-1}, S^k) \tag{100a}$$

$$= I(X^t; Y_t|Y^{t-1}) + I(S^k; Y_t|Y^{t-1}, X^t), \tag{100b}$$

where the second term on the right side of (100b) is equal to 0 since $Y_t - (Y^{t-1}, X^t) - S^k$ is a Markov chain. Thus,

$$I(S^k; Y_t|Y^{t-1}) = I(X^t; Y_t|Y^{t-1}) - I(X^t; Y_t|Y^{t-1}, S^k). \tag{101}$$

We expand $I(S^k; Y^{t_k})$ on the left side of (95) as

$$I(S^k; Y^{t_k}) = \sum_{t=1}^{t_k} I(S^k; Y_t|Y^{t-1}) \tag{102a}$$

$$= \sum_{t=1}^{t_k} I(X^t; Y_t|Y^{t-1}) - I(X^t; Y_t|Y^{t-1}, S^k) \tag{102b}$$

$$= t_k C - I(X^{t_k} \to Y^{t_k}|S^k), \tag{102c}$$

where (102a) is by the chain rule; (102b) is by plugging (101) into (102a); (102c) is by applying the definition of the directed information (94) to the second term of (102b) and plugging (31) and the fact that Y_i, $i = 1, \ldots, t_k$ are i.i.d. according to P_Y^* into the first term of (102b). The channel outputs $Y_1, Y_2, \ldots$ are independent since $Y_t - X_t - Y^{t-1}$ is a Markov chain and X_t is independent of Y^{t-1} (31). The channel outputs $Y_1, Y_2, \ldots$ are identically distributed according to P_Y^* since $X_1, X_2, \ldots$ follow the capacity-achieving distribution P_X^* (31).

E. Proof of Lemma 9

To show (98), we first upper bound the conditional directed information in (98) as a sum of conditional entropies, and upper bound each conditional entropy by a function of the source prior $\theta_{S^{N(t)}}(Y^{t-1})$. Then, we show that $\theta_{S^{N(t)}}(Y^{t-1})$ converges in probability to zero in time t for $t \in [1, t_k]$ as $k \to \infty$. Finally, we show that the convergence of the source prior leads to the convergence of the entropy sequence and conclude (98).

The conditional directed information in (98) can be upper bounded as

$$I(X^{t_k} \to Y^{t_k}|S^k) = \sum_{t=1}^{t_k} I(X^t; Y_t|Y^{t-1}, S^k) \tag{103a}$$

$$\leq \sum_{t=1}^{t_k} H(X_t|Y^{t-1}, S^k) \tag{103b}$$

$$= \sum_{t=1}^{t_k} H(X_t|Z_t, Y^{t-1}), \tag{103c}$$

where (103a) is by the chain rule, and (103c) holds since Z_t is a deterministic function of (Y^{t-1}, S^k) and $X_t - (Z_t, Y^{t-1}) - S^k$ is a Markov chain.

We upper bound each term in the sum of (103c) using $\theta_{S^{N(t)}}(Y^{t-1})$. Given that $Z_t = z$, $Y^{t-1} = y^{t-1}$, if $z \in \underline{\mathcal{X}}(y^{t-1})$, we use (30) to conclude

$$H(X_t|Z_t = z, Y^{t-1} = y^{t-1}) = 0. \tag{104}$$

If $z \in \overline{\mathcal{X}}(y^{t-1})$, we rearrange terms in (24) to obtain

$$1 - \frac{P_X^*(z)}{\pi_z(y^{t-1})} \leq 1 - \frac{P_X^*(z)}{P_X^*(z) + \min_{i \in \mathcal{G}_z(y^{t-1})} \theta_i(y^{t-1})} \tag{105a}$$

$$\leq \frac{\min_{i \in \mathcal{G}_z(y^{t-1})} \theta_i(y^{t-1})}{\min_{x \in \mathcal{X}} P_X^*(x)}. \tag{105b}$$

We upper bound $H(X_t|Z_t = z, Y^{t-1} = y^{t-1})$, $z \in \overline{\mathcal{X}}(y^{t-1})$ by

$$H(X_t|Z_t = z, Y^{t-1} = y^{t-1})$$

$$= \frac{P_X^*(z)}{\pi_z(y^{t-1})} \log \frac{\pi_z(y^{t-1})}{P_X^*(z)} + \sum_{x \in \underline{\mathcal{X}}(y^{t-1})} \frac{p_{z \to x}}{\pi_z(y^{t-1})} \log \frac{\pi_z(y^{t-1})}{p_{z \to x}} \tag{106a}$$

$$\leq \frac{P_X^*(z)}{\pi_z(y^{t-1})} \log \frac{\pi_z(y^{t-1})}{P_X^*(z)} + \left(1 - \frac{P_X^*(z)}{\pi_z(y^{t-1})}\right) \log \frac{|\mathcal{X}| - 1}{1 - \frac{P_X^*(z)}{\pi_z(y^{t-1})}} \tag{106b}$$

$$= \left(1 - \frac{P_X^*(z)}{\pi_z(y^{t-1})}\right) \log(|\mathcal{X}| - 1) + h\left(1 - \frac{P_X^*(z)}{\pi_z(y^{t-1})}\right) \tag{106c}$$

$$\leq \frac{\min_{i \in \mathcal{G}_z(y^{t-1})} \theta_i(y^{t-1})}{\min_{x \in \mathcal{X}} P_X^*(x)} \log(|\mathcal{X}| - 1) + 2\sqrt{\frac{\min_{i \in \mathcal{G}_z(y^{t-1})} \theta_i(y^{t-1})}{\min_{x \in \mathcal{X}} P_X^*(x)}}, \tag{106d}$$

where (106a) holds by (27) and (30); (106b) holds since the sum in the second term on the right side of (106a) is maximized if $p_{z \to x}$ is equiprobable on $\underline{\mathcal{X}}(y^{t-1})$, and $|\underline{\mathcal{X}}(y^{t-1})| \leq |\mathcal{X}| - 1$; (106c) holds by rearranging terms; (106d) holds by applying the upper bound $h(p) \leq 2\sqrt{p}$ to the binary entropy function in (106c) and plugging (105) into (106c). Therefore, each term in (103c) is upper bounded as

$$H(X_t|Z_t, Y^{t-1})$$

$$\leq \frac{\log(|\mathcal{X}| - 1)}{\min_{x \in \mathcal{X}} P_X^*(x)} \mathbb{E}\left[\min_{i \in \mathcal{G}_{Z_t}(Y^{t-1})} \theta_i(Y^{t-1})\right] + \frac{2}{\sqrt{\min_{x \in \mathcal{X}} P_X^*(x)}} \mathbb{E}\left[\sqrt{\min_{i \in \mathcal{G}_{Z_t}(Y^{t_k-1})} \theta_i(Y^{t-1})}\right] \tag{107a}$$

$$\leq \frac{\log(|\mathcal{X}| - 1)}{\min_{x \in \mathcal{X}} P_X^*(x)} \mathbb{E}\left[\theta_{S^{N(t)}}(Y^{t-1})\right] + \frac{2}{\sqrt{\min_{x \in \mathcal{X}} P_X^*(x)}} \mathbb{E}\left[\sqrt{\theta_{S^{N(t)}}(Y^{t-1})}\right] \tag{107b}$$

$$\leq \alpha \mathbb{E}\left[\sqrt{\theta_{S^{N(t)}}(Y^{t-1})}\right], \tag{107c}$$

where

$$\alpha \triangleq \max\left\{\frac{\log(|\mathcal{X}| - 1)}{\min_{x \in \mathcal{X}} P_X^*(x)}, \frac{2}{\sqrt{\min_{x \in \mathcal{X}} P_X^*(x)}}\right\}; \tag{108}$$

(107a) holds by (104) and (106); (107b) holds since $S^{N(t)} \in \mathcal{G}_{Z_t}(Y^{t_k-1})$.

To obtain the asymptotic behavior of $H(X_t|Z_t, Y^{t-1})$ in (107), we proceed to analyze the asymptotic behavior of $\theta_{S^{N(t)}}(Y^{t-1})$. The source prior $\theta_{S^{N(t)}}(Y^{t-1})$ in (107b) is upper bounded as

$$\theta_{S^{N(t)}}(Y^{t-1})$$

$$= P_{S^{N(t)}}(S^{N(t)}) \prod_{j=1}^{t-1} \frac{\sum_{x \in \mathcal{X}} P_{Y|X}(Y_j|x) P_{X_j|Z_j, Y^{j-1}}(x|Z_j, Y^{j-1})}{P_Y^*(Y_j)} \tag{109a}$$

$$\leq P_{S^{N(t)}}(S^{N(t)}) \prod_{j=1}^{t-1} \frac{p_{\max}}{P_Y^*(Y_j)}, \tag{109b}$$

where (109a) holds by (22) and (32); (109b) holds since the numerator in the product term of (109a) is upper bounded by $p_{\max}$ (33). Given a DSS in Lemma 9 with distinct symbol arriving times $d_{n'}$, $n' = 1, 2, \ldots$ (96) (n' is not bounded due to Lemma 8 (ii)), we denote the gap between the symbol arriving rate f and the threshold on the right side of (37) by

$$\gamma \triangleq f - \frac{1}{\underline{H}}\left(H(P_Y^*) - \log \frac{1}{p_{\max}}\right) \in (0, \infty). \tag{110}$$

For any $t \in [d_{n'}, d_{n'+1})$, $n' = 1, 2, \ldots$, the source prior $\theta_{S^{N(t)}}(Y^{t-1})$ (109) satisfies

$$\mathbb{P}\left[\frac{1}{t} \log \theta_{S^{N(t)}}(Y^{t-1}) \leq -\gamma \underline{H}\right] \tag{111a}$$

$$\geq \mathbb{P}\Bigg[-\frac{1}{t}\left(\log \frac{1}{P_{S^{N(d_{n'})}}\left(S^{N(d_{n'})}\right)}\right) + \frac{t-1}{t} \log p_{\max} + \frac{1}{t} \sum_{j=1}^{t-1} \log \frac{1}{P_Y^*(Y_j)} \leq -\gamma \underline{H}\Bigg] \tag{111b}$$

$$\geq \mathbb{P}\Bigg[\frac{N(d_{n'})}{d_{n'+1} - 1}\left(\frac{1}{N(d_{n'})} \log \frac{1}{P_{S^{N(d_{n'})}}\left(S^{N(d_{n'})}\right)}\right) \geq \log p_{\max} + H(P_Y^*) + \gamma \underline{H}, \frac{t-1}{t} \log p_{\max} + \frac{1}{t} \sum_{j=1}^{t-1} \log \frac{1}{P_Y^*(Y_j)} = \log p_{\max} + H(P_Y^*)\Bigg] \tag{111c}$$

$$\geq \mathbb{P}\Bigg[\frac{N(d_{n'})}{d_{n'+1} - 1}\left(\frac{1}{N(d_{n'})} \log \frac{1}{P_{S^{N(d_{n'})}}\left(S^{N(d_{n'})}\right)}\right) \geq \log p_{\max} + H(P_Y^*) + \gamma \underline{H}\Bigg] \tag{111d}$$

$$+ \mathbb{P}\left[\frac{t-1}{t} \log p_{\max} + \frac{1}{t} \sum_{j=1}^{t-1} \log \frac{1}{P_Y^*(Y_j)} = \log p_{\max} + H(P_Y^*)\right] \tag{111e}$$

$$- 1 \tag{111f}$$

$$\to 1, \tag{111g}$$

as $n' \to \infty$, where (111b) holds by plugging (109b) into (111a) and by replacing $N(t) \leftarrow N(d_{n'})$ since $t \in [d_{n'}, d_{n'+1})$; (111c) holds since $t \leq d_{n'+1} - 1$ and the event in (111b) is implied by the events in (111c); (111d)–(111f) hold by applying Fréchet inequalities [40] to the probability in (111c); (111g) holds since both probabilities in (111d)–(111e) converge to 1 as $n' \to \infty$: the probability in (111d)

converges to 1 as $n' \to \infty$ by Lemma 8 (i), the fact that $\liminf_{n'\to\infty} \frac{N(d_{n'})}{d_{n'}} \geq f$ since $\left\{\frac{N(d_{n'})}{d_{n'}}\right\}_{n'=1}^{\infty}$ is a subsequence of $\left\{\frac{n}{t_n}\right\}_{n=1}^{\infty}$, the lower bound on the symbol arriving rate (assumption (b)), the lower bound on the information in $S^{N(d_{n'})}$ (assumption (a)), and the fact that $N(d_{n'}) \to \infty$ as $n' \to \infty$ since $N(d_{n'}) \geq n'$; the probability in (111e) converges to 1 since the sum over the logarithms of i.i.d. random variables $Y_1, Y_2, \ldots$ (they are i.i.d. by the argument below (102c)) in (111e) converges to $H(P_Y^*)$ by the law of large numbers, and $t \to \infty$ as $n' \to \infty$ due to $t \geq d_{n'}$. Rearranging terms in (111a), we we conclude that for any $\delta \in (0,1)$, there exists $n_\delta \in \mathbb{Z}_+$, such that for all $n' \geq n_\delta$, $t \in [d_{n'}, d_{n'+1})$, the probability in (113) satisfies

$$\mathbb{P}[\theta_{S^{N(t)}}(Y^{t-1}) \leq e^{-\gamma \underline{H} t}] > 1 - \delta. \tag{112}$$

We analyze the asymptotic behavior of $H(X_t|Z_t, Y^{t-1})$ in (107) using (112). Using the boundedness of the source prior $\theta_{S^{N(t)}}(Y^{t-1}) \in [0,1]$, we upper bound the expectations in the right side of (107c) as

$$\begin{aligned}\mathbb{E}\left[\sqrt{\theta_{S^{N(t)}}(Y^{t-1})}\right] &\leq \mathbb{P}\left[\theta_{S^{N(t)}}(Y^{t-1}) > e^{-\gamma \underline{H} t}\right] \\ &\quad + e^{-\frac{\gamma \underline{H}}{2} t} \mathbb{P}\left[\theta_{S^{N(t)}}(Y^{t-1}) \leq e^{-\gamma \underline{H} t}\right] \qquad (113) \\ &< \delta + e^{-\frac{\gamma \underline{H}}{2} d_{n_\delta}}(1-\delta), \qquad (114)\end{aligned}$$

$\forall t \in [d_{n'}, d_{n'+1})$, where (114) holds due to (112) and the fact that the function $f(p) = p + \beta(1-p)$, $\beta < 1$, is monotonically increasing on $p \in [0,1]$.

Plugging (114) into (107c), we conclude that for all $n' \geq n_\delta$, it holds that $\forall t \in [d_{n'}, d_{n'+1})$,

$$H(X_t|Z_t, Y^{t-1}) < \alpha\left(\delta + e^{-\frac{\gamma \underline{H}}{2} d_{n_\delta}}(1-\delta)\right). \tag{115}$$

We proceed to show (98) using (115). Dividing both sides of (103) by t_k and taking $k \to \infty$, we upper bound the left side of (103) as

$$\begin{aligned}&\limsup_{k\to\infty} \frac{1}{t_k} I(X^{t_k} \to Y^{t_k}|S^k) \\ \leq\ &\limsup_{k\to\infty} \frac{1}{t_k} \sum_{t=1}^{t_k} H(X_t|Z_t, Y^{t-1}) \qquad (116a) \\ <\ &\limsup_{k\to\infty} \frac{1}{t_k} \Bigg(|t_k - d_{n_\delta}| \alpha \left(\delta + e^{-\frac{\gamma \underline{H}}{2} d_{n_\delta}}(1-\delta)\right) \\ &\quad + d_{n_\delta} \log|\mathcal{X}| \Bigg) \qquad (116b) \\ =\ &\alpha\left(\delta + e^{-\frac{\gamma \underline{H}}{2} d_{n_\delta}}(1-\delta)\right), \qquad (116c)\end{aligned}$$

where (116a) holds by (103c); (116b) holds by upper bounding $H(X_t|Z_t, Y^{t-1}) \leq \log|\mathcal{X}|$ for $t \leq d_{n_\delta}$ and upper bounding $H(X_t|Z_t, Y^{t-1})$ by (115) for $t > d_{n_\delta}$; (116c) holds since Lemma 8 (i) implies that $d_{n_\delta} < \infty$ for some $n_\delta \in \mathbb{Z}_+$, and $f < \infty$ implies that $t_k \to \infty$ as $k \to \infty$.

Since δ can be made arbitrarily small while d_{n_δ} can be made arbitrarily large, we conclude (98).

Appendix F
Decoding before the final arrival time

For transmitting the first k source symbols of a $(q, \{t_n\}_{n=1}^{\infty})$ DSS with $p_{S,\max} < 1$, we show that if we decode before the final arrival time t_k, then the error probability $\mathbb{P}[S^k \neq \hat{S}_t^k]$, $t < t_k$ will not vanish with k for any code with instantaneous encoding.

For any $t < t_k$, $y^t \in \mathcal{Y}^t$, we lower bound the conditional error probability as

$$\mathbb{P}[S^k \neq \hat{S}_t^k | Y^t = y^t] \geq 1 - \max_{i \in [q]^k} P_{S^k|Y^t}(i|y^t), \tag{117}$$

where the equality is attained by the MAP decoder. Taking an expectation of both sides of (117), we obtain

$$\begin{aligned}&\mathbb{P}[S^k \neq \hat{S}_t^k] \\ \geq\ &1 - \mathbb{E}\left[\max_{i\in[q]^k} P_{S^k|Y^t}(i|Y^t)\right] \qquad (118a) \\ =\ &1 - \mathbb{E}\left[\max_{i\in[q]^k} \sum_{j\in[q]^{N(t)}} P_{S^k|S^{N(t)}}(i|j) P_{S^{N(t)}|Y^t}(j|Y^t)\right] \qquad (118b) \\ \geq\ &1 - \max_{i\in[q]^k, j\in[q]^{N(t)}} P_{S^k|S^{N(t)}}(i|j) \qquad (118c) \\ \geq\ &1 - \prod_{n=N(t)+1}^{k} \max_{s\in[q], s'\in[q]^{n-1}} P_{S_n|S^{n-1}}(s|s') \qquad (118d) \\ \geq\ &1 - (p_{S,\max})^{k-N(t)} \qquad (118e) \\ >\ &0, \qquad (118f)\end{aligned}$$

where (118b) holds since $S^k - S^{N(t)} - Y^t$ is a Markov chain; (118c) holds by upper bounding $P_{S^k|S^{N(t)}}(i|j)$ in (118b) by its maximum; (118d) holds by writing $P_{S^k|S^{N(t)}}(\cdot|\cdot)$ as a product of probabilities $\{P_{S_n|S^{n-1}}(\cdot|\cdot)\}_{n=N(t)+1}^{k}$ and maximizing each term in the product; (118e) holds by upper bounding each term in the product by $p_{S,\max}$ (35); (118f) holds by the assumption $p_{S,\max} < 1$.

Appendix G
Proof of Remark 1

We show that after the instantaneous encoding phase drops the randomization step (25)–(30) and only transmits Z_t (29) as the channel input, it continues to satisfy the sufficient condition in (41) under assumption (b′). To this end, we first write $I(S^k; Y^{t_k})$ in (41) as a sum of mutual informations. Then, we show that all source priors converge pointwise to zero in time during the symbol arriving period $[1, t_k]$ as $k \to \infty$; this implies that group priors converge pointwise to the capacity-achieving probabilities. Finally, we show that the convergence of the group priors implies that the summands of $I(S^k; Y^{t_k})$ converge to the capacity C and conclude (41). Given channel outputs $y^{t-1} \in \mathcal{Y}^{t-1}$, we denote the source sequence in $[q]^{N(t)}$ that has the maximum source prior by

$$i^* \triangleq \arg\max_{i\in[q]^{N(t)}} \theta_i(y^{t-1}). \tag{119}$$

To expand $I(S^k; Y^{t_k})$ in (41), we first notice that (100)–(102b) continue to hold, thus $I(S^k; Y^{t_k})$ is equal to (102b).

The second term on the right side of (102b) is equal to zero since X_t is a deterministic function of (Y^{t-1}, S^k), thus,

$$I(S^k; Y^{t_k}) = \sum_{t=1}^{t_k} I(X_t; Y_t | Y^{t-1}). \tag{120}$$

We proceed to analyze the asymptotic behavior of $\theta_{i^*}(y^{t-1})$ (119). Since the encoder drops the randomization step (25)–(30) and only transmits Z_t (29) as the channel input, the posterior update (32) becomes (45). Upper bounding $P_{S^{N(t)}|S^{N(t-1)}}(\cdot|\cdot)$ in the prior update (22) by the maximum symbol arriving probability $p_{S,\max}^{N(t)-N(t-1)}$ (35), and upper bounding the numerator by $p_{\max}$ and the denominator by $p_{\min}$ in the fraction on the right side of (45), we obtain an upper bound on the source prior $\theta_{i^*}(y^{t-1})$ as

$$\theta_{i^*}(y^{t-1}) \le p_{S,\max}^{N(t)} \left(\frac{p_{\max}}{p_{\min}}\right)^{t-1} \tag{121}$$

for all $i \in [q]^{N(t)}$, $y^{t-1} \in \mathcal{Y}^{t-1}$. Given a DSS that satisfies assumption (b′) with $f < \infty$ and distinct symbol arriving times $d_{n'}$, $n' = 1, 2, \ldots$ (96) (n' is not bounded due to Lemma 8), similar to (110), we denote the gap between the symbol arriving rate f and the threshold in assumption (b′) by

$$\gamma' \triangleq f - \frac{1}{\log \frac{1}{p_{S,\max}}} \left(\log \frac{1}{p_{\min}} - \log \frac{1}{p_{\max}}\right). \tag{122}$$

For any $t \in [d_{n'}, d_{n'+1})$, $n' = 1, 2, \ldots$, the source prior $\theta_{i^*}(y^{t-1})$ for any $i^* \in [q]^{N(t)}$, $y^{t-1} \in \mathcal{Y}^{t-1}$ in (121) satisfies

$$\begin{aligned}
&\limsup_{n'\to\infty} \frac{1}{t} \log \theta_{i^*}(y^{t-1}) \\
\le\ & -\left(\liminf_{n'\to\infty} \frac{N(t)}{t} \log \frac{1}{p_{S,\max}}\right) + \log \frac{p_{\max}}{p_{\min}} && \text{(123a)} \\
\le\ & -\left(\liminf_{n'\to\infty} \frac{N(d_{n'})}{d_{n'+1}-1} \log \frac{1}{p_{S,\max}}\right) + \log \frac{p_{\max}}{p_{\min}} && \text{(123b)} \\
\le\ & -f \log \frac{1}{p_{S,\max}} + \log \frac{p_{\max}}{p_{\min}} && \text{(123c)} \\
=\ & -\gamma' \log \frac{1}{p_{S,\max}}, && \text{(123d)}
\end{aligned}$$

where (123a) is by taking the logarithm, dividing by t, and taking n' to infinity on both sides of (121); (123b) holds since $\frac{N(d_{n'})}{d_{n'+1}-1} \le \frac{N(t)}{t}$ for all $t \in [d_{n'}, d_{n'+1})$; (123c) holds due to Lemma 8 (i) and the fact that $\left\{\frac{N(d_{n'})}{d_{n'}}\right\}_{n'=1}^{\infty}$ is a subsequence of $\left\{\frac{n}{t_n}\right\}_{n=1}^{\infty}$; (123d) holds by plugging (122) into (123c). Rearranging terms of (123), we conclude that the maximum source prior (119) converges pointwise: for any $y^{t-1} \in \mathcal{Y}^{t-1}$,

$$\lim_{n'\to\infty} \theta_{i^*}(y^{t-1}) = 0, \ \forall t \in [d_{n'}, d_{n'+1}), \tag{124}$$

where $t \to \infty$ for any $t \in [d_{n'}, d_{n'+1})$ as $n' \to \infty$.

The convergence of the source prior (124) implies the convergence of the group prior. The partitioning rule in (24) ensures that the group prior $\pi_x(y^{t-1})$, $\forall x \in \mathcal{X}$ is simultaneously upper and lower bounded as

$$\begin{aligned}
P_X^*(x) + \theta_{i^*}(y^{t-1}) &\ge \pi_x(y^{t-1}) && \text{(125a)} \\
&\ge P_X^*(x) - |\mathcal{X}| \theta_{i^*}(y^{t-1}), && \text{(125b)}
\end{aligned}$$

where the upper bound (125a) holds by (24) and (119); the lower bound (125b) holds since all $|\mathcal{X}|$ group priors are upper bounded by (125a). From (124) and (125), we conclude that for all $x \in \mathcal{X}$, $y^{t-1} \in \mathcal{Y}^{t-1}$,

$$\lim_{n'\to\infty} \pi_x(y^{t-1}) = P_X^*(x), \ t \in [d_{n'}, d_{n'+1}). \tag{126}$$

Next, we show the convergence of the group prior (126) implies the convergence of the mutual information $I(X_t; Y_t | Y^{t-1})$ in the sum of (120). We expand the mutual information $I(X_t; Y_t | Y^{t-1})$ as

$$\begin{aligned}
I(X_t; Y_t | Y^{t-1}) = &\sum_{y^{t-1} \in \mathcal{Y}^{t-1}} P_{Y^{t-1}}(y^{t-1}) \\
&\sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_{Y|X}(y|x) \pi_x(y^{t-1}) \log \frac{P_{Y|X}(y|x)}{\sum_{x'\in\mathcal{X}} P_{Y|X}(y|x') \pi_{x'}(y^{t-1})},
\end{aligned} \tag{127}$$

which achieves the channel capacity C if $\pi_x(y^{t-1}) = P_X^*(x)$ for all $x \in \mathcal{X}$, $y^{t-1} \in \mathcal{Y}^{t-1}$. Using (126) and (127), we conclude

$$\lim_{n'\to\infty} I(X_t; Y_t | Y^{t-1}) = C, \ t \in [d_{n'}, d_{n'+1}). \tag{128}$$

Since $I(X_t; Y_t | Y^{t-1}) \le C$, one can write the equivalent of (128) as: for all $\epsilon > 0$, there exists a $n_\epsilon \in \mathbb{N}$, such that for all $n' \ge n_\epsilon$, it holds that

$$I(X_t; Y_t | Y^{t-1}) > C - \epsilon, \ \forall t \in [d_{n'}, d_{n'+1}). \tag{129}$$

We proceed to show (41) using (120) and (129). Dividing both sides of (120) by t_k and taking $k \to \infty$, we lower bound the left side of (120) as

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{t_k} I(S^k; Y^{t_k}) &= \lim_{k\to\infty} \frac{1}{t_k} \sum_{t=1}^{t_k} I(X_t; Y_t | Y^{t-1}) && \text{(130a)} \\
&> \lim_{k\to\infty} \frac{1}{t_k} (t_k - d_{n_\epsilon})(C - \epsilon) && \text{(130b)} \\
&= C - \epsilon, && \text{(130c)}
\end{aligned}$$

where (130b) holds by lower bounding $I(X_t; Y_t | Y^{t-1})$ by (129) for $t > d_{n_\epsilon}$, and lower bounding $I(X_t; Y_t | Y^{t-1})$ by zero for $t \le d_{n_\epsilon}$.

Since ϵ in (130c) can be made arbitrarily small, and $\lim_{k\to\infty} \frac{1}{t_k} I(S^k; Y^{t_k}) \le C$ by data processing, we conclude by the squeeze theorem that under assumption (b′), the instantaneous encoding phase satisfies (41) even if it does not randomize the channel input.

Appendix H
The approximating instantaneous SED rule ensures (54)

We show that the approximating instantaneous SED rule in step (iii′) ensures (54). Since the left side of (54) is equal to

the minimum value on the right side of (53a), it suffices to show that the latter is upper bounded by $\theta_{\mathcal{S}_j}(y^{t-1})$.

We denote

$$c_n \triangleq (\pi_0(y^{t-1}) - n\theta_{\mathcal{S}_j}(y^{t-1})) - (\pi_1(y^{t-1}) + n\theta_{\mathcal{S}_j}(y^{t-1})) \tag{131a}$$

$$= 2\pi_0(y^{t-1}) - 1 - 2n\theta_{\mathcal{S}_j}(y^{t-1}), \tag{131b}$$

and we rewrite the minimization problem in (53a) as

$$\min_{n\in\{\underline{n},\bar{n}\}} |c_n|. \tag{132}$$

By definitions of $\underline{n}$ (53b) and $\bar{n}$ (53c), it holds that $\bar{n}-\underline{n}=1$. Thus

$$c_{\underline{n}} - c_{\bar{n}} = 2\theta_{\mathcal{S}_j}(y^{t-1}). \tag{133}$$

Since $c_{\underline{n}} \geq 0$ and $c_{\bar{n}} \leq 0$, we conclude from (133) that

$$\min\{c_{\underline{n}}, |c_{\bar{n}}|\} \leq \theta_{\mathcal{S}_j}(y^{t-1}), \tag{134}$$

which means that (132) is upper bounded by $\theta_{\mathcal{S}_j}(y^{t-1})$.

Appendix I
Cardinality of Common Randomness

We adapt the proof in [13, Theorem 19] to our codes with instantaneous encoding to show that for any $\langle k, R, \epsilon\rangle$ code with instantaneous encoding that allows $|\mathcal{U}|=\infty$, there exists a $\langle k, R, \epsilon\rangle$ code with instantaneous encoding that allows $|\mathcal{U}| \leq 2$. Fixing a source length k, for $u = 1, 2, \ldots, \infty$, we define $\mathcal{G}_u \subseteq \mathbb{R}^2$ as

$$\mathcal{G}_u \triangleq \{(R,\epsilon) : \exists\ \langle k, R, \epsilon\rangle \text{ code with instantaneous encoding that allows } |\mathcal{U}| \leq u\}. \tag{135}$$

We show that $\mathcal{G}_1$ is a connected set. To see this, we arbitrarily select two elements in $\mathcal{G}_1$, denoted by $\Lambda_1 \triangleq (R_1, \epsilon_1)$ and $\Lambda_2 \triangleq (R_2, \epsilon_2)$. We denote $\Lambda_3 \triangleq (\min\{R_1, R_2\}, \max\{\epsilon_1, \epsilon_2\})$. According to the rate and the error constraints in (14)–(15), $\Lambda_i \in \mathcal{G}_1$, $i \in \{1, 2\}$, indicates that all elements (R, ϵ) that simultaneously satisfy $R \leq R_i$ and $\epsilon \geq \epsilon_i$ belong to $\mathcal{G}_1$ (see the shaded region in Fig. 10). As a result, the line segments $L_i \triangleq \{\lambda\Lambda_i + (1-\lambda)\Lambda_3, \lambda \in [0,1]\}$, $i = 1, 2$, belong to $\mathcal{G}_1$, and the arc $L_1 \cup L_2$ joins Λ_1 and Λ_2.

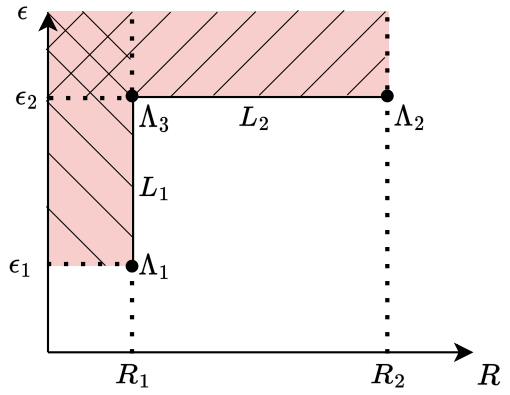


Fig. 10. Elements Λ_1 and Λ_2 are jointed by the arc $L_1 \cup L_2$.

Since $\mathcal{G}_1 \subseteq \mathbb{R}^2$, $\mathcal{G}_1$ is a connected set, and $\mathcal{G}_\infty$ is a convex hull of $\mathcal{G}_1$, by Fenchel-Eggleston-Carathéodory's theorem for connected sets [41, Theorem 18(ii)], any element in $\mathcal{G}_\infty$ can be represented as a convex combination of 2 elements in $\mathcal{G}_1$, in other words, $\mathcal{G}_2 = \mathcal{G}_\infty$.

Appendix J
Zero-error code for degenerate DMCs

In Appendix J-A, we present our zero-error code with instantaneous encoding and common randomness for transmitting k symbols of a DSS over a degenerate DMC. In Appendix J-B, we present the proof that the code in Appendix J-A achieves zero error for any rate asymptotically below $\frac{C}{H}$.

For a degenerate DMC (9) in Theorem 3, we denote by $P_{Y|X}\colon \mathcal{X} \to \mathcal{Y}$ its single-letter transition probability and denote by P_X^* its capacity-achieving distribution. We relabel x in (9a) by ACK, and relabel x' in (9b) by NACK. We denote by $E_G(P_{Y|X}, R_c)$ Gallager's error exponent [33], where R_c is the channel coding rate in nats per channel use[4]. We denote by $R(\ell)$ the rate of the code used in the communication phase of the ℓ-th block, and we denote by $\hat{S}^k(\ell)$ the estimate formed at the end of the communication phase of the ℓ-th block.

A. Zero-error code with instantaneous encoding and common randomness

Similar to [5][6][7][17], our code is divided into blocks. Each block contains a communication phase and a confirmation phase. The first block is different from the blocks after it, since it uses a Shannon limit-achieving code in the communication phase, whereas the blocks after the first block use random coding for all source sequences in alphabet $[q]^k$. We introduce the first block and the ℓ-th block, $\ell \geq 2$, respectively.

The first block is transmitted according to steps i)–ii) below. See Fig. 11 (a) below for the diagram of the time division of transmitted blocks. See Fig. 11 (b)–(c) for the diagram of the first block.

i) Communication phase. The first k symbols S^k of the DSS in Theorem 3 is transmitted via a Shannon limit-achieving code with instantaneous encoding and common randomness at rate $R(1) < \frac{C}{H}$ symbols per channel use. (Such a code has been presented in the proof sketch of Theorem 3. Namely, if $f = \infty$, we use a buffer-then-transmit code that implements the block encoding scheme in [30, Theorem 2]; if $f < \infty$, we precede the block encoding scheme in [30, Theorem 2] by an instantaneous encoding phase that satisfies (41).) At the end of the communication phase, the decoder yields an estimate $\hat{S}^k(1)$ of the source S^k using the channel outputs that it has received in this phase.

ii) Confirmation phase. The encoder knows $\hat{S}^k(1)$ since it knows the channel outputs through the noiseless feedback. The encoder repeatedly transmits ACK if $S^k = \hat{S}^k(1)$, and

[4]For the consistency of notation, we use the same unit (i.e., nats per channel use) for R_c as that in Gallager's paper [33]. The unit of all other rates in this paper is symbols per channel use.

transmits NACK if $S^k \neq \hat{S}^k(1)$, for n_k channel uses. We pick n_k as

$$n_k = \delta k, \tag{136}$$

where $\delta \in (0,1)$ can be made arbitrarily small. At the end of the confirmation phase, if the decoder receives a y,

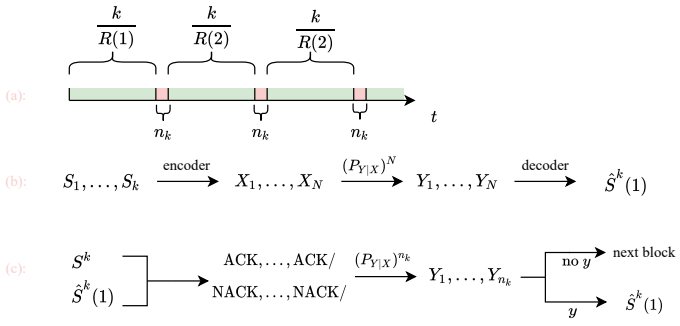


Fig. 11. (a) Time division of the transmitted blocks. The green regions represent the communication phases, and the red regions represent the confirmation phases. The *expected* length of the first communication phase is $\frac{k}{R(1)}$. The length of the ℓ-th communication phase, $\ell \geq 2$, is $\frac{k}{R(2)}$ since the random coding scheme has a fixed length. The length of the confirmation phase is n_k (136). (b) Communication phase of the first block. The codeword length N can be random with expectation $\mathbb{E}[N] = \frac{k}{R(1)}$. (c) Confirmation phase of the first block.

The ℓ-th block, $\ell \geq 2$, is transmitted according to steps iii)–iv) below.

iii) Communication phase. For every sequence in the alphabet $[q]^k$ of S^k, the encoder generates a codeword via random coding according to the capacity-achieving distribution P_X^* at rate $R(2) < \frac{C}{\log q}$ symbols per channel use. At the end of the communication phase, the maximum likelihood (ML) decoder yields an estimate $\hat{S}^k(\ell)$ of the source symbols S^k using the channel outputs that it has received in this phase.

iv) Confirmation phase. The encoder, the decoder, and the stopping rule are the same as those in the first block with $\hat{S}^k(1) \leftarrow \hat{S}^k(\ell)$.

The random codebook is refreshed in every retransmitted block and is known by the decoder. This gives rise to the following observations:

1) The codewords transmitted in the communication phases of the $\ell = 1, 2, \ldots$ blocks are independent from each other;
2) As a result of 1), the channel outputs of the $\ell = 1, 2, \ldots$ blocks are independent from each other;
3) The codewords transmitted in the communication phase of the $\ell = 2, 3, \ldots$ blocks are i.i.d. random vectors. (The codeword in the first block is excluded since the first block need not use random coding in the communication phase);
4) As a result of 3), the channel outputs of the $\ell = 2, 3, \ldots$ blocks are i.i.d. random vectors.

We will use observations 2) and 4) in the proof below.

B. Proof of Theorem 3

Fix any $R < \frac{C}{H}$. We show that by adjusting $R(1)$, the rate of the Shannon limit-achieving code in the communication phase of the first block, to R, the code in Appendix J-A achieves zero error with rate converging to R (56).

We denote by η_k and T_k the stopping time and the number of blocks transmitted after the first block until the stopping time, respectively. We denote by A_ℓ the event that no y is received in the confirmation phase of the ℓ-th block.

Since the decoder will never receive y if ACK is transmitted in the confirmation phase, the error probability of the code in Appendix J-A is zero, i.e.,

$$\mathbb{P}[S^k \neq \hat{S}^k(1 + T_k)] = 0, \tag{137}$$

where $1 + T_k$ represents the total number of blocks transmitted until the stopping time, and T_k is almost surely finite as a result of Lemmas 10 and 11 below. This confirms that the code in Section J-A achieves zero error (15).

To analyze the behavior of the rate $R_k = \frac{k}{\mathbb{E}[\eta_k]}$, we first observe that since the expected length of the first block is $\frac{k}{R(1)} + \delta k$ and the (fixed) length of the ℓ-th block, $\ell \geq 2$, is $\frac{k}{R(2)} + \delta k$, the expected decoding time $\mathbb{E}[\eta_k]$ is equal to

$$\mathbb{E}[\eta_k] = \frac{k}{R(1)} + \delta k + \mathbb{E}[T_k] \left(\frac{k}{R(2)} + \delta k \right). \tag{138}$$

We bound the expected number of blocks T_k transmitted after the first block using Lemmas 10 and 11, stated next.

Lemma 10. *The number of blocks T_k transmitted after the first block satisfies*

$$\mathbb{E}[T_k] \leq \frac{\mathbb{P}[S^k \neq \hat{S}^k(1)] + (1 - P_{Y|X}(y|\mathrm{ACK}))^{\delta k}}{1 - \mathbb{P}[S^k \neq \hat{S}^k(\ell)] - (1 - P_{Y|X}(y|\mathrm{ACK}))^{\delta k}}. \tag{139}$$

Proof. Appendix J-C. ∎

Lemma 11. *Given a DSS with entropy rate $H > 0$ satisfying assumptions (a)–(b) in Theorem 1, the probability of erroneously decoding S^k at the end of the communication phase of the ℓ-th block is upper bounded as*

$$\mathbb{P}[S^k \neq \hat{S}^k(1)] \leq e^{-\frac{k}{R(1)} \left(\frac{C}{1+o(1)} - \frac{H(S^k)}{k} R(1) \right)}, \tag{140a}$$

$$\mathbb{P}[S^k \neq \hat{S}^k(\ell)] \leq e^{-\frac{k}{R(2)} E_G\left(P_{Y|X}, R(2) \log q\right)}, \ \ell = 2, 3, \ldots \tag{140b}$$

Proof. Since the block encoding scheme [30, Theorem 2] satisfies Lemma 5 with $C_1 \leftarrow C$, one can follow Appendices E-B–E-C with $C_1 \leftarrow C$ to upper bound the expected decoding time of the Shannon limit-achieving code in [30, Theorem 2] and thereby obtain (140a). The error probability (140b) holds since the random encoder together with the ML decoder attains Gallager's error exponent [33] for channel coding rate (nats per channel use) below C. This holds regardless of the distribution of the message because Gallager's error exponent holds under the maximum error probability criterion. ∎

Plugging (139) and (140) into the right side of (138), we obtain the asymptotic behavior of the rate as

$$\lim_{k\to\infty} R_k = \lim_{k\to\infty} \frac{k}{\mathbb{E}[\eta_k]} \tag{141a}$$

$$\geq R(1)\frac{1}{1+R(1)\delta}. \tag{141b}$$

Letting $R(1)$ be arbitrarily close to $\frac{C}{H}$ and taking δ to an arbitrarily small number, we conclude (56).

C. Proof of Lemma 10

We establish the pmf of T_k using the probabilities $\mathbb{P}[A_\ell]$, $\ell = 1, 2, \ldots$ The complementary cdf of T_k is given by

$$\mathbb{P}[T_k > 0] \ = \mathbb{P}[A_1], \tag{142}$$

where (142) holds by the definition of A_1 and the stopping rule of the code. We proceed to show the pmf at $T_k = t$, $t \geq 1$ conditioned on $T_k > 0$:

$$\begin{aligned}
&\mathbb{P}[T_k = t | T_k > 0] \\
=\ &\mathbb{P}[A_2 \cap \cdots \cap A_t \cap A_{t+1}^c | A_1] && \text{(143a)} \\
=\ &\left(\prod_{i=2}^{t} \mathbb{P}[A_i | A_1, \ldots, A_{i-1}]\right) \mathbb{P}[A_{t+1}^c | A_1, \ldots, A_t] && \text{(143b)} \\
=\ &(\mathbb{P}[A_2])^{t-1}(1 - \mathbb{P}[A_2]), && \text{(143c)}
\end{aligned}$$

where (143a) is by the stopping rule of the code; (143b) is by expanding (143a); (143c) is by observations 2) and 4) in Appendix J-A: observation 2) implies that event A_i and its complementary event A_i^c are both independent of $A_1, \ldots, A_{i-1}$, $i \geq 2$, observation 4) implies that $\mathbb{P}[A_i] = \mathbb{P}[A_2]$, $i \geq 2$. Since the conditional pmf $\mathbb{P}[T_k = t | T_k > 0]$ in (143) follows a geometric distribution with success probability $1 - \mathbb{P}[A_2]$, its mean is given by

$$\mathbb{E}[T_k | T_k > 0] = \frac{1}{1 - \mathbb{P}[A_2]}. \tag{144}$$

Using (142) and (144), we obtain $\mathbb{E}[T_k]$ as

$$\mathbb{E}[T_k] = \frac{\mathbb{P}[A_1]}{1 - \mathbb{P}[A_2]}. \tag{145}$$

It remains to compute the probability of event A_ℓ in (145) to conclude (139). In the confirmation phase of the ℓ-th block, $\ell = 1, 2, \ldots$, conditioned on $S^k = \hat{S}^k(\ell)$, the probability of event A_ℓ is given by[5]

$$\mathbb{P}[A_\ell | S^k = \hat{S}^k(\ell)] = (1 - P_{Y|X}(y|\text{ACK}))^{\delta k}. \tag{146a}$$

The probability of event A_ℓ is upper bounded as

$$\begin{aligned}
\mathbb{P}[A_\ell] &= \mathbb{P}[A_\ell | S^k \neq \hat{S}^k(\ell)]\mathbb{P}[S^k \neq \hat{S}^k(\ell)] \\
&\quad + \mathbb{P}[A_\ell | S^k = \hat{S}^k(\ell)]\mathbb{P}[S^k = \hat{S}^k(\ell)] && \text{(147a)} \\
&\leq \mathbb{P}[S^k \neq \hat{S}^k(\ell)] + \mathbb{P}[A_\ell | S^k = \hat{S}^k(\ell)], && \text{(147b)}
\end{aligned}$$

where (147b) holds by upper bounding the first and the last probabilities on the right side of (147a) by 1. Plugging the upper bound in (147b) into the right side of (145), we obtain (139).

[5]For practical implementations, one can choose ACK as the channel input that achieves the maximum transition probability $\max_{x\in\mathcal{X}} P_{Y|X}(y|x)$ to increase the probability of receiving a y.

REFERENCES

[1] M. Horstein, "Sequential transmission using noiseless feedback," in *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, July 1963.

[2] M. V. Burnashev and K. S. Zigangirov, "An interval estimation problem for controlled observations," in *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 51–61, 1974.

[3] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," in *IEEE Transactions on Information Theory*, vol. 12, pp. 172–182, Apr. 1966.

[4] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," in *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 118–1222, Mar. 2011.

[5] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," in *Problemy Peredachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1976.

[6] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," in *IEEE Transactions on Information Theory*, vol. 25, pp. 729–733, 1979.

[7] G. Caire, S. Shamai and S. Verdú, "Propagation, feedback and belief," in *4th International Symposium on Turbo Codes and Related Topics; 6th International ITG-Conference on Source and Channel Coding*, pp. 1–6, Apr. 2006.

[8] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen–Shannon divergence: Applications to variable-length coding," in *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2148–2164, Apr. 2015.

[9] A. Antonini, H. Yang and R. D. Wesel, "Low complexity algorithms for transmission of short blocks over the BSC with full feedback," in *2020 IEEE International Symposium on Information Theory*, Los Angeles, CA, USA, pp. 2173–2178, July 2020.

[10] H. Yang, M. Pan, A. Antonini, R. D. Wesel, "Sequential transmission over binary asymmetric channels with feedback," in *IEEE Transactions on Information Theory*, 2022.

[11] N. Guo and V. Kostina, "Instantaneous SED coding over a DMC," in *IEEE International Symposium on Information Theory*, Melbourne, Victoria, Australia, pp. 148–153, July 2021.

[12] C. E. Shannon, "The zero error capacity of a noisy channel," in *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," in *IEEE Transactions on Information Theory,* vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[15] I. Csiszàr, "Joint source-channel error exponent," in *Problems of Control and Information Theory*, vol. 9, pp. 315–328, 1980.

[16] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source–channel coding error exponent for discrete memoryless systems," in *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1450–1468, Apr. 2006.

[17] L. V. Truong and V. Y. F. Tan, "The reliability function of variable-length lossy joint source-channel coding with feedback," in *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5028–5042, Aug. 2019.

[18] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link–Part I: scalar systems," in *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[19] L. J. Schulman, "Coding for interactive communication," in *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1745–1756, Nov. 1996.

[20] R. T. Sukhavasi and B. Hassibi, "Linear error correcting codes with anytime reliability," in *IEEE International Symposium on Information Theory*, St. Petersburg, Russia, pp. 1748–1752, July 2011.

[21] A. Khina, W. Halbawi and B. Hassibi, "(Almost) practical tree codes," in *IEEE International Symposium on Information Theory*, Barcelona, Spain, pp. 2404–2408, Aug. 2016.

[22] A. Lalitha, A. Khina, T. Javidi and V. Kostina, "Real-time binary posterior matching," in *IEEE International Symposium on Information Theory*, Paris, France, pp. 2239–2243, July 2019.

[23] C. Chang and A. Sahai, "Error exponents for joint source-channel coding with delay-constraints," in *44-th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Sep. 2006.

[24] A. Antonini, R. Gimelshein, and R. D. Wesel, "Causal (progressive) encoding over binary symmetric channels with noiseless feedback," in *IEEE International Symposium on Information Theory*, Melbourne, Victoria, Australia, pp. 142–147 July 2021.

[25] S. C. Draper and A. Khisti, "Truncated tree codes for streaming data: Infinite-memory reliability using finite memory," in *IEEE International*

Symposium on Wireless Communication Systems, Aachen, Germany, pp. 136–140, Nov. 2011.

[26] S. C. Draper, C. Chang and A. Sahai, "Lossless coding for distributed streaming sources," in *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1447–1474, Mar. 2014.

[27] S.-H. Lee, V. Y. F. Tan and A. Khisti, "Exact moderate deviation asymptotics in streaming data transmission," in *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2726–2736, May 2017.

[28] S.-H. Lee, V. Y. F. Tan, and A. Khisti, "Streaming data transmission in the moderate deviations and central limit regimes," in *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6816–6830, Dec. 2016.

[29] T. T. Kadota, "On the information stability of stationary ergodic processes." in *SIAM Journal on Applied Mathematics*, vol. 26, no. 1, pp. 176–182, Jan. 1974.

[30] V. Kostina, Y. Polyanskiy and S. Verdú, "Joint source-channel coding with feedback", in *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3502–3515, June 2017.

[31] P. Berlin, B. Nakiboğlu, B. Rimoldi and E. Telatar, "A simple converse of Burnashev's reliability function," in *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3074–3080, July 2009.

[32] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," in *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[33] R. G. Gallager, "Simple derivation of the coding theorem and some applications," in *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[34] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction To Algorithms (3rd ed.)*, Cambridge, MA: The MIT Press, 2009.

[35] R. E. Korf, "From approximate to optimal solutions: a case study of number partitioning", in *14th International Joint Conference on Artificial Intelligence*, vol. 1, pp. 266–272, Aug. 1995.

[36] R. Waeber, P. I. Frazier and S. G. Henderson, "Bisection search with noisy responses," in *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2261–2279, 2013.

[37] T. M. Cover, *Elements of Information Theory*. John Wiley and Sons, 1999.

[38] D. Williams, *Probability With Martingales*. Cambridge University Press, 1991.

[39] G. H. Hardy and J. E. Littlewood, "Some problems of diophantine approximation: Part II. The trigonometrical series associated with the elliptic θ-functions", in *Acta Mathematica*, vol. 37, no. 1, pp. 155–191, 1914.

[40] Maurice Fréchet, "Généralisations du théorème des probabilités totales," in *Fundamenta Mathematicae*, vol. 25, no. 1, pp. 379–387, 1935.

[41] H. G. Eggleston, *Convexity*. Cambridge: Cambridge University Press, 1958.

[42] J. Massey, "Causality, feedback and directed information," in *Proceedings International Symposium on Information Theory and its Applications*, pp. 303–305, Nov. 1990.

[43] G. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," in *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.

[44] N. Guo and V. Kostina, "Reliability function for streaming over a DMC with feedback", in *IEEE International Symposium on Information Theory*, Espoo, Finland, pp. 3204–3209, June 2022.