AALBORG UNIVERSITY

Common-Message Broadcast Channels with Feedback in the Nonasymptotic Regime

Full Feedback

Trillingsgaard, Kasper Fløe; Yang, Wei; Durisi, Giuseppe; Popovski, Petar

Common-Message Broadcast Channels with Feedback in the Nonasymptotic Regime: Full Feedback

Kasper Fløe Trillingsgaard, *Member, IEEE*, Wei Yang, *Member, IEEE*, Giuseppe Durisi, *Senior Member, IEEE*, and Petar Popovski, *Fellow, IEEE*

Abstract—We investigate the maximum coding rate achievable on a two-user broadcast channel for the case where a common message is transmitted with feedback using either fixed-blocklength codes or variable-length codes. For the fixed-blocklength-code setup, we establish nonasymptotic converse and achievability bounds. An asymptotic analysis of these bounds reveals that feedback improves the second-order term compared to the no-feedback case. In particular, for a certain class of antisymmetric broadcast channels, we show that the dispersion is halved. For the variable-length-code setup, we demonstrate that the channel dispersion is zero.

Index Terms—Broadcast channel with common-message, finite blocklength regime, full feedback, channel dispersion, variable-length coding.

I. INTRODUCTION

WE consider a two-user common-message discrete-time memoryless broadcast channel (CM-DMBC) with full feedback. When fixed-length codes are used, it is well-known that feedback does not improve capacity (as in the point-to-point setup), which is given by [2]

$$C = \sup_P \min_{k \in \{1,2\}} I(P, W_k). \tag{1}$$

Here, W_1 and W_2 denote the component channels from the encoder to the two decoders and the supremum is over all input distributions P. In this paper, we show that feedback can improve the speed at which the maximum coding rate approaches capacity as the blocklength increases, and that the improvement in the speed of convergence differs depending on whether one uses fixed-length or variable-length codes.

For the no-feedback case, a CM-DMBC is equivalent to a compound channel, whose capacity was characterized in [3], [4], and whose second-order coding rate (i.e., the second-order

expansion of the maximum coding rate in the limit of large blocklength) was found in [5]. Specifically, it was proven in [5] that the second-order term in the asymptotic expansion of the logarithm of the maximum number of codewords $M^*_{\text{no-f}}(n, \epsilon)$ (here "no-f" stands for no feedback) in the limit of large blocklength n and for a fixed average error probability $\epsilon \in (0, 1)$ is in general a function of both the *dispersion* [6, Eq. (222)] of the individual component channels and of the directional derivatives of their mutual informations computed at the capacity-achieving input distribution (CAID) P^*.

In contrast, when feedback is present, the capacity of a CM-DMBC differs in general from that of the compound channel [7]. This is because, for the compound channel with feedback, the transmitter can send a training sequence and learn the state of the compound channel via the feedback link. It can then adapt the input distribution to maximize the mutual information of that specific component channel. This communication scheme cannot be used in the CM-DMBC because both decoders are required to decode the message.

In the fixed-length-code setup, it is known that feedback does not improve the second-order term for point-to-point discrete memoryless channels (DMCs) under certain symmetry conditions. Specifically, it was shown in [8] that feedback does not improve the second-order term for weakly input-symmetric DMCs. This result was extended to a broader class of DMCs in [9], where it was shown that the same holds when the conditional information variance is constant for all input symbols.

When *variable-length codes* are used, feedback is known to improve the speed at which the maximum coding rate converges to capacity for the point-to-point setup. This was first demonstrated by Burnashev [10], who proved that the error exponent with variable-length codes and full feedback is given by

$$E(R) = \frac{\widetilde{C}_1}{\widetilde{C}} (\widetilde{C} - R) \tag{2}$$

for all $0 < R < \widetilde{C}$. Here, $\widetilde{C}$ denotes the point-to-point capacity and $\widetilde{C}_1$ denotes the maximum relative entropy between conditional output distributions. Yamamoto and Itoh [11] proposed a two-phase scheme that achieves the error exponent in (2) and Berlin *et al.* [12] provided an alternative and simpler proof of the converse. For the CM-DMBC, Truong and Tan [13] established upper and lower bounds on the error exponent using arguments similar to those of Burnashev, Yamamoto, and Itoh. The bounds in [13] are tight when the broadcast channel is stochastically degraded.

For the regime of large average blocklength and fixed error probability, it was shown in [8] that there is no square-root penalty in the asymptotic expansion of the maximum coding rate achievable with variable-length codes for point-to-point DMCs, a result known as *zero-dispersion*. Moreover, this fast convergence to capacity can be achieved using only *stop feedback* (also known as *decision feedback*). Namely, the feedback link is used only to stop the transmitter. For CM-DMBCs, however, stop feedback is not sufficient to achieve zero dispersion. Specifically, we recently showed that the asymptotic expansion of the maximum coding rate with variable-length codes and stop feedback contains a square-root penalty term, provided that some mild technical conditions are satisfied [14].

Contributions: We show that the presence of full feedback improves the second-order term in the asymptotic expansions of the maximal coding rate for a general class of CM-DMBCs. Specifically, we prove nonasymptotic achievability and converse bounds on the maximal number of codewords $M_\text{f}^*(n, \epsilon)$ and $M_\text{vf}^*(\mathsf{n}, \epsilon)$ (here "f" and "vf" stand for *feedback* and *variable-length plus feedback*, respectively) that can be transmitted over a CM-DMBC with feedback and with reliability $1 - \epsilon$ using fixed-length codes and variable-length codes, respectively. Here, n denotes the blocklength in the fixed-length setup whereas n stands for the average blocklength in the variable-length setup. Through an asymptotic analysis of our bounds, we obtain the following results.

- For the fixed-length case, we show that the second-order term depends on the directional derivatives of the mutual informations of the two component channels evaluated at the CAID through a channel-dependent parameter that we shall denote by $\eta \in (0, 1)$. Specifically, we show that the second-order term is larger or equal to $\sqrt{\eta^2 V_1 + (1 - \eta)^2 V_2} Q^{-1}(2\epsilon)$ (here, $\epsilon < 1/2$), where V_1 and V_2 denote the conditional information variances of the two component channels evaluated at P^*. Furthermore, if the weighted sum of the conditional information variances given a specific input symbol x takes the same value for all x, the lower bound $\sqrt{\eta^2 V_1 + (1 - \eta)^2 V_2} Q^{-1}(2\epsilon)$ is tight. For CM-DMBCs satisfying $V_1 = V_2$ and a certain "antisymmetric property" (which will be made precise in Definition 3 on page 6), the parameter η is equal to $1/2$ and the second-order term simplifies to $\sqrt{V_1/2} Q^{-1}(2\epsilon)$, thereby demonstrating that, in this case, full feedback halves the dispersion compared to the no-feedback case.
- For the variable-length case, we show that full feedback yields zero dispersion. In light of our previous result for stop-feedback codes in [14], this novel result shows that variable-length codes combined with full feedback attain a second-order term that is strictly better than the one of variable-length codes with stop-feedback.

Intuition: Feedback allows the encoder to compute the accumulated information density at both decoders and to adapt the input distribution accordingly. Specifically, the encoder makes small adjustments to the input distribution in order to favor the decoder with the smallest information density and to drive both information densities close to their arithmetic mean. This strategy is helpful in both the fixed-length coding and variable-length

coding setup as explained next.

Using this strategy in the fixed-length setup, one transforms the problem of computing the maximum coding rate into that of computing the ϵ-quantile of the arithmetic mean of the two information densities. The desired result follows because the arithmetic mean of the two information densities has variance $V_1/2$ when $V_1 = V_2$. Hence, the dispersion is halved. Furthermore, the improvement from $Q^{-1}(\epsilon)$ to $Q^{-1}(2\epsilon)$ is achieved as follows: If the arithmetic mean of the information densities is below a suitably chosen threshold shortly before the end of the transmission, the encoder changes the input distribution to either the CAID of W_1 or to the one of W_2. In this way, it can ensure that at least one of the two decoders is successful with high probability.

When variable-length codes are used, each decoder attempts decoding as soon as the information density at that decoder exceeds a certain threshold. Since the information densities at both decoders are driven towards their arithmetic mean using the above mentioned feedback scheme, both decoders attempt decoding at approximately the same time. This means that the stochastic overshoot of the arithmetic mean of the two information densities that results in the square-root penalty when fixed-length codes are used can be virtually eliminated by using variable-length codes. Hence, the dispersion is zero as in the single-user variable-length setup [8]. Note that the zero-dispersion result depends crucially on the availability of full feedback at the encoder, which allows the encoder to steer the information densities between the transmitter and the two users towards their arithmetic mean. This is not possible if the encoder is only provided with stop feedback.

Organization: In Section II, we introduce the system model. Our nonasymptotic bounds and asymptotic expansions for fixed-length codes with feedback and variable-length codes with feedback are presented in Section III and Section IV, respectively. Section V concludes the paper.

Notation: We denote vectors by boldface letters; their entries are denoted by roman letters. Upper case, lower case, and calligraphic letters denote random variables (RVs), deterministic quantities, and sets, respectively. The cardinality of a set is denoted by $|\cdot|$. The inner product between two vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, and the l_1-norm and the l_2-norm are denoted by $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively. The vectors $\mathbf{0}_d$ and $\mathbf{1}_d$ denote a d-dimensional column vectors whose entries are all zero and all one, respectively. We let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_+$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the set of natural numbers, the set of integers, the set of nonnegative integers, the set of real numbers, and the set of nonnegative real numbers, respectively. We also denote by $\mathbb{R}_0^d$, the set of d-dimensional real vectors whose entries sum to zero. We let $\mathbb{c}$ denote an arbitrary positive constant whose value may change at each occurrence and $|\cdot|_+$ denotes $\max\{0, \cdot\}$. Furthermore, we denote the base e logarithm by $\log(\cdot)$. Throughout the paper, the index k is used to designate one of the two decoders. Hence, it always belongs to the set $\{1, 2\}$, although this is sometimes not explicitly mentioned. We also set $\bar{k} = 3 - k$. For a probability distribution P, we let P^n be the joint probability distribution of the vector $[X_1, \ldots, X_n]$, whose entries $\{X_i\}$ are independently and identically distributed (i.i.d.) according to P. We let $\mathbb{E}[\cdot]$ and $\mathbb{Var}[\cdot]$ denote the expectation and the variance, respectively.

For a probability distribution Q, we denote the expectation with respect to Q by $\mathbb{E}_Q[\cdot]$. For a RV X with probability distribution P on $\mathcal{X}$ and a RV Y which is the output of a channel transition matrix $W : \mathcal{X} \mapsto \mathcal{Y}$, we let $PW(\cdot)$ denote the induced output distribution on $\mathcal{Y}$; furthermore, we denote the joint probability distribution of (X, Y) by $P \times W$. Given two probability distributions P and Q on a common measurable space, we define, for every $\alpha \in (0, 1)$, the Neyman-Pearson function $\beta_{1-\alpha}(P, Q)$ as the minimum type-II error probability of a binary hypothesis test between P and Q subject to the constraint that the type-I error probability does not exceed α. Finally, for two functions $f(\cdot)$ and $g(\cdot)$, the notation $f(x) = \mathcal{O}(g(x))$, as $x \to \infty$, means that $\lim_{x\to\infty} |f(x)/g(x)| < \infty$ and $f(x) = o(g(x))$, as $x \to \infty$, means that $\lim_{x\to\infty} |f(x)/g(x)| = 0$.

II. System Model

We consider a CM-DMBC with input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$.[1] We assume that the channel outputs at any given time i are conditionally independent given the input, namely

$$W(y_{1,i}, y_{2,i}|x_i) \triangleq W_1(y_{1,i}|x_i)W_2(y_{2,i}|x_i). \quad (3)$$

The capacity of the CM-DMBC is given in (1). We shall assume that it is achieved by P^* and that this CAID is unique.[2] We denote the capacities of the two component channels by C_1 and C_2, respectively.

For every input distribution P and every $n \in \mathbb{N}$, we denote the information density between the vectors x^n and y_k^n as

$$\imath_{P,W_k}(x^n; y_k^n) \triangleq \sum_{i=1}^{n} \log \frac{W_k(y_{k,i}|x_i)}{PW_k(y_{k,i})}. \quad (4)$$

Furthermore, we let

$$I_k(P) \triangleq \mathbb{E}_{P \times W_k}[\imath_{P,W_k}(X; Y_k)] \quad (5)$$

be the mutual information and

$$V_k(P) \triangleq \mathbb{E}_P[\mathbb{V}\mathrm{ar}_{W_k}[\imath_{P,W_k}(X; Y_k)|X]] \quad (6)$$

be the conditional information variance. Finally, we let $V_k \triangleq V_k(P^*)$ and we let $\langle \nabla I_k, \mathbf{v} \rangle$ denote the directional derivative of the mutual information $I_k(\cdot)$ along the direction $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$ at the point P^*

$$\langle \nabla I_k, \mathbf{v} \rangle \triangleq \sum_{x \in \mathcal{X}} v_x D(W_k(\cdot|x) \| P_{Y_k}^*) \quad (7)$$

(the entries of $\mathbf{v}$ should sum to zero because the mutual information is only defined on the $|\mathcal{X}|$-dimensional probability simplex).[3] Here, $D(\cdot\|\cdot)$ denotes the relative entropy, $P_{Y_k}^* = P^*W_k(\cdot)$ is the capacity-achieving output distribution (CAOD). Without loss of generality, we have also assumed in (7) that $\mathcal{X} = \{1, \ldots, N\}$ where $N = |\mathcal{X}|$.

Besides (3) and the uniqueness of P^*, we also assume that the channel laws satisfy the following conditions

1) $P^*(x) > 0$ for all $x \in \mathcal{X}$.
2) $V_k > 0$ for $k \in \{1, 2\}$.
3) $C_k > C$ for $k \in \{1, 2\}$.

The following lemma provides a relation between $\langle \nabla I_1, \cdot \rangle$ and $\langle \nabla I_2, \cdot \rangle$ that will be important in the asymptotic analyses of our converse and achievability bounds.

Lemma 1: There exists a unique constant $\eta \in (0, 1)$ such that, for all $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$,

$$\eta \langle \nabla I_1, \mathbf{v} \rangle + (1 - \eta) \langle \nabla I_2, \mathbf{v} \rangle = 0. \quad (8)$$

Proof: Let $\overline{\nabla}_k \in \mathbb{R}^{|\mathcal{X}|-1}$ be the gradient of $\langle \nabla I_k, [\bar{\mathbf{v}}, -\langle \mathbf{1}_{|\mathcal{X}|-1}, \bar{\mathbf{v}} \rangle] \rangle$ with respect to $\bar{\mathbf{v}} \in \mathbb{R}^{|\mathcal{X}|-1}$. Note that the concavity of the mutual information and the assumption $C_k > C$ imply that $\overline{\nabla}_k$ cannot be the zero vector. Since P^* maximizes $P \mapsto \min_k I_k(P)$, it follows that $\langle \overline{\nabla}_1, \bar{\mathbf{v}} \rangle \langle \overline{\nabla}_2, \bar{\mathbf{v}} \rangle \leq 0$ for all $\bar{\mathbf{v}}$. Suppose on the contrary that there exists a $\tilde{\mathbf{v}} \in \mathbb{R}^{|\mathcal{X}|-1}$ such that $\langle \overline{\nabla}_1, \tilde{\mathbf{v}} \rangle \langle \overline{\nabla}_2, \tilde{\mathbf{v}} \rangle > 0$. Then, by differentiability of $P \mapsto I_k(P)$ at P^*, there exists a sufficiently small constant $\xi \neq 0$ satisfying

$$\min_k I_k(P^* + \xi[\tilde{\mathbf{v}}, -\langle \mathbf{1}_{|\mathcal{X}|-1}, \tilde{\mathbf{v}} \rangle])$$

$$\geq \min_k \left\{ I(P^*, W_k) + \frac{\xi}{2} \langle \overline{\nabla}_k, \tilde{\mathbf{v}} \rangle \right\} \quad (9)$$

$$> \min_k I(P^*, W_k). \quad (10)$$

Note that the constant ξ is positive if $\langle \overline{\nabla}_k, \tilde{\mathbf{v}} \rangle > 0$ for $k \in \{1, 2\}$ and negative if $\langle \overline{\nabla}_k, \tilde{\mathbf{v}} \rangle < 0$ for $k \in \{1, 2\}$. Now, (10) contradicts that P^* maximizes $P \mapsto \min_k I(P, W_k)$, implying that $\langle \overline{\nabla}_1, \bar{\mathbf{v}} \rangle \langle \overline{\nabla}_2, \bar{\mathbf{v}} \rangle \leq 0$ for all $\bar{\mathbf{v}}$ as desired. As a result, we have that

$$\frac{\overline{\nabla}_1}{\|\overline{\nabla}_1\|_2} + \frac{\overline{\nabla}_2}{\|\overline{\nabla}_2\|_2} = \mathbf{0}_{|\mathcal{X}|-1}. \quad (11)$$

Indeed, assume that

$$\frac{\overline{\nabla}_1}{\|\overline{\nabla}_1\|_2} + \frac{\overline{\nabla}_2}{\|\overline{\nabla}_2\|_2} = \mathbf{w} \neq \mathbf{0}_{|\mathcal{X}|-1}. \quad (12)$$

Then, it follows from (12) and the Cauchy-Schwarz inequality that

$$\langle \overline{\nabla}_1, \mathbf{w} \rangle = \|\overline{\nabla}_1\|_2 + \frac{\langle \overline{\nabla}_1, \overline{\nabla}_2 \rangle}{\|\overline{\nabla}_2\|_2} \geq 0 \quad (13)$$

and similarly that $\langle \overline{\nabla}_2, \mathbf{w} \rangle \geq 0$. Here, the inequalities hold with equality if and only if (11) holds, i.e., when $\overline{\nabla}_1$ and $\overline{\nabla}_2$ are linearly dependent and point in opposite directions. Hence, since $\mathbf{w} \neq \mathbf{0}_{|\mathcal{X}|-1}$, we conclude that $\langle \overline{\nabla}_1, \mathbf{w} \rangle > 0$ and $\langle \overline{\nabla}_2, \mathbf{w} \rangle > 0$. But this contradicts that $\langle \overline{\nabla}_1, \mathbf{w} \rangle \langle \overline{\nabla}_2, \mathbf{w} \rangle \leq 0$. Now, (8) follows from (11) with $\eta = (1/\|\overline{\nabla}_1\|_2)/(1/\|\overline{\nabla}_1\|_2 + 1/\|\overline{\nabla}_2\|_2)$. Uniqueness of η follows trivially. ∎

We use (8) to define the channel-dependent constant $\eta \in (0, 1)$ as follows:

$$\eta = \frac{\langle \nabla I_2, \mathbf{v} \rangle}{\langle \nabla I_2, \mathbf{v} \rangle - \langle \nabla I_1, \mathbf{v} \rangle}. \quad (14)$$

Here, $\mathbf{v}$ is an arbitrary vector in $\mathbb{R}_0^{|\mathcal{X}|}$ satisfying $\langle \nabla I_1, \mathbf{v} \rangle \neq 0$. Note that by Lemma 1, the value of η does not depend on $\mathbf{v}$. For notational convenience, we set $\eta_1 = \eta$ and $\eta_2 = 1 - \eta$ in the remainder of the paper.

---

[1] The assumption $\mathcal{Y} = \mathcal{Y}_1 = \mathcal{Y}_2$ comes without loss of generality.

[2] The assumption that P^* is unique is used in the proof of our converse results. Our achievability results can be readily extended to CM-DMBCs with nonunique CAIDs.

[3] It is sometimes required in literature that $\mathbf{v}$ is a unit vector, but we do not assume this here.
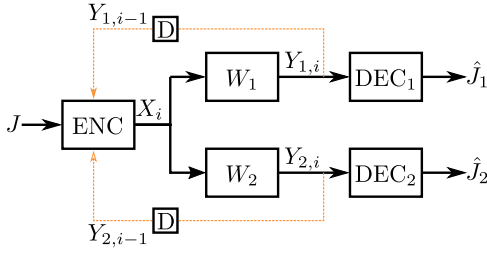
Fig. 1. The system model considered in this paper. Here, D stands for a delay of one unit of time.

Next, we define the notions of fixed-length feedback (FLF) codes and variable-length feedback (VLF) codes.

Definition 1: An (n, M, ϵ)-FLF code for the CM-DMBC consists of:

1) n encoding functions $f_i : \mathcal{M} \times \mathcal{Y}^{i-1} \times \mathcal{Y}^{i-1} \mapsto \mathcal{X}$, $i = 1, \ldots, n$ (possibly randomized), mapping the message J, drawn uniformly from $\mathcal{M} \triangleq \{1, \ldots, M\}$, and the past channel outputs Y_1^{i-1}, Y_2^{i-1} to the channel input $X_i = f_i(J, Y_1^{i-1}, Y_2^{i-1})$.

2) Two decoders $g_k : \mathcal{Y}^n \mapsto \mathcal{M}$ satisfying

$$\mathbb{P}\big[g_k(Y_k^n) \neq J\big] \leq \epsilon, \quad k \in \{1, 2\}. \quad (15)$$

The maximum code size achievable with blocklength n and average error probability not exceeding ϵ is denoted by

$$M_{\mathrm{f}}^*(n, \epsilon) \triangleq \max\{M : \exists (n, M, \epsilon)\text{-FLF code}\}. \quad (16)$$

Definition 2: An $(\mathbb{n}, M, \epsilon)$-VLF code for the CM-DMBC consists of:

1) A RV $U \in \mathcal{U}$ with $|\mathcal{U}| \leq 3$, which is known at both the encoder and the decoders.

2) A sequence of encoding functions $f_i : \mathcal{U} \times \mathcal{M} \times \mathcal{Y}^{i-1} \times \mathcal{Y}^{i-1} \mapsto \mathcal{X}$, each one mapping the message J, drawn uniformly at random from the set $\mathcal{M}$, the past channel outputs Y_1^{i-1}, Y_2^{i-1}, and the auxiliary RV U to the channel input $X_i = f_i(U, J, Y_1^{i-1}, Y_2^{i-1})$.

3) Two nonnegative integer-valued RVs τ_1 and τ_2 that are stopping times with respect to the filtrations $\mathcal{F}_{1,i} \triangleq \sigma(U, Y_1^i)$ and $\mathcal{F}_{2,i} \triangleq \sigma(U, Y_2^i)$, respectively, and satisfy

$$\mathbb{E}\big[\max_k \tau_k\big] \leq \mathbb{n}. \quad (17)$$

4) A sequence of decoders $g_{k,i} : \mathcal{U} \times \mathcal{Y}^i \mapsto \mathcal{M}$ satisfying

$$\mathbb{P}\big[J \neq g_{k,\tau_k}(U, Y_k^{\tau_k})\big] \leq \epsilon, \quad k \in \{1, 2\}. \quad (18)$$

The maximum code size achievable with average blocklength $\mathbb{n}$ and average error probability not exceeding ϵ is denoted by

$$M_{\mathrm{vf}}^*(\mathbb{n}, \epsilon) \triangleq \max\{M : \exists (\mathbb{n}, M, \epsilon)\text{-VLF code}\}. \quad (19)$$

We depict our system model in Fig. 1.

III. FIXED-LENGTH FEEDBACK CODES

In this section, we present nonasymptotic achievability and converse bounds on $M_{\mathrm{f}}^*(n, \epsilon)$ for FLF codes. These bounds are then analyzed in the large-n regime. Under some technical assumptions, the asymptotic expansions of our bounds are shown to match up to the second order. We then introduce a class of antisymmetric CM-DMBCs for which the second-order term in the asymptotic expansion of $M_{\mathrm{f}}^*(n, \epsilon)$ takes on a simple expression, which allows for an insightful comparison with the no-feedback case.

A. Nonasymptotic bounds

We first state a Verdú-Han-type converse bound for FLF codes.

Theorem 2: Every (n, M, ϵ)-FLF code for the CM-DMBC satisfies[4]

$$\epsilon \geq \frac{1}{2}\mathbb{P}_{J,Y_1^n,Y_2^n}\left[\sum_k \eta_k \sum_{i=1}^n \imath_{P^*, W_k}\left(Y_{k,i}; f_i(J, Y_1^{i-1}, Y_2^{i-1})\right)\right.$$
$$\left. \leq \log M - \lambda\right] - e^{-\lambda} \quad (20)$$

for every $\lambda > 0$, where

$$\mathbb{P}_{J,Y_1^n,Y_2^n}(j, y_1^n, y_2^n)$$
$$\triangleq \frac{1}{M}\prod_{i=1}^n\prod_k W_k(y_{k,i}|f_i(j, y_1^{i-1}, y_2^{i-1})). \quad (21)$$

Proof: Fix an (n, M, ϵ)-FLF code with encoding functions $\{f_i\}$ and decoding functions $\{g_k\}$. Define the following auxiliary probability distributions on $\mathcal{M} \times \mathcal{Y}^n \times \mathcal{Y}^n$

$$\mathbb{Q}_{J,Y_1^n,Y_2^n}^{(k)}(j, y_1^n, y_2^n)$$
$$\triangleq \frac{1}{M}\prod_{i=1}^n W_{\bar{k}}(y_{\bar{k},i}|f_i(j, y_1^{i-1}, y_2^{i-1}))P^*W_k(y_{k,i}). \quad (22)$$

Recall that $\bar{k} = 1$ when $k = 2$ and $\bar{k} = 2$ when $k = 1$. When $(J, Y_1^n, Y_2^n) \sim \mathbb{P}_{J,Y_1^n,Y_2^n}$, both decoders have an average error probability not exceeding ϵ because of (15). When $(J, Y_1^n, Y_2^n) \sim \mathbb{Q}_{J,Y_1^n,Y_2^n}^{(k)}$, the average error probability at decoder k is $1 - 1/M$. The meta-converse theorem in [6, Th. 27] and the inequality [6, Eq. (106)] then imply that

$$\frac{1}{M} \geq \beta_{1-\epsilon}(\mathbb{P}_{J,Y_1^n,Y_2^n}, \mathbb{Q}_{J,Y_1^n,Y_2^n}^{(k)}) \quad (23)$$

$$\geq \frac{1}{\gamma}\left(\mathbb{P}_{J,Y_1^n,Y_2^n}\left[\log\frac{\mathrm{d}\mathbb{P}_{J,Y_1^n,Y_2^n}}{\mathrm{d}\mathbb{Q}_{J,Y_1^n,Y_2^n}^{(k)}} \leq \log\gamma\right] - \epsilon\right) \quad (24)$$

for every $\gamma > 0$ and $k \in \{1, 2\}$. Let

$$A_k(j, y_1^n, y_2^n) \triangleq \sum_{i=1}^n \imath_{P^*, W_k}(f_i(j, y_1^{i-1}, y_2^{i-1}); y_{k,i}). \quad (25)$$

[4]Recall that $\eta_1 = \eta$ and $\eta_2 = 1 - \eta$, where η was defined in (14).

By setting $\log \gamma \triangleq \log M - \lambda$ and by using (21) and (22) in (24), we obtain

$$\epsilon \geq \max_k \mathbb{P}_{J,Y_1^n,Y_2^n}\left[A_k(J,Y_1^n,Y_2^n) \leq \log M - \lambda\right] - e^{-\lambda}. \tag{26}$$

The desired result (20) is obtained from (26) through the following chain of inequalities:

$$\epsilon + e^{-\lambda}$$
$$\geq \frac{1}{2}\sum_k \mathbb{P}_{J,Y_1^n,Y_2^n}\left[A_k(J,Y_1^n,Y_2^n) \leq \log M - \lambda\right] \tag{27}$$

$$\geq \frac{1}{2}\mathbb{P}_{J,Y_1^n,Y_2^n}\left[\bigcup_k \left\{A_k(J,Y_1^n,Y_2^n) \leq \log M - \lambda\right\}\right] \tag{28}$$

$$\geq \frac{1}{2}\mathbb{P}_{J,Y_1^n,Y_2^n}\left[\sum_k \eta_k A_k(J,Y_1^n,Y_2^n) \leq \log M - \lambda\right]. \tag{29}$$

Here, (27) follows because $\max_k a_k \geq \frac{1}{2}\sum_k a_k$ for every pair $(a_1,a_2) \in \mathbb{R}_+^2$, (28) follows from the union bound, and (29) follows because $\sum_k \eta_k b_k \leq \zeta$ imply that either $b_1 \leq \zeta$ or $b_2 \leq \zeta$ for every triple $(b_1,b_2,\zeta) \in \mathbb{R}^3$. ∎

Before stating our achievability bound, we need to introduce some notation. We let $S \geq 2$, L, and m denote arbitrary integers. We also let $\mathcal{S} \triangleq \{1,\ldots,S\}$ and $\mathcal{L} \triangleq \{1,\ldots,L\}$. Furthermore, we index each element $\varkappa$ of $\mathcal{X}^{SLm}$ as follows:

$$\varkappa = (\varkappa_1(1),\ldots,\varkappa_1(S),\ldots\ldots,\varkappa_L(1),\ldots,\varkappa_L(S)). \tag{30}$$

Here, $\varkappa_\ell(s) = (\varkappa_{\ell,1}(s),\ldots,\varkappa_{\ell,m}(s)) \in \mathcal{X}^m$. For a given vector $b^\ell \in \mathcal{S}^\ell$, we let $\varkappa^\ell(b^\ell) \triangleq (\varkappa_1(b_1),\ldots,\varkappa_\ell(b_\ell))$. We use a similar notation for the elements in the set $\mathcal{Y}^{SLm}$. Finally, we define the minimum error probability achievable on an CM-DMBC using fixed-length codes with no feedback of blocklength n_B and with S^L codewords:

$$\epsilon^*(n_B, S^L)$$
$$\triangleq \min\{\epsilon : \exists (n_B, S^L, \epsilon) \text{ finite-length no-feedback code}\}. \tag{31}$$

The intuition behind our achievability bound is as follows. The codebook consists of codewords of length Lm. Each codeword is divided into L blocks of length m, which are of constant composition. Within each block ℓ, feedback is used to compute the index B_ℓ, which selects one out of the S available subcodewords (the vectors $\varkappa_\ell(B_\ell)$, $B_\ell = 1,\ldots,S$). We then communicate the sequence B^L to the decoders using a code with S^L codewords of length n_B. The following achievability bound for FLF codes is then a consequence of the achievability bound [5, Th. 3] for compound DMCs without feedback.

Theorem 3: Let $P_1,\ldots,P_S$ be types of sequences in $\mathcal{X}^m$, let $\tau \in (0,\epsilon)$, let $\zeta > 0$, and let $\{h_\ell\}_{\ell=1}^L$ be arbitrary mappings from $\mathcal{X}^{(\ell-1)m} \times \mathcal{Y}^{(\ell-1)m} \times \mathcal{Y}^{(\ell-1)m}$ to $\mathcal{S}$. Then

$$\log M_{\mathrm{f}}^*(Lm + n_B, \epsilon + \epsilon^*(n_B, S^L))$$
$$\geq \sup\Big\{\gamma : \max_k \mathbb{P}\big[\imath_k(\varkappa^L(B^L);\mathbb{Y}_k^L(B^L)) \leq \gamma\big]$$
$$< \epsilon - \tau - e^{-\zeta}\Big\}$$
$$+ \log\frac{\tau}{2} - SL|\mathcal{X}|\log(1+m) - L\log S - \zeta. \tag{32}$$

Here, $\varkappa \in \mathcal{X}^{SLm}$ is an arbitrary element in the set

$$\mathbb{F} \triangleq \Big\{\varkappa \in \mathcal{X}^{SLm} : \varkappa_\ell(s) \text{ has type } P_s, \quad \forall s \in \mathcal{S}, \forall \ell \in \mathcal{L}\Big\} \tag{33}$$

and $\mathbb{P}_{\mathbb{Y}_1,\mathbb{Y}_2,B^L}$ denotes the probability distribution of $(\mathbb{Y}_1,\mathbb{Y}_2,B^L)$ on $\mathbb{B} \times \mathbb{B} \times \mathcal{S}^L$ defined by

$$\mathbb{P}_{\mathbb{Y}_1,\mathbb{Y}_2,B^L}(\mathbb{y}_1,\mathbb{y}_2,b^L) \triangleq$$
$$\prod_{\ell=1}^L\left(\prod_{s=1}^S\prod_k W_k^m(\mathbb{y}_{k,\ell}(s)|\varkappa_\ell(s))\right)$$
$$\times \mathbb{1}\Big\{h_\ell\big(\varkappa^{\ell-1}(b^{\ell-1}),\mathbb{y}_1^{\ell-1}(b^{\ell-1}),\mathbb{y}_2^{\ell-1}(b^{\ell-1})\big) = b_\ell\Big\} \tag{34}$$

for $\mathbb{y}_{k,\ell}(b) \in \mathcal{Y}^m$. Finally,

$$\imath_k(\varkappa_\ell(b_\ell);\mathbb{y}_{k,\ell}(b_\ell)) \triangleq \sum_{i=1}^m \imath_{P_{b_\ell},W_k}(\varkappa_{\ell,i}(b_\ell);\mathbb{y}_{k,\ell,i}(b_\ell)) \tag{35}$$

and $\imath_k(\cdot;\cdot)$ in (32) is defined as follows:

$$\imath_k(\varkappa^L(b^L);\mathbb{y}_k^L(b^L)) \triangleq \sum_{\ell=1}^L \imath_k(\varkappa_\ell(b_\ell);\mathbb{y}_{k,\ell}(b_\ell)). \tag{36}$$

Proof: See Appendix I. ∎

Remark 1: The achievability bound in Theorem 3 holds also under the more stringent error probability constraint $\max_{j \in \mathcal{M}} \mathbb{P}[g_k(Y_k^n) \neq j|J = j] \leq \epsilon$.

Remark 2: The term $-SL|\mathcal{X}|\log(1+m)$ in (32) represents the penalty due to the use of constant composition codes.

Remark 3: We shall apply Theorem 3 in the following way: The types P_1 and P_2 are chosen close to P^* but such that they slightly favor decoder 1 and decoder 2, respectively. The types P_3, P_4, and P_5 are set equal to P^*, P_1^*, and P_2^*, respectively. By the end of the $(\ell-1)$th block, for $\ell \in \{1,\ldots,L-1\}$, the encoder computes the information densities at the decoders and chooses $B_\ell = k$ if decoder k has accumulated the smallest amount of information density. By the end of the $(L-1)$th block, the encoder chooses $B_L = 3$ if the arithmetic mean of the information densities at the decoders exceeds a certain threshold and it sets B_L equal to 4 or 5, each with probability $1/2$, otherwise.

B. Asymptotic analysis

By analyzing Theorem 2 and Theorem 3 in the large-blocklength limit, we will next establish asymptotic expansions for our converse and achievability bounds that match up to the second order provided that the assumption (38) is satisfied. In particular, we show in the next theorem that in this case the channel dispersion [6, Def. 1] is

$$V \triangleq \eta^2 V_1 + (1-\eta)^2 V_2. \tag{37}$$

Theorem 4: Suppose that

$$\eta^2 \mathbb{Var}[\imath_{P^*,W_1}(x;Y_1)|X = x]$$
$$+ (1-\eta)^2 \mathbb{Var}[\imath_{P^*,W_2}(x;Y_2)|X = x] = V \tag{38}$$

for all $x \in \mathcal{X}$, where V is defined in (37). Then, for every $\epsilon \in (0,1/2)$, we have

$$\log M_{\mathrm{f}}^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(2\epsilon) + \mathcal{O}(n^{1/3}\log n). \tag{39}$$

Proof: See Appendix II for the proof of the converse part and Appendix III for the proof of the achievability part. ∎

Remark 4: The assumption $\epsilon \in (0, 1/2)$ is crucial for (39) to hold. Indeed, when $\epsilon > 1/2$, one can achieve $\log M_{\mathrm{f}}^*(n, \epsilon) = n \min_k C_k + o(n)$ by using a standard point-to-point fixed-length code of rate $R < \min_k C_k$ and codewords generated i.i.d. according to P_1^* (the CAID of W_1) with probability $1/2$ and another code of rate R and codewords generated i.i.d. according to P_2^* (the CAID of W_2) with probability $1/2$. This implies that the strong converse does not hold.

Remark 5: The assumption (38) is a multiuser analogue of the assumption in [9, Th. 2] under which the authors of [9] show that feedback does not improve the second-order terms of point-to-point DMCs. The assumption (38) is only needed in the proof of the converse part, and our achievability result continues to hold without it.

If (38) does not hold, our bounds imply only the weaker asymptotic expansion given in the next theorem.

Theorem 5: For every $\epsilon \in (0, 1/2)$, there exists a positive constant c such that

$$nC + \mathcal{O}(\sqrt{n}) = \log M_{\mathrm{f}}^*(n, \epsilon)$$
$$\geq nC - \sqrt{nV} Q^{-1}(2\epsilon) + \mathcal{O}(n^{1/3} \log n). \quad (40)$$

Proof: See Appendix II for the proof of the converse part and Appendix III for the proof of the achievability part. ∎
In the next section, we introduce a general class of CM-DMBCs for which the assumption (38) holds, and the second-order term can be characterized using (39).

If the CAID is nonunique and (38) does not hold, then the second-order term can be further improved compared to what is reported in (40) by using the feedback scheme in [9], provided that the conditional information variance takes different values when evaluated for the different CAIDs. The intuition behind the feedback scheme proposed in [9] is as follows: The encoder starts by using the CAID that minimizes the conditional information variance. During the transmission, the encoder can use the feedback to track the information density accumulated at the decoder and thereby compute the conditional error probability given the realized channel uses. If the conditional error probability becomes sufficiently large, it eventually becomes favorable to change the input distribution to the CAID maximizing the conditional information variance.

C. Antisymmetric CM-DMBCs

We shall apply our results to the class of antisymmetric CM-DMBCs defined as follows.

Definition 3: A CM-DMBC is antisymmetric if $|\mathcal{X}|$ is even, $\mathcal{Y}$ can be decomposed in disjoint sets $\overline{\mathcal{Y}}_1, \ldots, \overline{\mathcal{Y}}_r$, and the channel transition matrices can be decomposed as follows:

$$W_1 = \begin{bmatrix} p_{11}\overline{W}_{11} & \cdots & p_{1r}\overline{W}_{1r} \\ p_{21}\overline{W}_{21} & \cdots & p_{2r}\overline{W}_{2r} \end{bmatrix} \quad (41)$$

and

$$W_2 = \begin{bmatrix} p_{21}\overline{W}_{21} & \cdots & p_{2r}\overline{W}_{2r} \\ p_{11}\overline{W}_{11} & \cdots & p_{1r}\overline{W}_{1r} \end{bmatrix}. \quad (42)$$

Here, $\sum_{i=1}^r p_{ki} = 1$ for $k \in \{1, 2\}$ and $\{\overline{W}_{1i}\}$ and $\{\overline{W}_{2i}\}$ are weakly symmetric channel transition matrices[5] with dimensions $\frac{|\mathcal{X}|}{2} \times |\overline{\mathcal{Y}}_i|$. For notational convenience, we let $W^{(1)} \triangleq [p_{11}\overline{W}_{11}, \cdots, p_{1r}\overline{W}_{1r}]$ and $W^{(2)} \triangleq [p_{21}\overline{W}_{21}, \cdots, p_{2r}\overline{W}_{2r}]$.

Remark 6: The CM-DMBC composed of antisymmetric Z-channels is an example of an antisymmetric CM-DMBC, where $r = 2$, $\overline{W}_{11} = \overline{W}_{12} = \overline{W}_{21} = \overline{W}_{22} = 1$, $1 - p_{11} = p_{12} = q$, and $p_{21} = 1 - p_{22} = 0$.

We showed in [1] that feedback halves the dispersion for the CM-DMBC composed of antisymmetric Z-channels described in Remark 6. The following corollary of Theorem 4 and Theorem 5 extends this result to the more general class of antisymmetric CM-DMBCs.

Before stating the corollary, we shall briefly recall the intuition behind the half-dispersion result. The key point is that full feedback allows the encoder to compute the accumulated information densities at both decoders and make small adjustments in the input distribution in order to favor the decoder with the smallest accumulated information density. By doing so, both information densities are driven towards their arithmetic mean and the computation of the maximum coding rate roughly becomes equivalent to the computation of the ϵ-quantile of the arithmetic mean of the two information densities. Now, because the encoder only makes small adjustments in the input distribution, the arithmetic mean of the information densities is well-approximated by a Gaussian distribution with mean nC and variance $nV_1/2$. As a result, the dispersion is halved.

Corollary 6: For every antisymmetric CM-DMBC satisfying $\min_k C_k > C$ and for every $\epsilon \in (0, 1/2)$, we have that

$$\log M_{\mathrm{f}}^*(n, \epsilon) = nC - \sqrt{\frac{nV_1}{2}} Q^{-1}(2\epsilon) + \mathcal{O}(n^{1/3} \log n). \quad (43)$$

Proof: We need to show that, for every antisymmetric CM-DMBC, we have that $\eta_1 = \eta_2 = 1/2$, that $V_1 = V_2$, and that

$$\mathbb{V}\mathrm{ar}[\imath_{P^*, W_1}(x; Y_1) + \imath_{P^*, W_2}(x; Y_2) | X = x] = 2V_1 \quad (44)$$

for all $x \in \mathcal{X}$. We first show that P^* is the uniform distribution on $\mathcal{X}$. In the following, we let $\alpha_P = \sum_{x=1}^{|\mathcal{X}|/2} P(x)$ and $\overline{\alpha}_P = 1 - \alpha_P$ for every input distribution P on $\mathcal{X}$. Note that

$$H(Y_1|X) = \alpha_P H(\mathbf{r}_1) + \overline{\alpha}_P H(\mathbf{r}_2) \quad (45)$$
$$H(Y_2|X) = \overline{\alpha}_P H(\mathbf{r}_1) + \alpha_P H(\mathbf{r}_2) \quad (46)$$

where $\mathbf{r}_k$ denotes the first row of W_k. Define

$$\overline{P}_{Y_1}(y)$$
$$= \left[\frac{\alpha_P p_{11} + \overline{\alpha}_P p_{21}}{|\overline{\mathcal{Y}}_1|} \mathbf{1}_{|\overline{\mathcal{Y}}_1|}, \cdots, \frac{\alpha_P p_{1r} + \overline{\alpha}_P p_{2r}}{|\overline{\mathcal{Y}}_r|} \mathbf{1}_{|\overline{\mathcal{Y}}_r|} \right]. \quad (47)$$

It follows that [15, Th. 2.6.4]

$$0 \leq D(P_{Y_1} || \overline{P}_{Y_1})$$
$$= \sum_{j=1}^r (\alpha_P p_{1j} + \overline{\alpha}_P p_{2j}) \log \frac{|\overline{\mathcal{Y}}_j|}{\alpha_P p_{1j} + \overline{\alpha}_P p_{2j}} - H(Y_1) \quad (48)$$

---

[5]A channel transition matrix is weakly symmetric if the rows are permutations of each other and all column sums are equal. The CAID and CAOD of a weakly symmetric channel are uniform [15, pp. 189–190].

which implies

$$H(Y_1) \le \sum_{j=1}^{r} (\overline{\alpha}_P p_{1j} + \alpha_P p_{2j}) \log \frac{|\overline{\mathcal{Y}}_j|}{\overline{\alpha}_P p_{1j} + \alpha_P p_{2j}}. \quad (49)$$

Similarly,

$$H(Y_2) \le \sum_{j=1}^{r} (\overline{\alpha}_P p_{1j} + \alpha_P p_{2j}) \log \frac{|\overline{\mathcal{Y}}_j|}{\overline{\alpha}_P p_{1j} + \alpha_P p_{2j}}. \quad (50)$$

As a result of (45)–(46), (49)–(50), and because of symmetry, we have

$$\min_k I_k(P)$$
$$\le \frac{1}{2} \left(\sum_{j=1}^{r} (p_{1j} + p_{2j}) \log \frac{2|\overline{\mathcal{Y}}_j|}{p_{1j} + p_{2j}} - H(\mathbf{r}_1) - H(\mathbf{r}_2) \right). \quad (51)$$

This upper bound holds with equality when P is uniform. Hence, we have shown that P^* is the uniform distribution on $\mathcal{X}$, and consequently, we also have that $V_1 = V_2$, that $P_{Y_1}^* = P_{Y_2}^*$ and that (44) holds.

We next argue that $\eta_1 = \eta_2 = 1/2$. Define the permutation π of $\{1, \ldots, |\mathcal{X}|\}$ as follows: $\pi \triangleq [\lfloor \mathcal{X}/2 \rfloor + 1, \ldots, |\mathcal{X}|, 1, \ldots, |\mathcal{X}/2|]$. We first show that for every $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$

$$\langle \nabla I_1, \mathbf{v}_\pi \rangle = \sum_{x \in \mathcal{X}} v_\pi(x) D(W_1(\cdot|x) \| P_{Y_1}^*) \quad (52)$$
$$= \sum_{x \in \mathcal{X}} v_\pi(x) D(W_2(\cdot|\pi(x)) \| P_{Y_2}^*) \quad (53)$$
$$= \sum_{x \in \mathcal{X}} v(x) D(W_2(\cdot|x) \| P_{Y_2}^*) \quad (54)$$
$$= \langle \nabla I_2, \mathbf{v} \rangle. \quad (55)$$

Here, $\mathbf{v}_\pi$ denotes the vector $\mathbf{v}$ whose entries are permuted according to π. In particular, since $\pi(\pi(x)) = x$ for all $x \in \mathcal{X}$, (55) implies that

$$\langle \nabla I_1, \mathbf{v} + \mathbf{v}_\pi \rangle = \langle \nabla I_2, \mathbf{v} + \mathbf{v}_\pi \rangle. \quad (56)$$

By the uniqueness of P^* and by (41)–(42), there cannot exist a vector $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$ such that $\langle \nabla I_1, \mathbf{v} + \mathbf{v}_\pi \rangle$ and $\langle \nabla I_2, \mathbf{v} + \mathbf{v}_\pi \rangle$ are simultaneously positive or simultaneously negative, and hence $\langle \nabla I_1, \mathbf{v} + \mathbf{v}_\pi \rangle$ must be equal to zero for all $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$. Therefore, by using (55), we conclude that

$$0 = \langle \nabla I_1, \mathbf{v} + \mathbf{v}_\pi \rangle = \langle \nabla I_1, \mathbf{v} \rangle + \langle \nabla I_1, \mathbf{v}_\pi \rangle$$
$$= \langle \nabla I_1, \mathbf{v} \rangle + \langle \nabla I_2, \mathbf{v} \rangle. \quad (57)$$

Since η in (8) is unique, (57) implies that $\eta_1 = \eta_2 = 1/2$. The desired result is then established by invoking Theorem 4 and Theorem 5. ∎

D. A specific CM-DMBC

We shall next consider the antisymmetric CM-DMBC depicted in Fig. 2, which we shall use for evaluating numerically the accuracy of the asymptotic expansion provided in Corollary 6. For this specific class of antisymmetric CM-DMBC, we shall derive next an alternative achievability bound that yields tighter
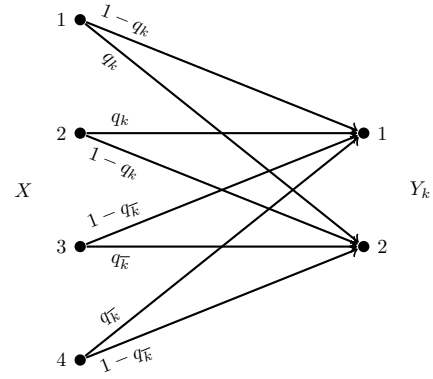

Fig. 2. An antisymmetric CM-DMBC composed of two binary symmetric channels. This figure illustrates the component channel W_k.

numerical results for small values of n than the generic achievability bound in Theorem 3. This bound, which is a random-coding union (RCU)-type of achievability bound, is based on a minimum distance mismatched decoder.

Theorem 7: For the CM-DMBC depicted in Fig. 2, there exists an (n, M, ϵ)-FLF code satisfying

$$\epsilon \le \sum_{t=0}^{n} P_{Z_1^{(n)}}(t) \min\left\{ 1, (M-1) \sum_{k=0}^{t} \binom{n}{k} 2^{-n} \right\}. \quad (58)$$

Here, $P_{Z_1^{(n)}}$ denotes the probability mass function of the RV $Z_1^{(n)}$ that is defined through the following recursion

$$Z_k^{(0)} \triangleq 0 \quad (59)$$
$$Z_k^{(i)} \triangleq Z_k^{(i-1)}$$
$$+ \begin{cases} E_1^{(i)} & \text{if } Z_1^{(i-1)} > Z_2^{(i-1)} \\ & \quad \text{or } \left(Z_1^{(i-1)} = Z_2^{(i-1)} \text{ and } \overline{Z}^{(i)} = 0 \right) \\ E_2^{(i)} & \text{if } Z_1^{(i-1)} < Z_2^{(i-1)} \\ & \quad \text{or } \left(Z_1^{(i-1)} = Z_2^{(i-1)} \text{ and } \overline{Z}^{(i)} = 1 \right) \end{cases} \quad (60)$$

for $i \in \{1, \ldots, n\}$, where $E_k^{(i)} \sim \text{Bern}(q_k)$ and $\overline{Z}^{(i)} \sim \text{Bern}(1/2)$ are independent RVs.

Proof: See Appendix IV. ∎

Let W_1 and W_2 be as depicted in Fig. 2 with $q_1 = 0.05$ and $q_2 = 0.10$. This channel is an antisymmetric CM-DMBC and, hence, it satisfies the conditions of Corollary 6. We plot the achievability bound given in Theorem 7, the converse bound given in Theorem 2, and the first two terms of the asymptotic expansion (normal approximation) in Corollary 6. We observe that the normal approximation is an accurate proxy for the maximum coding rate for the considered range of n. We would like to point out that the achievability scheme in Theorem 7 is not capacity-achieving since it is based on minimum-distance decoding. Specifically, the capacity of the channel depicted in Fig. 2 is given by $1 - \frac{1}{2}(h(q_1) + h(q_2)) \approx 0.622$, where $h(x) \triangleq -x \log x - (1-x) \log(1-x)$ is the binary entropy function. In contrast, the rate achievable by minimum-distance decoding is only $1 - h(\frac{1}{2}(q_1 + q_2)) \approx 0.616$. The plotted achievability bound is thus expected to be accurate only for small to moderate values of n. The generic achievability bound
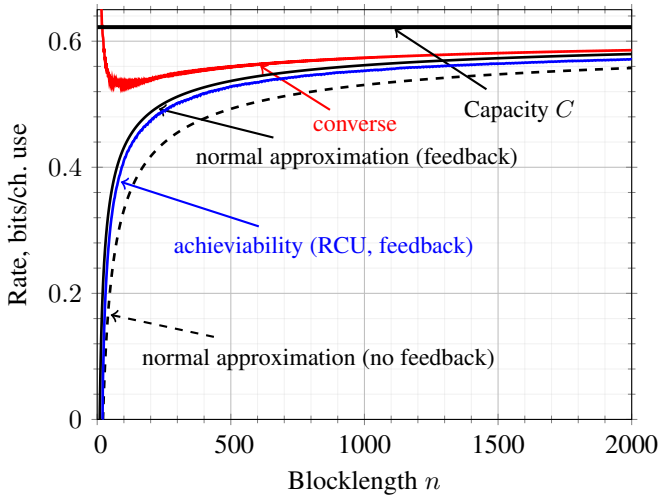
Fig. 3. Comparison between the RCU-type achievability bound in (58), the converse bound in (20), and the normal approximation in (43) for the CM-DMBC depicted in Fig. 2 with $q_1 = 0.05$, $q_2 = 0.10$, and $\epsilon = 10^{-3}$. The normal approximation (feedback) is the expression in (43) with the $\mathcal{O}(\cdot)$ term omitted. The normal approximation (no feedback) is the expression in [5, Eq. (12)] with the $o(\sqrt{n})$ term omitted.

in Theorem 3 is not plotted since it is not accurate numerically for moderate values of n. This is because of the use of constant composition codes and because the sequence B^L needs to be communicated explicitly.

IV. VARIABLE-LENGTH FEEDBACK CODES

We first present a Fano-type converse bound and a nonasymptotic achievability bound for VLF codes. We then show that these bounds imply that the second-order term in the asymptotic expansion of $\log M_{\text{vf}}^*(\mathrm{n}, \epsilon)$ is zero.

A. Nonasymptotic bounds

Our converse result for VLF codes is based on Lemma 1 and on [10, Lem 1 and Lem 2].

Theorem 8: Every $(\mathrm{n}, M, \epsilon)$-VLF code with $0 < \epsilon \leq 1 - 1/M$ satisfies

$$\log M \leq \frac{\mathrm{n}C + h(\epsilon)}{1 - \epsilon} \tag{61}$$

where $h(x) \triangleq -x \log x - (1-x) \log(1-x)$ is the binary entropy function.

Proof: See Appendix V. ∎

Our nonasymptotic achievability bound for VLF codes leverages ideas similar to in Theorem 3 and in [8, Th. 3]. The achievability scheme consists of two phases: a fixed-length transmission phase via an FLF code and a variable-length transmission phase via a variable-length stop-feedback code.

Theorem 9: Let $P_1, \ldots, P_S$ be probability distributions on $\mathcal{X}$, let $\gamma \geq 0$ and $0 \leq q \leq 1$ be arbitrary scalars, and $\tau_{\max}$ be an

arbitrary integer larger than Lm. The stopping times τ_k and $\bar{\tau}_k$ for $k \in \{1, 2\}$ are defined as follows:

$$\tau_k \triangleq \min\{\inf\{n \geq Lm : S_{k,n}(X^n, Y_k^n; B^L) \geq \gamma\}, \tau_{\max}\} \tag{62}$$

$$\bar{\tau}_k \triangleq \min\{\inf\{n \geq Lm : S_{k,n}(\bar{X}^n, Y_k^n; B^L) \geq \gamma\}, \tau_{\max}\}. \tag{63}$$

Here, we have defined for $n \geq Lm$

$$S_{k,n}(x^n, y_k^n; b^L) \triangleq \sum_{\ell=1}^{L} \sum_{i=1}^{m} \imath_{P_{b_\ell}, W_k}(x_{\nu(\ell,i)}; y_{k,\nu(\ell,i)})$$
$$+ \sum_{i=Lm+1}^{n} \imath_{P^*, W_k}(x_i; y_{k,i}) \tag{64}$$

where $\nu(\ell, i) \triangleq (\ell - 1)m + i$, and the joint probability distribution of $(X^n, \bar{X}^n, Y_1^n, Y_2^n, B^L)$ is

$$\mathbb{P}_{X^n, \bar{X}^n, Y_1^n, Y_2^n, B^L}(x^n, \bar{x}^n, y_1^n, y_2^n, b^L)$$
$$= \left(\prod_k W_k^n(y_k^n | x^n) \right)$$
$$\times \left(\prod_{\ell=1}^{L} \left(\prod_{i=1}^{m} P_{b_\ell}(x_{\nu(\ell,i)}) P_{b_\ell}(\bar{x}_{\nu(\ell,i)}) \right) \right.$$
$$\left. \times \mathbb{1}\left\{ b_\ell = h_\ell(x^{\ell-1}, y_1^{\ell-1}, y_2^{\ell-1}) \right\} \right)$$
$$\times (P^*)^{n-Lm}(x_{Lm+1}^n)(P^*)^{n-Lm}(\bar{x}_{Lm+1}^n). \tag{65}$$

Then, for every positive integer M and n_B, there exists an $(\mathrm{n}, M, \epsilon)$-VLF code with

$$\mathrm{n} \leq (1-q)\left(\mathbb{E}\left[\max_k \tau_k \right] + n_B \right) \tag{66}$$

$$\epsilon \leq q + (1-q)\epsilon^*(n_B, S^L)$$
$$+ \max_k \left\{ (1-q)(M-1)\mathbb{P}[\tau_k \geq \bar{\tau}_k] \right\}. \tag{67}$$

Remark 7: Following steps similar to those in [8, Eqs. (111)-(118)], one can further upper-bound ϵ in (67) as follows:

$$\epsilon \leq q + (1-q)\left(\epsilon^*(n_B, S^L) + \exp(-\gamma) + \max_k \mathbb{P}[\tau_k = \tau_{\max}] \right). \tag{68}$$

Here, we have treated the event $\{\tau_k = \tau_{\max}\}$ as an error.

Remark 8: Compared to the nonasymptotic achievability bound for VLF codes for DMCs reported in [8], in our bounds, the stopping times τ_k and $\bar{\tau}_k$ are lower- and upper-bounded by Lm and $\tau_{\max}$, respectively. These bounds simplify the asymptotic analysis of the achievability bound.

Proof: See Appendix VI. ∎

B. Asymptotic analysis

The following result reveals that one can achieve zero dispersion by using VLF codes.

Theorem 10: For every $\epsilon \in (0, 1)$, we have

$$\log M_{\text{vf}}^*(\mathrm{n}, \epsilon) \geq \frac{\mathrm{n}C}{1 - \epsilon} + \mathcal{O}(\mathrm{n}^{1/3} \log \mathrm{n}). \tag{69}$$

Proof: See Appendix VII. ∎

The intuition behind this result is as follows: The encoder uses a feedback scheme similar to the one used for our fixed-blocklength result. This implies that the information densities at both decoders are driven towards their arithmetic mean. Each decoder sends a stop signal as soon as the information density at that decoder exceeds a threshold. Since the information densities at both decoders are well-approximated by their arithmetic mean, both decoders send a stop signal at approximately the same time, which yields zero dispersion

We would like to emphasize that we have previously shown that the dispersion is positive when only stop feedback is available [14]. Our result in Theorem 10 implies that there is a fundamental difference between full feedback and stop feedback in terms of speed of convergence to capacity. To our knowledge, this is the first result in the literature on fixed-error asymptotics that explicitly shows that the speed of convergence to capacity is different with stop feedback and full feedback. For comparison, for the single-decoder case, both stop feedback and full feedback yield the zero dispersion.

V. CONCLUSION

In this paper, we investigated the maximal coding rate for the CM-DMBC with feedback. When fixed-length codes are employed, we demonstrated that the second-order term is reduced by a factor of more than $1/\sqrt{2}$ under mild technical conditions. Under an additional symmetry condition, we also proved that the asymptotic expansions of the our converse and achievability bounds match up to the second order. This improved second-order term is in contrast to the point-to-point setup, where it is shown that the second-order term is unaltered under similar symmetry conditions [9]. We also found that the second-order term vanishes under certain mild technical conditions if we allow the use of variable-length codes. This result together with the one previously reported in [14] implies that zero-dispersion is achievable when feedback is available and variable-length codes are used but that stop-feedback is not sufficient to attain it.

APPENDIX I
PROOF OF THEOREM 3

We consider feedback schemes in which the encoders can be represented as follows: There exist M vectors in $\mathcal{X}^{SLm}$, denoted by $\{\varkappa^{(1)}, \ldots, \varkappa^{(M)}\}$, such that

$$f_i(j, Y_1^{i-1}, Y_2^{i-1}) = \varkappa^{(j)}_{\ell(i), i-m(\ell(i)-1)}(B_{\ell(i)}) \qquad (70)$$

for $i \in \{1, \ldots, Lm\}$. Here, $\ell(i) \triangleq \lceil i/m \rceil$ and

$$B_\ell \triangleq h_\ell(X^{(\ell-1)m}, Y_1^{(\ell-1)m}, Y_2^{(\ell-1)m}) \qquad (71)$$

where $X^{(\ell-1)m}$ denotes the sequence of channel inputs transmitted up to time $(\ell-1)m$.

In the remainder of the proof, we shall use $\mathbf{Y}_{k,\ell}$ to denote $(Y_{k,(\ell-1)m+1}, \ldots, Y_{k,\ell m})$ and we let $\mathbf{Y}_k$ denote

$(\mathbf{Y}_{k,1}, \ldots, \mathbf{Y}_{k,L})$. The conditional probability distribution of $(\mathbf{Y}_1, \mathbf{Y}_2, B^L)$ given that $\mathbb{X} = \varkappa \in \mathcal{X}^{SLm}$ is then given by

$$\mathbb{P}_{\mathbf{Y}_1, \mathbf{Y}_2, B^L | \mathbb{X}}(\mathbf{y}_1, \mathbf{y}_2, b^L | \varkappa)$$
$$= \prod_{\ell=1}^{L} \left(\prod_k W_k^m(\mathbf{y}_{k,\ell} | \varkappa_\ell(b_\ell)) \right)$$
$$\times \mathbb{1}\left\{ h_\ell\left(\varkappa^{\ell-1}(b^{\ell-1}), \mathbf{y}_1^{\ell-1}, \mathbf{y}_2^{\ell-1} \right) = b_\ell \right\}. \qquad (72)$$

Next, we shall apply the achievability bound for the compound channel with channel state information at the receiver reported in [5, Th. 3]. In particular, we let $\mathcal{X}^{SLm}$ be the input alphabet, $\mathbb{B} \triangleq \mathcal{Y}^{Lm} \times \mathcal{S}^L$ be the output alphabets, and consider the compound channel with transition probability

$$\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}}(\mathbf{y}_k, b^L | \varkappa)$$
$$\triangleq \sum_{\mathbf{y}_{\bar{k}} \in \mathcal{Y}^{Lm}} \prod_{\ell=1}^{L} \left(\prod_k W_k^m(\mathbf{y}_{k,\ell} | \varkappa_\ell(b_\ell)) \right)$$
$$\times \mathbb{1}\left\{ h_\ell\left(\varkappa^{\ell-1}(b^{\ell-1}), \mathbf{y}_1^{\ell-1}, \mathbf{y}_2^{\ell-1} \right) = b_\ell \right\} \qquad (73)$$
$$= \left(\prod_{\ell=1}^{L} W_k^m(\mathbf{y}_{k,\ell} | \varkappa_\ell(b_\ell)) \right) P_{B^L | \mathbb{X}}(b^L | \varkappa). \qquad (74)$$

Moreover, we set $\mathbb{F}$ as in (33) and take as auxiliary channel

$$\mathbb{Q}_{\mathbf{Y}_k, B^L}(\mathbf{y}_k, b^L) \triangleq \frac{1}{S^L} \prod_{\ell=1}^{L} (P_{b_\ell} W_k)^m(\mathbf{y}_{k,\ell}). \qquad (75)$$

It will turn out convenient to define also the following probability measure:

$$\mathbb{Q}_{\mathbb{X}}(\varkappa) \triangleq \prod_{\ell=1}^{L} \prod_{s=1}^{S} (P_s)^m(\varkappa_\ell(s)). \qquad (76)$$

Note that

$$\sum_{\varkappa \in \mathcal{X}^{SLm}} \mathbb{Q}_{\mathbb{X}}(\varkappa) \mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}}(\mathbf{y}_k, b^L | \varkappa)$$
$$\leq \sum_{\varkappa \in \mathcal{X}^{SLm}} \mathbb{Q}_{\mathbb{X}}(\varkappa) \prod_{\ell=1}^{L} W_k^m(\mathbf{y}_{k,\ell} | \varkappa_\ell(b_\ell)) \qquad (77)$$
$$= \prod_{\ell=1}^{L} (P_{b_\ell} W_k)^m(\mathbf{y}_{k,\ell}) \qquad (78)$$
$$= S^L \mathbb{Q}_{\mathbf{Y}_k, B^L}(\mathbf{y}_k, b^L). \qquad (79)$$

We observe that this compound channnel is equivalent to a CM-DMBC with encoding functions satisfying (70), provided that the decoders have knowledge about B^L. We can readily provide both decoders with knowledge of B^L using a finite blocklength code with n_B channel uses and an error probability $\epsilon^*(n_B, S^L)$ defined in (31) at the end of the transmission. Hence, we apply [5, Th. 3] and conclude that

$$M_{\mathrm{f}}^*(Lm + n_B, \epsilon + \epsilon^*(n_B, S^L)) \geq \frac{\kappa_\tau}{\beta_{1-\epsilon+\tau}} \qquad (80)$$

where

$$\beta_{1-\epsilon+\tau} = \sup_{\varkappa \in \mathbb{F}, k \in \{1,2\}} \beta_{1-\epsilon+\tau}(\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}=\varkappa}, \mathbb{Q}_{\mathbf{Y}_k, B^L}) \qquad (81)$$
$$\kappa_\tau = \inf_E \max_{k \in \{1,2\}} \mathbb{Q}_{\mathbf{Y}_k, B^L}[E] \qquad (82)$$

and the infimum in the definition of κ_τ is with respect to all sets $E \subseteq \mathbb{B}$ satisfying

$$\forall \varkappa \in \mathbb{F}, \exists k : \mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X} = \varkappa}[E] \geq \tau. \tag{83}$$

We conclude the proof by providing an upper bound on $\beta_{1-\epsilon+\tau}$ and a lower bound on κ_τ.

Upper bound on $\beta_{1-\epsilon+\tau}$: By symmetry, $\beta_{1-\epsilon+\tau}(\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X} = \varkappa}, \mathbb{Q}_{\mathbf{Y}_k, B^L})$ takes the same value for all $\varkappa \in \mathbb{F}$. To upper-bound $\beta_{1-\epsilon+\tau}$, we apply [6, Eq. (103)] to obtain

$$\log \beta_{1-\epsilon+\tau} \leq -\sup\Bigg\{ \gamma_0 :$$
$$\max_k \mathbb{P}_\varkappa \Bigg[\log \frac{\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}}(\mathbf{Y}_k, B^L | \varkappa)}{\mathbb{Q}_{\mathbf{Y}_k, B^L}(\mathbf{Y}_k, B^L)} < \gamma_0 \Bigg] \leq \epsilon - \tau \Bigg\}. \tag{84}$$

Here, $\varkappa$ is an arbitrary element in $\mathbb{F}$. We further upper-bound (84) by lower-bounding $\log \frac{\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}}(\mathbf{Y}_k, B^L | \varkappa)}{\mathbb{Q}_{\mathbf{Y}_k, B^L}(\mathbf{Y}_k, B^L)}$ as follows:

$$\mathbb{P}_\varkappa \Bigg[\log \frac{\mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X}}(\mathbf{Y}_k, B^L | \varkappa)}{\mathbb{Q}_{\mathbf{Y}_k, B^L}(\mathbf{Y}_k, B^L)} \geq \gamma_0 \Bigg]$$
$$= \mathbb{P}_\varkappa \big[\imath_k(\varkappa(B^L); \mathbf{Y}_k) + \log\big(S^L P_{B^L | \mathbb{X}}(B^L | \varkappa)\big) \geq \gamma_0 \big] \tag{85}$$
$$\geq \mathbb{P}_\varkappa \big[\imath_k(\varkappa(B^L); \mathbf{Y}_k) + \log\big(S^L P_{B^L | \mathbb{X}}(B^L | \varkappa)\big) \geq \gamma_0$$
$$, S^L P_{B^L | \mathbb{X}}(B^L | \varkappa) \geq e^{-\zeta} \big] \tag{86}$$
$$\geq \mathbb{P}_\varkappa \big[\imath_k(\varkappa(B^L); \mathbf{Y}_k) - \zeta \geq \gamma_0$$
$$, S^L P_{B^L | \mathbb{X}}(B^L | \varkappa) \geq e^{-\zeta} \big] \tag{87}$$
$$\geq \mathbb{P}_\varkappa \big[\imath_k(\varkappa(B^L); \mathbf{Y}_k) - \zeta \geq \gamma_0 \big]$$
$$- \mathbb{P}_\varkappa \big[S^L P_{B^L | \mathbb{X}}(B^L | \varkappa) < e^{-\zeta} \big] \tag{88}$$
$$\geq \mathbb{P}_\varkappa \big[\imath_k(\varkappa(B^L); \mathbf{Y}_k) - \zeta \geq \gamma_0 \big] - e^{-\zeta}. \tag{89}$$

Here, in (88), we have used that $\mathbb{P}[\mathcal{A}, \mathcal{B}] = \mathbb{P}[\mathcal{A}] - \mathbb{P}[\mathcal{A} \setminus \mathcal{B}] \geq \mathbb{P}[\mathcal{A}] - \mathbb{P}[\mathcal{B}^\complement]$ for any events $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{B}^\complement$ being the complement of the set $\mathcal{B}$, and (89) follows because

$$\mathbb{P}_\varkappa \big[P_{B^L | \mathbb{X}}(B^L | \varkappa) < e^{-\zeta} / S^L \big]$$
$$= \sum_{b^L \in \mathcal{S}^L} P_{B^L | \mathbb{X}}(b^L | \varkappa) \mathbb{1}\Bigg\{ P_{B^L | \mathbb{X}}(b^L | \varkappa) < \frac{e^{-\zeta}}{S^L} \Bigg\} \tag{90}$$
$$\leq \sum_{b^L \in \mathcal{S}^L} \frac{e^{-\zeta}}{S^L} \mathbb{1}\Bigg\{ P_{B^L | \mathbb{X}}(b^L | \varkappa) < \frac{e^{-\zeta}}{S^L} \Bigg\} \tag{91}$$
$$\leq e^{-\zeta}. \tag{92}$$

Lower bound on κ_τ: We follow steps similar to [5, Eq. (28)–(29)]. Specifically, for any set E satisfying (83), we have

$$\sum_k \mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X} = \varkappa}[E] \geq \tau \mathbb{1}\{\varkappa \in \mathbb{F}\}. \tag{93}$$

By averaging (93) over $\mathbb{X} \sim \mathbb{Q}_\mathbb{X}$ in (76) and by using (79), we obtain

$$\tau \mathbb{Q}_\mathbb{X}[\mathbb{F}] \leq \sum_k \sum_{\varkappa \in \mathcal{X}^{SLm}} \mathbb{Q}_\mathbb{X}(\varkappa) \mathbb{P}_{\mathbf{Y}_k, B^L | \mathbb{X} = \varkappa}[E] \tag{94}$$
$$\leq S^L \sum_k \mathbb{Q}_{\mathbf{Y}_k, B^L}(E). \tag{95}$$

This implies that

$$\max_k \mathbb{Q}_{\mathbf{Y}_k, B^L}(E) \geq \frac{\tau \mathbb{Q}_\mathbb{X}(\mathbb{F})}{2 S^L}. \tag{96}$$

Since E is arbitrary, we have shown that

$$\kappa_\tau \geq \frac{\tau \mathbb{Q}_\mathbb{X}(\mathbb{F})}{2 S^L}. \tag{97}$$

Next, by using that [16, Lem 2.6]

$$\mathbb{Q}_\mathbb{X}(\mathbb{F}) \geq (1 + m)^{-SL|\mathcal{X}|} \tag{98}$$

we conclude that

$$\log \kappa_\tau \geq \log \frac{\tau}{2} - SL|\mathcal{X}| \log(1 + m) - L \log S. \tag{99}$$

We establish the desired result (32) by substituting (89) and (99) in (80).

APPENDIX II
PROOF OF THEOREM 4 (CONVERSE) AND OF THEOREM 5 (CONVERSE)

By Theorem 2, every (n, M, ϵ)-FLF satisfies

$$\epsilon \geq \frac{1}{2} \mathbb{P}_{J, Y_1^n, Y_2^n}[A(J, Y_1^n, Y_2^n) \leq \log M - \lambda] - e^{-\lambda} \tag{100}$$

where

$$A(j, y_1^n, y_2^n) \triangleq \sum_k \eta_k \sum_{i=1}^n \imath_{P^*, W_k}\Big(y_{k,i}; f_i(j, y_1^{i-1}, y_2^{i-1})\Big). \tag{101}$$

Observe also that

$$\mathbb{E}_{\mathbb{P}_{Y_{1,i}, Y_{2,i} | Y_1^{i-1}, Y_2^{i-1}, J=j}}\Bigg[$$
$$\sum_k \eta_k \log \frac{W_k(Y_{k,i} | f_i(j, Y_1^{i-1}, Y_2^{i-1}))}{P_{Y_k}^*(Y_{k,i})} \Bigg]$$
$$= \sum_k \eta_k D(W_k(\cdot | f_i(j, Y_1^{i-1}, Y_2^{i-1})) || P_{Y_k}^*) \tag{102}$$
$$\leq \max_{x \in \mathcal{X}} \Bigg\{ \sum_k \eta_k D(W_k(\cdot | x) || P_{Y_k}^*) \Bigg\} \tag{103}$$
$$= C + \max_{x \in \mathcal{X}} \Bigg\{ \sum_k \eta_k \langle \nabla I_k, \mathbb{1}_x - P^* \rangle \Bigg\} \tag{104}$$
$$= C. \tag{105}$$

Here, in (104), we let $\mathbb{1}_x$ be the $|\mathcal{X}|$-dimensional vector that has a one in the xth entry and zeroes in all other entries and we have used (7). Furthermore, (105) follows from an application of Lemma 1.

Proof of Theorem 4 (converse): Next, we apply a Berry-Esseen-type central-limit theorem for martingales due to Bolthausen to estimate the probability $\mathbb{P}_{Y_1^n, Y_2^n | J=j}[A(j, Y_1^n, Y_2^n) \leq \log M - \eta]$ [17, p. 2] (see also [18]) in a manner similar to [9, Th. 2].

Theorem 11: Let $0 < \gamma < \infty$ and let $\{X_i\}_{i=1}^n, n \geq 2$, be a martingale difference sequence[6] with respect to the filtration

---

[6]A sequence $\{X_i\}_{i=1}^\infty$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_i\}_{i=0}^\infty$ if it satisfies the following two conditions for all $t \in \mathbb{N}$: X_i is $\mathcal{F}_i$-measurable and $\mathbb{E}[X_i | \mathcal{F}_{i-1}] = 0$ almost surely [17].

$\{\mathcal{F}_i\}_{i=0}^n$ where $|X_i| \leq \gamma$ almost surely $i = 1, \ldots, n$. Let $\sigma_i^2 \triangleq \mathbb{E}[X_i^2 | \mathcal{F}_{i-1}]$ and $\bar{\sigma}^2 \triangleq \mathbb{E}[X_i^2]$. Furthermore, assume that

$$s_n^2 \triangleq \sum_{i=1}^n \bar{\sigma}_i^2 = \sum_{i=1}^n \sigma_i^2 \qquad (106)$$

almost surely. Then, there exists a constant $0 < L(\gamma) < \infty$ that depends only on γ, so that

$$\sup_t \left| \mathbb{P}\left[\frac{1}{s_n} \sum_{i=1}^n X_i \leq t \right] - Q(-t) \right| \leq \frac{nL(\gamma) \log n}{s_n^3}. \qquad (107)$$

To apply this result, we note that, for every $j \in \mathcal{M}$,

$$\left\{ \frac{1}{\sqrt{V}} \left(\left(\sum_k \eta_k \log \frac{W_k(Y_{k,i}|f_i(j, Y_1^{i-1}, Y_2^{i-1}))}{P_{Y_k}^*(Y_{k,i})} \right) \right. \right.$$
$$\left. \left. - \mathbb{E}_{\mathbb{P}_{Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}, J=j}} \left[\right. \right. \right.$$
$$\left. \left. \left. \sum_k \eta_k \log \frac{W_k(Y_{k,i}|f_i(j, Y_1^{i-1}, Y_2^{i-1}))}{P_{Y_k}^*(Y_{k,i})} \right] \right) \right\}_{i=1}^n \qquad (108)$$

is a martingale difference sequence with respect to the filtration $\{\sigma(Y_1^i, Y_2^i)\}_{i=1}^n$ which is uniformly bounded because $W_k(\cdot|\cdot)$ and $P_{Y_k^*}(\cdot)$ only take a finite number of different values. Furthermore, the conditional variance σ_i^2 (defined as in Theorem 11) of the RVs in (108) is equal to V because of (38). It then follows from Theorem 11 and from (105) that

$$\mathbb{P}_{Y_1^n, Y_2^n | J=j}[A(j, Y_1^n, Y_2^n) \leq \log M - \lambda]$$
$$\geq Q\left(\frac{nC - \log M + \lambda}{\sqrt{nV}} \right) - \frac{\mathbb{c} \log n}{\sqrt{n}}. \qquad (109)$$

Here, the constant in the last term depends only on the component channels $\{W_k\}$. Substituting (109) into (100), we conclude that

$$\epsilon \geq \frac{1}{2} Q\left(\frac{nC - \log M + \lambda}{\sqrt{nV}} \right) - \frac{\mathbb{c} \log n}{\sqrt{n}} - e^{-\lambda}. \qquad (110)$$

By solving for $\log M$, by setting $\lambda = \log n$, and by performing a Taylor expansion of $Q^{-1}(\cdot)$ around 2ϵ, we obtain

$$\log M$$
$$\leq nC - \sqrt{nV} Q^{-1}\left(2\left(\epsilon + \frac{\mathbb{c} \log n}{\sqrt{n}} + e^{-\lambda} \right) \right) + \lambda \qquad (111)$$
$$= nC - \sqrt{nV} Q^{-1}(2\epsilon) + \mathcal{O}(\log n) \qquad (112)$$

which is the desired result.

Proof of Theorem 5 (converse): When (38) is not satisfied, we resort to Chebyshev inequality [19, Eq. (3.1.1)] instead of the central limit theorem. Specifically, since $\mathcal{X}$ and $\mathcal{Y}$ are finite-cardinality alphabets, there exists a constant $K < \infty$ such that

$$\mathbb{Var}\left[\sum_k \eta_k \log \frac{W_k(Y_{k,i}|f_i(J, Y_1^{i-1}, Y_2^{i-1}))}{P_{Y_k}^*(Y_{k,i})} \middle| J, Y_1^{i-1}, Y_2^{i-1} \right]$$
$$\leq K \qquad (113)$$

for $i \in \{1, \ldots, t\}$. It follows from (100), (105), (113), and Chebyshev inequality that

$$2(\epsilon + e^{-\lambda}) \geq \mathbb{P}_{J, Y_1^n, Y_2^n}[A(J, Y_1^n, Y_2^n) \leq \log M - \lambda] \qquad (114)$$
$$\geq 1 - \frac{Kn}{(\log M - nC - \lambda)^2}. \qquad (115)$$

By solving for $\log M$ and by setting $\lambda = \log n$, we obtain that for all $\epsilon \in (0, 1/2)$ and for all sufficiently large n

$$\log M \leq nC + \sqrt{\frac{Kn}{1 - 2(\epsilon + 1/n)}} + \log n \qquad (116)$$
$$= nC + \mathcal{O}(\sqrt{n}). \qquad (117)$$

APPENDIX III
PROOF OF THEOREM 4 (ACHIEVABILITY) AND OF THEOREM 5 (ACHIEVABILITY)

To establish that the right-hand side of (40) in Theorem 5 is achievable (which implies achievability in Theorem 4 as well), we start by setting in Theorem 3 the parameter S to 5 (recall that S controls the number of subcodewords available in each of the L blocks). Let now κ be the smallest integer larger than $C^{-1} \log 5$. Furthermore, set the number of blocks and the number of channel uses per block in Theorem 3 to

$$L = L_n \triangleq \lfloor n^{1/3} \rfloor \quad \text{and} \quad m = m_n \triangleq \lfloor n/L_n \rfloor - \kappa. \qquad (118)$$

We use the remaining κL_n channel uses to communicate the sequence $B^{L_n} \in \{1, \ldots, 5\}^{L_n}$ to the decoders. Since $\kappa > C^{-1} \log 5$, the rate $\log 5/\kappa$ of the code is smaller than C. Hence, its error probability can be made to decay exponentially in L_n

$$\epsilon^*(\kappa L_n, 5^{L_n}) \leq \exp(-\omega L_n) \qquad (119)$$

for some positive constant ω [16, Th. 10.10]. In Theorem 3, we also set $\tau = 1/\sqrt{n}$, $\zeta = n^{1/3}$.

Next, we specify the types $\{P_b\}_{b=1}^5$ in Theorem 3. Let $\rho > 0$ (a constant that we shall specify later) and let $\overline{\mathbf{v}}_0$ be an arbitrary nonzero vector in $\mathbb{R}_0^{|\mathcal{X}|}$ satisfying

$$\langle \nabla I_1, \overline{\mathbf{v}}_0 \rangle = \rho. \qquad (120)$$

We next define the following probability distributions

$$\overline{P}_b^{(n)} \triangleq P^* - (-1)^b \overline{\mathbf{v}}_0 n^{-1/3}, \quad b \in \{1, 2\} \qquad (121)$$
$$\overline{P}_3^{(n)} \triangleq P^*, \quad \overline{P}_4^{(n)} \triangleq P_1^*, \quad \overline{P}_5^{(n)} \triangleq P_2^*. \qquad (122)$$

Since $P^*(x) > 0$ for all $x \in \mathcal{X}$, the entries of $\overline{P}_b^{(n)}$ are nonnegative and sum to one for all sufficiently large n. Therefore, $\overline{P}_b^{(n)}$, $b \in \{1, 2\}$, are legitimate probability distributions that lie in the neighborhood of P^*.

Now, let $P_b^{(n)}$ be the type $Q^{(n)} \in \mathcal{P}_{m_n}(\mathcal{X})$ that minimizes $\|\overline{P}_b^{(n)} - Q^{(n)}\|$ (recall that $\mathcal{P}_{m_n}(\mathcal{X})$ denotes the set of types of m_n-dimensional sequences and $\|\cdot\|$ denotes the Euclidean distance). We then choose the types in Theorem 3 as $P_b = P_b^{(n)}$.

Observe that $\|P_b^{(n)} - \overline{P}_b^{(n)}\|_1 = \mathcal{O}(1/n)$ and that $I_k(\cdot)$ is differentiable. Hence, we have

$$I_k(P_b^{(n)}) = I_k(\overline{P}_b^{(n)}) + \mathcal{O}(1/n) \qquad (123)$$

for every $b \in \{1, \ldots, 5\}$.

The mappings $\{h_\ell\}_{\ell=1}^{L_n}$, which use the feedback to determine which subcodeword to transmit in each of the L_n blocks, are chosen as follows: In the first $L_n - 1$ blocks, we let the transmitter choose between subcodewords of type $P_1^{(n)}$ and of type $P_2^{(n)}$ depending on which decoder has the largest accumulated information density. This balances the information densities at the two

decoders so that the difference is tightly concentrated around zero (as we shall see later). In the last block, the transmitter chooses a subcodeword of type $P_3^{(n)}$ (which approximates the CAID P^*) if the arithmetic average of the information densities is above a suitably chosen threshold γ_1 (specified in (147)). Otherwise, it chooses uniformly at random between the subcodeword of type $P_4^{(n)}$ (which approximates the CAID of W_1) and the one of type $P_5^{(n)}$ (which approximates the CAID of W_2).

Mathematically, we can express the chosen mappings $\{h_\ell\}_{\ell=1}^{L_n}$ as follows: $h_1 = 2$ and for $\ell = 2, \ldots, L_n - 1$,

$$
\begin{aligned}
&h_\ell(\varkappa^{\ell-1}(B^{\ell-1}), \mathbb{Y}_1^{\ell-1}(B^{\ell-1}), \mathbb{Y}_2^{\ell-1}(B^{\ell-1})) \\
&= 1 + \mathbb{1}\Big\{ \imath_1(\varkappa^{\ell-1}(B^{\ell-1}); \mathbb{Y}_1^{\ell-1}(B^{\ell-1})) \\
&\qquad\qquad \geq \imath_2(\varkappa^{\ell-1}(B^{\ell-1}); \mathbb{Y}_2^{\ell-1}(B^{\ell-1})) \Big\}. \quad (124)
\end{aligned}
$$

For $\ell = L_n$, we let T be a uniformly distributed RV on $\{1, 2\}$ and set

$$
\begin{aligned}
&h_\ell(\varkappa^{\ell-1}(B^{\ell-1}), \mathbb{Y}_1^{\ell-1}(B^{\ell-1}), \mathbb{Y}_2^{\ell-1}(B^{\ell-1})) \\
&\triangleq \begin{cases}
3, \text{if } \sum_k \eta_k \imath_k(\varkappa^{\ell-1}(B^{\ell-1}); \mathbb{Y}_k^{\ell-1}(B^{\ell-1})) \geq \gamma_1 \\
4, \text{if } \sum_k \eta_k \imath_k(\varkappa^{\ell-1}(B^{\ell-1}); \mathbb{Y}_k^{\ell-1}(B^{\ell-1})) < \gamma_1 \\
\quad \text{and } T = 1 \\
5, \text{if } \sum_k \eta_k \imath_k(\varkappa^{\ell-1}(B^{\ell-1}); \mathbb{Y}_k^{\ell-1}(B^{\ell-1})) < \gamma_1 \\
\quad \text{and } T = 2.
\end{cases} \quad (125)
\end{aligned}
$$

Roughly speaking, the threshold γ_1 is chosen so that $P_3^{(n)}$, $P_4^{(n)}$, and $P_5^{(n)}$ are used with probability $(1-2\epsilon)$, ϵ, and ϵ, respectively.

Using Theorem 3 with the parameters listed above, we obtain

$$
\begin{aligned}
&\log M_{\mathrm{f}}^*(n, \epsilon) \\
&\geq \log M_{\mathrm{f}}^*(L_n(m_n + \kappa), \epsilon) \quad (126) \\
&\geq \sup\Big\{ \gamma : \max_k \mathbb{P}\big[\imath_k(\varkappa(B^{L_n}); \mathbb{Y}_k(B^{L_n})) < \gamma\big] \\
&\qquad\qquad \leq \epsilon - \epsilon^*(\kappa L_n, S^{L_n}) - 1/\sqrt{n} - e^{-n^{1/3}} \Big\} \\
&\quad + \log \frac{1}{2\sqrt{n}} - SL_n |\mathcal{X}| \log(1 + m_n) \\
&\quad - L_n \log S - n^{1/3} \quad (127) \\
&\geq \sup\Big\{ \gamma : \max_k \mathbb{P}\big[\imath_k(\varkappa(B^{L_n}); \mathbb{Y}_k(B^{L_n})) < \gamma\big] \\
&\qquad\qquad \leq \epsilon - c_1 n^{-1/2} \log n \Big\} \\
&\quad + \mathcal{O}(n^{1/3} \log n) \quad (128)
\end{aligned}
$$

for some $c_1 > 0$. In (128), we used that

$$
\begin{aligned}
&\epsilon^*(\kappa L_n, S^{L_n}) + 1/\sqrt{n} + e^{-n^{1/3}} \\
&\leq e^{-\omega L_n} + 1/\sqrt{n} + e^{-n^{1/3}} \quad (129) \\
&\leq c_1 n^{-1/2} \log n \quad (130)
\end{aligned}
$$

for sufficiently large n. To conclude the proof, we show that

$$
\begin{aligned}
\gamma &\triangleq L_n m_n C - \sqrt{(L_n - 1)m_n V} \\
&\quad \times Q^{-1}\big(2\epsilon - 2\sqrt{2}(c_1 + c_2) n^{-1/2} \log n\big) \quad (131)
\end{aligned}
$$

(the constant $c_2 > 0$ will be defined shortly) satisfies

$$
\mathbb{P}\big[\imath_k(\varkappa(B^{L_n}); \mathbb{Y}_k(B^{L_n})) < \gamma\big] \leq \epsilon - c_1 n^{-1/2} \log n \quad (132)
$$

for all sufficiently large n. The desired result (40) then follows by using (132) in (128) to further lower-bound $\log M_{\mathrm{f}}^*(n, \epsilon)$, by observing that $(L_n - 1)m_n = n + \mathcal{O}(n^{2/3})$, and by performing a Taylor expansion of $Q^{-1}(\cdot)$ in (131) around 2ϵ.

Proof of (132): To simplify the notation, we set

$$
Z_{k,\ell,i}^{(b)} \triangleq \imath_{P_b^{(n)}, W_k}(\varkappa_{\ell,i}(b); \mathbb{Y}_{k,\ell,i}(b)) \quad (133)
$$

$$
\overline{Z}_{k,\ell_1,\ell_2} \triangleq \sum_{\ell=\ell_1}^{\ell_2} \sum_{i=1}^{m_n} Z_{k,\ell,i}^{(B_\ell)}. \quad (134)
$$

The RVs $Z_{k,\ell,i}^{(b)}$ are independent for all k, b, ℓ, and i. Note also that $\overline{Z}_{k,1,L_n} = \imath_k(\varkappa(B^{L_n}); \mathbb{Y}_k(B^{L_n}))$. Hence, the left-hand side of (132) can be rewritten as $\mathbb{P}\big[\overline{Z}_{k,1,L_n} < \gamma\big]$.

Let $\ell(i) \triangleq \lceil i/m_n \rceil$. It will turn out convenient to define the following random quantities as well:

$$
\begin{aligned}
Z_i^* &\triangleq \sum_{k \in \{1,2\}} \eta_k \imath_{P^*, W_k}\Big(\varkappa_{\ell(i), i-(\ell(i)-1)m_n}(B_\ell); \\
&\qquad\qquad\qquad \mathbb{Y}_{k,\ell(i), i-(\ell(i)-1)m_n}(B_\ell)\Big) \quad (135)
\end{aligned}
$$

$$
\overline{Z}^* \triangleq \sum_{i=1}^{(L_n-1)m_n} Z_i^* \quad (136)
$$

$$
A \triangleq \sum_{\ell=1}^{L_n-1} \sum_{i=1}^{m_n} \sum_{k \in \{1,2\}} \eta_k \log \frac{P_{B_\ell}^{(n)} W_k(\mathbb{Y}_{k,\ell,i}(B_\ell))}{P_{Y_k}^*(\mathbb{Y}_{k,\ell,i}(B_\ell))} \quad (137)
$$

$$
E \triangleq \overline{Z}_{1,1,L_n-1} - \overline{Z}_{2,1,L_n-1}. \quad (138)
$$

In the remainder of the proof, we shall make use of the following decomposition:

$$
\overline{Z}_{k,1,L_n-1} = \overline{Z}^* - A - \eta_{\bar{k}}(-1)^k E. \quad (139)
$$

We next show that $\overline{Z}^*$ is accurately approximated by a normal distribution with mean $(L_n - 1)m_n C$ and variance $(L_n - 1)m_n V$, whereas both A and E are of order $n^{1/3}$. To prove these claims, we rely on the Berry-Esseen-type central limit theorem for martingales given in Theorem 11, and also on the Hoeffding inequality [20], on the Azuma-Hoeffding inequality [19, Th. 2.2.1], and a stabilization lemma, which we state in Lemma 14.

Lemma 12 (Hoeffding inequality [20]): If $X_1, \ldots, X_n$ are independent and $a_i \leq X_i \leq b_i$, for $i = 1, \ldots, n$, then for $t > 0$,

$$
\mathbb{P}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq nt\right] \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \quad (140)
$$

Lemma 13 (Azuma-Hoeffding inequality [19]): Let $\{X_i, \mathcal{F}_i\}_{i=0}^n$ be a real-valued martingale sequence. Suppose that there exist nonnegative real numbers $d_1, \ldots, d_n$, such that $|X_i - X_{i-1}| \leq d_k$ almost surely for all $k \in \{1, \ldots, n\}$. Then, for every $t > 0$,

$$
\mathbb{P}[|X_n - X_0| > t] \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n d_i^2}\right). \quad (141)
$$

Lemma 14 (Stabilization lemma): Let $\{X_\ell^{(1)}\}$ and $\{X_\ell^{(2)}\}$ be i.i.d. RVs with mean $\mu_1 > 0 > \mu_2$ satisfying

$$
\mathbb{P}\big[X_\ell^{(b)} \geq v\big] \leq e^{-\beta |v - \mu_b|_+^2} \quad (142)
$$

and

$$\mathbb{P}\left[X_\ell^{(b)} \leq v\right] \leq e^{-\beta|\mu_b - v|_+^2} \qquad (143)$$

for all $b \in \{1, 2\}$ and all $\ell \in \mathbb{N}$. Define the sequence $\{Y_\ell\}$ as follows: $Y_1 = X_1^{(2)}$ and

$$Y_\ell = \begin{cases} Y_{\ell-1} + X_\ell^{(1)} & \text{for} \quad Y_{\ell-1} < 0 \\ Y_{\ell-1} + X_\ell^{(2)} & \text{for} \quad Y_{\ell-1} \geq 0 \end{cases} \qquad (144)$$

for $\ell \geq 2$. Let $c \geq 1$ satisfy

$$\min\{\mu_1, -\mu_2\} \geq \sqrt{\frac{\pi}{\beta}} \exp\left(\frac{c^2}{4}\right). \qquad (145)$$

Then

$$\mathbb{P}[|Y_\ell| \geq v] \leq 2e^{-c\sqrt{\beta}(v - \mu_1 + \mu_2)}. \qquad (146)$$

The proof of this lemma is delegated to Appendix VIII.

Let now $c_3 > 0$ be an arbitrary constant and define the thresholds

$$\gamma_1 \triangleq \gamma - m_n C + n^{1/3} \log n \qquad (147)$$
$$\gamma_2 \triangleq (L_n - 1) m_n C - c_3 \sqrt{n} \log n. \qquad (148)$$

Here, γ is given in (131). Roughly speaking, γ_1 is chosen so as to be the 2ϵ-quantile of $\overline{Z}^* - A$ for large n, whereas γ_2 is needed to upper-bound the probability of the (rare) event that $\overline{Z}^* - A$ is far smaller than its mean. Note also that $\gamma_1 > \gamma_2$ for sufficiently large n. Define the four disjoint events

$$\mathcal{E}_{1,m} \triangleq \left\{\overline{Z}^* - A \in [\gamma_2, \gamma_1), T = m\right\}, \quad m \in \{1, 2\} \qquad (149)$$
$$\mathcal{E}_2 \triangleq \left\{\overline{Z}^* - A \geq \gamma_1\right\} \qquad (150)$$
$$\mathcal{E}_3 \triangleq \left\{\overline{Z}^* - A < \gamma_2\right\}. \qquad (151)$$

Using the events defined above, we obtain the following upper bound by an application of the union bound and the inequality $\mathbb{P}[A \cap B] \leq \mathbb{P}[A]$, which holds for any two events A and B,

$$\mathbb{P}\left[\overline{Z}_{k,1,L_n} < \gamma\right]$$
$$= \mathbb{P}\Big[\left(\mathcal{E}_{1,k} \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right) \cup \left(\mathcal{E}_{1,\bar{k}} \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right)$$
$$\cup \left(\mathcal{E}_2 \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right) \cup \left(\mathcal{E}_3 \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right)\Big] \ (152)$$
$$\leq \mathbb{P}\left[\mathcal{E}_{1,\bar{k}}\right]$$
$$+ \mathbb{P}\left[\mathcal{E}_{1,k} \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right] + \mathbb{P}\left[\mathcal{E}_2 \cap \left\{\overline{Z}_{k,1,L_n} < \gamma\right\}\right]$$
$$+ \mathbb{P}\left[\mathcal{E}_3\right]. \qquad (153)$$

In the following, we shall upper-bound each of these four probabilities. It turns out that only $\mathbb{P}\left[\mathcal{E}_{1,\bar{k}}\right]$ yields a nonvanishing contribution, which converges to ϵ from below at a rate $\mathcal{O}(n^{-1/2} \log n)$ as $n \to \infty$. The remaining three probabilities all vanish at a rate of $\mathcal{O}(1/n)$. This implies that there exists a constant $c_1 > 0$ such that the inequality (132) holds for all sufficiently large n.

Bound on $\mathbb{P}\left[\mathcal{E}_{1,\bar{k}}\right]$: We first establish a bound on the upper-tail probability of A in (149). For $b \in \{1, 2\}$, we have

$$D(P_b^{(n)} W_k || P_{Y_k}^*)$$
$$= D(W_k || P_{Y_k}^* | \overline{P}_b^{(n)}) - I_k(\overline{P}_b^{(n)}) + \mathcal{O}(1/n) \qquad (154)$$
$$= \left(D(W_k || P_{Y_k}^* | P^*) - (-1)^b \langle \nabla I_k, \overline{\mathbf{v}}_0 \rangle n^{-1/3}\right)$$
$$\quad - \left(I_k(P^*) - (-1)^b \langle \nabla I_k, \overline{\mathbf{v}}_0 \rangle n^{-1/3}\right) + \mathcal{O}(n^{-2/3}) \ (155)$$
$$= \mathcal{O}(n^{-2/3}). \qquad (156)$$

Here, (154) follows from the Golden formula $I(P_X, P_{Y|X}) = D(P_{Y|X} || Q_Y | P_X) - D(P_Y || Q_Y)$ and from (123); (155) follows from the differentiability of $I_k(\cdot)$ and of $D(W_k || P_{Y_k}^* | \cdot)$ at P^*, by Taylor-expanding $I_k(\cdot)$ and $D(W_k || P_{Y_k}^* | \cdot)$ around P^*, and by using (121); finally, (156) follows because $D(W_k || P_{Y_k}^* | P^*) = I_k(P^*) = C$ for $k \in \{1, 2\}$.

Next, it follows from (121) and from the differentiability of $\log(1 + x)$ around $x = 0$ that there exists a constant $c_4 > 0$ such that

$$\left|\sum_k \eta_k \log \frac{P_b^{(n)} W_k(y)}{P_{Y_k}^*(y)}\right|$$
$$= \left|\sum_k \eta_k \log\left(1 - n^{-1/3}(-1)^b \frac{\sum_{x \in \mathcal{X}} \overline{v}_{0,x} W_k(y|x)}{P_{Y_k}^*(y)}\right)\right|$$
$$\leq c_4 n^{-1/3} \qquad (157)$$

for sufficiently large n. Hence, for sufficiently large n,

$$\mathbb{P}\left[A \geq n^{1/3} \log n\right]$$
$$= \mathbb{P}\left[\sum_{\ell=1}^{L_n - 1} \sum_{i=1}^{m_n} \sum_k \eta_k \log \frac{P_{B_\ell}^{(n)} W_k(\mathbb{Y}_{k,\ell,i}(B_\ell))}{P_{Y_k}^*(\mathbb{Y}_{k,\ell,i}(B_\ell))}\right.$$
$$\left. \geq n^{1/3} \log n\right] \qquad (158)$$
$$\leq \exp\left(-\frac{\left|n^{1/3} \log n - \mathbb{c}\, n^{-2/3}(L_n - 1) m_n\right|_+^2}{2 c_4^2 n^{-2/3}(L_n - 1) m_n}\right) \qquad (159)$$
$$\leq \exp\left(-\mathbb{c} n^{1/3} \log^2 n\right) \qquad (160)$$
$$= \mathcal{O}\left(\frac{1}{n}\right). \qquad (161)$$

Here, (159) follows from the Azuma-Hoeffding inequality (see Lemma 13) applied to the bounded martingale difference sequence (162), shown in the top of the next page, with respect to the filtration

$$\overline{\mathcal{F}}_i \triangleq \sigma\Big\{B_1, \ldots, B_{\ell(i+1)},$$
$$\mathbb{Y}_{1,1}, \ldots, \mathbb{Y}_{1,\ell(i)-1}, \mathbb{Y}_{1,\ell(i),1}, \ldots, \mathbb{Y}_{1,\ell(i),i-(\ell(i)-1)m_n},$$
$$\mathbb{Y}_{2,1}, \ldots, \mathbb{Y}_{2,\ell(i)-1}, \mathbb{Y}_{2,\ell(i),1}, \ldots, \mathbb{Y}_{2,\ell(i),i-(\ell(i)-1)m_n}\Big\}. (163)$$

In the application of the Azuma-Hoeffding inequality (see Lemma 13), we also need (156), which implies (164), shown in the top of the next page.

$$\left\{ \sum_k \eta_k \log \frac{P_{B_{\ell(i)}}^{(n)} W_k(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))}{P_{Y_k}^*(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))} - \mathbb{E}\left[\sum_k \eta_k \log \frac{P_{B_{\ell(i)}}^{(n)} W_k(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))}{P_{Y_k}^*(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))} \middle| \overline{\mathcal{F}}_{i-1} \right] \right\}_{i=1}^{(L_n-1)m_n} \quad (162)$$

$$\sum_{i=1}^{(L_n-1)m_n} \mathbb{E}\left[\sum_k \eta_k \log \frac{P_{B_{\ell(i)}}^{(n)} W_k(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))}{P_{Y_k}^*(\mathbb{Y}_{k,\ell(i),i-(\ell(i)-1)m_n}(B_{\ell(i)}))} \middle| \overline{\mathcal{F}}_{i-1} \right] \le \mathbb{c}(L_n-1)m_n n^{-2/3}. \quad (164)$$

Next, we shall invoke the Berry-Esseen-type central limit theorem for martingales given in Theorem 11. Specifically, we define the increasing filtration

$$\mathcal{F}_i \triangleq \sigma(Z_1^*, \dots, Z_i^*, B_1, \dots, B_{\ell(i+1)}) \quad (165)$$

and the sequence

$$D_i \triangleq \sqrt{\frac{V}{V_{B_{\ell(i)}}^{(n)}}} \left(Z_i^* - \mathbb{E}[Z_i^* | \mathcal{F}_{i-1}] \right) \quad (166)$$

where

$$V_b^{(n)} \triangleq \mathbb{E}_{P_b^{(n)} \times W_k} \left[\mathbb{V}\mathrm{ar}_{W_k} \left[\sum_k \eta_k \iota_{P^*, W_k}(X; Y_k) \middle| X \right] \right] \quad (167)$$

$$= V + \mathcal{O}(n^{-1/3}). \quad (168)$$

One can readily verify that $\{D_i\}$ is a martingale difference sequence since D_i is $\mathcal{F}_i$-measurable ($B_{\ell(i)}$ is $\mathcal{F}_{i-1}$-measurable, which implies that $\mathbb{E}[D_i | \mathcal{F}_{i-1}] = \sqrt{V/V_{B_{\ell(i)}}^{(n)}} \mathbb{E}[Z_i^* - \mathbb{E}[Z_i^* | \mathcal{F}_{i-1}] | \mathcal{F}_{i-1}] = 0$). Furthermore, $|D_i|$ is bounded. Next, for all $\ell \in \{1, \dots, L_n - 1\}$, because $\mathbb{x}_\ell(b)$ is of constant composition for $b \in \{1, 2\}$ and $\ell \in \{1, \dots, L_n\}$, it follows that

$$\sum_{i=(\ell-1)m_n+1}^{\ell m_n} \mathbb{E}[Z_i^* | F_{i-1}]$$

$$= m_n \sum_k \eta_k D(W_k \| P_{Y_k}^* | P_b^{(n)}) \quad (169)$$

$$= m_n \sum_k \eta_k \left(C + \langle \nabla I_k, P_b^{(n)} - P^* \rangle \right) \quad (170)$$

$$= m_n C. \quad (171)$$

Here, (170) follows because $D(W_k \| P_{Y_k}^* | P)$ is linear in $P \in \mathcal{P}(\mathcal{X})$ and (171) follows from Lemma 1. It also follows that

$$\sum_{i=(\ell-1)m_n+1}^{\ell m_n} \mathbb{E}[D_i^2 | \mathcal{F}_{i-1}]$$

$$= \frac{V}{V_{B_\ell}^{(n)}} \sum_{i=(\ell-1)m_n+1}^{\ell m_n} \mathbb{E}\left[(Z_i^* - \mathbb{E}[Z_i^* | \mathcal{F}_{i-1}])^2 \middle| \mathcal{F}_{i-1} \right] \quad (172)$$

$$= m_n V. \quad (173)$$

Recall here the definition of V in (37) and of Z_i^* in (135). Moreover, (168) implies that there exists a constant $d_1 > 0$ such that

$$\frac{V}{V_b^{(n)}} \le \frac{1}{1 - d_1 n^{-1/3}} \quad (174)$$

for all sufficiently large n and for $b \in \{1, 2\}$.

We next upper-bound $\mathbb{P}[\mathcal{E}_{1,\bar{k}}]$ in (175)–(179), shown in the top of the next page. Here, (175) follows from (149), from (161), and because T is uniform on $\{1, 2\}$ and is independent of $(\mathbb{Y}_1, \mathbb{Y}_2)$; (176) follows from (171); and (177) follows from (174). By applying Theorem 11 to the martingale difference sequence $\{D_i\}$ in (177), we observe that, for sufficiently large n, there exists a constant $c_2 > 0$ satisfying

$$\mathbb{P}[\mathcal{E}_{1,\bar{k}}] \le \frac{1}{2} Q\left(\frac{(L_n-1)m_n C - \gamma_1 - n^{1/3}\log n}{\sqrt{V(L_n-1)m_n(1 - d_1 n^{-1/3})}} \right)$$

$$+ \frac{\sqrt{2}c_2 \log n}{\sqrt{n}} \quad (178)$$

$$\le \epsilon - \frac{\sqrt{2}c_1 \log n}{\sqrt{n}} \quad (179)$$

where (179) follows from (131) and (147).

Bound on $\mathbb{P}[\mathcal{E}_{1,k} \cap \{\overline{Z}_{k,1,L_n} < \gamma\}]$: We shall first show that E in (139) is sufficiently concentrated around 0. Specifically, let

$$F_\ell^{(b)} \triangleq \frac{1}{2} \sum_{i=1}^{m_n} (Z_{1,\ell,i}^{(b)} - Z_{2,\ell,i}^{(b)}) \quad (180)$$

for $b \in \{1, 2\}$. The $\{F_\ell^{(b)}\}$ are independent RVs and $E = \sum_{\ell=1}^{L_n-1} F_\ell^{(B_\ell)}$. It follows from (120)–(121), (123), and from Taylor's theorem that there exist convergent sequences[7] $\{\xi_b^{(n)}\}$ such that

$$I_1(P_b^{(n)}) - I_2(P_b^{(n)}) = -2(-1)^b(\rho n^{-1/3} + \xi_b^{(n)} n^{-2/3}). \quad (181)$$

Consequently, since the $\{\mathbb{x}_{\ell,i}(b)\}_{i=1}^{m_n}$ are of constant composition, it follows that $\mathbb{E}[F_\ell^{(b)}] = \mu_b^{(n)}$, where

$$\mu_b^{(n)} \triangleq -(-1)^b(\rho m_n n^{-1/3} + \xi_b^{(n)} m_n n^{-2/3}) = \mathcal{O}(n^{1/3}). \quad (182)$$

Since $F_\ell^{(b)}$ is a sum of independent RVs, it follows from Hoeffding inequality that

$$\mathbb{P}[F_\ell^{(b)} \ge v] \le e^{-\frac{1}{2m_n c_5^2} \left| v - \mu_b^{(n)} \right|_+^2} \quad (183)$$

and

$$\mathbb{P}[F_\ell^{(b)} \le v] \le e^{-\frac{1}{2m_n c_5^2} \left| \mu_b^{(n)} - v \right|_+^2}. \quad (184)$$

Here, $c_5 \ge \max_{\ell,i} \max\{1, |Z_{1,\ell,i} - Z_{2,\ell,i}|\}$. We now choose ρ so that $\rho > \sqrt{4\pi c_5^2} e^{c_5^2}$. Next, we apply Lemma 14 with $X_\ell^{(b)} =$

---

[7]We say that a sequence is convergent if its limit exists and is finite.

$$\mathbb{P}\big[\mathcal{E}_{1,\bar{k}}\big] \le \frac{1}{2}\mathbb{P}\Big[\overline{Z}^* < \gamma_1 + n^{1/3}\log n\Big] + \mathcal{O}(1/n) \tag{175}$$

$$\le \frac{1}{2}\mathbb{P}\left[\sum_{i=1}^{(L_n-1)m_n} \frac{Z_i^* - \mathbb{E}[Z_i^*|\mathcal{F}_{i-1}]}{\sqrt{1-d_1 n^{-1/3}}} < \frac{\gamma_1 + n^{1/3}\log n - (L_n-1)m_n C}{\sqrt{1-d_1 n^{-1/3}}}\right] + \mathcal{O}(1/n) \tag{176}$$

$$\le \frac{1}{2}\mathbb{P}\left[\frac{1}{\sqrt{V(L_n-1)m_n}} \sum_{i=1}^{(L_n-1)m_n} D_i < \frac{\gamma_1 + n^{1/3}\log n - (L_n-1)m_n C}{\sqrt{V(L_n-1)m_n(1-d_1 n^{-1/3})}}\right] + \mathcal{O}(1/n). \tag{177}$$

$F_\ell^{(b)}$, $\beta = 1/(2m_n c_5^2)$, and $c = 2c_5$ and conclude that, for sufficiently large n,

$$\mathbb{P}\big[|E| \ge n^{1/3}\log n\big]$$
$$\le 2e^{-\sqrt{\frac{2}{m_n}}(n^{1/3}\log n - \mu_1^{(n)} + \mu_2^{(n)})} \le 3/n. \tag{185}$$

In the last step, we used that $\xi_n m_n n^{-2/3} = \mathcal{O}(1)$.

Define

$$\eta_{\max} = \max_k \eta_k. \tag{186}$$

It now follows that there exists a constant $c_6 > 0$ such that

$$\mathbb{P}\big[\mathcal{E}_{1,k} \cap \{\overline{Z}_{k,1,L_n} < \gamma\}\big]$$
$$\le \mathbb{P}\big[\mathcal{E}_{1,k} \cap \{|E| \le n^{1/3}\log n, \overline{Z}_{k,1,L_n} < \gamma\}\big] + 3/n \tag{187}$$
$$\le \mathbb{P}\Big[\overline{Z}_{k,L_n,L_n} < \gamma - \gamma_2 + \eta_{\max} n^{1/3}\log n \big| B_{L_n} = 3+k\Big]$$
$$\quad + 3/n \tag{188}$$
$$\le \exp\Big(-\frac{c_6}{m_n}\big(m_n(C_k - C) - 2c_3\sqrt{n}\log n\big)^2\Big) + 3/n \tag{189}$$
$$= \mathcal{O}(1/n). \tag{190}$$

Here, (187) follows from of (185); (188) follows because if $\mathcal{E}_{1,k}$ occurs then $B_{L_n} = 3+k$ (see (125)), because $\overline{Z}^* - A \ge \gamma_2$, and because $\overline{Z}_{k,1,L_n} = \overline{Z}^* - A - \eta_{\bar{k}}(-1)^k E + \overline{Z}_{k,L_n,L_n}$. Finally, (189) follows from (131), from (148), and from Hoeffding inequality applied to $\overline{Z}_{k,L_n,L_n}$.

Bound on $\mathbb{P}\big[\mathcal{E}_2 \cap \{\overline{Z}_{k,1,L_n} < \gamma\}\big]$: The upper bound on $\mathbb{P}\big[\mathcal{E}_2 \cap \{\overline{Z}_{k,1,L_n} < \gamma\}\big]$ follows by an argument similar to the one we used to establish (190). Specifically,

$$\mathbb{P}\big[\mathcal{E}_2 \cap \{\overline{Z}_{k,1,L_n} < \gamma\}\big]$$
$$\le \mathbb{P}\Big[\overline{Z}^* - A \ge \gamma_1, |E| \le n^{1/3}\log n, \overline{Z}_{k,1,L_n} < \gamma\Big]$$
$$\quad + 3/n \tag{191}$$
$$= \mathbb{P}\Big[\overline{Z}_{k,1,L_n-1} + \eta_{\bar{k}}(-1)^k E \ge \gamma_1$$
$$\quad , |E| \le n^{1/3}\log n, \overline{Z}_{k,1,L_n} < \gamma, B_{L_n} = 3\Big] + 3/n \tag{192}$$
$$\le \mathbb{P}\Big[\overline{Z}_{k,1,L_n-1} \ge \gamma_1 - \eta_{\max} n^{1/3}\log n$$
$$\quad , \overline{Z}_{k,1,L_n} < \gamma, B_{L_n} = 3\Big] + 3/n \tag{193}$$
$$\le \mathbb{P}\Big[\overline{Z}_{k,L_n,L_n} < \gamma - \gamma_1 + \eta_{\max} n^{1/3}\log n, B_{L_n} = 3\Big]$$
$$\quad + 3/n \tag{194}$$
$$\le \mathbb{P}\Big[\overline{Z}_{k,L_n,L_n} < m_n C - (1-\eta_{\max})n^{1/3}\log n \big| B_{L_n} = 3\Big]$$
$$\quad + 3/n. \tag{195}$$

Here, (191) follows from (150), from (185), and from the inequality $\mathbb{P}[A] \le \mathbb{P}[A,B] + \mathbb{P}\big[B^{\complement}\big]$ for events A and B with $B^{\complement}$ being the complement of B; (192) follows from (139) and because $\overline{Z}^* - A \ge \gamma_1$ implies $B_{L_n} = 3$ (see (125) and (150)); (192) follows because $\max_k \eta_k \le \eta_{\max}$; (194) follows because $\overline{Z}_{k,1,L_n} - \overline{Z}_{k,1,L_n-1} = \overline{Z}_{k,L_n,L_n}$; and (195) follows from (131) and (147).

Finally, we apply Hoeffding inequality to the right-hand side of (195) to obtain the desired result:

$$\mathbb{P}\big[\mathcal{E}_2 \cap \{\overline{Z}_{k,1,L_n} < \gamma\}\big] \le \exp\Big(-\frac{\mathbb{c} n^{2/3}\log^2 n}{m_n}\Big) + 3/n \tag{196}$$
$$= \mathcal{O}(1/n). \tag{197}$$

In the application of Hoeffding inequality, we have used that $\eta_{\max} < 1$ and that the conditional expectation of $\overline{Z}_{k,L_n,L_n}$ given $B_{L_n} = 3$ is $m_n C$.

Bound on $\mathbb{P}[\mathcal{E}_3]$: The upper bound on $\mathbb{P}[\mathcal{E}_3]$ follows by an application of Azuma-Hoeffding inequality (see Lemma 13) applied to the martingale difference sequence $\big\{Z_i^* - \mathbb{E}[Z_i^*|\mathcal{F}_{i-1}]\big\}_{i=1}^{(L_n-1)m_n}$ for the filtration

$$\{\mathcal{F}_i = \sigma(Z_1^*, \ldots, Z_i^*, B_1, \ldots, B_{\ell(i+1)})\}_i. \tag{198}$$

Specifically,

$$\mathbb{P}[\mathcal{E}_3]$$
$$= \mathbb{P}\Big[\overline{Z}^* - A < (L_n-1)m_n C - c_3\sqrt{n}\log n\Big] \tag{199}$$
$$\le \mathbb{P}\Bigg[\sum_{i=1}^{(L_n-1)m_n}\Big(Z_i^* - \mathbb{E}[Z_i^*|\mathcal{F}_{i-1}]\Big)$$
$$\quad < -c_3\sqrt{n}\log n + n^{1/3}\log n\Bigg] + \mathcal{O}(1/n) \tag{200}$$
$$\le \exp\Big(-\frac{\mathbb{c}(c_3\sqrt{n}\log n - n^{1/3}\log n)}{(L_n-1)m_n}\Big) + \mathcal{O}(1/n) \tag{201}$$
$$= \mathcal{O}(1/n). \tag{202}$$

In (199), we have used (148), (200) follows from (161), and (201) follows from Azuma-Hoeffding inequality.

APPENDIX IV
PROOF OF THEOREM 7

We first construct an auxiliary channel that embeds our feedback scheme. To this end, we write the input X_i of the channel as

$$X_i = \overline{X}_i + 2B_i \tag{203}$$

where $i \in \{1, \ldots, n\}$, $\overline{X}_i \in \{1, 2\}$, and

$$B_i = \begin{cases} 0 & \text{if } d(\overline{X}^{i-1}, Y_1^{i-1}) > d(\overline{X}^{i-1}, Y_2^{i-1}) \\ 1 & \text{if } d(\overline{X}^{i-1}, Y_1^{i-1}) < d(\overline{X}^{i-1}, Y_2^{i-1}) \\ \overline{Z}^{(i)} & \text{if } d(\overline{X}^{i-1}, Y_1^{i-1}) = d(\overline{X}^{i-1}, Y_2^{i-1}). \end{cases} \quad (204)$$

Here, $d(\cdot, \cdot)$ denotes the Hamming distance. Through (203), we define the auxiliary component channels $P_{Y_k^n|\overline{X}^n} = P_{X^n|\overline{X}^n} \circ P_{Y_k^n|X^n}$. We note that, by symmetry, the component channels $P_{Y_1^n|\overline{X}^n}$ and $P_{Y_2^n|\overline{X}^n}$ are identical, i.e., $P_{Y_1^n|\overline{X}^n}(y^n|\overline{x}^n) = P_{Y_2^n|\overline{X}^n}(y^n|\overline{x}^n)$ for all $\overline{x}^n \in \{1, 2\}^n$. Hence, the average error probabilities at both decoders are the same. We next apply the RCU bound under minimum distance decoding to the component channel $P_{Y_1^n|\overline{X}^n}$.

The remaining steps follow closely the proof of the RCU bound in [6, Th. 16]. Consider first the average error probability of a binary codebook $\underline{\mathbf{C}} = (\mathbf{C}(1), \ldots, \mathbf{C}(M)) \in \{1, 2\}^{M \times n}$ with a minimum Hamming distance decoder averaged over all randomly generated codebooks with i.i.d. entries uniformly distributed on $\{1, 2\}$:

$$\mathbb{E}[\epsilon_{\underline{\mathbf{C}}}] \leq \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\Bigg[\mathbb{P}\bigg[\bigcup_{j=1; j \neq m}^{M} \Big\{ d(\mathbf{C}(j), Y_1^n) \\ \leq d(\mathbf{C}(m), Y_1^n) \Big\} \Big| \overline{X}^n = \mathbf{C}(m) \bigg] \Bigg]. \quad (205)$$

The symmetry in the probability distribution of the codewords in $\underline{\mathbf{C}}$ implies that all terms in the summation on the right-hand side of (205) are identical. Hence, we obtain the inequalities (206)–(209), shown in the top of the next page. Here,

$$\mathbb{P}_{\overline{X}^n(1), \ldots, \overline{X}^n(M), Y_1^n}(\overline{x}^n(1), \ldots, \overline{x}^n(M), y_1^n) \\ = 2^{-nM} P_{Y_1^n|\overline{X}^n(1)}(y_1^n|\overline{x}^n(1)) \quad (210)$$

and, in (208), we have bounded the probability by one and used the union bound. The desired result follows by noting that $d(\overline{X}^n(2), Y_1^n)$ is a Binomial RV with parameters n and $1/2$. Indeed, $\overline{X}^n(2)$ is uniformly distributed on $\{1, 2\}^n$, independent of $\overline{X}^n(1)$ and Y_1^n.

Appendix V
Proof of Theorem 8

We first establish the result for deterministic VLF codes with $|\mathcal{U}| = 1$. At the end of the proof, we show that (61) holds also when $|\mathcal{U}| > 1$ by using the same argument as in [8, Th. 4].

It will turn out convenient in the remainder of the proof to use the notations of conditional entropy and conditional mutual information given a sigma-algebra $\mathcal{F}$. For two random variables Z and T belonging to discrete spaces $\mathcal{Z}$ and $\mathcal{T}$, these two quantities are defined as follows:

$$\mathcal{H}(Z|\mathcal{F}) \triangleq - \sum_{z \in \mathcal{Z}} \mathbb{P}[Z = z|\mathcal{F}] \log \mathbb{P}[Z = z|\mathcal{F}] \quad (211)$$

$$\mathcal{I}(Z; T|\mathcal{F}) \triangleq \\ \sum_{(z,t) \in \mathcal{Z} \times \mathcal{T}} \mathbb{P}[Z = z, T = t|\mathcal{F}] \log \frac{\mathbb{P}[Z = z, T = t|\mathcal{F}]}{\mathbb{P}[Z = z|\mathcal{F}] \mathbb{P}[T = t|\mathcal{F}]}. \quad (212)$$

As in [13], we simply write Y^n in place of $\sigma(Y^n)$ when $\mathcal{F} = \sigma(Y^n)$.

Fix an arbitrary $(\mathbb{n}, M, \epsilon)$-VLF code with $0 \leq \epsilon \leq 1 - 1/M$ and $|\mathcal{U}| = 1$. Now, by [10, Lem. 1] (see also [13, Lem. 3]), we have that

$$\mathbb{E}[\mathcal{H}(J|Y_k^{\tau_k})] \leq h(\epsilon) + \epsilon \log M. \quad (213)$$

Here, the expectation is with respect to $Y_1^\infty, Y_2^\infty, \tau_1$ and τ_2. Using that $\log M = H(J) = \mathbb{E}[\mathcal{H}(J|Y_k^0)]$ and that $\min_{k \in \{1,2\}} a_k \leq \sum_k \eta_k a_k$ for all $(a_1, a_2) \in \mathbb{R}_+^2$, we obtain

$$(1 - \epsilon) \log M$$
$$\leq \min_k \Big\{ \mathbb{E}[\mathcal{H}(J|Y_k^0) - \mathcal{H}(J|Y_k^{\tau_k})] \Big\} + h(\epsilon) \quad (214)$$
$$\leq \sum_k \eta_k \Big(\mathbb{E}[\mathcal{H}(J|Y_k^0) - \mathcal{H}(J|Y_k^{\tau_k})] \Big) + h(\epsilon). \quad (215)$$

We continue from (215) by using the telescoping sum $\sum_{i=1}^{\tau_k} \big(\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big) = \mathcal{H}(J|Y_k^0) - \mathcal{H}(J|Y_k^{\tau_k})$ to obtain the following chain of inequalities:

$$(1 - \epsilon) \log M - h(\epsilon)$$
$$\leq \sum_k \eta_k \mathbb{E}\Bigg[\sum_{i=1}^{\tau_k} \big(\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big) \Bigg] \quad (216)$$
$$= \sum_k \eta_k \mathbb{E}\Bigg[\sum_{i=1}^{\infty} \mathbb{1}\{\tau_k \geq i\} \big(\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big) \Bigg] \quad (217)$$
$$= \sum_{i=1}^{\infty} \mathbb{E}\Bigg[\sum_k \eta_k \mathbb{1}\{\tau_k \geq i\} \big(\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big) \Bigg]. \quad (218)$$

Here, (218) follows from Fubini's theorem [21, Th. 18.3] because $\sum_{i=1}^{\infty} |\mathbb{1}\{\tau_k \geq i\} (\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i))| \leq \tau_k \log M$ implies that $\mathbb{1}\{\tau_k \geq i\} (\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i))$ is integrable. Next, we apply to (218) the law of total expectation, where the outer expectation is with respect to (Y_1^{i-1}, Y_2^{i-1}) and the inner expectation is with respect to $(Y_{1,i}, Y_{2,i})$, and use that the event $\{\tau_k \geq i\}$ is $\mathcal{F}_{k,i-1}$-measurable to obtain

$$(1 - \epsilon) \log M - h(\epsilon)$$
$$\leq \sum_{i=1}^{\infty} \mathbb{E}\Bigg[\mathbb{E}\bigg[\sum_k \eta_k \mathbb{1}\{\tau_k \geq i\} \\ \times \big(\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big) \Big| Y_1^{i-1}, Y_2^{i-1} \bigg] \Bigg] \quad (219)$$
$$= \sum_{i=1}^{\infty} \mathbb{E}\Bigg[\sum_k \eta_k \mathbb{1}\{\tau_k \geq i\} \\ \times \mathbb{E}\big[\mathcal{H}(J|Y_k^{i-1}) - \mathcal{H}(J|Y_k^i) \big| Y_1^{i-1}, Y_2^{i-1} \big] \Bigg] \quad (220)$$
$$= \sum_{i=1}^{\infty} \mathbb{E}\Bigg[\sum_k \eta_k \mathbb{1}\{\tau_k \geq i\} \mathcal{I}(J; Y_{k,i}|Y_k^{i-1}) \Bigg] \quad (221)$$
$$= \sum_{i=1}^{\infty} \Bigg(\mathbb{P}[\tau_1 \geq i, \tau_2 \geq i] \\ \times \mathbb{E}\bigg[\sum_k \eta_k \mathcal{I}(J; Y_{k,i}|Y_k^{i-1}) \Big| \tau_1 \geq i, \tau_2 \geq i \bigg] \\ + \sum_k \eta_k \mathbb{P}[\tau_k \geq i, \tau_{\overline{k}} < i]$$

$$\mathbb{E}\big[\epsilon_{\underline{\mathbf{C}}}\big] \le \mathbb{P}_{\overline{X}^n(1),\dots,\overline{X}^n(M),Y_1^n}\left[\bigcup_{j=2}^M \Big\{ d(\overline{X}^n(j), Y_1^n) \le d(\overline{X}^n(1), Y_1^n) \Big\}\right] \tag{206}$$

$$= \mathbb{E}_{\overline{X}^n(1),Y_1^n}\left[\mathbb{P}_{\overline{X}^n(2),\dots,\overline{X}^n(M)}\left[\bigcup_{j=2}^M \Big\{ d(\overline{X}^n(j), Y_1^n) \le d(\overline{X}_1^n, Y_1^n) \Big\}\Big|\overline{X}_1^n, Y_1^n\right]\right] \tag{207}$$

$$\le \mathbb{E}_{\overline{X}^n(1),Y_1^n}\left[\min\Big\{1, (M-1)\mathbb{P}_{\overline{X}^n(2)}\Big[d(\overline{X}^n(2), Y_1^n) \le d(\overline{X}^n(1), Y_1^n)\Big|\overline{X}^n(1), Y_1^n\Big]\Big\}\right] \tag{208}$$

$$= \sum_{t=0}^n P_{Z_1^{(n)}}(t) \min\Big\{1, (M-1)\mathbb{P}\Big[d(\overline{X}^n(2), Y_1^n) \le t\Big]\Big\}. \tag{209}$$

$$\times \mathbb{E}\Big[\mathcal{I}(J; Y_{k,i}|Y_k^{i-1})\Big|\tau_k \ge i, \tau_{\bar{k}} < i\Big]\Big). \tag{222}$$

Now, the Markov chain $(Y_1^{i-1}, Y_2^{i-1}, J) - X_i - (Y_{1,i}, Y_{2,i})$ and the data-processing inequality imply that

$$\mathcal{I}(J; Y_{k,i}|Y_k^{i-1}) \le \mathcal{I}(X_i; Y_{k,i}|Y_k^{i-1}). \tag{223}$$

We also note that the chain of inequalities (224)–(227) in the top of the next page hold. Here, (225) follows from Jensen's inequality and from the concavity of mutual information, (226) follows again from the concavity of mutual information, and (227) follows from Lemma 1. Similarly, we also have that $\mathcal{I}(X_i, Y_{k,i}|Y_k^{i-1}) \le C_k$. Substituting (223) and (227) in (222), we find that

$$(1-\epsilon)\log M - h(\epsilon)$$
$$\le \sum_{i=1}^\infty \mathbb{P}[\tau_1 \ge i, \tau_2 \ge i]\, C$$
$$\quad + \sum_k \eta_k C_k \sum_{i=1}^\infty \mathbb{P}[\tau_k \ge i, \tau_{\bar{k}} < i] \tag{228}$$

$$= \mathbb{E}\Big[\min_k \tau_k\Big]C + \sum_k \mathbb{E}\big[|\tau_k - \tau_{\bar{k}}|_+\big]\eta_k C_k \tag{229}$$

$$\le \mathbb{E}\Big[\min_k \tau_k\Big]C + \mathbb{E}\big[|\tau_k - \tau_{\bar{k}}|\big]\max_k \eta_k C_k \tag{230}$$

$$= \mathbb{E}\Big[\max_k \tau_k\Big]C + \mathbb{E}\big[|\tau_k - \tau_{\bar{k}}|\big]\Big(\max_k \eta_k C_k - C\Big) \tag{231}$$

$$\le \mathsf{n}C. \tag{232}$$

In (230), we have used that $\sum_k \mathbb{E}\big[|\tau_k - \tau_{\bar{k}}|_+\big] = \mathbb{E}\big[|\tau_k - \tau_{\bar{k}}|\big]$ and in (232), we have used (17) and that

$$\eta_1 C_1 \le \eta_1 C + \eta_1 \langle \nabla I_1, P_1^* - P^* \rangle \tag{233}$$
$$= \eta_1 C - \eta_2 \langle \nabla I_2, P_1^* - P^* \rangle \tag{234}$$
$$\le \eta_1 C + \eta_2 C \tag{235}$$
$$= C \tag{236}$$

where (233) follows from the concavity of mutual information, (234) follows from Lemma 1, and (235) follows because, by the concavity of mutual information, $C + \langle \nabla I_2, P_1^* - P^* \rangle \ge I_2(P_1^*) \ge 0$. The same argument shows that $\eta_2 C_2 \le C$.

Now, consider an arbitrary $(\mathsf{n}, M, \epsilon)$-VLF code with $0 \le \epsilon \le 1 - 1/M$ but without a constraint on $|\mathcal{U}|$. Then, the argument

outlined above shows that

$$\log M \le C\mathbb{E}\big[\max_k \tau_k \big| U = u\big]$$
$$\quad + \sum_k \eta_k \Big[h\Big(\mathbb{P}\big[J \ne \hat{J}_k | U = u\big]\Big)$$
$$\quad + \mathbb{P}\big[J \ne \hat{J}_k | U = u\big]\log M\Big] \tag{237}$$

for every $u \in \mathcal{U}$. By taking expectations with respect to U on both sides and by applying Jensen's inequality to the binary entropy function, we obtain

$$\log M$$
$$\le C\mathbb{E}\big[\max_k \tau_k\big]$$
$$\quad + \sum_k \eta_k \Big(h\Big(\mathbb{P}\big[J \ne \hat{J}_k\big]\Big) + \mathbb{P}\big[J \ne \hat{J}_k\big]\log M\Big) \tag{238}$$
$$\le C\mathbb{E}\big[\max_k \tau_k\big] + h(\epsilon) + \epsilon \log M. \tag{239}$$

Here, we have also used that $\mathbb{P}\big[J \ne \hat{J}_k\big] \le \epsilon \le 1 - 1/M$, that $h(x) + x\log M$ is an increasing function of $x \in [0, 1 - 1/M]$, and that $\eta_1 + \eta_2 = 1$. Solving (239) for $\log M$ establishes the desired result in (61).

APPENDIX VI
PROOF OF THEOREM 9

The proof follows closely the proofs of [8, Th. 3] and [14, Th. 1]. We let Υ be a Bernoulli RV with $\mathbb{P}[\Upsilon = 1] = q$ and let its probability mass function be given by P_Υ. The RV U has the following domain $\mathcal{U}$ and probability mass function P_U

$$\mathcal{U} \triangleq \{0,1\} \times \underbrace{\mathcal{X}^m \times \cdots \times \mathcal{X}^m}_{SLm \text{ times}} \times \mathcal{X}^\infty \tag{240}$$

$$P_U \triangleq P_\Upsilon \times \underbrace{P_1 \times \cdots \times P_1}_{Lm \text{ times}} \times \cdots \times \underbrace{P_\Upsilon \times \cdots \times P_\Upsilon}_{Lm \text{ times}}$$
$$\quad \times (P^*)^\infty. \tag{241}$$

The realization of $U = u$ implicitly defines codebooks $\{\mathbf{C}_{j,\ell}^{(u)}(b) \in \mathcal{X}^m\}$ and $\{\bar{\mathbf{C}}_j^{(u)} \in \mathcal{X}^\infty\}$ for $(j, b, \ell) \in \mathcal{M} \times \{1, 2\} \times \{1, \dots, L\}$. The encoding functions are defined as

$$f_{\nu(\ell,i)}(u, j, y_1^{\nu(\ell,i-1)}, y_2^{\nu(\ell,i-1)})$$
$$\triangleq \Big(\mathbf{C}_{j,\ell}^{(u)}\Big(h_\ell\Big(x^{\nu(\ell,0)}, y_1^{\nu(\ell,0)}, y_2^{\nu(\ell,0)}\Big)\Big)\Big)_i \tag{242}$$
$$f_{Lm+n_B+i}(u, j, \cdot, \cdot) \triangleq (\bar{\mathbf{C}}_j^{(u)})_i \tag{243}$$

$$\mathbb{E}\left[\sum_k \eta_k \mathcal{I}(X_i; Y_{k,i}|Y_k^{i-1})\Big|\tau_1 \geq i, \tau_2 \geq i\right] = \sum_k \eta_k \mathbb{E}\left[I_k\left(P_{X_i|Y_1^{i-1},Y_2^{i-1}}\left(\cdot|Y_1^{i-1},Y_2^{i-1}\right)\right)\Big|\tau_1 \geq i, \tau_2 \geq i\right] \tag{224}$$

$$\leq \sum_k \eta_k I_k\left(\mathbb{E}\left[P_{X_i|Y_1^{i-1},Y_2^{i-1}}\left(\cdot|Y_1^{i-1},Y_2^{i-1}\right)\Big|\tau_1 \geq i, \tau_2 \geq i\right]\right) \tag{225}$$

$$\leq C + \sum_k \eta_k \left\langle \nabla I_k, \mathbb{E}\left[P_{X_i|Y_1^{i-1},Y_2^{i-1}}\left(\cdot|Y_1^{i-1},Y_2^{i-1}\right)\Big|\tau_1 \geq i, \tau_2 \geq i\right] - P^*\right\rangle \tag{226}$$

$$= C. \tag{227}$$

---

for $\ell \in \{1,\ldots,L\}$, $i \in \{1,\ldots,m\}$, and $i \in \{1,\ldots,\infty\}$ (recall that $\nu(\ell,i) = (\ell-1)m + i$). At the time indices $Lm+1,\ldots,Lm+n_B$, the transmitter sends B^L using a fixed-blocklength code. To keep the notation compact, we omit the superscript (u) in the remaining part of the proof.

At time $n \geq Lm + n_B$, decoder k computes the information densities

$$A_{k,n}(j)$$
$$\triangleq \sum_{\ell=1}^L \sum_{i=1}^m \imath_{P_{B_\ell},W_k}\left(f_{\nu(\ell,i)}(U, Y_1^{\nu(\ell,i)-1}, Y_2^{\nu(\ell,i)-1}); Y_{k,\nu(\ell,i)}\right)$$
$$+ \sum_{i=Lm+n_B+1}^n \imath_{P^*,W_k}(f_i(U,j,\cdot,\cdot); Y_{k,i}) \tag{244}$$

for $j \in \mathcal{M}$. Define the stopping times

$$\tau_k(j) = \mathbb{1}\{\Upsilon=0\}\min\Big\{\inf\{n \geq Lm+n_B : A_{k,n}(j) \geq \gamma\}$$
$$, \tau_{\max} + n_B\Big\} \tag{245}$$

and let τ_k^* be the time at which decoder k makes the final decision:

$$\tau_k^* \triangleq \min_{j \in \mathcal{M}} \tau_k(j). \tag{246}$$

The output of decoder k at time τ_k^* is

$$g_{k,\tau_k^*}(U, Y_k^{\tau_k^*}, B^L) \triangleq \max\{j \in \mathcal{M} : \tau_k(j) = \tau_k^*\}. \tag{247}$$

When $\Upsilon = 1$, we have $\tau_k^* = 0$, and hence the decoder $g_{k,\tau_k^*}(U, Y_k^{\tau_k^*})$ outputs M. The average blocklength is then given by

$$\mathbb{E}\left[\max_k \tau_k^*\right]$$
$$= (1-q)\mathbb{E}\left[\max_k \tau_k^*|\Upsilon=0\right] \tag{248}$$
$$\leq (1-q)\frac{1}{M}\sum_{j=1}^M \mathbb{E}\left[\max_k \tau_k(j)|J=j,\Upsilon=0\right] \tag{249}$$
$$= (1-q)\mathbb{E}[\max_k \tau_k(1)|J=1,\Upsilon=0] \tag{250}$$
$$= (1-q)\left(\mathbb{E}\left[\max_k \tau_k\right] + n_B\right) \tag{251}$$

where the expectation is over J, Y_1^∞, Y_2^∞, and U (B^L is deterministic given these RVs). Here, (249) follows from (246); (250)

follows from symmetry; and (251) follows from the definition of τ_k in (62). We next bound the average error probability:

$$\mathbb{P}\left[g_{k,\tau_k^*}(U, Y_k^{\tau_k^*}, B^L) \neq J\right]$$
$$\leq q + (1-q)\mathbb{P}\left[g_{k,\tau_k^*}(U, Y_k^{\tau_k^*}, B^L) \neq J|\Upsilon=0\right]$$
$$+ (1-q)\epsilon^*(n_B, S^L) \tag{252}$$
$$\leq q + (1-q)\mathbb{P}\left[g_{k,\tau_k^*}(U, Y_k^{\tau_k^*}, B^L) \neq 1|J=1,\Upsilon=0\right]$$
$$+ (1-q)\epsilon^*(n_B, S^L). \tag{253}$$

Here, (253) follows from (247). We proceed with bounding probability term in (253):

$$\mathbb{P}\left[g_{k,\tau_k^*}(U, Y_k^{\tau_k^*}, B^L) \neq 1|J=1,\Upsilon=0\right]$$
$$\leq \mathbb{P}[\tau_k(1) \geq \tau_k^*|\Upsilon=0] \tag{254}$$
$$= \mathbb{P}\left[\bigcup_{j=2}^M \{\tau_k(1) \geq \tau_k(j)\}\Big|\Upsilon=0\right] \tag{255}$$
$$\leq (M-1)\mathbb{P}[\tau_k(1) \geq \tau_k(2)|\Upsilon=0] \tag{256}$$
$$= (M-1)\mathbb{P}[\tau_k \geq \bar{\tau}_k]. \tag{257}$$

Here, (257) follows by noting that, given $J=1$, the RVs

$$(A_{k,n}(1), A_{k,n}(2), \tau_k(1), \tau_k(2)) \tag{258}$$

where the RVs $\{A_{k,n}(j)\}$ are defined in (244), have the same joint distribution as

$$\left(A_{k,n}(X^n; Y_k^n), A_{k,n}(\bar{X}^n; Y_k^n), \tau_k + n_B, \bar{\tau}_k + n_B\right). \tag{259}$$

We conclude the proof by noting that (251), (253), and (257) imply that the tuple $(U, h_\ell, f_n, \{g_{k,n}\}, \{\tau_k^*\})$ defines an $(\mathbb{n}, M, \epsilon)$-VLSF code satisfying (66) and (67).

APPENDIX VII
PROOF OF THEOREM 10

We let $\bar{\mathbb{n}}$ be a variable related to the average blocklength $\mathbb{n}$ (it is roughly equal to $(1-\epsilon)\bar{\mathbb{n}}$), and we analyze the achievability bound in Theorem 9 in the limit $\bar{\mathbb{n}} \to \infty$. We set $S=2$ and let κ be the smallest integer strictly larger than $C^{-1}\log 2$. We also set

$$L = L_{\bar{\mathbb{n}}} \triangleq \lfloor(\bar{\mathbb{n}} - \sqrt{\bar{\mathbb{n}}}\log\bar{\mathbb{n}})^{1/3}\rfloor \tag{260}$$

$$m = m_{\bar{\mathbb{n}}} \triangleq \left\lfloor\frac{(\bar{\mathbb{n}} - \sqrt{\bar{\mathbb{n}}}\log\bar{\mathbb{n}})^{1/3}}{L_{\bar{\mathbb{n}}}}\right\rfloor \tag{261}$$

$$\bar{\mathbb{n}}_- \triangleq L_{\bar{\mathbb{n}}}m_{\bar{\mathbb{n}}} \tag{262}$$

$$\tau_{\max} = \bar{\mathbb{n}}_+ \triangleq \lfloor\bar{\mathbb{n}} + \sqrt{\bar{\mathbb{n}}}\log\bar{\mathbb{n}}\rfloor \tag{263}$$

and $n_B \triangleq \kappa L_{\bar{\mathbb{n}}} = \mathcal{O}(\bar{\mathbb{n}}^{1/3})$. Note that $\bar{\mathbb{n}}_+ - \bar{\mathbb{n}}_- = \mathcal{O}(\sqrt{\bar{\mathbb{n}}} \log \bar{\mathbb{n}})$. As in the proof of Theorem 5, we have that

$$\epsilon^*(n_B, 2^{L_{\bar{\mathbb{n}}}}) \leq \exp(-\omega L_{\bar{\mathbb{n}}}) \triangleq \epsilon_{\bar{\mathbb{n}}}^* \tag{264}$$

for some positive constant ω. In Appendix VII-A, we prove that there exists a constant c_1 such that, for all sufficiently large $\bar{\mathbb{n}}$,

$$\max_k \mathbb{P}[\tau_k = \bar{\mathbb{n}}_+] \leq \frac{c_1}{\bar{\mathbb{n}}}. \tag{265}$$

Next, we let $\overline{\mathbf{v}}_0 \in \mathbb{R}_0^{|\mathcal{X}|}$ be an arbitrary vector satisfying $\langle \nabla I_1, \overline{\mathbf{v}}_0 \rangle > 0 > \langle \nabla I_2, \overline{\mathbf{v}}_0 \rangle$. We define the probability distributions P_1 and P_2 as follows

$$P_b^{(\mathbb{n})} \triangleq P^* - (-1)^b \overline{\mathbf{v}}_0 n^{-1/3} \tag{266}$$

for $b \in \{1, 2\}$. For sufficiently large $\mathbb{n}$, $P_b^{(\bar{\mathbb{n}})}$ are in the neighborhood of P^* and are legitimate probability distributions (recall that $P^*(x) > 0, \forall x \in \mathcal{X}$). Moreover, for $\ell \in \{1, \ldots, L_{\bar{\mathbb{n}}}\}$, we recursively define $\{h_\ell\}_{\ell=1}^{L_{\bar{\mathbb{n}}}}$ and $\{B_\ell\}_{\ell=1}^{L_{\bar{\mathbb{n}}}}$ as follows

$$h_\ell(X^{\nu(\ell,0)}, Y_1^{\nu(\ell,0)}, Y_2^{\nu(\ell,0)})$$
$$\triangleq 1 + \mathbb{1}\left\{ \overline{Z}_{1,\nu(\ell,0)} \geq \overline{Z}_{2,\nu(\ell,0)} \right\} \tag{267}$$

$$B_\ell \triangleq h_\ell(X^{\nu(\ell,0)}, Y_1^{\nu(\ell,0)}, Y_2^{\nu(\ell,0)}). \tag{268}$$

Here, $\nu(\ell, i) = (\ell-1)m_n + i$ and

$$Z_{k,t} \triangleq \begin{cases} \imath_{P_{B_{\ell(t)}}^{(\bar{\mathbb{n}})}, W_k}(X_t; Y_{k,t}) & \text{for } t \leq L_{\bar{\mathbb{n}}} m_{\bar{\mathbb{n}}} \\ \imath_{P^*, W_k}(X_t; Y_{k,t}) & \text{for } t > L_{\bar{\mathbb{n}}} m_{\bar{\mathbb{n}}} \end{cases} \tag{269}$$

$$\overline{Z}_{k,t} \triangleq \sum_{i=1}^t Z_{k,i}. \tag{270}$$

In view of Theorem 9, we set for all integers $\bar{\mathbb{n}} > 0$

$$\gamma_{\bar{\mathbb{n}}} \triangleq C\bar{\mathbb{n}} - 2\bar{\mathbb{n}}^{1/3} \log \bar{\mathbb{n}} \tag{271}$$

$$q_{\bar{\mathbb{n}}} \triangleq \frac{\bar{\mathbb{n}}(\epsilon - \epsilon_{\bar{\mathbb{n}}}^* - c_1/\bar{\mathbb{n}}) - 1}{\bar{\mathbb{n}} - 1} \tag{272}$$

$$M_{\bar{\mathbb{n}}} \triangleq \lfloor \exp(\gamma_{\bar{\mathbb{n}}} - \log \bar{\mathbb{n}}) \rfloor. \tag{273}$$

Thus, we conclude that

$$q_{\bar{\mathbb{n}}} + (1 - q_{\bar{\mathbb{n}}}) \Big((M_{\bar{\mathbb{n}}} - 1) \exp(-\gamma_{\bar{\mathbb{n}}}) + \epsilon^*(n_B, 2^{L_{\bar{\mathbb{n}}}})$$
$$+ \max_k \mathbb{P}[\tau_k = \tau_{\max}] \Big)$$
$$\leq \frac{\bar{\mathbb{n}}(\epsilon - \epsilon_{\bar{\mathbb{n}}}^* - c_1/\bar{\mathbb{n}}) - 1}{\bar{\mathbb{n}} - 1}$$
$$+ \frac{\bar{\mathbb{n}}(1 - \epsilon + \epsilon_{\bar{\mathbb{n}}}^* + c_1/\bar{\mathbb{n}})}{\bar{\mathbb{n}} - 1} \frac{1}{\bar{\mathbb{n}}} + \epsilon_{\bar{\mathbb{n}}}^* + c_1/\bar{\mathbb{n}} \tag{274}$$
$$= \epsilon. \tag{275}$$

In Appendix VII-B, we shall prove that there exists an integer $\bar{\mathbb{n}}_0 \geq 0$ such that, for all $\bar{\mathbb{n}} \geq \bar{\mathbb{n}}_0$,

$$\mathbb{E}\left[\max_k \tau_k \right] \leq \bar{\mathbb{n}}. \tag{276}$$

Here, τ_k is defined by (62) with γ replaced by $\gamma_{\bar{\mathbb{n}}}$. It then follows that

$$(1 - q_{\bar{\mathbb{n}}}) \Big(\mathbb{E}\left[\max_k \tau_k \right] + n_B \Big)$$
$$\leq \underbrace{\frac{\bar{\mathbb{n}}(1 - \epsilon + \epsilon_{\bar{\mathbb{n}}}^* + c_1/\bar{\mathbb{n}})}{\bar{\mathbb{n}} - 1}(\bar{\mathbb{n}} + \kappa L_{\bar{\mathbb{n}}})}_{\triangleq \mathbb{n}_{\bar{\mathbb{n}}}}. \tag{277}$$

We conclude the proof by invoking Theorem 9 with $q = q_{\bar{\mathbb{n}}}$, $\gamma = \gamma_{\bar{\mathbb{n}}}$, and $M = M_{\bar{\mathbb{n}}}$, which implies that there exists a sequence of $(\mathbb{n}_{\bar{\mathbb{n}}}, M_{\bar{\mathbb{n}}}, \epsilon)$-VLF codes for all $\bar{\mathbb{n}} \geq \bar{\mathbb{n}}_0$ satisfying

$$\log M_{\bar{\mathbb{n}}} \geq C\bar{\mathbb{n}} - 2\bar{\mathbb{n}}^{1/3} \log \bar{\mathbb{n}} - \log \bar{\mathbb{n}} \tag{278}$$
$$= \frac{\mathbb{n}_{\bar{\mathbb{n}}} C}{1 - \epsilon} + \mathcal{O}(\mathbb{n}_{\bar{\mathbb{n}}}^{1/3} \log \mathbb{n}_{\bar{\mathbb{n}}}). \tag{279}$$

Before proceeding to the proofs of (265) and (276), we first introduce some notation. It follows from Taylor's theorem and from Lemma 1 that there exist positive convergent sequences $\{\zeta_{\bar{\mathbb{n}}}^{(1)}\}$ and $\{\zeta_{\bar{\mathbb{n}}}^{(2)}\}$ such that

$$\sum_k \eta_k I_k(P_b^{(\bar{\mathbb{n}})}) = C - \bar{\mathbb{n}}^{-2/3} \zeta_{\bar{\mathbb{n}}}^{(b)} \tag{280}$$

for $b \in \{1, 2\}$. Let also $\zeta_{\bar{\mathbb{n}}}^{(\max)} \triangleq \max_{b \in \{0,1\}} \zeta_{\bar{\mathbb{n}}}^{(b)}$ and let $\mu_{b,\bar{\mathbb{n}}}$ for $b \in \{1, 2\}$ be defined as

$$\mu_{b,\bar{\mathbb{n}}} \triangleq m_{\bar{\mathbb{n}}} \left[I_1(P_b^{(\bar{\mathbb{n}})}) - I_2(P_b^{(\bar{\mathbb{n}})}) \right] = \mathcal{O}(\bar{\mathbb{n}}^{1/3}). \tag{281}$$

We shall use the following decomposition for $t \geq \bar{\mathbb{n}}_-$

$$\overline{Z}_{k,t} = \overline{Z}_t - \eta_{\bar{k}}(-1)^k \overline{E}_t. \tag{282}$$

Here,

$$Z_t \triangleq \sum_k \eta_k Z_{k,t} \tag{283}$$

$$\overline{Z}_t \triangleq \sum_{i=1}^t Z_i \tag{284}$$

$$\overline{E}_t \triangleq \sum_{i=1}^t (Z_{1,i} - Z_{2,i}). \tag{285}$$

Some observations are in order. It follows from (280) that $\{\overline{Z}_t - tC + \bar{\mathbb{n}}^{-2/3} t \zeta_{\bar{\mathbb{n}}}^{(\max)}\}$ is a sub-martingale and that $\{\overline{Z}_t - tC\}$ is a super-martingale with respect to the filtration $\{\sigma(\overline{Z}_t)\}_t$ (notice that $\sum_k \eta_k Z_{k,t} \leq C$ by concavity of mutual information and by the definition of $\{\eta_k\}$). Next, observe that $\overline{E}_t - \overline{E}_{\bar{\mathbb{n}}_-}$ is a sum of i.i.d. RVs and that, by applying Lemma 14 as in the proof of Theorem 5, we obtain the following concentration inequality for $\overline{E}_{\bar{\mathbb{n}}_-}$

$$\mathbb{P}\left[|\overline{E}_{\bar{\mathbb{n}}_-}| \geq \xi \right] \leq 2e^{-\sqrt{\frac{2}{m_{\bar{\mathbb{n}}}}} |\xi - \mu_{1,\bar{\mathbb{n}}} + \mu_{2,\bar{\mathbb{n}}}|_+}. \tag{286}$$

A. Proof of (265)

As shown in the chain of inequalities (287)–(292) in the top of the next page, (265) readily follows by an application of the Azuma-Hoeffding inequality (see Lemma 13). Here, (288) follows from the decomposition (282), (289) follows by setting $\eta_{\max} = \max_k \eta_k$, (290) follows from (286), which implies that

$$\mathbb{P}\left[\eta_{\max} |E_{\bar{\mathbb{n}}_-}| \geq \frac{1}{2} \bar{\mathbb{n}}^{1/3} \log \bar{\mathbb{n}} \right] \leq \mathcal{O}\left(\frac{1}{\bar{\mathbb{n}}} \right). \tag{293}$$

Finally, (291) follows from the Azuma-Hoeffding inequality (see Lemma 13) applied to the sub-martingale

$$\Big\{ \overline{Z}_{\min\{t,\bar{\mathbb{n}}_-\}} + \mathbb{1}\{t > \bar{\mathbb{n}}_-\} (\overline{Z}_{k,t} - \overline{Z}_{k,\bar{\mathbb{n}}_-})$$
$$- tC + t\bar{\mathbb{n}}^{-2/3} \zeta_{\bar{\mathbb{n}}}^{(\max)} \Big\}_{t=1}^{\bar{\mathbb{n}}_+ - 1}. \tag{294}$$

$$\max_k \mathbb{P}\Big[\tau_k = \bar{\mathbb{n}}_+\Big] \le \max_k \mathbb{P}\Big[\overline{Z}_{k,\bar{\mathbb{n}}_+-1} < \gamma_{\bar{\mathbb{n}}}\Big] \tag{287}$$

$$= \max_k \mathbb{P}\Big[\min_k\Big\{\overline{Z}_{\bar{\mathbb{n}}_-} - \eta_k(-1)^k \overline{E}_{\bar{\mathbb{n}}_-} + (\overline{Z}_{k,\bar{\mathbb{n}}_+-1} - \overline{Z}_{k,\bar{\mathbb{n}}_-})\Big\} < \gamma_{\bar{\mathbb{n}}}\Big] \tag{288}$$

$$\le \max_k \mathbb{P}\Big[\overline{Z}_{\bar{\mathbb{n}}_-} + \min_k\{\overline{Z}_{k,\bar{\mathbb{n}}_+-1} - \overline{Z}_{k,\bar{\mathbb{n}}_-}\} < \gamma_{\bar{\mathbb{n}}} + \eta_{\max}|\overline{E}_{\bar{\mathbb{n}}_-}|\Big] \tag{289}$$

$$\le \max_k \mathbb{P}\Big[\overline{Z}_{\bar{\mathbb{n}}_-} + \overline{Z}_{k,\bar{\mathbb{n}}_+-1} - \overline{Z}_{k,\bar{\mathbb{n}}_-} < \gamma_{\bar{\mathbb{n}}} + \frac{1}{2}\bar{\mathbb{n}}^{1/3}\log\bar{\mathbb{n}}\Big] + \mathcal{O}(1/\bar{\mathbb{n}}) \tag{290}$$

$$\le 2\exp\left(-\frac{\mathbb{c}\big|(\bar{\mathbb{n}}_+-1)C - (\bar{\mathbb{n}}_+-1)\bar{\mathbb{n}}^{-2/3}\zeta_{\bar{\mathbb{n}}}^{(\max)} - \gamma_{\bar{\mathbb{n}}} - \frac{1}{2}\bar{\mathbb{n}}^{1/3}\log\bar{\mathbb{n}}\big|_+^2}{\bar{\mathbb{n}}_+ - 1}\right) + \mathcal{O}(1/\bar{\mathbb{n}}) \tag{291}$$

$$= \mathcal{O}(1/\bar{\mathbb{n}}). \tag{292}$$

B. Proof of (276)

The idea is to upper-bound $\mathbb{E}[\max_k \tau_k]$ via the following stopping time:

$$\widetilde{\tau} \triangleq \inf\{t \ge \bar{\mathbb{n}}_- : \overline{Z}_t \ge \widetilde{\gamma}_{\bar{\mathbb{n}}} \text{ or } t = \bar{\mathbb{n}}_+\} \tag{295}$$

where

$$\widetilde{\gamma}_{\bar{\mathbb{n}}} \triangleq \gamma_{\bar{\mathbb{n}}} + \bar{\mathbb{n}}^{1/3}\log\bar{\mathbb{n}}. \tag{296}$$

Intuitively, since $\widetilde{\gamma}_{\bar{\mathbb{n}}} - \gamma_{\bar{\mathbb{n}}} = \bar{\mathbb{n}}^{1/3}\log\bar{\mathbb{n}}$, the stopping time $\widetilde{\tau}$ is rarely smaller than $\max_k \tau_k$. In addition, the expectation of $\widetilde{\tau}$ can be upper-bounded by means of Doob's optional stopping theorem. This is simpler than upper-bounding the expectation of the maximum $\max_k \tau_k$ of the two stopping times. Formally, by using that $\max \tau_k \le \mathbb{n}_+$, we upper-bound $\mathbb{E}[\max_k \tau_k]$ in terms of $\mathbb{E}[\widetilde{\tau}]$ as follows:

$$\mathbb{E}\Big[\max_k \tau_k\Big] \le \mathbb{E}[\widetilde{\tau}] + \bar{\mathbb{n}}_+ \mathbb{P}\Big[\widetilde{\tau} < \max_k \tau_k\Big]. \tag{297}$$

In the remaining part of the proof, we demonstrate that there exist constants c_2 and c_3, which do not depend on $\bar{\mathbb{n}}$, such that the following upper bounds hold for sufficiently large $\bar{\mathbb{n}}$

$$\mathbb{E}[\widetilde{\tau}] \le \frac{\widetilde{\gamma}_{\bar{\mathbb{n}}}}{C} + c_2 \tag{298}$$

$$\bar{\mathbb{n}}_+ \mathbb{P}\Big[\widetilde{\tau} < \max_k \tau_k\Big] \le c_3. \tag{299}$$

Consequently, we have that

$$\mathbb{E}\Big[\max_k \tau_k\Big] \le \frac{\widetilde{\gamma}_{\bar{\mathbb{n}}}}{C} + c_2 + c_3 \le \bar{\mathbb{n}} \tag{300}$$

for all sufficiently large $\bar{\mathbb{n}}$, as desired.

Proof of (298): Applying Doob's optional stopping theorem [22, Th. 10.10] to the sub-martingale $\{\overline{Z}_t - tC + \mathbb{n}^{-2/3}t\zeta_{\bar{\mathbb{n}}}^{(\max)}\}_t$ and the stopping time $\widetilde{\tau}$, we obtain

$$0 \le \mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\big] - \mathbb{E}[\widetilde{\tau}]\Big(C - \bar{\mathbb{n}}^{-2/3}\zeta_{\mathbb{n}}^{(\max)}\Big). \tag{301}$$

Hence, we conclude that there exists a constant c_1 such that, for all sufficiently large $\bar{\mathbb{n}}$,

$$\mathbb{E}[\widetilde{\tau}] \le \frac{\mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\big]}{C}\Big(1 + c_1\bar{\mathbb{n}}^{-2/3}\Big). \tag{302}$$

In order to upper-bound $\mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\big]$, we shall use the decomposition

$$\mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\big]$$
$$= \mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\mathbb{1}\{\overline{Z}_{\bar{\mathbb{n}}_-} < \widetilde{\gamma}_{\bar{\mathbb{n}}}\}\big] + \mathbb{E}\big[\overline{Z}_{\bar{\mathbb{n}}_-}\mathbb{1}\{\overline{Z}_{\bar{\mathbb{n}}_-} \ge \widetilde{\gamma}_{\bar{\mathbb{n}}}\}\big]. \tag{303}$$

This decomposition holds because (295) implies that $\widetilde{\tau} \in \{\bar{\mathbb{n}}_-, \ldots, \bar{\mathbb{n}}_+\}$ and because $\overline{Z}_{\widetilde{\tau}} = \overline{Z}_{\bar{\mathbb{n}}_-}$ if and only if $\overline{Z}_{\bar{\mathbb{n}}_-} \ge \widetilde{\gamma}_{\bar{\mathbb{n}}}$. The first term on the right-hand side of (303) is trivially upper-bounded by $\widetilde{\gamma}_{\bar{\mathbb{n}}}$. The second term is upper-bounded by an application of the following lemma, which relates $\mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\mathbb{1}\{\overline{Z}_{\bar{\mathbb{n}}_-} \ge \xi\}\big]$ to the tail probability of $\overline{Z}_{\bar{\mathbb{n}}_-}$.

Lemma 15: Suppose that $\mathbb{P}[X \ge \xi] \le \exp(-a|\xi - b|_+^2)$ for $a > 0$, $b \in \mathbb{R}$, and $\xi \in \mathbb{R}$. Then, for all $c \ge b$, we have

$$\mathbb{E}[X\mathbb{1}\{X \ge c\}] \le (c + \sqrt{\pi/a})\exp(-a|c - b|_+^2). \tag{304}$$

Proof: The result readily follows using integration and the upper bound $Q(x) \le e^{-x^2/2}$ for all $x \ge 0$. ∎

In order to apply Lemma 15, we need to upper-bound the tail probability $\mathbb{P}\big[\overline{Z}_{\widetilde{\tau}} \ge \xi\big]$. To do so, we apply the Azuma-Hoeffding inequality (see Lemma 13) to the super-martingale $\{\overline{Z}_t - tC\}$ to obtain for $\xi \ge \bar{\mathbb{n}}_- C$

$$\mathbb{P}\big[\overline{Z}_{\bar{\mathbb{n}}_-} \ge \xi\big] \le \exp\left(-\frac{\mathbb{c}|\xi - \bar{\mathbb{n}}_- C|_+^2}{\bar{\mathbb{n}}_-}\right). \tag{305}$$

Let a_0 be a constant not depending on $\mathbb{n}$ such that $|Z_{k,t}| \le a_0$ with probability one. It follows that there exists a positive constant c_2 such that

$$\mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\big]$$
$$= \mathbb{E}\big[\overline{Z}_{\widetilde{\tau}}\mathbb{1}\{\overline{Z}_{\bar{\mathbb{n}}_-} < \widetilde{\gamma}_{\bar{\mathbb{n}}}\}\big] + \mathbb{E}\big[\overline{Z}_{\bar{\mathbb{n}}_-}\mathbb{1}\{\overline{Z}_{\bar{\mathbb{n}}_-} \ge \widetilde{\gamma}_{\bar{\mathbb{n}}}\}\big] \tag{306}$$
$$\le \widetilde{\gamma}_{\bar{\mathbb{n}}} + a_0 + (\widetilde{\gamma}_{\bar{\mathbb{n}}} + \mathbb{c}\sqrt{\bar{\mathbb{n}}_-})\exp\left(-\frac{\mathbb{c}|\widetilde{\gamma}_{\bar{\mathbb{n}}} - \bar{\mathbb{n}}_- C|_+^2}{\bar{\mathbb{n}}_-}\right) \tag{307}$$
$$\le \widetilde{\gamma}_{\bar{\mathbb{n}}} + c_2 \tag{308}$$

for sufficiently large $\bar{\mathbb{n}}$. Here, (307) follows because $|Z_t| \le a_0$ and from Lemma 15, and (308) follows because $\widetilde{\gamma}_{\bar{\mathbb{n}}} - \bar{\mathbb{n}}_- C = \mathcal{O}(\sqrt{\bar{\mathbb{n}}}\log\bar{\mathbb{n}})$.

Proof of (299): We need to prove that $\mathbb{P}\big[\widetilde{\tau} < \max_k \tau_k\big] = \mathcal{O}(1/\mathbb{n})$ as $\mathbb{n} \to \infty$. To establish this, observe that $\widetilde{\tau} < \max_k \tau_k$

implies $\min_k S_{k,\widetilde{\tau}}(X^{\widetilde{\tau}}, Y_k^{\widetilde{\tau}}; B^L) < \gamma_{\bar{n}}$ (see (62)). Now, it follows from $\max_k \tau_k \leq \bar{n}_+$, from (64), and from (282) that

$$\mathbb{P}\left[\widetilde{\tau} < \max_k \tau_k\right]$$

$$= \mathbb{P}\left[\widetilde{\tau} < \max_k \tau_k, \widetilde{\tau} < \bar{n}_+\right] \tag{309}$$

$$\leq \mathbb{P}\left[\min_k\left\{\overline{Z}_{\widetilde{\tau}} - \eta_{\bar{k}}(-1)^k \overline{E}_{\widetilde{\tau}}\right\} < \gamma_{\bar{n}}, \widetilde{\tau} < \bar{n}_+\right] \tag{310}$$

$$\leq \mathbb{P}\left[\min_k\left\{\widetilde{\gamma}_{\bar{n}} - \eta_{\bar{k}}(-1)^k \overline{E}_{\widetilde{\tau}}\right\} < \gamma_{\bar{n}}\right] \tag{311}$$

$$\leq \mathbb{P}\left[\eta_{\max}|\overline{E}_{\widetilde{\tau}}| > \widetilde{\gamma}_{\bar{n}} - \gamma_{\bar{n}}\right] \tag{312}$$

where (311) follows from (295) because, in combination with $\widetilde{\tau} < \bar{n}_+$, the definition in (295) implies that $\overline{Z}_{\widetilde{\tau}} \geq \widetilde{\gamma}_{\bar{n}}$. We continue from (312) as follows

$$\mathbb{P}\left[\widetilde{\tau} < \max_k \tau_k\right]$$

$$\leq \mathbb{P}\left[\eta_{\max}(|\overline{E}_{\bar{n}_-}| + |\overline{E}_{\widetilde{\tau}} - \overline{E}_{\bar{n}_-}|) > \bar{n}^{1/3}\log\bar{n}\right] \tag{313}$$

$$\leq \mathbb{P}\left[\eta_{\max}|\overline{E}_{\widetilde{\tau}} - \overline{E}_{\bar{n}_-}| > \frac{1}{2}\bar{n}^{1/3}\log\bar{n}\right] + \mathcal{O}(1/\bar{n}) \tag{314}$$

$$\leq \mathbb{P}\left[\eta_{\max}\max_{t\in\{\bar{n}_-,\ldots,\bar{n}_+\}}|\overline{E}_t - \overline{E}_{\bar{n}_-}| > \frac{1}{2}\bar{n}^{1/3}\log\bar{n}\right]$$
$$+ \mathcal{O}(1/\bar{n}) \tag{315}$$

$$\leq (\bar{n}_+ - \bar{n}_-) \tag{316}$$

$$\times \max_{t\in\{\bar{n}_-+1,\ldots,\bar{n}_+\}}\mathbb{P}\left[\left|\frac{\overline{E}_t - \overline{E}_{\bar{n}_-}}{t - \bar{n}_-}\right| \geq \frac{\bar{n}^{1/3}\log\bar{n}}{2\eta_{\max}(t-\bar{n}_-)}\right]. \tag{317}$$

Here, (313) follows by the triangle inequality and because $\widetilde{\gamma}_{\bar{n}} - \gamma_{\bar{n}} = \bar{n}^{1/3}\log\bar{n}$; (314) follows from (293); finally, (317) follows from the union bound. We obtain the desired result by applying Hoeffding inequality to $\overline{E}_t - \overline{E}_{\bar{n}_-}$ (indeed, $\overline{E}_t - \overline{E}_{\bar{n}_-}$ is a sum of bounded i.i.d. RVs with zero mean):

$$\mathbb{P}\left[\widetilde{\tau} < \max_k \tau_k\right]$$

$$\leq 2(\bar{n}_+ - \bar{n}_-)\max_{t\in\{\bar{n}_-+1,\ldots,\bar{n}_+\}}\exp\left(-\frac{c\bar{n}^{2/3}\log^2\bar{n}}{t-\bar{n}_-}\right) \tag{318}$$

$$\leq \mathcal{O}(\sqrt{\bar{n}}\log\bar{n})\exp\left(-c\bar{n}^{1/6}\log\bar{n}\right) \tag{319}$$

$$= \mathcal{O}\left(\frac{1}{\bar{n}}\right). \tag{320}$$

We conclude that there exists a positive constant c_3 such that (299) holds.

APPENDIX VIII
PROOF OF LEMMA 14

For convenience, we set $\overline{\mu} \triangleq \mu_1 - \mu_2$ and $\alpha \triangleq c\sqrt{\beta}$. We shall prove the desired result in (146) by mathematical induction. To do so, we start by using (142)–(144) to prove the base case; namely that $\mathbb{P}[Y_1 \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ and that $\mathbb{P}[Y_1 \leq v] \leq e^{-\alpha(-v-\overline{\mu})}$ for all $v \in \mathbb{R}$. Then, we prove the inductive step; namely that the statement $\mathbb{P}[Y_{\ell-1} \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ implies $\mathbb{P}[Y_\ell \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ for all integers $\ell \geq 2$ and $v \in \mathbb{R}$. The inductive step hinges on the assumptions in (142)–(145). Finally, (146) follows by applying an analogous argument to $\mathbb{P}[Y_\ell \leq -v]$ and by mathematical induction.

Initial step: By using (142)–(143) and (144), we have

$$\mathbb{P}[Y_1 \geq v] = \mathbb{P}\left[X_1^{(2)} \geq v\right] \tag{321}$$

$$\leq e^{-\beta|v-\mu_2|_+^2} \tag{322}$$

$$\leq e^{-\alpha(v-\mu_2)+\frac{\alpha^2}{4\beta}} \tag{323}$$

$$\leq e^{-\alpha v} \tag{324}$$

$$\leq e^{-\alpha(v-\overline{\mu})}. \tag{325}$$

Here, (323) follows because $-\beta|x|_+^2 + \alpha x - \alpha^2/(4\beta) \leq 0$ for all $x \in \mathbb{R}$, (324) follows because $\alpha^2/(4\beta) = c^2/4 \leq \sqrt{\pi}e^{c^2/4} = \frac{\alpha}{c}\sqrt{\pi/\beta}e^{c^2/4} \leq -\alpha\mu_2/c \leq -\alpha\mu_2$. Using the similar inequality $\alpha^2/(4\beta) \leq \alpha\mu_1$, we also have that

$$\mathbb{P}[Y_1 \leq v] = \mathbb{P}\left[X_1^{(2)} \leq v\right] \tag{326}$$

$$\leq e^{-\beta|\mu_2-v|_+^2} \tag{327}$$

$$\leq e^{-\alpha(\mu_2-v)+\frac{\alpha^2}{4\beta}} \tag{328}$$

$$\leq e^{-\alpha(-v-\overline{\mu})}. \tag{329}$$

Inductive step: We need to show that $\mathbb{P}[Y_{\ell-1} \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ implies that $\mathbb{P}[Y_\ell \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ for all integers $\ell \geq 2$. Suppose that

$$\mathbb{P}[Y_{\ell-1} \geq v] \leq e^{-\alpha(v-\overline{\mu})} \tag{330}$$

for some integer $\ell \geq 2$. Then, by relating $Y_{\ell-1}$ and Y_ℓ using (144) and by applying (142)–(143), we obtain

$$\mathbb{P}[Y_\ell \geq v]$$

$$= \mathbb{E}\left[\mathbb{1}\{Y_{\ell-1} \geq 0\}\mathbb{P}\left[X_\ell^{(2)} \geq v - Y_{\ell-1}|Y_{\ell-1}\right]\right.$$
$$\left. + \mathbb{1}\{Y_{\ell-1} < 0\}\mathbb{P}\left[X_\ell^{(1)} \geq v - Y_{\ell-1}|Y_{\ell-1}\right]\right] \tag{331}$$

$$\leq \mathbb{E}\left[\mathbb{1}\{Y_{\ell-1} \geq 0\}\exp\left(-\beta|v - Y_{\ell-1} - \mu_2|_+^2\right)\right.$$
$$\left. + \mathbb{1}\{Y_{\ell-1} < 0\}\exp\left(-\beta|v - Y_{\ell-1} - \mu_1|_+^2\right)\right]. \tag{332}$$

To further upper-bound the right-hand side of (332), we maximize over all distributions $Y_{\ell-1}$ satisfying (330):

$$\mathbb{P}[Y_\ell \geq v]$$

$$\leq \sup_{\substack{Z: \\ \mathbb{P}[Z \geq v] \leq \exp(-\alpha(v-\overline{\mu}))}}$$
$$\mathbb{E}\left[\mathbb{1}\{Z \geq 0\}\exp\left(-\beta|v - Z - \mu_2|_+^2\right)\right.$$
$$\left. + \mathbb{1}\{Z < 0\}\exp\left(-\beta|v - \mu_1|_+^2\right)\right] \tag{333}$$

$$\leq \sup_{\substack{Z: \\ \mathbb{P}[Z \geq v] \leq \exp(-\alpha(v-\overline{\mu}))}} \mathbb{E}\left[\exp\left(-\beta|v - Z - \mu_2|_+^2\right)\right] \tag{334}$$

$$= \mathbb{E}\left[\exp\left(-\beta|v - \overline{Z} - \mu_2|_+^2\right)\right]. \tag{335}$$

Here, (334) follows since $\exp\left(-\beta|v - \mu_1|_+^2\right) \leq \exp\left(-\beta|v - z - \mu_2|_+^2\right)$ for all $z \geq \overline{\mu}$ and because Z satisfying $\mathbb{P}[Z \geq \overline{\mu}] = 1$ is a feasible solution to the maximization problem. Moreover, the RV $\overline{Z}$ has cumulative distribution function $F_{\overline{Z}}(z) = 1 - e^{-\alpha|z-\overline{\mu}|_+}$ and (335) thus follows because $\overline{Z}$ stochastically dominates all RVs Z satisfying $\mathbb{P}[Z \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ and because the function

$$f(z) \triangleq \exp\left(-\beta|v - z - \mu_2|_+^2\right) \tag{336}$$

is monotonically nondecreasing in z, which imply that $\mathbb{E}\big[f(\overline{Z})\big] \geq \mathbb{E}\big[f(Z)\big]$.

To establish the inductive step, it remains to show that $\mathbb{E}\big[\exp\big(-\beta|v - \overline{Z} - \mu_2|_+^2\big)\big] \leq e^{-\alpha(v-\overline{\mu})}$ for $v \in \mathbb{R}$. Observe that this inequality is trivially satisfied when $\overline{\mu} \geq v$ implying that we can focus only on the case $v \geq \overline{\mu}$. Now, for $v \geq \overline{\mu}$, we have

$$\mathbb{E}\big[\exp\big(-\beta|v - \overline{Z} - \mu_2|_+^2\big)\big]$$

$$\leq \alpha \int_{\overline{\mu}}^{v-\mu_2} e^{-\beta(v-\mu_2-z)^2 - \alpha(z-\overline{\mu})}\, \mathrm{d}z$$

$$+ \alpha \int_{v-\mu_2}^{\infty} e^{-\alpha(z-\overline{\mu})}\, \mathrm{d}z \tag{337}$$

$$\leq \alpha \int_{-\infty}^{\infty} e^{-\left(\sqrt{\beta}(z-v+\mu_2)+\frac{\alpha}{2\sqrt{\beta}}\right)^2 - \alpha(v-\mu_2-\overline{\mu})+\frac{\alpha^2}{4\beta}}\, \mathrm{d}z$$

$$+ e^{-\alpha(v-\mu_2-\overline{\mu})}. \tag{338}$$

Here, (337) holds because $v - \mu_2 \geq \overline{\mu}$ (recall that $\mu_2 < 0$) and (338) follows from the algebraic identity

$$\beta(v-\mu_2-z)^2 + \alpha(z-\overline{\mu})$$

$$= \left(\sqrt{\beta}(z-v+\mu_2)+\frac{\alpha}{2\sqrt{\beta}}\right)^2 + \alpha(v-\mu_2-\overline{\mu}) - \frac{\alpha^2}{4\beta}. \tag{339}$$

Now, by algebraic manipulations, we obtain from (338)

$$\mathbb{E}\big[\exp\big(-\beta|v - \overline{Z} - \mu_2|_+^2\big)\big]$$

$$\leq \alpha e^{\frac{\alpha^2}{4\beta}} e^{-\alpha(v-\mu_2-\overline{\mu})} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\sqrt{2\beta}z)^2}\, \mathrm{d}z + e^{-\alpha(v-\mu_2-\overline{\mu})} \tag{340}$$

$$= \left(1 + \frac{\sqrt{\pi}\alpha e^{\frac{\alpha^2}{4\beta}}}{\sqrt{\beta}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\xi^2}\, \mathrm{d}\xi\right) e^{-\alpha(v-\mu_2-\overline{\mu})} \tag{341}$$

$$= \left(1 + \frac{\sqrt{\pi}\alpha e^{\frac{\alpha^2}{4\beta}}}{\sqrt{\beta}}\right) e^{\alpha\mu_2} e^{-(v-\overline{\mu})} \tag{342}$$

$$\leq e^{-\alpha(v-\overline{\mu})} \tag{343}$$

where (343) follows from the inequality $\log(1 + z) \leq z$ and (145) because

$$1 + \frac{\sqrt{\pi}\alpha e^{\frac{\alpha^2}{4\beta}}}{\sqrt{\beta}} \leq \exp\left(\frac{\sqrt{\pi}\alpha e^{\frac{\alpha^2}{4\beta}}}{\sqrt{\beta}}\right) \tag{344}$$

$$= \exp\left(\frac{\sqrt{\pi}\alpha e^{\frac{c^2}{4}}}{\sqrt{\beta}}\right) \tag{345}$$

$$\leq \exp(-\alpha\mu_2). \tag{346}$$

Combining (335) and (343), we establish the inductive step, i.e., that $\mathbb{P}[Y_{\ell-1} \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ implies $\mathbb{P}[Y_\ell \geq v] \leq e^{-\alpha(v-\overline{\mu})}$ for all integers $\ell \geq 2$ and $v \in \mathbb{R}$. An analogous argument shows that $\mathbb{P}[Y_{\ell-1} \leq -v] \leq e^{-\alpha(v-\overline{\mu})}$ implies that $\mathbb{P}[Y_\ell \leq -v] \leq e^{-\alpha(v-\overline{\mu})}$ for all integers $\ell \geq 2$ and $v \in \mathbb{R}$.

REFERENCES

[1] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, "Feedback halves the dispersion for some common-message broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, 2017.

[2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge Univ. Press, 2011.

[3] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, pp. 1229–1241, 1959.

[4] J. Wolfowitz, *Coding Theorems of Information Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[5] Y. Polyanskiy, "On dispersion of compound DMCs," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, 2013, pp. 26–32.

[6] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[7] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3629–3644, 2009.

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

[9] Y. Altug and A. B. Wagner, "Feedback can improve the second-order coding performance in discrete memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, Jul. 2014.

[10] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random Transmission time," *Probl. Inf. Transm.*, vol. 12, no. 4, pp. 10–30, 1976.

[11] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

[12] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Teletar, "A simple converse of Burnashev's reliability function," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

[13] L. V. Truong and V. Y. F. Tan, "On the reliability function of the common-message broadcast channel with variable-length feedback," Jan. 2017, arXiv:1701.01530 [cs.IT].

[14] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, "Common-message broadcast channels with feedback in the nonasymptotic regime: Stop feedback," *IEEE Trans. Inf. Theory*, Jul. 2018, to appear.

[15] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd ed.* Hoboken, NJ, USA: John Wiley & Sons, 2012.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems, 2nd ed.* New York, NY, USA: Cambridge Univ. Press, 2012.

[17] M. E. Machkouri and L. Ouchti, "Exact convergence rates in the central limit theorem for a class of martingales," *Bernoulli*, vol. 13, no. 4, pp. 981–999, Nov. 2007.

[18] E. Bolthausen, "Exact convergence rates in some martingale central limit theorems," *The Annals of Probability*, vol. 10, no. 3, pp. 672–688, 1982.

[19] M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding, 2nd ed.* Delft, Netherlands: Now Publisher, 2014.

[20] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[21] P. Billingsley, *Probability and Measure, Anniversary Ed.* Hoboken, NJ, USA: Wiley, 2012.

[22] D. Williams, *Probability with Martingales*. New York, NY, USA: Cambridge Univ. Press, 1991.

Kasper Fløe Trillingsgaard (S'12) received his B.Sc. degree in electrical engineering, his M.Sc. degree in wireless communications, and his Ph.D. degree in electrical engineering from Aalborg University, Denmark, in 2011, 2013, and 2017, respectively. He is currently a postdoctoral researcher at the same institution. He was a visiting student at New Jersey Institute of Technology, NJ, USA, in 2012 and at Chalmers University of Technology, Sweden, in 2014. His research interests are in the areas of information and communication theory.

Wei Yang (S'09–M'15) received the B.E. degree in communication engineering and M.E. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, and the Ph.D. degree in Electrical Engineering from Chalmers University of Technology, Gothenburg, Sweden, in 2015. In the summers of 2012 and 2014, he was a visiting student at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. From 2015 to 2017, he was a postdoctoral research associate at Princeton University, Princeton, NJ. In Sep. 2017, he joined Qualcomm Research, San Diego, CA, where he is now a senior engineer.

Giuseppe Durisi (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2006 to 2010 he was a postdoctoral researcher at ETH Zurich, Zurich, Switzerland. In 2010, he joined Chalmers University of Technology, Gothenburg, Sweden, where he is now professor and co-director of Chalmers information and communication technology Area of Advance.

Dr. Durisi is a senior member of the IEEE. He is the recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and is co-author of a paper that won a "student paper award" at the 2012 International Symposium on Information Theory, and of a paper that won the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. In 2015, he joined the editorial board of the IEEE TRANSACTIONS ON COMMUNICATIONS as associate editor. From 2011 to 2014, he served as publications editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. His research interests are in the areas of communication theory, information theory, and machine learning.

Petar Popovski (S'97–A'98–M'04–SM'10–F'16) is a Professor of Wireless Communications with Aalborg University. He received the Dipl. Ing. degree in electrical engineering and the Magister Ing. degree in communication engineering from the "Sts. Cyril and Methodius" University, Skopje, Republic of Macedonia, in 1997 and 2000, respectively, and the Ph.D. degree from Aalborg University, Denmark, in 2004. He has over 300 publications in journals, conference proceedings, and edited books. He holds over 30 patents and patent applications. He received an ERC Consolidator Grant (2015), the Danish Elite Researcher award (2016), the IEEE Fred W. Ellersick prize (2016), and the IEEE Stephen O. Rice prize (2018). He is currently a Steering Committee Member of IEEE SmartGridComm and previously served as a Steering Committee Member of the IEEE INTERNET OF THINGS JOURNAL. He is also an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. His research interests are in the area of wireless communication and networking, and communication/information theory.