(article starts on next page)

Variable-Length Coding with Stop-Feedback for the Common-Message Broadcast Channel

Kasper Fløe Trillingsgaard*, Wei Yang†, Giuseppe Durisi‡, and Petar Popovski*

* Aalborg University, 9220, Aalborg Øst, Denmark
† Princeton University, 08544, Princeton, USA
‡ Chalmers University of Technology, 41296, Gothenburg, Sweden

Abstract— **This paper investigates the maximum coding rate over a K-user discrete memoryless broadcast channel for the scenario where a common message is transmitted using variable-length stop-feedback codes. Specifically, upon decoding the common message, each decoder sends a stop signal to the encoder, which transmits continuously until it receives all K stop signals. We present nonasymptotic achievability and converse bounds for the maximum coding rate, which strengthen and generalize the bounds previously reported in Trillingsgaard *et al.* (2015) for the two-user case. An asymptotic analysis of these bounds reveal that—contrary to the point-to-point case—the second-order term in the asymptotic expansion of the maximum coding rate decays inversely proportional to the square root of the average blocklength. This holds for certain nontrivial common-message broadcast channels, such as the binary symmetric broadcast channel. Furthermore, we identify conditions under which our converse and achievability bounds are tight up to the second order. Through numerical evaluations, we illustrate that our second-order asymptotic expansion approximates accurately the maximum coding rate and that the speed of convergence to capacity is indeed slower than for the point-to-point case.**

I. INTRODUCTION

We are concerned with the setup in which an encoder wishes to convey a common message over a discrete memoryless broadcast channel with feedback to K decoders. Similarly to the single-decoder case, *full feedback* combined with *fixed-blocklength codes* does not improve capacity, which is given by [1, p. 126]

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

Here, $W_1, \cdots, W_k$ denote the channels to the decoders $1, \cdots, K$, respectively, and the supremum is over all input distributions P. If no feedback is available, the speed at which C is approached as the blocklength n increases is of the order $1/\sqrt{n}$ (see [2]), which is the same as in the single-decoder (point-to-point) case [3].

For point-to-point channels, although full feedback does not increase capacity, it improves dramatically the error exponent, provided that *variable-length codes* are used. This was demonstrated by Burnashev who found that the error exponent for the full-feedback case is given by [4]

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

for all rates $0 < R < \widetilde{C}$. Here, $\widetilde{C}$ denotes the channel capacity for the point-to-point case and $\widetilde{C}_1$ denotes the maximum relative entropy between conditional output distributions. Yamamoto and Itoh [5] proposed a two-phase scheme that attains the error exponent in (2) and [6] provides an alternative and simpler converse proof to (2), which parallels the two-phase scheme proposed in [5].

In this paper, we shall be concerned with the scenario in which the feedback channel is used only to stop transmissions (*stop/decision feedback*). Following [2], we shall refer to variable-length coding schemes relying on stop feedback as variable-length stop-feedback (VLSF) codes. It was shown in [2], [7] that, using only VLSF codes, one can achieve an error exponent $E(R) = \widetilde{C} - R$. No converse result is available.

In the fixed-error regime, Polyanskiy *et al.* [2] found that the speed at which the maximum coding rate convergences to capacity is significantly improved in the presence of full feedback and variable-length codes. Specifically, they showed that

$$\frac{1}{\ell} \log \widetilde{M}_{\mathrm{f}}^*(\ell, \epsilon) = \frac{\widetilde{C}}{1 - \epsilon} - \mathcal{O}\left(\frac{\log \ell}{\ell}\right) \tag{3}$$

where ℓ stands for the average blocklength (average transmission time) and $\widetilde{M}_{\mathrm{f}}^*(\ell, \epsilon)$ is the maximum number of codewords in the point-to-point case. One sees from (3) that no square-root penalty occurs (*zero dispersion*), which implies a fast convergence to the asymptotic limit. It turns out that (3) can be achieved using VLSF codes. In other words, stop feedback is enough to obtain zero dispersion.

This result, however, does not extend to the common-message discrete memoryless broadcast channel (CM-DMBC). Specifically, it was shown in [8] that the dispersion of such channel, computed for the case of VLSF codes (i.e., stop feedback), may be positive. More precisely, the second term in the asymptotic expansion of the maximum coding rate is of order $1/\sqrt{\ell}$ in some cases (cf., (3)). The analysis in [8] is limited to the two-user case, and relies on the restrictive assumption that there exists a unique input distribution P^* that simultaneously maximizes $I(P, W_1)$ and $I(P, W_2)$. Furthermore, the upper and lower bounds on the maximum coding rate provided in [8] do not match up to second order. The purpose of this paper is to refine the results obtained in [8] and to extend them to a broader class of CM-DMBCs.

Contribution: Focusing on VLSF codes, we obtain nonasymptotic achievability and converse bounds on the maximum number of codewords $M_{\mathrm{sf}}^*(\ell, \epsilon)$ with average

blocklength ℓ that can be transmitted on a CM-DMBC with reliability $1 - \epsilon$. Here, the subscript "sf" stands for stop feedback. By analyzing these bounds in the large-ℓ regime, we prove that when the K subchannels are independent (in the sense made precise in (4)) and when the mutual information evaluated at the capacity-achieving input distribution equals C for two or more subchannels, then the asymptotic expansion of $\log M_{\mathrm{sf}}^*(\ell, \epsilon)$ contains a square-root penalty provided that some mild technical conditions are satisfied. Thus, we cannot expect the same fast convergence as in the single-decoder case. We also obtain upper and lower bounds on the second-order term in the asymptotic expansion of $\log M_{\mathrm{sf}}^*(\ell, \epsilon)$, which generalize and tighten the ones reported in [8]. The bounds turn out to match for certain special cases, including the case when the capacity-achieving distribution P^* simultaneously maximizes $I(P, W_1)$ and $I(P, W_2)$ treated in [8]. All proofs are omitted for space constraint; we refer the interested reader to [9].

Notation: We denote the n-dimensional all-zero vector by $\mathbf{0}_n$. Vectors are denoted by boldface letters while their entries are denoted by roman letters. The length of a vector is denoted by $\mathrm{len}(\cdot)$. For a differentiable function $f(\cdot)$, we let $f'(\cdot)$ denote its derivative. Upper case, lower case, and calligraphic letters indicate random variables (RV), deterministic quantities, and sets, respectively. The cardinality of a set is denoted by $|\cdot|$. We let x_m^n denote the tuple $(x_m, \cdots, x_n)$. When $m = 1$, the subscript is sometimes omitted. We denote the set of probability distributions on $\mathcal{A}$ by $\mathcal{P}(\mathcal{A})$. For a random variable X with probability distribution P, we let P^n denote the the joint probability distribution of the vector $[X_1, \cdots, X_n]$, where $X_i \stackrel{\text{i.i.d.}}{\sim} P$. We let $\Phi(x)$ be the cumulative distribution function for the Gaussian distribution. Throughout the paper, the index k belongs always to the set $\mathcal{K} \triangleq \{1, \cdots, K\}$, although this is sometimes not explicitly mentioned. Finally, $\mathbb{N}$ denotes the set of positive integers, $\mathbb{Z}_+ \triangleq \mathbb{N} \cup \{0\}$, the symbol $\mathbb{R}$ indicates the set of real numbers, and $\mathbb{R}_0^n$ denotes the set $\left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0 \right\}$.

II. System Model

A CM-DMBC with K decoders consists of a finite-cardinality input alphabet $\mathcal{X}$, and output alphabets $\{\mathcal{Y}_k\}$, along with K stochastic matrices $\{W_k\}$, where $W_k(y_k|x)$ denotes the probability that $y_k \in \mathcal{Y}_k$ is observed at decoder k given the channel input $x \in \mathcal{X}$. We assume without loss of generality that $\mathcal{X} = \{1, \cdots, |\mathcal{X}|\}$. The outputs at time t are assumed to be conditionally independent given the input, i.e.,

$$P_{Y_{1,t}, \cdots, Y_{K,t}|X_t}(y_{1,t}, \cdots, y_{K,t}|x_t) \triangleq \prod_k W_k(y_{k,t}|x_t). \quad (4)$$

The assumption (4) comes without loss of generality if one is interested in infinite-blocklength analysis with vanishing error probability. However, this ceases to be true in the fixed-error-probability setup considered in the present paper. Let $\mathcal{P}(\mathcal{X})$ be the set of all probability distributions on $\mathcal{X}$. Let also $P \times W_k : (x, y_k) \mapsto P(x)W_k(y_k|x)$ denote the joint distribution of input and output at decoder k. Finally, let $PW_k : y_k \mapsto \sum_{x \in \mathcal{X}} P(x)W_k(y_k|x)$ denote the induced marginal distribution

on $\mathcal{Y}_k$. For every $P \in \mathcal{P}(\mathcal{X})$ and $n \in \mathbb{N}$, the information density is defined as

$$i_{P,W_k}(x^n; y_k^n) \triangleq \sum_{i=1}^n \log \frac{W_k(y_{k,i}|x_i)}{PW_k(y_{k,i})}. \quad (5)$$

We let $I_k(P) \triangleq \mathbb{E}_{P \times W_k}[i_{P,W_k}(X; Y_k)]$ be the mutual information and $V_k(P) \triangleq \mathrm{Var}_{P \times W_k}[i_{P,W_k}(X; Y_k)|X]$ be the conditional information variance.

The capacity of the CM-DMBC is given by (1), where the supremum is over all probability distributions $P \in \mathcal{P}(\mathcal{X})$. We restrict ourselves to the case where the supremum in (1) is achieved by a unique probability distribution P^*. The corresponding (unique) capacity-achieving output distributions are denoted by $P_{Y_k}^*$. Furthermore, the individual capacities of each of the discrete memoryless subchannels $\{W_k\}$ are denoted by $C_k \triangleq \sup_{P \in \mathcal{P}(\mathcal{X})} I_k(P)$. Finally, we let $V_k \triangleq V_k(P^*)$ and let $\nabla I_k(\mathbf{v})$ denote the directional derivative of the mutual informations $I_k(P)$ along the direction $\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}$ at the point P^*

$$\nabla I_k(\mathbf{v}) \triangleq \sum_{x \in \mathcal{X}} v_x D(W_k(\cdot|x)||P_{Y_k}^*) \quad (6)$$

where $D(\cdot||\cdot)$ denotes the Kullback-Liebler divergence.

In addition to (4) and to the uniqueness of P^*, we shall assume that the channel laws $\{W_k\}$ satisfy the following conditions:

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

The first condition is not critical (see [9]), and it is added only to simplify the statement of our results.

We are now ready to formally define a VLSF code for a CM-DMBC.

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

1) A RV $U \in \mathcal{U}$, with $|\mathcal{U}| \leq K + 1$, which is known at both the encoder and the decoders.
2) A sequence of encoders $f_n : \mathcal{U} \times \mathcal{M} \to \mathcal{X}$, each one mapping the message J, drawn uniformly at random from the set $\mathcal{M} \triangleq \{1, \ldots, M\}$, to the channel input $X_n = f_n(U, J)$.
3) Nonnegative integer-valued RVs $\tau_1, \cdots, \tau_K$ that are stopping times with respect to the filtrations $\mathcal{F}_{k,n} = \sigma\{U, Y_k^n\}$ and satisfy $\mathbb{E}\left[\max_k \tau_k \right] \leq \ell$.
4) A sequence of decoders $g_{k,n} : \mathcal{U} \times \mathcal{Y}_k^n \to \mathcal{M}$ satisfying

$$\mathrm{P}\left[J \neq g_{k,\tau_k}(U, Y_k^{\tau_k}) \right] \leq \epsilon, \qquad k \in \mathcal{K}. \quad (7)$$

Remark 1: VLSF codes require a feedback link from each decoder to the encoder. This feedback consists of a 1-bit "stop signal" per decoder, which is sent by decoder k at time τ_k. The encoder continuously transmits until all decoders have fed back a stop signal. Hence, the blocklength is $\max_k \tau_k$. Note also that, differently from the full-feedback case, the encoder output at time n depends on the message and on the common randomness U, but does not depend on the past output signals $\{Y_k^{n-1}\}$.

Remark 2: The RV U serves as common randomness between the transmitter and all receivers, and enables the use of

randomized codes [10]. To establish the bound on the cardinality of U provided in Definition 1, one can proceed as in [2, Th. 19] and use Caratheodory theorem to show that $|\mathcal{U}| \leq K + 2$. This bound can be further improved to $|\mathcal{U}| \leq K + 1$ by using the Fenchel-Eggleston theorem [11, p. 35] in place of Caratheodory theorem.

The maximum number of codewords with average length ℓ and error probability not exceeding ϵ is denoted by

$$M_{\mathrm{sf}}^*(\ell, \epsilon) = \max\{M : \exists(\ell, M, \epsilon)\text{-VLSF code}\}. \quad (8)$$

III. MAIN RESULTS

A. Nonasymptotic Achievability Bound

We provide below a K-user generalization of the nonasymptotic achievability bound reported in [8, Th. 1].[1]

Theorem 1: Fix a probability distribution P_{X^∞} on $\mathcal{X}^\infty$. Let $\gamma \geq 0$ and $0 \leq q \leq 1$ be arbitrary scalars. Let the stopping times τ_k and $\bar{\tau}_k$, $k \in \mathcal{K}$, be defined as follows:

$$\tau_k \triangleq \inf\{n \geq 0 : i_{P_{X^n}, W_k^n}(X^n; Y_k^n) \geq \gamma\} \quad (9)$$

$$\bar{\tau}_k \triangleq \inf\{n \geq 0 : i_{P_{X^n}, W_k^n}(\bar{X}^n; Y_k^n) \geq \gamma\}. \quad (10)$$

Here, the joint probability distribution of $(X^n, \bar{X}^n, Y_1^n, \cdots, Y_K^n)$ is

$$P_{X^n, \bar{X}^n, Y_1^n, \cdots, Y_K^n}(x^n, \bar{x}^n, y_1^n, \cdots, y_K^n)$$
$$= P_{Y_1^n, \cdots, Y_K^n | X^n}(y_1^n, \cdots, y_K^n | x^n) P_{X^n}(x^n) P_{X^n}(\bar{x}^n). \quad (11)$$

For every M, there exists an (ℓ, M, ϵ)-VLSF code such that

$$\ell \leq (1 - q)\mathbb{E}\left[\max_k \tau_k\right] \quad (12)$$

$$\epsilon \leq \max_k\{q + (1 - q)(M - 1)\mathrm{P}[\tau_k \geq \bar{\tau}_k]\}. \quad (13)$$

Remark 3: The constant q in Theorem 1 is used to enable time-sharing. With probability q, the decoders simultaneously send stop signals to the encoder at time 0. The common randomness U can be used to enable this weak form of cooperation among the decoders.

Remark 4: Following the same steps as in [2, Eqs. (111)–(118)], we can further upper-bound ϵ in (13) as follows:

$$\epsilon \leq q + (1 - q)(M - 1)\exp\{-\gamma\}. \quad (14)$$

This bound is easier to evaluate numerically and to analyze asymptotically.

B. Nonasymptotic Converse Bound

Let $\mathfrak{Y}_k$ denote all possible sequences (of arbitrary length) of symbols from $\mathcal{Y}_k$, i.e., $\mathfrak{Y}_k \triangleq \{[]\} \cup \bigcup_{n=1}^\infty \mathcal{Y}_k^n$, where $[]$ stands for the vector of length 0. A subset $\bar{\mathfrak{Y}}_k$ of $\mathfrak{Y}_k$ is called complete prefix-free [2] if and only if, for every $\mathbf{y} \in \mathcal{Y}_k^\infty$, there exists a unique $\bar{\mathbf{y}} \in \bar{\mathfrak{Y}}_k$ such that $\bar{\mathbf{y}}$ is a prefix to $\mathbf{y}$. Let $Q_k^{(\infty)}$ be an arbitrary probability measure on $\mathcal{Y}_k^\infty$ and define the mapping $Q_k : \mathfrak{Y}_k \mapsto [0, 1]$ as follows:

$$Q_k(\bar{\mathbf{y}}) \triangleq \sum_{\substack{\mathbf{y} \in \mathcal{Y}_k^\infty: \\ \bar{\mathbf{y}} = [y_1, \cdots, y_{\mathrm{len}(\bar{y})}]}} Q_k^{(\infty)}(\mathbf{y}), \qquad \bar{\mathbf{y}} \in \mathfrak{Y}_k. \quad (15)$$

We shall use the convention that $[y_1, \cdots, y_{\mathrm{len}(\bar{\mathbf{y}})}] = []$ when $\mathrm{len}(\bar{\mathbf{y}}) = 0$. For every complete prefix-free subset $\bar{\mathfrak{Y}}_k \subset \mathfrak{Y}_k$, we observe that $Q_k(\cdot)$ defines a probability measure on $\bar{\mathfrak{Y}}_k$. Indeed,

$$1 = \sum_{\bar{\mathbf{y}} \in \bar{\mathfrak{Y}}_k} \sum_{\substack{\mathbf{y} \in \mathcal{Y}_k^\infty: \\ \bar{\mathbf{y}} = [y_1, \cdots, y_{\mathrm{len}(\bar{\mathbf{y}})}]}} Q_k^{(\infty)}(\mathbf{y}) = \sum_{\bar{\mathbf{y}} \in \bar{\mathfrak{Y}}_k} Q_k(\bar{\mathbf{y}}). \quad (16)$$

Based on $Q_k(\cdot)$, we define the "mismatched" information density

$$i_k(x^n; y_k^n) \triangleq \log \frac{P_{Y_k^n | X^n}(y_k^n | x^n)}{Q_k(y_k^n)} \quad (17)$$

for $x^n \in \mathcal{X}^n$, $y_k^n \in \mathcal{Y}^n$, and $n \in \mathbb{N}$ (with the convention that $i_k([]; []) = 0$).

We are now ready to state our converse bound.

Theorem 2: For arbitrary probability measure $Q_k^{(\infty)}$ on $\mathcal{Y}_k^\infty$, and arbitrary $M \in \mathbb{N}$, $t \in \mathbb{Z}_+$, $\eta > 0$, and $\varepsilon_k \in (0, 1)$, $k \in \mathcal{K}$, define the following function:

$$L_t(\varepsilon_1, \cdots, \varepsilon_K) \triangleq \max_{x^t \in \mathcal{X}^t} \prod_k \min\Big\{1,$$

$$\mathrm{P}\Big[\max_{0 \leq n \leq t} i_k(x^n; Y_k^n) \geq \log M + \log \eta\Big] + \varepsilon_k\Big\}. \quad (18)$$

Here, the vector x^n contains the first n entries of x^t, and $Y_k^n \sim P_{Y_k^n | X^n = x^n}$. Then, every (ℓ, M, ϵ)-VLSF code must satisfy

$$\ell \geq \min_{\substack{P_U \in \mathcal{P}(\mathcal{U}), \varepsilon_k^{(u)} \in [0,1]: \\ \mathbb{E}_U\left[\varepsilon_k^{(U)}\right] \leq \epsilon + \eta}} \mathbb{E}_U\left[\sum_{t=0}^\infty \Big(1 - L_t(\varepsilon_1^{(U)}, \cdots, \varepsilon_K^{(U)})\Big)\right].$$

$$(19)$$

For the case when $\{W_k\}$ are identical and symmetric,[3] we have the following particularization of Theorem 2.

Corollary 3: For arbitrary $M \in \mathbb{N}$, $t \in \mathbb{Z}_+$, $\eta > 0$, and an arbitrary sequence $\mathbf{x} \in \mathcal{X}^\infty$, let

$$\bar{L}_t = \mathrm{P}\Big[\max_{0 \leq n \leq t} i_{P^*, W_1}(x^n; Y_1^n) \geq \log M + \log \eta\Big] \quad (20)$$

where $Y_k^n \sim P_{Y_k^n | X^n = x^n}$. When $W_1 = \cdots = W_K$ and W_1 is symmetric, every (ℓ, M, ϵ)-VLSF code must satisfy

$$\ell \geq \min_{\substack{P(u), \varepsilon^{(u)} \in [0,1]: \\ \mathbb{E}_U\left[\varepsilon^{(U)}\right] \leq \epsilon + \eta}} \mathbb{E}_U\left[\sum_{t=0}^\infty \Big(1 - \min\Big\{1, \bar{L}_t + \varepsilon^{(U)}\Big\}^K\Big)\right].$$

$$(21)$$

C. Asymptotic Expansion

Analyzing (14) and (19) in the limit $\ell \to \infty$, we obtain the following asymptotic characterization of $M_{\mathrm{sf}}^*(\ell, \epsilon)$.

Theorem 4: Let $V \triangleq (\prod_k V_k)^{1/K}$ and $\varrho_k \triangleq \sqrt{V_k/V}$ and assume, without loss of generality, that $C_1 \geq \cdots \geq C_K$. For every CM-DMBC satisfying

$$1/C_i + i/C_K > i/C \quad (22)$$

---

[1] Note that there is a typo in [8, Eq. (12)]: a maximization over k is missing.
[2] FiXme Note: Remember to find reference for this

[3] A channel is symmetric if the rows and columns of the stochastic channel matrix are permutations of each other [12, p. 189].

for $i \in \{1, \cdots, K-1\}$, and for every $\epsilon \in (0, 1)$, we have[4]

$$\frac{C\ell}{1-\epsilon} - \sqrt{\frac{V\ell}{1-\epsilon}}\Xi_a + o(\sqrt{\ell}) \leq \log M_{\text{sf}}^*(\ell, \epsilon)$$

$$\leq \frac{C\ell}{1-\epsilon} - \sqrt{\frac{V\ell}{1-\epsilon}}\Xi_c + o(\sqrt{\ell}). \quad (23)$$

Here,

$$\Xi_a \triangleq \min_{\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}} \mathbb{E}\left[\max_k \nabla I_k(\mathbf{v}) + \varrho_k Z_k\right] \quad (24)$$

with $Z_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ and

$$\Xi_c \triangleq \mathbb{E}\left[\max_k H_k\right] \quad (25)$$

with the RVs $\{H_k\}$ being independent and with cumulative distribution functions

$$F_{H_k}(w) \triangleq \Phi\left(\frac{w + \nabla I_k(\hat{\mathbf{v}}(w))}{\varrho_k}\right). \quad (26)$$

The function $\hat{\mathbf{v}}(\cdot)$ is defined as follows:[5]

$$\hat{\mathbf{v}}(w) \triangleq \arg\max_{\mathbf{v} \in \mathbb{R}_0^{|\mathcal{X}|}} \prod_k \Phi\left(\frac{w + \nabla I_k(\mathbf{v})}{\varrho_k}\right). \quad (27)$$

Remark 5: The condition (22) is needed only for the converse bound in (23).

Remark 6: When $K = 2$, the condition (22) reduces to $1/C_1 + 1/C_2 > 1/C$. Note that for all CM-DMBC, we have $1/C_1 + 1/C_2 \geq 1/C$. Indeed, suppose that $1/C_1 + 1/C_2 < 1/C$. Then one can achieve a rate larger than C:

$$\max_{\alpha \in [0,1]} \min\{\alpha C_1, (1-\alpha)C_2\} = 1/(1/C_1 + 1/C_2) > C. \quad (28)$$

Theorem 4 does not hold for the special case $1/C_1 + 1/C_2 = 1/C$. In this case, since time-sharing is capacity achieving, one can sequentially transmit to the two users and achieve zero dispersion.

Remark 7: One can verify that $\Xi_a \geq \Xi_c > 0$. See [9] for details.

There are cases where $\Xi_a = \Xi_c$, and hence (23) provides a complete second-order characterization of $\log M_{\text{sf}}^*(\ell, \epsilon)$. This happens when $\mathbf{0}_{|\mathcal{X}|}$ maximizes (27). This occurs for example when P^* simultaneously maximizes $I_k(P)$ for all $k \in \mathcal{K}$. In the following corollary, we provide necessary and sufficient conditions for Theorem 4 to yield an asymptotic expansion that is tight up to the second order.

Corollary 5: We have that $\Xi_a = \Xi_c$ and, hence,

$$\log M_{\text{sf}}^*(\ell, \epsilon) = \frac{C\ell}{1-\epsilon} - \sqrt{\frac{V\ell}{1-\epsilon}}\mathbb{E}\left[\max_k Z_k\right] - o(\sqrt{\ell}) \quad (29)$$

if and only if the following holds

1) The capacity-achieving input distribution P^* simultaneously maximizes $I_k(P)$ for all $k \in \mathcal{K}$, or

2) $V_1 = \cdots = V_K$ and $\sum_k \nabla I_k(\mathbf{e}_{|\mathcal{X}|}(i)) = 0$ for $i \in \{1, \cdots, |\mathcal{X}|\}$. Here, $\mathbf{e}_{|\mathcal{X}|}(i)$ stands for the $|\mathcal{X}|$-dimensional vector whose ith entry is equal to one and the remaining entries are equal to zero.

For broadcast channels that do not satisfy the conditions of Corollary 5, we can tighten the left-hand side of (23) by using an input distribution that is not stationary memoryless.

Theorem 6: Let $V \triangleq (\prod_k V_k)^{1/K}$ and $\varrho_k \triangleq \sqrt{V_k/V}$. Fix a differentiable function $\bar{\mathbf{v}} : \mathbb{R} \mapsto \mathbb{R}_0^{|\mathcal{X}|}$ such that $P^* + C\bar{\mathbf{v}}'(w) \in \mathcal{P}(\mathcal{X})$ for all $w \in \mathbb{R}$. Additionally, define

$$E_k(s) \triangleq C - I_k(P^* + C\bar{\mathbf{v}}'(s)) + \nabla I_k(C\bar{\mathbf{v}}'(s)) \quad (30)$$

and assume that $\int_{-\infty}^{\infty} E_k(s)ds < \infty$ and that $|E_k'(s)| < \infty$. Then, for every CM-DMBC, we have

$$\log M_{\text{sf}}^*(\ell, \epsilon) \geq \frac{C\ell}{1-\epsilon} - \sqrt{\frac{V\ell}{1-\epsilon}}\bar{\Xi}_a - o(\sqrt{\ell}). \quad (31)$$

Here, $\bar{\Xi}_a \triangleq \mathbb{E}\left[\max_k \bar{H}_k\right]$, where the independent RVs $\{\bar{H}_k\}$ are defined by the cumulative distribution functions

$$F_{\bar{H}_k}(w) \triangleq \Phi\left(\frac{1}{\varrho_k}\left(w + \nabla I_k(\bar{\mathbf{v}}(w)) - \int_{-\infty}^{w} \frac{E_k(s)}{C^2}ds\right)\right). \quad (32)$$

If one sets $\bar{\mathbf{v}}(\cdot)$ in Theorem 6 equal to (27), the gap between Ξ_c and $\bar{\Xi}_a$ is only due to the "error" term $E_k(s)$. Interestingly, there are channels for which $E_k(s) = 0$ and, hence, a complete second-order characterization of $\log M_{\text{sf}}^*(\ell, \epsilon)$ is available. The next corollary describes a class of channels for which this is the case.

Corollary 7: Let $\mathcal{X}_1, \cdots, \mathcal{X}_L$ be disjoint sets and let $\mathcal{X} = \cup_{l=1}^{L} \mathcal{X}_l$. Moreover, for $k \in \mathcal{K}$ and $l \in \{1, \cdots, L\}$, let $W_{k,l}$ be a channel from $\mathcal{X}_l$ to $\mathcal{Y}_k$ with capacity-achieving input distribution P_l^* (independent of k), capacity-achieving output distribution $P_{Y_k}^*$ (independent of l), and capacity $C_{k,l}$. Define for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}_k$ the channel $W_k(y|x) = W_{k,l(x)}(y|x)$, where the function $l : \mathcal{X} \mapsto \{1, \cdots, L\}$ is such that $x \in \mathcal{X}_{l(x)}$. Assume that $C_1 \geq \cdots \geq C_K$ and that (22) is satisfied. Define

$$\boldsymbol{\beta}(w) \triangleq \arg\max_{\boldsymbol{\beta} \in \mathbb{R}_0^L} \prod_k \Phi\left(\frac{1}{\varrho_k}\left(w + \sum_{l=1}^{L} \beta_l C_{k,l}\right)\right) \quad (33)$$

and assume that $P^*(x) + CP_{l(x)}^*(x)\beta_{l(x)}'(w) \in [0, 1]$ for every $w \in \mathbb{R}$. Then, for every $\epsilon \in (0, 1)$,

$$\log M_{\text{sf}}^*(\ell, \epsilon) = \frac{C\ell}{1-\epsilon} - \sqrt{\frac{V\ell}{1-\epsilon}}\Xi_c + o(\sqrt{\ell}). \quad (34)$$

The following lemma shows that there exist nontrivial channels that satisfy the conditions of Corollary 7.

Lemma 8: Let $L = 2$, $K = 2$, and $\Delta_1 \triangleq C_{11} - C_{12} > 0 > C_{21} - C_{22} \triangleq \Delta_2$. Let also

$$D \triangleq -\frac{\Delta_1 \varrho_2^2 + \Delta_2 \varrho_1^2}{\Delta_1^2 \varrho_2^2 + \Delta_2^2 \varrho_1^2}. \quad (35)$$

Then the condition $P^*(x) + CP_{l(x)}^*(x)\beta_{l(x)}'(w) \in [0, 1]$ holds for all $x \in \mathcal{X}$ and $w \in \mathbb{R}$ provided that

$$P^*(x) + (-1)^{l(x)+1}CP_{l(x)}^*(x)D \in [0, 1] \quad (36)$$

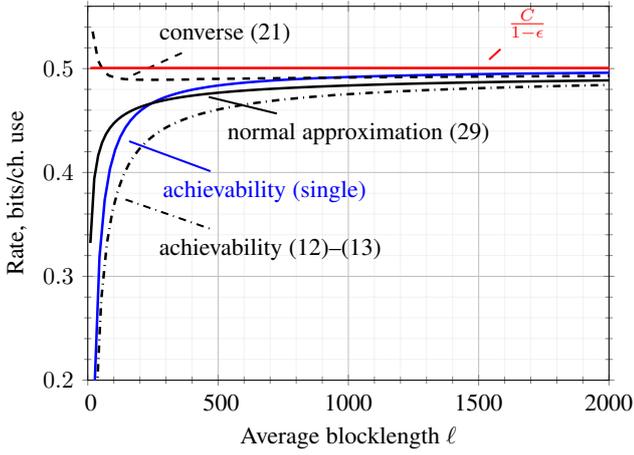for all $x \in \mathcal{X}$ and $(\Delta_1/\varrho_1 + \Delta_2/\varrho_2)(\varrho_2 - \varrho_1) \geq 0$.

---

[4]The subscripts "a" and "c" in Ξ_a and Ξ_c stand for achievability and converse, respectively.

[5]If the maximizer in (27) is not unique, $\hat{\mathbf{v}}(w)$ is chosen arbitrarily from the set of maximizers.

Fig. 1. Comparison between achievability bound (12)-(13), converse bound (19), and normal approximation (29) for $\epsilon = 10^{-3}$ and $K = 2$ users. The individual channels in the CM-DMBC are BSCs with crossover probability 0.11. The normal approximation corresponds to the asymptotic expansion in (29) with the $o(\cdot)$ term neglected. The blue curve labeled "achievability (single)" corresponds to the single-user achievability bound in [2, Th. 3] evaluated for a BSC with crossover probability 0.11 and $\epsilon = 10^{-3}$.
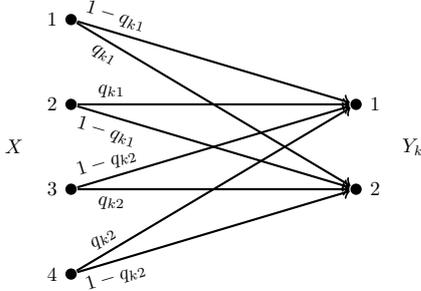

Fig. 2. Asymmetric channels $\{W_k\}$ that obey the conditions in Corollary 7 for the channel parameters $q_{11} = 0.01$, $q_{12} = 0.40$, $q_{21} = 0.15$, and $q_{22} = 0.10$. The channels consist of two BSCs with common outputs.

IV. NUMERICAL EXAMPLES

A. Binary Symmetric Channels

Let W_1 and W_2 be two BSCs, each with crossover probability. Note that W_1 and W_2 are symmetric [12, p. 185] and have the same capacity-achieving input distribution. In this section, we evaluate the bounds presented in Theorem 1, Corollary 3, and Corollary 5. The bounds are depicted in Fig. 1 for the case $q = 0.11$ and $\epsilon = 10^{-3}$. The capacity-achieving input distribution P^* of the individual BSCs is Bern$(1/2)$, their capacity is given by $1 - h_b(q)$, where $h_b(\cdot)$ denotes the binary entropy function, and the directional derivatives $\nabla I_k(\cdot)$ of the mutual information at P^* are zero. Furthermore, for $Y_k^n \sim P_{Y_k^n|X^n = x^n}$, the information densities $i_{P^*, W_k}(x^n; Y_k^n)$ satisfy

$$i_{P^*, W_k}(x^n; Y_k^n) \sim n \log(2 - 2q) + \log \frac{q}{1-q} \sum_{j=1}^{n} Z_{k,j} \quad (37)$$

where $Z_{k,j} \overset{\text{i.i.d.}}{\sim} \text{Bern}(q)$. We observe that the distribution of the information density in (37) is independent of x^n. For the converse bound in Corollary 3, the value of $\eta > 0$ is chosen by numerical optimization. For the achievability bound in Theorem 1, we fix $P_{X^\infty} = (P^*)^\infty$ and note that $\tau_1, \cdots, \tau_K$ are independent RVs. This allows us to compute $\mathbb{E}[\max_k \tau_k]$ directly from the cumulative distribution function $F_{\tau_k}(\cdot)$ of τ_k. To compute the upper bound (13), we use the change of measure technique in [2, p. 4911].

We observe that, in the two-user case, the speed of convergence to the asymptotic limit is indeed slower than for the single-user case (the curve marked "achievability (single)" in Fig. 1, which is the point-to-point achievability bound reported in [2, Th. 3]). In particular, for $\ell \geq 1000$ and $K = 2$, our converse bound is strictly below the single-user achievability bound.

B. Asymmetric Channels

Next, we illustrate through an example that Theorem 6 indeed improves over Theorem 4. We consider W_1 and W_2 with $q_{11} = 0.01, q_{12} = 0.40, q_{21} = 0.15$, and $q_{22} = 0.10$ depicted in Fig. 2. Both consist of two BSCs with common outputs. If we let $\mathcal{X}_1 = \{1, 2\}$, $\mathcal{X}_2 = \{3, 4\}$, and W_{kl} be BSCs with crossover probability q_{kl} for $k \in \{1, 2\}$ and $l \in \{1, 2\}$, then we can invoke Corollary 7. One can verify that the condition $P^*(x) + C P_{l(x)}^*(x) \beta'_{l(x)}(w) \in [0, 1]$ for $x \in \mathcal{X}$ and $w \in \mathbb{R}$ in Corollary 7 is satisfied using Lemma 8. Therefore, the asymptotic expansion in (34) holds, i.e., the converse bound in Theorem 4 is tight up to second order. The same is not true for (19). Indeed, by computing (24) and (25), we find that $\Xi_c = 0.2630$ and $\Xi_a = 0.3175$.

REFERENCES

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

[2] 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.

[3] ——, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[4] 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.

[5] 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, 1979.

[6] 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.

[7] A. Tchamkerten and E. Teletar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, 2006.

[8] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, "Broadcasting a common message with variable-length stop-feedback codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 2505–2509.

[9] ——, "Variable-length coding with stop-feedback for the common-message broadcast channel (extended version)," Feb. 2016. [Online]. Available: http://arxiv.org/

[10] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[11] H. G. Eggleston, *Convexity*. New York: Cambridge Univ. Press, 2009.

[12] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd*. Hoboken: Wiley Interescience, 2012.