



# CHALMERS

## Chalmers Publication Library

### **Feedback Halves the Dispersion for Some Two-User Broadcast Channels with Common Message**

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

**Proc. IEEE Int. Symp. Inf. Theory (ISIT)**

Citation for the published paper:

Trillingsgaard, K. ; Yang, W. ; Durisi, G. et al. (2017) "Feedback Halves the Dispersion for Some Two-User Broadcast Channels with Common Message". Proc. IEEE Int. Symp. Inf. Theory (ISIT)

Downloaded from: <http://publications.lib.chalmers.se/publication/249349>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

# Feedback Halves the Dispersion for Some Two-User Broadcast Channels with Common Message

Kasper Fløe Trillingsgaard<sup>1</sup>, Wei Yang<sup>2</sup>, Giuseppe Durisi<sup>3</sup>, and Petar Popovski<sup>1</sup>

<sup>1</sup>Aalborg University, 9220, Aalborg Øst, Denmark

<sup>2</sup>Princeton University, 08544, Princeton, USA

<sup>3</sup>Chalmers University of Technology, 41296 Gothenburg, Sweden

**Abstract**—We investigate the maximum coding rate achievable on a two-user broadcast channel for the case where a common message is transmitted using fixed-blocklength codes with feedback. Specifically, we focus on a family of broadcast channels composed of two antisymmetric Z-channels. For this setup, we obtain matching upper and lower bounds on the dispersion term in the asymptotic expansion of the maximum coding rate. These bounds reveal that the dispersion is halved compared to the no-feedback case.

## I. INTRODUCTION

We consider a two-user common-message discrete-time memoryless broadcast channel (CM-BC) with full feedback. As for the point-to-point setup, it is well known that feedback does not improve capacity, which is given by

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

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 there exist CM-BCs for which feedback improves the speed at which the maximum coding rate achievable with *fixed-blocklength codes* approaches capacity as the blocklength increases. Specifically, we shall focus (for simplicity) on the class of CM-BCs consisting of two antisymmetric Z-channels  $W_1$  and  $W_2$  defined as follows:

$$W_1(0|0) = 1 - W_1(1|0) = W_2(1|1) = 1 - W_2(0|1) = 1 \quad (2)$$

$$W_1(0|1) = 1 - W_1(1|1) = W_2(1|0) = 1 - W_2(0|0) = \delta. \quad (3)$$

Here,  $\delta \in (0, 1)$ . Note that for this class of CM-BCs, which we shall refer to as AZ-CM-BCs (AZ stands for antisymmetric Z), the capacity-achieving input distribution (CAID)  $P^*$  is the uniform distribution.

For the no-feedback case, a CM-BC is equivalent to a compound channel, whose second-order coding rates (i.e., the second-order expansion of the maximum coding rate in the limit of large blocklength) were investigated in [1]. There, it was proven that the second-order term in the expansion of the logarithm of maximum number of codewords  $M^*(n, \epsilon)$  as a function of the blocklength  $n$  and the average error probability  $\epsilon \in (0, 1)$  is in general a function of both the *dispersions* [2, Eq. (222)] of the individual component channels and of the directional derivatives of their mutual informations computed at the CAID.

The particular structure of the antisymmetric Z-channel setup considered in this paper yields, however, the following drastic simplification to the second-order expansion of  $\log M^*(n, \epsilon)$  for the no-feedback case:

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + o(\sqrt{n}). \quad (4)$$

Here,  $V$  is the conditional information variance (see [2, Eq. (242)]) evaluated at  $P^*$  which is the same for  $W_1$  and  $W_2$ .

When *variable-length codes* are used, feedback is known to improve the speed at which the maximum coding rate converges to capacity. Specifically, it was shown in [3] that the square-root term in the asymptotic expansion of the maximum coding rate (cf. (4)) vanishes for point-to-point discrete memoryless channels (DMCs), a result known as *zero-dispersion*. Furthermore, this fast convergence to capacity can be achieved using only stop feedback (also known as decision feedback). Namely, the feedback channel is used only to stop transmissions. For CM-BCs, 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 [4].

In the fixed-blocklength-code setup, it is known that feedback does not improve the second-order term for point-to-point DMCs under certain symmetry conditions. Specifically, it was shown in [3] 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 [5], where it was shown that the same holds when the conditional information variance is constant for all input symbols.

*Contributions:* We show that, in the fixed-blocklength-code setup, the availability of feedback yields an improvement to the second-order term of the asymptotic expansion of  $\log M^*(n, \epsilon)$  for CM-BCs composed of two antisymmetric Z-channels. Specifically, we demonstrate that, when  $\epsilon \in (0, 1/2)$ , the second-order term in (4) can be improved to  $\sqrt{nV/2}Q^{-1}(2\epsilon)$  if feedback is available, and that this value is optimal.

The intuition behind our result is as follows: 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. The problem of computing the maximum coding rate then becomes roughly equivalent to the one of computing the  $\epsilon$ -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  $V/2$ . 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 can change the input distribution to either the CAID of  $W_1$  or the one of  $W_2$ . In this way, it can ensure that at least one of the two decoders is successful.

*Extensions:* Our results can be extended to a broader class of CM-BCs. In particular, a feedback scheme similar to the one considered in this paper yields an improved second-order term for all CM-BC for which  $P^*$  is the unique maximizer in (1), it satisfies  $P^*(x) > 0$  for all inputs  $x$ , and it is neither the CAID of  $W_1$  nor of  $W_2$ . Our converse result requires certain additional symmetry conditions.

## II. SYSTEM MODEL

We consider the family of AZ-CM-BCs described in Section I. We also assume that the channel outputs at any given time  $i$  are conditionally independent given the input, namely

$$P_{Y_{1,i}, Y_{2,i} | X_i}(y_{1,i}, y_{2,i} | x_i) \triangleq W_1(y_{1,i} | x_i) W_2(y_{2,i} | x_i). \quad (5)$$

An  $(n, M, \epsilon)$ -feedback code for this channel consists of:

- 1)  $n$  encoding functions  $f_i : \{1, \dots, M\} \times \{0, 1\}^{i-1} \times \{0, 1\}^{i-1} \mapsto \{0, 1\}$ ,  $i = 1, \dots, n$ , mapping the message  $J$ , drawn uniformly from  $\{1, \dots, M\}$ , and the past channel outputs to the channel input  $X_i = f_i(J, Y_1^{i-1}, Y_2^{i-1})$ .
- 2) Two decoders  $g_k : \{0, 1\}^n \mapsto \{1, \dots, M\}$  satisfying<sup>1</sup>

$$\max_{j \in \{1, \dots, M\}} \mathbb{P}[g_k(Y_k^n) \neq j | J = j] \leq \epsilon, \quad k \in \{1, 2\}. \quad (6)$$

The maximum number of codewords with blocklength  $n$  and maximum error probability not exceeding  $\epsilon$  is denoted by

$$M_f^*(n, \epsilon) \triangleq \max\{M : \exists(n, M, \epsilon)\text{-feedback code}\}. \quad (7)$$

Let  $PW_k(\cdot)$  and  $P \times W_k(\cdot)$  be the marginal output distribution and the joint probability distribution of  $(X, Y_k)$ , respectively, induced by the input distribution  $P$ . For every input distribution  $P$  and  $n \in \mathbb{N}$ , we denote the information density between the vectors  $x^n$  and  $y_k^n$  as

$$\iota_{P, W_k}(x^n; y_k^n) \triangleq \sum_{i=1}^n \log \frac{W_k(y_{k,i} | x_i)}{PW_k(y_{k,i})}. \quad (8)$$

We let  $I_k(P) \triangleq \mathbb{E}_{P \times W_k}[\iota_{P, W_k}(X; Y_k)]$  be the mutual information and  $V_k(P) \triangleq \mathbb{E}_P[\text{Var}_{W_k}[\iota_{P, W_k}(X; Y_k) | X]]$  be the conditional information variance. The capacity of the AZ-CM-BC is given in (1) and is achieved by the uniform distribution  $P^*$ . We denote by  $C_1 = C_2 = \tilde{C}$  the capacities of the two component channels. Finally, we let  $V \triangleq V_1(P^*) = V_2(P^*)$  and let  $I'_k$

<sup>1</sup>Throughout the paper, the index  $k$  belongs always to the set  $\{1, 2\}$ , although this is sometimes not explicitly mentioned.

be the derivative of  $I_k([p, 1 - p])$  with respect to  $p$  evaluated at  $p = 1/2$ . The following two properties, which are satisfied for the AZ-CM-BC because of antisymmetry, are crucial for establishing our asymptotic result:

$$I'_1 + I'_2 = 0 \quad (9)$$

and, for  $x \in \{0, 1\}$ ,

$$\text{Var}[\iota_{P^*, W_1}(x; Y_1) + \iota_{P^*, W_2}(x; Y_2) | X = x] = 2V. \quad (10)$$

## III. NONASYMPTOTIC BOUNDS

We first state a Verdú-Han-type converse bound for feedback codes. Its proof (which is omitted) relies on the meta-converse theorem [2, Th. 27] and on the inequality [2, Eq. (106)].

*Theorem 1:* Every  $(n, M, \epsilon)$ -feedback code for the AZ-CM-BC satisfies

$$\epsilon \geq \max_k \mathbb{P} \left[ \sum_{i=1}^n \iota_{P^*, W_k}(Y_{k,i}; f_i(J, Y_1^{i-1}, Y_2^{i-1})) \leq \log M - \eta \right] - e^{-\eta}, \quad \forall \eta > 0. \quad (11)$$

Before stating our achievability bound, we need to introduce some notation. Throughout this section, we let  $S \geq 2$ ,  $L$ ,  $m$ , and  $n_B$  denote arbitrary positive integers. We also let  $\mathcal{S} \triangleq \{1, \dots, S\}$  and  $\mathcal{L} \triangleq \{1, \dots, L\}$ . Furthermore, we set  $\mathbb{A} \triangleq \{0, 1\}^{S L m}$  and index an element  $\mathfrak{x}$  of  $\mathbb{A}$  as follows

$$\mathfrak{x} = (\mathfrak{x}_1(1), \dots, \mathfrak{x}_1(S), \dots, \mathfrak{x}_L(1), \dots, \mathfrak{x}_L(S)). \quad (12)$$

Here,  $\mathfrak{x}_\ell(s) = (\mathfrak{x}_{\ell,1}(s), \dots, \mathfrak{x}_{\ell,m}(s)) \in \{0, 1\}^m$ . For a given vector  $b^\ell \in \mathcal{S}^\ell$ , we let  $\mathfrak{x}^\ell(b^\ell) \triangleq (\mathfrak{x}_1(b_1), \dots, \mathfrak{x}_\ell(b_\ell))$ . Finally, we let  $\epsilon^*(n_B, S^L)$  denote the minimum error probability achievable on an AZ-CM-BC using fixed-blocklength codes with no feedback of blocklength  $n_B$  and with  $S^L$  codewords.

Next, we state our achievability bound.

*Theorem 2:* Let  $P_1, \dots, P_S$  be types of sequences in  $\{0, 1\}^m$ , let  $\tau \in (0, \epsilon)$  and  $\zeta > 0$  be arbitrary, and let  $\{h_\ell\}_{\ell=0}^{L-1}$  be arbitrary mappings from  $\{0, 1\}^{\ell m} \times \{0, 1\}^{\ell m} \times \{0, 1\}^{\ell m}$  to  $\mathcal{S}$ . Then,

$$\begin{aligned} & \log M_f^*(Lm + n_B, \epsilon + \epsilon^*(n_B, S^L)) \\ & \geq \sup \left\{ \gamma : \min_k \mathbb{P}[\iota_k(\mathfrak{x}^L(B^L); \mathbb{Y}_k^L(B^L)) \leq \gamma] < \right. \\ & \quad \left. \epsilon - \tau - e^{-\zeta} \right\} \\ & \quad + \log \tau - SL \log(1 + m) - (L + 1) \log S - \zeta. \end{aligned} \quad (13)$$

Here,  $\mathfrak{x} \in \mathbb{A}$  is an arbitrary element of

$$\mathbb{F} \triangleq \left\{ \mathfrak{x} \in \mathbb{A} : \mathfrak{x}_\ell(b) \text{ has type } P_b, \quad \forall b \in \mathcal{S}, \ell \in \mathcal{L} \right\} \quad (14)$$

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

$$\begin{aligned} \mathbb{P}(y_1, y_2, b^L) & \triangleq \left( \prod_{\ell=1}^L \prod_{s=1}^S \prod_k W_k^m(y_{k,\ell}(s) | \mathfrak{x}_\ell(s)) \right) \\ & \quad \times \prod_{\ell=0}^{L-1} \mathbb{1} \left\{ h_\ell(\mathfrak{x}^\ell(b^\ell), \mathbb{Y}_1^\ell(b^\ell), \mathbb{Y}_2^\ell(b^\ell)) = b_{\ell+1} \right\} \end{aligned} \quad (15)$$

for  $y_{k,\ell}(b) \in \{0, 1\}^m$ . Finally, we have defined

$$v_k(\mathbf{x}_\ell(b_\ell); y_{k,\ell}(b_\ell)) \triangleq \sum_{i=1}^m v_{P_{b_\ell}, W_k}(\mathbf{x}_{\ell,i}(b_\ell); y_{k,\ell,i}(b_\ell)) \quad (16)$$

and

$$v_k(\mathbf{x}^L(b^L); y_k^L(b^L)) \triangleq \sum_{\ell=1}^L v_k(\mathbf{x}_\ell(b_\ell); y_{k,\ell}(b_\ell)). \quad (17)$$

*Sketch of the proof:* Each codeword of length  $Lm$  is divided into  $L$  blocks of length  $m$ . Within each block  $\ell$ , feedback is used to compute the index  $b_\ell$ , which selects one out of the  $S$  available subcodewords (the vectors  $\mathbf{x}_\ell(b_\ell)$ ,  $b_\ell = 1, \dots, S$ ). We then communicate the sequence  $b^L$  to the decoders using a code with  $S^L$  codewords of length  $n_B$ . The desired bound then follows by applying the achievability bound [1, Th. 3] for compound DMCs without feedback.

#### IV. ASYMPTOTIC ANALYSIS

By analyzing Theorem 1 and Theorem 2 in the large-blocklength limit, we obtain the following asymptotic expansion for  $\log M_f^*(n, \epsilon)$ .

*Theorem 3:* For every  $\epsilon \in (0, 1/2)$ , we have

$$\log M_f^*(n, \epsilon) = nC - \sqrt{\frac{nV}{2}} Q^{-1}(2\epsilon) + \mathcal{O}(n^{1/3} \log n). \quad (18)$$

*Remark 1:* The assumption  $\epsilon \in (0, 1/2)$  is crucial for (18) to hold. Indeed, when  $\epsilon > 1/2$ , one can achieve  $\log M_f^*(n, \epsilon) = n\tilde{C} + o(n)$  by using a standard point-to-point fixed blocklength code of rate  $\tilde{C}$  and input distribution  $P_1^*$  (the CAID of  $W_1$ ) with probability  $1/2$  and another code of rate  $\tilde{C}$  but input distribution  $P_2^*$  (the CAID of  $W_2$ ) with probability  $1/2$ . This implies that the strong converse does not hold.

*Remark 2:* If the error probability constraint (6) is replaced with  $\mathbb{P}[(g_1(Y_1^n), g_2(Y_2^n)) \neq (J, J)] \leq \epsilon$ , then the asymptotic expansion in (18) continues to hold except that  $Q^{-1}(2\epsilon)$  is replaced with  $Q^{-1}(\epsilon)$ .

##### A. Proof of Theorem 3: Converse

Let

$$A_k(j, y_1^n, y_2^n) \triangleq \sum_{i=1}^n v_{P^*, W_k}(f_i(j, y_1^{i-1}, y_2^{i-1}); y_{k,i}) \quad (19)$$

and

$$A(j, y_1^n, y_2^n) \triangleq \frac{1}{2} (A_1(j, y_1^n, y_2^n) + A_2(j, y_1^n, y_2^n)). \quad (20)$$

By Theorem 1, every  $(n, M, \epsilon)$ -feedback code satisfies

$$\epsilon + e^{-\eta} \geq \max_k \mathbb{P}[A_k(j, y_1^n, y_2^n) \leq \log M - \eta] \quad (21)$$

$$\geq \frac{1}{2} \sum_k \mathbb{P}[A_k(j, y_1^n, y_2^n) \leq \log M - \eta] \quad (22)$$

$$\geq \frac{1}{2} \mathbb{P} \left[ \bigcup_k \{A_k(j, y_1^n, y_2^n) \leq \log M - \eta\} \right] \quad (23)$$

$$\geq \frac{1}{2} \mathbb{P}[A(J, Y_1^n, Y_2^n) \leq \log M - \eta]. \quad (24)$$

We observe now that, by symmetry, the probability distribution of  $v_{P^*, W_1}(x; Y_1) + v_{P^*, W_2}(x; Y_2)$  does not depend on  $x \in \{0, 1\}$ . This implies that the probability distribution of  $A(j, Y_1^n, Y_2^n)$  does not depend on  $j$  and coincides with the one of a sum of bounded i.i.d. random variables (RVs) with mean  $C$  and variance  $V/2$  (see (10)). Thus, by the Berry-Esseen central limit theorem [6, Th. V.4], we conclude that

$$\begin{aligned} \mathbb{P}[A(J, Y_1^n, Y_2^n) \leq \log M - \eta] \\ \geq Q\left(\frac{nC - \log M + \eta}{\sqrt{nV/2}}\right) - \frac{c}{\sqrt{n}}. \end{aligned} \quad (25)$$

Here,  $c$  denotes a nonnegative constant, which depends only on the channels  $\{W_k\}$ . We obtain the desired result by substituting (25) in (24), by solving for  $\log M$ , by setting  $\eta = \log n$ , and by performing a Taylor expansion of  $Q^{-1}(\cdot)$  around  $2\epsilon$ .

##### B. Proof of Theorem 3: Achievability

To establish the achievability part, we start by setting in Theorem 2 the parameter  $S$  (which controls the number of subcodewords available per block) to 5. Let  $\kappa \triangleq \lfloor C^{-1} \log 5 \rfloor + 1$  and set the number of blocks and the number of channel uses per block in Theorem 2 to

$$L = L_n \triangleq \lfloor n^{1/3} \rfloor \quad \text{and} \quad m = m_n \triangleq \lfloor n/L_n \rfloor - \kappa. \quad (26)$$

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

$$\epsilon^*(\kappa L_n, 5^{L_n}) \leq \exp(-\omega L_n) \quad (27)$$

for some positive constant  $\omega$ .

Next, we specify the types  $\{P_b\}_{b=1}^5$ . Let  $\rho > 0$  (a constant we shall specify later) and define the following 5 sequences:

$$p_{1,n} \triangleq \frac{1}{2} + \frac{\rho n^{-1/3}}{I_1'}, \quad p_{2,n} \triangleq 1 - p_{1,n} = \frac{1}{2} + \frac{\rho n^{-1/3}}{I_2'} \quad (28)$$

$p_{3,n} \triangleq 1/2$ ,  $p_{4,n} \triangleq p_1^*$ , and  $p_{5,n} \triangleq p_2^*$ . Here,  $p_1^*$  and  $p_2^*$  parameterize the CAIDs of  $W_1$  and  $W_2$ , respectively. Note that  $p_{1,n}$  and  $p_{2,n}$  are in the interval  $(0, 1)$  for sufficiently large  $n$ . We now let  $\bar{p}_{b,n}$  ( $b \in \{1, 3, 4, 5\}$ ) be the element of the set  $\{i/m_n : i \in \{0, \dots, m_n\}\}$  that minimizes  $|\bar{p}_{b,n} - p_{b,n}|$ . Furthermore, we set  $\bar{p}_{b,2} = 1 - \bar{p}_{b,1}$ . We then choose the types in Theorem 2 as  $P_b = \bar{P}_{b,n} \triangleq [\bar{p}_{b,n}, 1 - \bar{p}_{b,n}]$ .

Some observations are in order. Let  $P_{b,n} = [p_{b,n}, 1 - p_{b,n}]$ ; then  $\|\bar{P}_{b,n} - P_{b,n}\|_1 = \mathcal{O}(1/n)$ . Hence, since  $I_k(\cdot)$  is differentiable,

$$I_k(\bar{P}_{b,n}) = I_k(P_{b,n}) + \mathcal{O}(1/n), \quad b \in \{1, \dots, 5\}. \quad (29)$$

In Theorem 2, we also set  $\tau = 1/\sqrt{n}$ ,  $\zeta = n^{1/3}$ , and take  $\mathbf{x}$  as an arbitrary element of  $\mathbb{F}$  in (14) satisfying

$$\mathbf{x}_{\ell,i}(1) = 1 - \mathbf{x}_{\ell,i}(2) \quad (30)$$

for all  $\ell \in \{1, \dots, L_n\}$  and  $i \in \{1, \dots, m_n\}$ . This choice is possible because  $\bar{p}_{1,n} = 1 - \bar{p}_{2,n}$ .

The mappings  $\{h_\ell\}_{\ell=0}^{L_n-1}$ , which use the feedback to determine which subcodeword to transmit in each block, are chosen as follows: in the first  $L_n - 1$  blocks, we let the transmitter choose between subcodewords of type  $\bar{P}_{1,n}$  and of type  $\bar{P}_{2,n}$  depending on which decoder has the largest accumulated information density. This balances the information densities at the two decoders and makes the difference tightly concentrated around zero (as we shall see later). In the last block, the transmitter chooses a subcodeword of type  $\bar{P}_{3,n}$  (which approximates the CAID  $P^*$ ) if the arithmetic average of the information densities is above a suitably chosen threshold  $\gamma_1$ , specified later in (45). Otherwise it chooses randomly between the subcodeword of type  $\bar{P}_{4,n}$  (which approximates the CAID of  $W_1$ ) and the one of type  $\bar{P}_{5,n}$  (which approximates the CAID of  $W_2$ ).

Specifically, we choose the mappings  $\{h_\ell\}_{\ell=0}^{L_n-1}$  as follows:  $h_0 = 2$  and for  $\ell = 1, \dots, L_n - 2$ ,

$$h_\ell(\mathfrak{z}^\ell(B^\ell), \mathfrak{Y}_1^\ell(B^\ell), \mathfrak{Y}_2^\ell(B^\ell)) = 1 + \mathbf{1}\left\{v_1(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_1^\ell(B^\ell)) \geq v_2(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_2^\ell(B^\ell))\right\}. \quad (31)$$

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

$$h_\ell(\mathfrak{z}^\ell(B^\ell), \mathfrak{Y}_1^\ell(B^\ell), \mathfrak{Y}_2^\ell(B^\ell)) \triangleq \begin{cases} 3, & \text{if } v_1(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_1^\ell(B^\ell)) + v_2(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_2^\ell(B^\ell)) \geq 2\gamma_1 \\ 4, & \text{if } v_1(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_1^\ell(B^\ell)) + v_2(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_2^\ell(B^\ell)) < 2\gamma_1 \\ & \text{and } T = 1 \\ 5, & \text{if } v_1(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_1^\ell(B^\ell)) + v_2(\mathfrak{z}^\ell(B^\ell); \mathfrak{Y}_2^\ell(B^\ell)) < 2\gamma_1 \\ & \text{and } T = 2. \end{cases} \quad (32)$$

Roughly speaking, the threshold  $\gamma_1$  is chosen so that  $\bar{P}_{3,n}$ ,  $\bar{P}_{4,n}$ , and  $\bar{P}_{5,n}$  are used with probability  $(1 - 2\epsilon)$ ,  $\epsilon$ , and  $\epsilon$ , respectively.

Using Theorem 2 with the parameters listed above, we obtain

$$\begin{aligned} \log M_f^*(n, \epsilon) &\geq \log M_f^*(L_n m_n + \kappa L_n, \epsilon) \\ &\geq \sup\left\{\gamma : \min_k \mathbb{P}\left[v_k(\mathfrak{z}(B^{L_n}); \mathfrak{Y}_k(B^{L_n})) < \gamma\right] \leq \right. \\ &\quad \left. \epsilon - c_1 n^{-1/2}\right\} + \mathcal{O}(n^{1/3} \log n) \end{aligned} \quad (33)$$

for some  $c_1 > 0$ . In the last step, we used that  $(\epsilon^*(\kappa L_n, S^{L_n}) + 1/\sqrt{n} + e^{-n^{1/3}}) \leq (e^{-\omega L_n} + 1/\sqrt{n} + e^{-n^{1/3}}) \leq c_1 n^{-1/2}$  for sufficiently large  $n$ . Now set

$$\begin{aligned} \gamma &\triangleq L_n m_n C - \sqrt{(L_n - 1)m_n V/2} \\ &\quad \times Q^{-1}\left(2\epsilon - 2\sqrt{2}(c_1 + c_2)n^{-1/2}\right). \end{aligned} \quad (34)$$

The constant  $c_2 > 0$  will be defined shortly. The final step of the proof is to show that the following upper bound holds for sufficiently large  $n$

$$\mathbb{P}\left[v_k(\mathfrak{z}(B^{L_n}); \mathfrak{Y}_k(B^{L_n})) < \gamma\right] \leq \epsilon - c_1 n^{-1/2}. \quad (35)$$

The desired result (18) then follows by using (34) and (35) in (33), 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 (34) around  $2\epsilon$ .

*Proof of (35):* To simplify the notation, we set

$$Z_{k,\ell,i}^{(b)} \triangleq v_{\bar{P}_{b,n}, W_k}(\mathfrak{z}_{\ell,i}(b); \mathfrak{Y}_{k,\ell,i}(b)) \quad (36)$$

$$\bar{Z}_{k,\ell_1,\ell_2} \triangleq \sum_{\ell=\ell_1}^{\ell_2} \sum_{i=1}^{m_n} Z_{k,\ell,i}^{(B_\ell)}. \quad (37)$$

The RVs  $Z_{k,\ell,i}^{(b)}$  are independent for all  $k, b, \ell$ , and  $i$ . Note also that  $\bar{Z}_{k,1,L_n} = v_k(\mathfrak{z}(B^{L_n}); \mathfrak{Y}_k(B^{L_n}))$ . Hence, the left-hand side of (35) can be rewritten as  $\mathbb{P}[\bar{Z}_{k,1,L_n} < \gamma]$ .

It will turn out convenient to define the following quantities as well:

$$\bar{Z}^* \triangleq \frac{1}{2} \sum_{\ell=1}^{L_n-1} \sum_{i=1}^{m_n} \sum_{k \in \{1,2\}} v_{P^*, W_k}(\mathfrak{z}_{\ell,i}(B_\ell); \mathfrak{Y}_{k,\ell,i}(B_\ell)) \quad (38)$$

$$A \triangleq \frac{1}{2} \sum_{\ell=1}^{L_n-1} \sum_{i=1}^{m_n} \sum_{k \in \{1,2\}} \log \frac{\bar{P}_{B_\ell, n} W_k(\mathfrak{Y}_{k,\ell,i}(B_\ell))}{P_{Y_k}^*(\mathfrak{Y}_{k,\ell,i}(B_\ell))} \quad (39)$$

$$E \triangleq \frac{1}{2} (\bar{Z}_{1,1,L_n-1} - \bar{Z}_{2,1,L_n-1}). \quad (40)$$

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

$$\bar{Z}_{k,1,L_n-1} = \bar{Z}^* - A - (-1)^k E. \quad (41)$$

We next show that  $\bar{Z}^*$  is accurately approximated by a normal distribution with mean  $(L_n - 1)m_n C$  and variance  $(L_n - 1)m_n V/2$ , whereas both  $A$  and  $E$  are of order  $n^{1/3}$ . To prove these claims, we rely on the Berry-Esseen central limit theorem, on Hoeffding inequality [7], and on the following stabilization lemma, whose proof (omitted) is by mathematical induction.

*Lemma 4 (stabilization lemma):* Let  $\{X_\ell^{(1)}\}$  and  $\{X_\ell^{(2)}\}$  be i.i.d. RVs with mean  $\mu_1 > 0 > \mu_2$  satisfying

$$\mathbb{P}\left[X_\ell^{(b)} \geq v\right] \leq e^{-\beta|v-\mu_b|_+^2}, \mathbb{P}\left[X_\ell^{(b)} \leq v\right] \leq e^{-\beta|\mu_b-v|_+^2} \quad (42)$$

for all  $b \in \{1, 2\}$ , all  $\ell \in \mathbb{N}$ , and some  $\beta > 0$ . Here,  $|\cdot|_+ \triangleq \max\{0, \cdot\}$ . Define the sequence  $\{Y_\ell\}$  as follows:  $Y_0 = 0$  and

$$Y_\ell = \begin{cases} Y_{\ell-1} + X_\ell^{(1)} & \text{for } Y_{\ell-1} < 0 \\ Y_{\ell-1} + X_\ell^{(2)} & \text{for } Y_{\ell-1} \geq 0. \end{cases} \quad (43)$$

Let  $c \geq 1$  satisfy  $\min\{\mu_1, -\mu_2\} \geq \sqrt{\pi/\beta} \exp(c^2/4)$ . Then

$$\mathbb{P}[|Y_\ell| \geq v] \leq 2e^{-c\sqrt{\beta}(v-\mu_1+\mu_2)}. \quad (44)$$

Let now  $c_4 > 0$  be arbitrary and define the thresholds

$$\gamma_1 \triangleq \gamma - m_n C + n^{1/3} \log n \quad (45)$$

$$\gamma_2 \triangleq (L_n - 1)m_n C - c_4 \sqrt{n} \log n. \quad (46)$$

Here,  $\gamma$  was defined in (34). Intuitively,  $\gamma_1$  is chosen so as to be the  $2\epsilon$ -quantile of  $\bar{Z}^* - A$  for large  $n$ , whereas  $\gamma_2$  is needed to upper-bound the probability of the (rare) event that  $\bar{Z}^* - A$  is far smaller than its mean. Note also that  $\gamma_1 > \gamma_2$  for sufficiently large  $n$ . Define the four disjoint events

$$\mathcal{E}_{1,m} \triangleq \left\{\bar{Z}^* - A \in [\gamma_2, \gamma_1], T = m\right\}, \quad m \in \{1, 2\} \quad (47)$$

$$\mathcal{E}_2 \triangleq \left\{\bar{Z}^* - A \geq \gamma_1\right\}, \quad \mathcal{E}_3 \triangleq \left\{\bar{Z}^* - A < \gamma_2\right\}. \quad (48)$$

Let  $\bar{k} = 1$  when  $k = 2$  and  $\bar{k} = 2$  when  $k = 1$ . Using the events defined above, we readily obtain the following upper bound

$$\begin{aligned} \mathbb{P}[\bar{Z}_{k,1,L_n} < \gamma] &\leq \mathbb{P}[\mathcal{E}_{1,\bar{k}}] + \mathbb{P}[\mathcal{E}_{1,k} \cap \{\bar{Z}_{k,1,L_n} < \gamma\}] \\ &\quad + \mathbb{P}[\mathcal{E}_2 \cap \{\bar{Z}_{k,1,L_n} < \gamma\}] + \mathbb{P}[\mathcal{E}_3]. \end{aligned} \quad (49)$$

In the following, we upper-bound each of these four probabilities. It turns out that only  $\mathbb{P}[\mathcal{E}_{1,\bar{k}}]$  yields a nonvanishing contribution.

**Bound on  $\mathbb{P}[\mathcal{E}_{1,\bar{k}}]$ :** We first establish a bound on the tail probability of  $A$  in (47). Since  $\bar{P}_{1,n} = 1 - \bar{P}_{2,n}$ , we can find a sequence  $\{\zeta_n\}$  such that

$$I_1(\bar{P}_{b,n}) + I_2(\bar{P}_{b,n}) = 2(C - \zeta_n n^{-2/3}), \quad b \in \{1, 2\}. \quad (50)$$

It follows from the concavity of mutual information, from (28)–(29), from Taylor’s theorem, and from (9) that the sequence  $\zeta_n$  is positive and convergent. Since  $D(W_1(\cdot|x)||P_{Y_1}^*) + D(W_2(\cdot|x)||P_{Y_2}^*) = 2C$  for every  $x \in \{0, 1\}$ , we must also have that

$$D(\bar{P}_{b,n} W_1 || P_{Y_1}^*) + D(\bar{P}_{b,n} W_2 || P_{Y_2}^*) = 2\zeta_n n^{-2/3} \quad (51)$$

for all  $b \in \{1, 2\}$ . Since  $\|\bar{P}_{b,n} - P^*\|_1 = \mathcal{O}(n^{-1/3})$ , there exists a constant  $c_3 > 0$  such that  $|\sum_k \log \bar{P}_{b,n} W_k(y) / P_{Y_k}^*(y)| \leq c_3 n^{-1/3}$  for sufficiently large  $n$ . Hence, for sufficiently large  $n$ ,

$$\mathbb{P}[A \geq n^{1/3} \log n] \leq \exp(-cn^{1/3} \log^2 n) = \mathcal{O}(1/n). \quad (52)$$

Here, we have used (30), that  $\bar{P}_{2,n} = 1 - \bar{P}_{1,n}$ , Hoeffding inequality, and that  $\zeta_n$  is convergent.

We next upper-bound  $\mathbb{P}[\mathcal{E}_{1,\bar{k}}]$  as follows:

$$\mathbb{P}[\mathcal{E}_{1,\bar{k}}] \leq \frac{1}{2} \mathbb{P}[\bar{Z}^* < \gamma_1 + n^{1/3} \log n] + \mathcal{O}(1/n) \quad (53)$$

$$\leq \frac{1}{2} Q\left(\frac{(L_n - 1)m_n C - \gamma_1 - n^{1/3} \log n}{\sqrt{V(L_n - 1)m_n}/2}\right) + \frac{\sqrt{2}c_2}{\sqrt{n}} \quad (54)$$

$$\leq \epsilon - \sqrt{2}c_1/\sqrt{n}. \quad (55)$$

Here, (53) follows from (47) and (52), and because  $T$  is equiprobable on  $\{1, 2\}$  and independent of  $(Y_1, Y_2)$ ; (54), which holds for some constant  $c_2 > 0$  and sufficiently large  $n$ , follows from the Berry-Esseen central limit theorem [6, Th. V.4]; finally, (55) follows from (34) and (45).

**Bound on  $\mathbb{P}[\mathcal{E}_{1,k} \cap \{\bar{Z}_{k,1,L_n} < \gamma\}]$ :** We show that  $E$  in (41) is sufficiently concentrated around 0. To do this, let

$$F_\ell^{(b)} \triangleq \frac{1}{2} \sum_{i=1}^{m_n} (Z_{1,\ell,i}^{(b)} - Z_{2,\ell,i}^{(b)}) \quad (56)$$

for  $b \in \{1, 2\}$ . The  $F_\ell^{(b)}$  are independent RVs. Furthermore, observe that  $E = \sum_{\ell=1}^{L_n-1} F_\ell^{(B_\ell)}$ . Since  $\bar{P}_{1,n} = 1 - \bar{P}_{2,n}$ , there must exist a convergent sequence  $\{\xi_n\}$  such that

$$I_1(\bar{P}_{b,n}) - I_2(\bar{P}_{b,n}) = -2(-1)^b(\rho n^{-1/3} + \xi_n n^{-2/3}). \quad (57)$$

Consequently, since  $\{\mathcal{X}_{\ell,i}(b)\}_{i=1}^{m_n}$  is of constant composition, it follows that  $\mathbb{E}[F_\ell^{(b)}] = -(-1)^b \mu_n$ , where  $\mu_n \triangleq \rho m_n n^{-1/3} + \xi_n m_n n^{-2/3} = \mathcal{O}(n^{1/3})$  (recall that  $\rho > 0$  is a constant that we have not chosen yet). Since  $F_\ell^{(b)}$  can be written as a sum

of independent random variables, we conclude, by applying Hoeffding inequality, that

$$\mathbb{P}[F_\ell^{(b)} \geq v] \leq e^{-\frac{1}{2m_n c_5^2} |v + \mu_n (-1)^b|^2} \quad (58)$$

and

$$\mathbb{P}[F_\ell^{(b)} \leq v] \leq e^{-\frac{1}{2m_n c_5^2} |\mu_n (-1)^b - v|^2}. \quad (59)$$

Here,  $c_5 \geq \max\{1, |Z_{1,\ell,i} - Z_{2,\ell,i}|\}$ . We now choose  $\rho$  so that  $\rho > \sqrt{4\pi c_5^2} e^{c_5^2}$ . Next, we apply Lemma 4 with  $X_\ell^{(b)} = F_\ell^{(b)}$ ,  $\beta = 1/(2m_n c_5^2)$  and  $c = 2c_5$  and obtain

$$\mathbb{P}[|E| \geq n^{1/3} \log n] \leq 2e^{-\sqrt{\frac{2}{m_n}} (n^{1/3} \log n - 2\mu_n)} \leq 3/n. \quad (60)$$

In the last step, which holds for sufficiently large  $n$ , we used that  $\xi_n m_n n^{-2/3} = \mathcal{O}(1)$ . Consider now

$$\begin{aligned} &\mathbb{P}[\mathcal{E}_{1,k} \cap \{\bar{Z}_{k,1,L_n} < \gamma\}] \\ &\leq \mathbb{P}[\mathcal{E}_{1,k} \cap \{|E| \leq n^{1/3} \log n, \bar{Z}_{k,1,L_n} < \gamma\}] + 3/n \end{aligned} \quad (61)$$

$$\begin{aligned} &\leq \mathbb{P}[\bar{Z}_{k,L_n,L_n} < \gamma - \gamma_2 + n^{1/3} \log n | B_{L_n} = 3 + k] + 3/n \\ &\quad (62) \end{aligned}$$

$$\begin{aligned} &\leq \exp\left(-\frac{c_6}{m_n} (m_n(C_k - C) - 2c_4 \sqrt{n} \log n)^2\right) + 3/n \quad (63) \\ &= \mathcal{O}(1/n) \quad (64) \end{aligned}$$

where  $c_6 > 0$  is some constant. Here, (61) follows from (60); (62) follows because if  $\mathcal{E}_{1,k}$  occurs then  $B_{L_n} = 3 + k$  (see (32)), because  $\bar{Z}^* - A \geq \gamma_2$ , and because  $\bar{Z}_{k,1,L_n} = \bar{Z}^* - A - (-1)^k E + \bar{Z}_{k,L_n,L_n}$ . Finally, (63) follows from (34), (46), and Hoeffding inequality applied to  $\bar{Z}_{k,L_n,L_n}$ .

**Bound on  $\mathbb{P}[\mathcal{E}_2 \cap \{\bar{Z}_{k,1,L_n} < \gamma\}]$  and on  $\mathbb{P}[\mathcal{E}_3]$ :** We show that  $\mathbb{P}[\mathcal{E}_2 \cap \{\bar{Z}_{k,1,L_n} < \gamma\}] = \mathcal{O}(1/n)$  using an argument similar to the one leading to (64) and that  $\mathbb{P}[\mathcal{E}_3] = \mathcal{O}(1/n)$  using (46), (52), and Hoeffding inequality applied to  $\bar{Z}^*$ . Substituting these last two results as well as (55) and (64) into (49), we establish the desired upper bound (35).

## V. ACKNOWLEDGEMENTS

The work of K. F. Trillingsgaard and P. Popovski was supported in part by the European Research Council (ERC Consolidator Grant Nr. 648382 WILLOW) within the Horizon 2020 Program. The work of G. Durisi was supported by the Swedish Research Council under the grant 2016-03293.

## REFERENCES

- [1] Y. Polyanskiy, “On dispersion of compound DMCs,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, 2013, pp. 26–32.
- [2] 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.
- [3] —, “Feedback in the non-asymptotic regime,” *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.
- [4] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, “Variable-length coding with stop-feedback for the common-message broadcast channel,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, 2016.
- [5] 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.
- [6] V. V. Petrov, *Sums of Independent Random Variables*. Berlin: Springer, 1975, translated from the Russian by A. A. Brown.
- [7] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.