CHALMERS

Chalmers Publication Library

Broadcasting a Common Message with Variable-Length Stop-Feedback Codes

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:

2015 IEEE International Symposium on Information Theory

Citation for the published paper:
Trillingsgaard, K. ; Yang, W. ; Durisi, G. et al. (2015) "Broadcasting a Common Message with Variable-Length Stop-Feedback Codes". 2015 IEEE International Symposium on Information Theory

Downloaded from: http://publications.lib.chalmers.se/publication/215766

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)

Broadcasting a Common Message with Variable-Length Stop-Feedback Codes

Kasper Fløe Trillingsgaard*, Wei Yang†, Giuseppe Durisi†, and Petar Popovski*
* Aalborg University, 9220 Aalborg, Denmark
† Chalmers University of Technology, 41296 Gothenburg, Sweden

***Abstract*—We investigate the maximum coding rate achievable over a two-user 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 both stop signals. For the point-to-point case, Polyanskiy, Poor, and Verdú (2011) recently demonstrated that variable-length coding combined with stop feedback significantly increases the speed at which the maximum coding rate converges to capacity. This speed-up manifests itself in the absence of a square-root penalty in the asymptotic expansion of the maximum coding rate for large blocklengths, a result a.k.a. *zero dispersion*. In this paper, we show that this speed-up does not necessarily occur for the broadcast channel with common message. Specifically, there exist scenarios for which variable-length stop-feedback codes yield a positive dispersion.**

I. Introduction

We consider the setup where an encoder wishes to convey a common message over a broadcast channel with noiseless feedback to two decoders. Similarly to the single-decoder (SD) case, noiseless feedback combined with fixed-blocklength codes does not improve capacity, which is given by [1, p. 126]

$$C = \sup_{P} \min\{I(P, W_1), I(P, W_2)\}\,. \tag{1}$$

Here, W_1 and W_2 denote the channels to decoder 1 and 2, respectively, and the supremum is over all input distributions P. For the case when there is no feedback, the speed at which C is approached as the blocklength n increases is of the order $1/\sqrt{n}$ [2] (same as in the SD case). The constant factor associated to the $1/\sqrt{n}$ term is commonly referred to as channel *dispersion*.

For the SD case, noiseless feedback combined with variable-length codes improve significantly the speed of convergence to capacity. Specifically, it was shown in [3] that

$$\frac{1}{l}\log \widetilde{M}_{\mathrm{f}}^{*}(l,\epsilon) = \frac{\widetilde{C}}{1-\epsilon} - \mathcal{O}\left(\frac{\log l}{l}\right) \tag{2}$$

where l stands for the average blocklength (average transmission time), $\widetilde{M}_{\mathrm{f}}^{*}(l,\epsilon)$ is the maximum number of codewords in the SD case, and $\widetilde{C}$ denotes the corresponding capacity. One sees from (2) that no square-root penalty occurs (zero dispersion), which implies a fast convergence to the asymptotic limit. This fast convergence is demonstrated numerically in [3] by means of nonasymptotic bounds. Variable-length stop-feedback (VLSF) codes, i.e., coding schemes where the feedback is used only to stop transmissions, are sufficient to achieve (2).

The purpose of this paper is to investigate whether a similar result holds for the broadcast channel with common message.

Contribution: We consider the subclass of discrete memoryless broadcast channels for which $I(P, W_1)$ and $I(P, W_2)$ are maximized by the same input distribution P^*, which we assume to be unique. In this case, $C = \min\{I(P^*, W_1), I(P^*, W_2)\}$. Focusing on the case when VLSF codes are used, we obtain nonasymptotic achievability and converse bounds on the maximum number of codewords $M_{\mathrm{sf}}^{*}(l,\epsilon)$ with average blocklength l that can be transmitted 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 two subchannels are independent and have the same capacity and the same dispersion, and when $\epsilon \leq 0.1968$, the asymptotic expansion of $M_{\mathrm{sf}}^{*}(l,\epsilon)$ contains a square-root penalty (see (18) and (22) for a precise statement of this result). Hence, the fast convergence to the asymptotic limit experienced in the SD case cannot be expected.

The intuition behind this result is as follows: in the SD case, the stochastic variations of the information density that result in the square-root penalty can be virtually eliminated by using variable-length coding with stop-feedback. Indeed, decoding is stopped after the information density exceeds a certain threshold, which yields only negligible stochastic variations. In the broadcast setup, however, the stochastic variations in the difference between the stopping times at the two decoders make the square-root penalty reappear. Note that our result does not necessarily imply that feedback is useless. It only shows that VLSF codes cannot be used to speed-up convergence to the same level as in the SD case.

Proof techniques: The achievability bound is an extension of [3, Th 3]; the converse bound is based on an optimal stopping problem, where the probability that the stopping time exceeds a given threshold is minimized under a constraint on the "stopped" information density process. The asymptotic analysis of the converse bound relies on Hoeffding's inequality and on the Berry-Esseen central limit theorem, whereas the asymptotic analysis of the achievability bound relies on asymptotic results for random walks [4] and on a Berry-Esseen-type theorem that holds for random summations [5].

Notation: Upper case, lower case, and calligraphic letters denote random variables (RV), deterministic quantities, and sets, respectively. The probability density function of a standard Gaussian RV is denoted by $\phi(x)$. Furthermore, $\Phi(x) \triangleq 1 - Q(x)$ is its cumulative distribution, where $Q(x)$ is the Q-function. We let x^+ and x^- denote $\max(0,x)$ and $\min\{0,x\}$, respectively. Throughout the paper, the index k belongs always to the set $\{1,2\}$, although this is sometimes omitted. Furthermore, $\bar{k} \triangleq 3-k$. We adopt the convention that $\sum_{i=j}^{j-1} a_i = 0$ for all $\{a_i\}$ and all integers j. We use "$\mathbb{c}$" to denote a finite nonnegative constant. Its value may change at each occurrence. Finally, $\mathbb{N}$ denotes the set of positive integers and $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$.

II. System Model

A common-message discrete memoryless broadcast channel with two decoders is defined by the finite input alphabet $\mathcal{X}$ and the finite output alphabets $\mathcal{Y}_k$, along with the 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 $x \in \mathcal{X}$. We assume that the outputs at each time i are conditionally independent given the input, i.e.,

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

Define the set of probability distributions on $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. Let $P \times W_k : (x, y_k) \to P(x)W(y_k|x)$ denote the joint distribution of input and output at decoder k, and let $PW_k : y_k \to \sum_{x\in\mathcal{X}} P(x)W_k(y_k|x)$ denote the marginal distribution on $\mathcal{Y}_k$. For every $P \in \mathcal{P}(\mathcal{X})$, the information density is defined as

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

We let $I(P, W_k) \triangleq \mathbb{E}_{P\times W_k}[\imath_{P,W_k}(X; Y_k)]$ be the mutual information, $V(P, W_k) \triangleq \mathrm{Var}_{P\times W_k}[\imath_{P,W_k}(X; Y_k)]$ be the (unconditional) information variance, and $T(P, W_k) \triangleq \mathbb{E}_{P\times W_k}\left[|\imath_{P,W_k}(X; Y_k) - I(P, W_k)|^3\right]$ be the third absolute moment of the information density. We restrict ourselves to the case, where there exists a unique probability distribution $P^* \in \mathcal{P}(\mathcal{X})$ that maximizes simultaneously both $I(P, W_1)$ and $I(P, W_2)$. In this case, the capacity is given by

$$C \triangleq \min\{C_1, C_2\} \quad (5)$$

where $C_k \triangleq I(P^*, W_k)$. The corresponding (unique) capacity-achieving output distributions are denoted by $P^*_{Y_k}$. Finally, we also define the dispersions $V_k \triangleq V(P^*, W_k)$.

We are now ready to formally define a VLSF code for the broadcast channel with common message.

Definition 1: An (l, M, ϵ)-VLSF code for the broadcast channel with common message consists of:

1) A RV $U \in \mathcal{U}$, with $|\mathcal{U}| \le 3$, which is known by the encoder and by both decoders.
2) A sequence of encoders $f_n : \mathcal{U} \times \mathcal{M} \to \mathcal{X}$, each one mapping the message $J \in \mathcal{M} = \{1, \dots, M\}$, drawn uniformly at random, to the channel input according to $X_n = f_n(U, J)$.
3) Two nonnegative integer-valued RVs τ_1 and τ_2 that are stopping times with respect to the filtrations $\mathcal{F}(U, Y_1^n)$ and $\mathcal{F}(U, Y_2^n)$, respectively, and which satisfy
$$\mathbb{E}[\max\{\tau_1, \tau_2\}] \le l. \quad (6)$$
4) A sequence of decoders $g_{k,n} : \mathcal{U} \times \mathcal{Y}_i^n \to \mathcal{M}$ satisfying
$$\Pr[J \ne g_{k,\tau_k}(U, Y_k^{\tau_k})] \le \epsilon, \qquad k \in \{1, 2\}. \quad (7)$$

Remark 1: The RV U serves as common randomness, and enables the use of randomized codes [6]. To establish the cardinality bound on U, we proceed as in [3, Th. 19] to show that $|\mathcal{U}| \le 4$ is sufficient. This bound can be further improved to $|\mathcal{U}| \le 3$ by using the Fenchel-Eggleston theorem [7, p. 35].

Remark 2: VLSF codes require a feedback link from the decoders 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 both decoders have fed back a stop signal. Hence, the blocklength is $\max\{\tau_1, \tau_2\}$.

Our aim is to characterize the largest number of codewords $M^*_{\mathrm{sf}}(l, \epsilon)$, whose average length is l, that can be transmitted with reliability $1 - \epsilon$ using a VLSF code.

III. Main Results

A. Achievability bound

We first present an achievability bound. Its proof (omitted) follows closely the proof of [3, Th. 3].

Theorem 1: Fix $P \in \mathcal{P}(\mathcal{X})$. Let $\gamma_1, \gamma_2 \ge 0$ and $0 \le q \le 1$ be arbitrary scalars. Let the stopping times τ_k and $\bar{\tau}_k$, $k \in \{1, 2\}$, be defined as

$$\tau_k \triangleq \inf\{n \ge 0 : \imath_{P,W_k}(X^n; Y_k^n) \ge \gamma_k\} \quad (8)$$
$$\bar{\tau}_k \triangleq \inf\{n \ge 0 : \imath_{P,W_k}(\bar{X}^n; Y_k^n) \ge \gamma_k\} \quad (9)$$

where $(X^n, \bar{X}^n, Y_1^n, Y_2^n)$ are jointly distributed according to

$$P_{X^n,\bar{X}^n,Y_1^n,Y_2^n}(x^n, \bar{x}^n, y_1^n, y_2^n) = P_{Y_1^n,Y_2^n|X^n}(y_1^n, y_2^n|x^n) \prod_{i=1}^{n} P(x_i)P(\bar{x}_i). \quad (10)$$

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

$$l \le (1 - q)\mathbb{E}[\max\{\tau_1, \tau_2\}] \quad (11)$$

and

$$\epsilon \le q + (1 - q)(M - 1)\Pr[\tau_k \ge \bar{\tau}_k]\,. \quad (12)$$

Remark 3: Following the same steps as in [3, Eq. (111)–(118)], ϵ in (12) can be further upper-bounded as

$$\epsilon \le q + (1 - q)(M - 1)\exp\{-\gamma_k\}\,. \quad (13)$$

This bound is easier to evaluate and to analyze asymptotically.

B. Converse bound

Let $P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X})$ be the type [8, Def. 2.1] of the sequence $\mathbf{x}^n \in \mathcal{X}^n$. We are now ready to state our converse bound.

Theorem 2: For every M, $t \in \mathbb{Z}_+$ and $\delta > 0$, let

$$\lambda_t \triangleq \log M - \log\log M - \delta - (|\mathcal{X}| - 1)\log(t + 1) \quad (14)$$

and let

$$L_t \triangleq \prod_{k=1}^{2} \max_{\mathbf{x}^t\in\mathcal{X}^t} \left\{\Pr\left[\imath_{P_{\mathbf{x}^t},W_k}(\mathbf{x}^t; Y_k^t) > \lambda_t\right]\right\} + \varepsilon_M \left(1 + \min_k \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_{P_{\mathbf{x}^t},W_k}(\mathbf{x}^t; Y_k^t) > \lambda_t\right]\right) \quad (15)$$

where $\varepsilon_M = \epsilon + (\log M)^{-1}$. Then, for every (l, M, ϵ)-VLSF code, we have

$$l \geq \sum_{t=0}^{\infty} (1 - L_t)^+ . \quad (16)$$

Proof: See Section IV. ■

C. Asymptotic expansion

Analyzing (13) and (16) in the limit $l \to \infty$, we obtain the following asymptotic characterization of $M^*_{\mathrm{sf}}(l, \epsilon)$.

Theorem 3: Let $Z_k \sim \mathcal{N}(0, 1)$, $V = \sqrt{V_1 V_2}$, $\varrho_k = (V_k/V_{\bar{k}})^{1/4}$, and let $y = \tilde{Q}^{-1}(x)$ be the solution of

$$\prod_{k=1}^{2} Q(-\varrho_k y) + x \left(1 + \min_k Q(-\varrho_k y)\right) = 1. \quad (17)$$

For every discrete memoryless broadcast channel with $C_1 = C_2$ and every $\epsilon \in (0, 1)$, we have

$$\frac{Cl}{1-\epsilon} - \Xi_{\mathrm{a}}\sqrt{l} - \mathcal{O}\left(l^{1/4+\delta}\right) \leq \log M^*_{\mathrm{sf}}(l, \epsilon) \leq \frac{Cl}{1-\epsilon} - \Xi_{\mathrm{c}}\sqrt{l} + \mathcal{O}(\log l) \quad (18)$$

where $\delta > 0$ is an arbitrarily small constant,

$$\Xi_{\mathrm{a}} \triangleq \sqrt{\frac{V_1 + V_2}{2\pi(1-\epsilon)}} \quad (19)$$

and

$$\Xi_{\mathrm{c}} \triangleq \sqrt{\frac{V}{(1-\epsilon)^3}} \left(\mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon), \max_k \varrho_k Z_k\right\}\right] - \epsilon\left(2\tilde{Q}^{-1}(\epsilon) - \min_k \mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon), \varrho_k Z_k\right\}\right]\right)\right). \quad (20)$$

Proof: The converse bound in (18) is proved in Section V and the achievability bound is proved in Section VI. ■

Remark 4: When $C_1 \neq C_2$, it can be shown that the square-root penalty on the LHS of (18) vanishes. In this case, the problem reduces to the point-to-point transmission to the weakest decoder, for which the zero-dispersion result in [3] applies.

Remark 5: For the case when $P_{Y_{1,i},Y_{2,i}|X_i}$ does not satisfy (3), a bound similar to the LHS of (18) can be obtained by replacing Ξ_{a} in (19) with

$$\sqrt{\frac{V_1 + V_2 - 2\mathrm{Cov}(\imath_{P^*,W_1}(X;Y_1), \imath_{P^*,W_2}(X;Y_2))}{2\pi(1-\epsilon)}}. \quad (21)$$

Remark 6: When $\varrho_1 = \varrho_2 = 1$ (and, hence, $V_1 = V_2$), one can simplify the RHS of (18) as follows:

$$\log M^*_{\mathrm{sf}}(l, \epsilon) \leq \frac{Cl}{1-\epsilon} - \sqrt{\frac{Vl}{(1-\epsilon)^3}} \times \left(\frac{1}{\sqrt{\pi}}\left(1 - Q\left(\sqrt{2}Q^{-1}(\epsilon)\right)\right) + (\epsilon - 2)\phi(Q^{-1}(\epsilon))\right) - \mathcal{O}(\log l). \quad (22)$$

The second-order term in (22) is strictly negative for all $\epsilon \leq 0.1968$. This implies that, when $C_1 = C_2$, $V_1 = V_2$, and $\epsilon \leq 0.1968$, the asymptotic expansion of $\log M^*_{\mathrm{sf}}(l, \epsilon)$ contains a square-root penalty.

IV. Proof of Theorem 2

Fix M and ϵ. To establish Theorem 2, we derive a lower bound on l that holds for all VLSF codes having M codewords and probability of error no larger than ϵ. Since,

$$l \geq \mathbb{E}[\max\{\tau_1, \tau_2\}] = \sum_{t=0}^{\infty} (1 - \Pr[\max\{\tau_1, \tau_2\} \leq t]) \quad (23)$$

we can lower-bound l by upper-bounding $\Pr[\max\{\tau_1, \tau_2\} \leq t]$ for every $t \in \mathbb{Z}_+$. The following property turns out to be useful.

Property 1: Fix $t \in \mathbb{Z}_+$ and $\alpha \in [0, 1]$, and suppose there exists an (l, M, ϵ)-VLSF code with $\Pr[\max\{\tau_1, \tau_2\} \leq t] \leq \alpha$. Then there exists an (l', M, ϵ)-VLSF code for some $l' \geq l$, for which $\Pr[\max\{\tau_1, \tau_2\} \leq t] \leq \alpha$ and $\tau_1, \tau_2 \in \{t, t+1, \ldots\}$.

Fix an arbitrary (l, M, ϵ)-VLSF code, defined by the tuple $(f_n, g_{1,n}, g_{2,n}, \tau_1, \tau_2, U)$. By Property 1, it is sufficient to consider codes for which $\tau_1, \tau_2 \in \{t, t+1, \cdots\}$. Let $\epsilon_k^{(u)}$, $u \in \mathcal{U}$, be constants in $[0, 1]$ such that $\sum_{u\in\mathcal{U}} P_U(u)\epsilon_k^{(u)} \leq \epsilon$ and $\Pr[J \neq g_{k,\tau_k}(U, Y_k^{\tau_k})|U = u] \leq \epsilon_k^{(u)}$.

Since $\{\tau_k = n\} \in \mathcal{F}(U, Y_k^n)$, we can define a sequence of binary functions $\varphi_k \triangleq \{\varphi_{k,t}, \varphi_{k,t+1}, \cdots\}$ such that $\varphi_{k,n}(u, y_k^n) \triangleq \mathbb{1}\{\tau_k = n\}$. Let $P_{\mathbf{X}}^{(u)}$ be the conditional probability measure on $\mathcal{X}^\infty$ induced by the encoder given $U = u$. Define for $u \in \mathcal{U}$ the set $\bar{\mathcal{Y}}_k^{(u)} \triangleq \{y^n \in \mathcal{Y}_k^n : \varphi_{k,n}(u, y^n) = 1\}$. Note that we must have $Y_k^{\tau_k} \in \bar{\mathcal{Y}}^{(u)}$. Let the length of a sequence of channel outputs $\bar{y} \in \bar{\mathcal{Y}}_k^{(u)}$ be denoted by $|\bar{y}|$. On $\bar{\mathcal{Y}}_k^{(u)}$, define the conditional probability measure $\mathbb{P}_{\bar{Y}|\mathbf{X}}^{(k,u)}$, given $\mathbf{x} \in \mathcal{X}^\infty$ and $u \in \mathcal{U}$, as

$$\mathbb{P}_{\bar{Y}|\mathbf{X}}^{(k,u)}(\bar{y}|\mathbf{x}) \triangleq \prod_{i=1}^{|\bar{y}|} W(\bar{y}_i|\mathbf{x}_i) \quad (24)$$

and the probability measure $\mathbb{P}_{\bar{Y},\mathbf{X}}^{(k,u)}(\bar{y}, \mathbf{x}) \triangleq \mathbb{P}_{\bar{Y}|\mathbf{X}}^{(k,u)}(\bar{y}|\mathbf{x})P_{\mathbf{X}}^{(u)}(\mathbf{x})$ on $\bar{\mathcal{Y}}^{(u)} \times \mathcal{X}^\infty$. We also need the following auxiliary probability measure $\mathbb{Q}_{\bar{Y}}^{(k,u)}$ on $\bar{\mathcal{Y}}_k^{(u)}$

$$\mathbb{Q}_{\bar{Y}}^{(k,u)}(\bar{y}) \triangleq \sum_{P_{\mathbf{x}^t} \in \mathcal{P}_t(\mathcal{X})} \left(\frac{1}{|\mathcal{P}_t(\mathcal{X})|} \prod_{i=1}^{t} P_{\mathbf{x}^t} W_k(\bar{y}_i) \prod_{i=t+1}^{|\bar{y}|} P^*_{Y_k}(\bar{y}_i)\right) \quad (25)$$

and the probability measure $\mathbb{Q}_{\bar{Y},\mathbf{X}}^{(k,u)}(\bar{y}, \mathbf{x}) = \mathbb{Q}_{\bar{Y}}^{(k)}(\bar{y})\, P_{\mathbf{X}}^{(u)}(\mathbf{x})$ on $\bar{\mathcal{Y}}^{(u)} \times \mathcal{X}^\infty$. Here, $\mathcal{P}_t(\mathcal{X}) \subseteq \mathcal{P}(\mathcal{X})$ denotes the set of types formed by length-t sequences.

Using the meta-converse theorem [9, Th. 27], the inequality [9, Eq. (102)], the fact that $\mathbb{Q}_{\bar{Y}_k,\mathbf{X}}^{(k,u)}$ is a convex combination of distributions [10, Lem. 3], and the upper bound $|\mathcal{P}_t(\mathcal{X})| \leq (t+1)^{|\mathcal{X}|-1}$ [11, Lem. 1.1], we conclude that (see details in [12, App. I-B])

$$\mathbb{P}_{\bar{Y},\mathbf{X}}^{(k,u)}\left[\tilde{\imath}_k^{(u)}(\mathbf{X}; \bar{Y}_k) \leq \lambda_t\right] \leq \varepsilon_{k,M}^{(u)} \quad (26)$$

where $\varepsilon_{k,M}^{(u)} \triangleq \epsilon_k^{(u)} + (\log M)^{-1}$ and λ_t is defined in (14). Here,

$$\tilde{\imath}_k^{(u)}(\mathbf{x};\bar{y}) \triangleq \imath_k(\mathbf{x}^t;y^t) + \sum_{i=t+1}^{|\bar{y}|} \log \frac{W_k(y_i|\mathbf{x}_i)}{P^*_{Y_k}(y_i)} \tag{27}$$

where $\imath_k(\mathbf{x}^t;y^t) \triangleq \imath_{P_{\mathbf{x}^t},W_k}(\mathbf{x}^t,y^t)$. Next, we minimize $\Pr[\tau_k \le t|U=u]$ over all stopping times τ_k satisfying (26):

$$\begin{aligned}
&\Pr[\tau_k \le t|U=u] = \mathbb{P}^{(k,u)}_{\bar{Y},\mathbf{X}}\left[|\bar{Y}|=t\right] \\
&= \mathbb{P}^{(k,u)}_{\bar{Y},\mathbf{X}}\left[\tilde{\imath}_k^{(u)}(\mathbf{X};\bar{Y}_k) > \lambda_t, |\bar{Y}|=t\right] \\
&\quad + \mathbb{P}^{(k,u)}_{\bar{Y},\mathbf{X}}\left[\tilde{\imath}_k^{(u)}(\mathbf{X};\bar{Y}_k) \le \lambda_t, |\bar{Y}|=t\right] &(28)\\
&\le \min\left\{1, \mathbb{P}^{(k,u)}_{\bar{Y},\mathbf{X}}\left[\tilde{\imath}_k^{(u)}(\mathbf{X};\bar{Y}_k) > \lambda_t, |\bar{Y}|=t\right] + \varepsilon_{k,M}^{(u)}\right\} &(29)\\
&\le \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda_t\right] \\
&\quad + \min\left\{\varepsilon_{k,M}^{(u)}, 1 - \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda_t\right]\right\}. &(30)
\end{aligned}$$

Here, (29) follows from (26). Since the stopping times τ_1 and τ_2 are conditional independent given $U=u$, (30) implies that

$$\begin{aligned}
&\Pr[\max\{\tau_1,\tau_2\} \le t|U=u] = \prod_{k=1}^{2} \mathbb{P}^{(k,u)}_{\bar{Y},\mathbf{X}}\left[|\bar{Y}_k|=t\right] &(31)\\
&\le \prod_{k=1}^{2} \max_{\mathbf{x}^t\in\mathcal{X}^t} \left\{\Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda_t\right]\right\} \\
&\quad + \min_k \left\{\varepsilon_{\bar{k},M}^{(u)} + \varepsilon_{k,M}^{(u)} \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_{\bar{k}}^t) > \lambda_t\right]\right\}. &(32)
\end{aligned}$$

Note that (32) holds for all τ_k that satisfies (26). Averaging (32) over $u\in\mathcal{U}$ and using the inequality $\sum_{u\in\mathcal{U}} P_U(u)\varepsilon_{k,M}^{(u)} \le \epsilon + (\log M)^{-1} = \varepsilon_M$, we obtain (15). The proof is concluded using (23).

V. Asymptotic Analysis: Converse Bound

We analyze L_t in (15) in the limit $l\to\infty$. By (16),

$$l \ge \sum_{t=0}^{\infty}(1-L_t)^+ \ge \sum_{t=0}^{\lfloor\beta\rfloor}(1-L_t)^+ \ge \sum_{t=0}^{\lfloor\beta\rfloor}(1-L_t) \tag{33}$$

where $\beta > 0$ will be specified shortly. Let $\lambda \triangleq \log M - \log\log M - \delta - (|\mathcal{X}|-1)\log(\beta+1)$. For all $t\le\beta$,

$$\max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda_t\right] \le \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right]. \tag{34}$$

The key step is to establish an asymptotic upper bound on $\max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda]$ for every $t\in\mathbb{Z}_+$ as $\lambda\to\infty$.

Let $\alpha \triangleq \frac{\lambda}{C} - \sqrt{\frac{V\lambda}{C^3}}\log\lambda$ and let β be the solution of

$$(\lambda-\beta C)/\sqrt{\beta V} = -\tilde{Q}^{-1}(\epsilon) \tag{35}$$

where C is given in (5), V is defined in Theorem 3, and $\tilde{Q}^{-1}(\epsilon)$ in (17). We divide the asymptotic analysis of $\max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda]$ into three cases: the "large deviations regime" $t\in[0,\alpha)$, where we use Hoeffding's inequality, the "central regime" $t\in[\alpha,\beta)$, where Berry-Esseen central limit theorem is applied, and the case $t\ge\beta$, where the trivial upper bound $\max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda] \le 1$ suffices.

In the first case, invoking Hoeffding's inequality [13, Th. 2] and using that $I(P_{\mathbf{x}^t}, W_k)$ is upper-bounded by C uniformly, we obtain (see [12, App. II-A])

$$\sum_{t=0}^{\lfloor\alpha\rfloor} \max_{\mathbf{x}^t\in\mathcal{X}^\infty} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right] = o(1), \quad \lambda\to\infty \tag{36}$$

and

$$\sum_{t=0}^{\lfloor\alpha\rfloor} \prod_{k=1}^{2} \max_{\mathbf{x}^t\in\mathcal{X}^t} \left\{\Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right]\right\} = o(1),\ \lambda\to\infty. \tag{37}$$

In the central regime, we use the Berry-Esseen central limit theorem [14, Th. V.3] to show that

$$\Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right] \le Q\left(\frac{\lambda - tI(P_{\mathbf{x}^t},W_k)}{\sqrt{tV(P_{\mathbf{x}^t},W_k)}}\right) + \frac{\mathbb{c}}{\sqrt{t}}. \tag{38}$$

We next maximize (38) over $\mathbf{x}^t \in \mathcal{X}^t$ following the approach in [10, Prop. 8]. Specifically, we use continuity properties of $I(P,W_k)$ and $V(P,W_k)$ for probability distributions $P\in\mathcal{P}(\mathcal{X})$ close to P^* to show that (see [12, App. II-B])

$$\begin{aligned}
&\sum_{t=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right] \\
&\le \sqrt{\frac{V\lambda}{C^3}}\Big(\tilde{Q}^{-1}(\epsilon) - \mathbb{E}\Big[\min\Big\{\tilde{Q}^{-1}(\epsilon), \varrho_k Z_k\Big\}\Big]\Big) + \mathcal{O}(\log\lambda) &(39)
\end{aligned}$$

where ϱ_k are defined in Theorem 3 and $Z_k\sim\mathcal{N}(0,1)$. Similarly, we obtain

$$\begin{aligned}
&\sum_{t=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} \prod_{k=1}^{2} \max_{\mathbf{x}^t\in\mathcal{X}^t} \Pr\left[\imath_k(\mathbf{x}^t;Y_k^t) > \lambda\right] \\
&\le \sqrt{\frac{V\lambda}{C^3}}\left(\tilde{Q}^{-1}(\epsilon) - \mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon), \max_k \varrho_k Z_k\right\}\right]\right) \\
&\quad + \mathcal{O}(\log\lambda). &(40)
\end{aligned}$$

Using (33), (36), (37), (39), and (40), we obtain

$$\begin{aligned}
l &\ge \sum_{t=0}^{\lfloor\beta\rfloor}(1-L_t) &(41)\\
&\ge \frac{\lambda(1-\varepsilon_M)}{C} + \sqrt{\frac{V\lambda}{C^3}}\Bigg(\mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon), \max_k \varrho_k Z_k\right\}\right] \\
&\quad -\varepsilon_M\left(2\tilde{Q}^{-1}(\epsilon) - \min_k \mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon), \varrho_k Z_k\right\}\right]\right)\Bigg) \\
&\quad -\mathcal{O}(\log\lambda) &(42)
\end{aligned}$$

as $\lambda \to \infty$. Finally, we have that

$$\lambda = \log M - \log\log M - \delta - (|\mathcal{X}| - 1)\log(\beta + 1) \quad (43)$$
$$\le \frac{Cl}{1-\varepsilon_M} - \sqrt{\frac{Vl}{(1-\varepsilon_M)^3}}\left(\mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon)\,, \max_k \varrho_k Z_k\right\}\right]\right) - \varepsilon_M\left(2\tilde{Q}^{-1}(\epsilon) - \min_k \mathbb{E}\left[\min\left\{\tilde{Q}^{-1}(\epsilon)\,, \varrho_k Z_k\right\}\right]\right)\right) + \mathcal{O}(\log l) \quad (44)$$

as $l \to \infty$. The final result in (18) is obtained through algebraic manipulations.

VI. Asymptotic Analysis: Achievability Bound

Set $P = P^*$, and fix $r \in \mathbb{N}$, $q = \frac{l'\epsilon - 1}{l'-1}$, and $l' > 0$, a parameter that will be related to the average blocklength. Let the thresholds be chosen as follows:

$$\gamma \triangleq \gamma_k \triangleq C\left(l' - g(Cl')\right). \quad (45)$$

Here,

$$g(x) \triangleq \sqrt{\frac{V_1 + V_2}{2\pi C^2}}\sqrt{\frac{x}{C}} + b_1 x^{\frac{r+1}{4r+2}} \log x \quad (46)$$

where b_1 will be specified later. If we choose a code with a number of codewords $\tilde{M}$ that satisfies

$$\log \tilde{M} \triangleq C\left(l' - g(Cl')\right) - \log l' \quad (47)$$

we have $(\tilde{M} - 1)\exp\{-\gamma\} \le 1/l'$. Furthermore, by Remark 3, the average probability of error is upper-bounded by

$$q + (1-q)(\tilde{M} - 1)\exp\{-\gamma_k\} \le \frac{l'\epsilon - 1}{l' - 1} + \frac{l'(1-\epsilon)}{l'-1}\frac{1}{l'} = \epsilon. \quad (48)$$

Suppose it can be shown that

$$\mathbb{E}[\max\{\tau_1, \tau_2\}] \le l' \quad (49)$$

for sufficiently large l'. Then the average blocklength is

$$(1-q)\mathbb{E}[\max\{\tau_1, \tau_2\}] \le \frac{l'(1-\epsilon)}{l'-1} l' \triangleq l. \quad (50)$$

Consequently, by Theorem 1, there exists an (l, M, ϵ)-VLSF code with

$$\log M \ge \log \tilde{M} \quad (51)$$
$$= C\left(l' - g(Cl')\right) - \log l' \quad (52)$$
$$= \frac{Cl}{1-\epsilon} - \sqrt{\frac{V_1 + V_2}{2\pi(1-\epsilon)}}\sqrt{l} - \mathcal{O}(l^{\frac{r+1}{4r+2}} \log l) \quad (53)$$

where the last step follows because

$$l = \frac{(l')^2(1-\epsilon)}{l'-1} = l'(1-\epsilon) + o(1). \quad (54)$$

To establish (49), we proceed as follows. Let $W_n = \imath_{P,W_1}(X_n; Y_{1,n})$ and $Z_n = \imath_{P,W_2}(X_n; Y_{2,n})$. We can then upper-bound $\mathbb{E}[\max\{\tau_1, \tau_2\}]$ using the following lemma, which is proved using asymptotic results for random walks [4] and a Berry-Esseen-type theorem that holds for random summations (see proof in [12, App. III]).

Lemma 1: Let $\{W_n\}$ and $\{Z_n\}$, $n \ge 1$, be i.i.d. discrete RVs with $(W_1, Z_1) \sim P_{W,Z}$, positive mean $\mu_W \triangleq \mathbb{E}[W_1]$ and $\mu_Z \triangleq \mathbb{E}[Z_1]$, respectively, and finite moments of order $r \ge 3$, i.e., $\mathbb{E}[|W_1|^r] < \infty$, and $\mathbb{E}[|Z_1|^r] < \infty$. Define the random walks $U_n \triangleq \sum_{i=1}^n W_i$ and $V_n \triangleq \sum_{i=1}^n Z_i$, and the stopping times $\tau_1 \triangleq \inf\{n \ge 0 : U_n \ge \gamma\}$ and $\tau_2 \triangleq \inf\{n \ge 0 : V_n \ge \gamma\}$ for every $\gamma \in \mathbb{R}$. Then

$$\mathbb{E}[\max\{\tau_1, \tau_2\}] \le \frac{\gamma}{\min\{\mu_W, \mu_Z\}} + \frac{\sigma}{\sqrt{2\pi}}\sqrt{\frac{\gamma}{\mu_W}}\mathbb{1}\{\mu_W = \mu_Z\} + \mathcal{O}\left(\gamma^{\frac{r+1}{4r+2}} \log \gamma\right) \quad (55)$$

as $\gamma \to \infty$, where $\sigma^2 \triangleq \mathrm{Var}\left[\frac{W_1}{\mu_W} - \frac{Z_1}{\mu_Z}\right]$.

Lemma 1 implies that there exists a constant b_1 such that

$$\mathbb{E}[\max\{\tau_1(\gamma), \tau_2(\gamma)\}] \le \frac{\gamma}{C} + g(\gamma) \quad (56)$$

for sufficiently large γ. The conditional average blocklength of the VLSF code can be bounded as follows

$$\mathbb{E}[\max\{\tau_1, \tau_2\}] = \mathbb{E}[\max\{\tau_1(\gamma), \tau_2(\gamma)\}] \quad (57)$$
$$\le \frac{\gamma}{C} + g(\gamma) \quad (58)$$
$$= l' - g(Cl') + g(Cl' - Cg(Cl')) \le l'. \quad (59)$$

Here, (58) holds by (56), and (59) follows by the definition of γ in (45) and the fact that $g(x)$ is nonnegative and nondecreasing.

References

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

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

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

[4] A. Gut, *Stopped random walks: Limit Theorems and Applications*. New York: Springer, 2009.

[5] D. Landers and L. Rogge, "The exact approximation order in the central-limit-theorem for random summation," *Z. Wahrsch. verw. Gebiete*, vol. 36, pp. 269–283, 1976.

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

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

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems*. New York: Cambridge Univ. Press, 2012.

[9] 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, May 2010.

[10] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov. 2013.

[11] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends in Comm. and Inf. Theory*, vol. 10, no. 4, pp. 1–184, 2014.

[12] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, "Broadcasting a common message with variable-length stop-feedback codes (extended version)," Apr. 2015. [Online]. Available: http://arxiv.org/

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

[14] V. V. Petrov, *Sums of Independent Random Variables*. Berlin: Springer, 1975, translated from the Russian by A. A. Brown.