Bounds on the Effective-length of Optimal Codes for Interference Channel with Feedback

Mohsen Heidari
EECS Department
University of Michigan
Ann Arbor, USA
Email: mohsenhd@umich.edu

Farhad Shirani
Department of Electrical
and Computer Engineering
New York University
New York, New York
Email: fsc265@nyu.edu

S. Sandeep Pradhan
EECS Department
University of Michigan
Ann Arbor, USA
Email: pradhanv@umich.edu

***Abstract*—In this paper, we investigate the necessity of finite blocklength codes in distributed transmission of independent message sets over channels with feedback. We provide two examples of three user interference channels with feedback where codes with asymptotically large effective lengths are sub-optimal. As a result, we conclude that coded transmission using finite effective length codes is necessary to achieve optimality. We argue that the sub-optimal performance of large effective length codes is due to their inefficiency in preserving the correlation between the inputs to the distributed terminals in the communication system. This correlation is made available by the presence of feedback at the terminals and is used as a means for coordination between them when using finite effective length coding strategies.**

I. Introduction

Most of the coding strategies developed in information theory are based on random code ensembles which are constructed using independent and identically distributed (IID) sequences of random variables [1]–[4]. The codes associated with different terminals in the network are mutually independent. Moreover, the blocklengths associated with these codes are asymptotically large. This allows the application of the laws of large numbers and concentration of measure theorems when analyzing the performance of coding strategies; and leads to characterizations of their achievable regions in terms of information quantities that are the functionals of the underlying distribution used to construct the codes. These characterizations are often called single-letter characterizations. Although the original problem is to optimize the performance of codes with asymptotically large blocklengths, the solution is characterized by a functional (such as mutual information) of just one realization of the source or the channel under consideration. It is well-known that unstructured random codes with asymptotically large blocklength can be used to achieve optimality in terms of achievable rates in point-to-point communications. In fact, it can be shown that large blocklength codes are necessary to approach optimal performance. At a high level, this is due to the fact that the efficiency of fundamental tasks of communication such as covering and packing increases as the input dimension is increased [5].

This research was supported in part by NSF grant CCF 1717299.

In network communication, one needs to (a) remove redundancy among correlated information sources [2], [4] in a distributed manner in the source coding problems, and (b) induce redundancy among distributed terminals to facilitate [1], [3] cooperation among them. For example, in the network source coding problems such as distributed source coding and multiple description coding, the objective is to exploit the statistical correlation of the distributed information sources. Similarly, in the network channel coding problems, such as the interference channel and broadcast channel, correlation of information among different terminals is induced for better cooperation among them. At a high level, in addition to the basic objectives of efficient packing and covering at every terminal, the network coding strategies need to exploit statistical correlation among distributed information sources or induce statistical correlation among information accessed by terminals in the network.

Witsenhausen [6] and Gács-Körner [7] made the observation that distributed processing of pairs of sequences of random variables leads to outputs which are less correlated than the original input sequences. In the network communications context, this implies that the outputs of encoding functions at different terminals in a network are less correlated with each other than the original input sequences. In [8], [9], we built upon these observations and showed that the correlation between the outputs of pairs of encoding functions operating over correlated sequences is inversely proportional to the effective length of the encoding functions. Based on these results, it can be concluded that while random unstructured coding strategies with asymptotically large blocklengths are efficient in performing the tasks of covering and packing, they are inefficient in facilitating coordination between different terminals. Using these results, we showed that finite effective codes are necessary to achieve optimality in various setups involving the transmission of correlated sources. Particularly, we showed that the effective length of optimality achieving codes is bounded from above in the distributed source coding problem as well as the problem of transmission of correlated sources over the multiple access channel (MAC) and the interference channel (IC) [8], [10].

So far, all of the results showing the necessity of finite

978-1-5386-4780-6/18/$31.00©2018 IEEE

effective length codes pertain to situations involving the distributed transmission of sources over channels and distributed compression of sources. However, the question of whether such codes are necessary in multi-terminal channel coding has remained open. The reason is that the application of the results in [8], [9] requires the presence of correlated inputs in different terminals of the network. In the case of distributed processing of sources, such correlation is readily available in the form of the distributed source. Whereas, in distributed transmission of independent messages it is unclear how such a correlation can be created and exploited. In this work, we argue that in channel coding with feedback, correlation is induced because of the feedback link. More precisely, the feedback sequence at one terminal is correlated with the message in the other terminal. In order to exploit this correlation efficiently, finite effective length codes are necessary. The contributions of this paper can be summarized as follows. We provide two examples of interference channels with feedback where finite effective length codes are necessary to approach optimality. For each of these examples, we provide an outer bound on the achievable region as a function of the effective-length of the encoding functions used at the transmitters. Furthermore, we use finite effective length codes to prove the achievability of certain rate vectors which lie outside of the outer bound when the effective length is large. The combination of these two results shows that in these examples any coding strategy which uses encoding functions with asymptotically large effective lengths is sub-optimal.

The rest of the paper is organized as follows: In Section II, we introduce the problem formulation. Section III provides the prior results which are used in this paper. Section IV explains our main results. Finally, section V concludes the paper.

II. System Model

A. Notations

Calligraphic letters such as $\mathcal{C}$ and $\mathcal{M}$ are used to represent sets. Random variables are denoted using capital letters such as X, Y. The random vector $(X_1, X_2, ..., X_n)$ is represented by X^n. For shorthand, vectors are sometimes represented using underline letters without any superscript such as $\underline{X}$, and $\underline{f}$. We denote the set $\{1, 2, \ldots, m\}$ by $[1, m]$, where m is an integer. The Hamming weight of a vector $\underline{x}$ is denoted by $w_H(\underline{x})$. For any vector $\mathbf{i} \in \{0,1\}^n$, let $Z_{\mathbf{i}} = \{Z_j : i_j = 1\}$; for example, take $n = 3$, then Z_{101} represents (Z_1, Z_3).

B. Model

The problem of Interference Channel with Feedback (IC-FB) has been studied extensively [11], [12]. A three-user interference channel with generalized feedback is characterized by three input alphabets $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$, three output alphabets $(\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3)$, three feedback alphabets $(\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3)$, and transition probability distributions $(Q_{\underline{Y}|\underline{X}}, P_{\underline{Z}|\underline{Y}})$. We assume that all the alphabets are finite and that the channel is memoryless. Let $x_i^n, y_i^n, z_i^n, i \in [1,3]$, be the channel inputs, outputs and the channel feedback after n uses of the channel, respectively. The memoryless property implies that

$$P\Big\{y_{j,n}, z_{j,n}\ ,\ j \in [1,3] \ \Big|\ y_i^{n-1}, z_i^{n-1}, x_i^n, i \in [1,3]\Big\} = Q_{\underline{Y}|\underline{X}}(y_{1,n}, y_{2,n}, y_{3,n}|x_{1,n}, x_{2,n}, x_{3,n}) \times P_{\underline{Z}|\underline{Y}}(z_{1,n}, z_{2,n}, z_{3,n}|y_{1,n}, y_{2,n}, y_{3,n}).$$

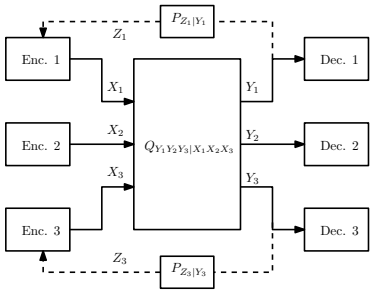

Fig. 1: An instance of the three-user IC with generalized feedback. Here transmitters 1 and 3 receive noisy feedback, whereas transmitter 2 does not receive feedback.

In this setup, there are three transmitters and three receivers. The ith transmitter, $i \in [1,3]$, intends to transmit a message index W_i to the ith receiver. The channel's feedback $Z_i, i \in [1,3]$ is causally available at transmitter i with one unit of delay. An example of such a setup is depicted in Figure 1. In this figure, Z_2 is a trivial random variable (i.e. the second transmitter does not receive any feedback) and $P_{Z_1,Z_3|Y_1,Y_2,Y_3} = P_{Z_1|Y_1} P_{Z_3|Y_3}$.

In what follows, we provide formal definitions for achievability and coding scheme. Let M_1, M_2, M_3 and N be arbitrary positive integers.

Definition 1. *An (M_1, M_2, M_3, N) feedback-block-code for the three user IC-FB consists of*

- *Three sets of messages $\mathcal{M}_i = \{1, 2, ..., M_i\}, i \in [1,3]$.*
- *Three sequences of encoding functions*

$$f_{i,n} : \mathcal{M}_i \times \mathcal{Z}_i^{n-1} \to \mathcal{X}_i, \quad 1 \le n \le N,$$

- *Three decoding functions:*

$$g_i : \mathcal{Y}_i^N \to \mathcal{M}_i, \quad i \in [1,3].$$

The message for transmitter i is denoted by a random variable W_i which takes values from the message set $\mathcal{M}_i, i \in [1,3]$ with uniform distribution. Furthermore, it is assumed that the messages $W_i, i \in [1,3]$ are mutually independent. The output of the ith transmitter at the nth use of the channel is denoted by $X_{i,n} = f_{i,n}(W_i, Z_i^{n-1})$. The rate-triple of a (M_1, M_2, M_3, N) code is defined as $R_i = \frac{\log_2 M_i}{N}, i \in [1,3]$. The probability of the error is defined as

$$P_e \triangleq P\Big\{(W_1, W_2, W_3) \neq (g_1(Y_1^n), g_2(Y_2^n), g_3(Y_3^n))\Big\}.$$

One can consider, a randomized coding strategy for which the encoding functions are selected randomly according to a probability measure defined over the set of all encoding

functions $f_i^N, i \in [1,3]$ as in Definition 1. The following defines a randomized coding strategy.

Definition 2. *An (M_1, M_2, M_3, N)-randomized coding strategy is characterized by a probability measure P_N on the set of all functions $(f_i^N), i \in [1,3]$ described in Definition 1.*

Definition 3. *For $\epsilon > 0$, a rate-triple (R_1, R_2, R_3) is said to be ϵ-achievable by a feedback-block-code with parameters (M_1, M_2, M_3, N), if the following conditions are satisfied*

$$P_e \leq \epsilon, \qquad \frac{1}{N} \log_2 M_i \geq R_i - \epsilon, \ i \in [1,3].$$

Definition 4. *For $\epsilon > 0$, a rate-triple (R_1, R_2, R_3) is said to be ϵ-achievable by a (M_1, M_2, M_3, N)-randomized coding strategy with probability measure P_N, if, with probability one with respect to P_N, there exists a feedback-block-code for which (R_1, R_2, R_3) is ϵ-achievable.*

Definition 5. *For $\epsilon > 0$, a rate-triple (R_1, R_2, R_3) is said to be ϵ-achievable, if there exist a (M_1, M_2, M_3, N) feedback-block-code (randomized coding strategy) for which (R_1, R_2, R_3) is ϵ-achievable.*

Definition 6. *A rate-triple (R_1, R_2, R_3) is said to be achievable, if it is ϵ-achievable for any $\epsilon > 0$. Given an IC-FB, the set of all achievable rate-triples is called the feedback-capacity.*

III. Preliminaries

We investigate the case in which the input of the channel is binary. For such a setup, at any time instance $n+1$, each encoder is modeled as a Boolean function $e : \mathcal{Z}^n \to \{0,1\}$, where $\mathcal{Z}$ is the input. In this section, we summarize the results in [8] and [9] on the correlation between the outputs of Boolean functions of pairs of sequences of random variables. These results are used in the next section to prove the necessity of finite effective length codes.

Suppose Z^n is a sequence of IID random variables each taking values from $\mathcal{Z}$. Let $P\{e(Z^n) = 1\} = q$. For the pair of e and q, define the real-valued analogue of e as

$$\tilde{e}(\mathbf{z}^n) = \begin{cases} 1-q & \text{if } e(\mathbf{z}^n) = 1, \\ -q & \text{otherwise.} \end{cases} \tag{1}$$

With this definition, $\mathbb{E}[\tilde{e}(Z^n)] = 0$ and $Var(\tilde{e}(Z^n)) = q(1-q)$. For $\tilde{e}$ and Z^n define the following function

$$\tilde{e}_{\mathbf{i}} = \mathbb{E}_{Z^n|Z_{\mathbf{i}}}[\tilde{e}(Z^n)|Z_{\mathbf{i}}] - \sum_{\mathbf{j} \in \{0,1\}^n:\ \mathbf{j} < \mathbf{i}} \tilde{e}_{\mathbf{j}}, \tag{2}$$

where the condition $\mathbf{j} < \mathbf{i}$ means $j_k < i_k, \forall k \in [1,n]$. Note that $\tilde{e}_{\mathbf{i}}$ is understood as the component of $\tilde{e}$ which is only a function of $Z_{\mathbf{i}}$. If $\mathbf{i}$ is such that $w_H(\mathbf{i}) = k$, then $\tilde{e}_{\mathbf{i}}$ is called a k-letter component of $\tilde{e}$.

Definition 7. *For a Boolean function e, the collection $(\tilde{e}_{\mathbf{i}})_{\mathbf{i} \in \{0,1\}^n}$ is called the real value decomposition of e.*

For each function $\tilde{e}_{\mathbf{i}}$, define $\mathbf{P}_{\mathbf{i}} = Var(\tilde{e}_{\mathbf{i}}(Z^n))$. Then, we have the following definition.

Definition 8. *For a Boolean function e with the real value decomposition $(\tilde{e}_{\mathbf{i}})_{\mathbf{i} \in \{0,1\}^n}$, the vector of all variances $(\mathbf{P}_{\mathbf{i}})_{\mathbf{i} \in \{0,1\}^n}$ is called the dependency spectrum of e.*

The following Lemma provides a lower bound on the correlation between the output of two Boolean functions of a pair of random sequences. The result of this Lemma is used in the next Section.

Lemma 1. *Let (Z^n, Y^n) denote a pair of sequences of IID random variables each with joint distribution P_{ZY} and with values in $\mathcal{Z} \times \mathcal{Y}$. Let $e : \mathcal{X}^n \to \{0,1\}$, and $f : \mathcal{Y}^n \to \{0,1\}$ be two Boolean functions with dependency spectrum $(\mathbf{P}_{\mathbf{i}})$, and $(\mathbf{Q}_{\mathbf{i}})$, respectively. Then, the following bound holds*

$$P\{e(X^n) \neq f(Y^n)\} \geq 2\Big[\sum_{\mathbf{i},\mathbf{j}} \mathbf{P}_{\mathbf{i}}\mathbf{Q}_{\mathbf{j}}\Big]^{\frac{1}{2}} - 2\sum_{\mathbf{i}} \rho^{w_H(\mathbf{i})} \mathbf{P}_{\mathbf{i}}^{\frac{1}{2}} \mathbf{Q}_{\mathbf{i}}^{\frac{1}{2}},$$

where ρ is defined as $\rho \triangleq \sup_{g,h} \mathbb{E}[g(Z)h(Y)]$, the supremum is taken over all functions $g : \mathcal{Z} \to \mathbb{R}$, and $h : \mathcal{Y} \to \mathbb{R}$ such that $g(Z)$ and $h(Y)$ have unit variance and zero mean.

Remark 1. Note that the term $\rho^{w_H(\mathbf{i})}$ is decreasing in w_H. Therefore, $P\{e(X^n) \neq f(Y^n)\}$ is minimized when most of the variance $\mathbf{P}_{\mathbf{i}}$ is distributed on $\mathbf{i}$ with smaller Hamming weight. This implies that Boolean functions with smaller effective-lengths can have higher correlation between their outputs.

IV. Main Results

In this section, we introduce two examples of three user IC-FBs where finite effective length codes are necessary to approach optimality.

Example 1. Consider the setup shown in Figure 2. Here, $(X_{11}, X_{12}), X_2, (X_{32}, X_{33})$ are the outputs of Encoder 1, 2, and 3, respectively. The channel outputs Y_1, (Y_2, Y_2'), and Y_3 are received at Decoders 1, 2, and 3, respectively. The channel corresponding to the transition probability $P_{Y_2'|X_{12}X_{32}}$ is described by the following relation:

$$Y_2' = X_{12} + N_\delta + (X_{12} \oplus X_{32}) \wedge E,$$

where N_δ and E are independent Bernoulli random variables with $P(N_\delta = 1) = \delta$ and $P(E = 1) = \frac{1}{2}$. Also, the random variables N_ϵ, and N_p in Figure 2 are Bernoulli random variables with $P(N_\epsilon = 1) = \epsilon$ and $P(N_p = 1) = p$, respectively. The variables N_δ, E, N_ϵ and N_P are mutually independent. In this setup, feedback is only available at encoder 1 and 3. The feedback at the first transmitter is $Z_1 = Y_1$ with probability one. The feedback at the third transmitter is $Z_3 = Y_3$ with probability one.

The following Theorem provides an outer bound on the achievable region of any coding strategy with blocklength n.

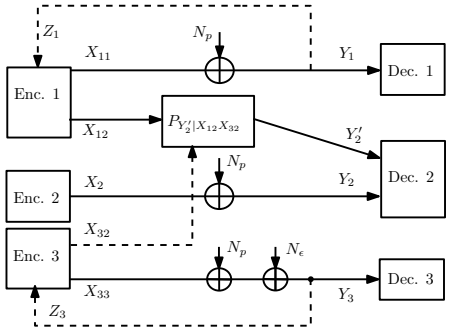

Fig. 2: The diagram of the IC-FB given in Example 1. In this setup, Z_1 is the feedback at Transmitter 1, and Z_3 is the feedback at transmitter 3.

Theorem 1. *For any (M_1, M_2, M_3, n)-randomized coding strategy achieving (R_1, R_2, R_3), the following inequalities are satisfied:*

$$R_1 \le 1 - h_b(p),$$
$$R_2 \le 1 - \Big|h_b(p) - (1 - h_b(\delta))(\frac{1}{n}\sum_{i=1}^{n} P\{X_{12,i} = X_{32,i}\})\Big|^+,$$
$$R_3 \le 1 - h_b(p * \epsilon),$$

where $p * \epsilon = p(1-\epsilon) + (1-p)\epsilon$, and h_b is the binary entropy function.

Proof. The proof is given in Appendix A. □

Corollary 1. *Define the set $\mathcal{R}^*$ as the union of all rate-triples (R_1, R_2, R_3) such that*

$$R_1 \le 1 - h_b(p),$$
$$R_2 \le 1 - |h_b(p) - (1 - h_b(\delta))|^+,$$
$$R_3 \le 1 - h_b(p * \epsilon).$$

Then, the feedback-capacity of the channel in Example 1 is contained in $\mathcal{R}^*$.

Corollary 2. *For the channel in Example 1, assume $\epsilon = 0$, and p, δ are such that $h(p) \le 1 - h(\delta)$. Then, the feedback-capacity of the channel is characterized by the following*

$$R_1 \le 1 - h(p), \quad R_2 \le 1, \quad R_3 \le 1 - h(p).$$

Lemma 2. *Let C_ϵ denote the feedback-capacity region (as a function of the parameter ϵ) of the IC-FB in Example 1 . For any $(R_1, R_2, R_3) \in C_0$, there exists a continuous function $\zeta(\epsilon)$ such that, for sufficiently small $\epsilon > 0$, $(R_1 - \zeta(\epsilon), R_2 - \zeta(\epsilon), R_3 - \zeta(\epsilon)) \in C_\epsilon$, where $\zeta(\epsilon) \to 0$, as $\epsilon \to 0$.*

The proof of Lemma 2 and Corollary 2 is given in [14] .

Theorem 2. *Let $\delta = p$. There exist $\gamma > 0$ and $\epsilon > 0$, such that for any coding strategy achieving the rate-triple $(1 - h_b(p), 1 - \gamma, 1 - h_b(p))$ the effective length of the encoding functions producing X_{12} and X_{32} are bounded from above by a constant. Furthermore, the effective length is greater than 1 (i.e. uncoded transmission is not optimal).*

proof outline. From Theorem 1, and the assumption that $(1 - h_b(p), 1 - \gamma, 1 - h_b(p))$ is achievable we obtain

$$R_2 = 1 - \gamma \le 1 - h_b(p)(1 - \frac{1}{n}\sum_{i=1}^{n} P\{X_{12,i} = X_{32,i}\}).$$

This implies, for small enough γ, that $\frac{1}{N}\sum_{i=1}^{N} P(X_{12,i} = X_{32,i}) \approx 1$. Hence, for almost all $i \in [1, n]$, $P(X_{12,i} = X_{32,i}) \approx 1$. Therefore, by Lemma 1, this requires that the effective length be bounded from above. If the effective length is equal to 1, then $\mathbf{P}_{\mathbf{i}_n} \approx 1$ for all $n \in \mathbb{N}$. As a result, with probability $\mathsf{P}_N \approx 1$, the first encoding function satisfies $F_{12,i}(m_1, z_1^{i-1}) \approx \tilde{F}_{12,i}(N_{p,i-1})$. Thus, $P(Y'_{2,n} = N_p + N_\delta) \approx 1$. However,

$$\frac{1}{n}H(Y_2^n | X_2^n, Y_2'^n) \approx \frac{1}{n}H(N_p^n | \tilde{F}_{12}^n(N_p^n) + N_\delta^n)$$
$$\le \frac{1}{n}H(N_p^n | N_p^n + N_\delta^n) \approx (2h_b(p) - h_b(p * p))$$

As a result, from Fano's inequality as in (3), $R_2 \lesssim 1 + h_b(p*p) - 2h_b(p) < 1$. This inequality contradicts with the assumption $R_2 = 1 - \gamma$, for sufficiently small γ. In other words, using uncoded transmission, it is not possible to reconstruct N_p at Decoder 2.

□

Example 2. Consider the IC shown in Figure 3. In this setup, (X_{11}, X_{12}), X_2, and $(X_{32}, X_{33}, X'_{33})$ denote the outputs of Encoder 1, 2, and 3, respectively. In addition, Z_1 and Z_2 represent the feedback available at Encoder 1 and Encoder 3, respectively. the inputs and outputs of the channel are taken to be binary; except Y_1 which is ternary. The noise random variables N_1, $N_3, N_\delta, N_\epsilon$ and E are mutually independent Bernoulli random variables with parameter $p_1, p_3, \delta, \epsilon$, and $1/2$, respectively. Finally, it is assumed that $p_1, p_3, \delta, \epsilon < 1/2$.

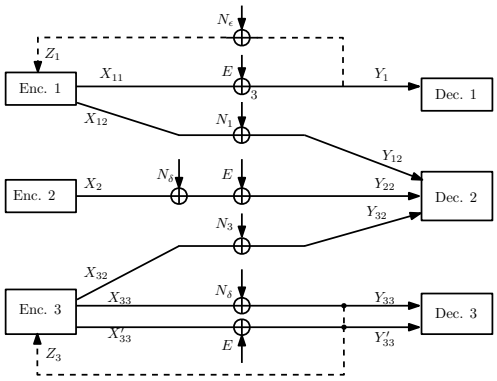

Fig. 3: The diagram of the IC-FB in Example 2. In this setup, Z_1, the feedback at Transmitter 1, is a noisy version of Y_1.

In Lemma below, we provide an achievable rate for $\epsilon = 0$.

Lemma 3. *For $\epsilon = 0$, the rate-triple $(\log_2 3 - 1, 1 - h_b(d), 1 - h_b(\delta))$ is achievable, where $d = h_b^{-1}(|h_b(p_1 * \delta) + h(p_3) - 1|^+)$.*

proof outline. Achievability for R_1, R_3 follows from the standard arguments as in point-to-point channel coding. Next,

we show $R_2 = 1 - h_b(d)$ is achievable. Upon receiving the feedback, Encoder 1 and 3 recover E and (E, N_δ), respectively. Encoderss 1 and 3 employ a source-channel coding scheme to transmite the sources E, N_δ such that Decoder 2 reconstructs $E + N_\delta$ within a Hamming distortion d. This problem is similar to the Common-Bit One-Help-One Problem introduced in [13]. Using Theorem 3 in [13], we can show that Decoder 2 is able to reconstruct $E + N_\delta$ within Hamming distortion d, if the bounds $R_{32} \geq h_b(\Delta * \delta) - h_b(d)$, and $R_{12} \geq 1 - h_b(\Delta)$ hold for some $0 \leq \Delta \leq 1/2$. Suppose that the transmitted codewords from Encoders 1 and 3 are decoded at Decoder 2 with small probability of error. From standard channel coding arguments, the inequalities $R_{12} \leq 1 - h_b(p_1) - \zeta$, and $R_{32} \leq 1 - h_b(p_3) - \zeta$ hold, where $\zeta > 0$ is arbitrarily small. The proof follows by setting $\Delta \approx p_1$, and d as in the statement of the Lemma. □

For the case when $\epsilon > 0$, there is no common information between Encoders 1 and 3. From the discontinuity argument as in [13], we can show that the minimum distortion level d is discontinuous as a function of ϵ. This implies that the achievable rates using single letter coding schemes decreases discontinuously comparing to the case when $\epsilon = 0$. Using this argument, we establish the following Lemma.

Lemma 4. *There exist $\gamma > 0$ and $\epsilon > 0$, such that for any coding strategy achieving the rate-triple $(\log_3 -1, 1 - h_b(d) - \gamma, 1 - h_b(\delta))$ the effective length of the encoding functions producing X_{12} and X_{32} are bounded from above by a constant. Furthermore, the effective length is greater than 1 (i.e. uncoded transmission is not optimal).*

The proof follows by a similar argument as in Theorem 2.

V. Conclusion

We provided two examples of channel coding with feedback over interference networks where finite effective length coding is necessary to achieve optimal performance. We showed that in these examples, optimality achieving coding strategies utilize the feedback available in different terminals to coordinate their outputs. We showed that coding strategies with asymptotically large effective lengths are inefficient in preserving the correlation among their outputs and are hence unable to coordinate their inputs to the channel effectively.

Appendix A
Proof of Theorem 1

Proof. The bounds $R_1 \leq 1 - h_b(p)$ and $R_3 \leq 1 - h_b(p)$ follow from standard arguments as in point-to-point channel coding problem. Next, we apply Fano's inequality as follows

$$\begin{aligned} nR_2 \leq H(W_2) &\overset{(a)}{=} H(W_2|Y_2'^n) \overset{(b)}{\leq} I(W_2; Y_2^n|Y_2'^n) + n\zeta_n \\ &= H(Y_2^n|Y_2'^n) - H(Y_2^n|W_2, Y_2'^n) + n\zeta_n \\ &\overset{(c)}{=} H(Y_2^n|Y_2'^n) - H(Y_2^n|W_2, X_2^n, Y_2'^n) + n\zeta_n \\ &\overset{(d)}{\leq} n - H(Y_2^n|W_2, X_2^n, Y_2'^n) + n\zeta_n \\ &\overset{(e)}{=} n - H(N_p^n|Y_2'^n) + n\zeta_n, \end{aligned} \tag{3}$$

where (a) and (e) hold since $X_{12}^n, X_{32}^n, Y_2'^n$ are independent of W_2. The equality at (c) holds, because the second transmitter does not receive feedback and, hence, X_2^n is a function of W_2. (b) follows from Fano's inequality and (d) follows from the fact that Y_2 is binary. Let $Z_i, i \in [1, n]$ be the indicator function of the event $\{X_{12,i} = X_{32,i}\}$. Then,

$$\begin{aligned} H(N_p^n|Y_2'^n) &\geq H(N_p^n|Y_2'^n, Z^n) \\ &= \sum_{\underline{z} \in \{0,1\}^n} p(Z^n = \underline{z}) H(N_p^n|Y_2'^n, \underline{z}). \end{aligned}$$

Next, we have:

$$\begin{aligned} H(N_p^n|Y_2'^n, \underline{z}) &\overset{(a)}{=} H(N_p^n|X_{12}^{\underline{z}} \oplus N_\delta^{\underline{z}}) \\ &= H(N_p^n, X_{12}^{\underline{z}} \oplus N_\delta^{\underline{z}}) - H(X_{12}^{\underline{z}} \oplus N_\delta^{\underline{z}}) \\ &\overset{(b)}{\geq} H(N_p^n, X_{12}^{\underline{z}} \oplus N_\delta^{\underline{z}}) - w_H(\underline{z}) \\ &\overset{(c)}{\geq} H(N_p^n, N_\delta^{\underline{z}}) - w_H(\underline{z}) \\ &= H(N_p^n) + H(N_\delta^{\underline{z}}) - w_H(\underline{z}) \\ &= nh_b(p) - w_H(\underline{z})(1 - h_b(\delta)), \end{aligned} \tag{4}$$

where (a) follows from the definition of Y_2', (b) follows from the fact that the binary entropy is upper bounded by one and (c) is true as X_{12} is independent of N_δ. Finally, the proof is completed by combining equations (3) and (4). □

References

[1] Te. Han and K. Kobayashi, "A new achievable rate region for the interference channel," IEEE Trans. on Inf. Theory, 27(1):49–60, 1981.

[2] S. Y. Tung, "Multiterminal Source Coding," PhD thesis, Cornell University, Ithaca, NY, 1978.

[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," IEEE Trans. on Inf. Theory, 25(3):306–311, May 1979.

[4] Z. Zhang and T. Berger, "New results in binary multiple descriptions," IEEE Trans. on Inf. Theory, 33(4):502–521, Jul 1987.

[5] I. Csiszár and J. Körner, "Information Theory: Coding Theorems for Discrete Memoryless Systems," Academic Press Inc. Ltd., 1981.

[6] H. S. Witsenhausen, "On sequences of pair of dependent random variables," SIAM Journal of Applied Mathematics, 28(1):100–113, 1975.

[7] P. Gacs and J. Körner, "Common information is far less than mutual information," Problems of Control and Inf. Theory, 2(2):119–162, 1972.

[8] F. Shirani and S. Pradhan, "On the sub-optimality of single-letter coding in multi-terminal communications," arXiv:1702.01376, 2017.

[9] F. Shirani and S. Pradhan, "On the correlation between boolean functions of sequences of random variables," IEEE International Symposium on Inf. Theory (ISIT), pages 1301–1305, 2017.

[10] F. Shirani and S. S. Pradhan,"Finite block-length gains in distributed source coding," IEEE International Symposium on Inf. Theory (ISIT), pages 1702–1706, 2014.

[11] S. Yang and D. Tuninetti, "Interference channel with generalized feedback (a.k.a. with source cooperation): part I: achievable region," IEEE Trans. on Inf. Theory, vol. 57, no. 5, pp. 2686-2710, May 2011.

[12] G. Kramer, "Feedback strategies for white Gaussian interference networks," IIEEE Trans. Inf. Theory, vol. 48, no. 6, pp. 1423âĂŞ1438, 2002.

[13] A. B. Wagner, B. G. Kelly and Y. Altug, "Distributed Rate-Distortion With Common Components," IEEE Trans. on Inf. Theory, vol. 57, no. 7, pp. 4035-4057, July 2011.

[14] M. Heidari, F. Shirani, and S. Pradhan, "Bounds on the Effective-length of Optimal Codes for Interference Channel with Feedback," arXiv:1801.05294, 2018.